[SATA-IP Application Note 1]

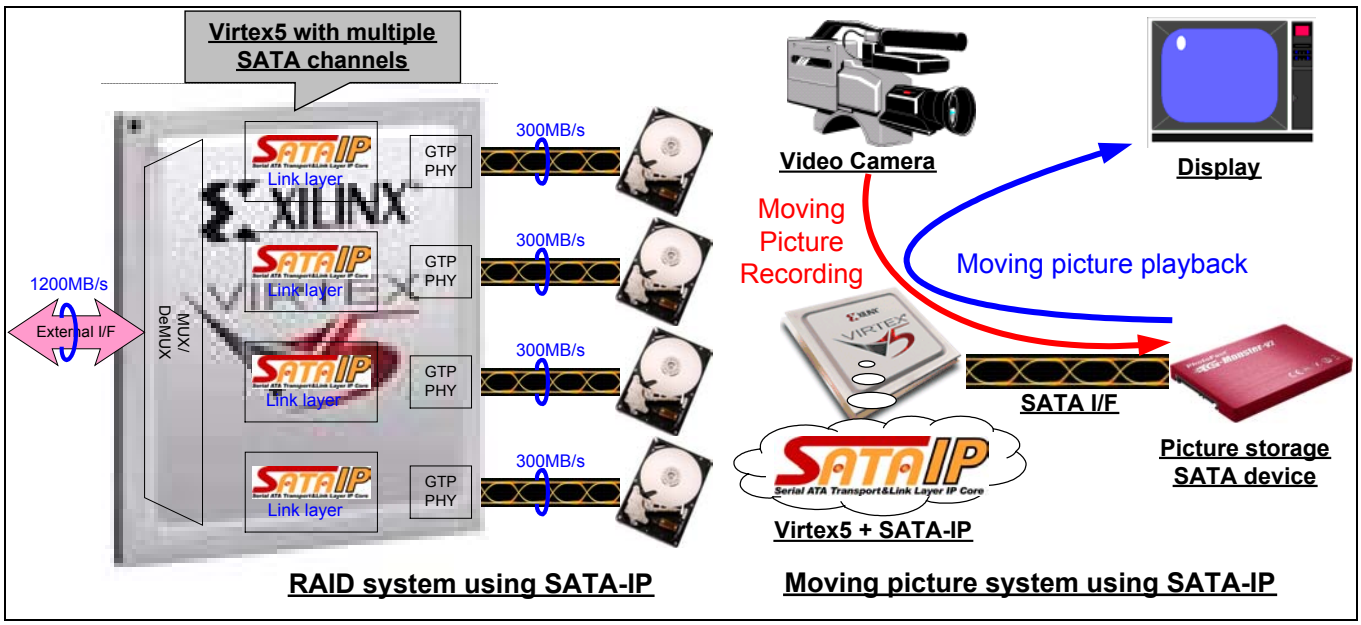
SSD Performance Report

[Rev 1.2] 23 March, 2009

This document reports transfer performance test result of latest SSD drive using SATA-IP.

1. Summary

Because of its expanded storage capacity and reduced cost of the recent storage device, it is general to equip FPGA embedded system with SATA device. In such application, by using SATA-IP, it is possible to develop high-speed and large-capacity RAID system or low-cost and function-rich moving picture system within a very short term and release them into the market.



[Figure1-1] SATA-IP Application System

Generally speaking, HDD is applied in such SATA device embedded system. However, due to the drastic Flash device cost reduction and capacity increase, SSD is rapidly replacing HDD for storage application. SSD is superior to HDD for vibration tolerance, and provides better consecutive burst data transfer speed.

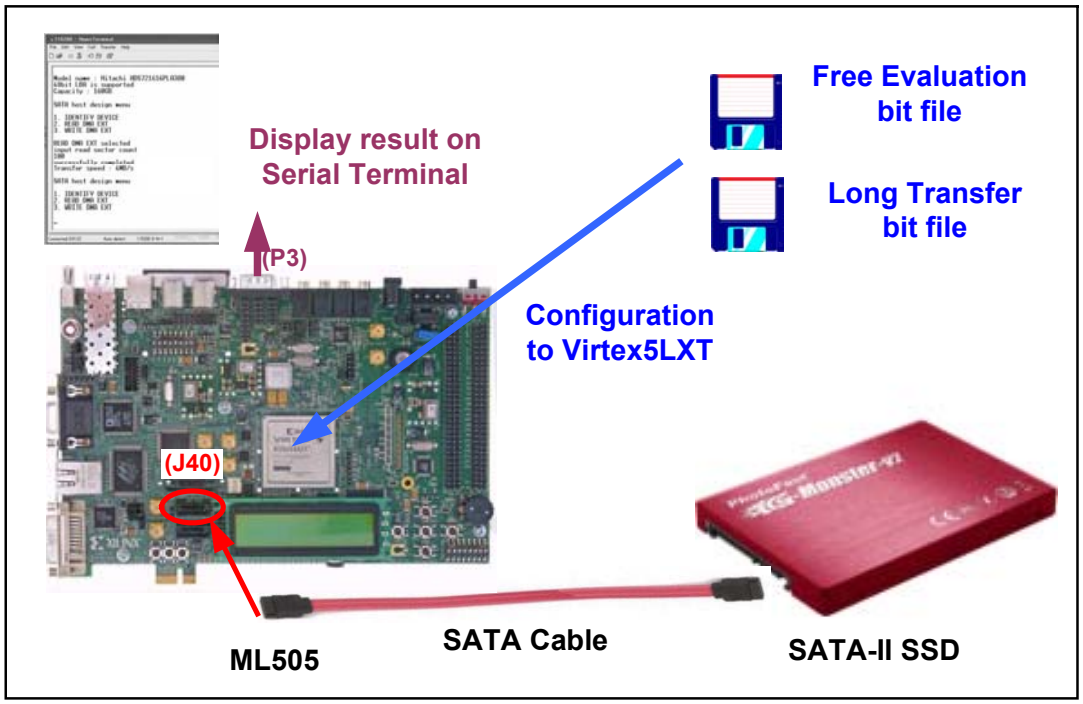
Therefore, DesignGateway evaluated SATA-IP performance using latest SSD available from market. This document reports SSD performance result of the SATA-IP application.

2. Evaluation Condition

2.1 Evaluation Environment

Evaluation environment is shown at figure2-1. This evaluation measures read and write transfer speed between Virtex5LXT on ML505 and SSD drive. Virtex5LXT is configured by evaluation test circuit. Timer counter built in the FPGA measures transfer duration, and displays measured transfer performance on the serial terminal.

Free Evaluation bit file available from SATA-IP web site can display transfer speed after read/write operation, but one read/write access is limited to 32MByte(65,536 sectors) at maximum. However, for example of moving picture system, much more data transfer size in one operation is necessary. Thus, this evaluation prepares another test circuit that can transfer very large size data (Long Transfer bit file) in addition to the Free Evaluation bit file.



[Figure2-1] Measurement environment of transfer performance

2.2 Evaluated SSD

Table 2-1 shows specification of two latest SSD for this evaluation.

	X25-E Extreme	G-Monster V2
Outline		
Vendor	Intel	PhotoFast
Product No.	SSDSA2SH032G1	PF25S128GSSDV2
Capacity	32GB	128GB
Market Price (*)	43,000JPY (480US$)	39,800JPY (445US$)
Write Speed	170MB/s	160MB/s
Read Speed	250MB/s	230MB/s

[Table2-1] SSD specification

(*) Market Price is the lowest price data in Japanese SSD market on February-2009.

In addition, two low-cost SSDs shown at Table2-2 below are evaluated. These SSDs of MLC type can provide very good cost performance to the users. However, because these 32GB SSDs do not support 48bitLBA mode of READ/WRITE DMA EXT command (25H/35H), legacy 28bitLBA mode of READ/WRITE DMA command (C8H/CAH) must be used for data access instead.

	Transcend MLC	Buffalo MLC
Outline		
Vendor	Transcend	Buffalo
Product No.	TS32GSSD25S-M	SHD-NSUM30G
Capacity	32GB	30GB
Market Price (*)	7,980JPY (84US$)	7,800JPY (80US$)
Write Speed	60MB/s	
Read Speed	123MB/s	

[Table2-2] 32GB/MLC type SSD specification of additional evaluation

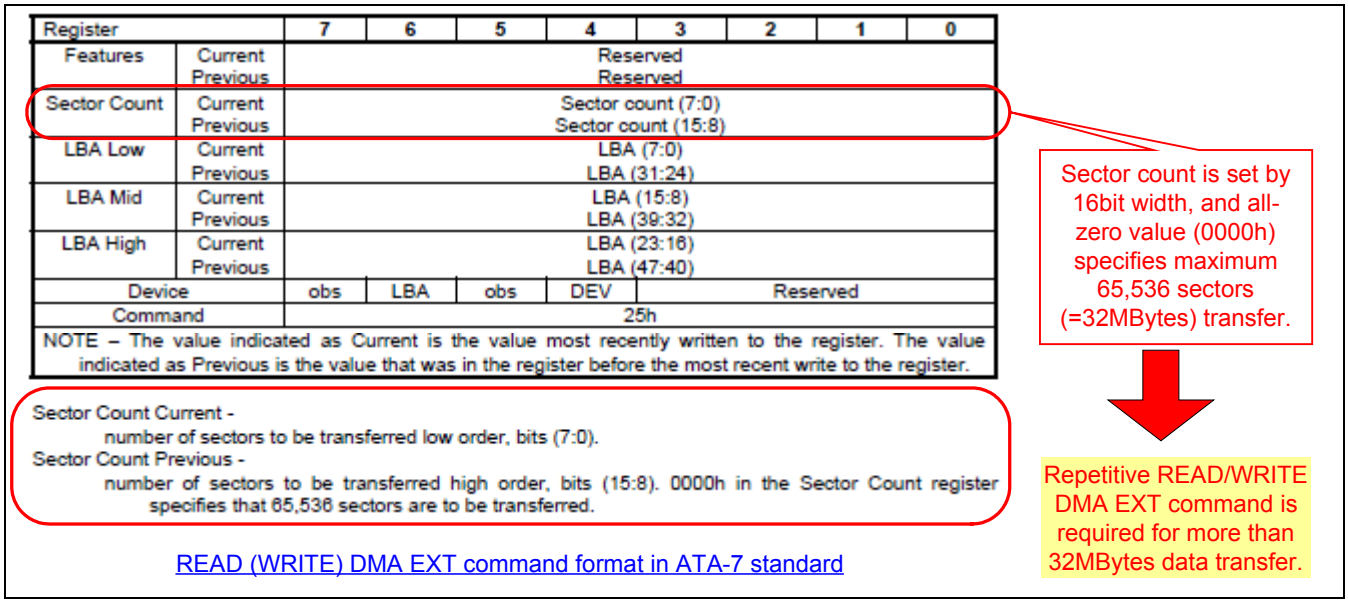
Section 7 of [Low-cost SSD additional evaluation] describes evaluation result of two low-cost SSDs using Long Transfer bit file modified for 28bitLBA support.

3. Evaluation Circuit Implementation

3.1 SATA command format

Generally speaking, data read/write access of the SATA application is done by READ DMA EXT / WRITE DMA EXT command defined in the ATA/ATAPI-7 standard. READ DMA EXT command format is shown at figure 3-1, WRITE DMA EXT command format is almost identical.

This command specifies transfer data count by Sector Count register that has 16bit width. And when command sets this register value to all-zero (0000h), it executes maximum 65,536 sectors (= 32MBytes) data transfer. Therefore, if more than 32MBytes data continuous process is necessary, repetitive command issue with 32MBytes data transfer is required. And it is also necessary to update LBA (target address of data access) parameter value by each command issue.



Register		7	6	5	4	3	2	1	0
Features	Current	Reserved							
	Previous	Reserved							
Sector Count	Current	Sector count (7:0)							
	Previous	Sector count (15:8)							
LBA Low	Current	LBA (7:0)							
	Previous	LBA (31:24)							
LBA Mid	Current	LBA (15:8)							
	Previous	LBA (39:32)							
LBA High	Current	LBA (23:16)							
	Previous	LBA (47:40)							
Device		obs	LBA	obs	DEV	Reserved			
Command		25h							

NOTE – The value indicated as Current is the value most recently written to the register. The value indicated as Previous is the value that was in the register before the most recent write to the register.

Sector Count Current -
number of sectors to be transferred low order, bits (7:0).
Sector Count Previous -
number of sectors to be transferred high order, bits (15:8). 0000h in the Sector Count register specifies that 65,536 sectors are to be transferred.

READ (WRITE) DMA EXT command format in ATA-7 standard

[Figure3-1] SATA command format

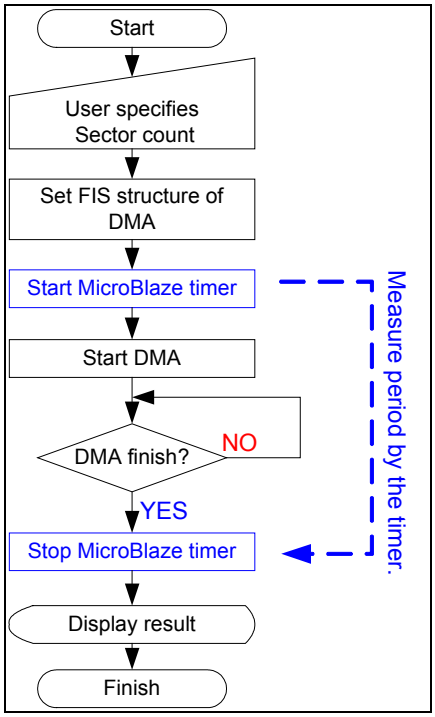
On the other hand, READ DMA / WRITE DMA command is required for such SSD that only supports legacy 28bitLBA mode. In this case, one command issue can cover 128Kbytes consecutive data transfer at maximum.

3.2 Free Evaluation bit file

In the Free Evaluation bit file, internal controller circuit executes only one READ/WRITE DMA EXT command of figure 3-1 when user specifies read/write access. Figure 3-2 below shows flowchart of this controller circuit operation. It sets FIS parameter such as LBA or Sector count value by MicroBlaze firmware before command execution. MicroBlaze timer begins its count up operation just before DMA start, and stops just after DMA finish is detected, so that measured transfer performance by this timer does not include command overhead of MicroBlaze process operation. (Exactly speaking, DMA start process and DMA finish check polling process are done by MicroBlaze, so this timer result includes very little overhead by MicroBlaze firmware.)

Thus, measured transfer performance by this bit file indicates “SATA-IP hardware logic performance including data flow control by SSD drive”. For example of write operation, when SSD internal data buffer is filled by the transferred data, SSD will request data suspend by HOLDp primitive to avoid data overflow. So transfer performance result of this bit file includes such data flow control.

In this evaluation, sector size is set to the maximum sector count value of 65,536 sectors (32MBytes).



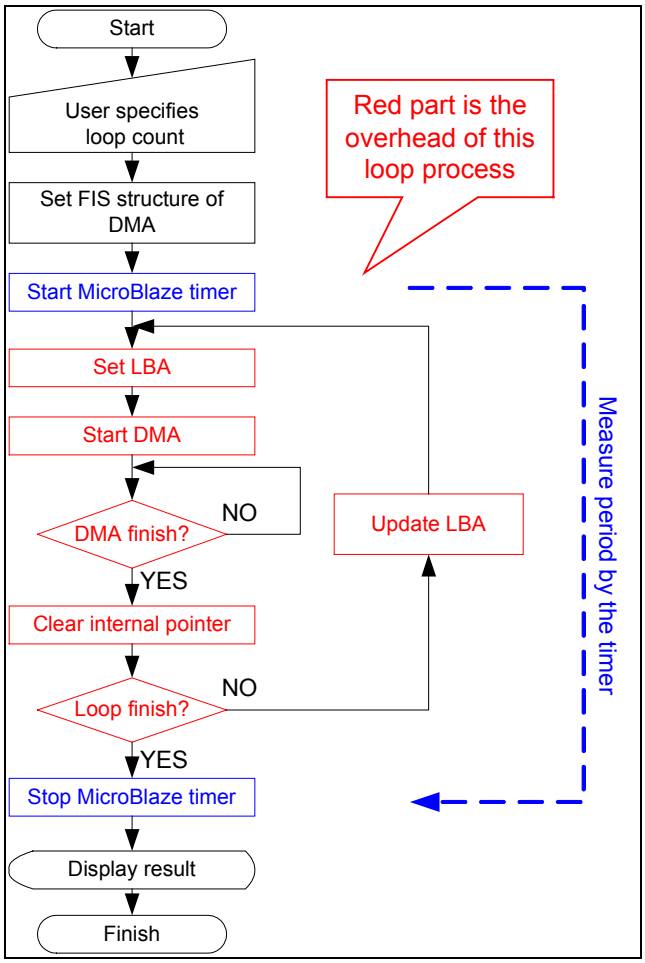
[Figure3-2] Free Evaluation bit file flowchart

3.3 Long Transfer bit file

In the Long Transfer bit file, internal controller executes large size continuous data process by repetitive READ/WRITE DMA EXT command issue with Sector count value=0000h (32MBytes). User can set loop count of this repetitive process prior to this operation. Figure 3-3 below shows flowchart of this controller circuit operation.

Timer measurement period of this controller includes command overhead of FIS parameter update by the MicroBlaze firmware because FIS parameter update is necessary for consecutive command issue. So that this bit file can emulate practical read/write operation of large size data transfer, and can evaluate “real” sustained data transfer speed.

This bit file executes 32MBytes data transfer per one loop, so for example when user specifies 512 loop count, it executes sequential 16GByte data transfer.

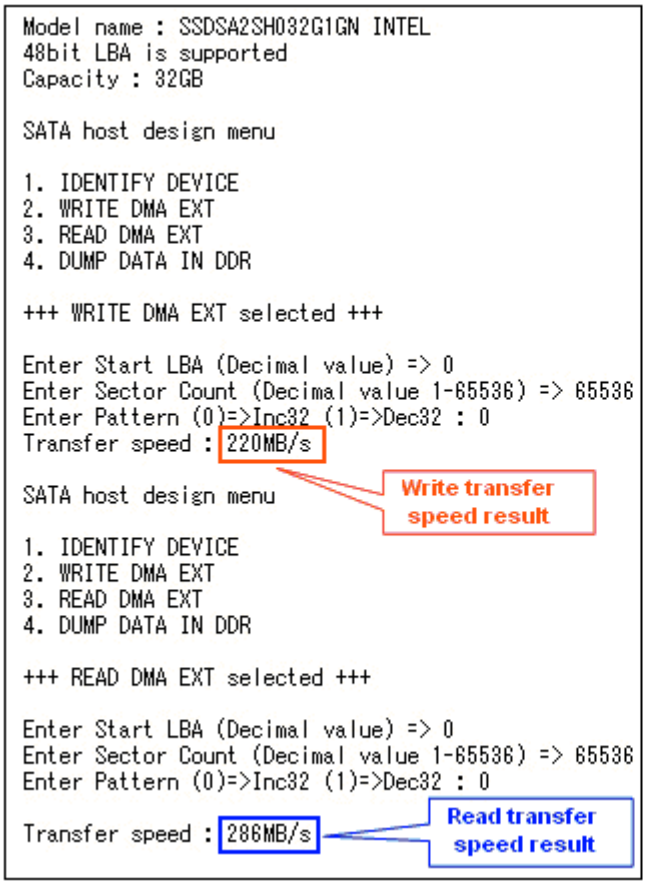


[Figure3-3] Long Transfer bit file flowchart

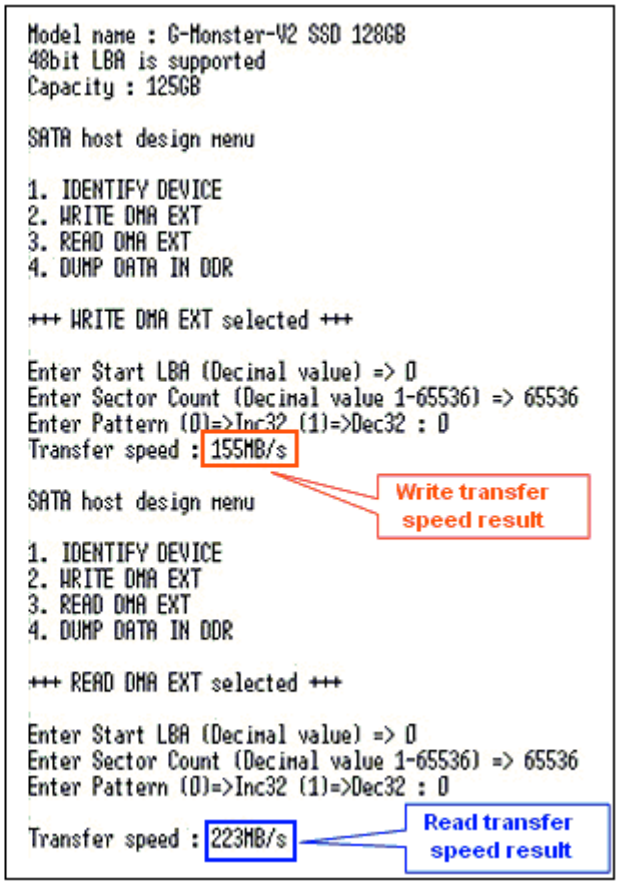
4. Free Evaluation bit file result

Figure 4-1 and 4-2 shows Free Evaluation bit file performance result of X25-E Extreme and G-Monster V2, respectively. In the X25-E Extreme performance, read speed result of 286MB/s is close to the maximum speed of 300MB/s in SATA-II standard, so it seems that data flow control is scarcely occurred during read data transfer. On the other hand, 220MB/s of write speed in figure 4-1 indicates that some data flow is generated during write data transfer.

Read=223MB and write=155MB/s result of G-Monster V2 is a little slower than that of X25-E Extreme, but it can almost achieve vendor specification performance of read=230MB/s and write=160MB/s.



[Figure4-1] X25E Extreme result



[Figure4-2] G-Monster V2 result

5. Long Transfer bit file result

For the Long Transfer bit file, evaluation is done by 10 loop count settings of 1,2,4,8,16,32,64,128,256, and 512. Evaluation result of X25E Extreme and G-Monster V2 is shown at table 5-1 and 5-2, respectively. From this evaluation result, there is a tendency that large size data such as 16GB transfer speed is a little slower than small data size, but transfer performance is almost stable in all transfer size result.

As a conclusion of this result, even very large size data such as 16GB transfer can still keep high transfer speed as well as small data size.

Loop cnt	xfr byte cnt	Write result	Read result
1	32MB	217.04 [MB/s]	286.60 [MB/s]
2	64MB	218.11 [MB/s]	286.35 [MB/s]
4	128MB	216.52 [MB/s]	286.36 [MB/s]
8	256MB	214.40 [MB/s]	286.23 [MB/s]
16	512MB	212.45 [MB/s]	286.40 [MB/s]
32	1GB	215.18 [MB/s]	285.99 [MB/s]
64	2GB	210.15 [MB/s]	285.89 [MB/s]
128	4GB	210.22 [MB/s]	286.32 [MB/s]
256	8GB	209.40 [MB/s]	283.93 [MB/s]
512	16GB	209.73 [MB/s]	285.25 [MB/s]
Average		213.32 [MB/s]	285.93 [MB/s]

[Table5-1] X25E Extreme result

Loop cnt	xfr byte cnt	Write result	Read result
1	32MB	157.39 [MB/s]	223.25 [MB/s]
2	64MB	157.18 [MB/s]	223.24 [MB/s]
4	128MB	157.16 [MB/s]	223.23 [MB/s]
8	256MB	157.76 [MB/s]	223.22 [MB/s]
16	512MB	155.15 [MB/s]	223.22 [MB/s]
32	1GB	155.99 [MB/s]	223.21 [MB/s]
64	2GB	155.45 [MB/s]	223.20 [MB/s]
128	4GB	153.32 [MB/s]	217.64 [MB/s]
256	8GB	154.07 [MB/s]	220.93 [MB/s]
512	16GB	154.31 [MB/s]	222.02 [MB/s]
Average		155.78 [MB/s]	222.32 [MB/s]

[Table5-2] G-Monster V2 result

6. Command Overhead evaluation

6.1 State-Machine Controller evaluation

For large size data transfer, this evaluation added test design of State-Machine based controller to compare FIS parameter update overhead between software and hardware process. MicroBlaze firmware that executes loop process in the red part of figure 3-3 is replaced by the pure hard-wired logic of this State-Machine based controller. So State-Machine based controller design can minimize elapsed time of overhead process.

In this evaluation, transfer byte count is fixed to 16GBytes which is equivalent to 512 loop count of Long Transfer bit file, and checked by G-Monster V2 SSD. Table 6-1 shows this State-Machine based controller design result. (Table 6-1 also includes 16GBytes result of Long Transfer bit file for comparison.)

Circuit	Loop process	xfer size	Write result	Read result
Long Transfer bit file	MicroBlaze firmware	16GB	154.31 [MB/s]	222.02 [MB/s]
State-Machine controller	Pure logic hardware	16GB	159.93 [MB/s]	223.22 [MB/s]

[Figure6-1] 16GBytes large size data transfer result of G-Monster V2

This result shows that difference of FIS parameter settings overhead between MicroBlaze firmware and State-Machine hardware is very little. This is because READ/WRITE DMA EXT command can process 32MBytes data in one time, so command overhead duration is quite small compared with whole data transfer time.

6.2 Command process duration measurement

The final evaluation is to measure real command overhead process duration of READ/WRITE DMA EXT in each of Long Transfer bit file and State-Machine controller test design. Table 6-2 shows this evaluation result.

Circuit	WRITE	READ
Long Transfer bit file	1600us	3.0us
State-Machine controller	450us	1.0us

[Table6-2] Real command overhead duration result

In write operation, when transfer speed is 150MB/s, 32MBytes data transfer time will be 213ms. On the other hand, ratio of write command overhead to transfer time is 0.21% in State-Machine controller and is 0.75% in Long Transfer bit file. So anyway, they are almost negligible compared with data transfer time that will occupy more than 99% of total command execution time.

In read operation, this tendency is more remarkable. When read transfer speed is 220MB/s, 32MBytes data transfer time will be 145ms. So ratio of read command overhead to transfer time is only 0.0007% in State-Machine controller and is still 0.002% in Long Transfer bit file.

As a conclusion of this result, when READ/WRITE DMA EXT command with 32MBytes data transfer is used, command overhead is negligible even by the MicroBlaze firmware. So State-Machine controller design can't get so much performance improvement.

Because DesignGateway provides SATA-IP customer with SATA-Host reference design that can be a template of MicroBlaze based controller, customer can minimize product development time by starting from this reference design. Moreover, MicroBlaze firmware can save ISE compilation of logic circuit that consumes long time, it can also be a strong merit of flexibility and immediate response of possible bug fix.

Therefore, if MicroBlaze implementation is possible in the product, MicroBlaze controller should be the best way because performance penalty is negligible and should not influence transfer performance.

7. Low-cost SSD additional evaluation

Additional evaluation of low-cost MLC type SSD (shown at Table 2-2) is done in March-2009. These SSDs can provide advantage of high cost performance. Because these 32GBytes low-cost SSD do not support 48bitLBA mode, they can only be accessed through legacy 28bitLBA mode.

For this evaluation, Long Transfer bit file is modified to support 28bitLBA. This modified test circuit is limited to 256 sectors consecutive data transfer at maximum, so command process overhead is multiplied by 256 from that of 48bitLBA mode.

Total data transfer size of this evaluation is 32GBytes that is equal to the whole storage area of SSD. Table 7-1 below shows evaluation result.

SSD	data size	sector size	WRITE	READ
Transcend MLC	32GB	62,586,880	91 [MB/s]	147 [MB/s]
Buffalo MLC	32GB	62,586,880	89 [MB/s]	147 [MB/s]

[Table7-1] Real command overhead duration result

This evaluation result indicates that both MLC type SSD can achieve similar performance (around 90MB/s for WRITE and 147MB/s for READ) under large data process condition of whole storage area. Actually speaking, measured data transfer performance of low-cost SSD is a little inferior to that of upper grade SSD shown in Table 2-1. However for the cost-value product, even low-cost MLC type SSD can still offer attractive enchantment to the users if this performance is acceptable.

8. Conclusion

With the latest SSD and SATA-IP, customer can build high-speed and large-capacity data storage system without speed down penalty at large data transfer

And by using READ/WRITE DMA EXT command that can ignore command overhead, MicroBlaze firmware controller is practical so that SATA-IP reference design template can minimize product development time.

For cost-value products, low-cost MLC type SSD application can enable SATA-IP customers to build price competitive products with sufficient data transfer performance.

9. Revision History

Revision	Date	Description
1.0	3-Feb-2009	Initial version release
1.1	7-Feb-2009	Added command overhead measurement result
1.2	23-Mar-2009	Added low-cost MLC SSD evaluation result