Flash Memory Summit

SOLIDIGM

Aligning Storage to emerging CXL Peer to Peer Transaction Models

Sam Bradshaw
Solidigm Pathfinding
Sam.Bradshaw@solidigm.com

Disclaimers

All product plans, roadmaps, specifications, and product descriptions are subject to change without notice.

Nothing herein is intended to create any express or implied warranty, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, or any warranty arising from course of performance, course of dealing, or usage in trade.

Your costs and results may vary.

For copies of this document, documents that are referenced within, or other Solidigm literature, please contact your Solidigm representative.

CXL: Implications for Storage

Economics of DRAM on CXL drives ecosystem uptake. Rack is the new server.

Clusters of resources carved out targeting specific workloads

- Faster, denser local storage may displace SAN
- "Close" memory enables an opportunity to cost reduce SSDs

Multi-level CXL switch topology puts some resources closer and others farther away

Tiering solutions needed

- Memory cost arbitrage for frigid tiers
- Access introspection assists

Optimized cross-switch P2P flows for performance and to decongest upstream CXL lanes

- *Unordered-IO* and *Back-Invalidate Snoop*

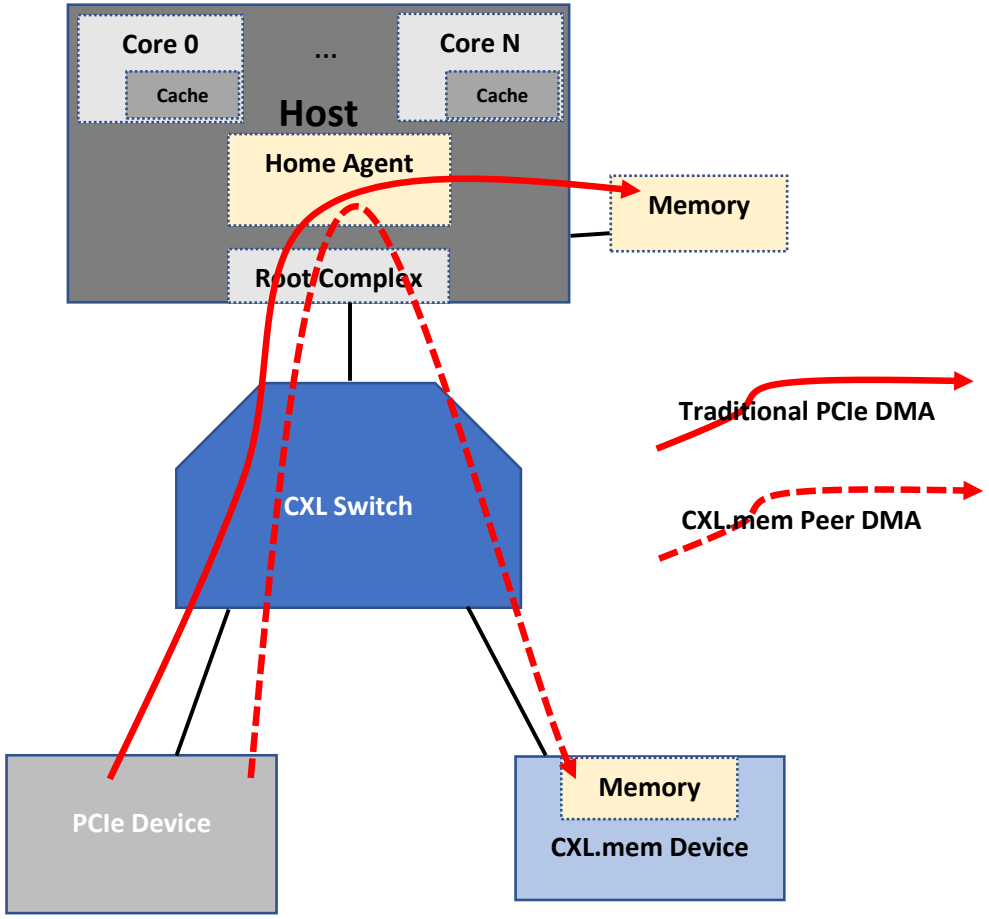
Near vs Far Memory Implications

Before CXL, coherent memory always existed "*on the other side of the root complex*"

CXL Memory is now one switch hop away from an accelerator or storage device

Because the Home Agent tracks the coherence state of peer CXL memory, DMA transactions must route through it to resolve state.

- Congests upstream CXL lanes
- Traversal latency



What's the Solution?

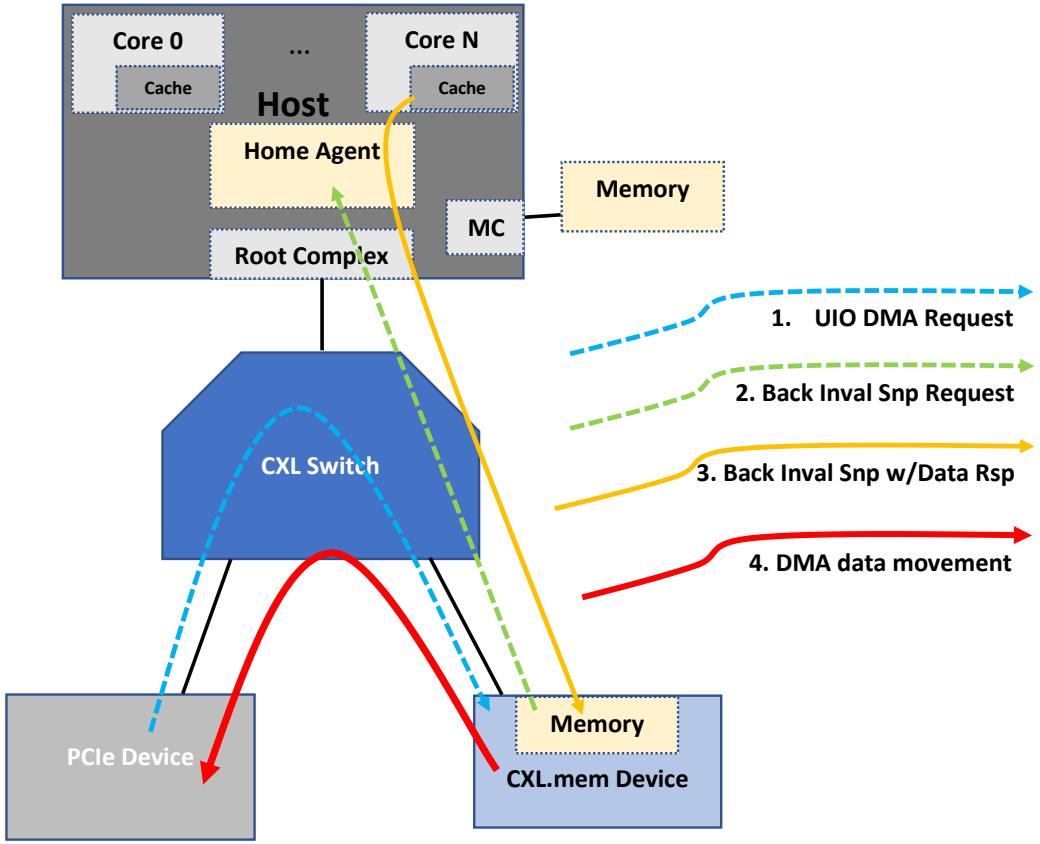
PCI-SIG has introduced a new TLP format called ***Unordered IO***

- Enables direct DMA P2P routing through switches

CXL has introduced a new request type called ***Back Invalidate Snoop***

- Enables subordinate CXL.mem devices to initiate coherence state update to itself

Now P2P DMA can produce coherent data



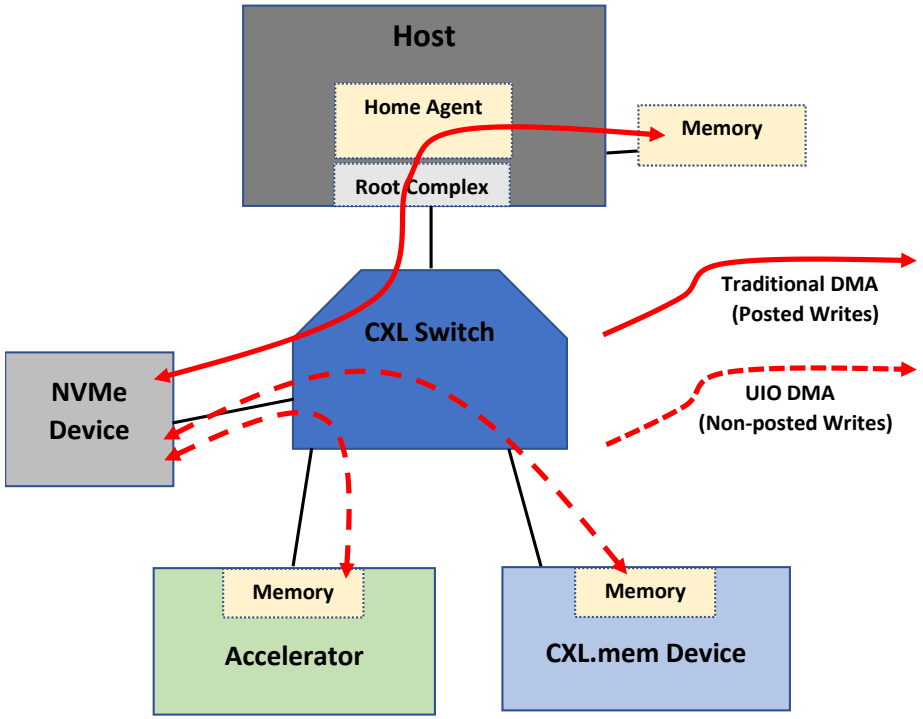
The Link to NVMe

U-IO writes are non-posted with performance penalty

NVMe device should only initiate U-IO DMA if:

1. DMA operation is exclusively targeted to address on switch peer
2. CXL switch fabric supports direct U-IO P2P routing
3. The target CXL.mem devices supports Back Invalidate

System and Protocol enablement work is required.



Call to Action

U-IO is a new PCIe Gen6 TLP format

- Call your NVMe/PCIe IP vendor for support roadmap

Protocol support for “U-IO Selectability” in NVMe

- Need consortium-wide collaboration towards TPAR

Thank you!


SOLIDIGM™