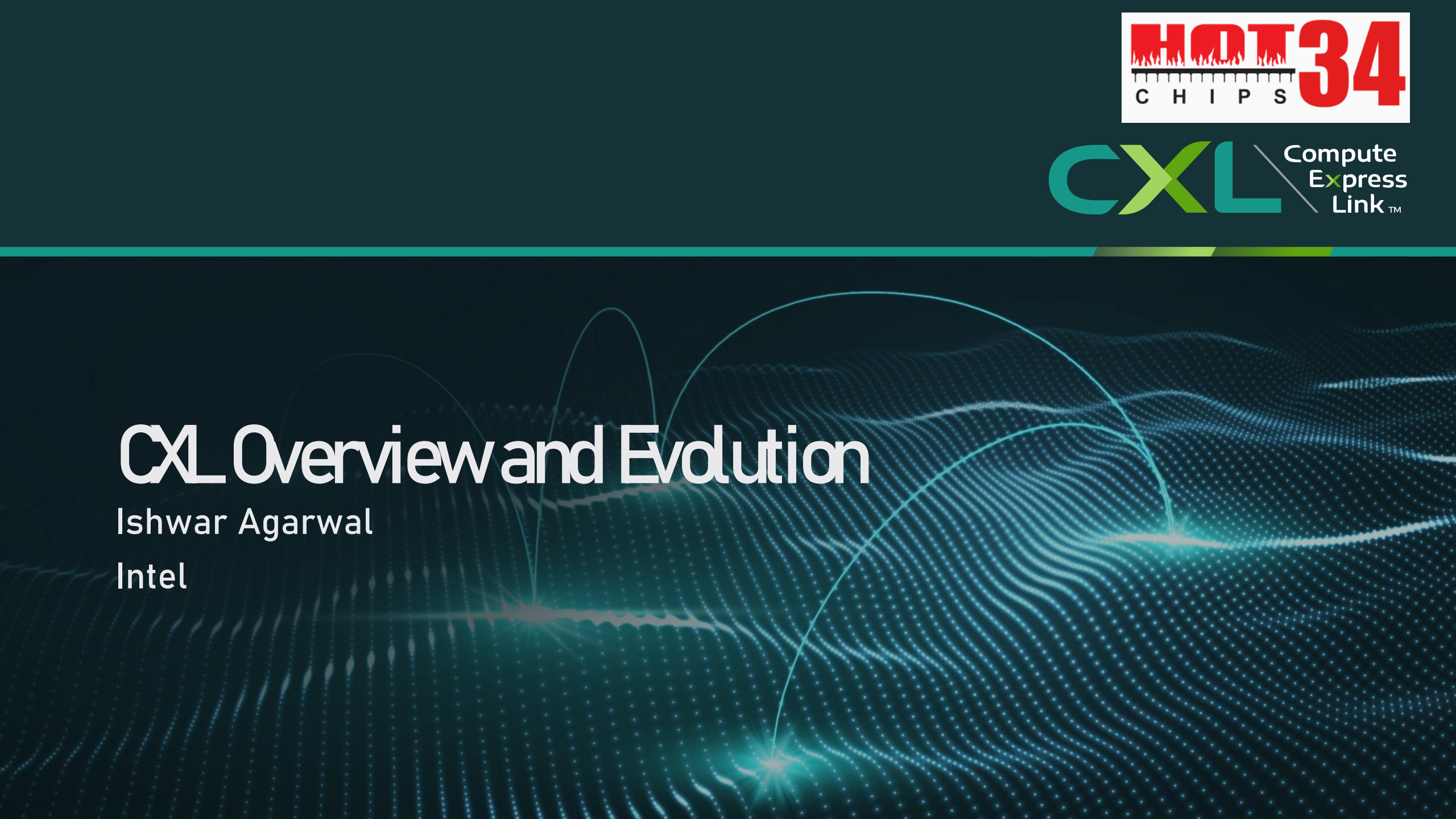


CXL Overview and Evolution

Ishwar Agarwal

Intel

Alibaba Group
AMD
arm
cisco
DELL EMC
Google
Hewlett Packard Enterprise
HUAWEI
IBM
intel
Meta
Microsoft
Micron
NVIDIA
SAMSUNG

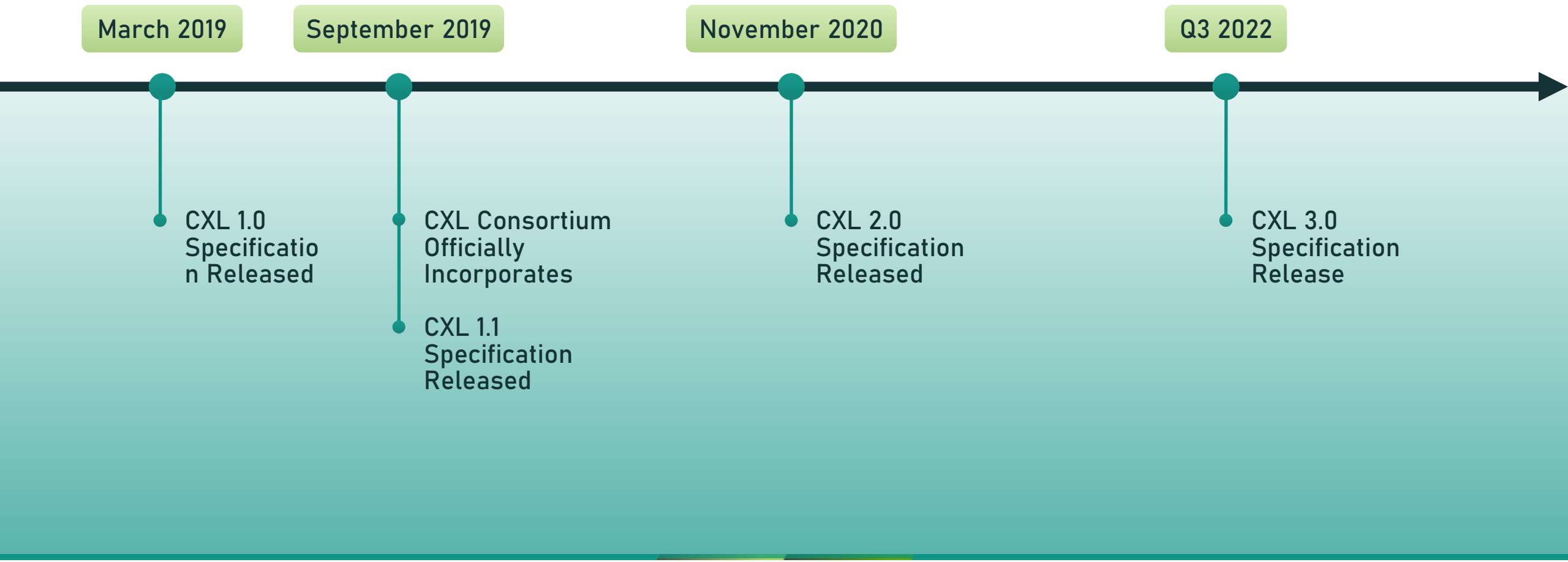
CXL Board of Directors

CXL Compute Express Link™

Industry Open Standard for High Speed Communications

180+ Member Companies

CXL Specification Release Timeline

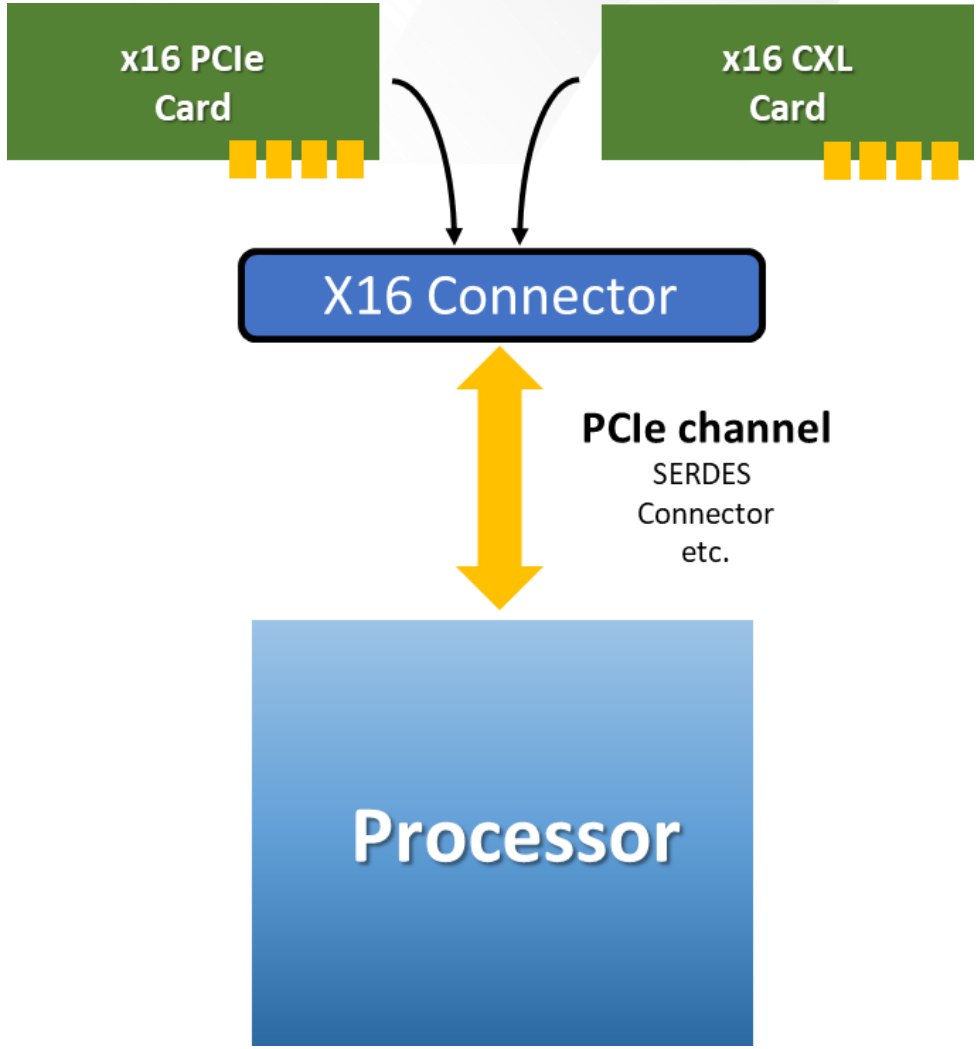


CXL Overview

- New breakthrough high-speed interconnect
  - Enables a high-speed, efficient interconnect between CPU, memory and accelerators
  - Builds upon PCI Express® 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 with heterogeneous processing
    - Enables memory disaggregation, memory pooling and sharing, persistent memory and emerging memory media
- Delivered as an open industry standard
  - CXL Specification 3.0 is available now with full backward compatibility with CXL 2.0 and CXL 1.1
  - Future CXL Specification generations will continue to innovate to meet industry needs with backward compatibility

What is CXL?

- Alternate protocol that runs across the standard PCIe physical layer
- Uses a flexible processor port that can auto-negotiate to either the standard PCIe transaction protocol or the alternate CXL transaction protocols
- CXL 2.0 and CXL 1.1 align to 32 GT/s PCIe 5.0
- CXL 3.0 aligns to 64GT/s PCIe 6.0 and is backward compatible



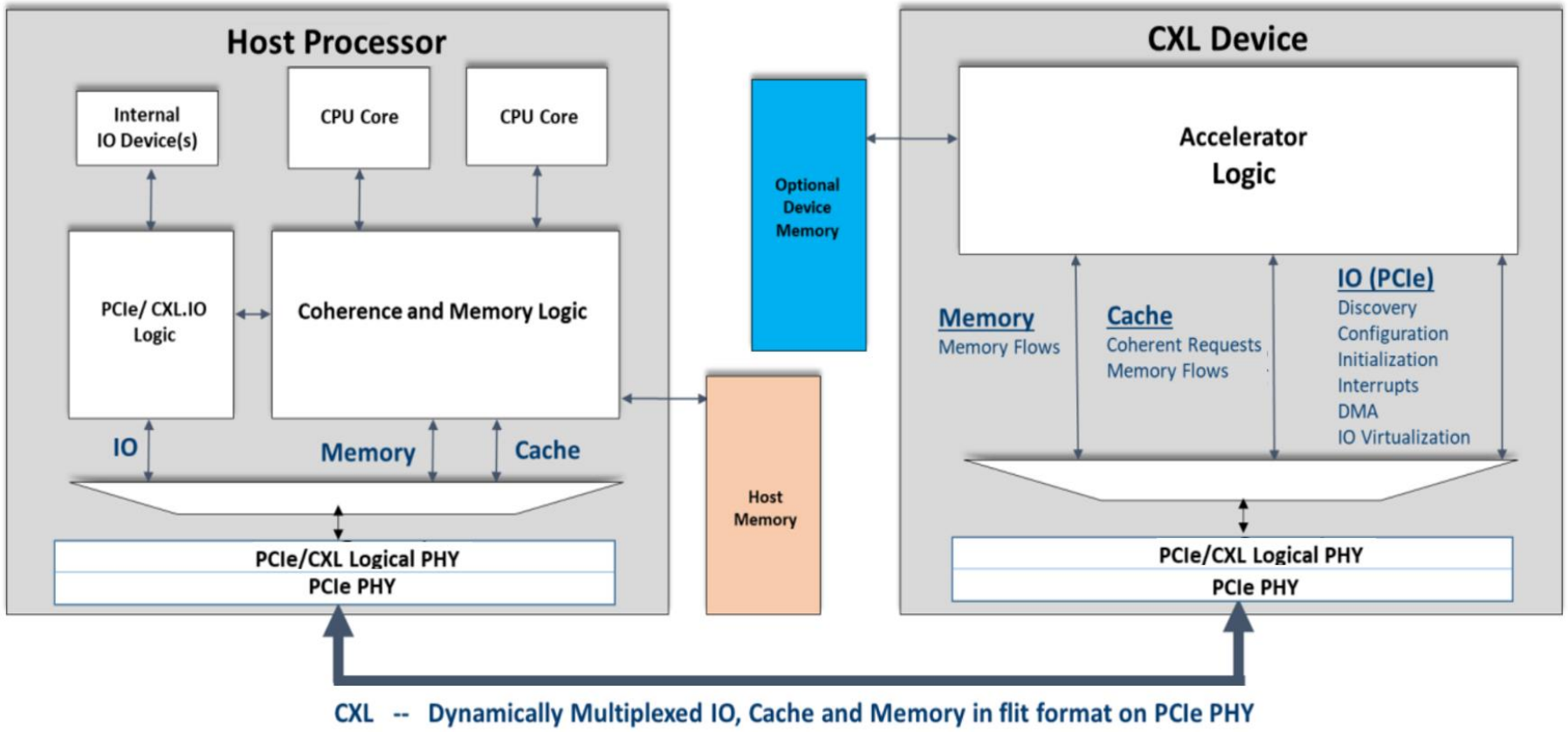
CXL Protocols

- The CXL transaction layer is compromised of three dynamically multiplexed sub-protocols on a single link:

CXL.io
Discovery, configuration, register access, interrupts, etc.

CXL.cache
Device access to processor memory

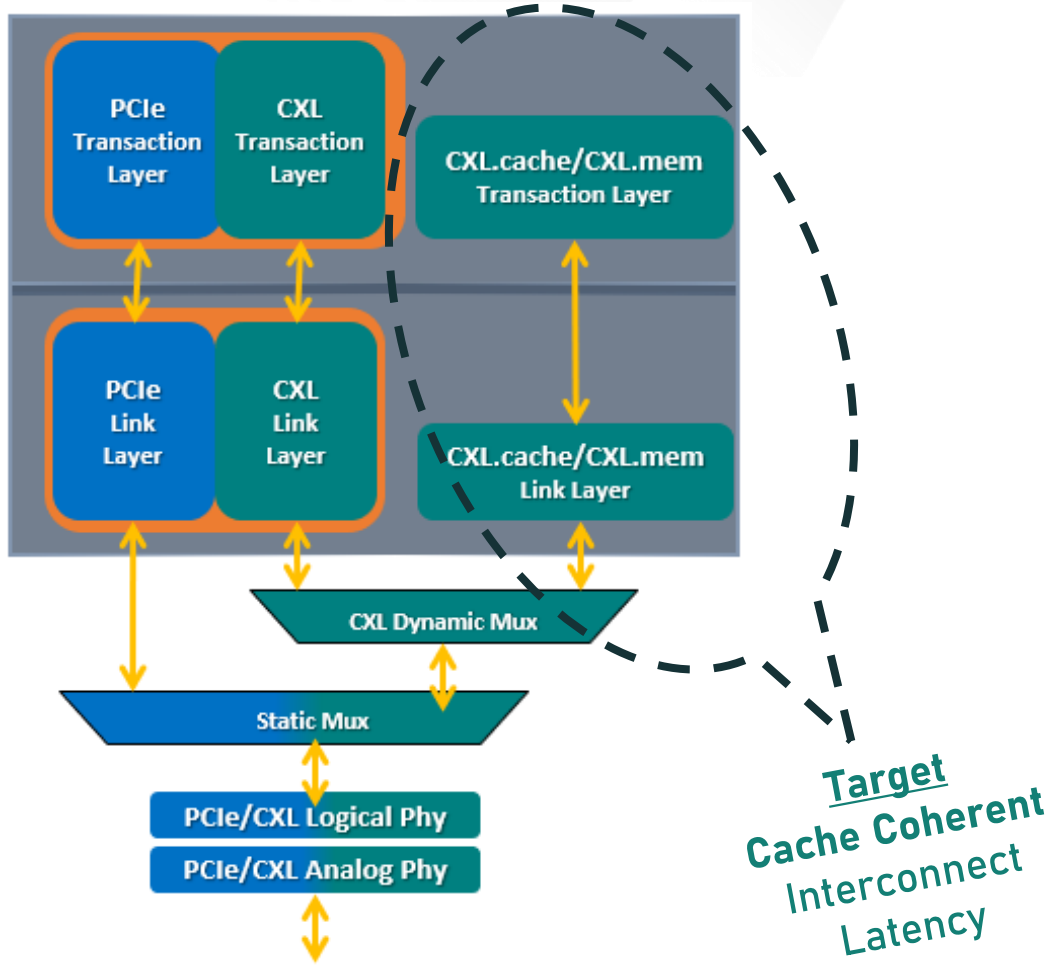
CXL.Memory
Processor access to device attached memory



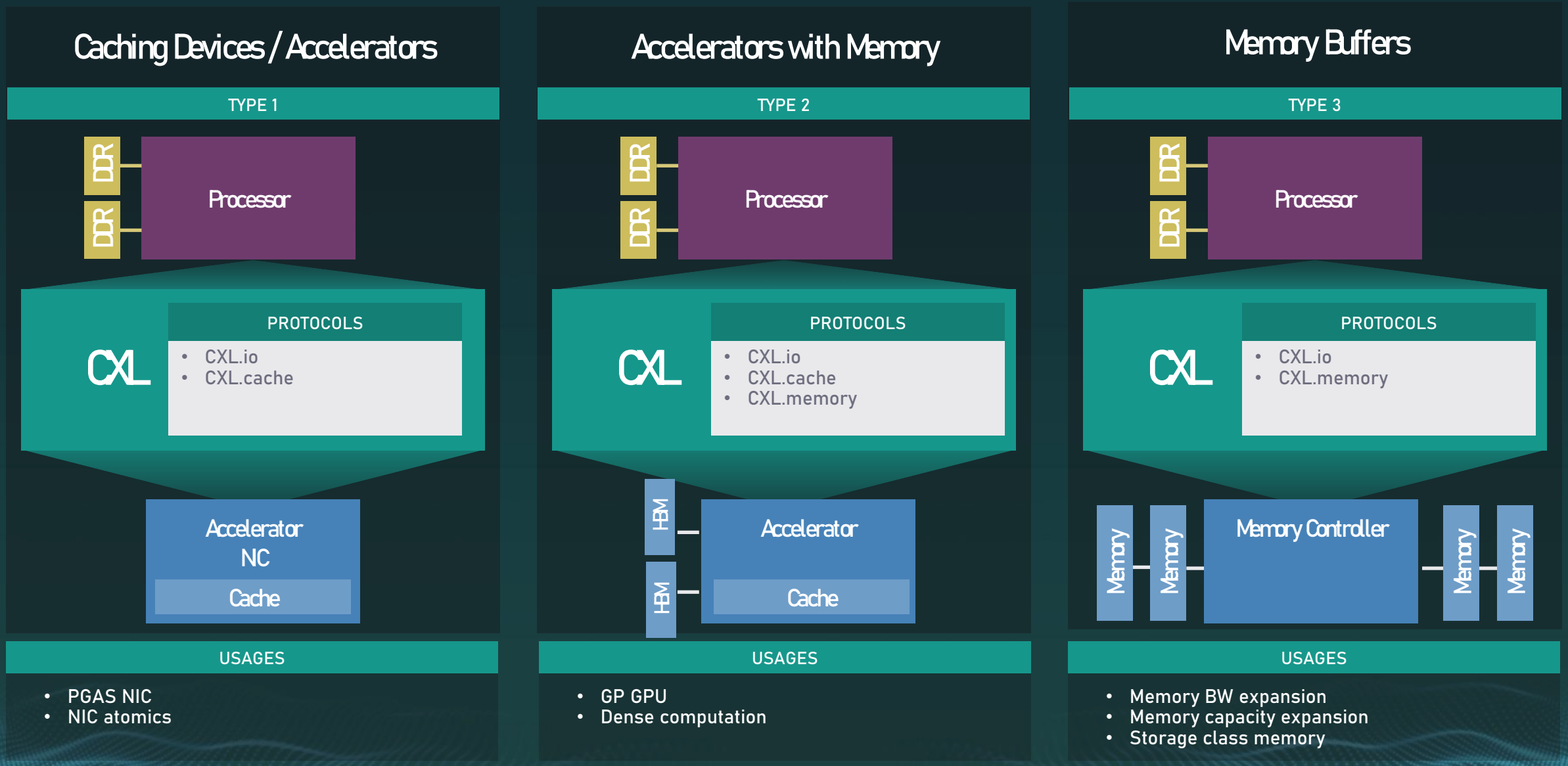
CXL Stack
Designed for Low Latency

- All 3 representative usages have latency critical elements:
  - CXL.cache
  - CXL.memory
  - CXL.io
- CXL cache and memory stack is optimized for latency:
  - Separate transaction and link layer from IO
  - Fixed message framing
- CXL io flows pass through a stack that is largely identical a standard PCIe stack:
  - Dynamic framing
  - Transaction Layer Packet (TLP)/Data Link Layer Packet (DLLP) encapsulated in CXL flits

CXL Stack –
Low latency Cache and Mem Transactions



Representative CXL Usages



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 Spec Feature Summary

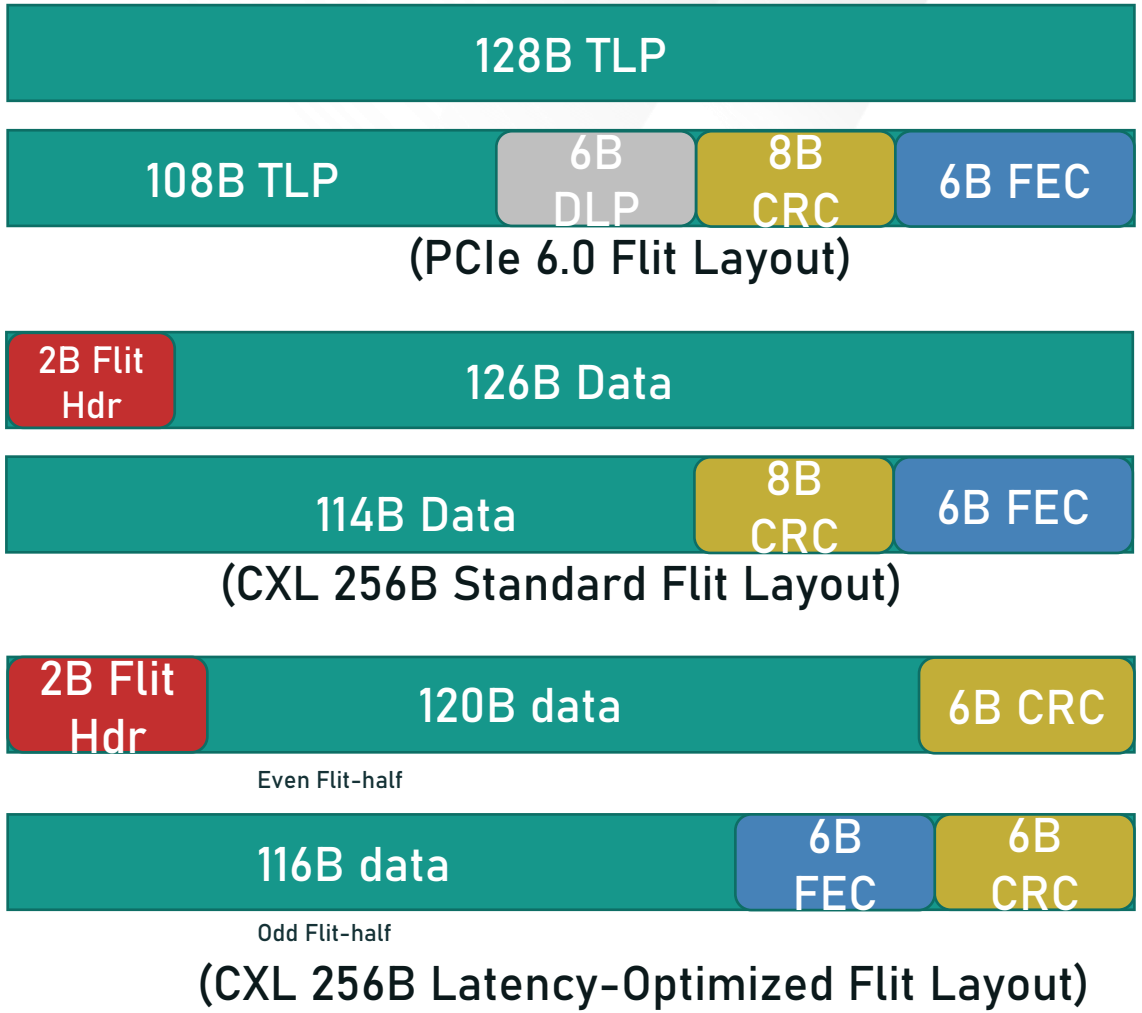
Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0
Release date	2019	2020	2022
Max link rate	32GTs	32GTs	64GTs
Flit 68 byte (up to 32 GTs)	✓	✓	✓
Flit 256 byte (up to 64 GTs)			✓
Type 1, Type 2 and Type 3 Devices	✓	✓	✓
Memory Pooling w/ MLDs		✓	✓
Global Persistent Flush		✓	✓
CXL IDE		✓	✓
Switching (Single-level)		✓	✓
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			✓
Fabrics (256 byte flit)			✓

Not supported

✓ Supported

CXL 3.0: Doubles bandwidth with same latency

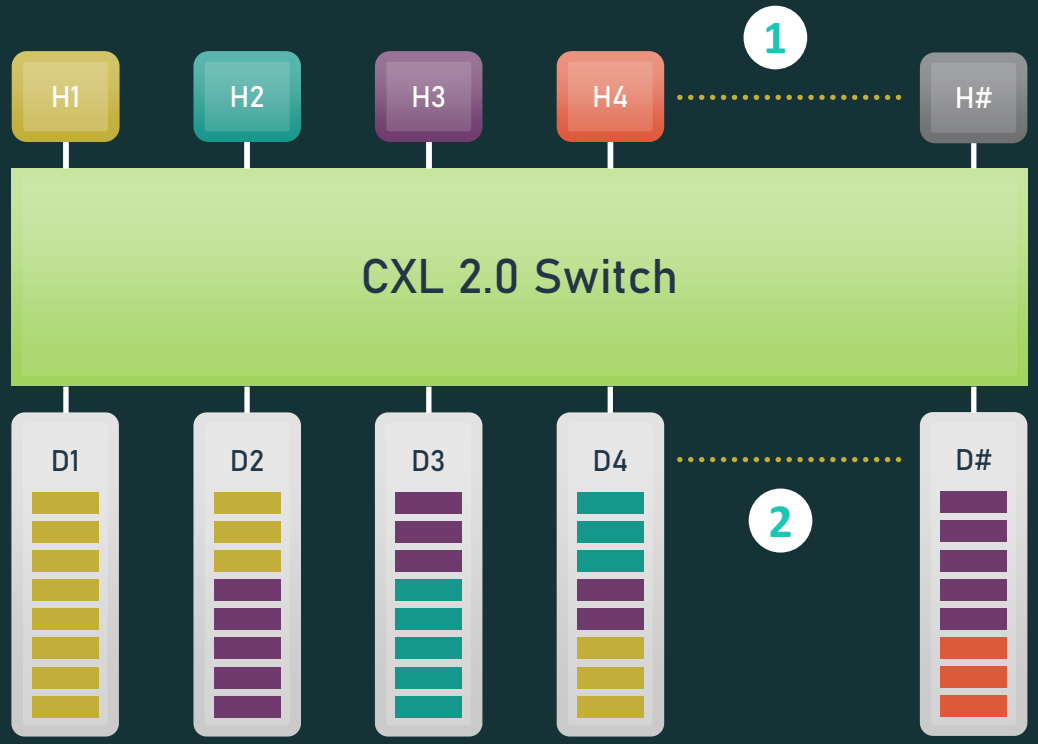
- Uses PCIe 6.0® PHY @ 64 GT/s
- PAM-4 and high BER mitigated by PCIe 6.0 FEC and CRC (different CRC for latency optimized)
- Standard 256B Flit along with an additional 256B Latency Optimized Flit (0-latency adder over CXL 2)
  - 0-latency adder trades off FIT (failure in time, 10^9 hours) from $5x10^{-8}$ 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)

RECAP: CXL 2.0 FEATURE SUMMARY

MEMORY POOLING

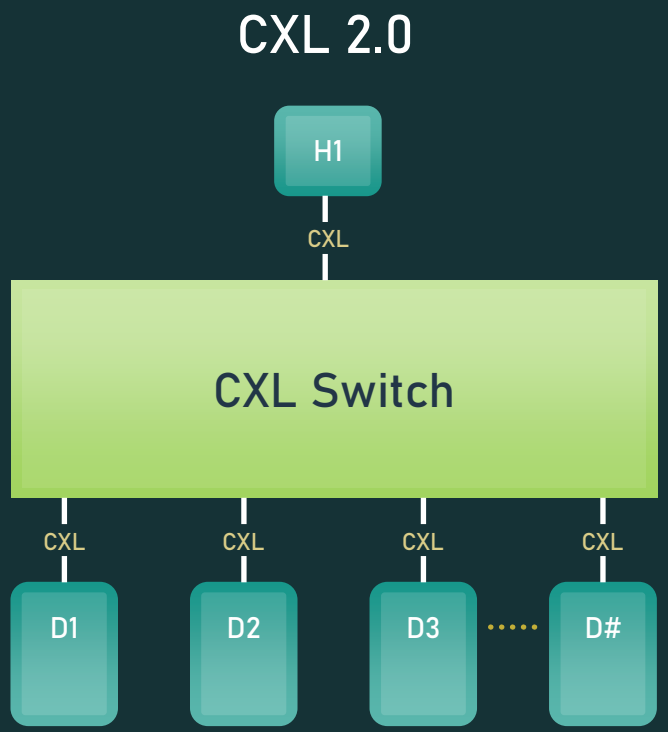


1. Device memory can be allocated across multiple hosts.

2. Multi Logical Devices allow for finer grain memory allocation

RECAP: CXL 2.0 FEATURE SUMMARY

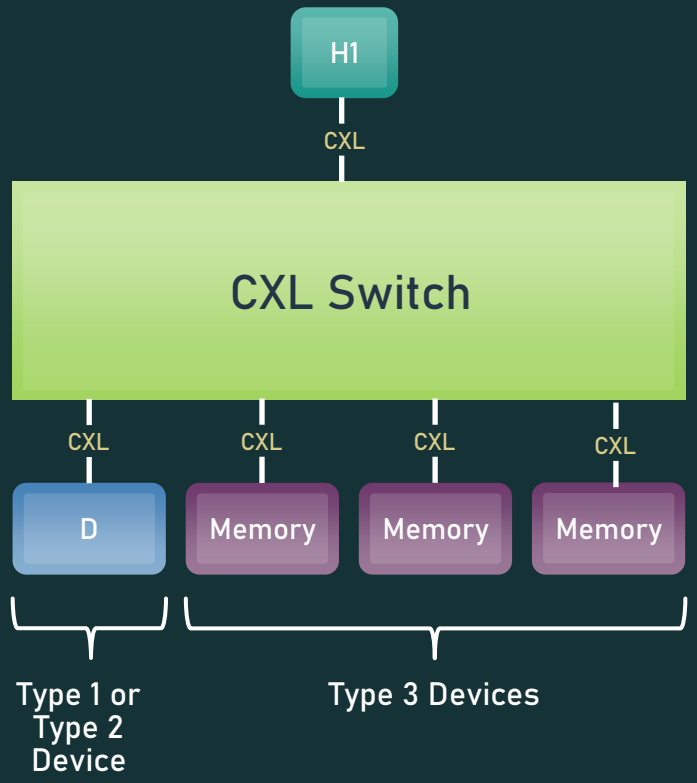
SWITCH CAPABILITY



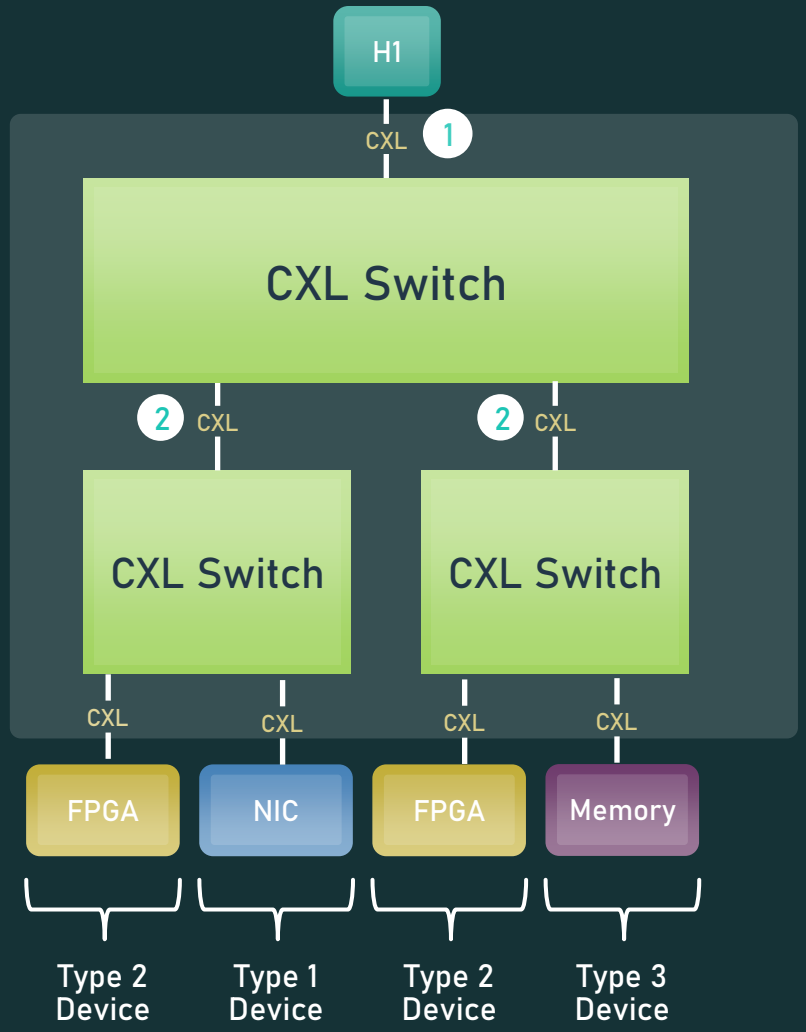
- Supports single-level switching
- Enables memory expansion and resource allocation

CXL 3.0: MULTIPLE LEVEL SWITCHING, MULTIPLE TYPE-1/2 Devices

CXL 2.0



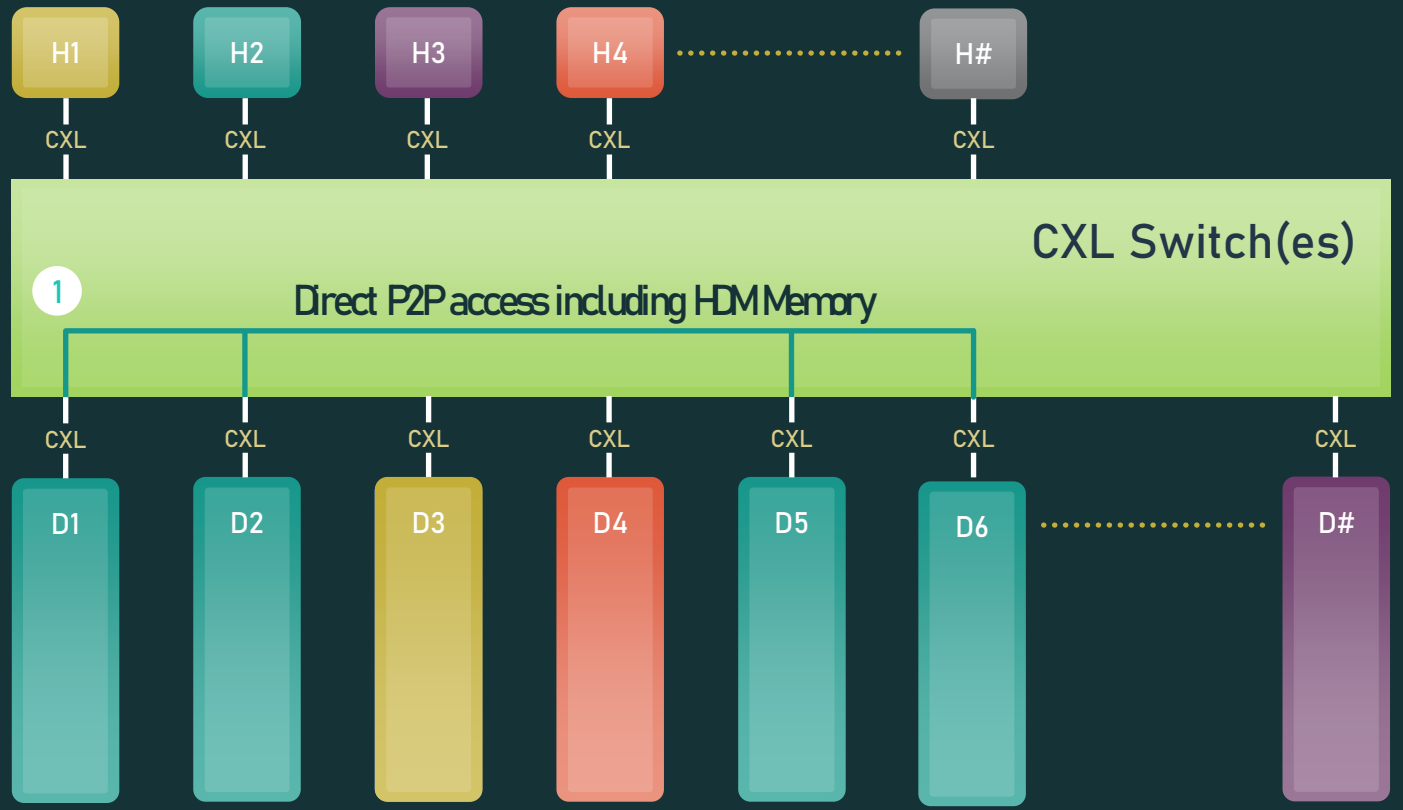
CXL 3.0



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

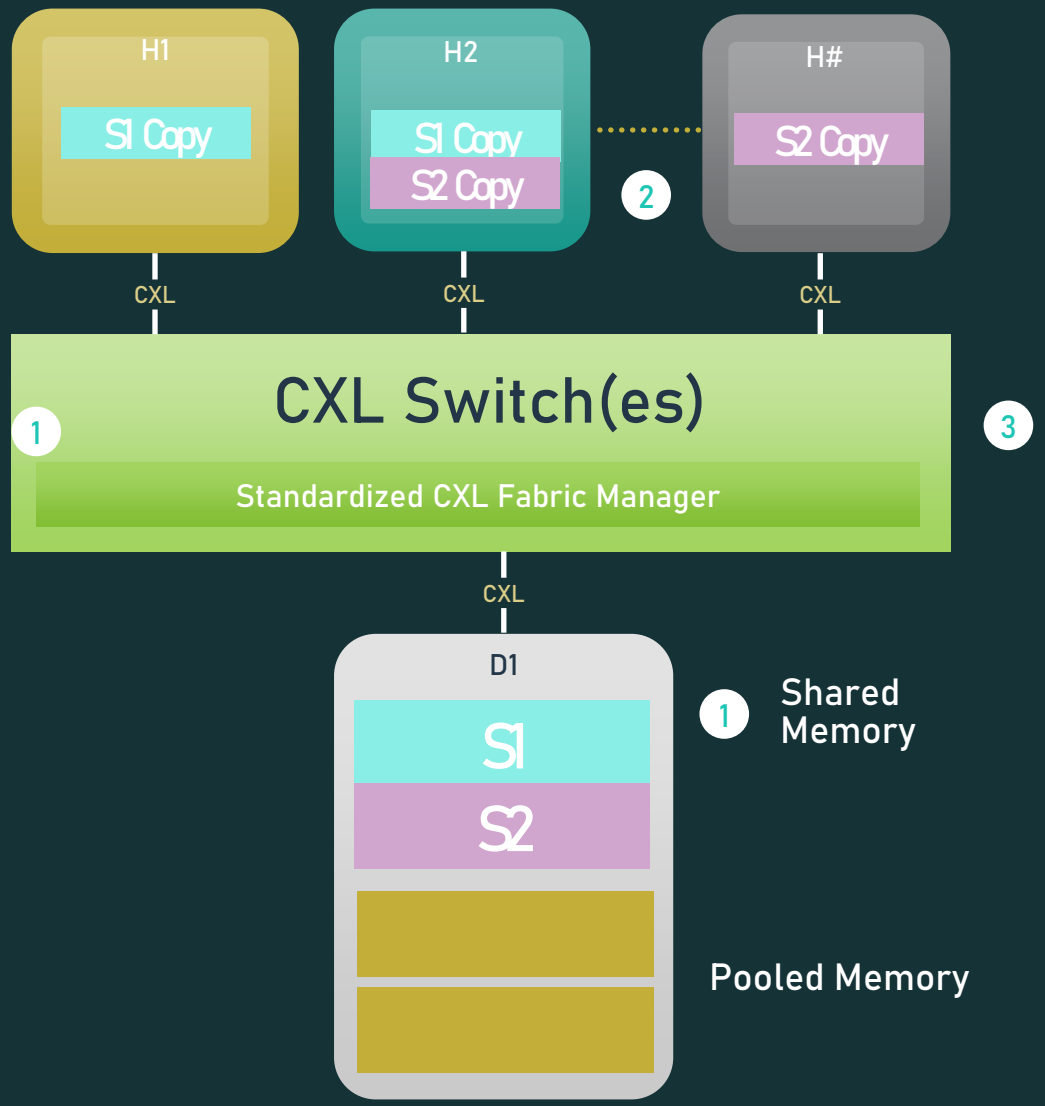
2. Multiple switch levels (aka cascade)
   - Supports fanout of all device types

CXL 3.0 PROTOCOL ENHANCEMENTS (UIO and BI) for DEVICE TO DEVICE CONNECTIVITY



1. 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: COHERENT MEMORY SHARING

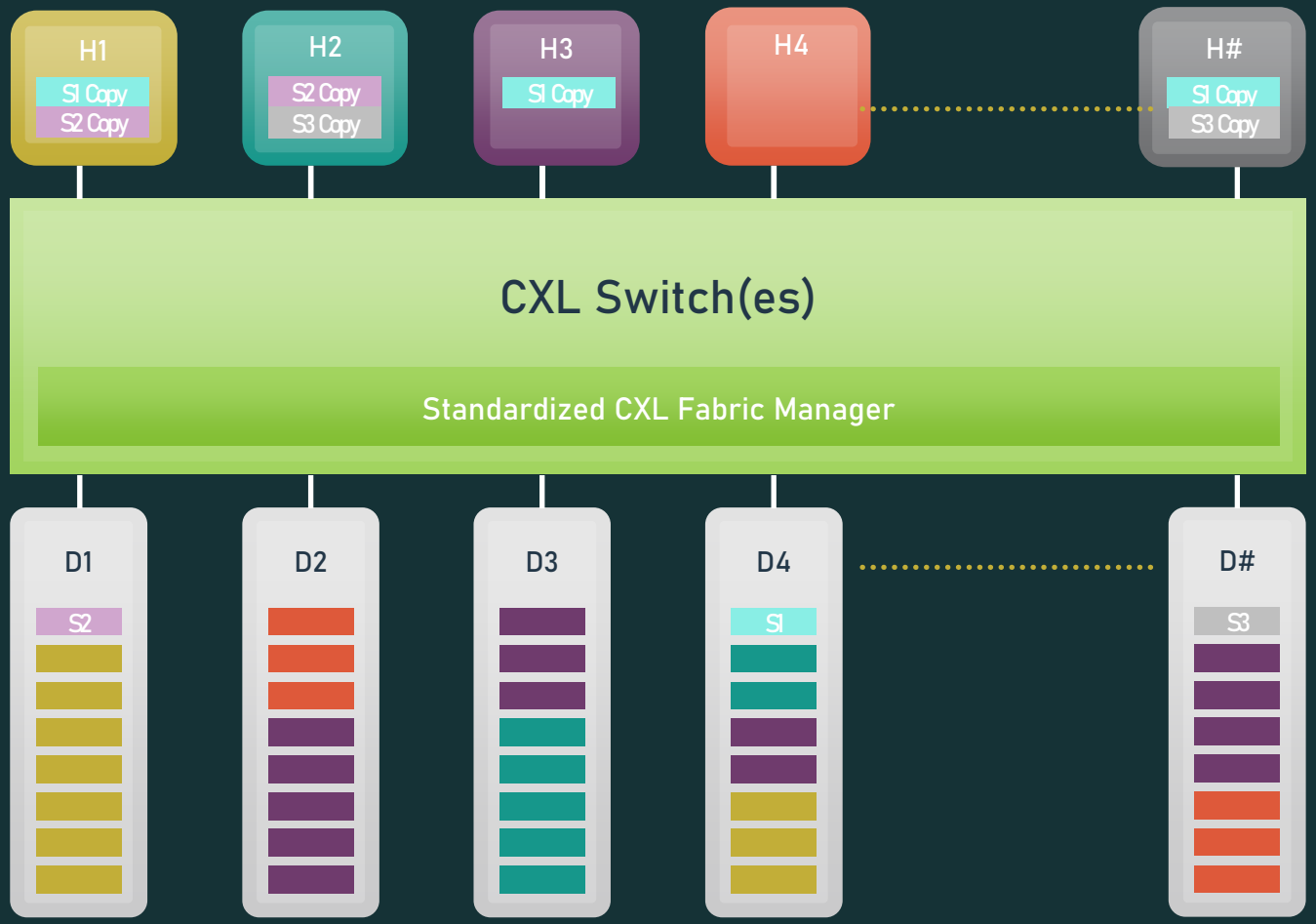


1. Device memory can be shared by all hosts to increase data flow efficiency and improve memory utilization

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

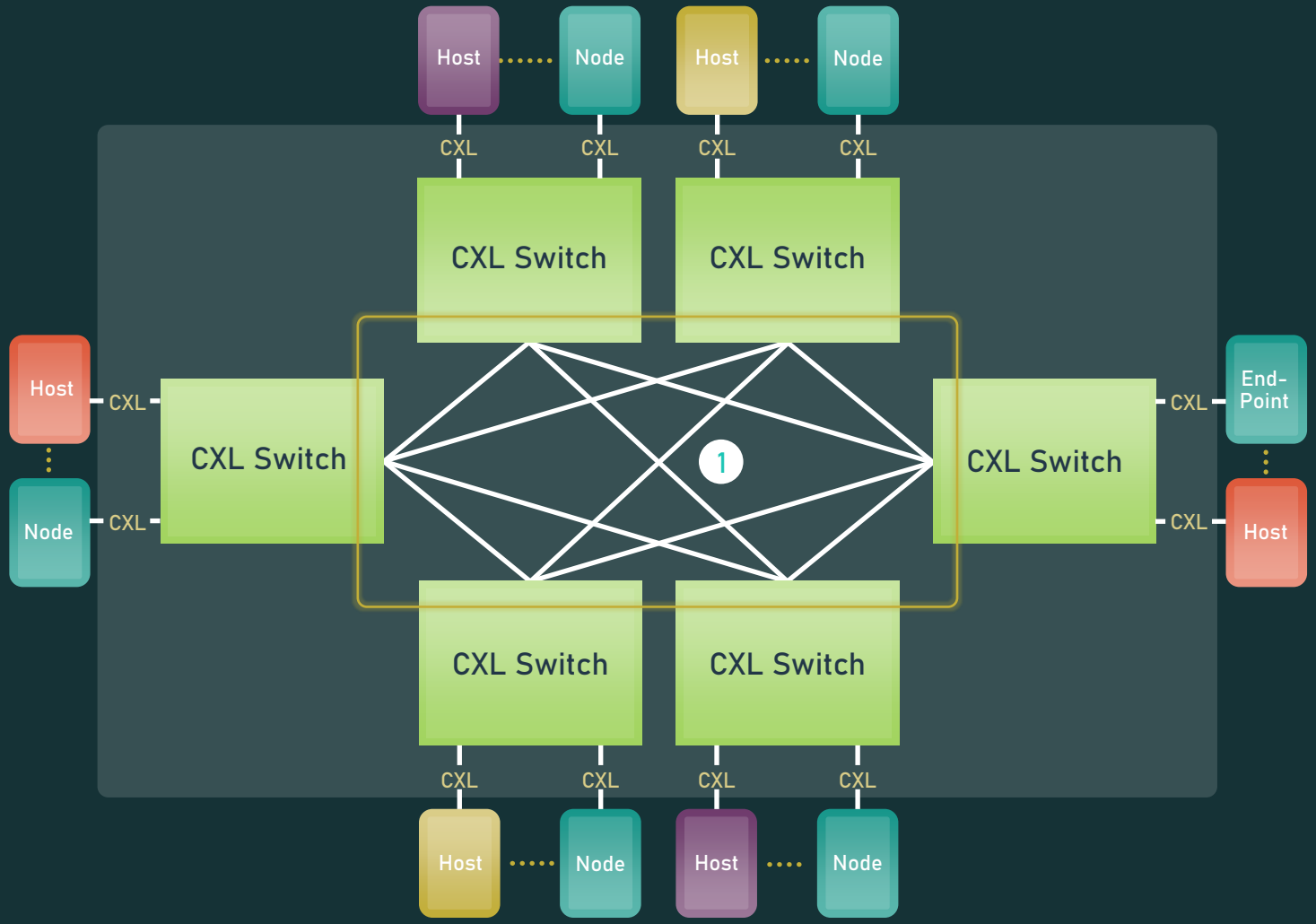
3. CXL 3.0 defined mechanisms to enforce hardware cache coherency between copies

CXL 3.0: 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

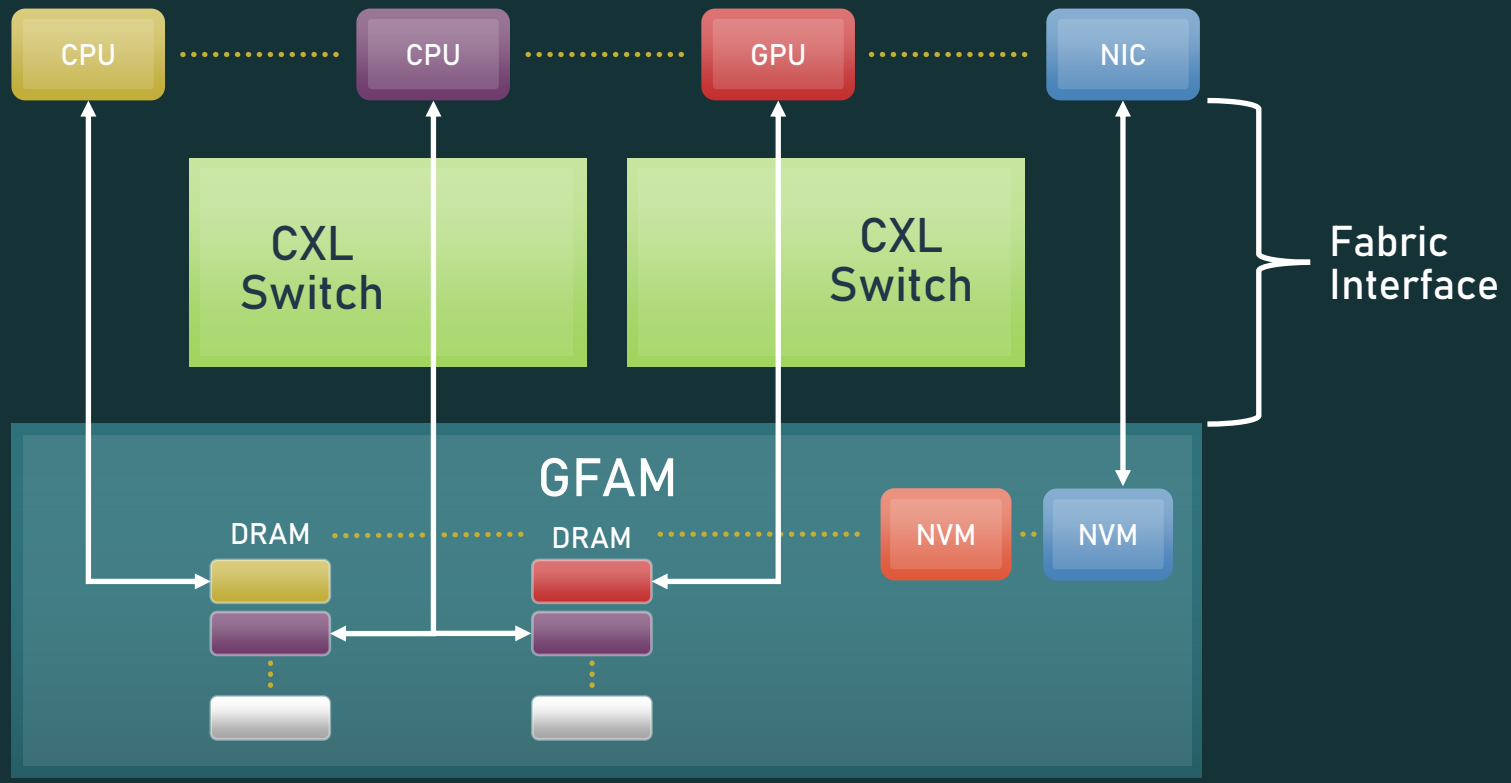
FABRICS Overview



1 Nodes can be any combination:

- Hosts
- Type 1 – Device with cache
  - Example: Smart NIC
- Type 2 – Device with cache and memory
  - Example: AI Accelerator
- Type 3 – Device with memory
  - Example: memory expander
- PCIe Device

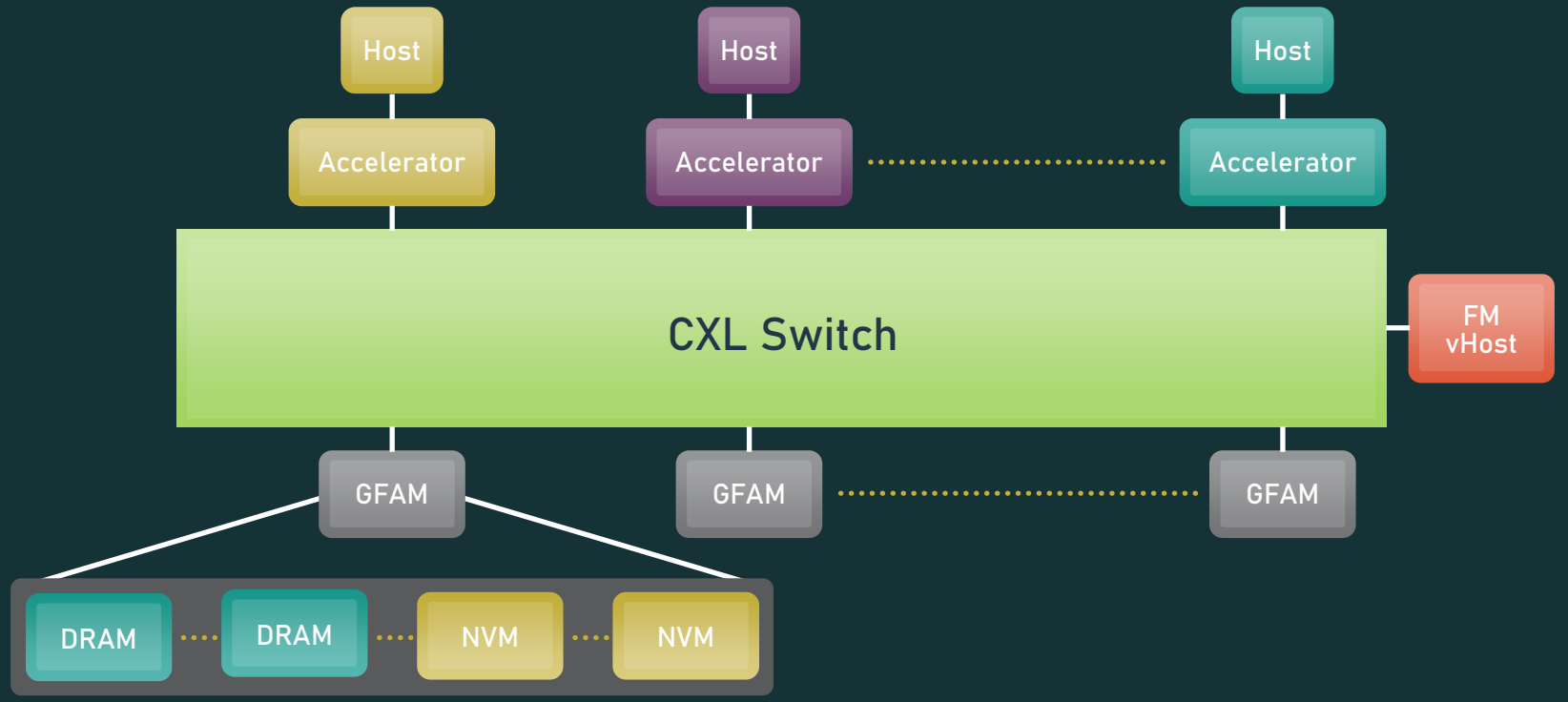
CXL 3.0: GLOBAL FABRIC ATTACHED MEMORY (GFAM) DEVICE



- CXL 3.0 enables Global Fabric Attached Memory (GFAM) architecture which differs from traditional processor centric architecture by disaggregating the memory from the processing unit and implements a shared large memory pool
- Memory can be of the same type or different types which can be accessed by multiple processors directly connected to GFAM or through a CXL switch

CXL 3.0: FABRICS EXAMPLE USE CASE

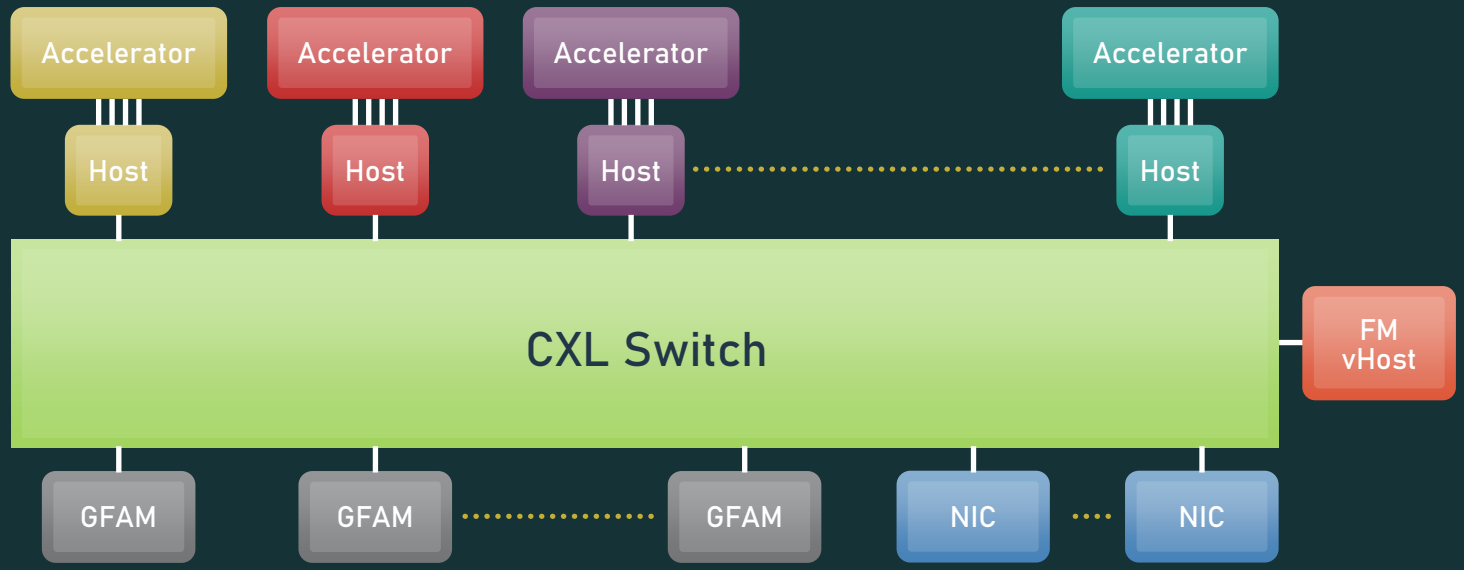
Machine Learning Accelerator and GFAM Device in a Fabric Architecture



GFAM enables multiple media types, i.e. DRAM, Flash, future memory types

CXL 3.0: FABRICS EXAMPLE USE CASE

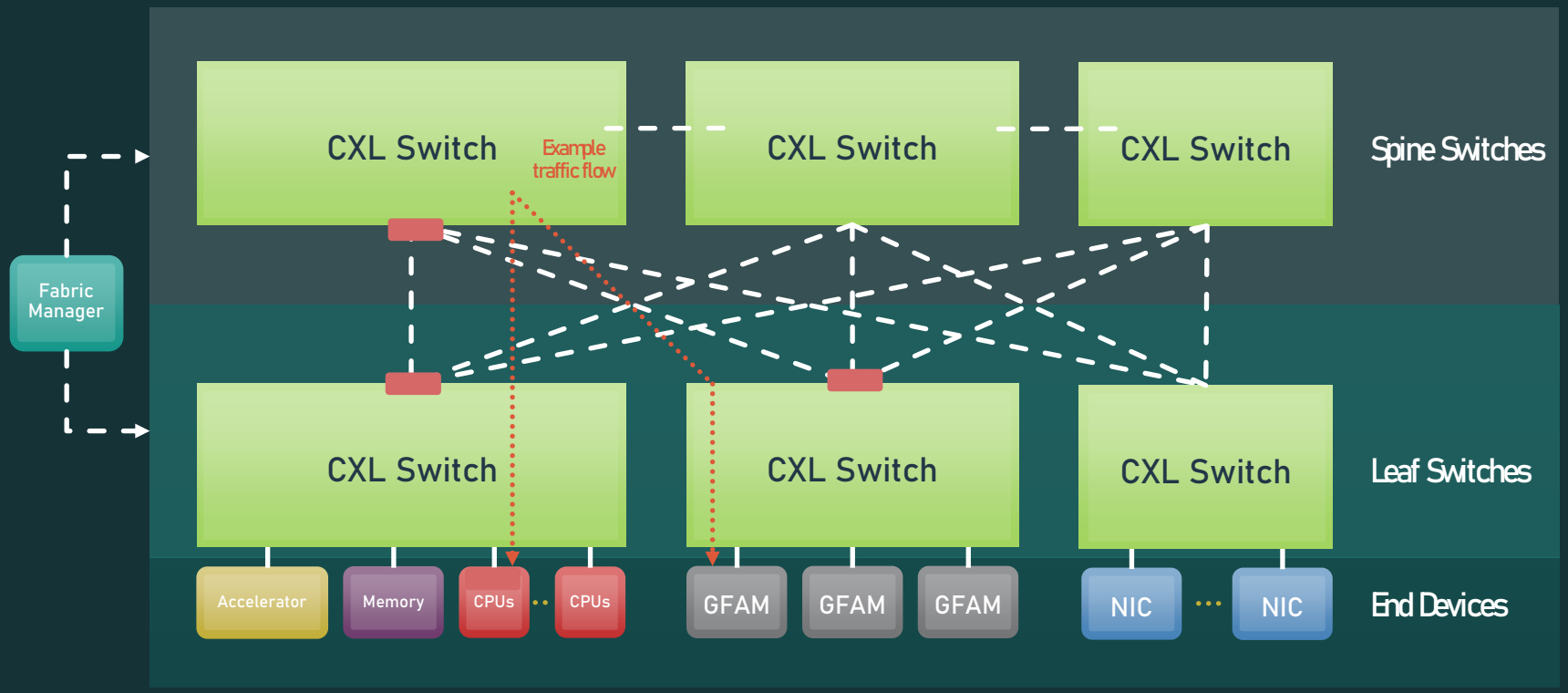
HPC/Analytics



Sharing memory and networking devices to reduce cost and improve efficiency

CXL 3.0: FABRICS EXAMPLE USE CASE

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 Summary

- CXL 3.0 features
  - Enhanced memory pooling and enables new memory usage models
  - Multi-level switching with multiple host and fabric capabilities and enhanced fabric management
  - New symmetric coherency capabilities
  - Improved software capabilities
- CXL 3.0 introduces new usage models
  - Delivers industry needs for higher bandwidth
  - Optimized system level flows with advanced switching, efficient peer-to-peer and fine-grained resource sharing across multiple domains
- Call to Action
  - Join CXL Consortium
  - Follow us on Twitter and LinkedIn for updates!

Thank You