

平成 29 年 6 月 21 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730026

研究課題名(和文) マルチFPGAシステムにおける任意のデータアクセス幅のキャッシュ機構の実現

研究課題名(英文) Arbitrary data-width cache memory architecture in multi-FPGA systems

研究代表者

金澤 健治 (Kanazawa, Kenji)

筑波大学・システム情報系・助教

研究者番号：40707874

交付決定額(研究期間全体)：(直接経費) 2,900,000 円

研究成果の概要(和文)：複数個のFPGAと、off-chipのDRAMから構成されるマルチFPGAシステムにおいて、DRAMアクセスレイテンシを効率よく隠蔽するための、任意のデータ幅を持ったキャッシュメモリ機構の研究を行った。同システムにおけるFPGAの個数と接続形態をパラメータとし、その性能を充足可能性問題、最大充足可能性問題、部分最大充足可能性問題、計算幾何の凸包算出、2次割当問題を用いて評価し、それぞれの問題において、どのような構成が最も有効であるかを評価した。

研究成果の概要(英文)：We studied arbitrary data-width cache memory architecture in multi-FPGA systems which consist of multiple FPGAs and off-chip DRAMs. Using the number of FPGAs and topologies of the interconnect of the FPGAs as parameters, we evaluated its performance on several applications including the satisfiability (SAT), the maximum satisfiability (MaxSAT), the partial-MaxSAT, calculation of convex hulls and the quadratic assignment problems, and then made it clear which configuration achieved the maximum performance for each application.

研究分野：Reconfigurable computing

キーワード：FPGA キャッシュ

1. 研究開始当初の背景

FPGA に代表される再構成可能な LSI は、問題に最も適した回路を構成することにより、最新のマイクロプロセッサを大きく上回る性能を実現できることが、数多くの研究により示されている。FPGA を用いた高速計算システムにおいては、優れた性能を持ち、かつ並列性の高い解法アルゴリズムを、その並列性を十分に利用しながら実行することが重要である小規模な問題であれば、データを全て FPGA 上に保持することにより、並列度を最大化することができる。一方、問題が大規模になるにしたがって、より多くのデータを外部メモリ上に置かざるを得なくなり、有効な並列度は外部メモリの実効データ転送速度により制限される。

外部メモリとしては、大容量化が容易であることから DRAM の利用が一般的である。最新の DRAM では、デュアルデータレートおよびバーストアクセスにより、非常に高いピーク性能が実現されている。しかし、ランダムアクセスが頻繁に発生する場合には、その高いデータ転送速度を利用することができず、システムの処理速度は著しく低下する。また、1 チップあたりのデータアクセス幅が固定されていることから、問題に合わせたきめ細かなデータ幅を実現することが難しい。本研究に先立ち、大規模な充足可能性問題をターゲットとし、1 チップの FPGA と数個の DRAM バンクからなるシステムにおいて、どの程度の高速化が可能であるかを検証した。また、このシステムにおいて、利用可能なハードウェアリソースを増やし、並列度を変化させた場合に、どの程度の性能向上が期待できるかを検討を行った。その結果、DRAM のデータ転送速度がピーク性能のおよそ 80% を超える場合には、概ね処理の並列度に比例した速度向上が見られた一方、並列に読み出されるデータ量が増えると、DRAM のバーストアクセスにより読み出されるデータの減少と、それに伴うランダムアクセスの増加により、速度向上が頭打ちとなることが判明した。また、問題によってはランダムアクセスされるデータの割合が高いなどの理由で、データ転送速度がピーク性能の 50% 程度まで低下し、並列度の増加が処理速度の向上に貢献しなくなる場合もあった。

以上のように、DRAM の実効データ転送速度により全体の処理速度が律速される状況下で高い処理速度を得るためには、単に処理並列度を上げるだけではなく、外部メモリの実効データ転送速度を引き上げる必要がある。

2. 研究の目的

本研究では、複数の Field Programmable Gate Array (FPGA) から構成される小規模な高速計算システムにおいて、

そのいずれかを専用のキャッシュメモリとして用い、その内部記憶領域である BlockRAM、および高速 I/O を用いた任意のデータアクセス幅のキャッシュ機構の実現と、それによる実応用問題の高速化を目指す。

FPGA の大容量化・高性能化は著しく、最新の FPGA は、専用の SRAM チップに匹敵する量の内部メモリ領域、すなわち BlockRAM を搭載しているほか、高いデータ転送速度を持った高速 I/O を備えている。BlockRAM は、データアクセス幅を自由に設定することが可能であり、その総量の範囲において、問題に応じたデータアクセス幅を実現することが可能である。これに高速 I/O を併用することにより、任意のアクセス幅のキャッシュメモリを実現することを目的とする。キャッシュメモリとして専用の SRAM を用いる方法も考えられるが、専用メモリチップが固定されたワード幅でのアクセスしかできないのに対して、提案手法では処理する問題に適したアクセス幅をとることが可能になると考えられる。

FPGA は大容量化してはいるが、その保持可能なデータ量は、DRAM と比較して大変少ないため、データ参照のパターンに応じて両者を使い分けることが重要である。ランダムアクセスが多く発生したり、DRAM の供給データ幅が過剰であり、実効データ転送速度が低下する場合には、提案手法を用いることにより、DRAM の実効データ転送速度を改善することができる。

大容量化著しい BlockRAM、および最大で数 Gbps に及ぶ内蔵の高速 I/O を併用し、数個の FPGA のうちいずれかを問題の特性に合わせたデータアクセス幅のキャッシュメモリとして使用する機構によりシステム全体の实効メモリ転送速度をどの程度向上させることが可能であり、それにより実応用問題に対してどの程度の高速化が可能であるかを明らかにする。

3. 研究の方法

まず、本研究以前の我々の研究成果に対して、提案する可変キャッシュ機構を導入した場合の性能をソフトウェアシミュレーションにより評価し、従来手法と比較してどの程度の性能向上が期待できるか確認する。

すなわち、充足可能性問題に対して、本機構により DRAM の実効データ転送速度の改善度合いと、それによりシステム全体の処理速度をどの程度高速化することができるかを評価する。その際、最新の FPGA の性能や、市販の FPGA ボードの構成等を考慮し、実現可能な構成と、性能との関係を明らかにする。充足可能性問題は、ハードウェア・ソフトウェア設計の形式的検証、自動テストパターン生成、プランニング・スケジューリング問題など、多くの応用がある。また、最大充足可能性問題など多くの拡張が存在し、それぞれ

に固有の応用があるため、その基礎となる充足可能性問題の高速化がその周辺に与える波及効果は大きい。

その後、マルチ FPGA 構成のボード、およびその開発環境を購入し、可変キャッシュ機構を備えたシステムの構築を行うとともに、本システム上に実応用問題を実装し、どの程度の高速化が可能であるかを実機評価する。実システムの構成にあたっては、カスタムな FPGA ボードを製作することも考えられるが、ボード設計の工数、提案手法の汎用性、拡張性、移植性などを考慮すると、市販のボードを用いる方が現実的である。提案手法を実現する最も簡単な構成は、2つの FPGA を高速 I/O で結合し、その一方をキャッシュとして、もう一方を演算回路として用い、外部メモリバンクは全て演算回路用の FPGA と接続する、というものである。これに類似の FPGA ボードは多数市販されていることから、提案手法を実現する最も汎用的な構成とも言える。この構成において、FPGA の回路規模、高速 I/O の性能や本数、問題規模等をパラメータとして、従来構成と比較してどの程度の速度向上が可能であるかを明らかにする。

続いて、使用する FPGA の個数を増やし、(個数が増えるとシステムが複雑化し、高い性能を得られなくなることが懸念される。市販のボードの構成等も考慮すると、4 個程度までが妥当であると思われる)、応用問題毎のデータアクセスパターン、キャッシュ機構の形態(データ幅、キャッシュとして使用する FPGA の数や利用可能な高速 I/O の本数、FPGA 間の接続形態等)をパラメータとして、どの応用問題に対してどのようなシステム構成が最も有効であるかを明らかにする。また、その結果をもとに提案手法の実装形態の類型化を行う。

更に、今後の FPGA の高集積化・高性能化や、高速 I/O の進歩により、提案手法の有効性がどの程度向上するかをソフトウェアシミュレーションにより検証し、提案手法の進展性を確認する。

4. 研究成果

平成 26 年度は、充足可能性問題および最大充足可能性問題を対象とし、その解法のひとつである WSAT アルゴリズムの改良を行い、その性能を詳細に検討した。その結果、これらの問題の応用のひとつであるハードウェア検証において特に有効であることを明らかにした。次に、本アルゴリズムを、1 個の FPGA と数バンクの DRAM からなるボードにマッピングした場合において、DRAM アクセス遅延を隠蔽するためのキャッシュの有効性をシミュレーションによって評価した。本機構は FPGA の内部リソースを用いて構成され、その連想度を DRAM からフェッチされるデータサイズに応じて変更することができ、本研究開始前の従来研究

と異なる点である。これにより、データサイズの変動に対してより頑健となり、効率よく DRAM アクセス遅延を隠蔽することができる。評価の結果、外部 DRAM のアクセス遅延を最大 60% 隠蔽することが可能であり、そのときおよそ 26% の速度向上が可能であることを確認した。

次に、複数 FPGA からなる可変キャッシュ機構の有効性を検証するため、FPGA の個数を 2 個に増やした場合について検討を行った。その結果、FPGA 間を接続する I/O のスループットは十分であるが、レイテンシがやや大きく、DRAM のアクセス遅延を隠蔽するには不十分であった。

これらの研究結果のうち、学術的に有効と思われるものに関して、国際会議において発表を行うとともに、論文誌への投稿準備を進めた。

平成 27 年度は、2 個の FPGA からなるシステムにおいて、可変キャッシュ機構の有効性を評価するとともに、FPGA の個数を更に増やした場合のシステム構成の検討を行った。

FPGA 間の接続に使用する高速 IO には、スループットとレイテンシにトレードオフの関係があり、両立が困難である。キャッシュとして必要とされるデータ供給能力は IO の速度以外にも主記憶の構成、特に主記憶のデータ幅に依存する。本システムの目的はキャッシュであるため、レイテンシを小さくすることをより重視した。この場合、データ供給能力を上げるためには IO の必要数が増加し、FPGA の個数を増やすことにより、FPGA 1 個あたりの IO の本数が限定されることから、キャッシュとしてのデータ供給能力は低下する。一方、FPGA の個数が増えることにより、キャッシュ容量増加によるミスヒット率の低減が期待される。また、構成の汎用性を考慮し、FPGA ベンダから供給される代表的なボード上の主記憶の構成を想定した。アプリケーションには、(1)充足可能性問題、(2)部分最大充足可能性問題、(3)計算幾何学の凸包導出処理を用いた。(1)(2)は比較的ランダムアクセス頻度が高いとみられたもの、(3)は(1)(2)との対比のための、比較的ランダムアクセスの少ないとみられた問題である。しかしながら、今回の評価ではいずれの問題においてもデータアクセスパターンのランダム性が比較的小さく、キャッシュミスヒットの影響がさほど大きくなかったことから、システムを構成する FPGA の個数は 2 個が最適である結果となった。

以上の研究を進める過程で得られた成果のうち、学術的に有効なものに関して、国際会議での発表を行った。

平成 28 年度は、2 個の FPGA からなるシステム(構成 A)において、可変キャッシュ機構の実機評価(アプリケーションとして部分最大充足可能性問題を使用)を行うとともに、

FPGA の個数を更に増やした場合(構成 B)の同機構の有効性の評価を、昨年度に引き続き実施した(アプリケーションとして充足可能性問題、最大充足可能性問題、部分最大充足可能性問題、2 次割当て問題を使用)。

構成 A については、現在、部分最大充足可能性問題の例題を用いてシミュレーション、および実機による評価を行っており、公開されている例題のうち中規模の問題において、シミュレーション評価で算出されたものとほぼ同程度の性能が得られる見通しがついている。今後、より多くの例題を用いて、所期の性能が得られることを確認する予定である。なお、シミュレーション評価に関しては、その一部を国際会議で発表した。

構成 B に関しては、昨年度に引き続き、アプリケーションを追加・変更して評価を行ったが、IO の本数と、データ転送速度の不足を補うようなシステム構成を発見することはできず、昨年度と同様、FPGA の個数が 2 個のとき、最も性能が出る結果となった。

これらの過程で得られた研究成果のうち、学術的に有効とみられるものについては、部分最大充足可能性問題の実機評価の結果に関して、論文誌への投稿を行った。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

(1)Kenji Kanazawa and Tsutomu Maruyama, ``An Approach for Solving SAT-encoded Formal Verification Problems on FPGA'', The IEICE Transactions on information and systems Vol.E100-DNo. 8, 2017 (accepted) (査読あり)

〔学会発表〕(計 3 件)

(全て査読あり)

(1)Shohei Sassa, Kenji Kanazawa, Shaowei Cai and Moritoshi Yasunaga, ``An FPGA Solver for Partial MaxSAT Problems Based on Stochastic Local Search'', International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART2016), HongKong. 2016

(2)Kenji Kanazawa, Kahori Kemmotsu, Yamato Mori, Noriyuki Aibe and Moritoshi Yasunaga, ``High-Speed Calculation of Convex Hull in 2D Images Using FPGA'', Parallel Computing with FPGAs (ParaFPGA2015), pp. 533-542, Edinburgh, Scotland, UK, 2015

(3)Kenji Kanazawa and Tsutomu Maruyama,

``FPGA Acceleration of SAT/MaxSAT Solving using Variable-way Cache'', International Conference on Field Programmable Logic and Applications (FPL2014), (4pages), Munich, Germany, 2014

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

<http://www.cs.tsukuba.ac.jp/~kanazawa/>

6. 研究組織

(1)研究代表者

金澤健治 (KANAZAWA, Kenji)

筑波大学システム情報系 助教

研究者番号: 40707874

(2)研究分担者

なし。

(3)連携研究者

なし。

(4)研究協力者

なし。