

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6673574号
(P6673574)

(45) 発行日 令和2年3月25日(2020.3.25)

(24) 登録日 令和2年3月9日(2020.3.9)

(51) Int. Cl. F I
G06F 9/315 (2006.01) G O 6 F 9/315 S
G06F 9/38 (2006.01) G O 6 F 9/38 3 7 O A

請求項の数 25 (全 38 頁)

(21) 出願番号	特願2017-528541 (P2017-528541)	(73) 特許権者	591003943 インテル・コーポレーション
(86) (22) 出願日	平成27年11月25日 (2015.11.25)		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
(65) 公表番号	特表2018-506096 (P2018-506096A)	(74) 代理人	110000877 龍華国際特許業務法人
(43) 公表日	平成30年3月1日 (2018.3.1)	(72) 発明者	ウルド・アハメド・ヴァル、エルムスタッフ 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
(86) 国際出願番号	PCT/US2015/062564		
(87) 国際公開番号	W02016/105819		
(87) 国際公開日	平成28年6月30日 (2016.6.30)		
審査請求日	平成30年11月20日 (2018.11.20)		
(31) 優先権主張番号	14/583, 636		
(32) 優先日	平成26年12月27日 (2014.12.27)		
(33) 優先権主張国・地域又は機関	米国 (US)		

最終頁に続く

(54) 【発明の名称】ベクトルビットシャッフルを実行するための方法および装置

(57) 【特許請求の範囲】

【請求項1】

複数のソースデータ要素を格納するための第1のベクトルレジスタと、
 複数の制御要素を格納するための第2のベクトルレジスタであって、前記複数の制御要素の各々は複数のビットフィールドを含んで前記第1のベクトルレジスタにおける前記複数のソースデータ要素のうちの別々の1つに対応し、各ビットフィールドはデスティネーションマスクレジスタ内の単一のビット位置に対応し、各ビットフィールドは、前記デスティネーションマスクレジスタ内の対応する前記単一のビット位置へコピーされるべき、対応する前記ソースデータ要素からの一のビットを識別する、第2のベクトルレジスタと

、
 前記対応するソースデータ要素からの一のビットを識別すべく前記第2のベクトルレジスタから各ビットフィールドを読み出し且つこれに応じて前記対応するソースデータ要素からの識別された前記ビットを、前記デスティネーションマスクレジスタ内の前記ビットフィールドに対応する単一のビット位置にコピーするためのベクトルビットシャッフルロジックと、を備え、

前記デスティネーションマスクレジスタのビット数は、前記複数のソースデータ要素の個数と、当該複数の制御要素の各々における前記複数のビットフィールドの個数との乗算値に等しい、プロセッサ。

【請求項2】

前記ベクトルビットシャッフルロジックは、前記複数の制御要素の各々における前記複

数のビットフィールドに従い、前記複数のソースデータ要素の各々からビットのセットを選択するための1または複数のマルチプレクサを含む、請求項1に記載のプロセッサ。

【請求項3】

前記複数のソースデータ要素の各々は64ビットデータ要素を含み、各ビットフィールドは、複数の前記64ビットデータ要素の各々からのビットを識別するための少なくとも6ビットを含む、請求項1または2に記載のプロセッサ。

【請求項4】

前記複数のビットフィールドの各々は制御バイトを含み、前記6ビットは複数の前記制御バイトの各々から選択され、複数の前記64ビットデータ要素の各々からの各ビットを識別する、請求項3に記載のプロセッサ。

10

【請求項5】

8個の前記制御バイトを使用して、各データ要素から8ビットが選択される、請求項4に記載のプロセッサ。

【請求項6】

各データ要素からの前記8ビットは、前記デスティネーションマスクレジスタ内で連結される、請求項5に記載のプロセッサ。

【請求項7】

前記第1のベクトルレジスタは、複数の前記64ビットデータ要素のうち8個の64ビットデータ要素を格納し、前記デスティネーションマスクレジスタは前記8個の64ビットデータ要素から選択された8個の対応する8ビット値を格納する、請求項6に記載のプロセッサ。

20

【請求項8】

前記デスティネーションマスクレジスタ内の前記ビットは、前記プロセッサによって実行される1または複数の後続の命令のためのマスク演算の実行に使用される、請求項7に記載のプロセッサ。

【請求項9】

前記ベクトルビットシャッフルロジックは、前記プロセッサ内のデコードロジックによってデコードされ、前記プロセッサ内の実行ロジックによって実行されるベクトルビットシャッフル命令にตอบสนองして動作する、請求項1から8のいずれか一項に記載のプロセッサ。

30

【請求項10】

複数のソースデータ要素を第1のベクトルレジスタ内に格納する段階と、

複数の制御要素を第2のベクトルレジスタ内に格納する段階であって、前記複数の制御要素の各々は複数のビットフィールドを含んで前記第1のベクトルレジスタにおける前記複数のソースデータ要素のうちの別々の1つに対応し、各ビットフィールドはデスティネーションマスクレジスタ内の単一のビット位置に対応し、各ビットフィールドは、前記デスティネーションマスクレジスタ内の対応する前記単一のビット位置へコピーされるべき、対応する前記ソースデータ要素からの一のビットを識別する、複数の制御要素を第2のベクトルレジスタ内に格納する段階と、

前記対応するソースデータ要素からの一のビットを識別すべく前記第2のベクトルレジスタから各ビットフィールドを読み出し且つこれに応じて前記対応するソースデータ要素からの識別された前記ビットを、前記デスティネーションマスクレジスタ内の前記ビットフィールドに対応する単一のビット位置にコピーする段階と、を備え、

40

前記デスティネーションマスクレジスタのビット数は、前記複数のソースデータ要素の個数と、当該複数の制御要素の各々における前記複数のビットフィールドの個数との乗算値に等しい、方法。

【請求項11】

前記複数の制御要素の各々における前記複数のビットフィールドに従い、1または複数のマルチプレクサを用いて、前記複数のソースデータ要素の各々からビットのセットを選択する段階をさらに備える、請求項10に記載の方法。

50

【請求項 1 2】

前記複数のソースデータ要素の各々は 6 4 ビットデータ要素を含み、各ビットフィールドは複数の前記 6 4 ビットデータ要素の各々からのビットを識別するための少なくとも 6 ビットを含む、請求項 1 0 または 1 1 に記載の方法。

【請求項 1 3】

前記複数のビットフィールドの各々は制御バイトを含み、前記 6 ビットは、複数の前記制御バイトの各々から選択され、複数の前記 6 4 ビットデータ要素の各々からの各ビットを識別する、請求項 1 2 に記載の方法。

【請求項 1 4】

8 個の前記制御バイトを使用して、各データ要素から 8 ビットが選択される、請求項 1 3 に記載の方法。

【請求項 1 5】

各データ要素からの前記 8 ビットは、前記デスティネーションマスクレジスタ内で連結される、請求項 1 4 に記載の方法。

【請求項 1 6】

前記第 1 のベクトルレジスタは複数の前記 6 4 ビットデータ要素のうち 8 個の 6 4 ビットデータ要素を格納し、前記デスティネーションマスクレジスタは前記 8 個の 6 4 ビットデータ要素から選択された 8 個の対応する 8 ビット値を格納する、請求項 1 5 に記載の方法。

【請求項 1 7】

前記デスティネーションマスクレジスタ内の前記ビットは、プロセッサによって実行される 1 または複数の後続の命令のためのマスク演算の実行に使用される、請求項 1 6 に記載の方法。

【請求項 1 8】

プログラムコードおよびデータを格納するためのメモリと、
指定されたキャッシュ管理ポリシーに従い前記プログラムコードおよびデータをキャッシュするための複数のキャッシュレベルを有するキャッシュ階層と、
ユーザからの入力を受信するための入力デバイスと、
前記プログラムコードを実行し且つ前記ユーザからの前記入力に応じて前記データを処理するためのプロセッサと、を備え、

前記プロセッサは、

複数のソースデータ要素を格納するための第 1 のベクトルレジスタと、

複数の制御要素を格納するための第 2 のベクトルレジスタであって、前記複数の制御要素の各々は複数のビットフィールドを含んで前記第 1 のベクトルレジスタにおける前記複数のソースデータ要素のうちの別々の 1 つに対応し、各ビットフィールドはデスティネーションマスクレジスタ内の単一のビット位置に対応し、各ビットフィールドは、前記デスティネーションマスクレジスタ内の対応する前記単一のビット位置へコピーされるべき、対応する前記ソースデータ要素からの一のビットを識別する、第 2 のベクトルレジスタと

、
前記対応するソースデータ要素からの一のビットを識別すべく前記第 2 のベクトルレジスタから各ビットフィールドを読み出し且つこれに応じて前記対応するソースデータ要素からの識別された前記ビットを、前記デスティネーションマスクレジスタ内の前記ビットフィールドに対応する単一のビット位置にコピーするためのベクトルビットシャッフルロジックと、を含み、

前記デスティネーションマスクレジスタのビット数は、前記複数のソースデータ要素の個数と、当該複数の制御要素の各々における前記複数のビットフィールドの個数との乗算値に等しい、システム。

【請求項 1 9】

前記ベクトルビットシャッフルロジックは、前記複数の制御要素の各々における前記複数のビットフィールドに従い、前記複数のソースデータ要素の各々からビットのセットを

10

20

30

40

50

選択するための1または複数のマルチプレクサを含む、請求項18に記載のシステム。

【請求項20】

前記複数のソースデータ要素の各々は64ビットデータ要素を含み、各ビットフィールドは複数の前記64ビットデータ要素の各々からのビットを識別するための少なくとも6ビットを含む、請求項18に記載のシステム。

【請求項21】

前記複数のビットフィールドの各々は制御バイトを含み、前記6ビットは複数の前記制御バイトの各々から選択され、複数の前記64ビットデータ要素の各々からの各ビットを識別する、請求項20に記載のシステム。

【請求項22】

8個の前記制御バイトを使用して、各データ要素から8ビットが選択される、請求項21に記載のシステム。

【請求項23】

各データ要素からの前記8ビットは、前記デスティネーションマスクレジスタ内で連結される、請求項22に記載のシステム。

【請求項24】

前記第1のベクトルレジスタは複数の前記64ビットデータ要素のうち8個の64ビットデータ要素を格納し、前記デスティネーションマスクレジスタは前記8個の64ビットデータ要素から選択された8個の対応する8ビット値を格納する、請求項23に記載のシステム。

【請求項25】

前記デスティネーションマスクレジスタ内の前記ビットは、前記プロセッサによって実行される1または複数の後続の命令のためのマスク演算の実行に使用される、請求項24に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、概してコンピュータのプロセッサ分野に関する。具体的には、本発明は、ベクトルビットシャッフルを実行するための方法および装置に関する。

【背景技術】

【0002】

命令セットまたは命令セットアーキテクチャ(ISA)は、ネイティブデータタイプ、命令、レジスタアーキテクチャ、アドレス指定モード、メモリアーキテクチャ、割り込みおよび例外処理並びに外部入力および出力(I/O)を含む、プログラミングに関するコンピュータアーキテクチャの一部である。本明細書において、「命令」という用語は概してマクロ命令を指すことに留意されたい。マクロ命令とは、実行のためにプロセッサに供給される命令であり、これに対し、マイクロ命令またはマイクロopとは、マクロ命令をデコーディングするプロセッサのデコーダの結果である。マイクロ命令またはマイクロopは、プロセッサの実行ユニットに対し、マクロ命令に関連するロジックを実装するための演算を実行するよう命令するように構成可能である。

【0003】

ISAは、命令セットの実装に使用される一連のプロセッサ設計技術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持つプロセッサは、共通の命令セットを共有可能である。例えば、インテル(登録商標)PENTIUM(登録商標)4プロセッサ、インテル(登録商標)コア(商標)プロセッサおよびカリフォルニア州サンニールのAdvanced Micro Devices社のプロセッサは、x86命令セット(より新しいバージョンに追加されたいくつかの拡張機能を持つ)とほぼ同一バージョンを実装するが、内部設計が異なる。例えば、ISAの同一のレジスタアーキテクチャは、専用の物理レジスタ、レジスタリネーミングメカニズムを使用(例えば、レジスタエイリアステーブル(RAT)、リオーダバッファ(ROB)およびリタイアメ

10

20

30

40

50

ントレジスタファイルの使用)して動的に割り当てられた1または複数の物理レジスタを含む周知の技術を使用し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てよい。別途の記載がない限り、本明細書において、レジスタアーキテクチャ、レジスタファイルおよびレジスタという文言は、ソフトウェア/プログラマに可視であるもの、および命令がレジスタを指定する方法を指すために使用される。区別が必要な場合、「論理」、「アーキテクチャ」または「ソフトウェアビジブル」なる形容詞が、レジスタアーキテクチャにおけるレジスタ/ファイルを示すために使用される一方で、異なる形容詞が、特定のマイクロアーキテクチャにおけるレジスタ(例えば、物理レジスタ、リオードバッファ、リタイアメントレジスタ、レジスタプール)を指すために使用される。

【0004】

命令セットは、1または複数の命令フォーマットを含む。特定の命令フォーマットは、とりわけ、実行されるべき演算およびその演算が実行されるべきオペランドを指定するための様々なフィールド(ビット数、ビット位置)を定義する。いくつかの命令フォーマットは、命令テンプレート(またはサブフォーマット)の定義を通して、さらに細分化されている。例えば、特定の命令フォーマットの命令テンプレートは、命令フォーマットのフィールドの異なるサブセットを有するように定義されてよく(含まれるフィールドは通常、同一順序であるが、少なくともいくつかは、含まれるフィールド数がより少ないので、異なるビット位置を有する)、および/または、異なって解釈される特定のフィールドを有するように定義されてよい。特定の命令は、特定の命令フォーマット(また、定義されている場合には、その命令フォーマットの命令テンプレートのうちの特定の1つにおいて)を使用して表現され、演算およびオペランドを指定する。命令ストリームとは、特定の命令シーケンスであり、シーケンス内の各命令は、命令フォーマット(また、定義されている場合には、その命令フォーマットの命令テンプレートのうちの特定の1つにおける)内の命令の出現である。

【図面の簡単な説明】**【0005】**

以下の詳細な説明に以下の添付図面を組み合わせると、本発明のより良い理解が得られる。

【0006】

【図1A】本発明の実施形態による汎用ベクトル向け命令フォーマットおよびその命令テンプレートを示すブロック図である。

【図1B】本発明の実施形態による汎用ベクトル向け命令フォーマットおよびその命令テンプレートを示すブロック図である。

【0007】

【図2A】本発明の実施形態による例示的な特定ベクトル向け命令フォーマットを示すブロック図である。

【図2B】本発明の実施形態による例示的な特定ベクトル向け命令フォーマットを示すブロック図である。

【図2C】本発明の実施形態による例示的な特定ベクトル向け命令フォーマットを示すブロック図である。

【図2D】本発明の実施形態による例示的な特定ベクトル向け命令フォーマットを示すブロック図である。

【0008】

【図3】本発明の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。

【0009】

【図4A】本発明の実施形態による、例示的なインオーダーフェッチ、デコード、リタイアパイプラインおよび例示的なレジスタリネーミング、アウトオブオーダー発行/実行パイプラインの両方を示すブロック図である。

【0010】

【図4B】本発明の実施形態によるプロセッサに含まれる、インオーダーフェッチ、デコー

10

20

30

40

50

ド、リタイアコアに係る例示的な実施形態および例示的なレジスタリネーミング、アウトオブオーダー発行/実行アーキテクチャコアの両方を示すブロック図である。

【0011】

【図5A】オンダイ相互接続ネットワークへの接続を伴う単一のプロセッサコアのブロック図である。

【0012】

【図5B】本発明の実施形態による図5A中のプロセッサコアの一部の拡大図を示す。

【0013】

【図6】本発明の実施形態による統合メモリコントローラおよびグラフィックを持つ単一のコアプロセッサおよびマルチコアプロセッサのブロック図である。

10

【0014】

【図7】本発明の一実施形態によるシステムのブロック図を示す。

【0015】

【図8】本発明の実施形態による第2のシステムのブロック図を示す。

【0016】

【図9】本発明の実施形態による第3のシステムのブロック図を示す。

【0017】

【図10】本発明の実施形態によるシステムオンチップ(SoC)のブロック図を示す。

【0018】

【図11】本発明の実施形態による、ソース命令セット内のバイナリ命令をターゲット命令セット内のバイナリ命令に変換するためのソフトウェア命令コンバータの使用を対比するブロック図を示す。

20

【0019】

【図12】本発明の実施形態が実装されてよい例示的なプロセッサを示す。

【0020】

【図13】本発明の一実施形態によるベクトルビットシャッフルロジックを示す。

【0021】

【図14】本発明の一実施形態による方法を示す。

【発明を実施するための形態】

【0022】

30

以下の詳細な説明には、後述の本発明の実施形態に係る完全な理解を共すべく、説明目的で多数の具体的な詳細が記載されている。しかしながら、本発明の実施形態は、これらの具体的な詳細の一部を省いても実施可能であることは当業者に自明なところである。他の例においては、本説明の実施形態に係る根本的な原理を曖昧にしないように、周知の構造およびデバイスはブロック図内に詳細に示されていない。

[例示的なプロセッサアーキテクチャおよびデータタイプ]

【0023】

命令セットは1または複数の命令フォーマットを含む。特定の命令フォーマットは、とりわけ、実行されるべき演算(オペコード)およびその演算が実行されるべきオペランドを指定するための様々なフィールド(ビット数、ビット位置)を定義する。いくつかの命令フォーマットは、命令テンプレート(またはサブフォーマット)の定義を通して、さらに細分化されている。例えば、特定の命令フォーマットの命令テンプレートは、命令フォーマットのフィールドの異なるサブセットを有するように定義されてよく(含まれるフィールドは通常、同一順序であるが、少なくともいくつかは、含まれるフィールド数がより少ないので、異なるビット位置を有する)、および/または、異なって解釈される特定のフィールドを有するように定義されてよい。故に、ISAの各命令は、特定の命令フォーマット(また、定義されている場合には、その命令フォーマットの命令テンプレートのうちの特定の1つにおいて)を使用して表現され、演算およびオペランドを指定するためのフィールドを含む。例えば、例示的なADD命令は、特定のオペコード並びにそのオペコードを指定するためのオペコードフィールドおよびオペランド(ソース1/デスティネー

40

50

ションおよびソース2)を選択するためのオペランドフィールドを含む命令フォーマットを有する。命令ストリーム内にこのADD命令が出現すると、特定のオペランドを選択するオペランドフィールド内に特定の内容を有することになる。アドバンスドベクトル拡張(AVX)(AVX1およびAVX2)と称され、ベクトル拡張(VEX)コーディングスキームを使用する一連のSIMD拡張機能がリリースおよび/または公開されている(例えば、2011年10月のインテル(登録商標)64およびIA-32アーキテクチャソフトウェアデベロッパーズマニュアル並びに2011年6月のインテル(登録商標)アドバンスドベクトル拡張プログラミングリファレンスを参照)。

[例示的な命令フォーマット]

【0024】

本明細書に記載の命令の実施形態は異なる形式で具現化されてよい。また、例示的なシステム、アーキテクチャおよびパイプラインについて詳細に後述する。本命令の実施形態は、このようなシステム、アーキテクチャおよびパイプライン上で実行されてよいが、本発明の実施形態はそれらの具体的な内容に限定されるわけではない。

A [汎用ベクトル向け命令フォーマット]

【0025】

ベクトル向け命令フォーマットとは、ベクトル命令に好適な命令フォーマットである(例えば、ベクトル演算に特有の特定のフィールドが存在する)。実施形態は、ベクトル演算およびスカラ演算の両方がベクトル向け命令フォーマットを通してサポートされるように記載されているものの、代替的な実施形態は、ベクトル向け命令フォーマットのベクトル演算のみを使用する。

【0026】

図1A~1Bは、本発明の実施形態による、汎用ベクトル向け命令フォーマットおよびその命令テンプレートを示すブロック図である。図1Aは、本発明の実施形態による汎用ベクトル向け命令フォーマットおよびそのクラスA命令テンプレートを示すブロック図であり、これに対し、図1Bは、本発明の実施形態による汎用ベクトル向け命令フォーマットおよびそのクラスB命令テンプレートを示すブロック図である。具体的には、汎用ベクトル向け命令フォーマット100に対し、クラスA命令テンプレートおよびクラスB命令テンプレートが定義され、クラスA命令テンプレートおよびクラスB命令テンプレートは両方とも、メモリアクセスなし105命令テンプレートおよびメモリアクセス120命令テンプレートを含む。ベクトル向け命令フォーマットの文脈における汎用(generic)という用語は、いずれの特定の命令セットにも関連付けられない命令フォーマットを指す。

【0027】

本発明の実施形態は、ベクトル向け命令フォーマットが次のものをサポートするように記載されている。すなわち、32ビット(4バイト)または64ビット(8バイト)データ要素幅(またはサイズ)を備えた64バイトベクトルオペランド長(またはサイズ)(つまり、64バイトベクトルは、16個のダブルワードサイズの要素または代替的に8個のクワッドワードサイズの要素のいずれかから成る);16ビット(2バイト)または8ビット(1バイト)データ要素幅(またはサイズ)を備えた64バイトベクトルオペランド長(またはサイズ);32ビット(4バイト)、64ビット(8バイト)、16ビット(2バイト)または8ビット(1バイト)データ要素幅(またはサイズ)を備えた32バイトベクトルオペランド長(またはサイズ);および32ビット(4バイト)、64ビット(8バイト)、16ビット(2バイト)または8ビット(1バイト)データ要素幅(またはサイズ)を備えた16バイトベクトルオペランド長(またはサイズ)。一方で、代替的な実施形態は、より多い、より少ない、または異なるデータ要素幅(例えば、128ビット(16バイト)データ要素幅)を備えたより多い、より少ない、および/または異なるベクトルオペランドサイズ(例えば、256バイトベクトルオペランド)をサポートしてよい。

【0028】

10

20

30

40

50

図1A中のクラスA命令テンプレートには次のものが含まれる。すなわち、1)メモリアクセスなし105命令テンプレート内に、メモリアクセスなし、完全ラウンド制御タイプ演算110命令テンプレートおよびメモリアクセスなし、データ変換タイプ演算115命令テンプレートが存在するように図示されている。2)メモリアクセス120命令テンプレート内に、メモリアクセス、一時的125命令テンプレートおよびメモリアクセス、非一時的130命令テンプレートが存在するように図示されている。図1B中のクラスB命令テンプレートには次のものが含まれる。すなわち、1)メモリアクセスなし105命令テンプレート内に、メモリアクセスなし、書き込みマスク制御、部分的なラウンド制御タイプ演算112命令テンプレートおよびメモリアクセスなし、書き込みマスク制御、vsizeタイプ演算117命令テンプレートが存在するように図示されている。2)メモリアクセス120命令テンプレート内に、メモリアクセス、書き込みマスク制御127命令テンプレートが存在するように図示されている。

10

【0029】

汎用ベクトル向け命令フォーマット100は、以下に挙げられるフィールドを図1Aおよび図1B中に図示される順序で含む。

【0030】

フォーマットフィールド140。このフィールド内の特定の値(命令フォーマット識別子の値)は、ベクトル向け命令フォーマットを一意に識別し、故に命令ストリーム内のベクトル向け命令フォーマットの命令の出現を一意に識別する。よって、このフィールドは、汎用ベクトル向け命令フォーマットのみを有する命令セットには不要であるという意味において任意的である。

20

【0031】

ベース演算フィールド142。その内容が、異なるベース演算を区別する。

【0032】

レジスタインデックスフィールド144。その内容が、直接的にまたはアドレス生成を介して、ソースオペランドおよびデスティネーションオペランドの位置を指定する。それらはレジスタ内またはメモリ内である。これらは、 $P \times Q$ (例えば、 32×512 、 16×128 、 32×1024 、 64×1024)レジスタファイルからN個のレジスタを選択するための十分なビット数を含む。一実施形態において、Nは最大3つのソースレジスタおよび1つのデスティネーションレジスタであってよく、一方で、代替的な実施形態は、それより多いまたは少ないソースレジスタおよびデスティネーションレジスタをサポートしてよい(例えば、最大2つのソースをサポートしてよく、この場合、これらのソースのうちの1つがデスティネーションとしても動作する。最大3つのソースをサポートしてよく、この場合、これらのソースのうちの1つがデスティネーションとしても動作する。最大2つのソースおよび1つのデスティネーションをサポートしてよい)。

30

【0033】

修飾子フィールド146。その内容が、汎用ベクトル命令フォーマットの、メモリアクセスを指定する命令の出現を、メモリアクセスを指定しないものから区別する。すなわち、メモリアクセスなし105命令テンプレートおよびメモリアクセス120命令テンプレート間を区別する。メモリアクセス操作はメモリ階層に対し、読み取りおよび/または書き込みを行う(場合によっては、レジスタ内の値を使用してソースアドレスおよび/またはデスティネーションアドレスを指定する)が、メモリアクセスなし操作はそれを行わない(例えば、ソースおよびデスティネーションはレジスタである)。一実施形態において、このフィールドはまたメモリアドレス計算を実行するための3つの異なる方法の中で選択をする一方で、代替的な実施形態は、メモリアドレス計算を実行するためのより多い、より少ないまたは異なる方法をサポートしてよい。

40

【0034】

拡張演算フィールド150。その内容が、ベース演算に加え、様々な異なる演算のうちどれが実行されるべきかを区別する。このフィールドは、コンテキストに特有のものである。本発明の一実施形態において、このフィールドは、クラスフィールド168、アルフ

50

ァフィールド 152 およびベータフィールド 154 に分割される。拡張演算フィールド 150 は、2、3 または 4 個の命令ではなく、単一の命令の中で共通の演算グループが実行されることを可能にする。

【0035】

スケールフィールド 160。その内容が、メモリアドレス生成のための（例えば、 $2^s \text{scale} * \text{インデックス} + \text{ベース}$ を使用するアドレス生成のための）インデックスフィールドの内容のスケールリングを可能にする。

【0036】

変位フィールド 162A。その内容が、メモリアドレス生成（例えば、 $2^s \text{scale} * \text{インデックス} + \text{ベース} + \text{変位}$ を使用するアドレス生成について）の一部として使用される。

10

【0037】

変位係数フィールド 162B（変位係数フィールド 162B の直接の上位に、変位フィールド 162A が並置されていることで、一方または他方が使用されることを示すことに留意されたい）。その内容が、アドレス生成の一部として使用される。その内容は、メモリアクセス（N）のサイズ分スケールされるべき変位係数を指定する。ここで N は、メモリアクセス（例えば、 $2^s \text{scale} * \text{インデックス} + \text{ベース} + \text{スケールされた変位}$ を使用するアドレス生成について）におけるバイト数である。冗長下位ビットは無視され、従って、変位係数フィールドの内容は、有効アドレスの計算に使用される最終的な変位を生成すべく、メモリオペランドの合計サイズ（N）によって乗算される。N の値は、フル

20

【0038】

データ要素幅フィールド 164。その内容が、複数のデータ要素幅のうちどれが使用されるべきかを区別する（いくつかの実施形態においては、すべての命令に対し、他の実施形態においては、命令の一部のみに対し）。1つのデータ要素幅のみがサポートされる、

30

【0039】

書き込みマスクフィールド 170。その内容が、データ要素位置単位で、デスティネーションベクトルオペランド内のそのデータ要素位置が、ベース演算および拡張演算の結果を反映するかを制御する。クラス A 命令テンプレートは、マージ 書き込みマスクをサポートする一方で、クラス B 命令テンプレートは、マージ 書き込みマスクおよびゼロイング 書き込みマスクの両方をサポートする。マージの場合、ベクトルマスクは、任意の演算の実行中、デスティネーション内のあらゆる要素セットが更新されないように保護されることを可能にする（ベース演算および拡張演算によって指定される）。他の一実施形態においては、対応するマスクビットが 0 を有する場合、デスティネーションの各要素の古い値が保持される。これと対照的に、ゼロイングの場合、ベクトルマスクは、任意の演算の実行中、デスティネーション内のあらゆる要素セットがゼロにされることを可能にする（ベース演算および拡張演算によって指定される）。一実施形態においては、対応するマスクビットが 0 値を有する場合、デスティネーションの要素は 0 に設定される。この機能のうちのサブセットで、実行される演算のベクトル長（すなわち、要素のスパンが第 1 のものから最後のものへと変更される）を制御できる。しかしながら、変更される要素は連続的であることは必要ではない。故に、書き込みマスクフィールド 170 は、ロード、ストア、算術、論理等を含む部分的なベクトル演算を可能にする。本発明の実施形態は、書

40

50

き込みマスクフィールド170の内容は、複数の書き込みマスクレジスタのうち使用されるべき書き込みマスクを含むものを選択（故に、書き込みマスクフィールド170の内容は、実行されるべきマスクングを間接的に識別する）するように記載されているものの、代替的な実施形態は、代替的または追加的に、マスク書き込みフィールド170の内容が、実行されるべきマスクングを直接指定することを可能にする。

【0040】

即値フィールド172。その内容が、即値の指定を可能にする。このフィールドは即値をサポートしない汎用ベクトル向けフォーマットの実装には存在しない、および、このフィールドは即値を使用しない命令内には存在しないという意味において、このフィールドは、任意的なものである。

10

【0041】

クラスフィールド168。その内容が、異なるクラスの命令間を区別する。図1Aおよび図1Bを参照すると、このフィールドの内容で、クラスA命令およびクラスB命令間を選択する。図1Aおよび図1B中、特定の値がフィールド内に存在することを示すために、隅が丸められた四角が使用されている（例えば、図1Aおよび図1B中、クラスフィールド168に対し、それぞれクラスA 168AおよびクラスB 168B）。

[クラスAの命令テンプレート]

【0042】

クラスAのメモリアクセスなし105命令テンプレートの場合、アルファフィールド152はRSフィールド152Aとして解釈され、RSフィールド152Aの内容が、異なる拡張演算タイプのうちどれが実行されるべきか（例えば、ラウンド152A.1およびデータ変換152A.2がそれぞれ、メモリアクセスなし、ラウンドタイプ演算110命令テンプレートおよびメモリアクセスなし、データ変換タイプ演算115命令テンプレートに対し指定される）を区別し、一方で、ベータフィールド154は指定されるタイプの演算のうちどれが実行されるべきかを区別する。メモリアクセスなし105命令テンプレートには、スケールフィールド160、変位フィールド162Aおよび変位スケールフィールド162Bは存在しない。

20

[メモリアクセスなし命令テンプレート 完全ラウンド制御タイプ演算]

【0043】

メモリアクセスなしの完全ラウンド制御タイプ演算110命令テンプレートでは、ベータフィールド154はラウンド制御フィールド154Aとして解釈され、ラウンド制御フィールド154Aの内容は静的ラウンドを提供する。本発明に記載の実施形態においては、ラウンド制御フィールド154Aは、すべての浮動小数点の例外を抑制（SAE）フィールド156およびラウンド演算制御フィールド158を含み、一方で、代替的な実施形態は、これら両方の概念をサポートしてよく、且つこれら両方の概念を同一フィールドにエンコードしてよく、または代替的な実施形態はこれらの概念/フィールドのうち的一方または他方のみを有してよい（例えば、ラウンド演算制御フィールド158のみを有してよい）。

30

【0044】

SAEフィールド156。その内容が、例外イベント報告を無効にするか否かを区別する。SAEフィールド156の内容が、抑制が有効になっていることを示す場合、特定の命令は、あらゆる種類の浮動小数点例外フラグを報告せず、浮動小数点例外ハンドラを発生させない。

40

【0045】

ラウンド演算制御フィールド158。その内容が、ラウンド演算グループ（例えば、切り上げ、切り捨て、ゼロへの丸めおよび最近値への丸め）のうちどれが実行されるかを区別する。故に、ラウンド演算制御フィールド158は、命令単位で、ラウンドモードの変更を可能にする。本発明の一実施形態において、プロセッサがラウンドモードを指定するための制御レジスタを含む場合、ラウンド演算制御フィールド150の内容で、そのレジスタ値を上書きする。

50

【 0 0 4 6 】

[メモリアクセスなし命令テンプレート データ変換タイプ演算]

【 0 0 4 7 】

メモリアクセスなしのデータ変換タイプ演算 1 1 5 命令テンプレートでは、ベータフィールド 1 5 4 はデータ変換フィールド 1 5 4 B として解釈され、データ変換フィールド 1 5 4 B の内容が、複数のデータ変換（例えば、データ変換なし、スウィズル、ブロードキャスト）のうちどれが実行されるべきかを区別する。

【 0 0 4 8 】

クラス A のメモリアクセス 1 2 0 命令テンプレートの場合、アルファフィールド 1 5 2 はエビクションヒントフィールド 1 5 2 B として解釈され、エビクションヒントフィールド 1 5 2 B の内容が、エビクションヒントのうちどれが使用されるべきかを区別し（図 1 A 中、一時的 1 5 2 B . 1 および非一時的 1 5 2 B . 2 がそれぞれ、メモリアクセスの一時的 1 2 5 命令テンプレートおよびメモリアクセスの非一時的 1 3 0 命令テンプレートに対し指定される）、一方で、ベータフィールド 1 5 4 はデータ操作フィールド 1 5 4 C として解釈され、データ操作フィールド 1 5 4 C の内容が、複数のデータ操作演算（プリミティブとしても知られる）のうちどれが実行されるべきかを区別する（例えば、操作なし、ブロードキャスト、ソースのアップコンバージョンおよびデスティネーションのダウンコンバージョン）。メモリアクセス 1 2 0 命令テンプレートは、スケールフィールド 1 6 0 を含み、随意で変位フィールド 1 6 2 A または変位スケールフィールド 1 6 2 B を含む。

10

20

【 0 0 4 9 】

ベクトルメモリ命令は、変換サポートを用いて、メモリからのベクトルロードおよびメモリへのベクトルストアを実行する。通常のベクトル命令の場合と同様、ベクトルメモリ命令は、データ要素全体でデータをメモリから / メモリへ転送し、実際に転送される要素は、書き込みマスクとして選択されるベクトルマスクの内容によって記述されている。

[メモリアクセス命令テンプレート 一時的]

【 0 0 5 0 】

一時的データとは、キャッシュの利益を十分得るべく、間もなく再使用される可能性の高いデータのことである。しかしながら、これはヒントであり、異なるプロセッサは、ヒントを完全に無視することを含め、それを異なる方法で実装してよい。

30

[メモリアクセス命令テンプレート 非一時的]

【 0 0 5 1 】

非一時的データとは、第 1 のレベルキャッシュにおけるキャッシュから利益を十分得るために、間もなく再利用される可能性の低いデータのことであり、エビクションのための優先度が付与されるべきである。しかしながら、これはヒントであり、異なるプロセッサは、ヒントを完全に無視することを含め、それを異なる方法で実装してよい。

[クラス B の命令テンプレート]

【 0 0 5 2 】

クラス B の命令テンプレートの場合、アルファフィールド 1 5 2 は書き込みマスク制御 (Z) フィールド 1 5 2 C として解釈され、書き込みマスク制御 (Z) フィールド 1 5 2 C の内容が、書き込みマスクフィールド 1 7 0 によって制御される書き込みマスキングが、マージであるべきか、またはゼロイングであるべきかを区別する。

40

【 0 0 5 3 】

クラス B のメモリアクセスなし 1 0 5 命令テンプレートの場合、ベータフィールド 1 5 4 の一部は R L フィールド 1 5 7 A として解釈され、R L フィールド 1 5 7 A の内容が、異なる拡張演算タイプのうちどれが実行されるべきかを区別し（例えば、ラウンド 1 5 7 A . 1 およびベクトル長 (V S I Z E) 1 5 7 A . 2 がそれぞれ、メモリアクセスなし、書き込みマスク制御、部分的なラウンド制御タイプ演算 1 1 2 命令テンプレートおよびメモリアクセスなし、書き込みマスク制御、V S I Z E タイプ演算 1 1 7 命令テンプレートに対し指定される）、一方で、ベータフィールド 1 5 4 の残部が、指定されるタイプの演

50

算のうちどれが実行されるべきかを区別する。メモリアクセスなし105命令テンプレートには、スケールフィールド160、変位フィールド162Aおよび変位スケールフィールド162Bが存在しない。

【0054】

メモリアクセスなし、書き込みマスク制御、部分的ラウンド制御タイプ演算110命令テンプレートでは、ベータフィールド154の残部はラウンド演算フィールド159Aとして解釈され、例外イベント報告が無効にされる（特定の命令は、あらゆる種類の浮動小数点例外フラグを報告せず、浮動小数点例外ハンドラを発生させない）。

【0055】

ラウンド演算制御フィールド159A。まさにラウンド演算制御フィールド158と同様、その内容が、ラウンド演算グループ（例えば、切り上げ、切り捨て、ゼロへの丸めおよび最近値への丸め）のうちどれが実行されるかを区別する。故に、ラウンド演算制御フィールド159Aは、命令単位で、ラウンドモードの変更を可能にする。プロセッサがラウンドモードを指定するための制御レジスタを含む場合の本発明の一実施形態において、ラウンド演算制御フィールド150の内容で、そのレジスタ値を上書きする。

10

【0056】

メモリアクセスなし、書き込みマスク制御、VSIZEタイプ演算117命令テンプレートでは、ベータフィールド154の残部はベクトル長フィールド159Bとして解釈され、ベクトル長フィールド159Bの内容が、複数のデータベクトル長のうちのどれ（例えば、128、256または512バイト）に実行されるべきかを区別する。

20

【0057】

クラスBのメモリアクセス120命令テンプレートの場合、ベータフィールド154の一部はブロードキャストフィールド157Bとして解釈され、ブロードキャストフィールド157Bの内容が、ブロードキャストタイプのデータ操作演算が実行されるか否かを区別し、一方で、ベータフィールド154の残部はベクトル長フィールド159Bとして解釈される。メモリアクセス120命令テンプレートは、スケールフィールド160を含み、随意で変位フィールド162Aまたは変位スケールフィールド162Bを含む。

【0058】

汎用ベクトル向け命令フォーマット100に関しては、フルオペコードフィールド174は、フォーマットフィールド140、ベース演算フィールド142およびデータ要素幅フィールド164を含むように表示されている。一実施形態は、フルオペコードフィールド174がこれらのフィールドのうちすべてを含むように示されているものの、これらのフィールドのすべてをサポートしない実施形態においては、フルオペコードフィールド174は、これらのフィールドのすべてより少ない数を含む。フルオペコードフィールド174は、オペレーションコード（オペコード）を提供する。

30

【0059】

拡張演算フィールド150、データ要素幅フィールド164および書き込みマスクフィールド170は、汎用ベクトル向け命令フォーマット内でこれらの機能が、命令単位で指定されることを可能にする。

【0060】

書き込みマスクフィールドおよびデータ要素幅フィールドの組み合わせで、異なるデータ要素幅に基づいてマスクが適用されることを可能にするタイプの命令を作成する。

40

【0061】

クラスAおよびクラスB内に存在する様々な命令テンプレートは、異なる状況において有益である。本発明のいくつかの実施形態において、あるプロセッサ内の異なる複数のプロセッサまたは異なるコアが、クラスAのみ、クラスBのみ、またはこれら両方のクラスをサポートしてよい。例えば、汎用コンピューティング向けの高性能な汎用アウトオブオーダーコアはクラスBのみをサポートしてよく、主にグラフィックおよび/または科学技術（スループット）コンピューティング向けのコアはクラスAのみをサポートしてよく、これら両方向けのコアは両方をサポートしてよい（もちろん、両方のクラスのテンプレート

50

および命令がいくつか混在したものを有するが、両方のクラスのすべてのテンプレートおよび命令を有さないコアは、本発明の範囲内に属する)。また、単一のプロセッサが複数のコアを含んでよく、それらのすべてが同一クラスをサポートし、またはそれらのうち異なるコアが異なるクラスをサポートする。例えば、別個のグラフィックコアおよび汎用コアを備えるプロセッサでは、主にグラフィックおよび/または科学技術コンピューティング向けのグラフィックコアのうちの1つはクラスAのみをサポートしてよく、一方で、汎用コアのうちの1または複数は、クラスBのみをサポートする、汎用コンピューティング向けのアウトオブオーダー実行およびレジスタリネーミングを備えた高性能な汎用コアであってよい。別個のグラフィックコアを有さない別のプロセッサは、クラスAおよびクラスBの両方をサポートする1または複数の汎用インオーダーまたはアウトオブオーダーコアを含んでよい。もちろん、本発明の異なる実施形態において、一方のクラスに属する諸機能が、他方のクラスに実装されてもよい。高水準言語で記述されるプログラムは、様々な異なる実行可能な形式になされるであろう(例えば、ジャストインタイムコンパイルまたは静的コンパイル)。それらの形式としては、1)実行のためにターゲットプロセッサによってサポートされるクラスの命令のみを有する形式、または2)すべてのクラスの命令の異なる組み合わせを使用して記述された代替的なルーチンを有し且つ現在コードを実行中のプロセッサによってサポートされる命令に基づき、実行するルーチンを選択する制御フローコードを有する形式が含まれる。

B. [例示的な特定ベクトル向け命令フォーマット]

【0062】

図2は、本発明の実施形態による、例示的な特定ベクトル向け命令フォーマットを示すブロック図である。図2は特定ベクトル向け命令フォーマット200を示す。特定ベクトル向け命令フォーマット200は、場所、サイズ、解釈およびフィールド順序に加え、これらのフィールドの一部の値を指定するという意味において特定のである。特定ベクトル向け命令フォーマット200は、x86命令セットを拡張するために使用されてよく、よって、当該フィールドのうちのいくつかは、既存のx86命令セットおよびその拡張機能(例えば、AVX)で使用されるフィールドと類似または同一である。このフォーマットは、いくつかの拡張機能を備えた既存のx86命令セットのプレフィクスエンコーディングフィールド、リアルオペコードバイトフィールド、MOD R/Mフィールド、SIBフィールド、変位フィールドおよび即値フィールドと、整合性が維持されている。図1のフィールドが図2のどのフィールドにマッピングされるかが図示されている。

【0063】

本発明の実施形態は、例示目的で、汎用ベクトル向け命令フォーマット100に照らし特定ベクトル向け命令フォーマット200に関し説明されているものの、本発明は特許請求される場合を除き、特定ベクトル向け命令フォーマット200には限定されないことを理解されたい。例えば、特定ベクトル向け命令フォーマット200は特定のサイズのフィールドを有するように図示されているものの、汎用ベクトル向け命令フォーマット100は、様々なフィールドについて様々な考え得るサイズを想定している。特定の例示であるが、データ要素幅フィールド164は、特定ベクトル向け命令フォーマット200では1ビットフィールドとして図示されているものの、本発明はそのようには限定されない(すなわち、汎用ベクトル向け命令フォーマット100は、データ要素幅フィールド164の他のサイズを想定している)。

【0064】

特定ベクトル向け命令フォーマット200は、以下に挙げられるフィールドを図2Aに図示される順序で含む。

【0065】

EVEXプレフィクス(バイト0 3)202。これは4バイト形式でエンコードされる。

【0066】

フォーマットフィールド140(EVEXバイト0、ビット[7:0])。第1のバイ

10

20

30

40

50

ト (E V E X バイト 0) はフォーマットフィールド 1 4 0 であり、フォーマットフィールド 1 4 0 は $0 \times 6 2$ を含む (本発明の一実施形態において、ベクトル向け命令フォーマットを区別するために使用される一意の値)。

【 0 0 6 7 】

第 2 から第 4 のバイト (E V E X バイト 1 3) は、特定の機能を提供する複数のビットフィールドを含む。

【 0 0 6 8 】

R E X フィールド 2 0 5 (E V E X バイト 1、ビット [7 5])。これは E V E X . R ビットフィールド (E V E X バイト 1、ビット [7] R)、E V E X . X ビットフィールド (E V E X バイト 1、ビット [6] X) および 1 5 7 B E X バイト 1、ビット [5] B から成る。E V E X . R ビットフィールド、E V E X . X ビットフィールドおよび E V E X . B ビットフィールドは、対応する V E X ビットフィールドと同一の機能を提供し、それらは 1 の補数形式を使用してエンコードされ、すなわち Z M M 0 は 1 1 1 1 B としてエンコードされ、Z M M 1 5 は 0 0 0 0 B としてエンコードされる。命令の他のフィールドは、レジスタインデックスの下位 3 ビットを当該技術分野で既知の方法 (r r r、x x x および b b b) でエンコードし、その結果、R r r r、X x x x および B b b b が、E V E X . R、E V E X . X および E V E X . B を追加することによって形成されてよい。

【 0 0 6 9 】

R E X' フィールド 1 1 0。これは R E X' フィールド 1 1 0 の第 1 の部分であり、拡張 3 2 レジスタセットの上位 1 6 または下位 1 6 のいずれかをエンコードするために使用される E V E X . R' ビットフィールド (E V E X バイト 1、ビット [4] R') である。本発明の一実施形態において、以下に示される他のものと共にこのビットは、ビット反転フォーマットで格納され、B O U N D 命令から区別 (周知の $x 8 6$ の 3 2 ビットモードで) される。B O U N D 命令のリアルオペコードバイトは 6 2 であるが、M O D R / M フィールド (後述) 内では、M O D フィールドの値 1 1 を受け付けない。本発明の代替的な実施形態は、このビットおよび後述される他のビットを反転フォーマットで格納しない。値 1 が使用され、下位 1 6 個のレジスタをエンコードする。換言すると、E V E X . R'、E V E X . R および他のフィールドの他の R R R を組み合わせると、R' R r r r r が形成される。

【 0 0 7 0 】

オペコードマップフィールド 2 1 5 (E V E X バイト 1、ビット [3 : 0] m m m m)。その内容が暗示される先頭オペコードバイト (O F、O F 3 8、または O F 3) をエンコードする。

【 0 0 7 1 】

データ要素幅フィールド 1 6 4 (E V E X バイト 2、ビット [7] W)。これは E V E X . W という表記で表される。E V E X . W が使用され、データタイプの粒度 (サイズ) を定義する (3 2 ビットデータ要素または 6 4 ビットデータ要素のいずれか)。

【 0 0 7 2 】

E V E X . v v v v 2 2 0 (E V E X バイト 2、ビット [6 : 3] v v v v)。E V E X . v v v v の役割は以下を含んでよい。1) E V E X . v v v v は第 1 のソースレジスタオペランドを指定された反転 (1 の補数) 形式にエンコードし、E V E X . v v v v は 2 またはそれより多いソースオペランドを持つ命令に対し有効である。2) E V E X . v v v v はデスティネーションレジスタオペランドを、特定のベクトルシフト用の指定された 1 の補数形式にエンコードする。または 3) E V E X . v v v v はいずれのオペランドもエンコードせず、当該フィールドは予約され、1 1 1 1 b を含むべきである。故に、E V E X . v v v v フィールド 2 2 0 は、反転 (1 の補数) 形式で格納された第 1 のソースレジスタ指定子の 4 つの下位ビットをエンコードする。命令に応じて、追加の異なる E V E X ビットフィールドが使用され、指定子サイズを 3 2 個のレジスタに拡張する。

【 0 0 7 3 】

EVEX.U 168クラスフィールド(EVEXバイト2、ビット[2] U)。EVEX.U = 0の場合、それはクラスAまたはEVEX.U 0を示す。EVEX.U = 1の場合、それはクラスBまたはEVEX.U 1を示す。

【0074】

プレフィクスエンコーディングフィールド225(EVEXバイト2、ビット[1:0] pp)。これは、ベース演算フィールドの追加のビットを提供する。EVEXプレフィクスフォーマットにおけるレガシSSE命令のサポートの提供に加え、これはまた、SIMDプレフィクスのコンパクト化の利点を有する(SIMDプレフィクスを表わすために1バイトを要求する代わりに、EVEXプレフィクスは2ビットのみを要求する)。一実施形態において、レガシフォーマットおよびEVEXプレフィクスフォーマットの両方において、SIMDプレフィクス(66H、F2H、F3H)を使用するレガシSSE命令をサポートすべく、これらのレガシSIMDプレフィクスは、SIMDプレフィクスエンコーディングフィールドにエンコードされる。これらのレガシSIMDプレフィクスは、デコーダのPLAに提供される前に、ランタイムにレガシSIMDプレフィクスに拡張される(よって、PLAは、変更なしで、これらのレガシ命令のレガシフォーマットおよびEVEXフォーマットの両方を実行可能である)。より新しい命令はEVEXプレフィクスエンコーディングフィールドの内容を直接オペコード拡張として使用できるものの、特定の実施形態は、整合性のために同様の方法で拡張させるが、これらのレガシSIMDプレフィクスによって指定される異なる手段を可能にする。代替的な実施形態は、2ビットSIMDプレフィクスエンコードをサポートするように、つまり拡張を要求しないように、PLAを再設計してよい。

10

20

【0075】

アルファフィールド152(EVEXバイト3、ビット[7] EH。これはEVEX.EH、EVEX.rs、EVEX.RL、EVEX.書き込みマスク制御およびEVEX.Nとしても知られる。また を用いて図示)。上記の通り、このフィールドはコンテキストに特有のものである。

【0076】

ベータフィールド154(EVEXバイト3、ビット[6:4] SSS。これはEVEX.s₂₋₀、EVEX.r₂₋₀、EVEX.rr1、EVEX.LL0、EVEX.LLBとしても知られる。また を用いて図示)。上記の通り、このフィールドはコンテキストに特有のものである。

30

【0077】

REX'フィールド110。これはREX'フィールドの残部であり、REX'フィールド110は、拡張された32個のレジスタセットの上位16個または下位16個のいずれかをエンコードするために使用され得るEVEX.V'ビットフィールド(EVEXバイト3、ビット[3] V')である。このビットは、ビット反転フォーマットで格納される。値1が使用され、下位16個のレジスタをエンコードする。換言すると、EVEX.V'、EVEX.vvvvを組み合わせることにより、V'VVVVが形成される。

【0078】

書き込みマスクフィールド170(EVEXバイト3、ビット[2:0] kkk)。上記の通り、その内容が書き込みマスクレジスタ内のレジスタのインデックスを指定する。本発明の一実施形態において、特定の値EVEX.kkk = 000は、特定の命令について書き込みマスクが使用されないことを暗示する特別な動作を有する(これは、すべて1にハードワイヤードされた書き込みマスクの使用またはマスキングハードウェアを迂回するハードウェアの使用を含む、様々な方法で実装されてよい)。

40

【0079】

リアルオペコードフィールド230(バイト4)は、オペコードバイトとしても知られる。このフィールドで、オペコードの一部が指定される。

【0080】

MOD R/Mフィールド240(バイト5)は、MODフィールド242、Regフ

50

フィールド244およびR/Mフィールド246を含む。上記の通り、MODフィールド242の内容が、メモリアクセス操作およびメモリアクセスなし操作間を区別する。Regフィールド244の役割は、デスティネーションレジスタオペランド若しくはソースレジスタオペランドのいずれかをエンコードすること、または、オペコード拡張として扱われ、命令オペランドをエンコードするために使用されないこと、という2つの状況に要約できる。R/Mフィールド246の役割としては、メモリアドレスを参照する命令オペランドをエンコードすること、またはデスティネーションレジスタオペランド若しくはソースレジスタオペランドのいずれかをエンコードすることが含まれてよい。

【0081】

スケール、インデックス、ベース(SIB)バイト(バイト6)。上記の通り、スケールフィールド150の内容は、メモリアドレス生成に使用される。SIB.xxx254およびSIB.bbb256。これらのフィールドの内容は、レジスタインデックスxxxおよびBbbbに関して記載済みである。

10

【0082】

変位フィールド162A(バイト710)。MODフィールド242に10が含まれる場合、バイト710は変位フィールド162Aであり、変位フィールド162Aはレガシ32ビット変位(disp32)と同様に動作し、バイト粒度で動作する。

【0083】

変位係数フィールド162B(バイト7)。MODフィールド242に01が含まれる場合、バイト7は変位係数フィールド162Bである。このフィールドの場所は、レガシx86命令セットの8ビット変位(disp8)の場所と同一であり、レガシx86命令セットの8ビット変位(disp8)はバイト粒度で動作する。disp8は符号拡張されるので、disp8は-128~127バイトオフセット間のアドレス指定のみ可能である。64バイトのキャッシュラインに関しては、disp8は4つの実際に有用な値、-128、-64、0および64のみに設定可能な8ビットを使用する。通常、さらに広い範囲が必要であるので、disp32が使用されるが、disp32は4バイトを必要とする。disp8およびdisp32と対照的に、変位係数フィールド162Bはdisp8と再解釈される。変位係数フィールド162Bを使用する場合、実際の変位は、メモリオペランドアクセス(N)のサイズで乗算された変位係数フィールドの内容によって決定される。このタイプの変位は、disp8xNと称される。これは、平均的な命令の長さ(変位に使用されるのは1バイトであるが、はるかにより広い範囲を備える)を低減する。このような圧縮された変位は、有効な変位は、メモリアクセスの粒度の倍数であり、従って、アドレスオフセットの冗長下位ビットはエンコードの必要がないという前提に基づいている。換言すると、変位係数フィールド162Bは、レガシx86命令セットの8ビット変位に置き換わる。故に、変位係数フィールド162Bは、disp8がdisp8xNにオーバーロードされる点のみを除いては、x86命令セットの8ビット変位と同じ方法でエンコードされる(よって、ModRM/SIBエンコードルールの変更はない)。換言すると、ハードウェアによる変位値の解釈のみを除き、エンコーディングルールまたはエンコーディング長に変更はない(バイト単位のアドレスオフセットを取得するために、メモリオペランドのサイズだけ変位をスケールリングする必要がある)。

20

30

40

【0084】

即値フィールド172は、上記の通り動作する。

[フルオペコードフィールド]

【0085】

図2Bは、本発明の一実施形態による、特定ベクトル向け命令フォーマット200のフルオペコードフィールド174を構成するフィールドを示すブロック図である。具体的には、フルオペコードフィールド174は、フォーマットフィールド140、ベース演算フィールド142およびデータ要素幅(W)フィールド164を含む。ベース演算フィールド142は、プレフィクスエンコーディングフィールド225、オペコードマップフィールド215およびリアルオペコードフィールド230を含む。

50

[レジスタインデックスフィールド]

【 0086 】

図2Cは、本発明の一実施形態による、特定ベクトル向け命令フォーマット200のレジスタインデックスフィールド144を構成するフィールドを示すブロック図である。具体的には、レジスタインデックスフィールド144は、REXフィールド205、REX'フィールド210、MODR/M.regフィールド244、MODR/M.r/mフィールド246、VVVVフィールド220、xxxフィールド254およびbbbフィールド256を含む。

[拡張演算フィールド]

【 0087 】

図2Dは、本発明の一実施形態による、特定ベクトル向け命令フォーマット200の拡張演算フィールド150を構成するフィールドを示すブロック図である。クラス(U)フィールド168が0を含む場合、それはEVEX.U0(クラスA 168A)を表わす。クラス(U)フィールド168が1を含む場合、それはEVEX.U1(クラスB 168B)を表わす。U=0で且つMODフィールド242が11を含む場合(メモリアクセスなし操作を意味)、アルファフィールド152(EVEXバイト3、ビット[7]EH)は、rsフィールド152Aとして解釈される。rsフィールド152Aが1を含む場合(ラウンド152A.1)、ベータフィールド154(EVEXバイト3、ビット[6:4]SSS)はラウンド制御フィールド154Aとして解釈される。ラウンド制御フィールド154Aは、1ビットのSAEフィールド156および2ビットのラウンド演算フィールド158を含む。rsフィールド152Aが0を含む場合(データ変換152A.2)、ベータフィールド154(EVEXバイト3、ビット[6:4]SSS)は3ビットのデータ変換フィールド154Bとして解釈される。U=0で且つMODフィールド242が00、01または10を含む場合(メモリアクセス操作を意味)、アルファフィールド152(EVEXバイト3、ビット[7]EH)は、エビクションヒント(EH)フィールド152Bとして解釈され、ベータフィールド154(EVEXバイト3、ビット[6:4]SSS)は3ビットのデータ操作フィールド154Cとして解釈される。

【 0088 】

U=1の場合、アルファフィールド152(EVEXバイト3、ビット[7]EH)は、書き込みマスク制御(Z)フィールド152Cとして解釈される。U=1で且つMODフィールド242が11を含む場合(メモリアクセスなし操作を意味)、ベータフィールド154の一部(EVEXバイト3、ビット[4]S₀)は、RLフィールド157Aとして解釈される。RLフィールド157Aが1を含む場合(ラウンド157A.1)、ベータフィールド154の残部(EVEXバイト3、ビット[6:5]S_{2:1})はラウンド演算フィールド159Aとして解釈され、一方で、RLフィールド157Aが0を含む場合(VSIZE 157.A2)、ベータフィールド154の残部(EVEXバイト3、ビット[6:5]S_{2:1})は、ベクトル長フィールド159B(EVEXバイト3、ビット[6:5]L_{1:0})として解釈される。U=1で且つMODフィールド242が00、01または10を含む場合(メモリアクセス操作を意味)、ベータフィールド154(EVEXバイト3、ビット[6:4]SSS)は、ベクトル長フィールド159B(EVEXバイト3、ビット[6:5]L_{1:0})およびブロードキャストフィールド157B(EVEXバイト3、ビット[4]B)として解釈される。

C. [例示的なレジスタアーキテクチャ]

【 0089 】

図3は、本発明の一実施形態による、レジスタアーキテクチャ300のブロック図である。図示される実施形態には、512ビット幅の32個のベクトルレジスタ310がある。これらのレジスタは、zmm0からzmm31と参照符号が付されている。下位16個のzmmレジスタの下位256ビットは、レジスタymm0~ymm16に重なっている。下位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

10

20

30

40

50

)は、レジスタ $xmm0 \sim xmm15$ に重なっている。特定ベクトル向け命令フォーマット 200 は、これらの重なったレジスタファイルに対し、以下の表に示されるように動作する。

【表 1】

調整可能 ベクトル長	クラス	演算	レジスタ
ベクトル長 フィールド159Bを 含まない 命令テンプレート	A(図1A; U=0)	110, 115, 125, 130	zmmレジスタ (ベクトル長は64バイト)
	B(図1B; U=1)	112	zmmレジスタ (ベクトル長は64バイト)
ベクトル長 フィールド159Bを 含む命令 テンプレート	B(図1B; U=1)	117, 127	ベクトル長フィールド159Bに 応じて、zmm、ymmまたは xmmレジスタ(ベクトル長は 64バイト、32バイトまたは 16バイト)

10

20

【0090】

換言すると、ベクトル長フィールド159Bは、最大長から1または複数の他のより短い長さまでの範囲内から選択する。ここで、当該より短い長さの各々は、1つ前の長さの半分であり、ベクトル長フィールド159Bを持たない命令テンプレートは、最大ベクトル長に対し演算を行う。さらに、一実施形態において、特定ベクトル向け命令フォーマット200のクラスB命令テンプレートは、パックド単精度/倍精度浮動小数点データまたはスカラ単精度/倍精度浮動小数点データおよびパックド整数データまたはスカラ整数データに対し、演算を行う。スカラ演算とは、zmm/ymm/xmmレジスタ内の最下位のデータ要素の位置で実行される演算である。実施形態に応じ、より上位のデータ要素の位置は、命令前と同じに保持されるか、ゼロにされるかのいずれかである。

30

【0091】

図示された実施形態中の書き込みマスクレジスタ315には、8個の書き込みマスクレジスタ(k0からk7)が存在し、各々64ビットのサイズである。代替的な実施形態において、書き込みマスクレジスタ315は、16ビットのサイズである。上記の通り、本発明の一実施形態において、ベクトルマスクレジスタk0は書き込みマスクとして使用不可である。通常k0を示すエンコーディングが書き込みマスクに使用される場合、それは0xFFFFのハードワイヤードされた書き込みマスクを選択し、有効にその命令に対し書き込みマスクを無効にする。

【0092】

図示された実施形態中の汎用レジスタ325には、メモリオペランドをアドレス指定するために既存のx86アドレス指定モードと共に使用される16個の64ビットの汎用レジスタが存在する。これらのレジスタは、RAX、RBX、RCX、RDX、RBP、RSI、RDI、RSPおよびR8~R15という名称で参照される。

40

【0093】

図示された実施形態中、スカラ浮動小数点スタックレジスタファイル(x87スタック)345について、MMXパックド整数フラットレジスタファイル350というエイリアスが示されているが、x87スタックは、x87命令セット拡張を使用して、32/64/80ビットの浮動小数点データにスカラ浮動小数点演算を実行するために使用される8個の要素のスタックである。MMXレジスタは、64ビットのパックド整数データに対し演算を実行するために使用されるが、MMXレジスタおよびXMMレジスタ間で実行され

50

るいくつかの演算のためのオペランドを保持するためにも使用される。

【 0 0 9 4 】

本発明の代替的な実施形態は、より範囲の広いまたは狭いレジスタを使用してよい。また、本発明の代替的な実施形態は、より多い、より少ないまたは異なるレジスタファイルおよびレジスタを使用してもよい。

D . [例示的なコアアーキテクチャ、プロセッサおよびコンピュータアーキテクチャ]

【 0 0 9 5 】

プロセッサコアは、異なる方法で、異なる目的のために、および異なるプロセッサ内に実装されてよい。例えば、このようなコアの実装としては次のようなものが含まれてよい。すなわち、1) 汎用コンピューティング用の汎用インオーダーコアインオーダーコア、2) 汎用コンピューティング用の高性能汎用アウトオブオーダーコア、3) 主にグラフィックおよび/または科学技術(スループット)コンピューティング用の専用コア。異なるプロセッサの実装としては、次のようなものが含まれてよい。すなわち、1) 汎用コンピューティング用の1または複数の汎用インオーダーコアおよび/または汎用コンピューティング用の1または複数の汎用アウトオブオーダーコアを含むCPU、および2) 主にグラフィックおよび/または科学技術(スループット)用の1または複数の専用コアを含むコプロセッサ。このような異なるプロセッサは、異なるコンピュータシステムアーキテクチャをもたらし、それには次のようなものが含まれてよい。すなわち、1) CPUとは別個のチップ上のコプロセッサ、2) CPUと同一パッケージ内の別個のダイ上にあるコプロセッサ、3) CPUと同一ダイ上のコプロセッサ(この場合、このようなコプロセッサは、統合グラフィックおよび/または科学技術(スループット)ロジック等の専用ロジック、または専用コアと呼ばれることがある)および、4) 同一のダイ上に上記CPU(アプリケーションコアまたはアプリケーションプロセッサと呼ばれることがある)、上記コプロセッサおよび追加の機能を含み得るシステムオンチップ。例示的なコアアーキテクチャが次に記載され、その後例示的なプロセッサおよびコンピュータアーキテクチャが続く。

【 0 0 9 6 】

図4Aは、本発明の実施形態による、例示的なインオーダーパイプラインおよび例示的なレジスタリネーミング、アウトオブオーダー発行/実行パイプラインの両方を示すブロック図である。図4Bは、本発明の実施形態による、プロセッサに含まれる、インオーダーアーキテクチャコアに係る例示的な実施形態および例示的なレジスタリネーミング、アウトオブオーダー発行/実行アーキテクチャコアの両方を示すブロック図である。図4A~図4B中の実線ボックスは、インオーダーパイプラインおよびインオーダーコアを示すが、オプションで追加される破線ボックスは、レジスタリネーミング、アウトオブオーダー発行/実行パイプラインおよびコアを示す。インオーダーの態様はアウトオブオーダー態様のサブセットであると想定して、アウトオブオーダー態様について以下記載する。

【 0 0 9 7 】

図4A中、プロセッサパイプライン400は、フェッチステージ402、長さデコードステージ404、デコードステージ406、割り当てステージ408、リネーミングステージ410、スケジューリング(ディスパッチまたは発行としても知られる)ステージ412、レジスタ読み取り/メモリ読み取りステージ414、実行ステージ416、ライトバック/メモリ書き込みステージ418、例外処理ステージ422およびコミットステージ424が含まれる。

【 0 0 9 8 】

図4Bは、実行エンジンユニット450に連結されたフロントエンドユニット430を含むプロセッサコア490を示し、フロントエンドユニット430および実行エンジンユニット450の両方はメモリユニット470に連結されている。コア490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若しくは代替的なコアタイプであってよい。さらなる別のオプションとして、コア490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

ィック処理ユニット (G P G P U) コア、グラフィックコア等のような専用コアであってよい。

【 0 0 9 9 】

フロントエンドユニット 4 3 0 は、命令キャッシュユニット 4 3 4 に連結された分岐予測ユニット 4 3 2 を含み、命令キャッシュユニット 4 3 4 は、命令トランシェーションルックアサイドバッファ (T L B) 4 3 6 に連結され、 T L B 4 3 6 は命令フェッチユニット 4 3 8 に連結され、命令フェッチユニット 4 3 8 はデコードユニット 4 4 0 に連結される。デコードユニット 4 4 0 (すなわちデコーダ) は命令をデコードしてよく、および、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てよく、これらは元の命令からデコードされ、あるいは元の命令を反映し、あるいは元の命令から派生する。デコードユニット 4 4 0 は、様々な異なるメカニズムを使用して実装されてよい。好適なメカニズムの例としては、限定はされないがルックアップテーブル、ハードウェア実装、プログラマブルロジックアレイ (P L A)、マイクロコードリードオンリメモリ (R O M) 等が含まれる。一実施形態において、コア 4 9 0 は、特定のマクロ命令のためのマイクロコードを格納 (例えば、デコードユニット 4 4 0 内またはフロントエンドユニット 4 3 0 内部) するマイクロコード R O M または他の媒体を含む。デコードユニット 4 4 0 は、実行エンジンユニット 4 5 0 内のリネーム / アロケータユニット 4 5 2 に連結される。

10

【 0 1 0 0 】

実行エンジンユニット 4 5 0 は、リタイアメントユニット 4 5 4 に連結されたリネーム / アロケータユニット 4 5 2 および 1 または複数のスケジューラユニット 4 5 6 のセットを含む。スケジューラユニット 4 5 6 は、予約ステーション、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わす。スケジューラユニット 4 5 6 は物理レジスタファイルユニット 4 5 8 に連結される。物理レジスタファイルユニット 4 5 8 の各々は、1 または複数の物理レジスタファイルを表わし、それらの異なる 1 つ 1 つは、1 または複数の異なるデータタイプを格納する。そのようなものとしては、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、状態 (例えば、実行される次の命令のアドレスである命令ポインタ) 等が挙げられる。一実施形態において、物理レジスタファイルユニット 4 5 8 はベクトルレジスタユニット、書き込みマスクレジスタユニットおよ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、アーキテクチャのベクトルレジスタ、ベクトルマスクレジスタおよび汎用レジスタを提供してよい。レジスタリネーミングおよびアウトオブオーダー実行が実装され得る様々な方法を示すため、物理レジスタファイルユニット 4 5 8 がリタイアメントユニット 4 5 4 に重ねられている (例えば、リオーダーバッファおよびリタイアメントレジスタファイルを使用する、将来のファイル、履歴バッファおよびリタイアメントレジスタファイルを使用する、レジスタマップおよびレジスタプールを使用する等)。リタイアメントユニット 4 5 4 および物理レジスタファイルユニット 4 5 8 は、実行クラスタ 4 6 0 に連結される。実行クラスタ 4 6 0 は、1 または複数の実行ユニット 4 6 2 のセットおよび 1 または複数のメモリアクセスユニット 4 6 4 のセットを含む。実行ユニット 4 6 2 は、様々な演算 (例えば、シフト、加算、減算、乗算) を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) に行ってよい。いくつかの実施形態は、特定の関数または関数のセットに専用割り当てられた複数の実行ユニットを含んでよく、一方で、他の実施形態は、1 つのみの実行ユニットまたは、それらすべてが全関数を実行する複数の実行ユニットを含んでよい。スケジューラユニット 4 5 6、物理レジスタファイルユニット 4 5 8 および実行クラスタ 4 6 0 が可能性として複数形で図示されているのは、特定の実施形態が特定のタイプのデータ / 演算のために別個のパイプライン (例えば、スカラ整数のパイプライン、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のパイプラインおよび / またはメモリアクセスパイプライン。これらの各々は独自のスケジューラユニット、物理レジスタファイルユニット、および / または実行クラスタを有する。別個

20

30

40

50

のメモリアクセスパイプラインの場合、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464を有する特定の実施形態が実装される)を形成するからである。別個のパイプラインが使用される場合、これらのパイプラインのうちの1または複数アウトオブオーダー発行/実行であってよく、残りはインオーダーであってよいことも理解されたい。

【0101】

メモリアクセスユニット464のセットがメモリユニット470に連結され、メモリユニット470はレベル2(L2)キャッシュユニット476に連結されたデータキャッシュユニット474に連結されたデータTLBユニット472を含む。一例示的な実施形態において、メモリアクセスユニット464は、ロードユニット、ストアアドレスユニット、およびストアデータユニットを含んでよく、これらの各々はメモリユニット470内のデータTLBユニット472に連結される。命令キャッシュユニット434は、メモリユニット470内のレベル2(L2)キャッシュユニット476にさらに連結される。L2キャッシュユニット476は、1または複数の他のレベルのキャッシュに連結され、最終的にメインメモリに連結される。

10

【0102】

例を挙げると、例示的なレジスタリネーミング、アウトオブオーダー発行/実行コアアーキテクチャは、パイプライン400を以下のように実装してよい。すなわち、1)命令フェッチ438がフェッチステージ402および長さデコーディングステージ404を実行する。2)デコードユニット440がデコードステージ406を実行する。3)リネーム/アロケータユニット452が割り当てステージ408およびリネーミングステージ410を実行する。4)スケジューラユニット456がスケジューリングステージ412を実行する。5)物理レジスタファイルユニット458およびメモリユニット470がレジスタ読み取り/メモリ読み取りステージ414を実行する。実行クラスタ460が実行ステージ416を実行する。6)メモリユニット470および物理レジスタファイルユニット458がライトバック/メモリ書き込みステージ418を実行する。7)様々なユニットが例外処理ステージ422に参与してよい。8)リタイアメントユニット454および物理レジスタファイルユニット458がコミットステージ424を実行する。

20

【0103】

コア490は、本明細書に記載の命令を含む、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に追加されたいいくつかの拡張を持つ)、カリフォルニア州サンニールのMIPS TechnologiesのMIPS命令セット、カリフォルニア州サンニールのARM HoldingsのARM命令セット(NEON等のオプションの追加拡張を持つ))をサポートしてよい。一実施形態において、コア490は、バックドデータ命令セット拡張(例えば、AVX1、AVX2)をサポートするロジックを含み、それにより、多くのマルチメディアアプリケーションによって使用される演算がバックドデータを使用して実行されることを可能にする。

30

【0104】

コアは、マルチスレッディング(演算またはスレッドの2または2より多い並列セットの実行)をサポートしてよく、様々な方法でマルチスレッディングを実行してよいことを理解されたい。そのようなものとしては、時分割マルチスレッディング、同時マルチスレッディング(この場合、単一の物理コアは、物理コアが同時にマルチスレッディングを行うスレッドの各々に対し、論理コアを提供する)、またはこれらの組み合わせ(例えば、時分割フェッチおよび時分割デコーディング並びにインテル(登録商標)ハイパースレッディング技術等のそれら以降の同時マルチスレッディング)が含まれる。

40

【0105】

レジスタリネーミングはアウトオブオーダー実行の文脈で説明されているが、レジスタリネーミングはインオーダーアーキテクチャで使用されてよいことを理解されたい。図示されたプロセッサの実施形態はまた、別個の命令キャッシュユニット434およびデータキャッシュユニット474並びに共有L2キャッシュユニット476を含むが、代替的な実施

50

形態は、命令およびデータの両方のための例えば、レベル1 (L1) 内部キャッシュまたは複数のレベルの内部キャッシュのような単一の内部キャッシュを有してよい。いくつかの実施形態において、システムは、内部キャッシュ並びにコアおよび/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んでよい。代替的に、すべてのキャッシュは、コアおよび/またはプロセッサの外部にあってよい。

【0106】

図5A~5Bは、より具体的な例示のインオーダーコアアーキテクチャのブロック図を示し、コア(同一タイプおよび/または異なるタイプの他のコアを含む)はチップ内のいくつかの論理ブロックの1つであろう。その適用に応じ、論理ブロックは、何らかの固有の機能ロジック、メモリI/Oインタフェースおよび他の必要なI/Oロジックを備えた高帯域幅の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。

10

【0107】

図5Aは、本発明の実施形態による、オンダイ相互接続ネットワーク502への接続を備え、且つ、レベル2(L2)キャッシュ504のローカルサブセットを備えた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500は、パックドデータ命令セット拡張を備えたx86命令セットをサポートする。L1キャッシュ506は、キャッシュメモリからスカラユニットおよびベクトルユニットへと読み出す低レイテンシアアクセスを可能にする。一実施形態(設計を簡易化した)において、スカラユニット508およびベクトルユニット510は、別個のレジスタセット(それぞれスカラレジスタ512およびベクトルレジスタ514)を使用し、それらの間で転送されたデータはメモリに書き込まれた後、レベル1(L1)キャッシュ506からリードバックされる一方で、本発明の代替的な実施形態は、異なるアプローチ(例えば、単一のレジスタセットを使用する、またはデータが書き込みおよびリードバックされることなく、2つのレジスタファイル間で転送されることを可能にする通信パスを含む)を使用してよい。

20

【0108】

L2キャッシュのローカルサブセット504は、1つのプロセッサコアにつき1つのローカルサブセットとして、別個の複数のローカルサブセットに分割されるグローバルL2キャッシュの一部である。各プロセッサコアは、プロセッサコア自身のL2キャッシュ504のローカルサブセットへのダイレクトアクセスパスを有する。プロセッサコアによって読み取られたデータは、そのL2キャッシュサブセット504に格納され、当該データは、他のプロセッサコアが、自身のローカルL2キャッシュサブセットにアクセスすると並列的に、迅速にアクセス可能である。プロセッサコアによって書き込まれたデータは、自身のL2キャッシュサブセット504に格納され、必要な場合、他のサブセットからはフラッシュされる。リングネットワークは、共有データのためのコピーレンシを保証する。リングネットワークは双方向であり、プロセッサコア、L2キャッシュおよび他の論理ブロック等のエージェントが、チップ内で互いに通信することを可能にする。各リングデータパスは、一方向当たり1012ビット幅である。

30

【0109】

図5Bは、本発明の実施形態による、図5Aのプロセッサコアの一部の拡大図である。図5Bには、L1キャッシュ504の一部であるL1データキャッシュ506Aに加え、ベクトルユニット510およびベクトルレジスタ514に関しより詳細なものが含まれる。具体的には、ベクトルユニット510は、16幅ベクトル処理ユニット(VPU)(16幅ALU 528を参照)であり、整数命令、単精度浮動命令および倍精度浮動命令のうち1または複数を実行する。VPUは、スイズルユニット520を用いるレジスタ入力のスウィズル、数値変換ユニット522A~Bを用いる数値変換およびメモリ入力での複製ユニット524を用いる複製をサポートする。書き込みマスクレジスタ526は、結果ベクトル書き込みのプレディケートを可能にする。

40

【0110】

図6は、本発明の実施形態による、プロセッサ600のブロック図であり、当該プロセッサは、2以上のコアを有してよく、統合メモリコントローラを有してよく、統合グラフ

50

ックを有してよい。図 6 中の実線ボックスは、単一のコア 6 0 2 A、システムエージェント 6 1 0、1 または複数のバスコントローラユニット 6 1 6 のセットを備えたプロセッサ 6 0 0 を示す一方で、破線ボックスのオプションの追加は、複数のコア 6 0 2 A ~ N、システムエージェントユニット 6 1 0 内の 1 または複数の統合メモリコントローラユニット 6 1 4 のセット、および専用ロジック 6 0 8 を備えた代替的なプロセッサ 6 0 0 を示す。

【 0 1 1 1 】

故に、プロセッサ 6 0 0 の異なる実装は、次のもの、すなわち 1) 統合グラフィックおよび / または科学技術 (スループット) ロジック (1 または複数のコアを含んでよい) である専用ロジック 6 0 8 と、1 または複数の汎用コアであるコア 6 0 2 A ~ N (例えば、汎用インオーダコア、汎用アウトオブオーダコア、それら 2 つの組み合わせ) を有する CPU、2) 主にグラフィックおよび / または科学技術 (スループット) 向けの多数の専用コアであるコア 6 0 2 A ~ N を有するコプロセッサ、並びに 3) 多数の汎用インオーダコアであるコア 6 0 2 A ~ N を有するコプロセッサ、を含んでよい。故に、プロセッサ 6 0 0 は、例えば、ネットワークプロセッサまたは通信プロセッサ、圧縮エンジン、グラフィックプロセッサ、G P G P U (汎用グラフィック処理ユニット)、高スループット多集積コア (M I C) コプロセッサ (3 0 または 3 0 より多いコアを含む)、組み込みプロセッサ等のような汎用プロセッサ、コプロセッサ、または専用プロセッサであってよい。プロセッサは、1 または複数のチップ上に実装されてよい。プロセッサ 6 0 0 は、例えば、B i C M O S、C M O S または N M O S 等の複数のプロセス技術のうちの任意のものを使用する 1 または複数の基板の一部であってよく、および / または当該基板上に実装されてよい。

【 0 1 1 2 】

メモリ階層は、コア内の 1 または複数のレベルのキャッシュ、共有キャッシュユニット 6 0 6 のセットまたは 1 若しくは複数の共有キャッシュユニット 6 0 6、および統合メモリコントローラユニット 6 1 4 のセットに連結された外部メモリ (不図示) を含む。共有キャッシュユニットのセット 6 0 6 は、レベル 2 (L 2)、レベル 3 (L 3)、レベル 4 (L 4) 等の 1 または複数の中レベルキャッシュまたは他のレベルのキャッシュ、ラストレベルキャッシュ (L L C) および / またはそれらの組み合わせを含んでよい。一実施形態において、リングベースの相互接続ユニット 6 1 2 は、統合グラフィックロジック 6 0 8、共有キャッシュユニット 6 0 6 のセットおよびシステムエージェントユニット 6 1 0 / 統合メモリコントローラユニット 6 1 4 を相互接続する一方で、代替的な実施形態は、このようなユニットを相互接続するための任意の数の周知技術を使用してよい。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 6 0 6 およびコア 6 0 2 A ~ N 間で維持される。

【 0 1 1 3 】

いくつかの実施形態において、コア 6 0 2 A ~ N のうちの 1 または複数は、マルチスレーディングが可能である。システムエージェント 6 1 0 は、コア 6 0 2 A ~ N を調整および操作するそれらのコンポーネントを含む。システムエージェントユニット 6 1 0 は、例えば、電力制御ユニット (P C U) およびディスプレイユニットを含んでよい。P C U は、コア 6 0 2 A ~ N および統合グラフィックロジック 6 0 8 の電力状態を統制するために必要なロジックおよびコンポーネントであってよい、またはそれらを含んでよい。ディスプレイユニットは、1 または複数の外部接続されたディスプレイを駆動するためのものである。

【 0 1 1 4 】

コア 6 0 2 A ~ N は、アーキテクチャ命令セットの観点から同種または異種であってよい。すなわち、コア 6 0 2 A ~ N のうち 2 または 2 より多くは、同一命令セットを実行可能であってよいが、他のものはその命令セットのサブセットのみまたは異なる命令セットを実行可能であってよい。

【 0 1 1 5 】

10

20

30

40

50

図7～図10は、例示的なコンピュータ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PC、携帯情報端末、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、デジタル信号プロセッサ(DSP)、グラフィックデバイス、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ーヤ、ハンドヘルドデバイスおよび様々な他の電子デバイスのための当該技術分野で既知の他のシステム設計および構成も好適である。一般的に、本明細書に開示のプロセッサおよび/または他の実行ロジックを組み込み可能な非常に多種多様なシステムまたは電子デバイスが概して好適である。

【0116】

ここで図7を参照すると、本発明の一実施形態によるシステム700のブロック図が示されている。システム700は、1または複数のプロセッサ710、715を含んでよく、当該1または複数のプロセッサ710、715は、コントローラハブ720に連結される。一実施形態において、コントローラハブ720は、グラフィックメモリコントローラハブ(GMCH)790および入/出力ハブ(IOH)750(別個のチップ上に存在してよい)を含む。GMCH790は、メモリ740およびコプロセッサ745が連結されたメモリコントローラおよびグラフィックコントローラを含む。IOH750は、入出力(I/O)デバイス760をGMCH790に連結する。代替的に、メモリコントローラおよびグラフィックコントローラ的一方または両方がプロセッサ内に統合され(本明細書に記載の通り)、メモリ740およびコプロセッサ745は、プロセッサ710と、単一のチップ内のIOH750を持つコントローラハブ720とに直接連結される。

【0117】

図7中、破線を用いて、追加のプロセッサ715がオプションの性質であることが示されている。各プロセッサ710、715は、本明細書に記載の処理コアのうちの1または複数を含んでよく、プロセッサ600の何らかのバージョンであってよい。

【0118】

メモリ740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)、またはこれら2つの組み合わせであってよい。少なくとも1つの実施形態について、コントローラハブ720は、フロントサイドバス(FSB)等のマルチドロップバス、QuickPathインターコネクタ(QPI)等のポイントツーポイントインタフェースまたは類似の接続795を介して、プロセッサ710、715と通信する。

【0119】

一実施形態において、コプロセッサ745は、例えば、高スループットMICプロセッサ、ネットワークプロセッサまたは通信プロセッサプロセッサ、圧縮エンジン、グラフィックプロセッサ、GPGPU、組み込みプロセッサ等のような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720は統合グラフィックアクセラレータを含んでよい。

【0120】

物理リソース710と715との間には、アーキテクチャ上、マイクロアーキテクチャ上、熱的、電力消費特性等を含む利点の様々な基準に関して、様々な差異が存在し得る。

【0121】

一実施形態において、プロセッサ710は、汎用タイプのデータ処理演算を制御する命令を実行する。コプロセッサ命令が命令内に埋め込まれてよい。プロセッサ710は、これらのコプロセッサ命令を取り付けられたコプロセッサ745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710はこれらのコプロセッサ命令(またはコプロセッサ命令を表わす制御信号)を、コプロセッサ745への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745はコプロセッサ命令を受け取り、受信されたコプロセッサ命令を実行する。

【0122】

ここで図 8 を参照すると、本発明の一実施形態による第 1 のより具体的な例示的システム 800 のブロック図を示す。図 8 に図示の通り、マルチプロセッサシステム 800 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 850 を介して連結された第 1 のプロセッサ 870 および第 2 のプロセッサ 880 を含む。プロセッサ 870 および 880 の各々は、プロセッサ 600 の何らかのバージョンであってよい。本発明の一実施形態において、プロセッサ 870 および 880 は、それぞれプロセッサ 710 および 715 である一方で、コプロセッサ 838 はコプロセッサ 745 である。別の実施形態においては、プロセッサ 870 および 880 は、それぞれプロセッサ 710 およびコプロセッサ 745 である。

【0123】

プロセッサ 870 および 880 は、それぞれ統合メモリコントローラ (IMC) ユニット 872 および 882 を含むように図示されている。プロセッサ 870 はまた、そ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) インタフェース 876 および 878 を含み、同様に第 2 のプロセッサ 880 は P-P インタフェース 886 および 888 を含む。プロセッサ 870、880 は、P-P インタフェース回路 878、888 を使用して、ポイントツーポイント (P-P) インタフェース 850 を介して情報を交換してよい。図 8 に図示の通り、IMC 872 および 882 はプロセッサをそれぞれのメモリ、すなわちメモリ 832 およびメモリ 834 に連結する。メモリ 832 およびメモリ 834 は、それぞれのプロセッサにローカルに取り付けられたメインメモリの一部であってよい。

【0124】

プロセッサ 870、880 はそれぞれ、ポイントツーポイントインタフェース回路 876、894、886、898 を使用して、個々の P-P インタフェース 852、854 を介して、チップセット 890 と情報を交換してよい。随意で、チップセット 890 は、高性能インタフェース 839 を介してコプロセッサ 838 と情報を交換してよい。一実施形態において、コプロセッサ 838 は、例えば、高スループット MIC プロセッサ、ネットワークプロセッサまたは通信プロセッサプロセッサ、圧縮エンジン、グラフィックプロセッサ、GPGPU、組み込みプロセッサ等のような専用プロセッサである。

【0125】

共有キャッシュ (不図示) が、いずれかのプロセッサの内部または両方のプロセッサの外部に含まれてよく、共有キャッシュはさらに当該プロセッサと P-P 相互接続を介して接続されていてよく、その結果、プロセッサが低電力モードの場合、いずれか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内に格納され得るようになる。

【0126】

チップセット 890 が、インタフェース 896 を介して第 1 のバス 816 に連結されてよい。一実施形態において、第 1 のバス 816 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、すなわち PCI Express バス若しくは別の第 3 世代 I/O 相互接続バス等のバスであってよいが、本発明の範囲はそのようには限定されない。

【0127】

図 8 に図示の通り、様々な I/O デバイス 814 がバスブリッジ 818 と共に第 1 のバス 816 に連結されてよく、バスブリッジ 818 は第 1 のバス 816 を第 2 のバス 820 に連結する。一実施形態において、コプロセッサ、高スループット MIC プロセッサ、GPGPU のアクセラレータ (例えば、グラフィックアクセラレータまたはデジタル信号処理 (DSP) ユニット等)、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 815 が第 1 のバス 816 に連結される。一実施形態において、第 2 のバス 820 はローピンカウント (LPC) バスであってよい。一実施形態において、様々なデバイスが第 2 のバス 820 に連結されてよく、そのようなものとしては、例えば、キーボードおよび/またはマウス 822、通信デバイス 827 および命令/コードおよびデータ 830 を含み得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 828 が含まれる。さらに、オーディオ I/O

10

20

30

40

50

0824が第2のバス820に連結されてよい。他のアーキテクチャも可能であることに留意されたい。例えば、図8のポイントツーポイントアーキテクチャの代わりに、システムはマルチドロップバスまたは他のこのようなアーキテクチャを実装してよい。

【0128】

ここで図9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900のブロック図が示されている。図8および図9中で同様の要素は同様の参照符号が付されており、図9の他の態様を不明瞭にするのを回避すべく、図8の特定の態様は図9で省略されている。

【0129】

図9は、プロセッサ870、880が統合メモリおよびI/O制御ロジック(「CL」)872および882をそれぞれ含んでよいことを示す。故に、CL872、882は、統合メモリコントローラユニットを含み、I/O制御ロジックを含む。図9は、メモリ832、834がCL872、882に連結されるだけでなく、I/Oデバイス914も制御ロジック872、882に連結されることも示している。レガシI/Oデバイス915がチップセット890に連結される。

【0130】

ここで図10を参照すると、本発明の一実施形態によるSoC1000のブロック図が示されている。図6中と同様の要素は同一の参照番号が付されている。また、破線ボックスは、より高度なSoC上でのオプションの機能である。図10中、相互接続ユニット1002は、アプリケーションプロセッサ1010と、システムエージェントユニット610と、バスコントローラユニット616と、統合メモリコントローラユニット614と、コプロセッサ1020のセットまたは1若しくは複数のコプロセッサ1020と、スタティックランダムアクセスメモリ(SRAM)ユニット1030と、ダイレクトメモリアクセス(DMA)ユニット1032と、1または複数の外部ディスプレイに連結するためのディスプレイユニット1040とに連結される。アプリケーションプロセッサ1010は、1または複数のコア202A~Nのセットおよび共有キャッシュユニット606を含む。コプロセッサ1020のセットまたは1若しくは複数のコプロセッサ1020は、統合グラフィックロジック、イメージプロセッサ、オーディオプロセッサおよびビデオプロセッサを含んでよい。一実施形態において、コプロセッサ1020は、例えば、ネットワークプロセッサまたは通信プロセッサ、圧縮エンジン、GPGPU、高スループットMICプロセッサ、組み込みプロセッサ等のような専用プロセッサを含む。

【0131】

本明細書に開示のメカニズムに係る実施形態は、ハードウェア、ソフトウェア、ファームウェアまたはこのような実装アプローチの組み合わせで実装されてよい。本発明の実施形態は、少なくとも1つのプロセッサ、ストレージシステム(揮発性および不揮発性のメモリ並びに/またはストレージ要素を含む)、少なくとも1つの入力デバイスおよび少なくとも1つの出力デバイスを備えるプログラム可能なシステム上で実行され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てよい。

【0132】

図8に図示されたコード830等のプログラムコードは、本明細書に記載の機能を実行するための命令を入力するため、および出力情報を生成するために適用されてよい。出力情報は、1または複数の出力デバイスに既知の態様で適用されてよい。本願の目的において、処理システムには、例えば、デジタル信号プロセッサ(DSP)、マイクロコントローラ、特定用途向け集積回路(ASIC)、またはマイクロプロセッサ等のプロセッサを有する任意のシステムが含まれる。

【0133】

プログラムコードは、処理システムと通信するために、高水準の手順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てよい。必要であれば、プログラムコードはまた、アセンブリ言語または機械言語で実装されてもよい。実際、本明細書に記載のメカニズムは、いずれの特定のプログラミング言語にも範囲限定されない。

10

20

30

40

50

いずれの場合においても、言語はコンパイル型言語または解釈型言語であってよい。

【0134】

少なくとも1つの実施形態に係る1または複数の態様は、機械可読媒体上に格納された、プロセッサ内で様々なロジックを表わす典型的命令によって実装されてよく、当該命令は機械による読み取り時に、機械に対し、本明細書に記載の技術を実行するためのロジックを生成させる。このような「IPコア」として知られる典型的なものが、有形の機械可読媒体上に格納され、様々な顧客または製造施設に供給され、実際にロジックまたはプロセッサを作成する製造機械にロードされてよい。

【0135】

このような機械可読記録媒体としては、限定はされないが、機械またはデバイスによって製造または形成される複数の物品から成る非一時的な有形の構成が含まれてよく、それらとしては、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）、およ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）等の半導体デバイス、磁気カード若しくは光カードといった記録媒体または電子的命令を格納するのに好適な任意の他のタイプの媒体が含まれる。

【0136】

従って、また、本発明の実施形態は、命令を含む、または本明細書に記載の構造、回路、装置、プロセッサおよび/またはシステム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な有形の機械可読媒体を含む。また、このような実施形態はプログラム製品としても称されてよい。

【0137】

いくつかの場合において、命令コンバータが使用され、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してよい。例えば、命令コンバータは、ある命令を、コアによって処理されるべき1または複数の他の命令へと、トランスレート（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使用して）、モーフィング、エミュレート、またはそれら以外の方法による変換を行ってよい。命令コンバータは、ソフトウェア、ハードウェア、ファームウェア、またはこれらの組み合わせで実装されてよい。命令コンバータは、プロセッサ内、プロセッサ外、または部分的にプロセッサ内または部分的にプロセッサ外に存在してよい。

【0138】

図11は、本発明の実施形態による、ソース命令セット内のバイナリ命令をターゲット命令セット内のバイナリ命令に変換するためのソフトウェア命令コンバータの使用を対比するブロック図である。図示された実施形態において、命令コンバータはソフトウェア命令コンバータであるものの、代替的に、命令コンバータはソフトウェア、ファームウェア、ハードウェアまたはこれらの様々な組み合わせで実装されてもよい。図11は、高水準言語1102のプログラムが、x86バイナリコード1106を生成するx86コンパイラ1104を使用してコンパイルされ得ることを示しており、当該x86バイナリコード1106は、少なくとも1つのx86命令セットコアを持つプロセッサ1116によってネイティブに実行されてよい。少なくとも1つのx86命令セットコアを持つプロセッサ1116は、少なくとも1つのx86命令セットコアを持つインテルプロセッサと実質的に同一の諸機能を実行できる任意のプロセッサを表わしており、これは次のように行う。すなわち、少なくとも1つのx86命令セットコアを持つインテルプロセッサと実質的に同一の結果を得るべく、(1)インテルx86命令セットコアの命令セットの大部分、または(2)少なくとも1つのx86命令セットコアを持つインテルプロセッサ上での実行を目的とするアプリケーションまたは他のソフトウェアのオブジェクトコードバージョン

10

20

30

40

50

、を互換性のある状態で実行またはそれ以外の方法で処理することによってである。x 86 コンパイラ 1104 は、x 86 バイナリコード 1106 (例えばオブジェクトコード) を生成するように動作可能なコンパイラを表わし、当該 x 86 バイナリコード 1106 は、追加のリンク処理と共に、または追加のリンク処理なしに、少なくとも1つの x 86 命令セットコアを持つプロセッサ 1116 上で実行可能である。同様に、図 11 は、高水準言語 1102 のプログラムが、代替的な命令セットバイナリコード 1110 を生成する代替的な命令セットコンパイラ 1108 を使用してコンパイルされ得ることを示しており、当該代替的な命令セットバイナリコード 1110 は、少なくとも1つの x 86 命令セットコアを持たないプロセッサ 1114 (例えば、カリフォルニア州サニーベールの MIPS Technologies の MIPS 命令セットを実行する、および/または、カリフォルニア州サニーベールの ARM Holdings の ARM 命令セットを実行するコアを持つプロセッサ) によってネイティブに実行されてよい。命令コンバータ 1112 は、x 86 バイナリコード 1106 を、x 86 命令セットコアを持たないプロセッサ 1114 によってネイティブに実行可能なコードに変換されるのに使用される。これが可能な命令コンバータの作成は難しいので、この変換されたコードは、代替的な命令セットバイナリコード 1110 と同じである可能性は低いが、しかしながら、変換されたコードは、一般的な演算を達成し、代替的な命令セットに属する命令で構成されるであろう。故に、命令コンバータ 1112 は、ソフトウェア、ファームウェア、ハードウェアまたはこれらの組み合わせを表わし、それらは、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介して、x 86 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、x 86 バイナリコード 1106 を実行できるようにする。

[ベクトルビットシャッフルを実行するための方法および装置]

【0139】

ベクトルビットシャッフル命令について以下説明する。当該命令は、第1のソースオペランドを制御として、第2のソースオペランドをデータとして使用し、ビットシャッフルを実行する。この命令は、複数のビット操作ルーチンを実装するために効率的に使用されてよい。例示であり限定ではないが、当該命令を使用して、可変ビット置換を実装し、現行の VEX 実装または EVEX 実装に対し、最大 8 倍の速度アップをもたらしてよい。

【0140】

図 12 に図示の通り、本発明の実施形態が実装されてよい例示的なプロセッサ 1255 は、汎用レジスタ (GPR) セット 1205、ベクトルレジスタセット 1206、およびマスクレジスタセット 1207 を含む。一実施形態において、複数のベクトルデータ要素が、2個の 256 ビット値、4個の 128 ビット値、8個の 64 ビット値、16個の 32 ビット値等を格納するための 512 ビット幅を有してよい各ベクトルレジスタ 1206 内にパックされる。しかしながら、本発明の根本的な原理は、いずれの特定のサイズ/タイプのベクトルデータにも限定されない。一実施形態において、マスクレジスタ 1207 は、ベクトルレジスタ 1206 内に格納された値にビットマスク演算を実行するために使用される 8個の 64 ビットオペランドマスクレジスタ (例えば、上記のマスクレジスタ k0 ~ k7 として実装される) を含む。しかしながら、本発明の根本的な原理は、いずれの特定のマスクレジスタのサイズ/タイプにも限定されない。

【0141】

簡単にするために、単一のプロセッサコア (「コア 0」) の詳細が図 12 中に示されている。しかしながら、図 12 に図示の各コアは、コア 0 と同一のロジックセットを有してよいことを理解されたい。例えば、各コアは、指定されたキャッシュ管理ポリシーに従い、命令およびデータをキャッシュするための専用のレベル 1 (L1) キャッシュ 1212 およびレベル 2 (L2) キャッシュ 1211 を含んでよい。L1 キャッシュ 1212 は、命令を格納するための別個の命令キャッシュ 1220 およびデータを格納するための別個のデータキャッシュ 1221 を含む。様々なプロセッサキャッシュ内に格納された命令およびデータは、固定サイズ (例えば、64、128、512 バイト長) であってよいキャッシュラインの粒度で管理される。この例示的な実施形態の各コアは、メインメモリ 12

10

20

30

40

50

00 および/または共有レベル3 (L3) キャッシュ1216から命令をフェッチするための命令フェッチユニット1210、命令をデコーディング(例えば、プログラム命令をマイクロオペレーションまたは「 μop 」へとデコーディング)するためのデコードユニット1220、命令を実行するための実行ユニット1240、および命令をリタイアし、結果をライトバックするためのライトバックユニット1250を有する。

【0142】

命令フェッチユニット1210は、メモリ1200(または複数のキャッシュのうちの1つ)からフェッチされる次の命令のアドレスを格納するための次の命令ポインタ1203、アドレス変換速度を改善すべく最近使用された仮想命令アドレスと物理命令アドレスのマッピングを格納するための命令トランシェーションルックアサイドバッファ(ITLB)1204、命令分岐アドレスを投機的に予測するための分岐予測ユニット1202、および分岐アドレスおよびターゲットアドレスを格納するための分岐ターゲットバッファ(BTB)1201を含む、様々な周知のコンポーネントを含む。いったんフェッチされた命令は、その後デコードユニット1230、実行ユニット1240およびライトバックユニット1250を含む命令パイプラインの残りのステージにストリームされる。これらのユニットの各々の構造および機能は当業者に十分理解されており、本発明の異なる実施形態の関連態様を不明瞭にするのを避けるべく、ここでは詳細に記載しない。

【0143】

一実施形態において、デコードユニット1230は、本明細書に記載のベクトルビットシャッフル命令を(例えば、一実施形態において、一連のマイクロオペレーションへと)デコーディングするためのベクトルビットシャッフルデコードロジック1231を含み、実行ユニット1240は、当該命令を実行するためのベクトルビットシャッフル実行ロジック1241を含む。

【0144】

上記の通り、一実施形態において、ベクトルビットシャッフル命令は、第1のソースを制御として、第2のソースをデータとして使用し、結果をデスティネーションレジスタ内に出力する、ビットギャザー選択を実行する。一実施形態において、デスティネーションの各ビットは、第1のソースからの6つの制御ビットを使用して、第2のソースから識別される。

【0145】

図13は例示的な実施形態を示し、そこには、制御ビットを格納するための第1のソースレジスタであるSRC2、ソースデータを格納するための第2のソースレジスタであるSRC3、およびベクトルビットシャッフル命令の結果を格納するためのデスティネーションレジスタであるDSTが含まれる。一実施形態において、SRC3は、512ビットベクトルレジスタ内にパックされた64ビットデータの8個のレーン0~7を含み、SRC2は、これもまた512ビットベクトルレジスタ内にパックされた8個の制御バイトから成る8個のセット1300~1302...1307を含み、DSTは、64ビットのマスキングレジスタ1320を含む。しかしながら、上記の通り、本発明の根本的な原理は、いずれの特定のサイズ/タイプのオペランドまたはレジスタにも限定されない。簡略にするため、図13には、SRC3内に格納されたデータおよびSRC2内に格納された制御ビットの一部のみが図示されていることに留意されたい。

【0146】

演算では、8個の制御バイトから成る各セットの各バイトが、その対応する64ビットレーン内の特定のビットを識別する。従って、8個の制御バイト1300の各々は、レーン0内のビットを識別し、8個の制御バイト1301の各々は、レーン1内のビットを識別するといった具合である。一実施形態において、各制御バイト1300~1302...1307において、8ビットのうち6ビットのみが使用され、SRC3内のビットを識別する($2^6 = 64$ なので、6ビットで十分である)。残りの2ビットは無視されてよい。

【0147】

10

20

30

40

50

一実施形態において、8個の制御バイト1300～1307の各々によるこれらの6ビットが、選択ロジック1310～1312...1317（例えば、マルチプレクサのセット）に適用され、各レーンから8ビットが選択される。従って、選択ロジック1310はレーン0から8ビットを選択し、選択ロジック1311はレーン1から8ビットを選択し、選択ロジック1312はレーン2から8ビットを選択し、選択ロジック1317はレーン7から8ビットを選択する（上記の通り、簡略化のため、レーン3～6および関連の選択ロジックは図示されていない）。

【0148】

最終結果は、8ビットの8セットがSRC3から読み出されるということである。一実施形態において、当該8ビットの8セットがDEST内で共に連結され、64ビットマスク値1320を形成する。いったん形成された64ビットマスク値1320は、後続のマスク演算に使用されてよい。

10

【0149】

図14に、本発明の一実施形態による方法が図示されている。方法は、上記のアーキテクチャの文脈において実行されてよいが、当該方法はいずれの特定のシステムアーキテクチャにも限定されない。

【0150】

1401において、ベクトルビットシャッフル命令がシステムメモリからフェッチされ、またはキャッシュ（例えば、L1、L2またはL3キャッシュ）から読み出される。1402において、ベクトルビットシャッフル命令のデコーディング/実行にตอบสนองして、シャッフルされる入力ベクトルデータが第1のソースレジスタ内に格納される。上記の通り、一実施形態において、第1のソースレジスタは512ビットベクトルレジスタであり、ベクトルデータは8個の64ビットデータレーンを含む。1403において、ベクトルビットシャッフルを実行するために必要な制御データが第2のソースレジスタ内に格納され、第2のソースレジスタは上記の通り、別の512ビットベクトルレジスタであってよい。

20

【0151】

1404において、第1のソースレジスタ内の各レーンからビットのセットが、第2のソースレジスタ内の関連付けられた制御ビットのセットを使用して、識別される。上記の通り、一実施形態において、各レーンに8個の制御バイトが提供され、各制御バイトのうち6ビットが使用され、対応するレーンのビットを識別する。最終結果は、第1のソースレジスタから8ビットの8セットが読み出されるということである。最後に、1405において、デスティネーションマスクレジスタ内で当該ビットのセットが連結される。上記の実施形態においては、例えば、8ビットの8セットが連結され、64ビットのマスク値を形成する。

30

【0152】

一実施形態において、EVEXエンコード実装においては、第1のソースオペランド、第2のソースのオペランドおよびデスティネーションオペランドはすべてZMMレジスタである。一実施形態において、ベクトルビットシャッフル命令は以下の形態を取る。ここでDESTはデスティネーション、SRC2は制御データを含むソースを含み、SRC3はシャッフルされるデータを含むソースを含む。

40

【数1】

VPSHUFBITQMB DEST, SRC2, SRC3

【0153】

以下の擬似コードは、本発明の一実施形態により実行される演算の典型例を提供する。

【数 2】

```

VSHUFBITQMB DEST, SRC2, SRC3
(KL, VL) = (16, 128), (32, 256), (64, 512)
FOR i:= 0 to KL/8 - 1; Qword
    FOR j:= 0 to 7; Byte
        IF EVEX.b AND SRC3 *is memory*
            THEN
                Data:= SRC3.qword[i];
            ELSE
                Data:= SRC3.qword[i];
                m:= SRC2.qword[i].byte[j] & 0x3F
                k1[i*8+j]:=Data.bit[m]
            END FOR
        END FOR
    END FOR
    k1[MAX_KL-1:KL]:=0

```

【0154】

従って、KL = 64 且つ VL = 512 と想定すると、外側 FOR ループ (i に基づく) は異なる 64 ビットレーン (Qワード) の各々を選択するために使用され、内側 FOR ループ (j に基づく) は制御バイトで指定された制御値を使用して各レーン内の 8 ビットを選択するために使用される。「EVEX.b AND SRC3 *is memory*」を持つ IF ステートメントは、「b」ビットが EVEX ビットフィールドに設定される場合 (通常、ソースブロードキャスト、ラウンド制御 (L'L との組み合わせ) または抑制例外に使用) 且つソースデータがシステムメモリから読み出されている場合、単一の 64 ビットソース値がすべてのレーン ((KL, VL) = (64, 512)) にコピーされることを示す。さもなければ、使用される Qワード / レーンは、i の現在の値 (Data:= SRC3.qword[i]) に基づいて選択される。また、各 64 ビットレーン内のビット値を識別するために、各制御バイトのうちの 6 ビットのみが使用されるので、0x3F の値はインデックス値 SRC2.qword[i].byte[j] と AND 演算される (すなわち、0x3F との AND 演算で上位 2 ビットを除去する) 。

【0155】

上記の明細書において、本発明の実施形態は、本発明の具体的な例示的实施形態を参照して記載されている。しかしながら、添付の特許請求の範囲に記載の本発明のより広範な精神および範囲から逸脱することなく、様々な修正および変更がそこに加えられ得ることは自明であろう。従って、明細書および図面は限定的な意味ではなく、例示的な意味において解釈されるべきである。

【0156】

本発明の実施形態は、上記の様々な段階を含んでよい。当該段階は機械で実行可能な命令に具現化されてよく、当該命令を使用して、汎用プロセッサまたは専用プロセッサに当該段階を実行させてよい。代替的に、これらの段階は具体的なハードウェアコンポーネン

トによって実行されてよく、当該ハードウェアコンポーネントは、当該段階を実行するためのハードワイヤードされたロジックを含む。またはこれらの段階はプログラムされたコンピュータコンポーネントおよびカスタムのハードウェアコンポーネントの任意の組み合わせによって実行されてよい。

【 0 1 5 7 】

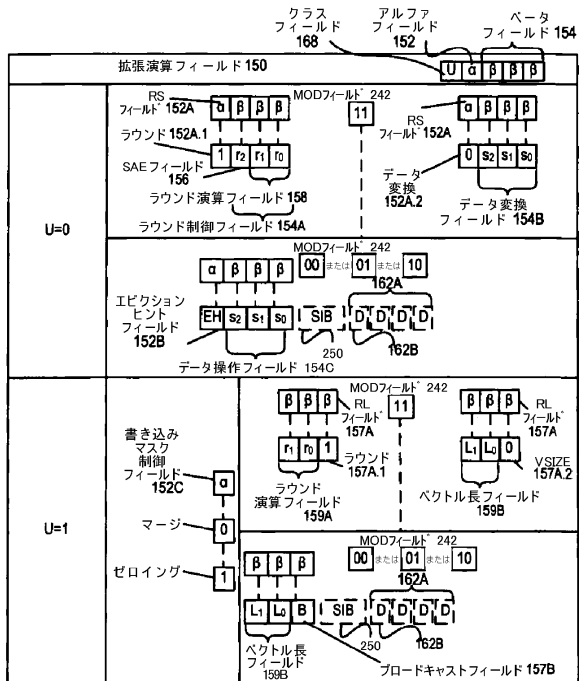
本明細書で上記の通り、命令とは、特定の処理を実行するように構成された若しくは予め定められた機能を有する特定用途向け集積回路（ASIC）等のハードウェアの特定の構成、または、非一時的コンピュータ可読媒体に具現化されたメモリ内に格納されたソフトウェア命令を指してよい。故に、図面中に図示された技術は、1または複数の電子デバイス（例えば、エンドステーション、ネットワーク要素等）に格納され、当該電子デバイス上で実行されるコードおよびデータを使用して実装可能である。このような電子デバイスは、コンピュータ機械可読媒体を使用してコードおよびデータを格納および通信し（内部的におよび/またはネットワーク経由で他の電子デバイスと共に）、このようなコンピュータ機械可読媒体としては、非一時的コンピュータ機械可読記録媒体（例えば、磁気ディスク、光ディスク、ランダムアクセスメモリ、リードオンリメモリ、フラッシュメモリデバイス、相変化メモリ）および一時的コンピュータ機械可読通信媒体（例えば、搬送波、赤外線信号、デジタル信号等、電気、光、音響または他の形態の伝搬信号）が挙げられる。また、このような電子デバイスは通常、1または複数のストレージデバイス（非一時的機械可読記録媒体）、ユーザ入力/出力デバイス（例えば、キーボード、タッチスクリーンおよび/またはディスプレイ）およびネットワーク接続等の1または複数の他のコンポーネントに連結された1または複数のプロセッサのセットを含む。プロセッサのセットと他のコンポーネントとの連結は通常、1または複数のバスおよびブリッジ（またバスコントローラとも呼ばれる）を介してなされる。ストレージデバイスおよびネットワークラフィックを搬送する信号はそれぞれ、1または複数の機械可読記録媒体および機械可読通信媒体を表わす。故に、特定の電子デバイスのストレージデバイスは通常、その電子デバイスの1または複数のプロセッサのセット上で実行されるためのコードおよび/またはデータを格納する。もちろん、本発明の実施形態に係る1または複数の部分が、ソフトウェア、ファームウェア、および/またはハードウェアの異なる組み合わせを使用して実装されてもよい。詳細な説明にわたり、本発明の完全な理解を共すべく、多数の具体的な詳細が説明目的で記載された。しかしながら、本発明はこれらの具体的な詳細の一部を省いても実施可能であることは当業者に自明なところである。特定の例においては、本発明の主題を不明瞭にするのを避けるべく、周知の構造および機能は詳細には記載されていない。従って、本発明の範囲および精神は以降の特許請求の範囲に照らし判断するものとする。

10

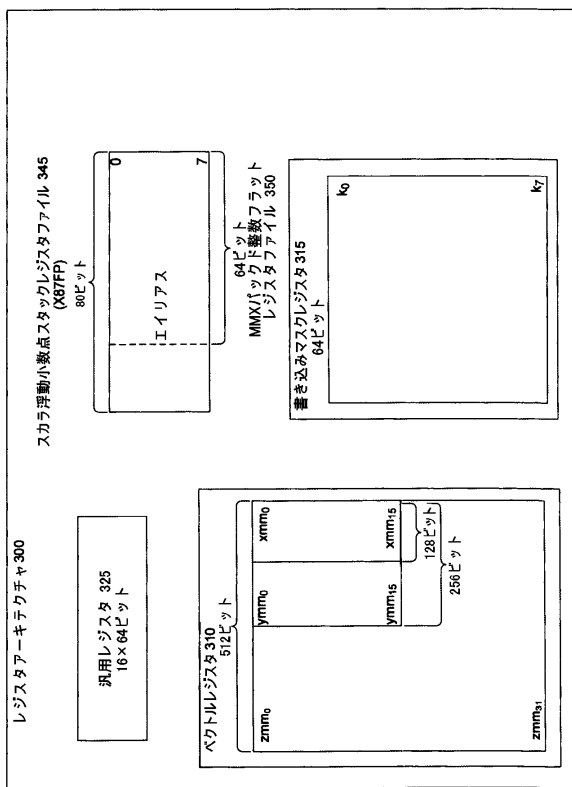
20

30

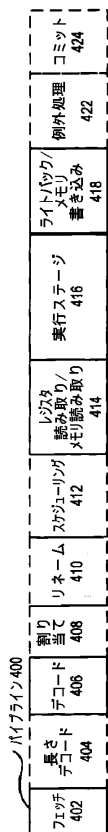
【図2D】



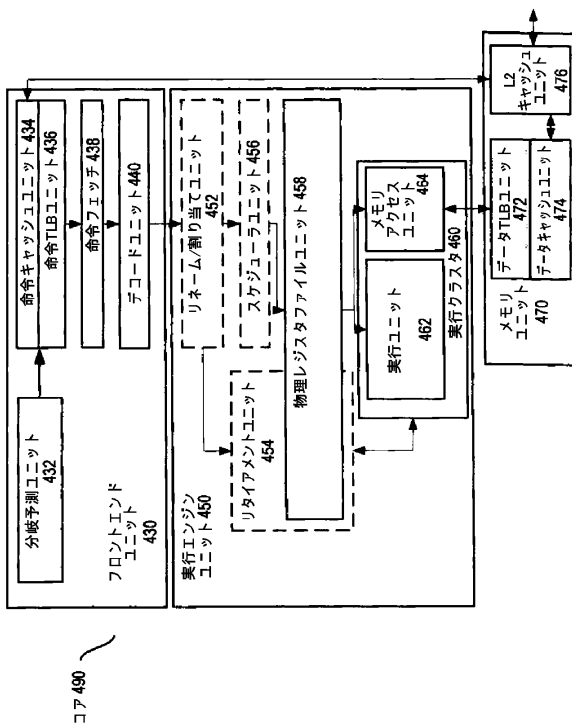
【図3】



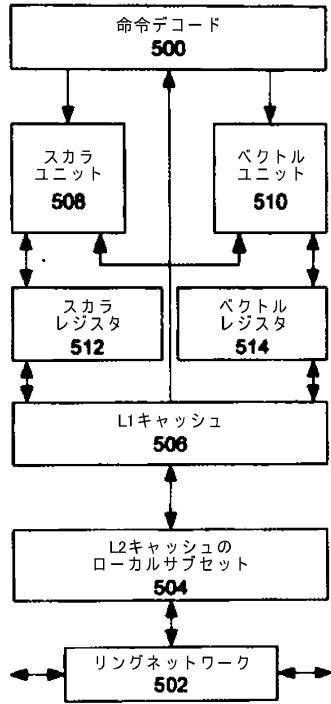
【図4A】



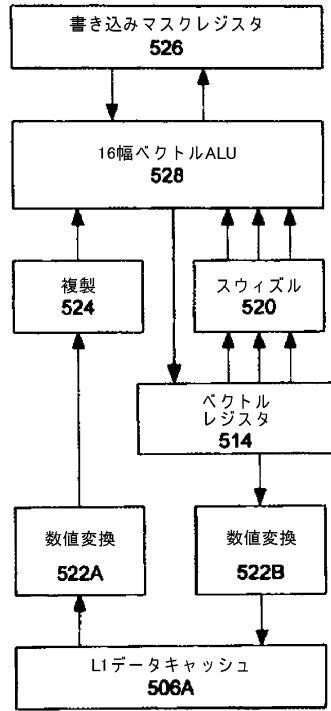
【図4B】



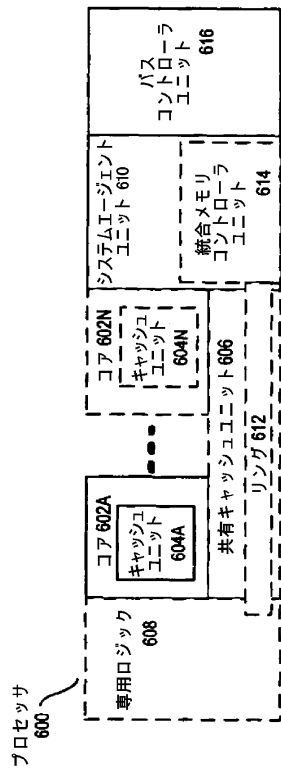
【図5A】



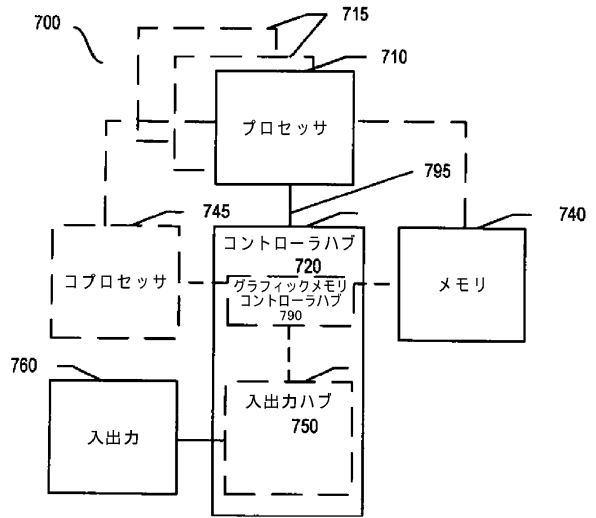
【図5B】



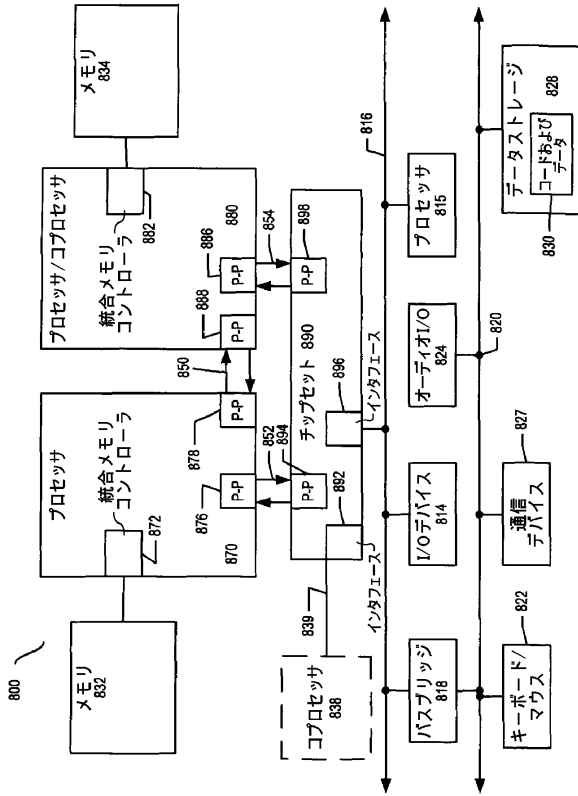
【図6】



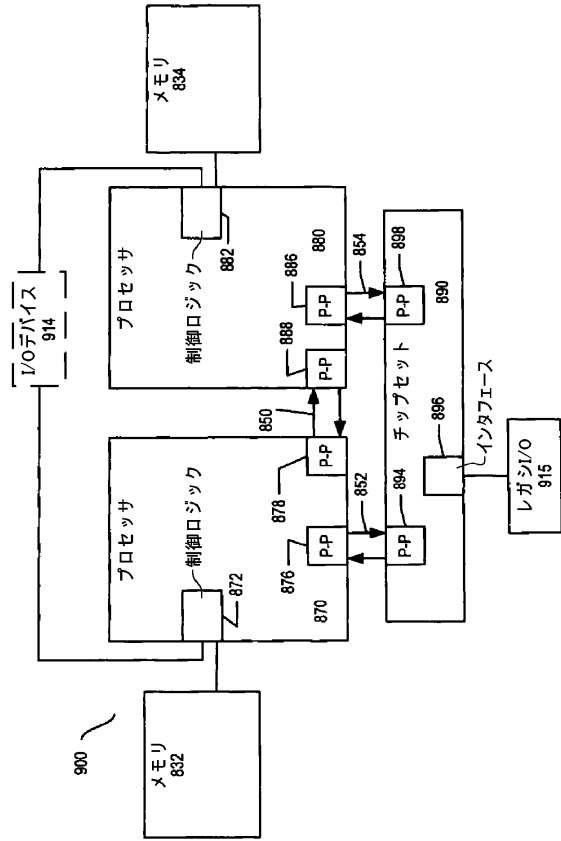
【図7】



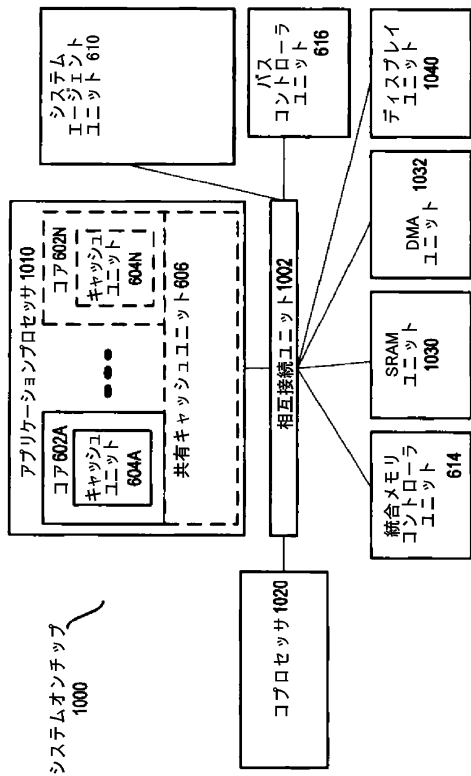
【図 8】



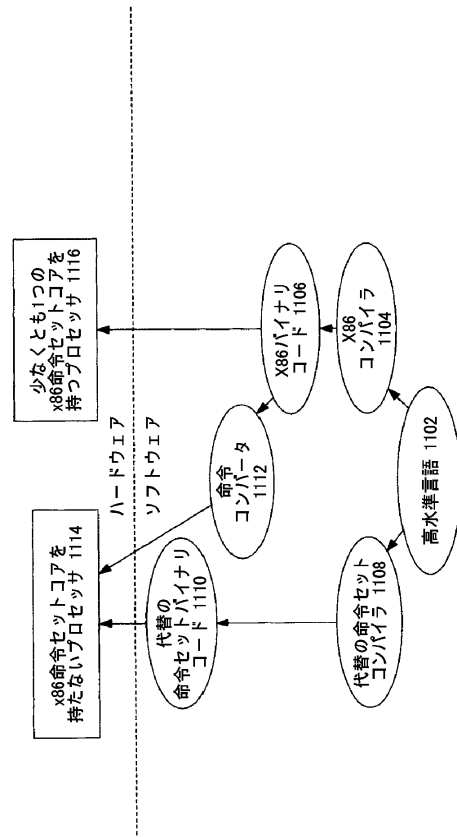
【図 9】



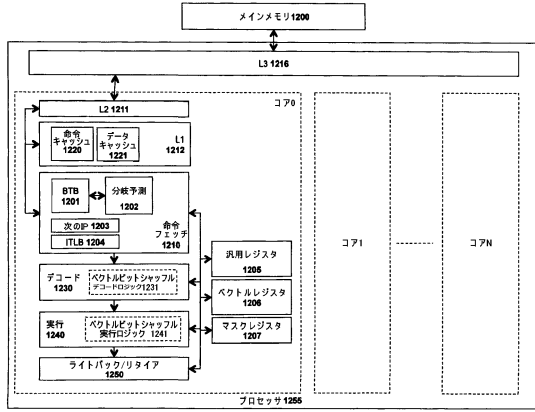
【図 10】



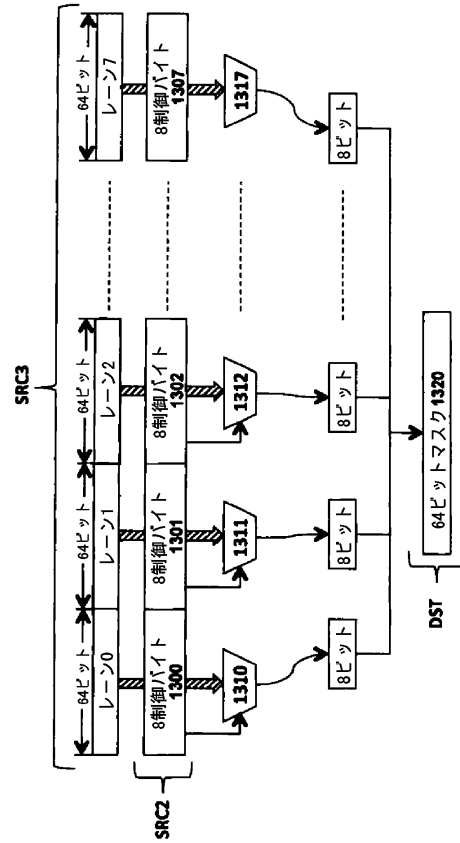
【図 11】



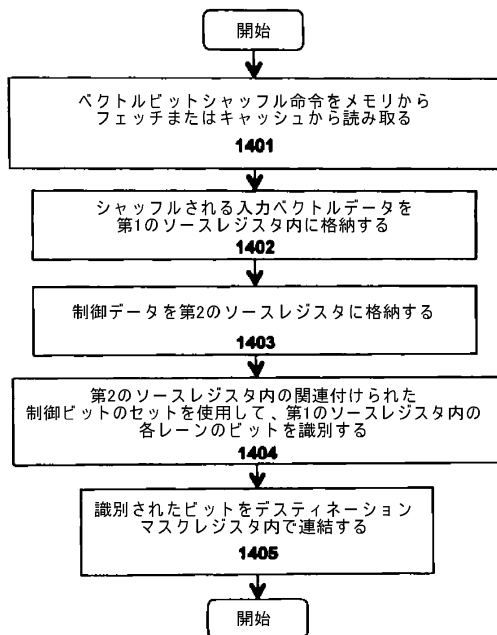
【 図 1 2 】



【 図 1 3 】



【 図 1 4 】



フロントページの続き

- (72)発明者 コーバル、ジーザス
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
バード・2200 インテル・コーポレーション内
- (72)発明者 バレンタイン、ロバート
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
バード・2200 インテル・コーポレーション内
- (72)発明者 チャーニー、マーク ジェイ.
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
バード・2200 インテル・コーポレーション内
- (72)発明者 ソレ、グイレム
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
バード・2200 インテル・コーポレーション内
- (72)発明者 エスパサ、ロジャー
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 特許第6526175(JP, B2)
特表2018-500666(JP, A)
特開2012-33032(JP, A)
特表2007-526536(JP, A)
米国特許出願公開第2014/0059323(US, A1)
米国特許出願公開第2014/0013088(US, A1)
米国特許出願公開第2013/0318328(US, A1)
米国特許出願公開第2013/0212360(US, A1)
米国特許出願公開第2005/0139647(US, A1)
米国特許出願公開第2002/0035678(US, A1)
米国特許第6963341(US, B1)
国際公開第2014/150913(WO, A2)

(58)調査した分野(Int.Cl., DB名)

G06F 9/30 - 9/355
G06F 9/38
G06F17/10 - 17/18