2L-09

FPGAを用いたストレージコントローラの実装と評価

紀野國祐太†
筑波大学大学院システム情報工学研究科†

山口佳樹‡§
筑波大学システム情報系‡
筑波大学計算科学研究センター§

1. はじめに

マイクロソフト社は、Catapult project[1]、Brainwave project [2]、Azure [3]など、FPGA導入した製品を次々と発表している。また、Amazon 社の Amazon EC2 F1 インスタンス [4]、IBM の SuperVessel [5]など、クラウド大手各社も、FPGAの利用を積極的に考えている。

筑波大学計算科学研究センターでは、運用していたスパコン「HA-PACS/TCA (PACS-8) 」において、密結合並列計算演算機構として FPGA 利用を提案しており、2019 年度より運用されるCygnus(PACS-X)では、この機構をさらに拡張し、FPGA により更なる加速をサポートすることが計画されている。

これらのシステムにおいて、演算性能は GPUを初めとする並列処理技術が、ノード間データ転送性能（ネットワーク性能）については FPGAなどによる高速通信により強化されてきたといえる。しかし、データストレージ部については、物理的制約もあり大きな性能向上がみられず、システム全体性能を高める上で大きな課題として残っている。

そこで、本研究では FPGA を用いたストレージアクセス加速に注目し、Solid State Drive(SSD)に代表される大容量半導体素子メモリストレージを FPGA に直接接続し、更には、演算部とストレージを密に接続することによって、大容量かつ高速なデータストレージと演算機構一体型システムの構築について提案する。本報告では、その中心的要素技術である FPGA によるストレージ制御機構について紹介する。

2. Serial ATA 規格

広く利用されているインタフェース規格、実効転送性能、および FPGA 1 個あたりに理論的に接続可能な台数を表１にまとめる。

表１ インタフェース規格（SATA, SAS, PCIe）

インタフェース規格	SATA Rev 3.x	SAS-3	PCIe 3.0 x4
実効転送性能 [MB/s]	600	1,200	**3,938**
高速シリアル IO 数 [本]	**1**	2	4
接続可能台数 [個数/FPGA]	**128**	64	32
総転送性能 [MB/s]	76,800	76,800	**126,016**
総容量 [TB] (4TB SSD 使用時)	**512**	256	128

Serial ATA 規格 [6]は、Rev 3.x において、実効転送性能は 600[MB/s]、SAS-3 は 1200[MB/s]、PCI Express 3.0 x4 は 3938[MB/s]となっている。これらを FPGA に接続する場合、高速トランシーバの利用が要求される。現在入手可能な Xilinx 社の最大規模の FPGA パッケージ [7] に搭載されているトランシーバ数は 128 であり、SATA は 128 台、SAS は 64 台、PCIe 3.0 x4 は 32 台接続可能であることがわかる。そこで総転送性能を考えると、FPGA 1 個で、PCIe なら約 123[GB/s]、SATA や SAS でも 75 [GB/s]の転送性能に達する。この性能は、Intel i9-7980XE などのハイエンドCPU のメモリ帯域 (85.2 [GB/s]) と比べてもそん色なく大変魅力的な数字である。

一方、補助記憶装置として重要視される要因の一つにストレージ容量がある。これを重視した場合、SATA (512[TB]) 、SAS (256[TB]) に比べ PCIe (128[TB]) は接続可能ストレージ数の関係から大きく見劣りする。また、FPGA 内でのPCIe Root Complex の配置（ハード IP では不足するため、ソフト IP での効率的配置）やリソース使用量の関係から期待される性能を出すことが難しいことが推測される。一方、SATA は FPGAへの接続に必要なトランシーバ数が最も少なく、数多くの台数を接続することが可能である。そのため、一デバイスあたりの実効転送性能は他のインタフェースに劣るが、FPGA 内部回路の設計を含め、全体の速度を向上させるという意味では最も望ましい。

そこで本報告では、フィジビリティスタディとし、FPGA に 7 台の SATA3.0 SSD を接続し、ストレージシステムを構築することを目標とした。

An FPGA implementation for a large-scale storage solution
Yuta KINOKUNI† and Yoshiki YAMAGUCHI‡§
Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Ten-ou-dai Tsukuba Ibaraki 305-8573, Japan†
Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Ten-ou-dai Tsukuba Ibaraki, 305-8573, Japan‡
Center for Computational Sciences, University of Tsukuba, 1-1-1 Ten-ou-dai, Tsukuba Ibaraki, 305-8577, Japan§

3. Serial ATA と FPGA

これまで提案されている FPGA 用 SATA IP コアの性能を表 2 に示す。

表２　SATA IP コアにおける先行研究

関連研究	[8]	[11]	[13]
インターフェス規格	SATA Rev.2.x	SATA Rev.3.x	SATA Rev.3.x
実効転送速度 [MB/s]	300	600	600
転送速度(Read) [MB/s]	279.0 (93.0 [%])	564 (94.0 [%])	531 (88.5 [%])
転送速度(Write) [MB/s]	242.5 (80.8 [%])	520 (86.7 [%])	505 (84.1 [%])

文献[8]では、XILINX 社製 Virtex-5 FPGA を対象とした SATA Rev 2.x コアが提案されている。しかし、非常に簡素なコアであり、Processor Local Bus (PLB) への接続や DMA などに十分に対応していたとは言い難い。文献[9]は、SATA Rev 2.x ではあるものの XILINX 社製 Virtex-6 FPGA をカバーし、PLB や DMA 機能への対応を充実させている。文献[10]は、Native Command Queuing (NCQ) は対応していないものの、SATA Rev 3.x を対象とし、XILINX 社製 FPGA だけでなく Altera 社製 FPGA にも対応している。商用ベースの IP にも目を向けると、[11][12][13]などが提案されている。ただ、いずれも大量のデバイスを考慮したものではなく、数個程度の想定であり、本報告で提案するような 100 台を超えるような利用は想定されていない。

4. 提案手法

PHY リンク下位層、リンク上位層、トランスポート層の 3 層からなる SATA IP コアをベースとし、その上位モジュールとしてコントローラを配置する。コントローラは、各 SATA コアからのスピードテストの結果に応じ、それぞれに送るデータ量を変更する。例えば、SATA SSD 2 台に HDD が 1 台つながる構成であれば、SSD 2 台により多くのデータを送ることで、ストレージ全体の性能を落とさないで済む。なお、耐故障性の観点からデータ配分方法ならびに冗長方式についても検討を行っているが、その詳細については紙面の都合により割愛する。

5. 実験結果

実装した SATA IP コアの各 SSD における実効転送速度を表 3 に示す。

表３　実装した SATA IP コアの性能

テストデータ	8GiB	8GiB	8GiB
SSD	ADATA SU800	SAMSUNG 860 EVO	GOODRAM CX300
実効転送速度 [MB/s]	600	600	600
転送速度(Write) [MB/s]	502 (83.6 [%])	506 (84.3 [%])	489 (81.5 [%])

それぞれ実効転送速度の 80%以上となっている。

これら計 7 台のストレージで、テストデータをそれぞれ 1GiB、2GiB、4GiB、8GiB としたときの結果を表 4 に示す。

表４　実験結果

テストデータサイズ	1GiB	2GiB	4GiB	8GiB
総転送速度(Write) [MB/s]	3,581	3,567	3,527	3,532

このように、転送速度の約 7 倍の値となり、コントローラの有効性が示された。

6. まとめ

本研究では、FPGA に 7 台の SATA3.0 SSD を接続し、実験を行った。接続する SSD が一台の場合、SSD の BIOS に関わらず通信・転送を行うことができる。今後は 7 台以上の SSD を接続し、コントローラの性能評価を行っていく予定である。

謝辞

本研究は JSPS 科研費 JP17H01707 および Xilinx University Program を通じ開発ソフトウェアの支援を受けておりここに謝意を表する.

参考文献

1. Adrian M. Caulfield, Eric S. Chung, Andrew Putnam, et al., “A Cloud-Scale Acceleration Architecture”, Annual IEEE/ACM Symposium on Microarchitecture, pp.1-13, Oct. 2016.
2. Eric S. Chung and Jeremy Fowers, “Accelerating Persistent Neural Networks at Detacenter Scale”, Hot Chips: A Symposium on High Performance Chips, August 2017.
3. Mark Russinovich, “Inside the Microsoft FPGA-based configurable cloud”, Build2017, May 2017.
4. Amazon Web Services, Inc., “Amazon EC2 F1 インスタンス”, (オンライン)(引用日:2018 年 7 月 21 日) https://aws.amazon.com/jp/ec2/instance-types/f1/
5. IBM, “Xilinx and IBM to Enable FPGA-Based Acceleration within SuperVessel OpenPOWER Development Cloud”, https://www.xilinx.com/ news/press/2016/xilinx-and-ibm-to-enable-fpga-based-acceleration-within-supervessel-openpower-development-cloud.html, Press Release, April 2016.
6. SATA-IO Board Members, “Serial ATA International Organization”, revision 3.3 (Gold revision), February 2016.
7. Virtex UltraScale+ (オンライン)(引用日:2018 年 7 月 21 日) https://japan.xilinx.com/products/silicon-devices/fpga/virtex-ultrascale-plus.html
8. Louis Woods and Ken Eguro, “Groundhog — A Serial ATA Host Bus Adapter (HBA) for FPGAs, IEEE International Symposium on Field-Programmable Custom Computing Machines, pp.220-223, April 2012.
9. Ashwin A. Mendon, Bin Huang, and Ron Sass, “A High Performance, Open Source SATA2 Core”, International Conference on Field Programmable Logic and Applications, pp.421-428, August 2012.
10. Patrick Lehmann, Thomas Frank, Oliver Knodel, Steffen Kohler, Thomas B. PreuBer, Rainer G. Spallek, “WEASEL: A Platform-Independent Streaming-Optimized SATA Controller”, International Conference on Field Programmable Logic and Applications, pp.1-4 (PS3-14), September 2013.
11. Design Gateway Co.,Ltd, “SATA IP Transport & Link Layer Core”, Product Specification, Rev 2.1, January 2016.
12. IntelliProp Inc., “IPC-SA101A-HI SATA Host App Core”, Datasheet, Rev 3.5, July 2013.
13. ASICSws, “Serial ATA I/II/III Host Controller IP Core”, Product Overview, Rev 3.1, June 2013.