

Beyond Conventional Two-Phase Cooling

A New Approach for Scalable AI Infrastructure

OCP Educational Webinar Series

May 19th 2026



IDTechEx



Community-driven hyperscale innovation for all

Beyond Conventional Two-Phase Cooling

A New Approach for Scalable AI Infrastructure

Cliff Grossner, Ph.D.
Chief Innovation Officer
OCP Foundation



Xiaoxi He, Ph.D.
Principal Research Associate
IDTechEx



Shahar Belkin
Chief Evangelist
ZutaCore



Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

Agenda

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

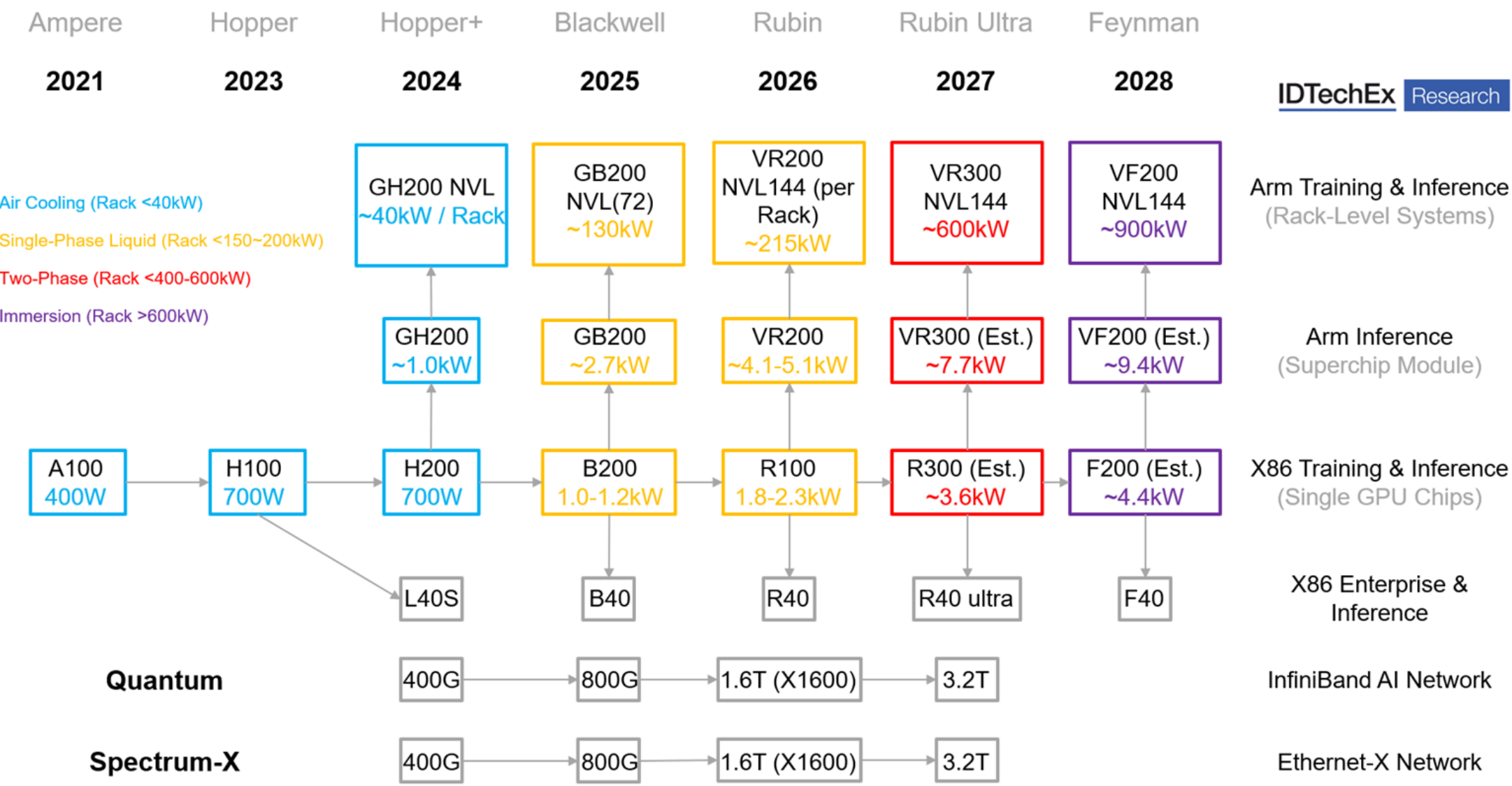
Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

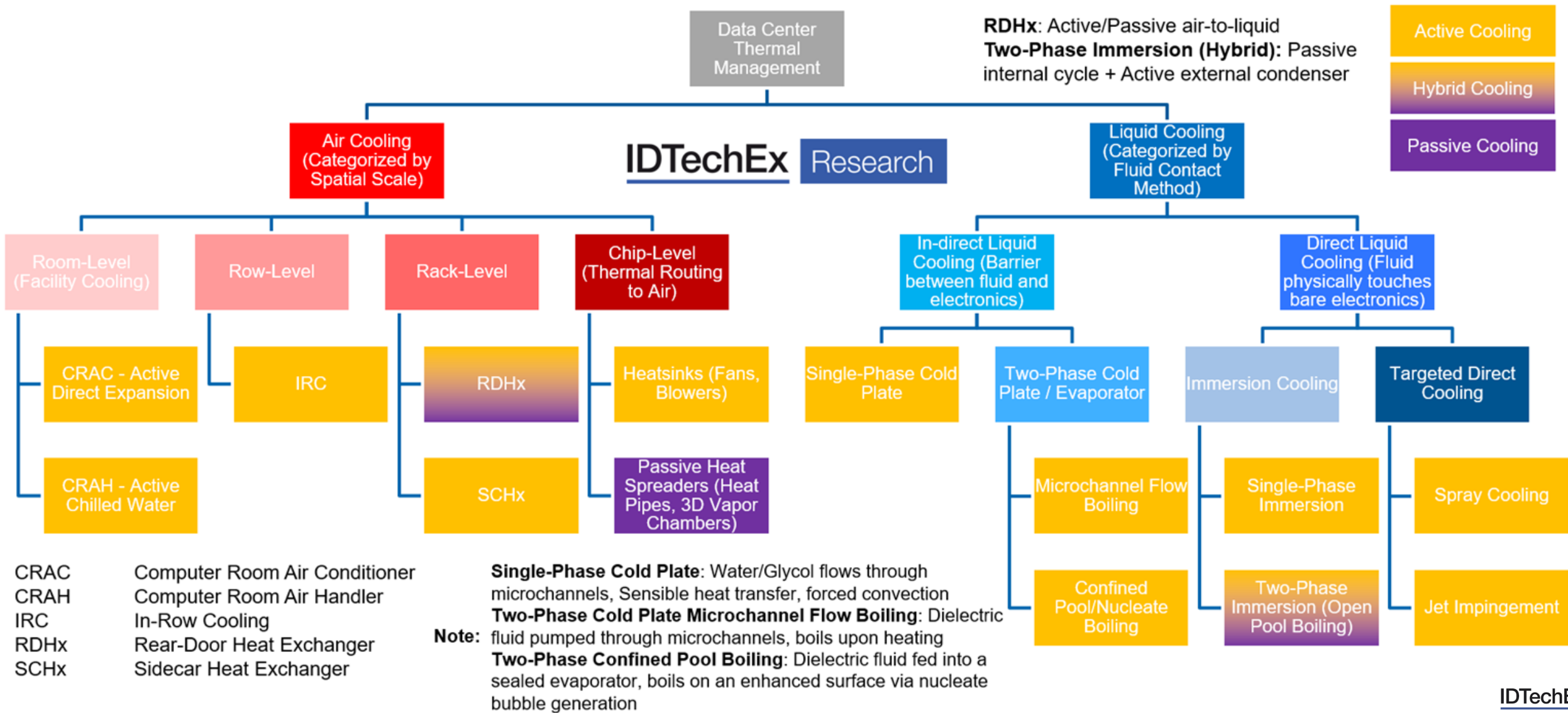
NVIDIA Architecture & Thermal Power Density Trajectory



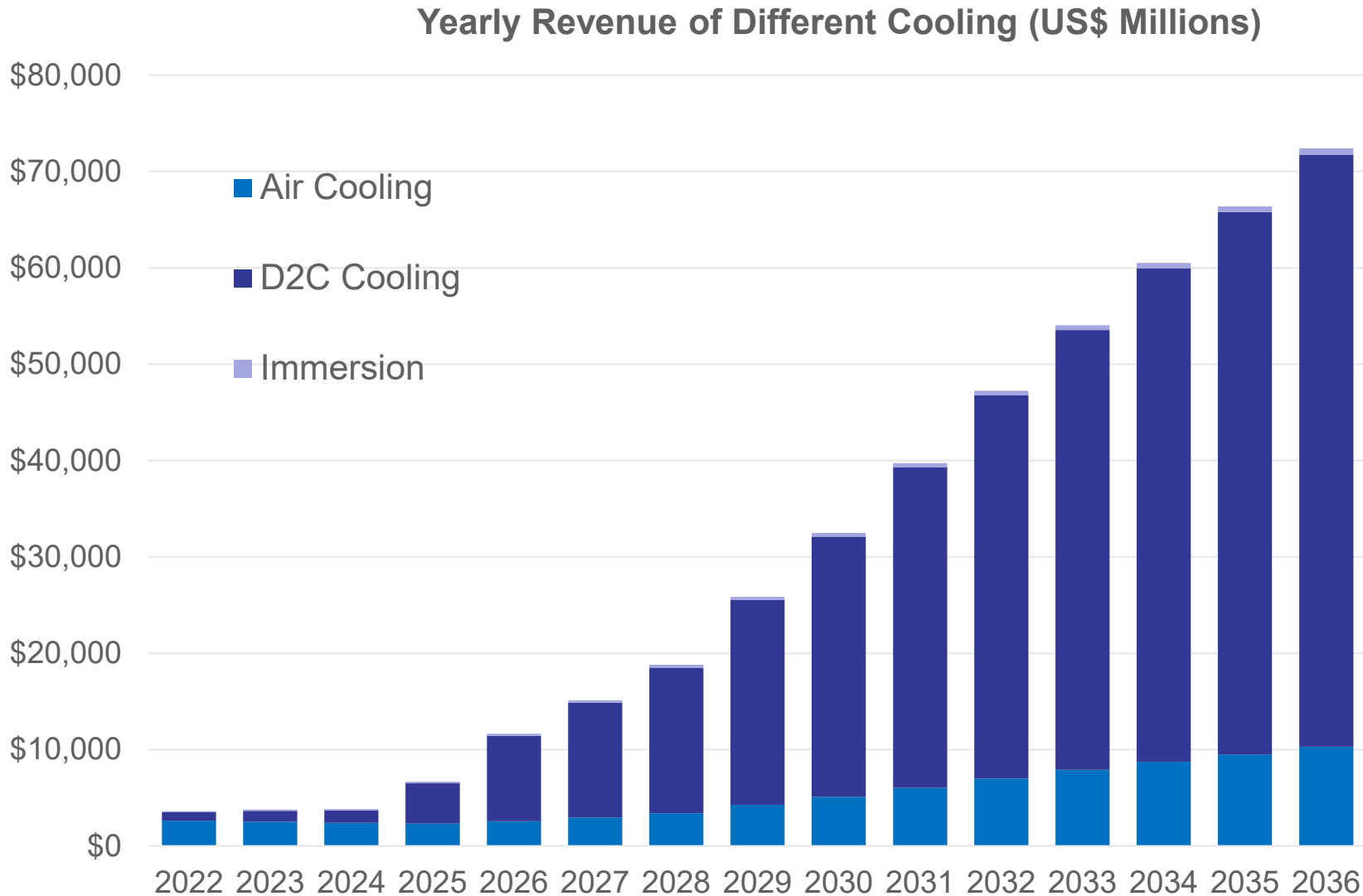
IDTechEx Research

IDTechEx

Thermal Management Technologies for Data Center



Market Forecast by Cooling Method



Immersion (both single and two-phase immersion):

- Immersion tanks
- Immersion Coolant
- CDUs for immersion
- Piping/Valves/Monitoring

Cold plate (both single- and two-phase) includes:

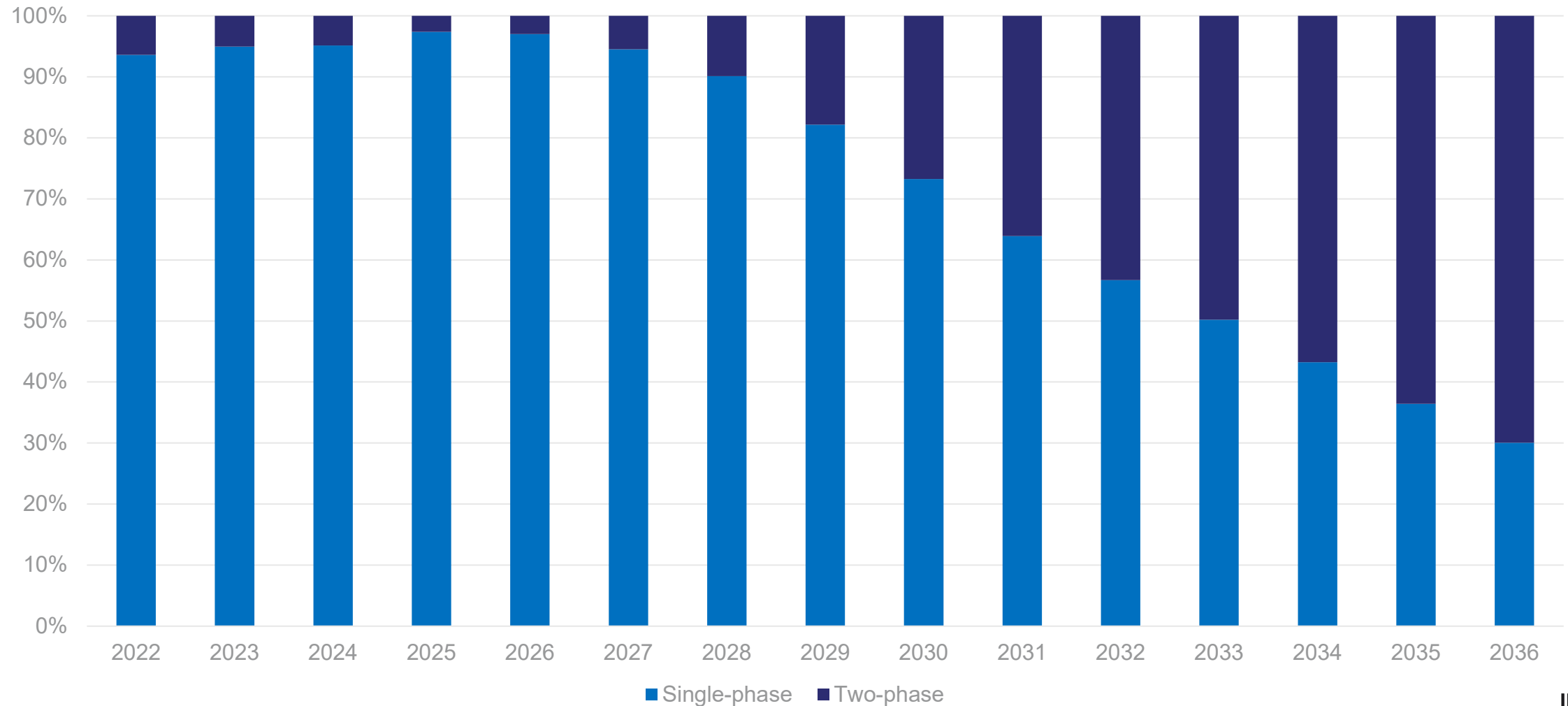
- Cold plate, tubing, hoses, piping, fluid distribution manifold in the servers)
- QDs
- CDUs for D2C
- Manifolds for D2C

Air cooling includes (air-cooling components for solely air-cooled DCs, D2C-cooled DCs and immersion-cooled DCs):

- RDHx
- Fans on servers and facility (if applicable)
- 3D VCs and heat sinks

Single vs Two Phase Markets

Percentage of Single and Two-Phase Cold Plate (Volume)



Move Towards Edge Computing

• PROCESS DATA CLOSER. RESPOND FASTER. TRANSFORM TOMORROW. •

1 EVOLUTION



2 BENEFITS



LOW LATENCY

Process data locally and respond in real time.



ENHANCED SECURITY

Keep sensitive data closer to where it's generated.



BANDWIDTH EFFICIENCY

Reduce data transfer and optimize bandwidth.

3 TRENDS

AI AT THE EDGE

Smarter decisions with real-time AI inference on edge devices.



SMART MANUFACTURING

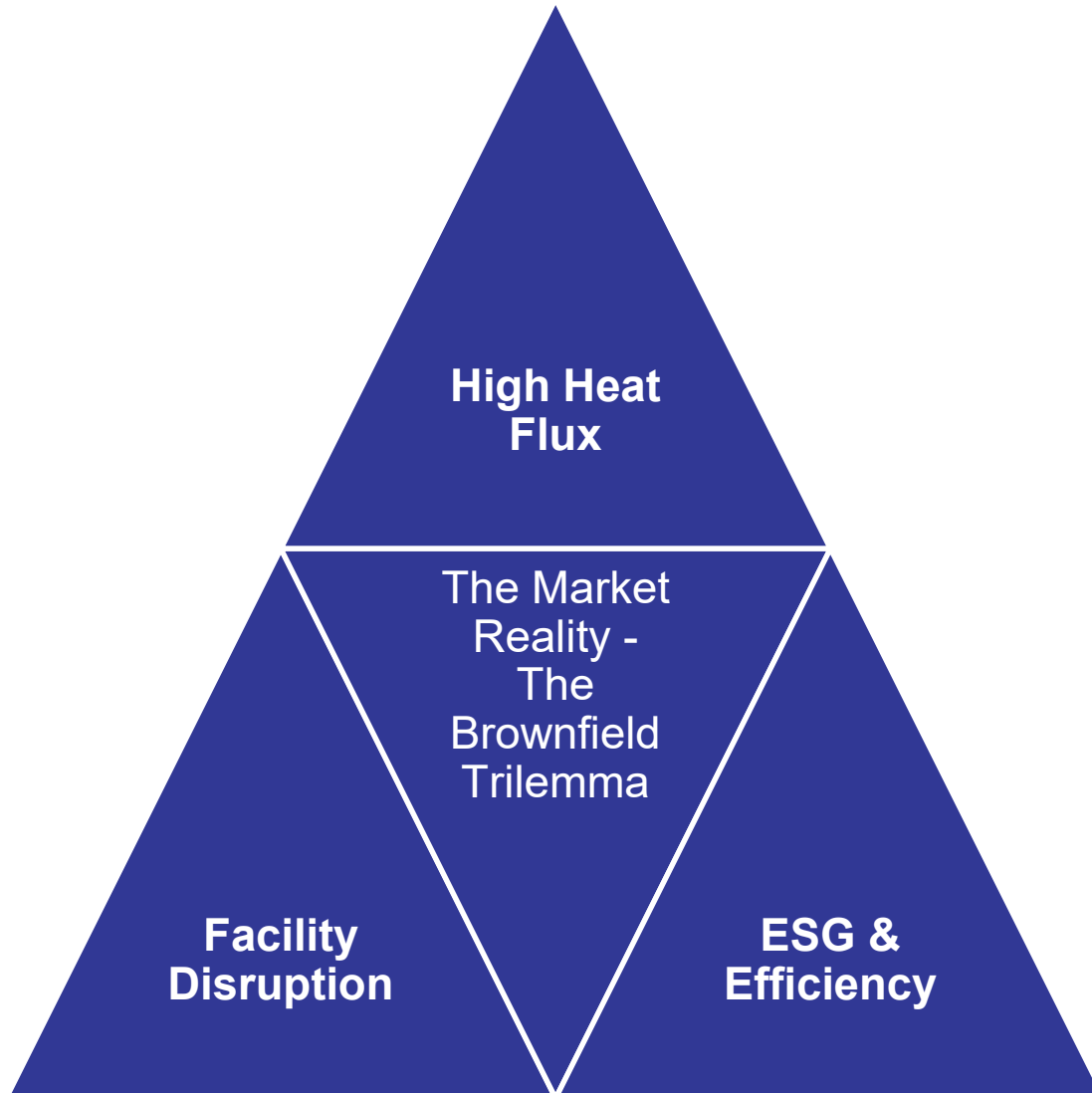
Intelligent factories powered by edge computing, IoT, and automation.



EDGE COMPUTING

THE FOUNDATION OF A CONNECTED, INTELLIGENT FUTURE.

Edge Constraints



Spatial & Structural Limits

Edge facilities lack the physical footprint for massive cooling infrastructure and the floor load capacity to support tons of circulating liquid.

Strict Power Ceilings

With rigid power grid quotas at the edge, every watt consumed by high-pressure cooling pumps is a watt stolen from AI computation.

Severe Water Scarcity

Urban locations and retrofitted cell towers simply do not have the municipal water supply required to operate traditional evaporative cooling systems.

Zero-Touch Reliability

Since edge sites are often unmanned, any risk of water leakage or complex maintenance renders traditional water-cooling catastrophically unviable, demanding dielectric safety.

Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

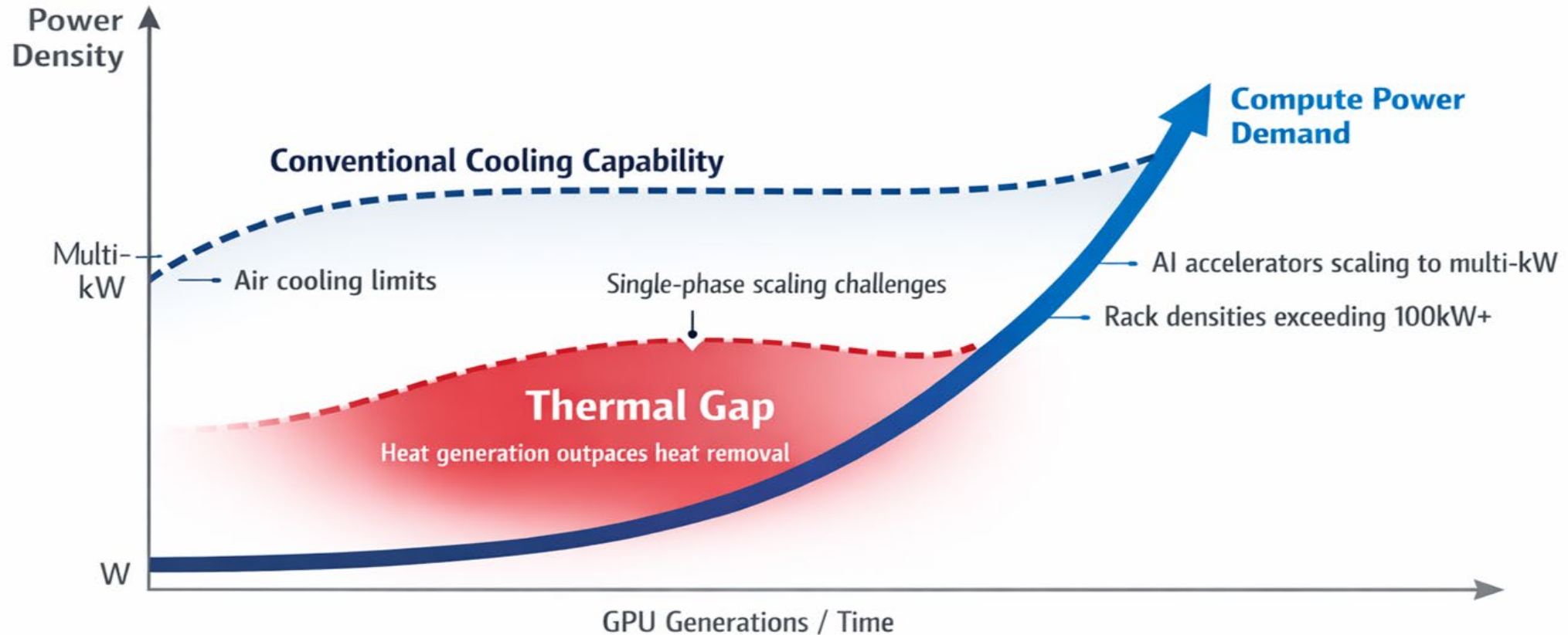
Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

AI Infrastructure Is Outpacing Thermal Design

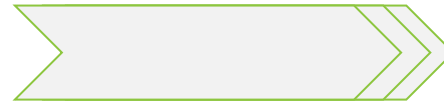


Cooling is no longer a support function — it is a **primary constraint on compute scaling**.

Why do Existing Cooling Methods Break?

AIR COOLING LIMITATIONS

- Limited heat transfer capacity at high densities
- Increasing airflow → **higher energy + mechanical complexity**
- Struggles with **hotspots and thermal variability**



SYSTEM-LEVEL IMPACT

- Thermal constraints → **performance throttling**
- Limits on rack density → **infrastructure sprawl**
- Increasing **power overhead (cooling vs compute)**

SINGLE-PHASE LIQUID LIMITATIONS

- Requires **high flow rates + pumping power**
- Increasing complexity at rack/system level
- Reduced efficiency at **very high heat flux**

Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

Cooling Must Move Closer to the Silicon

Focus on removing heat at the point of generation, not at the system level



PROXIMITY

Cooling applied directly at the silicon through targeted thermal transfer.



EFFICIENCY

Higher heat transfer vs. air-based methods. Transition from bulk air movement.



PRECISION

Targeted cooling of high-density hotspots

What Defines Next Generation Cooling?

HIGH HEAT FLUX HANDLING

- Capable of managing multi-kW device power
- Sustained performance under continuous load

REDUCED MECHANICAL OVERHEAD

- Less reliance on high airflow and large pumping systems
- Improved overall system efficiency

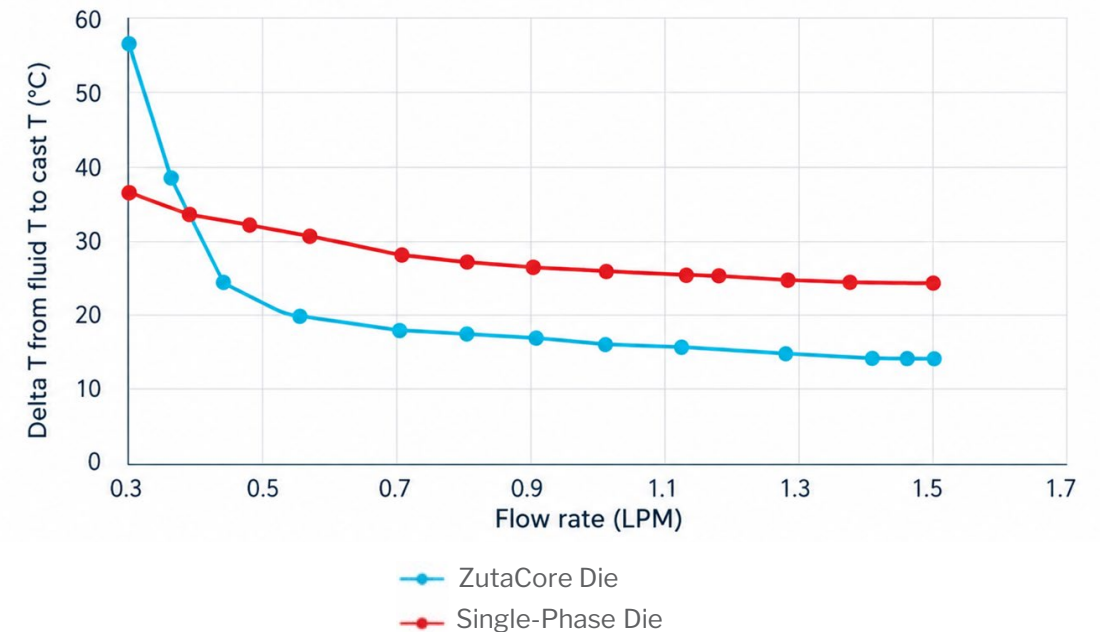
SCALABILITY AT RACK & DATA CENTER LEVEL

- Designed for 100kW+ rack densities
- Modular and repeatable deployment models

THERMAL STABILITY

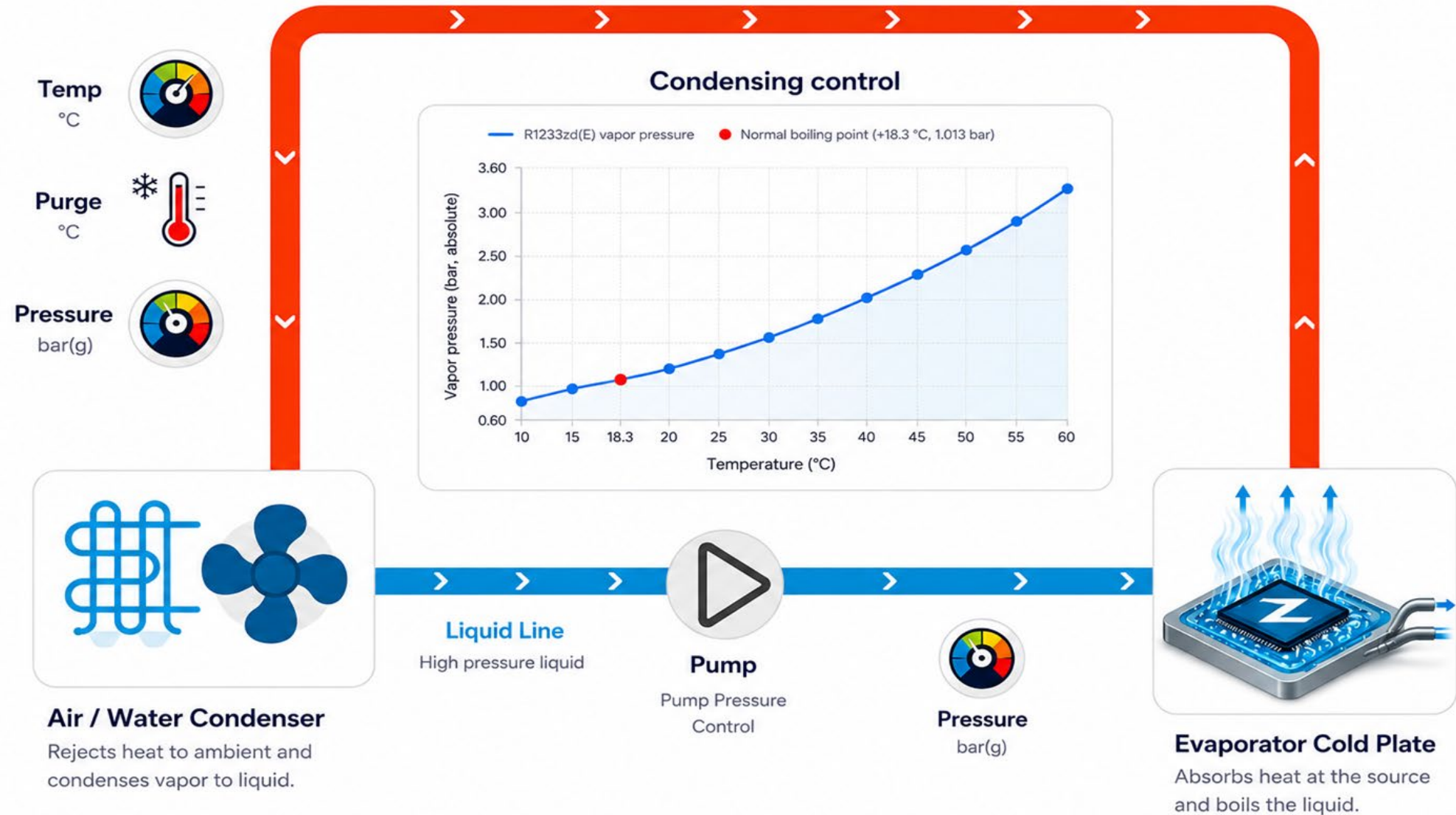
- Minimized hotspots and consistent operating conditions
- Enables predictable system performance

ZutaCore & Single-Phase: Delta T to Flow Rate

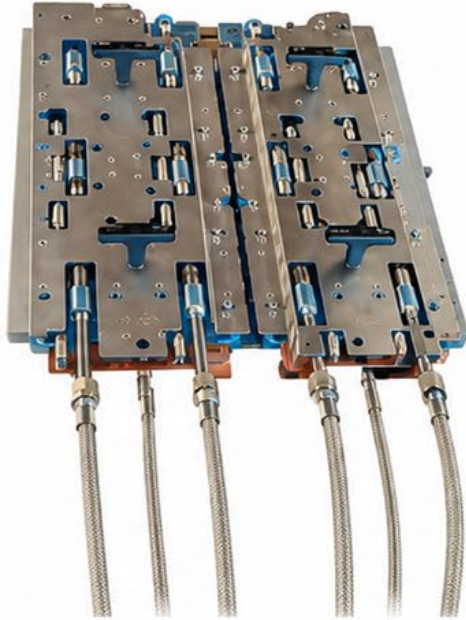


The next generation of cooling must deliver **capacity, stability, and scalability simultaneously.**

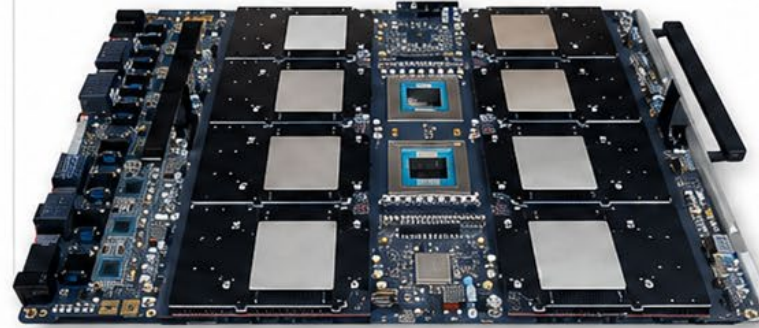
Two-Phase Controlled Cooling Loop



Direct Liquid Cooling within GPU Servers



Cold Plates



GPU Server

- GPU $8 \times 1,100 \text{ W}$
- NVLink $2 \times 425 \text{ W}$
- Retimer $2 \times 100 \text{ W}$
- CPU $2 \times 350 \text{ W}$

Total = 10,500 W



Server Interior



DLC removes ~90% of a 11.55kW Server Thermal Load

For Facilities Without Liquid Cooling Infrastructure

Air Based Side Car



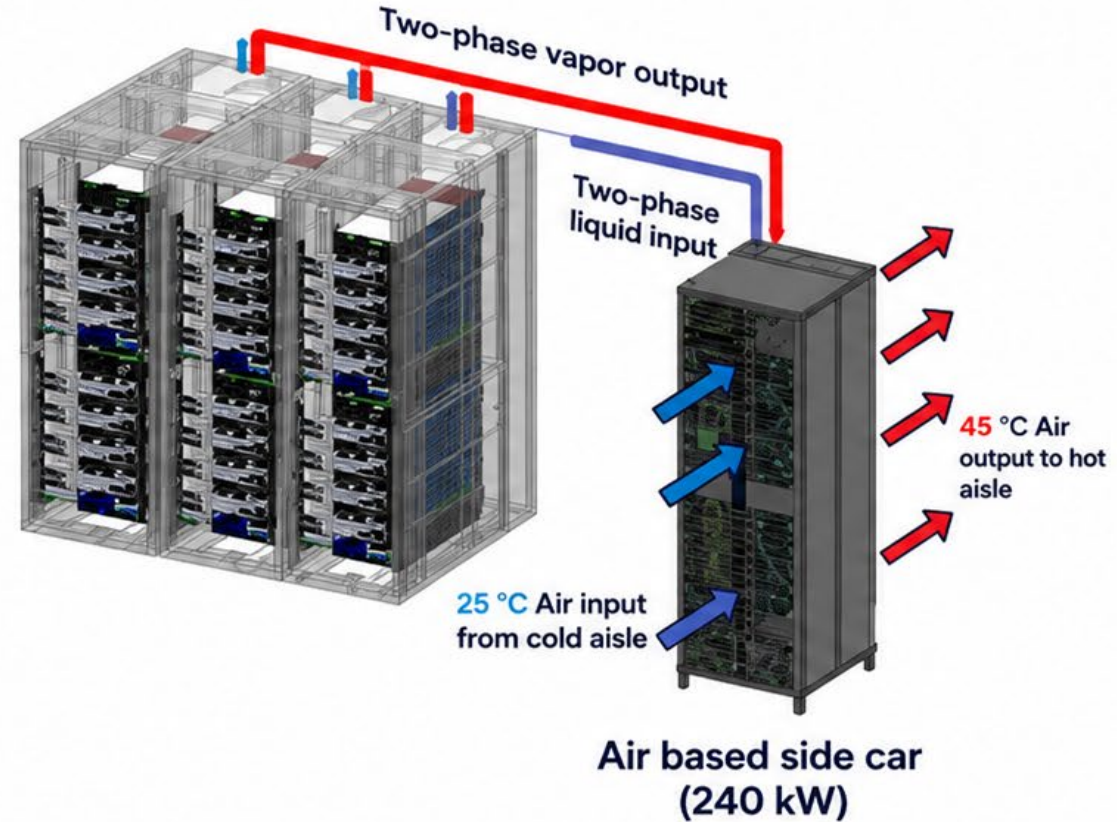
Most data centers still rely solely on air-cooling.

But as servers hit higher densities, air cooling becomes inefficient and uneconomical.



Solution:

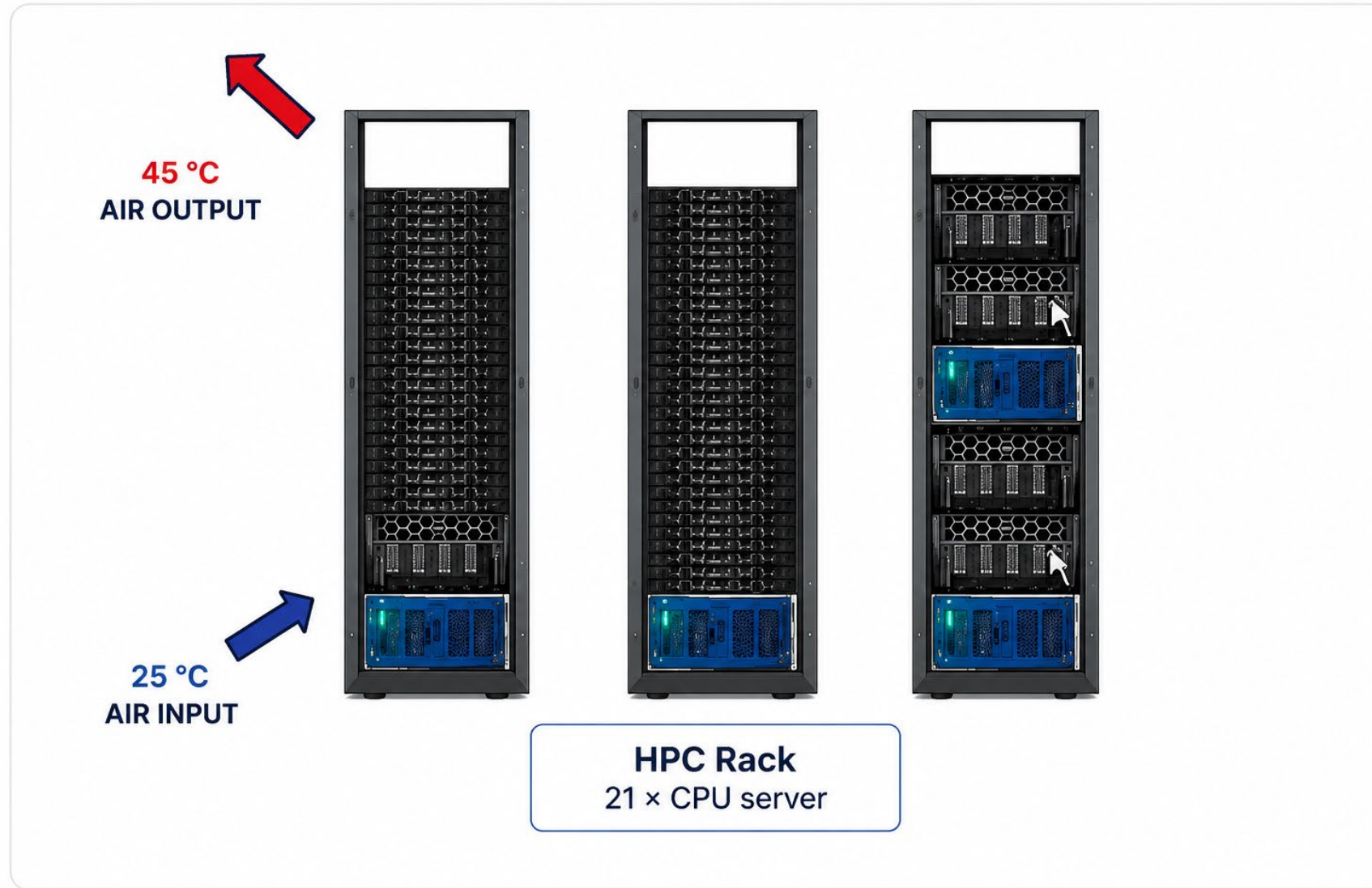
Reject heat to an air-based unit and expels it into the hot aisle — enabling higher rack densities within an air-cooled environment.



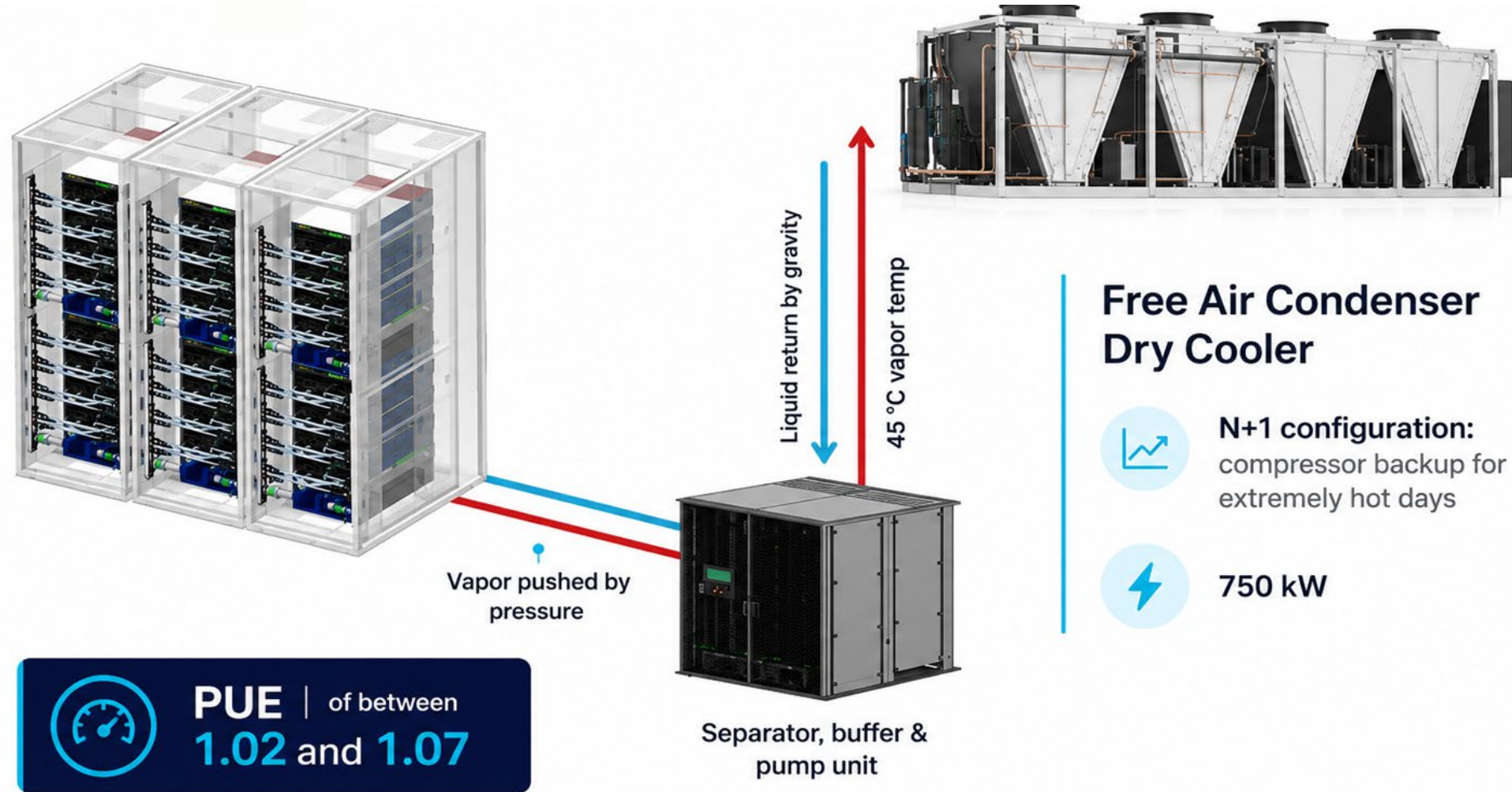
DLC removes ~90% of a 11.55kW Server Thermal Load

For Facilities Without Liquid Cooling Infrastructure

In-Rack CDU Air (16kW)



Single-Loop Two-Phase Cooling with Dry Cooler Condenser

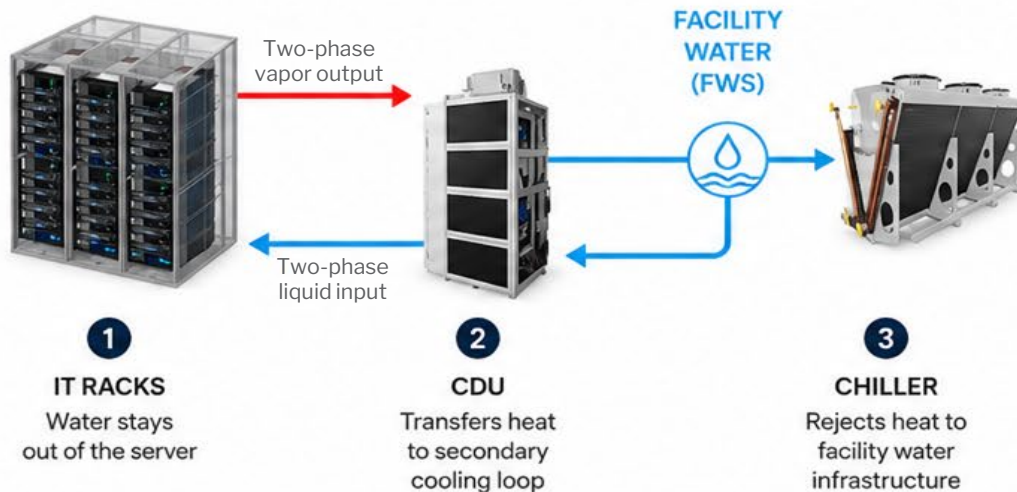


PUE | of between
1.02 and **1.07**

Leveraging Existing Facility Water Infrastructure

WATER-FREE SECONDARY COOLING LOOP

Facility water is used only for heat rejection through the Facility Water System (FWS).



Uses existing facility water infrastructure



Water-free secondary cooling loop

Keeping water out of the servers

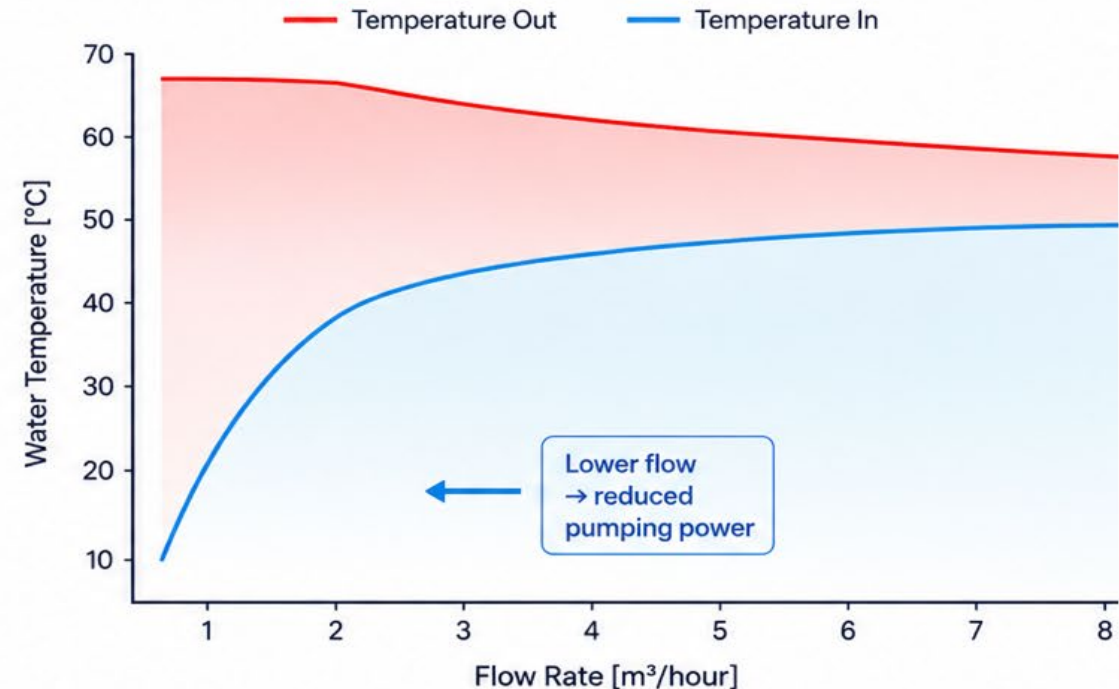


High-temperature operation

Reduces cooling overhead and operating costs

70kW Cooling at Low Flow Rates

Stable thermal performance with reduced pumping requirements



**Lower Flow.
Lower Power.**



Reduced
pumping
energy

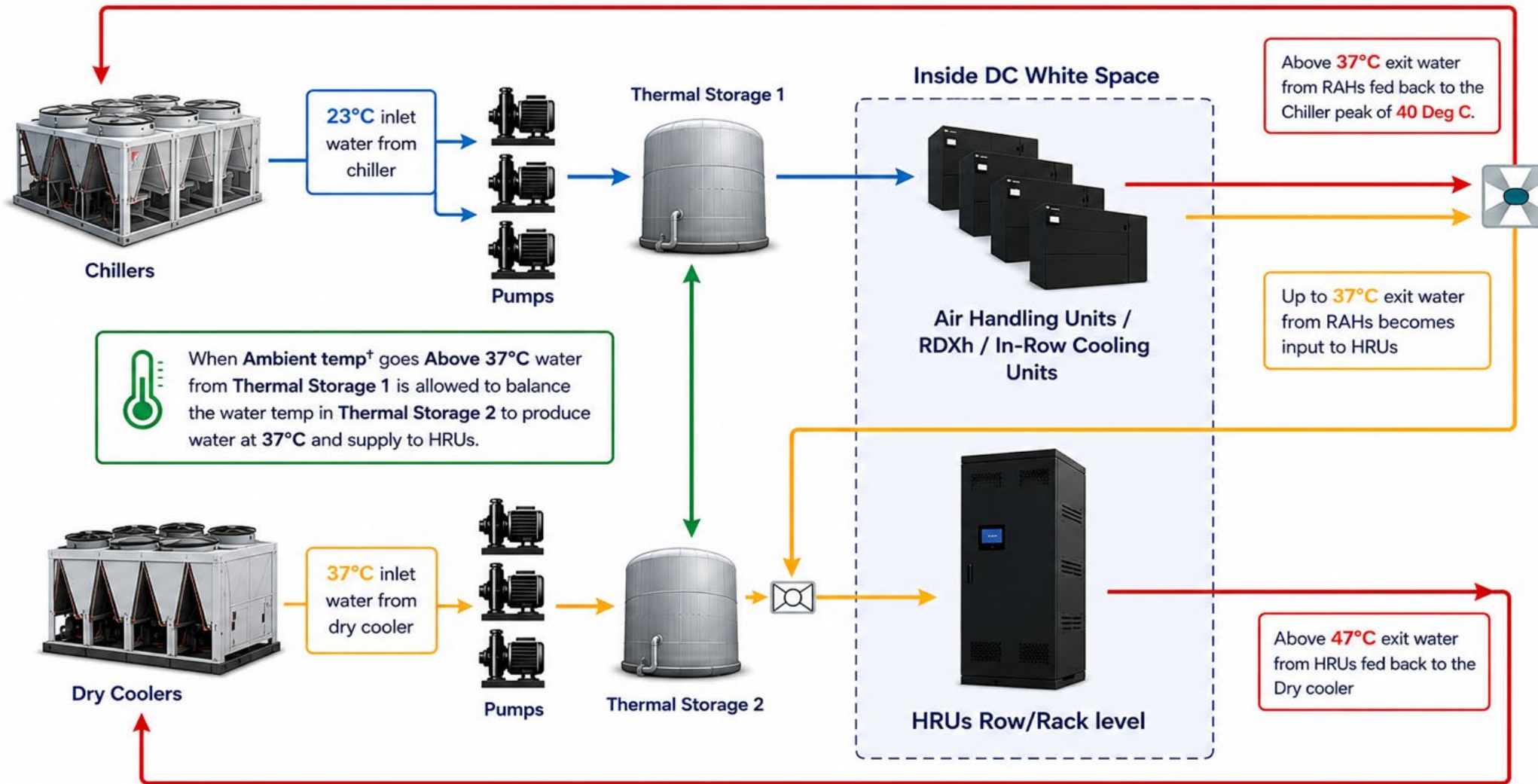


Stable
thermal
operation



Optimized
for AI
infrastructure

High level Data Center Reference Architecture



i * The high-level architecture assumes balanced water inflow and outflow to the chillers and dry coolers, as needed

Audience Poll Question

What is the timeline for implementing at scale two phase direct to chip liquid cooling in you organization

- A.** Already Implemented
- B.** Within 12 months
- C.** Within 24 months
- D.** Within 36 months
- E.** No Plan

Audience Poll Results

- Some discussion facilitated by host and analyst of the results with optional comments from sponsors



Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

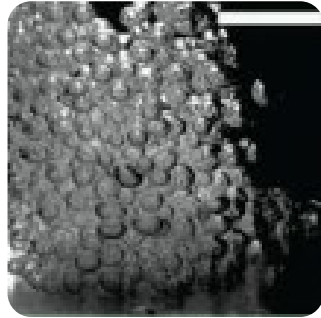
Sponsor Products

Conclusion

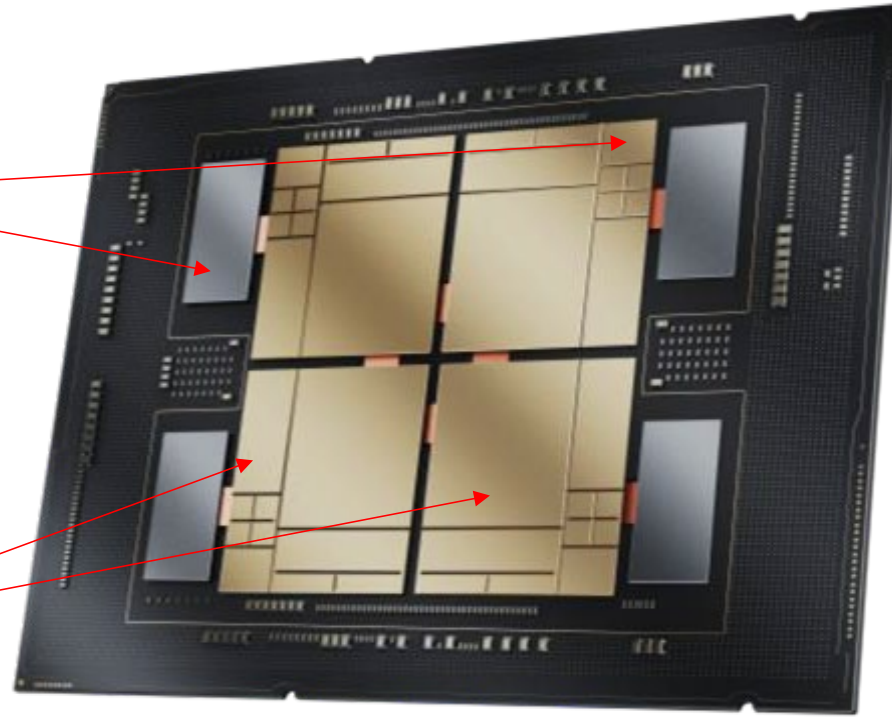
Live Q&A

How Two-Phase Pool-Boiling Works

High Heat, Heavy Boiling



Low Heat, Light Boiling



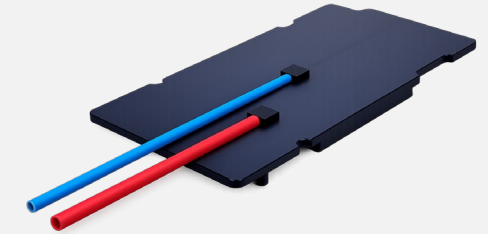
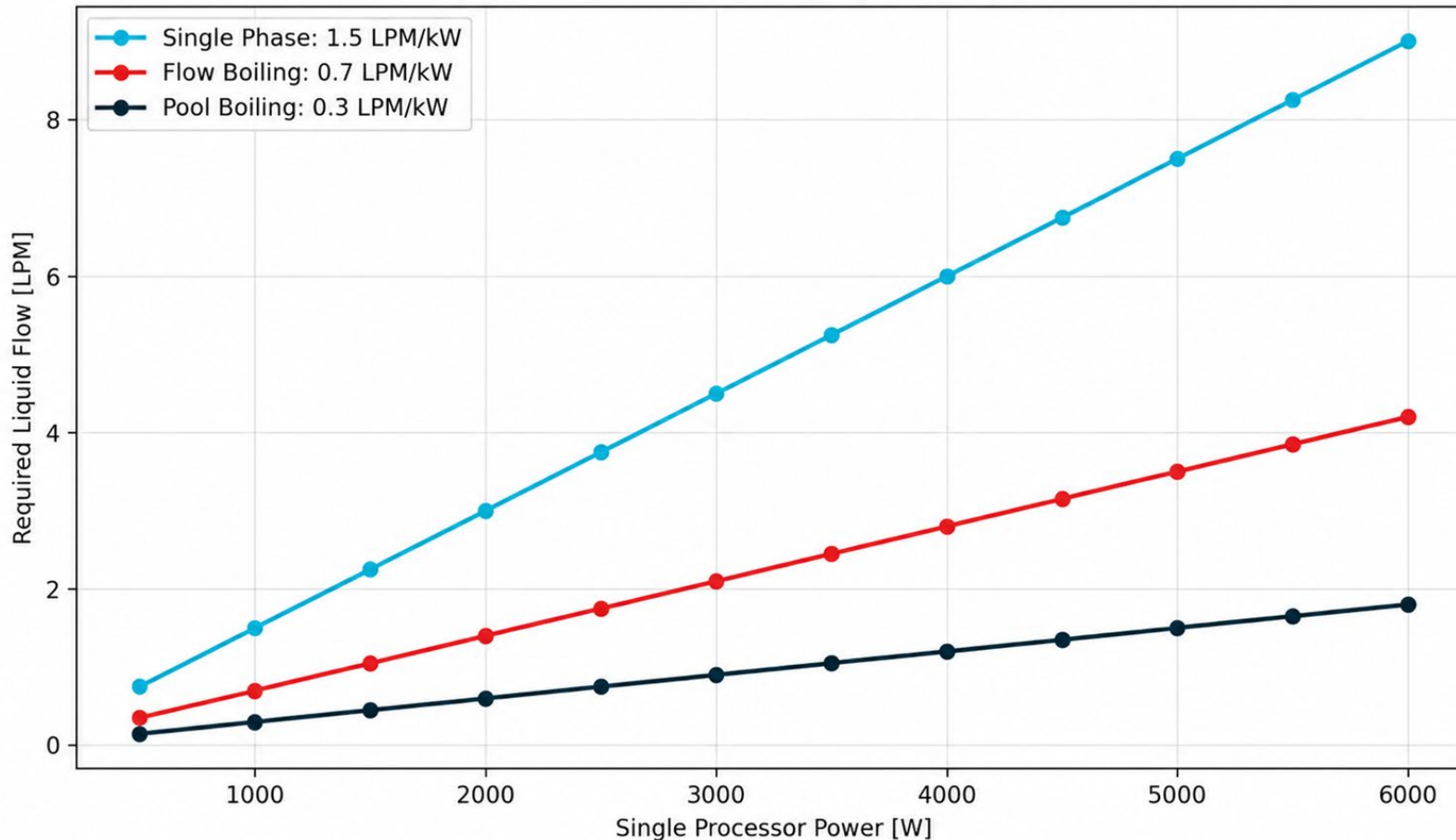
- Micro-structured fins increase surface area at the silicon interface
- Wick structure distributes liquid evenly across the die
- Capillary action continuously replenishes fluid to hot spots

PLAY VIDEO

ZutaCore OmniTherm™ vs Single-Phase

Higher Thermal Efficiency at Lower Flow Rates

Singer Processor Power vs Required Liquid Flow



- **Sustained 600W GPU Cooling** — Enables full NVIDIA RTX PRO™ 6000 Blackwell Server Edition performance without thermal throttling under continuous AI workloads.
- **Full GPU + Memory Coverage** — Cools both the GPU die and 96GB GDDR7 memory for stable operation during memory-intensive AI applications.
- **Waterless Two-Phase Cooling** — Uses phase-change heat transfer for high thermal efficiency and rapid heat removal, with no water inside the server.
- **Single-Slot, Enterprise-Ready Design** — Maintains standard PCIe server compatibility while supporting 24/7 operation in enterprise and AI cloud environments.

Patented Innovation. Engineered Advantage.

65+

GLOBAL PATENTS

01

COLD PLATES



Pool Boiling Technology

Advanced pool boiling for superior heat transfer.



OmniTherm™ Technology

Proprietary two-phase cooling for maximum thermal efficiency.

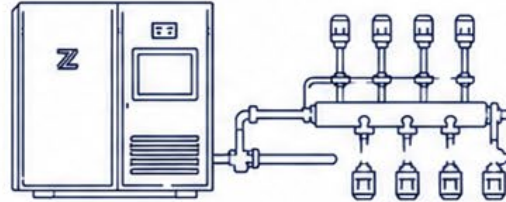


MultiZone Cold Plate

Zoned thermal management for GPU and memory optimization.

02

SYSTEM LEVEL



CDUs

Compact, efficient cooling distribution units.



Manifolds

Precision-engineered manifolds for optimal flow distribution.



ZutaCore HyperCool® Solution

Advanced, waterless cooling designed for modern AI data centers & HPC environments.

03

SUSTAINABILITY



HyperCool Cloud™

Intelligent thermal management at scale for AI data centers.



Heat Reuse & Power Savings

Capture and reuse heat for cooling and other applications—reducing power consumption and emissions.



65 PATENTS



3 TECHNOLOGY CATEGORIES



GLOBAL INNOVATION

Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

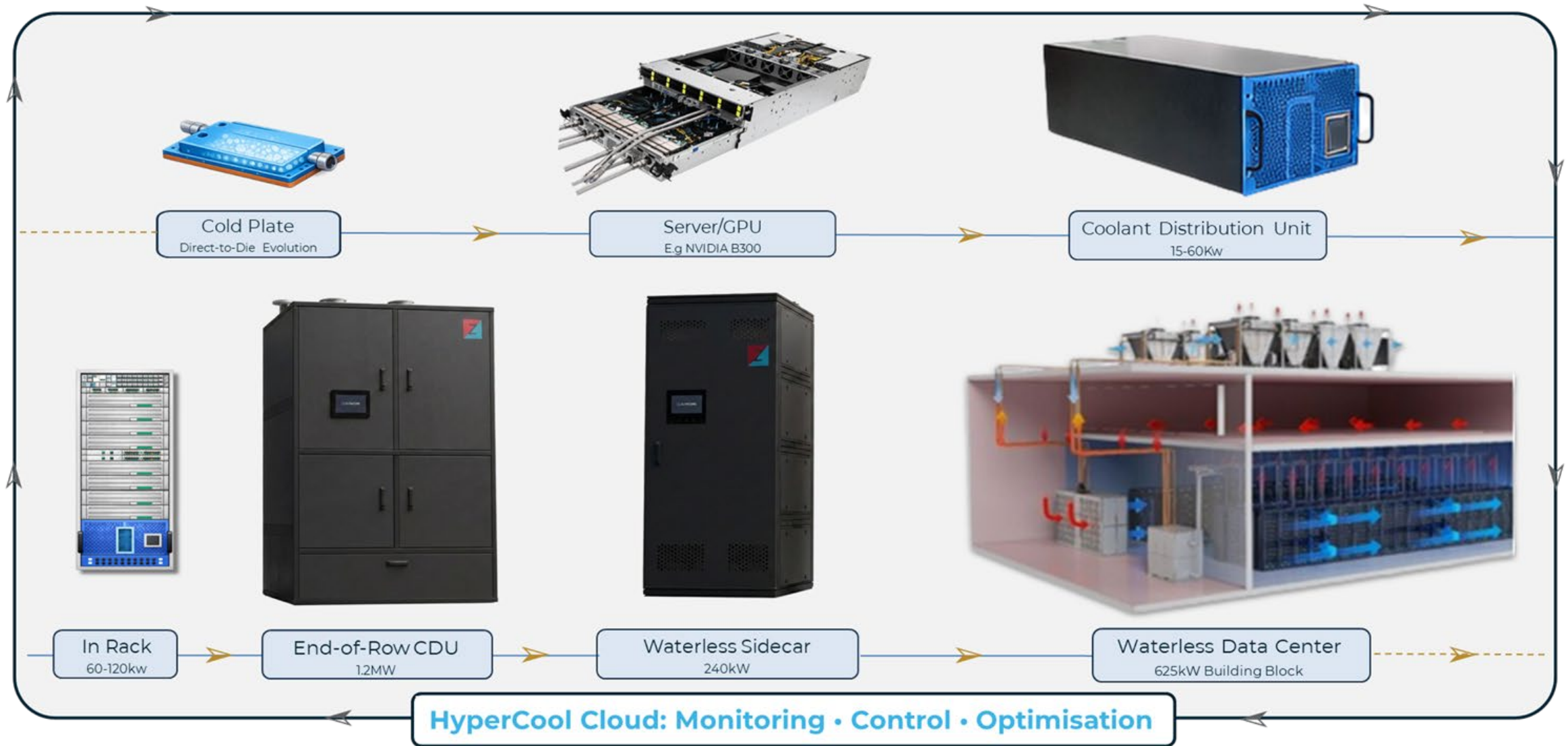
Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

From Rack to Megawatt AI



Oregon State University: Advancing HPC and AI with Two-Phase Liquid Cooling

THE CHALLENGE

Growing AI and HPC workloads pushed OSU's air-cooled data centers to their limits.

- Cooling capacity and energy efficiency constraints
- Limited space for additional cooling infrastructure
- Need to retrofit existing servers and racks

THE SOLUTION

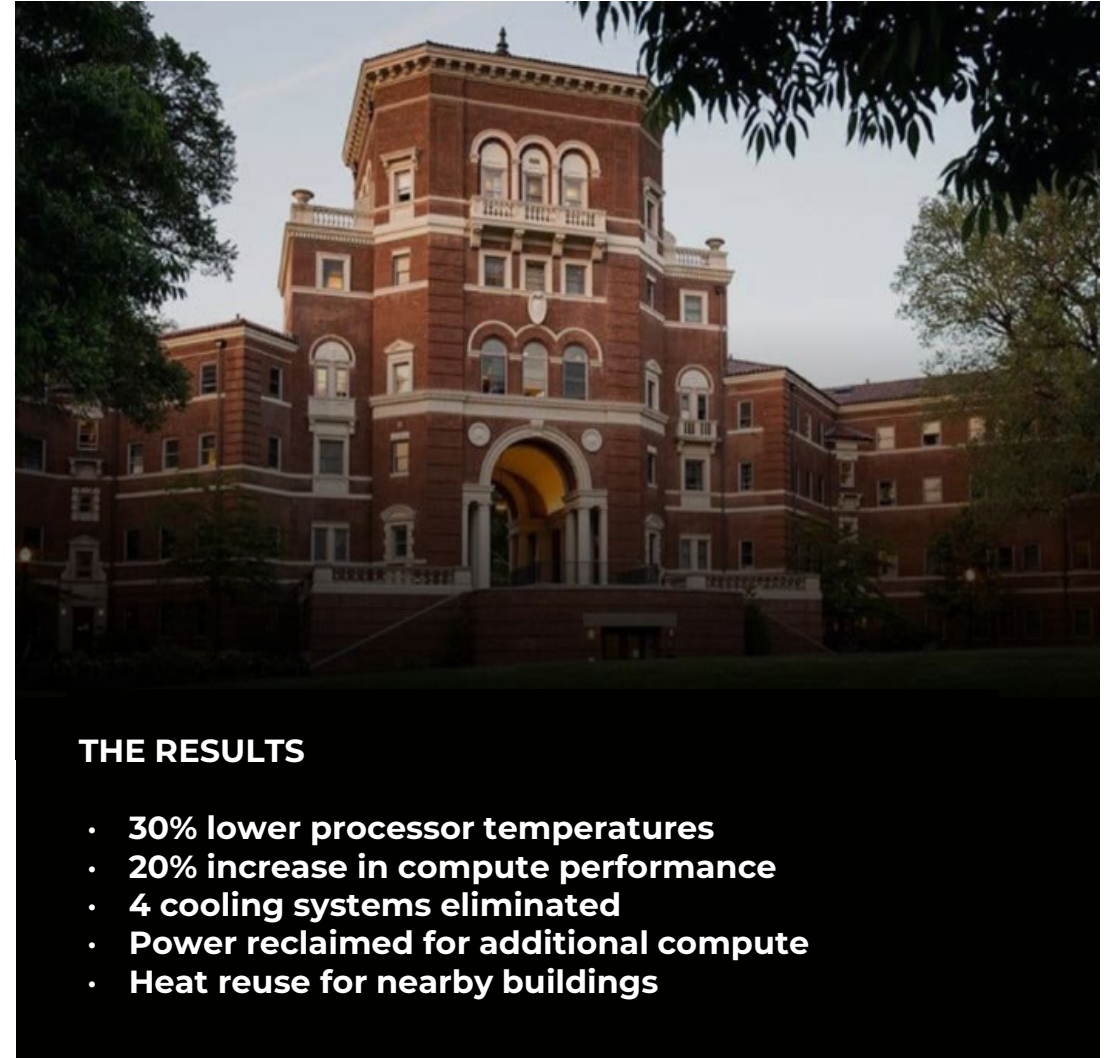
Deployment of ZutaCore® HyperCool® waterless two-phase direct-to-chip liquid cooling.

- Retrofit into existing server infrastructure
- Closed-loop pool boiling architecture
- Integrated with CPI high-density racks
- Compatible with existing chilled water systems

CUSTOMER INSIGHT

Christopher M. Sullivan - Director, Research & Academic Computing

"Two-phase direct-to-chip cooling allowed us to reduce heat load and redirect power previously used for cooling into additional compute capacity within the same space."



Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

ZutaCore® OmniTherm™ Cold Plate

KEY CAPABILITIES

Orientation-agnostic performance

Full thermal efficiency in vertical, horizontal, or inverted deployments

Built for next-gen architectures

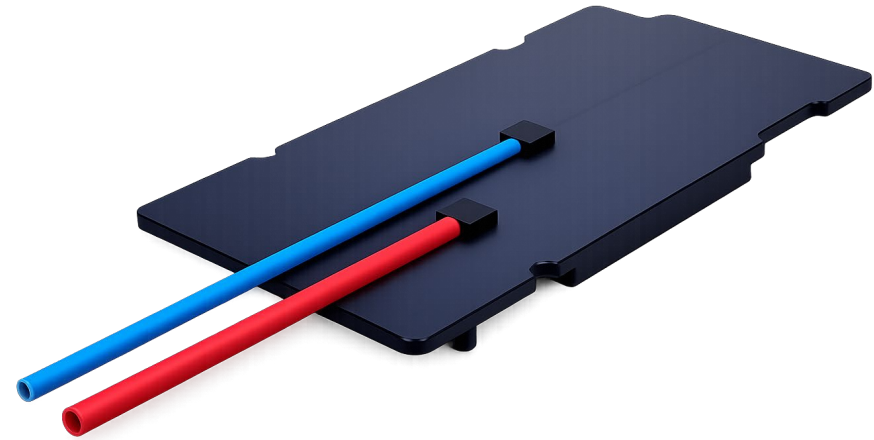
Supports back-side power delivery, PCIe GPUs, and bottom-mounted components

Ultra-thin, direct-to-silicon two-phase cooling

High heat flux removal with passive hotspot mitigation and no complex pumping

Waterless, sealed design

Zero water in the server for reduced risk and simplified deployment



BUSINESS IMPACT

- Maximize GPU performance with sustained power delivery
- Increase density without thermal compromise
- Reduce infrastructure risk and complexity
- Unlock new server designs without cooling limitations

ZutaCore HyperCool Cloud

Cloud-Based Monitoring & Performance Optimization for ZutaCore® HyperCool® Deployments

CENTRALISED MONITORING

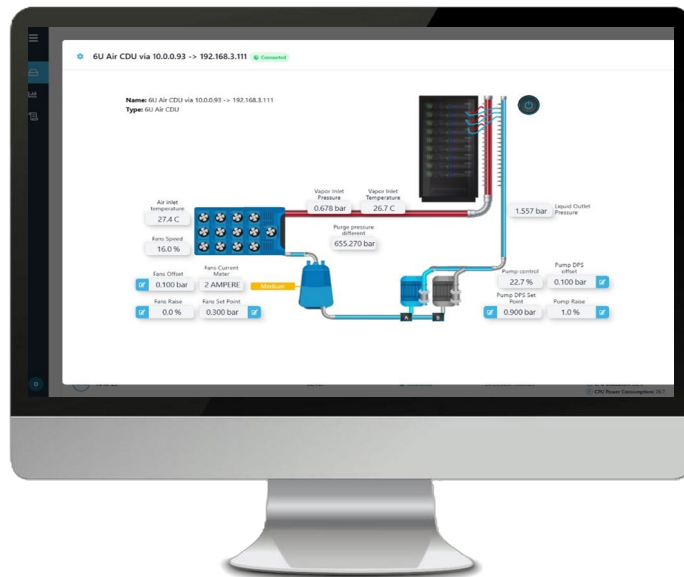
Real-time, multi-site visibility across cooling systems, power, and thermal performance.

PERFORMANCE OPTIMISATION

Controlled tuning capabilities that directly correlate cooling performance with compute output.

ENTERPRISE-GRADE SECURITY

Secure, gateway-based architecture with encrypted communication and role-based controls.



ECOSYSTEM INTEGRATION

API-ready architecture integrating server telemetry, throttling, and performance data.

SCALABLE CLOUD ARCHITECTURE

Designed to scale seamlessly from single racks to multi-megawatt AI deployments.

REMOTE DIAGNOSTIC SUPPORT

Enables proactive monitoring and faster issue resolution across distributed environments.



Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

Take Aways

Market Realities: The Cooling Imperative (The "Why")

- Driven by the aggressive NVIDIA power roadmap, advanced liquid cooling is no longer optional—propelling a projected >\$70 Billion global cooling market by 2036.
- Deploying high-density AI requires infrastructure that overcomes severe limitations in space, power, and water scarcity across both brownfield and edge environments.

The Physics Breakthrough: Why Pool Boiling Wins (The "How")

- Uses micro-structured fins to deliver coolant exactly where the silicon runs hottest, enabling handling multi-kilowatt heat fluxes with low flow rates.
- Crucially for edge and enterprise deployments with its entirely waterless inside the server.
- OmniTherm cold plates are orientation-agnostic and scalable up to 100kW per rack.

As traditional thermal architectures hit their physical limits, two-phase pool boiling emerges not just as a cooling technology, but as one of the key enablers for scalable, reliable next-generation AI Infrastructure.



Agenda Overview

Cooling AI Clusters: Market Overview

Current Cooling Distribution Limitations

New High Density Server & Facilities Cooling Solutions

New Data Center Cooling Technologies – Deeper Look

Large Scale Deployments Have Begun

Sponsor Products

Conclusion

Live Q&A

Beyond Conventional Two-Phase Cooling

A New Approach for Scalable AI Infrastructure

Cliff Grossner, Ph.D.
Chief Innovation Officer
OCP Foundation



Xiaoxi He, Ph.D.
Principal Research Associate
IDTechEx



Shahar Belkin
Chief Evangelist
ZutaCore



Thank you for attending

Beyond Conventional Two-Phase Cooling

A New Approach for Scalable AI Infrastructure

OCP Educational Webinar Series

May 19th 2026



IDTechEx



Community-driven hyperscale innovation for all