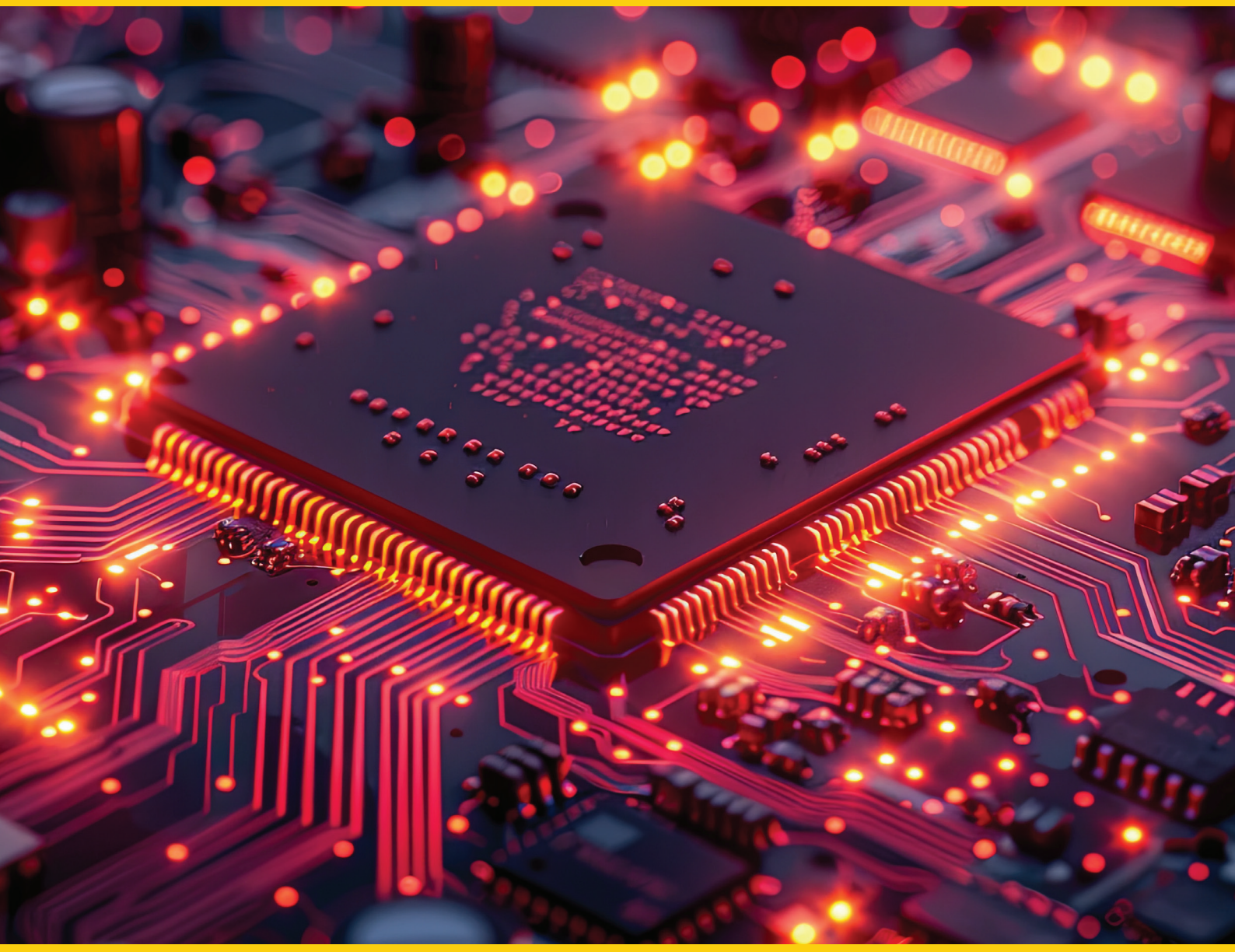
Revolutionizing semiconductor design in the sub-3nm era with chiplets & 3D ICs



About the authors:

Anam Haque,
VP of Engineering / Industry Leader / Engineering Delivery Head, Quest Global

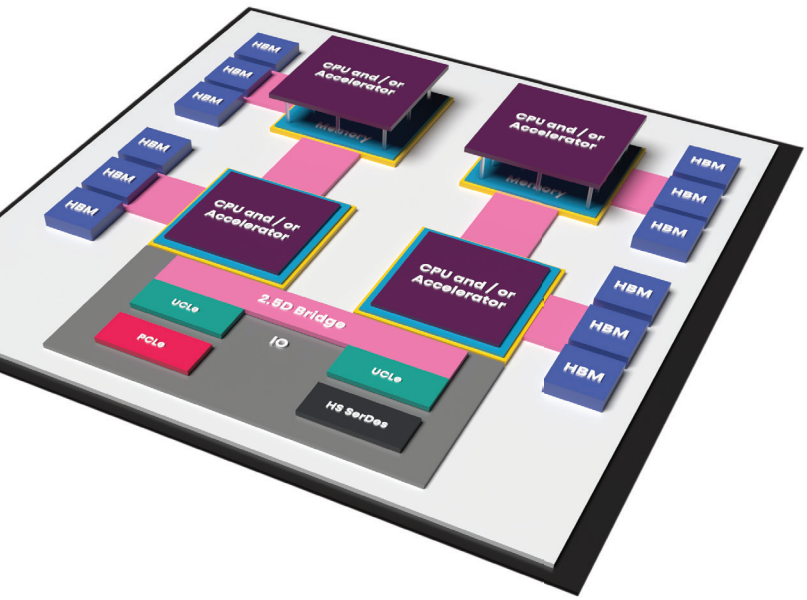
Vasan Karighattam,
Chief Architect, AVP Engineering, Quest Global

Introduction

As we approach the physical limits of traditional semiconductor scaling, the industry is witnessing a paradigm shift in chip design and manufacturing. This article explores the cutting-edge technologies of chiplets and 3D integrated circuits (ICs), which promise to extend Moore's Law and usher in a new era of semiconductor innovation. The insatiable demand for increased processing power, coupled with the need for energy efficiency, has pushed engineers to innovate beyond traditional planar chip designs. Enter the era of chiplets and 3D integrated circuits (ICs) – a paradigm shift that promises to reshape the landscape of silicon engineering. Chiplets represent a revolutionary approach to system-on-chip (SoC) design.

Before delving into full 3D integration, it's important to note the significant role of 2.5D technology in this evolution. 2.5D integration, which involves placing multiple chiplets side-by-side on an interposer, has already gained traction in the industry. Major players like TSMC (with their CoWoS technology), Samsung (with their I-Cube), and Intel (with their EMIB) have embraced 2.5D packaging solutions. These technologies have paved the way for more complex integration strategies by allowing high-bandwidth connections between chips and overcoming some of the limitations of traditional 2D designs. Notably, 2.5D chips have become crucial in today's applications as they allow chip designers to overcome the reticle limit - the maximum size of a single chip that can be manufactured using photolithography. The connection of multiple smaller chiplet on an interposer in 2.5D technology enables the creation of larger, more complex systems that would be impossible to manufacture as a single monolithic chip.

Open chiplet ecosystem with UCIe 2.0



The rise of chiplets

Chiplets are individual semiconductor components that can be integrated to form a complete system-on-chip (SoC) or a more complex semiconductor device. This modular approach disrupts traditional monolithic SoC design by deconstructing complex systems into specialized, miniaturized ICs, each optimized for a specific function. These individual semiconductor components can be integrated to form complex devices, effectively deconstructing traditional monolithic SoCs into specialized, miniaturized ICs.

Key Advantages of Chiplet Technology

1. Technical flexibility: Chiplets allow for independent design and fabrication using the most suitable process node for each function.

2. Manufacturing efficiency: Potential for lower overall production costs and improved yield.

3. Heterogeneous integration: Facilitates the integration of functionalities built on diverse process nodes and materials.

4. Performance boost: Tight integration of specialized chiplets often results in superior performance.

5. Improved yield: Individual chiplets can be tested and validated before integration.

6. Scalability: Functionality can be more easily scaled by adding or replacing chiplets.

Stacking for success–3D IC integration

3D IC integration involves vertically stacking multiple layers of semiconductor devices, creating much higher levels of integration and faster computing than previously possible.

Types of 3D integration

Package-on-Package (PoP): This technique involves stacking fully packaged chips on top of each other. Typically, a logic chip (like a processor) forms the base, with memory chips stacked above. Key features include:

- Vertical interconnects between packages using solder balls or through-mold vias
- Each chip can be tested individually before stacking, improving overall yield
- Allows for mixing of different chip technologies and manufacturers
- Commonly used in mobile devices to save horizontal space on the PCB
- Challenges include thermal management and limitations in interconnect density

Die-to-die bonding: This method involves directly bonding multiple bare dies together. It can be achieved through various techniques such as micro-bump bonding or hybrid bonding. Key aspects include:

- Higher interconnect density compared to PoP, allowing for more bandwidth between dies
- Can be done face-to-face or face-to-back, depending on the design requirements
- Enables heterogeneous integration of dies from different process nodes
- Requires precise alignment and bonding techniques
- Thermal management is critical due to the close proximity of active layers
- Examples include AMD's chiplet-based processors and HBM (High Bandwidth Memory) stacks

Die-to-wafer and wafer-to-wafer bonding: These techniques extend the concept of die-to-die bonding to larger scales, offering potential advantages in manufacturing efficiency and cost.

Die-to-wafer bonding:

- Involves bonding individual dies onto a full wafer
- Allows for known-good-die (KGD) selection, potentially improving overall yield
- Enables heterogeneous integration of dies from different wafer sizes or manufacturing processes
- Challenges include ensuring consistent bonding quality across the wafer and managing thermal expansion mismatches

Wafer-to-wafer bonding:

- Involves bonding two or more full wafers together before dicing
- Offers the highest throughput for high-volume manufacturing
- Provides excellent alignment accuracy across the entire wafer
- Challenges include yield management, as defects on either wafer can impact the final product
- Best suited for applications where dies from both wafers have similar sizes and high individual yields

Monolithic 3D integration: This advanced technique involves building multiple layers of devices on a single substrate, creating a truly three-dimensional IC. Key features include:

- Achieved by sequential processing of device layers on the same wafer
- Offers the highest possible vertical integration density
- Enables extremely short vertical interconnects, reducing power consumption and latency
- Challenges include managing thermal issues and developing fabrication processes that don't damage lower layers
- Still largely in research phase, with potential for future high-performance, low-power devices
- Examples include research on stacked CMOS layers and integration of logic with non-volatile memory

Each of these techniques represents a different approach to 3D integration, offering various trade-offs in terms of performance, manufacturing complexity, and cost.

Applications and use cases

Automotive: Enhancing ADAS and autonomous vehicles with efficient integration of high-performance processing and AI accelerators. Chiplet-based designs allow for the combination of sensors, processors, and AI cores in compact, power-efficient packages. This enables real-time processing of complex sensor data, improving safety features and paving the way for fully autonomous driving. The modular nature of chiplets also allows for easier upgrades and customization to meet varying levels of autonomy.

Mobile devices: Enabling customized solutions for smartphones and tablets, optimizing power consumption, graphics, and AI processing. Chiplet technology allows mobile device manufacturers to integrate best-in-class components for each function, such as high-performance CPU cores, efficient GPU units, and dedicated AI processors. This approach can lead to improved battery life, enhanced gaming experiences, and more powerful on-device AI capabilities, all within the tight thermal and space constraints of mobile devices.

Data centers: Offering enhanced performance, energy efficiency, and modular upgradability for servers. Chiplet architectures enable data centers to mix and match components for optimal performance and power

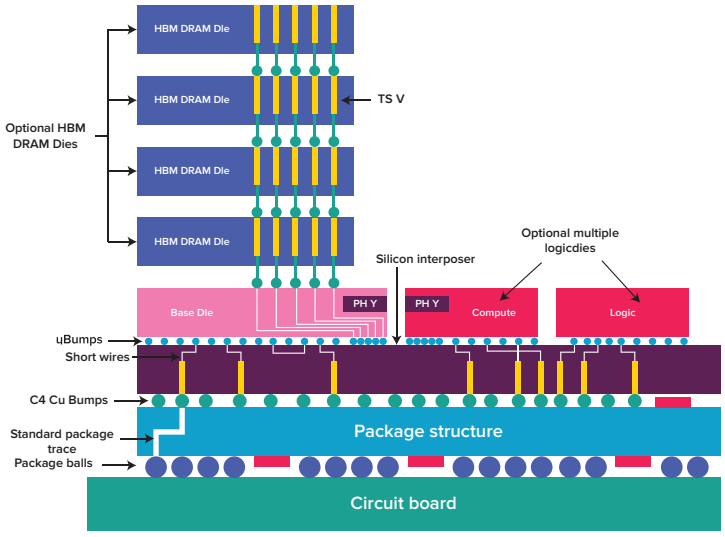
efficiency. This flexibility allows for specialized configurations tailored to specific workloads, such as AI training, database management, or high-performance computing. The modular approach also facilitates easier upgrades and maintenance, potentially extending the lifespan of data center hardware.

AI and machine learning: Delivering tailored acceleration for specific workloads, improving inferencing and training capabilities. Chiplet-based AI accelerators can be customized for different types of neural networks and machine learning algorithms. This specialization allows for significant performance improvements and energy efficiency gains compared to general-purpose processors. The modular nature of chiplets also enables easier scaling of AI systems, from edge devices to massive data center installations, by combining multiple AI accelerator chiplets as needed.

Wearables: Chiplet technology is revolutionizing wearable devices by enabling more sophisticated functionalities in compact form factors. The ability to integrate specialized chiplets for sensors, low-power processing, and wireless communication allows for the development of advanced health monitoring systems, smart clothing, and next-generation fitness trackers. This modular approach facilitates longer battery life and more accurate data processing, crucial for continuous health monitoring and user comfort.

AR/VR devices: Augmented and Virtual Reality devices benefit significantly from chiplet architecture. The technology enables the integration of high-performance graphics processing, spatial computing, and sensor fusion in compact, head-mounted displays. Chiplets allow for optimized thermal management and power efficiency, critical for comfortable and extended use of AR/VR devices. Additionally, the modular nature of chiplets facilitates easier upgrades to keep pace with rapidly evolving AR/VR technologies, potentially extending the lifespan of these devices.

2.5D /3D Solutions for the data center



Advanced packaging and system-level considerations

Package-level implications

1. **Increasing package sizes:** Trend towards larger packages to accommodate multiple chiplets, necessitating new approaches to system integration and thermal management.
2. **I/O density and management:** Challenges in routing and signal integrity with high I/O counts, requiring advanced PCB design techniques and materials.
3. **Package-level simulation:** Necessity for comprehensive simulations to capture complex interactions, including electrical, thermal, and mechanical aspects of multi-chiplet packages.

High-speed communication

1. **Advanced communication standards:** Implementation of 112G and 224G standards, pushing the boundaries of signal integrity and requiring sophisticated equalization techniques.
2. **Signal integrity and reach management:** Maintaining signal quality across stacked cards and boards, involving advanced channel modeling and pre-emphasis/de-emphasis strategies.

Thermal management

1. **Heat dissipation challenges:** Dealing with concentrated heat sources and implementing advanced cooling solutions, such as microfluidic cooling or phase-change materials.
2. **Thermal simulation and design:** Integration of thermal considerations into the overall design process, using advanced computational fluid dynamics (CFD) simulations.

System-level challenges

1. **High port count systems:** Designing and managing systems with 1000 to 6000 ports, requiring sophisticated signal routing and power distribution networks.
2. **Miniaturization and performance:** Balancing performance with footprint constraints, often necessitating 3D integration techniques.
3. **AI-based array-level chips:** Design considerations for highly parallel, AI-focused architectures, including optimized data flow and specialized memory hierarchies.
4. **Power optimization:** Techniques for reducing geometries and addressing the power equation ($P = \frac{1}{2}CV^2$), including voltage scaling and advanced power gating strategies.

Interposer-level integrations

1. **High Bandwidth Memory (HBM):** Integration using standardized interfaces, enabling high-performance, low-power memory solutions for data-intensive applications.
2. **Chiplet-to-chiplet and chiplet-to-memory connectivity:** Use of mature node substrates and silicon passive interposers, allowing for high-density interconnects and reduced signal path lengths.
3. **Industry standards and initiatives:** Intel's EMIB, UCIe standard, and TSMC's advanced packaging technologies, driving interoperability and ecosystem development.

Cost and manufacturing considerations

1. **Manufacturing complexity:** Challenges in producing defect-free silicon at large scales, requiring advanced process control and yield management techniques.
2. **Cost implications:** High costs associated with advanced packaging technologies, necessitating careful cost-benefit analysis for different application domains.
3. **Market segmentation:** Focus on high-end applications where costs can be justified, with potential for trickle-down to mainstream markets as technologies mature.
4. **Reliability and testing:** Importance of comprehensive testing and redundancy, including built-in self-test (BIST) and advanced fault tolerance mechanisms.

Simulation and modeling challenges

Increased model complexity: In chiplet-based designs, multiple interacting models must be considered simultaneously, significantly increasing the complexity of simulation and modeling tasks. This includes:

- **Electrical models:** Accounting for power delivery, signal integrity, and electromagnetic interactions between chiplets
- **Thermal models:** Simulating heat generation and dissipation across the entire package, including interactions between chiplets
- **Mechanical models:** Analyzing stress and strain in the package, especially important for 3D stacked designs
- **Functional models:** Ensuring correct operation of the entire system composed of multiple chiplets

These models must not only be accurate individually but also interact correctly to provide a holistic view of the system behavior. This multi-physics, multi-scale modeling approach requires sophisticated simulation tools and significant computational resources.

Exponential growth of scenarios: As the number of chiplets in a design increases, the number of possible interaction scenarios grows exponentially. This includes:

- **Power state combinations:** Each chiplet may have multiple power states, leading to a combinatorial explosion of system-wide power scenarios
- **Data flow patterns:** The variety of possible data exchange patterns between chiplets increases dramatically with chiplet count
- **Fault scenarios:** The need to simulate various fault conditions and their propagation through the system becomes more complex
- **Manufacturing variations:** Accounting for the impact of process variations across multiple chiplets adds another layer of complexity

This explosion in scenarios makes exhaustive testing practically impossible, necessitating advanced statistical and machine-learning techniques to identify and focus on the most critical scenarios.

Inter-chiplet communication: Accurate modeling of communication pathways and protocols between chiplets remains crucial for system performance and reliability. While standards like UCIe have made significant strides, some interoperability challenges persist. This involves:

- **High-speed serial link modeling:** Simulating the behavior of multi-gigabit transceivers, including effects like jitter, crosstalk, and signal loss
- **Protocol-level simulation:** Ensuring correct implementation of communication protocols, including error handling and flow control
- **Bandwidth and latency analysis:** Predicting system-level performance based on inter-chiplet communication characteristics
- **Power analysis:** Estimating the power consumption of communication interfaces, which can be a significant portion of overall system power

- **Interoperability testing:** Developing comprehensive test methodologies to ensure seamless integration of chiplets from different vendors, similar to the challenges faced in PCI Express implementations

The challenges are compounded by the need to model different interconnect technologies (e.g., silicon interposers, organic substrates) and emerging standards like UCIe. While UCIe has addressed many standardization concerns, the industry is still working on defining robust interoperability tests and methodologies. This gap in the ecosystem highlights the ongoing need for collaboration and standardization efforts beyond initial protocol definitions.

Scalability issues: As the number of chiplets in a design increases, simulation complexity grows exponentially, presenting significant scalability challenges:

- **Computation time:** Full-system simulations can become prohibitively time-consuming, potentially taking days or weeks for complex designs
- **Memory requirements:** Storing and manipulating data for large-scale simulations can exceed the capabilities of typical workstations
- **Model integration:** Combining models of different chiplets, potentially from different vendors or using different modeling approaches, becomes increasingly difficult
- **Abstraction level management:** Balancing the need for detailed, low-level simulations with higher-level system modeling to manage complexity

To address these scalability issues, the industry is exploring various approaches:

- **Hierarchical simulation techniques:** Using different levels of abstraction for different parts of the system
- **Distributed and cloud-based simulation:** Leveraging cloud computing resources to parallelize simulations
- **AI-assisted modeling:** Using machine learning techniques to predict system behavior and focus simulation efforts on critical areas
- **Reduced-order modeling:** Creating simplified models that capture essential system behavior without the full complexity of detailed simulations

These simulation and modeling challenges are at the forefront of chiplet-based design, and overcoming them is crucial for realizing the full potential of this technology. As the industry continues to develop more advanced tools and methodologies, we can expect to see improvements in our ability to design and validate increasingly complex chiplet-based systems.

Emerging trends and future directions

Ecosystem development: The growth of chiplet ecosystems is accelerating, with various players contributing specialized components. This trend is bringing about a more diverse and innovative semiconductor landscape. Companies are developing niche expertise in specific types of chiplets, such as AI accelerators, high-speed I/O, or specialized analog components. This specialization allows for rapid advancement in specific areas without the need for each company to master every aspect of chip design. As the ecosystem matures, we can expect to see a marketplace of chiplets, where system integrators can select the best components from multiple vendors to create highly optimized and customized solutions.

Mobile applications: The integration of High Bandwidth Memory (HBM) in mobile chips is poised to revolutionize smartphone and tablet performance. HBM offers significantly higher bandwidth and lower power consumption compared to traditional LPDDR memory. This integration enables mobile devices to handle more complex AI tasks, improves graphics rendering, and speeds data processing. However, challenges remain in terms of packaging size and thermal management for mobile form factors. Future developments may include specialized HBM configurations optimized for mobile use cases, potentially enabling desktop-class performance in handheld devices.

Artificial General Intelligence (AGI) implications: The pursuit of Artificial General Intelligence is driving innovation in chip design and potentially leading to cost reductions. AGI research demands unprecedented computational power, pushing the boundaries of chip design. This drive is leading to novel architectures that could have spillover benefits for broader chip applications. For instance, neuromorphic computing designs inspired by AGI research could lead to more energy-efficient chips for edge computing. Additionally, AGI could potentially optimize chip design processes themselves, using advanced algorithms to explore design spaces far more efficiently than human engineers, potentially reducing development costs and time-to-market for new chip designs.

Power infrastructure: The development of more efficient power delivery networks is becoming crucial as chip complexity increases. Future directions include on-chip voltage regulators with higher efficiency and responsiveness, enabling finer-grained power management. Advanced materials like Gallium Nitride (GaN) and Silicon Carbide (SiC) are being explored for more efficient power conversion. Additionally, there's growing interest in photonic power delivery systems, which could potentially provide power to chips using light, offering benefits in terms of isolation and reduced electromagnetic interference.

Scaling transistor count: The push towards systems with 1 Trillion transistors is driving innovation in chip design and manufacturing. One of the most recent examples is the Nvidia Grace-Blackwell GB200 GPU with 208B transistors. It uses a custom 4nm process named 4NP (a variation of the standard N4P with, possibly, more metal layers). This massive scale introduces challenges in design complexity, verification, and yield management. To address these issues, advancements in AI-assisted design tools are being developed to handle the complexity of multi-billion-transistor systems. New lithography techniques, such as high-NA EUV, are being explored to enable even finer feature sizes. Additionally, vertical stacking technologies are being refined to increase transistor density without relying solely on planar scaling.

Memory structures and communication: Advancements in chiplet-to-memory and chiplet-to-chiplet communication are critical for realizing the full potential of modular chip designs. Novel memory architectures, such as Compute Express Link (CXL), are being developed to provide coherent, high-bandwidth connections between processors and memory pools. Optical interconnects are being explored for ultra-high bandwidth, low latency communication between chiplets. Additionally, advancements in packaging technologies, like TSMC's 3DFabric and Intel's Foveros, are enabling tighter integration of memory and logic chiplets, blurring the line between on-chip and off-chip memory.

These emerging trends and future directions highlight the dynamic and rapidly evolving nature of the semiconductor industry as it embraces chiplet and 3D IC technologies. The interplay between these trends will shape the future of computing, enabling new applications and pushing the boundaries of what's possible in electronic systems.

Industry collaboration and standards

The complexity of chiplet-based designs necessitates increased collaboration across the industry, ushering a new era of open innovation in semiconductor design. Standards like Universal Chiplet Interconnect Express (UCIe) are crucial in enabling a robust ecosystem where multiple players can contribute specialized chiplets. UCIe, backed by industry giants such as Intel, AMD, Arm, TSMC, and Samsung, defines a die-to-die interconnect standard that ensures interoperability between chiplets from different vendors. This standardization is pivotal in breaking down proprietary barriers and allowing for mix-and-match chiplet designs.

Initiatives like the Open Compute Project (OCP) and CHIPS Alliance are pushing for open-source hardware designs and verification tools, respectively. These collaborative efforts are democratizing chiplet technology, enabling smaller companies and academic institutions to participate in and contribute to cutting-edge semiconductor development. The industry can tackle complex challenges more effectively by pooling resources and expertise, from advanced packaging techniques to system-level integration issues.

TSMC and EDA vendors have come up with an open standard called the 3Dblox 2.0 to provide a uniform packaging solution for chiplets and memory for all 3D IC techniques available today. This will enable top-down design methodology and allow chiplet reuse across different products and different vendors. This collaborative approach is essential for driving innovation and reducing costs in the long term. It accelerates the development cycle by leveraging specialized expertise from various players, potentially leading to faster time-to-market for new technologies. Moreover, the ability to reuse and combine chiplets from different sources can significantly reduce design and manufacturing costs, making advanced semiconductor solutions more accessible across various market segments. As the chiplet ecosystem matures, we can expect to see a proliferation of innovative designs that push the boundaries of performance, energy efficiency, and functionality in ways that were previously unattainable with monolithic architectures.

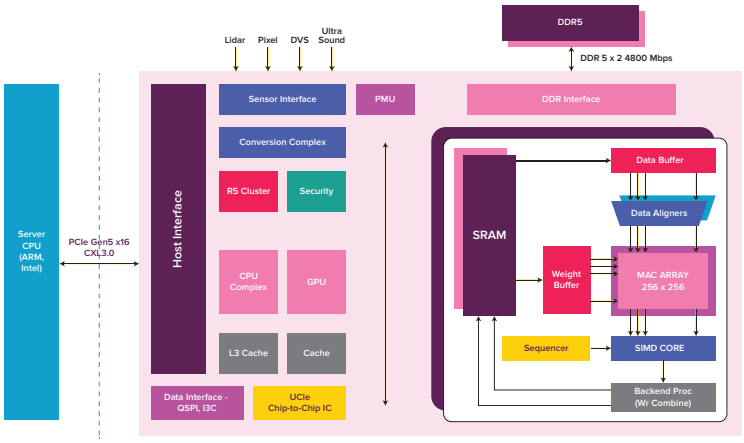
Partnering with Quest Global to demystify chiplets

The adoption of chiplets and advanced packaging technologies represents a paradigm shift in the semiconductor industry. While these technologies offer exceptional opportunities for performance improvement and functional integration, they also introduce significant challenges in platforms, SW/HW partitioning, design, manufacturing, and cost management. As the industry continues to push the boundaries of Moore's Law, success will depend on addressing these multifaceted challenges through innovative engineering solutions, standardization efforts, and collaborative ecosystems.

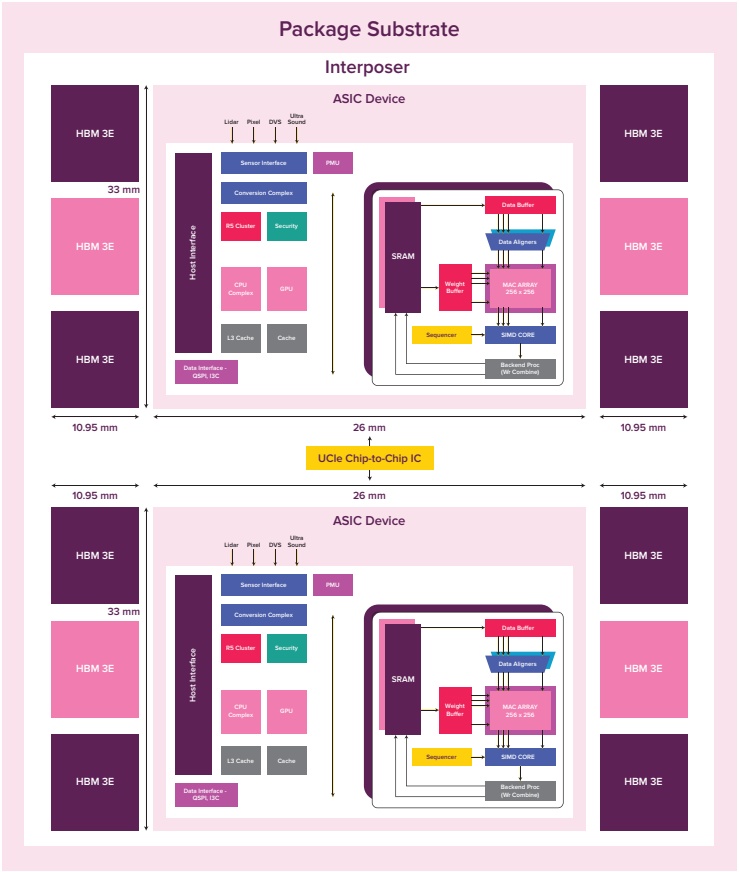
Companies that can navigate these complexities and deliver optimized chipset-based solutions will be well-positioned to lead in the next era of semiconductor technology. Quest Global's expertise in providing end-to-end solutions - from silicon to systems to cloud - is instrumental in helping companies capitalize on the immense opportunities of chiplets & 3D ICs technology. With Quest Global as a partner, semiconductor companies can confidently step into the future of chip design, armed with the knowledge, skills, and support needed to thrive in this new paradigm.

Quest Global, with our pioneering Silicon to Systems to Cloud (S2S2C) approach and extensive expertise in semiconductor engineering, is actively shaping the 3D transition. As a trusted partner to leading semiconductor companies, we leverage our deep domain knowledge across rail, automotive, aerospace, defense, hi-tech, communications, energy, medtech, healthcare, and semiconductors to offer holistic solutions. This cross-industry expertise allows us to provide invaluable insights and innovative approaches at every stage of chiplet and advanced packaging development. Our in-depth understanding of the entire value chain, from silicon design to systems integration and cloud solutions, enables us to address complex challenges with a unique, multidisciplinary perspective. This breadth of knowledge and experience positions us to deliver tailored, cutting-edge solutions that drive innovation and efficiency across the semiconductor landscape and its diverse applications.

Typical ASIC device



Interposer device packaging and manufacturing



For further information or queries, please reach out to us at info@quest-global.com