

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5583893号
(P5583893)

(45) 発行日 平成26年9月3日(2014.9.3)

(24) 登録日 平成26年7月25日(2014.7.25)

(51) Int.Cl.

F I

G06F 12/08 (2006.01)

G06F 12/08 517C
G06F 12/08 503F
G06F 12/08 511Z
G06F 12/08 513
G06F 12/08 523B

請求項の数 4 (全 19 頁) 最終頁に続く

(21) 出願番号 特願2008-139986 (P2008-139986)
(22) 出願日 平成20年5月28日 (2008.5.28)
(65) 公開番号 特開2009-288977 (P2009-288977A)
(43) 公開日 平成21年12月10日 (2009.12.10)
審査請求日 平成23年2月17日 (2011.2.17)
審判番号 不服2013-9139 (P2013-9139/J1)
審判請求日 平成25年5月17日 (2013.5.17)

(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 100089118
弁理士 酒井 宏明
(72) 発明者 石村 直也
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 小島 広行
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】 演算処理装置及び演算処理装置の制御方法

(57) 【特許請求の範囲】

【請求項1】

データを記憶する主記憶装置に接続する演算処理装置において、
メモリアクセス命令をそれぞれ実行する複数の演算処理部と、
前記主記憶装置から読みだしたデータを複数のバンクに分割された前記主記憶装置に対応してそれぞれ保持する複数のキャッシュメモリと、
前記主記憶装置から読みだしたデータ又は前記キャッシュメモリから読みだしたデータのいずれかを選択する選択部と、

前記選択部が選択したデータを前記複数の演算処理部のいずれかに転送する転送部と、
前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前記複数のキャッシュメモリのいずれかに対して要求したデータがキャッシュミスが発生させた場合、前記キャッシュミスの発生に応じて前記主記憶装置の複数のバンクのいずれかから読みだしたデータを前記複数のキャッシュメモリのうち前記主記憶装置の複数のバンクのいずれかに対応するキャッシュメモリに保持させるとともに、前記主記憶装置の複数のバンクのいずれかから読みだした前記データを前記選択部に選択させ、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に対して転送させる制御部と、を有し、

前記制御部は、

前記メモリアクセス命令を複数のステージを有するパイプラインで処理するとともに、前記キャッシュメモリからデータを読み出す場合の前記パイプラインにおける前記キャッ

10

20

シユメモリからデータを読み出すタイミングと、前記複数の演算処理部のいずれかが前記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた場合の前記パイプラインにおける前記主記憶装置から読みだしたデータを前記選択部に出力させるタイミングとを一致させるとともに、メモリアクセスの実行を開始した後、前記キャッシュメモリに対するアクセスに必要な期間、または前記転送部によるデータ転送に必要な期間、だけ後続のメモリアクセス命令の実行を待機させる

ことを特徴とする演算処理装置。

【請求項 2】

前記演算処理装置はさらに、

前記主記憶装置から読みだしたデータを保持して前記キャッシュメモリに出力するキューを有し、

前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた場合、前記キャッシュミスの発生に応じて前記主記憶装置から読みだしたデータを、前記キューを介して前記キャッシュメモリに保持させるとともに、前記キューから読みだしたデータを前記選択部に選択させ、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に転送させる制御部を有することを特徴とする請求項 1 記載の演算処理装置。

【請求項 3】

前記制御部は、

前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた場合、前記キャッシュミスの発生に応じて前記主記憶装置から読みだしたデータを前記キャッシュメモリに保持させた後に、前記主記憶装置から読みだしたデータを前記選択部に選択させ、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に転送させることを特徴とする請求項 1 または 2 に記載の演算処理装置。

【請求項 4】

データを記憶する主記憶装置に接続するとともに、前記主記憶装置から読みだしたデータを複数のバンクに分割された前記主記憶装置に対応してそれぞれ保持する複数のキャッシュメモリを有する演算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複数の演算処理部が、メモリアクセス命令をそれぞれ実行し、

前記演算処理装置が有する選択部が、前記主記憶装置から読みだしたデータ又は前記キャッシュメモリから読みだしたデータのいずれかを選択し、

前記演算処理装置が有する転送部が、前記選択部が選択したデータを前記複数の演算処理部のいずれかに転送し、

前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前記複数のキャッシュメモリのいずれかに対して要求したデータがキャッシュミスが発生させた場合、前記演算処理装置が有する制御部が、前記キャッシュミスの発生に応じて前記主記憶装置の複数のバンクのいずれかから読みだしたデータを前記複数のキャッシュメモリのうち前記主記憶装置の複数のバンクのいずれかに対応するキャッシュメモリに保持させ、

前記制御部が、前記主記憶装置の複数のバンクのいずれかから読みだした前記データを前記選択部に選択させ、

前記制御部が、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に対して転送させ、

前記制御部が、前記メモリアクセス命令を複数のステージを有するパイプラインで処理するとともに、前記キャッシュメモリからデータを読み出す場合の前記パイプラインにおける前記キャッシュメモリからデータを読み出すタイミングと、前記複数の演算処理部のいずれかが前記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた

10

20

30

40

50

場合の前記パイプラインにおける前記主記憶装置から読みだしたデータを前記選択部に出
力させるタイミングとを一致させ、

前記制御部が、メモリアクセスの実行を開始した後、前記キャッシュメモリに対するア
クセスに必要な期間、または前記転送部によるデータ転送に必要な期間、だけ後続のメモ
リアクセス命令の実行を待機させる

ことを特徴とする演算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、演算処理装置及び演算処理装置の制御方法に関する。本発明は、回路規模を
増大させることなく、キャッシュミス時にコアがデータを取得するまでのレイテンシを短
縮することができる演算処理装置及び演算処理装置の制御方法を含む。

10

【背景技術】

【0002】

通常、CPU (Central Processing Unit) などのプロセッサコアを備えた情報処理装
置には、処理の高速化を図るために、キャッシュメモリが設けられる。キャッシュメモリ
は、メインメモリなどの主記憶部よりも高速にアクセス可能なメモリであり、主記憶部に
記憶されたデータのうちCPUが頻繁に使用するデータのみを記憶している。このため、
CPUは、様々な演算処理を実行する際、最初にキャッシュメモリにアクセスし、必要な
データを要求する。このとき、必要なデータがキャッシュメモリに記憶されていなければ
、キャッシュミスが発生したこととなり、必要なデータは主記憶部からキャッシュメモリ
へ転送される。すなわち、キャッシュメモリに対するリード (READ) が実行された結果、
キャッシュミスが発生すると、ムーブイン (MOVE-IN) により主記憶部からキャッシュメ
モリへデータが転送される。

20

【0003】

このように、キャッシュミスが発生する場合には、ムーブインによって必要なデータが
キャッシュメモリに格納されるため、CPUは、再度リードを実行して、キャッシュメモリ
からデータを読み出すことになる。このため、CPUがデータを取得するまでには、2
つのリードと1つのムーブインとが実行されることとなり、遅延時間 (以下「レイテンシ
」という) が長くなる。そこで、データ取得までの遅延を低減して情報処理装置の性能を
向上するために、主記憶部からキャッシュメモリへデータが転送されるのと同時に、デー
タをCPUへも転送することが検討されている (例えば特許文献1参照)。

30

【0004】

図7は、特許文献1に記載された情報処理装置の構成を示す図である。図7に示す情報
処理装置は、主記憶部1、命令制御部2、および記憶制御部3を有している。命令制御部
2がデータを要求すると、記憶制御部3においてデータのリードが実行される。すなわち
、命令制御部2からのデータ要求が記憶制御部3内のセクタ7を介してキャッシュ4へ
送られ、要求されたデータがキャッシュ4に記憶されている場合は (キャッシュヒット)
、該当するデータがセクタ5を介して命令制御部2内のバッファ8へ読み出される。こ
の場合、図8の上段に示すように、命令制御部2がデータを記憶制御部3から取得する
ことができ、データ取得までは1つのリード (図中「RD」と示す) が実行されるのみで
ある。したがって、命令制御部2がデータを取得するまでのレイテンシはほとんどない。

40

【0005】

一方、命令制御部2によって要求されたデータがキャッシュ4に記憶されていない場合
は (キャッシュミス)、主記憶部1に記憶されたデータをキャッシュ4へ転送するムーブ
インが実行される。すなわち、命令制御部2からのデータ要求がセクタ7を介して主記
憶部1へ送られ、要求されたデータがセクタ6を介してキャッシュ4へ転送される。通
常であれば、以後、再び命令制御部2がデータを要求し、記憶制御部3においてリードが
実行され、要求されたデータがキャッシュ4からセクタ5を介してバッファ8へ読み出
されることになる。この場合は、図8の中段に示すように、命令制御部2がデータを取得

50

するまでに、2つのリード(RD)と1つのムーブイン(図中「MI」と示す)が実行される。したがって、命令制御部2がデータを取得するまでのレイテンシが長くなる。

【0006】

しかし、特許文献1においては、主記憶部1からセクタ5を介して命令制御部2のバッファ8へ直接接続するラインLが設けられており、データが主記憶部1からセクタ6を介してキャッシュ4へ転送されるのと同時に、ラインLを経由してバッファ8へ読み出される。このため、図8の下段に示すように、命令制御部2は、キャッシュ4へのムーブイン(MI)と同時にデータを取得することができ、レイテンシを短縮することができる。

【0007】

【特許文献1】特開平10-111798号公報

【発明の開示】

【発明が解決しようとする課題】

【0008】

ところで、近年、1つのプロセッサコア(以下、単に「コア」という)を備えるシングルコアの半導体集積回路においては、消費電力の増大などの問題が無視できないものとなっており、性能向上の限界が近づいている。そこで、1つの基板上に複数のコアを備えるマルチコア化によって、半導体集積回路のさらなる性能向上が図られることがある。また、マルチコア化と同時に、キャッシュメモリや主記憶部を複数のバンクに分割することにより、各コアとキャッシュメモリや主記憶部との間のスループットの向上が図られることがある。

【0009】

このようにマルチコア化およびバンク分けされた半導体集積回路においては、複数のコア、キャッシュメモリ、および主記憶部に接続する主記憶制御部が基板の外縁に配置され、基板の中央には、全体のデータの転送を制御する制御部が配置される。そして、バンク分けされたそれぞれの主記憶部は、アドレスが異なるデータを記憶しているため、各コアは、基板上のすべての主記憶制御部に対してデータを要求することがある。したがって、上述した特許文献1のように、主記憶部とコアを直接接続する場合には、すべてのコアと主記憶制御部を互いに接続する必要がある、基板上の配線が複雑になってしまうという問題がある。

【0010】

すなわち、各主記憶制御部は、基板の中央に配置された制御部を挟んで対向するコアからもデータを要求される可能性があるため、コアによるデータ取得までのレイテンシを短縮するためには、基板上で遠く離れた主記憶制御部とコアについても直接接続する必要がある。結果として、基板上の配線を大幅に変更・増設する必要が生じ、半導体集積回路の大型化を招いてしまう。最近では、半導体集積回路を搭載する装置は、小型化の一途をたどっており、コアのデータ取得におけるレイテンシを短縮するために半導体集積回路が大型化するのとは現実的ではない。

【0011】

本発明はかかる点に鑑みてなされたものであり、回路規模を増大させることなく、キャッシュミス時にコアがデータを取得するまでのレイテンシを短縮することができる演算処理装置及び演算処理装置の制御方法を提供することを目的とする。

【課題を解決するための手段】

【0012】

上記課題を解決するために、演算処理装置は、データを記憶する主記憶装置に接続する演算処理装置において、メモリアクセス命令をそれぞれ実行する複数の演算処理部と、前記主記憶装置から読みだしたデータを保持するキャッシュメモリと、前記主記憶装置から読みだしたデータ又は前記キャッシュメモリから読みだしたデータのいずれかを選択する選択部と、前記選択部が選択したデータを前記複数の演算処理部のいずれかに転送する転送部と、前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前

10

20

30

40

50

記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた場合、前記キャッシュミスの発生に応じて前記主記憶装置から読みだしたデータを前記キャッシュメモリに保持させるとともに、前記主記憶装置から読みだしたデータを前記選択部に選択させ、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に転送させる制御部を有する構成を採る。

【0014】

また、演算処理装置の制御方法は、データを記憶する主記憶装置に接続するとともに、前記主記憶装置から読みだしたデータを保持するキャッシュメモリを有する演算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複数の演算処理部が、メモリアクセス命令をそれぞれ実行し、前記複数の演算処理部のいずれかが、前記メモリアクセス命令の実行に応じて前記キャッシュメモリに対して要求したデータがキャッシュミスが発生させた場合、前記演算処理装置が有する制御部が、前記キャッシュミスの発生に応じて前記主記憶装置から読みだしたデータを前記キャッシュメモリに保持させ、前記制御部が、前記主記憶装置から読みだしたデータを前記選択部に選択させ、前記制御部が、前記選択部に選択させたデータを前記転送部に前記複数の演算処理部のうち前記キャッシュミスが発生させたデータを要求した演算処理部に転送させるようにした。

【発明の効果】

【0015】

本明細書に開示された演算処理装置及び演算処理装置の制御方法によれば、回路規模を増大させることなく、キャッシュミス時にコアがデータを取得するまでのレイテンシを短縮することができる。

【発明を実施するための最良の形態】

【0016】

以下、本発明の一実施の形態について、図面を参照して詳細に説明する。以下においては、1つの基板上にコア、キャッシュメモリ、および主記憶制御部がそれぞれ複数配置される半導体集積回路について説明する。ただし、1つの基板上にコア、キャッシュメモリ、または主記憶制御部が1つのみ配置される半導体集積回路にも本発明を同様に適用することができる。

【0017】

図1は、本実施の形態に係る半導体集積回路100の概略構成を示す図である。図1に示す半導体集積回路100は、8つのコア#0～#7を有している。また、半導体集積回路100においては、図示しない4つにバンク分けされた主記憶部に対応して、キャッシュメモリおよび主記憶制御部が4つのデータメモリ#0～#3および4つのMAC (Memory Access Controller) #0～#3に分割されている。そして、コア#0～#7、データメモリ#0～#3、およびMAC#0～#3は、基板の外縁に配置されており、基板の中央には、制御部が配置されている。また、図1に示す半導体集積回路100は、各MAC#0～#3から各データメモリ#0～#3へのデータ転送や、各データメモリ#0～#3から各コア#0～#7へのデータ転送を制御するキャッシュメモリ制御装置を含んでいる。

【0018】

コア#0～#7は、制御部におけるリード (READ) の実行によって、データメモリ#0～#3に一時的に保持されたデータを取得し、様々な演算処理を実行する。また、コア#0～#7は、データメモリ#0～#3に必要なデータが保持されていない場合 (キャッシュミス)、該当するデータを記憶している主記憶部のバンクに対応するMAC#0～#3に対してデータを要求する。そして、コア#0～#7は、制御部におけるムーブイン (MOVE-IN) の実行によって、MAC#0～#3からデータメモリ#0～#3へデータが転送される際、データメモリ#0～#3へ入力されるデータと同一のデータを取得する。すなわち、コア#0～#7は、データメモリ#0～#3への入力部分から、データメモリ#0～#3に書き込まれるデータと同一のデータを取得する。

【 0 0 1 9 】

データメモリ # 0 ~ # 3 は、それぞれキャッシュメモリの一部を構成しており、コア # 0 ~ # 7 が使用するデータを一時的に保持する。具体的には、データメモリ # 0 ~ # 3 は、制御部においてリードが実行されると、コア # 0 ~ # 7 が要求するデータを出力する。また、データメモリ # 0 ~ # 3 は、制御部においてムーブインが実行されると、コア # 0 ~ # 7 が要求したデータを M A C # 0 ~ # 3 から取得して記憶する。

【 0 0 2 0 】

M A C # 0 ~ # 3 は、それぞれアドレスが異なるデータを記憶する図示しない主記憶部の 4 つのバンクに対応している。そして、M A C # 0 ~ # 3 は、キャッシュミスが生じた場合に、コア # 0 ~ # 7 が要求するデータを図示しない主記憶部から取得してデータメモリ # 0 ~ # 3 へ転送する。

10

【 0 0 2 1 】

制御部は、パイプライン処理によってリードやムーブインなどを実行する。すなわち、制御部は、コア # 0 ~ # 7 がデータを要求した場合、パイプラインにリード要求を投入し、図示しないタグメモリを検索して、要求されたデータがデータメモリ # 0 ~ # 3 に記憶されているか否かを判断する。そして、要求されたデータがデータメモリ # 0 ~ # 3 に記憶されていれば（キャッシュヒット）、制御部は、該当するデータをデータメモリ # 0 ~ # 3 からコア # 0 ~ # 7 へ出力させる。一方、要求されたデータがデータメモリ # 0 ~ # 3 に記憶されていなければ（キャッシュミス）、制御部は、パイプラインにムーブイン要求を投入し、該当するデータを M A C # 0 ~ # 3 からデータメモリ # 0 ~ # 3 へ転送させる。

20

【 0 0 2 2 】

なお、制御部は、リードおよびムーブイン以外にも、コア # 0 ~ # 7 によって演算処理されたデータをデータメモリ # 0 ~ # 3 に書き戻すライトバック（WRITE-BACK）や、データをデータメモリ # 0 ~ # 3 から図示しない主記憶部へ排出するムーブアウト（MOVE-OUT）などを実行する。また、制御部は、リード要求やムーブイン要求などの命令をパイプラインに投入する際、複数の命令が同時にデータメモリ # 0 ~ # 3 などの資源を使用する干渉が生じないように、適切な間隔をあけて命令をパイプラインに投入する。

【 0 0 2 3 】

図 2 は、本実施の形態に係る半導体集積回路 1 0 0 の要部構成を示す図である。図 2 においては、特に、図 1 に示した半導体集積回路 1 0 0 のコア # 0 ~ # 3、データメモリ # 0、# 1、および M A C # 0、# 1 に対応する部分を拡大して示している。図 2 に示す半導体集積回路 1 0 0 は、コア 1 0 1 - 0、1 0 1 - 1、制御部 1 0 2、ライトバックデータキュー（以下「W B D Q」と略記する）1 0 3、ムーブインデータキュー（以下「M I D Q」と略記する）1 0 4、セレクトア 1 0 5、データメモリ 1 0 6、セレクトア 1 0 7、1 0 8、データバス 1 0 9、および M A C 1 1 5 を備えている。なお、図 2 に示す半導体集積回路 1 0 0 は、左右対称に構成されているため、左半分についてのみ符号を付している。また、図 2 においては、主にデータメモリ 1 0 6 付近におけるデータの移動を矢印で示している。

30

【 0 0 2 4 】

コア 1 0 1 - 0、1 0 1 - 1 は、データメモリ 1 0 6 または図示しない主記憶部に記憶されたデータを使用して様々な演算処理を実行する。このとき、コア 1 0 1 - 0、1 0 1 - 1 は、演算処理に必要なデータを制御部 1 0 2 に対して要求し、該当するデータがデータメモリ 1 0 6 に記憶されていれば（キャッシュヒット）、リードの実行によりデータメモリ 1 0 6 から読み出されたデータを取得する。また、コア 1 0 1 - 0、1 0 1 - 1 は、必要なデータがデータメモリ 1 0 6 に記憶されていなければ（キャッシュミス）、ムーブインの実行により主記憶部から M A C 1 1 5 を介してデータメモリ 1 0 6 へ転送されるデータを取得する。より具体的には、コア 1 0 1 - 0、1 0 1 - 1 は、ムーブインの実行によりデータメモリ 1 0 6 の入力部分に設けられた M I D Q 1 0 4 へ転送されたデータを、データメモリ 1 0 6 への書き込み終了後に取得する。

40

50

【 0 0 2 5 】

制御部 1 0 2 は、パイプライン処理によりコア 1 0 1 - 0、1 0 1 - 1 が要求するデータのリードやムーブインを実行する。すなわち、制御部 1 0 2 は、コア 1 0 1 - 0、1 0 1 - 1 がデータを要求すると、パイプラインにリード要求を投入し、キャッシュヒットの場合には、データメモリ 1 0 6 からコア 1 0 1 - 0、1 0 1 - 1 へデータを転送させる。また、制御部 1 0 2 は、キャッシュミスの場合には、要求されたデータを記憶する主記憶部に対応する M A C 1 1 5 から M I D Q 1 0 4 へデータを転送させた後、パイプラインにムーブイン要求を投入し、要求されたデータを M I D Q 1 0 4 からデータメモリ 1 0 6 へ転送させる。

【 0 0 2 6 】

W B D Q 1 0 3 は、ライトバックの実行時に、コア 1 0 1 - 0、1 0 1 - 1 からデータメモリ 1 0 6 へ書き戻されるデータを一時的に保持するキューである。すなわち、W B D Q 1 0 3 は、コア 1 0 1 - 0、1 0 1 - 1 によって加工されたデータを一時的に保持し、セクタ 1 0 5 を介してデータメモリ 1 0 6 へ出力する。

【 0 0 2 7 】

M I D Q 1 0 4 は、データメモリ 1 0 6 の入力部分に設けられ、キャッシュミスが生じた場合のムーブインの実行時に、M A C 1 1 5 からデータメモリ 1 0 6 へ転送されるデータを一時的に保持するキューである。すなわち、M I D Q 1 0 4 は、コア 1 0 1 - 0、1 0 1 - 1 によって要求された M A C 1 1 5 からのデータを一時的に保持し、セクタ 1 0 5 を介してデータメモリ 1 0 6 へ出力する。また、M I D Q 1 0 4 は、データがデータメモリ 1 0 6 に書き込まれた後、同一のデータをセクタ 1 0 7 へも出力する。

【 0 0 2 8 】

セクタ 1 0 5 は、W B D Q 1 0 3 および M I D Q 1 0 4 に保持されたデータのいずれか一方を選択してデータメモリ 1 0 6 へ出力する。具体的には、セクタ 1 0 5 は、制御部 1 0 2 のパイプラインによってライトバックが実行されている場合は、W B D Q 1 0 3 に保持されたデータを選択する。また、セクタ 1 0 5 は、制御部 1 0 2 のパイプラインによってムーブインが実行されている場合は、M I D Q 1 0 4 に保持されたデータを選択する。

【 0 0 2 9 】

さらに、セクタ 1 0 5 は、ムーブインが実行されている場合は、M I D Q 1 0 4 に保持されたデータがデータメモリ 1 0 6 に書き込まれた後、同一のデータをライン L 0 を介してセクタ 1 0 7 へも出力する。つまり、セクタ 1 0 5 は、ムーブインの実行によりデータメモリ 1 0 6 に書き込まれたデータと同一のデータをライン L 0 へ出力し、セクタ 1 0 7 からコア 1 0 1 - 0、1 0 1 - 1 へ転送させる。

【 0 0 3 0 】

ここで、ライン L 0 は、M I D Q 1 0 4 およびデータメモリ 1 0 6 の中間とセクタ 1 0 7 とを接続する配線であり、データメモリ 1 0 6 の入力部分と出力部分を結んでいる。したがって、ライン L 0 は、データメモリ 1 0 6 の周囲にのみ設けられる非常に短いラインであり、新規にライン L 0 を設けても、周辺の配線に与える影響はほとんどない。そして、ライン L 0 は、データメモリ 1 0 6 へ入力するデータと同一のデータを、リードが実行された場合にデータメモリ 1 0 6 から出力するデータが経由するデータメモリ 1 0 6 の出力部分へ転送する。

【 0 0 3 1 】

データメモリ 1 0 6 は、キャッシュメモリの一部を構成するメモリであり、コア 1 0 1 - 0、1 0 1 - 1 によって頻繁に使用されるデータを一時的に記憶する。本実施の形態においては、キャッシュメモリは、バンク分けされたデータメモリとタグメモリとを備えており、それぞれのメモリとしては、例えば R A M (Random Access Memory) などを使用することが可能である。

【 0 0 3 2 】

セクタ 1 0 7 は、データメモリ 1 0 6 の出力部分に設けられ、M I D Q 1 0 4 から出

10

20

30

40

50

力されたデータとデータメモリ106から出力されたデータとのいずれか一方を選択して、セクタ108およびコア101-0、101-1以外のコアへ出力する。具体的には、セクタ107は、制御部102のパイプラインによってムーブインが実行されている場合は、MIDQ104から出力されたデータを選択する。また、セクタ107は、制御部102のパイプラインによってリードが実行されている場合は、データメモリ106から出力されたデータを選択する。

【0033】

換言すれば、セクタ107は、キャッシュヒット時にリードされるデータおよびキャッシュミス時にムーブインされるデータのいずれか一方をコア101-0、101-1へ転送するデータとして選択する。そして、キャッシュミス時には、セクタ107は、ムーブインの実行によりデータがデータメモリ106に書き込まれたとき、同一のデータをラインL0から取得して、コア101-0、101-1へ転送する。このように、セクタ107は、ムーブインの実行時にデータメモリ106へ入力するデータと同一のデータを、リードの実行時にデータメモリ106から出力するデータと同等に扱ってコア101-0、101-1へ転送する。これにより、ムーブインの実行によってデータメモリ106に入力したデータを、リードの実行により改めてデータメモリ106から出力しなくても、要求されたデータを迅速にコア101-0、101-1へ転送することが可能となる。

10

【0034】

なお、セクタ107は、リードが実行されている場合であってもムーブインが実行されている場合であっても、パイプライン処理の同じサイクルでデータを出力する。すなわち、ムーブイン時には、データメモリ106に対するデータ書き込みアクセス後に、データメモリ106に書き込まれたのと同じデータがMIDQ104からセクタ107へ出力される。また、リード時には、データメモリ106に対するデータ読み出しアクセス後に、データメモリ106から読み出されたデータがデータメモリ106からセクタ107へ出力される。したがって、ムーブイン時でもリード時でも、セクタ107にデータが到達するパイプライン処理のサイクルが同じになり、セクタ107からコア101-0、101-1へのデータ転送は、まったく同じ処理となる。

20

【0035】

セクタ108は、セクタ107から出力されたデータとデータメモリ106以外のデータメモリから出力されたデータとのいずれか1つを選択して、コア101-0、101-1へ出力する。具体的には、セクタ108は、コア101-0、101-1がデータメモリ106に記憶されたデータを要求した場合は、セクタ107から出力されたデータを選択する。また、セクタ108は、コア101-0、101-1がデータメモリ106以外のデータメモリに記憶されたデータを要求した場合は、データ要求先のデータメモリから出力されたデータを選択する。

30

【0036】

データバス109は、セクタ108とコア101-0、101-1とを接続し、データメモリ106またはMIDQ104から出力されてセクタ107およびセクタ108を経由したデータをコア101-0、101-1へ転送する。つまり、データバス109は、制御部102のパイプラインによってリードが実行される場合にもムーブインが実行される場合にも共通して使用されるため、それぞれの命令の実行期間中において所定サイクルの間占有されることになる。

40

【0037】

MAC115は、コア101-0、101-1が使用し得るすべてのデータを記憶する図示しない主記憶部に接続しており、キャッシュミスが生じた場合、キャッシュミスしたデータを主記憶部から取得し、MIDQ104へ転送する。MIDQ104へ転送されたデータは、ムーブインの実行により、MIDQ104からデータメモリ106へ転送される。

【0038】

50

次いで、本実施の形態に係るコア、キャッシュメモリ、およびMACの接続関係について、図3を参照しながらより詳細に説明する。図3は、コア101、データメモリ106とタグメモリ112を備えるキャッシュメモリ、およびMAC115の接続関係を模式的に示す図である。同図において、図2と同じ部分には同じ符号を付している。図3においては、図2に示した構成に加えてムーブインポート（以下「MIポート」と略記する）110、セクタ111、タグメモリ112、ムーブインバッファ（以下「MIバッファ」と略記する）113、およびムーブアウトデータキュー（以下「MODQ」と略記する）114を図示している。

【0039】

MIポート110は、コア101からデータの要求を受信し、制御部102に対してリードの実行を要求する。セクタ111は、リードの実行またはムーブインの実行を制御部102に要求する。すなわち、セクタ111は、MIポート110からリードの実行が要求された場合は、制御部102に対してリードの実行を要求し、MIバッファ113からムーブインの実行が要求された場合は、制御部102に対してムーブインの実行を要求する。

10

【0040】

タグメモリ112は、データメモリ106に記憶されたデータのアドレスを保持しており、制御部102のパイプラインにリード要求が投入されると、コア101が要求したデータがデータメモリ106に記憶されているか否かを判断する。つまり、要求されたデータのアドレスをタグメモリ112が保持していればキャッシュヒットとなり、要求されたデータのアドレスをタグメモリ112が保持していなければキャッシュミスとなる。

20

【0041】

MIバッファ113は、キャッシュミスが生じた場合、キャッシュミスしたデータをMIDQ104へ転送するようにMAC115に対して指示する。そして、MIバッファ113は、キャッシュミスしたデータがMIDQ104へ転送された後、制御部102に対してムーブインの実行を要求する。

【0042】

MODQ114は、キャッシュリプレース時に、データメモリ106からMAC115を介して図示しない主記憶部へ書き戻されるデータを一時的に保持する。

【0043】

以上のような構成において、コア101がデータを要求する際、データの要求は、MIポート110によって受信され、リード実行の要求がセクタ111を介して制御部102へ入力する。そして、制御部102においては、パイプラインにリード要求が投入され、タグメモリ112の検索により、要求されたデータがデータメモリ106に記憶されているか否かが判断される。判断の結果、要求されたデータがデータメモリ106に記憶されており、キャッシュヒットした場合は、該当するデータがデータメモリ106から出力され、セクタ107、108、およびデータバス109を介してコア101へ転送される。これにより、リードのパイプライン処理が完了する。

30

【0044】

一方、コア101が要求したデータのアドレスがタグメモリ112に保持されておらず、キャッシュミスした場合は、キャッシュミスしたデータのアドレスがタグメモリ112からMIバッファ113へ通知される。そして、MIバッファ113からMAC115に対して、キャッシュミスしたデータをMIDQ104へ転送するように指示される。すなわち、コア101が要求したデータのアドレスがタグメモリ112に保持されていない場合、MIバッファ113は、コア101が要求したデータをMIDQ104へ転送するようにMAC115へ指示する。

40

【0045】

MAC115は、MIバッファ113からの指示を受け、コア101が要求したデータを主記憶部から取得してMIDQ104へ転送する。コア101が要求したデータがMIDQ104へ転送された後、MIバッファ113によって、ムーブイン実行の要求がセ

50

クタ111を介して制御部112へ入力される。制御部102においては、パイプラインにムーブイン要求が投入され、ムーブイン要求のパイプライン処理に応じて、MIDQ104に一時的に保持されたデータがセクタ105を介してデータメモリ106へ出力され、データメモリ106に書き込まれる。MIDQ104から出力されたデータがデータメモリ106に書き込まれた後、同一のデータがMIDQ104からラインL0を介してセクタ107へ出力される。

【0046】

ここで、ラインL0を介してMIDQ104からセクタ107へデータが出力されるサイクルは、リード実行時にデータメモリ106からデータが出力されるのと同じサイクルである。つまり、本実施の形態においては、MIDQ104から出力されたデータがデータメモリ106に書き込まれるのと同時にラインL0を介してセクタ107へ出力されるのではなく、データの書き込み終了後にセクタ107へデータが出力される。これにより、リード実行時およびムーブイン実行時でセクタ107にデータが到達するパイプライン処理のサイクルが同じになり、セクタ107からコア101へのデータ転送を共通の制御で実現することができる。

【0047】

MIDQ104からのデータがラインL0を介してセクタ107に到達すると、データは、リード実行時と同様に、セクタ107、108、およびデータバス109を介してコア101へ転送される。これにより、ムーブインのパイプライン処理が完了する。つまり、本実施の形態においては、キャッシュミスが生じた場合でも、ムーブインの完了と同時に、要求されたデータがコア101によって取得される。換言すれば、ムーブインによってデータメモリ106に書き込まれたデータを、再度リードによって読み出してコア101へ転送する必要がない。結果として、キャッシュミス時にコア101がデータを取得するまでのレイテンシを1つのリード実行に相当する時間だけ短縮することができる。

【0048】

次に、キャッシュミス時のレイテンシの短縮について、図4および図5を参照しながら説明する。図4は、本実施の形態に係るキャッシュミス時の動作を示すシーケンス図である。

【0049】

コア101による演算などのためにデータが必要となると、MIポート110およびセクタ111を介してコア101から制御部102に対してリードの実行が要求される(ステップS101)。制御部102においては、パイプラインにリード要求が投入され、リード要求が処理されるとタグメモリ112の検索などが行われる。図4の例では、コア101によって要求されたデータがデータメモリ106に記憶されておらず、データのアドレスがタグメモリ112に保持されていない。すなわち、キャッシュミスが生じるため、制御部102からMIバッファ113を介してMAC115へデータの転送が指示される(ステップS102)。そして、コア101によって要求されたデータは、MAC115によって図示しない主記憶部から取得され、MIDQ104へ転送され(ステップS103)、MIDQ104に一時的に保持される。

【0050】

また、制御部102においては、リード要求が投入された後、所定サイクルの間隔をあけて、MIバッファ113からパイプラインにムーブイン要求が投入される。パイプラインへのリード要求の投入後、ムーブイン要求の投入までに所定サイクルの間隔をあけるのは、データメモリ106やデータバス109などの資源が複数の命令によって同時に使用されることを防ぐためである。

【0051】

なお、実際にキャッシュミスが生じた場合には、上述したように、MIバッファ113からMAC115に対してキャッシュミスしたデータをMIDQ104へ転送するように指示され、該当するデータがMAC115によってMIDQ104へ転送された後に、ムーブイン要求が投入される。したがって、パイプラインへムーブイン要求が投入される前

10

20

30

40

50

に、M I D Q 1 0 4 へのムーブイン対象のデータの転送を完了する必要がある。このため、パイプラインへのリード要求の投入からムーブイン要求の投入までの間隔は、データメモリ 1 0 6 やデータバス 1 0 9 などの資源が複数の命令によって同時に使用されることを防ぐために必要な間隔よりも長く（例えば 1 0 0 サイクル以上）なる。

【 0 0 5 2 】

キャッシュミスしたデータが M I D Q 1 0 4 に保持され、制御部 1 0 2 のパイプラインにムーブイン要求が投入されると、M I D Q 1 0 4 に保持されたデータは、セクタ 1 0 5 を介してデータメモリ 1 0 6 へ出力され（ステップ S 1 0 4 ）、データメモリ 1 0 6 に書き込まれる。データがデータメモリ 1 0 6 に書き込まれると、同一のデータが M I D Q 1 0 4 からライン L 0 を介してセクタ 1 0 7 へ出力され、セクタ 1 0 7、1 0 8、およびデータバス 1 0 9 を介してコア 1 0 1 へ転送される（ステップ S 1 0 5 ）。これにより、コア 1 0 1 は、ムーブインの完了と同時に要求したデータを取得することができる。ここで、もし、ライン L 0 およびセクタ 1 0 7 が配設されていなければ、ステップ S 1 0 5 の工程は実行不可能であり、コア 1 0 1 がデータを取得するまでには、ムーブインの完了後、さらに図 4 に破線で示した工程が必要となる。

10

【 0 0 5 3 】

すなわち、ステップ S 1 0 5 の処理が実行されることなく、M I バッファ 1 1 3 から制御部 1 0 2 に対して、再度リードの実行が要求される。制御部 1 0 2 においては、パイプラインにリード要求が投入され、リード要求が処理されるとタグメモリ 1 1 2 の検索などが行われる。ここでは、既に完了したムーブインにより、コア 1 0 1 によって要求されたデータがデータメモリ 1 0 6 に記憶されている。すなわち、キャッシュヒットするため、データメモリ 1 0 6 に対して、要求されたデータを出力するように指示される（ステップ S 2 0 1 ）。そして、データメモリ 1 0 6 からコア 1 0 1 へ要求されたデータが転送されることになる（ステップ S 2 0 2 ）。

20

【 0 0 5 4 】

以上のステップ S 2 0 1 ~ S 2 0 2 の工程は、ライン L 0 およびセクタ 1 0 7 を設けることにより、省略することが可能となり、制御部 1 0 2 のパイプラインにおいてリード 1 つを実行する時間だけレイテンシを短縮することができる。つまり、ライン L 0 およびセクタ 1 0 7 がない場合には、図 5 の上段に示すように、コア 1 0 1 がデータを取得するまでに、パイプラインにおいて 2 つのリード（R D ）と 1 つのムーブイン（M I ）が実行される。これに対して、本実施の形態に係るライン L 0 およびセクタ 1 0 7 がある場合には、図 5 の下段に示すように、コア 1 0 1 は、1 つのリード（R D ）と 1 つのムーブイン（M I ）の実行のみで、要求したデータを取得することができる。

30

【 0 0 5 5 】

したがって、本実施の形態においては、リード 1 つ分の時間だけレイテンシを短縮することができると同時に、パイプラインに投入されるリード要求の数が削減されるため、結果として、パイプラインの使用回数を削減することができる。

【 0 0 5 6 】

次に、制御部 1 0 2 におけるパイプライン処理の具体例について、図 6 を参照しながら説明する。図 6 は、リード要求、ムーブイン要求、リード要求の順で制御部 1 0 2 のパイプラインに命令が投入される場合のパイプライン処理の一例を示す図である。図 6 において、上部の数字は、サイクル数を示している。また、各サイクルを横切る太線は、それぞれのサイクルにおいて、データメモリ 1 0 6 およびデータバス 1 0 9 などの資源を使用した処理が実行されることを示している。ただし、図 6 に示すサイクル数は、説明の便宜上利用している数字であるため、実際のパイプライン処理におけるサイクル数は、図 6 と異なっても良い。また、以下においては、データの書き込みおよび読み出し時のデータメモリ 1 0 6 に対するアクセスには 4 サイクルが必要であるものとする。同様に、データバス 1 0 9 によるコア 1 0 1 へのデータの転送にも 4 サイクルが必要であるものとする。

40

【 0 0 5 7 】

第 0 サイクルで制御部 1 0 2 のパイプラインにリード要求が投入されると、データメモ

50

リ 1 0 6 およびデータバス 1 0 9 などの資源に関して、投入されたリード要求と次の命令とが干渉しないように、第 0 サイクルから第 3 サイクルの 4 サイクルの間パイプラインへの命令の投入が禁止される。すなわち、上述したように、リードの実行中におけるデータメモリ 1 0 6 からのデータの読み出しには 4 サイクルが必要であるため、リードの実行によって 4 サイクルの間はデータメモリ 1 0 6 が占有される。同様に、リードの実行に伴うデータの転送によって、4 サイクルの間はデータバス 1 0 9 が占有される。したがって、次に投入される命令との干渉が生じないように、第 0 サイクルから第 3 サイクルの 4 サイクルの間は、パイプラインによる新たな命令の実行開始が待機される。

【 0 0 5 8 】

また、第 0 サイクルでリード要求が投入されると、第 1 サイクルから第 4 サイクルにおいて、タグメモリ 1 1 2 におけるアドレスの検索などが実行され、第 5 サイクルから第 8 サイクルの 4 サイクルにおいて、データメモリ 1 0 6 からのデータの読み出しが実行される。そして、第 9 サイクルでは、データがデータメモリ 1 0 6 からセクタ 1 0 7 へ出力される。以後、第 1 0 サイクルから第 1 1 サイクルでデータがセクタ 1 0 7 からセクタ 1 0 8 へ転送され、第 1 2 サイクルから第 1 5 サイクルの 4 サイクルにおいて、データバス 1 0 9 によってデータがコア 1 0 1 へ転送される。これにより、最初のリードが完了する。

【 0 0 5 9 】

ところで、第 0 サイクルにおけるリード要求の投入により、第 0 サイクルから第 3 サイクルではパイプラインへの命令の投入が禁止されているが、第 4 サイクルになると、パイプラインに新たなムーブイン要求が投入される。ここでも、リード要求投入時と同様に、4 サイクルの間パイプラインへの命令の投入が禁止される。すなわち、上述したように、ムーブインの実行中におけるデータメモリ 1 0 6 へのデータの書き込みには、4 サイクルが必要であるため、ムーブインの実行によって、4 サイクルの間はデータメモリ 1 0 6 が占有される。同様に、ムーブインの実行に伴うデータの転送によって、4 サイクルの間はデータバス 1 0 9 が占有される。したがって、次に投入される命令のとの干渉が生じないように、第 4 サイクルから第 7 サイクルの 4 サイクルの間は、パイプラインによる新たな命令の実行開始が待機される。

【 0 0 6 0 】

また、第 4 サイクルでムーブイン要求が投入されると、第 5 サイクルから第 7 サイクルにおいて、タグメモリ 1 1 2 におけるアドレスの新規登録などが実行され、第 8 サイクルにおいて、M I D Q 1 0 4 に保持されたデータがセクタ 1 0 5 を介してデータメモリ 1 0 6 へ出力される。そして、第 9 サイクルから第 1 2 サイクルの 4 サイクルにおいて、データメモリ 1 0 6 へのデータの書き込みが実行される。また、第 1 3 サイクルでは、データメモリ 1 0 6 に書き込まれたデータと同一のデータが、M I D Q 1 0 4 からライン L 0 を介してセクタ 1 0 7 へ出力される。以後、第 1 4 サイクルから第 1 5 サイクルでデータがセクタ 1 0 7 からセクタ 1 0 8 へ転送され、第 1 6 サイクルから第 1 9 サイクルの 4 サイクルにおいて、データバス 1 0 9 によってデータがコア 1 0 1 へ転送される。これにより、ムーブインが完了する。

【 0 0 6 1 】

ここで注目すべきなのは、最初のリード実行時およびムーブイン実行時の両方において、命令が投入されてから 9 サイクル目でセクタ 1 0 7 へデータが出力されていることである。つまり、リード実行時には第 0 サイクルでリード要求が投入され、第 9 サイクルでデータがセクタ 1 0 7 へ出力されている。一方、ムーブイン実行時には第 4 サイクルでムーブイン要求が投入され、第 1 3 サイクルでデータがセクタ 1 0 7 へ出力されている。したがって、セクタ 1 0 7 からコア 1 0 1 までのデータ転送に関しては、リード実行時およびムーブイン実行時の両方に共通した制御を実行すれば良い。

【 0 0 6 2 】

このように、データメモリ 1 0 6 の入力部分と出力部分とを接続するライン L 0 を設けることにより、M I D Q 1 0 4 から出力されたデータをデータメモリ 1 0 6 から出力され

10

20

30

40

50

たデータと同等に扱うことが可能となる。すなわち、データメモリ106の入力部分からラインL0経由で出力部分へデータを転送するタイミングを、リード実行時にデータメモリ106から実際にデータが出力されるタイミングと一致させることができる。そして、結果として、リード実行時およびムーブイン実行時のセクタ107からコア101までのデータ転送に関する制御を共通化することができる。

【0063】

ところで、第4サイクルにおけるムーブイン要求の投入により、第4サイクルから第7サイクルではパイプラインへの命令の投入が禁止されているが、第8サイクルになると、パイプラインに新たなリード要求が投入される。以降、最初のリードと同様に、第13サイクルから第16サイクルにおいて、データメモリ106からデータが読み出され、リード要求の投入から9サイクル目に当たる第17サイクルでは、データメモリ106から読み出されたデータがセクタ107へ出力される。そして、第20サイクルから第23サイクルにおいて、データバス109によってデータがコア101へ転送される。これにより、2つ目のリードが完了する。

10

【0064】

以上の2つのリードおよび1つのムーブインは、互いに4サイクルの間隔があげられてパイプラインに投入されている。このため、図6に示すように、データメモリ106およびデータバス109は、3つの命令によって途切れることなく連続して使用されることになる。すなわち、データメモリ106は、最初のリードによって第5サイクルから第8サイクルにアクセスされており、ムーブインによって第9サイクルから第12サイクルでアクセスされており、2つ目のリードによって第13サイクルから第16サイクルにアクセスされている。また、データバス109は、最初のリードによって第12サイクルから第15サイクルにデータを転送し、ムーブインによって第16サイクルから第19サイクルでデータを転送し、2つ目のリードによって第20サイクルから第23サイクルでデータを転送している。

20

【0065】

このように、本実施の形態においては、データメモリ106およびデータバス109などの資源における命令間の干渉がなく、かつ、資源が途切れることなく連続して使用されており、資源利用の効率化を図ることができる。なお、図6に示した例では、リードおよびムーブインの双方がデータメモリ106およびデータバス109を4サイクルずつ占有するものとしたが、占有するサイクル数は、4サイクルに限定されない。また、必ずしもデータメモリ106およびデータバス109が占有されるサイクル数が等しくなくても良い。これらの場合には、連続する2つの命令間で資源の使用に関する干渉が発生しないように、最も長い資源の占有サイクル数をパイプラインへの命令投入禁止期間とすれば良い。

30

【0066】

以上のように、本実施の形態によれば、キャッシュメモリの一部を構成するデータメモリの入力部分と出力部分とを接続し、ムーブイン実行時に入力部分からデータメモリへ入力されるデータと同一のデータをデータメモリの出力部分にも転送する。このため、ムーブインによってデータメモリに書き込まれたデータをリード実行時にデータメモリから読み出されたデータと同等に扱うことができる。すなわち、ムーブインによってデータメモリに書き込まれたデータを改めてリードによって読み出す必要がなく、リードの実行を省略することができる。結果として、回路規模を増大させることなく、キャッシュミス時にコアがデータを取得するまでのレイテンシを短縮することができる。

40

【0067】

以上の実施の形態に関して、さらに以下の付記を開示する。

【0068】

(付記1) データを使用した演算処理を実行する演算処理手段と、

主記憶部によって記憶されるデータのうち前記演算処理手段によって使用されるデータを記憶するキャッシュメモリと、

50

前記キャッシュメモリの入力部分と出力部分とを接続する接続手段と、

主記憶部のデータが前記キャッシュメモリの入力部分から前記キャッシュメモリに入力される場合に、当該データを前記接続手段経由で前記キャッシュメモリの出力部分へ転送させる制御手段と、

前記制御手段の制御により前記キャッシュメモリの出力部分へ転送されたデータを前記演算処理手段へ転送する転送手段と

を有することを特徴とするキャッシュメモリ制御装置。

【0069】

(付記2) 前記キャッシュメモリの入力部分に設けられ、主記憶部のデータを前記キャッシュメモリに書き込む際に一時的に保持するキュー手段をさらに有し、

10

前記制御手段は、

前記キュー手段から出力されるデータを前記接続手段経由で前記キャッシュメモリの出力部分へ転送させることを特徴とする付記1記載のキャッシュメモリ制御装置。

【0070】

(付記3) 前記キャッシュメモリの出力部分に設けられ、前記キャッシュメモリから実際に出力されるデータおよび前記接続手段経由で転送されたデータのいずれか一方を選択する選択手段をさらに有し、

前記転送手段は、

前記選択手段によって選択されたデータを前記演算処理手段へ転送することを特徴とする付記1記載のキャッシュメモリ制御装置。

20

【0071】

(付記4) 前記選択手段は、

前記演算処理手段によって要求されるデータが前記キャッシュメモリに記憶されておらず、当該データが主記憶部から前記キャッシュメモリへ転送される場合に、前記接続手段経由で転送されたデータを選択することを特徴とする付記3記載のキャッシュメモリ制御装置。

【0072】

(付記5) 前記制御手段は、

前記キャッシュメモリに入力されたデータが前記キャッシュメモリに書き込まれた後に、当該データと同一のデータを前記接続手段経由で前記キャッシュメモリの出力部分へ転送させることを特徴とする付記1記載のキャッシュメモリ制御装置。

30

【0073】

(付記6) 前記制御手段は、

データの転送に関する命令をパイプライン処理するパイプライン処理手段を含み、

前記パイプライン処理手段のパイプライン処理において、前記キャッシュメモリから実際にデータが出力されるサイクルと、前記接続手段経由で前記キャッシュメモリの出力部分へデータが転送されるサイクルとを一致させることを特徴とする付記1記載のキャッシュメモリ制御装置。

【0074】

(付記7) 前記パイプライン処理手段は、

40

1つの命令の実行が開始された後、前記キャッシュメモリに対するアクセスが継続するサイクル数だけ次の命令の実行開始を待機することを特徴とする付記6記載のキャッシュメモリ制御装置。

【0075】

(付記8) 前記パイプライン処理手段は、

1つの命令の実行が開始された後、前記転送手段によるデータ転送に必要なサイクル数だけ次の命令の実行開始を待機することを特徴とする付記6記載のキャッシュメモリ制御装置。

【0076】

(付記9) データを使用した演算処理を実行するプロセッサコアと、

50

メインメモリに記憶されるデータを記憶するキャッシュメモリと、
 前記キャッシュメモリの入力部分と出力部分とを接続する接続ラインと、
 前記メインメモリのデータを前記キャッシュメモリに入力する場合に、当該データを前記接続ライン経由で前記キャッシュメモリの出力部分へ転送させる制御手段と、
 前記制御手段の制御により前記キャッシュメモリの出力部分へ転送されたデータを前記プロセッサコアへ転送するデータバスと
 を有することを特徴とする半導体集積回路。

【0077】

(付記10) 前記プロセッサコアを複数有し、
 前記キャッシュメモリは、
 それぞれ記憶するデータのアドレスが異なる前記メインメモリの複数のバンクに対応して分割されることを特徴とする付記9記載の半導体集積回路。

10

【0078】

(付記11) データを使用した演算処理を実行する演算処理手段と、主記憶部によって記憶されるデータのうち前記演算処理手段によって使用されるデータを記憶するキャッシュメモリとを備えるキャッシュメモリ制御装置におけるキャッシュメモリ制御方法であって、

主記憶部のデータを前記キャッシュメモリの入力部分から入力して前記キャッシュメモリに書き込む書込ステップと、

前記書込ステップにて前記キャッシュメモリに書き込まれたデータを前記キャッシュメモリの入力部分から前記キャッシュメモリの出力部分へ転送する第1転送ステップと、

前記第1転送ステップにて転送されたデータを前記キャッシュメモリの出力部分から前記演算処理手段へ転送する第2転送ステップと

を有することを特徴とするキャッシュメモリ制御方法。

20

【図面の簡単な説明】

【0079】

【図1】一実施の形態に係る半導体集積回路の概略構成を示す図である。

【図2】一実施の形態に係る半導体集積回路の要部構成を示す図である。

【図3】一実施の形態に係るキャッシュメモリ制御装置における接続関係を示す図である

30

【図4】一実施の形態に係るキャッシュミス時の動作を示すシーケンス図である。

【図5】一実施の形態に係るレイテンシの短縮を説明する図である。

【図6】一実施の形態に係るパイプライン処理の一例を示す図である。

【図7】情報処理装置の構成の一例を示す図である。

【図8】情報処理装置におけるレイテンシを説明する図である。

【符号の説明】

【0080】

100 半導体集積回路

101、101-0、101-1 コア

102 制御部

103 WBDQ

104 MIDQ

105、107、108、111 セレクタ

106 データメモリ

109 データバス

110 MIポート

112 タグメモリ

113 MIバッファ

114 MODQ

115 MAC

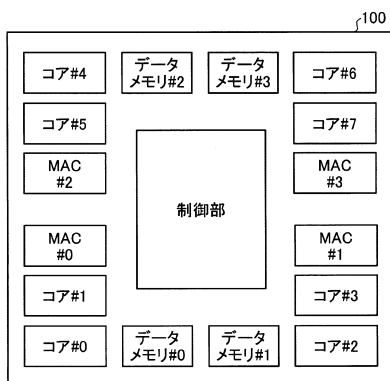
40

50

L 0 ライン

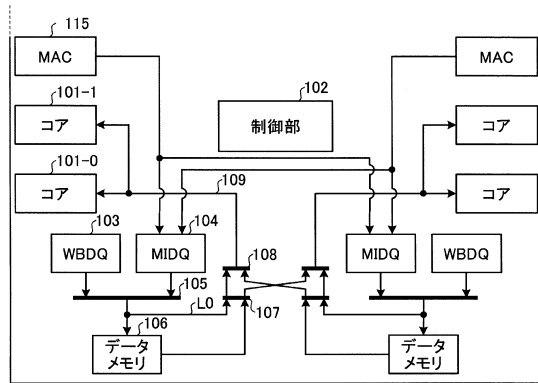
【 図 1 】

一実施の形態に係る半導体集積回路の概略構成を示す図



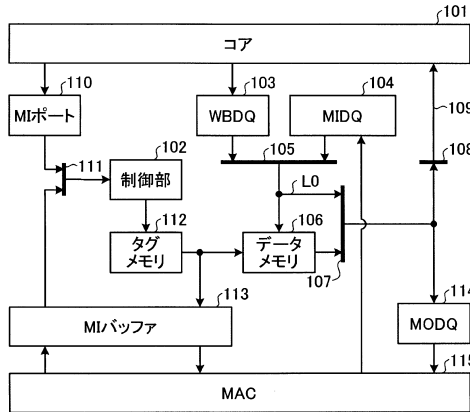
【 図 2 】

一実施の形態に係る半導体集積回路の要部構成を示す図



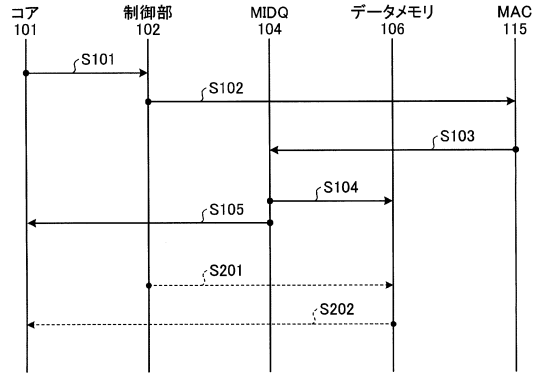
【 図 3 】

一実施の形態に係るキャッシュメモリ制御装置における接続関係を示す図



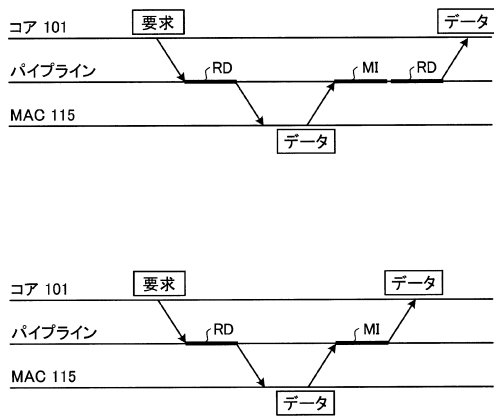
【 図 4 】

一実施の形態に係るキャッシュミス時の動作を示すシーケンス図



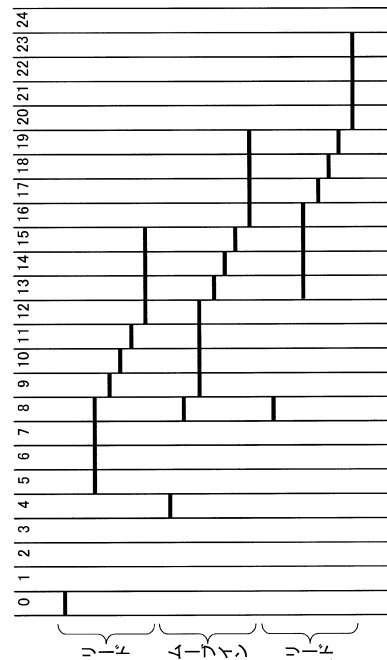
【 図 5 】

一実施の形態に係るレイテンシの短縮を説明する図



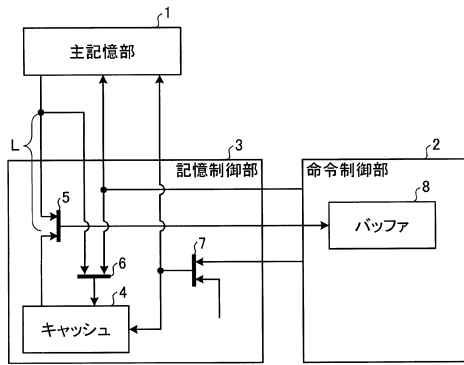
【 図 6 】

一実施の形態に係るパイプライン処理の一例を示す図



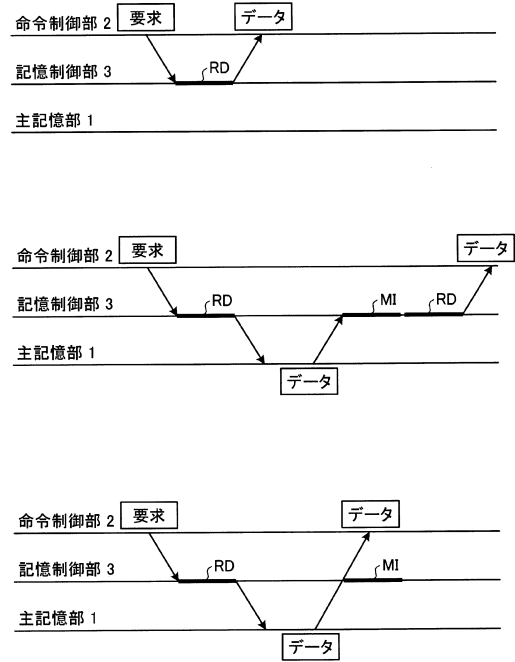
【図7】

情報処理装置の構成の一例を示す図



【図8】

情報処理装置におけるレイテンシを説明する図



フロントページの続き

(51)Int.Cl.

F I

G 0 6 F 12/08 5 7 3

合議体

審判長 和田 志郎

審判官 山田 正文

審判官 千葉 輝久

- (56)参考文献 特開平04 - 155525 (JP, A)
国際公開第2004/068361 (WO, A1)
特開2002 - 140232 (JP, A)
特開2007 - 156963 (JP, A)
特開昭59 - 71184 (JP, A)
特開2008 - 107983 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F12/08-12/12