

SMT プロセッサにおける再構成可能キャッシュモデルと FPGA による実装

小笠原 嘉泰[†] 加藤 義人[†] 大和 仁典[†]
佐藤 未来子[†] 笹田 耕一[†] 内倉 要[†]
並木 美太郎[†] 中條 拓伯[†]

近年, FPGA の高速化, 集積度の増加により, 大規模なプロセッサを実装することが可能となった. マルチスレッドプロセッサのキャッシュは, 複数のスレッドを実行するため, プログラムごとに最適な構成, スレッド間のブロック共有度などが変化する. 本稿では, キャッシュの再構成を念頭に, FPGA に SMT プロセッサを実装し, キャッシュ構成をプログラムごとに再構成するモデルについて検討した. その結果, プロセッサの SMT 化による性能向上を確認し, ハードウェア増加量はシングルスレッドプロセッサと比較して 1.3 倍程度に抑え, 動作周波数はおよそ 80MHz となった. さらに, プログラムに最適なキャッシュ構成を選定し, 再構成による性能向上を確認した.

A New Model of Reconfigurable Cache for an SMT Processor and its FPGA Implementation

YOSHIYASU OGASAWARA[†], NORITO KATO[†], MASANORI YAMATO[†], MIKIKO SATO[†],
KOICHI SASADA[†], KANAME UCHIKURA[†], MITARO NAMIKI[†] and HIRONORI NAKAJO[†]

SoC (System on a Chip), which can be utilized for various applications, can be implemented with FPGA. Therefore, many products and research testbeds which implement a processor in FPGA are emerging, and many softcore processors are getting released. Multithread architecture will be possibly adopted as a softcore processor. In this paper, we implement an SMT processor with FPGA and evaluate its performance, amount of hardware costs and its clock frequency. Moreover, from the characteristic of FPGA, since we can reconfigure cache statically, we propose a new model which reconfigures the configuration of cache according with a being executed program. As a result of implementing an SMT processor, though we have confirmed improvement of performance by SMT implementation with increased hardware of about 1.3 times compared with a conventional superscalar processor. SMT processor runs with clock frequency of about 80MHz. As a result of cache evaluation, we have found the optimal cache configurations for programs with our designed cache. And we have confirmed performance improvement with statically reconfigurable cache.

1. はじめに

近年, リコンフィギャラブルデバイスが注目されている. FPGA に代表されるリコンフィギャラブルデバイスは高速化, 集積度の増加により, 大規模なプロセッサを実装することが可能となった. そのため, 今後は, 組込みシステムを中心にプロセッサの高性能化が要求される.

プロセッサの高性能化において, プログラムが持つ命令レベル並列性 (Instruction Level Parallelism : ILP) 抽出の困難さから, スレッドレベル並列性 (Thread Level

Parallelism : TLP) の利用を目的としたアーキテクチャへの注目が増してきている. TLP の利用を目的としたアーキテクチャの一つに Simultaneous Multithreading (SMT) プロセッサ¹⁾がある. SMT プロセッサは, 複数のスレッドをプロセッサ中のリソースを共有しながら同時に実行させるという特徴を持つ. このような特徴により, 少ない追加ハードウェアで性能向上が実現できる. ところが, SMT プロセッサに対応したキャッシュは, 複数のスレッドの同時実行のため, シングルスレッドプロセッサのキャッシュに比べ, プログラムごとに最適な構成, スレッド間の共有度などが異なる. そこで, FPGA の再構成という特性から, キャッシュ構成を再構成する方式を提案する.

本稿では, まず FPGA に SEMP (Simple Efficient Multithreaded Processor) という SMT プロセッサを実装することで, 性能計算, ハードウェア量, 動作周波数の測定や従来のシングルスレッドプロセッサとの比較を行う. 次に, FPGA 実装という特性を生かし,

[†] 東京農工大学 大学院 工学教育部
Graduate School of Engineering, Tokyo University of Agriculture
and Technology
現在, ソニー株式会社
Presently with Sony Corporation
現在, 株式会社東芝
Presently with TOSHIBA Corporation

キャッシュ構成をプログラムごとに再構成するモデルについて検討する。それにより、設計したキャッシュ構成の中で実行するプログラムに最適な構成を選定し、性能向上について考慮する。

2. SEMP の構成

2.1 対象とするプロセッサ ISA

SEMP は、我々が研究・開発をすすめているオンチップマルチスレッドプロセッサである OChiMuS PE ISA に準ずる。

OChiMuS PE は、MIPS プロセッサ ISA をベースとし、マルチスレッドをサポートするため、モジュールやスレッド制御命令を新たに追加している。将来的には、本プロセッサを1つのプロセッシングエレメント (PE: Processing Element) とし、この PE を同一チップ上に複数搭載したオンチップマルチ SMT (OChiMuS) アーキテクチャとすることを検討している。詳細は参考文献 2) を参照されたい。

2.2 SEMP の設計

SEMP は OChiMuS PE ISA をベースにおき、スレッド制御命令が実現可能である。設計した SMT プロセッサ SEMP の概略図を図 1 に示す。灰色部分はパイプラインレジスタを表している。

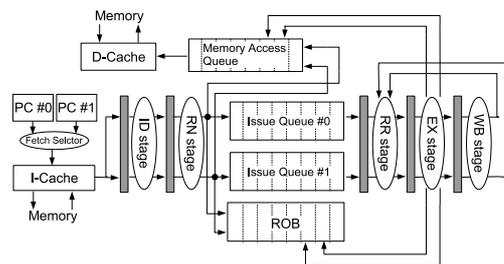


図 1 SEMP の概略図

このプロセッサの主な仕様を表 1 に示す。発行幅、リタイア幅、ハードウェアコンテキスト数は 2 のアウトオブオーダー実行を実現する SMT プロセッサである。なお、このプロセッサは Verilog-2001 により設計した。

パラメータ	説明
ハードウェアコンテキスト数	2
発行幅, リタイア幅	2
演算器	SimpleALU×2, ComplexALU×1, Load/Store Unit×1
物理レジスタ数	92
ROB	24
Issue Queue	各8
Memory Access Queue	8
分岐予測器	PHT512エントリ, GHR2bit
データ幅	32bit

3. 再構成可能キャッシュ

本章では、静的再構成可能キャッシュについて示し、キャッシュを設計する。なお、本稿は FPGA として、Xilinx 社の Spartan-3 XC3S1500⁴⁾ を使用した。

3.1 再構成の利点

キャッシュは、実行するプログラムの特性、プログラムの実行時間などにより最適な構成が異なる。ASIC のような固定構成の場合、キャッシュ構成は平均的・一般的な構成を取るが、プログラムによっては性能低下を招く場合がある。

リコンフィギャラブルデバイスを用いることで、キャッシュ構成を静的もしくは動的に変えることができる。SMT プロセッサなどのマルチスレッドプロセッサにおけるキャッシュは、実行するスレッド数、スレッド相性によってシングルスレッドプロセッサより、最適な構成が大きく変化することがある。そのため、SMT プロセッサにおいてキャッシュ構成を最適な構成に再構成すると、性能向上が期待できる。本稿では、静的再構成をプログラムを実行していない状態 (実行前、実行後) で再構成することと定義し、動的再構成をプログラム実行中に再構成することと定義する。ここでは FPGA 実装という特性を生かした静的に再構成するキャッシュについて考え、動的再構成可能キャッシュは今後の課題とする。

3.2 再構成可能キャッシュの概要

本稿で示す再構成可能キャッシュとは、動作させるプログラムによってキャッシュ構成を静的に再構成し、最適な構成にする方式である。プログラムに最適な構成を選定するために、キャッシュを実装する前に、設計した全てのキャッシュ構成の性能評価を行う。再構成する構成要素として、表 2 のようにウェイ数とブロックサイズを取り上げる。ウェイ数とブロックサイズはキャッシュの構成要素として、大きく性能に影響を及ぼすためである。ここで、再構成するキャッシュはデータキャッシュのみとし、命令キャッシュは再構成しない。

パラメータ	説明
再構成する構成要素	ウェイ, ブロックサイズ
再構成するタイミング	プログラム実行前
再構成するキャッシュメモリ	データキャッシュメモリ

また、再構成するタイミングはプログラム実行前とする。プログラムをメモリにロードすると同時に FPGA 内の構成情報を書き換え、キャッシュをそのプログラムに最適な構成に再構成する。

3.3 キャッシュの設計

SEMP に対応したキャッシュとして、表 3 のような様々な構成要素のキャッシュを Verilog-2001 により設

計した。様々なキャッシュ容量における最適な構成を選定するため、ウェイ数、ブロックサイズの他にキャッシュ容量もいくつか設計した。ここで、キャッシュはL1キャッシュのみとする。FPGAにキャッシュを実装するとき、キャッシュはアクセス速度の関係からFPGA内のBlock RAMに割り当てるのが妥当である。しかし、Block RAMは小容量であり、本稿で用いるSpartan-3デバイスではL2キャッシュまで実装する余裕が無い。その場合、L2キャッシュにFPGAチップ外部の記憶媒体を用いることも考えられるが、それではL1キャッシュとL2キャッシュ間のアクセス時間が増加してしまい、キャッシュミスペナルティが大きくなり、キャッシュの効果が期待できない。よって、キャッシュはL1キャッシュのみとする。メモリは本稿で用いるFPGAボード⁵⁾に組み込まれているDDR-SDRAMを用いる。その容量は4MBである。

表3 設計したキャッシュ構成要素

Parameter	Value
Capacity	1KB, 2KB, 4KB, 8KB, 16KB, 32KB
Way	1, 2, 4
Block Size	16B, 32B, 64B
Replacement Algorithm	LRU
Write Algorithm	Write Back
Cache Latency	1clock
Cache Miss Penalty	8clock, 12clock, 20clock

4. 性能評価

本章では設計したSMTプロセッサSEMPおよびキャッシュについて評価する。性能評価はRTL(Register Transfer Level)Verilog-HDLシミュレーションにより行った。使用したソフトウェアは、Synopsys社VCS7.1.1である。スレッドライブラリとしてOChiMuS PEスレッド制御命令に対応したMULiTh⁶⁾を使用した。また、ハードウェア量、動作周波数の評価としてXilinx社ISE6.2.03i及びCore Generator 6.2.03iを使用した。

4.1 SEMPの実行性能評価

設計したSEMPの実行性能評価を行う。実行したプログラムは以下の4つであり、明示的にマルチスレッド化されている。

- Radix ソート (Radix)⁸⁾
- LU 分解 (LU)⁸⁾
- 高速フーリエ展開 (FFT)⁸⁾
- 32 × 32 行列乗算 (Matrix)

プログラムの実行結果を図2に示す。各プログラムのIPCを表す。図2を見ると、全てのプログラムで、1スレッド動作時よりも、2スレッド動作時において

性能が向上し、最大17%、平均で10.5%ほどの性能向上がある。このことから、本プロセッサのような小規模なSMTプロセッサにおいても、SMT化による性能向上が得られることが確認できた。

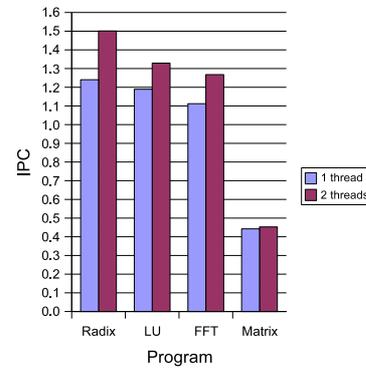


図2 SEMPの性能評価

4.2 SEMPのハードウェア量と動作周波数

表4にシングルスレッドプロセッサとSMTプロセッサの消費Slice数を示す。

表4 ハードウェア量

Slice数(Single)	Slice数(SMT)	増加率(SMT/Single)
6370	8366	1.31

SMT化により、全体で約1.3倍程度消費スライス数が増加している。POWER5³⁾では2スレッドのSMT化により、約24%程度チップ面積が増加したとされている。そのため、SEMPはPOWER5と同等かもしくはそれ以上のハードウェア量増加があることになる。

表5に論理合成後の動作周波数を示す。この結果は、理想的な配置配線ができた場合の結果である。比較対象としてMicroBlaze⁷⁾を取り上げると、Spartan-3実装時の動作周波数は80MHzである。これより、SEMPはMicroBlazeとほぼ同程度の動作周波数を達成できたことがわかる。

表5 動作周波数

動作周波数	79.662MHz
最長パス	12.506ns

4.3 再構成可能キャッシュの評価

SMTプロセッサにおける再構成するデータキャッシュの性能を調査するため、SEMPより取得したメモリアクセス命令に利用される有効アドレス、制御信号を用い、データキャッシュ単体で評価した。今回の実験で使用したプログラムは以下の5つである。

- 素数探索問題 (Prime)

Xilinx社のFPGA内部RAM
本研究は東京大学大規模集積システム設計教育研究センターを通し、シノプシス株式会社の協力で行われたものである。

線形ソート (Sort)
 Radix ソート (Radix)⁸⁾
 N クイーン問題 (Queens)
 32 × 32 行列乗算 (Matrix)

紙面の都合上、ここでは行列乗算の評価結果を示す。残りのプログラムについては選定した最適なキャッシュ構成のみを示す。

行列乗算の評価結果を図 3 に示す。グラフは、棒グラフの横軸がキャッシュ容量およびウェイト数を示し、縦軸は実行サイクル数を示す。折れ線グラフはヒット率を示す。

行列乗算はキャッシュ構成により、性能差が大きく生じることがわかる。キャッシュ容量が 1K バイトの時、ブロックサイズが性能に影響しているが、全体的にみるとウェイト数が 1 と 2 の間で大きな性能差が生じている。

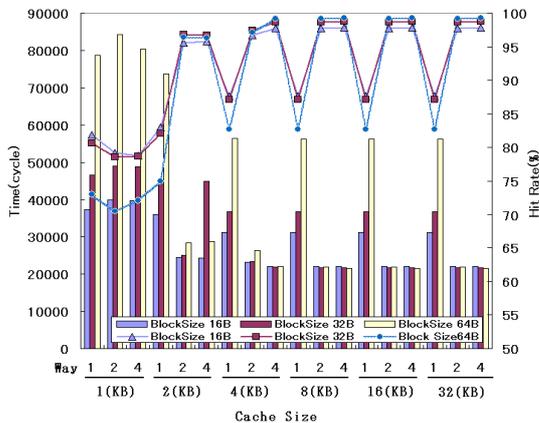


図 3 32 × 32 行列乗算の評価結果

キャッシュ容量が 16K バイトの時の各プログラムの最適なキャッシュ構成を表 6 に示す。なお、性能が同じ場合はより小さいウェイト数、ブロックサイズを最適な値とした。各プログラムの実行する時、このようなキャッシュ構成をデータキャッシュとして実装すれば、プログラムの性能は向上すると考えられる。

表 6 16KB の最適なキャッシュ構成	
プログラム	構成要素
線形探索問題	ウェイト:1, ブロックサイズ:32B
線形ソート	ウェイト:1, ブロックサイズ:32B
Radixソート	ウェイト:1, ブロックサイズ:64B
Nクイーン問題	ウェイト:1, ブロックサイズ:32B
行列乗算	ウェイト:2, ブロックサイズ:32B

5. おわりに

スレッド制御命令に対応した SMT プロセッサ SEMP

およびキャッシュについて設計した。

SEMP を設計した結果、小規模な SMT プロセッサにおいても SMT 化による性能向上が確認できた。ハードウェア量はシングルスレッドプロセッサに比べておよそ 1.3 倍増加した。動作周波数はおよそ 80MHz で、既存のソフトコアと同等の速さを実現した。

SMT に対応したキャッシュを実装するにあたって、設計したキャッシュ構成の中でプログラムに最適な構成を評価によって選定した。その結果、キャッシュを静的に再構成することによる性能向上が確認できた。性能向上率は実行するプログラムやキャッシュ容量、実行するスレッド数などによって異なる。キャッシュのハードウェア量はおよそ 1000 ~ 1500Slice になり、SEMP と共にキャッシュを Spartan-3 に実装できる。

本稿では、最適なキャッシュ構成を得るために、設計した全てのキャッシュについて評価した。最適なキャッシュ構成を得る手法は今後の課題の一つであり、コンパイラなどシステムソフトウェアとの協調を生かした方式、または、キャッシュのヒット率や共有度により、最適な構成ではなくとも、それにより近いキャッシュ構成を検知する回路の作成などが考えられる。

参考文献

- 1) D. Tullsen, S. Eggers, H. Levy : "Simultaneous Multithreading: Maximizing On-Chip Parallelism", *Proceedings of 22rd Annual International Symposium on Computer Architecture*, pp.392-403, 1995.
- 2) 河原章二, 佐藤未来, 並木美太郎, 中條拓伯 : "システムソフトウェアとの協調を目指すシングルチップマルチスレッドアーキテクチャの構想", *コンピュータシステムシンポジウム Vol.2002 No.18* pp.1-8, 2002.
- 3) R. Kalla, B. Sinharoy, J. Tendler: "POWER5: IBM's Next Generation POWER Microprocessor", *HOTChips-15*, 2003.
- 4) Spartan-3 Data Sheet (2004) : <http://www.xilinx.com/bvdocs/publications/ds099.pdf>.
- 5) Memec Spartan-3 FPGA-board : http://www.memec.co.jp/html/xilinx/eboard/docs/sp3/Memec_3SMB_UsersGuide_v2_0.pdf.
- 6) 笹田耕一, 佐藤未来子, 河原章二, 加藤義人, 大和仁典, 中條拓伯, 並木美太郎 : "マルチスレッドアーキテクチャにおけるスレッドライブラリの実装と評価", *情報処理学会論文誌*, Vol.44, No.SIG11(ACS3), pp.215-225, 2003.
- 7) MicroBlaze : http://www.xilinx.com/ipcenter/catalog/logicore/docs/microblaze_risc_32bit_proc_final.pdf.
- 8) S. C. Woo, M. Ohara, E. Torrie, J. P. Singh, A. Gupta : "The SPLASH-2 Programs: Characterization and Methodological Considerations", *ISCA-22*, pp.186-192, 1999.