Compute Express Link 3.0

Dr. Debendra Das Sharma, Intel Fellow and Director, I/O Technology and Standards, and
Ishwar Agarwal, Principal Hardware Engineer, Microsoft Azure Hardware Architecture
Co-Chairs, CXL BoD Technical Task Force, CXL Consortium

Compute Express Link™ (CXL™) is an open industry standard interconnect offering high-bandwidth, low-latency connectivity between host processor and devices such as accelerators, memory buffers, and smart I/O devices. It is designed to address the growing high-performance computational workloads by supporting heterogeneous processing and memory systems with applications in Artificial Intelligence, Machine Learning, Analytics, Cloud Infrastructure, Cloudification of the Network and Edge, communication systems, and High-Performance Computing by enabling coherency and memory semantics on top of the PCI Express® (PCIe®) based I/O semantics for optimized performance in evolving usage models. This is increasingly important as processing data in these emerging applications requires a diverse mix of scalar, vector, matrix and spatial architectures deployed in CPU, GPU, FPGA, smart NICs, and other accelerators.

CXL 1.0 debuted in March 2019 supporting dynamic multiplexing between a rich set of protocols that includes I/O (CXL.io, based on PCIe), caching (CXL.cache), and memory (CXL.memory) semantics. CXL maintains a unified, coherent memory space between the CPU (host processor) and any memory on the attached CXL device. This allows both the CPU and device to share resources and operate on the same memory region for higher performance, reduced data-movement, and reduced software stack complexity, resulting in three primary usages as demonstrated in Figure 1.

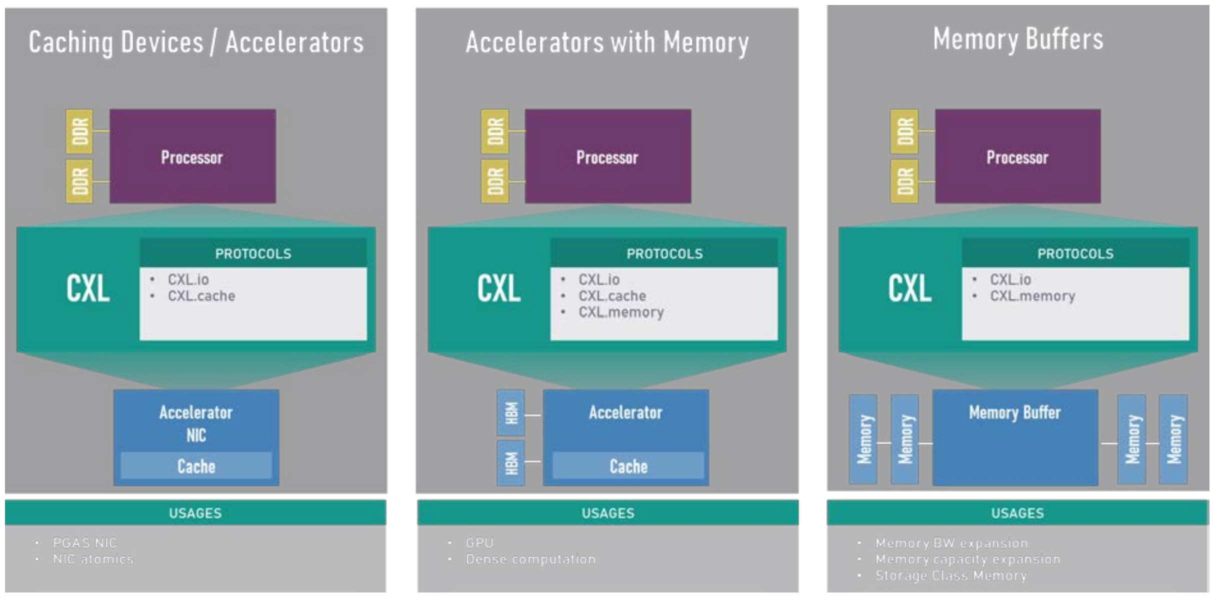
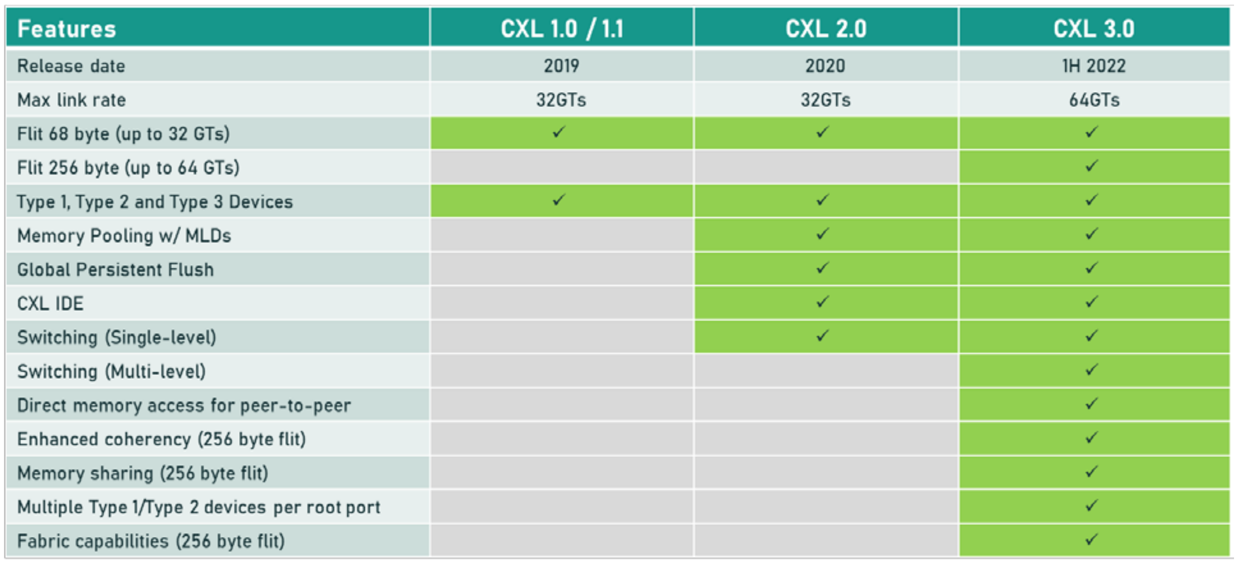


Figure 1: Representative Usage Cases enabled by CXL 1.0 and CXL 1.1

Building on the industry success and acceptance of CXL, as evidenced by the 180+ member companies' active participation, CXL 2.0 was released in November 2020 enabling additional usage models while maintaining full backward compatibility with CXL 1.1 and CXL 1.0. CXL 2.0 enhances the CXL 1.1 experience in three major areas: support for single-level switches, support for persistent memory,

memory pooling, and security. These features enable many devices in a platform to migrate to CXL, while maintaining compatibility with PCIe 5.0 and the low-latency characteristics of CXL.

Based on the resounding success of CXL 2.0 and feedback from the computer industry and the end-user community, the CXL consortium forged ahead to define the next generation of the specification to deliver even higher performance and encompass more use cases. The resulting CXL 3.0 specification is a major revision which unlocks the potential for transformative technologies and usages! Figure 2 summarizes the journey of CXL specifications with major enhancements and features.



Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0
Release date	2019	2020	1H 2022
Max link rate	32GTs	32GTs	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)			✓

Figure 2: CXL Features over Generations

CXL 3.0, which is based on PCIe 6.0 technology, *doubles* the transfer rate to 64GT/s *with no additional latency over previous generations.* This allows for aggregate raw bandwidth of up to 256GB/s for x16 width link. For low-latency transfers, CXL 3.0 leverages PCIe 6.0's combination of lightweight FEC and strong CRC for error free transmission with 256B flits on PAM-4 signaling to achieve 64GT/s. CXL 3.0, however, goes further in introducing a latency-optimized flit variant to further reduce 2-5ns of latency by breaking up the CRC in 128B sub-flit granular transfers to mitigate store-and-forward overheads in the physical layer. As with previous generations of CXL, the new 256B flit format is backward compatible with 8GT/s, 16GT/s and 32GT/s which eases the transition to CXL 3.0.

The combination of larger real-estate provided by CXL 3.0's 256B flit with doubling of signaling rate enables a number of protocol enhancements which unlocks several use-cases, a few of which we'll briefly describe next.

Enhanced Coherency – For Accelerators with Memory (refer to Figure 1), commonly known as CXL Type-2 devices, a major enhancement in CXL 3.0 is the ability to *back invalidate* the Host's caches. This model of maintaining coherency for host-managed device-attached memory (HDM) is called *enhanced coherency* and replaces Bias Based coherency introduced in previous generations. With symmetric coherency, a Type-2 device can implement a Snoop Filter for HDM address ranges which allows it to map and manage larger amounts of memory more efficiently than before. In addition, enhanced coherency semantics introduced in CXL 3.0 also enables direct peer access to HDM memory without going through host and the ability by the Type-2/ Type-3 device to back invalidate a cache line through the host processor. The latter is described in more detail in the next section.

Memory Pooling and Sharing – CXL 3.0 makes major enhancements to memory pooling which was first introduced in CXL 2.0. Memory pooling is the ability to treat CXL-attached memory as a fungible resource that can be flexibly allocated and deallocated to different servers (aka nodes or hosts) depending on the need. This enables system designers not to overprovision every server in the rack while obtaining best performance. An example of CXL 2.0 memory pooling is shown in Figure 3.

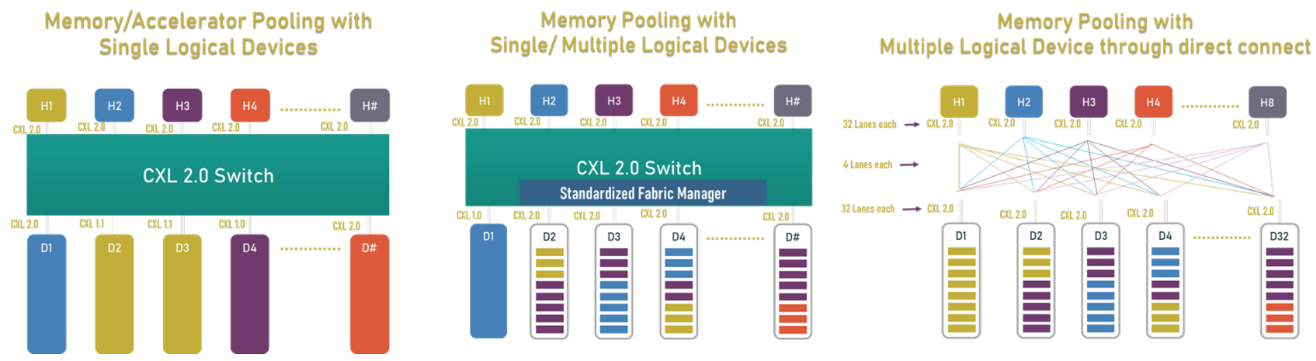

Memory/Accelerator Pooling with Single Logical Devices
CXL 2.0 Switch
Memory Pooling with Single/ Multiple Logical Devices
CXL 2.0 Switch
Standardized Fabric Manager
Memory Pooling with Multiple Logical Device through direct connect
32 Lanes each
4 Lanes each
H1 H2 H3 H4 H# H8
D1 D2 D3 D4 D# D32
CXL 2.0 CXL 1.1 CXL 1.0

Figure 3: Pooling of CXL devices with or without a CXL 2.0 switch

With CXL 3.0, in addition to memory pooling, we introduce the concept of *memory sharing*. Memory sharing is the ability of CXL-attached memory to be coherently shared across hosts using hardware coherency. Thus, unlike memory pooling, memory sharing allows the a given region of memory to be simultaneously accessible by more than one host and still guarantee that every host sees the most up to date data at that location, without the need for software-managed coordination. This allows system designs to build clusters of machines to solve large problems through shared memory constructs. An example of a shared and pooled memory topology is shown in Figure 4.

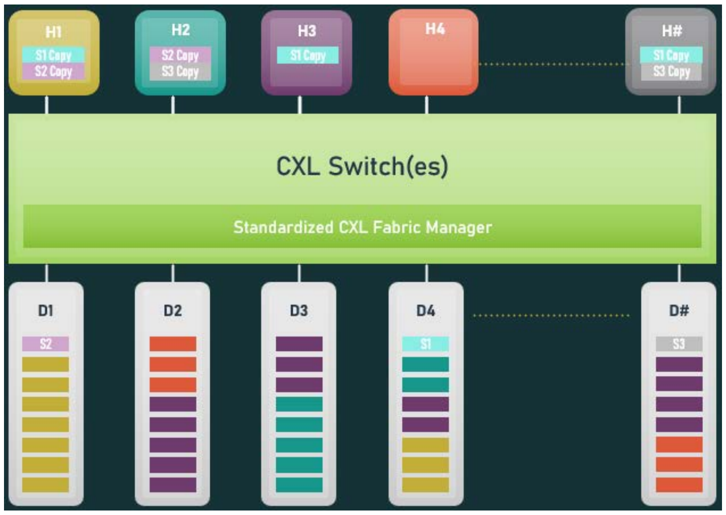


Figure 4: CXL 3.0 Memory Pooling and Sharing

Fabrics – For the first time, CXL 3.0 introduces fabric capabilities which goes beyond traditional tree-based architectural structures of PCIe and previous CXL generations. An example *non-tree topology* of a CXL fabric is shown in Figure 5.

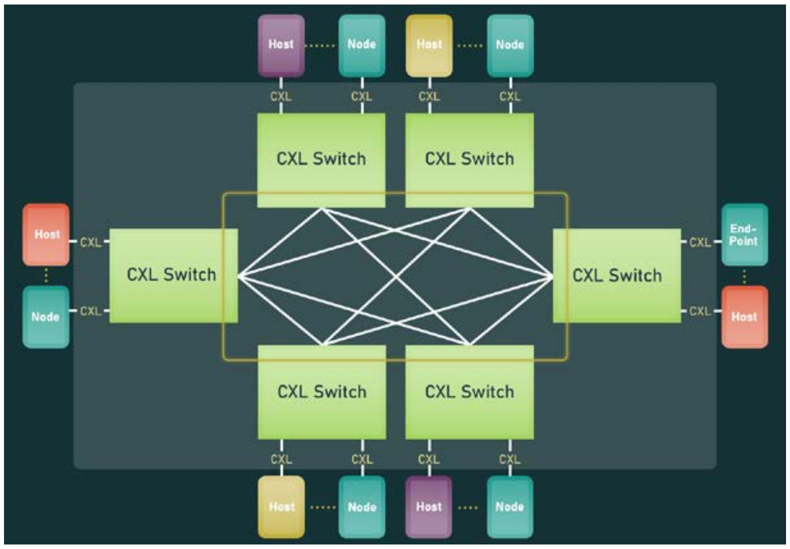
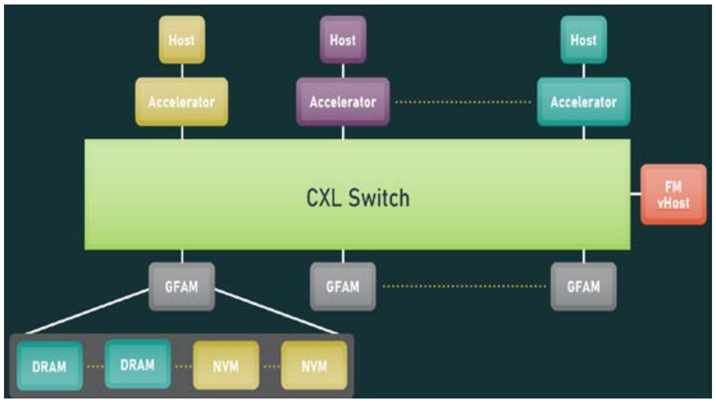


Figure 5: CXL Fabric Non-Tree Topology

The CXL fabric can support up to 4096 nodes that can communicate with each other using a new scalable addressing mechanism called *Port Based Routing (PBR)*. Here, a node can be a CPU Host, a CXL accelerator with or without memory, a PCIe device or a *Global Fabric Attached Memory (GFAM)* device. A GFAM device is like a traditional CXL Type-3 device, except it can be accessed by multiple nodes (up to 4095) in flexible ways using port-based routing. This architecture opens up a world of possibilities in constructing powerful systems comprising of compute and memory elements arranged to suffice the needs of particular workloads. A couple of example use-cases of CXL Fabrics is shown in Figure 6.



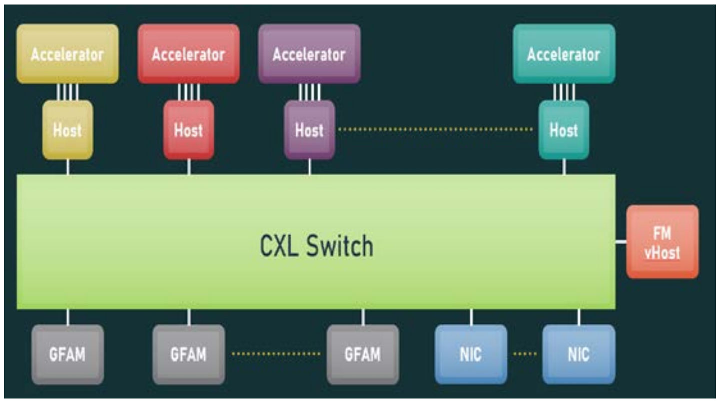


Figure 6: ML Accelerator & GFAM Device (left), HPC Analytics with Shared Memory and NIC (right)

CXL 3.0 marks a major milestone for the CXL consortium in its journey to accelerate compute, memory, and fabric architecture for the industry. With a host of exciting new features, CXL 3.0 provides fresh impetus to innovation in computer architecture. CXL will continue to encompass new requirements and feedback from the industry and deliver even higher performance and value. Please join us in this journey and make this experience even better!