SNIA DEVELOPER CONFERENCE

SDC 24

BY Developers FOR Developers

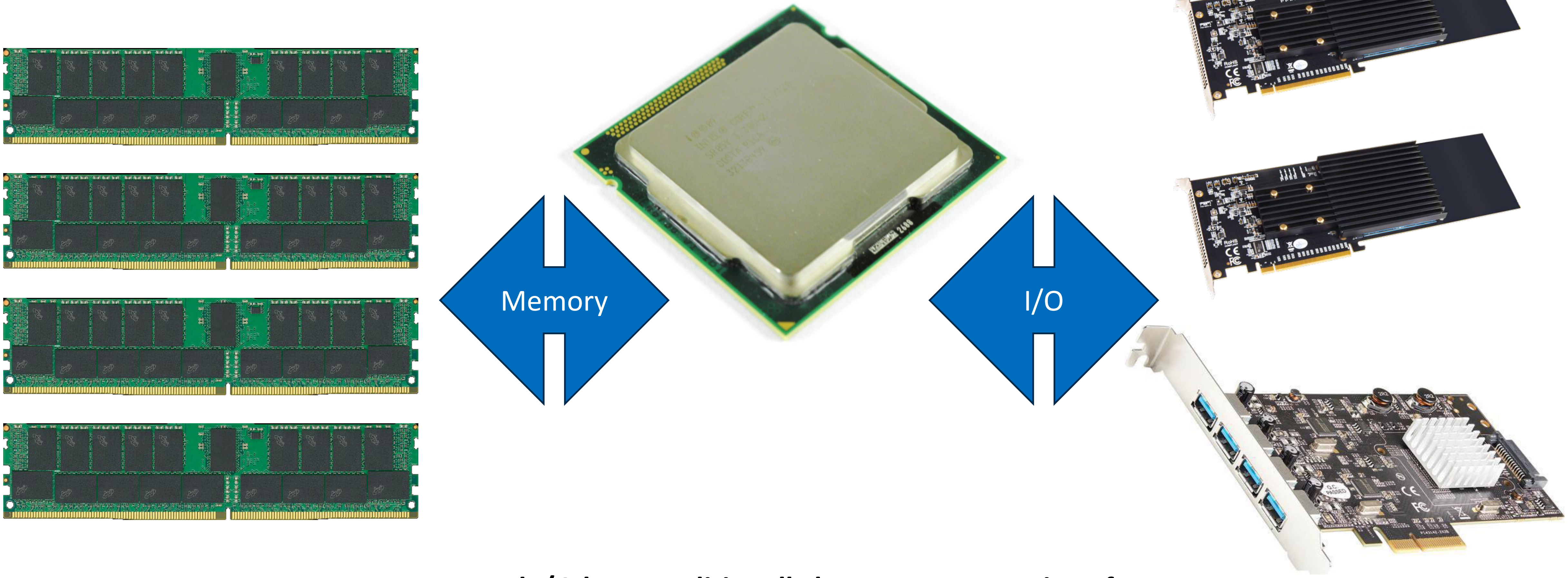
September 16-18, 2024
Santa Clara, CA

NVMe Over CXL is much more than just an SSD

Bill Gervasi, Principal Systems Architect

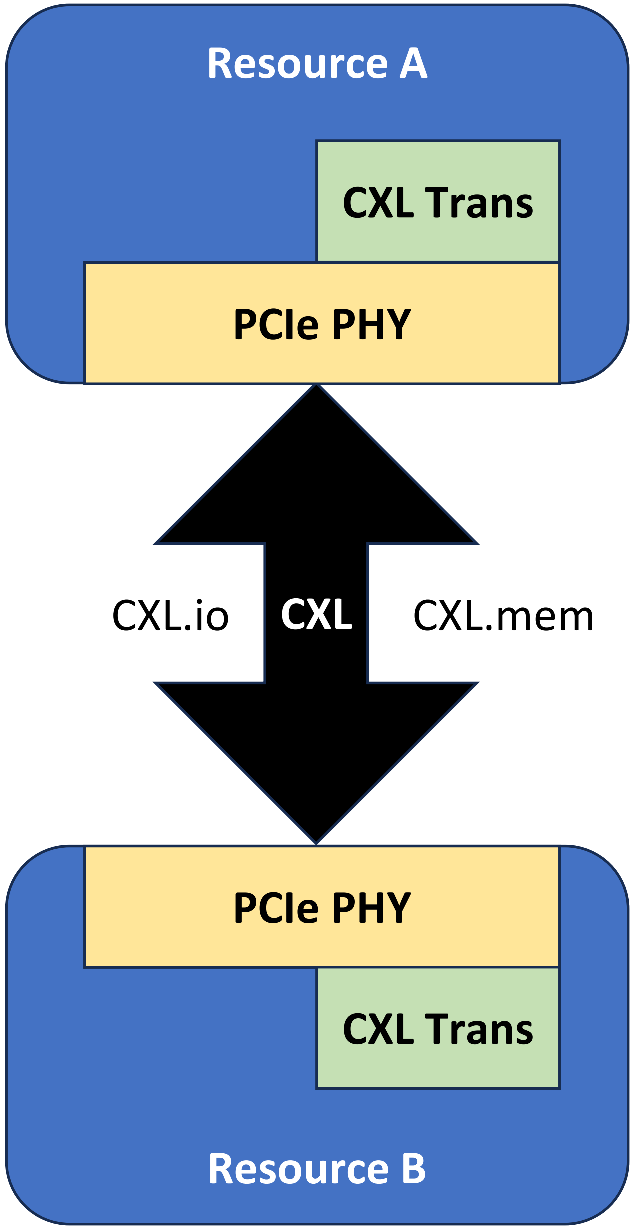
WOLLEY

Memory Versus Storage



Memory and I/O have traditionally been on separate interfaces

This has been our basic system architecture for thousands of years

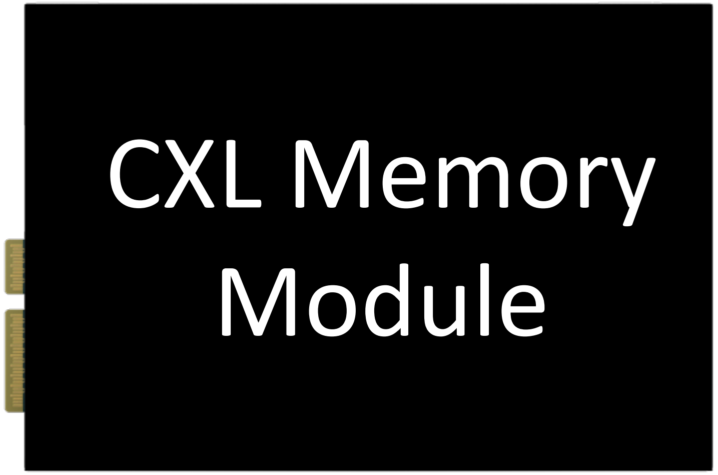


CXL is more than just another I/O bus

CXL allows the blending of processing, memory, storage, and I/O over a consistent protocol

This enables virtualizing resources in interesting new ways

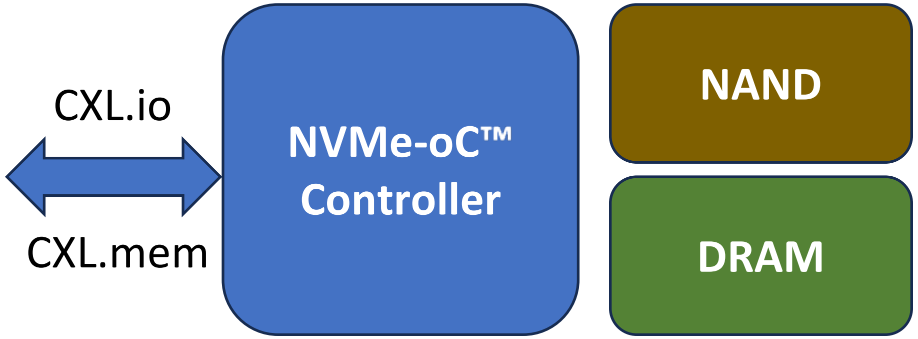
PCIe SSD



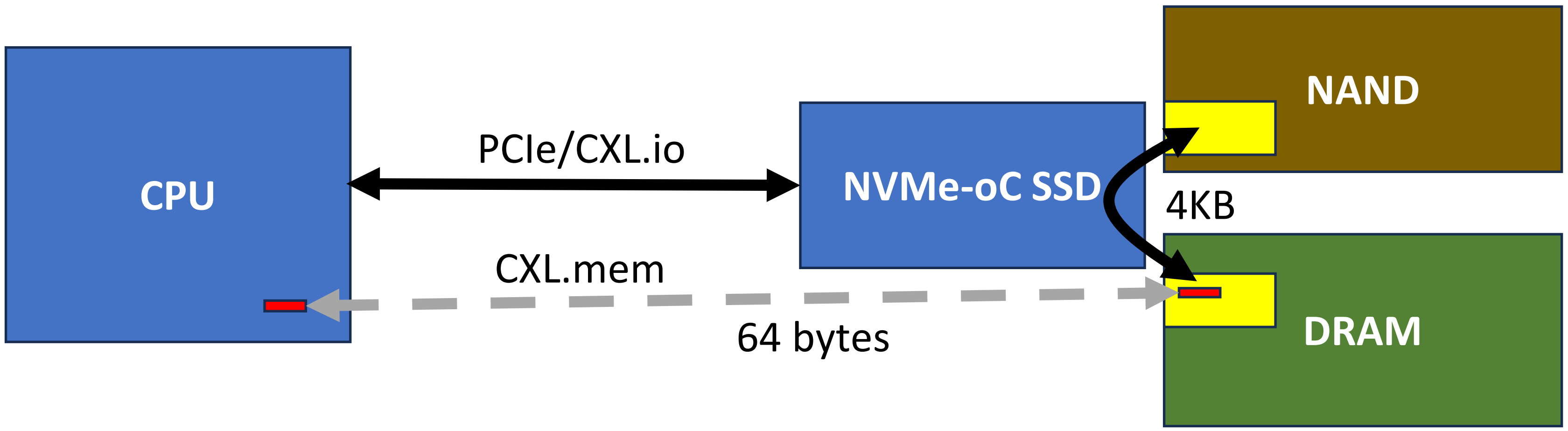
CXL Memory Module

NVMe-Over-CXL™ Merges memory and storage into a unified interface

NVMe-Over-CXL™



The NVMe Over CXL Solution: Only grab the FLITs you need



Dual-function device

Appears to the system as
an NVMe SSD and as
Host Directed Memory (HDM)

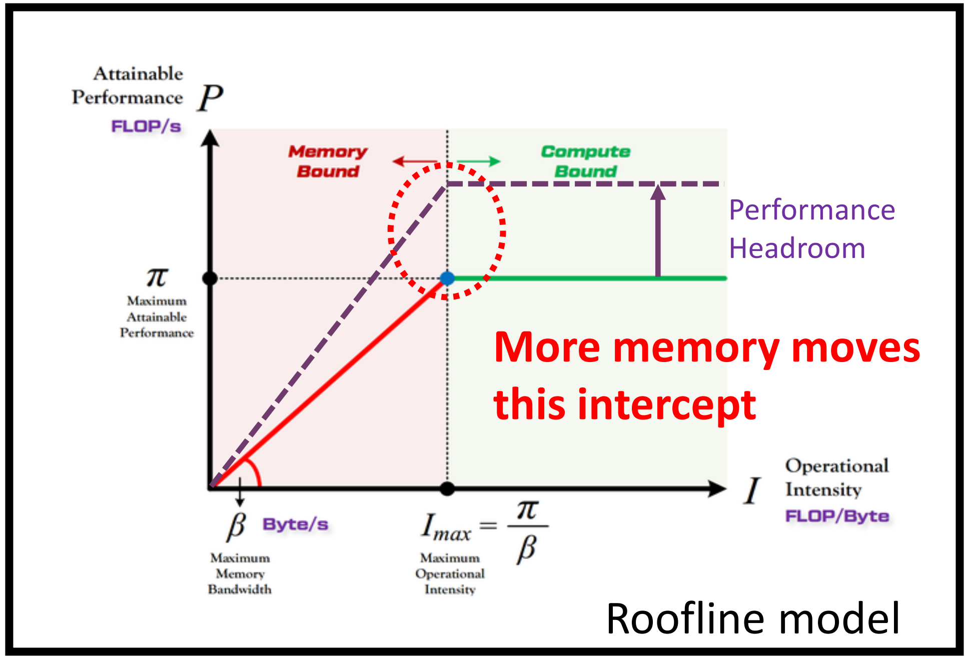
NVMe is just a cache protocol between NAND and DRAM

NVMe-oC places the controller memory buffer (CMB) in CXL space (HDM)

Processor grabs only the FLITs needed using CXL.mem

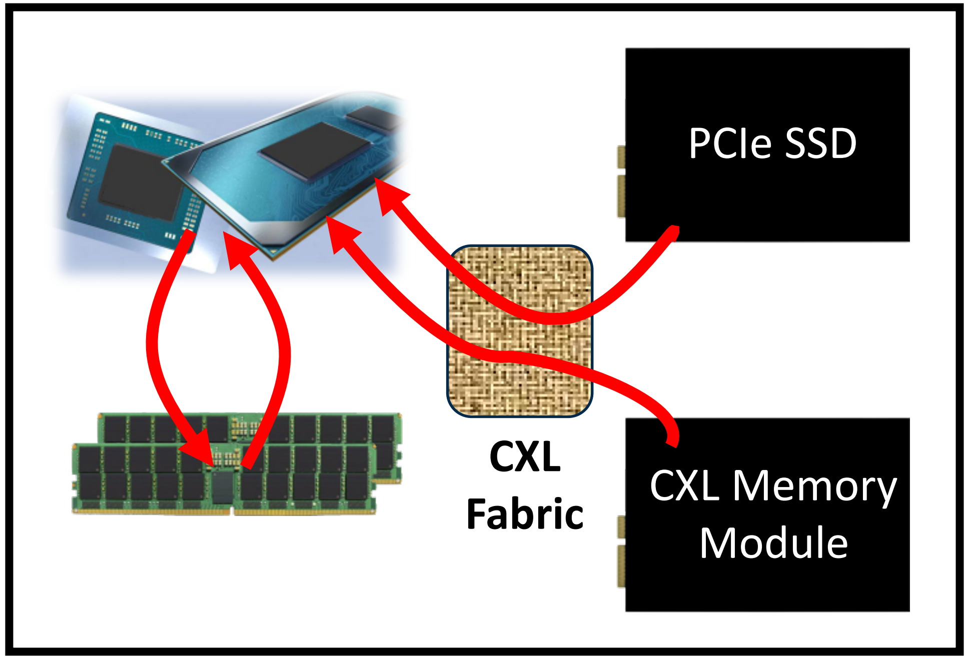
The rest of the CMB data (on average, 97%) remains where it is

This cache management scheme is expanded to create Virtual HDM



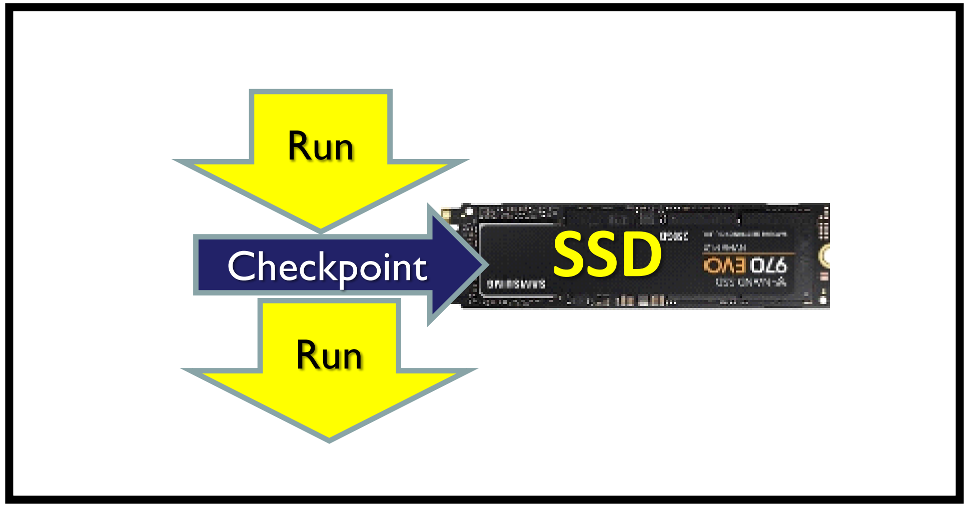
More memory moves this intercept

Roofline model



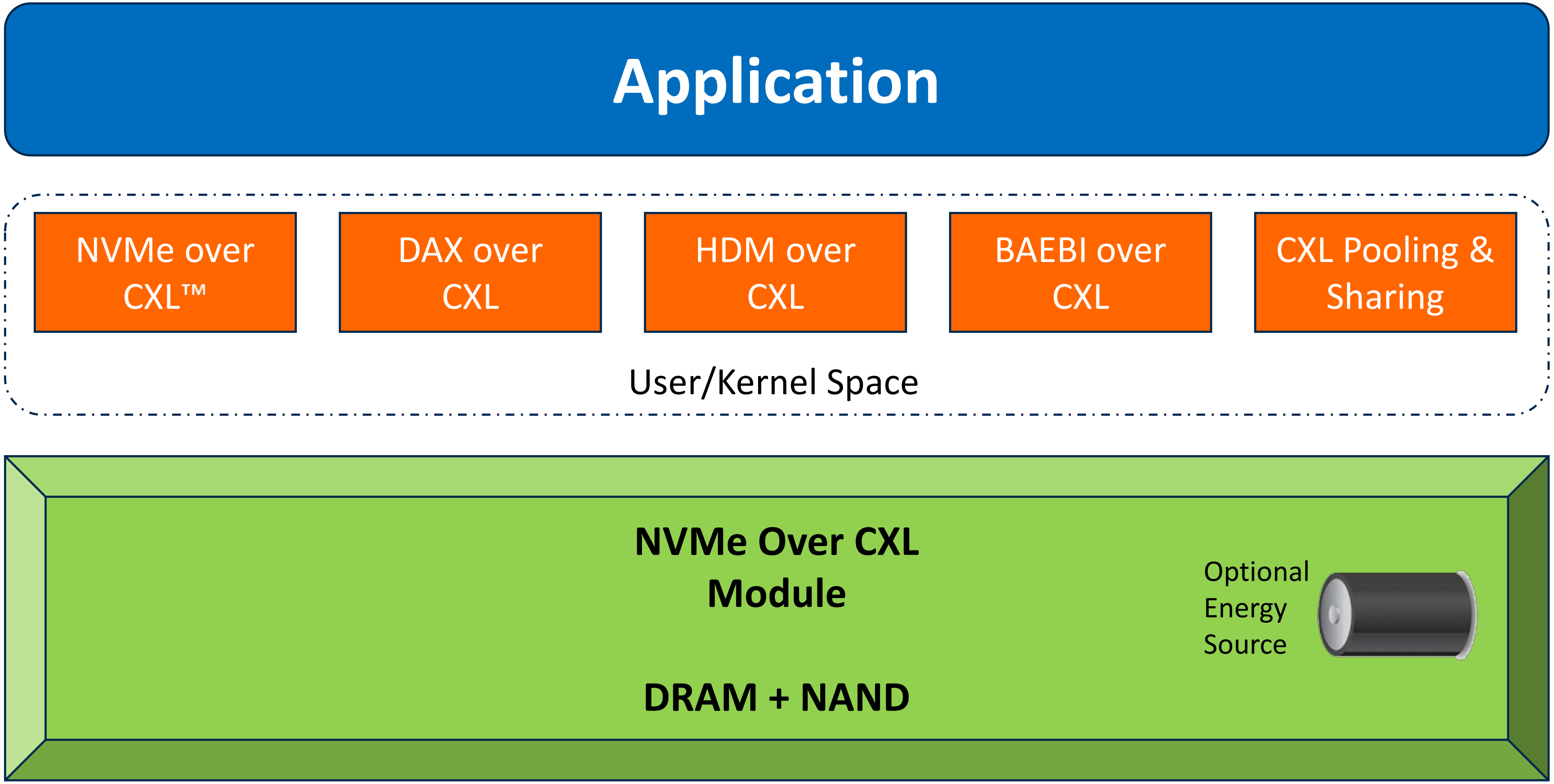
NVMe-oC addresses the memory wall which limits AI

Always let the Host decide where data belongs



NVMe-oC reduces wasted data traffic over the fabric by 30x or more

NVMe-oC supports persistence, allowing checkpoint elimination



Application

NVMe over CXL™

DAX over CXL

HDM over CXL

BAEBI over CXL

CXL Pooling & Sharing

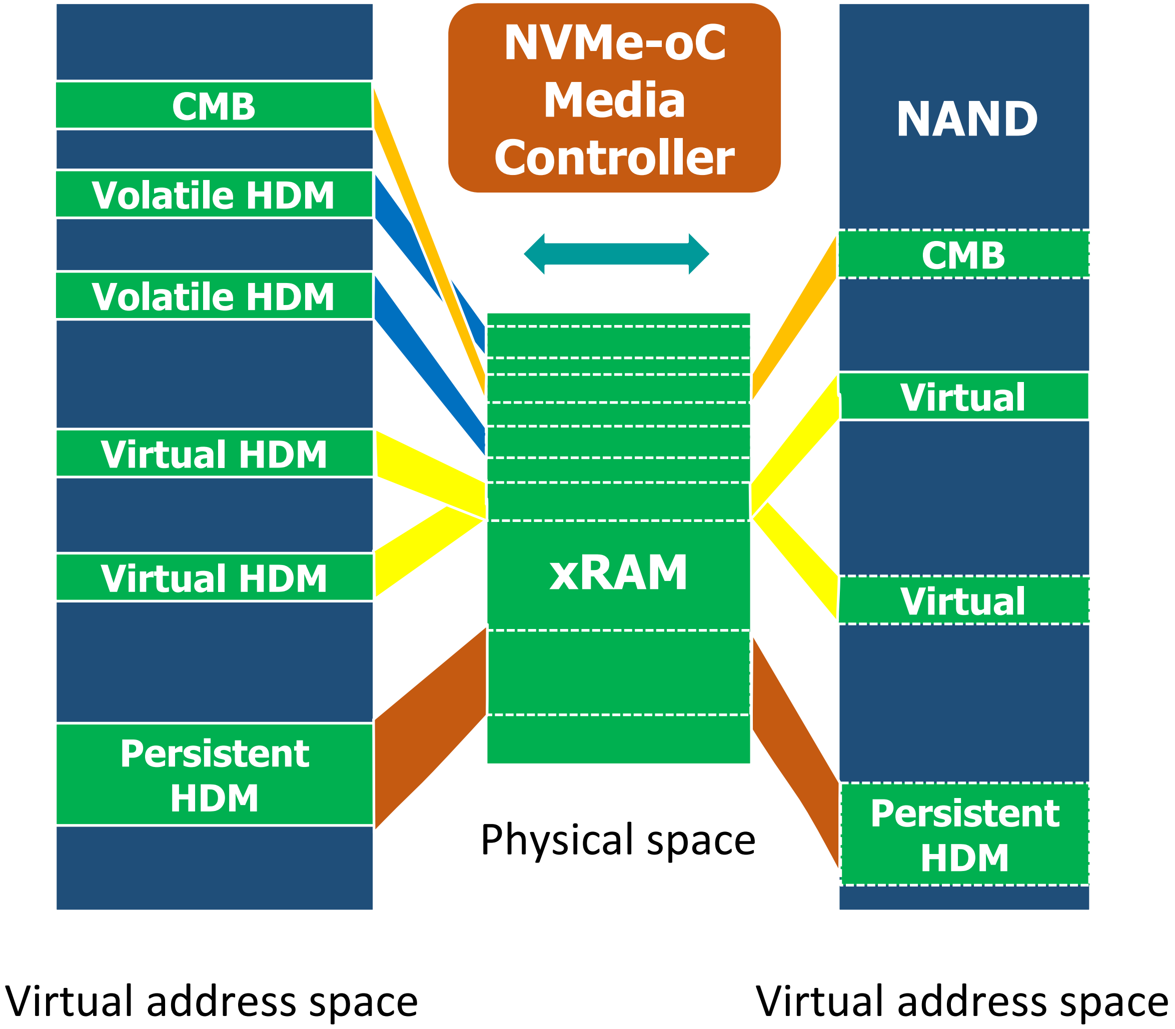
User/Kernel Space

NVMe Over CXL Module

DRAM + NAND

Optional Energy Source

Use existing software APIs where possible



NVMe-oC operates in all access modes simultaneously

xRAM always accessed as HDM

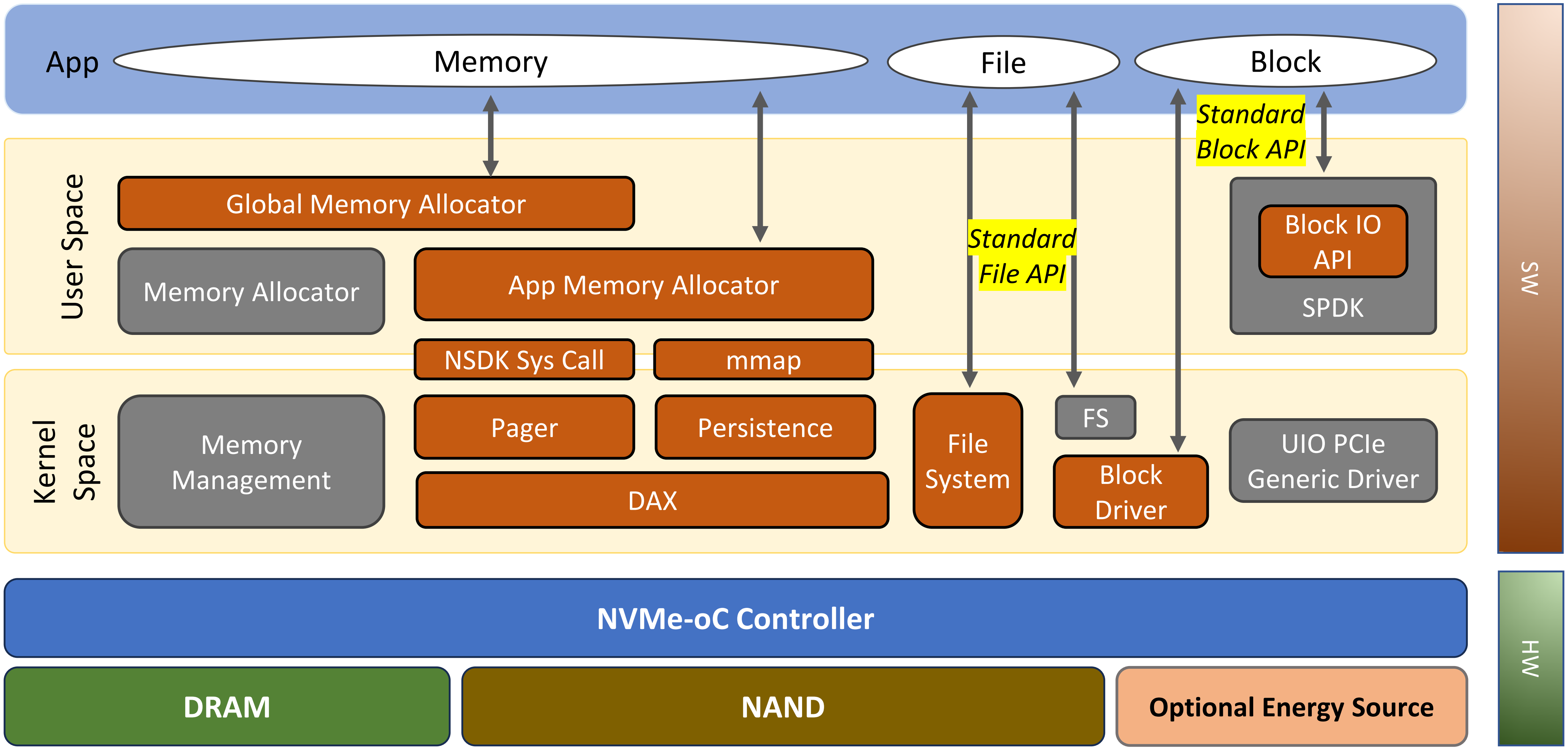
CMB, DAX, HDM all allowed

NAND to xRAM transfer schemes driven by host using NVMe commands

Persistence regions can be partial

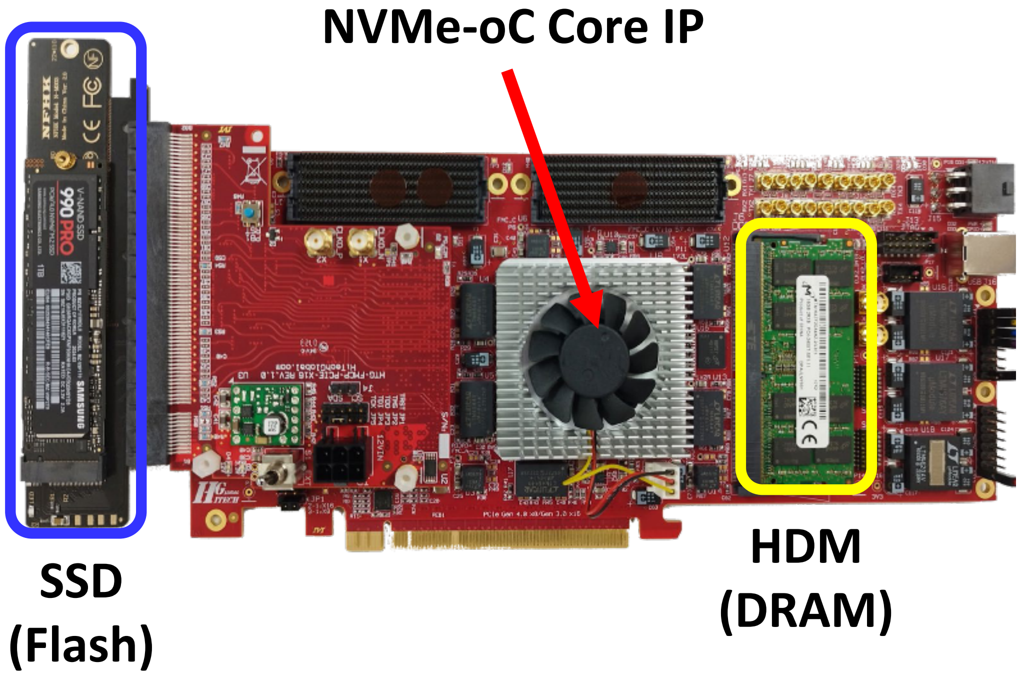
Cost savings from lower $/bit of NAND versus DRAM

NVMe-oC Software Development Kit (NSDK)



NVMe-oC Demonstration Platform

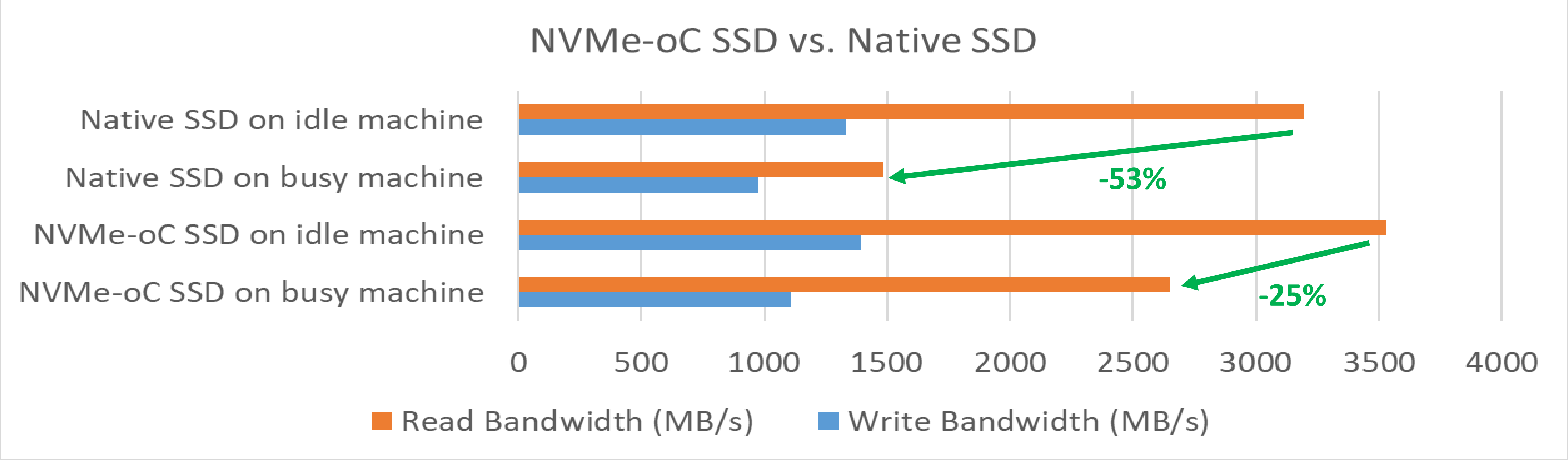
Device	
Host Interface	CXL 1.1/2.0 Gen3x8
HDM	16GB (DDR4-2000)
SSD	128GB ~ 1TB
System Clock	250MHz
NVMe	2.0
Operation Mode	Memory / Storage



Demonstrating Virtual HDM mode using NVMe-oC

Host	
CPU	Intel Granite Rapids, 2 processors, 288-cpu
Memory	128GB DRAM 6400MT
OS	Fedora release 40 (Forty)
Kernel	6.9.5

NVMe-oC Versus Traditional SSD Impact of Traffic Reduction

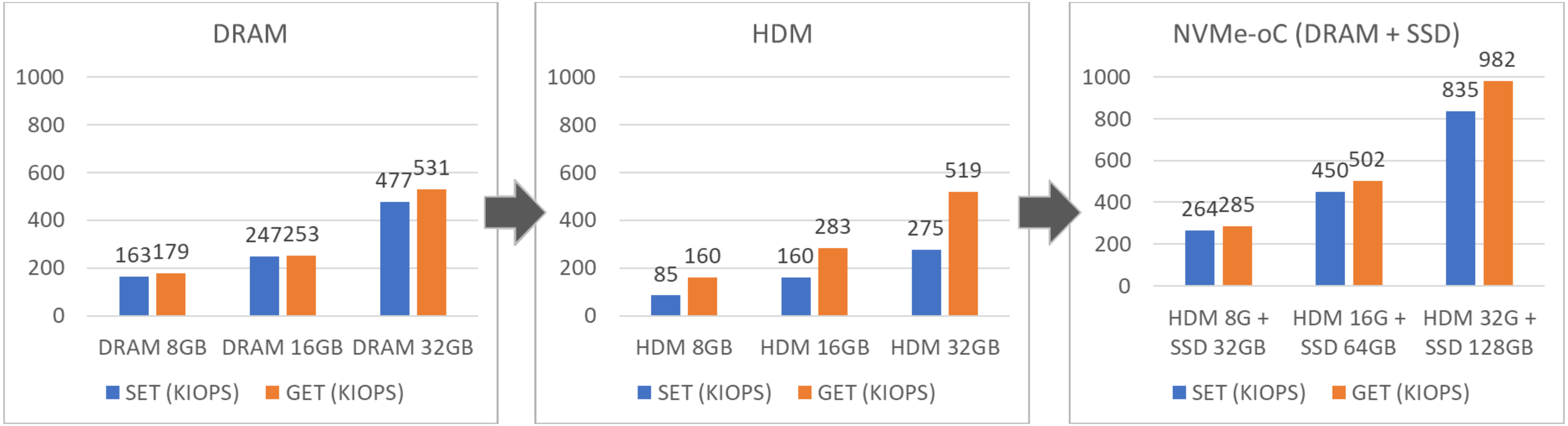


STREAM (memory benchmark) with 256 background threads

Results:

2.8X reduced impact on read performance

Virtual HDM Mode Redis Performance Versus HDM Impact of Large Memory Footprint



Unmodified Redis
(In-memory Key Value Store)

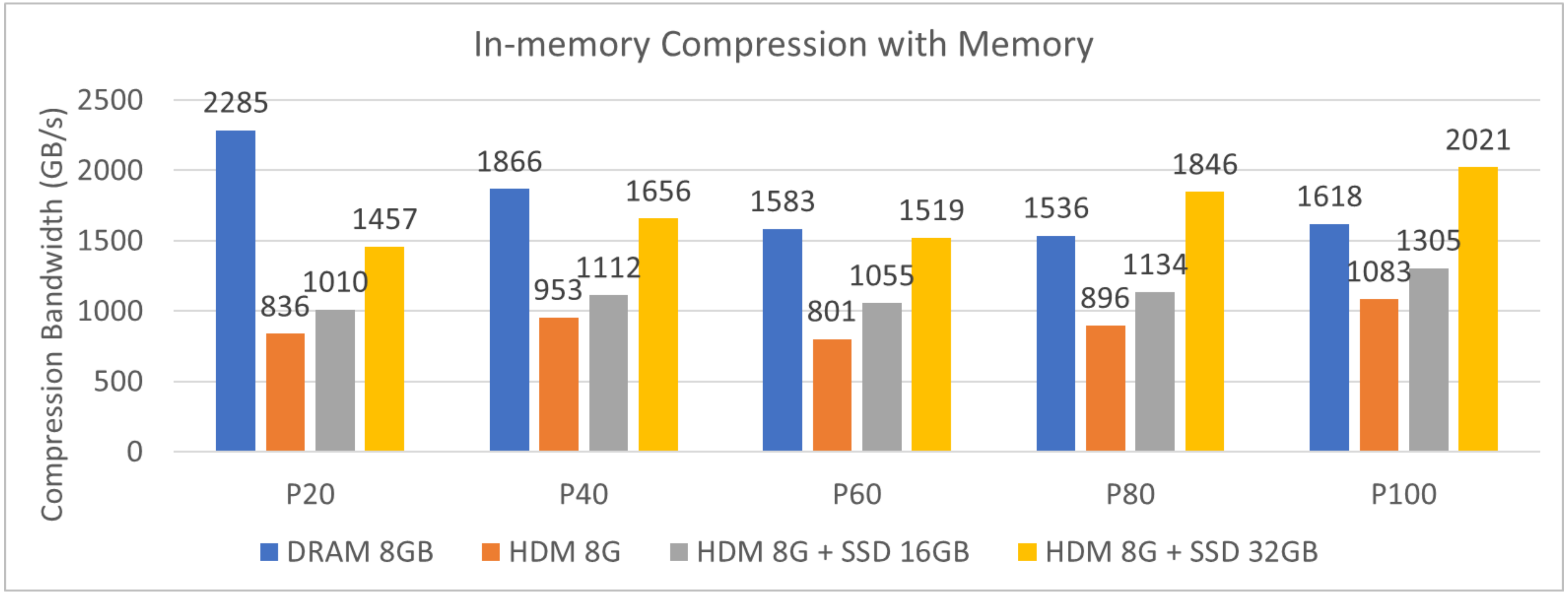
Results:

4X memory capacity

2X performance

90% cost reduction

Virtual HDM Mode Compression Performance Versus HDM Impact of Higher Thread Count



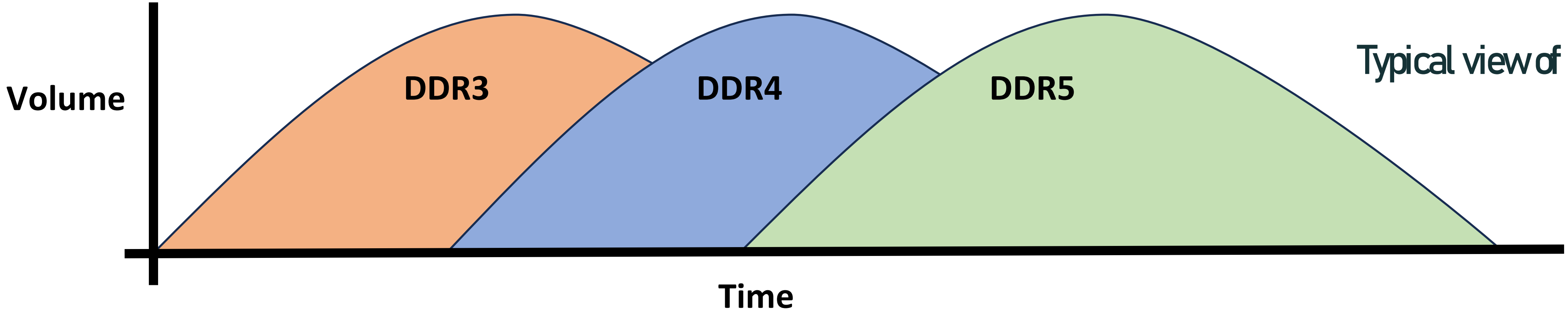
LZ77 lossless data compression method

Results:

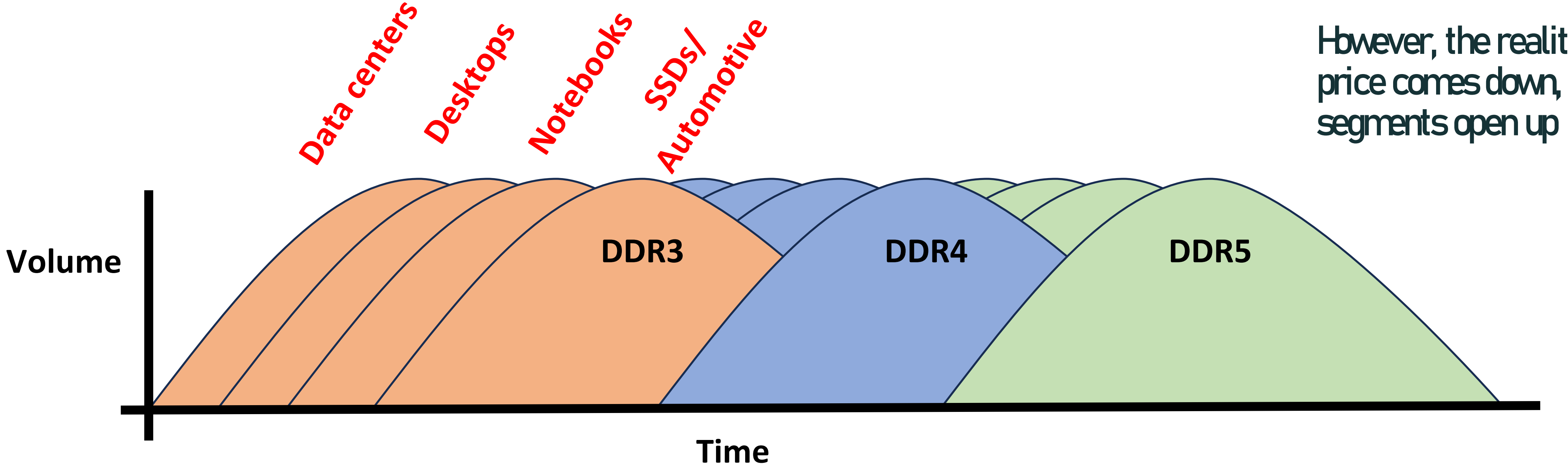
4X memory compression

Same performance as DRAM

More compression threads executed



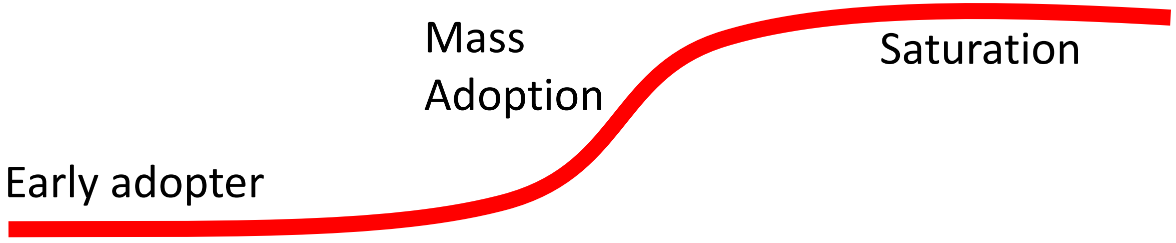
Typical view of the DRAM generations



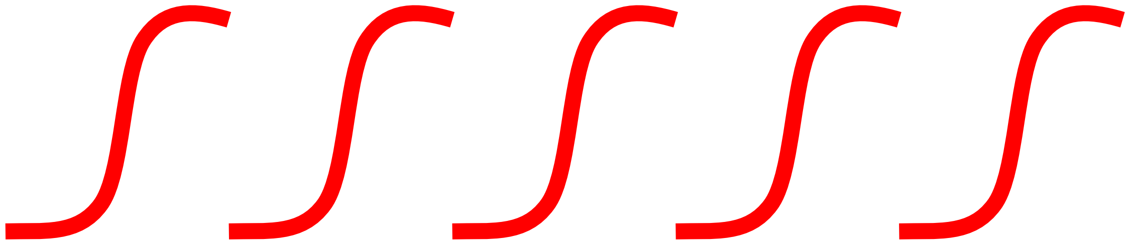
However, the reality is that as the price comes down, new market segments open up

It is essential for the success of CXL that it similarly follow broad adoption

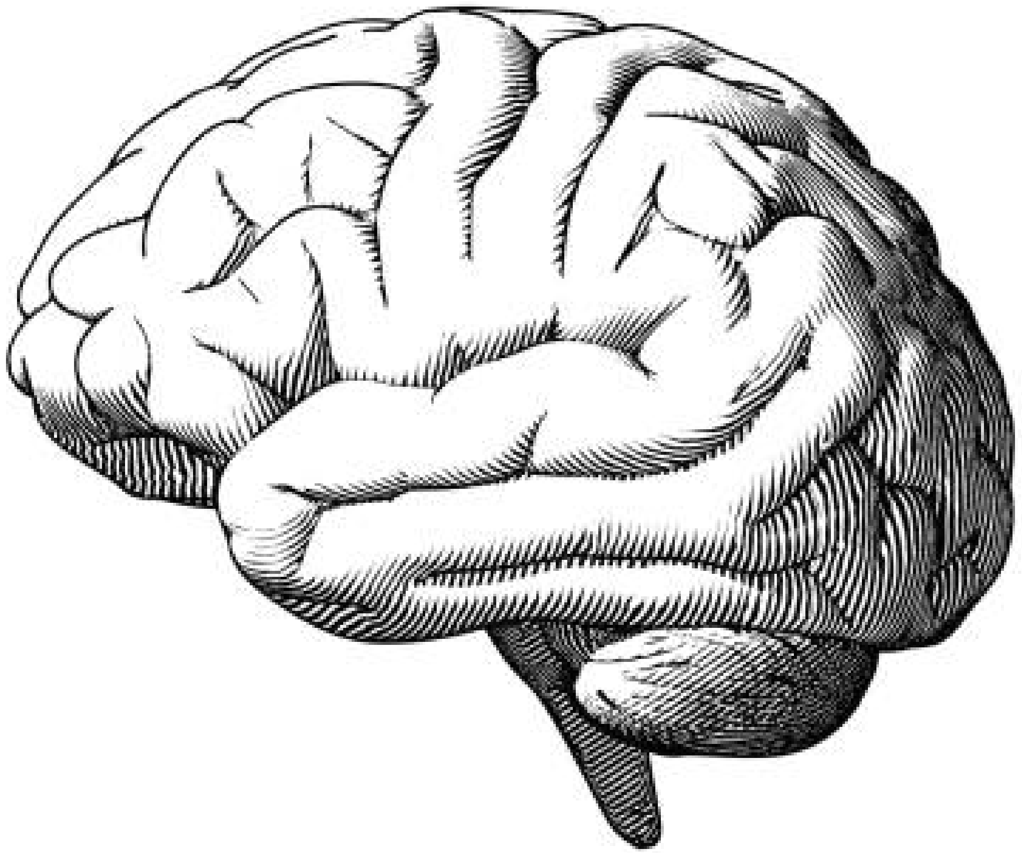
Maybe you’ve heard of artificial intelligence?



Typical tech S-curves are 7 years

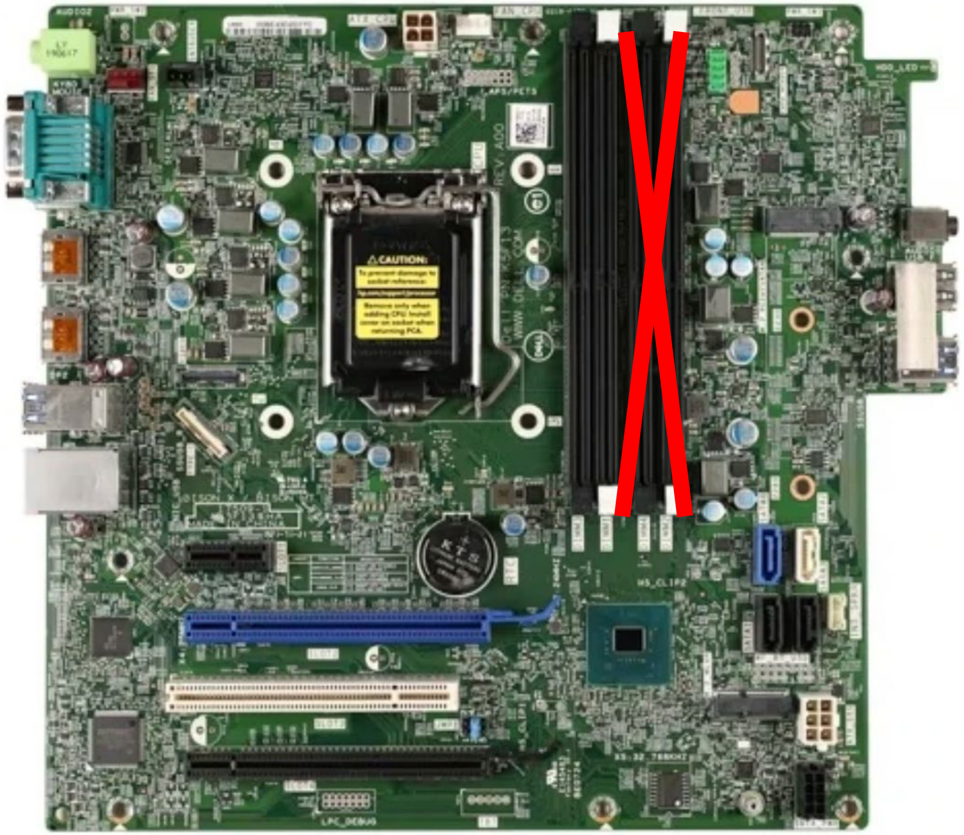


AI is on an 18 month S-curve
Who can keep up?



AI has an insatiable appetite for memory… 80GB to 240GB to 1.8TB and beyond

What is pulling on this rope?



DELL Technologies

DDR5 dropping down to one module per channel means system capacity is being cut in half – memory expansion on CXL can replace the lost DDR

AI on CXL can be upgraded

Industry innovation is stalled by closed designs

CXL on the motherboard is like PCI in 1992... opening the door to new ideas for next generation PCs

Summary

- CXL Enables Resource Virtualization
- NVMe Over CXL Virtualizes Memory and Storage
- Multi-mode: File, Memory, Persistence
- All Modes Available Simultaneously
- Always Use Host Direction Over Standard APIs
- Reduced Fabric Traffic Improves System Perf
- Virtual HDM Mode Improves Memory Hungry Apps
- Leverages Low Cost Per Bit of NAND to Reduce Cost
- CXL must grow beyond the data center

Thank you for your time

Bill Gervasi, Principal Systems Architect
bilge@wolleytech.com

Please take a moment to rate this session.

Your feedback is important to us.