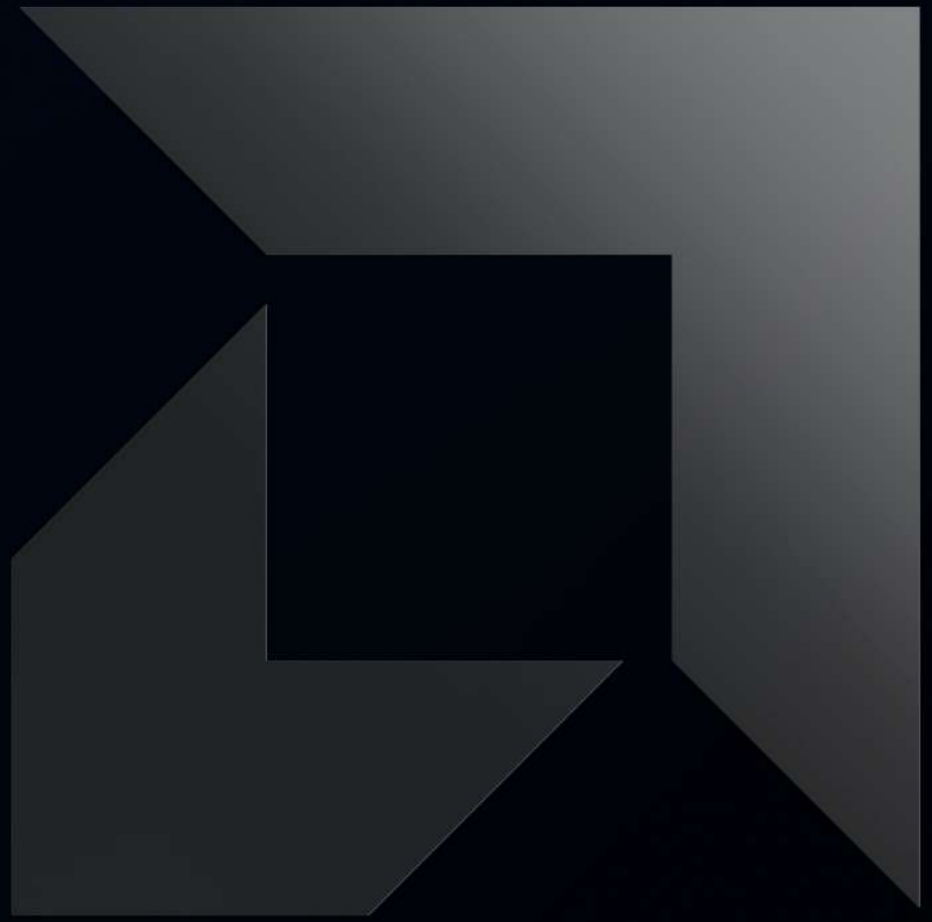


Co je nového u AMD v serverech: EPYC, Instinct a sítě

Michal Sztemon
AMD CZ/SK Marketing Consultant

AMD
together we advance_



AMD by the Numbers

AMD Market Cap **$748 Billions** (at 12 May 2026)

~$34,6B annual revenue

5X annual revenue growth over 10 years

18,800 patent assets worldwide[1]

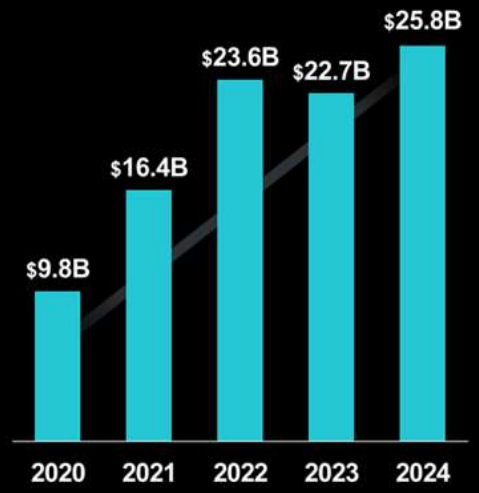
Spans hardware, software, AI, and related technologies

$6.5B annual R&D spend

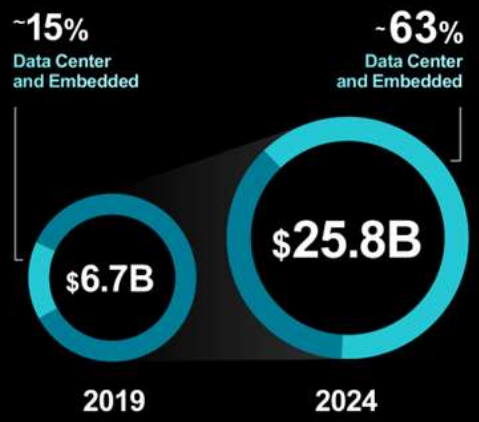
25% of revenue invested in R&D in 2024

Driving Growth from a Strong Financial Foundation

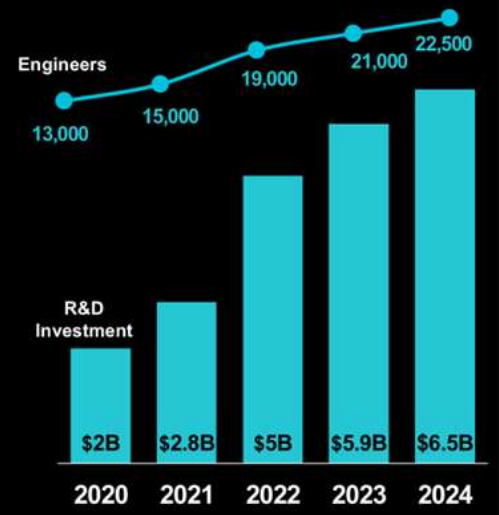
Accelerating Revenue Growth



Transforming Revenue Mix



Investing for Leadership



57 Years

Founded May 1, 1969
HQ Santa Clara, CA

28,000+ employees worldwide

~15,500+ in North America

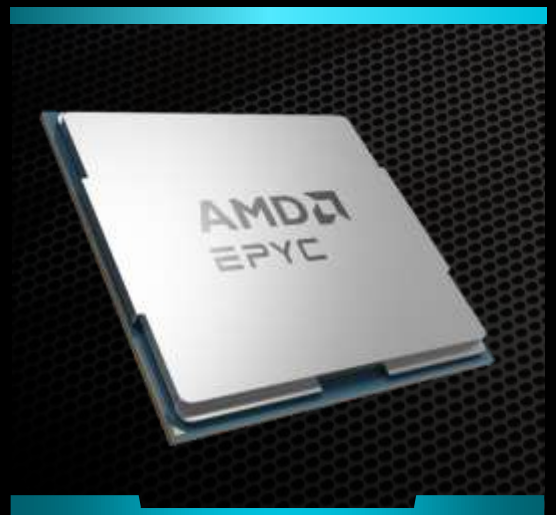
#167 Fortune 500

Among 8 companies with >30% annualized 10-year shareholder return

AMD 10-K annual report filing, February 5, 2025. Includes patents and pending patent applications

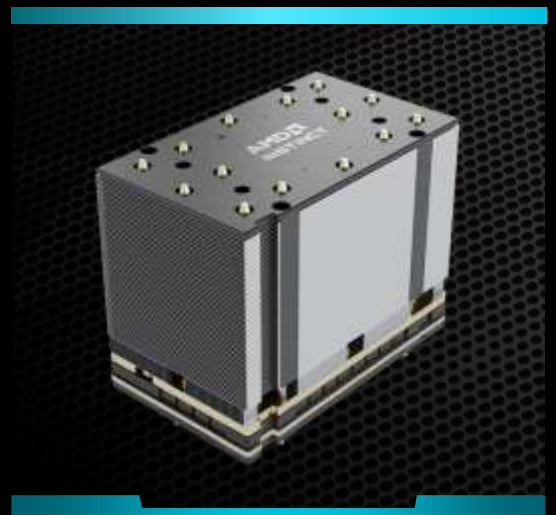
AMD
together we advance_

Leadership Data Center Solutions



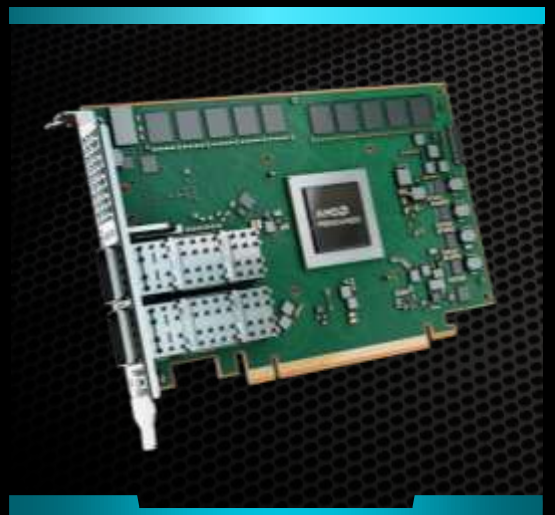
Server CPU Family

AMD EPYC



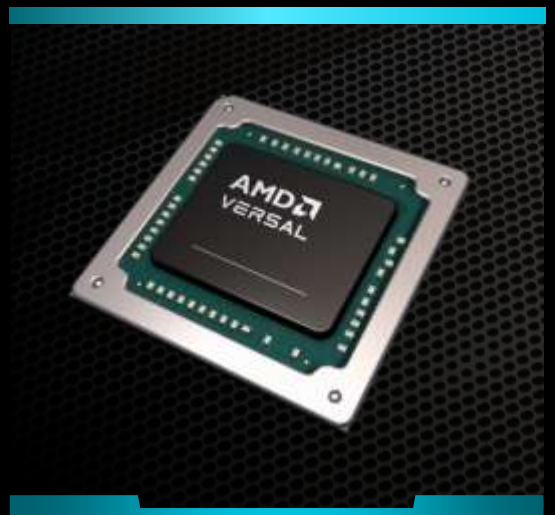
AI and HPC Accelerator

AMD INSTINCT



Networking and DPU

AMD ALVEO AMD PENSANDO



FPGA and Adaptive SoC

AMD VERSAL AMD ALVEO

AMD delivers the broadest technology portfolio to the data center

AMD
together we advance_

The Cost of Doing Nothing – Delaying Server Refresh

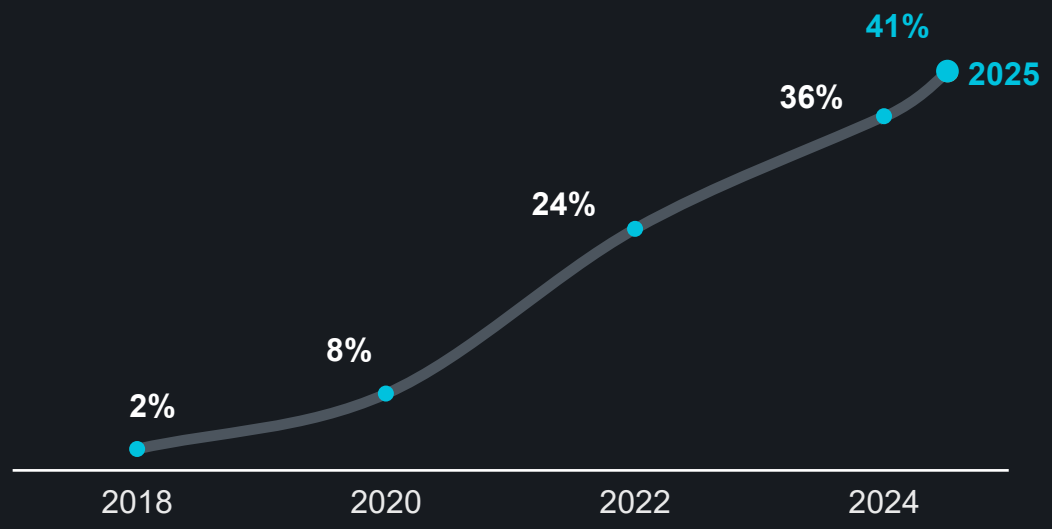


IT ability to support a competitive business is limited

AMD
together we advance_

AMD EPYC™ Server CPUs Grows to Over 40% Revenue Share

>18x Server CPU Revenue Share Growth



Industry Leaders Run on EPYC™

Cloud	aws, Microsoft, Google, ORACLE
Digital	NETFLIX, Uber, Meta, zoom
Enterprise	BEST BUY, IBM, Emirates NBD, NISSAN
OEM	DELL Technologies, HPE, Lenovo, SUPERMICRO, cisco

AMD Supported by Leading OEMs and ODMs, No Need to Switch Supplier

Source: Mercury Research Sell-In Revenue Shipment Estimates, Q4 2025. EPYC -055C

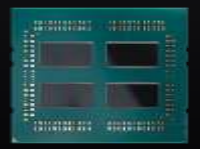
Use of third-party marks / logos/ products is for informational purposes only and no endorsement of or by AMD is intended or implied. GD-83

AMD
together we advance_

[Public]

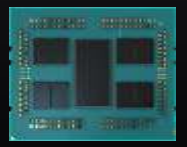
AMD EPYC™ CPU Reliable Execution Over 5 Generations

1st Gen EPYC™



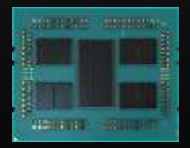
General Purpose
"Naples"

2nd Gen EPYC™

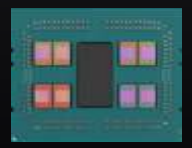


General Purpose
"Rome"

3rd Gen EPYC™

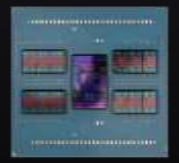


General Purpose
"Milan"



Technical
"Milan-X"

4th Gen EPYC™



General Purpose
"Genoa"



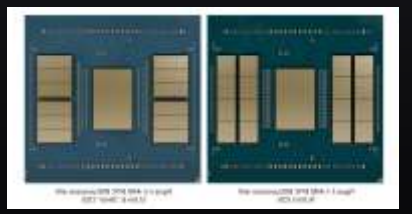
Cloud Native
"Bergamo"



Telco/Edge
"Siena"



5th Gen EPYC™



General Purpose
"Turin"

Cloud Native
"Turin-D"

Telco/Edge
"Sorano"

2017 → 2025

Five Generations of On-Time Execution, 6th and 7th currently in development

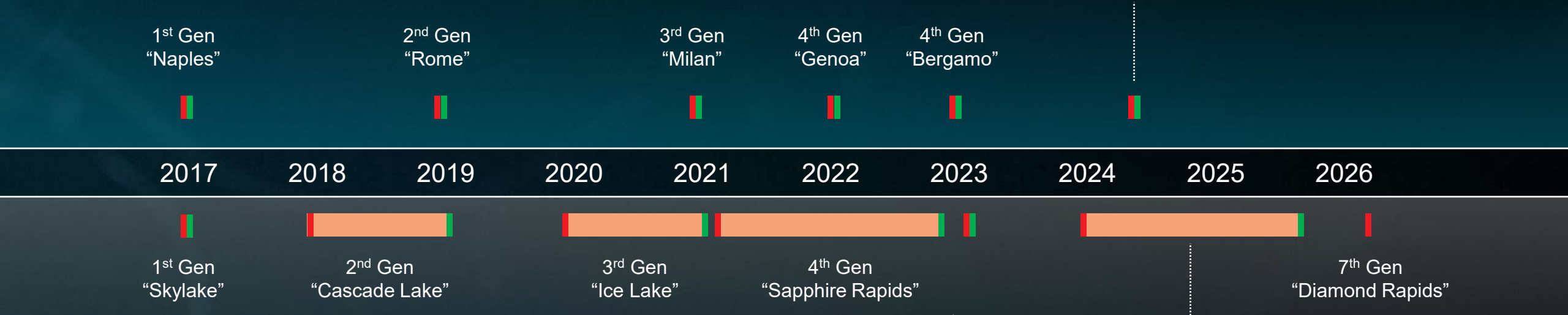
AMD
together we advance_

[Public]

Five Generations of Execution vs a Decade of Delay

AMD EPYC
Advancement Customers can Rely On

5th Gen AMD EPYC
Up to **11.3x** The Generational Performance



Intel Xeon
Consistent Record… of Delayed Products

"...the yields [of 18A] are not where we need them to be in order to drive the appropriate level of margins." – David Zinsner, CFO, Intel Q3'25 Earnings Call

Trickled GA Adoption
288C E-Core discontinued
Only 8.3x generational uplift

Expected GA

See Endnote 9xx5-069D. Launch dates based on AMD analysis. Generational uplift based on comparing the current generation server processor to first generation server processor's highest SPECrate®2017_int_base score during respective launch years.

Dated February 2026

[Public]

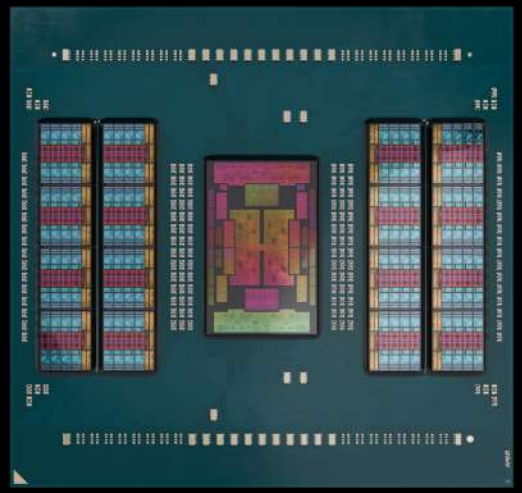
5th Gen AMD EPYC™ Generational Innovations

Compute

- "Zen5" up to **128 cores** / 256 threads
- "Zen5c" up to **192 cores** / 384 threads
- AVX-512 with **full 512b data path**
- New **500W** performance option
- Faster **5GHz** options
- **3/4nm** Zen cores

I/O & Platform

- 2P and 1P Configurations
- Up to 160 lanes of PCIe® Gen5
- **PCIe link encryption**
- SP5 Compatible with **"Genoa"**
- **CXL® 2.0[1]**



Memory

- 12 ch. DDR5 ECC **up to 6400[2] MT/s**
- Up to 2 DIMMs/channel capacity delivering up to **6TB/socket**
- **Dynamic Post Package Repair** (PPR) for x4 and x8 ECC RDIMMs

Security

- Hardware Root-of-Trust
- **Trusted I/O**
- **FIPS 140-3 in process**

1 - CXL Type 1&2 devices and PCIe link encryption support dependent upon ecosystem readiness
2 - See endnote 9xx5-083A
See endnotes 9xx5-048, 072, 083, GD-183A

AMD
together we advance_

AMD EPYC™ 9005 Series Processors

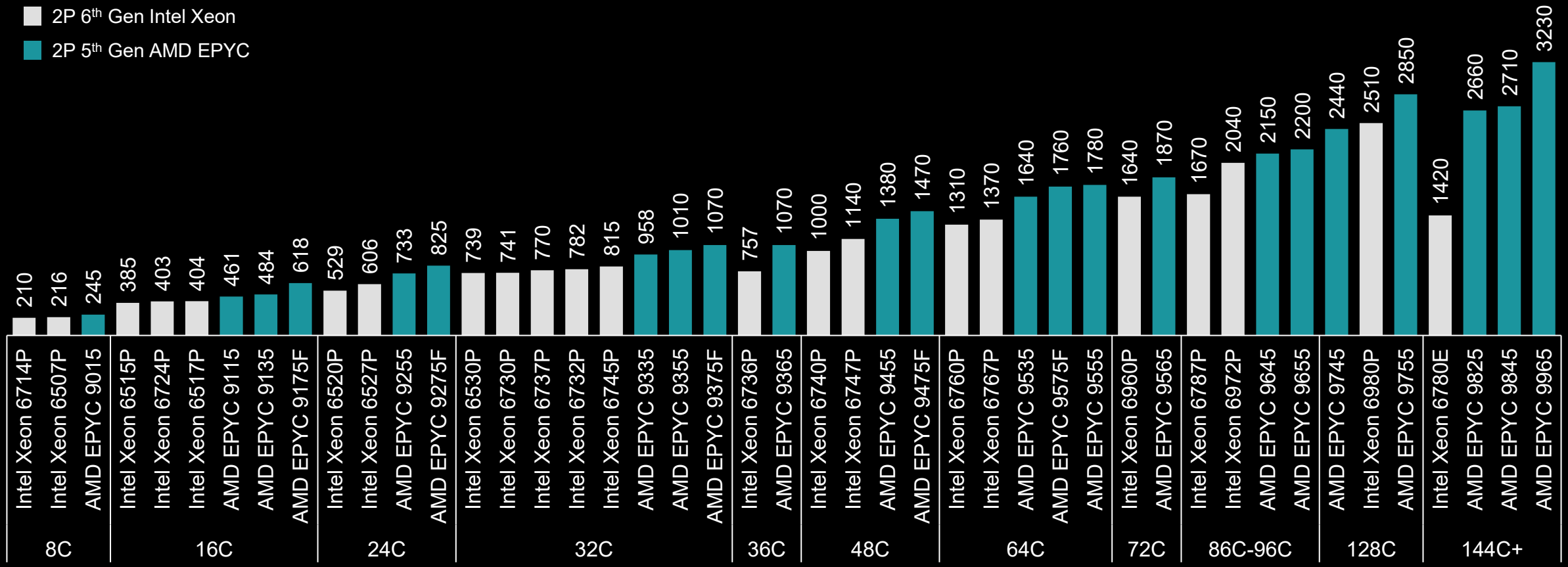
- Increased core density
- Energy efficient
- Broad OPN stack

Cores	AMD EPYC	CCD (Zen5/Zen5c)	Base/Boost* (up to GHz)	Default TDP (W)	L3 Cache (MB)	Price (1 KU, USD)
192 cores	9965	“Zen5c”	2.25 / 3.7	500W	384	$14,813
160 cores	9845	“Zen5c”	2.1 / 3.7	390W	320	$13,564
144 cores	9825	“Zen5c”	2.2 / 3.7	390W	384	$13,006
128 cores	9755 9745	“Zen5” “Zen5c”	2.7 / 4.1 2.4 / 3.7	500W 400W	512 256	$12,984 $12,141
96 cores	9655 9655P 9645	“Zen5” “Zen5” “Zen5c”	2.6 / 4.5 2.6 / 4.5 2.3 / 3.7	400W 400W 320W	384 384 384	$11,852 $10,811 $11,048
72 cores	9565	“Zen5”	3.15 / 4.3	400W	384	$10,486
64 cores	9575F 9555 9555P 9535	“Zen5” “Zen5” “Zen5” “Zen5”	3.3 / 5.0 3.2 / 4.4 3.2 / 4.4 2.4 / 4.3	400W 360W 360W 300W	256 256 256 256	$11,791 $9,826 $7,983 $8,992
48 cores	9475F 9455 9455P	“Zen5” “Zen5” “Zen5”	3.65 / 4.8 3.15 / 4.4 3.15 / 4.4	400W 300W 300W	256 192 192	$7,592 $5,412 $4,819
36 cores	9365	“Zen5”	3.4 / 4.3	300W	256	$4,341
32 cores	9375F 9355 9355P 9335	“Zen5” “Zen5” “Zen5” “Zen5”	3.8 / 4.8 3.55 / 4.4 3.55 / 4.4 3.0 / 4.4	320W 280W 280W 210W	256 256 256 256	$5,306 $3,694 $2,998 $3,178
24 cores	9275F 9255	“Zen5” “Zen5”	4.1 / 4.8 3.25 / 4.3	320W 200W	256 128	$3,439 $2,495
16 cores	9175F 9135 9115	“Zen5” “Zen5” “Zen5”	4.2 / 5.0 3.65 / 4.3 2.6 / 4.1	320W 200W 125W	512 64 64	$4,256 $1,214 $726
8 cores	9015	“Zen5”	3.6 / 4.1	125W	64	$527

AMD
together we advance_

Have the Confidence to Win at Every Level

AMD EPYC 9005 Series Processors Lead at Every Core Count



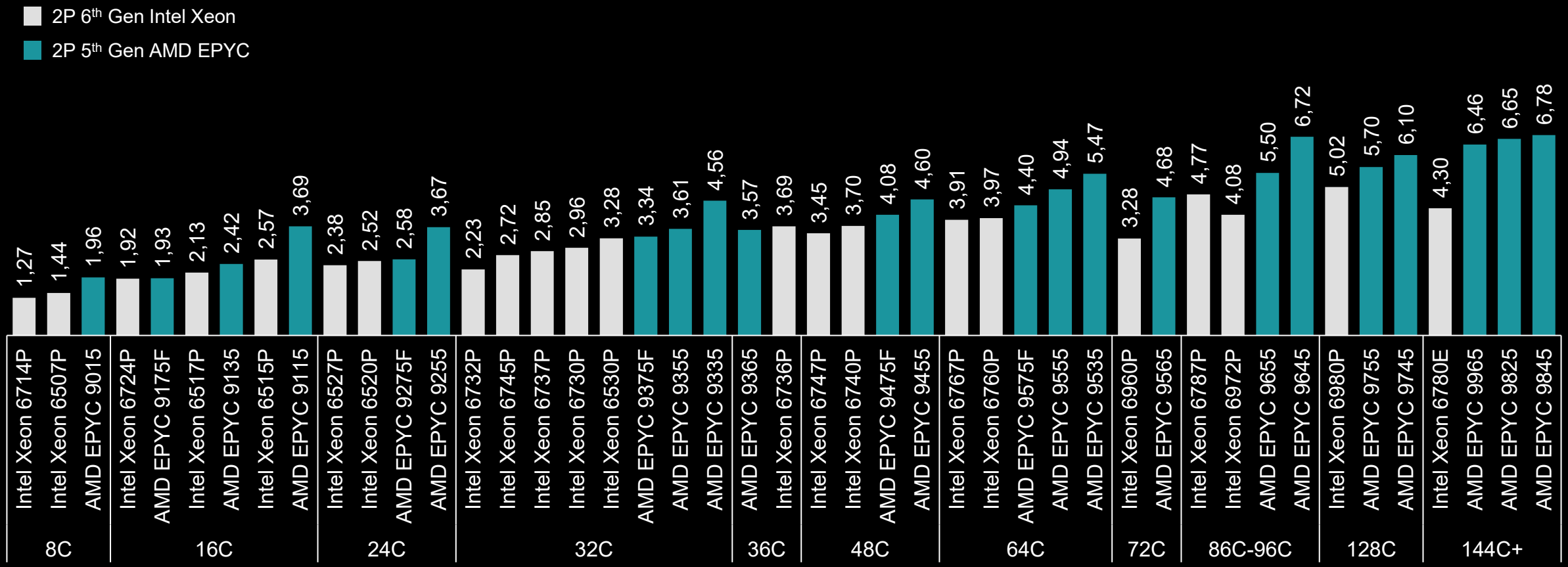
SPECrate®2017_int_base per core category for published dual socket servers

As of 12/11/2025. See Endnote 9xx5-173A. SPECrate®2017_int_base scores from spec.org. SPEC®, SPEC CPU®, and SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

AMD
together we advance_

AMD EPYC 9005 Efficiency Leadership

SPECrate®2017_int_base / CPU W for published dual socket servers



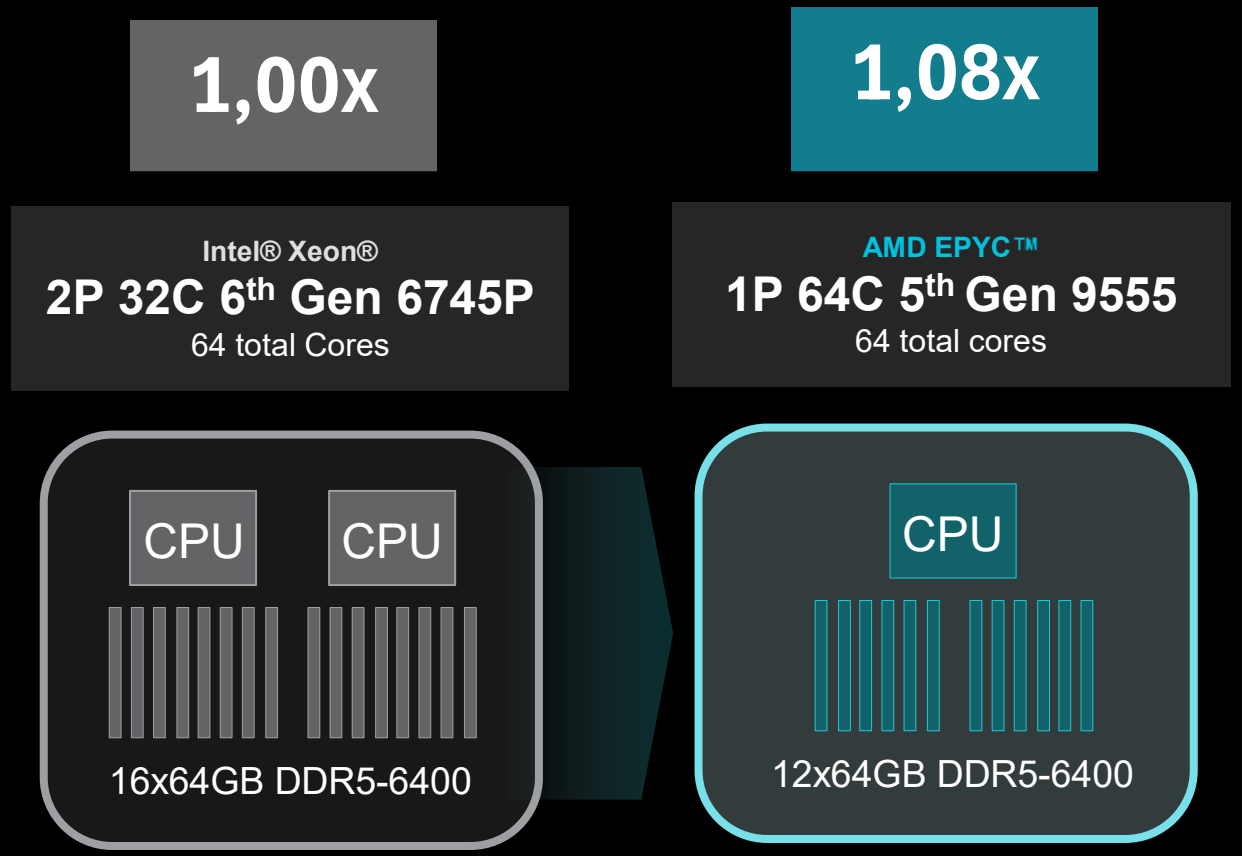
As of 12/11/2025. See Endnote 9xx5-173A. SPECrate®2017_int_base scores from spec.org. SPEC®, SPEC CPU®, and SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

AMD
together we advance_

Cut Costs, Keep Performance

Outperform the newest 2P Intel® Xeon® servers with higher power efficiency and lower cost

SPECrate®2017_int_base



1.8X
The Performance / CPU W

~25%
Lower Memory Cost per Server

~24%
Lower CPU Cost per Server

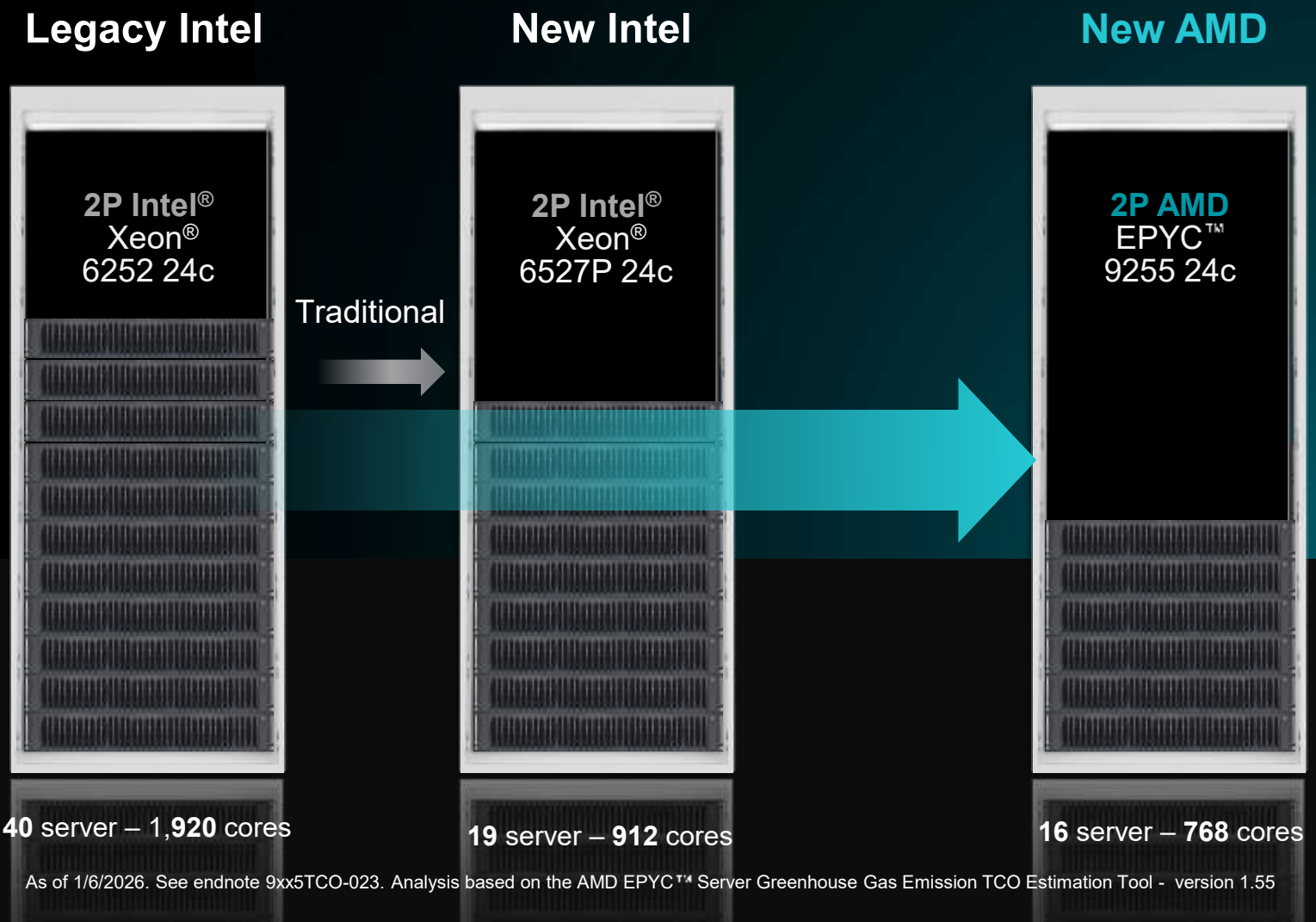
~50%
Lower Per-Socket Licensing Cost

See endnote 9xx5-277

AMD
together we advance_

Example: Upgrade to AMD EPYC™ Server CPUs

Dual CPU Server Example – 24 modern cores to 24 modern cores



Skip Intel® Xeon®

Choose AMD EPYC™ server CPU

Up to 16% Less Space

Up to 30% Less Power

Up to 21% More on TCO

AMD EPYC™ 9255 vs Intel Xeon 6527P

1.1x the performance and 22% less capex compared to Intel® Xeon® 6527P

11,480 SPECrate® 2017_int_base performance

As of 1/6/2026. See endnote 9xx5TCO-023. Analysis based on the AMD EPYC™ Server Greenhouse Gas Emission TCO Estimation Tool - version 1.55

AMD
together we advance_

5th Gen AMD EPYC™ 9575F CPU with 8x Nvidia B200

Up to 13%
Faster Overall Time to First Token
Relative Performance (Higher is Better)



Up to 6%
Faster Overall Time Per Output Token
Relative Performance (Higher is Better)



Up to 6%
More Overall Inference Throughput
Relative Performance (Higher is Better)



Inference performance measurement based on geomean across 8 models

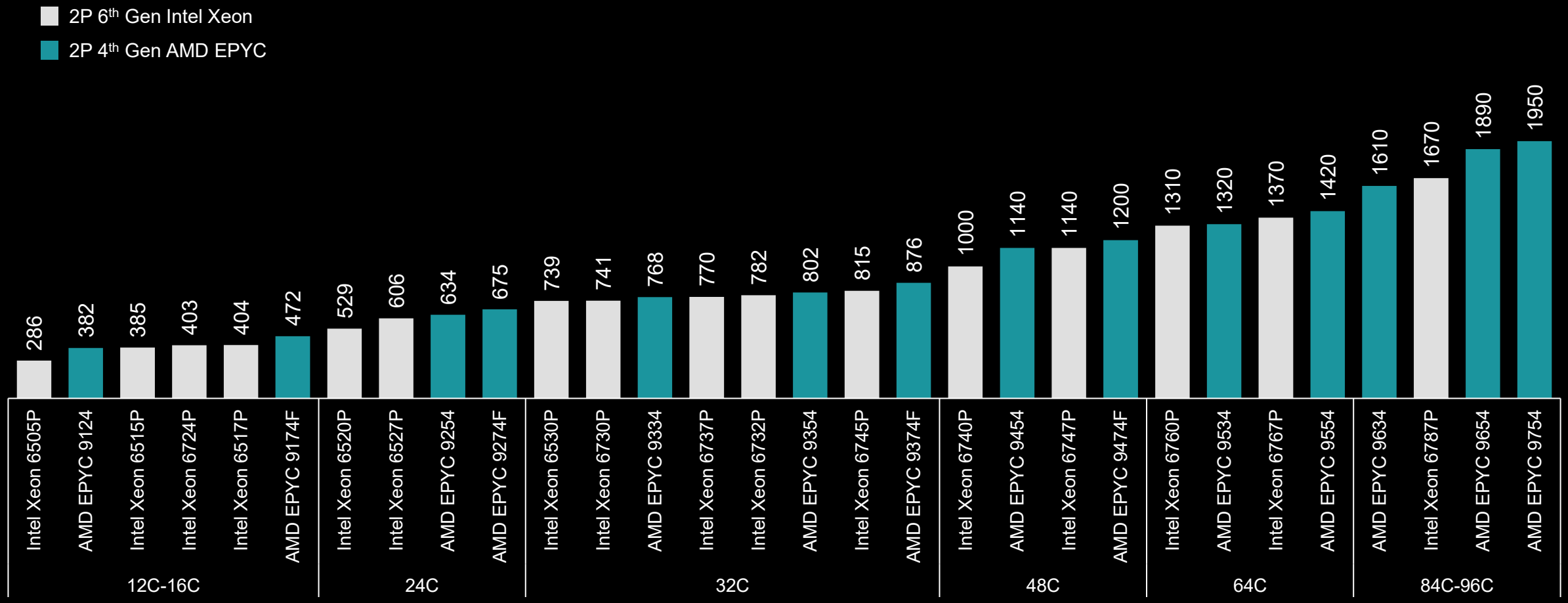
Enabling Maximum GPU System Performance as an AI Host Processor for AI Workloads

Test Configurations: 2P Intel® Xeon® 6960P (SYS-A22GA-NBRT) | 24×64 GB DDR5-6400 | NVMe Storage | Ubuntu 24.04 (Linux 6.8.0-85-generic) | NVIDIA Driver 580.95.05 (CUDA 13.0, NVML 580.95) | UCX v1.18.0 (built with rdmacm, verbs, mlx5, CMA; logging/debug disabled). 2P AMD EPYC 9575F (AS-A126GS-TNBR) | 24×64 GB DDR5-6400 | Samsung 3.84 TB NVMe | Ubuntu 24.04 (Linux 6.8.0-85-generic) | NVIDIA Driver 580.82.07 (CUDA 13.0, NVML 580.82) | Kernel amd_pstate=active, iommu=pt, mitigations=off. See Endnotes 9xx5-258, 9xx5-259, 9xx5-260

AMD
together we advance_

AMD EPYC™ 9004 CPU Performance Across Similar Cores

SPECrate®2017_int_base per core category for published dual socket servers



As of 12/1/2025. See SP5-317. SPECrate®2017_int_base scores from spec.org. SPEC®, SPEC CPU®, and SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

AMD
together we advance_

AMD Multi-year Investment Building Confidential Computing Features

CPU	Transparent Secure Memory Encryption	Secure Encrypted Virtualization			
		SEV	Encrypted State	Secure Nested Paging	Trusted I/O
AMD EPYC™ 7002	✓	✓	✓		
AMD EPYC™ 7003	✓	✓	✓	✓	
AMD EPYC™ 8004 & 9004	✓	✓	✓	✓	
AMD EPYC™ 9005	✓	✓	✓	✓	✓

AMD
together we advance_

OPEN and ROBUST Ecosystem Support

AMD INFINITY GUARD

TSME

Transparent Secure Memory Encryption

Encrypt System Memory with One Key

- System BIOS Setting
- OS-enabled SME on Linux® 4.14

SEV

Secure Encrypted Virtualization

Encrypt Each VM with Unique Keys

- Oracle Cloud Infrastructure
- Linux® 4.15 (guest)
- Linux® 4.16 (host)
- Linux® Distros
- Google Cloud
- Google Kubernetes Engine
- Open Stack (RHOSP 16.1)
- Enarx
- Anjuna
- Cysec
- Decentriq
- Kata Containers
- KubeVirt
- IBM

SEV-ES

SEV – Encrypted State

VM Integrity with Protected CPU Registers

- VMware ESXi ™ 7.0 U1
  - Photon OS 3.0
  - Tanzu
- Linux® 5.10 (guest)
- Linux® 5.11 (host)
- Linux® Distros

SEV-SNP

SEV – Secure Nested Paging

Hardware Protection Against Malicious Hypervisor

- Microsoft Azure Cloud
- Microsoft Azure Local
- Amazon Web Services
- Google Cloud Platforms
- Linux 5.19 (guest)
- Linux 6.11 (host)
- Red Hat – RHEL 9.5 (Tech preview)
- SUSE – Linux® Enterprise Server 15 SP6 – separate patch
- Canonical – Ubuntu® 24.04.2

Software enables the system to detect, protect, and comply!

AMD Infinity Guard features vary by EPYC ™ Processor generations and/or series. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at https://www.amd.com/en/technologies/infinity-guard . GD-183A

AMD
together we advance_

Modernize with AMD EPYC™ Server CPU

Cost Effective & Power Efficient



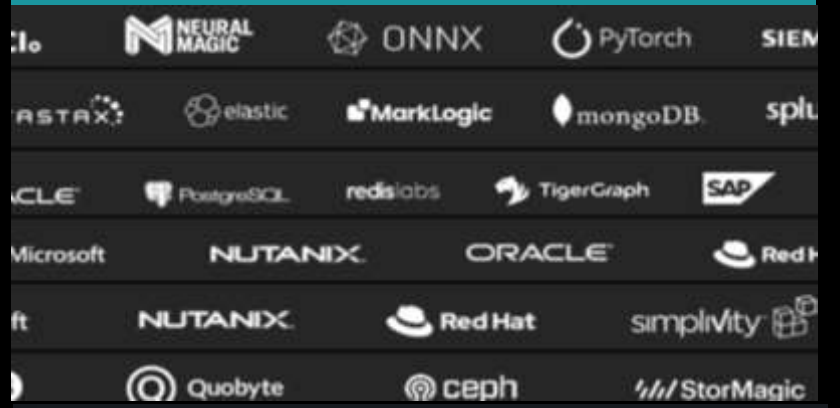
Gain compelling power, space and TCO advantages in server consolidation

Trusted & Reliable



Choose the server CPU relied on by many of the world's largest and most sensitive computing systems

Modernize with Ease



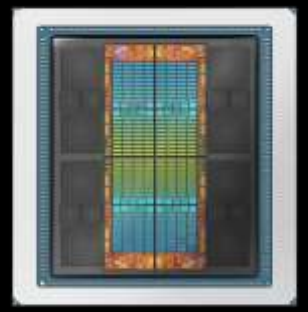
x86 compatible, mature ecosystem and migration tools allows hassle-free refresh to AMD

Confidently run General IT Workloads on AMD EPYC Server CPU

AMD
together we advance_

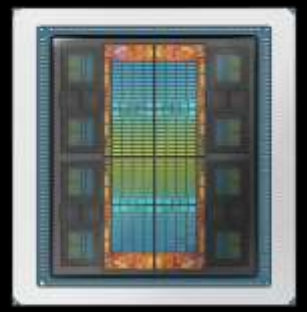
Delivering on Annual Roadmap Commitment

AMD Instinct™ MI300A/X



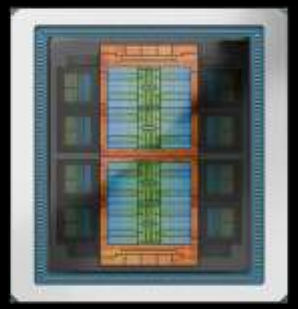
2023

AMD Instinct™ MI325X



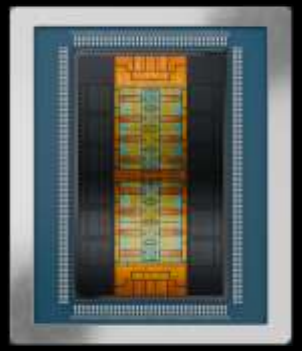
2024

AMD Instinct™ MI350 Series



2025

AMD Instinct™ MI400 Series



2026

AMD Instinct™ MI500 Series



2027

Roadmap subject to change

AMD
together we advance_

AMD INSTINCT

Growing Industry Adoption

8 of 10 Top AI Companies Use AMD Instinct™ GPUs

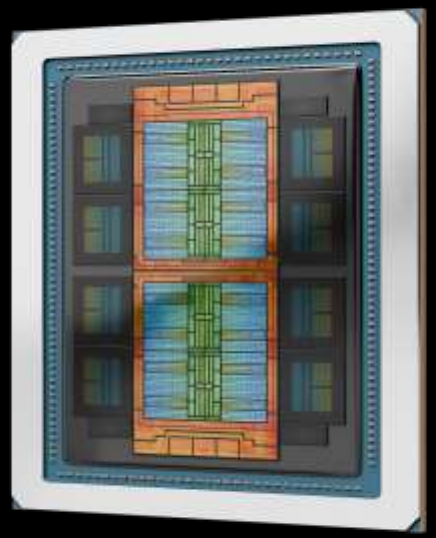


AMD
together we advance_

[AMD Confidential - AMD Attorney Privileged Information]

AMD Instinct™ MI350 Series GPUs

AMD Instinct™ MI350X GPU



AMD Instinct™ MI355X GPU

AMD CDNA 4

3nm	185B	MXFP4 & MXFP6 Support	288GB HBM3E
Process Node	Number of Transistors	New Gen AI Datatypes	Leadership capacity

See endnote MI350-023

[Public]

Industry's Highest Performing GPUs for AI and HPC

AMD INSTINCT MI355X GPU VS NVIDIA B200/GB200

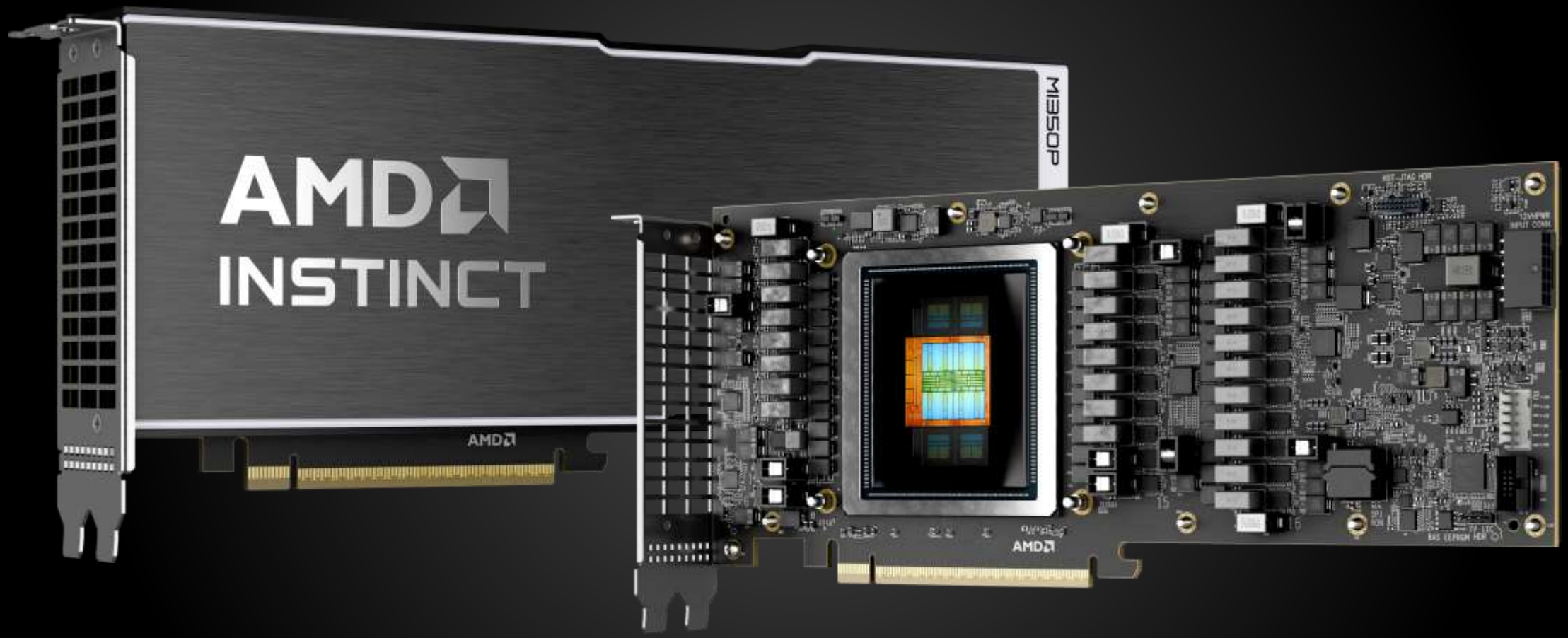
Instinct™ MI355X	vs. GB200	vs. B200
MEMORY CAPACITY	~1.6x	1.6x
MEMORY BANDWIDTH	1x	1x
PEAK FP64*/PEAK 32	~2x	~2x
PEAK FP16	1x	1.1x
PEAK FP8	1x	1.1x
PEAK FP6	2x	2.2x
PEAK FP4	1x	1.1x

Equal GPU Count for Architecture-Level Comparison with GB200

*All Matrix Core numbers except FP64 with sparsity

See endnote MI350-006, MI350-009

AMD
together we advance_



AMD
INSTINCT
MI350P

Enterprise AI – Ready Where You Are

[Public]

AMD AI Solution Portfolio – Air Cooled









	EPYC™ CPU	Radeon AI PRO™	AMD Instinct™ PCIe®	AMD Instinct™ UBB8
Positioning	Mixed Workload General purpose + AI Batch Inference	Mixed Workload Edge AI + Graphics	Low-latency, Memory-intensive Compute	High- throughput AI at Scale for Inference + Training
Use cases	Inference: Classical ML, RecSys, SLM/RAG, NLP	SLM Inference / RAG / MCP / Txt to Image / Video AI	SLM / MLM/ LLM Inference / RAG	Large-scale LLM Inference & Training
Model Size* (Max)	~ 30B per CPU	~ 40-50B per GPU	~ 200-250B per GPU	~up to 500B per GPU
AMD offering	AMD EPYC 5th Gen	R9700S | R9600D	MI350P[1]	MI350X | MI355X
GPUs/node	0	Up to 8GPUs per node	Up to 8 GPUs per node	8 GPU per node
GPU-to-GPU Interconnect	N/A	N/A	N/A	Fully-connected 8-GPU Infinity Fabric
Memory Capacity with Peak Bandwidth	614 GB/s per CPU 6400 MT/s	32G GDDR6 640 GB/s	144GB HBM3e 4 TB/s	2.3TB HBM3e 8TB/s per GPU

*Model size is estimated based on FP4/MXFP4. Estimated by AI

AMD
together we advance_

[Public]

AMD Instinct™ MI350P PCIe® Card – Platform Offering

Target Supported Platforms		
DELL	 PowerEdge XE7745	 PowerEdge R7725
CISCO	C845a M8 X series 580p UC245 M8	
HPE		 ProLiant DL385, 385a Gen 11 ProLiant DL345 Gen12
SUPERMICRO	AS -5126GS-TNRT AS -5126GS-TNRT2	AS -2026HS-TN AS -2116CS-TN
Lenovo	 ThinkSystem SR675/I v3	

AMD
together we advance_

[Public]

AMD ROCm

Open, modular software stack

AI Models and Algorithms

PyTorch ONNX TensorFlow

Support all major frameworks and models

Libraries

Compilers and Tools

Runtime

Expanded Gen AI optimizations

New algorithms

New libraries

Expanding platform support

AMD GPUs



AMD INSTINCT



AMD RADEON

Extended developer support

Use or mention of third-party marks, logos, products, services, or solutions herein is for informational purposes only and no endorsement by AMD is intended or implied. GD-83

AMD
together we advance_

AMD ROCm

Deepening Ecosystem Collaboration

Pytorch	Triton v3.3	Hugging Face 1.8 million models	vLLM v1, SGL, llm-d	LLaMA 4, Gemma 3, deepseek, QwQ-32B, Command R+, Grok, Mistral AI	ONNX, deepspeed, TensorFlow, OpenXLA, MLIR
Day 0 support daily performance CI	Performance focus	Nightly CI/CD, finetuning support	Serving leadership Distributed inference	Support for SOTA models	Expanding open-source footprint

AMD ROCm™ Software: Easing Transition of AI Workloads to AMD GPUs



* Use or mention of third-party marks, logos, products, services, or solutions, herein is for informational purposes only and no endorsement by AMD is intended or implied. GD-83

[AMD Confidential - Distribution Under NDA]

What's new in ROCm 7?



Performance Improvements

- Optimized Math Libraries
- Gen AI + ML Speedups
- Communications & Multi GPU
- Compiler Optimizations

Faster Training & Throughput



New Feature Additions

- Expanded Data Types
- Broader AI Framework support
- Domain-Specific Libraries
- HPC & Distributed Computing Features

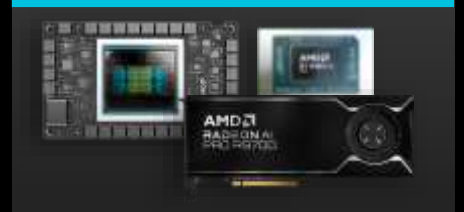
New Precisions & AI Capabilities



Developer Tool Improvements

- Compiler Toolchain Upgrades
- Debugging and Profiling Tools
- Profiling Infrastructure
- System Management

Deeper Profiling, Faster Compiles



Compatibility & Platform Support

- New GPU Architectures *MI350X, MI355X*
- Operating Systems
- Virtualization and Cloud
- Driver & Packaging Changes

Wider GPU, OS, Virtualization Support



API Programming Model

- HIP API Updates for Portability
- New HIP Runtime Capabilities
- Deprecations & Ecosystem Changes

More Portable, Richer HIP Features

AMD
together we advance_

[AMD Confidential - Distribution Under NDA]

AMD Instinct™ MI350 Series Rack Infrastructure

Fully Built on Open Standards





x86 CPU
5th Gen EPYC

+



Instinct GPU
MI350 series

+



UEC Scale-Out NIC
AMD Pollara NIC

OPEN Compute Project

OCP Design
Industry Standard Design

Ultra Ethernet Consortium

UEC Supported
Open Network Solution

AMD
together we advance_

AMD | Advancing AI Infrastructure on an Annual Cadence

2025

AMD EPYC
"TURIN"

AMD Instinct
MI350 SERIES

AMD Pensando
POLLARA 400



2026

AMD EPYC
"VENICE"

AMD Instinct
MI400 SERIES

AMD Pensando
"VULCANO"



"Helios"

2027

AMD EPYC
"VERANO"

AMD Instinct
MI500 SERIES

AMD Pensando
"VULCANO"



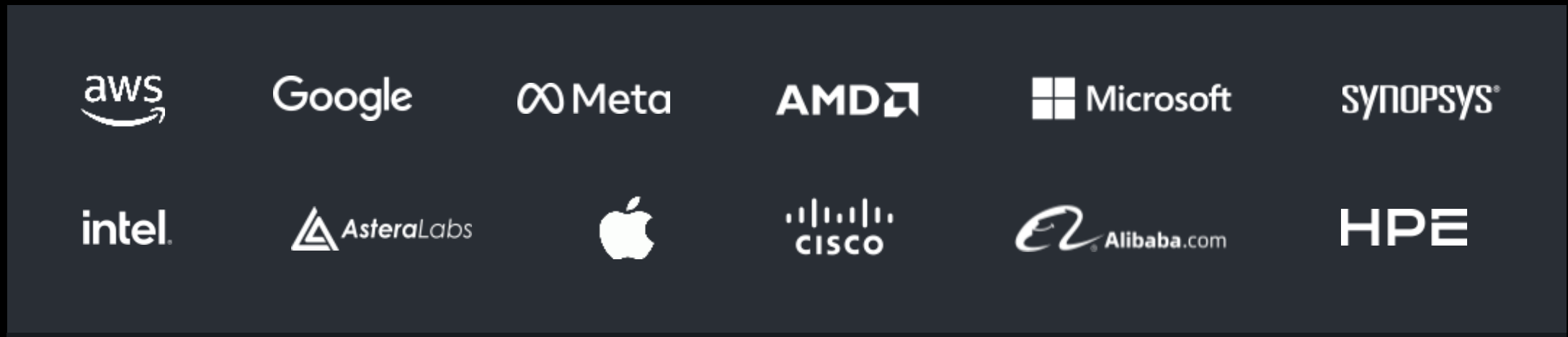
Next Gen AI Rack

Roadmap subject to change

Ultra Accelerator Link, the Truly Open Standard

ULTRA ACCELERATOR LINK™

	UALINK	NVLINK FUSION
LOW ROUND TRIP LATENCY	Yes	Yes
HIGH SPEED I/O	224 Gbps	224 Gbps
MAXIMUM SCALABILITY	1024 GPUs	576 GPUs
CPU USE	Any	Only Nvidia CPU
GPU USE	Any	Only to Nvidia GPU
MANAGEMENT SOFTWARE	Open	Nvidia Proprietary
SPECIFICATION	Fully Open	Closed



Promoters



Select Contributors

Ultra Ethernet Consortium

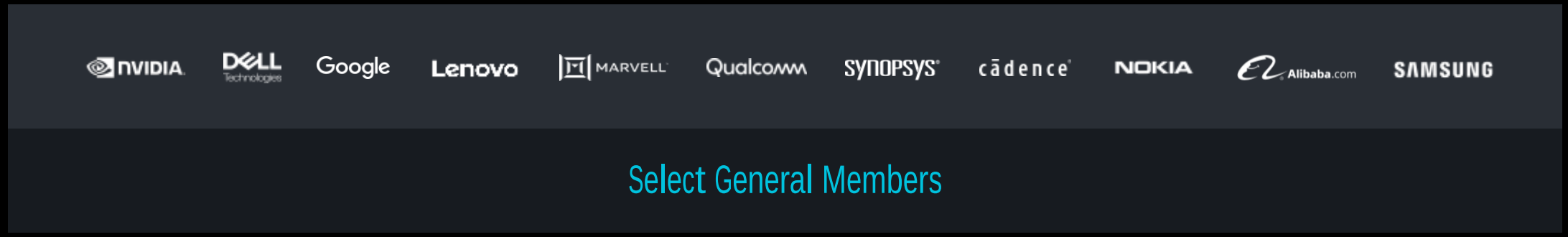
Open Architecture for Performant AI Scale-Out

Modern RDMA transport with automatic configuration & tuning

Optimized protocol for high fabric utilization & congestion control

Unparallel scalability to over a million GPUs





100+ Consortium Members

[Public]

AMD Pensando™ Pollara 400

Industry's first ultra ethernet consortium ready AI NIC



Ultra Ethernet Consortium

Programmable Hardware Pipeline	Up to 6x Performance Boost*	400 Gbps	Open Ecosystem UEC Ready RDMA Reduction in Job Completion Times High Availability

Based on comparison of RDMA vs. RoCEv2. * Reference: STrack: A Reliable Multipath Transport for AI/ML Clusters , July 2024

AMD
together we advance_

AMD Pensando™ “Vulcano”

Next Gen NIC for AI Clusters

3nm	800G	Up to 8x	UAL | PCIe®
Process Node	Network Throughput	Scale Out Bandwidth per GPU	Host Interface

Ultra Ethernet Consortium | Coming in 2026

[Public]

Comprehensive Open Ecosystem from CPU to Application

Software Solutions	AMD ROCm, AMD ZenDNN	Open-source software frameworks for developer freedom and performance optimization
Networking	Ultra Ethernet, UA Link | Ultra Accelerator Link	Open networking standards for long term infrastructure flexibility
DPU / FPGAs	AMD PENSANDO, AMD ALVEO	Network accelerators for high performance data processing and low-latency applications
Accelerators		AMD Instinct™ GPUs for exceptional AI workload acceleration
Servers		AMD EPYC™ Server Processors for AI, HPC, Enterprise, Edge, and SMB



Děkuji za pozornost!

Kontakty:
michal.sztemon@amd-czech.cz

AMD
together we advance_