Introducing the Compute Express Link™ 2.0 Specification

Dr Debendra Das Sharma
Intel Fellow, Director I/O Technology and Standards, Intel
Compute Express Link Board Technical Task Force Co-chair

Ahmad Danesh
Product Marketing and Strategy, Data Center Solutions, Microchip Technology
Compute Express Link Marketing Workgroup

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

Growth of AI & Analytics

Cloudification of the Network & Edge
CXL Board of Directors

Alibaba Group

AMD

arm

cisco.

cf

DELL EMC

facebook

Hewlett Packard Enterprise

HUAWEI

Google

IBM

intel®

Microchip

Microsoft

XILINX®

CXL

Industry Open Standard for High Speed Communications

130+ Member Companies

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

**Asymmetric Complexity**

Eases burdens of cache coherent interface designs
Data Center: Scope of CXL 2.0 over CXL 1.1

CXL 1.1
- Single Node
- Coherent interconnect

CXL 2.0
- Across Multiple Nodes inside a Rack/Chassis
- Supporting pooling of resources

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

**Caching Devices / Accelerators**
- **TYPE 1**
  - DDR
  - Processor
- **PROTOCOLS**
  - CXL.io
  - CXL.cache
- **USAGES**
  - PGAS NIC
  - NIC atomics

**Accelerators with Memory**
- **TYPE 2**
  - DDR
  - Processor
  - Accelerator
- **PROTOCOLS**
  - CXL.io
  - CXL.cache
  - CXL.memory
- **USAGES**
  - GP GPU
  - Dense computation

**Memory Buffers**
- **TYPE 3**
  - DDR
  - Processor
  - Memory Buffer
- **PROTOCOLS**
  - CXL.io
  - CXL.memory
- **USAGES**
  - Memory BW expansion
  - Memory capacity expansion
  - Storage class memory

Compute Express Link™ and CXL™ Consortium are trademarks of the Compute Express Link Consortium.
CXL 2.0 Introduces New Features and Usage Models

- Fully backward compatible with CXL 1.1 and 1.0
- Switching and pooling
- Hot-plug support
- Fabric Manager API
- Persistent memory support
- Security
- Built in Compliance & Interop program

Compute Express Link™ and CXL™ Consortium are trademarks of the Compute Express Link Consortium.
CXL Multiprotocol Support with Asymmetry

CXL.IO – PCIe based - discovery, register access, interrupts, initialization, I/O Virtualization, DMA
CXL.Cache – supports device caching of host memory with host processor orchestrating the coherency management
CXL.Memory – memory access protocol, host manages (coherency) device attached memory similar to host memory
Simple coherency flows for device w/ asymmetry => ease of adoption w/ backward compatibility for investment protection

CXL. -- Dynamically Multiplexed IO, Cache and Memory in flit format on PCIe PHY

PCIe PHY

IO (PCIe)
Discovery Configuration Initialization Interrupts DMA IO Virtualization

Cache
Coherent Requests Memory Flows

Memory
Memory Flows

Accelerator Logic

CXL Device

Host Memory

PCIe PHY

Coherence and Memory Logic

Optional Device Memory

CPU Core

CPU Core

Internal IO Device(s)

PCIe/ CXL.IO Logic

Host Processor

IO Device(s)

Compute Express Link™ and CXL™ Consortium are trademarks of the Compute Express Link Consortium.
## CXL 2.0 Scope: Hot-Plug, Persistence, Switching, and Disaggregation

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
</table>
| **CXL PCIe End-Point** | CXL device to be discovered as PCIe Endpoint  
Support of CXL 1.1 devices directly connected to Root-Port or Downstream Switch Port |
| **Switching** | Single level of switching with multiple Virtual Hierarchies (cascaded possible in a single hierarchy)  
CXL Memory Fan-Out & Pooling with Interleaving  
CXL.Cache is direct routed between CPU and device with a single caching device within a hierarchy. Downstream port must be capable of being PCIe. |
| **Resource Pooling** | Memory Pooling for Type3 device – Multiple Logical Device (MLD), a single device to be pooled across 16 Virtual Hierarchies. |
| **CXL.cache and CXL.mem enhancements** | Persistence (Global Persistence Flush), Managed Hot-Plug, Function Level Reset Scope Clarification, Enhanced FLR for CXL Cache/Mem, Memory Error Reporting and QoS Telemetry |
| **Security** | Authentication and Encryption – CXL.IO uses PCIe IDE, CXL defines similar capability for CXL.$Mem |
| **Software Infrastructure/API** | ACPI & UEFI ECNs to cover notification and management of CXL Ports and devices  
CXL Switch API for a multi-host or memory pooled CXL switch configuration and management |

CXL 2.0 is fully backwards compatible with CXL 1.0/1.1 (see next slide for details)  
Predictable spec release cadence by CXL consortium to help the ecosystem plan better.
# CXL 2.0: Backwards compatible with CXL 1.0/1.1

## CPU – Device Connectivity

<table>
<thead>
<tr>
<th></th>
<th>CXL 1.X (1.1/1.0) EP</th>
<th>CXL 2.0 EP</th>
<th>PCIe EP/Switch</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU with CXL 1.x</td>
<td>CXL 1.x</td>
<td>CXL 1.x</td>
<td>PCIe</td>
<td>CXL 2.0 EP needs to support both RCiEP and EP modes</td>
</tr>
<tr>
<td>CPU with CXL 2.0</td>
<td>CXL 1.x</td>
<td>CXL 2.0</td>
<td>PCIe</td>
<td>CXL 2.0 CPU also needs to be bi-modal for backwards compatibility</td>
</tr>
</tbody>
</table>

## CXL 2.0 Switch Connectivity

<table>
<thead>
<tr>
<th></th>
<th>Operation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upstream: CPU</td>
<td>Only CXL 2.0</td>
<td>Since switch definition is with CXL 2.0, the platform requirement is upstream port be a CXL 2.0 CPU</td>
</tr>
<tr>
<td>Downstream: CXL Device</td>
<td>CXL 1.x or CXL 2.0</td>
<td>All downstream CXL ports work as CXL 1.x or CXL 2.0 – mix and match to the device’s capability</td>
</tr>
<tr>
<td>Downstream: PCIe EP/Switch</td>
<td>PCIe</td>
<td>Any CXL switch downstream Port must be able to support a PCIe hierarchy, either an EP or a PCIe switch but assigned to <em>one</em> domain</td>
</tr>
<tr>
<td>Downstream: CXL Switch</td>
<td>N/A</td>
<td>CXL 2.0 is defined only as a single level switch for multiple virtual hierarchies (no cascading of CXL switches)</td>
</tr>
</tbody>
</table>
Benefit of CXL 2.0 Switching

Expansion

- Host
  - CXL 2.0 Switch
    - D1
    - D2
    - D3
    - D4
  - Host
    - CXL 2.0 Switch
      - D1
      - D2
      - D3
      - D4
    - D5
    - D6
Memory/Accelerator Pooling with Single Logical Devices

CXL 2.0 Switch

Memory Pooling with Multiple Logical Devices

CXL 2.0 Switch

Standardized CXL Fabric Manager
CXL 2.0 Pooling without a Switch

Memory Pooling with Single/Multiple Logical Devices

CXL 2.0 Switch

Standardized Fabric Manager

Memory Pooling with through direct connect

CXL 2.0 Switch

Compute Express Link™ and CXL™ Consortium are trademarks of the Compute Express Link Consortium.
Managed Hot Remove (Off-lining)

- Request host to remove device D3 from H1

Host H1 quiesces traffic from D3 and indicates device is safe to remove

Device D3 is removed (off-lined) from H1 and added to pool

Managed Hot Add (On-lining)

Host H2 requests a device

Fabric Manager assigns D3 through Switch and Initiates a Hot-add flow

Device D3 is added to H2

Normal traffic flow to/from D3 as a part of H2
CXL 2.0 Defines Memory Decode Mechanism

- Supports memory expansion
- Supports Interleave across devices
- 2 way, 4 way and 8 way Interleave options supported
- Determines the Downstream VPPB to route requests to
- Responses always routed to VPPB

CXL.IO same as PCIe decode

CXL.Cache – no fanout. Software must enable Single Device with CXL.Cache in Virtual Hierarchy

Switch does Address Look-Up and forwards request/response between Downstream Port(s) and RP for CXL.IO and CXL.Mem
CXL 2.0 Fabric Manager View

FM Binds a Logical Device to VCS

Architecture allows for > 16 RPs on a switch with flexible binding to Logical Devices on a link.

PCIe or CXL Device

Type 3 Pooled

LD FFFF

LD 0

LD 1

LD 15

PCIe or CXL Device

Type 3 Pooled

LD FFFF

LD 0

LD 1

LD 15

FM API Support
Benefits of CXL 2.0 and Persistent Memory

- Moves Persistent Memory from Controller to CXL
- Enables Standardized Management of the Memory and Interface
- Supports a Wide Variety of Industry Form Factors

- CXL 2.0
  - Persistent Memory: $10^2 - 10^3$ nanoseconds

- Memory
  - CPU: $10^9$ nanoseconds
  - DRAM CXL 1.1/1.0: $10^1$ nanoseconds

- Storage
  - Performance SSD: $10^4$ nanoseconds
  - Capacity SSD: $10^5$ nanoseconds
  - HDD: $10^6$ nanoseconds

CXL + PM Fills the Gap!
CXL 2.0 Security Benefits

CXL 2.0 provides Integrity and Data Encryption of traffic across all entities (Root Complex, Switch, Device)

CPU/SoC Root Complex
- IO Bridge
- IO MMU
- Coherent Bridge

Area of Protection

- CXL.memory
- CXL.cache
- CXL.io

CXL Device
- DTLB
- Coherent Cache (Optional)

CXL 2.0 Switch

Host Memory

Host Agent

MC

Memory

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

CXL Consortium momentum continues to grow

• 130+ members and growing
• Celebrating first anniversary of incorporation – second generation specification
• Responding to industry needs and challenges

CXL 2.0 introduces new features & usage models

• Switching, pooling, fabric manager API, persistent memory, security, hot-plug
• Fully backward compatible with CXL 1.1 and 1.0
• Built in Compliance & Interop program

Call to action

• Join CXL Consortium
• Follow us on Twitter and LinkedIn for more updates!

12/10/2020

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

Please share your questions in the Question Box
Thank You