



StorExpress

The Next Evolution of Solid State Storage

**Venkata Krishnan
Lynne Brocco
Tim Miller**

Revision 1.0

June 2009

**Dolphin Interconnect Solutions
www.dolphinics.com**

High performance, data intensive enterprise applications have been experiencing an increasing performance gap, much of it due to the bottleneck presented by traditional hard disk drive (HDD) storage. This performance gap has spurred the introduction of a variety of solid state storage solutions that have recently come to market.

Flash-based solid state drive (SSD) storage solutions are now being recognized as an effective way of removing application performance bottlenecks, either by replacing or enhancing existing HDD solutions.

Examples of areas where high performance solid state storage solutions could be employed are:

- Database OLTP applications – MySQL / Oracle / PostgreSQL)
- Data warehouse and business intelligence
- Video / animation editing and processing
- Real time / low latency markets (Telco, Finance, etc.)
- eCommerce, web stores, web hosting
- Defense / security – intelligence and data analysis
- Medical records and clinical documentation
- High availability using synchronous replication)

This paper introduces the latest in flash-based storage: Dolphin's StorExpress product family, an External PCI Express Solid State Storage solution addressing high performance applications.

Solid State Storage Advantage

The advantages of solid state storage have been well covered by the industry. Key factors and benefits include:

- Orders of magnitude increase in performance over hard disk drives
- Significant reduction in power consumption and cooling costs
- Significant reduction in area requirements
- No moving parts resulting in much higher mechanical reliability
- Enterprise-level electrical reliability

Many of the SSD form factors rely on the traditional HDD form factors, bus protocols and infrastructure, resulting in a solid state version of HDD. Utilizing the HDD architecture means that performance can still be limited by disk controllers, storage and network protocols, switches, and additional SAN or NAS hardware and software. To fully realize the performance gains available, as much of this legacy infrastructure as possible can be removed. PCI Express (PCIe) solid state storage does just that.

PCI Express Solid State Storage Performance

PCIe solid state storage provides the optimum performance in terms of

- Bandwidth – measures total amount of data transferred per second
- Latency – measures the amount of time required for data to reach its target
- IOPs – measures the number of disk access I/O operations per second

Since the storage device is connected directly to PCI Express, the high bandwidth provided by the PCIe bus is directly available without being throttled by intermediate protocols. This results in random read and write bandwidths measured in GBytes/s, as opposed to Kbytes/s with standard HDDs.

Additionally, latency is greatly reduced by removing the overhead of the storage controllers, intermediate storage protocols, etc. The latency of PCIe SSD is measured in microseconds once this overhead is removed, compared to milliseconds with non-PCIe storage solutions.

The exceptional latency and bandwidth of PCIe-based storage result in leading price per performance IOPs.

However, card-based PCIe SSD solutions require that the storage be located in the server, and uses one of the server's PCIe slots for each storage card inserted. Many application solutions may require that the persistent storage be located remotely from the server, or may not have sufficient PCIe slots available for the amount of storage desired. This is where a card-based PCIe solution falls short.

StorExpress – Next Step in PCIe Solid State Storage

Disaggregated or External PCI Express Solid State Disk Storage (E-PCIe SSD) is the next evolutionary step in the advancement of solid state storage, addressing the issues of in-server PCIe SSD solutions. Dolphin Interconnect Solutions has introduced the industry's first E-PCIe SSD solution with the StorExpress family of products.

StorExpress houses PCIe flash cards in an external chassis, which attaches directly to the server using PCI Express cabling. An adapter card provides the cable to slot conversion at the server. StorExpress retains the advantage of removal of the overhead of storage controllers and intermediate storage protocols, since it still connects directly to a server's PCIe interconnect. Performance is further enhanced over single card solutions by the ability to support multiple PCIe SSD cards, adding terabytes of solid state storage to the same server without consuming additional PCIe slots.



Figure 1 Dolphin's StorExpress Product

Table 1 shows an IOPs and bandwidth comparison between StorExpress, an enterprise-level hard drive, and an enterprise-level solid state drive. The performance advantage is graphically shown in Figure 2.

Table 1 StorExpress Performance Comparison

Storage Technology	Random Read IOPs	Random Write IOPs	Read Bandwidth (MB/s)	Write Bandwidth (MB/s)
Enterprise Hard Drives (HDD) ¹	320	320	50	50
Enterprise Solid State Drives (SSD) ² – SATA	35,000	35,000	220	120
StorExpress (SSD) – PCIe	270,000	270,000	2,800	2,800

¹ VelociRaptor 10K SATA

² Intel SSD SATA

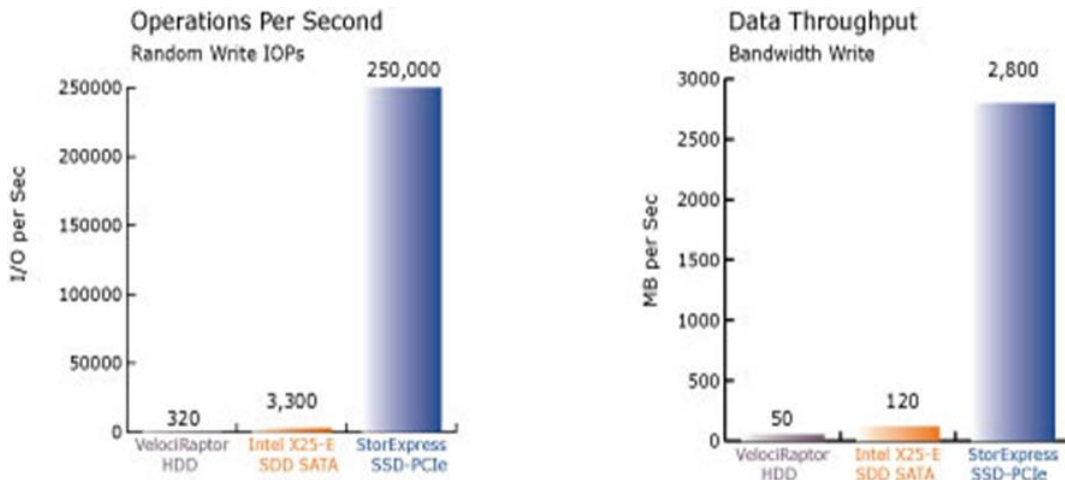


Figure 2 StorExpress Performance Graphical Comparison

Storage solutions are often scaled to try to satisfy the performance requirements of enterprise applications. This is when the cost advantage of a StorExpress solution becomes apparent. For example, Table 2 shows the cost of the hardware required to implement a storage solution that provides 250,000 IOPs of performance. The other solutions require significant scaling to achieve the same performance point of a single StorExpress solution. The StorExpress solution provides a much higher “dollar per IOPs” return.

Table 2 Solution Cost Comparison

Solution	Number of Devices	Total Solution Cost
Enterprise Hard Drives (HDD) ³	700	\$300K
Enterprise Solid State Drives (SSD) ⁴ – SATA	70	\$40K
StorExpress (SSD) – PCIe	1	\$20K

The following section lists the key features of StorExpress.

³ VelociRaptor 10K SATA

⁴ Intel SSD SATA

StorExpress Features

- External chassis
 - 19" rack mountable
 - 2U or 4U enclosure
 - Redundant power supplies
 - Supports up to two x8 PCIe connections
- Storage Capacity
 - .5TB, 1TB or 2TB with 2U chassis
 - 4TB with 4U chassis
- x8 PCI Express 1.1 host adapter
 - Single x8 connection with a single adapter
 - Dual x8 connection with two adapters
 - Supports connection to chassis-based servers & blade servers
- Server to StorExpress chassis cable connection, supporting up to
 - 10m using copper cable
 - 300m using fiber optic cable

Extend your DAS Solution with StorExpress

Although many would prefer to employ direct-attached storage (DAS) due to the desired advantages of relative simplicity and lower cost, storage area networks (SAN) has been seen as a requirement to implement server virtualization, increase utilization efficiency, and improve data integrity to support high performance applications. High performance HDDs, when placed in a RAID configuration, result in a solution with far more storage than is needed for a single server. Thus, a SAN implementation was used to spread this storage across multiple servers. StorExpress can provide the right amount of high performance storage to one or more servers running high performance applications, avoiding the need to migrate to a complex and costly SAN solution.

StorExpress provides storage in the form of multiple independent block devices. Depending on the needs of the application, this flexible block storage model can be configured by the operating system as RAID 0, RAID 1+0, RAID 5, single volume, etc.

A high performance virtualization solution can be realized using StorExpress, for example, by dedicating one or more block devices per virtual machine (VM). This enables high performance VMs and relieves the VM storage bottleneck, making StorExpress ideal for today's multi-core virtualized server platforms.

StorExpress supports a variety of configurations, providing the flexibility to scale from the simplest solution of a single server and single StorExpress chassis to a variety of multiple server and/or chassis configurations, such as:

- A dedicated StorExpress chassis for each server
- Multiple StorExpress chassis per server
- Two servers sharing a single StorExpress chassis
- Multiple servers and StorExpress chassis in a high availability solution

Single Server Configuration

Figure 3 shows the simplest configuration in which a single StorExpress chassis is connected to the server, providing up to 4TB of persistent direct-attached storage. The StorExpress chassis is connected to a PCIe adapter in the server providing up to x8 PCIe connectivity. The OS can optionally configure a RAID solution across the block devices (RAID 0 is shown in the figure).

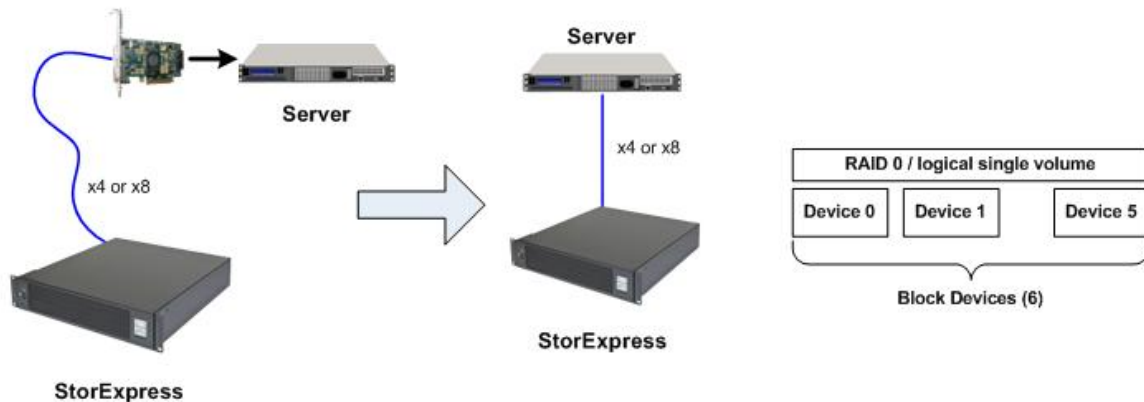


Figure 3 StorExpress Server DAS Expansion

Single Server Dual Connection Configuration

Bandwidth can be further increased by slightly modifying the single server/chassis configuration to add a second PCIe adapter to the server, as shown in Figure 4. This provides two parallel x8 connections, resulting in even higher throughputs of over 2700 Mbyte/s. See Figure 5 for an illustration of the performance that can be achieved using this dual x8 configuration.

Using the dual setup, half of the block devices can be assigned to one connection and the other half can be assigned to the second connection. As with the prior configuration, RAID 0 can be configured across the block devices. The redundant x8 connection may make a RAID 1 configuration between the two sets of block devices desirable.

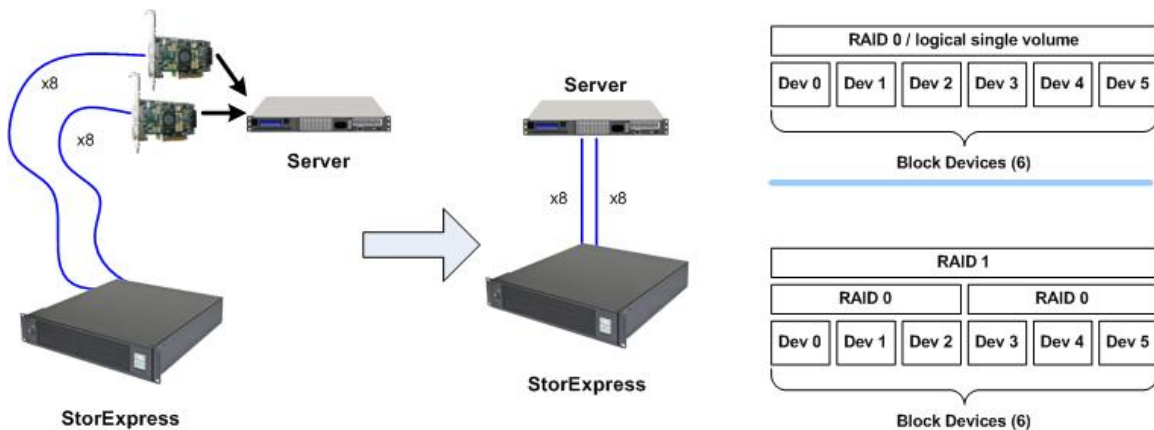


Figure 4 StorExpress Ultra High Performance Configuration

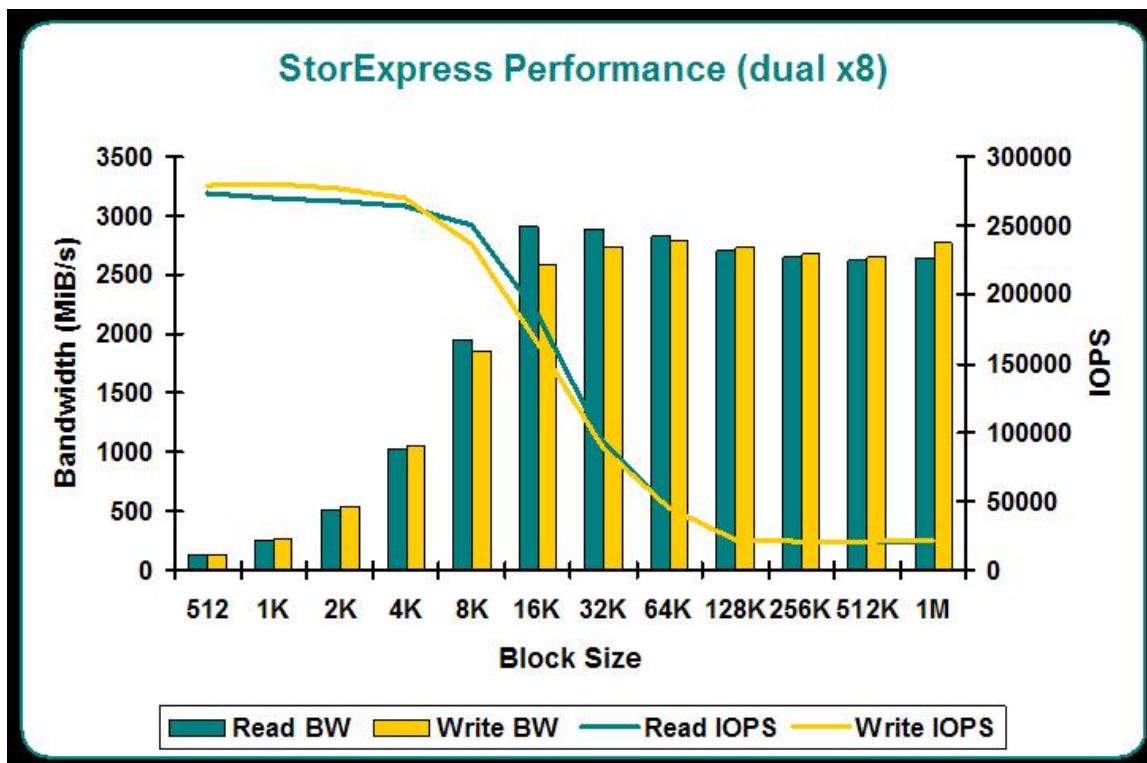


Figure 5 StorExpress Dual x8 Performance⁵

⁵ Using fio file system benchmark

Single Server Two Chassis Configuration

For additional storage up to 8TB, Figure 6 shows two StorExpress chassis connected to a single server. This configuration also provides increased redundancy, and RAID can be configured across the block devices in both chassis in the same way as with a single chassis.



Figure 6 StorExpress Single Server Scaling Out

Two Server Shared Chassis Configuration

To efficiently utilize storage, a single StorExpress chassis can be connected to multiple servers. A subset of block devices is assigned to each server. Figure 7 shows a two server single StorExpress chassis configuration.

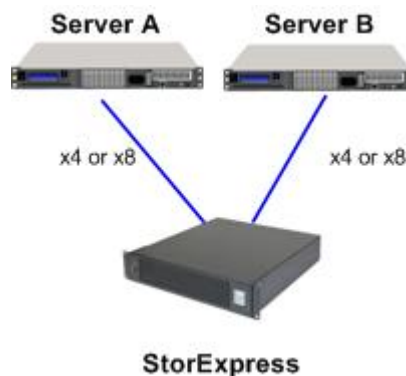


Figure 7 Shared StorExpress

Multiple Server and Chassis Redundant Configuration

The capability to connect a single StorExpress chassis to multiple servers, and multiple servers to a single StorExpress chassis, can be combined to provide a fully redundant configuration, as shown in Figure 8. Again, a subset of block devices in each chassis is assigned to each server. Thus, each server has storage allocated in both chassis. RAID can be configured across these block devices, or a synchronous replication solution can be employed. For more information about replication solutions, contact Dolphin or visit www.dolphinics.com.

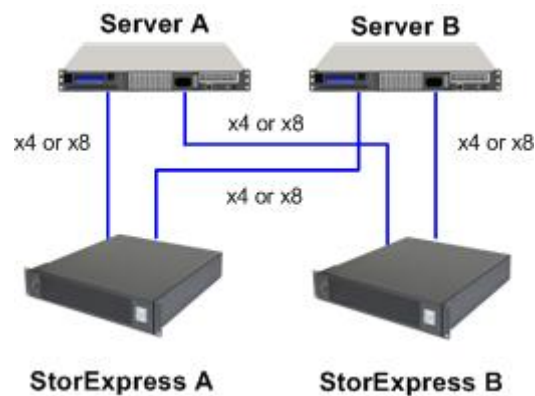


Figure 8 Redundant StorExpress Solution

Additional Capabilities

Dolphin Interconnect Solutions offers hardware and software to extend StorExpress capabilities beyond those described in this paper. These capabilities include

- Storage consolidation in a clustered environment
- Shared storage using a clustered file system
- Using StorExpress with Dolphin Express PCIe sockets clustering technology
- And more...

For information on enhanced configurations and capabilities, contact Dolphin or visit us at <http://www.dolphinics.com/>.

Reliability

Although the performance advantages of PCIe SSD have been well demonstrated, advances in reliability have now made this storage technology fit for the enterprise data center. In addition to the inherent advantage over HDDs of having no moving parts to wear out, advanced wear leveling algorithms have largely addressed concerns concerning write endurance of flash devices. Now, lifetimes of over 20 years can be expected using Flash Single-Level Cell (SLC) technology with wear-leveling techniques applied. Additionally, self-healing integrity checks monitor stored data, repair bit errors detected, and remove bad cells or blocks from the system, resulting in a slow and predictable rather than catastrophic wear-out. These techniques are implemented at the PCIe device level, and with the addition of RAID, replication, and other techniques at higher levels result in a solution that can be relied upon in the data center.

Summary

For data intensive applications, StorExpress provides the performance and reliability of a SAN at a fraction of the power, size and cost of traditional disk arrays. HDD-RAID arrays can be avoided or eliminated and servers can be better utilized or reduced.

The reduced hardware, area and cooling requirements of PCI Express solid state storage result in a lowered overall total cost of ownership.

For more information on StorExpress and E-PCIe Solid State technology, visit Dolphin's website at www.dolphinics.com.