

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3891994号
(P3891994)

(45) 発行日 平成19年3月14日(2007.3.14)

(24) 登録日 平成18年12月15日(2006.12.15)

(51) Int. Cl.

G06F 9/46 (2006.01)

F I

G06F 9/46 430

請求項の数 10 (全 16 頁)

(21) 出願番号 特願2004-126472 (P2004-126472)
 (22) 出願日 平成16年4月22日(2004.4.22)
 (65) 公開番号 特開2004-334863 (P2004-334863A)
 (43) 公開日 平成16年11月25日(2004.11.25)
 審査請求日 平成16年4月22日(2004.4.22)
 (31) 優先権主張番号 10/434685
 (32) 優先日 平成15年5月9日(2003.5.9)
 (33) 優先権主張国 米国(US)

(73) 特許権者 503003854
 ヒューレット・パカード デベロップメ
 ント カンパニー エル. ピー.
 アメリカ合衆国 テキサス州 77070
 ヒューストン 20555 ステイト
 ハイウェイ 249
 (74) 代理人 110000039
 特許業務法人アイ・ピー・エス
 (72) 発明者 クリス・グレア
 アメリカ合衆国テキサス州 アレン カレ
 ドニア1220

審査官 殿川 雅也

最終頁に続く

(54) 【発明の名称】 順番のある (in-order) キューをドレインする (drain) システムおよび方法

(57) 【特許請求の範囲】

【請求項1】

順番のあるキュー(111、121)をドレインする方法であって、

前記順番のあるキューに格納されたトランザクションのトランザクションフェーズに関
 して第1のキューループシーケンス(310、311)を実施することであって、該第1
 のキューループシーケンスは、前記トランザクションフェーズが終了しなかった前記順番
 のあるキューに格納された前記トランザクションに対する前記トランザクションフェーズ
 の終了をまねることと、

前記第1のキューループシーケンスの終了後に、前記順番のあるキューに格納された前
 記トランザクションの別のトランザクションフェーズに関して第2のキューループシーケ
 ンス(320、321、322、323)を実施することであって、該第2のキューループシーケ
 ンスは、前記別のトランザクションフェーズが終了しなかった前記順番のあるキュー
 に格納された前記トランザクションに対する前記別のトランザクションフェーズの終
 了をまねることと

を含み、

前記最初に示したトランザクションフェーズおよび前記別のトランザクションフェーズ
 それぞれは、

複数のフェーズ

を含み、

マルチプレクサに関連付けられる

10

20

順番のあるキューをドレインする (drain) 方法。

【請求項 2】

前記最初に示したトランザクションフェーズは、

スヌープフェーズ

を含み、

前記別のトランザクションフェーズは、

応答フェーズ

を含む

請求項 1 に記載の順番のあるキューをドレインする方法。

【請求項 3】

前記第 1 のキューループシーケンスと前記第 2 のキューループシーケンスとのうちの少なくとも 1 つは、

停滞状態 (3 1 1、3 2 1、3 2 2、3 2 3) を実施することであって、それにより前記まねることを特定のプロトコルに従わせるようにすること

を含む

請求項 1 に記載の順番のあるキューをドレインする方法。

【請求項 4】

前記 停滞状態 を実施することは、

複数の 停滞状態 (3 2 1、3 2 2、3 2 3) を実施すること

を含む

請求項 3 に記載の順番のあるキューをドレインする方法。

【請求項 5】

前記第 1 のキューループシーケンスの動作は、前記最初に示したトランザクションフェーズに関連する第 1 のトランザクションフェーズポインタ (4 3 2) が所定値に達するまで続き、

前記第 2 のキューループシーケンスの動作は、前記別のトランザクションフェーズに関連する第 2 のトランザクションフェーズポインタ (4 3 3) が前記所定値に達するまで続く

請求項 1 に記載の順番のあるキューをドレインする方法。

【請求項 6】

前記所定値は、前記最初に示したトランザクションフェーズおよび前記別のトランザクションフェーズとは異なるトランザクションフェーズに関連する第 3 のトランザクションフェーズポインタ (4 3 1) に対応し、

前記異なるトランザクションフェーズは、

複数のフェーズ

を含み、

前記第 1 のトランザクションフェーズポインタと前記第 3 のトランザクションフェーズポインタとが一致すると、順番のあるキューのすべてのトランザクションが、終了をまねられた前記最初に示したトランザクションフェーズを通して取り出され、

前記第 2 のトランザクションフェーズポインタと前記第 3 のトランザクションフェーズポインタとが一致すると、順番のあるキューのすべてのトランザクションが、終了をまねられた前記別のトランザクションフェーズを通して取り出される

請求項 5 に記載の順番のあるキューをドレインする方法。

【請求項 7】

重大なエラーを検出すること

をさらに含み、

前記第 1 のキューループシーケンスと前記第 2 のキューループシーケンスとを実施することは、前記重大なエラーを検出することに対応する

請求項 1 に記載の順番のあるキューをドレインする方法。

【請求項 8】

10

20

30

40

50

前記第 1 のキューループシーケンスと、前記第 2 のキューループシーケンスとを実施する間に、前記重大なエラーに関連するバスから特定のシステム資源を切り離すことをさらに含む請求項 7 に記載の順番のあるキューをドレインする方法。

【請求項 9】

順番のあるキュー（ 1 1 1、 1 2 1 ）をドレインするシステムであって、

第 1 のドレイン状態（ 3 1 0 ）と、

第 2 のドレイン状態（ 3 2 0 ）と

を有する状態機械（ 3 0 0 ） と、

第 1 のトランザクションフェーズポイントと、

第 2 のトランザクションフェーズポイントと

を具備し、

前記第 2 のドレイン状態は、前記第 1 のドレイン状態のトランザクションフェーズに続くトランザクションフェーズに関連し、

前記第 1 のトランザクションフェーズポイントおよび前記第 2 のトランザクションフェーズポイントそれぞれは、トランザクションをたどるために用いられ、

前記状態機械は、処理中のトランザクションに関連する第 1 のトランザクションフェーズポイントを、前記トランザクションに対するトランザクションフェーズの終了をまねる前記第 1 のドレイン状態と、この第 1 のドレイン状態とは別の状態とを交互に進めることによって、順番に進めるように動作可能であり、

前記状態機械は、前記処理中のトランザクションに関連する第 2 のトランザクションフェーズポイントを、前記トランザクションに対する別のトランザクションフェーズの終了をまねる前記第 2 のドレイン状態と、この第 2 のドレイン状態とは別の状態とを順次進めることによって、順番に進めるようにさらに動作可能である

順番のあるキューをドレインするシステム。

【請求項 10】

トランザクションをたどるために用いられる第 3 のトランザクションフェーズポイントをさらに具備し、

前記状態機械により前記第 1 のトランザクションフェーズポイントを順番に進めることは、該第 1 のトランザクションフェーズポイントが第 3 のトランザクションフェーズポイントに等しくなるまで続き、前記第 1 のトランザクションフェーズポイントと前記第 3 のトランザクションフェーズポイントとが等しくなると、前記第 1 のドレイン状態から前記第 2 のドレイン状態に移り、

前記状態機械により前記第 2 のトランザクションフェーズポイントを順番に進めることは、該第 2 のトランザクションフェーズポイントが前記第 3 のトランザクションフェーズポイントに等しくなるまで続く

請求項 9 に記載の順番のあるキューをドレインするシステム。

【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において、順番のあるキューをドレインするシステムおよび方法に関する。

【0002】

[関連出願の相互参照]

本出願は、「System and Method for Transaction Time-Out」と題する同時出願され同時係属中の本願と同一の譲受人に譲渡された米国特許出願第 10 / 435, 127 号と、「Systems and Methods Controlling Transaction Draining for Error Recovery」と題する同第 10 / 434, 920 号と、「System and Method For Error Capture and Logging in Computer Systems」と題する同第 10 / 434, 925 号と、に関連する。

【背景技術】

【0003】

10

20

30

40

50

今日、コンピュータシステムが、並列および／または分散処理を提供するように構成された複数のプロセッサをしばしば含み、極めて複雑であることは珍しいことではない。

たとえば、マルチプロセッサコンピュータシステムは、しばしば、複数の主処理ユニット(MPU)を含むのみではなく、メモリプロセッサ等の複数のサポートプロセッサまたはエージェントを含む場合もある。

これらのあらゆるプロセッサを、メモリ、入出力デバイス、ディスクデバイス等の他のシステム資源とともに、あらゆるバスによって可能になる通信によりコンピュータシステムを通して分散させることができる。

たとえば、コンピュータシステムは、主処理ユニット(MPU)、エージェントおよび／またはメモリ等の複数のシステム資源を有する、本明細書ではセルまたはセルカードと呼ぶ複数のサブモジュールと、それらの上に配置されたバスと、を備える場合がある。

サブモジュールのシステム資源は、他のシステム資源に対し要求を行いかつ／または他のシステム資源からの要求に対処することができる。

かかるシステム資源を、システムの同じサブモジュールおよび／または他のサブモジュールと連結することができる。

【0004】

複数のシステム資源からの要求に順序正しくかつ予測可能な方法で対処するために、システムは、あらゆるプロトコルおよびトランザクションキューを実施することができる。

たとえば、プロトコルは、例えば要求、スヌープおよび応答等の複数のトランザクションが実行されるべき順序および、恐らくは、かかるトランザクションの各々が終了するために提供されるべき複数のバス周期、を確定することができる。

同様に、トランザクションキューは、特定のシステム資源に関して「処理中の(in-process)」特定のトランザクションに関する情報を格納することができる。

たとえば、特定のトランザクションが関連するシステム資源によって適切な順序で確実に実施されるように、順番のある(in-order)キューを実施することができる。

したがって、順番のあるキューは、要求フェーズ(たとえば、データを要求すること)と、スヌープフェーズ(たとえば、要求されたデータがローカルに入手可能であるか否かを判断するためにローカルキャッシュに問合せること)と、応答フェーズ(たとえば、要求されたデータを含むデータパケットを提供すること)と、を含む場合があるような、それらの順番のあるフェーズを通して複数のトランザクション(たとえば、8つまでの未解決の(outstanding)トランザクション)と、をたどることができる。

【0005】

上述したシステム資源のうちの任意の1つに関連するような、どんな態様であれ、システムの動作中のエラーがシステムにより検出される場合、エラー信号を生成して、適切なシステム資源に通知することができる。

システム資源単位の動作に分離されておりかつ／または回復可能な動作に関連しているようなエラーは、重大でない可能性がある。

しかしながら、プロセッサバス全体の初期化を必要とし(本明細書では、バス初期化またはBINITエラーと呼ぶ)、したがってシステムバス上のシステム資源の初期化を必要とするようなエラーは実際に、重大である可能性がある。

【発明の開示】

【発明が解決しようとする課題】

【0006】

マルチプロセッサシステムにおけるバス初期化エラーまたは同様の重大なエラーにより、要求および応答を発行しおよび／またはそれらに応答する種々のシステム資源の相互依存性のために、障害が広がる可能性があり、システム全体の障害になる場合さえある。

単一プロセッサシステムは、すべての保留中のトランザクションをページし新たな命令をフェッチすることにより、バス初期化エラーから回復することができる可能性がある(たとえば、「ソフトブート(soft booting)」)、マルチプロセッサシステムのバス初期化では、「ハード」リセットが必要なシステム「ロックアップ(lock-up)」がもたら

10

20

30

40

50

される可能性があり、またはエラー源を分離するために有用なオペレーティングシステム状態ダンプを実行することが妨げられる可能性がある。

たとえば、バス初期化エラーにより、特定のシステム資源が、システム資源がその上でトランザクションを実行しているバスが初期化された時に、予想されるトランザクション応答を待って「ハングする(hang)」可能性がある。

初期化されたバス上のシステム資源は、それらの関連する順番のあるキューをたどるのをやめると、初期化されたバスにはないシステム資源に対して、待ちの応答を提供するのをやめることになる。

したがって、バス初期化エラーまたは同様のエラーにより、システム全体がデッドロックするカスケード故障がもたらされる可能性がある。

10

【0007】

バス初期化エラーがハードリセットを必要とするシステムロックアップをもたらす上述した状況は、高可用性(HA)システムでは望ましくない、ということを理解しなければならない。

さらに、かかる結果は、システム「クラッシュダウン(crash-down)」またはオペレーティングシステム(OS)がエラーの原因を特定するのを容易にするための、ディスクまたは他の媒体に対するすべてのメモリのダンプを行わず、代わりにシステム初期化を必要とし、それによりエラー時のシステムの状態を分析することができなくなる。

【課題を解決するための手段】

【0008】

20

順番のあるキューをドレインする方法であって、本方法は、順番のあるキューに格納されたトランザクションのトランザクションフェーズに関して第1のキューループシーケンスを実施することであって、第1のキューループシーケンスが、トランザクションフェーズが終了しなかった順番のあるキューに格納されたトランザクションに対するトランザクションフェーズの終了をスプーフ(spooof)する、実施すること、および第1のキューループシーケンスの終了後に、順番のあるキューに格納されたトランザクションの別のトランザクションフェーズに関して第2のキューループシーケンスを実施することであって、第2のキューループシーケンスが、別のトランザクションフェーズが終了しなかった順番のあるキューに格納されたトランザクションに対する別のトランザクションフェーズの終了をスプーフする、実施することを含む。

30

【0009】

順番のあるキューをドレインするシステムであって、本システムは、第1のドレイン状態と第2のドレイン状態とを有する状態機械を備え、この状態機械は、処理中のトランザクションに関連する第1のトランザクションフェーズポインタを、トランザクションに対するトランザクションフェーズの終了をスプーフする第1のドレイン状態を用いて、順番に進めるように動作可能であり、状態機械は、処理中のトランザクションに関連する第2のトランザクションフェーズポインタを、トランザクションに対する別のトランザクションフェーズの終了をスプーフする第2のドレイン状態を用いて、順番に進めるようにさらに動作可能である。

【0010】

40

順番のあるキューをドレインするためのコンピュータプログラムロジックが記録されたコンピュータ読取可能媒体を有するコンピュータプログラム製品であって、本コンピュータプログラム製品は、順番のあるキューに格納されたトランザクションのトランザクションフェーズに関して第1のキューループシーケンスを実施するコードであって、第1のキューループシーケンスが、第1のトランザクションフェーズポインタが所定値に達するまで、トランザクションフェーズが終了しなかった順番のあるキューに格納されたトランザクションに対するトランザクションフェーズの終了をスプーフするものである、コードと、順番のあるキューに格納されたトランザクションの別のトランザクションフェーズに関して第2のキューループシーケンスを実施するコードであって、第2のキューループシーケンスが、第2のトランザクションフェーズポインタが所定値に達するまで、別のトラン

50

ザクションフェーズが終了しなかった順番のあるキューに格納されたトランザクションに対する別のトランザクションフェーズの終了をスプーフするものである、コードと、を含む。

【発明を実施するための最良の形態】

【0011】

本発明の実施形態は、重大なバス初期化エラー等の特定のシステムエラーを認識し、システムエラーによって直接影響を受けるシステム資源に関連する1つまたは複数のキューを空にするように動作する、システムおよび方法を提供する。

一実施形態によれば、バス初期化エラーを検出し、初期化されるバス上のプロセッサを、初期化の前にバスから離れた状態にしている間に、それらのプロセッサに関連する順番のあるキューのドレインを行う。

10

本発明の実施形態によるキューのドレインは、キューに入れられたトランザクションにしたがって、初期化されるバスに関連しないものを含むあらゆるシステム資源に対し、応答、たとえばヘッダおよび/または任意のデータ要素を含むデータパケットを提供し、それにより、あらゆるシステム資源に対しトランザクションの終了をスプーフする。

【0012】

本発明の実施形態によるキューのドレインは、順番のあるキューに示されるトランザクションフェーズがシステムの動作プロトコルと一貫性のある順序でキューからドレインされることを確実にする、階層的な方法で行う。

たとえば、本発明の実施形態により、初期のトランザクションフェーズに関連するすべてのトランザクションを、後のトランザクションフェーズに関連するいかなるトランザクションよりも前にキューからドレインする。

20

一実施形態による動作時、キューに入れられたスヌープトランザクションフェーズからのパケット(packets off)の生成に続いてキューに入れられた応答トランザクションフェーズからパケットを生成するように、順番のあるキューをドレインする。

【0013】

本発明の実施形態は、バス上の種々のトランザクションをたどるために通常動作時に用いられる、順番のあるキューのトランザクションポイントについて、キューに入れられたトランザクションに歩を進めシステムプロトコルに準拠するパケットを生成するための、状態機械を実施する。

30

状態機械は、たとえば、まずまだスヌープフェーズを通過していないスヌープフェーズの各トランザクションに歩を進め、適切なシステム資源にパケットを発行してもよい。

その後、状態機械は、応答フェーズを通過していない各トランザクションに歩を進め、適切なパケットを発行してもよい。

本発明の実施形態によるかかる状態機械の動作により、システム内部のカウンタが同期するように保持する。

さらに、かかる状態機械の動作を利用して、キューが詰る(wedge)ことがないようにしてもよく、それによりシステム資源がタイムアウトするかまたは他の方法で処理を継続しなくなることが防止される。

【0014】

40

本発明の実施形態によってキューをドレインすることは、システムの正常な(graceful)終了および/または回復を提供することに関し、特に重大なシステムエラーの原因の診断を容易にするために、特に有用である、ということを理解しなければならない。

たとえば、本発明の実施形態により、オペレーティングシステム(OS)が、バス初期化エラーの原因の分析に対して有用な、メモリ状態をディスク等にダンプすることが容易になる。

【0015】

図1Aを参照すると、本発明の実施形態が実施されるコンピュータシステムの一部が、サブモジュールまたはセルコントローラ100として示されている。

サブモジュール100が一部を形成するホストシステムは、複数のIntel ITA

50

NIUMプロセッサを実施するHewlett Packard rxまたはsxシリーズサーバシステム等のマルチプロセッサシステムを備えてもよい。

【0016】

サブモジュール100の図示する実施形態は、主処理ユニット(MPU)、メモリプロセッサ、メモリおよび/または同様のもの(すべて図示せず)等のあらゆるシステム資源間の通信を可能にするために利用してもよいような、ここではバス101および102として示す複数のバスを含む。

サブモジュール100はさらに、関連する1つまたは複数のプロセッサ(たとえばメモリプロセッサ)と関連するバス(たとえば、バス101および102のうちの対応するもの)との間のデータ通信を調停してもよいような、複数のプロセッサインタフェース、こ

10

こではプロセッサインタフェース110および120を含む。
かかる通信を、サブモジュール100上に配置されるかホストシステムの別の場所(たとえば、サブモジュール100に対応する1つまたは複数の他のサブモジュール)に配置されるかに係らず、上述した関連するプロセッサと、ホストシステムのMPU、メモリプロセッサ、メモリ等の他のあらゆるシステム資源と、の間の通信を可能にするために利用してもよい。

【0017】

図示する実施形態のプロセッサインタフェース110および120は、プロセッサ入力ブロック(PIN)、ここでは各々プロセッサインタフェースブロック113および123と、プロセッサ出力ブロック(POUT)、ここでは各々プロセッサ出力ブロック114および124と、プロセッサインタフェースデータバス(PIDP)、ここではプロセッサインタフェースデータバス115および125と、を備える。

20

プロセッサインタフェース110および120の図示する実施形態は、さらに、各々プロセッサインタフェース順番のあるキュー(PIQ)112および122を備え、あらゆるトランザクションたどることにより、システムプロトコルを実施しおよび/または他の方法で上述した通信の制御されたおよび予測可能な調停を提供する。

図示する実施形態のプロセッサインタフェース順番のあるキュー112および122は、各々順番のあるキュー(IOQ)111および121を含む。

図示する実施形態の順番のあるキュー111および121は、オーバーライディングプロトコルによる順序でトランザクションの終了を確実にするためにトランザクションフェーズをたどる。

30

【0018】

図1B-1および図1B-2は、図1Aのサブモジュール100のプロセッサインタフェースの実施形態に関する詳細をさらに示す。

図1B-1および図1B-2の実施形態において分かるように、プロセッサインタフェース順番のあるキュー112は、順番のあるキュー111に入れられるトランザクションのフェーズ(要求フェーズ、スヌープフェーズおよび応答フェーズ)を制御するように動作可能な制御信号要求ポインタ(req_ptr)、スヌープポインタ(snp_ptr)および応答ポインタ(rsp_ptr)を提供する順番のあるキュー制御モジュール131を含む。

【0019】

40

図2に注目すると、一実施形態による、順番のあるキュー111のエントリの一部、たとえば、8エントリ順番のあるキューのスライスが示されている。

特に、図示する実施形態の順番のあるキュー111は、関連するプロセッサにより終了のあらゆるフェーズにおけるトランザクションに応答してもよいような、エントリ201および202を含む。

通常動作時、順番のあるキュー111は、スヌープフェーズおよび応答フェーズからシステムパケットを生成するためにプロセッサ入力ブロック113およびプロセッサ出力ブロック114とインタフェースし、これにより同時にシステムに複数のパケットを生成する能力を有する。

【0020】

50

図 2 の順番のあるキュー 1 1 1 は、上述した制御信号の制御下で、トランザクションを適切なフェーズで伝播させ最終的に順番のあるキュー 1 1 1 から出すように動作可能なマルチプレクサ 2 1 1 および 2 1 2 を含む。

たとえば、マルチプレクサ 2 1 1 を、初期のトランザクションフェーズ、ここでは「スヌープフェーズ」に関連付けてもよく、マルチプレクサ 2 1 2 を、後のトランザクションフェーズ、ここでは「応答フェーズ」に関連付けてもよい。

マルチプレクサ 2 1 1 および 2 1 2 を、キューに入れられたトランザクションに付随するパケットを生成するように独立して、好ましくは同時に動作させてもよいが、特定のトランザクションは、その応答フェーズにおける処理の前にスヌープフェーズを終了しなければならない、これにより順番のあるキュープロトコルを確定する。

10

【 0 0 2 1 】

通常動作時、上述したポインタは、順番のあるキュー 1 1 1 のトランザクションがある種々のフェーズをたどる。

これらのポインタは、適切なパケットが生成されトランザクションのフェーズに関連付けられるにしたがってキュー内の至るところを移動する (move up and down)。

しかしながら、バス初期化 (B I N I T) エラー等の重大なエラーが検出された場合、システム資源のうちの 1 つまたは複数がりセット状態になる可能性がある。

たとえば、バス 1 0 1 に関してバス初期化エラーが発行されると、プロセッサインタフェース 1 1 0 に関連するプロセッサがりセット状態になる可能性があり、それにより、いかなる現データもダンプし動作をリスタートするためにコードをフェッチする。

20

したがって、順番のあるキュー 1 1 1 に入れられたいかなるトランザクションも、関連するプロセッサの動作に関して無意味になる。

従来のロジックにしたがって、順番のあるキュー 1 1 1 もまた初期化される。

しかしながら、本発明の実施形態は、重大なエラー後に順番のあるキューをドレインし、それによりパケットをスプーフし、したがって他のシステム資源がロックアップするかまたは他の方法で重大なエラー後の処理を継続できなくなることを防止するように動作する。

【 0 0 2 2 】

図 3 に注目すると、本発明の実施形態によって順番のあるキューをドレインする、順番のあるキュー制御モジュール 1 3 1 内で実施されてもよいような状態機械 3 0 0 が示されている。

30

当技術分野において既知であるように、概して、状態機械は、所与の時点で何かのステータスを格納し、入力に対して操作することにより、ステータスを変更するようにおよび/または任意の所与の変更に対してアクションまたは出力を発生させることができる任意のデバイスである。

【 0 0 2 3 】

通常システム動作中、状態機械 3 0 0 はアイドルドレイン状態 3 0 1 で保持される。

しかしながら、バス初期化エラーが検出されると、状態機械 3 0 0 は、アイドルドレイン状態 3 0 1 から、本発明の教示にしたがってキューをドレインするように動作可能な第 1 の状態 (ここでは、スヌードレイン状態 3 1 0) に移る。

40

本発明の実施形態による動作時、バス初期化エラーが検出されたバスのプロセッサおよび/または他のシステム資源を、順番のあるキューのドレイン中に初期化されないようにする。

たとえば、バス初期化エラーが検出されたバス上のシステム資源を、アイドルドレイン状態 3 0 1 を除く状態機械 3 0 0 のすべての状態の間に新たな要求を行わないようにしてもよい。

特定の動作の終了中に特定のバスからシステム資源を切り離すことに関するさらなる詳細は、「Systems and Methods Controlling Transaction Draining for Error Recovery」と題する先に参照した特許出願に記載されている。

【 0 0 2 4 】

50

スヌープドレイン状態 3 1 0 は、システムトランザクションプロトコルの初期トランザクションフェーズ、この例では「スヌープフェーズ」に関連する。

したがって、状態機械 3 0 0 は、初期のトランザクションフェーズ「スヌープ」で保留中の次の待機トランザクションが選択されるようにし、恐らくは任意のデータ要素を含むが適切にフォーマットされた 1 つまたは複数のパケットが、スヌープドレイン状態 3 1 0 において生成されるようにする。

生成されたスヌープフェーズパケットは、通常動作と同様にシステム内を伝播し、それによりそのトランザクションに関してスヌープフェーズの終了をスプーフする。

【 0 0 2 5 】

本発明の実施形態は、システムプロトコルにしたがって順番のあるキューをドレインするように動作する。

たとえば、システムプロトコルは、キュードレインから切り離されたダウンストリームロジック、たとえばプロセッサ入力ブロック 1 1 3 が通常通りに動作可能になるなどのために、スヌープパケットがせいぜい 1 周期おきの頻度で生成される、ということを確認してもよい。

したがって、図示する実施形態の状態機械 3 0 0 は、システムプロトコルにしたがってキュー内ドレインを容易にするように、スヌープドレイン状態 3 1 0 後に入るスヌープドレインストール状態 3 1 1 を含む。

【 0 0 2 6 】

図示する実施形態によれば、状態機械 3 0 0 は、初期トランザクションフェーズが十分にドレインされるまで、スヌープドレイン状態 3 1 0 とスヌープドレインストール状態 3 1 1 との間を往復する。

たとえば、スヌープドレイン状態 3 1 0 とスヌープドレインストール状態 3 1 1 とを交互に行ってもよく、それにより、スヌープポインタが要求ポインタに等しくなるまで、1 周期おきにスヌープフェーズパケットを生成する。

スヌープトランザクションフェーズポインタが要求ポインタ（初期トランザクションフェーズポインタ）に等しくなると、順番のあるキューのすべてのトランザクションがスプーフされたスヌープフェーズを通して取出されたことになる。

したがって、スヌープポインタが要求ポインタに等しくなると、状態機械 3 0 0 の図示する実施形態は、システムトランザクションプロトコルの後のトランザクションフェーズ、この例では「応答フェーズ」に関連するトランザクションフェーズのキューのドレインを提供するように動作可能な状態（ここでは、応答ドレイン状態 3 2 0 ）に移る。

【 0 0 2 7 】

応答ドレイン状態 3 2 0 は、スヌープドレイン状態 3 1 0 ののみに続くシステムトランザクションプロトコルのトランザクションフェーズに関連する。

したがって、本発明の実施形態により順番のあるキューをドレインする動作は、特定のトランザクションフェーズを、通常動作中に処理される可能性のある順序でスプーフするように動作しなくてもよいが、本発明の実施形態は、各トランザクションに関連するフェーズがそれらの適切な順序でスプーフされることを確実にする。

図示する実施形態では、すべてのスヌープフェーズトランザクションがいかなる応答フェーズトランザクションをスプーフする前にもスプーフされるため、各トランザクションに関連するフェーズが所望の順序でスプーフされることを理解しなければならない。

【 0 0 2 8 】

図示する実施形態による動作時、状態機械 3 0 0 は、「応答」フェーズを受取る次の待機トランザクションが選択されるようにし、恐らくは任意のデータ要素を含むが適切にフォーマットされた 1 つまたは複数のパケットが、応答ドレイン状態 3 2 0 において生成されるようにする。

上述したスヌープフェーズパケットと同様に、生成された応答フェーズパケットは、通常動作の場合と同様にシステム内を伝播し、それによってそのトランザクションに関する応答フェーズの終了をスプーフする。

10

20

30

40

50

【 0 0 2 9 】

上述したように、本発明の実施形態は、システムプロトコルにしたがって順番のあるキューをドレインするように動作する。

したがって、図示する実施形態の状態機械 3 0 0 は、システムプロトコルにしたがって順番のあるキューのドレインを容易にするように、応答ドレイン状態 3 2 0 後に入る応答ドレインストール状態 3 2 1 ~ 3 2 3 を含む。

【 0 0 3 0 】

図示する実施形態によれば、状態機械 3 0 0 は、後のトランザクションフェーズが十分にドレインされるまで、応答ドレイン状態 3 2 0 と、応答ドレインストール状態 3 2 1 と、応答ドレインストール状態 3 2 2 と、応答ドレインストール状態 3 2 3 と、に順次入る

10

。たとえば、図示する実施形態では、応答ポインタが要求ポインタに等しくなるまで、4クロック周期おきに応答パケットがスプーフされるように、応答ドレインストール状態が提供される。

応答トランザクションフェーズポインタが要求ポインタ（初期トランザクションフェーズポインタ）に等しくなると、順番のあるキューのすべてのトランザクションがスプーフされた応答フェーズを通して取出されたことになる、ということを理解しなければならない。

応答ポインタが要求ポインタに等しくなると、本発明の状態機械 3 0 0 は、システムトランザクションプロトコルのさらに後のトランザクションフェーズに関連するトランザク

20

ションフェーズのキューのドレインを提供するように動作可能な状態等、後続する所望の状態に移ってもよい。

しかしながら、図示する実施形態では、ドレインされるトランザクションフェーズは2つしか存在せず、したがって図示する実施形態の状態機械 3 0 0 は、初期エラーが検出されたバス上のプロセッサがリセットから解放され初期化が可能になるため等、アイドルドレイン状態 3 0 1 に戻る。

【 0 0 3 1 】

図 4 に注目すると、状態機械 3 0 0 にしたがう動作を示すタイミング図が示されている。

4 0 1 において、システムにバス初期化エラーが存在する。

30

したがって、4 0 1 の直後に、4 0 2 においてプロセッサインタフェースキューグループのバス初期化状態信号グループが、アイドル状態からドレイン状態に移る。

4 1 1 ~ 4 1 5 は、第 1 のトランザクションフェーズ、たとえば上述したスヌープフェーズに関するドレイン状態のシーケンスを表す。

4 3 1 および 4 3 2 において分かるように、4 1 1 ~ 4 1 5 を経過した後、第 2 のトランザクションフェーズポインタ、ここではスヌープポインタは、第 1 のトランザクションフェーズポインタ、ここでは要求ポインタに等しい。

したがって、4 0 3 において、バス初期化状態信号グループは、第 2 のドレイン状態に移る。

4 2 1 ~ 4 2 6 は、第 2 のトランザクションフェーズ、たとえば上述した応答フェーズに関するドレイン状態のシーケンスを表す。

40

4 3 1 および 4 3 3 において分かるように、4 2 1 ~ 4 2 6 を経過した後、第 3 のトランザクションフェーズポインタ、ここでは応答ポインタは、第 1 のトランザクションフェーズポインタ、ここでは要求ポインタに等しい。

【 0 0 3 2 】

図示する実施形態に示すような上述したドレイン状態およびドレインストール状態の実施態様は、本発明によって採用してもよい状態を例示するものである、ということを理解しなければならない。

特定の状態を採用すべきか否かと採用する特定の状態の数とは、システムおよび/またはそれと使用するプロトコルの属性によって変化してもよい。

50

たとえば、図示する実施形態は、２つのトランザクションフェーズに関連する２つのドレイン状態を採用し、それらに対し、トランザクションフェーズを終了することができないことにより、キューの詰りおよび／またはシステム資源ロックアップがもたらされる可能性がある。

本発明の実施形態は、必要に応じて、いかなる数のかかるドレイン状態を実施してもよい。

【 0 0 3 3 】

同様に、図示する実施形態は、特定のドレイン状態に関して３つの応答ドレインストール状態を採用することにより、バス上の、トランザクションが応答フェーズにありデータがプロセッサ無しに、たとえばキャッシュラインで駆動している場合に適応する。

10

応答フェーズパケットスプーフに関して実施される４つの周期により、かかるデータが、重大なエラーが検出された場合のトランザクションの実際のステータスに係りなくシステムプロトコルにしたがって、図示する実施形態のプロセッサ入力ブロック 1 1 3 および／またはプロセッサインタフェースデータパスブロック 1 1 5 等のシステムを適切に伝播することが確実になる。

本発明の実施形態は、必要に応じて、いかなる数のかかるドレインストール状態（０を含む）を特定のドレイン状態に関していかなる組合せで実施してもよい。

【 0 0 3 4 】

本発明の実施形態によるキューのドレインは、種々のパケットに適応するように適合される、ということを理解しなければならない。

20

たとえば通常システム動作は、特定のシステム資源に向けられたパケットのみを採用するのではなく、マルチキャストパケット等のマルチポイントパケットを採用してもよい。

システムのセルのうちの１つのプロセッサバスは、たとえば、保護領域にあるすべてのメモリプロセッサ（すなわち、特定のオペレーティングシステムの制御下で動作可能なすべてのセル）等複数のシステム資源に送出されるキャッシュ同期パケットを発行してもよい。

保護領域の各プロセッサバスは、キャッシュ同期を行うためにキャッシュ同期パケットを調べる必要がある。

したがって、通常動作時、プロセッサインタフェースは、関連するプロセッサからかかるマルチキャストパケット要求を受取り、それを、送出する各要求パケットに対する肯定応答を期待して、その保護領域にあるすべてのセルに送出する。

30

しかしながら、上述したように、セルは、複数のプロセッサおよび／または複数のバスを含む可能性がある。

したがって、単一セル内の複数のシステム資源が、最終的にマルチキャストパケットキャッシュ同期に影響するが、要求しているシステム資源は、各セルから単一の肯定応答のみを期待してもよい。

【 0 0 3 5 】

図示する実施形態のプロセッサ連絡（processor junction（PJ））ブロック 1 3 0 は、通常動作においてかかるマルチキャスト肯定応答の調整を提供する。

特に、プロセッサ連絡ブロック 1 3 0 は、特定のパケットがマルチキャストパケットであることを認識し、各プロセッサ入力ブロック 1 1 3 および 1 2 3 と協議することにより最新のもののみがトランザクションを終了して期待された肯定応答を実際に送出することができるようにする。

40

図示する実施形態によれば、プロセッサ連絡ブロック 1 3 0 には、これらの状況を扱うためにカウンタがある。

本発明の実施形態による順番のあるキューのドレインの動作は、キューに入れられたトランザクションフェーズの各々を順序正しくスプーフして、上述したプロセッサ連絡ブロック 1 3 0 カウンタを同期させた状態にし、それにより、あらゆるタイプの保留中のトランザクションに関してもシステムを安定した状態に維持する。

【 0 0 3 6 】

50

本発明の図示する実施形態のドレイン状態は、システムプロトコルに準拠するパケットを生成するように動作するが、バス初期化エラーが検出されたバス上のシステム資源がリセット状態に保持されるため、生成されたパケットは、無効なデータ要素を含む可能性が高い、ということを理解しなければならない。

たとえば、特定のデータパケットは、すべて0等のヌルデータを含んでもよい。

したがって、本発明の実施形態による動作により、重大なエラーを有していたバスに関連するシステム資源によってトランザクションが終了されるのを待って、キューが詰るかまたはシステム資源がロックアップする状況は回避されるが、スプーフされたトランザクションフェーズの結果として、システムを通してシステム資源に1つまたは複数のエラーがもたらされる可能性がある。

10

しかしながら、これらの動作エラーは、それほど小さくなく、たとえば重大でないと予期され、したがってシステム処理は順序正しいクラッシュ、コアダンプおよび/またはソフト初期化を持続してもよい。

したがって、本発明の実施形態は、好ましくは、「System and Method for Error Capture and Logging in Computer Systems」と題する先に参照した特許出願に示されているようなエラー収集および/またはロギングを実施する。

【0037】

本発明によるキューのドレインの動作を、追加の重大障害回復技法と組合せて実施してもよい、ということを理解しなければならない。

たとえば、本発明の実施形態による順番のあるキューのドレインを、「Systems and Methods Controlling Transaction Draining for Error Recovery」と題する先に参照した特許出願に示すような順番通りでない(out of order)バッファドレインも含むシステム障害回復プロセスで実施してもよい。

20

一実施形態によれば、順番のあるキューのドレインを実施し終了した後、アウトオブオーダーバッファドレインを実施して終了し、その後通常のシステム動作に戻ってもよい。

したがって、本発明の実施形態による順番のあるキュードレインは、関連する動作を調整するために、順番のあるキュードレイン中にプロセッサ出力ブロック114等のシステム資源または他の機能ブロックに対し、「piqドレイン」信号等の信号をアサートしてもよい。

【0038】

30

実施形態を、保護領域のシステム資源に関して上述したが、本発明の実施形態の利点は、システム資源が保護領域間で排他的に区分されるマルチプロセッサ環境に限定されない、ということを理解しなければならない。

たとえば、本発明の実施形態による順番のあるキュードレインは、システム資源が保護領域の区分に互って利用される世界共有メモリ(globally shared memory(GSM))を実施するシステムに関して特に有用で有り得る。

本発明の実施形態による順番のあるキュードレインを実施することにより、1つの保護領域のクラッシュが別の保護領域を、それらの保護領域がメモリを共有する場合であっても、ダウンさせることがない。

1つの保護領域が正常にクラッシュし残りの保護領域が処理を継続することができるようにして、キューの詰りおよび/またはシステム資源ロックアップが回避されるためである。

40

【0039】

本発明の要素をソフトウェアで実施する場合、それらは本質的に、本明細書で説明したようなタスクを実行するコードセグメントである。

プログラムまたはコードセグメントをコンピュータ読取可能媒体に格納し、または搬送波に具体化されるコンピュータデータ信号かまたは搬送波によって変調される信号により、伝送媒体を介して送信することができる。

コンピュータ読取可能媒体は、情報を格納しまたは転送することができるいかなる媒体を含んでもよい。

50

コンピュータ読取可能媒体の例には、電子回路、半導体メモリデバイス、ROM、フラッシュメモリ、消去可能ROM（EROM）、フロッピー（登録商標）ディスク、コンパクトディスクCD-ROM、光ディスク、ハードディスク、光ファイバ媒体、無線周波数（RF）リンク等がある。

コンピュータデータ信号は、電子ネットワークチャネル、光ファイバ、空気、電磁気、RFリンク等の伝送媒体を介して伝播することができるいかなる信号を含んでもよい。

コードセグメントを、インターネット、イントラネット等のコンピュータネットワークを介してダウンロードしてもよい。

【図面の簡単な説明】

【0040】

10

【図1A】本発明の実施形態が実施されるマルチプロセッサシステムの一部を示す図である。

【図1B-1】本発明の実施形態が実施されるマルチプロセッサシステムの一部を示す図である。

【図1B-2】本発明の実施形態が実施されるマルチプロセッサシステムの一部を示す図である。

【図2】本発明の実施形態によって動作可能な順番のあるキューに関して詳細を示す図である。

【図3】本発明の実施形態により順番のあるキュードレインを実施する状態機械を示す図である。

20

【図4】図3の状態機械の動作に対応するタイミング図である。

【符号の説明】

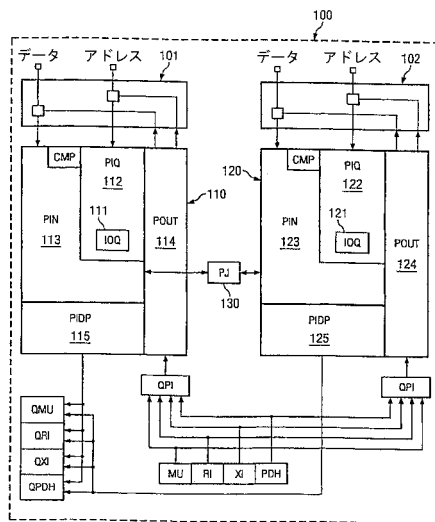
【0041】

100・・・サブモジュール、
 101, 102・・・バス、
 110, 120・・・プロセッサインタフェース、
 113, 123・・・プロセッサインタフェースブロック、
 114, 124・・・プロセッサ出力ブロック、
 115, 125・・・プロセッサインタフェースデータパス、
 112, 122・・・キュー（PIQ）、
 111, 121・・・キュー（IOQ）、
 131・・・キュー制御モジュール、
 201, 202・・・エントリ、
 211, 212・・・マルチプレクサ、
 300・・・状態機械、
 301・・・アイドルドレイン状態、
 310・・・スヌープドレイン状態、
 311・・・スヌープドレインストール状態、
 320・・・応答ドレイン状態、
 321～321・・・応答ドレインストール状態、

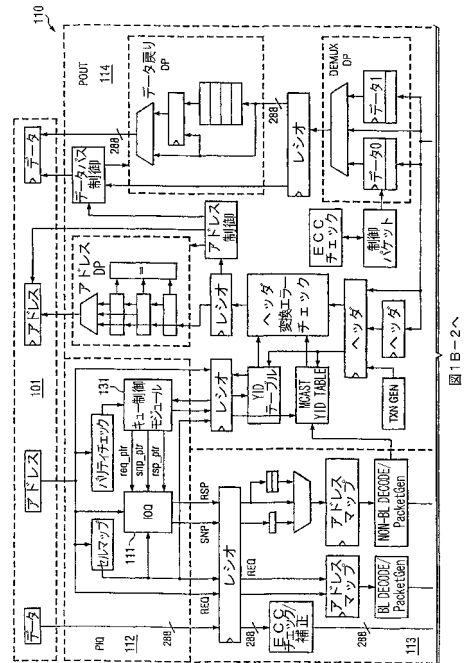
30

40

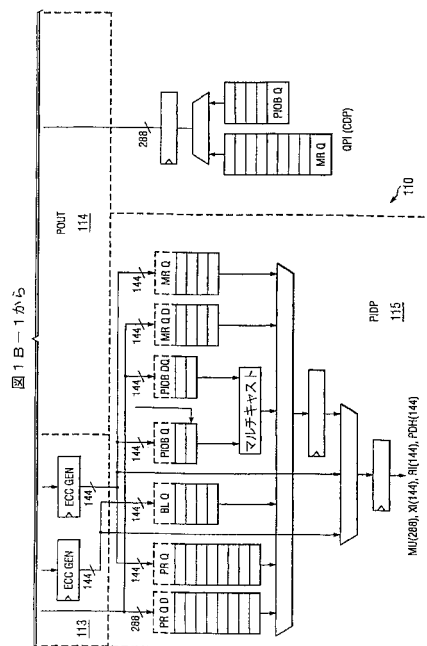
【図 1 A】



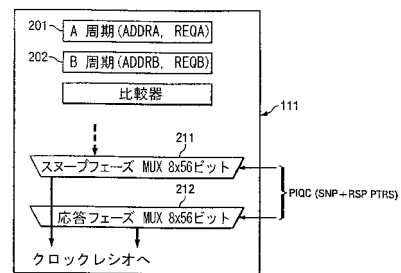
【図 1 B - 1】



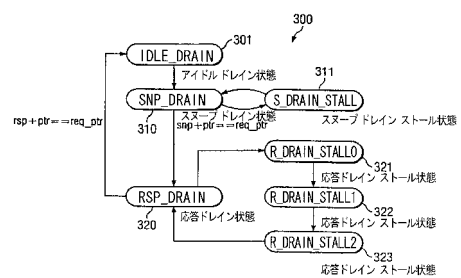
【図 1 B - 2】



【図 2】



【図 3】



[illegible]

フロントページの続き

(56)参考文献 特開平09-251412(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 9/46 - 9/54