

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3594082号

(P3594082)

(45) 発行日 平成16年11月24日(2004.11.24)

(24) 登録日 平成16年9月10日(2004.9.10)

(51) Int. Cl.⁷

F I

G O 6 F 12/10

G O 6 F 12/10 5 O 1 Z

G O 6 F 12/08

G O 6 F 12/10 5 5 7

G O 6 F 15/177

G O 6 F 12/08 5 O 5 Z

G O 6 F 15/177 6 7 6 A

請求項の数 7 (全 24 頁)

(21) 出願番号 特願2001-239590 (P2001-239590)
 (22) 出願日 平成13年8月7日(2001.8.7)
 (65) 公開番号 特開2003-50743 (P2003-50743A)
 (43) 公開日 平成15年2月21日(2003.2.21)
 審査請求日 平成14年7月19日(2002.7.19)

(73) 特許権者 000004237
 日本電気株式会社
 東京都港区芝五丁目7番1号
 (74) 代理人 100123788
 弁理士 宮崎 昭夫
 (74) 代理人 100088328
 弁理士 金田 暢之
 (74) 代理人 100106297
 弁理士 伊藤 克博
 (74) 代理人 100106138
 弁理士 石橋 政幸
 (72) 発明者 四宮 潔
 東京都港区芝五丁目7番1号 日本電気株
 式会社内

最終頁に続く

(54) 【発明の名称】 仮想アドレス間データ転送方式

(57) 【特許請求の範囲】

【請求項1】

中央処理装置と、データを記憶する主記憶装置と、仮想アドレスを実アドレスに変換するための変換情報を保持するアドレス変換表と、前記アドレス変換表の変換情報の一部を保持する T L B が組み込まれた前記主記憶装置のデータを読み出して複数の計算機を接続する交換網に送信する送信装置と、前記アドレス変換表の変換情報の一部を保持する T L B が組み込まれた前記交換網からのデータを受信して前記主記憶装置に書き込む受信装置とが実装された複数の計算機を有し、前記複数の計算機間で、データ転送元及び転送先のアドレスを前記仮想アドレスで指定して前記主記憶装置のデータを前記交換網を介して互いに転送するとともに、前記計算機内の前記送信装置及び前記受信装置の内部で、データ転送元及び転送先の仮想アドレスに対応する変換情報を前記アドレス変換表から前記 T L B に登録し、該 T L B に登録した変換情報を用いて前記仮想アドレスから前記実アドレスへのアドレス変換を行う仮想アドレス間データ転送方式において、
 前記送信装置は、当該送信装置が実装された計算機内の主記憶装置のデータを他の計算機内の受信装置に送信する場合、当該送信装置が実装された計算機内の主記憶装置からデータを読み出すことと平行して、前記他の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報を前記他の計算機内の T L B に事前登録することを指示することを特徴とする仮想アドレス間データ転送方式。

10

【請求項2】

前記受信装置は、当該受信装置が実装された計算機内の主記憶装置のデータを要求元の計

20

算機の主記憶装置に転送するリモートリード命令を該要求元の計算機内の送信装置から指示された場合、当該受信装置が実装された計算機内の主記憶装置からのデータ読み出しと平行して、前記要求元の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報を前記要求元の計算機内の T L B に事前登録することを指示することを特徴とする請求項 1 に記載の仮想アドレス間データ転送方式。

【請求項 3】

前記送信装置及び前記受信装置は、前記主記憶装置における予め決められた記憶空間の大きさであるページ毎に前記アドレス変換を行うことを特徴とする請求項 2 に記載の仮想アドレス間データ転送方式。

【請求項 4】

前記送信装置は、当該送信装置が実装された計算機内の主記憶装置のデータを他の計算機内の受信装置に送信する場合、データ送信開始時に前記他の計算機内の受信装置に対して前記 T L B への事前登録を指示し、更に、前記他の計算機内の前記主記憶装置の転送先の仮想アドレスがページ境界越えを起こした時にも、該他の計算機内の受信装置に対して前記 T L B への事前登録を指示することを特徴とする請求項 3 に記載の仮想アドレス間データ転送方式。

【請求項 5】

前記送信装置は、他の計算機の主記憶装置内のデータを前記送信装置が実装された計算機内の主記憶装置を転送先として転送するために、他の計算機に転送要求を送信する際、転送要求元である前記転送先の仮想アドレスがページ境界越えを起こした場合に、他の計算機の受信装置に対してアドレス変換情報の事前登録を指示するフラグをパケットに付加した転送要求を送信し、前記他の計算機の受信装置は、前記先読みを指示するフラグがパケットに付加されている場合にのみ、前記転送要求を送信した計算機の受信装置に対して前記 T L B への事前登録を指示するパケットを送信することを特徴とする請求項 3 または請求項 4 に記載の仮想アドレス間データ転送方式。

【請求項 6】

前記送信装置は、前記 T L B への事前登録を指示する際に、データ転送先の仮想アドレスを含む T L B 先読みパケットを送信することを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の仮想アドレス間データ転送方式。

【請求項 7】

前記受信装置は、前記 T L B への事前登録を指示する際に、データ転送先の仮想アドレスを含む T L B 先読みパケットを送信することを特徴とする請求項 2 乃至 6 のいずれか 1 項に記載の仮想アドレス間データ転送方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、データ転送元及び転送先のアドレスを仮想アドレスで指定して互いにデータ転送を行う複数の計算機を有する仮想アドレス間データ転送方式に関し、特に、各計算機内の送信装置及び受信装置が、仮想アドレスを実アドレスに変換するための変換情報の一部を保持する T L B をそれぞれ具備し、前記 T L B に保持された変換情報を用いて仮想アドレスを実アドレスに変換する仮想アドレス間データ転送方式に関する。

【0002】

【従来の技術】

近年、単一のプロセッサの処理能力を超える処理が計算機システムに要求され、更に、計算機システムに要求される処理能力の向上分が単一のプロセッサの性能向上分をも上回る傾向にある。

【0003】

このため、計算機システムとして、複数の C P U を使用し、複数の C P U でメモリ空間を共有するメモリ共有型のマルチプロセッサ装置が、より広範囲に使用されつつある。

【0004】

10

20

30

40

50

しかしながら、メモリ共有型のマルチプロセッサ装置においては、プロセッサに組み込まれているメモリアクセス性能の向上のためのキャッシュ間の整合性、いわゆるキャッシュ・コヒーレンシを維持しようとする、その他の性能が低下する場合があります、特に大規模のマルチプロセッサ装置でその傾向が高い。このため、メモリ共有型のマルチプロセッサ装置においては、メモリ空間を共有するプロセッサの台数を無制限に増加できないという第1の問題がある。

【0005】

前記第1の問題を解決するために、研究用途等に用いる大規模数値演算装置として、複数の計算機を交換網で接続することにより、複数の計算機でメモリ空間を共有する必要がないマルチ計算機装置が使用されている。最近では、一般の計算機においても、キャッシュの整合性を保証しない複数の計算機を交換網で接続したマルチ計算機装置を利用する機会が増えている。

10

【0006】

しかしながら、上記のマルチ計算機装置においては、複数の計算機のそれぞれで処理したデータを別の計算機に転送する必要があるため、データ量が増加した場合には、データ処理をするためのプロセッサにおける計算機間のデータ転送に消費する時間が増加し、マルチ計算機装置全体としての処理性能が低下するという第2の問題が発生する。

【0007】

前記第2の問題を解決するために、前記マルチ計算機装置においては、各計算機内にデータ転送専用の転送装置を組み込み、前記転送装置によりプロセッサが一度に処理するデータサイズよりも大きな単位でデータ転送を行うことで、転送性能を向上させるとともにプロセッサがデータ転送に消費する時間を短縮し、マルチホスト装置全体の性能を向上させる装置が開発されている。

20

【0008】

一方、オペレーティング・システムは、初期は小型で簡単な構成であったが、最近は大規模で複雑な構成となっており、ユーザー・プロセスからオペレーティング・システム・プロセス(OSプロセス)に対して、上記のデータ転送を依頼する時のプロセス間の移行のためのコンテキストスイッチのオーバーヘッドを無視することができないという第3の問題がある。

【0009】

前記第3の問題を解決するために、データ転送時の転送元/転送先アドレスをユーザ・プロセスのアドレス空間である仮想アドレスで指定し、上記の転送装置にて仮想アドレスを実アドレスに変換する方式が開発されている。

30

【0010】

前記方式によれば、データ転送指示を行う度に、毎回オペレーティングシステムに仮想アドレスを実アドレスに変換する依頼の必要性を無くすことが可能となるため、コンテキストスイッチの頻度を下げることが可能となる。

【0011】

前記転送装置は、仮想アドレスを実アドレスに変換するための変換情報を保持するアドレス変換機能を備えるが、仮想アドレス全体の変換情報は巨大であるため、転送装置内に仮想アドレス全体の変換情報を保持するアドレス変換表を実装することは困難であるという第4の問題がある。

40

【0012】

前記第4の問題を解決するため、転送装置の外部に、仮想アドレス全体の変換情報を保持するアドレス変換表を配置し、前記転送装置内に、アドレス変換表の中から最近使用された一部の仮想アドレスに対する変換情報のみを保持する小さなアドレス変換表、すなわちTLB(Translation look aside buffer)を組み込んだ方式が広く使用されている。

【0013】

【発明が解決しようとする課題】

50

しかしながら、T L Bを使用した方式では、T L B内に必要な仮想アドレスの変換情報が保持されていない場合、すなわちT L Bミスである場合は、転送装置外部のアドレス変換表から必要な変換情報を読み出す必要があるが、前記読み出し速度が低速であるため、アドレス変換を高速に行うことができないことにより、T L Bミス時の性能が低下するという第5の問題がある。

【0014】

さらに、データ転送開始時や、仮想アドレスの管理単位であるページ境界越えを起こした時には、T L Bミスを発生する可能性が高いため、仮想アドレス間のデータ転送機能を使用する際の性能低下が発生するという第6の問題がある。

【0015】

前記第5の問題及び第6の問題を解決するため、仮想アドレスを用いてデータ転送を行う仮想アドレス間転送の性能を、物理アドレスを用いてデータ転送を行う物理アドレス間転送の性能と同等にする方式が望まれている。

【0016】

本発明は、前記第5の問題及び第6の問題を解決するため、転送データがデータの受信装置に届く前にアドレス変換情報の先読み要求を事前に送信することにより、アドレス変換に要する時間をデータ送信部のデータ読み出し時間の中に隠蔽し、アドレス変換を見かけ上高速に実行し、データ転送に要する時間を短縮するとともに、データ転送開始時及びページ境界越え時にT L Bミスが発生する可能性を低減させる仮想アドレス間データ転送方式を提供することを目的とする。

【0017】

【課題を解決するための手段】

前記目的を達成するために本発明は、中央処理装置と、データを記憶する主記憶装置と、仮想アドレスを実アドレスに変換するための変換情報を保持するアドレス変換表と、前記アドレス変換表の変換情報の一部を保持するT L Bが組み込まれた前記主記憶装置のデータを読み出して交換網に送信する送信装置と、前記アドレス変換表の変換情報の一部を保持するT L Bが組み込まれた前記交換網からのデータを受信して前記主記憶装置に書き込む受信装置とが実装された複数の計算機を有し、前記複数の計算機間で、データ転送元及び転送先のアドレスを前記仮想アドレスで指定して前記主記憶装置のデータを前記交換網を介して互いに転送するとともに、前記計算機内の前記送信装置及び前記受信装置の内部で、データ転送元及び転送先の仮想アドレスに対応する変換情報を前記アドレス変換表から前記T L Bに登録し、該T L Bに登録した変換情報を用いて前記仮想アドレスから前記実アドレスへのアドレス変換を行う仮想アドレス間データ転送方式において、前記送信装置は、当該送信装置が実装された計算機内の主記憶装置のデータを他の計算機内の受信装置に送信する場合、当該送信装置が実装された計算機内の主記憶装置からデータ読み出しと平行して、前記他の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報を前記他の計算機内のT L Bに事前登録することを指示することを特徴とする。

【0018】

また、要求元となる他の計算機内の送信装置から、自計算機内の主記憶装置のデータを前記要求元計算機内の主記憶装置に転送するリモート・リード命令を受け付けた前記受信装置は、前記自計算機内の主記憶装置からのデータ読み出しと平行して、前記要求元の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報を前記要求元の計算機内のT L Bに事前登録することを指示することを特徴とする。

【0019】

また、前記送信装置及び前記受信装置は、前記主記憶装置における予め決められた記憶空間の大きさであるページ毎に前記アドレス変換を行うことを特徴とする。

【0020】

また、前記送信装置は、当該送信装置が実装された計算機内の主記憶装置のデータを他の計算機内の受信装置に送信する場合、データ送信開始時に前記他の計算機内の受信装置に

10

20

30

40

50

対して前記 T L B への事前登録を指示し、更に、前記他の計算機内の前記主記憶装置の転送先の仮想アドレスがページ境界越えを起こした時にも、該他の計算機内の受信装置に対して前記 T L B への事前登録を指示することを特徴とする。

【 0 0 2 1 】

また、要求元となる他の計算機内の送信装置から、自計算機内の主記憶装置のデータを前記要求元計算機内の主記憶装置に転送するリモート・リード命令を受け付けた前記受信装置は、前記要求元の計算機の仮想アドレスがページ境界越えを起こした時に、該要求元の計算機内の受信装置に対して前記 T L B への事前登録を指示することを特徴とする。

【 0 0 2 2 】

また、前記送信装置は、前記 T L B への事前登録を指示する際に、計算機間に専用線を設けて行う方式、あるいはデータ転送先の仮想アドレスを含む T L B 先読みパケットを交換網を通して前記受信装置に送信する。

【 0 0 2 3 】

また、前記受信装置は、前記 T L B への事前登録を指示する際に、計算機間に専用線を設けて行う方式、あるいはデータ転送先の仮想アドレスを含む T L B 先読みパケットを交換網を通して前記受信装置に送信する。

【 0 0 2 4 】

(作用)

本発明においては、データ送信元の計算機内の送信装置において、データ送信先の計算機内の受信装置に対してデータを送信する場合、データ送信元の計算機内の主記憶装置からのデータ読み出しと平行して、データ送信先の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報を T L B へ事前登録することを指示することにより、データ送信元の計算機における主記憶装置からのデータ読み出しと、データ送信先の計算機におけるアドレス変換表から T L B への変換情報の読み出しとが並行して行われることになるため、データ受信装置におけるアドレス変換のためのアドレス変換表読み出しが、データ送信装置におけるデータ読み出し時間に隠蔽されるため、見かけ上のアドレス変換時間が短縮され、データ転送にかかる総所要時間が短縮される。

【 0 0 2 5 】

【発明の実施の形態】

以下に、本発明の実施の形態について図面を参照して説明する。

【 0 0 2 6 】

(第1の実施の形態)

図1は、本発明の仮想アドレス間データ転送方式の第1の実施の形態を示す図である。

【 0 0 2 7 】

図1に示すように本実施形態においては、交換網600を介して互いにデータ交換を行う計算機700a, 700b(以下、それぞれ「計算機1」「計算機2」と称する場合もある)が設けられている。

【 0 0 2 8 】

以下に、計算機700aの内部構成について説明する。なお、図1においては、計算機700bの内部構成が省略されているが、計算機700bの内部構成も計算機700aと同様である。

【 0 0 2 9 】

計算機700aは、種々の命令を処理する中央処理装置であるCPU100と、種々の命令及びデータを保持する主記憶装置300と、交換網600を介して他計算機700bとの間でデータを送受信する転送装置400と、CPU100、主記憶装置300及び転送装置400に接続され、主記憶装置300へのデータの書き込み及び読み出しを行うシステム制御装置200と、転送装置400に接続され、データ転送元及び転送先の論理アドレス(仮想アドレス)を物理アドレス(実アドレス)に変換するための変換情報として実アドレスベース511(図3、図4等参照)を保持するアドレス変換表500とから構成されている。

10

20

30

40

50

【0030】

図2を参照して転送装置400の構成について説明する。

【0031】

転送装置400は、交換網600を介して他計算機700bにデータを送信する送信装置410と、他計算機700bから交換網600を介して送信されてきたデータを受信する受信装置420とから構成されている。

【0032】

送信装置410及び受信装置420には、アドレス変換表500内部のアドレス変換表エントリのうち最近使用された一部のアドレス変換表エントリをアドレス変換に使用した仮想アドレスの一部である仮想アドレスベースと共にTLBエントリとして保持するアドレス変換用バッファTLB430, 440がそれぞれ組み込まれている。TLBはアドレス変換用の小容量のバッファを示し、Translation Look Aside Bufferの頭文字の略号である。以降ではTLBをアドレス変換バッファを示すものとして使用する。

10

【0033】

図3を参照して転送装置400内の送信装置410の構成について説明する。

【0034】

送信装置410は、システム制御装置を経由して主記憶装置300からのデータを登録するデータバッファ411と、CPU100からの転送指示をシステム制御装置100を経由して登録する転送指示処理部412と、ユーザプロセスの仮想アドレスで指定された転送元アドレスを、主記憶装置300上のアドレスである実アドレスに変換するアドレス変換用バッファTLB430と、転送指示処理部412からの仮想アドレスとTLB内の各エントリの仮想アドレスベース431を比較し、一致した場合には実アドレスベース432から実アドレスを生成して転送指示処理部412に返却し、不一致の場合にはアドレス変換表500内の対応する実アドレスベースを読み出し、前記転送指示処理部から受け取った仮想アドレスと前記アドレス変換表500から読み出した実アドレスベース511を1組のTLBエントリとしてTLBに登録し、再度TLBを索引した後に一致した仮想アドレスベース431を持つTLBエントリの実アドレスベース432から実アドレスを生成し転送指示処理部412に返却する、TLB430内に実装されているアドレス比較部433と、転送指示処理部412から指示された実アドレスの主記憶装置300上のデータをシステム制御装置200を経由して読み出す要求を発行するデータ読出部418と、転送指示処理部412からの指示とデータバッファ411から受け取ったデータからパケットを生成し、交換網600に出力するデータ出力部417とから構成されている。

20

30

【0035】

データバッファ411は、主記憶装置300から読み出されたデータをシステム制御装置200を経由して受け付け、一時的に保持し、前記読み出されたデータをデータ出力部417に転送する。

【0036】

転送指示処理部412は、CPU100からの転送指示をシステム制御装置200を経由して受け付け、前記転送指示を命令800に保持する。命令800は命令コード801、転送先仮想アドレス(RV)413、転送回数414(LEN)、転送元仮想アドレス(SV)415、及び宛先計算機番号(DST)416から構成される。転送指示処理部412は命令800に従ってデータ出力部417に転送パケットの生成及び交換網600への出力を指示する。

40

【0037】

また、転送指示処理部412は、予め決められたデータ長分のデータを転送回数414で指示された回数分転送するまでの間、1パケット分のデータを転送する度に転送先仮想アドレス413及び転送元仮想アドレス415を再計算する。なお、転送先仮想アドレス413及び転送元仮想アドレス415の再計算は、あらかじめ定められた1パケットのデータ転送長を元に再計算され、例えば1回のデータ転送長を64バイトとすると、新仮想ア

50

ドレス = (旧仮想アドレス + 64) により計算する。

【0038】

T L B 4 3 0 は、転送指示処理部 4 1 2 内の転送元仮想アドレス 4 1 5 を実アドレスに変換するために、アドレス変換表 5 0 0 内のアドレス変換表エントリである実アドレスベース 5 1 1 の一部のアドレス変換表エントリと、前記アドレス変換表エントリである実アドレスベース 4 4 2 を読み出す際に使用した仮想アドレスの一部である仮想アドレスベースを 1 組とする T L B エントリとして保持する。各 T L B エントリは、仮想アドレスの一部である仮想アドレス・ベース 4 3 1 及び実アドレスの一部である実アドレス・ベース 4 3 2 として保持される。

【0039】

アドレス比較部 4 3 3 は、T L B 4 3 0 の仮想アドレス・ベース 4 3 1 と転送元仮想アドレス 4 1 5 の一部である仮想アドレス・ベースとを比較し、該比較結果に基づいて T L B ヒット/ミスと判定する。転送元仮想アドレス 4 1 5 の一部である仮想アドレス・ベースと一致する T L B エントリが仮想アドレス・ベース 4 3 1 に登録されていた場合を T L B ヒットとし、登録されていない場合を T L B ミスとする。

【0040】

データ読出部 4 1 8 は、アドレス比較部 4 3 3 にて T L B ヒットと判定された場合に、T L B 4 3 0 の実アドレス・ベース 4 3 2 と転送元仮想アドレス 4 1 5 の一部であるオフセットとから実アドレスを生成し、主記憶装置 3 0 0 の前記生成された実アドレスからのデータ読み出しをシステム制御装置 2 0 0 に要求する。

【0041】

データ出力部 4 1 7 は、転送指示処理部 4 1 2 にて受け付けられた命令 8 0 0 に従い、データバッファ 4 1 1 からデータ出力部 4 1 7 に転送されるデータと、転送指示処理部 4 1 2 内の命令コード 8 0 1 と、転送先アドレス 4 1 3 と、転送元アドレス 4 1 5 と、宛先計算機番号 4 1 6 とを用いてリモート・ライト・パケット、リモート・リード・パケット及び T L B 先読みパケットを生成し、交換網 6 0 0 に送信する。なお、リモート・ライト・パケット、リモート・リード・パケット及び T L B 先読みパケットについての詳細な説明は後述する。

【0042】

図 3 を参照して、アドレス変換表 5 0 0 について説明する。

【0043】

アドレス変換表 5 0 0 は複数のアドレス変換表エントリから構成され、アドレス変換表エントリは実アドレスベース 5 1 1 を保持する。複数の実アドレスベース 5 1 1 はそれぞれの実アドレスベース番号が付けられ、送信装置 4 1 0 あるいは受信装置 4 2 0 からの実アドレスベース読み出し要求内の仮想アドレスベースを実アドレスベース番号として実アドレスベースを読み出し、読み出した実アドレスベースを返却される。

【0044】

図 4 を参照して転送装置 4 0 0 内の受信装置 4 2 0 の構成について説明する。

【0045】

受信装置 4 2 0 は、交換網 6 0 0 からの転送されてきたパケットのデータを受け取るデータバッファ 4 2 1 と、交換網 6 0 0 から転送されてきたパケットのデータ以外の部分を受け取り、受け取ったパケット内の命令項目の内容がリモートライト命令であった場合には、宛先仮想アドレスを T L B 4 4 0 を使用して実アドレスに変換し、該変換した実アドレスに対してデータバッファ 4 2 1 が交換網 6 0 0 から受け取ったデータをメモリアクセス部 4 2 3 を経由してシステム制御装置 2 0 0 に主記憶装置 3 0 0 に書き込むことを指示し、前記受け取ったパケット内の命令項目の内容がリモートリード命令であった場合には、メモリアクセス部 4 2 3 に対してシステム制御装置 2 0 0 を経由して主記憶装置 3 0 0 内のデータの読み出しを指示し、前記読み出したデータを交換網出力部 4 2 5 を経由して交換網 6 0 0 に転送することを指示し、前記受け取ったパケット内の命令項目が T L B 先読みであった場合には、前記受け取ったパケット内の宛先仮想アドレスを実アドレスに変換

10

20

30

40

50

することをTLB440に指示し、前記仮想アドレスを実アドレスに変換するためのTLBエントリを事前に登録することを指示するコマンドアドレスバッファ422と、コマンドアドレスバッファ422からの指示によりデータバッファ421からのデータをシステム制御装置200を経由して主記憶装置300にライトするか、システム制御装置200を経由して主記憶装置300内のデータを読み出し、前記読み出したデータをアドレス・コマンド・バッファ422に返却するメモリアクセス部423と、コマンド・アドレスバッファ422からの指示により交換網600にパケットを出力する交換出力部425と、コマンド・アドレスバッファ422の指示により仮想アドレスを実アドレスに変換するTLB440とから構成されている。

【0046】

TLB440はコマンド・アドレス・バッファ422からの指示によりTLBエントリを検索し、ヒットであればTLB内の実アドレスベース442を用いて実アドレスを生成し、TLBミスした場合、アドレス変換表500から読み出した実アドレスベースにより、実アドレスを生成し、前記生成した実アドレスをコマンドアドレスバッファ422に返却するアドレス比較部443と、仮想アドレスベース441と実アドレスベース442を1組のTLBエントリとし、複数あるいは単数のTLBエントリを保持するTLBエントリ配列444とから構成される。

【0047】

データバッファ421は、他計算機700bから交換網600を介して送信されてきたデータを受信し、一時的に保持し、前記受信データをメモリアクセス部423に転送する。

【0048】

TLB440は、コマンドアドレスバッファ422に受信された転送先仮想アドレスを実アドレスに変換するために、アドレス変換表500におけるエントリの一部を保持する。各エントリは、仮想アドレスの一部である仮想アドレス・ベース441及び実アドレスの一部である実アドレス・ベース442として保持される。

【0049】

アドレス比較部443は、TLB440の仮想アドレス・ベース441とコマンドアドレスバッファ422に受信された転送先仮想アドレスの一部である仮想アドレス・ベースとを比較し、その比較結果に基づいてTLBヒット/ミス进行判定する。コマンドアドレスバッファ422に受信された転送先仮想アドレスの一部である仮想アドレス・ベースと一致するエントリが仮想アドレス・ベース441に存在した場合をTLBヒットとし、存在しない場合をTLBミスとする。

【0050】

メモリアクセス部423は、アドレス比較部443にてTLBヒットと判定された場合に、TLB440の実アドレス・ベース442とコマンドアドレスバッファ422に受信された転送先仮想アドレスの一部であるオフセットとから実アドレスを生成し、主記憶装置300の実アドレスへのデータ書き込みをシステム制御装置200に要求する。

【0051】

交換網出力部425は、コマンドアドレスバッファ422に受信された転送先仮想アドレス等を用いてリモート・ライト・パケット及びTLB先読みパケットを生成し、交換網600に送信する。なお、リモート・ライト・パケット及びTLB先読みパケットについての詳細な説明は後述する。

【0052】

図5を参照して、本実施形態における仮想アドレスを実アドレスに変換するアドレス変換動作を説明する。

【0053】

仮想アドレスを実アドレスに変換する場合は、送信装置410及び受信装置420にそれぞれ組み込まれているTLB430, 440を用いる。

【0054】

TLB430, 440は、TLBエントリ配列434, 444を持ち、複数或いは単数の

10

20

30

40

50

T L B エントリを仮想アドレス・ベース 4 3 1 , 4 3 2、実アドレス・ベース 4 4 1 , 4 4 2 として保持している。

【 0 0 5 5 】

図 5 では仮想アドレスがビット 3 1 からビット 1 3 までの 1 9 ビットの仮想アドレスベースとビット 1 2 からビット 0 までの 1 3 ビットのオフセットで構成されているが、前記ビット数は説明のためであり、仮想アドレスベースとオフセットのビット数及び仮想アドレス全体のビット数は任意のビット数を使用することができる。

【 0 0 5 6 】

T L B 索引時には、全ての T L B エントリの仮想アドレス・ベース 0 ~ N と、実アドレスに変換しようとする変換対象の仮想アドレスの仮想アドレス・ベースとが比較され、前記変換対象の仮想アドレスの仮想アドレス・ベースと一致する T L B エントリが仮想アドレス・ベース 0 ~ N のいずれかに存在した場合を T L B ヒットとし、存在しない場合を T L B ミスとする。

10

【 0 0 5 7 】

T L B ヒットの場合は、上記の一致した T L B エントリ内の実アドレス・ベースを変換後の実アドレスの実アドレス・ベースとし、前記変換対象の仮想アドレスの下位 1 3 ビットのオフセットを変換せずにそのまま上記の実アドレス・ベースと結合したものを変換後の実アドレスとする。

【 0 0 5 8 】

一方、T L B ミスの場合は、アドレス変換表 5 0 0 内のアドレス変換表エントリの中から前記変換対象仮想アドレスの仮想アドレスベースの値を実アドレスベース番号とし、アドレス変換表エントリを読み出し、前記読み出したアドレス変換表エントリの実アドレスベース 5 1 1 を T L B 4 3 0 , 4 4 0 の T L B エントリに登録した後、再度 T L B 4 3 0 , 4 4 0 を索引して前記アドレス変換を行う。前記 T L B エントリ登録後のアドレス再度のアドレス変換は、エントリ登録済みであるため、必ず T L B ヒットとなる。

20

【 0 0 5 9 】

次に、図 6 を参照して、アドレス変換表 5 0 0 におけるアドレス変換表エントリを T L B 4 3 0 , 4 4 0 の T L B エントリに登録する動作について説明する。

【 0 0 6 0 】

アドレス変換表 5 0 0 におけるアドレス変換表エントリ (図 6 では「実アドレス・ベース 1 ~ N 」) を変換対象の仮想アドレスの仮想アドレス・ベースを実アドレスベース番号として索引し、前記仮想アドレス・ベースに対応する実アドレス・ベース 5 1 1 を読み出す。

30

【 0 0 6 1 】

続いて、T L B 4 3 0 , 4 4 0 の T L B エントリから L R U 等の予め決められた手順で T L B エントリを選択し、前記選択された T L B エントリに、前記読み出された実アドレス・ベースと該実アドレス・ベースの読み出しに用いた仮想アドレス・ベースとを 1 組の T L B エントリとして登録する。本 T L B エントリ登録動作により、T L B 4 3 0 , 4 4 0 の値が更新される。

【 0 0 6 2 】

なお、登録 T L B エントリの選択方法としては、ランダム、L R U (L e a s t R e c e n t l y U s e d)、或いはラウンドロビン等の幾つかの方法を用いることができる。

40

【 0 0 6 3 】

また、アドレス変換表 5 0 0 は、転送装置 4 0 0 を使用するユーザ・プロセスによる初期化時に、図示しない経路で更新される。

【 0 0 6 4 】

次に、図 7 を参照して、自計算機データを他計算機に書き込むリモート・ライト動作について説明する。

【 0 0 6 5 】

50

図7は計算機1のデータを計算機2に書き込む場合のリモート・ライト動作を示す。

【0066】

まず、計算機1のCPU100及び計算機2のCPU100は、計算機1、計算機2内のそれぞれのアドレス変換表500に、データ転送を命令する各プロセスのアドレス変換に必要なアドレス変換表500内の実アドレス・ベース511を設定する(順序1)。

【0067】

次に、計算機1のCPU100は、送信装置410に対してリモート・ライト命令を、転送指示装置412内の命令800に、命令コード801、転送元仮想アドレス415、転送先仮想アドレス413、及びデータ転送長414、宛先計算機番号416として設定する(順序2)。

10

【0068】

次に、計算機1の送信装置410は、順序2で設定された転送先仮想アドレスを仮想アドレスとして含むTLB先読みパケットを生成し、計算機2の受信装置420へ転送することにより、計算機2の受信装置420にTLB先読みを要求する(順序3)。

【0069】

次に、計算機1の送信装置410は、順序2で設定された転送元仮想アドレスをTLB430を使用して転送元実アドレスに変換する(順序4)

また、順序4においては、計算機2の受信装置420は、順序3で計算機1の送信装置410から転送されてきたTLB先読みパケットを受信し、前記TLB先読みパケット内の仮想アドレスでTLB440を索引して仮想アドレスを実アドレスに変換する。

20

【0070】

順序4におけるTLB索引がTLBミスである場合、計算機1の送信装置410はアドレス変換表500の実アドレス・ベース511の中から転送元仮想アドレスに対応する実アドレス・ベースを読み出し、該読み出した実アドレス・ベースと前記転送元仮想アドレスの仮想アドレス・ベースとを1組のTLBエントリとしてTLB430に登録し、再度TLB430を索引してアドレス変換を行う(順序5)。TLBエントリ登録後のTLB再索引は、必要な仮想アドレス・ベースに登録した直後に行われるため、TLBヒットとなり、アドレス変換が完了する。

【0071】

また、順序5においては、計算機2の受信装置420は、順序4におけるTLB索引がTLBミスである場合、アドレス変換表500の実アドレス・ベース511の中から仮想アドレスに対応する実アドレス・ベースを読み出し、読み出した実アドレス・ベースと仮想アドレスの仮想アドレス・ベースとを1組のTLBエントリとしてTLB440に登録し、再度TLB440を索引してアドレス変換を行う。前記TLB再索引は、必要な仮想アドレス・ベースに登録した直後に行われるため、TLBヒットとなり、アドレス変換が完了する。

30

【0072】

次に、計算機1の送信装置410は、システム制御装置200を経由して主記憶装置300の転送元実アドレスから書き込みデータを読み出す(順序6)。

【0073】

次に、計算機1の送信装置410は、順序2で設定された転送先仮想アドレスと、順序6で読み出した書き込みデータとを含むリモート・ライト・パケットを生成し、計算機2の受信装置420へ転送する(順序7)。

40

【0074】

次に、計算機2の受信装置420は、順序7で計算機1の送信装置410から転送されたリモート・ライト・パケットを受信し、該リモート・ライト・パケット内の転送先仮想アドレスでTLB440を索引し、転送先仮想アドレスを転送先実アドレスに変換する(順序8)。

【0075】

順序8におけるTLB索引がTLBミスである場合、計算機2の受信装置420はアドレ

50

ス変換表 5 0 0 の実アドレス・ベース 5 1 1 の中から転送先仮想アドレスに対応する実アドレス・ベースを読み出し、読み出した実アドレス・ベースと転送元仮想アドレスの仮想アドレス・ベースとを 1 組の T L B エントリとして T L B 4 4 0 に登録し、再度 T L B 4 4 0 を索引してアドレス変換を行う。前記 T L B 再索引は、必要な仮想アドレス・ベースを登録した直後に行われるため、T L B ヒットとなり、アドレス変換が完了する（順序 9）。

【 0 0 7 6 】

順序 7 で受信されたリモート・ライト・パケット内の書き込みデータを計算機 2 の受信装置 4 2 0 がシステム制御装置 2 0 0 を経由して主記憶装置 3 0 0 の転送先実アドレスに書き込む（順序 1 0）。

10

【 0 0 7 7 】

なお、計算機 1 の送信装置 4 1 0 と計算機 2 の受信装置 4 2 0 との間では、コマンド及びデータがパケット単位で転送される。

【 0 0 7 8 】

図 8 を参照して、計算機 1 , 2 間で転送されるパケットの形式を説明する。

【 0 0 7 9 】

本実施形態においては、リモート・リード・パケットと、リモート・ライト・パケットと、T L B 先読みパケットとの 3 つのパケット形式を定義する。

【 0 0 8 0 】

図 8 (a) は、リモート・ライト・パケットのパケット形式を示す。

20

【 0 0 8 1 】

リモート・ライト・パケットは、命令項目の命令が W r i t e 、命令項目のデータ長がデータ項目の個数である 4 、計算機項目が本パケットの送信（発信）計算機番号及び受信（宛先）計算機番号、アドレス項目が転送先仮想アドレス、データ項目が 4 つの転送データである。データ項目の数は転送するデータ項目の個数によって代わり、本実施形態におけるデータ項目の個数 4 は一例である。

【 0 0 8 2 】

図 8 (b) は、リモート・リード・パケットのパケット形式を示す。

【 0 0 8 3 】

リモート・リード・パケットは、命令項目の命令が R e a d 、命令項目のデータ長が 1 、計算機項目が本パケットの送信（発信）計算機番号及び受信（宛先）計算機番号、アドレス項目が転送元仮想アドレス、データ項目が転送先仮想アドレスである。

30

【 0 0 8 4 】

図 8 (c) は、T L B 先読みパケットのパケット形式を示す。

【 0 0 8 5 】

T L B 先読みパケットは、命令項目の命令が T L B 、データ長が 0 、計算機項目が本パケットの送信（発信）計算機番号及び受信（宛先）計算機番号、アドレス項目が仮想アドレスであり、データ項目は存在しない。

【 0 0 8 6 】

図 8 (d) は、図 8 (a) ~ (c) に示した各パケットによる命令の動作を示す。

40

【 0 0 8 7 】

リモート・ライト・パケットは、命令発行元の計算機のデータを宛先計算機へ転送する。

【 0 0 8 8 】

また、リモート・リード・パケットは、宛先計算機のデータを命令発行元の計算機内に転送する。

【 0 0 8 9 】

また、T L B 先読みパケットは、宛先計算機の受信装置に対し、仮想アドレスで T L B を索引し、T L B ミスの場合はアドレス変換表 5 0 0 から必要な変換情報を読み出し、T L B 4 4 0 に登録するよう命令する。

【 0 0 9 0 】

50

次に、図9を参照して、他計算機からデータを読み出し、前記データを自計算機に書き込むリモート・リード動作について説明する。

【0091】

図9は、計算機2からデータを読み出し、前記データを計算機1に書き込む場合のリモート・リード動作を説明する。

【0092】

まず、計算機1のCPU100及び計算機2のCPU100は、計算機1、計算機2内のそれぞれのアドレス変換表500に、各プロセスのアドレス変換に必要な実アドレス・ベース511を設定する(順序1)。

【0093】

次に、計算機1のCPU100は、送信装置410に対してリモート・リード命令を、転送指示装置412内の命令800に、命令コード801、転送元仮想アドレス415、転送先仮想アドレス413、及びデータ転送長414、宛先計算機番号416として設定する(順序2)。

【0094】

次に、計算機1の送信装置410は、順序2で設定された命令800内の命令コード801、宛先計算機番号416、転送元仮想アドレス415及び転送先仮想アドレス413を含むリモート・リード・パケットを生成し、計算機2の受信装置420へ交換網600を経由して転送する(順序3)。

【0095】

次に、計算機2の受信装置420は、順序3で計算機1の送信装置410から転送されてきたリモート・リード・パケットを受信し、前記リモート・リード・パケット内の転送先仮想アドレスを宛先仮想アドレスとして含むTLB先読みパケットを生成し、前記リモート・リード・パケットの送信元である計算機1へ転送する(順序4)。

【0096】

次に、計算機2の受信装置420は、順序4で受信したリモート・リード・パケット内の転送元仮想アドレスでTLB440を索引して転送元仮想アドレスを転送元実アドレスに変換する(順序5)。

【0097】

また、順序5においては、計算機1の受信装置420は、順序4で計算機2の受信装置420から転送されてきたTLB先読みパケットを受信し、前記TLB先読みパケット内の宛先仮想アドレスでTLB430を索引して仮想アドレスを実アドレスに変換する。

【0098】

順序5におけるTLB索引がTLBミスであった場合、計算機1の受信装置420はアドレス変換表500の実アドレス・ベース511の中から転送元仮想アドレスに対応する実アドレス・ベースを読み出し、前記読み出した実アドレス・ベースと仮想アドレスの仮想アドレス・ベースとを1組のTLBエントリとしてTLB430に登録し、再度TLB430を索引してアドレス変換を行う(順序6)。前記TLB再索引は、必要な仮想アドレス・ベースに登録した直後に行われるため、TLBヒットとなり、アドレス変換が完了する。

【0099】

また、順序6においては、計算機2の受信装置420は、順序5におけるTLB索引がTLBミスであった場合、アドレス変換表500の実アドレス・ベース511の中から仮想アドレスに対応する実アドレス・ベースを読み出し、読み出した実アドレス・ベースと転送元仮想アドレスの仮想アドレス・ベースとを1組のTLBエントリとしてTLB440に登録し、再度TLB440を索引してアドレス変換を行う。前記TLB再索引は、必要な仮想アドレス・ベースに登録した直後に行われるため、TLBヒットとなり、アドレス変換が完了する。

【0100】

次に、計算機2の受信装置420は、システム制御装置200を経由して主記憶装置30

10

20

30

40

50

0の前記アドレス変換後の転送元実アドレスから転送データを読み出す(順序7)。

【0101】

次に、計算機2の受信装置420は、順序4で受信したリモート・リード・パケット内の転送先仮想アドレスと、順序7で主記憶装置300から読み出した転送データとを含むリモート・ライト・パケットを生成し、リモート・リード・パケットの送信元である計算機1へ転送する(順序8)。

【0102】

次に、計算機1の受信装置420は、順序8で計算機2の受信装置420から転送されてきたリモート・ライト・パケットを受信し、前記リモート・ライト・パケット内の転送先仮想アドレスでTLB430を索引し、転送先仮想アドレスを転送先実アドレスに変換する(順序9)。

10

【0103】

順序9におけるTLB索引がTLBミスであった場合、計算機1の受信装置420はアドレス変換表500の実アドレス・ベース511の中から転送先仮想アドレスに対応する実アドレス・ベースを読み出し、読み出した実アドレス・ベースと転送先仮想アドレスの仮想アドレス・ベースとを1組のTLBエントリとしてTLB430に登録し、再度TLB430を索引してアドレス変換を行う(順序10)。前記TLB再索引は、必要な仮想アドレス・ベースに登録した直後に行われるため、TLBヒットとなり、アドレス変換が完了する。

【0104】

20

その後、計算機1の受信装置420は、順序8で受信したリモート・ライト・パケット内の読み出しデータをシステム制御装置200を経由して主記憶装置300の前記アドレス変換後の転送先実アドレスに書き込む(順序11)。

【0105】

図10を参照して、送信装置410の動作について詳細に説明する。

【0106】

まず、転送指示処理部412がCPU100からの転送命令800を受け付け(ステップS110)、前記転送命令800を、命令コード801、転送先仮想アドレス(RV)413、転送回数414(LEN)、転送元仮想アドレス(SV)415、及び宛先計算機番号(DST)416として保持する。

30

【0107】

次に、転送指示処理部412に受け付けられた転送指示がリモート・ライト命令であるかが判断され(ステップS120)、リモート・ライト命令である場合、データ出力部417は、転送指示処理部412内の転送先仮想アドレス413及び宛先計算機番号416と、自計算機の送信計算機番号とを含むTLB先読みパケットを生成し、交換網600に送信する(ステップS130)。前記TLB先読みパケットは、交換網600により宛先計算機番号416の他計算機内の受信装置420に転送される。

【0108】

次に、アドレス比較部433は、転送指示処理部412内の転送元仮想アドレス415でTLB430を索引し(ステップS140)、前記索引がTLBミスであるか否かを判定する(ステップS150)。TLB索引時、アドレス比較部433は、転送元仮想アドレス415の一部である仮想アドレス・ベースと、TLB430の仮想アドレス・ベース431とを比較し、前記比較結果が全てのTLBエントリにおいて不一致であった場合はTLBミスと判定し、一致したTLBエントリがある場合はTLBヒットと判定する。

40

【0109】

ステップS150にてTLBミスと判定されると、図6に示したように、アドレス変換表500の実アドレス・ベース511の中から転送元仮想アドレス415に対応する実アドレス・ベースが読み出され、読み出された実アドレス・ベースと転送元仮想アドレス415の一部である仮想アドレス・ベースとが、それぞれ実アドレス・ベース432及び仮想アドレス・ベース431に1組のTLBエントリとして登録される(ステップS160)

50

。TLBエントリ登録後に再度、TLB430を索引すると、必要な仮想アドレス・ベースを登録した直後であるため、TLBヒットとなる。

【0110】

ステップS150 あるいはステップS160にてTLBヒットと判定されると、データ読出部418は、図5に示したように、TLB430の実アドレス・ベース432のうち転送元仮想アドレス415の仮想アドレス・ベースに一致した実アドレス・ベースと、転送元仮想アドレス415の一部であるオフセットとから、転送元実アドレスを生成し、システム制御装置200を経由して主記憶装置300の前記転送元実アドレスから転送データを読み出し、データバッファ411に登録する(ステップS170)。

【0111】

次に、データ出力部417は、データバッファ411内の転送データと、転送指示処理部412内の転送先仮想アドレス413及び宛先計算機番号416と、自計算機の送信計算機番号とを含むリモート・ライト・パケットを生成し、交換網600に送信する(ステップS180)。前記リモート・ライト・パケットは、交換網600を経由して宛先計算機番号416の他計算機内の受信装置420に転送される。

【0112】

ステップS180における転送が終了すると、転送指示処理部412は、転送回数414を1減算して転送回数414を更新し(ステップS190)、更新した転送回数414が0よりも大きいかなかを判断する(ステップS200)。

【0113】

ステップS200にて転送回数414が0である場合は転送命令を終了し、また、転送回数414が0よりも大きい場合、すなわち転送命令を継続する場合は、転送先仮想アドレス413及び転送元仮想アドレス415を1パケットのデータ長を加算することで更新した後(ステップS210)、ステップS140における処理に戻る。例えば、1パケットのデータ長が16バイトの場合、新仮想アドレス=(旧仮想アドレス+16)により更新する。

【0114】

一方、ステップS120にて転送指示処理部412に受け付けられた転送命令がリモート・ライト命令ではなく、リモート・リード命令であると判断された場合、データ出力部417は、転送指示処理部412内の転送先仮想アドレス413、転送元仮想アドレス415及び宛先計算機番号416と、自計算機の送信計算機番号とを含むリモート・リード・パケットを生成し、交換網600に送信する(ステップS300)。前記リモート・リード・パケットは、交換網600により宛先ホスト番号416の他計算機内の受信装置420に転送される。

【0115】

ステップS300における転送が終了すると、転送指示処理部412は、転送回数414を1減算して転送回数414を更新し(ステップS310)、更新した転送回数414が0よりも大きいかなかを判断する(ステップS320)。

【0116】

ステップS320にて転送回数414が0である場合は転送命令を終了し、また、転送回数414が0よりも大きい場合、すなわち転送命令を継続する場合は、転送先仮想アドレス413及び転送元仮想アドレス415を1パケットのデータ長を加算することで更新した後(ステップS330)、ステップS300における処理に戻る。例えば、1パケットのデータ長が16バイトの場合、新仮想アドレス=(旧仮想アドレス+16)により更新する。

【0117】

図11を参照して、受信装置420の動作について詳細に説明する。

【0118】

まず、コマンドアドレスバッファ422が交換網600からパケットを受信する(ステップR110)。

10

20

30

40

50

【 0 1 1 9 】

次に、コマンドアドレスバッファ 4 2 2 に受信されたパケットがリモート・リード・パケットであるか否かが判断され（ステップ R 1 2 0）、リモート・リード・パケットである場合、交換網出力部 4 2 5 は、リモート・リード・パケット内の発信元計算機番号及び転送先仮想アドレスをそれぞれ宛先計算機番号及び仮想アドレスとして含む T L B 先読みパケットを生成し、交換網 6 0 0 に送信する（ステップ R 1 3 0）。前記 T L B 先読みパケットは、交換網 6 0 0 により宛先計算機番号の他計算機内の受信装置 4 2 0 に転送される。

【 0 1 2 0 】

次に、アドレス比較部 4 4 3 は、リモート・リード・パケット内の転送元仮想アドレスで T L B 4 4 0 を索引し（ステップ R 1 4 0）、前記索引が T L B ミスであるか否かを判定する（ステップ R 1 5 0）。ステップ R 1 5 0 の T L B 索引時、アドレス比較部 4 4 3 は、リモート・リード・パケット内の転送元仮想アドレスの一部である仮想アドレス・ベースと、T L B 4 4 0 の仮想アドレス・ベース 4 4 1 とを比較し、前記比較結果が全ての T L B エントリにおいて不一致であった場合は T L B ミスと判定し、一致した T L B エントリがある場合は T L B ヒットと判定する。

10

【 0 1 2 1 】

ステップ R 1 5 0 にて T L B ミスと判定された場合は、図 6 に示したように、アドレス変換表 5 0 0 の実アドレス・ベース 5 1 1 の中からリモート・リード・パケット内の転送元仮想アドレスに対応する実アドレス・ベースが読み出され、読み出された実アドレス・ベースとリモート・リード・パケット内の転送元仮想アドレスに対応する仮想アドレスの一部である仮想アドレス・ベースとが、それぞれ実アドレス・ベース 4 4 2 及び仮想アドレス・ベース 4 4 1 に 1 組の T L B エントリとして登録される（ステップ R 1 6 0）。前記 T L B エントリ登録後、再度、T L B 4 4 0 を索引すると、必要な仮想アドレス・ベースを登録した直後であるため、T L B ヒットとなる。

20

【 0 1 2 2 】

ステップ R 1 5 0 或いはステップ R 1 6 0 にて T L B ヒットと判定されると、メモリアクセス部 4 2 3 は、図 5 に示したように、T L B 4 4 0 の実アドレス・ベース 4 4 2 のうちリモート・リード・パケット内の転送元仮想アドレスの仮想アドレス・ベースに一致した実アドレス・ベースと、リモート・リード・パケット内の転送元仮想アドレスの一部であるオフセットとから、転送元実アドレスを生成し、システム制御装置 2 0 0 を経由して主記憶装置 3 0 0 の前記転送元実アドレスから転送データを読み出す（ステップ R 1 7 0）。

30

【 0 1 2 3 】

交換網出力部 4 2 5 は、主記憶装置 3 0 0 から読み出した転送データと、宛先計算機番号（リモート・リード・パケット内の発信計算機番号）と、リモート・リード・パケット内の転送先仮想アドレスとを含むリモート・ライト・パケットを生成し、交換網 6 0 0 に送信する（ステップ R 1 8 0）。前記リモート・ライト・パケットは、交換網 6 0 0 により宛先計算機番号の計算機内の受信装置 4 2 0 に転送される。

【 0 1 2 4 】

以上で、リモート・リード・パケットの処理が完了する。

40

【 0 1 2 5 】

一方、ステップ R 1 2 0 にてコマンドアドレスバッファ 4 2 2 に受信されたパケットがリモート・リード・パケット以外であると判断された場合、そのパケットがリモート・ライト・パケットであるか否かが判断される（ステップ R 3 0 0）。

【 0 1 2 6 】

ステップ R 3 0 0 にてコマンドアドレスバッファ 4 2 2 に受信されたパケットがリモート・ライト・パケットであると判断された場合、アドレス比較部 4 4 3 は、リモート・ライト・パケット内の転送先仮想アドレスで T L B 4 4 0 を索引し（ステップ R 3 1 0）、前記索引が T L B ミスであるか否かを判定する（ステップ R 3 2 0）。前記 T L B 索引時、

50

アドレス比較部 4 4 3 は、リモート・ライト・パケット内の転送先仮想アドレスの一部である仮想アドレス・ベースと、T L B 4 4 0 の仮想アドレス・ベース 4 4 1 とを比較し、前記比較結果が全ての T L B エントリにおいて不一致であった場合は T L B ミスと判定し、一致したエントリがある場合は T L B ヒットと判定する。

【 0 1 2 7 】

ステップ R 3 2 0 にて T L B ミスと判定された場合は、図 6 に示したように、アドレス変換表 5 0 0 の実アドレス・ベース 5 1 1 の中からリモート・ライト・パケット内の転送先仮想アドレスに対応する実アドレス・ベースが読み出され、読み出された実アドレス・ベースとリモート・ライト・パケット内の転送先仮想アドレスの一部である仮想アドレス・ベースとがそれぞれ実アドレス・ベース 4 4 2 及び仮想アドレス・ベース 4 4 1 に 1 組の T L B エントリとして登録される（ステップ R 3 3 0）。T L B 4 4 0 の再索引では、必要な仮想アドレス・ベースを登録した直後であるため、T L B ヒットとなる。

10

【 0 1 2 8 】

ステップ R 3 2 0 或いはステップ R 3 3 0 にて T L B ヒットと判定されると、メモリアクセス部 4 2 3 は、T L B 4 4 0 の実アドレス・ベース 4 4 2 のうちリモート・ライト・パケット内の転送先仮想アドレスの仮想アドレス・ベースに一致した実アドレス・ベースと、リモート・ライト・パケット内の転送先仮想アドレスの一部であるオフセットとから、図 5 に示したように転送先実アドレスを生成し、システム制御装置 2 0 0 を経由して主記憶装置 3 0 0 の前記転送先実アドレスにデータバッファ 4 2 1 内の転送データをメモリアクセス部 4 2 3 に転送し、メモリアクセス部 4 2 3 はシステム制御装置 2 0 0 を経由して主記憶装置 3 0 0 に書き込む（ステップ R 3 4 0）。

20

【 0 1 2 9 】

以上で、リモート・ライト・パケットの処理が完了する。

【 0 1 3 0 】

一方、ステップ R 3 0 0 にてコマンドアドレスバッファ 4 2 2 に受信されたパケットがリモート・ライト・パケットでなく、T L B 先読みパケットであると判断された場合、アドレス比較部 4 4 3 は、T L B 先読みパケット内の仮想アドレスで T L B 4 4 0 を索引し（ステップ R 4 0 0）、前記索引が T L B ミスであるか否かを判定する（ステップ R 4 1 0）。前記 T L B 索引時、アドレス比較部 4 4 3 は、T L B 先読みパケット内の仮想アドレスの仮想アドレス・ベースと、T L B 4 4 0 の仮想アドレス・ベース 4 4 1 とを比較し、前記比較結果が全ての T L B エントリにおいて不一致であった場合は T L B ミスと判定し、一致したエントリがある場合は T L B ヒットと判定する。

30

【 0 1 3 1 】

ステップ R 4 1 0 にて T L B ミスと判定された場合は、図 6 に示したように、アドレス変換表 5 0 0 の実アドレス・ベース 5 1 1 の中から T L B 先読みパケット内の仮想アドレスに対応する実アドレス・ベースが読み出され、読み出された実アドレス・ベースと T L B 先読みパケット内の仮想アドレスの一部である仮想アドレス・ベースとが、それぞれ実アドレス・ベース 4 4 2 及び仮想アドレス・ベース 4 4 1 に 1 組の T L B エントリとして登録される（ステップ R 4 2 0）。

【 0 1 3 2 】

以上で、T L B 先読みパケットの処理が完了する。

40

【 0 1 3 3 】

（第 2 の実施の形態）

上述した第 1 の実施形態においては、送信装置 4 1 0 は、C P U 1 0 0 からのリモート・ライト指示を受け付けた後、データ送信開始時にのみ T L B 先読みパケットを送信し、また、受信装置 4 2 0 は、全てのリモート・リード・パケットに対して T L B 先読みパケットを送信している。

【 0 1 3 4 】

しかしながら、データ送信開始時にのみ T L B 先読みパケットを送信する構成では、アドレス変換をページと呼ばれる予め決められた主記憶空間の大きさ毎に行う場合に、データ

50

送信中に仮想アドレス・ベースが変化する時、即ちページ境界越え時にTLBミスとなる可能性が高いため、ページ境界越え時にデータ転送速度が低下する問題がある。

【0135】

また、全てのリモート・リード・パケットに対してTLB先読みパケットを送信する構成では、交換網の帯域幅の浪費、及びTLB索引頻度の無意味な増加を引き起こす問題がある。

【0136】

本実施形態は、前記2つの問題を解決するために、送信装置410は、データ送信開始時、及びデータ転送先の仮想アドレスがページ越えをした時に、TLB先読みパケットを送信し、また、受信装置420は、データ転送先の仮想アドレスがページ越えをした時のみ、リモート・リード・パケットに対してTLB先読みパケットを送信する構成とすることで、ページ境界越え時の性能低下、交換網帯域幅浪費及びTLB索引頻度増の問題を解消するものである。

10

【0137】

図12に、第1の実施形態からの相違点を明確にするために、送信装置410内の転送指示処理部412の転送先仮想アドレス計算部分のみを示す。

【0138】

第1の実施形態における転送指示処理部412は、転送先仮想アドレス413と1パケットのデータ長であるパケット当たりのデータ長710とを仮想アドレス加算部711で加算し、仮想アドレス加算部711の出力を転送先仮想アドレス413としている。

20

【0139】

これに対して、本実施形態における転送指示処理部412は、仮想アドレス加算部711の出力を転送先仮想アドレス413とするとともに、ページ境界越え検出信号712としてデータ出力部417に入力する。

【0140】

すなわち、転送先仮想アドレス413がページ境界越えを起こした場合には、更新された転送先仮想アドレス413によるTLB先読み要求をページ境界越え検出信号712としてデータ出力部417に送信する。

【0141】

ページ境界検出機能により、データ出力部417は、リモート・ライト命令の最初のリモート・ライト・パケット送信時、あるいは転送指示処理部412からのTLB先読み要求を受け付けた場合にも、TLB先読みパケットを交換網600を経由して宛先計算機に送信する。

30

【0142】

また、データ出力部417は、リモート・リード・パケット送信中に、転送指示処理部412からのTLB先読み要求を受け付けた場合には、TLB先読み要求フラグをセットしたリモート・リード・パケットを送信し、また、リモート・リード・パケット送信中に、TLB先読み要求を受け付けてない場合には、TLB先読み要求フラグをセットしないリモート・リード・パケットを送信する。

【0143】

TLB先読み要求フラグ付きのリモート・リード・パケットを図8(e)に示す。

40

【0144】

図8(e)ではTLB先読み要求フラグPFが命令項目に追加されており、前記PF=1のリモート・リード・パケットを受け付けた受信装置420は、前記リモート・リード・パケット送信元にTLBの先読みパケットを送信し、前記PF=0のリモート・リード・パケットを受け付けた受信装置420は前記TLB先読みパケットを送信しない。

【0145】

第2の実施形態の受信装置420の動作を第1の実施形態の相違点のみを示すフローチャートを図13に示す。なお、図13において、「R+数字」のステップが第1の実施形態のステップを示し、「RX+数字」のステップが第2の実施形態で修正されたステップを

50

示す。

【0146】

図13を参照して、第2の実施形態における受信装置420の動作を説明する。

【0147】

受信装置420がリモート・リード・パケットを受信した場合、(ステップR120)、ステップRX100において、TLB先読み要求フラグPFを確認し、前記PF=1の場合はステップR130に遷移する。ステップRX100において、前記PF=0の場合はステップR140に遷移する。

【0148】

【発明の効果】

以上説明したように本発明においては、計算機間のデータを転送する場合、データ送信元の計算機内の主記憶装置からのデータ読み出しと同時に、データ送信先の計算機内の受信装置に対して、データ転送先の仮想アドレスに対応する変換情報をTLBへ事前登録することを指示する構成とし、データ送信元の計算機における主記憶装置からのデータ読み出しと、データ送信先の計算機におけるアドレス変換表からTLBへの仮想アドレス変換情報の読み出しとが並行して行うことにより、データ送信先の計算機でデータが受信されてからアドレス変換情報を読み出しTLBへの登録を行う従来方式と比較して、アドレス変換時間を見かけ上短縮する第1の効果がある。

【0149】

更に、データ送信先の計算機におけるTLB読み出し時間の一部或いは全てが、データ送信元の計算機におけるデータ読み出し時間に隠蔽されることから、データ送信先の計算機内では、受信したデータを一時的に保持するためのデータバッファにてデータがアドレス変換表読み出しを待つ時間が短縮されるため、データ送信先の計算機内でのデータ待ちバッファの容量を縮小することができ、使用するハードウェアの量を削減することができる第2の効果がある。

【0150】

また、データ送信元の計算機内の受信装置において、データ送信先の計算機内の送信装置からの指示に従って該データ送信先の計算機内の受信装置に対してデータを送信する場合にも、データ送信元の計算機内の主記憶装置からのデータ読み出しと同時に、データ送信先の計算機内の受信装置に対して、当該データの転送先の仮想アドレスに対応する変換情報をTLBへ事前登録することを指示する構成とすることで、上記と同様の効果が得られる第3の効果がある。

【0151】

また、データ送信元の計算機内の送信装置において、データ送信先の計算機内の受信装置にデータを送信する場合、データ送信開始時にデータ送信先の計算機内の受信装置にTLBへの事前登録を指示し、更に、データ送信先の計算機内の主記憶装置の仮想アドレスがページ境界越えを起こした時にもTLBへの事前登録を指示する構成とすることにより、データ送信開始時及びページ境界越え時に、TLB内に必要な仮想アドレスの変換情報が登録されていない状態、すなわちTLBミスが発生することによる性能低下の可能性を低減することができるという第4の効果がある。

【0152】

また、データ送信元の計算機内の受信装置において、データ送信先の計算機内の送信装置からの指示に従って該データ送信先の計算機内の受信装置に対してデータを送信する場合、データ送信先の計算機内の主記憶装置の仮想アドレスがページ境界越えを起こした時のみ、TLBへの事前登録を指示する構成とすることにより、TLB索引頻度の過剰な増加を抑制することができるとともに、交換網の帯域幅を低下させることができるという第5の効果がある。

【図面の簡単な説明】

【図1】本発明における仮想アドレス間データ転送方式の第1の実施の形態を示す図である。

10

20

30

40

50

【図 2】図 1 に示した転送装置の構成を説明するための図である。

【図 3】図 2 に示した送信装置の構成を説明するための図である。

【図 4】図 2 に示した受信装置の構成を説明するための図である。

【図 5】図 1 に示した仮想アドレス間データ転送方式における、T L B を用いたアドレス変換動作を説明するための図である。

【図 6】図 1 に示した仮想アドレス間データ転送方式における、アドレス変換表の読み出し動作を説明するための図である。

【図 7】図 1 に示した仮想アドレス間データ転送方式におけるリモート・ライト動作を説明するための図である。

【図 8】本発明の仮想アドレス間データ転送方式に用いられるパケットのパケット形式を説明するための図であり、(a) はリモート・ライト・パケットのパケット形式を示す図、(b) はリモート・リード・パケットのパケット形式を示す図、(c) は T L B 先読みパケットのパケット形式を示す図、(d) は (a) ~ (c) に示した各パケットによる命令の動作を示す図、(e) は先読み指示フラグ付きのリモート・リードパケットを示す図である。

【図 9】図 1 に示した仮想アドレス間データ転送方式におけるリモート・リード動作を説明するための図である。

【図 10】図 3 に示した送信装置の動作を説明するためのフローチャートである。

【図 11】図 4 に示した受信装置の動作を説明するためのフローチャートである。

【図 12】本発明の仮想アドレス間データ転送方式の第 2 の実施の形態に用いられる送信装置内の転送処理部を説明するための図である。

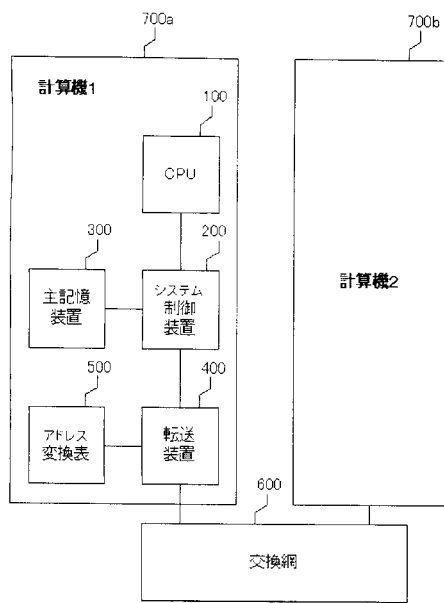
【図 13】本発明の仮想アドレス間データ転送方式の第 2 の実施の形態に用いられる受信装置の動作を説明するためのフローチャートである。

【符合の説明】

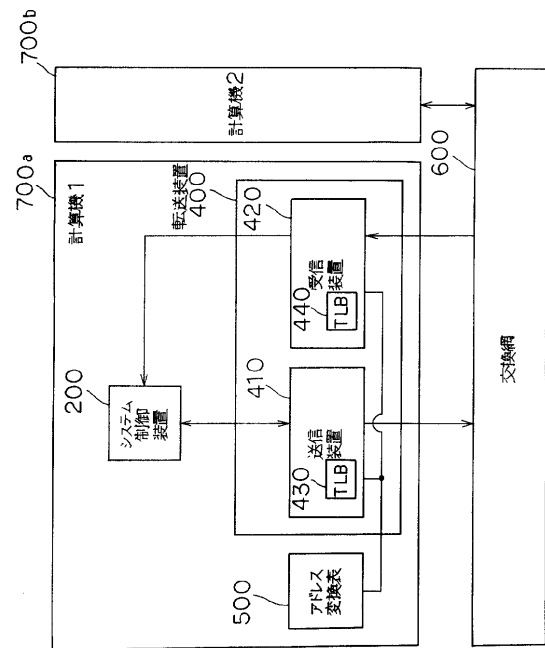
1 0 0	中央処理装置	
2 0 0	システム制御装置	
3 0 0	主記憶装置	
4 0 0	データ転送装置	
4 1 0	送信装置	
4 1 1	送信装置のデータバッファ	30
4 1 2	送信装置の転送指示処理部	
4 1 3	転送先仮想アドレス	
4 1 4	転送回数	
4 1 5	転送元仮想アドレス	
4 1 6	宛先計算機番号	
4 1 7	送信装置のデータ出力部	
4 1 8	送信装置のデータ読出部	
4 3 3	送信装置のアドレス比較部	
4 2 0	受信装置	
4 2 1	受信装置のデータバッファ	40
4 2 2	受信装置のコマンドアドレスバッファ	
4 2 3	受信装置のメモリアクセス部	
4 4 3	受信装置のアドレス比較部	
4 2 5	受信装置の交換網出力部	
4 3 0	送信装置の T L B	
4 3 1	送信装置の T L B の仮想アドレス・ベース部	
4 3 2	送信装置の T L B の実アドレス・ベース部	
4 4 0	受信装置の T L B	
4 4 1	受信装置の T L B の仮想アドレス・ベース部	
4 4 2	受信装置の T L B の実アドレス・ベース部	50

- 500 アドレス変換表
- 600 交換網
- 700 a , 700 b 計算機
- 710 データ長
- 711 仮想アドレス加算部
- 712 ページ境界越え検出信号

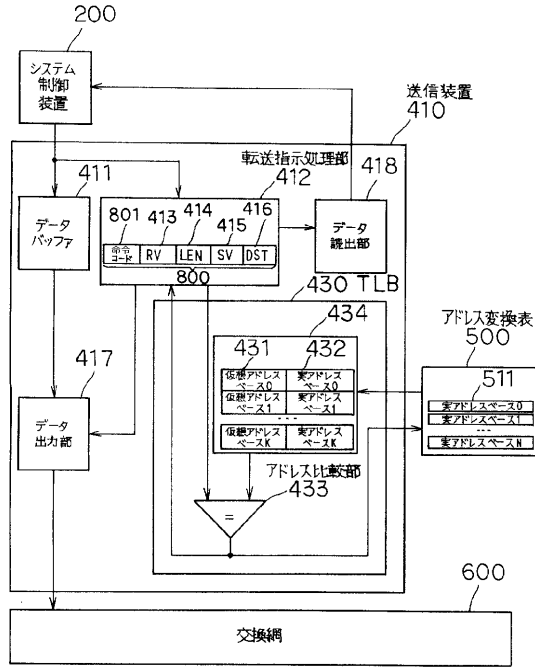
【図1】



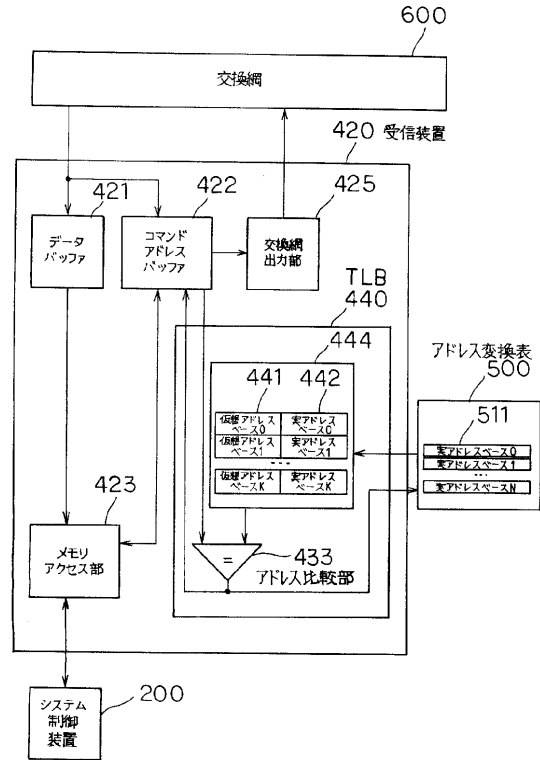
【図2】



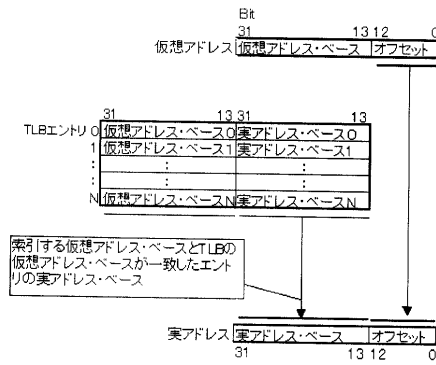
【図3】



【図4】



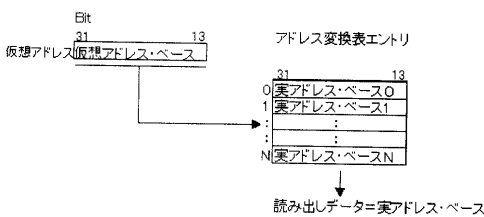
【図5】



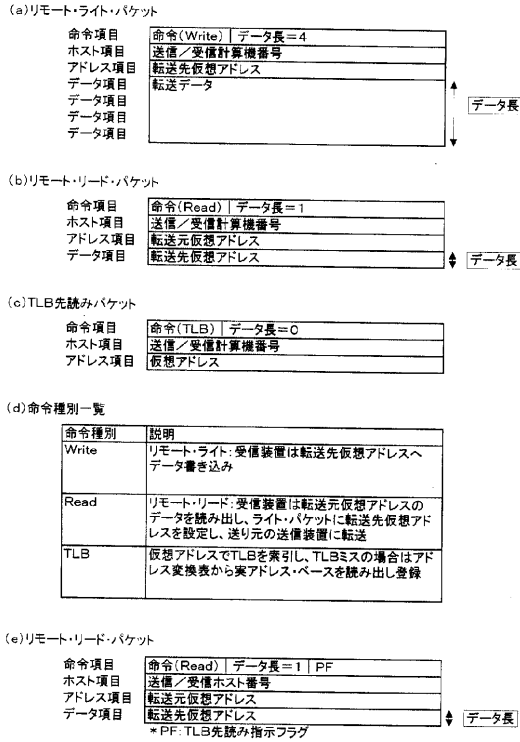
【図7】

	計算機1		計算機2	
	CPU100	送信装置410	受信装置420	CPU100
順序1	アドレス変換表500設定			アドレス変換表500設定
順序2		リモートライト命令		
順序3		転送先仮想アドレスを仮想アドレスとして含むTLB先読みバケットを計算機2へ転送		
順序4		転送元仮想アドレスでTLBを参照し転送元実アドレスに変換	TLB先読みバケット内の仮想アドレスでTLBを参照し実アドレスに変換	
順序5	TLBミスの場合は、アドレス変換表500を用いて転送元実アドレスに変換		TLBミスの場合は、アドレス変換表500を用いて実アドレスに変換	
順序6	主記憶装置300の転送元実アドレスから書き込みデータを読み込む			
順序7		書き込みデータと転送先仮想アドレスを含むライトバケットを計算機2へ転送		
順序8			ライトバケット内の転送先仮想アドレスでTLBを参照し転送先実アドレスに変換	
順序9			TLBミスの場合は、アドレス変換表500を用いて転送先実アドレスに変換	
順序10				書き込みデータを主記憶装置300の転送先実アドレスに書き込む

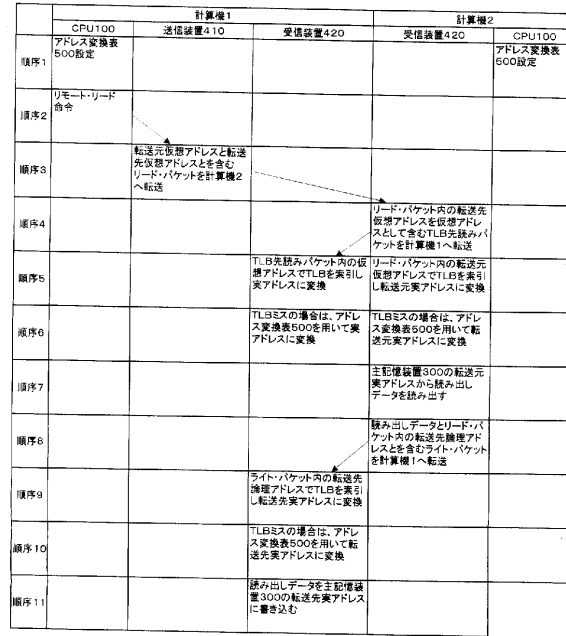
【図6】



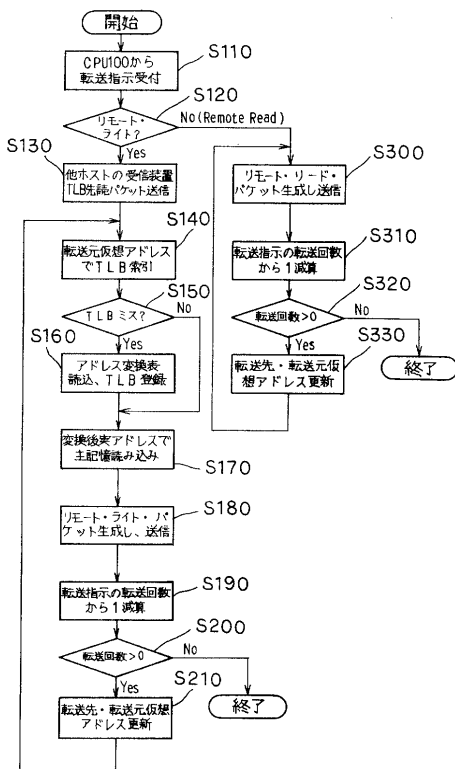
【図 8】



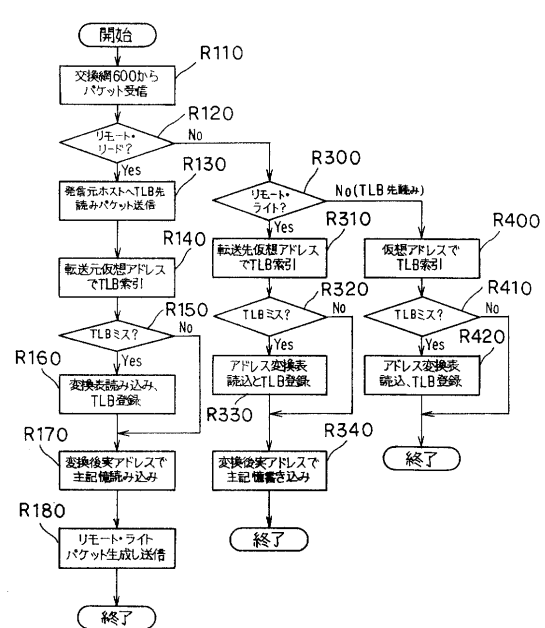
【図 9】



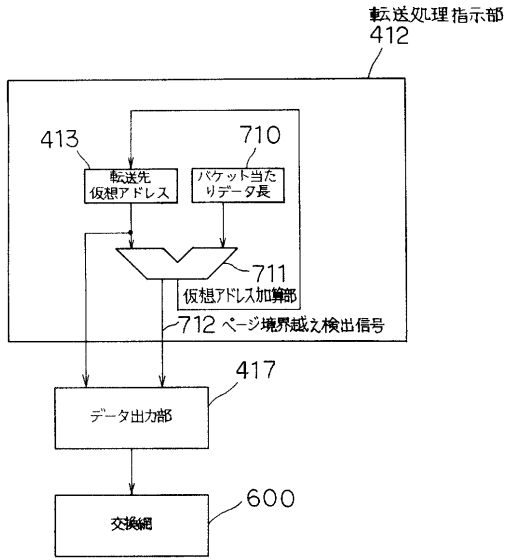
【図 10】



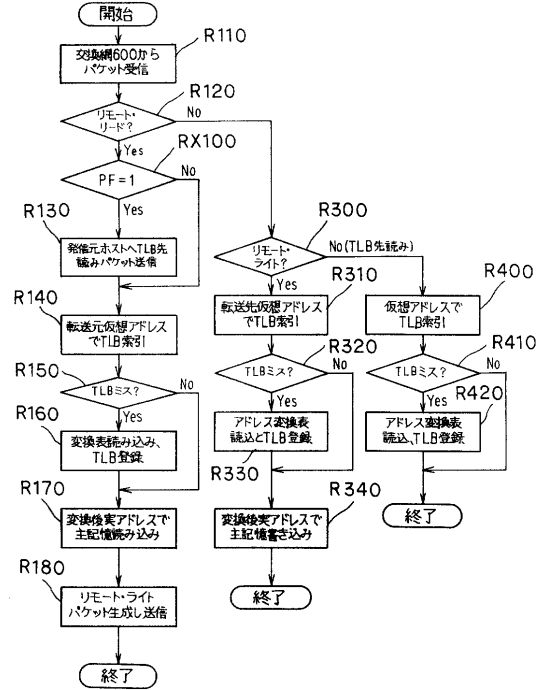
【図 11】



【 図 1 2 】



【 図 1 3 】



フロントページの続き

審査官 清木 泰

- (56)参考文献 特開平6 - 19785 (JP, A)
特開平10 - 275129 (JP, A)
特開2000 - 330960 (JP, A)
特開2000 - 67009 (JP, A)
特開平8 - 241293 (JP, A)
特開平5 - 181751 (JP, A)
特開平5 - 89056 (JP, A)
加納健、外4名、並列コンピュータCenju-4のユーザレベルメッセージ通信機構、並列処理シンポジウムJSP'99論文集、日本、社団法人情報処理学会、1999年6月9日、p. 7 - 14
松岡浩司、外7名、超並列計算機RWC-1における記憶構成、情報処理学会研究報告、日本、社団法人情報処理学会、1993年8月19日、第93巻、第71号、(93-ARC-101)、p. 17 - 24
國澤亮太、外2名、アドレス変換ハードウェアで支援されたメモリベース通信の性能評価、電子情報通信学会技術研究報告、日本、社団法人電子情報通信学会、1998年8月4日、第98巻、第233号、(CPSY98-48~59)、p. 61 - 66
中條拓伯、外3名、ネットワーク結合型並列計算機上の仮想共有メモリシステムにおける無矛盾化プロトコルの性能評価と、並列処理シンポジウムJSP'91論文集、日本、社団法人情報処理学会、1991年5月、p. 45 - 52、(論文タイトルの続き)ハードウェアによる実現
清水謙多郎、分散処理&分散OSの基礎知識、インターフェース、日本、CQ出版株式会社、1991年10月1日、第17巻、第10号、p. 124 - 139
工藤知宏、外4名、Network based Parallel ComputingのためのNetwork Interfaceの評価、電子情報通信学会技術研究報告、日本、社団法人電子情報通信学会、1998年8月5日、第98巻、第234号、(CPSY98-60~73)、p. 1 - 8
Creve Maples, A HIGH-PERFORMANCE, MEMORY-BASED INTERCONNECTION SYSTEM FOR MULTICOMPUTER ENVIRONMENTS, Proceedings of Supercomputing'90, IEEE, 1990年11月12日, pages:295-304

(58)調査した分野(Int.Cl.⁷, DB名)

G06F12/08-12/12
G06F15/16-15/177
G06F13/00-13/14
G06F13/20-13/42
G06F 9/46- 9/54