OPEN ACCESS

EDITORIAL

Cooling next-generation electronics: From emerging semiconductors to data centers

Zhihu Wu[1] and Zuankai Wang[1,*]

[1]Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong SAR, China

*Correspondence: Email: zk.wang@polyu.edu.hk

Received: September 20, 2025; Accepted: October 10, 2025; Published Online: October 10, 2025; https://doi.org/10.59717/j.xinn-energy.2025.100125

© 2025 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Wu Z. and Wang Z. (2025). Cooling next-generation electronics: From emerging semiconductors to data centers. *The Innovation Energy* **2**:100125.

With the significant progresses in lithography, advanced packaging technologies, and the integration of emerging semiconductor materials, modern electronics have revolutionized numerous fields, including data center, electric vehicles, and wireless communication. However, the accompanying rise in thermal design power and heat flux, which are over kilowatts and kilowatts per square centimeter respectively, poses a significant challenge for thermal management. This cooling crisis is further exacerbated by the exponential growth of data centers driven by artificial intelligence (AI) workloads, which now consume over 2% of global electricity, with cooling alone accounting for nearly 40% of this energy. Without transformative thermal management strategies, the next generation of electronics risks being limited by its own heat. This editorial traces the thermal challenge across scales—from chip-level, enabled by emerging semiconductor materials, to rack-level infrastructure. We begin by dissecting the thermal bottlenecks of wide-bandgap and ultra-wide-bandgap semiconductors, where soaring heat fluxes clash with poor thermal conduction. Next, we explore the heat transfer complexities of advanced packaging technology, where heterogeneous integration amplifies thermal crosstalk. Finally, we scale up to data centers, where liquid cooling and waste heat recovery must improve to support the global decarbonization efforts. Through this lens, we emphasize for a thermal-centric redesign of electronics, making cooling innovation a driver of sustainable progress.

COOLING WIDE-BANDGAP SEMICONDUCTORS

Wide-bandgap (WBG) semiconductors, notably gallium nitride (GaN) and silicon carbide (SiC), offer electrical performance far beyond that of conventional silicon (Si). For example, their bandgaps—3.4 eV for GaN and 3.3 eV for SiC—are roughly three times that of Si (1.1 eV), enabling much higher critical breakdown fields (3.3 MV/cm for GaN vs. 0.3 MV/cm for Si). These intrinsic advantages translate directly into devices with higher power density, faster switching speeds, and lower conduction losses, making WBG transistors the front-runners for applications ranging from electric-vehicle inverters and 5G base stations to high-efficiency renewable - energy converters. However, the corresponding ultra-high heat fluxes of over kW/cm^2 creates severe thermal challenges. without effective cooling, the unacceptable junction temperatures will degrade device performance, reliability, and lifespan. Traditional air cooling, limited by the low heat-transfer coefficients, cannot evacuate heat fluxes above roughly 100 W/cm^2. Liquid cold plates improve convective heat transfer but still face limitations due to high interfacial thermal resistance and significant conduction resistance within the chip package. In fact, under conventional packaging, the resistance of the heat conducting through solids and across interfaces including device layers, substrates, packaging material and heat sink, typically accounting for over 50 percent of the total thermal resistance from an on-chip hotspot to the ambient.

Many studies have addressed to reduce the conduction resistance, including the development of high-thermal-conductivity interface materials and the replacement of conventional substrates like silicon, which has a thermal conductivity of about 140 W/m·K, with materials such as diamond (2000 W/m·K) via epitaxial-transfer techniques. A more radical and promising strategy involves embedding microfluidic structures directly into the chip substrate (Figure 1A). This method has demonstrated excellent cooling performance by bringing the coolant to within approximately 100 μm of the active electrical layer, thereby significantly lowering conduction resistance. Beginning in 2012, DARPA's Intra/Interchip Enhanced Cooling (ICECool) Program[1] aims to develop advanced embedding cooling to mitigate thermal limitations on the operation of military electronic systems, while significantly reducing size, weight, and pumping power consumption. With the integration of high-performance microfluidic cooling structures, such as microjets and manifold microchannels, with diamond substrates, the program successfully demonstrated heat dissipation of up to 1000 W/cm^2 for GaN chips. Although, thermal management solutions were typically implemented only after chip design and fabrication, leaving considerable room for further improvements in cooling performance. In 2020, a research group[2] ingeniously co-designed the electronic architecture and embedded microfluidic cooling structures, effectively reducing the conduction thermal resistance to nearly zero. This breakthrough enabled a GaN device to dissipate a heat flux exceeding 1700 W/cm^2. Furthermore, combined with excellent microfluidic structure design,[3] the heat dissipation of embedded cooling can reach 3000 W/cm^2. Indeed, the growing thermal challenges demand a "thermal-first" mindset, where thermal management considerations influence device design and process flow from the outset. This principle was further emphasized in DARPA's Technologies for Heat Removal in Electronics at the Device Scale (THREADS) program started in 2023. With an expected investment of $60 million, the program aims to develop technologies that overcome transistor thermal limitations and dramatically increase the power output of GaN-based radio frequency devices from below 10 W/mm to 81 W/mm. To achieve this ambitious goal, thermal issue must be carefully considered during the transistor epilayer design, gate layouts, and multi-finger transistor topologies.

COOLING ULTRA-WIDE-BANDGAP SEMICONDUCTORS

While WBG semiconductors have transformed power electronics, attention is increasingly turning to ultra-wide-bandgap (UWBG) materials such as gallium oxide (Ga_2O_3) for their potential to push performance even further. With a bandgap of approximately 4.8 eV and a theoretical critical electric field exceeding 8 MV/cm, Ga_2O_3 offers exceptional potential for high-voltage, high-power-density devices with significantly smaller footprints than WBG counterparts. Strikingly, Ga_2O_3 can be grown from the melt using techniques such as edge-defined film-fed growth (EFG), making it more cost-effective than compound semiconductors that rely on expensive vapor-phase epitaxy.

However, realizing these benefits requires overcoming severe thermal management challenges. One of the most limiting factors is the intrinsically low and anisotropic thermal conductivity of β- Ga_2O_3, which is approximately 11–27 W/m·K—an order of magnitude lower than GaN and SiC. This poor heat conduction significantly hampers the extraction of heat from hot spots. Moreover, the thermal anisotropy further complicates heat spreading, making conventional thermal design less effective. Nevertheless, recent studies have shown that Ga_2O_3 devices can maintain stable electrical performance even at elevated operating temperatures of 300 °C, owing to their robust thermal stability and wide bandgap. This high-temperature tolerance may relax some of the strict cooling requirements.

Recent thermal management progress in Ga_2O_3 chips has primarily focused on improving heat spreading and localized heat extraction at the device level. Researchers have explored the integration of high-thermal-conductivity substrates such as diamond or SiC to improve vertical thermal transport, as well as thinning Ga_2O_3 epitaxial layers to reduce internal thermal resistance. Flip-chip and double-side packaging are also benefit to improve the heat conduction path. In addition, embedded microfluidic cooling is being explored to further enhance overall thermal performance and energy efficiency. However, its effectiveness is only realized after the Ga_2O_3 substrate is sufficiently thinned to reduce internal thermal resistance. Compared to WBG devices, the thermal challenges associated with UWBG semiconductors are even more critical to unlocking their full performance potential.

THERMAL CHALLENGES IN ADVANCED PACKAGING

As WBG and UWBG semiconductors continue to push the boundaries of electronic performance, packaging technologies are evolving rapidly, particularly as Moore's Law approaches its physical and economic limits. Modern

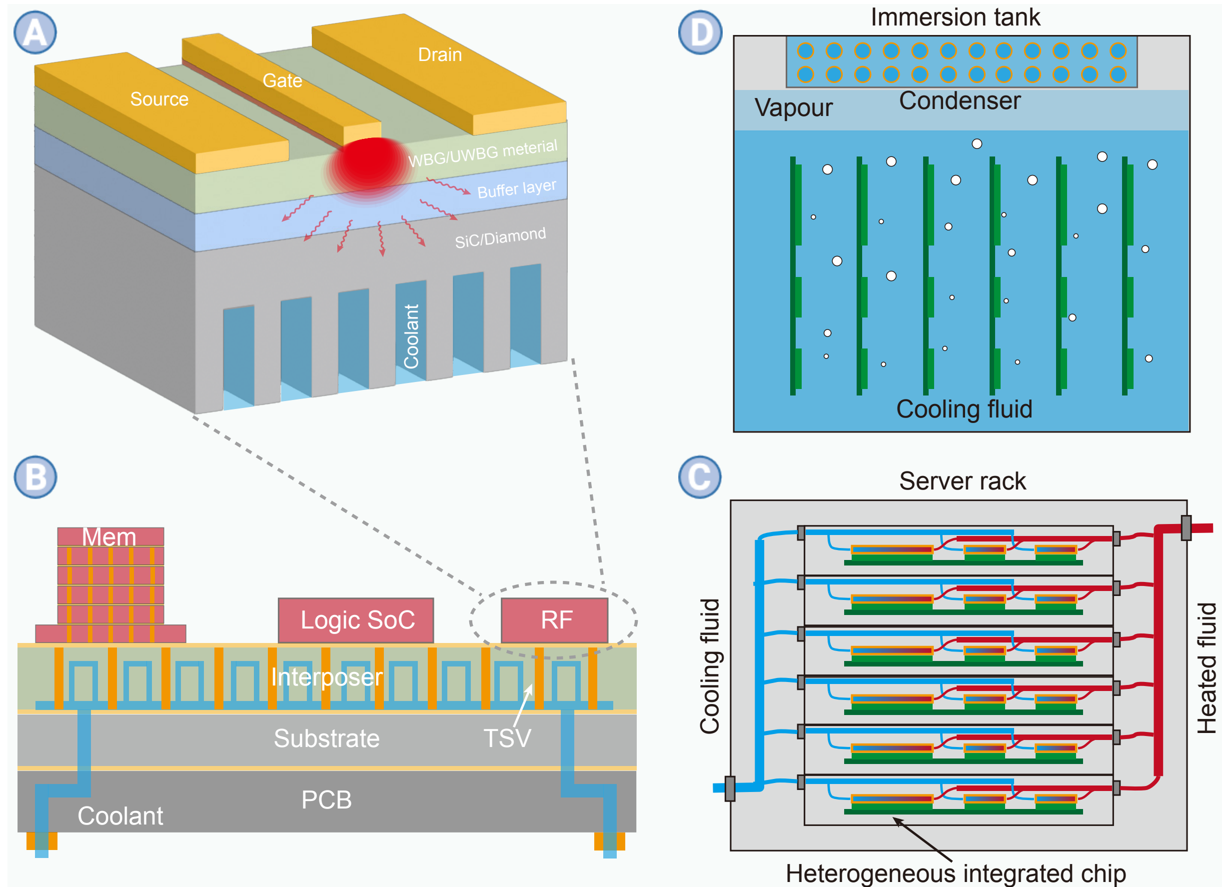

Figure 1. Thermal management across from chip to rack-level infrastructure (A) Embedded microfluidic cooling for electronics based on wide-bandgap or ultra-wide-bandgap semiconductors. (B) Co-design of electronic and microfluidic structure for the heterogeneous integration packaging. (C) Liquid cold plate cooling for server racks in data centers. (D) Two-phase immersion cooling for high-power data center infrastructure.

chip packaging has advanced far beyond traditional single-die configurations. Today, techniques such as 2.5D and 3D integration enable multiple chiplets—with distinct functionalities and fabrication processes—to be assembled into a single package. This heterogeneous integration allows logic units, memory blocks, and other functional blocks (including radio frequency devices) to be densely co-located, promises substantial improvements in capabilities of future microsystems. Representative examples include NVIDIA's B300 chips and AMD's MI300A/X chips, both of which are built on TSMC's Chip-on-Wafer-on-Substrate (CoWoS) platform. CoWoS employs high-density silicon interposers to support wide bandwidth and low-latency communication among chiplets, while facilitating greater integration scale and performance per unit volume. However, the resulting increases in power density and interconnect complexity creating significant thermal management challenges. These include efficient in-tier and cross-tier heat removal, effective thermal isolation between adjacent functional blocks, and mechanical reliability issues arising from thermal expansion mismatch across complex multilayer structures. Recognizing the urgency of these issues, DARPA launched the Minitherms3D program in 2023, committing $69 million to develop miniature thermal management systems for 3D heterogeneous integration (3DHI).

To manage these thermal challenges, a coupled thermal and electrical co-design strategy must be employed throughout the entire chip design and fabrication process. Firstly, the chiplet placement must be optimized to prevent thermal hotspots and mitigate heat accumulation. This is particularly important for in-tier heat spreading, as many interior chip tiers lack direct access to external heat sinks. In addition, the arrangement of through-silicon vias (TSVs) and redistribution layers (RDLs) plays a pivotal role in thermal optimization. Leveraging the high thermal conductivity of copper, these structures can be strategically distributed to form effective heat conduction pathways across both vertical and lateral dimensions. Embedded microfluidic cooling is also emerging as a critical solution for next-generation heterogeneous systems. Researchers have demonstrated the integration of microfluidic structure within silicon interposers and even inside 3D chip stacks, enabling localized and high-efficiency heat removal near the source (Figure 1B). However, compared to conventional single-layer chips, the microfluidic design with 3D architectures is far more complex due to the presence of intricate device layers, diverse materials, and dense interconnects. For instance, taller TSVs can enhance the pin-fin effect for improved microfluidic cooling, but simultaneously increase electrical signal delay, creating design trade-offs that must be carefully balanced.

DATA CENTER COOLING: FROM CHIPS TO RACKS

The rapid advancement of process node technologies, semiconductor materials, and heterogeneous integration packaging is also driving transformative changes in data centers. Modern data centers are evolving beyond traditional roles of storage and general-purpose computing toward AI-driven workloads, high-performance computing (HPC), and real-time data analytics. This shift demands unprecedented levels of computational power, the majority of which is ultimately dissipated as heat—posing formidable thermal management challenges. For instance, the thermal design power (TDP) of NVIDIA's latest GB300 GPU already exceeds 1000 W, expects to push total rack-level power densities beyond 100 kW. Traditional air-cooling systems, constrained by insufficient heat transfer capacity, high energy consumption, and increasing operational noise, are inadequate once rack power exceeds approximately 30 kW. In recent years, data center cooling has been undergoing a paradigm shift from traditional air-based methods to liquid-based solutions.[4]

Compared with air cooling, liquid cooling technologies such as cold plates and immersion cooling (Figure 1C & D) offer significantly higher heat dissipa-

tion capabilities, raising the cooling heat flux limit from several tens of watts per square centimeter to several hundred. Moreover, liquid cooling can reduce the power usage effectiveness (PUE) from values above 1.5 to below 1.1, marking a substantial gain in energy efficiency. Beyond thermal performance and efficiency, liquid cooling solution also eliminates the need for bulky fans, freeing up rack space and enabling greater system integration density. Despite current dominance of air cooling due to its lower cost and maturity, the rapid adoption of advanced semiconductor materials and heterogeneous integration in modern chips will inevitably make liquid cooling the mainstream solution in the near future.

Looking ahead, several key challenges must be addressed to ensure liquid cooling systems are powerful, reliable, efficient, and scalable for widespread deployment in data centers. First, cooling modules for high-power components such as CPUs and GPUs must be redesigned to enhance thermal performance and efficiency. Advanced microfluidic structures offer promising solutions: microjets can target localized hotspots with high convective heat transfer capability, while manifolds can reduce the flow path and lower pumping power requirements. Modularity and adaptability are also essential, particularly for heterogeneous integration platforms where multiple dies with distinct power profiles are co-packaged. Phase-change cooling technologies[5] that leverage the latent heat of fluids, combined with electrically insulating coolants, promise to significantly enhance heat transfer while maintaining system safety. Second, rapid and intelligent thermal control becomes increasingly vital in the liquid-cooled era. Unlike air cooling, liquid systems offer less margin for error in the event of coolant leakage, especially when using electrically conductive fluids. Additionally, precise thermal monitoring and dynamic coolant flow control can prevent local overheating of racks while avoiding unnecessary overcooling, thus enhancing energy efficiency. Third, waste heat recovery presents a major opportunity, particularly in the pursuit of carbon-neutral data centers. Instead of simply rejecting heat to the ambient, exhaust heat can be repurposed for district heating, boiler feedwater preheating, or low-grade thermoelectric power generation. Finally, decoupled system-level design considerations such as standardized interfaces, modular piping networks, and robust leak containment architectures are essential to accelerate adoption. A decoupled, plug-and-play cooling ecosystem will be critical to supporting the diverse and fast-paced development of future data centers.

THERMAL-CENTRIC DESIGN: THE BLUEPRINT FOR NEXT-GENERATION ELECTRONICS

From wide- and ultra-wide-bandgap semiconductors to advanced packaging and data-center-scale infrastructure, the thermal challenges in modern electronics span vastly different length scales—but converge on a single imperative: heat must be treated as a main design constraint of modern electronics, not an afterthought. As power densities rise and system integration deepens, traditional cooling approaches are being pushed to their limits. Meeting the demands of next-generation electronics requires a fundamental shift toward effective, scalable, and energy-efficient thermal management strategies. Promising directions include embedded microfluidic cooling that largely shortens thermal pathways, advanced microfluidic structures that enhance convective heat transfer and reduce hydraulic resistance, and phase-change methods that exploit latent heat for high flux removal.[5] But even the most advanced cooling technologies cannot succeed in isolation. A co-design paradigm, in which electrical, packaging, and thermal architectures are developed in tandem, is essential. This “co-design” mindset must extend from transistor layout and interposer design to system integration and facility-level planning. With the continuous advancement in electronic complexity and performance, adopting a thermal-centric design approach has become imperative—it serves as the blueprint for unlocking the full potential of next-generation systems.

REFERENCES

1. Bar-Cohen A., Maurer J. J. and Altman D. H. (2019). Embedded cooling for wide bandgap power amplifiers: A review. *J. Electron. Packag.* **141**:040803. DOI:10.1115/1.4043404
2. van Erp R., Soleimanzadeh R., Nela L., et al. (2020) Co-designing electronics with microfluidics for more sustainable cooling. *Nature* **585**:211–216. DOI:10.1038/s41586-020-2666-1
3. Wu Z.H., Xiao W., He H.Y., et al. (2025). Jet-enhanced manifold microchannels for cooling electronics up to a heat flux of 3,000 W cm^{-2}. *Nat. Electron.* DOI:10.1038/s41928-025-01449-4
4. Alissa H., Nick T., Raniwala A., et al. (2025). Using life cycle assessment to drive innovation for sustainable cool clouds. *Nature* **641**:331–338. DOI:10.1038/s41586-025-08832-3
5. Jiang M.N., Wang Y., Liu F. Y., et al. (2022). Inhibiting the Leidenfrost effect above 1,000 °C for sustained thermal cooling. *Nature* **601**:568–572. DOI:10.1038/s41586-021-04307-3

FUNDING AND ACKNOWLEDGMENTS

The authors acknowledge the financial support from the National Natural Science Foundation of China (Nos. T2293694, 52333015, 52206100), National Key Research and Development Program of China (No. 2023YFE0209900), Research Grants Council of Hong Kong (Nos.15237824, SRFS2223-1S01, 11215523, N_PolyU5172/24), the Innovation and Technology Commission of Hong Kong (No. MHP/025/23), Meituan Foundation through the Green Tech Award, and Research, Academic and Industry Sectors One-plus Scheme (No. RAI/23/1/094A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

Zuankai Wang is an Editorial Board member of The Innovation Energy and was blinded from reviewing or making final decisions on the manuscript. Peer review was handled independently of this member and their research group. The other authors declare no conflicts of interest.