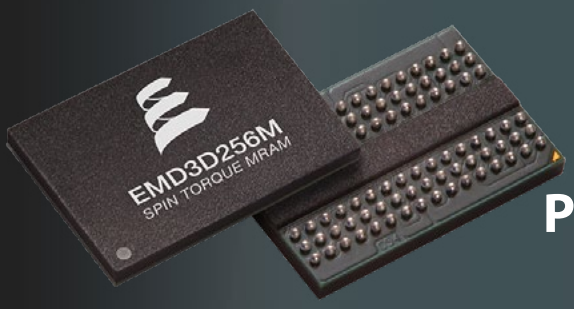
EVERSPIN® TECHNOLOGIES

App Note



Utilizing Everspin STT-MRAM in Enterprise SSDs to Simplify Power-fail Protection and Increase Density, Performance, and Endurance

Introduction

As enterprise solid state drives (SSDs) continue to push the envelope in terms of system performance and smaller form factors, SSD solution providers are facing greater challenges. The need to increase density, endurance, performance and add significant new functionality, while also continuing to protect data-in-flight from power failures.

Next generation SSDs will rapidly grow to 32TB and beyond by using more flash channels with faster interface speeds, and higher density flash devices. If a traditional architecture of a controller with DRAM working memory is employed, this significantly increases the need for energy storage to provide power fail protection, which in turn reduces space available for the storage array in a fixed form factor. These next generation devices will also require new functionality including advanced CMB buffering, in-line encryption, deduplication, and compression.

This application note explores the SSD architecture benefits of employing Everspin Spin-transfer Torque Magnetoresistive Random Access Memory (STT-MRAM) and DRAM residing on the DDR bus of the SSD controller and using STT-MRAM as a replacement for DRAM to provide a high speed, non-volatile cache. These applications enable designs to:

- Reduce write amplification in the flash array
- Increase performance and reduce latency
- Reduce power fail energy storage without impacting data protection
- Provide advanced "on device" persistent CMB buffering
- Provide non-volatile storage for advanced applications including deduplication and compression

Current Enterprise SSD Controller Architecture

An SSD has three key components: controller, volatile DRAM working memory, and the nonvolatile flash memory array. Because of the limitations of flash performance and endurance, a number of advanced algorithms are needed to manage the flash array in order to create a robust storage subsystem. These algorithms include garbage collection, error correction, and wear leveling. The controller is an embedded processor that manages accesses to the flash array through the use of these algorithms. The DRAM working memory is necessary to buffer reads and writes from the host as well as to enable quick access to the flash translation table (FTL) and metadata.

A key difference between enterprise class SSDs and consumer class SSDs is that in enterprise class SSDs, all data-in-flight is protected in the event of power loss or interruption. In order to accomplish this, a power fail detection and isolation circuit, that includes hold up energy storage, works in conjunction with logic in the SSD controller to flush all volatile data not committed to the nonvolatile array in the event of a power failure. Because of the relatively slow write speed of NAND Flash, the controller and memory system must be held up for hundreds of milliseconds.

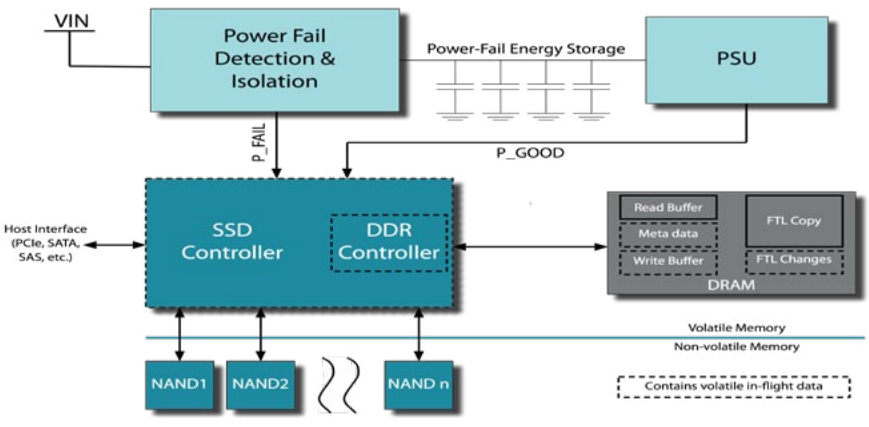


Figure 1: Conventional SSD Architecture with Power Fail Protection

Heterogeneous DDR Controller Architecture for SSDs

As a solution to overcome the limitations of the existing NAND Flash based SSDs, Everspin offers STT-MRAM with ST-DDR3 and ST-DDR4 interfaces that can be used to improve the system performance and reliability of SSDs by providing high-speed nonvolatile storage for in-flight data. By adding STT-MRAM to complement or replace the volatile DRAM on the SSD controller's DDR bus (Figure 2), the SSD controller can now utilize this high speed nonvolatile memory for the write buffer and any other critical data in flight that was previously volatile.

For an enterprise SSD, the design of the power management system is important. The system must detect power failing, isolate the drive from the host and hold up the drive with enough energy storage to allow any in-flight data to be committed to non-volatile memory in order to guarantee data integrity. The holdup energy required to accomplish this is directly proportional to the amount of data-in-flight, the speed of the non-volatile memory and the power consumption of the system. The amount of time this hold up energy storage provides can be considered the power fail window or the time available to store unprotected data before the holdup energy is exhausted.

In order a to support a heterogeneous DDR architecture comprised of separate ranks of different memory types, the DDR controller contained in the SSD controller ideally needs to support handling differing timing and addressing requirements for STT-MRAM for optimal performance. The SSD controller must also employ additional logic to properly manage the small amounts of resident data in flight within the DDR controller buffers to ensure the pipeline is flushed to the STT-MRAM and any open pages in the STT-MRAM are closed prior to power loss.

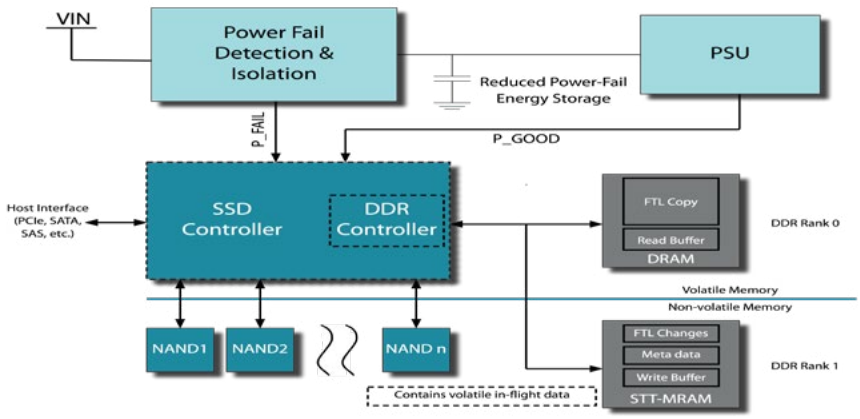


Figure 2: Hybrid DDR/STT-MRAM SSD Architecture with Power Fail Protection

The in-flight data present in the volatile memories is on the order of tens of megabytes in today's SSDs due to the balance of flash performance, power consumption and energy storage. By employing STT-MRAM in the system, the majority of the data in flight can now be buffered in this persistent memory; thus greatly reducing the amount of unprotected data in flight.

The following benefits can be realized with this architecture:

Benefit: A high-speed NV write cache made possible by STT-MRAM and cooperative data management can reduce write amplification to NAND, reducing the amount of overprovisioning needed

By using STT-MRAM as a high speed nonvolatile cache in an SSD, added benefits can be imparted on the flash array. A large high-speed nonvolatile cache allows for write amplification to be reduced by allowing more uncommitted data to be coalesced before writing to the flash array. This in turn provides more endurance for the overall SSD as writes can be organized and managed more efficiently in the larger high speed nonvolatile buffer that STT-MRAM provides. Simulation has shown that write amplification can be reduced in some cases by up to 38% and a reduction of the copy out of write data by garbage collection of up to 51.8%. This significantly increases overall drive endurance or significantly reduces the amount of over provisioning needed to meet drive performance targets, hence, increasing the effective drive size and reducing the overall cost in terms of $/GB. (Source: E. Lee, J. Kim, H. Bahn, S. Noh "Reducing Write Amplification of Flash Storage through Cooperative Data Management with NVM" MSST 2016.)

Benefit: STT-MRAM improves reliability and manufacturability by greatly reducing the need for large energy storage devices like supercapacitors or batteries

A typical system implements the energy storage with banks of capacitors or batteries that consume a large amount of PCB area and enclosure volume. These energy storage devices have a relatively high sensitivity to temperature which greatly reduces their lifetime and requires a large amount of derating to use in an enterprise environment.

SSD architectures that use MRAM greatly reduce the amount of energy storage needed, thus freeing up the area and reducing the component count. This improves manufacturability by reducing points of failure and streamlining assembly by removing the solder reflow constraints imposed by batteries or super capacitors.

Benefit: STT-MRAM increases burst write performance with a larger write buffer

Because of the speed and non-volatility of STT-MRAM, the write buffer can be significantly increased which provides much higher burst performance in the system. STT-MRAM can be written to much faster than the host interfaces of today's SSDs. Also, the effect of having a large low latency write buffer has been shown to reduce the variation in latency and provide significant QOS improvements. This performance increase is also seen at Queue Depth 1 (QD=1) where most of the current SSDs struggle and STT-MRAM solutions excel.

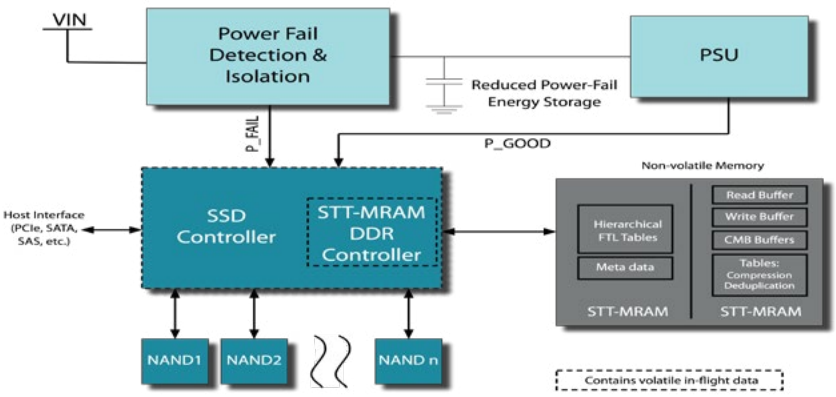


Figure 3: STT-MRAM Replaces DDR in Advanced SSD Architecture

Advanced SSD Architecture: STT-MRAM Replaces DRAM

As STT-MRAM density increases, STT-MRAM has the opportunity to completely replace DRAM in SSD architectures (Figure 3) allowing for even more advanced capabilities that are only possible with a large amount of persistent memory that operates at DRAM speeds.

Smaller SSD's, these typically use flat FTL tables which required approximately 1GB of ST-DDR for each Terabyte of flash storage. The trend for larger SSDs is to use multi-level hierarchical tables to reduce the table size held in DDR. However these tables do still have to be periodically written to cache along with other controller metadata which complicates the controller design. These hierarchical solutions do provide the added advantage of faster recover time upon power up since the controller does not have to wait until the whole flat table is loaded from flash.

The implementation of advanced functions including deduplication and compression requires maintaining and storing the required tables in ST-DDR. These tables also need to be maintained thru power interruption and consequently have to be flushed to flash which adds further complexity to the controller design.

Persistent CMB Buffering

As described in the NVMe Specification 1.2+, an optional feature is a Controller Memory Buffer (CMB) which is a region of general purpose read/write memory on the controller that may be used for a variety of purposes. The controller indicates which purposes the memory may be used for by setting support flags in the CMBSZ register. Submission Queues in host memory require the controller to perform a PCI Express read from host memory in order to fetch the queue entries.

Submission Queues in controller memory enable host software to directly write the entire Submission Queue Entry to the controller's internal memory space, avoiding one read from the controller to the host. This approach reduces latency in command execution and improves efficiency in a PCI Express fabric topology that may include multiple switches.

Similarly, PRP Lists or SGLs require separate fetches across the PCI Express fabric, which may be avoided by writing the PRP or SGL to the Controller Memory Buffer. Completion Queues in the Controller Memory Buffer may be used for peer to peer or other applications.

For writes of small amounts of data, it may be advantageous to have the host write the data and/or metadata to the Controller Memory Buffer rather than have the controller fetch it from host memory.

A controller memory based queue is used in the same manner as a host memory based queue – the difference is the memory address used is located within the controller's own memory rather than in the host memory. The Admin or I/O Queues may also be placed in the Controller Memory Buffer. The host ensures that all writes to the CMB that are needed for a command have been sent before updating the SQ Tail doorbell register.

Implementing a CMB with persistent memory allows the completion queue to be immediately updated once the write data has been submitted to the CMB and thus lowers latency.

The following additional benefits can be realized with this architecture:

Benefit: STT-MRAM eliminates the need to periodically flush dirty hierarchical FTL data to flash

STT-MRAM based solutions simplify flash controller design since large SSDs using hierarchical FTL tables will no longer have to periodically flush dirty FTL tables and other data to flash.

Benefit: STT-MRAM significantly reduces start up recovery time

Since the hierarchical FTL tables are stored directly in STT-MRAM they are available immediately at power up.

Benefit: STT-MRAM provides capability to implement non-volatile CMB buffering

CMB buffering inside the SSD device using STT-MRAM provides a non-volatile solution for high speed, low latency buffering required by RDMA devices making the SSD suitable in a wide range of low latency applications.

Benefit: STT-MRAM provides capability for storage of other non-volatile tables required to implement advanced functionality

STT-MRAM offers non-volatile tables to be stored for advanced functions including deduplication, and compression. The implementation complexity of this functionality is significantly reduced since the tables are persistent and it is no longer necessary to keep additional copies in Flash memory.

Conclusion

Everspin's STT-MRAM provides a superior alternative to DRAM-based SSD controller architectures in next generation enterprise SSDs. STT-MRAM allows SSD designers to optimize footprint, performance, endurance, and reliability, while at the same time reducing complexity and enabling advanced functionality.

STT-MRAM enables a larger write buffer in SSDs with less hold up energy storage. By greatly reducing the amount of energy storage required for power fail protection, the energy storage can be implemented with highly reliable aluminum or tantalum capacitors as opposed to supercapacitors or batteries. Eliminating the need for temperature sensitive and lifetime limiting supercapacitors and batteries improves the reliability and manufacturability of the SSDs. This removes the need for additional manufacturing constraints or potential field service issues that arise due to use of supercapacitors and batteries.

Beyond power fail protection, the larger write buffer implemented in low latency STT-MRAM can greatly improve burst write performance during normal operation by allowing much more data to be buffered before requiring back pressure to the host. The increased write buffer size also allows for more write coalescing, thereby reducing write amplification and improving the flash array's endurance.

STT-MRAM provides non-volatile storage of hierarchical FTLs and simplifies the controller design while reducing recovery time following a power fail. Use of STT-MRAM provides for persistent storage for local low latency CMB implementations besides supporting compression and deduplication tables and at the same time easing the task of adding this broad capability to the SSD drive.

Everspin's STT-MRAM is now in production and ready to accelerate the performance, reliability and density of SSDs.

How to Reach Us:

www.everspin.com

E-Mail:

support@everspin.com

orders@everspin.com

sales@everspin.com

USA/Canada/South and Central America

Everspin Technologies

5670 W. Chandler Road, Suite 100

Chandler, Arizona 85226

+1-877-347-MRAM (6726)

+1-480-347-1111

Europe, Middle East and Africa

support.europe@everspin.com

Japan

support.japan@everspin.com

Asia Pacific

support.asia@everspin.com

Everspin Technologies, Inc.