

Power and Thermal Constraints in AI Accelerator Packages

July 17, 2026 | Chetan Arvind Patil, Marvell Technology

Estimated reading time: 5 minutes



Chetan Patil

For more than five decades, semiconductor progress was driven primarily by transistor scaling. Moore's Law provided a predictable path for increasing transistor density, while process technology advancements consistently improved performance, power efficiency, and manufacturing cost. During this period, package development evolved alongside silicon, supporting higher pin counts, faster interfaces, and improved thermal performance, but it rarely dictated overall system capability.

As processors became more powerful, package power gradually increased, requiring better heat spreaders, improved substrates, and more efficient cooling

solutions. Even so, power delivery and thermal management remained largely incremental engineering challenges because performance gains originated from advances in silicon technology. The package functioned primarily as an electrical and mechanical interface rather than a performance enabler.

The industry also benefited from a relatively straightforward system architecture. Most processors were implemented as monolithic dies with external memory connected through package pins and PCB traces. As a result, package design focused on enabling electrical connectivity, protecting the silicon, and supporting manufacturability. Although package technologies evolved to accommodate higher I/O counts and improved signal integrity, they rarely influenced processor architecture or overall system performance.

This long-standing relationship began to change as traditional transistor scaling matured. At the same time, leading semiconductor manufacturers are advancing device technologies through innovations such as Intel's 18A and 18A-P process nodes, TSMC's A16 technology, and IBM's recent demonstrations of sub-1 nm transistor research. These advancements extend silicon performance and efficiency, but they also depend on sophisticated manufacturing technologies, including High-NA Extreme Ultraviolet (EUV) lithography, more complex process integration, and tighter design rules.

The result is substantially higher manufacturing complexity, capital investment, and development costs, while system-level performance gains become increasingly difficult to achieve through process scaling alone. Consequently, semiconductor companies have increasingly complemented silicon innovation with architectural advances, heterogeneous integration, and advanced packaging to deliver the next generation of computing performance.

As packaging assumed a greater role in integrating compute, memory, analog, mixed signal, and power functions, it evolved from a supporting technology into a critical system enabler. Performance was no longer determined solely by transistor density, but increasingly by how efficiently the package could distribute power, transfer data, and dissipate heat. This marked the beginning of the industry's transition from transistor-centric scaling toward package-centric system design.

The Transition From Compute Centric to Power and Thermal Scaling

As advanced packaging became the foundation for system-level integration, the industry's design priorities also began to change. The challenge was no longer limited to integrating more transistors or connecting multiple silicon devices within a package. Instead, the focus shifted toward delivering increasing amounts of power while efficiently removing the heat generated by highly integrated computing systems. Package architecture became central to sustaining performance rather than simply enabling integration.

AI accelerated this transition. Modern AI accelerators achieve higher performance by integrating more compute engines, larger memory capacity, and significantly higher memory bandwidth within a single package. While this architecture dramatically increases computational throughput, it also drives unprecedented increases in power consumption and heat generation.

Table 1: Power and thermal constraints in AI accelerator packages

Characteristic	Traditional Processor Era	AI Accelerator Era
Package Power	10-150 W	500-1500+ W
Memory Location	Off-package DDR	Integrated HBM
Package Architecture	Single die	Multi-die heterogeneous integration
Current Demand	Moderate	Extremely high
Cooling Method	Air cooling	Advanced air and liquid cooling
Primary Limitation	Compute performance	Power and thermal density
Package Function	Electrical interface	System integration platform

As compute density increases, every additional processing element requires more power to be delivered and more heat to be dissipated. Advanced packaging technologies, including high bandwidth memory (HBM), chiplets, silicon interposers, and dense die-to-die interconnects, enable this level of integration but also concentrate electrical and thermal loads into a much smaller physical footprint. The package has consequently become the primary location where power delivery and thermal management challenges converge.

The industry is therefore transitioning from a compute-centric scaling model to one increasingly governed by power and thermal constraints. Future AI performance will depend not only on adding more computational resources but on how efficiently advanced packages can distribute power, maintain electrical integrity, and dissipate heat. In this new era, power and thermal scaling have become as important as compute scaling itself.

Reasons Accelerator Packages Create Extreme Power and Thermal Challenges

Modern AI accelerators consume significantly more power than traditional processors while operating at increasingly lower voltages to improve energy efficiency. This combination requires packages to deliver exceptionally high current with minimal electrical loss while maintaining stable voltage across multiple compute and memory devices. As package power rises, efficient power delivery has become a fundamental requirement for achieving consistent performance and reliability.

The challenge is compounded by the dynamic nature of AI workloads. Unlike conventional

Premier IPC Training & Certification

Also offered in Spanish!

2026 Course Calendar

Earn credentials on 6 IPC certification programs including the 620 Space Addendum



CLICK FOR COURSE SCHEDULE

BLACKFOX
Premier Training & Certification

INSULECTRO

ADVANCED MATERIALS FOR ADVANCED ELECTRONICS

Breakthrough Metal Organic Decomposition Chemistry

Learn more >

949.587.3200 insulelectro.com

PCB can't take the heat?

Electronics Protection

Advanced Materials for Long-Term Reliability

LEARN MORE

MacDermid Alpha

ASC

Where Advanced Packaging Meets Proven PCB Manufacturing

mks

NEW HANDBOOK RELEASE

Process Technologies in ADVANCED PACKAGING

This book introduces the basics of Semiconductor Advanced Packaging, its process technologies, key products and applications that Optimize the Interconnect.

across thousands of processing elements. These sudden changes create large fluctuations in power demand that must be supported without affecting device operation. Maintaining stable electrical performance under these conditions has become more difficult as both compute density and package complexity grow.

The integration of HBM further intensifies both power and thermal challenges. Placing HBM stacks adjacent to compute dies dramatically increases memory bandwidth while reducing communication latency. However, this tightly integrated architecture also concentrates power consumption within a confined package area. Heat generated by the compute dies directly influences HBM operating temperatures, while the memory itself contributes additional thermal loading, creating strong thermal interactions throughout the package.

Increasing package complexity amplifies these challenges even further. Modern AI accelerator packages combine multiple compute chiplets, HBM stacks, silicon interposers, advanced substrates, and dense die-to-die interconnects within a single package. The result is a highly heterogeneous thermal environment with multiple localized hotspots, each with distinct power densities and heat dissipation characteristics. As AI workloads continuously change during operation, hotspot locations shift dynamically, making thermal management increasingly complex. Consequently, leading AI accelerator packages are approaching or exceeding 1000W Thermal Design Power (TDP), pushing conventional air cooling beyond its practical limits and accelerating the adoption of cold plates, direct liquid cooling, and advanced thermal interface materials (TIMs) to sustain performance and reliability.

The consequences extend well beyond electrical performance and temperature. Large thermal gradients introduce mechanical stresses across dies, substrates, solder joints, and interconnect structures, while repeated thermal cycling accelerates material fatigue throughout the package. Consequently, power delivery, thermal behavior, and mechanical reliability have become tightly coupled engineering challenges. Successfully scaling future AI accelerators will require package designs that optimize these factors together rather than addressing each independently.

Power and Thermal Constraints Will Define Future AI Package Scaling

In summary, power and thermal management are no longer secondary considerations in AI accelerator development. As compute density, memory bandwidth, and heterogeneous integration continue to increase, the package has become the point where electrical power, thermal energy, and system performance converge. Every increase in computational capability brings higher current demand, greater heat generation, and more stringent requirements for package design.

This shift is changing the way AI accelerators are engineered. Simply integrating more compute chiplets or larger HBM configurations is no longer sufficient to improve overall system performance. Package architecture must simultaneously support efficient power delivery, effective heat dissipation, and reliable operation across increasingly complex heterogeneous systems.

Consequently, package design has evolved into a multidisciplinary engineering effort in which electrical, thermal, optical, and mechanical considerations are optimized together. Advanced substrates, TIMs, cooling technologies, and package layouts are no longer independent design elements but interconnected components that collectively determine package performance and reliability.

The semiconductor industry, has long relied on transistor scaling to drive innovation. However, AI accelerators are demonstrating that package-level power delivery and thermal management have become equally important. Advanced packaging is no longer simply connecting silicon devices, it is enabling them to operate efficiently, reliably, and at the performance levels demanded by modern AI workloads.

Share on:



Subscribe

Stay ahead of the technologies shaping the future of electronics with our latest newsletter, **Advanced Electronics Packaging Digest**. Get expert insights on advanced packaging, materials, and system-level innovation, delivered straight to your inbox.

[Subscribe now to stay informed, competitive, and connected.](#)

Suggested Items

[Hitachi Energy, Eve Air Mobility Accelerate Electrification Infrastructure for Urban Air Mobility](#)

07/17/2026 | Hitachi Energy

Hitachi Energy, a global technology leader in electrification, and Eve Air Mobility, an Embraer-backed electric vertical takeoff and landing (eVTOL) aircraft manufacturer, have signed a Memorandum of Understanding to advance the electrical infrastructure needed to support urban air mobility at scale.

[Infineon, LS ELECTRIC Team Up on DC Power Solutions for AI Data Centers](#)

07/16/2026 | Infineon

Infineon Technologies AG and LS ELECTRIC have signed a Memorandum of Understanding (MoU) to collaborate on high-efficiency direct current (DC) power infrastructure solutions for AI data centers and next-generation power grids.

[Gartner Identifies the Companies to Beat in the AI Semiconductor Vendor Race](#)

07/16/2026 | Gartner, Inc.

To meet the increasing AI demands for computing power, the competition is intensifying among the Companies to Beat in the AI semiconductor vendor race, according to Gartner, Inc., a business and technology insights company.

[The AI Data Center Boom: Technology, Capacity, and Society](#)

07/17/2026 | Marcy LaPort, I-Connect007

AI begins with advanced chips, but delivering AI at scale depends on an ecosystem of technologies that extends far beyond the semiconductor. In this interview, Andrew Bradner, senior vice president of Schneider Electric's cooling business, explains how AI is transforming data center design, why liquid cooling has become essential, and what these shifts mean for advanced packaging, power delivery, thermal management, and the broader electronics supply chain supporting next-generation AI infrastructure.

[New York Imposes One-Year Moratorium on New AI Data Centers](#)

07/15/2026 | I-Connect007 Editorial Team

New York has become the first U.S. state to impose a temporary statewide moratorium on new hyperscale data centers, following an executive order signed July 14 by Gov. Kathy Hochul.



- DESIGN007 WEEK
- MILAERO007 WEEK
- ADVANCED ELECTRONICS PACKAGING DIGEST
- I-CONNECT007 MAGAZINE
- SMT007 MAGAZINE
- PCB007 MAGAZINE
- DESIGN007 MAGAZINE

[About Us](#) | [Your Privacy](#) | [Contact Us](#) | [Site Map](#) | [Media Kit](#) | [RSS](#)