

Review

Immersion Cooling Technologies in High-Density Computing: Review and Future Perspectives

Ghassan Adnan Abd, Ahmed J. Shkarah, Mustafa M. Mansour *

Department of Mechanical Engineering, College of Engineering, University of Thi-Qar, Thi-Qar 64001, Iraq; E-Mails: ghassanadnan77@utq.edu.iq; shkarah@utq.edu.iq; mustafa.muhammedali@utq.edu.iq

* **Correspondence:** Mustafa M. Mansour; E-Mail: mustafa.muhammedali@utq.edu.iq

Academic Editor: Orlando Corigliano*Recent Prog Sci Eng*

2026, volume 2, issue 4

doi:10.21926/rpse.2604024

Received: July 20, 2026**Accepted:** October 12, 2026**Published:** October 26, 2026

Abstract

Air cooling is being pushed to its limits in data centres, with racks now pushing tens of kilowatts and localised areas of the accelerator reaching at 100 W cm^{-2} . The architecture most frequently suggested to replace it is immersion cooling, where electronics are in direct contact with a dielectric liquid. Still, the literature backing this replacement is very disjointed—results are reported at varying spatial scales, reported in different system boundaries, and with incomplete property and hydraulic data, making direct comparison risky. This review does not aim to be yet another descriptive survey, but to tackle this fragmentation. 90 studies published from 2015 to 2026 were identified via a documented search protocol, screened for explicit criteria and analysed within five levels of chip, package, server, tank, and facility scales. The comparative performance of single-phase (SPIC) and two-phase (TPIC) immersion cooling is systematically compared using twelve engineering criteria and the quantitative data extracted are integrated into thirteen comparison tables between coolant properties, heat-transfer coefficients, thermal resistances, pressure drops, enhancement penalties, energy metrics with explicit boundary tags, cost and environmental indicators, and long-duration test evidence. Three findings emerge. First, the reported performance envelopes of SPIC and TPIC overlap significantly more than is suggested by their mechanisms: Forced SPIC has achieved heat-transfer coefficients of $0.4\text{--}2 \text{ kW m}^{-2} \text{ K}^{-1}$, and jet or locally actuated SPIC has achieved



© 2026 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

values between 2 and 15.3 kW m⁻² K⁻¹, which spans much of the range of heat-transfer coefficient for plain-surface nucleate boiling from 1.5 to 12 kW m⁻² K⁻¹; thus, the choice of architecture depends less on the heat-transfer coefficient achieved, and more on the critical heat flux margin, vapour management and complexity of operation. Second, apparent contradictions in the literature can be traced to five identifiable sources: declared system boundaries were different, unreported coolant properties, bulk flow rate as a substitute for local availability of coolant, mismatched scales, and steady uniform heating as a substitute for realistic workloads, and once these issues are controlled, the disagreements are largely eliminated. Third, the evidence base is very unbalanced: what is known about thermal performance—measured from the chip to the server—is well developed, but multi-year cooling ageing, material compatibility, and facility-scale demonstration and standardised test protocols are relatively limited or non-existent—this is the current limiting constraint for deployment, not thermal performance. On this basis, we suggest a prioritised research agenda and minimum reporting requirements for future immersion-cooling studies.

Keywords

Immersion cooling; two-phase immersion cooling; high-density computing; AI data centres; dielectric coolants; critical heat flux; thermal-hydraulic optimization; power usage effectiveness; long-term reliability

1. Introduction

1.1 Thermal Drivers in High-Density and AI Computing

High-density computing focuses electrical power at chip, server, rack, and facility levels. The major thermodynamic factors limiting IT are local heat flux and its spatial variation, rather than the average facility power [1, 2].

Air cooling remains mature, and higher density increases the required thermal conductivity, volumetric heat capacity, and heat-transfer area, increasing electrical and physical infrastructure needs [3, 4].

A typical current enterprise processor is 0.5-1.5 W cm⁻² on a package-average basis, while an AI/HPC accelerator is 3-10 W cm⁻² with higher local hotspots, and AI racks are typically 40-130 kW. With air-side h limited to about 250 W m⁻² K⁻¹, these loads increasingly require liquid cooling [5-7].

1.2 Immersion Cooling Within the Liquid-Cooling Landscape

Indirect liquid cooling is done with cold plates or heat exchangers without touching the electronics. Immersion cooling involves submerging the electronics in an electrically insulating liquid, thus eliminating the air-side resistance and reducing the heat transfer path [4, 5, 8].

SPIC uses a liquid coolant and depends upon natural or forced convection, while TPIC relies upon boiling and condensation where latent heat plays a role in heat transfer [1, 9]. Figure 1 is a comparison of both architectures with air and direct-to-chip cooling.

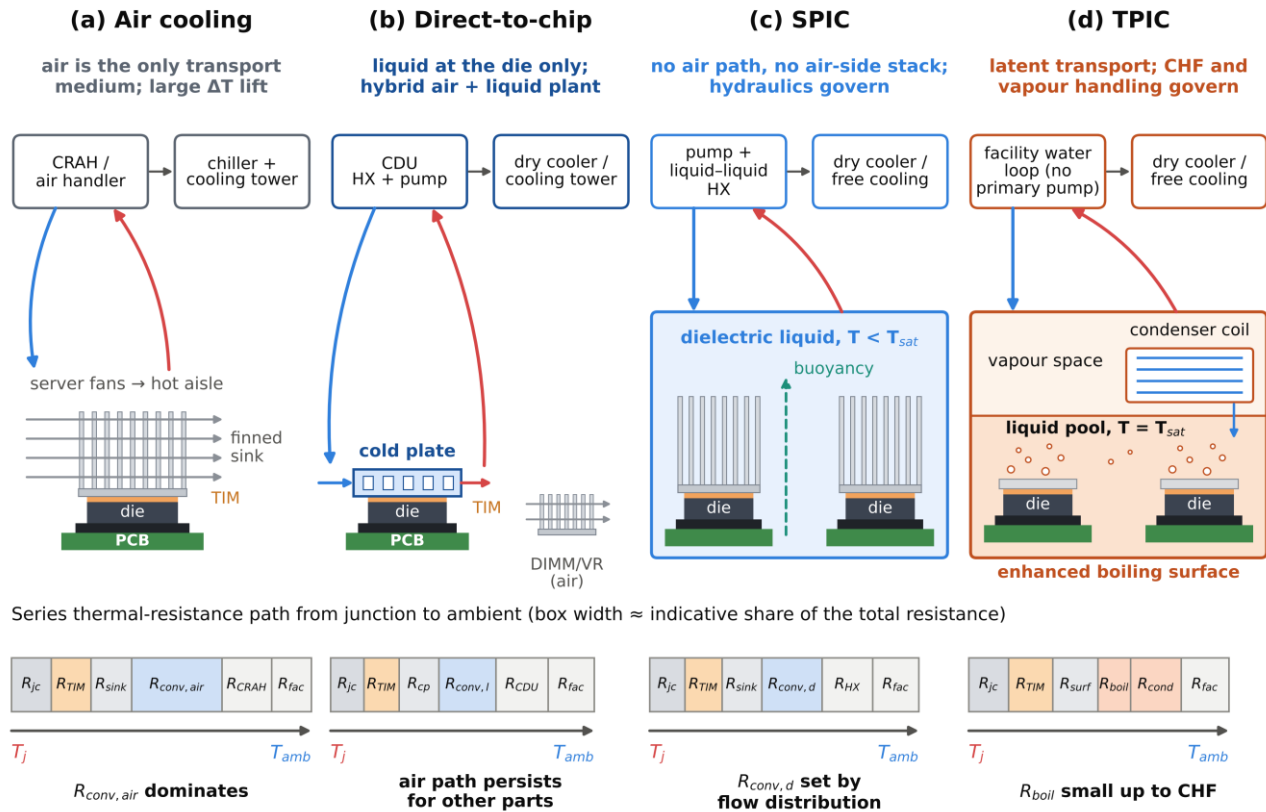


Figure 1 High-density cooling architectures and series thermal-resistance paths, (a) air cooling, (b) direct-to-chip cold-plate cooling, (c) single-phase immersion cooling (SPIC) and (d) two-phase immersion cooling (TPIC).

The upper row depicts the heat-transfer path from the die to the facility heat-rejection equipment; the lower row breaks down the same path into series resistances, with the width of the boxes representing the approximate proportion of the total resistance of the junction-to-ambient path. Immersion architectures eliminate air-side convective resistance and the associated air-handling infrastructure, but shift the performance-limiting role to coolant distribution in SPIC and the boiling stability, vapour transport and condensation in TPIC.

Immersion removes the dominant air-side resistance, while the limiting step becomes coolant distribution and local convection for SPIC, and nucleation, vapour removal, and condensation for TPIC. Various resistance chains are why cross-architecture comparisons need to be based on the same criteria and declared scales.

1.3 What Previous Reviews Have Established, and What They Have Not

Previous reviews have defined the general role of immersion cooling, but most are descriptive. Typical weaknesses include a lack of a reproducible search protocol, a lack of a criterion-based comparison of SPIC/TPIC, lack of scale-resolved interpretation, lack of boundary-tagged energy metrics, contradiction analysis and lack of long-term deployment evidence [4, 6, 10-12] (Table 1).

Table 1 Position of this review relative to recent reviews of data-centre and immersion cooling. Full coverage is indicated by “yes”, partial or qualitative coverage by “partial”, and absent coverage by “no”.

Review (year)	Scope	Documented search protocol	SPIC vs TPIC criterion matrix	Quantitative Comparison tables	Energy metrics tagged to system boundary	Scale-resolved analysis	Long-term reliability and ageing	Contradiction analysis
Cai and Gou (2026) [4]	Passive and active data-centre cooling, all architectures	No	Partial	Partial	No	No	No	No
Al Kez et al. (2025) [6]	AI-driven cooling technologies for HPC data centres	No	Partial	Partial	No	Partial	No	No
Zhou et al. (2024) [10]	Immersion cooling development status, SPIC and TPIC	No	Partial	Partial	No	Partial	Partial	No
Zheng et al. (2025) [11]	Single-phase immersion cooling, comprehensive	No	No (SPIC only)	Yes	No	Partial	Partial	No
Lua and Wang (2026) [12]	SPIC advances, control and sustainability metrics	No	No (SPIC only)	Partial	Partial	Partial	Partial	No
This review	Immersion cooling for high-density and AI computing	Yes	Yes (12 criteria)	Yes (13 tables)	Yes (B1-B4)	Yes (5 levels)	Yes (Section 10.3)	Yes (Section 6.4)

Notes: Assessments are based on the explicit presence of seven features: reproducing search methodology; criterion-based comparison of SPIC/TPIC; quantitative data with test conditions; explicit definition of energy-system boundaries; reporting on the spatial scale; long-term ageing and material compatibility; and analysis of conflicting findings. Partial means partial or not complete coverage.

The gap is then in lack of a common framework rather than lack of information. The use of various coolants, loads, surfaces, scales, and system boundaries makes it hard to establish which architecture is appropriate for a specific heat flux or rack density, what enhancements are hydraulic, and the strength of the evidence regarding long-term operation.

1.4 Novelty, Scope and Contributions

This review addresses these gaps by five connected contributions: the reproducible literature-search and screening process; the five-level chip-to-facility framework; the 12-criterion SPIC/TPIC comparison; harmonised quantitative tables that preserve the test conditions and energy boundaries; and explicit analysis of contradictions, evidence maturity, and deployment readiness.

Immersion cooling of electronic computing equipment for dense and AI systems. Only as a reference architecture, air and direct-to-chip cooling are used. Batteries, photovoltaic modules, and power-electronic applications unrelated to batteries are excluded unless quantitative thermal, hydraulic, energy, economic, or reliability data can be extracted in a relevant manner, as are studies lacking in quantitative thermal, hydraulic, energy, economic, or reliability data.

2. Review Methodology

2.1 Search Strategy

The databases searched were Scopus, Web of Science Core Collection and ScienceDirect, and the last search was made in March 2026. Searches targeted the title, abstract, and keyword fields of all the documents in Table 2, with backward and forward citation chasing. The final corpus consists of 90 studies, out of which 53% were published in 2024-2026.

Table 2 Search protocol: databases, search strings, filters, and inclusion and exclusion criteria applied at each screening stage.

Element	Specification
Databases	Scopus; Web of Science Core Collection; ScienceDirect
Publication window	January 2015-March 2026 (last search executed March 2026)
Fields searched	Title, abstract, author keywords
Primary search string	("immersion cooling" OR "immersion-cooled" OR "liquid immersion" OR "two-phase immersion" OR "single-phase immersion") AND ("data center*" OR server OR CPU OR GPU OR "high-performance computing" OR HPC OR "artificial intelligence" OR electronics)
Secondary strings	("dielectric coolant" OR "dielectric fluid") AND (server OR "data center*"); ("pool boiling" OR "nucleate boiling") AND ("dielectric" OR "HFE" OR "Novec" OR "FC-72") AND electronics; ("direct-to-chip" OR "cold plate") AND "data cent*" (reference architecture only)
Supplementary sources	Backward and forward citation chasing on every included study; reference lists of the reviews in Table 1

Inclusion criteria	(i) immersion cooling, or a directly comparable liquid-cooling architecture used as a reference case, applied to electronic computing hardware; (ii) at least one extractable quantitative thermal, hydraulic, energy, economic or reliability result; (iii) coolant, heating configuration and load stated in sufficient detail for the result to be interpreted
Exclusion criteria	Immersion of batteries, photovoltaic modules or other non-computing hardware without transferable findings; purely descriptive, promotional or market-survey material; performance reported without any statement of coolant or load conditions; datasets duplicated from the same group; conference papers superseded by an archival journal version
Data extracted	Coolant and reported properties; heating surface and enhancement; heat load, heat flux and its reference area; flow condition, flow rate, inlet temperature; h , R_{th} , ΔT ; Δp and pumping power; energy metric and its system boundary (B1-B4); spatial scale; test duration; reported uncertainty
Quality appraisal	Completeness of reported conditions; statement of uncertainty; validation of numerical models against experiment; independence of the performance claim from the enhancement promoted
Screening outcome	1148 records identified; 806 after de-duplication; 312 after title/abstract screening; 173 full-text assessed; 90 included

Notes: Screening counts correspond to the search executed in March 2026. Search strings are reported in the form used in Scopus and were adapted to the syntax of the other databases.

2.2 Inclusion and Exclusion Criteria

We selected studies that selected if they discussed computing-related immersion cooling (or a directly comparable reference architecture), presented at least one quantitative result, and provided sufficient details about the immersion cooling coolant, heating configuration, and load to be analyzed. Non-computing applications, those without any transferable findings, promotional or descriptive material, incomplete datasets, and superseded conference papers were not included.

2.3 Data Extraction and Harmonisation

Properties of the coolant, heating surface and enhancement, heat load/flux and reference area, flow condition, inlet temperature, h , R_{th} , ΔT , pressure drop, pumping power, energy metric and boundary, spatial scale, duration, and uncertainty were recorded for each study. Harmonization of heat-flux bases, when possible, and tagging of energy metrics to boundaries B1-B4; no estimation of unreported quantities.

2.4 Quality Appraisal

The completeness of the operating conditions, the reporting of uncertainty, numerical validation, and independence of the performance claims from the promoted enhancement were studied. No study was rejected on quality grounds; appraisal was used to weight the results and identify claims lacking hydraulic or auxiliary-power data.

2.5 Multi-Scale Evaluation Framework

Five scales were used to interpret the results: chip/die, package/heat sink, board/server, tank/rack and facility. Metrics and governing variables vary by scale, and constraints flow both ways: local thermal limits limit the system; and the temperature of heat rejection at a facility and ambient conditions limit chip-level performance. Figure 2 shows the multi-scale evaluation framework used in this review.

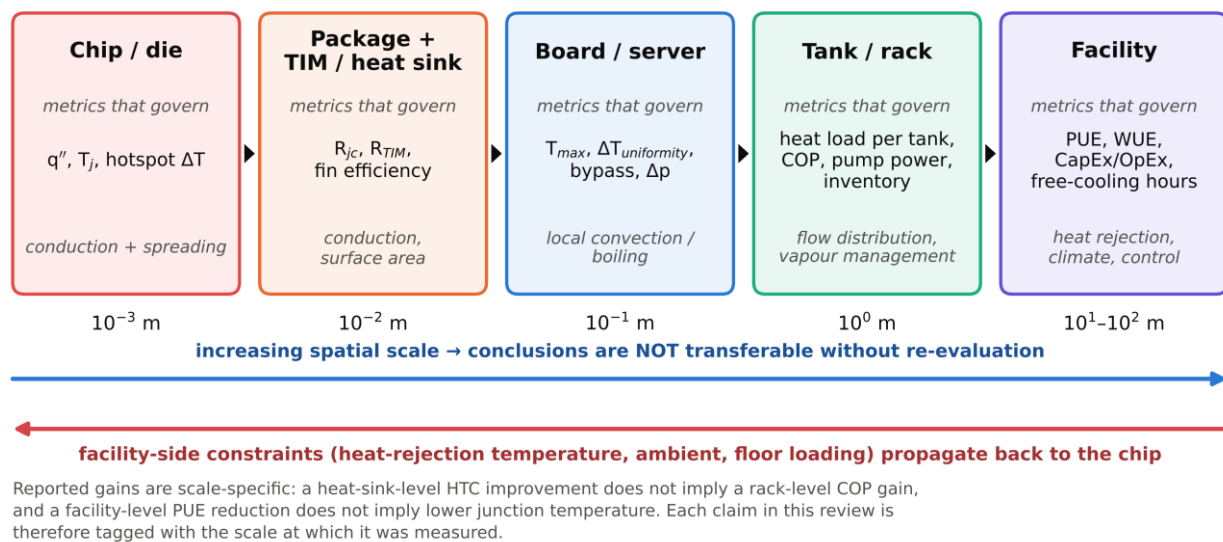


Figure 2 Multi-scale evaluation framework applied in this review.

There are multiple physics associated with each scale, and valid performance metrics for each scale; constraints that move upward from the die and downward from the facility heat-rejection system. Each quantitative finding mentioned in Sections 6-9 is identified as having occurred at which scale.

3. Thermal Requirements of High-Density and AI Computing

With high-density computing, high power and small package sizes challenge density, reliability, and operation [6]. The literature shows five design-relevant differences between existing servers: higher heat flux, higher spatial non-uniformity, multiple interacting heat sources, faster load transients and a much higher power density of the racks (Table 3).

Table 3 Thermal characteristics of modern high-density and AI computing hardware, and their implications for the selection and design of immersion-cooling systems.

Characteristic	Conventional enterprise hardware	Current AI/HPC hardware	Consequence for immersion-cooling selection and design
Package power	150-300 W per CPU	700-1400 W per accelerator	Sets the required junction-to-coolant conductance; at the upper end, natural-circulation SPIC is generally insufficient and forced circulation or phase change is required
Area-average surface heat flux	0.5-1.5 W cm ⁻²	3-10 W cm ⁻²	Determines which mechanism (Section 3.1) can operate at an acceptable ΔT ; excludes forced-air cooling and narrows the CHF margin available to plain-surface TPIC
Local (hotspot) heat flux	1.5-3 × area average	3-10 × area average over compute tiles and memory stacks	Area-average h becomes an incomplete descriptor; local coolant availability, not bulk circulation, governs the peak temperature
Spatial distribution	One or two dominant heat sources per board	Multiple dissimilar dies, chiplets and HBM stacks in close proximity	Components interact thermally through the coolant; component arrangement and inter-board spacing become thermal design variables
Temporal behaviour	Slow, load-following variation	Step transients from synchronised collective operations, seconds time scale	Steady uniform heating does not characterise the architecture; SPIC damps transients but responds slowly, TPIC is near-isothermal but approaches CHF rapidly under a load step
Rack power density	5-15 kW	40-130 kW and rising	Heat rejection, structural loading, fluid inventory and service access, rather than server-level cooling, frequently become the binding constraint
Dominant Design objective	Junction Temperature below limit	Junction temperature, temperature uniformity, transient margin and auxiliary energy simultaneously	Requires the multi-objective treatment of Section 7.6 rather than single-objective temperature minimisation

Notes: Power and heat-flux values are representative of currently deployed hardware classes and are intended to define the design context; they are not drawn from a single source and will continue to shift upward.

3.1 Heat Flux and the Air-Cooling Ceiling

The 700-1400 W dissipations of AI/HPC accelerator packages are approximately $3\text{--}10\text{ W cm}^{-2}$ at the package-average level and local levels even higher. If one chooses $300\text{ W m}^{-2}\text{ K}^{-1}$, then the temperature rise for air becomes too large, but forced SPIC and nucleate boiling can be used in this range of flux. These regimes are plotted on a common heat-flux basis in Figure 3.

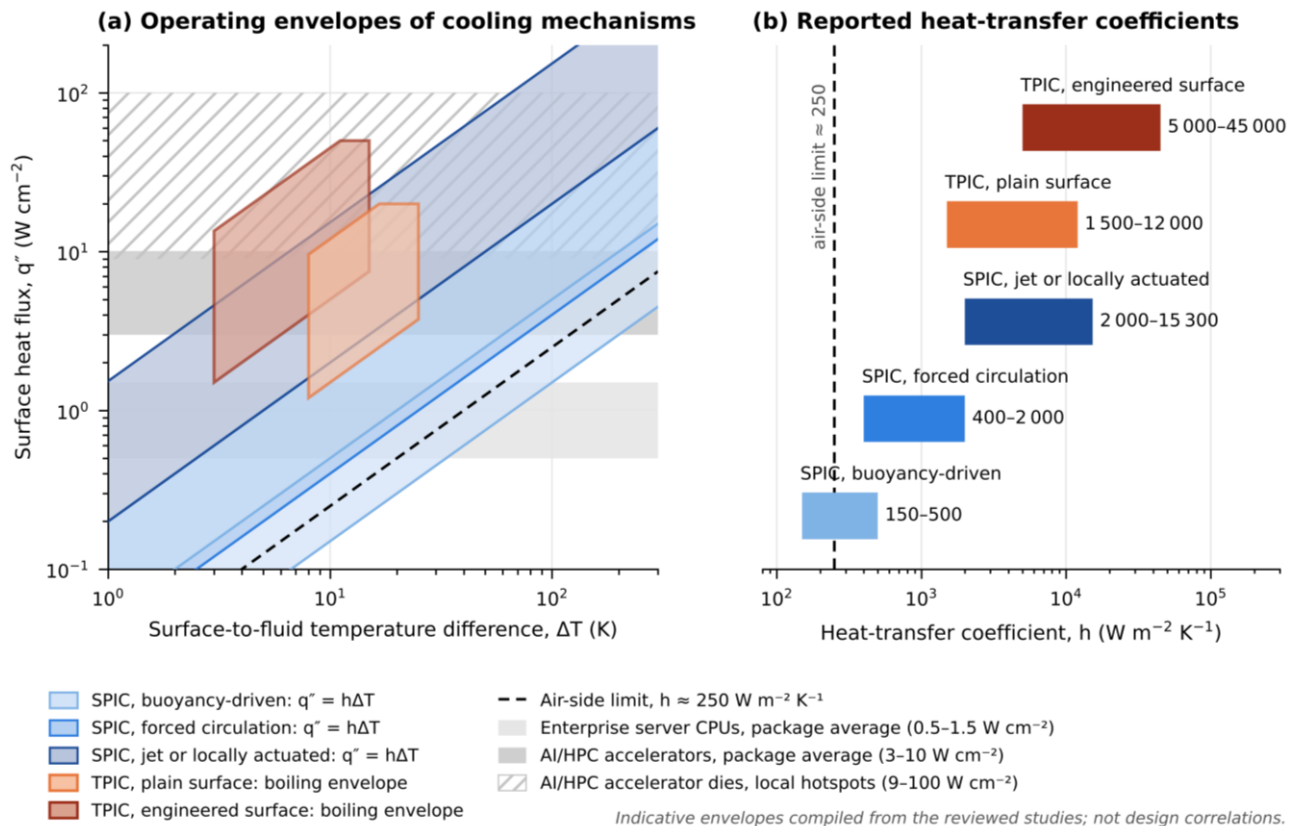


Figure 3 Plotted on the same set of axes, competing cooling mechanisms operate in different regions, (a) the surface heat flux is plotted as a function of the surface-to-fluid temperature difference, with single-phase convective envelopes for the reported ranges of heat-transfer coefficient, nucleate-boiling envelopes for plain and enhanced surfaces bounded by the reported ranges of wall superheat and heat flux, the air-side limit stated in Section 1.1, and the heat-flux bands of mainstream server processors and current AI/HPC accelerators (package average and local hotspots), with forced or jet-assisted SPIC and plain-surface TPIC among the envelopes that intersect these bands; and (b) the heat-transfer coefficient ranges reported from the reviewed studies for each cooling mode. Envelopes are ranges of values derived from the literature reviewed and are not design correlations.

Cooling must provide acceptable temperature control for processors, memory, storage, and power electronics, and to reject heat to the building. So, the maximum temperature, uniformity, the capability of transferring heat, and the cooling energy should be taken into account simultaneously [13].

3.2 Localised Hotspots and Non-Uniform Heat Generation

Temperature control may not be manifested in the package-average h for modern accelerators with local hotspots and dissimilar chiplets. The availability of local coolant is more significant than bulk circulation; component arrangement and inter-board spacing are thermal design variables.

The same applies to adjacent hardware studies: heat spreading can be used to spread the heat of the processor onto larger rejection surfaces, and dense memory arrangements need heat to be spread away from confined areas [14, 15]. Both methods are effective in bringing heat to the areas where there is coolant in immersion systems.

Thermal interaction between the accelerators, memory, and voltage-regulation components occurs at the board scale through coolant action. These studies of component arrangement and fin geometry on the server level recently confirmed the importance of server-level evaluation over single-device evaluation [16-20].

3.3 Dynamic Workloads and Transient Behaviour

There is a large power transient associated with AI/HPC workloads, but most immersion testing practices rely on uniform heating. The thermal inertia is large for SPIC, and the response is slow, while the response is near-isothermal for stable boiling and can reach CHF quickly after a load step for TPIC. Future validations should therefore include representative transient workloads.

3.4 Rack-Level Power Density and Facility Coupling

The power of the rack has increased from about 5-15 kW of conventional enterprise systems to 40-130 kW for AI racks. In addition to server-level heat transfer limitations, at this scale, heat rejection, structural loading, fluid inventory, service access and service redundancy can become more limiting.

Therefore, the evaluation criterion for immersion systems should be based on maximum temperature, temperature uniformity, heat-transfer capability, cooling power and equipment density that can be achieved, instead of a single thermal value.

4. Heat-Transfer Mechanisms and Operating Envelopes

First, we set up governing mechanisms for air and direct-to-chip reference architectures, then for SPIC and TPIC, and finally a common engineering comparison.

4.1 Reference Architectures: Air and Direct-To-Chip Cooling

Air cooling removes heat by circulating air through the server and the room/facility air-handling equipment. Non-uniform airflow can also cause hotspots [7, 21], and higher loads require more airflow and fan pressure but consume more infrastructure volume [22].

Direct-to-chip cooling reduces the thermal path of heat through a cold plate and liquid loop, but still has thermal interfaces and typically keeps memory, voltage regulators, and storage on an air path. Performance is thus sensitive to cold-plate design, flow distribution, pressure drop, pump power, and incorporation of the facility [23-29].

Immersion changes the architecture in a more basic way: immersed components are in direct contact with the dielectric coolant, and are less dependent on server airflow and traditional air-handling systems [30-34].

4.2 Single-Phase Immersion Cooling

In SPIC, the heat is transferred by convection and the coolant temperature stays below its boiling point. A key advantage of natural circulation is that it requires no primary pump, but its velocity is limited by buoyancy and geometry, whereas forced circulation increases local velocity and h but also increases pressure drop and pumping power [30, 35].

$$Q = \dot{m}c_p\Delta T_{coolant} \quad (1)$$

where Q is the heat-removal rate, $\dot{m}$ is the mass flow rate, c_p is specific heat, and ΔT is coolant temperature rise.

$$Q = hA\Delta T \quad (2)$$

$$R_{th} = (T_j - T_{coolant})/Q \quad (3)$$

Bulk flow is not enough because of equations (1)-(3) as follows: An increase in $\dot{m}$ increases transport capacity, but for the local h to be adequate, it must be augmented by flow over high-power surfaces. Thus, the performance of SPIC is influenced by the property of the coolant, geometry of the heat sinks, arrangement and distribution of servers [36, 37].

4.3 Two-Phase Immersion Cooling

The boiling of the dielectric in TPIC takes place on heated surfaces; vapor is then transported to the condenser where it condenses and releases the latent heat and flows back to the bath [16, 38]. The high heat transfer rates of latent transport depend on the very low temperature differences, but they rely on nucleation, bubble departure, replenishment of liquid, vapour transport, and condensation.

$$Q = \dot{m}[c_p(T_{sat} - T_{in}) + xh_{fg}] \quad (4)$$

For a coolant entering below saturation, $Q = \dot{m}[c_p(T_{sat} - T_{in}) + xh_{fg}]$, where T_{sat} is saturation temperature, x is vapour quality, and h_{fg} is latent heat.

The strong dependence of TPIC on bubble interaction, surface orientation, source spacing, and vapour collection [39, 40] is due to multiple heat sources. These effects differ from behaviour in a single-heater boiling situation, especially in terms of server scale.

4.4 Operating Envelopes and the Critical-Heat-Flux Constraint

The operating envelope is the practical difference. SPIC can deliver increasing heat flux, but at increasing pumping penalty and temperature; TPIC can achieve high flux at small superheat but will have a harder CHF limit beyond which rapid temperature rise can result from vapour blanketing.

There is significant overlap in the reported values of h ; forced SPIC lies in the range of $0.4\text{--}2\text{ kW m}^{-2}\text{ K}^{-1}$, jet/local SPIC $2\text{--}15.3\text{ kW m}^{-2}\text{ K}^{-1}$, and plain-surface TPIC $1.5\text{--}12\text{ kW m}^{-2}\text{ K}^{-1}$. The next step is to design the boiling surfaces to achieve heat-transfer coefficients of about $5\text{--}45\text{ kW m}^{-2}\text{ K}^{-1}$. In fact, the choice of architecture is sensitive not only to CHF margin but also to vapour handling, containment and operating complexity [38, 39, 41–44].

4.5 Systematic Comparison of SPIC and TPIC

SPIC conveys sensible heat; TPIC conveys sensible and latent heat by means of boiling and condensation. SPIC is typically easier to operate while TPIC may offer higher flux, but requires stable boiling, vapour management and reliable condensation [16, 40].

A comparison of the 12 engineering criteria is continued in Table 4. The main findings are that it is more attractive to be close to the top of the heat-flux envelope with TPIC and that SPIC is more predictable, easier to manage fluid, and has better server/facility evidence. Evidence comparisons should be done on the same scale because there are more SPIC systems at the server scale and more TPIC systems at the chip scale.

Table 4 Comparative analysis of single-phase (SPIC) and two-phase (TPIC) immersion cooling based on 12 engineering criteria, including quantitative ranges, as reported in the literature. Values are indicative ranges determined from the studies conducted in various coolants, surfaces, and scales; the limiting condition is specified when a meaningful numerical range cannot be specified.

Criterion	SPIC	TPIC	Which architecture is favoured, and why
1. Transport mechanism	Sensible heat, forced or buoyancy-driven convection	Sensible plus latent heat; nucleate boiling and condensation	Neither; the mechanisms set different limits rather than different capabilities
2. Reported heat-transfer coefficient	$150\text{--}500\text{ W m}^{-2}\text{ K}^{-1}$ buoyancy-driven; $400\text{--}2000\text{ W m}^{-2}\text{ K}^{-1}$ forced; $2000\text{--}15300\text{ W m}^{-2}\text{ K}^{-1}$ jet or locally actuated	$1.5\text{--}12\text{ kW m}^{-2}\text{ K}^{-1}$ plain surface; $5\text{--}45\text{ kW m}^{-2}\text{ K}^{-1}$ engineered surface	TPIC at the top of the range, but the envelopes overlap over most of the design space (Figure 3(b))
3. Admissible surface heat flux	No intrinsic ceiling; temperature penalty rises linearly with flux	High flux at small superheat, terminated by CHF	TPIC below CHF; SPIC above it, since SPIC degrades gracefully whereas TPIC fails abruptly
4. Temperature uniformity	Set by flow-path geometry and component arrangement; 20–46%	Near-isothermal at T_{sat} during stable nucleate boiling;	TPIC where boiling is stable and uniform; SPIC is more

	improvements reported from flow redistribution	degraded by local vapour accumulation	predictable and more designable
5. Pressure drop and pumping demand	Primary penalty; rises steeply with flow rate and with viscous fluids	Low primary-loop demand; phase change and buoyancy drive circulation	TPIC; this is its clearest operational advantage
6. Auxiliary power and energy overhead	Pump power is continuous and load-dependent; recoverable by geometric flow management and hybrid circulation	Condenser and secondary loop dominate; primary pumping largely avoided	TPIC on primary loop; comparable at facility level once heat rejection is included (Section 8.1)
7. Control complexity	Flow rate and inlet temperature; large thermal inertia, slow response, temperature feedback usable	Pressure, filling ratio, condenser capacity, vapour venting; little temperature warning before CHF	SPIC; the control problem is better conditioned and failure is not abrupt
8. Fluid inventory and containment	Liquid retention only; hydrocarbon and ester fluids at low unit cost	Liquid and vapour retention; volatile fluorinated fluids at high unit cost; loss at each intervention	SPIC, on both capital and recurrent fluid cost (Section 9.1)
9. Material and ageing exposure	Continuous liquid contact; oxidation and moisture uptake are the main ageing pathways	Liquid, vapour and condensate contact; thermal decomposition at high local surface temperature; fouling of microstructured boiling surfaces	SPIC; the exposure pathways are fewer and the fluids more chemically characterised in service
10. Maintenance and serviceability	Drain, drip-off and cleaning; no fluid loss penalty beyond carry-over	As SPIC plus vapour loss at each lid opening, making intervention frequency an economic variable	SPIC; the difference grows with maintenance frequency
11. Scalability to facility scale	Tank and cabinet studies exist; coupling to secondary loop and heat rejection studied; structural	Condenser capacity, vapour handling and containment must scale with IT load; very few multi-server	SPIC; the scaling path is demonstrated further, and its constraints are

	loading and inventory scale with tank volume	or facility-scale studies	hydraulic rather than phase-change related
12. Evidence maturity	More server-scale, tank-scale and facility-coupled data; one five-month exposure study	More chip-scale boiling data; very little server-scale or long-duration evidence	SPIC; a literature-wide comparison is biased by this asymmetry unless scale is controlled

Notes: Quantitative ranges are compiled across the reviewed studies, which differ in coolant, surface, scale, and instrumentation, and are therefore indicative envelopes rather than design values. CHF = critical heat flux.

5. Dielectric Coolants: An Engineering Selection Perspective

Selecting a dielectric coolant has to take into consideration thermal, hydraulic, electrical, compatibility, safety, economic, and environmental considerations [42, 45].

5.1 Coupled Property Requirements

- In SPIC, conductivity and volumetric heat capacity help to transport heat, and viscosity helps to increase hydraulic demand. Saturation temperature, latent heat, vapour density and surface tension control boiling in TPIC. After ageing, the dielectric properties must be satisfactory, and compatibility with boards, solder, TIMs, polymers, adhesives, seals and elastomers.
- One of the most important properties that can be used to distinguish SPIC and are significantly influenced by temperature is viscosity. Comparative studies of fluorinated liquids show that differences in thermal-hydraulic properties can follow the viscosity more closely than conductivity [42, 46].
- Coolant choice influences deployability, which is increasingly dependent on fluid inventory, unit cost, global warming potential, PFAS status, biodegradability, and end-of-life route.

5.2 Property Data for the Candidate Fluid Families

In Table 5, representative properties from the literature of the fluid families found in the reviewed literature are compiled, such as mineral and synthetic hydrocarbon oils, natural and synthetic esters, polyalphaolefins, and the hydrofluoroether and perfluorocarbon families of fluids used primarily in TPIC [47]. The values are provided at a specified reference temperature because these properties, especially viscosity, are highly temperature dependent; the table should be used by comparison with families of fluids and the values of the properties of any candidate fluid used should be obtained from the current data sheet of the manufacturer.

Table 5 Some representative characteristics of the families of dielectric coolants for immersion cooling, at about 25°C unless otherwise specified. The values are indicative and are taken from manufacturer data sheets and literature and may differ between product types within the family and with temperature, and should be checked against current data sheets for design purposes. n.a. = not applicable; n.r. = not consistently reported.

Property (at ≈25°C unless stated)	Mineral oil (transformer grade)	Synthetic ester	Polyalpha-olefin (PAO)	HFE-7000 (Novec 7000)	HFE-7100 (Novec 7100)	FC-72	FC-40
Typical architecture	SPIC	SPIC	SPIC	TPIC	TPIC	TPIC	SPIC
Boiling/saturation temperature (°C)	≈300 (b.p.)	>300 (b.p.)	>300 (b.p.)	34	61	56	155
Density, ρ (kg m ⁻³)	≈860	≈970	≈800-820	1400	1510	1680	1870
Specific heat, c_p (J kg ⁻¹ K ⁻¹)	≈1900	≈1880	≈2100-2200	1300	1183	1100	1100
Thermal conductivity, k (W m ⁻¹ K ⁻¹)	≈0.13	≈0.14	≈0.14	0.075	0.069	0.057	0.065
Dynamic viscosity, μ (mPa·s)	≈15-30	≈25-35	≈15-30	0.61	0.58	0.64	≈4.1
Latent heat, h_{fg} (kJ kg ⁻¹)	n.a.	n.a.	n.a.	142	112	88	n.a.
Surface tension, σ (mN m ⁻¹)	≈30	≈30	≈28	12.4	13.6	10	≈16
Dielectric strength (kV, 2.5 mm gap)	≈30-70	>75	≈40-50	≈40	≈40	≈38	≈46
Relative permittivity (-)	≈2.2	≈3.2	≈2.1	7.4	7.4	1.8	1.9
Flash point (°C)	≈150-200	≈250-320	≈200-260	none	none	none	none
Global warming potential (100 yr)	low	low	low	≈530	≈320	≈9300	≈10000
PFAS classification	no	no	no	yes	yes	yes	yes

Relative fluid cost	low	low-moderate	moderate	high	high	high	high
Principal engineering limitation	High viscosity; oxidation and additive depletion	High viscosity; hygroscopic	Cost; Elastomer compatibility	Low T_{sat} constrains heat-rejection temperature; PFAS	Low latent heat; PFAS	Very high GWP; low latent heat	High viscosity for its family; very high GWP

Note: These values are typical for each family and are based on manufacturer data sheets and literature. Every property, and especially the viscosity, is highly temperature dependent, and products within a family differ; the property values must be verified against the current product data sheet before using in the design. b.p. = boiling point, n.a. = not applicable (non-volatile in the service range), GWP values are order of magnitude indications. No table even provides the value of water ($k \approx 0.6 \text{ W m}^{-1} \text{ K}^{-1}$, $c_p \approx 4180 \text{ J kg}^{-1} \text{ K}^{-1}$, $h_{\text{fg}} \approx 2260 \text{ kJ kg}^{-1}$) as it is not dielectric, but its property values are what make the dielectric requirement a significant thermal penalty.

Table 5 illustrates the simple trade-off; hydrocarbon/ester fluids generally are more conductive and more heat-capacitive, but also have much higher viscosity, favouring SPIC, whereas the fluorinated fluids have low saturation temperatures favouring TPIC, but lower latent heat and greater fluid-management and environmental constraints.

5.3 Figures of Merit and the Failure of Single-Property Selection

Fluid rank depends on the circulation mode. Forced SPIC gives credit for conductivity and volumetric heat capacity, and penalises the viscosity; buoyancy-driven systems penalise viscosity more heavily. The concepts of latent heat, surface tension and vapour properties come into play in case of boiling, which introduces a different set of merits. Coolant selection needs to be considered a multi-criteria architecture decision and not a ranking of a single property [45, 46, 48].

5.4 Nanofluids and Modified Coolants

Adding more particles to nanofluids increases effective thermal conductivity; however, viscosity and hydraulic resistance also rise with the number of particles added [49, 50].

Other concerns for immersion use include long-term suspension stability, dielectric strength, deposition, loading and pump wear. Thermal and hydraulic drawbacks are moderate, while maintenance drawbacks are significant, keeping nanofluids in the research stage.

5.5 Environmental and Regulatory Constraints

Fluid families also have their environmental and regulatory factors. The GWP and PFAS regulation risk of the volatile fluorinated fluids for TPIC are high, with hydrocarbon and ester fluids posing other ageing and flammability concerns. Fluid loss and end-of-life impacts should, therefore, be part of the life cycle assessment, alongside operational energy.

6. Thermal Performance: Quantitative Synthesis

In an immersion configuration, electronics couple directly to the dielectric coolant, and performance is assessed based on maximum temperature, uniformity, thermal resistance, h , acceptable heat load, and cooling power [41, 51]. Sections 6.1-6.4 summarise the quantitative evidence and its paradoxes.

6.1 Single-Phase Immersion Cooling: Reported Thermal-Hydraulic Performance

Local coolant velocity and distribution significantly affect SPIC performance. Low thermal resistance may arise from direct contact, while poor flow paths may result in high power regions under-cooled [35, 41, 52] (Table 6).

Table 6 Reported performances for single-phase immersion cooling (SPIC) has been quantified, with test conditions noted. The coolant, test article and scale, thermal load, circulation condition and the reported thermal and hydraulic results are stated in each row. If quantities were not reported from the source, they are indicated by the n.r. which is also an observation of this review.

Configuration class	Coolants covered	Scale and load range	Reported h ($W m^{-2} K^{-1}$)	Reported R_{th} or ΔT	Hydraulic cost reported?	Controlling variable identified	n
Buoyancy-driven SPIC, single device or board	Mineral oil, coconut oil, PAO, FC-40	Chip to single server; 50-500 W per device	150-500	$\Delta T_{surface-to-fluid}$ 15-40 K	Not applicable (no pump)	Fluid viscosity and tank geometry; performance limited by the Rayleigh number	7
Forced-circulation SPIC, single server	Mineral oil, synthetic ester, FC-40, fluorinated liquids	Single server; 0.5-3 kW per server	400-2000	R_{th} 0.08-0.4 $K W^{-1}$ where reported	In 5 of 14 studies	Local coolant velocity over the high-power devices, not bulk flow rate	14
SPIC with flow-guiding structures	Mineral oil, fluorinated liquids	Server to cabinet; 1-20 kW per tank	500-1800	20-46% improvement in temperature uniformity	In 6 of 9 studies	Baffle angle, equalisation-plate porosity, inlet/outlet placement	9
Jet-assisted and locally actuated SPIC	Fluorinated liquids, mineral oil	Chip to server; 100-800 W per device	2000-15300	Convective R_{th} down to 0.08 $K W^{-1}$	In 5 of 6 studies	Jet Reynolds number, nozzle-to-surface spacing, actuator power	6
Optimised immersion heat sinks	Mineral oil, fluorinated liquids	Package to server; 200-1000 W per device	600-2500	10-28% reduction in peak temperature	In 7 of 8 studies	Fin pitch and height; oblique or staggered arrangement for uniformity	8

Tank- and cabinet-scale SPIC	Mineral oil, synthetic ester	Multi-server tank; 10-60 kW per tank	n.r. (system-level metrics reported)	Cabinet ΔT 5-15 K between servers	In 8 of 11 studies	Inter-server flow distribution; coupling to secondary loop and heat rejection	11
------------------------------	------------------------------	--------------------------------------	--------------------------------------	---	--------------------	---	----

Notes: The ranges are based on different published studies across the configuration class and account for variations in coolant, geometry, load and instrumentation and are not single study results, but should not be used as design values. n = number of included studies that give rise to the range in the row. The value in the “hydraulic cost reported?” column indicates the number of studies in each class that reported pressure drop or pumping power along with the thermal result; n.r. = not reported, is a discovery of this review: fewer than half of the forced-circulation studies report the auxiliary-energy cost of the performance they claim. The individual values for all 90 studies included are provided in a study-by-study extraction sheet available from the corresponding author.

- Natural-circulation h values are concentrated in the region of $150\text{--}500\text{ W m}^{-2}\text{ K}^{-1}$, whereas forced circulation h values are concentrated in the region of $400\text{--}2000\text{ W m}^{-2}\text{ K}^{-1}$ and jet/local actuation h values can reach up to $15.3\text{ kW m}^{-2}\text{ K}^{-1}$. The benefits of forced flow must be balanced against pumping power, which is not well documented in the literature.
- The thermal benefit of a change of the flow through a system, such as providing more equalising plates, baffles or inlet/outlet changes, is greater than increasing the total flow of the system beyond a certain point, as Eq. (2) is a function of local h rather than bulk $\dot{m}$.
- Peak temperature and uniformity are mildly related but very different goals: peak temperature is mostly dependent on local velocity and inlet temperature; uniformity is mostly dependent on flow path, placement of components and depth of immersion.
- Coolant rankings also change when using different flow modes of circulation, since the importance of viscosity differs between buoyancy and forced flow. The apparent differences between fluid studies may therefore be because they are operating in different regimes, not different physics.

6.2 Two-Phase Immersion Cooling: Reported Boiling Performance

Latent heat provides high local heat transfer rates at almost constant bulk temperature; however, TPIC has a limited range of application beyond CHF and dependent on the coolant, surface, orientation, and operating conditions [39, 40, 53] (Table 7).

Table 7 A compilation of reported two-phase immersion cooling (TPIC) performance with test conditions labeled: coolant, boiling surface and enhancement, orientation, range of heat flux, superheat of the wall, reported heat transfer coefficient and measured critical heat flux, and system-level metric, where applicable.

Configuration class	Coolants covered	Surface/enhancement	q'' range (W cm ⁻²)	Wall superheat ΔT_{sat} (K)	Reported h (kW m ⁻² K ⁻¹)	CHF reported?	Controlling variable identified	n
Plain-surface pool boiling, single heater	HFE-7000, HFE-7100, FC-72	Polished or as-machined metal	0.5-20	8-25	1.5-12	In 5 of 8 studies	Saturation temperature and latent heat of the fluid; surface roughness	8
Structured and porous boiling surfaces	HFE-7000, HFE-7100	Porous coatings, micro-structures, pin fins	1-50	3-15	5-45	In 4 of 7 studies	Nucleation-site density combined with liquid-replenishment path	7
Wettability-modified surfaces	HFE fluids	Chemical or textured wettability modification	1-30	3-12	4-25	In 2 of 3 studies	Contact angle and its durability under prolonged boiling	3
Multi-chip and board-level TPIC	HFE-7000, HFE-7100	Plain and finned, vertical and horizontal orientation	0.5-15	6-20	1-10	In 2 of 6 studies	Bubble interaction between sources; vapour accumulation; board spacing and orientation	6
Tank-scale TPIC with condenser	HFE fluids	Server boards in a sealed tank with condenser coil	n.r. (system-level)	n.r.	n.r.	Rarely	Condenser capacity, fluid inventory, filling ratio and operating pressure	9

Notes: The ranges are indicative and are compiled across the studies in each configuration class, but are not design values. The “CHF reported?” column records how many studies stated the critical heat flux of the configuration tested; because CHF and not heat-transfer coefficient is the design-limiting quantity in TPIC, the low reporting frequency is a significant weakness of this literature. n = number of included studies contributing to the row; n.r. = not reported.

- Surface engineering generally produces larger TPIC gains than switching among fluids within the same family. The h can be raised by a factor of about two to four by porous, microstructured and pin-finned surfaces, as well as by modification of the wettability [31, 53, 54].
- The most important TPIC design limit is CHF, but it is less consistently reported than h . AI/HPC heat-flux requirements may overlap with plain-surface CHF, and so it is sometimes necessary to enhance the surface for high-flux TPIC.
- The differences between multi-chip boiling and single-heater tests are that there is bubble interaction with increased load, vapour accumulation at higher loads, and confined spacing that affects the local heat transfer [39, 40].
- The condenser may become the system limit: the low latent heat of the available dielectric fluids leads to high vapour mass flow, which must be taken into consideration when designing filling ratio, operating pressure, condenser capacity and liquid return with the boiling surface.

6.3 Temperature Uniformity, Hotspot Control and Distributed Components

Immersion depth, heat-source arrangement, flow direction, and coolant distribution are the most important factors for maximum temperature and uniformity at server scale [55, 56].

An immersed processor, Memory, voltage regulators, etc. are cooled simultaneously through immersion; however, stagnant areas, close board spacing and inappropriate inlets/outlets can counterbalance direct liquid contact [57, 58].

The best conclusion across Table 6 and Table 7 is that thermal performance is related to the spatial relationship between the heat generated and the coolant delivered by the servers. The arrangement of the components, the spacing between boards, obstructions, and inlet/outlet location are, therefore, as important as the properties of the coolant.

6.4 Reconciling Contradictory Findings

The major differences in the literature, shown in Table 8, are methodological rather than physical. Recurring causes are: undeclared energy boundaries, incomplete reporting of properties of the coolant, bulk flow assumed as the local flow of coolant, mismatching of spatial scales, and steady uniform heating applied as realistic workloads.

Table 8 Apparently contradictory findings in the immersion-cooling literature, and their resolution. In both instances, the disagreement can be traced to a lack of a common methodology, not to the physics.

Apparent contradiction in the literature	Position A	Position B	Cause of the disagreement	Resolution
Does increasing coolant flow rate improve SPIC performance?	Substantial reduction in component temperature with flow rate	Diminishing or negligible benefit beyond a threshold	Bulk flow rate used as a proxy for local coolant availability; test articles differ in the degree of flow bypass	Both are correct for their configuration. Eq. (1) governs transport capacity, Eq. (2) governs local temperature; where bypass exists, additional flow raises the former without the latter
Is TPIC superior to SPIC in heat-transfer capability?	TPIC by an order of magnitude	Comparable, with overlapping ranges	Comparisons drawn between engineered TPIC surfaces and buoyancy-driven SPIC, i.e. across both architecture and enhancement level	The envelopes overlap substantially (Figure 3(b)). TPIC's advantage is concentrated above the flux at which forced or jet-assisted SPIC becomes impractical, not across the range
Which dielectric fluid performs best in SPIC?	Fluids with the highest thermal conductivity and specific heat	Fluids with the lowest viscosity	Circulation mode differs between studies; viscosity is weakly penalised in forced convection and heavily penalised in buoyancy-driven convection	The rankings are not in conflict. They apply to different merit groups (Section 5.3) and are valid only within the circulation mode in which they were obtained
Do nanofluids improve immersion-cooling performance?	Improvement rises with particle concentration	Optimum concentration beyond which performance degrades	Studies that report only the thermal result versus studies that also report hydraulic performance and pumping power	The thermal enhancement is real but is offset by a viscosity penalty, so a net optimum exists; studies reporting only thermal data cannot locate it

How large is the energy benefit of immersion cooling?	Partial PUE as low as 1.03-1.04	Facility PUE of 1.1-1.2	Undeclared or differing system boundaries (Section 8.1)	The same physical system yields both values on different boundaries. Values must be compared only within the same boundary, which is why every energy metric compiled in Section 8.1 is tagged B1-B4
Does immersion cooling reduce hotspot temperature or only mean temperature?	Marked improvement in temperature uniformity	Hotspots persist and can worsen locally	Uniform steady heating versus non-uniform or multi-source heating; single-device versus board-level test articles	Uniform-heating tests cannot resolve hotspot behaviour. Under non-uniform heating, uniformity depends on the spatial match between heat generation and coolant delivery, not on the coolant alone.
Is buoyancy-driven SPIC adequate for high-density hardware?	Adequate, and avoids pump power entirely	Inadequate above moderate loads	Mismatched scales and loads: single low-power devices versus multi-kilowatt servers	Both hold within their tested range. Buoyancy-driven operation is viable at the device power levels tested in those studies and insufficient at current accelerator power

Notes: Position A and Position B refer to classes of findings present in the examined corpus and not to individual studies, as in both cases several studies are located at both sides of the positions.

Most apparent contradictions can be attributed to these five causes. Declared boundaries allow energy measures to be compared, complete fluid data allow ranking, local flow measurements remove the need to extrapolate from the chip to the facility, scale tags ensure the metrics are not extrapolated from the chip to the facility, and realistic load profiles reveal transient behaviours.

Overall, thermal resistance decreases, hotspots decrease, and density increases, although this depends on the configuration. Cooling, hydraulics and facility heat rejection should therefore be considered as a whole.

7. Heat-Transfer Enhancement and System Optimization

Enhancement strategies involve increasing local heat transfer, maintaining uniformity and controlling hotspots without causing an undue increase in the pressure drop or auxiliary power requirement. Although Table 9 and Figure 4 compare techniques in terms of net benefit, this is not only in terms of thermal gain.

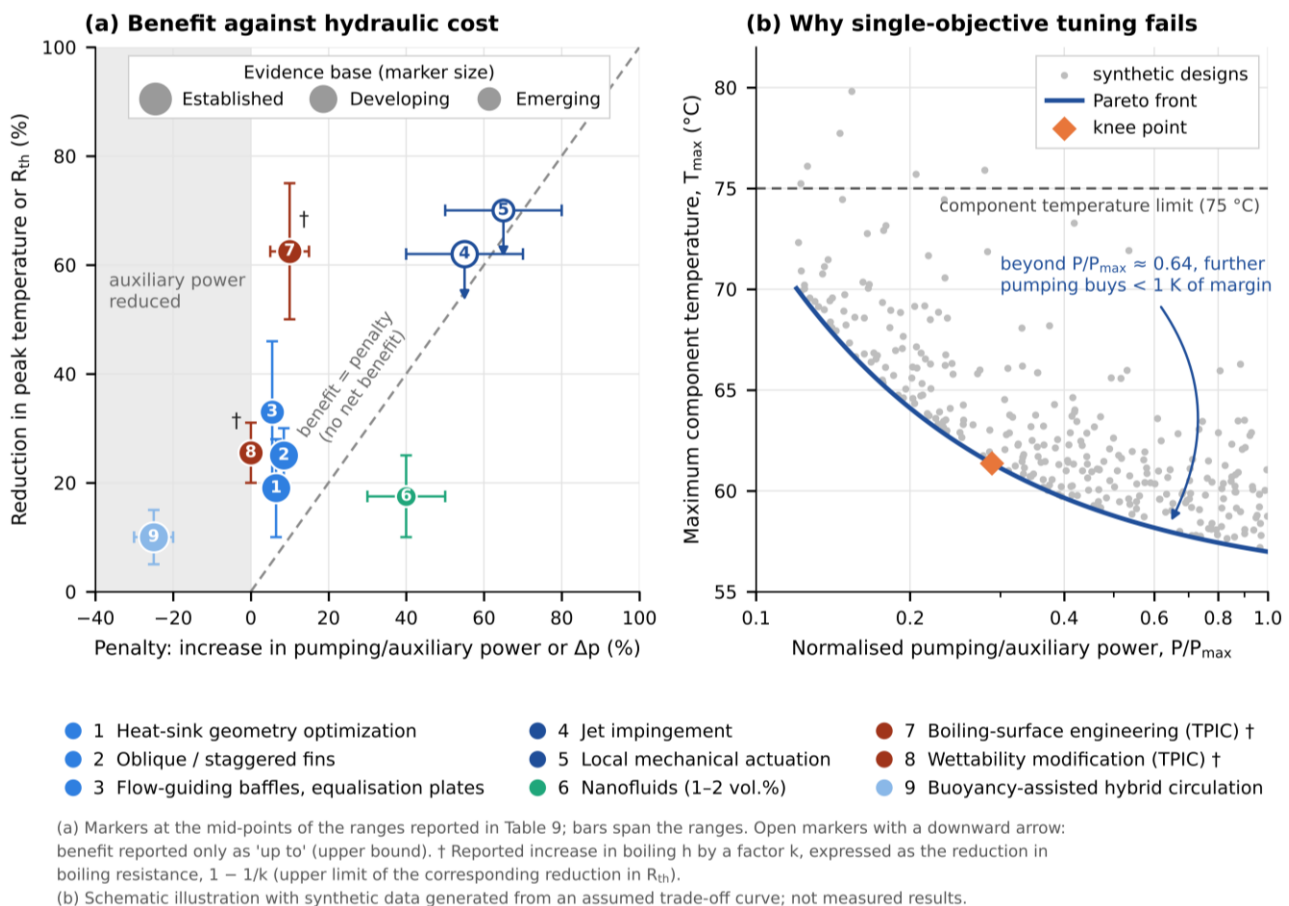


Figure 4 (a) Techniques plotted by their reported reduction in peak temperature or thermal resistance against their hydraulic or auxiliary penalty (Table 9), with markers at the mid-points of the reported ranges (bars span the ranges) or, where only an 'up to' value is reported, at that upper bound (open markers), gains in boiling heat-transfer coefficient expressed as reductions in boiling resistance, and the size of the marker reflecting the maturity of supporting evidence; techniques to the right of (below) the diagonal are less thermally beneficial than the increase in pumping, auxiliary or actuator power they incur; and the shaded area is for techniques that result in a decrease in auxiliary power. (b) Why single-objective tuning is not good - beyond the knee of the trade-off curve, each additional amount of pumping power will buy minimal temperature margin at the expense of consuming auxiliary energy constantly (schematic illustration with synthetic data; not measured results).

Table 9 Mechanism, thermal benefit, hydraulic or auxiliary energy penalty, scale, and principal limitation and maturity of the supporting evidence for heat-transfer enhancement techniques for immersion cooling.

Enhancement technique	Mechanism	Reported thermal benefit	Hydraulic/Auxiliary penalty	Scale	Principal limitation	Evidence
Heat-sink geometry optimization (fin pitch, height, thickness)	Added wetted area with preserved coolant access	10-28% lower peak temperature	3-10% higher Δp	Package, server	Optimum fin pitch is fluid- and circulation-dependent; area added beyond the flow's ability to wet it is inactive	Established
Oblique and staggered fin arrangements	Periodic boundary-layer interruption and inter-passage flow redirection	20-30% lower peak temperature; markedly improved processor-to-processor uniformity	5-12% higher Δp	Package, server	Manufacturing complexity; benefit depends on flow direction relative to the fin orientation	Established
Flow-guiding baffles and equalisation plates	Redistribution of existing flow towards thermally critical components; bypass suppression	20-46% improvement in temperature uniformity or peak temperature	3-8% higher Δp ; no additional pump duty	Server, tank	Configuration-specific; requires knowledge of the heat-source layout at design time	Developing
Jet impingement (immersed jets)	High local velocity and thin boundary layer at the target surface	Local h of 2-15.3 kW m ⁻² K ⁻¹ ; up to 62% lower thermal resistance	40-70% higher pumping power	Chip, server	Benefit is confined to the impingement footprint; continuous auxiliary power	Developing
Local mechanical actuation (ducted propellers, synthetic jets)	Directed local circulation without raising tank-wide flow	Up to 70% lower local thermal resistance	50-80% higher local auxiliary power	Chip	Moving or piezoelectric components immersed in the coolant; service-life reliability unaddressed	Emerging

Nanofluids (1-2 vol.%)	Increased effective thermal conductivity of the coolant	10-25% improvement at the optimum concentration	30-50% higher pumping power from increased viscosity	Chip, server	Long-term suspension stability; reduced dielectric strength; erosion, filter loading and deposition	Emerging
Boiling-surface engineering (porous, microstructured, pin-finned)	Increased nucleation-site density with maintained liquid replenishment	2-4 × higher boiling h; CHF delayed	5-15% (vapour-escape resistance)	Chip, server	Fouling of small-scale features by particulates and degradation products; performance demonstrated on pristine surfaces only	Developing
Wettability modification	Altered surface energy, changing nucleation and rewetting behaviour	25-45% higher boiling h	Negligible	Chip	Durability of the treatment under prolonged boiling and thermal cycling	Developing
Buoyancy-assisted hybrid circulation	Buoyancy substituted for pumping over part of the operating range	5-15% (thermal benefit is modest by design)	20-30% reduction in auxiliary power	Server, tank	Requires a control strategy able to identify the transition point; benefit is load-dependent	Established

Note: These benefit and benefit ranges are combined from the studies reviewed and are reported as a percentage of the baseline employed in each study (which varies from one study to another) and are therefore representative and plotted accordingly in Figure 4(a). If a study did not report the thermal benefit, but did report the cost of the hydraulic or auxiliary power, the penalty number is taken from the other studies in the same class, and the reporting requirement is proposed in Section 11.3. This is one of the reasons for reporting the cost of the hydraulic or auxiliary power.

7.1 Heat-Sink and Fin-Geometry Optimization

The trade-offs in immersion heat sinks are between heat-sink area and flow obstruction. Fin spacing, height, thickness and orientation should be designed to ensure the added area is effectively wetted, as recent oblique and staggered geometries redirect flow, improving temperature uniformity [17-19, 59-61].

7.2 Flow Guiding and Coolant Distribution

To increase the coolant flow to high-power devices and minimize bypass flow, baffles, ducts, channels, and modified inlet arrangements are used [57, 58, 62]. This tends to yield a higher benefit/penalty ratio compared to increasing total circulation.

7.3 Jet Impingement and Locally Actuated Enhancement

Jet impingement focuses high-momentum coolant on critical surfaces, making it a desirable approach for high-flux devices [38, 43, 63, 64].

Jet and locally actuated methods can offer some of the highest SPIC local gains (up to h of order $15 \text{ kW m}^{-2} \text{ K}^{-1}$) but will need extra power to drive the actuators or pump devices and could introduce moving parts or flow interactions [44, 65, 66].

7.4 Boiling-Surface Engineering and Wettability Modification

TPIC surface engineering encourages nucleation and keeps the replenishment of the liquid and evacuation of vapours. Increased area and nucleation sites are provided by fins, pin fins, porous structures and microstructures; but their help is dependent on the coupled system of fluid and surface [31, 53, 54].

Wettability modification is an attractive option as it can improve boiling without a significant hydraulic penalty or auxiliary power, but issues of durability, contamination and reproducibility are not well addressed [53].

7.5 Passive and Hybrid Circulation Strategies

Passive or active circulation trade control against energy use. Natural circulation lowers pump power but generally offers lower capacity, while forced circulation provides higher cooling but higher hydraulic demand [30, 67].

If inventory and elevation are matched with the condenser geometry, then the primary pump can be eliminated using a gravity-driven two-phase loop [68, 69]. Buoyancy is used at low load, with active flow when needed, to provide reduced auxiliary power and control.

7.6 Multi-Objective Optimization

Immersion optimisation should also try to reduce the maximum temperature, non-uniformity, pressure drop, auxiliary power, and coolant inventory at the same time, not just temperature [70, 71].

The trade-off is non-linear: In the beginning, small increases in pumping power or in actuator power can bring substantial thermal improvements, but after a “knee point”, further increments in

power will provide minimal improvements in temperature. That is why multi-objective optimisation is better single-objective tuning [72].

8. Energy Efficiency and Facility Integration

Immersion eliminates or reduces the number of server fans, changing the entire cooling chain, but still requires pumps, heat exchangers, condensers (if applicable), and heat rejection from the facility [1, 73].

8.1 Metrics, System Boundaries and the Comparability Problem

$$PUE = P_{total}/P_{IT} \quad (5)$$

where P_{total} is the total facility power and P_{IT} is the power delivered to the IT equipment.

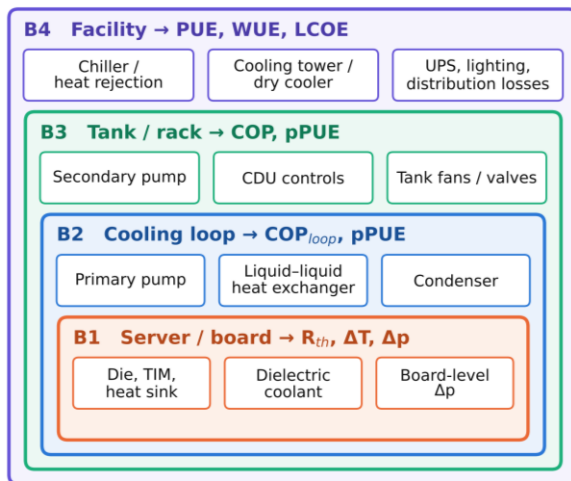
PUE measures facility overhead but does not identify individual cooling contributions.

$$COP = Q/P_{cooling} \quad (6)$$

where Q is the heat removed by the cooling system and $P_{cooling}$ is the electrical power consumed by the cooling equipment within the declared boundary.

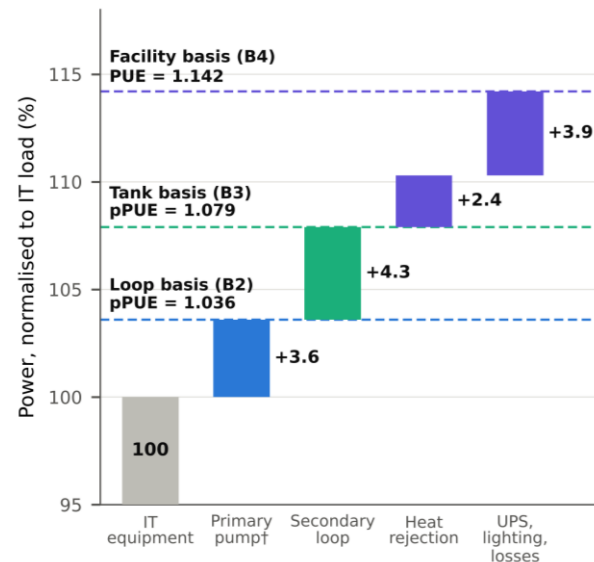
PUE and COP are only meaningful if there is a declared system boundary. This review is based on B1 server/board, B2 primary loop, B3 tank/rack and B4 facility boundaries (Figure 5).

(a) Nested system boundaries



A COP measured on B2 and one measured on B3 differ by the secondary-loop auxiliaries; a partial PUE on B3 is not a facility PUE.

(b) One system, three different numbers



Illustrative accounting, not measured data: the physical system is identical in all three cases (axis starts at 95 %).
 † Primary-pump power includes the pressure drop of the liquid-liquid heat exchanger and condenser.

Figure 5 Energy metrics and system boundaries: (a) four nested boundaries (B1-B4) used in this review, with the components accounted for in each and the measures (partial PUE or PUE) that are appropriately defined on each; (b) illustrative accounting for one physical system, where the reported measure of PUE or partial PUE increases from 1.036 to 1.142 simply as the accounting boundary expands.

All energy metrics in Table 10 are tagged with the boundary at which they are reported.

Table 10 Results of energy performance reported for the immersion cooling architecture and reference cooling architecture, both labeled with the system boundary B1-B4 on which they were tested, and with the IT load and auxiliary power/heat-rejection configuration that was tested. It is not possible to compare values across different boundaries.

Architecture and study type	Boundary	Metric and reported value	IT load	Auxiliary components inside the boundary	Heat rejection/ ambient	Comparability caution
TPIC pilot tank, experimental	B3	pPUE 1.053 → 1.037 as server power rose	1127 → 1577 W	Condenser, secondary loop, tank auxiliaries	Facility water loop; ambient not varied	Partial PUE on a tank boundary; not comparable with a facility PUE
TPIC with multi-mode condenser, simulation	B3	Energy-saving potential reported relative to a single-mode condenser	Modelled tank load	Condenser modes, secondary loop	Modelled climate	Relative result; the absolute value depends on the reference case chosen
SPIC tank with alternative economisers, annual simulation	B4	Energy use, cost and emissions differ simultaneously with economiser choice	Modelled facility load	Pumps, HX, economiser, chiller, facility overheads	Climate-dependent; multiple economiser types compared	Facility boundary; the economiser choice dominates the result
TPIC, experimental thermoeconomic analysis	B3-B4	System performance varies with the operating and economic assumptions applied	Tank-scale load	Condenser, secondary loop, capital charge	Stated operating conditions	Economic result; sensitive to tariff and capital assumptions
Immersed power-supply unit, experimental	B1	Unit efficiency improved by transferring PSU losses directly to the dielectric liquid; fan eliminated	Single PSU	None beyond the immersed unit	Not applicable	Component-level result; does not imply a facility-level benefit

Loop-heat-pipe hybrid CPU cooling, analysis	B4	Annual energy saving 34.9-52.4%; dynamic payback 2.48-6.64 years	Facility-scale model	Loop heat pipe, pumps, heat rejection, facility overheads	Climate- and tariff-dependent	Not an immersion architecture; included as a reference for the magnitude of liquid-cooling savings
Reservoir thermal energy storage for data-centre cooling	B4	Summer-peak COP 16.5 over a modelled 20-year lifetime	5 MW cooling load	Storage, pumps, heat rejection	Site-specific geology and climate	Storage-assisted COP; not comparable with a cooling-loop COP
Trans-critical compressed-CO ₂ storage using data-centre waste heat	B4	Round-trip efficiency 64.67% and 67.41% for the two configurations	Modelled facility	Storage cycle, compressors, heat exchangers	Modelled conditions	An energy-storage metric, not a cooling metric; reported here for completeness

Notes: Boundaries B1-B4 are shown in Figure 5(a): B1 server/board; B2 primary cooling loop; B3 tank/rack and secondary cooling loop auxiliaries; B4 facility. It is not valid to compare values on separate boundaries. Entries for which the source does not mention the IT load, the components of the auxiliary equipment located within the boundary, or the heat-rejection arrangement are reported as reported, one of the purposes of which is to be motivated by the requirements for reporting of Section 11.3.

A benefit of immersion is that it eliminates fan energy overhead on the server, and experimental results indicate that immersed power electronics can be more efficient [74].

A study using TPIC reported that as server power was raised from 1127 W to 1577 W, the partial PUE was reduced from 1.053 to 1.037 [73]. These partial-PUE figures are not directly comparable to whole-facility PUE.

If thermal resistance is lower, more flow or more heat-rejection equipment will be needed - and the energy performance needs to take all the additional loads within a declared boundary.

8.2 Heat-Rejection Temperature and Free Cooling

Liquid cooling can reject heat at higher temperatures than chilled-air systems, expanding the potential for economiser and free-cooling.

The usable temperature for heat rejection for immersion depends on the secondary-loop design, ambient temperature and component limits, and the heat-exchanger approach and operating temperature of the coolant [75].

Free-cooling benefit is not a function of climate alone and should be calculated for a representative year, not just for a particular operating point [76].

8.3 Dynamic Operation and Workload Variation

Adaptive operation can adjust coolant flow and heat rejection based on the number of chips being loaded and the ambient conditions, thereby minimizing the use of auxiliary energy at partial load [77].

This is especially relevant for the case of AI/HPC workloads that experience significant fluctuations in the number of processors active and heat generation.

SPIC control is mainly affected by the transport delay & thermal inertia of the coolant, while for TPIC control, the pressure, filling ratio, condenser capacity & CHF margin are important.

8.4 Waste-Heat Recovery

Liquid cooling concentrates waste heat and can facilitate integration with external thermal systems [78, 79]. The recovered heat should be considered as useful energy based on the heat-pump power, distribution losses, and temporal mismatch. Higher operating temperatures of the SPIC fluids may allow low-temperature heat reuse without the heat pump, while the saturation temperatures of the common fluorinated fluids limit TPIC.

8.5 Thermal and Energy Storage

Thermal storage and energy storage can help to separate cooling needs from electricity availability. Attempts have been made to use liquid-air cooling storage and other storage concepts to this end [80].

Reservoir thermal energy storage has been related to a summer-peak COP of 16.5 and reductions in electricity consumption, cost and emissions for a modelled 5 MW cooling load [81].

Energy storage round-trip efficiency for a waste heat-assisted transcritical CO₂ system is reported to be approximately 64.67-67.41% [82].

Storage strategies can serve several purposes, including storing cooling capacity and/or recovering rejected heat, and can be valuable depending on capital cost, losses, and system complexity.

9. Economic and Environmental Assessment

Deployment capability is not solely thermal. Immersion equipment, avoided air infrastructure, coolant inventory, maintainability, replacement, containment and end-of-life costs are all necessary parts of the economics.

9.1 Cost Structure

Avoidance of fans, air handlers, containment and chiller infrastructure (where appropriate) in new-built facilities can be significant, and the economic case is strongest in these situations. The fluid stock inventory is a capital cost that increases with tank volume, and there is limited quantification of replacement, maintenance or disposal costs. The principal cost items, their direction relative to an air-cooled baseline and their coverage in the literature are summarised in Table 11.

Table 11 The differences in the immersion cooling and air-cooled baseline cost items and their direction, their main drivers, and their literature coverage.

Cost item	Direction relative to an air-cooled baseline	Principal drivers	Magnitude	Coverage in the reviewed literature
Immersion tanks, coolant distribution and heat exchangers	Increase	Tank count and size; redundancy; secondary-loop design	Major capital item	Well covered in techno-economic studies
Dielectric fluid inventory (initial charge)	Increase	Tank volume rather than IT load; fluid unit cost (Table 5); immersion depth; thermal expansion allowance	Major capital item; an order of magnitude greater per litre for fluorinated fluids	Unit costs reported; inventory requirements rarely reported
Server fans and internal air-cooling hardware	Decrease	Number of servers; fan specification	Moderate saving; also, a continuous energy saving	Well covered
Air-handling infrastructure (CRAH, ducting,	Decrease	Facility design; applies fully to new build and only partly to retrofit	Major saving for new build	Covered qualitatively; rarely quantified

containment, raised floor)		separately from the chiller plant		
Chiller plant	Decrease or eliminated	Coolant supply temperature; climate; free-cooling hours (Section 8.2)	Major saving where dry cooling is feasible	Covered in economiser-comparison studies
Cooling electricity (pumps, condensers, heat rejection)	Decrease	Architecture; flow management; control strategy; climate	Principal operating saving	Well covered, but frequently on an undeclared boundary (Table 10)
Water consumption	Decrease with dry rejection	Heat-rejection method; ambient conditions	Material in water-stressed regions	Sparse; WUE is reported far less often than PUE
Fluid make-up and replacement	Increase	Evaporative and handling losses, higher in TPIC; ageing-driven replacement interval; maintenance frequency	Recurrent; potentially significant over the service life for fluorinated fluids	Gap: replacement intervals are not established (Section 10.3)
Maintenance labour and serviceability	Increase	Drain, drip-off and cleaning procedures; hardware orientation; lifting equipment	Recurrent; raises mean time to repair	Gap: not quantified in the reviewed corpus
Structural works (floor loading, spill containment, ventilation)	Increase	Tank mass when filled; fluid classification; fire strategy	Site-specific; can be significant in retrofit	Gap: mentioned qualitatively only
Hardware qualification and warranty	Increase	Vendor support for immersion service; component and label compatibility	Commercial rather than technical	Gap: not addressed in the technical literature
Fluid disposal, reclamation and residual value at end of life	Increase	Fluid chemistry and regulatory classification; reclamation route availability	Uncertain; PFAS classification is a material risk (Section 5.5)	Gap: essentially absent from the reviewed corpus

Note: Direction and magnitude are reported from an air-cooled reference with an equivalent IT capacity. The magnitude column is qualitative because the reviewed studies use different base years, currencies, etc., and different scales for the facilities and it would be misleading to have a single normalised cost comparison across the studies. Any items indicated as gaps are addressed in Section 11.

9.2 Reported Savings and Payback

Cooling-electricity savings need to be balanced against new liquid-cooling equipment costs. The result depends heavily on climate, IT load, coolant temperature, heat rejection, auxiliary equipment, and economisation.

Liquid cooling studies have been reported, and potentially significant energy savings have been achieved, depending on the baseline and climate. The other connected cooling study indicated a 2.48-6.64 years dynamic payback, and 34.9-52.4% annual energy savings [83-85].

Payback values are indicative only, as tariffs, climate and service-life costs for excluded services, particularly fluid replacement and disposal, will significantly influence them.

9.3 Fluid Inventory, Replacement and Containment Cost

Significant fluid inventories are required for large immersion systems, influencing tank size, structural loading, initial system costs and servicing.

With volatile fluid coolants, TPIC is more likely to experience recurrent fluid loss exposure due to the volatilization of fluid (loss during maintenance and/or through containment pathways), and in the case of high fluid cost, intervention frequency becomes an economic decision.

9.4 Environmental Impact and End of Life

During the design of operations, consider energy, water, fluid GWP/PFAS status, make-up fluid treatment, and end-of-life treatment. Comparisons of operational energy alone may bias the results in favour of TPIC, which do not include the burden of fluid [86].

Regulatory uncertainty for PFAS is independent of thermal performance and strengthens the case for compliant TPIC fluids.

10. Long-Term Reliability and Deployment Readiness

Thermal capability is well proven at chip-to-server scale, but only multi-year time series of data exist on fluid and hardware survival. Short laboratory tests do not capture the slow failure modes of materials, fluids and seals.

10.1 Material Compatibility

Immersion means continuous contact with the coolant of PCBs, connectors, solder, TIMs, cable insulation, polymers, adhesives, seals and elastomers. Failure modes include swelling or extraction of plasticiser, TIM degradation, solder/intermetallic attack, coating removal and corrosion, which can affect mechanical, electrical, or thermal properties over months to years [46].

10.2 Coolant Ageing and Contamination

Thermal cycling, oxidation, moisture absorption, contamination, dissolved material and interaction with hardware [42, 45, 46] can change coolant properties.

Dielectric strength, moisture, acid number, viscosity, particles and dissolved metals are the major ageing indicators. Oxidation can increase viscosity and acidity, moisture can lead to a lowering of the dielectric strength, deposits can contaminate surfaces, and local thermal decomposition can lead to corrosive products.

10.3 Available Long-Duration Evidence and Its Limits

Long-duration testing should not only validate the thermal performance, but also provide evidence of the stability of the fluid and hardware under realistic service conditions.

The five-month mineral oil and virgin coconut oil exposure study showed no damage to the components, but it is not generalizable due to the duration and fluid set used [87] (Table 12).

Table 12 Evidence of immersion cooling from the reviewed corpus: test duration, type of fluid, hardware that was immersed, quantities that were monitored, reported outcome, limitation that does not allow for generalization.

Study type	Duration	Fluid(s)	Hardware exposed	Quantities monitored	Reported outcome	Limitation preventing generalisation
Extended SPIC exposure trial	5 months	Mineral oil; virgin coconut oil	Complete server hardware	Visual and functional condition of components; thermal performance	No component damage observed over the test period	Longest exposure in the corpus, yet an order of magnitude shorter than the service life; two fluids only; fluid properties not tracked over time
Comparative coolant tests with TIM interaction	Hours to days	Several dielectric fluids	Processor-TIM-coolant interface	Thermal behaviour attributable to fluid-material interaction	Interaction with TIM materials measurably affects thermal behaviour	Demonstrates the mechanism but not its rate; no ageing trajectory established
Dielectric-property characterisation	Single point	Fluorinated and hydrocarbon fluids	Fluid only	Dielectric strength and thermophysical properties of virgin fluid	Virgin-fluid properties characterised	Properties of the virgin fluid are not those of the fluid in service; no exposure history
Boiling-surface durability	Hours	HFE fluids	Engineered boiling surfaces	Boiling heat transfer and CHF on pristine surfaces	Substantial enhancement on pristine surfaces	No data on fouling, degradation-product deposition or treatment durability over months
Facility-scale operational monitoring	-	-	-	-	No study in the corpus reports audited multi-year operational data for an immersion facility	This is the single largest gap between the thermal evidence and the deployment evidence

Notes: Across the reviewed corpus, no long-term exposure reports the quantities that determine long-term viability such as evolution of dielectric strength, moisture content, acid number, viscosity drift, particulate and dissolved-metal loading, elastomer and adhesive properties, or integrity of solder joints. TIM = thermal interface material, CHF = critical heat flux.

The longest exposure found was five months, whereas the hardware service lives are multi-year. Few reports describe the evolution of dielectric strength, acid number, elastomer properties and in-service fluid condition for multi-year operation.

10.4 Hydraulic, Containment and Scale-Up Constraints

The distribution of flow is influenced by forced SPIC by the presence or absence of adequate coolant delivery, inlet/outlet arrangement, board orientation, spacing, heat sinks and obstructions which produce the flow field [48, 88].

If distribution is poor, the capacity of the pump can be increased to a certain degree, but it can also increase auxiliary energy consumption; in general, geometric flow management is better.

Stable nucleation, vapour removal, condensation and liquid return must be maintained in TPIC. The effect of the interaction between bubbles and the accumulation of vapour increases as the load increases [39, 40, 89, 90].

The condenser capacity also depends on the vapour generator capacity and cannot be determined solely by the size of the boiling surface.

Scale-up coolant distribution, tank volume, structural loading, heat exchanger, secondary loops and redundancy and service access [33, 36, 88].

The extent of evidence at multi-server and facility scale is considerably less advanced than evidence at the component level.

Serviceability is also undervalued: draining, drip-off, cleaning, lifting and, in the case of TPIC, fluid loss affect maintenance and total ownership cost.

The key deployment issues for SPIC are flow distribution and pumping energy, whereas stable boiling, vapour handling, condensation and containment are the key issues for TPIC. Both need to be reliable in heat rejection, monitoring, maintenance and redundancy.

11. Research Gaps, Agenda and Reporting Requirements

11.1 Where the Evidence Is and Is Not Sufficient

There is a highly imbalanced evidence base for each topic and spatial scale. Thermal performance and enhancement are relatively well developed at the chip, package, and server scale, and partial to well developed at a small scale but sparse at a tank scale and absent at a facility scale, with regard to coolant thermophysics. Energy efficiency, facility coupling and techno-economics are partial at tank and facility scale, and almost non-existent at component scale due to the lack of energy reporting in studies at this scale. Only a few of the scales reviewed include indications for material compatibility, multi-year coolant ageing, or standardized test protocols. The bottom line is that the field has developed a good quantity of evidence on the effectiveness of immersion cooling for heat removal, but much less evidence on the performance of an immersion installation over several years. The latter would entail long-duration exposure data, property evolution of the coolant and materials, and standardisation of test protocols. The gaps are summarized in Table 12 and converted to specific research questions and needed approaches in Table 13.

Table 13 Research gaps identified in this review, sorted by level of observation, the question to be answered, the experimental or analytical method needed, and priority assigned.

Scale	Gap	Question to be answered	Approach required	Priority
Chip	Transient response to realistic load profiles	How do SPIC and TPIC respond to the synchronised power transients characteristic of AI training?	Programmable multi-heater test articles driven by recorded workload traces; transient junction-temperature and, for TPIC, CHF-margin measurement	High
Chip	CHF margin under non-uniform heating	How much CHF margin remains at current accelerator hotspot fluxes on realistic surfaces?	CHF measurement on multi-source, non-uniformly heated articles with aged rather than pristine surfaces	High
Package/ heat sink	Enhancement performance on aged surfaces	Do the gains from structured, porous and wettability-modified surfaces survive fouling and thermal cycling?	Accelerated ageing followed by repeat boiling characterisation on the same articles	High
Server/ board	Component arrangement as a thermal variable	How should heat sources be arranged on an immersed board, given coolant pre-heating along the flow path?	Parametric experimental and numerical studies of board layout, inter-board spacing and obstruction at constant total power	High
Server/ board	Auxiliary cost of enhancement	What is the pumping or actuator power associated with each reported thermal gain?	Adoption of the reporting requirements of Section 11.3; re-measurement of Δp and pump power in existing configurations	High
Tank/rack	Server-to-rack scaling	How do server-level results transfer to a multi-server tank sharing coolant and heat rejection?	Scaling studies using identical hardware and instrumentation at both scales	High
Tank/rack	Serviceability and mean time to repair	What is the maintenance time, fluid loss and cost per intervention for each architecture?	Instrumented maintenance trials on operating tanks, reporting fluid loss per intervention	Medium

Facility	Audited facility-scale performance	What energy, water and cost performance does a full immersion facility actually deliver?	Facility demonstrators with audited metering on a declared boundary over at least one full design year	High
Facility	Water consumption and WUE	How does immersion cooling with dry heat rejection affect water usage effectiveness by climate?	Annual simulation and measurement of WUE alongside PUE for matched configurations	Medium
All	Multi-year fluid and material ageing	How do dielectric strength, viscosity, acidity, contamination and elastomer properties evolve over years of service?	Multi-year exposure campaigns with scheduled fluid analysis and periodic hardware extraction for metallurgical and electrical examination	Critical
All	Standardised test protocol	How can results from different laboratories be made comparable?	Round-robin protocol specifying fluid, surface preparation, heater geometry, load profile, instrumentation and uncertainty reporting	Critical
All	PFAS-compliant boiling fluids	Can fluids outside the PFAS class provide the saturation temperature, latent heat and dielectric properties that TPIC requires?	Fluid development coupled with boiling and material-compatibility characterisation	Critical

Notes: This is a measure of how much closing the gap is a requirement for deployment instead of a small improvement in performance. CHF = critical heat flux; WUE = water usage effectiveness.

11.2 Prioritised Research Agenda

Near-term priorities are simple but consequential: report IT load, Auxiliary power, and System boundary for each energy metric; report full coolant properties at operating temperature; report pressure drop or actuator power for each enhancement result.

Medium-term priorities are a round-robin immersion test protocol, matched server-to-rack scaling studies and representative AI/HPC transient load profiles.

Long-term priorities are multi-year fluid-hardware ageing campaigns, PFAS-compliant boiling fluids and facility demonstrators with audited energy and water and cost accounting.

11.3 Minimum Reporting Requirements for Future Studies

The minimum reporting standard should include: coolant and properties, surface preparation, heat load and flux basis, flow conditions, temperature measurements, pressure drop and auxiliary powers, spatial scale, source count, duration and uncertainty. In addition, energy studies should include IT energy loads, all auxiliary energy loads, energy rejected, and the complete system boundary.

12. Conclusions

The 90 studies included in this review were selected from the published literature and span 2015 to 2026, following a documented search protocol, a five-level scale framework, a 12-criterion SPIC/TPIC comparison and harmonised quantitative tables. The main conclusions are dependent on the scale.

The air-side limitation is eliminated at the chip/package scale. Forced SPIC ($\approx 0.4\text{--}2 \text{ kW m}^{-2} \text{ K}^{-1}$) and plain-surface TPIC ($\approx 1.5\text{--}12 \text{ kW m}^{-2} \text{ K}^{-1}$) cover the heat-flux range of current AI/HPC accelerators, while jet SPIC and engineered boiling surfaces reach the high limits. Thus, the choice of architecture is dictated by CHF margin, vapour handling, containment, and operating complexity as well as h.

At the server/board level, the most important variable is the heat-to-coolant balance. Beyond increasing total flow, these factors can achieve better thermal uniformity more efficiently: flow redistribution, inlet/outlet design, fin arrangement, and component arrangement.

In the case of tank/facility scale, the boundaries of PUE and COP need to be clear. Energy, heat rejection, climate, cost, water, and IT load need to be evaluated as a single system because the reported values for these quantities will vary depending on the accounting boundary.

Overall, SPIC is the more predictable and easier path to deployment, and has fully developed hydraulic design, as well as a wider base of server/facility evidence. TPIC provides higher high-flux, near-isothermal boiling, and introduces CHF, vapour, condenser, containment, fluid-cost/regulatory constraints.

The biggest remaining gap is deployment evidence - multi-year ageing, material compatibility, standardised testing, operation at the facility level, with audited results. What's needed right now is not more one-off, peak performance showcases, but disciplined, similar reporting and long-duration validation.

Nomenclature

<i>A</i>	effective heat-transfer area (m^2)
<i>B1</i>	Server/board system boundary
<i>B2</i>	Primary cooling-loop system boundary
<i>B3</i>	Tank/rack system boundary
<i>B4</i>	Facility system boundary
<i>c_p</i>	specific heat capacity ($\text{J kg}^{-1} \text{ K}^{-1}$)
<i>COP</i>	coefficient of performance (-)
<i>h</i>	heat-transfer coefficient ($\text{W m}^{-2} \text{ K}^{-1}$)
<i>h_{fg}</i>	latent heat of vaporisation (J kg^{-1})
<i>k</i>	thermal conductivity ($\text{W m}^{-1} \text{ K}^{-1}$)

$\dot{m}$	coolant mass-flow rate (kg s^{-1})
P	electrical power (W)
Q	heat-removal rate (W)
q''	surface heat flux (W cm^{-2})
R_{th}	junction-to-coolant thermal resistance (K W^{-1})
T	temperature ($^{\circ}\text{C}$ or K)
T_j	junction (or representative surface) temperature ($^{\circ}\text{C}$)
T_{sat}	saturation temperature (K)
x	vapour quality (-)
dp	pressure drop (Pa)
dT	temperature difference (K)
μ	dynamic viscosity ($\text{Pa}\cdot\text{s}$)
ρ	density (kg m^{-3})
σ	surface tension (N m^{-1})

Abbreviations

<i>CHF</i>	critical heat flux
<i>CDU</i>	coolant distribution unit
<i>CRAH</i>	computer-room air handler
<i>GWP</i>	global warming potential
<i>HPC</i>	high-performance computing
<i>LCA</i>	life-cycle assessment
<i>ONB</i>	onset of nucleate boiling
<i>PFAS</i>	per- and polyfluoroalkyl substances
<i>pPUE</i>	partial power usage effectiveness
<i>PUE</i>	power usage effectiveness
<i>SPIC</i>	single-phase immersion cooling
<i>TIM</i>	thermal interface material
<i>TPIC</i>	two-phase immersion cooling
<i>WUE</i>	water usage effectiveness

Author Contributions

Ghassan Adnan Abd: conceptualization of the review framework, literature search and screening, data extraction and harmonisation, preparation of the figures and tables, and writing of the original draft. Ahmed J. Shkarah: methodology of the search protocol and quality appraisal, verification of the extracted quantitative data, and review and editing of the manuscript. Mustafa M. Mansour: conceptualization, supervision and project administration, critical analysis and interpretation of the comparative findings, and review and editing of the final manuscript. All authors have read and approved the final version of the manuscript.

Image Processing Statement

All figures in this article (Figures 1-5) are original graphics prepared by the authors; no photographs, micrographs or third-party images were used, reproduced or adapted. All five figures were produced programmatically in Python using the Matplotlib library: the schematic diagrams (Figure 1, Figure 2 and Figure 5(a)) were built from Matplotlib shape patches, arrows and text annotations, and the plots (Figure 3, Figure 4 and Figure 5(b)) were generated with Matplotlib line, bar, area-fill and scatter functions; Figure 4(b) and Figure 5(b) are illustrative examples, as stated in the figures. Figures 3 and 4(a) are plotted directly from data files containing the values given in Tables 3, 4, 7 and 9 and the air-side limit stated in Section 1.1; quantities derived from these values (the hotspot band, mid-points of reported ranges, upper bounds and the conversion of boiling heat-transfer gains) follow rules stated in the figures and the data files. The box widths in Figure 1 are qualitative, indicative shares of the junction-to-ambient resistance rather than computed values. The layouts of all five figures, and the indicative or illustrative values in Figures 1, 4(b) and 5(b), are based on earlier versions of the figures, which had been generated with Claude (Anthropic) from the authors' descriptions and have now been replaced. The Python scripts and data files, which regenerate all five figures, were prepared with the assistance of Claude and are available from the corresponding author. Each figure was exported from Matplotlib as a PNG image at 600 dpi and at the text width, and was inserted into the manuscript at that size without cropping. No part of any image was modified after export: no element was retouched, cloned, erased, added or moved, and no brightness, contrast or colour adjustment was applied to the whole or any part of any figure.

Competing Interests

The authors have declared that no competing interests exist.

AI-Assisted Technologies Statement

Artificial intelligence (AI) tools were used in the preparation of this manuscript for two purposes. OpenAI's ChatGPT was employed for basic grammar correction and language refinement, to improve the readability and linguistic clarity of the English text. Claude (Anthropic) was used to assist in preparing the Python scripts and data files from which all five figures are produced and in drafting the revised figure captions and figure-related statements; earlier versions of the figures that had been generated with Claude have been replaced (see the Image Processing Statement). All scientific content, data interpretation, and conclusions were developed independently by the authors. The authors have thoroughly reviewed and edited the AI-assisted text to ensure its accuracy and accept full responsibility for the content of the manuscript.

References

1. Zhang C, Sun X, Han Z, Li X, Dong J. Energy saving potential analysis of two-phase immersion cooling system with multi-mode condenser. *Appl Therm Eng.* 2023; 219: 119614.
2. Shrigondekar H, Lin YC, Wang CC. Investigations on performance of single-phase immersion cooling system. *Int J Heat Mass Transfer.* 2023; 206: 123961.

3. Taddeo P, Romaní J, Summers J, Gustafsson J, Martorell I, Salom J. Experimental and numerical analysis of the thermal behaviour of a single-phase immersion-cooled data centre. *Appl Therm Eng.* 2023; 234: 121260.
4. Cai S, Gou Z. Towards energy-efficient data centers: A comprehensive review of passive and active cooling strategies. *Energy Built Environ.* 2024; 7: 206-226.
5. He W, Li C, Xu C, Yang Y, Liu S, Cao B, et al. Performance of air-liquid integrated cold plate heat sinks in servers. *Appl Therm Eng.* 2025; 281: 128734.
6. Al Kez D, Foley AM, Wong FW, 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: 104511.
7. Omrani M, Ghassemi M. AI-driven optimization of fan-wall cooling system in a medium-density data center. *Int J Heat Mass Transfer.* 2025; 247: 127159.
8. Yang Z, Wang Y, Yuan X, Wang X, Yang X, Li Q. Design and optimization of a staggered counter-current microchannel liquid-cooled plate. *Appl Therm Eng.* 2025; 279: 127891.
9. Sun X, Han Z, Li X. Simulation study on cooling effect of two-phase liquid-immersion cabinet in data center. *Appl Therm Eng.* 2022; 207: 118142.
10. Zhou K, Yu X, Xie B, Xie H, Fu W. Immersion cooling technology development status of data center. *Sci Technol Energy Transit.* 2024; 79: 41.
11. Zheng S, Su C, Yang X, Zhang Y, Duan K, Zhang Y, et al. A comprehensive review of single-phase immersion cooling in data centres. *Appl Therm Eng.* 2025; 272: 126385.
12. 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: 116786.
13. Xue ZH, Ai BC, Qu W. Data center energy conservation study utilizing loop heat pipes as a chip-level cooling technique and its industrial application. *Appl Therm Eng.* 2023; 220: 119715.
14. Shah D, Pandya N, Kim SK, Mistofsky K, Eamons A. Thermal impact of ultra-narrow DIMM pitch and air-cooled heat pipe solution on optimizing DDR5 cooling in high density HPC systems. *Therm Sci Eng Prog.* 2025; 68: 104206.
15. Li C, Li J. Thermal characteristics of a flat plate pulsating heat pipe module for onsite cooling of high power server CPUs. *Therm Sci Eng Prog.* 2023; 37: 101542.
16. Khoshvaght-Aliabadi M, Ghodrati P, Shin JY, Kang YT. Impact of coolant distribution design on server-level thermal management in data centers. *Energy.* 2025; 330: 136936.
17. Khoshvaght-Aliabadi M, Ghodrati P, Nasrolahzadeh A, Kang YT. Optimization of heat sinks for data center server CPUs cooled via single-phase immersion cooling. *Appl Therm Eng.* 2025; 280: 127790.
18. Cheng SW, Zhang YD, Chien LH, Wang CC. A comparative study of a single-phase immersion-cooled server with a pin-fin heat sink for mitigation of the flow bypass effect. *Processes.* 2026; 14: 2209.
19. Khoshvaght-Aliabadi M, Nasrolahzadeh A, Ghodrati P, Kang YT. Oblique-fin heat sink design for uniform CPU cooling in immersion-cooled servers. *Appl Therm Eng.* 2025; 288: 129628.
20. Khoshvaght-Aliabadi M, Hojjati F, Hassani S. Influence of component arrangement on thermal management in immersion-cooled server boards. *Energy.* 2026; 347: 140323.
21. Liu P, Kandasamy R, Ho JY, Wong TN, Toh KC. Dynamic performance analysis and thermal modelling of a novel two-phase spray cooled rack system for data center cooling. *Energy.* 2023; 269: 126835.

22. Cho J. Numerical analysis of rack-based data center cooling with rear door heat exchanger (RDHx): Interrelationship between thermal performance and energy efficiency. *Case Stud Therm Eng.* 2024; 63: 105247.
23. Heydari A, Gharaibeh AR, Tradat M, Manaserh Y, Radmard V, Eslami B, et al. Experimental evaluation of direct-to-chip cold plate liquid cooling for high-heat-density data centers. *Appl Therm Eng.* 2024; 239: 122122.
24. Hu H, Chen C, Li C, Pan M. Experimental investigation of roll bond liquid cooling plates for server chip heat dissipation. *Appl Therm Eng.* 2023; 226: 120284.
25. Samal SK, Chang HC, Fulpagare Y, Wang CC. Thermal management of data centers: Chip-scale cooling using novel distributed inlet-outlet jet impingement liquid cold plate. *Appl Therm Eng.* 2025; 271: 126360.
26. Jiao J, Duan F, Gong L. Multi-objective optimal design of hybrid liquid cooling system with manifold microchannels for data center server thermal management. *Energy.* 2025; 337: 138648.
27. Cho J, Moon JH. Numerical coupling of energy efficiency and thermal performance for cold plate cooling optimization in high-density compute AI data centers. *Energy Build.* 2025; 348: 116441.
28. Li J, Luo X, Wang M, Chen C, Yang C, Zhang Z, et al. Study of flow and heat transfer characteristics of tandem cold plates for data center cooling. *Case Stud Therm Eng.* 2025; 73: 106590.
29. Qu S, Wang C, Liu Y, Feng Y, Han Y, Xing Z. System design, control strategy development and performance comparison analysis for a novel air-liquid hybrid data center cooling system based on cold plates. *Int J Refrig.* 2025; 180: 222-239.
30. 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.
31. 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.
32. Liu Z, Sun X, Ji S, Yuan K. A multi-objective optimization of a single-phase immersion cooling system at cabinet level. *Energy Build.* 2024; 312: 114221.
33. Kim J, Choi H, Lee S, Lee H. Computational study of single-phase immersion cooling for high-energy density server rack for data centers. *Appl Therm Eng.* 2025; 264: 125476.
34. Wang J, Yuan C, Li Y, Li C, Wang Y, Wei X. Direct-to-chip immersion liquid cooling for high-power vertical-cavity surface-emitting laser (VCSEL). *Appl Therm Eng.* 2025; 269: 126137.
35. Sun X, Liu Z, Ji S, Yuan K. Experimental study on thermal performance of a single-phase immersion cooling unit for high-density computing power data center. *Int J Heat Fluid Flow.* 2025; 112: 109735.
36. Lionello M, Rampazzo M, Beghi A, Varagnolo D, Vesterlund M. Graph-based modelling and simulation of liquid immersion cooling systems. *Energy.* 2020; 207: 118238.
37. Gheni M, Kerskes H, Stergiaropoulos K. Operational analysis of the cooling system in a direct liquid-cooled data center: A measurement and simulation study on the impact of supply water temperature. *Appl Energy.* 2026; 403: 127061.
38. Liu C, Huang Y, Zhang C. Efficient jet-assisted single-phase immersion liquid cooling for high heat-flux servers. *Appl Therm Eng.* 2025; 259: 124935.

39. Krishnadasan VB, Suresh P, Balaji C. Experimental investigations on the performance and bubble dynamics of two-phase immersion cooling system with multiple chips. *Int J Heat Mass Transf.* 2025; 245: 126977.
40. 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: 126575.
41. Huang Y, Liu B, Xu S, Bao C, Zhong Y, Zhang C. Experimental study on the immersion liquid cooling performance of high-power data center servers. *Energy.* 2024; 297: 131195.
42. Chen X, Huang Y, Xu S, Bao C, Zhong Y, Chen Y, et al. Thermal performance evaluation of electronic fluorinated liquids for single-phase immersion liquid cooling. *Int J Heat Mass Transf.* 2024; 220: 124951.
43. Huang Y, Liu C, Zhong Y, Zhang C. Experimental study on jet-enhanced immersion liquid cooling for energy-efficient data centers. *Energy.* 2025; 334: 137584.
44. Azarifar M, Arik M. Liquid synthetic jets for high flux electronics cooling. *Appl Therm Eng.* 2025; 261: 125007.
45. Li X, Guo S, Sun H, Liu S, Wang X, Wang L. Assessment of coolants used single-phase immersion cooling data center. *Energy.* 2025; 340: 139240.
46. Ruichek A, Hnayno M, Chehade A, Fardoun F, Moussa T, Polidori G, et al. Immersion cooling fluid comparison using a single-phase immersion/liquid technique for data centers. *Appl Therm Eng.* 2025; 280: 128355.
47. Koucheh AB, Sharbati P, Ünlü C, Koşar A, Sadaghiani A. Comprehensive evaluation of battery cooling mechanisms including two-phase immersion with 3M™ NOVEC™-7000 and 7100. *Energy Convers Manage.* 2026; 348: 120780.
48. Rongyang Y, Yu Z, Liang R, Su W, Wei J. Numerical investigation on thermal performance of single-phase immersion cooling systems using oil coolant. *Front Heat Mass Transfer.* 2025; 23: 279-298.
49. Luo Q, Wang C, Wen H, Liu L. Research and optimization of thermophysical properties of sic oil-based nanofluids for data center immersion cooling. *Int Commun Heat Mass Transf.* 2022; 131: 105863.
50. Alkasmoul FS, Almogbel AM, Shahzad MW, Al-Damook AJ. Thermal performance and optimum concentration of different nanofluids in immersion cooling in data center servers. *Results Eng.* 2025; 25: 103699.
51. Gonzalez RC, Diaz AJ, Rojas D, Sarmiento-Laurel C. Assessing solar absorption chillers for data center cooling across diverse climatic conditions. *Energy Build.* 2025; 331: 115376.
52. Muneeshwaran M, Lin YC, Wang CC. Performance analysis of single-phase immersion cooling system of data center using FC-40 dielectric fluid. *Int Commun Heat Mass Transf.* 2023; 145: 106843.
53. Kalantarpour R, Vafai K. Investigating the two-phase immersion cooling performance by surface wettability modification. *Int J Heat Mass Transf.* 2026; 255: 127796.
54. Zhou G, Zhou J, Huai X, Zhou F, Jiang Y. A two-phase liquid immersion cooling strategy utilizing vapor chamber heat spreader for data center servers. *Appl Therm Eng.* 2022; 210: 118289.
55. Ding B, Xia F, He L, Huang Y. A novel concept for performance enhancement of immersion-cooled data center servers. *Int Commun Heat Mass Transf.* 2024; 159: 107980.

56. Suresh P, Chang KW, Lua KB, Wang CC. Deep reinforcement learning for energy-efficient thermal management in 2U air-cooled server systems. *Appl Energy*. 2026; 404: 127168.
57. Ni D, Yu F. CFD simulation study of flow equalisation plate model in single-phase immersion liquid cooling for servers. *Therm Sci Eng Prog*. 2024; 47: 102268.
58. Wang H, Yuan X, Zhang K, Lang X, Chen H, Yu H, et al. Performance evaluation and optimization of data center servers using single-phase immersion cooling. *Int J Heat Mass Transf*. 2024; 221: 125057.
59. Ozer RA. Heat sink optimization with response surface methodology for single phase immersion cooling. *Int J Heat Fluid Flow*. 2025; 112: 109745.
60. Li X, Xu Z, Liu S, Zhang X, Sun H. Server performance optimization for single-phase immersion cooling data center. *Appl Therm Eng*. 2023; 224: 120080.
61. Pandey V, Liang SS, Law LL, Kuppusamy NR, Ching-Chiuan Y, Lee PS, et al. Experimental evaluation of 3D printed topology optimized liquid and hybrid-cooled heat sinks for data centers. *Int J Heat Mass Transf*. 2026; 256: 128106.
62. Xu Y, Yang Y, Tian M, Li D, Deng T, Hu K, et al. An investigation into the effects of various IT scheduling algorithms on thermal control and energy efficiency in liquid-cooled data centers. *J Build Eng*. 2025; 107: 112798.
63. Huang Y, Liu C, Ding B, Deng Z, Gao W. Machine learning-optimized jet-enhanced immersion liquid cooling for high-power data centers. *Int Commun Heat Mass Transf*. 2025; 169: 109654.
64. Jin P, Zhang X, Zhang X, Tan Z, Xie G. Immersed jet impingement heat transfer for servers in immersion cooling data center: A parametric study. *Appl Therm Eng*. 2025; 274: 126628.
65. He W, Liu C, Du L, Yang Y, Liu S, Zeng M, et al. Structural optimization on vertical jet impingement cooler used for sever heat dissipation in data centers. *Appl Therm Eng*. 2025; 280: 128432.
66. Jin P, Yang J, Zhang X, Zhang X, Xie G. Heat transfer enhancement for high power chips in immersion cooling by ducted propellers. *Appl Therm Eng*. 2025; 273: 126469.
67. Charaf H, Maalouly M, Chehade A, Al Moussawi H, Moussa T, Polidori G, et al. Passive hydraulic cooling: A CDU-free solution for data center efficiency. *Appl Therm Eng*. 2025; 280: 128116.
68. Wang L, Chen L, Wang X, Zheng H, Chen X, Li Z. Design and experimental study of a gravity-driven two-phase loop system for multi-elevation heat source cooling. *J Energy Storage*. 2026; 142: 119517.
69. Zhang Y, Gao Y, Wang Y, Ma X, Yang X, Wei J. Immersion phase-change thermal management system coupled with loop thermosyphons for utilization in data centers cooling. *Energy Convers Manage*. 2026; 348: 120688.
70. Zou S, Zhang J, Wang Y, Xiao T, Huang J. Multi-objective optimization of server component temperatures and coolant filling ratio in single-phase immersion tank based on Taguchi grey correlation analysis. *Case Stud Therm Eng*. 2025; 72: 106363.
71. Liu HM, Chang Z. Multi-objective optimization of temperature uniformity in the immersion liquid cooling cabinet with Taguchi-based grey relational analysis. *Int Commun Heat Mass Transf*. 2024; 154: 107395.
72. Qin X, Shen A, Duan H, Liu H, Wang D, Liu H, et al. Investigating 4E analyses and multi-objective optimization of a novel data center compression/ejection transcritical CO₂ cooling system. *Appl Therm Eng*. 2025; 264: 125477.

73. Ma Y, Bao Y, Li J. Heat transfer dependence of power usage effectiveness of an augmented two-phase immersion cooling system for high-power servers. *Energy*. 2025; 323: 135853.
74. Liu X, Li M, Wang F, Zhang H, Wang H, Wen B, et al. Thermal performance analysis of single-phase immersion cooling for power device: An experimental study. *Appl Therm Eng*. 2025; 284: 129142.
75. He W, Zhang J, Li H, Liu S, Wang Y, Lv B, et al. Optimal thermal management of server cooling system based cooling tower under different ambient temperatures. *Appl Therm Eng*. 2022; 207: 118176.
76. Liu S, Guo S, Sun H, Xu Z, Li X, Wang X. Evaluation of energy, economic, and pollution emission for single-phase immersion cooling data center with different economizers. *Appl Therm Eng*. 2025; 260: 125049.
77. Wang C, Yu H, Xian L, Liu G, Chen L, Tao WQ. Study on the operation strategy of the water cooling system in the single-phase immersion cooling data center considering cooling tower performance. *Energy Build*. 2024; 324: 114890.
78. Huang G, Xie M, Zhao Y, Jia T, Dai Y. Experimental study on waste heat recovery process in a water-cooled data center for dehumidification application using desiccant-coated heat exchangers. *Energy Build*. 2025; 346: 116196.
79. Fu C, Zhang W, Zhou X, Shen Q, Wu T. Stochastic optimization of photovoltaic-integrated data centers with hybrid cooling and waste heat recovery for district energy supply. *Renew Energy*. 2025; 256: 124248.
80. Liu C, Hao N, Zhang T, Wang D, Li Z, Bian W. Optimization of data-center immersion cooling using liquid air energy storage. *J Energy Storage*. 2024; 90: 111806.
81. Oh H, Jin W, Peng P, Winick JA, Sickinger D, Sartor D, et al. Techno-economic performance of reservoir thermal energy storage for data center cooling system. *Appl Energy*. 2025; 391: 125858.
82. Li F, Li P, Ding R, Su W, Ou S, Xing L, et al. Thermo-economic analysis on trans-critical compressed CO₂ energy storage system integrated with the waste heat of liquid-cooled data center. *J Energy Storage*. 2024; 103: 114292.
83. Ren H, Li Y, Lin W, Sun Y. Energy and economic analysis of photovoltaic thermal night cooling and adaptive reset control for ultra-efficient data center immersion cooling. *Renew Energy*. 2025; 255: 123821.
84. Du S, Zhang Q, Zou S, Meng F, Liu L. Simulation analysis on energy consumption and economy of CPU cooling system based on loop heat pipe for data center. *Therm Sci Eng Prog*. 2023; 45: 102115.
85. Kanbur BB, Wu C, Fan S, Tong W, Duan F. Two-phase liquid-immersion data center cooling system: Experimental performance and thermoeconomic analysis. *Int J Refrig*. 2020; 118: 290-301.
86. Zhang M, Carbajales-Dale M, Ma X, Guo L, Fan C. Cleaner grid or smarter cooling? Environmental impact trade-offs of a data center using the life cycle assessment method. *Clean Energy Syst*. 2025; 12: 100223.
87. Pambudi NA, Yusuf AM, Sarifudin A. The use of single-phase immersion cooling by using two types of dielectric fluid for data center energy savings. *Energy Eng*. 2021; 119: 275-286.

88. Liu S, Xu Z, Wang Z, Li X, Sun H, Zhang X, et al. Optimization and comprehensive evaluation of liquid cooling tank for single-phase immersion cooling data center. *Appl Therm Eng.* 2024; 245: 122864.
89. He Z, Smith TW, Lei Z, Taleghani AD. Repurposing abandoned mine lands for cooling data centers. *Resour Conserv Recycl.* 2026; 225: 108640.
90. Ejaz F, Kwon B. Two-phase active immersion cooling for vertically mounted electronics with interchip component-assisted bubble departure. *Int Commun Heat Mass Transfer.* 2024; 159: 107981.