Technology Brief

Data Center | CXL™ and Intel® Optane™ Persistent Memory

Migration from Direct-Attached Intel® Optane™ Persistent Memory to CXL™-Attached Memory

Intel Optane persistent memory and Intel® Xeon® Scalable processors offer a practical migration path to memory expansion, tiering, and pooling with Compute Express Link (CXL)-attached memory devices

Contents

Executive Summary

Enterprises are faced with ever-increasing data. They need a memory solution that balances total cost of ownership (TCO) and performance. Also, data centers are increasingly running heterogeneous workloads—including artificial intelligence (AI) and high-performance computing—that have varied infrastructure, performance, and availability requirements. As a result, the concept of "system memory" is transforming. The industry is turning to memory expansion, tiering, and pooling to optimize cost, capacity, and performance. Intel developed and productized Intel® Optane™ persistent memory (PMem) to accelerate adoption of these new usage models of memory.

In 2019, the Compute Express Link™ (CXL™) Consortium was formed, of which Intel is a founding member, to develop an open standard for next-generation memory capabilities. Intel donated the code that formed the basis of the CXL 1.0 specification published in March of 2019. Subsequent versions were published in June 2019 (1.1) and in November of 2020 (2.0). It takes several years to develop products and usage models around I/O standards. CXL usage models and software solutions are comparable to the usage models and software that Intel Optane PMem supports. As the CXL 2.0 ecosystem matures over the next two to three years, enterprises can begin the transition to CXL-based memory solutions that support their usage models. Therefore, enterprises can confidently use Intel Optane PMem today to affordably meet memory expansion and tiering needs, and then migrate to CXL-based solutions when they are proven in the marketplace.

This technology brief provides information on how enterprises can achieve a low infrastructure TCO by incorporating Intel Optane PMem into server deployments today; benefit from hierarchical memory expansion, tiering, and pooling; and add CXL-capable servers to the data center once they become available. Intel's long-term collaboration with hardware and software developers will help customers to migrate from today's Intel Optane PMem-based data and memory tiering solutions to tomorrow's usage models with CXL devices.

Key Points

- The IT industry desperately needs cost-effective memory expansion, tiering, and pooling to keep up with data and compute growth as well as heterogeneous workloads.
- Companies can continuously modernize their data centers and scale up workloads with advantageous total cost of ownership (TCO) savings with Intel® Optane™ persistent memory (PMem) today and add servers with CXL™ capabilities in the future.

Surmounting Memory Capacity and Latency Limitations

Massive data growth is fundamentally changing data center workloads and infrastructure. To accommodate workloads like real-time analytics at the edge, cloudification of the network, machine learning, artificial intelligence (AI), and high-performance computing, enterprises want lower latency (i.e., faster data access rates) and access to larger amounts of data than ever before. Heterogeneous workloads that use different types of processors and have varying requirements for performance and availability are driving data center infrastructure and server architecture changes—especially for system memory.

The traditional approach of simply buying more DRAM doesn't work. First, as processor core counts grow, there are not enough dedicated DRAM slots to scale accessible memory sufficiently. Unfortunately, increases in DRAM density have slowed from 4x every three years to only 2x every four years,[1] elevating memory limitations to a critical architecture and compute issue.

Today, Intel® Optane™ persistent memory (PMem) provides additional capacity in those slots that DRAM can't. Future offerings of Compute Express Link™ (CXL™) can provide additional slots beyond the dedicated DDR slots to attach memory.

Second, installing and accessing more DRAM via higher-capacity DDR5 slots is often cost-prohibitive.

Together, Intel Optane PMem and CXL Solve Memory Challenges Now and Tomorrow

Intel recognized the performance and capacity gaps between DRAM and storage, and began pursuing solutions to address the issue. In 2015, the development of Intel 3D XPoint technology was announced, which was the foundation of Intel Optane technology. By 2017, the first Intel Optane SSD was commercially available, followed by Intel Optane PMem in 2019—the same year that Intel donated the CXL 1.0 code and helped form the CXL Consortium. This product offering helped to address pressing memory cost and storage performance challenges, while at the same time enabling the pursuit of a much broader, long-term solution by helping define and mature the open industry CXL standard.

Providing Business Value Today and in the Near Future with Intel Optane Technology

While the CXL standard and ecosystem mature over the next few years, Intel customers can help solve their system memory limitations by deploying Intel Optane PMem in their data centers today. Here are several compelling reasons to consider this approach:

- **Accelerate time to results and insights.** Organizations can modernize their infrastructure and achieve data insights faster for workloads like analytics, AI, and real-time processing. Adding Intel Optane PMem to DRAM reduces latencies because more data can be stored in memory and doesn't need to be fetched from storage. Real-time analytics and AI workloads require processing of large datasets. By deploying Intel Optane PMem with 3rd Gen Intel® Xeon® processors, you can double the performance of these workloads.[2]
- **Reduce memory and storage infrastructure costs.** Increase infrastructure agility, optimize resource utilization, and consolidate servers to help reduce costs in the data center. For example, you can double your memory capacity per server and lower your costs per VM by up to 25% with Intel Optane PMem and SSDs.[3]
- **Rely on a trusted ecosystem.** Intel Optane technology is already helping accelerate the pace of transformation across many industries, including finance, telco, retail, healthcare, transportation, and gaming. Intel's broad partner ecosystem includes Dell Technologies, Oracle, HPE, Cisco, VMware, and Nutanix, as well as major cloud service providers like Microsoft Azure, Google Cloud, Amazon Web Services, Meta, and more. These successful relationships help optimize, validate, and deploy Intel Optane technology for customers who want to modernize their data centers.

For more information about the business benefits of Intel Optane PMem, read the Intel® Optane™ Persistent Memory 200 Series Brief.

Paving the Way to Future Memory Capabilities: The CXL Standard

Intel has long recognized the need for memory expansion and tiering beyond Intel Optane PMem's capabilities. Intel, along with its customers, need these capabilities because they are fundamental to capturing total cost of ownership (TCO) savings in a distributed compute world. Thus, the push for memory pooling support in CXL 2.0 and memory sharing in CXL 3.0 continues.

Intel has a lengthy history of initiating open industry standards (see Figure 1). One example is the now-taken-for-granted floating-point computing standard (more specifically, the IEEE Standard 754 for Binary Floating-point Arithmetic). In the storage realm, Intel provided thought leadership first for the SATA standard, then later for innovations like the PCIe and NVMe standards. Intel has also worked with the ecosystem to establish connectivity standards, like the USB specification, 802.11, and Ethernet. More recently, Intel has jumpstarted the Universal Chiplet Interconnect Express (UCIe) standard.

Intel is a board member of the CXL Consortium and has contributed to each new release of the CXL specification. CXL is now widely supported in the industry.

To accelerate the maturation of the ecosystem, Intel is working with leading technology companies to test engineering samples of CXL 2.0 memory devices. However, it will take time for the ecosystem to catch up, even to the CXL 2.0 specification. In the meantime, enterprises can use Intel Optane PMem as a bridge to CXL.

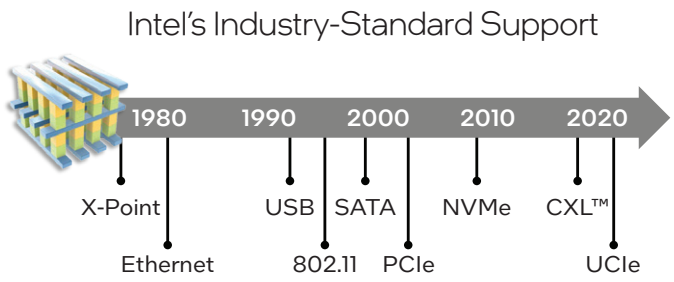

Figure 1. Intel's long history of supporting the development of industry standards has helped transform the IT industry over the years, leading to new solutions and sustained business value.

A Closer Look at the Evolution of CXL[4]

CXL is a new processor-to-device link that runs on top of existing PCIe physical connections. CXL provides a memory-coherent connection between the host and the device, allowing for lower latency of data transfers without involving OS and direct memory access setups, which are necessary for a pure PCIe-based connection. This new standard delivers a more efficient interconnect between the CPU and the traditional I/O interface, which helps solve data center scaling challenges in terms of latency, bandwidth, and capacity. The standard is optimized for high transfer rates and low latency. It is specifically designed to handle trends like increasingly heterogeneous computing, massive datasets, and demanding workloads. A CXL-attached device can either be an accelerator or a memory buffer.

The following benefits are the result of the memory-coherent links offered by CXL:

- **Efficient resource sharing** (such as between the CPU and an accelerator device) that can improve data handling and reduce I/O bottlenecks while contributing to TCO savings.
- **Memory bandwidth and capacity expansion,** with efficient access across shared memory.

The development of CXL can be viewed as a crawl-walk-run scenario:

- CXL 1.0 and the first productized version, 1.1, were released in 2019. These versions introduced and defined the CXL I/O protocol, memory protocol, and coherency interface, focusing on a single host with a single device or with multiple devices.
- The CXL 2.0 specification, released in 2021, includes support for switching for fan-out to connect to more devices; provides memory pooling for increased memory utilization efficiency and memory capacity on demand; and adds support for persistent memory. With standardized management support, CXL 2.0 switches and interfaces make pools of memory or storage available system-wide instead of only to a single host. Support for pooling multiple memory devices enables servers to assign and relinquish resources based on workload needs. In addition, the 2.0 specification removed the necessity of a shared DDR bus for DRAM and persistent memory, which helps resolve performance issues. Lastly, CXL 2.0 is fully backward compatible with CXL 1.1 and 1.0.
- The CXL 3.0 specification was completed August of 2022, introducing memory-centric fabric architectures and expanded capabilities. These new capabilities include improved memory pooling and sharing, enhanced coherency, and efficient peer-to-peer communication. This specification doubles the data rate to 64 gigatransfers per second. CXL 3.0 is fully backward compatible with CXL 2.0 and 1.1.[5]

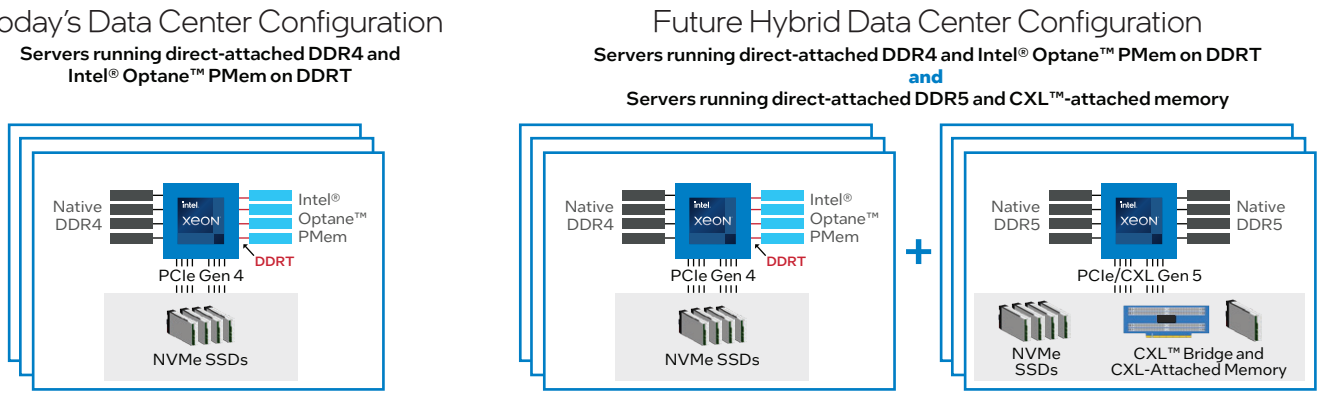

Figure 2. Sample data center migration from Intel® Optane™ PMem to CXL™.

Migration Path from Intel Optane PMem to CXL

Intel Optane PMem provides a proven, solid foundation for memory expansion and tiering. Enterprises can deploy innovative services and products that require high-performance analytics using an established, trusted technology supported by advanced hardware-based security. Enterprises can continue to experience the performance and TCO benefits of Intel Optane technology over the next few years, and can gradually migrate to CXL as production-grade products become available in the future. The recommended migration path is as follows (see Figure 2):

- **Save money today**: Replace DRAM-only servers with servers equipped with Intel Optane PMem in addition to DRAM.
- **Expand into the future**: Add newer servers with CXL-attached memory as they become available.

Servers with Intel Optane PMem and servers with CXL memory solutions can coexist in the data center, maximizing infrastructure ROI. The migration over time allows businesses to take advantage of the commercial and computing benefits of memory and storage tiering today, while anticipating open-standard, software-based solutions in the future.

Conclusion

Intel has established a track record of identifying customer pain points early, and then spearheading both products and standards to build an ecosystem that can address market needs. Intel continues to work with the ecosystem so that the CXL standard and programming models enable a smooth migration for the ecosystem from today's Intel Optane PMem to tomorrow's CXL-attached memory products. The hardware and electrical interfaces are changing, but the memory tiering programming model remains consistent. The business and computing challenges that have existed for more than a decade and are answered by Intel Optane PMem can remain resolved with Intel Optane technology deployments until CXL-based memory solutions become commercially viable.

Learn More

You may also find the following resources helpful:

- Intel® Optane™ persistent memory
- Intel® Xeon® Scalable processors
- CXL™ Consortium
- CXL™ Consortium Pressroom

Find the solution that is right for your organization. Contact your Intel representative or visit **CXL Consortium website** and **intel.com/optane**.

intel®

[1] Source: "DRAM scaling: 3D NAND Technology – Implications for Enterprise Storage Applications," by J. Yoon, IBM; 2015 Flash Memory Summit.
[2] Source: Claim 4 at "Intel® Optane™ PMem 200 Series – Performance Index." edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-optane-persistent-memory-200-series.
[3] Source: Claim 3 at "Intel® Optane™ PMem 200 Series – Performance Index." edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-optane-persistent-memory-200-series.
[4] Source for CXL 1.1 and 2.0: "CXL 1.1 vs. CXL 2.0 – What's the difference?" video, published by the CXL Consortium, June 2022 (youtube.com/watch?v=w2cZn-Tzrmo).
[5] Source for CXL 3.0: "Introducing Compute Express Link (CXL) 3.0" video, published by the CXL Consortium, August 2022 (youtube.com/watch?v=_HWWh7cBMq4).