ECE 631 - Microelectronic Processing Technology

University of Waterloo

Department of Electrical and Computer Engineering

3D STACKING TECHNOLOGY

Tirth Halvadia (ID: 21100463)

Date: April 1, 2024

Abstract

This project explores 3D stacking technology, highlighting its space-saving, speed-enhancing, and power-efficient features. It explains briefly how chips are made by stacking thin layers on top of each other and talks about four different stacking ways. It also talks about specific circuits and devices that work well with 3D stacking and how tiny tunnels called Through-Silicon Vias (TSVs) are integral part of the 3D stacking. Big companies like Intel, Cadence, and IBM are working together to make 3D stacking even better, which could change how we make computers and gadgets in the future.

1 Introduction

The semiconductor industry's unwavering pursuit of realizing Moore's Law scaling has allowed it to experience unparalleled growth over the past 60 years. The semiconductor industry benefited from Moore's law as feature sizes shrank and the cost per transistor decreased. However, transistor shrinkage has its limits, and this approach cannot accommodate additional integration and due to that the industry is shifting to design-technology and system-technology co-optimization (STCO) paradigms, where added value is achieved through heterogeneous integration of different technologies targeted towards specific end-applications. An example of a technology that offers a way to surpass these limitations is three-dimensional (3D) integrated circuit technology.

The concept of 3-dimensional integration is not entirely novel. In fact, expanding Integrated Circuit (IC) designs which are uniform in 2D format to a stacked 3D configuration is relatively straightforward. For instance, CMOS image sensors, DRAM

memories , and NAND Flash memories have all utilized the 3D integration technology [1].

The semiconductor industry has recently created many three-dimensional (3D) chip stacking techniques in response to the needs of steadily expanding functionality and mobility in electronic device microelectronic packaging. High bandwidth, low latency, low power dissipation, and tiny form factor are benefitting that 3D chip stacking integration may offer for improving the performance of integrated circuit systems and many applications in electronic devices. Furthermore, 3D IC technology is a very useful method for highly integrated parts with a range of heterogeneous electronics. In light of this, 3D chip stacking technology is a viable packaging technique for high-performance consumer electronics. To achieve the objective of boosting compute density, these three-dimensional integration techniques aid in the vertical integration of many layers of active transistors and interconnects.

Although this technique appears promising, it has some disadvantages, including thermals (hard to cool bottom layers), complicated package layouts, and higher packaging costs. But the need to go into three dimensions comes from all of its benefits, which, aside from the ones mentioned above, outweighs the disadvantages and exceed those provided by existing two-dimensional technology. Some of these benefits includes:

- 3D integration facilitates a reduction in interconnect delay and power consumption by bringing functional blocks into closer proximity.
- It conserves space by enabling the partitioning of large dies into smaller ones, when necessary, thus lowering costs and enhancing yield.
- 3D technology can provide extra communication channels in third dimension.

Considering such advantages of this technology, certain circuits or devices that

are highlighted in [2] as particularly advantageous if developed using this technology are as follows:

- **Sensor systems:** Sensor systems were among the earliest 3D chips to be demonstrated, primarily because of the growing demand for heterogeneous integration and parallel processing. For instance, in imaging systems, the process involves detecting light using a sensor array on the top layer, which then converts it into electrical signals. These signals are subsequently amplified, if necessary, and converted into digital format in the middle layers. Finally, they are sent to digital layers for further processing. Parallel processing algorithms can be effectively implemented due to the availability of more communication channels in third dimension. This enables efficient processing of data in parallel, enhancing overall system performance and throughput.

- **Processor-Memory Systems:** Processor-memory systems consistently require greater numbers of processing elements and larger storage capacities. 3D stacking offers a solution by not only allowing for more logic and memory devices to be accommodated in the third dimension but also facilitating heterogeneous integration. This means that within a single stacked structure, various types of components with different functionalities can be integrated together. For example, processing elements such as CPUs and GPUs can be stacked alongside different types of memory modules like DRAM and NAND Flash. This heterogeneous integration enables more efficient communication between different components, leading to improved performance and functionality in processor-memory systems.

- **Network-on-chip:** The advantage of implementing a 3D-NoC (Three-Dimensional Network-on-Chip) is evident in its ability to reduce average hop counts for pack-

ets compared to their counterparts in a 2D-NoC (Two-Dimensional Network-on-Chip). This reduction in hop counts translates to lower average latency in 3D-NoCs. This efficiency improvement arises from the vertical integration of routing nodes, enabling more direct pathways and shorter communication distances between components within the network.

- **Field-Programmable Gate Arrays (FPGAs):** Transitioning to a 3D design can decrease the length of interconnects and conserve power, even with minimal alterations to the architecture. Additionally, with the availability of emerging memory devices, memory cel ls can be organized more densely. This novel 3D-FPGA architecture has the potential to further diminish the chip area and interconnect length, rendering it well-suited for 3D technology.

2 Scientific/Technical Background

Three-dimensional (3D) technology has two objectives: to establish vertical integration and to enhance the functionality and performance of electronic systems and devices. Four forms of 3D technology that are employed to help in achieving these objectives includes the following [2] :

1. **Package Stacking:** Package stacking is an established method where chips undergo individual packaging before being stacked. The spacing between 3D interconnects typically falls within the range of hundreds of micrometers, thereby resulting in a limited number of available interconnects. Despite this limitation, package stacking offers advantages in terms of cost-effectiveness and reduced design time.

2. **Die Stacking:** In contrast, die stacking uses Through silicon Vias (TSVs) and

micro bumps to connect dies. It allows for the integration of diverse functionality and high-density memory architectures by directly stacking individual semiconductor dies.

3. **Wafer stacking:** Wafer stacking uses same process as die stacking for joining processed wafers before slicing them into different dies, this technique enables the integration of several functional layers onto a single die and the creation of creative packaging solutions.

4. **Device stacking:** This is more complex and aggressive technique giving benefits in more communication channels between active layers which makes it very beneficial when it comes to parallel processing and high-density storage.

Furthermore, in order to effectively profit from 3D stacking, one must innovate in architecture design and learn from and improve upon current 2D and 2.5D technologies. Since architecture design is the first step in the 3D integration process, it is extremely important. Without a doubt, the TSVs are the most integral and significant component of the architectural design that affects 3D stacking since they control the communication bandwidth between stacked layers based on the density and operating frequency of the vias. The presence of parasitic resistance and capacitance in the TSVs are also a great deal of importance when it comes to efficiency of the chip. According to the research paper [2] reducing parasitic resistance and increasing channel yield may be achieved by switching out a single via for many vias. Also, there are trade-offs since additional chip space will be required for many vias. Therefore, TSVs can be considered as a limited resource.

Having discussed through-silicon vias (TSVs), the other major type of wafer bonding involves Cu-Cu connections. Cu-Cu connections entail copper-to-copper bonding between stacked ICs, often used in TSVs. The Cu-Cu non-thermo compression

bonding, also known as direct bonding, requires very flat surfaces with minimal surface irregularities between the copper (Cu) pads and the surrounding oxide. This method of wafer-to-wafer non-thermo compression bonding offers high throughput, allowing for processing rates of up to 14 wafers per hour. Its sub-micrometer precision alignment capability enables efficient high-volume manufacturing. Additionally, Cu-Cu non-thermo compression bonding results in very low contact resistivity, facilitating effective electrical conduction through the Cu-Cu bonding interface, particularly after low-temperature post-bond annealing. The data presented in this paper [3] supports the Cu-Cu non-thermo compression bonding technique as a promising option for wafer stacking technology in 3D integration applications.

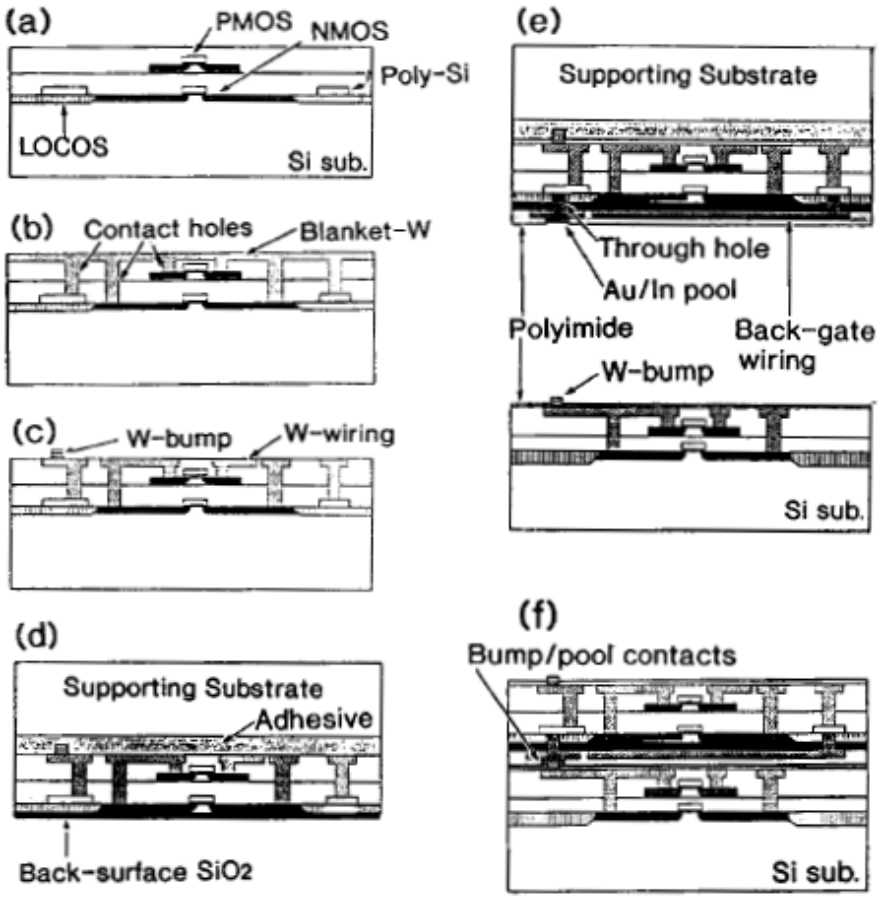

Figure 1: Sequence of the key process steps [4]

Additionally, digging a bit deeper to know about the architectural as well as struc-

tural design and some fabrication techniques let's consider an example of 3D-IC where stacking of thin film dual-cmos layers is discussed briefly[4]. Here, on the Si substrate, nMOSFETs are first constructed, and on the SOI crystal that is created by laser beam annealing, pMOSFETs are then fabricated as shown in the figure 1. Followed by this is the process of connecting contact holes to electrodes of both the MOSFETs, after which Ti/TiW and tungsten(W) is deposited using sputtering and blanket W-CVD techniques respectively. These layers are patterned to form DUAL-CMOS function blocks, and the surface is smoothed using an "etch-back" technique. Tungsten bumps are then created on the Ti/TiW/W wirings through a selective W-CVD technique. The IC substrate is then attached to a supporting substrate and thinned down to the back surface using mechano-chemical polishing.

A passivation film of silicon dioxide (SiO2) is deposited on the back surface at a low temperature, providing insulation and protection. Through holes and back gate wirings are created, followed by the fabrication of gold/indium pools for connection purposes. Using infrared microscopy, the thin film block is aligned with another DUAL-CMOS IC substrate and bonded together at high temperature. Afterward, the supporting substrate and adhesive are removed to obtain a 3D-IC structure. This 3D-IC contains multiple function blocks stacked on top of each other, isolated by a polyimide inner layer, and interconnected through the bump/pool contacts. Input/output (I/O) pads for power supply, ground, and signals are placed on the upper layer.

Overall, this fabrication process enables the creation of high-density DUAL-CMOS function blocks stacked together efficiently, resulting in a short process turnaround time.

3 Current Challenges

The 3D stacking technology does look promising, however it does come with many challenges with it, for instance since the fabrication methods are quite different than the traditional methods used in the current 2D and 2.5 technology re-designing of the structural design is often difficult.Similarly, one more significant challenge addressed in [5], when it comes to advanced 3D chip stacking, is the requirement of high temperature (more than 250℃) when making high density interconnections. The utilization of high temperatures poses a risk of device function failure due to potential chip warpage and damage to the chip's interconnects. However, to overcome this issue anisotropic conductive film (ACF) is used as a transitional layer for interconnection and to fill the gap between substrate and the chip during the bonding process. The use of ACF is highly beneficial compared to conventional method of solder joining process, since it helps in reducing the temperature to 200℃. Moreover, this process also reduces the number of steps required.

Considering the high-density 3D designs at block and gate level one more challenge that adds on to the list is the lack of good EDA tools/software which can help to innovate and implement such 3D design and thereby help in the architecture of the 3D stacking process. Contemporary 3DIC designs operate under the premise that fully functional systems will be stacked in three dimensions. Consequently, existing tools lack support for automated 3D partitioning or optimization of cross-tier placement, timing, and routing. However, companies like Cadence are working towards solving this issue as discussed in the next section

Furthermore, in the case of die stacking and wafer stacking, fabricating TSVs with high aspect ratios (the ratio of depth to width) presents technical challenges, particularly in terms of achieving uniformity and reliability. The aspect ratio limita-

tions can constrain the density of TSVs that can be realized in practice. Despite these challenges, the density of TSVs can still reach impressive levels, with dimensions measured in micrometers. Additionally, the bandwidth between the layers in 3D stacked configurations is notably superior to that achievable with conventional package stacking methods. Some technical challenges mentioned in [6] include the thermal treatment problem and increase in design complicity problem.

- Thermal management poses a significant challenge in high-density electronic packages such as 3D stacked packages. As electronic devices become increasingly compact and densely packed, the dissipation of heat generated by these components becomes more difficult. Despite advancements in miniaturization and energy efficiency, the power consumption of electronic devices continues to rise or remain constant, exacerbating thermal issues. To address these challenges, greater attention needs to be focused on thermal design considerations in electronic packaging. This includes the use of advanced thermal interface materials, such as thermal interface pastes and pads, to improve heat transfer between components and heat sinks. Additionally, innovative cooling solutions, such as micro-channels, heat pipes, and vapor chambers, can be employed to enhance heat dissipation efficiency within compact electronic packages.

- When integrating multiple dies or packages in a high-density method, and incorporating various functional materials in a 3D package, the distribution of electromagnetic fields becomes significantly more complex. This complexity can lead to various electrical problems, including serious isolation ratio issues, signal waveform distortion, and electromagnetic interference (EMI) or electromagnetic compatibility (EMC) issues. To address these challenges, high-quality designs that effectively manage electromagnetic interference are essential. This

requires employing appropriate design methodologies, such as careful routing and layout techniques, optimized grounding and shielding strategies, and thorough electromagnetic simulation and analysis.

4 Recent Advances

Currently, numerous notable advancements are underway in the realm of 3D stacking, with several prominent ones being highlighted and discussed herein.

IBM Research and Tokyo Electron have teamed up to pioneer a breakthrough in chipmaking, streamlining the manufacturing process for wafers with 3D chip stacking technology [7]. Together, they've developed a novel 300 mm module utilizing an infrared laser for a debonding process that works seamlessly with silicon, eliminating the need for glass wafers typically used as carriers. Their innovative approach allows silicon wafers to be directly bonded to one another, eliminating the introduction of glass into the production line. This not only simplifies the supply chain but also brings several fabrication benefits. By processing bonded silicon wafer pairs, issues related to tool compatibility and chucking are avoided, resulting in fewer defects and process challenges often encountered with dissimilar wafer pairs. Additionally, it enables inline testing of thinned wafers, enhancing quality control. Furthermore, this technology paves the way for advanced FOWL (Fan-Out Wafer-Level) and 3D chiplet technologies utilizing silicon interconnects, which exhibit excellent scalability. In essence, this collaborative effort between IBM Research and Tokyo Electron marks a significant step forward in chip manufacturing, making the process more efficient and enabling the development of cutting-edge semiconductor technologies.

Intel's most recent breakthrough in transistor research, unveiled at IEDM 2023, marks a significant milestone for the industry [8]: the pioneering capability to ver-

tically stack complementary field-effect transistors (CFET) with a scaled gate pitch as small as 60 nanometers (nm). This achievement highlights Intel's leadership in gate-all-around transistors and demonstrates the company's capacity for innovation beyond RibbonFET, positioning it ahead of competitors in the field.

Cadence has introduced a new suite of software tools named the Integrity 3D-IC platform, aimed at aiding engineers in creating faster and more energy-efficient chips utilizing 3D packaging technologies [9]. This comprehensive solution from Cadence encompasses 3D design planning, implementation, and system analysis, all within a unified interface. It caters to the diverse needs of 3D-IC design, covering digital SoCs, analog/mixed-signal designs, and complete systems. The platform facilitates hardware and software co-verification, along with comprehensive power analysis, utilizing emulation, prototyping, and chipset based PHY IP for connectivity. This integration is optimized for various factors such as latency, bandwidth, and power, ensuring an enhanced performance and efficiency in chip development.

5 Conclusion

The potential of 3D stacking technology appears promising, given its significant advantages in terms of space-saving and reduction of interconnect delay and power consumption. It has become an integral part in revolutionizing the semiconductor industry, both presently and in the foreseeable future. Despite the presence of various challenges, such as architectural redesign and managing temperature during the fabrication of high-density interconnections, engineers are actively working on solutions.

In addressing these challenges, companies like Intel, Cadence, and IBM are at the forefront of innovation and investment. They recognize the importance of 3D

stacking technology and are committed to advancing its capabilities. For example, Cadence's advanced CAD tools are specifically designed to assist engineers in developing faster and more energy-efficient chips using 3D packaging technologies. Similarly, IBM and Intel are continually exploring new methods and efficient approaches to produce 3D stacked chips.

Through ongoing research and development efforts, these companies are paving the way for the widespread adoption of 3D stacking technology across various industries. By overcoming technical hurdles and optimizing design processes, they are driving innovation and shaping the future of semiconductor manufacturing. As a result, the potential of 3D stacking technology to revolutionize semiconductor design and manufacturing remains promising, with far-reaching implications for the advancement of technology and society as a whole.

References

[1] Saurabh Sinha et al. *Stack up your chips: Betting on 3D integration to augment Moore's Law scaling*. 2020. arXiv: 2005.10866 [cs.AR].

[2] Yi-Chang Lu. "3D technology based circuit and architecture design". In: *2009 International Conference on Communications, Circuits and Systems*. IEEE. 2009, pp. 1124–1128.

[3] I. Radu et al. "Recent developments of Cu-Cu non-thermo compression bonding for wafer-to-wafer 3D stacking". In: *2010 IEEE International 3D Systems Integration Conference (3DIC)*. 2010, pp. 1–6. DOI: 10.1109/3DIC.2010.5751454.

[4] Y. Hayashi et al. "A new three dimensional IC fabrication technology, stacking thin film DUAL-CMOS layers". In: *International Electron Devices Meeting 1991 [Technical Digest]*. 1991, pp. 657–660. DOI: 10.1109/IEDM.1991.235336.

[5] Yu-Wei Huang et al. “Assembly of 3D chip stack with 30m-pitch micro interconnects using novel arrayed-particles anisotropic conductive film”. In: *2013 IEEE 63rd Electronic Components and Technology Conference*. 2013, pp. 71–76. DOI: 10.1109/ECTC.2013.6575552.

[6] Jianyong Zheng et al. “3D Stacked Package Technology and Its Application Prospects”. In: *2009 International Conference on New Trends in Information and Service Science*. 2009, pp. 528–533. DOI: 10.1109/NISS.2009.100.

[7] Dale McHerron. *Breakthrough could simplify the 3D chipmaking supply chain*. Jan. 2023. URL: https://research.ibm.com/blog/IBM-TEL-chip-debonding.

[8] By. *Intel demonstrates breakthroughs in next-generation transistor...* URL: https://www.intel.com/content/www/us/en/newsroom/news/research-advancements-extend-moore-law.html#gs.77g73q..

[9] URL: https://www.cadence.com/en_US/home/explore/what-is-3dic.html.