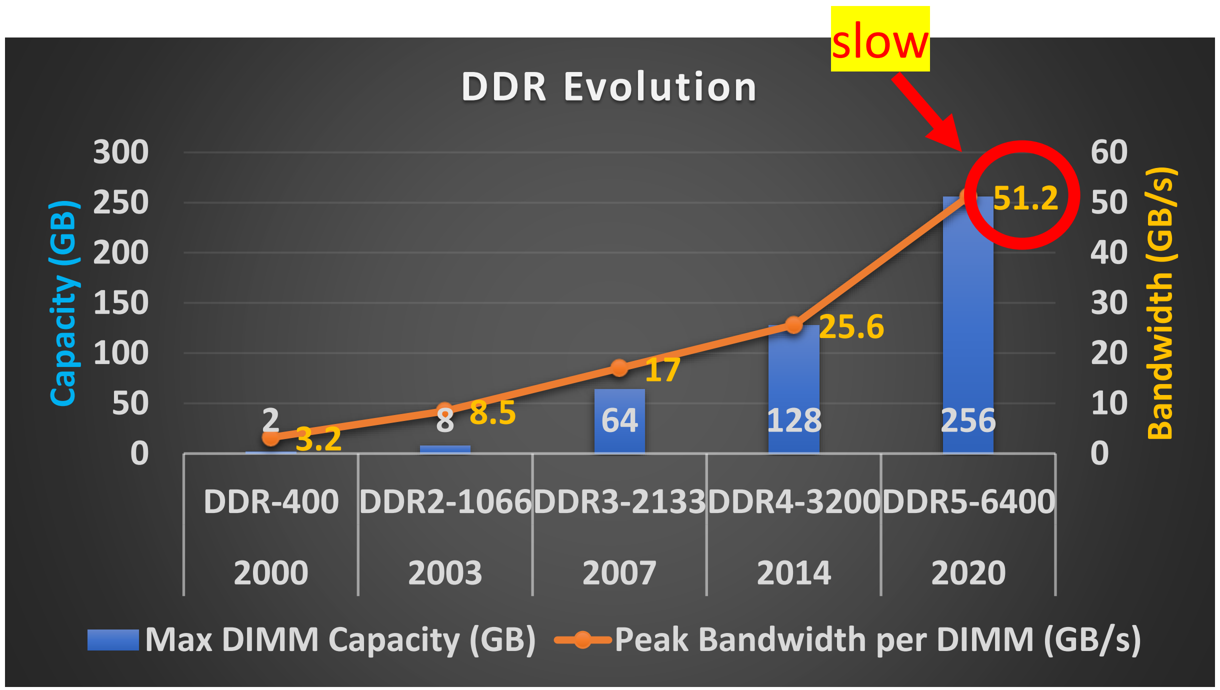
System Test Results with NVMe-over-CXL (NVMe-oC) Memory Mode

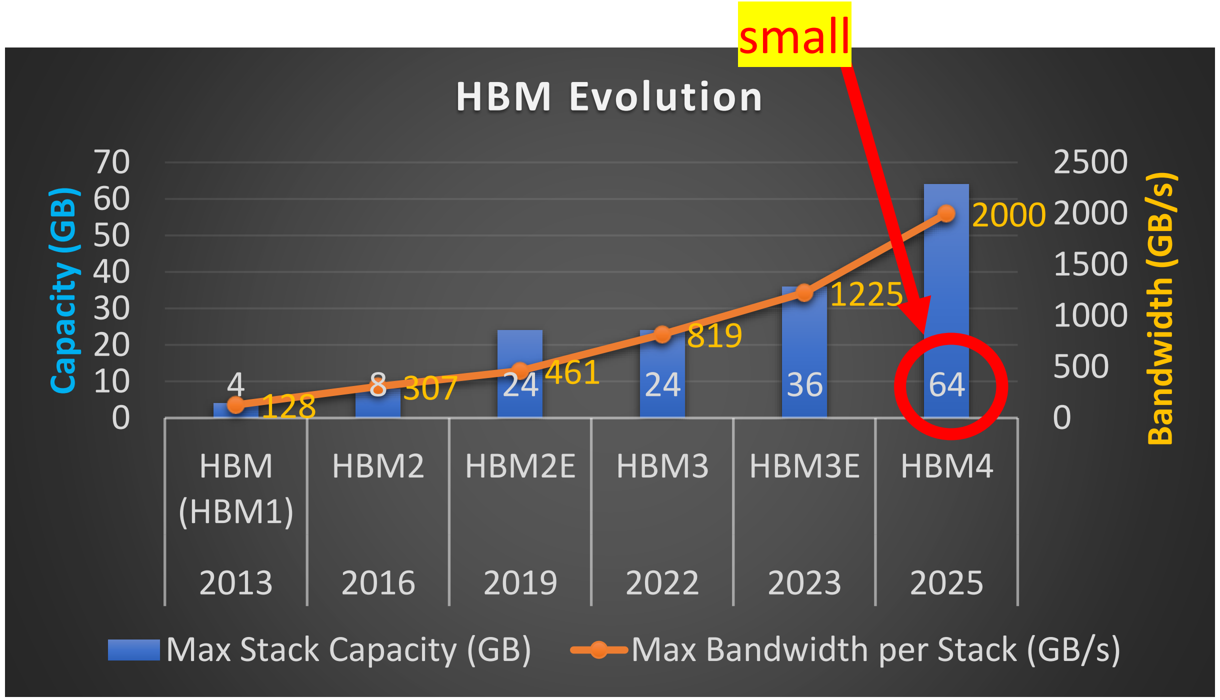
San Chang and Bernard Shung
Wolley Inc.

©2025 Conference Concepts, Inc. All Rights Reserved

Memory Is Not Scaling Fast Enough

- DRAM bandwidth grows slowly; HBM capacity gains are incremental
- Emerging workloads such as AI demand memory systems far beyond today's limits



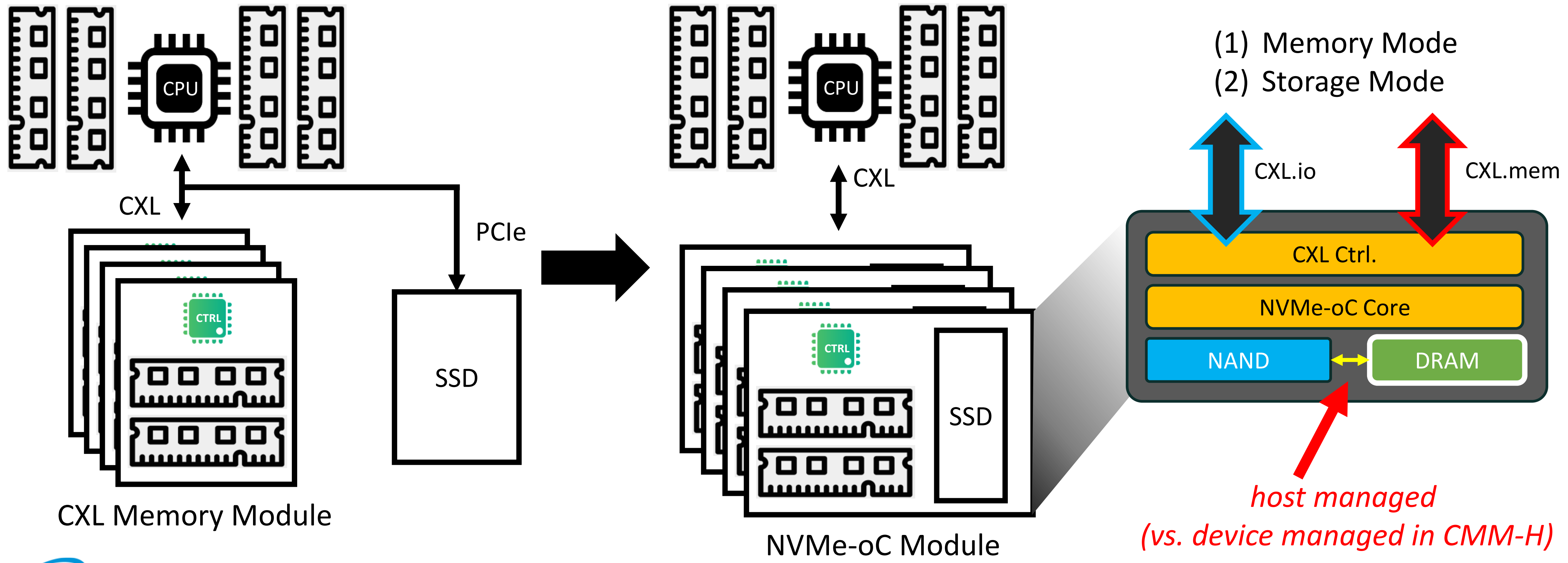


the Future of Memory and Storage

©2025 Conference Concepts, Inc. All Rights Reserved

What is NVMe-over-CXL (NVMe-oC)?

“Combines DRAM and NAND into a unified and CXL-attached memory module”

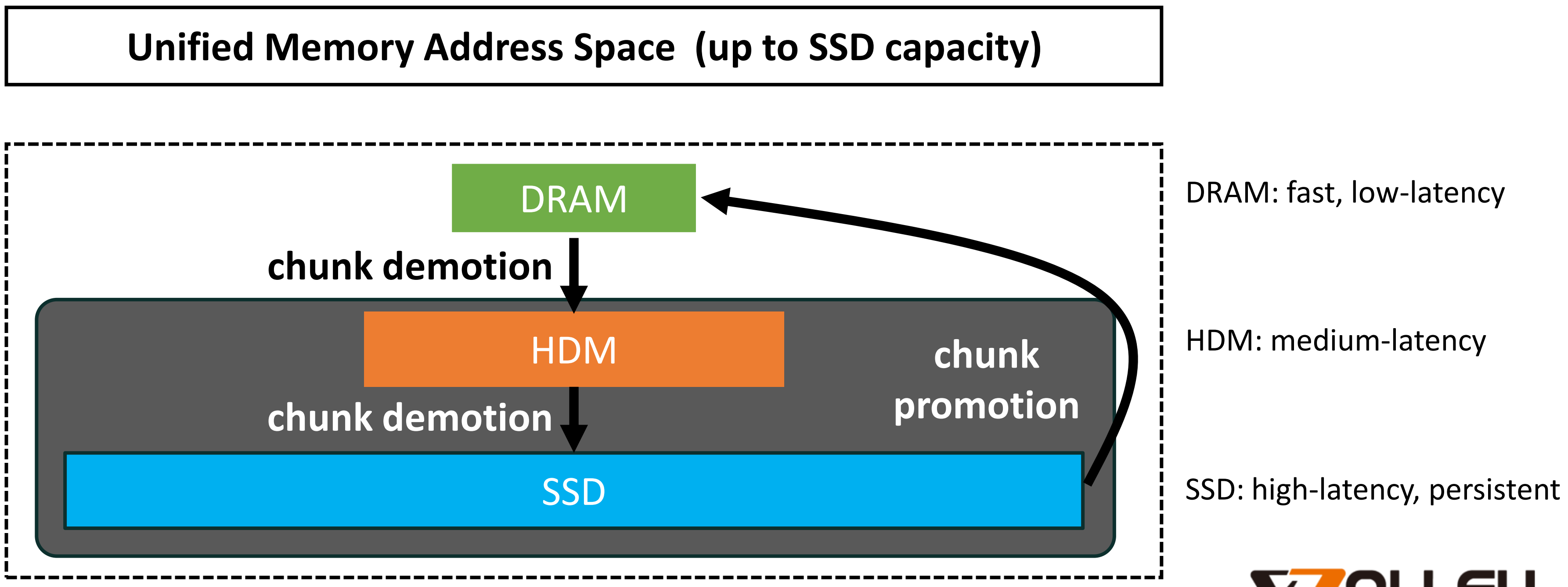


©2025 Conference Concepts, Inc. All Rights Reserved

NVMe-oC Memory Mode

- Memory Access (load/store) will either fall into either DRAM or HDM
- Hierarchical Caching with Unified Address Space
  - DRAM <-> HDM: exclusive cache pair (no duplicates, only one valid copy)
  - HDM <-> SSD: write-back strategy

(powered by DAX-tiering)



the Future of Memory and Storage

©2025 Conference Concepts, Inc. All Rights Reserved

Experiment Setup & Benchmark

Item	Description
CPU	Intel Xeon (Model 173), 144 cores (72 cores per socket × 2 sockets)
Memory	2 DDR5 RDIMM, 6400 MT/s, 32GB each (total bandwidth: 100GB/s)
NVMe-oC	PCIe Gen3 x8 (8GB/s), 16GB HDM + 64GB SSD
OS	Fedora Linux 40 (Workstation Edition)
kernel	6.8.5-301.fc40.x86_64

Capacity: 64GB

NVMe-oC Memory Mode Configurations

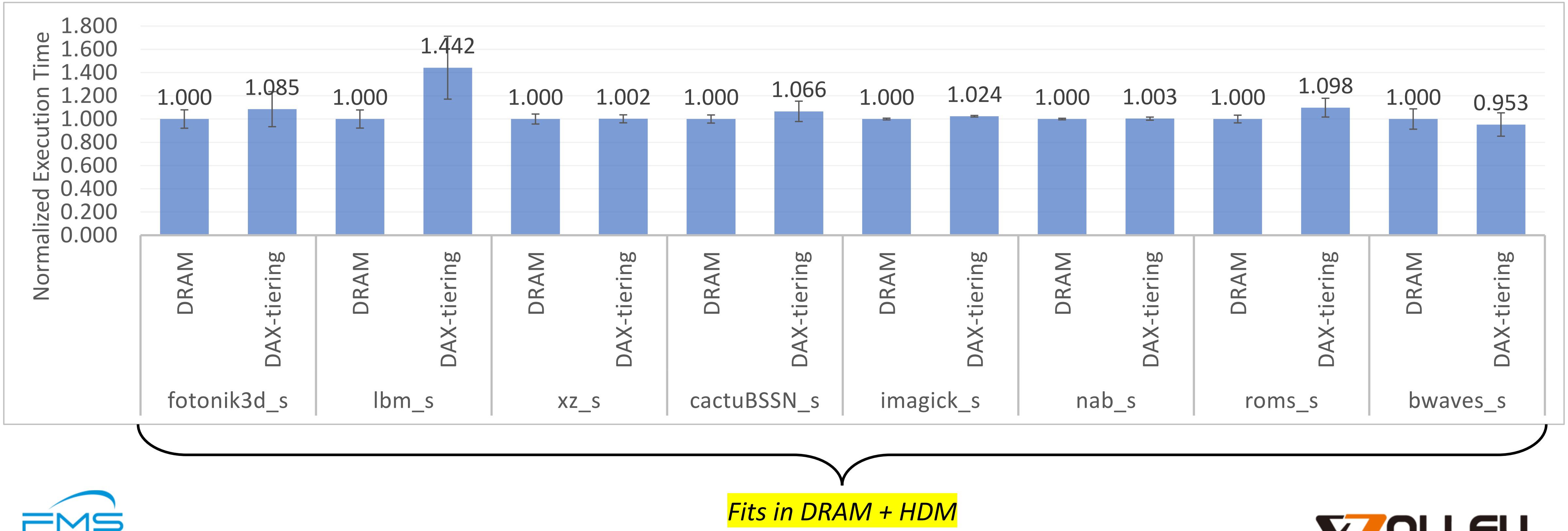
DAX-tiering	16GB DRAM + 16GB HDM + 64GB SSD

Capacity: 64GB

Benchmark Suite	Domain	Workload Style
SPEC 2017	CPU & memory compute	Single/multi-core integer/FP jobs
multichase/multiload	Memory subsystem	STREAM-like copy/compute/write microkernel with NT stores
XSBench	Scientific computing / HPC	Irregular memory access patterns
YCSB	Cloud NoSQL / key-value stores	Configurable DB access workloads

SPEC 2017 – DRAM vs. NVMe-oC Memory

- NVMe-oC achieves near-DRAM performance with average CPU stall of ~1.08 (vs. DDR5: 1.0, CMM-H: ~1.7 based on Samsung paper), delivering lower stall time and outperforming CMM-H



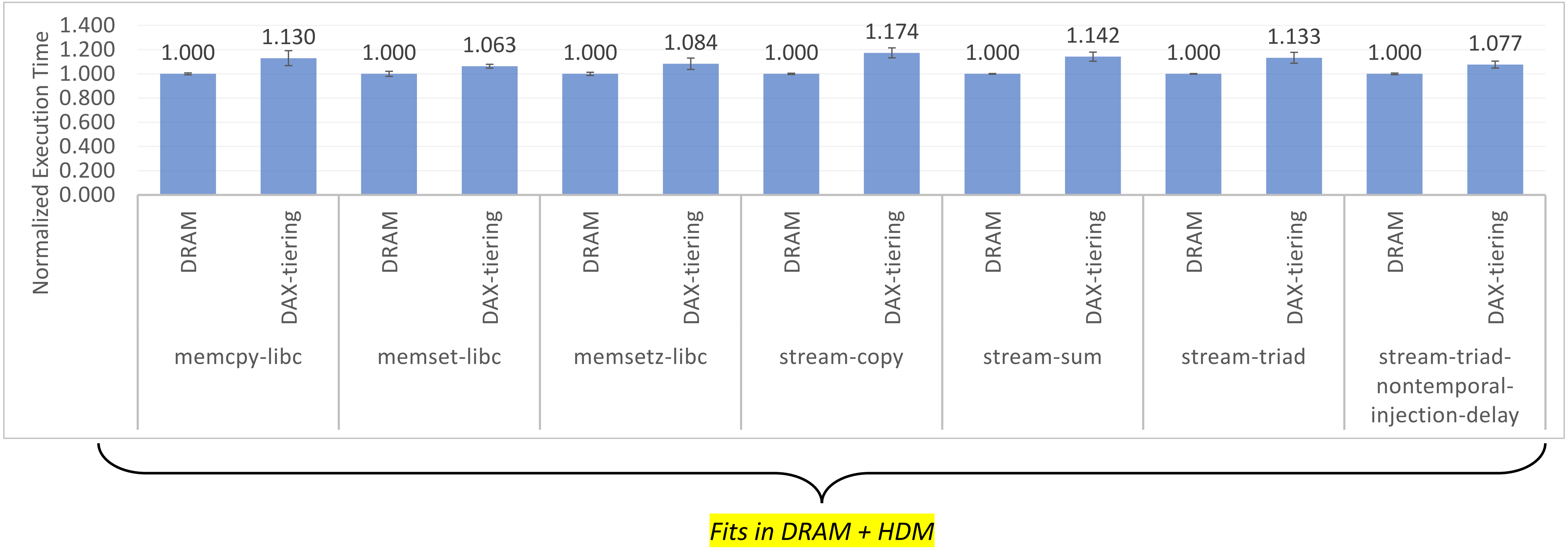
FMS the Future of Memory and Storage

©2025 Conference Concepts, Inc. All Rights Reserved

VOLLEY

multichase/multiload – DRAM vs. NVMe-oC Memory

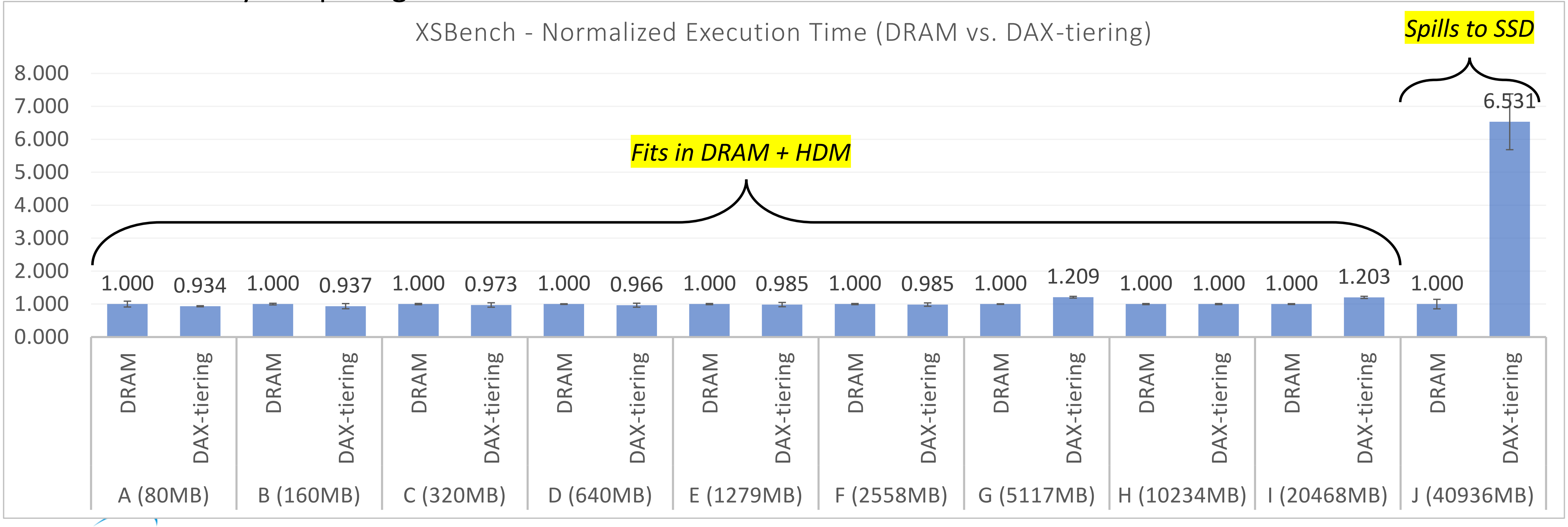
Able to provide near-DRAM performance when the working set fits within DRAM + HDM



©2025 Conference Concepts, Inc. All Rights Reserved

XSBench (DRAM vs. NVMe-oC DAX-tiering)

- Our NVMe-oC with DAX-tiering maintains near-DRAM performance up to workload I, while CMM-H, without host DRAM support, slows down by 1.58× to 5.56× as the memory footprint grows from 80 MB to 40 GB



FMS the Future of Memory and Storage

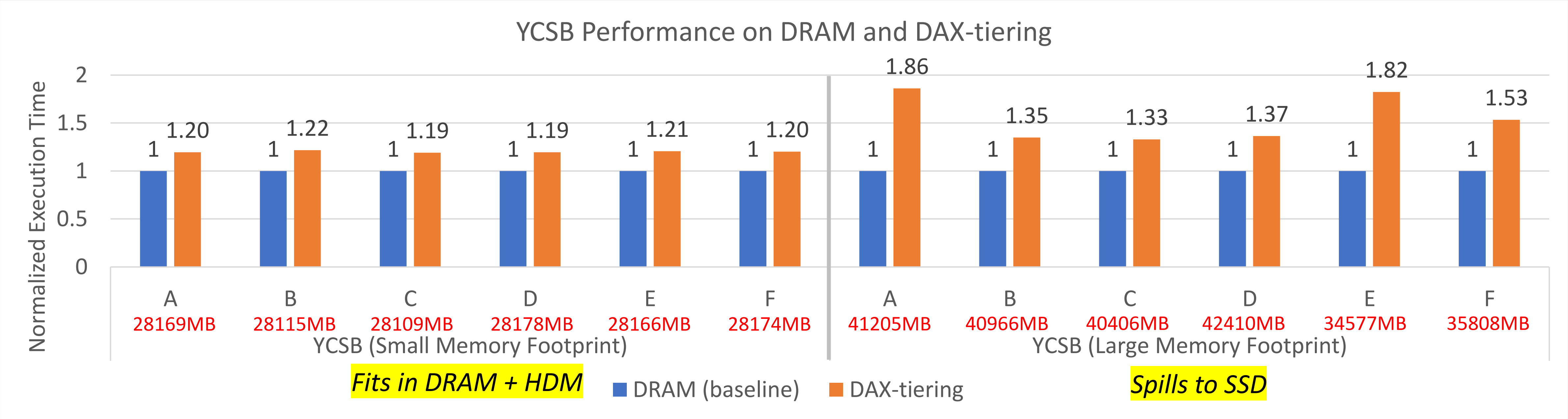
©2025 Conference Concepts, Inc. All Rights Reserved

VOLLEY

YCSB Performance (DDR5 vs. NVMe-oC DAX-Tiering)

- Workload A (Update Heavy, 50% read / 50% update)
- Workload B (Read Heavy, 95% read / 5% update)
- Workload C (Read Only, 100% read)
- Workload D (Read-Modify-Write, 95% read / 5% insert)
- Workload E (Scan Heavy, 95% scan / 5% insert)
- Workload F (Read-Modify-Write, 50% read / 50% RMW)

- **DAX-tiering delivers 83% and 65% of DDR5 performance** under small and large memory footprints respectively, while costing less than half of DDR5. This results in **performance-per-dollar gains of 67% (small footprint) and 30% (large footprint)**
  - Small memory footprint workloads (fits in DRAM + HDM) show an average of 1.2× slowdown
  - Large memory footprint workloads (spill over to SSD) show an average of 1.54× slowdown

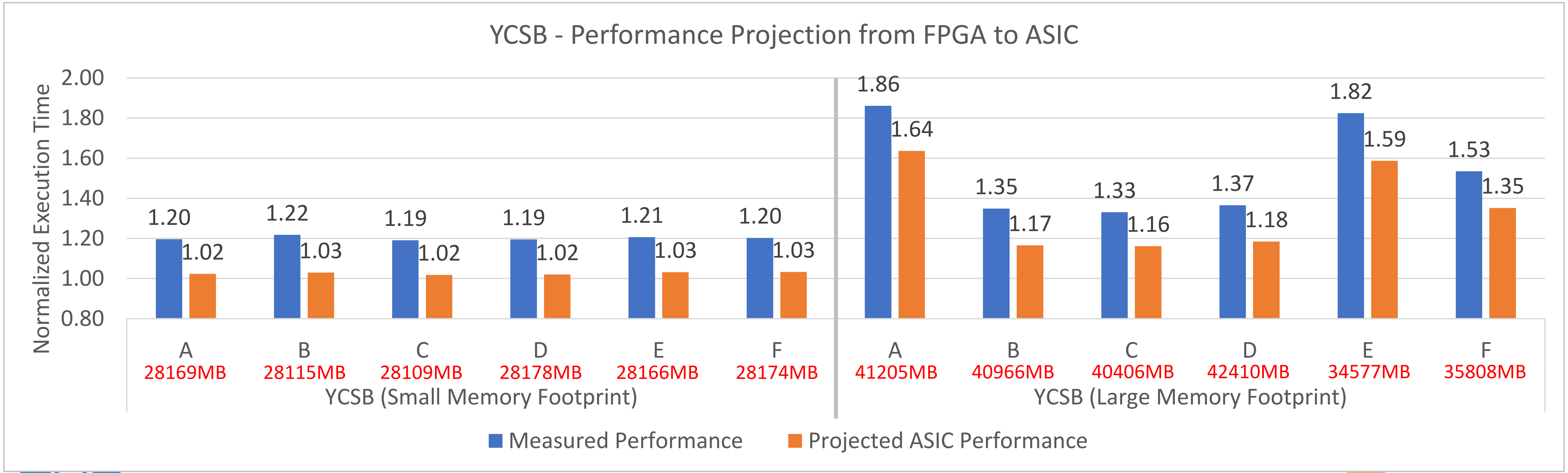


©2025 Conference Concepts, Inc. All Rights Reserved

YCSB Performance Projection from FPGA to ASIC

- Workload A (Update Heavy, 50% read / 50% update)
- Workload B (Read Heavy, 95% read / 5% update)
- Workload C (Read Only, 100% read)
- Workload D (Read-Modify-Write, 95% read / 5% insert)
- Workload E (Scan Heavy, 95% scan / 5% insert)
- Workload F (Read-Modify-Write, 50% read / 50% RMW)

- With ASIC, **DAX-tiering delivers 97% (vs. 83%) and 74% (vs. 65%) of DDR5 performance** under small and large footprints, at less than half the cost. This leads to the **performance-per-dollar gains of 94% (small footprint) and 48% (large footprint)**



FMS the Future of Memory and Storage

©2025 Conference Concepts, Inc. All Rights Reserved

VOLLEY

Performance Comparison (NVMe-oC vs. CMM-H)

- NVMe-oC (with 16GB DDR4-2000)
  - (SPEC CPU2017) NVMe-oC, which combines host DRAM and device DRAM as cache, shows an average 8.4% slowdown compared to DDR5-L
  - (XSBench) Performance degrades by 1.0× to 6.5× as the memory footprint increases from 80 MB to 40 GB
- CMM-H (with 16GB DDR4-2666)
  - (SPEC CPU2017) shows an average 70% performance degradation compared to DDR5-L (local DRAM)
  - (XSBench) Performance degrades by 1.58× to 5.56× as the memory footprint increases from 80 MB to 40 GB

Zeng, Jianping, et al. "Performance Characterizations and Usage Guidelines of Samsung CXL Memory Module Hybrid Prototype." *arXiv preprint arXiv:2503.22017* (2025).

©2025 Conference Concepts, Inc. All Rights Reserved

Feature Comparison (NVMe-oC vs. CMM-H)

	CMM-H	NVMe-oC
Host Interface	PCIe Gen5x8	PCIe Gen5x8
Media	DRAM + SSD	DRAM + SSD
Memory Mode Capacity	Up to SSD total capacity	Up to SSD total capacity
Memory Mode Namespace	Single	Multiple (user configurable) ✅
HDM Management in Memory Mode	Device Cache Controller	Host CPU
Dual Mode (Memory/Storage) Online	NA	Support ✅
Insert Application Intelligence	Limited	Support ✅
Direct SSD Data to Host DRAM (Hot Data)	NA	Support ✅
Dual-cache (Host DRAM + Device DRAM)	NA	Support ✅
Data Prefetching	Device Cache Controller	Support ✅
Hot/Cold Data Detection	Device Cache Controller	Driver (TRACE)
Cache Policy	8-way, 4KB management, LRU, MRU (HW implementation)	N-way, M KB management, LRU, MRU, etc., (software defined)

©2025 Conference Concepts, Inc. All Rights Reserved

Takeaways

- NVMe-oC memory mode enables cost-effective memory expansion over CXL for data-intensive workloads
  - Jointly manages host DRAM, device DRAM (HDM) and NAND with flexible caching
  - No code changes to the existing applications
- NVMe-oC memory mode value proposition
  - Provides near-DRAM performance when the working set fits within DRAM plus HDM, serving as a standard CXL memory module for memory expansion
  - Delivers better performance/$ (30~90% improvement) when the working set exceeds DRAM plus HDM, outperforming CMM-H through host-managed caching intelligence

San Chang, san@wolleytech.com

©2025 Conference Concepts, Inc. All Rights Reserved