

**OPEN**
Compute Project

WHITEPAPER THERMAL BOUNDARIES IN AI INFRASTRUCTURE

*Matching cooling technologies to compute through the power
inflection*

Authors:

Rolf Brink (Promersion), Mohan J. Kumar (Oracle), Steve Mills (Meta), Kelley Mullick (Avayla)

Version History

Date	Version #	Author	Description
August 27, 2026	0.98	Rolf Brink	Public Review Version
September 30, 2026	0.99	Rolf Brink	Community Review Incorporated
October 7, 2026	0.995	Rolf Brink	CE Leadership Review Incorporated
October 12, 2026	1.00	Rolf Brink	Published Version

Acknowledgments

This paper was written to be checked, and the community checked it. The authors thank the reviewers of the public comment round, whose questions hardened the definitions, whose challenges moved the figures, and whose contributions are visible throughout this final version: Neil Edmunds, Brian Kinkade, Junming Li, Andrew Miller, Amin Zeinvand Moghaddam, Jussi Myllyluoma, Stephanie Schmidt, Jouk van Veen, Andries Louw Wolthuizen, and Herb Zien. The full review record, each comment with its resolution, remains part of the project archive.

The review also surfaced themes that reach beyond this paper's scope: commissioning as a discipline in its own right, the acceptance of dynamic behavior at the handover, and practices worth carrying across work streams. The authors intend to pursue these within the OCP community, together with the reviewers who raised them.

Abstract

Artificial intelligence has pushed processor power into a regime where cooling, long a supporting utility, now sets the pace at which compute can be deployed. This paper offers a way to reason about the cooling methods competing to meet that shift and to choose among them. It sets out three boundaries that bound every method, of efficiency, of economics and of physics, and a single dynamic that decides when the industry actually adopts a new one. That dynamic, which we develop as the central argument of the paper, is that a new cooling method is taken up only when the current one can no longer do the job, because each new method brings an architectural dependency the industry has good reason to defer. Read through that lens, each technology, air, cold plate and immersion in their single and two-phase forms and emerging microfluidics, occupies a different and shifting envelope, and none answers the whole problem.

The conclusion the evidence forces is a multi-technology roadmap, a prepared and interoperable toolbox whose methods are matched to the heat, the site and the moment, with the supply chains and standards for each made ready before they are needed.

Contents

Version History.....	2
Acknowledgments	2
Abstract.....	3
Contents	4
1. Introduction	7
1.1. Cooling becomes the constraint.....	7
1.2. Runaway power curves are not new	7
1.3. Why there is no rescue this time	9
1.4. A different kind of transition.....	10
1.5. How the paper is organized	11
2. Open Compute Project Tenets.....	12
2.1. Openness	12
2.2. Efficiency	12
2.3. Impact.....	13
2.4. Scale.....	13
2.5. Sustainability.....	13
3. The physics that sets the boundaries	14
3.1. The governing equation	14
3.2. Why liquids outperform air	14
3.3. The shrinking temperature budget	15
3.4. Flux and total power.....	16
4. The three boundaries	18
4.1. The efficiency boundary	18
4.2. The economic boundary	18
4.3. The physics boundary	18
4.4. The avoidance zone.....	19
4.5. Reading a technology by its boundaries.....	20
5. The adoption law.....	21
5.1. Change is resisted for good reasons.....	21
5.2. Why liquid cooling is an unusually strong case	22
5.3. The chip vendor's dilemma	22
5.4. Who feels the limit and who can act.....	23
5.5. Two clocks.....	23
5.6. Forced or gradual	24

5.7.	The cost of waiting	25
6.	The shape of the heat	27
6.1.	Concentrated heat, and where the move to liquid began	27
6.2.	Distributed heat, the challenge left behind	27
7.	The thermal path.....	29
7.1.	Acquisition, transport and rejection	29
7.2.	The IT envelope and the facility envelope	30
7.3.	A difference in lifecycle	31
7.4.	A common confusion	31
8.	The cooling technologies	32
8.1.	Air	32
8.2.	Single-phase cold plate.....	35
8.3.	Cold plate reliability quantified	37
8.4.	Two-phase cold plate	42
8.5.	Single-phase immersion.....	44
8.6.	Immersion reliability quantified	45
8.7.	Two-phase immersion.....	51
8.8.	Microfluidics	52
9.	The economics of cooling	54
9.1.	The case behind the numbers.....	55
9.2.	Cost in the IT envelope	55
9.3.	Cost in the facility envelope	57
9.4.	Reliability, the swing factor.....	58
9.5.	The comparative picture.....	59
10.	The toolbox, a multi-technology roadmap	61
10.1.	Matching tools to jobs	61
10.2.	A second forcing function, the facility	62
10.3.	Two businesses, one seam.....	62
10.4.	Toward an intelligent cooling subsystem.....	63
10.5.	Preparing the toolbox.....	63
11.	Epilogue: Commissioning the AI datacenter	65
11.1.	Commissioning becomes the bottleneck.....	65
11.2.	From single-vendor startup to multi-party choreography	66
11.3.	What commissioning must verify.....	66
11.4.	The cost of skipping ahead	67

12.	Conclusion	69
Appendix A	Detailed source material for graphs in this paper	71
Appendix B	Thermohydraulic Limits of Single Phase Cold Plate	84
Appendix C	Reader's reference	99
Appendix D	License	103
About Open Compute Foundation.....		104

1. Introduction

1.1. Cooling becomes the constraint

For most of the history of the data center, cooling was a background concern. Designers sized fans and chillers to whatever the silicon produced, confident that the silicon would stay within a familiar envelope. That confidence has run out. The processors driving modern AI now dissipate so much power in so small an area that cooling them at acceptable cost is no longer the hard question. The hard question is whether a given method can cool them at all. Cooling has moved from the supporting cast to the constraint that decides how much compute a site can hold.

1.2. Runaway power curves are not new

A runaway power curve is not new in computing, and twice before the industry has been rescued from one. The first rescue came in the 1980s. The dominant chip logic of the day drew current continuously, whether or not it was doing useful work, so a chip ran hot simply by being switched on, and the largest machines had to be water cooled to survive. The way out was a different kind of transistor, complementary metal-oxide-semiconductor, or CMOS, which had been understood for two decades but not yet used widely for high performance. A CMOS gate draws meaningful current only at the instant it switches state and sits almost idle otherwise, so for the same work it produced far less heat. Its adoption reset the power curve and improved performance at the same time, and the industry took it up readily because it asked nothing of the system around the chip. It was a straight replacement that made the product better and added no new requirement.

Module-level heat flux of high-end processor modules

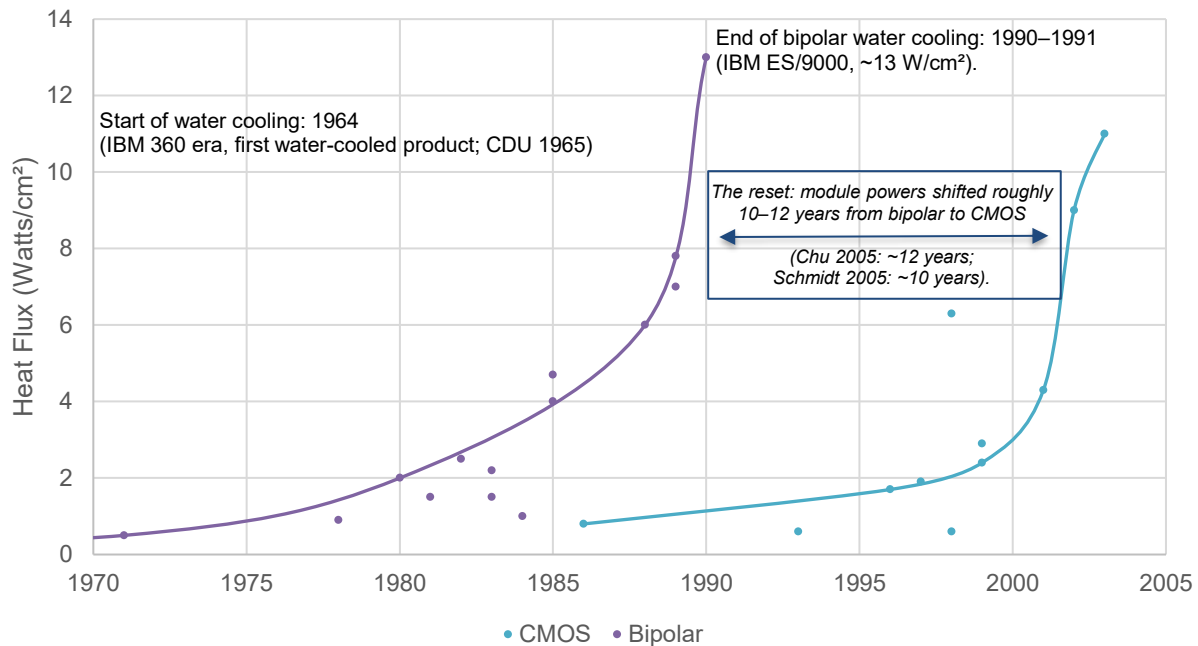


Figure 1 Module-level heat flux of high-end processor modules, bipolar and CMOS eras, W/cm². Sources in A.1.

The second rescue came from architecture instead of a new device. Around 2005 a long-standing rule began to break. For decades, shrinking a transistor had also lowered the power it needed, so each generation could pack in more transistors without raising the power per unit of area. That balance failed once the operating voltage could no longer be reduced without the transistor leaking too much current while idle. Clock speeds, which had risen for years, stalled near a few gigahertz, and chipmakers turned to placing more processing cores side by side to keep performance climbing.

Even that route narrowed. Through the middle 2010s the steady shrinking of features slowed, as transistors approached the size of a handful of atoms and each new manufacturing generation returned less benefit for more cost. With faster clocks and affordable size-reduction efforts both exhausted, the industry had one lever left for more performance: spending more power. From around 2017 the power draw of leading processors began to rise steeply, and it has not stopped since. The lower panel of the figure carries the counterpart record for the same devices, die-level heat flux, which the physics chapter reads closely.

Leading chip TDP and flux trends

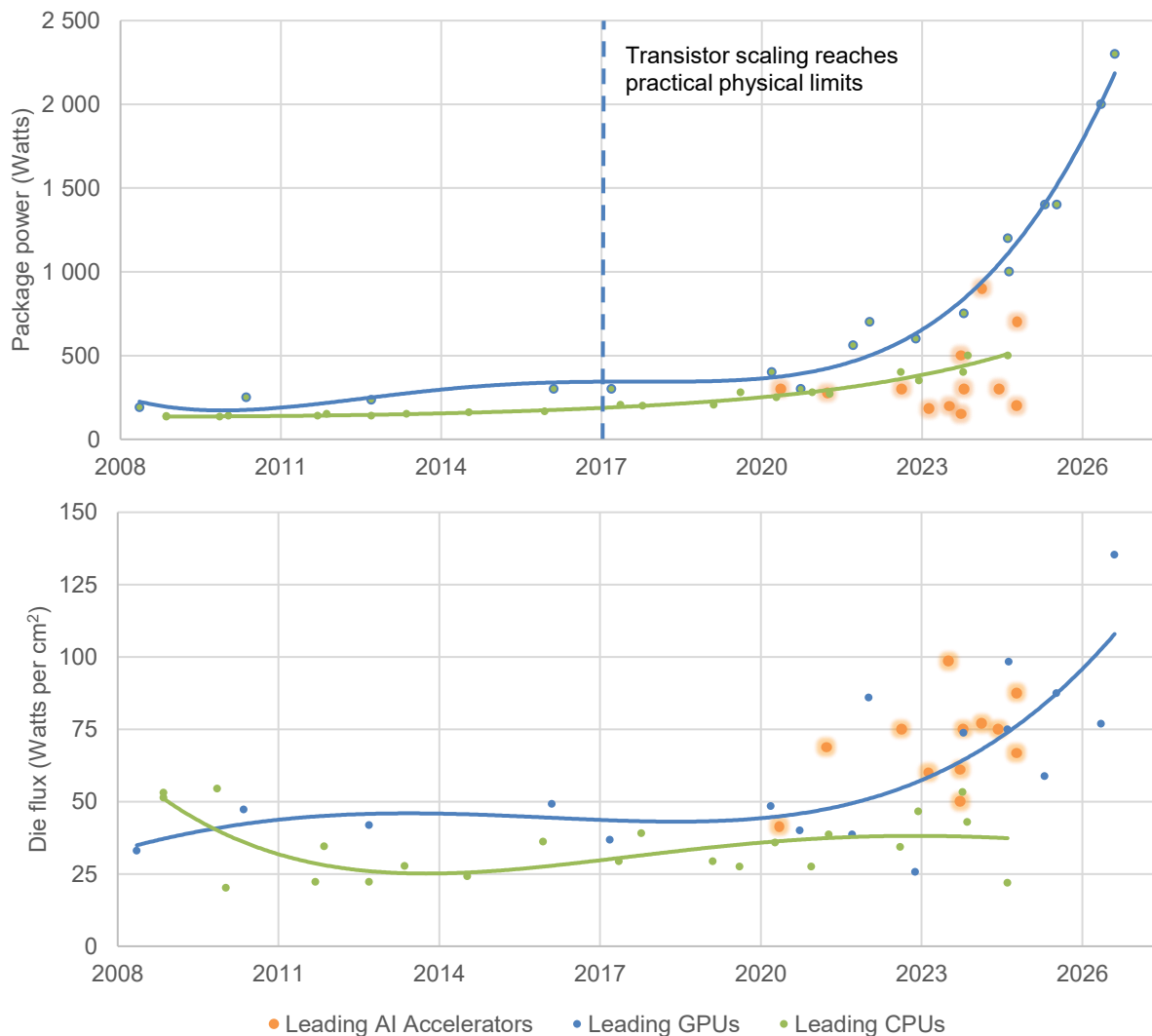


Figure 2 Leading chip package power and die-level heat flux, CPUs, GPUs and AI accelerators, 2008–2026. Sources in A.2.

1.3. Why there is no rescue this time

What is different this time is that no comparable rescue is in sight. Engineers continue to refine the transistor and the way it is built. Newer designs wrap the control gate fully around the channel to cut leakage, route power up through the back of the silicon wafer to reduce losses and assemble a processor from several smaller pieces, called chiplets, in place of one large one. Each of these helps, yet they bend the curve gently downward without resetting it, and the savings they offer are offset by the rapidly rising cost of foundries to build them.

None of the technologies put forward as a successor changes the outlook either. Specialized accelerators are highly efficient, though only for the particular task they are designed for.

Quantum computers tackle a narrow set of problems, need an entirely different way of programming and carry an extreme cooling burden of their own.

Photonic computing performs arithmetic in light and is fast and efficient for one narrow class of work, the linear operations at the heart of neural-network inference, but it hands the data back to electronics for memory, control and the nonlinear steps, so it accelerates part of a workload and leaves the processor that runs the rest in place. The optical links now appearing between chips are a separate and more mature use of the same physics, lowering the energy of moving data without changing what does the computing. Every one of these adds to the conventional processor, and none of them lowers its power.

	<i>Workload scope</i>	<i>Programming model</i>	<i>Maturity</i>	<i>Cooling burden</i>
Specialized accelerator	Narrow, task-specific	Familiar, incremental	Deployed at scale	Conventional, on the roadmap
Quantum processor	Narrow problem classes	Fundamentally different	Early, research-to-pilot	Extreme, often cryogenic
Photonic compute	Linear ops only, inference	Hybrid with electronics	Early commercial	Conventional for the electronics it still needs

Table 1 Proposed successor computing paradigms compared on workload scope, programming model, maturity and cooling burden.

The successors are not hidden. Like CMOS before them, they are visible well in advance. The difference is that none of them, once examined, resets the curve, so seeing them coming offers no rescue.

1.4. A different kind of transition

So the power of mainstream compute keeps climbing with no device in view to bring it down, and the consequence falls on cooling. Here the present moment departs from the earlier ones in a way that matters. The two historical rescues were changes made inside the chip, invisible to the rest of the data center, and they could be adopted without touching anything else.

The cooling transition now under way is not contained inside the chip. Moving from air to liquid raises the power a system can handle, which is genuinely useful, yet it brings new requirements with it: liquid carried into the heart of the IT space, a closer tie between the server and the building, unfamiliar supply chains and new ways of servicing equipment. These dependencies reach into parts of the ecosystem the chip vendor neither owns nor can see, and that reach, more than any single technical hurdle, is what makes the change hard to set in motion.

So the industry has sound reasons to take on each such dependency only when it must, which is why these transitions tend to be deferred until they can no longer be avoided. That pattern of delay, and the price it exacts, is the central thread of this paper.

Every earlier leap in computing improved the chip in place and left the surrounding system untouched. Liquid cooling is the first to reshape the environment the chip lives in.

1.5. How the paper is organized

We begin with the physics that fixes where any method starts to lose efficiency, then set out the three boundaries that describe each method's useful range. We then develop the adoption dynamic that decides when the industry actually moves, the argument that runs through the rest of the paper. With that lens in place we add two framing ideas the technologies depend on, the shape of the heat a system presents and the path that heat travels to leave the building, then examine each cooling method in turn, its physics, its boundaries and its place in the pattern of adoption. The paper closes where the evidence leads, on a multi-technology roadmap and on what the industry must prepare now to be ready for it. Throughout, the framework reads each method at steady state and inside the controlled environment of the data center; how systems behave while moving between operating points is treated in the commissioning discussion that closes the paper, and the uncontrolled edge, where much of the calculus inverts, is deliberately left outside the scope.

2. Open Compute Project Tenets

This document is a whitepaper, not a hardware or software contribution. What it offers the community is a framework: three boundaries that describe every cooling method's useful range, an adoption law that explains when the industry actually moves, and the multi-technology toolbox they point to. The subsections below describe how that offering aligns with each tenet.

2.1. Openness

The framework of this paper is built to be reused and challenged. The three boundaries, the least-disruption law and the toolbox they point to are offered as a shared vocabulary rather than a settled answer. The reliability models are presented with their assumptions, arithmetic and source data in the open (Appendix A, Detailed source material for graphs in this paper), so any reader can rerun, correct or extend them; where ranges remain, in the reliability brackets above all, they are explicit placeholders for the fleet data the community is better positioned to supply than any single contributor. The paper is authored across four companies, licensed under Creative Commons Attribution-ShareAlike 4.0, and closes with a direct invitation to the OCP Open Systems for AI Strategic Initiative to pressure-test the models and carry the standards work forward in the open.

2.2. Efficiency

Efficiency is not an attribute this paper claims for a product. It is the first of the paper's three boundaries and the lens the whole argument runs on. The framework defines the point where cooling stops being a good use of energy, when the power spent moving heat grows into a visible share of the power spent computing, and shows where that point lies for each technology and how it shifts as processor power rises. Read this way, the paper gives operators a method for selecting the most efficient cooling technology for a given heat shape instead of defaulting to the incumbent, and it identifies the levers that matter beyond the chip: compute packed into a constrained power envelope and heat captured warm enough to reuse.

2.3. Impact

The paper aims to change how the industry times and prepares cooling transitions. The least-disruption law explains why transitions arrive forced rather than gradual, and preparation debt names the cost of meeting one unprepared, a bill the move to cold plate has already presented. The practical consequence is a call to keep the supply chains, standards and skills for every method in the toolbox ready before the next forcing function arrives, from the chip or from the facility, so the next transition is a deployment and not a scramble. The framework is offered as the shared vocabulary for exactly the pre-competitive collaboration OCP exists to host, and the epilogue extends the same logic to commissioning, the readiness of the site itself.

2.4. Scale

The analysis is written for hyperscale and works at any size below it. The toolbox it argues for is additive by nature: methods are paired to the heat a system actually presents and extended as density grows, rather than replaced wholesale. The paper locates the real barrier to scale in interfaces, not in any single technology, and argues for open, agreed interfaces at the IT-to-facility boundary, so that one server can be cooled by many facilities, one facility can host many kinds of server, and every method can be manufactured and deployed by many parties in parallel. The commissioning epilogue addresses the remaining condition of scale, bringing multi-technology thermal systems into service reliably at the pace hyperscale construction now demands.

2.5. Sustainability

Cooling choice is treated throughout as a first-order sustainability decision. The paper frames the power, water and land a site can command as a second forcing function alongside the chip, favoring the denser methods and the ones that yield heat warm enough to reuse. Energy enters the framework twice, at the efficiency boundary and again in the total cost of ownership, so wasteful cooling is visible in the same models that guide adoption. The technology chapters are transparent about environmental trade-offs, including the PFAS and regulatory questions attached to two-phase fluids, and the reliability analysis counts the hardware failures that liquid environments can remove. Regulation that rewards efficient, heat-reusing cooling is treated as part of the adoption timeline, not an afterthought.

3. The physics that sets the boundaries

3.1. The governing equation

Every cooling method is judged against the same physics. That judgment begins where heat leaves the package, at the lid or bare die; everything inside, conduction stack and throttling alike, is the silicon's domain and taken here as given. Heat leaves a processor by being carried away in a fluid, whether that fluid is air or a liquid, and the rate at which a fluid carries heat away is governed by a single relationship. The heat removed equals the mass flow of the fluid, multiplied by its heat capacity and by the temperature rise it is allowed to take on.

$$Q = \dot{m} \cdot c_p \cdot \Delta T$$

Here Q is the heat removed, $\dot{m}$ the mass flow rate of the fluid, c_p its specific heat capacity, and ΔT the temperature rise the fluid is allowed to take on.

Everything that follows comes back to this equation, from the advantage of liquids over air to the penalties that appear at high density and the point at which a method stops making sense. The equation offers only two ways to remove more heat. Move more fluid, or let the fluid warm up more. Both reach their limits quickly at high power density. Allowing a larger temperature rise is held at the top by the junction temperature the silicon can tolerate and at the bottom by how cold the incoming fluid can be made without condensation or excessive energy cost, so the usable window is narrow and it does not widen as the problem grows. The ceiling itself does not move: junction tolerances are flat to falling across generations, so every added watt must be bought with lower thermal resistance rather than borrowed from the silicon.

That leaves mass flow as the lever the industry actually pulls, and mass flow is expensive. The power needed to move a fluid does not rise in step with the flow. It rises far faster, so each additional increment of cooling costs disproportionately more than the one before it.

3.2. Why liquids outperform air

The heat capacity term explains why liquids displace air as density climbs.

Fluid type	Specific heat capacity, J/(kg·K)	Volumetric heat capacity, kJ/(m ³ ·K)
Air	1,005	1.2
Pure water	4,184	4,172
25% propylene glycol / water (PG25)	3,950	4,048
Fluorocarbon-based dielectric fluids	1,050	1,785
Oil-based dielectric fluids	1,850	1,610

Table 2 Heat capacity of common cooling fluids.

Volume is what ducts and pipes actually move, and a given volume of water carries roughly three to four thousand times the heat of the same volume of air for the same temperature rise. A liquid therefore removes a given quantity of heat while moving far less volume, which keeps the flow penalty manageable to much higher power. This advantage is the reason liquids enter the picture at all. It is finite, and the same flow penalty that constrains air reappears for liquids further along the curve.

3.3. The shrinking temperature budget

A second relationship matters as much as the first. Between the silicon junction and the fluid sits a chain of thermal resistances, from the die to its case, from the case to the cooling surface and from that surface into the fluid. The junction temperature is the fluid temperature plus the heat flow multiplied by the sum of these resistances.

$$T_j = T_{fluid} + Q \cdot (R_{jc} + R_{cs} + R_{sa})$$

Here T_j is the junction temperature, T_{fluid} the fluid temperature, Q the heat flow, and R_{jc} , R_{cs} and R_{sa} the thermal resistances from junction to case, from case to cooling surface and from surface to fluid.

As device power climbs toward and past a kilowatt while the die area stays roughly fixed, the heat flux through that chain rises sharply and the temperature budget available to absorb any imperfection shrinks. At low power a mediocre thermal interface or an uneven flow distribution is a minor loss. At high flux the same imperfection consumes the entire margin, and small engineering inefficiencies that were once ignored become the factors that decide whether the part stays within specification. This is a statement about margin, not about total power on its own. A low-power device with a tight rated junction temperature, a thin package or a poor thermal interface can exhaust its own budget and hit its limit sooner than a higher-power part that was designed with more headroom, so power alone is not a reliable predictor of which device fails first. The boundaries in this paper bind through flux and the temperature budget, and the next section separates those from the headline power number.

3.4. Flux and total power

Two quantities are easy to conflate here, and separating them matters. Total package power sets the load the facility must ultimately move and reject. Heat flux, the power divided by the area it passes through, sets the difficulty of getting that heat off the component in the first place.

$$q = \frac{P}{A}$$

Here q is the heat flux, P the device power and A the area it passes through.

Reticle limits hold the die area within a narrow range, so as per-reticle power rises the flux rises with it, and it is the flux, concentrated at the hottest points of the die, that defeats a cooling method first. A method can have ample capacity for a chip's total watts and still fail to pull them off the surface fast enough to hold the junction temperature. This distinction is more than a technicality, and it runs through the rest of the paper. Flux is the difficulty of getting the heat off the component, the step that fails first, while total power is what the system must then carry away and reject. A chip's headline power tells only part of the cooling story. Denser nodes sharpen this problem rather than easing it. Packing more transistors into the same area, as the latest sub-1 nanometer research illustrates, raises the heat per unit of die area, which is exactly the flux that acquisition must overcome.

Leading chip flux trends

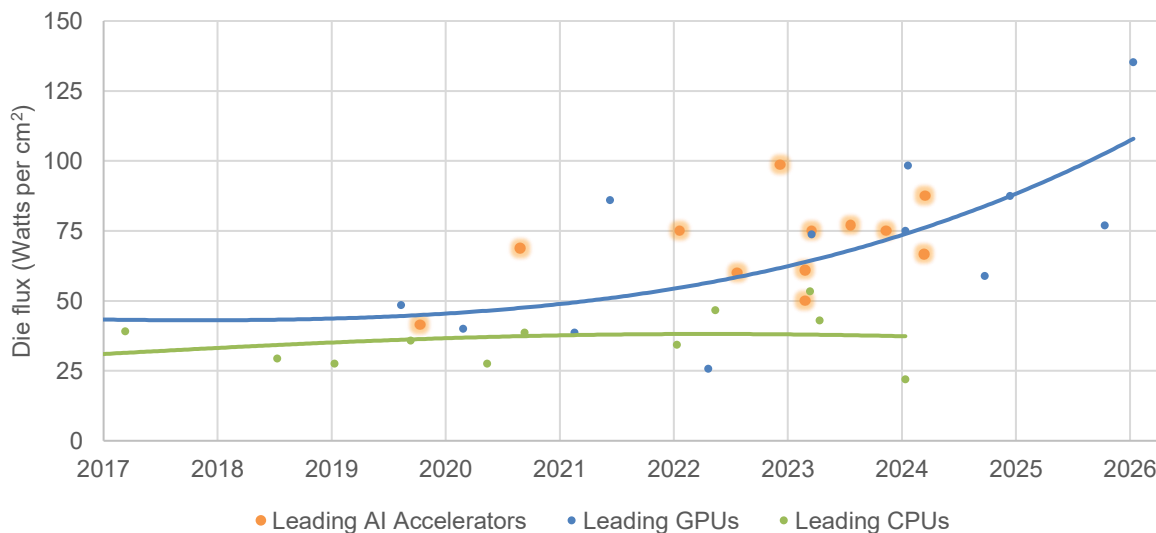


Figure 3 Die-level heat flux, 2017 to 2026; haloed markers denote the AI accelerator class. Sources in A.2.

The record shows how deliberately the industry has used the remaining lever. Across the accelerators until 2024, die-level flux held nearly flat, roughly 50 to 90 W/cm², while package power tripled: the growth was delivered by area, not intensity, because area is the cheaper axis for as long as the reticle allows it. That plateau is no artifact of the window; the longer record of Figure 2 shows the same discipline holding across the CPU era. The first announced breakout above that band arrives in 2024, the signal that the area lever is ending and flux is beginning to climb.

At high density a cooling method fails at the surface, where it can no longer move heat away fast enough to hold the junction temperature, well before it runs short of raw capacity.

4. The three boundaries

The physics gives every cooling method the same overall shape, a comfortable range that ends in a sequence of limits. It helps to name those limits, because they are reached for different reasons and they matter to different people. A method passes the **efficiency**, **economic** and **physics** boundaries in turn, and the distance between them tells us most of what we need to know about where the method belongs and how it will be replaced.

4.1. The efficiency boundary

The first limit is the efficiency boundary. It is reached when the energy spent on cooling grows into a visible share of the energy spent on computing. Plotted against device power, that share follows a curve, and the boundary is not a fixed percentage but the knee of that curve: the point where the cooling share turns visibly superlinear and each further watt of compute buys disproportionately more cooling. Below it, the cooling system is a quiet overhead. Above it, the overhead begins to eat into the useful output of the facility, and every further watt of compute carries a growing tax in fan power or pump power. The method still works and it still holds the silicon in specification. What has changed is that it has stopped being a good use of energy. This is the boundary an efficiency-minded operator feels first, through the power bill and through the share of a site given over to moving heat.

4.2. The economic boundary

The second limit is the **economic boundary**, reached when the full cost of running the method turns against it. Energy is part of that cost and only part. Reliability enters here, because added complexity brings added failure and the cost of failure falls straight onto operations. So does the capital tied up in extra hardware, the labor of building and commissioning it, and the floor space it consumes. A method can sit well inside its **physics boundary** on a tally of watts and still be the wrong choice once these wider costs are counted. The **economic boundary** is where the total cost of ownership, with energy as one term among several, says it is time to change.

4.3. The physics boundary

The third limit is the **physics boundary**, and it is the hard one. It is reached when no achievable flow, surface or temperature can hold the junction within its rated range. Here the method does not become wasteful or expensive. It becomes impossible. The chip cannot be run inside the method at all, and the only choices left are to slow the chip or to change the method. This is the limit that cannot be negotiated, because thermodynamics sets it and budgets do not.

4.4. The avoidance zone

The revealing space is the gap between the **economic boundary** and the **physics boundary**, the **avoidance zone**. Inside it a method is past the point of good economics and still physically capable, so it can be pushed further at rising cost to put off a change. The width of that gap differs sharply from one method to another, and it shapes behavior. A narrow gap leaves little room to delay, because impossibility follows soon after the economics turn, and the change, when it comes, is abrupt. A wide gap invites delay, because the method keeps working long after it has stopped paying its way, and the change, when it comes, is gradual and grudging.

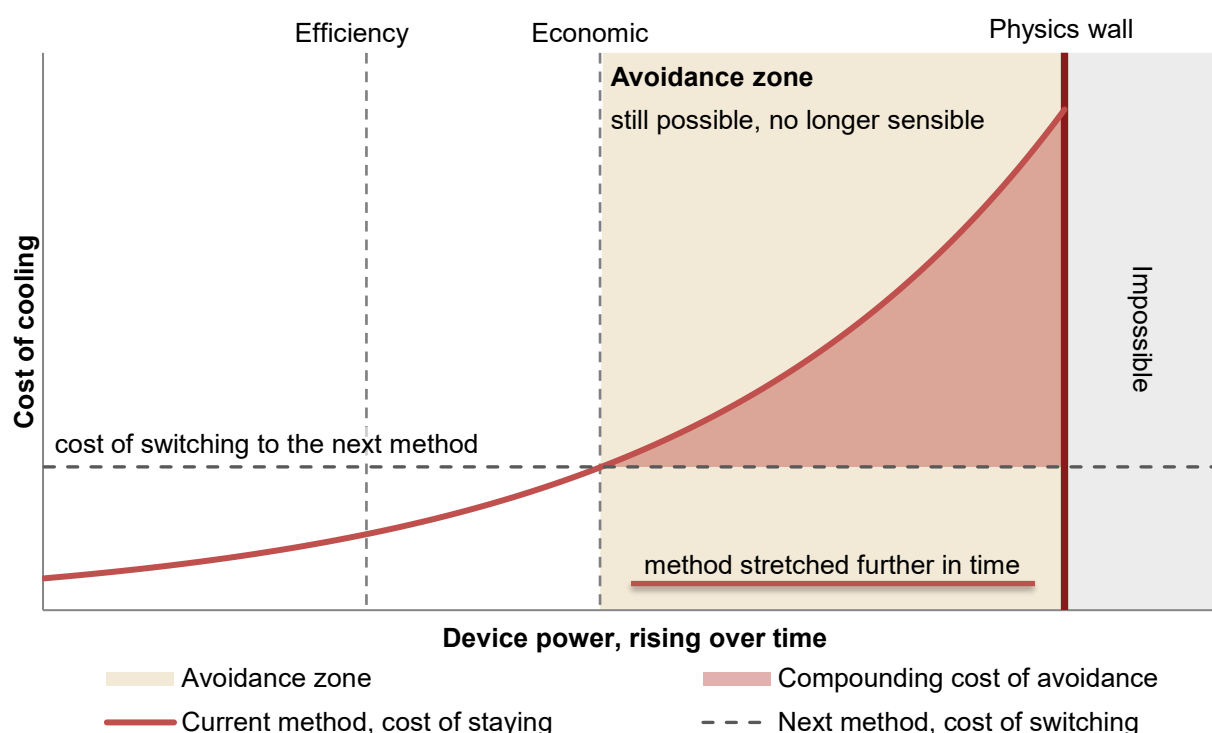


Figure 4 The three boundaries and the avoidance zone.

Whether the industry stops at the **economic boundary** or runs on toward the **physics boundary** is not a thermodynamic question. It is a question of incentives. The width of the **avoidance zone**, and who bears the cost of staying inside it, decides how and when each method gives way, and it is the subject which this paper leans on.

A cooling method usually stops being worth using long before it stops being able to work. The space between those two points is where most of the industry's cost and risk sits.

4.5. Reading a technology by its boundaries

These three boundaries give a compact way to describe any cooling method. Where does it stop being efficient, where does it stop being economic, and where does it stop being possible? The spacing between the answers is the method's character. A method whose efficiency and **physics boundaries** sit almost on top of each other has little usable margin and fails abruptly when it fails. A method with a wide span between its economic and **physics boundaries** can be stretched a long way, for good or ill, and tends to be left in service well past its prime.

5. The adoption law

The boundaries explain when a method should give way. They do not explain when it actually does, and the two are seldom the same. The gap between them is the subject of this section, because it decides whether the industry meets a transition prepared or caught short, and it follows a pattern consistent enough to state as a rule. This rule is the spine of the paper, and the technologies that follow are read through it.

5.1. Change is resisted for good reasons

The Least Disruption Law

A disruptive technology is rarely adopted the moment it becomes the better choice on paper. It is adopted when the incumbent can no longer do the job. The reason is not inertia or poor judgment. It is that a disruptive technology arrives with strings attached, new dependencies that ripple outward through a reliability-driven ecosystem organized around the old way of working. The gains may be real, yet the strings reach into parts of the system the *decision-maker neither owns nor controls*, and the consequences of pulling them come back on the one who pulls. A dependency forced before it is unavoidable raises the burden of adoption for everyone downstream, exposes the mover to rivals who ask less, and sets off second-order effects no one can fully foresee. The rational course, for each participant on its own, is to defer the disruption for as long as the incumbent can be made to work. We call this the **least-disruption law**: an ecosystem as a whole exhausts every option that preserves its current architecture before it accepts one that changes it. Cooling is a sharp case of this, and the next section takes it up.

A dependency forced downstream lands on parties the decision-maker neither owns nor controls, and what it sets in motion comes back on the one who forced it.

The Least Disruption Law versus Mover Advantage

The dynamic is easily mistaken for the first-mover and second-mover advantage of innovation economics, and the distinction is worth drawing. Mover advantage is a strategy choice available to an individual supplier or adopter, a weighing of the cost of early entry against the market position it may buy. The choice only has meaning relative to a tipping point in industry adoption, since the tipping point defines who counts as first and who as second. Mover advantage is, in other words, a strategy played against an adoption curve that already exists.

This paper describes the adoption behavior itself, the layer beneath that choice.

The least-disruption law explains when the tipping point arrives, only when the incumbent method can no longer do the job, and why it arrives later than the economics alone would set it. A consequence follows for the strategy layer. Because the inflection is forced and lands on everyone at once, the reward for waiting that mover strategy assumes is not reliably there, and the scramble that followed the forced move to cold plate is the record of what waiting bought. Deferral is individually rational and collectively expensive, which is why the paper treats it as a source of **preparation debt** (section 5.7, The cost of waiting) rather than a strategy.

The least-disruption law. The ecosystem as a whole exhausts every option that preserves its current architecture before it accepts one that changes it. Inflections are forced, not chosen.

5.2. Why liquid cooling is an unusually strong case

The law holds for any disruptive change, and liquid cooling is an unusually strong instance of it, for the reason seeded in the introduction. The two historical rescues, CMOS and the move to many cores, happened inside the chip and were invisible to everything around it. Liquid cooling is not contained inside the chip. It reaches outward into the server, the rack, the facility loop, the supply chain and the service procedure, parts of the system the chip vendor neither owns nor controls. A device-level improvement can be shipped on the vendor's own authority. A cooling dependency cannot, because it asks a long chain of other parties to change at the same time, and the vendor cannot see far enough down that chain to know what its demand will cost. That reach is what makes liquid cooling the hardest kind of transition to set in motion, and the most tempting to defer.

5.3. The chip vendor's dilemma

The pressure is sharpest on the maker of the processor, and its position is worth following closely, because it sets the clock for everyone else. A chip vendor that forces a new cooling dependency onto the market attaches a string to its product. The string raises the burden of adopting that product, and a customer weighing two processors will notice that one demands a costly infrastructure change and the other does not. In a competitive market that difference can move a sale. So the vendor that moves first risks handing its customers to a rival that asks less of them, and risks alienating the base it depends on.

This gives every chip maker a powerful reason to squeeze the last drop of performance out of the current cooling approach before requiring a new one. It will push air to its limit and then push the first liquid method to its limit, before it imposes the next dependency. From the outside the behavior looks like conservatism. It is closer to self-preservation, and it is rational. The downstream effects of forcing a change are large, they fall on parties the vendor cannot coordinate, and they can rebound on the vendor as lost share. Everything in its power is therefore bent toward avoiding the change for as long as the product can still be made to work.

A chip vendor that forces a cooling change attaches a string to its product, and a competitor that attaches none can take the customer. So the string is held back until the silicon leaves no choice.

5.4. Who feels the limit and who can act

The delay is made worse by a mismatch between the party that feels a limit and the party that can act on it. The operator feels the **economic boundary**, through the cost of energy, of failures and of complexity. The chip vendor holds the design that any cooling change has to fit, and it feels the **physics boundary**, the point where its product cannot ship and perform. Because the one who carries the running cost is not the one who holds the design, the economic signal does not move the architecture. The change waits for the physics, which arrives a generation after the economics first asked for it. We call this the **agency mismatch**, and it is why cooling transitions lag their own business case for years.

5.5. Two clocks

There is an exception, and it is instructive. When the same organization designs the silicon and operates the data center, the **agency mismatch** disappears. Such an organization feels the **economic boundary** and holds the design at once, so it can act on cost without waiting for impossibility. This is why operators with their own silicon are able to move to new cooling earlier than the wider market that buys its processors from a third party. The industry therefore may run on **two clocks**, a faster one for the vertically integrated and a slower one for everyone dependent on a separate supplier, and the distance between them shapes who leads each transition and who follows.

5.6. Forced or gradual

The law sets when a transition happens. The shape of the boundaries decides how it happens, and the two transitions this paper cares about most could not be more different in character.

Pushed by a wall

The move from air to cold plate was forced and abrupt. Air ran into its **physics boundary** on the hottest accelerators, the hard wall that cannot be negotiated, and once a part could not be cooled on air at all, the change had to happen at once. A generation arrived that the incumbent method simply could not hold, and the industry switched within a single product cycle.

Pulled by economics

The next move will look nothing like that. Single-phase cold plate does have a physics boundary of its own, projected at roughly 4,000 watts for a reticle-class die on current silicon roadmaps, and that wall is farther away in watts than in time: frontier package power doubles roughly every two years, and announced parts are already climbing toward its neighborhood. What the wall cannot do is force the industry over at once, because in front of it lies an avoidance zone air never had. Cold plate can be extended, plate by plate and component by component, across more and more of the system, and each extension dodges the physics argument for a while longer. This stretches its **avoidance trajectory** far to the right, so there is no near-term moment where it becomes impossible and the industry is forced over in a season. The pressure to move comes instead from the **economic boundary**, from efficiency and total cost growing too large to ignore, and from the slow erosion of cold plate's own best argument. Its strongest selling point was simple, serviceable hardware. As it is plumbed across the whole system to dodge its limits, the connectors multiply, the reliability penalty grows, and that serviceability argument quietly loses its value. The result is a gradual, grudging transition, pulled by economics where the last one was pushed by a wall.

The stretch no one chooses

There is a deeper reason the move is slow, and it is not that anyone is choosing to stretch cold plate. It is that no single party can do otherwise. The heat is unbalanced and spread across many components, each owned by a different vendor, and none of them is strong enough to mandate an infrastructure change for the rest. A memory maker cannot require the industry to re-cool its servers, so it plugs into whatever is already there, and cold plates on solid-state drives and network cards are becoming routine for the same reason. Each component that crosses its own small threshold adds another plate and another connector, and the cost accretes with no one able to call a halt.

This is **complexity debt**, the compounding price of a solution stretched across a system it was never meant to cover, and it is why the cold plate **avoidance zone** runs so much wider than air's ever did. The trigger for the next transition cannot be one part hitting a wall. It has to be collective pain, which is diffuse and slow, so the stretch goes on.

The ratchet, and the next in line

Cold plate, once forced into place, became the new baseline that the next method has to overcome. This is the **ratchet** by which each transition resets the definition of normal, and it is why the other mature liquid cooling technologies meet resistance now even though they are proven, because they are simply the next dependency in line. Immersion is drawn by a different mix of forces than cold plate was. It answers the distributed load and not the single hotspot, it is pulled forward by reliability and by the holistic coverage of cooling everything at once, and increasingly by the facility, where power and space are scarce. These are economic and operational pulls, not a physics wall, which is one more reason its adoption is more gradual and uneven across the market.

Air hit its wall within a single generation. Cold plate's wall sits much further out, far enough that economics, not physics, will decide when the industry gives it up.

5.7. The cost of waiting

Waiting until a method becomes impossible carries a hidden cost. The disruptive alternative is not ready on the day it is suddenly needed, because no one prepared it while the old method still worked. Supply chains, standards and skills take years to build, and an industry that defers every transition to the last possible moment reaches each one unprepared. We call this the **preparation debt**, the bill that comes due when a forced transition arrives ahead of the readiness to meet it. Business language already carries technical debt and capability debt, an actor's own shortcuts or gaps, incurred inside its own walls and repaid on its own schedule. **Preparation debt** is an ecosystem's unbuilt readiness for a deferred transition, incurred by no one in particular and due all at once when the change is forced.

The move to liquid for the highest-power accelerators is the clearest recent example, and it ran exactly to this pattern. Air held a generation of parts at a few hundred watts without trouble. The next generation, at a much higher power, could not be held on air at all, and the switch to cold plate had to happen at once. Only a thin layer of downstream vendors was ready for it. The supply of cold plates, manifolds, connectors and coolant-distribution units fell short of the sudden demand, deployment schedules slipped against the shortage, and facilities designed around air had their blueprints rewritten on the fly to bring liquid to the rack. The thermal milestone was visible from the outside, and so was the scramble that followed it.

Deferring a transition does not remove its cost. It moves the cost to the moment of least preparation, and the move to cold plate has already sent the bill.

6. The shape of the heat

Two simple frames are worth setting out before the technologies, because much of the confusion in this field comes from skipping them. One is the shape of the heat a system presents. The other is the path that heat travels to leave the building. Both cut across every method, and both help decide which method fits.

6.1. Concentrated heat, and where the move to liquid began

Some heat is concentrated. A large processor pushes a great deal of power through a small, flat, well-defined surface, and the difficulty is the intensity of the flux at that surface. This is the cleanest kind of cooling problem, because the heat arrives in one place with a known geometry and a method can be designed against it precisely. It is measured in power per unit of die area, and it governs whether a surface method such as a cold plate can keep up.

This concentrated problem is where the move to liquid began, and it is central to the story. When air could no longer hold the hottest accelerators within their **physics boundary**, the answer was to bring a liquid directly onto that surface, in the form of a cold plate. The shift to liquid started at the hottest chip. The distributed load discussed below is a second and different challenge, and it matters a great deal, but it is not what triggered the transition.

The concentrated problem is also not standing still. Device power keeps rising, and two newer pressures sharpen it. The die needs a roughly uniform temperature across its surface, so simply letting the coolant warm up more is not a free way to extend a method, because it widens the temperature difference across the chip. At the same time, stacked high-bandwidth memory placed beside and above the logic throws the concentration of heat across the package out of balance, creating local peaks a flat cold surface struggles to follow. These pressures are what draw two-phase operation into the picture for the highest-flux parts, a thread the technology sections pick up.

6.2. Distributed heat, the challenge left behind

Other heat is distributed. Memory, voltage regulators, network silicon and optics are spread across the board, each modest on its own, none offering a clean surface, but together a large and growing load. This is harder to solve, because there is no single hot point to design against and the parts shade one another in the airflow. The right measure is not the flux at any one device but the heat gathered across a slice of rack height, the load a method has to carry away from the system as a whole.

Distributed non-xPU heat per rack unit, AI platform

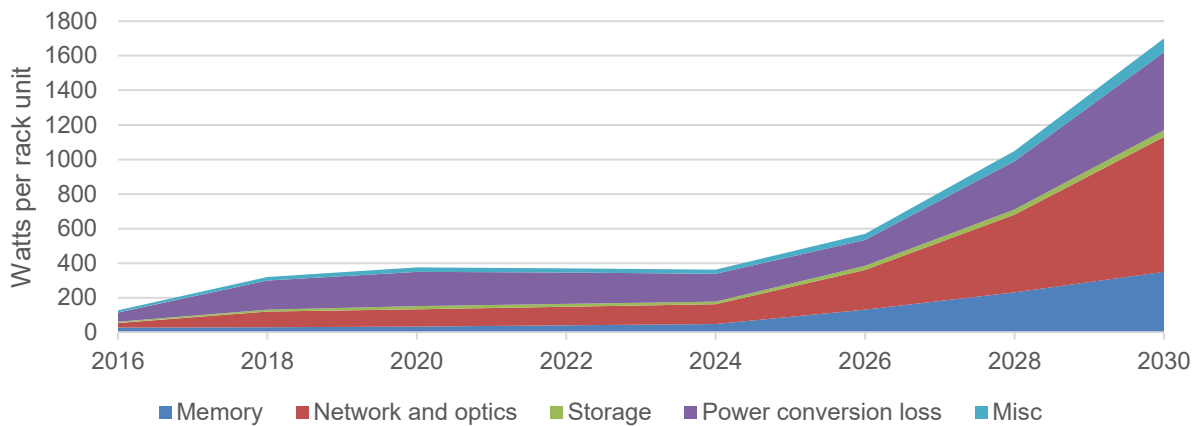


Figure 5 Distributed non-xPU heat per rack unit for the representative AI platform of each era, at published and announced rack powers; modeled composition. Sources in A.3.

Once the hottest processors are moved onto a surface method, the concentrated problem is handled and the binding constraint shifts to everything left behind. The scattered load keeps climbing, and at some point, air can no longer carry it even while the processors themselves are comfortable. The system then faces a choice between treating the whole volume at once or attaching a surface method to every component in turn, and that choice, more than any single device power, sets the timing of the next transition. Treating the whole volume at once is, at bottom, what air already does, cooling everything in the box through the medium it sits in. Immersion is the same idea carried into a denser medium for the whole system, with a surface method kept for the few devices that need still more.

Where this load falls matters as much as how large it grows. The chip supplier largely does not, because its own part is already taken care of by a cold plate, so it has little reason to drive a change. The pressure is felt instead by the platform builders, who have to design cooling for the memory, the networking and everything else, and who run into the difficulty of fitting a cold plate to each of them. The party that feels the distributed problem is not the party that controls the silicon, a sign of the coordination problem the adoption chapter set out. Where a silicon provider integrates vertically and sells at rack scale, it takes the distributed problem onto its own books, exactly like the vertically integrated end users; the coordination problem does not disappear, it moves inside one organization.

Liquid cooling began at the hottest chip. It did not narrow the problem, it multiplied it, reaching deeper into the hotspot and outward across everything around it.

7. The thermal path

7.1. Acquisition, transport and rejection

Cooling is usually spoken of as one activity and it is really three. Heat has to be taken off the component, carried to the edge of the rack, and then rejected to the outside world. The first step, acquisition, is the one the physics of the earlier sections governs and the one that fails first as density rises. The second, transport, moves the captured heat to the boundary of the rack through air, through a fluid loop or through a mix of the two. The third, rejection, hands the heat to the atmosphere, to water or to a reuse system. A cooling architecture is a particular set of choices across all three, and a method that excels at one step can be useless at another. This is where the two quantities separated in the physics divide cleanly. Flux governs acquisition, the hardest step, while total power sets the load on transport and rejection.

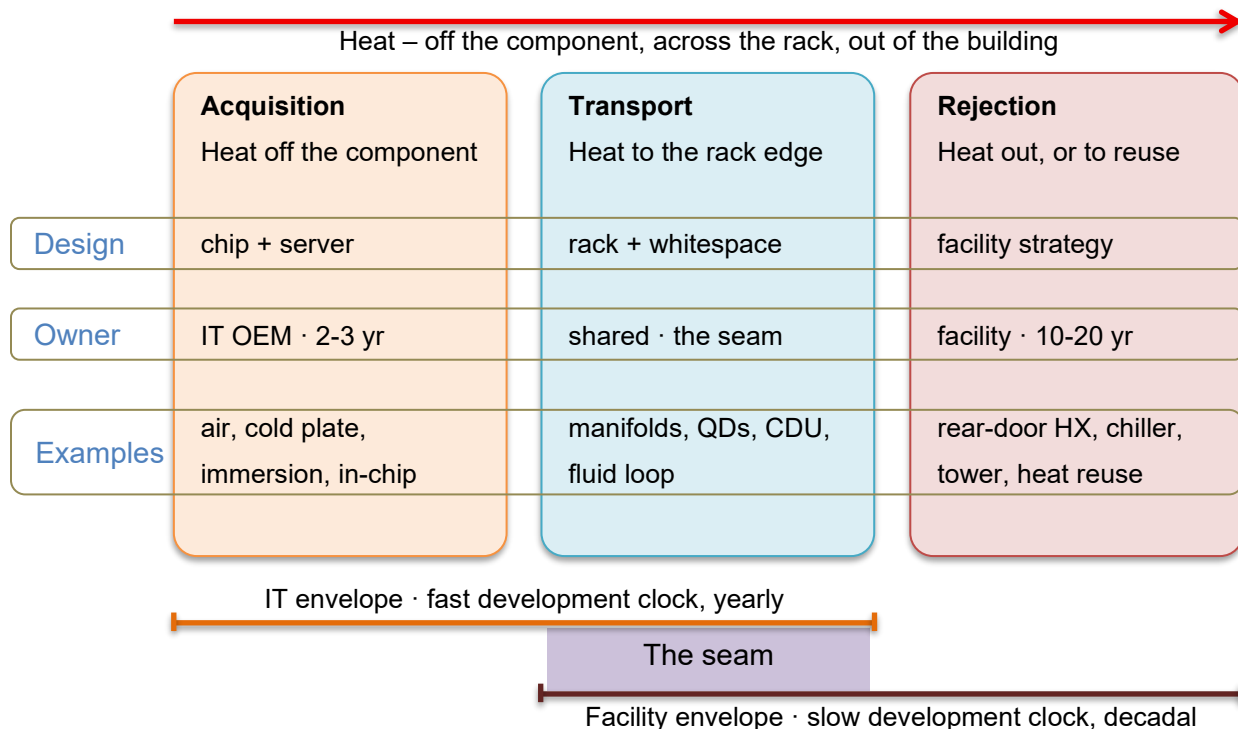


Figure 6 The thermal path segmented into acquisition, transport and rejection, aligned to the IT and facility envelopes and their lifecycles.

7.2. The IT envelope and the facility envelope

These steps do not sit in the same commercial world. Acquisition belongs to the IT envelope. It is bound to the silicon, designed by the people who build servers and refreshed on the two or three year cadence of the hardware. Rejection belongs to the facility envelope. It is bound to the building, designed by infrastructure and mechanical engineers and refreshed on the ten or twenty year cadence of a data center. Transport is the seam between them, the loops, manifolds and connectors where the two worlds have to meet. The envelopes have different cultures, different buyers and different standards, and a great deal of trouble follows from forgetting that they are different. Nor is the seam a single line. In a typical liquid path it is a corridor of hand-offs, half a dozen boundaries from the facility water interface to the cooling surface, each usually made by a different party that holds its own test record, the first half living on the facility clock and the rest on the IT clock. Figure 7 draws the consequence. In the air era the two worlds met at a single handoff. Liquid cooling redraws the roles, four parties now co-design across the seam, and every crossing in the figure is one of the strings of the adoption law, a dependency someone must accept before the technology can ship, and a downstream effect someone else did not choose.

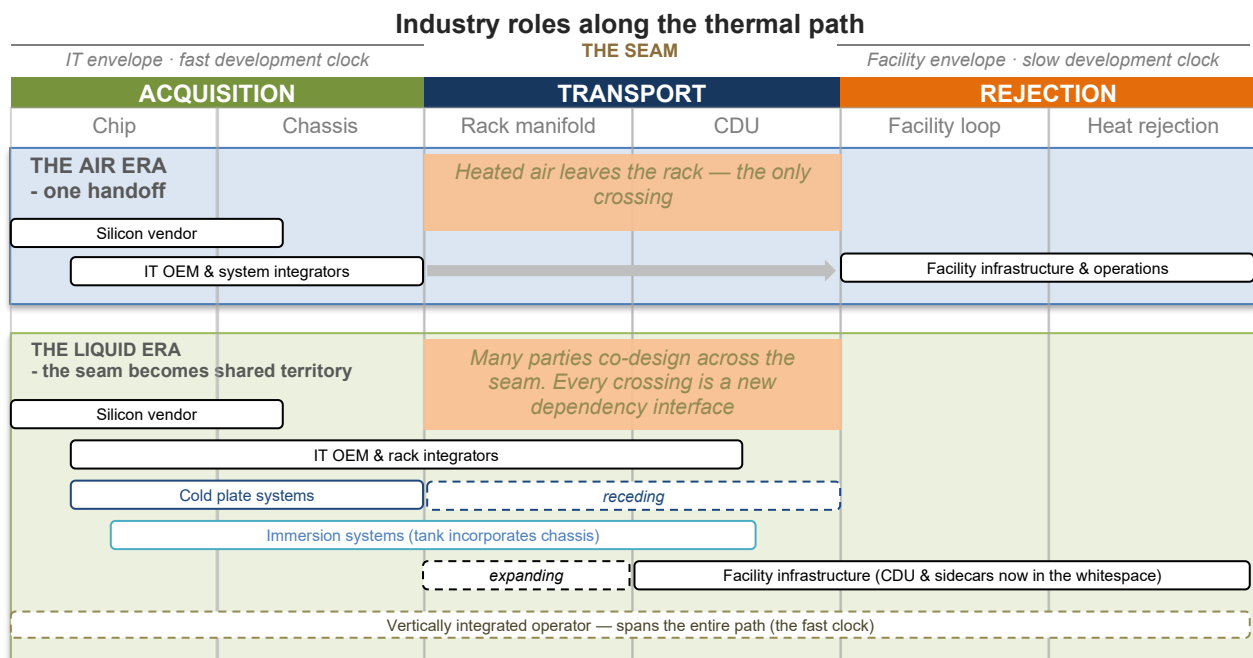


Figure 7 Industry roles along the thermal path, before and during the liquid transition; roles indicative, not exhaustive.

7.3. A difference in lifecycle

The two envelopes also run on different clocks, and conflating them creates a blind spot. Rejection lives in the building, durable and renewed on a timescale of decades. Acquisition is co-developed with the chip and bundled with the IT, which turns over every year. Because device power is climbing steeply, the flux, the thermal spread and the way heat accumulates across the system all shift from one generation to the next, so a cold-plate design can need a near-complete redesign within a year. That demands manufacturing flexibility and a readiness to retool that an industry treating acquisition with a facility-style sense of permanence has not planned for. This lifecycle mismatch is a string in its own right. The industry moving to liquid has not yet taken its full measure.

Acquisition turns over with the chip every year. Rejection lasts for decades. Designing the fast layer with the habits of the slow one is the blind spot.

7.4. A common confusion

A single example shows why the path matters. A rear-door heat exchanger, a coil on the back of a rack, can absorb a large amount of heat, and it is sometimes offered as proof that air has plenty of headroom left. It proves nothing of the sort. A rear-door unit works on **rejection**. It takes heat the air has already carried out of the chassis and hands it to a water loop. It does nothing for **acquisition**, for the harder job of getting heat off a dense component and into the air in the first place, and **acquisition is the step that runs out**. Mistaking a **rejection** capability for an **acquisition** capability is how an organization persuades itself it has more runway than it has.

Rejecting heat and acquiring heat are different jobs in different parts of the business. Being able to do one says nothing about being able to do the other.

8. The cooling technologies

With the segmentation in place, the technologies can be read for what they are, partial answers to a widening problem, each strong in one region and weak in another, each adopted on the timetable the law predicts. For every method we set out the heat it suits, the physics of how it works, the three boundaries that bound it, and where it sits in the pattern of adoption. We begin with air, the incumbent the rest are measured against.

8.1. Air

Air cooling moves heat by drawing a large volume of it across finned heatsinks and out of the chassis. It suits a distributed, moderate heat load well, because it touches every component in the box at once and needs no plumbing, no special fluids and no change to how a server is serviced. Its weakness is written into the physics of the earlier sections. Air has very low heat capacity, so cooling a high load means moving a great deal of it, and the fan power needed to move air rises with roughly the cube of the flow. Doubling the airflow to a hotter generation of parts can demand close to eight times the fan power.

The efficiency boundary

This is where air meets its efficiency boundary, the knee of the fan-power curve the figure marks. As silicon power climbs into the high hundreds of watts, the fan power required to keep it in specification grows into a visible fraction of the power the silicon itself draws. The figure carries that reading in its styling: past the knee the curve runs on muted, still possible, no longer sensible.

Air's efficiency boundary: cooling overhead climbs with device power

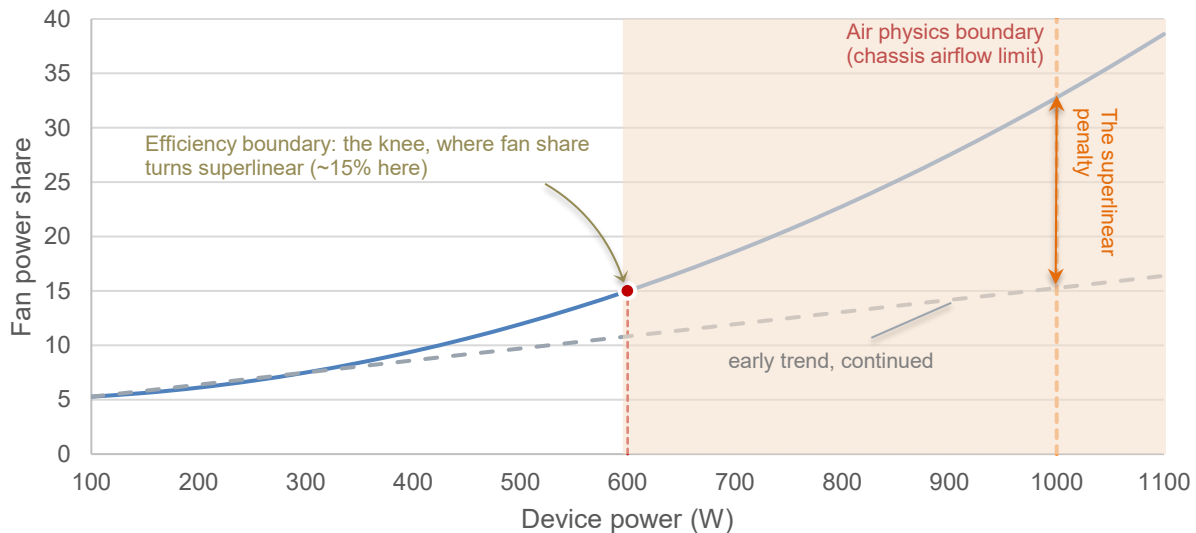


Figure 8 Fan power to hold junction temperature, as a share of the combined device and fan power, for a fixed reference thermal solution; the knee marks the efficiency boundary. Model in Appendix A.

The stretch, paid in space

Air did not give way at that point. It gave way long after. For years the industry ran air well into its **avoidance zone**, past the density where fan power made it a poor use of energy, because the only alternative carried a dependency no one wished to take on before they had to. Air in its upper margins is the first and plainest case of the **least-disruption law** at work, an incumbent kept in service past its prime precisely because the replacement asked too much of the system around it.

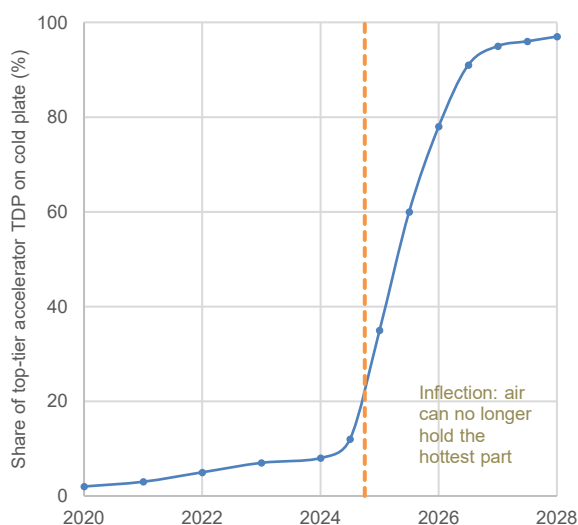
The stretching took two forms. Larger heatsinks and taller chassis genuinely raised what air could acquire, from a few hundred watts to the kilowatt class, but every step was paid for in space, more rack units per device, which is the compounding cost of staying, rendered in physical form. Colder supply air and rear-door heat exchangers worked the other end of the path, buying rejection capacity that moved no **acquisition** limit at all. When rack density itself became the product, the currency of the first form ran out, and the hottest devices arrived at air's **physics boundary** with nothing left to spend.

The wall that forced cold plate

What makes that boundary final is the arithmetic of the thermal path. To remove more heat without moving still more air, the incoming air would have to be made colder, and that runs into condensation and the operating range facilities are required to hold. With the inlet temperature effectively fixed and the junction temperature fixed by the silicon, airflow was the only lever left, and once the hottest accelerators needed more airflow than a practical chassis could deliver, air could not cool them at all. The most recent accelerator generations show the transition cleanly. A class of parts at a few hundred watts was deployed on air across the industry, while the following class, at a markedly higher power envelope, could not be held on air at scale. That is the wall that forced cold plate.

The Cold Plate inflection and the distributed-load trend it leaves behind

Cold plate's forced, abrupt adoption



The load left behind keeps climbing

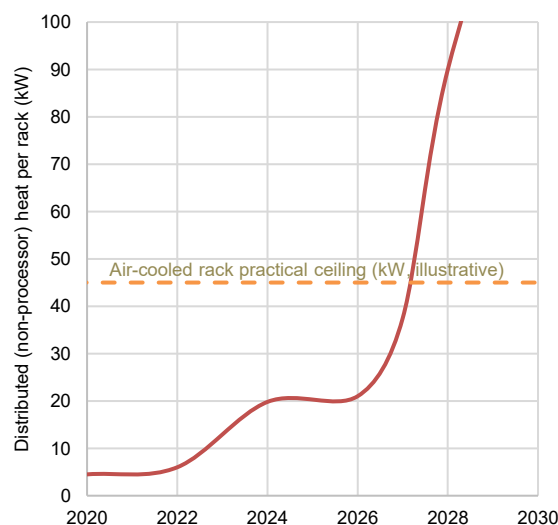


Figure 9 Cold-plate adoption for top-tier accelerators, with the distributed non-processor heat trend left behind.

The load left behind

Where air fails carries an important subtlety. The highest flux devices reach the limit first, and these are also the parts most readily taken over by a cold plate. What air is then left to cool, in systems where the hot chips have moved to liquid, is everything else: the memory, the voltage regulators, the network silicon and the optics spread across the board. This distributed load is harder to characterize than a single hot die, and it is subject to thermal shadowing, where air that has already passed one hot component reaches the next with less of its temperature budget left.

Adding cold plates to densely packed memory while keeping it reliable and serviceable is hard to do all at once, which is part of why the distributed load resists the same treatment the hot chips receive. And the load is not standing still. Published rack-scale designs leave roughly fifteen percent of rack power outside the cold plates, and on announced rack powers that remainder reaches the practical ceiling of an air-cooled rack, roughly forty-five kilowatts, close to one kilowatt per rack unit, in the 2027 to 2029 window. The heat air is left holding is on course to fill what an entire air-cooled rack once carried.

Air's place in the toolbox

Air's enduring advantages are real and easy to undervalue. It is simple, well understood, inexpensive at low and moderate density, undemanding to service, non-conductive barring condensation, and inert barring pollution. Its limitation is equally clear. A decade and a half ago a two hundred watt processor was considered the practical ceiling for air, and steady improvement in heatsinks, fans and airflow management has carried that ceiling toward a kilowatt for a single well-placed device. That progress is genuine and it is nearly exhausted. Air remains the right tool for a large installed base and for the many systems that will never approach extreme density, and it will increasingly share the rack with liquid methods that take on the loads it can no longer hold.

Air has been engineered almost to its physical ceiling. Liquid cooling has barely begun to explore its own.

8.2. Single-phase cold plate

A cold plate is a metal block with internal channels, clamped to a hot component, through which a coolant is pumped. It is a surface method aimed squarely at concentrated heat, and it is very good at it. By bringing a liquid with thousands of times the heat capacity of air into close contact with the package, a cold plate carries a high flux away from a small area while the coolant itself moves in modest volume. For the hottest processors and accelerators, it is the natural successor to the air-cooled heatsink.

Already the default

It is no longer where liquid cooling is heading. It is where high-performance computing already is. When air crossed its **physics boundary** on the hottest accelerators, cold plate became the overnight default, the forced and abrupt transition described in the adoption chapter. Its adoption was not a considered migration but a scramble, and the **preparation debt** that followed is the clearest evidence the paper has for the law.

The efficiency boundary

Within its range a cold plate is efficient. The pump power needed to push coolant through a well-designed channel is small next to the heat it carries, far smaller than the fan power air would need for the same duty, so its **efficiency boundary** arrives later than air's and is seldom the binding limit for a single hot device.

The physics boundary, far in watts and near in time

Its **physics boundary** is set by what happens when the flux climbs higher still. To hold a hotter, denser package the coolant has to move faster through the channels, and beyond a point that fast flow begins to wear the channels, to invite erosion and corrosion and to demand pressures the materials and joints resist. Two newer pressures push in the same direction. The die needs a uniform surface temperature, so the operator cannot simply let the coolant warm up more to gain range, and stacked high-bandwidth memory unbalances the heat across the package, creating local peaks a single flat surface struggles to follow. These are the limits that draw two-phase operation in for the highest-flux parts.

Unlike air's, this wall is measured in thousands of watts. Projected against current silicon roadmaps, single-phase cold plate's physics boundary sits at roughly 4,000 watts for a reticle-class die (about 850 mm²), several times what shipping parts draw. The wall is flux-governed, and the 4,000-watt figure is its per-reticle expression: the survivable flux drawn across a full reticle of die area. That projection is **derived against a constrained design boundary** rather than asserted. Appendix B works the boundary from first principles, with three independent constraints, calorimetric, hydraulic and mechanical, converging in the same region. Measured in time, the distance is shorter than it looks. Package power at the frontier has been doubling roughly every two years, and reported figures for 2027-class parts already approach the wall's neighborhood, so a boundary that seems remote in watts sits only a few product generations away. What keeps the transition gradual is the width of the avoidance zone in front of the wall.

The wall sits at thousands of watts, a few product generations away. What governs the transition is not the wall but the width of the avoidance zone in front of it.

Single-phase cold plate: where the three boundaries sit by power per reticle-class device

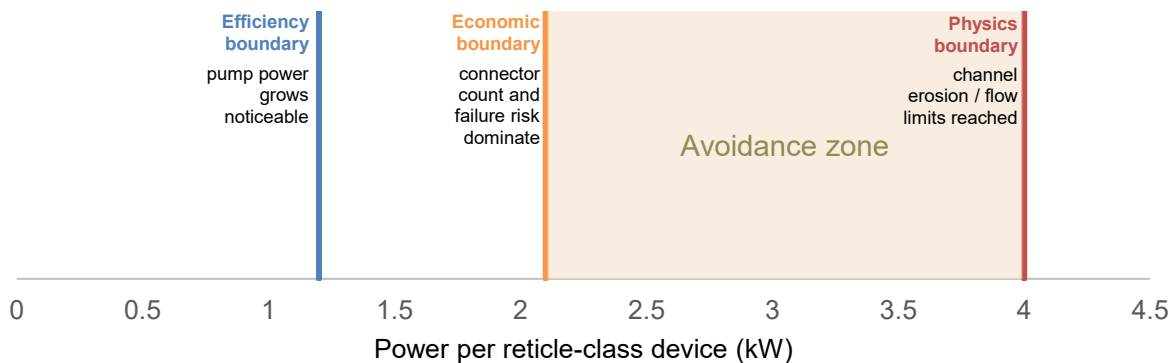


Figure 10 The three boundaries for single-phase cold plate, by power per reticle-class device (3.3); physics wall per the thermo-hydraulic analysis in Appendix B, efficiency and economic positions directional.

The economic boundary binds first

Long before that, the **economic boundary** binds, and it is a matter of complexity and reliability more than of energy. A cold plate cools the component it is clamped to and nothing else. As the distributed load around the processor grows, the temptation is to add more cold plates to more components until much of the system is plumbed. Every plate is another fitting, another connector, another possible leak, and the count of these rises quickly. Each joint is a place where reliability can fall, and in a dense system the cost of a single failure is high.

8.3. Cold plate reliability quantified

The **economic boundary** just described is a matter of complexity and reliability, and it is worth putting a number on. In an air-cooled rack, reliability is governed chiefly by the server fans within each compute tray, which typically run in parallel redundancy so that the loss of one fan does not take down the tray. In a fully liquid-cooled rack, that risk is traded for a different one: the chassis hydraulic components, the cold plates, quick disconnects and manifolds that now sit in the coolant path, and, at the rack level, the coolant-distribution unit that serves them all. To compare the two architectures on equal terms, the dominant failure contributors at the chassis level are the fan subsystem and the GPU itself.

The failure model

Tray fan failure follows standard k-out-of-n redundancy, where each tray carries n fans of which at least k must remain functional:

$$P_{Trayfanfailure} = \sum_{x=n-k+1}^n {}^nC_x * p_{fans}^x * (1 - p_{fans})^{(n-x)}$$

For a rack of N trays, each carrying m GPUs, the failure probability of the air-cooled rack follows directly from the GPU failure rate, the tray fan failure rate just derived, and the redundant power and network paths every rack already carries:

$$P_{Rack Failure (Air)} = 1 - \left[(1 - P_{GPU})^{N_{trays} \cdot m} * (1 - P_{TrayFan Failure})^{N_{trays}} * (1 - P_{PDU}) * (1 - P_{TOR}) \right]$$

The liquid-cooled case takes the same form, with the fan term replaced by a server-cooling failure term and a common coolant-distribution-unit term added at the rack level. Because the liquid-cooled GPU runs at a lower junction temperature, its own failure probability $P_{GPU,DLC}$ is lower than P_{GPU} on air, reflecting the reduced thermal stress a cold plate provides:

$$P_{Rack Failure (DLC)} = 1 - \left[(1 - P_{GPU, DLC})^{N_{trays} \cdot m} * (1 - P_{ServerCooling Failure})^{N_{trays}} * (1 - P_{CDU Failure}) * (1 - P_{PDU}) * (1 - P_{TOR}) \right]$$

Server cooling failure, in turn, is set by the quick disconnects, the manifold and the cold plates themselves, with six quick-disconnect fittings and two cold plates per GB200-class compute tray:

$$P_{ServerCoolingFailure} = 1 - \left[(1 - P_{qd})^6 * (1 - P_{Manifold}) * (1 - P_{CP})^2 \right]$$

For the small failure probabilities involved, $(1 - p)^n \approx 1 - np$, and the product of several such terms approximates to one minus their sum. That turns the rack expressions into sums of component rates, read from here on as expected component failures per year (assuming small probabilities). Applying that simplification to both rack equations isolates the terms that matter: the air-cooled rack's failure rate is set by the GPUs and the tray fans, while the liquid-cooled rack adds the cold plates and quick disconnects but starts from a lower GPU failure rate.

The design point today

Working through both equations with baseline annualized failure rates for each component, shown in Table 3, and a representative rack of 18 compute trays with 4 GPUs each, gives an expected 1.89 component failures per year on air against 1.30 on liquid cooling, a roughly 31 percent reduction. The lower junction temperature a cold plate holds is doing real work here: it more than offsets the extra hardware, fittings and connectors the liquid path adds. The balance is configuration-sensitive. The junction-temperature gain is fixed, while the connector term grows with every component plumbed, and a rack-scale system already carries on the order of 126 coupling pairs. Plumbed far enough, the connector sum overtakes the gain, which is the accretion cost this section closes on and the economics section carries as the swing factor.

Component / subsystem	Typical AFR	Primary failure mode / mechanism
GPU (air-cooled)	2.5%	High thermal stress / junction temperature cycling
GPU (DLC)	1.5%	Reduced junction temperature
Fan subsystem (per tray)	0.5%	Bearing wear, mechanical fan degradation per tray
Cold plate (per GPU)	0.1%	Microchannel clogging, internal seal leak, or micro-corrosion
Quick disconnect (per fitting)	0.15%	O-ring degradation, dry-break mating leak, lock failure
Manifold section (per tray)	0.05%	Structural joint leak or stress cracking
Central CDU	0.8%	PLC controller, heat exchanger, valves

Table 3 Baseline annualized failure rate assumptions, air- and liquid-cooled components.

At today's design point, liquid cooling fails less often than air. The lower junction temperature buys more reliability than the plumbing spends, at six couplings per tray.

Where the trend leads

The design point above is not where the trend is heading. Cold plates are already spreading beyond the processors, onto network silicon, memory and storage, one subsystem at a time, the accretion the adoption chapter describes. Extending the same equations to a fully plumbed tray, eleven plate assemblies and some two dozen couplings at the table's own rates, consumes the entire junction-temperature dividend: the expected failure count rises to 1.96 against air's 1.89, and the advantage is gone. The dividend buys roughly eight plumbed assemblies, and a full tray spends eleven. At coupling rates just twice the table's baseline, a stress case within reported field experience, the count reaches 2.61, a failure rate some forty percent above air, an MTBF factor near 0.7. The reliability gain at the design point and the penalty at the end of the trend are the same arithmetic, read at two points along complexity debt's accumulation.

Read as a design budget, the arithmetic fixes a ceiling: at the table's rates the junction dividend pays for roughly eight plumbed assemblies, about eighteen coupling pairs per tray. The budget spends in both directions: every interface added beyond it costs reliability the technology no longer earns back, and every pair which is engineered out of it, earns it back. Any proposal that moves the coupling count, whatever it delivers in exchange, can be judged against that budget before it ships.

The junction dividend is bounded. The connector count is not. Stretched across the whole system, cold plate's reliability advantage must invert.

Rack-level outage risk

Component-level reliability is not the whole picture, though, and the rack-level view tells a different story. An air-cooled rack has no cooling-related single point of failure. The only single point of failure is the shared power and network feed. A liquid-cooled rack adds one: the coolant-distribution unit that serves every tray on the loop. Of the CDU's 0.8 percent component rate in the baseline table, only its shared, non-redundant modes threaten the whole rack. Even with redundant, N+1 pumps inside the CDU, its shared controls, heat exchanger and valves remain common to the whole rack, and annual outage risk rises accordingly:

$$\lambda_{\text{Rack Outage (Air)}} = \text{AFR}_{\text{Dual PDU}} + \text{AFR}_{\text{Dual TOR}} \approx \mathbf{0.0001} \text{ (0.01\% / year)}$$

$$P_{\text{Rack Outage (Air)}} = 1 - e^{-0.0001} \approx \mathbf{0.01\%}$$

Against the DLC rack's, where the CDU's own failure modes replace the dual power and network feed as the dominant term:

$$\lambda_{\text{Rack Outage (DLC)}} = \text{AFR}_{\text{Dual Pump}} + \text{AFR}_{\text{Controls/Sensors}} + \text{AFR}_{\text{HX/Valves}} + \text{AFR}_{\text{Dual PDU/TOR}}$$

$$\lambda_{\text{Rack Outage (DLC)}} = 0.0004 + 0.0015 + 0.0005 + 0.0001 = \mathbf{0.0025} \text{ (0.25\% / year)}$$

$$P_{\text{Rack Outage (DLC)}} = 1 - e^{-0.0025} \approx \mathbf{0.25\%}$$

Even with pump redundancy inside the CDU, the shared coolant-distribution unit raises the relative risk of a rack-level outage by a factor of roughly 25 relative to air, from about 0.01 percent to about 0.25 percent per year.

Liquid cooling trades many small failures for one shared point of failure. A rack that fails less often as parts fails more often as a whole.

Two different failure modes

The two results are not in tension, they describe different failure modes. Cold plate lowers the rate at which individual components fail, because it holds the silicon at a lower, steadier junction temperature. It raises the rate at which an entire rack goes down at once, because it introduces a shared piece of infrastructure, the CDU, that air-cooled racks never needed. The decision to move to liquid cooling turns on the higher rack power density air cannot provide and on the improved facility-level efficiency liquid delivers, and both are real advantages. The rack-level outage risk that comes with them is not a reason to avoid liquid cooling. It is a design constraint the datacenter and the software that runs on it have to account for, through CDU redundancy, fault isolation and workload placement that assumes a rack, not just a component, can go down.

Reliability moves the boundaries

This reliability picture sharpens the point the **economic boundary** made in general terms. The wide span between cold plate's economic and **physics boundaries** is exactly the **avoidance zone** the law warns about. Because the method keeps working as it is plumbed across the system, it can be extended far past the point where the economics have turned, dodging its physics boundary for years. No one is choosing this. Because the heat is spread across many separately owned components and no single vendor can mandate a change for the rest, the accretion has no natural stopping point, a mechanism the adoption chapter sets out and the economics section quantifies as **complexity debt**.

That is why the move off cold plate will be gradual where the move onto it was abrupt. It also dissolves cold plate's best argument. Its appeal was simple, serviceable hardware, and as the connectors multiply to cover the whole system, the serviceability fades and the reliability penalty grows. Cold plate is the right tool for concentrated heat, it builds on a familiar rack and its supply chain is maturing fast. It was never meant to cool everything, and pressed into that role it carries a rising burden the alternatives are measured against.

A cold plate is excellent at cooling one hot chip. The difficulty begins when it is asked to cool the whole system.

8.4. Two-phase cold plate

A two-phase cold plate resembles its single-phase cousin but works on a different principle. Instead of warming a liquid as it passes, it lets the coolant boil inside the channels. Boiling absorbs a large amount of heat at a nearly constant temperature, because the energy is spent turning liquid into vapor and the fluid barely warms while it does so. The magnitude is the reason two-phase operation is worth the added complexity. Single-phase cooling relies on sensible heating, governed by the same relationship set out earlier:

Sensible heat against latent heat

Sensible heat, single-phase:

$$Q = \dot{m} \cdot c_p \cdot \Delta T$$

Latent heat, two-phase:

$$Q = \dot{m} \cdot h_{fg}$$

Here $\dot{m}$ is the mass flow rate as before and h_{fg} the fluid's latent heat of vaporization.

Boiling trades the temperature-rise term for the fluid's latent heat of vaporization, h_{fg} . For the dielectric fluids used in electronics cooling, h_{fg} is roughly ten to twenty times larger than $c_p \cdot \Delta T$ for a practical single-phase temperature rise, so the same mass flow carries many times the heat. This is what lets two-phase surfaces push the flux ceiling outward instead of simply moving more fluid, and it brings two advantages that answer the pressures just described. It carries far more heat for a given flow, and it holds the surface at a steadier, more uniform temperature, which suits both the highest-flux devices and the unbalanced heat of stacked memory.

Two-phase cold plate operates as flow boiling. Coolant is pumped through a channel and boils as it goes, and the imposed flow buys a high, controllable heat transfer coefficient at the cost of pumping power and of managing vapor in a moving stream. It suits a single, well-defined hot surface.

The physics boundary, pushed outward

By trading a temperature rise for a change of phase, two-phase operation pushes the **physics boundary** outward. The flux a surface can handle climbs, and the fast-flow problems that limit single-phase cold plate ease, because less mass has to move to carry the same heat. For concentrated heat it reaches further up the curve than any single-phase surface method can. The parts now reported approaching single-phase's wall are its first natural candidates.

Two-phase cold plate becomes viable where single-phase runs out of room

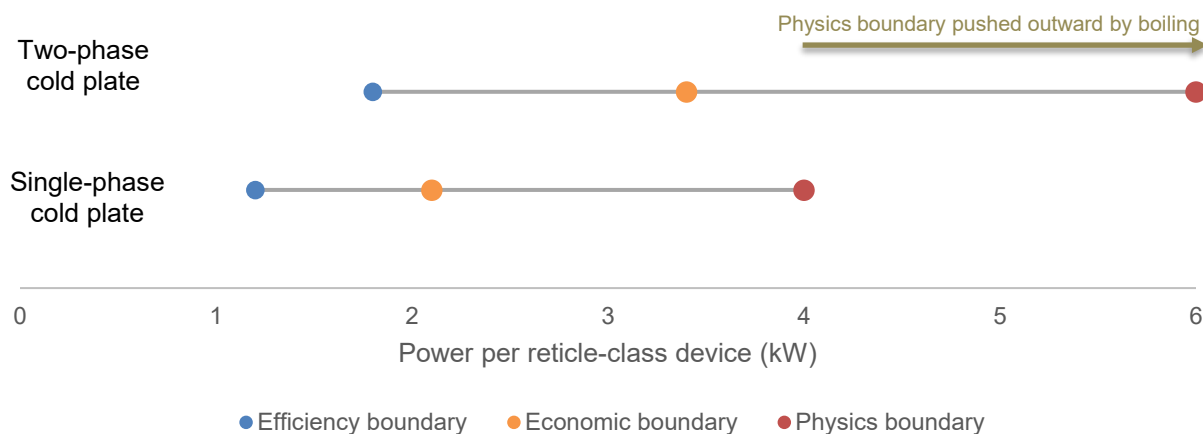


Figure 11 Illustrative single-phase and two-phase cold plate boundary positions, by power per reticle-class device; physics walls derived in Appendix B, two-phase position directional.

Only the cold plate family appears in this chart since these are the surface methods that contend for the point load, and the distributed load is a different job on the thermal path.

Raising the flux budget

This advantage reaches past cooling and into chip design. Processors today are designed against a fairly flat flux budget, the heat a given area of die is allowed to shed, because the cooling sets that ceiling. Lift the ceiling and the constraint loosens. Two-phase cold plate outperforms any other surface method on flux, so it hands chip designers room to pack more performance into the same area while keeping the part coolable. The cooling method becomes an enabler of denser silicon, and not only a response to it.

Two-phase cooling does not just remove more heat. It raises the flux budget chips are designed against, an enabler of denser silicon rather than a response to it.

Early in its own maturing

It is also early in its own maturing, and that matters to how it should be judged. Its present limits are a stage in a development cycle, not a fixed ceiling. As designs are refined and as adoption builds the experience and the supply base behind it, two-phase cold plate will grow more effective and more routine, the same path every method here has followed. The price today is in control and in the fluid. Boiling is harder to manage than simple flow, the system has to handle vapor and avoid dry-out where a surface loses its liquid film, and the fluids that boil at useful temperatures bring concerns of cost, containment and a tightening regulatory picture around certain coolant chemistries.

Strings of its own

There is also a dependency of its own. The single-phase cold plate now rolling out at scale is water-based, and moving to a boiling fluid is an infrastructure change and not a drop-in. It touches the facility loop, the fluid handling and the service model, so two-phase cold plate arrives with its own strings and downstream effects and falls under the same law as every other transition.

Even the cooling family's own next step arrives with strings attached. A boiling fluid is an infrastructure change, and the law applies to it as to everything else.

Two-phase cold plate is best read as the high-flux extension of the cold plate family, earning its added complexity where a single device pushes more heat through its surface, or spreads it more unevenly than a single-phase plate can take.

8.5. Single-phase immersion

Immersion takes a different path. Instead of attaching a cooler to each hot component, it submerges the whole system in a tank of dielectric fluid, a liquid that does not conduct electricity, and lets that fluid carry heat away from everything at once. It is a volume method aimed at the distributed problem the surface methods leave behind. Where a cold plate asks which components to plumb, immersion treats the entire board as one thermal body and cools the hot chips and the scattered load together. In that sense immersion is closer to air than it first looks. Both cool the whole system through the medium the electronics sit in, air in one case and a dense dielectric in the other. It is not a new philosophy replacing an old one, it is the same whole-system approach in a better medium, with a surface method kept on hand to strengthen the few hottest devices.

The tank is the form most deployments take, but not the only one: chassis-level immersion seals the fluid volume around a single system and plumbs it into the rack's manifolds like a cold plate. Read by the boundaries, it carries immersion's whole-system physics inside the enclosure and the surface methods' coupling economics outside it, and the connector arithmetic of the reliability chapter travels with that choice.

That whole-domain strength comes with a trade-off at the hotspot, and it is fundamental. A dielectric fluid carries less than half the heat of plain water by volume, as Table 2 shows, so on the single hottest, highest-flux device immersion cannot reach what a water-cooled cold plate can. The distributed reach is bought partly at the expense of peak hotspot capability.

The gap is narrower than it sounds, because the bath is not the whole story. Precision cooling, which directs and enhances the fluid flow at the hottest devices inside the tank, has been validated on chips around 1500 watts. A flux limit still has to be drawn, and it should be drawn honestly. The industry has not yet settled the fluid formulations or the immersion-optimized heatsink designs that a specific methodology would rest on. For the analyses in this paper, immersion's point-load wall is therefore taken directionally at 2,000 watts, roughly half of cold plate's, purely derived from the fluids' heat capacity per unit volume. Immersion's strength lies on the other axis, where the same bath carries the distributed load entirely.

The decisive advantage is that this hot device is cooled in the same operation as everything around it, so the hotspot and the distributed load are handled together in one system, not split across two. That combination, a surface method for the very highest flux and immersion for the holistic rest, is one of the strongest arguments in the paper for using more than one technology at once.

Immersion cools everything at once, but its dielectric carries less than half the heat of water, so the very hottest chip is where cold plate and immersion meet, each doing what the other cannot.

8.6. Immersion reliability quantified

Failure mechanisms

Immersion tends to raise reliability where heavy plumbing lowers it, and the mechanism deserves the same scrutiny the cold plate model received. The gain does not come from counting hardware. Relative to air, immersion removes the tray fans, a few percent of the arithmetic, and adds only a handful of tank-level couplings. It comes from what the bath removes from every component at once.

Thermal cycling, the fatigue driver for solder joints and interconnects, collapses in amplitude inside a thermally massive fluid. Vibration ends with the fans. Dust and fouling end with the airflow. And because the hardware sits in a fluid that excludes oxygen and moisture, the oxidation, corrosion and humidity mechanisms that age electronics are largely shut off. Each of these acts on the baseline failure rate of every component in the system, not on one hot device, which is why the effect compounds.

The subtraction is not free of structure. An immersion system carries its own shared dependency, the pump and filtration skid that serves a tank or row, and it enters the availability arithmetic exactly as the coolant-distribution unit does for a liquid-cooled rack. Table 4 sets the mechanisms side by side.

Failure mechanism	Air	Single-phase cold plate	Single-phase immersion
Thermal cycling fatigue (solder joints, interconnects)	Present	Reduced at the plated device, unchanged elsewhere	Largely eliminated, cycling amplitude collapses in the bath
Vibration (fans, air movers)	Present	Present for the unplated load	Eliminated
Airborne dust and fouling	Present	Present	Eliminated on the IT side; fluid-side cleanliness managed by filtration
Humidity, oxidation and corrosion	Present	Present	Largely eliminated, the fluid excludes oxygen and moisture
Couplings in the coolant path	None	Grow with coverage (Section 8.3)	Few, at tank level
Shared-loop single point of failure	None	CDU per rack	Pump and filtration skid per tank or row

Table 4 Failure mechanisms by cooling environment, air, single-phase cold plate and single-phase immersion.

The failure model

The magnitude can be estimated with the same discipline the cold plate model used, by making the assumptions explicit and countable. Decompose each component's baseline failure rate into the shares carried by the mechanisms of the table above, thermal cycling, the environmental group of humidity, oxidation, corrosion and dust, mechanical wear from vibration, and a residual of electrical, firmware and latent-defect failures that no cooling environment touches. Immersion applies an elimination factor to each share, and adds back its own hydraulics, the tank couplings and the shared skid:

$$\lambda_{imm} = \lambda_{comp} \sum_i s_i (1 - e_i) + \lambda_{hyd}$$

Here λ_{comp} is the rack's summed component failure rate on air excluding fans, s the mechanism shares, e the elimination factors, and λ_{hyd} the added tank couplings and skid. The summation runs over the four mechanism groups of the table, cycling (cyc), environmental (env), mechanical (mech) and residual (res); the residual's elimination factor is zero by definition, the residual being precisely what no cooling environment removes.

$$M = \frac{\lambda_{air}}{\lambda_{imm}}$$

Here M is the MTBF multiplier relative to the air rack and λ_{air} its total failure rate including the fan subsystem that immersion eliminates, 1.89 per year in the Section 8.3 configuration.

The estimate and the industry record

Working the Section 8.3 rack through, the air baseline is 1.89 failures per year, of which 0.09 is the fan subsystem that immersion removes outright, leaving 1.80 of component rate to decompose, with roughly 0.014 added back for four tank-level couplings and the shared skid. A conservative case, shares of 25 percent cycling, 50 percent environmental, 5 percent mechanical and 20 percent residual with eliminations of 80, 90 and 100 percent, leaves 30 percent of the component rate and yields M of about 3.4. A moderate case, shares of 30, 50, 5 and 15 percent with eliminations of 90, 95 and 100 percent, leaves about 20 percent and yields M near 5.

The shares are declared assumptions, but they are not free to take any value. Microsoft's Project Natick, a sealed nitrogen-atmosphere data center rather than an immersion system, removed only the environmental and handling mechanisms, its servers kept their fans, and reported roughly one eighth the failure rate of its land-based twin¹, which anchors the environmental share at the large end of these assumptions.

And the fleet observation reported in this community, more than three thousand immersion-cooled servers over thirty months with gains up to twelvefold², implies, read backward through the model, a residual share near one in twelve. The estimate of three to five is therefore the conservative end of a range whose far end the industry has already reported.

¹ ["Microsoft finds underwater datacenters are reliable, practical and use energy sustainably"](#), Microsoft, September 2020. Project Natick Phase 2

² ["Immersion Server Demonstrated Reliability at Datacenter Scale Deployment"](#), presented at OCP EMEA summit in 2025 by Austin Hipes (UNICOM Engineering)

The truth for a given deployment lies between the model's estimate and the fleet's observation, and closing that gap is precisely the reliability data this paper's conclusion invites the community to contribute. The two immersion entries bracket the expected range rather than contradict one another: the model estimate reads upward from its conservative floor, while the fleet observation reads downward from the reported ceiling toward that same floor, and fleet-scale validation of the intervening range remains the open community ask.

Directionally the conclusion is already firm. The mechanisms move immersion's **economic boundary** outward, while heavy plumbing moves the cold-plate boundary inward. Figure 12 places all five entries on one scale.

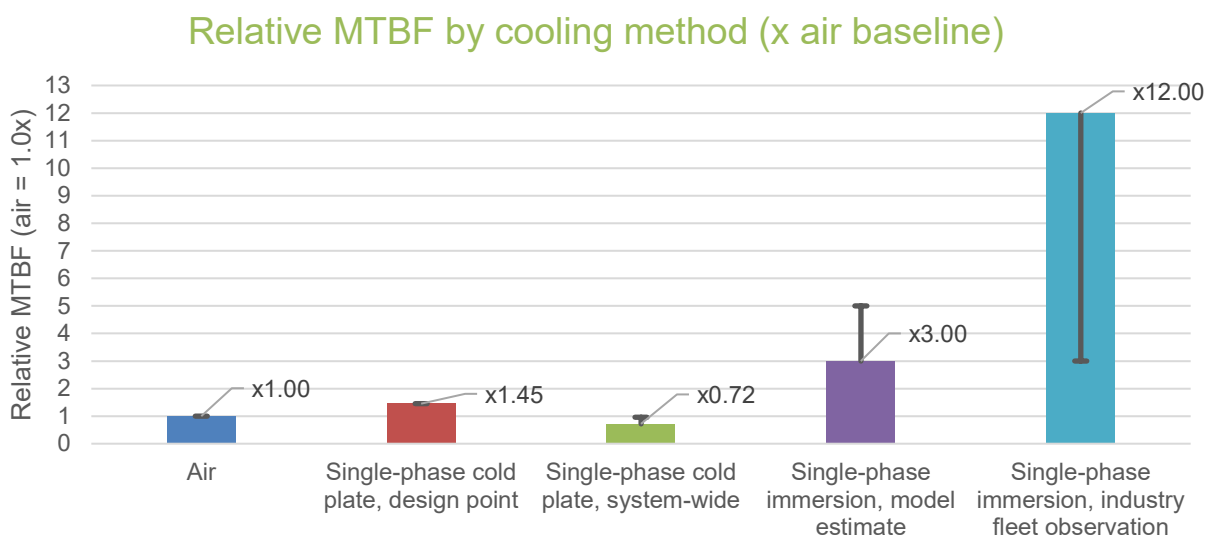


Figure 12 Relative MTBF by cooling method.

Immersion's reliability gain is subtractive, the bath removes failure mechanisms from every component at once. The model puts the floor at three to five times air; the fleet record reports up to twelve.

Immersion's preparation debt

One condition underlies all of it. These reliability gains require materials that are selected, or designed, in combination with the fluid they sit in. There is no reference fluid for equipment manufacturers to design against, so it remains common practice to retrofit hardware built for air, and every such program passes through pilot stages to rule out chemistry effects between fluid, polymers, metals and markings.

This is one of immersion's strings in the full sense the adoption chapter gives the word, a dependency that reaches across vendor boundaries, and it feeds the same loop the law creates everywhere else. Design for purpose and reference chemistry come with adoption, adoption is deferred while the incumbent can still be stretched, and the deferral holds at bay the very maturity it is waiting for. Furthermore, high-speed interfaces validated in air are not automatically valid in the bath: the fluid's higher dielectric constant shifts channel impedance and raises insertion loss.

This is a design input, not a wall, but until validation in fluid is routine, every high-speed interface enters immersion as a qualification exercise. The compatibility groundwork that would make these gains routine is part of the **preparation debt** the industry has yet to pay down.

Costs, and how immersion arrives

Immersion has real costs of its own. It is a single-phase method at heart, so the flow penalty that limits all sensible-heat cooling returns at high enough density. It asks for new fluids and the handling they need, a different way of servicing hardware that now sits in a bath, and facility changes for tanks and fluid management that the IT and facility teams have to learn together. Its capability has moved quickly, from a few hundred watts once thought a sensible ceiling to well beyond it, and the practical range keeps moving.

There is a structural cost as well, and it is best seen through the thermal path. Cold plate keeps acquisition and transport separable: the plates and manifolds live in the IT envelope, the coolant-distribution unit is a discrete product across the seam, and each side can be specified, sourced and standardized on its own. Immersion does not divide there.

The tank is the acquisition environment, but pumping, filtration and heat exchange, the transport functions a CDU carries elsewhere, are designed into or tightly around that same tank, so an immersion solution crosses from acquisition into transport as one co-designed product. The seam that cold plate draws between two products runs through the middle of an immersion product.

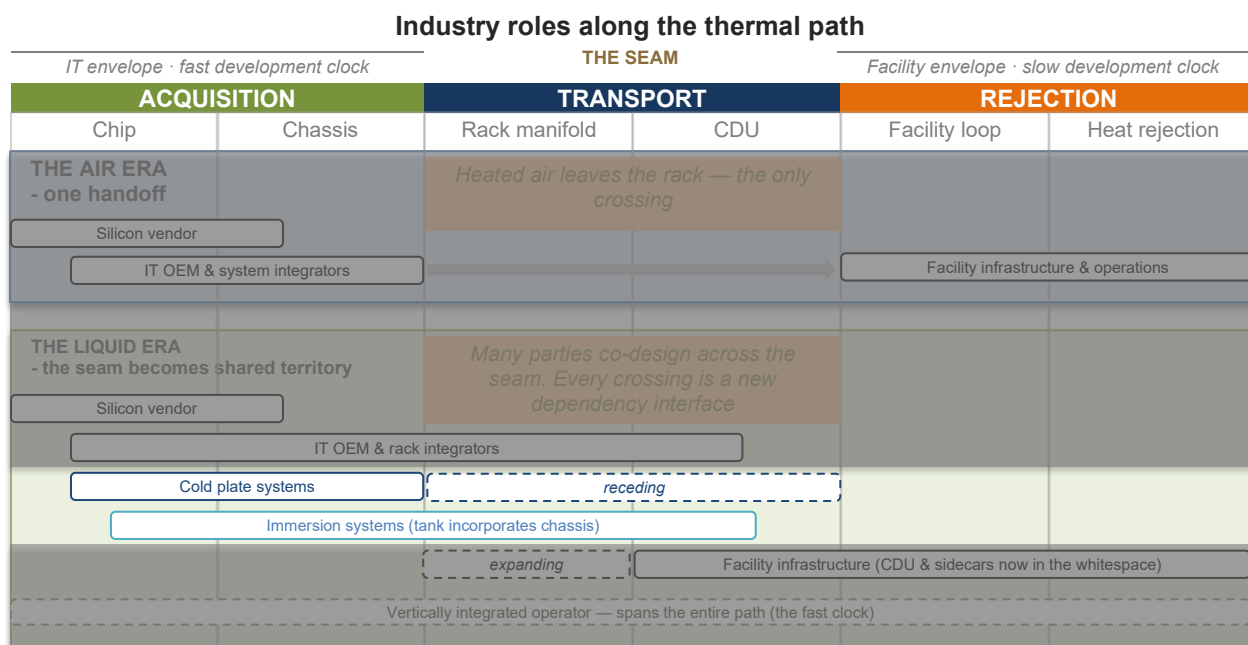


Figure 13 The ecosystem boundary immersion crosses, spanning the IT and facility arenas.

The consequence is ecosystem-shaped rather than thermal: there are fewer clean interfaces to standardize, integration stays vendor-specific for longer, and every deployment carries one more of the strings the adoption chapter describes. Where cold plate concentrates its complexity at the seam, immersion internalizes the seam, and the coordination burden that lives there comes with it.

Cold plate splits acquisition from transport at a seam the ecosystem can standardize. Immersion builds both into one product, so the complexity does not disappear, it moves inside, where standards reach last.

How immersion will actually arrive, pulled by economics and the facility on a slower clock than cold plate's abrupt switch, follows from the adoption chapter and is taken up again in the synthesis.

8.7. Two-phase immersion

Two-phase immersion combines the volume approach with boiling. The mechanism is the pool boiling regime, fluid boiling in an open bath with no imposed flow, the counterpart of the flow boiling inside a two-phase cold plate. The hardware sits in a fluid that boils on the hottest surfaces, the vapor rises and condenses on a coil above, and the condensate falls back as liquid. Pool boiling needs no pump at all and suits a whole system submerged at once, at the cost of less control over where and how fast the vapor forms. In pure thermal terms it is the strongest method here by a wide margin. It removes heat at the highest rate of any method fielded today and holds every component at a single, uniform temperature, because anything hotter simply boils the fluid off at its fixed boiling point. That stability is a genuine performance advantage, since a chip held at a steady, known temperature can be run closer to its limits with confidence.³ None of this is speculative. A hyperscale operator ran a two-phase immersion tank in a production data center from 2021, reporting server power reductions of five to fifteen percent from the removal of fans and the thermal headroom the bath provides.⁴

Held back by everything around the cooling

It is not held back by immaturity or by any shortfall in cooling. It is held back by everything around the cooling. The fluids are costly. The hardware cannot be accessed while it is running, because the system is a sealed, boiling bath. The vapor has to be sealed in, the materials have to tolerate it, and the operational safety case around a working fluid in vapor form is demanding. Above all, the coolant chemistries that perform best are fluorocarbons that raise serious environmental and health concerns across many regions. When a leading fluid manufacturer announced its exit from PFAS production by the end of 2025⁵, the supply of the working fluid itself came into question, and that, more than any shortfall in the technology, has confined its advancement to geographies where these chemistries remain permitted.

³ M. Jalili et al., "[Cost-Efficient Overclocking in Immersion-Cooled Datacenters](#)", ISCA 2021, 48th International Symposium on Computer Architecture

⁴ "[To cool datacenter servers, Microsoft turns to boiling liquid](#)", Microsoft, April 2021

⁵ "[3M to Exit PFAS Manufacturing by the End of 2025](#)", 3M news release, 20 December 2022

The inverted case

Read by its boundaries, two-phase immersion is the inverted case. Its physics boundary sits farther out than any other method in this paper, and its economics would let a single technology answer the hotspot and the distributed load at once. What pulls the boundary inward is neither physics nor cost but the chemistry: the uncertainty around regulatory and health liabilities of the fluids that perform best. Until fluids of comparable capability emerge without those liabilities, that remains the challenge that binds this technology.

In pure thermal terms, two-phase immersion wins by a wide margin. It is everything around the thermals, the fluids most of all, that has brought it to a halt.

8.8. Microfluidics

Microfluidics is different from the methods before it. Instead of cooling the package from outside, it brings the coolant inside the chip, through microscopic channels etched into the silicon itself or into the structure just beneath it. By removing the layers of material that normally sit between the hot transistors and the coolant, it attacks the flux problem at its source, where the heat is made, and cuts the thermal resistance that tightens the temperature budget as flux climbs.

An amplifier, not a rival

This makes it more an amplifier of the other methods than a rival to them. A chip cooled from the inside presents an easier surface to whatever cools it from the outside, so an external method paired with in-chip channels can reach further than either could alone. The pairing becomes most compelling as two-phase cooling becomes the norm. By relieving the flux inside the processor, microfluidics lets the part carry a higher heat load without overheating and presents a high-flux surface at the lid, which a two-phase method is well suited to take from there. In-chip acquisition handing off to two-phase transport is one of the strongest combinations on the horizon.

Microfluidics attacks flux at its source, inside the silicon. What every other method must fight through the package, it removes before the package begins.

The temperature budget, spent two ways

The enablement is worth stating precisely, because it applies to every method in the toolbox, not only the strongest pairing. What limits each external technology at the hotspot is not the total heat but the resistance stack between the junction and the coolant, and that stack is exactly what microfluidics removes. The temperature budget it recovers can be spent in either direction. Spent on power, it lets a cold plate serve a device whose flux would otherwise cross the plate's own limits, and it softens immersion's one structural weakness by reducing the hotspot to what the bath, or the precision flow within it, can carry. Spent on temperature, it lets the same device run on warmer coolant, and every degree recovered at the die is a degree the facility no longer has to produce, which moves the efficiency boundary of the entire chain. Either way, the external method is left with the problem it is good at, moving bulk heat, while the problem that sets its physics boundary, extreme flux through a package, is solved where it originates. That is why microfluidics reads differently: it does not draw boundaries of its own so much as push outward the boundaries of whichever method sits downstream of it.

Crossing into the chip

Its difficulty is that it crosses into the chip, the most tightly held boundary in the system. The experimental approaches so far lean on external infrastructure to feed the in-chip channels, which adds strings of its own. Chip suppliers hold intellectual property across several forms of microfluidics but have not deemed it necessary yet, partly because adopting it would extend their design into the chassis and the infrastructure which is a domain they, so far, have no reason to enter. What could change that calculation is vertical integration. As devices stack, from sheets within a transistor toward stacked active dies, heat becomes buried under layers no surface method can reach, and the only acquisition route left runs inside the stack. The greatest value comes when the cooling is co-designed with the silicon, built with the same processes that form and release the stacked sheets so the heat is taken at its source. Channel dimensions hold this at the scale of interleaved cooling tiers rather than the sheets themselves, and it remains a research frontier, but it is the point where the acquisition problem and the chip suppliers' hold on its solution both reach their limit, asking the parties on both sides of that boundary to act as one. Microfluidics is the least deployed and among the most promising of these approaches, and it may have significant impact on how far cold plate, immersion and two-phase can ultimately go.

Microfluidics does not compete for a place in the toolbox. It extends how far every tool in it can reach.

9. The economics of cooling

The earlier sections defined the **economic boundary** as the point where the full cost of a method turns against it. This section is where that boundary is quantified across the cooling domains, because the choice between methods, and the case for using several at once, finally rests on cost as much as on physics.

NOTE: All figures in this chapter are modelled generically and serve a directional purpose. Each scenario is deliberately unrealistic in the same way: it assumes one hundred percent saturation of the compared technology, which no real facility will show, so the comparisons expose the TCO effects that separate the methods rather than absolute outcomes. The appendices disclose how the modelling is done; the technologies are treated identically, with harmonized assumptions and the same cost categories assigned to each.

The analysis follows the thermal path the paper segmented earlier: the cost that lands in acquisition, inside the IT envelope; the cost that lands in transport, the seam where the envelopes meet; and the cost that lands in rejection and the facility. The segmentation matters because the money is owned differently in each zone. Transport is purchase-order money, visible on day one in a single budget, while the facility layer is diffuse money on a potentially different owner's fifteen-year clock, and reliability swings the outcome between them. That asymmetry, more than the totals, is why the capital cost of liquid cooling is perceived as high even where the system as a whole is cheaper.

9.1. The case behind the numbers

The comparison is built on one declared case. That case is a construction, not a description: it corresponds to no data center, built or planned, of any author's organization or anyone else, and every price, rate and configuration in it is a declared assumption held identical across the three deployments so that the only variable is the cooling choice. The results are properties of this hypothetical case, not observations of a facility, and appendix section A.6 publishes every input so readers can substitute their own.

A fixed facility power envelope of 200 megawatts hosts, in turn, a full air-cooled, a full cold-plate and a full immersion deployment of the same current-generation AI platform: forty-five kilowatt racks of eight-GPU servers under air, the air ceiling the paper has already established, and 250 kilowatt rack-scale systems under liquid. Each deployment delivers the same IT capacity from the same envelope, so IT spend is normalized across the three and the differences that remain are the costs the cooling choice itself causes. Facility capital is annualized over fifteen years and IT capital over five, at the same rate, and energy is carried at a flat contracted rate.

Reliability enters the comparison the way the reliability chapters left it: each technology at its most defensible figure, cold plate at its design point and immersion at its modeled floor, with the ranges of those chapters carried as whiskers on the figures. The asymmetry runs conservative: the design point is cold plate's most favorable published configuration, while the floor is immersion's lower bound, with the stronger fleet evidence carried only as a whisker. The figures are produced with TCO modeling, derived from and built upon OCP-contributed models.

9.2. Cost in the IT envelope

The IT envelope carries the acquisition cost, the cooling that is built into the server and bought with it. Here the methods diverge sharply. Air adds little beyond fans. Cold plate adds plates, connectors and manifolds, and the cost climbs with every component they are extended to, with the coolant-distribution unit waiting just across the seam, which is where the **complexity debt** of the previous sections shows up as a number.

Immersion moves much of the cost out of the server and into the tank and the fluid, changing the shape of the spend and not only its size. And because acquisition turns over with the IT on a yearly cycle while its thermal target keeps shifting, the **lifecycle mismatch** adds a cost the static view misses, the retooling and redesign a fast-moving thermal spec demands.

The seam is where the technologies declare themselves. Air crosses it with almost nothing, a rack and its containment. Cold plate stations its coolant-distribution units there, and immersion moves its entire cooling identity into it, the tank carrying acquisition and transport as one product, as the thermal-path chapter anticipated. This is the money a buyer actually sees quoted, which is why the segment looms larger than it is: the transport column is the visible tenth of a comparison whose bulk sits in the facility layer.

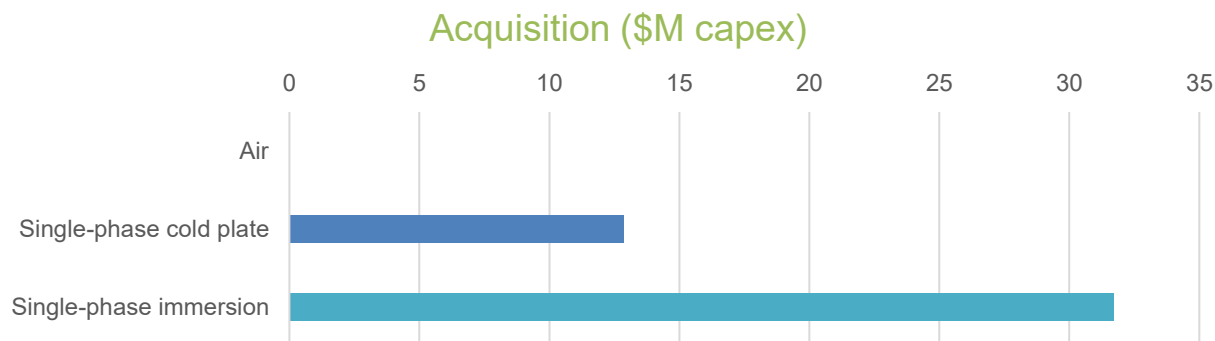


Figure 14 IT-side capex delta versus the air baseline, \$M; air IT capex 5.3B.

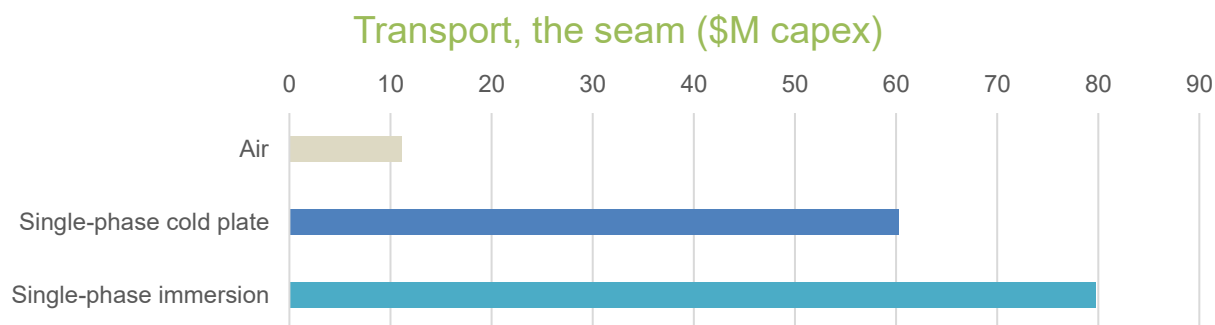


Figure 15 Capex in the transport segment, the seam, \$M; bar scales differ per segment figure.

9.3. Cost in the facility envelope

The facility envelope carries the rejection cost, the loops, the heat exchangers and the plant that move heat out of the building. This cost lives on the long lifecycle of the building and is slow and expensive to change, which is why any change that reaches this layer, a different working fluid for two-phase or the re-plumbed rejection loop behind immersion's tanks, weighs so heavily in the decision. The liquid methods spend less here for a physical reason: warmer return liquid rejects heat with smaller plant and less mechanical cooling, and denser racks need fewer square meters of building around the same power envelope, so the segment's capital falls as the cooling identity moves toward the rack. It is also where value can run the other way. The higher-grade heat that liquid methods capture can be reused, in a district network or an industrial process, turning a disposal cost into recovered value that air, venting low-grade warmth, cannot offer. Whether that value is realizable turns on one number, the temperature at which the loop returns heat at the facility boundary, and a loop tuned for the lowest cooling temperature often returns its heat with diminished economic value.

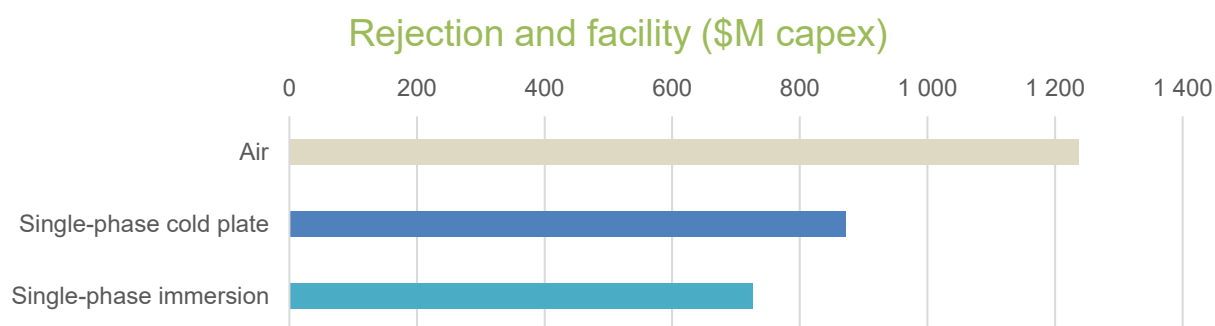


Figure 16 Capex in the rejection and facility segment, \$M.

9.4. Reliability, the swing factor

Reliability is the term that most often decides the comparison, and it lands on the IT. Every failure carries a cost, in the replacement, the labor and the lost service, and that cost runs straight into operating expense. The methods move differently. Cold plate improves on air at its published design point, because the junction dividend outweighs the handful of connectors a shipped tray adds. But every connector is a place the system can fail, and as plumbing spreads beyond the processors the advantage erodes, the fully stretched system crossing below air, exactly the trajectory the cold plate chapter derived. Immersion does the reverse, removing thermal cycling, vibration and dust and cutting the connector count, so it tends to raise mean time between failures. These two effects move the **economic boundary** in opposite directions, cold plate's inward and immersion's outward, and at scale the difference is large enough to dominate the rest of the comparison.

The figure carries that motion as ranges. The bars stand at each technology's most defensible figure and the whiskers show the direction of its unresolved uncertainty, upward in cost for cold plate as coverage spreads, downward for immersion as fleet evidence accumulates. One assumption does the heavy lifting in the failure line and is declared here: repairs outside warranty are costed at a fifth of server value, which on current AI hardware makes the failure line a repair-cost line before it is a downtime line.

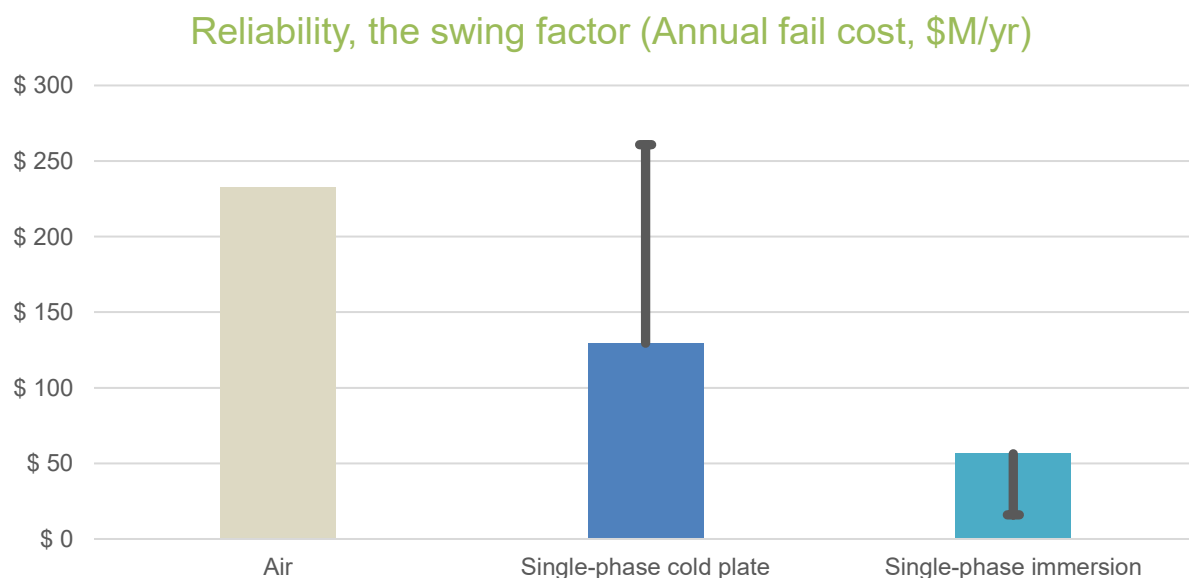


Figure 17 Annual IT failure cost by cooling method, \$M/yr; whiskers carry the ranges of the reliability chapters.

9.5. The comparative picture

Set side by side, the cost result echoes the thermal one. No method wins on every axis. Air is cheapest to acquire and the most limited. Cold plate is moderate at first and grows expensive as it is stretched, carrying both the **complexity debt** and the reliability penalty. Immersion costs more to stand up and earns it back through reliability and efficiency. The two-phase methods reach furthest and pay for it in fluids, containment and regulation. Mapped onto the three boundaries, these costs fix where each method's **economic boundary** sits and how wide its **avoidance zone** runs, and they turn the avoidance-zone diagram from an illustration into a measured statement. The operating lines follow the same shape: energy falls with the efficiency differences already established, maintenance tracks the capital it maintains, and real estate follows density, because a 250-kilowatt rack needs fewer square meters per megawatt than a 45-kilowatt one.

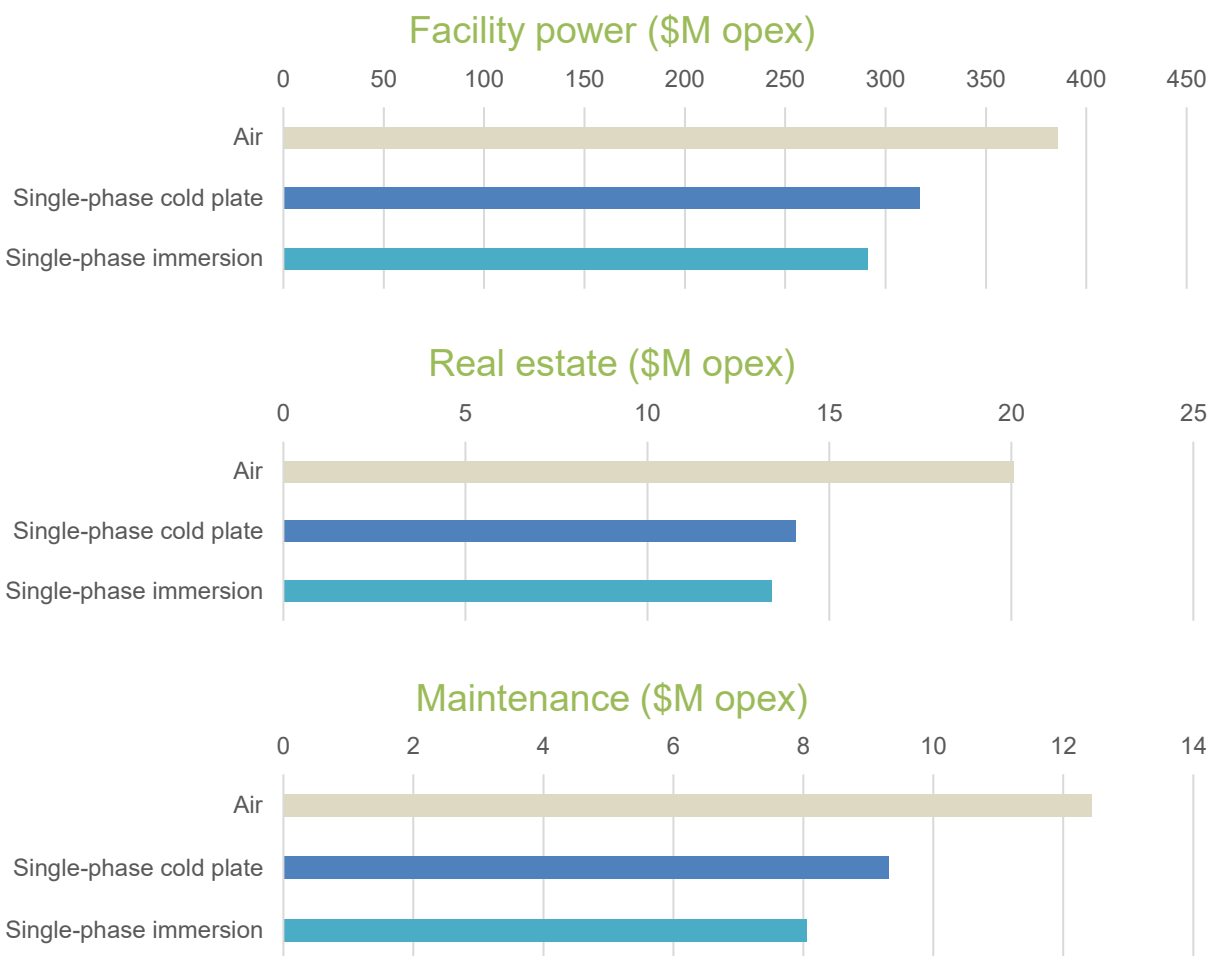


Figure 18 Annual facility opex by element, \$M/yr: energy, real estate, maintenance.

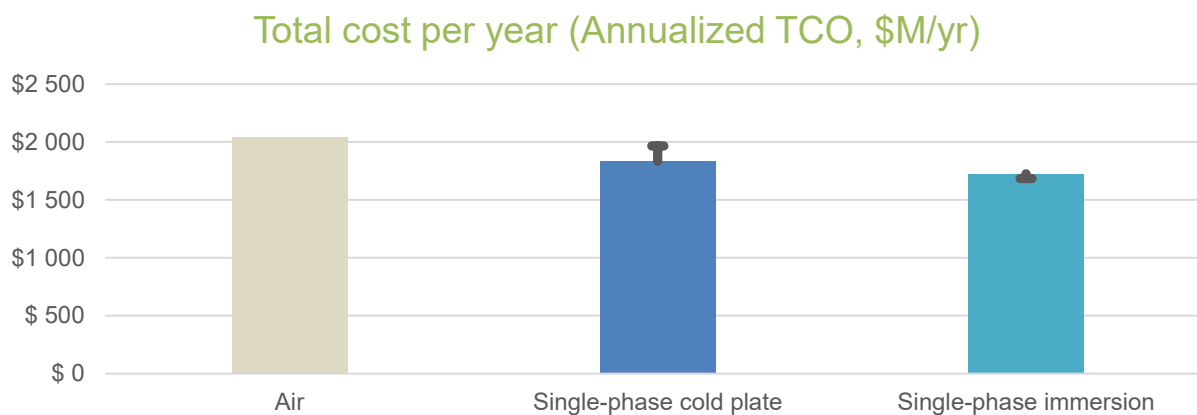


Figure 19 Total annualized cost per year by cooling method, \$M/yr; whiskers as in the failure-cost figure.

The conclusion they support is the one the physics reached by another road. The right answer is a combination, matched to the work.

On cost as on physics, no method wins on every axis. The cheapest method changes with the part of the problem, which is the economic case for using several at once.

10. The toolbox, a multi-technology roadmap

No single cooling method answers the problem. Each of the methods just surveyed is strong in one region and weak in another. Air is exhausted at the top and still right at the bottom. Cold plate owns the concentrated hotspot and falters when stretched across the system. Immersion owns the distributed volume and gives ground at the very highest flux. The two-phase methods reach furthest in pure thermal terms and are held back, one by complexity and one by its fluids. Microfluidics extends the others and does not stand alone. Set side by side, these are not rivals for a single throne. They are pieces of one puzzle, and the heat in a modern system is diverse enough to need several of them at once.

There is no winning cooling technology, because there is no single cooling problem. There is a widening set of them, and each method answers a different part.

10.1. Matching tools to jobs

The practical question is which pieces fit which job, and the answer increasingly involves combining them. The heat in a high-end AI system is not one problem but several stacked together, a handful of extreme hotspots, a large distributed load and a facility with its own limits, and the best architecture often pairs methods to match. A surface method on the very highest-flux accelerator with immersion for the holistic rest covers both the hotspot and the distributed load in a single system. Two-phase cold plate earns its place where stacked memory and rising flux defeat a single-phase plate. In-chip microfluidics, where it can be co-designed with the silicon, lifts the ceiling of whatever external method it is paired with.

Different application domains will weight these differently. Dense training clusters lean on the combination of extreme-hotspot and distributed cooling. Edge and high-throughput systems value serviceability and simplicity. Facilities under hard power and water limits move toward the densest, heat-reusing options. The point is not a single recommended stack. It is that the methods are complementary, and a serious design treats them as a toolbox matched to the work, not a contest with one survivor.

10.2. A second forcing function, the facility

One force has stayed in the background and deserves naming, because it adds a second **forcing function** alongside the chip. For most of this story the binding limit has been the ability to cool the silicon. Increasingly it is the power, the water and the land a site can command. Where a region cannot supply more electricity to a campus, the value of every watt changes, and compute packed into the available power, together with heat captured warm enough to reuse, becomes worth far more than before. This favors the denser methods and the ones that yield usable heat, and it can pull a transition forward on its own, ahead of anything happening on the die. Regulation works on the same long timescale, rewarding efficient, heat-reusing cooling while constraining particular coolant chemistries. The facility is, in effect, a second clock, slower than the chip but able to force a change the chip alone would not. This section is deliberately short for now. Facility-side specifics on grid, water, land and heat reuse will be expanded with input from the operating co-authors in the next revision.

10.3. Two businesses, one seam

The split introduced earlier between acquisition and rejection, between the IT envelope and the facility envelope, returns here as a matter of strategy. These are two different businesses, with different customers, design cycles and standards, and a capability in one does not carry into the other. Framed objectively, the cooling market is really two markets split along this line, acquisition inside the IT equipment and rejection inside the facility, each with its own buyers, design cycles and standards. Consolidation that treats them as one market, buying across the line on the assumption that a capability in one carries into the other, is buying an adjacency that the split shows does not exist, and the combined business tends to remain two parts under one roof rather than becoming more than the sum of them.

If the two envelopes are genuinely separate, durable value forms at the seam between them, the transport layer of loops, manifolds and connectors where the two worlds have to interoperate. Whoever makes it possible for one server to be cooled by many facilities, and one facility to host many kinds of server, holds a position neither envelope alone can reach. This is where standards work pays off, and where an open approach serves the whole industry and not any single firm.

Acquiring heat and rejecting heat are two businesses, not one. The lasting value is at the seam between them, not inside either.

10.4. Toward an intelligent cooling subsystem

Future cooling architectures will increasingly integrate sensing, telemetry, controls, leak detection, predictive maintenance, automated isolation and communication between IT and OT systems. Cooling evolves from passive infrastructure into an intelligent subsystem of the AI factory, one that reports its own state, anticipates failure and coordinates its actions with the compute it serves rather than simply reacting to the heat that compute produces.

This is more than instrumentation for its own sake. A toolbox assembled from several cooling technologies, as the previous section argues for, is only as trustworthy as the telemetry that tells an operator which technology is doing what, where a leak has started, and when a component is drifting toward failure. Sensing and controls are what let acquisition, transport and rejection, designed and owned by different parties on different clocks, behave as one coordinated system instead of three loosely coupled ones.

The same interoperability that lets the industry share a supply chain for hardware, developed through bodies such as **OCP**, has to extend to the telemetry and control interfaces that let that hardware be run safely and well. Readiness is not only a matter of having the plates, the fluids and the tanks on hand. It is the ability to see and act across all of them at once.

10.5. Preparing the toolbox

All of this points to one course of action, and it follows directly from the law. Because transitions are forced and arrive faster than the unprepared can answer, and because no single method will serve, the task is not to pick a winner. It is to keep the whole toolbox ready. The supply chains, the standards and the skills for cold plate, for immersion, for the two-phase methods and for in-chip cooling have to be kept alive and interoperable, so that when a **forcing function** arrives, from the chip or from the grid, the answer is a deployment and not a scramble. The cost of readying an option early is small when compared to delaying a more economic solution.

Unlike the timing of the next transition, readiness is something the industry can choose. In practice this means the chip, server and component makers designing for these methods before the economic or physics walls arrive, and not after. Doing so lets the ecosystem mature in a controlled way, gives each technology a track record and a supply chain before it is urgently needed, and turns the next transition from a scramble into a planned step. By the time a method reaches the point of necessity it is then proven instead of improvised, and the disruption is a fraction of what the move to cold plate cost.

The major challenge is that readiness is an industry problem, not a company problem. No single firm commands the full chain from chip to chiller. The final user, such as a hyperscaler, that designs a bespoke cooling interface still needs ODMs to manufacture at scale and component vendors to supply. Proprietary solutions lead to a thin supply chain and slow adoption.

Interoperability is what breaks this cycle. When a cooling method arrives with open, agreed interfaces at the boundary between IT and facility, it can be manufactured by many, deployed by many, and matured across the industry in parallel. Without interoperability, each new technology must rebuild its own supply chain from scratch when it is needed, which is exactly the scramble the preparation debt warns against. Interoperability does not slow innovation but accelerates adoption when compared to a bespoke ecosystem.

The **Open Compute Project** ([OCP](#)) exists to solve precisely this class of problem. Its model of open-source specifications developed by competitors collaborating pre-competitively, lets the industry agree on the things that do not differentiate (boundary interfaces, telemetry protocols, mechanical form factors, commissioning procedures) so that every participant can then add value with the confidence that the ecosystem will follow. OCP has already done this for server design, rack architecture, networking and power delivery. Now is the time to apply it to the cooling demands for 2028 and beyond, so industry is ready when the next forcing function arrives.

11. Epilogue: Commissioning the AI datacenter

Everything this paper has argued about cooling technology converges on a question the industry has not yet given the same rigor it gives efficiency, economics and physics: whether a finished thermal ecosystem can be brought into service reliably, on schedule, and at the pace hyperscale construction now demands. Commissioning, the step between building a system and trusting it with production load, has quietly become a constraint as real as the three boundaries described earlier, binding through schedule rather than physics. It deserves to be treated as its own subject rather than an afterthought to design, and a treatment with that rigor is a full paper in its own right. This epilogue does not attempt it. It closes the present paper by sketching, at a high level, why commissioning now belongs among the first-order constraints, and leaves the depth the subject deserves to future work.

11.1. Commissioning becomes the bottleneck

A single air-cooled rack was, for most of the data center's history, a commodity to assembly and commission. It arrived largely pre-assembled in the facility, it drew power and moved air, and validating it was a short, well-understood procedure repeated many times over. A rack assembled around a multi-technology cooling toolbox is a different kind of object. It may combine a facility fluid loop, cold plates from one supplier, a coolant-distribution unit from another, leak-sensing hardware, and firmware that has to agree with all of it about temperatures, flow rates and fault states before load is ever applied. Each of these pieces can be verified in isolation and still fail together, because commissioning a thermal ecosystem is not the sum of commissioning its parts. It is the verification that the parts behave as one system under the conditions production will actually impose.

This is the same shift in kind that the rest of the paper has described for cooling technology itself, applied to the process of bringing that technology into service. Just as a cold plate cools one component well and a whole system only with difficulty, a commissioning procedure built for a single, well-understood method does not scale cleanly to a toolbox of several methods deployed together. The complexity debt the paper has traced through a stretched cold-plate system, and the economics chapter priced, shows up again here, earlier in the lifecycle, as the burden of proving the system is safe to operate at all.

11.2. From single-vendor startup to multi-party choreography

The **agency mismatch** described earlier, between the party that feels a cost and the party that can act on it, has a direct counterpart in commissioning. No single vendor commissions an AI datacenter's cooling. The chip and server makers are responsible for what happens inside the IT envelope, the facility engineers for what happens inside the building, and the operator for the interfaces where the two have to agree. Each party can certify its own piece and still leave the seam, the manifolds, quick disconnects and coolant-distribution units introduced earlier as the meeting point of the two envelopes, inadequately tested, because no one party owns the seam outright. A leak at a connector, a mismatched flow rate between a server's expectation and a facility loop's delivery, or a control system that cannot see across the IT and facility boundary are exactly the failures a commissioning process must catch, and exactly the failures a single-vendor process is not positioned to find.

Interoperability, the standards-based answer the previous chapter proposed for supply chains, is also the answer here. A cooling method that arrives with open, agreed interfaces at the IT and facility boundary is not only easier to manufacture and deploy at scale, but easier to commission, because the acceptance criteria at the seam are shared and known in advance rather than negotiated fresh on every project. Commissioning is where the cost of a proprietary, bespoke interface is paid a second time, first in the supply chain and again on the site.

11.3. What commissioning must verify

A commissioning procedure for a modern thermal ecosystem has to answer more than whether the system holds temperature at steady state. It has to demonstrate that leak detection and automated isolation work under realistic fault conditions, not only on the bench. It has to witness the response to a change in load, not only the stable condition at full load: a steady temperature at one operating point proves that point, not the path into it. The record of that step, its size, its ramp, the excursions it produced at each scale and the time back to the agreed band, is kept in the same units for every method, agreed before the test, witnessed, and signed by the power side and the cooling side together, because at these densities a load step is as much a power event as a thermal one.

Acceptance criteria are the static half of the evidence trail; the load-step record is the dynamic half, and neither is real until witnessed. It has to verify that flow and pressure are balanced across every rack a shared facility loop serves, not just the one nearest the pump. It has to confirm that telemetry from acquisition, transport and rejection actually reaches the control system responsible for acting on it, and that a fault in one technology, a cold plate connector, an immersion tank sensor, a two-phase pressure relief path, is legible to an operator who is not a specialist in that particular technology.

And because the toolbox this paper argues for combines several cooling methods in the same system, commissioning has to prove the handoffs between them, the point where a surface method meets the volume it sits inside, or where an in-chip channel hands its heat to whatever cools the package from outside. The full procedure, item by item and criterion by criterion, is the substance of the paper this subject deserves. What this sketch establishes is its shape: a verification that is system-level and multi-technology by nature.

None of this is exotic. Every item on this list has an analogue in mature industries that commission complex, safety-relevant systems, aviation and process manufacturing among them. What is new is the pace. The **lifecycle mismatch** between IT hardware, refreshed yearly, and facility infrastructure, refreshed over decades, means the commissioning procedure itself has to be revisited nearly as often as the hardware it validates, at the same time that hyperscale build schedules leave less room to do so.

11.4. The cost of skipping ahead

The preparation debt the adoption chapter describes for supply chains has an exact counterpart in commissioning skill and procedure. An industry that defers the discipline of commissioning until a forced transition is already under way repeats, at the level of a single site, the same scramble the paper has described at the level of an entire industry. A cold plate deployment brought in ahead of trained commissioning staff and a validated procedure is not fully de-risked by the hardware being sound. The failures that follow, a manifold under-torqued, a sensor miscalibrated, a control interlock left in its default state, are not failures of cooling technology. Each is, in the end, a missing record at a known boundary: a torque log, a calibration certificate, a timed test of the interlock. They are failures of readiness, and they carry the same lesson the rest of this paper has drawn from hardware: the cost of preparing early is small next to the cost of being unprepared when the transition can no longer wait.

Treating commissioning as a discipline in its own right, with its own standards, its own trained workforce and its own place in the toolbox the industry is asked to keep ready, is not a departure from this paper's argument. It is the same argument, applied to the step where a thermal ecosystem stops being a design and starts being a system an AI factory can depend on.

A cooling method that cannot be commissioned reliably is not yet ready, whatever its physics promise. Commissioning is where the toolbox is proven, not merely assembled.

12. Conclusion

The thread of this paper is that cooling has become a first-order constraint on computing, at the moment the usual escape routes have closed. No new device waits to reset the power curve as CMOS once did, and the paradigms put forward as successors add to the conventional processor without the ability to lower its power. The heat has to be dealt with, and dealing with it now reaches into the server, the building and the supply chain.

How the industry deals with it is governed by the **least-disruption law**. A new method is adopted only when the old one can no longer do the job, because each new method carries a dependency the ecosystem has good reason to defer, and liquid cooling carries the heaviest dependency of all. Air was forced over its **physics boundary** in a single generation. Cold plate has a physics boundary of its own, projected near 4,000 watts for a reticle-class die on current silicon roadmaps. That wall is far in watts but near in time: frontier package power doubles roughly every two years, and announced parts are already climbing toward its neighborhood. What keeps the transition gradual is not the distance to the wall but the width of the avoidance zone in front of it, across which cold plate will be stretched as its economics turn. Each transition leaves a **preparation debt** for the industry that waited, and the move to cold plate has already shown what that bill looks like.

No single method answers the diverse and growing problem, so the destination is a toolbox. Air, cold plate, immersion, the two-phase methods and microfluidics each hold part of the answer, and the strongest designs combine them, matched to the heat, the site and the moment. Two facts sharpen the choice. Air, after decades of refinement, is near the end of what it can do. Liquid cooling is early, and its limits move with every advance in fluids, surfaces, flow control and integration. The methods that look most constrained today are the ones with the most room left to grow.

The decisive question is not which technology wins. It is whether the supply chains, the standards and the skills for the next one are ready before the next technology is needed. That readiness is the one thing in this story the industry fully controls, and it is the most useful thing it can do.

That is the work this paper is written to start, not finish. The three boundaries, the least-disruption law and the toolbox they point to are offered as a shared vocabulary, not a settled answer, and the boundary, adoption and cost figures are drawn from published data and a declared cost methodology precisely so they can be checked. Where ranges remain, in the reliability brackets above all, they are placeholders for the fleet data the community is better positioned to supply than any single contributor.

The next step belongs to the OCP Open Systems for AI Strategic Initiative and the wider community it serves: pressure-test the boundary and cost models against real deployments, contribute the reliability and commissioning data that will sharpen Section 8 and Section 11, and begin the standards work, on telemetry, on IT-to-facility interfaces, on commissioning procedure, that Section 10 and the epilogue argue has to be ready before the next forcing function arrives. **Open Compute Project** exists to carry that kind of work forward in the open, and this paper is an invitation to do so.

Air cooling is finishing. Liquid cooling is just beginning. No single method wins, and the only real choice is to be ready for all of them.

Appendix A Detailed source material for graphs in this paper

Basis flags used throughout this appendix: (p) published or derived from published figures; (a) declared assumption or author assessment; (r) reported figures not yet vendor-published; (c) value digitized from a published curve; (m) modeling output.

A.1 Figure 1 - Module-level heat flux

The figure's dataset is reproduced here in full. It plots module-level heat flux, module power over substrate footprint, in W/cm^2 , and it follows the curated dataset published by IBM's own thermal engineering leads, Chu, Simons, Ellsworth, Schmidt and Cozzolino, the people who ran the machines the figure describes. Where an independent published anchor exists for a machine's power or geometry, it is listed per row. Values digitized from the published curve are marked (c); anchored or computed values are marked (p). The rows through the ES/9000 form the bipolar series, liquid-cooled in every high-end machine from 1964 onward; the remaining rows form the CMOS series, the air-cooled return and the same climb begun again.

Machine	Year	W/cm ²	Basis	Anchor and public source
Vacuum tube era	1955	0.2	(c)	Chu 2005 chart, p. 8; cseweb.ucsd.edu/~kuan (open PDF)
IBM 360	1964	0.3	(c)	First water-cooled product 1964, CDU 1965; Schmidt, Electronics Cooling, Aug 2005
IBM 370	1971	0.5	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM 3033	1978	0.9	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM 3081 TCM	1980	2.0	(c)	118 chips, 90 mm substrate, up to 300 W; Blodgett and Barbour, IBM J. Res. Dev. 26(1) 1982; doi.org/10.1147/rd.261.0030
Fujitsu M380	1981	1.5	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
CDC Cyber 205	1982	2.5	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM 4381	1983	2.2	(c)	Oktay and Kammerer, IBM J. Res. Dev. 26(1) 1982; doi.org/10.1147/rd.261.0055
NEC SX-2 (LCM)	1983	1.5	(c)	Watari and Murano, IEEE Trans. CHMT 8(4) 1985; doi.org/10.1109/TCHMT.1985.1136535
Hitachi S-810	1984	1.0	(c)	Water-cooled supercomputer; en.wikipedia.org/wiki/HITAC_S-810
IBM 3090 TCM	1985	4.0	(p)	525 W module; Computer History Museum catalog 102741297; Simons, IEEE Trans. CPMT-A 18(4) 1995
Fujitsu M-780	1985	4.7	(c)	Water jet impingement; Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
NTT module	1988	6.0	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM 3090S TCM	1989	7.0	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
Fujitsu VP2000	1989	7.8	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM ES/9000	1990	13.0	(p)	End of bipolar water cooling; Delia et al., IBM J. Res. Dev. 36(4) 1992; ieeeexplore.ieee.org/document/5389685
Honeywell DPS-88	1986	0.8	(c)	Chu 2005 chart (early CMOS-adjacent low band)
IBM RX3	1993	0.6	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM RX4	1996	1.7	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM RY4	1997	1.9	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
Pentium II (Deschutes)	1998	0.6	(c)	TDP 27-43 W; Intel datasheet (chipdb.org mirror)
IBM RY5	1998	6.3	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
Pulsar (RS64-III)	1999	2.4	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855

Date: October 2026


This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

IBM RY6	1999	2.9	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
IBM RY7	2001	4.3	(c)	Chu 2004 review; doi.org/10.1109/TDMR.2004.840855
Itanium 2 (McKinley)	2002	9.0	(p)	130 W TDP; Intel documentation; cpu-world.com/CPUs/Itanium2
Pentium 4 Xeon DP	2003	11.0	(p)	~90-103 W over ~31x31 mm spreader, computed; Intel datasheet

Table 5 Module-level heat flux source data, bipolar and CMOS series, with per-machine basis flags and public anchors.

What the figure carries

Three annotations are carried on the figure itself. Water cooling begins in 1964 with the IBM 360 era, the first water-cooled product, the coolant-distribution unit following in 1965. Bipolar water cooling ends in 1990 with the ES/9000 at roughly 13 W/cm². And the reset between them is measured: the shift from bipolar to CMOS bought the industry roughly ten to twelve years of module power, Chu's estimate at twelve years and Schmidt's at ten, which is the citable measure of what a technology reset is worth.

Sources

The primary works behind the dataset are carried in the References list of the reader's reference appendix: the Chu 2004 review and 2005 seminar that publish the curve itself, the IBM Journal papers that anchor the individual machines, and the museum record for the ES/9000 module. All source records have been click-checked: the Chu review and the 1982 IBM Journal DOIs resolve to their papers, the NEC and Simons papers to their IEEE records, and the ES/9000 cooling paper is carried via its IEEE Xplore record. Intel's legacy datasheets are no longer hosted by Intel and are carried by the chipdb.org mirror.

A.2 Figure 2 - Leading chip TDP and flux trends, CPUs, GPUs and AI accelerators

This data set consists of package power at introduction for the leading server part of each generation, CPUs, GPUs and AI accelerators, 2008 through 2026. CPU values are the top-SKU TDP of each generation; accelerator values are the flagship module's rated power. On the package-power side, every shipped part is (p), taken from vendor specifications, Intel ARK, AMD product pages and NVIDIA datasheets, and the one exception is deliberate: the 2026 accelerator point is the announced Vera Rubin generation at reported rather than vendor-published power, and it carries that qualifier in the figure caption. The envelope curves on the figure are guides to the eye that end on data, and the CPU curve ends at its last shipped part, because a guide that runs past its evidence would be a claim.

Class	Vendor	Generation or part	Year	Package power (W)	Die area (mm ²)	Flux (W/cm ²)	Basis
CPU	Intel	Nehalem	2009	135	263	51.3	(p) vendor spec
CPU	Intel	Westmere	2010	135	248	54.4	(p) vendor spec
CPU	Intel	Sandy Bridge	2012	150	434	34.6	(p) vendor spec
CPU	Intel	Ivy Bridge	2013	150	541	27.7	(p) vendor spec
CPU	Intel	Haswell	2014	160	661	24.2	(p) vendor spec
CPU	Intel	Broadwell	2016	165	456	36.2	(p) vendor spec
CPU	Intel	Skylake	2017	205	698	29.4	(p) vendor spec
CPU	Intel	Cascade Lake	2019	205	698	29.4	(p) vendor spec
CPU	Intel	Cooper Lake	2020	250	698	35.8	(p) vendor spec
CPU	Intel	Ice Lake	2021	270	698	38.7	(p) vendor spec
CPU	Intel	Sapphire Rapids	2023	350	750	46.7	(p) vendor spec
CPU	Intel	Emerald Rapids	2023	400	750	53.3	(p) vendor spec
CPU	Intel	Granite Rapids	2024	500	2276	22.0	(p) vendor spec
CPU	AMD	Shanghai	2009	137	258	53.1	(p) vendor spec
CPU	AMD	Magny-Cours	2010	140	692	20.2	(p) vendor spec
CPU	AMD	Interlagos	2011	140	630	22.2	(p) vendor spec
CPU	AMD	Abu Dhabi	2012	140	630	22.2	(p) vendor spec
CPU	AMD	Naples	2017	200	512	39.1	(p) vendor spec
CPU	AMD	Rome	2019	280	1016	27.6	(p) vendor spec
CPU	AMD	Milan	2021	280	1016	27.6	(p) vendor spec
CPU	AMD	Genoa	2022	400	1165	34.3	(p) vendor spec
CPU	AMD	Turin	2024	500	1165	42.9	(p) vendor spec
GPU	Intel	Ponte Vecchio (Max 1550)	2023	600	2330	25.8	(p)/(r) total silicon 2,330 mm2, 47 tiles, TechInsights/reported
GPU	AMD	Instinct MI100	2020	300	750	40.0	(p)/(r) die size ~750 mm2, CDNA1, teardown/TechPowerUp reporting
GPU	AMD	Instinct MI250X	2021	560	1448	38.7	(p)/(r) 2x 724 mm2 GCDs (Aldebaran), reported
GPU	AMD	Instinct MI300X	2023	750	1017	73.7	(p)/(r) total silicon ~1,017 mm2, teardown reporting
GPU	AMD	Instinct MI325X	2024	1000	1017	98.3	(p)/(r) same silicon as MI300X (1,017 mm2)
GPU	AMD	Instinct MI355X	2025	1400	2380	58.8	(p)/(a) total silicon implied from source flux; CDNA4 multi-die
GPU	AMD	MI455X	2026	2000	2600	76.9	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
GPU	NVIDIA	GT200	2008	190	576	33.0	(p) vendor spec
GPU	NVIDIA	Fermi	2010	250	529	47.3	(p) vendor spec
GPU	NVIDIA	Kepler	2012	235	561	41.9	(p) vendor spec
GPU	NVIDIA	Pascal	2016	300	610	49.2	(p) vendor spec
GPU	NVIDIA	Volta	2017	300	815	36.8	(p) vendor spec
GPU	NVIDIA	Ampere	2020	400	826	48.4	(p) vendor spec
GPU	NVIDIA	Hopper	2022	700	814	86.0	(p) vendor spec

GPU	NVIDIA	Blackwell	2024	1200	1600	75.0	(p) vendor spec; 1,200 W GB200 configuration (1,000 W standalone)
GPU	NVIDIA	Blackwell Ultra	2025	1400	1600	87.5	(p)/(r) vendor spec, reported die
GPU	NVIDIA	Vera Rubin	2026	2300	1700	135.3	(a) announced; reported 1,800-2,300 W, Max-P taken (supply-chain reporting)
Accelerator	Intel	Gaudi 3	2024	900	1170	76.9	(p)/(r) vendor white paper: 900 W air-cooled OAM max; two TSMC N5 compute dies ~880 mm ² plus ~290 mm ² PHY and D2D, ~1,170 mm ² total
Accelerator	Google	TPU v4	2021	275	400	68.8	(r) ADOR paper (arXiv:2503.04253) Table I: 275 W max TDP, 400 mm ² ; Google publishes no TDP (ISCA 2023 reports 192 W max measured)
Accelerator	Google	TPU v5e	2023	197	200	98.5	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
Accelerator	Google	TPU v5p	2023	300	400	75.0	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
Accelerator	Google	TPU v6e (Trillium)	2024	200	300	66.7	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
Accelerator	AWS	Trainium 1	2022	300	400	75.0	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
Accelerator	AWS	Inferentia 2	2023	180	300	60.0	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure
Accelerator	AWS	Trainium 2	2024	700	800	87.5	(a)/(r) implied from source flux
Accelerator	Groq	GroqChip v1	2020	300	725	41.4	(r) ADOR paper (arXiv:2503.04253): 300 W max TDP, 725 mm ² , 14 nm; ~185 W operating reported (SemiAnalysis)
Accelerator	Groq	GroqChip v2	2024	300	400	75.0	(a) pair chosen to reproduce the contributed source flux; no vendor disclosure; published sources cover the 14 nm TSP only
Accelerator	Qualcomm	100 Ultra	2023	150	300	50.0	(p)/(a) die total implied
Accelerator	Microsoft	Maia 100	2023	500	820	61.0	(p) vendor spec

Table 6 Leading chip package power and die-level heat flux, CPUs, GPUs and AI accelerators, 2008–2026, with per-device basis flags; flux computed as TDP over total die area.

Date: October 2026

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



The flux figures draw on the same device set, extended with total die areas, and are folded into the single table above. Basis flags per point: (p) vendor-disclosed values, (r) reported or teardown figures, (a) TDP and die-area pairs reconstructed to match the contributed source dataset where no vendor disclosure exists. Flux is computed as TDP over total die area. All devices are catalogued in the dataset and announced parts are carried as announced. Reconstructed points are carried transparently pending vendor disclosure.

A.3 Figure 5 - Distributed non-xPU heat

The model and its validation

The model behind the figure is the representative AI platform of each era: the 8-GPU DGX/HGX-class node through 2024, and the rack-scale platform from 2026, when rack scale became the typical deployment. Values are normalized per rack unit of chassis height. Three exclusions keep the measure honest to its purpose: xPU heat is excluded because the figure describes everything else, HBM is excluded as concentrated load, and fans are excluded as cooling overhead rather than IT load. From 2026 the network class carries a declared allocation of in-rack scale-up fabric, the NVL-class switch trays, per compute rack unit.

What disciplines the composition is the datasheet residual. Each anchor system's published maximum power, minus its xPU rated powers, minus a declared fan estimate of five to ten percent for an air-cooled chassis, leaves a residual the classes must sum to, so the envelope cannot be tuned to the narrative. Power conversion loss is the closing term against that residual, since PSU, VRM and distribution losses are not publicly itemizable. The residual check column shows the datasheet-derived value at each anchor year; 2026 shows the in-node residual of 430 W per rack unit with the fabric allocation of roughly 123 W per rack unit on top.

Year	Era anchor	Memory	Network and optics	Storage	Power conversion	Misc.	Total non-xPU	Residual check
2016	DGX-1 (P100), 3U, 3.2 kW	27	27	7	54	12	127	123
2018	DGX-2 (V100, in-chassis NVSwitch), 10U, 10 kW	29	90	12	168	20	319	319
2020	DGX A100, 6U, 6.5 kW	32	100	20	198	25	375	375
2022	H100 transition (interpolated)	40	108	17	180	25	370	-
2024	DGX H100/H200, 8U, 10.2 kW	48	115	15	160	25	363	363
2026	GB300 NVL72 rack-scale era, ~140 kW rack	130	230	25	150	35	570	568
2028	VR200 / Rubin Ultra fleet era, ~250-300 kW rack (announced)	230	450	30	280	60	1050	1000
2030	Kyber-class era, ~600 kW rack (extrapolated)	350	780	40	450	80	1700	1800

Table 7 Distributed non-xPU heat per rack unit by class and era, W/RU, with datasheet-residual checks.

Class	Anchor and typical power	Basis	Public sources
System anchors	DGX-1 3.2 kW 3U; DGX-2 10 kW 10U; DGX A100 6.5 kW 6U; DGX H100 10.2 kW 8U; DGX B200 14.3 kW 10U	(p)	NVIDIA DGX datasheets and user guides, nvidia.com, docs.nvidia.com
Memory	DDR4 512 GB (DGX-1) to 2 TB (A100); DDR5 2-3 TB (H100/B200); ~5-6 W per DDR4 RDIMM, ~10-14 W per DDR5 RDIMM	(p)	Micron TN-40-07 and DDR5 whitepapers, micron.com; DGX datasheets for capacities
Network and optics	1:1 NIC:GPU rail-optimized from ~2020; ConnectX-4 to -8 ~18-40 W; 2x BlueField-3 at 75 W; NVSwitch in-chassis from DGX-2; optics 4 W (QSFP28) to ~15-16 W (800G); 1.6T-era and fabric allocation (a)	(p)/(a)	NVIDIA product briefs; Coherent transceiver specifications; CES 2026 Rubin coverage (NVLink-6 switch, liquid-cooled)
Storage	4-8x NVMe; Gen4 ~12-15 W to Gen5 ~19-25 W	(p)	Kioxia CM6/CM7 datasheets; Solidigm datasheets
Power conversion	Closing term against the datasheet residual; PSU 92-96 percent (80 PLUS Titanium), OCP ORv3 97 percent peak, multiphase VRM ~90 percent	(a)	80 PLUS program data; opencompute.org; Infineon and TI application notes
Miscellaneous	BMC, board, PCIe retimers	(a)	ASPEED documentation; retimer vendor datasheets

Table 8 Class anchors and public sources for the distributed-heat composition.

What the composition shows

Two findings come out of the composition rather than going into it. Rack-scale NVL-class compute trays concentrate roughly 1.5 to 2.5 times these per-RU values, with 1U trays, SOCAMM memory and dense fabric, and such configurations cross one kilowatt per rack unit inside the 2027 to 2029 window, while the base node series crosses around 2031. And networking becomes the largest class from about 2026, near 45 percent of the stack by 2030, an outcome of one-to-one NIC-to-GPU counts, DPUs, optics watts per port and the fabric allocation, not an assumption placed there.

A.4 Figure 8 - The air efficiency boundary model

The fan-share model is quadratic in device power with the share defined as fan power over the combined device and fan power: a base share of five percent plus a rise of ten points calibrated at 600 W, read against a chassis airflow limit taken at 1,000 W. The knee is anchored at the calibration point, where the share reaches base plus rise; the early-trend reference is the secant through the model's 100 W and 300 W points, continued; and the shaded region begins at the knee. All three constructions are parameter-derived and move with the model. The floor sits near five percent, rising with the square of device power per the fan affinity laws, and the modeled span is consistent with measured shares of five to thirty percent across shipped systems.

A.5 Figure 9 - Cold Plate inflection and the distributed-load trend

The adoption series

The adoption series measures the share of top-tier accelerator TDP cooled by cold plate. All values are (a), the authors' industry assessment corroborated by analyst liquid-cooling penetration reporting from TrendForce and Omdia, and values from 2027 onward are projections. The shape is the finding: single-digit penetration through the first half of 2024, while the 700-watt class still shipped predominantly on air; the knee arriving with the liquid-only rack-scale generation from late 2024; above ninety percent by mid-2026. Under ten percent to above ninety inside roughly twenty-four months is the forced, overnight default of the adoption chapter, drawn empirically.

Year	Share of top-tier accelerator TDP on cold plate (%)
2018	0.5%
2019	1%
2020	2%
2021	3%
2022	5%
2023	7%
2024	8%
2024.5	12%
2025	35%
2025.5	60%
2026	78%
2026.5	91%
2027	95%
2027.5	96%
2028	97%

Table 9 Cold plate adoption: share of top-tier accelerator TDP on cold plate, by year.

The rack-power series

The rack-power series tracks leading AI rack power demand over time, with the heat left behind for air computed as rack power times a residual share. The residual parameter is fifteen percent, at the conservative edge of the published capture-ratio range of fifteen to twenty-eight percent, anchored by the Childdyne 72/28 reference design with the Vertiv-NVIDIA blueprint family toward the higher capture end. The air-cooled rack practical ceiling is carried at 45 kW, extendable toward 60 to 80 kW with rear-door exchangers, which buy rejection only. xPU power is defined as the complement, rack power minus the heat left behind, so the whole split is governed by a single residual parameter, and the roughly 85/15 share holds back through the early rows: the DGX-1 was already about 83 percent xPU by TDP arithmetic.

Year	Leading AI rack power (kW)	xPU power (kW)	Heat left behind (kW)	Basis
2016	13	11.1	2.0	(a) representative multi-node AI rack
2018	20	17	3	(a)
2020	30	25.5	4.5	(a) DGX A100 pod era
2022	40	34	6	(a) HGX H100 air-rack era
2024	132	112.2	19.8	(p) GB200 NVL72, published
2026	140	119	21	(p) GB300 NVL72 class, reported
2027	250	212.5	37.5	(p) VR200 NVL144, reported
2028	600	510	90	(p) Rubin Ultra Kyber-class, announced
2030	1000	850	150	(a) announced megawatt-rack direction

Table 10 Leading AI rack power and the heat left behind for air, by year, with basis.

The anchors are published and announced rack powers: GB200 NVL72 at 120 to 132 kW in vendor datasheets, the GB300 class near 140 kW as reported, VR200 NVL144 at roughly 230 to 250 kW as reported, the Kyber-class Rubin Ultra near 600 kW as announced, and 2030 carried (a) per the announced megawatt-rack direction. Read against the air ceiling, the cold plate inflection is the 2024 step, and the whole-volume window opens where the heat left behind alone approaches an entire air-cooled rack, at 250 to 300 kW racks in 2027 to 2029, equivalently about one kilowatt per rack unit. The windows are bands, not points, and both series are reproduced in full in the tables of this appendix, with the residual carried as a declared value.

A.6 The economics of cooling

Lineage and the comparison basis

The figures of the economics chapter are produced with Promersion's TCO modeling, derived from and built upon OCP-contributed models. The OCP-contributed cost-model lineage supplies the facility cost framework; Promersion's methodologies extend it with power-envelope normalization, reliability integration, and the thermal-path allocation used in this paper. The contributed base models are listed in the References; they do not themselves support cross-technology comparison, and that capability is the derivation, which holds facility-side infrastructure elements identical across the cooling circuits. The model computes the full annualized cost of a complete single-technology facility, air-cooled, single-phase cold plate and single-phase immersion, under one declared input set. No other cost source is used, and every figure in the chapter traces to the same two model runs, differing only in the reliability multipliers; nothing is re-tuned per figure. The comparison is computed for one declared scenario, the parameter set declared in this appendix and claims no generality beyond it; the sensitivity treatment of the economics chapter and the contributed base models in the References are the standing invitation to recompute under other assumptions.

One fixed facility power envelope hosts each technology in turn. Air IT load is the envelope divided by the air baseline PUE; the liquid technologies convert their overhead reductions, fans removed, the air path narrowed, the heat-rejection path simplified, into additional delivered IT within the same envelope. This is the comparison a fixed grid connection imposes in practice: same envelope, same delivered work, and the differences that remain are the costs the cooling choice itself causes. Every technology is priced by the same formulas with technology-specific parameters; there are no per-technology special cases in the cost structure.

What the model covers

Facility capital covers the cooling plant, coolant-distribution units, immersion solutions including the tank cooling system, racks, rack fit-out, containment, power and data infrastructure, and real estate. Operating cost covers energy at the flat contracted rate, real estate, and maintenance carried as fixed shares of the capital lines being maintained, applied identically across technologies. IT capital covers servers and network switches, plus the cooling-attributable IT deltas, cold plates and in-chassis plumbing for cold plate, immersion-ready packaging for immersion. IT failure cost is modeled from reliability as described below. Capital is annualized by the standard payment formula at one rate, facility capital over fifteen years and IT capital over five, reflecting roughly three IT refreshes per facility life; annualized capital plus operating cost gives cost per year.

The cost of IT

Server capital is anchored at the air rack and normalized per kilowatt of delivered IT capacity, so each technology buys the same computing for the same money: higher-density liquid racks cost proportionally more per rack while facility-level IT spend stays within 0.6 percent across the three technologies. This construction removes IT price from the comparison by design and leaves only the cooling-attributable deltas, which are carried explicitly and shown in the figures. The chapter's differences therefore come from the facility, the seam, and reliability, not from asserted IT price differences.

Reliability integration

Failure events are computed from a per-server MTBF for the air baseline, derived from the paper's own component failure rates, and scaled per technology by the multipliers derived in the reliability chapters, cold plate at its published design point with the system-wide case as stress, immersion at its modeled floor with the fleet observation as ceiling. Each event is costed as the warranty-dependent repair share of server value plus downtime at the declared rate for the declared time to repair, with events accumulated over the five-year IT life. Each technology enters at its most defensible figure; the ranges appear as whiskers on the figures with the direction of each technology's unresolved uncertainty, and the fleet-observed case is never presented standalone. The out-of-warranty repair share is the load-bearing assumption of the failure line and is disclosed with the results.

Parameter	Value	Basis
Declared inputs		
Total facility power envelope	200 MW	(a) declared case
Air baseline PUE	1.3	(a) declared case
Energy cost	0.220 per kWh, flat	(a) contracted rate
Grid carbon intensity	0.74 kg CO ₂ per kWh	(p) regional grid reference
Air rack	45 kW, three 15 kW eight-GPU servers	(p) the air ceiling established in the air chapter
Liquid rack, cold plate and immersion	250 kW	(p) announced rack-scale generation
Server capital per air rack	1,500,000	(a) anchored to reported market pricing
Network switches	2 per rack at 30,000 each	(a)
xPU sockets per server	8	(p)
IT capital annualization	5 years at 6 percent	(a) convention, roughly three IT refreshes per facility life
Facility capital annualization	15 years at 6 percent	(a) convention
Air per-server MTBF	4.9 years	(p) the paper's component failure rates: eight GPUs at 2.5 percent plus 0.5 percent fan tray
Reliability multiplier, cold plate	1.45 design point; 0.72 system-wide stress case	(p) derived in the reliability chapters
Reliability multiplier, immersion	3.4 modeled floor; 12 fleet-observed ceiling	(p) derived in the reliability chapters
Warranty and repair	1 year; repair at 5 percent of server value in warranty, 20 percent out of warranty	(a) current vendor terms; the out-of-warranty share is the load-bearing assumption of the failure line
Downtime	250 per server-hour; time to repair 7 h air and cold plate, 8 h immersion	(a)

Table 11 Declared inputs of the cost comparison, with basis flags.

Thermal-path allocation

Acquisition carries the IT-side capital deltas versus air. Transport carries racks, rack fit-out, containment, coolant-distribution units, and immersion solutions including the tank cooling system; CDUs are megawatt-class units serving seven liquid racks each, costed per rack from unit cost and capacity, with maintenance prorated by capital share. Rejection and facility carries the cooling plant excluding CDUs and tanks, plus power and data infrastructure. Two allocation choices had a direction and both were resolved against the newer technologies' favor: air's room airflow equipment cannot be separated from plant in the cost structure and is kept in the facility layer, understating air's transport, and cold plate retains the air-side containment carry-over by design, since its unplated load still requires the air path.

Model variables and multipliers applied

For direct scrutiny of the inputs, the engine's harmonized variables are reproduced below. They originate in the TCO workstream of the OCP Cooling Environments Immersion sub-project and have been part of the contributed models for many years; community vetting of the base models remains ongoing work in that workstream. The values listed are taken from the originally contributed model, TCO modelling v0, whose ready-to-use multipliers the proprietary engine expands on; the References link the contributed models that source every listed variable. Where a variable applies to every technology that uses the element, it acts as a shared multiplier: adjusting it moves the absolute figures, not the standing of one technology against another, and the reader is invited to substitute their own values. The few technology-assigned entries are visible in the table and carried conservatively.

Engine variable	Value
FWS ΔT with liquid cooling	8°C
Fan power per 100 L/s internal air transport	55 W
ΔT air	20°C
Pump power per 0.03 L/s internal water transport	1 W
ΔT free air tolerated	15°C
Water additive specific heat	2,600 J/kgK
Water additive density	1.13 kg/l
Water additive mixture percentage	25.0%
Power loss (switchgear, PDUs, DCIM, transport)	4.0%
UPS loss	4.0%
Average PSU efficiency	88.0%
Floorspace per m ² (incl. security, fire suppression)	\$500
Racks incl. PDU and cables, capex per rack	\$2,500
Cold plate manifold capex per rack	\$3,000
250 kW Immersion tank reference (Incl. CDU/infra), per tank	\$150,000
Immersion rack length, per rack	2.50 m
Immersion rack width or depth, per rack	1.00 m
Power / data infrastructure per kW	\$1,400
Power / data infrastructure per air server	\$150
Liquid infrastructure per cold plate server	\$500
Power / data infrastructure per immersed server	\$100
Aisle containment / raised flooring per rack	\$750
Battery per kW	\$1,300
Genset per kW	\$1,000
Chiller per kW	\$2,000
Adiabatic add-on and water conditioning per kW	\$1,500

Dry-cool-only, standard low temperature, per kW	\$1,250
Dry-cool-only, standard high temperature, per kW	\$625
Power / data infrastructure maintenance (% of capex)	1.0%
Fit-out maintenance	1.0%
Cooling infrastructure maintenance (% of capex)	1.0%

Table 12 Harmonized engine variables and multipliers of the contributed base models, applied identically across technologies.

Consistency and bounds

The construction is verifiable from the published figures alone. The air case is identical in both model runs, so reliability assumptions cannot leak into capital. Capital lines are run-invariant, so the difference between the two runs' totals equals the difference in failure cost exactly, per technology. The three path segments sum to total facility capital plus the IT-side deltas, and the operations segments sum to total operating cost. IT capital agrees across technologies within 0.6 percent by construction. Energy and carbon use one rate and one grid factor for all three technologies. Any of these identities can be re-added from the figures as published, and the itemized blocks behind them are preserved in the project archive.

The scope is bounded deliberately: single-technology facilities only, no hybrids; a flat energy price; no fluid replenishment line; no phased build-out; no discounting of operating cost. Each exclusion bounds the claims; none changes the direction of a comparison, and where an approximation had a direction, it is named above together with the side it favors.

Appendix B Thermohydraulic Limits of Single Phase Cold Plate

This appendix derives the single-phase physics boundary stated in the cold plate chapter. It evaluates five representative cold-plate geometries at a common reference condition and shows that the approximately 4 kW figure emerges from the convergence of thermal, hydraulic and mechanical constraints rather than from any single one of them. Its conclusion is the mechanism of the boundary, not its repeal: beyond the envelope, further scaling requires changing the architecture.

As AI accelerator package power approaches and exceeds several kilowatts, single-phase direct-to-chip liquid cooling increasingly becomes a coupled thermo-hydraulic and mechanical design problem. Increasing convective heat-transfer coefficient through smaller channels can raise local heat-transfer capability, but it also increases flow velocity and pressure drop. At the same time, single-phase water cooling is fundamentally constrained by the sensible heat capacity of the coolant: for a specified flow rate, the coolant temperature must rise in proportion to the absorbed heat.

This analysis evaluates five representative copper cold-plate geometries under a common operating condition of 5 LPM water flow, 35°C inlet temperature, and a 30-mm flow length. The results demonstrate that progressively finer conventional microchannels can increase nominal heat-removal capability from hundreds of watts to approximately 4 kW, but only at rapidly increasing hydraulic cost. A 0.10-mm channel geometry reaches approximately 4.0 kW of calculated heat-removal capability but requires approximately 16.96 bar of pressure drop and produces an 11.52°C coolant temperature rise, exceeding the specified 10°C thermal budget. A high-aspect-ratio skived-fin architecture provides a substantially different tradeoff. By increasing wetted area rather than relying primarily on extreme fluid velocity, the reference skived-fin design dissipates approximately 2.30 kW at only 0.386 bar pressure drop and a 6.64°C coolant temperature rise.

The results identify an approximate 3–4 kW transition region for the specified operating envelope in which conventional single-phase cold plates increasingly struggle to satisfy thermal, hydraulic, mechanical, and infrastructure constraints simultaneously. The relevant engineering boundary is therefore a constrained design envelope rather than a single universal power limit.

B.1 Introduction

AI accelerator thermal design is entering a regime in which package power is no longer adequately characterized by the thermal capabilities of conventional server cooling technologies. As accelerator thermal design power (TDP) approaches and exceeds 2,000 W, and future packages move toward several kilowatts, the cold plate becomes a critical component of the complete power-delivery and thermal-management architecture. Single-phase liquid cooling remains attractive because water has high specific heat capacity, high thermal conductivity, low cost, and mature infrastructure. However, increasing package power creates competing requirements.

A cold plate must simultaneously provide:

1. sufficient heat-transfer coefficient to maintain acceptable silicon temperature;
2. sufficient wetted area to transfer the required heat;
3. sufficiently low pressure drop to remain compatible with pumps, manifolds, and quick-disconnects;
4. sufficiently low coolant temperature rise to prevent excessive spatial thermal gradients;
5. sufficient mechanical stiffness to maintain package-to-cold-plate contact;
6. acceptable reliability and manufacturability.

These requirements are strongly coupled. Reducing channel dimensions can increase heat-transfer coefficient, but it also increases flow velocity and pressure drop. Increasing flow rate reduces coolant temperature rise, but pressure drop generally rises rapidly with flow rate.

Increasing channel surface area can improve thermal performance without requiring extreme velocity, but it increases manufacturing complexity and can introduce its own hydraulic penalties. The central question is therefore not simply “How much heat can a single-phase cold plate dissipate?”. Rather, the relevant engineering question is:

How much heat can a single-phase cold plate dissipate while simultaneously satisfying thermal, hydraulic, mechanical, and infrastructure constraints?

This analysis evaluates that question using a common 5-LPM water-cooling baseline.

B.2 Governing Thermo-Hydraulic Relationships

The performance of a single-phase microchannel cold plate is governed by three coupled relationships.

Convective heat transfer

The heat transferred from the solid cold-plate structure to the coolant can be approximated as

$$Q = hA_w\Delta T_{LM}$$

where:

- Q is heat transfer rate;
- h is the convective heat-transfer coefficient;
- A_w is total wetted surface area;
- ΔT_{LM} is the appropriate mean temperature difference between the wall and coolant.

For a simplified lumped analysis:

$$Q \approx hA_w(T_{surface} - T_{fluid,avg}).$$

This relationship illustrates two fundamental methods of increasing heat-removal capability:

1. increase h ;
2. increase A_w .

Extreme microchannel refinement primarily attempts to increase h , whereas high-aspect-ratio skived-fin structures can increase A_w substantially without requiring equivalent increases in fluid velocity.

Hydraulic pressure drop

For internal flow, the pressure drop can be approximated by

$$\Delta P = f \frac{L}{D_h} \frac{\rho u^2}{2}$$

where:

- f is the Darcy friction factor;
- L is channel length;
- D_h is hydraulic diameter;
- ρ is coolant density;
- u is mean channel velocity.

For a rectangular channel,

$$D_h = \frac{2wH}{w + H}.$$

The equation illustrates the central hydraulic problem associated with channel refinement.

Reducing channel dimensions increases the number of channels required to pass a fixed volumetric flow rate, but the resulting velocity and hydraulic resistance can increase rapidly.

Consequently, increasing heat-transfer coefficient through geometric refinement can produce a disproportionately large increase in pressure drop.

Coolant temperature rise

For single-phase cooling, the coolant absorbs heat through sensible heating:

$$\Delta T_{fluid} = \frac{Q}{\dot{m}C_p}$$

with

$$\dot{m} = \rho \dot{V}.$$

Therefore,

$$\Delta T_{fluid} = \frac{Q}{\rho \dot{V} C_p}$$

where:

- $\dot{m}$ is mass flow rate;
- $\dot{V}$ is volumetric flow rate;
- C_p is coolant specific heat capacity.

Unlike a phase-change system, the coolant temperature necessarily increases as it travels through the cold plate. This creates a second fundamental constraint independent of the local heat-transfer coefficient.

B.3 Reference Operating Condition

Five representative cold-plate geometries are evaluated under a common operating condition.

Coolant Water at 35°C inlet temperature:

$$\rho = 994 \text{ kg/m}^3$$

$$C_p = 4178 \text{ J/(kg} \cdot \text{K)}.$$

Volumetric flow

$$\dot{V} = 5.0 \text{ L/min}$$

or

$$\dot{V} = 8.333 \times 10^{-5} \text{ m}^3/\text{s}.$$

The corresponding mass flow rate is

$$\dot{m} = 994(8.333 \times 10^{-5}) = 0.08283 \text{ kg/s}.$$

Channel length

$$L = 30 \text{ mm}.$$

Reference coolant-temperature-rise constraint

$$\Delta T_{fluid,max} = 10^\circ\text{C}.$$

This 10°C value is treated as a **design constraint for the reference architecture**, not as a universal physical limit on all liquid-cooling systems.

B.4 Comparative Cold-Plate Geometries

The five geometries span conventional channels through progressively finer microchannels and a high-aspect-ratio skived-fin architecture.

Parameter	Conventional	Fine-Channel	Microchannel	Skived Fin	Very Fine
Channel width w	1.00 mm	0.50 mm	0.20 mm	0.15 mm	0.10 mm
Channel height H	3.00 mm	1.50 mm	0.60 mm	1.50 mm	0.30 mm
Aspect ratio H/w	3:1	3:1	3:1	10:1	3:1
Channel count	14	27	67	89	134
Velocity	1.98 m/s	4.11 m/s	10.36 m/s	4.16 m/s	20.73 m/s
Reynolds number	4,107	4,259	4,291	1,567	4,291
h	11,832 W/m ² K	24,613 W/m ² K	62,006 W/m ² K	26,371 W/m ² K	123,991 W/m ² K
Wetted area	3,360 mm ²	3,240 mm ²	3,216 mm ²	8,712 mm ²	3,216 mm ²
Heat dissipated	398 W	797 W	1,994 W	2,298 W	3,988 W
Pressure drop	0.016 bar	0.134 bar	2.12 bar	0.386 bar	16.96 bar
Coolant rise	1.15°C	2.30°C	5.76°C	6.64°C	11.52°C
Ideal hydraulic power	0.13 W	1.11 W	17.7 W	3.22 W	141.4 W

Table 13 Comparative single-phase cold-plate geometries under the reference operating condition.

These velocities are computed consequences of channel refinement at the reference flow, not operating practice: field guidance holds loop velocities well below these values, and the erosion and corrosion behavior that motivates those limits is an active work area in OCP and ASHRAE. That practical designs cannot approach the computed velocities is itself part of why continued refinement fails.

The results reveal two distinctly different strategies for increasing thermal performance.

- The microchannel and Very Fine designs primarily increase h , whereas the skived-fin design substantially increases wetted area.
- This distinction becomes important as package power approaches several kilowatts.

B.5 The First Constraint: Hydraulic Pressure

A common way of evaluating liquid cooling is to consider pump power:

$$P_{\text{pump}} = \Delta P \dot{V}.$$

For the Very Fine geometry:

$$P_{\text{pump}} = 1.696 \times 10^6 (8.333 \times 10^{-5}) \approx 141.4 \text{ W}.$$

Relative to approximately 4 kW of heat dissipation, this corresponds to only several percent of the thermal power. Thus, in this example, pump energy is not necessarily the first limiting factor. The more consequential constraint is **pressure itself**.

The Very Fine geometry requires approximately

$$\Delta P = 16.96 \text{ bar}$$

to pass 5 LPM through the modeled cold plate.

This is dramatically higher than the pressure-drop budget typically allocated to an individual cold plate in many practical cooling architectures. However, the allowable pressure drop is system-specific and must be established from the actual CDU, manifold, quick disconnect couplings, tubing, cold-plate, and mechanical design. Consequently, the appropriate statement is not that 16.96 bar violates a universal datacenter pressure limit. Rather:

A pressure drop approaching 17 bar represents a major architectural constraint and is likely incompatible with many conventional liquid-cooling loops unless the entire hydraulic system is specifically designed for it.

This distinction is important because component pressure ratings are not universal.

B.6 Mechanical Consequences of High Internal Pressure

Pressure also creates a mechanical load on the cold plate. For a simplified 30 mm × 30 mm cavity:

$$A_{cavity} = 0.03 \times 0.03 = 9 \times 10^{-4} \text{ m}^2.$$

$$At$$

$$P = 1.696 \times 10^6 \text{ Pa},$$

the resulting pressure force is

$$F_{pressure} = PA$$

$$= 1.696 \times 10^6 (9 \times 10^{-4})$$

$$\approx 1526 \text{ N}.$$

Thus, approximately 1.5 kN of resultant pressure force acts over the nominal cavity area.

This does **not**, by itself, establish a particular cold-plate deflection.

Actual displacement depends on:

- copper thickness;
- cover thickness;
- support geometry;
- bolt pattern;
- cavity dimensions;
- material properties;
- pressure distribution;
- package attachment conditions.

A mechanical or finite-element model is therefore required to determine the resulting deflection.

Nevertheless, the magnitude of the load demonstrates why extreme pressure-drop designs cannot be evaluated solely from fluid mechanics. The cold plate becomes a pressure vessel as well as a heat exchanger.

For thin TIM bondlines, even tens of micrometers of relative deflection can produce a substantial increase in thermal resistance.

The approximate conduction relationship is

$$R_{TIM} = \frac{t_{TIM}}{k_{TIM}A}.$$

Consequently,

$$\frac{R_{new}}{R_{old}} = \frac{t_{old} + \Delta t}{t_{old}}.$$

For example, if the original effective bondline is 15 μm , an additional 15 μm of thickness would approximately double the thickness-dependent component of TIM resistance. Thus, hydraulic pressure can indirectly become a package-level thermal constraint through mechanical deformation.

B.7 The Second Constraint: Coolant Sensible Heat Capacity

The second fundamental constraint arises directly from energy conservation.

At 5 LPM,

$$\dot{m} = 0.08283 \text{ kg/s}.$$

If the maximum allowable coolant temperature rise is 10°C,

$$Q_{max} = \dot{m}C_p\Delta T_{max}.$$

Therefore,

$$Q_{max} = 0.08283(4178)(10)$$

$$Q_{max} \approx 3461 \text{ W}$$

or approximately

$$3.46 \text{ kW}.$$

This is a genuine thermodynamic constraint for the specified flow rate and coolant. It is important, however, to distinguish this result from a universal single-phase cooling limit. The result means:

At 5 LPM of water, a 10°C coolant-temperature-rise constraint limits sensible heat transport to approximately 3.46 kW.

Increasing the flow rate changes that limit.

For example, at 7 LPM:

$$\dot{m} \approx 0.116 \text{ kg/s}$$

and

$$Q_{max} \approx 0.116(4178)(10) \approx 4.85 \text{ kW}.$$

Therefore, single-phase cooling itself has not reached a fundamental thermodynamic limit at 3.46 kW. Instead, the system has encountered a **flow-dependent calorimetric constraint**.

B.8 The Flow-Rate Paradox

Increasing flow rate appears to provide an obvious solution.

Because

$$\Delta T_{fluid} = \frac{Q}{\dot{m}C_p},$$

increasing $\dot{V}$ decreases coolant temperature rise approximately in proportion to flow rate.

However, the hydraulic penalty can be severe. For a fixed geometry, pressure drop generally increases strongly with flow velocity and can approximately follow:

$$\Delta P \propto u^2$$

in regimes where the friction factor changes relatively slowly.

Thus, increasing flow can simultaneously produce:

$$\Delta T_{fluid} \downarrow$$

while

$$\Delta P \uparrow \uparrow.$$

This produces a fundamental engineering tradeoff. For the Very Fine geometry, increasing the flow from 5 LPM to approximately 7 LPM can reduce the coolant temperature rise into the acceptable range, but the corresponding hydraulic penalty becomes substantially larger.

The resulting design problem is therefore not "Can more water solve the thermal problem?" It is: "Can enough water be delivered without exceeding the hydraulic and mechanical constraints of the complete cooling loop?"

This distinction becomes increasingly important at multi-kilowatt package power.

B.9 Why Extreme Microchannel Refinement Eventually Becomes Inefficient

The five geometries illustrate the diminishing returns of pursuing increasingly high h . The Fine-Channel geometry approximately doubles the heat-transfer coefficient relative to the Conventional geometry.

The Microchannel geometry increases h to approximately

$$62,006 \text{ W/m}^2\text{K}.$$

The Very Fine geometry reaches approximately

$$123,991 \text{ W/m}^2\text{K}.$$

However, the corresponding pressure drop increases from:

$$0.016 \text{ bar}$$

for the Conventional design to

$$16.96 \text{ bar}$$

for the Very Fine design.

The thermal benefit therefore comes with an extremely nonlinear hydraulic penalty. This illustrates an important principle:

Maximizing h is not equivalent to maximizing cold-plate performance.

The relevant objective is the simultaneous optimization of:

h , A_w , ΔP , ΔT_{fluid} , and mechanical integrity.

B.10 The High-Aspect-Ratio Skived-Fin Alternative

The skived-fin geometry demonstrates a fundamentally different approach. Instead of aggressively shrinking both channel dimensions, the reference design uses:

$$H = 1.5 \text{ mm}$$

and

$$w = 0.15 \text{ mm},$$

producing an aspect ratio of approximately

$$10:1.$$

Its wetted area is:

$$A_w = 8712 \text{ mm}^2$$

This is approximately 2.6 times the conventional design in the supplied model. The resulting heat-transfer coefficient is approximately:

$$h = \frac{26,371 \text{ W}}{\text{m}^2\text{K}}.$$

This is substantially below the Very Fine value of approximately 124,000 W/m²K. Nevertheless, the skived-fin design dissipates:

$$2298 \text{ W}$$

with only:

$$\Delta P = 0.386 \text{ bar}$$

and:

$$\Delta T_{fluid} = 6.64^\circ\text{C}$$

The skived-fin architecture does not maximize the heat-transfer coefficient. Instead, it achieves a useful balance between heat-transfer coefficient and wetted area while avoiding the extreme velocities associated with the smallest channels. In other words:

Increasing wetted area can provide a more favorable thermal-hydraulic tradeoff than maximizing convective coefficient alone.

B.11 The Axial Temperature-Gradient Problem

The coolant-temperature-rise calculation also highlights an important limitation of the lumped model. The coolant does not experience a uniform temperature throughout the cold plate. Instead:

$$T_f = T_f(x).$$

An axial energy balance can be written approximately as:

$$\frac{dT_f}{dx} = \frac{q'(x)}{\dot{m}C_p}$$

where $q'(x)$ is the heat deposited into the coolant per unit channel length.

The local thermal driving force is therefore:

$$\Delta T(x) = T_s(x) - T_f(x).$$

As coolant progresses downstream, $T_f(x)$ increases.

Consequently, downstream regions of the package have less temperature head available for heat transfer unless the wall temperature also rises. For large AI packages with spatially nonuniform power density, this effect can become significant. A full thermal design should therefore use an axial or three-dimensional thermal model rather than relying exclusively on a single average temperature difference. The present calculations should consequently be interpreted as a system-level benchmark rather than a complete package thermal simulation.

B.12 Transition-Regime Considerations

The calculated Reynolds numbers for several of the 3:1 geometries are approximately 4,100–4,300. These values lie in a regime where simplistic laminar-versus-turbulent classification can become questionable for rectangular microchannels. The actual flow behavior depends on:

- aspect ratio;
- hydraulic diameter;
- entrance length;
- surface roughness;
- developing-flow effects;
- channel geometry;
- manifold distribution;
- and the heat-transfer/friction correlation being used.

The skived-fin design, at approximately

$$Re = 1567,$$

remains in a substantially lower Reynolds-number regime.

The transition-region designs should therefore be treated with geometry-specific correlations or experimental validation rather than assuming a universal transition Reynolds number. This is particularly important because both h and f directly influence the conclusions of the analysis.

B.13 The Multi-Constraint Design Envelope

The results suggest that there is no single universal "single-phase limit." Instead, the maximum practical package power is determined by the intersection of several constraints:

$$\Delta T_{fluid} \leq \Delta T_{max}$$

$$\Delta P \leq \Delta P_{max}$$

$$T_j \leq T_{j,max}$$

and

$$\delta_{mechanical} \leq \delta_{max}.$$

The maximum feasible heat load can therefore be represented conceptually as:

$$Q_{max} = \max_{geometry, \forall} Q$$

subject to the complete set of thermal, hydraulic, and mechanical constraints.

This is better understood as a **feasible design envelope** than as a single hard power threshold.

B.14 Interpreting the Approximately 4-kW Boundary

Under the reference conditions used in this analysis, three important values converge in the same region.

First, the 5-LPM calorimetric limit at a 10°C coolant rise is:

$$Q_{calorimetric} \approx 3.46 \text{ kW}.$$

Second, the Very Fine geometry reaches approximately:

$$Q \approx 3.99 \text{ kW},$$

but requires:

$$\Delta P \approx 16.96 \text{ bar}$$

and produces:

$$\Delta T_{fluid} \approx 11.52^\circ\text{C}.$$

Third, increasing flow to recover the coolant-temperature-rise margin substantially increases hydraulic pressure. These results indicate that approximately 3–4 kW is an important **design-transition region** for the specified architecture. However, it should not be interpreted as a universal physical limit.

A different system could exceed 4 kW through:

- higher flow rate;
- larger cold-plate area;
- multiple parallel flow paths;
- shorter hydraulic paths;
- improved manifold distribution;
- alternative channel architectures;
- different coolant conditions;
- higher allowable coolant temperature rise;
- multiple cold plates;
- or two-sided cooling.

The appropriate conclusion is therefore:

Approximately 4 kW represents a practical architectural boundary for the reference 5-LPM single-phase cold-plate configuration, rather than a fundamental thermodynamic ceiling on single-phase liquid cooling.

B.15 Conclusions

This analysis demonstrates that the limiting behavior of high-power single-phase cold plates is governed by the interaction of three primary effects.

1. First, increasing convective heat-transfer coefficient through aggressive microchannel refinement requires increasingly high fluid velocities and produces rapidly increasing pressure drop.
2. Second, single-phase water cooling is fundamentally limited by sensible heat capacity at a specified flow rate. At 5 LPM and a 10°C maximum coolant rise, the available sensible heat capacity corresponds to approximately:

3.46 kW

3. Third, extreme hydraulic pressure can become a mechanical and system-level constraint even when the corresponding pump power remains a relatively small fraction of package power.

The comparison of the five geometries demonstrates the resulting tradeoff clearly. The Very Fine design reaches approximately 4 kW of nominal heat-removal capability but requires approximately 16.96 bar pressure drop and exceeds the 10°C coolant-rise constraint. In contrast, the skived-fin design reaches approximately 2.30 kW while maintaining a substantially lower pressure drop of approximately 0.386 bar and a 6.64°C coolant rise.

The principal conclusion is therefore not that single-phase cooling has an intrinsic 4-kW limit. Rather:

"Single-phase cooling approaches a constrained design boundary at several kW/package."

For the reference architecture considered here, approximately 3–4 kW represents a region where conventional channel refinement increasingly fails to satisfy thermal and hydraulic requirements simultaneously.

Beyond this region, continued scaling requires architectural changes rather than simply smaller channels or higher flow rates.

Appendix C Reader's reference

C.1 Glossary

Terms are defined as this paper uses them; where the paper coins a term, the definition here is canonical for the paper. The Section column links each term to the passage that defines or principally treats it. For harmonized industry terminology, see the [OCP Cooling Environments Harmonized Glossary](#).

Term	Definition	Section
Acquisition	The first segment of the thermal path: capturing heat at its source, inside the IT envelope.	7.1
Agency mismatch	The party that feels a limit is not the party that can act on it, so transitions lag their business case.	5.4
Annualized failure rate (AFR)	Expected failures per component per year; the unit of the paper's reliability models.	8.3
Avoidance zone	The span between a method's economic and physics boundaries, across which an incumbent is stretched rather than replaced.	4.4
Cold plate (single-phase)	A channeled metal block clamped to a hot component, carrying heat away in a pumped liquid by sensible heating.	8.2
Complexity debt	The compounding cost of stretching a solution across a system it was not designed to cover; cold plate's accreting plumbing is the worked example.	5.6
Concentrated heat	The high-flux load of the hottest devices; the target of the surface methods.	6.1
Containment	The airflow-separation infrastructure of the air path; in the cost analysis, air's transport-segment element.	9.2
Coolant-distribution unit (CDU)	The transport-segment unit separating the facility loop from the IT loop, pumping coolant and exchanging heat across the seam.	8.2
Dielectric fluid	An electrically non-conductive coolant that can contact live electronics.	8.5
Distributed heat	The non-processor load spread across memory, network, storage and power conversion; the target of the volume methods.	6.2
Economic boundary	The point where a method's full cost, capital, operations and reliability together, turns against it even though it still works.	4.2
Efficiency boundary	The point where the power spent moving heat becomes a visible share of the power spent computing.	4.1
Facility envelope	The building-side domain of the thermal path, refreshed on the slow facility clock.	7.2
Flow boiling	Boiling in an imposed flow through a channel; the mechanism of the two-phase cold plate.	8.4
Heat flux	Heat per unit area, the quantity that binds cooling at the die; per-reticle power serves as its frontier proxy.	3.4
Immersion (single-phase)	A volume method submerging the whole system in a dielectric bath, cooling every component at once by sensible heating.	8.5
Inflection	The moment a forced transition arrives, when the incumbent can no longer hold the load; inflections are forced, not chosen.	5.6
IT envelope	The server-and-rack domain of the thermal path, refreshed on the fast IT clock.	7.2
Least-disruption law	An ecosystem as a whole exhausts every option that preserves its current architecture before it accepts one that changes it.	5
Lifecycle mismatch (two clocks)	IT hardware turns over yearly while facility infrastructure lasts decades, so the two sides of the seam evolve at different speeds.	5.5
Manifold	The distribution piping carrying coolant between the CDU and the trays within a rack.	8.2
Mean time between failures (MTBF)	Average operating time between failures; the paper states reliability as multiples of the air-cooled baseline.	8.3

Microfluidics	In-chip cooling through microscopic channels etched into or beneath the silicon, removing the resistance stack between junction and coolant.	8.8
PFAS	Per- and polyfluoroalkyl substances, the fluorocarbon chemistry family of the best-performing two-phase fluids, whose regulatory status is unresolved.	0
Physics boundary	The hard limit past which a method cannot remove the heat at all; it cannot be negotiated, only approached.	4.3
Pool boiling	Boiling in an open bath with no imposed flow; the mechanism of two-phase immersion.	0
Precision cooling	Directed and enhanced fluid flow at the hottest devices inside an immersion tank.	8.5
Preparation debt	The unbuilt supply chains, standards and skills that come due all at once when a deferred transition is forced; owed by an ecosystem, not by any one actor.	5.7
Quick disconnect (QD)	A dripless coupling pair that lets a liquid line be separated for service; counted per mated pair in the reliability model.	8.3
Rack unit (RU)	The standard vertical slot of a rack; the paper measures distributed heat per rack unit.	6.2
Rear-door heat exchanger	A rejection-side coil on the rack's exhaust, buying rejection capacity without moving the acquisition limit.	8.1
Rejection	The final segment of the thermal path: discharging the heat from the building.	7.1
Seam	The boundary where the IT and facility envelopes meet; no single party owns it, and liquid cooling earns its complexity there.	7.2
Sensible and latent heat	Heat carried by warming a fluid versus by changing its phase; the distinction that separates single-phase from two-phase methods.	8.4
Strings	The dependencies a technology forces on parts of the system its adopter does not control.	5.1
Temperature budget	The allowable temperature drop between the junction and the cooling medium, which shrinks as flux climbs.	3.3
Thermal path	The route heat travels from die to environment: acquisition, transport and rejection.	7
Thermal shadowing	Air pre-heated by one component reaching the next with less of its temperature budget left.	8.1
Toolbox	The paper's destination: several cooling methods deployed together, matched to the heat, the site and the moment.	10
Transport	The middle segment of the thermal path, moving heat from the IT side to the facility side across the seam.	7.1
Two-phase cold plate	A cold plate in which the coolant boils in the channel, trading a temperature rise for latent heat.	8.4
Two-phase immersion	A volume method whose bath boils on the hottest surfaces and condenses on a coil above.	0
xPU	A processor socket of any class, CPU or accelerator.	6.2

C.2 References

- [1] Microsoft, "Microsoft finds underwater datacenters are reliable, practical and use energy sustainably," September 2020. [Online]. Available: <https://news.microsoft.com/source/features/sustainability/project-natick-underwater-datacenter/>.
- [2] A. Hipes, "Immersion Server Demonstrated Reliability at Datacenter Scale Deployment," in *OCP EMEA Summit*, 2025.
- [3] M. Jalili, "Cost-Efficient Overclocking in Immersion-Cooled Datacenters," in *ISCA 2021, 48th International Symposium on Computer Architecture*, 2021.
- [4] Microsoft, "To cool datacenter servers, Microsoft turns to boiling liquid," April 2021. [Online]. Available: <https://news.microsoft.com/source/features/innovation/datacenter-liquid-cooling/>.
- [5] 3M, "3M to Exit PFAS Manufacturing by the End of 2025," 20 December 2022. [Online]. Available: <https://news.3m.com/2022-12-20-3M-to-Exit-PFAS-Manufacturing-by-the-End-of-2025>.
- [6] R. C. Chu, R. E. Simons, M. J. Ellsworth, R. R. Schmidt and V. Cozzolino, "Review of Cooling Technologies for Computer Products," *IEEE Transactions on Device and Materials Reliability*, 4(4), 2004.
- [7] R. C. Chu, *Review of Cooling Technologies for Computer Products: Past, Present and Future (CMRR Seminar, UC San Diego)*, 2005.
- [8] R. E. Simons, "The evolution of IBM high performance cooling technology," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A*, 18(4), 1995.
- [9] R. R. Schmidt, "Liquid Cooling is Back," *Electronics Cooling*, August 2005.
- [10] A. J. Blodgett and D. R. Barbour, "Thermal Conduction Module: A High-Performance Multilayer Ceramic Package," *IBM Journal of Research and Development*, 26(1), 1982.
- [11] S. Oktay and H. C. Kammerer, "A Conduction-Cooled Module for High-Performance LSI Devices," *IBM Journal of Research and Development*, 26(1), 1982.
- [12] D. J. Delia, "System Cooling Design for the Water-Cooled IBM Enterprise System/9000 Processors," *IBM Journal of Research and Development*, 36(4), 1992.
- [13] T. Watari and H. Murano, "Packaging Technology for the NEC SX Supercomputer," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, 8(4), 1985.
- [14] Computer History Museum, "ES9000 Thermal Conduction Module, catalog 102741297," 1990. [Online]. Available: <https://www.computerhistory.org/collections/catalog/102741297>.
- [15] Open Compute Project, "OCP TCO tools, contributed cost-model iterations (Cooling Environments)," 2026. [Online]. Available: <https://drive.google.com/open?id=1SzT0w9IJsumeVf5i9gpSHuqe4G3y3TjQ>.

C.3 Table of Figures

Figure 1 Module-level heat flux of high-end processor modules, bipolar and CMOS eras, W/cm ² . Sources in A.1.	8
Figure 2 Leading chip package power and die-level heat flux, CPUs, GPUs and AI accelerators, 2008–2026. Sources in A.2.	9
Figure 3 Die-level heat flux, 2017 to 2026; haloed markers denote the AI accelerator class. Sources in A.2.	16
Figure 4 The three boundaries and the avoidance zone.	19
Figure 5 Distributed non-xPU heat per rack unit for the representative AI platform of each era, at published and announced rack powers; modeled composition. Sources in A.3.	28
Figure 6 The thermal path segmented into acquisition, transport and rejection, aligned to the IT and facility envelopes and their lifecycles.	29
Figure 7 Industry roles along the thermal path, before and during the liquid transition; roles indicative, not exhaustive.	30
Figure 8 Fan power to hold junction temperature, as a share of the combined device and fan power, for a fixed reference thermal solution; the knee marks the efficiency boundary. Model in Appendix A.	33
Figure 9 Cold-plate adoption for top-tier accelerators, with the distributed non-processor heat trend left behind.	34
Figure 10 The three boundaries for single-phase cold plate, by power per reticle-class device (3.3); physics wall per the thermo-hydraulic analysis in Appendix B, efficiency and economic positions directional.	37
Figure 11 Illustrative single-phase and two-phase cold plate boundary positions, by power per reticle-class device; physics walls derived in Appendix B, two-phase position directional.	43
Figure 12 Relative MTBF by cooling method.	48
Figure 13 The ecosystem boundary immersion crosses, spanning the IT and facility arenas.	50
Figure 14 IT-side capex delta versus the air baseline, \$M; air IT capex 5.3B.	56
Figure 15 Capex in the transport segment, the seam, \$M; bar scales differ per segment figure.	56
Figure 16 Capex in the rejection and facility segment, \$M.	57
Figure 17 Annual IT failure cost by cooling method, \$M/yr; whiskers carry the ranges of the reliability chapters.	58
Figure 18 Annual facility opex by element, \$M/yr: energy, real estate, maintenance.	59
Figure 19 Total annualized cost per year by cooling method, \$M/yr; whiskers as in the failure-cost figure.	60

C.4 List of Tables

Table 1 Proposed successor computing paradigms compared on workload scope, programming model, maturity and cooling burden.	10
Table 2 Heat capacity of common cooling fluids.	14
Table 3 Baseline annualized failure rate assumptions, air- and liquid-cooled components.	39
Table 4 Failure mechanisms by cooling environment, air, single-phase cold plate and single-phase immersion.	46
Table 5 Module-level heat flux source data, bipolar and CMOS series, with per-machine basis flags and public anchors.	72
Table 6 Leading chip package power and die-level heat flux, CPUs, GPUs and AI accelerators, 2008–2026, with per-device basis flags; flux computed as TDP over total die area.	74
Table 7 Distributed non-xPU heat per rack unit by class and era, W/RU, with datasheet-residual checks.	76
Table 8 Class anchors and public sources for the distributed-heat composition.	76
Table 9 Cold plate adoption: share of top-tier accelerator TDP on cold plate, by year.	78
Table 10 Leading AI rack power and the heat left behind for air, by year, with basis.	78
Table 11 Declared inputs of the cost comparison, with basis flags.	81
Table 12 Harmonized engine variables and multipliers of the contributed base models, applied identically across technologies.	83
Table 13 Comparative single-phase cold-plate geometries under the reference operating condition.	88

Appendix D License

OCP encourages participants to share their proposals, specifications and designs with the community. This is to promote openness and encourage continuous and open feedback. It is important to remember that by providing feedback for any such documents, whether in written or verbal form, that the contributor or the contributor's organization grants OCP and its members irrevocable right to use this feedback for any purpose without any further obligation.

It is acknowledged that any such documentation and any ancillary materials that are provided to OCP in connection with this document, including without limitation any white papers, articles, photographs, studies, diagrams, contact information (together, "Materials") are made available under the Creative Commons Attribution-ShareAlike 4.0 International License found here:

<https://creativecommons.org/licenses/by-sa/4.0/>, or any later version, and without limiting the foregoing, OCP may make the Materials available under such terms.

As a contributor to this document, all members represent that they have the authority to grant the rights and licenses herein. They further represent and warrant that the Materials do not and will not violate the copyrights or misappropriate the trade secret rights of any third party, including without limitation rights in intellectual property. The contributor(s) also represent that, to the extent the Materials include materials protected by copyright or trade secret rights that are owned or created by any third-party, they have obtained permission for its use consistent with the foregoing. They will provide OCP evidence of such permission upon OCP's request. This document and any "Materials" are published on the respective project's wiki page and are open to the public in accordance with OCP's Bylaws and IP Policy. This can be found at <http://www.opencompute.org/participate/legal-documents/>. If you have any questions, please contact OCP.

About Open Compute Foundation

The Open Compute Project (OCP) brings at-scale innovations and hyperscaler best practices to all, spanning technology domains from the data center to the edge, and the technology stack from silicon, to systems, to site facilities and services. The international OCP Community is made up of organizations and people from hyperscale and tier-2 cloud data center operators, communications providers, colocation providers, diverse enterprises, and technology vendors. The OCP Foundation fosters collaboration within the Community to tackle market challenges in an array of OCP Projects that create open contributions such as specifications, designs, and more. The OCP Solution Provider Program then recognizes the application of those contributions in compliant products, solutions, and data center facilities and services with designations like OCP Inspired, OCP Accepted and OCP Ready. With the tenets of openness, impact, efficiency, scale and sustainability, the OCP engages and educates thousands of engineers every year through many events and webinars. Across many initiatives the OCP Foundation and Community are meeting the market today and shaping the future.

Learn more at: www.opencompute.org.