MEDIA RELEASE

Burning Cash: AI Factories Are Losing Over 30% of Compute Power Due to Heat

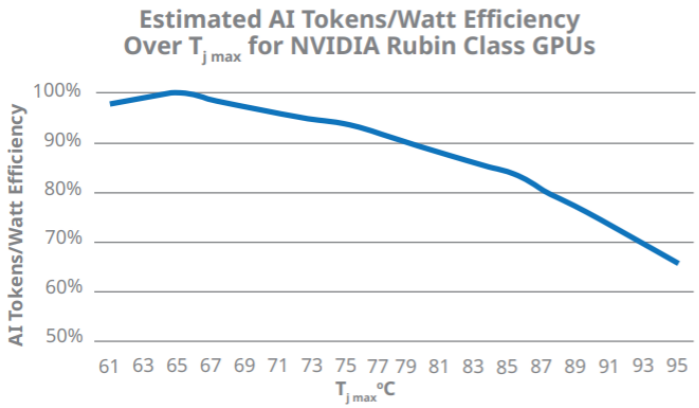
Chip to Chiller, a new white paper from Frore Systems shows the Thermal Stack is now the largest untapped lever for AI Factory economics

SAN JOSE, California — July 29, 2026 — Frore Systems, a Silicon Valley thermal solutions company, today released "Chip to Chiller: The Thermal Stack's Impact on AI Factory Efficiency," a white paper making the case that thermal management has become the largest remaining lever available to AI Factory hyperscalers seeking more compute from a power budget.

The paper's central finding: a well-engineered Thermal Stack can boost an AI Factory's Tokens/Watt — the metric that is the true economic scorecard for AI infrastructure — by over 30%. Because AI Factories are constrained by power, more than compute or memory, every percentage point improvement in Tokens/Watt converts directly into a percentage point increase in AI Factory profitability.

"GPU silicon performance is fixed the moment it's manufactured," the white paper states. "The only lever left in the hyperscaler's control is the Thermal Stack." As GPU junction temperature rises, leakage power increases exponentially, triggering performance degradation that silently erodes the Token output a hyperscaler has already paid for in power and hardware.

Tokens/Watt is determined by heat, or GPU Die Max junction temperature, which is managed by the Thermal Stack.



Among the paper's findings: Frore's own LiquidJet coldplate — a short-loop, multi-stage design built using semiconductor manufacturing techniques rather than conventional skiving — reduces GPU die temperature by 6 to 12°C compared to a conventional coldplate, translating to a 10 to 25% improvement in Tokens/Watt.

The AI industry has spent years optimizing everything upstream of the rack — AI model architecture, GPU chip design, power procurement — while treating thermal as a facilities afterthought," said Seshu Madhavapeddy, CEO and founder of Frore Systems. "This research shows that's backwards. At gigawatt scale, thermal isn't a line item. It's a competitive moat."

The paper positions thermal as a capital allocation decision that should be made before an AI Factory's budget is set, explaining that operators who treat the Thermal Stack as first-class infrastructure will extract materially more Tokens/Watt from the same GPU investment than those who don't, a gap that will widen as GPU power densities continue climbing.

Download the full White Paper here: Chip to Chiller – The Thermal Stack's Impact on AI Factory Efficiency

---

About Frore Systems: Frore Systems is a pioneer in advanced thermal technologies that unleash performance across data centers and edge devices. The company's flagship solutions include **LiquidJet®,** a multi-stage 3D short-loop jetchannel liquid cooling coldplate for data centers delivering higher GPU performance, AI token throughput, improved PUE and reduced TCO; **LiquidJet®** Nexus, a lightweight integrated coldplate system that integrates multiple LiquidJets and eliminates all hoses, connectors and manifolds enabling ½U compute trays and **AirJet®**, the world's first solid-state active air-cooling chip used in consumer, industrial, and IoT markets delivering higher performance in ultra-compact, silent, light, dustproof and water-resistant edge devices. Frore Systems' patented cooling technologies are integrated into products from major OEMs and system builders worldwide. Headquartered in Silicon Valley, with manufacturing operations in Taiwan, Frore Systems is redefining thermal architecture for the AI era. For more information, visit: www.froresystems.com

Media Contact:

Sue Ryan - VP Marketing, Frore Systems
sue@froresystems.com
Cell: +1 314 914 5008