How Two-Phase Liquid Cooling Future-Proofs Your Facility

Danny Bradbury, Independent IT Journalist

ACCELSIUS™
contact@accelsius.com
www.accelsius.com

INTRODUCTION

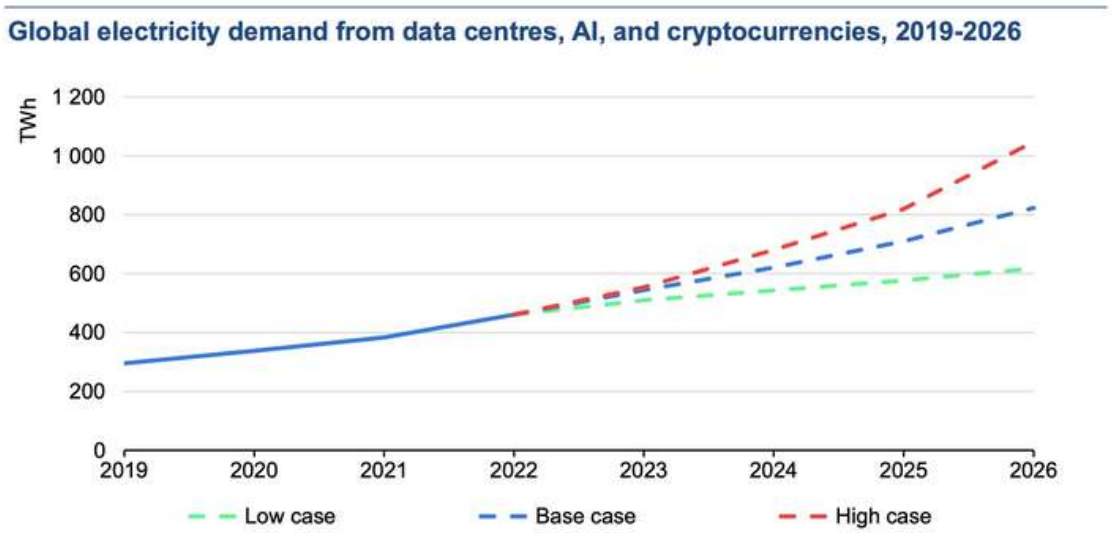
When Benjamin Franklin flew his famous kite in 1752 and discovered lightning's power, he couldn't have known what the world had in store for electrical energy. Today, the data center industry is one of the biggest users of electricity. The pressure is on for facility builders and operators as they face rising power demands. Delays in power procurement and long greenfield build times compound the problem.

With cooling representing around a third of the average data center's energy consumption, the race is on to find more energy-efficient ways of shedding heat from IT hardware. Liquid cooling — and specifically two-phase direct-to-chip (2PD2C) cooling — has emerged as the most promising way to cope with these increased energy pressures.

The Problem: Data Centers Aren't Keeping Pace with Chip Development

Data center operators are already straining under increased energy pressures after witnessing rack densities double from 6 to 12kW in the last six years. If you think that's bad, stick around and watch artificial intelligence (AI) usage turn up the pressure far further.

AI was always power-hungry, but its latest iteration is far more demanding. The large language models (LLMs) like ChatGPT and Claude that drive generative AI are insatiable. They become more accurate as they consume more content for their compute-intensive training processes, which has created a race to consume increasing amounts of data. The inference phase in which customers use the trained model is also becoming increasingly power-hungry.



Source: International Energy Agency

This phenomenon has substantially accelerated demands for computing power. According to the International Energy Agency (IEA), demand for AI servers is set to grow tenfold between 2023 and 2026. However, the pressure from AI comes not just from the volume of AI-focused graphics processing units (GPUs) entering data centers but from their velocity. GPU-processing capabilities and power consumption are developing at breakneck speed. NVIDIA, which currently leads the field in AI chips, has increased the cadence of its GPU releases to one each year.

NVIDIA Release Cadence

Processor	Production (Complete Systems)	Power Consumption
A100	2000	400W
H100	2023	700W
H200	2024	700W
B100	2024-5	700W
B200	2025	1000W
Rubin	2026	Unknown
Rubin Ultra	2027	Unknown

Compiled from multiple sources

Analysts are projecting a massive increase in data center energy consumption based on this development. Goldman Sachs estimated a total 400TWh of global data center energy usage in 2023, only 12TWh of which was AI. By 2030, it anticipates 1018TWh of energy usage, of which 209TWh will serve AI hardware.

These rapidly rising demands present a headache for data center builders and operators who take between three and five years on average to build a facility. Getting enough power from the local utility to serve these data centers takes up to seven years for hyperscale builds, according to analysts at TD Cowen.

Power is becoming increasingly constrained. As governments compel utilities to become more sustainable, demand to power everything from data centers to electric vehicles is skyrocketing. This is creating some worrying predictions. For example, TD Cowen expects Northern Virginia (the largest US data center market in the U.S.) to run out of reliability-rated power in winter 2027.

For data center operators, coping with this problem is like hitting a moving bullet. They must think years ahead when designing modern facilities to accommodate these accelerating energy demands. To do that, they will increasingly turn to liquid cooling.

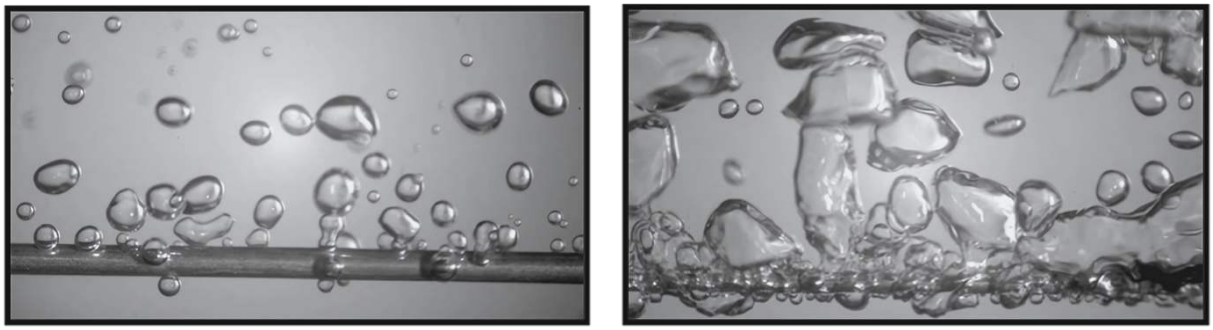
Liquid Cooling is the Answer

Liquid is simply more efficient than air at cooling hot processors, which is why NVIDIA began seriously supporting liquid cooling solutions in 2022 with its A100 GPU. In 2024, CEO Jenson Huang suggested that the company's next generation (the Blackwell) would be primarily liquid cooled.

However, data center designers and operators should beware; not all liquid cooling is the same. There are two main types. The first is immersion, where components are bathed in fluid, typically in a tank. The second is direct-to-chip (D2C) cooling, which requires far fewer server modifications than immersive tanks. D2C also requires less coolant because it funnels the liquid directly to the chip.

D2C cooling also comes in two categories of its own. The first, single-phase cooling (1P2DC), uses a coolant (typically water or a propylene glycol-water mixture) in a constant liquid state to transfer heat away from the processor.

Two-phase cooling (2PD2C), however, uses a different coolant with a low boiling point, enabling it to phase change from a liquid to a gas to increase cooling efficiency. Pumps pressurize the liquid flowing into a cooling plate mounted on the processor. The coolant then boils as it absorbs the chip's heat, passing back into a condenser that cools it back into a liquid, transferring heat either into facility water under the rack or alternatively via rear-door heat extraction.



Two-phase coolant undergoing phase change from liquid to gas.

Concerns remain regarding the polyfluoroalkyl substances (PFAS) commonly used as coolants in 2PD2C systems. These have a reputation as “forever chemicals” that don't break down in the environment. However, not all PFAS chemicals are the same. A fourth-generation PFAS coolant known as R1233zd used in the Accelsius NeuCool system converts into a naturally occurring compound called trifluoroacetic acid (TFA).

Why 2PD2C?

This targeted approach to cooling delivers several benefits:

Greater efficiency

A boiling liquid transfers more heat than one that doesn't change state. That's because a boiling liquid remains at the same temperature as it draws heat from its surroundings until it turns entirely to vapor. A constant flow of new coolant into the micro-channel-based cooling plate constantly extracts heat from the chip.

This high efficiency allows 2PD2C to cool high-wattage chips without reaching heat flux (the point where a coolant can't absorb any more heat and the chip begins to overheat). That's a critical advantage as the thermal output of GPUs rises in line with their power consumption. For example, tests with NVIDIA's H100 GPUs showed two-phase cooling kept chip temperatures below 57°C, compared to 66-71°C with single-phase water cooling.

Cost effectiveness

High resilience is a key factor in the financial benefits of 2PD2C. More resilient systems mean less unplanned downtime, enabling data center operators to generate more computational output. They also mean less planned downtime. 2PD2C's use of dielectric fluid avoids the corrosion and bacterial growth that force water-based 1PD2C operators to regularly replace and test the coolant.

Another financial advantage of 2PD2C is its ability to maintain maximum system performance. With less efficient cooling solutions, processors must 'throttle' their performance, reducing their computational output to avoid overheating. In some applications, such as financial services, throttling processors costs millions in lost productivity.

Future proofing

2PD2C also provides more strategic benefits, especially when it comes to long-term data center efficiency. It enables builders and operators to plan far into the future by increasing a facility's cooling capacity, making way for ongoing advances in silicon development.

What Happens Next?

As data center builders and operators acknowledge these pressures, expect to see liquid cooling make its way into facilities relatively quickly. Analyst firm Dell'Oro has already predicted that liquid cooling will hit a third of the data center cooling market by 2028.

In spite of its lower efficiency and more complex deployment requirements, Dell'Oro expects 1PD2C cooling to be the initial choice for many adopters. However, with data centers rapidly needing all the cooling capacity they can get, Accelsius believes that 1PD2C cooling will simply be a stepping stone. Smart companies will leap that transitory step altogether, turning instead to more efficient 2PD2C systems that provide extra cooling efficiency and system resilience.

Find out more about what the Accelsius NeuCool 2PD2C system can do for your data center by contacting us today.

ACCELSIUS™

info@accelsius.com www.accelsius.com