



2023 OFA Virtual Workshop

Introducing Compute Express Link™ (CXL™) 3.0: Expanding Fabric Capabilities and Management

Mahesh Wagh, AMD Sr. Fellow, CXL Consortium TTF Co-Chair

CXL Consortium





CXL Board of Directors



Industry Open Standard for
High Speed Communications

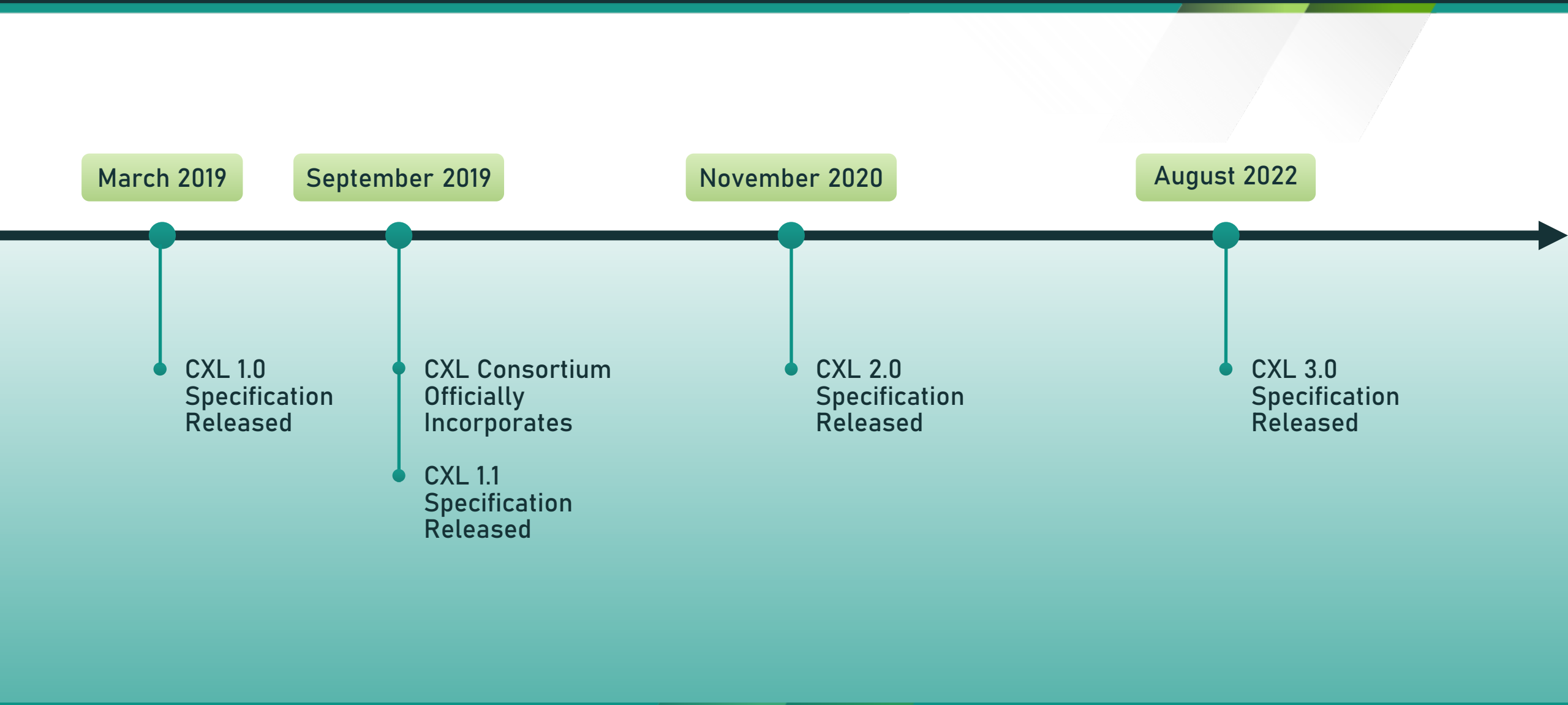
240+ Member Companies

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.



- **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 3.0 specification is fully backward compatible with CXL 1.1 and CXL 2.0
 - Future CXL Specification generations will include continuous innovation to meet industry needs and support new technologies

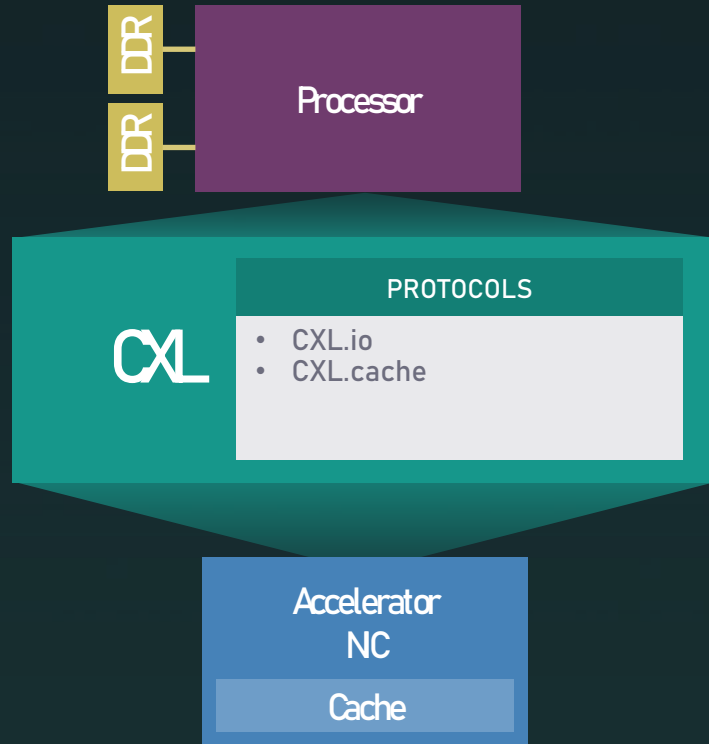
CXL Specification Release Timeline



Recap: CXL 1.0/1.1 Usages

Caching Devices / Accelerators

TYPE 1

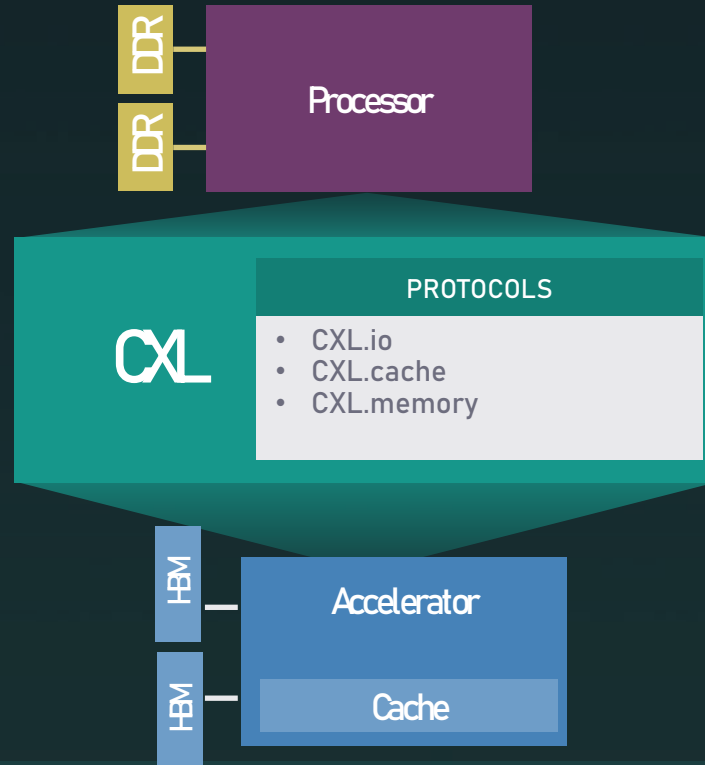


USAGES

- PGAS NIC
- NIC atomics

Accelerators with Memory

TYPE 2

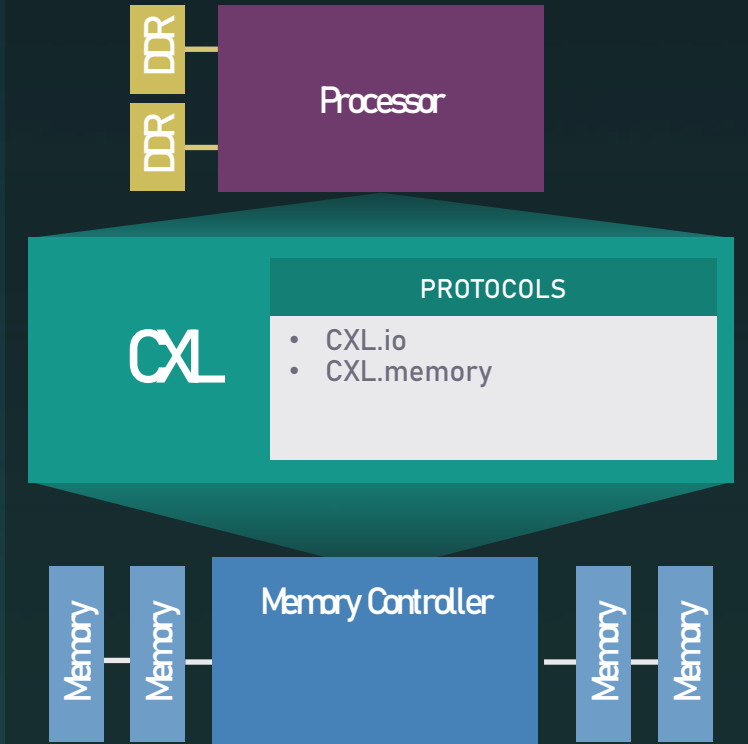


USAGES

- GP GPU
- Dense computation

Memory Buffers

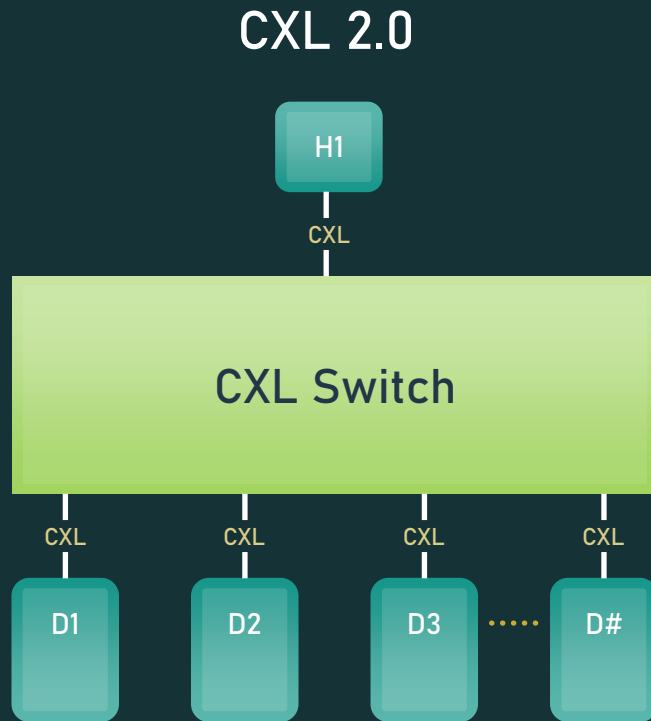
TYPE 3



USAGES

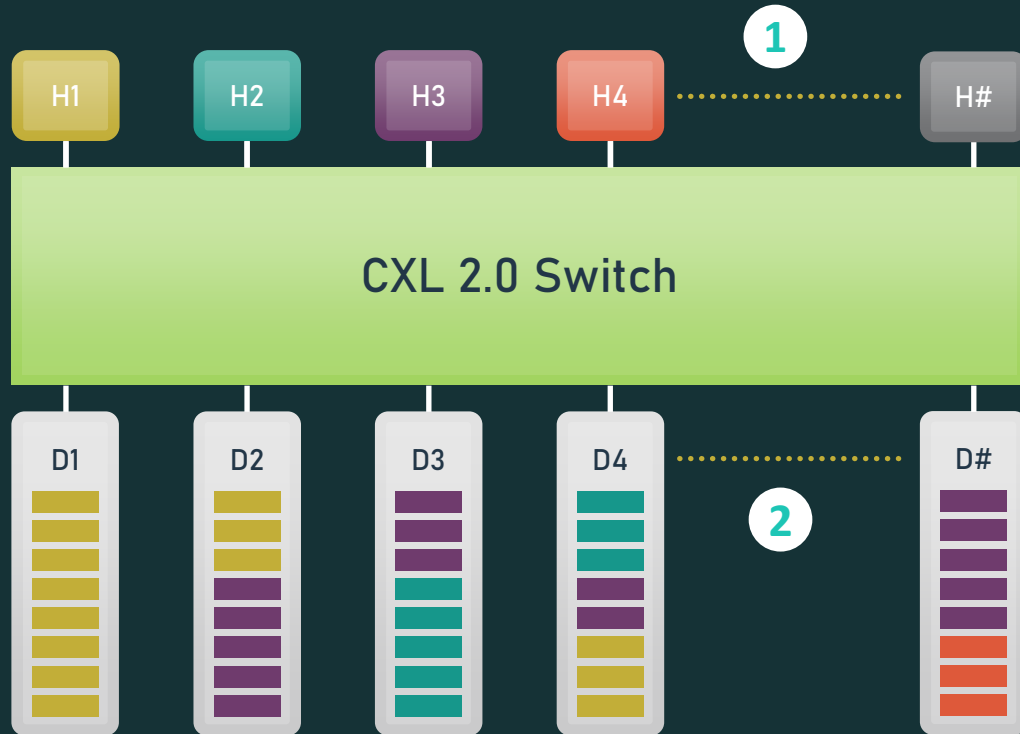
- Memory BW expansion
- Memory capacity expansion
- Storage class memory

Recap: CXL 2.0 Switch Capability



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

Recap: CXL 2.0 Memory Pooling



- 1 Device memory can be allocated across multiple hosts.
- 2 Multi Logical Devices allow for finer grain memory allocation
- 3 Persistence Flows
- 4 Pooling of accelerators
Hot-plug flows

CXL 3.0 Specification

Industry trends

- Use cases driving need for higher bandwidth: e.g., high performance accelerators, system memory, SmartNIC etc.
- CPU capability requiring more memory capacity and bandwidth per core
- Efficient peer-to-peer resource sharing/ messaging across multiple domains
- Memory bottlenecks due to CPU pin and thermal constraints needs to be overcome

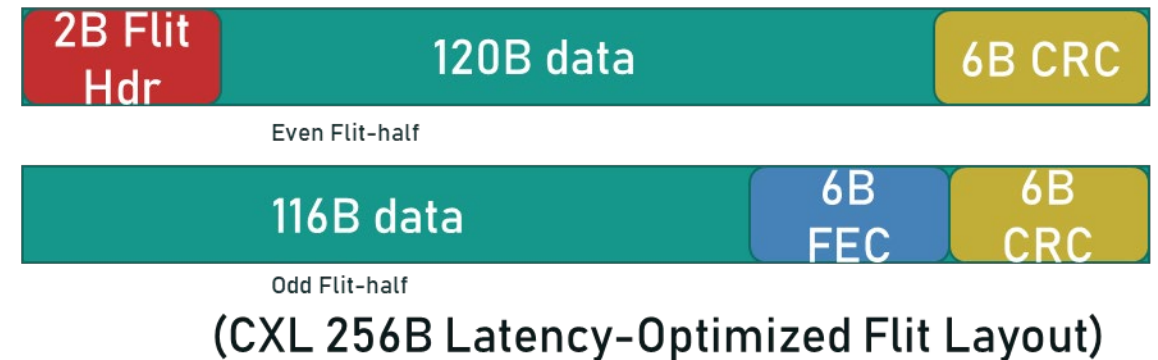
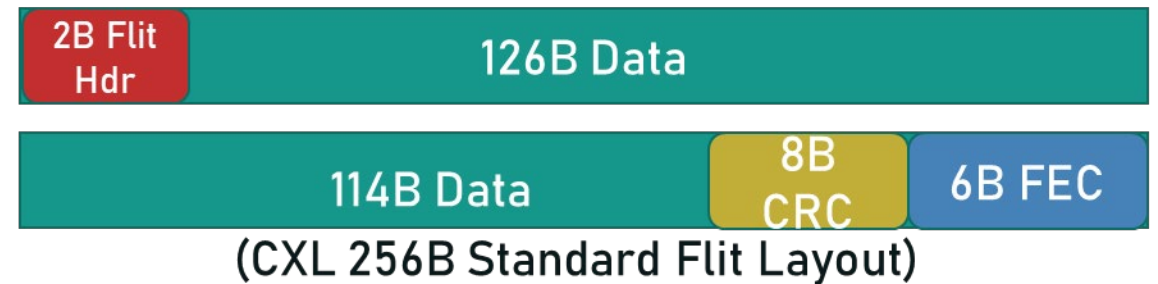
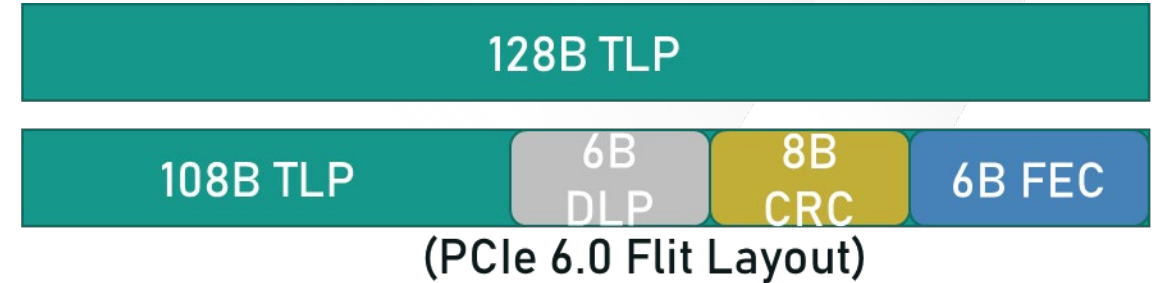
CXL 3.0 introduces...

- Double the bandwidth
 - *Zero added latency over CXL 2.0*
- 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
 - Direct memory/ Peer-to-Peer accesses by devices
 - New symmetric memory capabilities
 - Improved software capabilities
- Full backward compatibility with CXL 2.0, CXL 1.1, and CXL 1.0

CXL 3.0 is a huge step function with fabric capabilities while maintaining full backward compatibility with prior generations

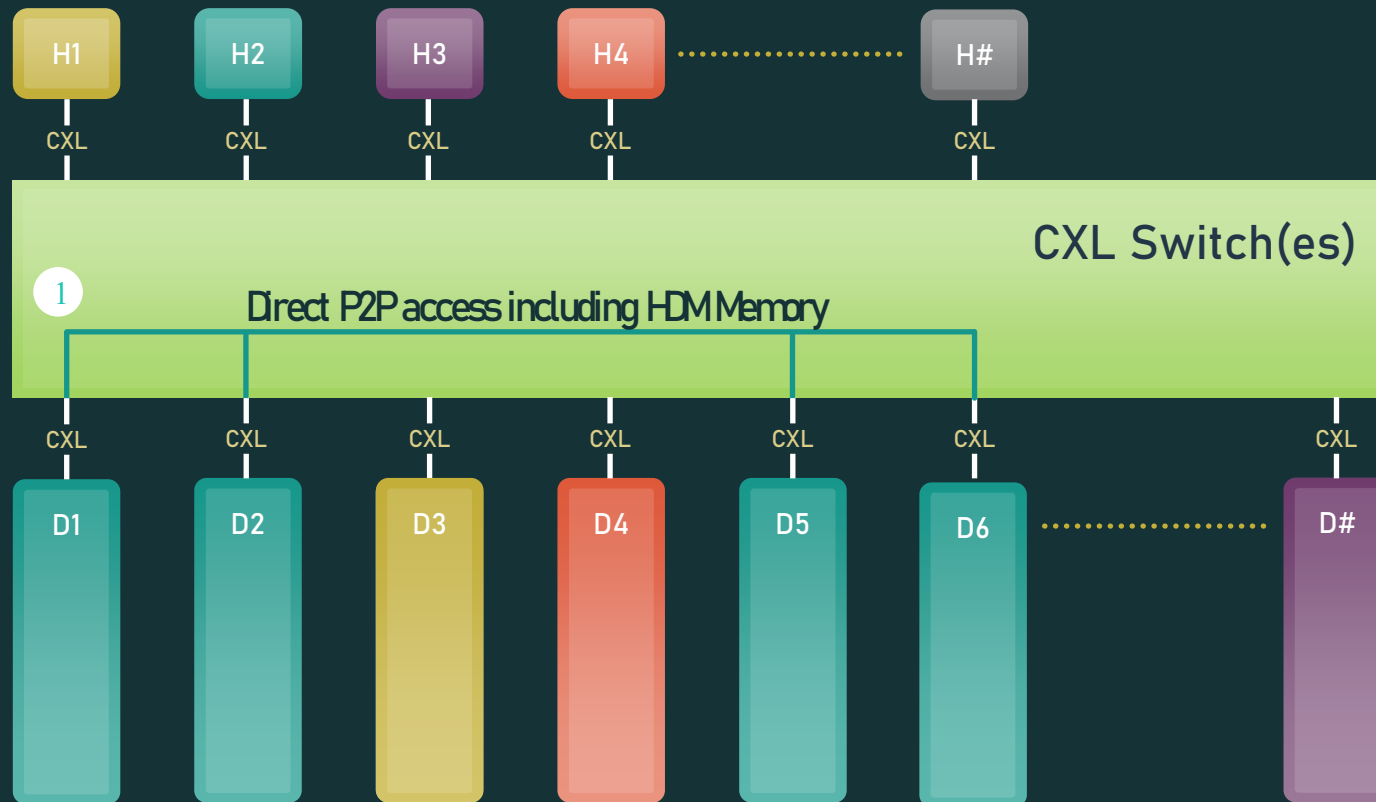
CXL 3.0: Doubles Bandwidth with Same Latency

- 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 (UO and BI) for Device to Memory Connectivity

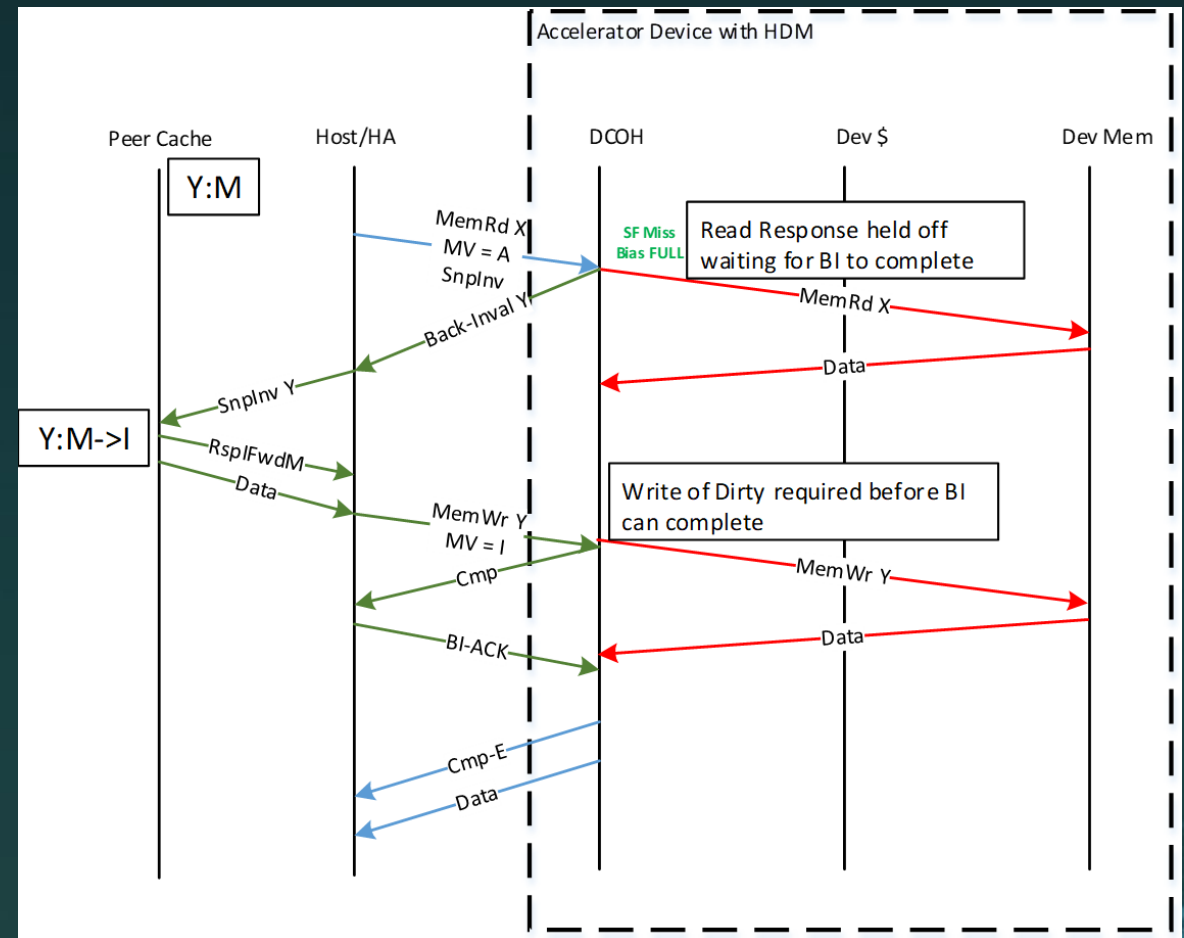
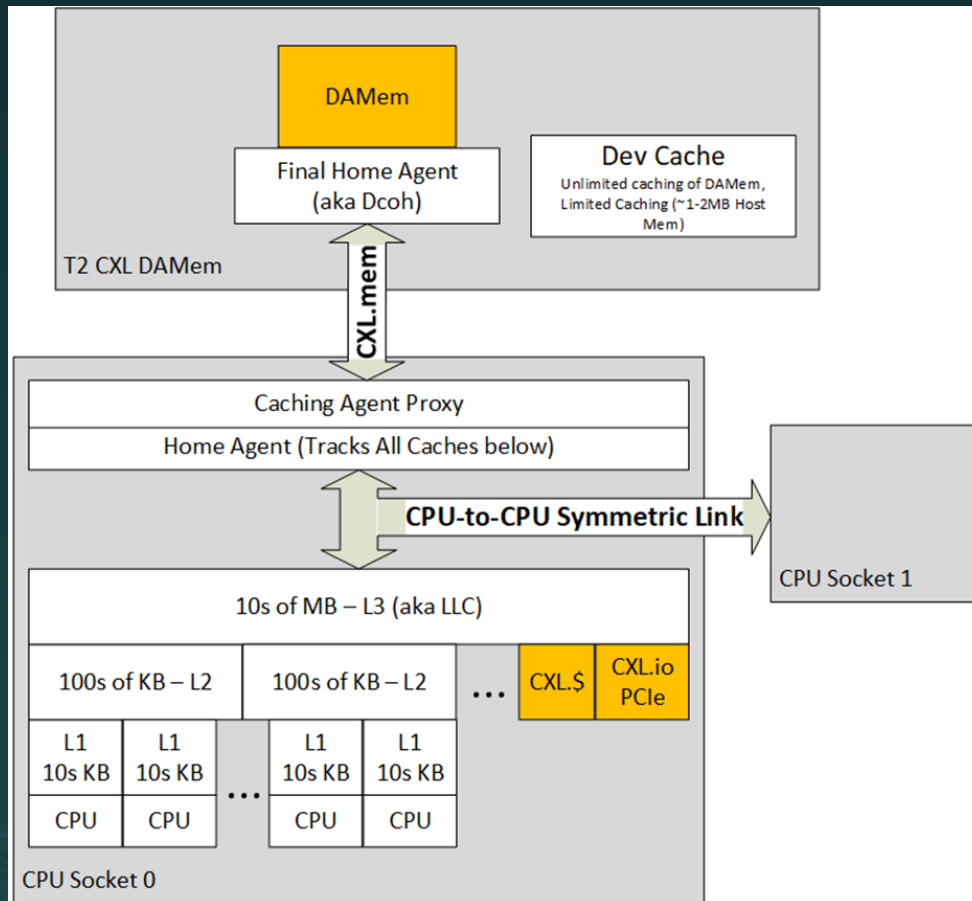


CXL 3.0 enables **non-tree topologies and peer-to-peer communication (P2P)** within a virtual hierarchy of devices

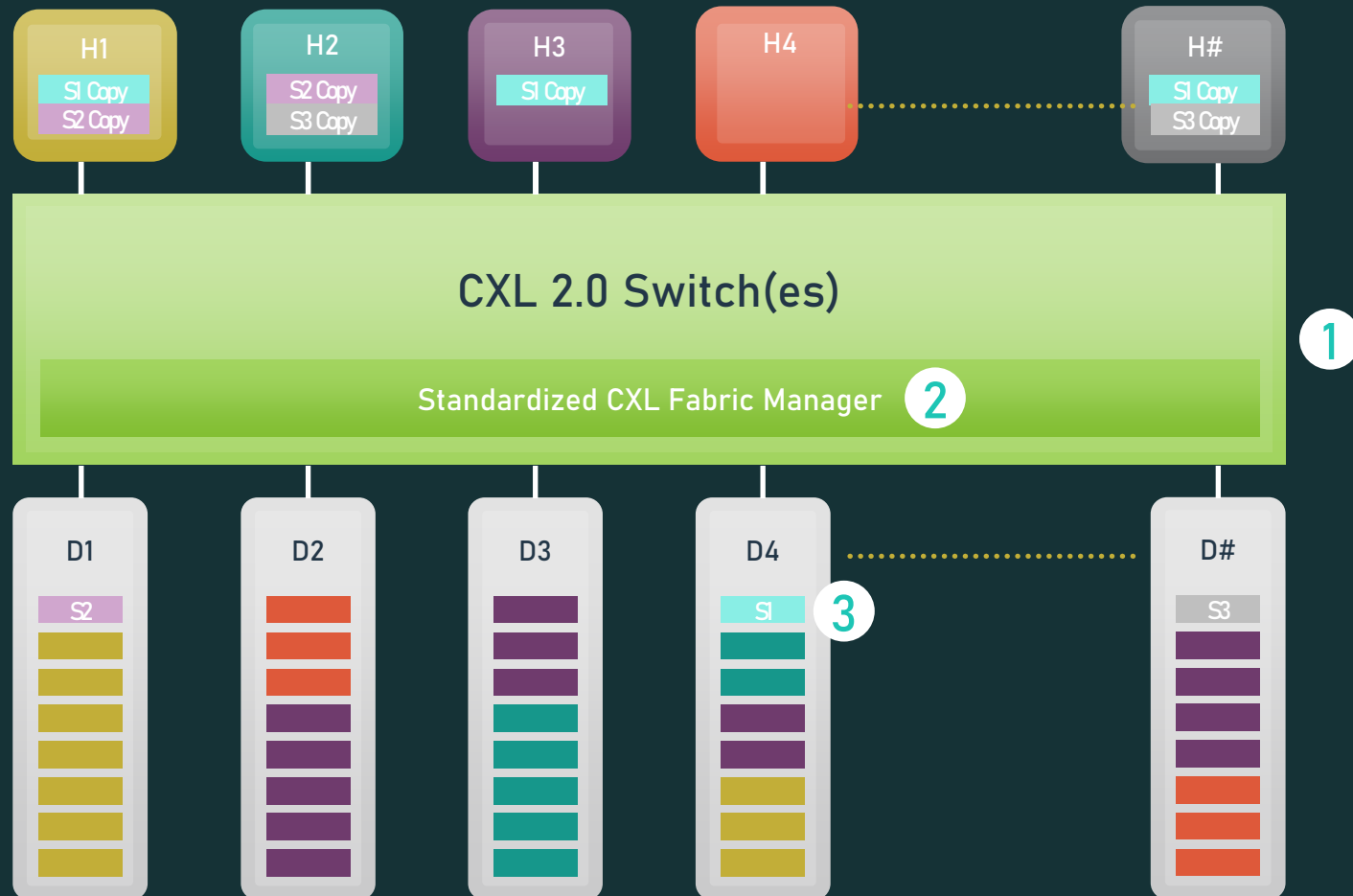
- Virtual hierarchies are associations of devices that maintains a coherency domain
- P2P to HDM-DB memory is I/O Coherent: a new Unordered I/O (UIO) Flow in CXL.io – the Type-2/3 device that hosts the memory will generate a new Back-Invalidation flow (CXL.Mem) to the host to ensure coherency if there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

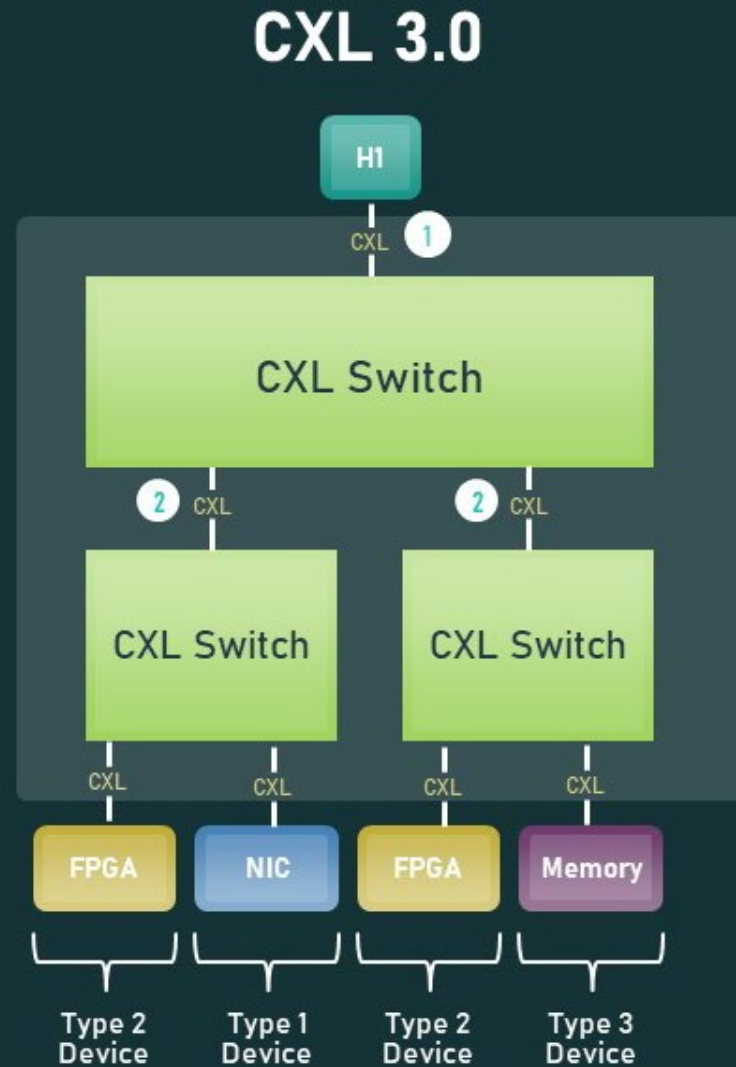
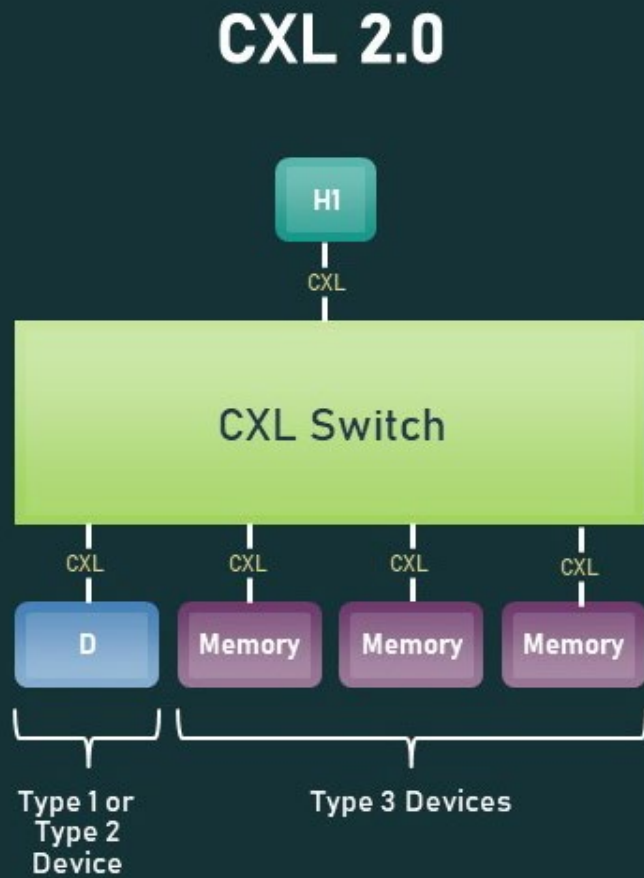


CXL 3.0: Memory Pooling & Sharing



- 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

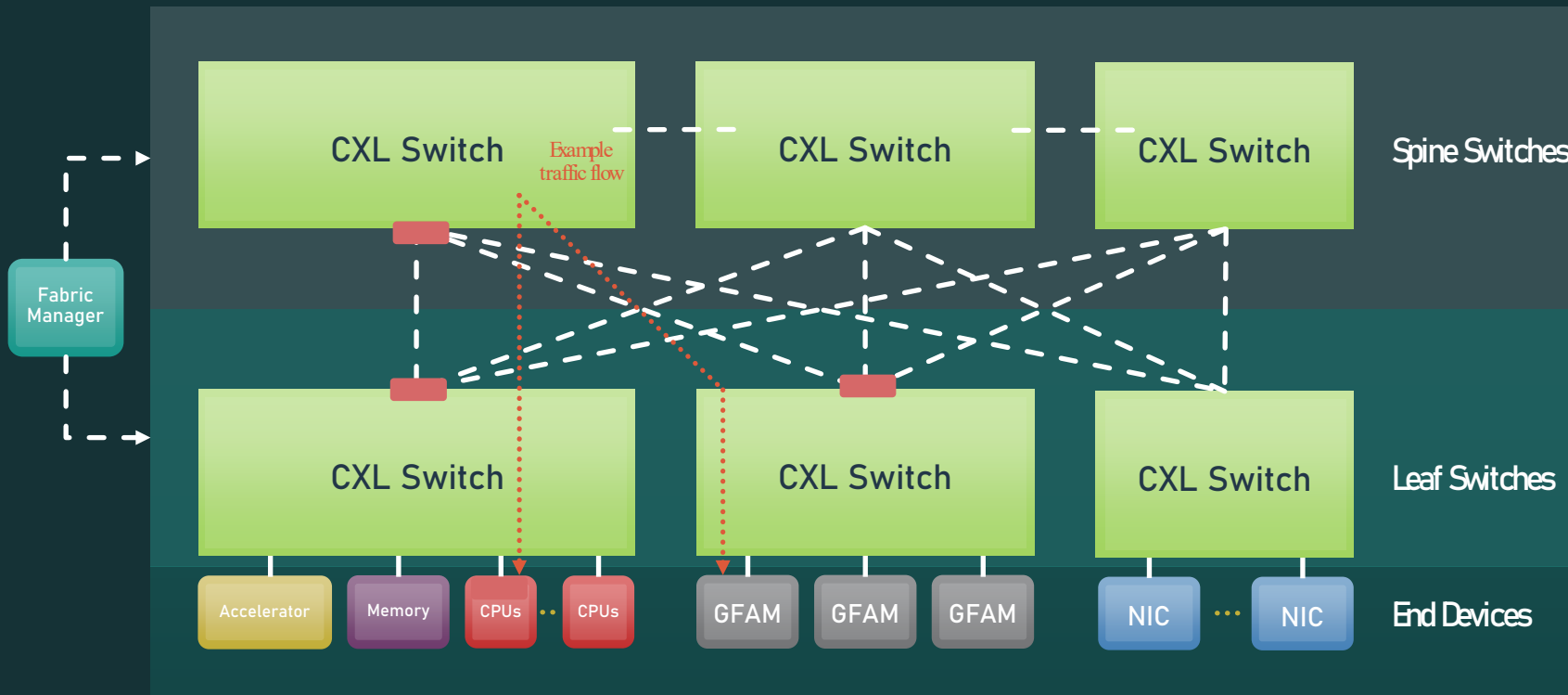
CXL 3.0: Multiple Level Switching, Multiple Type-1/2



- 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 Fabrics

Composable Systems with Spine/Leaf Architecture



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



CXL 3.0 features

- Full fabric capabilities and fabric management
- Expanded switching topologies
- Enhanced 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
- 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!



Thank You

[www. ComputeExpressLink.org](http://www.ComputeExpressLink.org)