Native PCIe SSD Controllers
A Next-Generation Enterprise Architecture
For Scalable I/O Performance

Shawn Kung
Director of Product Marketing
Marvell

January 2012
Overview

Next-generation system architectures are taking advantage of solid state drives (SSDs) to eliminate a growing storage I/O bottleneck exacerbated by the rapid growth of cloud computing and unstructured data. While CPU, network bandwidth and storage capacity have continued to keep up with infrastructure growth, storage IOPS (input/output operations per second) from hard disk drives (HDDs) has not. As a result, a significant and growing storage I/O gap has emerged that hinders the scalability and quality of service for cloud computing, database and virtualization environments.

SSDs offer breakthrough performance advantages over traditional HDDs. For this reason, enterprise and cloud computing infrastructures are rapidly adopting SSDs for applications such as large-scale distributed web caches, virtual desktop infrastructure (VDI) and online transaction processing (OLTP) databases. PCI Express (PCIe) offers the lowest-latency, highest throughput interface for applications running on host CPU to take direct advantage of NAND flash. However, current generation PCIe-SAS/SATA and FPGA-based PCIe SSDs have architectural limitations that may result in performance limitations and higher costs.

This paper is targeted at SSD vendors, server and storage OEMs, and cloud computing architects who are looking for next-generation SSD controller technologies that will provide improved performance and scalability.

A next-generation solution is offered in the Marvell® 88NV9145, the world’s first modularly scalable native PCIe-NAND controller. With industry-leading native PCIe SSD performance, the 88NV9145 solves for the limitations faced by existing bridge and FPGA-based PCIe SSDs. Furthermore, the 88NV9145 offers a highly configurable architecture for customers to build flexible configurations to meet the cost, capacity and performance needs of their target customers and market segments.

SSD Market Evolution

SSDs are rapidly becoming adopted in enterprise and cloud computing infrastructures for applications such as large-scale distributed web caches, VDI and OLTP databases.

While there has been plenty of industry buzz around SSDs, the best days of market growth are still ahead. Based on a 2011 market analysis by IMEX® Research, the cumulative SSD market over the next five years is expected to approach $8.6 billion, with the PCIe interface as the largest and fastest-growing segment. Another analysis by Hitachi® Global Storage estimates SSD unit sales to grow from under 2M units in 2011 to nearly 7M units by 2015 – during this same period, the PCIe SSD segment is projected to grow annual shipments from the low 100K units to nearly 2M units.

SSDs offer several performance advantages over traditional spinning HDDs, including higher IOPS, lower latency and higher throughput, as represented by Figures 1, 2 and 3.

---

1 Slide 26, [http://www.snia.org/sites/default/files/AnilVasudeva_Are_SSDs_Ready_Enterprise_Storage_SystemsV4.pdf](http://www.snia.org/sites/default/files/AnilVasudeva_Are_SSDs_Ready_Enterprise_Storage_SystemsV4.pdf)
Native PCIe SSD Controllers – A Next-Generation Enterprise Architecture For Highly Scalable I/O

FIGURE 1: CURRENT SSDS TYPICALLY ACHIEVE 100X-1000X HIGHER IOPS THAN HDDS

FIGURE 2: CURRENT SSDS TYPICALLY ACHIEVE 10X-100X LOWER LATENCY THAN HDDS

FIGURE 3: CURRENT SSDS TYPICALLY ACHIEVE 2X-5X HIGHER THROUGHPUT THAN HDDS
At a system level, the disruptive nature of SSDs is highly beneficial to application workloads where random I/O is the performance bottleneck. These applications include virtualization, OLTP databases, big data, content rendering, collaboration suites (e.g. Exchange server), and large-scale web content caches (e.g. Memcached). SSDs are used typically either as a transient cache or persistent data store to accelerate I/O performance. Many storage systems vendors provide software for automatic caching or data tiering that migrates hot data from HDDs to SSDs (e.g. based on spatial LBA-range or temporal data forensics).

In virtualization environments (e.g. Citrix® XenServer, VMware® ESX, Microsoft® Hyper-V), the increasing density of virtual machines (VMs) has created a storage I/O bottleneck commonly referred to as the I/O blender effect. Both public cloud service providers and private cloud enterprise data center organizations are facing growing pressure to consolidate more VMs on less hardware to lower costs. As the number of VMs grows to meet multi-tenant or VDI scaling requirements, the problems with storage I/O worsen. The reason is while CPU, network bandwidth and storage capacity have continued to keep up with infrastructure growth, storage IOPS from disk drives has not. As a result, a significant and growing disk I/O gap has emerged. Fortunately SSDs are very well positioned to solve for this pain.

Limitations of Existing SSD Solutions

Current SSD suppliers fall into three general segments:

- **NAND Vendors** – Micron®, Toshiba®, SanDisk®, Intel®, Samsung®, etc.
- **HDD Vendors** – Western Digital®, Seagate®, Toshiba®
- **Independent Vendors** – OCZ®, Smart Modular®, Fusion-io®, etc.

Within these segments, several SSD form factors and interfaces are available, including 2.5” SATA or SAS SSDs in a drive form factor and PCIe SSDs in a host adapter form factor. Among these options, PCIe SSDs currently offer the highest I/O and throughput performance in the market. By connecting directly to x86 host processors, PCIe SSDs provide the most direct, low latency path for applications to access data on these devices. In addition, the adapter form factor of PCIe SSDs allot more space for NAND flash and associated controllers, maximizing both SSD performance and available capacity.

Existing PCIe SSDs fall into one of the following categories:

- **Bridge-based** – PCIe-SAS/SATA RAID controller (RoC) + SAS/SATA SSDs
- **Native** – PCIe to NAND flash; either FPGA or ASIC

While both types of PCIe SSD architectures offer performance advantages over a single SATA or SAS SSD, there are several advantages of native PCIe SSDs that make them even more attractive for next-generation server architectures:

<table>
<thead>
<tr>
<th>Bridged PCIe SSD</th>
<th>Native PCIe SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Latency</strong></td>
<td>Low. SAS/SATA protocol overhead</td>
</tr>
<tr>
<td><strong>IOPS</strong></td>
<td>Medium-High. Possible RoC bottleneck</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>Medium. Added RoC power</td>
</tr>
</tbody>
</table>
Current native PCIe SSD solutions are primarily based on an FPGA design. While useful for early adoption, the future of native PCIe SSDs will almost certainly move away from FPGAs to lower-cost and higher-performing ASICs. This is because FPGAs simply cannot scale with the increasing demands of next-generation enterprise server computing technologies. For example, a PCIe 1.0 x8 FPGA-based SSD offers 1.6GB/sec (after overhead). Assuming a 256-bit internal data path, this implies the internal clock needs to run at 100MHz, which is already stressing current-generation FPGAs. At PCIe 2.0 x8, a CPU clock inside the FPGA would need to run at 200MHz and with new-generation servers supporting PCIe 3.0, an FPGA design just won't be powerful enough.

Aside from FPGA performance risks and their high cost, another fundamental challenge with an FPGA-based design is the significant host CPU and memory overhead required to handle wear-leveling algorithms, garbage collection and other SSD management tasks. In an ASIC-based design, the flash translation layer (FTL) and all NAND processing is offloaded onto powerful embedded CPUs. Instead of double-digit host CPU overhead (15-30 percent is typical for FPGA designs), an ASIC-based native PCIe SSD uses an ultra-thin driver for low single digit host CPU and memory overhead to ensure host resources are fully available for application processing.
Native PCIe SSDs – Introducing the Marvell 88NV9145

The Marvell 88NV9145 is the world’s first native PCIe-NAND controller that enables incrementally scalable performance for the enterprise. Marvell has been a trusted supplier to many of the world’s leading SSD suppliers for years and is the first company to offer a native PCIe-NAND controller product to SSD and OEM systems vendors. The 88NV9145 is the latest ground breaking SoC by Marvell that enables customers to scale to over 1 million 4K random IOPS on a single PCIe SSD.

A brief feature summary of the Marvell 88NV9145:

- Native PCIe-NAND controller
- Performance
  - 4K Random Read: 93,000 IOPS
  - 4K Random Write: 70,000 IOPS
  - Latency: <50 us
- PCIe 2.0 x1 end point
- 4 NAND channels, 4 chip select per channel (ONFI2, Toggle)
- Marvell ARM-based CPU
- 128/256-bit AES on-the-fly encryption
- Integrated DMA and ECC engines
- External DDR memory
- Host OS: Linux reference driver
- Power: <1 watt
- Package: 289-pin TFGA

Each 88NV9145 is part of a scalable, modular architecture. A single 88NV9145 can be designed as a native PCIe SSD module (see Figure 6), complete with NAND flash and external DDR. Two or more 88NV9145 modules can be aggregated behind a low-latency, non-blocking PCIe switch to create an incrementally scalable native PCIe SSD. For example, Marvell has developed a reference design (shown in Figures 7 and 8) that aggregates 8x 88NV9145 modules behind a PCIe 2.0 x8 switch. Of course, 88NV9145 controllers also can be designed in non-module configurations such as soldering directly on a printed circuit board.
The 88NV9145 offers industry-leading native PCIe SSD performance. By aggregating 16x 88NV9145 modules and a PCIe switch on a full-height, half-length PCIe card, customers can achieve 1.4 million 4K random read IOPS and 1 million 4K random write IOPS. Figure 9 shows the incremental scalability benefits of the 88NV9145, enabling flexible SSD configurations to match the cost, capacity and performance requirements for target market segments ranging from entry-level to high-end.
Native PCIe SSD Controllers – A Next-Generation Enterprise Architecture For Highly Scalable I/O

4k Random Read

<table>
<thead>
<tr>
<th>Modules</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Module</td>
<td>93 k</td>
</tr>
<tr>
<td>2 Module</td>
<td>186 k</td>
</tr>
<tr>
<td>4 Module</td>
<td>371 k</td>
</tr>
<tr>
<td>8 Module</td>
<td>730 k</td>
</tr>
<tr>
<td>16 Module</td>
<td>1.4 Million</td>
</tr>
</tbody>
</table>

4k Random Write (clean drive)

<table>
<thead>
<tr>
<th>Modules</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Module</td>
<td>70 k</td>
</tr>
<tr>
<td>2 Module</td>
<td>140 k</td>
</tr>
<tr>
<td>4 Module</td>
<td>277 k</td>
</tr>
<tr>
<td>8 Module</td>
<td>530 k</td>
</tr>
<tr>
<td>16 Module</td>
<td>1.04 Million</td>
</tr>
</tbody>
</table>

FIGURE 9: FLEXIBLE, INCREMENTALLY SCALABLE CONFIGURATIONS USING MARVELL 88NV9145

Summary

Next-generation system architectures are taking advantage of SSDs to eliminate a rapidly growing storage I/O bottleneck. SSD manufacturers, server and storage OEMs, cloud computing providers and many others are adopting SSDs to maximize IOPS and throughput, while minimizing application latency.

PCIe offers the most direct, highest performance interface for server applications to take advantage of NAND flash as either a cache or storage tier. However, current generation PCIe-SAS/SATA and FPGA-based PCIe SSDs have architectural limitations that result in performance limitations and higher costs. As depicted by Figure 10, new generation x86 and ARM server processors are projected to shift to native PCIe SSDs to resolve these limitations and maximize application-processing performance.

FIGURE 10: ENTERPRISE SHIFT TO NATIVE PCIe SSD ARCHITECTURES

The Marvell 88NV9145 is a native PCIe-NAND controller that offers industry-leading native PCIe SSD performance to solve for the limitations faced by existing bridge and FPGA-based PCIe SSDs. Furthermore, the 88NV9145 offers a highly configurable architecture for customers to build flexible designs that meet the cost, capacity and performance needs of their target customers and market segments.
As new PCIe SSD interface standards such as SCSI Express and NVMe are defined, the world’s largest SSD and OEM providers will look to industry leaders like Marvell to provide high-performance, reliable SSD ASIC technology to support the evolution of native PCIe SSDs.

To learn more about the Marvell 88NV9145 native PCIe-NAND controller, please visit us at http://www.marvell.com/storage/system-solutions/native-pcie-ssd-controller/.

About the Author:
Shawn Kung
Director of Product Marketing

Shawn Kung is Director of Product Marketing at Marvell. He has P&L responsibility for enterprise and consumer storage products, including innovative systems-level and software solutions in solid-state storage technology. With more than 12 years of Silicon Valley product management experience, Kung has spearheaded the product vision for several award-winning products in data storage and data management. Prior to Marvell, he was senior director of Product Management at Aster Data Systems, where his vision for integrating MapReduce with MPP clustered databases led the company from an early stage nine-engineer startup to a $300M "big data" acquisition by Teradata. Prior to Aster Data Systems, Kung held senior product management positions in core systems at NetApp and enterprise applications at Oracle Corporation.

Kung received his Bachelor of Arts from the Woodrow Wilson School at Princeton University and a Master of Science degree in Management Science and Engineering from Stanford University where he was a Valentine fellow.