

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2008-502069

(P2008-502069A)

(43) 公表日 平成20年1月24日(2008.1.24)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 12/08 (2006.01)	G06F 12/08 531Z	5B005
	G06F 12/08 525Z	
	G06F 12/08 507E	
	G06F 12/08 511B	

審査請求 未請求 予備審査請求 未請求 (全 25 頁)

(21) 出願番号 特願2007-526629 (P2007-526629)
 (86) (22) 出願日 平成17年5月31日 (2005.5.31)
 (85) 翻訳文提出日 平成19年1月10日 (2007.1.10)
 (86) 国際出願番号 PCT/IB2005/051774
 (87) 国際公開番号 W02005/121966
 (87) 国際公開日 平成17年12月22日 (2005.12.22)
 (31) 優先権主張番号 04013507.1
 (32) 優先日 平成16年6月8日 (2004.6.8)
 (33) 優先権主張国 欧州特許庁 (EP)

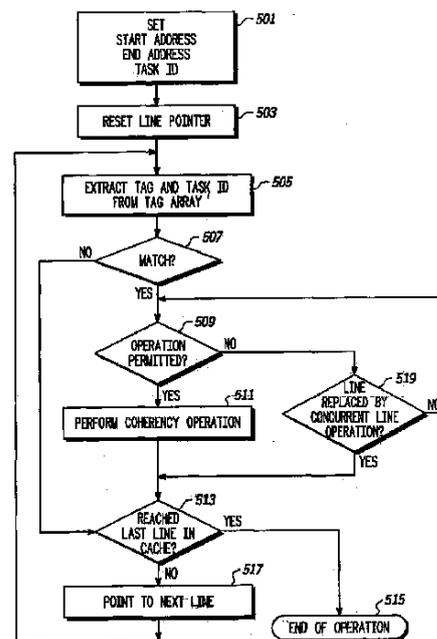
(71) 出願人 504199127
 フリースケール セミコンダクター イン
 コーポレイテッド
 アメリカ合衆国 78735 テキサス州
 オースティン ウィリアム キャノン
 ドライブ ウェスト 6501
 (74) 代理人 100089705
 弁理士 社本 一夫
 (74) 代理人 100140109
 弁理士 小野 新次郎
 (74) 代理人 100075270
 弁理士 小林 泰
 (74) 代理人 100080137
 弁理士 千葉 昭男

最終頁に続く

(54) 【発明の名称】 メモリ・キャッシュ制御装置及びそのためのコヒーレンシ動作を実行する方法

(57) 【要約】

メモリ・キャッシュ(105)についてコヒーレンシ動作を実行するメモリ・キャッシュ制御装置は、メイン・メモリ(103)と関連した複数のアドレスを備えるアドレス・グループに関するアドレス・グループ指示を受け取る受信プロセッサ(301)を備える。アドレス・グループ指示が、タスク識別、及びメイン・メモリ(103)のメモリ・ブロックに対応するアドレス範囲を指示する。制御ユニット(303)が、1群のキャッシュ・ラインの各キャッシュ・ラインを順次に処理する。詳細には、各キャッシュ・ラインがアドレス・グループのアドレスと関連するかどうか、マッチ判定基準を評価することにより決定される。マッチ判定基準に適合する場合、コヒーレンシ動作が、キャッシュ・ラインについて実行される。競合が前記コヒーレンシ動作と別のメモリ動作との間に存在する場合、コヒーレンシ手段が、コヒーレンシ動作を禁止する。本発明は、キャッシュのコヒーレンシ動作の持続時間の低減を可能にする。その持続時間は更に、コヒーレンシ動作によりカバーされるメイン・メモリ・アドレス空間のサイズとは無関係であ



【特許請求の範囲】**【請求項 1】**

メモリ・キャッシュ(105)についてコヒーレンシ動作を実行するメモリ・キャッシュ制御装置であって、メイン・メモリ(103)と関連した複数のアドレスを備えるアドレス・グループに関するアドレス・グループ指示を受け取る手段(301)を備えるメモリ・キャッシュ制御装置において、

1群のキャッシュ・ラインの各キャッシュ・ラインを順次に処理する処理手段(303)を備え、

当該処理手段(303)が、

キャッシュ・ラインがアドレス・グループのアドレスと関連するかどうかをマッチ判定基準を評価することにより決定する手段と、 10

前記マッチ判定基準に適合する場合前記キャッシュ・ラインについてコヒーレンシ動作を実行するコヒーレンシ手段と、

競合が前記コヒーレンシ動作と別のメモリ動作との間に存在するかどうかを決定する手段と、を備え、

前記コヒーレンシ手段が、競合が存在する場合前記コヒーレンシ動作を禁止するよう動作可能である

ことを特徴とするメモリ・キャッシュ制御装置。

【請求項 2】

前記競合が、前記コヒーレンシ動作と前記別のメモリ動作との間で共用される資源に関連する請求項1記載のメモリ・キャッシュ制御装置。 20

【請求項 3】

競合が存在するかどうかを決定する前記手段が、前記コヒーレンシ動作と前記別のメモリ動作とが同じキャッシュ資源に対する実質的に同時のアクセスをもたらす場合競合が存在すると決定するよう動作可能である請求項1又は2記載のメモリ・キャッシュ制御装置。

【請求項 4】

前記コヒーレンシ手段が、前記コヒーレンシ動作と前記別のメモリ動作とのうちの1つを遅延させることにより前記コヒーレンシ動作を禁止するよう動作可能である請求項1から3のいずれか一項に記載のメモリ・キャッシュ制御装置。 30

【請求項 5】

前記マッチ判定基準が、前記キャッシュ・ラインと関連したメイン・メモリ・アドレスが前記アドレス・グループに属するかどうかの評価を備える請求項1から4のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 6】

前記アドレス・グループ指示が、前記メイン・メモリのメモリ・ブロックの開始アドレス及び終了アドレスを備え、

前記マッチ判定基準が、前記メイン・メモリ・アドレスが前記メモリ・ブロックに属するかどうかを決定することを備える

請求項1から5のいずれか一項に記載のメモリ・キャッシュ制御装置。 40

【請求項 7】

前記開始アドレス及び終了アドレスが、仮想メモリ・アドレスである請求項1から6のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 8】

前記のマッチ手段が、キャッシュ・ライン・タグ及びキャッシュ・ライン・インデックスに応答して前記メイン・メモリ・アドレスを決定するよう動作可能である請求項5から7のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 9】

前記メモリ・キャッシュ(105)が、セットアソシエイティブ・メモリ・キャッシュであり、

前記 1 群のキャッシュ・ラインが、異なる組の前記セットアソシエイティブ・メモリのキャッシュ・ラインを備える

請求項 1 から 8 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 10】

前記処理手段 (303) が、複数の組の前記セットアソシエイティブ・メモリ・キャッシュを順次に処理するよう動作可能である請求項 9 記載のメモリ・キャッシュ制御装置。

【請求項 11】

前記アドレス・グループ指示が、少なくとも 1 つのタスク識別の指示を備え、

前記マッチ判定基準が、第 1 のキャッシュ・ラインと関連したタスク識別が前記少なくとも 1 つのタスク識別と一致するかどうかの評価を備える

請求項 1 から 10 のいずれか一項に記載のメモリ・キャッシュ制御装置。

10

【請求項 12】

前記アドレス・グループ指示が、タスク識別から成る請求項 11 記載のメモリ・キャッシュ制御装置。

【請求項 13】

前記 1 群のキャッシュ・ラインが、前記メモリ・キャッシュの全てのキャッシュ・ラインを備える請求項 1 から 12 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 14】

前記コヒーレンシ動作が、無効化動作である請求項 1 から 13 のいずれか一項に記載のメモリ・キャッシュ制御装置。

20

【請求項 15】

前記コヒーレンシ動作が、同期化動作である請求項 1 から 14 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 16】

前記コヒーレンシ動作が、フラッシュ動作である請求項 1 から 15 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 17】

前記処理手段 (303) が、前記 1 群のキャッシュ・ラインの全てのキャッシュ・ラインが処理されてしまったことを決定することに対応して終了指示を設定する手段を備える請求項 1 から 16 のいずれか一項に記載のメモリ・キャッシュ制御装置。

30

【請求項 18】

前記終了指示が、割り込み指示である請求項 17 記載のメモリ・キャッシュ制御装置。

【請求項 19】

前記メモリ・キャッシュが、命令キャッシュである請求項 1 から 18 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 20】

前記メモリ・キャッシュが、データ・キャッシュである請求項 1 から 19 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 21】

請求項 1 から 20 のいずれか一項に記載のメモリ・キャッシュ制御装置を備えるメモリ・キャッシュ・システム。

40

【請求項 22】

プロセッサと、

メイン・メモリと、

前記プロセッサ及び前記メイン・メモリに結合されたキャッシュ・メモリと、

請求項 1 から 20 のいずれか一項に記載のメモリ・キャッシュ制御装置とを備える処理システム。

【請求項 23】

メモリ・キャッシュについてコヒーレンシ動作を実行する方法であって、メイン・メモリ (103) と関連した複数のアドレスを備えるアドレス・グループに関するアドレス・

50

グループ指示を受け取るステップを備える、前記コヒーレンシ動作を実行する方法において、

1群のキャッシュ・ラインの各ラインを順次に処理するステップを備え、

当該処理するステップが、前記1群のキャッシュ・ラインの各キャッシュ・ラインに関して、

第1のキャッシュ・ラインが前記アドレス・グループのアドレスと関連するかどうかをマッチ判定基準を評価することにより決定するステップと、

前記マッチ判定基準に適合する場合前記第1のキャッシュ・ラインについてコヒーレンシ動作を実行するステップと、

競合が前記コヒーレンシ動作と別のメモリ動作との間に存在するかどうかを決定するステップと、を備え、

前記のコヒーレンシ手段が、競合が存在する場合前記コヒーレンシ動作を禁止するよう動作可能である

ことを特徴とするメモリ・キャッシュについてコヒーレンシ動作を実行する方法。

【請求項24】

前記競合が、前記コヒーレンシ動作と前記別のメモリ動作との間で共用される資源に関連する請求項23記載のメモリ・キャッシュについてコヒーレンシ動作を実行する方法。

【請求項25】

競合が存在するかどうかを決定する前記ステップが、前記コヒーレンシ動作と前記別のメモリ動作とが同じキャッシュ資源に対する実質的に同時のアクセスをもたらす場合競合が存在すると決定するステップを備えることを更に特徴とする請求項23又は24記載のメモリ・キャッシュについてコヒーレンシ動作を実行する方法。

【請求項26】

前記コヒーレンシ動作と前記別のメモリ動作とのうちの1つを遅延させることにより、前記のコヒーレンシ手段による前記コヒーレンシ動作を禁止するステップを更に備える請求項23から25のいずれか一項に記載のメモリ・キャッシュについてコヒーレンシ動作を実行する方法。

【請求項27】

請求項23記載の方法を実行するようプロセッサを制御するプロセッサ実行可能命令を格納する記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、メモリ・キャッシュ制御装置及びそのためのコヒーレンシ(coherency)動作を実行する方法に関する。

【背景技術】

【0002】

デジタル・データ処理システムは、例えば、データ処理システム、家庭電気機器、コンピュータ、自動車等を含む多くの応用において用いられている。例えば、パーソナル・コンピュータ(PC)は、複雑なデジタル処理機能を用いて、広範囲の多種多様なユーザ応用のためのプラットフォームを提供する。

【0003】

デジタル・データ処理システムは、典型的には、入力/出力機能、命令及びデータ・メモリ、及び、例えば、マイクロコントローラ、マイクロプロセッサ又はデジタル信号プロセッサのような1又はそれより多くのデータ・プロセッサを備える。

【0004】

処理システムの性能の重要なパラメータは、メモリ性能である。最適な性能に関しては、メモリが大きく、早く、そして好ましくは安価であることが望ましい。不都合にも、これらの特性は、競合する要件となりがちであり、適切なトレードオフが、デジタル・システムを設計するとき要求される。

【0005】

処理システムのメモリ性能を改善するため、異なるタイプのメモリの個々の利点を活用しようとする複雑なメモリ構造が、開発されてきた。特に、高速のキャッシュ・メモリをより大きく、比較的遅く、より安価なメイン・メモリと関連して用いることが一般的になってきた。

【0006】

例えば、PCにおいては、メモリは、典型的には異なるサイズと速度のメモリを備える記憶階層で組織化される。従って、PCは、典型的には、大きく低コストであるが遅いメイン・メモリを備え、その上、比較的小さく高価であるが高速のメモリを備える1又はそれより多くのキャッシュ・メモリ・レベルを有する。動作中に、メイン・メモリからのデータは、早い読み出しサイクルを可能にするため、キャッシュ・メモリの中に動的にコピーされる。同様に、データは、メイン・メモリではなくキャッシュ・メモリに書き込まれ、それにより早い書き込みサイクルを可能にする。

10

【0007】

従って、キャッシュ・メモリは、メイン・メモリの異なる記憶場所と動的に関連し、そしてメイン・メモリとキャッシュ・メモリの間のインターフェース及び相互作用が許容可能な性能にとって重要であることが明らかである。従って、キャッシュ動作に対する著しい研究が行われてきて、そしてデータがメイン・メモリではなくてキャッシュ・メモリへ書き込まれ又はそれから読み出されるとき並びにデータがキャッシュ・メモリとメイン・メモリとの間で転送されるときを制御する様々な方法及びアルゴリズムが、開発されてきた。

20

【0008】

典型的には、プロセッサが読み出し動作を実行するときにはいつでも、キャッシュ・メモリ・システムは、最初に、対応するメイン・メモリ・アドレスがキャッシュと現在関連しているかをチェックする。キャッシュ・メモリがメイン・メモリ・アドレスに関して有効なデータ値を含む場合、このデータ値は、システムのデータ・バス上にキャッシュにより入力され、そして読み出しサイクルは、いずれの待ちサイクル無しに実行する。しかしながら、キャッシュ・メモリがメイン・メモリ・アドレスに関して有効なデータ値を含まない場合、メイン・メモリ読み出しサイクルが、実行され、そしてデータは、メイン・メモリから検索される。典型的には、メイン・メモリ読み出しサイクルは、1又はそれより多くの待ち状態を含むことによりプロセスの速度を遅くする。

30

【0009】

プロセッサがデータをキャッシュ・メモリから受け取ることができるメモリ動作は、通常、キャッシュ・ヒットと呼ばれ、そしてプロセッサがデータをキャッシュ・メモリから受け取ることができないメモリ動作は、通常、キャッシュ・ミスと呼ばれる。典型的には、キャッシュ・ミスは、プロセッサがデータをメイン・メモリから検索することをもたらすばかりでなく、メイン・メモリとキャッシュとの間の多数のデータ転送をもたらす。例えば、所与のアドレスがアクセスされてキャッシュ・ミスをもたらす場合、後続の記憶場所は、キャッシュ・メモリに転送される。プロセッサが連続した記憶場所に頻繁にアクセスするので、所望のデータを備えるキャッシュ・メモリの確率は、それにより通常増大する。

40

【0010】

キャッシュ・メモリ・システムは、典型的には、キャッシュ・メモリの分解能に対応するキャッシュ・ラインに分割される。セットアソシエイティブ・キャッシュ・システムとして知られるキャッシュ・システムにおいては、多数のキャッシュ・ラインが、各組がメイン・メモリ・アドレスの下位データ・ビットに対する固定のマッピングに対応する異なる組と一緒にグループ化される。組を形成する各キャッシュ・ラインの極端なケースは、直接マップ方式キャッシュとして知られ、そして各メイン・メモリ・アドレスが1つの特定のキャッシュ・ラインにマッピングされることをもたらす。全てのキャッシュ・ラインが単一の組に属する他の極端なケースは、フルアソシエイティブ方式キャッシュとして知

50

られ、そしてこれは、各キャッシュ・ラインをいずれのメイン・メモリの記憶場所にマッピングするのを可能にする。

【0011】

各キャッシュ・ラインがどのメイン・メモリ・アドレス（ある場合）と関連しているかを追跡するため、キャッシュ・メモリ・システムは、典型的には、各キャッシュ・ラインのため、そのラインとメイン・メモリとの間の現在のマッピングを指示するデータを保持するデータ・アレイを備える。特に、データ・アレイは、典型的には、関連のメイン・メモリ・アドレスの上位データ・ビットを備える。この情報は、通常、タグとして知られ、そしてデータ・アレイは、タグ・アレイとして知られる。

【0012】

キャッシュ・メモリの制御は、非常に重要であることは明らかであり、特に、メイン・メモリとキャッシュ・メモリとの間の対応を管理することが不可欠であることが明らかである。例えば、データが、キャッシュ・メモリの対応のデータが更新されない又は無効なデータとして指示されずに、メイン・メモリで変更される場合、壊滅的な結果が生じる。同様に、キャッシュ・メモリに書き込まれたデータが、そのデータがキャッシュでオーバーライトされる前に又はメイン・メモリの対応の記憶場所が直接アクセスされる前に、メイン・メモリへ転送されない場合、データの不一致が、エラーをもたらす。従って、処理システムの信頼性は、キャッシュの制御に非常に依存している。従って、コヒーレンシ動作は、キャッシュ・メモリとメイン・メモリとの不一致が望ましくない効果をもたらさない確率を排除又は低減するために適切な瞬間に実行される。

10

20

【0013】

例えば、直接メモリ・アクセス（DMA）モジュールは、メイン・メモリに直接アクセスすることができる。DMAは、例えば、ハード・ディスク・インターフェースの一部であり、そしてデータをメイン・メモリからハード・ディスクへハード・ディスク書き込み動作中に転送するため用いられる。DMA動作が実行されることができるようになる前に、キャッシュ・メモリに書き込まれた全てのデータがメイン・メモリに転送されてしまうことが重要である。従って、ハード・ディスク書き込み動作の前に、プロセッサ・システムは、コヒーレンシ動作を実行することが好ましく、ここでは、メイン・メモリではなくてキャッシュ・メモリに書き込まれてしまった全てのデータがメイン・メモリに転送される。コヒーレンシ動作は、システムを正常動作のため解放するため、またシステムの計算負荷を低減するため、出来るだけ少ない複雑さ及び時間消費でおそらく実行される。

30

【0014】

しかしながら、一般的に、そのようなコヒーレンシ動作は、複雑で、時間を消費し、電力を消費し、及び/又は複雑なハードウェアを必要とすることによりコストが増大する。例えば、メイン・メモリの所与のアドレス・ブロックがハード・ディスクに転送されることになる場合、従来のアプローチは、メイン・メモリの各記憶場所を通るようステップを進めること、及びキャッシュがこの記憶場所に関する更新された値を備えるかどうかをチェックすることを備える。メイン・メモリ・アドレス・ブロックが非常に大きくなるにつれ、これは、非常に面倒なプロセスであり、それは、典型的には、ソフトウェアの実行のため非常に時間を消費し、そしてハードウェアの実現のため非常に複雑な要件を有する。

40

【0015】

一般的に、ハードウェア及びソフトウェアのコヒーレンシ機構であるコヒーレンシ機能性を実現するには2つのアプローチがある。ハードウェアのアプローチは、各キャッシュ・ベースのシステムに対してスヌーピング機構を追加することを含む。スヌーピング機構は、他のマスタ（例えば、DMAプロセッサのようなもの）により行われるメイン・メモリに対する全てのアクセスを追跡する。スヌーピング機構がキャッシュの中の有効なデータへのアクセスを検出したとき、スヌーピング機構は、メイン・メモリに知らせる。メイン・メモリへの書き込みの際には、キャッシュ・データは、自動的に無効化されることができ、そして読み出しの際には、データは、メイン・メモリではなくキャッシュによりリクエストへ供給されることができ。コヒーレンシに対するソフトウェアのアプローチは

50

、ユーザがソフトウェアによりキャッシュをフラッシュし、無効化し、そして同期化するのを可能にすることに基づいている。これは、これらの動作をソフトウェア形態により実行するコントローラを追加することにより行われる。ハードウェア・コヒーレンシ機構の主な利点は、それが自動的に行われ、即ち、ユーザがその動作を管理する必要がないことである。ハードウェア・コヒーレンシ機構の主な欠点は、実現するのに非常に複雑で、高い電力を消費し、そして半導体の追加の範囲を使うことである。デジタル信号プロセッサ(DPS)のような低コスト低電力システムにおいては、ハードウェアの解決法は、適切でない。

【0016】

キャッシュ・コヒーレンシ動作の例は、ヨーロッパ特許出願EP1182566A1に記載されている。その特許出願は、メイン・メモリ・ブロックの開始及び終了アドレスを定義し、その結果、キャッシュ・ラインの解像度でその範囲の全てのアドレスを通るようステップを進めることに基づくキャッシュ・メンテナンス動作を記載する。各ステップについて、メイン・メモリ・アドレスは、キャッシュ・メモリ・タグ・アレイに格納された全ての値と比較され、一致が検出された場合、コヒーレンシ動作が実行される。しかしながら、これは、非常に時間を消費するプロセスをもたらす。更に、プロセス時間がメイン・メモリ・アドレスとタグ・アレイとの間の並列ハードウェア比較を導入することにより低減されることができるとも拘わらず、これは、ハードウェアの複雑さを増大し、従ってコストを増大する。

10

【0017】

更に、コヒーレンシ動作の持続時間は、処理されるメモリ・ブロックの大きさに依存する。従って、メモリ範囲の大きさが増大するにつれ、増大した数のアドレスを通るようステップしなければならず、それにより持続時間が増大する。これは、特に、プロセス持続時間の不確かさが異なるプロセスのリアルタイム管理を著しく複雑にするリアルタイム・システムでは著しい欠点である。

20

【0018】

米国特許明細書No. 2002-065980は、各プロセッサと関連したプライベート・レベル1のキャッシュと、1エントリ当たり幾つかのセグメントを有する共用のレベル2のキャッシュと、レベル3の物理的メモリを含む幾つかのプロセッサを有するデジタル・システムを記載する。米国特許明細書No. 2002-065980は、キャッシュ・ライン上の「一致(マッチ)」を定義するため2つのクォリファイアを用いる機構を開示する。

30

【0019】

ヨーロッパ特許No. 1030243は、キャッシュの中の「ダーティ・エントリ」を文脈切り替え中にクリーンにするため割り込み可能ハードウェア・クリーン機能を用いる仮想インデックス、仮想タグ・キャッシュを記載する。MAXカウンタ及びMINレジスタは、ダーティである或る範囲のキャッシュ場所を定義する。ハードウェア・クリーン機能を実行中に、MAXカウンタは、下方にカウントし、一方、MAXカウンタにより与えられるアドレスでのキャッシュ・エントリは、そのエントリがダーティとマーキングされる場合メイン・メモリへ書き込まれる。とりわけ、割り込みが生じた場合、MAXカウンタは、その割り込みがサービス提供された後で、後続のクリーン要求が発行されるまで使用不能にされる。

40

【発明の開示】

【発明が解決しようとする課題】

【0020】

従って、メモリ・キャッシュについてコヒーレンシ動作を実行する改善されたメモリ・キャッシュ制御装置、処理システム、及び方法が有利であり、特に、柔軟性を増大し、複雑さを低減し、消費時間を低減し、コストを低減し、信頼性を増大し、及び/又は性能を改善することを可能にするシステムが、有利であろう。

【課題を解決するための手段】

50

【0021】

本発明は、添付の特許請求の範囲に記載されるメモリ・キャッシュ制御装置、メモリ・キャッシュ・システム、処理システム、及び記憶媒体を提供する。

従って、本発明は、好ましくは、前述の欠点の1つ又はそれより多くの欠点を単独で又はいずれかの組み合わせで緩和、軽減又は排除しようとするものである。

【発明を実施するための最良の形態】

【0022】

本発明の例示的实施形態が、ここで、添付図面を参照して説明されるであろう。

図1は、本発明の一実施形態に従ったキャッシュ・メモリ・システムを備えるプロセッサ・システムの説明図である。

10

【0023】

プロセッサ・システム100は、プロセッサ101及びメイン・メモリ103を備え、当該メイン・メモリ103は、アプリケーションを実行中にプロセッサ101により用いられる命令及びデータを格納する。プロセッサ101は、例えば、マイクロプロセッサ又はデジタル信号プロセッサであり、そしてメイン・メモリ103は、この実施形態においては、ダイナミックRAM(ランダム・アクセス・メモリ)である。メイン・メモリ103は、比較的大きく、そして例えば、1ギガバイトのオーダである。プロセッサ101及びメイン・メモリ103は、キャッシュ・メモリ・システム105に結合され、当該キャッシュ・メモリ・システム105は、メイン・メモリ103と一緒に、プロセッサ・システム100のための階層的記憶構造を形成する。

20

【0024】

キャッシュ・メモリ・システム105は、キャッシュ・メモリ107及びキャッシュ・コントローラ109を備える。キャッシュ・メモリ107は、記載される実施形態においては、メイン・メモリ103により用いられるダイナミックRAMより著しく早いステイックRAMである。しかしながら、キャッシュ・メモリ107は、メイン・メモリ103より相当に小さく、例えば、256キロバイトのオーダである。キャッシュ・コントローラ109は、階層的メモリ・システムの動作を制御し、特に、キャッシュ・メモリ・システム105の動作及びメイン・メモリ103のアクセスを制御する。

【0025】

動作において、プロセッサ101により実行されるタスクは、メイン・メモリ103のアドレス空間内の記憶場所をアドレスすることによりメモリにアクセスする。これらのメモリ・アクセスは、キャッシュ・メモリ・システム105により行われてもよく、又はメイン・メモリ103へのメモリ・アクセスをもたらしてもよい。特に、読み出し動作については、キャッシュ・コントローラ109は、キャッシュ・メモリ107が指定されたメイン・メモリ・アドレスのための有効なデータを含むかどうかを決定し、そして含む場合この値が、検索され、そしてプロセッサ101へフィードバックされる。特に、キャッシュ・マッチが検出された場合、キャッシュ・メモリ・システム105は、適切なデータをデータ・バス上に置く。キャッシュ・コントローラ109がキャッシュ・メモリ107が指定されたメイン・メモリ・アドレスのための有効なデータを含まないと決定した場合、キャッシュ・コントローラ109は、適切なデータをメイン・メモリ103から検索する。特に、キャッシュ・コントローラ109は、メイン・メモリ103が適切なデータをデータ・バス上に置くようにさせる。

30

40

【0026】

キャッシュ・ミスが生じたとき、キャッシュ・コントローラ109は更に、同じメイン・メモリ・アドレスが多くの場合前のアクセス後の短い期間に再びアクセスされるので、メイン・メモリ103から検索されたデータをキャッシュ・メモリ107にロードする。メイン・メモリ103の緩慢な応答時間に起因して、待ち信号が、典型的には、アクセスされ、それにより追加の待ち状態を読み出しプロセスに導入する。従って、キャッシュ・ヒットは、キャッシュ・ミスについての場合より早いメモリ・アクセスをもたらすであろう。更に、現在の記憶場所近くの記憶場所がアクセスされる確率が増大するので、キャッ

50

シュ・コントローラ 109 は、典型的には、その記憶場所に隣接した記憶場所からデータを転送する。

【0027】

実施形態がキャッシュ・コントローラ 109 を参照して単一の分離された機能モジュールとして説明されるが、これは単に説明の簡潔さ及び明瞭化のためであり、キャッシュ・コントローラ 109 がいずれの適切な方法で実現され得ることが認められるであろう。特に、キャッシュ・コントローラ 109 は、ハードウェア、ソフトウェア、又はそれらの組み合わせで実現され得る。その上、キャッシュ・コントローラ 109 は、例えば、キャッシュ・メモリ 107 又はプロセッサ 101 と統合化され得て、又は別個のモジュールであってもよい。特に、キャッシュ・コントローラ 109 の全部又は一部は、プロセッサ 101 上でランするソフトウェアで、又は別個のプロセッサ又はメモリ管理ユニットで完全に又は部分的に実現されてもよい。

10

【0028】

図 2 は、キャッシュ・メモリ 107 の構造の説明図である。この例においては、キャッシュ・メモリ 107 は、 2^K 個のキャッシュ・ラインを備える直接マップ方式キャッシュ・メモリである。この例においては、各キャッシュ・ラインは、4 データ・バイトを備え、メイン・メモリ・アドレッシングの解像度は、1 バイトである。図示の例では、 $K = 3$ であり、従って、キャッシュは、32 バイトを備える。実際のキャッシュは、典型的には、これより著しく大きいことが認められるであろう。例えば、PC 用の現在のキャッシュ・メモリは、典型的には、各キャッシュ・ラインにおいて 16 から 32 バイト、及び例えば、8192 個のキャッシュ・ライン（即ち、 $K = 13$ ）を有するキャッシュを備える。

20

【0029】

単純化のため、メイン・メモリ 103 は、特定の事例においては、10 ビット・アドレス空間に対応する 1 キロバイトを備えると考えられる。実際のメイン・メモリは、典型的には遙かに大きく、そして著しく長いアドレスを有することが認められるであろう。この例では、プロセッサ 101 によりアドレス・バス上に置かれたメイン・メモリ・アドレスは、次のように、バイナリ値により表される。

【0030】

$b_9, b_8, b_7, b_6, b_5, b_4, b_3, b_2, b_1, b_0$

この例においては、キャッシュ記憶場所へのマッピングは、アドレス・ビットとキャッシュ記憶場所との間の固定のマッピングにより達成される。従って、この例では、 b_1, b_0 は、キャッシュ・ライン内のバイト場所を決定し、 b_4, b_3, b_2 は、キャッシュ・ライン・アドレスを決定し、これはインデックスとして知られている。従って、 $b_1, b_0 = 1, 0$ 及び $b_4, b_3, b_2 = 1, 0, 1$ を有するアドレスは、キャッシュ・ライン $101_b = 5$ の記憶場所 10_b へマッピングする。直接マップ方式キャッシュの例では、 $b_1, b_0 = 1, 0$ 及び $b_4, b_3, b_2 = 1, 0, 1$ を有する全てのメイン・データ・アドレスは、このキャッシュ場所へマッピングするのである。

30

【0031】

キャッシュ・メモリ・システム 105 は、所与のキャッシュ・ラインがどの記憶場所と関連しているかを、並びにキャッシュ・ラインに保持されたデータの状態の経過を連続的に追う。詳細には、キャッシュ・コントローラ 109 は、キャッシュ・ラインが現在関連しているメイン・メモリ・アドレスの上位アドレス・ビットの値を格納する。上位アドレス・ビットは、この場合、タグとして知られ、そしてキャッシュ・コントローラ 109 は、タグ・アレイを維持する。タグ・アレイは、各キャッシュ・ラインに関するエントリを備え、各エントリは、キャッシュ・ラインを選択するため用いられる K 個のデータ・ビット（インデックス）によりアドレスされる。キャッシュ・ラインが新しいメイン・メモリ・アドレスと関連付けられたとき、前のタグ・アレイは、その新しいメイン・メモリ・アドレスの上位アドレス・ビットにより、即ち、特定の事例ではデータ・ビット b_9, b_8, b_7, b_6, b_5 によりオーバーライトされる。

40

【0032】

50

従って、プロセッサ101が読み出し動作を実行するときは常に、キャッシュ・メモリ・システム105は、インデックス(b_4, b_3, b_2)を用いてタグ・アレイにアクセスし、且つ格納されたタグを現在のアドレス(b_9, b_8, b_7, b_6, b_5)の上位アドレス・ビットと比較することにより、対応の値がキャッシュ・メモリに存在するかどうかを決定する。タグがアドレスと一致し、そしてフラグが格納されたキャッシュ・データが有効であることを指示する場合、キャッシュ・メモリからのデータ値が、データ・バス上に置かれ、その結果低い待ち時間の読み出し動作をもたらす。

【0033】

直接マップ方式キャッシュの欠点は、各メイン・メモリ・アドレスが単一のキャッシュ・ラインと関連付けられることができるだけで、その結果異なるメイン・メモリ・アドレス間の競合の確率が増大し、そして非常に軽くロードされたキャッシュに対してすら重要であることをもたらすことである。例えば、たとえ大きいキャッシュ・メモリの唯1つのキャッシュ・ラインが所与のメイン・メモリ・アドレスと関連したとしても、これがたまたま同じインデックスを既に関連したメイン・メモリ・アドレスとしてもたらす場合、第2のメイン・メモリ・アドレスをキャッシュと関連付けることは不可能である。

10

【0034】

フルアソシエイティブ方式キャッシュは、各キャッシュ・ラインをいずれのメイン・メモリ・アドレスと関連付けることを可能にすることにより著しく高い柔軟性を与える。詳細には、これは、インデックスがゼロ・ビットを備え、且つタグがキャッシュ・ライン内の1つの場所をアドレスするためには用いられない全てのアドレス・ビットを備えることと等価と考えられる。

20

【0035】

セットアソシエイティブ・キャッシュは、直接マップ方式キャッシュとフルアソシエイティブ方式キャッシュとの間の中間と見ることができる。セットアソシエイティブ・キャッシュにおいては、キャッシュ・メモリのブロックは、直接マップ方式メモリ・キャッシュに関して特定の下位アドレス・ビットと関連付けられる。しかしながら、直接マップ方式キャッシュとは異なり、複数のキャッシュ・ブロックは、同じアドレスにマッピングされる。例えば、上記の例では、3ビット b_4, b_3, b_2 のインデックスを有するよりむしろ、セットアソシエイティブ・キャッシュは、2ビット b_3, b_2 のインデックスを用いるだけである。従って、8個のキャッシュ・メモリの単一のブロックを有する代わりに、キャッシュ・メモリは、ここで、1ブロック4個のキャッシュ・ラインを2ブロック備える。従って、各メイン・メモリは、2つの異なるキャッシュ・ラインと関連付けられる。

30

【0036】

従って、キャッシュ・メモリ・システム105は、所与のインデックスに対して複数のエントリを有するタグ・アレイを維持する。従って、例えば、読み出し動作が生じたとき、直接マップ方式キャッシュに関するまさに単一のエントリとは異なりタグ・アレイの中の複数のエントリをチェックすることが必要である。しかしながら、チェックしなければならないエントリの数は、相変わらず比較的小さく、そして動作は、並列処理により促進され得る。

40

【0037】

従って、キャッシュ・メモリ・システム105がメモリ・アクセスがキャッシュ・メモリ107と関連するか又はメイン・メモリ103と関連するかを決定するため、キャッシュ・メモリ・システム105は、各キャッシュ・ラインについて、メイン・メモリ103に対する関連(連想)を指示するデータを備えるデータ・アレイ(タグ・アレイ)を維持する。その上、キャッシュ・メモリ・システム105は、キャッシュ・ラインのデータの状態の経過を追う。特に、キャッシュ・メモリ・システム105は、新しいデータ・バスがメイン・メモリではなくて所与のキャッシュ・ラインに書き込まれたかどうかを指示する状態指示を維持する。そうである場合、キャッシュ・メモリ107のデータとメイン・メモリ103のデータとの不一致が存在し、そしてキャッシュ・メモリ107のデータは

50

、そのデータがキャッシュから落とされる、又はメイン・メモリ103が直接アクセスされる前に、メイン・メモリ103に書き込まなければならない。この指示は、ダーティ・ビット指示と呼ばれる。

【0038】

同様に、読み出し動作について、有効な指示を用いて、キャッシュ・ラインがメイン・メモリ103から検索された有効なデータを備えるかどうかを指示する。

状態指示は、一部の実施形態においては、キャッシュ・ライン全体と関連し、又はキャッシュ・ラインの中の各場所に関する個々の状態指示が、例えば、維持されることが認められるであろう。

【0039】

階層的メモリ・システムを管理するため、コヒーレンシ（メンテナンス）動作が要求されることが認められるであろう。そのようなコヒーレンシ動作は、キャッシュ・メモリ107とメイン・メモリ103との間のコヒーレンシを維持する動作を含み、この動作には、メンテナンス書き込み動作、読み出し動作、同期化動作等が含まれる。

【0040】

図3は、本発明の一実施形態に従ったキャッシュ・メモリ・システム105をより詳細に示す。図示及び記載は、簡潔さ及び明瞭化のため、この実施形態を説明するため要求される機能に焦点を当てる。特に、記載は、直接マップ方式キャッシュに関するコヒーレンシ動作を実行するときのキャッシュ・メモリ・システム105の動作に焦点を当てる。

【0041】

実施形態においては、キャッシュ・メモリ・システム105は、命令をプロセッサ101から受け取る受信プロセッサ301を備える。受信プロセッサ301は、制御ユニット303に結合され、当該制御ユニット303は、キャッシュ・メモリ・システム105のコヒーレンシ動作を制御する。制御ユニット303は更に、タグ・アレイ305、並びにキャッシュ・メモリ107、及びメイン・メモリ103に結合される。

【0042】

本発明の実施形態によれば、コヒーレンシ動作は、アドレス・グループ指示をプロセッサ101から受け取る受信プロセッサ301により開始される。アドレス・グループ指示は、メイン・メモリ103の中の1群の記憶場所を識別する。記載される実施形態においては、このグループは、開始アドレスで開始し、終了アドレスで終了する連続のブロックの記憶場所から成る。しかしながら、他の実施形態及び他の応用においては、アドレス・グループが、メイン・メモリ103の交わらないアドレス範囲を含む他のグループのアドレスに対応することが認められるであろう。

【0043】

記載される実施形態においては、受信プロセッサ301は、こうして、開始アドレス及び終了アドレスから成るアドレス・グループ指示を受け取る。受信プロセッサ301は更に、特定のコヒーレンシ動作が指定されたアドレス範囲について実行されることになる指示を受け取る。例えば、アドレス範囲は、所与のアプリケーションに対応し、そしてコヒーレンシ動作は、アプリケーションが終了することに起因して引き起こされる。別の例としては、DMA動作は、メイン・メモリ103の指定されたアドレス範囲を直接アクセスするようセットアップされ、そしてコヒーレンシ動作は、このアドレス範囲に関してキャッシュに書き込まれた全データがDMA動作の前にメイン・メモリ103に転送されることを保証するため引き起こされる。

【0044】

受信プロセッサ301は、開始アドレス及び終了アドレスを制御ユニット303に供給し、当該制御ユニット303は、これらの値を格納する。次いで、制御ユニット303は、コヒーレンシ・プロセスを実行することを進める。しかしながら、従来のアプローチとは反対に、制御ユニット303は、キャッシュ・エントリが周波数範囲の各アドレスに関して存在するかどうかを決定するため、アドレス範囲のメイン・メモリ・アドレスを通るようステップを進めることはしない。むしろ、この実施形態においては、制御ユニット3

10

20

30

40

50

03は、タグ・アレイ305を通るようステップを進め、そして各エントリに関して、キャッシュ・ラインが適切なマッチ判定基準に従ってメイン・メモリ・アドレス範囲と関連付けられるかどうか決定することにより各キャッシュ・ラインを順次に処理する。キャッシュ・ラインがメイン・メモリ・アドレス範囲と関連付けられていることが分かった場合、制御ユニット303は、キャッシュ・ラインについて要求されたコヒーレンシ動作を実行する。

【0045】

例えば、制御ユニット303は、最初に、ゼロ・インデックスについて格納されたタグを検索する。対応のメイン・メモリ・アドレスは、タグとインデックスとを組み合わせることにより決定され、そしてその結果得られたアドレスは、開始及び終了アドレスと比較される。アドレスが上記範囲内に入る場合、コヒーレンシ動作が、そのキャッシュ・ラインについて実行される。例えば、コヒーレンシ動作がアドレス範囲と関連付けられたキャッシュの構成要素をフラッシュすることを備える場合、制御ユニット303は、キャッシュ・ラインのデータをメイン・メモリ103に書き込むようにする。次いで、制御ユニット303は、次のインデックスについて、即ち、1のインデックスについて格納されたタグを検索することを進め、次いでこのキャッシュ・ラインについてプロセスを繰り返す。

【0046】

従って、制御ユニット303は、一時に1つのキャッシュ・ラインでタグ・アレイ305を通るようステップを進め、そして各ラインについて、キャッシュ・ラインが指定されたメモリ範囲と関連付けられる場合キャッシュ・メモリ107について要求されたコヒーレンシ動作を実行する。

【0047】

説明されたアプローチは、従来技術を超えた多数の利点を提供し、そして柔軟性があり、複雑さが少なく、低コストで高信頼性のキャッシュ・メモリ・システムを促進又は可能にする。

【0048】

詳細には、メイン・メモリ・アドレス範囲が典型的にはキャッシュ・サイズより遙かに大きいので、比較サイクルの数がより少なくてよいと考えられる。換言すると、マッチ判定基準を評価し且つコヒーレンシ動作を条件付きで実行するループの反復の回数が、著しく低減される。これは、典型的には、コヒーレンシ・プロセスの持続時間を著しく低減し、それにより計算負荷を低減し、そしてシステムを他の活動のために解放する。

【0049】

更に、コヒーレンシ動作の持続時間は、アドレス範囲の大きさよりむしろキャッシュの大きさに依存する。これは、コヒーレンシ・プロセスのため必要とされる時間を低減しようとするばかりでなく、それが限度内に留められ、アドレス範囲と無関係となることをもたらす。これは、特に、リアルタイム処理システムでは著しい利点であり、そしてそのようなシステムの時間管理を容易にする。

【0050】

更に、アプローチが、比較的単純で、そして複雑さが少ないハードウェア、ソフトウェア、ファームウェア、又はそれらの組み合わせにより実現される。特に、制御ユニット303の機能は、プロセッサ101のファームウェア・ルーチンとして少なくとも部分的に実行され得る。

【0051】

明瞭化のため上記の説明は、キャッシュ・ラインのデータの状態の評価を考慮していないことが認められるであろう。しかしながら、制御ユニット303は、キャッシュ・ラインのデータの状態を決定することが好ましい。従って、マッチ判定基準は、キャッシュ・ライン・データの状態の考慮を備えることが好ましく、及び/又はコヒーレンシ動作は、キャッシュ・ライン・データの状態に回答して実行される。例えば、データは、状態指示がダーティ・ビット状態に対応する場合メイン・メモリ103に書き込まれるのみである。

【0052】

10

20

30

40

50

上記の説明はキャッシュ・ライン評価を詳細に考慮したにも拘わらず、プロセスはまた、キャッシュ・ラインの異なる構成要素の間を分離し得ることが認められるであろう。例えば、開始及び/又は終了アドレスは、キャッシュ・ライン分割と一致する必要がなく、キャッシュ・ライン内のデータ要素に対応してもよい。また、データの状態は、個々の要素と関連し、そしてコヒーレンシ動作は、個々の要素を考慮してもよい。例えば、状態指示は、キャッシュ・ラインの中の個々のデータ・バイトと関連し、そしてダーティ・ビット指示が設定されるデータ・バイトのみがメイン・メモリ103に書き込まれる。

【0053】

また、制御ユニット303は、一時に1つのキャッシュ・ラインでキャッシュ・メモリ全体を通るようステップを進めることが好ましいにも拘わらず、一部の実施形態において、キャッシュ・ラインのサブセットのみを通るようステップを進めることが有利であり、そしてこのサブセットは、例えば、事前定義、又は動的に決定し得ることが認められるであろう。

10

【0054】

コヒーレンシ・プロセス及び動作は、いずれの適切なコヒーレンシ・プロセス及び動作であってよい。

詳細には、コヒーレンシ動作は、無効化動作であり得る。無効化動作は、指定されたアドレス範囲と関連した全てのキャッシュ・ラインを無効化することが好ましい。従って、制御ユニット303は、キャッシュを通るようステップを進め、そしてアドレス範囲に対応する全てのキャッシュ・ラインについて無効にするよう状態指示を設定する。この動作は、例えば、データがメイン・メモリ103において(DMAにより)更新された状況、又はキャッシュが一時的変数を必要としないようなタスクの終わりに無効化されることができ、当該一時的変数をキャッシュ・メモリ107の中に保持する状況で有利である。

20

【0055】

代替として、又は追加で、コヒーレンシ動作は、同期化動作であってよい。同期化動作は、指定されたアドレス範囲と関連した全てのキャッシュ・ラインを同期化する。従って、制御ユニット303は、キャッシュを通るようステップを進め、メイン・メモリ103にダーティ・セクションを書き込み、そしてアドレス範囲に対応する全てのキャッシュ・ラインに関して有効な指示を保ちながらダーティ指示を否定する。

【0056】

この動作は、例えば、メモリ・セクションが、後の使用のためキャッシュ・メモリ107の中のデータの有効性を維持しながら、DMAによりメイン・メモリ103から読み出されることになる状況では有利である。同期化動作の別の使用は、自由サイクル(free cycles)を利用して、キャッシュ・メモリ107の中のダーティ・セクションの数を低減する。

30

【0057】

代替として、又は追加で、コヒーレンシ動作は、フラッシュ動作であってよい。フラッシュ動作は、指定されたアドレス範囲と関連した全てのキャッシュ・ラインをフラッシュする。従って、制御ユニット303は、キャッシュを通るようステップを進め、そしてアドレス範囲に対応し且つダーティ・ビット指示を有する全てのキャッシュ・ラインのデータをメイン・メモリ103に書き込み、次いでキャッシュ・ラインを無効化する。この動作は、例えば、キャッシュ・メモリ・システム105の介入無しで且つデータがプロセッサ101により用いられることを期待されないとき、記憶動作がメイン・メモリ103について直接実行されようとしている状況で有利である。

40

【0058】

以下においては、セットアソシエイティブ・メモリに適用される本発明の一実施形態が説明される。この実施形態においては、キャッシュ・メモリ107は、4つの組に編成される。メイン・メモリ・アドレスは、これらの組のうちのいずれかと関連付けられ、従って、各メイン・メモリ場所に関して4つのあり得るキャッシュ・ラインがある。実施形態は、図2に示されるキャッシュ・メモリ・システム105と互換性があり、そしてこれを

50

参照して説明されるであろう。

【0059】

実施形態においては、プロセッサによるアドレッシングは、仮想メモリ・アドレッシングを採用する。詳細には、プロセッサ101上でランする各タスクは、メモリ管理ユニットによりメイン・メモリ103の中の所与の物理的メモリ範囲にマッピングされる標準アドレス空間を用いる。ランしている各タスクは、メイン・メモリ103にマッピングするとき、メモリ管理ユニットにより用いられるタスク識別に割り当てられる。例えば、最初のタスクの命令は、範囲 $[0, FFF F_h]$ の中のメモリをアドレスする。メモリ管理ユニットは、このタスクをタスク識別1に割り当て、そしてその範囲を $[10000_h, 10FFF F_h]$ の物理的メモリ範囲にマッピングする。第2のタスクの命令はまた、メモリを範囲 $[0, FFF F_h]$ でアドレスする。メモリ管理ユニットは、このタスクをタスク識別2に割り当て、そしてその範囲を $[08000_h, 08FFF F_h]$ の物理的メモリ範囲にマッピングする。

10

【0060】

図4は、この実施形態に従ったキャッシュ・メモリ・システム105のためのタグ・アレイ400の一例を示す。タグ・アレイ400は、4つの別個のデータ構造401、403、405、407を備え、それぞれのデータ構造は、セットアソシエイティブ・キャッシュの4つの組のうちの一つに対応する。従って、エントリは、各キャッシュ・ラインについてのタグ・アレイに存在する。実施形態においては、各エントリは、プロセッサ101により用いられる仮想アドレスの上位ビットに対応するタグを備える。その上、各エントリは、キャッシュ・ラインがどのタスクと関連付けられるかを指示するタスク識別を備える。従って、タグ・アレイにおけるエントリは、キャッシュ・ラインと関連した物理的メイン・メモリ・アドレスを示す。

20

【0061】

図5は、本発明のこの実施形態に従ったキャッシュ・メモリ・コヒーレンシ動作を実行する方法のフロー・チャートを示す。記載される実施形態においては、この方法は、1又はそれより多くのアプリケーションをサポートするマイクロプロセッサ、中央処理装置(CPU)又はデジタル信号プロセッサ(DSP)のようなプロセッサにより実行される。図5の方法は、ユーザ・アプリケーションの処理に対してバックグラウンドで実行される。

30

【0062】

本方法は、ステップ501で開始し、そこにおいては、制御ユニット303がコヒーレンシ動作が実行されることになるアドレス範囲を定義する開始アドレス及び終了アドレスを用いて初期化される。開始アドレス及び終了アドレスは、所与のタスクにより用いられる仮想アドレスとして指定される。例えば、最初のタスクが範囲 $[0, FFF F_h]$ の中のメモリをアドレスするケースについては、開始アドレス及び終了アドレスは、この範囲内にある。仮想アドレスを物理的メイン・メモリ103のアドレス範囲と関連付けるため、制御ユニット303は更に、タスク識別(タスクID)により初期化される。特定の例では、コヒーレンシ動作は、最初のタスクのため仮想メモリ間隔 $[10000_h, 17FFF_h]$ と関連付けられる。従って、制御ユニット303は、ステップ501において、開始アドレスを 10000_h に、また終了アドレスを $17FFF_h$ に、更にタスクIDを1に設定することにより初期化される。

40

【0063】

本方法は、ステップ503に継続し、そこにおいて、キャッシュ・ライン・ポインタは、タグ・アレイ400の中の第1の組に関する第1のエントリ401に対応する第1のキャッシュ・ラインに設定される。

【0064】

ステップ503には、ステップ505が続き、そこにおいては、タグ及びタスク識別が、タグ・アレイ400から検索される。従って、現在のタグ(0, 0)及びタスクID(0, 0)が、タグ・アレイ400から検索される。

50

【0065】

ステップ505には、ステップ507が続き、そこにおいては、制御ユニット303は、第1のエントリ401に対応するキャッシュ・ラインがコヒーレンシ動作が実行されるべきであるアドレスと関連付けられるかどうかを決定する。詳細には、制御ユニット303は、検索されたタグをタグに関するインデックスと組み合わせることによりアドレスを発生する。従って、フル仮想アドレスが、タグからのアドレス・ビットをインデックスのアドレス・ビットと組み合わせることにより第1のエントリ401に対して発生される。

【0066】

発生されたアドレスは、開始及び終了アドレスと比較され、そして制御ユニット303は、検索されたタスクIDが指定されたタスクIDと一致するかどうかを決定する。従って、1のタスクIDがタスクID(0,0)に格納されているかどうか決定される。発生されたアドレスが指定されたアドレス範囲内であり、且つタスクIDが一致している場合、一致が、指示され、従って、対応のキャッシュ・ラインについてコヒーレンシ動作を実行することが望ましい。このケースにおいては、本方法は、ステップ509に継続され、そして他の場合には、本方法は、ステップ513に継続される。

10

【0067】

ステップ509において、コヒーレンシ動作を実行することが現在実際的であるかどうか決定される。詳細には、制御ユニット303は、競合がコヒーレンシ動作と別のメモリ動作との間に存在するかどうかを決定する。制御ユニット303は、例えば、コヒーレンシ動作と他のメモリ動作との間で共用される資源が他のメモリ動作により現在使用されているかどうかを決定する。例えば、キャッシュ・メモリ107が通常のキャッシュ動作(キャッシュ・ライン再割り当て)とコヒーレンシ動作との間で共用されている資源にアクセスする場合、より高い優先順位が、競合が上記2つの間に存在するとき上記通常のキャッシュ動作に与えられる。

20

【0068】

競合がステップ509において存在すると決定された場合、現在の実施形態においては、制御ユニット303は、コヒーレンシ動作を禁止するよう進める。特に、制御ユニット303は、他のメモリ動作が終了されるまでコヒーレンシ動作を遅延させることによりそのコヒーレンシ動作を禁止する。これは、ステップ519において、ラインが同時ライン動作(*concurrent line operation*)により達成される。ラインがステップ519において置換されてしまった場合、本方法は、ステップ513へ動く。ラインがステップ519において置換されなかった場合、プロセスは、ステップ509に戻って、コヒーレンシ動作を実行することが現在実際的であるかどうかを決定する。

30

【0069】

従って、スイープ・セグメント取り消し判定基準(ステップ519における)は、マッチ判定基準がステップ507において前にチェックされたので、スイープ・セグメント(掃引セグメント)と関連したキャッシュ・ラインが既に置換されてしまったかどうかを識別する。

【0070】

競合がステップ509において無いと決定されたとき、本方法は、ステップ511に進み、そこにおいて、制御ユニット303は、対応のキャッシュ・ラインについて所望のコヒーレンシ動作を実行する。前述したように、コヒーレンシ動作は、例えば、フラッシュ動作、無効化動作、又は同期化動作であり得る。

40

【0071】

ステップ511には、ステップ513が続き、そこにおいて、制御ユニット303は、それがキャッシュ全体を通るようステップを進んだどうかを決定する。そのようにした場合、本方法は、ステップ515に継続し、そこにおいて、プロセスは、終了する。他の場合は、本方法は、ステップ517に継続され、そこにおいて、ポインタは、次のキャッシュ・ラインを参照するため更新される。次いで、本方法は、次のキャッシュ・ラインを処理することによりステップ505に継続される。次のキャッシュ・ラインは、上記の組の

50

中の後続のキャッシュ・ラインとして決定される。組の最後のキャッシュ・ラインに達したとき、次のキャッシュ・ラインは、次の組の最初のキャッシュ・ラインとして決定される。最後の組の最後のキャッシュ・ラインに達したとき、このことが、ステップ513において検出され、その結果本方法が終了することになる。

【0072】

従って、本方法は、結果として、順次に各個々の組のキャッシュ・ラインを通るようステップを進め、並びに順次に個々の組を通るようステップを進めることになる。従って、実施形態においては、キャッシュの全てのキャッシュ・ラインが、順次に処理され、そして各キャッシュ・ラインに関して、コヒーレンシ動作が適切かどうか決定され、そして適切である場合、コヒーレンシ動作が実行される。

10

【0073】

詳細には、図4のタグ・アレイ400は、最初にセット0の最初のエン트리409を評価し、続いてセット0(組0)の次のエン트리411を評価し、以下セット0の最後のエン트리413に達するまで同様に評価することによりステップするよう進められる。次いで、本方法は、セット1(組1)の最初のエン트리415を指すことによりセット1にステップを進める。同様に、セット1の最後のエン트리417に、セット2(組2)の最初のエン트리419が続き、そしてセット2の最後のエン트리421に、セット3(組3)の最初のエン트리423が続く。セット3の最後のエン트리425に達したとき、コヒーレンシ動作の実行が完了する。

【0074】

説明された実施形態はステップが説明の順序で順次実行される実現形態を記載したが、並列動作及び/又はステップの別の順序も等しく適切に適用されることが認められるであろう。特に、ステップ505、507、509、517は、ステップ511と並列に実行されてもよい。従って、キャッシュ・ラインについてコヒーレンシ動作を実行する間に、コントローラは、次の1又は複数のキャッシュ・ラインを評価してもよい。

20

【0075】

制御ユニット303は、プロセスがステップ515において終了するとき終了指示を設定するのが好ましい。詳細には、制御ユニット303は、プロセッサに割り込みシーケンスをもたらす割り込み指示を設定するようにする。割り込み指示は、ソフトウェア割り込み指示でもよく、又は信号をプロセッサ101の割り込み信号入力上に設定することのよう

30

【0076】

上記の実施形態は、指定されたアドレス範囲に基づく単一のマッチ判定基準に回答して一致が決定されることに焦点を当てたものである。しかしながら、他の実施形態においては、他の判定基準を用いてもよく、及び/又は複数の判定基準を用いてもよい。例えば、アドレス・グループ指示は、タスク識別から成り、そしてマッチ判定基準は、各キャッシュ・ラインがそのタスク識別と一致するかどうかを単純に決定してもよい。従って、コヒーレンシ動作は、対応のタスク識別を指定することにより、所与のタスクについて単純に実行される。

40

【0077】

制御ユニット303は、複数のマッチ判定基準間で選択するよう動作可能であることが好ましく、特に、制御ユニット303は、プロセッサ101から受け取られたデータに回答して異なるマッチ判定基準間を選択するよう動作可能であってもよい。

【0078】

例えば、制御ユニット303がコヒーレンシ・プロセス推進コマンドに関連して開始アドレス、終了アドレス及びタスク識別を受け取る場合、制御ユニット303は、タグ・アレイのエントリが3つの全ての要件に一致するデータを備えるかどうかを評価するマッチ判定基準を用いることにより処理を進める。しかしながら、開始アドレス及び終了アドレスのみをコヒーレンシ・プロセス推進コマンドに関連して受け取る場合、格納されたアド

50

レス・タグのみが、マッチ判定基準により考慮されるであろう。これは、どのタスクが特定のメモリ範囲を用いているかに関係なく、所与のメモリ範囲について単純なコヒーレンシ動作を可能にする。更に、制御ユニット303が、コヒーレンシ・プロセス推進コマンドを有するタスク識別のみを受け取る場合、マッチ判定基準は、そのタスク識別が一致するかどうかのみを決定する。これは、特定のタスクについて単純なコヒーレンシ動作を可能にする。最後に、特定のデータがコヒーレンシ・プロセス推進コマンドに関連して受け取られない場合、制御ユニット303は、キャッシュ・メモリ107とメイン・メモリ103との間の関連（連想）と関係なく、キャッシュ・メモリ107全体についてコヒーレンシ動作を実行する。

【0079】

上記の記載はデータ・メモリ・キャッシュに特に適切であるが、本発明はまた、例えば、命令メモリ・キャッシュに適用し得ることが認められるであろう。

従って、本発明の好適な実施形態は、同時のCPU及びキャッシュ・スリーピング・プロセスを扱う機構を説明する。いずれのスリーピング及びクリーニング動作も、幾つかのセグメントを含む。とりわけ、各セグメントは、特定のキャッシュ・ラインについて動作を実行する。

【0080】

本発明の好適な実施形態においては、スリープ・セグメント遅延又は取り消しの管理は、内部機構によりセグメント毎のベースで処理される。これは、シームレス並列CPU及びキャッシュ・スリープ動作を可能にする。これは、CPUが出来るだけ多くアクティブである（機能停止しない、又は待ちモードでない）ことを可能にする点で明らかな利点を提供する。従って、CPUがアクティブである一方、キャッシュ・スリープ動作がアクティブであり、そしてこの並列動作により生じるいずれの競合が、内部的に管理される。

【0081】

本発明は、1サイクル当たり1つの比較のみを実行することに限定されるものではなく、例えば、複数の比較が、並列に実行されともよいことが認められるであろう。

本発明の実施形態の特定の好適な実行が上記で説明されたが、当業者は、そのような発明概念の変化及び変更を容易に適用することができるであろうことが明らかである。

【0082】

特に、上記の記載は、明瞭化のため、処理システムの異なる機能ユニットを参照して本発明の実施形態を説明したことが認められるであろう。しかしながら、異なる機能ユニット間の機能のいずれの適切な分配が、本発明から逸脱することなしに用いられ得ることが明らかであろう。従って、特定の機能ユニットに対する参照は、厳密な論理的又は物理的構造、組織又は仕切りを示すのではなくて、記載された機能を提供する適切な手段に対する参照としてのみ見るべきである。例えば、キャッシュ・コントローラは、統合化され、そしてプロセッサと組み合わせられ、又はこれの一部であってもよい。

【0083】

本発明は、ハードウェア、ソフトウェア、ファームウェア、又はそれらのいずれの組み合わせを含むいずれの適切な形態で実行することができる。しかしながら、本発明は、1又はそれより多くのデータ・プロセッサ上でランするコンピュータ・ソフトウェアとして実行されるのが好ましい。

【図面の簡単な説明】**【0084】**

【図1】図1は、本発明の一実施形態に従ったキャッシュ・メモリ・システムを備えるプロセッサ・システムの説明図である。

【図2】図2は、キャッシュ・メモリの構造の説明図である。

【図3】図3は、本発明の一実施形態に従ったキャッシュ・メモリ・システムを示す。

【図4】図4は、本発明の一実施形態に従ったキャッシュ・メモリ・システムのためのタグ・アレイの一例を示す。

【図5】図5は、本発明の一実施形態に従ったキャッシュ・メモリ・コヒーレンシ動作を

10

20

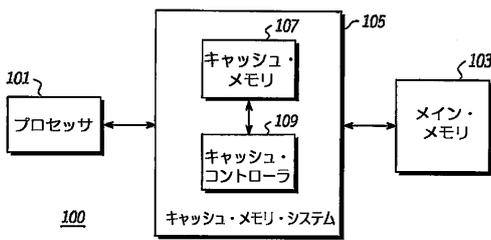
30

40

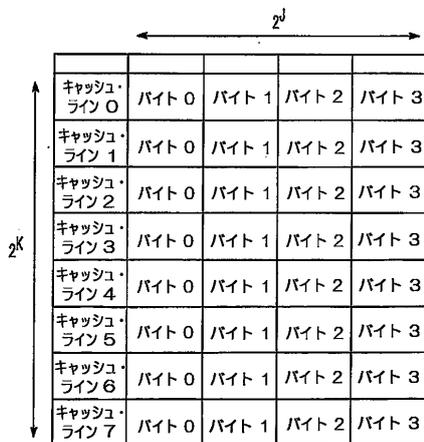
50

実行する方法のフロー・チャートを示す。

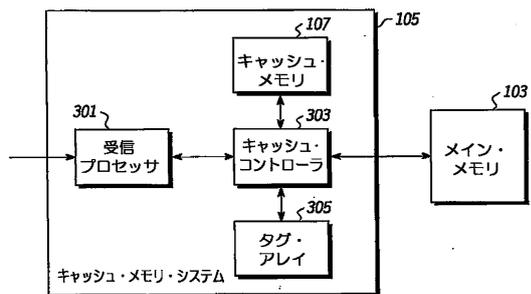
【 図 1 】



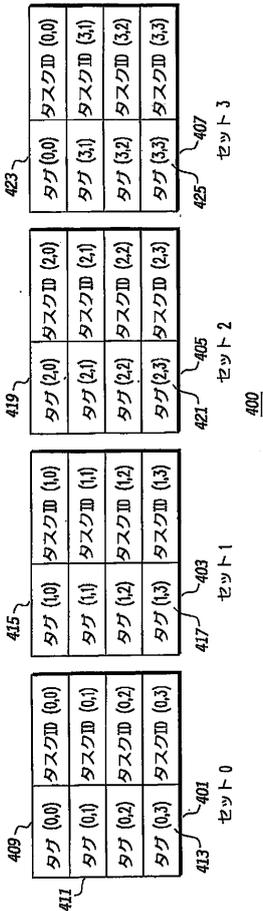
【 図 2 】



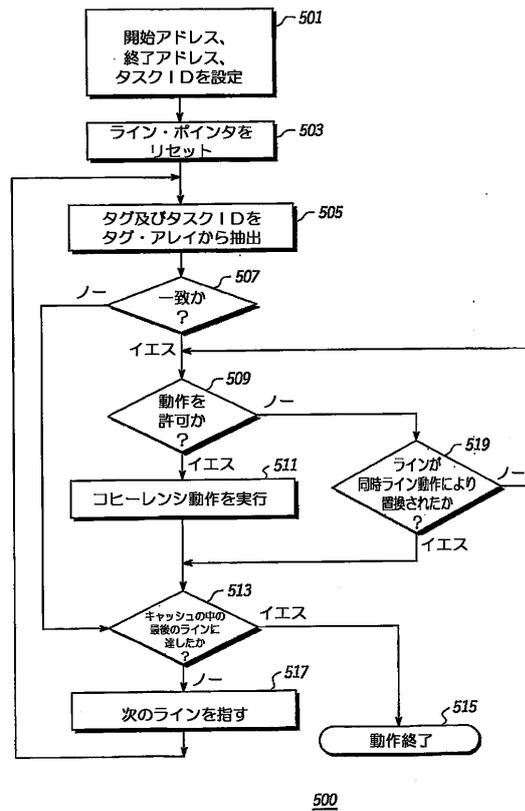
【 図 3 】



【 図 4 】



【 図 5 】



【 手続補正書 】

【 提出日 】平成 19年 8月 24日 (2007.8.24)

【 手続補正 1 】

【 補正対象書類名 】特許請求の範囲

【 補正対象項目名 】全文

【 補正方法 】変更

【 補正の内容 】

【 特許請求の範囲 】

【 請求項 1 】

メモリ・キャッシュ (1 0 5) についてコピーレンシ動作を実行するメモリ・キャッシュ制御装置であって、メイン・メモリ (1 0 3) と関連した複数のアドレスを備えるアドレス・グループに関するアドレス・グループ指示を受け取る手段 (3 0 1) を備えるメモリ・キャッシュ制御装置において、

1 群のキャッシュ・ラインの各キャッシュ・ラインを順次に処理する処理手段 (3 0 3) を備え、

当該処理手段 (3 0 3) が、

キャッシュ・ラインがアドレス・グループのアドレスと関連するかどうかをマッチ判定基準を評価することにより決定する手段と、

前記マッチ判定基準に適合する場合前記キャッシュ・ラインについてコピーレンシ動作を実行するコピーレンシ手段と、

競合が前記コピーレンシ動作と別のメモリ動作との間に存在するかどうかを決定する手段と、を備え、

前記コピーレンシ手段が、競合が存在する場合前記コピーレンシ動作を禁止するよう動作可能である

ことを特徴とするメモリ・キャッシュ制御装置。

【請求項 2】

前記マッチ判定基準が、前記キャッシュ・ラインと関連したメイン・メモリ・アドレスが前記アドレス・グループに属するかどうかの評価を備える請求項 1 記載のメモリ・キャッシュ制御装置。

【請求項 3】

前記アドレス・グループ指示が、前記メイン・メモリのメモリ・ブロックの開始アドレス及び終了アドレスを備え、

前記マッチ判定基準が、前記メイン・メモリ・アドレスが前記メモリ・ブロックに属するかどうかを決定することを備える

請求項 2 記載のメモリ・キャッシュ制御装置。

【請求項 4】

前記開始アドレス及び終了アドレスが、仮想メモリ・アドレスである請求項 3 記載のメモリ・キャッシュ制御装置。

【請求項 5】

前記のマッチ手段が、キャッシュ・ライン・タグ及びキャッシュ・ライン・インデックスにตอบสนองして前記メイン・メモリ・アドレスを決定するよう動作可能である請求項 2 から 4 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 6】

前記メモリ・キャッシュ(105)が、セットアソシエイティブ・メモリ・キャッシュであり、

前記 1 群のキャッシュ・ラインが、異なる組の前記セットアソシエイティブ・メモリのキャッシュ・ラインを備える

請求項 1 から 5 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 7】

前記処理手段(303)が、複数の組の前記セットアソシエイティブ・メモリ・キャッシュを順次に処理するよう動作可能である請求項 6 記載のメモリ・キャッシュ制御装置。

【請求項 8】

前記アドレス・グループ指示が、少なくとも 1 つのタスク識別の指示を備え、

前記マッチ判定基準が、第 1 のキャッシュ・ラインと関連したタスク識別が前記少なくとも 1 つのタスク識別と一致するかどうかの評価を備える

請求項 1 から 7 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 9】

前記アドレス・グループ指示が、タスク識別から成る請求項 8 記載のメモリ・キャッシュ制御装置。

【請求項 10】

前記 1 群のキャッシュ・ラインが、前記メモリ・キャッシュの全てのキャッシュ・ラインを備える請求項 1 から 9 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 11】

前記競合が、前記コヒーレンシ動作と前記別のメモリ動作との間で共用される資源に関連する請求項 1 記載のメモリ・キャッシュ制御装置。

【請求項 12】

競合が存在するかどうかを決定する前記手段が、前記コヒーレンシ動作と前記別のメモリ動作とが同じキャッシュ資源に対する実質的に同時のアクセスをもたらす場合競合が存在すると決定するよう動作可能である請求項 1 又は 11 記載のメモリ・キャッシュ制御装置。

【請求項 13】

前記コヒーレンシ手段が、前記コヒーレンシ動作と前記別のメモリ動作とのうちの 1 つを遅延させることにより前記コヒーレンシ動作を禁止するよう動作可能である請求項 1、11 及び 12 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 14】

前記コヒーレンシ動作が、無効化動作である請求項 1 から 1 3 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 1 5】

前記コヒーレンシ動作が、同期化動作である請求項 1 から 1 4 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 1 6】

前記コヒーレンシ動作が、フラッシュ動作である請求項 1 から 1 5 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 1 7】

前記処理手段(303)が、前記 1 群のキャッシュ・ラインの全てのキャッシュ・ラインが処理されてしまったことを決定することに応答して終了指示を設定する手段を備える請求項 1 から 1 6 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 1 8】

前記終了指示が、割り込み指示である請求項 1 7 記載のメモリ・キャッシュ制御装置。

【請求項 1 9】

前記メモリ・キャッシュが、命令キャッシュである請求項 1 から 1 8 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 2 0】

前記メモリ・キャッシュが、データ・キャッシュである請求項 1 から 1 9 のいずれか一項に記載のメモリ・キャッシュ制御装置。

【請求項 2 1】

請求項 1 から 1 9 のいずれか一項に記載のメモリ・キャッシュ制御装置を備えるメモリ・キャッシュ・システム。

【請求項 2 2】

プロセッサと、
メイン・メモリと、
前記プロセッサ及び前記メイン・メモリに結合されたキャッシュ・メモリと、
請求項 1 から 1 9 のいずれか一項に記載のメモリ・キャッシュ制御装置と
を備える処理システム。

【請求項 2 3】

メモリ・キャッシュについてコヒーレンシ動作を実行する方法であって、メイン・メモリ(103)と関連した複数のアドレスを備えるアドレス・グループに関するアドレス・グループ指示を受け取るステップを備える、前記コヒーレンシ動作を実行する方法において、

1 群のキャッシュ・ラインの各ラインを順次に処理するステップを備え、

当該処理するステップが、前記 1 群のキャッシュ・ラインの各キャッシュ・ラインに関して、

第 1 のキャッシュ・ラインが前記アドレス・グループのアドレスと関連するかどうかをマッチ判定基準を評価することにより決定するステップと、

前記マッチ判定基準に適合する場合前記第 1 のキャッシュ・ラインについてコヒーレンシ動作を実行するステップと、

競合が前記コヒーレンシ動作と別のメモリ動作との間に存在するかどうかを決定するステップと、

競合が存在する場合、前記のコヒーレンシ手段により前記コヒーレンシ動作を禁止するステップと、を備える

ことを特徴とするメモリ・キャッシュについてコヒーレンシ動作を実行する方法。

【請求項 2 4】

請求項 2 3 記載の方法を実行するようプロセッサを制御するプロセッサ実行可能命令を格納する記憶媒体。

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No PCT/IB2005/051774
A. CLASSIFICATION OF SUBJECT MATTER INV. G06F12/08		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 182 559 A (TEXAS INSTRUMENTS FRANCE ; TEXAS INSTRUMENTS INC (US) 27 February 2002 (2002-02-27) abstract; figures 2,12-19; table 2 paragraphs [0006], [0015], [0016], [0026], [0027], [0034] - [0050], [0058], [0061], [0076] - [0091] ----- -/--	1-14, 16-27
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the International filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed ** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search 21 April 2006		Date of mailing of the international search report 03/05/2006
Name and mailing address of the ISA/ European Patent Office, P.B. 5618 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx. 31 851 epo nl, Fax: (+31-70) 340-3016		Authorized officer Breche, P

INTERNATIONAL SEARCH REPORT

Int'l application No PCT/IB2005/051774

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>US 2002/065980 A1 (TEXAS INSTRUMENTS INC (USA)) 30 May 2002 (2002-05-30)</p> <p>abstract; figures 4,7,9; table 2 paragraphs [0006], [0007], [0010], [0011], [0028], [0041], [0044], [0057], [0061], [0063], [0065], [0066], [0076], [0080] - [0094], [0097] - [0099] paragraphs [0103] - [0105], [0113] - [0117], [0120], [0125]</p>	<p>1,3-14, 16-23, 25-27</p>
X	<p>EP 1 030 243 A (TEXAS INSTRUMENTS FRANCE ; TEXAS INSTRUMENTS INC (US)) 23 August 2000 (2000-08-23)</p> <p>abstract; figures 3-6 paragraphs [0002] - [0009], [0018] - [0040]</p>	<p>1-10,13, 14,18-27</p>
X	<p>EP 1 182 563 A (TEXAS INSTRUMENTS FRANCE) 27 February 2002 (2002-02-27)</p> <p>abstract; figures 1,3,4,10-12; table 2 paragraphs [0014] - [0016], [0024] - [0031], [0055] - [0089], [0103] - [0106], [0110] - [0115]</p>	<p>1-6,9, 10,13, 14,16-27</p>

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No
PCT/IB2005/051774

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 1182559	A	27-02-2002	US 2002073282 A1	13-06-2002
US 2002065980	A1	30-05-2002	NONE	
EP 1030243	A	23-08-2000	DE 69903707 D1	05-12-2002
			DE 69903707 T2	10-07-2003
			JP 2000242556 A	08-09-2000
			US 6606687 B1	12-08-2003
EP 1182563	A	27-02-2002	US 2002069330 A1	06-06-2002

 フロントページの続き

(81) 指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(74) 代理人 100096013

弁理士 富田 博行

(72) 発明者 ベレド, イタイ

イスラエル国 8 4 8 0 7 ピアー - シェバ, ミシヨル・エクロン 2 5

(72) 発明者 アンシエル, モシエ

イスラエル国 4 4 4 1 8 カフル - サベ, シャロム・ハリヘム 2 4

(72) 発明者 エフラト, ヤコブ

イスラエル国 クファー - サバ, ピカット・ベイト・ハネトファ 1 3 エイ

(72) 発明者 エルダー, アロン

イスラエル国 4 3 3 5 4 ラアナナ, ヘルツェル 7 8

F ターム(参考) 5B005 JJ13 MM02 MM03 PP21 TT02 TT03

【要約の続き】

る。