Heterogeneous Integration

From Packaging Innovation to the Architectural Foundation of the Chiplet Era

The Semiconductor Industry's New Scaling Strategy

Rozalia Beica
Field CTO Packaging Technologies, Rapidus Design Solutions
Contributor, IEEE EPS Heterogeneous Integration Roadmap (HIR), Chair of Wafer Level Packaging (WLP) Technical Working Group
Member of IEEE EPS 3D TSV and Substrate Technical Committees

Abstract
Heterogeneous integration (HI) has moved from an advanced packaging option to a central strategy for sustaining semiconductor innovation beyond the economic and physical limits of monolithic scaling. As artificial intelligence, high-performance computing, automotive electronics, communications, and edge intelligence demand more compute, memory bandwidth, energy efficiency, and technology diversity, the industry is increasingly shifting from scaling transistors to scaling systems. This article explains how HI has evolved from multi-chip modules and System-in-Package approaches into an architectural framework for chiplet-based design. It examines the economic logic of disaggregation, the rise of die-to-die interconnect standards such as UCIe, the impact of AI and high-bandwidth memory, and the growing importance of silicon-package co-design, thermal management, power delivery, testing, and ecosystem collaboration. The article concludes that the next phase of semiconductor innovation will be defined not only by the best transistors, but by the ability to integrate diverse technologies into optimized systems-of-chiplets.

Introduction: From Scaling Transistors to Scaling Systems

For more than five decades, semiconductor innovation was driven primarily by transistor scaling. Since Gordon Moore's 1965 observation, the industry has relied on increasing transistor density to deliver higher performance, lower power consumption, greater functionality, and lower cost per function [1]. This model created the foundation for the modern digital economy.

Today, however, the industry is operating at a new inflection point. Transistor scaling continues, but the economics and physics of leading-edge nodes have changed. Advanced technology development requires massive capital investment, EUV lithography, new device structures, increasingly complex design methodologies, and much tighter interactions between process technology, design, and manufacturing [2], [3]. Reticle limits, yield management, power density, memory bandwidth, and design complexity have become as important as raw transistor density. Continued innovations in device architecture, materials, lithography, and manufacturing remain essential because heterogeneous integration can only deliver maximum value when combined with advances in underlying silicon technology.

At the same time, emerging applications - AI training and inference, high-performance computing (HPC), autonomous systems, advanced communications, aerospace, defense, and cloud infrastructure - require unprecedented system performance. These platforms demand not only more compute capability, but also higher memory bandwidth, lower latency, lower energy per bit, better thermal paths, and integration of technologies that do not belong on the same process node.

The result is a strategic shift from scaling transistors to scaling systems. At the center of this transition is heterogeneous integration (HI). The Heterogeneous Integration Roadmap defines HI as the integration of separately manufactured components into a higher-level assembly that provides enhanced functionality and improved operating characteristics [4]. The definition is important because it reframes packaging from an end-of-line assembly step into a system architecture discipline.

Figure 1 summarizes this shift. It shows how semiconductor scaling is moving from transistor density and monolithic SoCs toward system-level integration, where chiplets, advanced packaging, memory proximity, interconnect, substrates, and ecosystem coordination become the new scaling vectors.

While today's heterogeneous systems are primarily implemented using 2.5D integration, the industry is already moving toward a future that combines chiplets, advanced packaging, and heterogeneous 3D integration. These approaches extend system scaling beyond traditional planar architectures by enabling both horizontal and vertical integration of specialized functions, creating new opportunities for performance, energy efficiency, and functional diversity.

Innovation is shifting from making one die denser to making many optimized dies work as one system

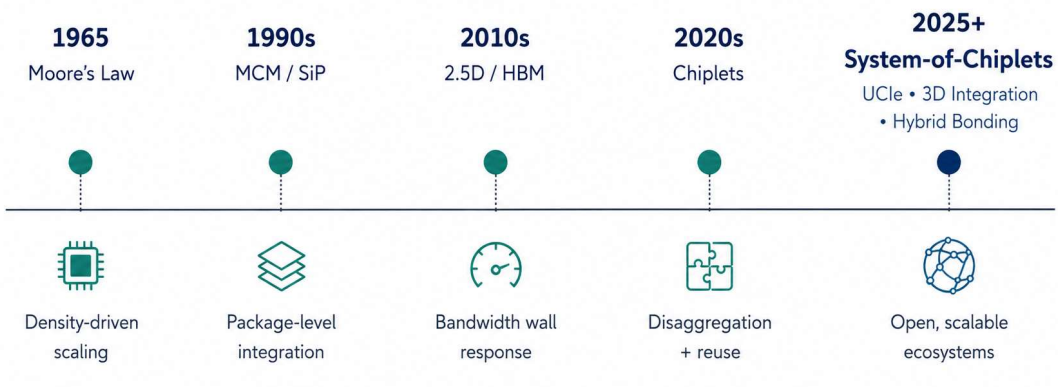

Figure 1. From transistor scaling to system scaling:

From Multi-Chip Modules to Heterogeneous Integration

Although heterogeneous integration is often viewed as a recent development, its roots extend back several decades and its role has changed dramatically. Early forms appeared in hybrid microelectronics and multi-chip modules (MCMs), which combined multiple integrated circuits on a common substrate to achieve improved electrical performance and reduce system size [6]. During the 1990s and 2000s, System-in-Package (SiP) approaches expanded the concept by integrating logic, memory, RF, passives, sensors, and other functions into compact modules [7].

As scaling challenges became increasingly evident during the late 2000s and 2010s, it became clear that future performance improvements would require more than transistor miniaturization, and that systems could not be optimized by placing every function on the most advanced logic node. Logic, SRAM, DRAM, RF, analog, power management, photonics, MEMS, and wide-bandgap power devices each have different materials, design rules, cost structures, and performance targets. Heterogeneous integration provides a practical way to combine these optimized technologies into a single system while allowing each function to be manufactured where it makes the most technical and economic sense.

The Heterogeneous Integration Roadmap (HIR), sponsored by IEEE EPS, SEMI, IEEE EDS, IEEE Photonics Society, and ASME EPPD, created a common language and framework for this transition and highlighted the

cross-disciplinary challenges that must be solved across design, packaging, materials, test, thermal, power delivery, and reliability [4], [5]. Heterogeneous integration was no longer viewed solely as a packaging activity. Instead, it became recognized as a strategic system-level approach involving multiple technologies, process nodes, materials, and ecosystem participants [4], [5]. The roadmap's message is clear: **HI is not just a packaging roadmap. It is a system integration roadmap.**

Why Heterogeneous Integration Became Necessary

The strongest drivers for heterogeneous integration are economic, architectural, and physical. At advanced nodes, large monolithic dies become increasingly expensive to design, fabricate, yield, and validate. Even when a monolithic SoC is technically possible, it may not be the best economic choice. Functions that do not benefit from the leading-edge node - such as analog I/O, RF, power management, embedded non-volatile memory, photonics, or specialized memory - can add cost and design risk without delivering proportional performance benefit.

Heterogeneous integration enables designers to partition the system across multiple dies, process nodes, and suppliers. Advanced logic can remain on the leading edge, while I/O, analog, power, memory, or security functions can be implemented on mature or specialized processes. This approach improves design reuse, reduces time-to-market, supports portfolio flexibility, and makes it easier to tailor products for different customers or market segments. While heterogeneous integration changes how systems are assembled, it does not eliminate the importance of advanced process technology. The most demanding AI and HPC applications continue to require improvements in transistor density, power efficiency, and performance. Heterogeneous integration and leading-edge silicon therefore evolve together, with system-level gains emerging from their combination rather than from either approach alone.

Importantly, heterogeneous integration does not diminish the value of advanced-node logic. On the contrary, leading-edge CPUs, GPUs, and AI accelerators increasingly depend on the performance, power efficiency, and transistor density delivered by advanced silicon, while heterogeneous integration enables those capabilities to be combined with complementary technologies.

Table 1 summarizes why different functions increasingly belong on different technologies:

Function	Typical technology choice	Reason
CPU / GPU / AI logic	Advanced CMOS	Highest logic density and performance
SRAM cache	Advanced CMOS, sometimes partitioned	Fast local memory, but scaling benefit is slowing
DRAM / HBM	Specialized memory processes	Bandwidth and capacity optimized independently
RF front end	RF SOI / SiGe / specialty CMOS / GaN	RF performance and isolation; GaN adds high power density and high-frequency capability
Power management	BCD / analog CMOS	Voltage conversion, robustness, cost
Silicon photonics	Photonic platforms	Optical I/O and co-packaged optics
Power electronics	SiC / GaN	High voltage, high efficiency, thermal robustness

The Chiplet Revolution

The most visible expression of heterogeneous integration is the rise of chiplets. Chiplets challenge the traditional assumption that a semiconductor system must be implemented as one large monolithic die. Instead, system functionality is partitioned into smaller building blocks - compute chiplets, I/O chiplets, memory interfaces, AI accelerators, security engines, photonic interfaces, or power-management functions - that are assembled using advanced packaging technologies [8], [10], [11].

Chiplets are attractive because they provide both technical and economic advantages. Technically, they allow designers to place memory closer to compute, shorten interconnect distances, increase bandwidth density, and combine technologies that would be difficult or inefficient to fabricate together. Economically, they improve yield and reuse. A defect on a very large monolithic die can destroy the entire product; a defect on one smaller die may only remove that die from the known-good-die pool. By partitioning a large system into smaller dies, companies can improve wafer utilization and product flexibility. The final cost advantage depends on the balance between higher die yield and added packaging, test, substrate, and assembly costs.

Although current chiplet implementations are predominantly arranged side-by-side within advanced packages, the long-term evolution of chiplet architectures is expected to include increasing levels of vertical integration. Heterogeneous 3D integration allows logic, memory, photonics, sensing, and specialized accelerators to be stacked using technologies such as hybrid bonding, wafer-to-wafer integration, and die-to-wafer assembly. These approaches further shorten communication distances, increase interconnect density, and reduce energy consumed by data movement.

From Wafer to Panel: Scaling the Manufacturing Format

Chiplets also change the physical format of advanced packaging. As more compute dies, HBM stacks, I/O dies, bridges, photonic interfaces, and power-delivery structures are integrated in a single package, package and interposer dimensions continue to grow. This creates a practical tension: the system-level benefits of disaggregation push packages larger, while lithography reticle limits, wafer geometry, carrier utilization, and warpage control constrain how economically those packages can be manufactured.

For high-performance AI and HPC devices, the transition from wafer-level to panel-level packaging is therefore becoming more than a cost-reduction topic. It is an enabling manufacturing transition. Panel processing provides a square or rectangular carrier format that can improve area utilization for very large interposers and substrates, reduce edge loss versus circular wafers, and support package dimensions that are increasingly difficult to scale within traditional wafer-based flows [21], [22].

This transition does not eliminate the challenges of advanced packaging. Panel-level manufacturing must still solve tight overlay, die shift, warpage, thermal-mechanical reliability, fine-line redistribution, inspection, repair, and yield-learning issues across larger formats. However, the strategic direction is clear: as the package becomes the system, manufacturing must scale not only in x-y interconnect density, but also in carrier size, substrate size and complexity, new infrastructure, and end-to-end process control.

As shown in Figure 2, the shift from 300 mm wafers to larger panel formats connects directly to the economics of chiplet-based integration. Reticle-size constraints make very large monolithic SoCs increasingly unattractive, particularly for AI and HPC systems where compute, memory, I/O, and accelerator functions must be integrated at very high density. Chiplets address part of this challenge by partitioning silicon into smaller, higher-yielding dies, while panel-level packaging addresses the next manufacturing bottleneck by enabling larger system-in-package assemblies with improved carrier utilization and the potential for lower cost per package at high volume.

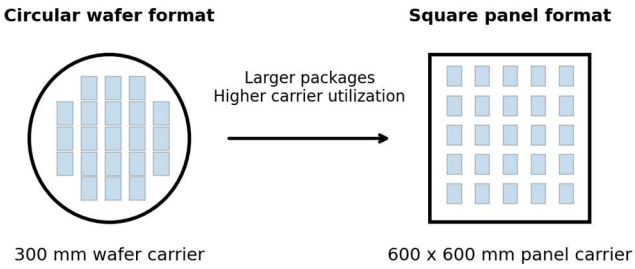

Figure 2. Wafer-to-panel transition for large heterogeneous systems.

Panel-level processing becomes increasingly attractive as chiplet-based AI and HPC packages grow beyond traditional reticle and wafer-utilization economics.

Interconnect: The New Scaling Frontier

In a monolithic SoC, much of the communication occurs through on-chip wiring. In a chiplet-based system, data must cross die boundaries while maintaining high bandwidth density, low latency, excellent signal integrity, low energy per bit, reliability, security, and manufacturability. Package-level interconnects therefore become extensions of on-chip wiring, and the package becomes part of the system fabric.

This system-level perspective is illustrated in Figure 3, which presents a conceptual System-of-Chiplets architecture integrating compute, memory, networking, storage, security, photonics, and I/O functions through a common die-to-die interconnect fabric. The figure highlights how interconnect has evolved beyond a physical connection between components to become the communications backbone that enables heterogeneous chiplets to operate as a cohesive system. As semiconductor architectures become increasingly modular and functionally diverse, the performance of the interconnect fabric—including latency, bandwidth, power efficiency, and protocol interoperability—plays a central role in determining overall system performance, scalability, and design flexibility. Advanced packaging technologies provide the physical foundation for this integration, enabling chiplets manufactured in different process nodes and technologies to be combined within a single package.

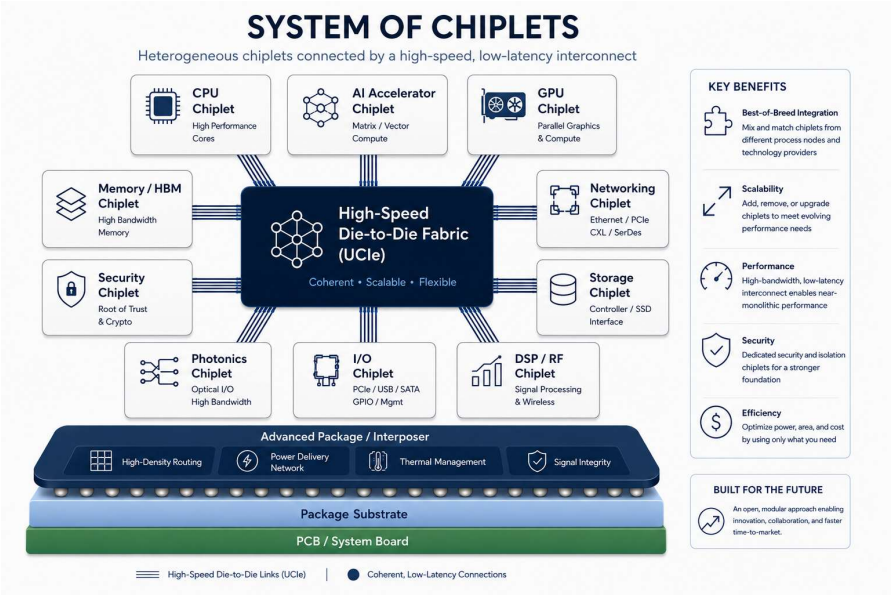

Figure 3. Representative System-of-Chiplets architecture integrating heterogeneous functional chiplets through a common high-speed die-to-die interconnect fabric.

This is why die-to-die interconnect standards are so important. The Universal Chiplet Interconnect Express (UCIe) specification was created to enable an open chiplet ecosystem by defining a complete die-to-die interconnect, including the physical layer, protocol stack, software model, and compliance testing [12]. UCIe 2.0 added support for 3D packaging and manageability in 2024, including a standardized management architecture that enables testing, telemetry, debug, and system-level monitoring across multiple chiplets, while UCIe 3.0, released in August 2025, increased performance to 48 and 64 GT/s and added further power-efficiency and manageability enhancements such as runtime recalibration and extended sideband reach [13, 30].

The growing adoption of UCIe also reflects a broader architectural transition. Future systems will increasingly combine both 2.5D and 3D integration approaches, requiring standardized communication frameworks capable of operating across multiple packaging topologies. Open die-to-die standards therefore become increasingly important as heterogeneous systems evolve from collections of chiplets into highly integrated three-dimensional architectures.

Open standards do not eliminate the need for proprietary optimization. Leading AI and HPC products will continue to use highly optimized internal fabrics. However, standards such as UCIe are important because they reduce ecosystem friction, support design reuse, and create a path toward interoperable chiplets sourced from multiple suppliers. This may ultimately change the structure of the semiconductor value chain as profoundly as standard interfaces changed the PC and server ecosystems.

AI as the Accelerator of Heterogeneous Integration

Artificial intelligence has become the strongest near-term catalyst for heterogeneous integration. AI systems are fundamentally bandwidth- and energy-limited. Training and inference require massive matrix operations, high memory bandwidth, low-latency data movement, and efficient scaling across accelerators. The memory wall is now one of the central constraints in AI hardware design.

Modern AI accelerators combine advanced logic, high-bandwidth memory (HBM), interposers or bridge technologies, advanced substrates, high-current power delivery, sophisticated thermal solutions, and high-speed networking. NVIDIA's Blackwell architecture illustrates the direction of travel: the company describes Blackwell GPUs as using two reticle-limited dies connected by a 10 TB/s chip-to-chip interconnect, with 208 billion transistors on a custom TSMC 4NP process [14]. The key point is not one product specification, but the architectural pattern: performance is increasingly delivered by package-scale integration.

Future AI systems are expected to push integration density even further through advanced memory stacking, hybrid-bonded logic-memory interfaces, and co-packaged optical interconnects. As AI infrastructure scales from package-level integration toward rack-scale systems, the cost and power associated with data movement increasingly become dominant design constraints. Heterogeneous integration therefore extends beyond the package itself and increasingly influences system architecture at the board, server, and data-center levels.

As power envelopes rise and memory bandwidth becomes a system bottleneck, heterogeneous integration becomes inseparable from thermal design, power delivery, and system-level co-optimization. Advanced packaging is no longer a passive carrier. Together with advances in silicon technology, it has become a performance-critical platform for compute-memory proximity, bandwidth density, signal integrity, power integrity, and cooling.

Key takeaway

The AI era is turning advanced packaging into a compute architecture problem. The limiting questions are no longer only “How many transistors?” but also “How close is memory?”, “How much bandwidth per watt?”, “How is heat removed?”, and “How quickly can the architecture be iterated?”

From Packaging to Architectural Co-Design

Perhaps the most important change is organizational. Historically, packaging often occurred after chip design. The package was selected to meet I/O, thermal, mechanical, and cost requirements that were largely downstream from the silicon architecture. In the chiplet era, that sequence no longer works.

Packaging decisions now influence system architecture from the earliest design stages. Chip partitioning, bump pitch, bridge technology, interposer choice, substrate material, HBM configuration, thermal interface materials, power delivery, warpage control, test access, known-good-die strategy, and reliability qualification must be considered together. A poor partitioning decision can erase the economic benefit of chiplets. A weak thermal or power-delivery strategy can limit product performance even when the silicon is excellent.

This creates a new discipline: silicon-package-system co-design. In this model, the package is not simply an enclosure. It is an architectural layer that determines bandwidth, energy efficiency, manufacturability, yield, cost, and time to market.

The growing complexity of heterogeneous systems is also driving new requirements for design automation. Future EDA platforms must optimize not only silicon implementation, but also chiplet partitioning, placement, thermal interactions, power-delivery networks, manufacturing constraints, and package architecture. Artificial intelligence is expected to play an increasingly important role in exploring this expanded design space and accelerating system-level optimization.

A critical research and standardization need is the development of interoperable Assembly Design Kits (ADKs) and Chiplet Design Kits (CDKs). Analogous to the role of process design kits (PDKs) in monolithic IC design, these kits must provide machine-readable assembly rules and validated electrical, thermal, mechanical, reliability, and test models across dies, interposers, substrates, bonding technologies, and manufacturing flows. OCP's Open Domain-Specific Architecture work, including its ADK framework and Chiplet Data Exchange (CDXML), is helping define information exchange for chiplet reuse and integration [31], [32]. Standardized design collateral is essential if EDA tools are to support early multi-physics analysis, automated design-rule checking, portable verification, and signoff across multiple foundries, OSATs, and chiplet suppliers. Without trusted, portable models, an open chiplet marketplace will remain difficult to scale.

Heterogeneous Integration as an Ecosystem Strategy

Heterogeneous integration also changes the semiconductor supply chain. Monolithic scaling favored companies that could optimize within relatively linear flows: architecture, design, wafer fabrication, packaging, test, and system integration. As shown in Figure 4, future heterogeneous systems require much tighter collaboration among foundries, OSATs, EDA suppliers, IP providers, equipment manufacturers, substrate suppliers, materials companies, test houses, standards bodies, cloud providers, and system companies.

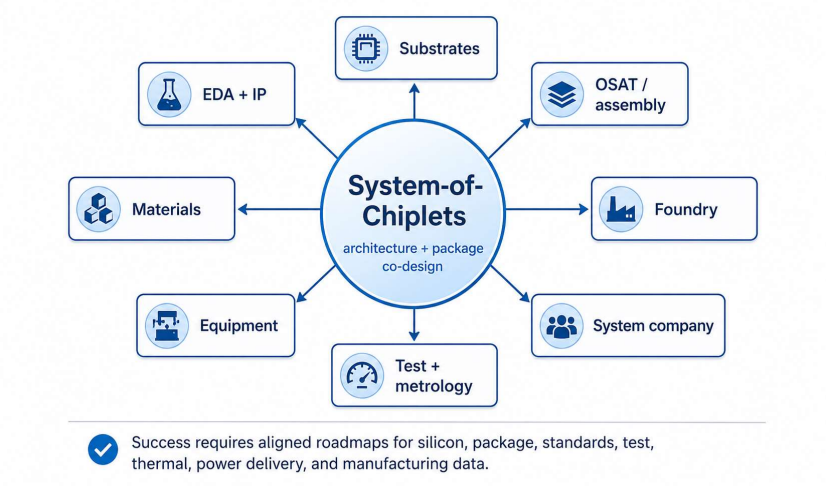

Figure 4. Heterogeneous integration depends on ecosystem-level coordination.

The challenge is no longer simply building transistors. It is orchestrating ecosystems. A successful chiplet product requires aligned design rules, interface standards, thermal models, power-delivery assumptions, substrate capacity, test strategies, reliability data, and supply-chain accountability. This is why industry roadmaps and open standards are increasingly important: they create shared assumptions that allow different parts of the ecosystem to innovate together [4], [5], [29].

This coordination requires standards at several complementary layers. UCIe defines a broad die-to-die connectivity framework, while Bunch of Wires (BoW) and the Advanced Interface Bus (AIB) provide open physical-interface approaches [33], [34]. Physical integration data must also be exchanged consistently: the emerging IEEE P3537 project builds on the 3Dblox description language for components and 2.5D/3D stacks [35]. Testability is equally important. IEEE 1838 defines test-access architecture for dies used in 3D stacks, and the developing IEEE P3405 standard addresses chiplet-interconnect test, repair, redundancy, and high-volume-manufacturing support [36], [37]. Together with ADKs and CDKs, these standards can reduce integration risk and enable chiplets from different suppliers to be designed, assembled, tested, and qualified within a more open ecosystem.

Advanced packaging is also becoming a strategic differentiation layer. The market is no longer defined only by OSAT assembly capacity or foundry wafer output. Foundries, OSATs, substrate manufacturers, memory suppliers, EDA companies, equipment vendors, and system companies are all increasing their activity because advanced packaging increasingly determines performance, power, bandwidth, form factor, time to market, and supply-chain resilience.

This creates a more crowded and more competitive landscape. Foundries are expanding proprietary advanced-packaging platforms, OSATs are investing in fan-out, 2.5D, 3D, and panel-level capabilities, substrate suppliers are pursuing glass and next-generation organic platforms, and system companies are becoming more involved in package and substrate architecture decisions. In this environment, differentiation shifts from a single technology module to the ability to deliver a complete co-optimized platform: silicon, interconnect, substrate, assembly, thermal solution, test, reliability, software-visible telemetry, and ecosystem access.

In parallel, the industry is investing heavily in glass-core substrates, which offer potential advantages in dimensional stability, routing density, and scalability for future large-area heterogeneous systems.

The strategic implication is significant: advanced packaging is becoming one of the few places where companies can still create meaningful system-level differentiation when transistor scaling alone is no longer

sufficient. Capacity, yield learning, design enablement, known-good-die strategy, substrate availability, and ecosystem trust may become as decisive as process-node leadership.

A New Manufacturing Paradigm for Heterogeneous Systems

As heterogeneous integration evolves and systems become more complex, the industry challenge extends far beyond integrating additional functions into a package; it must also rethink manufacturing models. Increasing system complexity is fundamentally changing how semiconductor innovation occurs.

Historically, semiconductor development followed relatively sequential workflows in which design, fabrication, packaging, testing, and system integration were optimized independently. In the heterogeneous integration era, these disciplines increasingly require simultaneous optimization.

A leading-edge AI accelerator today may combine advanced logic, HBM, specialized AI engines, photonic interfaces, power-management devices, advanced substrates, embedded bridges or interposers, and sophisticated thermal-management solutions. These components may come from different process nodes, technology platforms, and supply-chain partners [4], [5]. Sequential handoffs between design, wafer fabrication, packaging, test, and system validation are too slow for markets where AI models, workloads, and system architectures evolve continuously. The integration problem is therefore both technical and operational. As complexity increases, innovation cycles must accelerate.

This creates a new competitive requirement: innovation speed. The ability to rapidly prototype, evaluate, iterate, and optimize heterogeneous systems may become as important as process-node leadership.

This trend is driving renewed interest in manufacturing models that enables closer integration between front-end silicon development and advanced packaging. While wafer fabrication and packaging have traditionally operated as separate organizations and business models, future heterogeneous systems increasingly require co-optimization across silicon, package architectures, interconnects, thermal management, power delivery, system design, and manufacturing [4], [5].

Rapidus is an example of this emerging approach. Established specifically to address next-generation semiconductor challenges, the company has articulated a vision that combines advanced logic manufacturing, advanced packaging, and system-level integration within a common framework [18]. This model is particularly relevant because future performance improvements will increasingly emerge from system-level optimization. Equally important is speed. The AI era is characterized by unprecedented rates of innovation. New models, new architectures, and new workloads emerge continuously, placing pressure on semiconductor development cycles. Manufacturers capable of enabling rapid learning cycles, accelerated prototyping, and shorter design-to-silicon timelines may provide significant advantages to customers developing increasingly complex systems.

Rapidus has publicly emphasized rapid cycle-time objectives, extensive manufacturing automation, and the ambition to become an AI-enabled foundry. Combined with co-located advanced-node manufacturing and advanced packaging capabilities, these strengths are expected to support faster system-level iteration and more agile exploration of heterogeneous architectures [18]. Its Innovative Integration for Manufacturing (IIM) facility in Chitose, Hokkaido, together with its Rapid and Unified Manufacturing Service (RUMS), represents a next-generation foundry model focused on advanced logic, application-specific devices, and chiplet-based solutions. By integrating front-end and back-end manufacturing and leveraging accelerated turnaround times [15], [16] alongside AI-enabled design and optimization tools [17], [18]. Rapidus aims to enable a more responsive and efficient pathway from design to manufacturing. This model recognizes that future competitiveness depends not only on advanced packaging and system integration, but also on sustained leadership in transistor technology, process and manufacturing innovation and excellence.

The broader significance is not only one company's roadmap. It is the direction of the industry. Future semiconductor competitiveness will increasingly depend on how effectively companies connect front-end process development, package architecture, chiplet design, AI-enabled design automation, manufacturing data, test, and system validation into a faster learning loop [29].

In many respects, heterogeneous integration and artificial intelligence are mutually reinforcing. AI applications drive demand for increasingly complex heterogeneous systems, while AI-enabled design and manufacturing approaches create opportunities to accelerate development of those same systems.

The semiconductor industry's next chapter may therefore be defined not only by the technologies being integrated, but also by how quickly and effectively they can be brought together into optimized systems.

Looking Ahead: The System-of-Chiplets Era

The semiconductor industry is entering what may ultimately become known as the **System-of-Chiplets era**. The dominant design question is shifting from "How much functionality can be placed on one die?" to "What is the best system architecture across dies, packages, boards, and racks?"

Historically, innovation focused on System-on-Chip (SoC) architectures. Future performance scaling will increasingly come from architectural innovation, advanced packaging, memory proximity, optical and electrical interconnects, chiplet interoperability, power and thermal co-design, and ecosystem collaboration. Transistor scaling remains essential, but it is no longer sufficient by itself.

In this environment, packaging becomes a platform. It enables new product architectures, new business models, new ecosystem relationships, and new ways of optimizing performance per watt and cost per function. The most successful companies will be those that combine strong silicon technology with strong integration capability and ecosystem leadership.

The Next Frontier: Heterogeneous 3D Integration

While much of today's heterogeneous integration ecosystem is built around 2.5D packaging technologies, the next major scaling vector is expected to come from heterogeneous 3D integration. Technologies such as hybrid bonding, wafer-to-wafer stacking, die-to-wafer assembly, and advanced memory stacking enable much higher interconnect densities than conventional microbump-based approaches while significantly reducing interconnect length and energy per bit.

Heterogeneous 3D integration extends the chiplet concept beyond side-by-side assembly by enabling vertically integrated systems that combine logic, memory, photonics, sensing, power management, and specialized accelerators within tightly coupled stacks. These architectures offer a pathway toward near-monolithic performance while preserving the flexibility, reuse, and economic advantages of heterogeneous integration.

Significant challenges remain. Thermal management, power delivery, known-good-die strategies, repairability, test access, manufacturing yield, and supply-chain coordination become increasingly complex as integration density rises. Nevertheless, industry roadmaps consistently identify heterogeneous 3D integration as one of the most important long-term directions for semiconductor scaling [19], [25].

Looking further ahead, heterogeneous integration may extend beyond individual packages into tightly coupled systems spanning packages, boards, servers, and AI racks.

Current research programs illustrate both the opportunity and the remaining gaps. DARPA's Next-Generation Microelectronics Manufacturing (NGMM) program is establishing a U.S.-based center for accessible prototyping and manufacturing of advanced 3D heterogeneous microsystems [38], [39]. Its MiniTherms3D program focuses on removing heat from densely stacked, high-power logic, RF, and other functional tiers—one of the principal barriers to extending 3DHI beyond relatively low-power stacks [40]. These efforts highlight that progress in 3DHI depends on coordinated advances in manufacturing access, design infrastructure, materials, thermal management, test, and standards.

Author's Reflection: From Packaging to Platform

Having spent almost three decades working in advanced packaging, heterogeneous integration, and semiconductor ecosystem development, I have had the privilege of witnessing one of the most significant transformations in our industry.

When I began my career, packaging was often viewed as the final step in semiconductor manufacturing. Today, packaging directly influences system performance, power efficiency, functionality, manufacturability, cost, reliability, and supply-chain strategy. The emergence of chiplets and heterogeneous integration has accelerated this transformation and placed packaging at the center of semiconductor innovation.

My involvement with the Heterogeneous Integration Roadmap reinforced an important lesson: no single technology, company, or region will define the future of semiconductor innovation alone. The future belongs to ecosystems - and to the organizations capable of aligning silicon, package, standards, manufacturing, test, and system architecture around a common objective.

What is most exciting is that this transformation is still in its early stages. Chiplets, heterogeneous 3D integration, hybrid bonding, UCIe, co-packaged optics, glass-core substrates, panel-level packaging, advanced memory architectures, AI-driven design tools, and new manufacturing models are only beginning to reveal their full potential. For the first time in decades, semiconductor innovation is no longer driven exclusively by transistor scaling. Instead, progress increasingly depends on the combination of continued silicon innovation and the intelligent integration of diverse technologies into highly optimized systems.

Conclusion

Heterogeneous integration has evolved from a packaging methodology into the architectural foundation of modern semiconductor systems. Originally associated with integrating multiple functions within a package, it now encompasses advanced packaging, chiplet architectures, open standards, multi-node optimization, supply-chain collaboration, and system-level innovation.

The rise of chiplets and interconnect standards has accelerated this transition by creating new pathways for performance scaling and design reuse, while reshaping semiconductor business models and ecosystems. AI and HPC have made the transition urgent by pushing the limits of memory bandwidth, power delivery, thermal management, and manufacturing capacity.

In the emerging System-of-Chiplets era, success will belong not only to those who build the best transistors, but also to those who build the strongest integration platforms and ecosystems. Heterogeneous integration is therefore not simply the future of packaging; it is one of the foundations of the future semiconductor industry.

The next stage of heterogeneous integration will be defined by three simultaneous transitions: from chips to systems, from two-dimensional integration to heterogeneous 3D architectures, and from wafer-centric packaging flows to scalable panel- and substrate-centric manufacturing ecosystems. Future innovation will increasingly depend on the industry's ability to co-optimize architecture, silicon, packaging, interconnect,

power delivery, thermal management, manufacturing, and software as a unified system. Technologies such as hybrid bonding, co-packaged optics, glass-core substrates, AI-assisted design automation, and heterogeneous 3D integration are expected to become important contributors to this next phase of scaling.

References

[1] G. E. Moore, "Cramming More Components onto Integrated Circuits," Electronics, vol. 38, no. 8, pp. 114-117, 1965.

[2] International Roadmap for Devices and Systems (IRDS), 2024 Edition.

[3] National Academies of Sciences, Engineering, and Medicine, Heterogeneous Integration: Achieving Use from Chips to Systems, Washington, DC: National Academies Press, 2019.

[4] IEEE Electronics Packaging Society and SEMI, Heterogeneous Integration Roadmap, 2019-2025 editions.

[5] IEEE EPS, Heterogeneous Integration Roadmap: System Integration and Architecture chapters.

[6] J. H. Lau, Multi-Chip Modules and Related Technologies, McGraw-Hill, 1994.

[7] J. H. Lau, System-in-Package: Design and Assembly Process, McGraw-Hill, 2011.

[8] DARPA Microsystems Technology Office, CHIPS Program documentation.

[9] IEEE Transactions on Components, Packaging and Manufacturing Technology, special issues on heterogeneous integration.

[10] AMD technical publications on chiplet architectures and EPYC processor development.

[11] Intel Corporation, Ponte Vecchio and multi-tile architecture technical disclosures.

[12] UCIe Consortium, UCIe specifications and white papers.

[13] UCIe Consortium, "UCIe Consortium Releases 3.0 Specification With 64 GT/s Performance and Enhanced Manageability," Aug. 2025.

[14] NVIDIA, "NVIDIA Blackwell Architecture," product architecture overview.

[15] Rapidus Corporation, "Business and Technologies: RUMS Rapid and Unified Manufacturing Service."

[16] Rapidus Corporation, corporate technology and IIM announcements, 2026.

[17] Rapidus Corporation, public statements on AI design tools for advanced semiconductor manufacturing, 2025-2026.

[18] Rapidus Corporation, public funding and 2 nm mass-production target announcements, 2026.

[19] IMEC, "CMOS 2.0: Redefining the Future of Scaling Through Heterogeneous 3D System Integration," 2025.

[20] Semiconductor Industry Association, State of the Semiconductor Industry Report, 2024-2025.

[21] Semiconductor Engineering, "The Rise of Panel-Level Packaging," July 2025.

[22] Yole Group, "Is Panel-Level Packaging Finally Emerging?" March 2025.

[23] Rapidus Corporation, "NEDO Approves Rapidus' FY2026 Plan and Budget for 2nm Semiconductor Projects," April 2026.

[24] IBM, "Rapidus and IBM Expand Collaboration to Chiplet Packaging Technology for 2nm-Generation Semiconductors," June 2024.

[25] IEEE EPS Heterogeneous Integration Roadmap, 3D Integration Chapter.

[26] IMEC, 3D System Integration and Hybrid Bonding Roadmap.

[27] Intel Corporation, Foveros Direct and Hybrid Bonding Technical Briefs.

[28] Optica Industry Consortium, Co-Packaged Optics Roadmap for AI Infrastructure.

[29] P. R. Bressler, M. Töpper, and P. Ramm, “Europe’s pilot line for advanced packaging and heterogeneous integration of electronic components and systems (APECS),” *Nature Reviews Electrical Engineering*, vol. 2, pp. 4–5, 2025, doi: 10.1038/s44287-024-00134-6.

[30] UCIe Consortium, "UCIe 2.0 Specification Overview," 2024. Available: https://www.uciexpress.org/specifications

[31] Open Compute Project, "Assembly Design Kits (ADK)," Open Chiplet Economy technical contribution, 2025. Available: https://www.opencompute.org/documents/assembly-design-kits-adk-2-pdf

[32] Open Compute Project, "Chiplet Data Exchange in XML Format (CDXML)," Open Domain-Specific Architecture Chiplet Design Exchange Workstream. Available: https://www.opencompute.org/chiplets/31/chiplet-data-exchange-in-xml-format-cdxml

[33] Open Compute Project, "Bunch of Wires (BoW) PHY Specification 2.0." Available: https://www.opencompute.org/chiplets/26/bunch-of-wires-bow-phy-specification-20

[34] CHIPS Alliance, "Advanced Interface Bus (AIB) Specification." Available: https://github.com/chipsalliance/AIB-specification

[35] IEEE Standards Association, "IEEE P3537: Standard for a Description Language for 2.5D/3D Components and Systems," project page. Available: https://standards.ieee.org/ieee/3537/11871/

[36] IEEE Standards Association, IEEE Std 1838-2019, "IEEE Standard for Test Access Architecture for Three-Dimensional Stacked Integrated Circuits," 2019. Available: https://standards.ieee.org/ieee/1838/5073/

[37] IEEE Standards Association, "IEEE P3405: Standard for Chiplet Interconnect Test and Repair," project page. Available: https://standards.ieee.org/ieee/3405/11398/

[38] DARPA, "Next-Generation Microelectronics Manufacturing (NGMM)," program overview. Available: https://www.darpa.mil/research/programs/next-generation-microelectronics

[39] DARPA, "DARPA Brings Next-Gen U.S. Microelectronics Manufacturing Closer to Reality," July 18, 2024. Available: https://www.darpa.mil/news/2024/next-gen-microelectronics-manufacturing

[40] DARPA, "MiniTherms3D," program overview. Available: https://www.darpa.mil/research/programs/Minitherms3D