SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

Explore the Compute Express Link™ (CXL™) Device Ecosystem and Usage Models

Moderator: Kurtis Bowman (CXL Consortium TTF Co-Chair)

Panelists: Sandeep Dattaprasad (Astera Labs), George Apostol (Elastics.cloud), Kapil Sethi (Samsung), and Jianping Jiang (Xconn Technologies)

SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

CXL Overview

CXL Delivers the Right Features & Architecture

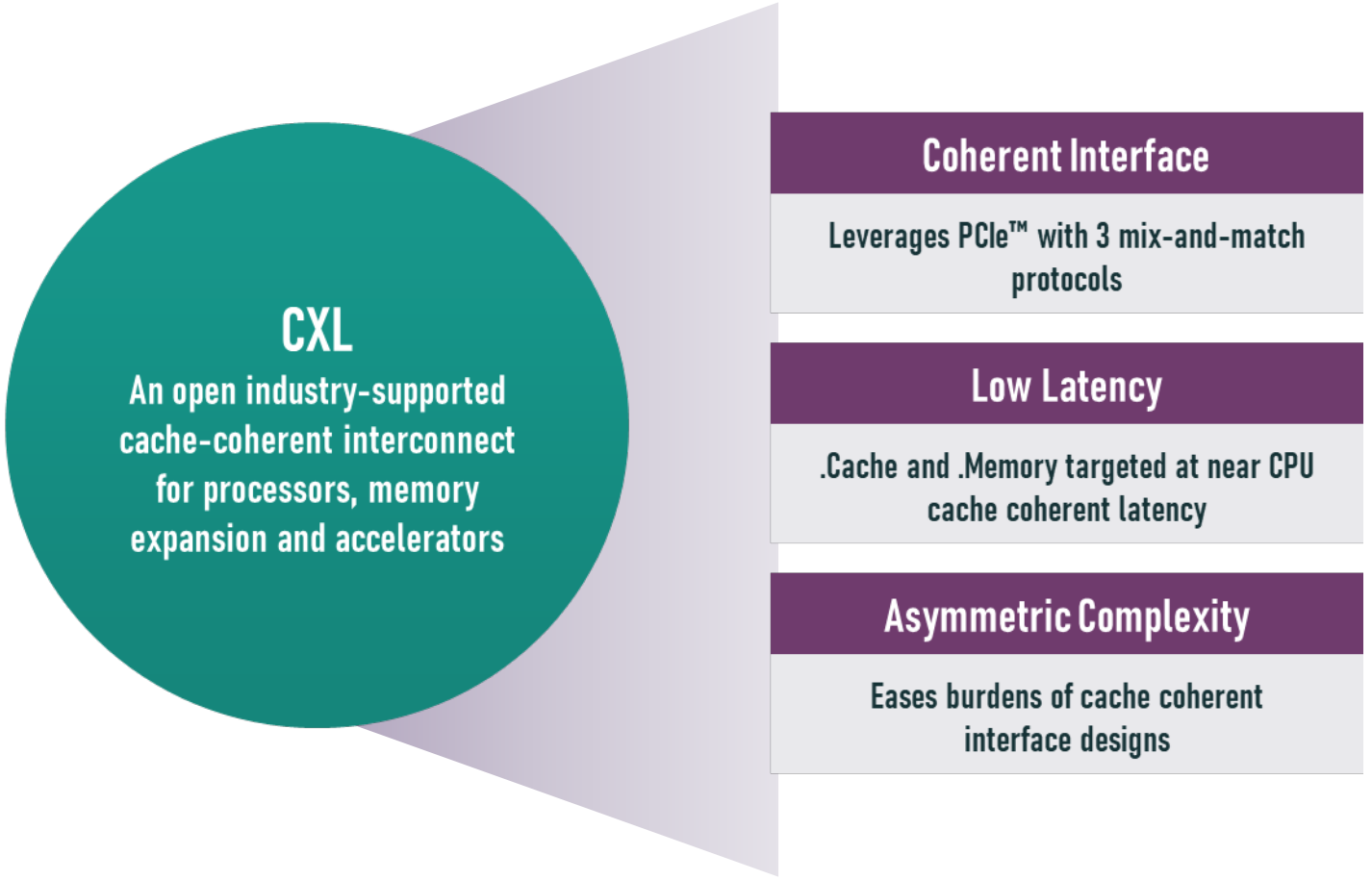
Challenges

Industry trends driving demand for faster data processing and next-gen data center performance

Increasing demand for heterogeneous computing and server disaggregation

Need for increased memory capacity and bandwidth

Lack of open industry standard to address next-gen interconnect challenges



Coherent Interface

Leverages PCIe™ with 3 mix-and-match protocols

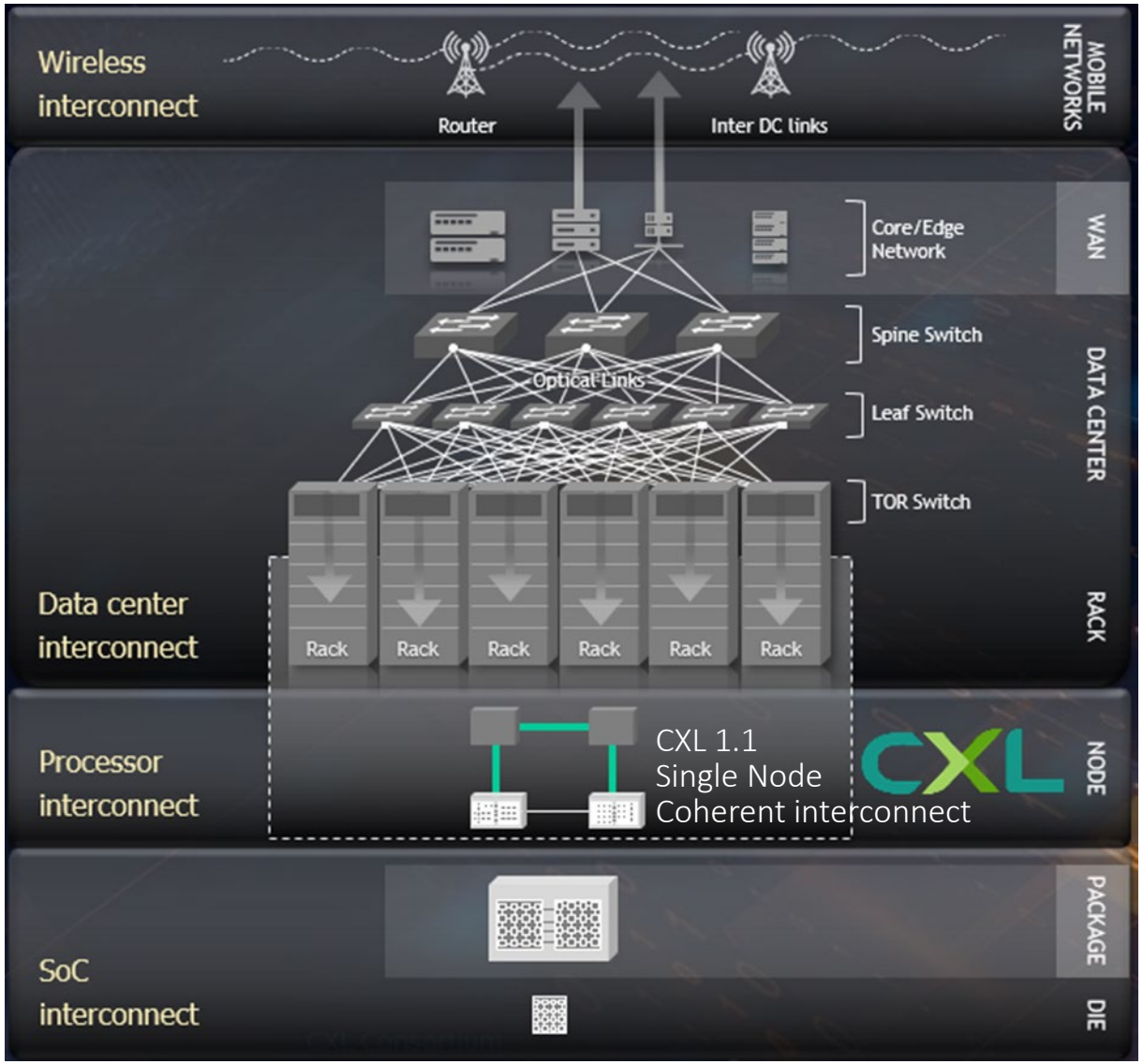
Low Latency

.Cache and .Memory targeted at near CPU cache coherent latency

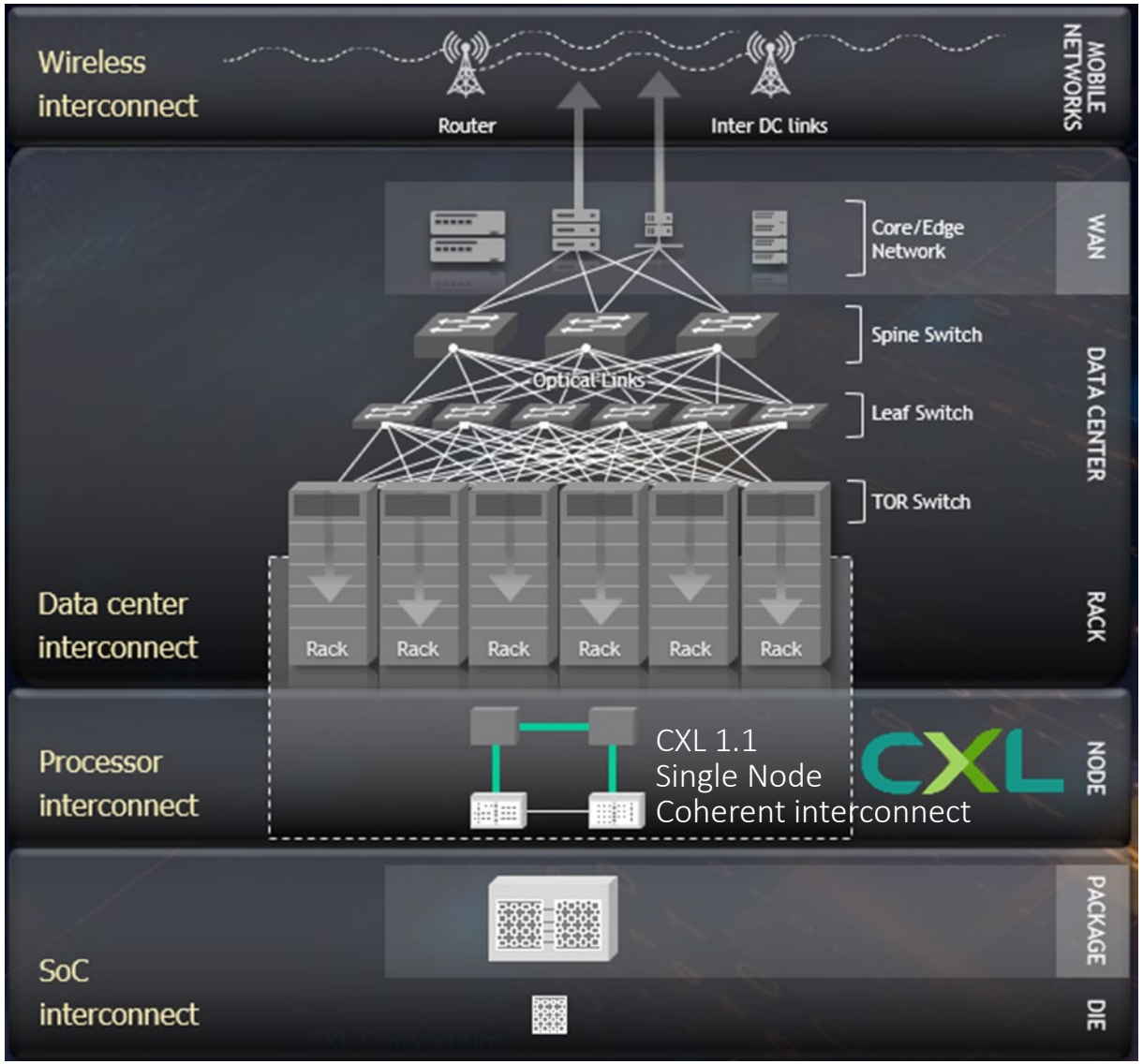
Asymmetric Complexity

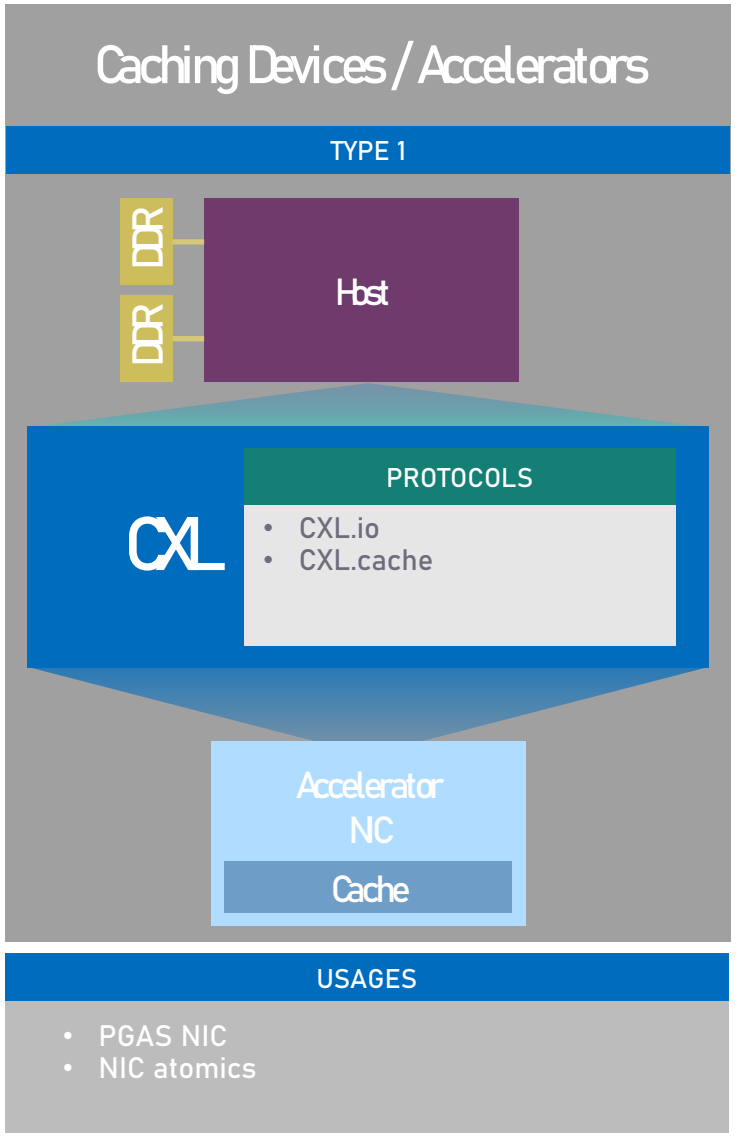
Eases burdens of cache coherent interface designs

Data Center: The Expanding Scope of CXL

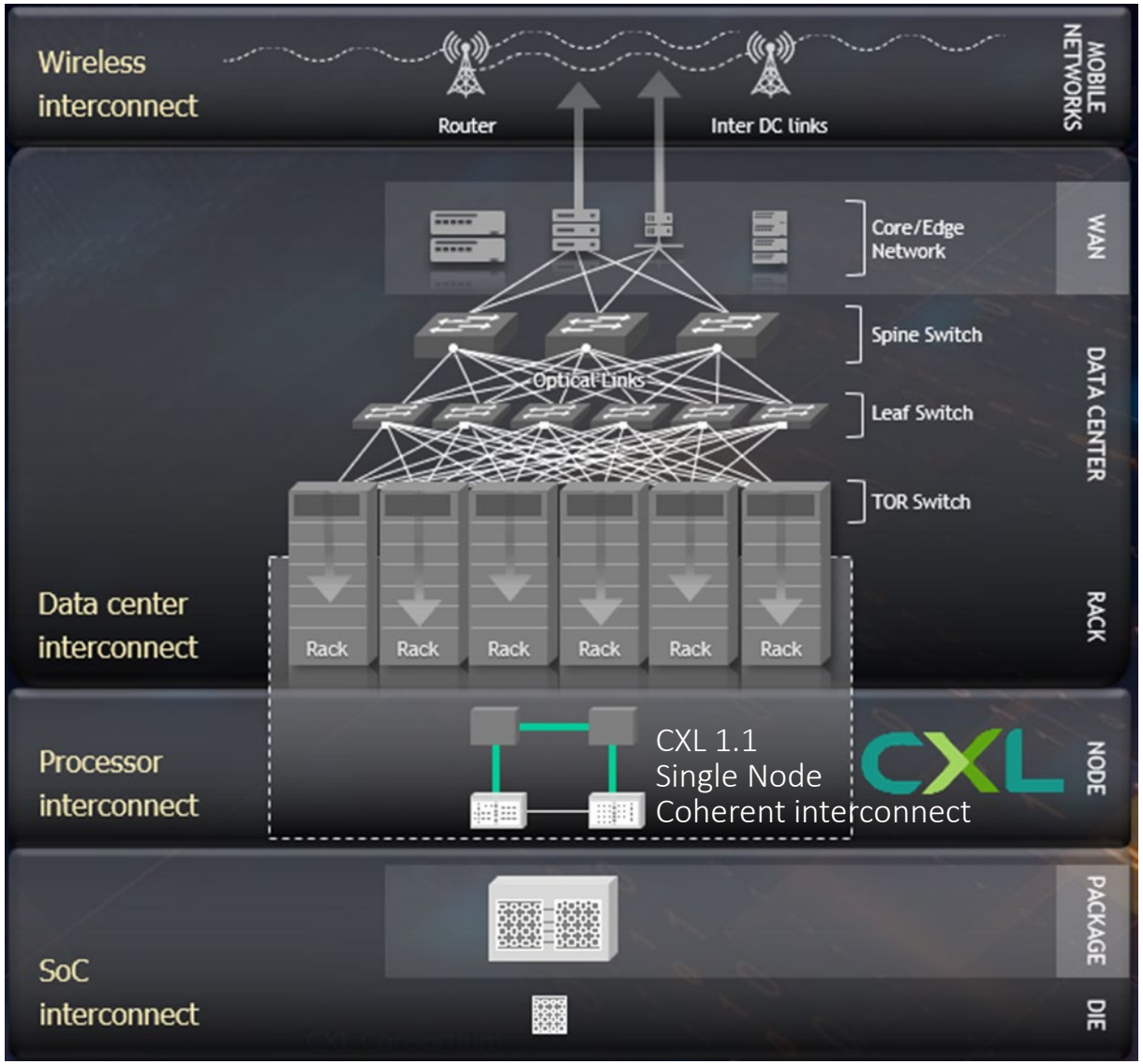


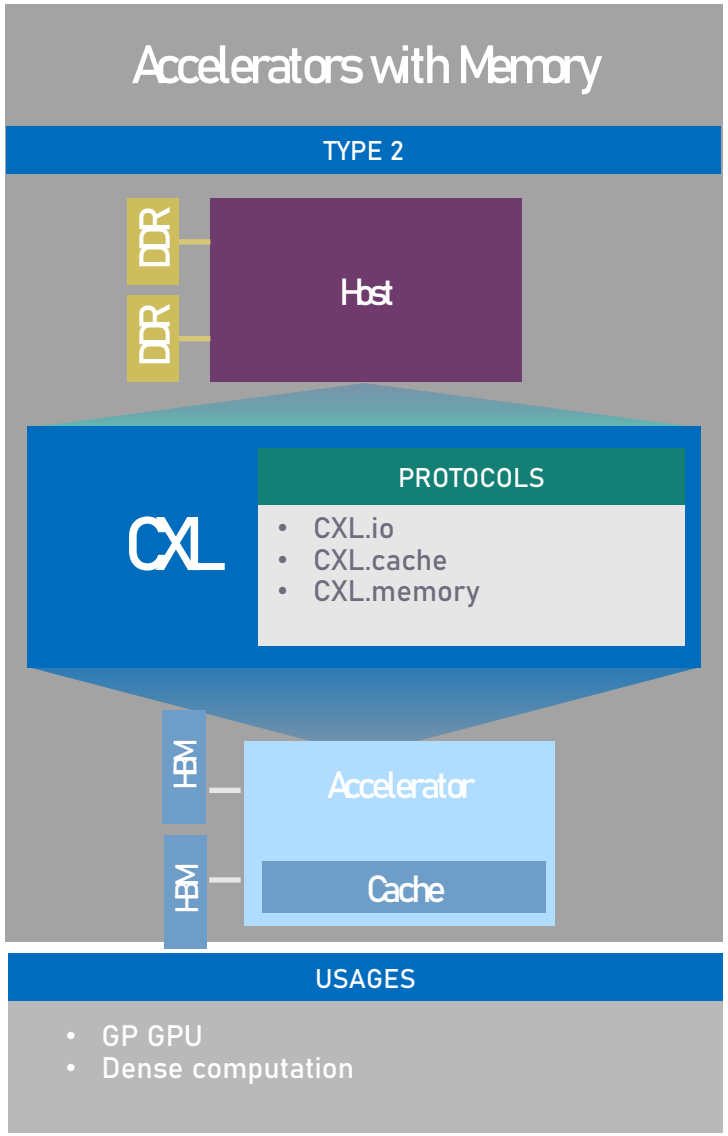
Data Center: The Expanding Scope of CXL



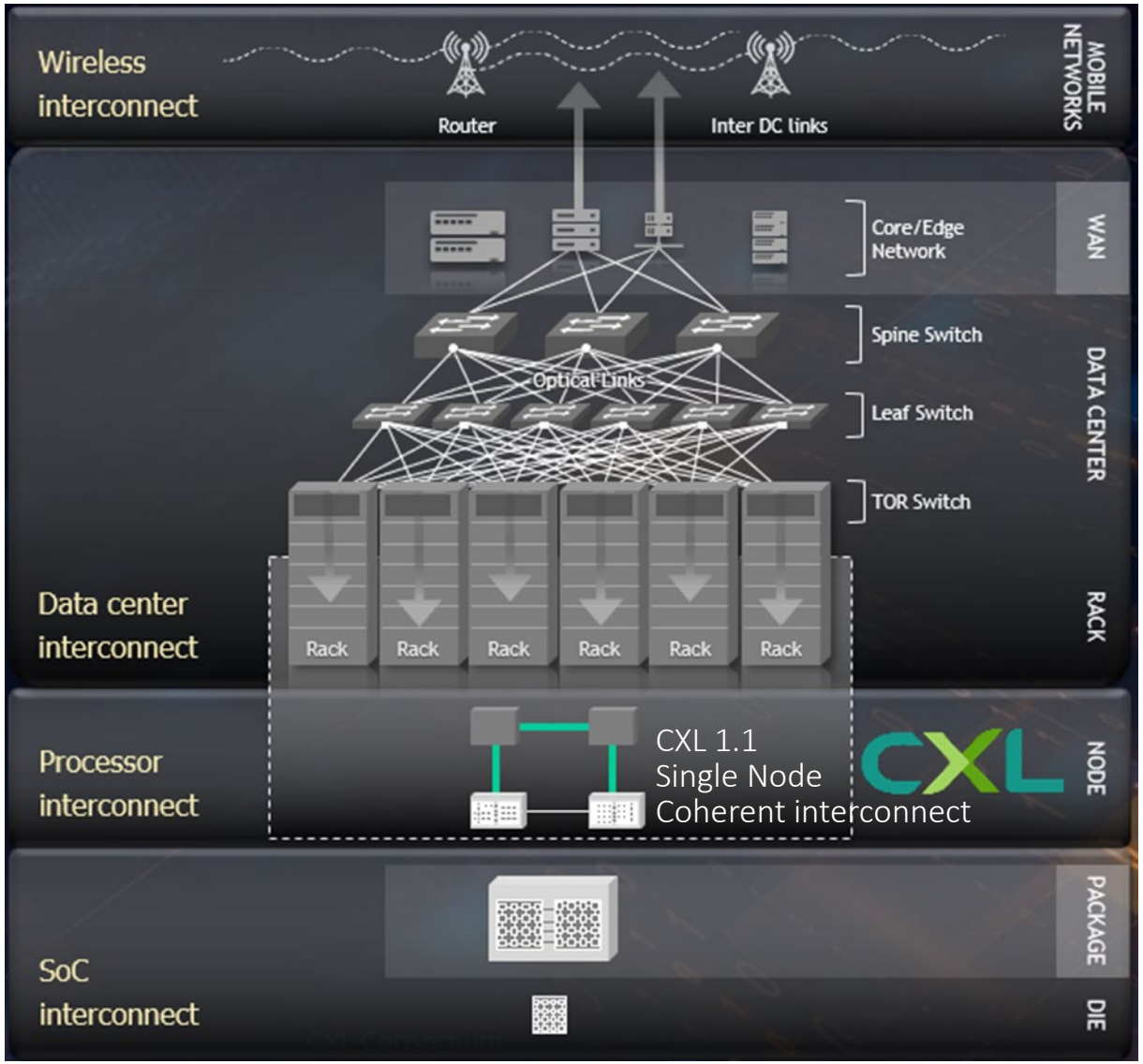


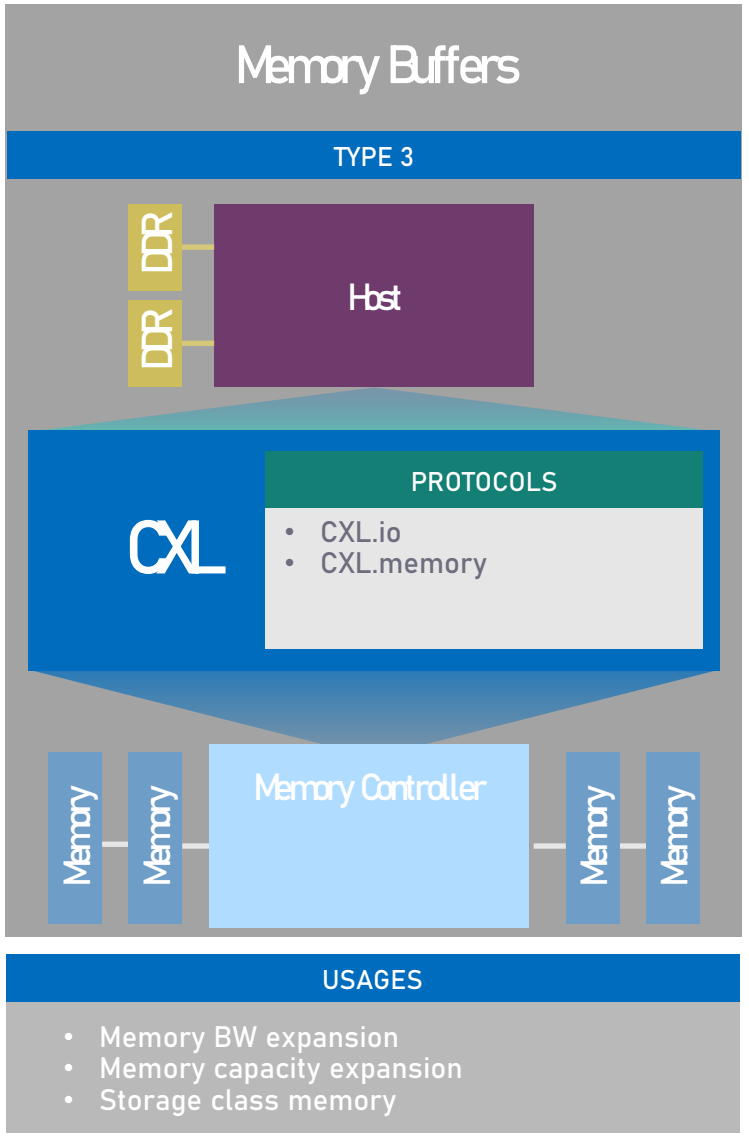
Data Center: The Expanding Scope of CXL



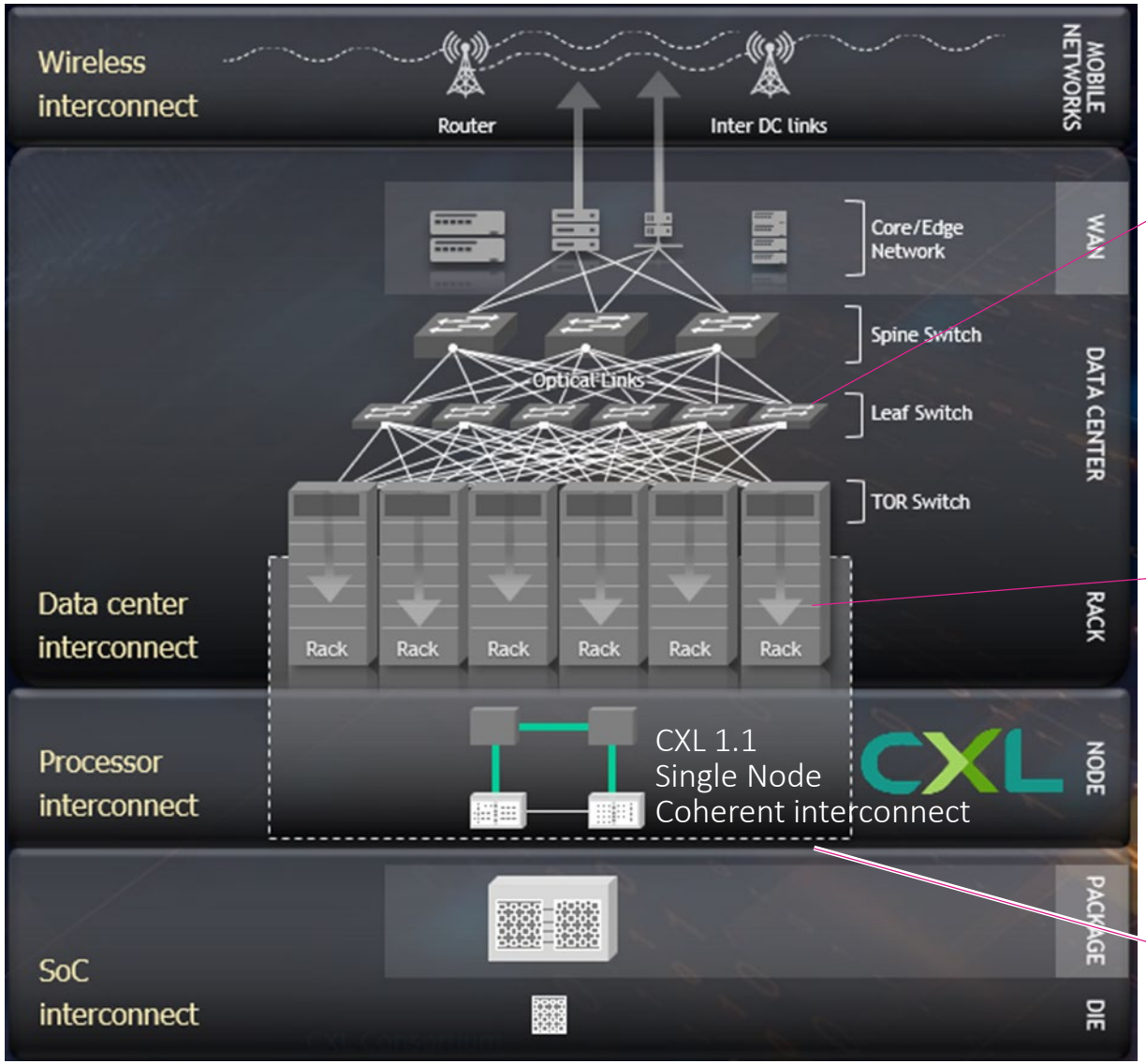


Data Center: The Expanding Scope of CXL

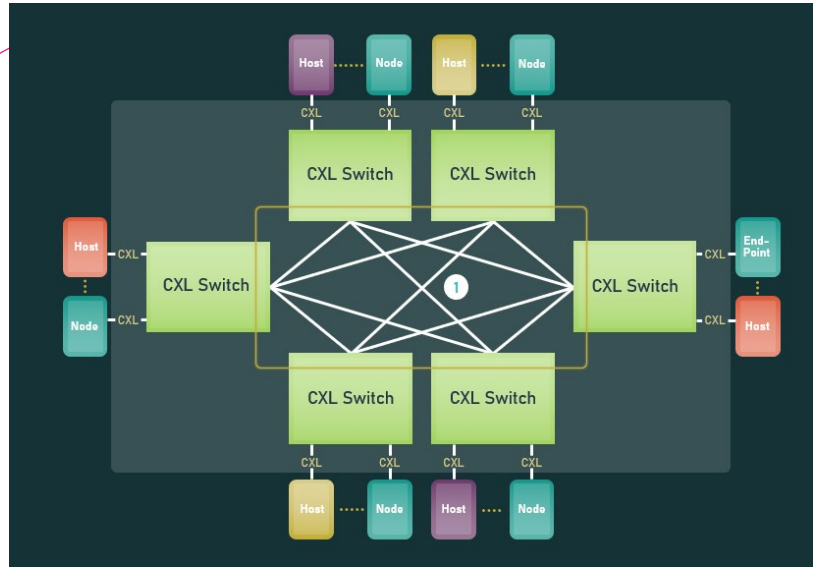




Data Center: The Expanding Scope of CXL



CXL 3.0

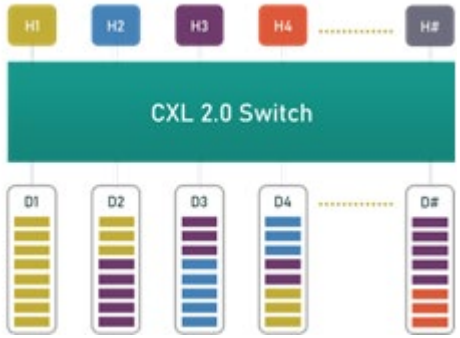


Composable Fabric growth for disaggregation/pooling/accelerator

CXL 2.0

Memory/Accelerator Pooling with Single Logical Devices

Memory Pooling with Multiple Logical Devices



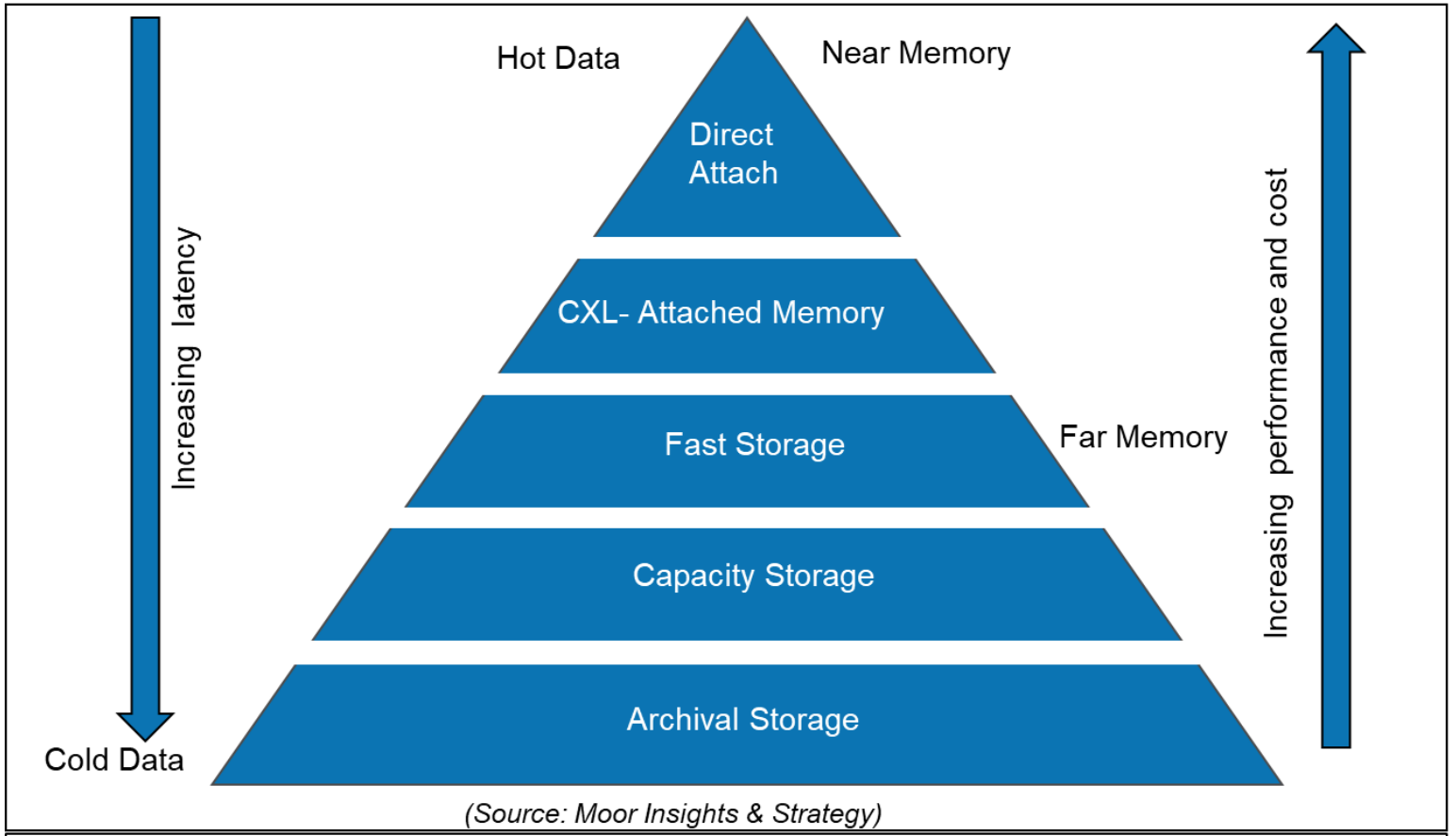
Multiple Nodes inside a Rack/Chassis supporting pooling of resources

SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

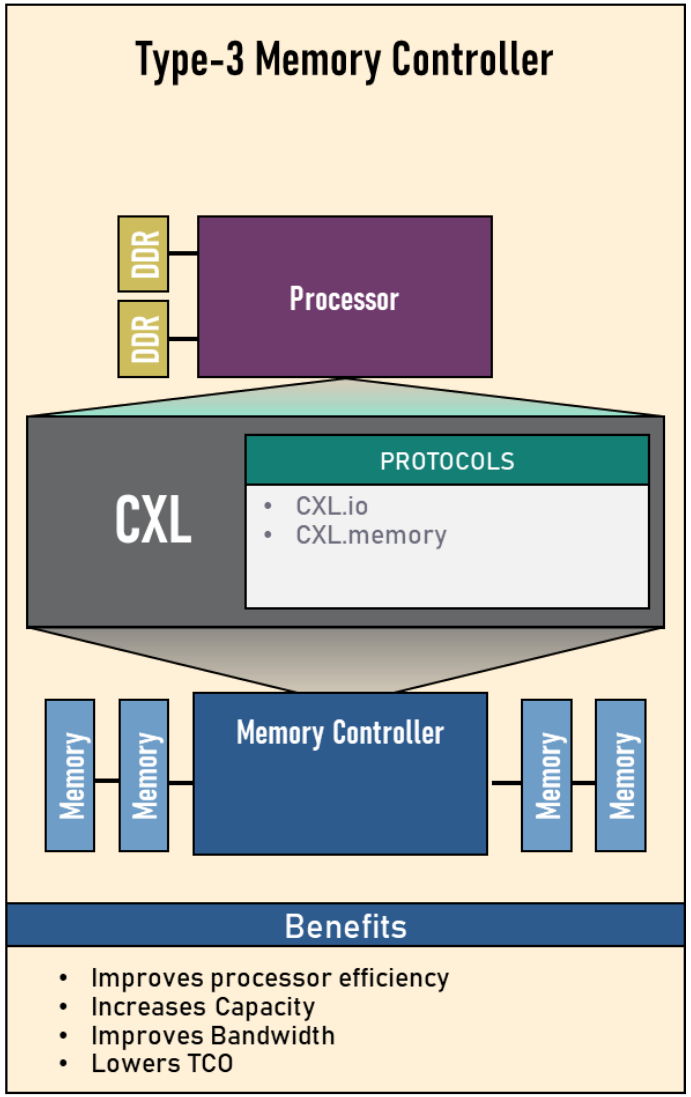
Introduction and Applications of CXL Type-3 Memory Controllers

Heterogenous Computing with CXL-Attached Memory



(Source: Moor Insights & Strategy)

- CXL Type-3 devices provide low latency byte addressable interconnect to memory devices
- Adds a layer between faster near memory and slower far memory providing optimal performance and cost benefits
- Provides composable architecture meeting AI/ML workload requirements



Technology Example: Leo Memory Connectivity Platform
Unleashing AI/ML Performance and Reducing TCO with DDR5 for Cloud Computing

Features

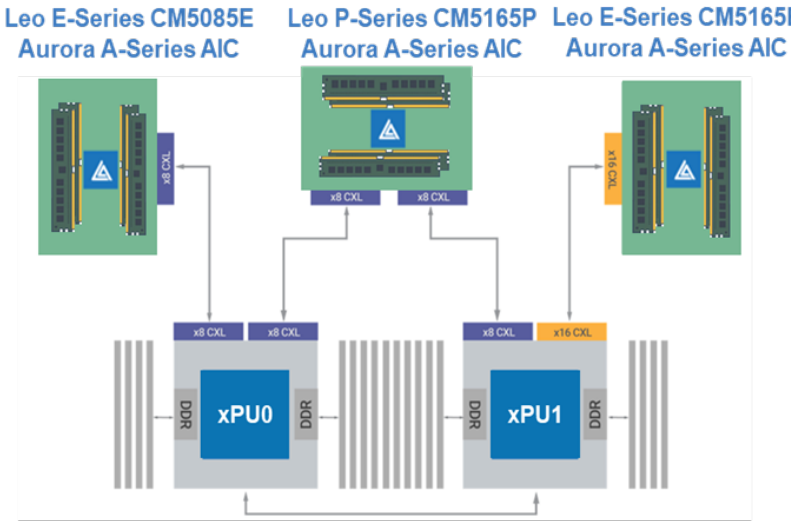
Overview

- Type-3 CXL 1.1/2.0 device purpose-built for cloud scale deployment
- CXL interface up to 32 GT/s per lane, up to 16 lanes
- Multiple DDR5 channels to increase memory capacity up to 2TB
- Server-grade customizable RAS and software APIs to integrate with fleet management services
- Best-in-class security features to ensure end-to-end data integrity and protection

Applications

- AI/ML Platform CPU-GPU Memory Expansion and Pooling use-cases

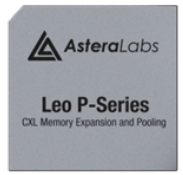
System Block Diagram



Product Offering

Leo Smart Memory Controllers

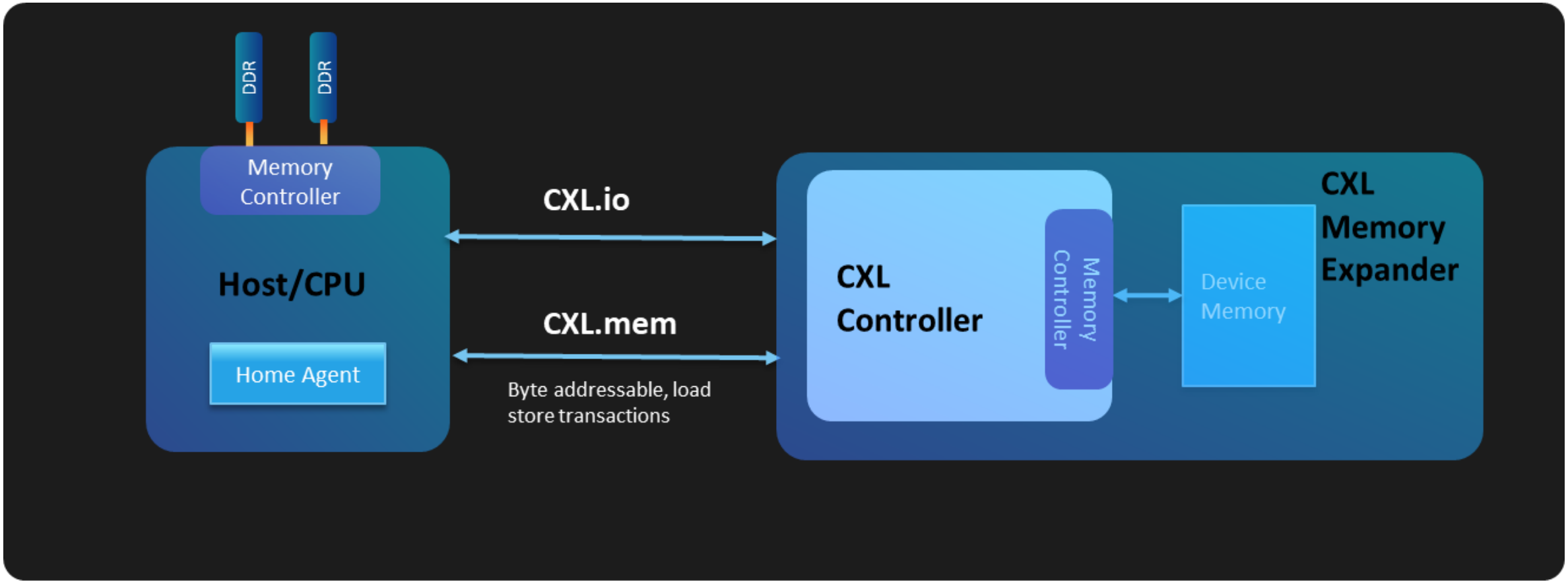
Leo E-Series
Memory Expansion



Leo P-Series
Memory Expansion & Pooling

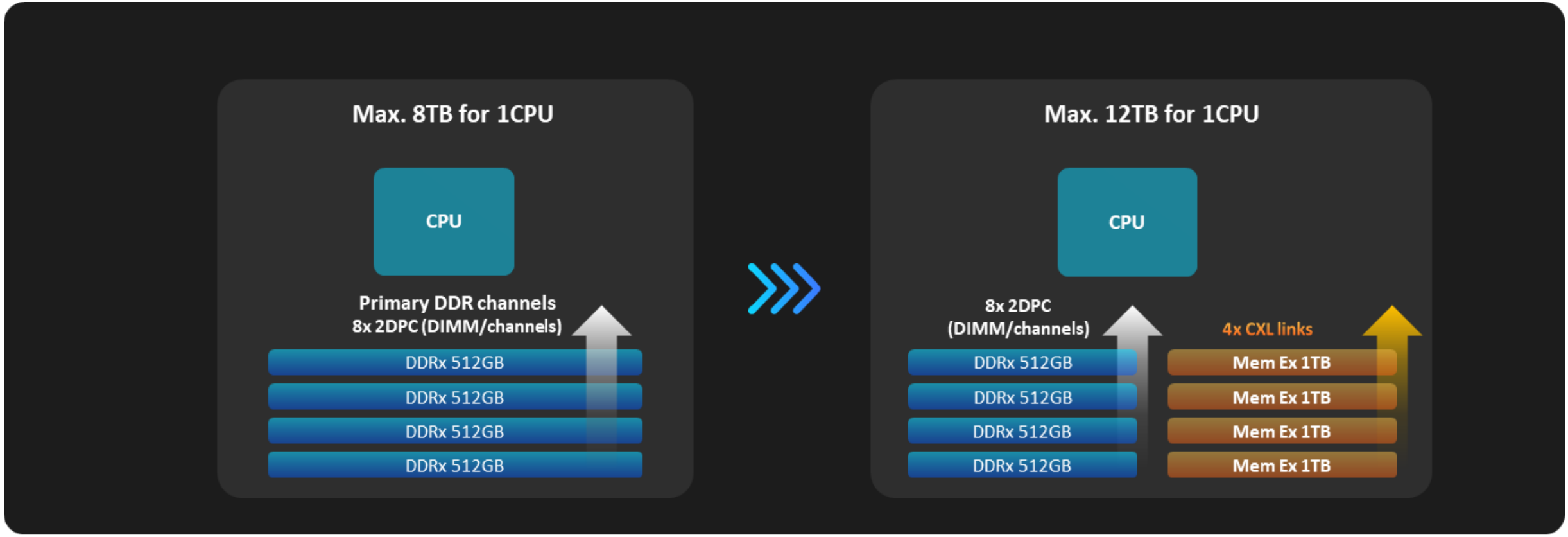
CXL Type 3 Device

CXL Type 3 Device



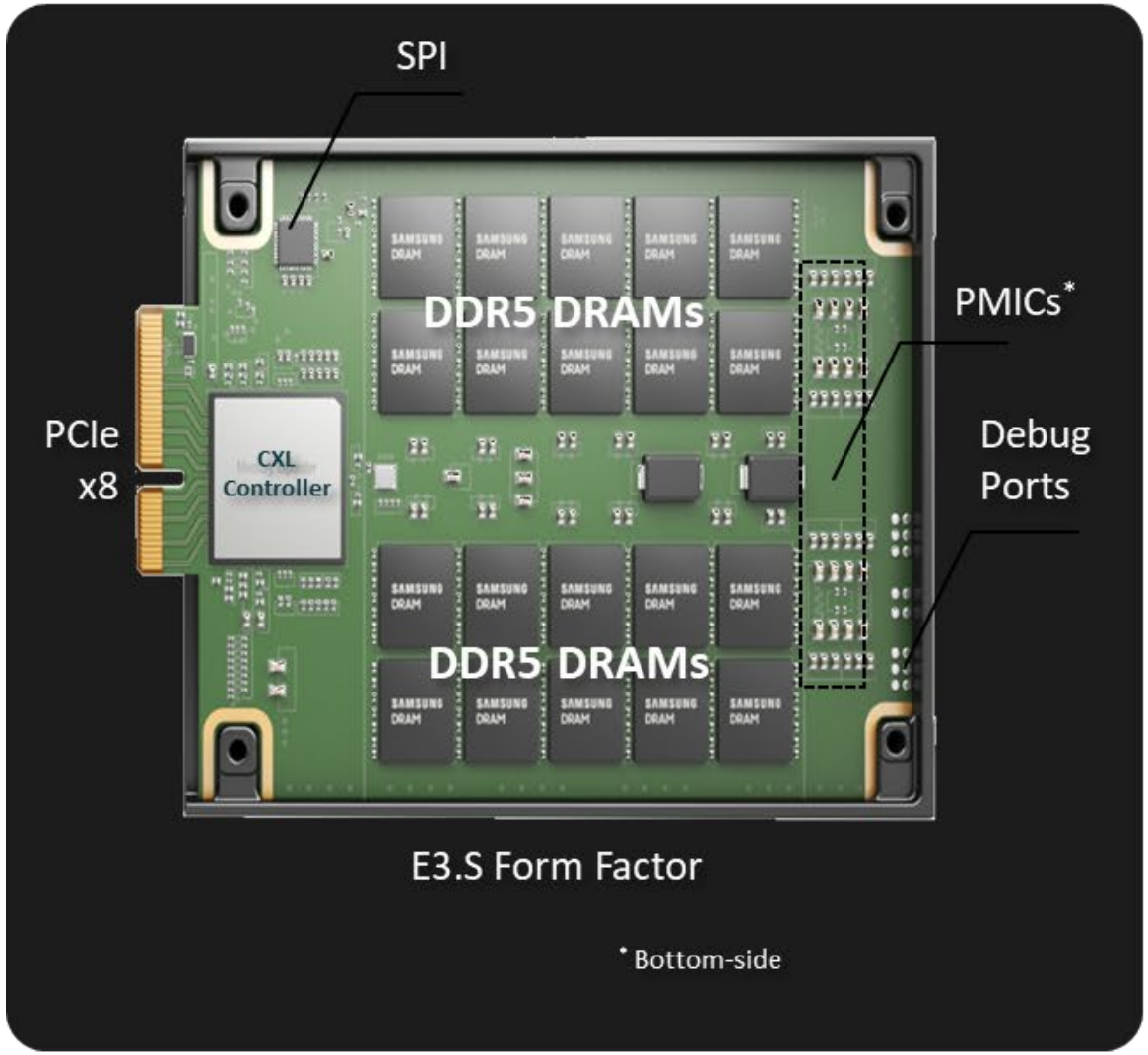
Memory attached to a CXL device is mapped coherently to the system address space

CXL Type 3 Device – Memory Expansion



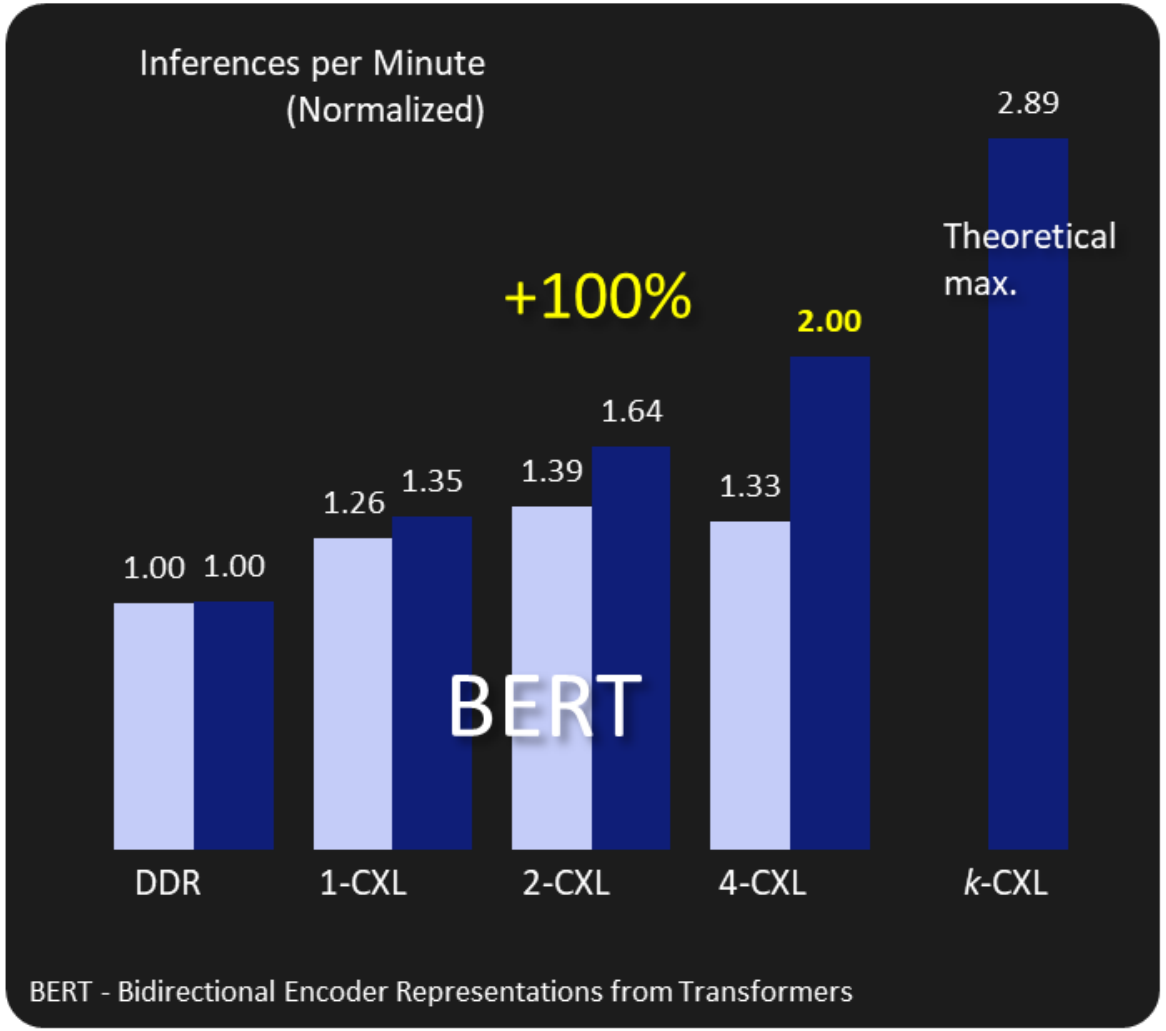
CXL enables systems to significantly scale memory capacity and bandwidth

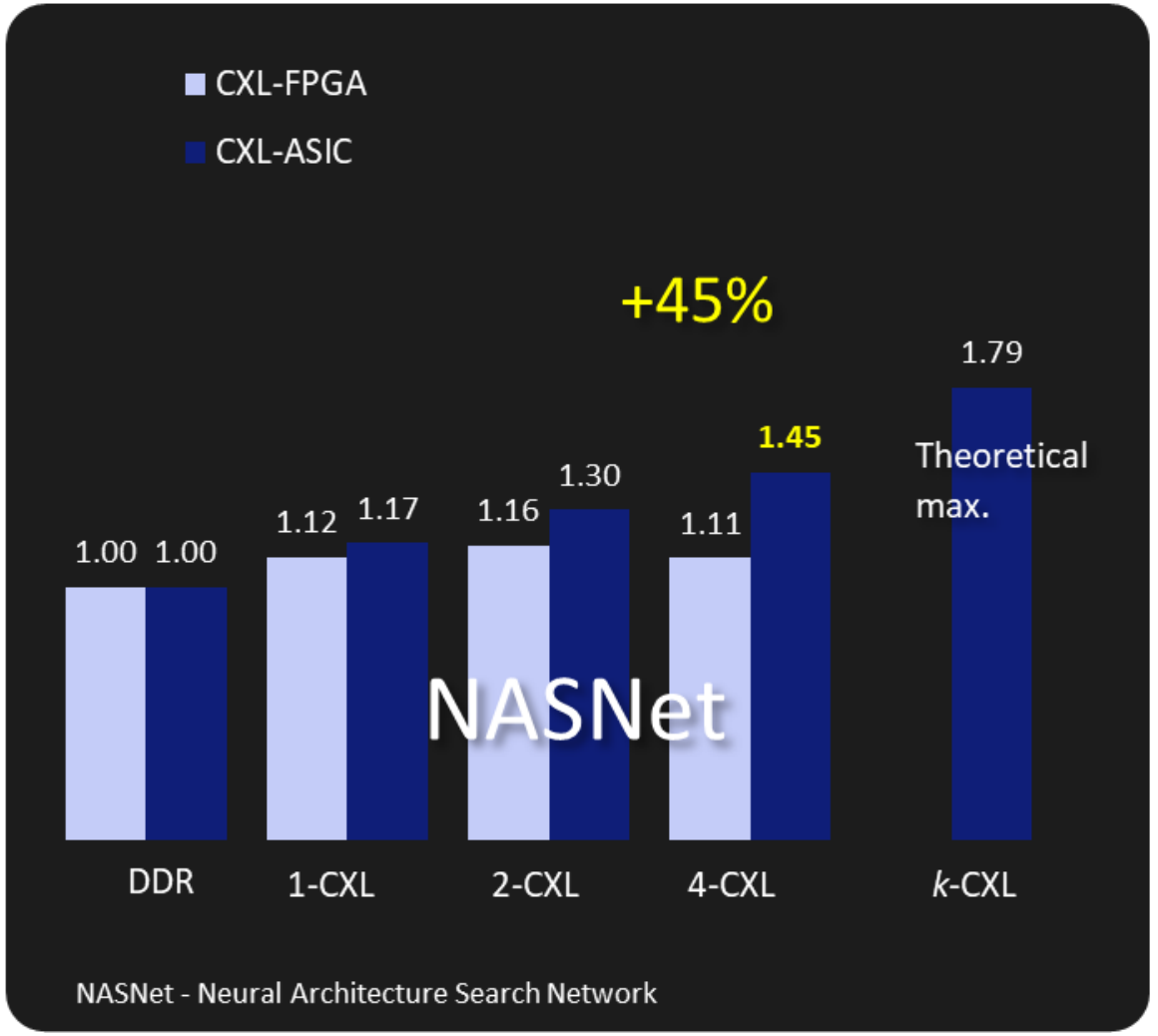
Samsung Memory Expander



- Form Factor – EDSFF (E3.S-2T)
- CXL 2.0 Support
- CXL Link Width - x8
- Media - DDR5
- Module Capacity – 128GB, 512GB
- Maximum CXL Bandwidth – 32GB/s
- Viral and Data Poisoning
- Memory Error Injection
- Multi-symbol ECC, Media scrubbing
- Availability – Currently available for evaluation/testing

System Test Results





SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

CXL 2.0 Switch from Xconn Technologies

Xconn Technologies' CXL 2.0 Switch

World's First CXL 2.0 and PCIe 5.0 switch

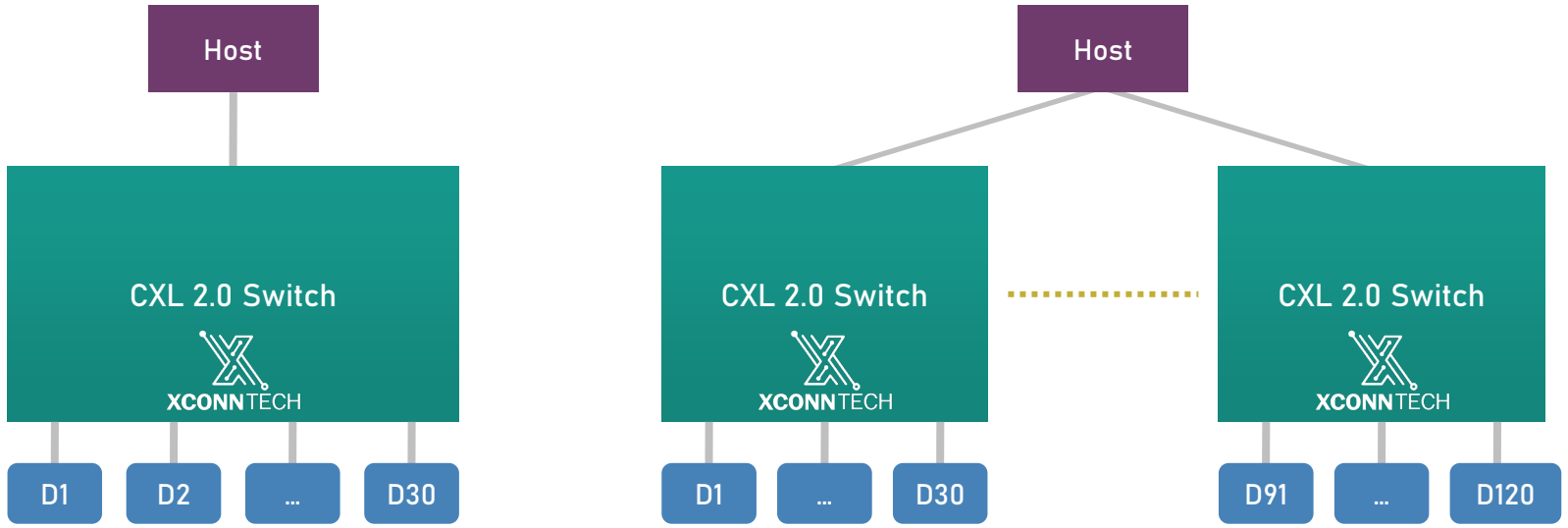
2,048 GB/s total BW with 256 lanes

Lowest port-to-port latency

Lowest power consumption/port

- XC50256 works with CXL 1.1 processors, CXL memory devices
- Future compatible with upcoming CXL 2.0 processors
- Also a PCIe 5.0 switch, works in hybrid mode (CXL/PCIe)
- Engineering samples available now

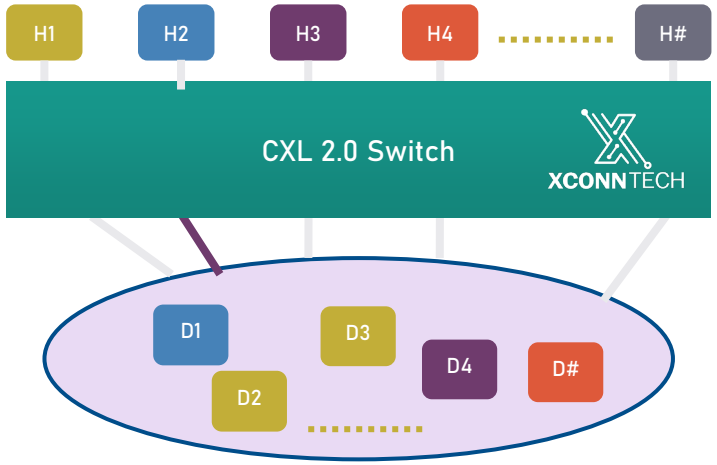
Scalable memory expansion enabled by CXL switch



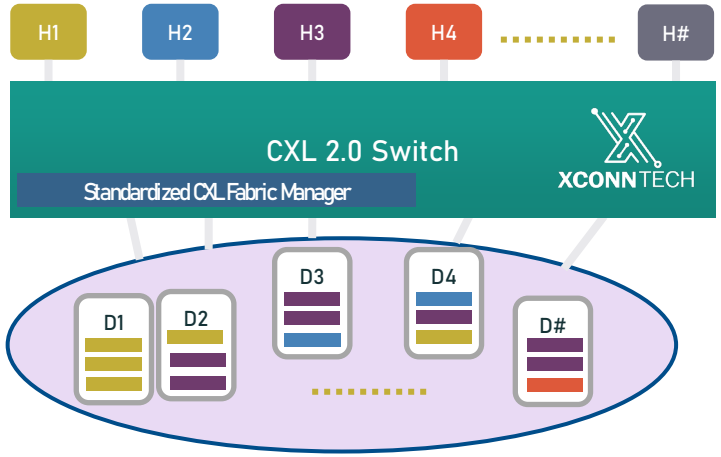
- Assume each CXL memory device is 0.5 TB
- With a single XC50256 (30 DSPs), up to 15TB memory expansion
- With 4x XC50256, up to 60TB memory expansion

Scalable memory pooling enabled by CXL switch

Memory/Accelerator Pooling with Single Logical Devices

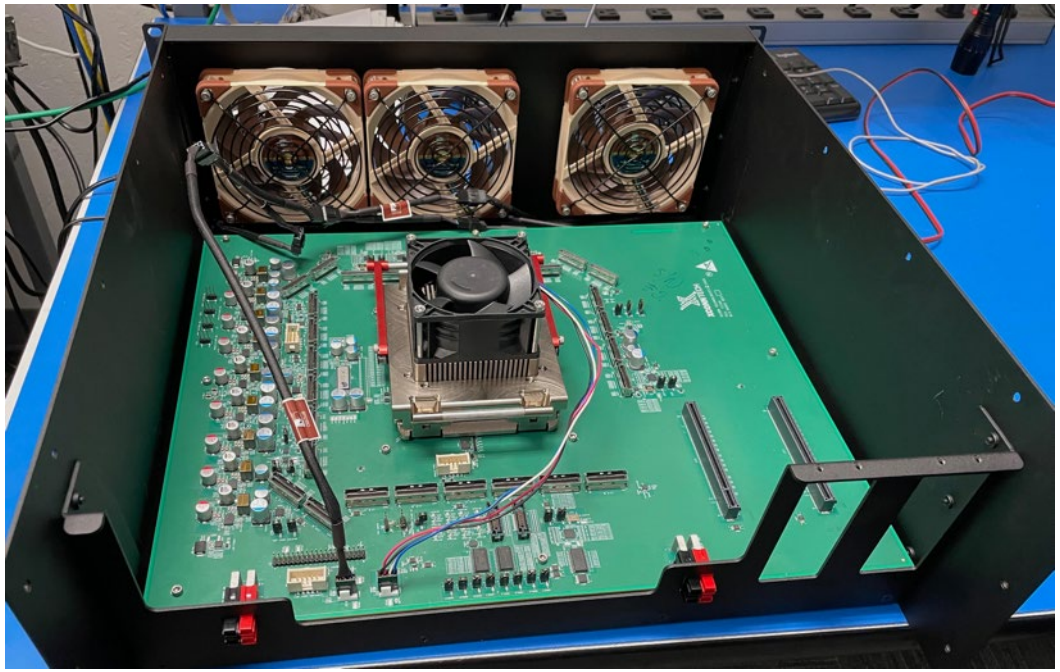


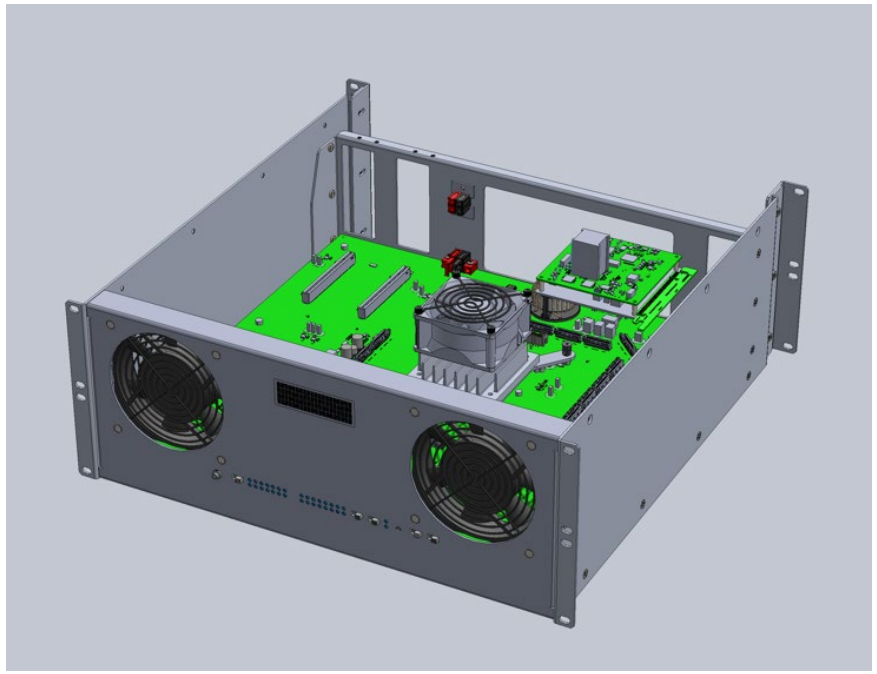
Memory Pooling with Multiple Logical Devices



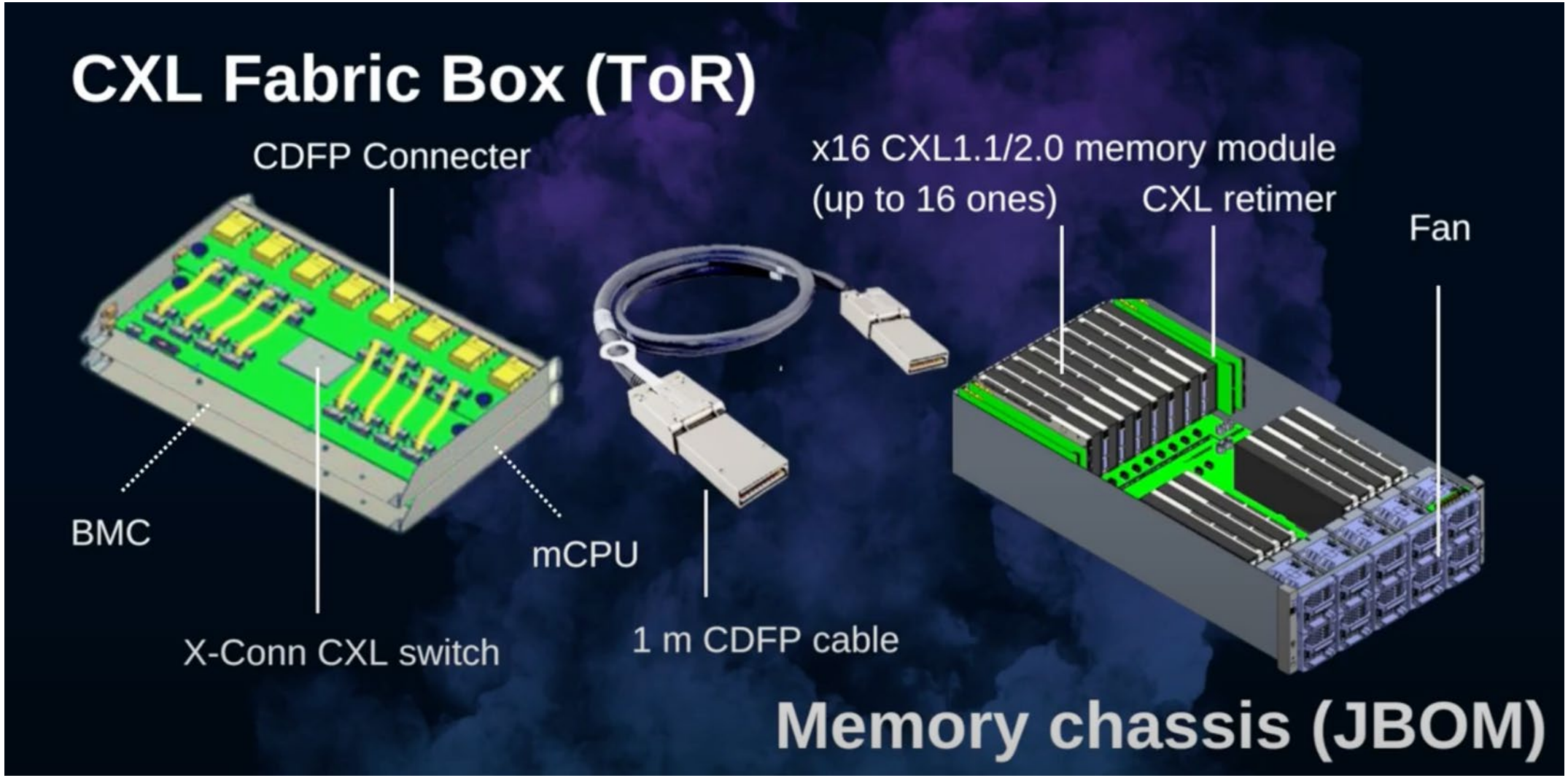
- One single XC50256 connects up to 32 combined hosts/devices
- Fully support CXL Fabric Manager
- Support switch cascading, enable a larger scale fabric

Xconn CXL Switch Silicon and Reference Boards





CXL Memory Pooling POC

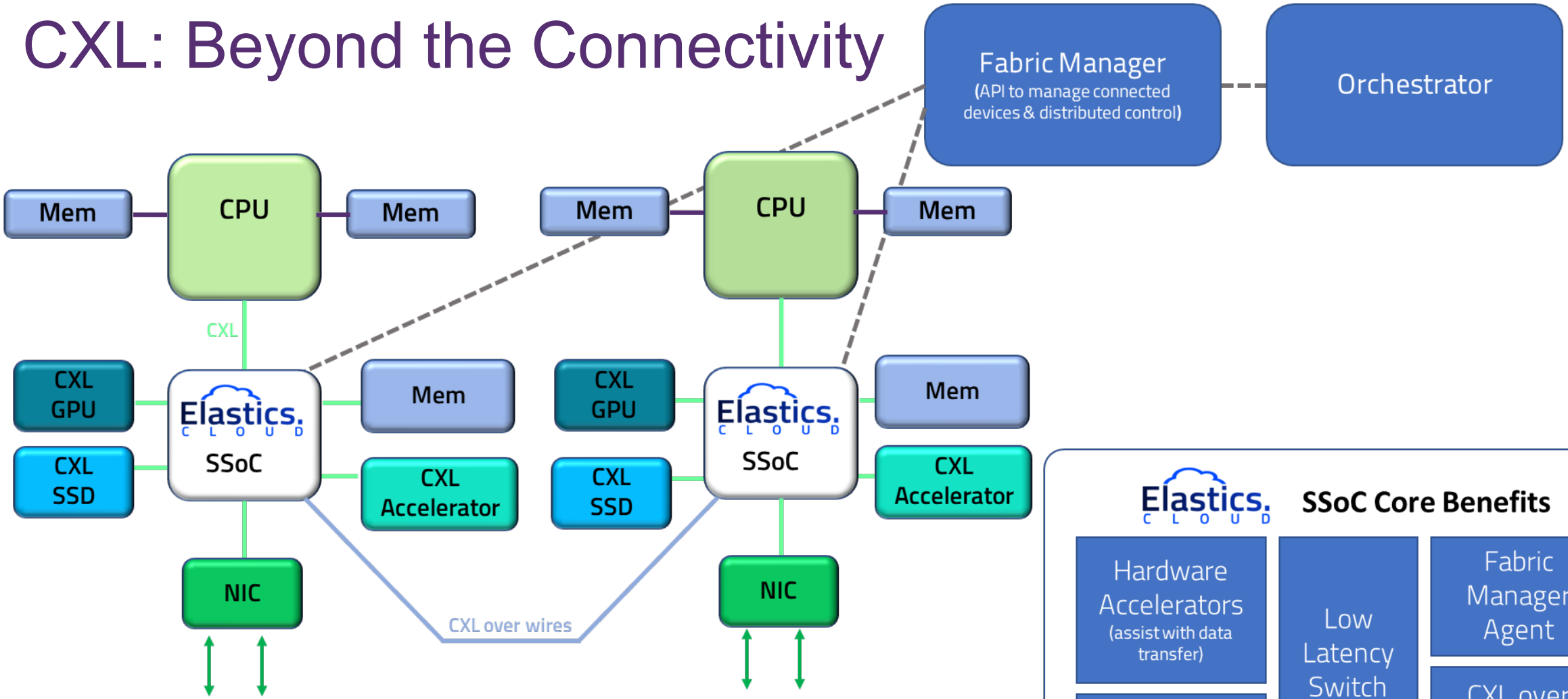


SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

CXL: Beyond the Connectivity

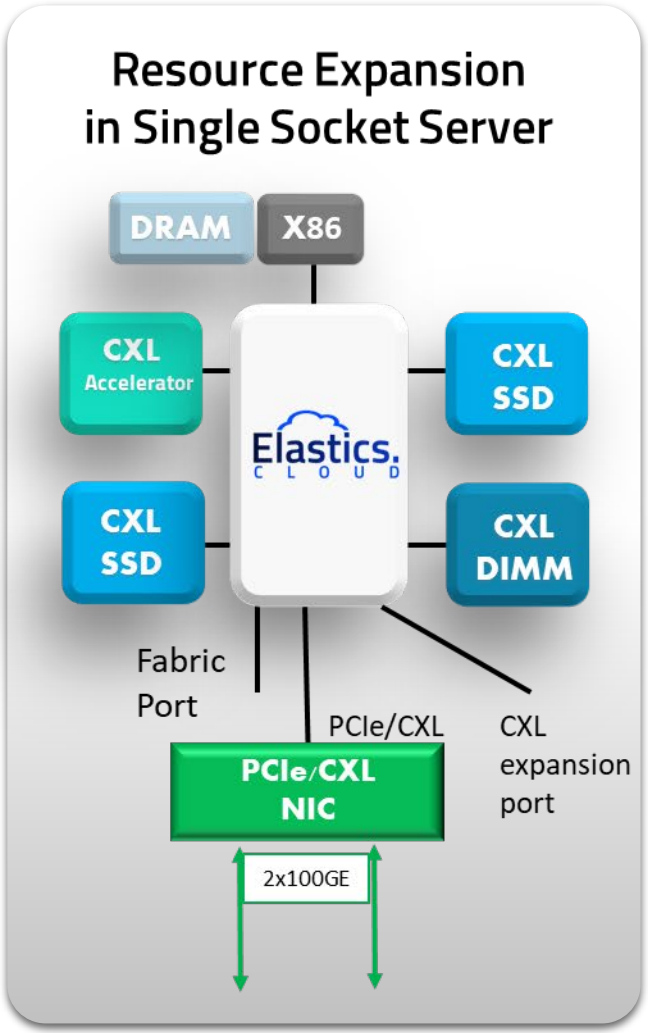
CXL: Beyond the Connectivity

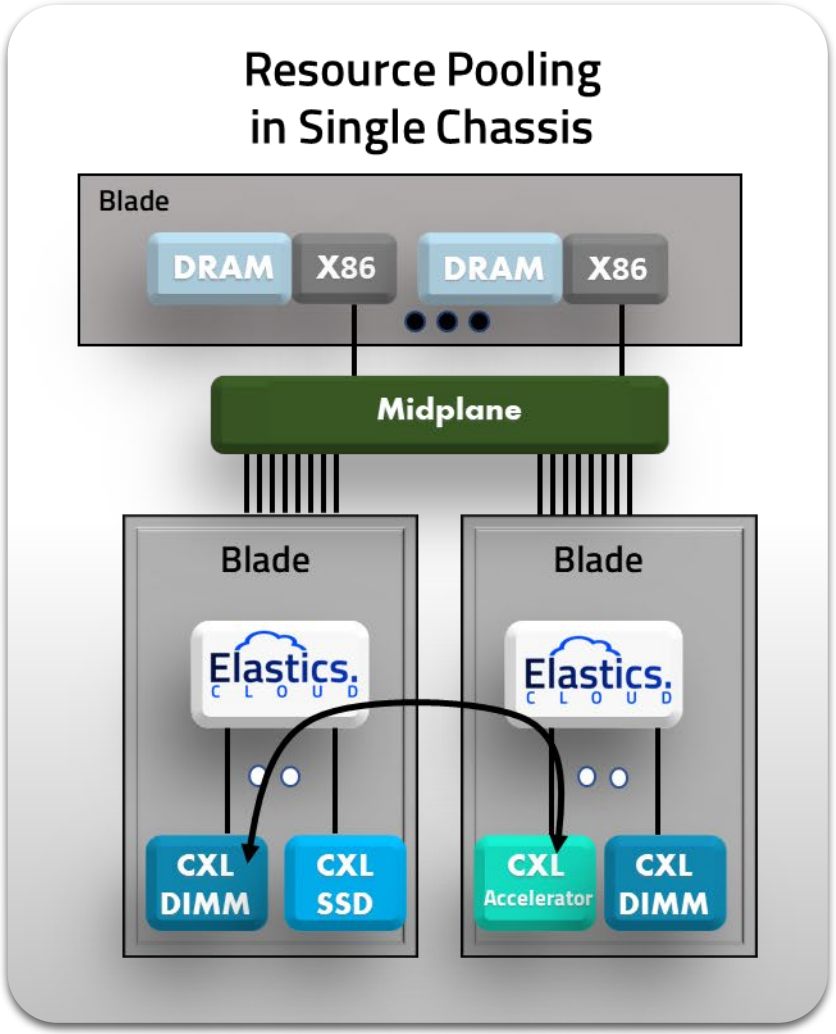


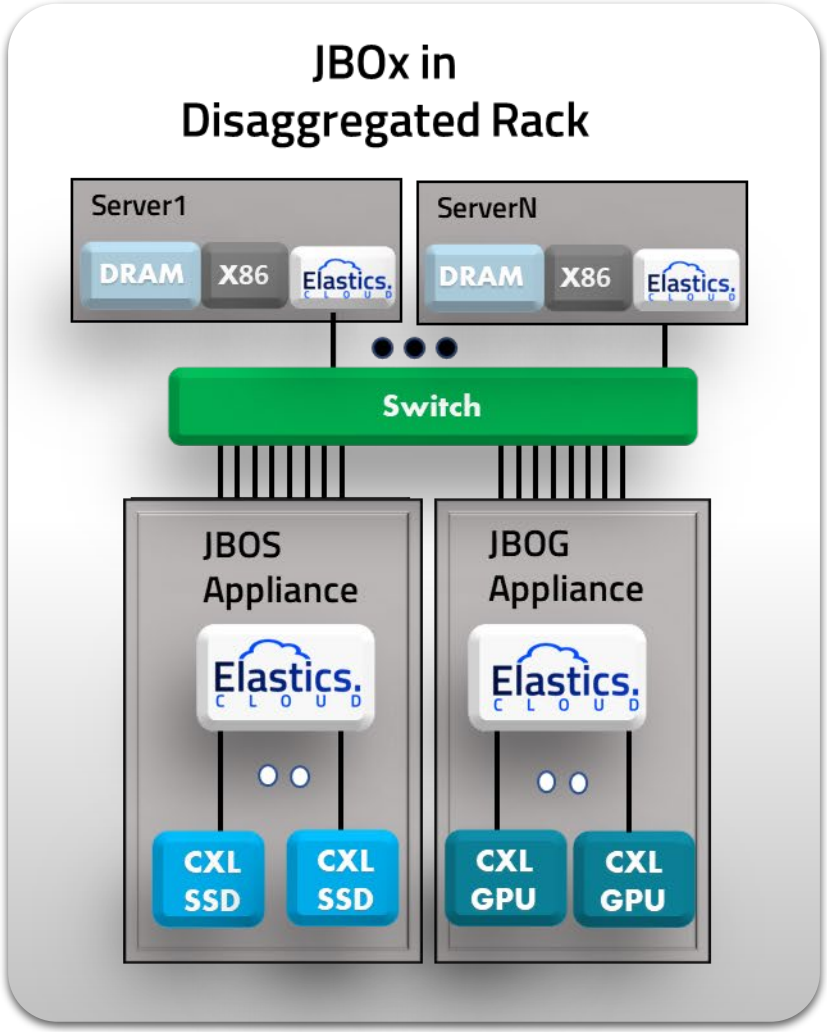
Elastics. CLOUD SSoC Core Benefits

- Hardware Accelerators (assist with data transfer)
- Smart Caching
- Low Latency Switch Fabric
- Fabric Manager Agent
- CXL over wires (expand reach rack to rack)

CXL: Extending and Expanding Device Pools







We are currently demonstrating these use cases in a CXL 1.1 compliant server

SNIA COMPUTE + MEMORY + STORAGE SUMMIT

Architectures, Solutions, and Community
VIRTUAL EVENT, APRIL 11-12, 2023

Please take a moment to rate this session.

Your feedback is important to us.

