Frontiers in Heat and Mass Transfer

Tech Science Press

https://doi.org/10.32604/fhmt.2026.087420

REVIEW

Electronic Cooling Technologies: A Comprehensive Review of Fundamentals, Advanced Strategies, and Future Directions

Wei Tong[1,2], Chuyi Peng[2] and Yugang Zhao[3,4,*]

[1]State Key Laboratory of Engines, School of Mechanical Engineering, Tianjin University, Tianjin, China
[2]Shenzhen Envicool Technology Co. Ltd., Shenzhen, China
[3]School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai, China
[4]Key Laboratory of Icing and Anti/De-icing, China Aerodynamics Research and Development Center, Mianyang, China

*Corresponding Author: Yugang Zhao. Email: ygzhao@usst.edu.cn

Received: 16 June 2026; Accepted: 03 August 2026; Published: 31 August 2026

ABSTRACT: The relentless pursuit of higher performance in electronics, driven by the demands of artificial intelligence and high-performance computing, has led to unprecedented power densities that exceed the capabilities of conventional cooling methods. This comprehensive review examines the landscape of electronic cooling technologies, spanning from fundamental heat generation mechanisms to state-of-the-art thermal management strategies. We analyze twelve distinct cooling technologies across the full performance spectrum, from conventional air cooling and heat pipes to advanced microchannel heat sinks, jet impingement, spray cooling, immersion cooling, and AI-optimized intelligent thermal management, comparing their heat transfer coefficients (10–250,000 $W/m^2\cdot K$), critical heat fluxes (100–1800 W/cm^2), and coefficient of performance values. By synthesizing recent advances from peer-reviewed studies (2015–2026), this review provides a unified quantitative framework for understanding and selecting electronic cooling technologies, offering actionable engineering guidelines and a forward-looking roadmap for next-generation thermal management. The findings underscore that no single technology can address all thermal challenges, and future cooling solutions will increasingly rely on hybrid architectures, AI-driven optimization, advanced materials, and sustainable coolants to ensure the performance, reliability, and longevity of future electronic systems.

KEYWORDS: Electronic cooling; thermal management; active cooling; passive cooling; AI-optimized cooling

1 Introduction

The exponential growth in computing demand, fueled by artificial intelligence, machine learning, and big data analytics, has placed unprecedented pressure on semiconductor technology. With the continuous scaling of integrated circuits following Moore's law, the power density of modern electronic devices has risen dramatically. Modern processors and AI accelerators, such as NVIDIA's B200 GPU, pack over 208 billion transistors with thermal design powers exceeding 1000 W, while high-performance computing (HPC) chips routinely operate at power densities of 100–300 W/cm^2, with next-generation designs projected to exceed 500 W/cm^2. The reliability of electronic components is exponentially dependent on operating temperature. According to the Arrhenius relationship, the failure rate of silicon chips approximately doubles for every 10°C increase in junction temperature [1]. Overheating accounts for nearly 55% of all electronic module failures. More recent studies [2,3] have refined this relationship, showing that for specific failure mechanisms (electromigration, time-dependent dielectric breakdown, negative bias temperature instability), the temperature sensitivity can vary, with failure rate increases of 5%–15% per 2°C depending on the dominant failure mode.

Over the past four decades, researchers have developed a diverse array of electronic cooling technologies, from conventional air cooling to sophisticated two-phase liquid systems, each with distinct performance characteristics and application domains [4–6]. Air cooling, with its heat transfer coefficient of 10–100 W/m^2·K, remains widely deployed due to simplicity and low cost but is inadequate for high-power-density devices [7,8]. Liquid cooling systems, including cold plates, microchannel heat sinks, and immersion cooling, achieve heat transfer coefficients of 1000–50,000 W/m^2·K and have gained widespread adoption in data centers and high-performance computing (HPC) clusters [9], with manifold microchannel designs achieving heat fluxes exceeding 1700 W/cm^2 [10]. Two-phase cooling technologies, flow boiling, jet impingement, spray cooling, and immersion cooling, leverage latent heat of vaporization to achieve heat transfer coefficients of 50,000–250,000 W/m^2·K and critical heat fluxes exceeding 2500 W/cm^2, with two-phase immersion cooling demonstrating power usage effectiveness (PUE) below 1.05 in data center applications [11–13]. Despite these remarkable advances, significant challenges persist, including flow instabilities and dryout in two-phase systems [14–16], agglomeration and sedimentation in nanofluids [17–19], the inherent trade-off between thermal conductivity and mechanical compliance in thermal interface materials [20–24], and the need for adaptive, energy-efficient cooling solutions for increasingly heterogeneous electronic systems [8,25].

The convergence of AI and HPC has pushed chip power densities beyond the limits of conventional cooling methods, creating an urgent need for more advanced thermal management. Recent breakthroughs across multiple cooling technologies make this an opportune moment to synthesize advances and provide critical engineering guidelines. This review provides a comprehensive, systematic, and up-to-date examination of electronic cooling technologies, spanning fundamental principles, practical implementations, performance comparisons, and future research directions, with the primary objectives of synthesizing the rapidly growing literature, providing clear performance benchmarks, identifying persistent challenges and emerging opportunities, and outlining a roadmap for future innovation. A systematic literature search was conducted in Web of Science, Scopus, and Google Scholar using keywords such as "electronic cooling," "thermal management," "heat pipe," and "AI cooling". Included in this review were important peer-reviewed journal articles (2015–2026), conference proceedings, and foundational pre-2015 works, provided they reported experimental data or validated models with clear performance metrics. Excluded were non-English publications, studies without quantitative data, duplicates, and non-electronic applications, yielding 104 selected publications.

This review distinguishes itself from prior works in four key aspects: (1) We provide the most comprehensive quantitative comparison to date, including heat transfer coefficients, heat flux limits, and coefficient of performance (COP = $Q_{cooling}/P_{input}$) ranges for 12 distinct cooling technologies; (2) We synthesize recent breakthrough developments (2023–2026) in two-phase immersion cooling with guest NCG microbubbles, AI-optimized cooling, and leaf-vein biomimetic microchannels that were merely covered in earlier reviews; (3) We critically analyze persistent challenges including flow instabilities, manufacturing scalability, and environmental sustainability through a unified framework; (4) We provide actionable engineering guidelines and a forward-looking roadmap grounded in the International Roadmap for Devices and Systems (IRDS) projections, identifying specific technology directions for the next decade.

Fig. 1 provides a unified overview of the electronic cooling landscape covered in this review, beginning with the growing thermal pressures from high-performance AI computing chips and summarizing internal heat generation mechanisms in integrated circuits alongside standardized performance evaluation metrics. The diagram systematically classifies cooling technologies into three categories, conventional passive air cooling, advanced single/two-phase liquid cooling, and emerging AI-optimized intelligent thermal management, while visualizing the core structures, heat transfer capacities, and inherent limitations of each approach. It further synthesizes the universal bottlenecks restricting current cooling performance and proposes future research directions, thereby establishing a structural framework for the subsequent in-depth discussion. Accordingly, the remainder of this review is organized as follows: Section 2 presents fundamental heat generation mechanisms and key performance metrics; Section 3 reviews conventional cooling methods including air cooling, heat pipes, and phase change materials; Section 4 examines advanced active cooling techniques such as microchannel heat sinks, jet impingement, spray cooling, immersion cooling, thermoelectric cooling, and AI-optimized cooling; and Section 5 concludes with a perspective on the future trajectory of electronic cooling technology and a summary of the major findings.

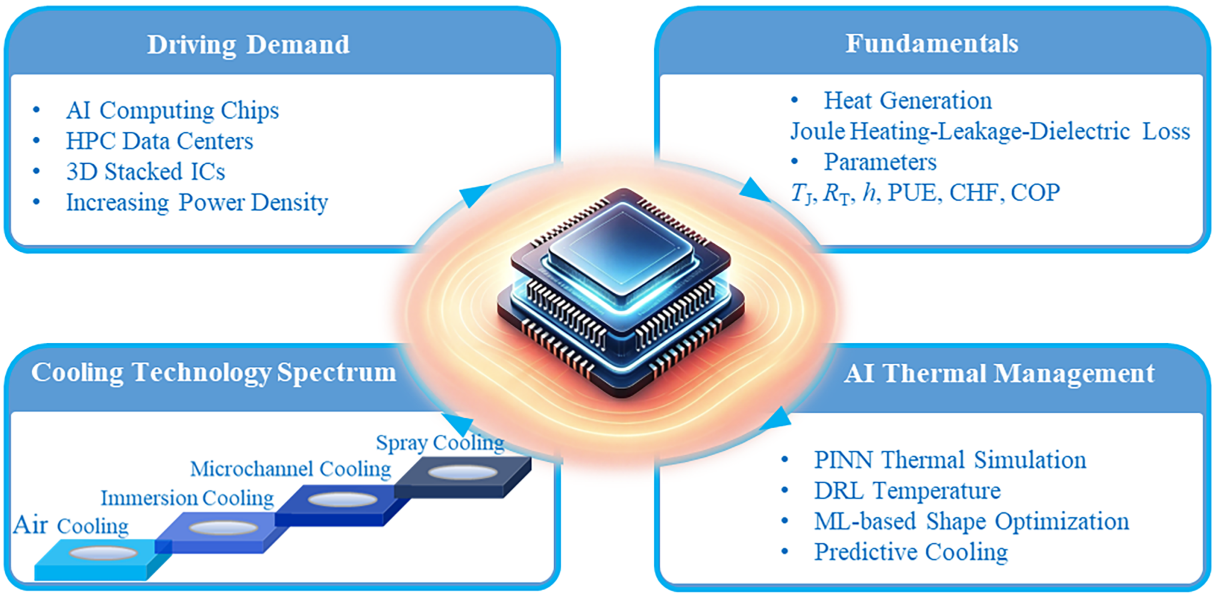

Figure 1: Comprehensive overview framework of electronic cooling technologies: driving demands, fundamental heat transfer mechanisms, full-spectrum thermal management techniques, existing challenges and future research directions.

2 Fundamentals of Heat Transfer in Electronic Systems

2.1 Heat Generation Mechanisms in Integrated Circuits

Heat generation in integrated circuits (ICs) arises from multiple interconnected physical processes that scale with device density and switching frequency [26]. Joule heating (resistive or Ohmic heating) occurs when electrical current traverses conductive pathways, including metal interconnects, vias, and transistor channels, and encounters electrical resistance, dissipating power as $P = I^2R$, with I and R being the electrical current and resistance of the pathways, respectively [8]. This mechanism dominates thermal loads in modern processors, particularly as resistance-capacitance delays and current densities increase with technology node scaling [27]. Switching losses originate from dynamic charging and discharging of parasitic gate capacitances during transistor state transitions, scaling linearly with frequency and quadratically with supply voltage [28]. Leakage currents, gate-oxide tunneling, and reverse-biased junction currents, persist even in the nominally “off” state and increase exponentially with junction temperature, becoming dominant in advanced nodes where threshold voltages are reduced to maintain performance [29]. Dielectric losses emerge in insulating materials (interlayer dielectrics, passivation layers) under time-varying electric fields, where polarization lag converts electrical energy into heat, significant in high-frequency and mixed-signal circuits [30]. In most three-dimensional integrated architectures, through-silicon vias (TSVs) and microbumps serve as dual-purpose electrical interconnects and vertical thermal conduits, introducing resistive losses while providing heat extraction pathways [31].

2.2 Key Performance Metrics

Quantitative assessment of electronic cooling technologies requires a standardized set of performance metrics that enable meaningful comparison across different approaches as shown in Table 1. The most fundamental metric is the junction temperature (T_j), which represents the maximum temperature at the active silicon die. While commercial integrated circuits typically specify safe operating limits between 85°C and 125°C, most devices can withstand junction temperatures up to approximately 150°C without immediate physical damage, though performance degradation and accelerated aging occur below this threshold [32]. Thermal resistance (R_T) in electronic cooling, expressed in the unit K/W, quantifies the temperature increase per unit of dissipated power. Note that this definition differs from the more general formulation in classical heat transfer, providing a practical figure of merit for evaluating passive heat sinks and thermal interface materials.

Table 1: Key performance metrics for electronic cooling.

Parameter	Definition	Typical Range	Significance
Junction temperature (T_j)	Maximum chip operating temperature	<85°C~105°C [32]	Directly determining device reliability
Thermal resistance (R_T)	Temperature rise per unit power ($\Delta T/P$)	0.1~10 K/W	Quantifying cooling system effectiveness
Heat transfer coefficient (h)	Heat flux per unit temperature difference	10~250,000 $W/m^2{\cdot}K$	Measuring convective cooling performance
Power usage effectiveness (PUE)	Ratio of facility power to IT equipment power	1.01~2 [35,36]	Energy useage efficiency
Critical heat flux (CHF)	Maximum heat flux before boiling crisis	100~1682 W/cm^2 [11,37]	Safety limit for phase-change cooling
Coefficient of performance (COP)	Cooling capacity divided by input power	5~30 [33,34]	Energy efficiency metric

For cooling systems, COP indicates energy efficiency. Single-phase liquid cooling systems typically achieve COP values between 10 and 100, while specific cooling technologies such as thermoelectric coolers operate at COP below 1.0 [33,34]. For data center applications, the power usage effectiveness (PUE), the ratio of total facility power to IT equipment power, has emerged as the dominant system-level metric, with conventional air-cooled data centers operating at PUE = 1.5–2.0 and single-phase immersion systems demonstrating PUE values below 1.10 [35,36]. For two-phase and boiling-based cooling systems, the critical heat flux (CHF) represents the maximum heat flux before vapor film formation triggers abrupt temperature rise, a safety limit that can cause catastrophic device failure within milliseconds if exceeded. For microchannel and jet impingement cooling, CHF values exceeding 1000 W/cm^2 have been reported in laboratory settings, though practical implementation typically operates well below these limits [32,37].

3 Conventional Cooling Methods

To facilitate a systematic comparison among the diverse thermal management strategies reviewed in this paper, Table 2 encompasses conventional passive solutions (fin heat sinks, heat pipes, vapor chambers, PCM buffers), established active air and liquid cooling systems (forced air, cold plates, single-phase and two-phase immersion), high-heat-flux techniques (single-phase and two-phase microchannels, jet impingement, and spray cooling), and emerging intelligent thermal management (AI-optimized cooling). For each technology, the table reports the representative energy source, length scale and control modality. These data are compiled from the peer-reviewed literature cited throughout this review and provide a quantitative framework for technology selection based on application-specific thermal, economic, and operation al constraints. Table 2 also provides a quantitative framework for technology selection, enabling engineers to evaluate trade-offs among thermal performance, energy efficiency, and operational constraints for specific applications.

Table 2: Multi-dimensional classification matrix of cooling technologies.

Category	h (W/m^2·K)	CHF (W/m^2)	COP	Complexity	Cost
Natural convection (air) [7]	5–25	<1	—	Low	Low
Forced convection (air) [38,39]	25–250	2–3	—	Low	Low
Cold plate [9,40]	1000–20,000	10–100	50–100	Medium	Medium
Single-phase microchannel [10,41]	10,000–50,000	50–200	100–1000	High	High
Two-phase microchannel [11,42]	50,000–200,000	173–325	695,600	High	High
Jet impingement [43]	30,000–200,000	100–300	10–50	High	High
Spray cooling [12,44]	50,000–250,000	400–1500	10–30	High	High
Single-phase immersion [35,45]	2000–5000	10–50	50–200	Medium	Medium
Two-phase immersion [46,47]	5000–20,000	50–100	4.5–7.6	Medium	High

3.1 Forced Air Cooling

Forced air cooling remains the most established solution in the data center industry, typically implemented through hot/cold aisle containment arrangements. Heated facility air is cooled by computer room air conditioning (CRAC) or air handling (CRAH) units and delivered to cold aisles via underfloor plenums as shown in Fig. 2a [7]. At the server level, electronics are cooled through direct forced convection with dedicated fans driving cold aisle air through individual servers, while heat sinks are required for high-heat-flux components. Air cooling offers advantages in versatility, facilitating easy retrofitting, installation, and reliable hot swapping of servers without the need for liquid coolants or fluidic connectors. These factors have rendered air cooling a prominent thermal management solution despite its inherent limitations. For low-power components such as hard disk drives (HDDs), memory modules, and power units, forced air cooling remains effective and primarily governed by fan rotation speed [48].

Nevertheless, air cooling suffers from fundamental heat transfer limitations due to the poor thermophysical properties of air, requiring lower coolant temperatures and higher flow rates relative to liquid cooling, both of which increase operating expenses [38]. Therefore, it is insufficient for high-performance computing applications where heat generation rates are substantially higher. Extending air cooling's capacity requires increasing heat sink surface area or reducing cold aisle temperatures, both of which are counterproductive given industry trends toward miniaturization and ASHRAE safety specifications. Nevertheless, recent work by Bünnagel et al. [39] demonstrates that strategic modifications to airflow distribution, specifically, the addition of a V-shaped air guide plate to redirect flow through heat sink fins, can achieve excellent temperature uniformity (temperature difference as low as 0.21°C across multiple modules) without increasing overall flow rate or heat sink footprint. This finding suggests that air cooling performance can be enhanced through intelligent flow management rather than brute-force scaling, potentially extending the viability of forced air cooling for certain power electronic applications where temperature uniformity, rather than absolute heat flux, is the primary constraint.

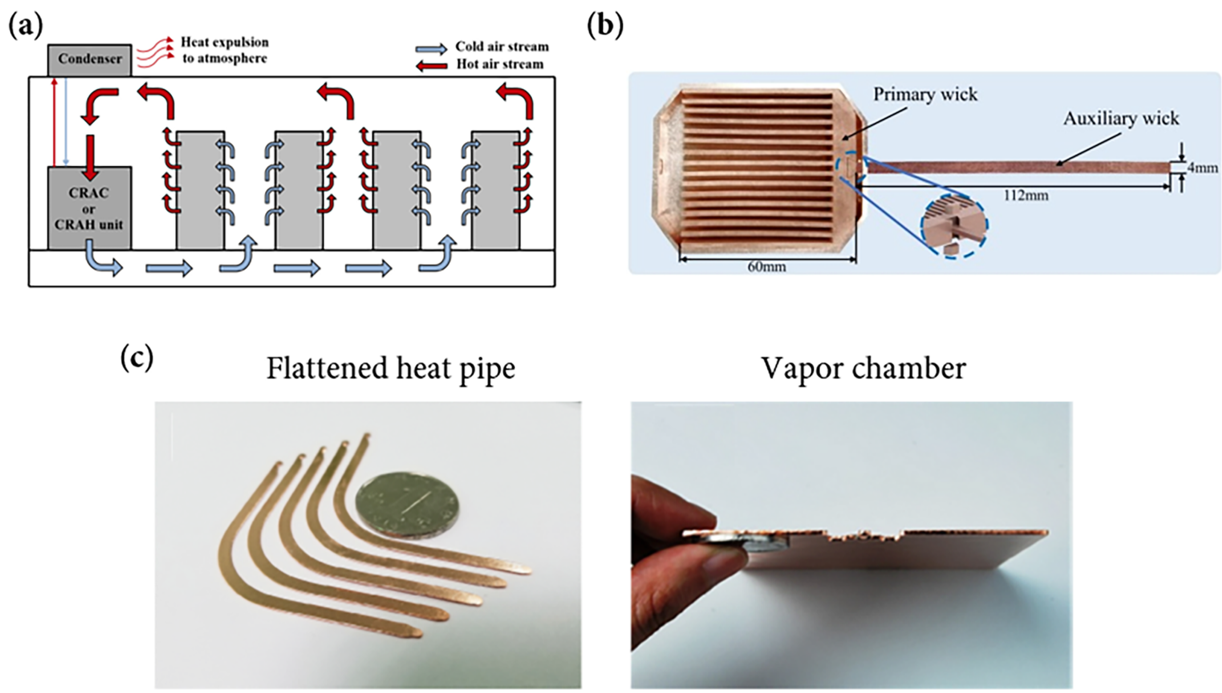

Figure 2: Configuration of conventional heat removal techniques. (**a**) Hot/cold aisle air cooling arrangement [7]; (**b**) The capture of loop heat pipes [49]; (**c**) Captures of flattened heat pipes and vapour chambers [50].

3.2 Fin Heat Sink

Fin heat sinks have undergone significant evolution in recent years, driven by advances in manufacturing techniques and the demand for higher thermal performance. Traditional fabrication methods include extrusion, skiving, and bonded-fin assembly. However, additive manufacturing technologies such as selective laser melting (SLM) now enable topology-optimized fin structures. For instance, Gyroid-based triply periodic minimal surface (TPMS) heat sinks fabricated via SLM have demonstrated superior thermal-hydraulic performance compared to conventional plate-fin designs [51]. The integration of phase change materials (PCMs) with metal foam structures further enhances passive thermal management.

The selection of fin geometry and manufacturing method directly impacts both thermal and economic performance. Circular pin-fin geometries have demonstrated the best thermal performance for two-phase immersion cooling applications, reducing surface temperature by approximately 35% compared to plain surfaces [52]. The porous-fin heat sink represents another promising innovation, where interconnected pores increase effective surface area and allow flow penetration, enhancing heat transfer rates by up to 36.9%, compared to solid-fin designs at similar Reynolds numbers [53]. These geometry-driven enhancements are critical for sustaining nucleate boiling and preventing vapor entrapment at high heat fluxes. Future heat sink development will likely focus on multi-material additive manufacturing, topology optimization, and hybrid passive-active cooling strategies to achieve both high performance and compact form factors.

3.3 Heat Pipes

Heat pipes are passive two-phase devices that exploit latent heat of vaporization and capillary action to achieve effective thermal conductivities of 5000–20,000 W/m·K, far exceeding solid conductors. The operating cycle involves evaporation at the hot end, vapor transport, condensation at the cold end, and liquid return via a wick structure. Compared to conventional fin heat sinks, heat pipes undergo passive operation, high reliability with no moving parts, and flexible form factors ranging from ultra-thin flattened pipes for smartphones to large-diameter pipes for aerospace applications. For loop heat pipes as shown in Fig. 2b, Tian et al. [49] proposed a dual gradient wicks relay design that achieved a cooling capacity of 288 W with junction temperature below 85°C, a minimum total thermal resistance R_T of 0.20°C/W, and a nominal PUE of 1.0015 while simultaneously generating up to 280 mW of electrical power. The self-venting annular wick represents another advancement, achieving a 29.2% enhancement in capillary limit over conventional designs.

Recent advances in additive manufacturing have also enabled complex wick geometries that overcome the trade-off between capillary pressure and permeability. Laser Powder Bed Fusion techniques have produced triangular, rectangular, trapezoidal, and arterial groove configurations, with trapezoidal and arterial grooved wicks demonstrating superior performance [54]. For silicon-based loop heat pipes targeting IC cooling, comparative studies revealed that in-line long rib wicks and staggered long rib wicks achieve maximum effective thermal conductivities of 848 and 860 W/m·K, respectively, substantially higher than conventional micropillar or microchannel designs [55]. Ultra-thin flattened heat pipes (UTHPs) with composite copper foam and mesh wicks have achieved maximum heat transport capacities of 9 W at 0.6 mm thickness as shown in Fig. 2c [50]. These structure-driven enhancements demonstrate that careful wick engineering can significantly extend heat pipe thermal transport limits, making them viable for high-performance computing, electric vehicle battery thermal management, and aerospace applications where passive, reliable, high-capacity heat transfer is critical. Future developments will likely focus on multi-material additive manufacturing, nanofluid-enhanced working fluids, and hybrid passive-active cooling strategies.

3.4 Vapor Chambers

Vapor chambers (VCs) are two-dimensional planar heat spreaders that operate on the same phase-change principle as heat pipes but spread heat radially rather than linearly, achieving effective thermal conductivities of 5000–20,000 W/m·K. The wick structure is the critical component governing VC performance, determining both capillary pressure (for liquid return) and permeability (for flow rate). Tan et al. [56] systematically compared hierarchical copper mesh wick designs, demonstrating that upward-coarser porous structures achieve 26.5% lower thermal resistance than reverse-gradient counterparts, with optimized wicks sustaining high-frequency bubble rupture at frequencies up to 2000 Hz under high heat flux. For ultra-thin flexible vapor chambers targeting foldable electronics, Gao et al. [57] fabricated 0.34-mm-thick graphene-film-based VCs using high-precision 3D printing, achieving an effective thermal conductivity of 5800 W/m·K and ultra-low thermal resistance of 1.2 K/W under 15 W heating power, with minimal performance degradation under bending deformation. The capillary performance factor (K/R_{eff}), which reaches 1.192×10^{-6} m for cauliflower-like electrodeposited wicks, directly determines the wick's ability to sustain liquid supply under high heat flux.

Recent advances in structural optimization have pushed vapor chamber performance beyond conventional wick designs. Huang et al. [58] introduced a novel step-gap structure above the boss area, creating local vapor acceleration and pressurization that reduced thermal resistance by 50.9% compared to conventional VC designs, with maximum heat load extended to 550 W. Functional support structures in L-shaped ultra-thin vapor chambers have achieved maximum thermal conductivity of 8233.82 W/m·K, representing a 334% improvement over conventional designs [59]. For large-area thermal management, leaf-inspired vapor chambers fabricated by roll-bond processing achieved equivalent thermal conductivity up to 43,188 W/m·K at 300 W power output, with thermal resistance as low as 0.022°C/W [60].

3.5 Phase Change Material Buffer

Phase change materials (PCMs) provide passive thermal buffering by absorbing or releasing large amounts of latent heat during solid-liquid phase transitions while maintaining a nearly constant temperature. Typical PCMs for electronics include paraffin waxes (melting points 30°C–60°C, latent heat 150–250 kJ/kg), fatty acids, and hydrated salts. Integrated into heat sinks or packaging, PCMs absorb transient thermal loads, delaying temperature rise and protecting components during peak power events. For electric vehicle batteries, composite PCMs with expanded graphite maintain battery temperatures below 50°C even at 5C discharge rates. However, the inherently low thermal conductivity of organic PCMs (0.1–0.6 W/m·K) limits heat transfer to small fluxes (typically <10 W/cm^2).

Enhancement strategies employ thermal conductivity enhancers including fins, metal foams, and nanoparticles. Triangular pin fins achieve superior thermal performance with an optimal volume fraction of approximately 9%, though excessive fins suppress natural convection as shown in Fig. 3a [61]. Copper and aluminum foams (75%–98% porosity) increase heat transfer rates by up to 36% compared to pure PCM as shown in Fig. 3b [62]. Dispersing nanoparticles creates nano-enhanced PCMs: 1% volume fraction of Cu nanoparticles increases thermal conductivity by 31%, while hybrid carbon-based nanoparticles achieve 96% enhancement [63]. Encapsulation in polymer shells or porous hosts prevents leakage, while hybrid systems combining PCMs with active cooling further improve performance.

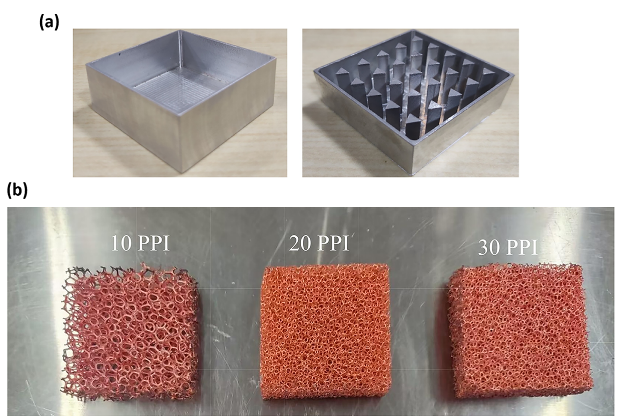

Figure 3: Physical cells that carry PCM. (**a**) Pictorial diagram the containers with and without fins [61]; (**b**) Open-cell copper foams with three pore sizes [62].

PCM emulsions represent an advanced class of thermal management fluids that combine high latent heat storage with the pumpability of conventional liquid coolants. These colloidal dispersions consist of PCM droplets suspended in a carrier fluid (usually water) with appropriate surfactants, offering high thermal energy storage density and minimal interface thermal resistance. Stable formulation requires careful selection of surfactants, preparation methods (ultrasonication or high-pressure homogenization), and nucleating agents to mitigate subcooling. Subcooling delayed solidification below the freezing point is particularly pronounced in small droplets (<20 μm). Adding nucleating agents such as higher-melting-point paraffins or carbon nanotubes reduces subcooling from over 15°C to as low as 1°C–2°C [64]. In practice, PCM based heat sinks can reduce electronic device operating temperatures by 6°C–10°C and provide more uniform temperature distributions, effectively delaying the thermal response during transient high-power operation [65].

4 Advanced Active Cooling Techniques

4.1 Cold Plate

The cold plate is one of the most widely implemented liquid-based active cooling devices for electronic systems, serving as the interface between heat-generating components and circulating coolant. Unlike heat pipes or vapor chambers which rely on passive capillary-driven phase change, cold plates operate on forced convection principles, where a liquid coolant (typically water, ethylene glycol/water mixture, or dielectric fluids) is pumped through internal channels to absorb and transport heat away from the component. The internal channel architecture is the critical design parameter determining both thermal resistance and hydraulic pressure drop. The coolant selection directly impacts thermal performance: water offers the highest specific heat capacity (4182 J/kg·K) and thermal conductivity (0.6 W/m·K), while dielectric fluids or refrigerants are required for applications where electrical conductivity poses safety concerns.

Recent advances in cold plate design have focused on enhanced channel geometries and topology optimization. Wang et al. [40] investigated cold plates with airfoil-shaped pin fins for prismatic lithium-ion battery packs, demonstrating that the airfoil design reduces recirculation zones and pressure losses, achieving a maximum battery temperature of 31.5°C and a temperature difference of 3.8°C under 3C discharge, representing a 68% reduction in temperature rise compared to conventional designs. Chainer et al. [66] proposed a liquid cooling system without a chiller using two-phase cooling technology compatible with 3D chips, demonstrating a significant reduction in cooling energy while maintaining chip temperatures below 85°C, with the transition from single-phase to two-phase flow achieving heat transfer coefficients 3–5 times higher than single-phase operation. For electric vehicle battery thermal management, hybrid systems combining cold plates with phase change materials provide continuous heat rejection during steady-state operation while buffering transient thermal spikes during aggressive driving cycles [67]. The continued push toward higher power densities will drive further innovation in cold plate design through additive manufacturing, topology optimization, and integration with hybrid thermal management strategies.

4.2 Single-Phase Microchannel Cooling

Single-phase microchannel cooling operates by forcing the liquid coolant through channels with hydraulic diameters below 1 mm, achieving heat dissipation of several hundred watts per square centimeter. The performance advantage arises from two fundamental effects: the area-to-volume ratio scales inversely with channel diameter, and the convective heat transfer coefficient scales inversely with hydraulic diameter for laminar flow. Unlike two-phase systems which exploit latent heat but suffer from flow instabilities and critical heat flux dryout, single-phase microchannel cooling offers stable, predictable operation with no risk of sudden thermal failure. This reliability makes it particularly attractive for applications such as data center processors and power electronics where uninterrupted operation is critical. A recent breakthrough in co-designing electronics with microfluidics has further enhanced performance by integrating power devices and cooling channels on the same substrate as shown in Fig. 4a, eliminating interfacial thermal resistance and improving overall system efficiency [10].

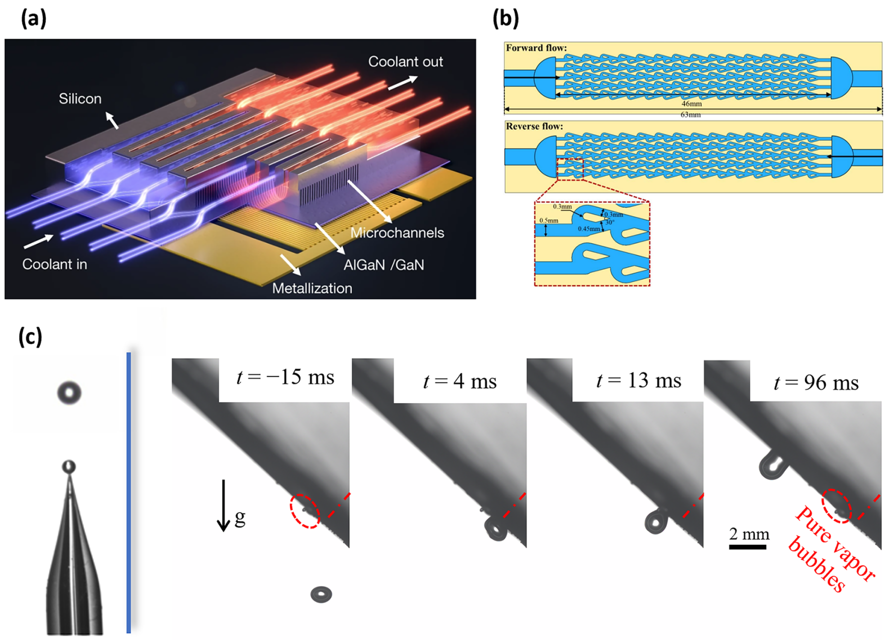

Figure 4: Schematic illustrations of advanced cooling technologies and their underlying fluid dynamics. (**a**) Schmatics show how the fluid flow through a monolithically integrated manifold microchannel heat sink in single-phase microchannel cooling [10]. (**b**) Configuration of reverse and forward Tesla-type microchannel in two-phase microchannel cooling [68]. (**c**) Air bubble dynamics when impinges onto the heating surface in two-phase immersion cooling [69].

Optimizing single-phase microchannel performance requires careful balancing of three interrelated factors: channel geometry, flow distribution, and coolant properties. Rectangular channels are most common due to fabrication simplicity, but increasing the aspect ratio (height-to-width) expands heat transfer area at the cost of higher pressure drop. Manifold microchannel designs overcome this trade-off by using multiple inlet and outlet manifolds to distribute flow across short parallel channels, keeping temperature rise uniform across the chip surface while reducing pressure drop by an order of magnitude compared to conventional serpentine designs [70]. For applications demanding even higher performance, wavy channels induce chaotic mixing and repeated boundary layer restarting, achieving 30%–80% heat transfer enhancement over straight channels with a moderate pressure drop penalty [71]. Where electrical isolation is required, such as in power electronics or immersion cooling, dielectric fluids are necessary, but their lower thermal conductivity (0.06–0.10 W/m·K) reduces performance, requiring higher flow rates or smaller channels at the expense of pumping power. A recent comprehensive review [72] specifically addressed microchannel heat sinks for cold plate liquid cooling in data centers, providing detailed evaluations of channel geometries, manifold designs, and system-level performance. While their work focuses exclusively on microchannel cold plates, our review extends the analysis to encompass the full spectrum of cooling technologies from air cooling to AI-optimized systems, enabling a holistic technology selection framework.

Beyond thermal performance and coolant selection, the practical deployment of single-phase microchannel heat sinks is critically constrained by manufacturing capabilities and scalability, with three primary fabrication methods dominating current practice: deep reactive ion etching (DRIE) for silicon microchannels (±2~5 μm tolerance), laser ablation for metal microchannels (±10~20 μm), and additive manufacturing for complex geometries (±25~50 μm) [73]. These fabrication tolerances directly impact thermal-hydraulic performance through surface roughness, which increases pressure drop by 10%–15% compared to smooth-channel predictions [74], and channel dimension non-uniformity, where a 10%

variation in hydraulic diameter can produce up to 25% variation in flow distribution among parallel channels, creating localized hotspots that negate the benefits of microchannel cooling [75]. Cost and scalability present additional barriers: DRIE offers excellent wafer-level scalability but entails high capital costs, making it viable primarily for high-value applications such as server processors; laser ablation provides a lower-cost alternative for metal components with moderate complexity; while additive manufacturing, though offering unparalleled design freedom, remains the most expensive option on a per-part basis. Scale-up from laboratory prototypes to production volumes introduces further challenges. Panel-level fabrication (300 mm × 300 mm or larger substrates) could reduce unit costs by 30%–50% but requires new equipment and process development [76], and integration with existing manufacturing supply chains demands standardization of form factors, fluidic connectors, and testing protocols, areas where industry consensus is still emerging [77,78].

Despite these manufacturing advances, several critical system-level challenges must be addressed before single-phase microchannel cooling achieves widespread deployment. Pumping power consumption can be significant in conventional designs, though advanced embedded liquid cooling architectures have demonstrated pumping power below 1% of total chip power through passive flow distribution to hotspots [41]. Future research directions should focus on additive manufacturing of triply periodic minimal surface structures that enhance heat transfer while reducing pressure drop, integration with AI-optimized flow control for adaptive thermal management, and hybrid designs that transition between single-phase and two-phase modes based on real-time thermal load. As heterogeneous integration and 3D chip stacking push heat fluxes beyond 1000 W/cm^2, further innovation in microchannel cooling architectures will be essential to maintain device reliability and performance.

4.3 Two-Phase Microchannel Cooling

Two-phase microchannel cooling leverages the latent heat of vaporization during flow boiling to achieve heat transfer coefficients substantially higher than single-phase systems, with reported cooling capacities of 950–990 W/cm^2 [42]. Unlike single-phase cooling where heat removal is limited by the coolant's sensible heat capacity, two-phase flow boiling absorbs thermal energy through liquid-to-vapor phase change, maintaining more uniform wall temperatures while handling extreme heat fluxes. However, a fundamental trade-off lies in flow instability: as vapor bubbles nucleate, grow, and coalesce within confined microchannels, they can cause flow reversal, pressure drop oscillations, and localized dry-out conditions where the liquid film on the channel wall evaporates completely, causing sudden temperature spikes that may lead to device failure. Therefore, suppressing flow instabilities while preserving high heat transfer performance has become the central challenge in two-phase microchannel cooling.

To address this challenge, recent innovations have focused on advanced channel geometries and passive flow control mechanisms. Tesla-type microchannels, inspired by Nikola Tesla's valvular conduit, have emerged as a promising solution. As shown in Fig. 4b, Yang et al. [68] demonstrated that the synergistic interaction between the main channel and secondary channels enhances heat transfer, strengthens flow stability, and promotes gas-liquid separation, achieving a 45.1% reduction in chip thermal-mechanical displacement compared to conventional parallel channels. For the more critical dry-out instability, Feng et al. [79] proposed a parallel microchannel system integrated with topological inlet restrictors and bypass channels, which generates periodic two-phase micro-jets at frequencies of 250–333 Hz. These high-frequency jets continuously replenish subcooled liquid, break up vapor slugs, and rewet the heated wall, limiting dry-out duration to only 0.5–1 ms while increasing the overall heat transfer coefficient by 37.4% and the critical heat flux by 28.5%. Beyond these designs, biomimetic structures have extended two-phase cooling research beyond conventional geometries. Ma et al. [80] found that an optimized leaf-vein biomimetic microchannel achieved a heat transfer coefficient 8 $kW/m^2 \cdot K$ higher than rectangular microchannels while reducing pressure drop by 9% through multi-objective genetic algorithm optimization. Collectively, these studies demonstrate that carefully engineered channel geometries can effectively mitigate flow instabilities while simultaneously enhancing thermal performance.

To date, several challenges remain before two-phase microchannel cooling achieves higher performance. Predictive modeling of pressure drops, a critical design parameter, remains difficult, with conventional homogeneous and separated flow models often under- or over-estimating values by up to 30%. Long term reliability under thermal cycling conditions, where repeated expansion and contraction of vapor bubbles can induce mechanical fatigue in channel walls, has not been thoroughly investigated. Scaling these designs from laboratory prototypes to industrial manufacturing also presents significant fabrication and cost barriers. Nevertheless, these advances position two-phase microchannel cooling as a critical enabling technology for next generation high-heat-flux electronics, particularly for applications such as 3D-stacked chips and high-power radar systems where heat fluxes routinely exceed 500 W/cm^2.

4.4 Single-Phase Immersion Cooling

Single-phase immersion cooling (SPIC) has emerged as a transformative direct liquid cooling paradigm that fully submerges electronic components in dielectric fluids, eliminating the need for fans, heat sinks, and intermediate thermal interfaces. Unlike indirect cold plate systems where significant heat still requires air cooling, SPIC achieves complete thermal management through direct contact between all heat generating components and the circulating coolant, substantially increasing the effective heat transfer area. SPIC systems are classified into two operating modes: natural convection (buoyancy-driven) and forced convection (pump-driven). Experimental studies have demonstrated that forced convection immersion cooling using SiC and white mineral oil nanofluids increases the maximum surface heat transfer coefficient by 11.3% at 100 W for bare heat sources, and by 11.7% at 300 W with heat sink auxiliary cooling [45]. Compared to two-phase immersion cooling, which requires complex pressure control and faces volatility challenges, SPIC offers simpler heat transfer mechanisms, excellent operational stability, and greater suitability for large-scale commercial deployment.

The selection of dielectric fluids is the most critical design decision governing SPIC thermal performance. Mineral oil offers high heat capacity, excellent electrical insulation, chemical stability, and high flash point, making it the preferred coolant for SPIC systems. To further enhance thermal performance, nanofluid modifications have been extensively investigated. For instance, surface modification of nanoparticles with hexadecyl trimethoxy silane (HDTMS) has been shown to enhance dispersion stability in mineral oil, with modified nanofluids maintaining 59.89% transmittance after 30 days of static settling and achieving interaction energy increases of 51.92% at 343 K [81]. Extending beyond mineral oil-based systems, h-BN nanosheet nanofluids in ester oils have demonstrated significant cooling capability for battery immersion cooling applications, achieving average battery pack temperatures of 26.55°C at 6 L/min flow rate, a 40.62% reduction compared to natural convection conditions.

Despite the promising thermal performance of single-phase immersion cooling, several practical challenges must be addressed to facilitate widespread commercial deployment. Material compatibility represents a primary concern, as prolonged contact between dielectric fluids and electronic materials can lead to degradation and failure. For example, EPDM rubber, commonly used for seals and O-rings, has been observed to swell by 15%–20% after one year of continuous immersion in mineral oil, compromising sealing integrity [15]. Similarly, solders and metallic interconnects are susceptible to corrosion, particularly when fluids contain trace moisture or acidic degradation byproducts, which can increase electrical contact resistance and lead to premature device failure. Polymer materials used in cable insulation may also experience plasticizer extraction or embrittlement, further complicating long-term reliability. Thermal stability of the dielectric fluid itself is another critical factor, as prolonged exposure to elevated temperatures promotes oxidative degradation, generating acidic byproducts, sludge, and particulate matter that alter the fluid's dielectric strength and heat transfer properties, with degradation products depositing on heat transfer surfaces and increasing thermal resistance over time [82]. Maintenance considerations add further complexity: unlike air-cooled systems where components are readily accessible, immersion cooling requires specialized procedures for component replacement, often necessitating system drainage, fluid filtration, and careful handling to prevent contamination. While *in-situ* filtration systems can extend fluid service life, periodic fluid sampling and analysis are essential to monitor oxidation, acid number increase, or moisture ingress, adding to operational overhead [83]. The capital expenditure for single-phase immersion cooling is significantly higher than conventional air cooling due to tank construction, fluid procurement (typically $15–30 per liter), and fluid handling infrastructure, with initial investments estimated at 2–5 times that of equivalent air-cooled systems. However, this higher upfront cost is often offset by reduced operating expenditure, as immersion systems eliminate the need for computer room air conditioning and allow higher cooling water temperatures, resulting in annual energy savings of 30%–50% and PUE values consistently below 1.10 compared to 1.5–2.0 for conventional air-cooled data centers [84]. The economic viability therefore depends critically on deployment scale, with payback periods typically ranging from 2 to 5 years for large-scale installations. Despite these challenges, ongoing research on compatible materials, advanced fluid formulations with enhanced oxidative stability, and standardized maintenance protocols is progressively mitigating these practical barriers, positioning single-phase immersion cooling as an increasingly attractive solution for high-density computing environments.

4.5 Two-Phase Immersion Cooling

Two-phase immersion cooling leverages the latent heat of vaporization during pool boiling to achieve heat transfer coefficients substantially higher than single-phase systems. Unlike single-phase immersion cooling where heat removal is limited by the coolant's sensible heat capacity, two-phase immersion cooling absorbs thermal energy through liquid-to-vapor phase change, maintaining more uniform device temperatures while handling extreme heat fluxes. At the data center level, experimental investigations have demonstrated that two-phase immersion cooling systems can maintain chip surface temperatures at approximately 68.3°C even at a high chip power of 600 W, with system pressure increasing by only 0.96 kPa when total server power rises from 0 to 1500 W, demonstrating excellent temperature control capability and system pressure stability [46]. However, a fundamental challenge lies in the presence of non-condensable gases (NCGs), which have long been believed to hinder heat transfer by blocking physical contact between the heating surface and the liquid coolant.

Contrary to this common belief, Luo et al. [69] demonstrated that guest NCG microbubbles can actually enhance heat transfer by up to one order of magnitude through three distinct mechanisms. The first mechanism is thermal boundary layer disturbance: upward-moving NCG microbubbles penetrate the thermal boundary layer (thickness ~ 0.1–1 mm, comparable to bubble size) and bring cold coolant directly to the heating surface, thereby reducing local temperature and enhancing heat transfer. The second mechanism is evaporation nuclei activation: as NCG microbubbles approach the heated surface, a sustained vapor concentration gradient is established between the wall temperature ($P_{w,sat}$) and the pool temperature ($P_{\infty,sat}$), driving continuous bubble growth. Unlike pure vapor bubbles generated from *in-situ* boiling nucleation which cease to grow once saturated at the wall temperature, guest air bubbles continue to grow appreciably as shown in Fig. 4c, reaching over eight times their original volume, due to the persistent presence of NCGs. The third mechanism is vapor film disruption: when NCG microbubbles collide with and merge into existing vapor films or large bubbles, they trigger buoyancy-assisted departure at a critical size. The merged bubble then slides along the heating surface, engulfing tiny pure vapor bubbles and leaving a clean trail behind, thereby enabling a new round of nucleation and preventing the formation of a sustained vapor film that would otherwise deteriorate heat transfer.

The selection of dielectric fluids and the optimization of boiling surfaces are the critical design decisions governing two-phase immersion cooling performance. For two-phase immersion cooling systems, CHF, the maximum sustainable heat flux before vapor film formation causes sudden temperature rise, is a key performance metric. For dielectric fluid YL-10 boiling on a smooth surface, the CHF is approximately 20.55 W/cm^2, with symmetrical vortexes observed above the chip during boiling that enhance heat transfer [46]. For vertically mounted electronics with interchip components such as capacitors and inductors, optimal printed circuit board topology achieves a maximum heat flux of 18 W/cm^2 and a heat transfer coefficient of 9650 $W/m^2 \cdot K$ using dielectric hydrofluoroether HFE-7100 at an excess wall temperature of 20 K [47]. Interchip components also play a critical role in deflecting bubble paths, promoting bubble coalescence, and reducing vapor coverage on chip surfaces.

Despite the exceptional heat transfer performance of two-phase immersion cooling, several engineering challenges must be overcome to enable large-scale commercial deployment. Vapor condensation recovery presents a primary concern, as the efficiency of external condenser systems, requiring water temperatures 5°C–10°C below the refrigerant condensation temperature, can limit free-cooling opportunities in warm climates [46]. Furthermore, non-condensable gas accumulation, which can enter through leaks or dissolved air release, significantly degrades condenser performance by creating a stagnant gas layer that increases thermal resistance, necessitating active management systems such as vacuum pumps or purge units [69]. Emission control poses another critical challenge, as many dielectric fluids have high global warming potentials and face increasingly stringent regulations under the Kigali Amendment to the Montreal Protocol. While hermetically sealed systems minimize emissions during normal operation, maintenance activities can release vapor, requiring capture systems or containment tents. The development of low-GWP alternatives, including hydrofluoroolefins, fluoroketones, and silicone-based fluids, has therefore become an urgent priority, though these substitutes often exhibit lower heat transfer coefficients or narrower operating temperature ranges. Pressure control adds further complexity, as two-phase systems operate at saturated conditions where system pressure is directly coupled to temperature. Fluctuating heat loads cause corresponding pressure variations that can challenge seal integrity and vessel walls, necessitating robust regulation strategies such as expansion tanks, pressure relief valves, or active control systems to maintain stable operation across varying server power demands. Safety considerations further influence design: while many dielectric fluids are non-flammable, some low-GWP alternatives exhibit higher flammability, requiring careful risk assessment and potentially additional fire suppression. Toxicity concerns, though generally low, necessitate proper handling procedures, personal protective equipment, and emergency response planning, particularly for larger installations where fluid inventory can exceed thousands of liters [85]. The complexity of these engineering challenges, spanning thermal management, environmental compliance, pressure control, and safety, highlights the need for integrated system design approaches and explains why two-phase immersion cooling, despite superior thermal performance, has seen slower commercial adoption compared to single-phase systems.

4.6 Jet Impingement Cooling and Spray Cooling

Jet impingement and spray cooling are two ultra-high-performance liquid based active cooling technologies that achieve among the highest heat transfer coefficients (30,000–250,000 $W/m^2 \cdot K$) and critical heat fluxes (400–1800 W/cm^2) of all cooling methods [5]. Both rely on direct liquid contact with heated surfaces, require high-pressure pumping, benefit from surface modifications, and need closed-loop coolant management. However, they differ fundamentally in their fluid delivery mechanisms. Jet impingement directs continuous high-velocity fluid jets perpendicular to the heated surface, achieving higher peak heat fluxes (up to 1800 W/cm^2) but suffers from poor temperature uniformity, intense cooling at the stagnation point creates thermal gradients. Spray cooling atomizes liquid into fine droplets (10–100 μm diameter) that form a thin evaporating film, offering superior temperature uniformity at somewhat lower peak heat fluxes (400–1500 W/cm^2). Consequently, jet impingement is preferred for targeted hotspot cooling (gas turbine blades, laser diodes, power modules), whereas spray cooling is favored for applications demanding uniform temperature distribution across larger areas (radar arrays, concentrated photovoltaics).

Jet impingement cooling directs high-velocity fluid jets perpendicular to heated surfaces, with heat transfer coefficients scaling with Reynolds number to the 0.5–0.7 power. Optimal performance for single jets occurs at nozzle-to-surface spacing ratios (H/D) of 5–8, while multiple jet arrays achieve optimal jet-to-jet spacing ratios (S/D) of approximately 3. Recent advances have pushed performance boundaries for high-power computing systems. Patel et al. [43] experimentally demonstrated multi-chip jet impingement cooling for a 2.5D integrated system containing a logic chip and four high-bandwidth-memory (HBM) chips, achieving a 1.86 kW thermal design power under a logic chip temperature constraint of 105°C, with a minimum thermal resistance of 0.183 cm^2·K/W and a logic chip heat flux of 252 W/cm^2. Surface temperature measurements over the 676 mm^2 logic chip revealed uniformity within 3°C even at the highest thermal design power. For cryogenic applications, Qenawy et al. [86] incorporated cryogen spray cooling with sweeping cold air jets, demonstrating a 26%–36% heat transfer coefficient improvement and a 52% increase in critical heat flux time compared to conventional cryogen spray cooling.

Spray cooling atomizes liquid into fine droplets (10–100 μm diameter) that impinge on heated surfaces, achieving exceptional performance through three simultaneous mechanisms: droplet impingement, thin-film evaporation, and nucleate boiling. Recent innovations have focused on surface structuring and multi-objective optimization. Inspired by leaf veins, researchers designed multi-scale structured surfaces (reticulated and parallel veins combined with graphene or carbon nanotube coatings) for dielectric liquid spray cooling using HFE-7100. The reticulated vein structure outperformed the parallel vein design due to superior liquid distribution, while graphene coating exhibited better heat transfer than carbon nanotube coating due to larger surface area and higher thermal conductivity. The multi-scale structured surface achieved a 116% increase in critical heat flux compared to a smooth surface, with evaporation efficiency reaching 60% at 80°C [44]. For high-evaporation-efficiency spray cooling, researchers achieved a critical heat flux of 143.2 W/cm^2 and an evaporation efficiency of 93.4% with HFE-7100 at 120 kPa, developing an empirical CHF correlation with a maximum deviation below 6.7% [87]. For electro-spray cooling at elevated flow rates (800 mL/h, 6 kV), a critical heat flux of 100.8 W/cm^2 and a maximum heat transfer coefficient of 14,416 W/m^2·K were achieved, with spray morphology classified into six distinct modes [88]. Despite their exceptional performance, practical implementation faces challenges: jet impingement requires precise nozzle alignment and risks surface erosion, while spray cooling struggles with nozzle clogging and overspray losses. Both technologies require high-pressure pumping and closed-loop coolant management, adding system complexity and cost that must be weighed against their unparalleled heat dissipation capabilities.

4.7 AI-Optimized Cooling

AI-optimized cooling represents an emerging paradigm that integrates machine learning and artificial intelligence with conventional cooling hardware to achieve unprecedented energy efficiency and adaptive thermal management. Unlike static or simple feedback-loop controllers (e.g., PID controllers), AI-optimized systems employ neural networks, reinforcement learning, and predictive models to anticipate thermal loads and proactively adjust cooling parameters. The core innovation lies in shifting from reactive cooling, responding to temperature thresholds, to predictive cooling, which anticipates thermal events before they occur, thereby minimizing both overcooling (energy waste) and undercooling (thermal risk) [4,42]. Fig. 5 provides a schematic overview of three representative AI-optimized cooling approaches across different scales, from chip-level prediction to device-level control and design optimization.

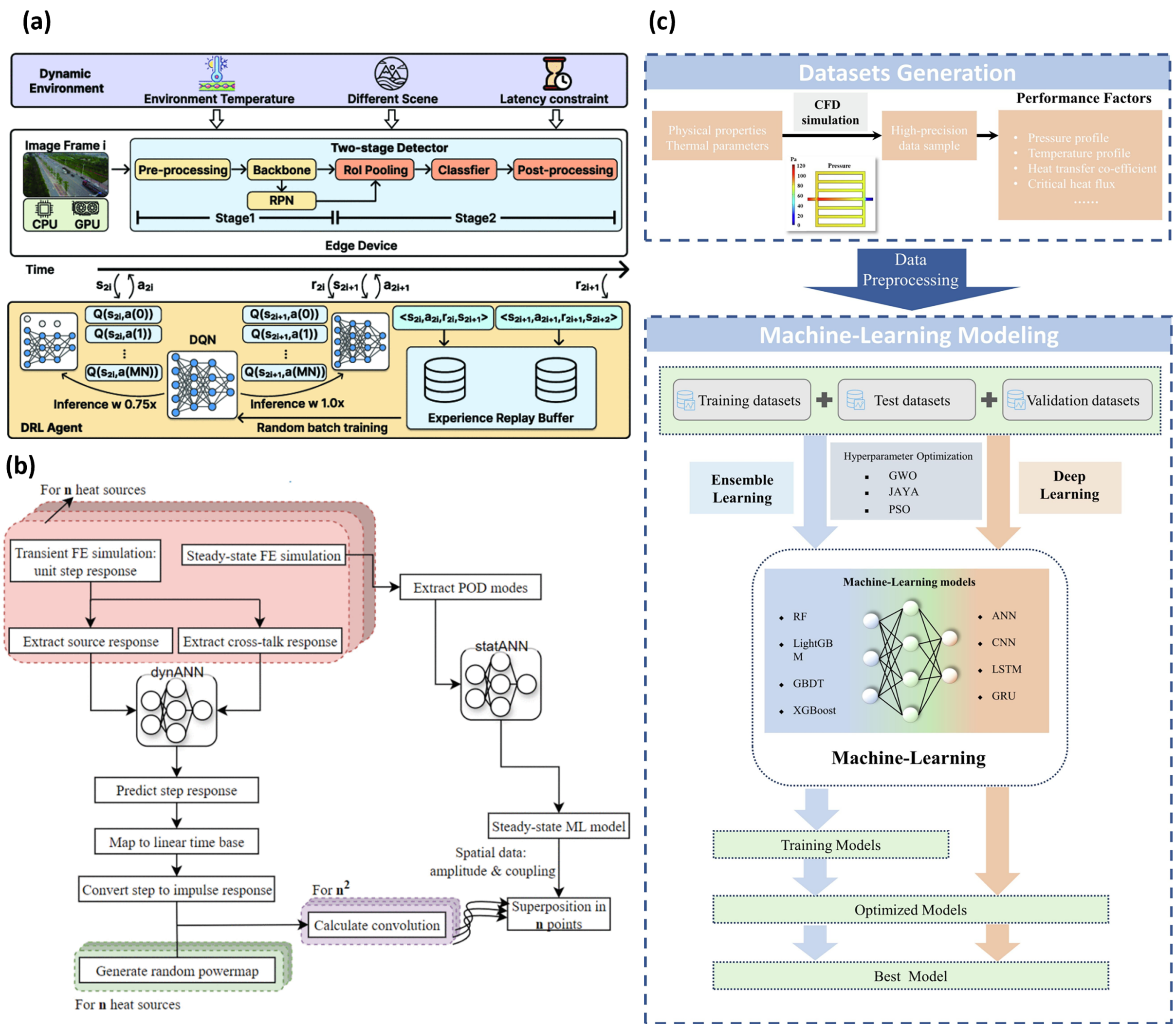

Figure 5: Schematic illustration of the principles of AI-optimized cooling across different scales. (**a**) Deep reinforcement learning (DRL) framework for joint CPU/GPU frequency scaling on edge devices [89]; (**b**) Physics-informed neural network (PINN) for chip-level dynamic thermal simulation [90]; (**c**) Process showing applying ML to the design of thermal management structures [42].

4.7.1 Deep Reinforcement Learning for Dynamic Control

Deep reinforcement learning (DRL) has emerged as a powerful approach for dynamic cooling control, enabling real-time flow regulation and frequency scaling in edge devices and data centers [89], as illustrated in Fig. 5a. In this framework, a DRL agent interacts with the cooling system by observing state variables, including component temperatures, operating frequencies, task latency, and coolant flow rates, and takes actions such as adjusting pump or fan speeds and frequency scaling decisions. The agent is trained to maximize a reward function that balances competing objectives: thermal safety (maintaining junction temperatures below critical thresholds, typically 85°C–100°C), computational performance (minimizing latency variation), and energy efficiency (minimizing pumping power). Training typically employs proximal policy optimization in simulation environments, with subsequent fine-tuning on physical hardware to account for system-specific dynamics. This approach has demonstrated 30.8% faster inference and 72.8% reduction in latency variation compared to conventional methods for edge device applications. For data center-scale dynamic flow regulation, DRL controllers have achieved 15%–25% energy savings compared to traditional PID controllers by proactively adjusting cooling delivery based on predicted rather than current thermal loads [91]. The advantage is most pronounced during transient workload conditions, where PID controllers exhibit significant overshoot and oscillation, whereas DRL agents learn smooth, anticipatory control policies that maintain stable temperatures with minimal energy expenditure.

4.7.2 Physics-Informed Neural Networks for Thermal Prediction

Physics-informed neural networks (PINNs) have emerged as a breakthrough technique for dynamic thermal simulation, enabling accurate and rapid prediction of temperature fields at both chip and data center scales [90], as illustrated in Fig. 5b. Unlike purely data-driven approaches that require extensive training data, PINNs incorporate fundamental physical laws, including energy conservation, the Navier-Stokes equations, and heat conduction principles, directly into the neural network loss function. This integration ensures predictions remain physically consistent even in regions with sparse training data, significantly improving generalization performance [92]. For chip-level applications, PINNs incorporate proper orthogonal decomposition modes and spatial-temporal decoupling to predict transient temperature fields under varying power maps and boundary conditions, achieving prediction errors below 2% for heat fluxes up to 100 W/cm^2, compared to 5% or higher for conventional machine learning approaches [93]. At the data center level, PINNs have been extended to model airflow and temperature distribution across entire server racks, with physical constraints ensuring predictions satisfy mass and energy conservation. The computational efficiency of PINNs, capable of predicting full temperature fields in milliseconds compared to hours for CFD simulations, enables real-time thermal state estimation and proactive hotspot anticipation, allowing cooling systems to respond before critical thresholds are reached. This predictive capability is particularly valuable for heterogeneous computing environments where workload patterns are highly dynamic and unpredictable.

4.7.3 Predictive Maintenance and Hardware Design Optimization

Beyond real-time control and prediction, AI is transforming both system maintenance and thermal hardware design. Fig. 5c depicts a machine learning workflow for optimizing thermal management structures [42]. In this data-driven paradigm, ML surrogate models (artificial neural networks, random forests, or Gaussian processes) are trained on datasets generated from computational fluid dynamics (CFD) simulations or experimental measurements. For predictive maintenance, advanced time-series models such as long short-term memory (LSTM) networks and transformer architectures analyze historical sensor data, including temperature, pressure, flow rate, and vibration, to identify patterns preceding equipment degradation or failure. Compared to conventional threshold-based methods that trigger alarms only when measurements exceed predetermined limits, AI-based predictive maintenance can provide early warnings hours or days in advance, enabling proactive maintenance scheduling and reducing unplanned downtime. LSTM networks excel at capturing long-term dependencies, making them effective for predicting gradual degradation such as pump wear or heat exchanger fouling, while transformer-based models detect subtle anomalies across multiple sensor streams that precede catastrophic failures. In data center applications, these systems have demonstrated 60%–80% reduction in false alarms and 30%–50% extension of equipment service life through optimized maintenance schedules [94]. For hardware design optimization, machine learning surrogate models, including artificial neural networks, random forests, and Gaussian processes, are trained on datasets from CFD simulations or experiments, then rapidly explore large design spaces encompassing fin geometries, microchannel layouts, manifold architectures, and nanofluid compositions. This approach identifies optimal configurations that minimize thermal resistance while satisfying constraints such as pressure drop, weight, and manufacturability, reducing design optimization time from weeks to hours while achieving performance improvements of 15%–30% compared to conventional heuristic-based designs [42]. The integration of predictive maintenance with DRL-based control systems further enhances reliability by allowing the controller to avoid operating conditions that accelerate degradation, creating a holistic approach to system health management.

4.7.4 Case Studies and Real-World Implementations

The most prominent real-world example of AI-optimized cooling is Google's DeepMind AI applied to data centers, which reduced cooling energy consumption by 40% and achieved a PUE of 1.12 compared to the previous 1.22 baseline [95]. The system uses an ensemble of deep neural networks trained on thousands of sensor measurements to predict future PUE and thermal conditions, with a DRL agent then recommending optimal cooling actions to operators. Microsoft has similarly implemented AI-driven data center control, achieving 15% cooling energy reduction through predictive load balancing and dynamic chiller optimization. At the academic level, researchers have demonstrated DRL-based thermal control for CPU-GPU heterogeneous systems, achieving 28% energy savings while maintaining performance constraints. For edge computing applications, lightweight AI models implemented on microcontroller units have demonstrated effective thermal management for mobile devices, with inference times below 10 ms and memory footprints under 100 KB [96]. Future research directions include edge-AI for on-device thermal management, federated learning for privacy-preserving optimization across multiple systems, and explainable AI to provide interpretable control decisions that build operator trust. The convergence of these approaches promises to deliver cooling systems that are not only more energy-efficient but also more reliable, adaptive, and sustainable than current solutions.

4.8 Emerging Active Cooling Technologies

Beyond the established active cooling methods discussed in the preceding sections, several emerging technologies are gaining attention for niche applications where conventional approaches face inherent limitations. Two promising candidates, electrohydrodynamic (EHD) conduction pumps and piezoelectric fans, offer unique advantages that may complement or, in specific cases, replace conventional cooling solutions.

Electrohydrodynamic conduction pumps operate on the principle of ion drag, where an applied electric field generates a net body force on a dielectric fluid, inducing flow without any moving mechanical parts. This principle eliminates mechanical wear, vibration, and noise, making EHD pumps attractive for applications requiring silent operation and high reliability. Typical heat transfer coefficients range from 500 to 5000 $W/m^2 \cdot K$, with COP values between 2 and 15 depending on the applied voltage and fluid properties [97,98]. However, practical implementation faces significant challenges: high voltage requirements (typically 5–20 kV) raise safety and insulation concerns, while performance is highly sensitive to dielectric fluid purity, with contaminants significantly degrading efficiency [99]. Additionally, relatively low flow rates limit EHD systems to low-to-moderate heat flux applications (typically <50 W/cm^2). Despite these limitations, recent advances in electrode geometry optimization and high-dielectric-strength fluids have demonstrated improved performance, positioning EHD cooling as a viable solution for compact electronics, LED lighting, and aerospace applications where silent, vibration-free operation is paramount.

Piezoelectric fans represent another emerging active cooling technology that leverages the inverse piezoelectric effect to generate oscillating airflow. A piezoelectric actuator vibrates at its resonant frequency (typically 20–200 Hz) when subjected to an alternating electric field, creating a synthetic jet that enhances convective heat transfer by disrupting the thermal boundary layer. Experimental studies have demonstrated 20%–50% heat transfer enhancement over natural convection, with remarkably low power consumption of 0.1–2 W, making piezoelectric fans highly energy-efficient for localized cooling [100]. Their compact form factor and silent operation (noise levels below 25 dB) make them particularly suitable for portable electronics, smartphones, and ultrabooks where space and acoustic noise are critical constraints. The primary limitations include low cooling capacity (typically <10 W), directional airflow requiring precise positioning, and reduced effectiveness in confined spaces where airflow recirculation occurs. Recent innovations in multi-blade arrays and dual-frequency driving strategies have achieved cooling enhancements exceeding 60% compared to natural convection [101].

A comparative assessment of these emerging technologies against established cooling methods reveals clear trade-offs. Piezoelectric fans offer the highest COP (50–200) for low-power applications due to minimal power consumption, while EHD pumps achieve moderate COP (2–15), comparable to thermoelectric coolers but lower than conventional mechanical pumps (50–100). For scalability, piezoelectric fans are readily manufacturable using standard PCB assembly processes and integrated into existing infrastructure at low unit costs ($0.50–2.00), making them cost-competitive for consumer electronics [102]. EHD pumps face higher manufacturing complexity due to precise electrode spacing and high-voltage isolation, resulting in higher costs ($5–20) and limited production scale, though MEMS fabrication may reduce costs in the future. Regarding application suitability, piezoelectric fans are best suited for low-power, space-constrained devices such as smartphones and wearables, where silent operation and compactness are critical. EHD pumps, with higher heat transfer coefficients and no moving parts, are more appropriate for avionics, medical devices, and LED lighting systems requiring reliability and vibration-free operation. However, neither technology can match the heat dissipation capacity of microchannel liquid cooling or spray cooling, and both are likely to serve as complementary solutions rather than replacements for high-performance computing applications, with hybrid systems combining these emerging technologies with conventional cooling methods representing a promising future direction [103].

5 Future Directions and Conclusions

5.1 Future Research Directions

Despite significant advances in electronic cooling technologies over the past decade, several key challenges remain unaddressed. Future research efforts should focus on four interconnected areas: materials innovation, advanced cooling architectures, intelligent thermal management, and sustainable coolants. These directions collectively address the escalating thermal demands of next-generation electronics while balancing performance, cost, and environmental impact.

Materials Innovation: Future research must prioritize the development of high-thermal-conductivity, low-cost thermal interface materials to minimize contact resistance between heat sources and cooling devices. Stable nanofluids with extended operational lifetimes are needed to prevent particle agglomeration and sedimentation that degrade long-term performance [64]. Composite phase change materials with enhanced thermal conductivity, through incorporation of metal foams, expanded graphite, or carbon nanomaterials, require further optimization to balance energy storage density and heat transfer rate [65]. Additionally, environmentally friendly dielectric fluids with low GWP values must be discovered and characterized for immersion cooling applications, as current fluorocarbon-based fluids face increasing regulatory pressure due to their high GWP values [64]. Building on these material advances, the next frontier lies in integrating them into novel cooling architectures that maximize their potential.

Advanced Cooling Architectures: Next-generation cooling systems demand innovative architectures that push beyond conventional designs. Manifold microchannel designs require optimization for uniform flow distribution across large chip areas, preventing localized hotspots from flow maldistribution. Hybrid cooling systems that combine multiple mechanisms, such as PCMs with heat pipes, thermoelectric coolers, or vapor compression systems, leverage the strengths of each approach while mitigating individual limitations, offering synergistic performance gains [65,104]. Passive two-phase systems for data center cooling promise energy-efficient operation without pumping power, but require further development to ensure stability under varying thermal loads. Additive manufacturing enables complex, topology-optimized heat sink geometries, including triply periodic minimal surface structures and functionally graded lattices that cannot be fabricated using conventional methods. For portable and wearable electronics, flexible heat pipes and vapor chambers that maintain thermal performance under repeated bending and twisting must be developed, with flexible PCM films (0.4 mm thick) incorporating paraffin wax with graphene-based films having already demonstrated temperature reductions of up to 14.3°C and a 32.4% extension of thermal control time [65]. As hardware architectures become more complex, intelligent control systems become essential to manage their dynamic thermal behavior.

Intelligent Thermal Management: The integration of artificial intelligence and machine learning into thermal management systems represents a transformative paradigm for real-time optimization. Future systems should implement real-time temperature sensing and predictive control using machine learning algorithms that anticipate thermal loads before they occur [64]. Adaptive cooling systems that respond dynamically to transient thermal loads, varying pump speeds, fan speeds, or coolant flow distribution based on instantaneous heat generation, can significantly reduce energy consumption while maintaining device safety. Reinforcement learning-based controllers for multi-variable cooling optimization (balancing temperature, pressure drop, power consumption, and acoustic noise) offer superior performance compared to conventional PID controllers, particularly in systems with complex, nonlinear dynamics. Digital twins, real-time virtual replicas of physical cooling systems, enable what-if analysis and predictive optimization, allowing operators to test control strategies without risking hardware damage. Ultimately, these intelligent systems must be paired with sustainable coolants to ensure that the environmental footprint of cooling does not undermine its thermal benefits.

Sustainable Coolants and System-Level Sustainability: Environmental concerns are increasingly shaping the future of electronic cooling, driving the development of sustainable working fluids and energy-efficient system designs. Next-generation coolants including dielectric nanofluids, ionanofluids (nanoparticles dispersed in ionic liquids), and PCM slurries offer the potential to simultaneously improve heat transfer and thermal storage capacity while reducing environmental impact [64,95]. Low-GWP refrigerants must be identified and validated for two-phase cooling applications, as conventional refrigerants face phase-down under international agreements (e.g., the Kigali Amendment to the Montreal Protocol). Biodegradable bio-PCMs derived from renewable sources, such as fatty acids and vegetable oils, offer sustainable alternatives to petroleum-based paraffins, though their thermal conductivity and long-term cycling stability require further improvement. Beyond working fluids, system-level sustainability metrics, including Carbon Usage Effectiveness (CUE), Water Usage Effectiveness (WUE), and Life Cycle Assessment (LCA), must complement traditional PUE to holistically evaluate cooling system environmental impact across manufacturing, operation, and end-of-life stages. By integrating sustainable coolants with intelligent control and advanced architectures, next-generation cooling systems can achieve both unprecedented thermal performance and environmental responsibility.

Cross-Cutting Integration and Roadmap (2026–2035): While each research direction is pursued independently, their integration is essential for achieving transformative breakthroughs. The convergence of advanced materials (diamond substrates, enhanced PCMs), innovative architectures (embedded microchannels, hybrid systems), intelligent control (PINNs, DRL), and sustainable coolants (low-GWP fluids, CO_2 systems) will enable chip power densities exceeding 1000 W/cm^2 by 2030 and 3000 W/cm^2 by 2035. However, realizing this vision requires sustained interdisciplinary collaboration among thermal scientists, materials engineers, AI researchers, semiconductor designers, and policymakers, a collaborative effort that must begin immediately to meet the accelerating thermal demands of next-generation computing systems.

5.2 Conclusions

This review has systematically examined the landscape of electronic cooling technologies, spanning from conventional fin heat sinks to advanced AI-optimized intelligent cooling. Below, we summarize the key findings, identify persistent technology limitations and gaps, provide actionable recommendations for engineering practice, and pose critical open questions to guide future research.

Key Findings. It is evident that fin heat sinks and forced air cooling, while simple and low-cost, are fundamentally inadequate for modern high-power-density devices where heat fluxes routinely exceed 100 W/cm^2. To address transient thermal loads, phase change materials offer effective thermal buffering, achieving temperature reductions of 6°C–16°C; however, their inherently low thermal conductivity necessitates enhancement through fins, metal foams, or nanoparticles. Similarly, heat pipes and vapor chambers provide excellent passive thermal transport for portable and space-constrained electronics, yet they cannot handle the highest heat fluxes encountered in high-performance computing. Therefore, passive solutions alone are insufficient for next-generation applications. Moving beyond passive approaches, liquid-based active cooling has become the standard for high-performance computing, data centers, and power electronics. Technologies such as cold plates, microchannels, and immersion cooling achieve heat transfer coefficients of 1000–50,000 W/m^2·K, with microchannel heat sinks handling heat fluxes exceeding 500 W/cm^2. Two-phase systems offer higher heat transfer coefficients and critical heat fluxes than single-phase counterparts, albeit at a lower COP and with increased system complexity. At the extreme end of the performance spectrum, jet impingement and spray cooling achieve heat transfer coefficients of 30,000–250,000 W/m^2·K and critical heat fluxes of 400–1500 W/cm^2. However, these technologies require careful design to mitigate nozzle clogging and non-uniform cooling patterns, highlighting the trade-off between performance and practicality. In parallel with these advances, AI-optimized cooling has emerged as a transformative paradigm, with machine learning-enabled dynamic control achieving a 40% reduction in cooling energy consumption in data center deployments, i.e., a major breakthrough with industry-wide implications.

Technology Limitations and Gaps. Despite the remarkable progress documented in this review, several fundamental gaps persist across all technology classes. For air-based cooling, the thermodynamic limit imposed by air's low specific heat and thermal conductivity establishes a practical maximum heat flux of approximately 100 W/cm^2, beyond which liquid cooling becomes mandatory regardless of engineering optimization. For microchannel cooling, manufacturing tolerances and flow maldistribution remain unresolved challenges, with a 10% variation in hydraulic diameter producing up to 25% variation in flow distribution among parallel channels, creating localized hotspots that undermine thermal performance. For two-phase systems, flow instabilities, pressure drop oscillations, and critical heat flux dryout continue to limit reliable operation at peak performance, with conventional predictive models under- or over-estimating pressure drops by up to 30%. For immersion cooling, long-term coolant degradation, material compatibility issues (seal swelling, corrosion), and the absence of standardized performance benchmarks hinder large-scale commercial deployment. For AI-optimized cooling, the lack of robust, generalizable control policies that perform reliably across varying workloads and hardware configurations remains a significant barrier to widespread adoption. Furthermore, the environmental impact of fluorocarbon-based dielectric fluids, with their high global warming potential, presents an increasingly urgent sustainability challenge that current research has only begun to address.

Actionable Recommendations for Engineering Practice. Based on the comprehensive analysis presented in this review, we offer the following specific recommendations for engineers and system designers. For applications with heat fluxes below 100 W/cm^2, forced air cooling with optimized fin geometries and intelligent flow management remains the most cost-effective solution, particularly when temperature uniformity rather than absolute heat flux is the primary constraint. For heat fluxes between 100 and 500 W/cm^2, single-phase liquid cooling, either cold plates or microchannel heat sinks, represents the optimal balance of performance, reliability, and cost, with manifold microchannel designs offering superior temperature uniformity and reduced pumping power. For heat fluxes exceeding 500 W/cm^2, two-phase cooling systems become necessary, but engineers should prioritize designs with passive flow stabilization mechanisms (e.g., Tesla-type microchannels, inlet restrictors, or bypass channels) to mitigate dry-out instability. For data center deployments, single-phase immersion cooling offers the most straightforward path to PUE values below 1.10, while two-phase immersion should be reserved for applications where maximum heat flux density is the overriding priority. In all cases, AI-optimized control should be implemented as a supplementary layer to adapt cooling delivery to dynamic thermal loads, but engineers should ensure that baseline hardware is capable of meeting thermal requirements even in the absence of AI control, treating AI as an enhancement rather than a critical dependency. Finally, when selecting coolants, a systematic multi-criteria decision-making approach should balance thermal performance, electrical safety, environmental impact (GWP, biodegradability), and lifecycle cost, with low-GWP dielectric fluids and biodegradable bio-PCMs representing the most sustainable forward-looking choices.

Critical Open Questions for Future Research. Beyond the specific research directions outlined in Section 5.1, several fundamental scientific questions remain unanswered and should guide future investigation. Can two-phase immersion cooling be stabilized without active pressure control, enabling passive, maintenance-free operation at scale? What is the maximum critical heat flux achievable on engineered surfaces before flow instabilities fundamentally limit performance, and what surface architectures are optimal across different operating conditions? Can AI controllers be developed that generalize across heterogeneous hardware configurations and workload patterns without requiring extensive retraining, and how can explainable AI provide interpretable control decisions that build operator trust? What are the long-term degradation mechanisms of emerging dielectric fluids under realistic operating conditions, and how can accelerated aging tests be standardized to predict 10-year system reliability? Can additive manufacturing techniques achieve the precision and cost-effectiveness required for volume production of topology-optimized heat sinks, or will conventional manufacturing methods continue to dominate? How can thermal management be co-designed with chip architecture to achieve holistic optimization rather than treating cooling as an afterthought? Answering these questions will require sustained interdisciplinary collaboration among thermal scientists, materials engineers, AI researchers, and semiconductor designers, a collaborative effort that will ultimately determine the trajectory of electronic cooling technology over the next decade. In conclusion, no single technology can solve all thermal challenges. As electronic power densities continue to escalate, the future of electronic cooling lies in hybrid architectures, AI-optimized designs, advanced materials, and sustainable solutions. The continued miniaturization and integration of electronic devices demand sustained innovation in thermal management to ensure performance, reliability, and longevity.

Acknowledgement: Not applicable.

Funding Statement: This study was supported by the National Natural Science Foundation of China (grant no. 52276079).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Wei Tong, Yugang Zhao and Chuyi Peng; methodology, Wei Tong; software, Wei Tong; validation, Wei Tong and Yugang Zhao; formal analysis, Wei Tong and Yugang Zhao; investigation, Wei Tong and Yugang Zhao; resources, Wei Tong and Yugang Zhao; data curation, Wei Tong and Yugang Zhao; writing—original draft preparation, Wei Tong, Yugang Zhao and Chuyi Peng; writing—review and editing, Wei Tong, Yugang Zhao and Chuyi Peng; visualization, Wei Tong, and Yugang Zhao; supervision, Yugang Zhao; project administration, Yugang Zhao; funding acquisition, Yugang Zhao. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Yugang Zhao, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Barnes JR. Thermal considerations. In: Robust electronic design reference book. New York, NY, USA: Springer; 2004. p. 868–91. doi:10.1007/1-4020-7830-7_39.
2. Lv YG, Chu WX, Wang QW. Thermal management systems for electronics using in deep downhole environment: a review. Int Commun Heat Mass Transf. 2022;139(5):106450. doi:10.1016/j.icheatmasstransfer.2022.106450.
3. Pambudi NA, Sarifudin A, Firdaus RA, Ulfa DK, Gandidi IM, Romadhon R. The immersion cooling technology: current and future development in energy saving. Alex Eng J. 2022;61(12):9509–27. doi:10.1016/j.aej.2022.02.059.
4. Du L, Hu W. An overview of heat transfer enhancement methods in microchannel heat sinks. Chem Eng Sci. 2023;280(2):119081. doi:10.1016/j.ces.2023.119081.
5. Riofrío MC, Caney N, Gruss JA. State of the art of efficient pumped two-phase flow cooling technologies. Appl Therm Eng. 2016;104:333–43. doi:10.1016/j.applthermaleng.2016.05.061.
6. Sohel Murshed SM, de Castro CA N. A critical review of traditional and emerging techniques and fluids for electronics cooling. Renew Sustain Energy Rev. 2017;78(1):821–33. doi:10.1016/j.rser.2017.04.112.
7. Kheirabadi AC, Groulx D. Cooling of server electronics: a design review of existing technology. Appl Therm Eng. 2016;105:622–38. doi:10.1016/j.applthermaleng.2016.03.056.
8. Moore AL, Shi L. Emerging challenges and materials for thermal management of electronics. Mater Today. 2014;17(4):163–74. doi:10.1016/j.mattod.2014.04.003.
9. Kandlikar SG, Grande WJ. Evolution of microchannel flow passages—thermohydraulic performance and fabrication technology. Heat Transf Eng. 2003;24(1):3–17. doi:10.1080/01457630304040.
10. van Erp R, Soleimanzadeh R, Nela L, Kampitsis G, Matioli E. Co-designing electronics with microfluidics for more sustainable cooling. Nature. 2020;585(7824):211–6. doi:10.1038/s41586-020-2666-1.
11. Huo S, Wang X, Sun B. Two-phase flow boiling for ultra-high heat flux SiC chip cooling. Int J Heat Mass Transf. 2026;255:127885. doi:10.1016/j.ijheatmasstransfer.2025.127885.
12. Liang G, Mudawar I. Review of spray cooling-Part 1: single-phase and nucleate boiling regimes, and critical heat flux. Int J Heat Mass Transf. 2017;115:1174–205. doi:10.1016/j.ijheatmasstransfer.2017.06.029.
13. Balakrishnan P. Optimization of nanofluid impinging jet systems for advanced electronic cooling: a critical review. Int J Heat Fluid Flow. 2026;117(17):110149. doi:10.1016/j.ijheatfluidflow.2025.110149.
14. Lu Y, Ling Y, Zhuang Y, Li C, Ouyang H, Han X. Flow boiling heat transfer characteristics and delayed dry-out ability of non-azeotropic mixtures R245fa/R134a in microchannels. J Therm Sci. 2024;33(3):1174–88. doi:10.1007/s11630-024-1959-3.

15. Shrigondekar H, Lin YC, Wang CC. Investigations on performance of single-phase immersion cooling system. Int J Heat Mass Transf. 2023;206(22):123961. doi:10.1016/j.ijheatmasstransfer.2023.123961.
16. Prajapati YK, Bhandari P. Flow boiling instabilities in microchannels and their promising solutions—a review. Exp Therm Fluid Sci. 2017;88(2):576–93. doi:10.1016/j.expthermflusci.2017.07.014.
17. Tao Q, Fu S, Yuan Y, Liu X, He D, Chen Y. Experimental study on enhanced nanofluid heat transfer limit and wetting predicted model in the microgroove under electric field. Int J Therm Sci. 2024;197:108829. doi:10.1016/j.ijthermalsci.2023.108829.
18. Yu Q, Song Y, Cui C, Cao F. Effects of nanoparticle aggregation on the thermal conductivity of nanofluids: a comprehensive review based on multiscale methods. Renew Sustain Energ Rev. 2026;226(4):116306. doi:10.1016/j.rser.2025.116306.
19. GaneshKumar P, Praveen Kumar G, Sivalingam V, Divya S, Oh TH. Revolutionizing microelectronics cooling: thermal management with nano-enhanced PCMs, hybrid cooling, conductive foams, and porous structures. Renew Sustain Energy Rev. 2025;222(2):115954. doi:10.1016/j.rser.2025.115954.
20. Rahman IU, Nardini S, Buonomo B, Manca O, Khan H, Siviero B. Thermal interface materials: a promising solution for passive heat dissipation in electronic appliances. Therm Sci Eng Prog. 2025;62(2):103673. doi:10.1016/j.tsep.2025.103673.
21. Yin M, Qiu M, Yuan Y, Zhu C, Zou W, Yang B. Advanced polymer-based thermal interface materials: a review on matrix optimization and filler engineering for superior performance. Eur Polym J. 2026;242(8):114468. doi:10.1016/j.eurpolymj.2025.114468.
22. Kaur N. Thermal Interface materials for efficient heat management in electronics. Appl Therm Eng. 2026;287(2):129415. doi:10.1016/j.applthermaleng.2025.129415.
23. Bahiraei M, Heshmatian S. Electronics cooling with nanofluids: a critical review. Energy Convers Manag. 2018;172(10):438–56. doi:10.1016/j.enconman.2018.07.047.
24. Hansson J, Nilsson TMJ, Ye L, Liu J. Novel nanostructured thermal interface materials: a review. Int Mater Rev. 2018;63(1):22–45. doi:10.1080/09506608.2017.1301014.
25. Kumar Swamy MC, Satyanarayan. A review of the performance and characterization of conventional and promising thermal interface materials for electronic package applications. J Electron Mater. 2019;48(12):7623–34. doi:10.1007/s11664-019-07623-7.
26. Pop E. Energy dissipation and transport in nanoscale devices. Nano Res. 2010;3(3):147–69. doi:10.1007/s12274-010-1019-z.
27. Banerjee K, Souri SJ, Kapur P, Saraswat KC. 3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration. Proc IEEE. 2001;89(5):602–33. doi:10.1109/5.929647.
28. Matioli E, Zhu H, Perera N, Nikoo MS, Jafari A, van Erp R. Switching losses in power devices: from dynamic on resistance to output capacitance hysteresis. In: 2023 25th European Conference on Power Electronics and Applications (EPE'23 ECCE Europe); 2023 Sep 4–8; Aalborg, Denmark. p. 1–7.
29. Roy K, Mukhopadhyay S, Mahmoodi-Meimand H. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. Proc IEEE. 2003;91(2):305–27. doi:10.1109/JPROC.2002.808156.
30. Zahidul Islam M, Fu Y, Deb H, Khalid Hasan M, Dong Y, Shi S. Polymer-based low dielectric constant and loss materials for high-speed communication network: dielectric constants and challenges. Eur Polym J. 2023;200(2):112543. doi:10.1016/j.eurpolymj.2023.112543.
31. Arora Y, Boora V, Dhiman R, Chandel R. Through silicon vias for 3D integration—a mini review. In: Agrawal Y, Mummaneni K, Sathyakam PU, editors. Interconnect technologies for integrated circuits and flexible electronics. Singapore: Springer Nature; 2024. p. 81–98. doi:10.1007/978-981-99-4476-7_6.
32. Zhao J, Dai Y, Tao W, Wang Y. A review of heat removal technologies across heat flux frontiers: from high to ultra high level. Int Commun Heat Mass Transf. 2026;172(21):110516. doi:10.1016/j.icheatmasstransfer.2026.110516.
33. Pandey J, Ansari MZ, Husain A. Optimisation of porous slab parameters for jet-impingement microchannel heat sink performance enhancement. Int J Numer Methods Heat Fluid Flow. 2022;32(8):2659–81. doi:10.1108/hff-08-2021-0523.

34. Alshamaileh D, Qasem M. Thermal performance and COP assessment of thermoelectric coolers using gallium and paraffin waxes as phase change materials under different pulsation durations: a comparative study. Discover Materials. 2026;6(1):175. doi:10.1007/s43939-026-00658-8.
35. Huang Y, Ge J, Chen Y, Zhang C. Natural and forced convection heat transfer characteristics of single-phase immersion cooling systems for data centers. Int J Heat Mass Transf. 2023;207:124023. doi:10.1016/j.ijheatmasstransfer.2023.124023.
36. Khosravi A, Sandoval OR, Taslimi MS, Sahrakorpi T, Amorim G, Garcia Pabon JJ. Review of energy efficiency and technological advancements in data center power systems. Energy Build. 2024;323:114834. doi:10.1016/j.enbuild.2024.114834.
37. Guo L, Camm J, Hirokawa T, Li H, Huang L. A detailed review on the role of critical heat flux in micro-channel dryout phenomena and strategies for heat transfer enhancement. Int J Heat Mass Transf. 2025;241(1):126740. doi:10.1016/j.ijheatmasstransfer.2025.126740.
38. Rodgers P, Eveloy V, Pecht MG. Limits of air-cooling: status and challenges. In: Semiconductor Thermal Measurement and Management IEEE Twenty First Annual IEEE Symposium; 2005 Mar 15–17; San Jose, CA, USA; 2005. p. 116–24. doi:10.1109/STHERM.2005.1412167.
39. Bünnagel C, Monir S, Sharp A, Anuchin A, Durieux O, Uria I, et al. Forced air cooled heat sink with uniformly distributed temperature of power electronic modules. Appl Therm Eng. 2021;199(2):117560. doi:10.1016/j.applthermaleng.2021.117560.
40. Wang L, Zuo H, Zhang B, Jia G. Effects of the cold plate with airfoil fins on the cooling performance enhancement of the prismatic $LiFePO_4$ battery pack. Energy. 2024;296(4):131210. doi:10.1016/j.energy.2024.131210.
41. Sharma CS, Tiwari MK, Zimmermann S, Brunschwiler T, Schlottig G, Michel B, et al. Energy efficient hotspot-targeted embedded liquid cooling of electronics. Appl Energy. 2015;138:414–22. doi:10.1016/j.apenergy.2014.10.068.
42. Miao Y, Zhang T, Xu X, Yang S, Chen H, Huang H. A comprehensive review on advanced microchannel cooling technologies: two-phase thermal management and machine learning applications. Appl Therm Eng. 2025;279:127697. doi:10.1016/j.applthermaleng.2025.127697.
43. Patel A, Yogi K, Sahu G, Wei T. Multi-chip Jet impingement cooling for heat dissipation in 2.5D integrated system with 1 kW+ thermal design power. Int J Heat Mass Transf. 2025;244:126978. doi:10.1016/j.ijheatmasstransfer.2025.126978.
44. Chen J, Liu X, Zhong F, Miao Q, Jia W, Zheng M, et al. Experimental study of dielectric liquid spray cooling on multi-scale structured surfaces inspired by leaf veins. Int J Heat Fluid Flow. 2024;109:109554. doi:10.1016/j.ijheatfluidflow.2024.109554.
45. Luo Q, Wang C, Wu C. Study on heat transfer performance of immersion system based on SiC/white mineral oil composite nanofluids. Int J Therm Sci. 2023;187:108203. doi:10.1016/j.ijthermalsci.2023.108203.
46. Wu X, Yang J, Liu Y, Zhuang Y, Luo S, Yan Y, et al. Investigations on heat dissipation performance and overall characteristics of two-phase liquid immersion cooling systems for data center. Int J Heat Mass Transf. 2025;239(11):126575. doi:10.1016/j.ijheatmasstransfer.2024.126575.
47. Chen X, Hu C, Liang J, Li J, Lu L, Wu Y, et al. Efficient two-phase liquid cooling and optimization technology on the chip side based on high-power data centers: summary and prospects. Renew Sustain Energy Rev. 2025;223(10):115994. doi:10.1016/j.rser.2025.115994.
48. Wang ZX, Tao WQ. Numerical study on an air-cooling and liquid-cooling combined server for high-temperature data center application. Appl Therm Eng. 2025;280:128111. doi:10.1016/j.applthermaleng.2025.128111.
49. Tian T, Yang X, Li J. To boost waste heat harvesting and power generation through a portable heat pipe battery during high efficient electronics cooling. Appl Energy. 2025;377:124397. doi:10.1016/j.apenergy.2024.124397.
50. Tang H, Tang Y, Wan Z, Li J, Yuan W, Lu L, et al. Review of applications and developments of ultra-thin micro heat pipes for electronic cooling. Appl Energy. 2018;223:383–400. doi:10.1016/j.apenergy.2018.04.072.
51. Zhang Y, Yang Y, Chen X, Liu B, Yang X, Gao Y, et al. High-efficiency heat sink design via SLM-processed diamond TPMS CuCrZr alloy: optimizing thermal-fluid performance through volume fraction control. Int J Therm Sci. 2026;223(8):110657. doi:10.1016/j.ijthermalsci.2026.110657.

52. Shete RK, Laskar N, Das MK. Geometry-driven high-density heat sink module for enhanced two-phase immersion cooling. Therm Sci Eng Prog. 2025;68:104237. doi:10.1016/j.tsep.2025.104237.
53. Fathi M, Heyhat MM, Zabetian Targhi M, Bigham S. Porous-fin microchannel heat sinks for future microelectronics cooling. Int J Heat Mass Transf. 2023;202:123662. doi:10.1016/j.ijheatmasstransfer.2022.123662.
54. Elkholy A, Narvan M, Thompson S, Durfee J, Kempers R. The effect of layer thickness on the geometry and capillary performance of strut-based heat pipe wicks manufactured by laser powder bed fusion. Int J Thermofluids. 2024;21(2):100543. doi:10.1016/j.ijft.2023.100543.
55. Zhou D, Guo L, Lan X, Zhang J, Wang X, Du W, et al. Fabrication and enhanced thermal performance of a self-rewetting wick of silicon-based loop heat pipe. Int Commun Heat Mass Transf. 2025;160:108378. doi:10.1016/j.icheatmasstransfer.2024.108378.
56. Tan Z, Jin P, Chen S, Xie G, Jian Y, Zheng J, et al. Vapor chambers with hierarchical wick structures: thermal performance and liquid film boiling visualization. Appl Therm Eng. 2025;280:128165. doi:10.1016/j.applthermaleng.2025.128165.
57. Gao J, Zhang J, Zong K, Meng J, Chen J, Fan W, et al. High-precision additive manufacturing enabled ultra-thin flexible vapor chamber for adaptive and efficient electronic cooling. Int J Heat Mass Transf. 2026;256:128044. doi:10.1016/j.ijheatmasstransfer.2025.128044.
58. Huang PJ, Chiu YC, Sha YH, Wang CC. Performance improvement of vapor chamber with a novel step gap. Appl Therm Eng. 2026;287(4):129471. doi:10.1016/j.applthermaleng.2025.129471.
59. Yin S, Zhou B, Shui Q, Zhao W, Tang Y, Ji W, et al. Functional support structure-based high thermal performance L-shaped ultra-thin vapor chamber design and evaluation. Appl Therm Eng. 2025;269:126013. doi:10.1016/j.applthermaleng.2025.126013.
60. Chai S, Lu L, Yi W, Zhong Y, Yang R, Li L, et al. Design, optimization, and fabrication of a leaf-inspired roll-bond large-area vapor chamber for thermal performance enhancement. Energy Convers Manag. 2026;353:121183. doi:10.1016/j.enconman.2026.121183.
61. Sivashankar M, Selvam C. Experimental investigation on the thermal performance of low-concentrated photovoltaic module using various pin-fin configurations of heat sink with phase change materials. J Energy Storage. 2022;55:105575. doi:10.1016/j.est.2022.105575.
62. Xiong J, Sun J, Chen Y, Nie Z. Study on the flow and heat transfer characteristics inside high-porosity open-cell copper foams: experimental and numerical explorations. Int J Heat Fluid Flow. 2024;109:109499. doi:10.1016/j.ijheatfluidflow.2024.109499.
63. Kumar A, Kothari R, Sahu SK, Kundalwal SI. Thermal performance of heat sink using nano-enhanced phase change material (NePCM) for cooling of electronic components. Microelectron Reliab. 2021;121(16):114144. doi:10.1016/j.microrel.2021.114144.
64. Cabaleiro D, Agresti F, Fedele L, Barison S, Hermida-Merino C, Losada-Barreiro S, et al. Review on phase change material emulsions for advanced thermal management: design, characterization and thermal performance. Renew Sustain Energy Rev. 2022;159:112238. doi:10.1016/j.rser.2022.112238.
65. Bianco V, De Rosa M, Vafai K. Phase-change materials for thermal management of electronic devices. Appl Therm Eng. 2022;214(1):118839. doi:10.1016/j.applthermaleng.2022.118839.
66. Chainer TJ, Schultz MD, Parida PR, Gaynes MA. Improving data center energy efficiency with advanced thermal management. IEEE Trans Compon Packag Manuf Technol. 2017;7(8):1228–39. doi:10.1109/TCPMT.2017.2661700.
67. Yi F, E. J, Zhang B, Zuo H, Wei K, Chen J, et al. Effects analysis on heat dissipation characteristics of lithium-ion battery thermal management system under the synergism of phase change material and liquid cooling method. Renew Energy. 2022;181(10):472–89. doi:10.1016/j.renene.2021.09.073.
68. Yang R, Li D, Deng Z, Zhang C, Chen Y. Two-phase cooling behaviors of Tesla-type microchannel using multiphysics field coupling. Int Commun Heat Mass Transf. 2026;172(4):110227. doi:10.1016/j.icheatmasstransfer.2025.110227.
69. Luo P, Tong W, Wei L, Zhang J, Zhao Y, Zhang H, et al. Enhanced electronic cooling with guest NCG microbubbles. Appl Therm Eng. 2025;272:126443. doi:10.1016/j.applthermaleng.2025.126443.

70. Sarangi S, Bodla KK, Garimella SV, Murthy JY. Manifold microchannel heat sink design using optimization under uncertainty. Int J Heat Mass Transf. 2014;69(17):92–105. doi:10.1016/j.ijheatmasstransfer.2013.09.067.
71. Pandey VK, Negi VPS, Ranganayakulu C. An in-depth comparison of straight and wavy microchannel heat exchangers. J Therm Sci Eng Appl. 2024;16(5):051007. doi:10.1115/1.4064985.
72. Zhou F, Gu W, Ma G. Microchannel heat sinks for cold plate liquid cooling in data centers: advances, evaluations and prospects. Renew Sustain Energy Rev. 2026;230:116729. doi:10.1016/j.rser.2026.116729.
73. Shah AK, Li H, Xin Z, Wu Z. A review on additive manufacturing of topology optimized microchannel heat sink. Heat Transf Eng. 2026;47(6):441–67. doi:10.1080/01457632.2025.2459981.
74. Everts M, Mahomed F. The effect of high values of relative surface roughness on heat transfer and pressure drop characteristics in the laminar, transitional, quasi-turbulent and turbulent flow regimes. Heat Mass Transf. 2024;60(7):1173–93. doi:10.1007/s00231-024-03472-1.
75. Alnaimat F, Mathew B. Flow distribution in microchannel devices with U-shaped manifolds. Int J Thermofluids. 2023;19:100391. doi:10.1016/j.ijft.2023.100391.
76. Yu ZQ, Li MT, Cao BY. A comprehensive review on microchannel heat sinks for electronics cooling. Int J Extreme Manuf. 2024;6(2):022005. doi:10.1088/2631-7990/ad12d4.
77. Wu Z, Xiao W, He H, Wang W, Song B. Jet-enhanced manifold microchannels for cooling electronics up to a heat flux of 3, 000?W?cm−2. Nat Electron. 2025;8(9):810–7. doi:10.1038/s41928-025-01449-4.
78. Obeng Y, Mansfield E, Davydov A, Barnes B, Vladar A. International Roadmap for Devices And Systems™ 2024 Metrology, International Roadmap for Devices and Systems (IRDS™) 2024 Edition [Internet]. [cited 2026 Jan 1]. Available from: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=959302.
79. Feng S, Yang C, Lai C, Tao D, Xiang Y, Yan P, et al. Improvement of flow boiling dry-out instability for parallel micro-channels using periodic two-phase micro-jets. Appl Therm Eng. 2026;285:129261. doi:10.1016/j.applthermaleng.2025.129261.
80. Ma J, Zhang D, Liu Z, Zhang W, Li Y, Gong L. Study on the influence of biomimetic leaf veins microchannel structure optimization on flow boiling heat transfer in high heat flux devices. Int Commun Heat Mass Transf. 2026;171:110121. doi:10.1016/j.icheatmasstransfer.2025.110121.
81. Mai Z, Zhang Z, Zhong K, Wu T, Wang C. Synergistically enhanced stability and thermal properties of oil-based nanofluids through multi-scale surface grafting of HDTMS on Al_2O_3 nanoparticles. Surf Interfaces. 2025;78(7):108114. doi:10.1016/j.surfin.2025.108114.
82. Liu J, Chen H, Yang M, Huang S, Wang K. Comparative study of natural ester oil and mineral oil on the applicability of the immersion cooling for a battery module. Renew Energy. 2024;224:120187. doi:10.1016/j.renene.2024.120187.
83. Lua KB, Wang CC. Single-phase immersion cooling for data centers and HPC: recent advances and system-level challenges. Renew Sustain Energy Rev. 2026;232(2):116786. doi:10.1016/j.rser.2026.116786.
84. Yang C, Zhu X, Zhang G, Pan G, Wang X. A review of the immersion liquid cooling technology for high-performance data centers. Int J Heat Fluid Flow. 2026;119(2):110282. doi:10.1016/j.ijheatfluidflow.2026.110282.
85. Yang CM, Muneeshwaran M, Hu Y, Pottker G, Yana Motta SF. Pool boiling heat transfer evaluation of next-generation dielectric fluid: opteon™ 2P50. Int J Thermofluids. 2025;29(1):101379. doi:10.1016/j.ijft.2025.101379.
86. Qenawy M, Chen Y, Zhu Y, Wang J, Tian J, Chen B. Improving fluid flow and heat transfer of cryogen spray cooling using sweeping cold air jet. Phys Fluids. 2024;36(4):043302. doi:10.1063/5.0201223.
87. Chen J, Liu X, Miao Q, Zheng M, Zhong F, Li Y, et al. Heat transfer characteristics of dielectric fluid spray cooling targeting high heat flux and high evaporation efficiency. Appl Therm Eng. 2024;257(3):124403. doi:10.1016/j.applthermaleng.2024.124403.
88. Shao Y, Wang J, Wang H, Zhou X, Xie Z, Zhu X, et al. Characterization of dynamics and heat transfer in electro-spray at elevated flow rates. Int J Heat Mass Transf. 2024;226(9):125473. doi:10.1016/j.ijheatmasstransfer.2024.125473.
89. Gong Y, Wu Y, Zhan Z, Zhao P, Liu L, Wu C, et al. LOTUS: learning-based online thermal and latency variation management for two-stage detectors on edge devices. In: Proceedings of the 61st ACM/IEEE Design Automation Conference; 2024 Jun 23–27. p. 1–6. doi:10.1145/3649329.3657310.

90. Coenen D, Oprins H. PINDAS: physics-informed decoupled spatiotemporal artificial neural network for dynamic thermal simulation. IEEE Trans Comput Aided Des Integr Circuits Syst. 2026;45(2):998–1006. doi:10.1109/TCAD.2025.3591411.
91. Kahil H, Sharma S, Välisuo P, Elmusrati M. Reinforcement learning for data center energy efficiency optimization: a systematic literature review and research roadmap. Appl Energy. 2025;389:125734.
92. Wang L, Fang K, Zheng K, Sun J, Wang P, Liu Z. Hybrid data-driven and physics-informed neural network modeling for regional spatiotemporal wave parameter prediction. Ocean Eng. 2026;358(2):125898. doi:10.1016/j.oceaneng.2026.125898.
93. Al Kez D, Foley AM, Hasan Wong FWBM, Dolfi A, Srinivasan G. AI-driven cooling technologies for high-performance data centres: state-of-the-art review and future directions. Sustain Energy Technol Assess. 2025;82(10):104511. doi:10.1016/j.seta.2025.104511.
94. Ma S, Liu Y, Liu Y, Wang J, Fang Q, Huang Y. Artificial intelligence-enabled predictive energy saving planning of liquid cooling system for data centers. Adv Eng Inform. 2025;65:103283. doi:10.1016/j.aei.2025.103283.
95. Chakraborty S, Shukla D, Kumar Panigrahi P. A review on coolant selection for thermal management of electronics and implementation of multiple-criteria decision-making approach. Appl Therm Eng. 2024;245:122807. doi:10.1016/j.applthermaleng.2024.122807.
96. Fridous VM, Agarwal A, Bhaskar KB, Nandagopal V, Sivakamasundari N. Optimization and benchmarking of lightweight neural networks for efficient embedded AI deployment. Engineering Reports. 2026;8(5):e70814. doi:10.1002/eng2.70814.
97. Vázquez PA, Talmor M, Seyed-Yagoobi J, Traoré P, Yazdani M. In-depth description of electrohydrodynamic conduction pumping of dielectric liquids: physical model and regime analysis. Phys Fluids. 2019;31(11):113601. doi:10.1063/1.5121164.
98. Luo Y, Schouten M, Shea H. Increasing the power density of electrohydrodynamic pumps by mapping the fluid landscape. Sci Adv. 2026;12(15):eaeb2623. doi:10.1126/sciadv.aeb2623.
99. Gao XL, Zhou FY, Wang FQ, Wu J, Luo K, Yi HL, et al. Flow control and heat transfer enhancement of modular electrohydrodynamic pump. Phys Fluids. 2025;37(9):093622. doi:10.1063/5.0289230.
100. Qiu YL, Wu CJ, Chen WF. Local heat transfer enhancement induced by a piezoelectric fan in a channel with axial flow. J Zhejiang Univ SCIENCE A. 2020;21(12):1008–22. doi:10.1631/jzus.A2000057.
101. Han C, Ma X, Xu J. Innovative configurations for heat sink integrated with piezoelectric fans. Int J Therm Sci. 2025;207:109383. doi:10.1016/j.ijthermalsci.2024.109383.
102. Peng B, Yi Z, Fan Y, Li X, Wei X, Wang C, et al. Piezoelectric active air-cooling devices. Chip. 2026;5(1):100163. doi:10.1016/j.chip.2025.100163.
103. Nandan R, Arumuru V, Das MK. Prospects of hybrid cooling technologies for thermal management of electronic gadgets: a review. Int J Heat Mass Transf. 2026;255:127766. doi:10.1016/j.ijheatmasstransfer.2025.127766.
104. Liang K, Li Z, Chen M, Jiang H. Comparisons between heat pipe, thermoelectric system, and vapour compression refrigeration system for electronics cooling. Appl Therm Eng. 2019;146(4):260–7. doi:10.1016/j.applthermaleng.2018.09.120.