[SATA-IP アプリケーション・ノート 1]

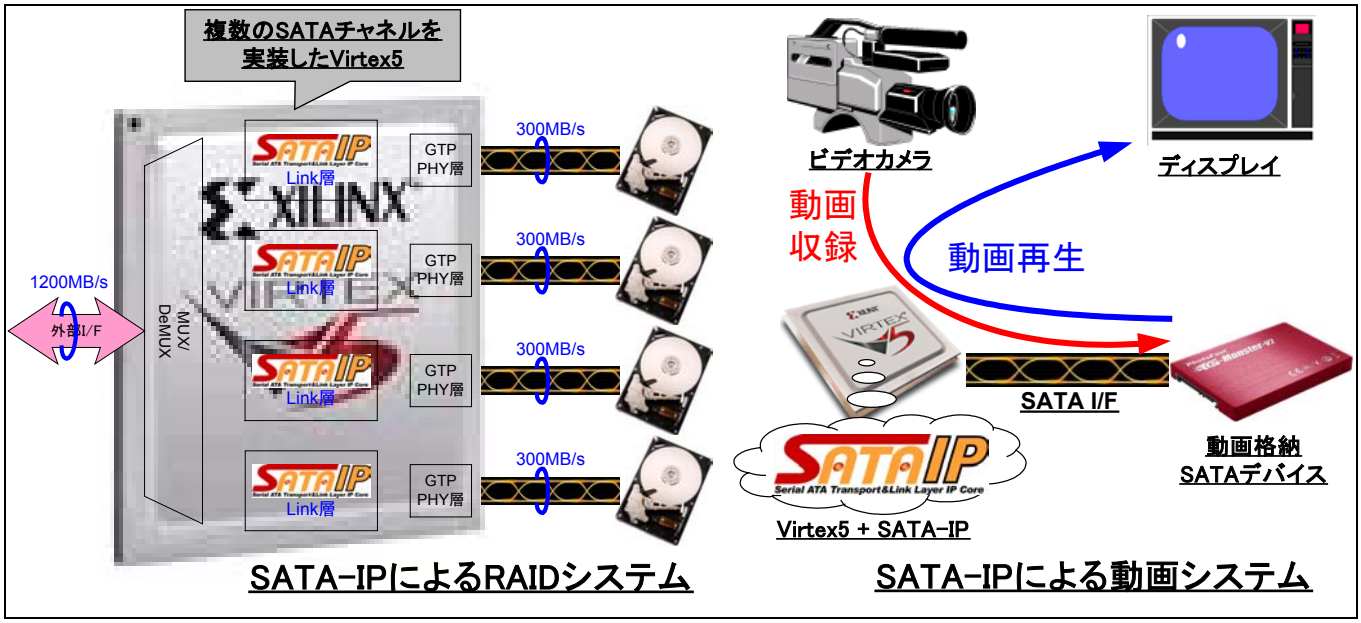
SSD パフォーマンス・レポート

Rev 1.2　2009 年 3 月 23 日

本ドキュメントは SATA-IP を最新の高速 SSD ドライブと接続し転送パフォーマンスを実測した結果レポートです。

1. 概要

近年のストレージ・デバイスの大容量化と低価格化に伴い、FPGA を使った組み込みシステムに SATA デバイスを応用するアプリケーションが一般的になりつつあります。 このようなアプリケーションにおいては、SATA-IP を活用することで、図 1-1 のような高速大容量の RAID システムや、低価格かつ高機能な動画システムを短期間で開発し、製品をいちはやく市場に投入することが可能となります。



[図 1-1] SATA-IP の応用システム例

従来、SATA デバイス組み込みシステムでは主に HDD が用いられてきました。 ところが Flash デバイスの急激な低価格化と大容量化によって、HDD にかわって SSD が使われるケースが飛躍的に増えています。SSD は HDD と比較して耐振性に優れ、またバーストデータの転送速度面でも有利です。 容量に対してのコスト面でも HDD と同等レベルに近づいており、SSD のメリットがより一層注目されています。

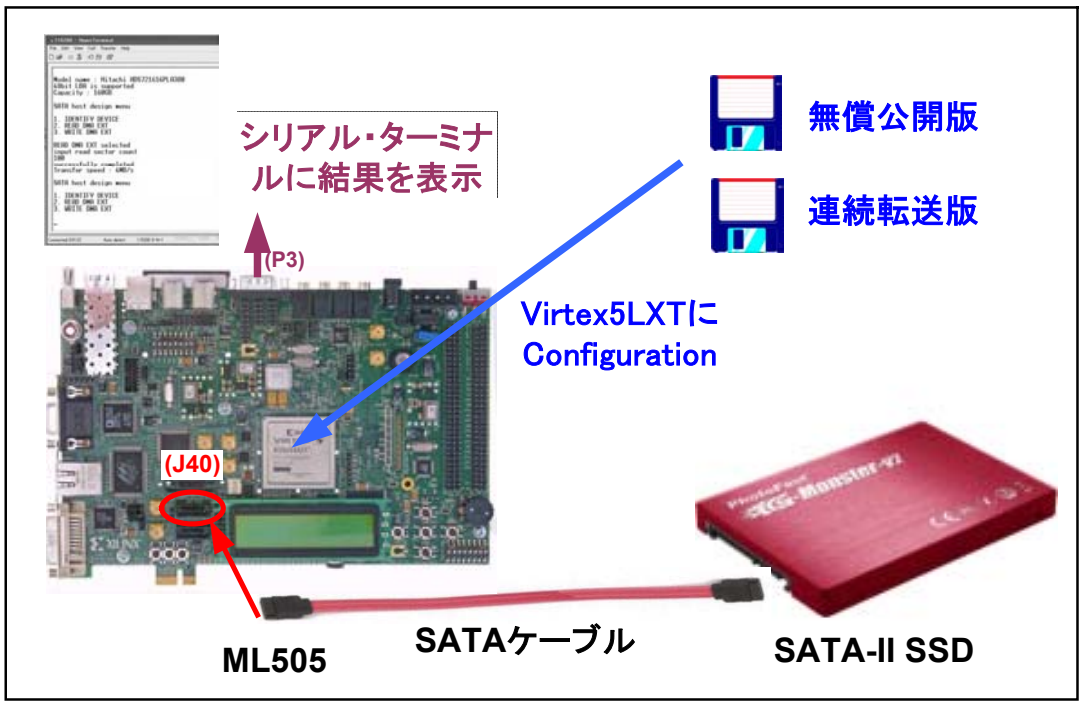
このような背景から、SATA-IP アプリケーションとして市場ニーズが高い SSD をターゲットとした SATA-IP の転送パフォーマンスを、ML505 ボードと最新 SSD を使って実測しました。

2. 評価条件

2.1 評価環境

今回の評価環境を図 2-1 に示します。 ML505 ボード上の Virtex5LXT に評価用の回路データをコンフィグレーションし、SATA 接続した SSD ドライブに対するリードライトの転送速度を評価します。 その転送所要時間は FPGA 内部タイマで計測されるので、その結果をシリアル・ターミナルで表示します。

評価用に無償公開しているビットファイル（無償公開版）でもリード・ライト実行後に転送速度を表示できますが、一度のリードライト・アクセスは 32MByte(65,536 セクタ)以下に制限されます。 しかし例えば動画システムなどでは、それ以上の大量データを連続転送するケースが一般的です。 そこで今回、評価用ビットファイルに加えて大量データを連続転送する回路（連続転送版）を追加して評価を実施しました。



[図 2-1] 転送パフォーマンスの測定環境

2.2 評価対象 SSD

今回の評価は下表 2-1 に示す(2009 年 2 月の時点で)最新の2点の SSD を使って実施しました。

	X25-E Extreme	G-Monster V2
外形		
メーカ	Intel	PhotoFast
型番	SSDSA2SH032G1	PF25S128GSSDV2
容量	32GB	128GB
市場価格 (※)	￥43,000	￥39,800
Write 公表値	170MB/s	160MB/s
Read 公表値	250MB/s	230MB/s

[表 2-1] 評価した SSD の仕様

(※) 市場価格は、2009 年 2 月上旬時点での[価格.com]における最低価格情報

また、現在 SSD 市場でも競争が激しくコスト・パフォーマンスが良い MLC タイプの SSD を対象として、下表 2-2 に示した2種類の低価格 SSD を追加で評価しました。これら 32GByte の低価格 SSD では調査したところ 48BitLBA モードの READ/WRITE DMA EXT コマンド(25H/35H)がサポートされていないため、従来の 28bitLBA モードの READ/WRITE DMA コマンド(C8H/CAH)でリードライト・アクセスを行う必要があります。

	Transcend MLC	Buffalo MLC
外形		
メーカ	Transcend	Buffalo
型番	TS32GSSD25S-M	SHD-NSUM30G
容量(※1)	32GB	30GB
市場価格 (※2)	￥7,980	￥7,800
Write 公表値	60MB/s	
Read 公表値	123MB/s	

[表 2-2] 追加で評価した低価格 SSD の仕様

(※1) 容量の公表値は Transcend/Buffalo で異なるが実容量(最大 LBA)は両社とも同じで 62,586,880 であった。
(※2) 市場価格は、2009 年 3 月下旬時点での[価格.com]における最低価格情報

追加評価は表 2-2 に示す2台の SSD を対象とし、28bitLBA サポートを追加した連続転送版で実施しました。 その結果を 7 章の[低価格 SSD の追加評価]で報告します。

3. 評価回路の実装

3.1 SATA のコマンド・フォーマット

SATA アプリケーションにおいて、大容量データのリード・ライトは ATA-7 規格で定義された READ DMA EXT / WRITE DMA EXT コマンドが使われます。 READ DMA EXT コマンドのフォーマットを下図 3-1 に示しますが WRITE DMA EXT コマンドのフォーマットもほぼ同様です。

このコマンドにおいて実行処理するデータ数は Sector Count レジスタで設定しますが、このレジスタは全部で 16bit 幅でありその値をオール・ゼロ(0000h)とした場合に最大の 65,536 セクタ= 32MByte (1 セクタ=512 バイト)が指定されます。 従って 32MByte 以上のデータを連続転送する場合、32MByte ごとに本コマンドを繰り返して連続発行する必要があります。 そしてコマンド発行するたびに LBA(アクセス先番地)も更新する必要があります。

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.

ATA-7規格によるREAD DMA EXTコマンドのフォーマット
(WRITE DMA EXTコマンドも同様)

セクタカウントは16bitで設定され0000hの場合最大の65,536セクタ(=32Mバイト)転送が設定される

32Mバイト以上の転送はREAD/WRITE EXT DMAコマンドを32Mバイトごとに繰り返して発行する必要がある

[図 3-1] SATA コマンドのフォーマット

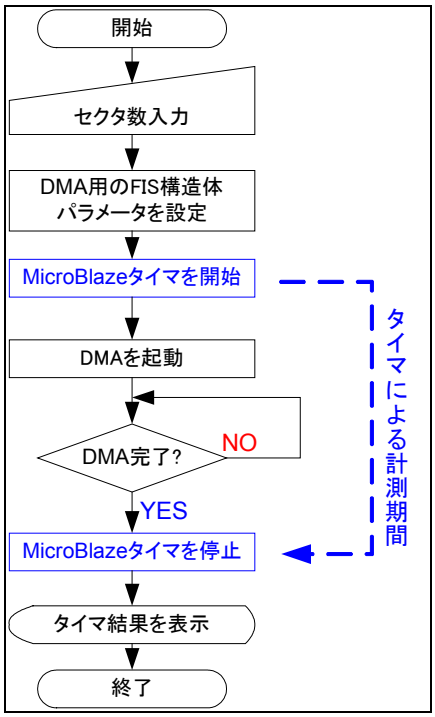
一方、28bitLBA のみ対応するドライブの場合、従来の READ DMA または WRITE DMA コマンドを使う必要があります。 この場合 1コマンドにおいて最大 256 セクタ = 128KByte までの連続データ転送が可能となります。

3.2 無償公開版

無償公開版の回路においては、一度のリードライト・アクセス指示に対して図 3-1 の READ/WRITE DMA EXT コマンドを1コマンドのみ発行します。下図 3-2 にこの回路の実行フローチャートを示します。コマンドを実行する前に、アクセス先の LBA や SectorCount 等 FIS パラメータを MicroBlaze のファームウエアによって設定します。 また、MicroBlaze タイマを起動した直後に DMA を起動し DMA 転送の完了直後にタイマが停止するため、タイマによって計測された転送パフォーマンスには MicroBlaze によるコマンド発行のためのオーバーヘッドが含まれません。

つまりこの回路による計測結果は、「SSD ドライブによるフロー制御を含めた、SATA-IP ハードウエア・ロジックによる転送パフォーマンス」となります。（ただし DMA 起動処理や DMA 完了待ちのポーリングは MicroBlaze によって行われるため、正確には若干のファームウエア処理オーバーヘッドが含まれます。）

例えばライトなどで SSD のデータバッファが溜まってきた場合、オーバーフローを防ぐため SSD から SATA-IP に対して HOLDp プリミティブによる転送の一時停止を要求してきますので、計測結果はそのようなフロー制御が含まれたパフォーマンスとなります。 今回はセクタ数を最大の 65,536 セクタ(32MByte)に固定して評価を行っています。



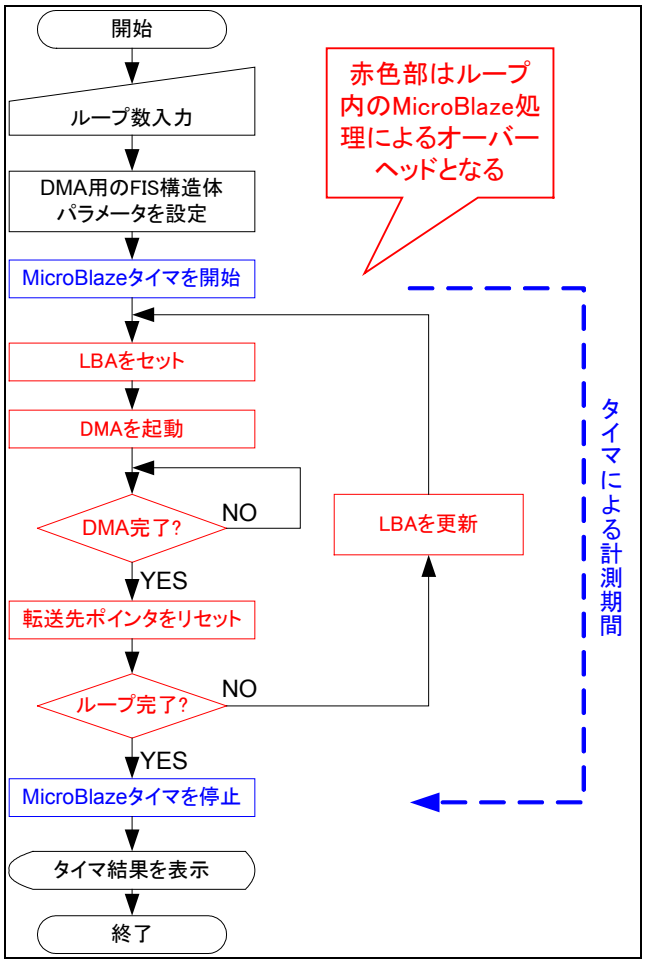
[図 3-2] 無償公開版の実行フローチャート

3.3 連続転送版

連続転送版の回路においては、図 3-1 の READ/WRITE DMA EXT コマンドを SectorCount=0000h の最大転送(32MByte)とした状態でユーザが指定したループ回数分繰り返すことで大量のデータ処理を連続実行します。 下図 3-3 に連続転送版の実行フローチャートを示します。

この回路においては、転送パフォーマンス測定用のタイマ計測中に、コマンドの連続発行で必要となる MicroBlaze によるパラメータの更新処理オーバーヘッドが含まれます。 従って、本回路は大量データの連続リード/ライトの実処理をエミュレートしたものとなり、実効転送パフォーマンスそのものの計測が可能となります。

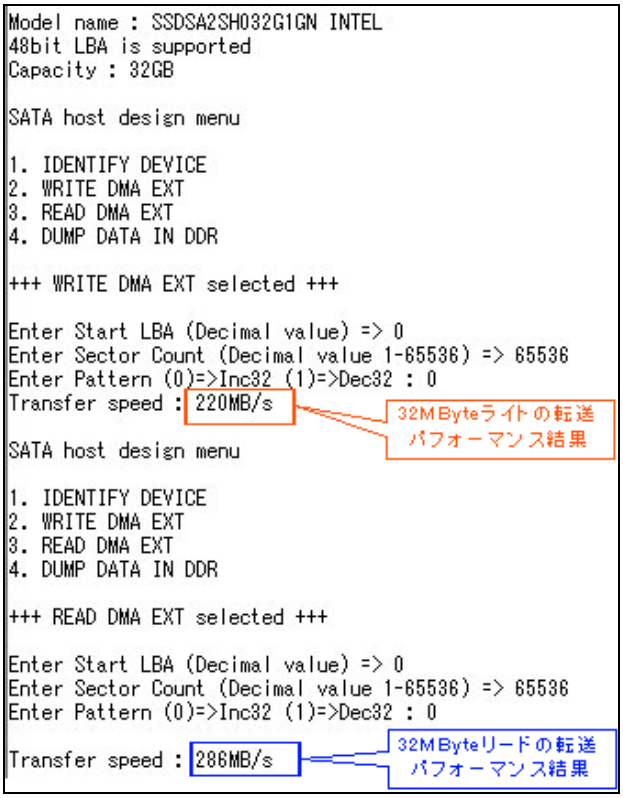
1ループにつき 32MByte 分の処理となるため、例えば 512 ループを指定した場合は 16GByte の連続処理が行われることになります。



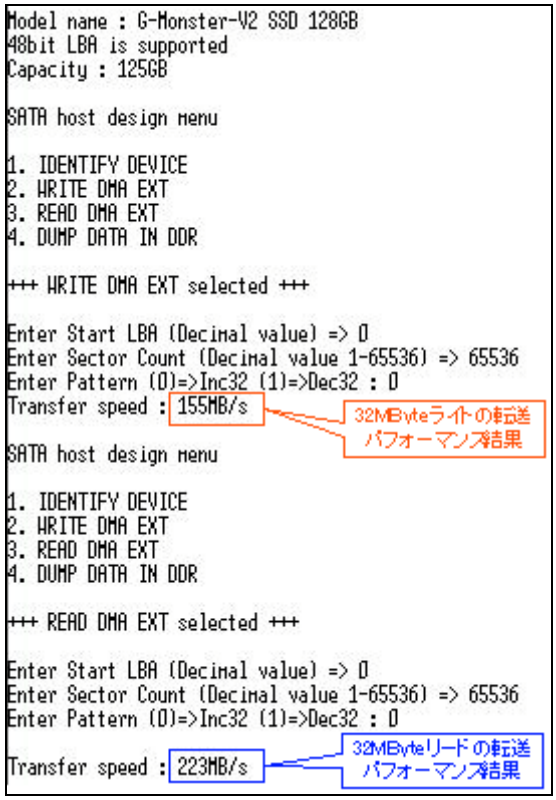
[図 3-3] 連続転送版の実行フローチャート

4. 無償評価版の結果

無償評価版による X25E Extreme と G-Monster V2 の評価結果をそれぞれ下図 4-1 と図 4-2 に示します。X25E Extreme にてリード時の 286MB/s は SATA-II 規格上最大の 300MB/s に近い結果となっていることから、リード転送中のフロー制御はほとんど発生しなかったものと考えられますがライトでは 220MB/s なので、SSD 側から若干のフロー制御が発生しています。 G-Monster V2 の結果は X25-E Extreme に比べると若干低くなっています(リード時 223MB/s ライト時 155MB/s)が、製品メーカの公表値(リード時 230MB/s ライト時 160MB/s)はほぼ達成できています。



[図 4-1] 無償評価版の X25E Extreme 評価結果



[図 4-2] 無償評価版の G-Monster V2 評価結果

5. 連続転送版の結果

連続転送版ではリードとライトそれぞれにおいて、ループ数を 1,2,4,8,16,32,64,128,256,512 の 10 通りとして評価を行いました。 X25E Extreme と G-Monster V2 の評価結果をそれぞれ下表 5-1 と表 5-2 に示します。

この結果から、両社とも転送データ量が増えるとパフォーマンスはほんの僅か低下する傾向が見られますが、ほぼ安定した結果が得られています。 即ち、これらの SSD では例えば 16GB 程度の大量データの連続リードや連続ライトにおいても、少量の場合と同等の高い転送速度を維持し続けることができるものと考えられます。

Loop 数	転送バイト数	Write 結果	Read 結果
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]
平均		213.32 [MB/s]	285.93 [MB/s]

[表 5-1] 連続転送版の X25E Extreme 評価結果

Loop 数	転送バイト数	Write 結果	Read 結果
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]
平均		155.78 [MB/s]	222.32 [MB/s]

[表 5-2] 連続転送版の G-Monster V2 評価結果

6. コマンド・オーバーヘッドの調査

6.1 ハード置換え版の評価

更に大量データの連続転送においてコマンド発行のたびに必要となる FIS パラメータ更新などのオーバーヘッドを、ソフトウエアのかわりにハードウエアで実装した場合のパフォーマンスについても評価回路(ハード置換え版)を試作して調査しました。

具体的には図 3-3 のフローチャートにて、ループ中に MicroBlaze によって行われる図中赤色のファームウエア処理を、ステートマシンによるハードウエアに置き換えて自動実行するテストロジックを実装しました。 この評価は転送バイト数を 16GB(連続転送版の 512 回のループと同じ)に固定して G-Monster V2 を対象として実施しています。その結果を下表 6-1 に示します。 (比較検討のためファームウエア処理による連続転送版の 16GB 結果を再掲しています。)

bit ファイル	Loop 処理	転送バイト数	Write 結果	Read 結果
連続転送版	ファームウエア	16GB	154.31 [MB/s]	222.02 [MB/s]
ハード置換え版	ステートマシン	16GB	159.93 [MB/s]	223.22 [MB/s]

[表 6-1] G-Monster V2 による 16GB の大量データ転送結果

この結果から、FIS パラメータ設定等によるコマンド・オーバーヘッドは、MicroBlaze によるファームウエアとステートマシンによるハードウエアであまり大きな差異はないことがわかります。 これはすなわち、READ/WRITE DMA EXT コマンドでは1コマンドで 32MByte ものデータを一度に転送できるため、データ転送時間に対してのコマンド・オーバーヘッド処理時間の割合が非常に小さいためだと考えられます。

6.2 コマンド処理時間の実測

最後に連続転送版とハード置換え版それぞれで、READ/WRITE DMA EXT コマンドの処理オーバーヘッド実時間を実測したのでその結果を下表 6-2 に示します。

bit ファイル	WRITE	READ
連続転送版	1600us	3.0us
ハード置換え版	450us	1.0us

[表 6-2] オーバーヘッド処理時間の計測結果

データ・ライトにおいて転送速度を 150MB/s とすると 32MByte のデータ転送時間は 213ms となります。 これに対してコマンド処理オーバーヘッドは、データ転送時間と比較してハード置換え版で 0.21%、連続転送版でも 0.75%にしかなりません。 いずれにしてもコマンド全実行時間でほとんど(99%以上)を占めるデータ転送時間に対して無視できるレベルです。

データ・リードの場合その傾向はもっと顕著です。 リード転送速度を 220MB/s とすると 32MByte のデータ転送時間は 145ms となりますので、ハード置換え版の 1.0us はデータ転送時間の 0.0007%、連続転送版の 3.0us でも 0.002%にしかなりません。

結論としては、1コマンドで 32MByte データを処理できる READ/WRITE DMA EXT コマンドを使う場合、コマンド・オーバーヘッドは MicroBlaze によるファームウエア処理であっても十分無視できるレベルであって、ステートマシン処理に変更した場合でもパフォーマンスの改善にはほとんど効果がないと考えられます。

SATA-IP では顧客提供のホスト向けリファレンス・デザインにて MicroBlaze ベースのホスト・コントローラのテンプレート（雛形）がすでに用意されているため、リファレンス・デザインをもとにして製品システムを構築するのが開発期間上最もメリットがあります。 また、MicroBlaze のファームウエアの場合、ハードウエア実装に比べてアップデートに時間のかかる ISE での回路再コンパイルが不要となるので、柔軟性や不具合発生時の即応性でも有利です。
従って、顧客の製品システムで MicroBlaze が実装可能な場合は、パフォーマンス・ペナルティは十分小さいためコマンド処理を MicroBlaze ファームウエアで行うのが最適であると考えられます。

7. 低価格 SSD の追加評価

更に、SSD 市場での激しいメーカ間競争でコスト・パフォーマンスが高い MLC タイプの低価格 SSD を 2009 年 3 月に追加して評価しました。 今回評価した 32GByte タイプでは 48bitLBA モードがサポートされていないため、従来の 28bitLBA モードでのみアクセスが可能です。
そこで連続転送版を改良し 28bitLBA コマンドに対応することで評価を実施しました。 この追加評価においては、1 コマンド当りの最大セクタ数が 256 セクタに制限されるため、コマンド処理オーバーヘッドは 48bitLBA 版に比べて 256 倍に増えることになります。
評価は転送バイト数=32GByte となる、全面ライト/全面リード・アクセスで実施しました。 その結果を下表 7-1 に示します。

SSD 種類	転送バイト数	処理セクタ数	Write 結果	Read 結果
Transcend MLC	32GB	62,586,880	91 [MB/s]	147 [MB/s]
Buffalo MLC	32GB	62,586,880	89 [MB/s]	147 [MB/s]

[表 7-1] 追加評価した低価格 SSD の評価結果

この結果から、どちらもほぼ同程度のパフォーマンス（ライト時 90MB/s 程度、リード時 147MB/s）が、全面リードライトのような大容量データ転送にて確認できました。 最初に調査した表 2-1 のような上位機種と比較するとパフォーマンスは確かに低下します。 しかしコスト面を重視したシステムであって表 7-1 のパフォーマンスで満足できるのであれば、安価な MLC タイプの SSD においても十分な実力を発揮できます。

8. 結論

最新の SSD を SATA-IP と組み合わせてストレージ応用製品を実装することで、大量データの高速リードライトを実現するシステムが構築できます。
さらに、処理オーバーヘッドの影響が事実上無視できる READ/WRITE DMA EXT コマンドを使うことで MicroBlaze ファームウエアでのシステム開発が可能となるため、SATA-IP リファレンス・デザインのテンプレートを活用することで短期間での製品開発に貢献します。
コスト重視のシステムにおいては、安価な MLC タイプの SSD を活用することで、十分なパフォーマンスを維持したままでの低コスト製品が SATA-IP によって実現可能となります。

9. 改版履歴

リビジョン	日付	内容
1.0	2009/2/3	初版作成
1.1	2009/2/4	コマンド・オーバーヘッドの実測結果を追加
1.2	2009/3/23	32GB/MLC の低価格 SSD 評価結果を追加