MiTAC Computing | AMD together we advance_

Scaling the Future of Open AI Infrastructure with Advanced Cooling

Raymond Huang
General Manager, MiTAC Computing Technology USA

Sachin Dhingra
Sr. Director, Enterprise AI Solutions, AMD

Oct.14, 2026

Scaling Open AI Infrastructure

Two Barriers. One Scalable Future.

ARCHITECTURE BARRIER

The challenges of closed ecosystems

- **Vendor lock-in**
  Limited choice and flexibility
- **Fragmented software**
  Complex integration and management
- **Management silos**
  Higher operational overhead

PHYSICAL BARRIER

The limits of today's infrastructure

- **Power**
  Rising energy demand
- **Heat**
  Thermal limits of air cooling
- **Rack density**
  Increasing compute per rack

OPEN INFRASTRUCTURE

Remove the architectural constraints

- Open hardware and standards
- Open software ecosystem
- Choice and interoperability
- Simplified operations

ADVANCED COOLING

Remove the physical constraints

- Higher density
- Stable and efficient operation
- Lower total cost of ownership

SCALABLE AI FACTORY

Higher Performance | Greater Flexibility | A More Sustainable Future

MiTAC Computing

AMD
together we advance_

Enabling Open Standard Hardware Architecture with MiTAC OCP Multi-node Servers

OPEN Compute Project®

Type	Air-cooled	Liquid-cooled	Air-cooled
Picture			
Model Name	C2810Z6/ C2610Z6 (AMD EPYC™ 9006)	C2611Z6 (AMD EPYC™ 9006)	Capri 3 Ultra Capri 3 Standard (AMD EPYC™ 9005/9004)
Form Factor	2OU2N	2OU4N	2OU3N

MiTAC Computing

AMD
together we advance_

MiTAC Diversified AI Infrastructure Solutions with AMD

MiTAC B8050/ MiTAC TN85-B8261

- Power-efficient, air-cooled for distributed AI inference
- Lower TCO for GenAI at the edge



Workstation/Server

AMD Radeon™
AI PRO R9700S

- Edge AI, retail
- Image generation
- Traditional AI (CNN, DLRM)
- SLM inference

MiTAC TN85-B8261

- Flexible enterprise AI inference
- High performance with efficient agent logic & multi-tasking



Server

AMD Instinct™
MI350P (PCIe)

- Mainstream LLM inference
- Agentic AI
- Content generation
- RAG

MiTAC G8825Z5 (Air-cooled)

- GPU intensive AI compute
- Air-cooled
- Massive memory & high-speed NVMe storage



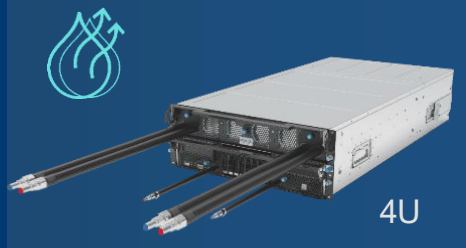
AI-Optimized Server

AMD Instinc™
MI350X x8 (UBB)

- LLM inference
- Model fine tuning
- SLM training
- Code generation

MiTAC G4826Z5 (Liquid-cooled)

- Extreme-scale AI compute
- Liquid-cooled
- High CPU core density, high memory & ultra-fast storage



AI-Optimized Server

AMD Instinct™
MI355X x8 (UBB)

- LLM inference
- Model fine tuning
- SLM training
- Code generation

MiTAC Rack-scale Solutions

- AI liquid-cooled rack
- AI air-cooled rack
- HPC liquid-cooled rack
- HPC air-cooled rack



Rack

AMD Instinct™
MI355X/MI350X x8 (UBB)

- Hyperscaler/CSP inference
- Frontier model training
- Low latency

AMD EPYC™ 9005 Series and Next-gen CPU

MiTAC Computing

AMD
together we advance_

MiTAC Rack-Scale Solutions with AMD

AMD EPYC | INSTINCT AMD PENSANDO AMD ROCm

AI Liquid-Cooled

AMD Instinct™ MI355X GPU



AMD EPYC | INSTINCT AMD PENSANDO AMD ROCm

AI Air-Cooled

AMD Instinct™ MI350X GPU



AMD EPYC

Storage Air-Cooled

Integrated with DDN Infinia



AMD EPYC OPEN COMPUTE PROJECT

HPC Liquid-Cooled

AMD EPYC™ 9005 CPU



AMD EPYC OPEN COMPUTE PROJECT

HPC Air-Cooled

AMD EPYC™ 9005 CPU



Open Ecosystem for AMD AI Infrastructure

From Open-Source FW to Full AI SW Stacks for AMD EPYC™ and Instinct™ MI355X Platforms

Open Source Stacks
From Firmware to Infra/POD Management

MiCoreView™
Infra / Rack / POD Management
- Monitoring
- Provisioning
- FW/OS Deployment
- Telemetry & Analytics
- POD Operations

MiRMC FW Coming Soon
MiOBMC™ | MiOPF™
Firmware Layer
OpenBMC-based | Open Platform Firmware
- Platform Management
- Security & Updates
- Hardware Control
- Rack Management
- Open Standards

AMD Platform
AMD
EPYC™ Server CPU
Instinct™ MI355X GPU

Open | Interoperable | Scalable

MiTAC MR1100
AMD Instinct™ MI355X GPU Rack Solution



High-Density | Liquid-Cooled | AI-Ready

AMD GPU / HPC Software Stack
Optimized for AI and HPC Workloads

RAFAY	GPU PaaS & Orchestration Kubernetes | Multi-Cluster | GPU as a Service	Service & Orchestration
AMD Enterprise AI	Enterprise AI Software Stack AI Workbench | AIMs | Resource Manager	AI Platform
AMD ROCm	GPU Compute Runtime Libraries | Tools | Optimization	Compute Runtime
ddn infinia	AI Data Intelligence Platform High-Performance | Scalable | Multi-Cloud	Data & Storage
Ubuntu	Linux OS for AI Infrastructure Stable | Secure | Cloud-Native	Operating System

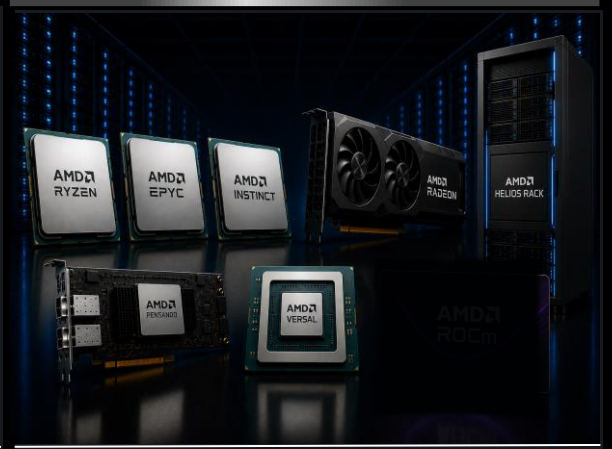
Open Software | Proven Partners | Accelerated AI

Public

High-Performance Computing Leadership



Leadership Foundational IP



Broad Product Portfolio



Open, Proven Software Ecosystem



Deep, Collaborative Partnerships

AMD
together we advance_

Public

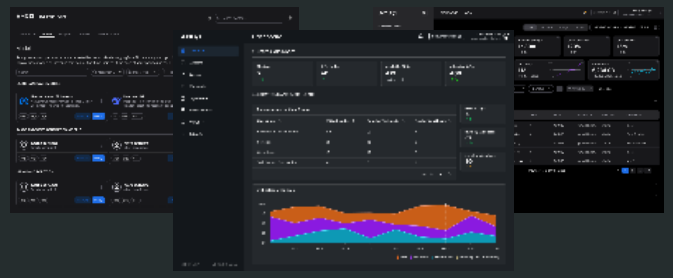
AMD Enterprise AI Solutions

Seamless Enablement of High Value Enterprise Applications



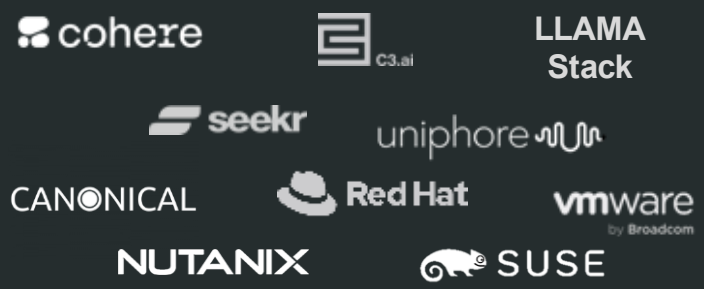
High Performance Platforms

Delivering Performance and TCO value; On-prem & Cloud



Enterprise AI Software

Complete solutions for deploying & customizing industry-standard AI models



Strategic Partnerships

Enable high-impact use cases optimized for AMD platforms

Complete Solutions = Hardware + Software + ISV partners

AMD
together we advance_

Public

Deploy the Right AMD Platform for Every Enterprise Workload

Start with the use case. Match the platform. Scale on one AMD software foundation.

LOCAL/ AGENTIC



AMD Ryzen™ AI Halo

Build local AI agents

1x users USER SCALE	
300B MODEL SIZE	192 GB MEMORY
Desktop FORM FACTOR	AIR | 120W SYSTEM POWER
Agent Apps	Sensitive Docs
Personal Assistant	Code Edits

EDGE/ PROTOTYPE



Radeon™ AI PRO R9700

Prototype. Deploy. Repeat

100x users USER SCALE	
35B MODEL SIZE	32 GB MEMORY
Wkstn/Server FORM FACTOR	AIR | 1-15kW SYSTEM POWER
Edge AI	Retail AI
Image Gen	SLM Inference

SCALE/ ENTERPRISE

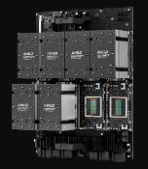


Instinct™ MI350P

Scale Enterprise inference

1,000x users USER SCALE	
175B MODEL SIZE	144 GB MEMORY
Server FORM FACTOR	AIR | 15-30kW RACK POWER
RAG Chat	Copilots
Content Gen	Production LLM

INDUSTRIALIZE/ AI FACTORY



Instinct™ MI355X x8

Build the AI factory

10,000x users USER SCALE	
1T+ MODEL SIZE	2.3 TB MEMORY
AI Server FORM FACTOR	AIR / LC | 30-60kW RACK POWER
Large LLM	Fine-Tuning
Code Gen	Training

HYPERSCALE/ RACKSCALE



AMD Helios

72× MI455X GPUs

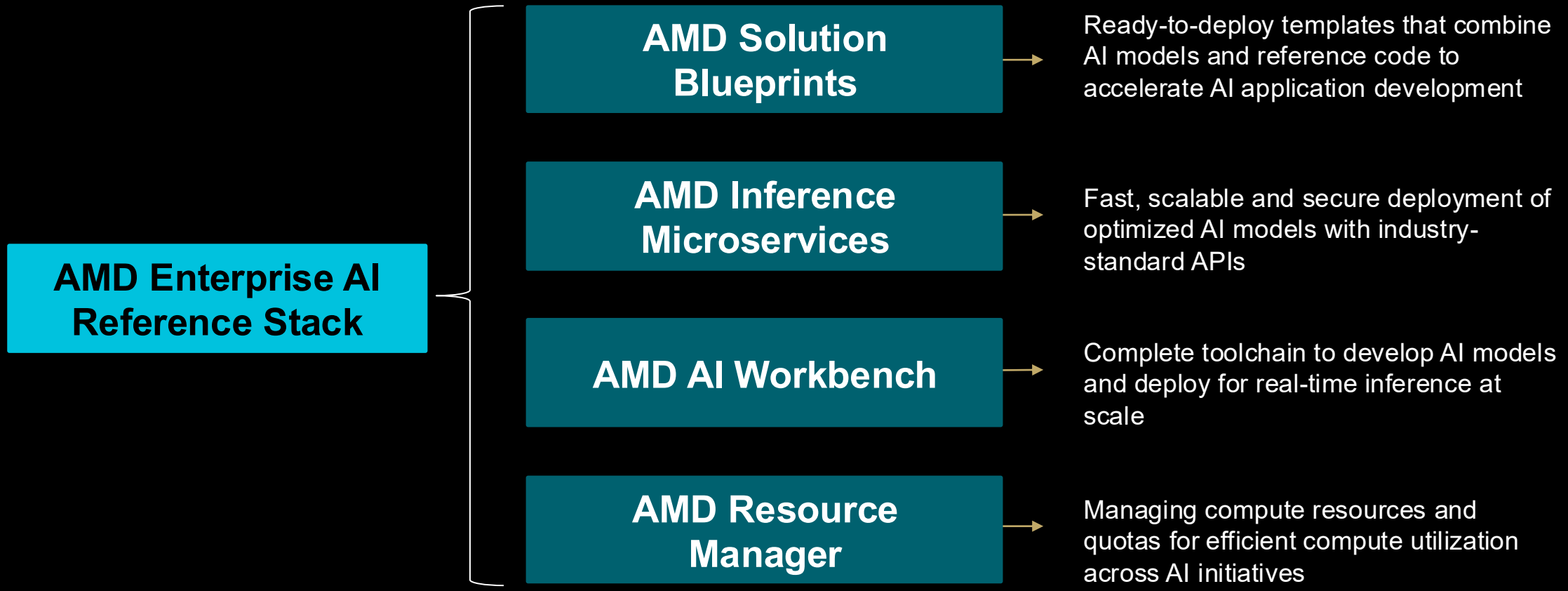
100,000+ users USER SCALE	
>1T MODEL SIZE	31 TB MEMORY
Rack FORM FACTOR	LC | 100kW+ RACK POWER
Hyperscale AI	Sovereign AI
Frontier Train	Low-Latency

AMD
together we advance_

Public

AMD Enterprise AI Software Stack

Removes underlying complexity, making AI deployments simple



AMD
together we advance_

MiTAC Diversified AI Infrastructure Solutions with AMD

MiTAC B8050/ MiTAC TN85-B8261

- Power-efficient, air-cooled for distributed AI inference
- Lower TCO for GenAI at the edge



Workstation/Server

AMD Radeon™
AI PRO R9700S

- Edge AI, retail
- Image generation
- Traditional AI (CNN, DLRM)
- SLM inference

MiTAC TN85-B8261

- Flexible enterprise AI inference
- High performance with efficient agent logic & multi-tasking



Server

AMD Instinct™
MI350P (PCIe)

- Mainstream LLM inference
- Agentic AI
- Content generation
- RAG

MiTAC G8825Z5 (Air-cooled)

- GPU intensive AI compute
- Air-cooled
- Massive memory & high-speed NVMe storage



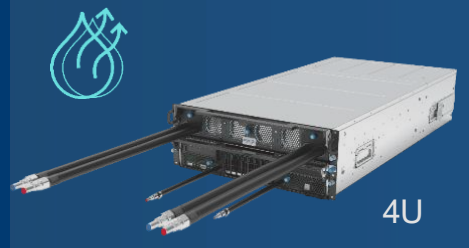
AI-Optimized Server

AMD Instinc™
MI350X x8 (UBB)

- LLM inference
- Model fine tuning
- SLM training
- Code generation

MiTAC G4826Z5 (Liquid-cooled)

- Extreme-scale AI compute
- Liquid-cooled
- High CPU core density, high memory & ultra-fast storage



AI-Optimized Server

AMD Instinct™
MI355X x8 (UBB)

- LLM inference
- Model fine tuning
- SLM training
- Code generation

MiTAC Rack-scale Solutions

- AI liquid-cooled rack
- AI air-cooled rack
- HPC liquid-cooled rack
- HPC air-cooled rack



Rack

AMD Instinct™
MI355X/MI350X x8 (UBB)

- Hyperscaler/CSP inference
- Frontier model training
- Low latency

AMD EPYC™ 9005 Series and Next-gen CPU

MiTAC Computing

AMD
together we advance_

Public



AMD Instinct MI350P

The Easiest Way to Add AI to Your Data Center

PERF/$ LEADERSHIP	LARGE MODEL CAPABLE	STANDARD SERVER COMPATIBLE
Up to **4.2x** Tokens/second/$ vs **RTX Pro 6000**	Up to **260 Billion** Model parameters @ FP4	**Air Cooled** 450-600W TBP

MI350P sweet spot: private, memory-intensive production inference in familiar enterprise servers.

See Endnote: MI350P-007

Open architecture gives us the flexibility to scale.

But higher compute density creates another constraint:

HEAT

MiTAC Computing

AMD
together we advance_

Open Infrastructure Extends Beyond Compute

It also means choice across the thermal ecosystem.

Cold Plate Choice

JWS / Cooler Master

COOLER MASTER | JWS



Proven designs for diverse AI workloads.

Coolant Choice

Castrol / PFX

qualified alternatives

Castrol CORE | PFX PERFORMANCE FLUID EXPERTS GROUP



Reliable, high-performance coolants for long-term sustainability.

Cooling Architecture Choice

Single-phase | Two-phase | Diamond

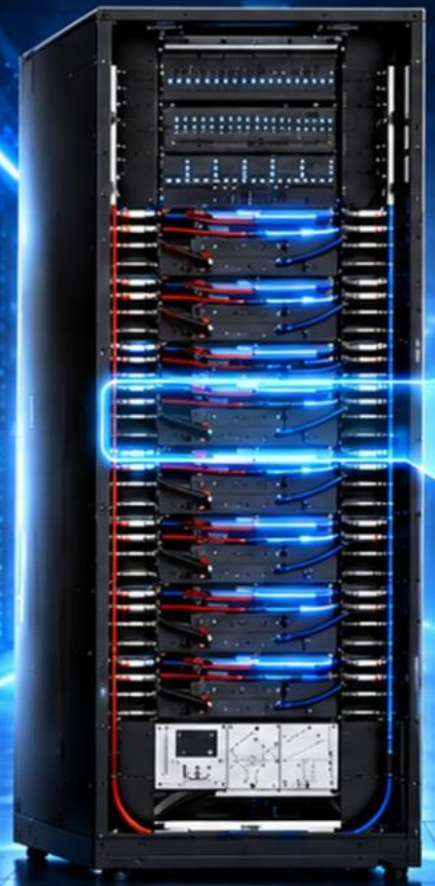
Single-phase	Two-phase	Diamond Cooling
CoolIT Systems	ZUTACORE	AKASH SYSTEMS
Nidec –All for dreams–		



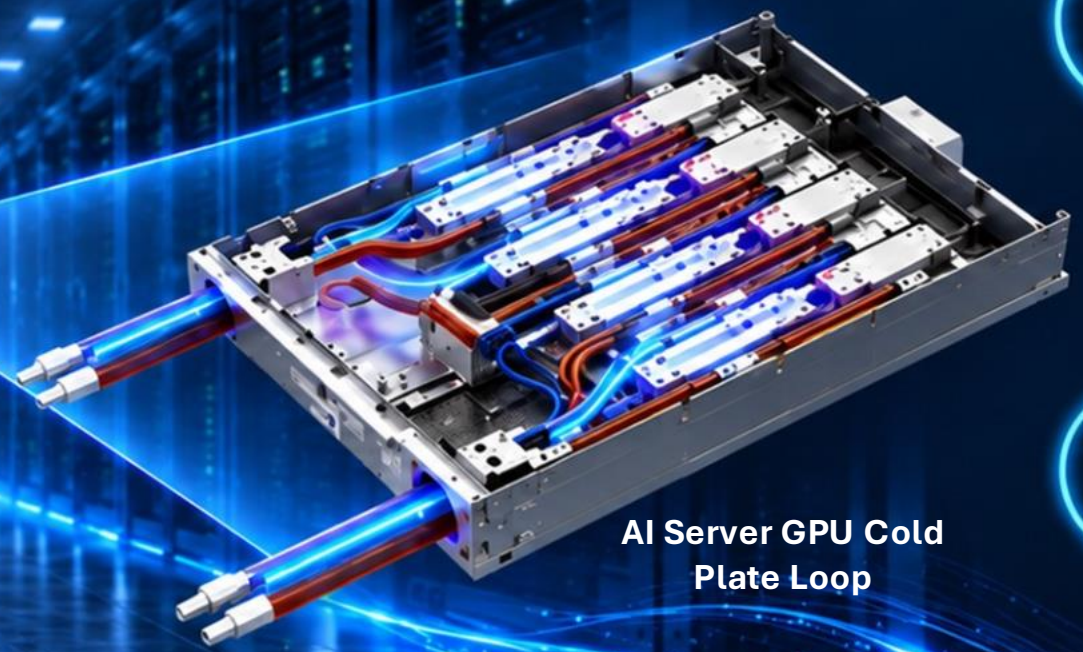
Flexible thermal architectures to support higher densities.

AI-Scale Liquid Cooling, Ready for Volume Deployment

From AMD-optimized components to large-scale deployment, Cooler Master enables high-density AI infrastructure.



AI Liquid-cooled Rack



AI Server GPU Cold Plate Loop

AMD-Optimized GPU Cold Plates

Precision-engineered cold plates optimized for AMD AI accelerators, enabling efficient heat transfer and sustained compute performance.

High-Capacity Rack Manifolds

High-flow coolant distribution engineered for uniform cooling across high-performance AI rack deployments.

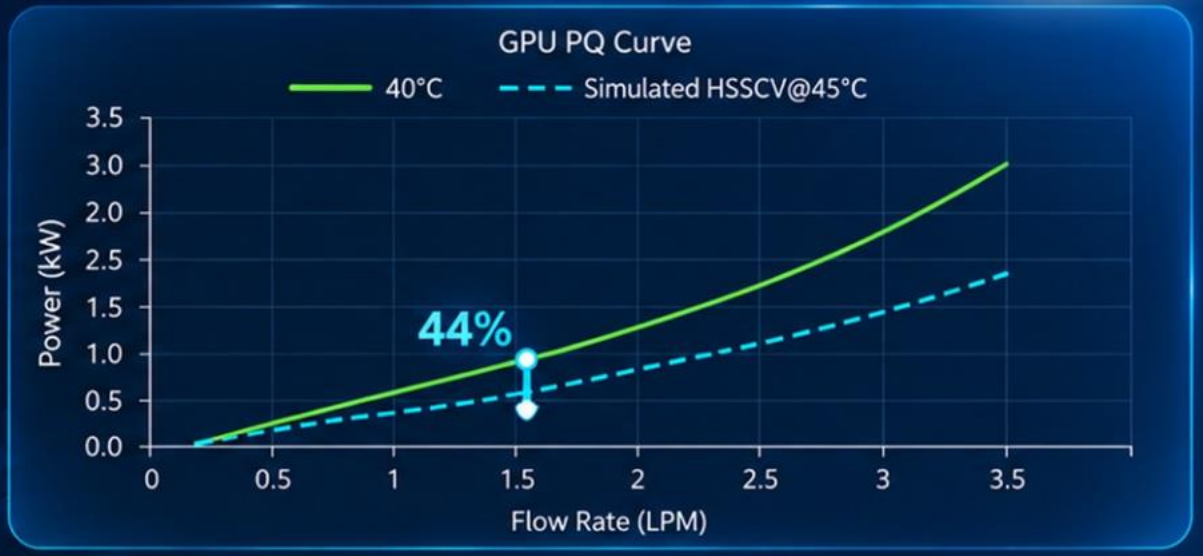
MiTAC Computing

AMD
together we advance_

Single-Phase Cooling with JWS

HSSVC Micro-Channel Optimization





More Uniform Temperature Distribution

Precision fluid paths maximize contact area directly over peak accelerator hotspots.

Minimized Pressure Drops

Eliminate 44% pressure drop at GPU cold plate.
Lowers pump power requirements across multi-node loops.

POC Results

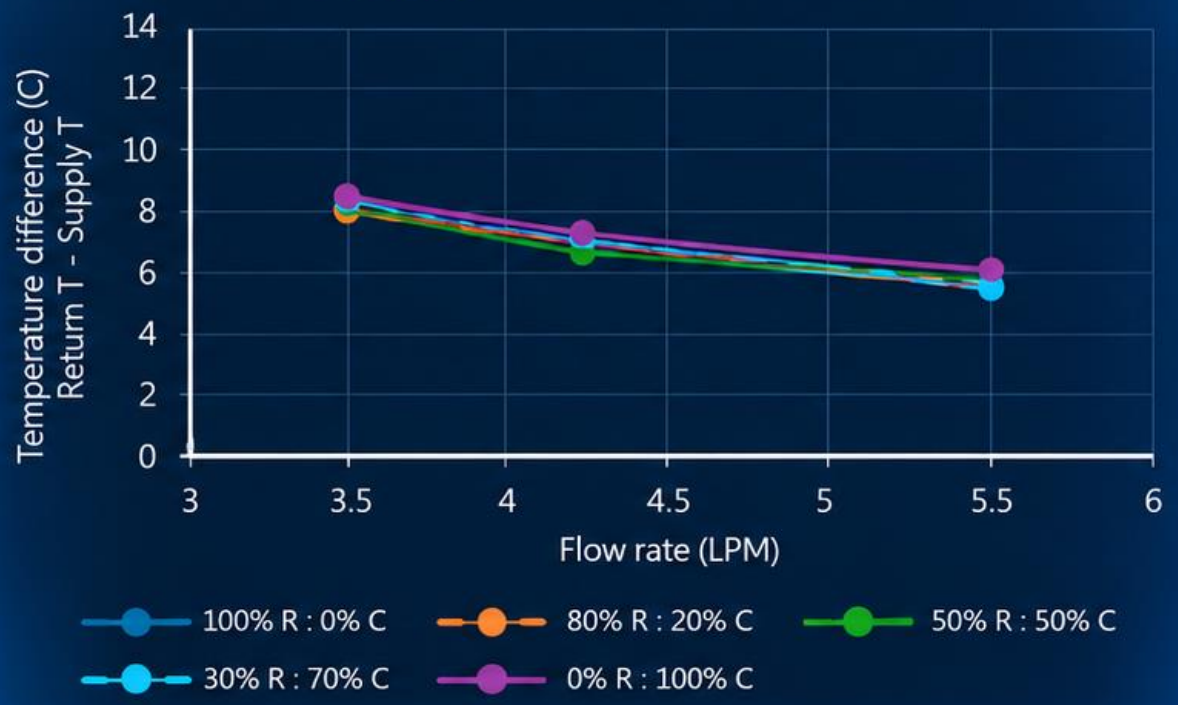
Minimizing pressure drop by 44% optimizes CDU power requirements and enhances overall heat dissipation efficiency.

Supply Chain Resilience: Castrol PG25 Coolant

A certified second source turns cooling chemistry into deployment assurance.



Flow rate & ΔT curve for each PG25 ratio
at 45 deg.C supply liquid temperature



Mixed Coolant Compatibility
Enables seamless multi-coolant operations.

Mitigating Single-Source Risks
Prevents deployment delays from proprietary coolant shortages.

Validated with AMD Instinct MI355X GPU Rack-Scale Solutions
Tested by MiTAC thermal engineering team.

Castrol as MiTAC Validated Coolant Supplier | Maintains proven cooling performance.

World's First Diamond Cooling®
G8825Z5 AMD 8-GPU AI Server

AKASH SYSTEMS



Throttle-Free
Performance

Up to 10°C
(50°F)
GPU Temperature Reduction

Up to 100%
Less Power to Cooling

Up to 22%
More FLOPs/Watt
at ~75°F Ambient
(Standard Data Center)

Up to 15%
Higher Token Throughput
at ~120°F Ambient
(High Ambient Data Center)

MiTAC Computing

AMD
together we advance_

Two-Phase DLC: How It Works & Measured Outcomes

ZUTACORE

A sealed, waterless loop that absorbs, transports, and condenses heat with extreme efficiency.

How It Works

A sealed, waterless loop with three key steps

1 Evaporate

Fluid boils on the hottest silicon.



2 Transport

Vapor carries the heat out (~0.3 L/min per kW).



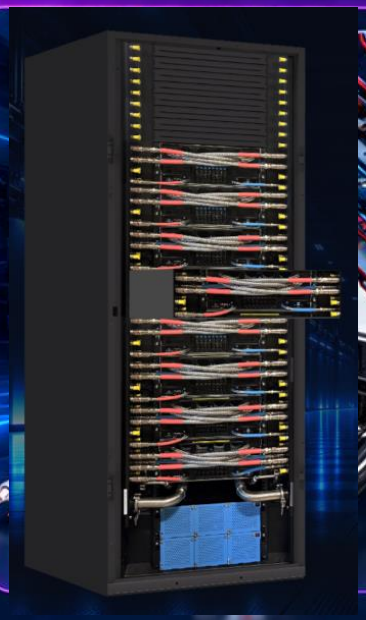
3 Condense

The CDU returns liquid to the cold plate.



Measured Outcomes

Proven performance for next-generation AI infrastructure





> 4,000W
Thermal Headroom

Engineered to cool next-generation AI processors exceeding 4,000 watts per chip.

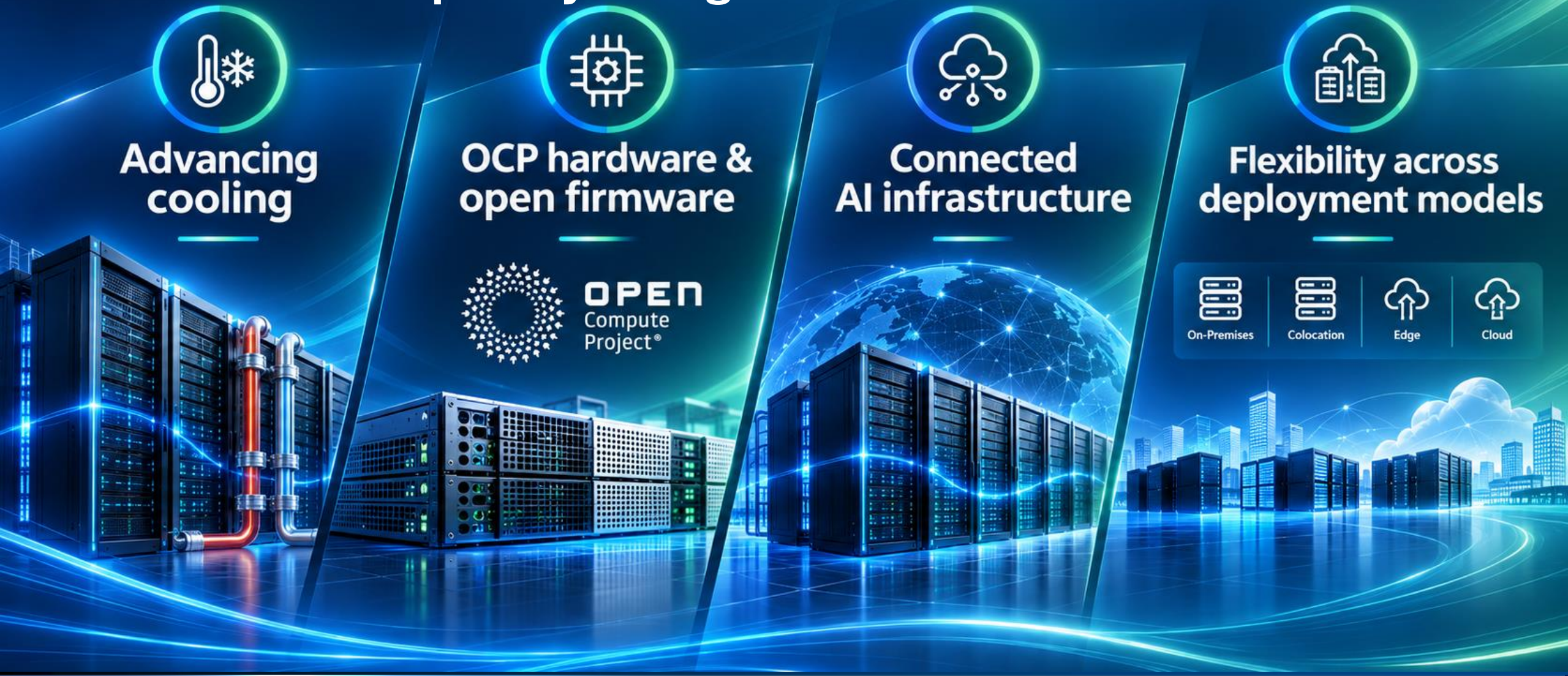
> 250 W/cm²
Heat Flux

Capable of absorbing an average surface heat flux exceeding 250 W/cm² reliably.

1.03
Annualized pPUE

Cuts cooling energy consumption by 74% compared to single-phase direct-to-chip cooling.

Scaling Open AI Infrastructure
Open by Design. Cooled to Scale.



MiTAC

MiTAC Computing

Booth

D22

