

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4286826号
(P4286826)

(45) 発行日 平成21年7月1日 (2009.7.1)

(24) 登録日 平成21年4月3日 (2009.4.3)

(51) Int. Cl.

F I

G 0 6 F 13/14 (2006.01)

G 0 6 F 12/08 (2006.01)

G 0 6 F 13/14 3 1 0 H

G 0 6 F 12/08 5 3 1 B

G 0 6 F 12/08 5 7 5

G 0 6 F 12/08 5 5 9 D

G 0 6 F 12/08 5 7 7

請求項の数 19 (全 23 頁) 最終頁に続く

(21) 出願番号 特願2005-300767 (P2005-300767)
 (22) 出願日 平成17年10月14日 (2005.10.14)
 (65) 公開番号 特開2006-120147 (P2006-120147A)
 (43) 公開日 平成18年5月11日 (2006.5.11)
 審査請求日 平成17年10月14日 (2005.10.14)
 (31) 優先権主張番号 60/619,069
 (32) 優先日 平成16年10月15日 (2004.10.15)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 395015319
 株式会社ソニー・コンピュータエンタテインメント
 東京都港区南青山二丁目6番21号
 (73) 特許権者 390009531
 インターナショナル・ビジネス・マシーンズ・コーポレーション
 INTERNATIONAL BUSINESS MACHINES CORPORATION
 アメリカ合衆国10504 ニューヨーク州 アーモンク ニュー オーチャードロード
 (74) 代理人 100099324
 弁理士 鈴木 正剛

最終頁に続く

(54) 【発明の名称】 マルチプロセッサシステムで多重構成をサポートする方法及び装置

(57) 【特許請求の範囲】

【請求項1】

マルチプロセッサであって、

1つ以上の通信バスを介して動作可能となるように相互に結合された複数のプロセッサと、

コンフィギュラブルなインターフェース回路と、を有し、

前記コンフィギュラブルなインターフェース回路は、第1インターフェースと第2インターフェースとを含み、これら第1及び第2インターフェースは、それぞれ独立して、(i) 前記マルチプロセッサと他のマルチプロセッサとの相互接続を行うことができ、かつ、前記マルチプロセッサの1つ以上のメモリと、前記他のマルチプロセッサの1つ以上のメモリと、の間のキャッシュのコヒーレンシーを維持することができる、コヒーレントな対称型インターフェースを提供する第1モードで動作するか、又は (i i) 前記マルチプロセッサと1つ以上の外部デバイスとを相互接続することができ、かつ、マルチプロセッサに少なくとも何らかのメモリ保護を与える、非コヒーレントなインターフェースを提供する第2モードで動作するように構成されている、マルチプロセッサ。

【請求項2】

前記コンフィギュラブルなインターフェース回路は、論理層と、トランスポート層と、物理層とを含む、請求項1記載のマルチプロセッサ。

【請求項3】

前記論理層は、前記第1モードで動作するためのコヒーレンシー・ルールと、前記第2

10

20

モードで動作するためのオーダリング・ルールとを定義するように構成されており、

前記トランスポート層は、前記マルチプロセッサと、前記1つ以上の外部デバイスと、の間で送信を行うためのコマンドとデータパケットの構成を定義するように構成されており、かつ、

前記物理層は、メモリ・アクセス・コマンドと、メモリ・スヌープ・リクエストと、前記マルチプロセッサと前記1つ以上の外部デバイスとの間のデータ伝送と、のタイミング及び電気的特徴を定義するように構成されている、請求項2記載のマルチプロセッサ。

【請求項4】

前記コンフィギュラブルなインターフェース回路は、メモリ・アクセス・コマンドと、メモリ・スヌープ・リクエストと、前記マルチプロセッサと前記1つ以上の外部デバイスとの間のデータ伝送を促進するよう動作可能である、請求項1記載のマルチプロセッサ。

【請求項5】

前記メモリ・アクセス・コマンドと、メモリ・スヌープ・リクエストと、データ伝送は、非同期の個別パケット形式である、請求項4記載のマルチプロセッサ。

【請求項6】

前記パケットは、所望のトランザクションを定義するアドレス情報と制御情報とを含む、請求項5記載のマルチプロセッサ。

【請求項7】

前記第1インターフェース及び第2インターフェースの物理層入力／出力帯域幅は、これらのインターフェースの各帯域幅の合計が、前記インターフェース回路の帯域幅の合計を超えない範囲で分割可能である、請求項1記載のマルチプロセッサ。

【請求項8】

各々が1つ以上の通信バスを介して、動作可能となるように相互に結合された複数のプロセッサを含む、1つ以上のマルチプロセッサと、

コンフィギュラブルなインターフェース回路と、を有し、

前記コンフィギュラブルなインターフェース回路は、第1インターフェースと第2インターフェースとを含み、これら第1及び第2インターフェースは、それぞれ独立して、(i) 前記マルチプロセッサと他のマルチプロセッサとの相互接続を行うことができ、かつ、前記マルチプロセッサの1つ以上のメモリと、前記他のマルチプロセッサの1つ以上のメモリと、の間のキャッシュのコヒーレンシーを維持することができる、コヒーレントな対称型インターフェースを提供する第1モードで動作するか、又は(ii) 前記マルチプロセッサと1つ以上の外部デバイスとを相互接続することができ、かつ、マルチプロセッサに少なくとも何らかのメモリ保護を与える、非コヒーレントなインターフェースを提供する第2モードで動作するように構成されている、システム。

【請求項9】

前記第2モードで動作する前記マルチプロセッサの前記第1インターフェースを介して、前記マルチプロセッサのうちの1つと結合されている第1外部デバイスと、

前記第2モードで動作する前記マルチプロセッサの前記第2インターフェースを介して、前記マイクロプロセッサのうちの1つと結合されている第2外部デバイス、とを更に含む、請求項8記載のシステム。

【請求項10】

各マルチプロセッサの前記第1モードで動作する第1インターフェースを介して相互接続された少なくとも2つのマルチプロセッサと、

前記第2モードで動作する前記マルチプロセッサの前記第2インターフェースを介して、前記少なくとも2つのマルチプロセッサのうちのひとつと結合されている第1外部デバイスと、

前記第2モードで動作する前記マルチプロセッサの前記第2インターフェースを介して、前記少なくとも2つのマルチプロセッサのうちの他のマルチプロセッサと結合されている第2外部デバイスとを更に有す、請求項8記載のシステム。

【請求項11】

前記第 1 モードで動作する前記マルチプロセッサの第 1 インターフェースを介して相互接続された、前記マルチプロセッサのうちの第 1 及び第 2 マルチプロセッサを有し、

前記マルチプロセッサのうちの前記第 1 プロセッサ及び第 3 マルチプロセッサを有し、前記第 1 プロセッサ及び前記第 3 のマルチプロセッサは、それぞれ前記第 1 モードで動作する第 2 インターフェースと第 1 インターフェースを介して相互接続されており、

前記第 2 及び第 3 マルチプロセッサの前記第 2 インターフェースは、1 つ以上の外部デバイスと相互接続するように動作可能である、請求項 8 記載のシステム。

【請求項 1 2】

前記第 2 モードで動作する前記マルチプロセッサの前記第 2 インターフェースを介して、前記第 2 及び第 3 マルチプロセッサのうちの一方と結合されている第 1 外部デバイスと、

10

前記第 2 モードで動作する前記マルチプロセッサの前記第 2 インターフェースを介して、前記第 2 及び第 3 マルチプロセッサのうちの他方と結合されている第 2 外部デバイスとを更に含む、請求項 1 1 記載のシステム。

【請求項 1 3】

マルチ・ポートのデータ・スイッチと、

前記第 1 モードで動作する前記マルチプロセッサの第 1 インターフェースを介して、前記スイッチと結合されている複数のマルチプロセッサとを更に含む、

前記マルチプロセッサの前記第 2 インターフェースは、1 つ以上の外部デバイスと相互接続するように動作可能である、請求項 8 記載のシステム。

20

【請求項 1 4】

前記第 2 モードで動作する前記マルチプロセッサの前記第 2 インターフェースを介して、前記マルチプロセッサのうちの 1 つと結合されている、少なくとも 1 つの外部デバイスを更に含む、請求項 1 3 記載のシステム。

【請求項 1 5】

前記第 1 インターフェース及び第 2 インターフェースの物理層入力／出力帯域幅は、これらのインターフェースの各帯域幅の合計が、前記インターフェース回路の帯域幅の合計を超えない範囲で分割可能である、請求項 1 記載のマルチプロセッサ。

【請求項 1 6】

1 つ以上の通信バスを介して動作可能となるように相互に結合された複数のプロセッサを用意するステップと、

30

第 1 インターフェースと第 2 インターフェースとを含むとともに、これら第 1 及び第 2 インターフェースは、それぞれ独立して、(i) コヒーレントな対称型インターフェースを提供する第 1 モードで動作するように、又は、(i i) 非コヒーレントなインターフェースを提供する第 2 モードで動作するものであるインターフェース回路を構成するステップとを含む、

前記コヒーレントな対称型インターフェースは、前記マルチプロセッサと他のマルチプロセッサとの相互接続が可能で、この相互接続は、前記マルチプロセッサの 1 つ以上のメモリと、前記他のマルチプロセッサの 1 つ以上のメモリと、の間のキャッシュのコヒーレンシーを維持することができるものであり、かつ、

40

前記非コヒーレントなインターフェースは、前記マルチプロセッサと 1 つ以上の外部デバイスとの相互接続を、前記マルチプロセッサに少なくとも何らかのメモリ保護を与えることができるように行うことができるものである、方法。

【請求項 1 7】

少なくとも 1 つの外部デバイスを前記インターフェースと結合するステップを更に含む、請求項 1 6 記載の方法。

【請求項 1 8】

メモリ・アクセス・コマンド、メモリ・スヌープ・リクエスト、及び／又は前記プロセッサと前記 1 つ以上の外部デバイスとの間のデータ伝送、のサービスを行うステップを更に有する、請求項 1 7 記載の方法。

50

【請求項 19】

前記第1インターフェース及び第2インターフェースの物理層入力／出力帯域幅は、これらのインターフェースの各帯域幅の合計が、前記インターフェース回路の帯域幅の合計を超えない範囲で分割可能である、請求項16記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はマルチプロセッサシステムアーキテクチャを用いて、マルチプロセッシング構成を実現するための方法及び装置に関する。

【背景技術】

10

【0002】

最先端のコンピュータアプリケーションは、リアルタイムのマルチメディア機能を伴っているために、近年はデータスループットが高くより高速なコンピュータが常に望まれている。グラフィックアプリケーションは処理システムへの要求が大きいアプリケーションの1つであり、その理由は、グラフィックアプリケーションが所望のビジュアル結果を実現するために、比較的短時間で非常に多くのデータアクセス、データの演算処理、及びデータ操作を要求するからである。これらのアプリケーションは、1秒間に数千メガビットのデータ処理等の非常に高速な処理速度を要求する。シングルプロセッサを採用し、高速の処理速度を実現している処理システムもある一方で、マルチプロセッサアーキテクチャを利用して実装されている処理システムもある。マルチプロセッサシステムでは、複数の

20

【発明の開示】

【発明が解決しようとする課題】

【0003】

マルチプロセッシングシステムには、処理のスループットや汎用性を高めるために、マトリックス構成においてインターフェースを介した相互接続を検討しているものもある。このような構成は米国特許公開公報2005/0097231号と、米国特許第6,526,491号とに開示されており、その開示の全てが本明細書に参照として組み込まれる。これらの文書に開示されている技術は様々なアプリケーションにおいて利用されうるが、これらの技術では、その他のアプリケーションで所望される柔軟性、及び／又は、プログラマビリティを与えることはできない。

30

【0004】

従って、1つ以上のマルチプロセッサシステムを、1つ以上の外部デバイスと相互接続し、高い処理機能を実現するための新たな方法や装置に対する技術が必要とされている。

【課題を解決するための手段】

【0005】

本発明の1つ以上の態様によれば、プロセッシングエレメント（PE：Processing Element,なお、PEでは複数の異なるパラレルプロセッサが採用されている）は、ブロードバンドインターフェースコントローラ（BIC：Broadband Interface Controller）を備えており、該BICは他のPEやメモリサブシステム、スイッチ、ブリッジチップなどを取り付けるために、コヒーレントの、あるいは非コヒーレントの高性能の相互接続を提供する。BICは種々のシステム要件を満たすよう、様々なプロトコルや帯域幅を、2つのフレキシブルなインターフェースに提供する。インターフェースは、2つのI/Oインターフェース（IOIF0/1）として、あるいはI/OとコヒーレントSMPインターフェース（IOIF及びBIF）のいずれかとして構成されうる。BICがコヒーレントSMPインターフェースとして動作するように設定されている場合、BICは高性能でコヒーレントな相互接続をPEに提供する。BICがI/Oインターフェースとして動作するように設定されている場合、BICはPEに高性能（非コヒーレント）の相互接続を提供する。

40

【0006】

50

B I Cは論理層、トランスポート層、データリンク層、及び物理リンク層を有している。論理層（及び、実施形態によってはトランスポート層）は、コヒーレントS M Pインターフェース（B I F）と非コヒーレントインターフェース（I O I F）間のB I Cの動作を変更するように構成されうる。論理層はオーダリングやコヒーレントルールを含む、B I F又はI O I Fの基本動作を定義する。トランスポート層はデバイス間にコマンドやデータパケットがどのように転送されるかを定義する。コマンドやデータパケットは好ましくは、データリンク層へ送るために、物理層群（P L G：Physical Layer Groups）と呼ばれる小さなユニットに分けられる。データリンク層は送信側と受信側の間に（実質的に）情報を間違いなく確実に送信する機構を定義する。物理層はI/Oドライバの電気的特徴やタイミングを定義し、また、データリンクエンベロープが物理層を通じてどのように送信されるかを記述する。物理リンク層は好ましくは、2セットまでの論理／トランスポート／データリンク層の同時並行処理をサポートし、また、その2つの間の物理層の、利用可能な帯域幅の割当てが設定可能な方法をサポートする。

10

【0007】

B I Cの論理層、トランスポート層、データリンク層、及び物理層の機能や動作は好ましくは以下の通りである。インターフェースの帯域幅の合計が物理層の最大帯域幅を超えない範囲で、物理層の帯域幅が2つのインターフェース間に分割される。一例では、物理層の出力帯域幅自体の合計値は35GB/s、入力帯域幅自体の合計値は25GB/sとなり得る。

【0008】

20

本発明の1つ以上の更なる実施形態によれば、B I Cのフレキシブルなインターフェースにより、1つ以上のプロセッサエレメントが配置されるシステム構成を実質的にフレキシブルなものとし得る。例えばB I Cは、P Eと2つのデバイス間に対応の非コヒーレントインターフェースを設けるために、デュアルI/Oインターフェース（I O I F 0及びI O I F 1）を実装するように動作することができる。B I Cの物理層入力／出力帯域幅は、2つのインターフェースの合計値が物理層の帯域幅の合計値を超えない範囲で、2つのI O I Fインターフェース間に分割されうる（例：出力30GB/s、入力25GB/s）。

【0009】

別の実施形態によれば、2つのプロセッサエレメントはコヒーレント対称型マルチプロセッサ（S M P：symmetric multiprocessor）インターフェース（又は、B I F）構成において、その対応のB I Cを採用している各々により、カスケード接続されうる。各プロセッシングエレメントのコヒーレントS M Pインターフェース（B I F）は、その間にコヒーレントインターフェースを設けるために相互に接続されうる。各プロセッシングエレメントのI O I Fは、非コヒーレントに他のデバイスとデータを送受信する。同様に、各B I Cの物理層入力／出力帯域幅が、その2つのインターフェース間に分割されうる。

30

【0010】

更なる別の実施形態によると、2つ以上のプロセッサエレメントが、コヒーレントS M Pインターフェース（B I F）構成において、その対応のB I Cを採用している各々により、カスケード接続されうる。中央に配置されているプロセッサエレメントは、2つのB I Fを持つB I Cを採用しうる。末端に配置されているプロセッサエレメントのペアは、中央に配置されているプロセッサエレメントを構成し、また、それぞれが1つのB I Fと1つのI O I Fを持つB I Cを採用している。各プロセッシングエレメントのB I Fは、その間にコヒーレントインターフェースを設けるために、相互に接続されうる。端のプロセッシングエレメントのI O I Fは、非コヒーレント法に他のデバイスとのデータを送受信しうる。

40

【0011】

本発明の更なる別の実施形態によれば、2つ以上のプロセッサエレメントは、I/O及びコヒーレントS M Pインターフェース（I O I F及びB I F）構成において、その対応のB I Cを採用している各々により、カスケード接続されうる。各プロセッシングエレ

50

ントのコヒーレントSMPインターフェースは、プロセッシングエレメントを相互に効果的に結合するスイッチと結合され、その間にコヒーレントインターフェースを設けることができる。各プロセッシングエレメントのIOIFは非コヒーレントにシステムの他のデバイスとのデータの送受信をしよう。

【発明を実施するための最良の形態】

【0012】

添付の図面を参照しつつ、ここに記載する本発明の説明を読めば、他の態様、機能および利点などは当業者に自明となるであろう。

【0013】

本発明を説明するために、現在の好ましい形態を図面の形式に示すが、本発明は図示したとおりの構成ならびに手段に限定されないことを理解されたい。

【0014】

以下に本明細書で説明している1つ以上の特徴を実行するのに適した、マルチプロセッサシステムのための好ましいコンピュータアーキテクチャを説明する。1つ以上の実施形態によれば、マルチプロセッサシステムは、ゲームシステム、家庭用端末、PCシステム、サーバーシステム、及びワークステーションなどのメディアリッチアプリケーションを、スタンドアローン処理、及び／又は分散処理するために動作することができる、シングルチップソリューションとして実装されうる。ゲームシステムや家庭用端末などのいくつかのアプリケーションでは、リアルタイムの演算処理は必須である。例えば、リアルタイムの分散ゲームアプリケーションでは、ユーザーにリアルタイムの経験をしていると思わせる程速く、1つ以上のネットワークイメージの復元、3Dコンピュータグラフィック、オーディオ生成、ネットワーク通信、物理的シミュレーション、及び人工知能処理が実行される必要がある。従って、マルチプロセッサシステムの各プロセッサは、短時間で、かつ予測可能時間でタスクを完了する必要がある。

【0015】

このために、また、本コンピュータアーキテクチャによれば、マルチプロセッシングコンピュータシステムの全プロセッサは、共通の演算モジュール（あるいはセル）から構成される。この共通の演算モジュールは、構造が一貫しており、また好ましくは、同じ命令セットアーキテクチャを採用している。マルチプロセッシングコンピュータシステムは、1つ以上のクライアント、サーバー、PC、モバイルコンピュータ、ゲームマシン、PD 30 A、セットトップボックス、電気器具、デジタルテレビ、及びコンピュータプロセッサを使用する他のデバイスから形成されうる。

【0016】

複数のコンピュータシステムもまた、所望に応じてネットワークのメンバーとなりうる。一貫モジュール構造により、マルチプロセッシングコンピュータシステムによるアプリケーション及びデータの効率的な高速処理が可能になる。またネットワークが採用される場合は、ネットワーク上にアプリケーション及びデータの高速送信が可能にする。この構造はまた、大きさや処理能力が様々なネットワークのメンバーの構築を単純化し、また、これらのメンバーが処理するアプリケーションの準備を単純化する。

【0017】

図1と図2を参照すると、基本的な処理モジュールはプロセッサエレメント(PE) 500である。PE 500はI/Oインターフェース502、プロセッシングユニット(PU) 504、及び複数のサブプロセッシングユニット508、すなわち、サブプロセッシングユニット508A、サブプロセッシングユニット508B、サブプロセッシングユニット508C、及びサブプロセッシングユニット508Dを備えている。なお、好適には、PUとしてパワーPC(PPE:Power PC Element)を、SPUとしてシナジスティックプロセッシングエレメント(SPE:Synergistic Processing Element)を用いる。ローカル(あるいは内部)PEバス512は、データ及びアプリケーションを、PU 504、サブプロセッシングユニット508、及びメモリインターフェース511間に送信する。ローカルPEバス512は、例えば従来のアーキテクチャを備えることができ、又は、 50

パケットスイッチネットワークとして実装されうる。パケットスイッチネットワークとして実装される場合は、更なるハードウェアが必要であるものの、利用可能な帯域幅を増やす。

【0018】

PE500はデジタル論理回路を実装するよう様々な方法を用いて構成されうる。しかしながら、好ましくは、PE500はSOI基板を用いた集積回路として構成でき、あるいは、シリコン基板に相補性金属酸化膜半導体(CMOS:Complementary Metal Oxide Semiconductor)を用いた単一の集積回路とすることも好適な構成である。基板の他の材料には、ガリウムヒ素、ガリウムアルミニウムヒ素、及び、様々なドーパントを採用している他の、いわゆる、III-B化合物を含む。PE500はまた、高速単一磁束量子(RSFQ:Rapid Single-flux-Quantum)論理回路などの超電導デバイスを用いて実装されうる。

10

【0019】

PE500は高帯域のメモリ接続516を介して、共有(メイン)メモリ514と密接に結合するよう構成できる。なお、メモリ514をオンチップ化してもよい。好ましくは、メモリ514はダイナミックランダムアクセスメモリ(DRAM:Dynamic Random Access Memory)であるが、メモリ514は例えば、スタティックランダムアクセスメモリ(SRAM:Static Random Access Memory)、磁気ランダムアクセスメモリ(MRAM:Magnetic Random Access Memory)、光メモリ、ホログラフィックメモリなどとして、他の方法を用いて実装されうる。

20

【0020】

PU504とサブプロセッシングユニット508は好ましくは、それぞれダイレクトメモリアクセス(DMA)の機能を備えたメモリフローコントローラ(MFC:Memory Flow Controller)と結合されており、該コントローラはメモリインターフェース511と共に、PE500のDRAM514とサブプロセッシングユニット508、PU504間のデータ転送を促進する。DMAC及び/又はメモリインターフェース511は、サブプロセッシングユニット508及びPU504に一体化して、別個に配置されうる。更に、DMACの機能及び/又はメモリインターフェース511の機能は、1つ以上の(好ましくは全ての)サブプロセッシングユニット508及びPU504に統合することができる。例えば、DRAM514は、実例で示しているように、チップ外に配置しても、あるいは一体化してオンチップ配置としてもよい。

30

【0021】

PU504はデータ及びアプリケーションをスタンドアロン処理できる標準プロセッサなどでありうる。作動時、PU504は、好ましくは、サブプロセッシングユニットによるデータ及びアプリケーション処理をスケジューリングし、調整を行う。サブプロセッシングユニットは好ましくは、単一命令複数データ(SIMD:Single Instruction Multiple Data)プロセッサである。PU504の管理下、サブプロセッシングユニットは並列で、かつ独立して、これらのデータ及びアプリケーション処理を行う。PU504は好ましくは、RISC(Reduced Instruction Set Computing)技術を採用しているマイクロプロセッサアーキテクチャであるパワーPC(PowerPC)コアを用いて実装される。RISCは、単純な命令の組合せを用いて、より複雑な命令を実行する。従って、プロセッサのタイミングは、単純で高速の動作に基づくものであり、マイクロプロセッサがより多くの命令を所定のクロック速度で実行できるようにする。

40

【0022】

PU504はサブプロセッシングユニット508により、データ及びアプリケーション処理をスケジューリングし調整を行う、メインプロセッシングユニットの役割を果たしているサブプロセッシングユニット508のうちの、1つのサブプロセッシングユニットにより実装されうる。更に、プロセッサエレメント500内には1つ以上の実装されたPUが存在しうる。なお、オンチップのPUを複数設けるようにしてもよい。

【0023】

50

本モジュール構造によれば、特定のコンピュータシステムにおけるPE500の数は、そのシステムが要求する処理能力に基づく。例えば、サーバーにおけるPE500の数は4、ワークステーションにおけるPE500の数は2、PDAにおけるPE500の数は1とすることができる。特定のソフトウェアセルの処理に割当てられるPE500のサブプロセッシングユニット数は、セル内のプログラムやデータの複雑度や規模により決定される。このように、PEはモジュール構造を有していることから拡張性が高く、搭載するシステムのスケール、パフォーマンスに応じて容易に拡張することができる。

【0024】

モジュラーインターコネクトバス(MIB:Modular Interconnect Bus)512はコヒーレントバスであり、それぞれが複数の同時データ転送をサポートする多数の(ハーフレート)リングとして構成される。

10

【0025】

MIC511は、PEと共有メモリ514を実装する複数のメモリバンク間の通信を促進するように動作することが出来る。MIC511は好ましくは、プロセッサ及びI/Oインターフェースに対して、非同期的に動作する。

【0026】

BIC513はMIB512を論理的拡張部であり、このBIC513によってMIB512とI/Oインターフェース502との間に非同期の相互接続を提供する。BIC513は、他のPE、メモリサブシステム、スイッチ、ブリッジチップなどを取り付けるために、コヒーレントあるいは非コヒーレントの高性能の相互接続を提供する。BIC513は、種々のシステム要件を満たすよう、様々なプロトコルや帯域幅を、2つのフレキシブルなインターフェースに提供する。

20

インターフェースは、2つのI/Oインターフェース(IOIF0/1)として、あるいはI/OとコヒーレントSMPインターフェース(IOIF及びBIF)のいずれかとしてとして構成されうる。フレキシブルなインターフェースは7送信バイト、5受信バイトで動作する。BICがコヒーレントSMPインターフェースとして動作するように構成されている場合、BICは高性能でコヒーレントな相互接続をPEに提供する。BIC513がI/Oインターフェースとして動作するように構成されている場合、BIC513は高性能(非コヒーレント)の相互接続をPEに提供する。(BIFあるいはIOIFとして動作中の)BIC513は、高速インターフェースを要求する他のアプリケーションにも使用されうる。

30

【0027】

BIC513はPEと他のPE、メモリサブシステム、スイッチ、ブリッジチップなどとの間のトランザクションの実施を促進する。BIFやIOIFトランザクションは通常、メモリアクセスリクエスト(データに対するリクエスト)である。メモリアクセスリクエストは、PE内のローカルキャッシュ階層あるいはPEと接続されている外部デバイスによってはサービスすることのできないデータトランザクションの結果もたらされるものである。メモリアクセスリクエストは、1つ以上のトランザクションを要求しうる。トランザクションはマスタデバイス又はキャッシュコヒーレントコントローラ(スヌープ)により開始され、マスタとスレーブ間に一連のパケット転送をもたらしうる。BIF及びIOIFトランザクションは3つのフェーズに分けられる。即ち、それらはコマンド(ロードやストアなど)、スヌープ、及びデータ(しかしながら全トランザクションがデータフェーズを要求するわけではない)の各フェーズである。

40

【0028】

BIC513のBIFやIOIFの特徴は、多くの異なるシステム構成や次世代のコンプライアンスプロセッサをサポートするために、スケーラブルでフレキシブルに動作することである。BIC513の特徴には、

(i) キャッシュのコヒーレンシーとデータの同期化をサポートするパケットプロトコル(BIFとして動作している場合)、

(ii) オーダリングとコヒーレンシーのためのフラッグを備えたパケットプロトコル(

50

I O I Fとして動作している場合)、

(iii) 完全にパイプライン化したコマンドトランザクション、データトランザクション、及びレスポンス／応答トランザクション、

(iv) スプリットトランザクション、及び (v) クレジットベースのコマンドやデータのサポート、

が含まれる。

【0029】

図3を参照すると、論理層（少なくとも2つの論理層0、1を含む）、トランスポート層（同様に少なくとも2つのトランスポート層0、1を含む）、データリンク層（同様に少なくとも2つのデータリンク層0、1を含む）、及び物理リンク層を備えている、B I C 5 1 3の1つ以上の態様のブロック図が示されている。論理層（及び実施形態によってはトランスポート層）は、コヒーレントSMPインターフェース（B I F）と非コヒーレントインターフェース（I O I F）間のB I C 5 1 3の動作を変更するように構成される。

10

【0030】

論理層は、オーダリングやコヒーレントルールを含む、B I F又はI O I Fの基本動作を定義する。従って、B I F又はI O I Fを使用するP Eに取り付けられているデバイスは、論理層の仕様に完全に対応している必要がある。しかしながらアプリケーションによっては、論理層の仕様のサブセットを実装しても、なおB I C 5 1 3を介してP Eと動作できるものもある。論理層の情報は、基本コマンド（アドレス）、データ、及び応答パケットの概要を表す。論理層がコヒーレントSMPインターフェースに対して構成されている場合は、スヌープ応答パケットが許容される。論理層が非コヒーレントインターフェースに対して構成されている場合は、応答パケットのみが許容される。

20

【0031】

トランスポート層は、デバイス間にコマンドとデータパケットがどのように転送されるかを定義する。好ましくは、コマンド及びデータパケットは、データリンク層に送るために物理層群（P L G s : Physical Layer Groups）と呼ばれる小さなユニットに分けられる。同様に、トランスポート層は、P L Gの分配ペースの決定に使用される、つまり、P L Gがどのように分配されるかを決定する、フロー制御機構の定義を含む。トランスポート層は好ましくは、システム又はアプリケーションの必要性に合うようにカスタマイズされる。

30

【0032】

データリンク層は、送信機と受信機間に情報を間違いなく確実に送信する機構を定義する。同様に、データリンク層には、物理リンクに対するトレーニングシーケンス又は初期化が含まれる。また、データリンク層は好ましくは、システム又はアプリケーションの必要性に合うようにカスタマイズされる。

【0033】

物理層はI/Oドライバの電気的特徴やタイミングを定義し、また、データリンクエンベロープが物理リンクを通してどのように送信されるかを記述する。好ましくは、物理リンク層は2セットまでの論理／トランスポート／データリンク層の同時並行処理をサポートし、また、その2つの間の物理層の利用可能な帯域幅の割当が設定可能な方法をサポートする。物理層はまた、プリント回路基板（P C B : Printed Circuit Board）のルーティングやパッケージングのガイドラインを定義する。物理層の目的には、I/Oドライバの物理的特徴（速度、単方向性対双方向性、I/O数など）を隠蔽することと、データリンク層に一貫したインターフェースを与えること、がある。入力／出力機能は、実際に帯域幅のサポートが可能な、R a m b u s R R A C I/Oを用いて実現される。フレキシビリティを高めるために、R R A Cの送信機と受信機はプロセッサとメモリに対して非同期的に動作し、利用可能な帯域幅をその2つのインターフェース間に設定可能としている。

40

【0034】

50

上述のB I Cの論理層、トランスポート層、データリンク層、及び物理層の機能や動作を考えると、別のシステム構成と同じく、相対的に高いP Eの帯域幅要件がサポートされる。例えば、物理層はペアにつき5 G B / sで実行するように、また、出力帯域幅自体の合計値が3 5 G B、入力帯域幅自体の合計値が2 5 G Bを有するように、動作することができる。物理層の帯域幅は、最大帯域幅が出力3 0 G B / s、入力2 5 G B / sである、2つのインターフェース間に分割されうる。各インターフェースの帯域幅は、5 G B / sインクリメントで構成されうる。好ましくは、2つのインターフェースの合計値は、物理層の帯域幅の合計値を超えることはできない。

【0035】

B I C 5 1 3のインターフェースの更なる詳細を以下に説明する。B I C 5 1 3はP E、メモリサブシステム、スイッチ、ブリッジチップ等の間のポイントツーポイントバスであり、また、M I B 5 1 2の論理的拡張である。B I C 5 1 3はブリッジチップやスイッチを備えた多くのデバイスの取付けをサポートする。単一の物理デバイスは複数のデバイスタイプのタスクを実行するように動作することができる。これらのデバイスタイプとしては、マスタ、スヌーパ、メモリ、バスアダプタ、I / Oブリッジ、が挙げられる。

マスタは、例えば、コマンドバスを調停し駆動するバスデバイスなどであり、スヌーパは、例えば、他のシステムのキャッシュと、キャッシュデータのコヒーレントを維持するために、コマンドバス上の動きを監視するバスデバイスなどである。バスアダプタあるいはI / Oブリッジはキャッシュを有することができ、その場合は、スヌーパのように機能し、例えばキャッシュデータと他のシステムキャッシュ間のコヒーレントを維持する。

スレーブは、例えば、メモリの読出しあるいは書込みコマンドに応答するバスデバイスなどである。スレーブはメモリ、あるいはI / Oレジスタ、あるいはその両方を有することが出来る。メモリデバイスはスレーブの一例である。

メモリは、例えば、メモリの読出しあるいは書込みに応答し、コヒーレントオペレーションに対する肯定応答を処理するバスデバイスなどである。メモリの一部がリモートバスに取り付けられている場合、バスアダプタはそのリモートメモリ空間へのメモリアクセスに対して、メモリとしての役割を果たす。

バスアダプタは、例えば、他のバスへのゲートウェイなどであり、同一の、あるいは異なるバスアーキテクチャを有し、また、好ましくはリターンプロトコル（あるいは再実行プロトコル）を用いて、コヒーレントオペレーションをリモートバスへ送る。

I / Oブリッジは、例えば、I / Oバスへのゲートウェイなどであり、排他状態あるいは変更状態においてデータをキャッシュしない。ブリッジではI / Oバスにコヒーレンシーを与えないであろう。しかしながら、ブリッジは、好ましくは、I / Oデバイスにより共有状態でキャッシュされたデータに対してI / Oディレクトリを有し、従って、I / Oバスへコヒーレントオペレーションを送るために再実行プロトコルを使用しない。ブリッジはプログラムI / O（P I O）あるいはメモリマップI / Oデバイスをサポートしうる。

【0036】

B I C 5 1 3のアーキテクチャは好ましくは、別々のコマンド、データ、及び（スヌープ）応答パケットに基づく。好ましくは、これらのパケットは独立して実行され、コマンドパケットがデータパケットに先行することが好ましい場合を除いて、コマンドパケットとデータパケット間に時差相関がないスプリットランザクションを可能にする。リクエスト及び応答はタグ付けされ、アウトオブオーダー応答を可能にする。このアウトオブオーダー応答は、他のバスへのI / Oや、又は非一様メモリアクセス（N U M A : Non-Uniform Mmemory access）環境では一般的である。

【0037】

コヒーレントS M F構成では、コマンドパケットはB I F上で実施されるランザクションを説明するアドレス及び制御情報を有している。アドレスコンセンレータはコマンドパケットを受信し、処理するコマンドの順番を決定し、コマンドを選択する。選択されたコマンドパケットは、マスタデバイスにより転送コマンド形式でB I Fのスレーブデ

10

20

30

40

50

バイスへ転送（送信）される。反映コマンドパケットの受信後、スレーブはスヌープ応答パケット形式でマスタへ応答を送る。スヌープ応答パケットは反映コマンドパケットの受入あるいは拒絶を示す。場合によっては、スレーブはトランザクションに対する最終送信先ではない。これらの場合、スレーブは最終送信先へリクエストを送る責任があり、また、スヌープ応答パケットを生成しない。一般に、コマンドパケットはデータトランザクションに対するリクエストである。コヒーレントの管理や同期化などのリクエストに対して、コマンドパケットはコンプリートトランザクションである。リクエストがデータトランザクションに対するものである場合、制御情報及びリクエストデータを有するデータパケットが、マスタとスレーブ間で転送される。トランスポート層の定義に応じて、コマンドやデータパケットがB I F上の両デバイスにより、同時に送受信されうる。

10

【0038】

非コヒーレント構成では、コマンドパケットは好ましくはI O I F上に実行されるトランザクションを説明するアドレス及び制御情報を有している。コマンドパケットはI O I Fコマンド形式でマスタによりI O I F上のスレーブデバイスへ送られる。コマンドパケットの受信後、スレーブはI O I F応答パケット形式でマスタへ応答を送る。応答パケットはI O I Fコマンドパケットの受入あるいは拒絶を示す。場合によっては、スレーブはトランザクションに対する最終送信先ではないこともある。これらの場合に対して、スレーブは最終送信先へリクエストを送るよう応答できる。一般にI O I Fコマンドパケットはデータトランザクションに対するリクエストである。割込みリクエストや割込み再送信オペレーションリクエストなどに対しては、コマンドパケットは、コンプリートトランザクションである。リクエストがデータトランザクションに対するものである場合、制御情報やリクエストデータを含むデータパケットはマスタとスレーブ間に転送される。トランスポート層の定義に応じて、コマンドやデータパケットが、I O I Fの両デバイスにより同時に送受信されうる。

20

【0039】

B I C 5 1 3はM I BとI / Oインターフェース間に非同期インターフェースを提供する。これによりB I Cは、速度一致S R A Mバッファ、論理、及び3つのクロックドメインを有する。プロセッサ側はハーフレートで動作し、I / O側はR R A Cの3分の1の速度で、また小さなディストリビューションネットワークはR R A Cの半分の速度で動作する。送信機と受信機が高速であるために、R R A CとB I C 5 1 3は較正を必要とする。B I C 5 1 3の較正には、インターフェースを備えているバイト間の歪みをなくすために、エラスティックバッファが用いられる。

30

【0040】

上述のように、B I C 5 1 3は2つの柔軟なインターフェース、即ち（i）デュアルI / Oインターフェース（I O I F 0 / 1）、及び（i i）I / O及びコヒーレントS M Pインターフェース（I O I F及び& B I F）を提供する。これにより、1つ以上のプロセッサエレメントが配置されるシステム構成が非常にフレキシブルになる。

【0041】

例えば、図4に例示しているように、B I C 5 1 3はP E 5 0 0と2つのデバイス、つまりデバイス0とデバイス1間にそれぞれの非コヒーレントインターフェースを設けるために、デュアルI / Oインターフェース（I O I F 0及びI O I F 1）を実装することができる。この構成では、単一のP E 5 0 0がI O I F 0とI O I F 1のそれぞれとデータの送受信をし得る。

40

【0042】

上述のように、B I C 5 1 3の物理層入力／出力帯域幅は、2つのインターフェースの合計が、物理層の帯域幅の合計（出力30GB/s、入力25GB/s、など）を超えない限りは、2つのインターフェース（B I F-B I F、I O I F-I O I F、及び／又はB I F-I O I F）間に分割されうる。デバイス0がグラフィックエンジンなどのスループットが高いデバイス、デバイス1がI / Oブリッジなどのスループットが低いデバイスであると仮定すると、B I C 5 1 3の帯域幅は、適切な構成を実現するために最新の方法

50

で分割されうる。例えば、グラフィックエンジン（デバイス0）へのI/OIF0非コヒーレントインターフェースは、出力30GB/s、入力20GB/sであることができ、一方でI/Oブリッジ（デバイス1）へのI/OIF1非コヒーレントインターフェースは（2.5GB/sのインクリメントが可能であると仮定すると）出力2.5GB/s、入力2.5GB/sでありうる。

【0043】

図5に例示しているように、2つのプロセッサエレメント500は、コヒーレントSMPインターフェース（BIF）構成において、その対応のBIC513を採用している各々によりカスケード接続されうる。各プロセッシングエレメント500のコヒーレントSMPインターフェース（BIF）は、その間にコヒーレントインターフェースを設けるために相互に接続される。各プロセッシングエレメント500のI/OIFは、非コヒーレントに他のデバイスとのデータの送受信をしうる。

10

【0044】

同様に、BIC513の物理層入力/出力帯域幅がその2つのインターフェース間に分割されうる。デバイス0とデバイス1が、I/Oブリッジなどの相対的にスループットが低いデバイスであると仮定すると、それぞれのBIC513の帯域幅は適切な構成を実現するために、最新の方法で分割されうる。例えば、デバイス0へのI/OIF0非コヒーレントインターフェースは、出力5GB/s、入力5GB/s、デバイス1へのI/OIF1非コヒーレントインターフェースは、出力5GB/s、入力5GB/s、また、プロセッシングエレメント500間のコヒーレントBIFインターフェースは、入力20GB/s、出力20GB/sでありうる。

20

【0045】

図6に例示しているように、2つ以上のプロセッサエレメント500は、コヒーレントSMPインターフェース（BIF）構成において、その対応するBIC513を採用している各々によりカスケード接続されうる。中央のプロセッサエレメント500は、2つのBIFを有するBIC513を採用している。各プロセッシングエレメント500のBIFは、その間にコヒーレントインターフェースを設けるために、相互に結合されている。端のプロセッシングエレメント500のI/OIFは非コヒーレントに他のデバイスとのデータの送受信を行う。

30

【0046】

図7に示すように、2つ以上のプロセッサエレメント500は、I/O及びコヒーレントSMPインターフェース（I/OIF及びBIF）構成で、その対応するBIC513を採用している各々によりカスケード接続されうる。各プロセッシングエレメント500のコヒーレントSMPインターフェース（BIF）は、プロセッシングエレメント500を相互に効果的に結合し、その間にコヒーレントインターフェースを設けるスイッチと結合されうる。各プロセッシングエレメント500のI/OIFは非コヒーレントにシステムの他のデバイスとデータを送受信し得る。

【0047】

SPUは、変換された、また、保護されたコヒーレントDMAを通じてメモリシステムをPPUと共有するが、データや命令は、各SPU専用の256kのローカルストレージ（LS）によりサポートされる、専用の実アドレススペースに格納される。SPUはプロセッサエレメントのコンピュータ性能の多くを提供する。8台のプロセッサの各々は、倍精度浮動小数点以外の全演算に対し完全にパイプライン化されている、128ビット幅の2命令同時発行SIMDデータフローを有している。オペランドは128ビット128エントリの統一されたレジスタファイルにより提供される。各SPUはMIBへのフル帯域幅の同時読出しや書込みDMAアクセス、16バイトのSPUのロード及びストア、及び命令（プレ）フェッチ、をサポートする256kBのシングルポートのLSを有している。SPUは関連のMFCへの有効アドレス（EA:Effective Address）を有するDMAコマンドを発行することにより、メインストレージにアクセスする。MFCはEAへ標準のパワーアーキテクチャアドレス変換を採用し、ローカルストレージとメインストレージ間

40

50

にデータを非同期に転送する。これにより、オーバーラッピング通信と演算処理が出来るようになり、また、リアルタイムの演算を容易にする。DMA、大きなレジスタファイル、及び、標準の順次実行動作を介した、共有メモリへのSPUアクセスは、多目的のストリーミングプログラミング環境を提供する。各SPUは、そのリソースが有効プログラムによってのみアクセスされうるような手法で動作するように動的に構成されうる。

【0048】

図8に一般的なサブプロセッシングユニット（SPU）508の更なる詳細を例示する。SPU508アーキテクチャは好ましくは、多目的プロセッサ（平均して高性能を広範なアプリケーションに実現するように設計されているもの）と、特殊目的プロセッサ（高性能を単一のアプリケーションに実現するように設計されているもの）間の間隙を埋める。SPU508は、ゲームアプリケーション、メディアアプリケーション、ブロードバンドシステムなどに高性能を実現するように、また、リアルタイムアプリケーションのプログラマーに高度な制御を提供するように設計される。SPU508は、グラフィックジオメトリパイプライン、サーフェースサブディビジョン、高速フーリエ変換、画像処理キーワード、ストリームプロセッシング、MPEGのエンコード／デコード、エンクリプション、デクリプション、デバイスドライバの拡張、モデリング、ゲーム物理学、コンテンツ制作、音響合成及び処理が可能である。

【0049】

サブプロセッシングユニット508は2つの基本機能ユニットを有し、それらはSPUコア510A及びメモリフローコントローラ（MFC）510Bである。SPUコア510Aはプログラムの実行、データ操作、などを行い、一方でMFC510BはシステムのSPUコア510AとDRAM514の間のデータ転送に関連する関数を実施する。

【0050】

SPUコア510Aはローカルメモリ550、命令ユニット（IU：Instruction Unit）552、レジスタ554、1つ以上の浮動小数点実行ステージ556、及び1つ以上の固定小数点実行ステージ558を有している。ローカルメモリ550は好ましくは、SRAMなどの、シングルポートのランダムメモリアクセスを用いて実装される。殆どすべてのプロセッサはキャッシュの導入により、メモリへのレイテンシを小さくする一方で、SPUコア510Aはキャッシュより小さいローカルメモリ550を実装している。更に、リアルタイムアプリケーション（及び本明細書に述べているように、他のアプリケーション）のプログラマーたちに一貫した、予測可能なメモリアクセスレイテンシを提供するため、SPU508A内のキャッシュメモリアーキテクチャは好ましくない。キャッシュメモリのキャッシュヒット／ミスという特徴のために、数サイクルから数百サイクルまでの、予測困難なメモリアクセス時間が生じる。そのような予測困難性により、例えばリアルタイムアプリケーションのプログラミングに望ましい、アクセス時間の予測可能性が低下する。DMA転送をデータの演算処理にオーバーラップさせることで、ローカルメモリSRAM550においてレイテンシの隠蔽を実現しうる。これにより、リアルタイムアプリケーションのプログラミングが制御しやすくなる。DMAの転送に関連するレイテンシと命令のオーバーヘッドが、キャッシュミスにサービスしているレイテンシのオーバーヘッドを超過していることから、DMAの転送サイズが十分に大きく、十分に予測可能な場合（例えば、データが必要とされる前にDMAコマンドが発行される場合）に、このSRAMのローカルメモリ手法による利点を得られる。

【0051】

サブプロセッシングユニット508のうちの、所定の1つのサブプロセッシングユニット上で実行しているプログラムは、ローカルアドレスを使用している関連のローカルメモリ550を参照する。しかしながら、ローカルメモリ550のそれぞれの場所はまた、システムのメモリマップ全体内に実アドレス（RA：Real Address）も割当てられる。これにより、プリビレッジソフトウェア（Privilege Software）はローカルメモリ550をプロセスの有効アドレス（EA：Effective Address）にマッピングする、ローカルメモリ550と別のローカルメモリ550間のDMA転送を促進する。PU504はまた、有効

アドレスを用いてローカルメモリ 550 に直接アクセスすることができる。好ましい実施形態では、ローカルメモリ 550 は 556 キロバイトのストレージを有し、またレジスタ 552 の容量は 128×128 ビットである。

【0052】

S P U コア 504 A は、好ましくは、論理命令がパイプライン式で処理される、プロセッシングパイプラインを用いて実装される。パイプラインは命令が処理されるいずれの数のステージに分けられうるが、一般にパイプラインは 1 つ以上の命令のフェッチ、命令のデコード、命令間の依存度チェック、命令の発行、及び、命令の実行ステップを有している。これに関連して、I U 552 は命令バッファ、命令デコード回路、依存度チェック回路、及び命令発行回路、を有する。

10

【0053】

命令バッファは、好ましくは、ローカルメモリ 550 と結合され、また、フェッチされる際に一時的に命令を格納するように動作できる、複数のレジスタを備えている。命令バッファは好ましくは、全ての命令が一つのグループとしてレジスタから出て行く、つまり、実質的に同時に出て行くように動作する。命令バッファはいずれの大きさでありうるが、好ましくは、2 あるいは 3 レジスタよりは大きくないサイズである。

【0054】

一般に、デコード回路は命令を壊し、対応する命令の関数を実施する論理的マイクロオペレーションを生成する。例えば、論理的マイクロオペレーションは、算術論理演算、ローカルメモリ 550 へのロード及びストアオペレーション、レジスタソースオペランド、及び／又は即値データオペランドを特定しうる。デコード回路はまた、ターゲットレジスタアドレス、構造リソース、機能ユニット、及び／又はバスなど、命令がどのリソースを使用するかを示しうる。デコード回路はまた、リソースが要求される命令パイプラインステージを示す情報を与えることが出来る。命令デコード回路は好ましくは、命令バッファのレジスタ数に等しい数の命令を実質的に同時にデコードするように動作する。

20

【0055】

依存度チェック回路は、所定の命令のオペランドがパイプラインの他の命令のオペランドに依存しているかどうかを判断するために試験を行う、デジタル論理回路を含む。その場合、所定の命令はそのような他のオペランドが（例えば、他の命令が実行の完了を許容することにより）アップデートされるまで、実行されることができない。依存度チェック回路は好ましくは、デコーダー回路 112 から同時に送られる複数の命令の依存度を判断する。

30

【0056】

命令発行回路は浮動小数点実行ステージ 556、及び／または固定小数点実行ステージ 558 へ命令を発行するように動作することができる。

【0057】

レジスタ 554 は好ましくは、128 エントリのレジスタファイルなどの、相対的に大きな統一レジスタファイルとして実装される。これにより、レジスタが足りなくなる状態を回避するよう、レジスタの名前の変更を必要としない、深くパイプライン化された高周波数の実装品が可能になる。一般に、ハードウェアの名前変更には、処理システムのかなりの割合の領域と電力を消費する。その結果、ソフトウェアのループ展開、又は他のインターリーブ技術によりレイテンシがカバーされると、最新のオペレーションが実現される。

40

【0058】

好ましくは、S P U コア 510 A はスーパースカラアーキテクチャであり、これにより 1 つ以上の命令がクロックサイクル毎に発行される。S P U コア 510 A は好ましくは、命令バッファから送られる同時命令の数、例えば 2 ～ 3 命令（各クロックサイクル毎に 2 命令あるいは 3 命令が発行されることを意味する）に対応する程度まで、スーパースカラとして動作する。所望の処理能力に応じて、多数の、あるいは少数の浮動小数点実行ステージ 556 と、固定小数点実行ステージ 558 が採用される。好ましい実施形態では、浮

50

動小数点実行ステージ556は1秒あたり320億の浮動小数点演算速度で演算し(32 GFLOPS)、また、固定小数点実行ステージ558は演算速度が1秒あたり320億回(32 GOPS)となっている。

【0059】

MFC510Bは、好ましくは、バスインターフェースユニット(BIU:Bus Interface Unit)564、メモリ管理ユニット(MMU:Memory Management Unit)562、及びダイレクトメモリアクセスコントローラ(DMAC:Direct Memory Access Controller)560を備えている。DMAC560は例外として、MFC510Bは好ましくは、低電力化設計とするため、SPUコア510Aやバス512と比べて半分の周波数で(半分の速度で)実行する。MFC510Bはバス512からSPU508に入力されるデータや命令を処理するように動作することができ、DMACに対しアドレス変換を行い、また、データコヒーレンシーに対しスヌープオペレーションを提供する。BIU564はバス512とMMU562及びDMAC560間にインターフェースを提供する。従って、SPU508(SPUコア510A及びMFC510Bを含む)及びDMAC560は、バス512と物理的に、及び/又は論理的に結合されている。

10

【0060】

MMU562は、好ましくは、メモリアクセスのために、実アドレスに有効アドレスを変換するように動作することができる。例えば、MMU562は、有効アドレスの上位ビットを実アドレスビットに変換しうる。しかしながら下位のアドレスビットは好ましくは変換不能であり、また、実アドレスの形成及びメモリへのアクセスリクエストに使用する場合には、ともに論理的及び物理的なものと考えられる。1つ以上の実施形態では、MMU562は、64ビットのメモリ管理モデルに基づいて実装され、また、4K-、64K-、1M-、及び16M-バイトのページサイズを有する 2^{64} バイトの有効アドレススペースと、256MBのセグメントサイズを提供しうる。MMU562は好ましくは、DMAコマンドに対し、 2^{65} バイトまでの仮想メモリ、 2^{42} バイト(4テラバイト)までの物理メモリをサポートするように動作することが出来る。MMU562のハードウェアは、8-エントリでフルアソシエィティブのSLBと、256-エントリと、4ウェイセットアソシエィティブのTLBと、TLBに対してハードウェアTLBのミスハンドリングに使用される 4×4 リプレースメント管理テーブル(RMT:Replacement Management Table)と、を含む。

20

30

【0061】

DMAC560は、好ましくは、SPUコア510Aや、PU504、及び/又は他のSPUなどの、1つ以上の他のデバイスからのDMAコマンドを管理するように動作することができる。DMAコマンドには3つのカテゴリが存在し、それらは、プットコマンド、ゲットコマンド、及びストレージ制御コマンドである。プットコマンドは、ローカルメモリ550から共有メモリ514へデータを移動させるよう動作する。ゲットコマンドは、共有メモリ514からローカルメモリ550へデータを移動させるよう動作する。また、ストレージ制御コマンドには、SLIコマンドと同期化コマンドが含まれる。この同期化コマンドは、アトミックコマンド(atomic command)、信号送信コマンド、及び専用バリアコマンドを有しうる。DMAコマンドに応答して、MMU562は有効アドレスを実アドレスに変換し、実アドレスはBIU564へ送られる。

40

【0062】

SPUコア510Aは、好ましくは、DMAC560内のインターフェースと通信(DMAコマンド、ステータスなどを送る)するために、チャンネルインターフェース及びデータインターフェースを使用する。SPUコア510Aはチャンネルインターフェースを介して、DMAC560のDMAキューへDMAコマンドを送る。DMAコマンドがDMAキューに存在すると、そのコマンドはDMAC560内の発行及び完了論理により処理される。DMAコマンドに対する全てのバストランザクションが終了すると、完了信号がチャンネルインターフェースを越えて、SPUコア510Aへ送られる。

【0063】

50

プロセッサエレメントは、64ビットのプロセッシングユニット504（又は、パワーアーキテクチャプロセッサのファミリーに対応の、パワープロセッシングユニット（PPU））を備えており、パワーアーキテクチャの整数、浮動小数点、VMX及びMMUユニットを備えたデュアルスレッドコアとして実装される。プロセッサは32kBの命令及びデータキャッシュ、512kBのL2キャッシュ、及びオンチップバスインターフェースロジックを有す。プロセッサは拡張パイプラインを有する、新たに作られた実装品であり、SPUとマッチングするように、低FO4を実現することができる。コアは適度の長さのパイプラインを持つ、性能が向上したインオーダー設計であり、最新のパフォーマンスキャパビリティを提供する。PPUはリアルタイムオペレーションをサポートするために、キャッシュ及び変換テーブル用のリソース管理テーブルで拡張されている。メモリがマッピングされたI/O制御レジスタを通じて、PPUはまたSPUの代わりにDMAリクエストを開始し、SPUのメールボックスとの通信をサポートできる。PPUはまた、パワーアーキテクチャハイパーバイザー拡張も実装しており、スレッド管理サポートを通じて複数の同時並行オペレーティングシステムが、その上で同時に実行されることが

10

【0064】

図9はPU504の一般的な構造及び機能を例示している。PU504は2つの機能ユニットを有しており、それらはPUコア504Aとメモリフローコントローラ（MFC）504Bである。PUコア504Aは、プログラム実行、データ操作、マルチプロセッサマネージメント関数などを実施し、一方でMFC504Bはシステム100のPUコア504Aとメモリスペース間のデータ転送に関連する機能を実行する。

20

【0065】

PUコア504AはL1キャッシュ570、命令ユニット572、レジスタ574、1つ以上の浮動小数点実行ステージ576、及び1つ以上の固定小数点実行ステージ578を有することができる。L1キャッシュは、共有メモリ106、プロセッサ102、又はMFC504Bを介してメモリスペースの他の部分、から受信したデータに対するデータキャッシングの機能を提供する。PUコア504Aが好ましくはスーパーパイプラインとして実装されるので、命令ユニット572は好ましくは、フェッチ、デコード、依存度チェック、発行、などを含む、多くのステージを備えた命令パイプラインとして実装される。PUコア504はまた好ましくは、スーパースカラ構成であり、一方で1つ以上の命令がクロックサイクル毎に命令ユニット572から発行される。高度な処理（演算）能力を実現するために、浮動小数点実行ステージ576と固定小数点実行ステージ578は、パイプライン構成で複数のステージを有する。要求される処理能力に応じて、多数の、又は少数の浮動小数点実行ステージ576と固定小数点実行ステージ578が採用されうる。

30

【0066】

MFC504Bは、バスインターフェースユニット（BIU）580、L2キャッシュメモリ、キャッシュ不可能なユニット（NCU：Non-Cachable Unit）584、コアインターフェースユニット（CIU：Core Interface Unit）586、及びメモリ管理ユニット（MMU）588を備えている。殆どのMFC504Bは、低電力化設計とするために、PUコア504Aとバス108と比べて、半分の周波数（半分の速度）で実行する。

40

【0067】

BIU580はバス108とL2キャッシュ582とNCU584論理ブロック間にインターフェースを提供する。このために、BIU580はバス108上で、十分にコヒーレントなメモリオペレーションを実施するために、マスタデバイスとして、また同様にスレーブデバイスとして機能する。マスタデバイスとして、BIU580はL2キャッシュ582とNCU584のために機能するため、バス108へロード／ストアリクエストを供給する。BIU580はまた、バス108へ送信されうるコマンドの合計数を制限するコマンドに対し、フロー制御機構を実装しうる。バス108のデータオペレーションは、8ビット要するように設計され、そのために、BIU580は好ましくは128バイトキャッシュラインを有するように設計され、また、コヒーレンシーと同期化の粒度単位は1

50

28KBである。

【0068】

L2キャッシュメモリ582（及びサポートハードウェア論理回路）は、好ましくは、512KBのデータをキャッシュするように設計されている。例えば、L2キャッシュ582はキャッシュ可能なロード／ストア、データプレフェッチ、命令プレフェッチ、命令プレフェッチ、キャッシュオペレーション、及びバリアオペレーションを処理する。L2キャッシュ582は好ましくは8ウェイのセットアソシエティブシステムである。L2キャッシュ582は6つのキャストアウトキュー（6つのRCマシンなど）と一致する6つのリロードキューと、8つ（64バイト幅）のストアキューを備える。L2キャッシュ582はL1キャッシュ570において、一部の、あるいは全てのデータのコピーをバックアップするように動作する。この点は、処理ノードがホットスワップである場合に状態を回復するのに便利である。この構成により、L1キャッシュ570が少ないポート数でより速く動作することができ、また、より速くキャッシュツーキャッシュ転送ができる（リクエストがL2キャッシュ582でストップしうするため）。この構成はまた、キャッシュコヒーレンシー管理をL2キャッシュメモリ582へ送るための機構も提供する。

10

【0069】

NCU584は、CIU586、L2キャッシュメモリ582、及びBIU580と連動しており、通常は、PUコア504Aとメモリシステム間のキャッシュ不可能なオペレーションに対して、キューイング／バッファリング回路として機能する。NCU584は好ましくは、キャッシュ抑制ロード／ストア、バリアオペレーション、及びキャッシュコヒーレンシーオペレーションなどの、L2キャッシュ582により処理されないPUコア504Aとの全ての通信を処理する。NCU584は好ましくは、上述の低電力化目的を満たすように、半分の速度で実行されう。

20

【0070】

CIU586は、MFC504BとPUコア504Aの境界に配置され、実行ステージ576、578、命令ユニット572、及びMMUユニット588からのリクエストに対し、また、L2キャッシュ582及びNCU584へのリクエストに対し、ルーティング、アービトレーション、及びフロー制御ポイントして機能する。PUコア504A及びMMU588は好ましくはフルスピードで実行され、一方でL2キャッシュ582及びNCU584は2：1の速度比で動作することができる。従って、周波数の境界がCIU586に存在し、その機能の一つは、2つの周波数ドメイン間でリクエストの送信及びデータのリロードを行いながら、周波数の差を適切に処理することである。

30

【0071】

CIU586は3つの機能ブロックを有しており、それらは、ロードユニット、ストアユニット、及びリロードユニットである。更に、データプレフェッチ関数がCIU586により実施され、また好ましくは、ロードユニットの機能部である。CIU586は、好ましくは、

- (i) PUコア504AとMMU588からのロード及びストアリクエストを受ける、
- (ii) フルスピードのクロック周波数をハーフスピードに変換する（2：1のクロック周波数変換）、
- (iii) キャッシュ可能なリクエストをL2キャッシュ582へ送り、キャッシュ不可能なリクエストをNCU584へ送る、
- (iv) L2キャッシュ582に対するリクエストとNCU584に対するリクエストを公正に調停する、
- (v) ターゲットウィンドウでリクエストが受信されてオーバーフローが回避されるように、L2キャッシュ582とNCU584に対する転送のフロー制御を提供する、
- (vi) ロードリターンデータを受信し、そのデータを実行ステージ576、578、命令ユニット572、又はMMU588へ送る、
- (vii) スヌープリクエストを実行ステージ576、578、命令ユニット572、又

40

50

はMMU 588へ送る、

(viii) ロードリターンデータとスヌープトラフィックを、ハーフスピードからフルスピードへ変換する、
ように動作可能である。

【0072】

MMU 588は、好ましくはPUコア540Aに対して、第2レベルのアドレス変換機能などによりアドレス変換を行う。第1レベルの変換は好ましくは、MMU 588よりも小型で高速でありうる、別々の命令及びデータERAT (Effective to Real Address Translation) アレイにより、PUコア504Aにおいて提供されうる。

【0073】

10

好ましい実施形態では、PUコア504は、64ビットの実装品で、4-6GHz、10F04で動作する。レジスタは好ましくは64ビット長（1つ以上の特殊用途のレジスタは小型でありうるが）であり、また、有効アドレスは64ビット長である。命令ユニット570、レジスタ572、及び実行ステージ574と576は好ましくは、(RISC) 演算技術を実現するために、PowerPCステージ技術を用いて実装される。

【0074】

本コンピュータシステムのモジュール構造に関する更なる詳細は、米国特許第6,526,491号に解説されており、該特許は参照として本願に組込まれる。

【0075】

本発明の少なくとも1つの更なる態様によれば、上述の方法及び装置は、図面において例示しているような、適切なハードウェアを利用して実現されうる。そのようなハードウェアは標準デジタル回路などのいずれの従来技術、ソフトウェア、及び／またはファームウェアプログラムを実行するように動作できるいずれの従来のプロセッサ、プログラム可能なROM (PROM)、プログラム可能なアレイ論理デバイス (PAL: Programmable Array Logic) などの、1つ以上のプログラム可能なデジタルデバイスあるいはシステム、を用いて実装されうる。更に、図示している装置は、特定の機能ブロックに分割されて示されているが、そのようなブロックは別々の回路を用いて、及び／あるいは1つ以上の機能ユニットに組み合わせて実装されうる。更に、本発明の様々な態様は、輸送及び／又は配布のために、(フロッピーディスク、メモリチップ、などの) 適切な記憶媒体に格納されうる、ソフトウェア、及び／又はファームウェアプログラムを通じて実装されうる。

20

30

【0076】

本明細書において、具体的な実施形態を用いて本発明を記載したが、これらの実施形態は本発明の原理および用途の例を示すものに過ぎないことを理解されたい。このため、添付の請求の範囲に記載した本発明の趣旨および範囲から逸脱することなく、これら例示的な実施形態を種々に変更したり、上記以外の構成を考案し得ることが理解されよう。

【図面の簡単な説明】

【0077】

【図1】 本発明の1つ以上の実施形態を採用したマルチプロセッシングシステムの構造を例示したブロック図。

【図2】 図1のシステムにより採用されうる更なる特徴を例示したブロック図。

40

【図3】 本発明の1つ以上の態様によるプロセッシングシステムでの使用に適したインターフェースコントローラのブロック図。

【図4】 プロセッシング構成を実現するために、マルチプロセッサシステムの1つ以上の態様が採用される方法の一例を例示したブロック図。

【図5】 更なるプロセッシング構成を実現するために、マルチプロセッサシステムの1つ以上の態様が採用される方法の一例を例示したブロック図。

【図6】 また更なるプロセッシング構成を実現するために、マルチプロセッサシステムの1つ以上の態様が採用される方法の一例を例示したブロック図。

【図7】 また更なるプロセッシング構成を実現するために、マルチプロセッサシステムの1つ以上の態様が採用される方法の一例を例示したブロック図。

50

【図8】本発明の1つ以上の更なる態様により採用されうる、図1の一般的なサブプロセッシングユニット（SPU）の構造を例示した説明図。

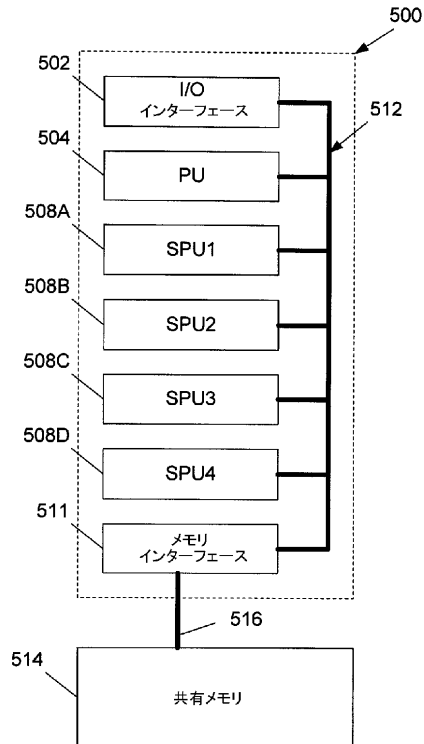
【図9】本発明の1つ以上の更なる態様により採用されうる、図1の一般的なプロセッシングユニット（PU）又はパワープロセッシングユニット（PPU）の構造を例示した説明図。

【符号の説明】

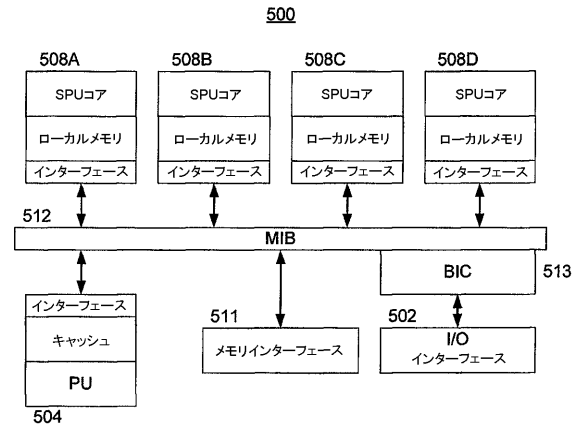
【0078】

100	システム	
102	プロセッサ	
106	共有メモリ	10
108	バス	
112	デコーダー回路	
500	プロセッサエレメント	
504	プロセッシングユニット	
504A	コア	
508	サブプロセッシングユニット	
510A	コア	
511	メモリインターフェース	
512	バス	
514	共有メモリ	20
540A	コア	
550	ローカルメモリ	
554	レジスタ	
570	命令ユニット	
572	命令ユニット	
582	キャッシュ	

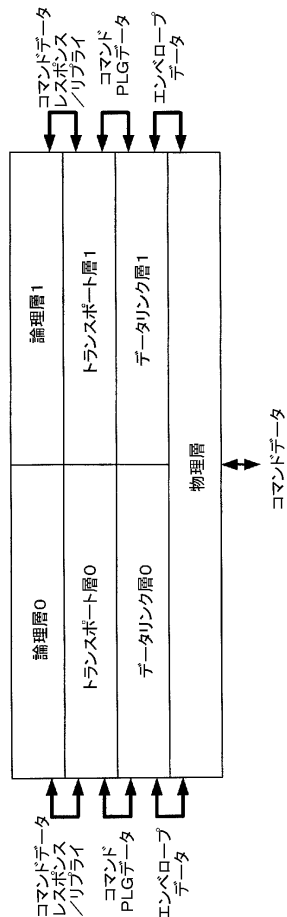
【図 1】



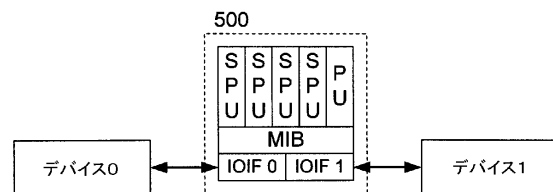
【図 2】



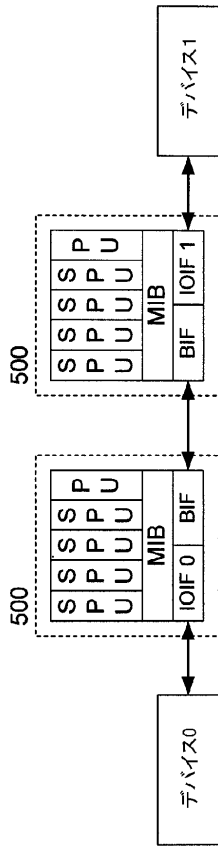
【図 3】



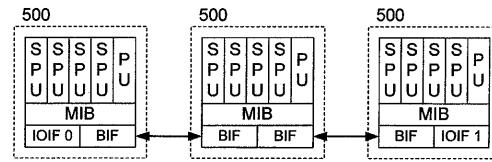
【図 4】



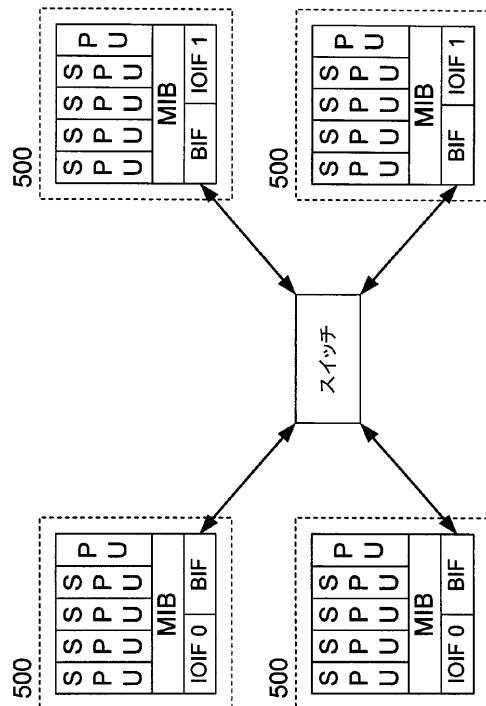
【図 5】



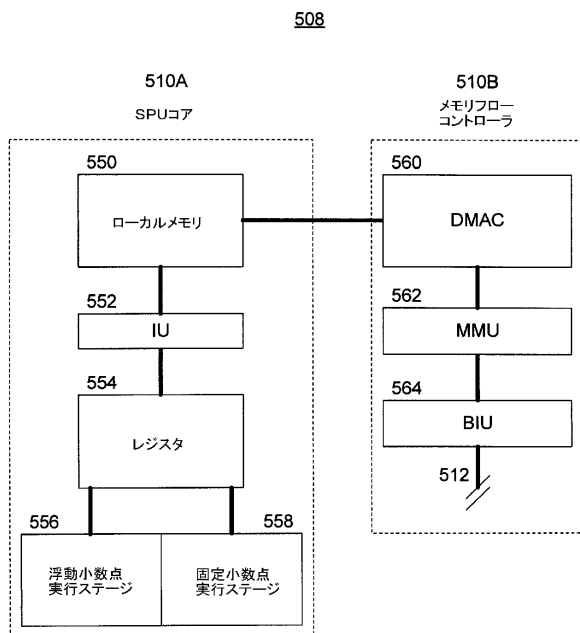
【図 6】



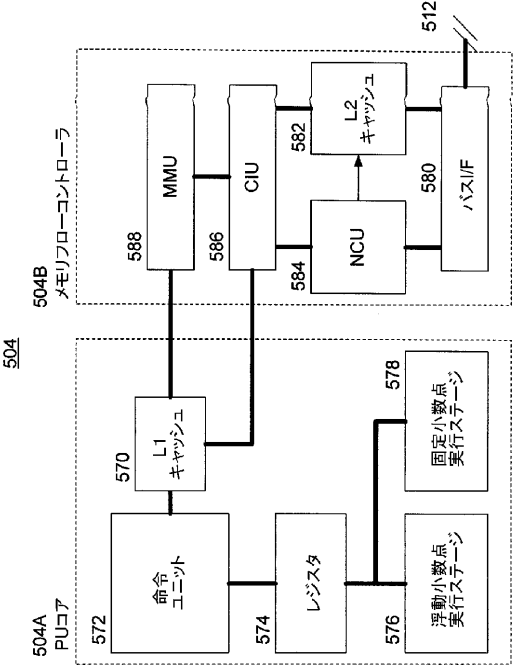
【図 7】



【図 8】



【図 9】



フロントページの続き

- (51)Int.Cl. F I
G 0 6 F 12/08 5 5 1 C
G 0 6 F 12/08 5 3 1 C
- (74)代理人 100108604
弁理士 村松 義人
- (74)代理人 100111615
弁理士 佐野 良太
- (72)発明者 山崎 剛
東京都港区南青山二丁目6番21号 株式会社ソニー・コンピュータエンタテインメント内
- (72)発明者 スコット ダグラス クラーク
アメリカ合衆国、ミネソタ州 55901-2328、ロチェスター、フォース アベニュー ノースウエスト 2922
- (72)発明者 チャールズ レイ ジョーンズ
アメリカ合衆国、テキサス州 78759、オースティン、カシア ドライブ 10703
- (72)発明者 ジェイムズ アラン カール
アメリカ合衆国、テキサス州 78731、オースティン、ハーロック ドライブ 5402
- 審査官 西出 隆二
- (56)参考文献 特表2003-508850 (JP, A)
特開平09-160892 (JP, A)
特開平10-143483 (JP, A)
- (58)調査した分野(Int.Cl., DB名)
G 0 6 F 13/14
G 0 6 F 12/08