

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6340097号
(P6340097)

(45) 発行日 平成30年6月6日(2018.6.6)

(24) 登録日 平成30年5月18日(2018.5.18)

(51) Int. Cl.		F I	
G06F	9/315 (2006.01)	G06F	9/30 340D
G06F	9/32 (2006.01)	G06F	9/32 330A
G06F	9/38 (2006.01)	G06F	9/38 370A
G06F	17/16 (2006.01)	G06F	17/16 E
G06F	8/41 (2018.01)	G06F	9/44 322G

請求項の数 12 (全 31 頁)

(21) 出願番号	特願2017-7593 (P2017-7593)
(22) 出願日	平成29年1月19日(2017.1.19)
(62) 分割の表示	特願2015-531915 (P2015-531915) の分割
原出願日	平成25年6月12日(2013.6.12)
(65) 公開番号	特開2017-107579 (P2017-107579A)
(43) 公開日	平成29年6月15日(2017.6.15)
審査請求日	平成29年1月21日(2017.1.21)
(31) 優先権主張番号	13/630, 118
(32) 優先日	平成24年9月28日(2012.9.28)
(33) 優先権主張国	米国 (US)

(73) 特許権者	591003943 インテル・コーポレーション アメリカ合衆国 95054 カリフォル ニア州・サンタクララ・ミッション カレ ッジ ブレーバード・2200
(74) 代理人	110000877 龍華国際特許業務法人
(72) 発明者	プロトニコフ、ミカイル アメリカ合衆国 95054 カリフォル ニア州・サンタクララ・ミッション カレ ッジ ブレーバード・2200 インテル ・コーポレーション内

最終頁に続く

(54) 【発明の名称】 リードマスク及びライトマスクにより制御されるベクトル移動命令

(57) 【特許請求の範囲】

【請求項1】

第1のマスクレジスタ、第2のマスクレジスタ、第1のソースレジスタおよび第2のソースレジスタを特定するための複数のフィールドを有する命令をデコードするデコード回路と、

デコードされた前記命令を実行し、前記第1のソースレジスタにおける1または複数のターゲットデータエレメントを前記第2のソースレジスタにおける1または複数のソースデータエレメントで置き換える実行回路と、を備え、

各ソースデータエレメントの位置は、第2の値を有する前記第2のマスクレジスタ内のマスク値に対応し、各ターゲットデータエレメントは、第1の値を有する前記第1のマスクレジスタ内のマスク値に対応する

装置。

【請求項2】

前記第1の値を有する前記第1のマスクレジスタの前記マスク値の各々に対し、前記実行回路は、前記1または複数のソースデータエレメントのうちの1つとして、前記第2のマスクレジスタ内の前記第2の値の、対応するマスク値を有するデータエレメントを探索するために前記第2のソースレジスタを検索する

請求項1に記載の装置。

【請求項3】

前記実行回路は、前記第1のソースレジスタ内の1または複数のデータエレメントに対

して再帰計算を実行し、前記第 1 の ソースレジスタ内に前記再帰計算の複数の結果を蓄積する

請求項 1 または 2 に記載の装置。

【請求項 4】

前記実行回路は、前記第 1 の ソースレジスタ内の 1 または複数のデータエレメントがさらに計算を必要としなくなるまで、前記第 1 の ソースレジスタ内の 1 または複数のデータエレメントに対して再帰計算を実行し、前記第 1 の マスクレジスタを用いて、前記 1 または複数のデータエレメントが前記 1 または複数のターゲットデータエレメントであると示す

請求項 1 から 3 の何れか一項に記載の装置。

10

【請求項 5】

前記第 1 の 値は前記第 2 の 値の逆である

請求項 1 から 4 の何れか一項に記載の装置。

【請求項 6】

前記第 1 の 値は前記第 2 の 値と同一である

請求項 1 から 4 のいずれか一項に記載の装置。

【請求項 7】

第 1 のマスクレジスタ、第 2 のマスクレジスタ、第 1 のソースレジスタおよび第 2 のソースレジスタを特定するための複数のフィールドを有する命令を、デコード回路を用いてデコードする段階と、

20

デコードされた前記命令を、実行回路を用いて実行し、前記第 1 のソースレジスタにおける 1 または複数のターゲットデータエレメントを前記第 2 のソースレジスタにおける 1 または複数のソースデータエレメントで置き換える段階と、を備え、

各ソースデータエレメントの位置は、第 2 の値を有する前記第 2 のマスクレジスタ内のマスク値に対応し、各ターゲットデータエレメントは、第 1 の値を有する前記第 1 のマスクレジスタ内のマスク値に対応する

方法。

【請求項 8】

前記第 1 の 値を有する前記第 1 のマスクレジスタの前記マスク値の各々に対し、前記実行回路が前記 1 または複数のソースデータエレメントのうちの 1 つとして、前記第 2 のマスクレジスタ内の前記第 2 の値の、対応するマスク値を有するデータエレメントを探索するために前記第 2 のソースレジスタを検索する

30

請求項 7 に記載の方法。

【請求項 9】

前記実行回路は、前記第 1 の ソースレジスタ内の 1 または複数のデータエレメントに対して再帰計算を実行し、前記第 1 の ソースレジスタ内に前記再帰計算の複数の結果を蓄積する

請求項 7 または 8 に記載の方法。

【請求項 10】

前記実行回路は、前記第 1 の ソースレジスタ内の 1 または複数のデータエレメントがさらに計算を必要としなくなるまで、前記第 1 の ソースレジスタ内の 1 または複数のデータエレメントに対し再帰計算を実行し、前記第 1 の マスクレジスタを用いて、前記 1 または複数のデータエレメントが前記 1 または複数のターゲットデータエレメントであると示す

40

請求項 7 から 9 の何れか一項に記載の方法。

【請求項 11】

前記第 1 の 値は前記第 2 の 値の逆である

請求項 7 から 10 の何れか一項に記載の方法。

【請求項 12】

前記第 1 の 値は前記第 2 の 値と同一である

請求項 7 から 10 の何れか一項に記載の方法。

50

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、プロセッサまたは他のプロセッシングロジックにより実行されると、論理演算、数学的演算、または他の関数演算を実行する、プロセッシングロジック、マイクロプロセッサ及び関連する命令セットアーキテクチャの分野に関する。

【背景技術】

【0002】

命令セット、すなわち命令セットアーキテクチャ (ISA) は、プログラミングに関連するコンピューターアーキテクチャの一部であり、複数のネイティブデータタイプ、複数の命令、レジスタアーキテクチャ、複数のアドレッシングモード、メモリアーキテクチャ、割り込み及び例外処理、及び外部入出力 (I/O) を含む。命令という用語は、本明細書においては、マクロ命令をデコードするプロセッサのデコードの結果であるマイクロ命令またはマイクロオペレーション (マイクロ - ops) とは対照的に、実行すべく、プロセッサ (または、命令を、プロセッサにより処理される 1 または複数の他の命令へと、変換 (例えば、静的バイナリ変換、動的コンパイルを含む動的バイナリ変換を用いて)、モーフィング、エミュレート、または別の方法でコンバートする命令変換部) に提供される命令である、マクロ命令を概して指す。

【0003】

ISA は、命令セットを実装するプロセッサの内部設計であるマイクロアーキテクチャとは区別される。異なるマイクロアーキテクチャを有する複数のプロセッサは共通の命令セットを共有できる。例えば、Intel (登録商標) Core (商標) プロセッサ及びカリフォルニア州サンベールのアドバンスト・マイクロ・デバイセズの複数のプロセッサは、ほとんど同じバージョンの x86 命令セット (複数のより新しいバージョンと共に追加された幾つかの拡張を有する) を実装しているが、内部設計は異なる。例えば、ISA の同一のレジスタアーキテクチャは、複数の専用の物理的レジスタ、レジスタリネーミングメカニズムを用いる、1 または複数の動的に割り当てられた物理的レジスタ等を含むよく知られた技法を用いて、異なるマイクロアーキテクチャに異なる方法で実装されうる。

【0004】

多くの最新の ISA は単一命令、複数データ (SIMD) オペレーションをサポートする。たった 1 つまたは 2 つのデータエレメントに対して実行するスカラ命令ではなく、ベクトル命令 (パックドデータ命令または SIMD 命令とも称される) は、同時に、またはパラレルに、複数のデータエレメントに、または、データエレメントの複数のペアに対して実行しうる。プロセッサは、ベクトル命令に回答するパラレルな実行ハードウェアを有し得て、同時に、またはパラレルに複数のオペレーションを実行する。SIMD オペレーションは、1 つのオペレーションにおいて、1 つのベクトルレジスタまたはメモリ位置内にパックされた複数のデータエレメントに対して実行する。これらのデータエレメントは、パックドデータまたはベクトルデータと称される。ベクトルエレメントのそれぞれは、他とは別個に独立して扱われうる、別個で独立した 1 つのデータ (例えば、ピクセルの色、など) を表しうる。

【0005】

幾つかのシナリオにおいて、SIMD オペレーションは、独立したベクトルデータエレメントに対して再帰的に実行し得て、その際データエレメントが異なればイタレーション回数は異なる。したがって、幾つかのデータエレメントに対する計算は、その他のデータエレメントがさらに多くのイタレーションを必要とする一方で、終了しうる。再帰計算の 1 つの例が、WHILE ループオペレーションである。この例において、N エレメントのデータアレイ $X[i]$ ($i = 0, \dots, N - 1$) は、条件 ($X[i]$) が真 (満たされる) である間、再帰計算が施される。 $X[i]$ のこの計算は、条件 ($X[i]$) が偽となる場合、終了する。その状況の一例は、 $X[i] > 0$ でありうる。

10

20

30

40

50

【数 1】

```

for (i=0; i<N; i++){
    while (condition(X[i])){
        X[i]=computation(X[i]);}
}

```

【0006】

上記計算は、X[i]の異なる複数のデータエレメントでWHILEループのイタレーション回数が異なれば、容易にベクトル化され得ない。1つのアプローチとしては、プロセッサが、その条件を満たさないそれらのエレメントに対して計算を実行し、それから、それらのエレメントから得られた結果を捨てるというアプローチが考えられる。しかしながら、プロセッサがこれらのエレメントに対して不必要な計算を実行するだけでなく、これらのエレメントが占めるベクトルレジスタスロットを使用することもできないので、このアプローチは低効率である。

10

【図面の簡単な説明】

【0007】

複数の実施形態は例として示されるものであって、添付の複数の図面における複数の図に限定されるものではない。

【0008】

20

【図1】一実施形態による、複数のベクトルレジスタ及び複数のマスクレジスタを含む命令処理装置のブロック図である。

【図2】一実施形態による、レジスタアーキテクチャのブロック図である。

【図3】一実施形態による、ベクトル演算シーケンスの一例を示す。

【図4A】一実施形態による、プロセッサにベクトルレジスタ及びマスクレジスタ上で複数のオペレーションを実行させる複数の命令の疑似コードの一例を示す。

【図4B】一実施形態による、図4Aの複数の命令を使用するためのコードセグメントの一例を示す。

【図5A】一実施形態による、マスク更新命令及びベクトル移動命令を用いるコードセグメントにตอบสนองして実行される複数のオペレーションを例示するフロー図である。

30

【図5B】一実施形態による、マスク更新命令にตอบสนองして実行されるオペレーションを例示するフロー図である。

【図5C】一実施形態による、ベクトル移動命令にตอบสนองして実行されるオペレーションを例示するフロー図である。

【図6】一実施形態による、ソース命令セットのバイナリ命令をターゲット命令セットのバイナリ命令に変換するソフトウェア命令変換器の使用を例示するブロック図である。

【図7A】一実施形態による、インオーダー及びアウトオブオーダーパイプラインのブロック図である。

【図7B】一実施形態による、インオーダー及びアウトオブオーダーコアのブロック図である。

40

【図8A】一実施形態による、より具体的な例示的インオーダーコアアーキテクチャのブロック図である。

【図8B】一実施形態による、より具体的で例示的なインオーダーコアアーキテクチャのブロック図である。

【図9】一実施形態による、プロセッサのブロック図である。

【図10】一実施形態による、システムのブロック図である。

【図11】一実施形態による、第2のシステムのブロック図である。

【図12】本発明のある実施形態による、第3のシステムのブロック図である。

【図13】一実施形態による、システムオンチップ(SoC)のブロック図である。

【発明を実施するための形態】

50

【0009】

以下の説明において、数々の具体的な詳細を記載する。しかしながら、本発明の複数の実施形態は、これらの具体的な詳細なしに実現可能であることが理解される。他の例において、公知の回路、構造及び技術については、説明を理解する妨げにならないよう、詳細には示していない。

【0010】

本明細書にて説明される複数の実施形態は、複数の独立したデータエレメントに対する再帰的なベクトル演算の効率性を向上させる複数の命令を提供する。その複数の命令は、ベクトルレジスタのペア及びマスクレジスタのペアを用いて再帰的なベクトル演算を実行する。第1のベクトルレジスタは、ベクトル演算結果を蓄積するアキュムレータとして機能し、第2のベクトルレジスタは、第1のベクトルレジスタの複数の未使用スロット（未使用または終了した複数のデータエレメント位置）を埋める複数の新しいデータエレメントを提供する。マスクレジスタは、その対応する複数のベクトルレジスタにおけるどのデータエレメントがさらに計算を必要としているかを示すべく用いられる。

10

【0011】

一実施形態において、第1のベクトルレジスタ（すなわちアキュムレータ）は、レジスタが完全ベクトルで満たされるまで、複数の入力データエレメントを蓄積する。プロセッサはそれから非マスク（すなわちデンス（dense）な）ベクトル演算を用いて、これらのデータエレメントに対して計算を実行する。計算後、アキュムレータ中の幾つかの（計算が終了した）エレメントは、メモリまたは他のストレージ場所に送り戻され得て、他の（計算が終了していない）エレメントは、イタレーション回数を追加すべく、アキュムレータ中に保たれ得る。アキュムレータ中の、計算が終了したデータエレメント位置を、同一の再帰計算もまた必要とする複数の新しいデータエレメントが使用できる。

20

【0012】

RWMASKUPDATE及びSPARSEMOVという2つの命令を本明細書にて説明する。これらの命令は、多くのシナリオにおいて、ベクトル化の効率性を向上させる。例えば、一シナリオにおいて、入力データエレメントは、1または複数のスパース（sparse）なベクトルデータセットに由来する。そのベクトルデータセットのそれぞれは、アキュムレータ全体（すなわち、第1のベクトルレジスタ）を埋めるのに十分なエレメントを有していない。さらに、異なるデータセットからの入力データエレメントは、計算中、異なるイタレーション回数を必要とし得る。したがって、これ以上計算を必要としない、それらのデータエレメントからの未使用スロットがアキュムレータに残される。本明細書にて説明される命令は、これらの未使用スロットが複数の有用なエレメントで埋められるようにするので、完全ベクトルに対して再帰計算が可能となる。以下でさらに詳細に説明するように、SPARSEMOV命令は、第2のベクトルレジスタからアキュムレータへ有用なデータエレメント（すなわち、計算を必要とするデータエレメント）を移動させるベクトル移動命令である。RWMASKUPDATE命令は、（第2のベクトルレジスタに関連付けられる）リードマスクレジスタ及び（アキュムレータに関連付けられる）ライトマスクレジスタの両方を更新し、これら2つのベクトルレジスタにおける有用なデータエレメントの位置を特定する。

30

40

【0013】

SPARSEMOVと組み合わせてRWMASKUPDATEを使用すれば、再帰計算法において必要になる命令の総数を低減し、第2のベクトルレジスタにおける複数の有用なデータエレメント（すなわち、複数のソースデータエレメント）の数が、第1のベクトルレジスタにおける未使用スロット（すなわち、ターゲット位置）の数と一致しない、オーバーフロー及びアンダーフローの場合を単純化する。更新されたリード及びライトマスクは、2つのベクトルレジスタ間のデータの移動を制御すべく用いられる。特に、0のライトマスクビットは、アキュムレータ中のターゲット位置を特定すべく用いられ、1のリードマスクビットは、第2のベクトルレジスタ中の複数のソースデータエレメントを特定すべく用いられる。ターゲット位置を特定すべく、反転させたライトマスクビットを用い

50

れば、スパースカつ再帰的な計算のベクトル化におけるデータ蓄積を単純化する。

【 0 0 1 4 】

図 1 は、R W M A S K U P D A T E 命令及び S P A R S E M O V 命令を含む命令を実行すべく動作可能な回路を含む実行ユニット 1 4 0 を有する命令処理装置 1 1 5 の実施形態のブロック図である。いくつかの実施形態では、命令処理装置 1 1 5 は、電子システムの、プロセッサ、マルチコアプロセッサのプロセッサコア、またはプロセッシングエレメントでありうる。

【 0 0 1 5 】

デコーダ 1 3 0 は、高レベルの機械語命令またはマクロ命令の形態の、入力命令を受信し、それらをデコードし、元の高レベルの命令を反映する、及び/または、元の高レベルの命令から得られる、低レベルのマイクロオペレーション、マイクロコードエントリーポイント、マイクロ命令、または、他の低レベルの命令または制御信号を生成する。低レベルの命令または制御信号は、低レベル（例えば、回路レベルまたはハードウェアレベル）のオペレーションを通して、高レベルの命令のオペレーションを実行しうる。デコーダ 1 3 0 は、様々な異なるメカニズムを用いて実装されうる。適切なメカニズムの複数の例は、限定されるものではないが、マイクロコード、ルックアップテーブル、ハードウェア実装、プログラム可能ロジックアレイ（ P L A ）、当技術分野において既知のデコーダを実装すべく用いられる他のメカニズム等を含む。

【 0 0 1 6 】

デコーダ 1 3 0 は、キャッシュ 1 1 0、メモリ 1 2 0、または他のソース用の、入力命令を受信しうる。デコードされた命令は、実行ユニット 1 4 0 へ送られる。実行ユニット 1 4 0 はデコーダ 1 3 0 から、受信した命令を反映する、または受信した命令から得られる、1 または複数の、マイクロオペレーション、マイクロコードエントリーポイント、マイクロ命令、他の命令、または他の制御信号を受信しうる。実行ユニット 1 4 0 は、レジスタファイル 1 7 0、キャッシュ 1 1 0、及び/またはメモリ 1 2 0 からのデータ入力を受信し、かつ、レジスタファイル 1 7 0、キャッシュ 1 1 0、及び/またはメモリ 1 2 0 へのデータ出力を生成する。

【 0 0 1 7 】

一実施形態において、レジスタファイル 1 7 0 は、レジスタとも称される、アーキテクチャレジスタを含む。別途特定しないか、または間違いなく明らかでない限り、複数のアーキテクチャレジスタ、レジスタファイル、及び複数のレジスタという文言は、ソフトウェア、及び/またはプログラマがアクセスできる（例えば、ソフトウェアビジブルな）複数のレジスタを、及び/またはオペランドを特定すべくマクロ命令が特定する複数のレジスタ、を指すべく本明細書中で用いられる。これらのレジスタは、与えられたマイクロアーキテクチャ上の他の複数の非アーキテクチャレジスタ（例えば、複数のテンポラリレジスタ、複数のリオーダーバッファ、複数のリタイアメントレジスタ等）とは対照的である。

【 0 0 1 8 】

説明を分かりにくくしないように、比較的簡易な命令処理装置 1 1 5 を示し、説明する。他の複数の実施形態においては、1 より多くの実行ユニットを有しうるということが理解されよう。例えば、装置 1 1 5 は、例えば、複数の演算ユニット、複数の演算ロジックユニット（ A L U ）、複数の整数ユニット、複数の浮動小数点ユニットなどの、複数の異なるタイプの実行ユニットを含み得る。命令処理装置またはプロセッサの、さらなる他の複数の実施形態は、複数のマルチコア、複数の論理プロセッサ、または複数の実行エンジンを有しうる。命令処理装置 1 1 5 の、幾つかの実施形態を、図 7 から図 1 3 に関して提供する。

【 0 0 1 9 】

一実施形態によると、レジスタファイル 1 7 0 は、1 セットのベクトルレジスタ 1 7 5 及び 1 セットのマスクレジスタ 1 8 5 を備え、その両方は、R W M A S K U P D A T E 命令及び S P A R S E M O V 命令の複数のオペランドをストアする。各ベクトルレジスタ 1 7 5 は、5 1 2 ビット幅、2 5 6 ビット幅、1 2 8 ビット幅であり得て、また、異なるベ

10

20

30

40

50

クトル幅が用いられる。各マスクレジスタ185は、幾つかのマスクビットを含み、その各マスクビットは、複数のベクトルレジスタ175のうちの1つのベクトルレジスタの1つのデータエレメントに対応している。各マスクビットは、ベクトルレジスタのデータエレメントをマスクすべく用いられるので、64ビットのマスクレジスタは、512ビットのレジスタの64個の8ビットデータエレメントをマスクすべく用いられ得る。異なる幅（例えば、256ビットまたは128ビット）、及び異なるサイズ（例えば、16ビット、32ビットまたは64ビット）の複数のデータエレメントを有するベクトルレジスタに対して、異なる数のマスクビットをベクトル演算に関連して用いてもよい。

【0020】

図2は、本明細書において説明される複数の命令をサポートする、根本的なレジスタアーキテクチャ200のある実施形態を示す。レジスタアーキテクチャ200は、x86、MMX（商標）、ストリーミングSIMD拡張（SSE）命令、SSE2命令、SSE3命令、SSE4.1命令、及びSSE4.2命令を含む命令セットも、アドバンスドベクトル拡張命令（AVX）（AVX1及びAVX2）を指すSIMD拡張命令の追加的なセットも実装する、Intel（登録商標）Core（商標）プロセッサに基づいている、しかしながら、異なるレジスタ長、異なるレジスタタイプ、及び/またはレジスタの異なる個数をサポートする、異なるレジスタアーキテクチャもまた用いることができるということが理解されよう。

【0021】

図示する実施形態において、512ビット幅の、32個のベクトルレジスタ210がある。これらのレジスタはzmm0からzmm31と参照される。下側の16個のzmmレジスタの下位256ビットは、レジスタymm0-16に重ねられている。下側の16個のzmmレジスタの下位128ビット（ymmレジスタの下位128ビット）は、レジスタxmm0-15に重ねられている。図示する実施形態において、8個のマスクレジスタ220（k0からk7）があり、それぞれ64ビットの長さがある。代替的な実施形態においては、マスクレジスタ220は16ビット幅である。

【0022】

図示する実施形態において、レジスタアーキテクチャ200は、16個の64ビット汎用（GP）レジスタ230をさらに備える。ある実施形態において、それらは、現存するx86アドレッシングモードと共に用いられ、メモリオペランドをアドレッシングする。その実施形態は、RFLAGSレジスタ260、RIPレジスタ270及びMXCSRレジスタ280も図示している。

【0023】

その実施形態は、スカラー浮動小数点（FP）スタックレジスタファイル（x87スタック）240も図示し、そのファイル上で、MMXパックド整数フラットレジスタファイル250がエイリアスされる。図示する実施形態において、x87スタックは、x87命令セットの拡張を使用して、32/64/80ビットの浮動小数点データ上で、複数のスカラー浮動小数点オペレーションを実行すべく用いられる8個のエレメントのスタックである。一方、MMXレジスタは、MMXとxmmレジスタとの間で実行される幾つかのオペレーションに対する複数のオペランドを保持すべく、また同様に、64ビットのパックド整数データ上で複数のオペレーションを実行すべく用いられる。

【0024】

本発明の複数の代替的实施形態は、より幅の広い、または狭いレジスタを使用しうる。

【0025】

追加的に、本発明の複数の代替的实施形態は、よりたくさんの、より少ない、または異なる、レジスタファイル及びレジスタを使用しうる。

【0026】

図3は、独立したデータエレメントに対する計算を効率的にベクトル化すべく、プロセッサ（例えば、命令処理装置115）が実行するオペレーションの一例を図示している。説明を単純化するべく、この例における各ベクトルレジスタは、8つのデータエレメント

10

20

30

40

50

のみを有するように示される。複数の代替的实施形態においては、複数のベクトルレジスタ内に、異なる数のデータエレメントを有してよい。ベクトルレジスタは、128ビット幅、256ビット幅、または512ビット幅（例えば、図2のxmm、ymm、またはzmmレジスタ）であることが可能であり、または異なる幅が用いられうる。各ベクトルレジスタには8つのデータエレメントがあるので、各ベクトルレジスタに関連して8つのマスクビットのみが用いられる。

【0027】

この例において、ベクトルレジスタV1はアキュムレータとして用いられ、ベクトルレジスタV2は、新しいデータエレメントをV1に提供すべく用いられる。マスクレジスタK1（ライトマスク）及びK2（リードマスク）はそれぞれ、V1及びV2のデータエレメントをマスクすべく用いられる。この例において、0のマスクビットは、対応するデータエレメントが計算からマスクされる（すなわち、さらなる計算は必要ない）ことを示し、1のマスクビットは、対応するデータエレメントがさらなる計算を必要とすることを示す。代替的实施形態においては、マスクビット値の意味は逆であり得て、例えば、1のマスクビットは、対応するデータエレメントがさらなる計算を必要としないことを示すべく用いられ、0のマスクビットは、対応するデータエレメントがさらなる計算を必要とすることを示すべく用いられうる。

【0028】

初めに、アキュムレータV1は、入力ベクトル：A及びBとして2セットのデータをストアし、そのそれぞれは、スパースなデータアレイの一部でありうると仮定する。A_j及びB_jの下付きの文字jは、データエレメントが受けたイタレーション回数を示す。例えば、A₀は、あらゆるイタレーションの前のエレメントAであり、A₁は第1のイタレーション310後のエレメントである。説明を単純化すべく、同一イタレーション内の同一データセットからの異なるデータエレメントは同一の識別子を有するように示される。例えば、入力ベクトルの、位置0のA₀及び位置2のA₀は、2つの異なるエレメントであり、同一または異なる値を有し得て、入力ベクトルの位置1のB₀及び位置3のB₀は、2つの異なるエレメントであり、同一または異なる値を取りうる。マスクレジスタK1の複数のマスクビットの初期値は全て1であり、V1の初期の入力ベクトルが完全ベクトルであり、かつV1の全エレメントがベクトル演算の第1のイタレーション310に關与できることを示している。

【0029】

この例において、各イタレーションは、再帰的なベクトル演算が実行されるWHILEループのイタレーションを表している。第1のイタレーション310の後、アキュムレータV1は複数のA及び複数のBからなる1セットを含み、ここで、下付き文字は、これらのエレメントが第1のイタレーションを終了したことを示す。AのエレメントはWHILEループの1回のイタレーションを、Bのエレメントは2回のイタレーションを必要とすると仮定しよう。そうすると、WHILEループの1回のイタレーション後、Bエレメントに対してはもう1回イタレーションが必要な一方、Aエレメントに対する計算は終了した。この時点で、複数のAエレメントのそれぞれに対する条件は偽であり（なぜなら、さらに計算するための条件を満たさないから）、複数のBエレメントのそれぞれに対する条件は真である（なぜなら、さらに計算するための条件を満たすから）。したがって、複数のAに対応するそれらのマスクビットに対し、K1の複数のマスクビット値は0に設定され、複数のBに対応するそれらのマスクビットに対しては1が設定される。

【0030】

一実施形態において、0のマスクビットは、対応するエレメント位置における結果が、ベクトルレジスタ全体（この場合、V1）に対するベクトル演算後に捨てられるであろうことを示す。複数の代替的实施形態において、0のマスクビットは、対応するエレメント位置に対する計算が以後実行されず、故にそのエレメント位置は使用されないことを示す。いずれのシナリオにおいても、アキュムレータV1に複数のA₁を保持しておくこと、ベクトルリソースの無駄使いであるし、ベクトル演算の効率性を低減する。よって、本発明

10

20

30

40

50

の一実施形態によると、第2のベクトルレジスタV2が用いられ、複数のA1が残した未使用スロット（すなわち、データエレメント位置）を埋めるべく、V1に新しいデータエレメントを提供する。複数のA1のデータエレメントは、メモリ、キャッシュ、または他のデータストレージに保存可能である。

【0031】

図3の例において、ベクトルレジスタV2は、別のスパースなベクトルアレイの一部でありうるデータセットCのエレメントをストアする。「*」でマーク付けされたV2中の位置は、「ドントケア」を表し、再帰的なベクトル演算の目的のための有用なデータエレメントを含まないことを意味する。Cの各データエレメントがWHILEループの3回のイタレーションを受ける必要があると仮定しよう。Cのエレメントに代えて、または追加して、V2は、WHILEループ（及び、故にさらなる計算）の1または複数のイタレーションを受ける必要がある、A及び/またはB（例えば、複数のA₀、複数のB₀、及び/または複数のB）の複数の新しいデータエレメントを提供しうる。さらなる計算を必要とするV2中のこれらのデータエレメントは、「ソースデータエレメント」と称される。V2中のこれらのソースデータエレメントは、複数のA（「ターゲットデータエレメント」と称される）が残したV1中の未使用スロットを埋めることができる。説明を簡単にすべく、さらなる計算を必要とする、V1及び/またはV2中のデータエレメントを「有用なデータエレメント」と呼ぶ。したがって、併合操作320が、V1及びV2中の複数の有用なデータエレメントをマージすべく実行され、V2中の複数のソースデータエレメントが、複数のターゲットデータエレメントによって占有されるV1の位置に移され、再帰計算は、V1中の複数の追加された有用なデータエレメントで、第2のイタレーション330に進むことができるようになる。

【0032】

そのような併合操作においては、オーバーフロー、アンダーフロー、及び完全一致の、3つのシナリオが起こりうる。完全一致とは、V2中の有用なデータエレメントが、V1に残された未使用スロットの数と同一数あることを示す。したがって、完全一致においては、V2中の複数のソースデータエレメントの全ては、V1に残された未使用スロットへ移動（すなわち、入れ替え）する。結果として、V1は次のイタレーションを開始する完全ベクトルを有し、K1は更新され、全て1を含む。V2にはこれ以上ソースデータエレメントは残っていないので、K2は更新されて、全て0を含む。

【0033】

併合操作320は、新しいデータエレメント（C₀）の数が、K1中の0値のマスクビットの数（すなわちA1の数）よりも大きい、オーバーフローのシナリオを示している。したがって、V2中の新しいデータエレメントの全てがV1へ移動するわけではない。この例において、V2の位置7の、丸で囲ったC₀はV2中に残される。一方、位置2、4および6の、その他のC₀はV1へ移動した。この実施形態において、V2の下位のエレメントがV1へ移され、複数の代替の実施形態においては、V2の上位のエレメントがV1へ移されうる。併合操作320は、K1及びK2における、対応するマスクビットの更新もする。

【0034】

併合操作320後、V1は8つのエレメントからなる完全ベクトルを含み、第2のイタレーション330を開始し、V2は位置7に残された1つのC₀を有するのみである。この時点（併合操作320後）における、対応するマスクレジスタK1は全て1を含み、K2は位置7に、1の値を有する、たった1つのマスクビットを含む。

【0035】

第2のイタレーション330後、アキュムレータV1は複数のB₂及び複数のC₁の組み合わせを含む。複数のBエレメントに対する計算はこのイタレーション後に終了したので、それらのB₂はメモリ、キャッシュ、または他のデータストレージへ保存できる。したがって、複数のBエレメントのそれぞれに対する条件は偽（さらに計算するための条件を満たさないから）であり、複数のCエレメントのそれぞれに対する条件は真（さらに計

10

20

30

40

50

算するための条件を満たすから)である。したがって、K 1の複数のマスクビット値は、複数のB 2に対応するそれらのマスクビットに対して0が設定され、複数のC 1に対応するそれらのマスクビットに対して1が設定される。

【0036】

複数のB 2によって残された未使用スロットは、V 2の残りの複数のソースデータエレメントが埋めることができ、この場合、それはV 2の位置7のC 0である。しかしながら、C 0の数は、B 2の数より少ないので、後に続く併合操作340においてアンダーフローが起きる。図3に示すアンダーフローのシナリオにおいて、V 1における最下位のB 2がC 0で置き換えられ、複数の代替の実施形態においては、V 1における最上位のB 2がC 0で置き換えられる。その併合操作340は、K 1及びK 2における、対応するマスクビットの更新もする。

10

【0037】

併合操作340後、アキュムレータV 1は完全には埋まっておらず、V 2は、V 1へ移動できる有用なデータエレメントをこれ以上有していない。この時点(併合操作340後)におけるマスクレジスタK 1は、複数のCエレメントに対応する複数の位置に1を含み、K 2は全て0を含む。複数の有用なデータエレメントの全てが処理され、V 2にもはやソースデータエレメントが残らなくなるまで、V 1へ移動させる、追加の有用なデータエレメントを、V 2はロードし得るし、320及び/または340の併合操作を繰り返すことが可能である。この時点において、V 1の複数のエレメントの全てが、必要とされるイタレーション回数に達するまで、V 1は、追加の回数のイタレーションを受けうる。

20

【0038】

0または1のマスクビット値の意味は、図3の例において示されるものと逆であり得る、例えば、0のマスクビット値は条件が満たされることを意味すべく用いられ得て、1のマスクビット値は条件が満たされないことを意味すべく用いられ得る、ということが理解されよう。いくつかの実施形態では、K 1マスクビット値の意味は、K 2マスクビット値の意味とは逆であり得る、例えば、1のK 1マスクビット値は条件が満たされないことを意味すべく用いられ得て、1のK 2マスクビット値は条件が満たされることを意味すべく用いられ得る。したがって、各マスクレジスタにおける各マスクビットの意味が、一貫性のある解釈を可能にすべく、矛盾なく定義される限り、同一のシナリオに対して、図3の例において異なるマスクビット値を使用可能である。

30

【0039】

本発明の一実施形態によると、図3に関連して説明されるオペレーションは、RWMA SKUPDATE命令及びSPARSEMOV命令を含む複数のベクトル命令に応答して、プロセッサ(例えば、命令処理装置115)が実行する。SPARSEMOV命令は、条件を満たさない、V 1中の複数のターゲットエレメント(例えば、もはや計算が必要ない複数のエレメント)を置き換えながら、複数のソースデータエレメントをベクトルレジスタV 2からベクトルレジスタV 1へと移動させるべく用いられ得る。RWMA SKUPDATE命令は、マスクレジスタK 1及びマスクレジスタK 2の更新に用いることができ、それにより、それぞれ、条件を満たすV 1及びV 2における複数のデータエレメント(例えば、もっと計算を必要とする複数のエレメント)の複数の位置を特定する。一実施形態において、RWMA SKUPDATEは、2つのオペランドK 1及びK 2を有し、SPARSEMOVは4つのオペランドK 1、V 1、K 2及びV 2を有する。複数の代替の実施形態においては、RWMA SKUPDATE及び/またはSPARSEMOVの複数のオペランドのうちの幾つかは黙示的である。

40

【0040】

図4Aは、一実施形態による、RWMA SKUPDATE命令及びSPARSEMOV命令用の疑似コード401および402の一例を示す。疑似コード401および402において、KLは、各ベクトルレジスタ(例えば、V 1及びV 2のそれぞれ)における、複数のデータエレメントの総数であるベクトルの長さを表す。zmmレジスタを、8ビットのデータエレメントを有するアキュムレータとして用いるならば、 $KL = 512 / 8 = 6$

50

4である。疑似コード401はRWMSKUPDATE命令を、疑似コード402はSPARSEMOV命令を記述する。プロセッサは、疑似コード401および402に示されるものとは異なる複数のオペレーションまたはロジックを有するRWMSKUPDATE命令及びSPARSEMOV命令を実装しうることに留意されたい。

【0041】

RWMSKUPDATE命令及びSPARSEMOV命令はそれぞれ、複数のマスクレジスタを更新し、及び複数のデータエレメントを複数のベクトルレジスタ間で移動させる。これらの命令の結果を用いるべく、複数の追加的な命令を実行できて、それにより、再帰的なベクトル演算をより効率的に実行できる。図4Bは、一実施形態による、RWMSKUPDATE命令及びSPARSEMOV命令を用いるコードセグメント400の一例を示す。プロセッサが実行すると、コードセグメント400は、プロセッサに、アレ
10
レイXの複数の独立したデータエレメントに対して再帰的なベクトル演算を実行させる。アレ
レイXは、メモリ、キャッシュ、または他のデータのストア位置にストアされうる。コードセグメント400は、初期化部410、初期マージ部420、後続マージ部430、演算部440、及び残余部450を備える。410-450の各部におけるオペレーションを、プロセッサ(例えば、図1の命令処理装置115)が実行する方法500の実施形態を示す図5Aのフロー図を参照し後述する。

【0042】

初期化部410において、マスクレジスタK1及びマスクレジスタK2の両方は、それら
20
らに対応するベクトルレジスタV1及びV2において有用なデータエレメントはないとい
うことを示す、0に初期化される。「複数の有用なデータエレメント」という用語は、計
算を必要とする複数のデータエレメントを意味する。イタレーションは初期マージ部42
0から始まる。420において、まず、有用なデータエレメントがV2に残っていないか
どうか決定すべく、K2をチェックする(ブロック531)。V2に有用なデータがなけれ
ば、複数の入力データエレメントがアレ
30
レイXからV2へとロードされ(ブロック532)、K2におけるそれらの対応するマスクビットがしかるべく設定される。

【0043】

後続マージ部430は、V2が複数の有用なデータエレメントを含むシナリオを処理す
る。複数の有用なデータエレメントが前のオーバーフローからV2に残され得て、ブロッ
ク532においてV2へとロードされうる。SPARSEMOV命令431に
30
応答して、V2中のこれらの有用なデータエレメントは、K1及びK2におけるマスクビットに従い、V1へと移される(ブロック533)。

【0044】

ブロック533における移動の後、RWMSKUPDATE命令433に
30
応答して、マスクレジスタK1及びマスクレジスタK2は更新され、それぞれ、V1及びV2中の複数の有用なデータエレメントの現在の位置を特定する(ブロック534)。

【0045】

後続マージ部430において、第2のSPARSEMOV命令432が
40
実行され、V2からV1へと移された、アレ
レイX中の複数のデータエレメントのインデックス(位置)を
ストアし、計算結果がアレ
レイX中のそれらの元の位置にストアし戻され得るようになる。

【0046】

演算部440は、(対応するマスクが全て1であることによって示されるような、すな
わち、IsFullMask(K1)が真である場合の)完全ベクトルのベクトル演算を
処理する。V1が、有用なデータエレメントの完全ベクトルを有さず(ブロック535)
、V1に、ロードされなかった複数の入力データエレメントがあれば(ブロック538)
、複数の追加的な入力データエレメントが、V2を介してV1へロードされうる(ブロッ
ク532-534)ことを示す。V1が完全ベクトルを有さず、V1にロードされる入力
データエレメントがもはや無いならば(ブロック538)、計算が終了し、複数の結果が
アレ
50
レイXに保存し戻されるまで、V1中の複数の残存するデータエレメントが計算される、残余部450(ブロック539)に、複数のオペレーションが進むことを示す。

【 0 0 4 7 】

V 1 が複数の有用なデータエレメントからなる完全ベクトルを有するならば（ブロック 5 3 5 ）、V 1 についてベクトル演算が実行され得る（ブロック 5 3 6 ）。V 1 中のデータエレメントがどれもこれ以上計算を必要としないならば、マスクレジスタ K 1 は更新される。1 または複数のデータエレメントが、アレイ X に保存し戻される時点（ブロック 5 3 7 ）において、V 1 中のそれらのデータエレメントがこれ以上計算を必要としなくなる（K 1 中の対応する 0 値のマスクビットで示されるように）まで、ベクトル演算は続く。示すようなその実施形態において、複数のデータエレメントは S C A T T E R 命令と共に保存され得て、K 1 中の複数の 0 値のマスクビットが、関数 k n o t (K 1) を用いて特定され得る。R W M A S K U P D A T E 命令及び S P A R S E M O V 命令を除いては、S C A T T E R、k n o t、I s F u l l M a s k、などのコードセグメント 4 0 0 において用いる複数の特定の命令及び関数を、複数の代替的な命令シーケンスがエミュレートできる。

10

【 0 0 4 8 】

V 2 を通して V 1 にロードされる入力データエレメントがもはやなくなる（ブロック 5 3 8 ）、すなわち、アレイ X 中の入力データエレメントの全てが V 2 にロードされ、V 2 中の有用なデータエレメントの全てが V 1 に移されたとき、まで、ブロック 5 3 1 5 3 7 のオペレーションは繰り返される。これが、残余部 4 5 0 が始まるときである。この時点で、V 1 は複数の有用なデータエレメントから成る完全ベクトルを有し得ないが、V 1 中のそれらのデータエレメントはさらに計算を必要とする。ベクトル演算は、V 1 中の残存するデータエレメントの全てが、必要とされるイタレーション回数に達するまで続く（ブロック 5 3 9 ）。この時点において、V 1 における演算結果がアレイ X に保存し戻され得る（例えば、S C A T T E R 命令を用いて）（ブロック 5 3 9 ）。

20

【 0 0 4 9 】

図 5 B は、一実施形態による、R W M A S K U P D A T E 命令を実行する方法 5 1 0 のブロックフロー図である。方法 5 1 0 は、プロセッサ（例えば、図 1 の命令処理装置 1 1 5 ）の、第 1 のマスクレジスタ及び第 2 のマスクレジスタを特定するマスク更新命令の受信で始まる（ブロック 5 1 1 ）。プロセッサはマスク更新命令をデコードする（ブロック 5 1 2 ）。そのデコードされたマスク更新命令に応答して、プロセッサは、第 1 のマスクレジスタ内の与えられた数のマスクビットを、例えばこれらのマスクビットを第 1 のビット値（例えば、0 ）から第 2 のビット値（例えば、1 ）に設定することによって反転させる段階（ブロック 5 1 3 ）と、第 2 のマスクレジスタ内の与えられた数のマスクビットを、例えばこれらのマスクビットを第 2 のビット値（例えば、1 ）から第 1 のビット値（例えば、0 ）に設定することによって反転させる段階（ブロック 5 1 4 ）と、を含む複数のオペレーションを実行する。与えられた数とは、第 1 のビット値を有する第 1 のマスクレジスタ内のマスクビットの数、及び、第 2 のビット値を有する第 2 のマスクレジスタ内のマスクビットの数のうちの小さい方である。代替的实施形態において、第 1 のビット値は 1 であり得て、第 2 のビット値は 0 であり得る。

30

【 0 0 5 0 】

図 5 C は、一実施形態による、S P A R S E M O V 命令を実行する方法 5 2 0 のブロックフロー図である。方法 5 2 0 は、プロセッサ（例えば、図 1 の命令処理装置 1 1 5 ）の、第 1 のマスクレジスタ、第 2 のマスクレジスタ、第 1 のベクトルレジスタ、及び第 2 のベクトルレジスタを特定するベクトル移動命令の受信から始まる（ブロック 5 2 1 ）。プロセッサは、ベクトル移動オペレーションをデコードする（ブロック 5 2 2 ）。デコードされたベクトル移動命令に応答して、かつ第 1 のマスクレジスタ及び第 2 のマスクレジスタ内の複数のマスクビット値に基づき、プロセッサは、第 1 のベクトルレジスタ内の与えられた数の複数のターゲットデータエレメントを、第 2 のベクトルレジスタ内の与えられた数の複数のソースデータエレメントで置き換える（ブロック 5 2 3 ）。一実施形態において、各ソースデータエレメントは、第 2 のビット値（例えば、1 ）を有する第 2 のマスクレジスタ内のマスクビットに対応し、各ターゲットデータエレメントは、第 1 のビット

40

50

値（例えば、0）を有する第1のマスキレジスタ内のマスクビットに対応する。代替的实施形態においては、第1のビット値は1であり得て、第2のビット値は0であり得る。与えられた数とは、第1のビット値を有する第1のマスキレジスタ内のマスクビットの数、及び、第2のビット値を有する第2のマスキレジスタ内のマスクビットの数のうちの小さい方である。

【0051】

様々な実施形態において、図5A - 5Cの方法は、汎用プロセッサ、専用プロセッサ（例えば、グラフィクスプロセッサまたはデジタルシグナルプロセッサ）、または別のタイプのデジタル論理デバイスまたは命令処理装置が実行しうる。いくつかの実施形態では、図5A - 5Cの方法は、図1の命令処理装置115、または図7 - 13に示す実施形態などの、同様のプロセッサ、装置またはシステムが実行しうる。さらに、図7 - 13に示すプロセッサ、装置及びシステムと同様、図1の命令処理装置115は、図5A - 5Cの複数の方法の実施形態と同一の、同様の、または異なる、複数のオペレーション及び複数の方法の実施形態を実行しうる。

10

【0052】

いくつかの実施形態では、図1の命令処理装置115は、命令をソース命令セットからターゲット命令セットへと変換する命令変換部と共に動作しうる。例えば、命令変換部は、命令を、変換（例えば、静的バイナリ変換、動的コンパイルを含む動的バイナリ変換を用いて）、モーフィング、エミュレート、または別の方法でコンバートし、コアが処理する1または複数の他の命令にする。命令変換部は、ソフトウェア、ハードウェア、ファームウェアまたはそれらの組み合わせにおいて実装されうる。命令変換部は、プロセッサ上、プロセッサ外、または、一部はプロセッサ上かつ一部はプロセッサ外でありうる。

20

【0053】

図6は、本発明の複数の実施形態による、ソフトウェア命令変換器の使用を対比させたブロック図である。図示した実施形態においては、命令変換部はソフトウェア命令変換器であるが、代替的に命令変換部はソフトウェア、ファームウェア、ハードウェア、またはそれらの組み合わせにおいて実装されてよい。図6は、高級言語602で書かれたプログラムを、x86コンパイラ604を用いてコンパイルできて、少なくとも1つのx86命令セットコアを有するプロセッサ616がネイティブで実行しうる、x86バイナリコード606を生成することを示している。少なくとも1つのx86命令セットコアを有するプロセッサ616は、少なくとも1つのx86命令セットコアを有するインテル社製プロセッサと実質的に同一な結果を得るべく、（1）インテル社製x86命令セットコアの命令セットのかなりの部分、または（2）複数のアプリケーションのオブジェクトコードバージョン、または少なくとも1つのx86命令セットコアを有するインテル社製プロセッサ上で実行することを目的とされる他のソフトウェア、を互換性のある状態で実行、または処理することにより、少なくとも1つのx86命令セットコアを有するインテル社製プロセッサと実質的に同一な複数の機能を実行可能な任意のプロセッサを表わす。x86コンパイラ604は、追加的なリンケージ処理有り、または無しで、少なくとも1つのx86命令セットコアを有するプロセッサ616上で実行され得る、x86バイナリコード606（例えば、オブジェクトコード）を生成すべく動作可能なコンパイラを表わす。

30

40

【0054】

同様に図6は、高級言語602で書かれたプログラムは、代替的な命令セットのコンパイラ608を用いてコンパイルされ得て、少なくとも1つのx86命令セットコアを有さないプロセッサ614（例えば、カリフォルニア州サニーベールにあるミップス・テクノロジーのMIPS命令セットを実行する、及び/または、カリフォルニア州サニーベールにあるARMホールディングスのARM命令セットを実行する、複数のコアを有するプロセッサ）がネイティブで実行しうる代替的な命令セットのバイナリコード610を生成することを示す。命令変換部612は、x86バイナリコード606を、x86命令セットコアを有さないプロセッサ614がネイティブで実行し得るコードへと変換すべく用いられる。変換後のコードは、代替的な命令セットのバイナリコード610と同一である可

50

能性が低い。なぜなら、これができる命令変換部は作るのが困難だからである。しかしながら、変換後のコードは一般的なオペレーションを達成し、代替的な命令セットからの複数の命令から生成されるであろう。したがって、命令変換部 612 は、エミュレーション、シミュレーション、または任意の他の処理を通して、x86 命令セットプロセッサまたはコアを有さないプロセッサまたは他の電子デバイスが、x86 バイナリコード 606 を実行できるようにする、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組み合わせを表わす。 < 例示的なコアアーキテクチャ > < インオーダ及びアウトオブオーダコアブロック図 >

【0055】

図 7A は、本発明の複数の実施形態による、例示的なインオーダパイプライン、及び例示的なレジスタリネームアウトオブオーダ発行/実行パイプラインの両方を図示するブロック図である。図 7B は、本発明の複数の実施形態による、例示的な、インオーダアーキテクチャコアの実施形態、及びプロセッサに含まれる、例示的な、レジスタリネームアウトオブオーダ発行/実行アーキテクチャコアの両方を図示するブロック図である。図 7A 及び図 7B 内の複数の実線で囲まれたボックスは、インオーダパイプライン及びインオーダコアを示し、一方で複数の点線で囲まれたボックスの任意の追加は、レジスタリネームアウトオブオーダ発行/実行パイプライン及びコアを示す。インオーダ態様がアウトオブオーダ態様のサブセットであるとして、アウトオブオーダ態様を説明する。

【0056】

図 7A において、プロセッサパイプライン 700 はフェッチ段階 702、レンス (length) デコード段階 704、デコード段階 706、割り当て段階 708、リネーム段階 710、スケジュール段階 (ディスパッチまたは発行としても知られる) 712、レジスタリード/メモリリード段階 714、実行段階 716、ライトバック/メモリライト段階 718、例外処理段階 722、及びコミット段階 724 を備える。

【0057】

図 7B は、実行エンジンユニット 750 に連結されるフロントエンドユニット 730 を含むプロセッサコア 790 を示し、実行エンジンユニット 750 及びフロントエンドユニット 730 の両方はメモリアリユニット 770 に連結されている。コア 790 は、縮小命令セットコンピューティング (RISC) コア、複合命令セットコンピューティング (CISC) コア、超長命令語 (VLIW) コア、またはハイブリッドコアタイプもしくは代替的なコアタイプであってよい。さらに別のオプションとして、コア 790 は、例えば、ネットワークコアまたは通信コア、圧縮エンジン、コプロセッサコア、汎用コンピューティンググラフィックスプロセッシングユニット (GPGPU) コア、グラフィックスコア等の専用コアであってよい。

【0058】

フロントエンドユニット 730 は、命令キャッシュユニット 734 に連結される分岐予測ユニット 732 を備え、この命令キャッシュユニットは命令トランシェーションルックアサイドバッファ (TLB) 736 に連結され、この TLB は命令フェッチユニット 738 に連結され、この命令フェッチユニットはデコードユニット 740 に連結されている。デコードユニット 740 (またはデコーダ) は、命令をデコードし、その元の命令からデコードされるか、または元の命令を反映するか、または元の命令から得られる 1 または複数のマイクロオペレーション、マイクロコードエントリーポイント、マイクロ命令、他の命令、または他の制御信号を出力として生成しうる。デコードユニット 740 は様々な異なるメカニズムを用いて実装されうる。適切なメカニズムの複数例には、限定されるものではないが、ルックアップテーブル、ハードウェア実装、プログラム可能ロジックアレイ (PLA)、マイクロコード読み出し専用メモリ (ROM) 等が挙げられる。一実施形態において、コア 790 は、(例えば、デコードユニット 740 内で、またはフロントエンドユニット 730 内において) 複数の特定のマクロ命令用のマイクロコードをストアするマイクロコード ROM または他の媒体を含む。デコードユニット 740 は、実行エンジンユニット 750 内のリネーム/アロケータユニット 752 に連結されている。

10

20

30

40

50

【 0 0 5 9 】

実行エンジンユニット750は、リタイアメントユニット754、及び1セットの、1または複数のスケジューラユニット756に連結されるリネーム/アロケータユニット752を含む。スケジューラユニット(複数)756は、複数のリザベーションステーション、中央命令ウィンドウ等を含む任意の数の異なるスケジューラを表す。スケジューラユニット(複数)756は物理レジスタファイル(複数)ユニット(複数)758に連結されている。複数の、物理レジスタファイル(複数)ユニット758のそれぞれは、1または複数の物理レジスタファイルを表し、その複数の物理レジスタファイルのうちの異なるものは、スカラ整数、スカラ浮動小数点、パックド整数、パックド浮動小数点、ベクトル整数、ベクトル浮動小数点、ステータス(例えば、次に実行される命令のアドレスである命令ポインタ)等の、1または複数の異なるデータタイプをストアする。一実施形態において、物理レジスタファイル(複数)ユニット758は、ベクトルレジスタユニット、ライトマスクレジスタユニット、及びスカラレジスタユニットを備える。これらのレジスタユニットは、複数のアーキテクチャベクトルレジスタ、複数のベクトルマスクレジスタ、及び複数の汎用レジスタを提供しうる。物理レジスタファイル(複数)ユニット(複数)758は、(例えば、リオードバッファ(複数)及びリタイアメントレジスタファイル(複数)を用いて; フューチャファイル(複数)、履歴バッファ(複数)、及びリタイアメントレジスタファイル(複数)を用いて; レジスタマップ及びレジスタのプールを用いて; 等により)レジスタリネーミング及びアウトオブオーダー実行が実装されうるさまざまな方法を示すべく、リタイアメントユニット754によって重ね合わされる。リタイアメントユニット754及び物理レジスタファイル(複数)ユニット(複数)758は、実行クラスタ(複数)760に連結されている。実行クラスタ(複数)760は、1セットの、1または複数の実行ユニット762、及び、1セットの、1または複数のメモリアクセスユニット764を備える。複数の実行ユニット762は、様々なタイプのデータ(例えば、スカラ浮動小数点、パックド整数、パックド浮動小数点、ベクトル整数、ベクトル浮動小数点)に対し、様々な演算(例えば、シフト、加算、減算、乗算)を実行しうる。幾つかの実施形態が、特定の複数の機能または特定の複数セットの機能専用の幾つかの実行ユニットを含みうる一方で、他の複数の実施形態は、全機能を全て実行する1つの実行ユニットのみを含むか、または複数の実行ユニットを含む。スケジューラユニット(複数)756、物理レジスタファイル(複数)ユニット(複数)758、及び実行クラスタ(複数)760については、複数個ある可能性があるように示した。なぜなら、複数の特定の実施形態は、複数の特定のタイプのデータ/オペレーションに対して別個の複数のパイプライン(例えば、スカラ整数パイプライン、スカラ浮動小数点/パックド整数/パックド浮動小数点/ベクトル整数/ベクトル浮動小数点パイプライン、及び/または、それぞれが自身のスケジューラユニット、物理レジスタファイル(複数)ユニット及び/または実行クラスタを有するメモリアクセスパイプライン、-及び別個のメモリアクセスパイプラインの場合においては、このパイプラインの実行クラスタのみがメモリアクセスユニット(複数)764を有する複数の特定の実施形態が実装される)を作成するからである。複数の別個のパイプラインが使用されると、これらのパイプラインのうちの1または複数はアウトオブオーダー発行/実行であり、残りはインオーダー発行/実行でありうることも理解されるべきである。

【 0 0 6 0 】

1セットのメモリアクセスユニット764はメモリユニット770に連結され、このメモリユニットは、レベル2(L2)キャッシュユニット776に連結されるデータキャッシュユニット774に連結されるデータTLBユニット772を備える。1つの例示的な実施形態において、複数のメモリアクセスユニット764は、ロードユニット、ストアアドレスユニット、及びストアデータユニットを含んでよく、それぞれは、メモリユニット770内のデータTLBユニット772に連結されている。命令キャッシュユニット734は、メモリユニット770内のレベル2(L2)キャッシュユニット776にさらに連結されている。L2キャッシュユニット776は、1または複数の他のレベルのキャッシ

10

20

30

40

50

ュに連結されており、最終的にはメインメモリに連結されている。

【 0 0 6 1 】

例として、例示的なレジスタリネームアウトオブオーダ発行/実行コアアーキテクチャは、以下のようなパイプライン700を実装しうる：1) 命令フェッチ738が、フェッチ段階702及びレンスデコード段階704を実行する；2) デコードユニット740が、デコード段階706を実行する；3) リネーム/アロケータユニット752が、割り当て段階708及びリネーム段階710を実行する；4) スケジューラユニット(複数)756が、スケジュール段階712を実行する；5) 物理レジスタファイル(複数)ユニット(複数)758及びメモリユニット770が、レジスタリード/メモリリード段階714を実行する；実行クラスタ760が実行段階716を実行する；6) メモリユニット770及び物理レジスタファイル(複数)ユニット(複数)758が、ライトバック/メモリライト段階718を実行する；7) 様々なユニットが例外処理段階722に関与しうる；及び8) リタイアメントユニット754及び物理レジスタファイル(複数)ユニット(複数)758がコミット段階724を実行する。

10

【 0 0 6 2 】

コア790は、本明細書で説明する命令(複数)を含む、1または複数の命令セット(例えば、x86命令セット(より新しいバージョンと共に追加された幾つかの拡張を有する)；カルフォルニア州サニーベールにあるミップス・テクノロジーズのMIPS命令セット；カルフォルニア州サニーベールにあるARMホールディングスのARM命令セット(NEONなどの任意追加の拡張を有する))をサポートしうる。一実施形態において、コア790はパックドデータ命令セット拡張(例えば、SSE、AVX1、AVX2等)をサポートすべくロジックを備え、それにより、多数のマルチメディアアプリケーションが用いる複数のオペレーションを、パックドデータを使用して実行できるようになる。

20

【 0 0 6 3 】

コアはマルチスレッド化(2またはそれ以上のパラレルなセットのオペレーションまたはスレッドを実行する)をサポートし得て、タイムスライスマルチスレッド化、同時マルチスレッド化(物理コアが同時にマルチスレッド化している複数のスレッドのそれぞれに対して、単一の物理コアが論理コアを提供する)、またはそれらの組み合わせ(例えば、Intel(登録商標)のハイパースレッディングテクノロジー中などのタイムスライスフェッチ及びデコード、並びにその後の同時マルチスレッド化)を含む様々な方法でそれを行いうることが理解されるべきである。

30

【 0 0 6 4 】

レジスタリネーミングがアウトオブオーダ実行との関連で説明される一方で、レジスタリネーミングはインオーダアーキテクチャにおいて使用されることが理解されるべきである。図示したプロセッサの実施形態が、別個の命令、及びデータキャッシュユニット734/774及び共有のL2キャッシュユニット776を備える一方で、代替的实施形態は、例えば、レベル1(L1)の内部キャッシュ、または複数レベルの内部キャッシュ等の、複数の命令及びデータの両方に対する単一の内部キャッシュを有しうる。いくつかの実施形態では、システムは、内部キャッシュ、及び、コア及び/またはプロセッサの外部にある外部キャッシュの組み合わせを含む。代替的に、全てのキャッシュがコア及び/またはプロセッサの外部にあってよい。<具体的で例示的なインオーダコアアーキテクチャ>

40

【 0 0 6 5 】

図8A-8Bは、より具体的な例示的インオーダコアアーキテクチャのブロック図を示す。そのコアは、チップ内の幾つかの論理ブロック(同一タイプ及び/または異なるタイプの複数の他のコアを含む)のうちの1つでありうる。複数の論理ブロックは、アプリケーションに応じて、ある固定のファンクションロジック、メモリI/Oインターフェース、及び他の必要なI/Oロジックと、高帯域幅相互接続ネットワーク(例えば、リングネットワーク)を介して通信する。

【 0 0 6 6 】

50

図 8 A は、本発明の複数の実施形態による、自身のオンダイ相互接続ネットワーク 8 0 2 への接続、及び自身のレベル 2 (L 2) のキャッシュのローカルサブセット 8 0 4 を伴う、単一のプロセッサコアのブロック図である。一実施形態において、命令デコーダ 8 0 0 は、パケットデータ命令セット拡張を有する x 8 6 命令セットをサポートする。L 1 キャッシュ 8 0 6 は、スカラユニット及びベクトルユニットへの、キャッシュメモリへの低レイテンシアアクセスを可能にする。一実施形態 (設計を単純化するための) において、スカラユニット 8 0 8 及びベクトルユニット 8 1 0 は、複数の別個のレジスタセット (それぞれ、複数のスカラレジスタ 8 1 2 及び複数のベクトルレジスタ 8 1 4) を使用し、それらの間を伝送されるデータがメモリへ書き込まれ、その後レベル 1 (L 1) キャッシュ 8 0 6 からリードバックされる一方で、本発明の複数の代替的实施形態は異なるアプローチ (例えば、単一のレジスタセットを用いるか、または、書き込まれもリードバックもされずに、2つのレジスタファイル間をデータが伝送され得るようにする通信パスを含む) を使用しうる。

10

【 0 0 6 7 】

L 2 キャッシュのローカルサブセット 8 0 4 は、1つのプロセッサコアにつき1つの、複数の別個のローカルサブセットに分割されるグローバル L 2 キャッシュの一部である。各プロセッサコアは、L 2 キャッシュの、自身のローカルサブセット 8 0 4 へのダイレクトアクセスパスを有する。プロセッサコアがリードするデータは、自身の L 2 キャッシュのサブセット 8 0 4 へストアされ、他のプロセッサコアが自身のローカル L 2 キャッシュのサブセットにアクセスすると同時に、高速にアクセス可能である。プロセッサコアが書き込むデータは、自身の L 2 キャッシュのサブセット 8 0 4 にストアされ、必要ならば複数の他のサブセットからフラッシュされる。リングネットワークは共有データに対するコヒーレンスを保証する。リングネットワークは、複数のプロセッサコア、複数の L 2 キャッシュ、及び複数の他の論理ブロックが、チップ内で互いに通信できるようにする双方向性のネットワークである。各リングデータパスは1つの方向につき 1 0 1 2 ビット幅である。

20

【 0 0 6 8 】

図 8 B は、本発明の複数の実施形態による、図 8 A のプロセッサコアの一部の拡大図である。図 8 B は、ベクトルユニット 8 1 0 及びベクトルレジスタ 8 1 4 に関するより詳細な内容と共に、L 1 キャッシュ 8 0 6 の L 1 データキャッシュ 8 0 6 A 部分を備える。

30

【 0 0 6 9 】

具体的に、ベクトルユニット 8 1 0 は、16ビットのベクトル処理ユニット (V P U) であり (1 6 幅 A L U 8 2 8 を参照) 、整数命令のうちの1または複数、単精度浮動小数点命令、及び倍精度浮動小数点命令を実行する。V P U は、スイズルユニット 8 2 0 による複数のレジスタ入力のスウィズル、複数の数値変換ユニット 8 2 2 A B による数値変換、及び、メモリ入力に対する複製ユニット 8 2 4 による複製をサポートする。ライトマスクレジスタ 8 2 6 は複数の結果のベクトル書き込みのプレディケートを可能にする。
< 統合型メモリコントローラ及び統合型グラフィックスを有するプロセッサ >

【 0 0 7 0 】

図 9 は、本発明の複数の実施形態による、2つ以上のコア、統合型メモリコントローラ、及び統合型グラフィックスを有しうるプロセッサ 9 0 0 のブロック図である。図 9 中の複数の実線で囲まれたボックスは、単一のコア 9 0 2 A 、システムエージェント 9 1 0 、1セットの1または複数のバスコントローラユニット 9 1 6 を有するプロセッサ 9 0 0 を図示する。複数の点線で囲まれたボックスの任意の追加は、複数のコア 9 0 2 A - N 、システムエージェントユニット 9 1 0 内の1セットの1または複数の統合型メモリコントローラユニット 9 1 4 、及び専用ロジック 9 0 8 を有する代替的なプロセッサ 9 0 0 を図示する。

40

【 0 0 7 1 】

したがって、プロセッサ 9 0 0 の複数の異なる実装は、1) 統合型グラフィックス及び/または科学的 (スループット) ロジック (1 または複数のコアを含み得る) である専用

50

ロジック 908、及び 1 または複数の汎用コア（例えば、複数の汎用インオーダーコア、複数の汎用アウトオーダーコア、その 2 つの組み合わせ）である複数のコア 902A N を有する CPU； 2）主にグラフィクス及び/または科学（スループット）向けの多数の専用コアである複数のコア 902A N を有するコプロセッサ；及び、 3）多数の汎用インオーダーコアである複数のコア 902A N を有するコプロセッサを含みうる。したがって、プロセッサ 900 は、例えば、ネットワークまたは通信プロセッサ、圧縮エンジン、グラフィクスプロセッサ、GPGPU（汎用グラフィックスプロセッシングユニット）、高スループットメニーインテグレートッドコア（MIC）コプロセッサ（30 以上のコアを含む）、組み込みプロセッサ等の、汎用プロセッサ、コプロセッサ、または専用プロセッサでありうる。プロセッサは、1 または複数のチップ上に実装されうる。プロセッサ 900 は、例えば、BiCMOS、CMOS、または NMOS などの幾つかのプロセス技術のうちの一つの任意のものを用いる 1 または複数の基板の一部であり得て、及び/または、1 または複数の基板に実装されうる。

10

【0072】

メモリ階層は、複数のコア内の、1 または複数レベルのキャッシュ、1 セットの 1 または複数の共有キャッシュユニット 906、及び、1 セットの統合型メモリコントローラユニット 914 に連結された外部メモリ（図示せず）を備える。1 セットの共有キャッシュユニット 906 は、レベル 2（L2）、レベル 3（L3）、レベル 4（L4）、または他の複数レベルのキャッシュ等の 1 または複数の中間レベルのキャッシュ、ラストレベルキャッシュ（LLC）、及び/またはそれらの組み合わせを含みうる。一実施形態において、リングベースの相互接続ユニット 912 は、統合型グラフィックスロジック 908、1 セットの共有キャッシュユニット 906、及びシステムエージェントユニット 910 / 統合型メモリコントローラユニット（複数）914 を相互接続する一方で、複数の代替的实施形態は、そのような複数のユニットを相互接続する、任意の数の公知の技術を使用しうる。一実施形態において、コヒーレンスは、1 または複数のキャッシュユニット 906 と複数のコア 902A N との間で維持される。

20

【0073】

いくつかの実施形態では、コア 902A N のうちの 1 または複数は、マルチスレッド化することが可能である。システムエージェント 910 は、複数のコア 902A N を調整し、操作する複数の構成要素を備える。システムエージェントユニット 910 は、例えば、電力制御ユニット（PCU）及びディスプレイユニットを含みうる。PCU は、複数のコア 902A N 及び統合型グラフィックスロジック 908 の電力状態の調整に必要なロジック及び複数の構成要素でありうるか、またはそれらを含みうる。ディスプレイユニットは、1 または複数の外部接続のディスプレイの駆動用である。

30

【0074】

コア 902A N は、アーキテクチャ命令セットに関して同種または異種でありうる。すなわち、2 またはそれ以上のコア 902A N は同一の命令セットを実行可能であり得て、一方で、他のコアはその命令セットのサブセットまたは異なる命令セットのみを実行可能でありうる。〈例示的なコンピューターアーキテクチャ〉

【0075】

図 10 - 図 13 は、例示的なコンピューターアーキテクチャのブロック図である。ラップトップ、デスクトップ、ハンドヘルド型 PC、携帯情報端末（PDA）、エンジニアリングワークステーション、サーバ、ネットワークデバイス、ネットワークハブ、スイッチ、組み込みプロセッサ、デジタルシグナルプロセッサ（DSP）、グラフィックスデバイス、ビデオゲームデバイス、セットトップボックス、マイクロコントローラ、携帯電話、ポータブルメディアプレイヤー、ハンドヘルドデバイス、及び様々な他の電子デバイスの当技術分野において既知の複数の他のシステム設計及び構成も適している。一般的には、本明細書で開示するような、プロセッサ及び/または他の実行ロジックを組み込み可能な多様なシステムまたは電子デバイスが概して適している。

40

【0076】

50

次に図10を参照すると、本発明の一実施形態による、システム1000のブロック図が示される。システム1000は、コントローラハブ1020に連結される、1または複数のプロセッサ1010、1015を含みうる。一実施形態において、コントローラハブ1020は、グラフィックスメモリコントローラハブ(GMCH)1090及び入出力ハブ(IOH)1050(複数の別個のチップ上にあってもよい)を含み、GMCH1090は、メモリ1040及びコプロセッサ1045に連結されるメモリ及びグラフィックスコントローラを含み、IOH1050は、入出力(I/O)デバイス1060をGMCH1090に連結する。代替的に、メモリ及びグラフィックスコントローラのうちの1つまたは両方は、(本明細書で説明するように)プロセッサ内に統合されており、メモリ1040及びコプロセッサ1045は、プロセッサ1010と、IOH1050を有する単一チップ内のコントローラハブ1020とに直接連結されている。

10

【0077】

追加のプロセッサ1015の、任意選択される特性が図10に破線で示される。各プロセッサ1010、1015は、本明細書で説明される、複数のプロセッサコアのうちの1または複数を含み得て、プロセッサ900のあるバージョンでありうる。

【0078】

メモリ1040は、例えば、ダイナミックランダムアクセスメモリ(DRAM)、相変化メモリ(PCM)、またはその2つの組み合わせでありうる。少なくとも1つの実施形態に対して、コントローラハブ1020は、フロントサイドバス(FSB)などのマルチドロップバス、クイックパスインターコネクト(QPI)などのポイントツーポイントインターフェース、または同様の接続1095を介してプロセッサ(複数)1010、1015と通信する。

20

【0079】

一実施形態において、コプロセッサ1045は、例えば、高スループットMICプロセッサ、ネットワークまたは通信プロセッサ、圧縮エンジン、グラフィックスプロセッサ、GPGPU、組み込みプロセッサ等の専用プロセッサである。一実施形態において、コントローラハブ1020は、統合型グラフィックスアクセラレータを含みうる。

【0080】

アーキテクチャ特性、マイクロアーキテクチャ特性、熱特性、電力消費特性等を含む、メリットのメトリクスの範囲に関して、複数の物理リソース1010、1015間の違いは様々あり得る。

30

【0081】

一実施形態において、プロセッサ1010は一般的なタイプの、データプロセッシングオペレーションを制御する複数の命令を実行する。複数のコプロセッサ命令がその複数の命令内に組み込まれうる。プロセッサ1010はこれらのコプロセッサ命令を、接続されたコプロセッサ1045によって実行されるべきタイプのものであると認識する。従って、プロセッサ1010はコプロセッサ1045に対して、これらのコプロセッサ命令(または、複数のコプロセッサ命令を表わす複数の制御信号)をコプロセッサバスまたは他の相互接続上に発行する。コプロセッサ(複数)1045は複数の受信したコプロセッサ命令を受け付け、実行する。

40

【0082】

次に図11を参照すると、本発明の実施形態による、第1のより具体的な例示的システム1100のブロック図が示される。図11に示されるように、マルチプロセッサシステム1100は、ポイントツーポイント相互接続システムであり、ポイントツーポイント相互接続1150を介して連結される、第1のプロセッサ1170及び第2のプロセッサ1180を備える。プロセッサ1170および1180のそれぞれは、プロセッサ900のあるバージョンでありうる。本発明の一実施形態において、プロセッサ1170および1180はそれぞれ、プロセッサ1010および1015であり、一方コプロセッサ1138はコプロセッサ1045である。別の実施形態では、プロセッサ1170および1180はそれぞれ、プロセッサ1010及びコプロセッサ1045である。

50

【0083】

プロセッサ1170および1180はそれぞれ、統合型メモリコントローラ（IMC）ユニット1172および1182を備えるように示されている。プロセッサ1170は、自身の複数のバスコントローラユニットの一部として、ポイントツーポイント（P-P）インターフェース1176および1178も備え、同様に、第2のプロセッサ1180はP-Pインターフェース1186および1188を備える。プロセッサ1170、1180は、ポイントツーポイント（P-P）インターフェース回路1178、1188を用いて、P-Pインターフェース1150を介して情報を交換しうる。図11に示されるように、IMC1172および1182は、複数のプロセッサを、各メモリ、すなわち、各プロセッサにローカルに接続されたメインメモリの複数部分でありうる、メモリ1132及びメモリ1134に連結する。

10

【0084】

プロセッサ1170、1180はそれぞれ、ポイントツーポイントインターフェース回路1176、1194、1186、1198を用いて、個々のP-Pインターフェース1152、1154を介してチップセット1190と各情報を交換しうる。チップセット1190は、高性能インターフェース1139を介してコプロセッサ1138と任意選択的に情報を交換しうる。一実施形態において、コプロセッサ1138は、例えば、高スループットMICプロセッサ、ネットワークまたは通信プロセッサ、圧縮エンジン、グラフィクスプロセッサ、GPGPU、組込みプロセッサ等の専用プロセッサである。

20

【0085】

共有キャッシュ（図示せず）は、いずれかのプロセッサ内、または両方のプロセッサの外部に含まれ得て、さらにP-P相互接続を介してその複数のプロセッサに接続され得て、プロセッサが低電力モードになった場合に、いずれかの、または両方のプロセッサのローカルキャッシュ情報が共有キャッシュにストアされるようにする。

【0086】

チップセット1190は、インターフェース1196を介して第1のバス1116に連結しうる。一実施形態において、第1のバス1116は、ペリフェラルコンポーネントインターコネクタ（PCI）バスであってもよいし、またはPCIエクスプレスバスまたは別の第3世代I/O相互接続バス等のバスであってもよいが、本発明の範囲はそのように限定されない。

30

【0087】

図11に示されるように、第1のバス1116を第2のバス1120に連結するバスブリッジ1118と共に、様々なI/Oデバイス1114を、第1のバス1116に連結しうる。一実施形態においては、コプロセッサ、高スループットMICプロセッサ、GPGPU、アクセラレータ（例えば、グラフィックスアクセラレータまたはデジタル信号処理（DSP）ユニットなどの）、フィールドプログラマブルゲートアレイ、または任意の他のプロセッサ等の1または複数の追加のプロセッサ1115を第1のバス1116に連結する。一実施形態において、第2のバス1120はローピンカウント（LPC）バスでありうる。

40

【0088】

一実施形態において、様々なデバイスを、例えば、キーボード及び/またはマウス1122、通信デバイス1127、及び、複数の命令/コード及びデータ1130を含みうるディスクドライブまたは他の大容量ストレージデバイス等のストレージユニット1128を含む第2のバス1120に連結しうる。さらに、オーディオI/O1124を、第2のバス1120に連結してよい。他のアーキテクチャが可能であることに留意されたい。例えば、システムは、図11のポイントツーポイントアーキテクチャではなく、マルチドロップバスまたは他のそのようなアーキテクチャを実装しうる。

【0089】

次に図12を参照すると、本発明の実施形態による、第2のより具体的な例示的システム1200のブロック図が示される。図11及び図12中の同様のエレメントには同様の

50

参照番号を付し、図 1 1 中の複数の特定の態様を図 1 2 から省き、図 1 2 中の複数の他の態様を分かりにくくしないようにした。

【 0 0 9 0 】

図 1 2 は、プロセッサ 1 1 7 0、1 1 8 0 がそれぞれ、統合されたメモリ及び I / O 制御ロジック (「 C L 」) 1 1 7 2 及び 1 1 8 2 を備えうることを示している。したがって、C L 1 1 7 2、1 1 8 2 は、複数の統合型メモリコントローラユニットを備え、I / O 制御ロジックを備える。図 1 2 は、C L 1 1 7 2、1 1 8 2 に連結されるのはメモリ 1 1 3 2、1 1 3 4 だけでなく、I O デバイス 1 2 1 4 も制御ロジック 1 1 7 2、1 1 8 2 に連結されることを示している。レガシー I / O デバイス 1 2 1 5 はチップセット 1 1 9 0 に連結される。

10

【 0 0 9 1 】

次に図 1 3 を参照すると、本発明の実施形態による、S o C 1 3 0 0 のブロック図が示される。図 9 中の同様のエレメントには同様の参照番号が付してある。また、点線で囲まれたボックスは、より高度な S o C 上の任意選択の機能である。図 1 3 において、相互接続ユニット (複数) 1 3 0 2 は、1 セットの 1 または複数のコア 2 0 2 A N 及び共有キャッシュユニット (複数) 9 0 6 を備えるアプリケーションプロセッサ 1 3 1 0 ; システムエージェントユニット 9 1 0 ; バスコントローラユニット (複数) 9 1 6 ; 統合型メモリコントローラユニット (複数) 9 1 4 ; 統合型グラフィックスロジック、画像プロセッサ、オーディオプロセッサ及びビデオプロセッサを含みうる 1 セットの 1 または複数のコプロセッサ 1 3 2 0 ; スタティックランダムアクセスメモリ (S R A M) ユニット 1 3 3 0 ; ダイレクトメモリアクセス (D M A) ユニット 1 3 3 2 ; 及び 1 または複数の外部のディスプレイに連結するためのディスプレイユニット 1 3 4 0 に連結される。一実施形態において、コプロセッサ (複数) 1 3 2 0 は、例えば、ネットワークまたは通信プロセッサ、圧縮エンジン、G P G P U、高スループット M I C プロセッサ、組込みプロセッサ等の専用プロセッサを備える。

20

【 0 0 9 2 】

本明細書にて開示される複数のメカニズムの複数の実施形態は、ハードウェア、ソフトウェア、ファームウェア、またはそのような実装手法の組み合わせにおいて実装されうる。本発明の複数の実施形態は、少なくとも 1 つのプロセッサ、ストレージシステム (揮発及び不揮発性メモリ、及び / またはストレージエレメントを含む)、少なくとも 1 つの入力デバイス、及び少なくとも 1 つの出力デバイスを備えるプログラマブルシステム上で実行するコンピュータプログラムまたはプログラムコードとして実装されうる。

30

【 0 0 9 3 】

図 1 1 に図示するコード 1 1 3 0 などのプログラムコードは、入力命令に適用され、本明細書で説明される機能を実行し、出力情報を生成する。出力情報は、既知の方法で 1 または複数の出力デバイスに適用されうる。こうした適用のために、プロセッシングシステムは、例えば、デジタルシグナルプロセッサ (D S P)、マイクロコントローラ、特定用途向け集積回路 (A S I C)、またはマイクロプロセッサなどのプロセッサを有する任意のシステムを備える。

【 0 0 9 4 】

プログラムコードは、高級手続き型プログラミング言語またはオブジェクト指向型プログラミング言語で実装され、プロセッシングシステムと通信する。プログラムコードは、望むならば、アセンブリ言語又は機械言語で実装されてもよい。実際には、本明細書で説明される複数のメカニズムは、何れ特定のプログラミング言語へも範囲を限定されるものではない。いずれの場合であっても、言語はコンパイラ型言語またはインタプリタ型言語でありうる。

40

【 0 0 9 5 】

少なくとも 1 つの実施形態の 1 または複数の態様は、機械がリードすると、機械に、本明細書で説明した複数の技術を実行するロジックを組み立てさせるプロセッサ内の様々なロジックを表わす、機械可読媒体上にストアされた複数の代表的な命令により実装されう

50

る。「IPコア」として知られるそのような表現は、有形のマシン可読媒体上にストアされ、様々な顧客または製造施設に供給されて、実際にロジックまたはプロセッサを作成する複数の製造機械にロードされうる。

【0096】

そのような機械可読記憶媒体は、限定はしないが、ハードディスク、フロッピー（登録商標）ディスク、光ディスク、コンパクトディスクリードオンリメモリ（CD-ROM）、コンパクトディスクリライタブル（CD-RW）、及び光磁気ディスクを含む任意の他のタイプのディスク、リードオンリメモリ（ROM）、ダイナミックランダムアクセスメモリ（DRAM）、スタティックランダムアクセスメモリ（SRAM）等のランダムアクセスメモリ（RAM）、消去可能プログラマブルリードオンリメモリ（EPROM）、フラッシュメモリ、電氣的消去可能プログラマブルリードオンリメモリ（EEPROM）、相変化メモリ（PCM）などの半導体デバイス、磁気カードまたは光カード、または電子命令をストアするのに適した任意の他のタイプの媒体等のストレージ媒体を含む、機械またはデバイスによって製造または形成される物品の、非一時的で有形な複数の構成を含みうる。

10

【0097】

従って、本発明の複数の実施形態は、複数の命令を含むか、または、本明細書にて説明する複数の構造、複数の回路、複数の装置、複数のプロセッサ、及び/または複数のシステムの機能を定義するハードウェア記述言語（HDL）等の設計データを含む、非一時的で有形な機械可読媒体も含む。そのような複数の実施形態はプログラム製品と称されることもある。

20

【0098】

複数の特定の例示的な実施形態が添付の複数の図面において説明され、示されてきたが、複数のそのような実施形態は、広義の発明の例示に過ぎず、それを限定するものではなく、かつ、本発明は、示され、説明される具体的な複数の構造及び複数の構成に限定されるものではないことが理解されるべきである。なぜなら、本開示を研究すれば、当業者は様々な他の変更形態を考え付き得るからである。成長が早く、さらなる進歩を容易には予測し得ないこのような技術の領域においては、複数の開示した実施形態は、本開示の原理、または添付の特許請求の範囲から逸脱することなく、技術的進歩を可能にすることによって容易になるような、構成及び詳細な内容を、たやすく変更可能でありうる。本明細書によれば、以下の各項目に記載の事項もまた開示される。

30

[項目1]

第1のマスキレジスタ、第2のマスキレジスタ、第1のベクトルレジスタ、及び第2のベクトルレジスタを含むレジスタファイルと、

前記レジスタファイルに連結される実行回路と、を備え、

前記実行回路は、命令を実行し、前記第1のベクトルレジスタ内の与えられた数の複数のターゲットデータエレメントを、前記第2のベクトルレジスタ内の前記与えられた数の複数のソースデータエレメントで置き換え、

前記複数のソースデータエレメントの各々は、第2のビット値を有する前記第2のマスキレジスタ内のマスクビットに対応し、前記複数のターゲットデータエレメントの各々は、第1のビット値を有する前記第1のマスキレジスタ内のマスクビットに対応する装置。

40

[項目2]

前記第1のビット値を有する前記第1のマスキレジスタの前記マスクビットの各々に対し、前記実行回路は、前記複数のソースデータエレメントのうちの1つとして、前記第2のマスキレジスタ内の前記第2のビット値の、対応するマスクビットを有するデータエレメントを探索するために前記第2のベクトルレジスタを検索する

項目1に記載の装置。

[項目3]

前記実行回路は、前記第1のベクトルレジスタ内の複数のデータエレメントに対して再

50

帰計算を実行し、前記第 1 のベクトルレジスタ内に前記再帰計算の複数の結果を蓄積する項目 1 または 2 に記載の装置。

[項目 4]

前記実行回路は、前記第 1 のベクトルレジスタ内の 1 または複数のデータエレメントがさらに計算を必要としなくなるまで、前記第 1 のベクトルレジスタ内の複数のデータエレメントに対して再帰計算を実行し、前記第 1 のマスクレジスタを用いて、前記 1 または複数のデータエレメントが前記複数のターゲットデータエレメントであると示す

項目 1 から 3 の何れか一項に記載の装置。

[項目 5]

前記第 1 のベクトルレジスタ内の前記与えられた数の複数のターゲットデータエレメントは、前記第 1 のベクトルレジスタ内の複数の下位のエレメントまたは複数の上位のエレメントのいずれかであり、前記第 2 のベクトルレジスタ内の前記与えられた数の複数のソースデータエレメントは、前記第 2 のベクトルレジスタ内の複数の下位のエレメントまたは複数の上位のエレメントのいずれかである

項目 1 から 4 の何れか一項に記載の装置。

10

[項目 6]

前記第 1 のビット値は前記第 2 のビット値の逆である

項目 1 から 5 の何れか一項に記載の装置。

[項目 7]

前記第 1 のビット値は前記第 2 のビット値と同一である

項目 1 から 6 のいずれか一項に記載の装置。

20

[項目 8]

前記与えられた数は、前記第 1 のビット値を有する前記第 1 のマスクレジスタ内のマスクビットの数、及び、前記第 2 のビット値を有する前記第 2 のマスクレジスタ内のマスクビットの数のうちの小さい方である

項目 1 から 7 の何れか一項に記載の装置。

[項目 9]

プロセッサが、第 1 のマスクレジスタ、第 2 のマスクレジスタ、第 1 のベクトルレジスタ、及び第 2 のベクトルレジスタを特定するベクトル移動命令を受信する段階と、

前記ベクトル移動命令に回答して、前記第 1 のベクトルレジスタ内の与えられた数の複数のターゲットデータエレメントを、前記第 2 のベクトルレジスタ内の前記与えられた数の複数のソースデータエレメントで置き換える段階と、を含み、

前記複数のソースデータエレメントの各々は、第 2 のビット値を有する前記第 2 のマスクレジスタ内のマスクビットに対応し、前記複数のターゲットデータエレメントの各々は、第 1 のビット値を有する前記第 1 のマスクレジスタ内のマスクビットに対応する方法。

30

[項目 10]

前記第 1 のビット値を有する前記第 1 のマスクレジスタの前記マスクビットの各々に対し、前記複数のソースデータエレメントのうちの一つとして、前記第 2 のマスクレジスタ内の前記第 2 のビット値の、対応するマスクビットを有するデータエレメントを探索するために前記第 2 のベクトルレジスタを検索する段階をさらに備える

項目 9 に記載の方法。

40

[項目 11]

前記第 1 のベクトルレジスタ内の複数のデータエレメントに対して再帰計算を実行する段階と、前記第 1 のベクトルレジスタ内に前記再帰計算の複数の結果を蓄積する段階と、をさらに備える

項目 9 または 10 に記載の方法。

[項目 12]

前記第 1 のベクトルレジスタ内の 1 または複数のデータエレメントがさらに計算を必要としなくなるまで、前記第 1 のベクトルレジスタ内の複数のデータエレメントに対し再帰

50

計算を実行する段階と、

前記第 1 のマスクレジスタを用いて、前記 1 または複数のデータエレメントが前記複数のターゲットデータエレメントであると示す段階と、をさらに備える

項目 9 から 1 1 の何れか一項に記載の方法。

[項目 1 3]

前記第 2 のベクトルレジスタ内の前記与えられた数の複数のソースデータエレメントは、前記第 2 のベクトルレジスタ内の複数の下位のデータエレメントまたは複数の上位のデータエレメントのいずれかであり、前記第 1 のベクトルレジスタ内の前記与えられた数の複数のターゲットデータエレメントは、前記第 1 のベクトルレジスタ内の複数の下位のデータエレメントまたは複数の上位のデータエレメントのいずれかである

10

項目 9 から 1 2 の何れか一項に記載の方法。

[項目 1 4]

前記第 1 のビット値は前記第 2 のビット値の逆である

項目 9 から 1 3 の何れか一項に記載の方法。

[項目 1 5]

前記第 1 のビット値は前記第 2 のビット値と同一である

項目 9 から 1 4 の何れか一項に記載の方法。

[項目 1 6]

前記与えられた数は、前記第 1 のビット値を有する前記第 1 のマスクレジスタ内のマスクビットの数、及び、前記第 2 のビット値を有する前記第 2 のマスクレジスタ内のマスクビットの数のうちの小さい方である

20

項目 9 から 1 5 の何れか一項に記載の方法。

[項目 1 7]

複数の入力データエレメントを含む入力データアレイをストアするメモリと、

第 1 のマスクレジスタ、第 2 のマスクレジスタ、第 1 のベクトルレジスタ、及び第 2 のベクトルレジスタを含むレジスタファイルと、

前記メモリ及び前記レジスタファイルに連結される実行回路と、を備え、前記実行回路は、複数のイタレーションの間、前記第 1 のベクトルレジスタに対し再帰計算を実行し、前記複数のイタレーションのうちの 2 回以上は、

前記入力データアレイから前記第 2 のベクトルレジスタへ前記複数の入力データエレメントのうちの少なくとも一部をロードするベクトルロードオペレーションと、

30

前記第 2 のベクトルレジスタ内の前記複数の入力データエレメントを前記第 1 のベクトルレジスタへ移動させるベクトル移動オペレーションと、

前記第 1 のマスクレジスタ及び前記第 2 のマスクレジスタを更新して、それぞれ、さらに計算を必要とする前記第 1 のベクトルレジスタ及び前記第 2 のベクトルレジスタ内の、複数のデータエレメントを特定するマスク更新オペレーションと、

前記第 1 のベクトルレジスタ内の前記複数のデータエレメントに対して実行するベクトル演算オペレーションと、

前記ベクトル演算オペレーションの複数の結果を前記メモリへストアするベクトルストアオペレーションからなる複数のオペレーションを含む

40

システム。

[項目 1 8]

前記実行回路は、前記第 1 のベクトルレジスタ内の与えられた数の複数のターゲットデータエレメントを、前記第 2 のベクトルレジスタ内の前記与えられた数の複数のデータエレメントで置き換える前記ベクトル移動オペレーションを実行し、各ソースデータエレメントは、第 2 のビット値を有する前記第 2 のマスクレジスタ内のマスクビットに対応し、各ターゲットデータエレメントは、第 1 のビット値を有する前記第 1 のマスクレジスタ内のマスクビットに対応する

項目 1 7 に記載のシステム。

[項目 1 9]

50

前記第 1 のビット値は前記第 2 のビット値の逆である
項目 17 に記載のシステム。

[項目 20]

前記第 1 のビット値は前記第 2 のビット値と同一である
項目 17 に記載のシステム。

[項目 21]

前記与えられた数は、前記第 1 のビット値を有する前記第 1 のマスクレジスタ内のマスクビットの数、及び、前記第 2 のビット値を有する前記第 2 のマスクレジスタ内のマスクビットの数のうちの小さい方である

項目 17 に記載のシステム。

10

[項目 22]

前記実行回路は前記マスク更新オペレーションを実行して、前記第 1 のマスクレジスタ及び前記第 2 のマスクレジスタを更新し、前記第 1 のマスクレジスタ及び前記第 2 のマスクレジスタの複数のマスクビットがそれぞれ、さらに計算を必要とする、前記第 1 のベクトルレジスタ及び前記第 2 のベクトルレジスタの、複数の対応するデータエレメントを特定するようにする

項目 17 から 21 の何れか一項に記載のシステム。

[項目 23]

前記第 1 のベクトルレジスタ、前記第 1 のマスクレジスタ、前記第 2 のベクトルレジスタ、及び前記第 2 のマスクレジスタを複数のオペランドとして特定するベクトル移動命令

に回答して、前記実行回路は前記ベクトル移動オペレーションを実行する

項目 17 から 22 の何れか一項に記載のシステム。

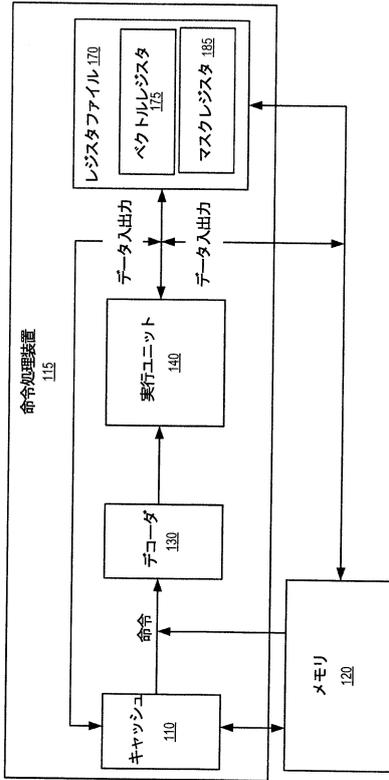
20

[項目 24]

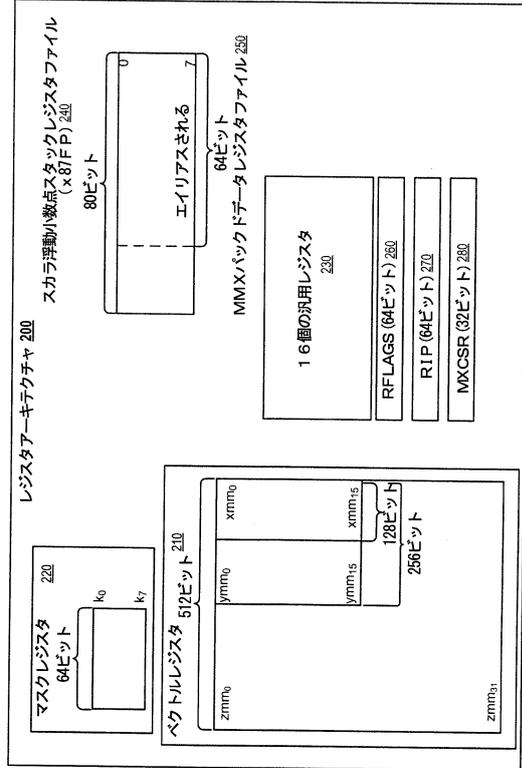
前記第 1 のマスクレジスタ及び前記第 2 のマスクレジスタを複数のオペランドとして特定するマスク更新命令に回答して、前記実行回路は前記マスク更新オペレーションを実行する

項目 17 から 23 の何れか一項に記載のシステム。

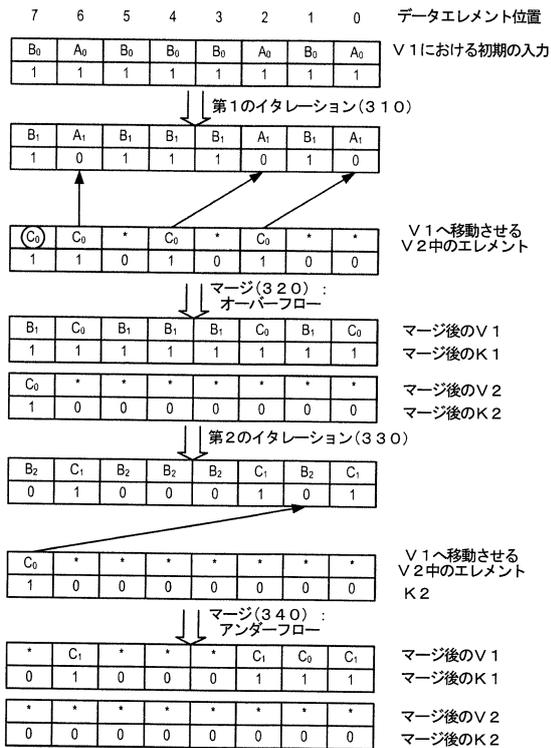
【図1】



【図2】



【図3】



【図4A】

```

401
rvmaskupdate(K1,K2)
for (i = 0, k = 0; i < KL; i++){
    if ((K1[i]) < (K2[i])){
        for (; k < KL; k++){
            if ((K2[k]) < (K1[k])){
                K2[k] = 0;
                K1[k] = 1;
                k++;
                break;
            }
        }
    }
}

402
sparsemov(K1,V1,K2,V2)
for (i = 0, k = 0; i < KL; i++){
    if ((K1[i]) < (K2[i])){
        for (; k < KL; k++){
            if ((K2[k]) < (K1[k])){
                V1[i] = V2[k];
                k++;
                break;
            }
        }
    }
}
    
```

【図4B】

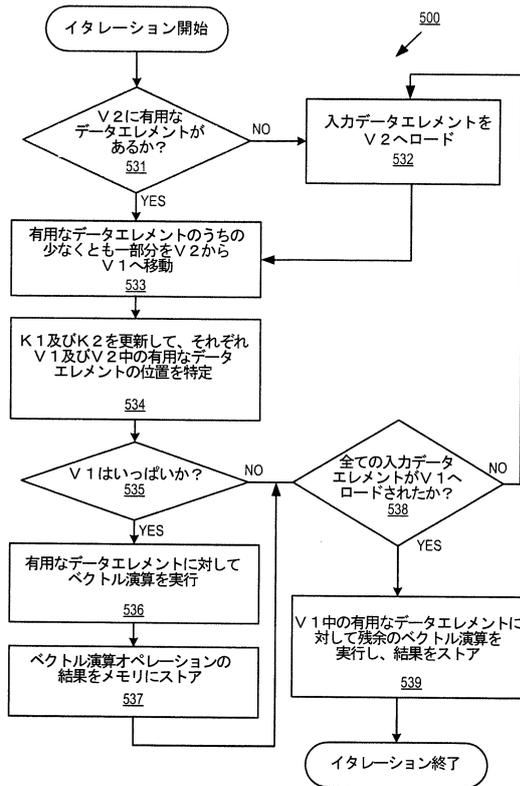
```

400
i = 0; // ループカウンタを初期化
v_index = -1; // インデックスの初期ベクトル
v_kl = KL; // インデックスのベクトルに対するインクリメント
K1 = 0; // アキュムレータは初期は空である
K2 = 0; // オーバーフローは来ない
do {
  if (K2 == 0) { // 前のオーバーフローから元素が1つも残されていないならば
    V2 = vector_load(X[i+KL-1:i]); // Xアレイの新しいKL元素をロード
    K2 = condition(V2); // 新しい元素に対してリードマスクを生成
    i += KL; // ループカウンタをインクリメント
    v_index += v_kl; // インデックスベクトルをインクリメント
  } // その他リードマスクK2で続行
431 sparsemov(K1, V1, K2, V2); // V1...に新しい元素を追加
432 sparsemov(K1, V3, K2, v_index); // V3にそれらのインデックスを追加
433 rwmaskupdate(K1, K2); // リードマスク及びライトマスクを更新

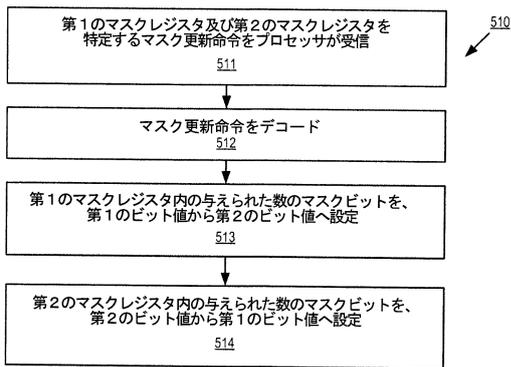
isFullMask(K1) { // いっぱいになったアキュムレータのみに対して
  do {
    V1 = computation(V1); // 蓄積されたデータに対するデンスコンピューテーション
    K1 = condition(V1); // 計算後、条件をチェック
    while(K1 == 0xFFFF) { // アキュムレータがまだいっぱい(k1が空で1)であるかどうかチェック
      scatter(knot(K1), V1, V3, X); // インデックスV3を有するV1からXアレイへ元素を分散
    }
  } while((i < N) || (K2 != 0)); // V2内にもっとも入力ストリームまたは新しいデータがある場合、続行
  // 残りの計算を開始
  K3 = K1; // 最終の分散用に残余のマスクをストア
450
  do {
    V1[K1] = computation(V1); // マスクK1下で計算
    K1 = condition(V1); // 計算後、条件をチェック
    while(K1 == 0) { // 成すべきことが何も残されていないかどうかチェック
      scatter(K3, V1, V3, X); // インデックスV3を有するV1からXアレイへ残余の元素を分散
    }
  }
}

```

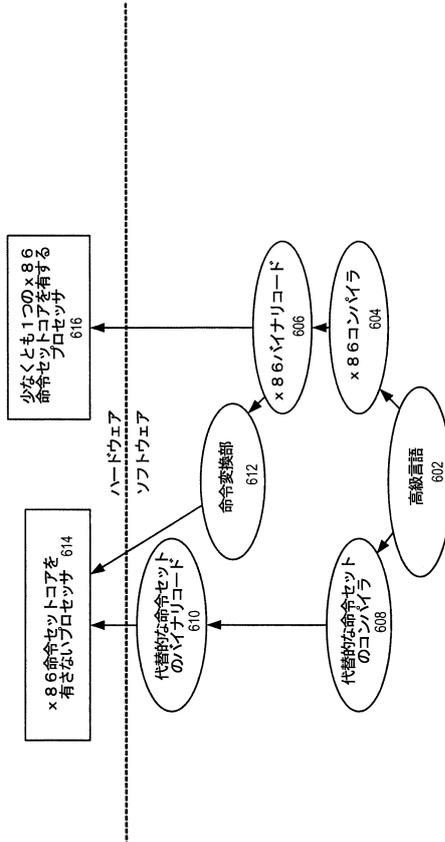
【図5A】



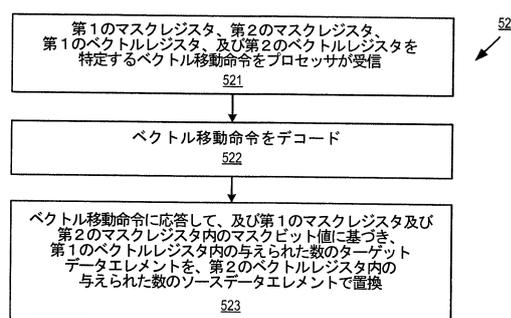
【図5B】



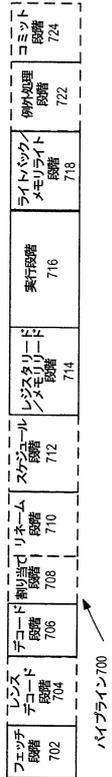
【図6】



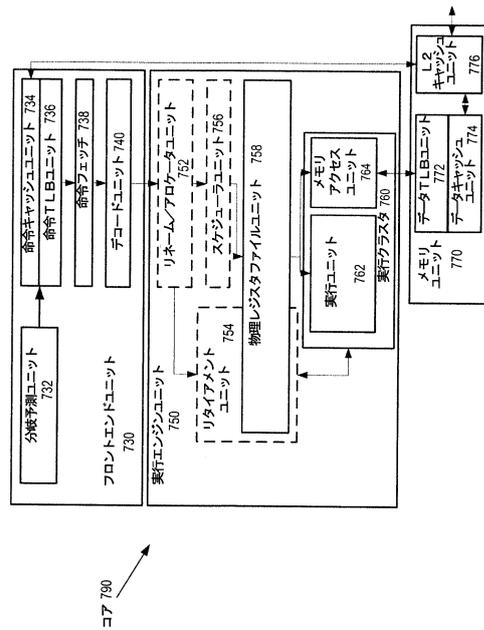
【図5C】



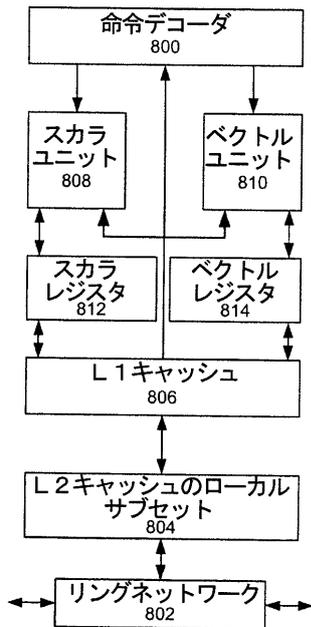
【図7A】



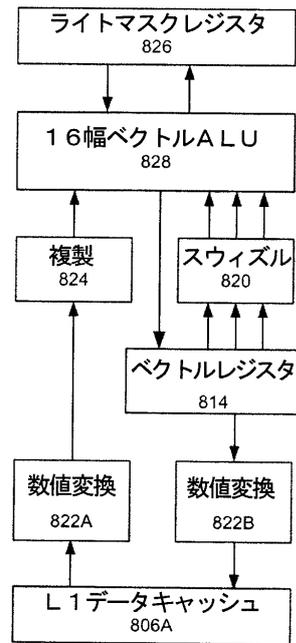
【図7B】



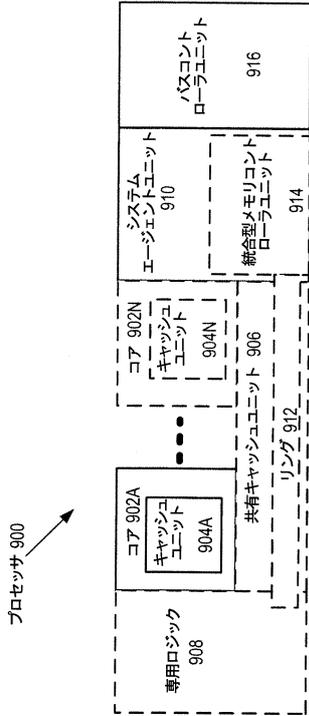
【図8A】



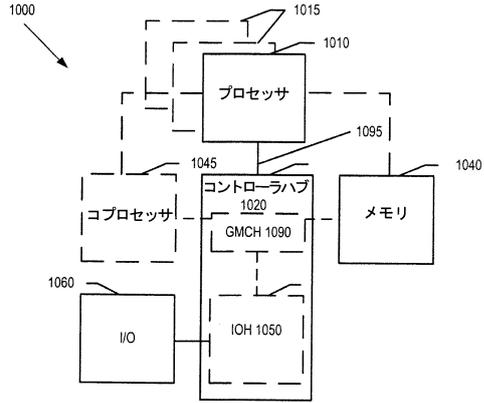
【図8B】



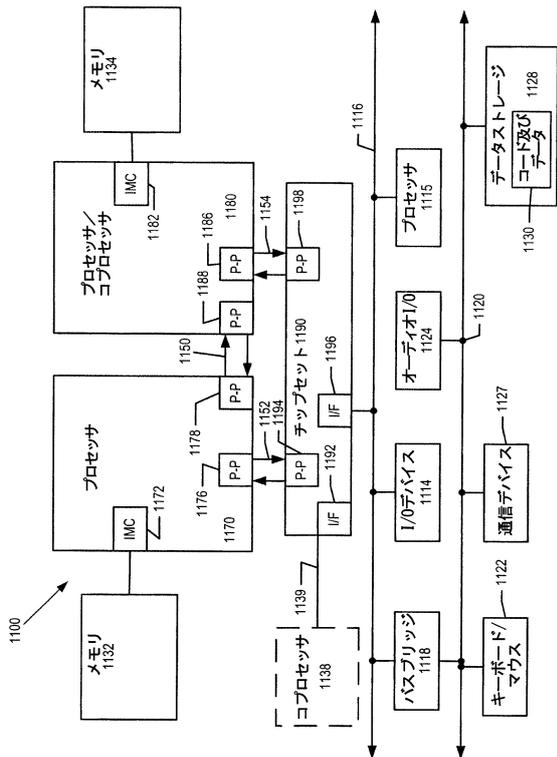
【図9】



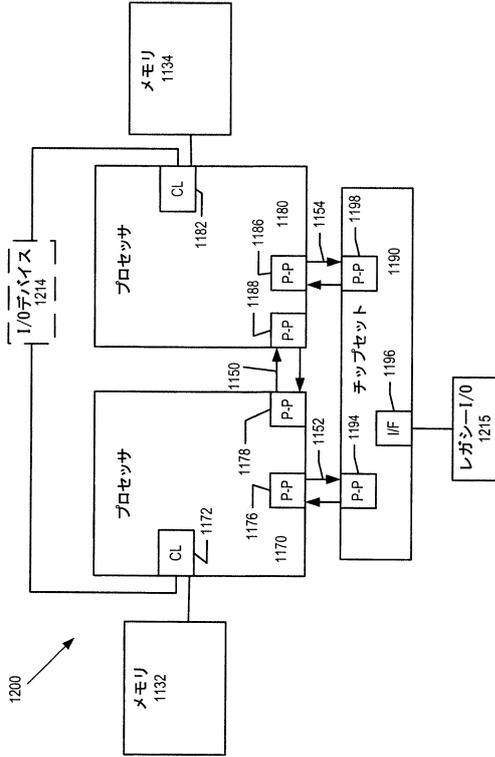
【図10】



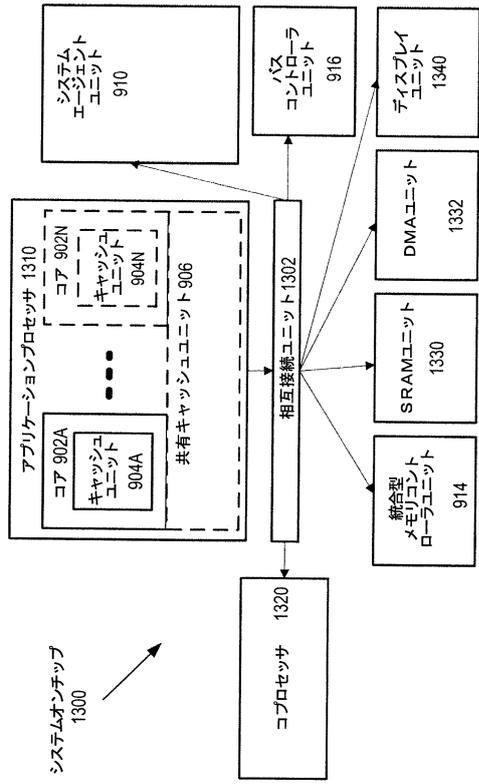
【図11】



【図12】



【 図 13 】



フロントページの続き

- (72)発明者 ナライキン、アンドレー
アメリカ合衆国 95054 カリフォルニア州・サンタクララ・ミッション カレッジ ブーレ
バード・2200 インテル・コーポレーション内
- (72)発明者 ヒューズ、クリストファー
アメリカ合衆国 95054 カリフォルニア州・サンタクララ・ミッション カレッジ ブーレ
バード・2200 インテル・コーポレーション内

審査官 清木 泰

- (56)参考文献 特表2015-524978(JP,A)
特開平11-126200(JP,A)
特開平11-3226(JP,A)
特開平7-141326(JP,A)
特開平6-44292(JP,A)
特開昭60-59469(JP,A)
米国特許第7480787(US,B1)
NECスーパーコンピュータSX-9シリーズ 中央処理装置機能説明書 UOM-G1AZ124-1, 日本電気株
式会社, 2008年 2月, 初版, Pages:219-220
廣松悠介, 黒田久泰, 金田康正, 複雑な制御構造を持つプログラムのSIMD命令セットによる最適
化, 情報処理学会論文誌, 日本, 社団法人情報処理学会, 2007年 3月15日, 第48巻, 第
SIG4(PR032)号, Pages:62-72

(58)調査した分野(Int.Cl., DB名)

G06F 9/30 - 9/42
G06F 17/16
G06F 8/41
G06F 15/80
G06F 7/57 - 7/575