(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5561622号
(P5561622)

(45)発行日　平成26年7月30日(2014.7.30)　(24)登録日　平成26年6月20日(2014.6.20)

(51)Int.Cl.		F I		
G06F 11/16	(2006.01)	G06F	11/16	310C
G06F 11/30	(2006.01)	G06F	11/30	305A
G06F 1/26	(2006.01)	G06F	1/00	330C
G06F 1/28	(2006.01)	G06F	1/00	333C

請求項の数 11　(全 22 頁)

(21)出願番号　特願2011-210049(P2011-210049)
(22)出願日　平成23年9月27日(2011.9.27)
(65)公開番号　特開2013-73289(P2013-73289A)
(43)公開日　平成25年4月22日(2013.4.22)
審査請求日　平成25年2月6日(2013.2.6)

(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　100102864
弁理士　工藤　実
(72)発明者　馬場　紀圭
東京都港区芝五丁目７番１号　日本電気株式会社内

審査官　▲高▼橋　正▲徳▼

最終頁に続く

(54)【発明の名称】多重化システム、データ通信カード、状態異常検出方法、及びプログラム

(57)【特許請求の範囲】
【請求項１】
複数の物理マシンと、
前記複数の物理マシンの各々に搭載され、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うためのデータ通信カードと
を含み、
前記データ通信カードは、
メイン電源の給電により駆動し、前記自系の物理マシン及び前記他系の物理マシンの内部状態を監視し、状態異常を検出する第１のＬＳＩ
を具備し、
前記第１のＬＳＩは、
前記メイン電源がＯＮの状態で、監視及び検出対象の物理マシンのメモリに対してデータの読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出する
多重化システム。
【請求項２】
請求項１に記載の多重化システムであって、
前記データ通信カードは、
搭載された物理マシンの基板上にチップとして搭載され、前記基板上にそれぞれ配置されているＣＰＵとＩ／Ｏチップとの間に設けられており、

前記第１のＬＳＩは、
前記メイン電源がＯＮの状態で、監視及び検出対象の物理マシンの前記Ｉ／Ｏチップに対して設定情報の読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出する
多重化システム。
【請求項３】
請求項１または２に記載の多重化システムであって、
前記データ通信カードは、
スタンバイ電源の給電により駆動し、前記メイン電源がＯＦＦの状態の時でも、前記自系の物理マシン及び前記他系の物理マシンの電源状態を監視し、電源の異常を検出する第２のＬＳＩ
を更に具備する
多重化システム。
【請求項４】
請求項３に記載の多重化システムであって、
前記第１のＬＳＩは、
前記データ通信カード内部を制御するプロセッサと、
前記自系の物理マシン上で動作するソフトウェア（ＳＷ）の状態を取得するＳＷ状態取得部と、
ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、前記自系の物理マシン内部のハードウェア及びＩ／Ｏチップと接続し、前記ハードウェア及び前記Ｉ／Ｏチップの状態を監視するＰＣＩ制御部と、
前記第２のＬＳＩ側とデータの送受信を行い、通信回線を介して前記他系の物理マシン側とデータの送受信を行う通信制御部と
を更に具備し、
前記第２のＬＳＩは、
前記自系の物理マシン及び前記他系の物理マシンの電源状態を監視する電源監視部と、
前記自系の物理マシン及び前記他系の物理マシンの電源状態を制御する電源制御部と、
ＳＭＢｕｓを介して、前記自系の物理マシン内部のＢＭＣ（Ｂａｓｅｂｏａｒｄ　Ｍａｎａｇｅｍｅｎｔ　Ｃｏｎｔｒｏｌｌｅｒ）と接続し、前記ＢＭＣから監視結果を取得するＳＭＢｕｓ制御部と、
前記第１のＬＳＩ側とデータの送受信を行い、通信回線を介して前記他系の物理マシン側とデータの送受信を行い、前記自系の物理マシンから電源の給電を受けられない場合、前記他系の物理マシンから電源の給電を受ける通信制御部と
を更に具備する
多重化システム。
【請求項５】
請求項１に記載の多重化システムであって、
前記データ通信カードは、
前記自系の物理マシンの基板上にチップとして搭載され、前記自系の物理マシン内部の前記基板上にそれぞれ配置されているＣＰＵとＩ／Ｏチップとの間に設けられている
多重化システム。
【請求項６】
複数の物理マシンの各々に搭載されたデータ通信カードであって、
通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行う手段と、
メイン電源の給電により駆動し、前記自系の物理マシン及び前記他系の物理マシンの内部状態を監視し、状態異常を検出する第１のＬＳＩと
を具備し、
前記第１のＬＳＩは、

前記メイン電源がＯＮの状態で、監視及び検出対象の物理マシンのメモリに対してデータの読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出する
データ通信カード。
【請求項７】
請求項６に記載のデータ通信カードであって、
前記データ通信カードは、
搭載された物理マシンの基板上にチップとして搭載され、前記基板上にそれぞれ配置されているＣＰＵとＩ／Ｏチップとの間に設けられており、
前記第１のＬＳＩは、
前記メイン電源がＯＮの状態で、前記監視及び検出対象の物理マシンの前記Ｉ／Ｏチップに対して設定情報の読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出する
データ通信カード。
【請求項８】
請求項６または７に記載のデータ通信カードであって、
スタンバイ電源の給電により駆動し、前記メイン電源がＯＦＦの状態の時でも、前記自系の物理マシン及び前記他系の物理マシンの電源状態を監視し、電源の異常を検出する第２のＬＳＩと
を更に具備する
データ通信カード。
【請求項９】
請求項８に記載のデータ通信カードであって、
前記第１のＬＳＩは、
前記データ通信カード内部を制御するプロセッサと、
前記自系の物理マシン上で動作するソフトウェア（ＳＷ）の状態を取得するＳＷ状態取得部と、
ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、前記自系の物理マシン内部のハードウェア及びＩ／Ｏチップと接続し、前記ハードウェア及び前記Ｉ／Ｏチップの状態を監視するＰＣＩ制御部と、
前記第２のＬＳＩ側とデータの送受信を行い、通信回線を介して前記他系の物理マシン側とデータの送受信を行う通信制御部と
を更に具備し、
前記第２のＬＳＩは、
前記自系の物理マシン及び前記他系の物理マシンの電源状態を監視する電源監視部と、
前記自系の物理マシン及び前記他系の物理マシンの電源状態を制御する電源制御部と、
ＳＭＢｕｓを介して、前記自系の物理マシン内部のＢＭＣ（Ｂａｓｅｂｏａｒｄ　Ｍａｎａｇｅｍｅｎｔ　Ｃｏｎｔｒｏｌｌｅｒ）と接続し、前記ＢＭＣから監視結果を取得するＳＭＢｕｓ制御部と、
前記第１のＬＳＩ側とデータの送受信を行い、通信回線を介して前記他系の物理マシン側とデータの送受信を行い、前記自系の物理マシンから電源の給電を受けられない場合、前記他系の物理マシンから電源の給電を受ける通信制御部と
を更に具備する
データ通信カード。
【請求項１０】
複数の物理マシンの各々に搭載されたデータ通信カードにより実施される状態異常検出方法であって、
前記データ通信カードは、
通信手段と、
メイン電源の給電により駆動する第１のＬＳＩと

を具備しており、
前記状態異常検出方法は、
前記通信手段が、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うステップと、
前記第１のＬＳＩが、前記自系の物理マシン及び前記他系の物理マシンの内部状態を監視し、状態異常を検出するステップと
を含み、
前記状態異常を検出するステップでは、
前記第１のＬＳＩが、前記メイン電源がＯＮの状態で、監視及び検出対象の物理マシンのメモリに対してデータの読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出する
状態異常検出方法。
【請求項１１】
複数の物理マシンの各々に搭載されたデータ通信カードにより実行されるプログラムであって、
前記データ通信カードは、
通信手段と、
メイン電源の給電により駆動する第１のＬＳＩと
を具備しており、
前記プログラムは、
前記通信手段が、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うステップと、
前記第１のＬＳＩが、前記自系の物理マシン及び前記他系の物理マシンの内部状態を監視し、状態異常を検出するステップと
を含み
前記状態異常を検出するステップでは、
前記第１のＬＳＩが、前記メイン電源がＯＮの状態で、監視及び検出対象の物理マシンのメモリに対してデータの読み出し要求を定期的に発行し、前記読み出し要求に対する応答を確認することで、前記監視及び検出対象の物理マシンの状態異常を検出すること
をデータ通信カードに実行させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、多重化システムに関し、特にフォールトトレラントシステムやクラスターシステムを実現する多重化システム、データ通信カード、状態異常検出方法、及びプログラムに関する。
【背景技術】
【０００２】
現在、物理マシン上で複数のＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）を動作させることを可能とした仮想化技術が広く用いられている。仮想化技術によって、利用率の低い複数の仮想マシンを、１つの物理マシン上にまとめて、物理マシン１台あたりの利用効率を高めるとともに、物理マシンの台数を減らして消費電力を抑えることが可能となる。
【０００３】
仮想化技術には、物理マシン上で稼働するホストＯＳ上に仮想マシンを動作させる層を設け、その層の上でゲストＯＳを動作させる方式や、ホストＯＳを介さず、ハードウェア（ＨＷ）上に仮想マシンを動作させるハイパーバイザ（ｈｙｐｅｒｖｉｓｏｒ）を設け、そのハイパーバイザ上でゲストＯＳを動作させる方式等がある。
【０００４】
また、現在、耐障害性の高いコンピュータシステムとして、フォールトトレラントシステム（Ｆａｕｌｔ　Ｔｏｌｅｒａｎｔ　Ｓｙｓｔｅｍ）が広く用いられている。

【０００５】
例えば、専用のハードウェア（ＨＷ）を介してロックステップ方式の動作を行い、多重化（多くは二重化）された主要なハードウェア（ＨＷ）を障害発生時に遅滞なく切り替えて動作し続けるハードウェア（ＨＷ）方式のフォールトトレラントシステムが、従来から知られている。
【０００６】
また、仮想マシンが動作している物理マシンでハードウェア（ＨＷ）上の故障等による障害が発生した場合に、その仮想マシンが行っている処理を、他の物理マシン上で待機している仮想マシンが継続して実行するソフトウェア（ＳＷ）方式のフォールトトレラントシステムが、近年研究されている。
【０００７】
ハードウェア（ＨＷ）方式のフォールトトレラントシステムでは、高価な専用のハードウェア（ＨＷ）を１システム毎に多重化する必要があるため、システムコストがかさむことになる。これに対して、ソフトウェア（ＳＷ）方式のフォールトトレラントシステムでは、汎用のハードウェア（ＨＷ）上で動作する複数の仮想マシンを用いて擬似的に多重化するため、専用のハードウェア（ＨＷ）が不要となり、システム毎にハードウェア（ＨＷ）を多重化する必要もないため、安価にシステムを構築・維持でき、システムコストを抑えることができる。
【０００８】
ソフトウェア（ＳＷ）方式のフォールトトレラントシステムでの処理の主体の切り替え方法の例として、ハードウェア（ＨＷ）方式のフォールトトレラントシステムで行われているロックステップ方式の動作や、メモリコピー方式の動作を、ソフトウェア（ＳＷ）により行い、障害発生時には瞬時に、処理の主体となる仮想マシンを切り替えるという方法がある。
【０００９】
しかし、一般的なＩＡサーバでソフトウェア（ＳＷ）方式のフォールトトレラントシステムやクラスターシステムを構築した場合、装置の故障検出を、一般的なハードウェア（ＨＷ）のエラー検出機構や専用のエラー検出ソフトウェア（ＳＷ）に依存することになる。
【００１０】
一般的なハードウェア（ＨＷ）では、機能を動作させた場合のみ故障を検出することができる。しかし、待機系となり動作していない場合には、故障を検出することができない。また、検出した故障を、ソフトウェア（ＳＷ）を介在せずに他系装置に通知する仕組みを持っていない。
【００１１】
また、専用のエラー検出ソフトウェア（ＳＷ）では、ソフトウェア（ＳＷ）によって動作させることができる機能については定期的なヘルスチェックを行うことで、早期に故障を検出したり、検出した装置を他系装置に通知したりできる。しかし、ソフトウェア（ＳＷ）が稼働するまでは、その機能を使用することができない。また、一般的に、ソフトウェア（ＳＷ）はタイムアウトによって故障を検出するため、他系装置に通知するには一定の時間を要する。したがって、フェイルオーバ（ｆａｉｌｏｖｅｒ）時間を短くすることが困難である。
【００１２】
関連する技術として、特許文献１（特許第４４６８４２６号公報）に高可用システム及び実行状態制御方法が開示されている。この関連技術では、第１の仮想計算機を管理する第１のハイパーバイザが備える収集部が、第1の仮想計算機について発生した、第1の仮想計算機に対する入力を伴うイベントに関する同期情報を収集する。また、第２の仮想計算機を管理する第２のハイパーバイザが備える制御部が、この同期情報に従って、第２の仮想計算機の入力に係る実行状態を、第１の仮想計算機の入力に係る実行状態と同一になるように制御する。これにより、独立した２台の計算機上でそれぞれ稼働する仮想計算機を

組み合わせて二重化を実現する。

【００１３】

また、特許文献２（特開２００９－０８０６９２号公報）に仮想計算機システム及び同システムにおけるサービス引き継ぎ制御方法が開示されている。この関連技術では、仮想マシンが動作しているサーバ計算機に障害が発生した場合、サーバ計算機の仮想マシンモニタは、障害発生時刻に最も近い時点でディスク装置に採取されたスナップショットに基づき、仮想マシンを仮想マシンとしてサーバ計算機上に再生成する。通信記録ユニットの状態再現部は、仮想マシンに対応付けられた通信履歴に基づき、スナップショットの採取時期から上記障害発生時刻までの期間における仮想マシンの状態を仮想マシンに再現させる。再起動部は、例えば仮想マシンの状態の再現に失敗した場合、仮想マシンをサーバ計算機上で再起動する。これにより、仮想マシンが動作している物理計算機に障害が発生した場合、別の物理計算機上で再生成または再起動される仮想マシンによりサービスを継続させる。

【００１４】

また、特許文献３（特開２００８－０３３４８３号公報）に計算機システム、計算機および計算機動作環境の移動方法が開示されている。この関連技術では、第１計算機の動作を中断する。次に、第１ディスク上のコピーイメージに含まれるファイルのリストを作成する。次に、第１計算機の実行コンテキストを、第２計算機にコピーする。次に、第２計算機において上記動作を再開させる。次に、上記リストを参照して、コピーイメージを第１ディスクから第２ディスクにコピーする。これにより、第１ディスクを使用する第１計算機の動作環境を、第２ディスクを使用する第２計算機に移動させる際に、業務の中断時間を短縮する。

【先行技術文献】

【特許文献】

【００１５】

【特許文献１】特許第４４６８４２６号公報

【特許文献２】特開２００９－０８０６９２号公報

【特許文献３】特開２００８－０３３４８３号公報

【発明の概要】

【発明が解決しようとする課題】

【００１６】

従来の多重化システムでは、汎用装置を用いて故障検出を行っているが、汎用装置の故障検出能力は低い。例えば、故障検出機能を動作させた時にしか故障を検出できないため、待機系等で装置や故障検出機能が停止している場合は検出できない。また、多重化用ＯＳの介在なく、外部に故障を通知することができないため、多重化した各装置のＯＳが起動していなければならない。更に、多重化用ＯＳの故障検出機能も、多重化用ＯＳが動作を開始するまでは機能しない。一般的に、ソフトウェア（ＳＷ）による故障検出では、タイムアウトで故障検出するため、検出まで時間がかかる。

【００１７】

本発明の目的は、一般的なＩＡサーバでソフトウェア（ＳＷ）方式のフォールトトレラントシステムやクラスターシステムを構築するために利用されるデータ通信カードに、自系装置のメモリやチップ等の状態監視機能と、他系装置への状態通知機能及び電源制御機能を追加した多重化システムを提供することである。

【課題を解決するための手段】

【００１８】

本発明に係る多重化システムは、複数の物理マシンと、複数の物理マシンの各々に搭載され、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うためのデータ通信カードとを含む。データ通信カードは、自律的に、自系の物理マシン及び他系の物理マシンの状態を監視し、状態異常を検出する。

【００１９】

本発明に係るデータ通信カードは、複数の物理マシンの各々に搭載されたデータ通信カードであって、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行う手段と、自律的に、自系の物理マシン及び他系の物理マシンの状態を監視し、状態異常を検出する手段とを具備する。

【００２０】

本発明に係る状態異常検出方法は、複数の物理マシンの各々に搭載されたデータ通信カードにより実施される状態異常検出方法であって、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うことと、自律的に、自系の物理マシン及び他系の物理マシンの状態を監視し、状態異常を検出することとを含む。

【００２１】

本発明に係るプログラムは、複数の物理マシンの各々に搭載されたデータ通信カードにより実行されるプログラムであって、通信回線を介して自系の物理マシンと他系の物理マシンとを接続し、相互にデータの送受信を行うステップと、自律的に、自系の物理マシン及び他系の物理マシンの状態を監視し、状態異常を検出するステップとをデータ通信カードに実行させるためのプログラムである。なお、本発明に係るプログラムは、記憶装置や記憶媒体に格納することが可能である。

【発明の効果】

【００２２】

データ通信カードが、物理マシンの状態に関わらず、定期的に物理マシン内部の状態を常に監視するため、早期のエラー検出が可能となる。また、多重化システムにおけるスタンバイ系（待機系、従系）装置等、普段は動作していない装置のエラー検出も可能となる。

【図面の簡単な説明】

【００２３】

【図１】本発明に係る多重化システムの基本構成を示す概念図である。
【図２】本発明に係る物理マシンの構成例を示す概念図である。
【図３】本発明に係るデータ通信カードの構成例を示す概念図である。
【図４】本発明に係るデータ通信カード内部の詳細を示すブロック図である。
【図５】本発明に係る多重化システムのシステム構成１を示す概念図である。
【図６】本発明に係る多重化システムのシステム構成２を示す概念図である。

【発明を実施するための形態】

【００２４】

＜実施形態＞

以下に、本発明の実施形態について添付図面を参照して説明する。

【００２５】

［基本構成］

図１に示すように、本発明に係る多重化システムは、複数の物理マシン１００（１００－ｉ，ｉ＝１～ｎ：ｎは台数）を含む。

【００２６】

ここでは、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の各々の例として、ＰＣ（パソコン）、アプライアンス（ａｐｐｌｉａｎｃｅ）、シンクライアントサーバ、ワークステーション、メインフレーム、スーパーコンピュータ等の計算機を想定している。なお、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）は、端末やサーバに限らず、中継機器でも良い。中継装置の例として、ネットワークスイッチ（ｎｅｔｗｏｒｋ　ｓｗｉｔｃｈ）、ルータ（ｒｏｕｔｅｒ）、プロキシ（ｐｒｏｘｙ）、ゲートウェイ（ｇａｔｅｗａｙ）、ファイアウォール（ｆｉｒｅｗａｌｌ）、ロードバランサ（ｌｏａｄ　ｂａｌａｎｃｅｒ：負荷分散装置）、帯域制御装置（ｐａｃｋｅｔ　ｓｈａｐｅｒ）、セキュリティ監視制御装置（ＳＣＡＤＡ：Ｓｕｐｅｒｖｉｓｏｒｙ　Ｃｏｎｔｒｏｌ　Ａｎｄ　Ｄａｔａ　Ａｃｑｕｉｓｉｔｉｏｎ）、ゲートキーパー（ｇａｔｅｋｅｅｐｅｒ）、基地局（ｂａｓｅ　ｓｔａｔｉｏｎ）、アクセスポイント（ＡＰ：Ａｃｃｅｓｓ　Ｐｏｉｎｔ）、或いは、

複数の通信ポートを有する計算機等が考えられる。

【００２７】

図示しないが、上記のような計算機や中継機器等は、プログラムに基づいて駆動し所定の処理を実行するプロセッサと、当該プログラムや各種データを記憶するメモリと、ネットワークインターフェースによって実現される。

【００２８】

上記のプロセッサの例として、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）、ネットワークプロセッサ（ＮＰ：Ｎｅｔｗｏｒｋ Ｐｒｏｃｅｓｓｏｒ）、マイクロプロセッサ（ｍｉｃｒｏｐｒｏｃｅｓｓｏｒ）、マイクロコントローラ、或いは、専用の機能を有する半導体集積回路（ＩＣ：Ｉｎｔｅｇｒａｔｅｄ Ｃｉｒｃｕｉｔ）等が考えられる。

【００２９】

上記のメモリの例として、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）、ＥＥＰＲＯＭ（Ｅｌｅｃｔｒｉｃａｌｌｙ Ｅｒａｓａｂｌｅ ａｎｄ Ｐｒｏｇｒａｍｍａｂｌｅ Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）やフラッシュメモリ等の半導体記憶装置等を想定している。ＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）やＳＳＤ（Ｓｏｌｉｄ Ｓｔａｔｅ Ｄｒｉｖｅ）等の補助記憶装置、又は、ＤＶＤ（Ｄｉｇｉｔａｌ Ｖｅｒｓａｔｉｌｅ Ｄｉｓｋ）等のリムーバブルディスクや、ＳＤメモリカード（Ｓｅｃｕｒｅ Ｄｉｇｉｔａｌ ｍｅｍｏｒｙ ｃａｒｄ）等の記憶媒体（メディア）等が考えられる。また、バッファ（ｂｕｆｆｅｒ）やレジスタ（ｒｅｇｉｓｔｅｒ）でも良い。或いは、ＤＡＳ（Ｄｉｒｅｃｔ Ａｔｔａｃｈｅｄ Ｓｔｏｒａｇｅ）、ＦＣ－ＳＡＮ（Ｆｉｂｒｅ Ｃｈａｎｎｅｌ － Ｓｔｏｒａｇｅ Ａｒｅａ Ｎｅｔｗｏｒｋ）、ＮＡＳ（Ｎｅｔｗｏｒｋ Ａｔｔａｃｈｅｄ Ｓｔｏｒａｇｅ）、ＩＰ－ＳＡＮ（ＩＰ － Ｓｔｏｒａｇｅ Ａｒｅａ Ｎｅｔｗｏｒｋ）等を用いたストレージ装置でも良い。

【００３０】

なお、上記のプロセッサ及び上記のメモリは、一体化していても良い。例えば、近年では、マイコン等の１チップ化が進んでいる。したがって、電子機器等に搭載される１チップマイコンが、上記のプロセッサ及び上記のメモリを備えている事例が考えられる。

【００３１】

上記のネットワークインターフェースの例として、ネットワーク通信に対応した基板（マザーボード、Ｉ／Ｏボード）やチップ等の半導体集積回路、ＮＩＣ（Ｎｅｔｗｏｒｋ Ｉｎｔｅｒｆａｃｅ Ｃａｒｄ）等のネットワークアダプタや同様の拡張カード、アンテナ等の通信装置、通信ポート等が考えられる。

【００３２】

また、ネットワークの例として、インターネット、ＬＡＮ（Ｌｏｃａｌ Ａｒｅａ Ｎｅｔｗｏｒｋ）、無線ＬＡＮ（Ｗｉｒｅｌｅｓｓ ＬＡＮ）、ＷＡＮ（Ｗｉｄｅ Ａｒｅａ Ｎｅｔｗｏｒｋ）、バックボーン（Ｂａｃｋｂｏｎｅ）、ケーブルテレビ（ＣＡＴＶ）回線、固定電話網、携帯電話網、ＷｉＭＡＸ（ＩＥＥＥ ８０２．１６ａ）、３Ｇ（３ｒｄ Ｇｅｎｅｒａｔｉｏｎ）、専用線（ｌｅａｓｅ ｌｉｎｅ）、ＩｒＤＡ（Ｉｎｆｒａｒｅｄ Ｄａｔａ Ａｓｓｏｃｉａｔｉｏｎ）、Ｂｌｕｅｔｏｏｔｈ（登録商標）、シリアル通信回線、データバス等が考えられる。

【００３３】

但し、実際には、これらの例に限定されない。

【００３４】

本発明では、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の各々には、データ通信カード１０が搭載されている。搭載の方法として、挿入、内蔵、組込、接続等が考えられる。但し、実際には、これらの例に限定されない。

【００３５】

データ通信カード１０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）に搭載可能な

高機能な拡張カードである。データ通信カード１０は、上記のネットワークインターフェースでも良い。データ通信カード１０は、ＬＳＩ（Ｌａｒｇｅ　Ｓｃａｌｅ　Ｉｎｔｅｇｒａｔｉｏｎ）を搭載している。データ通信カード１０は、他のデータ通信カードと通信回線を介して接続し、互いにデータの送受信を行う。また、データ通信カード１０は、障害検出機能を持ち、自身が搭載された物理マシン内部の障害や、接続相手のデータ通信カードが搭載された他の物理マシン内部の障害を検出する。他の物理マシン内部の障害については、接続相手のデータ通信カードからの通知により認識しても良い。

【００３６】

なお、データ通信カード１０の形状は、カード型に限らない。例えば、データ通信カード１０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板と一体化していても良い。但し、実際には、これらの例に限定されない。

【００３７】

物理マシン１００（１００－ｉ，ｉ＝１～ｎ）上に、データ通信カード１０を使用できる環境が最低限整っていて、対応したソフトウェアが導入されている／導入可能である場合、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）にデータ通信カード１０を挿すだけで、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）は、ＦＴサーバ（Ｆａｕｌｔ　Ｔｏｌｅｒａｎｔ　Ｓｅｒｖｅｒ）やクラスターサーバになる。ソフトウェア（ＳＷ）方式のフォールトトレラントシステムやクラスターシステムを実現する場合、データ通信カード１０自体が、仮想マシン用の設定情報やイメージファイル等を保持し、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）に提供するようにしていても良い。このとき、データ通信カード１０は、「ＦＴカード」や「クラスターカード」と呼ぶこともできる。

【００３８】

［物理マシンの内部構成］

図２を参照して、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の各々の内部構成の詳細について説明する。

【００３９】

物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の各々は、データ通信カード１０と、ハードウェア（ＨＷ：ｈａｒｄｗａｒｅ）２０と、ソフトウェア（ＳＷ：ｓｏｆｔｗａｒｅ）３０と、ドライバ４０と、Ｉ／Ｏ（Ｉｎｐｕｔ／Ｏｕｔｐｕｔ）チップ５０と、ＢＭＣ（Ｂａｓｅｂｏａｒｄ　Ｍａｎａｇｅｍｅｎｔ　Ｃｏｎｔｒｏｌｌｅｒ）６０を備える。

【００４０】

ここでは、２台の物理マシン（物理マシン１００－１、物理マシン１００－２）を例に説明する。例えば、物理マシン１００－１は、データ通信カード１０－１と、ハードウェア（ＨＷ）２０－１と、ソフトウェア（ＳＷ）３０－１と、ドライバ４０－１と、ＢＭＣ６０－１と、Ｉ／Ｏチップ５０－１を備える。また、物理マシン１００－２は、データ通信カード１０－２と、ハードウェア（ＨＷ）２０－２と、ソフトウェア（ＳＷ）３０－２と、ドライバ４０－２と、ＢＭＣ６０－２と、Ｉ／Ｏチップ５０－２を備える。

【００４１】

データ通信カード１０については、上記の通りである。

【００４２】

ハードウェア（ＨＷ）２０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のハードウェア（ＨＷ）である。通常、ハードウェア（ＨＷ）２０は、ＤＣ電源の給電（電力供給）を受けて駆動する。ハードウェア（ＨＷ）２０の例として、プロセッサ、メモリ、補助記憶装置、ネットワークインターフェース、ＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ　Ｃｏｍｐｏｎｅｎｔｓ　Ｉｎｔｅｒｃｏｎｎｅｃｔ　ｂｕｓ）スロット、及び電源装置、又はこれらの組み合わせ等が考えられる。なお、ハードウェア（ＨＷ）２０は、同一の物理マシン内部で多重化されていても良い。

【００４３】

ソフトウェア（ＳＷ）３０は、ハードウェア（ＨＷ）２０を利用し、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）上で動作するソフトウェア（ＳＷ）である。ソフトウェア（

ＳＷ）３０の例として、ＯＳやアプリケーションソフトウェア、ミドルウェア（ｍｉｄｄｌｅｗａｒｅ）等が考えられる。なお、ソフトウェア（ＳＷ）３０は、物理マシン上に構築された仮想マシン（ＶＭ：Ｖｉｒｔｕａｌ　Ｍａｃｈｉｎｅ）でも良い。

【００４４】

ドライバ４０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の内部に装着された装置や、外部に接続した機器を制御・操作するためのソフトウェア（ＳＷ）／デバイスドライバである。ドライバ４０は、ＯＳが上記のような機器を制御するための橋渡しを行う。なお、ドライバ４０は、ＯＳに組み込まれ、ＯＳの機能の一部として振舞うようにしても良い。すなわち、ドライバ４０は、ソフトウェア（ＳＷ）３０の一部でも良い。ドライバ４０は、ソフトウェア（ＳＷ）３０がＩ／Ｏチップ５０に接続された外部のデバイスを利用する際、ＯＳが提供する共通化されたＡＰＩ（アプリケーション・プログラミング・インターフェース）によってデバイスの機能を利用できるようにして、抽象化されたＡＰＩとデバイスとの間の対応を受け持つ。

【００４５】

Ｉ／Ｏチップ５０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）に搭載された接続口（物理ポート）であり、物理マシン本体と各種周辺機器に接続して、それらの機器とデータをやり取りするための入出力インターフェースである。入出力インターフェースの主な規格として、キーボードやマウスを接続するＰＳ／２、モデムやプリンタなどと双方向で通信を行うシリアルインターフェースのＲＳ２３２Ｃ、ハードディスクドライブ（ＨＤＤ）等と双方向接続するパラレルインターフェースのＳＣＳＩ、主に内蔵型ＨＤＤ等と双方向接続するパラレルインターフェースのＩＤＥ、本体と周辺機器全般とを双方向接続するシリアルインターフェースであるＵＳＢ、次世代の高速なＳＣＳＩ規格であるＩＥＥＥ１３９４といった規格が知られている。なお、Ｉ／Ｏチップ５０は、スーパーＩ／Ｏ（Ｓｕｐｅｒ　Ｉｎｐｕｔ／Ｏｕｔｐｕｔ）チップや、Ｉ／Ｏコントローラ・ハブ（ＩＣＨ：Ｉ／Ｏ　Ｃｏｎｔｒｏｌｌｅｒ　Ｈｕｂ）でも良い。

【００４６】

ＢＭＣ６０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部に設けられたコントローラである。ＢＭＣ６０は、ハードウェア（ＨＷ）２０の状態を常時監視し、ハードウェア（ＨＷ）エラーの発生をＯＳ等に通知する。具体的には、ＢＭＣ６０は、電源ユニットからの供給電圧や冷却ファンの回転数、プロセッサを含む各種パーツの温度、ＳＣＳＩターミネータの電源電圧等を常時監視している。例え本体の電源がＯＦＦでも、コンセントからの電源コードが電源ユニットに接続されている限り、ＢＭＣ６０には電力が供給されるため、ＢＭＣ６０は、本体の電源がＯＦＦ状態でも、ハードウェア（ＨＷ）２０の状態監視を継続する。

【００４７】

［データ通信カードの詳細］

図３を参照して、データ通信カード１０の詳細について説明する。

【００４８】

ここでは、データ通信カード１０に搭載されたＬＳＩの例として、ＦＰＧＡ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｇａｔｅ　Ａｒｒａｙ）と、ＣＰＬＤ（Ｃｏｍｐｌｅｘ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｌｏｇｉｃ　Ｄｅｖｉｃｅ）を使用して説明する。なお、ＦＰＧＡやＣＰＬＤは一例に過ぎない。実際には、他のＬＳＩでも良い。

【００４９】

データ通信カード１０は、ＦＰＧＡ１１と、ＣＰＬＤ１２を備える。

【００５０】

例えば、データ通信カード１０－１は、ＦＰＧＡ１１－１と、ＣＰＬＤ１２－１を備える。また、データ通信カード１０－２は、ＦＰＧＡ１１－２と、ＣＰＬＤ１２－２を備える。

【００５１】

ＦＰＧＡ１１は、第１のＬＳＩである。ＦＰＧＡ１１は、ハードウェア（ＨＷ）２０、

ソフトウェア（ＳＷ）３０、及びＩ／Ｏチップ５０の状態監視を行う。ＦＰＧＡ１１には、メイン電源（データ通信カード１０が受電したＤＣ電源）が給電されている。

【００５２】

ＣＰＬＤ１２は、第２のＬＳＩである。ＣＰＬＤ１２は、ＢＭＣ６０及び電源装置の状態監視を行う。ＣＰＬＤ１２には、スタンバイ電源（データ通信カード１０が受電したＡＣ電源から作成されるＤＣ電源）が給電されている。なお、スタンバイ電源とは、電源管理を行うため、常時一定出力を供給するための出力である。スタンバイ電源を出力する回路は、メイン電源をＯＦＦにしても動作している。

【００５３】

なお、ＦＰＧＡ１１とＣＰＬＤ１２は、相互に通信可能である。

【００５４】

［ＦＰＧＡ及びＣＰＬＤの詳細］

図４を参照して、ＦＰＧＡ１１及びＣＰＬＤ１２の詳細について説明する。

【００５５】

ＦＰＧＡ１１は、プロセッサ１１１と、ＳＷ状態取得部１１２と、ＰＣＩ制御部１１３と、通信制御部１１４を備える。

【００５６】

プロセッサ１１１は、データ通信カード１０内部の各部の制御やデータの計算・加工（演算処理）を行う。例えば、プロセッサ１１１は、データ通信カード１０内部のＲＡＭ等に記憶されたプログラムに基づいて駆動し、所定の処理を実行する。プロセッサ１１１は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のハードウェア（ＨＷ）２０に含まれるＣＰＵ等のプロセッサではなく、データ通信カード１０内部のＦＰＧＡ１１上に設けられている。また、プロセッサ１１１は、ＳＷ状態取得部１１２、ＰＣＩ制御部１１３、及び通信制御部１１４の動作を変更することが可能である。

【００５７】

ＳＷ状態取得部１１２は、ソフトウェア（ＳＷ）３０の状態を取得する。ＳＷ状態取得部１１２は、ドライバ４０から直接通知を受け取ることで、ソフトウェア（ＳＷ）３０の状態を取得しても良いし、ＰＣＩ制御部１１３を介して、ドライバ４０がメモリ上に設定したソフトウェア（ＳＷ）３０の状態を取得しても良い。

【００５８】

ＰＣＩ制御部１１３は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のハードウェア（ＨＷ）２０及びＩ／Ｏチップ５０と接続し、ハードウェア（ＨＷ）２０及びＩ／Ｏチップ５０の状態を監視する。

【００５９】

通信制御部１１４は、ＦＰＧＡ１１とＣＰＬＤ１２とを接続している。したがって、通信制御部１１４は、ＦＰＧＡ１１によるエラー検出結果を、ＣＰＬＤ１２側に通知することが可能である。また、通信制御部１１４は、他のデータ通信カードに搭載されたＦＰＧＡと、少なくとも１本のケーブルを介してリモート（ｒｅｍｏｔｅ：遠隔）接続する。ここでは、２本のケーブルを介して接続することとする。ケーブルの本数が複数の場合、物理的／論理的に１本に束ねることも可能である。複数のケーブルを１本に束ねることで、束ねた本数に応じて通信速度を倍増することができる。また、複数のケーブルを１本に束ねることで、いずれかのケーブルが故障した場合でも、残りのケーブルを使用して通信を継続することが可能である。通信制御部１１４は、この２本のケーブルを介して、ＦＰＧＡ１１によるエラー検出結果を、他のデータ通信カードに通知する。

【００６０】

ＣＰＬＤ１２は、電源監視部１２１と、電源制御部１２２と、ＳＭＢｕｓ（Ｓｙｓｔｅｍ　Ｍａｎａｇｅｍｅｎｔ　Ｂｕｓ）制御部１２３と、通信制御部１２４を備える。

【００６１】

電源監視部１２１は、自機（Ｌｏｃａｌ：自系）の物理マシン及びリモート接続先（Ｒｅｍｏｔｅ：他系）の物理マシンの電源状態を監視する。このとき、電源監視部１２１は

、自機の物理マシンの電源状態を直接監視しても良いし、ＳＭＢｕｓ制御部１２３を介して、ＢＭＣ６０による監視結果を取得しても良い。なお、電源監視部１２１は、メイン電源の有無／変化により、自機の物理マシンの電源状態を監視しても良い。また、電源監視部１２１は、通信制御部１２４を介して、リモート接続先の物理マシンの電源状態を直接監視しても良いし、リモート接続先の物理マシンのＳＭＢｕｓ制御部を介して、リモート接続先の物理マシンのＢＭＣによる監視結果を取得しても良い。

【００６２】

電源制御部１２２は、自機の物理マシン及びリモート接続先の物理マシンの電源状態を制御する。このとき、電源制御部１２２は、自機の物理マシンの電源状態を直接制御しても良いし、ＳＭＢｕｓ制御部１２３及びＢＭＣ６０を介して制御しても良い。また、電源制御部１２２は、通信制御部１２４を介して、リモート接続先の物理マシンの電源状態を直接制御しても良いし、リモート接続先の物理マシンのＳＭＢｕｓ制御部及びＢＭＣを介して制御しても良い。

【００６３】

ＳＭＢｕｓ制御部１２３は、ＳＭＢｕｓを介して、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のＢＭＣ６０と接続し、ＢＭＣ６０から監視結果を取得する。また、ＳＭＢｕｓ制御部１２３は、ＢＭＣ６０に対して制御指示を行うことも可能である。

【００６４】

通信制御部１２４は、ＦＰＧＡ１１とＣＰＬＤ１２とを接続している。したがって、通信制御部１２４は、ＣＰＬＤ１２によるエラー検出結果を、ＦＰＧＡ１１側に通知することが可能である。なお、ＦＰＧＡ１１の通信制御部１１４とＣＰＬＤ１２の通信制御部１２４とは、同一の装置／回路でも良い。また、通信制御部１２４は、他のデータ通信カードに搭載されたＣＰＬＤと、少なくとも１本のケーブルを介してリモート接続する。ここでは、１本のケーブルを介して接続することとする。ケーブルの本数が複数の場合、物理的／論理的に１本に束ねることが可能である。通信制御部１２４は、この１本のケーブルを介して、ＣＰＬＤ１２によるエラー検出結果を、他のデータ通信カードに通知する。また、通信制御部１２４は、この１本のケーブルを介して、リモート接続先の物理マシンから電源の給電を受ける（受電する）ようにしても良い。例えば、通信制御部１２４は、この１本のケーブルを構成する回線の一部を、リモート接続先の物理マシンからの電源の給電用に利用する。この場合、ＣＰＬＤ１２は、自機の物理マシンから電源の給電を受けることができなくなっても、リモート接続先の物理マシンから電源の給電を受けることで、動作を継続することができる。

【００６５】

［ハードウェア（ＨＷ）状態監視］

以下に、ハードウェア（ＨＷ）状態監視の動作の詳細について説明する。

【００６６】

ここでは、ハードウェア（ＨＷ）状態監視の一例として、メモリ状態監視の動作について説明する。なお、メモリ状態監視が行われるのは、メイン電源がＯＮ状態（メイン電源とスタンバイ電源の両方がＯＮの状態）の時のみである。

【００６７】

ＰＣＩ制御部１１３は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、メモリの読み出し要求（リードリクエスト）を発行し、ＣＰＵ等を介してメモリのデータの読み出しを行い、ＣＰＵ等からコンプリーション（ｃｏｍｐｌｅｔｉｏｎ：完了）の応答が正常に返ってくるかをチェックすることで、メモリが正常に動作しているか監視する。

【００６８】

例えば、ＰＣＩ制御部１１３は、一定間隔で定期的に、全メモリアドレスに対して順番に読み出し要求（リードリクエスト）を発行する。

【００６９】

なお、ＰＣＩ制御部１１３の動作は、プロセッサ１１１により変更することが可能である。

【００７０】
これにより、ＰＣＩ制御部１１３は、メモリ故障を検出することが可能となる。
【００７１】
従来は、実際にメモリの読み出しが必要となり、ＣＰＵ等からメモリの読み出し要求（リードリクエスト）が発行されるまで、エラー検出が不可能であった。
【００７２】
本発明では、データ通信カード１０が、メモリの読み出しの有無に関わらず、定期的にメモリの読み出し要求（リードリクエスト）を発行し、メモリの状態を常に監視するため、早期のエラー検出が可能となる。
【００７３】
また、多重化システムにおけるスタンバイ系（待機系、従系）装置等、普段は動作していない装置のエラー検出も可能となる。
【００７４】
［ソフトウェア（ＳＷ）状態監視］
以下に、ソフトウェア（ＳＷ）状態監視の動作の詳細について説明する。
【００７５】
ＳＷ状態取得部１１２は、ソフトウェア（ＳＷ）３０の状態を取得することで、ソフトウェア（ＳＷ）３０が正常に動作しているか監視する。
【００７６】
ＳＷ状態取得部１１２は、ドライバ４０から直接通知を受け取ることで、ソフトウェア（ＳＷ）３０の状態を取得しても良いし、ＰＣＩ制御部１１３を介して、ドライバ４０がメモリ上に設定したソフトウェア（ＳＷ）３０の状態を取得しても良い。
【００７７】
これにより、ＳＷ状態取得部１１２は、ソフトウェア（ＳＷ）３０の異常を検出することが可能となる。
【００７８】
通常、故障監視をしているソフトウェア（ＳＷ）自体を用いて、当該ソフトウェア（ＳＷ）の異常を検出することは困難である。
【００７９】
本発明では、データ通信カード１０が、ソフトウェア（ＳＷ）３０の状態を監視するため、故障監視をしているソフトウェア（ＳＷ）自体のエラー検出が可能となる。
【００８０】
［Ｉ／Ｏチップ状態監視機能］
以下に、Ｉ／Ｏチップ状態監視機能の動作の詳細について説明する。
【００８１】
Ｉ／Ｏチップ状態監視が行われるのは、メイン電源がＯＮ状態（メイン電源とスタンバイ電源の両方がＯＮの状態）の時のみである。
【００８２】
ＰＣＩ制御部１１３は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、Ｉ／Ｏチップ５０の設定情報（コンフィグ）の読み出し要求（リードリクエスト）を発行し、直接に／ＣＰＵ等を介してＩ／Ｏチップ５０の設定情報の読み出しを行い、Ｉ／Ｏチップ５０／ＣＰＵ等からコンプリーション（ｃｏｍｐｌｅｔｉｏｎ：完了）の応答が正常に返ってくるかをチェックすることで、Ｉ／Ｏチップ５０が正常に動作しているか監視する。
【００８３】
例えば、ＰＣＩ制御部１１３は、一定間隔で定期的に、全Ｉ／Ｏチップ５０に対して順番に読み出し要求（リードリクエスト）を発行する。
【００８４】
なお、ＰＣＩ制御部１１３の動作は、プロセッサ１１１により変更することが可能である。
【００８５】

これにより、ＰＣＩ制御部１１３は、Ｉ／Ｏチップ５０の故障を検出することが可能となる。

【００８６】

従来は、実際にＩ／Ｏチップ５０の読み出しが必要となり、ＣＰＵ等からＩ／Ｏチップ５０の読み出し要求（リードリクエスト）が発行されるまで、エラー検出が不可能であった。

【００８７】

本発明では、データ通信カード１０が、Ｉ／Ｏチップ５０の読み出しの有無に関わらず、定期的にＩ／Ｏチップ５０の読み出し要求（リードリクエスト）を発行し、Ｉ／Ｏチップ５０の状態を常に監視するため、早期のエラー検出が可能となる。

【００８８】

また、多重化システムにおけるスタンバイ系（待機系、従系）装置等、普段は動作していない装置のエラー検出も可能となる。

【００８９】

［ＢＭＣ状態監視］

以下に、ＢＭＣ状態監視の動作の詳細について説明する。

【００９０】

なお、ＢＭＣ状態監視は、メイン電源がＯＮ状態（メイン電源とスタンバイ電源の両方がＯＮの状態）／メイン電源がＯＦＦ状態（メイン電源がＯＦＦの状態で、スタンバイ電源のみＯＮの状態）のいずれの状態であっても行われる。

【００９１】

ＳＭＢｕｓ制御部１２３は、ＳＭＢｕｓを介して、ＢＭＣ６０にＳＭＢｕｓ読み出し要求（リードリクエスト）を発行し、ＢＭＣ６０が持っているレジスタの値の読み出しを行い、ＢＭＣ６０からコンプリーション（ｃｏｍｐｌｅｔｉｏｎ：完了）の応答が正常に返ってくるかをチェックすることで、ＢＭＣ６０が正常に動作しているか監視する。

【００９２】

例えば、ＳＭＢｕｓ制御部１２３は、一定間隔で定期的に、ＢＭＣ６０にＳＭＢｕｓ読み出し要求（リードリクエスト）を発行する。

【００９３】

ＢＭＣ６０は、ＳＭＢｕｓ制御部１２３に対して、ＳＭＢｕｓ書き込み要求（ライトリクエスト）を発行し、データ通信カード１０内にあるレジスタに対してデータの書き込みを行うことも可能である。ＳＭＢｕｓ制御部１２３は、ＢＭＣ６０から一定間隔でＳＭＢｕｓ書き込み要求（ライトリクエスト）が発行されるかどうかをチェックすることで、ＢＭＣ６０が正常に動作しているか監視しても良い。

【００９４】

なお、メイン電源がＯＮ状態であれば、プロセッサ１１１は、ＳＭＢｕｓ制御部１２３の動作を変更することが可能である。

【００９５】

これにより、ＳＭＢｕｓ制御部１２３は、ＢＭＣ６０の故障を検出することが可能となる。

【００９６】

本発明では、データ通信カード１０が、定期的にＢＭＣ６０のＳＭＢｕｓ読み出し要求（リードリクエスト）を発行し、ＢＭＣ６０の状態を常に監視するため、早期のエラー検出が可能となる。

【００９７】

［自機の電源状態監視］

以下に、自機の電源状態監視の動作の詳細について説明する。

【００９８】

なお、自機の電源状態監視は、メイン電源がＯＮ状態（メイン電源とスタンバイ電源の両方がＯＮの状態）／メイン電源がＯＦＦ状態（メイン電源がＯＦＦの状態で、スタンバ

イ電源のみＯＮの状態）のいずれの状態であっても行われる。
【００９９】
電源監視部１２１は、自機の物理マシンの電源状態を監視する。なお、電源監視部１２１は、スタンバイ電源の給電により駆動している。
【０１００】
（１）メイン電源がＯＮ状態時の動作
電源監視部１２１は、自機の物理マシンのメイン電源がＯＮ状態であることを検出する。また、電源監視部１２１は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源がＯＮ状態であることを検出する。
【０１０１】
電源制御部１２２は、リモート接続先の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源をＯＮ状態にする。
【０１０２】
なお、自機の物理マシンのメイン電源がＯＮ状態であることを検出した場合、電源制御部１２２は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源をＯＮ状態にすることが可能である。リモート接続先の物理マシンのメイン電源をＯＮ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとを同時に起動する場合に有用である。
【０１０３】
例えば、ソフトウェア（ＳＷ）方式のフォールトトレラントシステムを運用する場合、必ず２台の物理マシン（アクティブ系、スタンバイ系）のＤＣ電源がＯＮ状態になっている必要がある。そのため、１台の装置の電源スイッチを押した際、連動してもう片方の装置もＤＣ電源をＯＮ状態にすることに利点（メリット）がある。逆に、１台の装置を停止（Ｓｈｕｔｄｏｗｎ）してＤＣ電源をＯＦＦ状態にする際、連動してもう片方の装置もＤＣ電源をＯＦＦ状態にすることに利点がある。２台の物理マシンを連動させるかどうかは固定的ではなく、選択可能である。
【０１０４】
また、電源制御部１２２は、自機の物理マシンのメイン電源がＯＮ状態であることを検出した場合、リモート接続先の物理マシンのメイン電源をＯＦＦ状態にすることも可能である。リモート接続先の物理マシンのメイン電源をＯＦＦ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとの間で、アクティブ系（実行系、主系）とスタンバイ系（待機系、従系）を切り替える場合に有用である。
【０１０５】
（２）メイン電源がＯＦＦ状態時の動作
電源監視部１２１は、自機の物理マシンのメイン電源がＯＦＦ状態であることを検出する。また、電源監視部１２１は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源がＯＦＦ状態であることを検出する。
【０１０６】
電源制御部１２２は、リモート接続先の物理マシンのメイン電源がＯＮ状態であることを検出した場合、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源をＯＦＦ状態にする。
【０１０７】
なお、自機の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、電源制御部１２２は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源をＯＦＦ状態にすることが可能である。リモート接続先の物理マシンのメイン電源をＯＦＦ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとを同時に停止する場合に有用である。
【０１０８】

また、電源制御部１２２は、自機の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、リモート接続先の物理マシンのメイン電源をＯＮ状態にすることも可能である。リモート接続先の物理マシンのメイン電源をＯＮ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとの間で、アクティブ系（実行系、主系）とスタンバイ系（待機系、従系）を切り替える場合に有用である。

【０１０９】

これにより、電源監視部１２１は、電源故障を検出することが可能となる。

【０１１０】

通常、ソフトウェア（ＳＷ）制御により物理マシンの電源故障を検出することは困難である。

【０１１１】

自機の物理マシンのソフトウェア（ＳＷ）により自機の物理マシンの電源故障を検出することが困難な理由は、自機の物理マシンの電源が故障すると、自機の物理マシンのソフトウェア（ＳＷ）が停止する可能性が高いためである。

【０１１２】

本発明では、データ通信カード１０が、自機の物理マシン及びリモート接続先の物理マシンの電源状態を監視するため、各物理マシンの電源故障の検出、及びいずれの物理マシンの電源故障であるかの判別・特定が可能となる。

【０１１３】

［リモート接続先の電源状態監視］

以下に、リモート接続先の電源状態監視の動作の詳細について説明する。

【０１１４】

なお、リモート接続先の電源状態監視は、メイン電源がＯＮ状態（メイン電源とスタンバイ電源の両方がＯＮの状態）／メイン電源がＯＦＦ状態（メイン電源がＯＦＦの状態で、スタンバイ電源のみＯＮの状態）のいずれの状態であっても行われる。

【０１１５】

電源監視部１２１は、リモート接続先の物理マシンの電源状態を監視する。なお、電源監視部１２１は、スタンバイ電源の給電により駆動している。

【０１１６】

（１）メイン電源がＯＮ状態時の動作

電源監視部１２１は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源がＯＮ状態であることを検出する。また、電源監視部１２１は、自機の物理マシンのメイン電源がＯＮ状態であることを検出する。

【０１１７】

電源制御部１２２は、自機の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、自機の物理マシンのメイン電源をＯＮ状態にする。

【０１１８】

なお、リモート接続先の物理マシンのメイン電源がＯＮ状態であることを検出した場合、電源制御部１２２は、通信制御部１２４を介して、自機の物理マシンのメイン電源をＯＮ状態にすることが可能である。自機の物理マシンのメイン電源をＯＮ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとを同時に起動する場合に有用である。

【０１１９】

また、電源制御部１２２は、リモート接続先の物理マシンのメイン電源がＯＮ状態であることを検出した場合、自機の物理マシンのメイン電源をＯＦＦ状態にすることも可能である。自機の物理マシンのメイン電源をＯＦＦ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとの間で、アクティブ系（実行系、主系）とスタンバイ系（待機系、従系）を切り替える場合に有用である。

【０１２０】

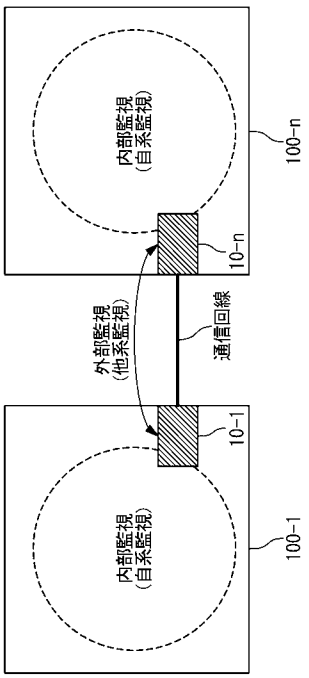
（２）メイン電源がＯＦＦ状態時の動作
電源監視部１２１は、通信制御部１２４を介して、リモート接続先の物理マシンのメイン電源がＯＦＦ状態であることを検出する。また、電源監視部１２１は、自機の物理マシンのメイン電源がＯＦＦ状態であることを検出する。
【０１２１】
電源制御部１２２は、自機の物理マシンのメイン電源がＯＮ状態であることを検出した場合、自機の物理マシンのメイン電源をＯＦＦ状態にする。
【０１２２】
なお、リモート接続先の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、電源制御部１２２は、通信制御部１２４を介して、自機の物理マシンのメイン電源をＯＦＦ状態にすることが可能である。自機の物理マシンのメイン電源をＯＦＦ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとを同時に停止する場合に有用である。
【０１２３】
また、電源制御部１２２は、リモート接続先の物理マシンのメイン電源がＯＦＦ状態であることを検出した場合、自機の物理マシンのメイン電源をＯＮ状態にすることも可能である。自機の物理マシンのメイン電源をＯＮ状態にするか否かについては、設定により変更可能である。自機の物理マシンとリモート接続先の物理マシンとの間で、アクティブ系（実行系、主系）とスタンバイ系（待機系、従系）を切り替える場合に有用である。
【０１２４】
これにより、電源監視部１２１は、電源故障を検出することが可能となる。
【０１２５】
通常、ソフトウェア（ＳＷ）制御により物理マシンの電源故障を検出することは困難である。
【０１２６】
自機の物理マシンのソフトウェア（ＳＷ）により自機の物理マシンの電源故障を検出することが困難な理由は、自機の物理マシンの電源が故障すると、自機の物理マシンのソフトウェア（ＳＷ）が停止する可能性が高いためである。
【０１２７】
自機の物理マシンのソフトウェア（ＳＷ）によりリモート接続先の物理マシンの電源故障を検出することが困難な理由は、リモート接続先の物理マシンの電源が壊れたのか、リモート接続先の物理マシンのソフトウェア（ＳＷ）の通信が止まっただけかを判断するのが困難なためである。
【０１２８】
本発明では、データ通信カード１０が、自機の物理マシン及びリモート接続先の物理マシンの電源状態を監視するため、早期のエラー検出が可能となる。
【０１２９】
［自律制御によるメイン電源のＯＦＦ］
以下に、自律制御によりメイン電源をＯＦＦ状態にする動作の詳細について説明する。
【０１３０】
データ通信カード１０は、上記の各状態監視の結果、いずれかの故障を検出した場合、ソフトウェア（ＳＷ）３０の介在なく、自律的に、故障箇所に係る装置のメイン電源をＯＦＦ状態にする。このとき、データ通信カード１０は、自機の物理マシン又はリモート接続先の物理マシン自体のメイン電源をＯＦＦ状態にしても良い。
【０１３１】
例えば、電源制御部１２２は、自機の物理マシン自体のメイン電源をＯＦＦ状態にすべきであれば、自機の物理マシン自体のメイン電源をＯＦＦ状態にする。
【０１３２】
また、電源制御部１２２は、リモート接続先の物理マシン自体のメイン電源をＯＦＦ状態にすべきであれば、通信制御部１２４を介して、リモート接続先の物理マシン自体のメ

イン電源をＯＦＦ状態にする。
【０１３３】
本発明では、データ通信カード１０が、自律的にメイン電源をＯＦＦ状態にする制御を行うため、エラー検出後、ソフトウェア（ＳＷ）の介在なく、物理マシン自体のメイン電源をＯＦＦ状態にすることが可能となる。
【０１３４】
［自律制御によるメイン電源のＯＮ］
以下に、自律制御によりメイン電源をＯＮ状態にする動作の詳細について説明する。
【０１３５】
データ通信カード１０は、リモート接続先の物理マシンにおける故障を検出した場合、自機の物理マシンのメイン電源がＯＦＦ状態であれば、自機の物理マシンのメイン電源をＯＮ状態にする。
【０１３６】
データ通信カード１０は、通信制御部１２４を介して、リモート接続先の物理マシンにおける故障を検出する。
【０１３７】
電源制御部１２２は、リモート接続先の物理マシンにおける故障が発生した場合、自機の物理マシンのメイン電源をＯＮ状態にする。
【０１３８】
本発明では、データ通信カード１０が、自律的にリモート接続先の物理マシンにおける故障を検出するため、リモート接続先の物理マシンのエラー検出後、ソフトウェア（ＳＷ）の介在なく、自機の物理マシンのメイン電源をＯＮ状態にすることが可能となる。
【０１３９】
これにより、アクティブ・スタンバイ方式の二重化システムにおいて、スタンバイ系（待機系、従系）装置のメイン電源をＯＮ状態にする必要がある時まで、メイン電源をＯＦＦ状態のまま待機させておくことが可能となる。したがって、スタンバイ系（待機系、従系）装置のメイン電源をＯＮ状態で待機させておく必要がなくなり、システム全体の消費電力を大幅に削減することが可能となる。
【０１４０】
［システム構成１（データ通信カード独立型）］
図５を参照して、データ通信カード１０が、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板から独立して存在している「システム構成１」について説明する。
【０１４１】
ここでは、データ通信カード１０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）のカードスロットに挿された拡張カードである。なお、データ通信カード１０の形状は、カード型に限らない。
【０１４２】
データ通信カード１０は、ハードウェア（ＨＷ）２０の１つであるプロセッサ（ＣＰＵ等）を介して、Ｉ／Ｏチップ５０と接続する。
【０１４３】
例えば、データ通信カード１０は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のハードウェア（ＨＷ）２０の１つであるプロセッサ（ＣＰＵ等）と接続する。このプロセッサ（ＣＰＵ等）は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、Ｉ／Ｏチップ５０と接続する。
【０１４４】
また、データ通信カード１０は、ＳＭＢｕｓを介して、ＢＭＣ６０と接続する。
【０１４５】
［システム構成２（データ通信カード一体型）］
図６を参照して、データ通信カード１０が、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板と一体化している「システム構成２」について説明する。

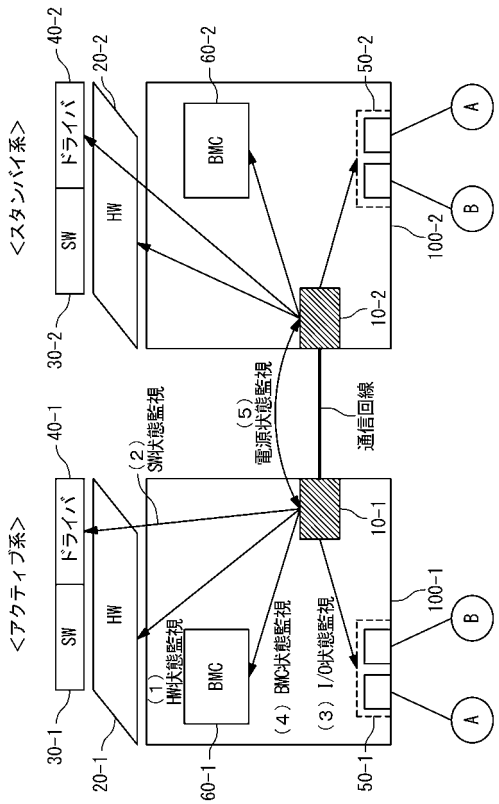
【０１４６】
ここでは、データ通信カード１０は、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板に搭載されたチップである。この場合、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板自体が、データ通信カード１０としての機能も持つことになる。すなわち、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）の基板自体が、データ通信カード１０に相当する。
【０１４７】
データ通信カード１０は、ハードウェア（ＨＷ）２０の１つであるプロセッサ（ＣＰＵ等）とＩ／Ｏチップ５０との間に存在し、このプロセッサ（ＣＰＵ等）とＩ／Ｏチップ５０との間の通信を監視する。
【０１４８】
例えば、物理マシン１００（１００－ｉ，ｉ＝１～ｎ）内部のハードウェア（ＨＷ）２０の１つであるプロセッサ（ＣＰＵ等）は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、データ通信カード１０と接続する。データ通信カード１０は、ＰＣＩ　Ｅｘｐｒｅｓｓバスを介して、Ｉ／Ｏチップ５０と接続する。
【０１４９】
また、データ通信カード１０は、ＳＭＢｕｓを介して、ＢＭＣ６０と接続する。
【０１５０】
＜本発明の特徴＞
以上のように、本発明は、一般的なＩＡサーバでソフトウェア方式のフォールトトレラントシステムやクラスターシステムを構築するために利用されるデータ通信カードに、スタンバイ電源から動作可能な装置の状態監視機能、他系装置への状態通知機能及び電源制御機能を追加することで、早期の故障検出とフェイルオーバ（ｆａｉｌｏｖｅｒ）及びコールドスタンバイ（ｃｏｌｄ　ｓｔａｎｄｂｙ）を実現する。
【０１５１】
上記の機能を追加したデータ通信カードを一般的なＩＡサーバに挿入するだけで、このデータ通信カードが自律的、定期的に、ＩＡサーバの主要コンポーネントが正常に動作しているか否かチェックを行うため、早期の故障検出が可能となる。また、このデータ通信カードは、検出した故障を他系装置に通知し、他系装置を即座にフェイルオーバ処理に遷移させることができる。
【０１５２】
また、このデータ通信カードを使用してアクティブ・スタンバイ構成の多重化システムを構築した場合、スタンバイ側の装置は、メイン電源がＯＦＦ状態であっても他系装置の状態を監視可能であり、他系装置の故障を検出した場合、自律的にメイン電源がＯＮ状態にすることも可能である。
【０１５３】
このように、このデータ通信カードは、早期の故障検出とフェイルオーバを実現し、第三者を介さない自律的なコールドスタンバイも実現する。
【０１５４】
＜備考＞
以上、本発明の実施形態を詳述してきたが、実際には、上記の実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の変更があっても本発明に含まれる。
【符号の説明】
【０１５５】
１０…　データ通信カード
１１…　ＦＰＧＡ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｇａｔｅ　Ａｒｒａｙ）
１１１…　プロセッサ
１１２…　ＳＷ（ｓｏｆｔｗａｒｅ）状態取得部
１１３…　ＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ　Ｃｏｍｐｏｎｅｎｔｓ　Ｉｎｔｅｒｃｏ

ｎｎｅｃｔ　ｂｕｓ）制御部
　１１４…　通信制御部
　１２…　ＣＰＬＤ（Ｃｏｍｐｌｅｘ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｌｏｇｉｃ　Ｄｅｖｉｃｅ）
　１２１…　電源監視部
　１２２…　電源制御部
　１２３…　ＳＭＢｕｓ（Ｓｙｓｔｅｍ　Ｍａｎａｇｅｍｅｎｔ　Ｂｕｓ）制御部
　１２４…　通信制御部
　２０…　ハードウェア（ＨＷ：ｈａｒｄｗａｒｅ）
　３０…　ソフトウェア（ＳＷ：ｓｏｆｔｗａｒｅ）
　４０…　ドライバ
　５０…　Ｉ／Ｏ（Ｉｎｐｕｔ／Ｏｕｔｐｕｔ）チップ
　６０…　ＢＭＣ（Ｂａｓｅｂｏａｒｄ　Ｍａｎａｇｅｍｅｎｔ　Ｃｏｎｔｒｏｌｌｅｒ）
　１００（－ｉ，ｉ＝１～ｎ）…　物理マシン

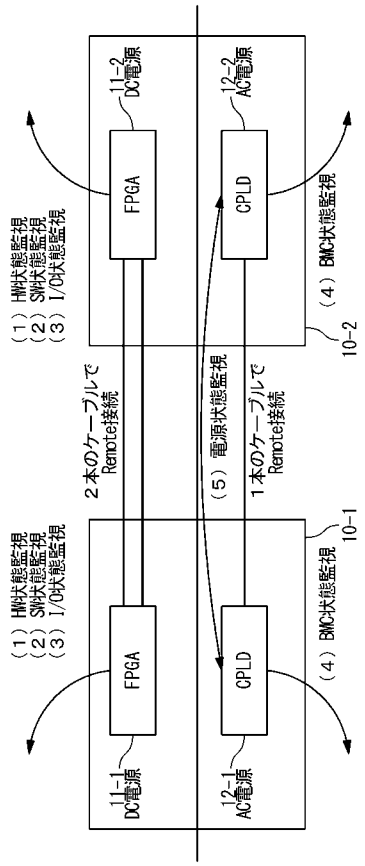
【図１】



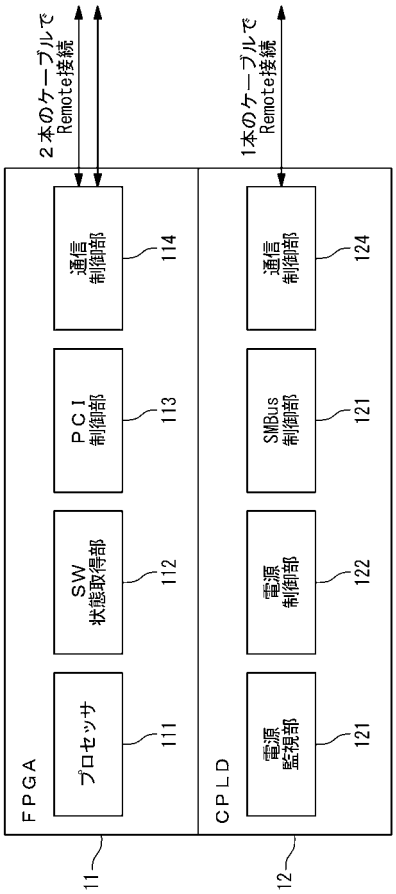
【図２】



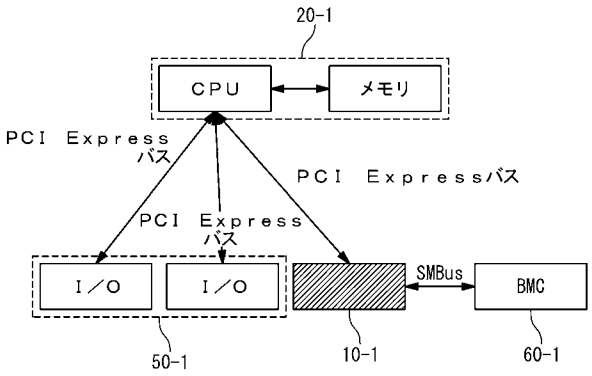
【図３】


（１）HW状態監視
（２）SW状態監視
（３）I/O状態監視
FPGA
CPLD
11-2 DC電源
12-2 AC電源
（４）BMC状態監視
10-2
２本のケーブルで Remote接続
（５）電源状態監視
１本のケーブルで Remote接続
10-1
11-1 DC電源
12-1 AC電源

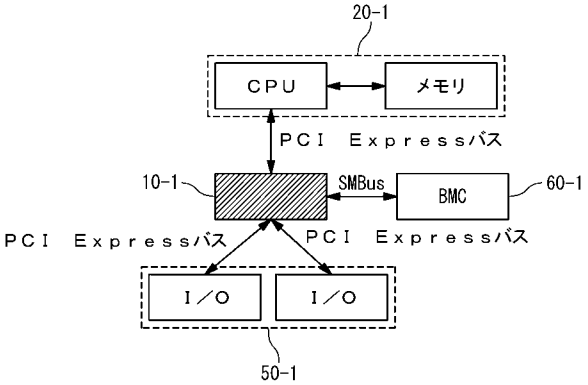
【図４】


２本のケーブルで Remote接続
１本のケーブルで Remote接続
通信制御部 114
PCI制御部 113
SW状態取得部 112
プロセッサ 111
FPGA
11
通信制御部 124
SMBus制御部 121
電源制御部 122
電源監視部 121
CPLD
12

【図５】


20-1
CPU
メモリ
PCI Express バス
PCI Expressバス
PCI Express バス
I／O
I／O
SMBus
BMC
50-1
10-1
60-1

【図６】


20-1
CPU
メモリ
PCI Expressバス
10-1
SMBus
BMC
60-1
PCI Expressバス
PCI Expressバス
I／O
I／O
50-1

フロントページの続き

(56)参考文献 特開２００１－０６０１６０（ＪＰ，Ａ）
特開２００６－１７２０５０（ＪＰ，Ａ）
特開２００９－２０５３３４（ＪＰ，Ａ）
特開２００８－２２５５６７（ＪＰ，Ａ）
特開平１０－１５０４７９（ＪＰ，Ａ）
特開２０１１－０２２９５７（ＪＰ，Ａ）
特開２０１１－１３４３１４（ＪＰ，Ａ）
特開平０８－２５１８１４（ＪＰ，Ａ）
特開２００８－０３３４８３（ＪＰ，Ａ）
特開２００９－０８０６９２（ＪＰ，Ａ）
特開２００９－０８０６９５（ＪＰ，Ａ）
特開２０１０－０７９７８９（ＪＰ，Ａ）
特開２０１０－０２０５０５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １１／１６－１１／２０，
Ｇ０６Ｆ １１／２８－１１／３４，
Ｇ０６Ｆ １／２６，１／２８