

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局

(43) 国際公開日  
2013年4月4日(04.04.2013)



(10) 国際公開番号  
WO 2013/046814 A1

- (51) 国際特許分類:  
G06F 13/36 (2006.01) G06F 13/38 (2006.01)  
G06F 1/26 (2006.01)
- (21) 国際出願番号: PCT/JP2012/065108
- (22) 国際出願日: 2012年6月13日(13.06.2012)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2011-210483 2011年9月27日(27.09.2011) JP
- (71) 出願人(米国を除く全ての指定国について): 日本電気株式会社(NEC Corporation) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 吉川 隆士 (YOSHIKAWA, Takashi) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 飛鷹 洋一 (HIDAKA, Youichi) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 鈴木 順 (SUZUKI,

Jun) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 樋口 淳一 (HIGUCHI, Junichi) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 稲葉 良幸, 外(INABA, Yoshiyuki et al.); 〒1066123 東京都港区六本木6-10-1 六本木ヒルズ森タワー23階 TMI 総合法律事務所 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: DISTRIBUTED COMPUTER SYSTEM

(54) 発明の名称: 分散型計算機システム

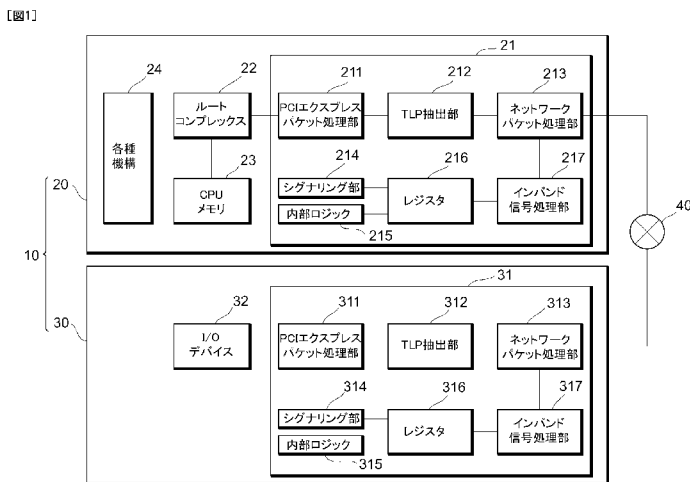
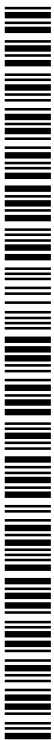


FIG. 1:  
 22 Route complex  
 23 CPU memory  
 24 Mechanisms  
 32 I/O device  
 211, 311 PCI Express packet processing unit  
 212, 312 TLP extraction unit  
 213, 313 Network packet processing unit  
 214, 314 Signaling unit  
 215, 315 Internal logic  
 216, 316 Register  
 217, 317 In-band signal processing unit

(57) Abstract: A local computer (20) and a remote I/O device (30) are connected on a network (40) via an upstream bridge (21) and a downstream bridge (31). The upstream bridge (21) and downstream bridge (31) are provided with: registers (216, 316) for referencing or writing a signal other than the main data stream transmitted in the network (40), said referencing or writing being done from both of said bridges; and mechanisms (217, 317) that perform in-band sending/receiving of the signal other than the main data stream, when the registers (216, 316) are referenced or written to.

(57) 要約: ローカルコンピュータ(20)及びリモートI/Oデバイス(30)は、上流側ブリッジ(21)及び下流側ブリッジ(31)を介してネットワーク(40)上で接続している。上流側ブリッジ(21)及び下流側ブリッジ(31)は、ネットワーク(40)に流れる主たるデータストリーム以外の信号を双方から参照し又は書き込むためのレジスタ(216、316)と、レジスタ(216、316)への参照時又は書き込み時に主たるデータストリーム以外の信号をインバンドで送受信する機構(217、317)とを備える。



WO 2013/046814 A1



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

## 明 細 書

発明の名称：分散型計算機システム

### 技術分野

[0001] 本発明は、ローカルコンピュータとリモートI/Oデバイスがネットワーク上に分散配置される分散型計算機システムに関する。

### 背景技術

[0002] 図3は従来例に係わる標準のPCIエクスプレススイッチ60の構成を示す機能ブロック図である。PCIエクスプレススイッチ60は、上流側（ルート側）を終端する上流側ブリッジ61を備えており、上流側から送られたデータを上流側データバッファ62に格納し、データパス以外の論理処理の制御に係わる制御信号を上流側レジスタ63に格納する。下流側（エンドポイント側）に関しても同様に、PCIエクスプレススイッチ60は、下流側ブリッジ64を備えており、下流側データバッファ65及び下流側レジスタ66が下流側へのデータ送信を行う。

[0003] 非特許文献1, 2には、PCIエクスプレスの機能を拡張するための計算機システムとして、ローカルコンピュータ（CPU、メモリ及びチップセット等を含むマザーボード）とリモートI/Oデバイスをネットワーク上に分散配置し、イーサネット上に仮想的なPCIエクスプレススイッチを構築するシステムが提案されている。この計算機システムでは、ローカルコンピュータに接続する上流側ブリッジ、リモートI/Oデバイスに接続する下流側ブリッジ、及びこれら二つのブリッジを接続するイーサネットを一体的に含めて、図3に示す従来例に係わる標準のPCIエクスプレススイッチ60と等価な動作及び機能が実現されているため、分散型の計算機システムであるにも関わらず、市販のマザーボードやI/Oデバイスを用いてハードウェアを構成できる利点を有する。また、ソフトウェアに関しても、市販のオペレーティングシステムやドライバがそのまま使用できる利点を有する。更に、上流側ブリッジと下流側ブリッジとの間を流れる主たるデータストリームに関

しても、標準のイーサネットパケットを使用してそのペイロード部分にPCIエクスプレスのTLP（トランザクション・レイヤ・パケット）を載せてPCIエクスプレスの制御及びデータ通信を行っているため、市販のイーサネットスイッチがそのまま使用できる利点を有する。また、この計算機システムを用いてネットワーク上にデバイスを分散配置することにより、デバイス間を自由に接続でき、距離の制約や筐体の制約を受けずにシステムを拡張できる。また、リモートI/OデバイスのグループIDがローカルコンピュータのグループIDと同じになるようにリモートI/OデバイスのグループIDを設定することにより、リモートI/Oデバイスをローカルコンピュータに所属させることができる。

## 先行技術文献

## 非特許文献

[0004] 非特許文献1：2006年電子情報通信学会通信ソサイエティ大会B-6-57 イーサネットを用いたシステム仮想化技術ExpressEtherの提案（1）システム概要とアーキテクチャ

非特許文献2：2006年電子情報通信学会通信ソサイエティ大会B-6-58 イーサネットを用いたシステム仮想化技術ExpressEtherの提案（2）I/O仮想化技術

## 発明の概要

## 発明が解決しようとする課題

[0005] しかし、図3に示す従来例に係わるPCIエクスプレススイッチ60は、特定の制御信号（例えば、処理の同期のためのタイミング信号、各論理回路での条件分岐等の制御信号、ステートやモード等の状態を示す信号）を上流側ブリッジ61及び下流側ブリッジ64の双方から読み書きするための機構として、共通レジスタ67を使用しており、この特定の制御信号をTLPに載せて、上流側ブリッジ61及び下流側ブリッジ64の双方から読み書きするように構成されていない。このため、非特許文献1，2にて提案されてい

るようにPCIエクスプレススイッチの機能を、ネットワークを跨いで上流側ブリッジと下流側ブリッジとに分散配置すると、特定の制御信号が利用できなくなるという問題が生じる。

[0006] また、PCIエクスプレススイッチネットワークを構成するローカルコンピュータとリモートI/Oデバイスとの間では、TLPに載せて伝送される主たるデータストリーム以外の信号として、サイドバンド信号が送受信されている。サイドバンド信号には例えば電源管理やカードのハードウェア状態の管理に関する信号が含まれており、具体的には、スロットにカードが存在するか否かを示すPresence、ローカルコンピュータの管理データをやりとりするSMBus、SMCIk、JTAG信号であるTCK、TDI、TDO、TMS、電源ON要求のためのWAKE#、カードの電源が入っていることを示すPWRGD等がある。これらのサイドバンド信号は、PCIエクスプレススロットのコネクタを通じてロジックレベルのHigh/Low信号で伝送されているため、イーサネット上をTLPに載せて伝送することができなかった。

[0007] また、リモートI/Oデバイスをローカルコンピュータからネットワーク越しに遠隔に置くと、電源管理の上で、次に述べるような不都合が生じる。例えば、リモートI/OデバイスがローカルコンピュータのPCIエクスプレススロットに挿入されている状態では、ローカルコンピュータの電源とリモートI/Oデバイスの電源は、常に連動してオン/オフされるが、リモートI/Oデバイスがネットワーク越しにローカルコンピュータに接続する場合には、ローカルコンピュータの電源とリモートI/Oデバイスの電源は、相互に関連付けがなく、個別にオン/オフしなければならない。リモートI/Oデバイス側にキーボード、マウス、グラフィックディスプレイ等が配置され、ネットワーク越しのローカルコンピュータ側にマザーボードが配置された場合には、リモートI/Oデバイス側からローカルコンピュータの電源のオン/オフを制御したい要望があるものの、両者の電源は関連付けがないため、リモートI/Oデバイス側からローカルコンピュータの電源を制御す

ることができないという不都合が生じる。

[0008] そこで、本発明は、ネットワークを介して接続されるローカルコンピュータとリモート I/O デバイスとの間のハードウェアレベルのリソース管理又は制御を、リモート I/O デバイスがネットワークを介さずにローカルコンピュータに接続している環境と同等のレベルで実現することを課題とする。

### 課題を解決するための手段

[0009] 上記の課題を解決するため、本発明に係わる分散型計算機システムは、ローカルコンピュータに接続する上流側ブリッジと、リモート I/O デバイスに接続する下流側ブリッジとを備える。ローカルコンピュータ及びリモート I/O デバイスは、上流側ブリッジ及び下流側ブリッジを介してネットワーク上で接続している。上流側ブリッジ及び下流側ブリッジは、ネットワークに流れる主たるデータストリーム以外の信号を双方から参照し又は書き込むためのレジスタと、レジスタへの参照時又は書き込み時に主たるデータストリーム以外の信号をインバンドで送受信する機構とを備える。

### 発明の効果

[0010] 本発明によれば、ネットワークを介して接続されるローカルコンピュータとリモート I/O デバイスとの間のハードウェアレベルのリソース管理又は制御を、リモート I/O デバイスがネットワークを介さずにローカルコンピュータに接続している環境と同等のレベルで実現できる。

### 図面の簡単な説明

[0011] [図1]実施例 1 に係わる分散型計算機システムの構成を示す機能ブロック図である。

[図2]実施例 2 に係わる分散型計算機システムの構成を示す機能ブロック図である。

[図3]従来例に係わる PC I エクスプレススイッチの構成を示す機能ブロック図である。

### 発明を実施するための形態

[0012] 以下、各図を参照しながら本発明の実施例について説明する。同一の機能ブロックには、同一の符号を付すものとし、重複する説明を省略する。

## 実施例 1

[0013] 図1は実施例1に係わる分散型計算機システム10の構成を示す機能ブロック図である。分散型計算機システム10は、ローカルコンピュータ20に接続する上流側ブリッジ21と、リモートI/Oデバイス30に接続する下流側ブリッジ31とを備える。ローカルコンピュータ20及びリモートI/Oデバイス30は、上流側ブリッジ21及び下流側ブリッジ31を介してネットワーク40上で接続している。ローカルコンピュータ20は、例えば、マザーボードであり、その構成部品として、CPU及びメモリ23と、メモリI/O制御チップセットとしてのルートコンプレックス22と、その他の各種機構24を備える。各種機構24には、例えば、電源、ATA、PCIエクスプレススロット、筐体等が含まれる。リモートI/Oデバイス30は、I/Oデバイス32を備える。I/Oデバイス32は、PCIエクスプレス接続可能なエンドポイントであり、例えば、グラフィックインタフェースカード、ハードディスクコントローラ、ネットワークコントローラ等を含む。I/Oデバイス32は、リモート側に所在する場合には、下流側ブリッジ31に接続されるが、ローカル側に所在する場合には、ローカルコンピュータ20のPCIエクスプレススロットに接続される。上流側ブリッジ21及び下流側ブリッジ31は、従来のPCIエクスプレススイッチの機能を有し、更にその機能を拡張した仮想的なスイッチとして機能するように構成されている。このため、I/Oデバイス32は、リモート側に所在する場合であっても、ローカルコンピュータ20のPCIエクスプレススロットに接続されているのと等価な環境でルートコンプレックス22に接続できる。また、ローカルコンピュータ20上で動作するソフトウェア、オペレーティングシステム、BIOSからは、I/Oデバイス32がリモート側に所在するのか、或いはローカル側に所在するのかの区別に関係なく、I/Oデバイス32を利用できる。

[0014] ここで、上流側ブリッジ21と下流側ブリッジ31との間の主たるデータストリームの伝送について説明する。上流側ブリッジ21は、ネットワーク40に流れる主たるデータストリームを下流側ブリッジ31との間で送受信するための機構として、PCIエクスプレスパケット処理部211、TLP抽出部212、及びネットワークパケット処理部213を備える。同様に、下流側ブリッジ31は、ネットワーク40に流れる主たるデータストリームを上流側ブリッジ21との間で送受信するための機構として、PCIエクスプレスパケット処理部311、TLP抽出部312、及びネットワークパケット処理部313を備える。

[0015] まず、上流側から下流側への主たるデータストリームの流れについて説明する。PCIエクスプレスパケット処理部211は、主たるデータストリームをそのペイロードに含むPCIエクスプレスパケットをルートコンプレックス22から受け取り、その物理レイヤ及びデータリンクレイヤを終端する。TLP抽出部212は、PCIエクスプレスパケットからトランザクションレイヤのパケットであるTLPを抽出する。ネットワークパケット処理部213は、抽出されたTLPをネットワーク40の通信プロトコルに従った形式にカプセル化し、これをネットワーク40経由で下流側ブリッジ31に送出する。ネットワークパケット処理部313は、受信したパケットのネットワークプロトコルを終端する。TLP抽出部312は、受信したパケットからTLPを抽出する。PCIエクスプレスパケット処理部311は、抽出されたTLPにPCIエクスプレスの物理レイヤ及びデータリンクレイヤの各ヘッダを付加してPCIエクスプレスパケットを生成し、これをI/Oデバイス32に送出する。PCIエクスプレスパケット処理部311からI/Oデバイス32へ送出されるPCIエクスプレスパケットは、PCIエクスプレスパケット処理部211がルートコンプレックス22から受け取るPCIエクスプレスパケットと等価である。

[0016] 次に、下流側から上流側への主たるデータストリームの流れについて説明する。PCIエクスプレスパケット処理部311は、主たるデータストリー



ムをそのペイロードに含むPCIエクスプレス packets を I/O デバイス 32 から受け取り、その物理レイヤ及びデータリンクレイヤを終端する。TLP 抽出部 312 は、PCIエクスプレス packets からトランザクションレイヤの packets である TLP を抽出する。ネットワーク packets 処理部 313 は、抽出された TLP をネットワーク 40 の通信プロトコルに従った形式にカプセル化し、これをネットワーク 40 経由で上流側ブリッジ 21 に送出する。ネットワーク packets 処理部 213 は、受信した packets のネットワークプロトコルを終端する。TLP 抽出部 212 は、受信した packets から TLP を抽出する。PCIエクスプレス packets 処理部 211 は、抽出された TLP に PCIエクスプレスの物理レイヤ及びデータリンクレイヤの各ヘッダを付加して PCIエクスプレス packets を生成し、これをルートコンプレックス 22 に送出する。PCIエクスプレス packets 処理部 211 からルートコンプレックス 22 へ送出される PCIエクスプレス packets は、PCIエクスプレス packets 処理部 311 が I/O デバイス 32 から受け取る PCIエクスプレス packets と等価である。

[0017] 上述の一連の動作により、I/O デバイス 32 は、リモート側に所在する場合であっても、ローカルコンピュータ 20 の PCIエクスプレス スロットに接続されているのと等価な環境でルートコンプレックス 22 に接続できる。言い換えれば、従来の PCIエクスプレス スイッチの機能を、ネットワーク 40 を介して上流側ブリッジ 21 と下流側ブリッジ 31 に分散して実装したことと等価である。

[0018] 次に、上流側ブリッジ 21 と下流側ブリッジ 31 との間でネットワーク 40 を介して特定の制御信号を読み書きするための機構について説明する。特定の制御信号は、ネットワーク 40 に流れる主たるデータストリーム以外の信号であり、例えば、処理の同期のためのタイミング信号、各論理回路での条件分岐等の制御信号、及びステータスやモード等の状態を示す信号等を含む。特定の制御信号の具体例として、データの同期に用いられるストロブ (STROBE)、データの有効性を示すバリッド (VALID)、及び論理

処理に係わるステートマシンのステート（STATE）等が含まれる。上流側ブリッジ21は、特定の制御信号を読み書きするための機構として、シグナリング部214、内部ロジック215、レジスタ216、及びインバンド信号処理部217を備える。同様に下流側ブリッジ31は、特定の制御信号を読み書きするための機構として、シグナリング部314、内部ロジック315、レジスタ316、及びインバンド信号処理部317を備える。これらのレジスタ216、316は、上流側ブリッジ21及び下流側ブリッジ31の双方からアクセス可能に構成され、特定の制御信号を読み書きするためのものである。

[0019] まず、上流側から下流側への特定の制御信号の流れについて説明する。シグナリング部214は、内部ロジック215の論理処理等に係わるステートが変更されたとき、或いは下流側ブリッジ31からの参照要求を受信したときを契機として、レジスタ216の内容（特定の制御信号）を書き換える。レジスタ216は、その内容が書き換えられると、書き換え後の信号をネットワーク40経由のインバンド方式で送受信するための信号形式に変換し、インバンド信号処理部217に転送する。書き換え後の信号をインバンド信号処理部217に転送する契機は、レジスタ216の書き換えを検出したときでもよく、或いはインバンド方式で送信すべき信号があることを示すフラグが立っていることを検出したときでもよい。インバンド信号処理部217は、転送する信号とレジスタアドレスとをセットにして、TLPと同等の packetsizeのサイズを有するインバンド信号用パケットを生成し、これをネットワークパケット処理部213に転送する。ネットワークパケット処理部213は、TLP抽出部212からの主たるデータストリームの中にインバンド信号用パケットを挿入し、これをネットワーク40の通信プロトコルに従った形式にカプセル化し、ネットワーク40経由で下流側ブリッジ31に送出する。ネットワークパケット処理部313は、受信したカプセルのカプセル化を解除し、そのペイロードがTLPである場合には、これをTLP抽出部312に転送し、そのペイロードがインバンド信号用パケットである場

合には、これをインバンド信号処理部 3 1 7 に転送する。インバンド信号処理部 3 1 7 は、T L P と同等の packets 長に変換されたインバンド信号用 packets から信号とレジスタアドレスとを復元し、これをレジスタ 3 1 6 に書き込む。レジスタ 3 1 6 は、その内容が書き換えられると、シグナリング部 3 1 4 に通知する。シグナリング部 3 1 4 に通知する契機は、レジスタ 3 1 6 の内容の書き換えを検出したときでもよく、或いはレジスタ 3 1 6 の書き換えを示すフラグが立っていることを検出したときでもよい。

[0020] 次に、下流側から上流側への特定の制御信号の流れについて説明する。シグナリング部 3 1 4 は、内部ロジック 3 1 5 の論理処理等に係わる状態が変更されたとき、或いは上流側ブリッジ 2 1 からの参照要求を受信したときを契機として、レジスタ 3 1 6 の内容（特定の制御信号）を書き換える。レジスタ 3 1 6 は、その内容が書き換えられると、書き換え後の信号をネットワーク 4 0 経由のインバンド方式で送受信するための信号形式に変換し、インバンド信号処理部 3 1 7 に転送する。書き換え後の信号をインバンド信号処理部 3 1 7 に転送する契機は、レジスタ 3 1 6 の書き換えを検出したときでもよく、或いはインバンド方式で送信すべき信号があることを示すフラグが立っていることを検出したときでもよい。インバンド信号処理部 3 1 7 は、転送する信号とレジスタアドレスとをセットにして、T L P と同等の packets 長のサイズを有するインバンド信号用 packets を生成し、これをネットワーク packets 処理部 3 1 3 に転送する。ネットワーク packets 処理部 3 1 3 は、T L P 抽出部 3 1 2 からの主たるデータストリームの中にインバンド信号用 packets を挿入し、これをネットワーク 4 0 の通信プロトコルに従った形式にカプセル化し、ネットワーク 4 0 経由で上流側ブリッジ 2 1 に送出する。ネットワーク packets 処理部 2 1 3 は、受信したカプセルのカプセル化を解除し、そのペイロードが T L P である場合には、これを T L P 抽出部 2 1 2 に転送し、そのペイロードがインバンド信号用 packets である場合には、これをインバンド信号処理部 2 1 7 に転送する。インバンド信号処理部 2 1 7 は、T L P と同等の packets 長に変換されたインバンド信号用パ

ケットから信号とレジスタアドレスとを復元し、これをレジスタ216に書き込む。レジスタ216は、その内容が書き換えられると、シグナリング部214に通知する。シグナリング部214に通知する契機は、レジスタ216の内容の書き換えを検出したときでもよく、或いはレジスタ216の書き換えを示すフラグが立っていることを検出したときでもよい。

[0021] 本実施例によれば、従来はPCIエクスプレススイッチ内で処理されていた特定の制御信号に関する処理を、市販のマザーボードやI/Oデバイスをそのまま用いてネットワーク40越しに実行できるように、特定の制御信号をTLPに載せて上流側ブリッジ21と下流側ブリッジ31との間でネットワーク40を介して矛盾なく送受信するための信号変換及び接続を行うように構成されているため、分散型計算機システム10の動作をより安定的なものとし、かつ複雑な動作を可能にできる。

## 実施例 2

[0022] 図2は実施例2に係わる分散型計算機システム70の構成を示す機能ブロック図である。分散型計算機システム70は、実施例1に係わる分散型計算機システム10の機能を更に拡張し、PCIエクスプレス規格のサイドバンド信号50を上流側ブリッジ21と下流側ブリッジ31との間でインバンド方式により送受信する機構を備える。サイドバンド信号50は、例えば電源管理やカードのハードウェア状態の管理に関する信号を含み、具体的には、スロットにカードが存在するか否かを示すPresence、ローカルコンピュータの管理データをやりとりするSMBus、SMCIk、JTAG信号であるTCK、TDI、TDO、TMS、電源ON要求のためのWAKE#、カードの電源が入っていることを示すPWRGD等がある。

[0023] なお、サイドバンド信号50を上流側ブリッジ21と下流側ブリッジ31との間でインバンド方式により送受信するための処理の流れは、実施例1で説明した特定の制御信号を上流側ブリッジ21と下流側ブリッジ31との間でインバンド方式により送受信するための処理の流れと同様であるため、その詳細な説明を省略する。

- [0024] リモート I/O デバイス 30 は、イーサネットの NIC の特殊機能である、ネットワーク越しにコンピュータの電源をオン/オフ制御するための WoL (Wake On LAN) 管理機能 33 と、PCI エクスプレス規格の一般的な電源管理機能 35 とを有している。シグナリング部 314 は、電源ボタン 34 のオン/オフ入力を検出すると、電源オン/オフを要求するサイドバンド信号 50 をレジスタ 316 に書き込む。このサイドバンド信号 50 は、下流側ブリッジ 31 からネットワーク 40 を介して上流側ブリッジ 21 に送信され、レジスタ 216 に書き込まれる。シグナリング部 214 は、サイドバンド信号 50 を管理バス 24 に出力し、PCI エクスプレススロットの WAKE # へ接続することにより、ローカルコンピュータ 20 の電源をオン/オフする。これにより、リモート I/O デバイス 30 からネットワーク 40 越しにローカルコンピュータ 20 の電源をオン/オフ制御することができる。
- [0025] リモート I/O デバイス 30 は、リセットボタン 36 を有している。シグナリング部 314 は、リセットボタン 36 の入力を検出すると、リセットを要求するサイドバンド信号 50 をレジスタ 316 に書き込む。このサイドバンド信号 50 は、下流側ブリッジ 31 からネットワーク 40 を介して上流側ブリッジ 21 に送信され、レジスタ 216 に書き込まれる。シグナリング部 214 は、サイドバンド信号 50 を管理バス 24 に出力し、PCI エクスプレススロットの PERST # へ接続することにより、ローカルコンピュータ 20 をリセットする。これにより、リモート I/O デバイス 30 からネットワーク 40 越しにローカルコンピュータ 20 をリセットすることができる。
- [0026] リモート I/O デバイス 30 は、JTAG 信号入力用のテストピン (又はコネクタ) 37 を有しており、テストピン 37 に入力された JTAG 信号を下流側ブリッジ 31 及び上流側ブリッジ 21 を介してインバンドでローカルコンピュータ 20 の管理バス 24 に送信することも可能である。
- [0027] 上述の説明は、上流側ブリッジ 21 と下流側ブリッジ 31 との間でサイドバンド信号 50 をインバンドで送受信する一例であり、本発明は上述の例に限られるものではなく本発明の範囲及び趣旨から逸脱しない様々な変更を含

む。例えば、ローカルコンピュータ 20 は、P C I エクスプレス規格のビーコンを検出すると、P M E (Power Management Event) を発行し、リモート I / O デバイス 30 の電源オンを要求する信号を上流側ブリッジ 21 及び下流側ブリッジ 31 を介してインバンドでリモート I / O デバイス 30 に送信してもよい。

[0028] なお、実施例 2 に係わる分散型計算機システム 70 は、実施例 1 に係わる分散型計算機システム 70 と同等の構成を備えているため、従来の P C I エクスプレススイッチと等価な機能を有するとともに、実施例 1 で説明した特定の制御信号をインバンド方式で送受信する機能をも有する。これらの機能は、実施例 1 と同様であるため、詳細な説明を省略する。

[0029] 本実施例によれば、ローカルコンピュータ 20 とリモート I / O デバイス 30 との間でインバンド方式により信号を送受信することにより、ネットワーク 40 越しに電源管理、リセット、J T A G 試験等の各種管理を行うことが可能であり、運用上の利便性を高めることができる。また、ネットワーク 40 越しに電源管理を行うことにより、不必要な電源投入を抑制し、低消費電力化を実現できる。

[0030] この出願は、2011年9月27日に提出された日本出願特願 2011-210483 を基礎とする優先権を主張し、その開示の全てをここに取り込む。

[0031] 以上、実施形態を参照して本願発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明の範囲内で当業者が理解し得る様々な変更をすることができる。

[0032] なお、上述の実施例 1, 2 の一部又は全部は、以下の付記のように記載され得るが、以下には限定されない。

(付記 1)

ローカルコンピュータに接続する上流側ブリッジと、リモート I / O デバイスに接続する下流側ブリッジとを備える分散型計算機システムであって、ローカルコンピュータ及びリモート I / O デバイスは、上流側ブリッジ及

び下流側ブリッジを介してネットワーク上で接続しており、

上流側ブリッジ及び下流側ブリッジは、ネットワークに流れる主たるデータストリーム以外の信号を双方から参照し又は書き込むためのレジスタと、レジスタへの参照時又は書き込み時に主たるデータストリーム以外の信号をインバンドで送受信する機構とを備える、分散型計算機システム。

(付記2)

付記1に記載の分散型計算機システムであって、

ネットワークに流れる主たるデータストリーム以外の信号は、PCIエクスプレス規格のサイドバンド信号である、分散型計算機システム。

(付記3)

付記2に記載の分散型計算機システムであって、

リモートI/Oデバイスは、電源ボタンを有しており、電源ボタンのオン/オフ入力を検出すると、電源オン/オフを要求するサイドバンド信号を下流側ブリッジ及び上流側ブリッジを介してインバンドでローカルコンピュータに送信し、

ローカルコンピュータは、電源オン/オフを要求するサイドバンド信号を受信すると、自機の電源をオン/オフする、分散型計算機システム。

(付記4)

付記2に記載の分散型計算機システムであって、

リモートI/Oデバイスは、リセットボタンを有しており、リセットボタンの入力を検出すると、リセットを要求するサイドバンド信号を下流側ブリッジ及び上流側ブリッジを介してインバンドでローカルコンピュータに送信し、

ローカルコンピュータは、リセットを要求するサイドバンド信号を受信すると、自機をリセットする、分散型計算機システム。

(付記5)

付記2に記載の分散型計算機システムであって、

リモートI/OデバイスはJTAG信号入力用のテストピンを有しており

、テストピンに入力されたJTAG信号を下流側ブリッジ及び上流側ブリッジを介してインバンドでローカルコンピュータに送信する、分散型計算機システム。

(付記6)

ローカルコンピュータに接続する上流側ブリッジと、リモートI/Oデバイスに接続する下流側ブリッジとを備える分散型計算機システムであって、

ローカルコンピュータ及びリモートI/Oデバイスは、上流側ブリッジ及び下流側ブリッジを介してネットワーク上で接続しており、

上流側ブリッジ及び下流側ブリッジは、ネットワークに流れる主たるデータストリーム以外の信号をインバンドで送受信する機構を備えており、

ローカルコンピュータは、PCIエクスプレス規格のビーコンを検出すると、PMEを発行し、リモートI/Oデバイスの電源オンを要求する信号を上流側ブリッジ及び下流側ブリッジを介してインバンドでリモートI/Oデバイスに送信する、分散型計算機システム。

## 符号の説明

- [0033] 10…分散型計算機システム  
20…ローカルコンピュータ  
21…上流側ブリッジ  
22…ルートコンプレックス  
23…CPU及びメモリ  
24…各種機構  
30…リモートI/Oデバイス  
31…下流側ブリッジ  
32…I/Oデバイス  
33…WOL管理機能  
34…電源ボタン  
35…電源管理機能  
36…リセットボタン



- 37…テストピン
- 50…サイドバンド信号
- 60…PCIエクスプレススイッチ
- 61…上流側ブリッジ
- 62…上流側データバッファ
- 63…上流側レジスタ
- 64…下流側ブリッジ
- 65…下流側データバッファ
- 66…下流側レジスタ
- 67…共通レジスタ
- 70…分散型計算機システム
- 211…PCIエクスプレスパケット処理部
- 212…TLP抽出部
- 213…ネットワークパケット処理部
- 214…シグナリング部
- 215…内部ロジック
- 216…レジスタ
- 217…インバンド信号処理部
- 311…PCIエクスプレスパケット処理部
- 312…TLP抽出部
- 313…ネットワークパケット処理部
- 314…シグナリング部
- 315…内部ロジック
- 316…レジスタ
- 317…インバンド信号処理部

## 請求の範囲

- [請求項1]           ローカルコンピュータに接続する上流側ブリッジと、リモート I/O デバイスに接続する下流側ブリッジとを備える分散型計算機システムであって、
- 前記ローカルコンピュータ及び前記リモート I/O デバイスは、前記上流側ブリッジ及び前記下流側ブリッジを介してネットワーク上で接続しており、
- 前記上流側ブリッジ及び前記下流側ブリッジは、前記ネットワークに流れる主たるデータストリーム以外の信号を双方から参照し又は書き込むためのレジスタと、前記レジスタへの参照時又は書き込み時に前記主たるデータストリーム以外の信号をインバンドで送受信する機構とを備える、分散型計算機システム。
- [請求項2]           請求項 1 に記載の分散型計算機システムであって、
- 前記ネットワークに流れる主たるデータストリーム以外の信号は、P C I エクスプレス規格のサイドバンド信号である、分散型計算機システム。
- [請求項3]           請求項 2 に記載の分散型計算機システムであって、
- 前記リモート I/O デバイスは、電源ボタンを有しており、前記電源ボタンのオン/オフ入力を検出すると、電源オン/オフを要求するサイドバンド信号を前記下流側ブリッジ及び前記上流側ブリッジを介してインバンドで前記ローカルコンピュータに送信し、
- 前記ローカルコンピュータは、前記電源オン/オフを要求するサイドバンド信号を受信すると、自機の電源をオン/オフする、分散型計算機システム。
- [請求項4]           請求項 2 に記載の分散型計算機システムであって、
- 前記リモート I/O デバイスは、リセットボタンを有しており、前記リセットボタンの入力を検出すると、リセットを要求するサイドバンド信号を前記下流側ブリッジ及び前記上流側ブリッジを介してイン

バンドで前記ローカルコンピュータに送信し、

前記ローカルコンピュータは、前記リセットを要求するサイドバンド信号を受信すると、自機をリセットする、分散型計算機システム。

[請求項5]

請求項2に記載の分散型計算機システムであって、

前記リモート I/O デバイスは JTAG 信号入力用のテストピンを有しており、前記テストピンに入力された JTAG 信号を前記下流側ブリッジ及び前記上流側ブリッジを介してインバンドで前記ローカルコンピュータに送信する、分散型計算機システム。

[請求項6]

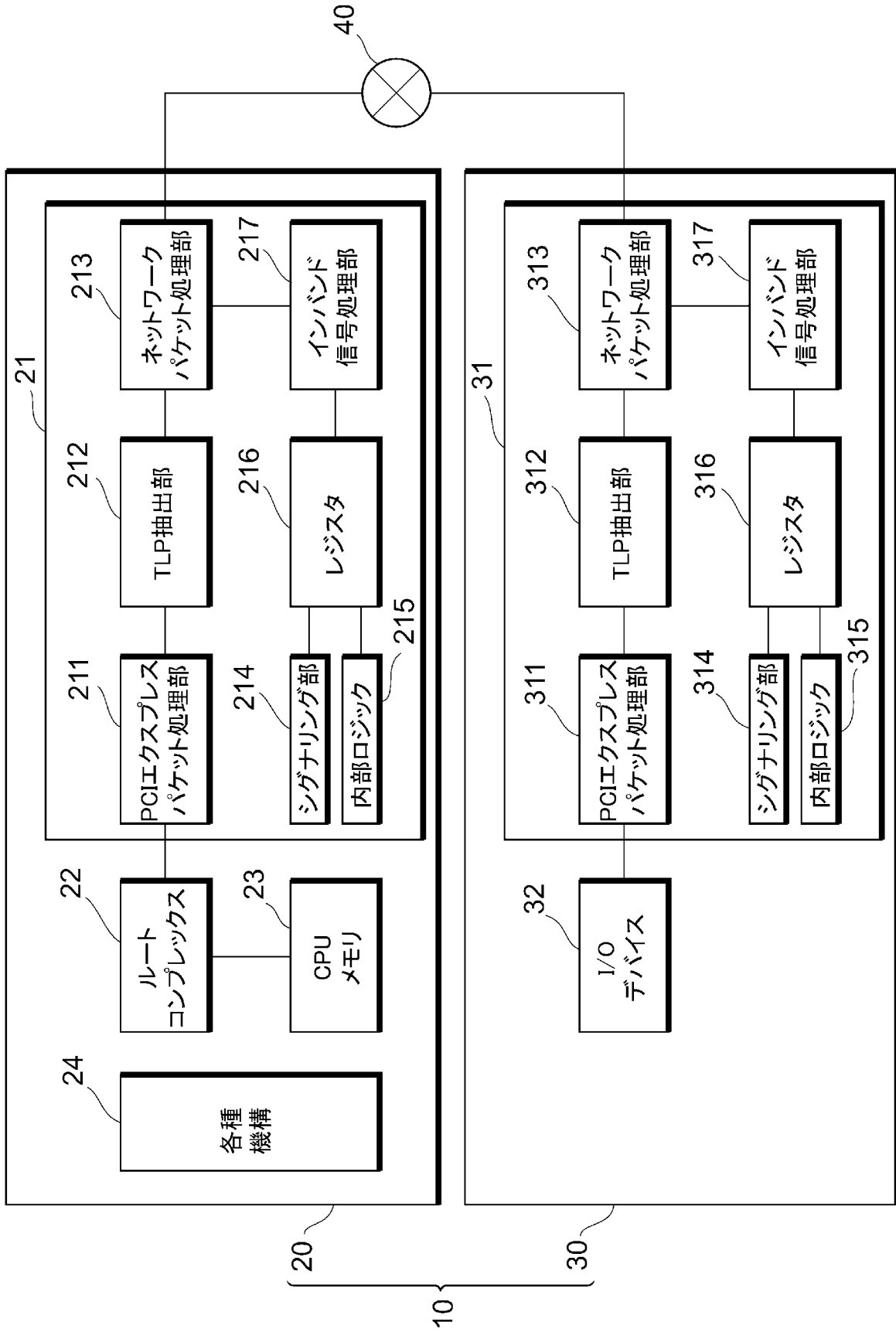
ローカルコンピュータに接続する上流側ブリッジと、リモート I/O デバイスに接続する下流側ブリッジとを備える分散型計算機システムであって、

前記ローカルコンピュータ及び前記リモート I/O デバイスは、前記上流側ブリッジ及び前記下流側ブリッジを介してネットワーク上で接続しており、

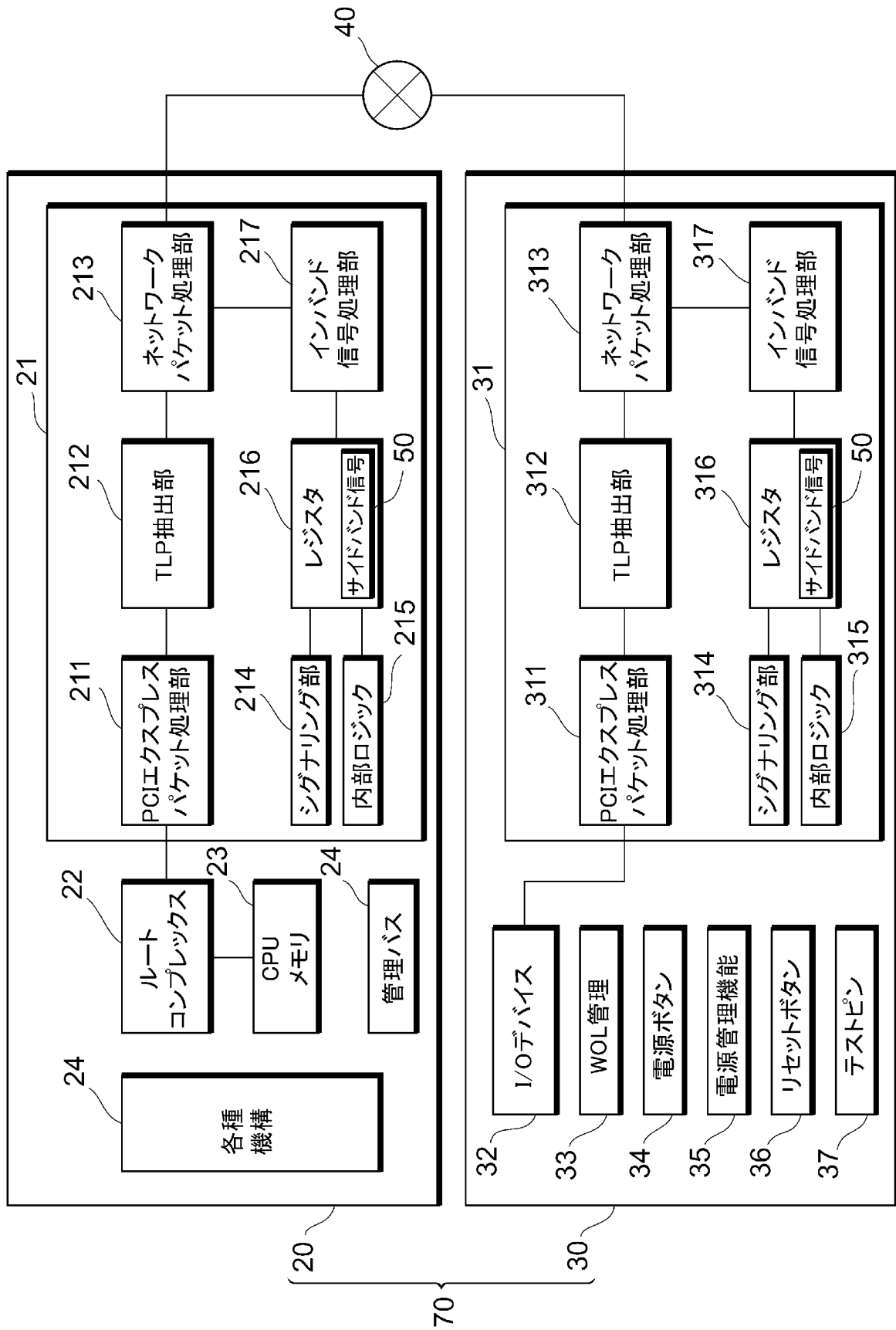
前記上流側ブリッジ及び前記下流側ブリッジは、前記ネットワークに流れる主たるデータストリーム以外の信号をインバンドで送受信する機構を備えており、

前記ローカルコンピュータは、PCI Express 規格のビーコンを検出すると、PME を発行し、前記リモート I/O デバイスの電源オンを要求する信号を前記上流側ブリッジ及び前記下流側ブリッジを介してインバンドで前記リモート I/O デバイスに送信する、分散型計算機システム。

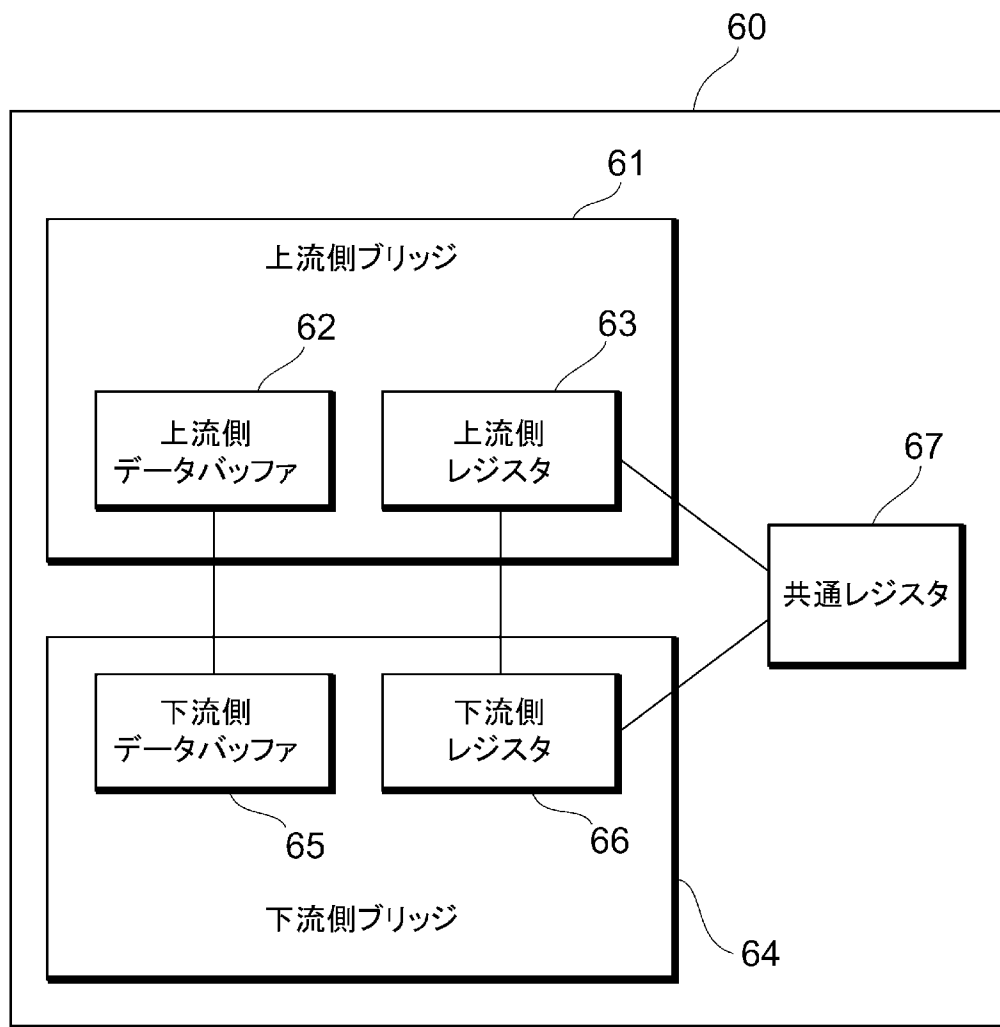
[図1]



[図2]



[図3]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/065108

## A. CLASSIFICATION OF SUBJECT MATTER

G06F13/36(2006.01)i, G06F1/26(2006.01)i, G06F13/38(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F13/36, G06F1/26, G06F13/38

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2012
Kokai Jitsuyo Shinan Koho	1971-2012	Toroku Jitsuyo Shinan Koho	1994-2012

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2008/053858 A2 (Graphin Co., Ltd.), 08 May 2008 (08.05.2008), paragraphs [0016] to [0022], [0025] to [0068]; fig. 1 to 12 (Family: none)	1-4, 6 5
Y	US 5764924 A (NCR Corp.), 09 June 1998 (09.06.1998), column 3, line 33 to column 6, line 67; fig. 2, 3 (Family: none)	5
A	WO 2011/090145 A1 (NEC Corp.), 28 July 2011 (28.07.2011), entire text; all drawings (Family: none)	1-6

 Further documents are listed in the continuation of Box C. See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search  
30 August, 2012 (30.08.12)Date of mailing of the international search report  
11 September, 2012 (11.09.12)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))  
 Int.Cl. G06F13/36(2006.01)i, G06F1/26(2006.01)i, G06F13/38(2006.01)i

B. 調査を行った分野  
 調査を行った最小限資料 (国際特許分類 (IPC))  
 Int.Cl. G06F13/36, G06F1/26, G06F13/38

最小限資料以外の資料で調査を行った分野に含まれるもの  
 日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2012年  
 日本国実用新案登録公報 1996-2012年  
 日本国登録実用新案公報 1994-2012年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y	WO 2008/053858 A2 (株式会社グラフィン) 2008.05.08, [0016]-[0022]、[0025]-[0068]、[図1]-[図12] (ファミリーなし)	1-4, 6 5
Y	US 5764924 A (NCR Corporation) 1998.06.09, 第3欄第33行目- 第6欄第67行目、第2図、第3図 (ファミリーなし)	5
A	WO 2011/090145 A1 (日本電気株式会社) 2011.07.28, 全文、全図 (ファミリーなし)	1-6

☐ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

\* 引用文献のカテゴリー  
 「A」特に関連のある文献ではなく、一般的技術水準を示すもの  
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」口頭による開示、使用、展示等に言及する文献  
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献  
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
 「&」同一パテントファミリー文献

国際調査を完了した日  
 30.08.2012

国際調査報告の発送日  
 11.09.2012

国際調査機関の名称及びあて先  
 日本国特許庁 (ISA/J P)  
 郵便番号100-8915  
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)  
 5 T | 9 8 5 7  
 木村 貴俊  
 電話番号 03-3581-1101 内線 3568