

# **X**itore, Inc.

**Introducing NVDIMM-X:  
Designed to be the World's Fastest NAND-Based SSD  
Architecture and a Platform for the  
Next Generation of New Media SSDs**

Doug Finke  
Director of Product Marketing  
September 2016

[www.xitore.com](http://www.xitore.com)

## Introduction

As computing technology is increasingly called upon to provide solutions for data rich applications such as analytics, the internet of things, deep learning, and other demanding applications, the bottleneck to overall system performance rests in the performance level of the storage subsystem. Not only are improvements needed in the raw bandwidth to storage, but also in the latency (response time), IOPs, and quality of service (consistency of the latency) to meet the needs of ever demanding customers. Several key factors that help to determine this performance include the transfer rate of the data bus, the overhead that is required to get the data to the microprocessor and the raw access time of the data media itself.

This paper will discuss a new architecture called NVDIMM-X that helps solve many of these problems. It takes advantage of the fastest bus in the computer system, the DRAM bus, and provides a platform for new media such as ReRAM, PCM, MRAM, 3D Xpoint, ST-RAM, and Memristor.

While the processing speeds of the latest microprocessors are constantly increasing, the performance levels of hard disk drives have stayed about the same for many years. An improvement occurred when NAND-based SAS and SATA interface SSDs were introduced and yet another improvement appeared as the NAND-based SSDs converted to the faster PCIe bus, but the raw performance of the NAND chip itself has also stayed pretty constant over the last several years.

The next logical step in the evolution is to use the DRAM interface as a storage interface. This interface is the fastest bus in today's computer system running at a speed of up to 3200 million transfers per second. It is wide and transfers 8 bytes of data at a time. In addition, it is very tightly coupled to the microprocessor and transfers data with very little overhead or wasted cycles. This interface provides over a 3X increase in raw bandwidth over today's NVMe solutions which depend upon the PCIe bus.

The initial steps to take advantage of the DRAM bus are already happening with a technology called the NVDIMM. The next section of this paper will describe the three versions of the NVDIMM that have already been defined by JEDEC. These are called NVDIMM-N, NVDIMM-F, and NVDIMM-P. These versions use architectures that use a combination of DRAM and NAND to provide greatly improved performance.

Following that section of this paper will be a description of a new version called NVDIMM-X, which provides significant improvements in features, performance, capacity, size, power usage, and cost. The initial version of NVDIMM-X uses media that consists of a combination of DRAM and NAND. However, NVDIMM-X will also serve as an ideal/host interface to use with the new/future media (Re-RAM, PCM, MRAM, 3D Xpoint, ST-RAM, Memristor, etc.) once they mature and show advantages over the DRAM and NAND media used today.

## Initial Versions of NVDIMM

JEDEC and SNIA have already defined three variations of NVDIMM. These can be seen as precursors to NVDIMM-X. However, NVDIMM-X contains additional innovations as explained later in this paper. [Table 1](#) below shows a comparison of the existing NVDIMM versions and an additional product called Memory1™.

**Table 1: Non-Volatile Dual-in-Line Memory Module (NVDIMM) and Memory1™**

Type	Storage or Persistent Memory	Main Technology	Requires External Components	Application Access?	Application Access to NV	Latency	DIMM Density Max	IOPS	Endurance	Queue	BIOS Change Needed	Note
NVDIMM-N	Persistent Memory	DRAM w/Flash as Back-up	Possibly	DRAM Byte Addressable	NO	10usec	64GB	2 MIOPS	N/A	1	Possibly	1, 2, 4
NVDIMM-F	Storage	Flash	NO	Flash Block Addressable	Through Addressable Window	10usec	128GB	100KIOPS	Flash	2	Possibly	
NVDIMM-P	Persistent Memory	DRAM & Flash	TBD	Byte Addressable	Yes	TBD	TBD	3MIOPS-N 100KIOPS-F	Flash or Other Media	TBD	Possibly	
Memory1	Neither	Flash	Yes. External DRAM Cache	DRAM Byte Addressable	No	100usec	256GB	N/A	N/A	1	Yes	
NVDIMM-X	Both	DRAM & Flash	Possibly	Flash Block Addressable and Byte Addressable in the future	Through Logical Address Space	2us	4TB	3-4MIOPS	10 <sup>5</sup> -10 <sup>6</sup> better than SLC	128	No	4,5

- 1) Although NVDIMM-N is byte addressable, drivers have been developed which can allow it to emulate a block storage device and be used for file systems
- 2) NVDIMM-N does not have endurance capabilities of storage solutions as flash is merely used as an emergency backup-system to store DRAM content during power loss only.
- 3) NVDIMM-X has higher endurance than any non-volatile memory solution due to its architecture and local cache.
- 4) If temporary backup power for the DRAM is not supplied by the system, an external SuperCAP will be required for the module to save the DRAM data to the non-volatile medium.
- 5) If a future version of NVDIMM-X were to also use a non-volatile memory technology for the cache, then no temporary backup power will be needed.

Following are details of each solution:

**NVDIMM-N** is not a storage solution but rather a standard DRAM DIMM with an additional local controller and additional flash, serving as an emergency back-up system. The NVDIMM-N is byte addressable with bandwidth, latency and capacity comparable to standard DRAM solutions. In addition, it features an additional local controller and an array of non-volatile memory (Flash) plus external SuperCAP to back-up DRAM content to its local flash array during power loss. The SuperCAP acts as a power source for the DIMM should system power be lost. The flash memory is only used by the local controller to back DRAM content but is not accessible by the host application. The flash density is equal to or larger than the DRAM density. The main purpose of this solution is not storage, but an emergency backup of the volatile DRAM data in case of a power outage making the data in the DRAM persistent. In addition, a BIOS change may be needed for the system to recognize this as a special type of DIMM.

**NVDIMM-F** is a storage solution that sits on a DRAM DIMM bus with multiple controllers and a bridge to convert the DRAM bus interface to multi-channel SATA protocols. On the back of the DIMM, one or more local controllers will convert SATA protocol to flash protocol for storage. This solution has the capability of block addressing, where a host system has the capability of using a moving address window to access the entire flash content. The aggregated bandwidth of this solution is gated by the number of local SATA controllers (each capable of max. 500MB/sec SATA-II interface). The latency of NVDIMM-F is directly correlated to the local SATA controller latency plus local DRAM-2-SATA bridge delays. IOPS performance is equivalent to flash-based SATA solutions and endurance is based on SATA controller firmware capabilities. The NVDIMM-F is capable of processing 1 or 2 queues at a time. NVDIMM-F performance is negatively impacted by multiple protocol conversions and associated BIOS/OS changes.

**NVDIMM-P** is a solution with specifications currently under active discussion at JEDEC for both the DDR4 and DDR5 technologies. Although NVDIMM-P was originally thought to be a combination of the NVDIMM-N and NVDIMM-F and offer both block and byte access, more recent proposals have moved to a concept of providing for a platform that can be a caching structure and use a non-deterministic protocol. The NVDIMM-P may also require BIOS changes to the system.

## Comparisons with the Memory1™

**Memory1** is neither a storage solution nor a persistent memory solution and the use cases for this product will be much different from the use cases for NVDIMMs. Rather, it is a large pool of NAND Flash acting as a DRAM. Although it has similarities to the NVDIMMs described above, the Memory1 product has a key difference that it will not retain its data when the power cuts off. So conceptually, it is a volatile memory, just like DRAM. It takes advantage of the fact that NAND chip capacities are currently about 4X the capacities of DRAM chips so that a DIMM can be configured with 128GB of memory instead of the current limit of 32GB with standard DDR4 chips. Memory1 requires its own custom device driver and system BIOS change. It also needs to occupy an extra DDR4 DIMM socket to serve as its front-end cache for all host application read and write requests. This additional DDR4 DIMM module acts as a cache for the data in Memory1. When the requested data is not in the cache or when the host system driver initiates a pre-fetch action, the Memory1 can take advantage of the DRAM bus speed to transfer the requested block of data from the Memory1 to the DDR4 DIMM module. During movement of data back and forth from Memory1 to the external DRAM DIMM Cache, the entire memory controller channel will be busy, thus the microprocessor will not be able to access any other socket on the same channel during that time.

## Innovations with NVDIMM-X

When thinking about how to achieve the best possible performance and value we set the following goals:

1. Achieve the fastest possible performance
2. Achieve costs similar to those achieved with traditional NAND-based NVMe and SAS SSDs
3. Use mature NAND and DRAM technology to minimize technical risks
4. Fit into an existing DDR4 sockets without any BIOS or operating system kernel changes
5. Achieve an extensible architecture that can support future memories with multiple terabytes per DIMM and not be limited to the 512GB inherent in today's standard DDR4 DIMM standard.
6. Provide hooks in the architecture so that new media like Re-RAM, PCM, MRAM, 3D Xpoint, Memristor, ST-RAM, etc. can be employed in the future.

With these goals in mind, Xitore created the NVDIMM-X architecture that will use the DDR4 DIMM slot and contain substantial NAND memory with a large, on-module DRAM cache in order to achieve low costs while maintaining excellent performance.

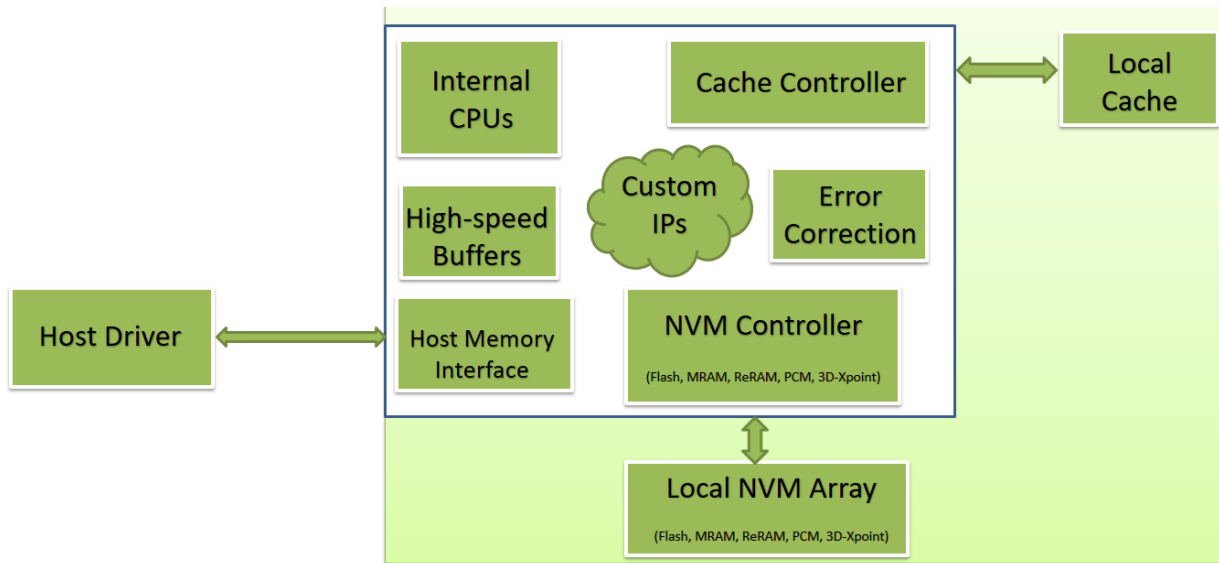
However, in order to use this architecture, we needed to solve one key problem. Specifically, the DDR4 DRAM bus has been designed using a deterministic protocol; thus, it does not have the capability of interfacing with heterogeneous memory subsystems that may have variable access times. And the cache-based architecture will produce variable access times depending upon whether the requested data is already in the DRAM cache or in the NAND.

To get around this limitation the initial version of the NVDIMM-X has been designed as a block mode device, i.e. an SSD on a DIMM. As discussed later on in this paper, future system architecture enhancements will allow the device to run in a byte access mode.

The NVDIMM-X architecture will use a new protocol that has several innovative features including:

1. The protocol is based upon command requests, returned status, and subsequent reads and writes that are communicated over the DDR4 bus. These transactions will be fully deterministic and avoid any problems that could violate the DDR4 protocol.
2. Provide a protocol that can accept a potential address space of  $2^{64}$  or 16 exabytes and provide headroom for increased capacity for many future generations of NAND or other media. Previous NVDIMM implementations use the standard DDR4 addressing protocol which communicates the address over the standard DRAM address bus and limits maximum capacity to  $2^{39}$  or about 512GB of data per DIMM. This is already a limitation that could prevent usage of high density NAND chips that are available right now.
3. The NVDIMM-X protocol will allow for the use of a 128 entry command queue. Among other things it would allow for out of order execution of the commands increasing the overall performance. This command queue capability would be another first for the NVDIMMs as previous implementations could only handle at most two commands at a time.
4. Because all transactions on the DDR4 bus occur in 64 byte chunks, when the DIMM does return the status, it will show the state of all possible 128 commands at one time (4 bits per command) providing additional efficiencies.
5. Since both the DRAM cache and the NAND are on the same module, any activity to store a cache line into the NAND or retrieve a cache line from the NAND to the DRAM will occur on-module and will not tie up the main DRAM bus. Unlike the NVDIMM-X, other implementations may require separate DRAM and NAND DIMM modules and which would consume extra bus cycles to transfer a line of data to or from the DRAM cache. The extra bus cycles in the other implementations decrease system performance because the DRAM bus cannot be used by the microprocessor during this time.

Figure 2 below shows a block diagram of the NVDIMM-X



## Transaction Diagrams

Figures 3, 4 and 5 below are some diagrams of how writes and reads are performed in the NVDIMM-X architecture. Note that all responses back from the DIMM to the host system will occur within 2 microseconds or less and conform to the requirements of the deterministic DRAM bus.

Figure 3 – Transaction Diagram for a Write Operation

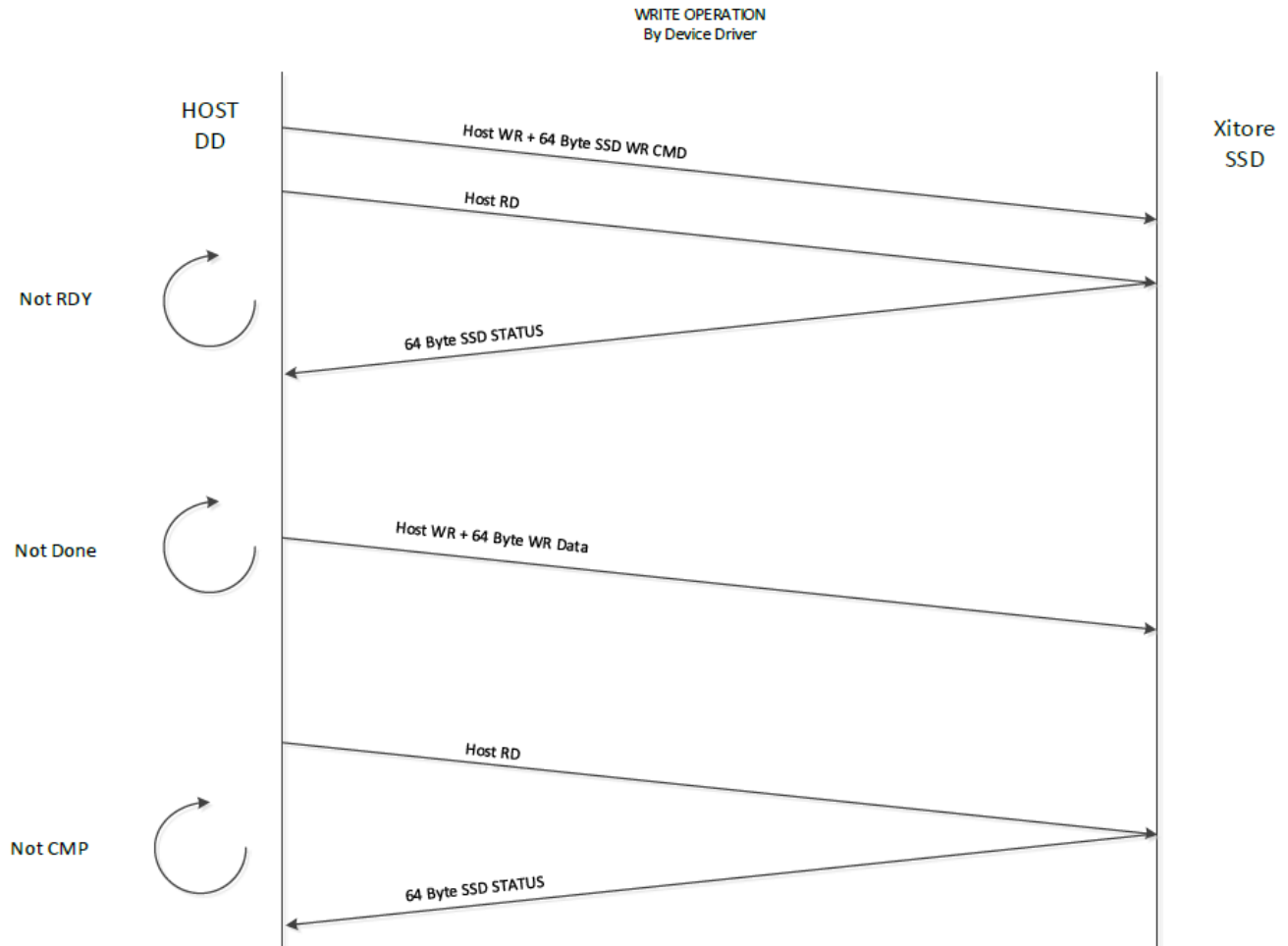


Figure 4 – Transaction Diagram for a Read Operation

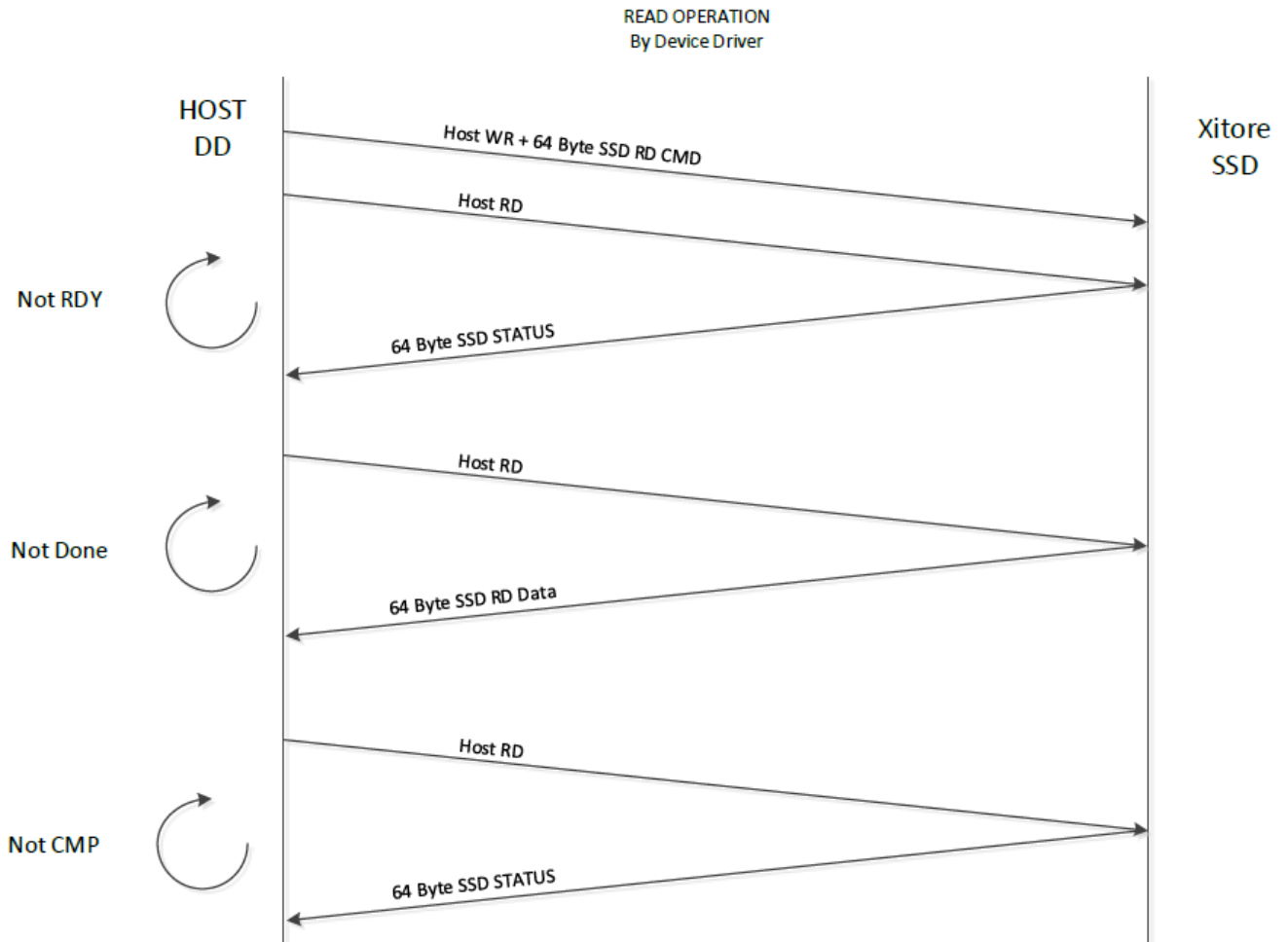
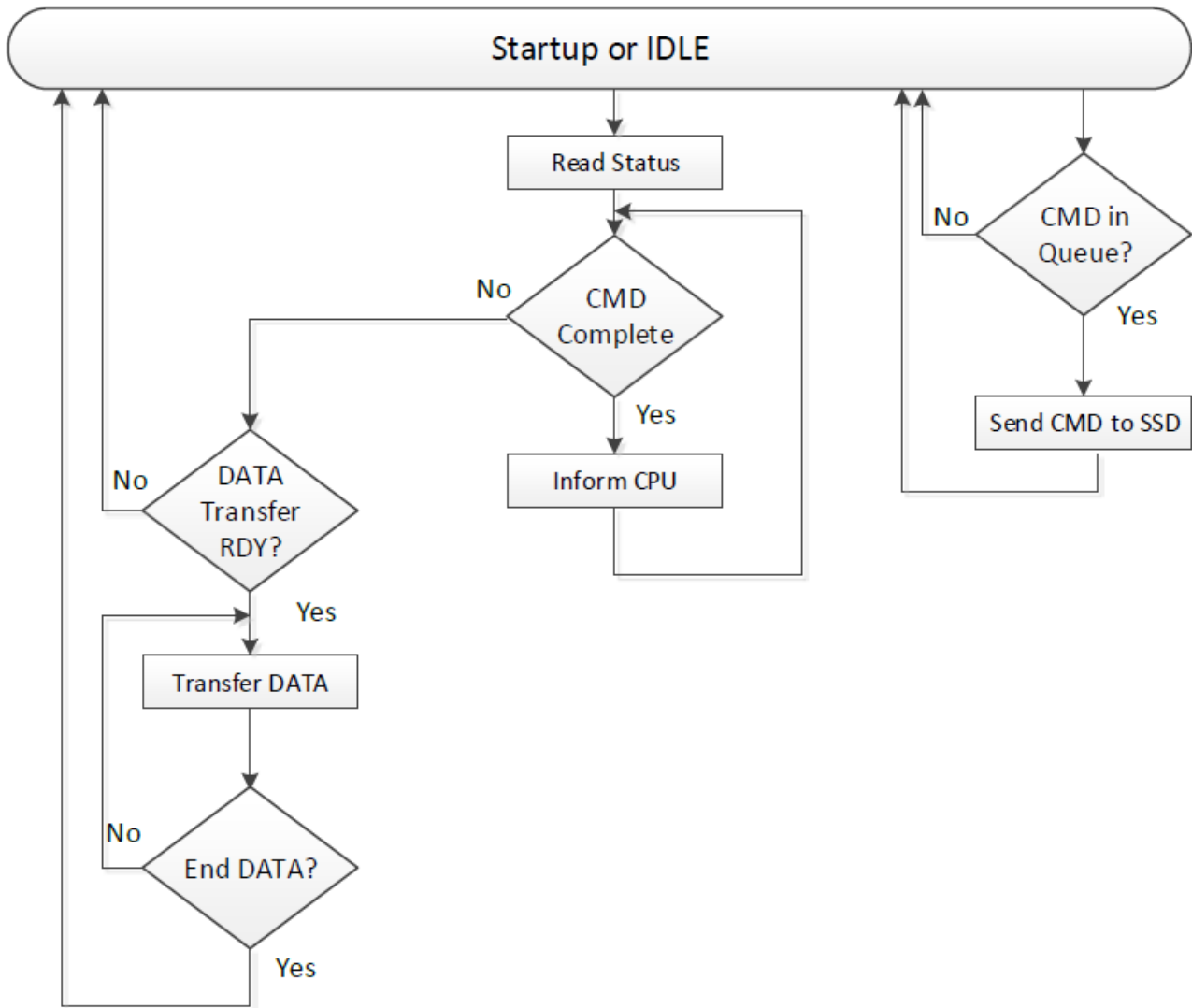




Figure 5 – Flow Chart of Driver Logic



## Future Improvements

Future product releases will continue to provide performance increases. When the new media technologies become mature enough to replace the NAND, the NVDIMM-X architecture will be the most logical interface to use. Future new media will be faster than NAND and it would not make sense to implement them with slower interfaces like NVMe, SAS, or SATA. However, even though these new media technologies are faster than NAND, most of them are still slower than DRAM. So it would still make sense to utilize the NVDIMM-X architecture that can provide for the on-module DRAM cache.

Another improvement will be to provide for byte mode access to the NVDIMM-X module. This can be done, but may require some additional enhancements to allow for variable access timing on the DRAM bus. Architectural enhancements will be needed in the interface between the DIMM and the microprocessor to enable this and discussions are already underway within the standards committees to determine the best implementation path. Providing byte level access will provide even more performance improvements because it will enable a direct access mode that will allow the application program to read and write directly to the DIMM, removing any inefficiencies introduced by the operating system or driver.

## Summary

The NVDIMM-X is a platform that will allow for the continued evolution of storage technology. In the initial implementation it will provide an immediate boost in performance over current solid state storage technologies using existing technology. But just as important, it will pave the way for use of next generation media and provide for the use of the most cost effective technology at any given time.

### Disclaimer

Xitore may make changes to specifications and product descriptions at any time, without notice. The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Any performance tests and ratings are measured using systems that reflect the approximate performance of Xitore products as measured by those tests. Any differences in software or hardware configuration may affect actual performance, and Xitore does not control the design or implementation of third party benchmarks or websites that may be referenced in this document. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to any changes in product and/or roadmap, component and hardware revision changes, new model and/or product releases, software changes, firmware changes, standard updates or the like. Xitore assumes no obligation to update or otherwise correct or revise this information. All trademarks are the property of their respective owners.

XITORE MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION. XITORE SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL XITORE BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF XITORE IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

© Xitore, Inc., 2016. All rights reserved.

[www.xitore.com](http://www.xitore.com)