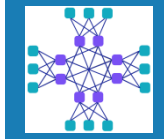
LYAMBDA



HPC - AI
Суперкомпьютеры
и искусственный интеллект

Новое поколение вычислительного оборудования для физики высоких энергий

Ноябрь 2024, для ОИЯИ

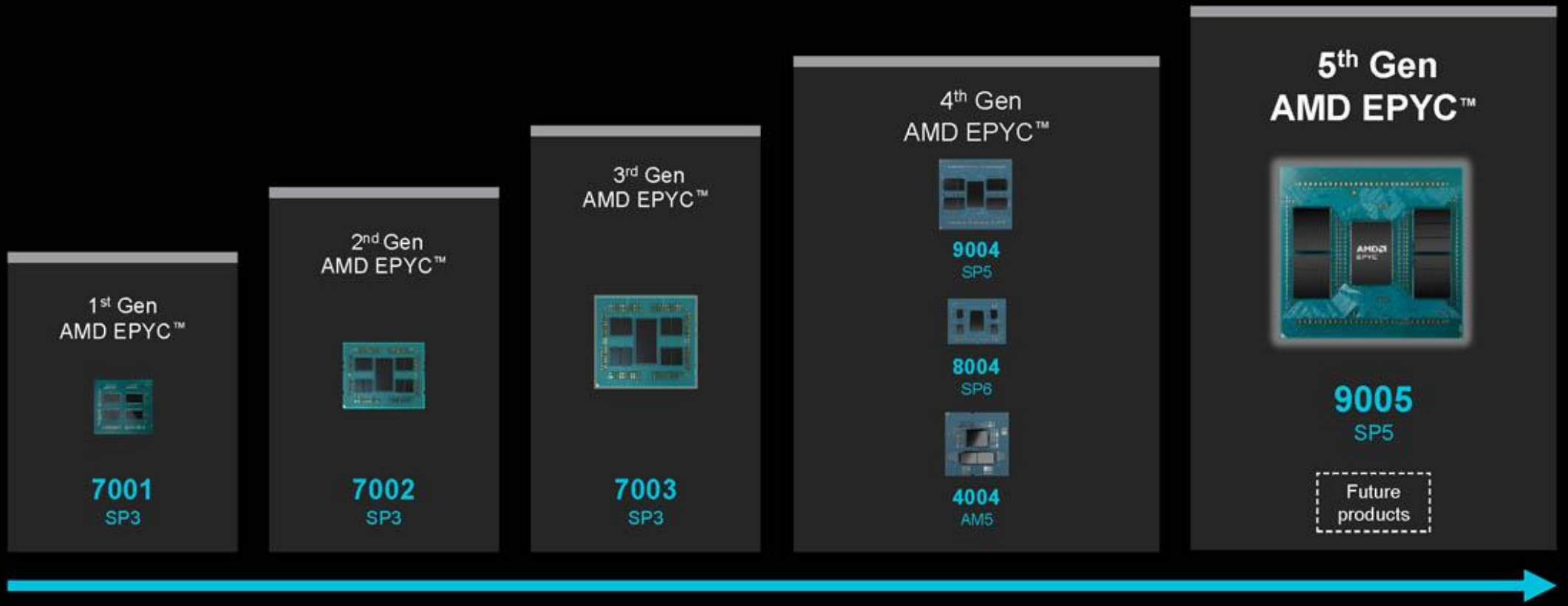

APC

Новое поколение процессоров

Процессоры AMD

AMD EPYC™ Processors

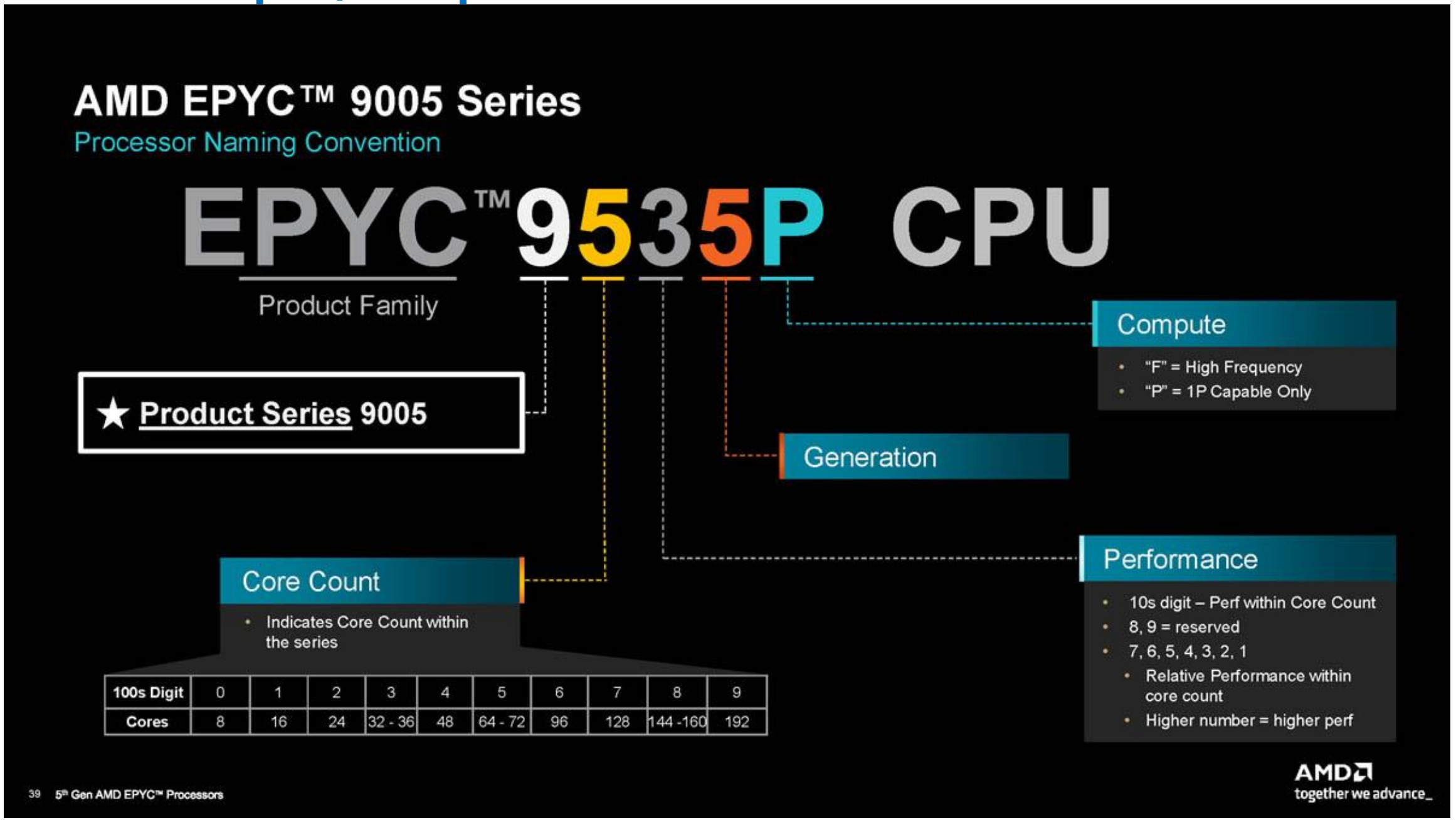
Five generations of on time technology innovation



All roadmaps are subject to change.

AMD
together we advance_

Процессоры AMD - система обозначений



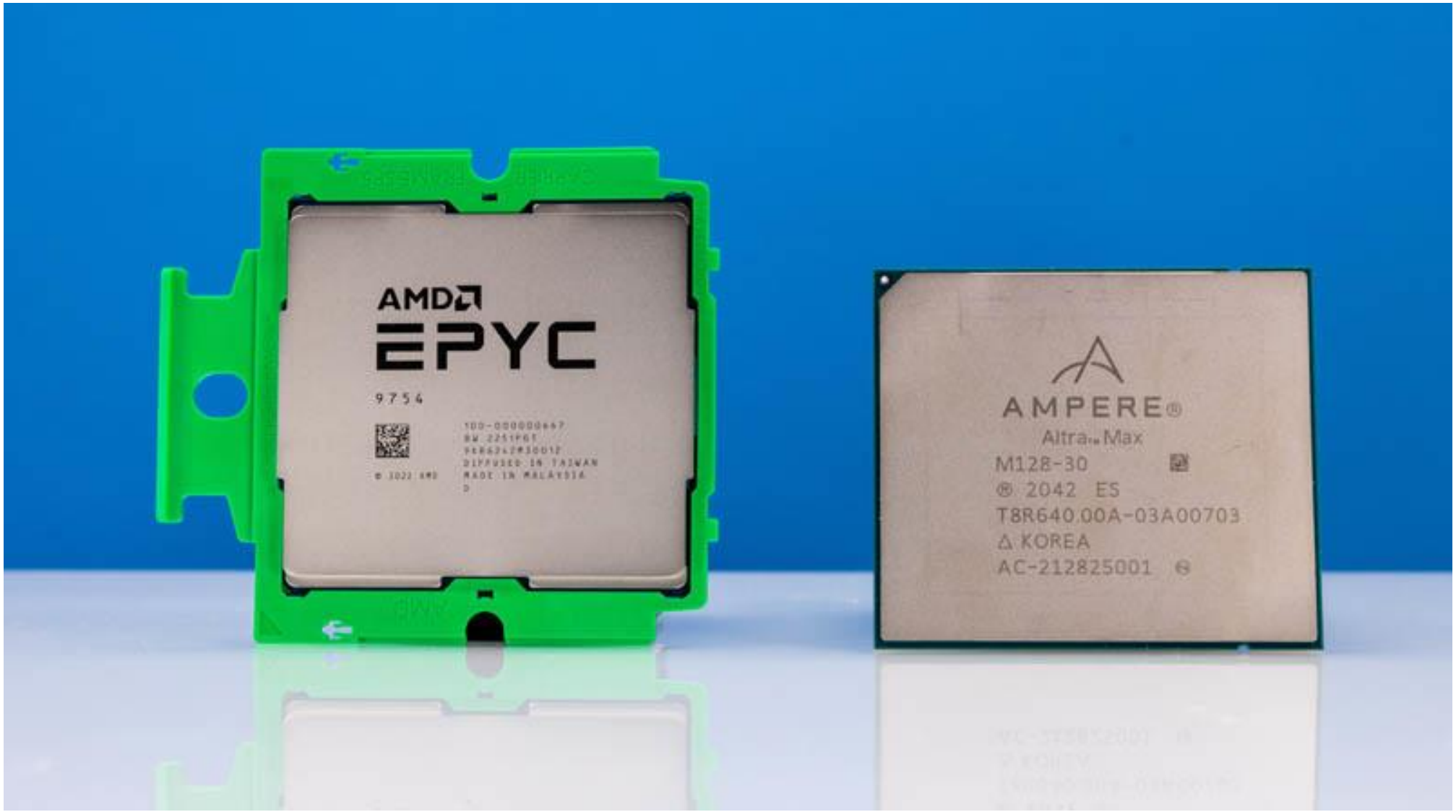
Процессоры AMD Genoa X
для HPC, ядра Zen4, «гигантский кэш» (~ 1GB)

Процессоры AMD Genoa X

4TH GEN AMD EPYC™ CPU WITH AMD 3D V-CACHE™ TECHNOLOGY PRODUCT STACK

MODEL	CORES	THREADS	DEFAULT TDP (W)	cTDP RANGE (W)	Fbase / Fboost*	L3 CACHE (MB)	DDR5 Channels	DDR5 CHANNELS
9684X	96	192	400	320 - 400	2.55 / 3.7	1,152	12	x128
9384X	32	64	320	320 – 400	3.1 / 3.9	768	12	x128
9184X	16	32	320	320 – 400	3.55 / 4.2	768	12	x128

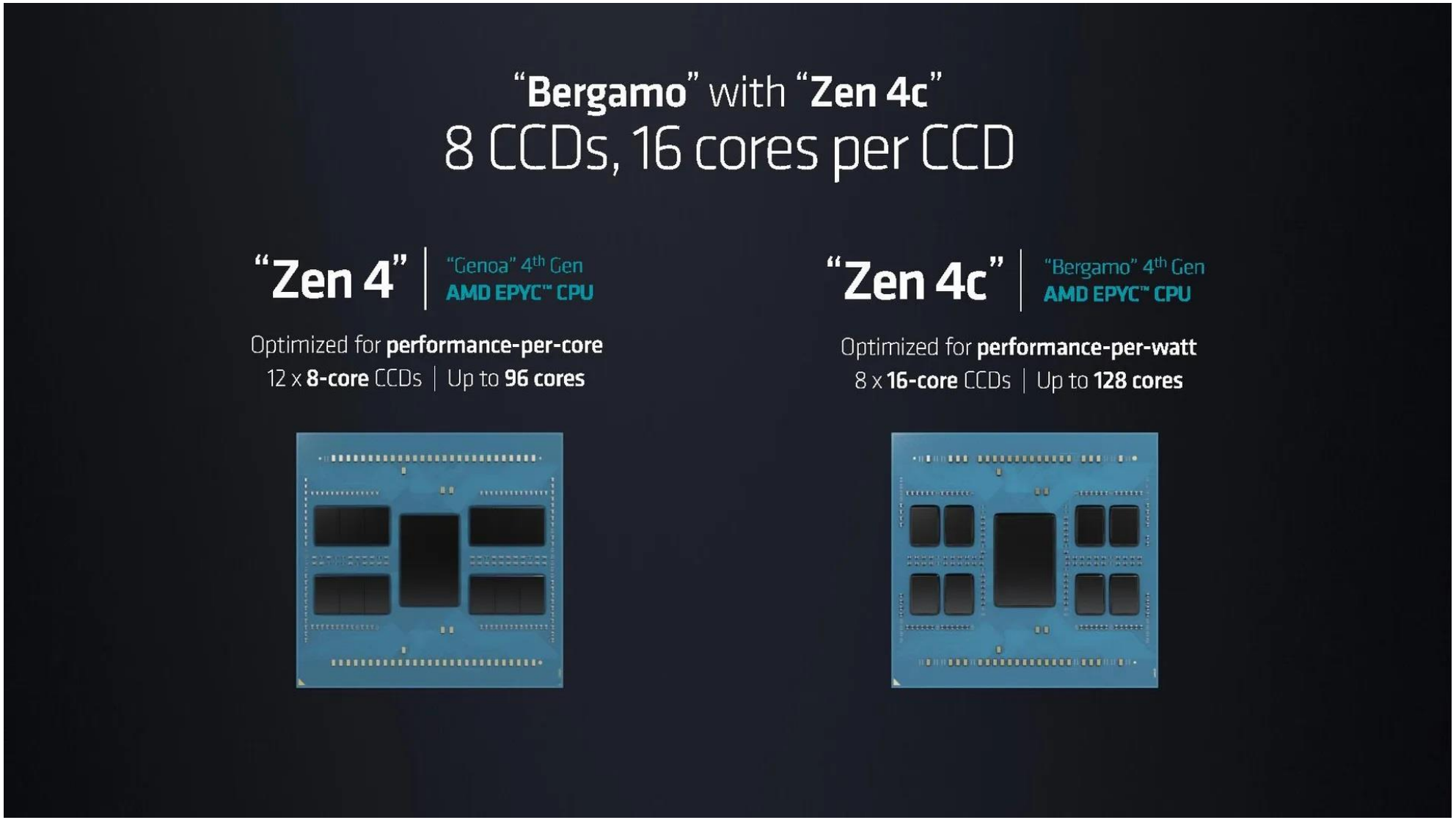
Процессоры AMD Bergamo (Zen4c)



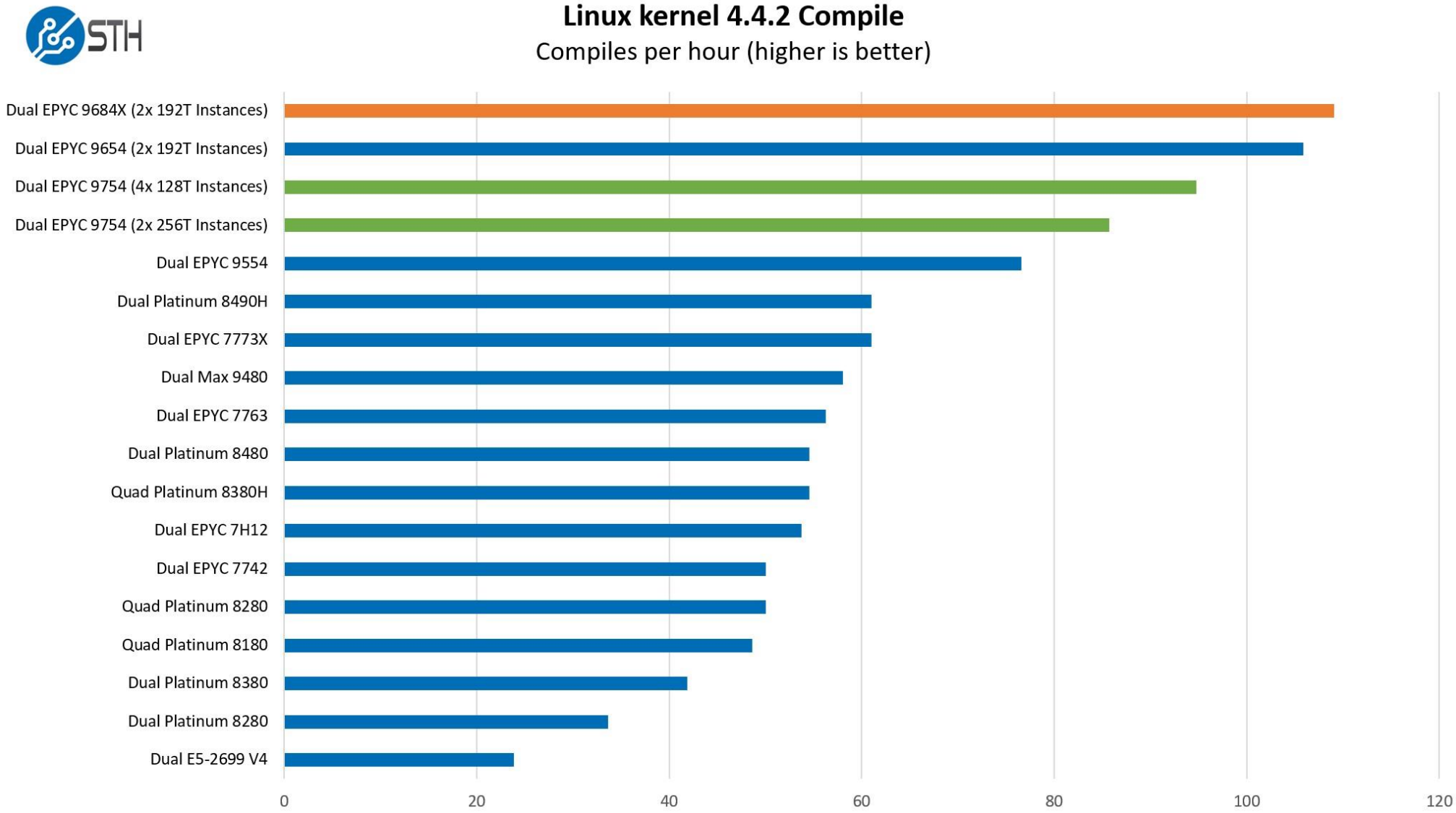
Процессоры AMD Bergamo (Zen4c)

	Cores / Max Threads	Base/Boost (GHz)	Default TDP	L3 Cache
9754	128 / 256	2.25 / 3.1	360W	256 MB
9754S	128 / 128	2.25 / 3.1	360W	256 MB
9734	112 / 224	2.2 / 3.0	320W	256 MB

Процессоры AMD Bergamo (Zen4c)

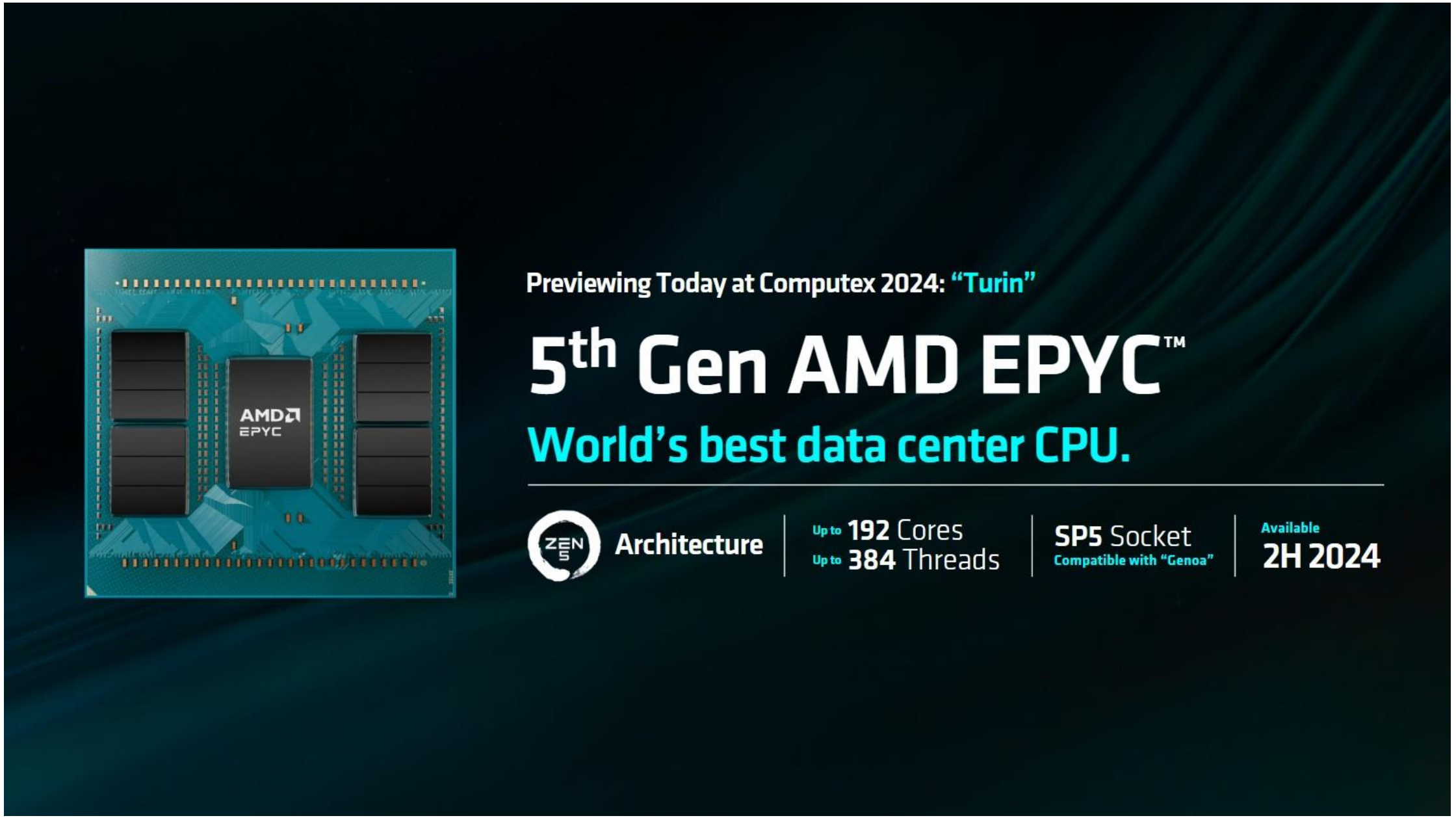


Сравнение производительностей процессоров AMD поколения Zen 4

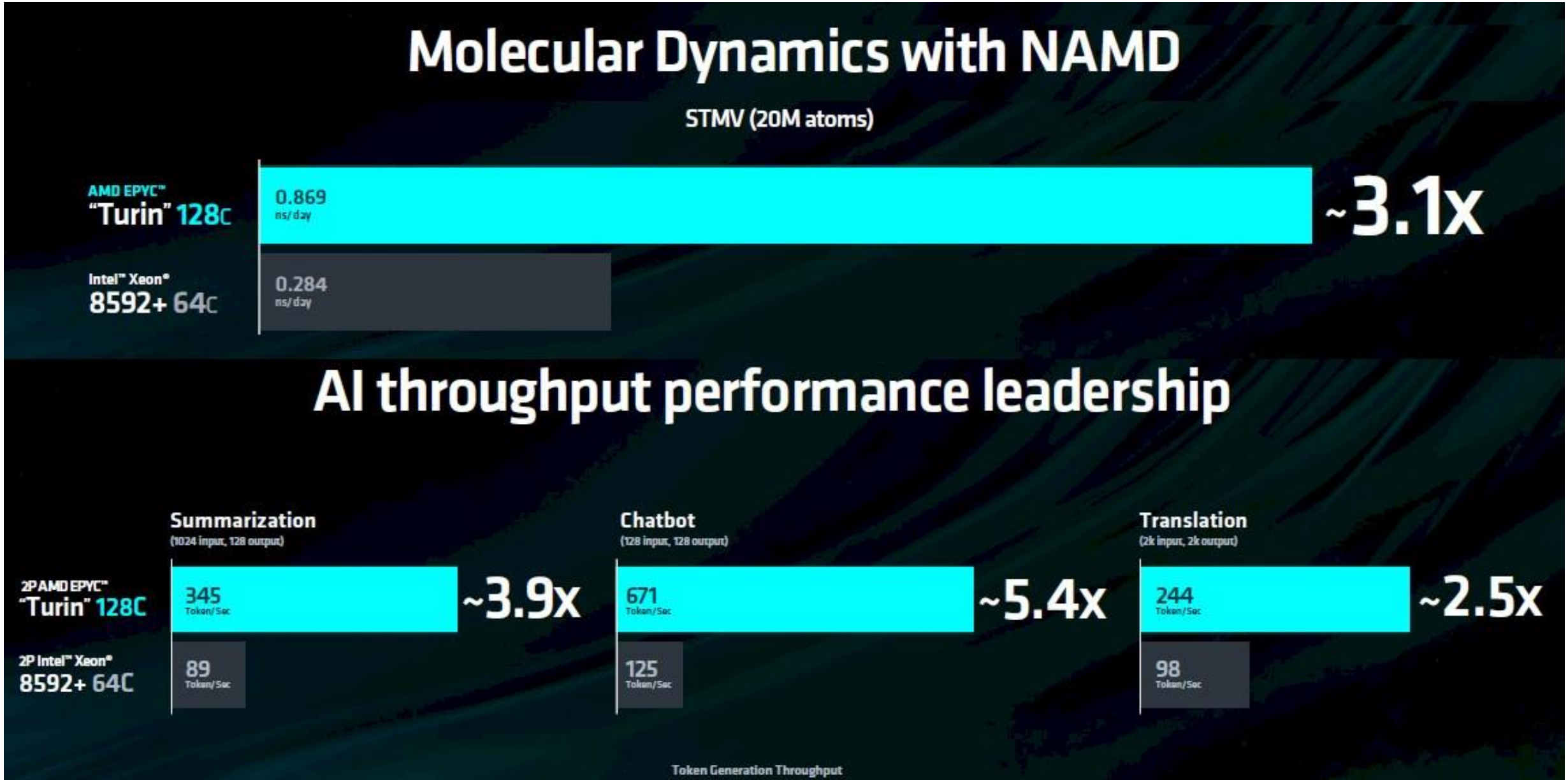


Процессоры AMD Turin (Zen 5)

Процессоры AMD Turin (Zen5)



Процессоры AMD Turin (Zen5)



Процессоры AMD Turin (Zen5)

Процессоры AMD Turin (Zen5)

AMD EPYC™ 9005 Series Processors

- Increased core density
- Energy efficient
- Broad OPN stack

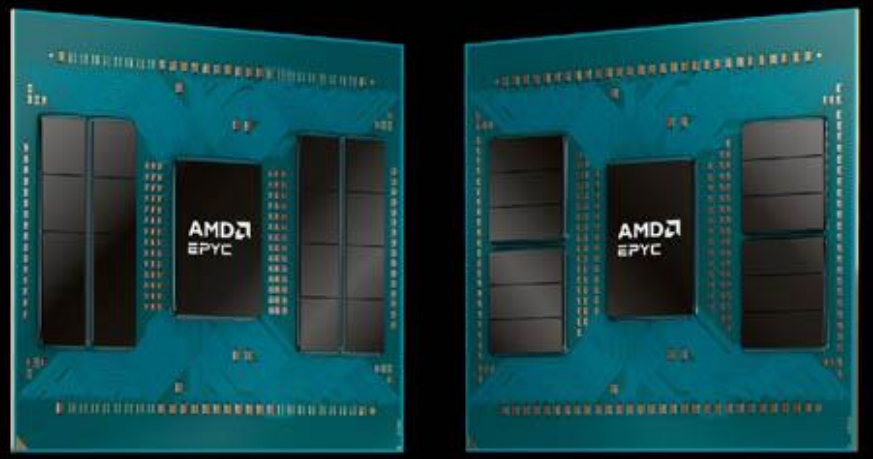
Cores	AMD EPYC	CCD (Zen5/Zen5c)	Base/Boost* (up to GHz)	Default TDP (W)	L3 Cache (MB)	Price (1 KU, USD)
192 cores	9965	"Zen5c"	2.25 / 3.7	500W	384	$14,813
160 cores	9845	"Zen5c"	2.1 / 3.7	390W	320	$13,564
144 cores	9825	"Zen5c"	2.2 / 3.7	390W	384	$13,006
128 cores	9755	"Zen5"	2.7 / 4.1	500W	512	$12,984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

Процессоры AMD Turin (Zen5)

“Turin” Continues to Deliver Technology Leadership

Scale-Up

Up to
16 “Zen 5” CCDs
128 Cores / 256 Threads



Scale-Out

Up to
12 “Zen 5c” CCDs
192 Cores / 384 Threads

Consistent features, ISA, & IPC uplift

SP5 Socket “Genoa” Compatible

8 to 192 Cores
155W to 500W

Up to
12Ch DDR5-6400*
128 PCIe 5.0/CXL 2.0

Confidential Compute with Trusted I/O

As of 10/1/2024. See endnote 9xx5-048.
*Standard roadmap offerings on AMD.com support 6000 MT/s, see endnotes 9xx5-083

Процессоры AMD Turin (Zen5)

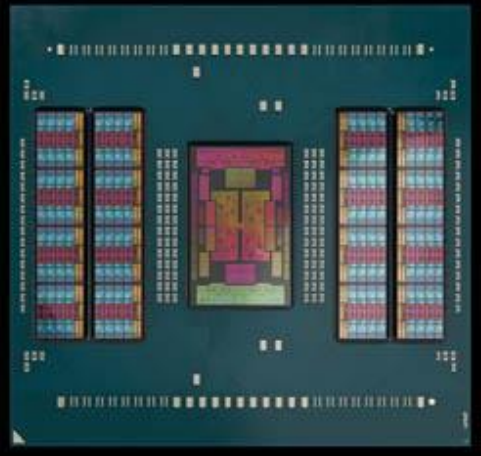
5th Gen AMD EPYC™ Generational Innovations

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* 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
*Standard roadmap offerings on AMD.com support 6000 MT/s
See endnotes 9xxS-048, 072, 083, GD-183A

Процессоры AMD Turin (Zen5)

5th Gen AMD EPYC™ SoC

Blue text indicates significant update from "Zen 4"

Compute

- AMD "Zen5/Zen5c" x86 cores
  - "Zen5" up to 16 CCDs / 128 cores / 256 threads
  - "Zen5c" up to 12 CCDs / 192 cores / 384 threads
- 1MB L2/Core, Up to 32MB L3 per CCD

- ISA Updates: AVX-512 with 512b data path

- Infinity Fabric: 3rd-gen MCM (32+Gbps die-to-die)

- Dynamic PPR for x4 and x8 ECC RDIMMs
- BMC MCA Crash-Dump
- MCA over APML

SP5 Platform

- BIOS Update Required
- 2P and 1P Configurations
- Up to 4 links of 32Gbps Gen 3 Infinity Fabric™
- Compatible with "Genoa IRM" Groups up to 400W
- 500W IRM option for maximum performance
- Flexible topology options
- Server Controller Hub (USB, UART, SPI, I2C, etc.)

Memory

- 12 channel DDR5 with ECC up to 6000 MT/s
- Support for 2, 4, 6, 8, 10, 12 channel memory interleaving
- RDIMM, 3DS RDIMM
- Up to 2 DIMMs/channel capacity of 6TB/socket (256GB 3DS RDIMMs)

I/O

- 2P: up to 160 lanes of PCIe Gen5 with speeds up to 32Gbps/lane
  - up to 12 bonus PCIe Gen3 lanes
- 1P: 128 lanes of PCIe Gen5 with speeds up to 32Gbps/lane
  - up to 8 bonus PCIe Gen3 lanes
- PCIe bifurcations supported: x16, x8, x4, x2 and x1
- PCIe link encryption[2]
- CXL® 2.0, 4 x16 Capable "P" links; Type 3, Type 1[2], Type 2[2] PoC
- Up to 32 IO lanes for SATA
- SDCI (Smart Data Cache Injection)

Security[1]

- Trusted IO
- Dedicated Security Subsystem with enhancements
- Hardware Root-of-Trust; Ciphertext-hiding capability

1 - 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

2 - CXL Type 1&2 devices and PCIe link encryption support dependent upon ecosystem readiness

AMD
together we advance_

Процессоры AMD Turin (Zen5)

5th Gen AMD EPYC™ CPU Chiplet Architecture

Up-to-16 CCDs capability

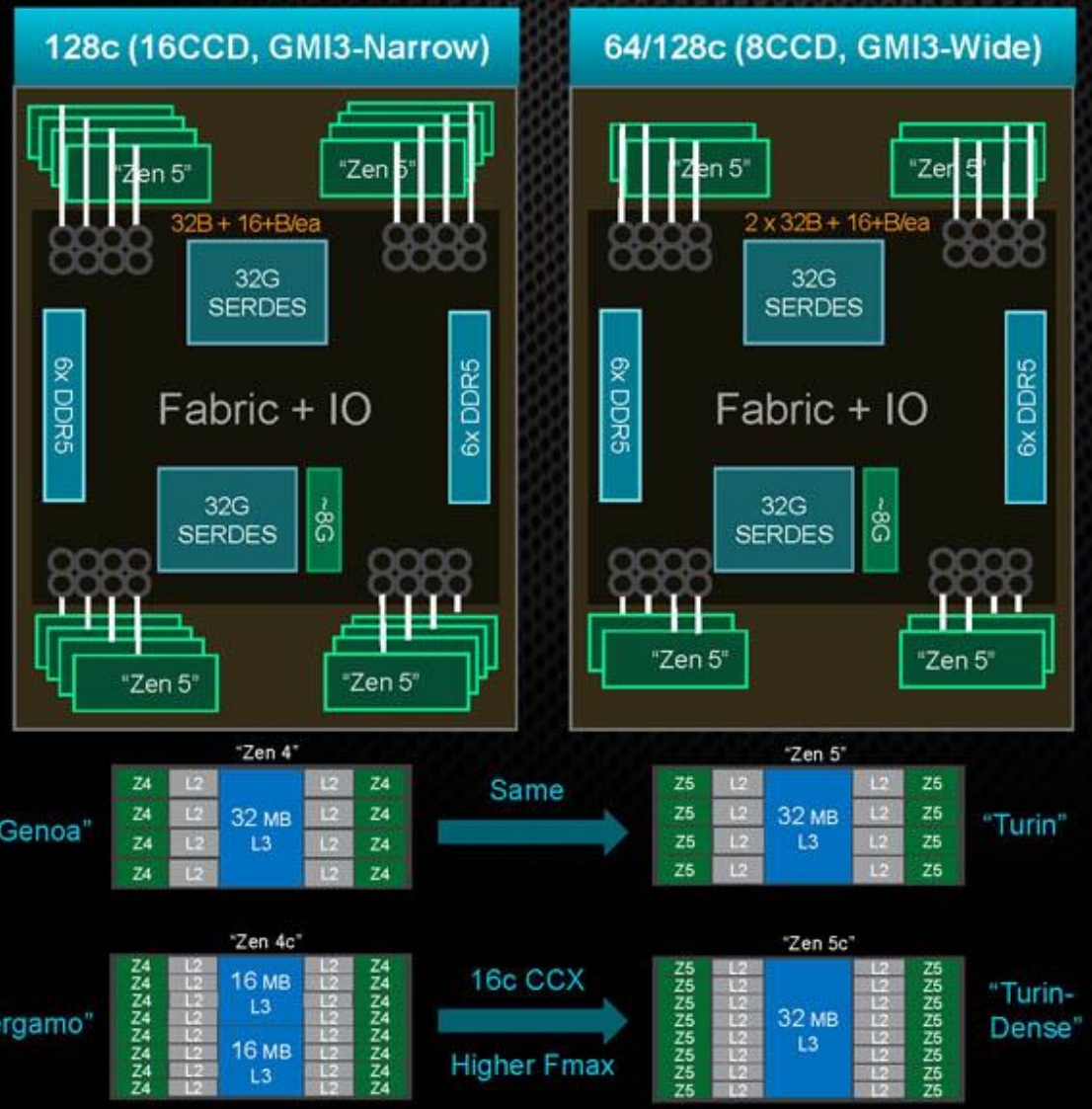
- Refined IOD/CCD and package co-design to enable

Enhanced "Turin"- Dense ("Zen 5c") vs "Bergamo" ("Zen 4c")

- Higher Fmax: 3.7GHz vs 3.1GHz
- 16c Shared L3 (CCX) for cache efficiency

GMI3 Chiplet interface

- Up-to-36Gbps, 20:1 with internal FCLK (1.8GHz max)
- Wide and Narrow connection options
  - >8CCD option use GMI3-Narrow
  - <=8CCD options can use GMI3-Wide
- 2x probe throughput vs GMI2 (3rd Gen AMD EPYC™)
- "16+B" CCD->IOD Datapath: Enhanced probe-response data (32B) and write-heavy traffic (25B) performance
- GMI "Folding" for power management (Half-width)



Процессоры AMD Turin (Zen5)

5th Gen AMD EPYC™ CPU Memory Capabilities

- 12ch DDR5/CPU; up to DDR5 6000
- 576GB/s peak theoretical BW (12ch * 8B * 6GTs)
- 1DIMM/ch and 2DIMM/ch capability
- x80 and x72 DIMMs
- RDIMM and 3DS RDIMM
- Up-to 6TB/socket capacity [1]
- AMDC (x4 DRAMs) and

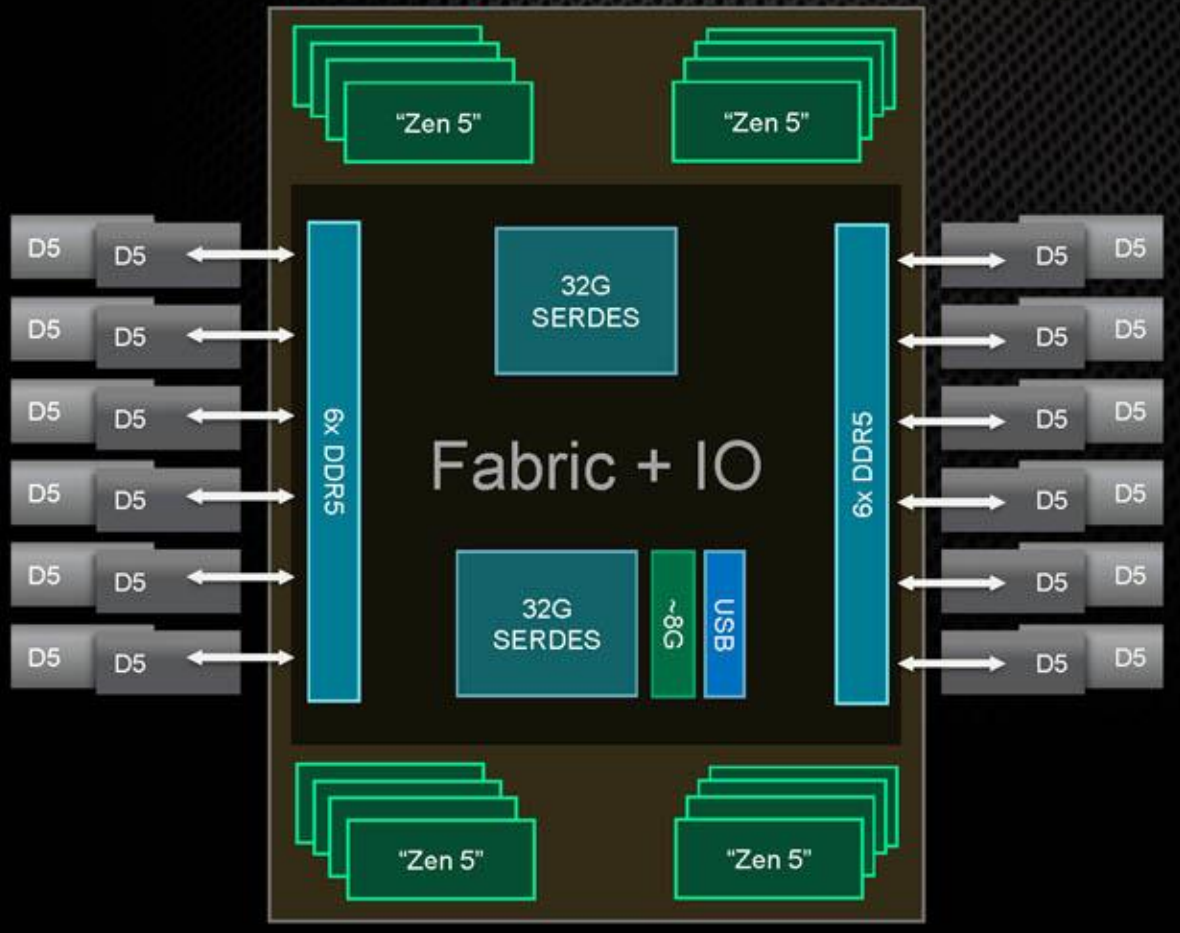
"Bounded Fault" DRAM ECC

- Read UECC retry capability
- DRAM Runtime Post-Package-Repair

(PPR) for x4 and x8 DIMMs

- Design focus on maximizing

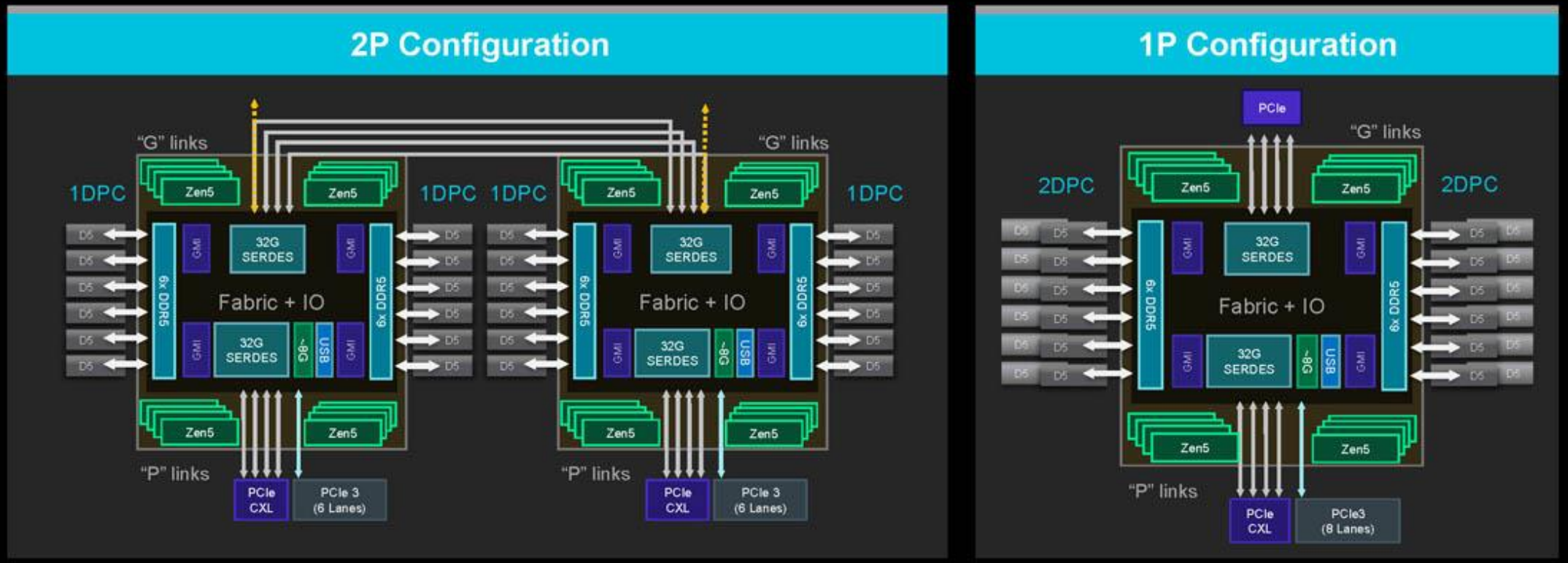
DRAM ECC bits for error detection and correction



1: 2x4R 3DS-RDIMM with 32Gbx4 Devices.

Процессоры AMD Turin (Zen5)

5th Gen AMD EPYC™ SoC Platform Overview



Strong upgrade to existing SP5 (4th Gen AMD EPYC™ CPU) in the same platforms

- 25% improvement in DRAM speed (4800 -> up to 6000) using JEDEC-standard (non-proprietary) DIMMs
- 1P PCIe® aggregate bandwidth improvement due to internal SOC topology changes
- Enhanced platform option for 500W TDP capability

Процессоры AMD Turin (Zen5)

5th Gen AMD EPYC™ CPU Memory Capabilities

- 12ch DDR5/CPU; up to DDR5 6000
- 576GB/s peak theoretical BW (12ch * 8B * 6GTs)
- 1DIMM/ch and 2DIMM/ch capability
- x80 and x72 DIMMs
- RDIMM and 3DS RDIMM
- Up-to 6TB/socket capacity [1]
- AMDC (x4 DRAMs) and

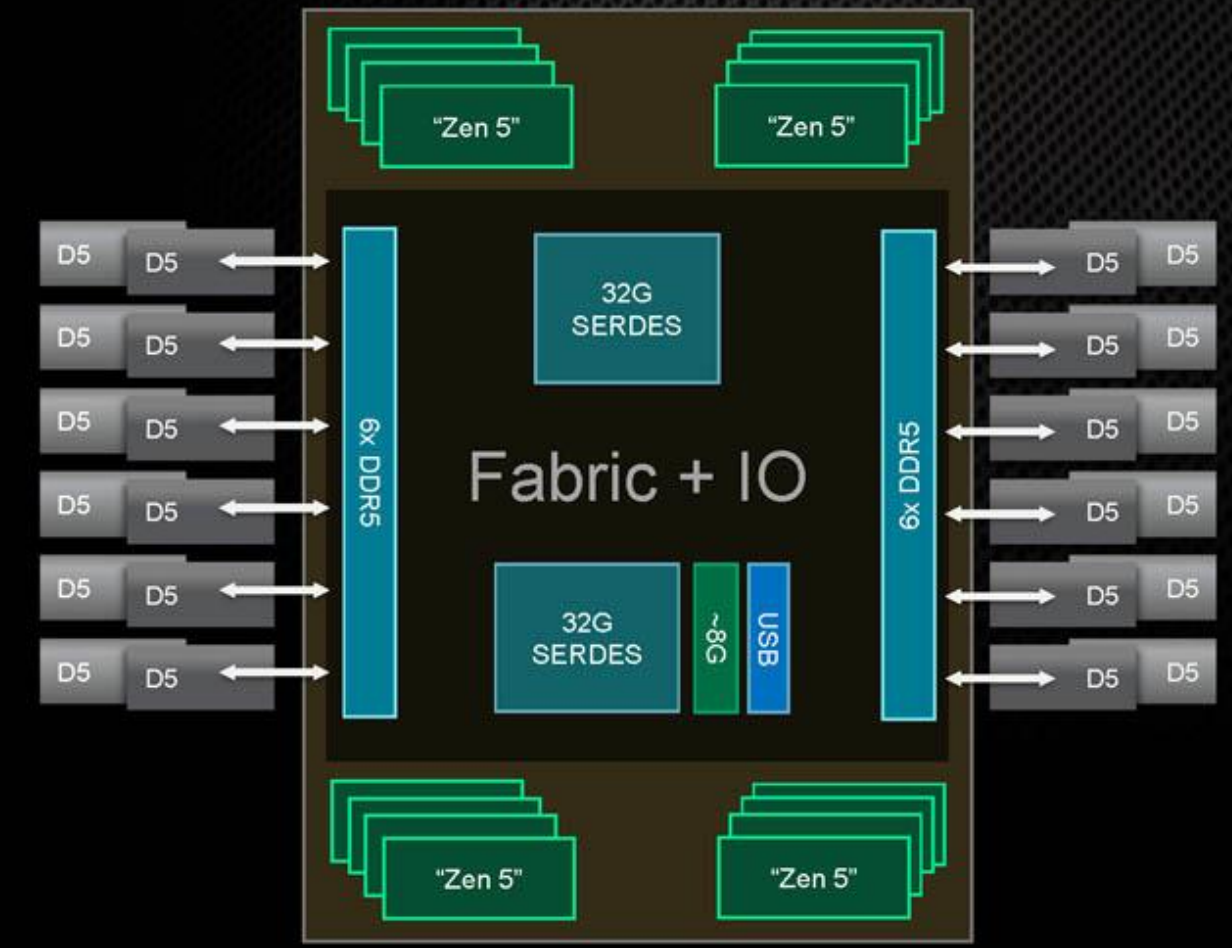
"Bounded Fault" DRAM ECC

- Read UECC retry capability
- DRAM Runtime Post-Package-Repair

(PPR) for x4 and x8 DIMMs

- Design focus on maximizing

DRAM ECC bits for error detection and correction



1: 2x4R 3DS-RDIMM with 32Gbx4 Devices.

Процессоры AMD Turin (Zen5)

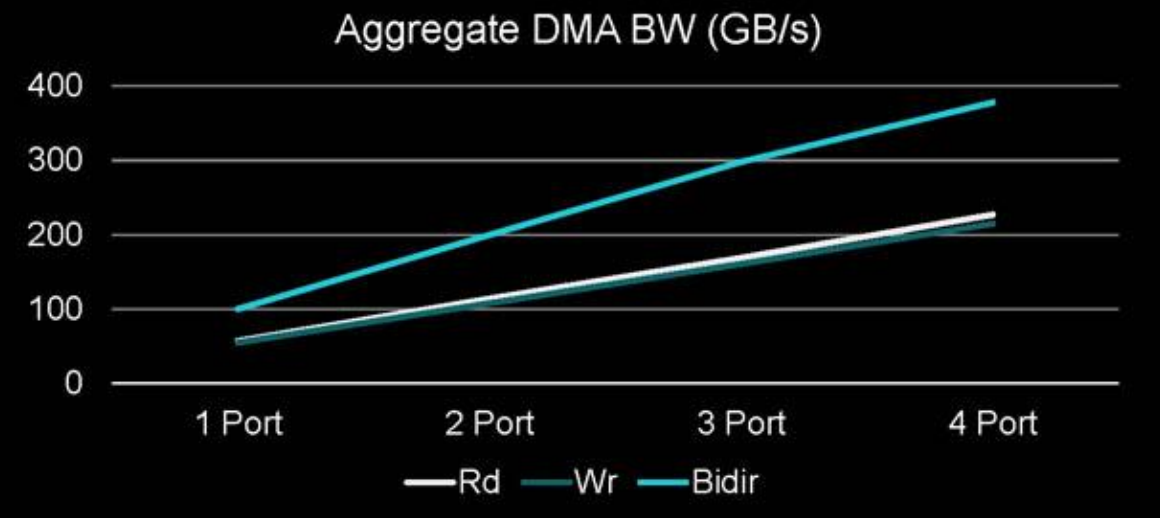
5th Gen AMD EPYC™ IO

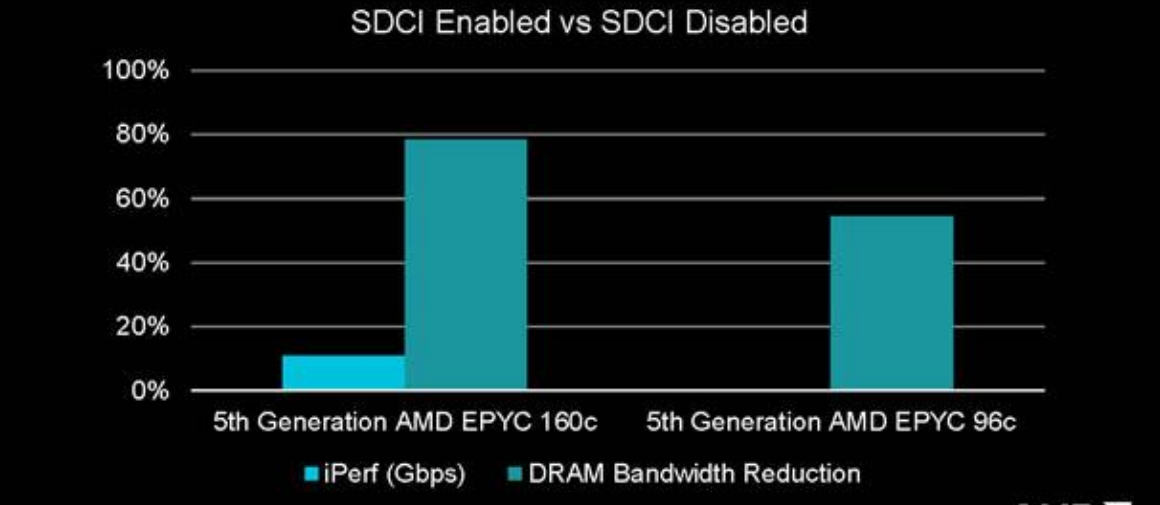
Advanced IO Capabilities

- IO performance is critical to advanced AI platform deployments, storage, and many other use models
- 5th Generation AMD EPYC™ provides strong DMA and Peer-to-Peer (P2P) capabilities for advanced system deployments
- For 1P 128L deployments leveraging AMD IO flexibility, 5th Generation AMD EPYC improves IO device performance uniformity vs the prior generation
- Trusted IO built on top of PCIe® Link Encryption for advanced security capability in SEV-SNP enabled Confidential VMs[1]

SDCI: Smart Data Cache Injection

- Enables an IO device to inject data directly into the cache hierarchy improving utilization and reducing DRAM BW for high IO workloads
- Utilizes PCIe Transaction Processing Hints (TPH) to facilitate cache injection and steering with additional capabilities for QoS management





1: Subject to Ecosystem Enablement

Процессоры AMD Turin (Zen5)

CXL® Overview

CXL industry standard Cache-Coherent Interconnect for Processors, Memory Expansion and Accelerators

5th Gen AMD EPYC™ CPU supports CXL® 2.0

- Types 1, 2 & 3 *

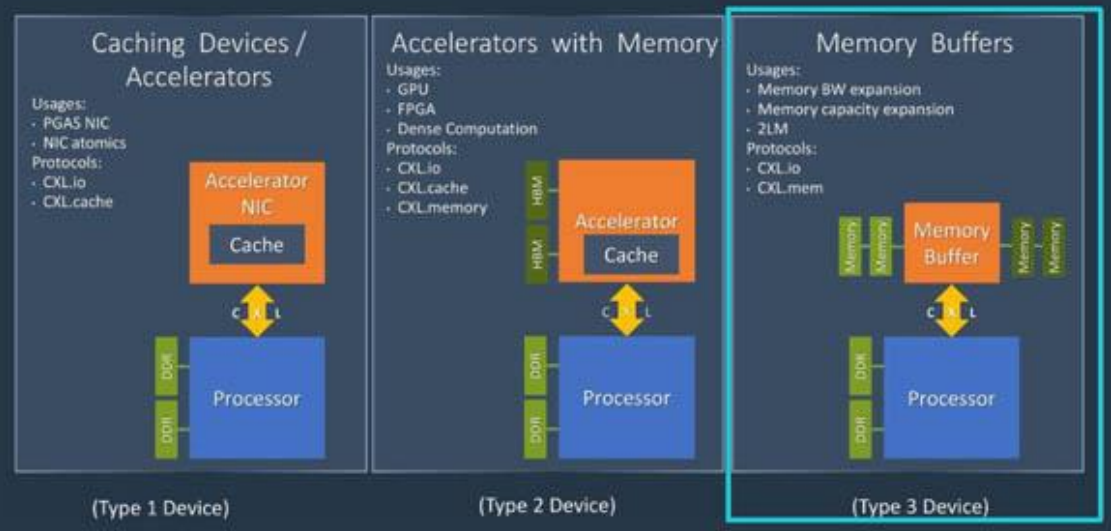
Additional CXL Capabilities

- Tiered memory support
  - Multiple CXL devices combined into single interleaved NUMA node
  - Headless NUMA nodes
  - "Pinned"-memory and Secure Memory datamover
  - Memory profiling by HV/Guest (IBS filtering, IBS virtualization)
- x86 QoS support
  - Separate x86 MBM BW limits for DRAM and CXL memory
- AMD SEV/SNP features fully supported on CXL Type 3 memory

Low latency memory attach

- Device-dependent; similar to remote 2P socket latency for local CXL-attach memory

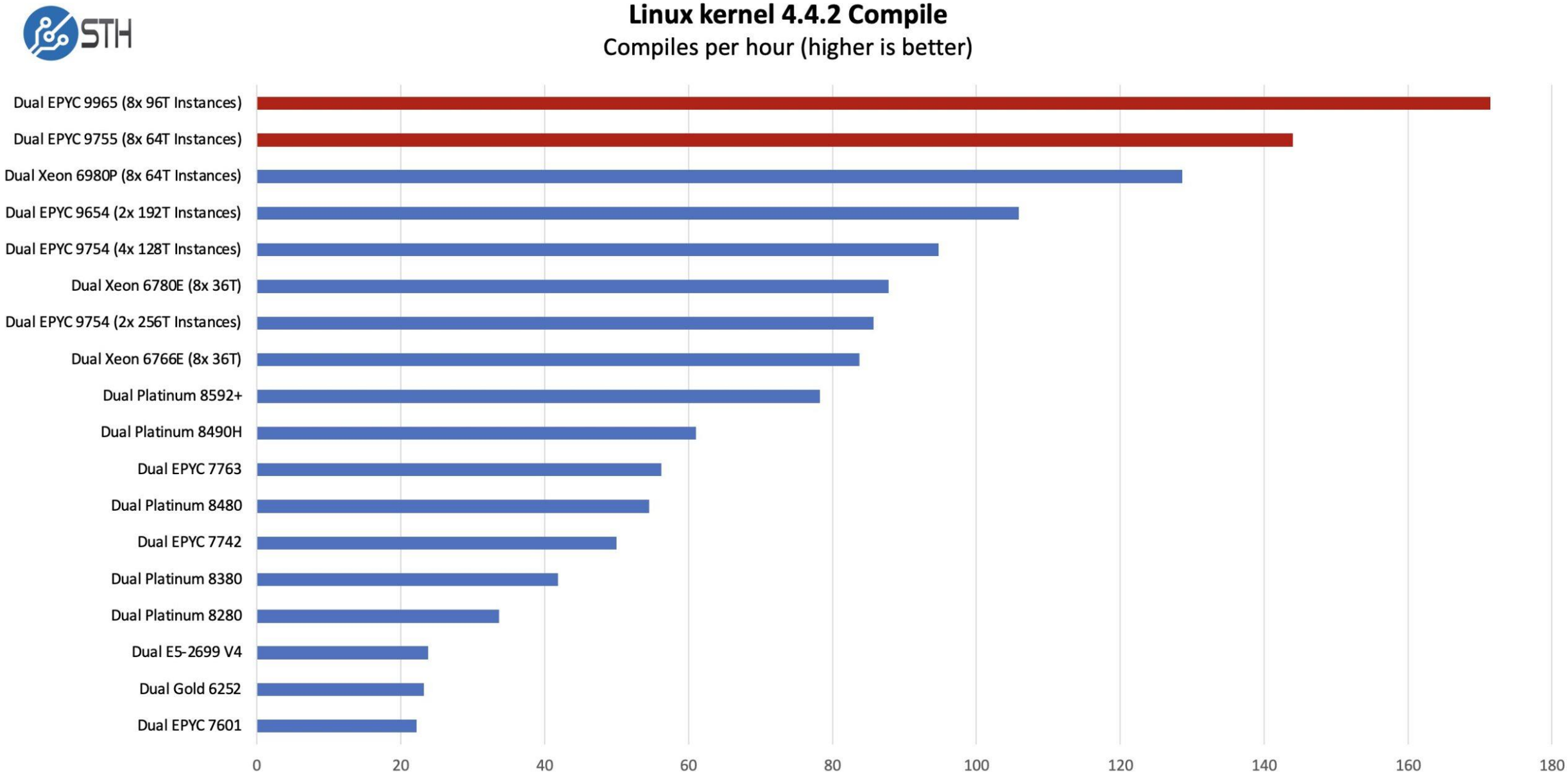
5th Gen AMD EPYC™ CPU Focus



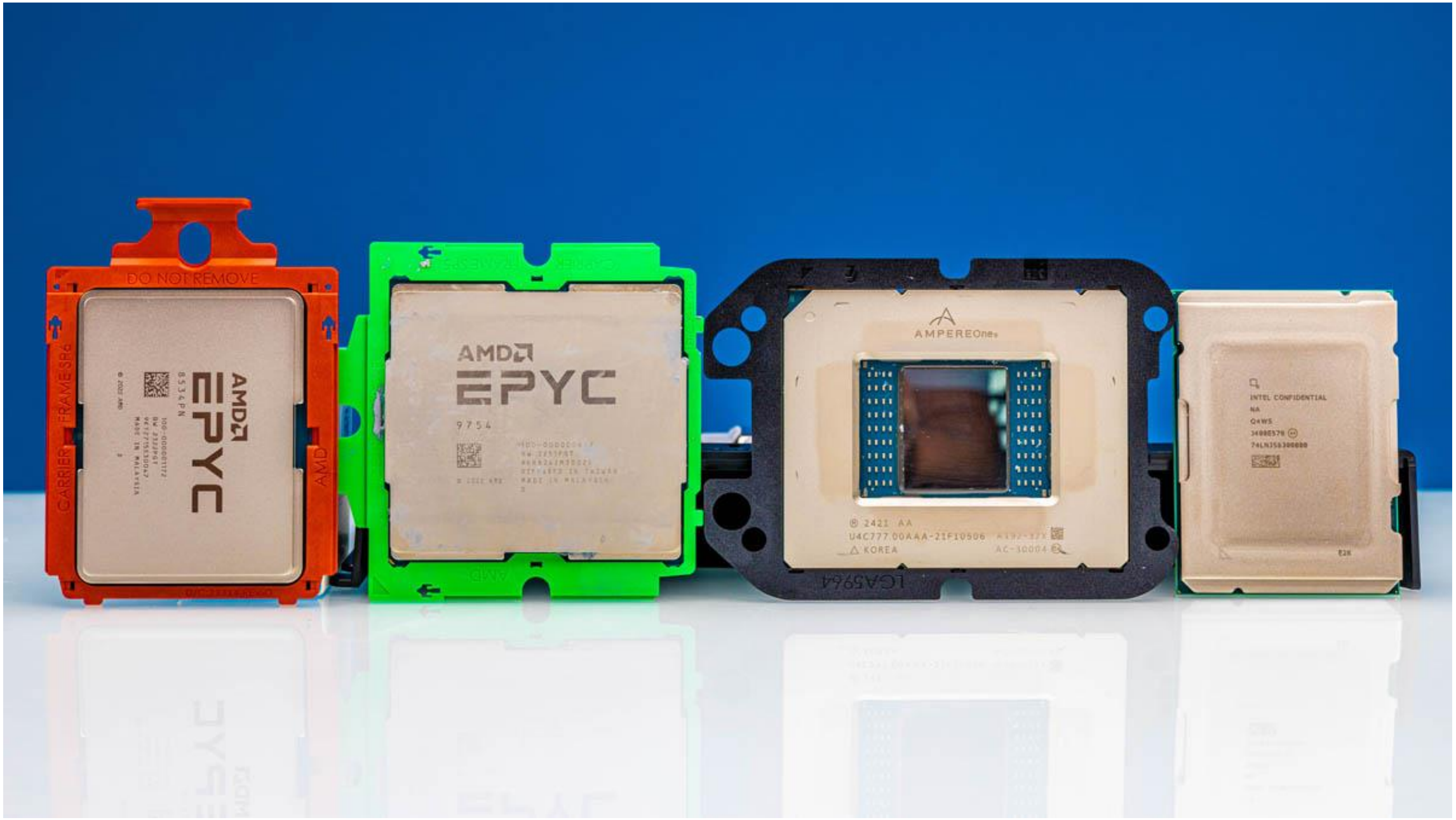
- **CXL.io:** Discovery, configuration, register access, interrupts, DMA, etc.
- **CXL.cache:** Device access to processor memory
- **CXL.memory:** Processor access to device attached memory

* CXL Type1&2 device support dependent on ecosystem readiness; type 2 PoC only

Процессоры AMD Turin (Zen5)

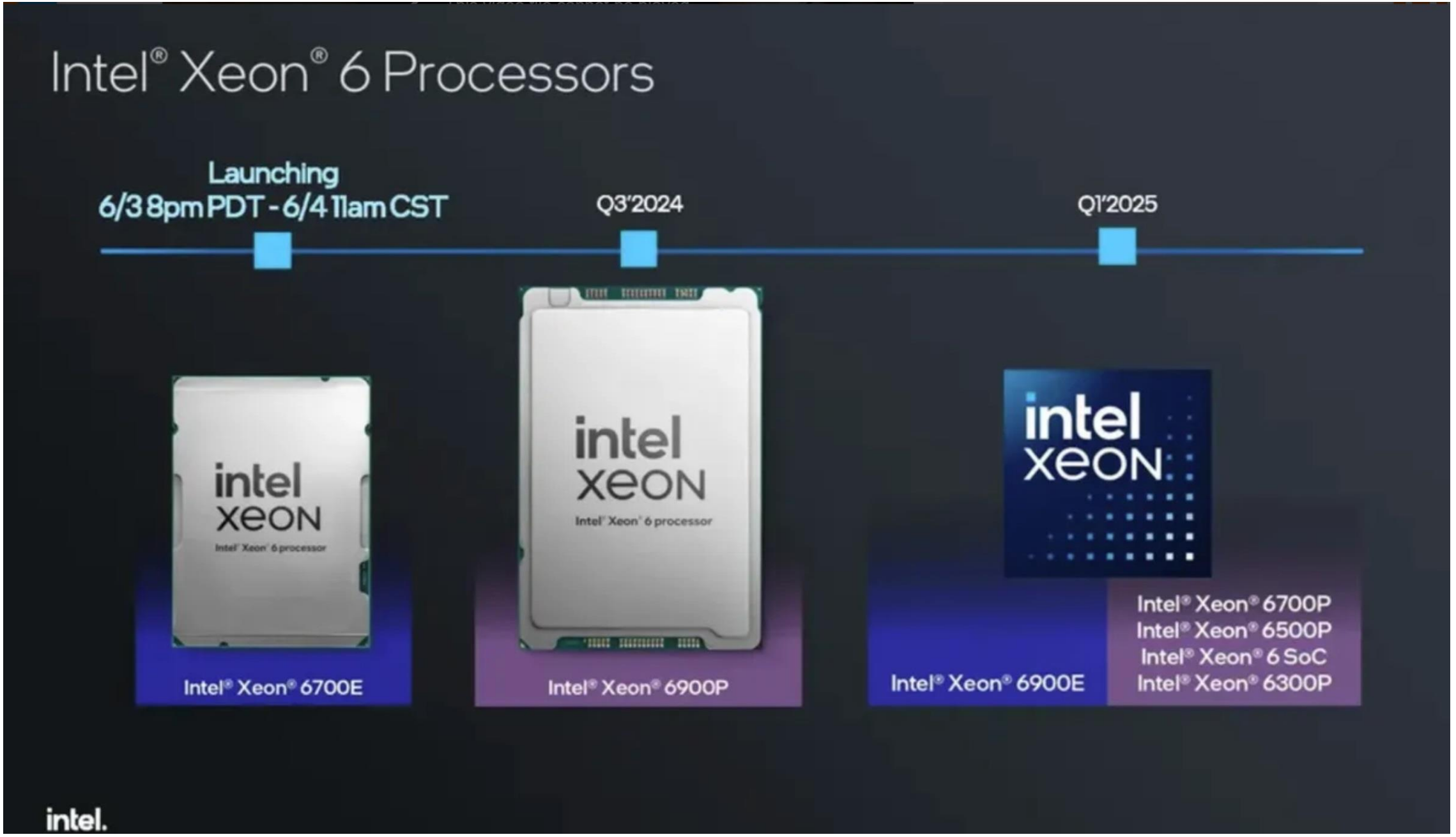


Процессоры AMD Turin (Zen5)



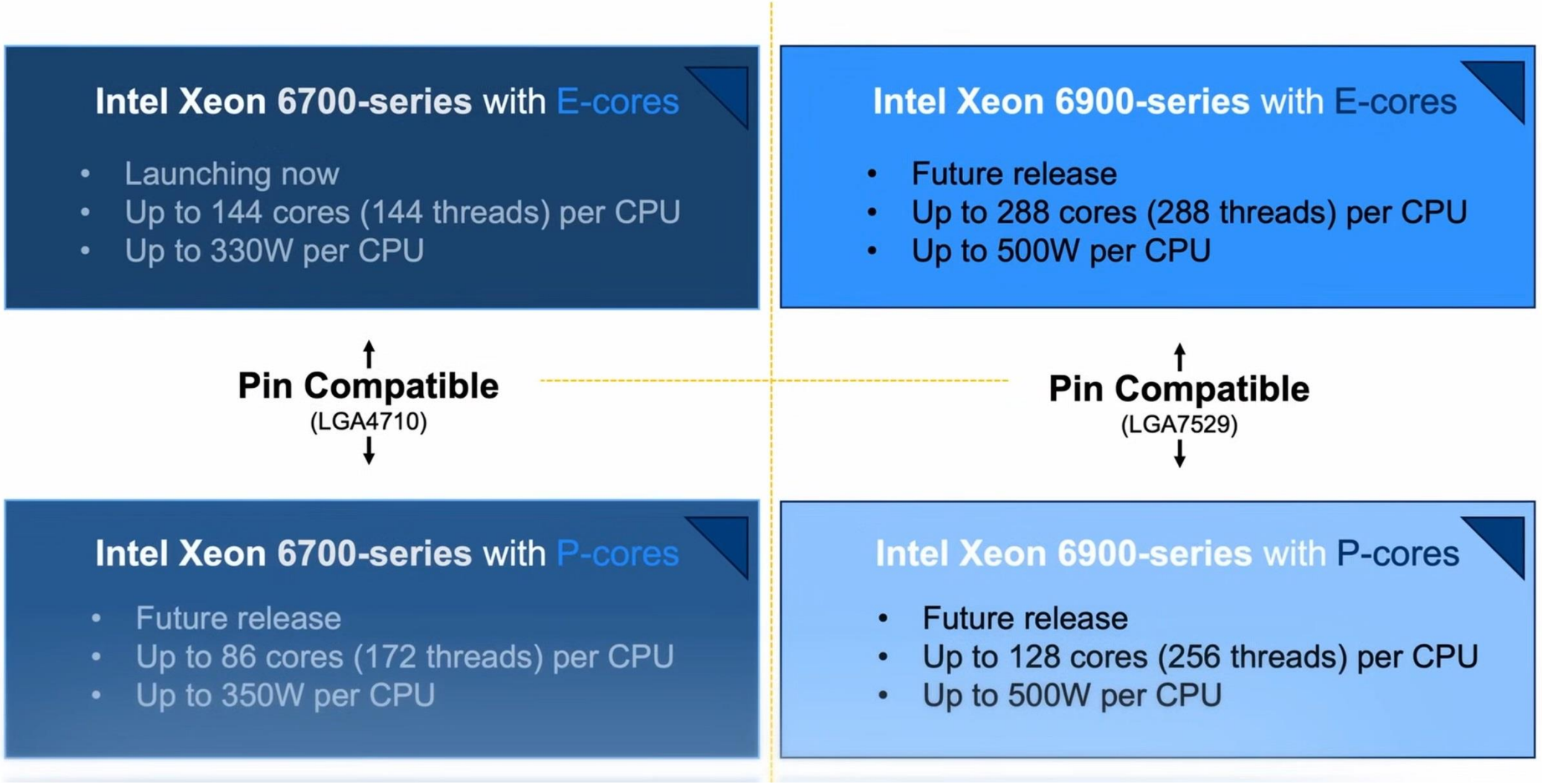
Процессоры Intel Xeon 6

Процессоры Intel Xeon 6



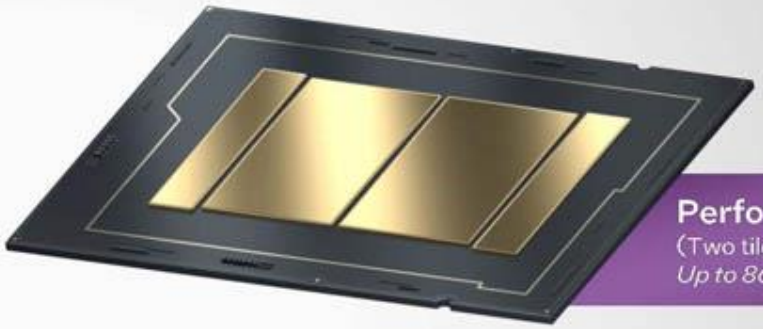
Процессоры Intel Xeon 6

Four categories of workload-optimized processors



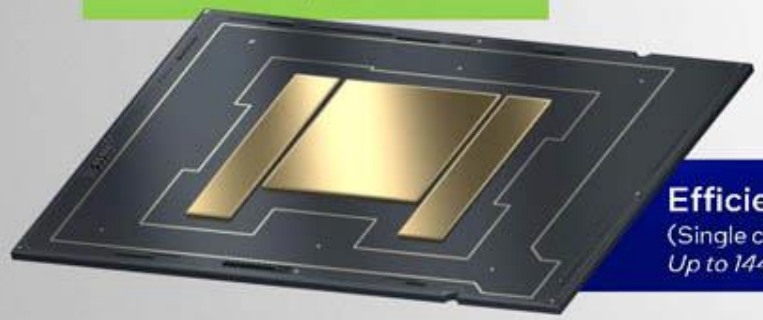
Процессоры Intel Xeon 6

Intel® Xeon® 6700 Series Die Packages



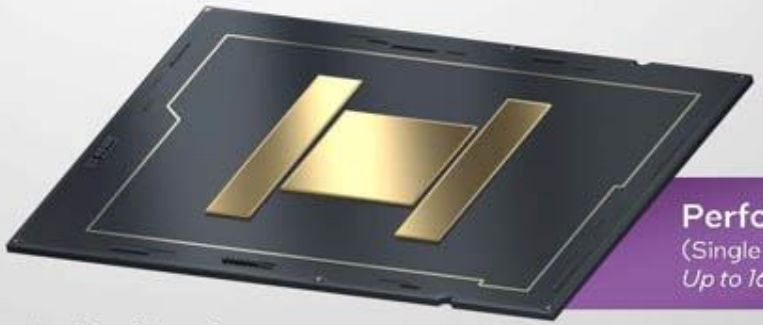
Performance Core
(Two tile compute die, XCC)
Up to 86 cores

Launching June 3, 8pm PDT
– June 4, 11am CST



Efficiency Core
(Single compute tile die)
Up to 144 cores

Performance Core
(Single compute tile die, HCC)
Up to 48 cores

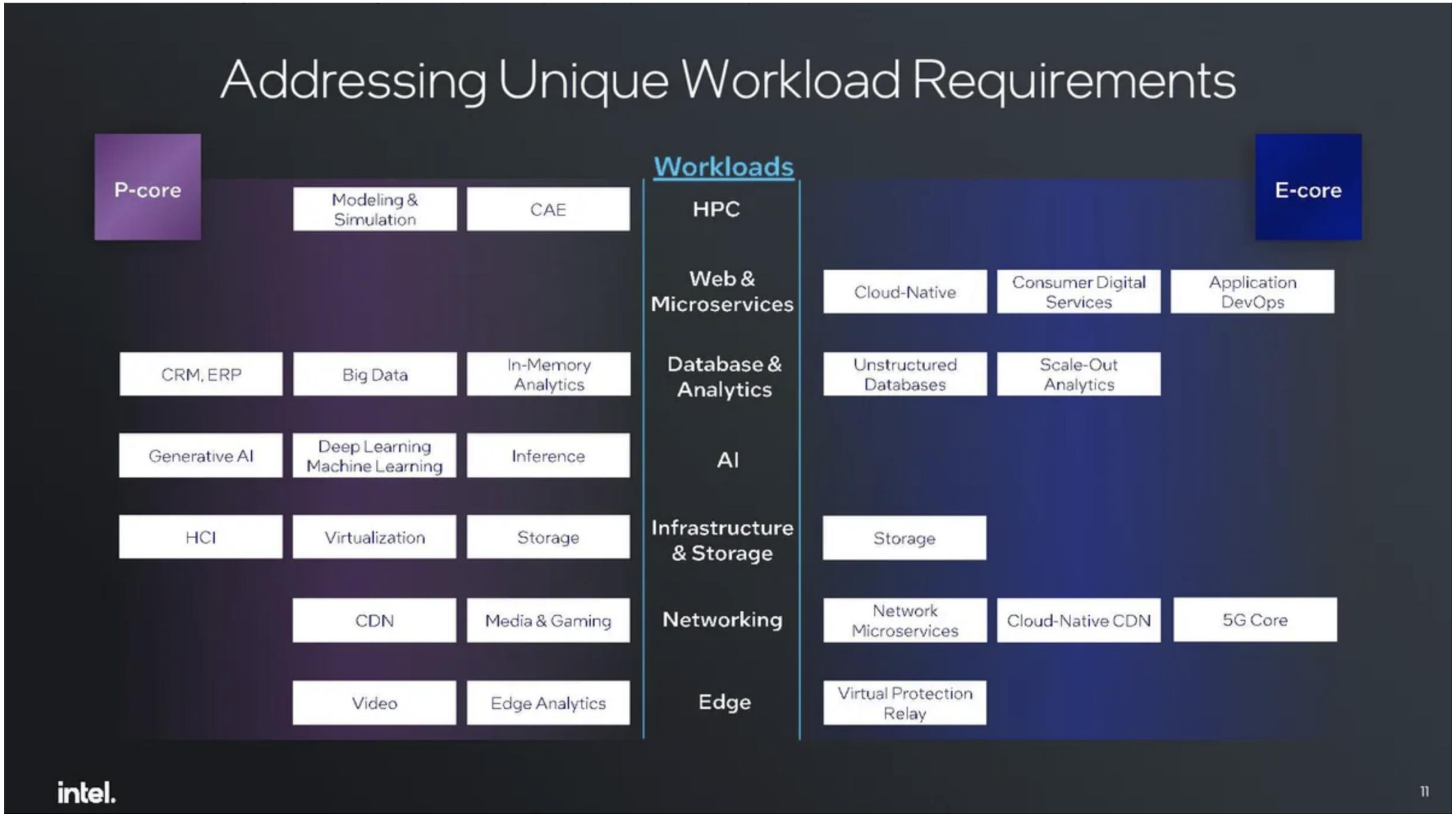


Performance Core
(Single compute tile die, LCC)
Up to 16 cores

Процессоры Intel Xeon 6

INTEL® XEON® 6 PROCESSORS Performance Core		INTEL® XEON® 6 PROCESSORS Efficient Core
up to **86 cores (6700)** or **128 cores (6900)**	Cores	up to **144 cores (6700)** or **288 cores (6900)**
1S, 2S, 4S, 8S	Sockets	**1S, 2S**
125 to 500W	Max TDP	205 to 500W
up to **12 channels DDR5 | MCR DIMM**	Memory	up to **12 channels DDR5**
6400 (1 DPC) | 5200 (2 DPC) | **8800 MCR (1 DPC)**	Max Memory Speed	6400 (1 DPC) | 5200 (2 DPC)
up to **6 UPI 2.0** | up to **24 GT/s** per lane	Intel® UPI	up to **4 UPI 2.0** | up to **24 GT/s** per lane
up to **96 lanes PCIe 5.0** (x16, x8, x4, x2) **R1S:** up to **136 lanes PCIe 5.0 for single socket designs**	PCI Express	up to **96 lanes PCIe 5.0** (x16, x8, x4, x2)
up to 64 lanes **CXL 2.0**	Compute Express Link	up to 64 lanes **CXL 2.0**
52/57	Physical/Virtual Address Bits	**52/48**
Intel Advanced Matrix Extensions (INT8, BF16, FP16) **Intel Advanced Vector Extensions 512** (VNNI/INT8)	AI Acceleration Intel® Deep Learning Boost	**Intel Advanced Vector Extensions 2** (VNNI/INT8)
Intel Software Guard Extensions, Intel Trusted Domain Extensions	Security	Intel Software Guard Extensions, Intel Trusted Domain Extensions
Vector AES, SHA2-256 extensions, VPMADD52	Crypto	Vector AES, SHA2-256 extensions, VPMADD52
Intel QuickAssist Technology, **Intel Dynamic Load Balancer, Intel Data Streaming Accelerator, Intel In-memory Analytics Accelerator**	Integrated Accelerators	Intel QuickAssist Technology, **Intel Dynamic Load Balancer, Intel Data Streaming Accelerator, Intel In-memory Analytics Accelerator**

Процессоры Intel Xeon 6



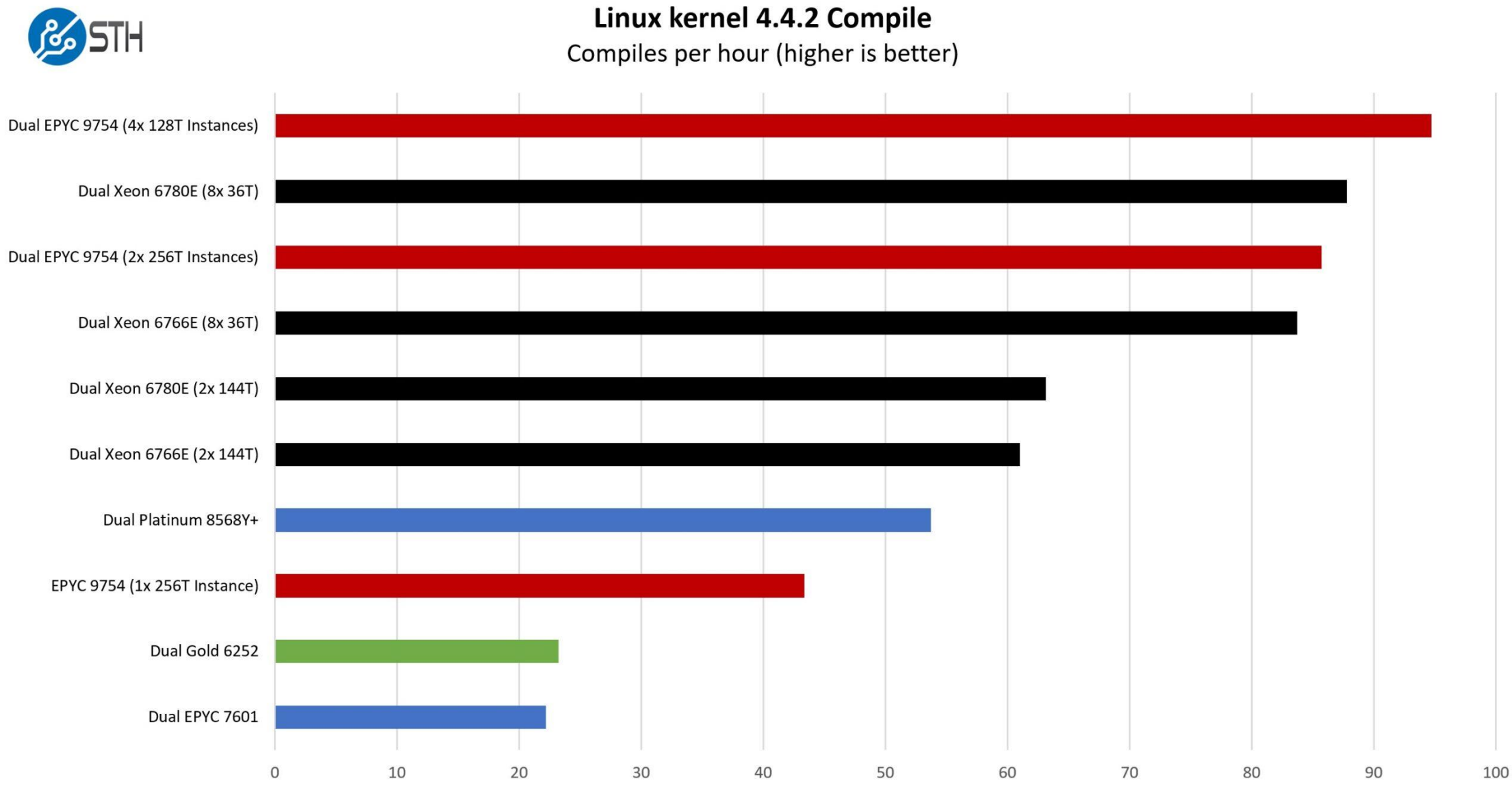
Процессоры Intel Xeon 6 (6700-series)

Intel® Xeon® 6 E-core SKU Map

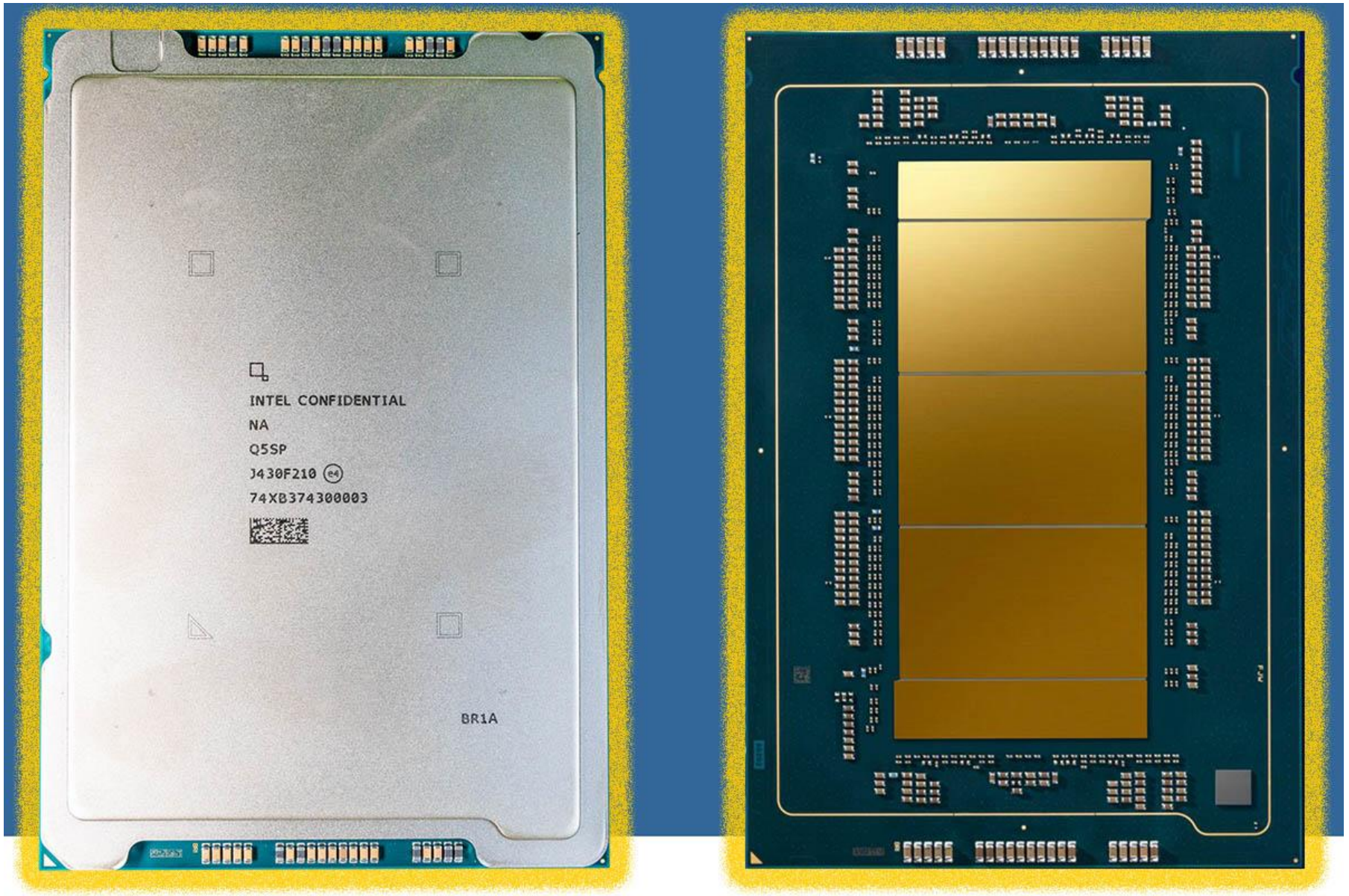
SKU	Cores	Micro Architecture	Base (GHz)	All Core Turbo (GHz)	Max Turbo (GHz)	L3 Cache (MB)	TDP (Watts)	Max Scala.	DDR5 Memory Speed (1 DPC)	Default Accelerator Devices	Intel TDX Keys (Per CPU)	Long Life Available*	UPI Links Enab.	PCIe5 Express Lanes/ CXL
6780E	144	E-core	2.2	3.0	3.0	108	330	2S	6400	2 Intel® DSA, 2 Intel® IAA, 2 Intel® QAT, 2 Intel® DLB	2048	✔	4	88
6766E	144	E-core	1.9	2.7	2.7	108	250	2S	6400	2 Intel DSA, 2 Intel IAA, 2 Intel QAT, 2 Intel DLB	1024	✔	4	88
6756E	128	E-core	1.8	2.6	2.6	96	225	2S	6400	2 Intel DSA, 2 Intel IAA, 2 Intel QAT, 2 Intel DLB	1024	-	4	88
6746E	112	E-core	2.0	2.7	2.7	96	250	2S	5600	2 Intel DSA, 2 Intel IAA, 2 Intel QAT, 2 Intel DLB	1024	-	4	88
6740E	96	E-core	2.4	3.2	3.2	96	250	2S	6400	2 Intel DSA, 2 Intel IAA, 4 Intel QAT, 4 Intel DLB	1024	✔	4	88
6731E	96	E-core	2.2	3.1	3.1	96	250	1S	5600	2 Intel DSA, 2 Intel IAA, 2 Intel QAT, 2 Intel DLB	1024	-	0	88
6710E	64	E-core	2.4	3.2	3.2	96	205	2S	5600	2 Intel DSA, 2 Intel IAA, 4 Intel QAT, 4 Intel DLB	1024	✔	4	88

*Long Life Availability: 7+ years

Процессоры Intel Xeon 6 (6700 series)



Процессоры Intel Xeon 6 (6900-series)



Процессоры Intel Xeon 6 (6900-series)

Intel Xeon 6 P-core SKU Map

Roadmap SKUs with customization options

PERFORMANCE SKUs

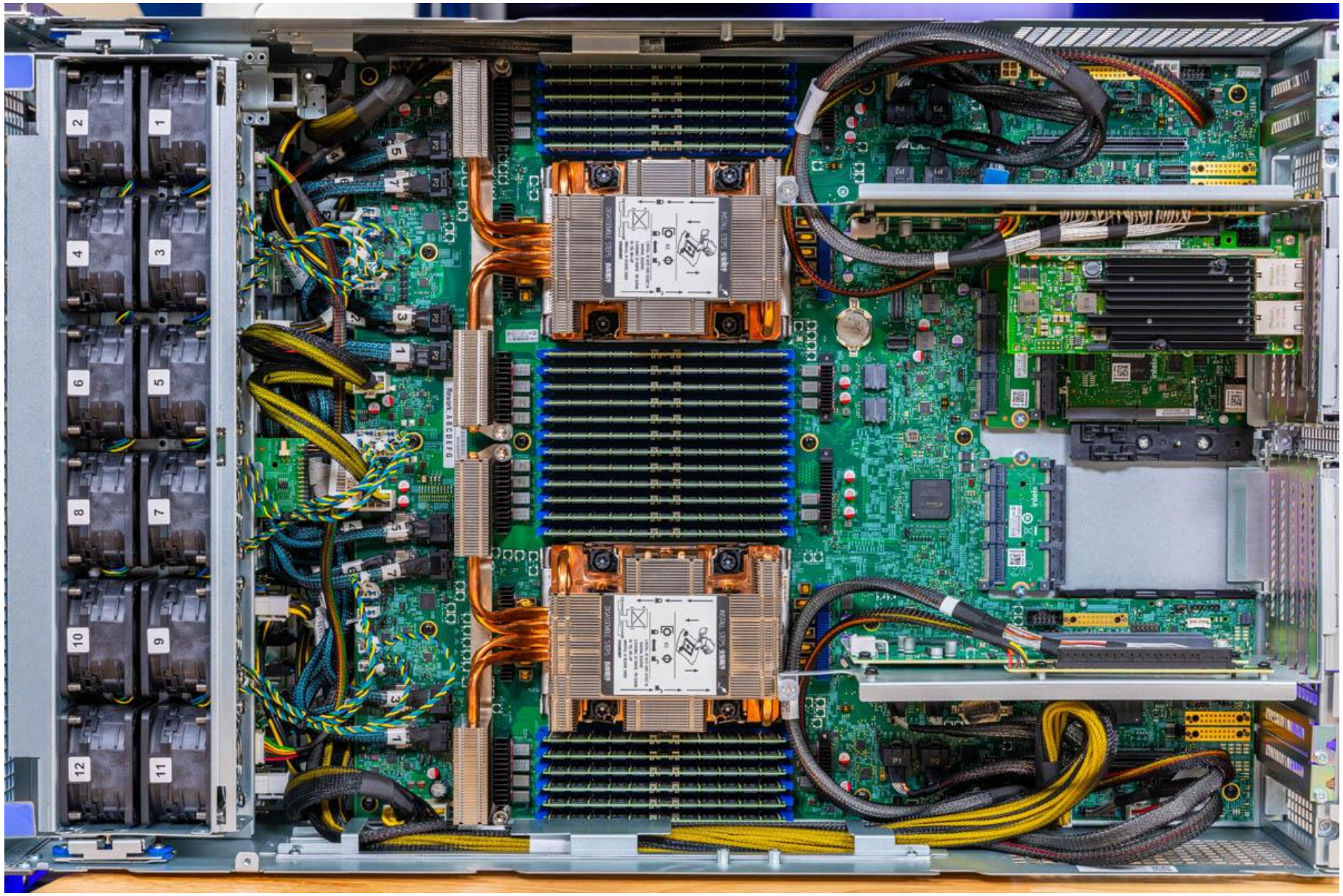
SKU	CORES	BASE (GHz)	ALL CORE TURBO (GHz)	Max TURBO (GHz)	L3 CACHE (MB)	TDP (Watts)	Max. Scala.	Memory Channels	DDR5 Memory Speed	MRDIMM Speed	Default Accel. Devices	Intel TDX Keys (Per CPU)	UPI Links Enab.	PCIe Lanes
6980P	128	2.0	3.2	3.9	504	500	2S	12	6400	8800	4/4/4/4	1024	6	96
6979P	120	2.1	3.2	3.9	504	500								
6972P	96	2.4	3.5	3.9	480	500								
6952P	96	2.1	3.2	3.9	480	400								
6960P	72	2.7	3.8	3.9	432	500								

Intel may make changes to specifications and product descriptions at any time, without notice.

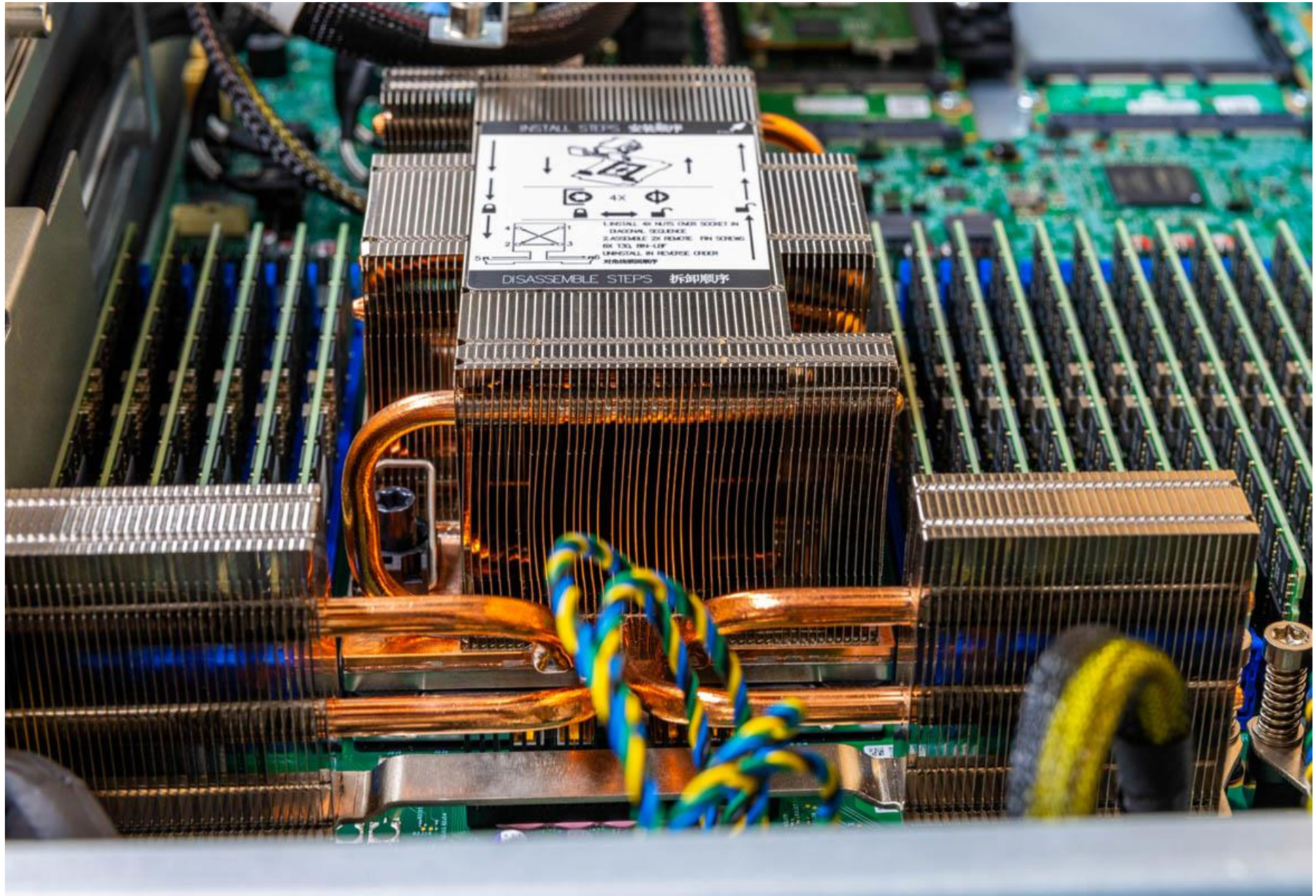
Please visit intel.com/xeon or contact your Intel representative to obtain the latest product specifications. Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. All processors support Intel Virtualization Technology (Intel VT-x).

*Accelerators List Order:
DSA, IAA, QAT, DLB
Intel® AMX featured in each core

Процессоры Intel Xeon 6 (6900-series)



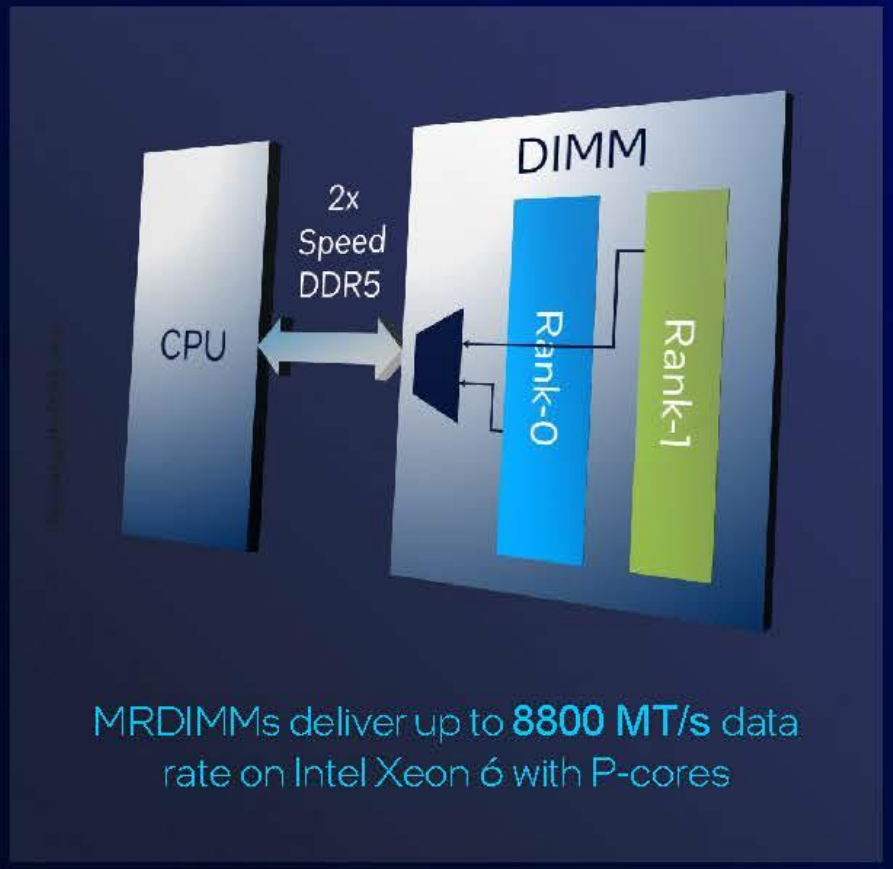
Процессоры Intel Xeon 6 (6900-series)



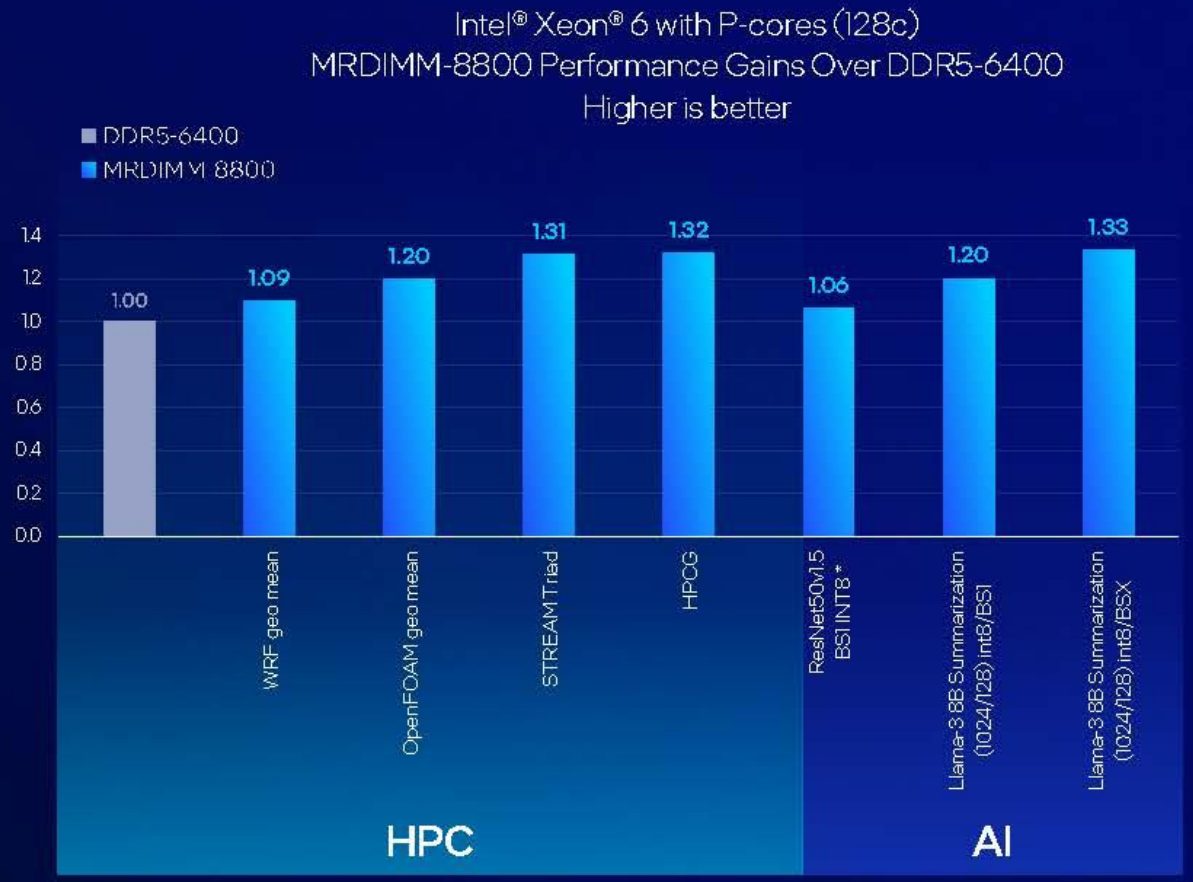
Процессоры Intel Xeon 6 (6900-series)

Multiplexed Rank DIMMs

First to market on Intel Xeon 6 processors with P-core



MRDIMMs deliver up to **8800 MT/s** data rate on Intel Xeon 6 with P-cores

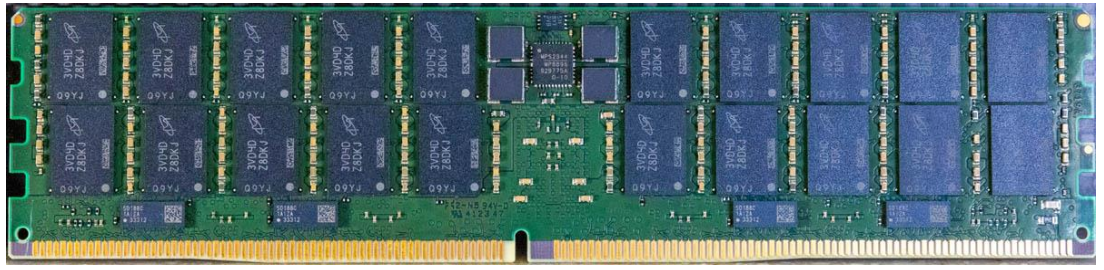


See intel.com/processorclaims: Intel Xeon 6. Results may vary. * 6972P (96c) used.
This offering is not approved or endorsed by OpenCFD Limited, producer and distributor of the OpenFOAM software via www.openfoam.com, and owner of the OPENFOAM® and OpenCFD® trademark.

Процессоры Intel Xeon 6 (6900-series)

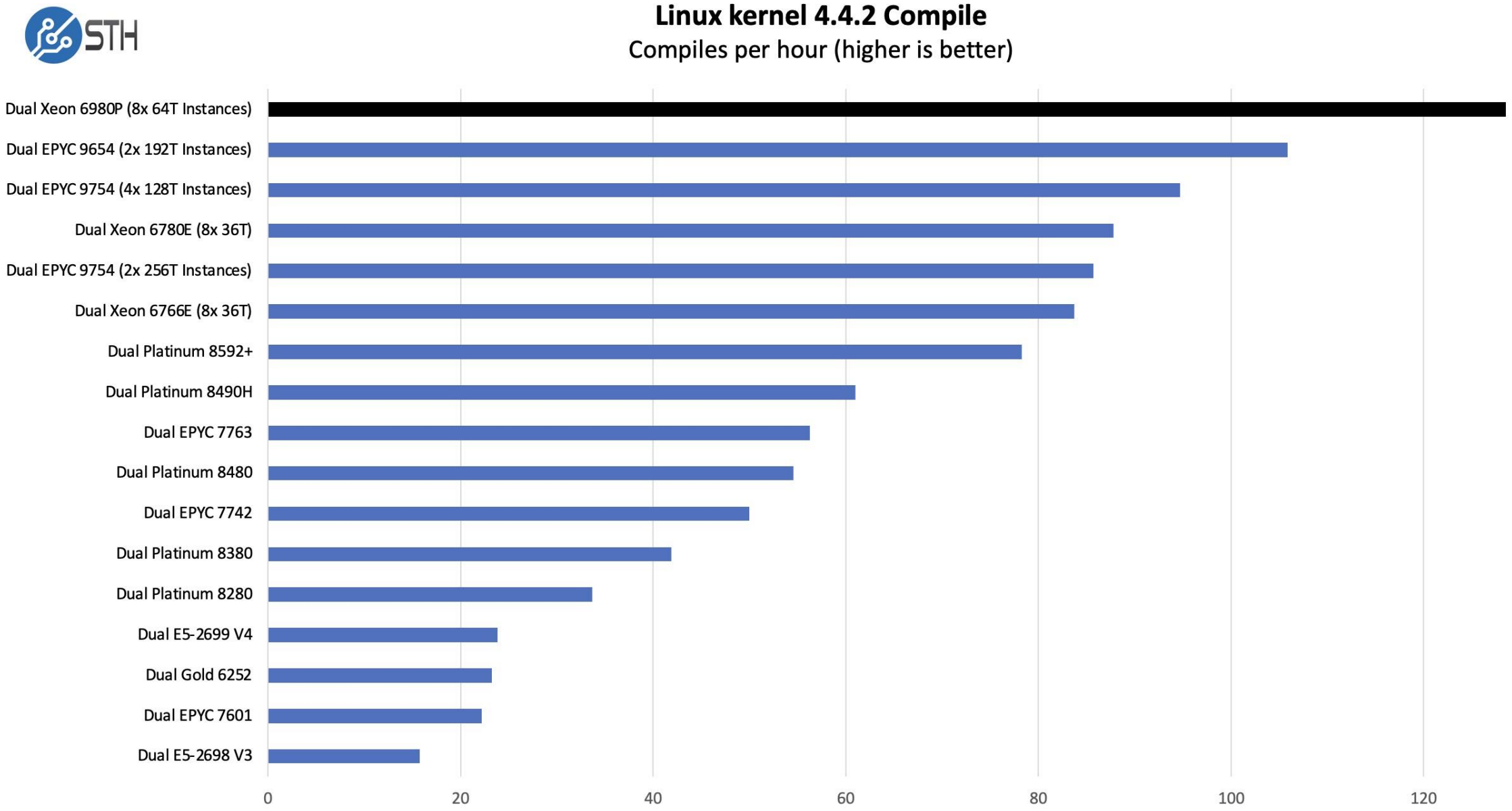
память MRDIMM

Память DDR5-6400



Память MRDIMM-8800

Процессоры Intel Xeon 6 (6900-series)

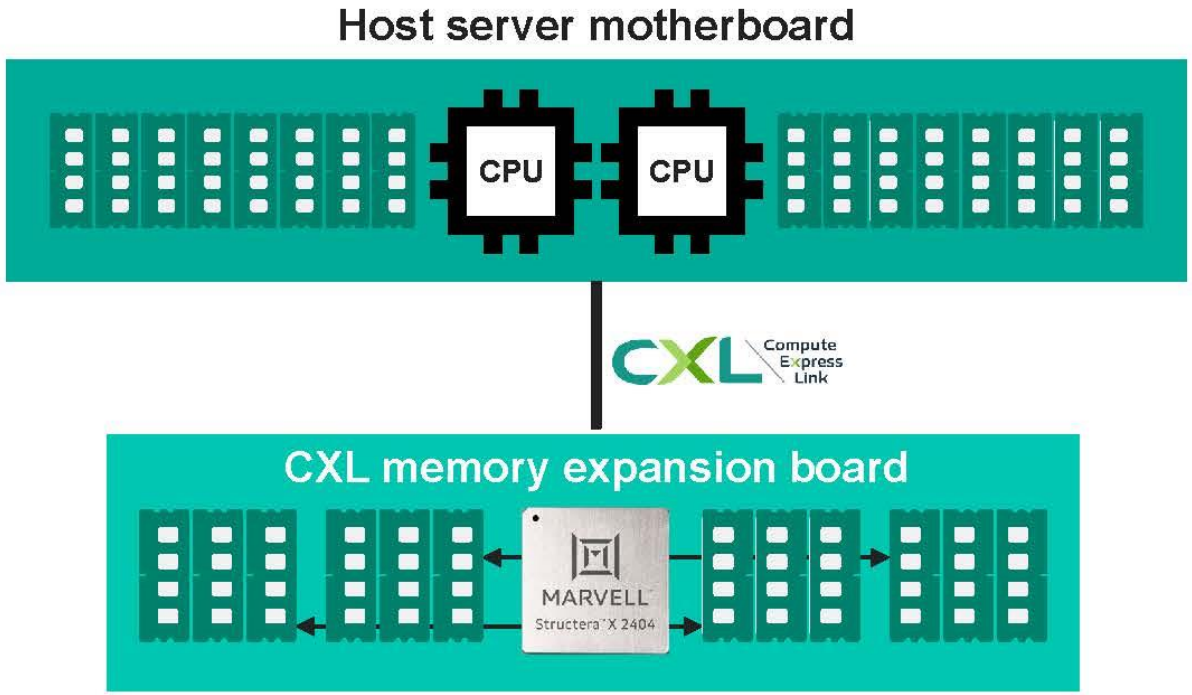


Процессоры Intel Xeon 6 (6900-series)

Процессоры Intel Xeon 6 (6900-series)

Процессоры Intel Xeon 6 (6900-series)

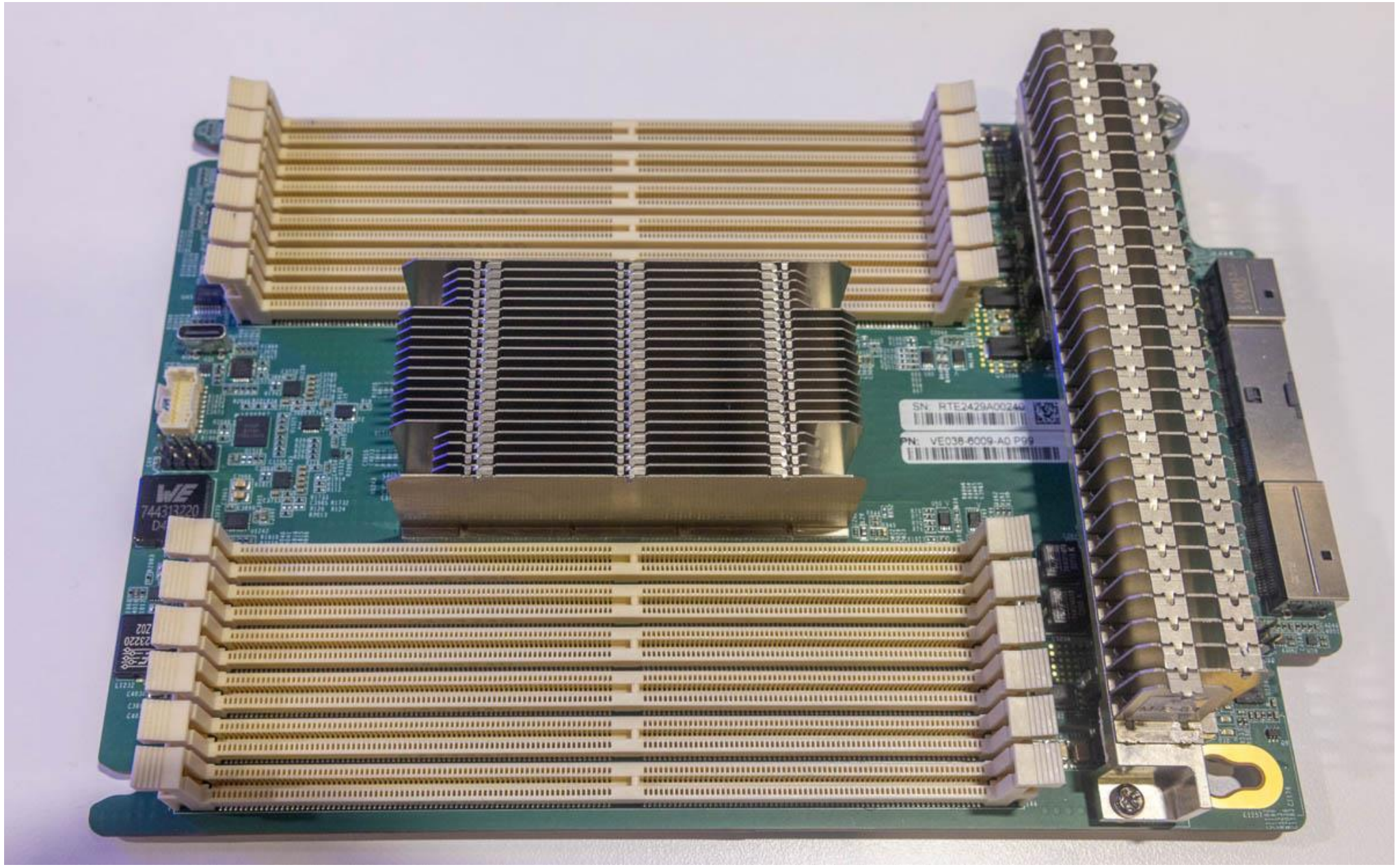
Structera X 2404 enables **DDR4 memory recycling**



Recycle up to **12 DDR4 DIMMs** per expander (up to 6TB)

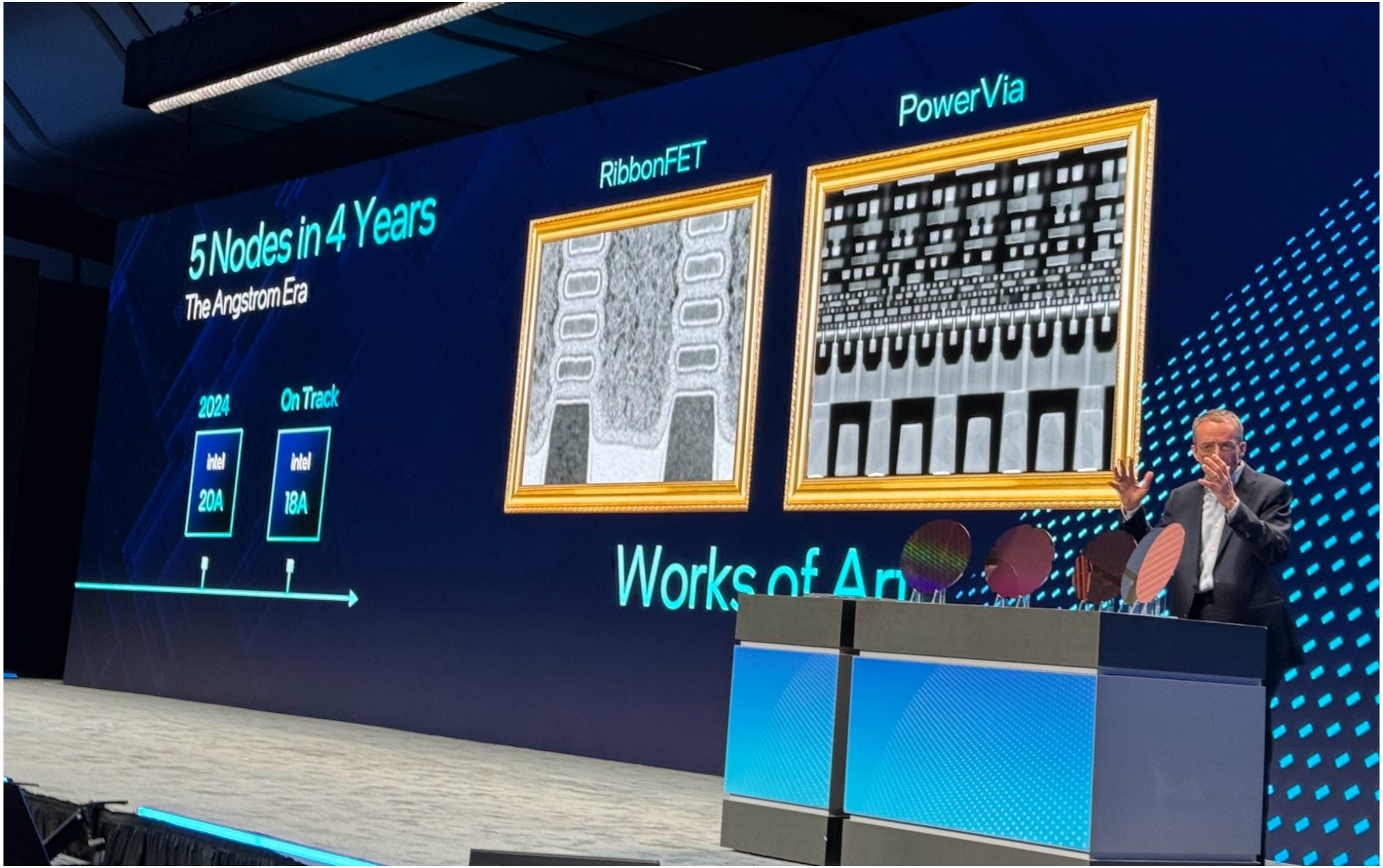
Increases server memory capacity with lower CAPEX and reduces e-waste

Процессоры Intel Xeon 6 (6900-series)



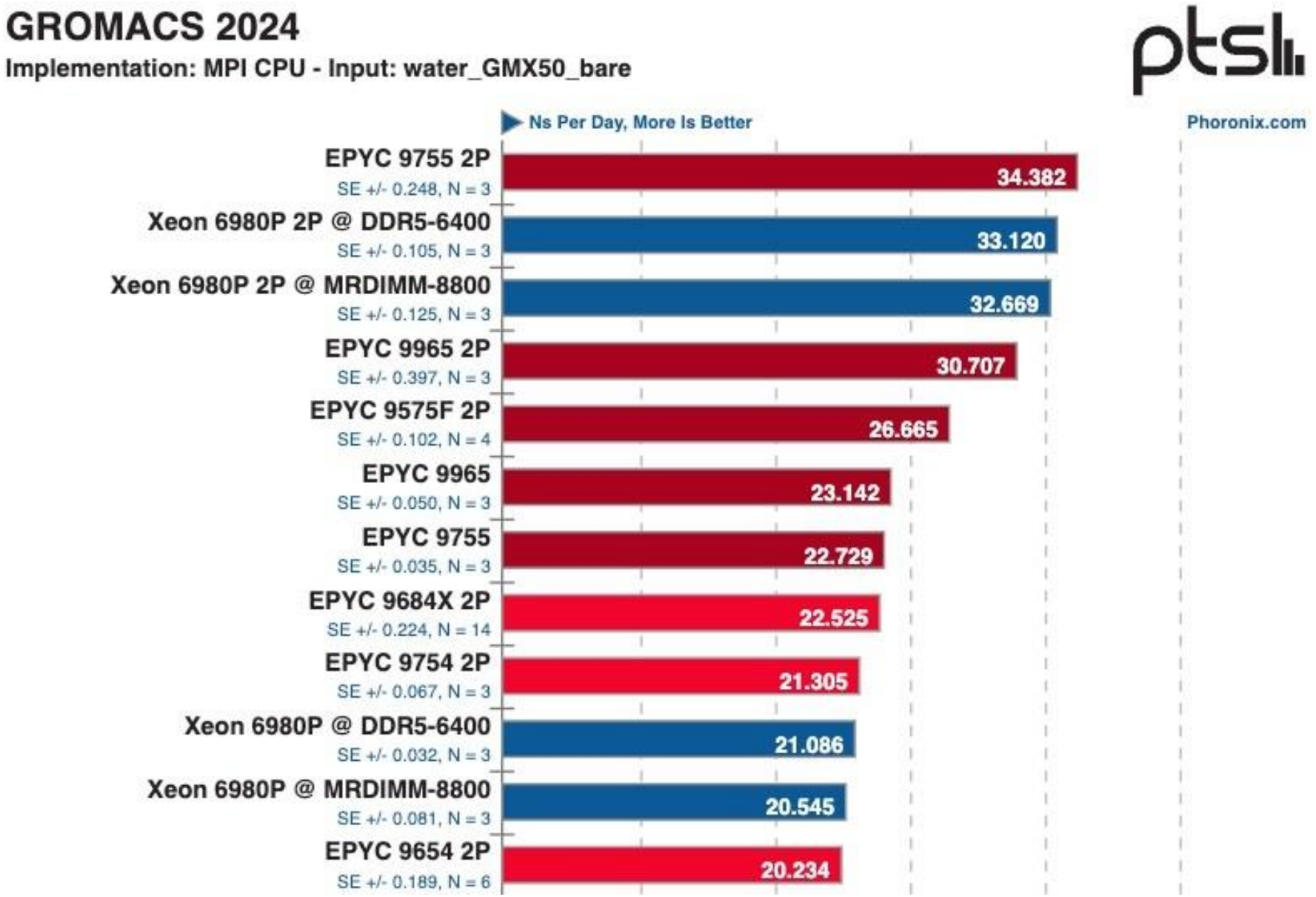
Процессоры Intel - роадмап

Процессоры Intel – новые технологии RibbonFET

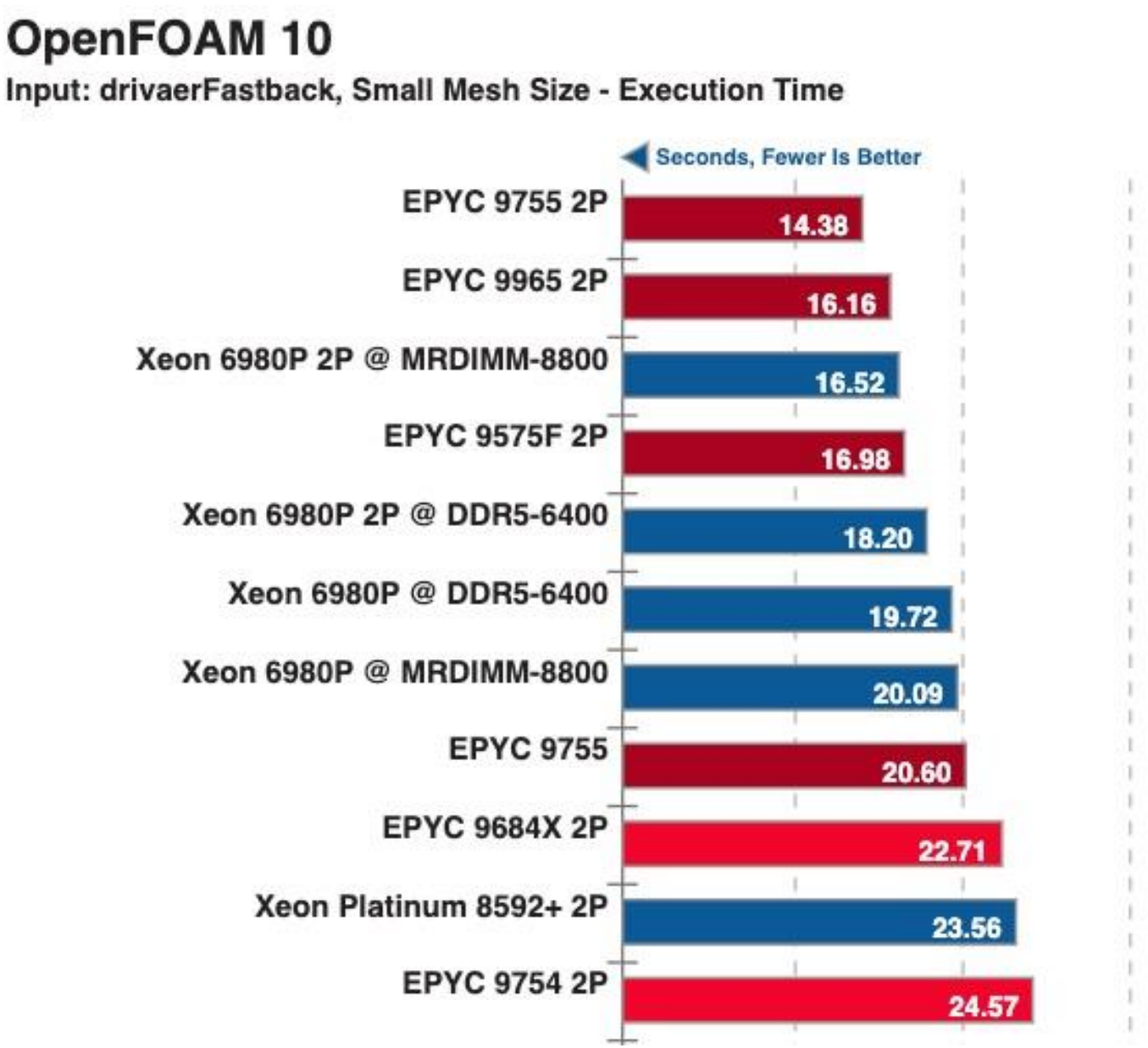


Сравнительные бенчмарки процессоров AMD Turin и Intel Xeon 6

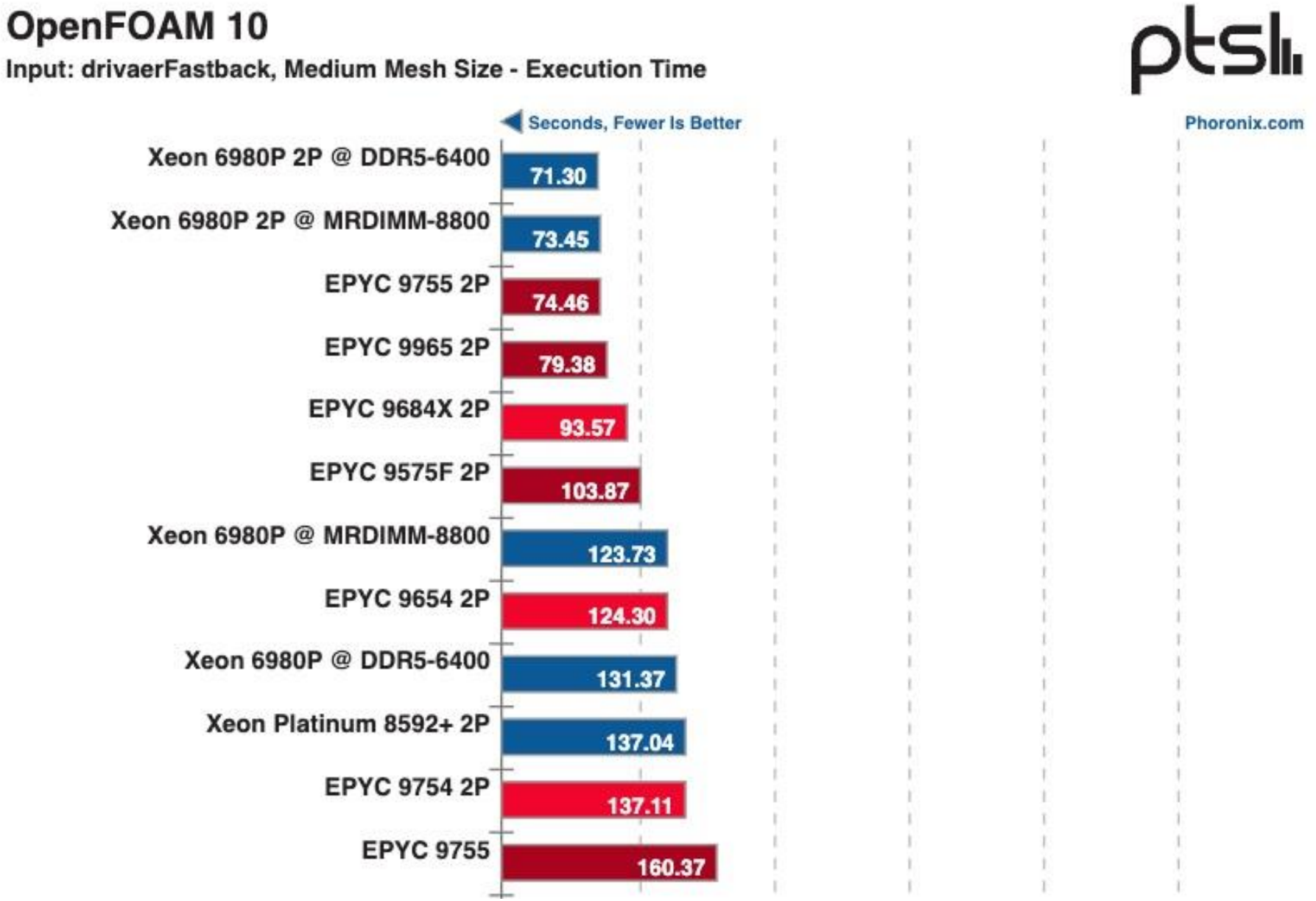
Молекулярная динамика



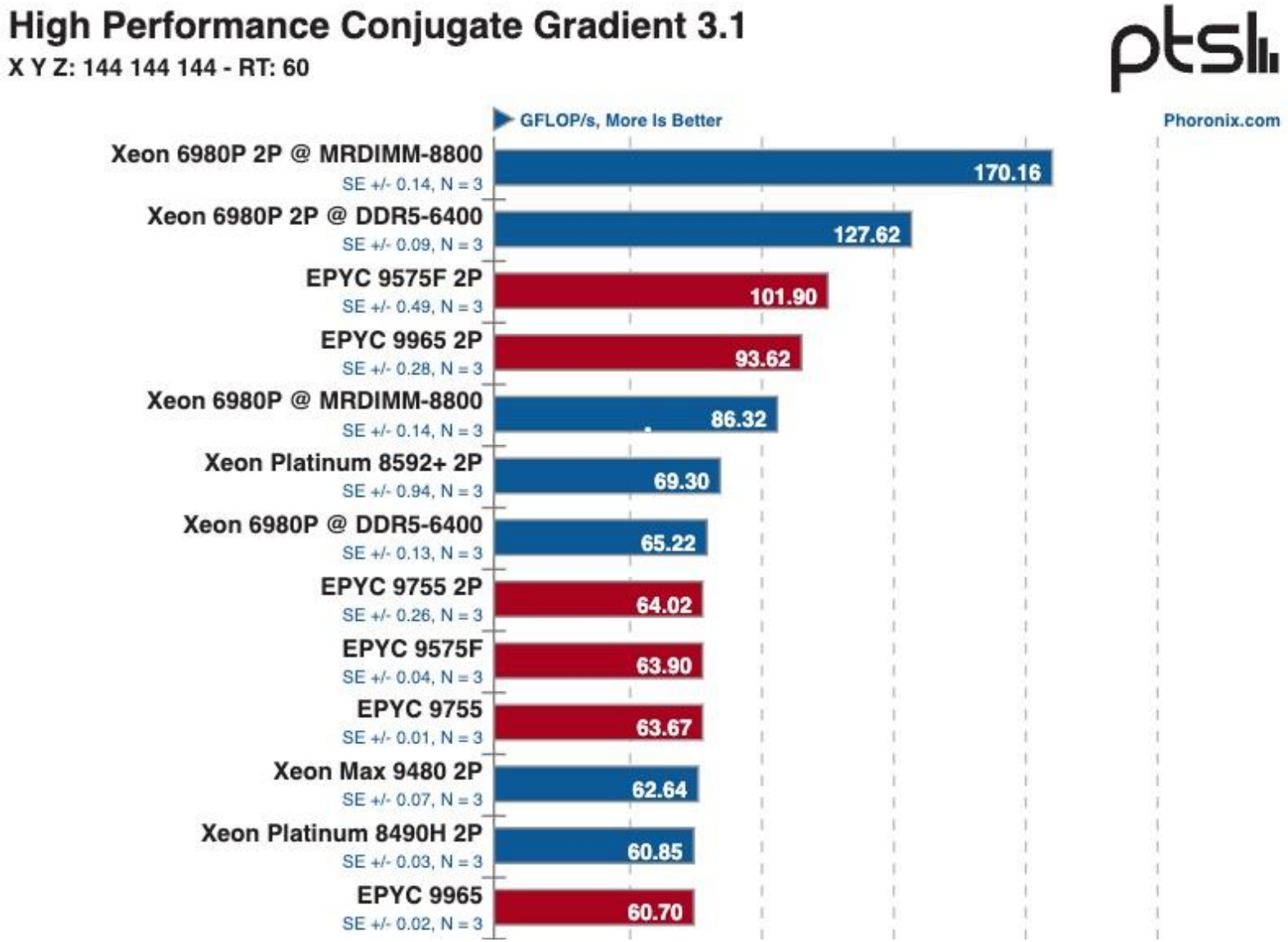
Computational Fluid Dynamics (CFD) – небольшая сетка



Computational Fluid Dynamics (CFD) – средняя сетка



HPCG benchmark – набор высокопроизводительных тестов



Компания Supermicro сегодня

Revenue	**$14B+** (FY2024 guidance) $7.1B (FY2023) $5.2B (FY2022)
Worldwide Presence	**6M+ Sq ft. Facilities Worldwide** 1. Silicon Valley (HQ), 2. Taiwan, 3. The Netherlands, 4. Malaysia and others
Production	**$18B/yr Production Capacity (CY23)** Top 5 Largest Server System Provider Worldwide (IDC & Gartner 2022), ~1.3M units annually
Human Resource in 4 Campuses	6000+ headcount Worldwide, ~50% Technical / R&D
Key Growth Matrix	**#1** in Generative AI and LLM Platforms 500%+ YoY Growth in Accel. Computing

GLOBAL PRESENCE

Production scale and cost optimization

Economy of Scale and Cost

- **Silicon Valley Green Computing Park B20-B23**
  - Rack-Scale Integration (Liquid Cooling)
  - Command Center
  - B2B/C Programs
  - Cloud Services
- **APAC Science and Tech Center B62**
  - Increase 2X-3X APAC capacity in FY24, 25, 26
- **Supermicro Malaysia Campus**
  - High Volume Subsystem and Rack-Scale Production by Q3 2024

Future Site Plans

- Additional Silicon Valley locations
- Additional Netherlands facility
- Mexico, Texas sites (in plan)

Industry's Most Comprehensive Portfolio

Hyper-E and Hyper
Best-in-Class Performance and Flexibility Rackmount Servers

Ultra and Ultra-E
High Performance & Flexibility Rackmount Systems for Enterprise Applications

Cloud DC
All-in-one Rackmount Platforms for Cloud Data Centers

WIO (UP)
Industry's Widest Variety of I/O Optimized Servers

Mainstream
Versatile Entry Level and Volume Servers for Enterprise Applications

BigTwin®
Highly Modular Multi-Node Systems with Tool-less Design

TwinPro®
Cost-effective 2U Multi-node Platforms

FatTwin®
Advanced Multi-node 4U Twin Architecture with 8 or 4 Nodes

SuperBlade®
High Density x86 Multi-node Server for Enterprise Cloud, HPC

SuperWorkstations
Workstations for High Performance Workloads

MP 4-Way Server
Highest Performance and Flexibility for Enterprise Applications

PCIe GPU Servers
High Density Systems for Double-width, Full Length PCIe GPUs

HGX GPU Servers
High Performance and Flexibility with Advanced Architecture and Thermal Design

SuperStorage®
Top-loading Server Optimized for Field Serviceability and Field Replacement

IOT/Embedded
High-efficiency, High-performance Compact Form Factor for 5G and Edge computing

Типы систем Supermicro

Supermicro X14 Systems Launching Now

Optimized for **Intel® Xeon® 6700-series processors with E-cores**

Rackmount

Range of systems optimized for flexibility and performance. Ideal for cloud-scale data center deployments

Hyper

CloudDC with DC-MHS

Multi-Node

High-density and efficiency resource-saving architectures with shared components

SuperBlade®

BigTwin®

GrandTwin®

Edge

Compact and short-depth form factors designed for maximum performance and efficiency at the intelligent edge

Compact Box

Hyper-E

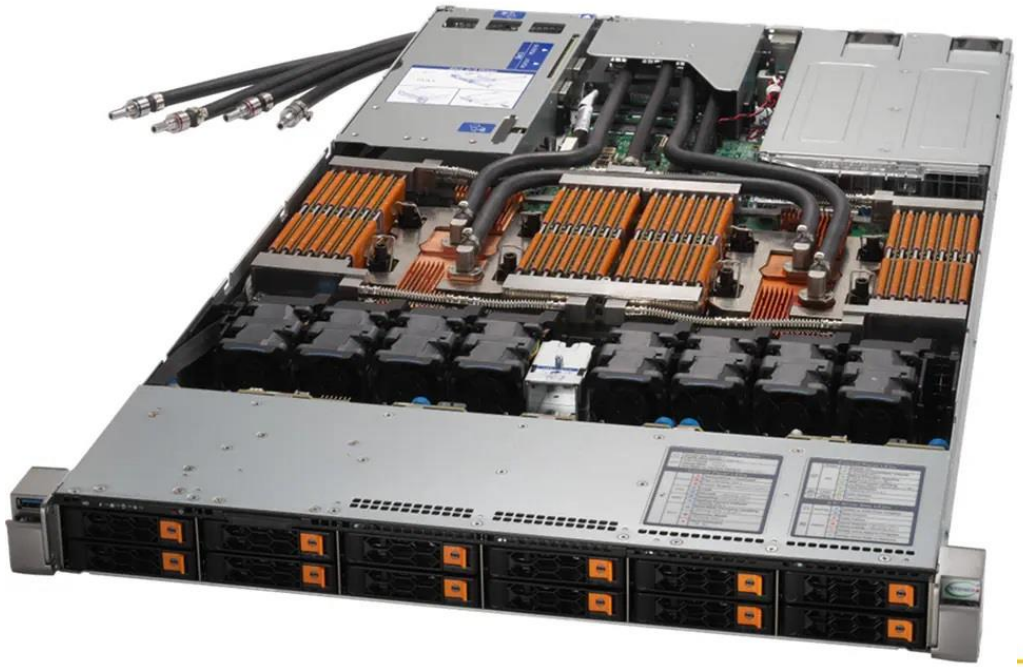
Telco/Edge

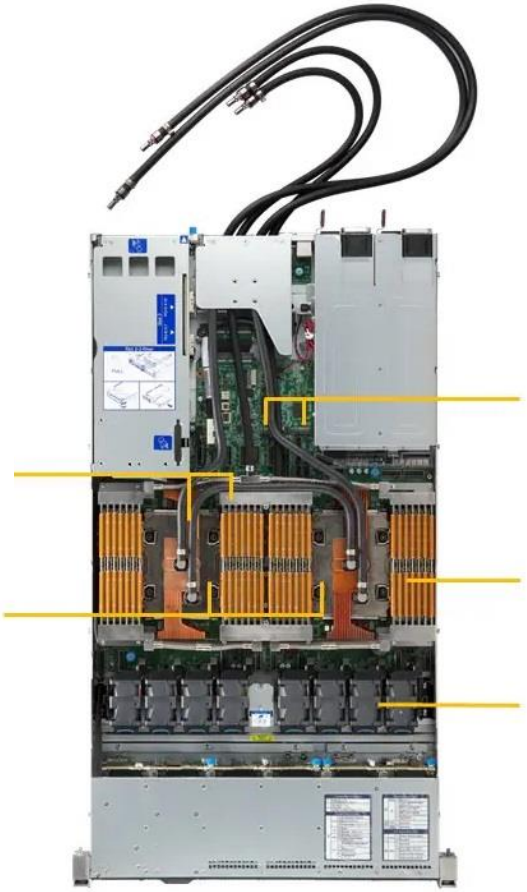
Классические системы 1U и 2U

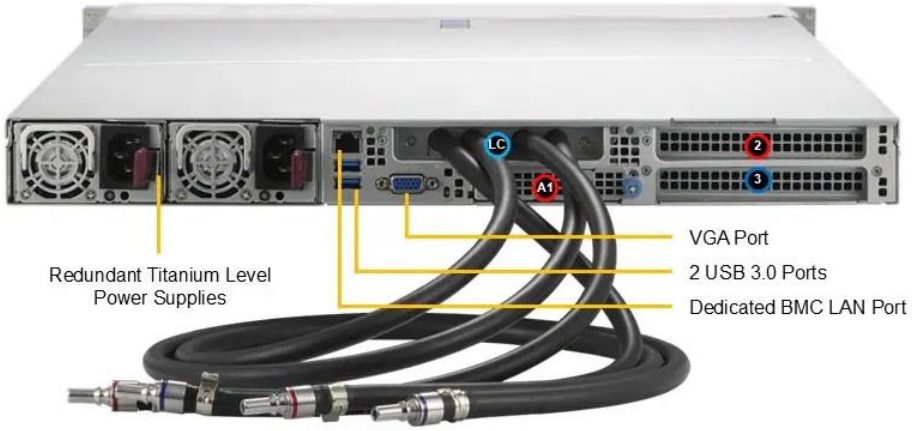
1U на процессорах:
Intel 6700 -> SYS-122H-TN
AMD 9004, 9005 -> AS-1125HS-TN

2U на процессорах:
Intel 6700 -> SYS-222H-TN
Intel 6900 -> SYS-222HA-TN
AMD 9004, 9005 -> AS-2025HS-TNR

Гибридная система 1U на Intel 6900







	Slot Description
2	PCIe 5.0 x16 Slot (FH, 10.5"L)
3	PCIe 5.0 x16 Slot (FH, 10.5"L)
A1	AIOM/OCP 3.0 NIC Slot
LC	Liquid Cooling Tubes

CPU1 CPU2 Liquid Cooling

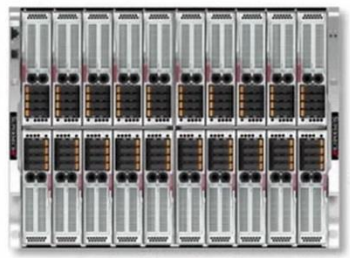
SYS-122HA-TN-LCC

X14 For High-Density Cloud

Powered by Intel® Xeon® 6 6700-series Processors with E-cores

- **Cloud Computing**
- **Content Delivery Networks**
- **Scale-out Object Storage**

Supermicro X14 Multi-node



SuperBlade®

Highest Density Multi-Node Architecture for Cloud Applications

BigTwin®

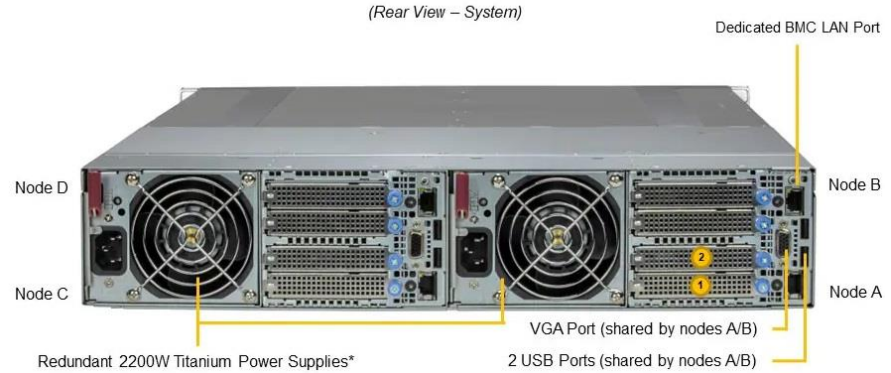
Multi-Node Architecture Optimized for Single-Processor Performance

GrandTwin®

Industry-leading Multi-node Architecture

1. **Maximum core density** up to 34,560 CPU cores per rack
2. **Shared power and cooling** for PUE as low as 1.05
3. **High throughput and density** E3.S NVMe storage (up to 32 drives in 2U)

4 x1P узла в 2U (2Twin2) на AMD 9004/9005 (Genoa, Turin)



Slot	Description
1	AIOM / OCP 3.0 PCIe 5.0 x16 Slot
2	AIOM / OCP 3.0 PCIe 5.0 x16 Slot

AS-2115GT-HNTR

Drive Bay (Node A-D)	Description
0 - 5	6x 2.5" Hot-swap NVMe/SATA Drive Bays

Блейд-системы 8U высокой плотности

SBI-422B-1NE14

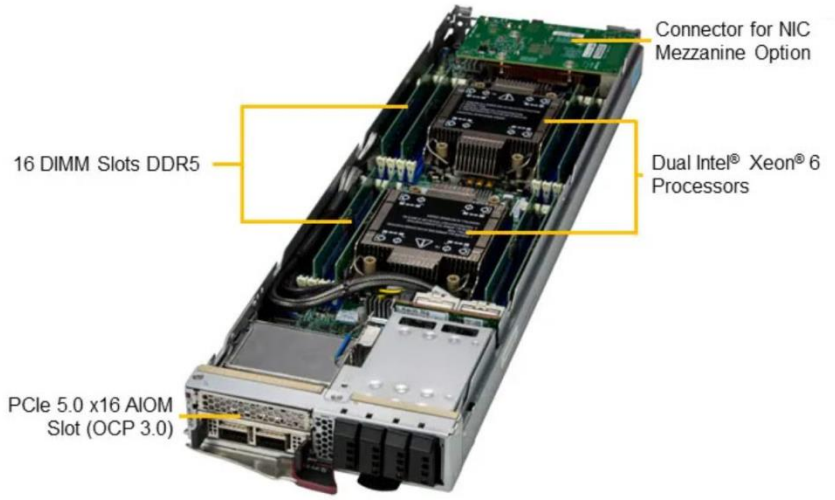
100 servers per 42U Rack. Dual Socket E2 (LGA-4710), Intel® Xeon® 6 processors

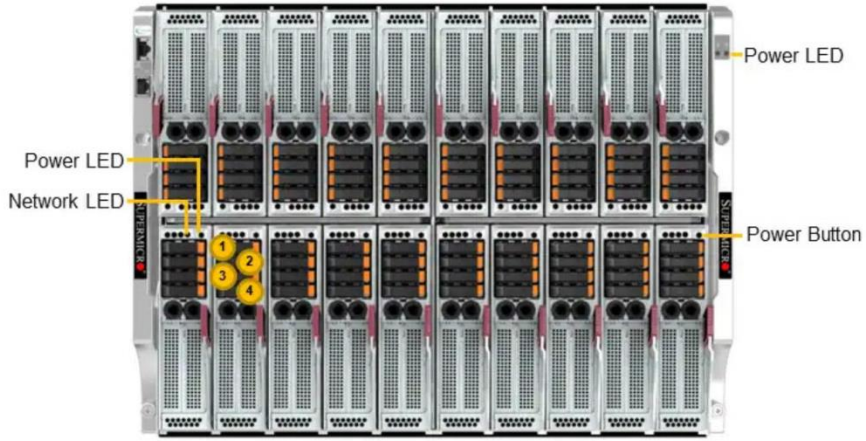
- 100 servers per 42U Rack
- Dual Socket E2 (LGA-4710), Intel® Xeon® 6 processors
- Up to 16 DIMMs, 1DPC with DDR5-6400MT/s ECC RDIMM/RDIMM 3DS
- Support up to 9 drives per blade
- Dual-port 25G Ethernet (LOM)
- Expansion slot for IB or Ethenret mezzanine card and for OCP 3.0 compliant card
- Liquid cooling solution is available

Key Applications

- Enterprise data center
- EDA, Cloud
- High-performance Computing (HPC)

Product Specification

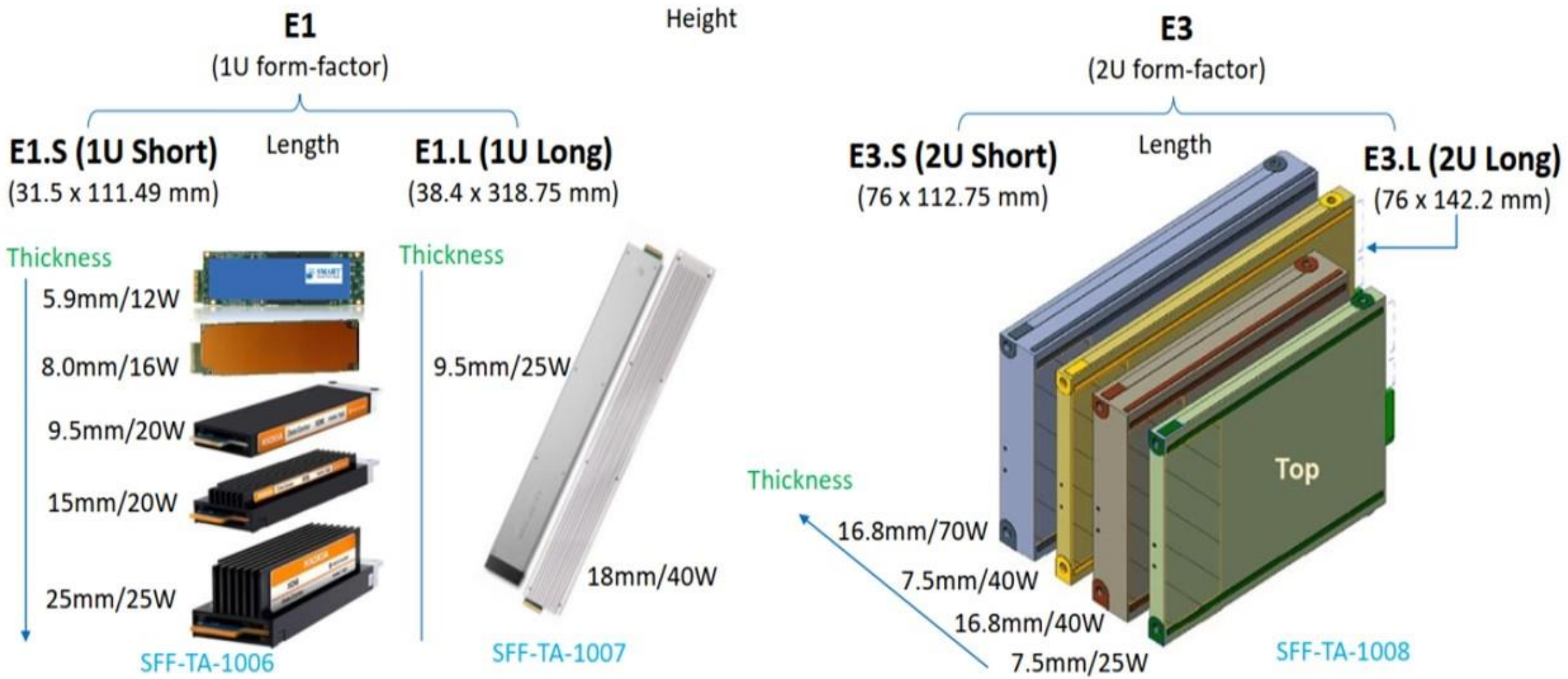




Drive Bay	Description
1 – 4	4 Hot-swap E1.S NVMe Drive Bays

8U 1 Node | 2 CPUs | 8 DIMMs | 4 Drives | 2x 25Gb | AIOM | Redundant

Современные топовые SSD диски для СХД



Современные топовые SSD диски для СХД

Современный топовые SSD диски для СХД

Solidigm™ D5-P5336 overview

Features

Product Name	Solidigm D5-P5336			
Media	192L QLC NAND (171GiB)			
Power off Retention	3 months @ 40°C			
Indirection Unit	16KB			
User Capacity	7.68TB	15.36TB	30.72TB	61.44TB
Endurance (5-yr DWPD)[1]	0.42	0.51	0.56	0.58
Endurance (PBW)[1]	5.9	14.1	31.5	65.2
Max Power	25 W			
Idle Power	<5 W			
UBER	< 1 Sector per 10^{17} Bits Read			
MTBF	2 Million Hours			
Features	OCP 2.0 support[2], NVMe 1.4 Compliance[3], FIPS 140-3 Level 2			

U.2 (15mm)

E3.S (7.5mm)

E1.L (9.5mm)

New Spec Alignment

NVMe 1.4c [3] (Target NVMe 2.0 support for PRQ2) and OCP 2.0 Support[2] (Latency monitor, FW history log, NSSR, Format Progress Indicator, NUSE specific etc.)

PCIe 4.0 Controller

Delivers better latency, expanded management capabilities, and critical new NVMe features compared to previous gen QLC SSD.

Data Center SSD features

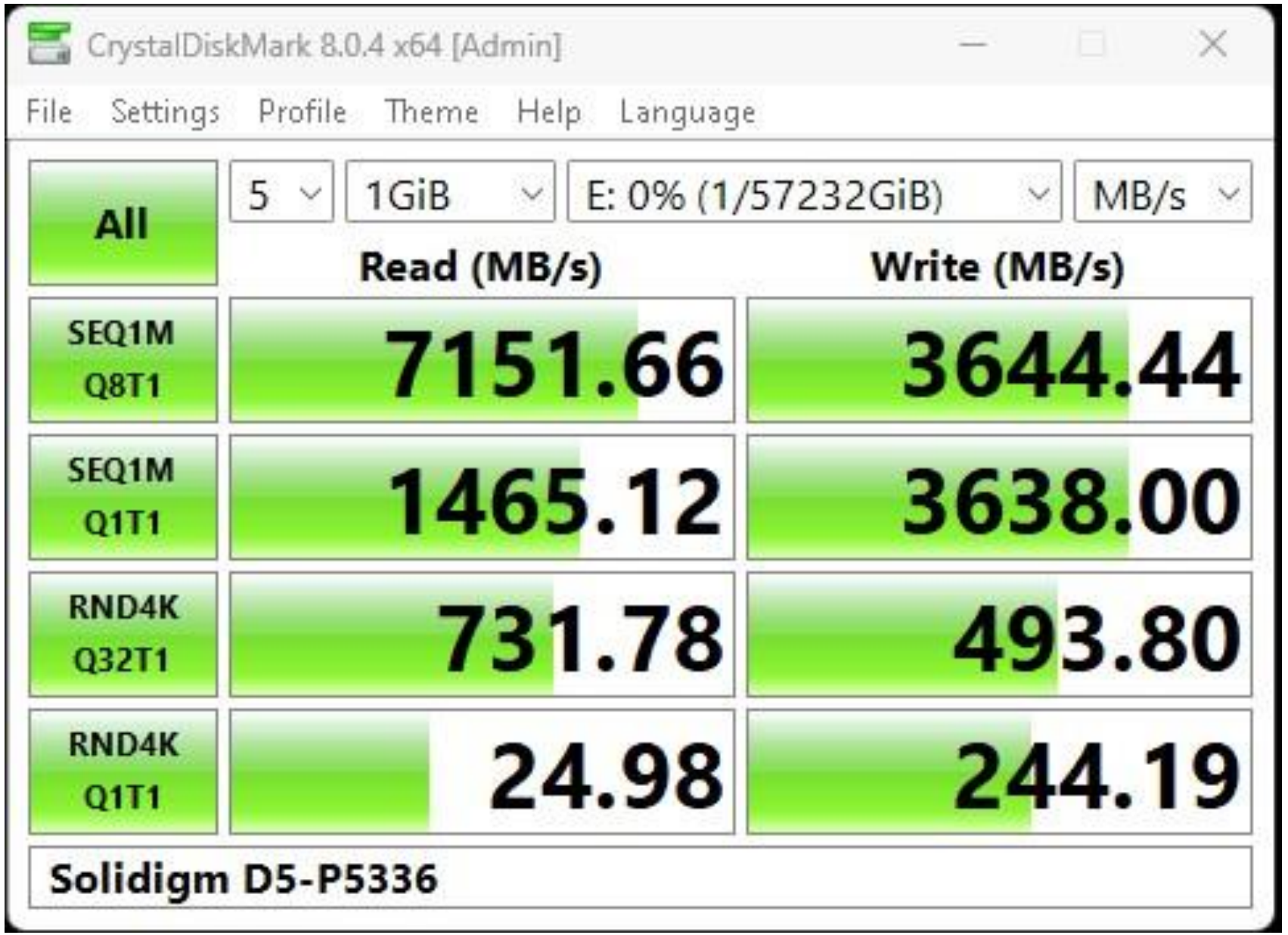
New Trim arch & improvements, new OCP 2.0 feature support, SGL, VSS, HDR with Opal, SGL with DSM, Opal Locking Range with MNS, Extended DST, FIPS 140-3 Level 2 (Future certification on generic SKUs), Telemetry, etc.

Performance

4K Random Read, IOPS, QD256	up to 1.005M
16K Random Write, IOPS, QD256	up to 43K
128K Seq. Read, MB/s, QD128	up to 7,000
128K Seq. Write, MB/s, QD128	up to 3,300

1. IU-Aligned Endurance. Based on 100% Random Write 16KB workload calculation.
2. See Solidigm D5-P5336 product specification for exceptions and modifications for compliance/support details
3. PRQ1 SKUs compliant with NVMe version 1.4 and NVMe MI 1.1. NVMe 2.0 and NVMe-MI 1.2 support to be targeted in PRQ2 and subsequent releases.

Современные топовые SSD диски для СХД



Современные топовые SSD диски для СХД

Современные топовые SSD диски для СХД

Solidigm™ D7-PS1010 and D7-PS1030 series detail

Features

Product Name	Solidigm D7-PS1010				Solidigm D7-PS1030			
Interface	PCIe 5.0							
Media	176L TLC 3D NAND (Charge Trap)							
User Capacity (TB)	1.92	3.84	7.68	15.36	1.6	3.2	6.4	12.8
Endurance Rating	Standard Endurance (SE)				Mid Endurance (ME)			
Endurance (5-yr)	1.0 DWPD				3.0 DWPD			
Max Lifetime PBW (5-yr)	28 PBW @ 15.36TB				70 PBW @ 12.8TB			
Max Avg Active Read & Write Power	23W (PCIe 5.0 and 4.0)							
Idle Power	5W (EU Lot 9-compliant)							
MTBF	↑ 2.5 Million Hours **(25% higher)**[1]							
UBER	↑ Tested to 1E-18 **(10x higher)**[1]							

Performance[1]

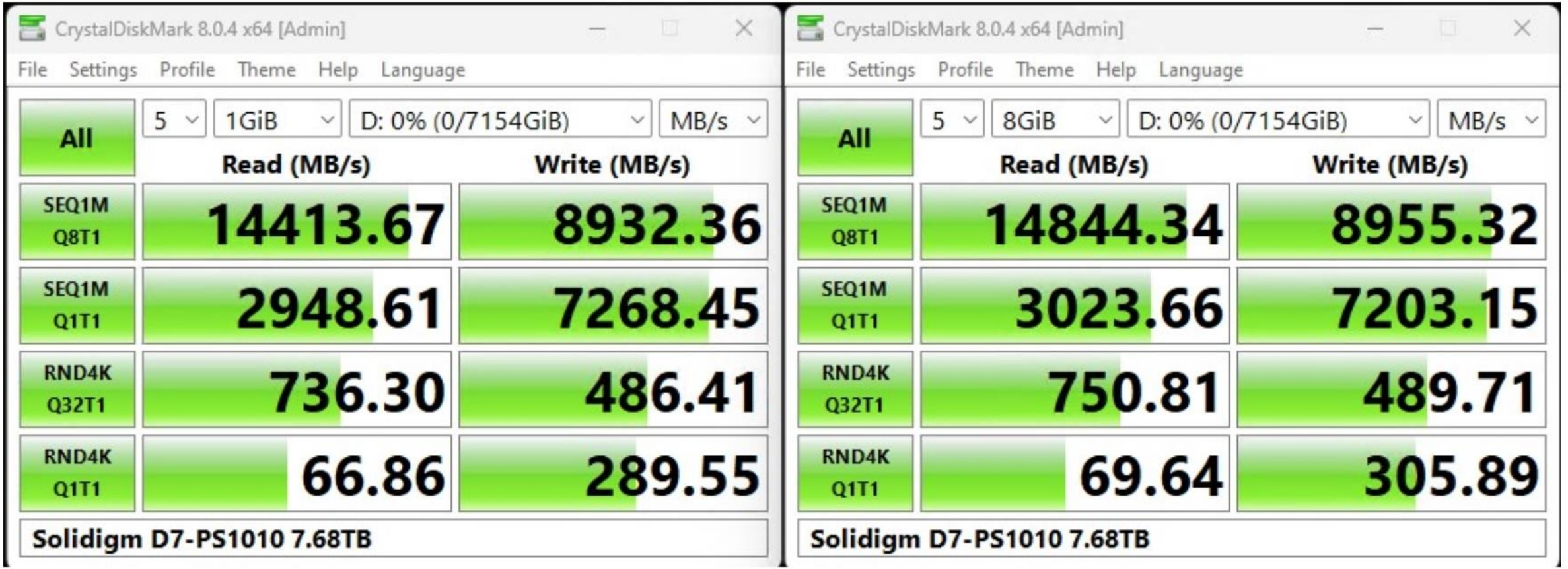
4K Random Read IOPS, QD256	↑ **2.8x** up to 3.1M	↑ **2.8x** up to 3.1M
4K Random Write IOPS, QD256	↑ **1.8x** up to 400K	↑ **2.1x** up to 800K
128K Seq. Read MB/s, QD128	↑ **2.0x** up to 14,500	↑ **2.0x** up to 14,500
128K Seq. Write MB/s, QD128	↑ **2.2x** up to 9,300	↑ **2.2x** up to 9,300

E3.S
7.5mm

U.2
15mm

[1] As compared to previous generation Solidigm D7-P5520. See Solidigm D7-PS1010/PS1030 product brief for performance, exceptions and modifications for compliance/support details.

Современные топовые SSD диски для СХД

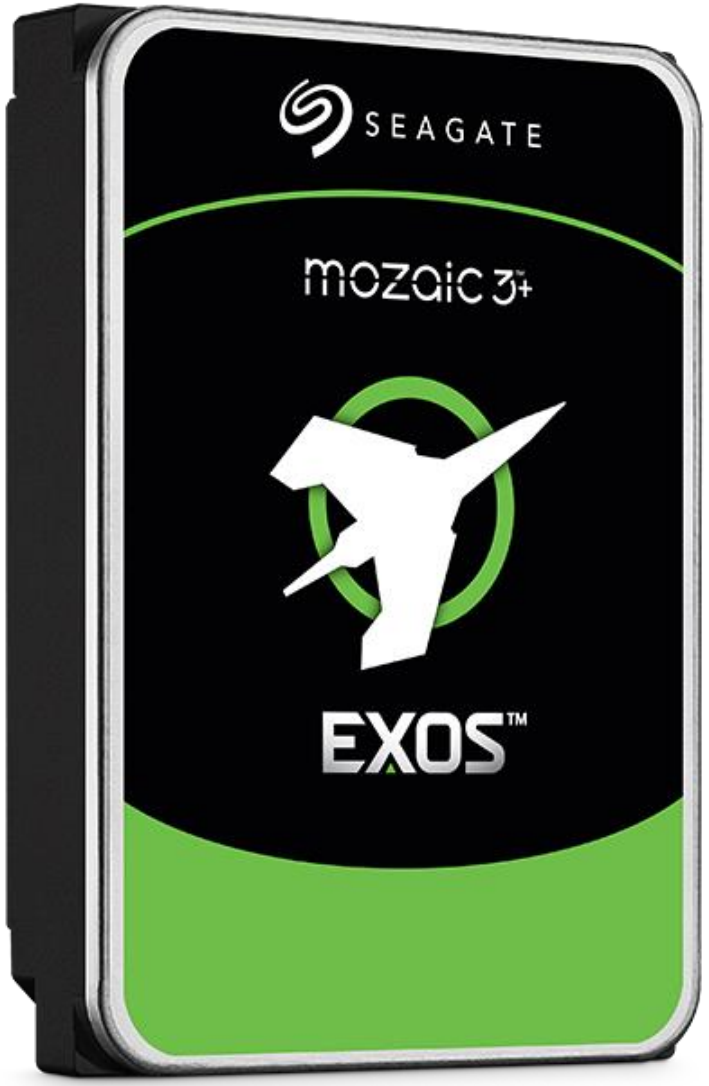


Solidigm D7 PS1010 7.68TB CrystalDiskMark 1GB And 8GB

Системы хранения на 60 дисков 3.5”

SSG-640SP-E1CR60

Современные топовые HDD диски для СХД



Шпиндельные диски 3.5” на 32ТБ:

- **WD Ultrastar DC HC690**
- **Segate Mozaic 3+**

Последние достижения сетевых технологий

NDR 400G INFINIBAND: NEXT-GENERATION MELLANOX INFINIBAND ARCHITECTURE

ADAPTER
NDR 400G InfiniBand
Programmable Datapath
In-Network Computing

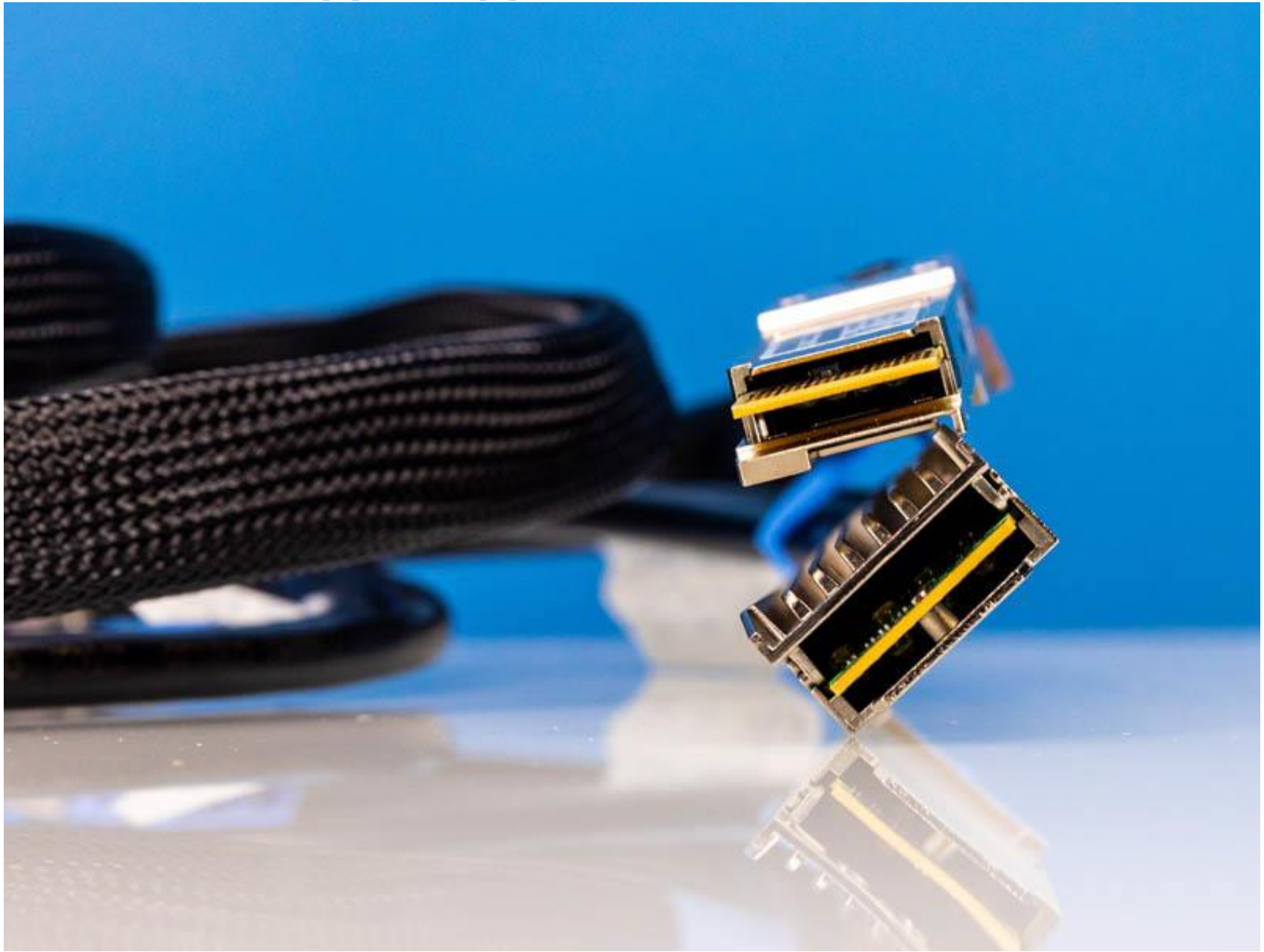
DPU
NDR 400G InfiniBand with Arm Cores
AI Application Accelerators
Programmable Datapath
In-Network Computing

SWITCH
64-ports NDR 400G InfiniBand
128-ports 200G NDR200
In-Network Computing

CABLE
Copper Cables
Active Copper Cables
Optical Transceivers

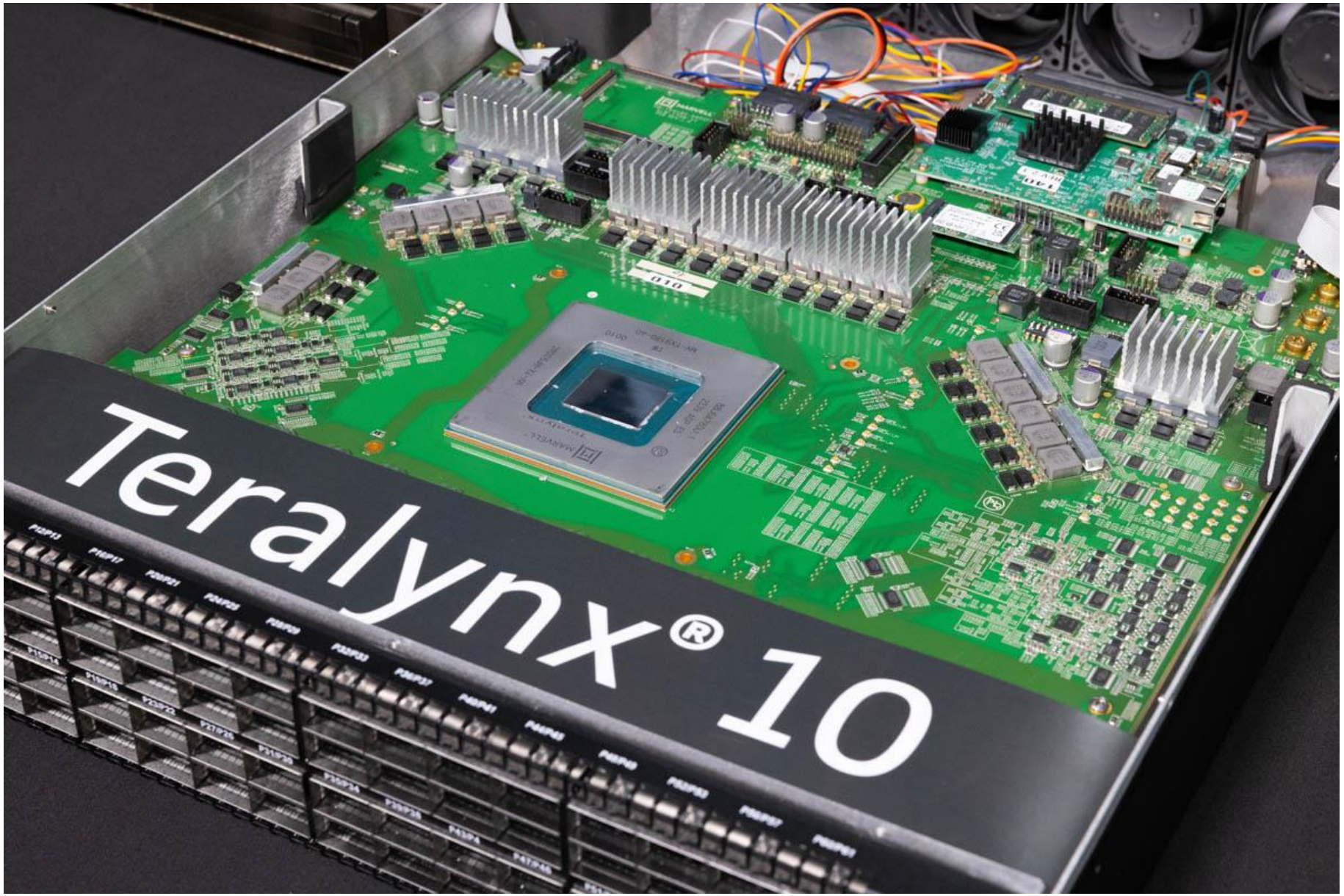
Последние достижения сетевых технологий

Последние достижения сетевых технлогий



OSFP-400
коннектор и кабель

Последние достижения сетевых технологий



Коммутатор
Marvell Terralynx 10
64 x 800 GbE
(или 512 x 100GbE)

Последние достижения сетевых технологий

Коммутатор
Marvell Terralynx 10
64 x 800 GbE
(или 512 x 100GbE)

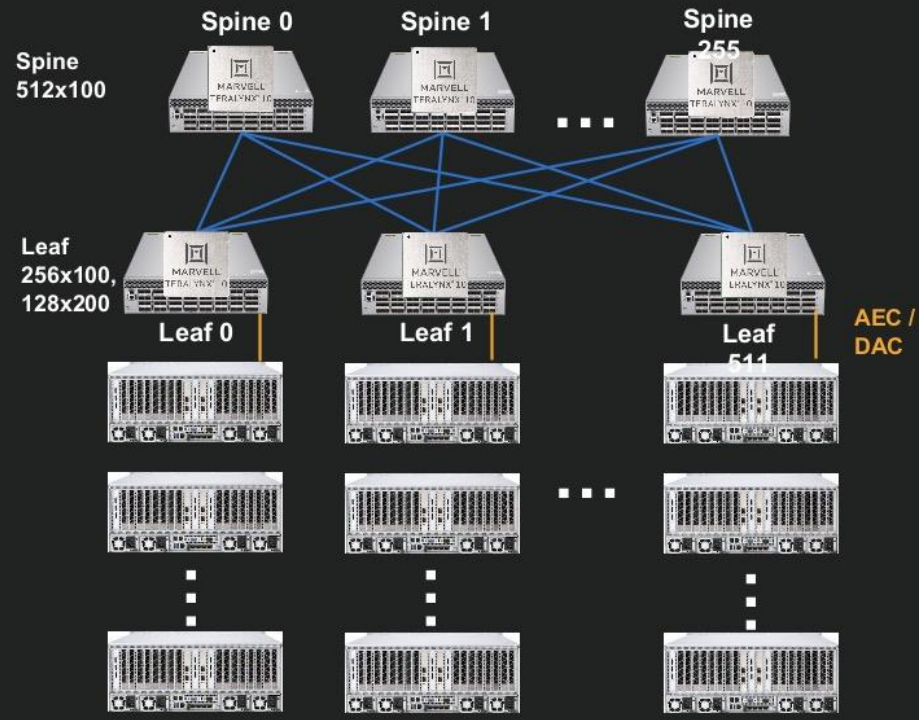
Последние достижения сетевых технологий

Коммутатор Marvell Terralynx 10 64 x 800 GbE (или 512 x 100GbE)

AI/ML with 512 radix

Power Estimation

Total Tiers	2
Total Nodes	64K
Total Switches	768
Total 800G Modules	128K
Total 800G AECs	16K
Total Estimated Power	2.09 MW



55% less power with Teralynx 10's 512 radix

Оригинальное жидкостное охлаждение от Supermicro

Supermicro Rack Integration Services provides a “one-stop-shop” for your data center needs

Optimized and Lab Tested Components for Superior Performance

Turn-Key Data Center

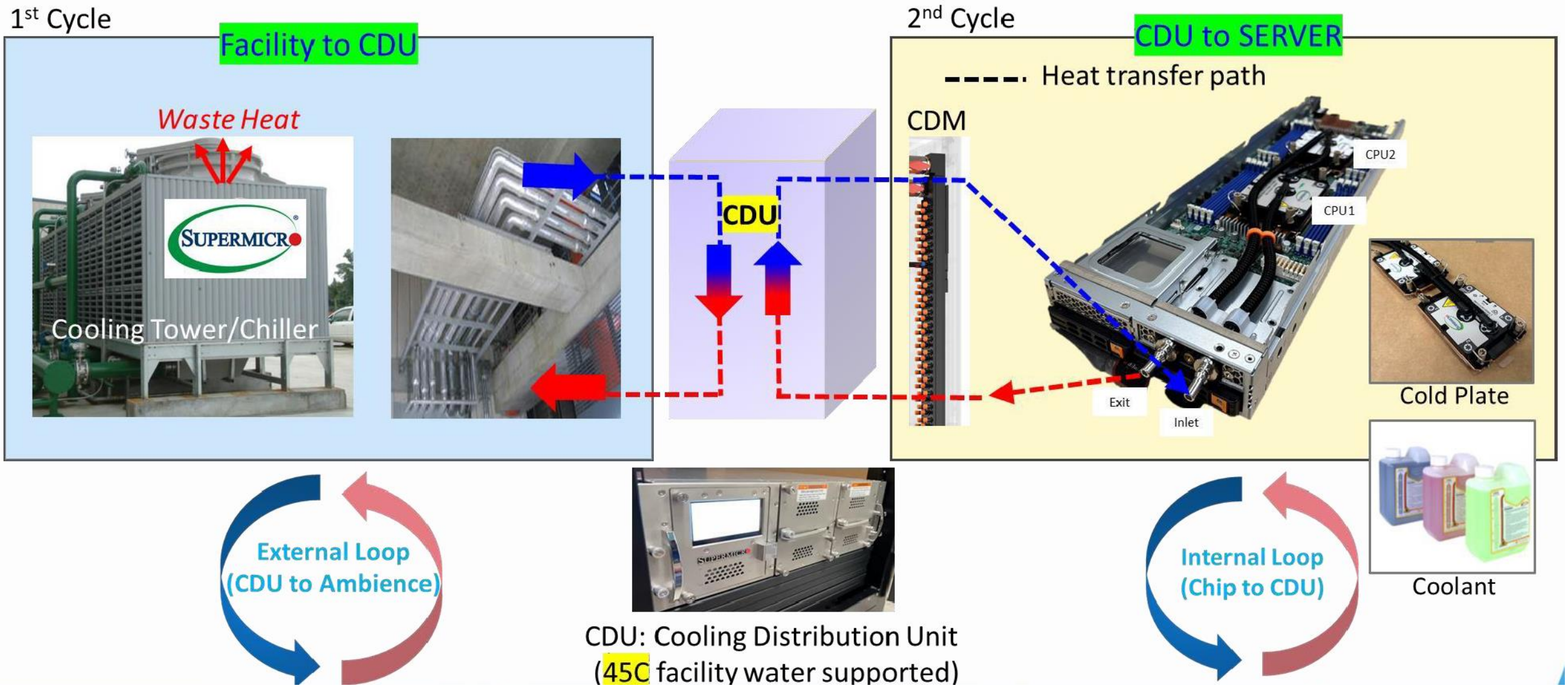
Accelerate Your Deployment

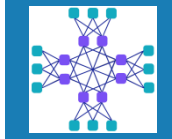
Professional Rack Level Design

Validation and Benchmarking

- Server
- Storage
- Network
- Software
- Cabling
- Power and Cooling
- Testing
- Benchmarking
- Full Rack Burn-in

Direct Liquid Cooling (DLC)





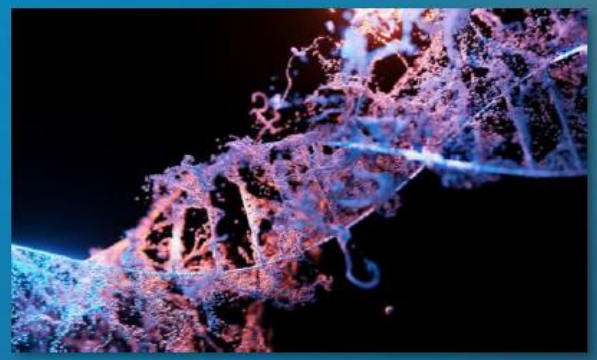
Системы с ускорителями GPU

Accelerate Everything

GPU Optimized Systems to Achieve 5X, 10X,… 100X Performance

Large Scale AI Training Workloads

Large language models, Generative AI training, autonomous driving, robotics



HPC/AI Workloads

Engineering simulation, scientific research, genomic sequencing, drug discovery

Enterprise AI Inference & Training

AI-enabled services/applications, chatbots, business automation

Visualization and Design

Graphical content development and automatic generation, digital twins, 3D collaboration

Content Delivery and Virtualization

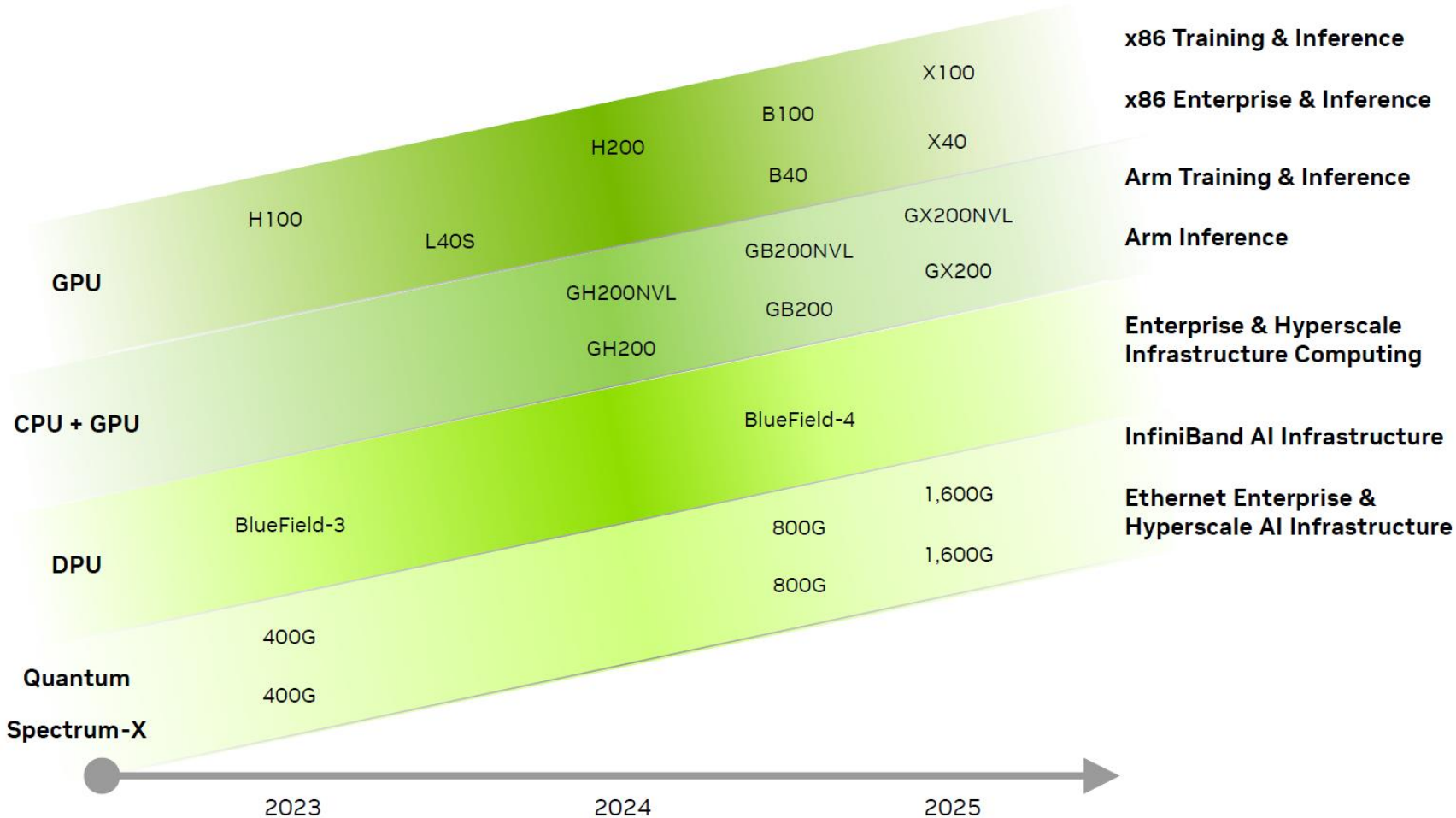
Content delivery networks (CDNs), video transcoding, live streaming, VDI

AI Edge

Retail automation, manufacturing/logistics automation, medical diagnosis/predictive care, security, and many more

AI - One Architecture | Train and Deploy Everywhere

One–Year Rhythm



NVIDIA

Supported By Supermicro

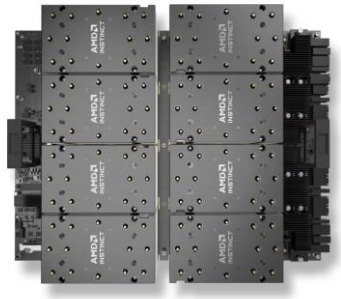
NVIDIA AMD intel

Multi socket

HGX
H100
H200
B100
B200

Coming Soon

CDNA3
MI300X

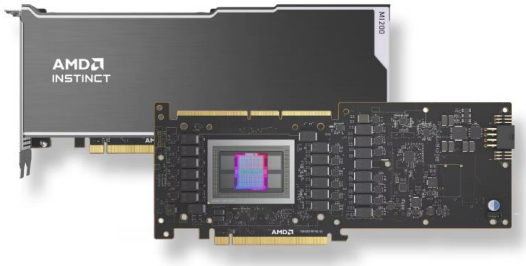


Gaudi3 UBB

PCIe

H100 NVL
L40S
L4

MI210



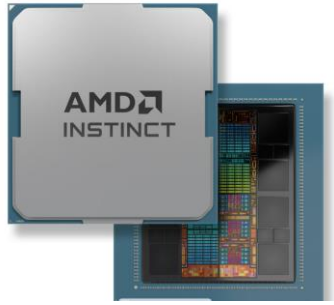
Gaudi3 PCIe

CPU+GPU

Grace Hooper
GH200
GB200

Coming Soon

CDNA3
MI300A



What GPU Fits The Best for Your Workload?

Manufacturer	GPU Model	Architecture	DL Training & DA	DL Inference	HPC / AI	Omniverse / Render Farms	AI Video	Far Edge Acceleration
NVIDIA	H200	*Multi Socket*	●	●	●			
NVIDIA	H100	*Multi Socket*	●	●	●			
AMD	MI300X	*Multi Socket*	●	●				
intel	GAUDI3	*Multi Socket*	●	●	●			
NVIDIA	H100NVL	*PCIe*	●	●				
NVIDIA	L40S	*PCIe*	●	●	●	●	●	
AMD	MI300A	*CPU+GPU*			●			
intel	GAUDI3	*PCIe*		●	●			
NVIDIA	L4	*PCIe*		●		●	●	●
NVIDIA	GH200	*CPU+GPU*	●	●	●			

● Best ● Better ● Good

Price-performance comparison relative across each entire workload column. This chart should be used in conjunction with measured data for targeted workloads.

NVIDIA GPU Platforms

Confidential

Highest Performance and Flexibility for AL/ML and HPC Applications

HGX Platforms

8U-8GPU SYS-821GE-TNHR

Integrated Performance, HGX H100 8-GPU

4U-8GPU SYS-421GE-TNHR2 LCC

Integrated Performance, HGX H100 8-GPU

5U/4U-4GPU SYS-521GU-TNXR

Scalable Performance, HGX H100 4-GPU

PCIe Gen5 Platforms

5U-10 GPU SYS-521GE-TNRT

Dual Root , PCIe GPU

4U-10 GPU SYS-421GE-TNRT

Dual Root, Direct Connect PCIe GPU

4U-4GPU SYS-741GE-TNRT

Flexible Solution, PCIe GPU

MGX Platforms

CG1, CG2, C2 Systems

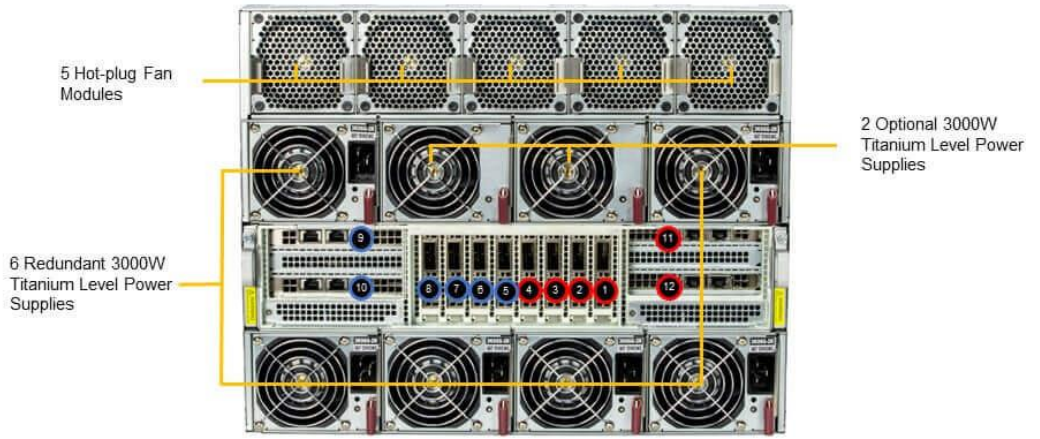
GPU-системы 8U на 8 карт GPU H100/H200 SMX

(Angled View – System)

16x 2.5" Hot-swap NVMe Drive Bays + 3x 2.5" Hot-swap SATA Drive Bays

На процессорах:
Intel -> SYS-821GE-TNHR
AMD -> AS-8125GS-TNHR

(Rear View – System)



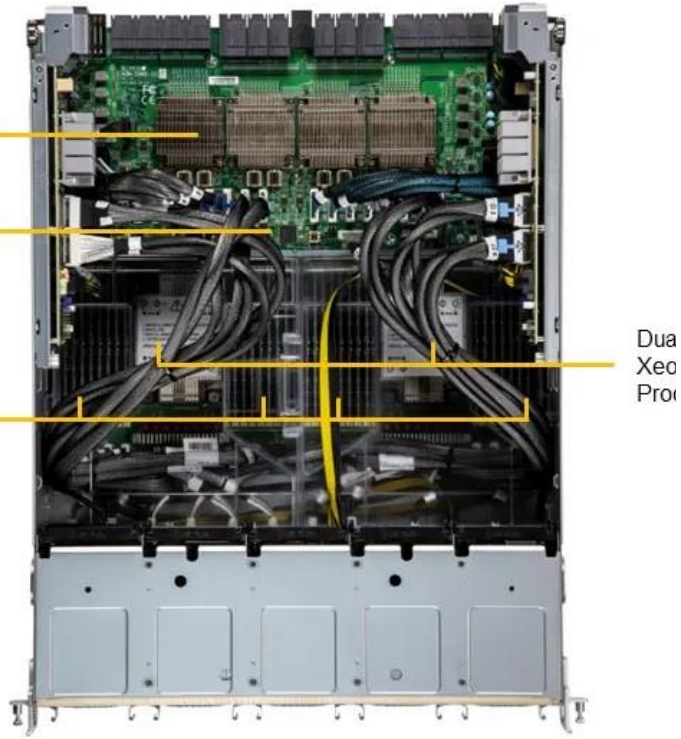
	Slot Description
1	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
2	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
3	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
4	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
5	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
6	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs

	Slot Description
7	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
8	PCIe 5.0 x16 (LP) from PLX switch linked to GPUs
9	PCIe 5.0 x16 (FHHL) via PLX switch (optional)
10	PCIe 5.0 x16 (FHHL) via PLX switch (optional)
11	PCIe 5.0 x16 (FHHL) via PLX switch
12	PCIe 5.0 x16 (FHHL) via PLX switch

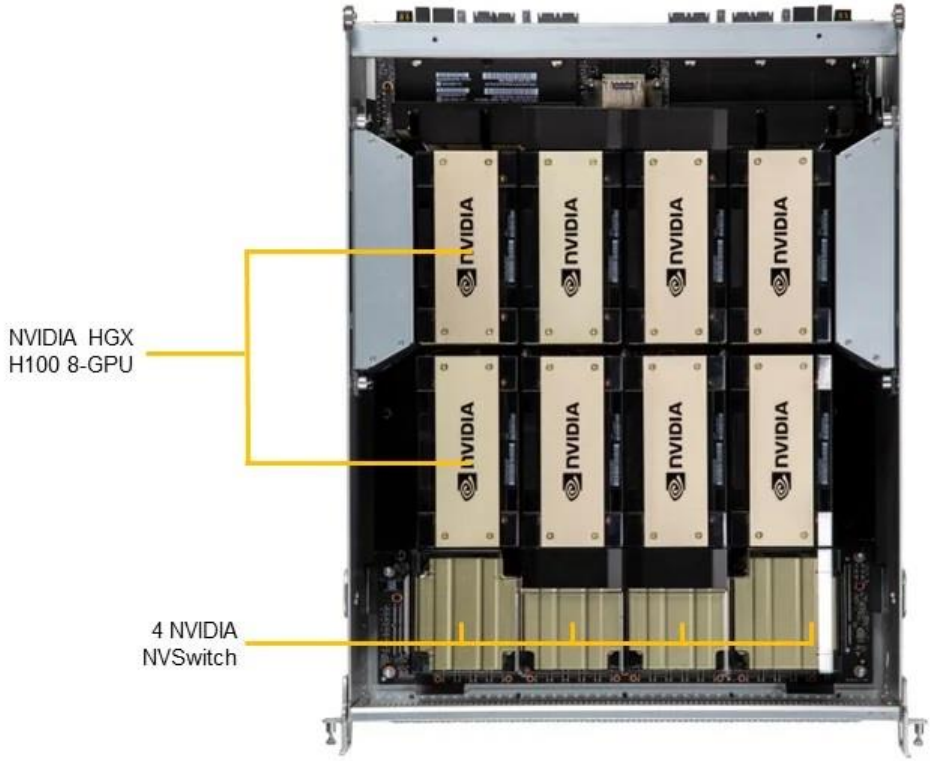
CPU1 CPU2

GPU-системы 8U на 8 карт GPU H100/H200 SMX

(Top View – System)



(Top View – System)



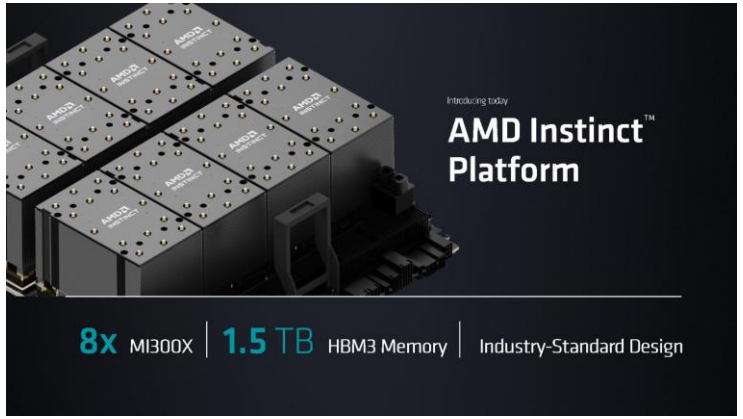
AS -8125GS-TNMR2

MI300X Solution

TITANIUM

Specifications

CPU – Dual EPYC 9004 Series processors, up to 256 cores/512 threads	**Memory –** 24x DIMM slots, ECC DDR5 supports up to 4800MHz
Drives – 2x onboard NVMe M.2 12x PCIe5.0 x4 NVMe U.2 (additional optional) 4x PCIe5.0 x4 NVMe U.2 2x SATA 2.5”	**Expansion –** 8x PCIe 5.0 x16 low profile 2x FHFL PCIe 5.0 x16 slots (Optional) 2x PCIe 5.0 x16 slots via additional PCIe switch
System Cooling 5x Front and 5x Rear counter-rotating fans with optimal fan speed control	**Power Supply –** 6x or 8x 3000-watt N+N Redundant Titanium Level Power Supplies
Accelerator Support - 8x MI300X	**Dimensions –** W x H x D 14.0” (H) x 17.2” (W) x 33.2” (D)



Recommend System Configuration:

Item#	Description	Q’ty
AS -8125GS-TNMR2	[NR]H13DSG-OM, CSE-GP801TS-D2 for MI300X	1
PSE-GEN9534-0799	Genoa 9534 DP/UP 64C/128T 2.45G 256M 280W SP5	2
MEM-DR564L-CL01-ER48	64GB DDR5-4800 2RX4 (10X4) LP (16Gb) ECC RDIMM	24
HDS-MMN-MTFDKBA960TFR-15	Micron 7450 PRO 960GB NVMe PCIe 4.0 M.2 22x80mm TCG Opal	2
AOC-CX766003N-SQ0	Nvidia 900-9X766-003N-SQ0 PCIe 1-port IB 400GE OSFP Gen5	8
GPU-AMD-MI300X-OAM-0045H	[NR] AMD Instinct MI300X 192GB 8 OAM + UBB	1
AOC-STGS-I2T-O	Std LP 2-port 10G RJ45, Intel X550 (Retail Pack)	1

SMCI Gaudi3 Product Spec

Confidential

Product Model : SYS-822GA-NGR3

Key Features

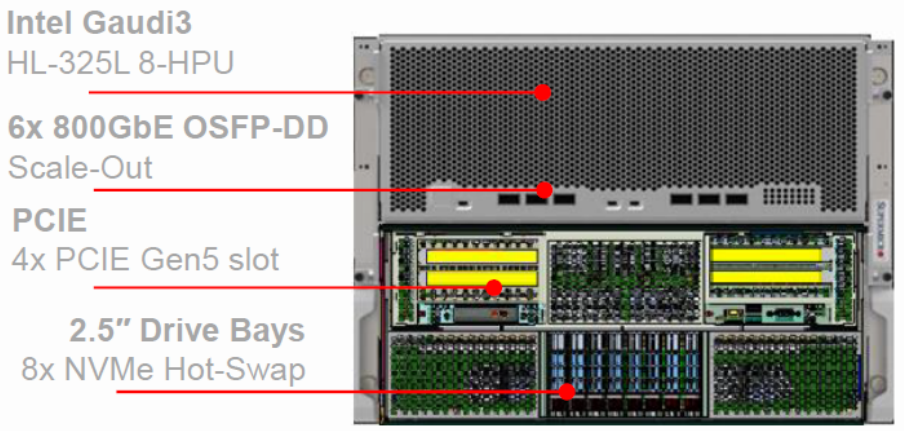
- Supports 8 Gaudi3 HL-325 OAM with 128GB HBM
- Direct connect on-board 800GbE OSFP-DD Scale-out
- Dual 6rd Gen Series Intel Xeon Scalable Processors

Key Application

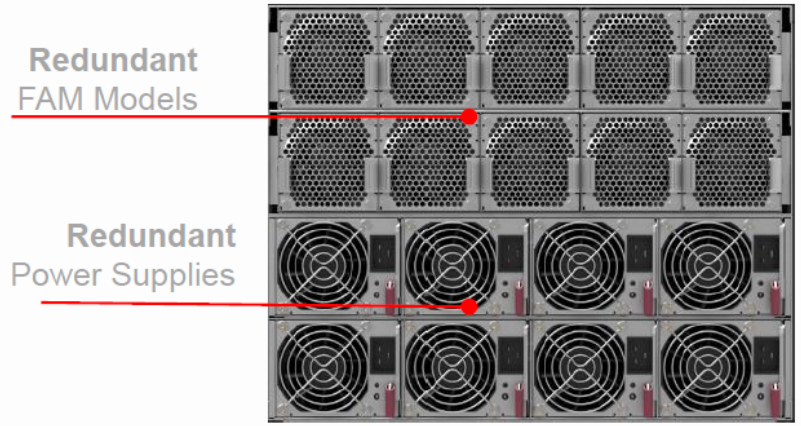
- AI Compute/Model Training/Deep Learning

Specification

CPU – Dual Socket Dual P-Cores Granite Rapids AP and E-Cores Clearwater Forest-AP processor up to 500W	**Memory –** 24 DIMM Slots 24x DIMM ECC DDR5 designed for up to 6400 MT/s (1DPC)
Drives – 8 Hot-Swap Bays 8x NVMe Gen 5 2x M.2 2280/22110	**Expansion –** 5 PCI-E Gen5 Slots 2xPCI-E 5.0 x16 (FHHL) 2xPCI-E 5.0 x 8 (FHHL) 1xAIOM with 10G LAN card (2x Intel i550)
High-Speed QSFP-DD– 800GbE 6x OSFP-DD ports 1x RJ45 1GbE IPMI	**Power Supply –** 4+4 Redundant Power 3000W Titanium Level



Front Side

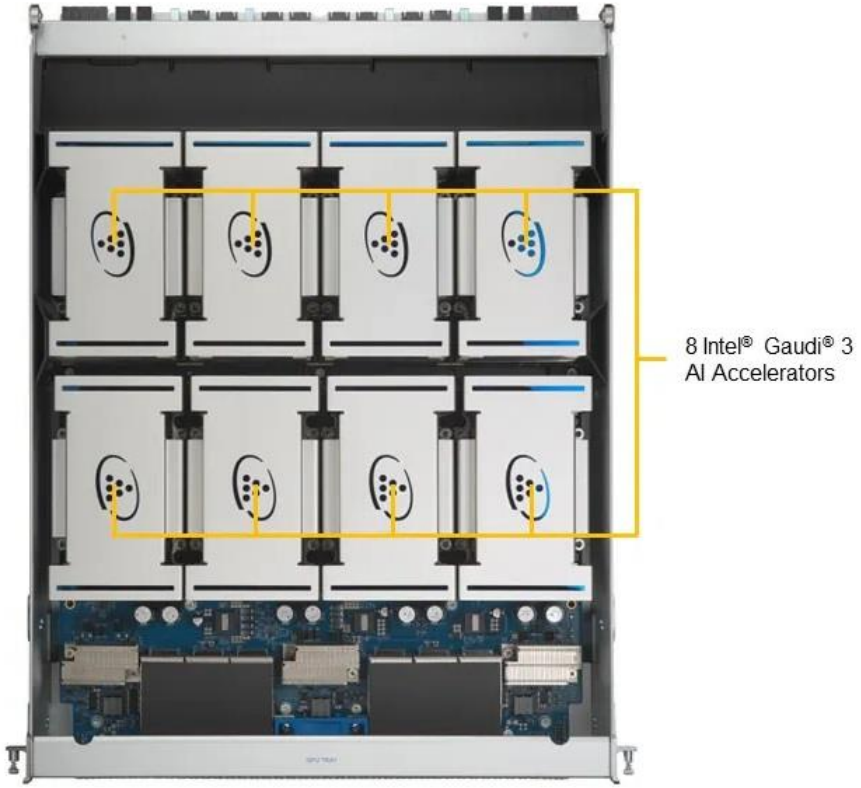


Rear Side

https://habana.ai/ , https://huggingface.co/hardware/habana

Система 8 x Intel Gaudi3 на процессорах Intel Xeon 6900

SYS-822GA-NGR3



Announcing NVIDIA H200 Tensor Core GPUs

Supercharging the Highest Performing Generative AI and HPC Platforms

Memory	Memory Bandwidth
141GB	4.8 TB/s
HBM3e	HBM3e

Llama 2 70B Inference	GPT-3 175B Inference	MILC HPC Simulation
1.9X	1.4X	110X
Performance vs H100	Performance vs H100	Performance vs x86 CPUs

8U-8GPU SYS-821GE-TNHR

Integrated Performance, HGX H100 8-GPU

4U-8GPU SYS-421GE-TNHR2 LCC

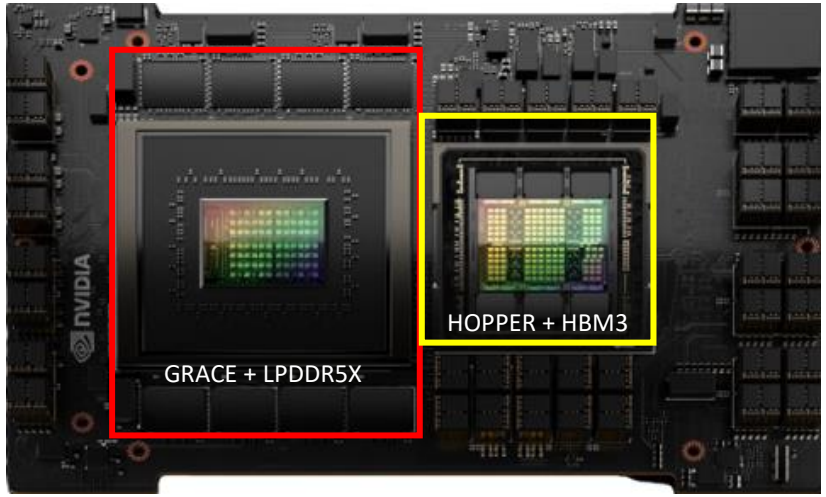
Integrated Performance, HGX H100 8-GPU

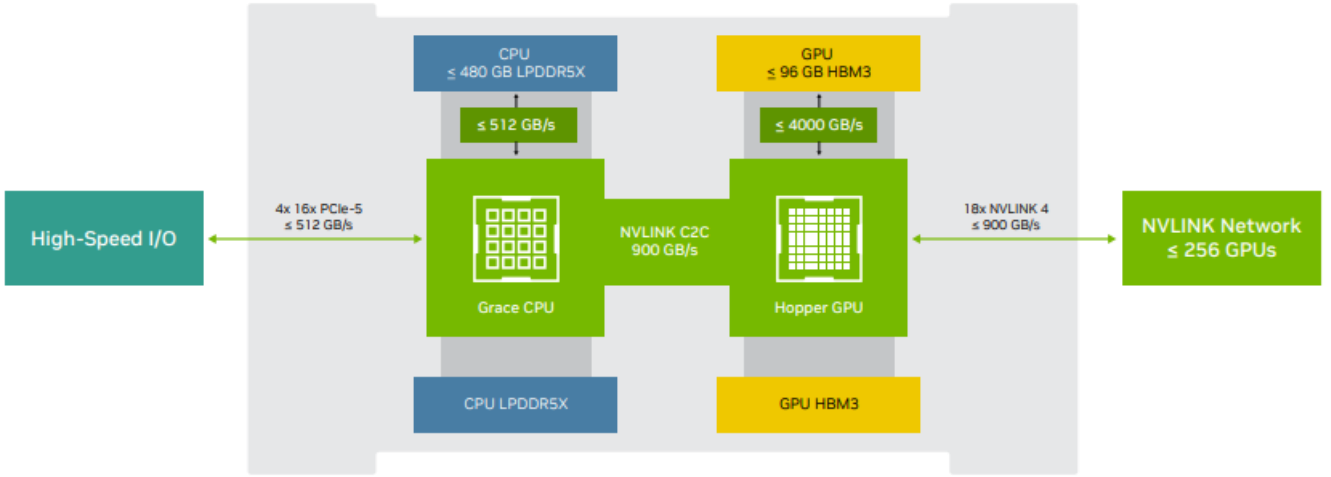
NVIDIA

NVIDIA GH200 Grace Hopper Superchip

- One Grace CPU with integrated LPDDR5X and one H100 Tensor Core GPU (Hopper) on mezzanine module
- Fast NVLink-C2C interface between CPU and GPU

Grace CPU	Feature
CPU core count	72 Arm Neoverse V2 cores
L1 cache	64KB i-cache + 64KB d-cache
L2 cache	1MB per core
L3 cache	117MB
LPDDR5X size	Up to 480GB
Memory bandwidth	Up to 512GB/s
PCIe links	Up to 4x PCIe x16 (Gen5)
Hopper H100 GPU	**Feature**
FP64	34 teraFLOPS
FP64 Tensor Core	67 teraFLOPS
FP32	67 teraFLOPS
TF32 Tensor Core	989 teraFLOPS* | 494 teraFLOPS
BFLOAT16 Tensor Core	1,979 teraFLOPS* | 990 teraFLOPS
FP16 Tensor Core	1,979 teraFLOPS* | 990 teraFLOPS
FP8 Tensor Core	3,958 teraFLOPS* | 1,979 teraFLOPS
INT8 Tensor Core	3,958 TOPS* | 1,979 TOPS
HBM3 size	Up to 96GB
Memory bandwidth	Up to 4TB/s
NVIDIA NVLink-C2C CPU-to-GPU bandwidth	900 GB/s bidirectional
Module thermal design power (TDP)	Programmable from 450W to 1000W (CPU + GPU + memory)
Form factor	Superchip module
Thermal solution	Air cooled or liquid cooled

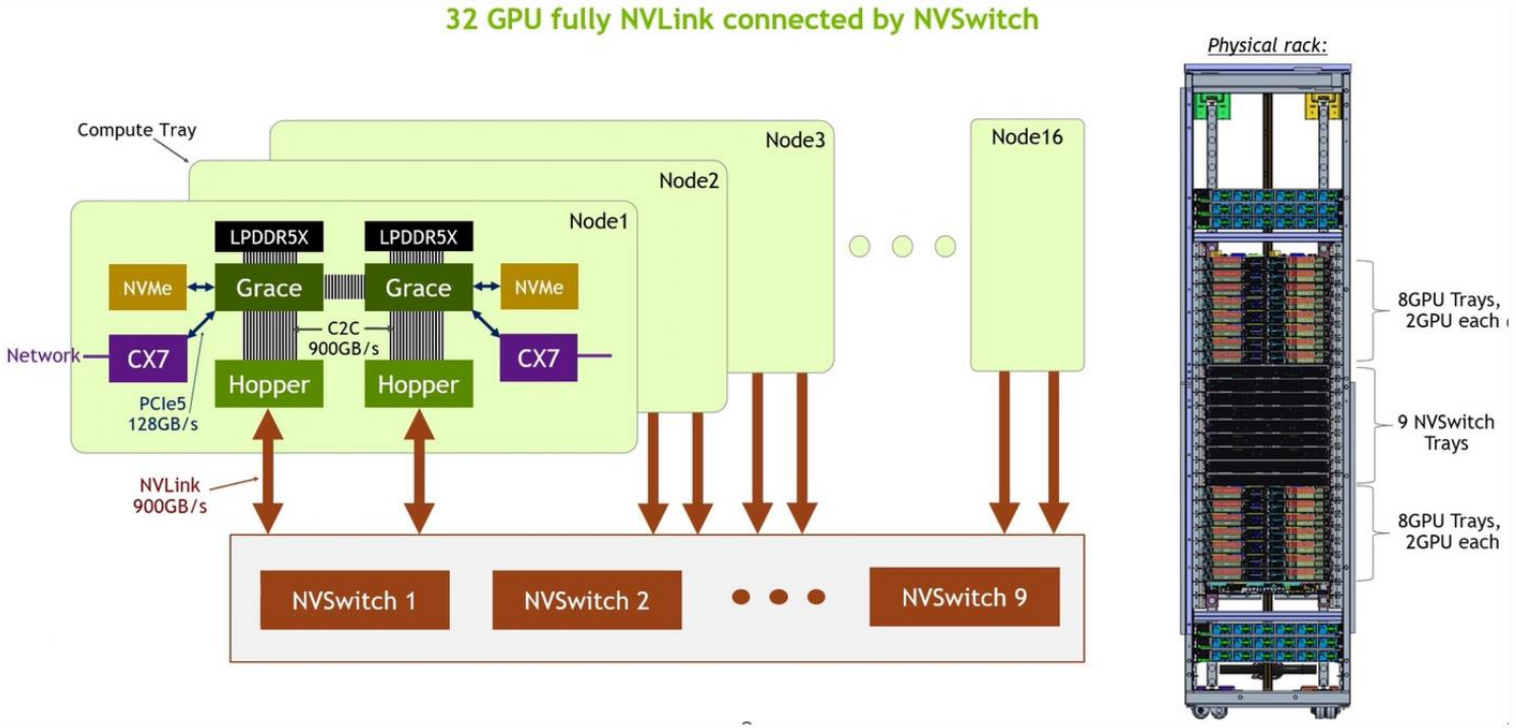




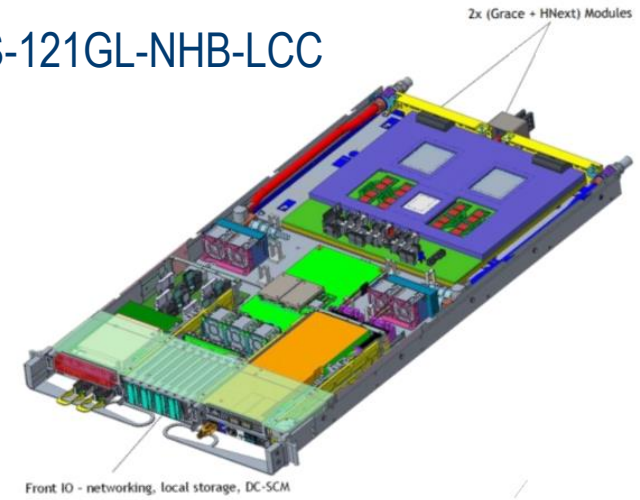
GH200 System And Rack Architecture

SMC Oberon or GH200NVL32

- NVLink Switches combine 256 GH200 superchips, allowing them to perform **as a single, GIANT GPU!**
- 1 Exaflop of performance
- 144 terabytes of shared memory



ARS-121GL-NHB-LCC



144TB GPU

Blackwell for Every Generative AI Use Case

Delivering the New Era of Performance for Every Data Center

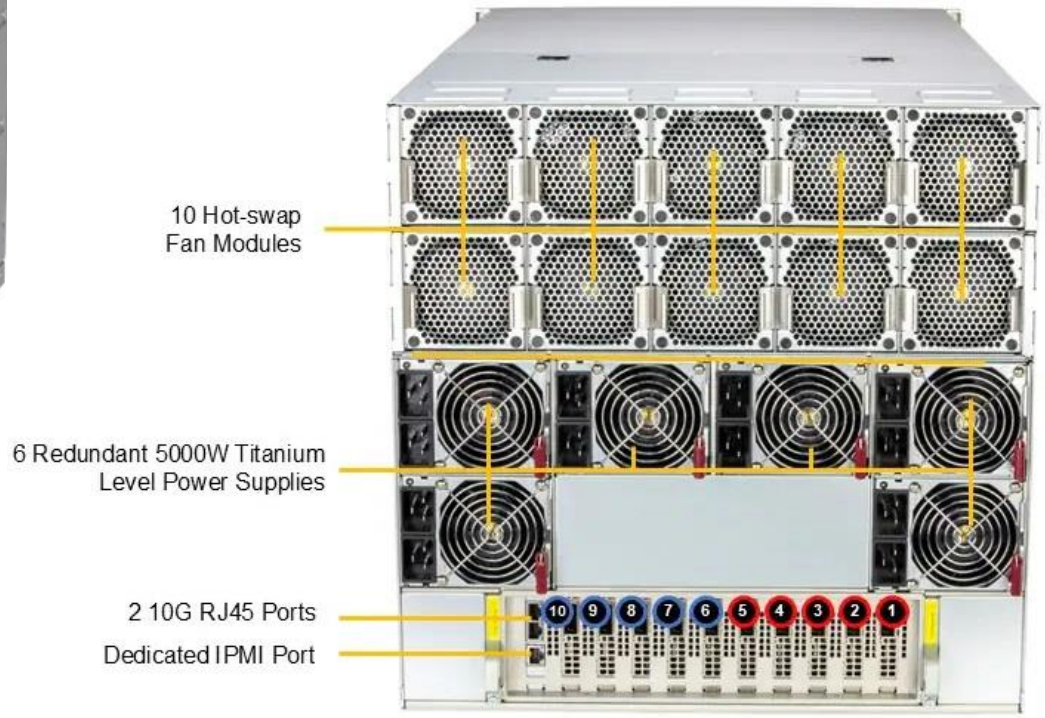
GB200 NVL72
Compute for Trillion Parameter Scale AI
Maximum Performance and Lowest TCO

HGX B200
Best Performance and TCO for HGX Platform

HGX B100
Drop-in Upgrade for Existing Hopper Infrastructure

Система 8 x B200 (180GB) на процессорах Intel Xeon 6900

SYS-A22GA-NBRT

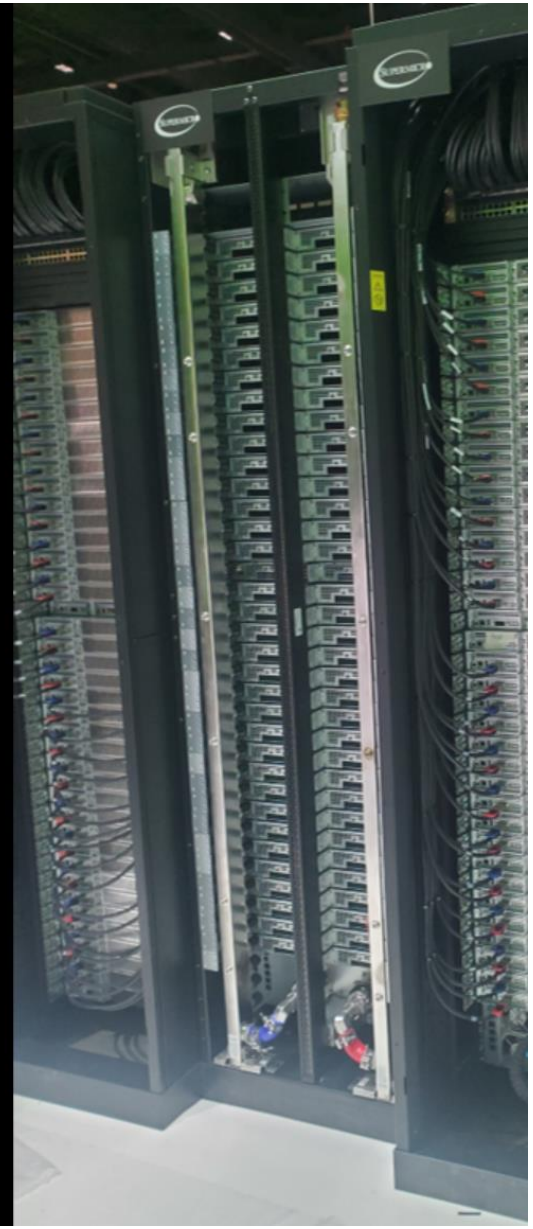


GB200 NVL72

SMCI Solution
as presented @ GTC 2024

GB200 NVL72
- 36 GRACE CPUs
- 72 BLACKWELL GPUs
- Fully Connected NVLink Switch Rack

Training	720 PFLOPs
Inference	1,440 PFLOPs
NVL Model Size	27T params
Multi-Node Bandwidth	130 TB/s
Multi-Node All-Reduce	260 TB/s



GB200 NVL72 Compute and Interconnect Nodes

Building Blocks for the GB200 NVL72 Rack

GB200 SUPERCHIP

40 PETAFLOPS FP4 AI INFERENCE
20 PETAFLOPS FP8 AI TRAINING
864GB FAST MEMORY

GB200 SUPERCHIP COMPUTE TRAY

2x GB200
80 PETAFLOPS FP4 AI INFERENCE
40 PETAFLOPS FP8 AI TRAINING
1728 GB FAST MEMORY
1U Liquid Cooled
18 Per Rack

NVLINK SWITCH TRAY

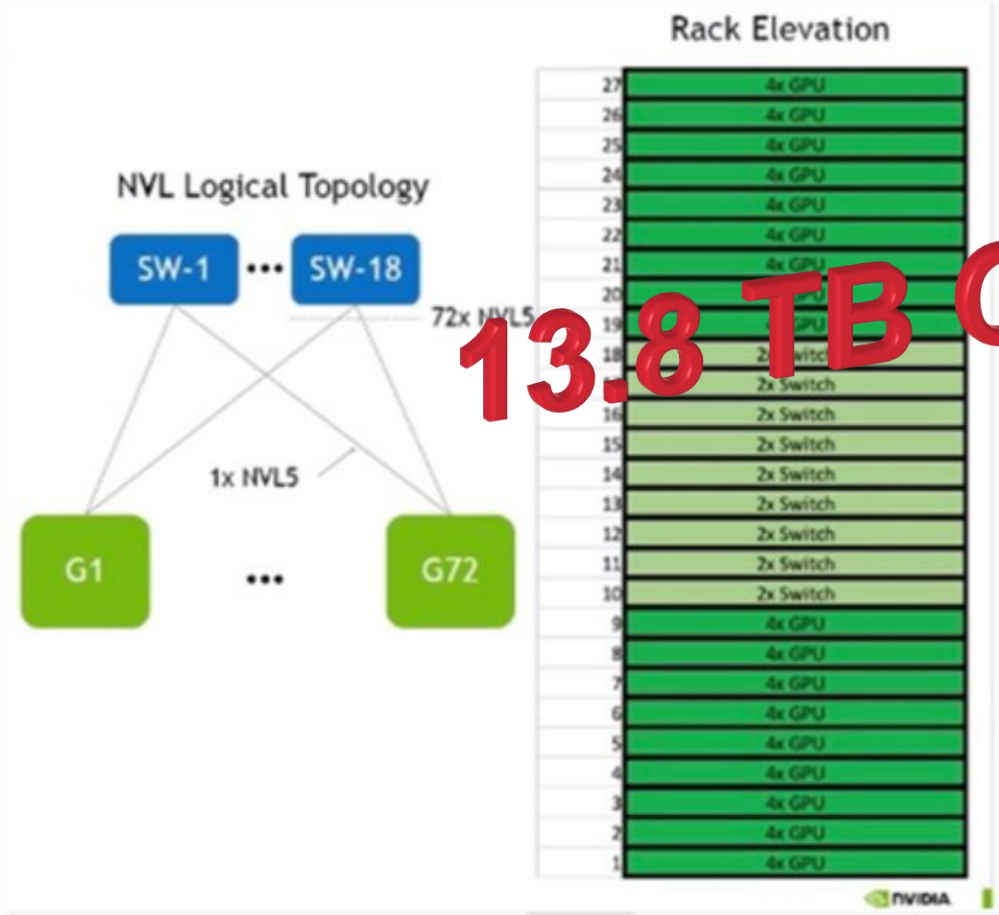
2x NVLINK SWITCH CHIP
14.4 TB/s Total Bandwidth
SHARPv4 FP64/32/16/8
1U Liquid Cooled
9 Per Rack

NVIDIA

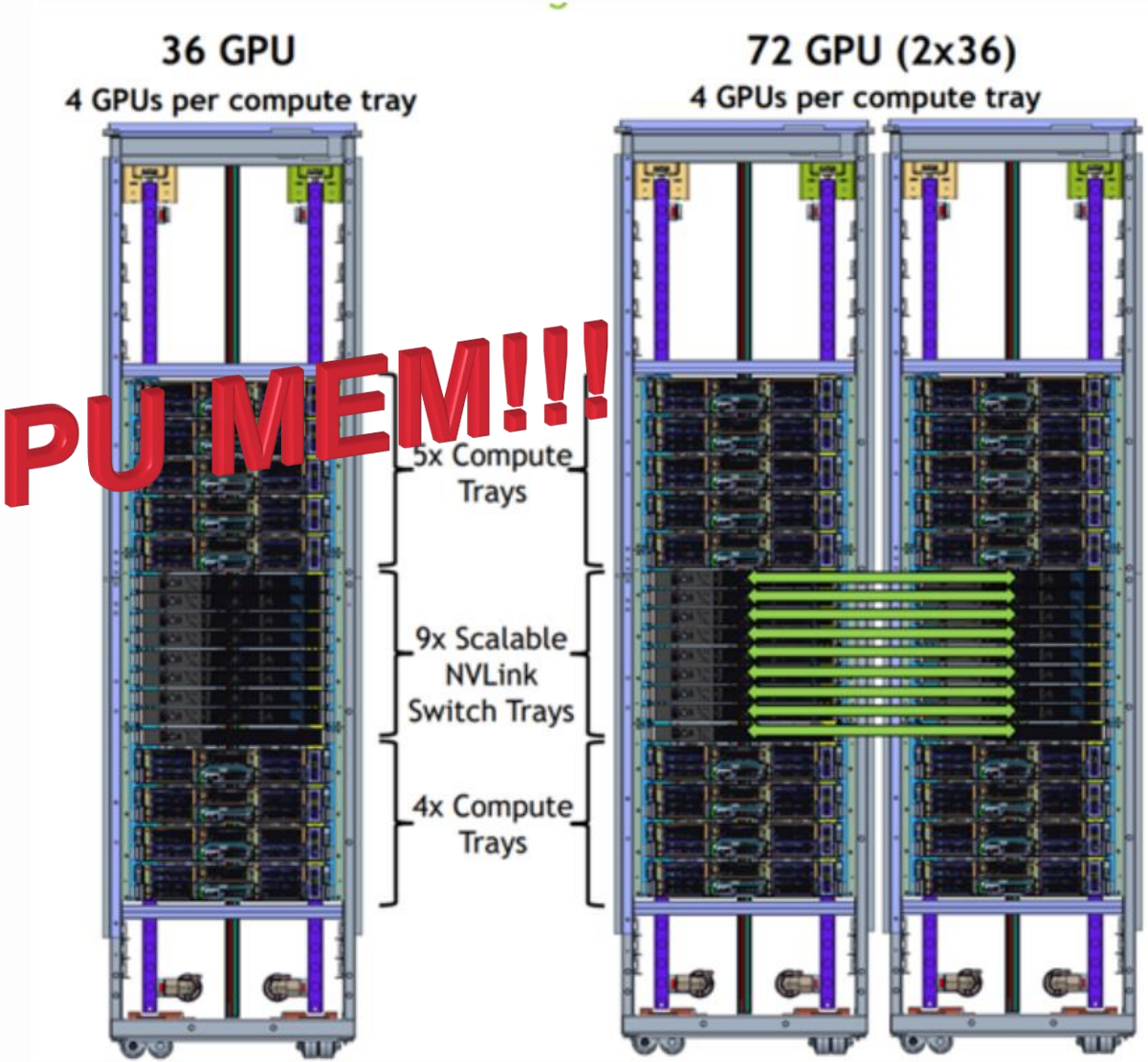
Oberon NVL72 GB200

SUPERMICRO
Confidential

- Liquid Cooled 192GB 1200 Watt GB 200 GPU
- Copper NVL to 72 GPU
- 400G CX7 and 800GB CX8 Q2, CY2025



13.8 TB GPU MEM!!!

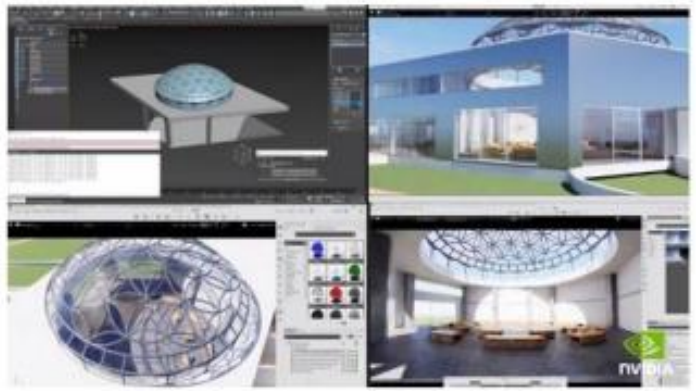


Digital Transformation Through Omniverse

Source: BMW Group

Digital Twins

Simulations and Rendering



Design and Collaboration

Global Connectivity and Collaboration

Increased Productivity and Maximize Creativity

Reduced Overhead Costs and Inefficient Workflows

NVIDIA OVX L40S

Scalable data center infrastructure for high-performance AI and graphics

OVX L40S

Reference Architecture

NVIDIA AI Enterprise I Omniverse Enterprise

Supermicro SYS-521GE
4 or 8X L40S
E/W Network: ConnectX-7
N/S Network: BlueField-3

Salable, High-Performance Network Fabric
Quantum I Spectrum-3 I Spectrum-X

NVIDIA L40S GPU

Ada Lovelace Architecture Features
New Streaming Multiprocessor
4th-Gen Tensor Cores
3rd-Gen RT Cores

Gen-AI, LLM Training, & Inference
Transformer Engine - FP8
1.5 petaFLOPS Tensor Performance

OVX Reference Architecture
Powerful AI and Graphics Performance at Scale
NVIDIA AI Enterprise I Omniverse Enterprise
Powered by L40S GPUS

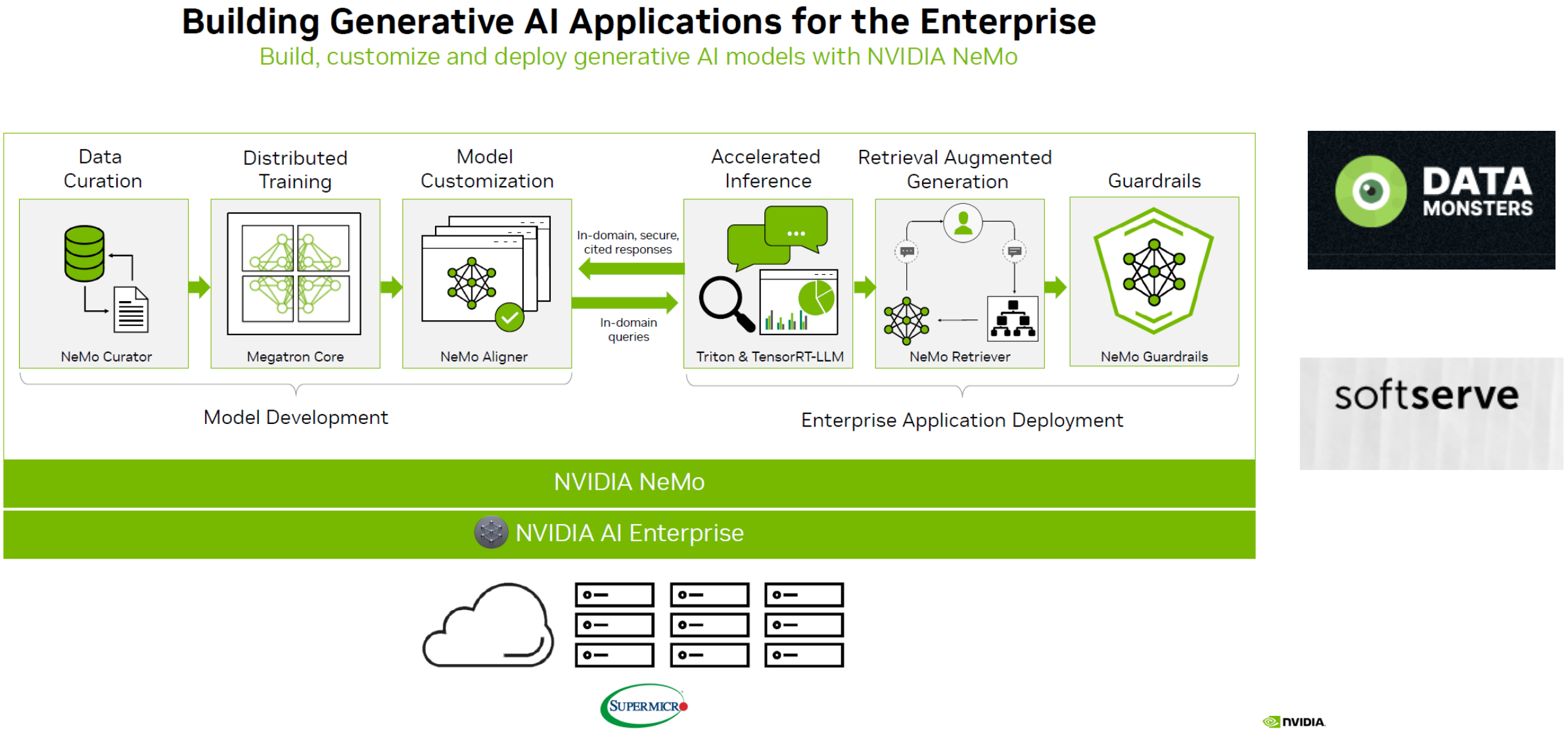
GPU SuperServer SYS-521GE-TNRT

OVX Systems Available from leading global OEMs

DP Intel 5U Dual-Root PCIe GPU System with up to 10 GPUs and extended thermal capacity

NVIDIA

Software Tools & Collaboration



Business Case: A University in Eastern Europe

Rack Layout Proposed Solution – IB Dragonfly +

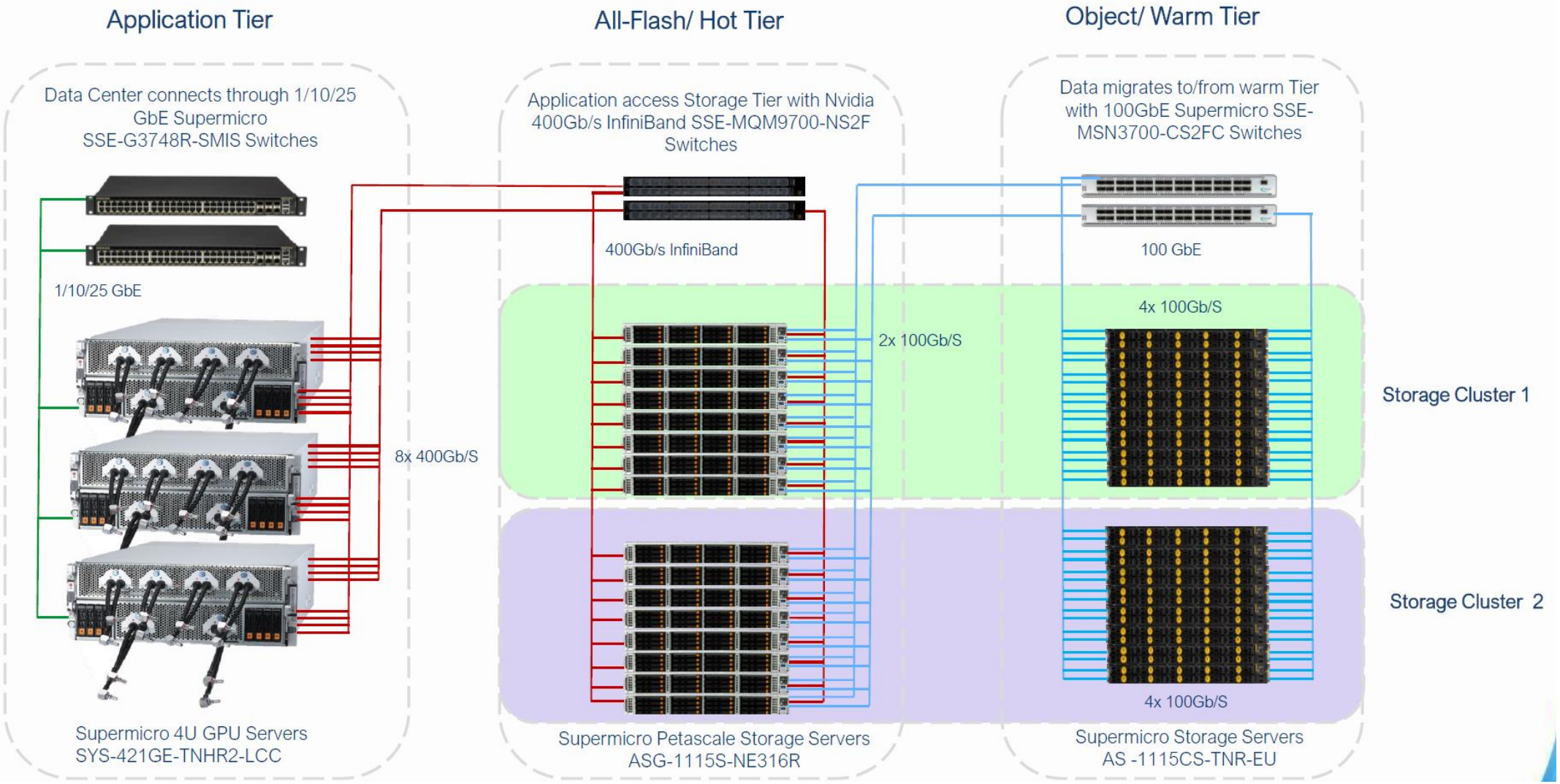
Overall Cluster Performance FP32 = 6 15 PFLOPs = 6,15 EFLOPs !!!

12 DLC + 2 Air RACKs
96 DLC GPU Systems with H200 HGX
8 PT of NVMe Storage
NDR 400Gbit Infiniband Network

Business Case: A University in Eastern Europe

WEKA

Storage Network Diagram

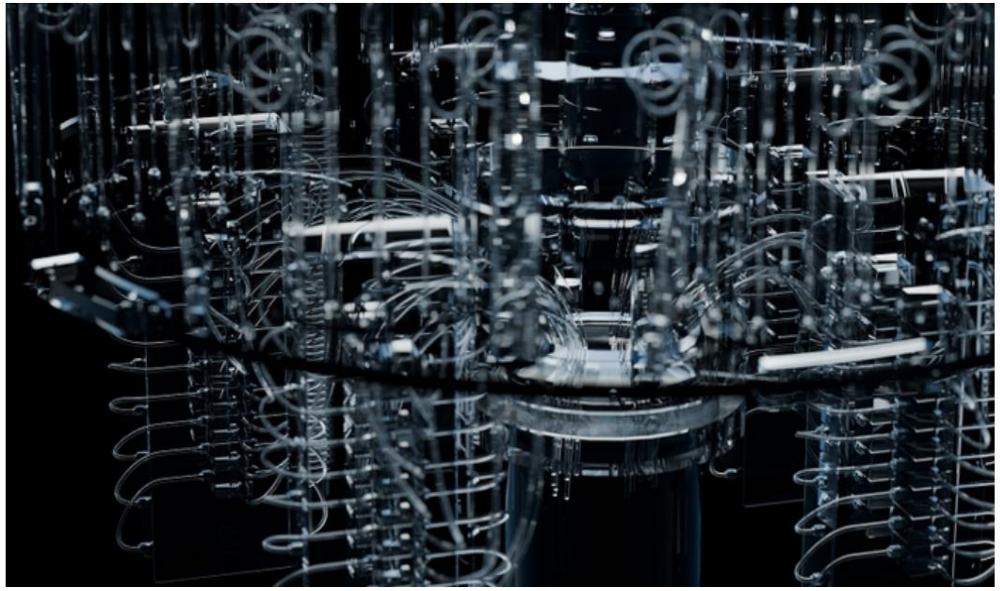


What's Next? Quantum Computing

Worldwide With CUDA-Q Platform

Supercomputers in Germany, Japan and Poland Incorporate Grace-Hopper and Quantum-Classical Accelerated Supercomputing Platform to Advance Quantum Computing Research

May 12, 2024



ISC -- NVIDIA today announced that it will accelerate quantum computing efforts at national supercomputing centers around the world with the open-source NVIDIA CUDA-Q™ platform.

Supercomputing sites in Germany, Japan and Poland will use the platform to power the quantum processing units (QPUs) inside their NVIDIA-accelerated high-performance computing systems.

QPUs are the brains of quantum computers that use the behavior of particles like electrons or photons to calculate differently than traditional processors, with the potential to make certain types of calculations faster.

Germany's Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich is installing a QPU built by IQM Quantum Computers as a complement to its JUPITER supercomputer, supercharged by the NVIDIA GH200 Grace Hopper™ Superchip.

The ABCI-Q supercomputer, located at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan, is designed to advance the nation's quantum computing initiative. Powered by the NVIDIA Hopper™ architecture, the system will add a QPU from QuEra.

Poland's Poznan Supercomputing and Networking Center (PSNC) has recently installed two photonic QPUs, built by ORCA Computing, connected to a new supercomputer partition accelerated by NVIDIA Hopper.

"Useful quantum computing will be enabled by the tight integration of quantum with GPU supercomputing," said Tim Costa, director of quantum and HPC at NVIDIA. "NVIDIA's quantum computing platform equips pioneers such as AIST, JSC and PSNC to push the boundaries of scientific discovery and advance the state of the art in quantum-integrated supercomputing."