CXL: Enabling A Heterogenous, Composable, Next-Generation Data Center

Summary

Data has become one of the most valuable and, at the same time, abundant resources in the world. More data has been created in the past three years than in the past two decades, and the biggest challenge has been to process this data into insight.

The straightforward solution to processing this data within the data center has been to focus on compute and increased central processing unit (CPU) core counts. In system architecture, however, memory bandwidth and speed to support these cores often failed to keep pace, which created a bottleneck in data processing.

Access to storage and memory bandwidth has become the limiting factor in the data center not only because of speed and bandwidth but because of the growth of cores in the CPU/GPU. HC is the use of different semiconductor core architectures in the same system (X86, GPU, ARM, FPGA, etc.)—sometimes even on the same die—to accelerate the processing of different types of data. Different core architectures have different memory and storage requirements.

While heterogeneous compute (HC) may offer improved performance advantages over traditional homogeneous computing for specific workloads, it adds a level of complexity to the system because of the current architecture's inability to detach memory from its respective CPU. For instance, ML training GPUs require on-board memory.

In today's world, as well as tomorrow's, memory needs to be independent of the CPU, composable, and flexible, like storage subsystems. Software-defined composable memory enables data center managers to scale memory as needed with other components of the systems, rather than simply being tied to the CPU.

Memory and storage innovation is pivotal to the next generation of data centers and data management. The Compute Express Link (CXL) interface serves as a great example of the kind of innovation the industry needs. This paper looks at how CXL can help enable of the next-generation data center.

WHAT IS CXL?

The CXL Consortium defines CXL succinctly as an "industry-supported cache-coherent interconnect for processors, memory expansions and accelerators." Cache coherence ensures shared resource data remains uniform while stored in multiple local memory cache locations.

Memory coherency is what allows CXL-enabled devices to be fully composable. CXL also has switching, enabling multiple host CPUs to switch between devices and resources. Switching and memory coherency enable data centers and hyperscalers (mostly large public cloud providers like Amazon Web Services (AWS), Microsoft's Azure, Google Cloud and Oracle Cloud) to offer flexible systems based on application and workload. Rather than paying for the max amount of resources a system may need for a workload, a data center could share resources between systems, unlocking significant performance at a more favorable total cost of ownership (TCO).

TABLE 1: CXL CONFIGURATIONS

Configuration	Protocol		Connectivity	Use case
# 1	CXL.io, CXL.cache		Connects the devices to the host processor's memory/cache.	Caching, accelerators without memory
#2	CXL.io, CXL.mem		Connects the host processor to the devices' memory/cache.	Memory buffer or expander
#3	CXL.io, CXL.mem, CXL.cache		Connects the devices to the host processor's memory/cache and connects the host processor to the devices' memory/cache.	Accelerators with memory

(Source: Moor Insights & Strategy)

CXL has three configurations to allow the host CPU to share memory between its devices and vice versa.

- The first allows HC devices to access the host CPU's memory.
- The second allows for the host CPU to access the memory of HC devices.
- The third allows for sharing HC devices and the host CPU between each other (both configurations #1 and #2).

CXL vs PCIe

The biggest difference between CXL and PCIe, another high-speed computer standard, is that CXL has three protocols while PCIe 5.0 has only one. CXL.cache and CXL.mem are the two protocols that allow the host processor and its devices to connect to their respective memory. CXL does not replace the PCIe 5.0 interconnect but rather builds on it. The CXL.io protocol is needed in all three types of configurations and based on PCIe PHY. Since PCIe is a fundamental part of CXL, Moor Insights & Strategy does not see CXL moving away from PCIe, but rather, as PCIe improves, inherently improving with it.

Demystifying the different CXL specifications

The CXL Consortium introduced three specifications within the past three years. As the CXL Consortium continues to add new features and capabilities, it is important to understand the difference between CXL 1.0/1.1, 2.0, and the latest release, 3.0.

Table 2: CXL Standards and Release Dates

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)			✓

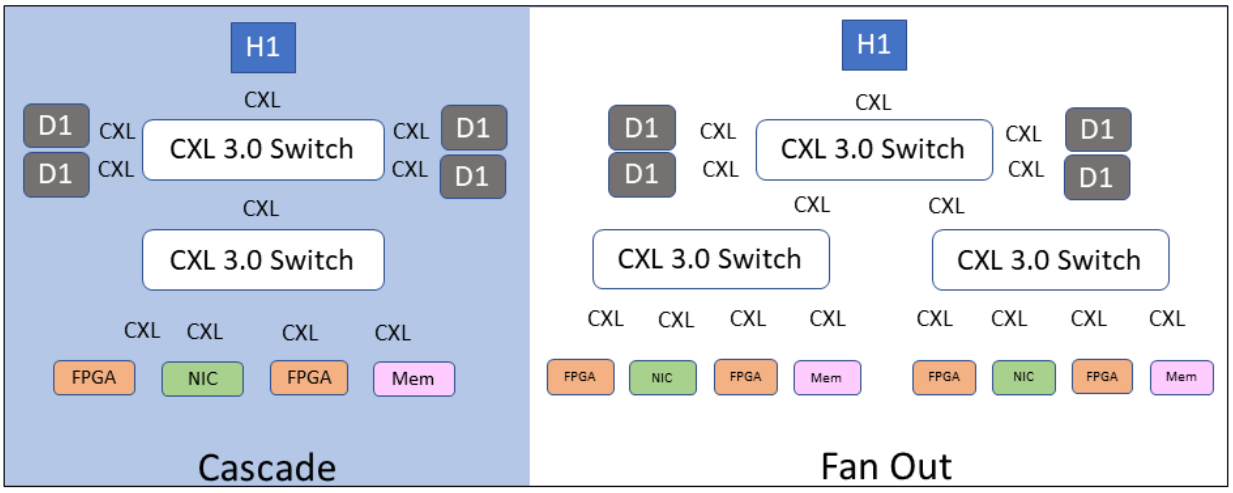
(Source: CXL Consortium)

CXL 1.0 introduced its multiple protocols on top of the PCIe 5.0 standard, allowing for coherent memory across the host processors and other devices. Devices with CXL 1.0 get the speeds and performance of PCIe 5.0 with the added benefit of memory coherency within the system. CXL 1.1 introduced compliance testing details for reliability, availability, and serviceability (RAS) and backward compatibility with CXL 1.0.

Memory coherency, low latency, and high speeds are significant improvements within the data center and give room for cloud service providers (CSPs) and large enterprises to create completely heterogeneous systems. While the goal is to create fully composable compute systems and memory pooling, not every data center component can be replaced at once. CXL 1.1 lays the foundation and focuses on accelerators and memory expansion use cases.

CXL 2.0 introduced switching, memory pooling, and more RAS within the standard. Switching allows multiple host CPUs to share and allocate devices depending on the workload, creating fully composable data centers. It allows for memory pooling where host CPUs are connected to a high-capacity memory device and allocate only the memory needed per workload, keeping the system from over-provisioning. The same is true for data accelerators, SmartNICs, and other composable devices. It also adds link-level integrity and data encryption so traffic on the CXL link has confidentiality, integrity, and replay protection.

FIGURE 1: CXL 3.0 MULTI-LEVEL SWITCHING CONFIGURATIONS



Host (H1) in blue, devices (D1) in gray.

(Source: Moor Insights & Strategy)

Released in August 2022, CXL 3.0 builds on the scalability and composability of CXL 2.0. It is based on PCIe 6.0, which doubles the transfer rate to 64GT/s without increasing latency. The CXL 3.0 standard has multi-level switching, allowing systems to level CXL 3.0 switches in either a cascading or fanout design. The CXL 3.0 standard also allows dynamic capacity devices, so that a switch could consist of different devices with different protocol types, as seen in Figure 1.

CXL 3.0 builds on CXL 2.0's memory pooling feature with memory sharing. Memory sharing allows more than one host to access a common section of memory simultaneously, enabling multiple machines to store, compute, and move data coherently.

CXL 3.0 also introduces CXL fabric, allowing devices to go beyond the tree-based topology. It uses port-based routing (PBR) to connect up to 4,096 nodes, which could be a CPU Host, an accelerator like a GPU, a PCIe device, or a Global Fabric Attached Memory (GFAM) device. GFAM devices could enable hardware disaggregation and could consist of different types of memory, even volatile and non-volatile memory.

Alongside these fabric non-tree topologies, CXL 3.0 enables peer-to-peer communication, allowing devices to communicate without returning to the shared host. CXL 3.0 standard is also backward compatible with CXL 2.0 and CXL 1.0 standards, which allows industries to continue innovating on top of existing technologies.

CXL is not the first protocol to have many of these features and capabilities. What makes the CXL protocol disruptive is that it makes these features and capabilities industry standard, agreed upon by CXL Consortium members and industry leaders.

The CXL standard is re-architecting the way data centers and cloud infrastructure handle data through memory and storage. With memory sharing and GFAM devices, memory—once an isolated subsystem dependent upon a host CPU—is now as dynamic and moveable as the data that moves through it. In the data center, memory has equal footing with processing and storage.

MEMORY IS NO LONGER THE LIMITER

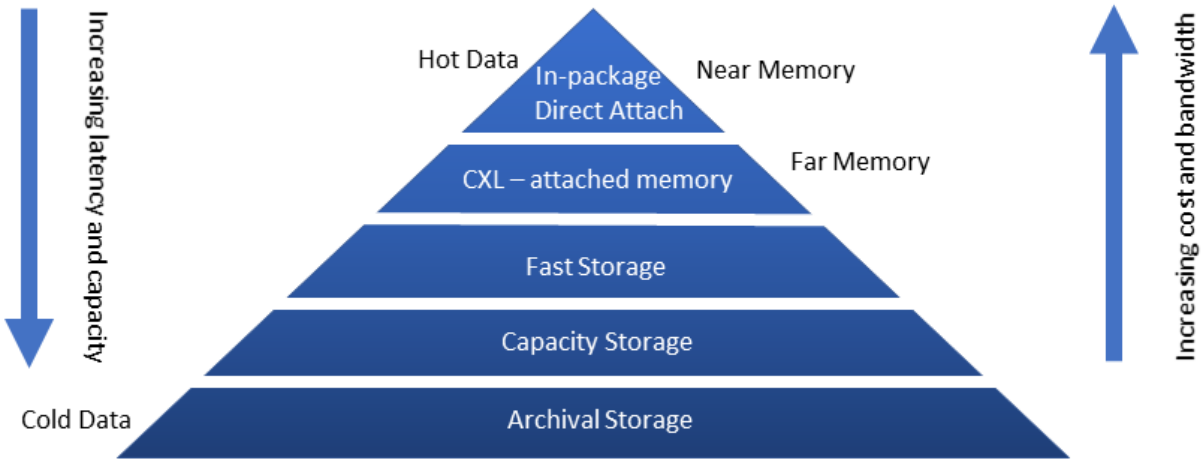
While it is apparent that next-generation HC and composable data centers may significantly affect how we process data at scale, CXL could also disrupt the memory and storage hierarchy pyramid. For the past three decades, storage and memory hierarchies have remained largely unchanged. Data centers usually store and archive data in storage, which has higher capacity but increasing latency the farther it must travel. Inversely, the higher you travel up the memory and storage hierarchy, memory has lower latency and increasing bandwidth. Then you have persistent memory and everything in between.

CXL adds a layer of far memory between near memory, traditional direct-attached memory, and fast storage. This far memory is slower and has higher latency but also higher capacity than near memory. Yet, it has higher bandwidth and faster speeds than

fast storage. Compared to memory stacking, it is more affordable due to its memory pooling capabilities.

Far memory's latency is minute, considering the capacity benefits it provides. The alternative to far memory would be stacking near memory and increasing density, which drives the non-linear cost per bit. This middle ground of far memory addresses diverse workloads that require a balance of performance and capacity. For data processing and AI and ML workloads, data is accessed more efficiently in hot data components.

FIGURE 2: MEMORY AND STORAGE HEIRACHY



(Source: Moor Insights & Strategy)

CXL enables platform architecture freedom, where a CXL-enabled system with a scale-out architecture has the memory benefits of a scale-up architecture. A scale-up architecture is defined by larger systems with high memory capacity and performance. Traditionally, data centers with scale-out architecture use many smaller machines and either span the memory with high-latency interfaces or break up the application into smaller bits. With CXL, devices are composable and can build optimized infrastructure while having the flexibility and scalability of a scale-out architecture. This flexibility creates new value propositions within the data center, depending on the workload and application. It's both a scale-up and scale-out data center.

Another benefit of CXL is its increased bandwidth—because CXL is built on PCIe, it offers high bandwidth of up to 64GT/s with no additional latency. This is important to hyperscalers. When memory and cache between the host CPU and HC devices are coherent and uniform, they are freed up within the system.

CXL IS AN OPEN, INDUSTRY STANDARD

The CXL Consortium is a successful open industry standards group with over 300 members, including every major company in the memory industry. While multiple leaders are pushing for the CXL standard within their respective industries, Micron is a leader in memory and storage, pushing CXL as the center of the next-generation data center.

FIGURE 3: CXL CONSORTIUM BOARD OF DIRECTORS

(Source: Moor Insights & Strategy)

Micron enables the ecosystem through its technology leadership and helps drive standards with JEDEC (Joint Electron Device Engineering Council), CXL Consortium, OCP (Open Compute Project), and NVMe (Non-Volatile Memory Express). Micron clearly sees promise in CXL to alleviate the limitations memory bestows upon homogeneous computing data centers and enable the next generation of heterogeneous computing data centers.

WRAPPING UP

We believe the next-generation data center will incorporate much more heterogeneous computing and will be composable—with resources defined in software. Turning data into insight puts memory and storage at the forefront of data center innovation. CXL is *the* standard that enables memory coherency, on-demand allocation of resources, and a slew of other benefits for diverse workloads and applications.

CSPs and large enterprises should thoroughly consider speedy CXL adoption within the data center value chain as they strategically plan the next generation of data centers. CXL could potentially transform the data center as we know it into on-demand computing, with memory at the core of its capabilities.

With its high-capacity memory allocation, increased bandwidth, and favorable TCO, data centers incorporating CXL should have no problem keeping future, data-specific workloads at the top of the memory and storage hierarchy. CXL could fundamentally change memory's role, from being the limiter within the data center to an enabler of heterogeneous and composable data centers. To learn more about memory and storage in the data center, visit micron.com/datacenter.

IMPORTANT INFORMATION ABOUT THIS PAPER

CONTRIBUTORS

Patrick Moorhead, CEO, Founder and Chief Analyst
Jacob Freyman, Contributor

PUBLISHER

Patrick Moorhead, CEO, Founder and Chief Analyst at Moor Insights & Strategy

INQUIRIES

Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING

DISCLOSURES

Micron Technology, Inc. commissioned this paper. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.