Mtron Flash SSD Single Drive and Raid Array Benchmarks

Tests Run November 25-29, 2007

Doug Dumitru – CTO EasyCo LLC

Executive Summary:

Mtron 3025-25 and 7025-25 Flash Solid State Disks were tested to measure random IO performance for read and write operations at varying block sizes. Single drive, 4-drive RAID-5, and 8-drive RAID-5 configurations were tested. The configurations were tested both with bare drives and with EasyCo's Managed Flash Technology driver in place.

Results for bare drives show that the Mtron drives are "best in class" delivering >20,000 read IOPS, >110 MB/sec read, and >80 MB/sec write performance. Write performance measured about 125 IOPS which is 3 to 10 times better than other Flash SSDs. Overall, the Mtron drives are 2-50 times as fast as 15K hard disks for most ranges of applications.

Results with the EasyCo MFT driver in place show that single drives are capable of about 15,000 4K read/write IOPS. Multiple drive arrays can deliver nearly 50,000 IOPS. This is a performance level that is "orders of magnitude" faster than traditional hard disks and represents the next generation of enterprise storage.

Test Hardware:

A number of low-level benchmarks were run with a collection of Mtron Flash Solid State disks. These tests should provide good predictions of application performance with similar hardware.

Drives: Mtron 7025 series (PRO) 32 GB SATA
Mtron 3025 series (MOBI) 32GB SATA

Server: MSI Motherboard
AMD Athlon 64x2 6000+
2GB Ram
Linux – Fedora 7 – 64-bit

Disk Controller: Highpoint PCI-e 2320

Test Software:

The 'bm-flash.c' program was used for all tests. This is a C program developed by EasyCo that does blocking, aligned IO with block sizes ranging from 512 bytes to 4 megabytes. Read and write operations are tested separately. This program also tests 1, 10, and 40 execution threads to show IO overlap performance. This program is designed to do exclusively "direct IO" so that Linux does not buffer or cache disk blocks.

Test Results for Single Bare Drives:

Mtron PRO series:

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW
512B	14422	7.0M	20751	10.1M	20756	10.1M	122	61.3K	121	60.5K		
4K	10337	40.3M	13223	51.6M	13226	51.6M	123	493.5K	123	495.5K		
8K	7851	61.3M	9417	73.5M	9422	73.6M	124	996.7K	124	994.3K		
16K	5168	80.7M	5839	91.2M	5842	91.2M	120	1932.7K	128	2.0M		
32K	3075	96.0M	3316	103.6M	3319	103.7M	125	3.9M	127	3.9M		
64K	1696	106.0M	1779	111.2M	1782	111.4M	121	7.6M	124	7.7M		
128K	876	109.6M	896	112.1M	899	112.4M	115	14.3M	115	14.4M		
256K	444	111.1M	450	112.6M			100	25.0M	101	25.3M		
512K	225	112.6M	226	113.1M			88	44.4M				
1M	113	113.3M	113	113.8M			70	70.5M				

This is the baseline for a Flash SSD. This drive is actually very fast. The linear read speed is rated at 120 MB/sec and this drive tests at 114 MB/sec. Linear writes are rated at 90 MB/sec and the drive actually tests at 83.5 MB/sec. With longer, sustained, transfers, this drive appears to meet its published specs.

Even more interesting is the drive's read "access times". For a single thread, 14422 IOPS at 512 bytes is .065 ms "access time". In that this drive does not support NCQ and the multi-thread numbers are a lot better, quite a bit of this access time has to be Linux overhead. Using the multi-thread numbers, the true access time for the drive is .040 ms. This is a shockingly good number. It also exceeds Mtron's published random read IOPS of 16,000 by more than 25%.

The 512 byte numbers really don't mean anything for real-world applications. Most filesystems do exclusively 4K IO. Most databases do either 4K or 8K IO. With 4K IO, the drive sustains over 13K IOPS or over 50 MB/sec. At 8K, the drive sustains over 9K IOPS and over 70 MB/sec. As a comparison, a 15K SAS drive will sustain about 225 4K or 8K IOPS making this drive 57 times as fast, drive for drive, as a single 15K SAS drive, when doing random 4K and 8K reads.

This test also shows typical write performance behavior for Flash SSDs. This drive basically does about 125 write IOPS for 64K and smaller writes. While 125 is a comparatively low number, it is actually very good for a Flash drive. SLC flash drives that we have tested in house at EasyCo range from 13-40 write IOPS. We even tested an MLC flash drive at 3.3 IOPS. Also of note is that when coupled with the Managed Flash Technology software driver, this drive does about 15,000 4K write IOPS. MFT benchmarks are included at the end of this paper.

Mtron MOBI Series:

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW
512B	13357	6.5M	18636	9.0M	18637	9.1M	119	59.9K	121	60.5K		
4K	9784	38.2M	12319	48.1M	12325	48.1M	124	496.7K	124	496.0K		
8K	7376	57.6M	8790	68.6M	8795	68.7M	125	1004.7K	119	958.3K		
16K	4806	75.1M	5392	84.2M	5394	84.2M	119	1908.7K	124	1.9M		
32K	2836	88.6M	3040	95.0M	3043	95.1M	119	3.7M	124	3.8M		
64K	1559	97.4M	1624	101.5M	1627	101.7M	117	7.3M	118	7.4M		
128K	802	100.2M	818	102.2M	821	102.6M	111	13.8M	116	14.6M		
256K	406	101.6M	410	102.7M			97	24.3M				
512K	205	102.7M	206	103.2M			79	39.6M				
1M	103	103.3M	103	103.8M			63	63.5M				

What is interesting about these numbers is that they are very close to the PRO series numbers. Overall, the MOBI drive appears to be about 8% slower than the PRO drives.

4 Drive RAID-5 Arrays:

We ran some tests with 4 drives setup RAID-5. These tests used the Linux "software raid" layer. This gives you an idea of what type of performance to expect for a small array:

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW	1 thread IOPS	1 thread BW	10 threads IOPS	10 threads BW	40 threads IOPS	40 threads BW
4K	10004	39.0M	41859	163.5M	50569	197.5M	248	992.0K	254	1016.7K		
8K	7669	59.9M	31254	244.1M	36110	282.1M	254	1.9M	259	2.0M		
16K	5089	79.5M	19539	305.2M	22414	350.2M	249	3.8M	255	3.9M		
32K	3009	94.0M	11211	350.3M	12734	397.9M	250	7.8M	253	7.9M		
64K	1690	105.6M	6032	377.0M	6861	428.8M	245	15.3M	246	15.3M		
128K	1656	207.0M	3359	419.9M	3493	436.6M	182	22.7M	181	22.6M		
256K	1555	388.7M	1776	444.0M	1778	444.7M	121	30.3M	123	30.8M		
512K	806	403.0M	892	446.3M	895	447.7M	106	53.0M	107	53.8M		
1M	416	416.5M	448	448.1M			92	92.3M				

These are quite good numbers as well. The drive scales to 50,000 4K read IOPS which is almost 4x the speed of a single drive. Another interesting characteristic is that the drive does writes 2x the speed of a single drive. With traditional RAID-5 arrays, you would expect random writes to be worse than this. Because flash drives do random reads so quickly, the random write speed of RAID-5 is identical to what you would expect from RAID-10.

8 Drive RAID-5 Arrays:

Our 8 drive tests used a mix of 4 7025-32 and 4 3025-32 drives. In that the 3025 MOBI series drives are about 8% slower, we would expect these numbers to be about 4% slower than a pure 8 drive 7025 array.

RAID-5 w/ Linux software raid driver

We ran two sets of test with 8 drives setup RAID-5. The first uses Linux "software raid" layer. Later we will run the same tests using Highpoint's RAID-5 driver. This gives you an idea of what type of performance to expect for a medium sized array:

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
4K	9763	38.1M	42978	167.8M	54100	211.3M	557	2.1M	562	2.1M	573	2.2M
8K	7429	58.0M	39584	309.2M	51352	401.1M	539	4.2M	571	4.4M	573	4.4M
16K	4908	76.6M	28166	440.0M	39423	615.9M	557	8.7M	565	8.8M	578	9.0M
32K	2919	91.2M	16562	517.5M	22369	699.0M	548	17.1M	559	17.4M	555	17.3M
64K	1611	100.7M	8984	561.5M	11908	744.2M	442	27.6M	532	33.2M	518	32.3M
128K	1556	194.5M	5533	691.6M	6051	756.4M	342	42.7M	367	45.9M		
256K	1489	372.2M	2940	735.0M	3036	759.0M	231	57.7M	228	57.0M		
512K	1256	628.1M	1531	765.6M	1537	768.9M	133	66.9M	135	67.9M		
1M	663	663.0M	775	775.2M	775	775.1M	111	111.1M	113	113.1M		

These test results are not quite as good as we would have hoped. With small random reads, the performance is only a little better multi-threaded than 4 drives. This implies that the controller card really does not like doing >50K IOPS. Then again, this is probably the first time this card has ever had such an array connected to it.

With large transfers and on writes, the 8 drive array does scale well producing nearly twice the linear read throughput and exactly twice the random write throughput.

RAID-5 w/ Highpoint software raid driver

This is the RAID-5 test repeated with the Highpoint driver. You should note that the 2320 Highpoint card is a "fake raid" card. It has no on-board processor. This implies that the performance should be about the same as with Linux software raid. While there are similarities, the differences are actually interesting.

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
4K	9047	35.3M	35505	138.6M	36567	142.8M	553	2.1M	582	2.2M	575	2.2M
8K	7023	54.8M	33665	263.0M	35575	277.9M	580	4.5M	570	4.4M	547	4.2M
16K	4734	73.9M	26757	418.0M	33440	522.5M	545	8.5M	575	8.9M	565	8.8M
32K	2861	89.4M	16372	511.6M	22349	698.4M	539	16.8M	558	17.4M	562	17.5M
64K	1593	99.6M	8962	560.1M	11909	744.3M	451	28.2M	530	33.1M	532	33.2M
128K	1554	194.3M	5287	660.9M	6062	757.7M	328	41.0M	344	43.0M		
256K	1486	371.6M	2897	724.2M	3066	766.6M	208	52.1M	204	51.0M		
512K	859	429.5M	1524	762.1M	1555	777.5M	134	67.2M	136	68.4M		
1M	549	549.5M	774	774.7M	783	783.0M	109	109.0M	117	117.5M		

The thing that stands out most here is that the maximum read IOPS saturates out at about 37,000, which is about 40% slower than the software raid setup. Otherwise, the numbers are nearly identical.

RAID-5 vs RAID-10 Write Performance

In an earlier white paper, I hypothesized that flash SSDs could use RAID-5 arrays without degrading random write performance as is noticed with traditional hard disk drives. With RAID-5, computing parity requires that the entire stripe be read before the parity sectors can be written. With hard disks, this extra read slows down random writes so that a RAID-5 array will tend to write at the same speed as a single drive. With flash SSDs, this read still happens, but because flash drives read almost instantaneously, this read should not have an impact on write performance.

These test show that 4 drives in a RAID-5 array write at about double the speed of a single drive and 8 drives in a RAID-5 array write at about four times the speed of a single drive. This is exactly what you would expect and matches the random write performance of RAID-10. With this behavior in mind, there appears to be no compelling reason to configure flash SSDs as RAID-10. Instead, RAID-5 appears to offer the same performance without the space penalty.

Controller Tests:

We re-ran the single drive tests with the drive connected to the motherboard SATA controller. Our experience with Highpoint boards is that they are very low latency, especially compared with some motherboard chipsets. This should show whether this holds for the 2230 board or not.

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
512B	9427	4.6M	10114	4.9M	10117	4.9M	119	59.8K	123	61.8K		
4K	7483	29.2M	7925	30.9M	7927	30.9M	124	497.5K	124	496.0K		
8K	6080	47.5M	6403	50.0M	6405	50.0M	125	1000.7K	125	1004.0K		
16K	4335	67.7M	4502	70.3M	4506	70.4M	125	1.9M	126	1.9M		
32K	2751	85.9M	2839	88.7M	2843	88.8M	124	3.8M	127	3.9M		
64K	1599	99.9M	1631	101.9M	1634	102.1M	122	7.6M	125	7.8M		
128K	866	108.3M	881	110.1M	884	110.5M	109	13.7M	116	14.5M		
256K	438	109.6M	444	111.0M			97	24.4M				
512K	219	109.8M	223	111.7M			82	41.0M				
1M	110	110.7M	112	112.5M			70	70.3M				

In terms of linear throughput, the on-board SATA controller is nearly identical to the Highpoint PCI-e card. But when looking at small random reads, the difference is remarkable. For 4K multi-threaded reads, this on-board controller is over 35% slower.

This test shows that controllers are very important. On the other hand, don't assume that every motherboard SATA interface is junk. We have tested other boards that do quite well.

Acceleration with the Managed Flash Technology driver:

We then set out to run the same tests with the MFT driver in place. To keep the listings short, we will only show three cases:

- Single drive (on the Highpoint controller)
- Four drives configured RAID-5
- Eight drives configured RAID-5

We chose Linux's software raid driver for the four and eight drive configurations.

Single Mtron PRO series drive with MFT:

	Random Read Tests:						Random Write Tests:					
Block	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
Size	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
4K	10134	39.5M	13294	51.9M	13297	51.9M	14987	58.5M	16180	63.2M	16087	62.8M
8K	7689	60.0M	9452	73.8M	9457	73.8M	7992	62.4M	8717	68.1M	8789	68.6M
16K	5095	79.6M	5847	91.3M	5850	91.4M	4293	67.0M	4496	70.2M	4539	70.9M
32K	3048	95.2M	3318	103.7M	3322	103.8M	2228	69.6M	2269	70.9M	2374	74.1M
64K	1684	105.3M	1780	111.2M	1783	111.4M	1116	69.7M	1170	73.1M	1230	76.8M
128K	871	108.9M	896	112.0M	899	112.4M	563	70.4M	590	73.8M	617	77.2M
256K	442	110.5M	450	112.5M			281	70.2M	306	76.5M		
512K	224	112.1M	226	113.1M			138	69.3M	154	77.2M		
1M	113	113.0M	113	113.7M			70	70.1M				

These results are typical of MFT storage sub-systems. Read performance is basically identical to the bare drive and random write performance uses 80-95% of the drives available linear bandwidth.

With this level of performance, running a 4K block size, 50% read / 50% write database, the Mtron PRO series flash drive would be about 25% faster than a 15K SCSI drive. The same flash drive with MFT is about 5000% faster.

Four Mtron PRO series drives configured for RAID-5 with MFT:

	Random Read Tests:						Random Write Tests:					
Block	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
Size	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
4K	9768	38.1M	39689	155.0M	49952	195.1M	29500	115.2M	29618	115.6M	29698	116.0M
8K	7419	57.9M	28635	223.7M	32830	256.4M	17856	139.5M	16293	127.2M	16099	125.7M
16K	4953	77.3M	18122	283.1M	20309	317.3M	9655	150.8M	8522	133.1M	8315	129.9M
32K	3091	96.6M	10470	327.2M	11657	364.2M	4847	151.4M	4257	133.0M	4336	135.5M
64K	1942	121.3M	5772	360.7M	6336	396.0M	2663	166.4M	2181	136.3M	2174	135.9M
128K	1445	180.6M	3185	398.1M	3348	418.5M	1158	144.7M	1093	136.7M	1153	144.1M
256K	1172	293.0M	1753	438.4M	1766	441.5M	635	158.9M	552	138.0M	640	160.1M
512K	638	319.2M	882	441.3M	889	444.9M	264	132.0M	273	136.5M		
1M	330	330.3M	444	444.5M			129	129.5M	151	151.5M		

Again, these are typical results. These numbers represent a small array capable of sustained 40,000+ read IOPS and 28,000+ write IOPS for 4K blocks. This level of performance would require nearly 200 15K SCSI drives in a RAID-10 array.

Eight Mtron PRO/MOBI drives configured RAID-5 with MFT:

Again, this test used 4 7025-32 drives and 4 3025-32 drives in the array.

	Random Read Tests:						Random Write Tests:					
Block Size	1 thread		10 threads		40 threads		1 thread		10 threads		40 threads	
	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW	IOPS	BW
4K	9528	37.2M	40972	160.0M	50980	199.1M	37696	147.2M	47256	184.5M	46542	181.8M
8K	7286	56.9M	33799	264.0M	40663	317.6M	24011	187.5M	27360	213.7M	27031	211.1M
16K	4821	75.3M	24169	377.6M	30628	478.5M	12680	198.1M	13899	217.1M	13517	211.2M
32K	3038	94.9M	15024	469.5M	19992	624.7M	6621	206.9M	8190	255.9M	7102	221.9M
64K	1887	117.9M	8745	546.6M	11016	688.5M	3496	218.5M	3961	247.5M	3717	232.3M
128K	1406	175.8M	5103	637.9M	5885	735.6M	1720	215.0M	1948	243.5M	1871	233.9M
256K	1188	297.0M	2846	711.6M	3096	774.1M	820	205.0M	951	237.9M	905	226.3M
512K	832	416.2M	1558	779.1M	1575	787.5M	430	215.2M	462	231.1M		
1M	458	458.0M	787	787.5M	792	792.1M	207	207.8M	218	218.0M		

As expected from the bare drive tests, the eight drive array does not scale linearly. Still, it yields about 50,000 4K read IOPS and 45,000 4K write IOPS. Considering that this array consists of only eight 2.5" drives and draws less than 30 watts of power, it is an amazing storage solution for mid-size servers.

Future Tests:

We are waiting for an LSI RAID-5 SAS controller to arrive. When it gets here, we plan on re-running the 4 and 8 drive tests. We will update this paper with new results as they come in.

Conclusion:

The Mtron drives perform very well. With 40 uS access times, the random and linear read performance is excellent. In terms of write performance, the linear performance is again excellent. The random write performance, when compared with other Flash SSDs is also "best in class", but still painfully slow compared with the other performance characteristics of the drive.

Bare Mtron Flash SSD vs. 15K HDD:

How fast the Mtron drives are compared with 15K HDD is dependent on the percentage of random reads to writes.

100% reads	Web Server / Static Databases	50x faster
90% reads / 10% writes	Light usage Databases / Desktops	5x faster
50% reads / 50% writes	Active Databases	Even

This is for random operations with small block sizes. At large block sizes, the Mtron drive is 10-40% faster than a 15K SSD, mostly depending on what part of the HDD you are accessing.

Drive w/ MFT vs 15K HDD:

When coupled with the MFT management layer, the Mtron drives are basically 50x faster than a 15K HDD regardless of the read/write mix.

About EasyCo

EasyCo is a software developer and integrator specializing in high performance storage subsystems. EasyCo also distributes Mtron Flash SSDs in North America.

http://easyco.com
http://mtron.easyco.com
(610) 237-2000 or (800) 470-2756
sales@easyco.com

About Managed Flash Technology

MFT is EasyCo's patent pending technology that optimizes storage subsystems for maximum random write performance. More information is available at EasyCo MFT website: http://managedflash.com