様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２７　年　　５　月　２６　日現在

機関番号：　１２６０８
研究種目：　若手研究(B)
研究期間：　2011～2014
課題番号：　２３７０００５２

研究課題名（和文）データ圧縮技術の適用による超並列プロセッサの低消費電力化と高性能化

研究課題名（英文）Improvement of power efficiency and performance of super parallel processors by application of data compression technology

研究代表者

金子　晴彦（Kaneko, Haruhiko）

東京工業大学・情報理工学（系）研究科・准教授

研究者番号：７０３９２８６８

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：超並列プロセッサのメモリアクセス性能向上を目的として，低遅延かつ高スループット処理が可能な圧縮アルゴリズムである，Periodic pattern coding（周期的パターン符号化）を提案した．GPUシミュレータを用いて圧縮率の評価を行い，従来の類似手法と比較して優れた圧縮率を有することを明らかにした．また，キャッシュミス率，IPC（1サイクルあたりの実行命令数），等を評価し，提案手法がプロセッサの性能向上に有効であることを示した．ハードウェア記述言語により圧縮・伸張回路を設計し，高いスループットで圧縮・伸張処理ができることを示した．

研究成果の概要（英文）：We proposed a low-latency and high-throughput lossless compression algorithm, named periodic pattern coding, to improve the memory access performance of super parallel processors. Compression ratio of the proposed method is evaluated by a GPU simulator, and result showed that the proposed method has higher compression ratio compared to conventional compression methods. Also, evaluations of the cache miss ratio and instructions per cycle (IPC) demonstrated that the proposed method is effective to improve the processor performance. Compression/decompression circuits are designed using hardware description language, and results showed that the circuit provides high processing throughputs.

研究分野：　データ圧縮

キーワード：　データ圧縮　キャッシュメモリ　並列プロセッサ　キャッシュミス率　GPU

１．研究開始当初の背景

大規模演算を高速に並列実行する手段として，GPU(graphics processing unit)などの超並列プロセッサを採用するケースが増加している．GPU は 1 チップに数百個以上のコアを搭載していることから，汎用プロセッサと比較して格段に高い演算能力を有し，CPU と GPU の演算性能(FLOPS)は 1 桁程度の差が出る場合がある．GPU は比較的安価な PC での使用が可能であり，また，一方で多数の GPU を用いた世界最速クラスのスーパーコンピュータの構築も行われており，今後も利用分野の拡大が予想される．

超並列プロセッサを搭載した計算機は高い演算能力を有する一方，消費電力が多く，例えば GPU を複数個搭載したサーバシステムの消費電力は 1000W を越えることがある．計算機システムで消費する電力のうち，メモリシステムが消費する電力の割合は近年増加しており，例えば，文献[1]によると，IBM POWER7 を搭載したサーバにおいて，システム全体の消費電力に対するプロセッサの消費電力の割合は 41%であるのに対し， メモリシステムの消費電力の割合は 46%を占める．このことから，メモリシステムの低消費電力化は環境問題の観点から重要な課題である．

CPU の性能向上を阻害する要因として，メモリウォールの問題が指摘されているが，超並列プロセッサにおいてはこの問題がさらに顕著となる恐れがある．すなわち，超並列プロセッサでは演算処理能力が飛躍的に向上しているのに対し，メモリからのデータ供給能力の向上は十分であるとは言えない．GPU ではスレッド並列化によりメモリアクセスの遅延を隠蔽しているが，スループットの不足を補うことは難しい．メモリアクセスのスループットを向上するために，オンチップキャッシュメモリを増設する手法も考えられるが，プロセッサダイの総面積に対するキャッシュメモリ面積の割合はすでに 5 割近くに達しているものもあり，さらなるキャッシュメモリの大容量化は容易ではない．

超並列プロセッサに関する上記の問題を解決する手段の一つとして，メモリシステム／データバスへのデータ圧縮技術の適用が考えられる．メモリシステムに対するデータ圧縮法に関する従来の関連研究の例として，以下が挙げられる．組み込みシステムのメモリ空間拡大と，メモリの消費電力削減を目的として，コンパイラを用いたデータ圧縮法[2]や, Huffman 符号を用いたリアルタイム圧縮/伸長法[3]が提案されている．また，サーバ等のメインメモリに対するデータ圧縮法として IBM-MXT[4]などが，キャッシュメモリに対するデータ圧縮法として X-Match[5]や C-Pack[6]などが提案されている．また，研究代表者らはこれまでにブロックソートを用いたデータ圧縮法をメインメモリに適用する手法を検討している[7,8]．しかし，従来の研究で想定されているコア数は一般に 10 個以下程度であり，GPU のような超並列プロセッサに対してデータ圧縮技術を適用した例は報告されていない．

２．研究の目的

本研究では，GPU などの超並列プロセッサの低消費電力化と高性能化を目的として，メモリシステム及びデータバスに対する効率的なデータ圧縮法を構築する．また，シミュレーションによりキャッシュミス率等の評価を行い，実用性の検証を行う．

３．研究の方法

超並列プロセッサを主な応用対象として以下の研究を行った．

(1) GPU シミュレータによるメモリシステム及びデータバス上のデータの収集と解析．

(2) データ解析に基づく低遅延・高スループットな並列データ圧縮／伸長アルゴリズムの構築．

(3) 上記の圧縮／伸張アルゴリズムを用いたメモリ制御回路の消費電力，回路量，及びプロセッサの演算性能，等のシミュレーションによる評価．

本研究は，情報理論，計算機設計論，等を理論的な基盤とし，併せて計算機シミュレーションと HDL による論理設計を行うことにより実施した．

４．研究成果

本研究では主に，(1)GPU シミュレータによるキャッシュメモリ上のデータ解析，(2)高速な可逆圧縮アルゴリズムである Periodic Pattern Coding の提案，(3)GPU シミュレータを用いた，圧縮率，キャッシュミス率，IPC の評価, HDL を用いた回路量, スループット，消費電力，等の評価，を行った．また，(4)科学技術シミュレーションで扱われることの多い，浮動小数点数の配列データに対する圧縮法の構築を行った．それぞれの概要を以下に示す．

(1) GPU シミュレータを用いたデータ解析:

GPU シミュレータである GPGPU-sim [9]を用いて GPU のキャッシュメモリ上のデータを解析した．例えば，以下のようなパターンのデータが多数存在することを確認した．

- 4 バイトまたは 8 バイト周期の系列
- 同一なバイト値の反復
- 値域の狭い浮動小数点数

キャッシュメモリ上のデータパターンの例を図１に示す．

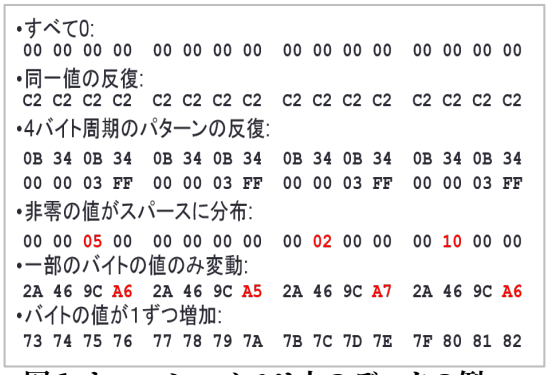

図 1.キャッシュメモリ上のデータの例

(2) Periodic Pattern Coding (PPC)

データの周期性を利用した高スループットな圧縮アルゴリズムとして Zero-base coding(ZBC)及び PPC を提案した．ZBC はデータをバイトごとに見たときに，非零の要素が少ないスパースなデータに適した圧縮法であり，非零の要素の位置と値を符号語とすることによりデータを圧縮するアルゴリズムである（図 2）．

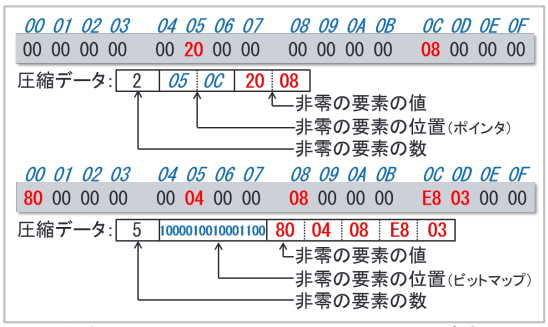

図 2．Zero-base coding(ZBC)の例

PPC は 4 バイトまたは 8 バイト周期で類似のパターンが反復して出現するデータに適した圧縮法である．符号語を，ベースパターン（反復して表れるパータン）と，ベースパターンと異なるバイトの位置及びその値の組合せとして表現することによりデータ列を圧縮する（図 3）．

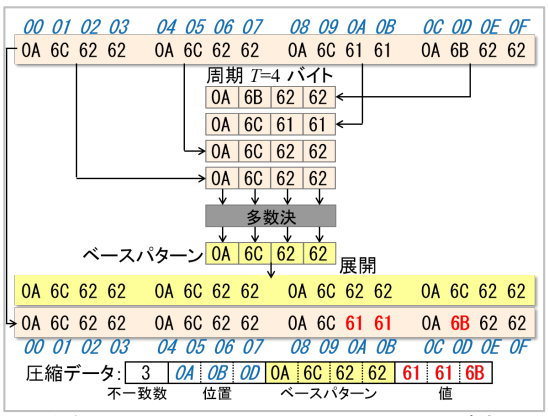

図 3．Periodic-pattern coding の例

(3) PPC の評価

GPU シミュレータ(GPGPU-sim)を用いて，①圧縮率及び②キャッシュミス率の評価を行った．また，HDL により圧縮／伸張回路を設計し，③回路量及び動作周波数の評価を行った．これらの評価により提案手法がキャッシュメモリに有効であることを明らかにした．

①圧縮率： 圧縮率の評価の結果，多くのベンチマークプログラムに対し，従来手法（C-Pack,X-MatchPro,LZSS,BΔI,FPC）よりも優れた圧縮率を有することが明らかになった（図 4）．

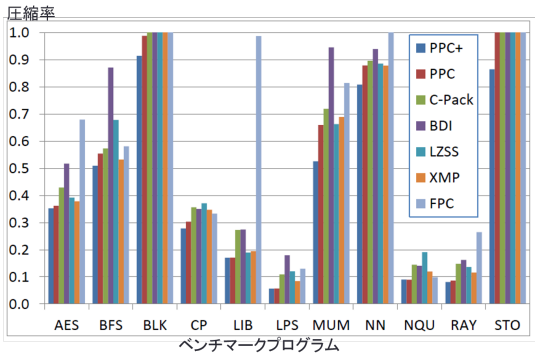

図 4．圧縮率の評価

②キャッシュミス率：GPU シミュレータを用いて，提案手法，従来手法(BΔI)及び圧縮を適用しない場合のキャッシュミス率を評価した．この結果，一部のベンチマークプログラムにおいて，キャッシュミス率を大きく低減できることが明らかになった（図 5）．

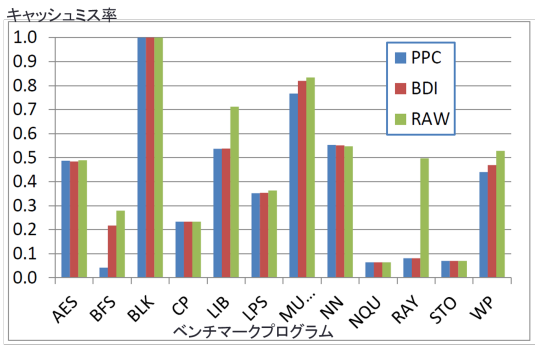

図 5．キャッシュミス率の評価

③回路量及び動作周波数の評価： 提案手法の復号回路を構成し，既存手法(C-Pack)と回路と動作周波数の比較を行った．この結果を表 1 に示す．

表 1．回路量と動作周波数の評価

	PPC	C-Pack
入出力幅(byte)	32	16
ASIC モデル(nm)	45	65
伸張回路サイズ(mm²)	0.060	0.043
動作周波数(GHz)	1.37	1.20
スループット(GB/s)	43.84	19.20
消費電力(mW)	48.79	24.14

(4) 浮動小数点数の配列データに対する圧縮法

スーパーコンピュータ等における大規模科学技術シミュレーションでは浮動小数点数の配列データを扱う場合が多いことから，このようなデータに特化した圧縮アルゴリズムを構築した．

本手法では，符号化済のデータ系列から符号化対象のデータの値を予測し，予測値とデータの値の誤差を符号化することにより，圧縮を行う（図 6）．

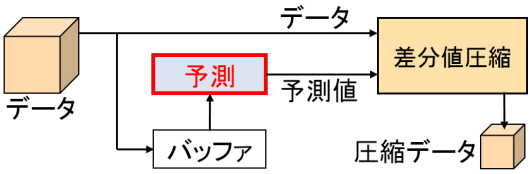

図 6．浮動小数点数配列データの圧縮

差分値の圧縮においては，浮動小数点数の構造（符号部，指数部，仮数部）を考慮し，予測値とデータの誤差が小さい場合，仮数部を効率的に圧縮できるようにしている（図 7）．

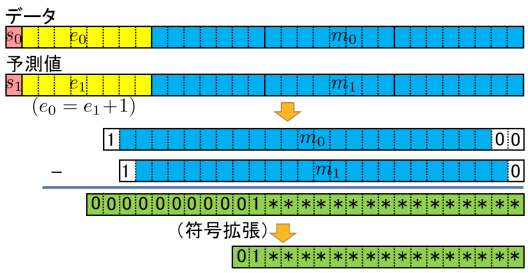

図 7．浮動小数点数の圧縮

提案手法を気象データ（気温，風速）及び流体力学計算データに適用した場合の圧縮率の評価を図 8 に示す．

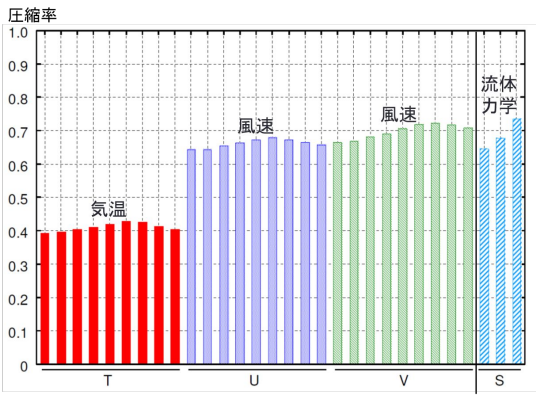

図 8．浮動小数点配列データの圧縮率

（参考文献）

[1] J. Carter and K. Rajamani, "Designing Energy-Efficient Servers and Data Centers," *Computer*, pp. 76-78, 2010.

[2] O. Ozturk, et al., "Compiler-guided data compression for reducing memory consumption of embedded applications," *Proc. Asia and South Pacific Conf. Design Automation*, pp. 814-819, 2006.

[3] A. Wolfe and A. Chanin, "Executing compressed programs on an embedded RISC architecture," *Proc. 25th Annual Int. Symp. Microarchitecture*, pp. 81-92, 1992.

[4] R. B. Tremaine, et al., "IBM Memory Expansion Technology," *IBM J. Research and Development*, vol. 45, no. 2, pp. 271-285, 2001.

[5] E. Ahn, et al., "Effective algorithms for cache-level compression," *Proc. 11th Great Lakes Symp. on VLSI*, pp. 89-92, 2001.

[6] X. Chen. et al., "Design and implementation of a high-performance microprocessor cache compression algorithm," *Proc. 2008 Data Compression Conf.*, pp.43--52, 2008.

[7] 東，金子，"主記憶上のデータに対するブロックソートを用いた圧縮法，" *第 8 回情報科学技術フォーラム講演論文集*, C-015, 2009.

[8] 新妻，金子，"キャッシュメモリから主記憶へのデータ転送に対する圧縮法の評価，" *第 33 回情報理論とその応用シンポジウム*, pp. 214-219, 2010.

[9] http://www.gpgpu-sim.org/

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 H. Kaneko, "Periodic Pattern Coding for Last Level Cache Data Compression," *IEICE Trans. Fundamentals*, 査読有, Vol. E96-A, No. 12, pp.2351-2359, Dec. 2013. DOI: 10.1587/transfun.E96.A.2351

〔学会発表〕（計　4　件）

① 金子晴彦，"データ圧縮技術のイクサフロップスマシンへの適用性，" *第 6 回アクセラレーション技術発表討論会*, 2014 年 6 月 20 日，沖縄県・国頭郡恩納村．

② H. Kaneko, S. Fujii, H. Sasaki, "Differential Base Pattern Coding for Cache Line Data Compression," *Proc. Data Compression Conference*, p. 499, 2013 年 3 月 21 日，ソルトレークシティ（米国）．

③ K. Ota, S. Fujii, H. Sasaki, H. Kaneko, "Differential Base Pattern Coding for Cache Line Data Compression," *Proc. 35th Symp. Information Theory and its applications*, pp. 514-519, 2012 年 12 月 12 日，大分県・別府市．

④ 藤井理史，金子晴彦，"GPGPU デバイス上のキャッシュメモリに対するデータ圧縮法，" *電子情報通信学会技術研究報告*, FIIS-12-321, 2012 年 3 月 9 日，石川県・野々市市．

６．研究組織

(1)研究代表者

金子 晴彦（KANEKO, Haruhiko）
東京工業大学・大学院情報理工学研究科・
准教授
研究者番号：７０３９２８６８