Innovative 3D Heterogeneous Chip Manufacturing Approach to the Problem of Approaching Physical Limits with Traditional Chip Manufacturing Technologies

Muhammed Conger*, İbrahim Develi

Electrical and Electronics Engineering / Faculty of Engineering, Erciyes University, Türkiye

*(1030210058@erciyes.edu.tr)

Abstract – The 3D heterogeneous chip manufacturing approach is an innovative technological approach that has significant future potential. This innovative manufacturing process enables different semiconductor materials to be used in a single chip. It also offers higher performance and more functionality than traditional chips. 3D heterogeneous chip fabrication is accomplished by combining different semiconductor materials. These materials may include semiconductors with different electronic properties. While all components are made of the same semiconductor material in traditional chip manufacturing processes, a more complex and advanced structure is obtained by using different materials together in 3D heterogeneous chip production. 3D heterogeneous chip production brings many advantages thanks to the combination of different materials. Another advantage is to increase energy efficiency with 3D heterogeneous chip production. Using different materials together and in a single chip makes it possible to manage energy more efficiently. Another advantage is that the more complex and dense logic circuits of 3D heterogeneous chip fabrication can be located on the same unit square and on top of each other. In this context, this study presents the technical details and potential risks of the proposed solution method, along with end-to-end chip production, for the resolution of the mentioned problem.

Keywords – Chip, Transistor, Quantum, Semiconductor, Nanoelectronics

I. INTRODUCTION
The transistor, an important invention found in nearly every electronic device today, is a cornerstone of the modern electronics industry. First invented in 1947 by William Shockley, John Bardeen, and Walter Brattain at Bell Laboratories, the transistor replaced earlier electron tubes, enabling the production of smaller and faster devices. Transistors, which serve as the starting point for the digital-based technologies we have achieved today, led to the emergence of the concept of integrated circuits through the realization of circuits designed with logic gates. The claim, put forward by Gordon Moore, the co-founder of semiconductor company Intel, known as Moore's Law, stated that the number of transistors on a processor could double every 18 months, allowing for twice as many transistors to fit in the same unit area (Figure 1). This statement remained valid until the past few years and was made possible through the continuous shrinking of transistors.

Fig. 1 Trend graph related to Moore's Law
Although the transistor, invented in 1947, was the size of an iron at the time, the transistor production technology for today's processors only reaches a
size of 5nm. However, the number of transistors within a chip has reached around 100 billion. Processors consisting of transistors in various sizes such as 130nm, 65nm, 22nm, and 14nm are widely used in the industry to meet the demands. However, especially due to the requirements of 5G technologies, graphics technologies, artificial intelligence technologies, and memory technologies, the expected performance from processors is increasing. One of the most crucial factors in enhancing performance is the size of the transistor. This is because reducing the size of transistors leads to faster processors and lower power consumption (Figure 2).

![Fig. 2 Microscope image of a nm-size processor](image)

However, further reducing the size of transistors is reaching physical limits. Therefore, it can be said that Moore’s Law is no longer valid [1]. The fundamental reason behind this problem lies in the size of the Silicon (Si) atom, which is used to construct transistors in current technology and has semiconductor properties. A Silicon atom has a size of 210 picometers (0.2 nanometers). Although transistors with lengths of 2nm, 1nm, 0.65nm, and 0.34nm have been created through laboratory studies, they are not usable (Figure 3). The underlying reason for this is directly related to quantum mechanics. Due to the close proximity of transistors, which are described as being in a state of 1 when given an electrical potential of 5V and in a state of 0 when given 0V, it is assumed that the transistor can hold both a 1 and a 0 simultaneously, and this is referred to as a "bit." In this case, it is not possible to know with certainty whether the transistor is in the 1 state or the 0 state. Due to the transistor being present in both states simultaneously, quantum mechanics comes into play.

![Fig. 3 The smallest functional transistor developed in a laboratory environment](image)

Considering that processor design technologies are based on logical gates, when the observation of both the 1 and 0 states simultaneously occurs, the resulting quantum mechanical state will disrupt all logical rules. The individual pieces of information in the resulting quantum mechanical state are called "qubits" instead of bits (Figure 4).

![Fig. 4 Comparison between logic bits and quantum qubits](image)

Due to the aforementioned physical constraints, it will not be feasible in the near future to produce chips that can meet the requirements of 5G technologies, graphics technologies, artificial intelligence technologies, and memory technologies. To solve this problem, either the transistor technology or the manufacturing technology needs to be changed. One possible solution is to develop quantum chips where information is transmitted using light instead of electrons, by changing the transistor technology. This would require the development of very small photodetectors, lenses, a new logic system that supports quantum mechanical states, and the integration of new chips with existing electronic systems.

Considering current technologies, it is not yet feasible to replace transistor technology with a quantum mechanical evolution. However, the proposed alternative method of changing the
manufacturing technology is easier and more feasible with current technologies compared to replacing transistor technology. There are two main methods being explored in the field of changing manufacturing technology to meet future needs.

The first approach, proposed to produce chips with more transistors to meet future demands, is to manufacture larger chips (Figure 5). However, simply increasing the transistor count by using more surface area while keeping the transistor size constant can lead to significant disadvantages. Using this method to produce much larger-scale chips compared to current ones would result in a significant decrease in performance and a substantial increase in power consumption. Additionally, the heat generated on the chip can reach levels that may damage the circuit board. Furthermore, the cooling systems for current processors may be insufficient to dissipate the generated heat.

Concerns regarding power consumption and performance may arise in this scenario. However, instead of expanding chips horizontally, vertically scaling them will surprisingly yield even higher performance than existing chips. For example, in the case of a "System on Chip" (SoC) structure where a processor core and an artificial intelligence accelerator need to communicate with each other, stacking these two devices on top of each other will significantly reduce the distance compared to their placement in current chip designs. Heterogeneous 3D chips arranged in this manner, as exemplified in the given scenario, will lead to increased performance and lower power consumption compared to existing chips, thanks to the substantial reduction in distances. Considering the increasingly widespread use of mobile devices in the present and future, it is crucial to significantly improve power efficiency. Furthermore, reducing the distances that units within a chip need to cover to communicate with each other will enable achieving very high performance values.

Thus, an innovative solution method has been proposed to address the problem of approaching the physical limits hindering the production of more powerful chips capable of meeting the requirements of advanced technologies.

II. MATERIALS AND METHOD

With current manufacturing technologies, the physical limits of chip production are being approached. To address this problem and envision the future of chip technology, a 3D heterogeneous chip manufacturing approach has been introduced. To understand this approach, it is essential to grasp chip architecture (Figure 7), chip design, and the existing chip manufacturing techniques.

A. Chip Architecture

To design a chip, the requirements must be determined first. After defining the requirements, the instruction set architecture (ISA) that matches
those requirements is determined, and then the instructions needed from the instruction set are identified. Once the architecture capable of meeting the expected requirements and the corresponding instructions are determined, a datapath capable of successfully executing all possible scenarios is constructed.

This datapath, tailored to the requirements, includes an arithmetic logic unit (ALU) for performing arithmetic and logic operations, a program counter (PC) that keeps track of when instructions should be fetched, a program memory that stores instructions, an expansion unit in case external inputs are shorter than the register width, a register file that determines where data is read from or written to, a data memory that stores information of running programs and allows reading from or writing to a given address, static and dynamic branch predictors for predicting potential branches in advance, structures for handling exceptional situations (e.g., interrupts), and pipelines that can be added to further enhance the chip's performance [3]. A control mechanism is also required to govern the actions the datapath should take under different circumstances. Therefore, the locations for control checkpoints on the datapath components are determined, followed by the design of the control unit, which ensures the necessary control. Another crucial aspect of chip design is memory. Although memory structures are often considered external to the chip, they are actually integrated into it. As technology advances, chip performance expectations increase rapidly. This particularly applies to memory structures since memory instructions often form the critical path, necessitating the partial integration of memory structures into the chip to enhance performance. The highest-speed memory type in terms of memory types, Static Random-Access Memory (SRAM), has been integrated into modern chips. However, these high-speed structures are considerably expensive compared to other memory types. Therefore, a trade-off is necessary. To achieve this trade-off, a memory hierarchy is established, consisting of SRAM structures at multiple levels, a main memory structure, and secondary storage areas, based on the requirements. (Figure 8).

In this hierarchy, the speed decreases towards secondary storage, but the cost decreases as well. Data that will be frequently accessed by the components within the chip is stored in SRAM structures or other intermediate structures closer to it, while less frequently accessed data is stored in the secondary storage or other intermediate structures closer to it at the end of the hierarchy [4]. This approach of providing memory for the components within the chip rather than fetching it from external sources leads to a significant increase in performance.

B. Chip Design

The design stages of a chip can be divided into front-end and back-end (Figure 9). The back-end design stage encompasses the physical design steps.
Once the requirements for the chip design are determined, a diagram illustrating the connections of the required architectural blocks should be prepared. Subsequently, the logical design that implements the emerged architecture is created using electronic design automation (EDA) tools. In order to create the logical design, hardware description languages (HDLs) are used. These languages can include Verilog, VHDL, and SystemVerilog. Unlike software programming languages that run on existing hardware, hardware description languages directly create the hardware structures using register transfer language (RTL). Thus, the architectural design that can meet the requirements is realized using the RTL structure created with hardware description languages [5]. Next, the generated RTL is synthesized using electronic design automation (EDA) tools. In this stage, the RTL is transformed into equivalent logic gate circuits. Depending on the capabilities of the design tools, optimizations, power estimations, and timing analyses can be performed.

Verification processes must be conducted at necessary points during the progression of stages. For this purpose, a testbench is created, and simulation tools are used to observe how the logic circuits will operate in the real world in the time domain.

After the error-free completion of the front-end design stages, the back-end design stages, which are the physical design steps, are initiated. In this section, detailed and precise operations are performed, such as the placement of the designed chip components on the silicon, floor planning, power and ground assignments, clock synthesis, signal routing, and physical verification.

Additionally, static timing analysis (STA) is carried out, especially for clock synthesis, to achieve the highest achievable performance of the chip. During the physical design stage, pre-designed macros, which are available for purchase or provided as open-source, are used to transform the created logic circuits into physical transistor circuits (Figure 10).

The term "macro structures" refers to the transistor-level circuit equivalent of any logic gate, combinational logic circuit, or memory unit’s logical design.

After successfully completing all the simplified stages related to chip design, the final version of the chip is outputted in the form of a geometric data base standard for information interchange (GDSII). With this output, the designed chip is ready for production.

C. Chip Manufacturing

Chip manufacturing is primarily dependent on Silicon (Si) atoms, which is why it is commonly referred to as silicon production. To create silicon, sand, which is abundant in silicon, is used as the main raw material. Although sand can be found in various regions worldwide, it may not contain the same amount of silicon as sand from another region.
By conducting necessary measurements and considering the silicon content and other parameters, the source of raw material for chip production is determined. The sand collected from the designated region is heated until it reaches a high purity level and then undergoes crystallization to solidify. This process separates the silicon present in the sand in a pure form. The resulting silicon ingot is then shaped into cylinders of predetermined dimensions based on requirements (Figure 11).

The diameter of the cylindrical pure silicon is important. Although production can be carried out in various sizes depending on the needs, the commonly used cylinder has a diameter of 300mm. This is because it allows for the production of chips with minimal waste.

The sand collected from the designated region is heated until it reaches a high purity level and then undergoes crystallization to solidify. This process separates the silicon present in the sand in a pure form. The resulting silicon ingot is then shaped into cylinders of predetermined dimensions based on requirements (Figure 11).

The diameter of the cylindrical pure silicon is important. Although production can be carried out in various sizes depending on the needs, the commonly used cylinder has a diameter of 300mm. This is because it allows for the production of chips with minimal waste.

The small size of silicon wafers for chip processing limits the number of chips that can fit on them. This leads to significantly higher production costs per unit chip. On the other hand, large silicon wafers are not desirable due to a disadvantage of the current production technology. This disadvantage arises from the decreasing performance of chips as one moves from the center to the edges of the silicon wafer. Due to these constraints, 300mm diameter silicon wafers are considered the optimal value for current production technologies. Another consideration is the cylindrical shape of silicon. In today's technology, where we are approaching physical limits, it is possible to commercially produce chips at the 5nm level.

However, for such production at the nanometer scale, a special geometry is required to fit as many chips as possible on the silicon wafer and minimize the margin of error during cutting. Placing chips on square, rectangular, or polygonal silicon plates can result in areas near the edges of the silicon plate being empty or exceeding the dimensions of the designed chips. Moreover, even the slightest cutting error during chip separation can render dozens of chips unusable. Therefore, the most suitable geometry is a circular shape. This allows for more efficient utilization of the areas near the edges of the silicon wafer, and cutting errors are minimized due to the geometry of the wafer. As a result of these production risks, pure crystalline silicon is formed in a cylindrical shape. The obtained cylindrical structure is sliced to obtain silicon wafers of a specific thickness according to the requirements, using special cutting techniques (Figure 12).

The surfaces of the sliced silicon wafers are rough and may contain defects. Therefore, polishing machines are used to smoothen the surface of the silicon wafers. This is because even the slightest imperfection or dust particle on the silicon wafer surface can become significant when considering that the transistors being fabricated will be at the nanometer scale. Consequently, these imperfections can disrupt the desired circuit structures.

The silicon wafer, which has been made smooth, is not yet conductive. A series of processes are applied to the silicon wafer to make it a semiconductor. Firstly, the wafer undergoes an oxidation process, creating an oxide film on its surface. Oxygen or water vapor is sprayed onto the surface for this purpose. This stage is crucial because the oxide film formed on the silicon wafer surface protects the silicon during subsequent processes. Additionally, this oxide film prevents any current leakage that may occur between the circuit structures to be formed. Next, a photosensitive resist layer, capable of responding to light, is created on
the oxide film-coated silicon wafer. The disk is now ready for transferring the chip design onto the silicon.

To transfer the chip design onto the silicon, layout plans and connections are made, and a mask is created. The silicon wafer, coated with the photosensitive resist layer, is positioned beneath the mask containing the pattern of the entire chip design. The chip design is then transferred onto the silicon wafer by exposing it to light through the mask. This process is called lithography. The areas not exposed to light, as defined by the mask, are then removed from the silicon wafer. This separation process is carried out using chemical solutions and gases or plasmas.

Subsequently, the successful transfer of the chip design onto the silicon is verified. The lithography, etching, and verification processes are repeated, taking into account the layers in the design. Thus, the process of transferring the chip design onto the silicon is completed (Figure 13).

![Figure 13 Chips transferred on silicon disc (wafer)](image)

To impart semiconductor properties to the fabricated circuits, ion implantation is performed by depositing materials at the molecular or atomic level onto a silicon wafer. In order for the design to be capable of carrying electrical signals, the created semiconductor wells need to allow the passage of electricity. For this purpose, a thin metal film, such as Aluminum, Titanium, or Tungsten, is deposited onto the silicon wafer. One might wonder why copper is not used for the propagation of electrical signals. Although copper conducts electricity with approximately 40% less resistance than aluminum, it is not preferred due to its rapid diffusion into silicon and its potential to disrupt the fabricated transistor structures. After all these processes, test procedures are applied to identify and discard defective chips on the silicon wafer that houses the chip designs.

In this stage, also known as energy dispersive spectroscopy (EDS), it can be observed that chips near the edges of the silicon disk show lower performance due to the larger diameter of the cylindrical silicon. Additionally, defective chips may also arise. In terms of production efficiency, it refers to the percentage of chips operating at maximum efficiency compared to the maximum number of chips on a single silicon wafer.

Following this stage, the chips on the silicon wafer are separated by cutting them individually. To mitigate potential cutting errors, the chips are tested again, and the validated chips proceed to the packaging stage. Here, fine wires are drawn from the relevant points on the chip's surface using a highly precise needle for its communication with the external world and power connections (Figure 14). Subsequently, the chip is packaged within a casing containing the necessary pins for its placement on electronic boards. Thus, the production process of the chip is completed [6]. With current technologies, the entire process from chip architecture design to chip manufacturing has been simplified and explained end-to-end. However, due to the approaching physical limits of current chip manufacturing technologies, it is necessary to reconstruct the entire process from chip architecture to chip production. In this regard, the most effective and feasible solution method proposed in the Introduction section is seen as the 3D heterogeneous chip manufacturing approach.

D. 3D Heterogeneous Chip Manufacturing

With the advancement of technology, the expectations for chip performance have been increasing. Especially the requirements for 5G technologies, graphics technologies, artificial intelligence technologies, and memory technologies have led to the need for more advanced and higher-performing chips. However, despite the nanometer-level manufacturing capabilities reached in transistor production with current chip manufacturing technologies, physical limits are being reached.
In recent years, many developments have been made to address this problem faced in the near future, and even today. While methods such as partially integrating memory structures into chips, consolidating different systems on a single chip, and developing structures capable of predicting branching in instructions have achieved significant progress, they are not sustainable methods to meet the requirements of new and future technologies. Therefore, this section analyzes the complex processes and methodologies related to the 3D heterogeneous chip manufacturing approach, which is considered the most effective and feasible method among the proposed solution methods in the introduction. The fundamental motivation for the development of 3D heterogeneous chips is to enable the positioning of devices with different electrical characteristics (heterogeneity) on top of each other (3D structure) and in a single package (System-in-Package, SiP) to solve the aforementioned problems [7]. This allows for the integration of various devices, ranging from sensor structures to photodetectors, memory units located outside of current chips, and analog units, in a single package without horizontally expanding the chip.

Even positioning devices with the same electrical characteristics on top of each other and in a single package is quite challenging. In this context, 3D heterogeneous chip manufacturing brings a series of specific challenges compared to traditional 2D chip manufacturing. These challenges can be listed as thermal heat management, electrical connections, surface flatness, topographic incompatibility, high integration, design complexity, and manufacturability.

In 3D heterogeneous chips, where different devices and components are densely integrated, thermal management is of great importance. Inadequate thermal heat management can lead to serious problems in terms of high-density integration, thermal resistance, and temperature distribution. In particular, thermal incompatibility between different devices can lead to thermal stress and temperature increase. Therefore, innovative thermal management solutions such as thermal interface materials and cooling techniques are necessary.

Another challenging issue is electrical connections. 3D heterogeneous chips require high-density electrical connections between different devices. These connections should provide low resistance, low loss, and high signal integrity. Consequently, highly sensitive and advanced manufacturing techniques need to be developed within this scope. Since devices composed of different materials will have different electrical characteristics, there may be a problem with signal integrity. This situation can lead to even meaningful signals being considered as noise and the generation of parasites. Surface flatness and topographic incompatibility issues can also arise in 3D heterogeneous chip manufacturing. 3D heterogeneous chips containing different devices will inevitably have different surface topographies. Different surface flatness and topographic incompatibilities can cause problems during silicon bonding and through-silicon via (TSV) processes. Precise processes and techniques such as surface correction or surface matching need to be employed to ensure proper surface flatness and compatibility. The challenges of high integration and design complexity indicate that 3D heterogeneous chip design and manufacturing will be significantly more complex than traditional chip stages. The integration of different devices and components will make the design highly complex in terms of determining their positioning and layers within the chip, routing signals, and power management [8]. In addition to these challenges, an unforeseen issue that may arise is electromagnetic interference. Having multiple devices with different tasks in such close proximity and within a single chip can lead to complex interference and the formation of high antenna effects. However, considering that current devices are located in discrete chips with large drivers, resistors, inductors, and capacitors,
reducing intermediate components and consolidating main components on a single chip can have a positive impact on electromagnetic interference contrary to what might be expected.

Despite all these challenges, when comparing 3D heterogeneous chips to current chips, it can be observed that product costs will primarily decrease. This is because various devices, including analog structures, digital structures, memories, and even sensors, can be produced on a single chip. Additionally, having multiple devices in a single chip without horizontal expansion allows for the creation of much smaller products, particularly making mobile devices much more compact. The unity of devices within the chip eliminates the need for large drivers and reduces the required number of resistors, inductors, and capacitors for interconnections between traditional devices, resulting in significant power consumption improvement.

Furthermore, with many units, especially memories that form critical paths in terms of instructions, being stacked directly above the central processor, the distance the signal lines need to traverse will be significantly reduced, thus pushing performance to higher levels. As a result, 3D heterogeneous chips will be capable of meeting the high-speed requirements of advanced technologies, exceeding 200 Gbps. A change in the chip manufacturing method, which is approaching physical limits with current production technologies, is inevitable. However, the situation does not only involve changing the architecture, design, and production methodologies in line with the 3D heterogeneous chip manufacturing approach. As a result, significant changes will be needed in analysis, static timing analysis, verification processes, integrity, and synthesis. Consequently, electronic design tools will also need to be reconfigured to enable these processes. In this context, it is essential to develop electronic design tools that encompass end-to-end stages, designed and implemented specifically for 3D heterogeneous chip manufacturing, in order to create chips that meet the requirements of future technologies.

### III. RESULTS

In recent years, rapid advancements in technology and electronics have increased the demand for smaller, faster, and more powerful chips. In response to this demand, the semiconductor industry is in search of new solution methods. As a result of this quest, the 3D heterogeneous chip manufacturing approach has emerged as a groundbreaking method for developing more complex and capable chips. 3D heterogeneous chips are chips where different materials and components are brought together. Unlike traditional 2D chips, the components in different layers of 3D heterogeneous chips can interact with each other and perform data transfer.

This approach translates to increased processing power, higher bandwidth, and lower power consumption. It is evident that 3D heterogeneous chips will outperform traditional chips due to the integration of different devices and the absence of horizontal expansion. Furthermore, the close proximity of different components in 3D heterogeneous chips allows for highly increased inter-device communication bandwidth.

### IV. DISCUSSION

In recent years, 3D heterogeneous chips have become a major focus of interest in the fields of electronics and computer science. This innovative approach offers greater flexibility, higher performance, and a significant potential for improvement in power consumption compared to traditional 2D chips. 3D heterogeneous chips are created by integrating different materials and components. This enables designers to create more complex and customizable chip designs. For example, components such as transistors, sensors, and photodetectors of different scales can be housed in a single package and within the same chip. The advancements and unprecedented high performance brought by this approach will greatly accelerate scientific research that requires high levels of complexity and processing power. Moreover, it will expedite discoveries in areas such as genomic analysis, universe mapping, and big data, where current chips face challenges in meeting demands within shorter time frames.

The integration of optical and electronic components can also be achieved with 3D heterogeneous chips, leading to a significant boost in optoelectronic research. Optoelectronic systems hold great importance in the scientific community. 3D heterogeneous chips allow for the combination of optical components at different scales with electronic components. This results in smaller, faster, and more powerful optoelectronic systems.
For instance, the integration of laser diodes and photodetectors enables high-speed data transmission in optical communication systems. Additionally, the merging of optical sensors with electronic processors facilitates high-precision and real-time data collection with significantly higher performance than traditional chips.

3D heterogeneous chips will bring forth numerous advanced applications and exploration potentials in the scientific world. For example, in the medical field, it will be possible to make biosensors and electronic devices in implants smaller and more integrated. In the fields of artificial intelligence and machine learning, 3D heterogeneous chips will greatly expedite progress by offering faster and more efficient computational capabilities.

As discussed, 3D heterogeneous chips hold great potential for scientific research and advancements. Factors such as advanced integration, high bandwidth, optoelectronic integration, power consumption, and advanced applications demonstrate the significant contribution 3D heterogeneous chips can make to science. These next-generation chips have the potential to drive new discoveries, enhance existing technologies, and provide more effective solutions. However, despite the contributions 3D heterogeneous chips can make to scientific research, there are also challenges. The design and manufacturing of 3D heterogeneous chips are more complex compared to traditional 2D chips. The integration of different components and the compatibility of various materials require more engineering and research. Furthermore, testing and characterizing 3D heterogeneous chips will be quite demanding. Nevertheless, considering the potential advantages of 3D heterogeneous chips, these challenges can be overcome and are manageable.

Scientists and researchers find this approach highly significant for advanced research and discoveries due to the potential advantages in flexibility, performance improvement, and power consumption. By embracing this advanced manufacturing approach, the scientific community can overcome the anticipated design and production challenges and achieve even greater technological successes in the future.

V. CONCLUSION

The overcoming of potential technical and mechanical challenges in the production processes of the proposed 3D heterogeneous chip manufacturing approach is expected to have a significant impact, particularly in areas that require very high chip performance and low power consumption. Although this emerging manufacturing approach has gained significant interest, it is still in its early stages. The lack of standardized and comprehensive definitions indicates that both chip manufacturing facilities and chip designers are not yet ready for this transformation. Therefore, in this study, information has been provided about new production techniques and potential challenges related to how the world of electronics and chip manufacturing will shape in the near future.

REFERENCES