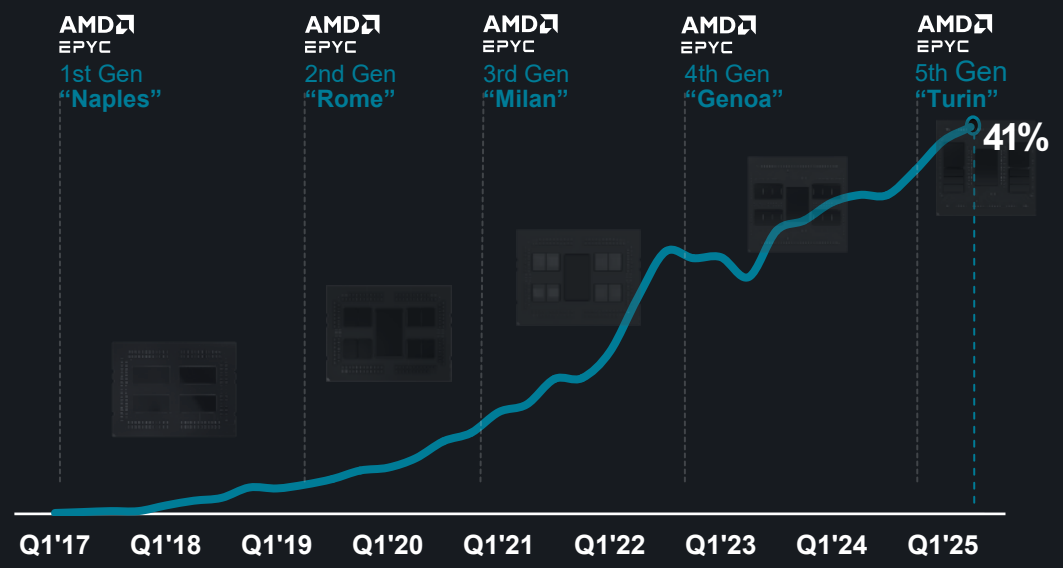
Scaling & Optimizing performance on AMD EPYC systems

Alexander Troshin

AMD
together we advance_

AMD EPYC™ CPUs – a Trusted Choice to Power World's Servers

>18x Server CPU Revenue Share Growth[1]



Industry Leaders Run on EPYC™

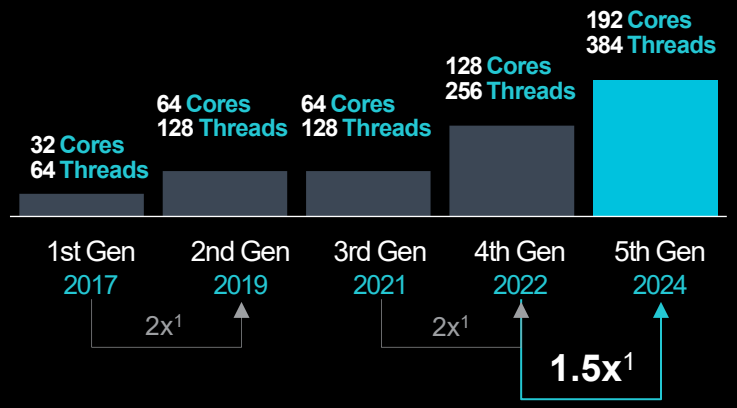
Cloud	aws, Microsoft, Google, ORACLE
Digital	NETFLIX, Uber, Meta, zoom
Enterprise	BEST BUY, IBM, Emirates NBD, NISSAN
OEM	DELL Technologies, Hewlett Packard Enterprise, Lenovo, SUPERMICRO, cisco

Consistently Expanding Customer Choice with Platform Availability & Market Coverage

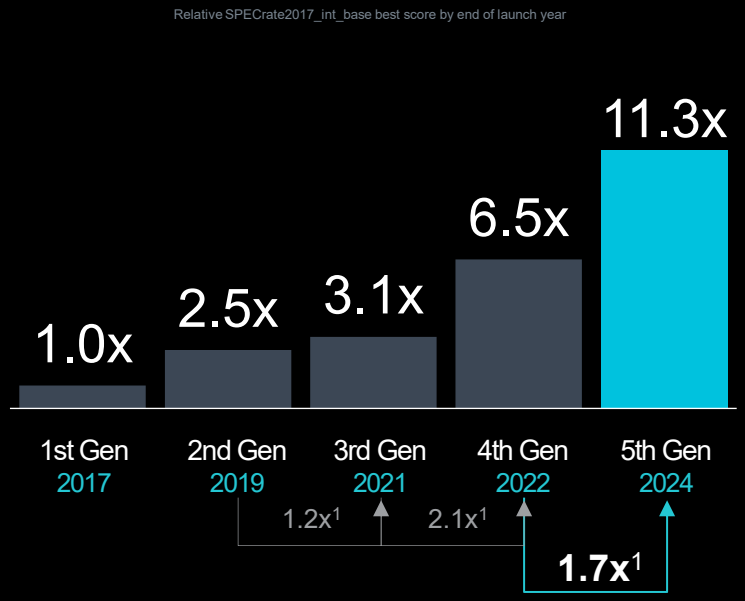
1 Market share is based on Mercury Research Sell-in Revenue Shipment Estimates, Q2 2025
See endnote: EPYC-055C

Benefits of Five Generations of AMD CPU Leadership

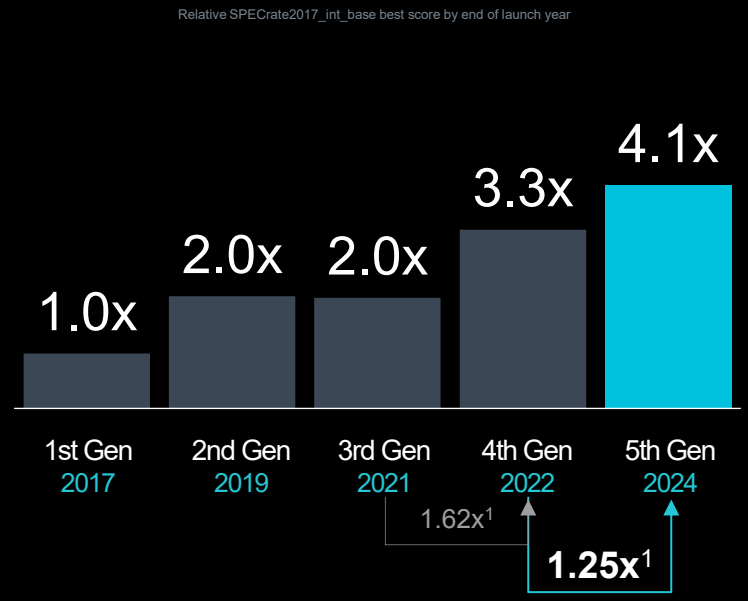
6x Core Count Across 5 Generations



11.3x Performance Across 5 Generations



4.1x Performance / CPU W Across 5 Generations



Total results as of 10/30/2024. See endnote 9xx5-069D

1 – comparison done using top of the stack CPU products, specifically: 1st Gen (7001 Series - 7601), 2nd Gen (7002 Series – 7742), 3rd Gen (7003 Series – 7763), 4th Gen (9004 Series – 9654), 5th Gen (9005 Series – 9965). All processor specifications are available on AMD.com website: Server Processor Specifications

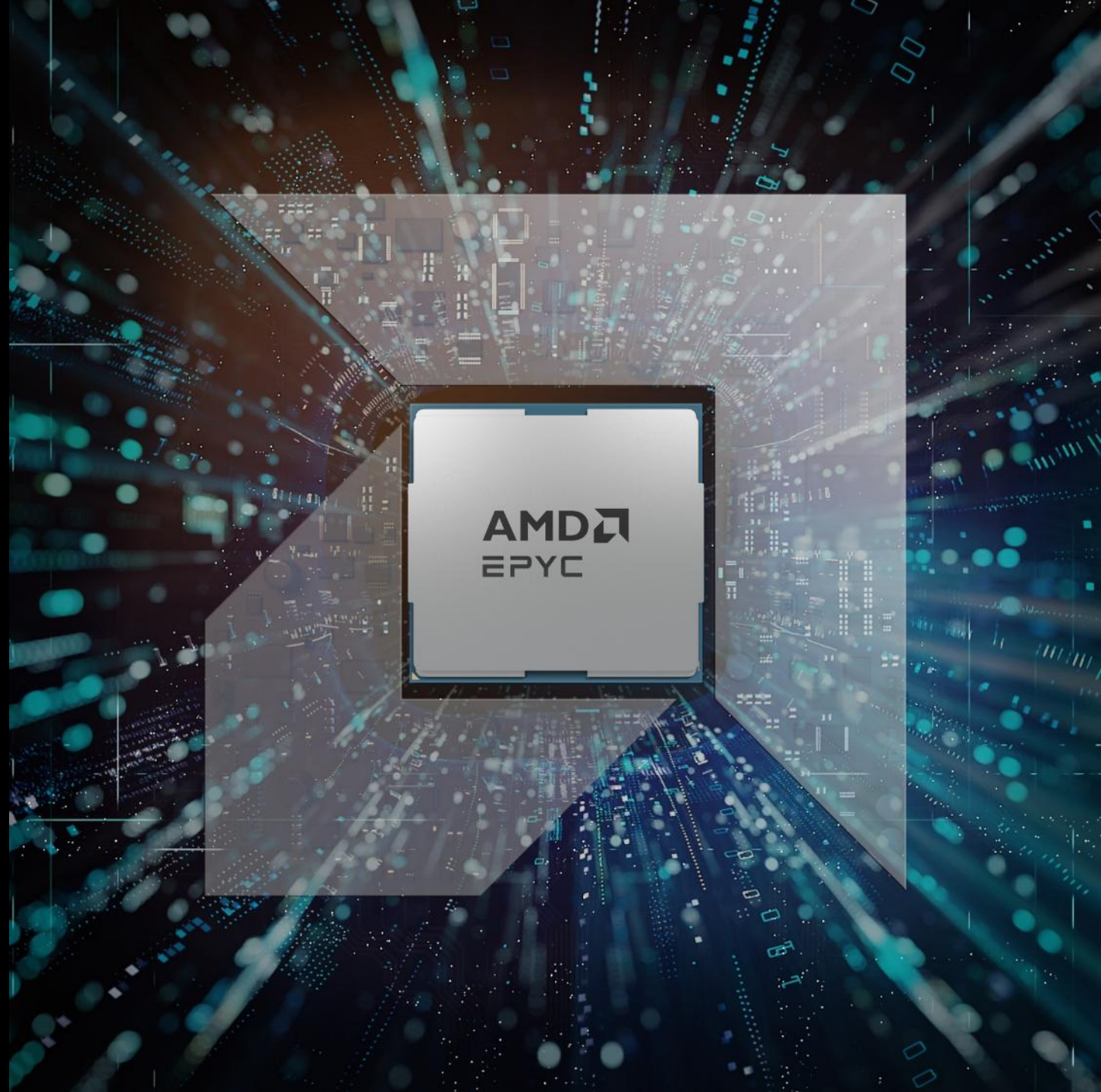
AMD EPYC™ Server CPUs

BUSINESS VALUE | PROVEN SOLUTIONS | SEAMLESS INNOVATION

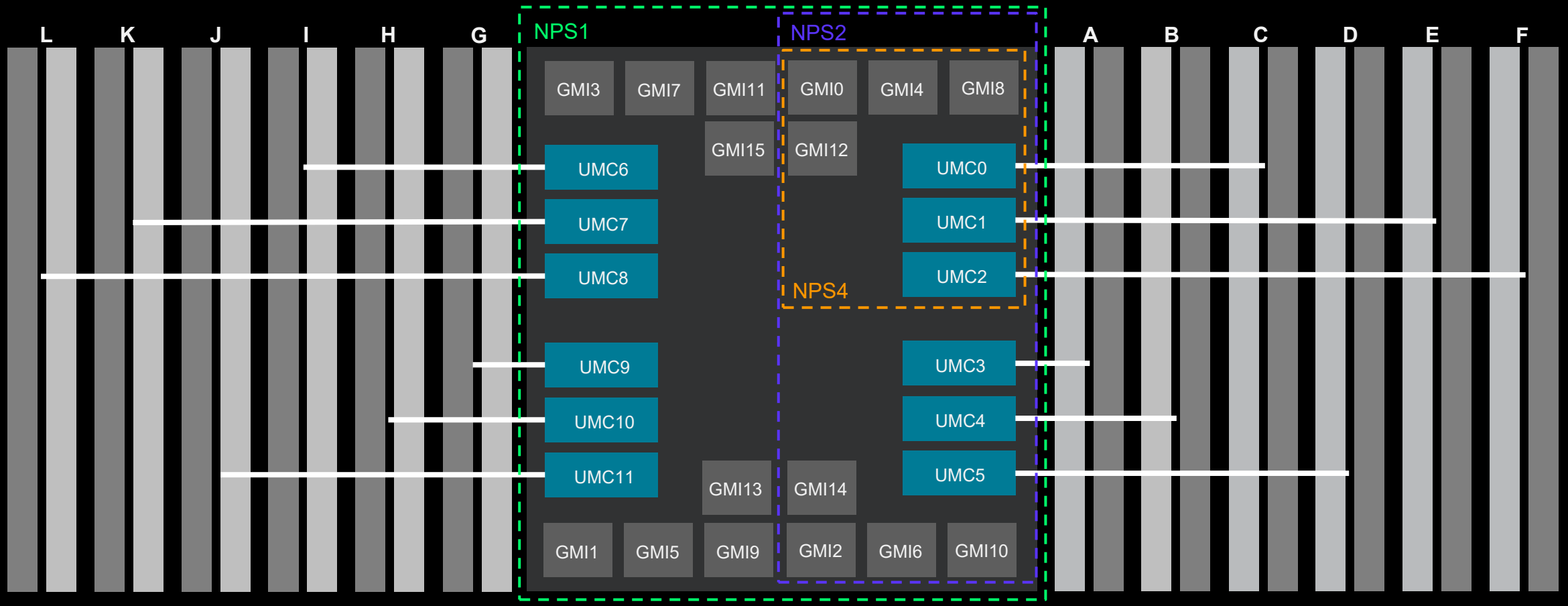
4000 Series	8000 Series	9000 Series
PRACTICAL	**INTELLIGENT EDGE**	**PERFORMANCE AND AI OPTIMIZED**
5th Gen / 4th Gen	4th Gen	5th Gen / 4th Gen
Attractively priced solutions for growing businesses and hosting services	Energy efficient solutions optimized for the Edge, Telco, Retail and Healthcare	Optimized performance for the most demanding AI-enabled data center workloads
4-16 cores | 1 Socket Only 2 Memory Channels 28 PCIe® 5.0 Lanes	8-64 Cores | 1 Socket Only 6 Memory Channels 96 PCIe® 5.0 Lanes	8-192 Cores | 1-2 Sockets 12 Memory Channels 128 (1P) / 160 (2P) PCIe® 5.0 Lanes

A complete portfolio of widely-deployed, optimized solutions, trusted by industry leaders

Memory Configurations



AMD EPYC™ 9005 CPUs



Multiple memory Nodes Per Socket (NPS) configurations for additional optimization

Server System DRAM Rules of Thumb*

Balance cost while meeting performance and scale needs

- Having enough memory for an application is *essential to performance and stability.*
- Provisioning more memory than needed is wasteful.
- Not all available memory channels *must* be populated on many systems.
- Evenly allocating memory across memory channels often provides optimal performance.

ASRock Rack TURIN2D48G 1 Photo courtesy of Serve The Home

Cost considerations especially critical in times of market constraint (today!)

*Consult system vendor documentation for specific platform on population guidance

Potential “Complications" in the Memory Selection Process

- **Memory module pricing**
  - Memory technologies are commoditized, but there are (**significant**) price and cost variabilities
    - Highest capacity DIMMs (typically 128GB and above) command significant price/GB premiums
    - Volumes matter at the bottom too: lower capacity DIMMs (16GB and below) can also have premiums
  - OEM server memory prices are typically significantly higher than spot market and channel retail prices

- **1 DIMM per channel (1DPC) or 2 DIMM per channel (2DPC)**
  - More memory channels allow more bandwidth to feed data to CPUs
  - 2DPC designs allow added capacity / flexibility, but are complex (bigger) system designs and incur performance hit

- **OEMs each provide supported or recommended memory configuration rules**
  - Some allow for heterogeneous or unbalanced memory configurations and use of fewer channels
  - Support and performance impact can be quite variable

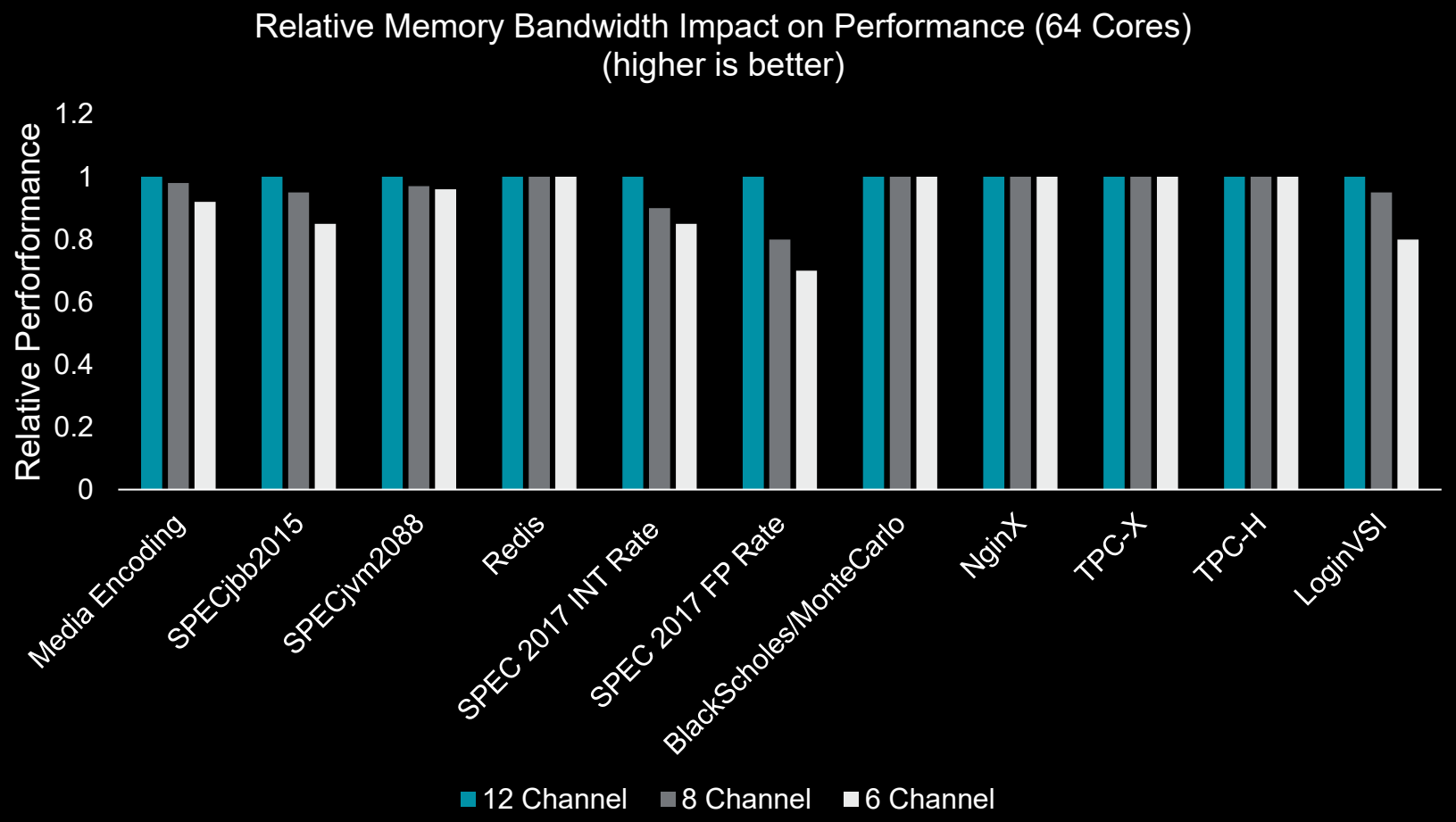
- **Non-standard memory options**
  - JEDEC is the industry group that creates memory standards to assure compatibility and interoperability
  - Choosing non-standard memory options creates support and availability risk and limits ability to re-deploy

Memory Configurations

8 memory channels vs 12 channels - Performance

Memory bandwidth impact is variable by workload:

- 30% on FP
- **Zero** on many workloads



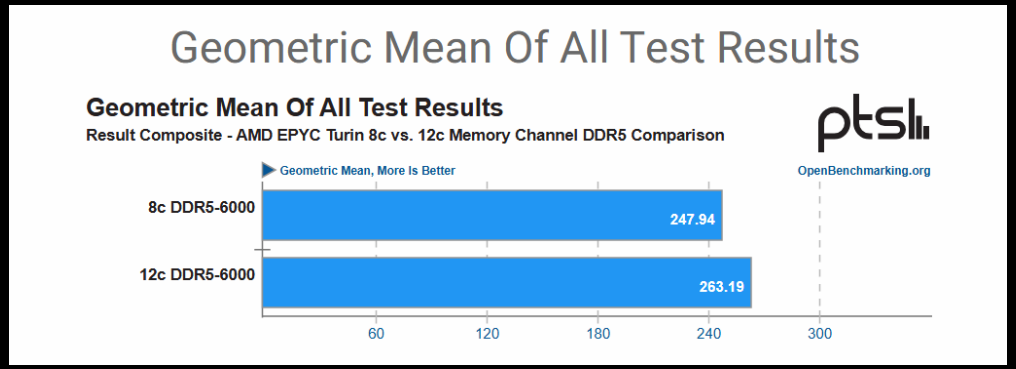
Source: AMD testing

8 memory channels vs 12 channels - Performance

In general, 12 channels is advantageous for performance - Geomean of testing with 100+ workloads shows roughly 5% better

12 channels allows the combination of more memory slots without the need to take the performance hit of utilizing 2DPC to get capacity

- 2S AMD EPYC 9005 platforms offer up to 24 DIMM slots with full 1DPC performance
- 2S Intel Xeon 6700/6500 platform only offer up to 16 DIMM slots before requiring move to 2DPC



Source: Phoronix: AMD EPYC Turin 8c Vs. 12c Memory Channel DDR5 Comparison Benchmarks - OpenBenchmarking.org

8 memory channels vs 12 channels – Options for Capacity

- EPYC 9005 processor provides more memory channels
- EPYC 9005 processor offers more flexible memory population that allows for additional capacity options

AMD EPYC 9005 Platform (12 ch)

Total Capacity (GB)	DIMM Capacity (GB)					
	16	32	64	96	128	256
128	8	4	2		1	
256	16	8	4		2	1
384	24	12	6	4		
512	32	16	8		4	2
576				6		
640	40	20	10			
768	48	24	12	8	6	
960				10		
1024		32	16		8	4
1152				12		
1280		40	20		10	
1536		48	24	16	12	6
1920				20		
2048			32		16	8
2304				24		
2560			40		20	10
3072			48	32	24	12
3840				40		
4096					32	16
4608				48		
5120					40	20
6144					48	24
8192						32
10240						40
12288						48

AMD EPYC 9005 1DPC	
Channels	12
Configuration Options	46
AMD EPYC 9005 2DPC	
Channels	12
Configuration Options	18

Supported at 1 DPC
Supported at 2 DPC
Cannot support

*Consult system vendor documentation for specific platform on population guidance
Additional info available in AMD EPYC 9005 Memory Population Recommendations Guide (58463) • Viewer • AMD Technical Information Portal

Net: A Prime Opportunity for a “How Much Memory” Discussion

- With memory costs gaining attention, it is a good time to assess REAL needs
- Customers often identify a target memory capacity to be allocated per core.
- GB/core ratios are often similar to the vCPU/memory ratios found in cloud
- Most common enterprise core count and capacity/core targets equal system capacities that are easily within range of AMD EPYC™ CPU 1DPC and memory pricing sweet spots.

Cores \ GB/Core	4	8	16	32
4	16	32	64	128
8	32	64	128	256
16	64	128	256	512
24	96	192	384	768
32	128	256	512	1024
48	192	384	768	1536
64	256	512	1024	2048
72	288	576	1152	2304
84	336	672	1344	2688
96	384	768	1536	3072
128	512	1024	2048	4096

Common DDR5 configuration / capacity choices

2DPC vs 1DPC Performance Considerations

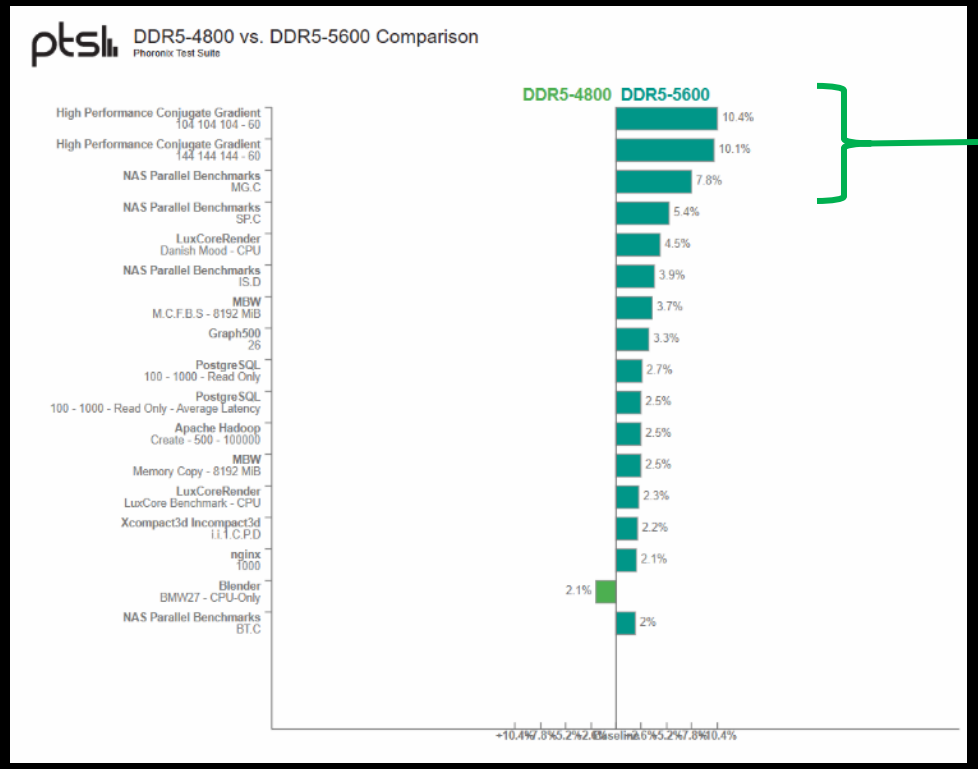
2DPC Performance

	AMD EPYC 9005	Intel Xeon 6 –SP
Host CPU options	8-192 Zen 5 w/SMT	E-core (no SMT) P-core w/SMT
Mem channels/CPU	12	8
Max RDIMM Frequency 1DPC, 2DPC	6400MT/s, 4000MT/s	6400MT/s, 5200MT/s
Max Mem BW 1DPC	1229 GB/s	819 GB/s
Max Mem BW 2DPC	768 GB/s	666 GB/s

- 5200MT/s is absolutely 30% higher than 4,000MT/s
- But ***Intel memory BW channel scale limitations undermine that advantage***
- Intel 2DPC aggregate BW 54% less than AMD 1DPC, 15% less than AMD 2DPC

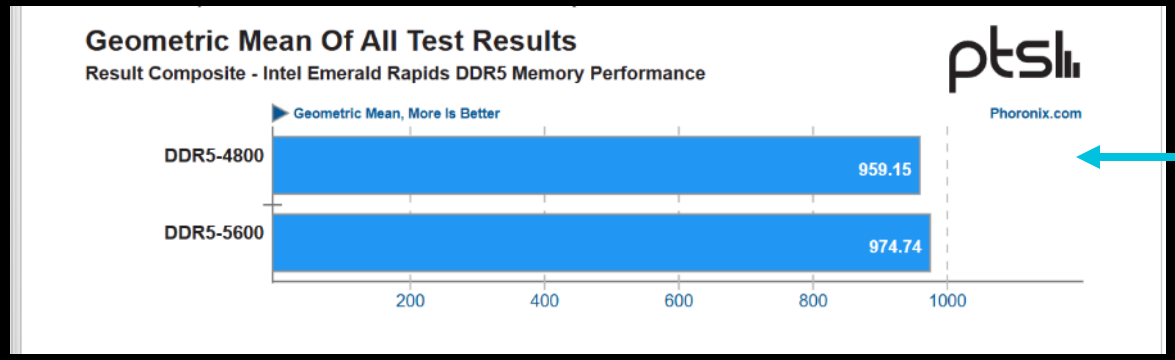
“Faster” Memory is Better: 5600MT/s v 4800 MT/s Example

- Mathematically, 5600MT/s DDR5 should deliver ~17% better throughput than 4800MT/s
- In practice, this is rarely achieved due to bus inefficiencies, CPU interactions and software memory utilization needs and more



Some workloads (especially in HPC and technical computing) show some performance gains with “faster” memory

None of the tests achieve theoretical levels of improvement (~17%) expected



The geomean performance gain across 37 benchmarks was ~2%

3rd Party Testing - Source: Phoronix: https://www.phoronix.com/review/intel-xeon-ddr5-5600/4

Confidential Computing

AMD Confidential Computing Journey

Multiple Generations of Security Leadership with AMD SEV

New in 5th Gen AMD EPYC™ CPUs

Secure Encrypted Virtualization — 2017
AMD launches Secure Encrypted Virtualization (SEV)

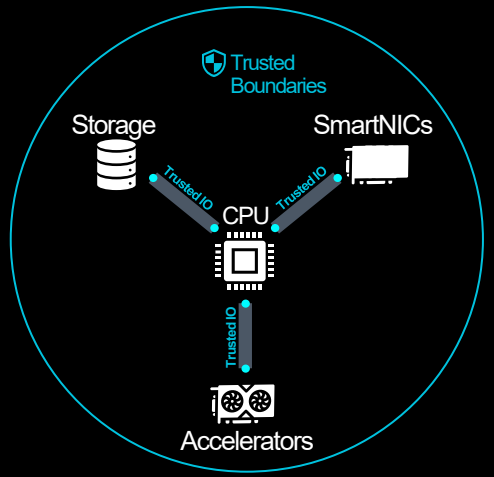
Encrypted State — 2019
SEV adds register encryption with Encrypted State

Secure Nested Paging — 2021
SEV adds malicious hypervisor protection with Secure Nested Paging

2022
SEV adds protection for CXL®-attached memory and stronger 256-bit encryption

2023
AMD publishes Secure Nested Paging source code

Trusted I/O — 2024
Trusted I/O extends the SEV Trusted Boundary to include PCIe®-attached devices



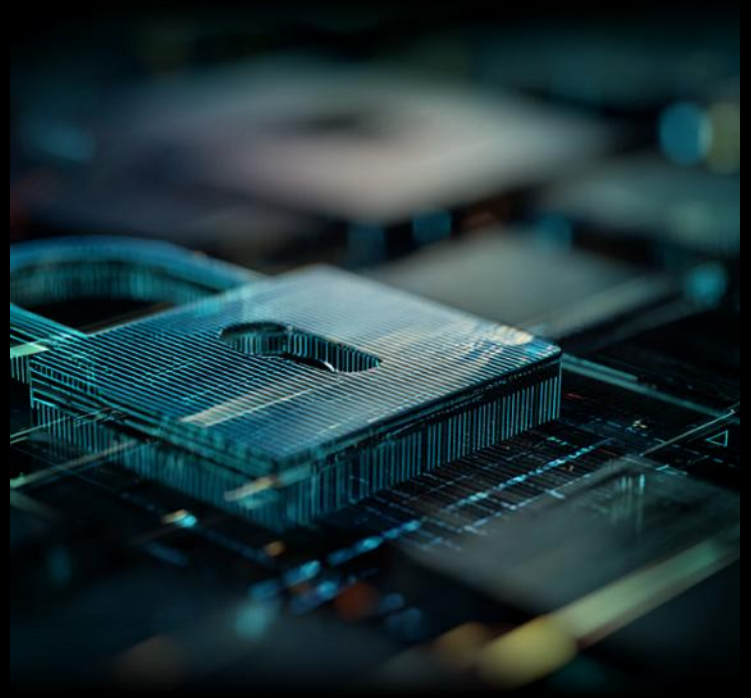
AMD Infinity Guard features vary by EPYC™ Processor series and/or generations. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at https://www.amd.com/en/technologies/infinity-guard.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

Building Confidential Computing Features

AMD Multi-year Investment

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	✓	✓	✓	✓	✓

Confidential Computing Starts with AMD EPYC™ Server CPUs



Visit the AMD Confidential Computing webpage

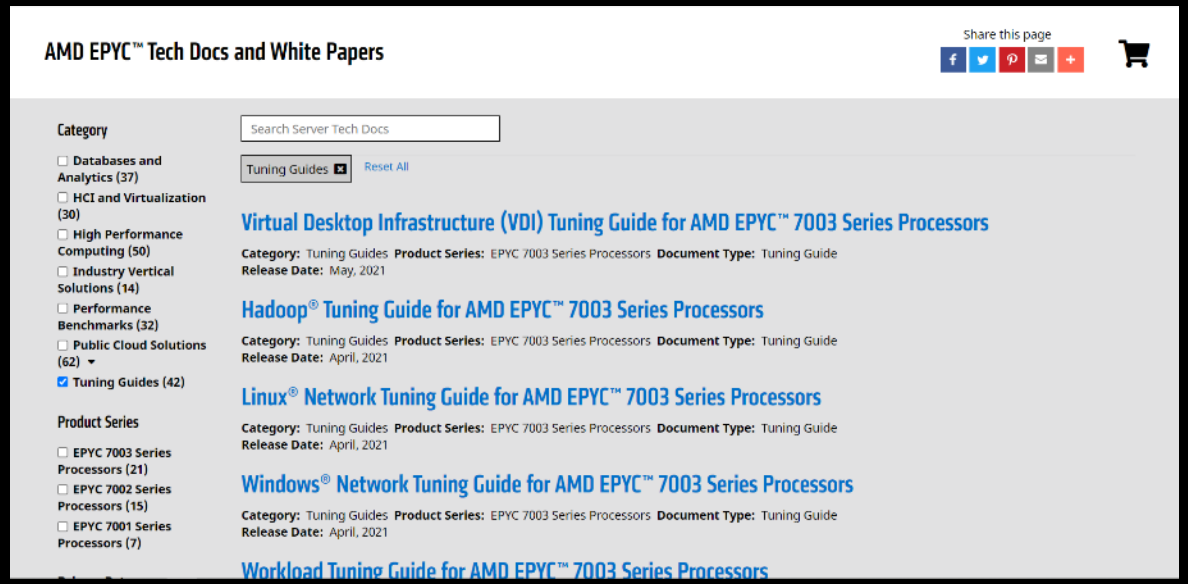
Check out AMD EPYC eBook: The confidential computing revolution

Learn more on how to deploy & use confidential computing features with the User Guide - Using SEV with AMD EPYC Processors (58207) • Viewer • AMD Technical Information Portal

EPYC™ SALES TOOLS

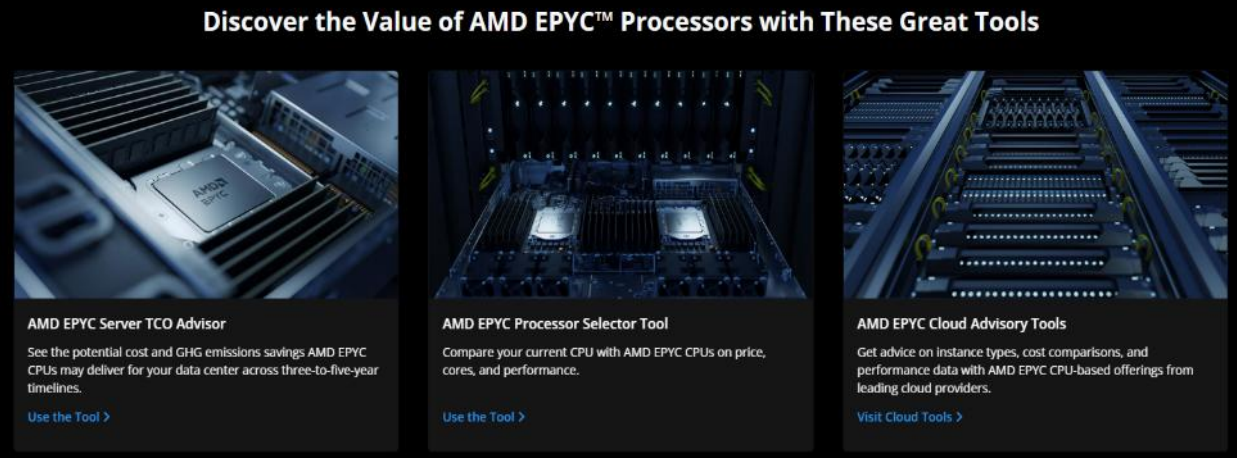
APPLICATION PERFORMANCE DOCUMENTATION

www.amd.com/en/processors/server-tech-docs/search



CPU SELECTOR AND TCO CALCULATORS

https://www.amd.com/en/resources/epyc-tools.html



ALSO AVAILABLE:

AMD PARTNER HUB

AMD'S "MEET THE EXPERTS" WEBINAR SERIES

TRAINING MODULES ON "AMD ARENA"

AMD EPYC™ CASE STUDIES

CPU PERFORMANCE COMPARISON TOOL

& MORE