Introducing the CXL 3.X Specification

Mahesh Natu

System and Software WG Co-Chair - CXL Consortium

Senior Principal Engineer and Director of Platform Architecture – Intel Corporation

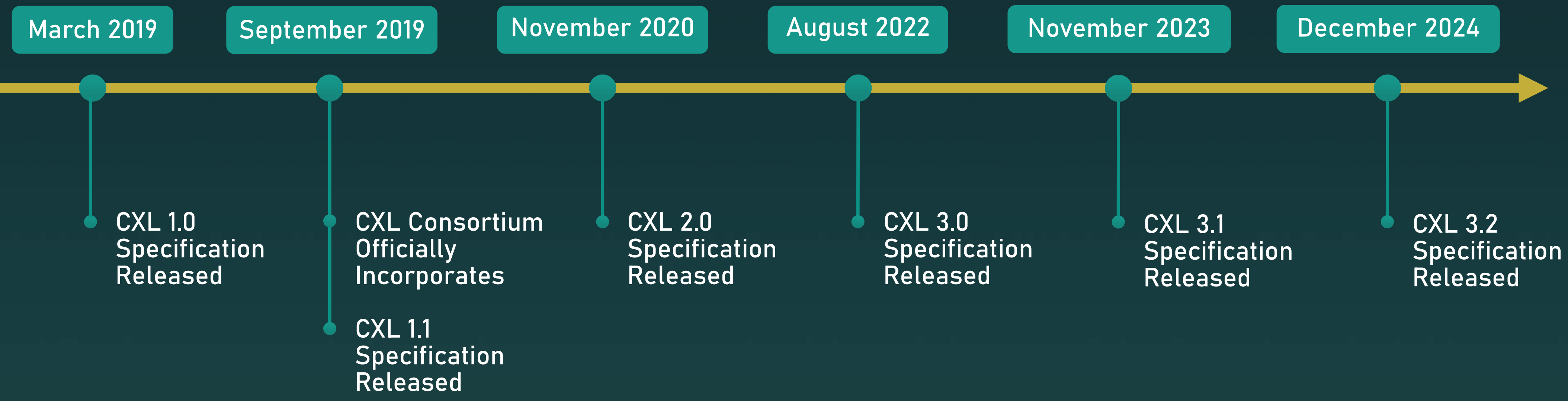
Agenda

- Industry Trends and CXL 3.X Themes
- CXL 3.X Features Progression
- CXL 3.2 – New Feature Enhancements
- Compliance Updates
- Summary
- Q&A

Industry Trends and CXL 3.X Themes

- AI and ML applications, heterogenous computing → 2X Bandwidth, Caching protocol enhancements, large fabric
- Disaggregation of memory from compute → standardize i/f for managing pooled and shared memory
- Lower-cost memory tiers deployed to decrease overall platform costs → standardize Hot-Page detection
- Confidential computing → TSP support for CXL memory devices and accelerators
- CXL becomes the industry choice for coherent IO (CCIX, OpenCAPI and Gen-Z assets transferred to CXL) → Cover use cases previously addressed by these standards such as large fabrics

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 Released

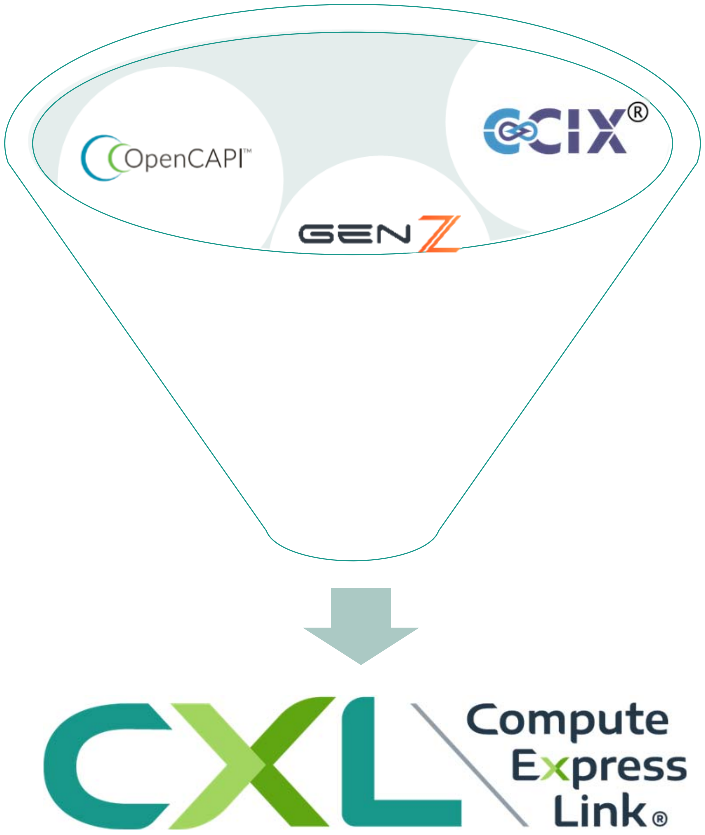
November 2023

CXL 3.1 Specification Released

December 2024

CXL 3.2 Specification Released

Industry Standards Converge



CXL becomes the industry choice for coherent IO

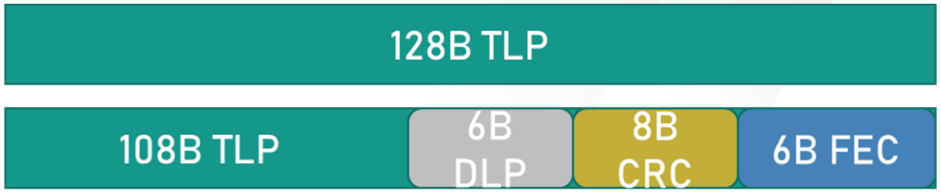
August 3, 2023, CXL Consortium and CCIX Consortium sign letter of intent to transfer CCIX specification and assets to the CXL Consortium

August 1, 2022, CXL Consortium and OpenCAPI Consortium Sign Letter of Intent to Transfer OpenCAPI Assets to CXL

February 2022, CXL Consortium and Gen-Z Consortium signed agreement to transfer Gen-Z specification and assets to CXL Consortium

CXL 3.0: Doubles Bandwidth with Same Latency

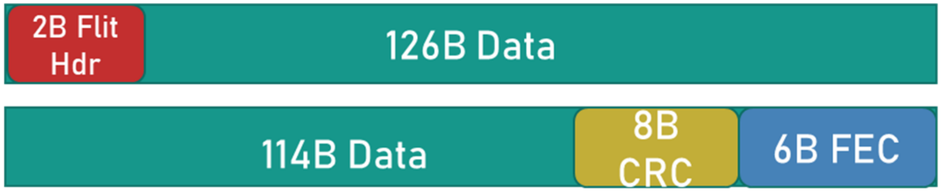
- Uses PCIe® 6.0 PHY @ 64 GT/s
- PAM-4 and high BER mitigated by PCIe 6.0 FEC and CRC (different CRC for latency optimized)
- Standard 256B Flit along with an additional 256B Latency Optimized Flit (0-latency adder over CXL 2)
  - 0-latency adder trades off FIT (failure in time, 109 hours) from 5x10-8 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



128B TLP

108B TLP | 6B DLP | 8B CRC | 6B FEC

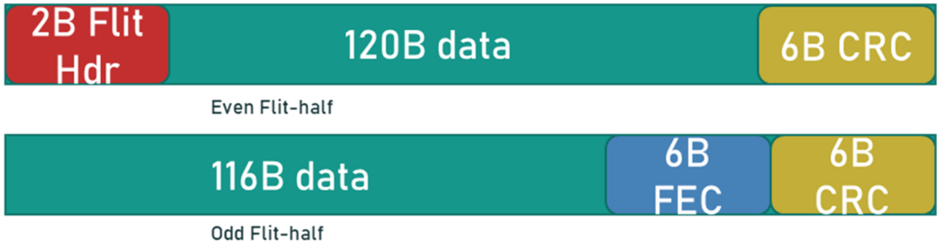
(PCIe 6.0 Flit Layout)



2B Flit Hdr | 126B Data

114B Data | 8B CRC | 6B FEC

(CXL 256B Standard Flit Layout)



2B Flit Hdr | 120B data | 6B CRC

Even Flit-half

116B data | 6B FEC | 6B CRC

Odd Flit-half

(CXL 256B Latency-Optimized Flit Layout)

[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.X Features Progression

CXL Scales New Heights

CXL 3.X

“Row Scale”

- Composable Fabric growth for disaggregation / pooling / accelerator
- use cases previously addressed by Gen-Z

CXL 2.0

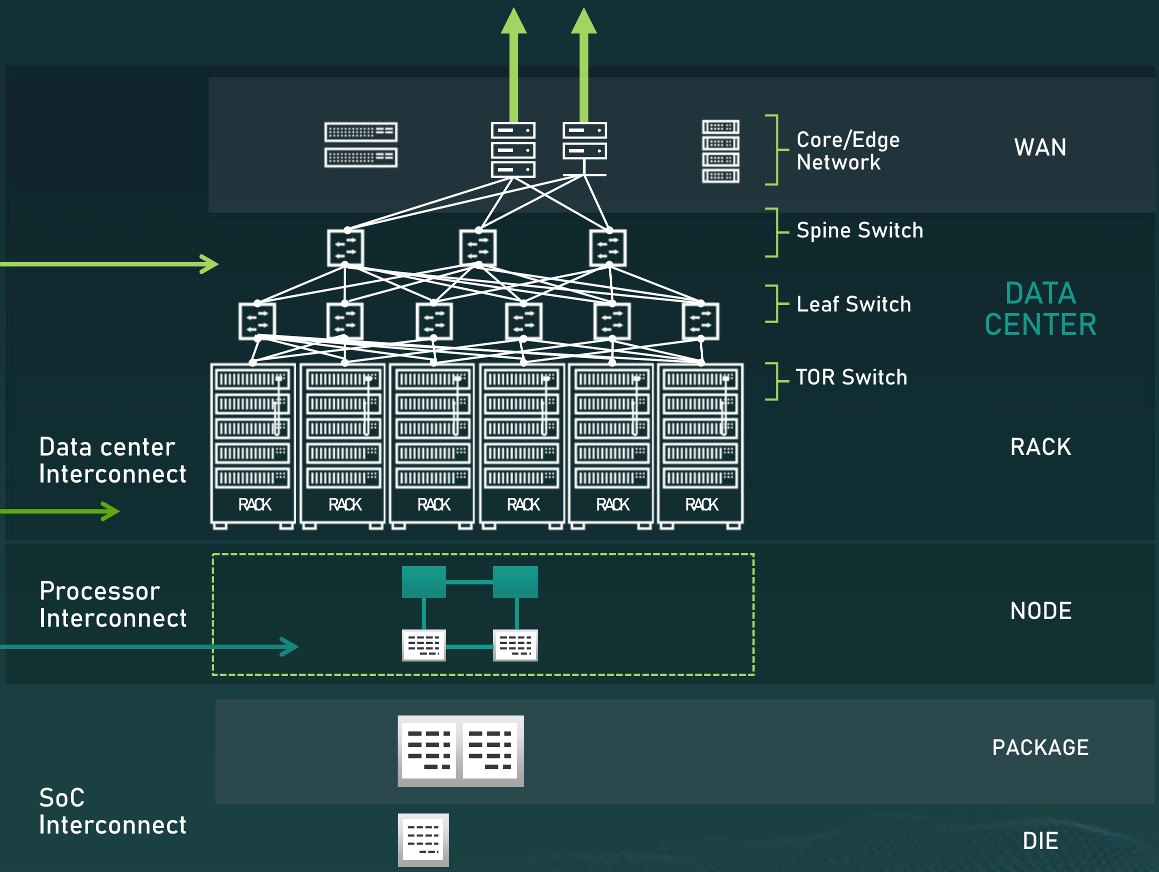
“Rack Scale”

- Multiple nodes inside a Rack/Chassis supporting pooling of resources
- Memory/accelerator pooling with single logical devices (SLD)
- Memory pooling with multiple logical devices (MLD)

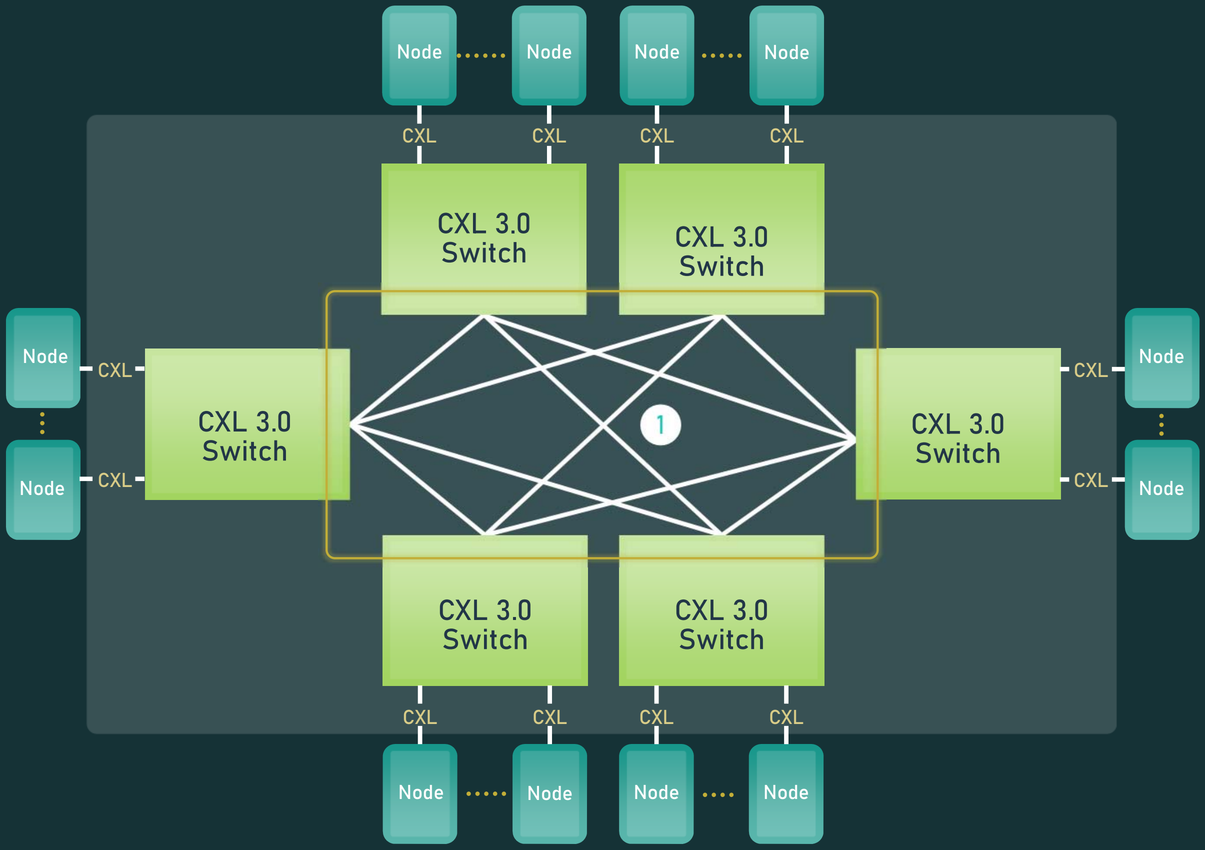
CXL 1.1

“Single server”

- Single Node Coherent interconnect



CXL Fabrics



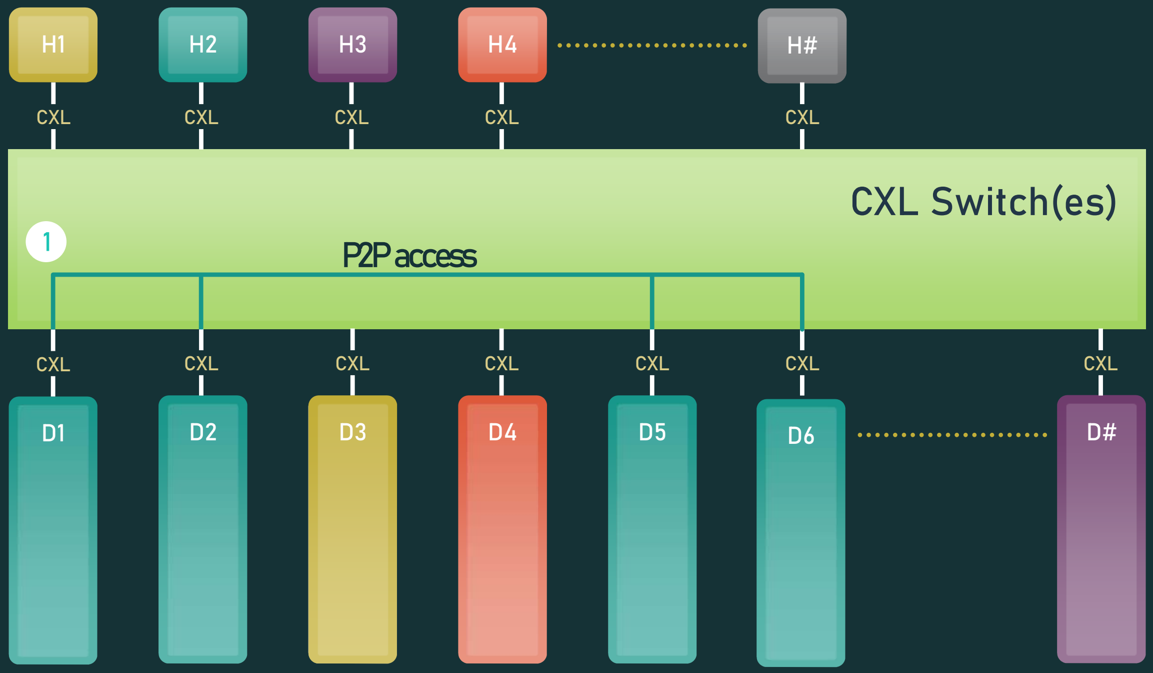
1 CXL 3.0 enables non-tree architectures

- Each node can be a CXL Host, CXL device or PCIe device

CXL 3.1 enables even larger fabrics via Port-based Routing, Fabric attached devices and Fabric Management APIs

Compute Express Link® and CXL® are registered trademarks of the Compute Express Link Consortium.

CXL 3.x Peer-to-peer Comms



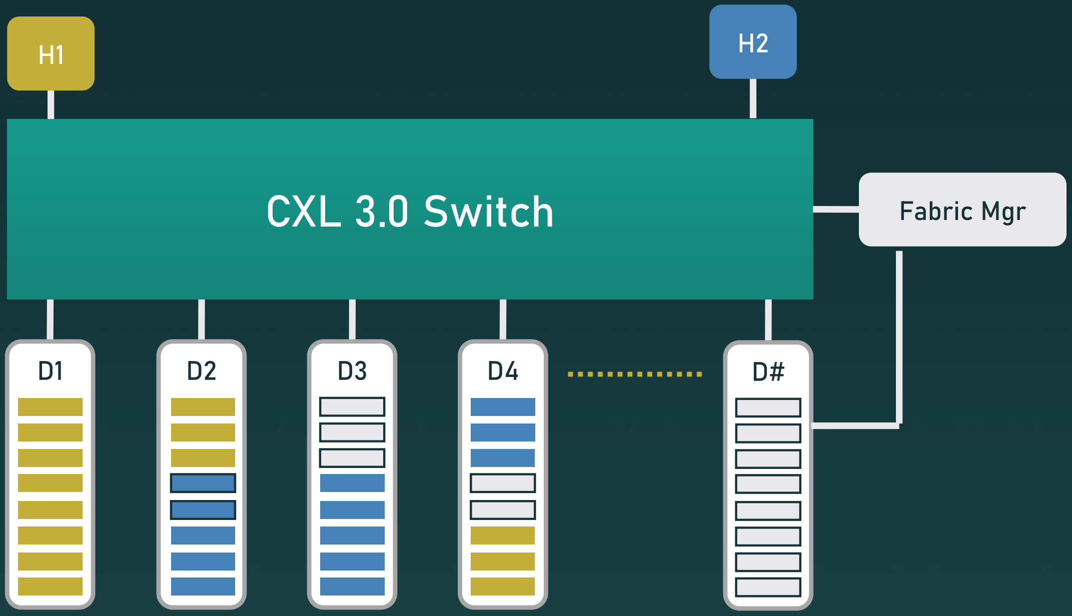
1 CXL 3.0 enables efficient peer-to-peer communication (P2P) between devices. Relies on PCIe Unordered I/O.

The target device that hosts the memory returns the latest copy, by using the Back-Invalidation protocol extension

CXL 3.1 adds peer-to-peer communication (P2P) using CXL.mem

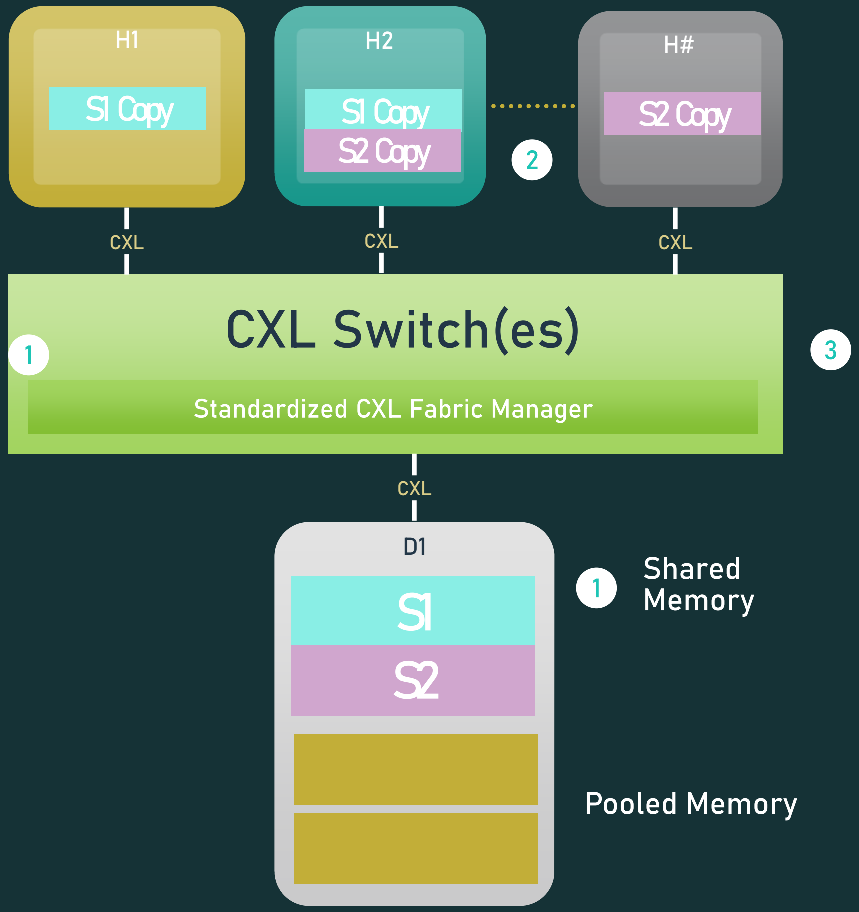
Compute Express Link® and CXL® are registered trademarks of the Compute Express Link Consortium.

CXL 3.X - Memory Pooling



- Memory Pooling allows a host to dynamically expand/shrink its memory capacity to match Workload
- Improves TCO by reducing stranded memory capacity
- CXL 3.0 standardized OS to device and Fabric Manager to device/switch interfaces
- CXL 3.1 expanded the scope to include Fabric attached devices

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

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

CXL Trusted Security Protocol (TSP)

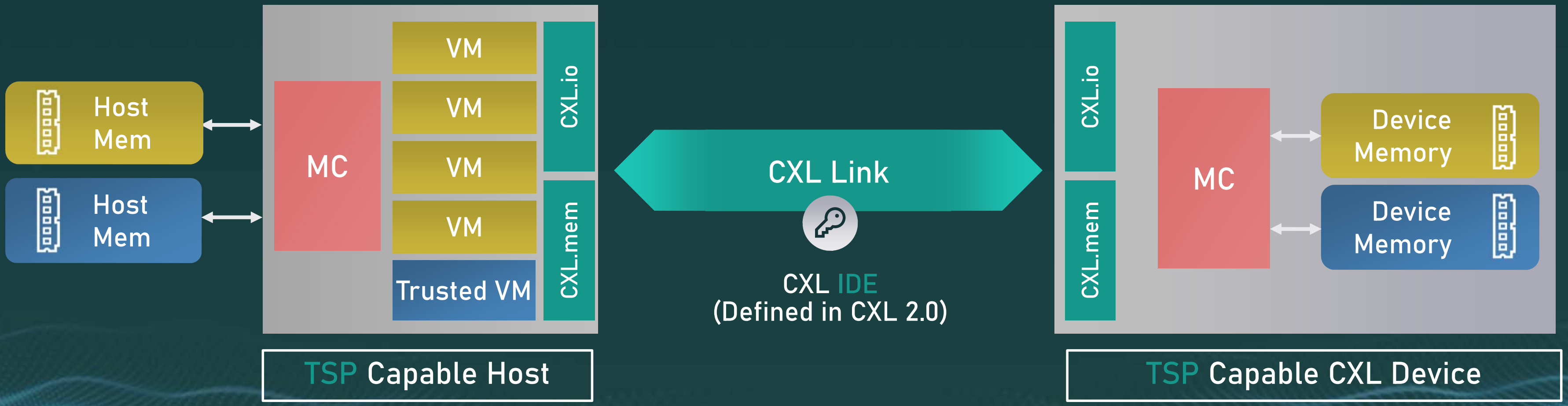
Allows for Virtualization-based, Trusted Execution Environments (TEEs)
to host Confidential Computing Workloads

Key Capabilities:

- Cryptographic Separation between Trusted VM & CSP infrastructure
- Support for memory devices and accelerators
- Encryption of sensitive data in Host & Device memory during use
- Cryptographically verify configuration of the computing environment

Benefits:

- Freedom to migrate sensitive WLs to TSP-enabled Clouds
- Collaboration with multiple parties without exposing secrets
- Conform to Compliance & Data sovereignty programs
- Strengthen Application security & Software IP protection



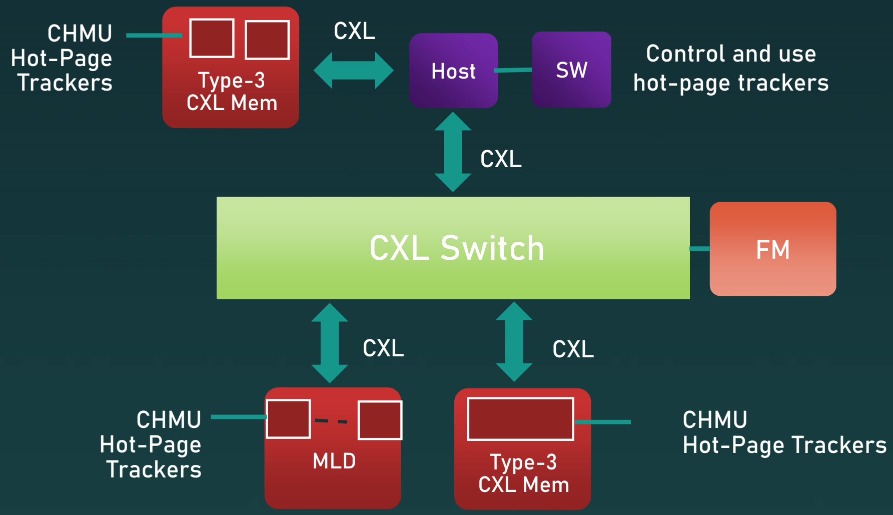
TSP Feature Progression

- TSP builds on top of CXL Integrity and Data Encryption (IDE) capability introduced in CXL 2.0
- CXL 3.0 introduces TSP for simple memory devices that rely on host for coherency management
- CXL 3.1 specification extended TSP to cover devices such as accelerators
- CXL 3.1 extended IDE protection to late poison messages
- CXL 3.2 specification added TSP compliance tests for improving interop

CXL 3.2 Specification

New Feature Enhancements

CXL Hot-Page Monitoring Unit (CHMU) for Memory Tiering



More efficient SW Memory Tiering
Better Perf, lower TCO

Challenges faced by the current SW tiering solutions

- Must trade-off accuracy against perf overhead
- Measurement polluted by cache hits
- CPU vendor specific

CHMU addresses these problems

- Works for simple and pooling memory devices
- Hot-page trackers implemented in CXL memory device, avoids host perf overhead
- Standardized interface, enables generic OS based solutions
- By design, counts memory accesses only, excludes cache hits
- Multiple CXL Hot-Page Monitoring Unit (CHMU) instances provides SW more flexibility.
- Allows counting at different granularity.
- Improves memory workload analysis

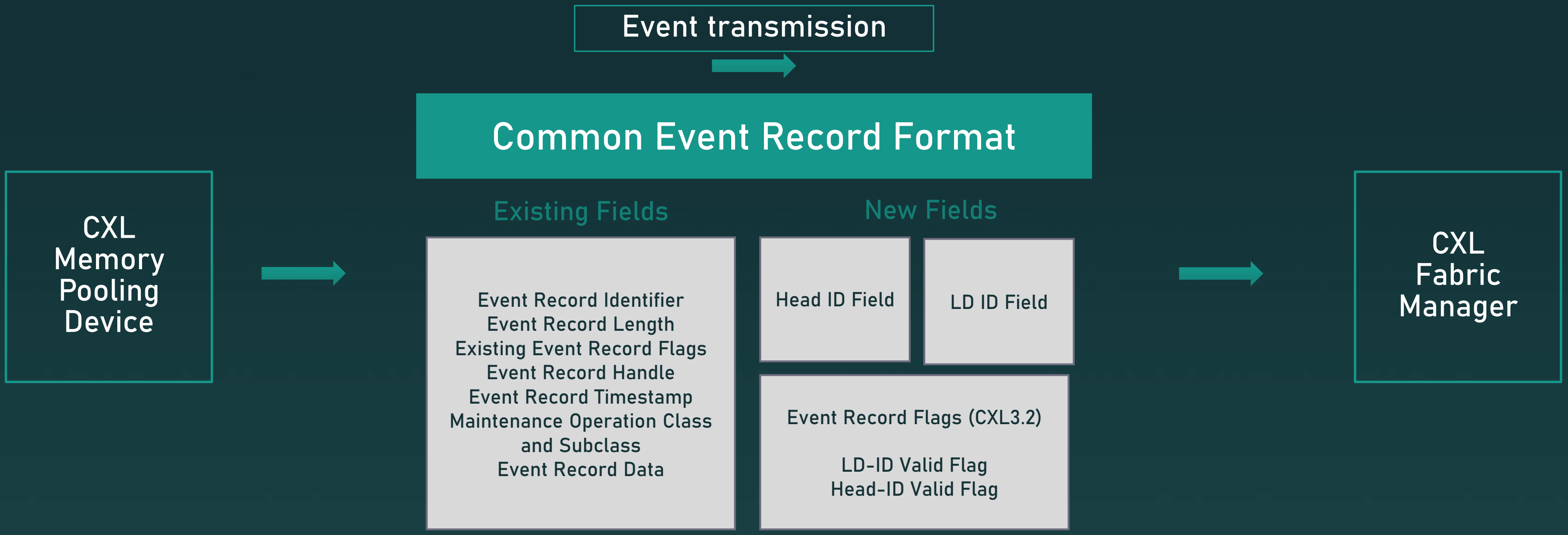
CHMU Overview

- Highly configurable, SW can make best use of these critical resources.
  - Counts accesses on specific DPA granularities called units; unit sizes is SW configurable
  - A unit is marked as hot if it encounters more accesses than software configurable threshold during an epoch. Epoch length is also SW configurable.
  - Access counting may be enabled on multiple address ranges with 256-MB granularity.
- Hot units are reported to SW thru' circular structure called Hotlist, the raw counters are not exposed to SW allowing device vendors to innovate
- SW can either poll for Hotlist or choose to be interrupted when Hotlist starts to become full
- SW chooses the types of CXL.mem requests that are counted.

Compatibility with the PCIe® MMPT ECN

- A great example of collaboration with PCI SIG
- Management Message Pass Through (MMPT) ECN was built on top of CXL 2.0 specification constructs and makes special accommodates for CXL backward compatibility
- Enables unified OS based management of CXL and PCIe devices, everybody wins!

CXL 3.2 Enhances Event Record



Event transmission

CXL Memory Pooling Device

Common Event Record Format

Existing Fields

Event Record Identifier
Event Record Length
Existing Event Record Flags
Event Record Handle
Event Record Timestamp
Maintenance Operation Class and Subclass
Event Record Data

New Fields

Head ID Field

LD ID Field

Event Record Flags (CXL3.2)

LD-ID Valid Flag
Head-ID Valid Flag

CXL Fabric Manager

More localized error handling of Memory Pooling devices
Limiting the error blast radius to fewer hosts.

CXL 3.2 Enhances functionality of CXL Memory Devices for OS and Application

Post Package Repair (PPR) enhancements

- Function: Enables PPR (Post Package Repair) at the hardware-level during initialization hPPR (Hardware Post Package Repair).
- Benefit: Extends RAS for CXL Memory Devices allowing seamless repair to the attached memory.

Addition of performance monitoring events for CXL Memory Devices

- Function: Adds CXL memory performance counters, events, and performance enhancements.
- Benefit: Provides memory usage analytics for OS/Application.

Meta-bits Storage Feature for Host-only Coherent Host-Managed Device Memory (HDM-H) address region

- Function: Allows the host to discover and control meta-data usage.
- Benefit: Dyanamic optimization of DRAM usage to match host requirements.

CXL Specification Feature Summary

Not Supported

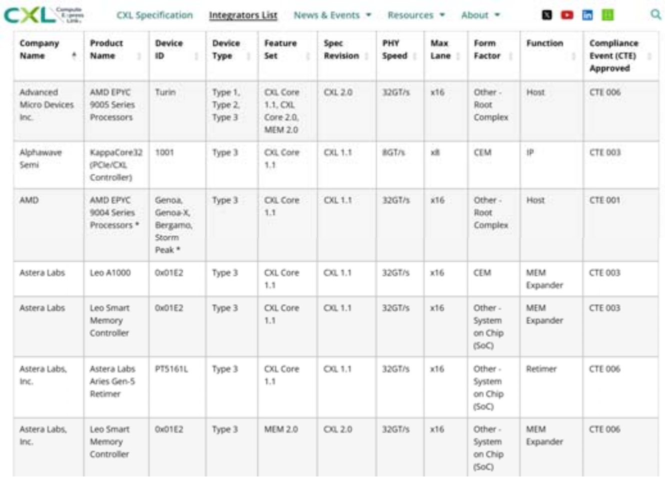
✓ Supported

Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0 / 3.1	CXL 3.2
Release date	2019	2020	2022 / 2023	November 2024
Max link rate	32GTs	32GTs	64GTs	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)			✓	✓
Back invalidate capabilities on Type 3 devices (HDM-DB)			✓	✓
Fabric Manager API definition for PBR Switch			✓	✓
Host-to-Host communication with Global Integrated Memory (GIM) concept			✓	✓
Trusted-Execution-Environment (TEE) Security Protocol			✓	✓
Memory expander enhancements (up to 32-bit of meta data, RAS capability enhancements)			✓	✓
Security, compliance, and CXL Memory Device enhancements				✓

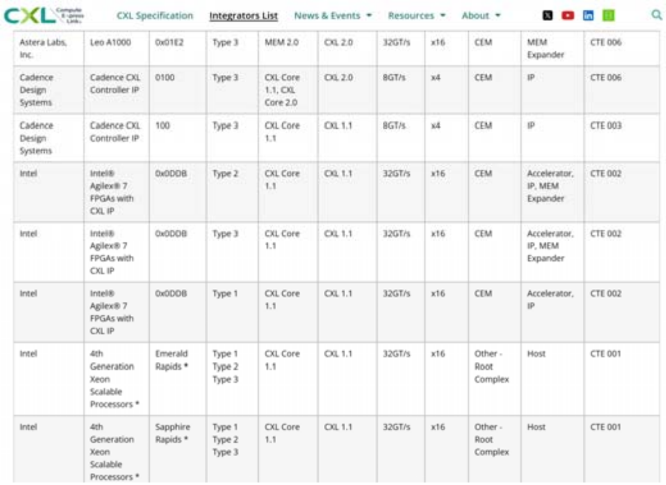
Compliance Updates

Official testing for CXL 2.0 kicked off in December 2024

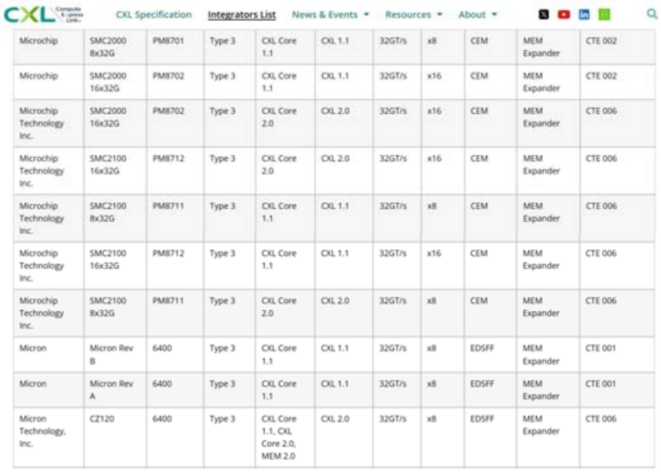
- CXL hosts multiple Test Events each year to provide Members with opportunities to test the functionality and interoperability of CXL devices and feature their devices on the CXL Integrators List
- The CXL Integrators List features over 48+ devices: https://computeexpresslink.org/integrators-list/



Company Name	Product Name	Device ID	Device Type	Feature Set	Spec Revision	PHY Speed	Max Lane	Form Factor	Function	Compliance Event (CTE) Approved
Advanced Micro Devices Inc.	AMD EPYC 9005 Series Processors	Turin	Type 1, Type 2, Type 3	CXL Core 1.1, CXL Core 2.0, MEM 2.0	CXL 2.0	32GT/s	x16	Other - Root Complex	Host	CTE 006
Alphawave Semi	KappaCore32 (PCIe/CXL Controller)	1001	Type 3	CXL Core 1.1	CXL 1.1	8GT/s	x8	CEM	IP	CTE 003
AMD	AMD EPYC 9004 Series Processors *	Genoa, Genoa-X, Bergamo, Storm Peak *	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	Other - Root Complex	Host	CTE 001
Astera Labs	Leo A1000	0x01E2	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	MEM Expander	CTE 003
Astera Labs	Leo Smart Memory Controller	0x01E2	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	Other - System on Chip (SoC)	MEM Expander	CTE 003
Astera Labs, Inc.	Astera Labs Aries Gen-5 Retimer	PT5161L	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	Other - System on Chip (SoC)	Retimer	CTE 006
Astera Labs, Inc.	Leo Smart Memory Controller	0x01E2	Type 3	MEM 2.0	CXL 2.0	32GT/s	x16	Other - System on Chip (SoC)	MEM Expander	CTE 006



Astera Labs, Inc.	Leo A1000	0x01E2	Type 3	MEM 2.0	CXL 2.0	32GT/s	x16	CEM	MEM Expander	CTE 006
Cadence Design Systems	Cadence CXL Controller IP	0100	Type 3	CXL Core 1.1, CXL Core 2.0	CXL 2.0	8GT/s	x4	CEM	IP	CTE 006
Cadence Design Systems	Cadence CXL Controller IP	100	Type 3	CXL Core 1.1	CXL 1.1	8GT/s	x4	CEM	IP	CTE 003
Intel	Intel® Agilex® 7 FPGAs with CXL IP	0x0DDB	Type 2	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	Accelerator, IP, MEM Expander	CTE 002
Intel	Intel® Agilex® 7 FPGAs with CXL IP	0x0DDB	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	Accelerator, IP, MEM Expander	CTE 002
Intel	Intel® Agilex® 7 FPGAs with CXL IP	0x0DDB	Type 1	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	Accelerator, IP	CTE 002
Intel	4th Generation Xeon Scalable Processors *	Emerald Rapids *	Type 1 Type 2 Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	Other - Root Complex	Host	CTE 001
Intel	4th Generation Xeon Scalable Processors *	Sapphire Rapids *	Type 1 Type 2 Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	Other - Root Complex	Host	CTE 001



Microchip	SMC2000 8x32G	PM8701	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x8	CEM	MEM Expander	CTE 002
Microchip	SMC2000 16x32G	PM8702	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	MEM Expander	CTE 002
Microchip Technology Inc.	SMC2000 16x32G	PM8702	Type 3	CXL Core 2.0	CXL 2.0	32GT/s	x16	CEM	MEM Expander	CTE 006
Microchip Technology Inc.	SMC2100 16x32G	PM8712	Type 3	CXL Core 2.0	CXL 2.0	32GT/s	x16	CEM	MEM Expander	CTE 006
Microchip Technology Inc.	SMC2100 8x32G	PM8711	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x8	CEM	MEM Expander	CTE 006
Microchip Technology Inc.	SMC2100 16x32G	PM8712	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x16	CEM	MEM Expander	CTE 006
Microchip Technology Inc.	SMC2100 8x32G	PM8711	Type 3	CXL Core 2.0	CXL 2.0	32GT/s	x8	CEM	MEM Expander	CTE 006
Micron	Micron Rev B	6400	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x8	EDSFF	MEM Expander	CTE 001
Micron	Micron Rev A	6400	Type 3	CXL Core 1.1	CXL 1.1	32GT/s	x8	EDSFF	MEM Expander	CTE 001
Micron Technology, Inc.	CZ120	6400	Type 3	CXL Core 1.1, CXL Core 2.0, MEM 2.0	CXL 2.0	32GT/s	x8	EDSFF	MEM Expander	CTE 006

Summary

- CXL 3.2 provides security, compliance, and CXL Memory Device enhancements
  - Optimizes CXL Memory Device Monitoring and Management
  - Enhances functionality of CXL Memory Devices for OS and Application
  - Extends security with TSP (Trusted Security Protocol)
    - IDE protection for late poison messages
    - Added for HDM-DB memory devices
    - Compliance testing

- Looking forward
  - CXL Consortium Technical Working Groups are developing the next CXL specification to increase speed and improve our features for AI workloads, memory expansion, security, and reliability.

- CXL 1.1 and 2.0 devices are available in the market today!
  - Scan the QR code to see the growing CXL device ecosystem

Q&A

Please share your questions in the Question Box

Thank You

www.ComputeExpressLink.org