Comparison of PCIe SLC Flash cards

Hussein N. El-Harake and Thomas Schoenemeyer
Swiss National Supercomputing Centre (CSCS), Manno, Switzerland
hussein@cscs.ch, schoenemeyer@cscs.ch

Abstract – We evaluated two PCIe attached SSD devices manufactured by Texas Memory Systems in terms of I/O bandwidth and IOPs. One of them is the recently launched "RamSan-70 Gorilla" card designed by Texas Memory Systems. This new card delivered the highest IOPS value ever measured at CSCS.

1. Introduction

In our paper from June 2011 [3] we presented performance numbers for two PCIe SLC Flash cards built by FusionIO and Virident, both with an x8 data lane configuration. Up to 200K IOPS and 1.6GB/s Bandwidth were reported in this earlier study. In August 2011, Texas Memory Systems provided the RamSan-20 card and the new RamSan-70 card shown in figure 1.

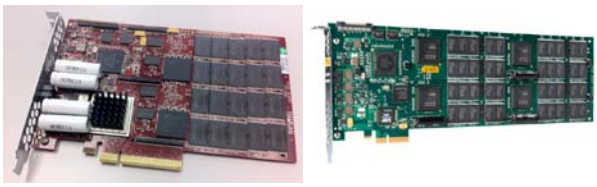

Figure 1: RamSan-20 (right) and RamSan-70 (left)

Both deploy Single Level Cell (SLC) NAND Flash, with a x4 data lane for the RamSan-20 and a x8 data lane configuration for the RamSan-70.

We repeated all benchmarks (IOR, FIO) described in our paper from June 2011 with these flash cards.

2. Methods

The devices were installed in a dual-socket Server manufactured by Supermicro. It deploys Intel Xeon X5670 processors with 6 cores and running at 2.93 GHz with 24 GB Main Memory. The server offers four x16 PCIe Gen2 slots. We used RedHat 5.5, kernel 2.6.18.

XFS was used for benchmarking.

a. Tested devices

We received two PCIe Flash Solid State Disks built by Texas Memory Systems (TMS). One of the key design differentiators of these flash cards is the on-board PowerPC processor that allows to offload tasks from the host managing the flash device (figure 2).

The RamSan-20 delivers 450GB of usable Flash, with fits in PCIe Gen2 slots with 4 lanes thus limited in the maximum bandwidth. It is advertised with a maximum bandwidth of 700 MB/s and 120K IOPS in read and write and consumes only 15 watts of power. The RamSan-20 comes with 640GB of "raw" flash storage, 40% more than that what is usable. This

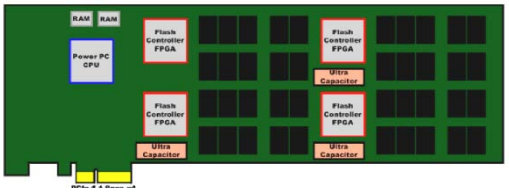

Figure 2: RamSan-20 board

additional 40% storage enhances performance and media endurance by a process called "wear leveling". The card

also uses a small amount of RAM to allow caching and make use of the lower data access time compared to Flash. Large capacitors on the card protect the data as it moves through the RAM buffers. Unplugging the card without warning will not cause any data losses. For all benchmarks, caching was disabled.

The RamSan-70 (figure 3) comes with an x8 data lane configuration and offers up to 900 GB of usable capacity. We tested the card with 450 GB of usable capacity. It also deploys Xilinx FPGAs and a PowerPC processor that moves the managing workload from the host computer. The smaller form factor allows a much easier integration in many servers. It has got two layers of flash memory protection, a proprietary ECC implementation within each 512-byte data set and Variable Stripe RAID (VSR) which is a patented technology from TMS.

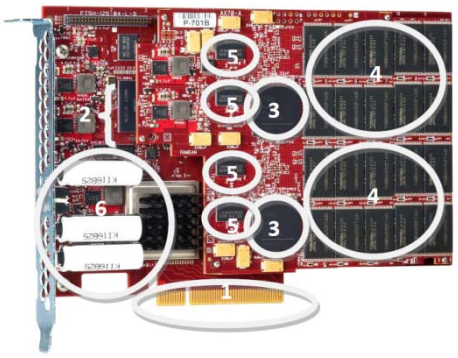

Figure 3: RamSan-70 card. 1=PCIe 2.0 x8, 2=PowerPC CPU, 3= FPGA, 4=450 usable Flash (650 raw), 5=RAM, 6=Super capacitors

Each flash memory chip contains 8 dies, each with 2 planes and 4096 blocks. If a block cannot be erased or written to, or if a read fails with uncorrectable ECC and within a plane, these planes can no longer be used to store data [1] errors, the PowerPC CPU in the Series-7 Flash Controller marks the block as bad attempts to recover its data, if possible. If more than 256 blocks have failures.

The RAID arrays stripe data across all chips, and the RamSan-70 uses 40 chips and 4 flash controllers in its base configuration. So each controller connects to ten flash chips and is initially set up to use a 9+1 rolling XOR RAID across all ten of its Flash chips to rebuild lost data in the event of uncorrectable data errors.

From the whitepaper [2] the following performance characteristics for the 900 GB device can be derived:

- Up to 600,000 IOPs at 4K block for reads
- Up to 160,000 IOPs for writes
- Bandwidth up to 2GB/s
- 30 microsec. access time writes
- 170 microsec. access time reads

From the 450GB card we can expect half the performance, since the number of SSDs is halfed.

RAM write caching is also available for the RamSan-70, the maximum size is limited to 2GB and can be configured by the system administrator in fixed steps between 0 and 2GB. Write caching was disabled for all tests.

b. Experiments

To evaluate the characteristics of the two devices, we repeated all benchmarks FIO and IOR described in [3] using a XFS file system.

3. Results

a. Bandwidth Measurements

We first discuss the random bandwidth measurements received from FIO and IOR for both devices, the results are shown in figures 4, 5, 6 and 7.

The RamSan-20 showed very little difference for peak read and write bandwidth at around 600MB/s which is

expected for a PCI-4x card and little less what is advertised in the datasheet (700). Nearly identical write and read performance is not typical for this sort of devices, so we conclude that the flash cells are only limited by the x4 data connection. Unlike the FusionIO and Viridient TachIOn devices, the bandwidth is almost independent of block size.

This behavior is similar for the RamSan-70, especially for the FIO benchmark. The peak read bandwidth of 1.1 GB/s is provided through all block sizes (figure 8) which is very different from what we reported for the FusionIO and the Viridient devices. In addition, the peak bandwidth is similar to that what is promised by TMS.

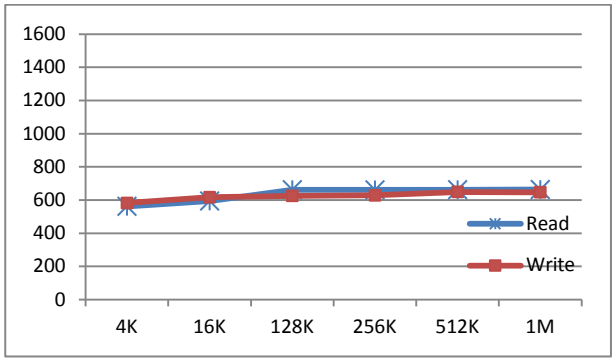

Figure 4: IOR Random Throughput on RamSan-20

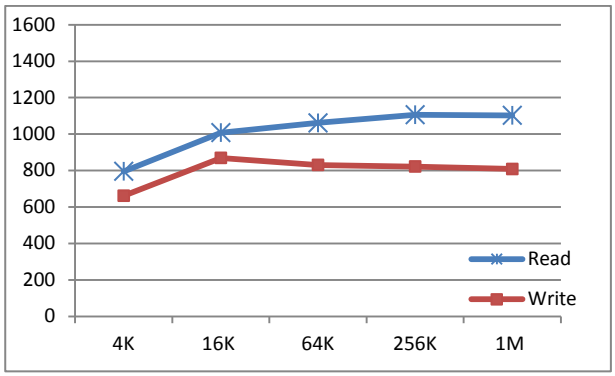

Figure 5: IOR Random Throughput on RamSan-70

Now, with the x8 PCIe connection, the RamSan-70 shows the expected performance difference in read and write with 26% less write bandwidth (as shown in figure 7 at 1M block size). These differences have also been measured for FusionIO and Viridient, but only for block sizes of larger than 64K.

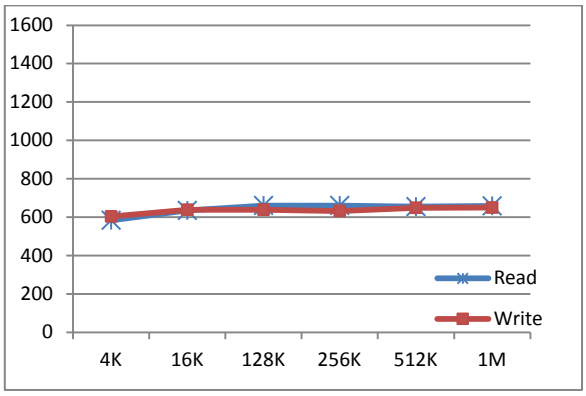

Figure 6: FIO Random Throughput on RamSan-20

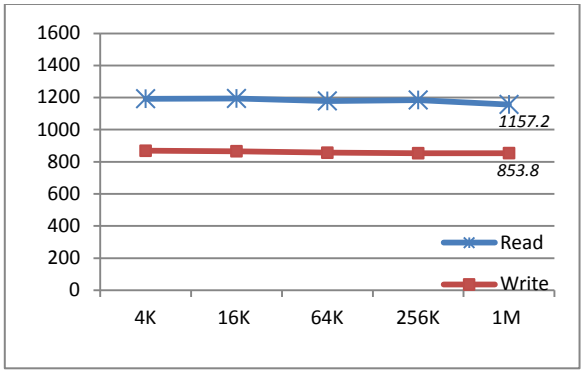

Figure 7: FIO Random Throughput on RamSan-70

b. IOPS measurements

Figures 8, 9, 10 and 11 summarize the results of the IOPS measurements for both flash devices.

The RamSan-20 should according to [4] deliver 120K IOPs for small block sizes. Our results exceeded this value for both benchmarks, IOR and FIO. We measured at 4K block size a peak of 154K IOPs (figure 10).

The RamSan-70 provided by far the best IOPS result we have ever measured at CSCS with a peak of more than 300,000 IOPS at 4K for FIO (figure 11) which is consistent to the expected value.

Again we see for the RamSan-20 no difference between read and write, whereas this difference amounts to up to 28% for the RamSan-70.

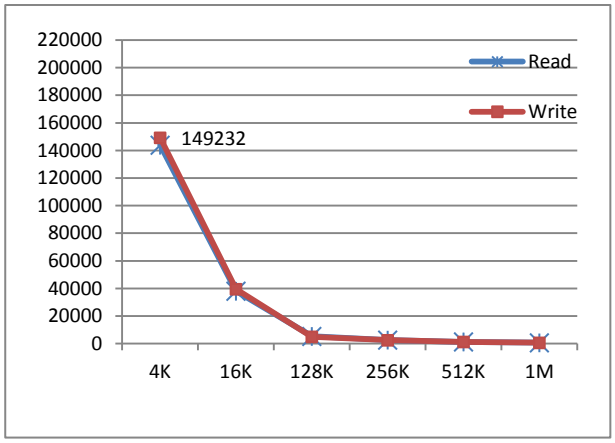

Figure 8: IOR Random IOPS on RamSan-20

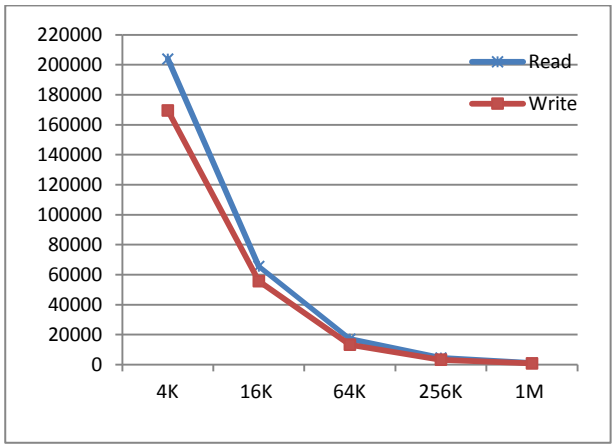

Figure 9: IOR Random IOPS on RamSan-70

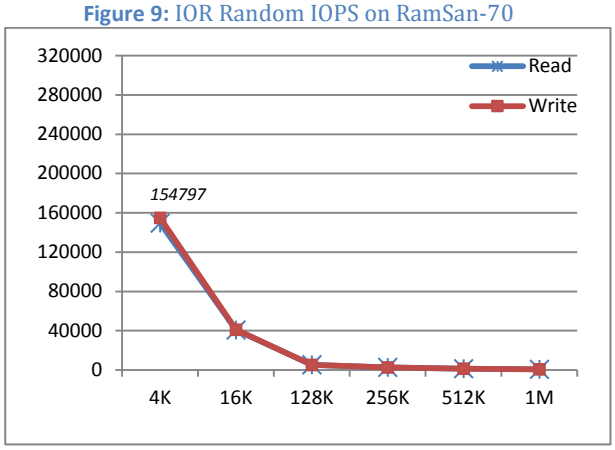

Figure 10: FIO Random IOPS on RamSan-20

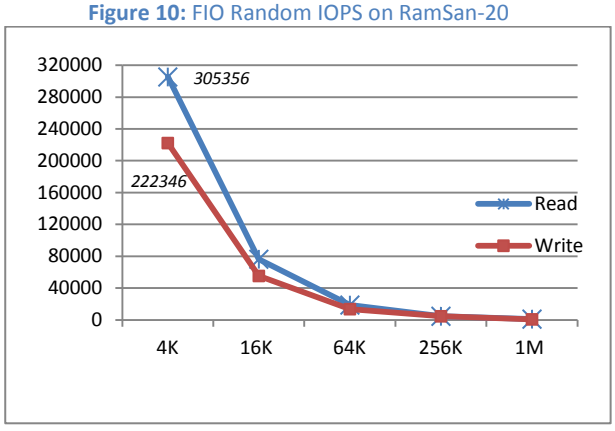

Figure 11: FIO Random IOPS on RamSan-70

4. Conclusion

We evaluated two enterprise class SSD devices based on SLC NAND using a PCIe connection. As test method we used FIO and IOR, two well-known open source in order to perform bandwidth and IOPS measurements with both devices.
The IOPS performance – compared to SATA drives or other PCIe devices – is outstanding.

The peak bandwidth is lower than for the FusionIO card, where we observed up to 1.5 GB/s for large block sizes. However we expect the RamSan-70 device to deliver a rate of 2GB/s because of the double number of SSDs in the device.

Both cards deliver a very stable IO bandwidth over a wide range of block sizes which is a clear differentiator to the cards tested in the previous study.

We will extend this study to other file systems such as Lustre and GPFS.

5. Literature

[1] Robbie Stevens, An In-Depth Look at Variable Stripe RAID (VSR). Texas Memory Systems Whitepaper, June 2011.

[2] Robbie Stevens, An In-Depth Look at the RamSan-70 PCIe Flash Solid State Disk, Texas Memory Systems Whitepaper, January 2011.

[3] Hussein N. Harake and Thomas Schoenemeyer: Detailed Analysis of Solid State Disks, Technical Paper, CSCS, July 2011.

[4] Levi Norman, RamSan-20 PCI Flash SSD, White Paper, Texas Memory Systems, May 2009.

[5] Press Release. TMS launches RamSan-70 Gorilla. http://www.theregister.co.uk/2011/05/17/tms_ramsan_70/