# Compute Express Link™ (CXL™): CXL 2.0 ECN

Ishwar Agarwal – CXL Technical Task Force Co-chair

Mahesh Natu – CXL Systems and Software WG Co-chair

December 9th, 2021

# Industry Landscape

Proliferation of
Cloud Computing

Growth of
AI & Analytics

Cloudification of
the Network & Edge


CXL Board of Directors


Industry Open Standard for High Speed Communications

170+ Member Companies

# 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

## CXL

An open industry-supported cache-coherent interconnect for processors, memory expansion and accelerators

## Coherent Interface

Leverages PCIe® with 3 mix-and-match protocols

## Low Latency

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

## Asymmetric Complexity

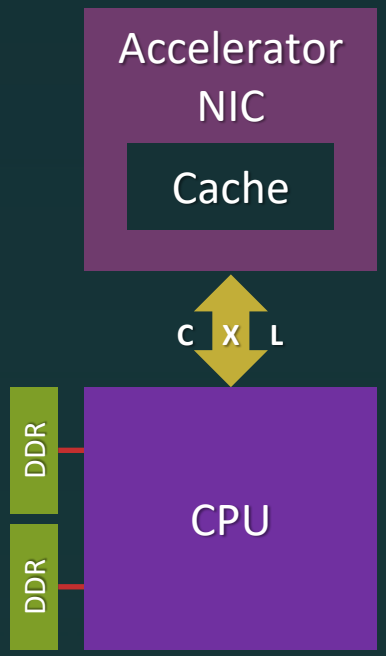
Eases burdens of cache coherent interface designs

# CXL 2.0 Usage Models – Recap

## Caching Devices / Accelerators

Usages:
- PGAS NIC
- NIC atomics

Protocols:
- CXL.io
- CXL.cache

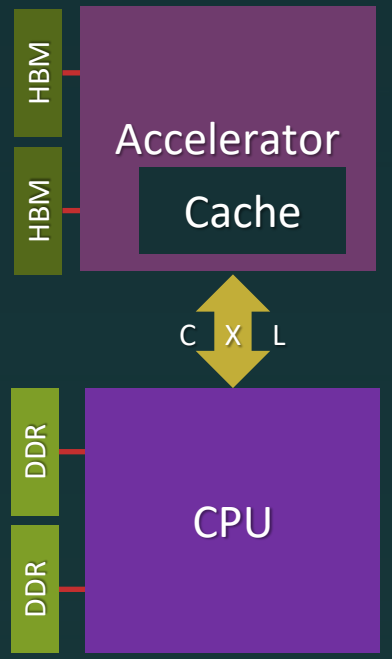



(Type 1 Device)

## Accelerators with Memory

Usages:
- GPU
- FPGA
- Dense Computation

Protocols:
- CXL.io
- CXL.cache
- CXL.memory

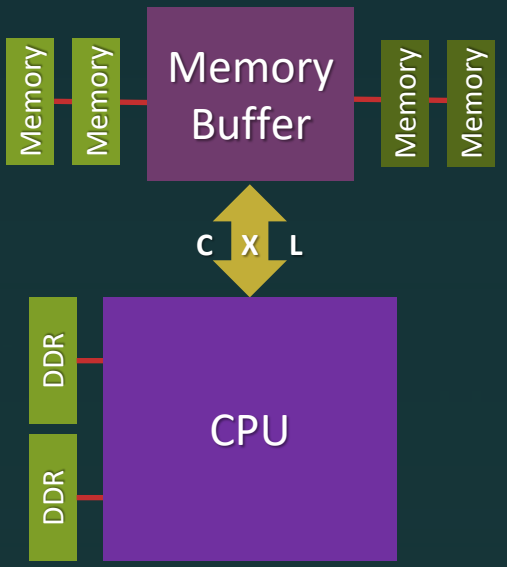



(Type 2 Device)

## Memory Buffers

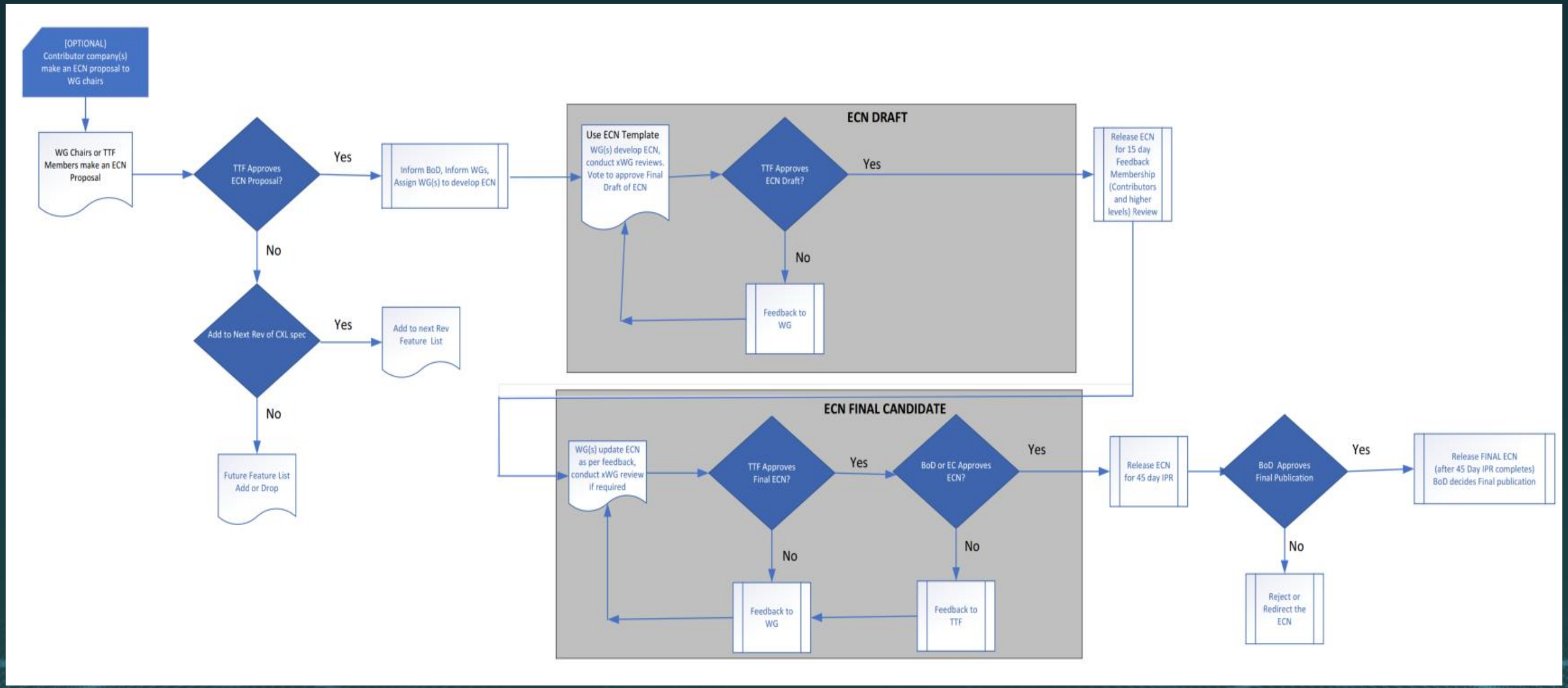
Usages:
- Memory BW expansion
- Memory capacity expansion
- Persistent Memory

Protocols:
- CXL.io
- CXL.mem



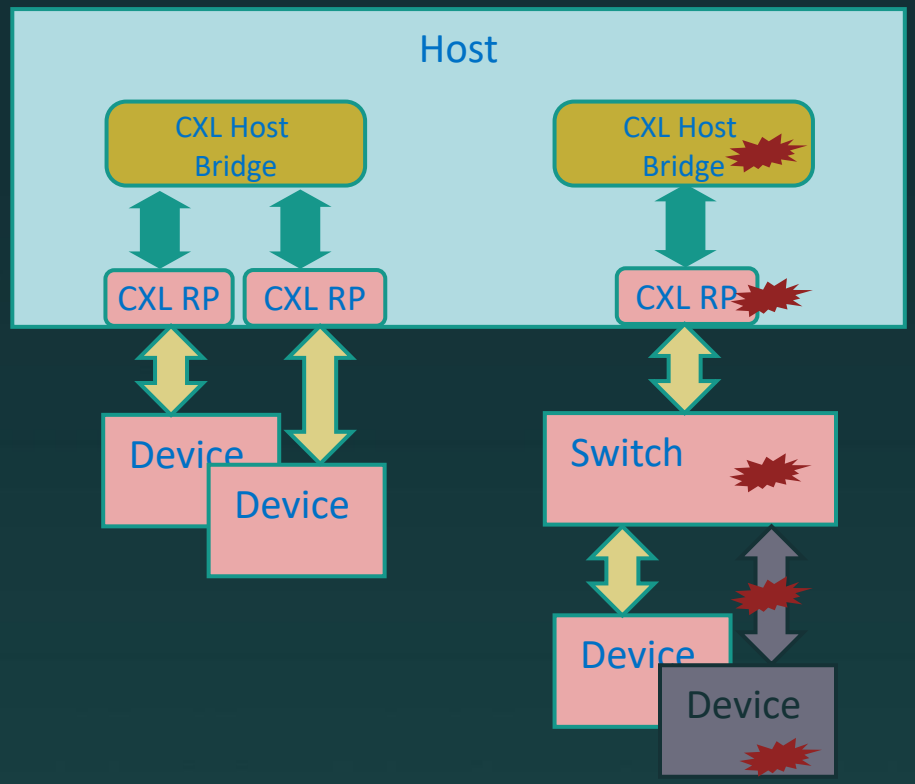

(Type 3 Device)

# CXL 2.0 ECN Process



# CXL Error Isolation – Problem Statement

- This ECN addresses graceful handling of uncorrectable errors at or below the CXL Root Port
- Prior to CXL Error Isolation, two mechanisms existed for error containment – Data Poisoning & Viral
- Limitations of existing mechanisms
  - Neither of these covered a variety of error types such as
    - Surprise link down (for any reason)
    - CXL device specific fatal errors
    - Security violation detected at a CXL memory controller
  - Such errors would lead to a full system crash
- CXL Error Isolation ECN allows such errors on CXL.cache and CXL.mem to be contained at the CXL Root Port and provides a mechanism for software to opportunistically recover without a full system reset



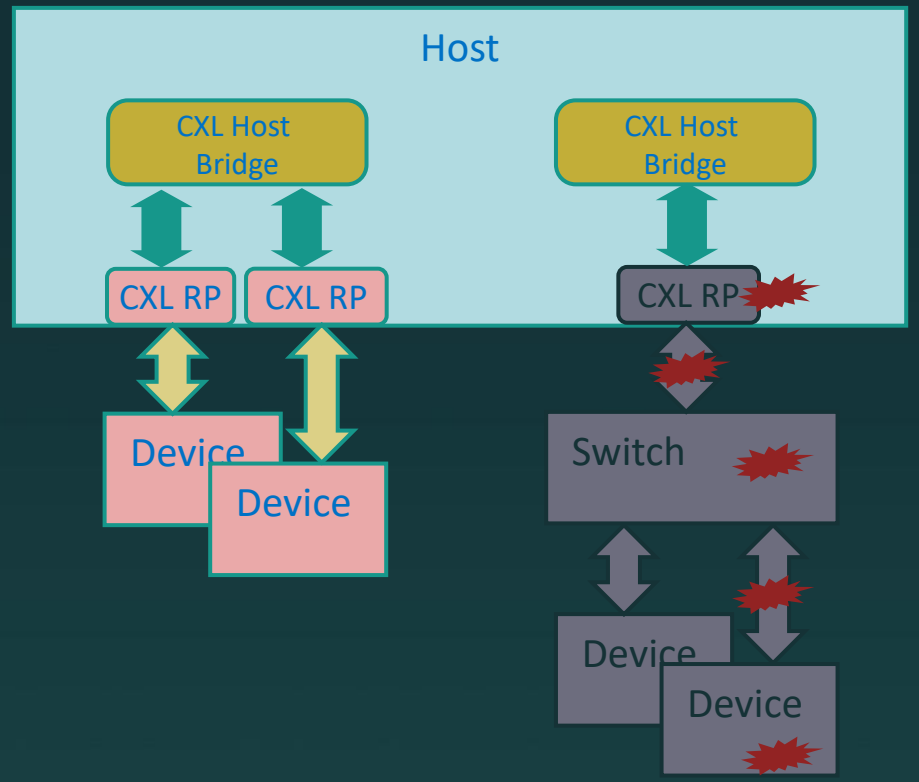
# CXL Error Isolation - Solution

- CXL.mem and/or CXL.cache Error Isolation independently triggered at a CXL Root Port when:
  - Request times out
  - Link goes down

- Once Error Isolation is triggered, the CXL Root Port synthesizes responses for all new and pending requests on that protocol
  - Reads return synchronous exception error
  - Writes and dropped and completed

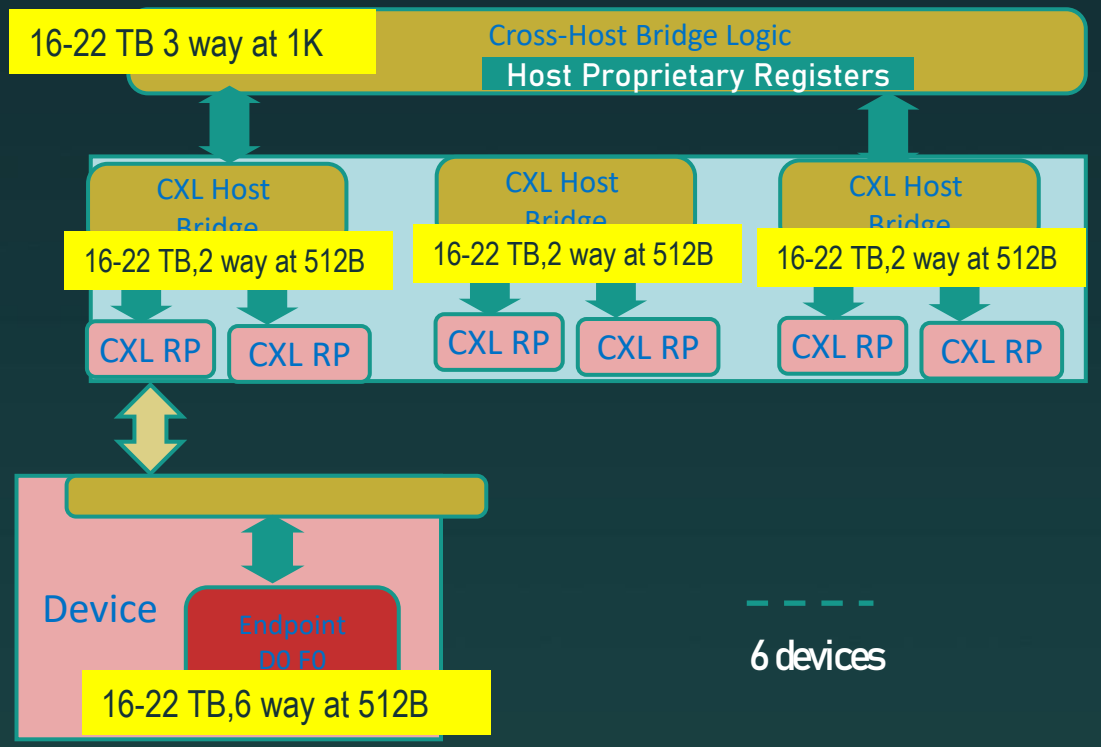
- This allows the error to be contained to the impacted CXL Root Port only and thus, the rest of the system can continue functioning. Software can optionally recover from this by terminating affected WLs and resetting the impacted device(s) and link(s)

- This ECN only impacts the Host. It requires no changes in the Device or in the Switch



# Memory Interleaving Enhancements

- CXL 2.0 specification defined 2, 4 and 8-way interleaving options
- This ECN introduces 3, 6, 12 and 16-way interleaving options
  - Matches interleaving options that are available with native DDR
  - Sweet spots that CXL 2.0 spec missed
- No impact to Switches
- 3, 6 and 12-way interleaving implies 3-way math
  - Hosts support mod3 to select one out of 3 (or 6 or 12) targets
  - The device needs to implement divide by 3 (or 6 or 12) operation on HPA
- The ECN limits the legal combinations
  - Limits scope
  - Avoid potential address aliasing when mixing 2 way and 3-way math
- 6 way interleaving across 6 Root Ports shown here



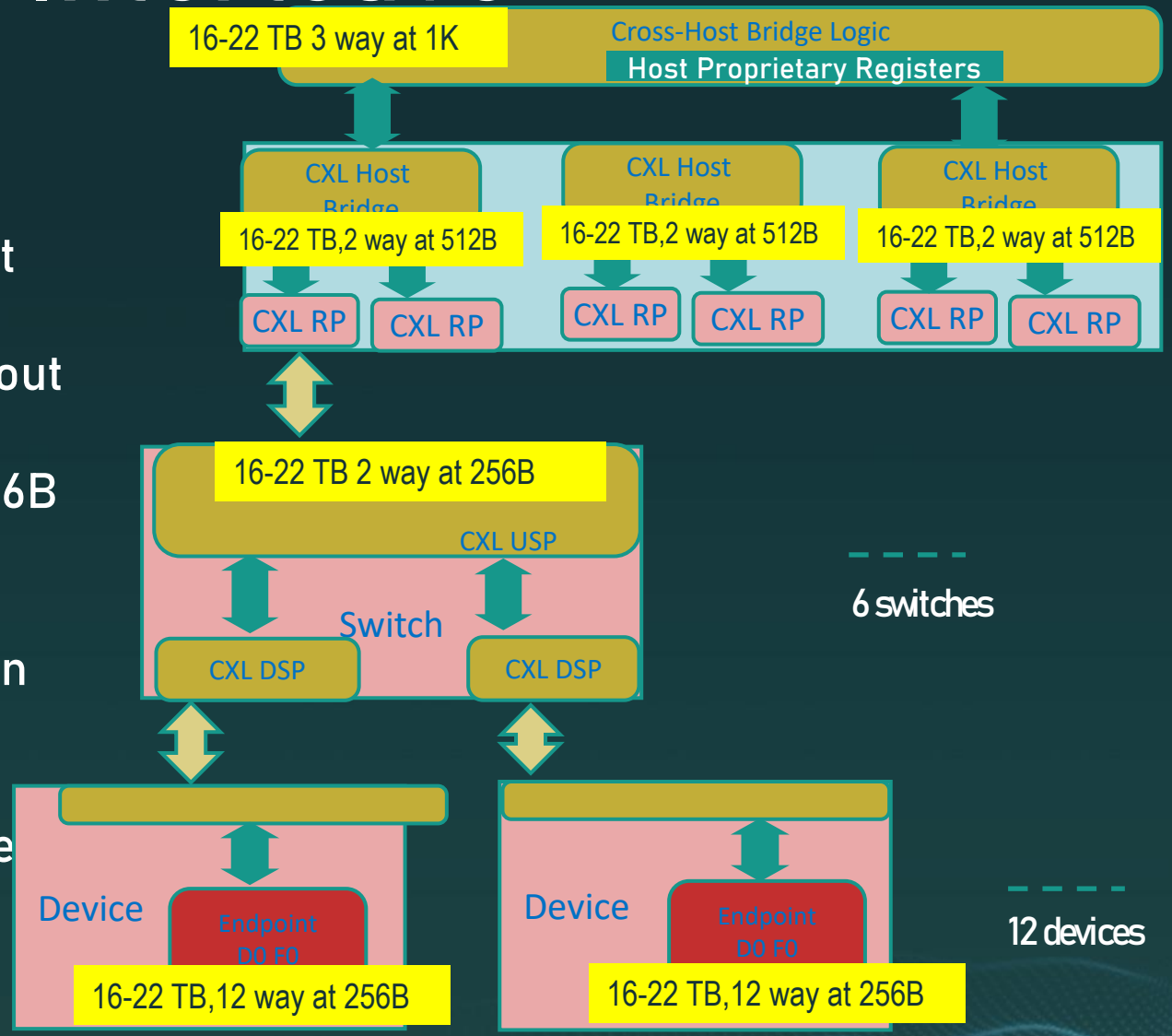
# 12-way Interleave

- A more complex, 12-way interleaving example shown here

- Cross-host Bridge logic picks one CXL Host Bridge out of 3 at 1K granularity
- Each CXL Host Bridge picks one Root Port out of 2 at 512B Granularity
- Each USP selects one DSP our of two at 256B granularity

- The device performs a divide by 12 operation during HPA to DPA conversion

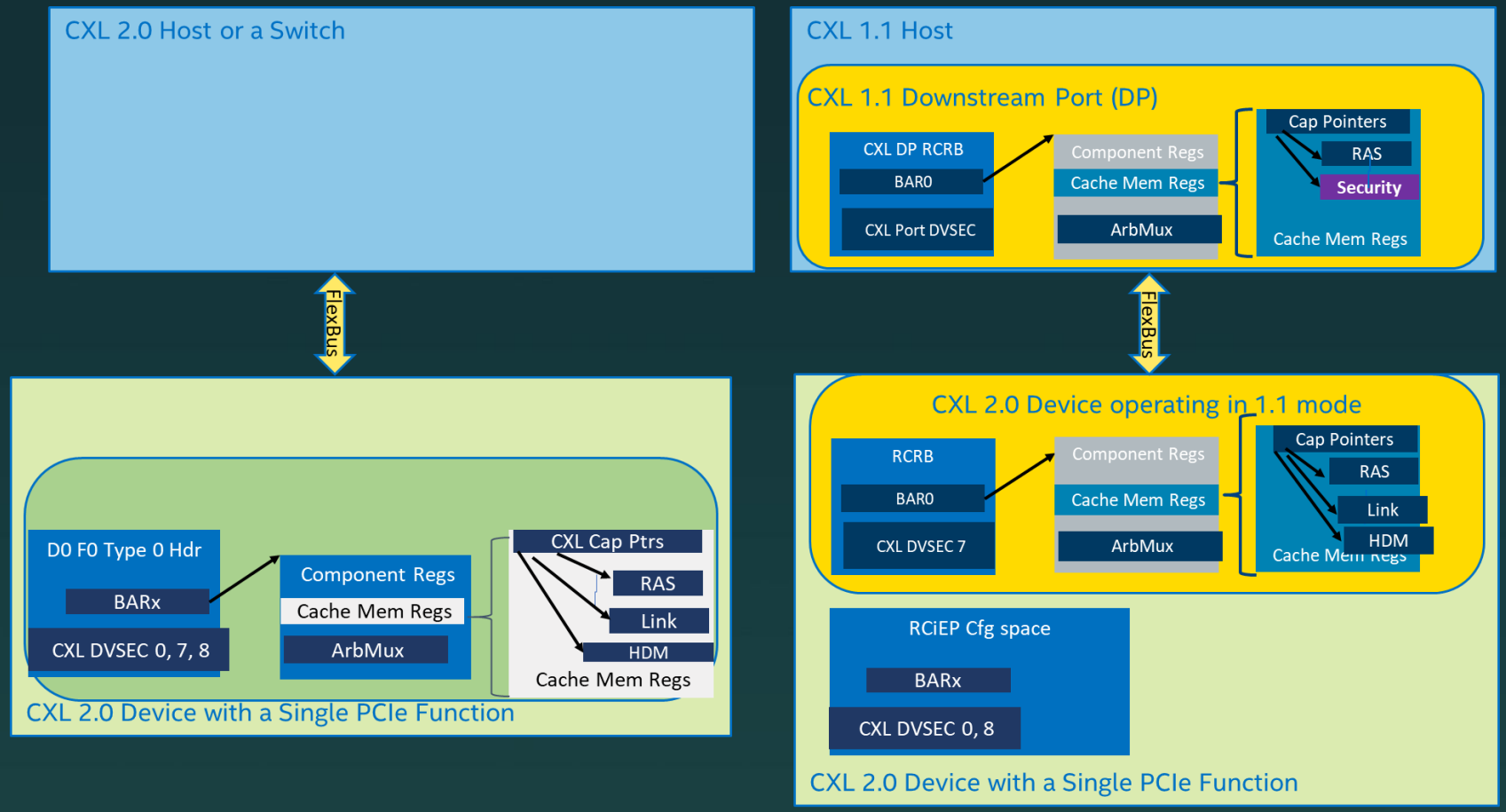
- Contiguous address bits used for interleave selection



# 1.1 Mode Operation without RCRB – Problem Statement

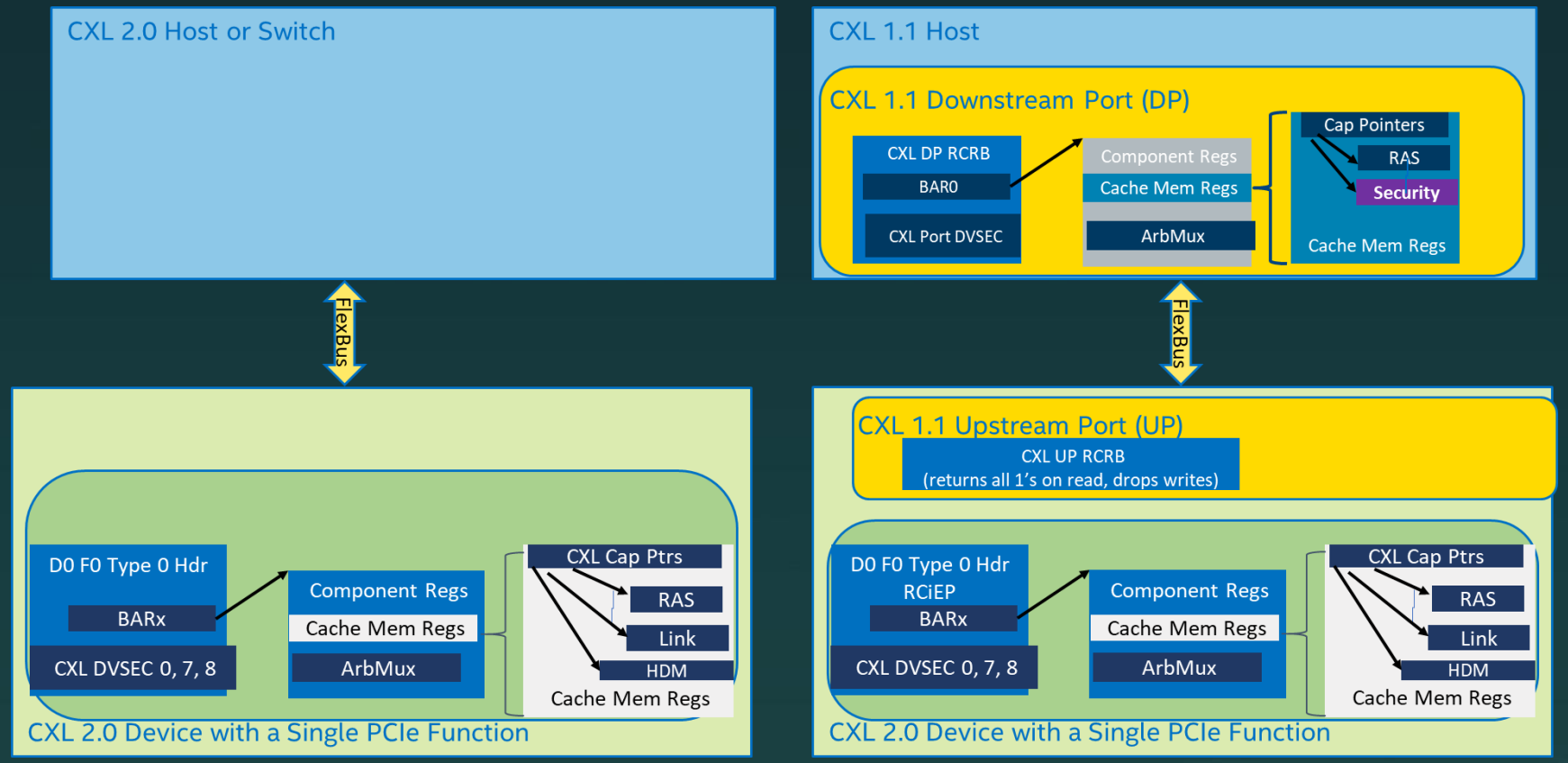
- A CXL 2.0 device is required to interop with CXL 1.1 host

- When connected to CXL 2.0 host, device registers are exposed via Endpoint config space and standard PCIe BAR
- When connected to a 1.1 host, several registers need to be exposed via RCRB and MEMBAR0
- The device needs to remap the register dynamically based on the outcome of training (CXL 2.0 vs. 1.1)
- The ecosystem indicated this was challenging ..



# 1.1 Mode Operation without RCRB - Solution

- Allows CXL 2.0 device operating in CXL 1.1 mode to have a register layout of a CXL 2.0 device, thus simplifying device designs
- Requires a change to CXL 1.1 enumeration software to detect these devices
- If read to RCRB base+4K returns all 1's, SW assume this device implements this ECN.

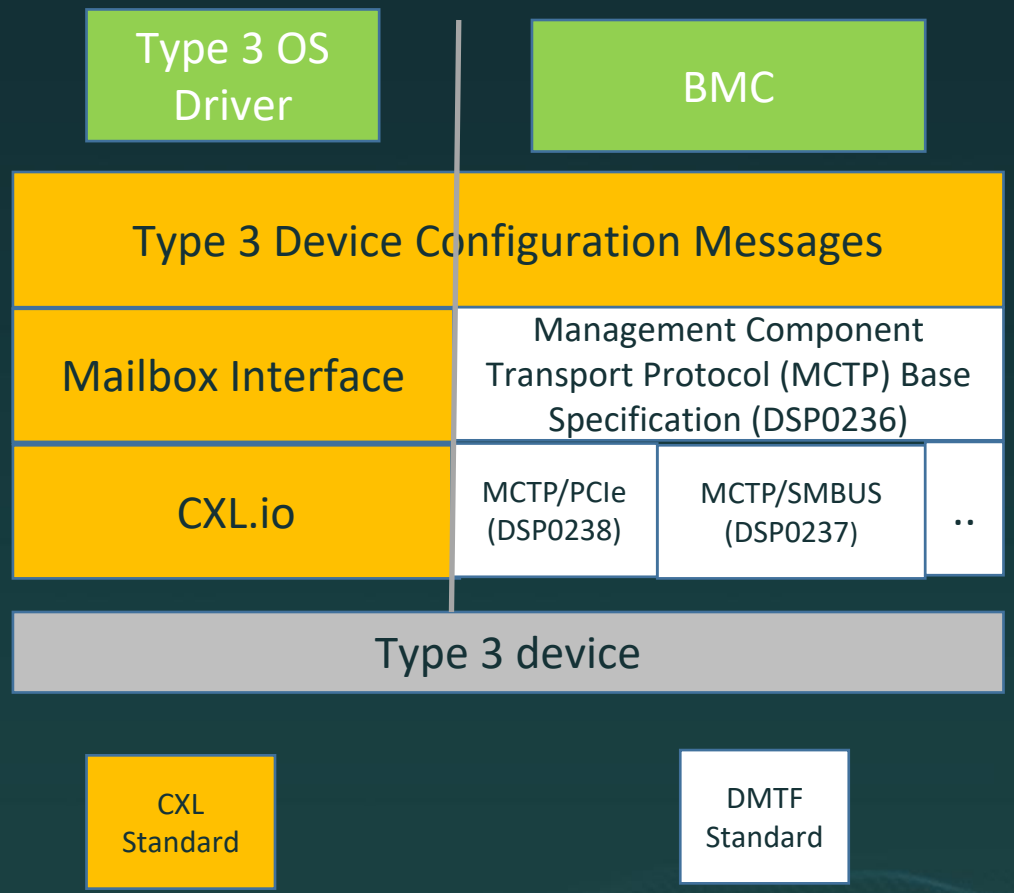


# MCTP CCI ECN

- The CXL ecosystem expectations from Type 3 devices
  - Standard-based and symmetric in-band and out-of-band configuration i/f
  - Time to market
  - Simple design to enable lower cost devices
  - Optimized BMC FW option

- No existing standard met the in-band configuration interface requirements, so CXL 2.0 specification defined one.
- Covers Media management, memory errors, FW update, data-at-rest security, ..

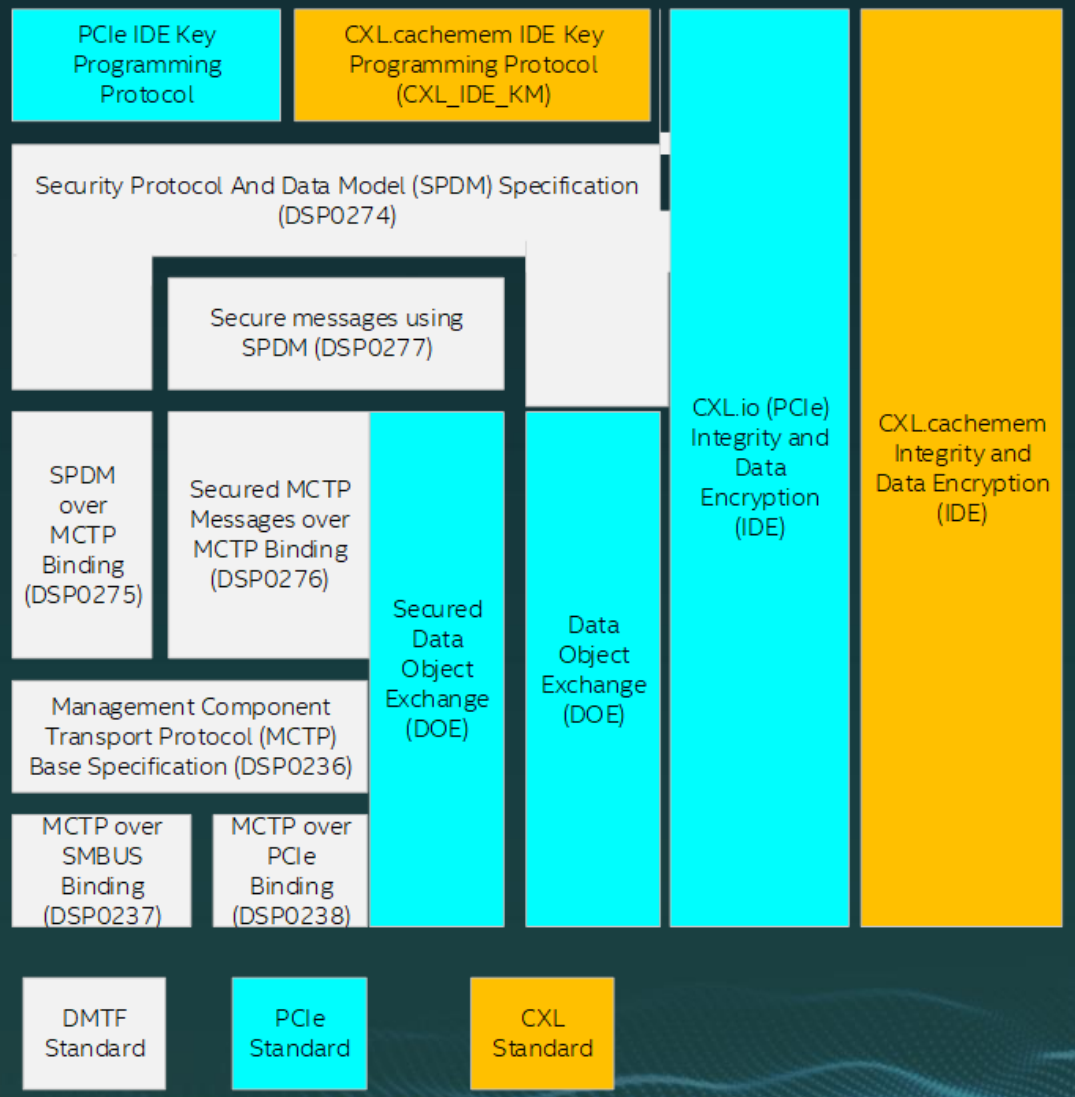
- MCTP CCI ECN allows these very same messages to be transported using a new MCTP message type (Type 8)



# IDE Establishment ECN

- CXL 2.0 specification defined the protocol level changes needed to enable CXL.cachemem IDE.
  - IDE encrypts and integrity protects the CXL.cachemem traffic on the link
- This ECN defines the software/firmware flows.
- Defines a set of messages, CXL_IDE_KM
  - Requests by host software (or another agent like BMC) to IDE capable component
  - These requests provision IDE keys and instruct the components to initiate or tear down CXL IDE
  - Requests and responses are confidentiality and integrity protected

- Enables architecture and code reuse by leveraging existing industry standards such
  - Suite of Security Protocol and Data Model (SPDM) specifications from DMTF
  - PCIe IDE ECN



# Compliance ECN Package

- Memory Device Error Injection
  - Inject poison in data and meta-data areas
  - Inject Health status change
- Compliance Tests for Viral Error Injection–
  - Injecting viral for purposes of testing
- Compliance DOE Return Values
  - Query Compliance DOE capabilities
- Compliance DOE 1B
  - Simplifies a compliance test
- QoS Telemetry Compliance Testcases
  - QoS compliance tests

# Summary of other ECNs

- CEDT CFMWS and QTG _DSM – FW/SW infrastructure
  - CFMWS: ACPI Table extensions to describe host specific memory address assignments
  - QTG: Software primitives to enable CXL 2.0 QoS features
- Mailbox Ready time – Design Flexibility
  - Accommodates devices that may take longer time to ready mailbox interface
- Register Locator DVSEC – Vendor extensions
  - Allows MMIO mapped vendor specific register blocks
- NULL capability ECN - Design Flexibility
  - Allows components to jump over a CXL.cachemem Capability Header Entry, may simplify component design
- Component State Dump Log – RAS/Debug
  - Standard commands for extraction of a component “crash log” and requesting the component to capture current state


# Thank You