Compute Express Link™ 2.0 White Paper

Dr. Debendra Das Sharma, Intel Fellow and Director, I/O Technology and Standards, and
Siamak Tavallaei, Principal Architect, Microsoft Azure Hardware Architecture
Co-Chairs, 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. It does this by enabling coherency and memory semantics on top of the PCI Express® (PCIe®) 5.0 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. Moreover, since the CPU is primarily responsible for coherency management, it can reduce device cost and complexity, as well as overhead traditionally associated with coherency across an I/O link. Introduced in July 2019, CXL 1.1 included the compliance testing details.

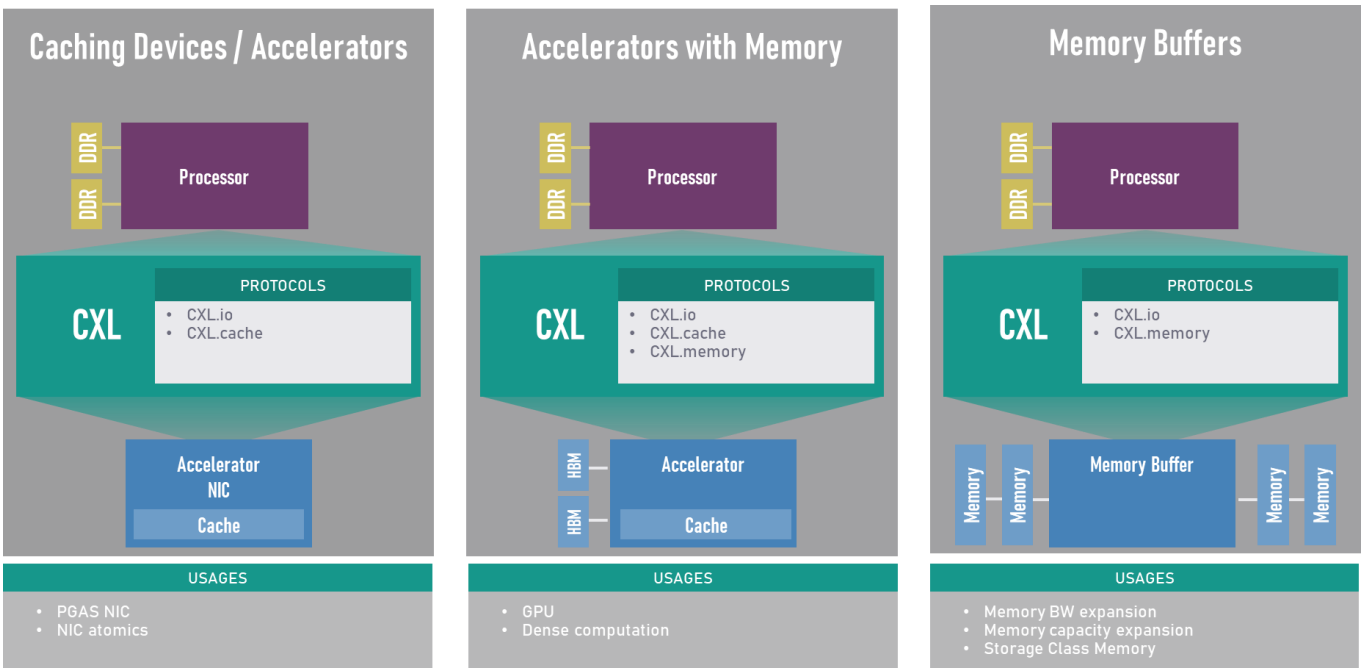

Figure 1: Representative Usage Cases enabled by CXL 1.0 and CXL 1.1

Building on the industry success and acceptance of CXL as evidenced by the 130+ member companies with active participation, we are pleased to announce the availability of CXL 2.0, about a year after CXL 1.1, 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 by introducing three major areas: CXL Switch, support for persistent memory, and security.

One of new CXL 2.0 features is the support for single level switching to enable fan-out to multiple devices as shown in Figure 2. This will enable many devices in a platform to migrate to CXL, while maintaining the backward compatibility and the low-latency characteristics of CXL.

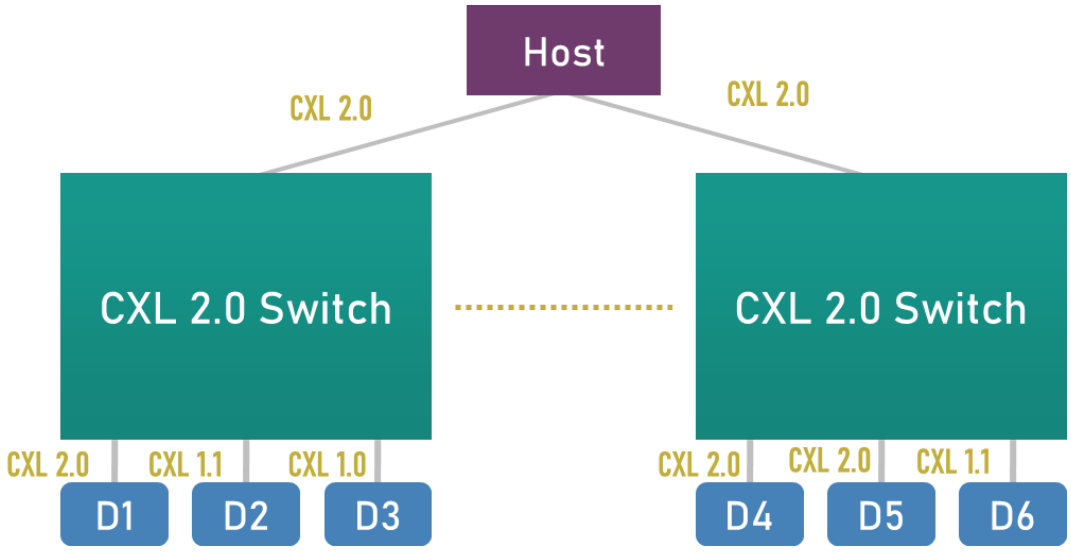

Figure 2: CXL 2.0 switch supports fan-out to multiple devices while maintaining backward compatibility

One of the important aspects of CXL 2.0 feature set is the support for pooling of multiple logical devices (MLD) as well as single logical device with the help of a CXL switch connected to several Hosts (Root Ports). This feature enables servers to pool resources such as accelerators and/or memory that can be assigned to different servers depending on the workload. Suppose a server needs two FPGAs and a GP-GPU, it can ask for those resources from the resource manager in the rack and obtain those if available and relinquish them when its job is done.

Similarly, memory 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. CXL 2.0 allows for pooling through the use of switches coupled to a Type-3 multiple logical device (MLD), as shown in Fig. 3 below. Each color in the Host (H) represents a domain or a server which defines a hierarchy.

A CXL 2.0 switch can handle multiple domains (up to 16 of such hierarchies may reach any one MLD). A Type-3 MLD device may support up to 16 domains on each of its CXL Ports. It is also possible for a Type-3 MLD device to partition its CXL resources and connect directly to multiple hosts, each with a dedicated CXL link, as shown in Fig. 3. This helps with performance identical to direct connect as the switch latency is eliminated, which is important for memory accesses.

CXL 2.0 accomplishes these goals by defining protocol enhancements capable of pooling while maintaining quality of service and reliability isolation requirements for different domains. It also defines managed hot-plug flows to add/ remove resources. Most importantly CXL 2.0 defines a standardized fabric manager to ensure that users have the same experience while pooling independent of the type of device, host, switch or the usage models they have.

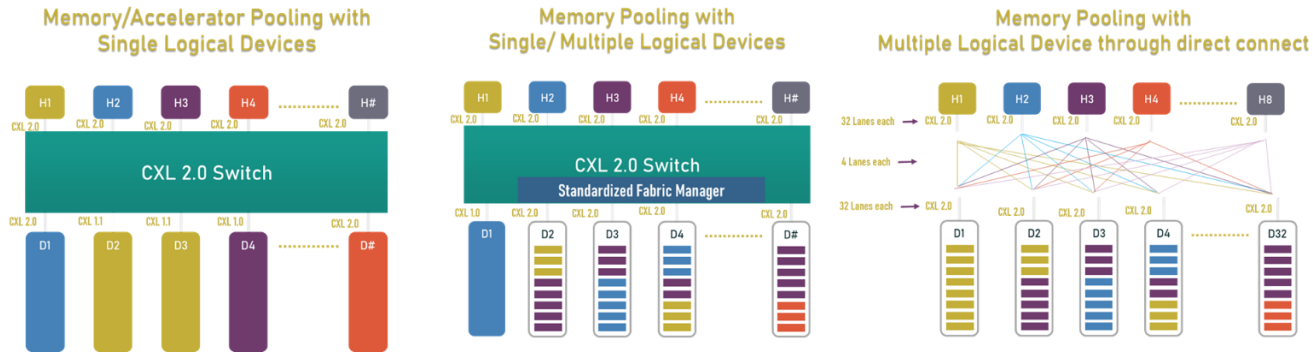

Figure 3: Pooling of CXL devices with or without a CXL 2.0 switch

One of the important innovations in the industry is around non-volatile memory, which is approaching DRAM like latency and bandwidth characteristics, while having the advantage of high capacity and persistent. This is enabling a lot of high-performance applications where the entire data set can be in memory.

One of the challenges of a load-store interconnect architecture such as CXL is guaranteeing persistence (committing a data store into persistent memory). CXL 2.0 meets that challenge through architected flow and standard memory management interface for software, enabling moving the persistent memory from a controller-based approach to direct memory management (Figure 4).

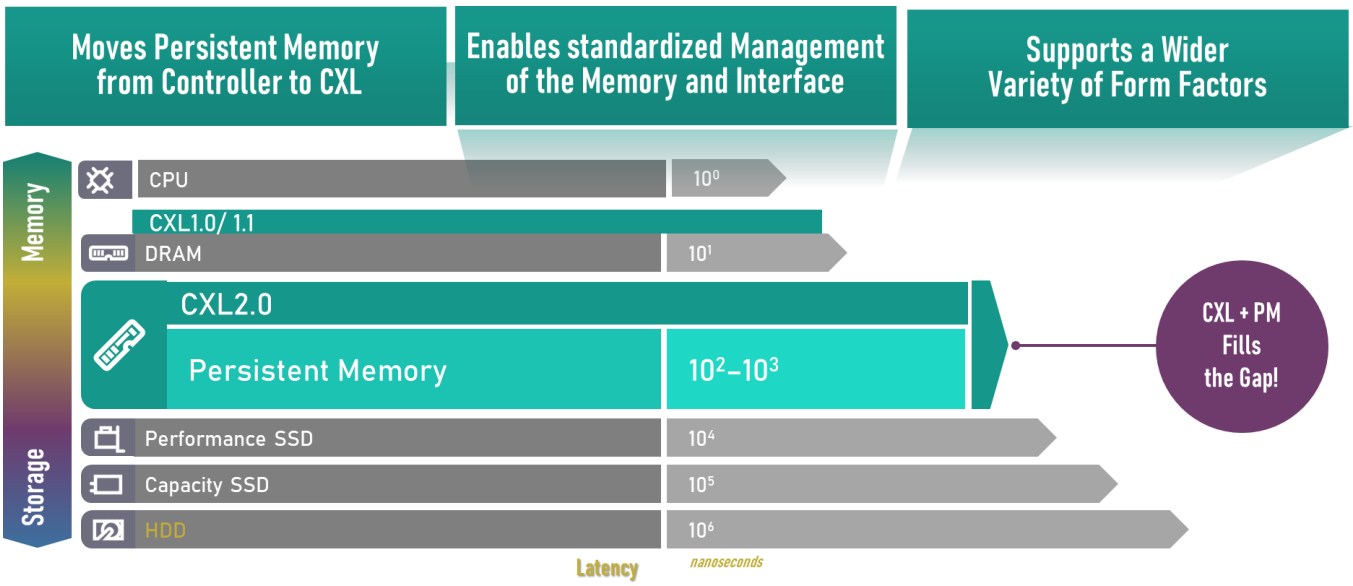

Figure 4: CXL 2.0 addresses persistence in Persistent Memory

Security is a key cornerstone for any technology to be successful, considering the vulnerability attacks that are so pervasive. CXL is making great strides in this regard, working collaboratively with other

industry-standard bodies such as PCI-SIG and DMTF to ensure that we have a seamless user experience while providing the best security mechanisms. CXL 2.0 enables encryption on the Link that works seamlessly with existing security mechanisms such as device TLB, as shown in Figure 5.

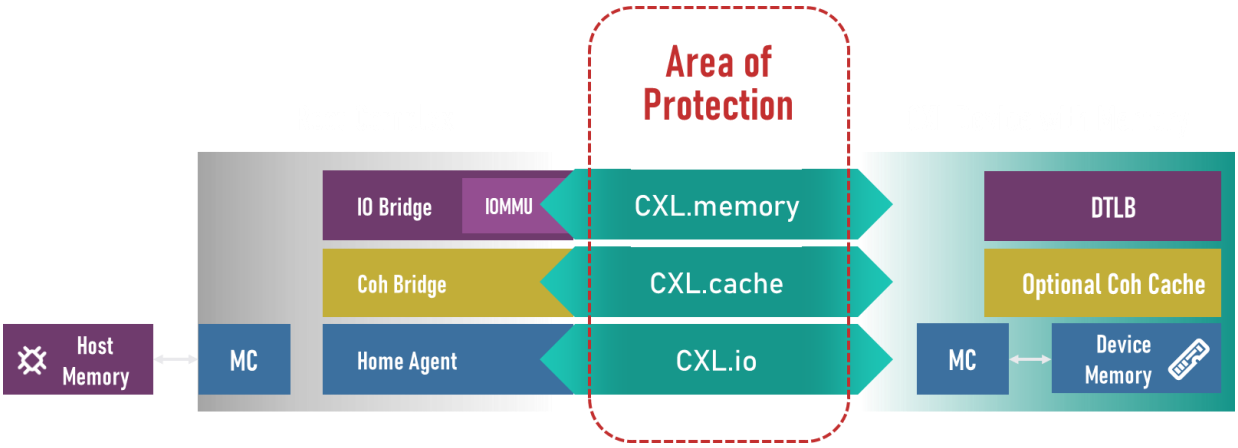

Figure 5: Security enhancements with CXL 2.0

Having published two versions of the specifications in one and a half years, the CXL Consortium is forging ahead beyond CXL 2.0. Based on the feedback from the computer industry and the end-user community, we are working toward the next revision of CXL 3.0 to encompass more useful scenarios and deliver even higher performance. Please join us in this exciting journey and make this experience even better!