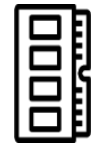
arm

Enabling CXL within the Data Center with Arm Solutions

Agenda

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

Market Dynamics - Disaggregation of compute and memory



Inefficiency of DRAM memory utilization

Memory channel bandwidth per core declining

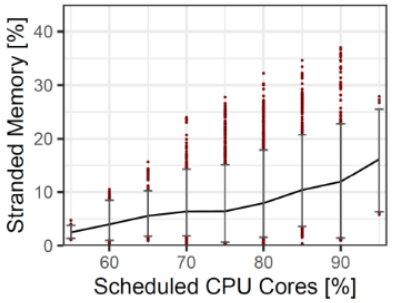
Reduce Total Cost of Ownership (TCO)

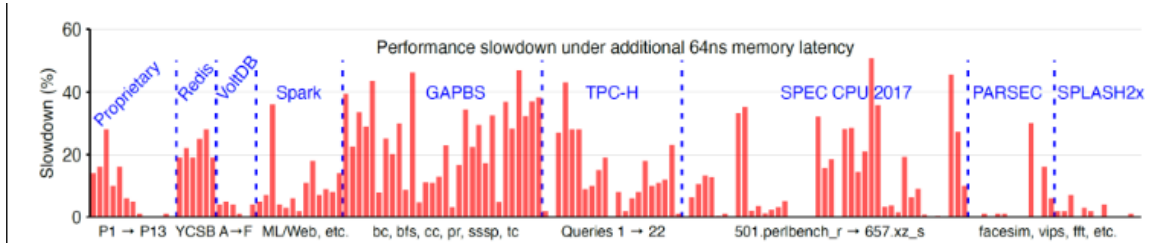
PCIe speeds equating single channel DRAM bandwidth

Workloads becoming more and more divergent

Hyperscaler's view on CXL

Azure

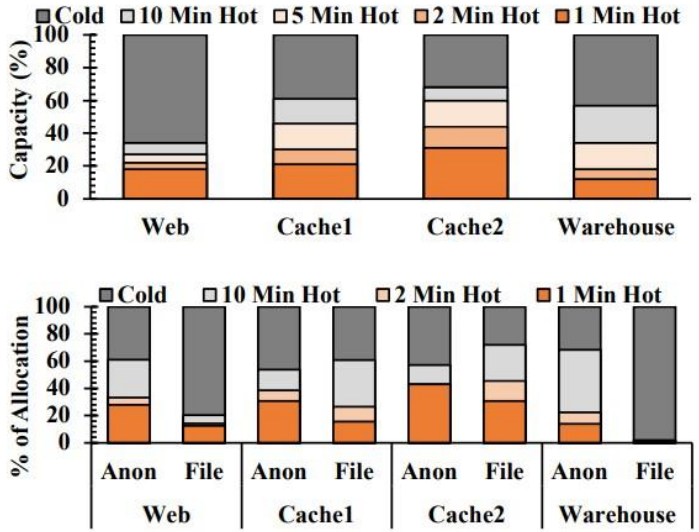




Source: https://arxiv.org/pdf/2203.00241.pdf

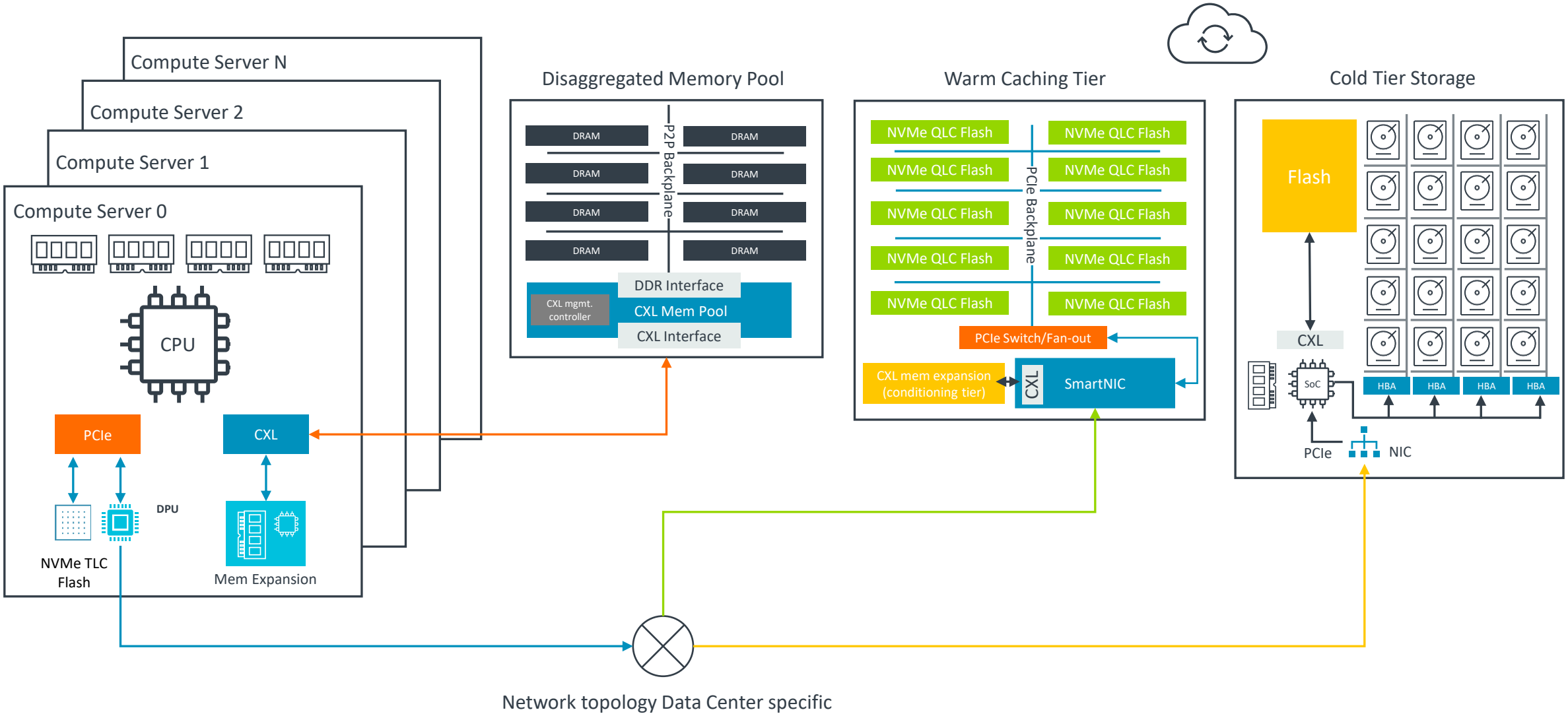
20% of workloads show NO performance impact due to additional CXL latency

Meta

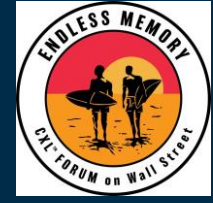


Source: https://www.nextplatform.com/2022/06/16/meta-platforms-hacks-cxl-memory-tier-into-linux/

Emerging Datacenter Architecture with CXL

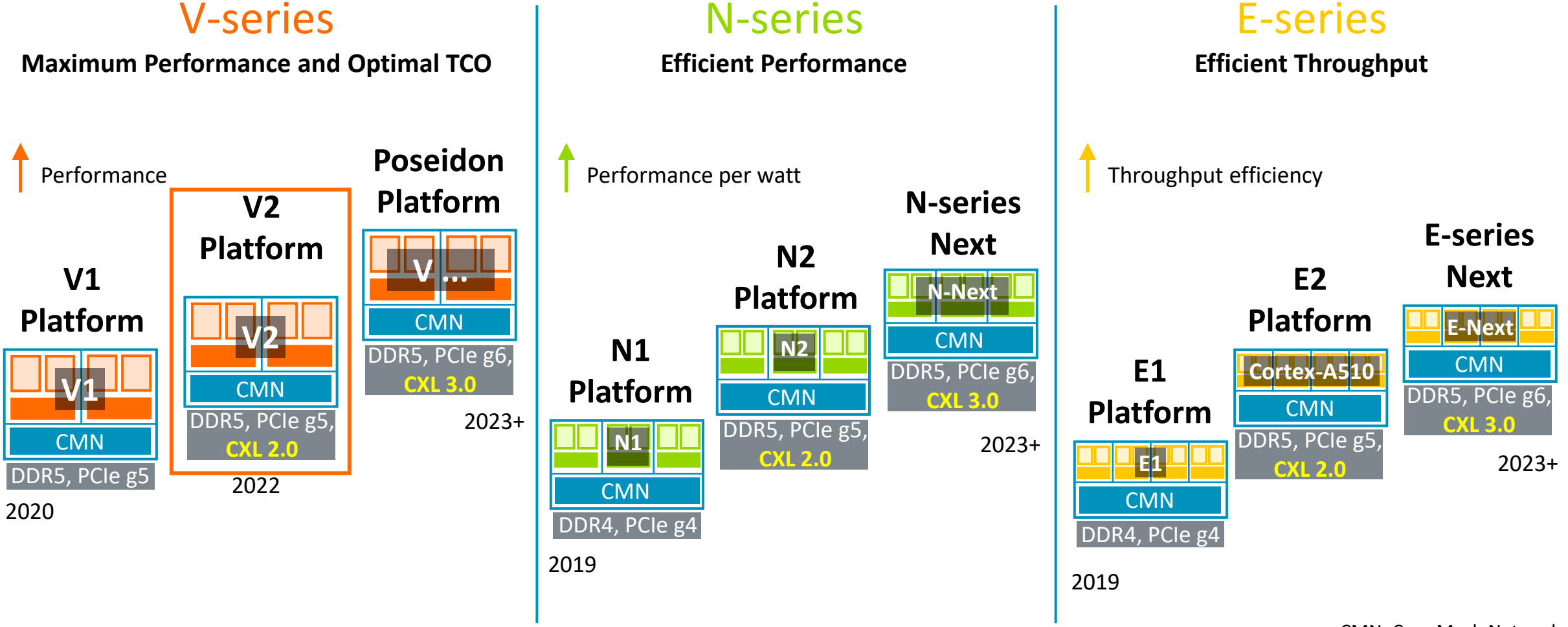


arm



CXL Product Enablement – Host Perspective

Arm is enabling Rapid Pace of Innovation with CXL

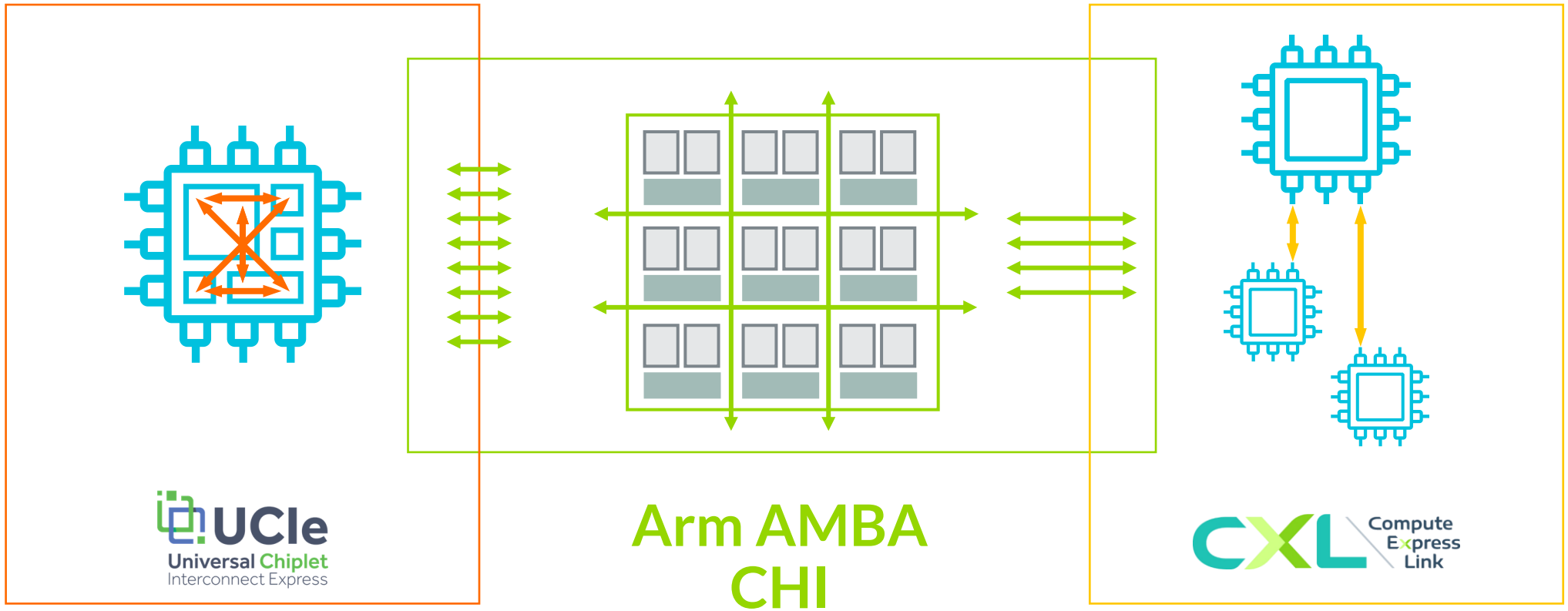


CMN: Core Mesh Network
CXL: Compute Express Link

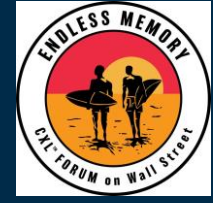
Arm Neoverse CPUs along with CMN enable CXL

Arm is Driving Technology Leadership Across Standards

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



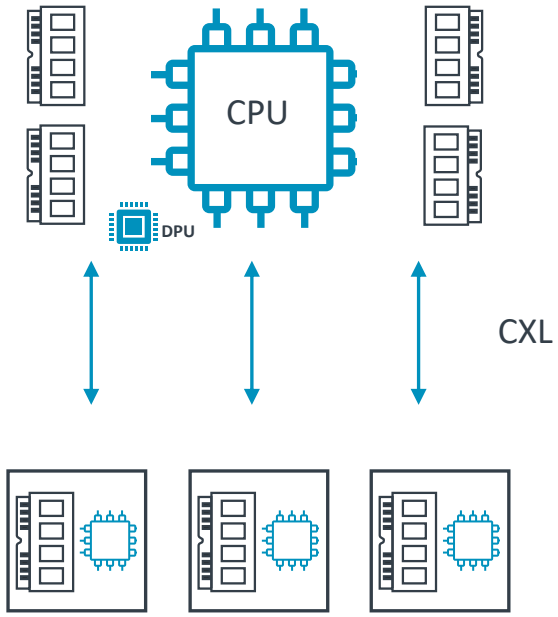
CMN: Coherent Mesh Network
CXL: Compute Express Link



CXL Product Enablement – Device Perspective

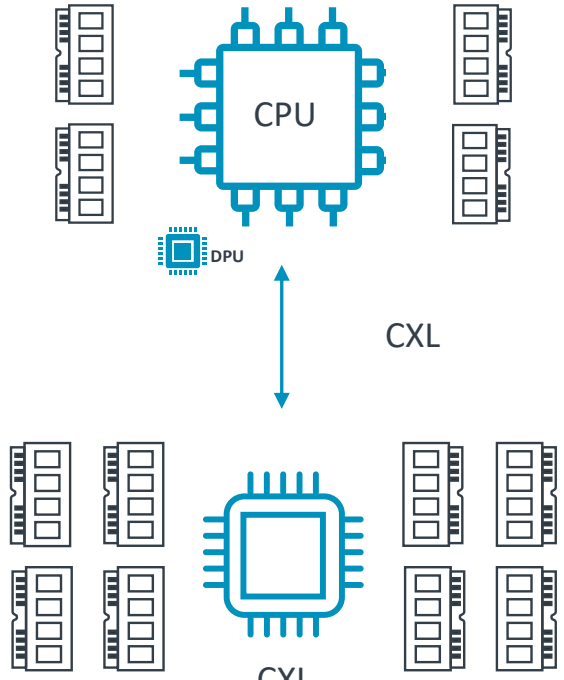
Solutions for Disaggregation of Compute and Memory

CXL Memory Expansion



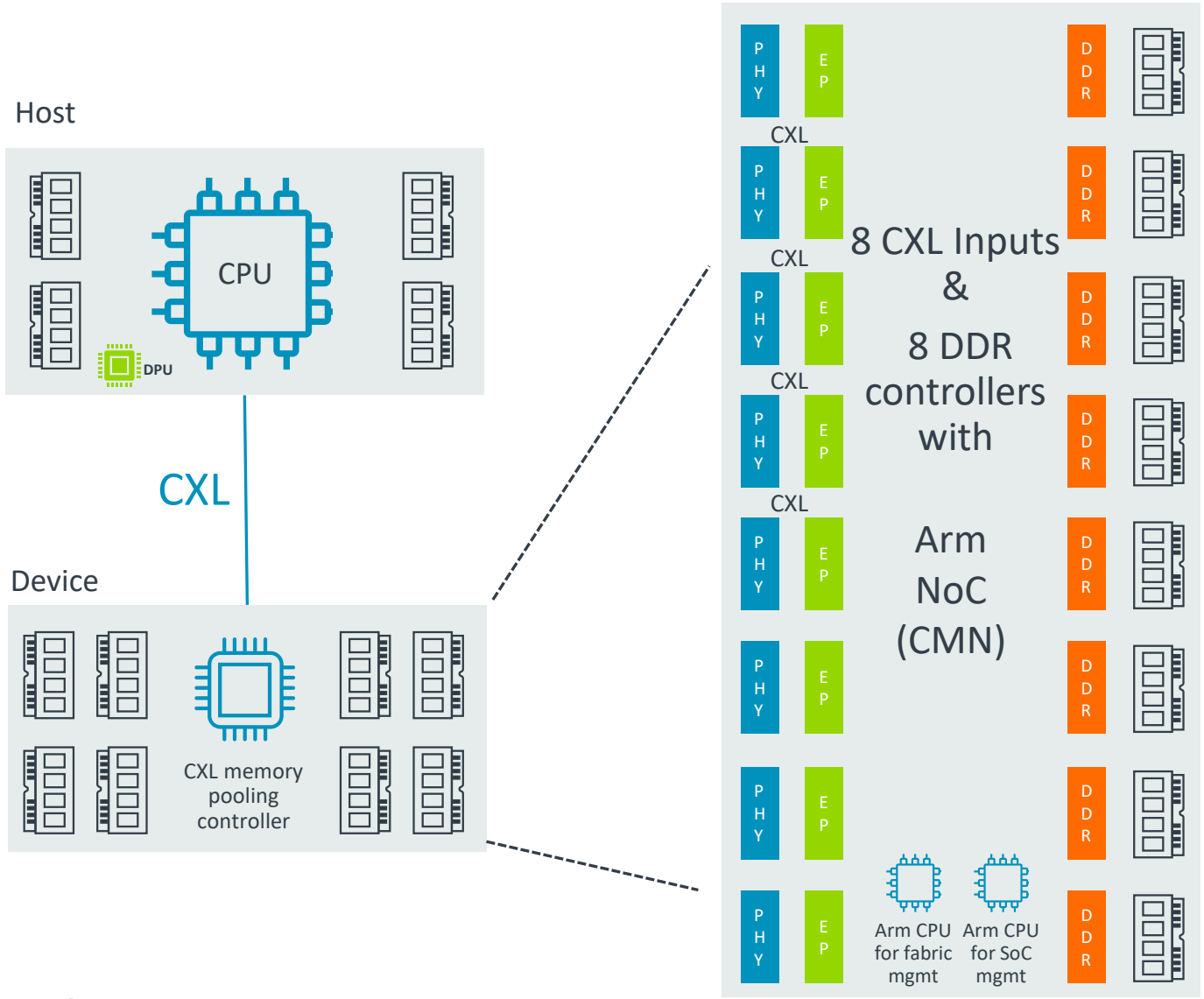
CXL memory expansion cards in the rack and separate racks

CXL Memory Pooling



CXL memory pooling controller

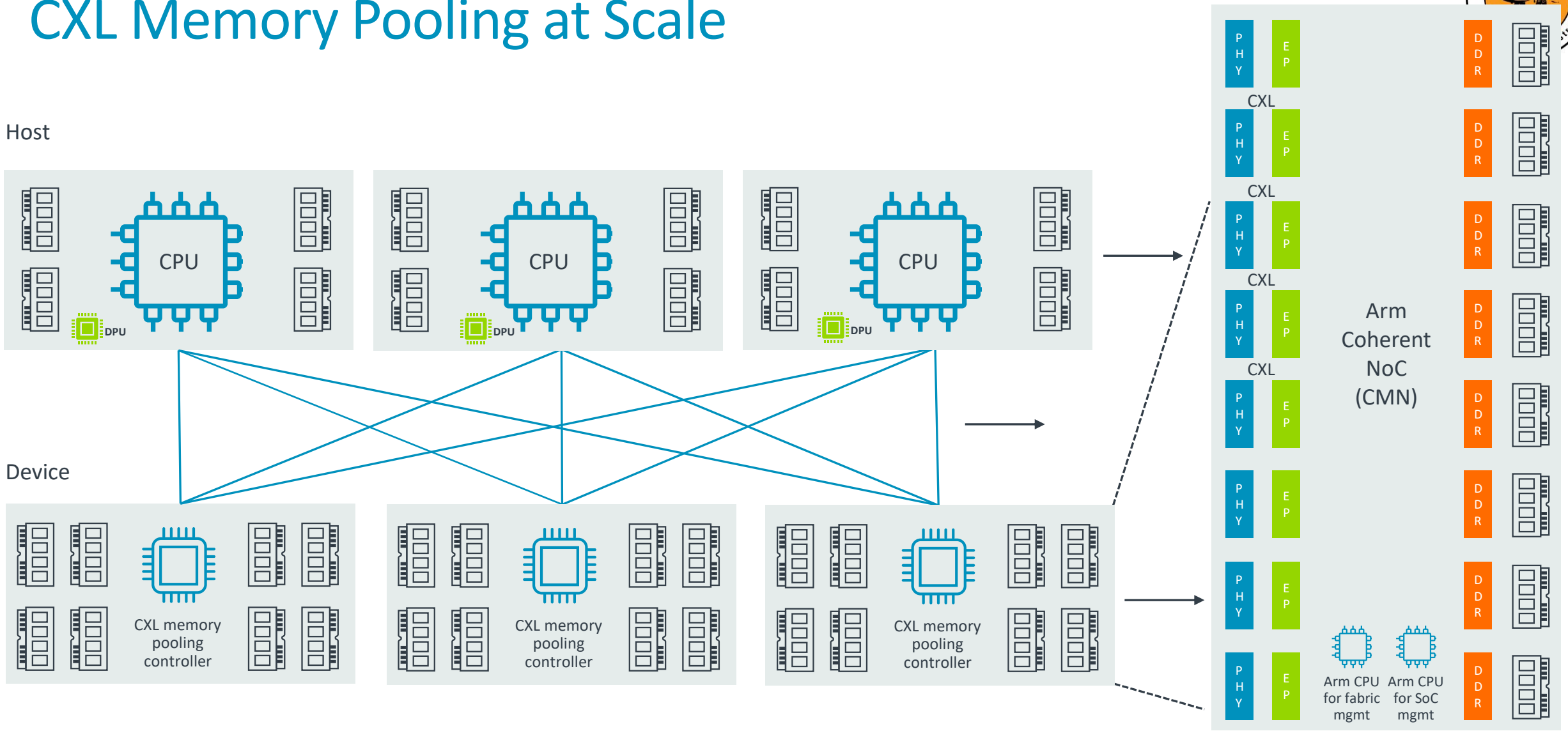
CXL Memory Pooling Solution



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

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

CXL Memory Pooling at Scale

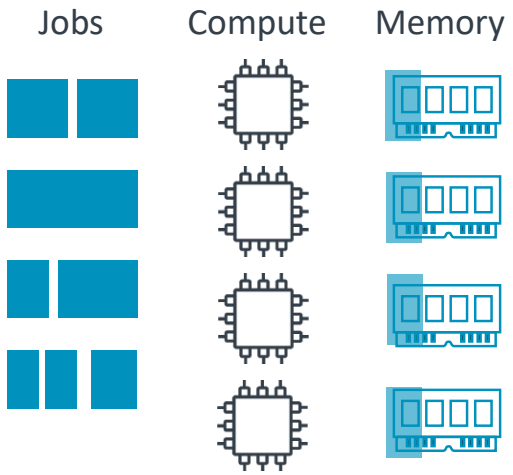


Arm is in a unique position to provide end-end CXL solution with Host and Device side knowledge

Economics Behind Memory Disaggregation

Disaggregated memory is worth doing, and benefit depends on optimization

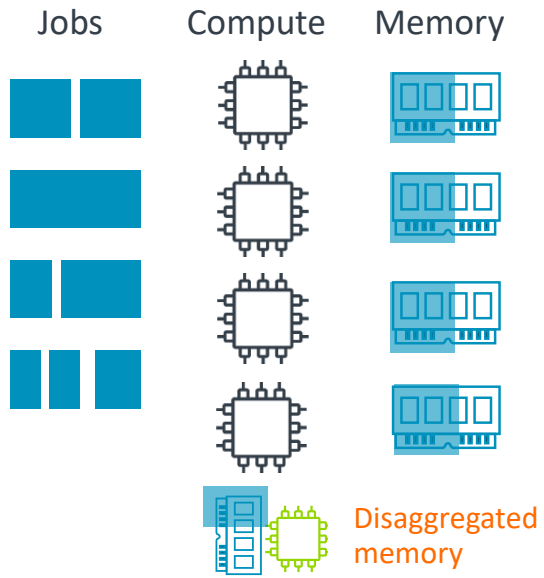
Baseline



Baseline performance and cost assumption:

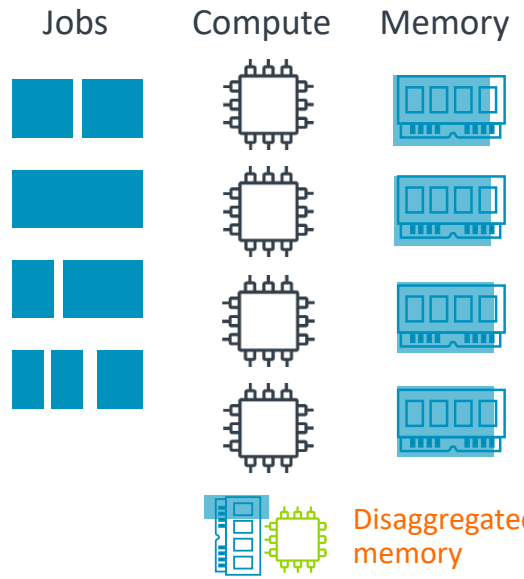
- 4GB/core
- 25-50% of DRAM is unused/stranded*

Disaggregated memory



- DRAM cost
- Average Latency
- 3%* better performance/TCO

Disaggregated memory with ideal tiering

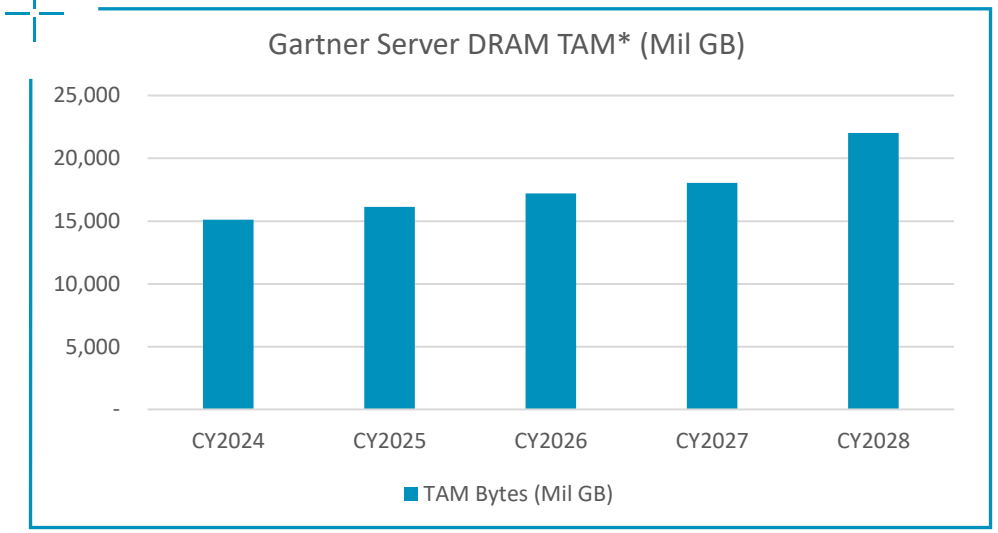


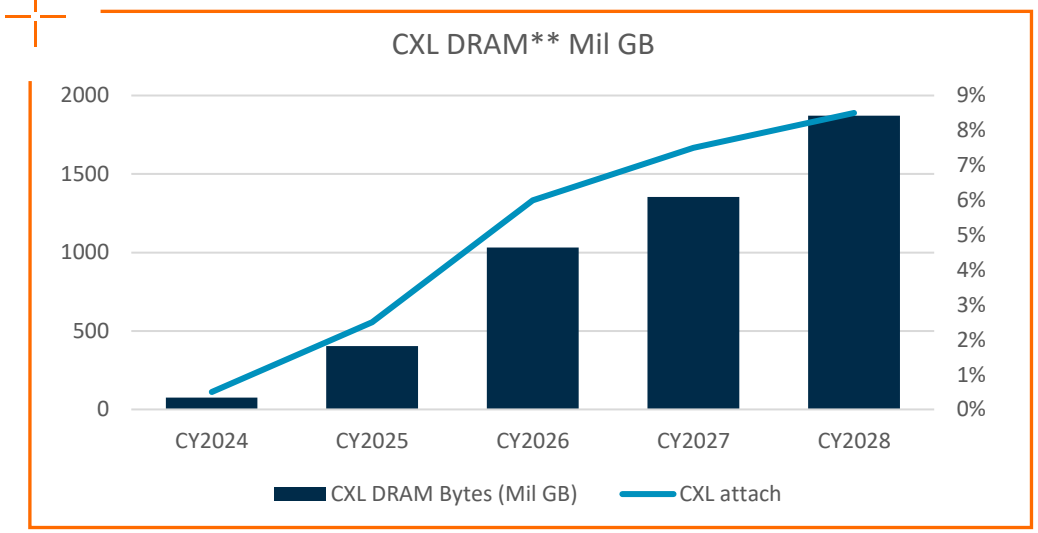
- DRAM cost
- Average Latency
- 26%* better performance/TCO

- Block size
- Migration rate
- Migration Algo
- Tiering cost

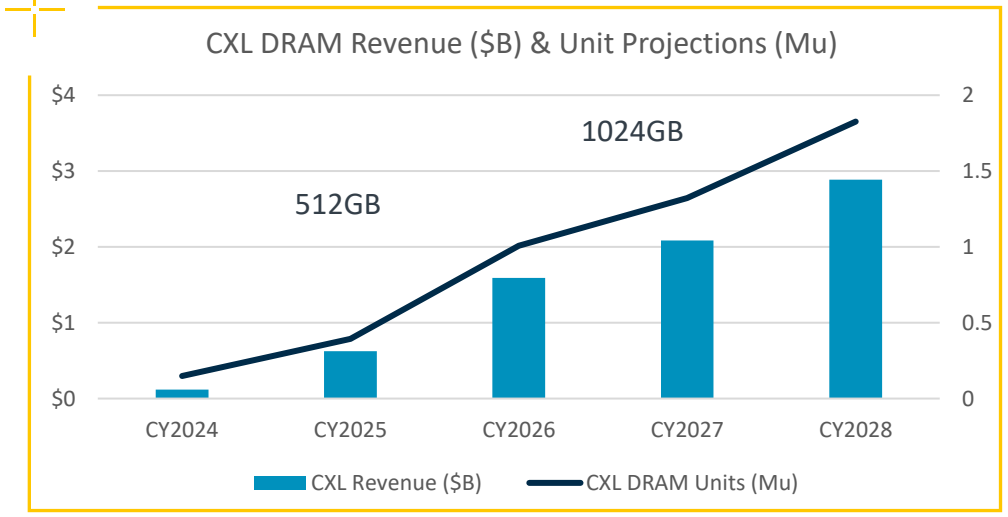
*Arm internal estimates based on public available information

CXL memory expansion and pooling TAM estimates

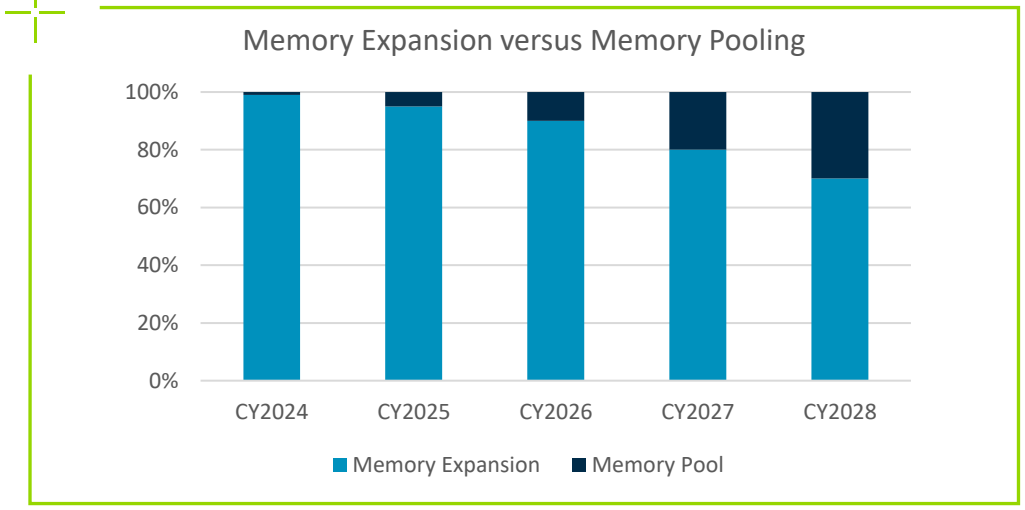




Server DRAM bits growing at 20% CAGR – CXL attach increasing



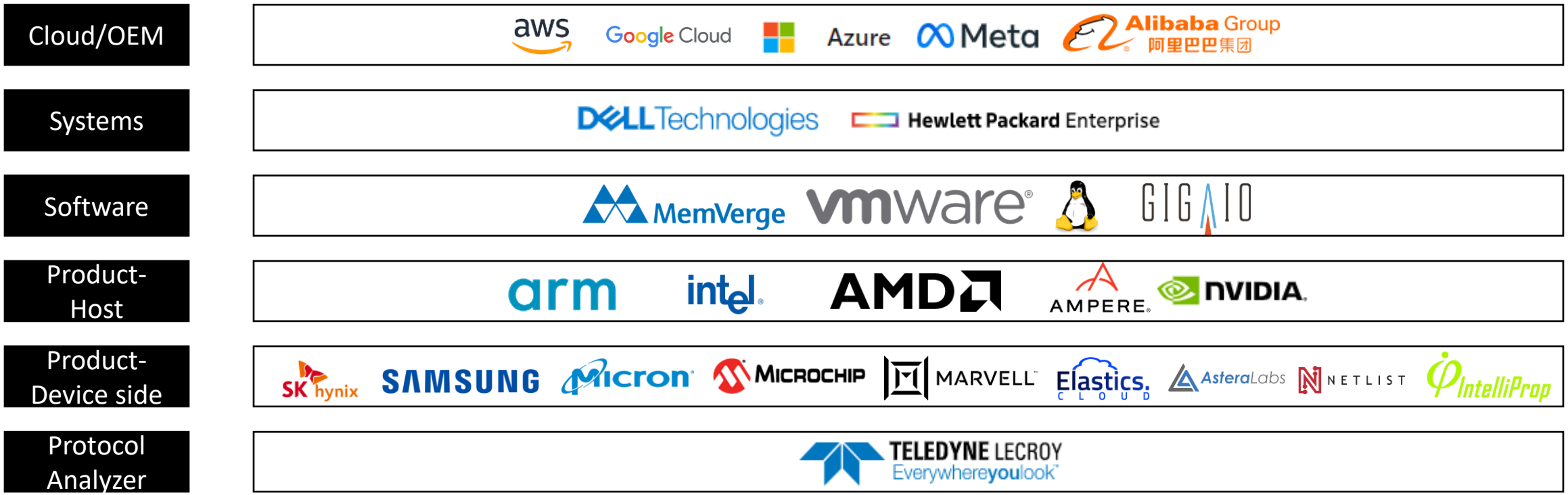
CXL DRAM TAM increasing initially with memory expansion and then pooling



* Gartner DRAM Market Statistics, Supply and Demand 2Q 2022

**Arm CXL attach estimates

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 hyperscalers and device partners
- Arm has demonstrated ecosystem leadership in many segments

Ecosystem Enablement

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

arm

Thank You
Danke
Gracias
Grazie
谢谢
ありがとう
Asante
Merci
감사합니다
धन्यवाद
Kiitos
شكرًا
ধন্যবাদ
תודה