

Enabling CXL within the Data Center with Arm Solutions



EMPOWERING OPEN.



OCP
GLOBAL
SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA





Agenda

- Disaggregation of compute and memory
- CXL Product Enablement - Host and Device perspective
- Summary



OCP
GLOBAL
SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA

EMPOWERING OPEN.

Market Dynamics - Disaggregation of compute and memory



Inefficiency of DRAM memory utilization



Memory channel bandwidth per core declining



Reduce Total Cost of Ownership (TCO)



PCIe speeds equating single channel DRAM bandwidth



Workloads becoming more and more divergent



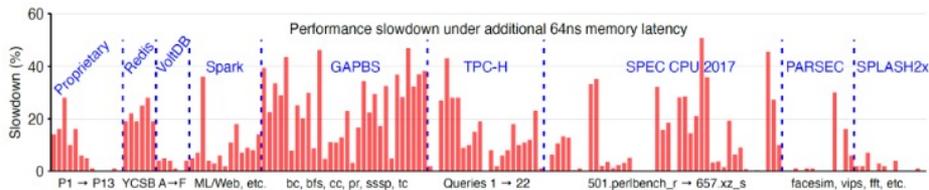
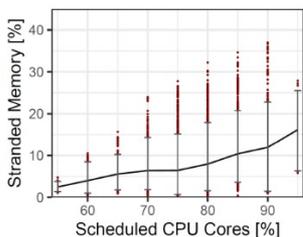
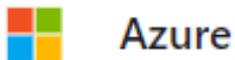
OCP
GLOBAL
SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA

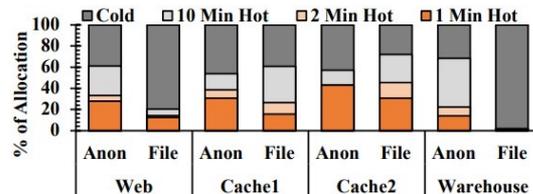
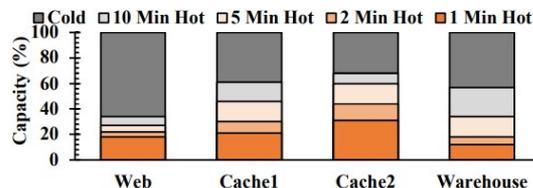
EMPOWERING OPEN.



Hyperscaler's view on CXL



Source: <https://arxiv.org/pdf/2203.00241.pdf>



Source: <https://www.nextplatform.com/2022/06/16/meta-platforms-hacks-cxl-memory-tier-into-linux/>

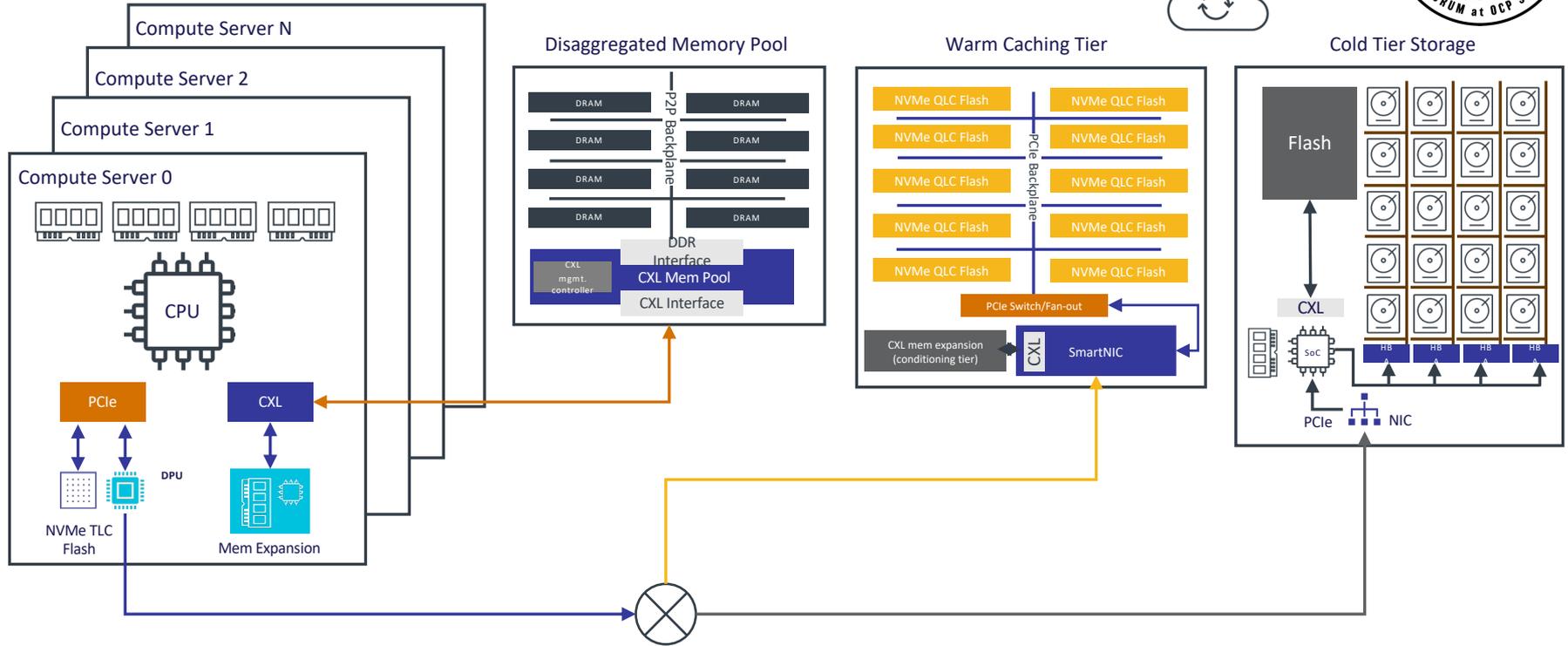
20% of workloads show NO performance impact due to additional CXL latency



OCTOBER 18-20, 2022
SAN JOSE, CA

EMPOWERING OPEN.

Emerging Datacenter Architecture with CXL



Network topology Data Center specific

CXL Product Enablement – Host Perspective



OCP
GLOBAL
SUMMIT

EMPOWERING OPEN.

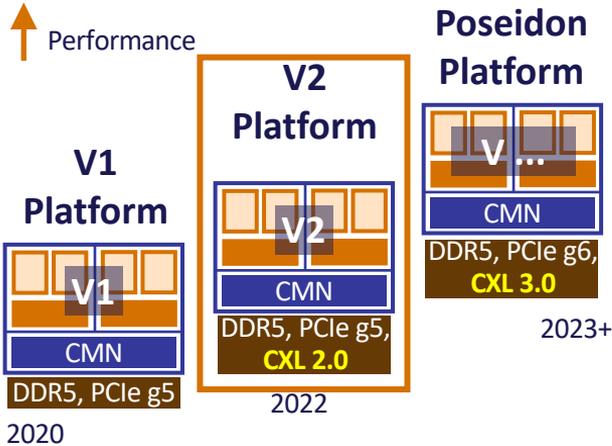
OCTOBER 18-20, 2022 | SAN JOSE, CA

Arm is enabling Rapid Pace of Innovation with CXL



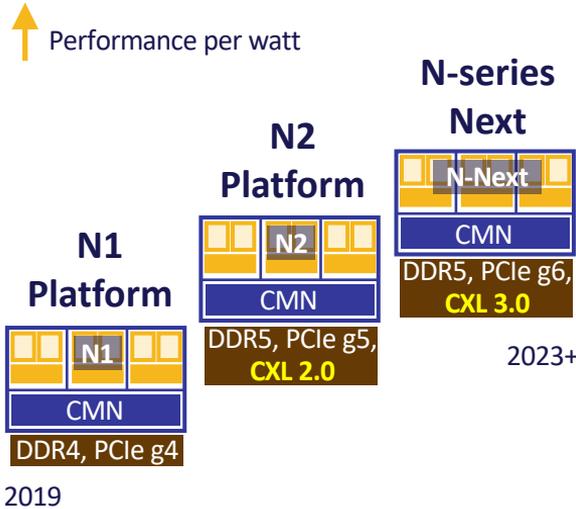
V-series

Maximum Performance and Optimal TCO



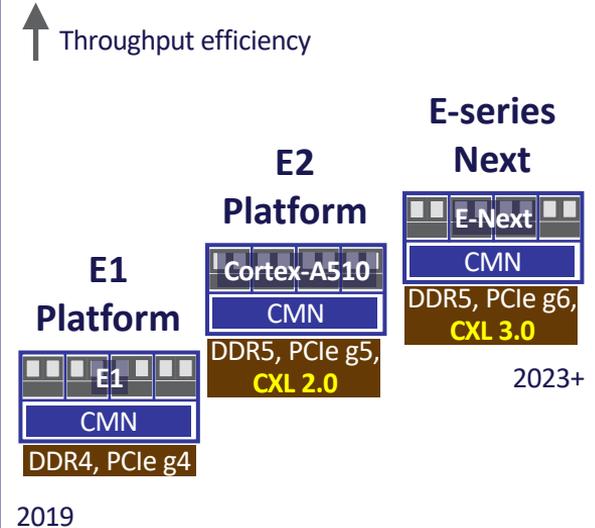
N-series

Efficient Performance



E-series

Efficient Throughput



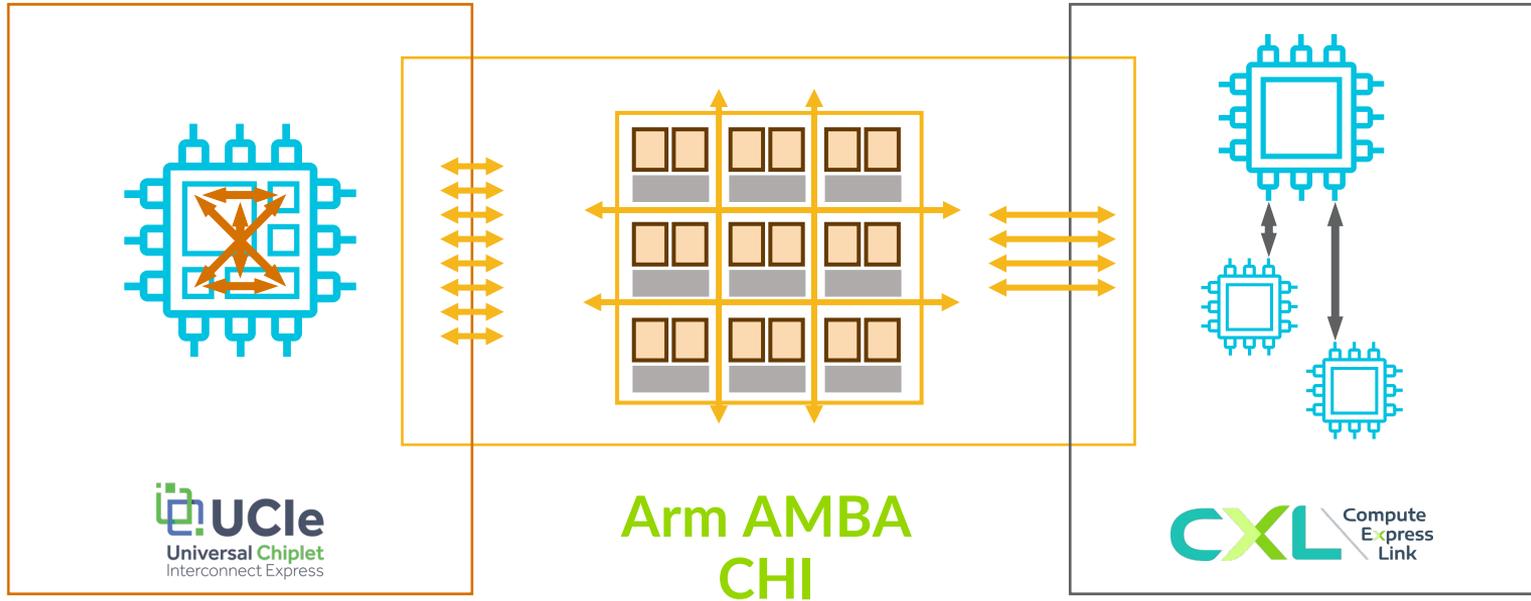
CMN: Coherent Mesh Network
CXL: Compute Express Link

Arm Neoverse CPUs along with CMN enable CXL

Arm is Driving Technology Leadership Across Standards



AMBA CHI, UCIe and CXL deliver leadership die-to-die and chip-to-chip fabric solutions



CMN: Coherent Mesh Network

CXL: Compute Express Link

CXL Product Enablement – Device Perspective



OCP
GLOBAL
SUMMIT

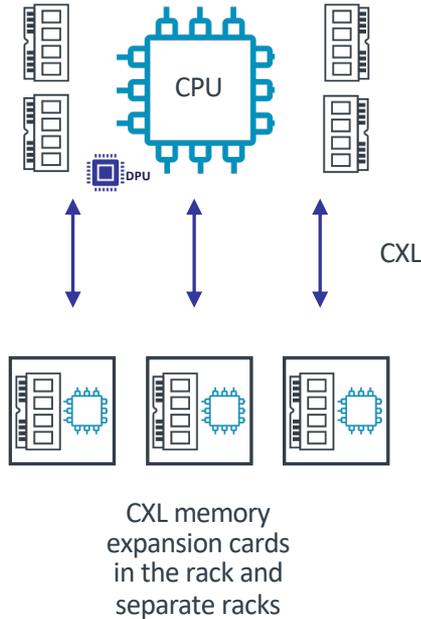
EMPOWERING OPEN.

OCTOBER 18-20, 2022 | SAN JOSE, CA

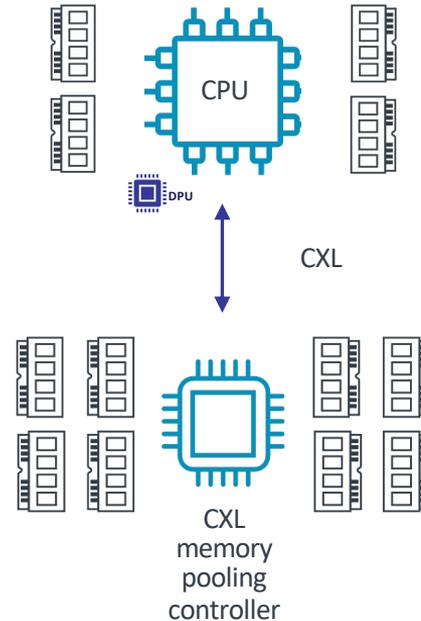
Solutions for Disaggregation of Compute and Memory



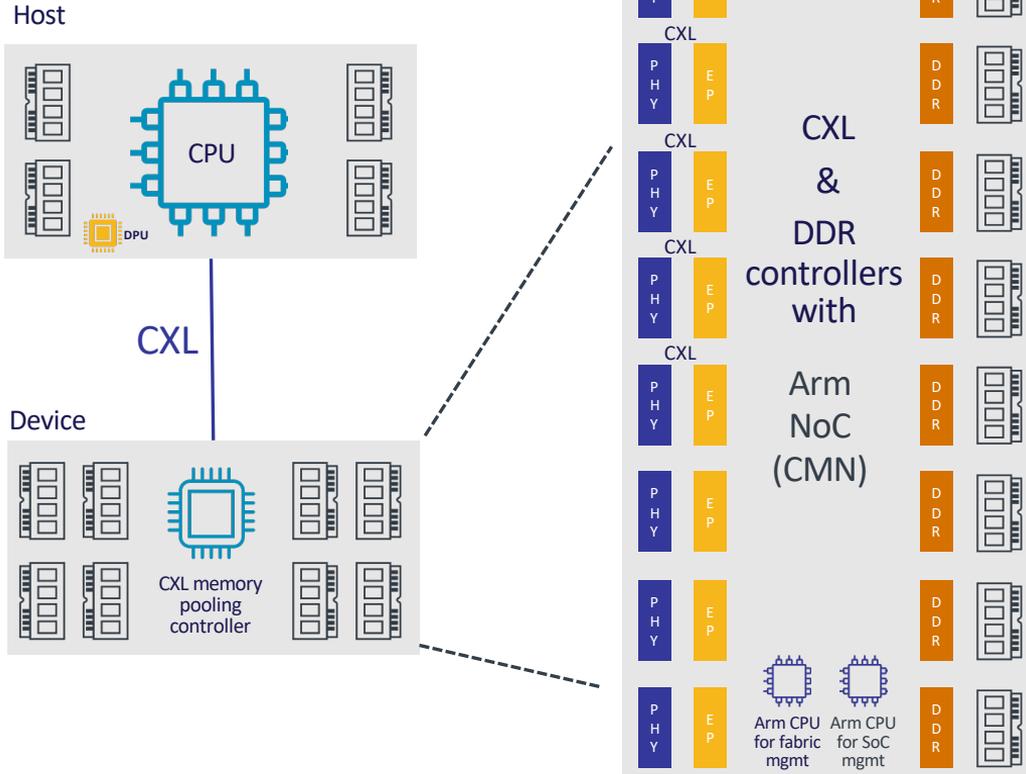
CXL Memory Expansion



CXL Memory Pooling



CXL Memory Pooling Solution

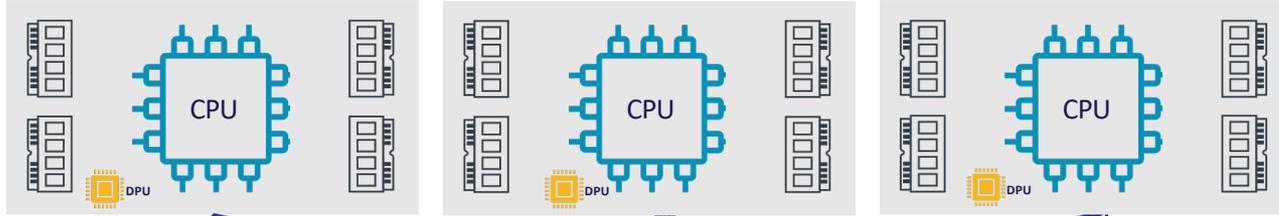


Arm is in a unique position to provide CXL memory pooling solution

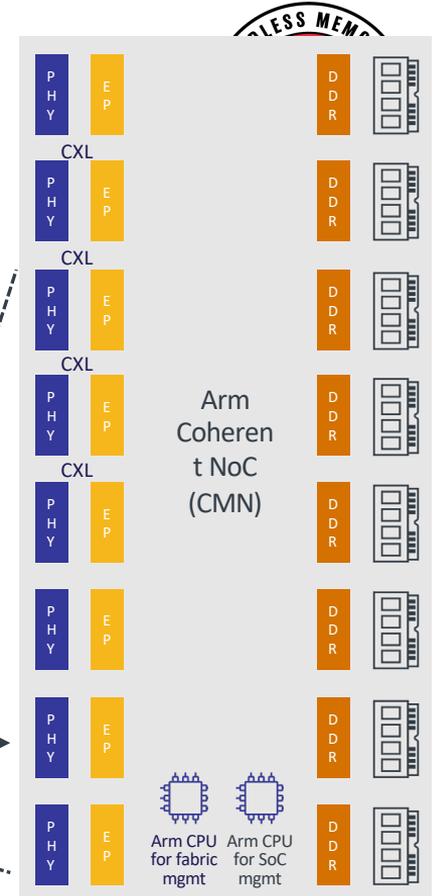
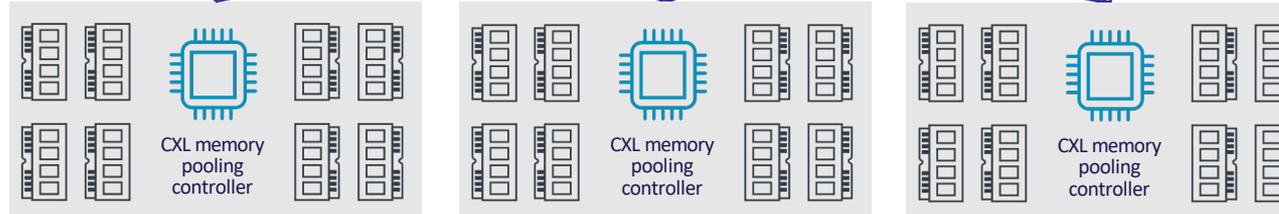
- Lowest latency Coherent Mesh Network (CMN) product
- Arm CPUs for Fabric and SoC management
- Provide end-end CXL optimized solutions with Host and Device side knowledge

CXL Memory Pooling at Scale

Host



Device



Arm is in a unique position to provide end-end CXL solution with Host and Device side knowledge

Economics Behind Memory Disaggregation

Disaggregated memory is worth doing, and benefit depends on optimization



Baseline

Jobs	Compute	Memory

Baseline performance and cost assumption:

- + 4GB/core
- + 25-50% of DRAM is unused/stranded*

Disaggregated memory

Jobs	Compute	Memory

+ DRAM cost

+ Average Latency

+ 3%* better performance/TCO

Disaggregated memory with ideal tiering

Jobs	Compute	Memory

+ DRAM cost

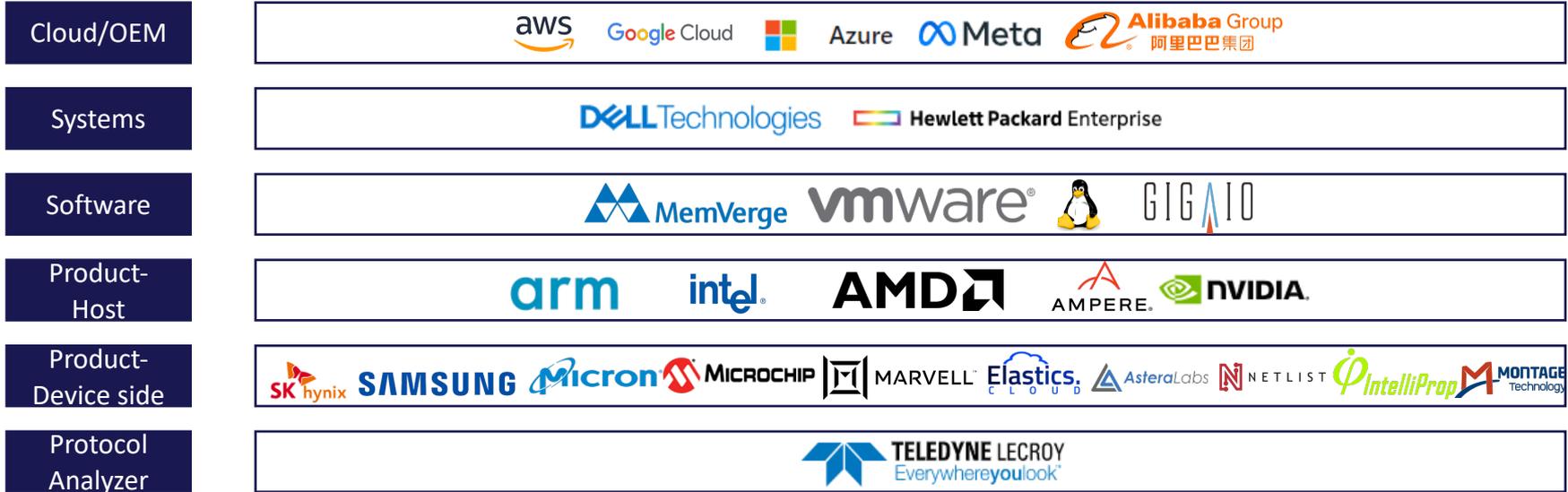
+ Average Latency

+ 26%* better performance/TCO

+ Block size
+ Migration rate
+ Migration Algo
+ Tiering cost

*Arm internal estimates based on public available information

CXL Ecosystem in development



Advantages on a partnership with Arm on CXL



Technology currently in development cycle

- Arm know-how in coherent device technology concepts due to CCIX efforts
- Arm has front-row seats to requirements via partners
- Arm has demonstrated ecosystem leadership in many segments

Ecosystem Enablement

- Arm has a presence in all major consortia giving unprecedented view of industry
- Arm neutrality critical for a fragmented ecosystem
- Arm is well versed in standardizing ecosystems



Thank you!



EMPOWERING OPEN.



OCP
GLOBAL
SUMMIT

OCTOBER 18-20, 2022
SAN JOSE, CA

