# High-Efficiency Two-Phase DI Water DLC for Next-Gen AI Servers


Ching-chih Wei
CreAlien
Kaohsiung, Taiwan
michael.wei@crealiens.com

Shi-bin Chen
CreAlien
Kaohsiung, Taiwan
ben.chen@crealiens.com

Ming-hsueh Wu
CreAlien
Kaohsiung, Taiwan
crealien@crealiens.com



**Abstract — As AI accelerator TDPs exceed 2300W, traditional single-phase cooling hits a Pump Power Wall, where cubic pumping power growth negates efficiency gains. This paper proposes a Water-Based Two-Phase architecture that leverages latent heat (2260 kJ/kg) to decouple cooling capacity from mass flow. To overcome water's extreme 12,000x vapor expansion under vacuum (50°C saturation), we introduce a confined vapor chamber with a 270x geometric expansion. This prevents choking by maintaining vapor velocities at a safe ~8.2 m/s. Prototype testing validates stable 300 W/cm² heat flux dissipation (2.5kW per chip) at a minimal flow rate of 0.4 LPM/kW. Ultimately, this PFAS-free solution enables high-ambient dry cooling strategies, achieving PUE < 1.1 and near-zero WUE for AI servers in a sustainable hyperscale AI infrastructure.**


## I.INTRODUCTION

As AI accelerator TDPs exceed 2300W, single-phase DLC faces a Pump Power Wall: managing heat via specific heat capacity requires a linear increase in mass flow, yet pumping power scales cubically with flow rate, eventually negating IT efficiency gains. To overcome this barrier, transitioning from Sensible Heat

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

,where Q is heating load, $\dot{m}$ is mass flow rate, $C_p$ is specific heat of liquid, $\Delta T$ is differences of temperature between exchanged liquid, to Latent Heat

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

,where $h_{fg}$ is enthalpy of vaporization, $x$ is mass vapor quality.
While two-phase cooling enables high heat flux dissipation (>1000 W/cm²), using water at high heat loads (>300 W/cm²) introduces significant challenges in flow stability and critical heat flux.

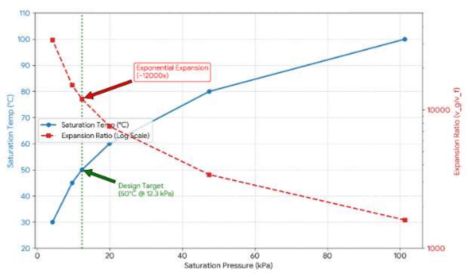


Fig. 1. The thermodynamic trade-off of water between saturation temperature, pressure and vapor expansion ratio.

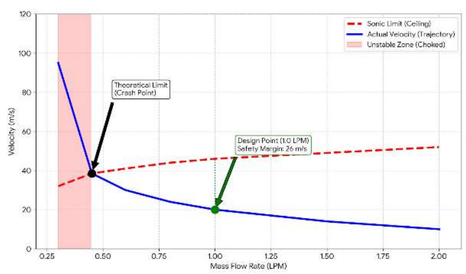


Fig. 2. Theoretical choking limit analysis defines the absolute minimum flow rate.

## II.THE THERMODYNAMIC TRADE-OFF

Water is chosen as the working fluid due to its high latent heat (2260 kJ/kg), about 20x greater than typical refrigerants. To keep the junction temperature below 80°C, the saturation temperature must remain under 50°C, requiring vacuum operation at 10–12 kPa. At this pressure, vapor specific volume increases ~12,000x causing rapid expansion. In constant-area microchannels, this leads to flow acceleration, choked flow, and severe pressure instability.

## III. DESIGN METHODOLOGY : NAVIGATING THE CHOKING LIMIT

To achieve >3 kW heat removal without choking, a physics-based approach was used to define flow limits and geometry.

**A. Theoretical Choking Limit Analysis**

- Sonic limit ↓: As vapor quality increases, mixture speed of sound drops sharply.
- Velocity ↑: Vapor expansion drives fluid velocity up nonlinearly.
- Critical point: Intersection at 0.45 LPM → below this, choking is unavoidable.
- Design choice: 1.0 LPM (equivalent to 0.4 LPM/kW for a 2.5 kW load), with velocity ~20 m/s vs. sonic limit ~46 m/s → ~26 m/s safety margin.

**B. Vapor Expansion Geometric Strategy**

- 45x expansion: At choking boundary (~50 m/s).
- 75x expansion: ~30 m/s, meets minimum safety.
- 270x expansion (final): ~8.2 m/s, well within safe zone → stable up to >5 kW.

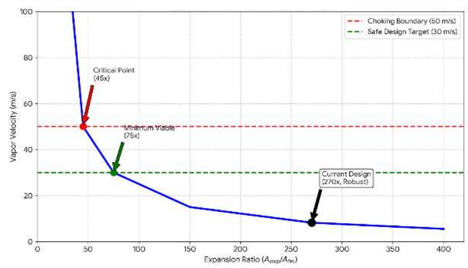


Fig. 3. Geometric selection strategy with relationship between vapor velocity and expansion ratio.

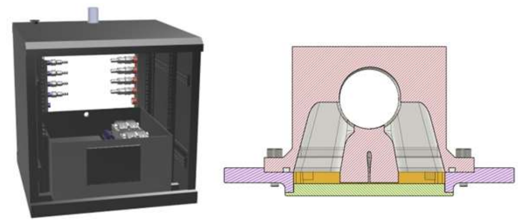

Fig. 4. A 2U server rack with two-phase cooling POC system and central-in-side-out structure coldplate .

## IV. SUSTAINABILITY & POC VALIDATION

A prototype 2U server rack utilizing our Two-Phase cooling system was tested:

1. Heat Flux: Successfully handled 300 W/cm², equivalent to a 2.5kW chip load under 35°C ambient free cooling conditions.
2. PUE: Real-world measured PUE was below 1.1, with a theoretical target of 1.03 for large-scale deployment.
3. Sustainability:
   - The system uses standard DI water, eliminating PFAS forever chemicals.
   - Operates at 50°C saturation temp. under high-ambient temp. using dry-cooler.
   - Allows operation with zero water consumption for the vast majority of the year.
4. Pump power wall: The system can effectively reduce mass flowrate to 0.4 LPM/kW to break the wall, eliminating the need for larger pumps as heat loads increase.
5. Future readiness: The architecture is designed to accommodate > 3kW per chip, scaling to the R200 era.

## V. CONCLUSION

This study resolves the pump power wall by leveraging latent heat to reduce mass flowrate to 0.4 LPM/kW. We validated that water's intrinsic over 12,000x vapor expansion is manageable via a 270x geometric expansion, achieving a stable heat flux of 300 W/cm². Furthermore, by elevating the saturation temperature to 50°C, we enable a high-ambient temp 35°C dry-cooler condensing structure with a PUE of lower than 1.1 and near-zero water consumption WUE. This demonstrates phase-change water cooling as a highly viable and sustainable solution for next-gen AI servers.


OCP FUTURE TECHNOLOGIES SYMPOSIUM
29–30 April, 2026
Barcelona, Spain
Leading the Future of AI.