

(19) 日本国特許庁(JP)

(12) 公 開 特 許 公 報(A)

(11) 特許出願公開番号
特開2004-30574
(P2004-30574A)

(43) 公開日 平成16年1月29日(2004.1.29)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
GO6F 12/08	GO6F 12/08 5 2 3 B	5 B 0 0 5
GO6F 15/16	GO6F 12/08 5 0 9 F	5 B 0 4 5
GO6F 15/173	GO6F 12/08 5 1 3	
	GO6F 12/08 5 4 3 B	
	GO6F 15/16 6 4 0 B	
審査請求 未請求 請求項の数 1 O L (全 10 頁) 最終頁に続く		

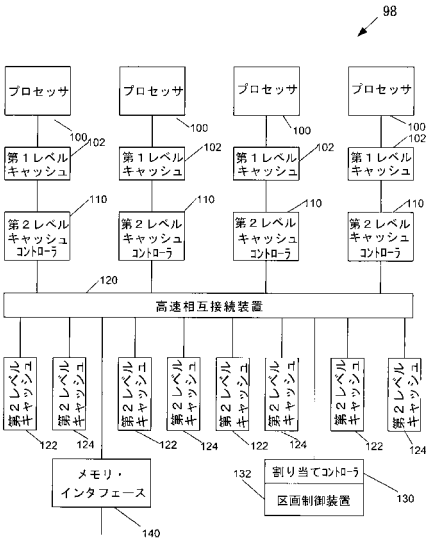
(21) 出願番号 特願2003-58097 (P2003-58097)	(71) 出願人 398038580
(22) 出願日 平成15年3月5日 (2003.3.5)	ヒューレット・パッカード・カンパニー
(31) 優先権主張番号 10/092, 645	HEWLETT-PACKARD COMPANY
(32) 優先日 平成14年3月6日 (2002.3.6)	アメリカ合衆国カリフォルニア州パロアルト
(33) 優先権主張国 米国 (US)	ハノーバー・ストリート 3000
(特許庁注：以下のものは登録商標)	(74) 代理人 100081721
U N I X	弁理士 岡田 次生
	(74) 代理人 100105393
	弁理士 伏見 直哉
	(74) 代理人 100111969
	弁理士 平野 ゆかり
	最終頁に続く

(54) 【発明の名称】 動的にキャッシュ・メモリの割り当てを行うプロセッサ集積回路

(57) 【要約】

【課題】プロセッサおよびキャッシュを動的分割するプロセッサ集積回路を提供すること。

【解決手段】 2以上の命令ストリームを実行するプロセッサ集積回路は、2以上のプロセッサを有する。各プロセッサは、キャッシュコントローラを介して、命令とデータにアクセスする。また、キャッシュメモリの複数ブロックも存在する。キャッシュメモリのいくつかのブロックは、任意に特定のキャッシュコントローラに直接接続される。キャッシュコントローラは、高速相互接続装置を介して、少なくともいくつかのキャッシュメモリの複数ブロックにアクセスする。これらのブロックの少なくとも1つは動的に、2以上のキャッシュコントローラに割り当て可能である。リソース割り当てコントローラは、少なくとも一つの動的割り当て可能キャッシュメモリブロックにアクセス用メモリコントローラを決定する。



【選択図】 図 1

【特許請求の範囲】

【請求項 1】

2 つ以上の命令ストリームを実行することが可能なプロセッサ集積回路であって、
第 1 のキャッシュ・メモリ・コントローラを介して命令を取り出し、データにアクセスするように接続された第 1 のプロセッサと、
第 2 のキャッシュ・メモリ・コントローラを介して命令を取り出し、データにアクセスするように接続された第 2 のプロセッサと、
それぞれがキャッシュ・データ・メモリを含む、複数のキャッシュ・メモリ・ブロックと

、
前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・メモリ・コントローラが、少なくとも 1 つの割り当て可能なキャッシュ・メモリ・ブロックの利用をイネーブルできるように、前記複数のキャッシュ・メモリ・ブロックを前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・メモリ・コントローラに接続する高速相互接続装置と、

前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・メモリ・コントローラからなる群からアクセス用キャッシュ・メモリ・コントローラを選択するように接続したリソース割り当てコントローラを含み、前記アクセス用キャッシュ・メモリ・コントローラが、前記割り当て可能なキャッシュ・メモリ・ブロックにアクセスする、
プロセッサ集積回路。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、マルチプロセッサ回路のような大規模で複雑な集積回路の設計及び実装の分野に関するものである。とりわけ、本発明は、複数 CPU 集積回路を動的に再分割して、クリティカル・パス・スレッドが、必要とするリソースを入手することができ、それによって、システム性能の最適化が可能になるようにするための装置及び方法に関するものである。

【0002】

【従来の技術】

高性能な最新のプロセッサ集積回路は、プロセッサ集積回路上に少なくともいくつかのキャッシュ・メモリを設けることによって製作される。一般に、キャッシュは、制御論理回路と共に、複数のブロックをなすメモリ・セルとして設計される。これらの回路の中には、製品の差別化並びに部分的に欠陥のある回路の販売を可能にする技法であって、キャッシュの一部をディスエーブルできるようにする、ボンディング・オプションを伴う設計を施されたものもあった。これらの回路の中には、キャッシュの欠陥セクションの代わりに利用することが可能なメモリの予備ブロックも備えている。一般に、キャッシュは、制御論理回路と共に、複数ブロックをなすメモリ・セルとして設計される。

【0003】

多くの最新ソフトウェアは、複数プロセッサ・マシンを利用するように書かれている。このソフトウェアは、一般に、複数スレッドを利用するように書かれている。各スレッドは、個別に実行スケジュールを決めることが可能な一連の命令を備えている。一般に、スレッドには、任意の特定の時間において、他のスレッドのアクションが完了するか、または、外部事象が生じるまで、実行を遅延する「待機」モードにあるものもあれば、すぐにでも実行可能になっているものもあり得る。

【0004】

ソフトウェアは、頻繁にスレッドに優先順位を付けることによって、ある特定の時間において、どのスレッドがほとんどのリソースを与えられるべきかを決定することも可能である。例えば、ウィンドウズ 2000（マイクロソフト社の登録商標）、VMS（ヒューレットパッカード社の登録商標）、及び、UNIXオペレーティング・システムは、全て、スレッドの優先順位を維持（アドミニストレータによって設定された基準優先順位から導

10

20

30

40

50

き出される場合が多い)する。これらのオペレーティング・システムは、これらの優先順位を利用して、どのスレッドを実行すべきかを決定したり、別のスレッドによって取って代わられる前に、各スレッドの実行時間量を決定する。

【0005】

複数プロセッサ・マシンの場合、各プロセッサ毎に、実行準備の整ったスレッドの中から異なるスレッドを実行するタスクを課すことが可能である。これらのスレッドは、同じか、または、異なるアプリケーション・プログラムに属する場合もあれば、システム・タスクに関連している場合もある。こうしたマシンは、単一プロセッサを備えたマシンよりも有用な作業を行うことが可能である場合が多い。

【0006】

マルチスレッド式プロセッサは、2つ以上の命令カウンタを備え、一般に、各命令カウンタに関連したレジスタ・セットを備え、2つ以上の命令ストリームを実行することが可能なプロセッサである。例えば、いくつかの命令ストリーム間において、単一パイプライン式実行装置に時分割を施すことが可能なマシンが知られている。実行装置が時分割されるので、各命令ストリームの実行は幾分遅くなりがちである。時分割される単一実行装置を備えたマルチスレッド・マシンは、ソフトウェアにとって、複数の独立したプロセッサのように見える。

【0007】

単一集積回路上に、それぞれ、数サイクル単位の複数命令をディスパッチすることが可能な複数プロセッサを備える、スーパー・スカラ性能のマシンが知られている。このタイプのマシンには、IBM Power-4及びPA8800が含まれる。一般に、これらの集積回路上の各プロセッサは、それ自体の専用セットをなす実行装置パイプライン及びキャッシュを備えている。これらの実行装置の場合、そのダイ面積、それ故のコストは、一般に、時分割マルチスレッド式マシンをはるかに超えることになる。これらのスーパー・スカラ複数プロセッサ回路は、やはり、複数スレッドを実行することが可能であり、高性能マルチスレッド・マシンの一形態とみなすことが可能である。

【0008】

集積回路をなす最新のプロセッサは、キャッシュ・メモリを備えるように製作される場合が多い。キャッシュ・メモリは、メイン・メモリに比べてかなり速いアクセスを可能にする。しかし、それはキャッシュ内で見つかる情報にのみ速いアクセスを可能にする。キャッシュにおいてメモリ参照が見つかる場合、キャッシュにおける「ヒット」と呼ばれ、一方、キャッシュにおいて参照が見つからない場合、キャッシュ「ミス」と呼ばれる。キャッシュ・ヒット数対全メモリ参照回数の比率は、「ヒット率」であり、キャッシュ・サイズ、キャッシュの連想「方式(ways)」数を含むキャッシュ・アーキテクチャ、及び、実行スレッドの性質の関数であることが分っている。

【0009】

既知のように、キャッシュ・ヒット率は、カウンタを用いて、キャッシュ・ヒットと、メモリ参照をカウントすることによって測定可能である。こうしたカウンタの値を読み取って、ヒット率を計算することができる。やはり既知のように、ヒット率が低いと、システム性能をひどく損なう可能性がある。

【0010】

既知のように、スレッドには、他のスレッドより高いヒット率を実現するのに、キャッシュ・サイズの拡大を必要とするものもある。やはり、既知のように、プロセッサの性能は、キャッシュにおけるヒット率が低いと、時折、深刻な悪影響を受ける可能性がある。従って、プロセッサの性能を最大限に発揮すべき場合には、全てのまたは大部分のスレッドについて、高ヒット率を支えるのに十分なキャッシュを設けることが必要になる。しかし、キャッシュ・サイズが大きくなると、コストが高くなる。従って、メーカーは、アプリケーション・プログラム、及び、キャッシュ要件が異なると予想される、それぞれに異なる市場に対してキャッシュ・サイズの異なる同様のプロセッサを備えた集積回路を出している。

10

20

30

40

50

【 0 0 1 1 】

複数プロセッサ集積回路のキャッシュは、一般に、処理コストによってサイズが制限される。大規模集積回路の製作コストは、一般に、その回路面積の指数関数であり、回路によっては、集積回路面積のほとんど半分が、キャッシュ、及び、キャッシュ・メモリ制御回路要素という場合もある。

【 0 0 1 2 】

複数プロセッサ集積回路は、一般に、各プロセッサ毎に所定のキャッシュ量が割り当てられている。従って、これらの回路は、一般に、プロセッサ数に、実行が予測される最もキャッシュ集約的なスレッドにおいて高ヒット率を達成するのに必要とされるキャッシュ量を掛けた値に等しい、全キャッシュ量を必要とする。

10

【 0 0 1 3 】

複数のオペレーティング・システムを同時に実行することが可能な、複数プロセッサ及びマルチスレッドによるマシンが知られている。これらは、分割可能マシンである。一般に、各オペレーティング・システムは、ある区画で実行されるが、この場合、1つの区画には、1つ以上のプロセッサ、メイン・メモリの適合するセクション、及び、他のシステムリソースが割り当てられている。各区画は、一般に、専用ディスクスペースを備えることができるか、または、ディスクスペースを他の区画と共用することが可能な、仮想マシンとして構成される。1つの区画ではウィンドウズNT（マイクロソフト社の商標）を実行し、一方、もう1つの区画ではUNIXを実行することが可能なマシンが存在する。各コピーを独立した区画で個別に実行するようにして、同じオペレーティング・システムのいくつかのコピーを同時に実行することが可能なマシンも存在する。これらのマシンは、各区画を特定のユーザ及びアプリケーション専用とすることが可能であり、また、ある区画で生じる問題（システム・クラッシュを含む）が、他の区画における操作に悪影響を及ぼすとは限らないという点において有利である。

20

【 0 0 1 4 】

既知のように、複数プロセッサ及びマルチスレッドによるマシンにおける実行時間は、各区画に割り当てられたプロセッサ数、メモリ量、及び、ディスクスペース量に基づいて課金することが可能である。やはり既知のように、分割可能マシンにおいて、1つ以上の複数プロセッサまたはマルチスレッドによる集積回路をプロセッサとして利用することも可能である。

30

【 0 0 1 5 】

【 発明が解決しようとする課題 】

本発明の目的は、分割可能マシンの一部をなす複数プロセッサ集積回路を含む、複数プロセッサ集積回路のプロセッサに動的にキャッシュを割り当てることにより、システムにおいて実行する各スレッドまたは区画毎に適切なキャッシュ量が与えられるようにすることにある。

【 0 0 1 6 】

【 課題を解決するための手段 】

高性能複数プロセッサ集積回路は、複数キャッシュ装置と、複数命令取り出し及び解読装置を備えており、各命令取り出し及び解読装置は、実プロセッサまたは仮想プロセッサに対応している。集積回路は、少なくとも1つの動的に割り当て可能なキャッシュ装置を備えており、特定のプロセッサに直接接続することが可能な追加キャッシュ装置を備えることも可能である。

40

【 0 0 1 7 】

集積回路は、2つ以上の実プロセッサまたは仮想プロセッサによる動的割り当て可能なキャッシュ装置の利用を可能にする高速相互接続装置、及び、キャッシュ割り当てレジスタも備えている。キャッシュ割り当てレジスタのフィールドによって、動的割り当て可能キャッシュ装置に対するアクセスを許される実プロセッサまたは仮想プロセッサが決定される。

【 0 0 1 8 】

50

実施態様の１つでは、動的割り当て可能キャッシュ装置は、システムにおける第２レベルのキャッシュの一部を形成している。この実施態様の場合、４つのプロセッサと、それぞれ、１つのプロセッサに対応する４つのサイズが固定された第１レベルのキャッシュと、４つの割り当てが固定された第２レベルのキャッシュ・ブロックと、４つの動的割り当て可能な第２レベルのキャッシュ・ブロックが設けられている。

【００１９】

第２レベルの・キャッシュ・システムは、各プロセッサに関連したヒット率モニタ装置が設けられている。オペレーティング・システム・ドライバは、各プロセッサに関連したヒット率をモニタし、ヒット率を追跡する。モニタされるヒット率は、追加割り当てによって恩恵を受けることができるスレッド区画の判定に役立つ。

10

【００２０】

ある特定の実施態様の場合、動的割り当て可能な第２レベルのキャッシュ・ブロックが、分割可能マシンのある特定の区画のプロセッサに割り当てられる。このマシンにおけるマシン時間は、各区画毎に動的割り当て可能キャッシュがどれだけ割り当てられるかに基づいて課金される。この実施態様の場合、動的割り当てキャッシュは、ブート時、または、区画の再構成時に割り当てられる；区画は、スケジュール並びにブート時間に従って再構成可能である。

【００２１】

きめ細かい代替実施態様の場合、各スレッドは、要求されるキャッシュ割り当てに関連している。この実施態様の場合、動的に割り当てられたキャッシュ・ブロックが、コンテキスト・スイッチ時に、その性能が最適化されるように再割り当てされる。この実施態様の場合、オペレーティング・システムは、前に割り当てられたキャッシュ割り当てで得られた、各スレッドに関連したヒット率を追跡し、各スレッド毎に、最適キャッシュ・サイズに基づいて要求されるキャッシュ割り当てを設定する。

20

【００２２】

動的に割り当てられたキャッシュが、１つの区画すなわちスレッドから割り当て解除されて、もう１つの区画に割り当てられる場合、まず、全ての「汚れた(dirty)」キャッシュ・ラインをメモリに書き込み、各キャッシュ・ライン毎に、「有効」ビットをクリアすることによって消去される。こうして、ある区画に関連したデータが別の区画で利用できないことを保証するため、キャッシュが有効に空にされる。

30

【００２３】

【発明の実施の形態】

分割可能複数プロセッサ集積回路９８（図１）は、いくつかのプロセッサ１００を備えている。各プロセッサは、第１レベル・キャッシュ１０２を介して、命令データについてメモリ参照を実施する。第１レベル・キャッシュでミスした参照は、第２レベル・キャッシュ・コントローラ１１０に振り当てられる。

【００２４】

各第２レベル・キャッシュ・コントローラ１１０は、高速相互接続装置１２０に接続されている。高速相互接続装置１２０によって、第２レベル・キャッシュ・コントローラ１１０は、それぞれ、複数のキャッシュ・メモリ・ブロック１２２及び１２４の１つ以上にアクセスできるようになる。これらのメモリ・ブロックのうちの少なくとも１つが、任意のキャッシュ・コントローラ１１０に割り当てることができる、割り当て可能キャッシュ・メモリ・ブロック１２４である。必要というわけではないが、割り当てが固定された１つ以上のキャッシュ・メモリ・ブロック１２２を設けることが可能である。割り当てコントローラ１３０は、もしあれば、動的に割り当てられるキャッシュ・メモリ・ブロック１２４のうちのどれが、各キャッシュ・コントローラ１１０によるアクセスを受けるかを決定する。区画制御装置１３２は、各システム区画に対応付けられるプロセッサを決定する働きをする。区画制御装置１３２及び割り当てコントローラ１３０は、従って、各割り当て可能キャッシュ・メモリ・ブロック１２２及び１２４が対応付けられる、第２レベル・キャッシュ・コントローラ１１０、プロセッサ１００、及び、システム区画を決定する。

40

50

【 0 0 2 5 】

第 1 レベル・キャッシュ 1 0 2 内において見つからず、第 2 レベル・キャッシュ内でも見つからないキャッシュ参照は、メモリ・インタフェース 1 4 0 を介してオフ・チップ・メイン・メモリ（不図示）に振り向けられる。

【 0 0 2 6 】

集積回路 1 9 6 の代替実施態様では、各動的割り当て可能キャッシュ・メモリ・ブロック 1 9 8（図 2）毎に、データ・メモリ 2 0 0、タグ・メモリ 2 0 2、及び、相互接続装置 2 0 6 に対するインタフェース 2 0 4 が組み込まれている。各キャッシュ・コントローラ 2 1 0 は、下位レベルのキャッシュでミスしたメモリ参照を受け取るためのインタフェース 2 1 2 を備えている。各キャッシュ・コントローラ 2 1 0 は、キャッシュ書き込みを追跡し、上位レベルのメモリ・インタフェース 2 1 5 を介して、メイン・メモリまたはより高レベルのキャッシュに対するライトバックを制御するためのライトバック待ち行列 2 1 4 を含む、当該技術分野において既知のキャッシュ制御論理回路、並びに、キャッシュ内で捜し出す参照を決定するためのタグ論理回路 2 1 8 及びヒット論理回路 2 1 6 も備えている。各キャッシュ・コントローラには、ヒット率を含むキャッシュ性能を測定するためのヒット率モニタ論理回路 2 1 9 も組み込まれている。キャッシュ・コントローラ 2 1 0 は、局部タグ・メモリ 2 2 0 及びデータ・メモリ 2 2 2 を備えている。タグ・メモリ 2 2 0 及びデータ・メモリ 2 2 2 は、十分に大きいので、小容量キャッシュを要求する命令ストリームについて、高いヒット率をもたらすことが可能である。

【 0 0 2 7 】

キャッシュ・コントローラ 2 1 0 は、割り当て制御装置 2 3 0 の制御下において、動的に割り当てられるキャッシュ・メモリ・ブロック 1 9 8 と連絡がとれるようにする、相互接続装置 2 0 6 に接続されている。割り当て制御装置 2 3 0 及び相互接続装置 2 0 6 は、ソフトウェアによって、ゼロまたは 1 つ以上の動的割り当てキャッシュ・メモリ・ブロック 1 9 8 を各キャッシュ・コントローラ 2 1 0 に接続するように構成されている。どの所定の時間においても、各動的割り当てキャッシュ・メモリ・ブロック 1 9 8 へのアクセスは、ただ 1 つのキャッシュ・コントローラ 2 1 0 だけしか実施することができない。

【 0 0 2 8 】

ある代替実施態様の場合、各キャッシュ・コントローラ 1 1 0 には、4 つものキャッシュ・メモリ・ブロック 1 2 2 及び 1 2 4 を制御するのに十分なタグが与えられている。この実施態様の場合、動的割り当てキャッシュ・メモリ・ブロック 1 2 4 には、キャッシュ・データ・メモリが含まれている。

【 0 0 2 9 】

1 つ以上のマルチプロセッサ集積回路 9 8（図 1）を中心として構築されたコンピュータ・システムを管理するための方法 3 0 0（図 3）には、キャッシュの性能をモニタするステップ 3 0 2 が含まれている（図 3）。モニタ・ステップでは、図 2 に示すヒット率モニタ 2 1 9、及び、集積回路に組み込むことができるような他のキャッシュ性能モニタ装置を利用して、区画、アプリケーション、及び、従来のキャッシュ割り当てに関する過去のキャッシュ性能が記録される。さまざまな条件下におけるこの過去の性能の記録は、1 つ以上のマルチプロセッサ集積回路 9 8 におけるキャッシュの可用性を考慮して、所望のシステム分割の決定 3 0 4 に利用される。ある特定の実施態様では、過去のプロセッサ間通信量も、各システム区画に配置すべきプロセッサの決定において考慮される。決定された分割は、区画割り当てテーブルをなすように構成される（3 0 6）。

【 0 0 3 0 】

システムは、各ブート時毎に再分割されるが、再分割がシステム性能全体の向上に適したものであれば、他の時点において、動的に再分割することも可能である。再分割には、影響を受ける各区画において実行されるオペレーティング・システムを停止する（3 0 8）ことが必要になる。各集積回路 9 8 のプロセッサ 1 0 0 は、区画割り当てテーブルに基づいて、区画に割り当てられる（3 1 0）。次に、動的割り当て可能キャッシュ・ブロック 1 2 4 が、各割り当てテーブルに基づいて、各区画のプロセッサ 1 0 0 に割り当てられる

10

20

30

40

50

(3 1 2)。次に、各区画のオペレーティング・システムが、ブート、すなわち、再起動され(3 1 4)、マシン時間、システム分割、及び、キャッシュ割り当ての課金レコードが、保持される(3 1 6)。これらの課金レコードによって、プロセッサ数、及び、そのアプリケーションに割り当てられたキャッシュ量に基づいて、顧客への請求が可能になる。この発明は、例として次のような実施形態を含む。

【 0 0 3 1 】

(1) 2 つ以上の命令ストリームを実行することが可能なプロセッサ集積回路 9 8 であって、

第 1 のキャッシュ・メモリ・コントローラを 1 1 0 介して命令を取り出し、データにアクセスするように接続された第 1 のプロセッサ 1 0 0 と、

10

第 2 のキャッシュ・メモリ・コントローラを 1 1 0 介して命令を取り出し、データにアクセスするように接続された第 2 のプロセッサ 1 0 0 と、

それぞれがキャッシュ・データ・メモリを含む、複数のキャッシュ・メモリ・ブロック 1 2 2、1 2 4 と、

前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・コントローラ 1 1 0 が、少なくとも 1 つの割り当て可能なキャッシュ・メモリ・ブロックの利用をイネーブルできるように、前記複数のキャッシュ・メモリ・ブロックを前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・コントローラ 1 1 0 に接続する高速相互接続装置 1 2 0 と、

前記第 1 のキャッシュ・メモリ・コントローラおよび第 2 のキャッシュ・メモリ・コントローラ 1 1 0 からなる群からアクセス用キャッシュ・メモリ・コントローラ 1 1 0 を選択するように接続したリソース割り当てコントローラ 1 3 0 を含み、前記アクセス用キャッシュ・メモリ・コントローラ 1 1 0 が、前記割り当て可能なキャッシュ・メモリ・ブロック 1 2 4 にアクセスする、

20

プロセッサ集積回路。

【 0 0 3 2 】

(2) さらに、複数の第 1 レベルのキャッシュ・システム 1 0 2 を含み、前記第 1 のプロセッサ 1 0 0 が、第 1 の第 1 レベル・キャッシュ・システム 1 0 2 を介して、前記第 1 のキャッシュ・メモリ・コントローラ 1 1 0 から命令を取り出し、データにアクセスし、前記第 2 のプロセッサ 1 0 0 が、第 2 の第 1 レベル・キャッシュ・システム 1 0 2 を介して、前記第 2 のキャッシュ・メモリ・コントローラ 1 1 0 から命令を取り出し、データにアクセスする、(1) に記載のプロセッサ集積回路。

30

【 0 0 3 3 】

(3) 前記キャッシュ・メモリ・ブロック 1 9 8 が、さらに、キャッシュ・タグ・メモリ 2 0 2 を有する、(1) に記載のプロセッサ集積回路。

【 0 0 3 4 】

(4) 各キャッシュ・メモリ・コントローラ 2 1 0 が、キャッシュ・ヒット率モニタ装置 2 1 9 を有する、(1) に記載のプロセッサ集積回路。

【 0 0 3 5 】

(5) 分割可能な複数プロセッサ・システムに用いられる複数プロセッサ集積回路 9 8 であって、

40

それぞれ、関連する第 1 レベルのキャッシュ 1 0 2 から命令を受信するように接続されている、複数のプロセッサ 1 0 0 と、

複数の割り当て可能上位レベル・キャッシュ・メモリ・ブロック 1 2 4 と、

各第 1 レベルのキャッシュ 1 0 2 におけるキャッシュ・ミスを、割り当てられた上位レベル・キャッシュ・メモリ・ブロック 1 2 4 に伝える相互接続装置 1 2 0 と、

上位レベル・キャッシュ・メモリ・ブロック 1 2 4 をプロセッサ 1 0 0 に割り当てるための割り当て装置 1 3 0 を含む、

複数プロセッサ集積回路 9 8 において、キャッシュを動的に割り当てる方法であって、

プロセッサ及び分割に関連した過去のキャッシュ性能をモニタするステップ 3 0 2 と、

50

分割するのが望ましいプロセッサ 100、及び、プロセッサ 100 に対する上位レベル・キャッシュ・ブロック 124 の割り当てを決定するステップ 304 と、前記システム 310 を再分割するステップを含み、前記システム 310 を再分割するステップは、前記複数プロセッサ集積回路 98 の少なくとも 1 つにおけるプロセッサ 100 への上位レベル・キャッシュ・ブロック 124 の割り当てを含む、方法。

【0036】

(6) 前記上位レベル・キャッシュ・ブロック 124 が、第 2 レベルのキャッシュ・ブロックである、(5) に記載の方法。

【0037】

(7) さらに、プロセッサ時間及び割り当てられたキャッシュに基づいて、顧客に請求するステップを含む、(5) に記載の方法。

【0038】

(8) 前記複数プロセッサ集積回路 98 が、さらに、複数の割り当て不能キャッシュ・メモリ・ブロック 222 を含む、(5) に記載の方法。

【0039】

(9) 各上位レベル割り当て可能キャッシュ・ブロック 124 が、さらに、タグ・メモリ 202 と、キャッシュ・データ・メモリ 200 を含む、(5) に記載の方法。

【0040】

(10) 前記相互接続装置 120 が、さらに、複数の上位レベル・キャッシュ・コントローラ 110 を含み、各上位レベル・キャッシュ・コントローラ 110 が、ライトバック・キャッシュとして、前記割り当て可能キャッシュ・メモリ・ブロック 124 の動作を制御可能である、(5) に記載の方法。

【0041】

本発明は、とりわけ、その特定の実施態様に関連して図示され、解説されてきたが、当業者には明らかなように、本発明の精神及び範囲を逸脱することなく、形態及び細部に関して他のさまざまな変更を加えることも可能である。もちろん、本明細書に開示され、付属の請求項に包含されている本発明のより一般的な概念から逸脱することなく、本発明をさまざまな実施態様に適応させる上において、さまざまな変更を加えることが可能である。

【図面の簡単な説明】

【図 1】 動的割り当て可能キャッシュの複数ブロックを備えた複数プロセッサ集積回路のブロック図である。

【図 2】 複数プロセッサ集積回路の用いられる動的割り当てキャッシュの複数ブロックを備えたキャッシュ・システムのブロック図である。

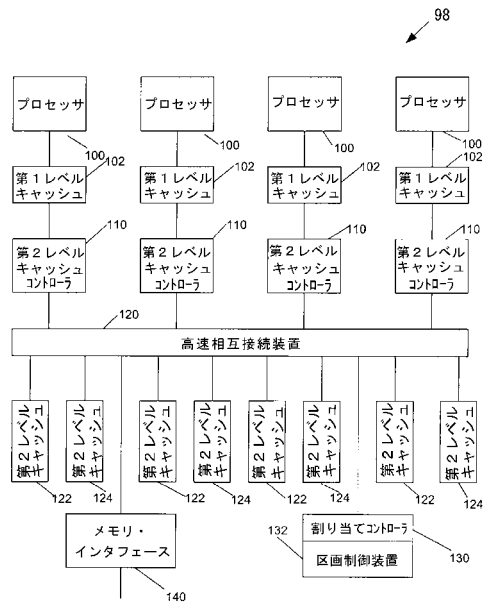
【図 3】 集積回路を利用するシステムにおけるキャッシュ・メモリの割り当て、及び、プロセッサ時間の課金を例示したフローチャートである。

10

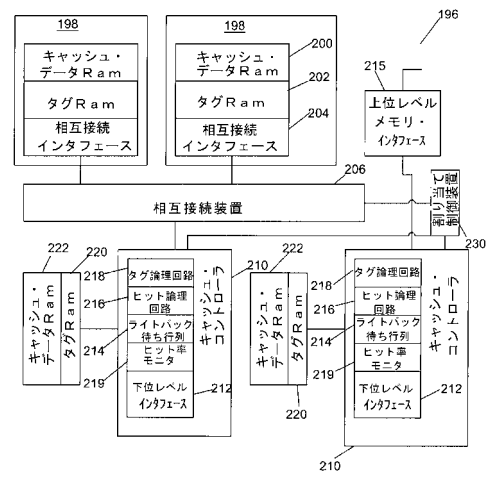
20

30

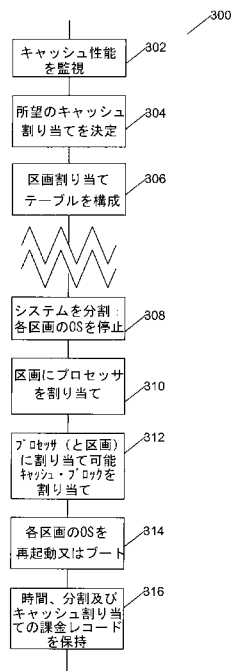
【図 1】



【図 2】



【図 3】



フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード(参考)
	G 0 6 F 15/16 6 4 5	
	G 0 6 F 15/173 Z	

(72)発明者 ドナルド・シー・ソルティス、ジュニア

アメリカ合衆国 8 0 5 2 6 コロラド州フォート・コリンズ、ローズゲート・コート 4 4 1 4

F ターム(参考) 5B005 JJ13 MM05 UU32 UU41

5B045 DD12 KK08