electronic design The Authority on Emerging Technologies for Design Solutions

print | close

The Evolution Of Solid-State Storage In Enterprise Servers

Electronic Design
Tom Heil
Tom Heil, LSI
Tue, 2013-04-23 16:20

Solid-state drives (SSDs) and PCI Express (PCIe) flash memory adapters are growing in popularity in enterprise, service provider, and cloud datacenters due to their ability to cost-effectively improve application-level performance. A PCIe flash adapter is a solid-state storage device that plugs directly into a PCIe slot of an individual server, placing fast, persistent storage near server processors to accelerate application-level performance.

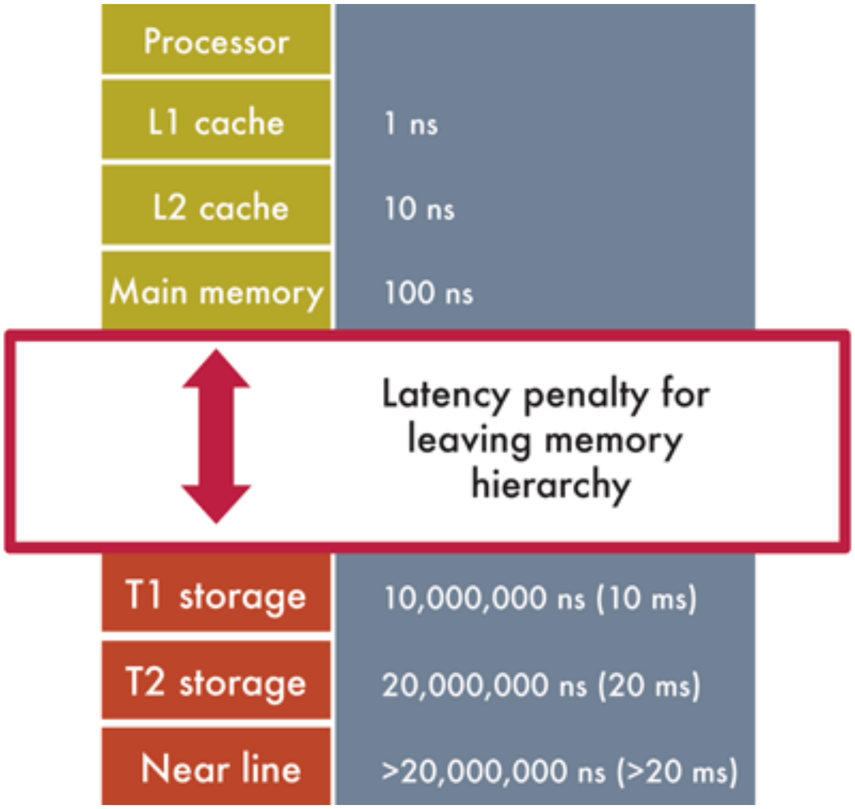
By placing storage closer to the server's CPU, PCIe flash adapters dramatically reduce latency in storage transactions compared to traditional hard-disk drive (HDD) storage. However, the configuration lacks standardization and critical storage device attributes like external serviceability with hot-pluggability.

To overcome these limitations, various organizations are developing PCIe storage standards that extend PCIe onto the server storage mid-plane to provide external serviceability. These new PCIe storage standards take full advantage of flash memory's low latency and provide an evolutionary path for its use in enterprise servers.

Table Of Contents

The Need For Speed

Many applications benefit considerably from the use of solid-state storage owing to the enormous latency gap that exists between the server's main memory and its direct-attached HDDs. Flash storage enables database applications, for example, to experience improvements of four to 10 times because access to main memory takes about 100 ns while input/output (I/O) to traditional rotating storage is on the order of 10 ms or more *(Fig. 1)*.



This access latency difference, approximately five orders of magnitude, has a profound adverse impact on application-level performance and response times. Latency to external storage area networks (SANs) and network-attached storage (NAS) is even higher owing to the intervening network infrastructure (e.g., Fibre Channel or Ethernet).

Related Articles

• NVM Express Changing The SSD Landscape
• Getting The Most Out Of SSD Arrays
• Reassess The Reliability Of Enterprise SSDs

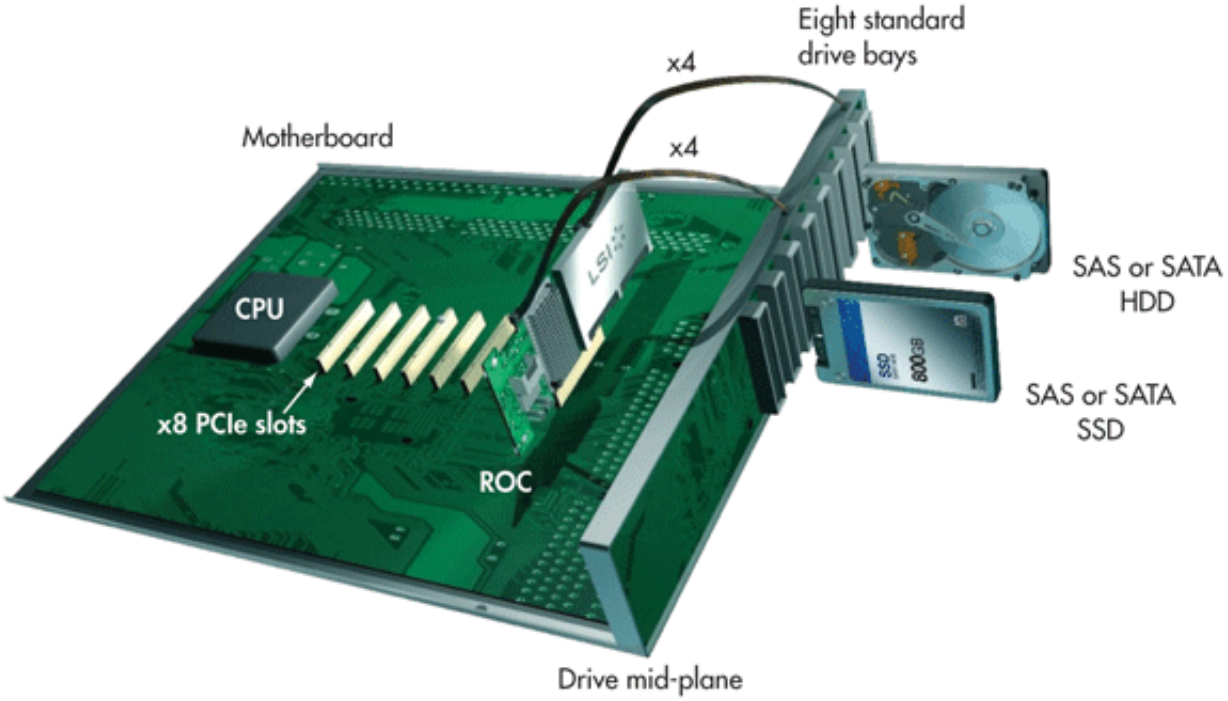
Flash memory provides a new high-performance storage tier that fills the gap between a server's dynamic random access memory (DRAM) and Tier 1 storage consisting of the fastest-spinning HDDs. This new "Tier 0" of solid-state storage, with latencies from 50 µs to several hundred microseconds, delivers dramatic gains in application-level performance while continuing to leverage rotating media's cost-per-gigabyte advantage in all lower tiers.

Because the need for speed is so pressing in many of today's applications, IT managers could not wait for new flash-optimized storage standards to be finalized and become commercially available. That's why SSDs supporting the existing SAS and SATA standards as well as proprietary PCIe-based flash adapters are already being deployed in datacenters. However, these existing solid-state storage solutions utilize very different configurations.

SAS And SATA SSDs

The norm today for direct-attached storage (DAS) is a rack-mount server with an externally accessible chassis having multiple 9-W storage bays capable of accepting a mix of SAS and SATA drives operating at up to 6 Gbits/s. The storage mid-plane typically interfaces with the server motherboard via a PCIe-based host redundant array of independent disks (RAID) adapter that has an embedded RAID-on-chip (ROC) controller

(Fig. 2).



While originally designed for HDDs, this configuration is ideal for SSDs that utilize 2.5-in. and 3.5-in. HDD disk form factors. Support for SAS and SATA HDDs and SSDs in various RAID configurations provides a number of benefits in DAS configurations, such as the ability to mix high-performance SAS drives with low-cost SATA drives in tiers of storage directly on the server. The fastest Tier 0 can utilize SAS SSDs, while the slowest tier utilizes SATA HDDs (or external SAN or NAS). In some configurations, firmware on the RAID adapter can transparently cache application data onto SSDs.

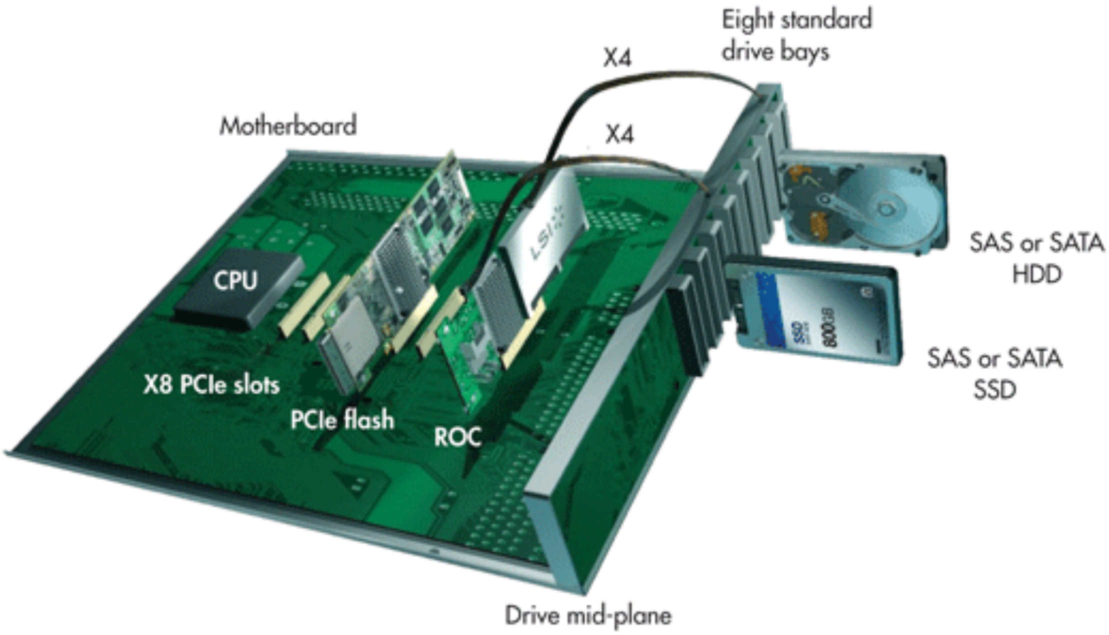
Being externally accessible and hot-pluggable, the configuration of disks can be changed as needed to improve performance by adding more SSDs, or to expand capacity in any tier, as well as to replace defective drives to restore full RAID-level data protection. Because the arrangement is fully standardized, any bay can support any SAS or SATA drive. Device connectivity is easily scaled via an in-server SAS expander or via SAS connections to external drive enclosures, commonly called JBODs for "just a bunch of disks."

The main advantage of deploying flash in HDD form factors using established SAS and SATA protocols is that it significantly accelerates application performance while leveraging mature standards and the existing infrastructure (both hardware and software). So, this configuration will remain popular well into the future in all but the most demanding latency-sensitive applications. Enhancements also continue to be made, including RAID adapters getting faster with PCIe version 3.0, and 12-Gbit/s SAS SSDs that are poised for broad deployment beginning in 2013.

Even with continual advances and enhancements, though, SAS and SATA cannot capitalize fully on flash memory's performance potential. The most obvious constraints are the limited power (9 W) and channel width (one or two lanes) available in a storage bay that was initially designed to accommodate rotating magnetic media, not flash. These constraints limit the performance possible with the amount of flash that can be deployed in a typical HDD form factor, and they are the driving force behind the emergence of PCIe flash adapters.

PCIe Flash Adapters

Instead of plugging into a storage bay, a flash adapter plugs directly into a PCIe bus slot on the server's motherboard, giving it direct access to the CPU and main memory *(Fig. 3)*. The result is a latency as low as 50 µs for (buffered) I/O operations to solid-state storage. Because there are no standards yet for PCIe storage devices, flash adapter vendors must supply a device driver to interface with the host's file system. In some cases, vendor-specific drivers are bundled with popular server operating systems.



Unlike storage bays that provide one or two lanes, server PCIe slots are typically four or eight lanes wide. An eight-lane (x8) PCIe (version 3.0) slot, for example, can provide a throughput of 8 Gbytes/s (eight lanes at 1 Gbyte/s each). By contrast, a SAS storage bay can scale to 3 Gbytes/s (two lanes at 12 Gbits/s or 1.5 Gbytes/s each). The higher bandwidth increases I/O operations per second (IOPs), which reduces the transaction latency experienced by some applications.

Another significant advantage of a PCIe slot is the higher power available, which enables larger flash arrays, as well as more parallel read/write operations to the array(s). The PCIe bus supports up to 25 W per slot, and if even more is needed, a separate connection can be made to the server's power supply, similar to the way high-end PCIe graphics cards are configured in workstations. For half-height, half-length (HHHL) cards today, 25 W is normally sufficient. Ultra-high-capacity full-height cards often require additional power.

A PCIe flash adapter can be utilized either as flash cache or as a primary storage solid-state drive. The more common configuration today is flash cache to accelerate I/O to DAS, SAN, or NAS rotating media. Adapters used as an SSD are often available with advanced capabilities, such as host-based RAID for data protection. But the PCIe bus isn't an ideal platform for primary storage due to its lack of external serviceability and hot-pluggability.

Flash Cache Acceleration Cards

Caching content to memory in a server is a proven technique for reducing latency and, thereby, improving application-level performance. But because the amount of memory possible in a server (measured in gigabytes) is only a small fraction of the capacity of even a single disk drive (measured in terabytes), achieving performance gains from this traditional form of caching is becoming difficult.

Flash memory breaks through the cache size limitation imposed by DRAM to again make caching a highly effective and cost-effective means for accelerating application-level performance. Flash memory is also non-volatile, giving it another important advantage over DRAM caches. As a result, PCIe-based flash cache adapters such as the LSI Nytro XD solution have already become popular for enhancing performance.

Solid-state memory typically delivers the highest performance gains when the flash cache is placed directly in the server on the PCIe bus. Embedded or host-based intelligent caching software is used to place “hot data” (the most frequently accessed data) in the low-latency, high-performance flash storage. Even though flash memory has a higher latency than DRAM, PCIe flash cache cards deliver superior performance for two reasons.

The first is the significantly higher capacity of flash memory, which dramatically increases the “hit rate” of the cache. Indeed, with some flash cards now supporting multiple terabytes of solid-state storage, there is often sufficient capacity to store entire databases or other datasets as “hot data.” The second reason involves the location of the flash cache: directly in the server on the PCIe bus. With no external connections and no intervening network to a SAN or NAS (that is also subject to frequent congestion and deep queues), the “hot data” is accessible in a flash (pun intended) in a deterministic manner under all circumstances.

Although the use of PCIe flash adapters can dramatically improve application performance, PCIe was not designed to accommodate storage devices directly. PCIe adapters are not externally serviceable, are not hot-pluggable, and are difficult to manage as part of an enterprise storage infrastructure. The proprietary nature of PCIe flash adapters also is an impediment to a robust, interoperable multi-party device ecosystem. Overcoming these limitations requires a new industry-standard PCIe storage solution.

Express Bay

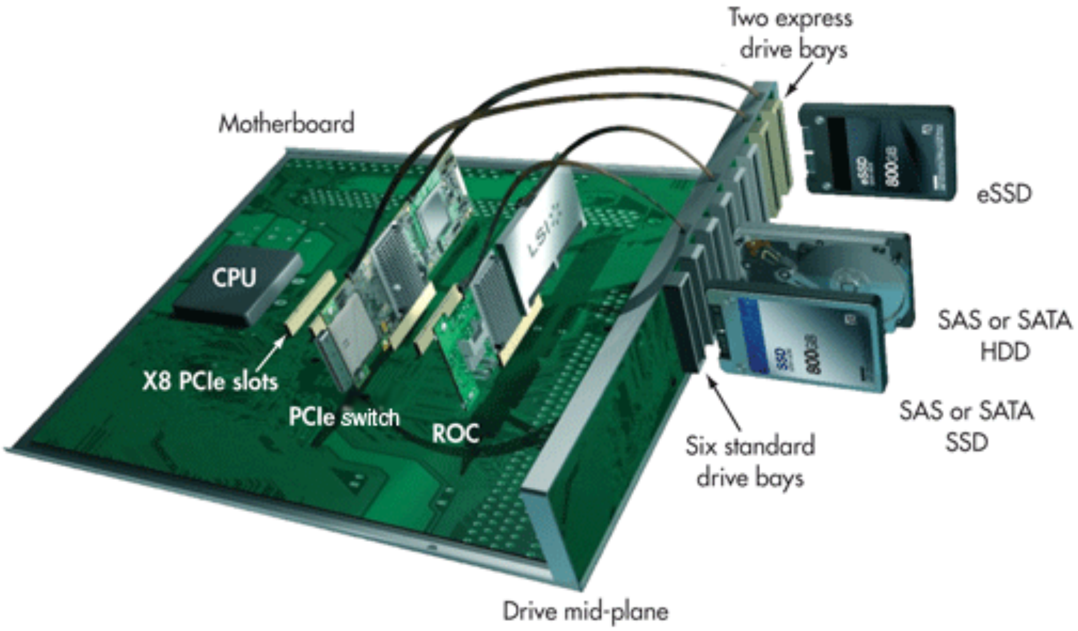
Support for the PCIe interface on an externally accessible storage mid-plane is emerging based on the Express Bay standard with the SFF-8639 connector. Express Bay provides four dedicated PCIe lanes and up to 25 W to accommodate ultra-high-performance, high-capacity Enterprise PCIe SSDs (eSSD) in a 2.5-in. or 3.5-in. disk drive form factor.

As a superset of today’s standard disk drive bay, Express Bay maintains backward compatibility with existing SAS and SATA devices. The SSD Form Factor Working Group is creating the Express Bay standard, Enterprise SSD Form Factor 1.0 Specification, in cooperation with the SFF Committee, the SCSI Trade Association, the PCI Special Interest Group, and the Serial ATA International Organization.

Enterprise SSDs for Express Bay will initially use vendor-specific protocols enabled by vendor-supplied host drivers. Enterprise SSDs compliant with the new NVM Express (NVMe) flash-optimized host interface protocol will emerge in 2013. The NVMe Work Group *(www.nvmexpress.org)* is defining NVMe for use in PCIe devices targeting both clients (PCs, ultrabooks, etc.) and servers. By 2014, standard NVMe host drivers should be available in all major operating systems, eliminating the need for vendor-specific drivers (except when a vendor supplies a driver to enable unique capabilities).

Also in 2014, Enterprise PCIe SSDs compliant with the new SCSI Express (SCSIe) host interface protocol are expected to make their debut. SCSIe SSDs will be optimized for enterprise applications and should fit seamlessly under existing enterprise storage applications based on the SCSI architecture and command set. SCSIe is being defined by the SCSI Trade Association and the InterNational Committee for Information Technology Standards (INCITS) Technical Committee T10 for SCSI Storage Interfaces.

Most mid-planes supporting the Express Bays will interface with the server via two separate PCIe-based cards: a PCIe switch to support high-performance Enterprise PCIe SSDs and a RAID adapter to support legacy SAS and SATA devices *(Fig. 4)*. Direct support for PCIe (through the PCIe switch) makes it possible to put flash cache acceleration solutions in the Express Bay.



This configuration is expected to become preferable over the flash adapters now being plugged directly into the server's PCIe bus. Nevertheless, PCIe flash adapters may continue to be used in ultra-high-performance or ultra-high-capacity applications that justify utilizing the wider x8 PCIe bus slots and/or additional power available only within the server.

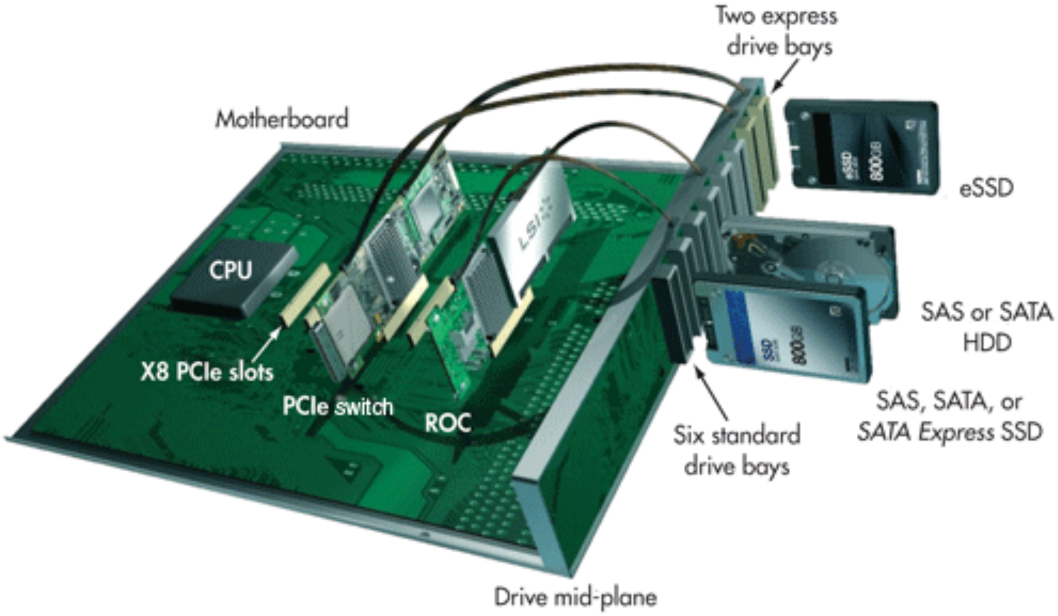
Because it is more expensive to provision an Express Bay than a standard drive bay, server vendors are likely to limit deployment of Express Bays until market demand for Enterprise PCIe SSDs increases. Early server configurations may support perhaps two or four Express Bays, with the remainder being standard bays. Server vendors may also offer some models with a high number of (or nothing but) Express Bays to target ultra-high-performance and ultra-high-capacity applications, especially those that require little or no rotating media storage.

SATA Express

PCIe flash storage also is expected to become common in client devices beginning in 2013 with the advent of the new SATA Express (SATAe) standard. Like SATA before them, SATAe devices are expected be adopted in the enterprise due to the low cost that inevitably results from the economics of high-volume client-focused technologies.

The SATAe series of standards includes a flash-only M.2 form factor (previously called the next-generation form factor or NGFF) for ultrabooks and netbooks and a 2.5-in. disk drive compatible form factor for laptop and desktop PCs. SATAe standards are being developed by the Serial ATA International Organization *(www.sata-io.org)*. Initial SATAe devices will use the current AHCI protocol to leverage industry-standard SATA host drivers, but will quickly move to NVMe once standard NVMe drivers become incorporated into major operating systems.

The SATAe 2.5-in. form factor is expected to play a significant role in enterprise storage. It is designed to plug into either an Express Bay or a standard drive bay. In both cases, the PCIe signals are multiplexed atop the existing SAS/SATA lanes. Either bay then can accommodate a SATAe SSD or a SAS or SATA drive *(Fig. 5)*. Of course, the Express Bay can additionally accommodate x4 Enterprise PCIe SSDs as previously discussed.



The configuration implies future RAID controller support for SATAe drives to supplement existing support for SAS and SATA drives. Note that although SATAe SSDs will outperform SATA SSDs, they will lag 12-Gbit/s SAS SSD performance (two lanes of 12 Gbits/s are faster than two lanes of 8-Gbit/s PCIe 3.0). The SATAe M.2 form factor will also be adopted in the enterprise in situations where a client-class PCIe SSD is warranted, but the flexibility and/or external serviceability of a storage form factor is not required.

Summary

With its ability to bridge the large gap in I/O latency between main memory and hard-disk drives, flash memory has exposed some limitations in existing storage standards. These standards have served the industry well, and

SAS and SATA HDDs and SDDs will continue to be deployed in enterprise and cloud applications well into the foreseeable future. Indeed, the new standards being developed all accommodate today's existing and proven standards, making the integration of solid-state storage seamless and evolutionary, not disruptive or revolutionary.

To take full advantage of flash memory's ultra-low latency, proprietary solutions that leverage the high performance of the PCIe bus have emerged in advance of the new storage standards. But while PCIe delivers the performance needed, it was never intended to be a storage architecture. In effect, the new storage standards extend the PCIe bus onto the server's externally accessible mid-plane, which was designed as a storage architecture.

Yogi Berra famously observed, "It's tough to make predictions, especially about the future." But because the new standards all preserve backwards compatibility, there is no need to predict a "winner" among them. In fact, all are likely to coexist, perhaps in perpetuity, because each is focused on specific and different needs in client and server storage. Fortunately, Express Bay supports both new and legacy standards, as well as proprietary solutions, all concurrently. This freedom of choice down to the level of an individual bay eliminates the need for the industry to choose only one as "the" standard.

Tom Heil *is a senior systems architect and Distinguished Engineer in LSI's Storage Products Division, where he is responsible for technology strategy, product line definition, and business planning. He is a 25-year veteran of the computer and storage industry and holds 17 patents in computer and I/O architecture. He can be reached at tom.heil@lsi.com.*

Source URL: http://electronicdesign.com/memory/evolution-solid-state-storage-enterprise-servers