

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4691105号
(P4691105)

(45) 発行日 平成23年6月1日(2011.6.1)

(24) 登録日 平成23年2月25日(2011.2.25)

(51) Int.Cl.	F I
G06F 11/16 (2006.01)	G06F 11/16 310D
G06F 11/18 (2006.01)	G06F 11/18 310C

請求項の数 28 (全 10 頁)

(21) 出願番号	特願2007-533793 (P2007-533793)	(73) 特許権者	591003943
(86) (22) 出願日	平成17年9月29日 (2005. 9. 29)		インテル・コーポレーション
(65) 公表番号	特表2008-515064 (P2008-515064A)		アメリカ合衆国 95052 カリフォル
(43) 公表日	平成20年5月8日 (2008. 5. 8)		ニア州・サンタクララ・ミッション カレ
(86) 国際出願番号	PCT/US2005/035375		ッジ ブレーバード・2200
(87) 国際公開番号	W02006/039595	(74) 代理人	100104156
(87) 国際公開日	平成18年4月13日 (2006. 4. 13)		弁理士 龍華 明裕
審査請求日	平成19年3月23日 (2007. 3. 23)	(72) 発明者	マッカジー、シャブヘンドウ
(31) 優先権主張番号	10/953, 887		アメリカ合衆国、01702 マサチュー
(32) 優先日	平成16年9月29日 (2004. 9. 29)		セッツ州、フレーミングハム、ゲーツ ス
(33) 優先権主張国	米国 (US)	(72) 発明者	エメル、ジョエル
			アメリカ合衆国、01720 マサチュー
			セッツ州、アクトン、ブラックホース ド
			ライブ 20

最終頁に続く

(54) 【発明の名称】 冗長マルチスレッド環境でのチェッカ命令の実行

(57) 【特許請求の範囲】

【請求項 1】

コンピュータが、互いに離れて実行される先行スレッドおよび追跡スレッド内に、それぞれのスレッドからの出力値を指定するチェッカ命令を生成し、前記先行スレッド内に、前記出力値を指定する選択された命令を残して前記チェッカ命令を追加し、前記追跡スレッド内の前記選択された命令を前記生成した前記チェッカ命令で置き換える段階と、

前記コンピュータが、前記先行スレッドおよび前記追跡スレッドからの対応するチェッカ命令を待つ段階と、

前記コンピュータが、前記先行スレッドおよび前記追跡スレッドからの前記対応するチェッカ命令を比較する段階とを備える方法。

【請求項 2】

前記先行スレッド内に前記チェッカ命令を追加する場合に、前記選択された命令の前に前記チェッカ命令を挿入する

請求項 1 に記載の方法。

【請求項 3】

前記チェッカ命令は前記先行スレッドおよび前記追跡スレッドの対応するパイプラインでパイプライン処理される

請求項 1 または 2 に記載の方法。

【請求項 4】

前記コンピュータが、前記先行スレッドおよび前記追跡スレッドからの前記対応するチェック命令の比較が一致した場合に、前記チェック命令をコミットする段階をさらに備える請求項 1 から 3 のいずれか一項に記載の方法。

【請求項 5】

前記コンピュータが、前記先行スレッドおよび前記追跡スレッドからの前記対応するチェック命令の比較が一致した場合に、前記先行スレッドの前記選択された命令に含まれる前記出力値を記憶する段階

をさらに備える請求項 4 に記載の方法。

【請求項 6】

前記先行スレッドおよび前記追跡スレッドは、単一のプロセッサによって実行される請求項 1 から 5 のいずれか一項に記載の方法。

10

【請求項 7】

前記先行スレッドおよび前記追跡スレッドは、複数のプロセッサによって実行される請求項 1 から 5 のいずれか一項に記載の方法。

【請求項 8】

前記コンピュータは、前記チェック命令によって、前記先行スレッドの命令をホールドすることなく、前記追跡スレッドの前記対応するチェック命令が現れるまでリタイアポイントをホールドする

請求項 1 から 7 のいずれか一項に記載の方法。

【請求項 9】

20

スレッドからの出力値を指定するチェック命令が当該出力値を指定する選択された命令を残してスレッド内に追加された先行スレッドを実行する回路である先行スレッド回路と、

スレッドからの出力値を指定するチェック命令でスレッド内の前記選択された命令が置き換えられた追跡スレッドを実行する回路である追跡スレッド回路と、

互いに離れて実行される前記先行スレッドおよび前記追跡スレッドからの対応するチェック命令を比較して、前記先行スレッドおよび前記追跡スレッドからの対応するチェック命令をコミットするコミットユニットと

を備える装置。

【請求項 10】

30

前記先行スレッド回路および前記追跡スレッド回路はパイプラインを有する請求項 9 に記載の装置。

【請求項 11】

前記先行スレッドおよび前記追跡スレッドは、単一のプロセッサによって実行される請求項 9 または 10 に記載の装置。

【請求項 12】

前記先行スレッドおよび前記追跡スレッドは、複数のプロセッサによって実行される請求項 9 または 10 に記載の装置。

【請求項 13】

前記先行スレッド回路および前記追跡スレッド回路に結合されるバッファをさらに備える請求項 9 から 12 のいずれか一項に記載の装置。

40

【請求項 14】

前記先行スレッドの前記チェック命令および前記追跡スレッドの前記チェック命令は、前記バッファにおいて対応するチェック命令を待つ請求項 13 に記載の装置。

【請求項 15】

前記選択された命令が指定する出力値は、前記先行スレッドおよび前記追跡スレッドのそれぞれの前記対応するチェック命令の比較が一致する場合に記憶される請求項 9 から 14 のいずれか一項に記載の装置。

【請求項 16】

50

前記コミットユニットは、前記対応するチェック命令が一致しない場合にエラーを生成する

請求項 9 から 15 のいずれか一項に記載の装置。

【請求項 17】

前記先行スレッドにおいて、前記チェック命令は前記選択された命令の前にバイナリトランスレータにより配置される

請求項 9 から 16 のいずれか一項に記載の装置。

【請求項 18】

前記選択された命令はストア命令である

請求項 9 から 17 のいずれか一項に記載の装置。

10

【請求項 19】

前記チェック命令は、前記先行スレッドの命令をホールドすることなく、前記追跡スレッドの前記対応するチェック命令が現れるまでリタイアポイントをホールドする

請求項 9 から 18 のいずれか一項に記載の装置。

【請求項 20】

第 1 プロセッサと、

第 2 プロセッサへの第 1 インタフェースと、

入 / 出力装置への第 2 インタフェースと、

前記第 2 インタフェースに結合されるオーディオ入出力装置と

を備え、

20

前記第 1 プロセッサは、

スレッドからの出力値を指定するチェック命令が当該出力値を指定する選択された命令を残してスレッド内に追加された先行スレッドを実行する回路である先行スレッド回路と

、
スレッドからの出力値を指定するチェック命令で、スレッド内の前記選択された命令が置き換えられた追跡スレッドを実行する回路である追跡スレッド回路と、

互いに離れて実行される前記先行スレッドおよび前記追跡スレッドからの対応するチェック命令を比較して、前記先行スレッドおよび前記追跡スレッドからの対応するチェック命令をリタイアさせるリタイアユニットと
を有するシステム。

30

【請求項 21】

前記先行スレッド回路および前記追跡スレッド回路はパイプラインを含む

請求項 20 に記載のシステム。

【請求項 22】

前記選択された命令が指定する出力値は、前記先行スレッドおよび前記追跡スレッドのそれぞれの前記対応するチェック命令の比較が一致する場合に記憶される

請求項 20 または 21 に記載のシステム。

【請求項 23】

前記先行スレッド回路および前記追跡スレッド回路に結合されるバッファ

をさらに備える請求項 20 から 22 のいずれか一項に記載のシステム。

40

【請求項 24】

前記リタイアユニットは、前記対応するチェック命令が一致しない場合にエラーを生成する

請求項 20 から 23 のいずれか一項に記載のシステム。

【請求項 25】

前記先行スレッド内において、前記チェック命令は前記選択された命令の前にバイナリトランスレータにより配置される

請求項 20 から 24 のいずれか一項に記載のシステム。

【請求項 26】

前記選択された命令はストア命令である

50

請求項 20 から 25 のいずれか一項に記載のシステム。

【請求項 27】

前記第 1 インタフェースおよび前記第 2 インタフェースはポイントツーポイントインタフェースである

請求項 20 から 26 のいずれか一項に記載のシステム。

【請求項 28】

前記チェッカ命令は、前記先行スレッドの命令をホールドすることなく、前記追跡スレッドの前記対応するチェッカ命令が現れるまでリタイアポイントをホールドする

請求項 20 から 27 のいずれか一項に記載のシステム。

【発明の詳細な説明】

10

【背景技術】

【0001】

現在の冗長実行システムは一般に、自己検査式であり且つハードウェアで実装されるチェッカ回路を採用している。チェッカ回路に類似するのは、2つのスレッド（たとえば、ストアアドレス及びデータ）からの結果を比較する比較命令である。両方のスレッドで比較命令を複製し、複製を介して自己検査の効果を得ることが可能であり得る。

【特許文献 1】米国特許第 6463579 号明細書

【発明の開示】

【発明が解決しようとする課題】

【0002】

20

不都合なことに、比較命令を複製することにより、アーキテクチャは冗長マルチスレッド（RMT）の性能利点を失うことになる。RMTの性能利点は、先行スレッドが追跡スレッドのキャッシュミス及び分岐予測をプリフェッチできるように十分に離れた先行スレッド及び追跡スレッドを有することから生じる。比較命令が複製される場合、キューの追加が必要なだけでなく（より高いオーバーヘッドを招く）、アーキテクチャは、両方向で求められる同期により2つのスレッドを十分離れた状態で保持することができなくなり得る。したがって、必要なのは、RMTの性能利点を犠牲にすることなくより低い故障率を実現できる命令である。

【課題を解決するための手段】

【0003】

30

本発明の各種特徴が、同様の参照番号が概して全図面を通して同じパーツを指す添付図面に示す好ましい実施形態の以下の説明から明らかになる。図面は必ずしも一定の縮尺で描かれておらず、それに代えて本発明の原理を示すことに重点が置かれている。

【発明を実施するための最良の形態】

【0004】

以下の説明では、限定ではなく説明のために、特定の構造、アーキテクチャ、インタフェース、技法等の特定の詳細を記して、本発明の各種態様の完全な理解を提供する。しかし、本発明の各種態様を他の例ではこれら特定の詳細から離れて実施可能なことが、本開示の恩恵を受ける当業者には明白であろう。特定の場では、本発明の説明を不必要な詳細で曖昧にしないように、既知の装置、回路、及び方法についての説明を省いている。

40

【0005】

冗長マルチスレッド環境でのチェッカ命令の方法及び装置について記載する。以下の説明では、説明のために、多くの特定の詳細を記して本発明の完全な理解を提供する。しかし、本発明をこれら特定の詳細なしで実施可能なことが当業者には明白であろう。

【0006】

図 1 は、冗長マルチスレッドアーキテクチャの一実施形態のブロック図である。冗長マルチスレッドアーキテクチャでは、故障は、プログラムの 2つのコピーを別個のスレッドとして実行することによって検出することができる。

【0007】

各スレッドには同一の入力が提供され、出力が比較されて、エラーが発生したか否かが

50

判断される。冗長マルチスレッドは、本明細書では「複製範囲 (sphere of replication)」と呼ぶ概念に関連して説明することができる。複製範囲とは、論理的又は物理的な冗長動作の境界である。

【0008】

複製範囲100内の構成要素（たとえば、先行スレッド105を実行しているプロセッサ及び追跡スレッド110を実行しているプロセッサ）が冗長実行の対象である。これとは対照的に、複製範囲100外の構成要素（たとえば、メモリ115）は冗長実行の対象ではない。故障保護は他の技法、たとえばメモリ115の誤り修正符号によって提供される。他の装置は複製範囲100外であってもよく、且つ/又は他の技法を使用して故障保護を複製範囲100外の装置に提供することができる。

10

【0009】

複製範囲100に入るデータは、データを複製し且つデータのコピーを先行スレッド105及び追跡スレッド110に送る入力複製エージェント120を通して入る。同様に、複製範囲100を出るデータは、データを比較し且つエラーが発生したか否かを判断する出力比較エージェント125を通して出る。複製範囲100の境界の変更は、性能とハードウェア量との間のトレードオフである。たとえば、メモリ115を複製すると、ストア命令の出力比較を避けることによりメモリへのより高速でのアクセスが可能になるが、システムでのメモリ量を2倍にすることによってシステムコストが増大する。

【0010】

本発明の一実施形態は、RMTのソフトウェア実装でのチェッカ回路を検査するメカニズムを提案する。RMTはコミットされた命令の出力を比較する（命令毎の比較を必要とする）ため、これはソフトウェアで実装することもできる。RMTのソフトウェア実装があらゆる命令を比較する場合、相当なオーバーヘッドが発生する。しかしそれに代えて、RMTは、ストア命令のみの比較及びロード命令のみの複製を可能にし、これはRMT実装のソフトウェアオーバーヘッドを大幅に低減することができる。

20

【0011】

図2は、チェッカ命令を生成する一方法を示す。まず、大半のコンピュータでのようにコンパイラが命令を生成する。コンパイラから、コンピュータはここで、一連のストア命令であり得るがこれに限定されないバイナリプログラムを有する（200）。次に、バイナリトランスレータがバイナリプログラム中の各ストア命令の前にチェッカ命令を挿入することができる（205）。バイナリトランスレータは当該技術分野において既知の任意のバイナリトランスレータであってもよい。バイナリプログラムの変換後、システムは先行スレッド及び追跡スレッドの両方にバイナリプログラムを作成する。先行スレッドのバイナリプログラムは、チェッカ命令をストア命令に追加する（210）。追跡スレッドのバイナリプログラムは、ストア命令を先行スレッドのピアチェッカ命令で置き換える（215）。

30

【0012】

図3は、チェッカ命令の一実施態様を示す。RMTが比較を要求すると、プロセッサはチェッカ命令を先行スレッド及び追跡スレッドの両方において発することができる（300）。各チェッカ命令は各スレッドから64ビット量を伝達することができる。チェッカ命令は、各パイプラインの終わりにあるバッファに達するまで各スレッドの個々のパイプラインを独立して移行することができる（305）。チェッカ命令は、ピアチェッカ命令をバッファで待つ（310）。これら2つのチェッカ命令は次いで、伝達している64ビット量の比較を行うことができる（315）。一致しない場合、両方がエラーを報告することができる。一致する場合、プロセッサにチェッカ命令をコミットさせることができる（320）。パイプラインは、CMP中の異なるプロセッサからであっても、又はSMTプロセッサと同じマルチスレッドプロセッサからであってもよい。

40

【0013】

この実施態様では、チェッカ命令は、命令キューからの処理命令からの先行する命令を中断しない。むしろ、リタイアポイントのみを、追跡スレッドからの対応するチェッカ命

50

令が現れるまで中断するだけである。また、環境が R M T 環境ではない場合、チェッカ命令は N O P として扱われることができる。

【 0 0 1 4 】

図 4 は、2 つのパイプラインを流れて流れるチェッカ命令の一例のブロック図である。ストア命令 R 1 [R 2] が、レジスタ R 1 中の値をレジスタ R 2 中のアドレスによってポイントされるメモリロケーションに記憶するものと仮定する。このストア命令は、チェッカ命令と共に先行スレッド及び追跡スレッドの両方で複製することができる。先行スレッドのストア命令は、チェッカ命令及びストア命令の両方を以下に示すように含むことができる。

Checkerinst R1

Checkerinst R2

Store: R1 [R2]

したがって、先行スレッドは、パイプライン 4 0 0 を移行する時にストア命令とともにチェッカ命令を含むことができる。

【 0 0 1 5 】

追跡スレッドのストア命令は、以下に示すようにチェッカ命令のみを含むことができる。

Checkerinst R1

Checkerinst R2

したがって、追跡スレッドはパイプライン 4 0 5 を通って移行するストア命令を有さない。

【 0 0 1 6 】

パイプライン 4 0 0 中の先行スレッドからの checkerinst R1 は、バッファ 4 1 0 においてピアチェッカ命令を待つ。パイプライン 4 0 5 中の追跡スレッドからの checkerinst R 1 は、バッファ 4 1 5 においてピアチェッカ命令を待つ。チェッカ命令は常にその相手方すなわちピアを探すか、又は待つ。ミラースレッドがある場合、チェッカ命令はバッファ 4 1 0、4 0 5 においてそのスレッドを探すか、又は待ち、ミラースレッドがそこにあることを確認してから、チェッカ命令を比較する。

【 0 0 1 7 】

先行スレッドからの checkerinst R1 及び追跡スレッドからの checkerinst R1 は、コミット順によりペアになることができ、レジスタ指示子と R 1 の値を比較して、レジスタがいずれのエラーも有さなかったことを保証する。エラーが見つからない場合、チェッカ命令はコミットされる (4 2 0)。チェッカ命令がコミットされると、R 1 の値が記憶される。R 1 の値は、コミットポイントを通してから記憶される。したがって、システムは、従来のように検査をストア毎に行うのではなくすべてのストアを同時に検査することができる。

【 0 0 1 8 】

図 5 は、マルチスレッドプロセッサ用の環境を提供することができるシステムのブロック図である。図 5 に示すシステムは或る範囲のシステムを表すことを意図される。代替のシステムは、より多数、より少数、且つ / 又は異なる構成要素を備えることができる。

【 0 0 1 9 】

システム 5 0 0 は、情報を伝達するバス 5 1 0 又は他の通信装置、及びバス 5 1 0 に結合されて情報を処理するプロセッサ (複数可) 5 2 0 を備える。システム 5 0 0 は、メモリコントローラ 5 3 0 を介してバス 5 1 0 に結合されて、情報及びプロセッサ (複数可) 5 2 0 が実行する命令を記憶するランダムアクセスメモリ (R A M) 又は他のダイナミックメモリ並びにスタティックメモリ、たとえばハードディスク又は他の記憶装置 5 3 5 (メモリと呼ぶ) をさらに備える。メモリ 5 3 5 はまた、プロセッサ (複数可) 5 2 0 による命令の実行中に一時変数又は他の中間情報を記憶するためにも使用することができる。メモリコントローラ 5 3 0 は、1 つ又は複数の種類のメモリ及び / 又は関連するメモリ装置を制御する 1 つ又は複数の構成要素を備えることができる。システム 5 0 0 は、バス 5

10

20

30

40

50

10に結合されてプロセッサ(複数可)520のためにスタティックな情報及び命令を記憶する読み取り専用メモリ(ROM)及び/又は他のスタティック記憶装置540も備える。

【0020】

システム500はまた、バス510を介して入/出力(I/O)インタフェース550にも結合することができる。I/Oインタフェース550はI/O装置555へのインタフェースを提供し、I/O装置555はたとえば、コンピュータユーザに情報を表示する陰極線管(CRT)又は液晶ディスプレイ(LCD)、英数字及び他のキーを含む英数字入力装置、並びに/或いはマウス、トラックボール、又はカーソル方向キー等のカーソル制御装置を含むことができる。システム500は、ローカルエリアネットワーク等のネットワークへのアクセスを有無線で提供するネットワークインタフェース560をさらに備える。

10

【0021】

命令は、メモリ535に磁気ディスク、読み取り専用メモリ(ROM)集積回路、CD-ROM、DVD等の記憶装置から、有無線のリモート接続(たとえば、ネットワークインタフェース860を介してネットワーク経由で)等を介して提供される。

【0022】

これより図6を参照すると、システム600は概して、プロセッサ、メモリ、及び入/出力装置が複数のポイントツーポイントインタフェースによって相互接続されるシステムを示す。システム600はいくつかのプロセッサも備えることができ、明瞭にするために、そのうちの2つプロセッサ605、610のみを示す。プロセッサ605、610はそれぞれ、ローカルメモリコントローラハブ(MCH)615、620を備えてメモリ625、630と接続することができる。プロセッサ605、610は、ポイントツーポイントインタフェース回路640、645を使用して、ポイントツーポイントインタフェース635を介してデータを交換することができる。プロセッサ605、610はそれぞれ、ポイントツーポイントインタフェース回路665、670、675、680を使用して、個々のポイントツーポイントインタフェース655、660を介してチップセット650とデータを交換することができる。チップセット650はまた、高性能グラフィックスインタフェース690を介して高性能グラフィックス回路685とデータを交換することもできる。

20

30

【0023】

チップセット650は、バスインタフェース695を介してバス616とデータを交換することができる。いずれのシステムでも、いくつかの実施形態では低性能グラフィックスコントローラ、ビデオコントローラ、及びネットワーキングコントローラを含む様々な入/出力I/O装置614がバス616上にあり得る。いくつかの実施形態では、別のバスブリッジ618を使用して、バス616とバス620との間でデータを交換できるようにすることができる。いくつかの実施形態では、バス620は、小型コンピュータシステムインタフェース(SCSI)バス、統合ドライブエレクトロニクス(IDE)バス、又はユニバーサルシリアルバス(USB)バスであることができる。追加のI/O装置をバス620に接続することができる。これらは、キーボード、マウスを含むカーソル制御装置622、オーディオI/O624、モデム及びネットワークインタフェースを含む通信装置626、及びデータ記憶装置628を含むことができる。ソフトウェアコード630は、データ記憶装置628に記憶することができる。いくつかの実施形態では、データ記憶装置628は、固定磁気ディスク、フロッピーディスクドライブ、光学ディスクドライブ、光磁気ディスクドライブ、磁気テープ、又はフラッシュメモリを含む不揮発性メモリであることができる。

40

【0024】

本明細書全体を通して、「命令」という語は、命令、マクロ命令、命令バンドル、又はプロセッサ動作を符号化するために使用される任意の数の他のメカニズムを全体的に指すために使用される。

50

【 0 0 2 5 】

以下の説明では、限定ではなく説明のために、特定の構造、アーキテクチャ、インタフェース、技法等の特定の詳細を記して、本発明の各種態様の完全な理解を提供する。しかし、本発明の各種態様を他の例ではこれら特定の詳細から離れて実施可能なことが、本開示の恩恵を受ける当業者には明白であろう。特定の場では、本発明の説明を不必要な詳細で曖昧にしないように、既知の装置、回路、及び方法についての説明を省いている。

【図面の簡単な説明】

【 0 0 2 6 】

【図 1】 マルチスレッドアーキテクチャの一実施形態のブロック図である。

【図 2】 チェッカ命令を生成する方法を示すフローチャートである。

10

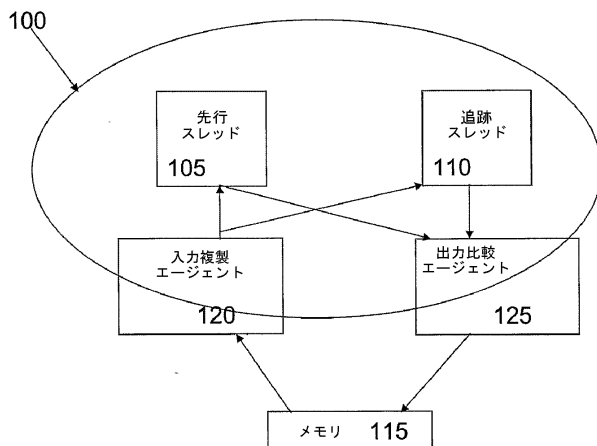
【図 3】 いずれかのスレッドでのチェッカ命令の一実施態様を示すフローチャートである。

【図 4】 チェッカ命令の一実施形態のブロック図である。

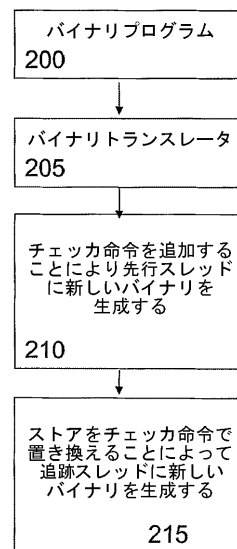
【図 5】 マルチスレッドプロセッサの環境を提供することができるシステムのブロック図である。

【図 6】 マルチスレッドプロセッサの環境を提供することができる代替のシステムのブロック図である。

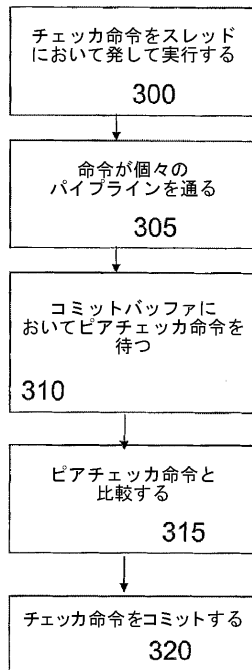
【図 1】



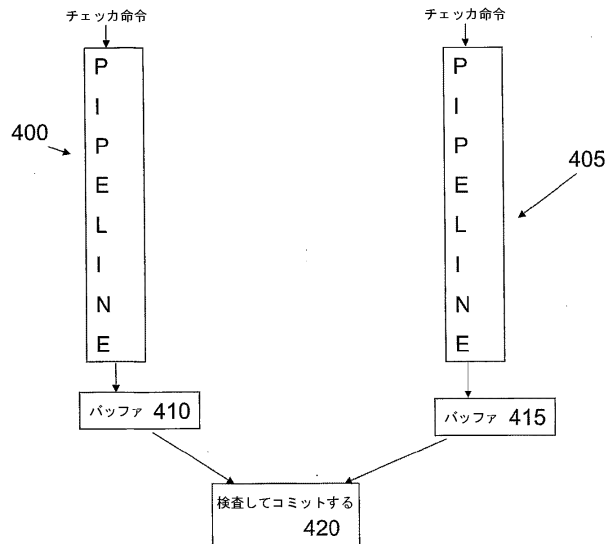
【図 2】



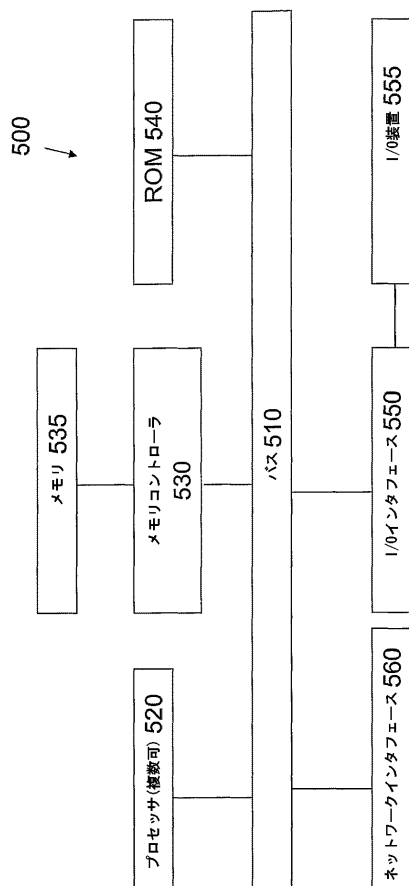
【図 3】



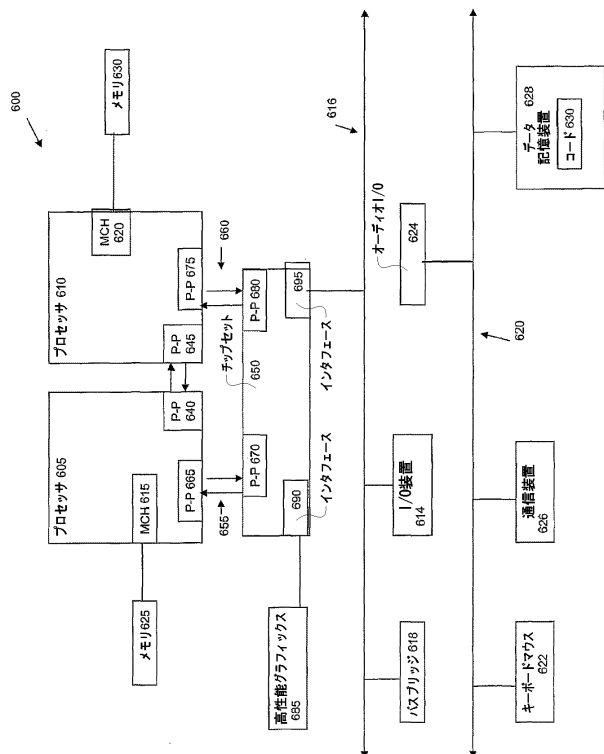
【図 4】



【図 5】



【図 6】



フロントページの続き

- (72)発明者 ラインハルト、スティーブン
アメリカ合衆国、4 8 1 0 5 ミシガン州、アナーバー、キルバン パーク シーアイアール、
3 1 5 8
- (72)発明者 ウィーバー、クリストファー
アメリカ合衆国、0 1 7 5 2 マサチューセッツ州、モールバラ、アップルブライアー レーン
1 0 0 9

審査官 漆原 孝治

- (56)参考文献 特開平11-015661(JP,A)
T.N.Vijaykumar et al., Transient-fault recovery using simultaneous multithreading, Proceedings of the 29th annual international symposium on Computer architecture, IEEE Computer Society, 2002年, p87-98
- (58)調査した分野(Int.Cl., DB名)
G06F 11/16
G06F 11/18