

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6227151号
(P6227151)

(45) 発行日 平成29年11月8日(2017.11.8)

(24) 登録日 平成29年10月20日(2017.10.20)

(51) Int.Cl.

F I

G O 6 F 12/08 (2016.01)
G O 6 F 9/46 (2006.01)G O 6 F 12/08 5 1 3
G O 6 F 12/08 5 6 5
G O 6 F 12/08 5 1 5 Z
G O 6 F 12/08 5 3 1 B
G O 6 F 9/46 4 1 0

請求項の数 26 (全 42 頁)

(21) 出願番号 特願2016-545961 (P2016-545961)
 (86) (22) 出願日 平成26年10月3日(2014.10.3)
 (65) 公表番号 特表2016-532233 (P2016-532233A)
 (43) 公表日 平成28年10月13日(2016.10.13)
 (86) 国際出願番号 PCT/US2014/059130
 (87) 国際公開番号 WO2015/048826
 (87) 国際公開日 平成27年4月2日(2015.4.2)
 審査請求日 平成28年3月24日(2016.3.24)

(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 リウ、エン・チェン
 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 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】 アドレスへの書き込みに対する監視命令を実行するスケーラブル機構

(57) 【特許請求の範囲】

【請求項1】

プロセッサであって、

分散キャッシュの第1のキャッシュ部に対応し、前記プロセッサの複数の論理プロセッサの総数より少ない総数であって各々が監視されるべきアドレスを格納する複数のキャッシュ側アドレス監視格納位置を有するキャッシュ側アドレス監視ユニットと、

第1のコアに対応し、前記第1のコアの1つまたは複数の論理プロセッサの数と同数であって各々が監視されるべきアドレス及び前記第1のコアの異なる対応する論理プロセッサの監視状態を格納する複数のコア側アドレス監視格納位置を有するコア側アドレス監視ユニットと、

前記第1のキャッシュ部に対応し、監視されるべき追加のアドレスを格納するのに利用可能な未使用キャッシュ側アドレス監視格納位置がない場合にアドレス監視格納オーバーフローポリシを強制するキャッシュ側アドレス監視格納オーバーフローユニットと

を備える

プロセッサ。

【請求項2】

前記第1のコアに対応し、前記コア側アドレス監視ユニットに連結され、対応するコア側アドレス監視格納位置が、トリガの準備ができてトリガイイベントが検出されたという監視状態を有する場合に、前記第1のコアの論理プロセッサをトリガするコア側トリガユニット

をさらに備える

請求項 1 に記載のプロセッサ。

【請求項 3】

前記キャッシュ側アドレス監視ユニットに連結され、複数の異なる論理プロセッサからの同一の監視アドレスに対する複数の監視要求を、共通キャッシュ側アドレス監視格納位置に記録するキャッシュ側アドレス監視格納位置再利用ユニット

をさらに備える

請求項 1 または 2 に記載のプロセッサ。

【請求項 4】

前記共通キャッシュ側アドレス監視格納位置は、前記同一の監視アドレスに対する前記複数の監視要求を提供した前記複数の異なる論理プロセッサを記録する構造を有する、

請求項 3 に記載のプロセッサ。

【請求項 5】

前記プロセッサは、40 より多くのハードウェアスレッドを有し、

前記第 1 のキャッシュ部に対応する前記キャッシュ側アドレス監視ユニットの前記複数のキャッシュ側アドレス監視格納位置の前記総数は、少なくとも 20 はあるが、前記 40 より多くのハードウェアスレッドの総数より少ない、

請求項 1 から 4 の何れか一項に記載のプロセッサ。

【請求項 6】

前記キャッシュ側アドレス監視ユニットの前記複数のキャッシュ側アドレス監視格納位置の前記総数は、前記複数のキャッシュ側アドレス監視格納位置のオーバーフローの確率が十万分の一以下となる前記プロセッサの前記複数の論理プロセッサの前記総数に対して十分である、

請求項 1 から 5 の何れか一項に記載のプロセッサ。

【請求項 7】

監視されるべき第 1 のアドレスを示す命令に応答して、

前記キャッシュ側アドレス監視ユニットは、前記第 1 のアドレスを、キャッシュ側アドレス監視格納位置に格納し、

前記コア側アドレス監視ユニットは、前記第 1 のアドレスを、コア側アドレス監視格納位置に格納する、

請求項 1 から 6 の何れか一項に記載のプロセッサ。

【請求項 8】

前記複数の論理プロセッサは、複数のハードウェアスレッドを含む、

請求項 1 から 6 の何れか一項に記載のプロセッサ。

【請求項 9】

前記キャッシュ側アドレス監視格納オーバーフローユニットは、複数の読み出しトランザクションに共有状態を用いるように強制することを含む前記アドレス監視格納オーバーフローポリシを強制する、

請求項 1 から 6 の何れか一項に記載のプロセッサ。

【請求項 10】

前記キャッシュ側アドレス監視格納オーバーフローユニットは、保留中の監視要求を有する可能性がある全てのコアへ複数の無効化要求を送信することを含む前記アドレス監視格納オーバーフローポリシを強制する、

請求項 1 から 6 の何れか一項に記載のプロセッサ。

【請求項 11】

前記キャッシュ側アドレス監視格納オーバーフローユニットは、どのコアに保留中の監視要求を有する可能性があるかを判断するオーバーフロー構造をチェックする、

請求項 10 に記載のプロセッサ。

【請求項 12】

複数の命令を処理するシステムであって、

10

20

30

40

50

相互接続と、

前記相互接続に連結され、かつ、キャッシュ部制御ユニットの第 1 のアドレス監視ユニット、コアインターフェースユニットの第 2 のアドレス監視ユニット、及び前記キャッシュ部制御ユニットのアドレス監視格納オーバーフローユニットを有するプロセッサと、

前記相互接続に連結されるダイナミックランダムアクセスメモリと、

前記相互接続に連結される無線通信デバイスと、

前記相互接続に連結される画像キャプチャデバイスと

を備え、

前記キャッシュ部制御ユニットの前記第 1 のアドレス監視ユニットは、分散キャッシュの第 1 のキャッシュ部に対応し、かつ、前記プロセッサの複数のハードウェアスレッドの総数より少ない総数であって各々が監視されるべきアドレスを格納する複数のアドレス監視格納位置を有し、

10

前記コアインターフェースユニットの前記第 2 のアドレス監視ユニットは、第 1 のコアに対応し、かつ、前記第 1 のコアの 1 つまたは複数のハードウェアスレッドの数と同数であって各々が監視されるべきアドレス及び前記第 1 のコアの異なる対応するハードウェアスレッドに対する監視状態を格納する、前記第 2 のアドレス監視ユニットの複数のアドレス監視格納位置を有し、

前記キャッシュ部制御ユニットの前記アドレス監視格納オーバーフローユニットは、前記第 1 のアドレス監視ユニットの全てのアドレス監視格納位置が用いられて監視要求に対してアドレスを格納するのに利用可能なものがない場合に、アドレス監視格納オーバーフローポリシを実施する、

20

システム。

【請求項 1 3】

前記アドレス監視格納オーバーフローユニットは、複数の読み出しトランザクションに共有状態を用いるように強制すること及び保留中の監視要求を有する可能性がある全てのコアへ複数の無効化要求を送信することを含む前記アドレス監視格納オーバーフローポリシを実施する、

請求項 1 2 に記載のシステム。

【請求項 1 4】

前記プロセッサは、40 より多くのハードウェアスレッドを有し、

30

前記第 1 のアドレス監視ユニットの前記複数のアドレス監視格納位置の前記総数は、少なくとも 20 はあるが、前記プロセッサの前記 40 より多くのハードウェアスレッドの前記総数より少ない、

請求項 1 2 または 1 3 に記載のシステム。

【請求項 1 5】

前記プロセッサは、複数の異なるハードウェアスレッドからの同一の監視アドレスに対する複数の監視要求を共通アドレス監視格納位置に記録する、前記キャッシュ部制御ユニットのアドレス監視格納位置再利用ユニットをさらに有する、

請求項 1 2 または 1 3 に記載のシステム。

【請求項 1 6】

40

プロセッサにおける方法であって、

アドレスを示し、かつ、マルチコアプロセッサの第 1 のコアの第 1 の論理プロセッサにおいて前記アドレスへの複数の書き込みに対して監視することを示す第 1 の命令を受信する段階と、

前記第 1 の命令に応答して、

前記第 1 のコアに対応し、かつ、数が前記第 1 のコアの複数の論理プロセッサの数に等しい複数のコア側アドレス監視格納位置の第 1 のコア側アドレス監視格納位置に、前記第 1 の命令によって示された前記アドレスを格納する段階と、

分散キャッシュの第 1 のキャッシュ部に対応し、かつ、総数が前記マルチコアプロセッサの複数の論理プロセッサの総数より少ない複数のキャッシュ側アドレス監視格納位置の

50

第 1 のキャッシュ側アドレス監視格納位置に、前記第 1 の命令によって示された前記アドレスを格納する段階と、

監視状態を投機状態に変更する段階と
を備える
方法。

【請求項 17】

前記アドレスをまた示し、第 2 のコアの第 2 の論理プロセッサにおいて前記アドレスへの複数の書き込みに対して監視することを示す第 2 の命令を受信する段階と、

前記第 2 のコアのための前記アドレスに対する監視要求を前記第 1 のキャッシュ側アドレス監視格納位置に記録する段階と

をさらに備える

請求項 16 に記載の方法。

【請求項 18】

前記第 2 のコアのための前記アドレスに対する前記監視要求を前記第 1 のキャッシュ側アドレス監視格納位置に記録する段階は、前記マルチコアプロセッサの各コアに対応する異なるビットを有するコアマスクにおけるビットを変更する段階を含む、

請求項 17 に記載の方法。

【請求項 19】

第 2 のアドレスを示し、前記第 1 の論理プロセッサにおいて前記第 2 のアドレスへの複数の書き込みに対して監視することを示す第 2 の命令を受信する段階と、

前記第 1 のキャッシュ部に対応する前記複数のキャッシュ側アドレス監視格納位置の中に利用可能な複数のキャッシュ側アドレス監視格納位置がないことを判断する段階と、

キャッシュ側アドレス監視格納位置オーバーフローモードに入ることを決定する段階と
をさらに備える

請求項 16 から 18 の何れか一項に記載の方法。

【請求項 20】

前記キャッシュ側アドレス監視格納位置オーバーフローモードにある間は、

前記第 1 のキャッシュ部に対応する全ての読み出しトランザクションに共有キャッシュコヒーレンシ状態を用いることを強制する段階と、

前記第 1 のキャッシュ部に対応する複数の無効化要求を、1 つまたは複数の保留中の監視要求を有する可能性がある、前記マルチコアプロセッサの全てのコアへ送信する段階と
をさらに備える

請求項 19 に記載の方法。

【請求項 21】

前記第 1 の論理プロセッサにおいて前記アドレスを示す第 2 の命令を受信する段階と、

前記第 2 の命令に応答して、前記監視状態をトリガ待ち状態に変更する段階と

をさらに備える

請求項 16 から 20 の何れか一項に記載の方法。

【請求項 22】

請求項 16 から 21 の何れかに一項に記載の方法を実行するプロセッサ。

【請求項 23】

請求項 16 から 21 の何れか一項に記載の方法を実行するための手段を含むプロセッサ。

【請求項 24】

コンピュータにより実行されるプログラムであって、

前記コンピュータにより実行された場合、前記コンピュータに請求項 16 から 21 の何れか一項に記載の方法を実行させるように動作可能である 1 つまたは複数の命令を格納する

プログラム。

【請求項 25】

10

20

30

40

50

相互接続、
前記相互接続に連結されるプロセッサ、
前記相互接続に連結される無線通信チップ、及び
前記相互接続に連結されるビデオカメラ
を備えるコンピュータシステムであって、
前記コンピュータシステムは、請求項 16 から 21 の何れか一項に記載の方法を実行する、
コンピュータシステム。

【請求項 26】

請求項 24 に記載のプログラムを格納するコンピュータ可読記憶媒体。

10

【発明の詳細な説明】

【技術分野】

【0001】

本明細書で説明される複数の実施形態は、複数のプロセッサに関する。特に、本明細書で説明される複数の実施形態は概して、アドレスへの書き込みに対して監視する命令を実行するように動作可能な複数のプロセッサに関する。

【背景技術】

【0002】

半導体処理及びロジック設計の進歩は、複数のプロセッサ及び複数の他の集積回路デバイスに含まれるロジックの量の増加を可能にしていた。結果として、現在、多くのプロセッサは、単一集積回路またはダイ上にモノリシックに統合される複数のコアに及ぶコアを有する。複数のコアは一般に、複数のソフトウェアスレッドまたは複数の他の作業量が同時に実行されることを可能にするのに役に立ち、一般に実行スループットを増加させるのに役に立つ。

20

【0003】

このような複数のコアプロセッサにおける 1 つの課題は、より高まる需要が、多くの場合、メモリからデータ及び / または複数の命令をキャッシュするのに用いられる複数のキャッシュに置かれる。ひとつには、そのような複数のキャッシュにおいてデータをアクセスするためのより高い相互接続帯域幅に対する需要が増え続ける傾向にある。複数のキャッシュに対する相互接続帯域幅を増加させるのに役に立つ 1 つの技術は、分散キャッシュを用いることを伴う。当該分散キャッシュは、複数の物理的個別または分散キャッシュスライス、もしくは複数の他のキャッシュ部を含み得る。そのような分散キャッシュは、共有相互接続を通して、キャッシュの複数の異なる分散部分への並行アクセスを可能にし得る。

30

【0004】

そのような複数のコアプロセッサにおける別の課題は、共有メモリに対するスレッド同期を提供する能力である。複数のオペレーティングシステムは、一般的に共有メモリに対するスレッド同期を処理する複数のアイドルグループを実行する。例えば、複数のメモリ位置のセットを用いるいくつかのビジュアループであってよい。第 1 のスレッドは、ループ内で待機して対応するメモリ位置をポーリングし得る。例として、メモリ位置は、当該第 1 のスレッドの作業キューを表し得、当該第 1 のスレッドは、実行可能な作業があるかどうかを判断すべく、当該作業キューをポーリングし得る。共有メモリ構成において、ビジュアループから出るとは、多くの場合、メモリ位置と関連付けられる状態変更起因して発生する。これらの状態変更は一般的に、別のコンポーネント（例えば、別のスレッドまたはコア）によるメモリ位置への複数の書き込みによって、引き起こされる。例えば、別のスレッドまたはコアは、第 1 のスレッドによって実行される作業を提供するように、当該メモリ位置において作業キューへ書き込み得る。

40

【0005】

複数の特定のプロセッサ（例えば、カリフォルニア州サンタクララ所在のインテル（登録商標）コーポレーションより入手可能なもの）は、共有メモリに対するスレッド同期を

50

実現するMONITOR及びMWAIT命令を用いることができる。ハードウェアスレッドまたは他の論理プロセッサは、監視ユニットによって監視されるべきリニアアドレス範囲を設定し、当該監視ユニットに準備させまたは作動させるべく、MONITOR命令を用いてよい。アドレスは、汎用レジスタを通して提供され得る。アドレス範囲は一般に、ライトバックキャッシング型である。監視ユニットは、当該監視ユニットをトリガするアドレス範囲内のアドレスへの複数の格納／書き込みを監視及び検出する。

【0006】

MWAIT命令は、プログラム順序に沿ってMONITOR命令に続いてよく、ハードウェアスレッドまたは他の論理プロセッサが命令実行を停止して実装依存状態に入ることができるヒントとして、機能してよい。例えば、論理プロセッサは、低減電力消費状態に入り得る。論理プロセッサは、MONITOR命令と関連付けられるクオリファイングイベントのセットのうちの1つが検出されるまで、その状態を保ち得る。先行のMONITOR命令によって準備されるアドレス範囲のアドレスへの書き込み／格納が、そのようなクオリファイングイベントの1つである。そのような複数の場合において、論理プロセッサは、当該状態から出てよく、プログラム順序に沿ってMWAIT命令に続く命令の実行を再開し得る。

【図面の簡単な説明】

【0007】

本発明は、複数の実施形態を例示するのに用いられる以下の説明及び複数の添付の図面を参照することによって、最もよく理解されることができる。複数の図面は以下の通りである。

【図1】プロセッサの実施形態のブロック図である。

【図2】キャッシュエージェントの実施形態のブロック図である。

【図3】監視有限状態機械の実施形態の複数の状態を示す図である。

【図4】複数の監視要求が同一アドレスを示す場合、複数のハードウェアスレッド及び／またはコアに対する単一キャッシュ側アドレス監視格納位置を再利用するように動作可能なオーバーフロー回避ロジックの実施形態のブロック図である。

【図5】複数の古くなった／旧式のキャッシュ側アドレス監視格納位置をチェックすることによってオーバーフローモードを回避することを任意選択的に試みる段階、及びそのような複数の古くなった／旧式の格納位置が発見されない場合にはそのオーバーフローモードに入る段階を備える方法の実施形態のブロックフロー図である。

【図6】オーバーフロー構造の実施形態のブロック図である。

【図7A】本発明の複数の実施形態に係る、例示的なインオーダーパイプライン及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7B】本発明の複数の実施形態に係るプロセッサに含まれるべきインオーダーアーキテクチャコアの例示的な実施形態及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8A】本発明の複数の実施形態に係る、オンダイ相互接続ネットワークとの接続及び自己のレベル2(L2)キャッシュローカルサブセットを有する単一プロセッサコアのブロック図である。

【図8B】本発明の複数の実施形態に係る図8Aのプロセッサコアの一部の拡大図である。

【図9】本発明の複数の実施形態に係る、1つより多く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、統合グラフィックスを有してよいプロセッサのブロック図である。

【図10】本発明の一実施形態に係るシステムのブロック図である。

【図11】本発明の実施形態に係る第1のより具体的で例示的なシステムのブロック図である。

【図12】本発明の実施形態に係る第2のより具体的で例示的なシステムのブロック図で

10

20

30

40

50

ある。

【図 1 3】本発明の実施形態に係る S o C のブロック図である。

【図 1 4】本発明の複数の実施形態に係る、ソース命令セットの複数のバイナリ命令をターゲット命令セットの複数のバイナリ命令に変換するソフトウェア命令コンバータの使用を対比させるブロック図である。

【発明を実施するための形態】

【 0 0 0 8 】

アドレスへの複数の書き込みを監視する命令をスケーラブルに実行する複数の方法、装置、及び複数のシステムは、本明細書に開示されている。以下の説明において、多数の具体的な詳細（例えば、複数の具体的な命令、複数の命令機能性、複数のプロセッサ構成、複数のマイクロアーキテクチャ詳細、複数の動作の複数のシーケンスなど）が記載される。しかしながら、複数の実施形態は、これら具体的な詳細を有することなく実施され得る。複数の他の例において、複数の周知の回路、複数の構造及び複数の技術は、説明の理解を不明瞭にすることを回避するために、詳細に示されていない。

【 0 0 0 9 】

図 1 は、プロセッサ 1 0 0 の実施形態のブロック図である。当該プロセッサは、物理プロセッサ、集積回路、またはダイを表す。いくつかの実施形態において、当該プロセッサは、汎用プロセッサ（例えば、デスクトップ、ラップトップ、及び複数の同様のコンピュータに用いられる種類の汎用マイクロプロセッサ）であってよい。代替的に、当該プロセッサは、特定用途向けプロセッサであってよい。複数の適した特定用途向けプロセッサの複数の例は、複数のネットワークプロセッサ、複数の通信プロセッサ、複数の暗号プロセッサ、複数のグラフィックスプロセッサ、複数のコプロセッサ、複数の組み込みプロセッサ、複数のデジタルシグナルプロセッサ（DSP）、及び複数のコントローラ（例えば、複数のマイクロコントローラ）を含むが、単にいくつかの例を挙げるためであって、それらに限定されない。プロセッサは、様々な複合命令セットコンピュータ（CISC）プロセッサ、様々な縮小命令セットコンピュータ（RISC）プロセッサ、様々な超長命令語（VLIW）プロセッサ、それらの様々なハイブリッド、または複数の全く他の種類のプロセッサのうちの何れかであってよい。

【 0 0 1 0 】

当該プロセッサは、複数のプロセッサコア 1 0 2 を有するマルチコアプロセッサである。示された例示的な実施形態において、当該プロセッサは、コア 0 1 0 2 - 0 からコア 7 1 0 2 - 7 を含む 8 つのコア（集合的には、複数のコア 1 0 2 ）を有する。しかし、複数の他の実施形態において、当該プロセッサは、例えば、2 から数百まで、多くの場合は、2 から数十まで（例えば、約 5 から約百まで）、任意の他の所望の数のコアを有してよい。複数のコアの各々は、単一ハードウェアスレッドと複数のハードウェアスレッドとを有してよく、またはいくつかのコアが単一ハードウェアスレッドを有し得る一方で他の複数のコアが複数のハードウェアスレッドを有し得る。例えば、一例示的な実施形態において、複数のコアの各々は、少なくとも 2 つのハードウェアスレッドを有してよいが、本発明の範囲はそれに限定されない。

【 0 0 1 1 】

コアという用語は、多くの場合、集積回路上に配置され、個々のアーキテクチャ状態（例えば、実行状態）を維持できるロジックを指し、当該個々に維持されるアーキテクチャ状態は、複数の専用実行リソースと関連付けられる。対照的に、ハードウェアスレッドという用語は、多くの場合は、集積回路上に配置され、個々のアーキテクチャ状態を維持できるロジックを指し、当該個々に維持されるアーキテクチャ状態は、自己が用いる複数の実行リソースへのアクセスを共有する。複数の特定のリソースがアーキテクチャ状態によって共有され、複数の他のリソースが当該アーキテクチャ状態に対して専用される場合、コアとハードウェアスレッドとの間の境界線は明確でなくなる。しかしながら、コア及びハードウェアスレッドは、多くの場合、オペレーティングシステムからは複数の個々の処理素子または複数の論理プロセッサとして見なされる。オペレーティングシステムは一般

10

20

30

40

50

に、複数のコア、複数のハードウェアスレッド、または複数の他の論理プロセッサまたは複数の処理素子の各々において個別に複数の動作をスケジューリングできる。言い換えれば、一実施形態において、複数の実行リソースが専用され、または共有され、もしくはそれらのいくつかの組み合わせであるか否かにかかわらず、処理素子または論理プロセッサは、ソフトウェアスレッド、オペレーティングシステム、アプリケーションのようなコード、または他のコードと個々に関連付けることができる任意のオンダイプロセッサロジックを表してよい。複数のハードウェアスレッド及び複数のコアに加えて、複数の論理プロセッサまたは複数の処理素子の複数の他の例は、複数のスレッドユニット、複数のスレッドスロット、複数の処理ユニット、複数のコンテキスト、及び/または状態を保持することとコードと個々に関連付けられることができる任意の他のロジックを含むが、それらに限定されない。

10

【0012】

複数のコア102は、1つまたは複数のオンダイ相互接続112によって、一緒に連結される。相互接続は、複数のコアの間で複数のメッセージ及びデータを送信するのに用いられ得る。多くの異なる種類の相互接続が適することは理解されるであろう。一実施形態において、リング相互接続は用いられ得る。複数の代替的な実施形態において、メッシュ、トーラス、クロスバー、ハイパーキューブ、他の相互接続構造、または複数のそのような相互接続のハイブリッドもしくは組み合わせは、用いられ得る。

【0013】

各コアは、例えば、1つまたは複数の下位レベルキャッシュ(図示せず)のようなローカル命令及び/またはデータストレージを含んでよい。例えば、各コアは、当該複数のコアに最も近い、対応する最下位レベルまたはレベル1(L1)キャッシュを含んでよく、かつ、当該複数のコアに次に最も近い、中位レベルまたはレベル2(L2)キャッシュを任意選択的に含んでよい。1つまたは複数の下位レベルキャッシュは、上位レベルキャッシュ(例えば、以下で説明される分散キャッシュ108)よりも、これらの対応する複数のコアに物理的に及び/または論理的に近いので、下位レベルとして称される。1つまたは複数のレベルのキャッシュの各々は、データ及び/または複数の命令をキャッシュし得る。

20

【0014】

複数のコア102はまた、上位レベルの分散キャッシュ108を共有し得る。上位レベルの分散キャッシュは、複数の物理的分散メモリまたは当該キャッシュの複数の部分を表し得る。示された例示的な実施形態において、分散キャッシュは、複数の(例えば、この場合は8)の物理的分散キャッシュ部108-0から108-7(集合的には、複数のキャッシュ部108)を含み、多くの場合、複数のキャッシュスライスとして称される。複数の他の実施形態において、分散キャッシュは、より少ないまたはより多くのキャッシュ部(例えば、プロセッサのコアの数と同数の分散キャッシュ部)を含んでよい。複数の分散キャッシュ部は、複数の異なるコア及び/またはスレッドによって共有されてよい。示されているように、各キャッシュ部は、それぞれのコアとより関連付けられてよく、及び/または任意選択的に物理的により近くダイ上にそれぞれのコアと共に配置(例えば、同一場所に配置)されてよい。例えば、キャッシュ部108-0は、複数の他のコアと比較して、コア0102-0とより関連付けられてよく、及び/または物理的により近くダイ上にコア0102-0と共に配置(例えば、同一場所に配置)されてよい。

30

40

【0015】

いくつかの実施形態において、各キャッシュ部は、複数のメモリアドレスの相互に排他的または重複しない範囲に対応してよくまたはマッピングされてよい。例えば、キャッシュ部108-0は、関連第1のセットのアドレスを有してよく、キャッシュ部108-1は、異なる関連第2のセットのアドレスを有しよく、等等。複数のアドレス範囲は、様々な異なる態様(例えば、複数の異なるハッシュ関数または複数の他のアルゴリズムを用いること)により分散キャッシュの複数の異なるキャッシュ部の中に分割されまたは割り当てられ得る。いくつかの実施形態において、必須ではないが、上位レベルの共有キャッ

50

シュは、データ及び／または複数の命令を格納するように動作可能なラストレベルキャッシュ（LLC）を表し得る。いくつかの実施形態において、分散キャッシュ（例えば、LLC）は、必須ではないが、キャッシュ階層の全ての下位レベルキャッシュを包括してよく、またはキャッシュ階層の最上位レベルの次のレベルのキャッシュ（例えば、L2キャッシュ）を包括してよい。いくつかの実施形態において、複数のコアは、データ及び／または複数の命令について1つまたは複数の下位レベルキャッシュを最初にチェックし得る。1つまたは複数の下位レベルキャッシュにおいて探されているデータ及び／または複数の命令が発見されない場合、当該複数のコアは次に、上位レベルの共有分散キャッシュをチェックすることを進める。

【0016】

示されているように、いくつかの実施形態において、コアインターフェース（I/F）ユニット104は、各々の対応するコア102に連結され得る。各コアインターフェースユニットはまた、相互接続112に連結されてよい。各コアインターフェースユニットは、対応するコアと複数の他のコアとの間、及び対応するコアと複数の分散キャッシュ部との間の中継ぎとして機能するように動作可能であってよい。さらに示されているように、いくつかの実施形態において、対応するキャッシュ制御ユニット106は、各キャッシュスライスまたは他の部分108に連結され得る。いくつかの実施形態において、各キャッシュ制御ユニットは、対応するキャッシュスライスと対応するコアとはおおよそ物理的に同一場所に配置されてよい。各キャッシュ制御ユニットは、相互接続112に連結されてよい。各キャッシュ制御ユニットは、対応する分散キャッシュ部に対してキャッシュコヒーレンシを提供することを制御及び支援するように動作可能であってよい。コアインターフェースユニット104及びキャッシュ制御ユニット106の各々の対応するペアは、対応するコア及び対応するキャッシュ部分を相互接続及び／または複数の他のコアに対してインターフェースするように動作可能なコア・キャッシュ部のインターフェースユニットを集合的に表し得る。複数のコアインターフェースユニット及び複数のキャッシュ制御ユニットは、ハードウェア（例えば、集積回路、複数の回路、複数のトランジスタなど）、ファームウェア（例えば、不揮発性メモリに格納される複数の命令）、ソフトウェア、またはそれらの組み合わせに実装されてよい。

【0017】

プロセッサはまた、当該プロセッサを第1のメモリ（図示せず）に連結させる第1のキャッシュコヒーレンシアウェアメモリコントローラ110 1、及び当該プロセッサを第2のメモリ（図示せず）に連結させる第2のキャッシュコヒーレンシアウェアメモリコントローラ110 2を含む。いくつかの実施形態において、各キャッシュコヒーレンシアウェアメモリコントローラは、キャッシュコヒーレンシを実行するように動作可能なホームエージェントロジック、及びメモリと連携するように動作可能な第2のメモリコントローラロジックを含んでよい。簡潔化のために、本説明において、このようなホームエージェント及びメモリコントローラの複数の機能性は、キャッシュコヒーレンシアウェアメモリコントローラとして称される。複数の他の実施形態は、より少ないまたはより多くのキャッシュコヒーレンシアウェアメモリコントローラを含み得る。さらに、示された実施形態において、複数のキャッシュコヒーレンシアウェアメモリコントローラがオンダイまたはオンプロセッサである一方で、複数の他の実施形態においては、それらは、その代わりに、オフダイ、またはオフプロセッサ（例えば、1つまたは複数のチップセットコンポーネントとして）であり得る。

【0018】

プロセッサがまた本明細書の様々な実施形態を理解するために必要ではない複数の他のコンポーネントを含み得ることは、理解されるであろう。例えば、プロセッサは、入力及び／または出力デバイスに対する1つまたは複数のインターフェース、システムインターフェース、ソケットツーソケット相互接続などを任意選択的に含み得る。上述されたように、複数の特定のプロセッサ（例えば、インテル（登録商標）コーポレーションから入手可能なもの）は、共有メモリに対するスレッド同期を実現するようにMONITOR及び

10

20

30

40

50

MWAIT命令を用いることができる。ハードウェアスレッドまたは他の論理プロセッサは、監視ユニットによって監視されるべきリニアアドレス範囲を設定し、監視ユニットに準備させまたは作動させるように、MONITOR命令を用いてよい。アドレスは、汎用レジスタ（例えば、EAX）を通して提供され得る。アドレス範囲は一般に、ライトバックキャッシング型である。監視ユニットは、当該監視ユニットをトリガするアドレス範囲内のアドレスへの複数の格納／書き込みを監視及び検出する。複数の他の汎用レジスタ（例えば、ECX及びEDX）は、監視ユニットへと他の情報を通信するのに用いられてよい。MWAIT命令は、プログラム順序に沿ってMONITOR命令に続いてよく、ハードウェアスレッドまたは他の論理プロセッサが命令実行を停止して実装依存状態に入ることが可能にするヒントとして機能し得る。例えば、論理プロセッサは、スリープ状態、電力C状態（Cステート）、または他の低減電力消費状態に入り得る。論理プロセッサは、MONITOR命令と関連付けられるクオリファイングイベントのセットのうちの1つが検出されるまでその状態を保ち得る。先行のMONITOR命令によって準備されたアドレス範囲のアドレスへの書き込み／格納は、そのようなクオリファイングイベントの1つである。複数のそのような場合において、論理プロセッサは、その状態から出てよく、プログラム順序に沿ってMWAIT命令に続く命令を実行することを再開し得る。複数の汎用レジスタ（例えば、EAX及びECX）は、監視ユニットへと他の情報（例えば、入るべき状態に関する情報）を通信するのに用いられ得る。

【0019】

図2は、キャッシュエージェント216の実施形態のブロック図である。いくつかの実施形態において、キャッシュエージェントは、図1のプロセッサに用いられ得る。しかし、図2のキャッシュエージェントが図1のプロセッサと異なる複数のプロセッサと共に用いられ得ることが理解されたい。

【0020】

キャッシュエージェント216は、コア202及びキャッシュ部208を含む。いくつかの実施形態において、コアは、マルチコアプロセッサの複数のコアのうちの1つであり得る。いくつかの実施形態において、キャッシュ部は、分散キャッシュの複数のキャッシュスライスまたは複数の他のキャッシュ部（例えば、分散LLC）のうちの1つであってよい。キャッシュエージェントはまた、コアインターフェースユニット204及びキャッシュ部制御ユニット206を含む。コアは、コアインターフェースユニットを通して相互接続212に連結される。キャッシュ部は、キャッシュ部制御ユニットを通して相互接続に連結される。コアインターフェースユニットは、コアとキャッシュ部制御ユニットとの間に連結される。キャッシュ部制御ユニットは、コアインターフェースとキャッシュ部との間に連結される。コア、キャッシュ部、コアインターフェースユニット、及びキャッシュ部制御ユニットは図1の、対応して名付けられている複数のコンポーネントと任意選択的に同様、または同一であってよい。この特定の例において、コアは、第1のハードウェアスレッド218-1及び第2のハードウェアスレッド218-2を含むマルチスレッドコアであるが、本発明の範囲はそれに限定されない。複数の他の実施形態において、コアは、単一スレッドであること、あるいは2つより多くのハードウェアスレッドを有することのいずれであってよい。

【0021】

キャッシュエージェント216は、1つまたは複数のアドレス（例えば、MONITOR命令によって示されたアドレス範囲）への書き込みに対して監視するのに用いられる監視命令（例えば、MONITOR命令）を実行するように動作可能な監視機構を含む。機構は、既存のキャッシュコヒーレンシ機構を利用またはレバレッジし得る（例えば、当該キャッシュコヒーレンシ機構を通して伝達されたアドレスへ書き込む意図の通信を利用し得る）。例示された実施形態において、監視機構は、キャッシュ側アドレス監視ユニット226、コア側アドレス監視ユニット220、コア側トリガユニット234、及びキャッシュ側格納オーバーフローユニット236を含む。本明細書で用いられているように、「コア側」という用語は、相互接続212の、コア202と同じ側上にあること、及び/ま

10

20

30

40

50

たはコアと相互接続との間に配置されること、並びにノもしくはキャッシュ部よりも論理的にコアに近いことを指す。同様に、「キャッシュ側」という用語は、相互接続212の、キャッシュ部208と同じ側上にあること、及びノまたはキャッシュ部と相互接続との間に配置されること、並びにノもしくはコアよりも論理的にキャッシュ部に近いことを指す。

【0022】

示された実施形態において、必須ではないが、キャッシュ側アドレス監視ユニット226及びキャッシュ側格納オーバーフローユニット236は両方、キャッシュ部制御ユニット206に実装される。複数の他の実施形態において、これらのユニットの1つまたは複数は、個別のキャッシュ側コンポーネントとして実装されてよい（例えば、キャッシュ制御ユニット及びノまたはキャッシュ部に連結される）。同様に、示された実施形態において、必須ではないが、コア側アドレス監視ユニット220及びコア側トリガユニット234は両方、コアインターフェースユニット204に実装される。複数の他の実施形態において、これらのユニットの1つまたは複数は、個別のコア側コンポーネントとして、実装され得る（例えば、コアインターフェースユニット及びノまたはコアに連結される）。

【0023】

キャッシュ側アドレス監視ユニット226は、分散キャッシュのスライスまたは他の部分であるキャッシュ部208に対応する。キャッシュ側アドレス監視ユニットは、複数の異なるキャッシュ側アドレス監視格納位置228を有する。示されているように、各キャッシュ側アドレス監視格納位置は、複数の書き込みに対して監視されるべきアドレス230を格納するのに用いられてよい。いくつかの実施形態において、各キャッシュ側アドレス監視格納位置はまた、アドレスに関するコアの表示（例えば、コア識別子、各異なるコアに対応する異なるビットを有するコアマスクなど）を格納し得る。例として、これら複数の格納位置は、ハードウェア実装テーブルにおける複数の異なるエントリを表し得る。示されたように、示された実施形態において、第1のキャッシュ側アドレス監視格納位置228-1から第Nのキャッシュ側アドレス監視格納位置228-Nがあつてよく、Nは、特定の实装に対して適切な数であつてよい。

【0024】

いくつかの実施形態において、キャッシュ部に対応するキャッシュ側アドレス監視ユニットにおけるキャッシュ側アドレス監視格納位置の総数は、プロセッサ及びノまたはプロセッサが実装されるソケットの、ハードウェアスレッド（または他の論理プロセッサ）の総数より少なくてもよい。いくつかの実施形態において、各ハードウェアスレッド（または他の論理プロセッサ）は、単一アドレスまたは複数のアドレスの単一範囲を監視すべく、監視命令（例えば、MONITOR命令）を用いるのに動作可能であつてよい。いくつかの場合において、そのような監視命令を用いた後、ハードウェアスレッドは、スリープ状態に置かれてよく、または別の低減電力消費状態に置かれてよい。1つ可能なアプローチは、監視されるべきアドレスを格納するように、各ハードウェアスレッド（または他の論理プロセッサ）に対して複数のキャッシュ側アドレス監視格納位置228を十分に提供することであろう。しかし、分散キャッシュが用いられている場合、各アドレスは、単一の対応するキャッシュスライスまたは他のキャッシュ部に対してのみ、ハッシュし得、あるいはマッピングし得る。例えば、アドレスのハッシュは、特定のハッシュ関数に従つて当該アドレスに対応する単一の対応するキャッシュスライスを選択し得る。従つて、このような分散キャッシュが用いられている場合、概して非常に低い可能性ではあるが、複数のハードウェアスレッド（または複数の他の論理プロセッサ）の全てに対して監視されるべき複数のアドレスの全ては、全部、同一の単一キャッシュスライスにハッシュされ得、あるいはマッピングされ得る可能性がある。

【0025】

そのような可能性を作るべく、1つ可能なアプローチは、プロセッサ及びノまたはソケットの複数のハードウェアスレッド（または複数の他の論理プロセッサ）の総数に等しい数のキャッシュ側アドレス監視格納位置228を有する各キャッシュ部の提供であろう。

例えば、各コアが2つのハードウェアスレッドを有するという8のコアのプロセッサにおいて、合計16（すなわち、コアの数にコア毎のスレッドの数を乗算された数）のキャッシュ側アドレス監視格納位置は、当該8のキャッシュスライスの各々に対して提供され得る。例として、ハードウェアスレッドの総数に等しい数のエントリを有するハードウェア実装テーブルは、含まれ得る。いくつかの場合において、各格納位置は、対応するハードウェアスレッドに対する固定対応関係またはアサインメントを有し得る。これは、全てのハードウェアスレッドが監視されるべきアドレスを格納することを可能にし得、これら複数のアドレスの全てができる限り同一のキャッシュ部に対してマッピングし得る可能性を作り得て、従って、そのキャッシュ部に対してローカルに格納される必要となるであろう。そのようなアプローチは、最悪の事態のために本質的に設計している。最悪の事態は、一般に考えにくい、仮に実際に発生してしまった場合に処理するための利用可能なアプローチがなかったため、これまで無視されることができなかったからである。

10

【0026】

そのようなアプローチの1つの欠点は、ハードウェアスレッド（または他の論理プロセッサ）の数及び/またはキャッシュ部の数が増加する場合、比較的にスケーラブルではない傾向にある。ハードウェアスレッドの数の増加は、各キャッシュ部に対して必要とされる格納位置の数を増加させる。さらに、キャッシュ部の数の増加は、各追加のキャッシュ部に対してそのような格納位置の追加のセットを加えることを伴う。単にいくつかの例を挙げるが、複数のプロセッサは、32のスレッド、36のスレッド、40のスレッド、56のスレッド、128のスレッド、または256のスレッドより多くのスレッドを有し得る。そのような大きい数のスレッドが用いられる場合、ストレージの量が相当膨大となり得ることは、当業者には容易に分かることであろう。ストレージのそのような実質的な量は、プロセッサの製造コスト、ストレージの提供に必要とされるオンダイの面積の広さ、及び/または当該ストレージによって生じる電力消費を増加させる傾向にある。

20

【0027】

もう1つのアプローチとして、いくつかの実施形態において、キャッシュ部208に対応するキャッシュ側アドレス監視ユニット226のキャッシュ側アドレス監視格納位置228の総数は、プロセッサ及び/またはソケットのハードウェアスレッド（または他の論理プロセッサ）の総数より少なくてもよい。アドレス監視格納オーバーフローの可能性を完全に回避するのに厳密に必要とされる数より少ないアドレス監視格納位置があってもよい。いくつかの実施形態において、各キャッシュ部は、自己と関連しており、オーバーフローを大抵の場合に回避するのに十分ではあるが、全ての例においてそのようなオーバーフローを完全に防ぐには不十分という数のアドレス監視格納位置を有してよい。いくつかの実施形態において、キャッシュ部毎のキャッシュ側アドレス監視格納位置の総数は、必要に応じてオーバーフローの確率が約十万分の一以下、約百万分の一以下、または約千万分の一以下となるようにプロセッサのハードウェアスレッドの総数に対して十分であり得る。いくつかの実施形態において、プロセッサは、約40より多くのハードウェアスレッドを有してよく、キャッシュ部毎のキャッシュ側アドレス監視格納位置の総数は、40より少なくてもよい（例えば、約20から約38まで）。いくつかの実施形態において、プロセッサは、50より多くのハードウェアスレッドを有してよく、キャッシュ部毎のキャッシュ側アドレス監視格納位置の総数は、約50より少なくてもよい（例えば、約20から約45まで、または約25から約40まで、もしくは、約30から約40まで）。いくつかの実施形態において、複数のキャッシュ側アドレス監視格納位置を複数の特定のハードウェアスレッドに指定するまたは割り当てる代わりに、複数の格納位置は、任意の特定のハードウェアスレッドに対応しなくてもよいが、むしろ任意の格納位置は、任意のハードウェアスレッドによって用いられてよい。有利には、プロセッサ及び/またはソケットのハードウェアスレッド（または他の論理プロセッサ）の総数より少ない総数であってキャッシュ部に対応するキャッシュ側アドレス監視ユニットのキャッシュ側アドレス監視格納位置を用いることは、複数の監視命令（例えば、MONITOR命令）を実行するよりスケーラブルなソリューションを提供するのに潜在的に役に立つ。しかしながら、本明細書に開示さ

30

40

50

れている複数の実施形態が、ハードウェアスレッド及び／またはコアの数と関係なしに、並びに／もしくはストレージの総量が大きいかまたは小さいかにかかわらず、ユーティリティを有することは、理解されるであろう。再び図2を参照すれば、キャッシュエージェントは、コア202に対応するコア側アドレス監視ユニット220を含む。コア側アドレス監視ユニットは、対応するコアの1つまたは複数のハードウェアスレッドの数と同数のコア側アドレス監視格納位置を有する。示された実施形態において、第1のコア側アドレス監視格納位置221-1は、第1のハードウェアスレッド218-1に対して固定対応関係を有し、第2のコア側アドレス監視格納位置221-2は、第2のハードウェアスレッド218-2に対して固定対応関係を有する。複数の他の実施形態において、他の数のスレッド及び格納位置が用いられ得る。各コア側アドレス監視格納位置は、対応するコアの対応するハードウェアスレッド218-1, 218-2に対して監視されるべきアドレス222-1, 222-2を格納するように動作可能であってよい。そのような固定対応関係がある場合、格納位置にアドレスを格納することは、アドレスを対応するハードウェアスレッドと関連付け得る。複数の他の実施形態において、仮に複数の格納位置及び複数のハードウェアスレッドの間に固定対応関係がない場合、そうして各格納位置は、監視されるべきアドレスに対応するハードウェアスレッド（例えば、ハードウェアスレッド識別子）の表示を格納するのに用いられ得る。いくつかの実施形態において、各コア側アドレス監視格納位置はまた、対応するコアの対応するハードウェアスレッド218-1, 218-2に対する監視状態224-1, 224-2を格納するように動作可能であってよい。いくつかの実施形態において、各監視状態は、監視有限状態機械（FSM）を表し得る。いくつかの実施形態において、MONITOR命令である場合、本発明の範囲はこれに限定されないが、監視状態は、アイドル状態、投機状態（例えば、監視ロード状態）、及びトリガ準備完了状態（例えば、トリガ待ち状態）のうち任意の1つであってよい。

【0028】

いくつかの実施形態において、キャッシュ側アドレス監視ユニット226及びコア側アドレス監視ユニット220は、1つまたは複数のアドレス（例えば、MONITOR命令によって示されたアドレス範囲のアドレス）への複数の書き込みに対して監視するように連携または一緒に作業し得る。複数の特定のコンセプトをさらに示すため、監視機構が如何にMONITOR及びMWAIT命令を実行し得るかの例を考慮する。第1のハードウェアスレッド218-1は、MONITOR命令を実行し得る。MONITOR命令は、書き込みに対して監視されるべきアドレスを示し得る。第1のハードウェアスレッドは、示された監視アドレスに対して対応するMONITOR要求を発行し得る。MONITOR要求は、第1のコア側アドレス監視ユニット220に、示された監視アドレス222-1を第1のコア側アドレス監視格納位置221-1に格納させ得る。監視状態224-1は、投機状態または監視ロード状態に設定され得る。MONITOR要求は、相互接続212上に、示された監視アドレスに対応するデータを格納すると想定される適切な分散キャッシュ部208へとルーティングされ得る。特定の示された監視アドレスに応じて、当該マッピングのために用いられるハッシュ関数または他のアルゴリズムに基づき、複数の分散キャッシュ部のうちの何れかであってよいことが留意されたい。キャッシュ側アドレス監視ユニットは、示された監視アドレスを、キャッシュ側アドレス監視格納位置230（例えば、位置230-1から230-Nのうち任意の利用可能な1つ）に格納してよい。第1のハードウェアスレッド218-1を有するコア202を識別するコア識別子はまた、コア識別子（ID）232として、キャッシュ側アドレス監視格納位置230に格納され得る。いくつかの実施形態において、コア識別子は、複数のコアのうちの1つを識別する複数のビットのセットであってよい。複数の他の実施形態において、単一格納位置が監視されている同一のアドレスに対して複数のコアによって共有され得るように、コアマスクは、任意選択的に用いられ得る。

【0029】

第1のスレッド218-1はその次に、監視されているアドレスをまた示し得るMWAIT命令を実行し得る。第1のハードウェアスレッドは、示された監視アドレスに対して

10

20

30

40

50

対応する M W A I T 信号を発行し得る。M W A I T 信号に応答して、コア側アドレス監視ユニット 2 2 0 は、トリガ準備ができた状態（例えば、トリガ待ち状態）で監視状態 2 2 4 1 を設定し得る。第 1 のハードウェアスレッドは、例えば、スリープ状態または他の低減電力消費状態のような異なる状態に任意選択的に置かれ得る。代表的には、第 1 のスレッドは、当該スレッドがスリープすべきであって次にスリープする場合に、自己の状態をコンテキストに格納し得る。

【 0 0 3 0 】

続いて、示された監視アドレスへ書き込む意図がある場合（例えば、オーナシップ要求に対する読み出し、示された監視アドレスに関係しているスヌープの無効化、共有状態から排他的状態へと変更する当該アドレスと関連付けられる状態遷移など）、キャッシュ側アドレス監視ユニットは、そのような、アドレスへの書き込む意図を検出し得る。アドレスは、複数の自己の格納位置のうちの 1 つにおける複数のアドレスのうちの 1 つに一致し得る。格納位置に対応する 1 つまたは複数のコアは、例えば、キャッシュ側アドレス監視格納位置に格納されているコア識別子またはコアマスクによって、判断され得る。キャッシュ側アドレス監視ユニットは、示された監視アドレスを格納するのに用いられるキャッシュ側アドレス監視格納位置をクリアし得る。また、例えば、対応するコアに対するスヌープ無効化を送信することによって、対応するコアに信号を送ってもよい。キャッシュ側アドレス監視ユニットは、アドレスへ書き込む意図の通知を、選択的に、そのアドレスを監視していると分かっているそれら 1 つまたは複数のコアに対してのみ、（例えば、オーナシップまたはスヌープ無効化の要求を通じて）指図するのに役に立つ、一種の高度フィルタとして、機能してよい。これら複数の通知は、アドレスを監視している複数のコアのサブセットに選択的に提供される「複数のヒント」を表し得る。有利には、これは、アドレスを監視していない複数のコアへ通知することを回避するのに役に立ち得て、複数の偽のウェイクアップを回避するのに及び／または相互接続上のトラフィックを低減するのに役に立ち得る。

【 0 0 3 1 】

信号を送られたコアにおけるコア側アドレス監視ユニット 2 2 0 は、当該信号を受信し得、当該信号に（例えば、スヌープ無効化に）示されたアドレスを、自己の複数のコア側アドレス監視格納位置における複数の監視アドレスと比較し得る。当該信号のアドレスが、第 1 のハードウェアスレッド 2 1 8 1 に対応する第 1 のコア側監視アドレス格納位置 2 2 1 1 における監視アドレス 2 2 2 1 に一致することを判断し得る。コア側アドレス監視ユニットは、第 1 のハードウェアスレッドが監視されているアドレスに対応することを知り得る。コア側アドレス監視ユニットは、コア側トリガユニット 2 3 4 に、監視されているアドレスへ書き込む意図が観察されていることを信号にて知らせ得る。それは、第 1 のコア側アドレス監視格納位置をクリアし得て、監視状態 2 2 4 1 をアイドル状態に変更し得る。コア側トリガユニットは、トリガ信号（例えば、警告、通知、またはウェイク信号）を、第 1 のハードウェアスレッドに提供するように動作可能であってよい。この実施形態において、コア側トリガユニットは、コア側にあり、ロジックを単純化するのに役に立ち得るが、キャッシュ側にもまた任意選択的に提供されてもよい。第 1 のハードウェアスレッドは、スリープ状態にあった場合は、ウェイクされ得る。

【 0 0 3 2 】

いくつかの実施形態において、キャッシュ側アドレス監視格納位置がオーバーフローし得る可能性がある。例えば、新たな監視要求は、キャッシュ側アドレス監視ユニットにおいて受信され得るが、キャッシュ側アドレス監視格納位置の全てが現在に使用されている場合があるので、新たな監視要求のアドレスを格納するための空き／利用可能なキャッシュ側アドレス監視格納位置がない。示されているように、いくつかの実施形態において、キャッシュ側アドレス監視ユニットは、キャッシュ部に対応するキャッシュ側アドレス監視格納オーバーフローユニット 2 3 6 に連結され得る。いくつかの実施形態において、キャッシュ側アドレス監視格納オーバーフローユニットは、新たな監視要求のアドレスを格納するのに利用可能な複数の空き／利用可能／未使用キャッシュ側アドレス監視格納位置

10

20

30

40

50

がない場合に、アドレス監視格納オーバーフローポリシを強制または実施するように動作可能であってよい。

【0033】

述べられたように、いくつかの実施形態において、コア側アドレス監視ユニットは、自己の対応するコアのハードウェアスレッドの数と同数のコア側アドレス監視格納位置を有し得る。同様に、いくつかの実施形態において、複数の他のコアのコア側アドレス監視ユニットは、これらの対応するコアのハードウェアスレッドの数と同数のコア側アドレス監視格納位置を有し得る。集合的には、これら複数のコア側アドレス監視格納位置は、プロセッサのハードウェアスレッド（または他の論理プロセッサ）の総数に及ぶコア側アドレス監視格納位置の1つのセットを表し得る。有利には、たとえキャッシュ側アドレス監視格納位置にオーバーフローが生じた場合であっても、コア側アドレス監視ユニットは、全てのハードウェアスレッド（または他の論理プロセッサ）に対して監視されているアドレスの全てを格納するのに十分なコア側アドレス監視格納位置を依然と有する。

【0034】

図3は、MONITOR命令及びMWAIT命令を実施するのに適した監視有限状態機械(FSM)347の実施形態の複数の状態を示す図である。実行スレッドからアドレスに対する監視要求を受信すると、監視FSMは、アイドルリング状態340から投機状態341へと遷移343を行い得る。アドレスに対応するデータを格納すべきキャッシュ部が当該アドレスに一致する書き込み要求を受信する場合、または実行スレッドから監視クリア要求が提供される場合、監視FSMが投機状態にある間に、当該監視FSMは、アイドルリング状態340に戻すように遷移344を行い得る。別の監視要求が同一の実行スレッドから提供された場合、監視FSMは、投機状態341に戻すように遷移343を行い得る。当該監視されたアドレスは、適切な場合は調整され得る。一方、投機状態341にある間にMWAIT要求が実行スレッドから提供される場合、監視FSMは、トリガ待ち状態342への遷移345を行い得る。投機状態は、たとえMWAIT要求を受信する前であっても、監視要求を受信されるときから複数のアドレスを追跡しながら、直近に監視されたアドレスに対してのみ複数の監視ウェイクイベントが送信されることを保証するのに役に立ち得る。監視FSMがトリガ待ち状態にある間に、アドレスに対応するデータを格納すべきキャッシュ部が、監視されているアドレスに一致する書き込み要求を受信する場合、監視ウェイクイベントは、実行スレッドへ送信され得る。一方、監視クリア要求は、監視FSMがトリガ待ち状態342にある間に、実行スレッドから提供され得る。このような場合、監視要求は、その実行スレッドに対して消去され得て、実行スレッドへ監視ウェイクイベントを送信する必要がない。しかし、これら2つの場合のいずれにおいても、監視FSMは、アイドルリング状態340に戻す遷移346を行い得る。

【0035】

図4は、複数の監視要求が同一アドレスを示している場合の複数のハードウェアスレッド及び/またはコアに対する単一キャッシュ側アドレス監視格納位置428を再利用するように動作可能なオーバーフロー回避ロジック460の実施形態のブロック図である。当該ロジックは、キャッシュ側アドレス監視格納位置428に連結されるキャッシュ側アドレス監視格納位置再利用ユニット464を含む。キャッシュ側アドレス監視格納位置再利用ユニットは、同一アドレスを示す複数の異なるハードウェアスレッド及び/またはコアから複数の監視要求462を受信し得る。1つ可能なアプローチは、複数の異なるキャッシュ側アドレス監視格納位置（例えば、ハードウェア実装テーブルの複数の異なるエントリ）におけるこの同一アドレスの複数の異なるコピーを格納することであろう。しかしながら、これは、複数またはいくつかの場合においては、多くのキャッシュ側アドレス監視格納位置を消費し得る。

【0036】

もう1つのアプローチとして、いくつかの実施形態において、単一キャッシュ側アドレス監視格納位置428は、監視されるべきアドレス430を格納するのに用いられ得、複数の異なるハードウェアスレッドからの複数の監視要求を表し得る。いくつかの実施形態

において、複数のコアを監視されるべきアドレスと関連付けることができる構造 4 3 2 はまた、キャッシュ側アドレス監視格納位置 4 2 8 に格納される。一例において、当該構造は、コアマスク構造 4 3 2 を含んでよい。コアマスクは、プロセッサのコアの総数と同数のビットを有し得、コアマスクの各ビットは、異なるコアに対して固定対応関係を有し得る。1 つ可能な慣行によれば、各ビットは、対応するコアがアドレスに対して保留中の監視要求を有しないということを示す第 1 の値（例えば、バイナリ 0 へとクリアされる）、または対応するコアがアドレスに対して保留中の監視要求を有するということを示す第 2 の値（例えば、バイナリ 1 へと設定される）を有し得る。その逆の慣行もまた可能である。対応するコアのビットは、キャッシュ側アドレス監視格納位置に格納されているアドレスに対する監視要求が、そのコアから受信されていること、またはアドレスへの書き込みが観察されてコア側ロジックに報告された場合にクリアされていることを示すように、設定され得る。キャッシュ側アドレス監視格納位置がスレッド識別子ではなくアドレスによって追跡されることが留意されたい。有利には、このようにして、複数の異なるコアからの同一アドレスに対する複数の監視要求は、同一の単一キャッシュ側アドレス監視格納位置内に折り畳まれ得る。そのような、複数の異なるスレッド / コアからの複数の要求に対する格納位置の再利用は、キャッシュ側アドレス監視格納位置オーバーフローを回避するのに役に立ち得る。

【 0 0 3 7 】

上述されたように、いくつかの例において、キャッシュ側アドレス監視格納位置の限定される数をオーバーフローすることが可能である。いくつかの実施形態において、オーバーフローモードまたは複数のポリシのセットは、たとえオーバーフローのイベント中であっても、監視機構が正確に動作することを可能にするように、提供され得る。

【 0 0 3 8 】

図 5 は、複数の古くなった / 旧式のキャッシュ側アドレス監視格納位置をチェックし、そのような複数の古くなった / 旧式の格納位置が発見されない場合にはオーバーフローモードに入ることによって、オーバーフローモードを回避することを任意選択的に試みる方法 5 7 0 の実施形態のブロックフロー図である。いくつかの実施形態において、図 5 の複数の動作及び / または方法は、図 1 のプロセッサ及び / または図 2 のキャッシュエージェントによって、及び / またはそれらの内に、実行され得る。図 1 のプロセッサ及び / または図 2 のキャッシュエージェントに対して本明細書で説明されている複数のコンポーネント、複数の機能、及び複数の具体的で選択的な詳細はまた、図 5 の複数の動作及び / または方法に任意選択的に適用する。代替的に、図 5 の複数の動作及び / または方法は、同様のまたは異なるプロセッサ及び / またはキャッシュエージェントによって、及び / または、それらの内に実行され得る。さらに、図 1 のプロセッサ及び / または図 2 のキャッシュエージェントは、図 5 の複数の動作及び / または方法と同一の、同様の、または異なるものを実行し得る。

【 0 0 3 9 】

方法は、ブロック 5 7 1 で、受信された監視要求を処理する利用可能 / 未使用キャッシュ側アドレス監視格納位置がないことを判断する段階を任意選択的に含む。例えば、監視要求は、キャッシュ側アドレス監視ユニット（例えば、キャッシュ側アドレス監視ユニット 2 2 6）において受信され得、キャッシュ側アドレス監視ユニットは、当該監視要求を処理する利用可能 / 未使用キャッシュ側アドレス監視格納位置がないことを判断し得る。例えば、キャッシュ側アドレス監視格納位置の全ては、監視されるべきアドレスを直ちに格納し得る。方法は、ブロック 5 7 2 で、古くなった / 旧式のキャッシュ側アドレス監視格納位置が存在して新たに受信された監視要求を処理するのに用いられることができるかを判断する段階を任意選択的に含む。いくつかの実施形態において、キャッシュ側アドレス監視ユニットは、アドレスを有するエントリを選択して古くなった及び / または旧式かを判断し得る。例として、古くなった / 旧式のアドレスは、依然と格納位置に格納されているが当該アドレスに対して現に有効な保留中の監視要求がないというアドレスを表し得る。例として、例えば、設定されているが準備されていないという監視に起因し

10

20

30

40

50

て、複数の偽の監視要求の場合があり得る。エントリは、エントリの時期に基づき、あるいは有効性の予測に基づき、ランダムに選択され得る。いくつかの実施形態において、格納位置が古くなったノ旧式か否かをチェックすべく、キャッシュ側アドレス監視ユニットは、関連アドレスに対するスヌープ要求を、アドレスに対する複数の監視要求を有することを示された（例えば、格納位置に格納されているコア識別子またはコアマスクに基づいて判断された）1つまたは複数のコアへ送信し得る。スヌープ要求を受信するコアの1つまたは複数のコア側アドレス監視ユニットは、アドレスが格納されているか否かを判断するように、これらの複数の対応するコア側アドレス監視格納位置をチェックし得る。次に、1つまたは複数のコア側アドレス監視ユニットの各々は、アドレスが依然と有効（例えば、対応するコアからの有効な監視要求に依然と対応している）か否かを示す応答を、キャッシュ側アドレス監視ユニットへ送り返し得る。1つまたは複数のコア側アドレス監視ユニットからの複数の応答が、アドレスに対する任意の複数のそのような有効で依然と保留中の監視要求を示す場合、アドレス及びノまたは格納位置は次に、古くなったノ旧式ではないと判断されてよい。そうでなければ、コア側アドレス監視ユニットがアドレスに対する有効で依然と保留中の監視要求を報告していない場合には、アドレス及びノまたは格納位置は次に、古くなったノ旧式と判断され得る。いくつかの実施形態において、単一格納位置及びノまたはアドレスのみが、そのようなアプローチを用いてチェックされ得る。代替的に、複数の格納位置及びノまたはアドレスは、そのようなアプローチを用いてチェックされ得る。

10

【0040】

20

図5を再び参照すると、ブロック572で新たに受信される監視要求を処理するのに用いられることができるそのような古くなったノ旧式のキャッシュ側アドレス監視格納位置があると判断される場合（すなわち、ブロック572の「はい」がその判断である）、方法は次に、ブロック573へ任意選択的に進めてよい。ブロック573では、古くなったノ旧式のキャッシュ側アドレス監視格納位置は、新たに受信された監視要求を処理するのに任意選択的に用いられ得る。有利には、この場合、オーバーフローモードは、古くなったノ旧式の格納位置を利用することによってこの時点で回避され得る。

【0041】

代替的に、ブロック572でそのような古くなったノ旧式のキャッシュ側アドレス監視格納位置がないと判断される場合（すなわち、ブロック572の「いいえ」がその判断である）、方法は次に、ブロック574へ進めてよい。ブロック574では、方法は、オーバーフローモードに入り得る。オーバーフローモードに入ることは、複数のオーバーフローポリシを強制または実施することを含んでよい。オーバーフローモードにおいて、性能はいく分低下され得る。しかしながら、多くの場合、オーバーフローモードが実行されることはまれにしか必要とされなく、通常、オーバーフローの状況が緩和されるまでの比較的短い期間においてのみ必要とされる。

30

【0042】

一オーバーフローポリシとして、ブロック575では、方法は、全ての読み出しトランザクションに共有キャッシュコヒーレンシ状態を用いることを強制する段階を含んでよい。概念的には、これは、全ての読み出しトランザクションを監視要求として取り扱うとされてよい。オーバーフローモードに入ると、キャッシュ側アドレス監視ユニットはもはや、専用ストレージで複数の監視要求ノアドレスを追跡することができなくなる。従って、キャッシュラインの排他的コピーを有することが可能なコアがない。例えば、キャッシュ側アドレス監視ユニットによって受信された任意の読み出し動作は、共有状態応答で処理され得る。対応するアドレスへ書き込む意図が、スヌープまたはブロードキャストが当該アドレスをキャッシュしたかもしれない複数のコアの全てに提供されるであろうということを保証するのに、そのような読み出しトランザクションに共有状態を用いることを強制することは、役に立ち得る。

40

【0043】

別のオーバーフローポリシとして、ブロック576では、方法は、任意の無効化要求を

50

、複数の保留中の監視要求を有する可能性のある全てのコアへ送信する段階を含む。いくつかの実施形態において、これは、任意の無効化要求が（例えば、所有の読み出し無効化要求またはスヌープ無効化要求などの検出を通じて）検出された場合、複数の保留中の監視要求を有する可能性がある、プロセッサの及び／または同一ソケット内の全てのコアのスヌープ無効化を含み得る。オーバーフローモードに入ると、キャッシュ側アドレス監視ユニットはもはや、専用ストレージで複数の監視要求／アドレスを追跡することができなくなる。従って、複数の保留中の監視要求を有する可能性のある全てのコアには、全ての無効化要求が通知されるはずである。スヌープは、そのような全てのコアの複数のコア側アドレス監視ユニットに到達し得、適切な場合は複数の監視トリガを、関連アドレスに対して有効な保留中の監視要求がある任意の複数のコアへ提供し得る。

10

【0044】

プロセッサの全てのコアに通知することが厳密に要求されなく、むしろ、複数の保留中の監視要求を有する可能性のある全てのコアのみに通知する必要があることが留意されるべきである。いくつかの実施形態において、構造は、オーバーフローが発生する場合に複数の保留中の監視要求を有する可能性のある全てのコアを把握するのに任意選択的に用いられ得る。そのような構造の一例は、選択的なオーバーフロー構造である。オーバーフロー構造は、オーバーフローが発生した場合にどのコアが複数の保留中の監視要求を有する可能性があるかを示し得る。一例において、オーバーフロー構造は、プロセッサのコアの総数と同数のビットを有し得、各ビットは、異なる対応するコアに対して固定対応関係を有し得る。1つ可能な慣行によれば、各ビットは、オーバーフローが発生した場合に対応するコアが保留中の監視要求を有する可能性があるということを示す第1の値（例えば、バイナリワン（binary one）へ設定される）を有し得、または、オーバーフローが発生した場合に対応するコアが保留中の監視要求を有する可能性がないということを示す第2の値（例えば、バイナリゼロ（binary zero）とクリアされる）を有し得る。

20

【0045】

一実施形態において、自己によるオーバーフロー構造は、オーバーフローが発生した場合に複数の保留中の監視要求を有する可能性のある複数のコアの全てを反映し得る。例えば、オーバーフローが発生した場合、オーバーフロー構造は、複数のキャッシュ側アドレス監視格納位置に現に格納されている任意の1つまたは複数のアドレスに対応する全てのコアを反映するように変更され得る。別の実施形態において、複数のキャッシュ側アドレス監視格納位置と組み合わせたオーバーフロー構造は、オーバーフローが発生した場合に複数の保留中の監視要求を有する可能性のある複数のコアの全てを反映し得る。例えば、オーバーフローが発生した場合、キャッシュ側アドレス監視格納位置が新たに受信された監視要求によって上書きされまたは消費されるたびに、当該上書きされまたは消費された複数のアドレスと関連付けられる複数のコアは、オーバーフロー構造に反映され得る。すなわち、オーバーフロー構造は、複数の保留中の監視要求を有する可能性のある複数のコアに関する情報を得るように、格納素子が上書きされるたびに、更新され得る。そのような実施形態において、オーバーフローが発生した場合にどのコアが保留中の監視要求を有する可能性があるかに関する情報は、キャッシュ側アドレス監視格納位置及びオーバーフロー構造の間に分割される。

30

40

【0046】

そのようなオーバーフロー構造または関連構造が用いられる実施形態において、いかなる受信された無効化要求も全てのコアへ送信することが必須ではない。むしろ単に、オーバーフローベクトル及び／または複数の格納位置によって複数の保留中の監視要求を有する可能性があることを示されたそれら複数のコアに送信する。いくつかのコアは、オーバーフローベクトル及び／または複数の格納位置において示されなくてよく、従って、オーバーフローが発生した場合にいかなる保留中の監視要求も有する可能性があるはずがない。従って、複数の無効化要求が送信される必要はない。しかしながら、そのようなオーバーフロー構造の使用は、選択的で必須ではない。

50

【 0 0 4 7 】

再び図5を参照すると、オーバーフローモードは、利用可能な格納位置がない限り、必要に応じてブロック575及び576を繰り返すことによって続いてよい。しかし、ブロック576で、オーバータイムの古くなったノ旧式のアドレス及びノまたは複数の格納位置は、スヌーピングすることによって、またはそうでなければ、任意の無効化要求を、複数の保留中の監視要求を有する可能性がある全てのコアへ送信することによって、能動的に取り除かれ得る。複数のコア側アドレス監視ユニットがスヌーブまたは無効化要求に対して有効で保留中の監視要求を有しない場合、それらは次にこれに関して報告し返し得、キャッシュ側アドレス監視ユニットにコアがアドレスを監視することに関知しないということを反映させる（例えば、コアマスクを更新する）こと、または、当該アドレスに関知する他のコアがない場合に当該格納位置をクリアすることを可能にし得る。様々な実施形態において、古くなったノ旧式の格納位置の除去は、特定のアドレス、特定のキャッシュ部、特定のコアなどに基づいて実行され得る。オーバーフローマスクはまた、複数の古くなったノ旧式の格納位置またはアドレスのクリーンアップを反映するように変更され得る。例えば、保留中の監視要求をもはや有しないコアは、オーバーフローマスクにおいてワン（one）の代わりに、ゼロ（zero）へと更新され得る。このようにして、ブロック576における複数のスヌーブまたは無効化要求は、ある時点でオーバーフローモードから出ることができるように、複数のオーバータイムの古くなったノ旧式の格納素子またはアドレスをクリーンアップするのに役に立ち得る。ブロック577に示されているように、オーバーフローモードから抜け出られ得る。

10

20

【 0 0 4 8 】

これは、単に1つの例示的な実施形態である。この実施形態における多くの変形は、企図される。例えば、ブロック572における判断は、選択的で必須ではない。他の実施形態において、オーバーフローモードは、可能な古くなったエントリノアドレスに対するチェックが無くても自動的に入られてよい。

【 0 0 4 9 】

図6は、オーバーフロー構造680の実施形態のブロック図である。オーバーフロー構造は、オーバーフローが発生した場合にどのコアが保留中の監視要求を有する可能性があるかを示すのに、単独か、またはキャッシュ側アドレス監視格納位置のセットとの組み合わせのどちらかが用いられ得る。この実施形態において、オーバーフロー構造は、各々がN+1のコア（例えば、コア0からコアN）のうち異なる1つに対して固定対応関係を有する、N+1のビットを含む。1つの可能な慣行によれば、各ビットは、オーバーフローが発生した場合に対応するコアが保留中の監視要求を有する可能性があるということを示す第1の値（例えば、バイナリワン（binary one）へ設定される）を有し得、または、オーバーフローが発生した場合に対応するコアが保留中の監視要求を有する可能性があるということを示す第2の値（例えば、バイナリゼロ（binary zero）へクリアされる）を有し得る。例えば、図において、コア0に対応する最左ビットは、コア0が保留中の監視要求を有しないということを示すバイナリゼロ（すなわち、0）を有し、コア1に対応する次の最左ビットは、コア1が保留中の監視要求を有するということを示すバイナリワン（すなわち、1）を有し、コアNに対応する最右ビットは、コアNが保留中の監視要求を有しないということを示すバイナリゼロ（すなわち、0）を有する。これは単に適したオーバーフロー構造の1つ例示的な例である。複数の他の構造が同一または同様の種類の情報を伝達するのに用いられ得ることが理解されたい。例えば、別の実施形態において、複数の保留中の監視要求を有する複数のコアIDのリストは、構造などに格納され得る。

30

40

【 0 0 5 0 】

本明細書に開示されている複数の監視機構、及びそれらの複数のユニットまたはコンポーネントのうちの何れかは、ハードウェア（例えば、集積回路、複数のトランジスタまたは複数の他の回路素子など）、ファームウェア（例えば、ROM、EPROM、フラッシュメモリ、または他の永続的もしくは不揮発性メモリ及びマイクロコード、複数のマイク

50

口命令、もしくはそれらに格納されている複数の他の下位レベル命令)、ソフトウェア(例えば、メモリに格納されている複数の上位レベル命令)、またはそれらの組み合わせ(例えば、1つまたは複数のファームウェア及び/またはソフトウェアと潜在的に組み合わせたハードウェア)に実装され得る。

【0051】

図1、3、4及び6の何れかについて説明されている複数のコンポーネント、複数の機能、及び複数の詳細はまた、図2及び5の何れかに任意選択的に用いられ得る。さらに、本装置の何れかに対して本明細書で説明された複数のコンポーネント、複数の機能、及び複数の詳細はまた、本明細書で説明された方法の何れかにも任意選択的に用いられ得、複数の実施形態においてそのような装置によって及び/またはそれと共に、実行されることが

10

【0052】

複数のプロセッサコアは、複数の異なる目的で、複数の異なる態様により複数の異なるプロセッサに実装され得る。例えば、このような複数のコアの複数の実装は、1)汎用コンピューティング向け汎用インオーダーコア、2)汎用コンピューティング向け高性能汎用アウトオブオーダーコア、及び3)主にグラフィックス及び/またはサイエンティフィック(スループット)コンピューティング向けの特定制用途向けコアを含んでよい。複数の異なるプロセッサの複数の実装は、1)1つまたは複数の汎用コンピューティング向け汎用インオーダーコア及び/または汎用コンピューティング向け1つまたは複数の汎用アウトオブオーダーコアを含むCPU、及び2)主にグラフィックス及び/またはサイエンティフィック(スループット)向けの1つまたは複数の特定制用途向けコアを含むコプロセッサを含んでよい。そのような複数の異なるプロセッサは、1)CPUとは別個のチップ上のコプロセッサ、2)CPUと同一のパッケージ内で個別のダイ上のコプロセッサ、3)CPUと同一のダイ上のコプロセッサ(この場合、そのようなコプロセッサは、場合によっては統合グラフィックス及び/またはサイエンティフィック(スループット)ロジックのような特定制用途向けロジック、または複数の特定制用途向けコアと称される)、及び4)同一のダイ上で、説明されたCPU(場合によっては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称される)、上述したコプロセッサ及び追加的な機能性を含み得るシステムオンチップを含み得る、複数の異なるコンピュータシステムアーキテクチャをもたらす。複数の例示的なコアアーキテクチャが次に説明され、続いて、複数の例示的なプロセッサ及びコンピュータアーキテクチャが説明される。

20

30

[例示的なコアアーキテクチャ][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]

【0053】

図7Aは、本発明の複数の実施形態に係る例示的なインオーダーパイプライン及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7Bは、本発明の複数の実施形態に係るプロセッサに含まれるべきインオーダーアーキテクチャコアの例示的な実施形態及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7A-Bの複数の実線のボックスは、インオーダーパイプライン及びインオーダーコアを示し、任意的追加の複数の破線のボックスは、レジスタリネーミング、アウトオブオーダー発行/実行パイプライン及びコアを示す。インオーダー態様がアウトオブオーダー態様のサブセットであるとして、アウトオブオーダー態様について説明する。

40

【0054】

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、長さ復号段階704、復号段階706、配分段階708、リネーミング段階710、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)段階712、レジスタ読み出し/メモリ読み出し段階714、実行段階716、ライトバック/メモリ書き込み段階718、例外処理段階722及びコミット段階724を含む。

【0055】

50

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、両方ともメモリユニット770に連結される。コア790は、縮小命令セットコンピュータ(RISC)コア、複合命令セットコンピュータ(CISC)コア、超長命令語(VLIW)コア、またはハイブリッドもしくは代替的なコア種類であってよい。また別の選択肢として、コア790は、例えば、ネットワークまたは通信コア、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処理ユニット(GPGPU)コア、グラフィックスコアなどのような特定用途向けコアであってよい。

【0056】

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に連結された分岐予測ユニット732を含み、命令キャッシュユニット734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、TLB736は、命令フェッチユニット738に連結され、命令フェッチユニット738は、復号ユニット740に連結される。復号ユニット740(またはデコーダ)は、複数の命令を復号し得、複数の元の命令から復号され、あるいは反映し、もしくは派生される、1つまたは複数のマイクロ動作、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の出力として生成し得る。復号ユニット740は、様々な異なる機構を用いて実装され得る。複数の適した機構の複数の例は、限定されるものではないが、複数のルックアップテーブル、複数のハードウェア実装、複数のプログラマブルロジックアレイ(PLA)、複数のマイクロコードリードオンリメモリ(ROM)などを含む。一実施形態において、コア790は、マイクロコードROMまたは複数の特定のマイクロ命令に対するマイクロコードを(例えば、復号ユニット740に、またはそうでなければフロントエンドユニット730内に)格納する他の媒体を含む。復号ユニット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に連結される。

【0057】

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つまたは複数のスケジューラユニット756のセットに連結されたりネーム/アロケータユニット752を含む。スケジューラユニット756は、予約ステーション、中央命令ウィンドウなど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756は、物理レジスタファイルユニット758に連結される。物理レジスタファイルユニット758の各々は、1つまたは複数の物理レジスタファイルを表し、それらのうちの複数の異なるものが、スカラー整数、スカラー浮動小数点、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)などのような、1つまたは複数の異なるデータ種類を格納する。一実施形態において、物理レジスタファイルユニット758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、及びスカラーレジスタユニットを含む。これら複数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てよい。物理レジスタファイルユニット758は、リタイアメントユニット754と重なり、レジスタリネーミング及びアウトオブオーダー実行が(例えば、リオーダーバッファ及びリタイアメントレジスタファイルを用いることにより、フューチャファイル、履歴バッファ、及びリタイアメントレジスタファイルを用いることにより、複数のレジスタのレジスタマップ及びプールを用いることにより、など)実装され得る様々な態様を示す。リタイアメントユニット754及び物理レジスタファイルユニット758は、実行クラスタ760に連結される。実行クラスタ760は、1つまたは複数の実行ユニット762のセット及び1つまたは複数のメモリアクセスユニット764のセットを含む。複数の実行ユニット762は、様々な動作(例えば、複数のシフト、加算、減算、乗算)及び様々な種類のデータ(例えば、スカラー浮動小数点、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)を実行し得る。いくつかの実施形態は、複数の具体的な関数または関数のセットに対する専用の多数の実行ユニットを含んでよい一方で、他の実施形態は、1つのみの実行ユニットまたは全ての関数を全てが実行する複数の

10

20

30

40

50

実行ユニットを含んでよい。スケジューラユニット 756、物理レジスタファイルユニット 758 及び実行クラスタ 760 は、複数として示される可能性があるが、その理由は、特定の実施形態は、複数の特定のデータ型 / オペレーションに対して個別のパイプラインを生成するからである（例えば、各々がこれら自己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及び / または実行クラスタを有する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 / またはメモリアクセスパイプライン、及び、個別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 764 を有する特定の実施形態が実装される）。複数の個別のパイプラインが用いられる場合、これらのパイプラインのうちの 1 つまたは複数がアウトオブオーダー発行 / 実行であり、残り

10

【0058】

メモリアクセスユニット 764 のセットは、データ TLB ユニット 772 を含むメモリユニット 770 に連結され、データ TLB ユニット 772 は、データキャッシュユニット 774 に連結され、データキャッシュユニット 774 は、レベル 2 (L2) キャッシュユニット 776 に連結される。例示的な一実施形態では、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、これらの各々は、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、さらに連結される。L2 キャッシュユニット 776 は、1 つまたは複数の他のレベルのキャッシュ、及び最終的にはメインメモリに連結される。

20

【0059】

例として、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 700 を以下のとおり実装してもよい。1) 命令フェッチ 738 が、フェッチ段階 702 及び長さ復号段階 704 を実行し、2) 復号ユニット 740 が、復号段階 706 を実行し、3) リネーム / アロケータユニット 752 が、配分段階 708 及びリネーミング段階 710 を実行し、4) スケジューラユニット 756 が、スケジューリング段階 712 を実行し、5) 物理レジスタファイルユニット 758 及びメモリユニット 770 が、レジスタ読み出し / メモリ読み出し段階 714 を実行し、実行クラスタ 760 が、実行段階 716 を実行し、6) メモリユニット 770 及び物理レジスタファイルユ

30

【0060】

コア 790 は、本明細書で説明される命令を含む 1 つまたは複数の命令セット（例えば、x86 命令セット（複数のより新しいバージョンが追加されているいくつかの拡張を有する）カリフォルニア州サニーベールの MIPS テクノロジーズの MIPS 命令セット、カリフォルニア州サニーベールの ARM ホールディングスの（NEON のような任意の複数の追加的な拡張を有する）ARM 命令セット）をサポートしてよい。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、AVX1、AVX2）をサ

40

【0061】

コアは、マルチスレッディング（複数のオペレーションまたは複数のスレッドの 2 つ又はそれより多くの並列セットを実行すること）をサポートしてもよく、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（単一の物理コアが複数のスレッドのそれぞれに対して論理コアを提供することにより、物理コアが同時にマルチスレッディングをすること）またはそれらの組み合わせ（例えば、時分割フェッチ及び復号化、その後、インテル（登録商標）ハイパースレッディングテクノロジなどでの同時マルチスレッディング）を含む様々な態様で、マルチスレッディングをしてもよいことを理解されるべきである。

50

【 0 0 6 2 】

レジスタリネーミングを、アウトオブオーダー実行との関連で説明しているが、インオーダーアーキテクチャにおいて、レジスタリネーミングが用いられてよいことは理解されるべきである。示されたプロセッサの実施形態は、個別の命令及びデータキャッシュユニット 7 3 4 / 7 7 4、及び共有の L 2 キャッシュユニット 7 7 6 をまた含む一方で、複数の代替的な実施形態は、複数の命令及びデータの両方のために、例えば、レベル 1 (L 1) 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、内部キャッシュとコア及び / またはプロセッサの外部にある外部キャッシュとの組み合わせを含んでよい。代替的に、キャッシュの全ては、コア及び / またはプロセッサの外部にあってよい。[具体的な例示的インオーダーコアアーキテクチャ]

10

【 0 0 6 3 】

図 8 A - B は、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を示し、ここで、コアは、チップ内のいくつかのロジックブロック (同一の種類及び / または異なる種類の他の複数のコアを含む) のうちの 1 つであってもよい。複数のロジックブロックは、用途に応じて、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、いくつかの固定関数ロジック、メモリ I / O インターフェース及び他の必要な I / O ロジックと通信を行う。

【 0 0 6 4 】

図 8 A は、本発明の複数の実施形態に係る単一プロセッサコアのブロック図であり、オンダイ相互接続ネットワーク 8 0 2 との接続及びレベル 2 (L 2) キャッシュローカルサブセット 8 0 4 と共に示される。一実施形態において、命令デコーダ 8 0 0 は、パックド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、スカラー及びベクトルユニット内にキャッシュメモリに対する低レイテンシアクセスを可能にする。一実施形態において (設計の単純化のために)、スカラーユニット 8 0 8 及びベクトルユニット 8 1 0 は、個別の複数のレジスタセット (それぞれ、複数のスカラーレジスタ 8 1 2 及び複数のベクトルレジスタ 8 1 4) を用い、これらの間で転送されるデータは、メモリに書き込まれ、次にレベル 1 (L 1) キャッシュ 8 0 6 から再読み出しされるが、本発明の複数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用い、または、データが書き込み及び再読み出しされることなく 2 つのレジスタフ

20

30

【 0 0 6 5 】

L 2 キャッシュローカルサブセット 8 0 4 は、1つのプロセッサコアあたり1つの個別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、自己の L 2 キャッシュローカルサブセット 8 0 4 に対するダイレクトアクセスパスを有する。プロセッサコアに読み出されたデータは、自己の L 2 キャッシュサブセット 8 0 4 に格納され、迅速かつ、これら自己の複数のローカル L 2 キャッシュサブセットにアクセスする他の複数のプロセッサコアと並列に、アクセスされることができる。プロセッサコアにより書き込まれたデータは、自己の L 2 キャッシュサブセット 8 0 4 に格納され、必要な場合には、他の複数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、共有のデータに対するコヒーレンシを保証する。リングネットワークが双方向であることにより、複数のプロセッサコア、複数の L 2 キャッシュ及び他の複数のロジックブロックのようなエージェントにチップ内で互いに通信を行わせる。各リングデータパスは、一方向あたり 1 0 1 2 ビット幅である。

40

【 0 0 6 6 】

図 8 B は、本発明の複数の実施形態に係る図 8 A のプロセッサコアの一部の拡大図である。図 8 B は、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、並びに、ベクトルユニット 8 1 0 及び複数のベクトルレジスタ 8 1 4 に関するより細かな詳細を含む。具体的には、ベクトルユニット 8 1 0 は、16 ワイドベクトル処理ユニット (V P U) (1 6 ワイド A L U 8 2 8 を参照) であり、1つまたは複数の整数、単精度浮動、及び倍

50

精度浮動命令を実行する。VPUは、スウィズルユニット820による複数のレジスタ入力のスウィズル、数値変換ユニット822A-Bによる数値変換、及び複製ユニット824によるメモリ入力に対する複製をサポートする。複数の書き込みマスクレジスタ826は、結果的な複数のベクトル書き込みを叙述することを可能にする。[統合メモリコントローラ及びグラフィックスを有するプロセッサ]

【0067】

図9は、本発明の複数の実施形態に係るプロセッサ900のブロック図であり、プロセッサ900は、1つより多く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、統合グラフィックスを有してもよい。図9の複数の実線のボックスは、単一のコア902A、システムエージェント910、1つまたは複数のバスコントローラユニット916のセットを有するプロセッサ900を示し、選択的に追加された複数の破線のボックスは、複数のコア902A-Nを有する代替的なプロセッサ900、システムエージェントユニット910内の1つまたは複数の統合メモリコントローラユニット914のセット及び特定用途向けロジック908を示す。

【0068】

こうして、プロセッサ900の異なる複数の実装は、1)(1つまたは複数のコアを含み得る)統合グラフィックス及び/またはサイエンティフィック(スループット)ロジックである特定用途向けロジック908、及び1つ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ブオーダーコア、これら2つの組み合わせ)である複数のコア902A-Nを有するCPU、2)主にグラフィックス及び/またはサイエンティフィック(スループット)向けの多数の特定用途向け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んでよい。このように、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、グラフィックスプロセッサ、GPGPU(汎用グラフィックス処理ユニット)、(30またはそれより多くのコアを含む)ハイスループット多数統合コア(MIC)コプロセッサ、組み込みプロセッサなどのような汎用プロセッサ、コプロセッサまたは特定用途向けプロセッサであり得る。プロセッサは、1つまたは複数のチップ上に実装され得る。プロセッサ900は、例えば、BICMOS、CMOSまたはNMOSのような多数の処理技術の何れかを用いて、1つまたは複数の基板の一部であってよく、及び/またはその上に実装されてもよい。

【0069】

メモリ階層は、複数のコア、1つまたは複数の共有キャッシュユニット906またはそのセット、及び複数の統合メモリコントローラユニット914のセットに連結される外部メモリ(図示せず)内に、1つまたは複数のレベルのキャッシュを含む。複数の共有キャッシュユニット906のセットは、レベル2(L2)、レベル3(L3)、レベル4(L4)または他の複数のレベルのキャッシュなどの1つまたは複数の中位レベルキャッシュ、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912が、統合グラフィックスロジック908、複数の共有キャッシュユニット906のセット及びシステムエージェントユニット910/統合メモリコントローラユニット914を相互接続する一方、複数の代替的な実施形態においては、そのような複数のユニットを相互接続するための任意の数の周知技術を用いてよい。一実施形態において、コヒーレンシは、1つまたは複数のキャッシュユニット904及びコア902A-Nの間に維持される。

【0070】

いくつかの実施形態において、複数のコア902A-Nのうちの1つまたは複数の、マルチスレッディングすることができる。システムエージェント910は、コア902A-Nを調整及び作動するそれら複数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んでよい。PCUは、複数のコア902A-N及び統合グラフィックスロジック908の電力状態を調整するために必要なロジック及び複数のコンポーネントであってよく、またはこれ

らを含んでもよい。ディスプレイユニットは、1つまたは複数の外部接続ディスプレイを駆動するためのものである。

【0071】

複数のコア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ってもよく、すなわち、複数のコア902A-Nのうちの2つまたはそれより多くは、同一の命令セットを実行可能であってよいが、他は、その命令セットまたは異なる命令セットのサブセットのみを実行可能であってもよい。[例示的なコンピュータアーキテクチャ]

【0072】

図10-13は、複数の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、ハンドヘルド型PC、パーソナルデジタルアシスタント、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用の当技術分野で既知の他の複数のシステム設計及び複数の構成はまた適切である。概して、本明細書で開示されるように、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、概して適切である。

【0073】

ここで図10を参照すると、本発明の一実施形態に係るシステム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つまたは複数のプロセッサ1010、1015を含んでよい。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及び入力/出力ハブ(IOH)1050(複数の別個のチップ上にあり得る)を含む。GMCH1090は、メモリコントローラ及びグラフィックスコントローラを含み、これらはメモリ1040及びコプロセッサ1045に連結される。IOH1050は、入力/出力(I/O)デバイス1060をGMCH1090に連結する。代替的に、メモリ及び複数のグラフィックスコントローラ的一方または両方は、プロセッサ内に統合され(本明細書に説明されているように)、メモリ1040及びコプロセッサ1045は、プロセッサ1010及び単一のチップ内でIOH1050を有するコントローラハブ1020に直接に連結される。

【0074】

複数の追加のプロセッサ1015の選択的な特性が、図10に複数の破線で示される。各プロセッサ1010、1015は、本明細書で説明される複数の処理コアのうちの1つまたは複数を含んでもよく、プロセッサ900のあるバージョンであってもよい。

【0075】

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)、位相変化メモリ(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1020は、フロントサイドバス(FSB)のようなマルチドロップバス、クイックバス相互接続(QPI)のような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1010、1015と通信を行う。

【0076】

一実施形態において、コプロセッサ1045は、例えば、ハイスループット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などのような特定用途向けプロセッサである。一実施形態において、コントローラハブ1020は、統合グラフィックスアクセラレータを含んでもよい。

【0077】

アーキテクチャ、マイクロアーキテクチャ、熱、電力消費の特性等を含む、性能の測定

10

20

30

40

50

基準の範囲の観点において、物理リソース 1010、1015の間には様々な差異が存在し得る。

【0078】

一実施形態において、プロセッサ1010は、一般的な種類の複数のデータ処理動作を制御する複数の命令を実行する。複数の命令内に、複数のコプロセッサ命令が組み込まれていてよい。プロセッサ1010は、これらの複数のコプロセッサ命令を、取り付けられたコプロセッサ1045によって実行されるべき種類のものと認識する。従って、プロセッサ1010は、これらの複数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で、コプロセッサ1045に対して発行する。コプロセッサ1045は、受信された複数のコプロセッサ命令を受け付けて実行する。

10

【0079】

ここで、図11を参照すると、本発明の実施形態に係る第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を含む。プロセッサ1170及び1180の各々は、プロセッサ900のある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1170及び1180は、それぞれプロセッサ1010及び1015であり、コプロセッサ1138は、コプロセッサ1045である。別の実施形態において、プロセッサ1170及び1180は、それぞれプロセッサ1010及びコプロセッサ1045である。

20

【0080】

プロセッサ1170及び1180は、統合メモリコントローラ(IMC)ユニット1172及び1182をそれぞれ含むものとして示される。プロセッサ1170はまた、自己のバスコントローラユニットの一部として、複数のポイントツーポイント(P-P)インターフェース1176及び1178をさらに含み、同様に、第2のプロセッサ1180は、複数のP-Pインターフェース1186及び1188を含む。複数のプロセッサ1170、1180は、複数のP-Pインターフェース回路1178、1188を用いたポイントツーポイント(P-P)インターフェース1150を介して、情報を交換してもよい。図11に示されるように、IMC1172及び1182は、複数のプロセッサをそれぞれのメモリ、すなわち、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部分となり得るメモリ1132及びメモリ1134に連結する。

30

【0081】

複数のプロセッサ1170、1180の各々は、複数のポイントツーポイントインターフェース回路1176、1194、1186、1198を用いる個々のポイントツーポイントインターフェース1152、1154を介して、チップセット1190と情報を交換してよい。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得る。一実施形態において、コプロセッサ1138は、例えば、ハイスループット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などのような特定用途向けプロセッサである。

40

【0082】

共有キャッシュ(図示せず)は、一方のプロセッサ内に含まれ、または両方のプロセッサの外側にあってよいが、プロセッサが低電力モード内に置かれた場合、一方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格納され得るように、P-P相互接続を介して複数のプロセッサに接続される。

【0083】

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に連結されてよい。一実施形態において、第1のバス1116は、周辺コンポーネント相互接続(PCI)バス、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

50

続バスのようなバスであってよいが、本発明の範囲はそれに限定されない。

【0084】

図11に示されるように、様々なI/Oデバイス1114が、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、第1のバス1116に連結されてよい。一実施形態において、コプロセッサ、ハイスループットMICプロセッサ、GPGPUのアクセラレータ（例えば、グラフィックスアクセラレータもしくはデジタル信号処理（DSP）ユニットなど）、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などの1つまたは複数の追加のプロセッサ1115が、第1のバス1116に連結される。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）バスであってよい。例えば、キーボード及び/またはマウス1122、通信デバイス1127、及びディスクドライブ、または複数の命令/コード及びデータ1130を含み得る他の大容量ストレージデバイスのような、ストレージユニット1128を含む様々なデバイスが、一実施形態においては、第2のバス1120に連結されてよい。さらに、オーディオI/O1124が、第2のバス1120に連結されてよい。複数の他のアーキテクチャが可能であることは留意されたい。例えば、図11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そのようなアーキテクチャを実装してよい。

10

【0085】

ここで、図12を参照すると、本発明の実施形態に係る第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12における同様の複数の素子には、同様の参照番号が付され、図11の複数の特定の態様は、図12の他の複数の態様の不明瞭を回避すべく、図12では省略されている。

20

【0086】

図12は、複数のプロセッサ1170、1180は、それぞれ統合メモリ及びI/O制御ロジック（「CL」）1172及び1182を含んでよいことを示す。こうして、CL1172、1182は、複数の統合メモリコントローラユニットを含み、かつ、I/O制御ロジックを含む。図12は、メモリ1132、1134がCL1172、1182に連結されるのみならず、複数のI/Oデバイス1214もまた制御ロジック1172、1182に連結されることを示す。複数のレガシI/Oデバイス1215は、チップセット1190に連結される。

30

【0087】

ここで、図13を参照すると、本発明の実施形態に係るSOC1300のブロック図が示される。図9における同様の複数の素子には、同様の参照番号が付される。また、複数の破線のボックスは、複数のより高度なSOC上の複数の任意の機能である。図13では、相互接続ユニット1302は、1つまたは複数のコア902A、N及び共有キャッシュユニット906のセットを含むアプリケーションプロセッサ1310、システムエージェントユニット910、バスコントローラユニット916、統合メモリコントローラユニット914、統合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得る1つまたは複数のコプロセッサ1320またはそのセット、スタティックランダムアクセスメモリ（SRAM）ユニット1330、ダイレクトメモリアクセス（DMA）ユニット1332及び1つ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1340に連結される。一実施形態において、コプロセッサ1320は、特定用途向けプロセッサ、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、GPGPU、ハイスループットMICプロセッサ、組み込みプロセッサ、またはその他のようなものを含む。

40

【0088】

本明細書に開示される複数の機構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、またはそのような複数の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（揮発性及び不揮発性メモリ及び/または複数の格納素子を含む）、少なくとも1つの入力デバ

50

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複数のシステム上で実行する複数の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。

【0089】

図11に示すコード1130のようなプログラムコードは、本明細書に説明される複数の関数を実行し、出力情報を生成するために、複数の入力命令に適用されてよい。出力情報は、1つまたは複数の出力デバイスに、既知の様式で適用されてよい。本願の複数の目的において、処理システムは、例えば、デジタルシグナルプロセッサ(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のようなプロセッサを有する任意のシステムを含む。

【0090】

プログラムコードは、処理システムと通信を行うために、高水準の手順型またはオブジェクト指向プログラミング言語で実装されてよい。プログラムコードはまた、必要に応じて、アセンブリまたは機械言語で実装され得る。実際に、本明細書に説明されている複数の機構は、いかなる特定のプログラミング言語にも範囲限定されない。いかなる場合も、言語はコンパイラ型言語またはインタープリタ型言語であってよい。

【0091】

少なくとも1つの実施形態の1つまたは複数の態様は、プロセッサ内の様々なロジックを表する機械可読媒体上に格納された複数の代表的命令により実装され、命令は、機械により読み出された場合、機械に本明細書で説明されている複数の技術を実行するロジックを作成させ得る。「複数のIPコア」として知られるそのような複数の表現は、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て、実際にロジックまたはプロセッサを作り出す製造機械にロードされてよい。

【0092】

そのような機械可読記憶媒体は、限定的ではないが、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)、リライタブルコンパクトディスク(CD-RW)及び光磁気ディスクを含む任意の他の種類のディスク、リードオンリメモリ(ROM)などの半導体デバイス、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)、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)、位相変化メモリ(PCM)、磁気または光カードなどのランダムアクセスメモリ(RAM)、または複数の電子的命令を格納するために適切な任意の他の種類の媒体などの記憶媒体を含む、機械またはデバイスによって製造もしくは形成される複数の物品の非一時的で有形の構成を含んでよい。従って、本発明の複数の実施形態はまた、非一時的で有形の機械可読媒体を含み、非一時的で有形の機械可読媒体は、複数の命令を含み、または本明細書で説明されている複数の構造、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機能を画定するハードウェア記述言語(HDL)のような設計データを含む。そのような複数の実施形態はまた、複数のプログラム製品と称されてよい。[エミュレーション(バイナリトランスレーション、コードモーフィングなどを含む)]

【0093】

いくつかの場合において、命令コンバータは、ソース命令セットからターゲット命令セットへ命令を変換するのに用いられてよい。例えば、命令コンバータは、命令をコアによって処理される1つまたは複数の他の命令に、トランスレート(例えば、スタティックバイナリトランスレーション、ダイナミックコンパイルを含むダイナミックバイナリトランスレーションを用いて)、モーフィング、エミュレート、あるいは変換してよい。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、またはそれらの組み合わせに実装されてよい。命令コンバータは、プロセッサ上にあってもよく、プロセッサ外にあってもよく、または一部がプロセッサ上で一部がプロセッサ外にあってもよい。

【0094】

図14は、本発明の複数の実施形態に係るソース命令セットの複数のバイナリ命令をタ

10

20

30

40

50

ーゲット命令セットの複数のバイナリ命令に変換するソフトウェア命令コンバータの使用を対比させるブロック図である。例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が、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それらの様々な組み合わせで実装されてもよい。図14は、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1416によってネイティブに実行され得るx86バイナリコード1406を生成するために、x86コンパイラ1404を用いてコンパイルされ得る高水準言語1402のプログラムを示す。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1416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(登録商標)プロセッサと実質的に同一の結果を実現すべく、(1)インテル(登録商標)x86命令セットコアの命令セットの大部分、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、もしくは少なくとも1つのx86命令セットコアを有するインテル(登録商標)プロセッサ上で動作することが想定された他のソフトウェアを互換可能に実行またはそうでなければ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(登録商標)プロセッサと実質的に複数の同一の関数を実行可能な任意のプロセッサを表す。x86コンパイラ1404は、追加のリンケージ処理の有無に関わらず、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1416上で実行可能なx86バイナリコード1406(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す。同様に、図14は、少なくとも1つのx86命令セットコア1414を有しないプロセッサ(例えば、カリフォルニア州サニーベールのMIPSテクノロジーズのMIPSの命令セットを実行及び/またはカリフォルニア州サニーベールの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)によってネイティブに実行され得る、代替的な命令セットバイナリコード1410を生成すべく、代替的な命令セットコンパイラ1408を用いてコンパイルされ得る高水準言語1402のプログラムを示す。命令コンバータ1412は、x86バイナリコード1406を、x86命令セットコアを有しないプロセッサ1414によってネイティブに実行され得るコードに変換するために用いられる。この変換されたコードは、このような変換ができる命令コンバータの製造は難しいので、代替的な命令セットバイナリコード1410と同一となる可能性は低い。しかしながら変換されたコードは、一般的な動作を達成し、代替的な命令セットからの複数の命令により補完される。こうして、命令コンバータ1412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、プロセッサまたはx86命令セットプロセッサもしくはコアを有しない他の電子デバイスにx86バイナリコード1406を実行させるソフトウェア、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す。

【0095】

説明及び特許請求の範囲で、「連結」及び/または「接続」という用語は、それらの複数の派生語と共に用いられてよい。これらの用語は、互いの類義語として意図されるものではないことは理解されるべきである。むしろ、複数の特定の実施形態において、「接続」は、2つまたはそれより多くの要素が、互いに直接物理的にまたは電氣的に接触することを示すために用いられてよい。「連結」は、2つまたはそれより多くの素子が、直接物理的にまたは電氣的に接触することを意味してよい。しかし、「連結」はまた、2つまたはそれより多くの素子が互いに直接接触しないものの、さらにまた互いに連動又は連携することを意味してよい。例えば、コアは、1つまたは複数の介在コンポーネントを通してキャッシュ部に連結され得る。複数の図において、複数の矢印は、複数の接続及び複数の連結を示すのに用いられる。

【0096】

説明及び/または複数の請求項において、「ロジック」、「ユニット」、「モジュール」または「コンポーネント」という複数の用語は、用いられてよい。これらはハードウェア、ファームウェア、ソフトウェア、またはそれらの様々な組み合わせを含んでよいことが理解されるであろう。これらの複数の例は、集積回路、複数の特定用途向け集積回路、複数のアナログ回路、複数のデジタル回路、複数のプログラミングされたロジックデ

10

20

30

40

50

バイス、複数の命令を含む複数のメモリデバイス等、及びそれらの複数の組み合わせを含む。いくつかの実施形態において、これらは、複数のトランジスタ及び／または複数のゲート及び／または複数の他の回路コンポーネントを含んでよい。

【0097】

上述の説明では、複数の実施形態への十分な理解を提供すべく、複数の具体的な詳細が記載されている。しかし、複数の他の実施形態は、これら複数の具体的な詳細のうちのいくつかがなくとも実施されてよい。本発明の範囲は、上で提供された複数の具体的な例によって決定されるものではなく、以下の特許請求の範囲によってのみ決定される。複数の他の例において、複数の周知の回路、複数の構造、複数のデバイス、及び複数の動作は、ブロック図形態に示され、または、説明の理解を不明瞭にすることを回避すべく詳細に示されていない。複数のコンポーネントが示されて説明されている複数の場合において、いくつかの例では、それらは代わりに単一コンポーネントとして一緒に統合され得る。単一コンポーネントが示されて説明されている複数の他の場合において、いくつかの例では、2つまたはそれより多くのコンポーネントに分離されてよい。

【0098】

様々な動作及び方法が説明されている。複数の方法のうちのいくつかは、複数のフロー図において比較的基本的な形態で説明されているが、複数の動作は任意選択的に、追加されてよく、及び／または複数の方法から取り除かれてよい。さらに、複数のフロー図は、複数の例示的な実施形態例による複数の動作の特定の順序を示しているが、その特定の順序は例示的なものである。複数の代替的な実施形態は、異なる順序で複数の動作を任意選択的に実行すること、複数の特定の動作を組み合わせること、複数の特定の動作を重複することなどをし得る。

【0099】

複数の特定の動作は、ハードウェアコンポーネントによって実行されてもよく、または当該複数の動作を実行する複数の命令を用いてプログラミングされた機械、回路、またはハードウェアコンポーネント（例えば、プロセッサ、プロセッサの部分、回路など）を生じさせ、及び／またはもたらすのに用いられ得る機械実行可能、または回路実行可能な複数の命令で具現化されてもよい。複数の動作はまた、ハードウェア及びソフトウェアの組み合わせによって任意選択的に実行されてよい。

【0100】

いくつかの実施形態は、非一時的機械可読記憶媒体を含む製造品（例えば、コンピュータプログラム製品）を含む。非一時的機械可読記憶媒体は、一時的伝搬される信号を含まない。非一時的機械可読記憶媒体は、機械によって可読の形態で情報を格納する機構を含み得る。機械可読記憶媒体は、仮に機械によって実行されたとき及び／またはその場合、当該機械に本明細書に開示されている1つまたは複数の動作、方法、または技術を実行させる及び／またはもたらせるように動作可能な命令または複数の命令のシーケンスを格納してよい。複数の適した機械の複数の例は、これらに限定されないが、複数のプロセッサ及び複数のコンピュータシステムまたはそのような複数のプロセッサを有する複数の他の電子デバイスを含む。様々な例として、非一時的機械可読記憶媒体は、フロッピー（登録商標）ディスク、光記憶媒体、光ディスク、光データストレージデバイス、CD-ROM、磁気ディスク、光磁気ディスク、リードオンリメモリ（ROM）、プログラマブルROM（PROM）、消去可能プログラマブルROM（EPROM）、電気的消去可能プログラマブルROM（EEPROM）、ランダムアクセスメモリ（RAM）、スタティックRAM（SRAM）、ダイナミックRAM（DRAM）、フラッシュメモリ、位相変化メモリ、位相変化データストレージ材料、不揮発性メモリ、不揮発性データストレージデバイス、非一時的メモリ、非一時的データストレージデバイス等を含んでよい。

【0101】

本明細書を通して、「一実施形態」、「実施形態」、「1つまたは複数の実施形態」、「いくつかの実施形態」という言及は、例えば、特定の機能が、本発明の実施に含まれてもよいが、必ずしも必要とされるものではないことを示す。同様に、説明において、様々

な複数の特徴は、場合によって、開示を簡素化し、様々な進歩的な複数の態様に対する理解を助けることを目的として、単一の実施形態において図またはそれらの説明の中で一緒にグループ化される。この開示の方法は、しかしながら、本発明が、各請求項で明確に記述されるよりも多くの機能を必要とするという意図を反映すると解釈されるものではない。むしろ、以下の複数の請求項が反映するように、進歩的な複数の態様は、単一の開示された実施形態の全ての機能の一部にある。こうして、詳細な説明に続く複数の請求項が本発明の個別の実施形態として各々の自己の請求項と共に、この詳細な説明に明確に組み込まれている。[例示的な実施形態]

【0102】

以下の複数の例は、さらなる複数の実施形態に関する。複数の例における具体的事柄は、1つまたは複数の実施形態の何れかの箇所でも用いられ得る。

10

【0103】

第1実施例は、分散キャッシュの第1のキャッシュ部に対応し、プロセッサの論理プロセッサの総数より少ない総数のキャッシュ側アドレス監視格納位置を有するキャッシュ側アドレス監視ユニットを含むプロセッサである。各キャッシュ側アドレス監視格納位置は、監視されるべきアドレスを格納する。プロセッサはまた、第1のコアに対応し、第1のコアの1つまたは複数の論理プロセッサの数と同数のコア側アドレス監視格納位置を有するコア側アドレス監視ユニットを含む。各コア側アドレス監視格納位置は、監視されるべきアドレス及び第1のコアの異なる対応する論理プロセッサに対する監視状態を格納する。プロセッサはまた、第1のキャッシュ部に対応し、監視されるべき追加のアドレスを格納するのに利用可能な未使用キャッシュ側アドレス監視格納位置がない場合にアドレス監視格納オーバーフローポリシを強制するキャッシュ側アドレス監視格納オーバーフローユニットを含む。

20

【0104】

第2実施例は、任意の先行の例のプロセッサを含み、かつ、第1のコアに対応し、コア側アドレス監視ユニットに連結されるコア側トリガユニットを任意選択的に含む。コア側トリガユニットは、対応するコア側アドレス監視格納位置が、トリガの準備ができてトリガイイベントが検出されたという監視状態を有する場合に第1のコアの論理プロセッサをトリガする。

【0105】

30

第3実施例は、任意の先行の例のプロセッサを含み、キャッシュ側アドレス監視ユニットに連結され、複数の異なる論理プロセッサからの同一の監視アドレスに対する複数の監視要求を、共通キャッシュ側アドレス監視格納位置に記録するキャッシュ側アドレス監視格納位置再利用ユニットを任意選択的に含む。

【0106】

第4実施例は、第3実施例のプロセッサを含み、共通キャッシュ側アドレス監視格納位置は、同一の監視アドレスに対して複数の監視要求を提供した複数の異なる論理プロセッサを記録する構造を含む。

【0107】

第5実施例は、任意の先行の例のプロセッサを含み、当該プロセッサは、40より多くのハードウェアスレッドを有し、第1のキャッシュ部に対応するキャッシュ側アドレス監視ユニットのキャッシュ側アドレス監視格納位置の総数は、少なくとも20のキャッシュ側アドレス監視格納位置はあるが、当該40より多くのハードウェアスレッドの総数より少ない。

40

【0108】

第6実施例は、任意の先行の例のプロセッサを含み、キャッシュ側アドレス監視ユニットのキャッシュ側アドレス監視格納位置の総数は、複数のキャッシュ側アドレス監視格納位置のオーバーフローの確率が十万分の一以下となるプロセッサの論理プロセッサの総数に対して十分である。

【0109】

50

第7実施例は、任意の先行の例のプロセッサを含み、監視されるべき第1のアドレスを示す命令にตอบสนองして、キャッシュ側アドレス監視ユニットは、当該第1のアドレスをキャッシュ側アドレス監視格納位置に格納する。また、コア側アドレス監視ユニットは、コア側アドレス監視格納位置に当該第1のアドレスを格納する。

【0110】

第8実施例は、任意の先行の例のプロセッサを含み、複数の論理プロセッサは、複数のハードウェアスレッドである。

【0111】

第9実施例は、任意の先行の例のプロセッサを含み、キャッシュ側アドレス監視格納オーバーフローユニットは、複数の読み出しトランザクションに共有状態を用いることを強制することを含みアドレス監視格納オーバーフローポリシを強制する。

10

【0112】

第10実施例は、任意の先行の例のプロセッサを含み、キャッシュ側アドレス監視格納オーバーフローユニットは、保留中の監視要求を有する可能性のある全てのコアへ複数の無効化要求を送信することを含みアドレス監視格納オーバーフローポリシを強制する。

【0113】

第11実施例は、第10実施例のプロセッサを含み、キャッシュ側アドレス監視格納オーバーフローユニットは、どのコアが保留中の監視要求を有する可能性があるかを判断するオーバーフロー構造をチェックする。

【0114】

20

第12実施例は、相互接続、及び当該相互接続に連結されるプロセッサを含む複数の命令を処理するシステムである。当該プロセッサは、分散キャッシュの第1のキャッシュ部に対応し、当該プロセッサのハードウェアスレッドの総数より少ない総数のアドレス監視格納位置を有する、キャッシュ部制御ユニットの第1のアドレス監視ユニットを含む。各アドレス監視格納位置は、監視されるべきアドレスを格納する。プロセッサはまた、第1のコアに対応し、第1のコアの1つまたは複数のハードウェアスレッドの数と同数のアドレス監視格納位置を有する、コアインターフェースユニットの第2のアドレス監視ユニットを含む。第2のアドレス監視ユニットの各アドレス監視格納位置は、監視されるべきアドレス及び第1のコアの異なる対応するハードウェアスレッドに対する監視状態を格納する。プロセッサは、第1のアドレス監視ユニットの全てのアドレス監視格納位置が用いられて監視要求に対するアドレスを格納するのに利用可能なものがない場合にアドレス監視格納オーバーフローポリシを実施する、キャッシュ部制御ユニットのアドレス監視格納オーバーフローユニットをさらに含む。システムはまた、相互接続に連結されるダイナミックランダムアクセスメモリ、相互接続に連結される無線通信デバイス、及び相互接続に連結される画像キャプチャデバイスを含む。

30

【0115】

第13実施例は、第12実施例のシステムを含み、アドレス監視格納オーバーフローユニットは、共有状態を用いる複数の読み出しトランザクションを実行すること、及び複数の無効化要求を、保留中の監視要求を有する可能性がある全てのコアへ送信することを含みアドレス監視格納オーバーフローポリシを実施する。

40

【0116】

第14実施例は、第12、13実施例のうちの何れかのシステムを含み、当該プロセッサは、40より多くのハードウェアスレッドを有し、第1のアドレス監視ユニットのアドレス監視格納位置の総数は、少なくとも20はあるが、プロセッサの40より多くのハードウェアスレッドの総数よりは少ない。

【0117】

第15実施例は、第12 - 14実施例のうちの何れかのシステムを含み、当該プロセッサは、複数の異なるハードウェアスレッドからの同一の監視アドレスに対する複数の監視要求を共通アドレス監視格納位置に記録する、キャッシュ部制御ユニットのアドレス監視格納位置再利用ユニットをさらに含む。

50

【 0 1 1 8 】

第 1 6 実施例は、プロセッサにおける方法であり、当該方法は、アドレスを示し、かつ、マルチコアプロセッサの第 1 のコアの第 1 の論理プロセッサにおいて当該アドレスへの複数の書き込みに対して監視することを示す第 1 の命令を受信する段階を含む。第 1 の命令に回答して、方法は、第 1 の命令によって示されたアドレスを、第 1 のコアに対応する複数のコア側アドレス監視格納位置の第 1 のコア側アドレス監視格納位置に格納する段階を含む。複数のコア側アドレス監視格納位置の数は、第 1 のコアの論理プロセッサの数に等しい。方法はまた、第 1 の命令によって示されたアドレスを、分散キャッシュの第 1 のキャッシュ部に対応する複数のキャッシュ側アドレス監視格納位置の第 1 のキャッシュ側アドレス監視格納位置に格納する段階を含む。複数のキャッシュ側アドレス監視格納位置の総数は、マルチコアプロセッサの論理プロセッサの総数より少ない。方法は、監視状態を投機状態に変更する段階をさらに含む。

10

【 0 1 1 9 】

第 1 7 実施例は、第 1 6 実施例の方法を含み、かつ、当該アドレスをまた示し、第 2 のコアの第 2 の論理プロセッサにおいて当該アドレスへの複数の書き込みに対して監視することを示す第 2 の命令を受信する段階、及び第 2 のコアに対してアドレスに対する監視要求を第 1 のキャッシュ側アドレス監視格納位置に記録する段階を任意選択的に含む。

【 0 1 2 0 】

第 1 8 実施例は、第 1 7 実施例の方法を含み、第 2 のコアに対してアドレスに対する監視要求を、第 1 のキャッシュ側アドレス監視格納位置に記録する段階は、マルチコアプロセッサの各コアに対応する異なるビットを有するコアマスクにおいてビットを変更する段階を含む。

20

【 0 1 2 1 】

第 1 9 実施例は、先行の例の何れかの方法を含み、かつ、第 2 のアドレスを示し、第 1 の論理プロセッサにおいて当該第 2 のアドレスへの複数の書き込みに対して監視することを示す第 2 の命令を受信する段階、第 1 のキャッシュ部に対応する複数のキャッシュ側アドレス監視格納位置中に利用可能なキャッシュ側アドレス監視格納位置がないことを判断する段階、及びキャッシュ側アドレス監視格納位置オーバーフローモードに入ることを決定する段階を任意選択的に備える。

【 0 1 2 2 】

第 2 0 実施例は、第 1 9 実施例の方法を含み、かつ、キャッシュ側アドレス監視格納位置オーバーフローモードにある際に、共有キャッシュコヒーレンシ状態を用いる第 1 のキャッシュ部に対応する全ての読み出しトランザクションを実行する段階、及び第 1 のキャッシュ部に対応する複数の無効化要求を、1 つまたは複数の保留中の監視要求を有する可能性がある、マルチコアプロセッサの全てのコアへ送信する段階を任意選択的に含む。

30

【 0 1 2 3 】

第 2 1 実施例は、任意の先行の例の方法を含み、かつ、第 1 の論理プロセッサにおいてアドレスを示す第 2 の命令を受信する段階、及び当該第 2 の命令に回答して、監視状態をトリガ待ち状態に変更する段階を任意選択的に備える。

【 0 1 2 4 】

第 2 2 実施例は、第 1 6 - 2 1 実施例のうちの何れかの方法を実行するプロセッサまたは他の装置を含む。

40

【 0 1 2 5 】

第 2 3 実施例は、第 1 6 - 2 1 実施例のうちの何れかの方法を実行するための手段を含むプロセッサまたは他の装置を有する。

【 0 1 2 6 】

第 2 4 実施例は、第 1 6 - 2 1 実施例のうちの何れかの方法を実行するプロセッサを含み、当該プロセッサは、集積回路及び / またはロジック並びに / もしくはユニット、及び / またはコンポーネント並びに / もしくはモジュール、及び / または手段、またはそれらの任意の組み合わせを含む。

50

【 0 1 2 7 】

第 2 5 実施例は、仮に機械によって実行されたとき及び／またはその場合に、第 1 6 - 2 1 実施例のうちの何れかの方法を機械に実行させるように動作可能な 1 つまたは複数の命令を任意選択的に格納あるいは提供する非一時的機械可読媒体を任意選択的に含む。

【 0 1 2 8 】

第 2 6 実施例は、相互接続と、当該相互接続に連結されるプロセッサと、D R A M、グラフィックスチップ、無線通信チップ、位相変化メモリ、及びビデオカメラのうちの少なくとも 1 つとを有するコンピュータシステムを備え、当該少なくとも 1 つは、当該相互接続に連結され、プロセッサ及び／またはコンピュータシステムは、第 1 6 - 2 1 実施例のうちの何れかの方法を実行する。

10

【 0 1 2 9 】

第 2 7 実施例は、実質的に本明細書で説明されたような 1 つまたは複数の動作または任意の方法を実行するプロセッサまたは他の装置を含む。

【 0 1 3 0 】

第 2 8 実施例は、実質的に本明細書で説明されたような 1 つまたは複数の動作または任意の方法を実行するための手段を含むプロセッサまたは他の装置を含む。第 2 9 実施例は、実質的に本明細書に開示されたような命令を実行するプロセッサまたは他の装置を含む。

【 0 1 3 1 】

第 3 0 実施例は、実質的に本明細書に開示されたような命令を実行するための手段を含むプロセッサまたは他の装置を有する。なお、本明細書によれば、以下の各項目もまた実施形態として開示される。

20

[項目 1]

プロセッサであって、

分散キャッシュの第 1 のキャッシュ部に対応し、前記プロセッサの複数の論理プロセッサの総数より少ない総数であって各々が監視されるべきアドレスを格納する複数のキャッシュ側アドレス監視格納位置を有するキャッシュ側アドレス監視ユニットと、

第 1 のコアに対応し、前記第 1 のコアの 1 つまたは複数の論理プロセッサの数と同数であって各々が監視されるべきアドレス及び前記第 1 のコアの異なる対応する論理プロセッサの監視状態を格納する複数のコア側アドレス監視格納位置を有するコア側アドレス監視ユニットと、

30

前記第 1 のキャッシュ部に対応し、監視されるべき追加のアドレスを格納するのに利用可能な未使用キャッシュ側アドレス監視格納位置がない場合にアドレス監視格納オーバーフローポリシを強制するキャッシュ側アドレス監視格納オーバーフローユニットと

を備える

プロセッサ。

[項目 2]

前記第 1 のコアに対応し、前記コア側アドレス監視ユニットに連結され、対応するコア側アドレス監視格納位置が、トリガの準備ができてトリガイイベントが検出されたという監視状態を有する場合に、前記第 1 のコアの論理プロセッサをトリガするコア側トリガユニット

40

をさらに備える

項目 1 に記載のプロセッサ。

[項目 3]

前記キャッシュ側アドレス監視ユニットに連結され、複数の異なる論理プロセッサからの同一の監視アドレスに対する複数の監視要求を、共通キャッシュ側アドレス監視格納位置に記録するキャッシュ側アドレス監視格納位置再利用ユニット

をさらに備える

項目 1 または 2 に記載のプロセッサ。

[項目 4]

50

前記共通キャッシュ側アドレス監視格納位置は、前記同一の監視アドレスに対する前記複数の監視要求を提供した前記複数の異なる論理プロセッサを記録する構造を有する、
項目 3 に記載のプロセッサ。

[項目 5]

前記プロセッサは、40 より多くのハードウェアスレッドを有し、
前記第 1 のキャッシュ部に対応する前記キャッシュ側アドレス監視ユニットの前記複数のキャッシュ側アドレス監視格納位置の前記総数は、少なくとも 20 はあるが、前記 40 より多くのハードウェアスレッドの総数より少ない、
項目 1 から 4 の何れか一項に記載のプロセッサ。

[項目 6]

前記キャッシュ側アドレス監視ユニットの前記複数のキャッシュ側アドレス監視格納位置の前記総数は、前記複数のキャッシュ側アドレス監視格納位置のオーバーフローの確率が十万分の一以下となる前記プロセッサの前記複数の論理プロセッサの前記総数に対して十分である、
項目 1 から 5 の何れか一項に記載のプロセッサ。

[項目 7]

監視されるべき第 1 のアドレスを示す命令に応答して、
前記キャッシュ側アドレス監視ユニットは、前記第 1 のアドレスを、キャッシュ側アドレス監視格納位置に格納し、
前記コア側アドレス監視ユニットは、前記第 1 のアドレスを、コア側アドレス監視格納位置に格納する、

項目 1 から 6 の何れか一項に記載のプロセッサ。

[項目 8]

前記複数の論理プロセッサは、複数のハードウェアスレッドを含む、
項目 1 から 6 の何れか一項に記載のプロセッサ。

[項目 9]

前記キャッシュ側アドレス監視格納オーバーフローユニットは、共有状態を用いるべく、複数の読み出しトランザクションを実行することを含む前記アドレス監視格納オーバーフローポリシーを強制する、
項目 1 から 6 の何れか一項に記載のプロセッサ。

[項目 10]

前記キャッシュ側アドレス監視格納オーバーフローユニットは、保留中の監視要求を有する可能性がある全てのコアへ複数の無効化要求を送信することを含む前記アドレス監視格納オーバーフローポリシーを強制する、
項目 1 から 6 の何れか一項に記載のプロセッサ。

[項目 11]

前記キャッシュ側アドレス監視格納オーバーフローユニットは、どのコアに保留中の監視要求を有する可能性があるかを判断するオーバーフロー構造をチェックする、
項目 10 に記載のプロセッサ。

[項目 12]

複数の命令を処理するシステムであって、
相互接続と、
前記相互接続に連結され、かつ、キャッシュ部制御ユニットの第 1 のアドレス監視ユニット、コアインターフェースユニットの第 2 のアドレス監視ユニット、及び前記キャッシュ部制御ユニットのアドレス監視格納オーバーフローユニットを有するプロセッサと、
前記相互接続に連結されるダイナミックランダムアクセスメモリと、
前記相互接続に連結される無線通信デバイスと、
前記相互接続に連結される画像キャプチャデバイスと
を備え、
前記キャッシュ部制御ユニットの前記第 1 のアドレス監視ユニットは、分散キャッシ

10

20

30

40

50

ユの第1のキャッシュ部に対応し、かつ、前記プロセッサの複数のハードウェアスレッドの総数より少ない総数であって各々が監視されるべきアドレスを格納する複数のアドレス監視格納位置を有し、

前記コアインターフェースユニットの前記第2のアドレス監視ユニットは、第1のコアに対応し、かつ、前記第1のコアの1つまたは複数のハードウェアスレッドの数と同数であって各々が監視されるべきアドレス及び前記第1のコアの異なる対応するハードウェアスレッドに対する監視状態を格納する、前記第2のアドレス監視ユニットの複数のアドレス監視格納位置を有し、

前記キャッシュ部制御ユニットの前記アドレス監視格納オーバーフローユニットは、前記第1のアドレス監視ユニットの全てのアドレス監視格納位置が用いられて監視要求に対してアドレスを格納するのに利用可能なものがない場合に、アドレス監視格納オーバーフローポリシを実施する、

システム。

[項目13]

前記アドレス監視格納オーバーフローユニットは、共有状態を用いるべく複数の読み出しトランザクションを実行すること及び保留中の監視要求を有する可能性がある全てのコアへ複数の無効化要求を送信することを含む前記アドレス監視格納オーバーフローポリシを実施する、

項目12に記載のシステム。

[項目14]

前記プロセッサは、40より多くのハードウェアスレッドを有し、

前記第1のアドレス監視ユニットの前記複数のアドレス監視格納位置の前記総数は、少なくとも20はあるが、前記プロセッサの前記40より多くのハードウェアスレッドの前記総数より少ない、

項目12または13に記載のシステム。

[項目15]

前記プロセッサは、複数の異なるハードウェアスレッドからの同一の監視アドレスに対する複数の監視要求を共通アドレス監視格納位置に記録する、前記キャッシュ部制御ユニットのアドレス監視格納位置再利用ユニットをさらに有する、

項目12または13に記載のシステム。

[項目16]

プロセッサにおける方法であって、

アドレスを示し、かつ、マルチコアプロセッサの第1のコアの第1の論理プロセッサにおいて前記アドレスへの複数の書き込みに対して監視することを示す第1の命令を受信する段階と、

前記第1の命令に応答して、

前記第1のコアに対応し、かつ、数が前記第1のコアの複数の論理プロセッサの数に等しい複数のコア側アドレス監視格納位置の第1のコア側アドレス監視格納位置に、前記第1の命令によって示された前記アドレスを格納する段階と、

分散キャッシュの第1のキャッシュ部に対応し、かつ、総数が前記マルチコアプロセッサの複数の論理プロセッサの総数より少ない複数のキャッシュ側アドレス監視格納位置の第1のキャッシュ側アドレス監視格納位置に、前記第1の命令によって示された前記アドレスを格納する段階と、

監視状態を投機状態に変更する段階と

を備える

方法。

[項目17]

前記アドレスをまた示し、第2のコアの第2の論理プロセッサにおいて前記アドレスへの複数の書き込みに対して監視することを示す第2の命令を受信する段階と、

前記第2のコアのための前記アドレスに対する監視要求を前記第1のキャッシュ側アド

10

20

30

40

50

レス監視格納位置に記録する段階と

をさらに備える

項目 16 に記載の方法。

[項目 18]

前記第 2 のコアのための前記アドレスに対する前記監視要求を前記第 1 のキャッシュ側
アドレス監視格納位置に記録する段階は、前記マルチコアプロセッサの各コアに対応する
異なるビットを有するコアマスクにおけるビットを変更する段階を含む、

項目 17 に記載の方法。

[項目 19]

第 2 のアドレスを示し、前記第 1 の論理プロセッサにおいて前記第 2 のアドレスへの複
数の書き込みに対して監視することを示す第 2 の命令を受信する段階と、

前記第 1 のキャッシュ部に対応する前記複数のキャッシュ側アドレス監視格納位置の中
に利用可能な複数のキャッシュ側アドレス監視格納位置がないことを判断する段階と、

キャッシュ側アドレス監視格納位置オーバーフローモードに入ることを決定する段階と
をさらに備える

項目 16 から 18 の何れか一項に記載の方法。

[項目 20]

前記キャッシュ側アドレス監視格納位置オーバーフローモードにある間は、

共有キャッシュコヒーレンシ状態を用いるべく、前記第 1 のキャッシュ部に対応する全
ての読み出しトランザクションを実行する段階と、

前記第 1 のキャッシュ部に対応する複数の無効化要求を、1 つまたは複数の保留中の監
視要求を有する可能性がある、前記マルチコアプロセッサの全てのコアへ送信する段階と
をさらに備える

項目 19 に記載の方法。

[項目 21]

前記第 1 の論理プロセッサにおいて前記アドレスを示す第 2 の命令を受信する段階と、
前記第 2 の命令に回答して、前記監視状態をトリガ待ち状態に変更する段階と

をさらに備える

項目 16 から 20 の何れか一項に記載の方法。

[項目 22]

項目 16 から 21 の何れかに一項に記載の方法を実行するプロセッサ。

[項目 23]

項目 16 から 21 の何れか一項に記載の方法を実行するための手段を含むプロセッサ。

[項目 24]

1 つまたは複数の命令を格納する非一時的機械可読媒体を含む製造品であって、

前記 1 つまたは複数の命令は、機械により実行された場合、前記機械に項目 16 から 2
1 の何れか一項に記載の方法を実行させるように動作可能である、

製造品。

[項目 25]

相互接続、

前記相互接続に連結されるプロセッサ、

前記相互接続に連結される無線通信チップ、及び

前記相互接続に連結されるビデオカメラ

を備えるコンピュータシステムであって、

前記コンピュータシステムは、項目 16 から 21 の何れか一項に記載の方法を実行する

、

コンピュータシステム。

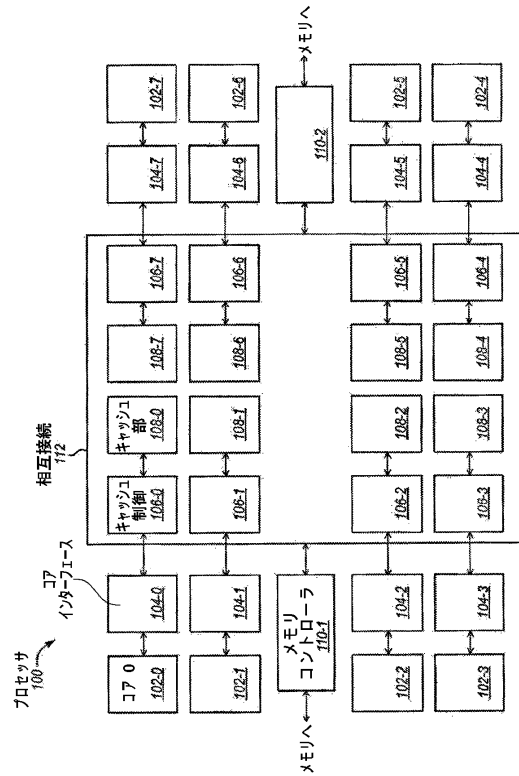
10

20

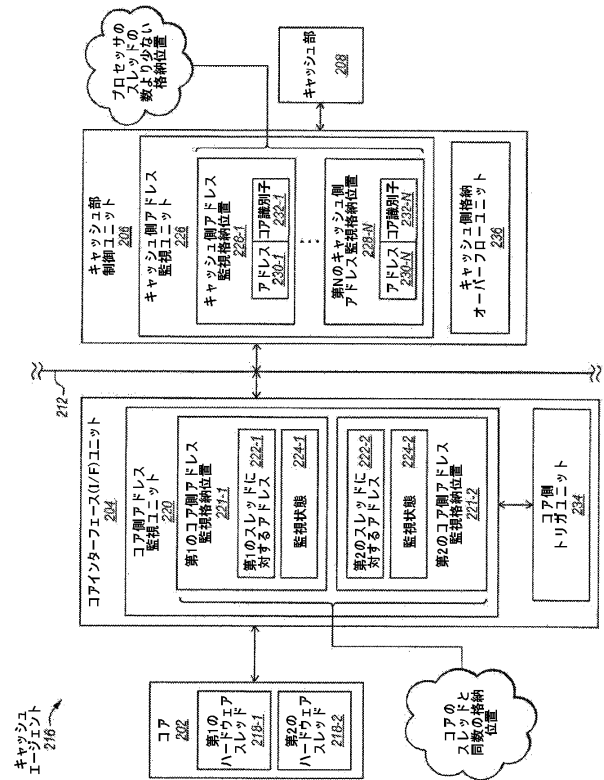
30

40

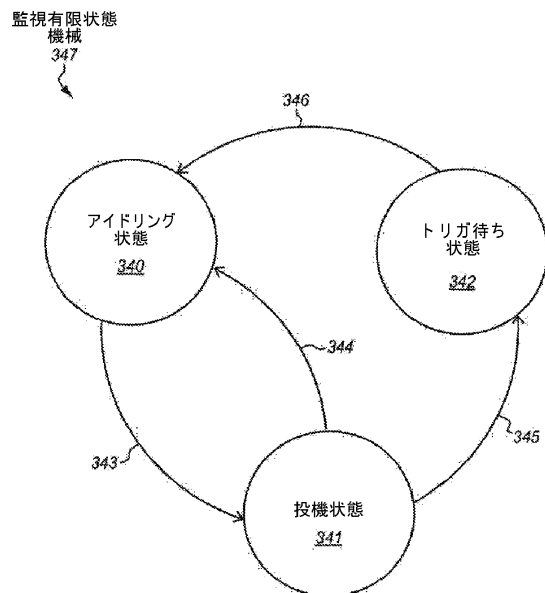
【 図 1 】



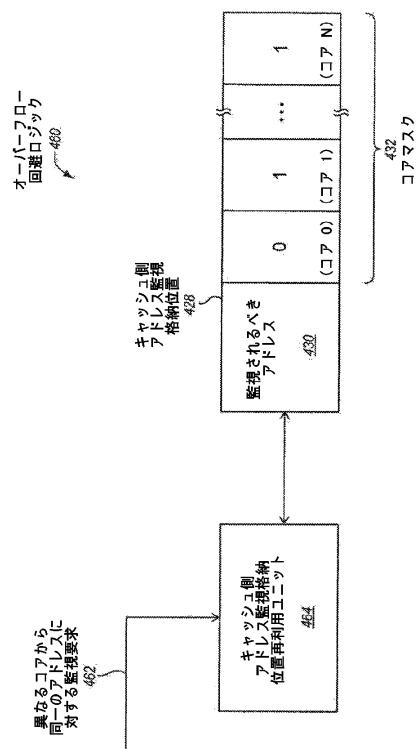
【 図 2 】



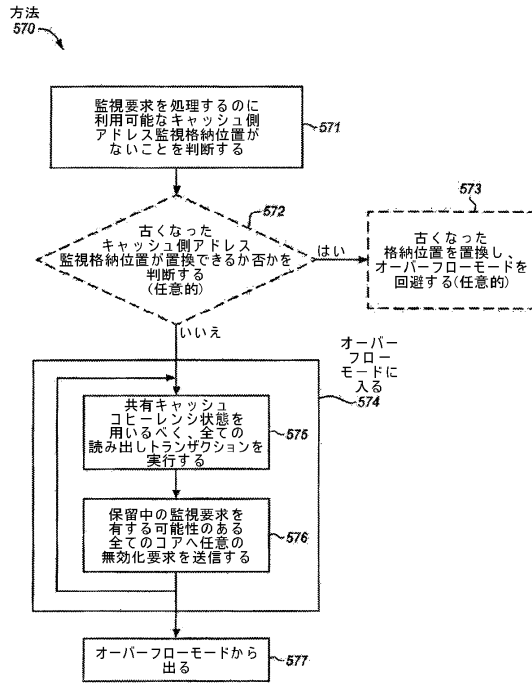
【 図 3 】



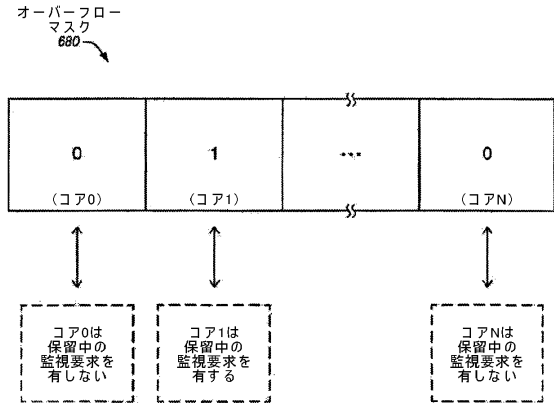
【 図 4 】



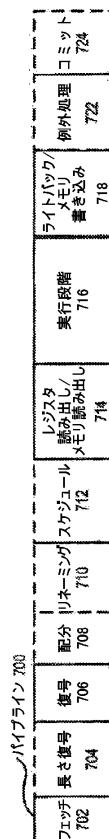
【図 5】



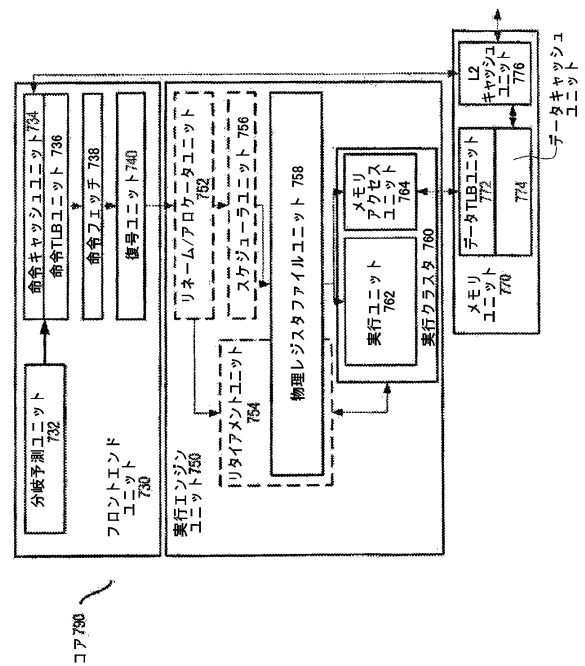
【図 6】



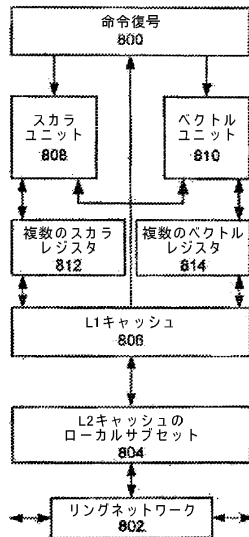
【図 7 A】



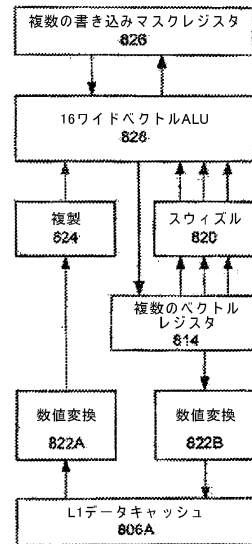
【図 7 B】



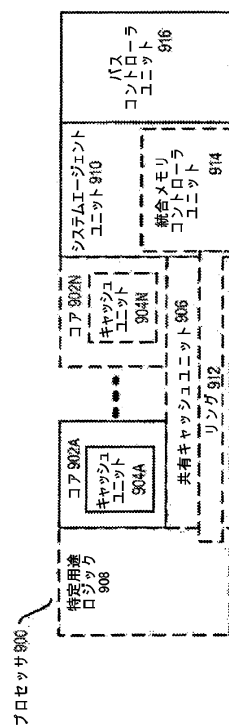
【図 8 A】



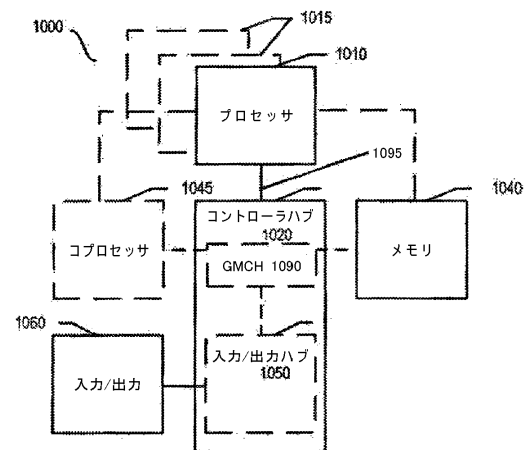
【図 8 B】



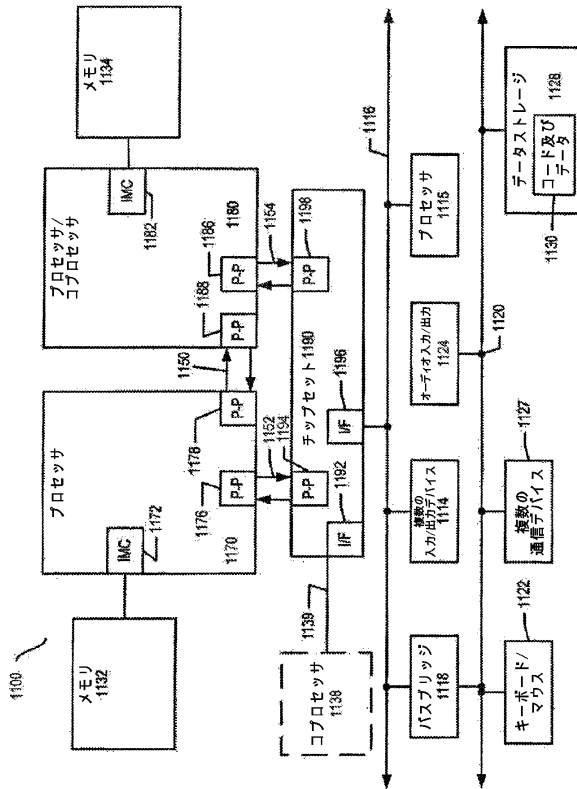
【図 9】



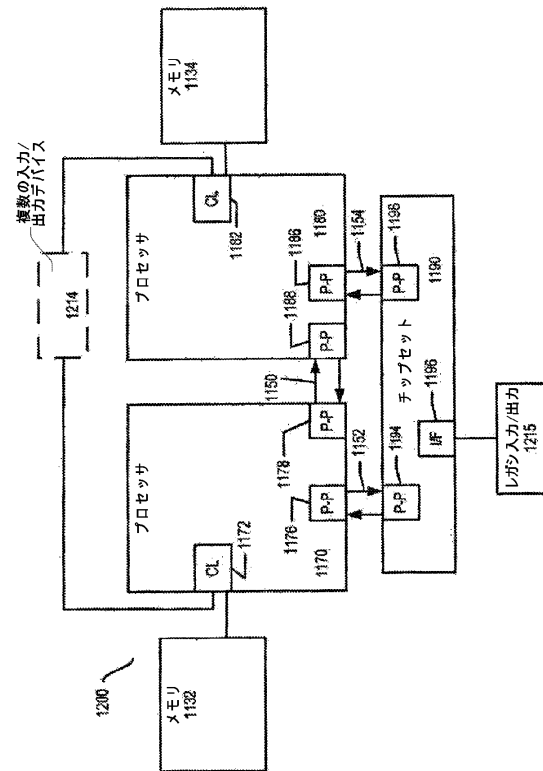
【図 10】



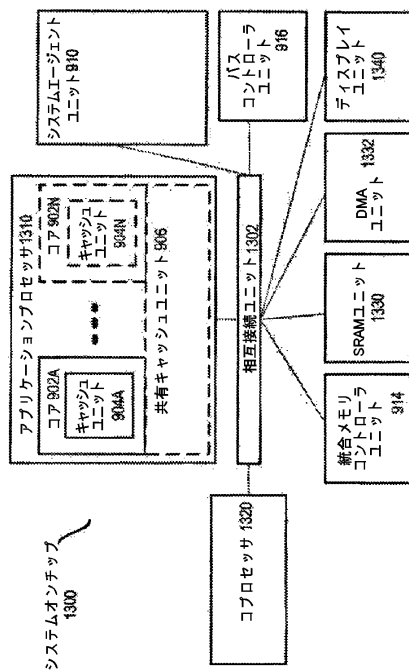
【 図 1 1 】



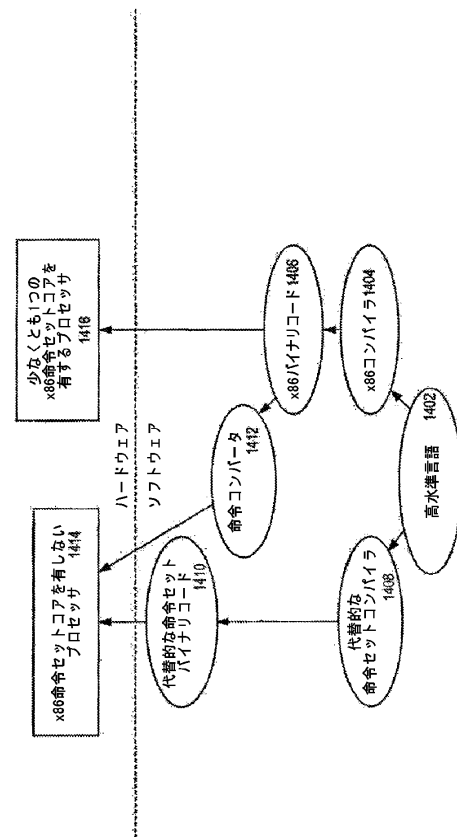
【 図 1 2 】



【 図 1 3 】



【 図 1 4 】



フロントページの続き

- (72)発明者 ファヒム、バハー
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 ハルノア、エリック ジー・
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 チャンベルライン、ジェフリー ディー・
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 バン ドレン、ステファン アール・
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 ブーレ
バード・2200 インテル・コーポレーション内
- (72)発明者 ジュアン、アントニオ
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 塚田 肇

- (56)参考文献 特表2009-540438(JP,A)
特表2007-520769(JP,A)
特表2006-500639(JP,A)
米国特許出願公開第2009/0172284(US,A1)
特表2009-537053(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 12/08
G06F 9/46