

The Case for CXL Memory Expansion

Yong Tian, Field CTO

Motivation To Scale Up Server



- HPC and Database use cases
- Enable processing of ever larger dataset
- Faster time to result
- Higher performance cost ratio

Example 1 - Metabat, a genomic workload

Running Environment	Time to Result
Cluster of 100 64GB nodes	>2 weeks per run, high chance of failure
Single 4TB node	4 days per run

Example 2 - SQL Server

Running Environment	QPS	Cost
Baseline Single 64GB node	1x	
Scale out Two 64GB nodes	2x	Additional server cost, Software license per CPU
Scale up 64GB node with 128 GB CXL expansion	2x	Memory expansion module

Challenges of Scale Up Server

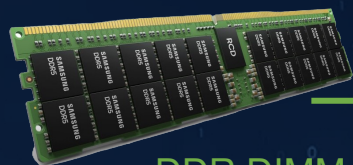
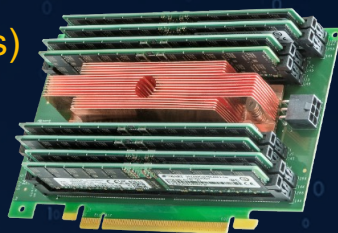


- Memory scale up challenges
 - Limited DIMM slots on the motherboard
 - Memory capacity and bandwidth
 - Cost
- Solution
 - CXL memory expansion
 - Memory Machine™ X software

CXL Memory Expansion

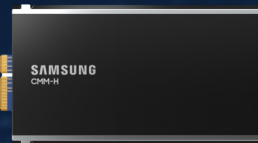


PCIe Add-In Cards (AICs)



DDR DIMMs

E3.S Memory Modules



Add-in Card (AIC)

- Flexible capacity, up to 2 TB per card
- Higher bandwidth, up to x16 PCIe5 lanes (~ 1x DDR5 channel)

E3.S Modules

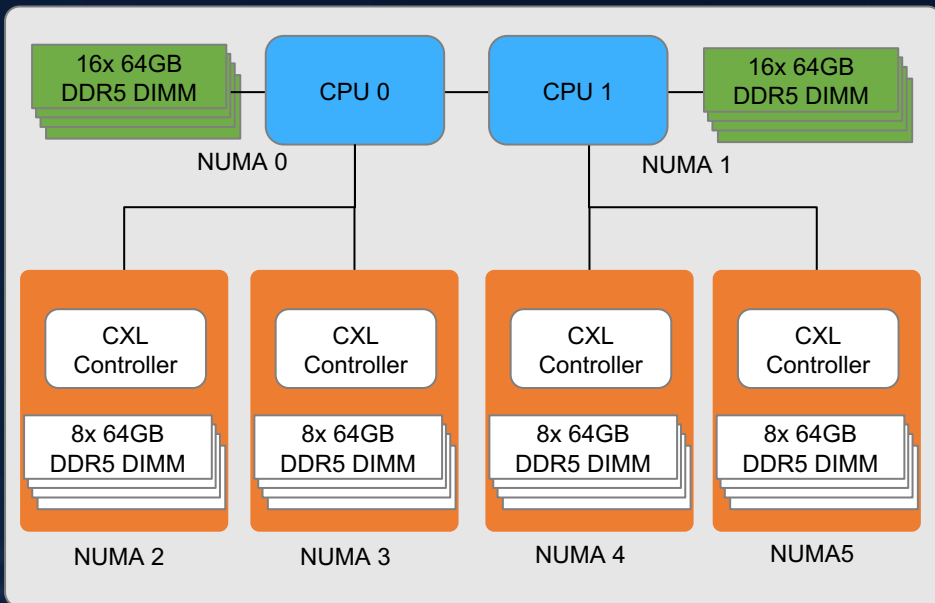
- Easy front loading, same as SSDs
- Fixed capacity – 128, 256, & 512 GB
- Lower bandwidth at x8 PCIe5 lanes

Memory Cost Comparison



Configurations (Dual CPU, unless otherwise noted)	Server Memory Spec			Socket DRAM - DDR 5				CXL DRAM					
	Total Size (GB)	Total Cost	Per GB Cost	DIMM size (GB)	# of DIMMs	Size Subtotal	Cost Subtotal	DIMM type	DIMM size (GB)	# of DIMMs per AIC	# of AICs	Size Subtotal	Cost Subtotal
A. Socket DRAM	4,096	\$46,080	\$11.25	128	32	4,096	\$46,080						
B. Socket & CXL DRAM	4,096	\$22,928	\$5.60	64	32	2,048	\$6,464	DDR5	64	8	4	2,048	\$16,464
C. Socket & CXL DRAM	4,096	\$20,072	\$4.90	64	32	2,048	\$6,464	DDR4	128	8	2	2,048	\$13,608
D. Socket DRAM, Quad CPU	8,192	\$92,160	\$11.25	128	64	8,192	\$92,160						
E. Socket & CXL DRAM	8,192	\$47,288	\$5.77	64	32	2,048	\$6,464	DDR4	128	8	6	6,144	\$40,824
F. Socket & CXL DRAM	11,264	\$62,656	\$5.56	96	32	3,072	\$12,224	DDR4	128	8	8	8,192	\$50,432
G. Socket & CXL DRAM, Quad CPU	32,768	\$425,600	\$12.99	256	64	16,384	\$204,800	DDR5	256	8	8	16,384	\$220,800

CXL Leads to Complex Heterogeneous Memory Hierarchy



NUMA	Capacity	Latency*	Bandwidth*
0	1 TB	90 ns	512 GB/s
1	1 TB	190 ns	Shared 62.4 GB/s with NUMA 4, 5
2	512 GB	190 ns	64 GB/s
3	512 GB	190 ns	64 GB/s
4	512 GB	290 ns	Shared 62.4 GB/s with NUMA 1, 5
5	512 GB	290 ns	Shared 62.4 GB/s with NUMA 1, 4

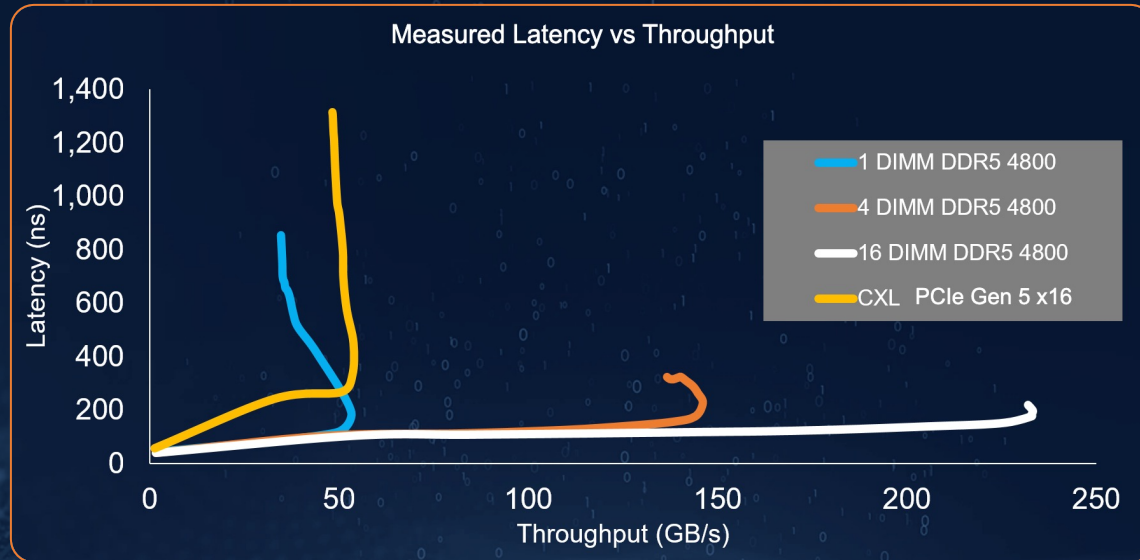
* Assume process running on NUMA 0

Application Performance Challenges



Application performance impact due to cross NUMA memory access

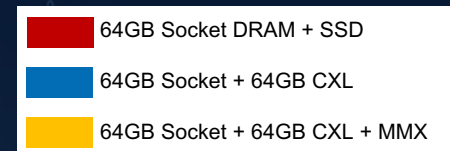
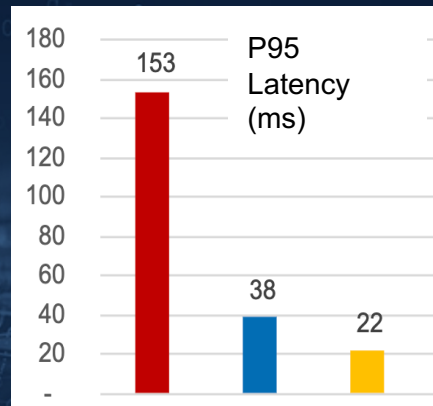
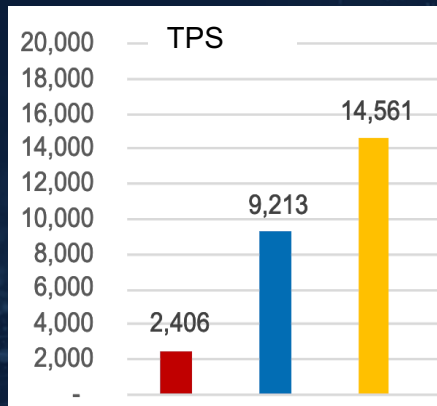
- Higher latency ~100 to 200 ns
- Bandwidth bottleneck



Memory Machine™ X Speeds Up Applications on CXL



- Memory Machine™ X (MMX) mitigate the CXL latency and bandwidth challenges, by
 - Continuously monitor app memory access pattern
 - Optimize data placement across NUMA nodes
- Transparent to application and Linux OS kernel
- Example - MySQL & TPC-C Benchmark



BACKUP



Reference Servers



- Model AS-2125HS-TNR
 - 2U, Dual AMD EPYC™ 9534 w/ 64-Core
 - Up to 4 CXL AIC cards
 - Cost **\$18,119** without any DRAM

- Lenovo SR650 / SR675 V3
 - 3U, Dual Intel SPR or AMD Genoa
 - Up to 8 CXL AIC cards

- Model AS-4125GS-TNRT
 - 4U, Dual AMD EPYC™ 9534 w/ 64-Core
 - Up to 8 CXL AIC cards
 - Cost **\$20,423** without any DRAM

- Model 241H-TNRTTP
 - 4U, Quad INTEL XEON 8454H w/ 32-Core
 - Up to 8 CXL AIC cards
 - Cost **\$42,391** without any DRAM

1600W Redundant Platinum Level Power Supplies**

VGA Port
2 USB Ports
Dedicated IPMI LAN Port

Slot	Description		Slot	Description	
1	Option 1*	Option 2*	3	Option 1	Option 2
	PCIe 5.0 x16 (FH, 10.5" L)	PCIe 5.0 x8 (FH, 10.5" L)		PCIe 5.0 x16 (FH, 10.5" L)	PCIe 5.0 x8 (FH, 10.5" L)
2	Not available	PCIe 5.0 x8 (FH, 10.5" L)	4	Not available	PCIe 5.0 x8 (FH, 10.5" L)

Slot	Description			Slot	Description
5	Option 1*	Option 2*	Option 3*	8	PCIe 5.0 x16 AICM Slot (NCSS)
	PCIe 5.0 x16 (FH, 10.5" L)	PCIe 5.0 x16 (FH, 10.5" L)	PCIe 5.0 x8 (FH, 10.5" L)		
6	Not available	Not available			
7	PCIe 5.0 x8 (FH, 10.5" L)	PCIe 5.0 x16 (FH, 10.5" L)			
8	PCIe 5.0 x8 (FH, 10.5" L)	Not available			

* Slots 1, 2, 5
** Full redundants



AIC Dimension, Power, and Availability



CXL AIC	Dimension	Power	Availability
AsteraLabs A1000 4x DDR5 DIMM	FHHL Double width	75 W (256 GB - 4x 64GB)	Off the shelf
Smart Modular 4x DDR5 DIMM	FHHL Single width	75 W (256 GB - 4x 64GB)	ES – Q1, 2024 CS – Q2, 2024
Smart Modular 8x DDR5 DIMM	FHHL Double width	135 W (512GB - 8x 64GB)	ES – Q1, 2024 CS – Q2, 2024
Vendor C 4x DDR4 DIMM	FHFL Single width	60 W (256 GB - 4x 64GB)	ES – Q1, 2024
Vendor C 8x DDR4 DIMM	FHFL Double width	100 W (512GB - 8x 64GB)	ES – Q1, 2024



ES-Engineering Sample; CS-Customer Sample

Memory Cost Optimization



Intel CPU, Quad (A) and Dual (B) Sockets

Configurations	Server Memory Spec			Socket DRAM - DDR 5				CXL DRAM					
	Total Size (GB)	Total Cost	Per GB Cost	DIMM size (GB)	# of DIMMs	Size Subtotal	Cost Subtotal	DIMM type	DIMM size (GB)	# of DIMMs per AIC	# of AICs	Size Subtotal	Cost Subtotal
A. Socket DRAM	8,192	\$92,160	\$11.25	128	64	8,192	\$92,160						
B. Socket & CXL DRAM	8,192	\$47,288	\$5.77	64	32	2,048	\$6,464	DDR4	128	8	6	6,144	\$40,824

AMD CPU, Dual Sockets

Configurations	Server Memory Spec			Socket DRAM - DDR 5				CXL DRAM					
	Total Size (GB)	Total Cost	Per GB Cost	DIMM size (GB)	# of DIMMs	Size Subtotal	Cost Subtotal	DIMM type	DIMM size (GB)	# of DIMMs per AIC	# of AICs	Size Subtotal	Cost Subtotal
C. Socket DRAM	6,144	\$76,800	\$12.50	256	24	6,144	\$76,800						
D. Socket & CXL DRAM	8,704	\$52,476	\$5.95	64	24	1,536	\$4,848	DDR4	128	8	7	7,168	\$47,628

* Cost estimate only

** Assume per AIC card cost \$2,500