

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6089891号
(P6089891)

(45) 発行日 平成29年3月8日(2017.3.8)

(24) 登録日 平成29年2月17日(2017.2.17)

(51) Int.Cl.

G06F 12/08 (2016.01)

F 1

G06F	12/08	5 3 1 E
G06F	12/08	5 5 1 C
G06F	12/08	5 3 1 B
G06F	12/08	5 7 9

請求項の数 7 (全 27 頁)

(21) 出願番号	特願2013-74711 (P2013-74711)
(22) 出願日	平成25年3月29日 (2013.3.29)
(65) 公開番号	特開2014-199576 (P2014-199576A)
(43) 公開日	平成26年10月23日 (2014.10.23)
審査請求日	平成27年12月4日 (2015.12.4)

(73) 特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番 1号
(74) 代理人	100113608 弁理士 平川 明
(74) 代理人	100105407 弁理士 高田 大輔
(72) 発明者	青柳 隆宏 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内
(72) 発明者	引地 徹 神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】演算処理装置、情報処理装置及び情報処理装置の制御方法

(57) 【特許請求の範囲】

【請求項 1】

他の演算処理装置に接続される演算処理装置において、自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理する演算処理部と、

前記第1のデータを記憶するメモリ部と、

前記第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶するデータ使用状況記憶部と、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第1のデータが他の演算処理装置に持ち出されている状態と前記第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示す指標部とを有し、

自身が実行するアプリケーションごとに前記指標部の動作を有効にする第1のモードまたは前記指標部の動作を無効にする第2のモードを設定する設定部と、

前記設定部が前記第1のモードに設定し、さらに前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略する制御部と、

を有することを特徴とする演算処理装置。

【請求項 2】

前記指標部は、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状

10

20

態を示す基準値からの指標値の増減によって前記遷移を示し、前記第1のデータが他の演算処理装置に持ち出されていない状態から他の演算処理装置に持ち出されている状態に遷移するときに前記指標値をインクリメントし、前記第1のデータが他の演算処理装置に持ち出されている状態から他の演算処理装置に持ち出されていない状態に遷移するときに前記指標値をデクリメントすることを特徴とする請求項1に記載の演算処理装置。

【請求項3】

前記演算処理装置は、前記指標部を複数有し、

前記演算処理装置において、各指標部は、前記メモリ部に記憶されるそれぞれ異なるデータのデータ使用状況について、他の演算処理装置に持ち出されている状態と他の演算処理装置に持ち出されていない状態との間の遷移を示す

10

ことを特徴とする請求項1または請求項2に記載の演算処理装置。

【請求項4】

前記演算処理装置は、前記制御部に所定の信号を供給する信号処理部をさらに有し、

前記制御部は、前記所定の信号に従って前記データ使用状況記憶部を参照し、

前記信号処理部は、前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記制御部に対する前記所定の信号の供給を停止する

ことを特徴とする請求項1から請求項3のいずれか1項に記載の演算処理装置。

【請求項5】

前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部には前記要求された第1のデータに関するデータ使用状況が存在しないことを前記制御部に通知するデータ使用状況通知部をさらに有することを特徴とする請求項1から請求項3のいずれか1項に記載の演算処理装置。

20

【請求項6】

他の演算処理装置と、前記他の演算処理装置に接続される演算処理装置とを有する情報処理装置において、

前記演算処理装置は、

自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理する演算処理部と、

30

前記第1のデータを記憶するメモリ部と、

前記第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶するデータ使用状況記憶部と、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第1のデータが他の演算処理装置に持ち出されている状態と前記第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示す指標部と、自身が実行するアプリケーションごとに前記指標部の動作を有効にする第1のモードまたは前記指標部の動作を無効にする第2のモードを設定する設定部とを有し、前記設定部が前記第1のモードに設定し、さらに前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略する制御部と、

40

を有する

ことを特徴とする情報処理装置。

【請求項7】

他の演算処理装置と、前記他の演算処理装置に接続されるとともに、自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理を行う演算処理部と、前記第1のデータを記憶するメモリ部とを含む演算処理装置とを有する情報処理装置の制御方法において、

前記演算処理装置が有するデータ使用状況記憶部が、前記第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶し、

50

前記演算処理装置が有する指標部が、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第1のデータが他の演算処理装置に持ち出されている状態と前記第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示し、

前記演算処理装置が有する設定部が、自身が実行するアプリケーションごとに前記指標部の動作を有効にする第1のモードまたは前記指標部の動作を無効にする第2のモードを設定し、

前記演算処理装置が有する制御部が、前記設定部が前記第1のモードに設定し、さらに前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略することを特徴とする情報処理装置の制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、演算処理装置、情報処理装置及び情報処理装置の制御方法に関する。

【背景技術】

【0002】

情報処理装置においては、複数の演算コア間でメモリのデータを共有する演算処理装置が実用に供されている。当該演算処理装置においては、演算コアとL1キャッシュの組が複数組集約された演算コア群が形成される。演算コア群に対しては、L2キャッシュ、L2キャッシュ制御部、メモリが接続されている。これら演算コア群、L2キャッシュ、L2キャッシュ制御部、メモリのセットをクラスタと呼ぶ。

【0003】

キャッシュとは、大容量のメモリに記憶されたデータのうち、頻繁に使用するデータを格納する小容量の記憶部である。メモリ内のデータをキャッシュに一時的に格納することにより、時間のかかるメモリへのアクセス頻度を低減する。キャッシュは階層構造を採り、高位層ほど高速であり、低位層ほど大容量である。

【0004】

ディレクトリベースのキャッシュコヒーレンス制御方式では、上記のL2キャッシュには、当該L2キャッシュが属するクラスタの演算コア群が要求したデータが格納されることが多い。そして、各演算コア群は、演算コア群に近いL2キャッシュをより頻繁に使用してデータを取得するように構成されていることが多い。また、データの整合性を保つため、1つのメモリに格納されているデータは当該メモリが属するクラスタによって管理される。また、この方式では、クラスタが、管理対象のメモリ内のデータが現在どのような状態でどのキャッシュに格納されているかを管理する。また、クラスタは、当該メモリに対するデータの要求を受けた場合に、データの状態に基づいてデータ取得要求に対して適切な処理を行う。そして、クラスタは、データ取得要求の処理後、当該データの状態に関する情報を更新する。

【0005】

ここで、特許文献1に示されるように、上記のクラスタ構成及び処理体系を有する演算処理装置において、メモリからデータを取得する際に、データの状態や書き戻し回数をカウンタ形式で管理する方法が提示されている。カウンタはディレクトリRAM内に設定されている。そして、クラスタは、ディレクトリRAMのカウンタを参照してデータ取得処理を行う。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2000-259596号公報

【発明の概要】

【発明が解決しようとする課題】

10

20

30

40

50

【0007】

上記の技術では、クラスタは、ディレクトリRAMの参照結果を待って、その後のデータ取得処理の動作を決定する。このため、ディレクトリRAMの参照処理の分だけデータ取得に伴うレイテンシが発生する。また、ディレクトリRAMの参照処理を行う分、消費電力も大きくなる。

【0008】

本件開示の技術は、上記の事情に鑑みてなされたものであり、その目的とするところは、メモリからのデータ取得に伴うレイテンシを低減し、消費電力を抑えることが可能な演算処理装置、情報処理装置及び情報処理装置の制御方法を提供することである。

【課題を解決するための手段】

10

【0009】

一実施形態による演算処理装置は、他の演算処理装置に接続される演算処理装置において、自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理する演算処理部と、第1のデータを記憶するメモリ部と、第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶するデータ使用状況記憶部と、データ使用状況記憶部に記憶されたデータ使用状況について、第1のデータが他の演算処理装置に持ち出されている状態と第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示す指標部とを有し、指標部が、第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、第1のデータを取得する要求が発生したときに、データ使用状況記憶部に対するデータ使用状況の参照を省略する制御部とを有する。

20

【発明の効果】**【0010】**

一実施形態によれば、メモリからのデータ取得に伴うレイテンシを低減し、消費電力を抑えることが可能な演算処理装置、情報処理装置及び情報処理装置の制御方法を実現できる。

【図面の簡単な説明】**【0011】**

【図1】図1は、比較例に係る情報処理装置における一部のクラスタ構成を示す図である。

30

【図2】図2は、比較例に係るL2キャッシュ制御部の概略の構成を示す図である。

【図3】図3は、比較例に係るクラスタにおいてデータ取得要求が発生した場合の動作を示す図である。

【図4】図4は、図3に示す動作例におけるL2キャッシュ制御部の動作を示す図である。

【図5】図5は、比較例に係るクラスタにおいてデータ取得要求が発生した場合の動作を示す図である。

【図6】図6は、図5に示す動作例におけるL2キャッシュ制御部の動作を示す図である。

【図7】図7は、本実施形態に係る情報処理装置における一部のクラスタ構成の概略を示す図である。

40

【図8】図8は、本実施形態に係るクラスタ内のL2キャッシュ制御部を示す図である。

【図9】図9は、本実施形態において、ディレクトリRAMにおけるエントリの更新処理の概略を示す図である。

【図10】図10は、本実施形態においてコントローラが有する回路を示す図である。

【図11】図11は、本実施形態においてコントローラが有する回路を示す図である。

【図12】図12は、本実施形態に係るクラスタにおいてデータ取得要求が発生した場合の動作を示す図である。

【図13】図13は、図12に示す動作例におけるL2キャッシュ制御部の動作を示す図である。

50

【図14】図14は、図12、13に示す動作例におけるタイミングチャートである。

【図15】図15は、本実施形態に係るコントローラの構成の一例を示す図である。

【図16】図16は、本実施形態に係るコントローラの構成の一例を示す図である。

【図17】図17は、本実施形態に係るカウンタ及びディレクトリRAMの構成の一例を示す図である。

【発明を実施するための形態】

【0012】

最初に、一実施形態に係る情報処理装置の比較例について、図面を参照しながら説明する。

【0013】

(比較例)

図1は、比較例に係る情報処理装置1における一部のクラスタ構成を示す。図1に示すように、クラスタ10は、演算コアとL2キャッシュの組をn組(nは自然数)有する演算コア群100、L2キャッシュ制御部101、メモリ102を有する。L2キャッシュ制御部101はL2キャッシュ103を有する。クラスタ20、30も、クラスタ10と同様、演算コア群200、300、L2キャッシュ制御部201、301、メモリ202、302、L2キャッシュ203、303をそれぞれ有する。

【0014】

以降の説明において、メモリに格納されるデータを要求している演算コアが属するクラスタをローカル(Local)と呼ぶ。また、要求されたデータが格納されているメモリが属するクラスタをホーム(Home)と呼ぶ。さらに、ローカルでないクラスタをリモート(Re mote)と呼ぶ。各クラスタは、データの要求元及び要求先に応じて、ローカルにもホームにもリモートにもなる。また、あるデータ取得要求の処理において、ローカルのクラスタは、ホームのクラスタを兼ねる場合がある。また、リモートのクラスタが、ホームのクラスタを兼ねる場合もある。さらに、ホームのクラスタが管理するホームのメモリに格納されているデータの状態情報をディレクトリ情報と呼ぶ。これらの詳細については後述する。

【0015】

図1に示すように、各クラスタはL2キャッシュ制御部が互いにバスあるいはインターフェクトによって接続されている。情報処理装置1内では、メモリ空間はいわゆるフラットであり、物理アドレスによってどのクラスタに属するメモリにどのデータが格納されているかが一意に決まる。

【0016】

例えば、クラスタ10が、クラスタ10内のメモリ102以外のメモリ202に格納されているデータを取得する場合、そのデータを保持するメモリ202が属するクラスタ20に対してデータの要求を行う。クラスタ20は、該当データの状態をチェックする。ここで、データの状態とは、データがどのクラスタにあるか、データが排他的に使用されているか否か、情報処理装置内におけるデータの同期状況等のデータの使用状況を意味する。そして、クラスタ20は、取得対象のデータが排他的に使用されていなければ、当該データを要求元のクラスタ10に送信する。そして、クラスタ20は、当該データの状態情報として、要求元のクラスタ10がデータを持っていることを記録する。

【0017】

図2は、L2キャッシュ制御部101の概略の構成を示す。L2キャッシュ制御部101は、コントローラ101aとL2キャッシュ103とディレクトリRAM104を備える。また、L2キャッシュ103は、タグRAM103aとデータRAM103bを備える。タグRAM103aは、データRAM103bが保持しているロックのタグ情報を保持する。タグ情報とは、コヒーレンスプロトコル制御における各データの使用状況に関する情報やメモリ内のアドレス等を意味する。ここで、複数のプロセッサを使用するマルチプロセッサ環境においては、プロセッサ間で同一のデータを共有してアクセスする可能性が高い。そこで、マルチプロセッサ環境では、各キャッシュ内に存在するデータの一貫

10

20

30

40

50

性を維持している。プロセッサ間の一貫性を維持するプロトコルをコヒーレンスプロトコルと呼ぶ。このようなプロトコルの一例として、M E S I プロトコルが挙げられる。以下の説明では、データの使用状況をModified、Exclusive、Shared、Invalidの4状態で管理するM E S I プロトコルを使用する。ただし、使用可能なプロトコルはこれに限られない。

【0018】

コントローラ101aは、タグRAM103aを使用して、メモリのブロックがデータRAM103bにどのような状態で存在しているかやデータの有無をチェックする。データRAM103bは、例えばメモリ102内のデータのコピーを保持するRAMである。ディレクトリRAM104は、ホームのクラスタに属するメモリのディレクトリ情報を扱うRAMである。10 ディレクトリ情報は巨大になるため、メモリに格納され、そのキャッシュがRAMに置かれことが多い。しかし、ここでは、ディレクトリRAM104にホームのクラスタに属するメモリのディレクトリ情報が格納されている。

【0019】

コントローラ101aは、演算コア、もしくは、別のクラスタのL2キャッシュ制御部のコントローラからの要求を受け付ける。コントローラ101aは、受け付けた要求内容に応じて、タグRAM103a、データRAM103b、ディレクトリRAM104、メモリ102、他のクラスタに対してそれぞれ動作要求を行う。そして、コントローラ101aは、要求された動作が完了すると、要求元にその結果を返す。

【0020】

図3は、クラスタ10においてデータ取得要求が発生した場合の動作の一例を示す図である。図3では、クラスタ10がローカルのクラスタである。また、クラスタ20がホームのクラスタである。図3では、クラスタ10において、クラスタ20に属するメモリ202に格納されるデータの取得要求が発生した場合を想定している。また、図3は、L2キャッシュ103においてキャッシュミスが発生したときの動作を説明する。なお、ここでは、L2キャッシュ制御部にデータ取得要求が届いた時点でL1キャッシュにおいてキャッシュミスが発生していることを前提として説明する。

【0021】

ローカルであるクラスタ10の演算コア群100の演算コアから、データの要求がL2キャッシュ制御部101に届く。このとき、当該要求には、要求するデータがクラスタ20のメモリ202に格納されるデータであることを示すアドレス情報も含まれる。クラスタ10のL2キャッシュ制御部101は、L2キャッシュ103が該当データを保持していない(miss)ことを確認する。そこで、クラスタ10は、ホームのクラスタであるクラスタ20に対して当該データのデータ取得要求を行う。クラスタ20のL2キャッシュ制御部201が、当該データ取得要求を受けると、L2キャッシュ203のディレクトリ情報をチェックする。L2キャッシュ制御部201のコントローラ201aは、L2キャッシュ203にも、リモートのクラスタ内のL2キャッシュにもデータがないことを確認すると(miss)、メモリ202に対して当該データのデータ取得要求を行う。

【0022】

メモリ202から当該データが返ってくると、L2キャッシュ制御部201は、L2キャッシュ203のディレクトリ情報を更新する。そして、L2キャッシュ制御部201は、データを要求元のローカルのクラスタ10に送る。クラスタ10のL2キャッシュ制御部101は、クラスタ20のL2キャッシュ制御部201から受け取ったデータをL2キャッシュ103に格納する。そして、L2キャッシュ制御部101は、当該データを演算コア群100要求元の演算コアに送る。

【0023】

図4は、図3に示す動作例におけるL2キャッシュ制御部101、201の動作を示す図である。ローカルのクラスタ10内のL2キャッシュ制御部101のコントローラ101aは、演算コア群100の演算コアからデータ取得要求を受け付ける。当該データ取得要求には、演算コアからの要求であることを示す情報とデータ取得要求の種類とメモリの40

10

20

30

40

50

アドレスが含まれる。コントローラ 101a は、要求内容に適切な処理を開始する。

【0024】

コントローラ 101a は、タグ RAM103a に対して、データ取得要求の対象となるデータを含むメモリのブロックのコピーがデータ RAM103b にあるか否かをチェックする。コントローラ 101a は、タグ RAM103a から当該コピーが「ない (miss)」という結果を受け取ると、ホームのクラスタ 20 に属する L2 キャッシュ制御部 201 のコントローラ 201a に対して、当該データのデータ取得要求を行う。

【0025】

コントローラ 201a は、当該データ取得要求を受け付けると、ディレクトリ RAM204 に対して、データ取得要求の対象となるデータがいずれかのクラスタの L2 キャッシュに格納されているか否かチェックする。コントローラ 201a は、ディレクトリ RAM204 から「どのクラスタも持ち出していない (miss)」という結果を受け取ると、メモリ 202 に対して当該データのデータ取得要求を行う。コントローラ 201a は、メモリ 202 から当該データが返ってくると、ディレクトリ RAM204 に、当該データの使用状況について「要求元のクラスタ 10 が持っている」ことを示す情報を登録する。そして、コントローラ 201a は、当該データを要求元のクラスタ 10 のコントローラ 101a に送る。データを受け取ったクラスタ 10 のコントローラ 101a は、当該データの使用状況 (Sharedなど) をタグ RAM103a に格納する。また、コントローラ 101a は、当該データをデータ RAM103b に格納する。そして、コントローラ 101a は、演算コア群 100 内の要求元の演算コアに当該データを送る。

10

20

【0026】

図 5 は、比較例において、図 3 及び図 4 に示す上記の処理が完了した後のクラスタ 20 におけるデータの取得処理の一例を示す図である。図 5 では、上記の説明の通り、クラスタ 10 が L2 キャッシュ 103 に持ち出しているデータに対して、クラスタ 20 の演算コア群 200 がデータ取得を要求する。図 5 における処理では、クラスタ 20 がローカルのクラスタでありホームのクラスタでもある。

【0027】

図 5 に示すように、まず演算コア群 200 が L2 キャッシュ制御部 201 に対してデータを要求する。このとき、当該要求には、要求するデータがクラスタ 20 のメモリ 202 に格納されるデータであることを示すアドレス情報も含まれる。L2 キャッシュ制御部 201 は、当該データが L2 キャッシュ 203 になく (miss)、クラスタ 10 が当該データを持ち出していることを確認する。次いで、L2 キャッシュ制御部 201 は、クラスタ 10 に対して当該データのデータ取得要求を行う。クラスタ 10 の L2 キャッシュ制御部 101 は、当該データ取得要求を受けると、L2 キャッシュ 103 に要求されているデータがあることを確認する。そして、L2 キャッシュ制御部 101 は、L2 キャッシュ 103 から当該データを取得して、当該データをクラスタ 20 に送る。

30

【0028】

L2 キャッシュ制御部 101 から当該データが返ってくると、L2 キャッシュ制御部 201 は、L2 キャッシュ 203 のディレクトリ情報を更新する。そして、L2 キャッシュ制御部 201 は、データを L2 キャッシュ 203 に格納する。次いで、L2 キャッシュ制御部 201 は、当該データを演算コア群 200 の要求元の演算コアに送る。

40

【0029】

図 6 は、図 5 に示す動作例における L2 キャッシュ制御部 101、201 の動作を示す図である。L2 キャッシュ制御部 201 のコントローラ 201a は、演算コア群 200 の演算コアからデータの取得要求を受ける。コントローラ 201a は、タグ RAM203a に対して、当該データがデータ RAM203b にあるか否かを確認する。コントローラ 201a は、当該データがデータ RAM203b にない (miss) ことを確認する。次に、コントローラ 201a は、ディレクトリ RAM204 に対して当該データのディレクトリ情報の読み取りを要求する。コントローラ 201a は、ディレクトリ RAM204 から受け取ったディレクトリ情報により、クラスタ 10 が当該データを持っていることを確認する

50

。そして、コントローラ201aは、コントローラ101aに対して、該当データのデータ取得要求を行う。

【0030】

要求を受け取ったホームのクラスタ10のコントローラ101aは、タグRAM103aに対して、当該データがデータRAM103bにあるか否かを確認する。コントローラ101aは、当該データがデータRAM103bにある(hit)ことを確認する。次に、コントローラ101aは、データRAM103bから当該データを取得する。そして、コントローラ101aは、当該データをクラスタ20のL2キャッシュ制御部201のコントローラ201aに送る。

【0031】

コントローラ201aは、当該データを取得すると、タグRAM203aに対して、当該データがデータRAM203bに格納されていることを示すよう情報の更新を要求する。また、当該データは、クラスタ10のデータRAM103bにも格納されている。そこで、コントローラ201aは、タグRAM203aに対して当該データの使用状況がSharedであることを示すよう情報の更新も要求する。タグRAM203aの情報の更新後、コントローラ201aは、当該データをデータRAM203bに格納する。コントローラ201aは、ディレクトリRAM204に対して当該データがローカルでもあるクラスタ20が持っていることを示すようデータの使用状況の更新を要求する。次に、コントローラ201aは、当該データを演算コア群200の当該データを要求している演算コアに送る。

10

【0032】

上記の情報処理装置において、例えば、各クラスタ内のメモリに格納されるデータに関するディレクトリ情報は、データとともにメモリに格納されるとする。そして、メモリからデータを取得する処理においては、一度のメモリ内のデータ参照処理で該当データとディレクトリ情報を取得できるようにする。そこで、メモリ内の同一ブロックに、データとディレクトリ情報を格納する。ただし、このような構成では、データが他のクラスタに持ち出されている場合、当該他のクラスタからデータを取得する際やデータをメモリに書き込む際に、メモリ内のディレクトリ情報を参照する処理が発生する。そして、クラスタは、当該ディレクトリ情報を確認した後で、各種処理を新たに実行する。このため、情報処理装置において、性能面での劣化、電力面での消費増加、メモリ帯域使用率の情報等が発生する。

20

【0033】

そこで、ディレクトリ情報をクラスタのL2キャッシュ制御部内のディレクトリRAMに格納する。当該ディレクトリRAMには、メモリから取得されたデータに関するディレクトリ情報が格納される。また、L2キャッシュ制御部内のディレクトリRAMにすべてのディレクトリ情報を格納することもある。この構成は、情報処理装置内のクラスタ数が少ない場合やメモリ容量が少ない場合等、ディレクトリ情報も少なく全ディレクトリ情報をL2キャッシュ制御部側に格納できる場合に採用される。

30

【0034】

したがって、上記の比較例の情報処理装置1では、ディレクトリRAMにディレクトリ情報の一部あるいは全体を格納している場合がある。この場合、クラスタは、ディレクトリRAMの参照結果を確認した後でデータ取得要求等の各種処理の詳細を決定する。例えば、データの取得要求が発生したときに、クラスタは、まず、ディレクトリRAMを参照して、当該データがディレクトリキャッシュミス、あるいはリモートのL2キャッシュが当該データを持っていないことを確認する。クラスタは、これらの確認の後で、メモリに対するデータの取得要求を実行する。したがって、この場合は、ディレクトリRAMを参照する処理の分、データ取得に伴うレイテンシが発生し、消費電力も増加する可能性がある。

40

【0035】

また、情報処理装置1内で実行するアプリケーションの実効性能を上げるために、各ク

50

ラスタの演算コア群が使用するデータに関する調整を行うこともある。すなわち、あるアプリケーションの処理を担当する各クラスタにおいて、クラスタ内の演算コア群が当該クラスタ内のメモリに格納されているデータを使用し、他のクラスタ内のメモリのデータを使用しないように調整する。これにより、クラスタのメモリに格納されているデータは、他のクラスタに持ち出されることがない。しかし、このようにアプリケーションを調整した後でも、上記の比較例では、データ取得時にディレクトリRAMの参照を行う。したがって、この場合でも、ディレクトリRAMを参照する処理の分、情報処理装置の性能が低下し、データ取得に伴うレイテンシが発生し、消費電力も増加する可能性がある。

【0036】

そこで、以上の比較例に関する説明を踏まえ、一実施形態に係る情報処理装置の例について、図面を参照しながら以下に説明する。 10

【実施例】

【0037】

図7は、本実施例としての情報処理装置2における一部のクラスタ構成の概略を示す。図7に示すように、情報処理装置2は、比較例と同様、クラスタ50、60、70を有する。なお、クラスタ50、60、70が演算処理装置の一例に相当する。また、ローカル、ホーム、リモートの違いも比較例において説明した通りであり、ここでは説明を省略する。クラスタ50は、演算コア群500、L2キャッシュ制御部501、メモリ502を有する。L2キャッシュ制御部501はL2キャッシュ503を有する。クラスタ60、70も、クラスタ50と同様、演算コア群600、700、L2キャッシュ制御部601、701、メモリ602、702、L2キャッシュ603、703をそれぞれ有する。なお、演算コア群500、600、700が演算処理部の一例に相当する。また、L2キャッシュ制御部501、601、701が制御部の一例に相当する。さらに、メモリ502、602、702がデータ記憶部の一例に相当する。 20

【0038】

また、図7に示すように、情報処理装置2はモードレジスタ80を有する。後述するように、L2キャッシュ制御部501、601、701は、カウンタ501b、601b、701bをそれぞれ有する。また、モードレジスタ80は、各カウンタのカウント処理を制御する。なお、モードレジスタ80が設定部の一例に相当する。また、カウンタ501b、601b、701bが指標部の一例に相当する。 30

【0039】

図7に示すように、各クラスタはL2キャッシュ制御部が互いにバスあるいはインターフェクトによって接続されている。情報処理装置2内では、メモリ空間はいわゆるフラットであり、物理アドレスによってどのクラスタに属するメモリにどのデータが格納されているかが一意に決まる。

【0040】

図8は、クラスタ50のL2キャッシュ制御部501を示す図である。L2キャッシュ制御部501は、コントローラ501aとカウンタ501bとL2キャッシュ503とディレクトリRAM504とを備える。また、L2キャッシュ503は、タグRAM503aとデータRAM503bを有する。また、ディレクトリRAM504がデータ使用状況記憶部の一例に相当する。なお、タグRAM503a、データRAM503b、ディレクトリRAM504は、それぞれ比較例と同様の機能を有するため、ここでは詳細な説明を省略する。 40

【0041】

カウンタ501bは、クラスタ50内のディレクトリRAM504内に格納されているエントリについて、エントリが管理するメモリのブロックのうち、他のクラスタに持ち出されているブロックの数をカウントする。例えば、ディレクトリRAM504のエントリ数を 2^N (Nは整数)とすると、カウンタ501bのビット数はN+1となる。例えば、クラスタ50が自身のメモリ502へのアクセスを処理する際に、カウンタ501bの値が0であるとする。これは、ディレクトリRAM504に「他(リモート)のクラスタに 50

持ち出されていることを示すエントリ」が存在しないことを示す。そこで、クラスタ 50 では、ディレクトリ RAM 504 の参照処理が省略され、メモリ 502 に対するデータ要求が実行される。

【0042】

モードレジスタ 80 は、本実施例に係る情報処理装置 2 内での各クラスタの動作モードを制御する。本実施例では、一例として、動作モードは「モードオン」と「モードオフ」の 2 つのモードを有する。具体的には、「モードオン」とは、各クラスタ内のかウンタの動作を有効にする動作モードである。また、「モードオフ」とは、各クラスタ内のかウンタの動作を無効にする動作モードである。なお、これらの動作モードにおける処理の詳細については後述する。本実施例では、情報処理装置 2 において、情報処理装置 2 の OS (Operating System) の起動前やアプリケーションの実行前に動作モードの切り換えを行う。さらに、本実施例では、情報処理装置 2 の OS がモードレジスタ 80 の動作モードの切り換えを制御する。なお、動作モードの切り換えは、情報処理装置 2 のユーザが明示的に OS に指示をして行ってもよいし、実行するアプリケーションのメモリ使用量等の情報に基づいて OS が自律的に行ってもよい。

10

【0043】

情報処理装置 2 において、クラスタ間通信の多いアプリケーションを実行する場合、カウンタが常に 1 以上になる可能性もある。この場合、カウンタ動作分だけ電力消費量が増加する。また、ディレクトリ RAM の参照処理が省略されないため、メモリからのデータ取得に伴って発生するレイテンシの低減や消費電力の削減がもたらされないことも考えられる。そこで、本実施例では、カウンタを動作させるか否かを設定するモードレジスタ 80 が設けられている。モードレジスタ 80 によってカウンタの動作を無効にする場合、動作モードが「モードオフ」となり、カウンタの動作が無効にされているクラスタの動作は比較例と同じ動作になる。

20

【0044】

また、本実施例においては、コントローラは、ディレクトリ RAM の更新処理時にカウンタのインクリメント又はデクリメントを行う。すなわち、コントローラは、ディレクトリ RAM のエントリ更新時に、更新対象のエントリを読み込んでから該当エントリの更新を行う。そして、コントローラは、更新前後の状態遷移に応じてカウンタのインクリメント又はデクリメントを行う。

30

【0045】

例えば、ディレクトリ RAM 内のあるディレクトリ情報が、更新対象のエントリに対応するデータが他（リモート）のクラスタに持ち出されていない状態を示すとする。このとき、当該ディレクトリ情報が指すエントリが、他のクラスタに持ち出されている状態に遷移するときに、コントローラがカウンタをインクリメントする。また、その逆の場合は、コントローラはカウンタをデクリメントする。すなわち、当該エントリが示すデータが他のクラスタに持ち出されている状態で、当該他のクラスタから当該データが戻される、つまり当該持ち出し状態が取り消される場合に、コントローラはカウンタをデクリメントする。なお、上記の通り、クラスタの動作モードが「モードオン」であり、カウンタの値が 0 である場合に、クラスタに対してデータ取得要求があったときは、ディレクトリ RAM に対する参照処理が省略される。

40

【0046】

図 9 は、本実施例において、カウンタのインクリメント又はデクリメントを行う際の、処理の概略を示す説明図である。図 9 には、ディレクトリ RAM 内のあるインデックスに対応するキャッシュラインを示す。当該キャッシュラインには、更新対象のエントリが含まれる。

【0047】

本実施例において、ディレクトリ RAM 504 のディレクトリ情報が、情報処理装置 2 内の各クラスタがメモリ 502 内のデータを持ち出しているか否かを示す。具体的には、各クラスタのタグ RAM には、他のクラスタから取得したデータについて、Modified、Ex

50

clusive、Shared、Invalidの4つのタイプコードが記憶される。そこで、ディレクトリRAM504には、持ち出し先のクラスタのタグRAMに記憶されたタイプコード等を基にして、情報処理装置2内のクラスタ毎のメモリ502内のデータについてのタイプコードが記憶される。例えば、情報処理装置2がクラスタ50、60、70で構成される場合を考える。この場合、クラスタ50に対してクラスタ60、70がリモートのクラスタになる。

【0048】

ここで、クラスタ70がクラスタ50からメモリ502に格納されるデータを持ち出したときに、ディレクトリRAM504には、リモートのクラスタ70がデータを持ち出したことが記憶される。例えば、クラスタ70が当該データを排他的データ取得により取得した場合は、ディレクトリRAM504に記憶されるタイプコードはExclusiveである。
また、クラスタ70によるデータ取得が排他的でない場合は、ディレクトリRAM504に記憶されるタイプコードはSharedである。また、クラスタ70において持ち出されたデータの内容が変更された場合、ディレクトリRAM504にはModifiedのタイプコードが記憶される。すなわち、メモリ502内のデータが他のクラスタに持ち出されたときに、ディレクトリRAM504に記憶されるタイプコードはInvalid以外である。そして、クラスタ70から当該データがクラスタ50に戻されるとき、ディレクトリRAM504に記憶されるタイプコードはInvalidである。

【0049】

このように、本実施例においては、クラスタ内のメモリに格納されるデータが持ち出されたり戻されたりする際に、そのクラスタ内のディレクトリRAMの更新処理が発生する。そして、ディレクトリRAMを更新する場合、まず、コントローラは、ディレクトリRAMから更新前のエントリの状態を示すデータを読み出す。次に、コントローラは、読み出したエントリの値、すなわち更新前のエントリの状態と、更新後のエントリの状態とを比較する。この比較結果に基づいて、コントローラは、カウンタの値のインクリメント又はデクリメントを行う。

【0050】

図10は、本実施例において、コントローラ501aが有する回路の一部を示す図である。本実施例においては、コントローラ501a、601a、701aが、この論理回路をそれぞれ有する。図10に示す制御回路により、コントローラ501aは、カウンタ501bに対して値のインクリメント及びデクリメントを行う。図10において、ORゲート501cは、更新前に読み出された、更新対象のL2キャッシュ503のエントリに対応するディレクトリ情報において、ローカル以外のクラスタについてタイプコード(TypeCode)に基づくOR演算を行う。このOR演算により、ディレクトリ情報更新前のデータ使用状況が、クラスタ50以外の各クラスタによって更新対象のエントリに対応するデータが持ち出されている状態であるか否かが確認される。

【0051】

ORゲート501cには、更新前のディレクトリ情報に含まれる、ローカル以外の各クラスタのタイプコードがそれぞれ入力される。そして、ORゲート501cは、各入力の少なくとも1つが「TypeCode!=I(Invalid)」を満たす場合、すなわちタイプコードがInvalidではない場合に、「1」を出力する。また、ORゲート501cは、それ以外の場合、すなわち、いずれの入力もタイプコードがInvalidである場合に「0」を出力する。

【0052】

また、ORゲート501dは、上記更新対象のL2キャッシュ503のエントリに対応する、更新後のディレクトリ情報において、ローカル以外のクラスタについてタイプコード(TypeCode)に基づくOR演算を行う。このOR演算により、ディレクトリ情報更新後のデータ使用状況が、クラスタ50以外の各クラスタによって更新対象のエントリに対応するデータが持ち出されている状態であるか否かが確認される。ORゲート501dには、更新後のディレクトリ情報に含まれる、ローカル以外の各クラスタのタイプコードがそれぞれ入力される。そして、ORゲート501dは、各入力の少なくとも1つが「TypeCo

10

20

30

40

50

`de!=l(Invalid)`」を満たす場合に「1」を出力する。また、ORゲート501dは、それ以外の場合に「0」を出力する。

【0053】

ANDゲート501gは、モードレジスタ80により動作モードが「モードオン」にされ、ORゲート501cの出力がインバータ501eで反転されて「1」となり、ORゲート501dの出力が「1」である場合に指示信号CountUpを出力する。カウンタ501bは、この指示信号に従って現在の値をインクリメントする。上述の通り、ここで「モードオン」とは、モードレジスタ80によってカウンタ501bの動作が有効にされている場合をいう。また、ANDゲート501hは、モードレジスタ80により動作モードが「モードオン」にされ、ORゲート501cの出力が「1」であり、ORゲート501dの出力がインバータ501fで反転されて「1」となる場合に指示信号CountDownを出力する。カウンタ501bは、この指示信号に従って現在の値をデクリメントする。
10

【0054】

図10に示すように、ディレクトリRAM504内の更新対象エントリについて、更新前の状態が他のクラスタに持ち出されていたことを示す場合、ORゲート501cから「1」が出力されてANDゲート501hに入力される。このとき、ORゲート501cの出力がインバータ501eにより反転されるため、ANDゲート501gには「0」が入力される。一方、ディレクトリRAM504内の更新対象エントリについて、更新前の状態が他のクラスタに持ち出されていないことを示す場合、ORゲート501cから「0」が出力されてANDゲート501hに入力される。このとき、ANDゲート501gには「1」が入力される。
20

【0055】

また、ディレクトリRAM504内の更新対象エントリについて、更新後の状態が他のクラスタに持ち出されていることを示す場合、ORゲート501dから「1」が出力されてANDゲート501gに入力される。このとき、ORゲート501dの出力がインバータ501fにより反転されるため、ANDゲート501hには「0」が入力される。一方、ディレクトリRAM504内の更新対象エントリについて、更新後の状態が他のクラスタに持ち出されていないことを示す場合、ORゲート501dから「0」が出力されてANDゲート501gに入力される。このとき、ANDゲート501hには「1」が入力される。
30

【0056】

一例として、本実施例において、コントローラ501aが、メモリ502に格納されるデータをクラスタ70に送信する場合を考える。ここでは、当該データは他のクラスタに持ち出されていない、すなわち、当該データはメモリ502又はデータRAM503bに存在する。そして、モードレジスタ80により動作モードが「モードオン」、すなわち、カウンタ501bの動作が有効に設定されているものとする。コントローラ501aは、メモリ502又はデータRAM503bからデータを取得する。そして、コントローラ501aは、ディレクトリRAM504に対して、当該データがリモートのクラスタ70が持っていることを示すよう情報の更新を要求する。したがって、コントローラ501aは、クラスタ70からのデータ取得要求が排他的であるか否かに応じて、クラスタ70による当該データの使用状況がExclusiveあるいはSharedであることを示すよう情報の更新を要求する。
40

【0057】

当該更新処理前において、ディレクトリRAM504内の当該データに関するディレクトリ情報は、当該データが他のいずれのクラスタにも持ち出されていないことを示す。すなわち、リモートのクラスタに関する当該データのタイプコードは、いずれもInvalidである。そのため、ORゲート501cは「0」を出力する。一方、更新後のディレクトリ情報は、当該データがリモートのクラスタ70に持ち出されていることを示す。すなわち、リモートのクラスタ70に関する当該データのタイプコードはSharedあるいはExclusiveである。したがって、ORゲート501dは「1」を出力する。
50

【0058】

このため、ANDゲート501gにはインバータ501e及びORゲート501dからともに「1」が入力される。さらに、モードレジスタ80によりカウンタ501bの動作モードが「モードオン」である。したがって、ANDゲート501gは、指示信号CountUpを出力する。そして、指示信号CountUpに従って、カウンタ501bの値がインクリメントされる。一方、ANDゲート501hには、ORゲート501c及びインバータ501fからともに「0」が入力される。このため、ANDゲート501hから指示信号CountDownは出力されない。

【0059】

次に、当該データがクラスタ70から戻された場合を考える。なお、ここでもモードレジスタ80により、カウンタ501bの動作が有効にされているものとする。コントローラ501aは、クラスタ70から当該データを取得し、ディレクトリRAM504に対して当該データがリモートのクラスタ70に持ち出されていないことを示すよう情報の更新を要求する。すなわち、コントローラ501aは、ディレクトリRAM504に対して、クラスタ70に関する当該データのタイプコードをInvalidにするよう要求する。10

【0060】

当該更新処理前において、ディレクトリRAM504内の当該データに関するディレクトリ情報は、クラスタ70に持ち出されていることを示す。すなわち、クラスタ70に関する当該データのタイプコードはInvalid以外の値である。そのため、ORゲート501cは「1」を出力する。一方、更新後のディレクトリ情報は、当該データが他のいずれのクラスタにも持ち出されていないことを示す。すなわち、リモートのクラスタに関する当該データのタイプコードは、いずれもInvalidである。したがって、ORゲート501dは「0」を出力する。20

【0061】

このため、ANDゲート501hにはORゲート501c及びインバータ501fからともに「1」が入力される。さらに、モードレジスタ80によりカウンタ501bの動作モードが「モードオン」である。したがって、ANDゲート501hは、指示信号CountDownを出力する。そして、指示信号CountDownに従って、カウンタ501bの値がデクリメントされる。一方、ANDゲート501gには、インバータ501e及びORゲート501dからともに「0」が入力される。このため、ANDゲート501gから指示信号CountUpは出力されない。30

【0062】

このように、本実施例においては、図10に示す制御回路により、ディレクトリRAM504内の更新対象エントリについて、更新前の状態と更新後の状態とが比較され、カウンタ501bの値のインクリメント又はデクリメントを行う処理が実現される。

【0063】

次に、図11に、カウンタの値が0である場合に、ディレクトリRAMの参照処理を省略して、メモリへの参照処理を行う制御を実行する論理回路を示す。本実施例においては、コントローラ501a、601a、701aが、この論理回路をそれぞれ有する。40

【0064】

図11において、ANDゲート501iは、モードレジスタ80によりクラスタ50が「モードオン」に設定され、カウンタ501bの値が0であり、クラスタ50に対するデータ取得要求が発生した場合に「1」を出力する。上述の通り、ここで「モードオン」とは、モードレジスタ80によってカウンタ501bの動作が有効とされている場合をいう。ANDゲート501iの出力は、ORゲート501jに入力される。また、ANDゲート501iの出力は、インバータ501kにより反転されて、ANDゲート501lに入力される。ORゲート501jは、ANDゲート501iの出力が「1」である場合に、メモリへのアクセスを実行する指示信号(LocalMemoryAccess2)を出力する。また、ORゲート501jは、上記の比較例に従ってメモリ502へのアクセスが発生した場合にも、メモリへのアクセスを実行する指示信号(LocalMemoryAccess2)を出力する。50

【0065】

ANDゲート5011は、ANDゲート501iの出力が「0」であり、上記の比較例に従ってディレクトリRAM504へのアクセスが発生した場合に、ディレクトリRAM504へのアクセスを行う指示信号(DirectoryRAMAccess2)を出力する。したがって、本実施例においては、クラスタ50の動作モードが「モードオフ」である場合は、上記の比較例に従って、ディレクトリRAM504へのアクセスやメモリ502へのアクセスが実行される。また、クラスタ50の動作モードが「モードオン」である場合は、クラスタ50に対してデータ取得処理が発生したときに、カウンタ501bの値が0であると、ディレクトリRAM504へのアクセスが発生しない。そして、メモリ502へのアクセスが発生し、データ取得要求により要求されているデータがメモリ502から取得される。

10

【0066】

また、クラスタの自身のL2キャッシュにデータが格納されている場合は、自身のメモリに格納されるデータを取得する際に、L2キャッシュにおいてキャッシュヒットが発生する。このため、当該データがL2キャッシュから取得されて演算コア群に送られる。以上より、L2キャッシュにおいてキャッシュミスが発生したときにカウンタが0であった場合、該当データが当該L2キャッシュ以外に存在しないという状態は生じない。

【0067】

また、上記の通り、本実施例においては、ディレクトリRAMの更新処理時に、コントローラは、カウンタのインクリメント処理又はデクリメント処理を行う。また、比較例においては、通常、プロトコルの妥当性確認等のため、ディレクトリRAMの更新処理時に、コントローラは、更新対象のエントリを読み込んでディレクトリ情報を確認する。したがって、本実施例のカウンタの構成を採用しても、比較例に比べてディレクトリRAMの参照回数が増えることはない。

20

【0068】

次に、図12は、本実施例において、クラスタ50がデータ取得を行う場合の動作の一例を示す図である。図12においては、カウンタ501bの値が「0」である場合の処理について説明する。すなわち、図12においては、メモリ502に格納されるデータは、データRAM503bではなく、他のクラスタにも持ち出されていない。したがって、当該データはメモリ502に格納されている。なお、カウンタ501bの値が「0」でない場合は、上記の比較例と同様に、ディレクトリRAM504のディレクトリ情報を参照して各種処理が実行される。また、図12においては、モードレジスタ80によって動作モードが「モードオン」に設定されている、すなわちカウンタ501bの動作が有効になっている。動作モードが「モードオフ」に設定されている場合は、カウンタ501bの動作が無効となり、クラスタ50は、上記の比較例と同様の処理を実行する。

30

【0069】

図13は、図12に示す動作例におけるL2キャッシュ制御部501の動作を示す図である。上記の通り、L2キャッシュ制御部501は、コントローラ501aとカウンタ501bとL2キャッシュ503とディレクトリRAM504をそれぞれ備える。また、L2キャッシュ503は、タグRAM503aとデータRAM503bをそれぞれ備える。

40

【0070】

図13に示すように、クラスタ50において、コントローラ501aが演算コア群500から、メモリ502に格納されるデータのデータ要求を受信する。次に、コントローラ501aは、タグRAM503aを参照し、要求されたデータがデータRAM503bに格納されているか否かを確認する。コントローラ501aは、データRAM503bに当該データがない(キャッシュミス)ことを確認すると、カウンタ501bの値を確認する。コントローラ501aは、カウンタ501bの値が「0」であることを確認すると、図11に示す制御回路の動作により、ディレクトリRAM504の参照処理を省略する。そして、コントローラ501aは、メモリ502から要求されているデータを取得する。コントローラ501aは、メモリ502から当該データを取得すると、ディレクトリRAM504に対して、クラスタ50がデータを持っていることを示す情報を記憶するよう要

50

求する。また、コントローラ501aは、データRAM503aに当該データがデータRAM503bに格納されていることを示す情報を記憶するよう要求する。さらに、コントローラ501aは、当該データをデータRAM503bに格納する。そして、コントローラ501aは、メモリ502から取得したデータを演算コア群500に送信する。

【0071】

図14は、図12、13に示す動作例におけるL2キャッシュ制御部501のタイミングチャートである。以下の説明において、チャート内のステップをSと略記する。S101において、コントローラ501aは、演算コア群500からメモリ502に格納されるデータのデータ要求を受信する。当該データ要求には、要求するデータの格納先を示すアドレス、すなわちメモリ502内の当該データの格納先を示すアドレスも含まれる。S102において、コントローラ501aは、タグRAM503aに対して、当該アドレスがデータ格納先であるデータがデータRAM503bに存在するか否かを確認する。S103において、タグRAM503aは、当該データはデータRAM503bにない(miss)ことをコントローラ501aに通知する。
10

【0072】

次に、S104において、コントローラ501aは、カウンタ501bの値を確認する。ここではカウンタ501bの値が0であるため、ディレクトリRAM504内に、当該データが他のクラスタに持ち出されたことを示すエントリは存在しない。したがって、コントローラ501aは、当該データの取得にあたり、ディレクトリRAM504のディレクトリ情報を確認する必要はない。そこで、コントローラ501aは、ディレクトリRAM504の参照処理を省略し、以下に説明するように、すぐにメモリ502に対してデータ取得要求を行うことができる。
20

【0073】

S105において、コントローラ501aは、図11に示す制御回路の動作により、メモリ502に対して当該データを要求する。S106において、メモリ502は要求されたデータをコントローラ501aに送信する。S107において、コントローラ501aは、タグRAM503aに対して、データRAM503bに当該データが存在することを示すよう情報の更新を要求する。また、コントローラ501aは、タグRAM503aに対して、当該データの使用状況がSharedであることを示すよう情報の更新を要求する。S108において、タグRAM503aは、コントローラ501aからの要求に従って情報を更新した後、当該更新処理が完了したことをコントローラ501aに通知する。
30

【0074】

S109において、コントローラ501aは、データRAM503bに対して、S105においてメモリ502から取得したデータを送信し、当該データを格納するよう要求する。S110において、データRAM503bは当該データの格納処理を行った後、当該格納処理が完了したことをコントローラ501aに通知する。S111において、コントローラ501aは、ディレクトリRAM504に対して、当該データはクラスタ50が持っていることを示すよう(Value=+Local)情報の更新を要求する。上記の通り、S103において当該データについてはキャッシュミスが発生している。また、S105においてカウンタ501bの値が0である。したがって、当該データは他のクラスタに持ち出されてもいない。そこで、ディレクトリRAM504に当該データのディレクトリ情報を示すエントリを新たに追加する。S112において、ディレクトリRAM504は、コントローラ501aからの要求に従って情報を更新した後、当該更新処理が完了したことをコントローラ501aに通知する。S113において、コントローラ501aは、当該データを演算コア群500に送信する。
40

【0075】

このように、本実施例においては、カウンタ501bの値が0であるため、ディレクトリRAM504内に他のクラスタがデータを持ち出したことを示すエントリは存在しない。さらに、当該データはデータRAM503bにも存在しないことも確認される。すなわち、当該データはメモリ502以外に存在しないことが確定する。本実施例では、このよ
50

うな場合に、コントローラ 501a が、ディレクトリ RAM 504 の参照処理を省略してメモリ 502 に対するアクセスを行う。したがって、情報処理装置 2 において、メモリ 502 に対するアクセス処理に伴うレイテンシを抑えるとともに、消費電力を削減することができる。

【0076】

また、カウンタ 501b の値のカウントに用いるビット数は、数ビットから多くとも數十ビットで収まると考えられる。すなわち、カウンタ 501b に用いる容量は、ディレクトリ RAM 504 の容量に比べて小さい容量で構成することができる。したがって、カウンタ 501b の参照処理に伴う情報処理量は、比較例において発生するディレクトリ RAM 504 の参照処理に伴う情報処理量に比べて小さいと考えられる。また、カウンタ 501b を設けることによって生じる消費電力量が、ディレクトリ RAM 504 の参照処理を省略することにより実現される消費電力の削減量を上回らないとも言える。10

【0077】

さらに、情報処理装置 2 において実行するアプリケーションに対して、クラスタ間通信が極力発生しないように調整を行うと、メモリに格納されたデータが他のクラスタに持ち出される頻度が低下する。したがって、カウンタの値が 0 になる可能性が高まる。このため、ディレクトリ RAM の参照処理が省略される回数が増加する。その結果、アクセス処理に伴うレイテンシを抑え、消費電力を削減し、情報処理装置 2 の性能を向上させる効果がより高まると考えられる。例えば、情報処理装置 2 において、各クラスタが自身のメモリのデータを使用するフェイズと、各クラスタがクラスタ間通信を行うフェイズを切り分けて実行するようアプリケーションを調整してもよい。これにより、特に、各クラスタが自身のメモリのデータを使用するフェイズにおいて、上記の効果が得られると考えられる。20

【0078】

ところで、本実施形態においてディレクトリ RAM は、メモリ又はデータ RAM 内に格納されている各データについて、各クラスタにおけるタイプコードを、各クラスタに対応するビットによって管理する。例えば、上記の説明では Modified、Exclusive、Shared、Invalid の 4 つのタイプコードが存在する。そこで、一例として、4 つのタイプコード Modified、Exclusive、Shared、Invalid に対して、「00」、「01」、「10」、「11」のビットをそれぞれ割り当て、各クラスタによる各データの持ち出し状況を管理する。ただし、ディレクトリ RAM におけるクラスタのデータの持ち出し状況を管理する構成は、上記に限られない。30

【0079】

以上が本実施形態に関する説明であるが、上記の情報処理装置の構成や処理は、上記の実施形態に限定されるものではなく、本発明の技術的思想と同一性を失わない範囲内において種々の変更が可能である。例えば、上記の説明では、モードレジスタ 80 がクラスタ 50、60、70 の外部に配置されているが、各クラスタの内部にモードレジスタを 1 つずつ設ける構成としてもよい。また、上記の説明では、L2 キャッシュ制御部内のディレクトリ RAM に、メモリに格納されているすべてのデータのディレクトリ情報が格納されている場合を想定したが、上記の実施形態は、そのような場合に限らず適用することができる。例えば、メモリにすべてのディレクトリ情報が格納されており、当該ディレクトリ情報のキャッシュがディレクトリ RAM に保持される構成としても、上記の実施形態の構成を適用することができる。40

【0080】

また、上記の実施例の情報処理装置 2 は、モードレジスタ 80 を備えないものであってもよい。例えば、情報処理装置 2 において、クラスタ間通信が比較的少ないアプリケーションを実行する場合には、メモリのデータが他のクラスタに持ち出されることが少ないと予想される。このようなアプリケーションを比較的多く実行する情報処理装置では、モードレジスタ 80 を省略してもよい。モードレジスタ 80 がない情報処理装置では、コントローラは、カウンタの値に従ってディレクトリ RAM を参照するか否かを判定すればよい50

。

【0081】

また、カウンタのビット数をできるだけ増やさない方法としては、ディレクトリRAM内の複数エントリを1まとまりとして、複数エントリ単位でカウントするようコントローラがカウンタを制御するという方法がある。具体的には、複数エントリの全てが持ち出されていない状態から複数エントリのうちいずれか1つでも持ち出された状態に遷移したときに、コントローラは、カウンタをインクリメントする。そして、その逆の遷移の場合は、コントローラは、カウンタをデクリメントする。なお、ディレクトリRAMから一度に読み出されるエントリ群とカウント処理における複数エントリ単位とを対応付けると、コントローラは、カウント処理を効率よく制御することができる。

10

【0082】

そして、ディレクトリRAMの参照処理を省略する方法として、コントローラがディレクトリRAMに対するクロックの供給を止める方法もある。図15は、上記の実施例において、コントローラ501aに追加することが可能な制御回路の一例である。図15に示すように、NANDゲート501mは、カウンタの値が0であり(Counter==0)、かつ、演算コア群500からデータ要求があった(RequestFrom==Local)場合に、「0」を出力する。このため、ディレクトリRAM504に対するクロックが発生しても、ANDゲート501nの出力が「0」となり、ディレクトリRAM504にはクロックは供給されない。なお、上記以外の場合は、NANDゲート501mは「1」を出力するため、ディレクトリRAM504にはクロックが供給される。なお、クロックが所定の信号の一例に相当する。また、NANDゲート501mとANDゲート501nが信号処理部の一例に相当する。

20

【0083】

あるいは、ディレクトリRAM504が備えるEnable信号を用いる方法もある。すなわち、上記の実施例において、コントローラ501aを、ディレクトリRAM504の参照結果を確認した後に各種動作を決定するように構成している場合は、セレクタを用いることができる。図16にその構成の一例を示す。図16に示す制御回路は、例えばディレクトリRAM504に設けることができる。カウンタの値が0のとき、NANDゲート501oは、カウンタの値が0であり(Counter==0)、かつ、演算コア群500からデータ要求があった(RequestFrom==Local)場合に、「0」を出力する。そして、NANDゲート501oの出力が「0」である場合、セレクタ501pは、Miss、Invalidを示す結果をコントローラ501a返す。この結果は、実質的に、ディレクトリRAM504において要求されているデータに関するディレクトリ情報がないことを意味する。そして、コントローラ501aは、当該結果を受け取ると、メモリ502に対するデータ要求を行う。このため、実質的に、コントローラ501aはディレクトリRAM504のディレクトリ情報の参照処理を省略して、メモリ502からデータを取得する。なお、NANDゲート501oとセレクタ501pがデータ使用状況通知部の一例に相当する。

30

【0084】

以上より、図15又は図16に示す制御回路をクラスタ内に設けることでも、上記の実施例と同様にディレクトリRAM504の参照処理を省略することができる。

40

【0085】

また、情報処理装置内でクラスタ間通信が頻繁に発生する場合、1つのディレクトリRAMに対し1つのカウンタを設ける構成においては、カウンタが0になりにくい可能性がある。そこで、1つのディレクトリRAMに対応するカウンタを複数設けてよい。例えば、高度科学技術計算(HPC:High Performance Computing)等では、各クラスタ内で使用されるデータとクラスタ間通信に使用されるデータとが明確に切り分けられることがある。このような場合に、情報処理装置のOS等と協調し、前者のデータに関する情報をディレクトリRAMの物理メモリ空間の前半アドレスに割り当てる。また、後者のデータに関する情報を、ディレクトリRAMの物理メモリ空間の後半アドレスに割り当てる。そして、カウンタを物理アドレスの前半部分と後半部分とでそれぞれ1つずつ設ける。これ

50

により、1つのディレクトリRAMに対して1つのカウンタを設ける場合に比べて、各クラスタ内で使用されるデータを使用する場合に、カウンタの値が0になる可能性を高めることができる。

【0086】

図17に、一例として、1つのディレクトリRAM804に対して2つのカウンタ801b、901bを設ける場合の概略の構成を示す。ディレクトリRAM804は、ディレクトリRAM804が属するクラスタ(図示せず)のメモリ(図示せず)の物理アドレス空間の前半部分のディレクトリ情報(前半アドレス)を管理する領域を有する。さらに、ディレクトリRAM804は、当該メモリの物理アドレス空間の後半部分のディレクトリ情報(後半アドレス)を管理する別の領域を有する。そして、クラスタ内のコントローラ(図示せず)は、カウンタ801bを制御して、ディレクトリRAM804の上記の前半アドレスに対応するディレクトリ情報を対象にカウント処理を行う。また、コントローラは、カウンタ901bを制御して、ディレクトリRAM804の上記の後半アドレスに対応するディレクトリ情報を対象にカウント処理を行う。なお、カウンタ801b、901bの値のインクリメント処理及びデクリメント処理は、上記の実施例と同様であるため、詳細な説明は省略する。

【0087】

ここで、当該メモリの物理アドレス空間の前半部分に格納されたデータに対するデータ要求が発生したとする。この場合、上記クラスタのコントローラは、カウンタ801bの値を確認する。そして、コントローラは、カウンタ801bの値が0である場合に、上記と同様に、ディレクトリRAM804の参照処理を省略し、メモリからデータを取得する。したがって、この構成では、カウンタ901bの値が0でなくてもディレクトリRAM804の参照処理が省略される。このため、上記の実施例の場合よりもディレクトリRAMの参照処理が省略される可能性が高まると考えられる。

【0088】

なお、上記の実施例ではディレクトリRAMの参照処理を省略することを念頭に説明した。ただし、各クラスタがディレクトリRAMを有さず、かつ、ディレクトリ情報がすべてメモリに格納されており、かつ、データとそのディレクトリ情報が別個のエントリに格納されている場合にも上記の構成を適用できる。すなわち、この場合、上記のカウンタをメモリ用に設け、コントローラがメモリ内のディレクトリ情報を参照してからデータ取得を行う際に、カウンタが上記のように制御される構成とする。これにより、コントローラは、カウンタの値に基づいてメモリ内のディレクトリ情報の参照を省略し、データ取得をより迅速かつ効率よく行うことができる。

【0089】

《コンピュータが読み取り可能な記録媒体》

コンピュータその他の機械、装置(以下、コンピュータ等)に上記情報処理装置の設定を行うための管理ツール、OSその他を実現させるプログラムをコンピュータ等が読み取り可能な記録媒体に記録することができる。ここで、設定とは、例えばレジスタの設定等を意味する。そして、コンピュータ等に、この記録媒体のプログラムを読み込ませて実行させることにより、その機能を提供させることができる。ここで、コンピュータは、例えば、クラスタやコントローラ等である。

【0090】

ここで、コンピュータ等が読み取り可能な記録媒体とは、データやプログラム等の情報を電気的、磁気的、光学的、機械的、または化学的作用によって蓄積し、コンピュータ等から読み取ることができる記録媒体をいう。このような記録媒体のうちコンピュータ等から取り外し可能なものとしては、例えばフレキシブルディスク、光磁気ディスク、CD-ROM、CD-R/W、DVD、ブルーレイディスク、DAT、8mmテープ、フラッシュユメモリ等のメモリカード等がある。また、コンピュータ等に固定された記録媒体としてハードディスクやROM等がある。

【0091】

10

20

30

40

50

以上の実施形態に関し、さらに以下の付記を開示する。

【0092】

(付記1)

他の演算処理装置に接続される演算処理装置において、自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理する演算処理部と、

前記第1のデータを記憶するメモリ部と、

前記第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶するデータ使用状況記憶部と、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第1のデータが他の演算処理装置に持ち出されている状態と前記第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示す指標部とを有し、前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略する制御部と、

を有することを特徴とする演算処理装置。

【0093】

(付記2)

前記演算処理装置が、前記指標部の動作を動作状態に設定する設定部をさらに有し、

前記設定部によって前記指標部が動作状態に設定されている場合に、前記指標部は前記遷移を示す

ことを特徴とする付記1に記載の演算処理装置。

【0094】

(付記3)

前記指標部は、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す基準値からの指標値の増減によって前記遷移を示し、前記第1のデータが他の演算処理装置に持ち出されていない状態から他の演算処理装置に持ち出されている状態に遷移するときに前記指標値をインクリメントし、前記第1のデータが他の演算処理装置に持ち出されている状態から他の演算処理装置に持ち出されていない状態に遷移するときに前記指標値をデクリメントすることを特徴とする付記1又は付記2に記載の演算処理装置。

【0095】

(付記4)

前記演算処理装置は、前記指標部を複数有し、

前記演算処理装置において、各指標部は、前記メモリ部に記憶されるそれぞれ異なるデータのデータ使用状況について、他の演算処理装置に持ち出されている状態と他の演算処理装置に持ち出されていない状態との間の遷移を示す

ことを特徴とする付記1から付記3のいずれかに記載の演算処理装置。

【0096】

(付記5)

前記演算処理装置は、前記制御部に所定の信号を供給する信号処理部をさらに有し、

前記制御部は、前記所定の信号に従って前記データ使用状況記憶部を参照し、

前記信号処理部は、前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記制御部に対する前記所定の信号の供給を停止する

ことを特徴とする付記1から付記4のいずれかに記載の演算処理装置。

【0097】

(付記6)

前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部には前記要求された第1のデータに関するデータ使用状況が存在しないことを前記制御部に通知するデータ使用状況通知部をさらに有することを特徴とする付記1から付

10

20

30

40

50

記 4 のいずれかに記載の演算処理装置。

【 0 0 9 8 】

(付記 7)

他の演算処理装置と、前記他の演算処理装置に接続される演算処理装置とを有する情報処理装置において、

前記演算処理装置は、

自身が管理する第 1 のデータと他の演算処理装置から取得した第 2 のデータとを用いて演算処理する演算処理部と、

前記第 1 のデータを記憶するメモリ部と、

前記第 1 のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶するデータ使用状況記憶部と、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第 1 のデータが他の演算処理装置に持ち出されている状態と前記第 1 のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示す指標部とを有し、前記指標部が、前記第 1 のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第 1 のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略する制御部と、

を有する

ことを特徴とする情報処理装置。

【 0 0 9 9 】

(付記 8)

20

前記演算処理装置が、前記指標部の動作を動作状態に設定する設定部をさらに有し、

前記設定部によって前記指標部が動作状態に設定されている場合に、前記指標部は前記遷移を示す

ことを特徴とする付記 7 に記載の情報処理装置。

【 0 1 0 0 】

(付記 9)

30

前記指標部は、前記第 1 のデータがいずれも他の演算処理装置に持ち出されていない状態を示す基準値からの指標値の増減によって前記遷移を示し、前記第 1 のデータが他の演算処理装置に持ち出されていない状態から他の演算処理装置に持ち出されている状態に遷移するときに前記指標値をインクリメントし、前記第 1 のデータが他の演算処理装置に持ち出されている状態から他の演算処理装置に持ち出されていない状態に遷移するときに前記指標値をデクリメントすることを特徴とする付記 7 又は付記 8 に記載の情報処理装置。

【 0 1 0 1 】

(付記 1 0)

前記演算処理装置は、前記指標部を複数有し、

前記演算処理装置において、各指標部は、前記メモリ部に記憶されるそれぞれ異なるデータのデータ使用状況について、他の演算処理装置に持ち出されている状態と他の演算処理装置に持ち出されていない状態との間の遷移を示す

ことを特徴とする付記 7 から付記 9 のいずれかに記載の情報処理装置。

【 0 1 0 2 】

(付記 1 1)

40

前記演算処理装置は、前記制御部に所定の信号を供給する信号処理部をさらに有し、前記制御部は、前記所定の信号に従って前記データ使用状況記憶部を参照し、

前記信号処理部は、前記指標部が、前記第 1 のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第 1 のデータを取得する要求が発生したときに、前記制御部に対する前記所定の信号の供給を停止する

ことを特徴とする付記 7 から付記 1 0 のいずれかに記載の情報処理装置。

【 0 1 0 3 】

(付記 1 2)

前記指標部が、前記第 1 のデータがいずれも他の演算処理装置に持ち出されていない状

50

態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部には前記要求された第1のデータに関するデータ使用状況が存在しないことを前記制御部に通知するデータ使用状況通知部をさらに有することを特徴とする付記7から付記10のいずれかに記載の情報処理装置。

【0104】

(付記13)

他の演算処理装置と、前記他の演算処理装置に接続されるとともに、自身が管理する第1のデータと他の演算処理装置から取得した第2のデータとを用いて演算処理を行う演算処理部と、前記第1のデータを記憶するメモリ部とを含む演算処理装置とを有する情報処理装置の制御方法において、10

前記演算処理装置が有するデータ使用状況記憶部が、前記第1のデータが他の演算処理装置に持ち出されているか否かを示すデータ使用状況を記憶し、

前記演算処理装置が有する指標部が、前記データ使用状況記憶部に記憶されたデータ使用状況について、前記第1のデータが他の演算処理装置に持ち出されている状態と前記第1のデータが他の演算処理装置に持ち出されていない状態との間の遷移を示し、

前記演算処理装置が有する制御部が、前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部に対するデータ使用状況の参照を省略することを特徴とする情報処理装置の制御方法。20

【0105】

(付記14)

前記演算処理装置が有する設定部が、前記指標部の動作を動作状態に設定し、

前記設定部が前記指標部を動作状態に設定した場合に、前記指標部は前記遷移を示すことを特徴とする付記13に記載の情報処理装置の制御方法。

【0106】

(付記15)

前記指標部は、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す基準値からの指標値の増減によって前記遷移を示し、前記第1のデータが他の演算処理装置に持ち出されていない状態から他の演算処理装置に持ち出されている状態に遷移するときに前記指標値をインクリメントし、前記第1のデータが他の演算処理装置に持ち出されている状態から他の演算処理装置に持ち出されていない状態に遷移するときに前記指標値をデクリメントすることを特徴とする付記13又は付記14に記載の情報処理装置の制御方法。30

【0107】

(付記16)

前記演算処理装置は、前記指標部を複数有し、

前記演算処理装置において、各指標部は、前記メモリ部に記憶されるそれぞれ異なるデータのデータ使用状況について、他の演算処理装置に持ち出されている状態と他の演算処理装置に持ち出されていない状態との間の遷移を示す

ことを特徴とする付記13から付記15のいずれかに記載の情報処理装置の制御方法。40

【0108】

(付記17)

前記演算処理装置が有する信号処理部が、前記制御部に所定の信号を供給し、

前記制御部は、前記所定の信号に従って前記データ使用状況記憶部を参照し、

前記信号処理部は、前記指標部が、前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記制御部に対する前記所定の信号の供給を停止する

ことを特徴とする付記13から付記16のいずれかに記載の情報処理装置の制御方法。

【0109】

(付記18)

10

20

30

40

50

前記演算処理装置が有するデータ使用状況通知部が、前記指標部が前記第1のデータがいずれも他の演算処理装置に持ち出されていない状態を示す場合に、前記第1のデータを取得する要求が発生したときに、前記データ使用状況記憶部には該データに関するデータ使用状況が存在しないことを前記記憶制御部に通知することを特徴とする付記13から付記16のいずれかに記載の情報処理装置の制御方法。

【符号の説明】

【0110】

1、2 情報処理装置

10、20、30、50、60、70 クラスタ

80 モードレジスタ

10

100、200、300、500、600、700 演算コア群

101、201、301、501、601、701 L2キャッシュ制御部

102、202、302、502、602、702 メモリ

103、203、303、503、603、703 L2キャッシュ

101a、201a、301a、501a、601a コントローラ

501b、801b、901b カウンタ

103a、203a、503a、603a タグRAM

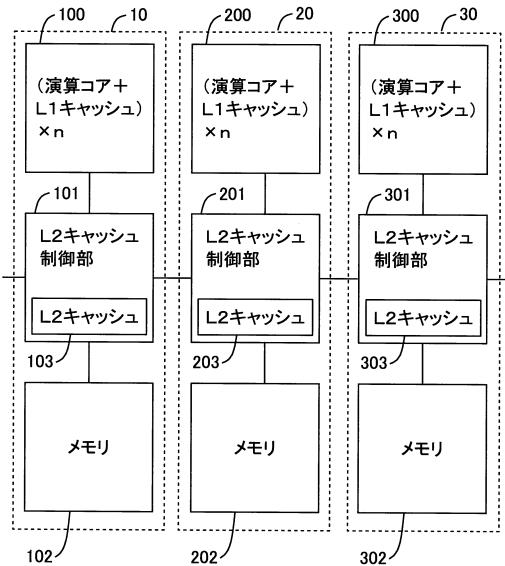
103b、203b、503b、603b データRAM

104、204、504、604、804 ディレクトリRAM

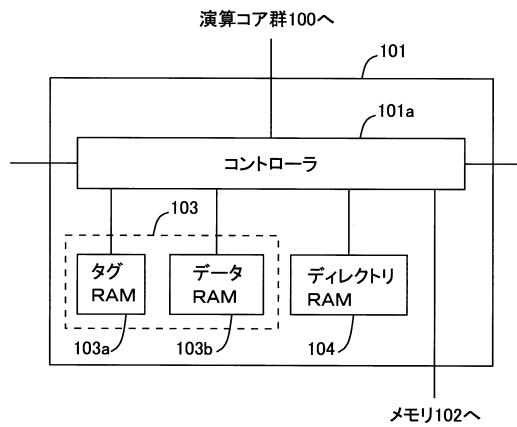
501p セレクタ

20

【図1】

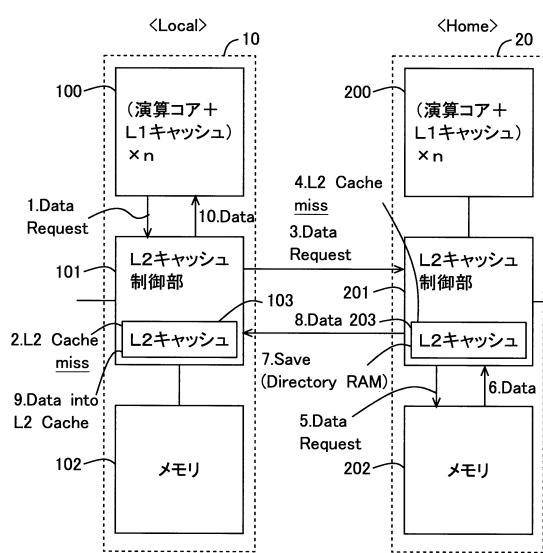


【図2】

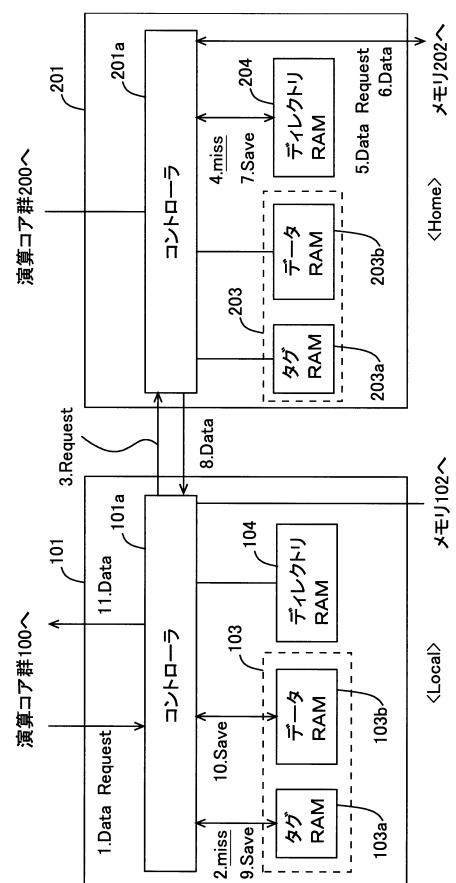


1

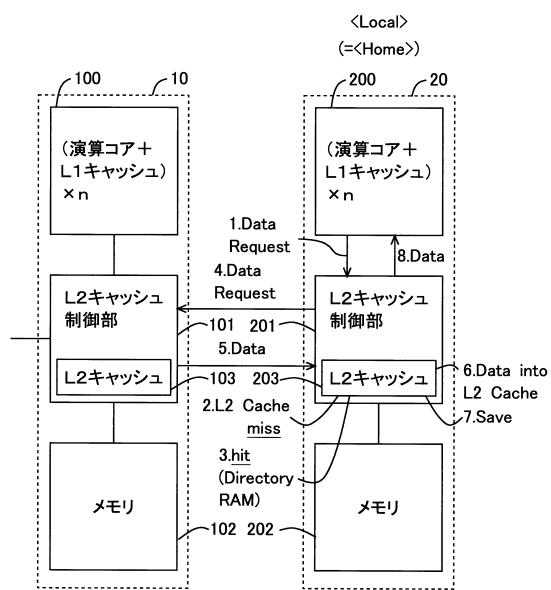
【図3】



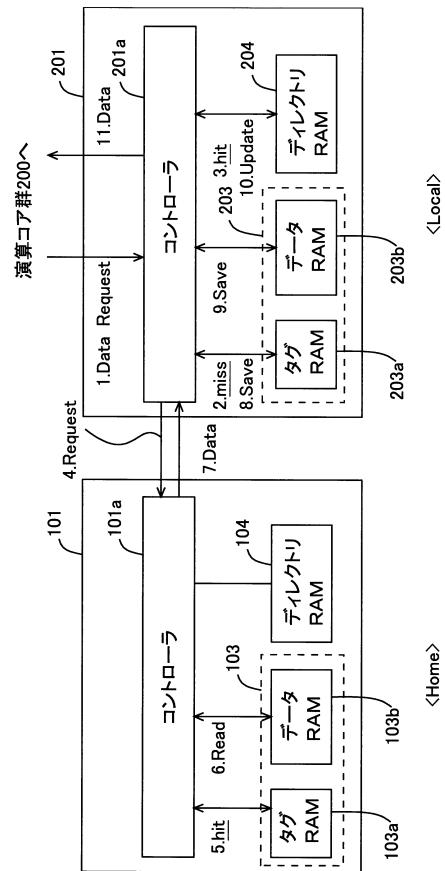
【図4】



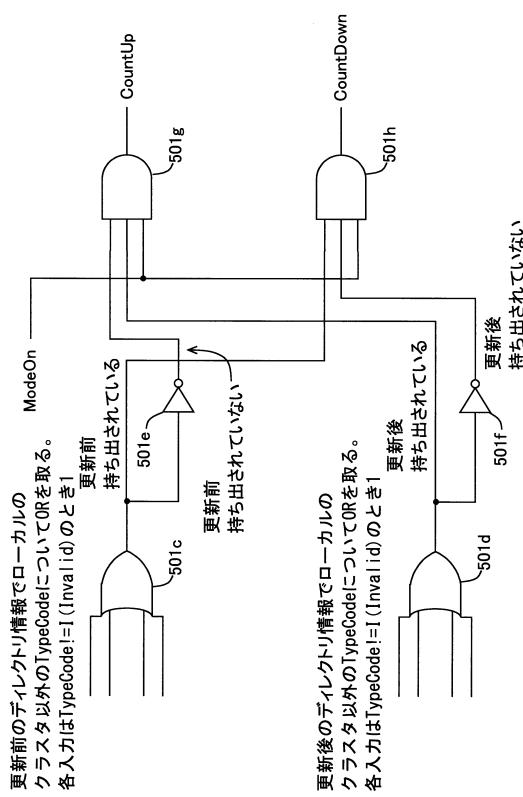
【図5】



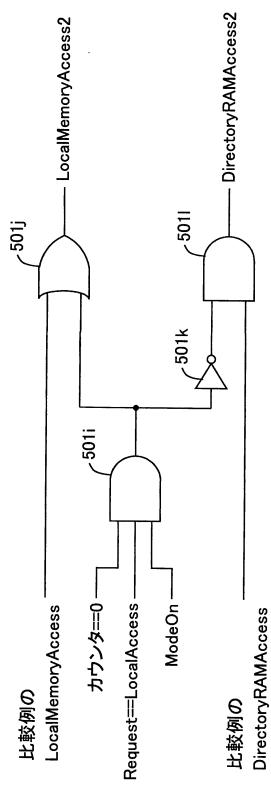
【図6】



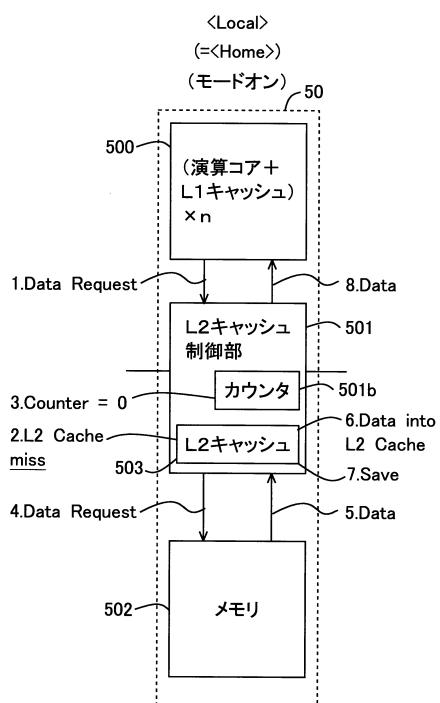
【図10】



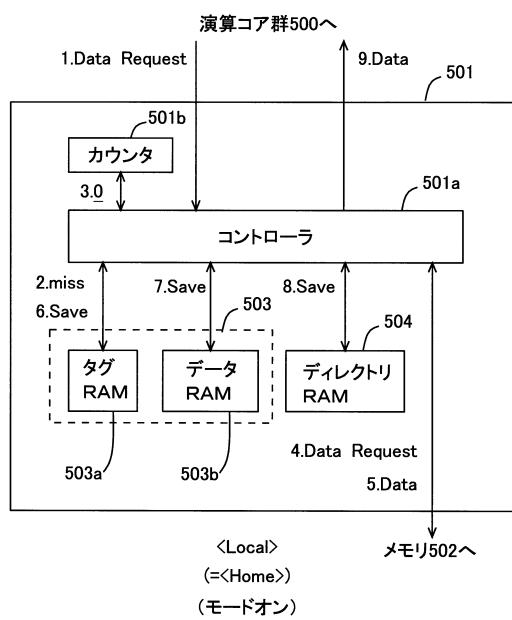
【図11】



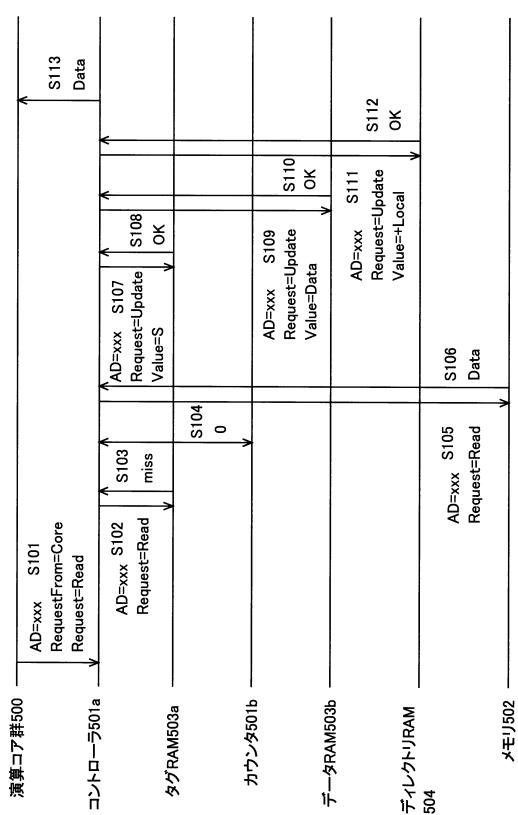
【図12】



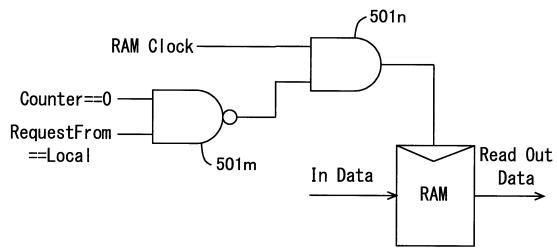
【図13】



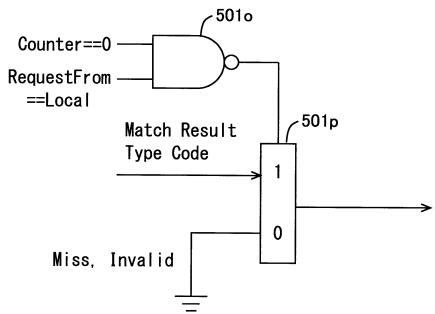
【図14】



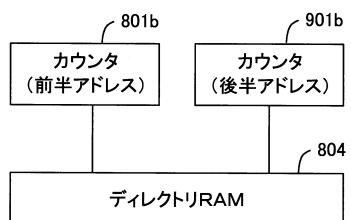
【図15】



【図16】



【図17】



フロントページの続き

審査官 後藤 彰

(56)参考文献 米国特許出願公開第2005/0120183(US,A1)
特開2005-234854(JP,A)
米国特許第06721856(US,B1)

(58)調査した分野(Int.Cl., DB名)

G 06 F 12 / 08