Science & Technology Facilities Council

Technical Report
DL-TR-2010-003

Working Notes of Distributed Computing Group: Comparison of Solid State Drives

DJ Cable

June 2010

Chadwick Library
Science and Technology Facilities Council
Daresbury Laboratory
Daresbury Science and Innovation Campus
Warrington
WA4 4AD

Tel: +44(0)1925 603397
Fax: +44(0)1925 603779
email: library@dl.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk/

ISSN 1362-0207

Working Notes of Distributed Computing Group

Comparison of Solid State Drives

Introduction

Flash-based solid-state disk drives (SSDs) are claimed to have significant performance benefits over regular hard drives. The Distributed Computing Group have been evaluating two different SSDs with a view to establishing how much of a benefit over regular SATA drives SSDs can deliver. Both drives use single-layer cell (SLC) technology, as opposed to multi-layer cell (MLC), which is slower and less reliable[1]. The drives evaluated were the MemoRight MR25.2 and the Intel X25-E, in 128GB and 64GB capacities respectively. The MemoRight drives were not selected on the basis of their potential performance, but rather on the basis of their availability from the supplier of the storage array. A 128GB version of the Intel drive is not yet available. This evaluation did not attempt to assess the reliability or longevity of the SSDs under test.

Test hardware

SuperMicro 24-bay storage server (SC846TQ-R900B)
SuperMicro X7DWA-N motherboard
2 2.5Ghz Intel Xeon 5420 quad-core CPUs
32GB (16GB) RAM
24 port 3Ware 9650SE-24M8 RAID controller
18 750GB Hitachi UltraStar A7K1000 SATA drives
6 SSDs (5 MemoRight MR25.2-S128G 128GB, 1 Intel X25-E 64GB)
2 port 3Ware 9650SE-2LP RAID controller + 2 Western Digital 2502ABYS 250GB SATA drives (RAID 1) (used for OS)
Mellanox MHQH29-XTC ConnectX Infiniband card (not used in this test)

Test environment

RedHat Enterprise Linux 5.3 (kernel 2.6.18-128.1.1.el5)
3Ware 3DM2 RAID controller software
Iozone 3.21 benchmark[2]

Test method

In each test a single disk of each type (MemoRight SSD, Intel SSD and Hitachi SATA) was partitioned with one primary partition. A shell script was used to format this partition with an ext2 filesystem, tune some filesystem parameters, mount the partition, run the Iozone benchmark (up to a file size of twice the RAM in the system), and run dd to test sequential read/writes. The tuning parameters were chosen by experimentation to deliver the best possible performance from a single drive on the 3Ware controller. Specifically, the read-ahead cache size was set to 16384 (default 256) bytes and the filesystem i/o elevator was set to "anticipatory" (default "cfq").

Throughout the evaluation the nfs server on the system was disabled and only the user running the tests was allowed to login.

To eliminate caching effects, it is desirable to tell Iozone to create a file twice the size of the system RAM. Since the Intel drive has a nominal capacity of 64GB, and after formatting only 56GB were available, it was not possible to run Iozone with a maximum file size of 64GB (twice system RAM). Thus 16GB of RAM were removed from the system and a maximum file size of 32GB used instead. The system remained in this configuration for the SATA test, and again a 32GB file was used.

The Iozone command used:

`./iozone -n 512k -g 32g -y 16k -q 2m -a -b $FILE` ("`-g 64g`" was used for the MemoRight SSD)

`$FILE` is an Excel spreadsheet containing the results.

[1] http://geekadviser.com/2008/12/mlc-vs-slc-ssd-drive-and-their-uses/
[2] http://www.iozone.org

The dd commands used:

```
/usr/bin/time -p /bin/dd if=/dev/zero of=/ssd/testfile bs=64k count=851968 \
>> $DDFILE 2>>$DDFILE
```
(sequential write)

```
/usr/bin/time -p /bin/dd if=/ssd/testfile of=/dev/null bs=64k count=851968 \
>> $DDFILE 2>>$DDFILE
```
(sequential read)

(count was 1048576 for the MemoRight SSD and Hitachi SATA drive)

`$DDFILE` contains both dd output and timings.

This script was run three times for each SSD and the results averaged. The test was run only once on the SATA drive as we were not measuring absolute performance but just using it for comparison.

Drive specifications

All drives were connected via a SATA II interface to the same controller in the same chassis, so it is the drive technology itself that creates any performance difference. Table 1 shows the manufacturers' claims for drive performance.

Drive	Sequential reads (MB/sec)	Sequential writes (MB/sec)	Comments
Hitachi SATA	85-42[1]	85-42	Variable because of drive technology. Higher figure used.
MemoRight SSD	100[2]	100	
Intel SSD	250[3]	170	

Table 1: Manufacturer's specifications for drive performance.

Test results

On sequential writes, both the SATA drive and the Intel SSD performed better in Iozone than with dd. The reverse was true of the MemoRight SSD (fig. 1). The Iozone results for both SSDs showed slightly variable speeds at different record sizes, whereas the SATA drive delivered a consistent write speed (fig. 2).

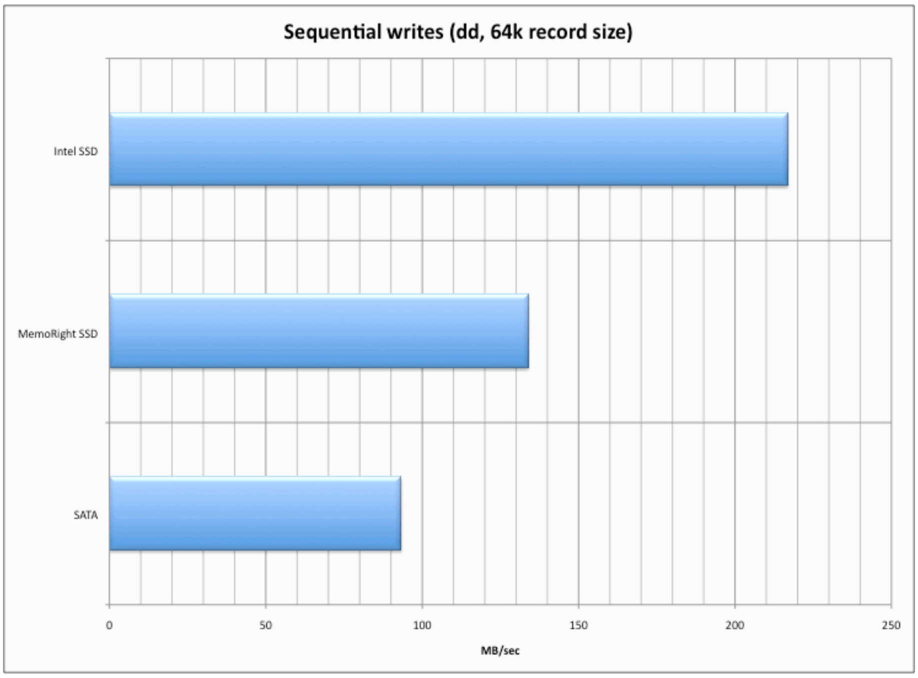

Figure 1: sequential writes using dd

1 Source: http://www.hitachigst.com/tech/techlib.nsf/techdocs/DF2EF568E18716F5862572C20067A757/$file/Ultrastar_A7K1000_final_DS.pdf
2 Source: http://www.memoright.com/uploadfile/MemoRight%20SSD%20GT%20series%20Product%20Specification.pdf
3 Source: http://www.intel.com/design/flash/nand/extreme/index.htm

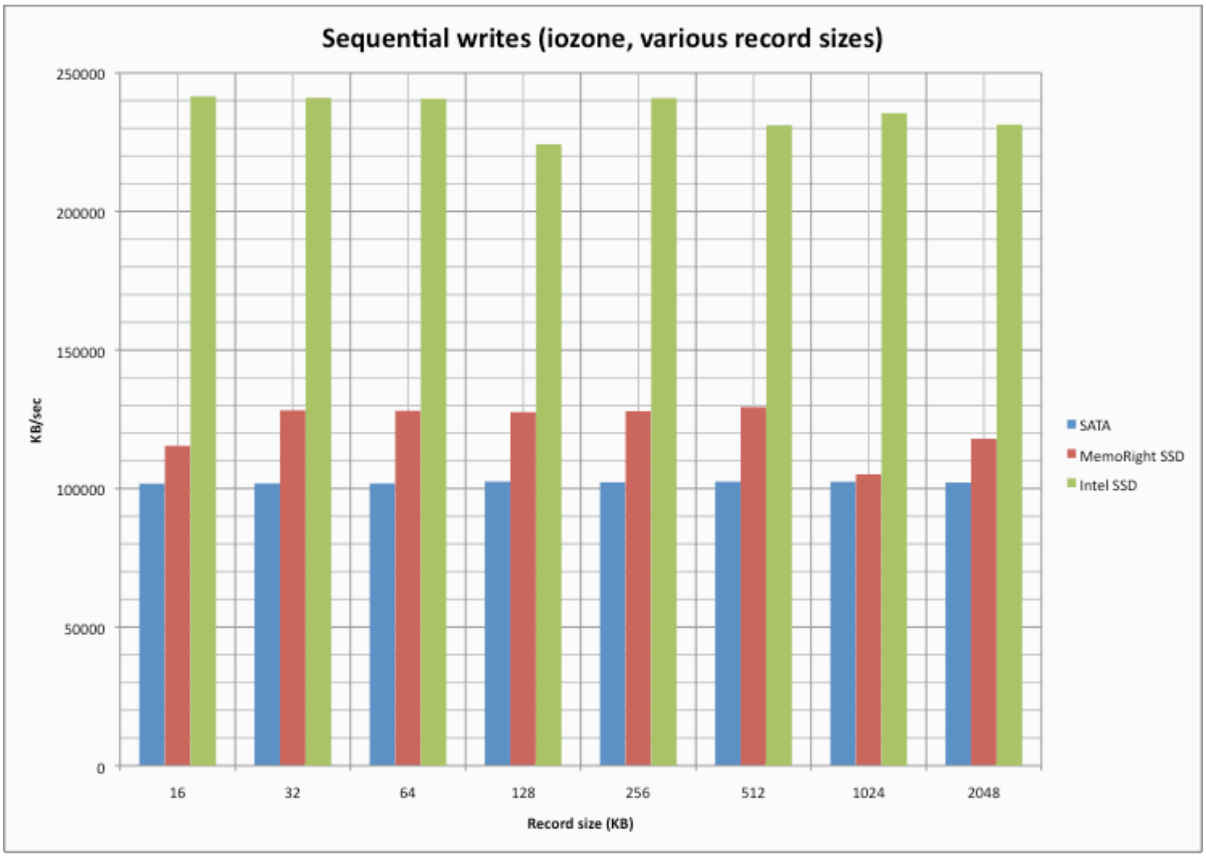

Figure 2: sequential writes using Iozone

The Intel SSD delivered a peak performance of 241MB/sec (Iozone) and the MemoRight 134MB/sec (dd).

On sequential reads, the Intel SSD performed best with Iozone, peaking at 236MB/sec (fig. 4). The MemoRight and SATA peaked at 101MB/sec and 86MB/sec respectively on the dd test (fig. 3). All three drives delivered a consistent performance across record sizes in Iozone (fig. 4).

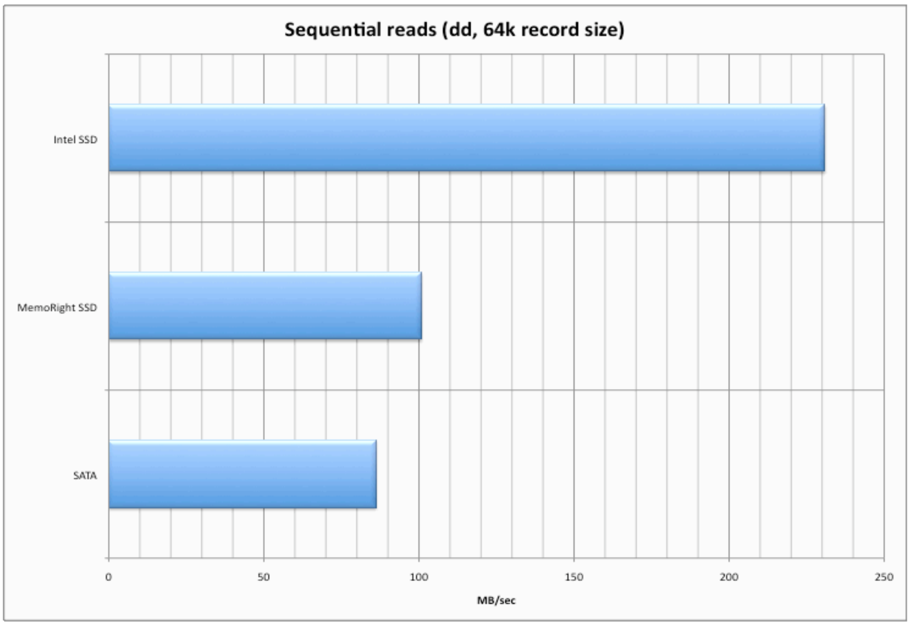

Figure 3: sequential reads using dd

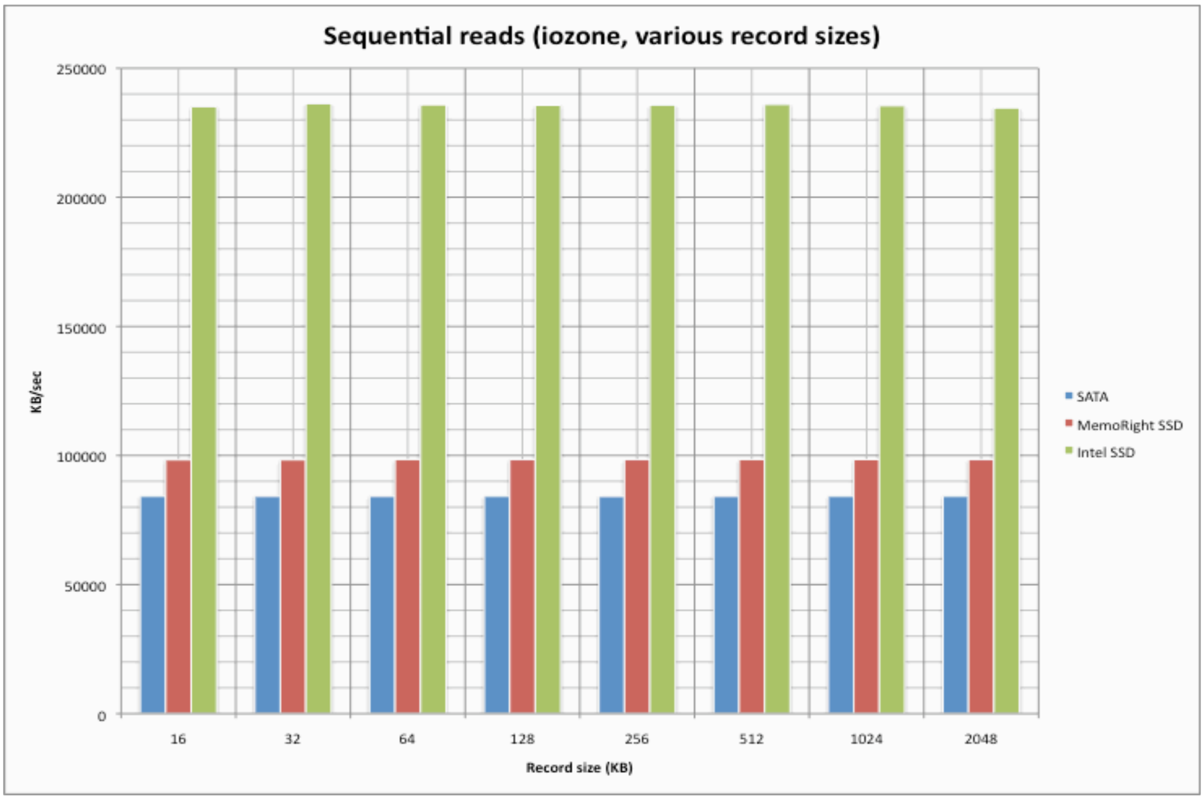

Figure 4: sequential reads using Iozone

With random writes the Intel SSD shows a very definite advantage (fig. 5). Interestingly, the SATA drive actually outperformed the MemoRight SSD on this workload throughout the range of record sizes. Table 2 shows how each drive's random write performance compared with its own sequential write performance at each record size. From this it is clear that the MemoRight SSD doesn't really perform until data is being written in blocks of 1MB, and even then it's still behind the SATA drive in outright speed.

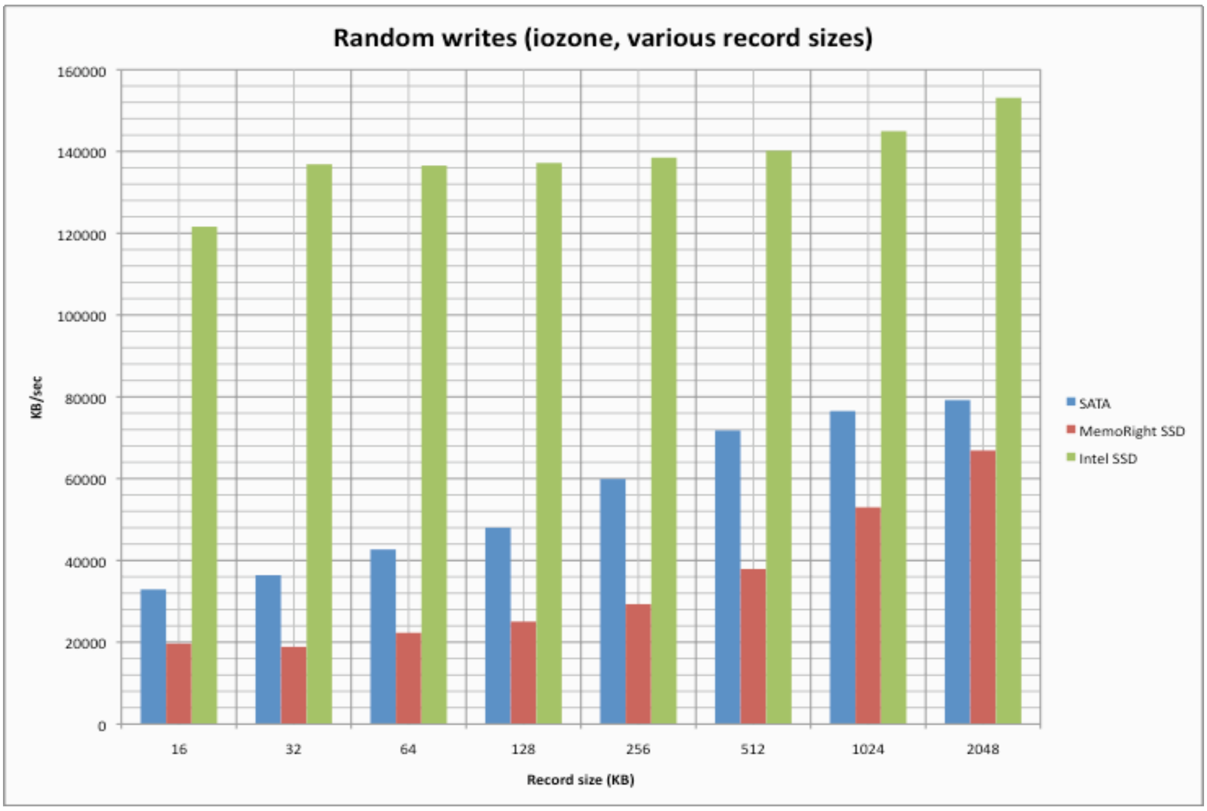

Figure 5: Random write performance using Iozone

	16KB	32KB	64KB	128KB	256KB	512KB	1024KB	2048KB
SATA	32.34	35.73	41.91	46.83	58.55	70.03	74.69	77.52
MemoRight SSD	17.05	14.73	17.42	19.64	22.92	29.29	50.40	56.72
Intel SSD	50.35	56.78	56.72	61.19	57.49	60.64	61.55	66.19

Table 2: Percentage of sequential write performance achieved on random write test

With random reads the MemoRight SSD is "efficient" throughout the range of record sizes (fig. 6) and actually exceeded both its sequential read and its rated performance at block sizes >= 512KB (table 3). The Intel SSD did not work well at small block sizes but with blocks >= 128KB it achieved good performance, more than three times that of the SATA drive (but see the comparison with rated speed below).

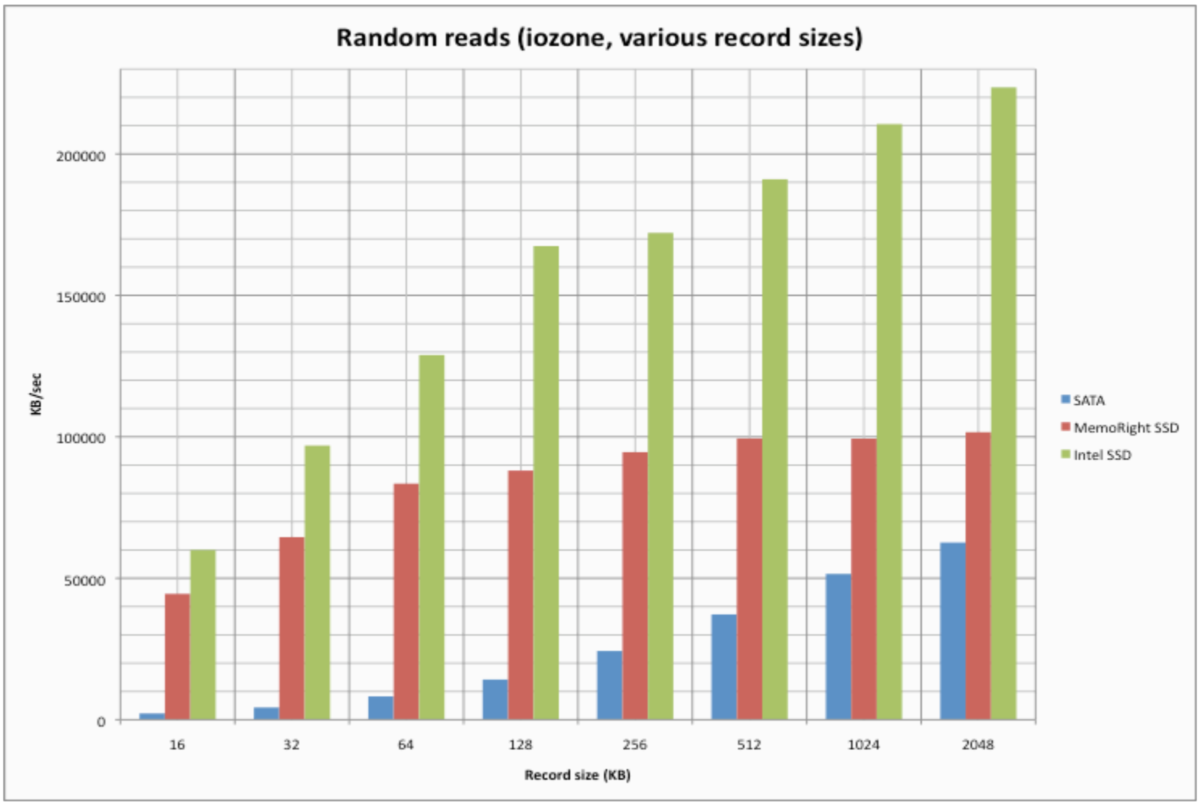

Figure 6: Random read performance using iozone

	16KB	32KB	64KB	128KB	256KB	512KB	1024KB	2048KB
SATA	2.65	5.16	9.78	16.86	28.87	44.22	61.20	74.35
MemoRight SSD	45.27	65.66	84.83	89.50	96.14	101.06	100.94	103.30
Intel SSD	25.53	41.02	54.67	71.06	73.01	80.97	89.43	95.28

Table 3: Percentage of sequential read performance achieved on random read test

Comparison with drive specifications

Table 1 shows the rated performance for each drive. It is interesting to compare this with the actual performance achieved. All three drives exceeded their specified speeds at sequential write tasks. The Intel drive actually delivered a maximum of 241MB/sec at record sizes from 16KB to 256KB, 141% of rated speed.

	MB/s	Sequential write, dd, 64KB blocks		Sequential write, Iozone, 1024KB blocks		Random write, Iozone, 1024KB blocks	
	Spec.	Observed	% of spec.	Observed	% of spec.	Observed	% of spec.
SATA	85	93.4	109.88	102.48	120.56	76.54	90.05
MemoRight SSD	100	134	134.00	105.17	105.17	53.00	53.00
Intel SSD	170	217	127.65	235.52	138.54	144.96	85.27

Table 4: Comparison of observed write performance with drive specifications

The situation is somewhat different when considering read performance (table 5). Both the SATA and MemoRight drives exceeded their rated speeds at smaller block sizes, and even at 1MB they were still very close. The Intel drive never managed its rated 250MB/sec read speed, peaking at around 235MB/sec.

	MB/s	Sequential read, dd, 64KB blocks		Sequential read, Iozone, 1024KB blocks		Random read, Iozone, 1024KB blocks	
	Spec.	Observed	% of spec.	Observed	% of spec.	Observed	% of spec.
SATA	85	86.3	101.53	84.22	99.08	51.55	60.65
MemoRight SSD	100	101	101.00	98.44	98.44	99	99.37
Intel SSD	250	231	92.40	235.46	94.18	211	84.23

Table 5: Comparison of observed read performance with drive specifications

Cost comparison

The cost of SSDs is still very volatile. The Intel drive was purchased in early June, so for a fair comparison the cost of the MemoRight drive (originally purchased in February) was checked with the supplier at the same time. They confirmed that the cost price had dropped substantially (~20%). It actually proved quite difficult to source the Intel drive. The mobile version (X25-M), which is an MLC design and does not perform as well, was readily available. It seems that most X25-Es are going directly to OEM manufacturers, not the retail channel.

Because of price volatility and commercial confidentiality, the prices paid for the SSDs do not appear in this report. However, it is possible to say that the Intel drive was approximately £1 per GB cheaper than the MemoRight drive.

A cost comparison between the SSDs and the SATA drive is not really sensible, because SAS and FC drives provide significantly better performance than SATA but are at the same time considerably more expensive. For the reader's interest only, the SATA drive cost us approximately 20p per GB – the SSD costs were measured in multiple pounds per GB.

Conclusion

Using SSDs to improve parallel filesystem performance makes sense today. For instance, SSDs could be used to hold filesystem metadata, leaving SATA or SAS drives to provide capacity for file data. Such an arrangement is likely to improve filesystem performance by a considerable margin. Another usage, and one which manufacturers such as Panasas[1] are starting to turn to, is to intelligently locate files based on their size, such that small files go on fast SSD and larger files are striped across SATA or SAS arrays. A similar effect can be achieved in software solutions such as GPFS[2], using filesystem policies and a range of storage pools.

This evaluation demonstrated that the performance characteristics of drives can vary significantly, so care must be taken when selecting a particular product.

The MemoRight SSD has only a slight performance advantage over the SATA drive on sequential reads/writes, and although it does well with random reads its random write performance is poor. So for HPC workloads it seems the MemoRight SSD is not a good choice, particularly when cost is taken into account. It may well be that this drive is designed primarily for use in netbooks/notebooks, and the manufacturer may have alternative, higher specification drives available.

The Intel drive is a better prospect, offering performance substantially greater than SATA on the workloads tested here. It is however also expensive, and at present limited to 64GB capacity. Intel is expected to launch a 160GB X25-E product very soon, along with a 320GB 'mobile' MLC version.

There are alternative products available from a number of manufacturers, including entire storage arrays specifically designed to use enterprise-class SSDs, and SSDs based on DRAM technology. It is hoped that the Distributed Computing Group may have the opportunity to evaluate some of these other products and/or use cases for SSDs.

Acknowledgements

Thanks to ClusterVision Ltd. for their help with current pricing information.

28/07/2009 version 1.2 (final)

[1] http://www.panasas.com/library.html
[2] http://www-03.ibm.com/systems/clusters/software/gpfs/index.html

D.J. Cable, Distributed Computing Group, Computational Science and Engineering Department, STFC Daresbury Laboratory, Daresbury, Warrington, Cheshire, WA4 4AD, UK