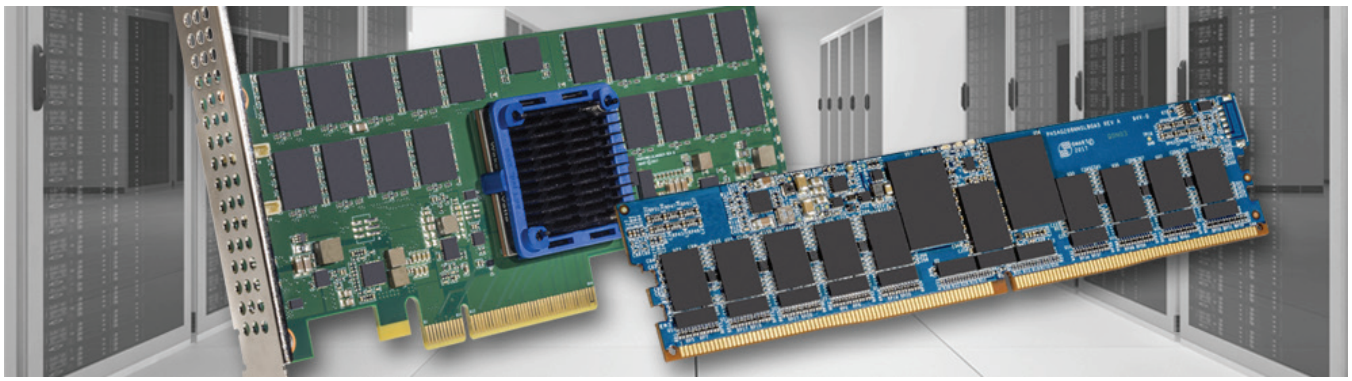
SMART®
Modular Technologies



Persistent Memory

High Speed and Low Latency

White Paper M-WP006

Corporate Headquarters: 39870 Eureka Dr., Newark, CA 94560, USA • Tel: (510) 623-1231 • Fax: (510) 623-1434 • E-mail: info@smartm.com
Customer Service: Tel: (+1) 978-303-8500 • Email: customers@smartm.com
Latin America: Tel: (+55) 11 4417-7200 • Email: sac@smartm.com
Europe: Tel: (+44) 7825-084427 • Email: euro.sales@smartm.com
Asia/Pacific: Tel: (+65) 6232-2858 • Email: infoasia@smartm.com

Table of Contents

Introduction **3**

nvNITRO™ **4**

NVDIMM **5**

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. This article discusses two such persistent memory products that create value for customers: The first one is used with file or block based IO applications in systems with a PCIe socket. The second one is used for memory mapped applications needing persistent data.

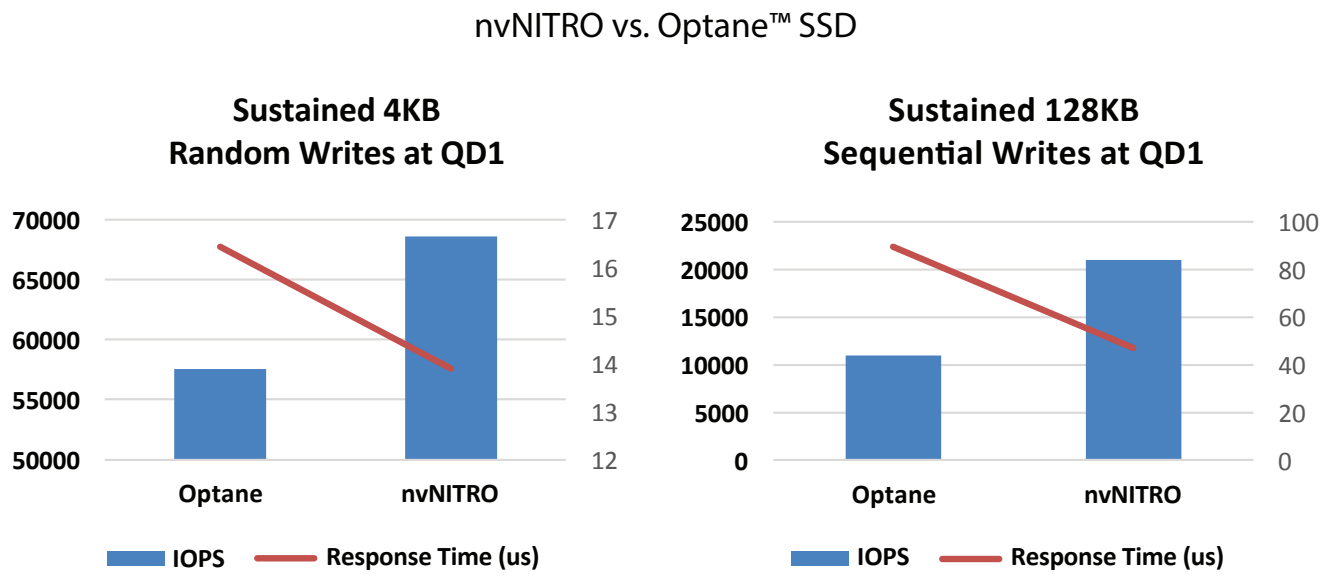
nvNITRO

The nvNITRO NVMe Storage Accelerator, or nvNITRO card, is a Half-Height Half-Length (HHHL) PCIe 1GB card that plugs into standard servers. It uses Everspin's STT-MRAM technology that is byte-addressable and persistent with very low latency, enabling dramatically improved system performance for many applications. The nvNITRO Accelerator Card fully exploits these unique MRAM attributes to provide a disruptive solution for industries where transaction processing and data recording performance are critical.

One specific application that can take advantage of the nvNITRO card is high-speed trading. The nvNITRO card is used as a high-speed write buffer or ring buffer. Data is written to the card in bytes and then pushed out to the SSDs in large blocks. The system can operate at maximum speed and maximum bandwidth without being slowed down by waiting for the SSDs to commit synchronous transaction logs. With systems today the solution is to use high-speed Enterprise SSDs. With the nvNITRO card, the system does not need to wait for the write logs to be saved into SSDs. The data is already protected and made persistent. The nvNITRO card reduces financial application wait times by reducing latency up to 90% over enterprise SSDs. In addition, there is no need to use more costly Enterprise SSDs for performance. Rather, lower cost Client SSDs can be used since the nvNITRO card is both accelerating and protecting the data from loss. Customers can utilize a single nvNITRO card as a front end cache connected to less expensive backend storage devices, providing a very cost effective solution for logging information and accessing terabytes of data at very high performance, while providing power loss protection for data. No super capacitors, UPS or batteries are required because nvNITRO is power fail safe through its use of STT-MRAM persistent memory technology.

Another important reason to use the nvNITRO card is due to financial trading regulations, where transaction logs, transaction communications, and all order status changes need to be saved.

As mentioned above at the start of this section, the nvNITRO card provides the capability to quickly store and fetch mission critical data. The chart below shows the benchmark test results of a 1GB nvNITRO NVMe card versus a 480GB Intel Optane NVMe SSD. The data was collected using a Supermicro X11DRi server board loaded with an ezfio benchmark tool. There is a dramatic increase in IOPs and a dramatic reduction in latency for sustained random and sequential writes for the nvNITRO card versus the Optane SSD.



The benefit of using Everspin's MRAM technology on a PCIe NVMe card is to dramatically accelerate system performance. No hardware modifications or file system are needed. BIOS and MRC (Memory Reference Code) changes are also not needed. Transaction logs can be written as fast as the system is capable of calculating or producing without risk of data loss. SMART's nvNITRO Storage Accelerator Card can help customers dramatically increase their system performance and save money doing it.

6 Reasons to Use nvNITRO for file or block based IO applications
1. Lowest latency (~6µs) and highest performing NVMe product for 100% 4KB RW IOPs (~1.5M)
2. Removes system bottlenecks by accelerating synchronous logging data being written to SSDs
3. No need for backup power
4. Plug and Play - No need for additional drivers
5. Save money by not needing to use Enterprise SSDs
6. Save money by not requiring a high core count CPU to increase performance

NVDIMM

The second persistent memory product pertinent to this paper is the NVDIMM. NVDIMMs are used in memory mapped applications where the high-speed access to memory is made persistent, or continually accessible with NVDIMMs, which dramatically improves system performance.

Where nvNITRO uses the NVMe protocol, which achieves a 6GB/s bandwidth and 6µs latency, NVDIMMs have no I/O overhead. Because of that, there is direct access from the CPU to the memory and NVDIMMs can achieve a 25.6GB/s bandwidth with DDR4-3200 and 50ns latency. With those two factors in mind, NVDIMMs are routinely and increasingly being used to improve system performance and reduce bottlenecks in high-availability, high-reliability server and storage applications. "Hot data" can now be put in portions and main memory made persistent with NVDIMMs without risk of loss. Now, critical data that would normally need to be moved to SSDs can be kept in memory for high speed access. Latency and performance-sensitive data, such as transaction logs and checkpoint images, can be stored and accessed from the persistent sections of main memory to dramatically improve application performance without the risk of data loss.

Two application examples of how system performance is dramatically improved with the use of NVDIMM-N include asynchronous transaction logging and check pointing. Transaction logging records changes to databases in a journal file and check pointing records the state of the database at a given moment in time. Instead of this data being stored into flash, such as an SSD, and taking hundreds of microseconds to store and access while going through layers of I/O, the data is now directly mapped into main memory and is accessed in hundreds of nanoseconds.

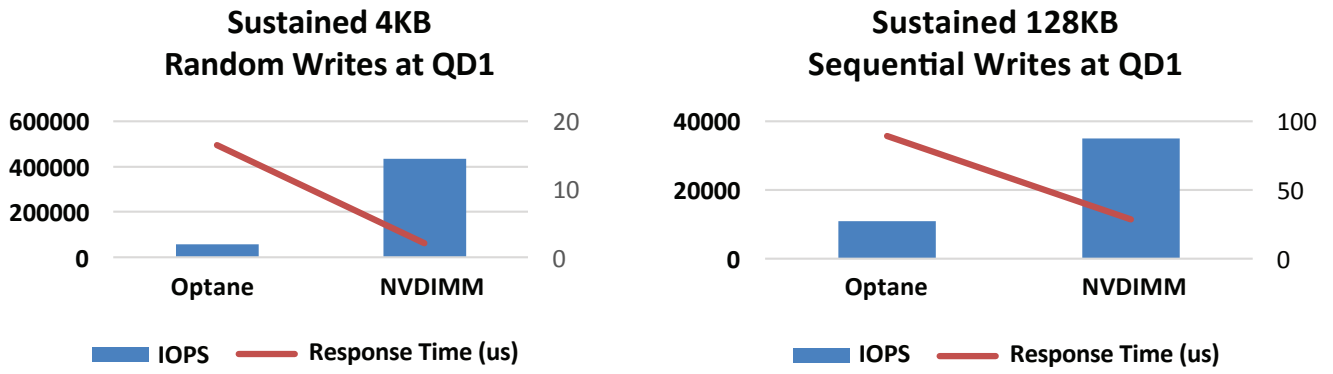
In systems that are configured to do asynchronous transaction logging, the system does not have to wait until the logging data is committed to start the next transaction. The advantage is very low latency and the system can run as fast as possible. The disadvantage is in the case of a power failure where all the transactions and order data in the memory are lost. By adopting NVDIMMs in these types of systems, logging data can be stored in main memory with NVDIMMs. There is no risk of losing log data, with a power failure, as noted above, and it eliminates UPS hold-up systems or the need for battery-backed NVRAM cards. System performance will be dramatically improved by more than four times.

NVDIMM-enabled systems include the following characteristics

- MRC that enables the NVDIMM
- BIOS that recognizes a NVDIMM when it is plugged in and maps the persistent memory regions
- ADR (Asynchronous DRAM Refresh) signal from the CPU that is wired to each DIMM socket for the DRAMs to be put into self-refresh before a power loss
- 12V power wired to the DIMM socket, which provides backup power to the NVDIMM long enough for the data to be transferred from the DRAMs to the Flash during a power loss
- NVDIMM-enabled OS - Linux 4.4+ or Windows Server 2016
- Application awareness of persistent memory

NVDIMMs provide a significant performance advantage by being able to store and retrieve "hot," or critical data at DRAM speeds without risk of losing the data. The chart below shows the benchmark test results of a 16GB DDR4 NVDIMM versus a 480GB Intel Optane NVMe SSD. The data was collected using a Supermicro X11DRi server board loaded with an ezfio benchmark tool. There is a dramatic increase in IOPs and a dramatic reduction in latency for sustained random and sequential writes for the NVDIMM versus the Optane SSD.

NVDIMM vs. Optane SSD

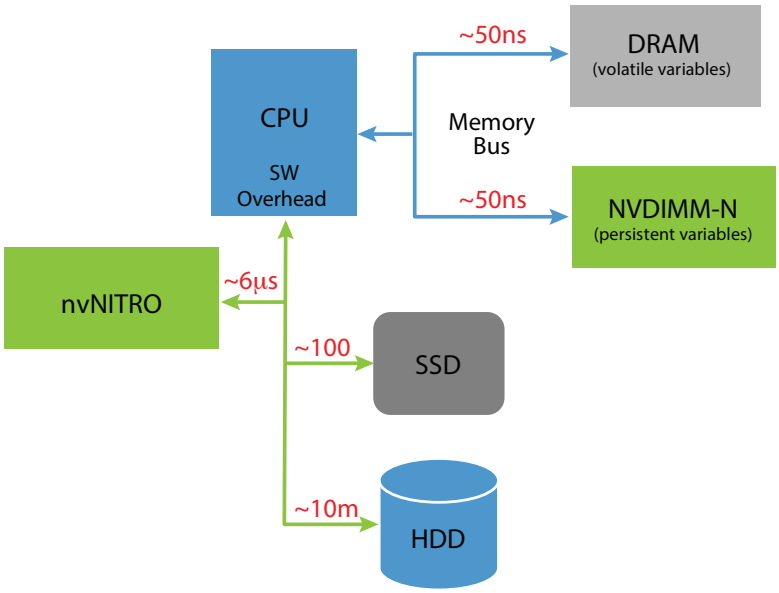


Significant application-level performance gains are achievable with the adoption of persistent memory products and the diagram above shows the concept of how nvNITRO and NVDIMM-Ns leverage their respective technologies to improve system performance.

The nvNITRO card uses MRAM technology to remove the bottlenecks in I/O logging and NVDIMM-Ns use DRAM and flash technology to improve system performance.

SMART is a leading designer, enabler, and supplier of advanced persistent memory products. SMART also contributes engineering and design support to JEDEC, OpenCAPI and Gen-Z industry organizations.

These two persistent memory products, the nvNITRO Storage Accelerator and NVDIMM, have clear and separate benefits that can improve application responsiveness and increase system performance for end users.



For file or block-based IO applications in systems with a PCIe socket, the nvNITRO card can be plugged in to achieve immediate latency, bandwidth and IOP benefits. For memory-mapped applications, NVDIMMs can be adopted to also achieve immediate latency, bandwidth and IOP benefits with the stipulation that the systems, OS, and applications are NVDIMM-enabled.

For more information please contact SMART Modular Technologies (www.smartm.com).