Compute Express Link™ (CXL™) Device Ecosystem and Usage Models

Flash Memory Summit 2023
Panelists

- Moderator: Kurtis Bowman, CXL Consortium MWG Co-Chair, and Director, Server System Performance at AMD
- Panelists:
  - Khurram Malik, Director of Product Marketing, CXL, Marvell
  - Timothy Pezarro, CXL Consortium member and Senior Product Manager, Microchip Technology
  - Mark Orthodoxou, VP of Marketing, Rambus
  - Kapil Sethi, Sr. Manager Product Planning, Samsung
CXL Overview

Khurram Malik
Director of Product Marketing, CXL, Marvell
This presentation contains forward-looking statements within the meaning of U.S. federal securities laws that involve risks and uncertainties. Forward-looking statements include, without limitation, any statement that may predict, forecast, indicate or imply future events or achievements. Actual events or results may differ materially from those contemplated in this presentation. Forward-looking statements are only predictions and are subject to risks, uncertainties and assumptions that are difficult to predict, including those described in the “Risk Factors” section of Marvell's Annual Reports on Form 10-K, Quarterly Reports on Form 10-Q and other documents filed by Marvell from time to time with the SEC. Forward-looking statements speak only as of the date they are made. Readers are cautioned not to put undue reliance on forward-looking statements, and no person assumes any obligation to update or revise any such forward-looking statements, whether as a result of new information, future events or otherwise.
What is CXL and why now?

Standard protocol
Memory semantics
Cache coherent
Low latency
Merged GenZ, CCIX

CXL leverages PCI Express’s ubiquity in the data center
Why does this matter to hyperscalers

Fixes scaling challenges

CPU core growth outpaces bandwidth

Maximizes utilization

DRAM key to unlocking TCO and perf for emerging workloads

Enables SKU flexibility

Pluggable, leverage and unify SSD FFs

©2023 Flash Memory Summit and Marvell. All Rights Reserved. Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
How does CXL impact memory in servers

(What is CXL trying to fix)

DRAM DIMM

- Limited performance and capacity
- Thermal / power constraints
- Not easily serviceable
- No telemetry, no acceleration
- Limited to memory supported by CPU

CXL memory module

- Greater performance and capacity
- Moved to outside for better thermals
- Easily plugged in and out (“hot swap”)
- Able to add telemetry and acceleration
- Supports greater range of memory types
CXL use cases

**Memory expansion**
- Scale performance and capacity
- Mix-and-match DRAM types
- Thermally optimized

**Memory acceleration**
- Coherent, efficient
- Accelerate analytics, ML, etc.
- Improves efficiency and TCO

**Memory pooling**
- Share memory across xPUs
- Improve utilization
- BW:cost optimized

**Switch**
- Low-latency fabric
- Supports optics
- Enables full composability
CXL vision: Optimal resource utilization

Existing
- TOR Switch
- Servers

Emerging
- TOR Switch
- Servers
- CXL appliances

Future
- TOR Switch
- Compute / acceleration (cores only)
- Memory
- I/O, storage

Rack fabric
- Ethernet

Disaggregation
- Storage and accelerators

©2023 Flash Memory Summit and Marvell. All Rights Reserved. Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
Comprehensive end-to-end CXL solutions

- Expanders
- Pooling
- Switch
- Accelerators
- Custom Compute
- DPUs / SmartNICs
- Electro-optics
- Re-timers
- SSD Controllers

CXL delivers a cutting-edge solution
Type-3 Devices

Timothy Pezarro
CXL Consortium Member
Senior Product Manager, Microchip Technology
Ecosystem Progress

- CXL™ attached memory
- Specified by CXL 1.1
- CXL 1.1+ hosts
- Several memory controller semiconductors and memory modules are in prototype
- CXL test event
  - Integrator’s list
    - https://www.computeexpresslink.org/integrators-list
- CXL test equipment
Type 3 Device AIC Implementation

- Two DDR channels
- One 64GB DDR4 RDIMM per channel
- 128GB total memory

Photo courtesy of Microchip Technology
Management & Performance

• Tools available today for use cases
  • NUMA nodes
  • Hot data
  • Interleaved

• Measured performance matches latency hierarchy prediction

<table>
<thead>
<tr>
<th>Memory Attachment Type</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>~100 ns</td>
</tr>
<tr>
<td>Single hop socket interconnect</td>
<td>~150 ns</td>
</tr>
<tr>
<td>CXL™ direct</td>
<td>~200 ns</td>
</tr>
</tbody>
</table>
Pooling and Sharing

Mark Orthodoxou
VP of Marketing, Rambus
Agenda

• Pooling vs. Sharing

• Key Concepts
  • Multi Logical Device (MLD)
  • Multi Headed Device (MHD)
  • Dynamic Capacity Device (DCD)

• Sharing and Back-Invalidate
Pooling vs. Sharing
Pooling vs. Sharing

- **Pooling**: Flexibly assigned pool of media capacity provided by any combination of switches, MLDs, and/or MHDs

- **Sharing**: Concurrent (or non-exclusive) multi-host access to same data provided via DCD framework
  - Advertised as “Shareable” – host is unable to prevent FM from sharing media after use

**2.0 – MLD and switch enabled pooling**

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
Why Pooling or Sharing?

• Pooling:
  • Provides more efficient utilization of memory resources at scale by allowing for dynamic allocation
  • Example use cases:
    • Dynamic allocation of memory resources when a hypervisor deploys a VM
    • Allocation of required memory at peak workload utilization

• Sharing:
  • Reduces aggregate memory requirements by providing multi-host access to the same data
  • Efficient data movement
  • Example use cases:
    • Clustered computing
    • Shared access to read-only kernel memory
    • VM migration

• Pooling & Sharing may exist together or separately

CXL 2.0 – MLD and switch enabled pooling

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

Pooling & Sharing
Key Concepts
Multi-Logical Device (MLD)

- Type 3 device with single CXL link
- Transactions across link carry ‘LD-ID’ to identify traffic to/from each host
- Requires a switch
  - Applies ‘LD-ID’ to traffic to device
  - Routes traffic from device to host based on ‘LD-ID’

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
Multi-Headed Device (MHD)

• Type 3 device with multiple host links (heads)
• Manages memory-to-LD allocation with differing head behavior
  • MH-SLD presents a 1-1 mapping of a single LD to each head
  • MH-MLD presents additional composability with up to 16 LDs mapped to each head
Dynamic Capacity Device (DCD)

- Allows memory capacity to change dynamically without reprogramming HDM decoders
- DCD presents its maximum capacity to each host
  - HDM decoders programmed for entire DPA range
  - DCD command set used to discover the actual memory allocation
- Fabric Manager (FM) uses the DCD command set to query and configure the DCD
- The DPA is divided into 1-8 separate regions, and each region is subdivided by the DCD into fixed-size blocks
Sharing and Back-invalidate
Back-Invalidation

- Provides cache coherent sharing of data
  - Host is notified to invalidate cache or write back modified data
- Allows multiple devices to work on same cacheline
  - Multi-host sharing for T3 devices
  - P2P UIO transactions from T1/T2 devices
- Simplifies sharing and exchange of data/control structures
  - Replaces data copies and doorbells/interrupts

*Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.*
Summary
Summary: Pooling vs. Sharing

- **Pooling**
  - Dynamic Switch Binding of SLDs
  - MHD/MLD/DCD Configured for Exclusive Host Access

- **Sharing**
  - Dynamic Allocation + Non-exclusive Multi-host Access
  - Statically Configured MHD/MLDs

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
Memory Media Device Types

Kapil Sethi
Sr. Manager New Business Planning, Samsung
• CXL Type 3 devices in the new Memory Hierarchy

  • CXL Protocol is agnostic of the underlying memory technology
    • Opportunity for volatile and persistent memory behind CXL interface
    • Byte addressable, load-store transactions
  • DRAM based CXL devices:
    • Memory Expansion
    • Memory Tiering/Pooling
  • SSD based CXL devices:
    • Large Memory Space
    • Persistent Memory

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
DRAM Based CXL Type 3 Devices

- CXL Memory Controller is a bridge/interface between the host and CXL memory.
  - CXL/PCIe (serial) interface to the host
  - DDR (parallel) interface to the DRAM
  - RAS, Management etc. features of CXL device
- DDR5 based CXL devices
  - DDR5 – up to 6.4 Gbps per IO
  - X16/x8 CXL Link ↔ 2/1 channel(s) of DDR5
- DDR4
  - DDR4 – up to 3.2 Gbps per IO
  - x8 CXL Link ↔ 2 channels of DDR4
- Other DRAM Technologies – LPDDR, HBM, GDDR, Future DRAM ??
  - Capacity, performance and cost consideration
SSD Based CXL Type 3 Devices

Memory-Semantic SSD™

1) Tiered Memory

With .mem and DRAM Cache

- Large memory space, lower TCO for future workloads which require exponential data increase
- Large memory space expansion for latency tolerant workloads
- Small granularity data access enables performance with cache hits

Compute Express Link™ and CXL™ are trademarks of the Compute Express Link Consortium.
2) Persistent Memory

With CXL GPF (Global Persistent Flush) and external battery

- Battery-backed DRAM with performance comparable to latest DRAM devices
- Persistency achieved with data dump to NAND flash
- Supports flush-on-fail with CXL 2.0 GPF (Global Persistent Flush) feature
Thank You

www.ComputeExpressLink.org