

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3872692号
(P3872692)

(45) 発行日 平成19年1月24日(2007. 1. 24)

(24) 登録日 平成18年10月27日(2006. 10. 27)

(51) Int. Cl.

G06F 12/16 (2006.01)

F I

G06F 12/16 320L

請求項の数 3 (全 24 頁)

(21) 出願番号	特願2002-3536 (P2002-3536)	(73) 特許権者	398038580
(22) 出願日	平成14年1月10日(2002. 1. 10)		ヒューレット・パカード・カンパニー
(65) 公開番号	特開2002-251331 (P2002-251331A)		HEWLETT-PACKARD COMPANY
(43) 公開日	平成14年9月6日(2002. 9. 6)		アメリカ合衆国カリフォルニア州パロアルト
審査請求日	平成16年4月13日(2004. 4. 13)		ハノーバー・ストリート 3000
(31) 優先権主張番号	09/759853	(74) 代理人	100087642
(32) 優先日	平成13年1月12日(2001. 1. 12)		弁理士 古谷 聡
(33) 優先権主張国	米国 (US)	(74) 代理人	100063897
			弁理士 古谷 馨
		(74) 代理人	100076680
			弁理士 溝部 孝彦
		(74) 代理人	100121061
			弁理士 西山 清春

最終頁に続く

(54) 【発明の名称】 データを回復するためにチェックサムを利用するシステムおよび方法

(57) 【特許請求の範囲】

【請求項 1】

データの記憶及び回復を行うシステムであって、

複数のメモリユニットのそれぞれがデータ値を記憶するための複数のメモリ位置を有し、前記メモリユニットのうちの1つにおける前記メモリ位置の1つが、他の前記メモリユニットに記憶されたデータ値に関連するチェックサムを有するように構成された、複数のメモリユニットと、

前記メモリユニット内のメモリ位置にアクセスするように構成された複数のメモリコントローラであって、前記メモリコントローラのうちの1つが、いずれの前記メモリコントローラが前記チェックサムを更新可能であることを示すインジケータを保持するように構成され、他のメモリコントローラから前記チェックサムを更新するためのチェックサム更新要求を取得し、該チェックサム更新要求に応じて前記インジケータを分析し、前記他のメモリコントローラが前記チェックサムを更新可能であることを前記インジケータが指示する場合にのみ前記チェックサム更新要求に基づいて前記チェックサムを更新するように更に構成され、前記他のメモリコントローラから前記メモリユニットに記憶された前記データ値のうちの1つを再構築するための再構築データ値を受信するように更に構成され、該再構築データ値に応じて前記再構築データ値を前記チェックサムと結合し、前記インジケータが前記他のメモリコントローラが前記チェックサムの更新をディセーブルされていることを示すように前記インジケータを更新するように更に構成された、複数のメモリコントローラと、

10

20

からなるシステム。

【請求項 2】

前記インジケータがビット列を含み、前記ビット列のそれぞれのビットが前記メモリコントローラのうちのいずれか 1 つに対応し、前記インジケータは、前記ビット列中の前記他のメモリコントローラに対応するビットが特定の論理状態を示すときにのみ、前記他のメモリコントローラが前記チェックサムを更新可能であることを示す、請求項 1 のシステム。

【請求項 3】

前記 1 つのメモリ位置がメモリユニットオフセットによって識別され、前記チェックサムに関する前記データ値のそれぞれが、前記メモリユニットオフセットによって識別されるメモリ位置に記憶される、請求項 1 のシステム。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、概して、データ処理技術に関し、特に、喪失データ値に関連するチェックサム値に基づいて喪失データ値を効率的に回復するシステムおよび方法に関する。

【0002】

【従来の技術】

大型コンピュータシステム（例えば、サーバ）は、しばしば、様々なアプリケーションのために十分な命令およびデータメモリを提供するために、複数のメモリユニットを採用する。各メモリユニットは、データを記憶することができる 1 つ以上のビット数の多数のメモリ位置を有し、各メモリ位置は、以降「メモリユニットアドレス」と呼ぶ特定のメモリアドレスに関連付けられ、これによって識別される。データを記憶する命令が実行される場合、この命令によって定義されるバスアドレスは、実際にデータが記憶されるメモリ位置を識別するメモリユニットアドレスを取得するのに利用される。この際、しばしば、バスアドレスをそのバスアドレスとは異なる値を有するメモリユニットアドレスにマップまたは変換するマップが使用される。異なるメモリユニットアドレスにマップされるバスアドレスを利用することに関しては、様々な利点がある。

20

【0003】

例えば、多くのコンピュータアプリケーションは、バスアドレスが連続的に使用されるようにプログラムされる。言換えれば、データを記憶するために最初に使用されるバスアドレスとして、バスアドレスのうちの 1 つが選択される。データを記憶するために新たなバスアドレスが利用される時、その新たなバスアドレスは、先に使用されたバスアドレスをインクリメントすることによって取得される。

30

【0004】

連続的なバスアドレスが同じメモリユニットのメモリユニットアドレスにマップされる場合、非効率になる可能性がある。この場合、メモリユニットへデータを記憶及びメモリユニットからデータを取得するための時間は限られている。2 つの連続的なデータ記憶が同じメモリユニットに対して発生する場合、第 2 のデータ記憶動作は、この第 2 のデータ記憶動作に先立って第 1 のデータ記憶動作が完了するまで待たなければならない場合がある。しかしながら、2 つの連続的なデータ記憶が異なるメモリユニットで生じる場合、第 2 のデータ記憶は第 1 のデータ記憶が完了する前に開始することができる。メモリ待ち時間を最小限にしつつメモリ帯域幅を最大限にするために、連続的なバスアドレスは可能な限り多くのメモリユニットにアクセスするべきである。これは、また、メモリインタリーブを最大限にすること、とも言うことができる。

40

【0005】

結果として、上述したマップは、しばしば、各連続的なバスアドレスが異なるメモリユニットのメモリユニットアドレスに変換されるように、バスアドレスをメモリユニットアドレスにマップするように設計される。例えば、第 1 の値を有するバスアドレスは、第 1 のメモリユニットの位置を識別するメモリユニットアドレスにマップされ、次に高い値を有

50

するバスアドレスは、第2のメモリユニットの位置を識別するメモリユニットアドレスにマップされる。従って、1つのコンピュータアプリケーションによる2つの連続的なデータ記憶が、同じメモリユニットにおいて行われなくてもありうる。言換えれば、コンピュータアプリケーションからの連続的なデータ記憶は、複数のメモリユニットにわたってインタリーブされることもありうる。

【0006】

メモリユニットの1つが障害を起こした場合にデータの回復を可能にするため、しばしばバックアップシステムが採用される。例えば、引用をもってその開示内容がすべて本明細書内に包含されたものとする米国特許第4,849,978号は、障害のあるメモリユニットのデータを回復するために使用することのできるチェックサムバックアップシステムを述べている。一般的なコンピュータシステムのメモリユニット内に記憶されるデータをバックアップするために、コンピュータシステムのメモリユニットのうちの1つが、チェックサムメモリユニットとして設計される。チェックサムメモリユニットの各メモリ位置は、他の非チェックサムメモリユニットの位置と関連付けられる。動作中、チェックサム値は、後により詳細に説明する技術に従って、チェックサムメモリユニットの各メモリ位置に保持される。各チェックサム値は、チェックサム値を記憶しているチェックサムメモリ位置と相関するメモリ位置のいずれかに記憶される非チェックサムデータ値のいずれかを回復するために利用することができる。チェックサムメモリ位置に記憶されるチェックサム値と、チェックサムメモリ位置と相関する位置に記憶される非チェックサム値のそれぞれを、本明細書ではまとめて「チェックサムセット」と呼ぶ。

【0007】

チェックサムメモリユニットの各位置は、0に初期化される。非チェックサムメモリユニットのうちの1つの位置に記憶されている各データ値は、その1つの非チェックサムメモリユニットの位置に予め記憶されたデータ値と排他的論理和がとられる。言換えれば、データ記憶動作によって記憶されているデータ値は、このデータ記憶動作によって上書きされるデータ値と排他的論理和がとられる。そして、この排他的論理和演算の結果は、チェックサムメモリユニットの相関するアドレスの「チェックサム」と呼ばれる値と排他的論理和がとられる。そして、上記排他的論理和演算の結果は、チェックサムメモリユニットの上記アドレスに、新たなチェックサム値として記憶される。

【0008】

メモリユニットに障害が発生した場合、障害が発生したメモリユニットの位置に記憶されているデータ値は、チェックサムメモリユニットの相関する位置のチェックサム値を、同様にそのチェックサムの位置と相関する位置に記憶されている他のメモリユニットの値の各々と排他的論理和をとることにより、回復することができる。チェックサムを保持しおよびチェックサムに基づいて喪失データ値を回復するプロセスは、本技術分野において一般的に知られている。

【0009】

【発明が解決しようとする課題】

不都合なことに、チェックサムセットの喪失データ値の回復中、大抵のコンピュータシステムは、同じチェックサムセットの他のデータ値を記憶するメモリ位置へのデータの書込みを行うことができない。この場合、メモリユニットのうちの1つへデータ値を書込むことにより、喪失データ値を回復するために利用されているチェックサム値への更新がもたらされる場合があり、従って、現チェックサムと矛盾するデータが、喪失した情報を再構築するために使用される可能性がある。かかる更新によりチェックサム情報が矛盾しないことを保証するために、追加のステップをとることができなければ、データ回復プロセスの誤りが発生する可能性がある。

【0010】

データ回復中に、コンピュータシステムが喪失データ値のチェックサムセットへのデータ記憶を続けることが許可される場合、データ回復で使用されているチェックサムに対する更新によるデータ回復誤りを防止する追加のステップをとることは、複雑化する可能性が

10

20

30

40

50

ある。このため、大抵のコンピュータシステムは、データ回復プロセスにおける誤りを防止するために、データ回復プロセスが開始されると、チェックサムセットの非チェックサムデータ値を記憶するいかなるメモリ位置へのデータ書込みも禁止する。データ回復プロセスが完了すると、先のメモリ位置へのデータ書込みは再びイネーブルされる。しかしながら、データ回復プロセス中にコンピュータシステムがチェックサムセットへの書込み要求をサービスすることができないようにすることは、コンピュータシステムの全体的な効率を低下させる。

【0011】

このように、本業界において、コンピュータシステムに対して、チェックサムセットのデータ値を上書きする書込み要求を一時的に規制することなく、コンピュータシステムの動作不能なメモリユニットからチェックサムセットのデータ値を回復するシステムおよび方法を提供する必要があり、それはまだ解決されていない。

10

【0012】

【課題を解決するための手段】

本発明は、上述したような従来技術の欠点および欠陥を克服する。概して、本発明は、メモリから喪失データ値を効率よく回復するシステムおよび方法を提供する。

【0013】

本発明のデータ回復システムは、構造上、データ値を記憶するメモリ位置を有する複数のメモリユニットを利用する。特定のチェックサムセットの複数のデータ値のチェックサムは、メモリ位置のうちの1つに保持される。複数のデータ値のうちの1つは、残りの複数のデータ値の各々をチェックサムと結合することにより、回復することができる。データ回復プロセス中に複数のデータ値のうちの1つを取得した後、取得されたデータ値の位置にアクセスしようとするいかなる後続の試みによっても、チェックサムに対する更新がもたらされない、ということを保証するステップがとられる。従って、データ回復プロセス中、データ回復プロセスに誤りをもたらしことなく、チェックサムセットのデータ値を記憶している位置にアクセスする（例えば、読出しまたは書込みを行う）ことができる。

20

【0014】

本発明の他の特徴および利点は、当業者には、以下の詳細な説明を添付の図面に関連して読み考察することにより明らかとなろう。かかる特徴および利点はすべて、本発明の範囲内に含まれ特許請求の範囲によって保護される、ということが意図されている。

30

【0015】

本発明は、以下の図面を参照してよりよく理解することができる。図面の要素は、必ずしも互いに一定の比率で描かれる必要はなく、本発明の原理を明確に例示することに重きがおかれている。更に、各図面を通して同じ参照番号は対応する部分を示している。

【0016】

【発明の実施の形態】

図1は、複数のメモリシステム19a~19cに連結された複数の処理ユニット17a~17eを含む従来の処理システム15を示す。図2に示すように、各処理ユニット17a~17eは、本技術分野において周知の技術によって命令を実行するように構成された1つまたは複数のプロセッサ22を含む。実行中は、しばしばデータを取得および記憶することが必要である。

40

【0017】

プロセッサ22のうちの1つによる命令の実行に応じてデータ値が記憶される時、プロセッサ22は、記憶要求をマップ26に送信する。記憶要求は、記憶されるデータ値及びそのデータ値が記憶される場所を示すバスアドレスを含む。マップ26は、バスアドレスを、メモリシステム識別子とメモリユニットオフセットとを含むメモリユニットアドレスにマップするように構成されている。メモリシステム識別子は、メモリシステム19a~19cのうちの1つを識別する。マップ26は、書込み要求をメモリシステム19a~19cの各々に送信するように構成されている。書込み要求は、記憶されるデータ値、メモリシステム識別子およびメモリユニットオフセットを含み、かつそれらによって定義される

50

。

【0018】

図3に示すように、メモリシステム19a~19cの各々は、それぞれメモリコントローラ32a~32cとそれぞれメモリユニット35a~35cを含む。メモリユニット35a~35cの各々は、データを記憶し取得することができる複数のメモリ位置を含む。前述の書込み要求に含まれるメモリシステム識別子によって識別されるメモリシステム19a~19cのメモリコントローラ32a~32cは、書込み要求のデータ値(すなわち、記憶されるデータ値)を、メモリユニットオフセットに基づいて識別されたメモリシステム19a~19cのメモリユニット35a~35cに記憶するように構成されている。これに関し、メモリユニットオフセットは、メモリユニット35a~35cの各々の1つの位置に対応し、識別されたメモリシステム19a~19cのメモリコントローラ32a~32cは、受信した書込み要求のデータ値をメモリユニットオフセットに対応するメモリユニット位置に記憶するように構成されている。

10

【0019】

処理システム15のパフォーマンスを向上させるために、マップ26は、連続したバスアドレスを異なるメモリシステム19a~19cにわたってインタリーブ方式でマップするように構成されている。例えば、プロセッサ22からの各バスアドレスは、8ビットの2進情報とした場合、00000000から11111111の範囲の値をとることができる。最初の4つの連続したバスアドレスは、00000000、00000001、00000010および00000011である。プログラマは、一般に、バスアドレスを連続した順序で利用する。従って、データ値を記憶するために最初にバスアドレス00000000が使用される場合、次にはアドレス00000001が、データを記憶するために使用される次の新たなバスアドレスとなる可能性が高い。そして、バスアドレス00000010が使用されると、次はバスアドレス00000011が使用される可能性が高い。このパターンは、それ以上新たなバスアドレスが必要とされなくなるか、またはバスアドレスのすべてが使用されるまで続けられる可能性が高い。

20

【0020】

上記連続したバスアドレスをメモリシステム19a~19cにわたってインタリーブ方式でマップするために、マップ26は、第1のバスアドレス(00000000)を、メモリシステム19aを識別するメモリシステム識別子を有するメモリユニットアドレスにマップする場合がある。そして、マップ26は、第2のバスアドレス(00000001)を、メモリシステム19bを識別するメモリシステム識別子を有するメモリユニットアドレスにマップする場合がある。次いで、マップ26は、第3のバスアドレス(00000010)を、残りのメモリシステム19cを識別するメモリシステム識別子を有するメモリユニットアドレスにマップする場合がある。このプロセスは、新たなバスアドレスがマップ26によって受信されるたびに繰返される。例えば、マップ26は、第4のバスアドレス(00000011)を含む記憶要求を受信すると、第4のバスアドレスを、メモリシステム19aを識別するメモリシステム識別子を有するメモリユニットアドレスにマップする。

30

【0021】

同じメモリシステム19a~19cにマップされる各バスアドレスは、好ましくは異なるメモリユニットオフセットを有するメモリユニットアドレスにマップされる。従って、バスアドレスのセットが同じメモリシステム19a~19cにマップされても、それらバスアドレスの各々は一意のメモリユニットアドレスにマップされる。したがって、上記セットの各バスアドレスは、同じメモリシステム19a~19cの異なる位置にマップされる。

40

。

【0022】

本明細書で上述した技術と同様の技術によって、データをメモリシステム19a~19cから取得することができる。これに関して、プロセッサ22のうちの1つがデータを取得する命令を実行すると、取得要求は、プロセッサ22によりプロセッサ22に連結された

50

マップ 26 に送信される。取得要求は、データが取得される場所を示すバスアドレスを含む。マップ 26 は、バスアドレスを、メモリシステム識別子とメモリユニットオフセットとを含むメモリユニットアドレスにマップする。メモリシステム識別子は、要求されたデータを記憶しているメモリシステム 19a ~ 19c を識別し、メモリユニットオフセットは、データが実際に記憶されている識別されたメモリシステム 19a ~ 19c のメモリユニット 35a ~ 35c 内の位置に対応する。

【0023】

マップ 26 は、取得要求に応じて読出し要求をメモリシステム 19a ~ 19c の各々に送信する。読出し要求は、取得要求のバスアドレスから変換された上記メモリユニットアドレスを含み、かつそれによって定義される。読出し要求のメモリシステム識別子によって識別されるメモリシステム 19a ~ 19c のメモリコントローラ 32a ~ 32c は、関連するメモリユニット 35a ~ 35c の読出し要求のメモリユニットオフセットに対応する位置においてデータ値を取得する。そして、識別されたメモリシステム 19a ~ 19c は、取得されたデータ値を、要求している処理ユニット 17a ~ 17e に送信する。この時、メモリコントローラ 32a ~ 32c は、取得されたデータ値をマップ 26 に返し、マップ 26 がこのデータ値を、要求しているプロセッサ 22 (すなわち、上述した取得要求を発行したプロセッサ 22) に送信してよく、あるいは、メモリコントローラ 32a ~ 32c は、データ値を、マップ 26 を利用せずに要求しているプロセッサ 22 に送信してもよい。

【0024】

図 1 に示すように、処理システム 15 はまた、処理ユニット 17a ~ 17e に連結されたチェックサムメモリシステム 39 も含む。図 3 に示すように、チェックサムメモリシステム 39 は、構造上、他のメモリシステム 19a ~ 19c と同様に構成されている。より詳細には、チェックサムメモリシステム 39 は、チェックサムメモリコントローラ 41 とチェックサムメモリユニット 43 とを含む。チェックサムメモリコントローラ 41 は、チェックサムメモリシステム 39 によって受信されたメモリユニットアドレスに基づいてデータを記憶し取得するように設計されている。同じメモリユニットオフセットに対応するメモリシステム 19a ~ 19c の各メモリユニット位置のチェックサムが、上記メモリユニットオフセットに対応するチェックサムメモリユニット 39 のメモリユニット位置に保持される。言い換えれば、チェックサムセットの各非チェックサムデータ値が、メモリシステム 19a ~ 19c のうちの 1 つの同じメモリユニットオフセットによって識別されるメモリ位置に記憶され、チェックサムセットのチェックサム値が、メモリシステム 39 のこの同じメモリユニットオフセットによって識別される位置に記憶される。

【0025】

チェックサムメモリユニット 43 の値は、0 に初期化される。マップ 26 によって記憶要求から変換される各書込み要求は、他のメモリシステム 19a ~ 19c に送信される。上述したように、送信された書込み要求のメモリユニットアドレスによって識別されるメモリシステム 19a ~ 19c のメモリコントローラ 32a ~ 32c は、書込み要求のデータ値を、メモリユニットアドレスのメモリユニットオフセットに対応するメモリユニット位置 (「宛先位置」と呼ばれる) に記憶するように構成されている。新たな値が宛先位置に記憶されるため、チェックサム更新は、宛先位置に記憶されるデータ値のチェックサムセットのチェックサムを更新するように実行されなければならない。従って、データ記憶を実行する際、識別されるメモリシステム 19a ~ 19c のメモリコントローラ 32a ~ 32c は、宛先位置に記憶されるデータ値を、宛先位置に予め記憶されているデータ値と (すなわち、書込み要求に応じて上書きされたデータ値と) 排他的論理和をとるようにも設計されている。そして、上記メモリコントローラ 32a ~ 32c は、チェックサムメモリシステム 39 に対し、この排他的論理演算の結果と、宛先位置と相関するメモリユニットオフセットとを送信することにより、チェックサム更新を実施する。チェックサム更新が完了するまで、宛先位置に対するいかなる読出しまたは書込み動作も発生を開始してはならない。

10

20

30

40

50

【0026】

チェックサムメモリコントローラ41は、この結果を、受信したメモリユニットオフセット（すなわち、宛先位置を識別するメモリユニットオフセット）に対応するチェックサムメモリユニット位置に現在記憶されているデータ値と排他的論理和をとるように構成されている。そして、チェックサムメモリコントローラ41は、この排他的論理和演算の結果を、チェックサムメモリユニット43の上記位置（すなわち、宛先位置のデータ値を含むチェックサムセットのチェックサムを記憶しているチェックサムメモリユニット位置）に記憶する。これが発生すると、チェックサム更新は完了し、完了したチェックサム更新によって更新されるチェックサムメモリユニット43の位置は、宛先位置のデータ値を含むチェックサムセットのチェックサムを記憶していなければならない。チェックサム更新は、チェックサムメモリコントローラ41が同時に、同じチェックサムセットの他のメモリコントローラ32a～32cからのメモリデータを再構築している場合であっても、終了することが許可されなければならない。システム15において発生する各データ記憶に対し上述した技術を実施することにより、特定のメモリユニットオフセットに対応するチェックサムメモリユニット43のそれぞれの位置は、その特定のメモリユニットオフセットに対応するメモリユニット35a～35cの位置に現在記憶されているデータ値のチェックサムを記憶していなければならない。

10

【0027】

メモリシステム19a～19cのいずれに障害が発生しても、障害が発生したメモリシステム19a～19cに記憶されたデータ値は、チェックサムメモリシステム39のデータ値を使用することによって回復することができる。障害が発生したメモリシステム19a～19cに記憶されているデータ値を回復するために、他の使用可能なメモリシステム19a～19cの各々の同じメモリユニットオフセットに対応するメモリユニット位置における各データ値と、チェックサムメモリユニット39の上記メモリユニットオフセットに対応する位置に記憶されたチェックサムデータ値との排他的論理和がとられる。言い換えれば、チェックサムセットの喪失データ値を除く各非チェックサムデータ値は、チェックサムセットのチェックサムと排他的論理和がとられる。これら排他的論理和演算の結果は、障害が発生したメモリシステム19a～19cの同じメモリユニットオフセットに対応するメモリユニット位置に記憶されたデータ値でなければならない。

20

【0028】

しかしながら、上記排他的論理和演算中にメモリユニット19a～19cのうちの1つへのデータ値の記憶のためにチェックサムが更新される場合、データ回復プロセスの誤りが発生する可能性がある。このため、従来技術の部分で説明したように、マップ26からチェックサムセットの非チェックサム値を記憶するメモリユニット位置への書込み要求の送信は、通常、チェックサムセット内のチェックサム値を利用するいかなるデータ回復プロセス中にも行われえない。これにより、回復プロセス中のあり得るデータ誤りが防止されるが、システム15の全効率が低下する。

30

【0029】

本発明は、概して、チェックサムに基づいてデータ値を回復する改良されたシステムおよび方法に関する。図4は、本発明を実現するために利用され得るコンピュータシステム112を示す。図4に示すように、コンピュータシステム112は、好ましくは処理システム115に含まれるメモリに記憶されるコンピュータアプリケーションの命令を実行するように構成された処理システム115を含む。処理システム115は、1つまたは複数のバスを含むことができるローカルインタフェース123を介して、コンピュータシステム112内の他の要素と通信しそれらを駆動する。更に、入力装置124、例えばキーボードまたはマウスは、システム112のユーザがデータを入力するために使用することができる。画面表示装置125またはプリンタ126は、ユーザにデータを出力するために使用することができる。ディスク記憶機構127は、不揮発性ディスク（例えば、磁氣的、光学的等）との間でデータを転送するためにローカルインタフェース123に接続されることが可能である。システム112は、システム112がネットワーク134とデータを交

40

50

換することができるようにするネットワークインタフェース 1 3 3 に接続されることが可能である。

【 0 0 3 0 】

また、コンピュータシステム 1 1 2 は、後により詳細に説明するように、好ましい実施の形態においてコンピュータシステム 1 1 2 の動作を制御するために使用されるシステムマネージャ 1 3 6 も含む。システムマネージャ 1 3 6 は、ソフトウェア、ハードウェアまたはそれらの組合せによって実現することができる。なお、システムマネージャ 1 3 6 は、ソフトウェアで実現される場合、コンピュータベースシステム等の命令実行システム、装置または機器、プロセッサ内蔵システム、あるいは、命令実行システムが装置かまたは機器からの命令をフェッチしその命令を実行することができる他のシステムによって、またはそれと接続して使用される、いかなるコンピュータ読取可能媒体にも記憶し転送することができる。例えば、システムマネージャ 1 3 6 によって定義される命令は、処理システム 1 1 5 の 1 つ以上のプロセッサによって実行することが可能である。

10

【 0 0 3 1 】

本明細書のコンテキストにおいて、「コンピュータ読取可能媒体」は、命令実行システム、装置または機器によってまたはそれと接続して使用されるプログラムを、含み、記憶し、通信し、伝搬または転送することができるいかなる手段であってもよい。コンピュータ読取可能媒体は、例えば、限定されないが、電子、磁気、光、電磁気、赤外線または半導体のシステム、装置、機器または伝搬媒体とすることができる。コンピュータ読取可能媒体のより特定の例（非網羅的リスト）には、以下のものが含まれる。すなわち、1 つまたは複数のワイヤを有する電氣的接続（電子）、ポータブルコンピュータディスク（磁気）、ランダムアクセスメモリ（RAM）（磁気）、リードオンリメモリ（ROM）（磁気）、消去可能プログラム可能リードオンリメモリ（EPROMまたはフラッシュメモリ）（磁気）、光ファイバ（光）およびポータブルコンパクトディスクリードオンリメモリ（CDROM）（光）である。なお、コンピュータ読取可能媒体は、プログラムが、例えば用紙または他の媒体の光学式走査により電子的に捕捉され、その後コンパイル、翻訳または必要な場合は適当な方法で他の方法で処理され、その後コンピュータメモリに記憶されることが可能であるならば、プログラムが印刷される用紙または他の適当な媒体でさえあってもよい。

20

【 0 0 3 2 】

図 5 に示すように、処理システム 1 1 5 は、複数のメモリシステム 1 3 9 a ~ 1 3 9 e に連結された複数の処理ユニット 1 3 7 a ~ 1 3 7 e を含む。図 6 に示すように、各処理ユニット 1 3 7 a ~ 1 3 7 e は、本技術分野において周知の技術により命令を実行するように構成された 1 つまたは複数のプロセッサ 1 4 2 を含む。これら命令は、好ましくは、メモリシステム 1 3 9 a ~ 1 3 9 e のうちの 1 つまたは複数に記憶されるコンピュータアプリケーションによって定義される。

30

【 0 0 3 3 】

プロセッサ 1 4 2 のうちの 1 つによる命令の実行に応じてデータ値が記憶される時、プロセッサ 1 4 2 は、記憶要求をマップ 1 4 6 に送信する。記憶要求は、記憶されるデータ値とそのデータ値が記憶される場所を示すバスアドレスとを含む。マップ 1 4 6 は、バスアドレスを、メモリシステム識別子及びメモリユニットオフセットからなるメモリユニットアドレスにマップするように構成される。メモリシステム識別子は、メモリシステム 1 3 9 a ~ 1 3 9 e の 1 つを識別する。マップ 1 4 6 は、記憶されるデータ値、メモリシステム識別子、およびメモリユニットオフセットを含む書込み要求を、メモリシステム 1 3 9 a ~ 1 3 9 e の各々に送信するように構成される。なお、従来のシステム 1 5 のマップ 2 6 と同様に、マップ 1 4 6 は、好ましくは連続したバスアドレスを異なるメモリシステム 1 3 9 a ~ 1 3 9 e にインタリーブ方式でマップする。

40

【 0 0 3 4 】

図 7 A および図 7 B に示すように、メモリシステム 1 3 9 a ~ 1 3 9 e の各々は、それぞれメモリコントローラ 1 5 2 a ~ 1 5 2 e とそれぞれメモリユニット 1 5 5 a ~ 1 5 5 e

50

とを含む。メモリユニット155a~155eの各々は、データを記憶し取得することができる複数のメモリ位置を含む。上述した書込み要求の上述したメモリシステム識別子によって識別されるメモリシステム139a~139eのメモリコントローラ152a~152eは、記憶されるデータ値(すなわち、メモリシステム識別子およびメモリユニットオフセットと共に受信されるデータ値)を、メモリユニットオフセットに基づいて、識別されたメモリシステム139a~139eのメモリユニット155a~155eに記憶するように構成される。これに関して、メモリユニットオフセットは、メモリユニット155a~155eのそれぞれにおいて1つのメモリ位置を識別し、識別されたメモリシステム139a~139eのメモリコントローラ152a~152eは、上記メモリユニット155a~155eのメモリユニットオフセットによって識別される位置に、受信したデータ値を記憶するように構成される。

10

【0035】

同じメモリシステム139a~139eにマップされる各バスアドレスは、好ましくは、異なるメモリユニットオフセットを有するメモリユニットアドレスにマップされる。従って、バスアドレスのセットが同じメモリシステム139a~139eにマップされても、バスアドレスの各々は一意的メモリユニットアドレスにマップされる。従って、同じメモリシステム139a~139eにマップされる各バスアドレスは、同じメモリシステム139a~139eの異なる位置にマップされる。

【0036】

本明細書で上述した技術と同様の技術により、メモリユニット155a~155eからデータが取得されてよい。これに関し、プロセッサ142のうちの1つがデータを取得する命令を実行すると、プロセッサ142により取得要求がプロセッサ142に連結されたマップ146に送信される。取得要求は、データが取得される場所を示すバスアドレスを含む。マップ146は、バスアドレスを、メモリシステム識別子およびメモリユニットオフセットを含むメモリユニットアドレスにマップする。メモリシステム識別子は、要求されたデータを記憶しているメモリシステム139a~139eを識別し、メモリユニットオフセットは、識別されたメモリシステム139a~139eのメモリユニット155a~155e内において、データが実際に記憶されている位置を識別する。

20

【0037】

マップ146は、取得要求に応じて、読出し要求をメモリシステム139a~139eの各々に送信する。読出し要求は、取得要求のバスアドレスから変換される上記メモリユニットアドレスを含む。識別されるメモリシステム139a~139eのメモリコントローラ152a~152eは、関連するメモリユニット155a~155eの、メモリユニットオフセットによって識別される位置からデータ値を取得し、このデータ値を、要求している処理ユニット137a~137eに送信する。これに関し、メモリコントローラ152a~152eは、上記データ値をマップ146に返し、マップ146が、このデータ値を要求している処理ユニット137a~137eに送信してよいし、あるいは、メモリコントローラ152a~152eは、マップ146を利用せずに要求しているプロセッサユニット137a~137eにデータ値を送信してもよい。

30

【0038】

図8は、メモリユニット155a~155eの各々のより詳細な図を示す。図8において、チェックサム値またはチェックサムは、CS1~CS8部分内に記憶され、非チェックサム値は、NCS1~NCS8部分内に記憶される。従来のシステム15と同様に、同じチェックサムセットのチェックサム値と非チェックサム値とは、同じメモリユニットオフセットを有するメモリ位置に記憶される。言い換えれば、メモリシステム139a~139eのうちの1つにおける各チェックサム値は、他のメモリシステム139a~139e内でありかつチェックサム値のメモリ位置と同じメモリユニットオフセットと相関するメモリ位置に記憶される、各データ値のチェックサムを表す。しかしながら、チェックサムは、例えば図8に示すように、異なるメモリシステム139a~139eにわたってインタリーブされる。

40

50

【0039】

図8を参照すると、メモリユニット155aのCS1部分(「チェックサム1」)内のメモリ位置が、メモリユニット155b~155eのNCS1部分(「非チェックサム1」)内のメモリ位置と同じメモリユニットオフセットに相関している、と仮定する。従って、NCS1部分内の、同じメモリユニットオフセットと相関するメモリ位置の値のチェックサムは、CS1部分の、上記メモリユニットオフセットと相関するメモリ位置に記憶されなければならない。同様に、CS2~CS8部分内のメモリ位置は、それぞれNCS2~NCS8部分内のメモリ位置と同じメモリユニットオフセットと相関するものと仮定する。このように、NCS2~NCS8部分内の値のチェックサムは、それぞれCS2~CS8部分に記憶されなければならない。例えば、同じメモリユニットオフセットと相関するメモリ位置のNCS2部分内の値のチェックサムは、CS2部分の、上記メモリユニットオフセットと相関するメモリ位置に記憶されなければならない。NCS3部分内の、同じメモリユニットオフセットと相関するメモリ位置の値のチェックサムは、CS3部分の、上記メモリユニットオフセットと相関するメモリ位置に記憶されなければならない。

10

【0040】

このように、チェックサムは、従来のシステム15のように1つのメモリユニット43(図3)内に記憶されるのではなく、異なるメモリユニット155a~155eに記憶される。その結果、チェックサム更新は、メモリユニット155a~155eに互ってインタリーブされることが可能である。引用をもってその開示内容がすべて本明細書内に包含されたものとする、「Data Storage System and Method」と題され2000年10月30日に出願された米国特許出願第09/699,877号は、図8に示すように、チェックサムがメモリユニット155a~155eにわたってインタリーブされ得るように、いかにしてシステム115を構成することができるか、をより詳細に説明している。このチェックサムを移動する技術は、後で詳細に説明を行うが、従来のシステム15に関して説明したときのチェックサムが同じメモリユニット43(図3)に記憶される場合についても用いることができることに留意されたい。

20

【0041】

チェックサムがメモリユニット155a~155e内にいかに保持されるかをよりよく例示するために、図9のブロック211において、メモリシステム139bが、メモリユニット155bのNCS1部分内のメモリ位置の1つにデータ値を記憶する書込み要求を受信するものと仮定する。後述するように、NCS1部分内の1つのメモリ位置(以下、「宛先」と呼ぶ)は、特定のメモリユニットオフセットに相関されていなければならない。書込み要求の発生前に宛先にすでに記憶されている値のチェックサムは、メモリユニット155aのCS1部分の、同じメモリユニットオフセットと相関するメモリ位置(以下、「チェックサム位置」と呼ぶ)に記憶されていなければならない。

30

【0042】

図9のブロック214に示すように、書込み要求に応じて、メモリコントローラ152bは、書込み要求のデータ値(すなわち、記憶されるデータ値)を、宛先に予め記憶されているデータ値と排他的論理和をとり、その後宛先に、書込み要求のデータ値を記憶する。そして、図9のブロック216および219に示すように、排他的論理和演算の結果は、メモリコントローラ152bによって、チェックサム位置を含むメモリシステム139aのメモリコントローラ152aを識別する第2の書込み要求(「チェックサム更新要求」)で送信される。この場合、このチェックサム更新要求は、少なくともメモリシステム139aを識別するメモリシステム識別子と、排他的論理和演算の結果と、宛先と相関するメモリユニットオフセットとを含む。

40

【0043】

このチェックサム更新要求は、メモリコントローラ152bによって他のメモリシステム139aおよび139c~139eの各々に送信される。チェックサム更新要求に応じて、メモリコントローラ152aは、図9のブロック225において、後により詳細に説明するブロック231の状態が満足されると、チェックサム位置に記憶されるチェックサム

50

を更新する。より詳細には、ブロック 2 2 5 において、メモリコントローラ 1 5 2 a は、チェックサム位置に記憶されるチェックサムを、図 9 のブロック 2 1 9 においてメモリコントローラ 1 5 2 b から送信される排他的論理和の結果と排他的論理和をとる。そして、メモリコントローラ 1 5 2 a によって実行される排他的論理和演算の結果は、チェックサム位置に記憶される。その結果、図 9 のブロック 2 1 4 において宛先に記憶されたデータ値に対するチェックサム位置のチェックサムの更新が完了する。

【 0 0 4 4 】

上述したように、メモリコントローラ 1 5 5 a ~ 1 5 5 e は、マップ 1 4 6 からの書込み要求に応じて宛先メモリ位置にデータ値を記憶する時、図 9 のブロック 2 1 4 において、記憶されるデータ値を宛先メモリ位置に予め記憶されているデータ値と排他的論理和をとらなければならない。そして、図 9 のブロック 2 1 9 において、この排他的論理和演算の結果は、チェックサムが更新され得るように、宛先位置に関連するチェックサムを記憶しているメモリシステム 1 3 9 a ~ 1 3 9 e に送信されなければならない。

10

【 0 0 4 5 】

いずれのメモリシステム 1 3 9 a ~ 1 3 9 e が宛先位置に関連するチェックサムを記憶しているかを識別することを可能にするために、メモリコントローラ 1 5 2 a ~ 1 5 2 e は、好ましくは、図 7 A および図 7 B に示すように、それぞれチェックサムインジケータ 1 6 1 a ~ 1 6 1 e を保持する。各チェックサムインジケータ 1 6 1 a ~ 1 6 1 e は、メモリシステム 1 3 9 a ~ 1 3 9 e 内の各特定のメモリ位置に対し、いずれのメモリシステム 1 3 9 a ~ 1 3 9 e がその特定のメモリ位置に現在記憶されているデータ値のチェックサムを記憶しているかを示す。

20

【 0 0 4 6 】

このように、上述した例では、メモリコントローラ 1 5 2 b によって保持されるチェックサムインジケータ 1 6 1 b は、メモリシステム 1 3 9 a が宛先に関連するチェックサムを記憶している（すなわち、宛先と同じメモリユニットオフセットと関連するメモリ位置にチェックサムを記憶している）ということを示さなければならない。従って、メモリコントローラ 1 5 2 b は、宛先に記憶されるデータ値を宛先に予め記憶されているデータ値と排他的論理和をとった後、ブロック 2 1 6 において、チェックサムインジケータ 1 6 1 b を分析することにより、他のメモリシステム 1 3 9 a および 1 3 9 c ~ 1 3 9 e のうちのいずれが宛先に関連するチェックサムを記憶しているかを判定する。メモリコントローラ 1 5 2 b は、チェックサムインジケータ 1 6 1 b に基づきメモリシステム 1 3 9 a を識別し、これによって、メモリコントローラ 1 5 2 a に対し、メモリシステム 1 3 9 a すなわちメモリコントローラ 1 5 2 a を識別するチェックサム更新要求を送信しなければならない。このチェックサム更新要求は、図 9 のブロック 2 1 4 において、メモリコントローラ 1 5 2 b によって実行される排他的論理和演算の結果を含む。これに応じて、メモリコントローラ 1 5 2 a は、図 9 のブロック 2 2 5 において、上述したように、排他的論理和の結果に基づいて宛先に関連するチェックサムを更新する。

30

【 0 0 4 7 】

チェックサムインジケータ 1 6 1 a ~ 1 6 1 e は、種々の技術によって実現することができる。図 1 0 に示す好ましい実施の形態では、各チェックサムインジケータ 1 6 1 a ~ 1 6 1 e は、複数のビット列 2 6 5 a ~ 2 6 5 f からなり、異なるメモリユニットオフセットを有するメモリ位置のそれぞれに対して 1 つのビット列 2 6 5 a ~ 2 6 5 f で実施される。ビット列 2 6 5 a ~ 2 6 5 f のそれぞれは、好ましくは n 個のビットからなり、n は、処理システム 1 1 5 内のメモリシステム 1 3 9 a ~ 1 3 9 e の数に対応する。図 5 に示す実施の形態では、システム 1 1 5 は、5 つのメモリシステム 1 3 9 a ~ 1 3 9 e を含むため、各チェックサムインジケータ 1 6 1 a ~ 1 6 1 e のビット列 2 6 5 a ~ 2 6 5 f のそれぞれは、5 つのアクティブビットを含まなければならない。ビット列 2 6 5 a ~ 2 6 5 e のそれぞれに含まれるそれぞれのアクティブビットは、メモリシステム 1 3 9 a ~ 1 3 9 e のうちの 1 つに対応する。

40

【 0 0 4 8 】

50

メモリシステム 139b がマップ 146 から書込み要求を受信する上述した例では、宛先と相関するチェックサムインジケータ 161b のビット列 265a ~ 265f は、メモリシステム 139a が宛先に関連するチェックサムを記憶している、ということを示すように適切にセットされなければならない。これに関して、チェックサムインジケータ 161b のビット列 265f が宛先に相関されていると仮定する。ビット列 265f 内で、システム 139a に対応するビットは、好ましくはアサートされ、ビット列 265f のビットの残りは、好ましくはデアサートされる。従って、ビット列 265f のビットのいずれがアサートされているかを判定することにより、メモリコントローラ 152b は、メモリシステム 139a が宛先に関連するチェックサムを記憶していると判定することができる。このように、メモリコントローラ 152b は、図 9 のブロック 214 において、メモリシステム 139a がメモリコントローラ 152b によって実行される排他的論理和演算の結果に基づいて更新されなければならないチェックサムを記憶している、ということを認識していなければならない。

10

【0049】

チェックサムを記憶しているメモリ位置と相関するいかなるビット列 265a ~ 265f も、好ましくは、それ自体のメモリシステム 139a ~ 139e を識別する。このように、上記例のチェックサムインジケータ 161a では、メモリユニット 155a のチェックサム位置と相関するビット列 265f は、好ましくは、メモリシステム 139a ~ 139e のそれぞれに対応する 5 つのビットを含む。メモリシステム 139a に対応するビットは、好ましくは、メモリシステム 139a が相関する位置（すなわち、チェックサム位置）にチェックサムを記憶していることを示すようにアサートされる。なお、チェックサム位置がチェックサム値の代りに非チェックサム値を記憶していた場合、上記ビットはデアサートされる。

20

【0050】

更に、メモリコントローラ 152a は、図 9 のブロック 219 において、メモリコントローラ 152b から送信される排他的論理和の結果を含むチェックサム更新要求を受信すると、チェックサムインジケータ 161a の上述したビット列 265f を分析する際に、受信したチェックサム更新要求が、そのチェックサム更新要求に含まれるデータ値（すなわち、排他的論理和の結果）を記憶するのではなくチェックサムを更新することを目的とする、と判定することができる。従って、メモリコントローラ 152a は、受信したチェックサム更新要求に含まれる排他的論理和の結果を記憶し他のメモリシステム 139b ~ 139e のチェックサムを更新しようとするのではなく、上述した技術に従ってチェックサム位置に記憶されるチェックサムを更新する。これに関し、メモリコントローラ 152a は、受信したチェックサム更新要求に含まれる排他的論理和の結果をチェックサム位置に予め記憶されているチェックサムと排他的論理和を取り、その後、チェックサム位置にこの排他的論理和演算の結果を記憶する。

30

【0051】

好ましい実施の形態では、チェックサムを記憶するメモリ位置と相関するビット列 265a ~ 265f はいずれも、いずれのメモリシステム 139a ~ 139e がアクティブであるかをも示す。これに関し、メモリシステム 139a ~ 139e が動作可能であり、ビット列 265a ~ 265f と相関するメモリ位置に記憶されたチェックサムの更新がイネーブルされている場合、ビット列 265a ~ 265f はメモリシステム 139a ~ 139e が「アクティブ」であることを示す。このように、上記例のチェックサムインジケータ 161a では、チェックサム位置と相関するビット列 265f の各ビットは、メモリシステム 139a ~ 139e がすべて動作可能であると仮定して、最初のアサートされる。メモリシステム 139a に対応するアサートされたビットは、メモリシステム 139a がビット列 265f と相関する位置にチェックサムを記憶していることを示し、メモリシステム 139b ~ 139e に対応しているアサートされたビットは、メモリシステム 139b ~ 139e がアクティブであることを示す。メモリシステム 139c および 139d がアクティブでありメモリシステム 139b が非アクティブ（例えば、動作不能）である場合、メ

40

50

モリスシステム 139 a、139 c および 139 d に対応しているビットのみがアサートされなければならない。かかる例において上記ビット列 265 f を分析することにより、モリスシステム 139 a がビット列 265 f と相関する位置にチェックサムを記憶していることと、モリスシステム 139 c および 139 d がアクティブであることと、モリスシステム 139 b が非アクティブであることと、が明らかとなるはずである。

【0052】

好ましい実施の形態では、チェックサム更新要求を受信するメモリコントローラ 152 a ~ 152 e は、チェックサムを更新する前に、最初にチェックサムのメモリ位置に関連するビット列 265 a ~ 265 f を分析する。上記ビット列 265 a ~ 265 f に基づいて、メモリコントローラ 152 a ~ 152 e は、チェックサム更新要求がアクティブなメモリスシステム 139 a ~ 139 e から送信されたか否かを判定する。要求がアクティブなメモリスシステム 139 a ~ 139 e から送信された場合、上記メモリコントローラ 152 a ~ 152 e はチェックサムを更新する。しかしながら、要求が、ビット列 265 a ~ 265 f によって非アクティブであると識別されるメモリスシステム 139 a ~ 139 e から送信された場合、上記メモリコントローラ 152 a ~ 152 e は、その要求を無視し、チェックサムを更新しない。

【0053】

後により詳細に説明する技術に従ってデータ回復を可能にするために、図 9 のブロック 219 においてメモリコントローラ 152 b からメモリコントローラ 152 a に送信されるチェックサム更新要求は、好ましくは、メモリスシステム 139 a のメモリスシステム識別子に加えて送信側メモリスシステム 139 b のメモリスシステム識別子を含む。メモリコントローラ 152 a は、メモリスシステム 139 a を識別するメモリスシステム識別子に基づいてチェックサム更新要求に回答し、メモリスシステム 139 b を識別するメモリスシステム識別子に基づいてメモリスシステム 155 b のメモリコントローラ 152 b がチェックサム更新要求を送信したことを知る。言い換えれば、上記メモリスシステム識別子は、いずれのメモリスシステム 139 b ~ 139 e がチェックサム更新要求を送信したかを識別する。

【0054】

図 9 のブロック 225 においてチェックサム位置に記憶されているチェックサム値を更新する前に、メモリコントローラ 152 a は、まず、図 9 のブロック 231 において、チェックサムインジケータ 161 a を分析することにより、チェックサム更新要求がアクティブなメモリスシステムから送信されたか判定する。より詳細には、書込み要求に含まれるメモリスシステム 139 b を識別するメモリスシステム識別子に基づき、メモリコントローラ 152 a は、チェックサム位置と相関するビット列 265 f のメモリスシステム 139 b に対応するビットを分析する。このビットがアサートされている場合、メモリスシステム 139 b は、アクティブであるとみなされる。従って、メモリコントローラ 152 a は、図 9 のブロック 225 に示すように、チェックサム更新要求に含まれる排他的論理和の結果に基づいてチェックサム位置のチェックサムを更新する。しかしながら、上記ビットがデアサートされている場合、メモリスシステム 139 b は、非アクティブであるとみなされる。従って、メモリコントローラ 152 a は、チェックサム更新要求を無視し、チェックサム更新要求に基づくチェックサム位置のチェックサム更新を行わない。言い換えれば、メモリコントローラ 152 a は、図 9 のブロック 225 をスキップする。

【0055】

メモリスシステム 139 a ~ 139 e のうちの 1 つに記憶されているデータ値が喪失すると（例えば、データ値を記憶しているメモリスシステム 139 a ~ 139 e に障害が発生した場合）、喪失データ値は、その喪失データ値のメモリ位置に関連するチェックサムを利用することにより、および他のアクティブなメモリスシステム 139 a ~ 139 e に記憶されているデータ値を利用することにより、回復することができる。これに関し、メモリスシステム 139 b において障害が発生するまで各メモリスシステム 139 a ~ 139 e が動作可能であると仮定する。更に、マップ 146 は、メモリスシステム 139 b の上述した宛先位置に記憶されているデータ値にアクセスするための記憶要求または取得要求を受信すると

10

20

30

40

50

仮定する。メモリシステム 139b は、すでに動作可能でないため、受信した要求を満足させることができない。更に、宛先位置のデータ値は喪失しているものとする。

【0056】

しかしながら、喪失データ値は、チェックサム位置（すなわち、メモリユニット 155a の、宛先位置と同じメモリユニットオフセットを有する位置）に記憶されているチェックサムを用いて回復することが可能である。より詳細には、喪失データ値は、チェックサム位置のチェックサムを、アクティブなメモリシステム 139c ~ 139e の、チェックサムおよび宛先位置と同じメモリユニットオフセットを有する位置に現在記憶されているデータ値と排他的論理和をとることにより、回復することができる。言い換えれば、喪失データ値は、喪失データ値のチェックサムセットの他の非チェックサム値をこのチェックサムセットのチェックサムと排他的論理和をとることによって、回復することができる。更に、後述するデータ回復技術を利用することにより、喪失データ値の回復は、回復プロセス中の動作可能なメモリユニット 155a および 155c ~ 155e へのデータ記憶動作を生じさせることなく、行うことができる。

【0057】

本発明のデータ回復プロセスをよりよく例示するために、上述したように、メモリシステム 139b に障害が発生した後、マップ 146 が、メモリユニット 155b の宛先位置に記憶されているデータ値にアクセスするための取得要求または記憶要求を受信すると仮定する。取得要求または記憶要求に応じて、マップ 146 は、取得要求または記憶要求のバスアドレスをメモリユニットアドレスに変換する。次に、マップ 146 は、読出し要求または書込み要求を、宛先位置に記憶されているデータ値の取得または宛先位置へのデータ値の記憶を要求しているメモリシステム 139b へ送信する。メモリシステム 139b に障害があるため、マップ 146 により送信される要求は満足されない。マップ 146 は、メモリシステム 139b が要求に応答しなかったことを検出し、これに応じて、喪失データ値（すなわち、障害が発生したメモリシステム 139b 内の宛先位置に記憶されているデータ値）に対するプロセスを開始する。

【0058】

これについて述べると、マップ 146 はまず、図 11 のブロック 272 に示すように、他のメモリシステム 139a および 139c ~ 139e のいずれが、喪失データ値のチェックサムを記憶しているか判定する。このため、マップ 146 は、好ましくは、メモリシステム 139a ~ 139e のいずれが、喪失データ値を回復するために使用され得るチェックサムを記憶しているかを示すチェックサムインジケータ 172 を保持する。本実施例では、喪失データ値のメモリ位置に関連するチェックサムはシステム 139a に記憶されているため、インジケータ 172 はメモリシステム 139a を識別しなければならない。喪失データ値のメモリ位置に関連するチェックサムを記憶している適当なメモリシステム 139a ~ 139e を識別するために採用することのできる様々な方法がある。

【0059】

例えば、各メモリユニットオフセットに対し、チェックサムインジケータ 172 は、メモリシステム 139a ~ 139e に記憶されるチェックサムインジケータ 161a ~ 161e と同様に、各メモリシステム 139a ~ 139e に対して 1 ビットを有するビット列（図示せず）を含むことができる。各ビット列は、異なるメモリユニットオフセットに相関され、喪失データ値のメモリ位置と同じメモリユニットオフセットと相関するビット列は、好ましくは、いずれのメモリシステム 139a ~ 139d が喪失データ値のチェックサムを記憶しているかを示すために使用される。これに関して、喪失データ値のチェックサムを記憶しているメモリシステム 139a ~ 139e に対応するビットは、好ましくはアサートされ、残りのビットは、好ましくはデアサートされる。従って、いかなる喪失データ値に対しても、マップ 146 は、チェックサムインジケータ 172 を分析して、いずれのメモリシステム 139a ~ 139e が喪失データ値のメモリ位置に関連するチェックサムを記憶しているかを判定することができる。このチェックサムは、識別されたメモリシステム 139a ~ 139e の喪失データ値の位置と同じメモリユニットオフセットを有す

10

20

30

40

50

るメモリ位置に記憶されなければならない。

【0060】

このように、喪失データ値のチェックサムがメモリユニット155aのメモリ位置(「チェックサム位置」)に記憶されている上記例では、図11のブロック272において、マップ146は、チェックサムインジケータ172を分析し、メモリシステム139aが喪失データ値のメモリ位置に関連するチェックサムを記憶していることを判定する。これに応じて、マップ146は、好ましくは、図11のブロック275に示すように、メモリシステム139aに対し、喪失データ値を再構築するよう要求を送信する。この再構築要求には、宛先のメモリユニットオフセット(すなわち、喪失データ値が記憶されていたメモリユニット155bの位置)と、喪失データ値を記憶していたメモリシステム139bを識別するメモリシステムインジケータとが含まれる。

10

【0061】

これに応じて、メモリコントローラ152aは、図11のブロック277に示すように、チェックサムインジケータ161aを分析して、チェックサム位置に関し他のメモリシステム139c~139eのうちのいずれがアクティブであるか判定する。本実施例では、メモリコントローラ152aが再構築要求を受信した時、メモリシステム139c~139eの各々がアクティブであると仮定する。好ましい実施の形態に対して上述した技術に従い、チェックサムインジケータ161a内のチェックサム位置に関連するビット列265fを分析することにより、アクティブなメモリシステム139c~139eの識別が行われる。なお、チェックサム位置は、マップ146から受信された再構築要求に含まれるメモリユニットオフセットに関連するメモリユニット155a内の位置でなければならない。

20

【0062】

図11のブロック277に示すように、メモリコントローラ152aは、他のアクティブなメモリシステム139c~139eを識別した後、アクティブなメモリシステム139c~139eにおいて、チェックサム位置と同じメモリユニットオフセットに対応する位置におけるデータ値の取得を要求する。言い換えれば、メモリコントローラ152aは、動作可能なメモリシステム139c~139eのそれぞれに対し、喪失データ値を回復するために使用されているチェックサムに対するチェックサムセットの非チェックサム値の取得を要求する。

30

【0063】

より詳細には、メモリコントローラ152aは、再構築のための読み出し要求をメモリコントローラ152cへ送信し、メモリユニット155cの、再構築要求に含まれるものと同じメモリユニットオフセット(すなわち、チェックサムと宛先位置との両方に関連するメモリユニットオフセット)と関連する位置に記憶されているデータ値を取得するように要求する。また、メモリコントローラ152aは、再構築のための読み出し要求をメモリコントローラ152dへ送信し、メモリユニット155dの、再構築要求に含まれるものと同じメモリユニットオフセットと関連する位置に記憶されているデータ値を取得するようにメモリコントローラ152dに要求する。さらに、メモリコントローラ152aは、再構築のための読み出し要求をメモリコントローラ152eへ送信し、メモリユニット155eの、再構築要求に含まれるものと同じメモリユニットオフセットと関連する位置に記憶されているデータ値を取得するようにメモリコントローラ152eに要求する。再構築のための読み出し要求のそれぞれは、パラレルに送信されてもシリアルに送信されてもよい。

40

【0064】

なお、「再構築のための読み出し要求」は、取得を行っているメモリコントローラ152a~152eに対し、メモリユニット135a~135eの1つに現在記憶されているデータ値を取得し(プロセッサキャッシュからより新しいコピーを取得する必要はない)、その取得されたデータ値(以降、「再構築データ値」と呼ぶ)を、再構築データ値に基づいてデータ値を再構築する責任を有するコンポーネントへ送信するように命令する読み出し要

50

求である。好ましい実施の形態では、再構築のための読出し要求を送信しているメモリコントローラ 152 a が、かかる再構築を行う責任を有するコンポーネントに相当する。このため、本実施例では、メモリコントローラ 152 c ~ 152 e のそれぞれは、上記再構築のための読出し要求のうちの 1 つに応じて、その個々の再構築データ値をメモリコントローラ 152 a へ送信する。好ましくは、メモリコントローラ 152 c ~ 152 e のそれぞれは、受信した再構築のための読み出し要求に応じて、その再構築データ値に加えて、メモリシステム識別子だけでなく、その再構築データ値が以前送信された再構築のための読出し要求に回答するものであることを示すデータも送信する。このため、メモリコントローラ 152 a は、再構築データ値を受信すると、(1) 受信したデータ値が喪失データ値を回復するために使用されるべきものであること、及び(2) いずれのメモリシステム 152 c ~ 152 e が受信したデータ値を送信したのかを判定することができる。

10

【0065】

なお、特定のメモリ位置を取得する再構築のための読出し要求は、以前に受信したその特定のメモリ位置へ書込むための書込み要求すべてが、それらの関連するチェックサムの更新も含めて完了するまで、全うされまたは満足されてはならない。同様に、特定のメモリ位置への後の書込み要求は、再構築のための読出しが完了するまでチェックサム更新を行ってはならず、チェックサムコントローラは、喪失データ値を回復するために使用されているチェックサムに対し、特定のメモリ位置に関連するメモリコントローラからのチェックサム更新はディスエーブルされなければならない。

【0066】

20

このため、図 11 のブロック 281 ~ 284 に示すように、メモリコントローラ 152 a は、再構築データ値のそれぞれに対し、チェックサムインジケータ 161 a を更新し(ブロック 282)、再構築データ値とチェックサム位置に記憶されている値との排他的論理和をとる(ブロック 283)。これに関し、ブロック 282 において、メモリコントローラ 152 a は、チェックサムインジケータ 161 a 内のチェックサム位置と相関するビット列 265 f を更新し、再構築データ値を送信したメモリコントローラ 152 c ~ 152 e がアクティブではなくなったことをビット列 265 f が指示するように変更する。好ましい実施の形態では、これは、上記ビット列 265 f 内の、受信した再構築データ値を送信したメモリシステム 139 c ~ 139 e に対応するビットをアサートすることによって達成される。

30

【0067】

このように、メモリコントローラ 152 a は、メモリコントローラ 152 c から送信された再構築データ値を受信すると、上記ビット列 265 f 内のメモリシステム 139 c に対応するビットをデアサートする。このビットがデアサートされるまで、チェックサムコントローラ 152 a は、再構築データ値を送信したメモリコントローラ 152 c からのチェックサム更新要求を受入れなければならない。従って、受入れたチェックサム更新要求に基づいてチェックサム位置を更新しなければならない。上記ビットがデアサートされると、メモリコントローラ 152 a は、図 11 のブロック 283 を実行する際に、再構築データ値をチェックサム位置に現在記憶されているデータ値と排他的論理和をとる。そして、メモリコントローラ 152 a は、この排他的論理和演算の結果をチェックサム位置に記憶することにより、ブロック 283 を完了する。

40

【0068】

更に、メモリコントローラ 152 a は、メモリコントローラ 152 d から送信された再構築データ値を受信すると、ブロック 282 において、上記ビット列 265 f 内のメモリシステム 139 d に対応するビットをデアサートする。このビットがデアサートされると、メモリコントローラ 152 a は、ブロック 283 において、再構築データをチェックサム位置に現在記憶されているデータ値と排他的論理和をとる。そして、メモリコントローラ 152 a は、この排他的論理和演算の結果をチェックサム位置に記憶することにより、ブロック 283 を完了する。

【0069】

50

更に、メモリコントローラ 152 a は、メモリコントローラ 152 e から送信された再構築データ値を受信すると、ブロック 282 において、上記ビット列 265 f 内のメモリシステム 139 e に対応するビットをデアサートする。このビットがデアサートされると、メモリコントローラ 152 a は、ブロック 283 において、再構築データをチェックサム位置に現在記憶されているデータ値と排他的論理和をとる。そして、メモリコントローラ 152 a は、この排他的論理和演算の結果をチェックサム位置に記憶することにより、ブロック 283 を完了する。

【0070】

上述したステップの各々が完了した後、チェックサム位置に記憶されているデータ値は、喪失データ値になるはずである。従って、このデータ値は、再構築要求に応じて回復されたデータ値としてマップ 146 に送信され、または他の方法で所望に応じて利用されることができる。好ましい実施の形態では、回復されたデータ値はチェックサム位置に記憶される。このため、データ回復プロセスは、チェックサム位置のチェックサムを回復されたデータ値と有効に置換える。システムマネージャ 136 は、好ましくは、マップ 146 に対しそのエントリを更新するよう命令し、それにより喪失データ値の位置に以前マップされていたバスアドレスが、前記チェックサム位置にマップして置き換えられる。

【0071】

なお、喪失データ値の回復が完了すると、処理ユニット 17 a ~ 17 e の各々に対し問合せすることにより、その問合せされた処理ユニット 17 a ~ 17 e が、チェックサム位置に記憶されているバージョンより新しいバージョンの喪失データを有する否かを判定することができる。処理ユニット 17 a ~ 17 e のいずれかがより最新のバージョンを有している場合、チェックサム位置の値は、そのより最新のバージョンに上書きされてよい。

【0072】

更に、図 11 のブロック 282 において、受信した再構築データ値を送信したメモリコントローラ 152 a ~ 152 e のアクティブな状態を変化させるステップは、図 11 に示すデータ回復プロセス中に、使用可能なメモリシステム 139 a および 139 c ~ 139 e に対するデータ記憶を可能にするために重要である。これに関して、上記ステップは、チェックサム更新がデータ回復プロセスにおける誤りをもたらすことを防止している。

【0073】

特に、データ回復プロセスにおける誤りを防止するために、図 9 のブロック 219 において特定のメモリコントローラ 152 a ~ 152 e によって送信される各チェックサム更新要求は、特定のメモリコントローラ 152 a ~ 152 e が再構築のための読み出し要求（すなわち、図 11 のブロック 277 において送信される要求）に応答する前に、喪失データ値を再構築しているメモリコントローラ 152 a ~ 152 e により満足されなければならない。このようにチェックサム更新要求を満足することによって、回復プロセスにおいて使用されるチェックサムがデータ回復中に特定のメモリコントローラ 152 a ~ 152 e によって送信される再構築データ値と一致することが保証される。

【0074】

しかしながら、図 9 のブロック 219 において特定のメモリコントローラ 152 a ~ 152 e によって送信される各チェックサム更新要求は、特定のメモリコントローラ 152 a ~ 152 e が再構築のための読み出し要求（すなわち、図 11 のブロック 277 において送信される要求）に応じた後、喪失データ値を再構築しているメモリコントローラ 152 a ~ 152 e によって満足されてはならない。かかる要求を満足させることにより、データ回復プロセスに誤りがもたらされる可能性が高い。更に、図 11 のブロック 282 と図 9 のブロック 231 とを実行することにより、かかる要求を満足することが防止され、それにより、データ回復中にメモリシステム 139 a ~ 139 e へのデータ記憶が許可される場合の潜在的なデータ回復誤りが防止される。

【0075】

ブロック 231 および 282 がかかるすべての要求を満足させることをうまく防止することを保証するために、メモリコントローラ 152 a ~ 152 e は、再構築のための読み出し

10

20

30

40

50

要求によって要求されるデータに対する事前のチェックサム更新要求のすべてが、再構築データに応答する前に完了していることを保証しなければならない。再構築のための読出し要求に応答するメモリコントローラ152a~152eは、データに対するいかなる新たな要求をも処理する前に、再構築のための読出しが完了することと、再構築データを受信するチェックサムコントローラがチェックサムセットのチェックサムを更新するチェックサム更新要求をそれ以上受入れないことと、をも保証しなければならない。

【0076】

本発明の上述した実施の形態、特にいかなる「好ましい」実施の形態も、単に本発明の原理を明確に理解するために示される実現可能な例にすぎない、ということは強調されなければならない。本発明の思想および原理から実質的に逸脱することなく、多くの変形および変更が本発明の上述した実施の形態に対して行うことができる。かかる変更および変形はすべて、本明細書において本発明のこの開示の範囲内に含まれ、特許請求の範囲によって保護される、ということが意図されている。

【0077】

以下においては、本発明の種々の構成要件の組み合わせからなる例示的な実施態様を示す。

1. データの記憶及び回復を行うシステム(115)であって、
複数のメモリユニット(155a~155e)のそれぞれがデータ値を記憶するための複数のメモリ位置を有し、前記メモリユニットのうち1つにおける前記メモリ位置の1つが、他の前記メモリユニットに記憶されたデータ値に関連するチェックサムを有するように構成される
複数のメモリユニットと、

前記メモリユニット内のメモリ位置にアクセスするように構成された複数のメモリコントローラ(152a~152e)であって、前記メモリコントローラのうち1つが、いずれの前記メモリコントローラが前記チェックサムを更新可能であることを示すことのできるインジケータ(161a~161e)を保持し、前記チェックサムを更新するためのチェックサム更新要求を他のメモリコントローラから取得し、前記チェックサム更新要求に応じて前記インジケータを分析して前記他のメモリコントローラが前記チェックサムを更新可能であることを前記インジケータが指示するときのみ前記チェックサム更新要求に基づいて前記チェックサムを更新し、さらに、前記他のメモリコントローラから前記メモリユニットに記憶された前記データ値のうち1つを再構築するための再構築データ値を受信し、前記再構築データ値に
応答して前記再構築データ値を前記チェックサムと結合し、前記他のメモリコントローラが前記チェックサムの更新をディセーブルされていることを示すように前記インジケータを更新するように構成される複数のメモリコントローラと、
からなるシステム。

2. 前記インジケータがビット列を含み、前記ビット列のそれぞれのビットが前記メモリコントローラのうちいずれか1つと対応し、前記インジケータは、前記他のメモリコントローラに対応する前記ビット列の前記ビットが特定の論理状態を示すときのみ、前記他のメモリコントローラが前記チェックサムを更新可能であることを示す、項番1のシステム。

3. 前記1つのメモリ位置がメモリユニットオフセットによって識別され、前記チェックサムに関する前記データ値のそれぞれが、前記メモリユニットオフセットによって識別されるメモリ位置に記憶される、項番1のシステム。

4. データの記憶及び回復を行うためのシステム(115)であって、
複数のメモリユニット(155a~155e)のそれぞれがデータ値を記憶するための複数のメモリ位置を有するように構成される複数のメモリユニットと、
複数のメモリコントローラ(152a~152e)のそれぞれが、前記メモリユニットのうちいずれか1つと関連し、前記メモリコントローラに関連する前記メモリユニット内に記憶されたデータ値にアクセスし、前記メモリコントローラに関連する前記メモリユニットに含まれるそれぞれのメモリ位置に対して異なるインジケータ(161a~161e)を保持するように構成され、非チェックサム値を記憶しているメモリ位置に関連する前記異なるインジケータの

10

20

30

40

50

それぞれは、いずれの前記メモリユニットが前記非チェックサム値のチェックサムを記憶しているか示し、チェックサムを記憶しているチェックサムメモリ位置に関連する前記異なるインジケータのそれぞれは、いずれの前記メモリユニットが前記チェックサムメモリ位置の前記チェックサムを更新可能であることを示すように構成される、複数のメモリコントローラと、
からなるシステム。

5．前記メモリコントローラのうち1つは、他の前記メモリコントローラから、前記1つのメモリコントローラに関連する前記メモリユニットの特定のチェックサムメモリ位置に記憶された特定のチェックサムを更新するためのチェックサム更新要求を受信し、前記チェックサム更新要求に応答して前記特定のチェックサムメモリ位置に関連する前記インジケータを分析し、前記他のメモリコントローラが前記特定のチェックサムを更新可能であることを前記1つのメモリコントローラに関連する前記インジケータが示すときのみ、前記チェックサム更新要求に基づいて前記特定のチェックサムを更新するように構成され、さらに、前記他のメモリコントローラから前記メモリユニットに記憶された前記データ値の1つを再構築するための再構築データ値を受信し、前記再構築データ値に応答して前記再構築データ値を前記特定のチェックサムと結合し、前記他のメモリコントローラが前記特定のチェックサムの更新をディセーブルされていることを示すように前記特定のチェックサムメモリ位置に関連する前記インジケータを更新するように構成される、項番4のシステム。

6．前記特定のチェックサムメモリ位置がメモリユニットオフセットによって識別され、前記特定のチェックサムが複数の前記データ値に基づいており、前記データ値のそれぞれが前記メモリユニットオフセットによって識別されるメモリ位置に記憶される、項番5のシステム。

7．データの記憶及び回復を行うための方法であって、
データを記憶するための複数のメモリ位置を有する複数のメモリユニット(155a~155e)を設けるステップと、

前記メモリユニットのうち1つにおけるデータ値をアクセスするための複数のメモリコントローラ(152a~152e)を設けるステップと、

前記メモリユニットのうち1つにおける前記メモリ位置の1つに、他の前記メモリユニットに記憶された複数のデータ値に基づくチェックサムを保持するステップと、

前記メモリコントローラのいずれが前記チェックサムを更新可能であることを指示できるインジケータ(161a~161e)を保持するステップと、

前記メモリコントローラのうち1つから前記チェックサムを更新するためのチェックサム更新要求を受信するステップと、

前記チェックサム更新要求に応答して前記インジケータを分析するステップと、

前記分析するステップにおいて前記1つのメモリコントローラが前記チェックサムを更新可能であることを前記インジケータが示す場合にのみ、前記チェックサムを更新するステップと、

前記1つのメモリコントローラから再構築データ値を受信するステップと、

前記再構築データ値を前記チェックサムと結合するステップと、

前記結合するステップに基づいて前記データ値の1つを回復し、前記再構築データ値を受信するステップに応答して、前記1つのメモリコントローラが前記チェックサムの更新をディセーブルされていることを示すように前記インジケータを更新するステップと、

からなる方法。

8．前記1つのメモリコントローラにおいて、前記複数のメモリ位置の特定のメモリ位置から前記再構築データ値を取得するための「再構築のための読出し要求」を受信するステップと、

前記1つのメモリコントローラにおいて、前記特定のメモリ位置への書き込みを行うための書き込み要求を受信するステップと、

前記「再構築のための読出し要求」を受信するステップから前記更新するステップまでの

10

20

30

40

50

間、前記書き込み要求に基づく前記特定のメモリ位置へのデータ記憶動作を防止するステップと、

をさらに含む項番 7 の方法。

9. 前記インジケータがビット列を含み、前記ビット列のそれぞれのビットが前記メモリコントローラのうちいずれか 1 つと対応し、前記インジケータは、前記 1 つのメモリコントローラに対応する前記ビット列の前記ビットが特定の論理状態を示す場合にのみ、前記 1 つのメモリコントローラが前記チェックサムを更新可能であることを示す、項番 7 の方法。

10. 前記チェックサム及び前記複数のデータ値が、それぞれ前記同じメモリユニットオフセットによって識別されるメモリ位置に記憶される、項番 7 の方法。

10

【0078】

【発明の効果】

本発明は、上記のように構成することにより、チェックサムセットのデータ値を上書きする書き込み要求を一時的に規制することなく、データ値の回復を可能にしている。

【図面の簡単な説明】

【図 1】従来の処理システムを示すブロック図である。

【図 2】図 1 に示す処理ユニットのより詳細な図を示すブロック図である。

【図 3】図 1 に示すメモリシステムのより詳細な図を示すブロック図である。

【図 4】本発明の原理を実現するために採用されてよいコンピュータシステムを示すブロック図である。

20

【図 5】図 4 に示す処理システムのより詳細な図を示すブロック図である。

【図 6】図 5 に示す処理ユニットのより詳細な図を示すブロック図である。

【図 7 A】図 5 に示すメモリシステムのより詳細な図を示すブロック図である。

【図 7 B】図 5 に示すメモリシステムのより詳細な図を示すブロック図である。

【図 8】図 7 A および図 7 B に示すメモリユニットのより詳細な図を示すブロック図である。

【図 9】図 5 のメモリシステムのうちの 1 つにデータ値を記憶する際の、図 4 のコンピュータシステムのアーキテクチャおよび機能を示すフローチャートである。

【図 10】図 7 A および図 7 B に示すチェックサムインジケータのいずれか 1 つのより詳細な図を示すブロック図である。

30

【図 11】喪失データ値を回復する際の、図 4 に示すコンピュータシステムのアーキテクチャおよび機能を示すフローチャートである。

【符号の説明】

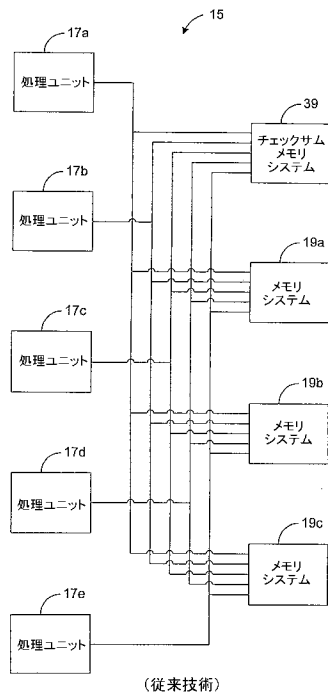
115 ... 処理システム

152 a ~ 152 e ... メモリコントローラ

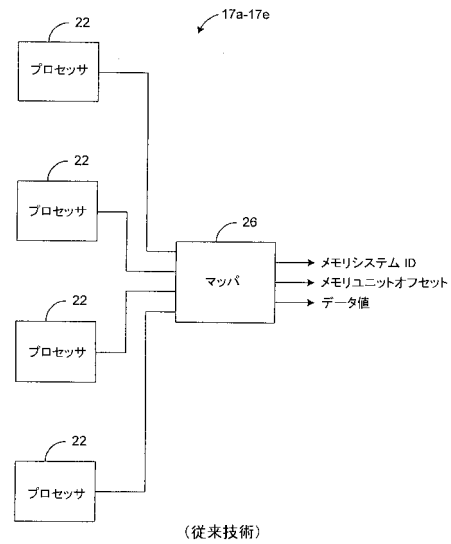
155 a ~ 155 e ... メモリユニット

161 a ~ 161 e ... チェックサムインジケータ

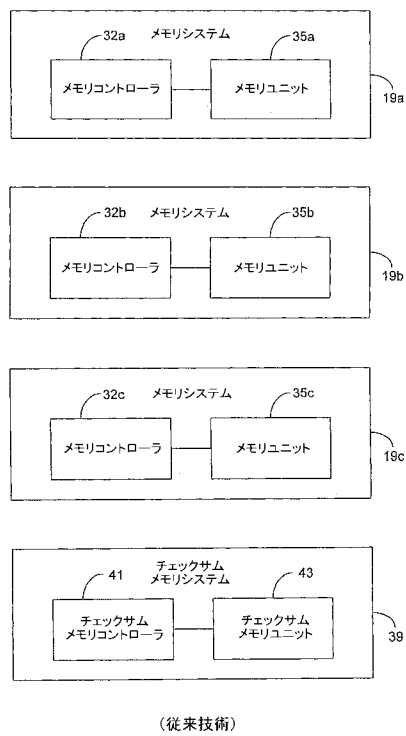
【図 1】



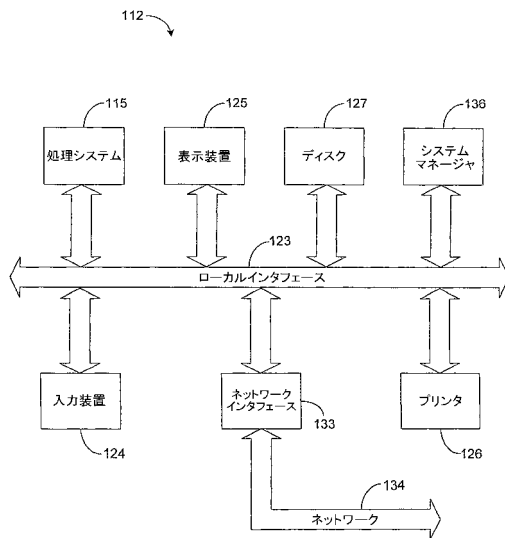
【図 2】



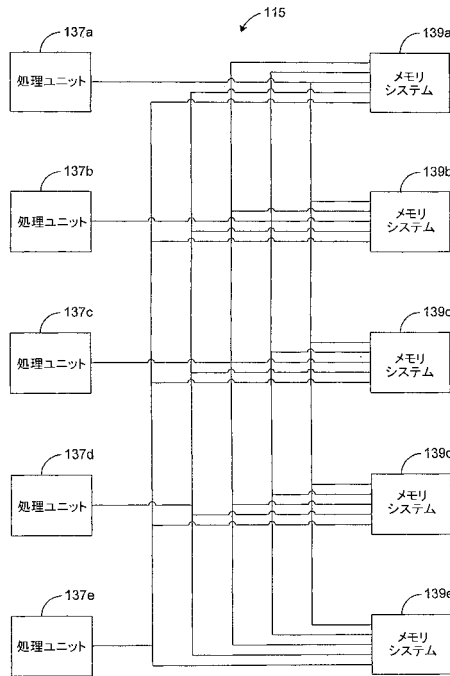
【図 3】



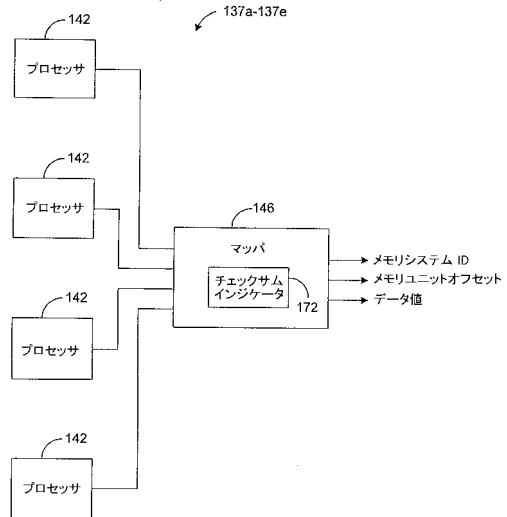
【図 4】



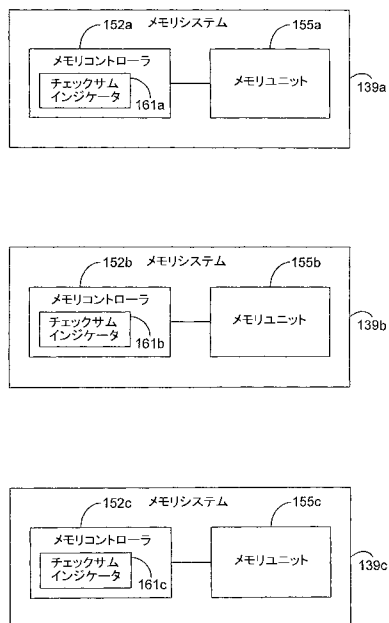
【図 5】



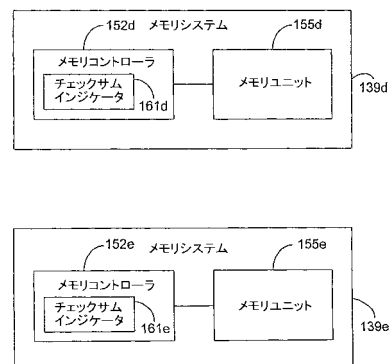
【図 6】



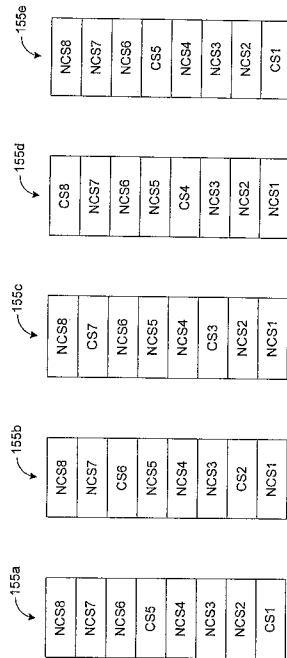
【図 7 A】



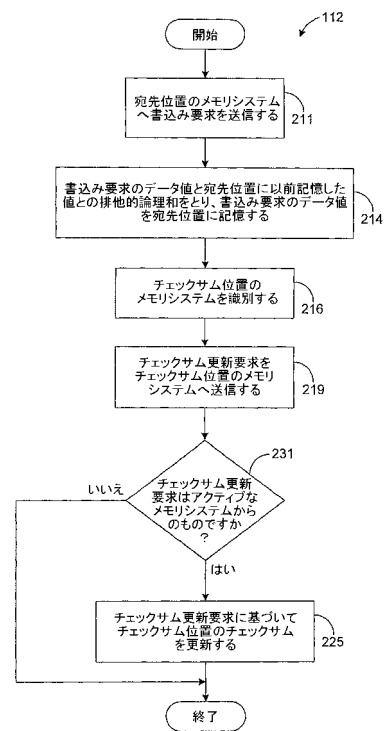
【図 7 B】



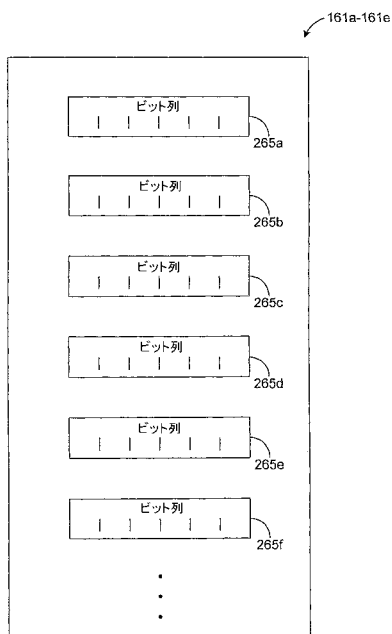
【図 8】



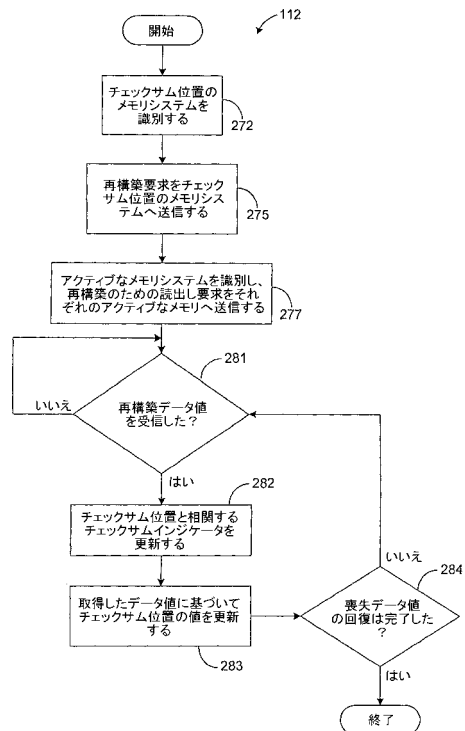
【図 9】



【図 10】



【図 11】



フロントページの続き

- (72)発明者 ブライアン・ホーニング
アメリカ合衆国テキサス州 75075, プラノ, クリークフィールド・ドライブ・1108
- (72)発明者 グレゴリー・エス・パルマー
アメリカ合衆国テキサス州 75025, プラノ, メゾン・ドライブ・3012
- (72)発明者 ボール・エフ・ボーゲル
アメリカ合衆国テキサス州 75044, ガーランド, プレア・オークス・ドライブ・3350

審査官 堀江 義隆

- (56)参考文献 特開平07-110776(JP, A)
特開昭64-021652(JP, A)
特開平08-087424(JP, A)
特開平05-143471(JP, A)
特開平08-286844(JP, A)
特開2000-010738(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/16

G06F 11/10

G06F 3/06