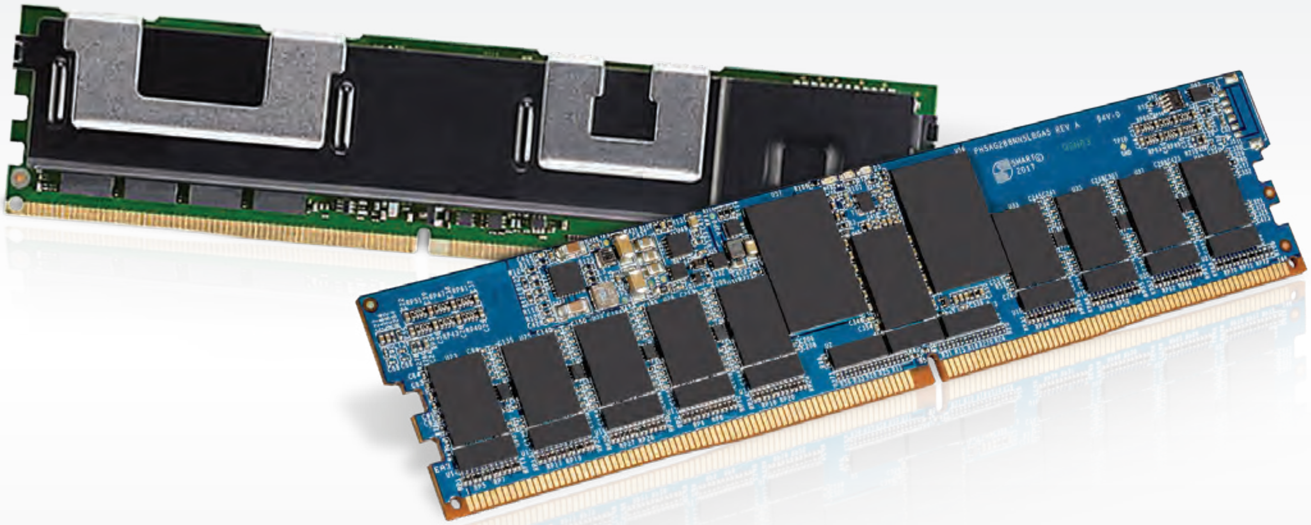
SMART®
Modular Technologies



White Paper M-WP007

Persistent Memory

NVDIMM-N and Optane™ DC DIMM™

Jan. 2021

Table of Contents

Introduction

Application responsiveness and system performance are key values for end users because every millisecond of latency can cost money. An excerpt from a recent article on high scalability in The GigaSpaces Technologies Blog (Insights into In-Memory Computing and Real-time Analytics).

”Latency matters, Amazon found every 100ms of latency cost them 1% in sales. Google found an extra 0.5 seconds in search page generation time dropped traffic by 20%. A broker could lose $4 million in revenues per millisecond if their electronic trading platform is 5 milliseconds behind the competition.”

Even more significant application level performance gains are achievable with the adoption of persistent memory. Persistent memory provides DRAM-speed access to memory, without the risk of losing data. This results in lower latency that can dramatically shorten data logging time, as just one example of its advantages.

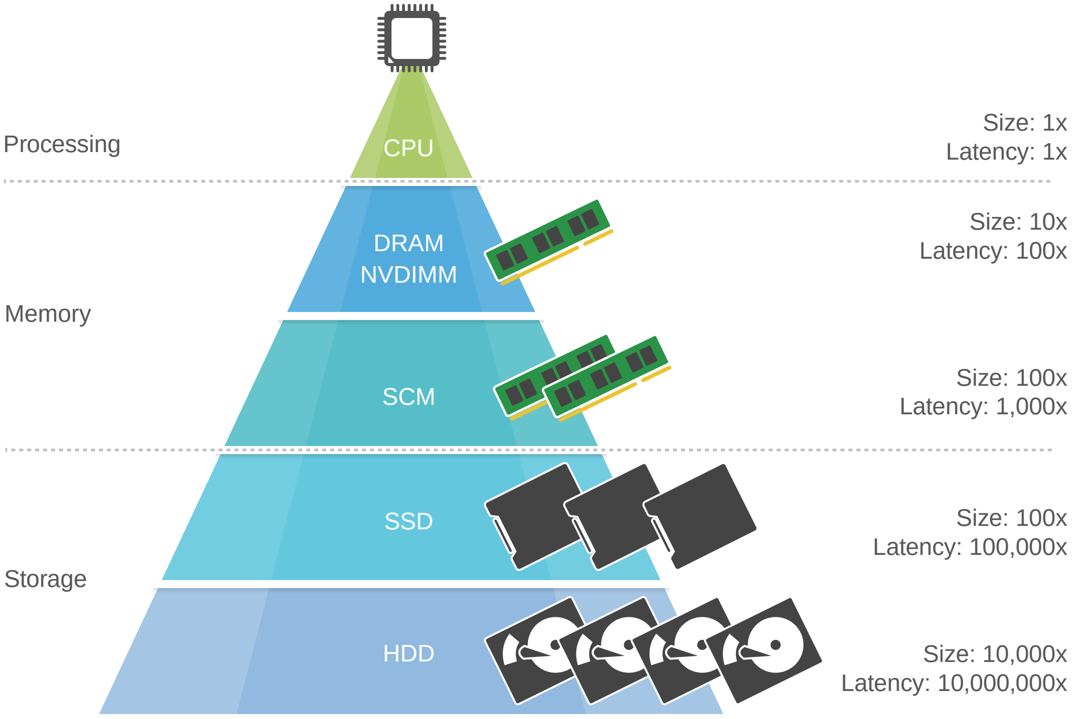
There are two types of DDR4 persistent memory modules available on the market today for data center servers. Although both modules can be installed in the same DDR4 systems and are used as byte-addressable, persistent memory, they have different characteristics and application use cases. This paper describes the two modules, identifies tradeoffs between them, and explains which data center server applications will benefit from using either one or both of these types of modules.

The goal of this paper is to objectively compare the two module technologies, highlight each of their benefits and costs, and provide guidance for determining the best server application use cases for the NVDIMM-N and Optane DC DIMM.

Persistent Memory and Memory Hierarchy

Persistent Memory for data center servers and storage servers is defined as nonvolatile, byte addressable (as opposed to block addressable) memory with latency close to DRAM level (10's of ns) and available in densities greater than or equal to DRAM 10's of GBs.

Persistent Memory improves server performance by eliminating the need to serialize or commit writes to lower level (slower) storage such as NAND Flash NVMe, SSDs, or HDDs. NAND Flash storage devices are block or page-based and have millisecond write speeds. When a server must write or commit data to NAND Flash, the application must wait 10's to 100's of milliseconds as opposed to waiting 10's to 100's nanoseconds for writes to persistent memory. Writing data to the next lower level in the memory hierarchy incurs a significant performance penalty.



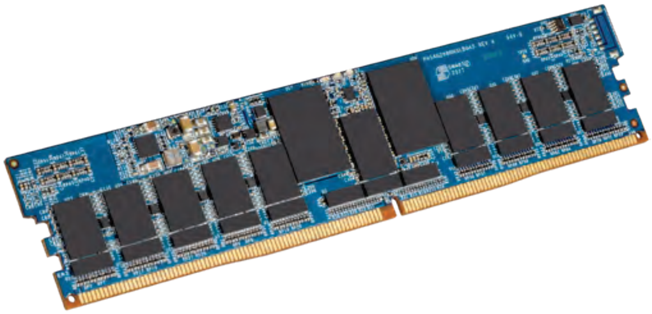
(Source: Adapted from SNIA PM Summit 2018)

The SNIA figure above, right, identifies the levels of the Memory Hierarchy with the NVDIMM-N represented at the DRAM level of the hierarchy and Optane DIMM appearing at the Storage Class Memory (SCM) level. Both are included in the Memory Level of the hierarchy because they are byte-addressable as opposed to the Storage Level, which is block addressable. The distinction is important because byte-addressable access to data down to the level of a cache line (usually 64 bytes) is the most efficient size of data for CPU L1, L2, and L3 cache movement operations.

What is a DDR4 NVDIMM-N Persistent Memory Module?

Overview

The NVDIMM (Non-Volatile Dual In-Line Memory Module) combines the use of volatile DRAMs and non-volatile Flash to retain data after a system power failure or reset by saving the DRAM contents to NAND Flash. During normal operation, the host system accesses the DRAM memory on the NVDIMM in the same way as an RDIMM.



JEDEC Standardization

The NVDIMM is a JEDEC defined module type with the host system recognizing the module as an NVDIMM during BIOS POST (Power On Self-Test) by reading the module's SPD (Serial Presence Detect) EEPROM in the same way as an RDIMM. Before the host system completes POST and boots an operating system, the BIOS will perform a restore operation reading data from the NVDIMM NAND Flash and writing it to DRAM. The BIOS will also report through ACPI (Advanced Configuration and Power Interface) tables the NVDIMM memory region as persistent, so the memory can be recognized and used by the applications as persistent memory. If there are multiple RDIMMs and NVDIMMs, the BIOS will interleave the RDIMMs and NVDIMMs separately to segregate the memory and optimize performance. NVDIMMs are supported on Intel, AMD, and other server platforms.

Capacity and Power

DDR4 NVDIMM is currently available in sizes of 16GB and 32GB which are mainstream RDIMM module capacities. To be able to perform a save on power failure, the NVDIMM must be provided an energy source. JEDEC defines two possible options, with either the host system providing the energy source via the two DDR4 slot 12V power pins, or the NVDIMM being tethered with a cable to a Backup Power Module (BPM).

Write Endurance

During normal operating mode, writes to the NVDIMM are only directed to DRAM memory, which has unlimited endurance. There is no wear difference between an NVDIMM and RDIMM during normal operation. The NVDIMM NAND Flash is written after a power loss or system reset event. Given the infrequency of these events during the lifetime of a server, the NVDIMM NAND Flash storage does not practically limit the endurance of an NVDIMM. NVDIMM designs will support multiple power loss or reset operations per day for 10 years or more. 99.999% availability or "five 9s" will result in not more than 1 or 2 such power loss or reset events per year.

Performance

The NVDIMM module runs at the same speed as RDIMM modules and can be used as Uniform Memory Access (UMA) memory in the system showing no performance difference. The NVDIMM provides DRAM speed write or commit operations accelerating write and logging operations that otherwise would have to wait to be written to the next level in the memory hierarchy. As shown on the previous page, the requirement to write to the next level of the memory hierarchy will result in a 10x or more performance penalty. The NVDIMM DRAM memory read and write latency is symmetrical, meaning that both read and write operations have the same performance level.

Optane DC DIMM (Persistent Memory Using 3D XPoint)

Overview

Optane is Intel's brand name for devices built using 3D XPoint™ media. Optane DIMMs provide a persistent direct CPU addressable storage tier that is faster than NAND, but slower than DRAM. In the previously shown Memory Hierarchy, Optane DC DIMMs reside at the Storage Class Memory (SCM) level.

Intel Proprietary

Intel has defined a proprietary signaling standard (DDR-T) that allows installing Optane DIMMs into Intel Xeon processor-based servers. This protocol allows the Optane memory to be accessible in the DDR4 memory slot without slowing down the faster RDIMM memory. Currently Optane DIMMs are only supported in Intel servers. SMART is partnering with Intel to provide a PCIe Add-In-Card (AIC) which will allow Optane DIMMs to be utilized on non-Intel Xeon server platforms.

Capacity and Power

Optane DIMMs are available in sizes of 128GB, 256GB and 512GB and do not require an energy source to guarantee persistency. It is important to note that DDR4 RDIMMs are only available up to a size of 256GB. Optane DIMMs, while providing several times the capacity of DRAM-based DIMMs, are estimated to consume a comparable amount of power to DRAM-based DIMMs, because Optane DIMMs can be used in the same DDR4 slots and servers as RDIMMs.

Write Endurance

The Optane DIMM storage technology has been reported by Intel to have a limited write endurance. This means, as with NAND Flash storage devices, the amount of wear on the storage devices must be managed and use cases with high write frequency or hot data may not be supportable.

Performance

Optane DIMM write latency has been reported by Intel to be 1000x faster than NAND storage and 10x slower than DRAM memory. Similar to a NAND device, Optane DIMM read latency has been reported by Intel to be lower or better than Optane DIMM write latency. The performance benefit from Optane DIMM is a result of the host CPU being able to write to the Optane DIMM SCM layer instead of the slower SSD NAND storage layer. Because the Optane DIMM has asymmetrical read and write latency, there is a more significant benefit to using the Optane DIMM for read-heavy or caching applications than write-heavy applications.

Application Use Cases for NVDIMM and Optane DIMMs

NVDIMMs used in Storage, Data Recovery, and GPU Servers

NVDIMMs are used in several different data center server applications. The NVDIMM is used widely in storage servers for write acceleration and commit logs. These applications benefit from the low DRAM write latency and unlimited endurance provided by NVDIMMs. The capacity required for write acceleration and commit logs typically does not exceed 1/8th of the amount of RDIMM memory in the server, making the NVDIMM cost and performance benefit acceptable for this application.

A second NVDIMM data center server application is data recovery following an operating system “hang” or crash. In this case, there is no power failure but data has been saved in the NVDIMM memory, which must be recovered to allow an application to continue running from where it left off before the operating system crash. During normal operation, the NVDIMM provides DRAM speed write and read access. After the operating system crash occurs, the NVDIMM is signaled to perform a save operation after a watchdog timeout, the BIOS performs the restore operation during POST, and the application data is available after the operation system has restarted.

The third NVDIMM application is using the NVDIMM in a GPU server as a location to store training and inference results. The system is I/O bound when fully loaded with GPU cards and GPU components, so being able to save results persistently to DRAM improves this Machine Learning (ML) application.

For the above applications, the defining characteristics that make the NVDIMM well-matched are these requirements:

- DRAM speed read-write latency
- Persistent memory
- One or two RDIMM module size capacity
- Unlimited write endurance

Optane DIMMs used in Storage Servers and In-Memory Database Applications

During the 2020 SNIA Persistent Memory Summit, Oracle presented use of the Optane DIMM memory in the X8M storage server. Oracle’s X8M database connects separate compute server and storage server systems with 100GE RDMA NIC (RNIC) and installs 1.5TB of Optane DIMMs into the storage server. Oracle X8M performance is improved with RNICs and Optane DIMMs for both OLTP Random Reads and Log Write/Redo Logging.

OLTP (On-Line Transaction Processing) Random Read Performance was improved by utilizing RDMA to write and read the 1.5TB Optane DIMM memory in the storage server as a read cache. Oracle utilized Optane DIMM as a read cache to improve on NAND access latency from 200us to 20us: 10x improvement.

Log Write or Redo Log performance was also improved by using RNICs and Optane DIMM memory. An 8x performance improvement was achieved. Whereas, the Optane read cache application used nearly all of 1.5TB Optane storage space, the Log Write function only used about 1GB of Optane Memory in persistent mode.

Optane DIMMs serve the Oracle X8M application well because most of the Optane memory is used for a large read cache, avoiding over-stressing the write endurance of Optane memory. This read-biased use of the Optane memory is complementary with the Optane memory higher relative read performance. The Log Write function requires only a small amount of Optane memory, which limits the number of persistent writes needed to Optane memory.

Traditional In-Memory Database and OLAP (On-Line Analytical Processing) applications store data in DRAM memory and the data set sizes for one server are limited to low single TBs. Several TB is the maximum amount of DRAM memory which can be installed in one DDR4 server. Using Optane DIMMs for In-Memory Database and OLAP servers increases by at least 4x the data set size which may be supported in a single server. For data sets of this size, 10x or more performance improvement can be realized by avoiding access to the lower SSD NAND next level in the Memory Hierarchy.

For the above applications, the defining characteristics that make the Optane DIMM well-matched are these requirements:

- Memory expansion 4x or more beyond DRAM modules
- Persistent memory with either low write frequency or small capacity

Complementary Nature of NVDIMM and Optane DIMM and Futures

SMART Modular has been designing and selling DDR4 NVDIMM-N modules since DDR4 servers were first introduced in 2014 and Intel Corporation began selling the Optane DC DIMM in 2019.

NVDIMMs and Optane DIMMs both provide persistent, byte-addressable memory, but nearly all similarities end with these characteristics. Within the Memory Hierarchy, the NVDIMM resides at the DRAM level and Optane DIMM at the SCM level.

The NVDIMM has 10x lower latency and unlimited endurance, while the Optane DIMM has 10x higher maximum capacity. These characteristics make the NVDIMM well-suited for write logging, write acceleration, and caching applications on the order of one or two DRAM memory module-sized capacity.

On the other hand, the Optane DIMM is matched for OLAP type applications, which require memory expansion and have a relatively low write frequency. The lower cost of Optane DIMM memory than DRAM and faster performance than NAND provides a compelling cost-benefit advantage for Optane DIMMs to be used for large in-memory database, read caching, and data analytics applications.

Feature	Optane DC DIMM	NVDIMM-N
Density	128GB to 512GB	8GB to 64GB
Endurance	10^6	Unlimited
Applications	In-Memory DB	Logging, checkpointing, caching
Standardization	DDR-T (Intel propriety)	JEDEC

The NVDIMM and Optane DIMM are different enough in characteristics that for systems to fully maximize performance for certain applications both solutions make sense to be implemented. For example, the Oracle X8M application mentioned previously could benefit from the Optane DIMM continuing to be used for a read cache for OLTP access while changing the Log Write or Redo Logging to reside in NVDIMM.

The relatively small size of the Redo Log and high-intensity write, low latency requirement are well-matched for an NVIDMM. In other similar application use cases, both module types could be used synergistically with the NVDIMM supporting hot persistent data result calculations in a relatively small amount of storage and Optane DIMMs supporting large cooler input data storage pools.

Intel will continue to increase capacity and performance of Optane DIMMs and work is being undertaken in JEDEC to define a standard DDR5 NVDIMM. There will likely be more data center server applications and use cases identified in servers of the 2020’s in which the integration of the NVDIMM and Optane DIMM are used in a complementary manner to provide a high performance, byte-addressable, persistent memory layer between CPU L1/L2/L3 caches and NVMe/SSD NAND block storage.

SMART® Modular Technologies

For more information, please visit: **www.smartm.com**

Headquarters/North America
T: (+1) 800-956-7627 • T: (+1) 510-623-1231
F: (+1) 510-623-1434 • E: info@smartm.com

Latin America
T: (+55) 11 4417-7200 • E: sales.br@smartm.com

Asia/Pacific
T: (+65) 6678-7670 • E: sales.asia@smartm.com

EMEA
T: (+44) 0 7826-064-745 • E: sales.euro@smartm.com

Customer Service
T: (+1) 978-303-8500 • E: customers@smartm.com