

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2010-521728
(P2010-521728A)

(43) 公表日 平成22年6月24日(2010.6.24)

(51) Int.Cl.

G06F 12/00 (2006.01)

F 1

G06F 12/00

560G

テーマコード(参考)

5B060

審査請求 未請求 予備審査請求 未請求 (全 23 頁)

(21) 出願番号 特願2009-553163 (P2009-553163)
 (86) (22) 出願日 平成20年3月14日 (2008.3.14)
 (85) 翻訳文提出日 平成21年11月2日 (2009.11.2)
 (86) 國際出願番号 PCT/EP2008/053133
 (87) 國際公開番号 WO2008/110633
 (87) 國際公開日 平成20年9月18日 (2008.9.18)
 (31) 優先権主張番号 0704976.0
 (32) 優先日 平成19年3月15日 (2007.3.15)
 (33) 優先権主張国 英国(GB)
 (31) 優先権主張番号 60/911,273
 (32) 優先日 平成19年4月11日 (2007.4.11)
 (33) 優先権主張国 米国(US)

(71) 出願人 509258418
 リニア アルジェブラ テクノロジーズ
 リミテッド
 アイルランド国 ダブリン マウントジョイ
 イスクエア イースト 19
 (74) 代理人 100116872
 弁理士 藤田 和子
 (72) 発明者 モロニー ディヴィッド
 アイルランド国 ダブリン グラスネヴィン
 アイオナ ロード 42 リニア アルジェブラ テクノロジーズ リミテッド
 F ターム(参考) 5B060 CB03

最終頁に続く

(54) 【発明の名称】データ圧縮のための回路及びこれを用いるプロセッサ

(57) 【要約】

【課題】本出願は、コンピュータ・システム設計における基本課題であるメモリ・アクセス・コストの最小化に對処する。

【解決手段】これは、メモリ技術又はプロセッサへの接続様式に関わらず、コンピュータ・システムの設計における基本的制約であり、所与の時間に転送可能なプロセッサとメモリとの間データについては、大きな制約、すなわち利用可能なメモリ帯域があり、利用可能なメモリ帯域によるコンピュータ計算力の制限は、しばしばメモリの壁と呼ばれている。本提供の解決においては、圧縮されるデータ構造体マップを生成し、該マップは該構造体における些少ではないデータ値の位置を表し(例えば、ゼロでない値)、圧縮構造体を提供するために該構造体から些少なデータ値を削除する。

【選択図】図1

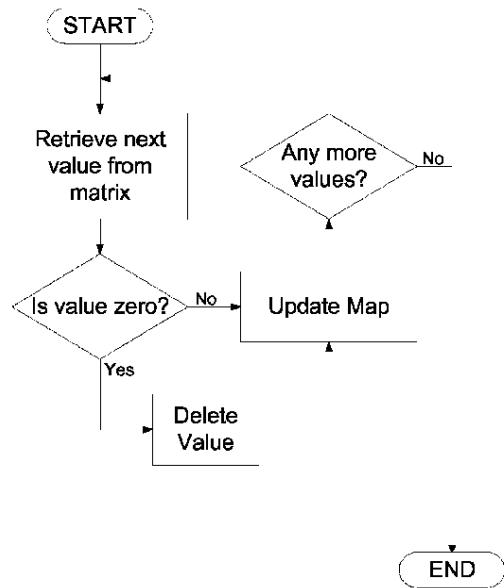


Figure 1

【特許請求の範囲】**【請求項 1】**

圧縮回路であって、

- a) 個々のデータ値の構造体を格納するためのデータメモリと、
- b) マップを格納するためのマップ・メモリであって、前記マップは前記構造体内においてゼロでない値の位置を表すマップ・メモリと、
- c) データ出力であって、前記回路は、前記データメモリからゼロでないデータを取り出し、前記取り出されたデータを前記データ出力において前記マップを表すデータを組み合わせた圧縮構造体として提供するよう構成されるデータ出力と、を含む圧縮回路。

【請求項 2】

前記データメモリは複数のレジスタを含む、請求項 1 に記載の回路。

【請求項 3】

前記データメモリはレジスタ・ファイルを含む、請求項 1 に記載の回路。

【請求項 4】

前記マップ・メモリはレジスタを含む、請求項 1 に記載の回路。

【請求項 5】

前記位置はビットマップとして格納される、請求項 4 に記載の回路。

【請求項 6】

前記ビットマップにおける各ビットは、前記格納された構造体における個々のデータの値に対応する、請求項 5 に記載の回路。

【請求項 7】

複数のコンパレータを更に含み、各コンパレータはデータ値がゼロでないかどうかを識別し、各コンパレータの出力は前記マップ・メモリへの入力として提供される、請求項 1 に記載の回路。

【請求項 8】

前記コンパレータ入力は、前記データメモリの読み込みポートによって設けられる、請求項 7 に記載の回路。

【請求項 9】

前記コンパレータは、前記データメモリの書き込みポートによって設けられる、請求項 7 に記載の回路。

【請求項 10】

前記コンパレータ入力は、プロセッサのロード／ストアポートによって設けられる、請求項 7 に記載の回路。

【請求項 11】

前記データ出力はデータバスを含み、前記回路は、前記メモリから前記データバスに前記圧縮した構造体を連続して出力するよう構成される、請求項 2 に記載の回路。

【請求項 12】

前記マップからゼロでない値の前記数を計算するために、少なくとも一つの加算器を更に含む、請求項 2 に記載の回路。

【請求項 13】

前記メモリから前記データ出力にゼロでない出力の前記書き込みを連続して可能にするためのロジックを更に含む、請求項 2 に記載の回路。

【請求項 14】

前記ロジックは加算器の構成を含む、請求項 13 に記載の回路。

【請求項 15】

前記構成におけるそれぞれの後続の加算器は、前記構成において先行する加算器の前記出力を、入力として有する、請求項 14 に記載の回路。

【請求項 16】

各加算器は前記構造体において付随するデータ値に対応し、各加算器は前記付随するデータ値に対応する前記マップから入力を受け付ける、請求項 2 に記載の回路。

10

20

30

40

50

【請求項 17】

更に整数コンパレータのツリーを含み、各整数コンパレータは二つの整数入力を比較するためのものであり、各コンパレータの第1の入力は前記加算器ツリーにおいて対応する加算器からの出力である、請求項14に記載の回路。

【請求項 18】

各コンパレータへの第2の入力は、配列信号である、請求項17に記載の回路。

【請求項 19】

前記データ出力に些少ではないデータを書き込む前記正しいシーケンスを確認するために、前記マップからの前記値を個々のコンパレータ出力と結合するためのコンバイナを更に含む、請求項17に記載の回路。

10

【請求項 20】

前記回路の前記動作を制御するためのコントローラを更に含む、請求項1に記載の回路。

【請求項 21】

前記データ値は、単精度浮動小数点数である、請求項1に記載の回路。

【請求項 22】

前記データ値は、倍精度浮動小数点数である、請求項1に記載の回路。

【請求項 23】

前記データ値は、拡張精度浮動小数点数である、請求項1に記載の回路。

20

【請求項 24】

前記データ値は、128ビット精度浮動小数点数である、請求項1に記載の回路。

【請求項 25】

前記データ値は、整数である、請求項1に記載の回路。

【請求項 26】

前記回路は、出力として前記マップの内容も提供するよう適合される、請求項1に記載の回路。

【請求項 27】

前記回路は、集積回路に設けられる、請求項1に記載の回路。

【請求項 28】

請求項1から請求項27のいずれか一つに記載の少なくとも一つの回路を含むプロセッサ。

30

【請求項 29】

前記回路の複数の実例が存在する、請求項28に記載のプロセッサ。

【請求項 30】

圧縮構造体から復元される構造体を提供するための復元回路であって、

a) 個々の些少ではないデータ値の圧縮構造体を受け付けるための入力と、

b) 復元構造体内において些少ではないデータ値の位置を識別するマップを受信するためのマップ・レジスタと、

c) 復元構造体を格納するためのメモリであって、前記回路は前記マップ・レジスタの内容にしたがって個々の入力された些少ではないデータ値を前記メモリに投入するメモリと、を含む回路。

40

【請求項 31】

前記メモリは、複数のレジスタを含む、請求項30に記載の回路。

【請求項 32】

前記メモリは、レジスタ・ファイルを含む、請求項30に記載の回路。

【請求項 33】

前記位置は、ビットマップとして格納される、請求項30に記載の回路。

【請求項 34】

前記ビットマップにおける各ビットは、復元構造体の個々のデータ値に対応する、請求項33に記載の回路。

50

【請求項 3 5】

前記データ入力はデータバスを含み、前記回路は前記データバスから前記メモリに連続して前記圧縮構造体を入力するよう構成される、請求項 3 0 に記載の回路。

【請求項 3 6】

前記マップから些少ではない値の数を計算するための少なくとも一つの加算器を更に含む、請求項 3 0 に記載の回路。

【請求項 3 7】

前記データ入力からメモリに些少ではないデータの前記書き込みを可能にするためのロジックを更に含む、請求項 3 0 に記載の回路。

【請求項 3 8】

前記ロジックは、加算器の構成を含む、請求項 3 7 に記載の回路。

【請求項 3 9】

前記構成のそれぞれの後続する加算器は、前記構成の前記先行する加算器の前記出力を入力として有する、請求項 3 8 に記載の回路。

【請求項 4 0】

各加算器は、非圧縮の前記構造体における関連データ値に対応し、各加算器は前記関連データ値に対応する前記マップからの入力を受け付ける、請求項 3 9 に記載の回路。

【請求項 4 1】

整数コンパレータの構成を更に含み、各整数コンパレータは二つの整数を比較するためのものであり、各コンパレータの第 1 の入力は前記加算器構成において対応する加算器からの出力である、請求項 4 0 に記載の回路。

【請求項 4 2】

各コンパレータへの第 2 の入力は、配列信号である、請求項 4 1 に記載の回路。

【請求項 4 3】

前記データ出力へのゼロでないデータの書き込みの前記正しいシーケンスを確認するために、前記マップからの前記値を個々のコンパレータ出力と結合するためのコンバイナを更に含む、請求項 4 1 に記載の回路。

【請求項 4 4】

前記回路の動作を制御するためのコントローラを更に含む、請求項 4 1 に記載の回路。

【請求項 4 5】

前記データ値は、単精度浮動小数点数である、請求項 3 0 に記載の回路。

【請求項 4 6】

前記データ値は、倍精度浮動小数点数である、請求項 3 0 に記載の回路。

【請求項 4 7】

前記データ値は、整数である、請求項 3 0 に記載の回路。

【請求項 4 8】

更にマップ入力を含み、前記回路は、前記マップ入力から前記マップ・レジスタへマップをロードする用適合される、請求項 3 0 に記載の回路。

【請求項 4 9】

前記回路は、集積回路に設けられる、請求項 3 0 に記載の回路。

【請求項 5 0】

些少なデータ値はゼロのデータ値であり、些少ではないデータ値はゼロでないデータ値である、請求項 3 0 から請求項 4 9 のいずれかに記載の回路。

【請求項 5 1】

請求項 3 0 から請求項 5 0 のいずれかに記載の少なくとも一つの回路を含むプロセッサ。

【請求項 5 2】

前記回路の複数の実例が存在する、請求項 5 1 に記載のプロセッサ。

【請求項 5 3】

データ構造体を格納する命令に応答する圧縮回路を含むプロセッサ・チップであって、

10

20

30

40

50

前記圧縮回路は、格納に対して圧縮フォーマットを提供するために前記構造体から些少な値を除去するよう適合される、プロセッサ・チップ。

【請求項 5 4】

前記圧縮回路は、前記構造体において些少な値の位置を識別するマップを提供するよう適合される、請求項 5 3 に記載のプロセッサ・チップ。

【請求項 5 5】

前記圧縮フォーマット・データをロードする命令に応答する復元回路を更に含み、前記復元回路は、前記圧縮データにロード時に些少な値を再投入するよう適合される、請求項 5 3 に記載のプロセッサ。

【請求項 5 6】

前記復元回路は、些少な値を再投入するためにマップを使用する、請求項 5 5 に記載のプロセッサ。

【請求項 5 7】

些少なデータ値はゼロのデータ値であり、些少ではないデータ値はゼロでないデータ値である、請求項 5 3 から請求項 5 6 のいずれかに記載のプロセッサ。

【請求項 5 8】

前記回路は、前記データ出力に並行して、前記取り出された個々のデータを多重に提供するよう構成される、請求項 1 に記載の回路。

【請求項 5 9】

前記個々のデータ値は x ビット長であり、前記データ出力は $n \times$ ビットのデータバスを含み、 n は 1 よりも大きな整数であり、 n データ値はある時刻に前記データバス上に配置される、請求項 5 8 に記載の回路。

【請求項 6 0】

前記回路は、個々の些少ではないデータ値の圧縮構造体を多重に受信するよう構成される、請求項 3 0 に記載の回路。

【請求項 6 1】

前記個々のデータ値は x ビット長であり、前記データ入力は $n \times$ ビットのデータバスを含み、 n は 1 よりも大きな整数であり、 n データ値はある時刻に前記データバスから取り出される、請求項 6 0 に記載の回路。

【請求項 6 2】

データ値の構造体を圧縮する方法であって、前記構造体内においてゼロの値の位置を識別するマップを生成するステップと、前記ゼロでない値及び前記マップのみからなる圧縮構造体を提供するために、前記構造体から前記識別された些少なエントリ値を除去するステップと、を含む方法。

【請求項 6 3】

前記データ値は、浮動小数点数である、請求項 6 2 に記載の方法。

【請求項 6 4】

前記データ値は、単精度又は倍精度浮動小数点数である、請求項 6 2 に記載の方法。

【請求項 6 5】

前記データ値は、拡張精度浮動小数点数又は 128 ビット精度浮動小数点数である、請求項 6 2 に記載の方法。

【請求項 6 6】

前記データ値は、整数である、請求項 6 2 に記載の方法。

【請求項 6 7】

前記マップは、個々のデータ値を表すビットマップの各ビットを有するビットマップを含む、請求項 6 2 ~ 6 6 のいずれかに記載の方法。

【請求項 6 8】

前記位置を識別する前記ステップは、ゼロでない値であるかどうか決定するために、各データ値を比較することを含む、請求項 6 2 ~ 6 7 のいずれかに記載の方法。

【請求項 6 9】

10

20

30

40

50

各比較の出力は、ゼロでない値の前記数に対して提供されるカウントに対して合計される、請求項 6 8 に記載の方法。

【請求項 7 0】

前記カウントは、前記圧縮構造体のサイズを決定するために用いられ、請求項 6 8 に記載の方法。

【請求項 7 1】

前記カウントは、前記圧縮構造体に設けられるエントリの前記数を決定するために用いる、請求項 6 8 に記載の方法。

【請求項 7 2】

各比較の前記出力は、前記圧縮構造体にデータ値の書き込みを有効化するために用いられる、請求項 6 8 に記載の方法。 10

【請求項 7 3】

前記構造体はマトリクスであり、前記マップは行及び列の前記数を識別する、請求項 6 2 から 7 2 のいずれかに記載の方法。

【請求項 7 4】

前記構造体は、行 - 列の構成に配列されるマトリクスを含む、請求項 6 2 に記載の方法。
。

【請求項 7 5】

圧縮データ構造体であって、複数のゼロでないデータ値と、前記構造体の非圧縮形態における複数の些少ではない値に対するゼロ・データ値の位置を表すマップとを含む、圧縮データ構造体。 20

【請求項 7 6】

前記マップはビットマップを含み、それぞれの独立したデータ値は独立したビットによって表される、請求項 7 5 に記載の圧縮データ構造体。

【請求項 7 7】

圧縮データ構造体を復元する方法であって、前記圧縮構造体は、複数のゼロでない値と、非圧縮構造体における前記ゼロでない値の位置を表すマップとを含み、前記方法は、未占有の非圧縮構造体を提供することと、

前記ゼロでない値を取り出すことと、

占有された復元データ構造体を提供するために、前記マップに表される位置にしたがって、前記未占有の構造体内の前記ゼロでない値を占有することと、
のステップを含む方法。 30

【請求項 7 8】

前記未占有のマトリクスの前記値は、前記ゼロでない値によって占有の前にゼロまで初期化される) 請求項 7 7 に記載の方法。

【請求項 7 9】

前記マップのゼロ値として識別される占有構造体における位置は、ゼロにセットされる、請求項 7 7 に記載の方法。

【請求項 8 0】

前記データ値は、浮動小数点数である、請求項 7 7 に記載の方法。 40

【請求項 8 1】

前記データ値は、単精度浮動小数点数である、請求項 7 7 に記載の方法。

【請求項 8 2】

前記データ値は、倍精度浮動小数点数である、請求項 7 7 に記載の方法。

【請求項 8 3】

前記データ値は、整数である、請求項 7 7 に記載の方法。

【請求項 8 4】

前記マップは、個々のデータ値を表す前記ビットマップの各ビットを有するビットマップからなる、請求項 7 7 から 8 3 のいずれかに記載の方法。

【請求項 8 5】

50

前記ビットマップの前記個々のビットは、前記圧縮構造体におけるゼロでない値の前記数のカウントを提供するために合計される、請求項 8 4 に記載の方法。

【請求項 8 6】

前記カウントは、前記未占有の構造体に読み込まれるデータ値の量を決定するために用いられる、請求項 8 5 に記載の方法。

【請求項 8 7】

前記マップは、前記非圧縮構造体へのデータ値の書き込みを有効化するために用いられる、請求項 7 7 から 8 6 のいずれかに記載の方法。

【請求項 9 9】

前記構造体は、行 - 列構成を有するマトリクスを含む、請求項 7 7 から 8 7 のいずれかに記載の方法。 10

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本出願は、データ圧縮方法に関し、特に構造体の圧縮に関する。

【背景技術】

【0 0 0 2】

メモリに格納されるデータ量を削減するための、圧縮方法及びアルゴリズムは、当技術分野において公知である。特に、例えば画像に対する J P E G 等、異なるデータの種類に対して、異なるアルゴリズムが知られている。 20

【0 0 0 3】

本出願は、マトリクス構造体の圧縮を目的とする。

【0 0 0 4】

以前に試みられた一つの方法（マローニー及びジェラティ、W O 2 0 0 6 1 2 0 6 6 4）は、非構造体化マトリクスから構造体化マトリクスへの変換により、マトリクス構造体を圧縮することを取り入れた。この圧縮方法は、マトリクスの主対角の上下にあるデータの重複を効果的に排除した。

【0 0 0 5】

マトリクス構造体を取扱う他の方法が米国特許第 6 5 9 1 0 1 9 号に提案されており、データ値のマトリクスは、b i t M a p テーブル、s i g n M a p テーブル及びd a t a M a p テーブルからなる構造体に圧縮される。ビットマップ・テーブルは、一連の 2 ビットのエントリを含み、それぞれのビットのエントリは、非圧縮マトリクス構造体のエントリに対応する。b i t M a p の各 2 ビットのエントリは、非圧縮のマトリクスの対応する値がゼロでもあるか、1 であるか、d a t a M a p にスケーリングされた形式で格納されているか、又はd a t a M a p に非圧縮形式で格納されているかを識別する。s i g n M a p テーブルは、非圧縮構造体における値の符号を識別する。この方法の不利な点は、値のスケーリングにおいて情報が失われるため、無損失ではないことである。該方法は、ソフトウェア実施に対してのみ実用的である。 30

【0 0 0 6】

上記の各方法は、メモリ内に構造体を格納するために必要なメモリがより少ない限りにおいては、従来技術に対して改良をもたらすが、プロセッサにおいて構造体を復元するときには、これらの方法ではまだ重大な計算コストを招く。加えて、上記の方法は、いずれも、圧縮及び復元に関する複雑さのために処理速度にも弱点があり、プロセッサにおける実行の際には、装置がパフォーマンスに関して比較的大で低速となる。 40

【発明の概要】

【発明が解決しようとする課題】

【0 0 0 7】

本出願は、実装コストを最小にし、実装の圧縮 / 復元速度を最大にすると共に、メモリ・アクセス・コストを最小にする、コンピュータ・システムの設計課題に対処することを目的とする。メモリ・アクセス・コスト（メモリ帯域幅）は、メモリ技術又はプロセッサ 50

接続様式に関わらず、コンピュータ・システムの設計における基本的な制約であり、プロセッサとメモリとの間で所与の時間内にどれだけのデータが転送できるかについては、最大の制約があり、これらは利用可能なメモリ帯域幅であり、利用可能なメモリ帯域幅によるコンピュータ能力の制約は、「メモリ・ウォール」と呼ばれることがある。

【0008】

本出願は、データを圧縮フォーマットに格納し、メモリ内に格納されなければならない巨大なデータセット上で動作する、例えばコンピュータ・グラフィック、剛体力学、有限要素分析及び他の科学及び工学アプリケーション等の多くのアプリケーションに用いられる、ブロック構造体データに適した圧縮及び復元手段を提供することにより、有効なメモリ帯域幅を増加させ、計算における「メモリ・ウォール」の制約を最小化することを目的とする。10

【課題を解決するための手段】

【0009】

本出願は、データ転送及び復元プロセッサ資源を占有せず、顕著の遅延がない一方で、実装コストに関しては低くしたまま、処理速度に関しては高速である、プロセッサのハードウェアに圧縮／復元回路をプロセッサのハードウェアに組み込むことによって、メモリ・ウォールの課題に対処する。比較的単純な圧縮／復元ハードウェアを許容する好ましい圧縮方法が提供される。

【0010】

第1実施形態は、個々のデータ値の構造体を格納するためのメモリと、構造体内の些少ではない（ゼロでない）値の位置を格納するためのマップ・メモリと、データ出力とを含む回路であって、本回路は該マップを用いてメモリから些少ではない（ゼロでない）データを取り出し、取り出したデータをデータ出力における圧縮構造体として提供するよう構成される回路を提供する。メモリは、レジスタ、複数のレジスタ及び／又はレジスタ・ファイルを適切に含む。位置は、ビットマップとして格納でき、ビットマップの各ビットは、格納された構造体における個々のデータ値に対応できる。複数のコンパレータを設けることが可能であり、各コンパレータは、データ値が些少ではないかどうかを識別し、各コンパレータの出力はマップへの入力として供給される。コンパレータ入力は、メモリの読み込みポート、メモリの書き込みポート、及び／又はプロセッサのロード／ストアポートに設けてもよい。20

【0011】

データ出力はデータバスを含んでもよく、本回路は、メモリからデータバスへ圧縮構造体を順次出力するよう構成される。本回路は、更に、マップからゼロでない値の数を算出するための少なくとも一つの加算器を含んでもよい。本回路は、更に、メモリからデータ出力までゼロでないデータを書き込むことを順次可能にするためのロジックを含んでもよい。該ロジックは、加算器の構成を含むことができ、該構成において後続の加算器は、該構成において先行する加算器の出力を入力として有する。各加算器は、構造体の関連データ値に対応でき、各加算器は関連データ値に対応するマップからの入力を受け付ける。本回路は、更に、整数コンパレータのツリーを含むことができ、各整数コンパレータは二つの整数入力を比較するためのものであり、各コンパレータにおける第1の入力は、加算器ツリーにおいて対応する加算器からの出力である。各コンパレータへの第2の入力は、配列信号である。本回路は、更に、些少ではないデータをデータ出力に書き込むという正しい配列を確実にするために、個々のコンパレータ出力を有するマップから値を結合するためのコンバイナを含んでもよい。本回路は、更に、本回路の動作を制御するためのコントローラを含んでもよい。本回路は、マップの内容を出力として提供するようにも適合できる。適切には、本回路は、集積回路上に設けることができる。本回路の一つ以上を含むプロセッサが提供されてもよい。30

【0012】

更なる態様において、圧縮構造体から復元構造体を抽出するための回路が提供される。本回路は、個々の些少ではないデータ値の圧縮構造体を受け付けるための入力と、復元構40

造体内の些少ではないデータ値の位置を識別するマップを受信するためのマップ・レジスタと、復元構造体を格納するためのメモリとを含み、本回路は、マップ・レジスタの内容にしたがって、メモリを些少ではないデータ値で占有するよう構成される。該メモリは、複数のレジスタ又はレジスタ・ファイルを含むことができる。該位置は、ビットマップとして格納でき、該ビットマップの各ビットは、復元構造体における個々のデータ値に対応する。

【0013】

データ入力はデータバスを含んでもよく、この場合に本回路は、データバスからメモリに圧縮構造体を順次入力するよう構成される。本回路は、更に、マップから些少ではない値の数を算出するための少なくとも一つの加算器を含んでもよい。回路は、更に、順次データ入力からメモリへの些少ではないデータの書き込みを順次可能にするための、好適には加算器の構成であるロジックを含んでもよい。この構成における後続の加算器は、この構成における先行の加算器の出力を入力として有することができる。各加算器は、非圧縮構造体における関連データ値に対応でき、各加算器は関連データ値に対応するマップからの入力を受け付ける。本回路は、更に、整数コンパレータの構成を含むことができ、各整数コンパレータは二つの整数入力を比較するため野茂のであり、各コンパレータの第1の入力は、加算器構成において対応する加算器からの出力である。各コンパレータへの第2の入力は、配列信号でもよい。本回路は、更に、些少ではないデータのデータ出力に対する書き込みの正しい順序を確実にするために、マップからの値を個別のコンパレータ出力と結合するためのコンバイナを含んでもよい。本回路は、更に、回路の動作を制御するためのコントローラを含んでもよい。本回路は、更に、マップ入力を含んでもよく、本回路は、マップ・レジスタに入力されるマップからマップをロードするよう適合される。適切には、本回路は、集積回路上に設けることができる。適切には、些少なデータ値はゼロ・データ値であり、些少ではないデータ値はゼロでないデータ値である。アプリケーションは、これらの回路の少なくとも一つを取り入れるプロセッサに拡張する。

10

20

30

40

【0014】

また更なる実施形態は、データ構造体を格納する指示に応答する圧縮回路を備えたプロセッサ・チップを提供し、この圧縮回路は、記憶部に圧縮フォーマットを提供するために構造体から些少な値を除去するよう適合される。圧縮回路は、構造体における些少な値の位置を識別するマップを提供するよう適合できる。プロセッサは、更に、圧縮フォーマット・データをロードする指示に応答する復元回路を含んでもよく、この復元回路は、些少な値をそれがロードされる際に圧縮データ内に再占有するよう適合される。復元回路は、些少な値が再占有するためのマップを使用できる。適切には、些少なデータ値はゼロ・データ値であり、些少ではないデータ値はゼロでないデータ値である。

【0015】

したがって、他の実施形態は、構造体内で些少な入力値の位置を識別する単一のマップを作成するステップと、一つのマップの情報だけから単に復元可能な圧縮構造体を提供するために、識別された些少な入力値を該構造体から除去するステップとを含む、データ値の構造体を圧縮する無損失な方法を提供する。データ値は、浮動小数点数（単精度又は倍精度）、拡張精度浮動小数点数、128ビット精度の浮動小数点数又は整数でもよい。

【0016】

適切には、マップは、個々のデータ値を表すビットマップの各ビットを有するビットマップを含む。位置を識別するステップは、それが些少ではないエントリ値であるかどうかを決定するために、各データ値を比較することを含むことができる。各比較の出力は合計され、些少ではないエントリ値のカウント数を提供できる。このカウントは、圧縮構造体のサイズ及び／又は圧縮構造体に提供されるエントリ数を決定するために使用できる。各比較の出力は、圧縮構造体へのデータ値の書き込みを可能にするために使用できる。構造体はマトリクスでもよく、マップは列及び行の数を識別できる。好み構成において、些少な入力値はゼロ値であり、些少ではない値はゼロでない。

【0017】

50

更なる実施形態は、複数の些少ではないデータ値と、複数の些少ではないデータ値に対する些少なデータ値の位置を表すマップとを含む、圧縮データ構造体を提供する。好ましい構成において、些少な入力値はゼロ値であり、些少ではない値はゼロでない。マップは、個々のビットによって表される個々のデータ値を有するビットマップを含んでもよい。

【0018】

更なる実施形態は、圧縮データ構造体を復元する方法を提供し、この圧縮構造体は、複数の些少ではないデータ値と、非圧縮構造体における些少ではない値の位置を表すマップとを含む。本方法は、占有された復元データ構造体を提供するために、未占有の非圧縮構造体を提供するステップと、些少ではない値を取り出すステップと、些少ではない値をマップに表されるその位置にしたがって当該未占有の構造体に占有させるステップとを含む。好ましい構成において、些少な入力値はゼロ値であり、些少ではない値はゼロでない。この場合、未占有のマトリクスの値は、ゼロでない値による占有の前に、ゼロに初期化してもよい。あるいは、マップにおいてゼロ値として識別される占有された構造体の位置が、ゼロにセットされてもよい。マップは、個々のデータ値を表しているビットマップの各ビットを有するビットマップを含んでもよい。ビットマップの個々のビットは合計され、圧縮構造体におけるゼロでない値のカウント数を提供できる。このカウントは、未占有の構造体に読み込まれるデータ値の量を決定するために使用できる。適切には、マップは、非圧縮構造体へのデータ値の書き込みを可能にするために使用される。構造体は、行列構成を有するマトリクスを含むことができる。

10

【0019】

これらの及び他の実施形態、特徴及び利点は、後述の例示的記載から明らかである。

20

【0020】

以下、添付図面を参照し、本発明を説明する。

【図面の簡単な説明】

【0021】

【図1】本出願に係る圧縮方法の例示的フローチャートである。

30

【図2】非圧縮構造体の例、及び図1の方法の結果として生成する圧縮構造体である。

【図3】本出願に係る復元方法の例示的フローチャートである。

【図4】本出願に係る例示的プロセッサである。

【図5】図4のプロセッサへの導入に適切な例示的復元回路である。

30

【図6】図5の動作を説明する例示的タイミング図である。

【図7】図4のプロセッサへの導入に適切な例示的復元回路である。

【図8】図7の動作を説明する例示的タイミング図である。

【図9】図5及び図7の回路がどのように結合できるかを説明する例示的構造体である。

【発明を実施するための形態】

【0022】

本出願は、3次元コンピュータ・グラフィック、ゲーム物理学（剛体力学）、有限要素解析（F E A）及び検索エンジン等に使用されるような、巨大な浮動小数点マトリクス・データセットに含まれるゼロ・データ値を利用する、新規な方法を採用する。しかしながら、本アプリケーションは他のデータ構造体に適切であり、単にマトリクス・ベースの構造体だけに限定されない。

40

【0023】

多くのマトリクスに対して、エントリの顕著な数は、浮動小数点表現における32ビット又は64ビットを全く占有しないゼロ充填であり、これはオンチップ又はオフチップのメモリから取り出されなければならず、メモリ又はレジスタから取り出されたゼロに対する些少な動作を実施することにプロセッサを忙殺する可能性があることを、発明者は認識してきた。

【0024】

本出願は、ゼロ・データ値が除去された圧縮構造体において、データ値の行及び列に構成される格納用マトリクス構造体を提供する。マトリクス構造体におけるゼロ及びゼロで

50

ない値の位置を識別し、マトリクス構造体の無損失の再構築を可能にする、圧縮構造体に対する関連するマップが提供される。値が些少な値として識別されなければ、それは些少ではなく、逆も同様であることは明らかであるので、このようなマップは、些少な値の位置を識別するもの、又は些少ではない値の位置を識別するものと見なすことができることはいうまでもない。本方法は、従来技術に対して多くの重要な効果を提供する。第1に、本方法は無損失であり、第2に、本方法は直ちにハードウェアに実装されるため、メモリへのロード時に圧縮構造体の復元に伴う計算負荷を低減できる。同様に、処理リソースの追加を招くことなく、データを圧縮形態に格納できる。

【0025】

このマップは、圧縮構造体に付随して、又は、複雑なデータ構造体へのランダムアクセスが必要な構造体の索引付けの一部として、メモリ内に格納できる。

10

【0026】

本発明の技術の利点は、32以下のデータ構造体エントリを表すために32ビットのビットマップが用いられることを仮定し、ゼロ値の数に対するメモリからの転送要求を、単精度値に対しては32ビットから1ビットに、又は倍精度値に対しては64ビットから1ビットに低減することである。一般性を全く失うことなく、32以上又は以下のエントリを有するデータ構造体を提供するため、ビットマップ・サイズが調節可能であることは明らかである。同じ効果は、整値を用いても得られる。

【0027】

以下、図1の例示的シーケンスのフローチャート、及び図2の16の32ビット単精度エントリを含む例示的 4×4 マトリクスを参照し、動作モードを記載する。

20

【0028】

この処理は、圧縮マトリクス構造体から第1のエントリを取り出すことから開始する。第1のエントリは、それがゼロ値かゼロでない値であるかどうかを決定するために比較される。比較の結果により、第1の値をゼロとして、又はゼロでない値として識別するために、マップの第1の位置にエントリが作られる。適切には、マップは、ビットマップ内に関連ビットを有するマトリクス構造体における各値を有するビットマップである。第1のエントリがゼロ値である場合には、このエントリは廃棄、削除又は後の廃棄又は削除のためにタグ付けされ（例えばマトリクス・データから圧縮ストレージへの転送中に）、例えば第1ビットをゼロにセットすることにより、ゼロ値が存在したことを示すため、マトリクスに対するマップの第1の位置にエントリが作られる。マトリクスにおいて、例えば図2の第1のエントリが1.0である等、第1の値がゼロでない値である場合は、マップにおける第1のエントリは1にセットされ、ゼロでない値であることを指示する。同時に、この第1のエントリは、圧縮構造体における第1のエントリとして格納される。

30

【0029】

次いで、マトリクスにおける残りのエントリのそれぞれに対して、この処理が繰り返される。図示の例示的マトリクスにおいて、エントリは行ごとにスキャンされるが、列ごとのスキャンも使用できること、又は単一のコンパレータを用いて順次的にするよりもむしろ複数のコンパレータを使用して单一操作でビットマップ全体の行／列を生成できることを考慮されたい。

40

【0030】

したがって、1.0、2.0、3.0及び4.0という値を有する第1の行に対しては、ゼロ値のエントリがないため対応するマップは1111であるが、第2の行においては2番目の値のみがゼロでない値(6.0)であるため、マップは0100である。同様に、第3の行に対しては、この行の最初の三つの値がゼロであるため、マップ内の最後のエントリのみが1である。

【0031】

この圧縮処理の結果、16のエントリとして（単精度に対してはそれぞれ32ビット）メモリに以前に格納されたマトリクスは、9の圧縮マトリクス（同様にそれぞれ32ビット）及びこれらのゼロでない値及び非圧縮マトリクスにおいて、対応するゼロ値を表す1

50

6 ビットのマップに置き換えることができる。各メモリの位置は 32 ビットであるため、該マップがマトリクスを用いてどこに格納されるかというマップの格納に対して、16 よりも 32 ビットが必要とされる。それにもかかわらず、 16×32 ビットの非圧縮マトリクス・フォーマット（全体で 512 ビット）は、 9×32 ビットの圧縮マトリクス及び 1×32 ビットのマップ（全体で 320 ビット）に置き換えられ、これは $(512 - 320) / 512 = 37.5\%$ の圧縮を表している。

【0032】

高密度マトリクス（ゼロ充填を含む）に対して一つ以上のエントリがゼロ（32 ビットのビットマップが使用される場合）である限り、圧縮は達成される。しかしながら、エントリあたりに 1 ビットのペナルティがあり、ゼロ充填なしで高密度のデータセットを格納するために、より多くのメモリが用いられる結果となる場合がある。実際には、3 次元コンピュータ・グラフィック、ゲーム物理学（剛体力学）、有限要素解析（FEA）及び検索エンジン（例えば Google（商標））を含む関心データセットは、まばらであり、大量のゼロ充填を含む。

10

【0033】

本出願は、複数の行及び列に関して記載されるが、单一の行又は列からなる構造体の圧縮にも適用できる。

【0034】

いうまでもなく、いくつかの状況において、マトリクスを再構成するためにマトリクスの列・行構造、すなわち行及び列の数を知ることが必要な場合がある。しかしながら、膨大な数のアプリケーションにおいて、コンテンツはメモリから取り出され、スカラー形態で、すなわち図 2 の單一カラムに示すようにレジスタに格納され、これは無関係であり、該データをマトリクスと見なすのは、ソフトウェア／プロセッサである。加えて、圧縮と復元との間の様式に整合性があれば、行・列構造の存在は不明確でもよい。状況によっては、列・行構造についての知識が必要とされが、これはマップ内に / マップに追加して含まれてもよい。

20

【0035】

図 3 に示すように、圧縮構造体を非圧縮構造体に展開する処理は、圧縮方法の逆である。この方法は、空の非圧縮構造体を提供し、マップをロードすることから始まる。次いで、マップの第 1 ビットは、ゼロであるかゼロでない値であるかを決定するために比較される。第 1 ビットが 1 である場合は、第 1 のデータ値は、メモリから非圧縮構造体の第 1 のエントリにロードされる。同様に、第 1 ビットがゼロである場合は、非圧縮構造体の第 1 のエントリにゼロ値を挿入する。非圧縮構造体の全ての値がゼロに初期化される場合には、ゼロ値を挿入する特別なステップは必要ではない。マップの終端に達するまで、又は全てのデータ値がロードされるまで、この処理が繰り返される。ロードするデータ値の数は、単にマップの個々のゼロでないビット（圧縮されたゼロでないマトリクス／ベクトル・エントリに対応する）を計数することによって算出できる。

30

【0036】

図 2 の圧縮構造体を復元する場合、マップの最初の 4 つの値は 1 であるため、最初の 4 つのデータ値が、圧縮構造体から非圧縮構造体の最初の 4 つのエントリにロードされる。マップの 5 番目のビットはゼロであるため、非圧縮構造体の 5 番目のエントリはゼロ値で占有され、以下同様である。

40

【0037】

圧縮構造体が非圧縮構造体に展開されるときでも、圧縮構造体に対するマップを保持して使用し、引き続き非常に効果を奏することが可能である。特に、マップは、出願人の同時係属中の出願に記載のように、乗算、加算又は他の算術演算の結果として、プロセッサ内で機能ユニットの制御に効果的に使用できる。

【0038】

効率に関しては、マトリクス構造体サイズが比較的大きい場合、これを複数の下位マトリクスに分割し、それぞれの下位マトリクスがマトリクス構造体の一部を表してもよい。

50

この構成においては、上記の方法で下位マトリクスのそれぞれを個々に圧縮できる。

【0039】

上記の方法は、ソフトウェアに実装でき、図4に示すように、データはプロセッサと往復する際にハードウェアにおいて復元、圧縮され、処理速度に関して非常に効果を得ることが可能である。選択された圧縮構造体は、特に、スケーリング及び他の操作を含む従来技術の方法とは対照的に、ハードウェア実装に適していることはいうまでもない。ハードウェア構成において、プロセッサは、例えばマトリクスの格納／取り出し等、ロック形式のデータの取り出し及び／又は格納のために特定の指示を有するよう適合できる。このような構成において、圧縮及び復元の方法／ハードウェアは、中央プロセッサ・コアに対して透過的とすることが可能であるが、上記に説明のようにプロセッサ自体の中でマップを使用することが非常に有効である場合があり、同様に、後述するように、プロセッサ・ハードウェア内に／これと組み合わせてマップを生成することが効果的である場合がある。いずれにしても、圧縮／復元の動作は、プログラマ／オペレーティング・ソフトウェアを実質的に透過的とする様式で効果的に実行できることはいうまでもない。

10

【0040】

本出願を、オンチップ又はオフチップのメモリからプロセッサにデータをロード及び格納することに関する記載してきたが、多様な他のアプリケーションが可能であることはいうまでもなく、例えばコプロセッサ間でのデータ共有を含むことを、更に考慮されたい。

【0041】

以下、メモリからプロセッサへのロードの際に、圧縮データ構造体の復元／展開を提供する例示的ハードウェア実装を記載する。説明のために、下記の表1に例示する比較的小さな 3×3 （9エントリ）の圧縮マトリクスを用い、ここにA、B、C、D及びEはゼロでないデータ値を表し、101010101は 3×3 マトリクスのゼロでない値に対するマップである。実際には、提案する技法は、復元データに対する転送先として機能し、 $N \times M$ エントリ以上を有する、いずれの任意の大きさの $N \times M$ マトリクス及びレジスタ・ファイルにも拡張できる。

20

【0042】

【表1】

A
B
C
D
E
101010101

30

【0043】

例示的な復元ロジックは、図5に示すように、圧縮ビットマップ・レジスタ、書込アドレス計算機、書込アドレス・コンパレータ、コンバイナ及び転送先オンチップ・メモリを含む。例示的装置において、これらの素子は、例えばアドレス生成ユニット（AGU）又はダイレクト・メモリ・アクセス（DMA）コントローラ等のコントローラの制御下にある。これらの素子の構成及び動作を、以下、図6の例示的なタイミング図を参照し、より詳細に説明する。

40

【0044】

圧縮ビットマップ・レジスタは、圧縮マトリクス構造体に対するマップを格納する十分なサイズのレジスタであり、実際には、これは、サポートする $N \times M$ 圧縮マトリクスにおける各エントリについて、1ビットのエントリに対応する。例示的装置において、圧縮ビットマップ・レジスタは、9ビット（MCB0 - MCB8）を含み、各ビットは非圧縮構造体の位置に対応する。展開処理の開始時に、コントローラは、オンチップ又はオフチップのメモリから圧縮ビットマップ・レジスタへマップをロードする。図5のビットマッ

50

プ・レジスタに示す値は、単に例示的な値であることを考慮されたい。同様に、値のデータは、圧縮マトリクス・データに対する読み出アドレスを生成するコントローラ制御下のバスにおいて、該データ全体にわたり、メモリからデータに順次転送される。あるいは、データは、ゼロでないエントリのグループ内にメモリから転送でき、例えば、 2×32 ビットの値を 64 ビットバスを介して転送でき、配線増加というコストはあるものの一般性を失うことはない。コントローラの挙動は、復元ロジック及び圧縮ビットマップ・レジスタにより制御され、書き込みアドレス計算機を用いて、クロックサイクルによりベースアドレスから開始するフェッチ対象バイト数(バイト・フェッチ)を生成できる。

【0045】

書き込みアドレス計算機は、オンチップ・メモリ(レジスタ・ファイル)上の各エントリに対する書き込みアドレス(WA)を計算するために、2の補足整数加算器のツリーを含む。ツリーの各加算器は、2つの入力を有する。第1の入力は、ビットマップ・レジスタからの対応する値であり、すなわち、ビットマップ・レジスタにおける第1のエントリ(MCB0)は、第1の加算器への入力である。同様に、ビットマップ・レジスタの最後のエントリ(MCB8)は、ツリーの最後の加算器への入力である。各加算器への第2の入力は、ツリーにおける以前の加算器からの出力を含むので、第1の加算器(WA0)の出力は、第2の加算器への入力であり、同様に、第8の加算器(WA7)の出力は最後の加算器への入力である。

【0046】

第1の加算器はツリーにおいて下位の加算器を有さないため、-1のベース入力が設けられ、これはレジスタ・ファイル内の展開構造体にデータを格納するときに確実に正確なアドレス割り当てを行うためである。小さい整数の加算器ツリーは、MAPレジスタ内のエントリに基づいて、レジスタ・ファイル転送先(書き込み)アドレスを生成する。図示の特定の実施例は、9エントリのレジスタ・ファイルであり、加算器ツリーのベースアドレスは-1に設定される。より大きなレジスタ・ファイル(9以上のエントリ)の場合には、 3×3 マトリクスは、加算器ツリーのベースアドレスを base_addr - 1 に設定することにより、ベースアドレスから開始して復元できる。後者の場合、 3×3 マトリクスは、位置0よりもむしろ特定のベースアドレスで開始するレジスタ・ファイルに配置されることになる。

【0047】

加えて、加算器のツリーにおける最後の加算器からの出力は、メモリから取り出す必要のあるゼロでない値がいくつあるかを(AGU/DMA)コントローラに指示する、バイト・フェッチ値として提供される。これらのバイトは、データバス(data_in)上ではシーケンシャルな様式でロードされる。より広いバスを介してペア又は4つのグループのいずれかでデータが転送される場合は、64又は128ビットワード数を転送することを示すために、最下位1又は2ビット(1sbs)は、バイト・フェッチ値から削除される(かつ丸められる)。

【0048】

書き込みアドレス・コンパレータは、ワードポインタ(wrd_ptr)値に対して書き込みアドレス(WA)を比較する2の補足整数コンパレータのツリーを含み、これはコントローラに設けられるカウンタ出力であり、加算器ツリー終端由来のバイト・フェッチ出力の値までゼロからインクリメントする。

【0049】

コンバイナは、データバスからのデータをレジスタ・ファイルの適切な場所にロードできる書き込み有効(WEN)信号を生成するために、その対応するビットマップ・レジスタ値を用いてそれぞれの書き込みアドレス・コンパレータ出力の論理積を取るためのANDゲートを含む。特に、このwrd_ptr値は、データがレジスタ・ファイルにラッチされるべき時点及び場所を決定するために、加算器ツリーにより生成されるWAアドレス(出力)に対して比較される。

【0050】

10

20

30

40

50

状況によっては、ビットマップ・レジスタのエントリがゼロ（例えば、M C B 1、M C B 3、M C B 5、M C B 7）である場合は、関連付けられるANDゲート（W E N 1、W E N 3、W E N 5、W E N 7）の出力は、決して有効化されず、レジスタ・ファイルにおいて対応する位置（r 1、r 3、r 5、r 7）には何のデータもロードされないことが明らかである。このため、それらはタイミング図には示されない。レジスタ・ファイルは、処理における最初のステップとして、クリア（全てゼロ値にセットされる）が可能である。コンパレータ値が論理積に一致し、対応するM C B ビットが1にセットされるときには、レジスタ・ファイルのW E N（書き込み有効）ビットはハイにセットされる。セットされると、W E N ビットにより、d a t a _ i n バスの内容は正しいレジスタ・ファイルのレジスタに書き込まれる。図示の例において、第1の書き込み有効信号（W E N 0）は、w r d _ p t r がゼロであるときに、すなわち第1のゼロでない要素をレジスタ・ファイルの位置r 0にロードするために、有効化される。同様に、W A 2 = M C B 2 { 7 } + W A 1 { = M C B I (0) + W A 0 (0) } = 1 であり、したがって第2のゼロでないデータ値（B）は、第3のレジスタ位置r 2にデータバスからロードされることになるため、第3の書き込み有効信号（W E N 2）は、w r d _ p t r 値が1であるときに有効化されることになる。

10

【0 0 5 1】

その結果、復元／展開処理の終了時において、レジスタ・ファイル内には非圧縮構造体が格納され、図示の例は、A、0、B、0、C、0、D、0、E、0を表す。

20

【0 0 5 2】

3×3 圧縮マトリクスに対してここで説明した同じ一般原理を用いて、いかなる任意の $N \times M$ 圧縮マトリクス及び復元マトリクスのエントリのためのいかなる任意のサイズのレジスタ・ファイル転送先もサポートできる。

30

【0 0 5 3】

類似の構成は、圧縮経路（圧縮回路）に使用できる。図7に示す圧縮経路の場合には、非圧縮データ構造体（すなわち各レジスタ・ファイルのエントリ）内の各データ値を0.0に比較し、レジスタ・ファイルのエントリがゼロでない場合にビットマップ・エントリ・ビットを1に設定するために、浮動小数点コンパレータのツリーが必要である。9入力レジスタ・ファイル9のためのIEEE单精度登録事項の例においては、このような浮動小数点コンパレータが必要である。

30

【0 0 5 4】

図7に示す例示的な圧縮回路は、32ビット・コンパレータのバンク、圧縮ビットマップ・レジスタ、読出アドレス計算機、読出アドレス・コンパレータ、及びコンバイナを含む。上述の展開回路と同様に、これらの素子は、例えば、アドレス生成ユニット（A G U）又はダイレクト・メモリ・アクセス（D M A）等のコントローラの制御下にある。以下、これらの素子の構造及び動作を、図8の例示的なタイミング図を参照し、より詳細に説明する。

40

【0 0 5 5】

例示的な構成において、非圧縮構造体（r 0 ~ r 8）はレジスタ・ファイルに格納され、レジスタ内のそれぞれの値は、コンパレータ・バンク内の対応するコンパレータに提供され、個々の値がゼロでない値かゼロ値であるかを決定するために比較が実行される。

40

【0 0 5 6】

コンパレータは、レジスタ・ファイルからの読出値として示しているが、レジスタ・ファイル内のエントリのいずれがゼロ値／ゼロでない値であるかを指示する、ゼロビット・レジスタを有することが、より効果的である場合がある。このレジスタは、データがレジスタ・ファイルの書き込みポートにラッチされる際に占有される。このゼロビット・レジスタは、他の目的（例えば、計算の高速化）に効果的に使用でき、同一出願人の同時係属出願において、より詳細に記載される。圧縮サイクル開始時に比較の必要がないため、この方法はクロックサイクルを節約する。加えて、レジスタ・ファイル上で並列比較を実施するために、配線と電力との両者の点で実装が非常に高価である9つの読出ポートを有する必

50

要性は、取り除かれる。この方法の更なる効果は、必要とする浮動小数点コンパレータ数の減少を含む（レジスタ・ファイルに対して3つの書きポートのある例示的構成においては $3 \times$ ）。

【0057】

複数のロード／ストアポートを有するプロセッサにおいて、圧縮／復元ロジックは、单一のロード／ストアポートを有するプロセッサの場合には共有することが可能であり、单一の共有レジスタ・ファイル又は複数の独立したレジスタ・ファイルのいずれが使用されても、それぞれの独立したロード／ストアポートに対して1回複製されなければならないことに注意すべきである。

【0058】

ただし、単に説明の容易さのために、図7の装置を参照して動作モードを説明するが、コンパレータ・バンクからの個々の出力は、圧縮ビットマップ・レジスタへの入力として設けられ、これは上述のように、圧縮／非圧縮マトリクス構造体に対するマップを格納する十分なサイズのレジスタである。しかしながら、いったんビットマップ・レジスタがロードされると、二つの異なる装置間では動作様式が同じであることはいうまでもない。コンパレータ出力は、圧縮処理の第1段階として、圧縮ビットマップ・レジスタにロードされる。図7のビットマップ・レジスタに示される値が、単に例示的な値であることはいうまでもない。

【0059】

コントローラの挙動は、圧縮ロジック及び圧縮ビットマップ・レジスタによって制御され、読み出アドレス計算機を用いて、クロックサイクルによりベースアドレスから開始する格納対象バイト数（バイト・ストア、後述する）を生成できる。

【0060】

読み出アドレス計算機は、オンチップ・メモリ（レジスタ・ファイル）上の各エントリに対する読み出アドレス（RA）を計算するために、2の補足（図示の例においては5ビット）整数加算器のツリーを含む。ツリーの各加算器は、2つの入力を有する。第1の入力は、ビットマップ・レジスタからの対応する値であり、すなわち、ビットマップ・レジスタにおける第1のエントリは、第1の加算器への入力である。同様に、ビットマップ・レジスタの最後のエントリは、ツリーの最後の加算器への入力である。各加算器への第2の入力は、ツリーにおける以前の加算器からの出力を含むので、第1の加算器（RA0）の出力は、第2の加算器への入力であり、同様に、第8の加算器（RA7）の出力は最後の加算器への入力である。

【0061】

第1の加算器はツリーにおいて下位の加算器を有さないため、0のベース入力（すなわち、入力なし）が設けられ、これはレジスタ・ファイルからのデータ読み出時に確実に正確なアドレス割り当てを行う。整数加算器のツリーは、マップ・レジスタのエントリに基づいて、レジスタ・ファイル転送先（読み出）アドレスを生成する。

【0062】

加算器ツリー内の最後の加算器からの出力は、（AGU/DMA）コントローラに対してオンチップ又はオフチップ・メモリ内の圧縮構造体に格納が必要なゼロでない値がいくつあるかを指示する、バイト・ストア値として提供される。これらのバイトは、データバス（data_out）を介して、シーケンシャルな様式で格納される。

【0063】

読み出アドレス・コンパレータは、ワードポインタ（rd_ptr）値に対して読み出アドレス（RA）を比較する2の補数整数コンパレータのツリーを含み、これはコントローラに設けられるカウンタ出力であり、加算器ツリー終端由来のバイト・ストア出力の値までゼロからインクリメントする。

【0064】

コンバイナは、ANDゲートのツリーを含む。各ANDゲートは、適切な位置でレジスタ・ファイルからデータを抽出できる読み出有效（REN）信号を生成するために、読み出ア

10

20

30

40

50

ドレス・コンパレータ出力をその対応するビットマップ・レジスタと結合する。特に、この $r_d_p_t_r$ 値は、データがレジスタ・ファイルから読み出されるべき時点及び場所を決定するために、加算器ツリーにより生成されるRAアドレス（出力）に対して比較される。

【0065】

状況によっては、ビットマップ・レジスタのエントリがゼロ（例えば、MCB1、MCB3、MCB5、MCB7）である場合は、関連付けられるANDゲート（REN1、REN3、REN5、REN7）の出力は有効化されず、レジスタ・ファイルにおいて対応する位置（r1、r3、r5、r7）からは何のデータも読み出されないことが明らかである。このため、それらはタイミング図には示されない。コンパレータ値が論理積に一致し、対応するMCBビットが1にセットされるときには、レジスタ・ファイルのREN（読み出有効）ビットはハイにセットされる。セットされると、RENビットにより、レジスタ・ファイル内の選択されたエントリの内容が、data_outバスに配置される。図示の例において、第1の読み出有効信号（REN0）は、 $r_d_p_t_r$ がゼロであるときに、すなわち第1のゼロでない要素（A）をレジスタ・ファイルの位置r0からdata_outに読み出すために、有効化される。同様に、 $RA_2 = MCB2\{7\} + RA1\{7\} = 2$ であり、したがって第2のゼロでないデータ値（B）は、第2のゼロでないデータ値（C）は第3のレジスタ位置r2からdata_outにラッチされることになるため、第3の読み出有効信号（REN2）は、 $r_d_p_t_r$ 値が2であるときに有効化されることになる。

10

20

【0066】

その結果、圧縮処理の終了後、レジスタ・ファイルは、マップ・レジスタ由来のマップと共にオンチップ又はオフチップ・メモリ内のコントローラによって格納されたdata_out上に、A、C、E、G及びIを配置することになる。

【0067】

3×3 非圧縮マトリクスに対して本願明細書において概説した同一の一般原理を用いて、いずれの任意の $N \times M$ 非圧縮マトリクスもサポートできることはいうまでもない。

【0068】

上述の説明から、コントローラの一般的制御下において同一のハードウェアを圧縮及び復元のために使用できることを、理解されたい。

30

【0069】

図9に、このような組み合わせ構成を示し、これは上述のレジスタ・ファイル書き込ポートに取り付けられるコンパレータを有する構成を取り入れている。この例では、レジスタ・ファイルに接続するプロセッサ・データ経路は、3つの書き込ポートを必要とする。実際には、この構成案は、任意の数のレジスタ・ファイル書き込ポート、及び任意の数のエントリを有するレジスタ・ファイルに容易に拡張できる。オンチップ又はオフチップ・ストレージにデータを圧縮するため、又はオンチップ又はオフチップ・ストレージからレジスタ・ファイルへのロード時にデータを展開するために、必要に応じてw_r_p_tr、r_d_p_tr及び他の信号が動作／受信されることはいうまでもない。

40

【0070】

base_addrが-1であることは、簡便な説明のために示していない。

【0071】

プロセッサの実行バスと直列的に浮動小数点コンパレータを書き戻す遅延が発生しないことを確実にする目的で、追加の実行パイプライン・ステージを必要とする場合があるが、さもなければ該プロセッサのクロックレートとこれによるFLOPSレートを制約する可能性があるためである。

【0072】

本出願に記載の新規な方法は、例えば、些少な値の格納に要するメモリを低減することによりプロセッサが必要とするメモリ帯域を削減すること、メモリを往復するオンチップ又はオフチップのバスを介する帯域に関して効率的に、圧縮又は非圧縮のいずれにおいて

50

も、スカラー、ベクトル又はマトリクス・データの移動に要するバスのサイズを削減すること、メモリ又はプロセッサ・レジスタから読み出されるときに、圧縮されたスカラー、ベクトル又はマトリクス・データ上で動作時に電力消費を削減し、効率的な処理パワー（F L O P S）を向上すること、及びプロセッサのレイテンシを低減することを含む多数の重要な効果を提供する。

【0073】

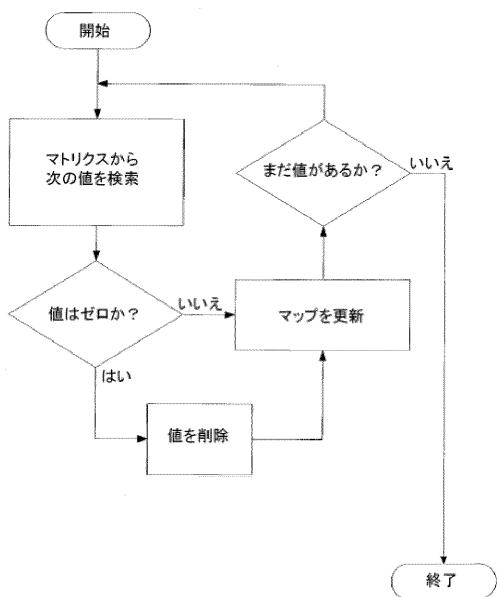
添付図面において、図示のラインは複数のラインに対応でき、これは隣接する番号を有する交差線によって表されており、例えば、図9におけるstore_dataバスは32ビット・バスを表す一方で、加算器ツリーからの各出力は5ビットである。

【0074】

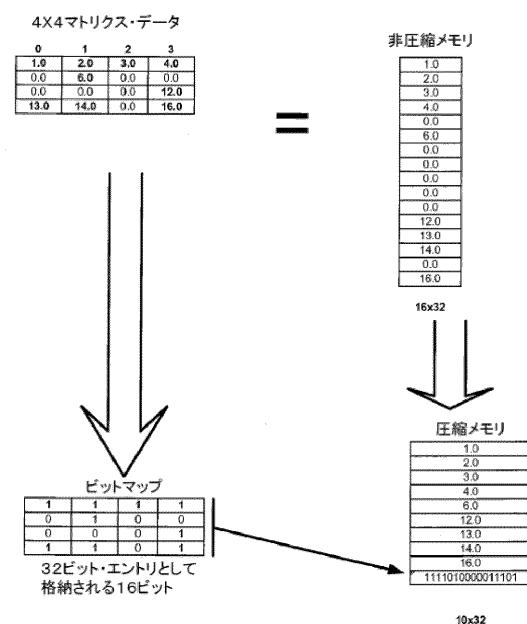
含む／からなるという言葉は、本明細書における使用時には、記載の特徴、整数、ステップ又は構成要素の存在を特定するためであるが、一以上の他の特徴、整数、ステップ、構成要素又はこれらの群の存在又は追加を排除するものではない。

10

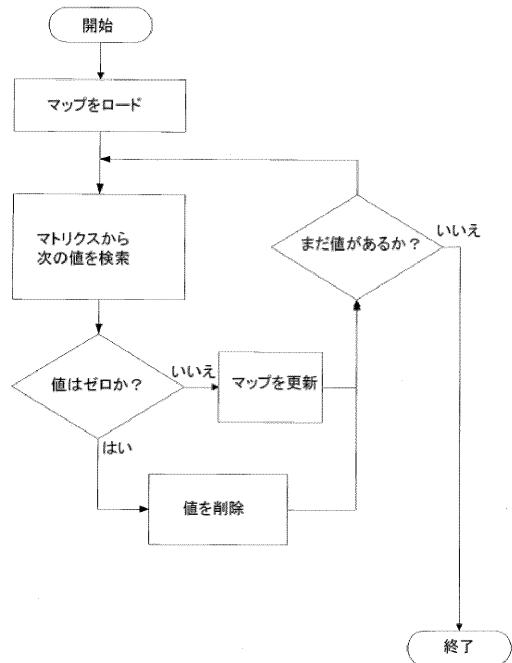
【図1】



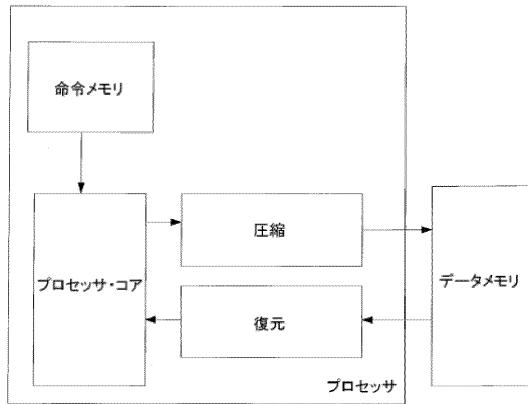
【図2】



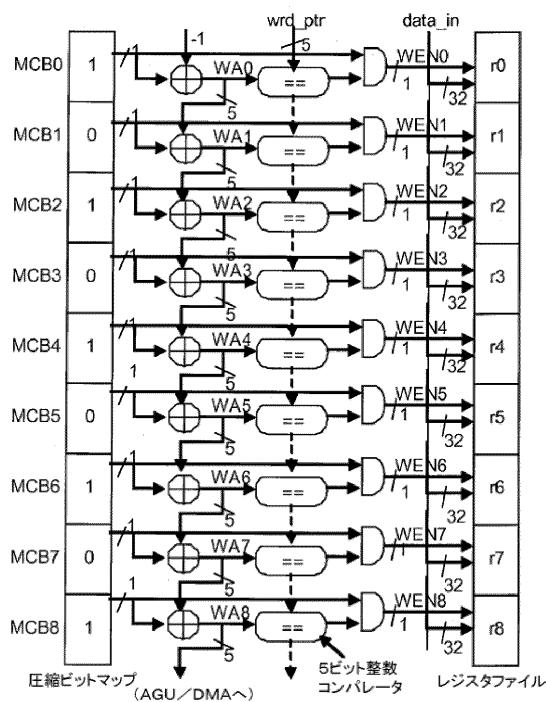
【図3】



【図4】



【図5】



【図6】

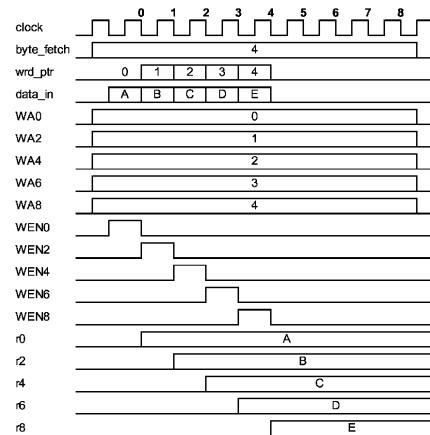
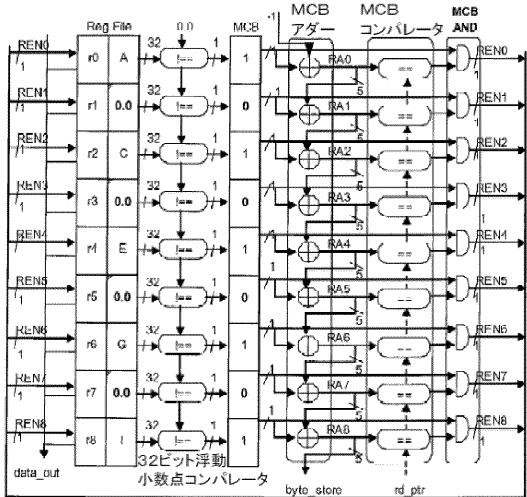


Figure 6

【図7】



【図8】

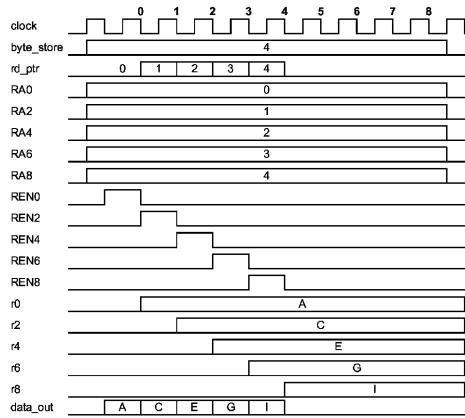
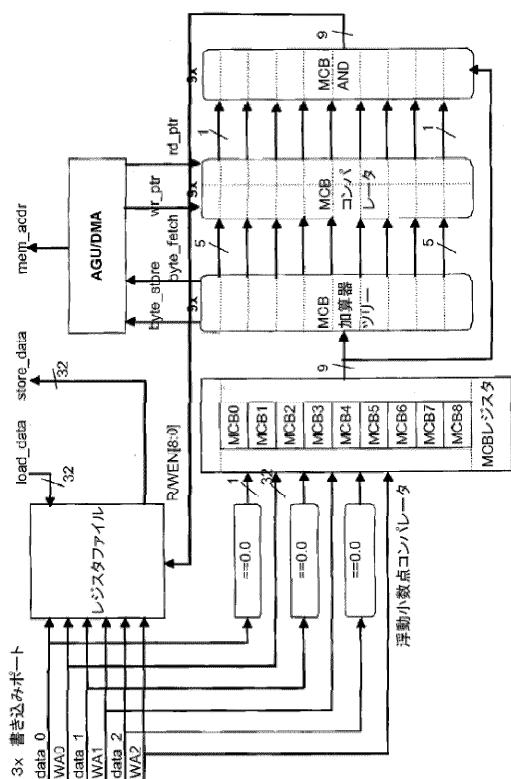


Figure 8

【図9】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/053133

A. CLASSIFICATION OF SUBJECT MATTER INV. H03M7/30		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) H03M		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, INSPEC		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/43074 A (NINTENDO CO LTD [JP]; COMAIR CLAUDE [CA]; LI XIN [US]; ABOU SAMRA SAMI) 14 June 2001 (2001-06-14) the whole document page 7 - page 9 figure 3	1-99
A	SALOMON D ED - SALOMON D: "DATA COMPRESSION THE COMPLETE REFERENCE, PASSAGE" DATA COMPRESSION : THE COMPLETE REFERENCE, NEW YORK, NY : SPRINGER, US, 1 January 1998 (1998-01-01), pages 6-10, XP002315914 ISBN: 978-0-387-98280-9 the whole document	1-99
<input type="checkbox"/> Further documents are listed in the continuation of Box C.		<input checked="" type="checkbox"/> See patent family annex.
<p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the International filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the International filing date but later than the priority date claimed</p> <p>"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search	Date of mailing of the International search report	
6 May 2008	14/05/2008	
Name and mailing address of the ISA/ European Patent Office, P.B. 5018 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fax: (+31-70) 340-3016	Authorized officer Rydyger, Kay	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2008/053133

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0143074	A 14-06-2001	AU	2060701 A	18-06-2001
		CA	2362680 A1	14-06-2001
		EP	1159712 A1	05-12-2001
		JP	2003516702 T	13-05-2003
		US	6591019 B1	08-07-2003

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LS,MW,MZ,NA,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MT,NL,NO,PL,PT,RO,SE,SI,SK,T
R),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BR,BW,BY,
BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IS,JP,KE,K
G,KM,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PG,PH,PL,PT
,RO,RS,RU,SC,SD,SE,SG,SK,SL,SM,SV,SY,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,ZA,ZM,ZW