Executive Summary

Violin Memory technologies offer powerful, reliable, and economical solutions to the data storage and application performance that federal government agencies face today.
Introduction

Violin has custom engineered our hardware and software flash Memory Array components to function as mission critical primary storage.

Violin Memory has approached the challenges posed by Federal Government data center primary storage in a revolutionary way. Unlike other companies rushing to enter the solid state storage marketplace by using commodity components to arrive sooner, Violin identified the underlying storage-related challenges faced by our Federal customers exploring virtualization, working with “Big Data,” and needing to accelerate their mission critical applications. We made these specific technical challenges our design objectives, then custom built an integrated hardware and software solution to solve them, rather than cobbling together off-the-shelf components. The result is a purpose-built flash memory solution with the performance, reliability, and cost-effectiveness to perform as primary storage in the most mission critical Federal environments.

The key to building one of the most successful and powerful primary storage solutions on the market was our decision to custom-engineer both the software and hardware components. This purpose-built approach allows a level of deep integration between our Violin Memory Operating System (vMOS) software and Flash Memory Fabric hardware that off-the-shelf integration cannot achieve. In this white paper we introduce the major vMOS and Flash Memory Fabric components, illustrate their unique solution features, and highlight some of the advantages to Federal organizations of our revolutionary purpose-built chip-to-chassis engineering approach.

Flash-Based Storage in the Datacenter

Using NAND flash in enterprise storage environments requires purpose-built technology that can manage reliability and performance challenges.

The core memory technology used by Violin is NAND Flash: Single Level Cell (SLC – the highest performance, highest endurance, highest cost) and Multi-level Cell (MLC – able to store 2 bits per cell, this lowering costs compared to SLC). MLC technology stores two bits per cell and provides the density and cost structure that enables the substitution of flash for hard disk drives (HDD) in many applications.

However, there are several issues with flash devices (chips):

- Writes (“flash programs”) are sequential and relatively slow (MLC = 900μsec).
- Erases require a whole section (“flash block”) to be erased and take considerable time (MLC often requires over 3,000μsec). During this time, nothing can be read or written.
- Reading can be very fast (<100μsec) and either random or sequential. However, only a single page can be read at a time.
- Pages can be damaged by repeated reading and handling.
- A single page can be erased only so many times before it wears out.

All flash supports the same basic set of operations: Read, Write, Erase. The key take-away from the table above is that Erase operations are relatively very slow: one full millisecond for SLC, and three milliseconds for MLC.

<table>
<thead>
<tr>
<th></th>
<th>Read Ops</th>
<th>Write Ops</th>
<th>Erase Ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC</td>
<td>25μs</td>
<td>250μs</td>
<td>1,000μs</td>
</tr>
<tr>
<td>MLC</td>
<td>50μs</td>
<td>900μs</td>
<td>3,000μs</td>
</tr>
</tbody>
</table>

NAND Flash Read, Write, and Erase Latency
MLC. Plus MLC typically has less than one-tenth the Erase cycles of SLC before it begins to degrade. This is the physics of flash and the implication is that delivering sustainable, low latency data storage using flash-based media requires some special handling of Erase operations.

Read/Write Speed: Write Cliffs

These issues are most obvious when measuring the sustained random Write performance of certain flash-based devices. The performance is initially good when the pages are clean, but drops dramatically over a “Write Cliff” when pages have to be recycled in a process called Garbage Collection (or grooming). The table below illustrates how both HDD form factor solid-state storage devices (SSD) and in-server PCIe cards suffer from significant Write Cliff performance degradation.

![Write Cliff Performance of Sample SSD and PCIe Products](image)

When SSDs and PCIe cards hit their Write Cliffs they begin cycling through already programmed flash dies. At this point, new Writes get stuck behind Erase operations, causing a dramatic drop in overall performance, up to 60%. The Write Cliff phenomenon doesn’t just affect Write latencies. Because of the way flash dies work, Erase operations also get in the way of Read operations on neighboring blocks. So Erase operations and the massive latency spike that come with them can seriously degrade both Read and Write latencies.

Traditional storage solutions address performance issues by short stroking, wide striping, and adding a few SSDs to an old array, among other techniques. Also, read-only flash caches may be added to existing storage arrays, in the network in front of the array, or in servers in front of the network. But these awkward solutions, like an old turtle uncomfortably strapped to a roller blade, are not fast or economical.
The Violin Flash Memory Fabric

Performance and reliability are engineered into the Violin Flash Memory Fabric, from the chip to the chassis.

For NAND flash to be viable in Federal data center applications, it requires a very different set of attributes compared to a PC or consumer device. Sustained and predictable performance is required for Federal data centers. Violin Memory has solved the inherent “behavioral issues” of both SLC and MLC NAND flash through innovative design and chip-to-chassis integration of hardware components known as the Flash Memory Fabric and software called the Violin Memory Operating System. Our Flash Memory Fabric architecture in particular enables thousands of flash devices to operate efficiently together, masks the chip level issues, and delivers reliable and sustained system performance:

- Writes are random and fast (20µsec).
- Erases are simple, fast, and hidden from the user.
- Reading is accelerated through parallel access to billions of pages.
- The page remains in crisp digital form with multiple copies for reliability.
- Repeated reading of the pages has no impact on the original.

Violin’s Flash Memory Fabric is woven from multiple layers of innovative hardware technologies, the result of an all silicon system approach with patented flash optimization algorithms implemented in hardware, operating at line rate.

- At the system’s core lies a resilient, highly available deep mesh of thousands of flash dies that work in concert to continuously optimize performance, latency, and longevity.

Violin Intelligent Memory Modules (VIMM) organize this mesh of individual dies into intelligent flash management units. VIMMs provide a hardware-based Flash Translation Layer with Garbage Collection, wear leveling, and error/fault management.

VIMMs are integrated into the patented Violin Switched Memory (vXM) architecture, designed from the ground-up for power efficiency and performance.

Finally, VIMMs and the vXM layer work in conjunction with vRAID, Violin’s patented hardware-based RAID algorithm specifically designed to increase reliability and reduce latency.

Collectively, the mesh of flash dies organized into VIMMs integrated into vXM and overlain by vRAID make up the Violin Flash Memory Fabric.

Violin’s Flash Memory Fabric works with the Violin Memory Operating System to enable profoundly reliable, highly available storage at the speed of memory that offers multiple industry leading benefits:

- **Spike-free Low Latency** - The Flash Memory Fabric delivers spike-free and predictable latency that is 95% lower than HDD and 70% lower than SSD and PCIe card solutions.

- **High Bandwidth** - A single Violin flash Memory Array supports over 4000 flash devices and 500 independent flash interfaces. This provides the bandwidth needed for outstanding flash performance with four times greater bandwidth than other storage systems and 50 times greater than most SSDs.

- **Extreme Reliability** – All active components of the Flash Memory Fabric are hot-swappable for enterprise grade reliability and serviceability.

The Flash Memory Fabric in Violin 6000 Series flash Memory Arrays is composed of up to 64 VIMMs and four active/active vRAID Control Modules (VCM). Except when the system is configured as PCIe direct attached, all inputs/outputs (I/O) are processed through Memory Gateways before being handed off to the VCMs, which
implement vRAID and orchestrate flash management operations across all the flash dies. Background Garbage Collection (GC) performs various flash optimization tasks, one of which is proactively erasing flash dies so they are ready for new incoming Writes. Performing GC involves reorganizing data placement and erasing flash dies within VIMMs. VCMs ensure that Garbage Collection is scheduled in such a way that there is always at least one VIMM protection group available for incoming Writes. Thus Violin flash Memory Arrays avoid flash Erase operations for incoming Writes, because we choose where to place these Writes.

For incoming Reads, we do not have the luxury of choosing where these reads can occur. So VCMs ensure that only one VIMM in each protection group is performing GC at any given point; this leaves four VIMMs available so that vRAID rebuilds can be performed to serve any Read request without delay. We call this “Erase hiding.” And because these algorithms are all implemented in the Flash Memory Fabric at line rate, we lead the industry in our ability to deliver sustainable spike free microsecond latencies for mixed workloads.

**Toshiba NAND Flash Memory Chips**

At the heart of the Flash Memory Fabric lie thousands of individual Toshiba dies. Violin and Toshiba have formed the Violin/Toshiba Strategic Supply and Roadmap Agreement, ensuring that Violin always has high priority access to most cost effective supplies of Toshiba flash chips. Our engineers collaborate to develop the most reliable and high performance flash technologies possible. Also, our unique roadmap sharing arrangements provide Violin engineers with deep understanding of Toshiba flash architectures, allowing us to optimize the performance and reliability of our Flash Memory Fabric through perfect
integration of our VIMM, vXM, and vRAID layers.

Violin Intelligent Memory Modules are the core building block of the Flash Memory Fabric, designed from the ground up by Violin Memory to be the highly resilient, hot swappable technology implementing our proprietary fabric level flash optimizations algorithms.

A single VIMM contains up to 128 flash dies. A 64 VIMM Flash Memory Fabric thus contains more than 8000 flash dies, managed as a single system by vRAID in the VCMs. Optimizing flash endurance, data placement, and performance across such a large number of dies is the key to Violin Memory’s unique ability to deliver sustainable performance, ultra low latency, and industry leading flash endurance. While a commodity SSD contains a flash controller optimizing flash across tens of dies within the SSD, the Flash Memory Fabric can leverage thousands of dies within which to implement optimization decisions.

Data is written and read to and from VIMMs. Each VIMM operates its own instance of a Flash Translation Layer. Data is written to VIMMs using the Logical Page Address (LPA), which assigns the page to a Physical Page Address (PPA) within the flash of the VIMM. Metadata is used to map between logical addresses and physical addresses. Each VIMM includes:

- High-performance logic-based flash memory controller
- Management processor
- DRAM for metadata
- NAND flash for storage.

VIMMs are designed to enable the scalability of large arrays of flash memory. The advantages of this architecture over a simple PCIe card are significant:

- Integrated Garbage Collection for sustained write performance
- Low latency access to DRAM metadata and flash memory
- Safe access and local storage of metadata for fault recovery
- Integrated monitoring and management of flash memory health
- Distributed Error Correction Code (ECC) correction for maximum bandwidth
- Hot swap and redundancy management
- 3 port design so that single port failures do not impact data access. Unlike most SSDs and PCIe cards, a failed flash device does not cause a VIMM to lose data or be removed from service. ECC and vRAID protection are used to manage the data.

VIMMs work in conjunction with Violin Memory’s patented vRAID to ensure a very high degree of system availability and data integrity using multiple techniques, including:

- **Data protection**: Robust ECC and Cyclic Redundancy Check (CRC) algorithms detect and correct bit errors in the system.

- **Data validation**: Extra data is stored with each block of flash so that invalid data is detected rather than passed on.

- **Data scrubbing**: All data in the system is read on a weekly basis and scanned for errors. Any errors found are then repaired. Violin does this without any noticeable impact to user performance. It greatly reduces data loss rates and increases data endurance.

- **Flash wear leveling**: The Violin Array distributes data Reads and Writes evenly to all the flash devices in all the modules of a system. No specific Logical Unit (LUN) is tied to a specific module and hence active LUNs do not wear out specific flash devices.
• **Flash monitoring:** All Read, Write, and error statistics are captured and reported. VIMMs behaving below specification are automatically removed from service and the error events logged. The data from that VIMM is moved to a spare VIMM using the vRAID algorithm to rebuild the data. This is done in the background without administrative intervention or any significant impact on access to user data. The system may have one to four spares and hence replacement of the module is not an urgent requirement and may be hot-swapped monthly or quarterly.

**vRAID: Hardware-based Data Protection and Performance Enhancement**

Other solid-state storage solutions and architectures, such as SSDs and PCIe cards, use processors and software to perform RAID, page mapping, and Garbage Collection. Violin implements these functions in hardware to reduce latency and dramatically increase sustained random Write input/outputs per second (IOPS) from less than 10,000 to more than a million.

Low-latency flash vRAID, Violin’s patent-pending flash technology designed specifically to enhance NAND flash system performance, provides full RAID data protection and a fundamentally more efficient and higher performance solution. Existing RAID 5 and 6 solutions rely on Read-Modify-Write operations that are unsuited to flash. Unlike inefficient RAID 1 (50% efficient) solutions, Violin’s vRAID enables 80% usable capacity and bandwidth.

vRAID guarantees spike-free latency under load by making sure there aren’t any Reads blocked by Erases. Notably, the microsecond latency of the Violin 6000 Series flash Memory Array is 80% lower than Tier-1 storage cache (DRAM) and significantly improves metrics such as file Read and Write, response, and query times.

vRAID delivers Fabric level flash optimization, dynamic wear leveling, and advanced ECC for fine grained flash endurance management, as well as Fabric orchestration of Garbage Collection / grooming to maximize system level performance. vRAID also protects the system from VIMM failures.

vRAID is implemented within the vRAID or Violin Control Module. Each VCM manages 15 VIMMs (4 VCMs for a total of 60 VIMMs plus 4 hot spares in a fully populated 6000 Series Array). Note that when VCMs fail, VIMMs are automatically reallocated across the remaining VCMs in the system such that a single VCM is actually able to manage all VIMMs in a system, though at the cost of degraded performance. For the purpose of data placement and protection, VIMMs are organized in groups of 5: any incoming 4KB Write is dynamically allocated to a VCM which itself dynamically selects the most appropriate group of 5 VIMMs to handle the Write. 4KB is written across 5 VIMMs as 4 * 1KB of data, plus 1KB of parity data for protection in case of a VIMM / flash failure.
This data placement algorithm is applied across the entire Flash Memory Fabric. Data is dynamically placed across the entire set of VIMMs based on real time flash conditions. The net result is automated, granular, LUN-wide striping across all available flash dies in the Fabric (all 8000 of them). This wide striping happens at all levels in the Fabric, across VCMs, across VIMMs, and across flash dies inside each VIMM. All operations are implemented in hardware, at line speed, ensuring that any data can be read from the Array with the lowest levels of latency. As a result, any LUN in the Array gets access to the full system bandwidth, by default. The system is Simply Fast.

The combination of VIMMs and vRAID resolve issues of flash errors, reliability, and wear and deliver an enterprise-grade flash system with a far greater life expectancy than typical HDD arrays. For example, Violin flash Memory Arrays using SLC NAND flash can sustain a write rate of 8TB/hour – greater than 2GB/s – for over 10 years. The truth is, in all the time that Violin Memory has been shipping flash arrays, no customer has worn out a VIMM. They have failed due to component failures, but typical enterprise use has resulted in wear rates of significantly less than 10% per year.

The real test for storage systems comes from handling a combination of media errors and module failures. As an example, 30 SATA disk drives in a RAID-5 configuration have long RAID rebuild times and high error rates that lead to a Mean Time to Data loss (MTDL) of less than a few years. The MTDL of a VIMM is estimated at 200 years, about 20 times higher than a rotating HDD. The RAID rebuild times for a Violin Array are typically between one and 24 hours, depending on user load and memory type. The RAID-5 rebuild time for large HDDs is measured in days, especially under load.

Violin Switched Memory

Violin has developed the industry’s first switched memory architecture. This patent-pending architecture is called Violin Switched Memory (vXM) and is implemented within our Flash Memory Fabric. With vXM, each VIMM is part of a switched array that supports large topologies and fault tolerance while enabling multiple memory types. Unlike other memory interconnects, vXM was designed from the ground up for power efficiency and performance. vXM is the industry’s first memory solution developed specifically for applications with large datasets, including large databases and image, video, scientific, and web content. These large dataset applications can leverage multiple vXM technology benefits:

- Unmatched scalability
- Highest performing I/O
- Memory fault tolerance
- Ultra-green power savings
- Flash and DRAM flexibility.

Previous memory solutions have not scaled because they relied on classical linear bus topologies. Bus speeds limit the number of registered devices that can be supported in a single channel to fewer than eight and, as speeds increase, to just one or two. But switched memory architectures do not suffer from these limitations. Memory architectures for enterprise-class applications and the networking industry are already benefiting from the evolution from memory buses to switched memory architectures that can truly meet their performance, scalability, manageability, and fault tolerance requirements.
Violin Memory Operating System

vMOS controls system operations and management within our Memory Arrays as well as providing an ever-expanding suite of data management features.

The Violin Memory Operating System runs on all 3000 Series and 6000 Series flash Memory Arrays, complementing their unique Flash Memory Fabric hardware architecture. vMOS delivers System Operations, Systems Management (including system wide flash optimization), and Data Management functionality, leveraging hardware accelerated logic whenever appropriate.

vMOS accomplishes its many tasks within the Violin flash Memory Array’s 100 microsecond system latency while maintaining its capability to perform a wide range of data management functions at any bandwidth and capacity scale. Key to these extraordinary capabilities is the deep software and hardware integration within the Memory Array ecosystem. This level of integration allows vMOS to take advantage of Violin’s patented hardware acceleration engines such as our switched memory architecture and unique flash-optimized vRAID protection mechanisms.

Reliable Data Protection

Enterprises depend on Violin flash Memory Arrays to store and manage the most mission critical data. vMOS offers enterprises a wide range of data management capabilities to reliably protect their data without affecting the impressive performance of Violin flash Memory Arrays. vMOS data protection features include space optimized snapshots, instant clones, asynchronous replication, and encryption.

Efficient Storage Optimization and Flexible Storage Connectivity

vMOS enables storage administrators to power more mission-critical applications with Violin’s high performance by driving greater storage efficiencies. Seamless integration of functionality across multiple layers of the Violin flash Memory Array enables vMOS to deliver hardware accelerated and flash optimized primary storage features such as thin provisioning, deduplication, and compression. vMOS also provides active-active storage connectivity through a pair of highly available Memory Gateways and integrates with all leading multi-pathing solutions. vMOS provides flexibility in storage connectivity with several high bandwidth options, including Fiber Channel, iSCSI, Infiniband, and PCIe direct connectivity.

Comprehensive Management

vMOS empowers administrators to monitor, manage, and configure Violin flash Memory Arrays anywhere, any time. The simple, elegant GUI offers a comprehensive view of the entire system, from provisioning and reporting to on-going management. vMOS provides administrators with access to all elements of the Array, enabling full visibility into the system. With support for multiple browsers and an iPad application, as well as integration into third party management environments like VMware vCenter, vMOS provides all the flexibility.
and simplicity expected of an enterprise-grade primary storage system, including several options to ease administration with live monitoring and dynamic administration capabilities such as Web-based and command line interfaces and REST API.
Proven Performance

Put Violin’s Flash Memory Fabric and vMOS together and you get a number of world records for industry standard performance benchmarks. Not from Violin Memory, but from companies like Cisco, IBM, and HP who have recognized that in order to show off the capabilities of their latest compute hardware, they must run their benchmarks on the highest performing storage systems available. They all independently concluded that Violin flash Memory Arrays was that storage.

<table>
<thead>
<tr>
<th>Date Submitted</th>
<th>Company</th>
<th>Performance (tpmC)</th>
<th>Price/tpmC</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/27/12</td>
<td>CISCO</td>
<td>1,609,186</td>
<td>.47 USD</td>
<td>Violin Memory</td>
</tr>
<tr>
<td>04/11/12</td>
<td>IBM</td>
<td>1,503,544</td>
<td>.53 USD</td>
<td>Violin Memory</td>
</tr>
<tr>
<td>03/27/12</td>
<td>ORACLE</td>
<td>5,055,888</td>
<td>.89 USD</td>
<td>Violin Memory</td>
</tr>
<tr>
<td>12/07/11</td>
<td>CISCO</td>
<td>1,053,100</td>
<td>.58 USD</td>
<td>Violin Memory</td>
</tr>
<tr>
<td>07/11/11</td>
<td>IBM</td>
<td>3,014,684</td>
<td>.59 USD</td>
<td>Violin Memory</td>
</tr>
<tr>
<td>05/23/11</td>
<td>HP</td>
<td>1,263,599</td>
<td>.51 USD</td>
<td>Violin Memory</td>
</tr>
</tbody>
</table>

Conclusion

The deep integration of the Violin Memory Operating System and Flash Memory Fabric achieves the performance, reliability, and cost-effectiveness demanded in Federal Government information system solutions.

By choosing to custom engineer our Memory Arrays from the chip to the chassis – from Toshiba flash dies through Flash Memory Fabric hardware to Violin Memory Operating System software – we solve, avoid, or dramatically improve every shortcoming of the basic flash medium itself and of less-thoroughly engineered solid state storage solutions. The benefits to Federal Government customers are clear – much higher system performance, ultra-low latency, decades-deep reliability, and easy installation and management. Perhaps most important, thanks to the efficiencies derived from our custom-engineered Flash Memory Fabric and vMOS, Violin Federal customers can get storage at the speed of memory for costs equivalent to enterprise hard disk drive arrays that soon will be obsolete – at any price.
About Violin Memory

Violin Memory is pioneering a new class of high-performance flash-based storage systems that are designed to bring storage performance in-line with high-speed applications, servers and networks. Violin Flash Memory Arrays are specifically designed at each level of the system architecture starting with memory and optimized through the array to leverage the inherent capabilities of flash memory and meet the sustained high-performance requirements of business critical applications, virtualized environments and Big Data solutions in enterprise data centers. Specifically designed for sustained performance with high reliability, Violin’s Flash Memory Arrays can scale to hundreds of terabytes and millions of IOPS with low, predictable latency. Founded in 2005, Violin Memory is headquartered in Mountain View, California.

For more information about Violin Memory products, visit www.vmem.com.

© 2013 Violin Memory. All rights reserved. All other trademarks and copyrights are property of their respective owners. Information provided in this paper may be subject to change. For more information, visit www.vmem.com