様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ５ 月 ３０ 日現在

機関番号：１２６０８
研究種目：基盤研究(C)（一般）
研究期間：2015～2018
課題番号：１５Ｋ００７１

研究課題名（和文）データインテンシブコンピューティングのためのデータバス用高スループット符号化法

研究課題名（英文）High-throughput data coding for data-bus in data intensive computing

研究代表者
金子　晴彦（Kaneko, Haruhiko）

東京工業大学・情報理工学院・准教授

研究者番号：７０３９２８６８

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：大規模計算機シミュレーションやビッグデータ解析などのデータインテンシブコンピューティングでは，大量のデータを高速にプロセッサに供給する必要があることから，メモリバスやノード間ネットワーク等のデータバスのスループット向上が求められる．本研究ではデータバスの実効スループットの向上を目的として，データバスに適した符号化技術（可逆データ圧縮法と同期誤り制御符号化法）を構築した．シミュレーションによる評価を行い，提案手法は従来手法と比較して高い圧縮率と，誤り制御能力を有することを示した．

研究成果の学術的意義や社会的意義
社会的意義：本研究成果は，大規模科学技術シミュレーションやビッグデータ解析等の処理能力向上に寄与できると考えられ，今後は，スーパーコンピュータ等のHPC分野への適用に向けてより詳細な検討を行うことが重要であると考えられる．また，提案した符号化技術を演算アクセラレータ，GPU，等のIOバスへ適用することにより，小規模な計算機システムの処理能力の向上も期待できる．
学術的意義：データバス向けの可逆データ圧縮アルゴリズムと同期誤り制御符号を新たに構築したことから，符号理論，情報理論の観点からも一定の意義を有する成果が得られたと考えられる．

研究成果の概要（英文）：Data intensive computing, such as large-scale computer simulation and big-data analysis, often requires high throughput transfer of large amount of data into processors. Therefore, improvement of throughput in data bus, e.g., memory bus, node interconnection bus, will be crucial. To improve the effective throughput of data buses, this research proposes new coding methods (lossless data compression and synchronization error control coding) suitable for high throughput data buses. Simulation results show that the proposed codings have higher compression ratio and error control capability, compared to existing methods.

研究分野：符号理論

キーワード：可逆データ圧縮　辞書式符号化　同期誤り訂正

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

大規模な科学技術シミュレーションの実行やビッグデータ解析等，大容量のデータを高速に処理するデータインテンシブコンピューティングの重要性が高まっている．データインテンシブコンピューティングでは，大量のデータを高スループットでプロセッサに供給することが求められる．例えば，次世代スーパーコンピュータの性能検討において，要求メモリバンド幅が 50PB/sec 以上である科学技術系のアプリケーションが複数存在することが示されている[1]．ビッグデータ解析の分散処理においても，処理ノード間で大量のデータを高速に転送する必要がある．

一方，現在の計算機システムの構成は，データインテンシブコンピューティングにおける要求と必ずしも適合していない．すなわち，プロセッサの演算能力は，マルチコア，メニーコアアーキテクチャの採用により向上し続けており，GPU では多数のコアによる並列処理により高い処理能力を達成している．また，ストレージの記憶容量は，熱アシスト磁気記憶やビットパターンドメディア(BPM)などの新技術により今後も容量の増大が期待される．このように，プロセッサ処理能力やストレージ記憶容量が向上し続けているのに対し，プロセッサとメモリ・ストレージとを結ぶデータバスのスループットの向上は不十分であると考えられ，特にデータインテンシブコンピューティングでは，データバスが処理のボトルネックとなることが懸念される．

データバスの高スループット化のためには，回路技術，信号処理技術，等に加えて適切な符号化技術（データ圧縮，誤り制御符号）の適用が重要である．すなわち，データ圧縮によりデータサイズを削減して実効的なスループットを向上でき，誤り制御符号によりノイズ，信号反射，クロストーク，等に対する耐性を高めてバスの高クロック化を図ることができる．しかし，従来の符号化技術は処理遅延や計算量の観点から，一般にデータバスへの応用には適していない．

[1] 文部科学省 HPCI 計画推進委員会 今後の HPCI 計画推進のあり方に関する検討 WG，システム検討サブ WG 報告書， 2013.

２．研究の目的

本研究ではデータバス高スループット化のための符号化技術として，データ圧縮法と誤り制御符号の構築を行う．データ圧縮法に関して，まず，圧縮対象となるデータの特性を解析する．大規模シミュレーションを考慮し，数値流体力学(CFD)データ，粒子計算データ，格子量子色力学(QCD)データ，等も用いる．圧縮アルゴリズムとして，圧縮率のみでなく，ハードウェア実装（並列化，パイプライン化）を考慮したアルゴリズムを構築する．

誤り制御符号に関して，シミュレーション等により高速データバスの通信路モデルを構築し，これをもとに，誤り制御符号の設計を行う．また，データ圧縮と誤り制御符号の符号化・復号回路を設計し，ハードウェア記述言語によりシミュレーションと評価を行う．本研究により，例えばデータバスのスループットを 1.5 倍から 4 倍程度向上することを目標とする．

３．研究の方法

本研究では，データバス高スループット化のためのデータ圧縮アルゴリズム構築と誤り制御符号(ECC)の設計，及び高速な符号化・復号回路の設計と実装を行う．データ圧縮アルゴリズム構築では，CFD 等のシミュレーションデータやセンサデータを収集し，これらの特性を考慮した圧縮アルゴリズムを構築する．アルゴリズムはハードウェア処理（並列化，パイプライン化）に適したものとする．また，ECC 設計のため，将来の高速データバスにおける誤り特性を推定し，これを基に通信路モデルを構築し，この通信路モデルに適合した ECC の設計を行う．許容される復号計算量に応じて，確率伝播アルゴリズムや限界距離復号等の各種復号法を想定した線形符号を検討する．

４．研究成果

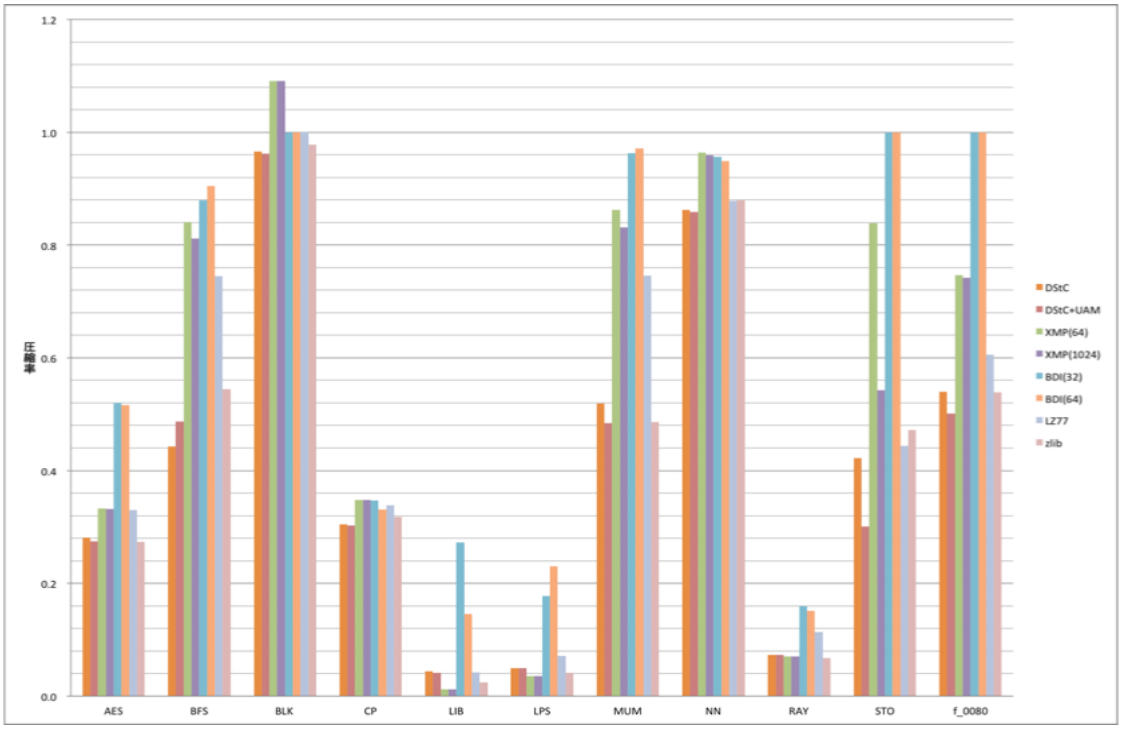
本研究ではデータバスの実効スループットの向上を目的として，データバスに適した符号化技術（データ圧縮法と誤り制御符号化法）を構築した．具体的には，数値データやその配列データ，記号列データ，等を高効率で圧縮可能なデータ圧縮法，データバスの高クロック化に有効な誤り制御符号の設計，等を行なった．高スループットな符号化・復号をハードウェア処理により実現するために，符号化・復号回路の設計ハードウェア記述言語を用いて行なった．主な成果は以下のとおりである．

(1) 以下の 2 つの可逆圧縮法を提案した．

i. 符号化処理を 2 つの処理に分割し，双方を並列に処理することによって遅延を削減した圧縮アルゴリズム DStC.

ii. テキストデータなどのようにデータのパターンが 4 バイトの倍数でアラインされていない系列に対して有効となるような圧縮アルゴリズムである UAM 符号化.

また，これら 2 つを組み合わせ，テキストデータなどを含めた多く系列に対して有効に作用する低遅延可逆圧縮法 DStC+UAM の構成を示した．シミュレーションにより，従来の手法と比べて多くのデータ系列に対して高い圧縮率が得られることを示した．圧縮率の例を下図に示す．ただし，圧縮率は「圧縮後データサイズ／圧縮前データサイズ」であり，値が小さい方が優れている．



(2) 上記(1)の結果を拡張し，データバスの帯域に合わせて圧縮法を動的に切り替えることにより，スループットを向上する手法を提案した．

(3) 高速データバスで懸念される同期誤りやシンボル間干渉を考慮した誤り制御符号について，以下の成果を得た．

i. 高速データ伝送等においては，ノイズや信号干渉などの影響により生じるランダム誤りに加えて，同期のずれによる挿入／削除誤りが生じる可能性があることから，これらの誤りを訂正する新たな挿入／削除／反転(IDS) 誤り制御符号を設計した．従来の IDS 誤り制御符号の多くは，挿入／削除誤りが統計的独立に生じることを仮定しているが，実際のデータバス等の通信路では，同期のずれが蓄積することにより挿入／削除誤りが発生すると考えるほうが自然である．そこで本研究では，このような誤りの発生を t/m ビット挿入／削除誤りやタイミングドリフトにより表現した同期誤り通信路モデルを新たに定義した．また，この通信路に対し従来のマーカーを用いた IDS 誤り制御符号を適用し，シミュレーションにより誤り率の評価を行った．この結果，従来の符号をそのまま適用するだけでは誤り率を十分に低減できないが，受信／送信クロック周期に対し微小な差異を与えることにより，誤り率を低減できる場合があることを明らかにした．

ii. 高速データバス等では同期誤りのみでなく，シンボル間干渉(ISI: inter-symbol interference) も生じることがある．本研究では，ISI と同期誤りの一種であるタイミングドリフト(TD) が同時に生じる通信路モデルとして，TD-ISI 通信路を定義した．また，送信ビットの事前確率を不均一とした場合の検出アルゴリズムとして，各送信ビットの尤度の近似値を求める計算法を，ファクターグラフを用いて示した．さらに，送信語と検出器出力の間の相互情報量の近似値をモンテカルロシミュレーションにより求め，事前確率分布と相互情報量の関係を明らかにした．

上記に示すとおり，本研究は特に大規模科学技術シミュレーションやビッグデータ解析等の処理能力向上に一定の貢献ができることが予想され，今後は，スーパーコンピュータ等の HPC 分野への適用に向けてより詳細な検討を行うことが重要であると考えられる．また，提案した符号化技術を演算アクセラレータ，GPU，等の IO バスへ適用することにより，小規模な計算機システムの処理能力の向上も期待できる．一方で，データバス向けのデータ圧縮アルゴリズムと誤り制御符号を新たに構築したことから，符号理論，情報理論の観点からも一定の意義を有する成果が得られたと考えられる．

５．主な発表論文等

〔雑誌論文〕（計　0　件）

〔学会発表〕（計　14　件）

（国際会議発表）
1. Haruhiko Kaneko. Slepian-Wolf Coding with Side-Information Having Insertion/Deletion Errors, Proc. 2018 International Symposium on Information Theory and Its Applications, pp. 607-611, Oct. 2018.
2. Yusei Suzuki, Haruhiko Kaneko. Correlated Insertion/Deletion Error Correction Coding for Bit-Patterned Media, Proc. 2017 IEEE International Conference on Consumer Electronics - Taiwan, pp. 7-8, Jun. 2017.
3. Haruhiko Kaneko. Timing-Drift Channel Model and Marker-Based Error Correction Coding, Proc. 2017 IEEE Int. Symp. Information Theory, pp. 1943-1947, Jun. 2017.
4. Haruhiko Kaneko, Yuki Katsu. Low-Latency Lossless Compression Using Dual-Stream Coding, Proc. 2016 Int. Symp. Information Theory and Its Applications, pp. 677-681, Oct. 2016.
5. Yuki Katsu, Haruhiko Kaneko. Low-Latency Lossless Compression Codec Design for High-Throughput Data-Buses, Proc. 2016 IEEE Int. Conf. Consumer Electronics-Taiwan, pp. 274-275, May. 2016.
6. Yuki Katsu, Haruhiko Kaneko. Low-Latency Lossless Compression for Data Bus Using Multiple-Type Dictionaries, Proc. 2016 Data Compression Conference, p. 610, Mar. 2016.

（国内学会・研究会発表）
1. 金子 晴彦. マーカーを用いた挿入／削除／反転誤り訂正符号化法, 2017 年電子情報通信学会ソサイエティ大会講演論文集, Sep. 2017.
2. 平野 武, 金子 晴彦. 不完全ビットシフトが生じる通信路に対するモジュレーション符号化法の検討, 電子情報通信学会技術研究報告, DC2017-16, Jul. 2017.
3. 勝 悠貴, 金子 晴彦. 複数のハッシュテーブルを用いた可逆圧縮アルゴリズムの並列化, 第 39 回情報理論とその応用シンポジウム予稿集, pp. 517-521, Dec. 2016.
4. 勝 悠貴, 金子 晴彦. テキストデータに対する並列圧縮アルゴリズムの圧縮・伸張回路構成の検討, 電子情報通信学会技術研究報告, Oct. 2016.
5. 金子 晴彦. t/m ビット同期誤り通信路モデルと誤り制御符号化法の検討, 電子情報通信学会技術研究報告, Aug. 2016.
6. 勝悠貴, 金子晴彦. コンテクストベース可変長符号化を用いた低遅延データ圧縮法の検討, 電子情報通信学会技術研究報告, Mar. 2016.
7. 鈴木悠生, 金子晴彦. 依存関係のある挿入／削除／反転誤り通信路モデルと検出アルゴリズムの検討, 電子情報通信学会技術研究報告, Mar. 2016.
8. 勝 悠貴, 金子 晴彦. 部分一致列探索が可能なハッシュテーブルを用いた辞書式圧縮法, 第 38 回情報理論とその応用シンポジウム予稿集, pp. 433-438, Nov. 2015.

〔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

６．研究組織

(1)研究分担者：　なし
(2)研究協力者：　なし