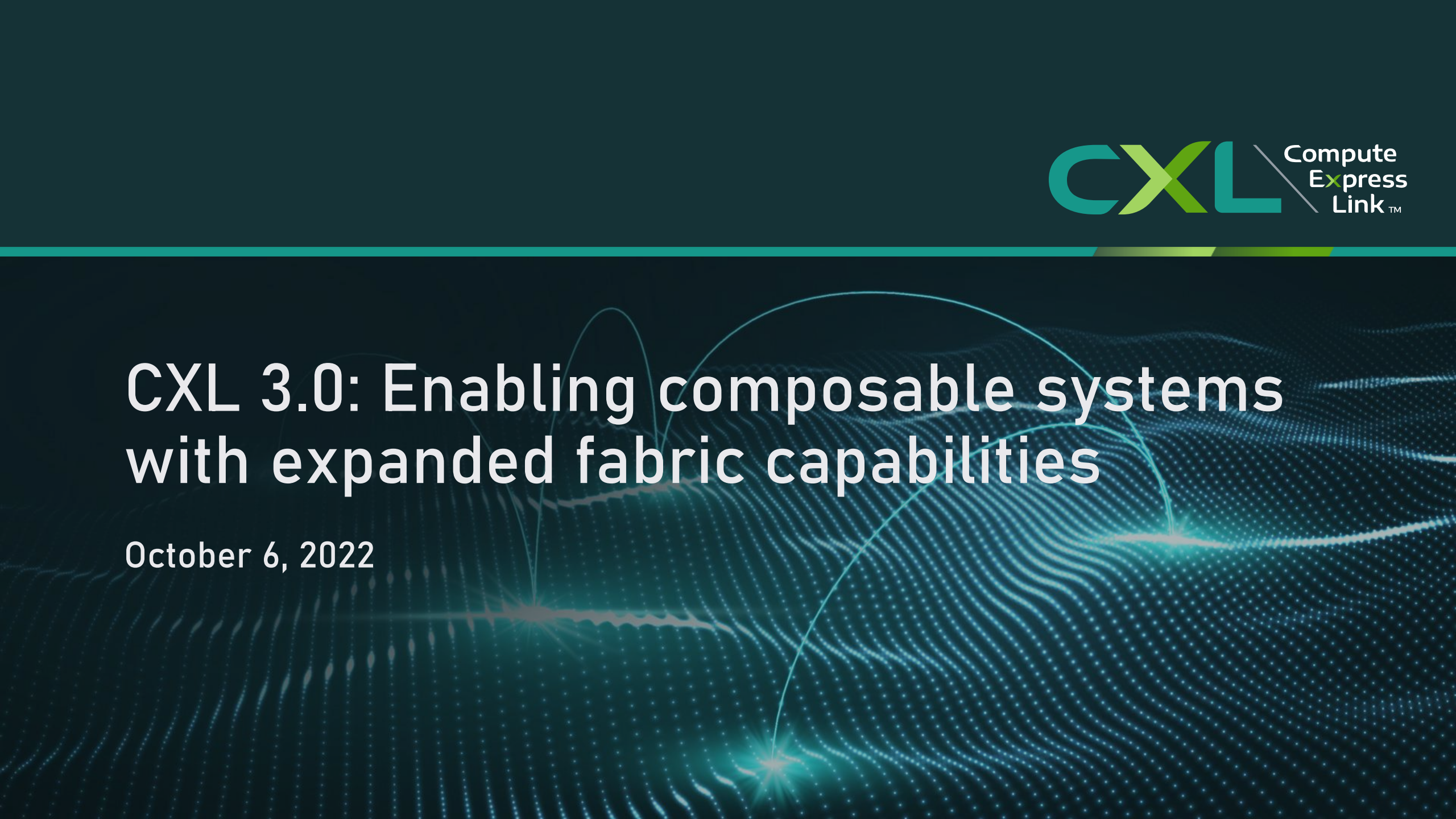


CXL 3.0: Enabling composable systems with expanded fabric capabilities

October 6, 2022

Meet the Presenters

Danny Moore

CXL Consortium MWG Contributor, Senior Manager Strategy and Product Management, Rambus

Dr. Debendra Das Sharma

CXL Consortium Technical Task Force Co-Chair and Intel Senior Fellow

CXL Board of Directors

Industry Open Standard for High Speed Communications

200+ Member Companies

CXL Specification Release Timeline

March 2019

- CXL 1.0 Specification Released

September 2019

- CXL Consortium Officially Incorporates
- CXL 1.1 Specification Released

November 2020

- CXL 2.0 Specification Released

August 2022

- CXL 3.0 Specification Release

Agenda

- Industry Landscape and CXL
- CXL 1.0 and CXL 2.0 recap
- CXL 3.0 Features
- Conclusions

Industry Landscape

Proliferation of
Cloud Computing

Growth of
AI & Analytics

Cloudification of
the Network & Edge

CXL Overview

- New breakthrough high-speed fabric
  - Enables a high-speed, efficient interconnect between CPU, memory and accelerators
  - Builds upon PCI Express® (PCIe®) infrastructure, leveraging the PCIe® physical and electrical interface
  - Maintains memory coherency between the CPU memory space and memory on CXL attached devices
    - Enables fine-grained resource sharing for higher performance in heterogeneous compute environments
    - Enables memory disaggregation, memory pooling and sharing, persistent memory and emerging memory media
- Delivered as an open industry standard
  - CXL Specification 3.0 will be available in August with full backward compatibility with CXL 1.1 and CXL 2.0
  - Future CXL Specification generations will include continuous innovation to meet industry needs and support new technologies

CXL Approach

Coherent Interface

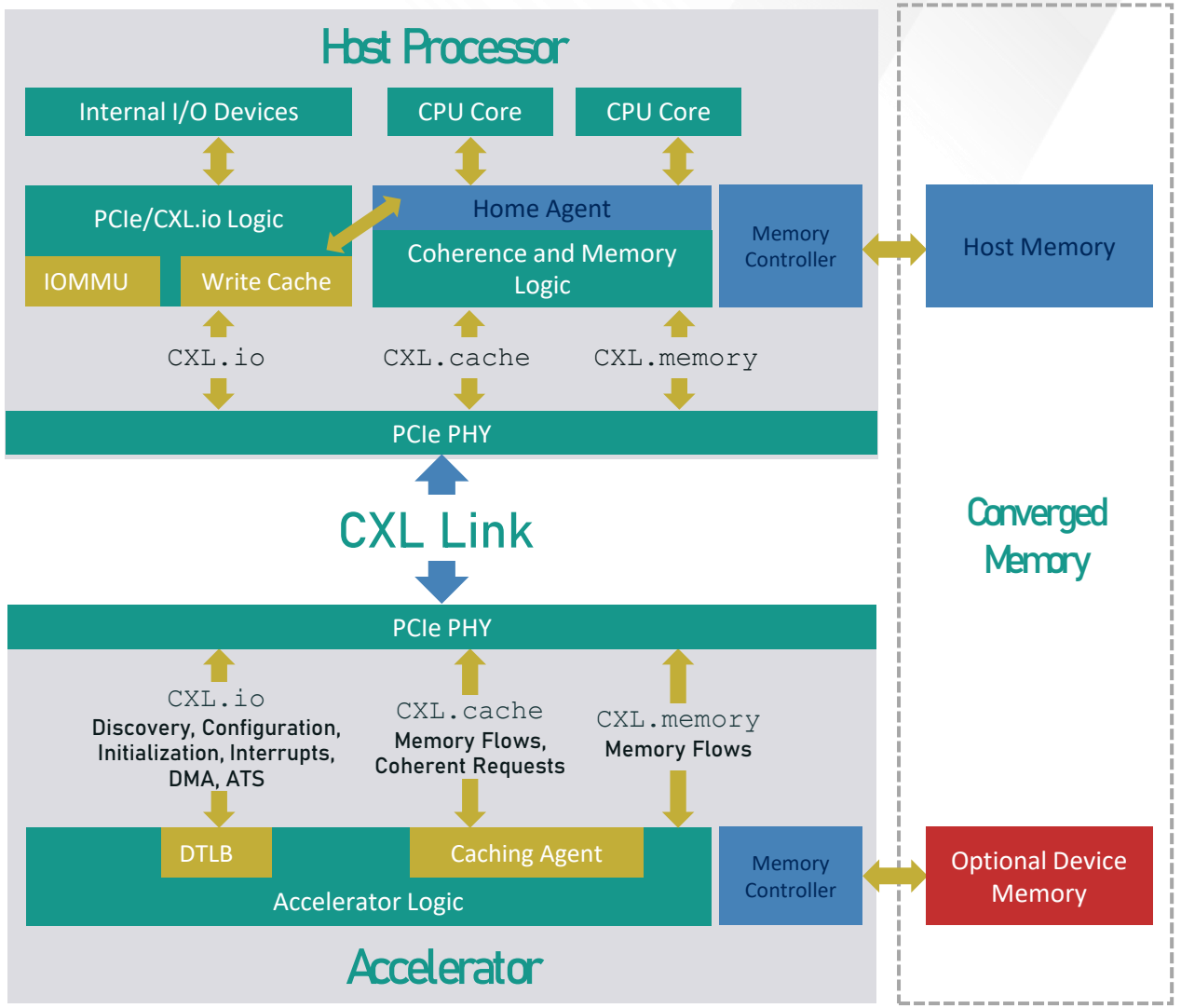
- Leverages PCIe with three multiplexed protocols
- Built on top of PCIe® infrastructure

Low Latency

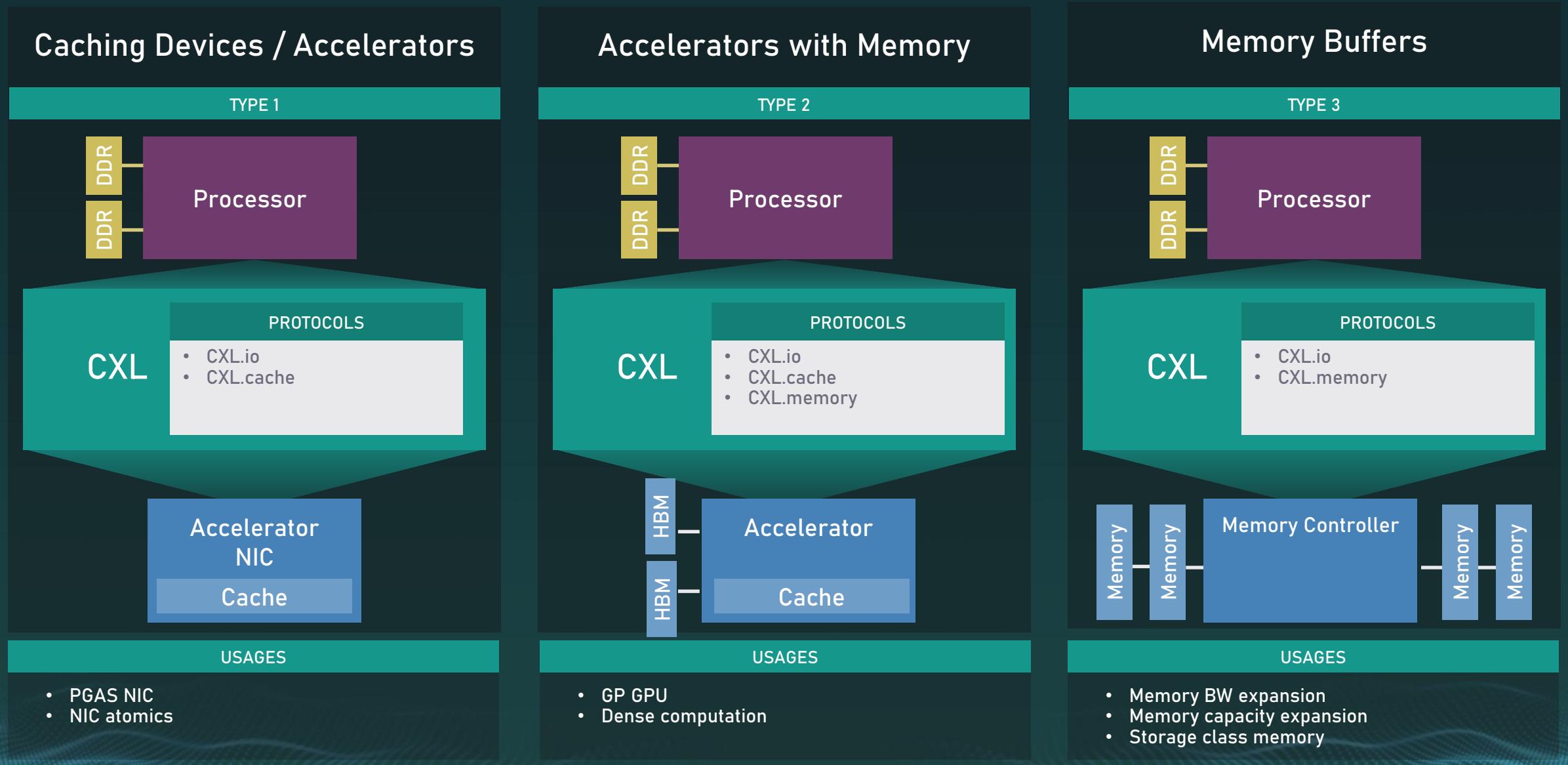
- `CXL.Cache/CXL.Memory` targets near CPU cache coherent latency (<200ns load to use)

Asymmetric Complexity

- Eases burdens of cache coherence interface designs for devices

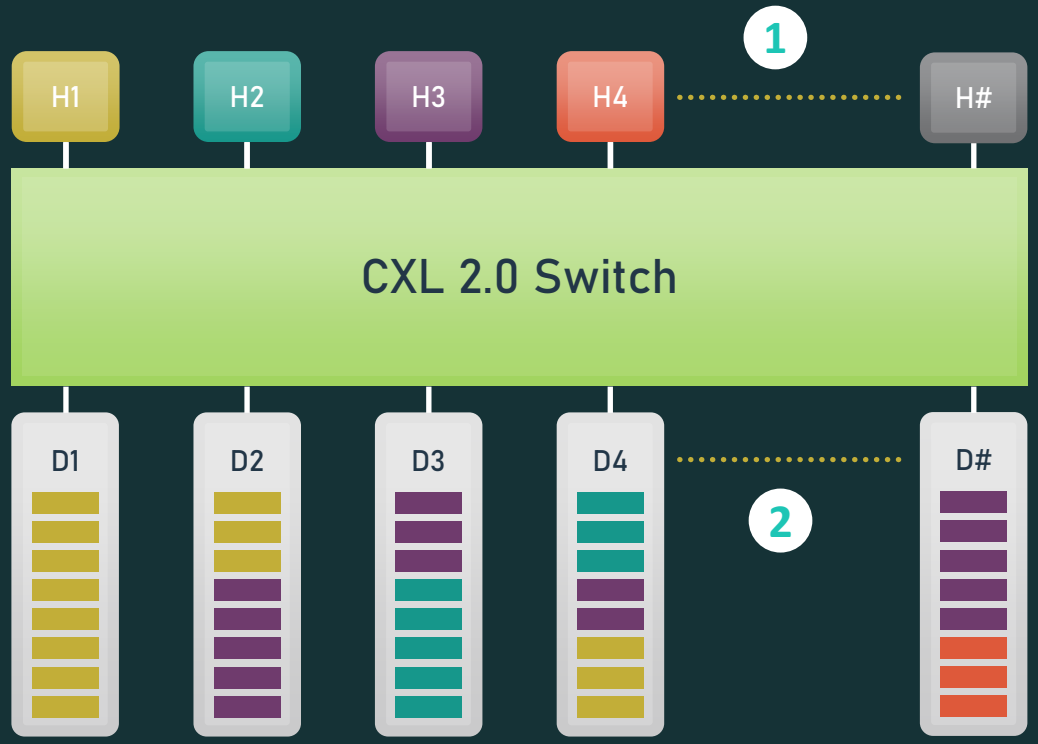


Representative CXL Usages with CXL 1.0



CXL 2.0 Feature Summary

MEMORY POOLING

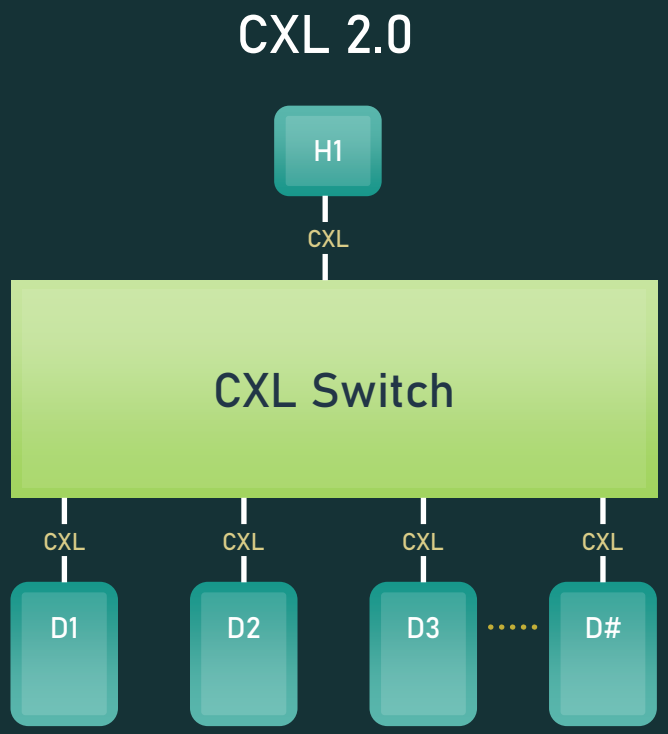


1. Device memory can be allocated across multiple hosts.

2. Multi Logical Devices allow for finer grain memory allocation

CXL 2.0 Feature Summary

SWITCH CAPABILITY



- Supports single-level switching
- Enables memory expansion and resource allocation

CXL 2.0: Resource Pooling at Rack Level, Persistent Memory Support and Enhanced Security

- Resource pooling/disaggregation
  - Managed hot-plug flows to move resources
  - Type-1/Type-2 device assigned to one host
  - Type-3 device (memory) pooling at rack level
  - Direct load-store, low-latency access – similar to memory attached in a neighboring CPU socket
    (vs. RDMA over network)
- Persistence flows for persistent memory
- Fabric Manager/API for managing resources
- Security: authentication, encryption
- Beyond node to rack-level connectivity!

Disaggregated system with CXL optimizes resource utilization delivering lower TCO and power efficiency

CXL 3.0 Specification

Industry trends

- Use cases driving need for higher bandwidth include: high performance accelerators, system memory, SmartNIC and leading edge networking
- CPU efficiency is declining due to reduced memory capacity and bandwidth per core
- Efficient peer-to-peer resource sharing across multiple domains
- Memory bottlenecks due to CPU pin and thermal constraints

CXL 3.0 introduces…

- Fabric capabilities
  - Multi-headed and fabric attached devices
  - Enhance fabric management
  - Composable disaggregated infrastructure
- Improved capability for better scalability and resource utilization
  - Enhanced memory pooling
  - Multi-level switching
  - New enhanced coherency capabilities
  - Improved software capabilities
- Double the bandwidth
- Zero added latency over CXL 2.0
- Full backward compatibility with CXL 2.0, CXL 1.1, and CXL 1.0

CXL 3.0 Specification

Fabric capabilities and management

Improved memory sharing and pooling

Enhanced coherency

Peer-to-peer

Expanded capabilities for increasing scale and optimizing resource utilization

- Fabric capabilities and fabric attached memory
- Enhance fabric management framework
- Memory pooling and sharing
- Peer-to-peer memory access
- Multi-level switching
- Near memory processing
- Multi-headed devices
- Multiple Type 1/Type 2 devices per root port
- Fully backward compatible to CXL 2.0, 1.1, and 1.0
- Supports PCIe® 6.0

CXL 3.0 Spec Feature Summary

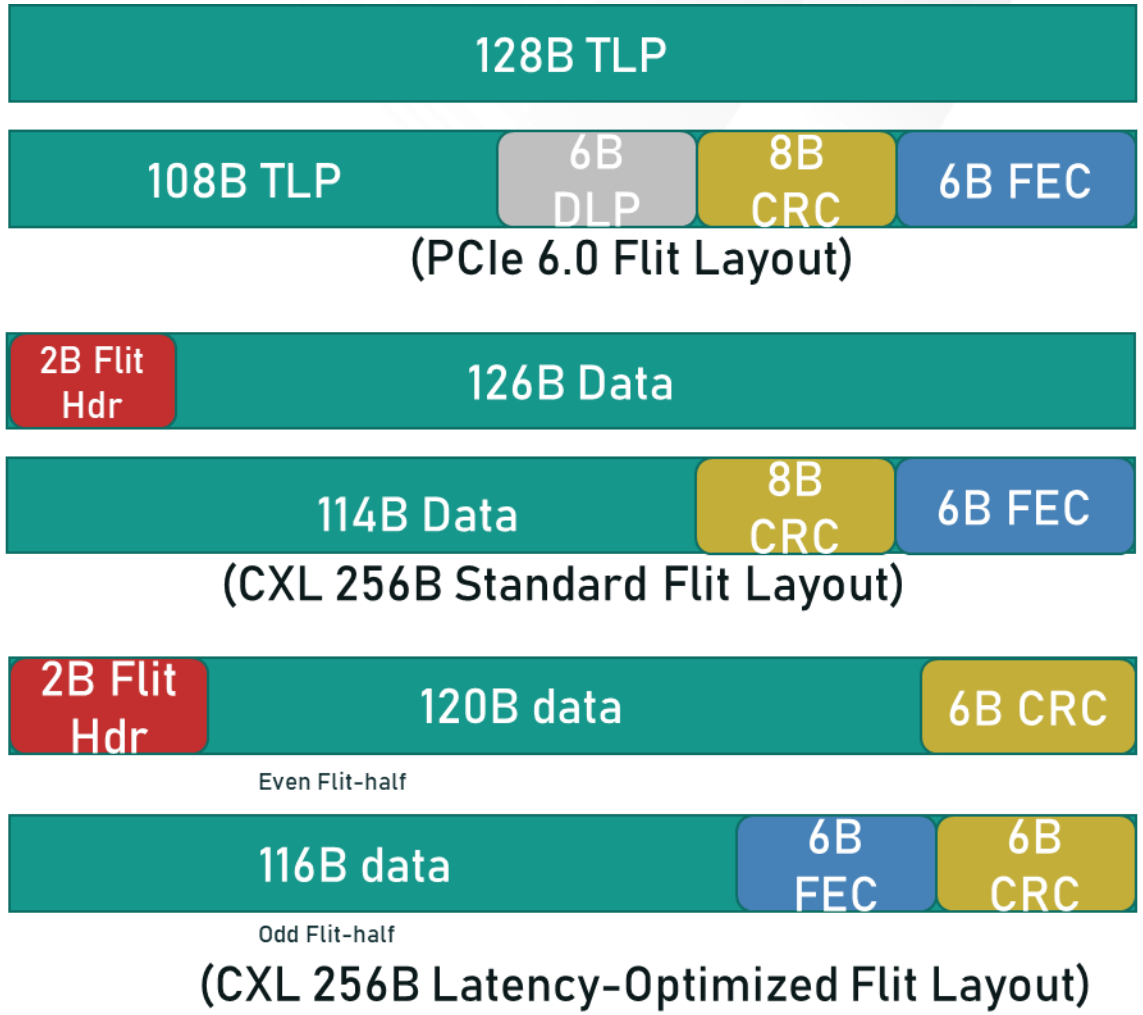
Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0
Release date	2019	2020	1H 2022
Max link rate	32GTs	32GTs	64GTs
Flit 68 byte (up to 32 GTs)	✓	✓	✓
Flit 256 byte (up to 64 GTs)			✓
Type 1, Type 2 and Type 3 Devices	✓	✓	✓
Memory Pooling w/ MLDs		✓	✓
Global Persistent Flush		✓	✓
CXL IDE		✓	✓
Switching (Single-level)		✓	✓
Switching (Multi-level)			✓
Direct memory access for peer-to-peer			✓
Enhanced coherency (256 byte flit)			✓
Memory sharing (256 byte flit)			✓
Multiple Type 1/Type 2 devices per root port			✓
Fabric capabilities (256 byte flit)			✓

Not supported

✓ Supported

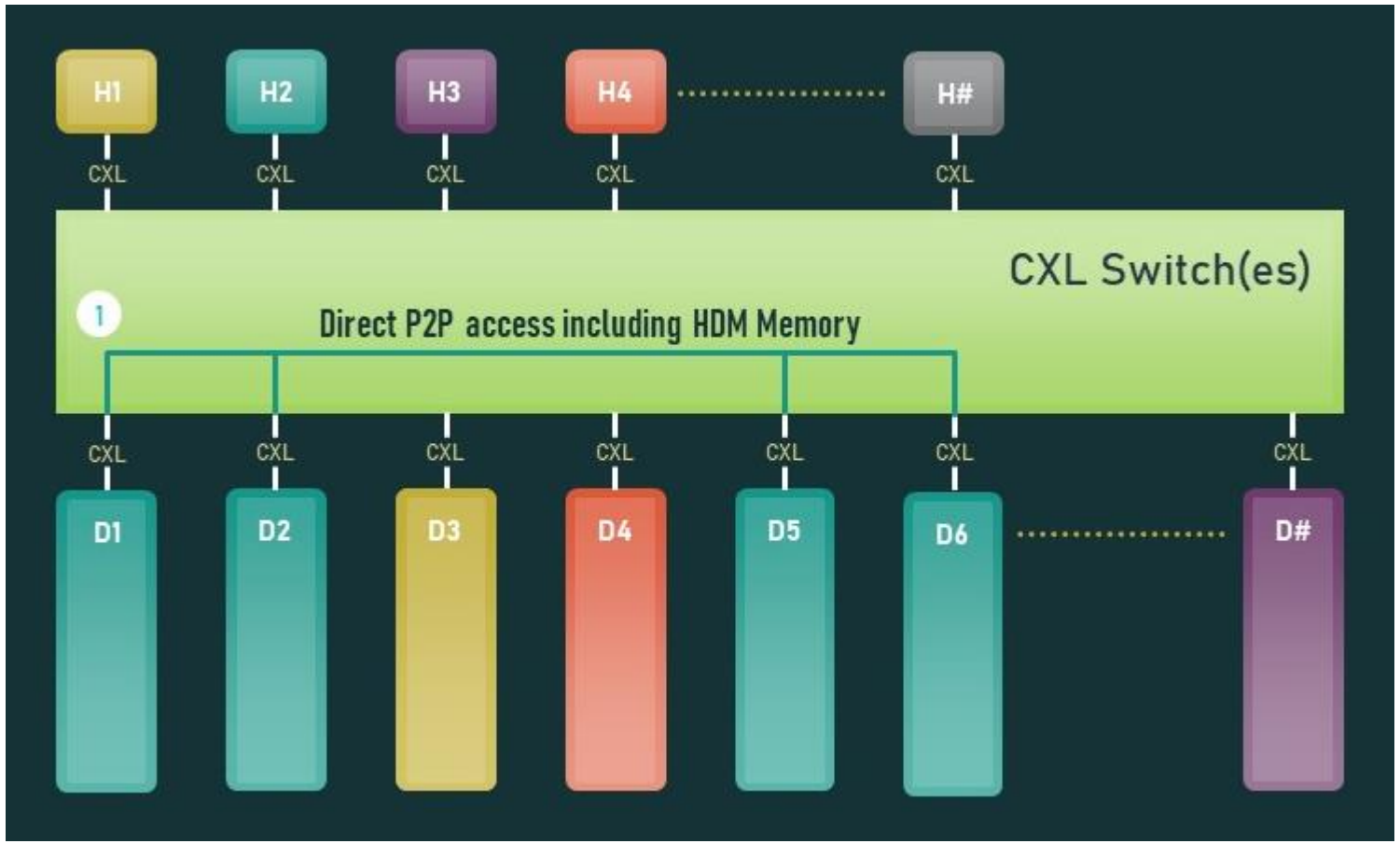
CXL 3.0: Doubles Bandwidth with Same Latency

- Uses PCIe 6.0® PHY @ 64 GT/s
- PAM-4 and high BER mitigated by PCIe 6.0 FEC and CRC (different CRC for latency optimized)
- Standard 256B Flit along with an additional 256B Latency Optimized Flit (0-latency adder over CXL 2)
  - 0-latency adder trades off FIT (failure in time, 109 hours) from 5x10-8 to 0.026 and Link efficiency impact from 0.94 to 0.92 for 2-5ns latency savings (x16 – x4)1
- Extends to lower data rates (8G, 16G, 32G)
- Enables several new CXL 3 protocol enhancements with the 256B Flit format



1: D. Das Sharma, "A Low-Latency and Low-Power Approach for Coherency and Memory Protocols on PCI Express 6.0 PHY at 64.0 GT/s with PAM-4 Signaling", IEEE Micro, Mar/ Apr 2022 (https://ieeexplore.ieee.org/document/9662217)

CXL 3.0 Protocol Enhancements (UIO and BI) for Device to Device Connectivity

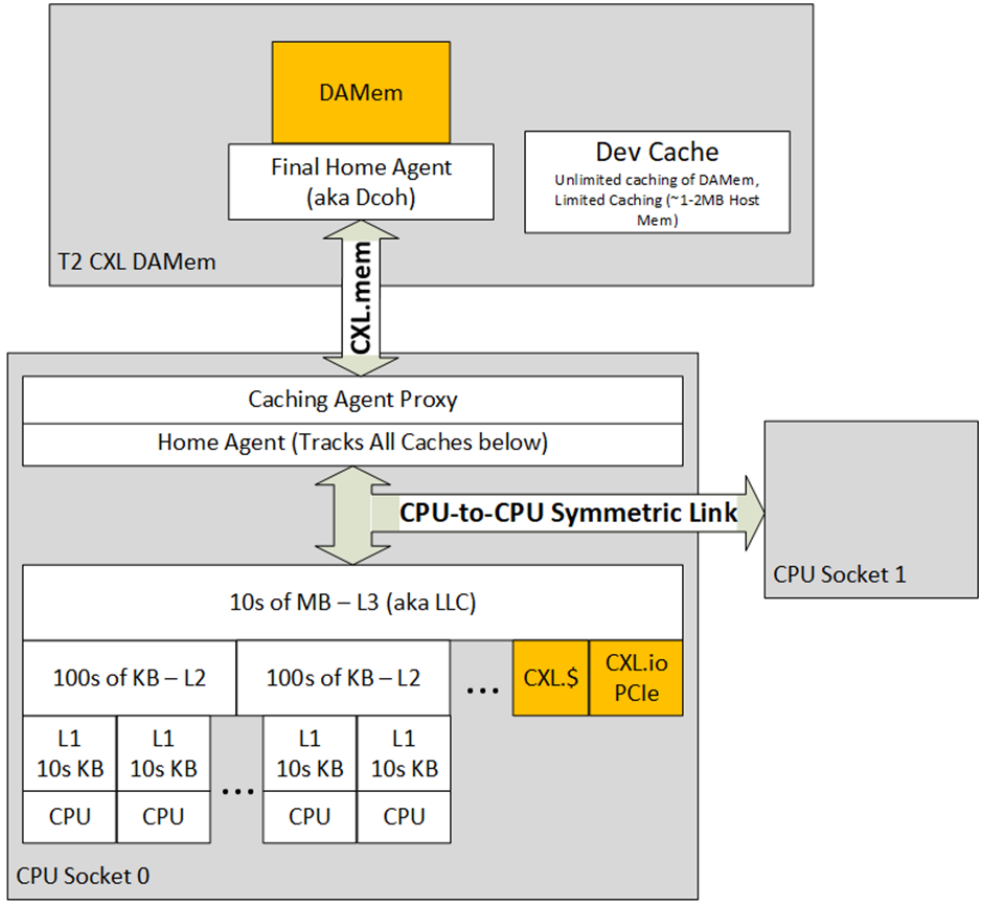


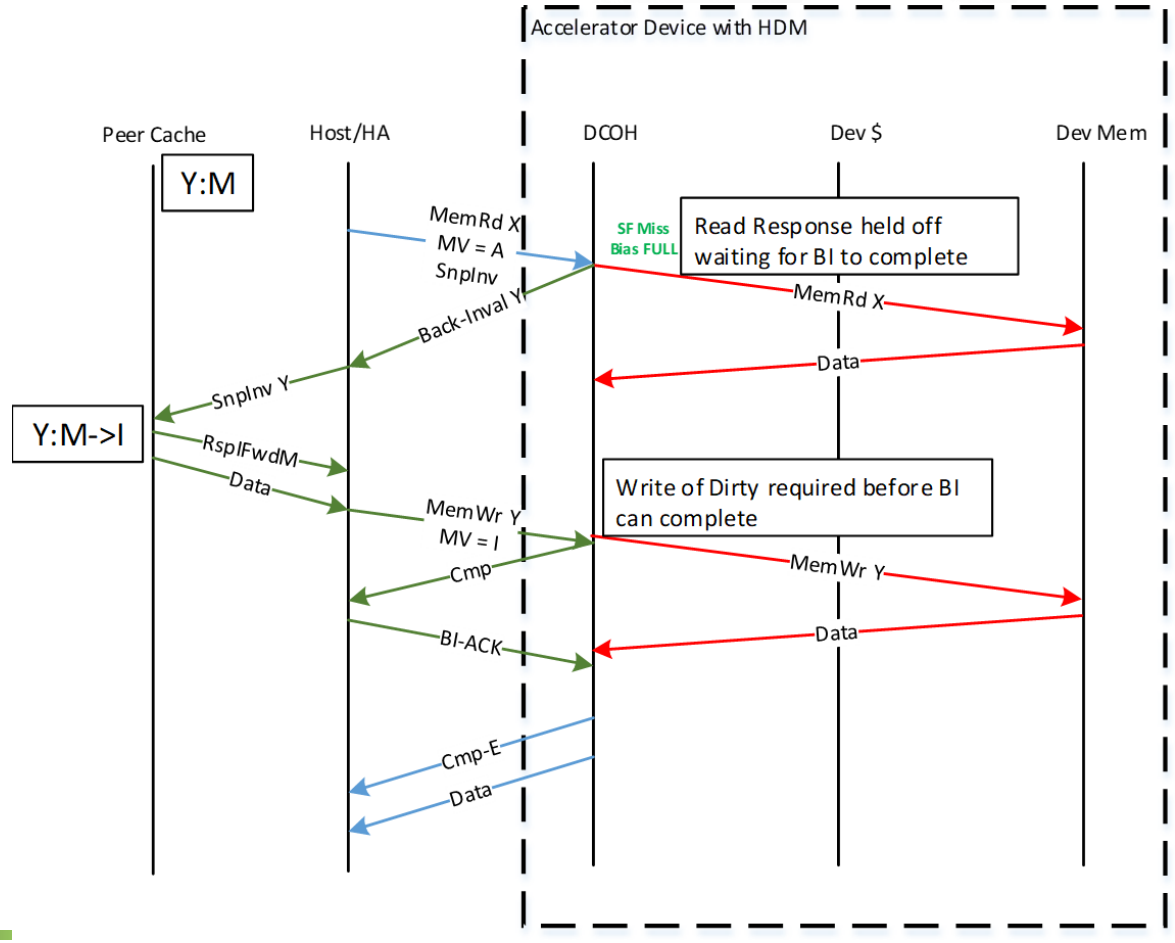
CXL 3.0 enables **non-tree topologies and peer-to-peer communication (P2P)** within a virtual hierarchy of devices

- Virtual hierarchies are associations of devices that maintains a coherency domain
- P2P to HDM-DB memory is I/O Coherent: a new Unordered I/O (UIO) Flow in CXL.io – the Type-2/3 device that hosts the memory will generate a new Back-Invalidation flow (CXL.Mem) to the host to ensure coherency if there is a coherency conflict

CXL 3.0 Protocol Enhancements: Mapping Large memory in Type-2 Devices to HDM with Back Invalidate

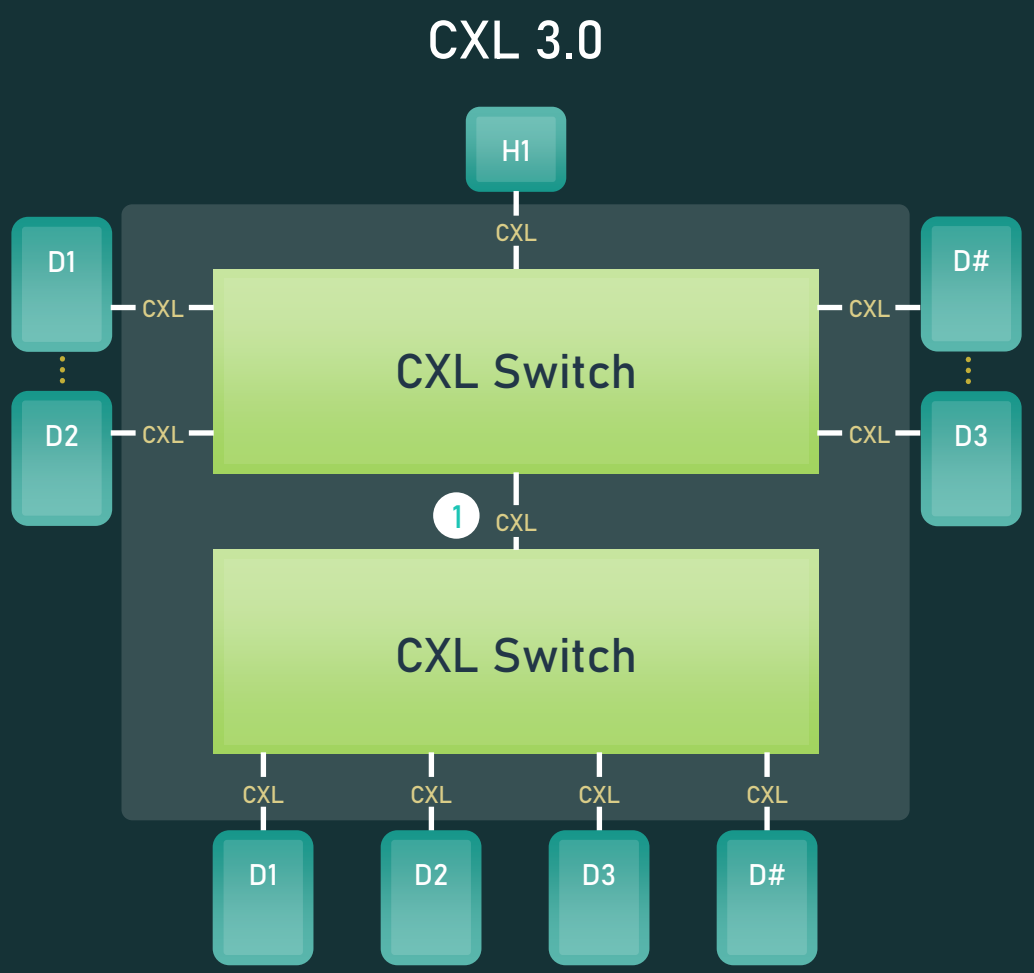
Existing Bias – Flip mechanism needed HDM to be tracked fully since device could not back snoop the host. Back Invalidate with CXL 3.0 enables snoop filter implementation resulting in large memory that can be mapped to HDM

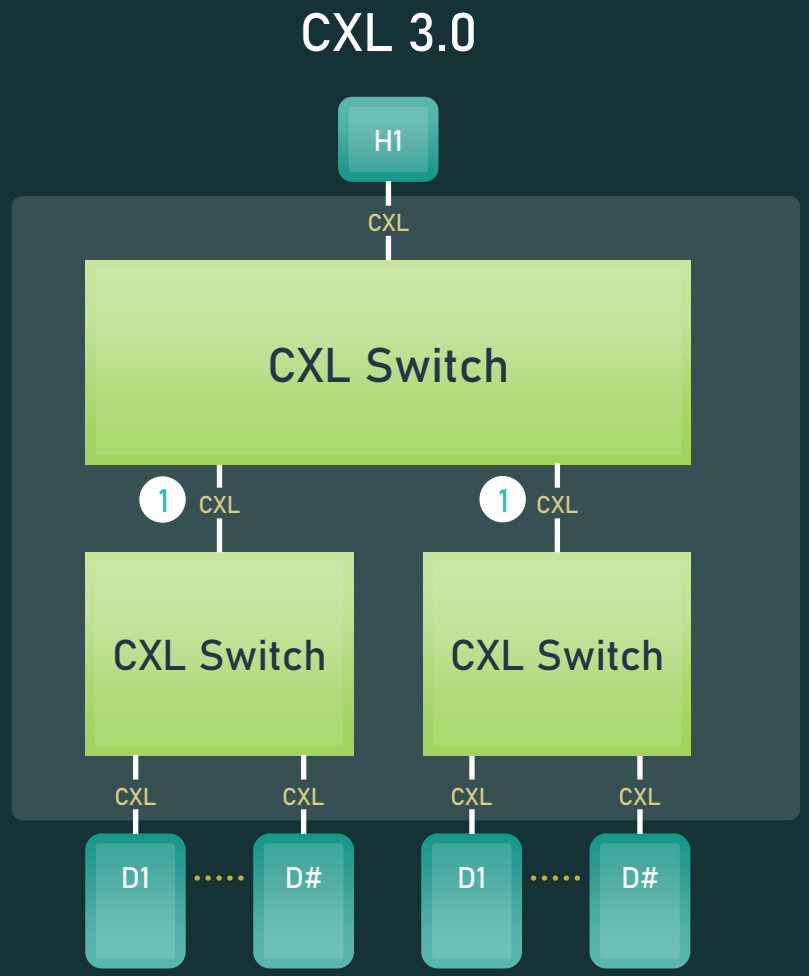




CXL 3.0: Switch Cascade/Fanout

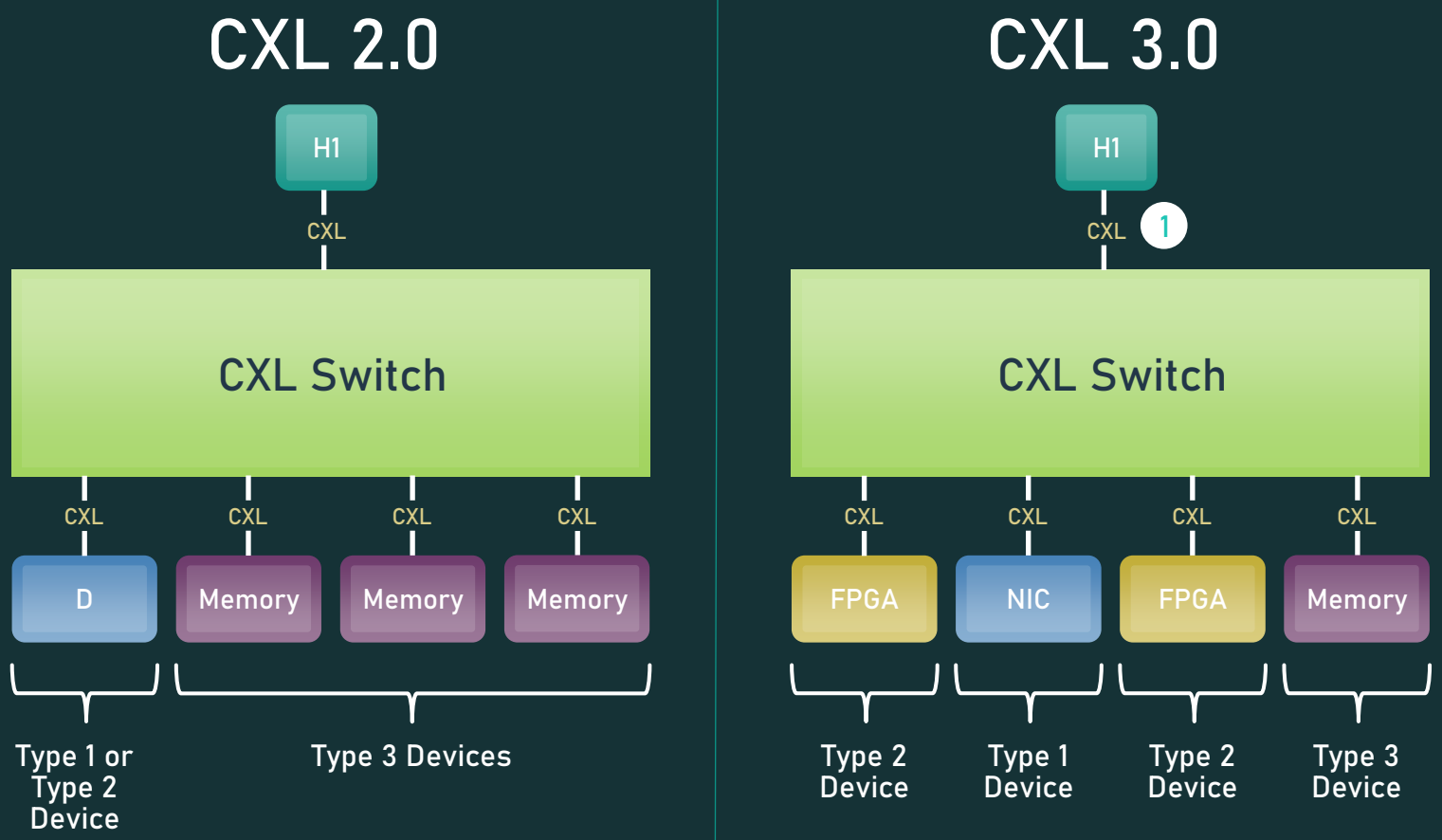
Supporting vast array of switch topologies





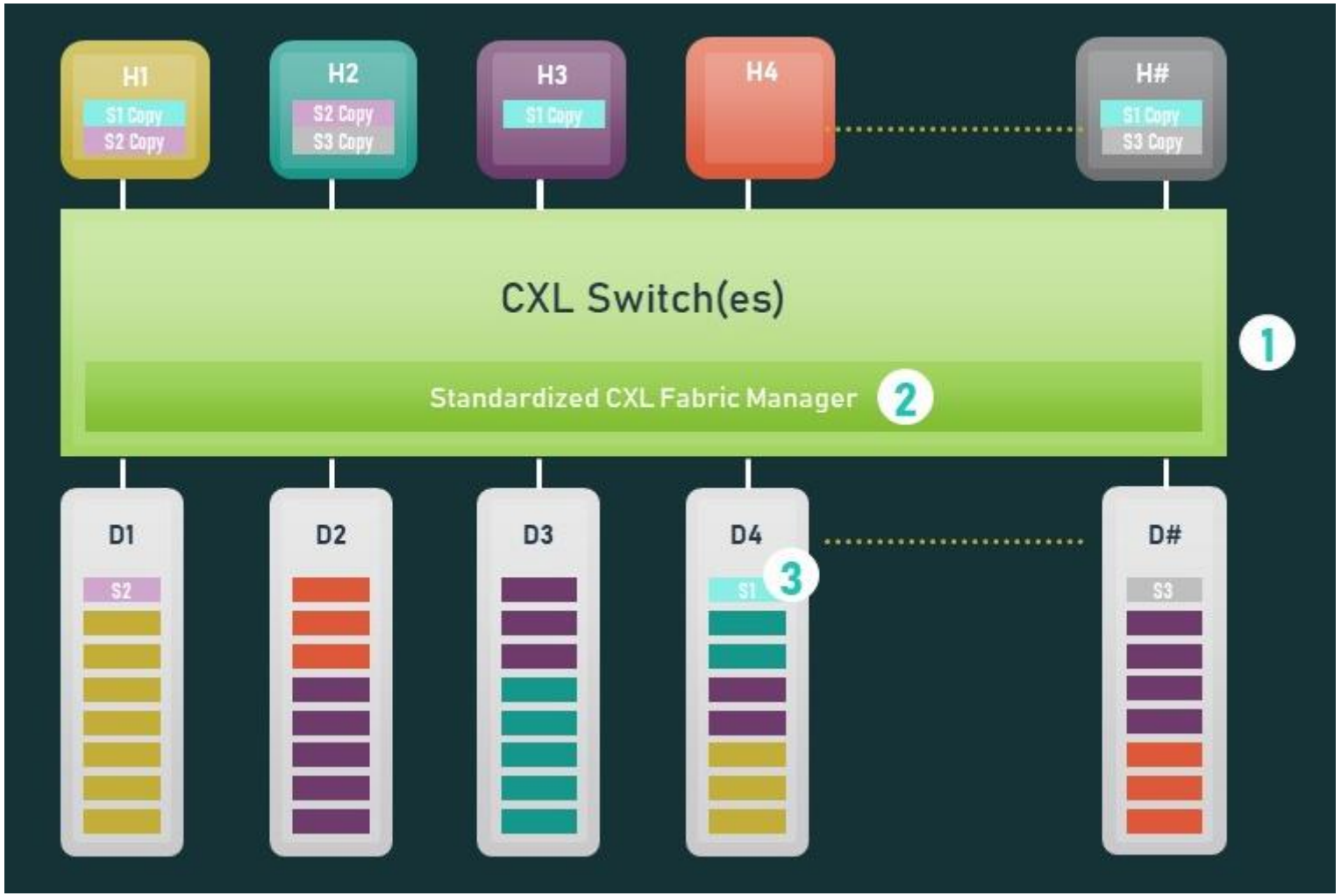
1. Multiple switch levels (aka cascade)
   - Supports fanout of all device types

CXL 3.0: Multiple Devices of all Types Per Root Port



1 Each host's root port can connect to more than one device type

CXL 3.0: Pooling & Sharing

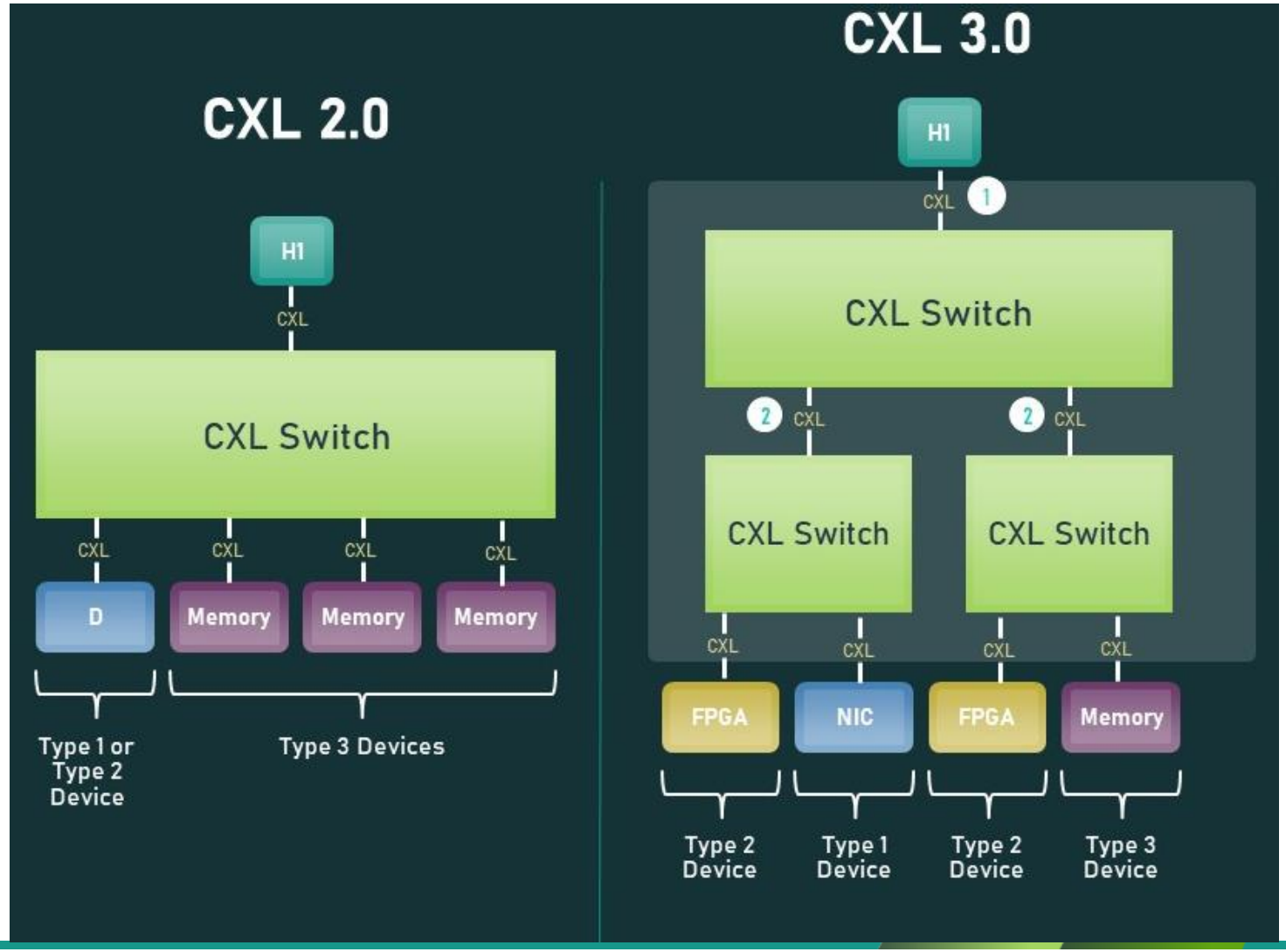


1. Expanded use case showing memory sharing and pooling

2. CXL Fabric Manager is available to setup, deploy, and modify the environment

3. Shared Coherent Memory across hosts using hardware coherency (directory + Back-Invalidate Flows). Allows one to build large clusters to solve large problems through shared memory constructs. Defines a Global Fabric Attached Memory (GFAM) which can provide access to up to 4095 entities

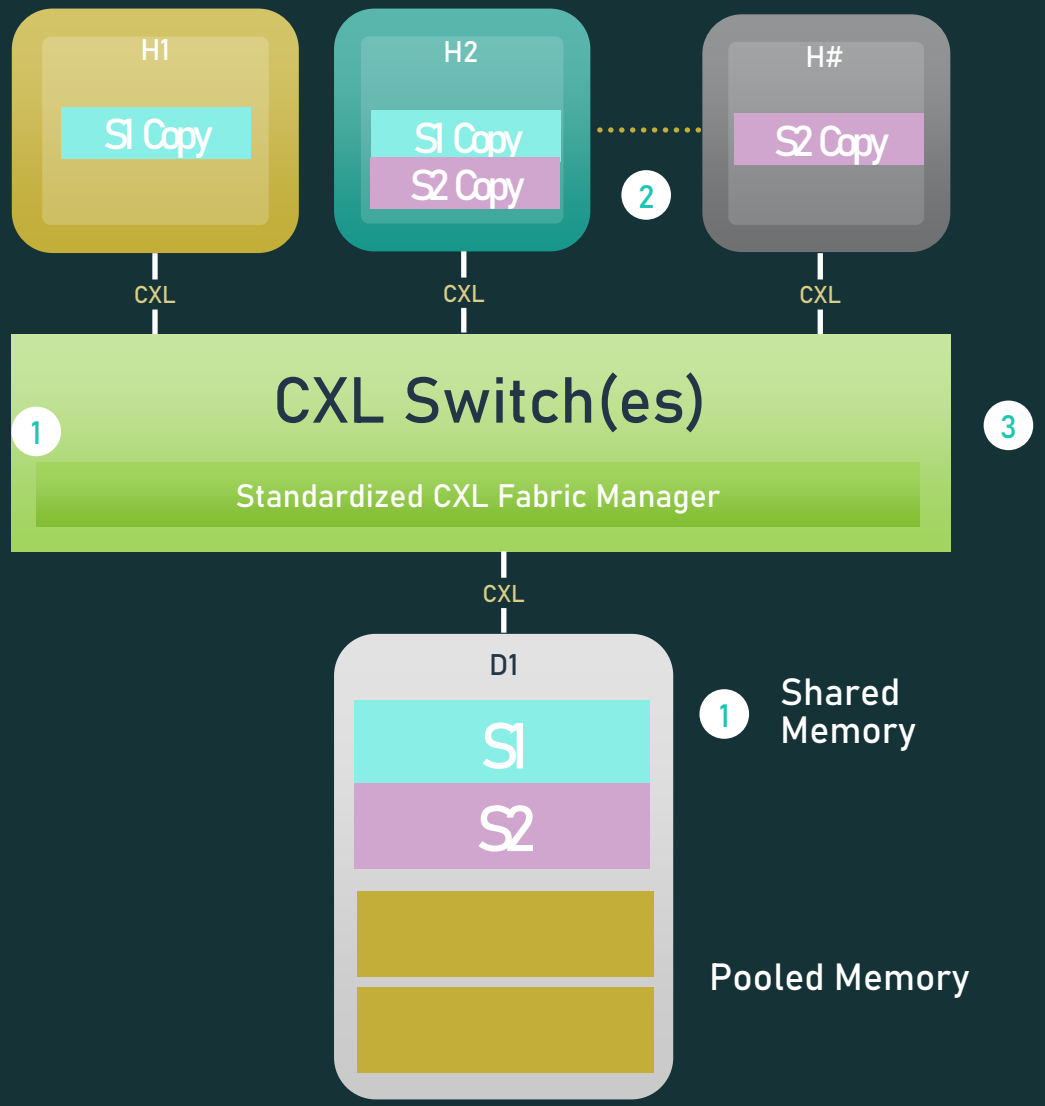
CXL 3.0: Multiple Level Switching, Multiple Type-1/2



1. Each host's root port can connect to more than one device type (up to 16 CXL.cache devices)

2. Multiple switch levels (aka cascade)
   - Supports fanout of all device types

CXL 3.0: Coherent Memory Sharing



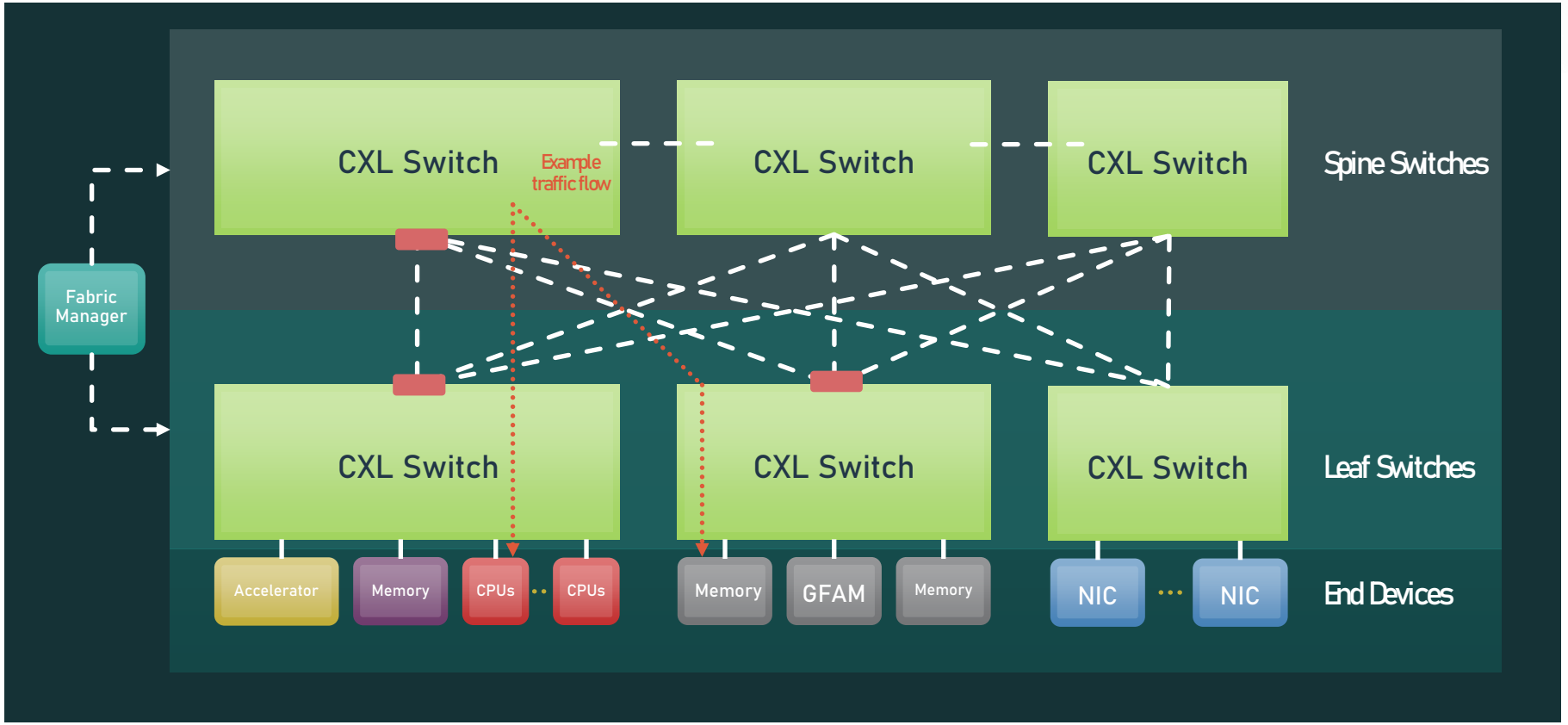
1. Device memory can be shared by all hosts to increase data flow efficiency and improve memory utilization

2. Host can have a coherent copy of the shared region or portions of shared region in host cache

3. CXL 3.0 defined mechanisms to enforce hardware cache coherency between copies

CXL 3.0 Fabrics

Composable Systems with Spine/Leaf Architecture at Rack/ Pod Level



CXL 3.0 Fabric Architecture

- Interconnected Spine Switch System
- Leaf Switch NIC Enclosure
- Leaf Switch CPU Enclosure
- Leaf Switch Accelerator Enclosure
- Leaf Switch Memory Enclosure

Conclusions

- CXL 3.0 features
  - Full fabric capabilities and fabric management
  - Expanded switching topologies
  - Symmetric coherency capabilities
  - Peer-to-peer resource sharing
  - Double the bandwidth and zero added latency compared to CXL 2.0
  - Full backward compatibility with CXL 2.0, CXL 1.1, and CXL 1.0
- Enabling new usage models
  - Memory sharing between hosts and peer devices
  - Support for multi-headed devices
  - [Symmetric coherency capabilities use case]
  - Expanded support for Type-1 and Type-2 devices
  - GFAM provides expansion capabilities for current and future memory
- Call to Action
  - Download the CXL 3.0 specification
  - Support future specification development by joining the CXL Consortium
  - Follow us on Twitter and LinkedIn for updates!

Q&A

Please share your questions in the Question Box

Thank You

Visit www.ComputeExpressLink.org to learn more!