intel foundry

White Paper

Intel Foundry | Power Delivery

Driving Power Delivery Innovations for the AI Data Center Era

Intel Foundry offers a broad portfolio of technologies for silicon and system architects to optimize power delivery to specific design requirements.

Authors

Kaladhar Radhakrishnan
Fellow, Semiconductor Technology

Vishal Javvaji
Process Integration Manager

Nicolas Butzen
Principal Engineer

Contents

Executive Summary

Power delivery has become a significant limiter for performance scaling in modern data center chips. The rapid rise of AI and high-performance computing (HPC) workloads places far greater demands on the power delivery network (PDN) than traditional solutions were designed to handle. Incremental improvements alone are no longer sufficient as rising current levels compound resistive losses and voltage droop across the PDN. Stacked-die multi-chip architectures exacerbate the problem by increasing effective current density, routing congestion, and reliability challenges.

First, we examine power delivery challenges across the accelerator, rack, and data center levels, highlighting how scaling trends are reshaping power delivery requirements throughout the system hierarchy.

Next, we describe Intel Foundry's portfolio of power delivery technologies that address these challenges. These include Intel Foundry's proprietary backside power delivery solution introduced on the Intel 18A node, and the use of advanced silicon capacitor solutions with best-in-class capacitance density on the die as well as the package.

In the final part, we cover Intel's pioneering work in integrated voltage regulator (IVR) technology, enabling higher voltages to be delivered to the package to mitigate current-density scaling challenges. This approach relies on building blocks such as integrated magnetic inductors, a fully capacitive IVR chiplet, and tightly coupled package integration schemes.

This paper does not advocate for a single, optimized power delivery solution. Instead, it outlines a set of approaches that senior architects and technologists can apply selectively based on system requirements and architectural tradeoffs.

The Power Delivery Challenge

Design trends across chips, servers, and data centers are pushing power delivery to its physical limits. As compute density increases, maintaining robust power integrity (PI) requires optimization across the entire PDN, from the board and package to the silicon.

Accelerator Package Design Trends

Modern accelerator packages are integrating unprecedented levels of compute, memory, and interconnect within a single package to meet the performance and bandwidth demands of AI and HPC workloads.

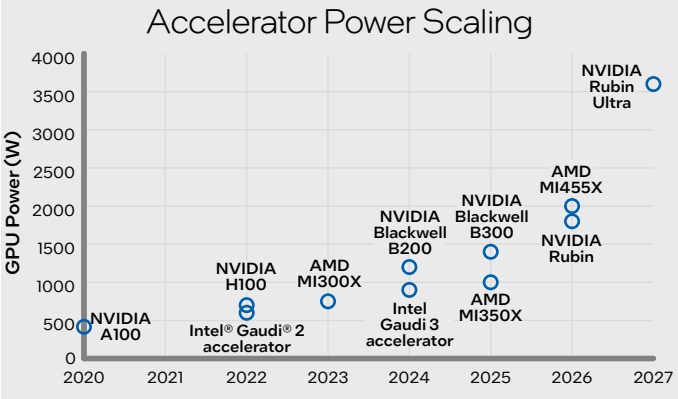

Figure 1. Data center accelerator power scaling over time, based on Intel Foundry analysis of publicly available information as of August, 2026.

AI accelerator devices typically operate at low supply voltages, requiring large currents to deliver high power (see Figure 1). As a result, modern integration levels now require package currents in the multi-kiloamp range. These trends create two fundamental power delivery challenges. First, the package must deliver an enormous amount of power within a limited footprint already occupied by the compute die, high bandwidth memory (HBM) stacks, high-speed interconnects, and cooling infrastructure. Second, it must do so efficiently while minimizing the static and dynamic losses in the PDN.

Delivering large currents efficiently is challenging, especially when voltage conversion remains physically distant from the load. Current must be transported across long resistive paths, increasing resistive losses and reducing overall power delivery efficiency. In addition, higher current and faster transients increase voltage droop, forcing designers to add voltage guard bands to maintain reliable operation. These guard bands increase power consumption and further reduce system efficiency.

Figure 2 illustrates how the efficiency of the PDN degrades as load current increases. While voltage conversion losses remain relatively constant, resistive losses and droop-related losses grow significantly at higher power levels. Consequently, traditional power delivery schemes require an increasingly large fraction of input power to the PDN rather than useful compute.

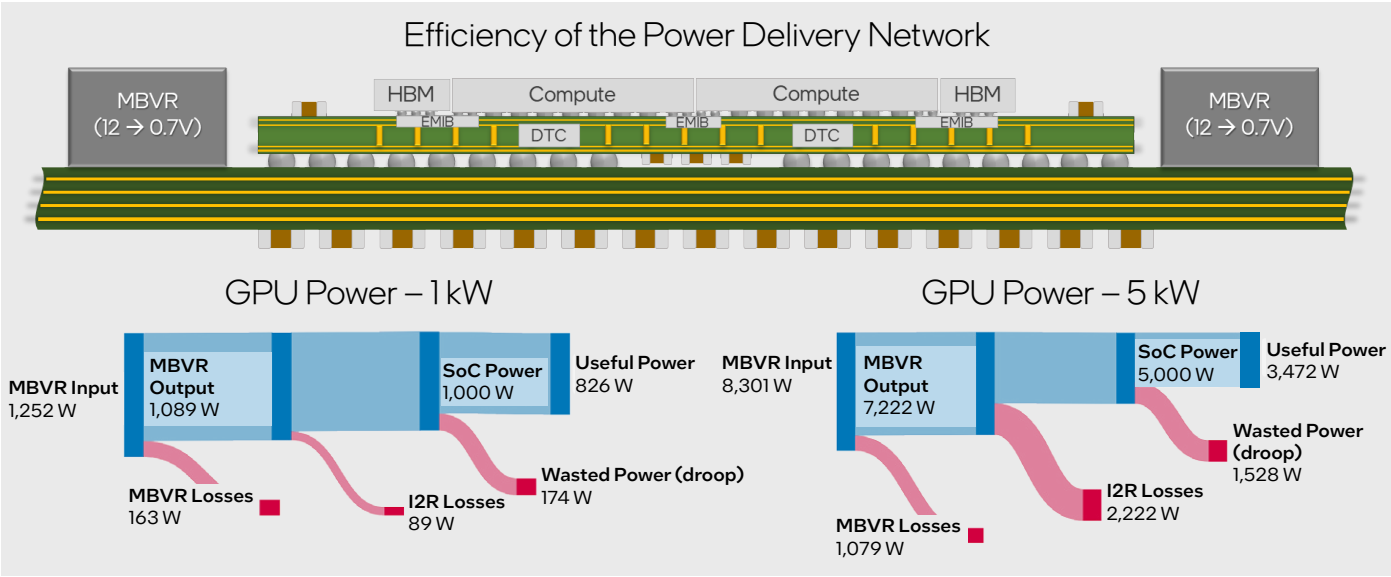

Figure 2. Degradation of power delivery network (PDN) efficiency at higher power levels.

Multi-Chip and Stacked Architectures

As traditional single-die scaling faces growing economic and technical challenges, the industry is turning to advanced packaging approaches, including chiplets, HBM integration, silicon bridges, and 3D stacking, to continue delivering system-level performance and bandwidth gains. While these multi-chip and stacked architectures enable higher levels of performance, bandwidth, and functional integration, they also introduce new challenges in power distribution, decoupling, and voltage regulation.

Unlike traditional monolithic designs, multi-chip architectures distribute compute, memory, and I/O functions across multiple silicon dies, each with different voltage, current, and transient-response requirements. Power must be delivered to each individual die through various package structures including interposers, silicon bridges, package substrates, microbumps, and through-silicon vias (TSVs). At the same time, HBM interfaces and die-to-die interconnects operate with increasingly tight voltage margins and are highly sensitive to power supply noise.

The growing complexity of heterogeneous accelerator packages creates an additional challenge within the silicon itself. The integration of additional compute tiles, memory interfaces, and high-speed die-to-die links puts growing pressure on signal-routing resources. In conventional frontside power delivery architectures, power rails and signal interconnects must share the same routing resources, creating a fundamental tradeoff between signal bandwidth and power delivery capability. To support growing signal density, advanced technologies continue to add routing layers and increasingly complex interconnect structures. However, these additional routing layers can also increase the effective resistance of the power delivery path by forcing current to traverse a larger number of vertical interconnect layers before reaching the active circuitry. At higher current levels, this additional resistance contributes to greater voltage drop (IR drop), reduced power delivery efficiency, and tighter voltage margins.

These trends are driving a fundamental architectural shift toward backside power delivery, where power distribution is moved to the backside of the wafer, freeing frontside routing resources for signals while reducing the resistance and congestion associated with conventional PDNs.

Data Center Power Trends

Power scaling at the individual rack level is reshaping power delivery requirements by imposing strict volumetric, thermal, and architectural constraints. Traditional data center racks are typically designed for approximately 10–20 kilowatts (kW) and rely almost exclusively on air cooling, with generous space allocated for power supplies, cabling, and airflow management. In contrast, modern AI racks already exceed 100 kW and have largely transitioned to liquid cooling to manage the resulting heat flux. Increasing rack power density reduces the available volume for power conversion, distribution, and cooling infrastructure while tightening efficiency requirements across the power chain. Looking ahead, industry roadmaps project rack power approaching 1 megawatt (MW) by the latter part of the decade, making volumetric efficiency, loss minimization, and close coupling between power delivery and cooling no longer optional, but essential design requirements.

Over the last decade, data center design assumptions have shifted dramatically. According to a June 2026 estimate from Lawrence Berkeley National Laboratory, U.S. data centers could account for 11.8% of total U.S. electricity consumption by 2030, with a projected range of 9.5% to 15.3% depending on AI deployment and infrastructure assumptions.[1] This represents an upward revision from the 2024 report estimate range of 6.7% to 12.0% of total U.S. electricity by 2028,[2] with growth driven primarily by increases in high-density compute in AI and accelerator-based devices.

> Driven by AI and accelerator-based computing, U.S. data centers could consume nearly 12% of the nation's electricity by 2030.[1]

At the same time, the increasing deployment of high-power accelerators is driving greater demands on power delivery and cooling infrastructure, with device roadmaps continuing to push package power levels upward. GPUs are one of the primary engines behind growing AI workloads, driving increases in both computational performance and power consumption. Higher power levels create disproportionately larger efficiency losses and thermal-management challenges. These effects increase operating costs, cooling requirements, and infrastructure complexity, making improvements in end-to-end power delivery efficiency increasingly important for future scaling.

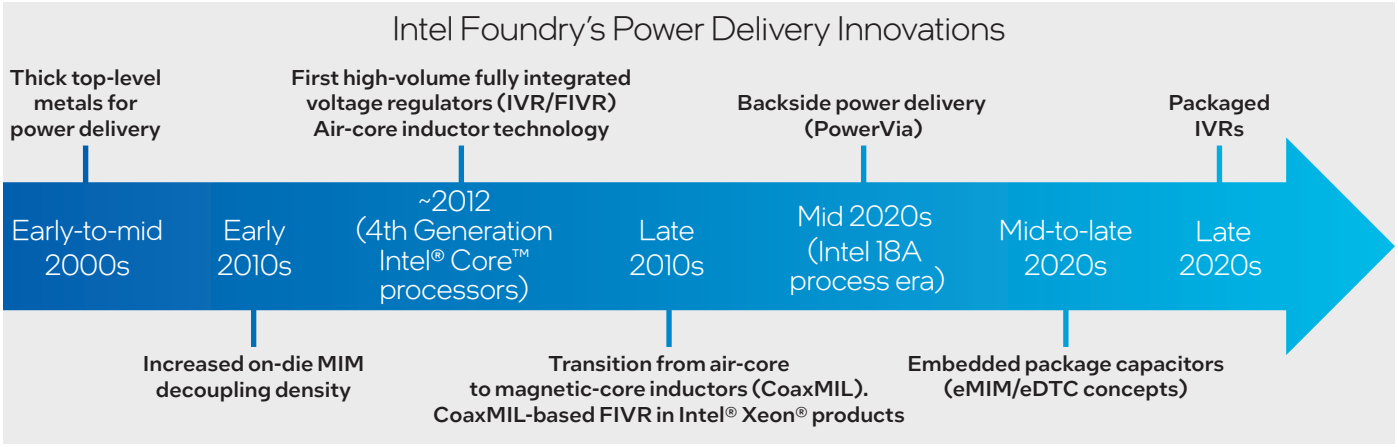

Figure 3. Intel Foundry's power delivery innovations over time.

Power Delivery Early Innovations to Today

Many of the power delivery challenges confronting today's AI accelerators first emerged in Intel's high-performance CPUs, where increasing current demand pushed conventional power delivery architecture to its limit. Figure 3 illustrates how Intel responded by pioneering a series of power delivery innovations that later became widely adopted across the industry. These include thick metal layers for improved power distribution, introduction of on-die metal-insulator-metal (MIM) capacitors, and early forms of integrated voltage regulators (IVRs).

The power requirements of modern AI accelerators are now driving a new generation of power delivery innovations. To address these challenges, Intel Foundry is advancing technologies such as backside power delivery, advanced package-level decoupling solutions, embedded bridge technologies with TSV, as well as next-generation IVR architectures that move power conversion closer to the load. These technologies are discussed in the following sections.

Backside Power Delivery: Separating Power from Signals

With conventional frontside power delivery architectures, power rails and signal interconnects compete for the same routing resources, creating a fundamental tradeoff between signal bandwidth and power delivery capability. Growing metal-stack complexity forces power to traverse additional routing layers before reaching the transistors, resulting in more IR drop. PowerVia reroutes primary power to the backside of the die, physically separating power routing from frontside signal interconnects (see Figure 4). This enables independent optimization of both networks by shortening resistive current paths while alleviating frontside routing congestion.

PowerVia enables a significant reduction in the on-die IR drop, more efficient use of frontside routing resources for signals, and improved standard cell utilization. For high-current logic, backside power delivery directly addresses many of the fundamental scaling limits of traditional frontside PDNs. Realizing these benefits requires close co-optimization across process integration, thermal management, and architecture.

Introduced on Intel 18A, PowerVia also provides the foundation for Power Boost on Intel 18A-P, a new dual-contact architecture that delivers over 10% higher frequency at matched capacitance.[3] PowerDirect, the second generation of Intel Foundry's backside power delivery technology planned for Intel 14A, will build upon the benefits of PowerVia by introducing additional area efficiencies and a more optimized power delivery architecture, further improving overall power, performance, and area (PPA).

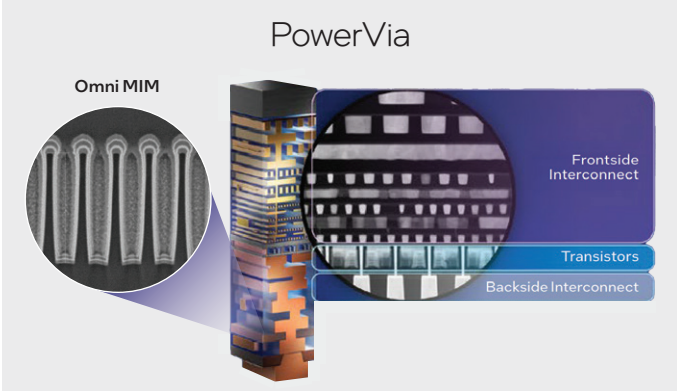

Figure 4. Backside power delivery using PowerVia (Omni MIM spotlight on left).

Omni MIM: On-Die Decoupling Capacitor

While backside power delivery minimizes on-die IR drop, achieving robust power integrity also requires advances in on-die decoupling to support the large transient currents generated by AI workloads. Intel was the first to introduce MIM capacitors into high-volume logic technologies, providing on-die capacitance to meet high-frequency decoupling needs.

The left spotlight in Figure 4 highlights Omni MIM, Intel Foundry's latest generation of on-die MIM capacitor technology. Early generations of MIM capacitors relied on planar capacitor structures, with successive generations increasing capacitance density by adding additional electrode layers. While effective, this approach provides only incremental improvements as capacitance scales

roughly with the number of plates. Omni MIM takes a different approach by extending the capacitor structure into the third dimension. By incorporating vertical trenches within the inter-layer dielectric, Omni MIM enables a substantial increase in capacitance density while maintaining close proximity to the active circuitry.

eMIM-T and eDTC: Advanced Package Decoupling Solutions

On-die MIM capacitors act as the first line of defense against fast current transients. However, their stored charge is finite and must be replenished by the next level of decoupling capacitors. Traditionally, this role has been served by discrete ceramic capacitors mounted on the land and die sides of the package. In high-density accelerator packages, nearly all of the top-side area is fully occupied by compute and memory dies. At the same time, the landside is densely populated with ball grid array (BGA) connections to support both high-current power delivery and off-package I/O. This leaves limited space for traditional surface-mounted decoupling capacitors, increasing the importance of embedded capacitor technologies.

Intel Foundry is developing two complementary embedded capacitor technologies to address these requirements:

- Intel Foundry's proprietary eMIM-T capacitor extends the Omni MIM technology into the package by integrating multiple layers of high-density MIM capacitors inside the package build-up layers (see Figure 5). By positioning capacitance closer to the load, eMIM-T improves high-frequency decoupling while maintaining DC delivery through TSV connections. In designs where on-die capacitance is limited by area or process constraints, the eMIM-T structures can help complement the on-die charge storage without consuming valuable silicon area.
- Integrated inside the core of the package, embedded deep trench capacitors (eDTC) provide yet another stage of decoupling capacitors to maximize the total capacitance available within the package, helping to address low- to mid-frequency decoupling needs.

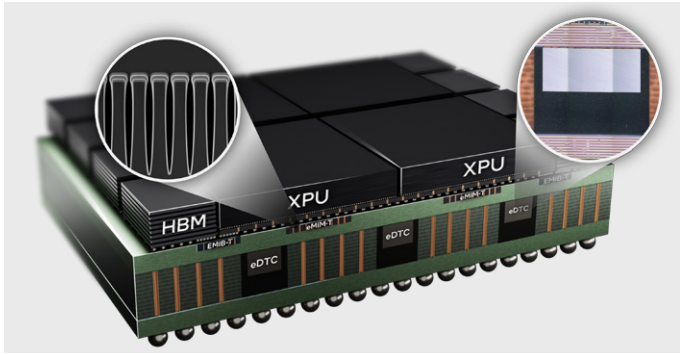

Figure 5. Package integrated decoupling for improved power integrity.

EMIB-T: Enabling HBM4 and High-Speed Die-to-Die Interconnects

Embedded Multi-die Interconnect Bridge (EMIB), shown on the left side of Figure 6, is one of Intel Foundry's advanced packaging technologies for enabling multi-chip architectures without the cost and complexity of a full silicon interposer. Since its introduction in 2017, EMIB has played a key role in scaling bandwidth and die-to-die connectivity in advanced multi-chip packages. As AI and HPC platforms adopt bandwidth-intensive interfaces such as HBM4 and next-generation Universal Chiplet Interconnect Express (UCIe) links, there is a corresponding increase in current demand. With traditional EMIB implementation, power is routed around the bridge structure, which increases the path resistance. As current density increases, IR drop can also increase and create current crowding at edge bumps. This introduces both efficiency and reliability concerns.

The newer EMIB-T architecture, shown on the right side of Figure 6, addresses these limitations by incorporating TSVs directly inside the bridge. These TSVs provide a more direct vertical current path to the load, allowing current to pass through the bridge rather than around it. This reduces resistive losses and distributes current more evenly across the bump array, helping mitigate edge bump reliability risks. EMIB-T also integrates MIM capacitors within the bridge structure to provide localized suppression of AC noise on high-speed I/O power rails. Together, these capabilities enable more robust power delivery for next-generation high-bandwidth chiplet architectures.

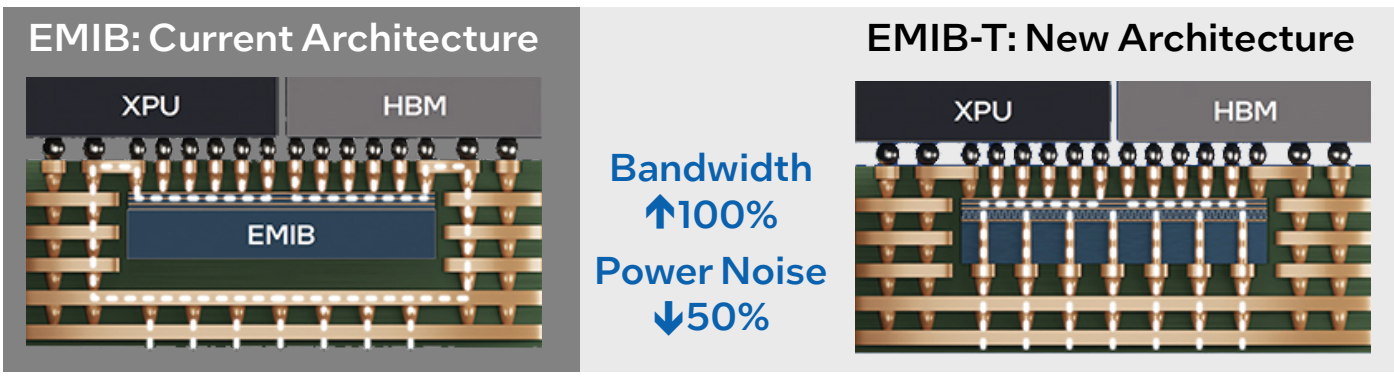

Figure 6. Embedded Multi-die Interconnect Bridge (EMIB) versus EMIB-T.[4]

Intregrated Voltage Regulators

As accelerator current requirements scale into the multi-kiloamp range, maintaining efficient power delivery and tight voltage regulation becomes increasingly challenging. In addition, regulator efficiency tends to degrade at higher load currents, requiring even more input power to sustain operation. The effects compound at scale, amplifying thermal and power delivery challenges.

IVR-based architectures help address these challenges by moving voltage conversion closer to the load. Shortening the distance between the regulator and the active circuitry reduces resistive losses and improves transient response. Higher local conversion efficiency at lower output voltages, combined with reduced distribution loss, allows more effective delivery of power where it is needed. Furthermore, minimizing the electrical path between the regulator and the load helps mitigate voltage droop during the rapid current transients that are common in modern accelerators. The following three sections describe several IVR-related technologies Intel has developed to support accelerator systems that have higher current and density.

FIVR with CoaxMIL

Intel was the first to deploy IVRs in high volume manufacturing over a decade ago. The initial fully integrated voltage regulator (FIVR) architecture integrated the power switches, controller, and output MIM capacitors directly on the processor die, while using air core inductors (ACI) in the package for energy storage. The approach was effective for early product generations, but higher current densities require smaller ACI footprints, which can degrade inductor quality and lead to a corresponding drop in conversion efficiency.

To address this limitation, Intel introduced CoaxMIL, a magnetic inductor solution first deployed on 4th Gen Intel® Xeon® Scalable processors (see Figure 8 on the following page). By significantly reducing inductor footprint, CoaxMIL enables higher current densities while delivering improved power conversion efficiency. Newer generations of CoaxMIL have been deployed across subsequent Intel Xeon processor families, providing continued improvements in conversion efficiency as well as introducing coupled-inductor configurations that enhance transient response while further reducing footprint.

Spotlight: Evolution of Voltage Regulation

Higher Bandwidth, Higher Current Density

Traditional motherboard voltage regulators (MBVRs) convert distribution voltages, typically around 12 V, to the lower voltages required by processors and accelerators (see the bottom-left image in Figure 7). To operate efficiently at these voltages, they use power devices that switch relatively slowly. Lower switching frequencies require larger inductors and capacitors to store and filter energy, which limits both current density and regulator bandwidth. As accelerators for AI workloads demand faster and larger current transients with tighter voltage margins, these limitations become more challenging.

Integrated voltage regulators (IVRs) use faster switching transistors that operate at much higher frequencies. Higher switching frequency enables significantly smaller inductors and capacitors, allowing more compact implementations with higher current density and faster response to load transients. The result is a regulator that can deliver large currents from a much smaller footprint while responding more quickly to changes in load demand.

Early IVR implementations used inductors integrated on the chiplet or package (see the middle image in Figure 7). More recent designs leverage advances in silicon capacitor technology to enable fully capacitive IVR chiplets (see the bottom-right image in Figure 7) that can be manufactured in extremely thin form factors and integrated close to the load.

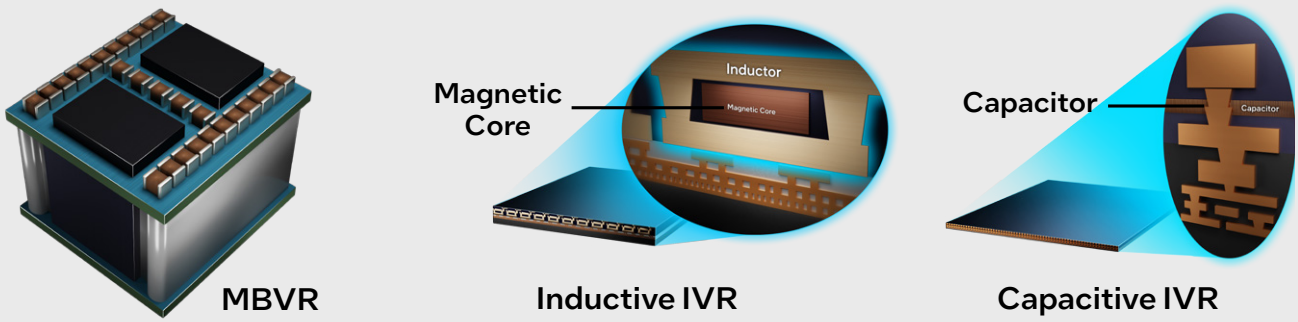

Figure 7. The evolution of voltage regulation.

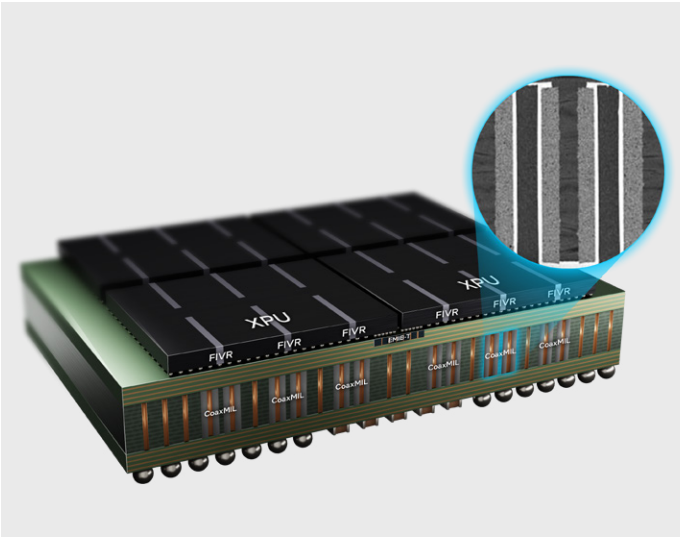

Figure 8. CoaxMIL magnetic inductor technology.

C2VR

The continuous capacitive voltage regulator (C2VR) is a VR chiplet that uses high-density MIM capacitors instead of inductors to convert an up to 2.4 V input to typical core domain voltages (see Figure 9). Because the MIM capacitors are integrated on the chiplet itself, the C2VR offers a comprehensive VR solution that can provide more than 90% efficiency and 10 A/mm^2 on a single piece of silicon.[5] Relying only on capacitors, the C2VR can increase output current almost instantaneously, exhibiting minimal voltage droop during high di/dt events transients.

The C2VR is scalable to multi-kiloamp current levels and supports placement on either the die side or landside of the package. Depending on the implementation, the C2VR can be embedded within package build-up layers or incorporated directly into the base die of stacked-die designs.

IVR Integration Architectures

A practical first step in moving voltage conversion closer to the load is to integrate the regulator on the landside of the package (shown in Figure 9). Compared to traditional MBVRs, a landside IVR significantly shortens the power delivery path and reduces distribution losses while avoiding many of the process-integration challenges associated with on-die voltage regulation. Despite their benefits, landside IVRs face significant electrical and thermo-mechanical challenges at higher current densities. The presence of a landside IVR, along with its cooling requirements, reduces the BGA and printed circuit board resources available for power delivery. This can increase routing losses on the input path, partially offsetting the benefit of near-load integration.

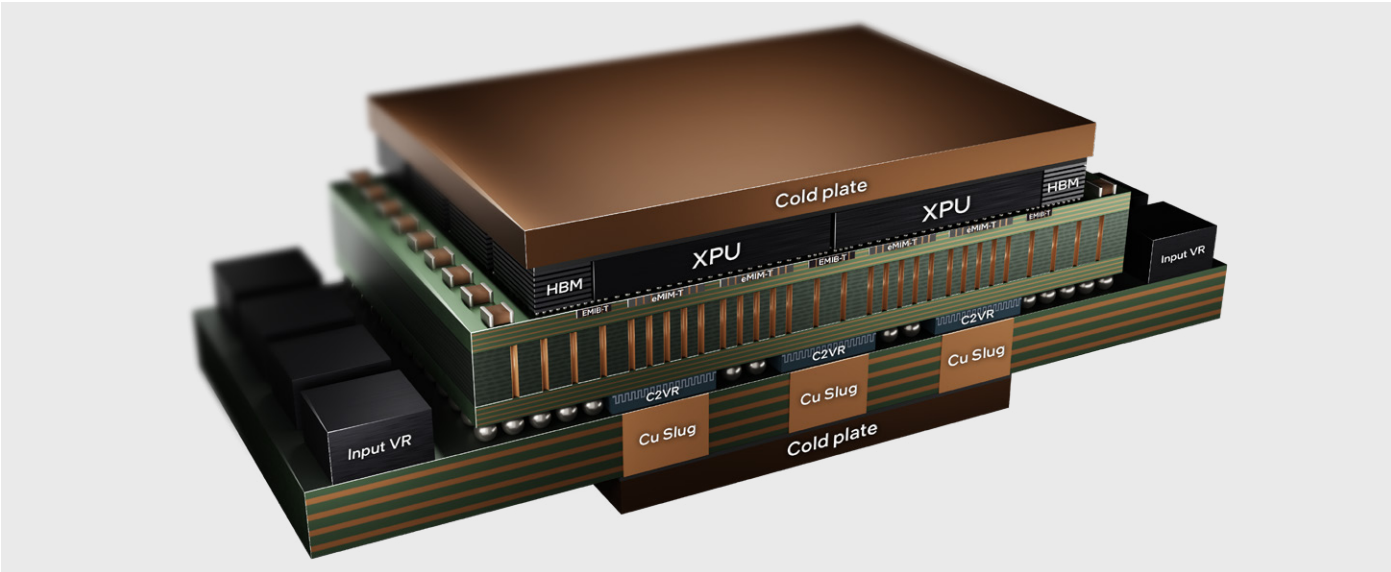

Figure 9. Landside implementation of IVR.

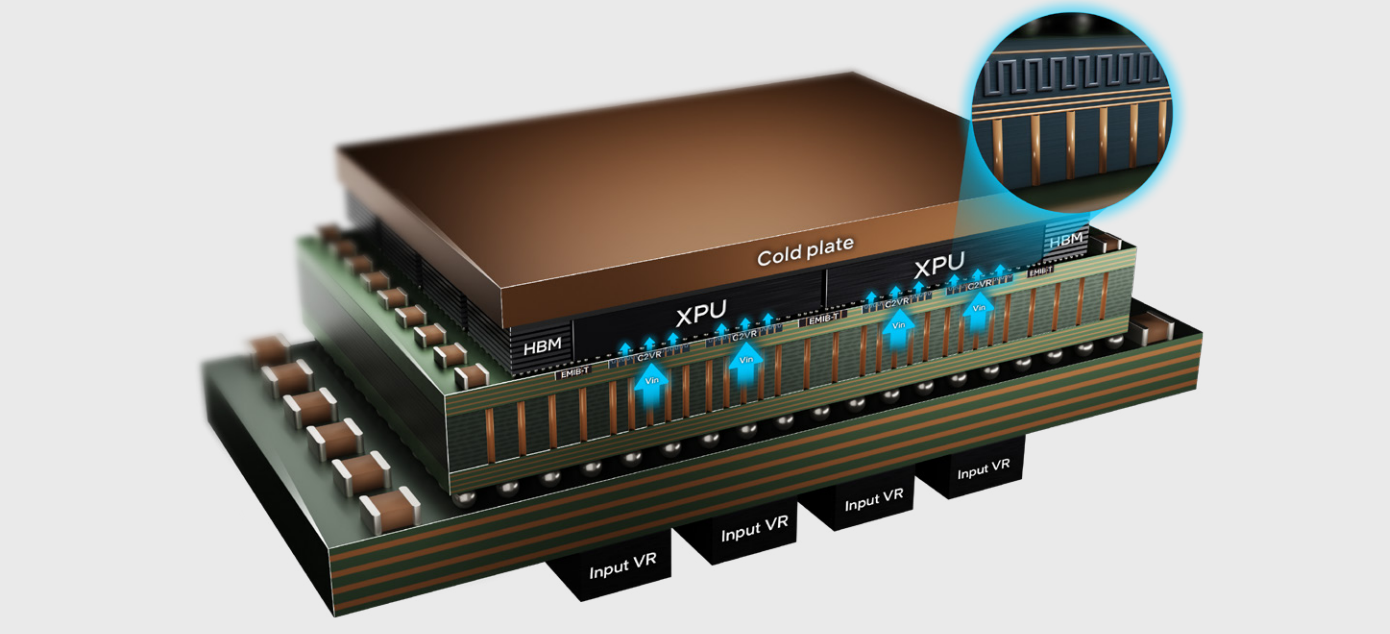

Figure 10. Embedded C2VR under the load die.

One way to overcome the limitation of landside implementation is by integrating the IVR directly beneath the compute die to enable vertical power delivery. However, this requires an IVR architecture with an extremely thin profile and the ability to route input power from the backside using TSVs. The C2VR is uniquely suited for this type of integration scheme because its fully capacitive construction can be thinned to less than 100 µm while maintaining functionality, enabling input power delivery through TSVs from the backside of the structure. This flexibility allows the C2VR to be embedded within package build-up layers (see Figure 10) or integrated directly into the base die in stacked-die architectures.

Delivering a Portfolio Approach to Power Delivery

As AI accelerators push toward multi-kilowatt power levels, power delivery has emerged as a key limiter to performance scaling. Every stage of the PDN—from the voltage regulator, to the different stages of decoupling capacitors, to the current distribution path—plays a direct role in how much of the available compute capability can be effectively utilized.

Intel's experience in developing HPC platforms with aggressive power and current density requirements has led to a broad portfolio of power delivery technologies designed to address these challenges at multiple levels of the system stack. The diversity of AI and HPC architectures demonstrates that there is no one-size-fits-all approach to power delivery. Different applications, package architectures, power levels, and integration requirements demand different combinations of technologies and optimization strategies.

Rather than pursuing a single architectural approach, Intel Foundry provides a toolkit of complementary power delivery technologies that allows architects to optimize power delivery based on the specific requirements of their design. Together, these innovations provide a scalable path for delivering the increasing current, efficiency, and performance demanded by future AI and HPC platforms.

To learn more about Intel Foundry's portfolio of power delivery technologies, visit intel.com/foundry or reach out to us at foundry.contact@intel.com.

intel foundry

[1] Lawrence Berkeley National Laboratory, "United States Data Center Energy Usage Report: 2025 Update," June 2026.

[2] Lawrence Berkeley National Laboratory, "United States Data Center Energy Usage Report: 2024 Update," December 2024; Pew Research Center, "Energy Use at U.S. Data Centers Amid the AI Boom," October 2025.

[3] A. Bowonder et al., "Intel 18A-P CMOS Technology Enhancement Featuring Advanced RibbonFET (GAA) Transistors and PowerVia for High-Performance Computing," IEEE/JSAP Symposium on VLSI Technology and Circuits, June 2026.

[4] Y. Mekonnen et al., "Enabling 12+ Gb/s HBM4e with EMIB-T Advanced Packaging Technology," IEEE Electronic Components and Technology Conference (ECTC), May 2026.

[5] N. Butzen et al., "A Monolithic 12.7 W/mm², 92% Peak-Efficiency Switched-Capacitor DC-DC Converter Using CSCR-First Topology," IEEE Journal of Solid-State Circuits, December 2024.

Performance varies by use, configuration, and other factors. Results may vary. Learn more at intel.com/performanceindex.

All product and service plans, roadmaps, and performance estimates are subject to change without notice. Projections about future node performance and other metrics are inherently uncertain. Performance results are based on testing as of dates shown. Your costs and results may vary.

This document contains forward-looking statements about Intel's future plans or expectations, including its process technology roadmaps. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

Intel® technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. No product or component can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

All information provided here is subject to change without notice. © Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others. 0826/DBAD/KC/PDF