Introduction

Data center workload capacity challenges can be handled by deploying additional Intel® Xeon® processors, by parallelizing workloads across multiple CPU cores, and by using hardware accelerators such as custom GPGPUs, ASICs, and FPGAs. Intel FPGA workload accelerators complement Intel Xeon processors in the data center to help meet varied performance needs across diverse workloads. FPGAs are often the accelerator of choice due to their reprogrammable nature, their relatively low power consumption, and their cost effectiveness.

Until now, FPGA accelerators deployed in data centers have mostly conformed to a nearly universal two-tiered external memory architecture. This traditional two-tiered external memory architecture includes main memory (Tier 1) and storage (Tier 2), usually implemented with DRAM for Tier 1 and hard disk drives (HDDs) and/or solid-state drives (SSDs) for Tier 2, respectively.

Currently, Tier 1 memory generally employs low-latency, high-bandwidth double-data-rate (DDR) SDRAMs, but DRAMs increasingly face density and cost challenges as server memory demands grow. Packing more memory into a server that has only a limited number of DIMM slots requires denser DRAMs. However, denser DRAMs present another challenge: higher-density DRAMs cost more per bit and DRAM cost represents a significant portion of a server's overall cost.

Consequently, it’s not economically or physically practical to achieve large Tier 1 memory capacities using DRAMs, which means that most of the data in a data center usually resides in Tier 2 storage until needed. If the server’s CPU requests data that is not in its Tier 1 memory, that data must be fetched from Tier 2 storage, which incurs a significant data-access delay because the latency performance gap between Tier 1 main memory and Tier 2 storage is very large. Tier 1 main memory latency is measured in tens of nanoseconds and Tier 2 storage latency is measured in tens of microseconds or milliseconds. That's at least a 1,000X latency difference.

For typical server configurations, Tier 1 main-memory densities are measured in gigabytes (GB) and Tier 2 storage densities are measured in terabytes (TB), because Tier 2 storage is much less expensive per bit than Tier 1 memory. That's another 1,000X difference between Tier 1 memory and Tier 2 storage. Typical characteristics of Tier 2 storage include relatively low bandwidth, long access latency, low cost per bit, and high capacity relative to Tier 1 memory.

Tier 2 storage differs from Tier 1 memory in another significant way: Tier 2 storage is persistent while Tier 1 memory based on DRAM is usually volatile. Tier 1 memory loses all of its stored data when the system power switches off, either due to a planned shutdown or due to a power failure. Tier 2 storage generally doesn't forget.
Intel® Optane™ DC Persistent Memory – A New Tier in the Memory Hierarchy

Intel has introduced a new memory category to address the bandwidth, latency, and capacity gaps between Tier 1 memory and Tier 2 storage. That new type of memory is Intel® Optane™ DC persistent memory DIMMs, shown in Figure 1, which creates another, intermediate tier in the memory hierarchy. You can use intermediate-tier memory based on Intel Optane DC persistent memory DIMMs to resolve a number of computing performance challenges.

Figure 1 shows where Intel Optane DC persistent memory fits in the memory hierarchy.

![Figure 1. Intel Optane DC persistent memory fits just below DRAM in the memory hierarchy.](image)

Intel Optane DC persistent memory modules (shown in Figure 2) are electrically compatible with DDR4 SDRAM sockets and can coexist with conventional DDR4 SDRAM DIMMs on the same CPU motherboard. Although electrically compatible, Intel Optane DC persistent memory DIMMs employ a different communications protocol called DDR-T, which adds commands specific to Intel Optane DC persistent memory DIMMs.

Consequently, Intel Optane DC persistent memory DIMMs require a specialized memory controller. The Intel Optane DC persistent memory module fits into standard DDR4 DIMM slots in motherboards designed for 2nd Generation Intel Xeon® Scalable processors, which incorporate appropriate memory controllers that can manage conventional DDR4 SDRAM DIMMs and Intel Optane DC persistent memory.

![Figure 2. Intel Optane DC persistent memory DIMMs can physically plug into DDR4 SDRAM sockets but require DDR-T signal timing protocols.](image)

On a per-DIMM basis, Intel Optane DC persistent memory offers denser (more memory capacity per device) that is only slightly slower in latency compared to fast SDRAM DIMMs. When using Intel Optane DC persistent memory DIMMs in an appropriately enabled server motherboard, system architects must carefully choose the mix of SDRAM and Intel Optane DC DIMMs to be plugged into the limited number of available DIMM slots on the board. The optimum balance between the two memory types depends on a variety of factors including memory-access patterns over a variety of expected workloads. Other computing factors or parameters may also affect the decision. Successfully predicting the appropriate mix of DRAM and Intel Optane DC persistent memory based on all of these dynamic factors can present quite a challenge.
Intel FPGAs are the first semiconductor devices besides Intel Xeon CPUs that can simultaneously support all three memory and storage tiers. Plugging an appropriately designed Intel FPGA-based memory/storage expansion card into a server motherboard’s expansion-card slot reduces or eliminates the need to populate some of the server CPU’s relatively limited number of DIMM sockets with Intel Optane DC DIMMs. Instead, the Intel Optane DC persistent memory DIMMs are inserted into DIMM sockets on the expansion card or cards, which communicate with the server CPU over a PCI Express* (PCIe*) interface or through Intel’s Ultra Path Interconnect (UPI) and compute express link (CXL) coherent protocol interfaces, as shown in Figure 3.

As the above figure shows, Intel Stratix® 10 DX FPGAs can connect to select Intel Xeon CPUs via the UPI coherent interconnect while the previously announced Intel Agilex™ FPGAs can use the coherent CXL interconnect. (Note: This White Paper strictly discusses the use of Intel Optane DC persistent memory with Intel Stratix 10 DX FPGAs.)

Plugging a suitably equipped expansion card into a server motherboard with a PCIe or UPI interface allows the Intel Xeon CPU to “see” the Intel FPGA-attached memory as host memory. Using this arrangement, system designers can populate all of the server CPU’s DIMM sockets with fast SDRAM to hold “hot” data, while storing “warm” data on the Intel Optane DC persistent memory located on suitably designed Intel FPGA expansion card(s).

Intel Optane DC persistent memory is available in capacities of 128 GB, 256 GB, and 512 GB and provides significantly higher memory capacity when compared to available SDRAM DIMMs, which are currently limited to a maximum capacity of 128 GB per DIMM. The new, flexible, intermediate-memory tier based on Intel Optane DC persistent memory provides designers and developers with large-capacity and affordable memory that is flexible and non-volatile while providing much faster performance relative to Tier 2 storage - on the order of 1,000X faster. Consistent, low-latency access times combined with high bandwidth, good Quality of Service (QoS), and endurance mean that the combination of Intel Optane DC persistent memory technology and the expansion capabilities of Intel FPGAs allows servers to provide more memory capacity and to support more virtual machines (VMs) for cloud-based and virtualized users.

Servers equipped with Intel Optane DC persistent memory can have much greater capacity to hold in-memory databases, but without incurring the prohibitive price tags of servers based exclusively on all-DRAM Tier 1 memory for storing those databases. Many mission-critical databases and other enterprise applications store large amounts of data in working memory. These mission-critical databases can also work well in non-volatile, fast-access, intermediate-tier memory based on Intel Optane memory technology. Memory-bound workloads benefit from Intel Optane DC persistent memory with its large capacity, high endurance, and greater bandwidth when compared to NAND SSDs.

Intel Optane DC persistent memory DIMMs also enable faster system restarts because the data stored in persistent memory is retained even during power outages. If a server goes down, either through a planned shutdown or because of accidental power loss, many hours may be needed to reload the server’s memory array with the lost database. Reloading a large database after a shutdown or a power loss greatly increases server and application downtime and application downtime can cost thousands of dollars per minute in lost business. For large retail operations, the costs associated with the lost revenue can climb to millions of dollars.

Intel Optane DC persistent memory retains data through a power cycle, so these mission-critical applications can return to service much faster by avoiding the need to reload the large database. Enterprise data centers, cloud providers, and communication service providers can consistently meet their service level agreements (SLAs) and avoid expensive system redundancy costs by adding Intel Optane DC persistent memory as an intermediate memory tier.
The Intel Stratix 10 FPGA Controller for Intel Optane DC Persistent Memory

Intel Optane DC persistent memory DIMMs use the DDR4 SDRAM electrical interface protocol but require somewhat different signal timing and different read-and-write protocols. Consequently, Intel Optane DC persistent memory DIMMs must be paired with an appropriate memory controller that has different or extended capabilities versus a memory controller designed to manage DDR4 SDRAM exclusively.

Intel offers a soft intellectual property (IP) memory controller, available within the Intel Quartus® Prime software, that’s specifically designed to manage Intel Optane DC persistent memory DIMMs operating at a maximum transfer rate of 2,400 MTransfers/sec (for Intel Stratix 10 DX FPGAs). Each instantiation of this soft IP memory controller in an FPGA can accommodate one 72 bit memory channel with two DIMMs per channel. Intel Stratix 10 DX FPGAs incorporate as many as four 72 bit DDR4-compatible channels, which allows these FPGAs to control as many as four banks of Intel Optane DC persistent memory per FPGA, with one or two Intel Optane DC persistent memory DIMMs per 72 bit memory channel.

Because of this memory controller IP – combined with on-chip, hardened DDR4 SDRAM memory controllers, SSD controller IP, and direct support for PCIe Gen4 x16 and UPI interfaces – Intel Stratix 10 DX FPGAs can uniquely provide universal memory- and storage-expansion services for server interfaces – Intel Stratix 10 DX FPGAs can uniquely provide controller IP, and direct support for PCIe Gen4 x16 and UPI on-chip, hardened DDR4 SDRAM memory controllers, SSD controller IP, and direct support for PCIe Gen4 x16 and UPI interfaces – Intel Stratix 10 DX FPGAs can uniquely provide universal memory- and storage-expansion services for server CPUs including Intel Xeon processors.

Intel Optane DC Persistent Memory Operational Modes

A server CPU accesses Intel Optane DC persistent memory DIMMs using one of two modes:

- **Memory Mode** – The Intel Optane DC persistent memory serves as the CPU’s primary memory, with a DRAM cache. The DRAM cache holds written data until it can be moved to Intel Optane DC persistent memory. In memory mode, the data stored in the Intel Optane DC persistent memory is not persistent due to the DRAM cache. That may seem counterintuitive, but the persistence comes from the Intel Optane DC persistent memory DIMM and the server CPU actually writes to the DRAM cache. So until the written data is transferred from the DRAM cache to the Intel Optane DC persistent memory, the data is not stored in a persistent manner. The Intel Optane DC persistent memory controller IP for the Intel Stratix 10 DX FPGA is not designed to use DRAM as a cache and therefore does not operate in Memory Mode.

- **App Direct Mode** – In App Direct mode, software running on the server CPU can directly access memory storage based on the Intel Optane DC persistent memory DIMMs using file-system application programming interface (API) calls or even more directly with memory loads and stores to a unique pool of system memory that has abilities and attributes that differ somewhat from SDRAM, such as non-volatility and somewhat slower write transactions. The Intel Optane DC persistent memory controller IP for the Intel Stratix 10 DX FPGA always operates in App Direct mode.

In App Direct Mode, software and applications can talk directly to the Intel Optane DC persistent memory through native file-system API calls but without the file-system software stack needed to communicate with Tier 2 storage devices. Consequently, file-system API calls directed at Intel Optane DC persistent memory storage complete much faster than do similar calls directed to SSDs or HDDs. An application program designed to talk to Tier 2 storage through the file system can run unmodified in this manner. For further details on Persistent Memory programming, please consult the pmem.io website.

These transactions occur much faster because Intel Optane DC persistent memory is much faster than an SSD or rotating-media drive and because the file-system I/O stack designed for Intel Optane DC persistent memory is much simpler and faster than the corresponding software stack needed to manage Tier 2 storage. In effect, this mode transforms the Intel Optane DC persistent memory into a very fast, virtual SSD while avoiding the performance bottlenecks that can occur by converting these API calls into non-volatile memory express (NVMe) and other drive-centric storage protocol transactions.

A server CPU can also access attached Intel Optane DC persistent memory DIMMs directly as system memory using App Direct mode. In this mode, the application software reads and writes persistent data in the same way that it reads and writes data to SDRAM using the appropriate addresses in the server CPU’s memory address space. There is no I/O stack at all for this access mode.

A memory pool based on Intel Optane DC persistent memory is persistent like storage, byte-addressable like Tier 1 SDRAM, and has consistently low memory latency, all of which support fast access to larger datasets. In this mode, called the direct-access (DAX) mode, the server CPU’s operating system sees Intel Optane DC persistent memory and DRAM as two separate memory pools in the server CPU’s address space with different speeds and capabilities.

**Note:** The Storage Networking Industry Association (SNIA) calls this configuration “non-uniform memory access” (NUMA) because the memory pools based on SDRAM and Intel Optane DC persistent memory have somewhat different behaviors although they are both accessed as locally attached system memory through the server CPU’s memory-address space.

To access Intel Optane DC persistent memory using the DAX mode, the application program must be aware of the difference between persistent and volatile memory and act accordingly. Such application programs are said to be “persistent-memory aware” (PM Aware) programs. Most application software is currently not written this way, because they were written prior to Intel Optane DC persistent memory availability, so existing software must usually be rewritten to make use of the DAX mode and realize the additional performance benefit. One exception is the SAP HANA database, which has been specifically optimized for Intel Optane DC persistent memory. Of course, new applications can be written to use the faster DAX mode from the very start.
**Intel Optane DC Persistent Memory Security**

Intel Optane DC persistent memory has built-in, 256-AES hardware encryption so you can rest easy knowing your data is more secure. In App Direct Mode, data stored in the Intel Optane DC persistent memory DIMMs is encrypted using a key stored on the DIMM. The data in memory is locked upon power loss and a passphrase is needed to unlock and access the stored data. The encryption key is stored in a security metadata region on the Intel Optane DC persistent memory DIMM that is only accessible to a server with a suitable Intel Optane DC persistent memory controller. For even more data security, if the Intel Optane DC persistent memory DIMM is repurposed or discarded, a secure cryptographic erase and DIMM over-write can prevent any possibility of improper data access.

**Intel Stratix 10 DX FPGAs Can Also Control SDRAM Banks and SSDs for Comprehensive Expansion**

Intel Stratix 10 FPGAs including Intel Stratix 10 DX devices support DDR4, DDR3, and DDR3L SDRAM protocols with multiple hardened memory controllers and hardened PHYs. In addition, Intel Stratix 10 FPGAs can support QDR-IV, QDR II + Xtreme, QDR II +, and QDR II SDRAMs using a soft memory controller in conjunction with the Intel Stratix 10 FPGAs’ hard PHYs.

In addition, Intel Stratix 10 devices including Intel Stratix 10 DX FPGAs can include soft IP instances for SSD control. There are multiple SSD interfaces including SATA, PCIe, and SAS, but for many systems NVMe is the interface of choice because of its high performance. Multiple 3rd-party IP vendors offer NVMe SSD controller IP for Intel Stratix 10 FPGAs including Intel Stratix 10 DX devices. One such NVMe controller IP core achieves better than 3,300 megabytes per second (Mbps) for read transactions and better than 2,100 Mbps for write transactions while consuming very few FPGA resources. Of course, Intel has developed a variety of NVMe SSD drivers for various operating systems.

Consequently, Intel Stratix 10 DX FPGAs can be used as the heart of memory/storage expansion card designs capable of supporting SDRAM, Intel Optane DC persistent memory, and SSD storage – and all using just one FPGA to control all of these memory and storage devices.

**The Intel Stratix 10 DX FPGA Can Process Data in Addition to Storing It**

The Intel Stratix 10 DX FPGA’s PCIe Gen4 and UPI interfaces make the device an ideal foundation and hub for a memory and storage expansion card, but using the FPGA in this manner leaves the FPGA’s logic fabric highly under-utilized. The FPGA’s on-chip logic fabric, digital signal processors (DSPs), and fast SRAM can be combined to provide a substantial amount of on-chip processing ability and system designers can realize significant additional performance benefits by structuring all of these resources (on-chip memory, attached DDR4 SDRAMs, attached Intel Optane DC persistent memory, and attached storage drives) to accelerate specific workloads that run locally on custom-designed processing engines instantiated in the FPGA’s programmable hardware.

Accelerating workloads using the FPGA can significantly offload the host CPU, freeing up CPU cycles that can be better used for other tasks. Partitioning workloads in this manner can minimize workload latency and, potentially, improves system-level performance. There are many existing, high-performance IP cores designed for FPGAs that implement myriad workloads, including:

- Compression/decompression
- Encryption/decryption
- Filtering
- Data analytics
- Artificial intelligence
- Machine learning
- Deep learning
- Delta merge operations for database updates

Another example of a suitable workload, one that’s frequently implemented using FPGAs, is a smart network interface controller (SmartNIC). Combining SmartNIC capabilities with persistent memory storage can decouple network transfers from the server CPU, which can speed overall throughput by minimizing inter-packet delays. Offloading network workloads allows system architects to more effectively disaggregate memory and storage from CPUs in data centers.

**Examples of FPGA Workload Processing with Persistent Memory Storage**

One important workload that’s well-suited to FPGA-based implementation is in-memory database acceleration. The growth of big data has driven the need for larger and larger, memory-resident databases. Using an FPGA as a memory expander devoted to Intel Optane persistent memory allows large amounts of warm data to be kept in non-volatile Intel Optane DC persistent memory while the host CPU works on hot data using its local DRAM storage.

At the same time, the FPGA on the memory-expansion card can be configured to perform various functions, such as compression or query acceleration, drawing data directly from the Intel Optane persistent memory. This arrangement offloads specific tasks from the server CPU and can boost overall system performance. (For an example of database acceleration using Intel Optane DC persistent memory, see the Baidu case study titled “Baidu Feed Stream Services Restructures Its In-Memory Database with Intel Optane Technology” and the associated Intel News Byte titled “Intel Optane DC Persistent Memory Improves Search, Reduces Costs in Baidu’s Feed Stream Services.”)
Storage replication, often employed by distributed database programs, is another excellent example of a workload that can benefit from the combination of FPGA acceleration and Intel Optane DC persistent memory. Triple replication of data is the gold standard for database redundancy in modern data center architectures. However, the remote-replication process is typically slow, especially when the storage media involved is a slow HDD located remotely, somewhere in the data center’s network. Data must pass over the network from the server to a storage subsystem where it passes through a storage network switch to a target HDD or SSD. An acknowledgement must then be returned back over the network to the initiating CPU, signaling that the requested activity has been completed. Only when the acknowledgement is received, can the initiating host continue to the next task. This is a relatively slow chain of events.

Coupling Intel Optane DC persistent memory with an Intel FPGA configured as a SmartNIC can accelerate this process dramatically from the server’s perspective. Instead of replicating data on a remote drive to achieve redundant data storage, the data can be shadowed in non-volatile Intel Optane DC persistent memory DIMMs plugged into the network card. This arrangement commits the replicated data to persistent memory, which is much faster than any HDD or SSD. Once the data is committed to non-volatile persistent storage, the SmartNIC can immediately acknowledge completion of the requested storage-replication task and free up the waiting server CPU. Then, in the background, the SmartNIC can complete the replication process by sending the data to be stored directly from the attached persistent memory to the designated storage drive. No data will be lost, even if an unexpected power outage occurs during this transfer, because the data is backed up in persistent memory before the SmartNIC issues the acknowledgement to the server CPU.

**Conclusion**

Intel Stratix 10 DX FPGAs offer server developers a radically new way to expand server memory capacity using Intel Optane DC persistent memory DIMMs plugged into an expansion card connected to the server CPU through a fast PCIe Gen4 or UPI interface. These FPGAs manage the Intel Optane DC persistent memory using a soft IP controller core. Memory-expansion cards based on Intel Stratix 10 DX FPGAs reduce or eliminate the necessity of devoting some portion of the server motherboard’s limited number of DIMM sockets to Intel Optane DC persistent memory DIMMs while retaining all of the performance advantages and other benefits provided by Intel Optane DC persistent memory, including fast restart times, database permanence, and blackout-resistant data caching.

**References**

- “Baidu Feed Stream Services Restructures Its In-Memory Database with Intel Optane Technology”, Tao Wang, Chief Architect, Baidu.
- “Intel Optane DC Persistent Memory Improves Search, Reduces Costs in Baidu’s Feed Stream Services”, Intel, August 29, 2019.

**Where to Find More Information**