

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3954389号
(P3954389)

(45) 発行日 平成19年8月8日(2007.8.8)

(24) 登録日 平成19年5月11日(2007.5.11)

(51) Int. Cl.	F I
G06F 12/16 (2006.01)	G06F 12/16 320L
G06F 11/10 (2006.01)	G06F 11/10 310B

請求項の数 3 (全 29 頁)

(21) 出願番号 特願2002-3154 (P2002-3154) (22) 出願日 平成14年1月10日(2002.1.10) (65) 公開番号 特開2002-251330 (P2002-251330A) (43) 公開日 平成14年9月6日(2002.9.6) 審査請求日 平成16年9月14日(2004.9.14) (31) 優先権主張番号 09/773, 124 (32) 優先日 平成13年1月31日(2001.1.31) (33) 優先権主張国 米国(US)	(73) 特許権者 398038580 ヒューレット・パカード・カンパニー HEWLETT-PACKARD COMPANY アメリカ合衆国カリフォルニア州パロアルト ハノーバー・ストリート 3000 (74) 代理人 100081721 弁理士 岡田 次生 (74) 代理人 100105393 弁理士 伏見 直哉 (74) 代理人 100111969 弁理士 平野 ゆかり
--	---

最終頁に続く

(54) 【発明の名称】 チェックサム構築システム

(57) 【特許請求の範囲】

【請求項1】

それぞれが複数の記憶位置を持つ複数のメモリ・ユニットと、

それぞれが前記複数のメモリ・ユニットのそれぞれ1つの内部の記憶位置にアクセスするよう構成されている複数のメモリ・コントローラであって、これらメモリ・コントローラの1つが前記記憶位置の1つにチェックサムを構築するよう構成されている、複数のメモリ・コントローラと、

前記複数のメモリ・コントローラのうち、前記1つの記憶位置の更新が可能であることを示すよう構成されている標識と、

を含み、

前記1つのメモリ・コントローラが、前記チェックサム値の構築の際に、

他の各メモリ・コントローラは前記1つの記憶位置の更新を禁止されていることを示すよう前記標識を設定するステップと、

前記設定ステップに続いて、前記他の各メモリ・コントローラに再構築用読出し要求を送信するステップと、

前記再構築用読出し要求に回答して、前記複数のメモリ・ユニットから取り出された再構築値を受信するステップと、

前記1つの記憶位置を前記再構築値でそれぞれ更新するステップと、

前記各再構築値に対し前記各再構築値の受信に回答して、前記各再構築値を送信したメモリ・コントローラは前記1つの記憶位置の更新が可能であることを示すよう前記標識

10

20

を変更するステップと、
の各ステップを実行するよう構成されている、チェックサムを構築するためのシステム。

【請求項 2】

それぞれが複数の記憶位置を持つ複数のメモリ・ユニットと、
それぞれが前記複数のメモリ・ユニットのそれぞれ 1 つの内部の記憶位置にアクセスする複数のメモリ・コントローラと、
前記複数のメモリ・コントローラのうち、前記複数の記憶位置の 1 つの更新が可能であるものを示す手段と、
各メモリ・コントローラは前記 1 つの記憶位置の更新を禁止されていることを示すよう前記示す手段を設定する手段と、
前記複数のメモリ・コントローラのそれぞれに再構築用読出し要求を送信する手段と、
前記再構築用読出し要求に応答して、前記複数のメモリ・コントローラによって前記複数のメモリ・ユニットから取り出された再構築値を受信する手段と、
前記 1 つの記憶位置を前記再構築値でそれぞれ更新する手段と、
前記各再構築値に対し前記各再構築値を受信する前記受信手段に응答して、前記各再構築値を送信したメモリ・コントローラは前記 1 つの記憶位置の更新が可能であることを示すよう前記示す手段を変更する手段と、
を含む、チェックサムを構築するためのシステム。

10

【請求項 3】

20

それぞれが複数の記憶位置を持つ複数のメモリ・ユニットを提供するステップと、
複数のメモリ・コントローラを提供するステップと、
前記複数のメモリ・コントローラのうち、前記複数のメモリ・ユニットの 1 つの中の前記記憶位置の 1 つの更新が可能であるものを示す標識を維持するステップと、
前記複数のメモリ・コントローラのそれぞれが前記 1 つの記憶位置の更新を禁止されていることを示すよう前記標識を設定するステップと、
前記設定ステップに続いて、前記複数のメモリ・コントローラのそれぞれに再構築用読出し要求を送信するステップと、
前記再構築用読出し要求に응答して、前記複数のメモリ・コントローラによって前記複数のメモリ・ユニットから取り出された再構築値を受信するステップと、
前記 1 つの記憶位置を前記各再構築値で更新するステップと、
前記各再構築値に対し前記各再構築値の受信に응答して、前記各再構築値を送信したメモリ・コントローラは前記 1 つの記憶位置の更新が可能であることを示すよう前記標識を変更するステップと、
を含む、チェックサムを構築する方法。

30

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は一般にデータ処理方法に関し、特にチェックサム・セット内の非チェックサム値のアクセス性を制限することなくチェックサム・セットのチェックサムを効率的に構築するためのシステム及び方法に関する。

40

【0002】

【従来の技術】

大規模なコンピュータ・システム(例えばサーバなど)では、多様なアプリケーションに十分な命令やデータ用のメモリを提供するために、複数のメモリ・ユニットを用いることが多い。各メモリ・ユニットは多数の 1 ビット以上の記憶位置を持ち、そこにデータを格納する。各記憶位置は、固有のメモリアドレスを持ち、このアドレスによって識別される。このアドレスを以後「メモリ・ユニット・アドレス」と呼ぶ。データを格納する命令が実行されると、その命令によって定義されるバス・アドレスを用いてメモリ・ユニット・アドレスが得られ、これによって実際にデータを格納すべき記憶位置を指定する。これにつ

50

いては、バス・アドレスをバス・アドレスとは値の異なるメモリ・ユニット・アドレスに変換するマッパーがよく使われる。バス・アドレスを異なるメモリ・ユニット・アドレスに変換することには様々な利点がある。

【0003】

例えば、多くのコンピュータ・アプリケーションは、バス・アドレスを連続して用いるようにプログラムされている。換言すれば、バス・アドレスを1つ選んでデータの格納に用いる最初のバス・アドレスとし、データの格納に新たなバス・アドレスを用いるときには、前に用いたバス・アドレスを増分することによって新たなバス・アドレスを得る。

【0004】

連続したバス・アドレスを同一メモリ・ユニット内のメモリ・ユニット・アドレスに変換すると、非効率が生じることがある。これに関して、メモリ・ユニットへのデータの格納と取出しには有限の時間が必要である。もし2つの連続したデータの格納が同一のメモリ・ユニットで起きると、二番目のデータ格納は最初のデータ格納が完了するまで待たなければならないことがある。しかし、もし2つの連続したデータ格納が別々のメモリ・ユニットで起きれば、二番目のデータの格納は最初のデータの格納が完了する前に開始できるだろう。メモリの待ち時間を最小化し、メモリの帯幅を最大化するために、連続したバス・アドレスはできるだけ多くのメモリ・ユニットにアクセスすべきである。これは、メモリインターリーブの最大化、と言うこともできる。

【0005】

結果として、前述のマッパーは連続したバス・アドレスをそれぞれ異なるメモリ・ユニット内のメモリ・ユニット・アドレスに変換するよう設計されることが多い。例えば、最初の値を持つバス・アドレスは最初のメモリ・ユニット内の位置を指定するメモリ・ユニット・アドレスに変換され、次の最高値を持つバス・アドレスは第二のメモリ・ユニット内の位置を指定するメモリ・ユニット・アドレスに変換される。従って、同一メモリ・ユニット内では単一のコンピュータ・アプリケーションからの2つの連続したデータの格納は起こらない可能性が高い。換言すれば、1つのコンピュータ・アプリケーションからの連続したデータ格納は、複数のメモリ・ユニットにわたってインターリーブされる可能性が高い。

【0006】

メモリ・ユニットの1つが故障してもデータを回復できるように、バックアップシステムが用いられることが多い。例えば、参照により援用される米国特許第4,849,978号は、故障したメモリ・ユニットのデータ回復に使用できるチェックサム・バックアップ・システムを開示している。典型的なコンピュータ・システムのメモリ・ユニット内に格納されたデータをバックアップするために、そのコンピュータ・システム内のメモリ・ユニットの1つをチェックサム・メモリ・ユニットとして指定する。チェックサム・メモリ・ユニット内の各位置は、他の非チェックサム・メモリ・ユニット内の位置と相関する。演算の際、チェックサム値は以下に詳述する技法に従ってチェックサム・メモリ・ユニットの各記憶位置に保持される。各チェックサム値は、そのチェックサム値を格納しているチェックサム記憶位置に相関する任意の記憶位置に格納された任意の非チェックサムデータ値の回復にも使用できる。ここで、チェックサム記憶位置に格納されたチェックサム値と、チェックサム記憶位置に相関する位置に格納された各非チェックサム値は、本明細書において全体で「チェックサム・セット」と呼ばれる。

【0007】

チェックサム・メモリ・ユニット内の各位置をゼロに初期化する。ある非チェックサム・メモリ・ユニットのある位置に格納しようとしている各データ値と、その非チェックサム・メモリ・ユニットのその位置に以前から格納されているデータ値の、排他的論理和をとる。換言すれば、データ格納演算によって格納しようとしているデータ値と、同じデータ格納演算で上書きしようとしているデータ値の排他的論理和をとる。次に、この排他的論理和演算の結果とチェックサム・メモリ・ユニットの相関アドレス内の(「チェックサム」と呼ばれる)値の排他的論理和をとる。次に、この排他的論理和の結果を前記のチェッ

10

20

30

40

50

クサム・メモリ・ユニットのアドレスに新たなチェックサム値として格納する。

【0008】

メモリ・ユニットが故障した場合、その故障したメモリ・ユニットのある位置に格納されていたデータ値を回復するには、チェックサム・メモリ・ユニットの相関位置のチェックサムとそのチェックサム位置の相関位置に格納された他のメモリ・ユニット内の各値の排他的論理和をとればよい。チェックサムを維持してチェックサムに基づいて消失データ値を回復する過程は、当技術分野では周知である。

【0009】

チェックサム・セット内の消失データを回復する際、多くのコンピュータ・システムではチェックサム・セットのチェックサムを回復されたデータ値で置き換える。このチェックサム・セットはもはやチェックサムを含まないので、チェックサム・セットのデータをバックアップする追加の手段をとらないかぎり、他のメモリ・ユニットが故障した場合にチェックサム・セット内のデータ値を回復できない。チェックサム・セットをバックアップするためにとれる追加の手段の一例は、もう1つメモリ・ユニットを設置してそこにチェックサム・セットのチェックサムを格納することである。

10

【0010】

しかし、チェックサムの構築中に、コンピュータ・システムがチェックサム・セットへのデータの格納の継続を許可する場合、チェックサムの構築は複雑になり得る。これに関して、チェックサム・セットの記憶位置へのデータ格納は、チェックサムの構築中にチェックサム・セットの値を変えることがある。このようなチェックサム・セットの更新に配慮してチェックサムの更新が適切になされるように注意しないと、チェックサムがチェックサム・セットの非チェックサム値と矛盾してしまう可能性がある。そこで、チェックサム構築中のエラーを防ぐため、ほとんどのコンピュータ・システムでは、一旦チェックサム構築過程が始まると、チェックサム・セットの非チェックサムデータ値を格納している記憶位置へのデータ書込みを禁止する。チェックサム構築過程が完了すると、チェックサム・セットの記憶位置へのデータ書込みが再び可能になる。しかし、チェックサム構築過程中的チェックサム・セットへの書込み要求にコンピュータ・システムが対応できないと、そのコンピュータ・システム全体の効率が低下する。

20

【0011】

【発明が解決しようとする課題】

30

そこで当分野において、チェックサム・セットのデータ値を上書きするような書込み要求の処理を一時的に止める必要なしに、コンピュータ・システム内にチェックサム・セットのチェックサムを構築するシステムと方法が必要とされている。

【0012】

【課題を解決するための手段】

本発明は前述のような従来技術の不適切な点や欠点を克服する。一般に、本発明は1つのコンピュータ・システムの別々のメモリ・ユニットに格納された様々なデータ値のチェックサムを効率的に構築するシステムと方法を提供する。チェックサム構築過程に、様々なデータ値を格納している記憶位置へのデータ格納が可能となり、そのためコンピュータ・システムの性能に顕著に影響を及ぼすことなくチェックサムを構築することができる。

40

【0013】

アーキテクチャについて述べると、本発明のチェックサム構築システムは、複数のメモリ・ユニット、複数のメモリ・コントローラ、及び1つの標識を用いる。各メモリ・ユニットはデータ格納用の複数の記憶位置を持ち、各メモリ・コントローラはそれぞれのメモリ・ユニット内の記憶位置にアクセスするよう構成される。メモリ・コントローラの1つは、その記憶位置の1つにチェックサムを構築するよう構成され、標識は、他のメモリ・コントローラのうちチェックサムの1つの記憶位置を更新できるものを示す。

【0014】

チェックサムを構築するとき、1つのメモリ・コントローラを以下のステップを実行するように構成してもよい。つまり、他の各メモリ・コントローラはその1つの記憶位置の更

50

新を禁止されていることを示すよう標識を設定するステップ、設定ステップに続いて、他の各メモリ・コントローラに再構築用読出し要求(read-for rebuild request)を送信するステップ、再構築用読出し要求に応じてメモリ・ユニットから再構築値を受信するステップ、前記1つの記憶位置を各再構築値で更新するステップ、及び、1つのメモリ・コントローラが受信した各再構築値について、送信中のメモリ・コントローラがその1つの記憶位置について更新可能であることを示すよう標識を変えるステップである。

【0015】

本発明の他の特徴と利点は、以下の詳細な説明を図面と共に検討すれば、当業者にとって明らかになるだろう。

【0016】

10

【発明の実施の形態】

図1は、複数のメモリ・システム19a~19cと接続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で示すように、各処理ユニット17a~17eは1つまたは複数のプロセッサ22を含み、これらは当技術分野で周知の技法によって命令を実行するよう構成されている。実行中はデータの取出しと格納が頻繁に必要である。

【0017】

プロセッサ22の1つによる命令の実行に応答してあるデータ値を格納しようとするとき、プロセッサ22はマッパー26に格納要求を送信する。格納要求は、格納すべきデータ値と、そのデータ値をどこに格納すべきかを示すバス・アドレスを含む。マッパー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・アドレスに変換するよう構成されている。メモリ・システム識別子はメモリ・システム19a~19cのうち1つを指定する。マッパー26は、書込み要求を各メモリ・システム19a~19cに送信する。書込み要求は、格納すべきデータ値、メモリ・システム識別子、メモリ・ユニット・オフセットを含み、これらによって定義される。

20

【0018】

図3で示すように、各メモリ・システム19a~19cは、それぞれメモリ・コントローラ32a~32cとメモリ・ユニット35a~35cを含む。各メモリ・ユニット35a~35cは、データを格納し取り出すことができる複数の記憶位置を含む。前述の書込み要求のメモリ・システム識別子によって指定されたメモリ・システム19a~19cのメモリ・コントローラ32a~32cは、メモリ・ユニット・オフセットに基づいて指定されたメモリ・システム19a~19cのメモリ・ユニット35a~35cに書込み要求のデータ値(すなわち格納すべきデータ値)を格納するよう構成されている。これに関して、メモリ・ユニット・オフセットは各メモリ・ユニット35a~35cの単一の位置に対応し、指定されたメモリ・システム19a~19c内のメモリ・コントローラ32a~32cはメモリ・ユニット・オフセットに対応するメモリ・ユニット位置に受け取った書込み要求のデータ値を格納するよう構成されている。

30

【0019】

処理システム15の性能を向上するため、マッパー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にインターリーブ方式で変換するよう構成されている。例えば、プロセッサ22からの各バス・アドレスは8ビットの二進数情報であり、00000000から11111111の範囲をとり得ると仮定しよう。最小の四つの連続したバス・アドレスは、00000000、00000001、00000010、00000011である。プログラマは通常、バス・アドレスを連続して用いる。従って、データ値の格納に最初にバス・アドレス00000000を使うと、次にデータの格納に使いそうな新しいバス・アドレスは00000001である。そして次はバス・アドレス00000010、その次はバス・アドレス00000011を使うだろう。このパターンは新しいバス・アドレスの必要がなくなるまで、またはすべてのバス・アドレスを使い果たすまで続く可能性が高い。

40

【0020】

前記の連続したバス・アドレスをインターリーブ方式でメモリ・システム19a~19cにわたって変換するため、マッパー26は、最初のバス・アドレス(00000000)を、メモリ・システム19aを指定するメモリ・システム識別子を持つメモリ・ユニット・アドレスに変換する

50

。マッパー26は、2番目のバス・アドレス(00000001)を、メモリ・システム19bを指定するメモリ・システム識別子を持つメモリ・ユニット・アドレスに変換する。そして、マッパー26は、3番目のバス・アドレス(00000010)を、残りのメモリ・システム19cを指定するメモリ・システム識別子を持つメモリ・ユニット・アドレスに変換する。このプロセスを、マッパー26の受け取る新しいバス・アドレスのそれぞれについて繰り返す。例えば、マッパー26が4番目のバス・アドレス(00000011)を含む格納要求を受け取ると、マッパー26は、この4番目のバス・アドレスを、メモリ・システム19aを指定するメモリ・システム識別子を持つメモリ・ユニット・アドレスに変換する。

【0021】

同一のメモリ・システム19a~19cに変換される各バス・アドレスは、異なるメモリ・ユニット・オフセットを持つメモリ・ユニット・アドレスに変換されることが好ましい。従って、バス・アドレスの集合が同一のメモリ・システム19a~19cに変換されるが、バス・アドレスのそれぞれは唯一のメモリ・ユニット・アドレスに変換される。これに応じて、前記の集合の各バス・アドレスは同一のメモリ・システム39a~39c内の異なる位置に変換される。

【0022】

前述した技法と同様の技法によって、データをメモリ・システム19a~19cから取り出すことができる。これに関して、プロセッサ22の1つがデータを取り出すための命令を実行するとき、取出し要求はプロセッサ22によってプロセッサ22に接続されたマッパー26に送信される。取出し要求は、データを取り出すべき場所を示すバス・アドレスを含む。マッパー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・アドレスに変換する。メモリ・システム識別子は要求されたデータを格納しているメモリ・システム19a~19cを指定し、メモリ・ユニット・オフセットはデータが実際に格納されている指定されたメモリ・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】

マッパー26は、取出し要求に応答して、メモリ・システム19a~19cのそれぞれに読出し要求を送信する。読出し要求は、取出し要求のバス・アドレスから変換された前記のメモリ・ユニット・アドレスを含み、こ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指定されたメモリ・システム19a~19cのメモリ・コントローラ32a~32cは、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で関連のメモリ・ユニット35a~35c内のデータ値を取り出す。指定されたメモリ・システム19a~19cは、次に、取り出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。これに関して、メモリ・コントローラ32a~32cが取り出されたデータ値をマッパー26に戻して、マッパー26がこのデータ値を要求しているプロセッサ22(すなわち前述の取出し要求を発したプロセッサ22)に送信してもよいし、またはマッパー26を用いずに、メモリ・コントローラ32a~32cが、要求しているプロセッサ22にデータ値を送信してもよい。

【0024】

図1で示したように、処理システム15も処理ユニット17a~17eに接続されたチェックサム・メモリ・システム39を含む。図3で示したように、チェックサムメモリ・システム39は、アーキテクチャに関して他のメモリ・システム19a~19cと同様に構成されている。より具体的には、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の受信したメモリ・ユニット・アドレスに基づいて、データを格納したり取り出したりするよう設計されている。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内の各メモリ・ユニット位置のチェックサムは、前記のメモリ・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39のメモリ・ユニット位置内に保持される。換言すれば、チェックサム・セットの各非チェックサムデータ値は、同じメモリ・ユニット・オフセットで指定された記憶位置でメモリ・システム19a~19cの1つに格納され、チェックサム・セットのチェックサム値

10

20

30

40

50

はこの同じメモリ・ユニット・オフセットで指定された位置でメモリ・システム39内に格納される。

【0025】

チェックサム・メモリ・ユニット43内の値はゼロに初期化される。マッパ-26によって格納要求から変換された各書込み要求は、他のメモリ・システム19a~19cに送信される。前述したように、送信された書込み要求のメモリ・ユニット・アドレスによって指定されたメモリ・システム19a~19cのメモリ・コントローラ32a~32cは、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)に書込み要求のデータ値を格納するように構成されている。新しい値が宛先位置に格納されるので、宛先位置に格納されたデータ値のチェックサム・セット内のチェックサムを更新するために、チェックサム更新を実行すべきである。従って、データ格納の実行においては、指定されたメモリ・システム19a~19cのメモリ・コントローラ32a~32cも、宛先位置に以前格納されたデータ値(すなわち、書込み要求に応じて書き込まれるデータ値)と宛先位置に格納されようとしているデータ値との排他的論理和をとるよう設計されている。前記のメモリ・コントローラ32a~32cは、次に、この排他的論理和演算の結果と宛先位置に相関したメモリ・ユニット・オフセットとをチェックサム・メモリ・システム39に送信することによってチェックサム更新を実行する。このチェックサム更新が完了するまで、宛先位置へのこれ以上の読み出し操作または書込み操作は開始すべきでない。

【0026】

チェックサム・コントローラ41は、この結果と、受け取ったメモリ・ユニット・オフセット(すなわち宛先位置を指定するメモリ・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格納されているデータ値との排他的論理和をとるよう構成されている。チェックサム・メモリ・コントローラ41は、次に、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の前記位置(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格納するチェックサム・メモリ・ユニット位置)に格納する。一旦これが起きると、チェックサム更新が完了し、完了したチェックサム更新によって更新されたチェックサム・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格納しているはずである。たとえば、チェックサムメモリ・コントローラ41が同じチェックサム・セット内の他のメモリ・コントローラ32a~32cから記憶データを同時に再構築していても、チェックサム更新は完了を許可されるべきである。システム15内で起きる各データ格納のための前述の技法を実行することによって、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内の各位置は、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の位置内に現在格納されているデータ値のチェックサムを格納しているはずである。

【0027】

もしメモリ・システム19a~19cの1つが故障した場合、チェックサム・メモリ・システム39内のデータ値を用いることによって、故障したメモリ・システム19a~19c内に格納されていたデータ値を回復することができる。故障したメモリ・システム19a~19c内に格納されていたデータ値を回復するには、同じメモリ・ユニット・オフセットに対応するメモリ・ユニット位置の他の各作動メモリ・システム19a~19c内の各データ値と、前記のメモリ・ユニット・オフセットに対応する位置のチェックサム・メモリ・ユニット39内に格納されたチェックサムデータ値との排他的論理和をとる。換言すれば、チェックサム・セットの消失データ値以外の各非チェックサムデータ値と、チェックサム・セットのチェックサムとの排他的論理和をとる。これらの排他的論理和演算の結果は、同じメモリ・ユニット・オフセットに対応するメモリ・ユニット位置で故障したメモリ・システム19a~19cに格納されていたデータ値となるはずである。

【0028】

しかし、もし前記の排他的論理和演算中にメモリ・ユニット19a~19cの1つの中のデータ値の格納を考慮してチェックサムを更新すると、データ回復過程でエラーの起きる可能性

10

20

30

40

50

がある。そこで、マッパー26からチェックサム・セットの非チェックサム値を格納しているメモリ・ユニット位置への書込み要求の送信は、チェックサム・セット内のチェックサム値を用いるデータ回復過程の最中は防止されるのが普通である。これは、回復過程中的数据エラーの可能性を防止するが、システム15の全体効率を低下させる。

【0029】

図4は、本発明の実施に使用できるコンピュータ・システム112を示す。図4で示したように、コンピュータ・システム112は、処理システム115に含まれるメモリに格納されることが好ましいコンピュータ・アプリケーションの命令を実行するよう構成された処理システム115を含む。処理システム115は、ローカルインタフェース123を通じてコンピュータ・システム112内の他の要素と通信し、それらを駆動する。ローカルインタフェース123は、1つまたは複数のバスを含むことができる。さらに、入力装置124(例えばキーボードやマウス)はシステム112のユーザからのデータ入力に使用でき、スクリーン・ディスプレイ125やプリンタ126はユーザへのデータ出力に使用できる。ディスク記憶機構127はローカルインタフェース123と接続して不揮発性ディスク(例えば磁気式、光学式など)とデータ転送を行う。システム112は、システム112とネットワーク134とのデータ交換を可能にするネットワーク・インタフェース133に接続することができる。

【0030】

コンピュータ・システム112は、システムマネージャ136も含む。これは、後で詳述するように、コンピュータ・システム112の演算を制御するために好ましい実施形態で使用される。システムマネージャ136は、ソフトウェア、ハードウェアまたはそれらの組合せで実現することができる。システムマネージャ136をソフトウェアで実現する場合、命令実行システム、機器、装置などから命令を取得してその命令を実行できるコンピュータベースシステム、プロセッサ含有システム、その他のシステムなどの命令実行システム、器具、装置などによる使用またはこれらとの関連での使用のために、システムマネージャ136をコンピュータ可読媒体に格納して移送することができる。例えば、システムマネージャ136によって定義された命令は、処理システム115内の1つまたは複数のプロセッサによって実行されることができる。

【0031】

本明細書の文脈において、「コンピュータ可読媒体」とは、命令実行システム、機器または装置によってまたはこれらとの関連において使用するためのプログラムを収容、格納、通信、伝搬、または移送できるいかなる手段でもあり得る。コンピュータ可読媒体は、例として、しかしこれらに限定することなく、電子、磁気、光学、電磁気、赤外線または半導体のシステム、機器、装置または伝搬媒体であり得る。コンピュータ可読媒体のより具体的な例(非網羅的リスト)は以下を含む。一本または複数本の電線を持つ電氣的接続(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)、読出し専用メモリ(ROM)(磁気)、消去可能プログラム可能読出し専用メモリ(EPROMまたはフラッシュメモリ)(磁気)、光ファイバー(光学)、ポータブル・コンパクト・ディスク読出し専用メモリ(CDROM)(光学)である。また、コンピュータ可読媒体は紙やその他のプログラムを印刷できる適切な媒体でさえあり得る。紙やその他の媒体を光学的に走査するなどしてプログラムを電子的に取得し、必要に応じて適切な方法でコンパイル、解釈その他の処理を施し、コンピュータのメモリに格納できるからである。

【0032】

図5で示すように、処理システム115は複数のメモリ・システム139a~139eに接続された複数の処理ユニット137a~137eを含む。図6で示すように、各処理ユニット137a~137eは1つまたは複数のプロセッサ142を含む。プロセッサ142は、当技術分野で周知の技術によって命令を実行するよう構成されている。これらの命令は、1つまたは複数のメモリ・システム139a~139e内に格納されたコンピュータ・アプリケーションによって定義されることが好ましい。

【0033】

10

20

30

40

50

あるデータ値をプロセッサ142の1つによる命令の実行に応答して格納しようとするとき、プロセッサ142はマッパー146に格納要求を送信する。格納要求は、格納すべきデータ値とそれをどこに格納すべきかを示すバス・アドレスとを含む。マッパー146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・アドレスに変換するよう構成されている。メモリ・システム識別子はメモリ・システム139a~139eの1つを指定する。マッパー146は、格納すべきデータ値とメモリ・システム識別子とメモリ・ユニットオフセットとを含む書込み要求を、メモリ・システム139a~139eのそれぞれに送信するよう構成されている。従来のシステム15のマッパー26と同様に、マッパー146は連続したバス・アドレスを異なるメモリ・システム139a~139eにインターリーブ方式で変換することが好ましい。

10

【0034】

図7に示すように、各メモリ・システム139a~139eはそれぞれメモリ・コントローラ152a~152eとメモリ・ユニット155a~155eを含む。各メモリ・ユニット155a~155eは、データを格納し取り出すことができる複数の記憶位置を持つ。前述の書込み要求の前述のメモリ・システム識別子によって指定されたメモリ・システム139a~139eのメモリ・コントローラ152a~152eは、格納すべきデータ(すなわちメモリ・システム識別子やメモリ・ユニット・オフセットと共に受け取ったデータ値)を、メモリ・ユニット・オフセットに基づいて指定されたメモリ・システム139a~139eのメモリ・ユニット155a~155eに格納するよう構成されている。これに関して、メモリ・ユニット・オフセットは各メモリ・ユニット155a~155e内の単一の記憶位置を指定し、指定されたメモリ・システム139a~139e内のメモリ・コントローラ152a~152eは前記のメモリ・ユニット155a~155e内にメモリ・ユニット・オフセットによって指定された位置に受信したデータ値を格納するよう構成されている。

20

【0035】

同じメモリ・システム139a~139eに変換された各バス・アドレスは、異なるメモリ・ユニット・オフセットを持つメモリ・ユニット・アドレスに変換されることが好ましい。従って、バス・アドレスの集合は同一のメモリ・システム139a~139eに変換されるが、バス・アドレスのそれぞれは一意なメモリ・ユニット・アドレスに変換される。それに応じて、同じメモリ・システム139a~139eに変換された各バス・アドレスは、同じメモリ・システム139a~139e内の異なる位置に変換される。

30

【0036】

前述した技法と同様の技法によって、メモリ・ユニット155a~155eからデータを取り出すことができる。これに関して、プロセッサ142の1つがデータを取り出す命令を実行するとき、プロセッサ142はプロセッサ142に接続されたマッパー146に取出し要求を送信する。この取出し要求は、データを取り出すべき場所を示すバス・アドレスを含む。マッパー146は、このバス・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・アドレスに変換する。メモリ・システム識別子は要求されたデータを格納しているメモリ・システム139a~139eを指定し、メモリ・ユニット・オフセットはデータが実際に格納されている指定されたメモリ・システム139a~139eのメモリ・ユニット155a~155e内の位置を指定する。

40

【0037】

マッパー146は、取出し要求に応じて各メモリ・システム139a~139eに読出し要求を送信する。この読出し要求は、取出し要求のバス・アドレスから変換された前記のメモリ・ユニット・アドレスを含む。指定されたメモリ・システム139a~139eのメモリ・コントローラ152a~152eは、メモリ・ユニット・オフセットで指定された位置で関連のメモリ・ユニット155a~155e内のデータ値を取り出し、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。これに関して、メモリ・コントローラ152a~152eが前記のデータ値をマッパー146に返し、マッパー146がこのデータ値を要求している処理ユニット137a~137eに送信しても良いし、あるいは、メモリ・コントローラ152a~152eが、マッパー146を使わずに、要求している処理ユニット137a~137eにデータ値を送信しても良い。

50

【 0 0 3 8 】

図 8 は各メモリ・ユニット155a～155eをより詳細に示す。図 8 では、チェックサム値またはチェックサムは部分CS1～CS8内に格納され、非チェックサム値は部分NCS1～NCS8内に格納される。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は同じメモリ・ユニット・オフセットを持つ記憶位置に格納される。換言すれば、1つのメモリ・システム139a～139e内の各チェックサム値は、他のメモリ・システム139a～139e内にある記憶位置であってチェックサム値の記憶位置として同じメモリ・ユニットオフセットに相関している記憶位置に格納された各データ値のチェックサムを表す。しかし、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にわたってインターリーブされることが好ましい。

10

【 0 0 3 9 】

図 8 について、メモリ・ユニット155aの部分CS1(「チェックサム1」)内の記憶位置はメモリ・ユニット155b～155eの部分NCS1(「非チェックサム1」)内の記憶位置と同じメモリ・ユニット・オフセットに相関している、と仮定しよう。従って、同じメモリ・ユニット・オフセットに相関する記憶位置で部分NCS1内の値のチェックサムは、前記のメモリ・ユニット・オフセットに相関する記憶位置で部分CS1内に格納されているはずである。同様に、部分CS2～CS8内の記憶位置はそれぞれ部分NCS2～NCS8内の記憶位置と同じメモリ・ユニット・オフセットに相関している、と仮定しよう。すると、部分NCS2～NCS8内の値のチェックサムはそれぞれ部分CS2～CS8内に格納されているはずである。例えば、同じメモリ・ユニット・オフセットに相関する記憶位置の部分NCS2内の値のチェックサムは、前記のメモリ・ユニット・オフセットに相関する記憶位置の部分CS2内に格納されているはずであり、また、同じメモリ・ユニット・オフセットに相関する記憶位置の部分NCS3内の値のチェックサムは、前記のメモリ・ユニット・オフセットに相関する記憶位置の部分CS3内に格納されているはずである。

20

【 0 0 4 0 】

こうして、従来のシステム15のように単一のメモリ・ユニット43(図 3)内にチェックサムを格納する代わりに、チェックサムを異なるメモリ・ユニット155a～155eに格納する。その結果、チェックサム更新はメモリ・ユニット155a～155eにわたってインターリーブされる。2000年10月30日に出願された米国特許出願第09/699,877号の"Data Storage System and Method"は、参照により本明細書に援用されるが、図 8 に示すようにチェックサムをメモリ・ユニット155a～155eにわたってインターリーブできるようにシステム115を構成する方法についてより詳しく述べている。ここで注意すべきことは、チェックサムに基づいてデータ値を回復する技法は、後にさらに詳しく述べるが、従来のシステム15について述べたようにチェックサムが同じメモリ・ユニット43(図 3)内に格納されるときにも用いることができる、ということである。

30

【 0 0 4 1 】

チェックサムがメモリ・ユニット155a～155e内でどのように維持されるかをさらに詳しく説明するため、図 9 のステップ211において、メモリ・システム139bがメモリ・ユニット155bの部分NCS1内の記憶位置の1つにデータ値を格納するための書込み要求を受け取る、と仮定しよう。前述したように、部分NCS1内の1つの記憶位置(以後「宛先」と呼ぶ)は、ある特定のメモリ・ユニット・オフセットに相関しているはずである。書込み要求の発生前に宛先にすでに格納されていた値のチェックサムは、同じメモリ・ユニット・オフセットに相関する記憶位置(以後「チェックサム位置」と呼ぶ)でメモリ・ユニット155aの部分CS1に格納されているはずである。

40

【 0 0 4 2 】

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(すなわち格納されるべきデータ値)と宛先に以前格納されたデータ値との排他的論理和をとり、書込み要求のデータ値を宛先に格納する(図 9 のステップ214)。この排他的論理和演算の結果は2番目の書込み要求(「チェックサム更新要求」)でメモリ・コントローラ152bによって送信される(図 9 のステップ216と219)。この2番目の書込み要求は、チェックサム位置を

50

含むメモリ・システム139aのメモリ・コントローラ152aを指定する。これに関して、このチェックサム更新要求は、メモリ・システム139aを指定するメモリ・システム識別子と、排他的論理和演算の結果と、宛先に関連したメモリ・ユニット・オフセットとを少なくとも含む。

【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、139c～139eのそれぞれに送信される。チェックサム更新要求に応答して、メモリ・コントローラ152aはステップ231内の条件(後で詳述する)が満足されればチェックサム位置内に格納されたチェックサムを更新する(図9のステップ225)。より具体的には、メモリ・コントローラ152aは、チェックサム位置に格納されたチェックサムと、図9のステップ219でメモリ・コントローラ152bから送信された排他的論理和結果との排他的論理和をとる(ステップ225)。メモリ・コントローラ152aによって実行された排他的論理和演算の結果は、チェックサム位置に格納される。結果として、チェックサム位置のチェックサムは図9のステップ214で宛先に格納されていたデータ値に対して更新される。

【0044】

上述のように、メモリ・コントローラ152a～152eがマッパ146からの書き込み要求に応じて宛先記憶位置にデータ値を格納するとき、メモリ・コントローラ152a～152eは格納すべきデータ値と宛先記憶位置に以前格納されたデータ値との排他的論理和をとる(図9のステップ214)。この排他的論理和演算の結果は、宛先位置と関連したチェックサムを格納しているメモリ・システム139a～139eに送信される(図9のステップ219)。こうして、チェックサムが更新される。

【0045】

どのメモリ・システム139a～139eが宛先位置に関連するチェックサムを格納しているかを特定できるために、メモリ・コントローラ152a～152eは、図7に示すようにそれぞれチェックサム標識161a～161eを保持することが好ましい。各チェックサム標識161a～161eは、メモリ・システム139a～139e内の各特定の記憶位置について、どのメモリ・システム139a～139eが特定の記憶位置に現在格納されているデータ値のチェックサムを格納しているかを示す。

【0046】

従って、上述の例では、メモリ・コントローラ152bによって保持されるチェックサム標識161bは、宛先に関連したチェックサムをメモリ・システム139aが格納している(すなわち宛先と同じメモリ・ユニット・オフセットに関連した記憶位置にチェックサムを格納している)ということを示すはずである。従って、メモリ・コントローラ152bが宛先に格納すべきデータ値と宛先に以前格納されたデータ値との排他的論理和をとった後、メモリ・コントローラ152bは、チェックサム標識161bを解析して他のメモリ・システム139a、139c～139eのうちどれが宛先に関連したチェックサムを格納しているかを決定する(ステップ216)。チェックサム標識161bに基づいて、メモリ・コントローラ152bはメモリ・システム139aを特定し、従って、メモリ・コントローラ152aに更新チェックサム要求を送信する。この要求は、メモリ・システム139aを、従ってメモリ・コントローラ152aを指定する。このチェックサム更新要求は、図9のステップ214でメモリ・コントローラ152bによって実行された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152aは、上述のように、排他的論理和に基づいて宛先に関連したチェックサムを更新する(図9のステップ225)。

【0047】

チェックサム標識161a～161eは様々な技法によって実施できる。図10で示した好ましい実施形態において、各チェックサム標識161a～161eは複数のビット列265a～265fとして実現されており、異なるメモリ・ユニット・オフセットを持つ各記憶位置について1つのビット列265a～265fがある。各ビット列265a～265fは、処理システム115内のメモリ・システム139a～139eの数に対応するビット数nを持つことが好ましい。図5で示す実施形態において、システム115の5つのメモリ・システム139a～139eを含み、従って、各チェック

10

20

30

40

50

サム標識161a～161eの各ビット列265a～265fは5つのアクティブビットを持つべきである。各ビット列265a～265e内の各アクティブビットは、メモリ・システム139a～139eの1つに対応する。

【0048】

前述の例ではメモリ・システム139bがマッパ146から書込み要求を受け取るが、宛先に相関するチェックサム標識161bのビット列265a～265fは、メモリ・システム139aが宛先に関連したチェックサムを格納している、ということを示すよう適切に設定されるべきである。これに関して、チェックサム標識161bのビット列265fが宛先に相関している、と仮定しよう。ビット列265f内では、システム139aに対応するビットはアサートされることが好ましく、ビット列265f内の他のビットはデアサートされることが望ましい。従って、ビット列265fのどのビットがアサートされているか決定することによって、メモリ・コントローラ152bはメモリ・システム139aが宛先に関連したチェックサムを格納していることを判定できる。従って、メモリ・コントローラ152bは、メモリ・システム139bが図9のステップ214でメモリ・コントローラ152bによって実行された排他的論理和演算の結果に基づいて更新されるべきチェックサムを格納していることを知るはずである。

10

【0049】

チェックサムを格納している記憶位置に相関しているビット列265a～265fは、どれも独自のメモリ・システム139a～139eを指定することが好ましい。例えば、前述の例のチェックサム標識161aにおいて、メモリ・ユニット155a内のチェックサム位置に相関するビット列265fは、メモリ・システム139a～139eにそれぞれ対応する5つのビット列を含むことが好ましい。メモリ・システム139aに対応するビットは、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)にチェックサムを格納している、ということを示すようアサートされることが好ましい。ここで、もしチェックサム位置がチェックサム値ではなく非チェックサム値を格納していたなら、前記のビットはデアサートされる、ということに注意して欲しい。

20

【0050】

さらに、メモリ・コントローラ152aが図9のステップ219でメモリ・コントローラ152bから送信された排他的論理和結果を含むチェックサム更新要求を受信するとき、メモリ・コントローラ152aは、前述のチェックサム標識161aのビット列265fを解析する際に、受信したチェックサム更新要求は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結果)の格納ではなくチェックサムを更新する目的のためである、ということを実行できる。従って、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結果を格納して他のメモリ・システム139b～139e内のチェックサムを更新しようと試みる代わりに、メモリ・コントローラ152aは、前述の技法に従ってチェックサム位置に格納されたチェックサムを更新する。これに関して、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結果とチェックサム位置に以前格納されたチェックサムとの排他的論理和をとり、そしてこの排他的論理和演算の結果をチェックサム位置に格納する。

30

【0051】

好ましい実施形態では、チェックサムを格納している記憶位置に相関するビット列265a～265fのどれもが、どのメモリ・システム139a～139eがアクティブであるかということも示す。これに関して、ビット列265a～265fは、メモリ・システム139a～139eが作動状態にあり、ビット列265a～265fに相関した記憶位置に格納されたチェックサムの更新が可能になるときに、メモリ・システム139a～139eは「アクティブ」である、と示す。従って、前述の例のチェックサム標識161aでは、チェックサム位置に相関したビット列265fの各ビットは、メモリ・システム139a～139eがすべて作動状態にある仮定して、最初はアサートされている。メモリ・システム139aに対応して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に相関した位置にチェックサムを格納している、ということを示し、メモリ・システム139b～139eに対応して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、ということを示す。もしメモリ・システム139cと139dがアクティブ

40

50

であり、メモリ・システム139bが非アクティブである(例えば作動不能)なら、メモリ・システム139a、139c、139dに対応するビットだけがアサートされるべきである。このような例のビット列265fを解析することによって、メモリ・システム139aがビット列265fに相關した位置にチェックサムを格納していること、メモリ・システム139cと139dがアクティブであること、及びメモリ・システム139bが非アクティブであることが明らかになるはずである。

【0052】

好ましい実施形態では、チェックサム更新要求を受信するメモリ・コントローラ152a~152eは、チェックサムを更新する前にまずチェックサムの記憶位置に関連したビット列265a~265fを解析する。前記のビット列265a~265fに基づいて、メモリ・コントローラ152a~152eは、そのチェックサム更新要求がアクティブなメモリ・システム139a~139eから送信されたかどうかを決定する。もし要求がアクティブなメモリ・システム139a~139eから送信されたものなら、前記メモリ・コントローラ152a~152eはチェックサムを更新する。しかし、もし要求がビット列265a~265fによって非アクティブと識別されたメモリ・システム139a~139eから送信されたものなら、前記メモリ・コントローラ152a~152eはその要求を無視し、チェックサムの更新をおこなわない。

【0053】

後に詳述する技法に従ってデータ回復を可能にするため、図9のステップ219でメモリ・コントローラ152bからメモリ・コントローラ152aへ送信されたチェックサム更新要求は、メモリ・システム139aのメモリ・システム識別子に加えて、送信しているメモリ・システム139bのメモリ・システム識別子をも含むことが好ましい。メモリ・システム139aを指定するメモリ・システム識別子に基づいて、メモリ・コントローラ152aはチェックサム更新要求に応答し、メモリ・システム139bを指定するメモリ・システム識別子に基づいて、メモリ・コントローラ152aは、メモリ・システム155bのメモリ・コントローラ152bがチェックサム更新要求を送信した、ということを知る。換言すれば、前述のメモリ・システム識別子は、どのメモリ・システム139a~139eがチェックサム更新要求を送信したかを指定する。

【0054】

図9のステップ225でチェックサム位置に格納されたチェックサム値を更新する前に、メモリ・コントローラ152aは、まず図9のステップ231のチェックサム標識161aを解析して、そのチェックサム更新要求がアクティブなメモリ・システムから送信されたものかどうかを決定する。より具体的には、書込み要求でメモリ・システム139bを指定するメモリ・システム識別子に基づいて、メモリ・コントローラ152aは、チェックサム位置に相關するビット列265f内のメモリ・システム139bに対応するビットを解析する。もしこのビットがアサートされていれば、メモリ・システム139bはアクティブとみなされる。従って、メモリ・コントローラ152aは、チェックサム更新要求内に含まれた排他的論理和結果に基づいてチェックサム位置内のチェックサムを更新する(図9のステップ225)。しかし、もし前記のビットがデアサートされていれば、メモリ・システム139bは非アクティブと見なされる。従って、メモリ・コントローラ152a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いてチェックサム位置のチェックサムの更新を行わない。換言すれば、メモリ・コントローラ152aは図9のステップ225をスキップする。

【0055】

メモリ・システム139a~139eの1つに格納されたデータ値が一旦失われると(例えば、そのデータ値を格納しているメモリ・システム139a~139eが故障すると)、消失データ値の記憶位置に関連したチェックサムを用いることと、他のアクティブなメモリ・システム139a~139e内に格納されたデータ値を用いることによって、消失データ値を回復することができる。これに関して、各メモリ・システム139a~139eはメモリ・システム139bに故障が起きるまで作動状態である、と仮定しよう。さらに、マッパー146は前述のメモリ・システム139b内の宛先位置に格納されたデータ値にアクセスするための取出し要求または格納要求を受け取る、と仮定しよう。メモリ・システム139bはもう作動状態でないので、受け

10

20

30

40

50

取った要求を満たすことは可能でない。さらに、宛先位置のデータ値は失われている。

【0056】

しかし、チェックサム位置(すなわち、宛先位置と同じメモリ・ユニット・オフセットを持つメモリ・ユニット155a内の位置)内に格納されたチェックサムによって消失データ値を回復することは可能である。より具体的には、チェックサム位置のチェックサムと、チェックサム・宛先位置と同じメモリ・ユニット・オフセットを持つ位置でアクティブなメモリ・システム139c~139eに現在格納されているデータとの排他的論理和をとることによって、消失データ値を回復できる。換言すれば、消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と、チェックサム・セットのチェックサムとの排他的論理和をとることによって、消失データ値を回復できる。さらに、後述するデータ回復技法を用いることによって、回復過程中に作動状態にあるメモリ・ユニット155a、155c~155eへのデータ格納の発生を妨げることなく消失データ値を回復できる。

10

【0057】

このデータ回復過程をより良く説明するため、マッパー146は上述のようにメモリ・システム139bが故障した後でメモリ・ユニット155bの宛先位置に格納されたデータ値にアクセスするための取出しまたは格納要求を受け取る、と仮定しよう。取出し要求または格納要求に応じて、マッパー146は取出し要求または格納要求のバス・アドレスをメモリ・ユニット・アドレスに変換する。そしてマッパー146は、読出し要求または書込み要求のどちらかをメモリ・システム139bに送信して、宛先位置に格納されたデータ値の取出しまたは宛先位置へのデータ値の格納のどちらかを要求する。メモリ・システム139bは故障しているので、マッパー146によって送信された要求は満たされない。マッパー146はメモリ・システム139bが要求に応答しないことを検知し、これに応答して、消失データ値(すなわち故障したメモリ・システム139b内の宛先位置に格納されていたデータ値)を回復するための過程が可能ならこのような回復過程を開始する。

20

【0058】

これに関して、マッパー146はこのようなデータ回復過程が可能かどうかを示すデータ回復標識171を含むことが好ましい。データ回復過程が可能かどうかを示すために使える方法には様々なものがあることに注意して欲しい。好ましい実施形態では、データ回復標識171は複数のビットを含み、これらは異なるメモリ・ユニット・オフセットそれぞれについて1ビット、従って異なるチェックサム・セットそれぞれについて1ビットである。もし特定のチェックサム・セット内のデータ値の回復が可能であるなら、そのチェックサム・セットに対応するデータ回復標識171内のビットはアサートされることが好ましい。しかし、特定のチェックサム・セット内のデータ値の回復が禁止されるなら、そのチェックサム・セットに対応するデータ回復標識171内のビットはデアサートされることが好ましい。

30

【0059】

好ましい実施形態では、データ回復標識171内の各ビットは最初、各チェックサム・セットについてデータ回復過程が可能であることを示すようアサートされる。しかし、後に詳述するように、チェックサム・セットのチェックサムはチェックサム・セットのデータ値の回復過程の最中に回復されたデータ値で置換される。従って、一旦データ回復過程が完了すると、回復されたデータ値のチェックサム・セットに対応するデータ回復標識171内のビットは、そのチェックサム・セット内のそれ以上のデータ回復を禁止するようデアサートされることが好ましい。

40

【0060】

さらに、マッパー146は、上述のようにメモリ・システム139bが要求されたデータ値にアクセスしないことを検知すると、要求されたデータ値(「消失データ値」と呼ばれる)のチェックサム・セットに対応するデータ回復標識171内のビットを解析する。もしこのビットがアサートされていれば、消失データ値を回復するための過程は可能にされ、データ回復過程は継続する。しかし、もしこのビットがデアサートされていれば、消失データ値を回復するための過程は禁止され、データ回復過程は終了される。現在の例では、消失デー

50

タ値を回復するためのデータ回復過程は可能にされ、従って、データ回復過程は継続する、と仮定しよう。

【 0 0 6 1 】

データ回復過程の実施において、マッパー146は、他のメモリ・システム139a、139c～139eのうちどれが消失データ値のチェックサムを格納しているかを決定する(図11のステップ272)。従って、マッパー146は、メモリ・システム139a～139eのうちどれが消失データ値の回復に使えるチェックサムを格納しているかを示すチェックサム標識172を維持することが好ましい。現在の例では、消失データ値の記憶位置に関連するチェックサムはシステム139aに格納されているので、標識172はメモリ・システム139aを指定すべきである。消失データ値の記憶位置に関連するチェックサムを格納している適切なメモリ・システム139a～139eを指定するのに使える方法は様々なものがある。

10

【 0 0 6 2 】

例えば、各メモリ・ユニット・オフセットについて、チェックサム標識172は、メモリ・システム139a～139eに格納されたチェックサム標識161a～161eに似た、各メモリ・システム139a～139eに対するビットを持つビット列(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・オフセットに相関し、消失データ値の記憶位置と同じメモリ・ユニット・オフセットに相関するビット列は、どのメモリ・システム139a～139eが消失データ値のチェックサムを格納しているかを示すのに使われることが好ましい。これに関して、消失データ値のチェックサムを格納しているメモリ・システム139a～139eに対応するビットはアサートされていることが好ましく、残りのビットはデアサートされていることが好ましい。従って、どのような消失データ値についても、マッパー146は、チェックサム標識172を解析することによって、どのメモリ・システム139a～139eが消失データ値の記憶位置に関連したチェックサムを格納しているかを決定できる。このチェックサムは、消失データ値の位置と同じメモリ・ユニットオフセットを持つ記憶位置で指定されたメモリ・システム139a～139eに格納されているべきである。

20

【 0 0 6 3 】

従って、消失データ値のチェックサムがメモリ・ユニット155aの記憶位置(「チェックサム位置」)に格納されている前述の例では、マッパー146はチェックサム標識172を解析し(図11のステップ272)、メモリ・システム139aが消失データ値の記憶位置に関連するチェックサムを格納しているということを決定する。これに回答して、マッパー146は、消失データ値を再構築するための要求をメモリ・システム139aに送信することが好ましい(図11のステップ275)。この再構築要求に含まれるのは、宛先(すなわち、消失データ値が格納されていたメモリ・ユニット155bの位置)のメモリ・ユニット・オフセットと、消失データ値を格納していたメモリ・システム139bを指定するメモリ・システム識別子である。

30

【 0 0 6 4 】

これに回答して、メモリ・コントローラ152aはチェックサム標識161aを解析して、他のメモリ・システム139c～139eのうちどれがチェックサム位置に関してアクティブかを決定する(図11のステップ277)。現在の例では、メモリ・コントローラ152aが再構築要求を受け取った時点でメモリ・システム139c～139eのそれぞれがアクティブである、と仮定する。好ましい実施形態について前述した技法に従って、アクティブなメモリ・システム139c～139eの指定は、チェックサム位置に相関するビット列265fをチェックサム標識161a内で解析することによって達成される。ここで、チェックサム位置はマッパー146から受け取った再構築要求に含まれるメモリ・ユニット・オフセットに相関するメモリ・ユニット155a内の位置である、ということに注意して欲しい。

40

【 0 0 6 5 】

他のアクティブなメモリ・システム139c～139eを識別した後、メモリ・コントローラ152aは、図11のステップ277に示した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でアクティブなメモリ・システム139c～139e内のデータ値の取出しを要求する。換言すれば、メモリ・コントローラ152aは、作動状態にあるメモリ・システム13

50

9c~139eのそれぞれから、消失データ値の回復に使用中の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出しを要求する。

【0066】

より具体的には、メモリ・コントローラ152aはメモリ・コントローラ152cに再構築用読出し要求(read-for-rebuild request)を送信し、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の両方に相関するメモリ・ユニット・オフセット)に相関する位置でメモリ・ユニット155cに格納されているデータ値を取り出すようメモリ・コントローラ152cに要求する。メモリ・コントローラ152aはメモリ・コントローラ152dにも再構築用読出し要求を送信し、再構築要求に含まれるものと同じメモリ・ユニット・オフセットに相関する位置でメモリ・ユニット155dに格納されているデータ値を取り出すようメモリ・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はメモリ・コントローラ152eにも再構築用読出し要求を送信し、再構築要求に含まれるものと同じメモリ・ユニット・オフセットに相関する位置でメモリ・ユニット155eに格納されているデータ値を取り出すようメモリ・コントローラ152eに要求する。各再構築用読出し要求は、パラレルでもシリアルでも送信できる。

【0067】

ここで、データ回復過程に送信される「再構築用読出し要求」は読出し要求であって、それが取り出すメモリ・コントローラ152a~152eに指示することは、メモリ・ユニット155a~155eの1つに現在格納されているデータ値を取り出すこと(プロセッサキャッシュからさらに新しいコピーを取り出す必要はない)と、取り出したデータ値(以後「再構築データ値」と呼ぶ)を再構築データ値に基づくデータ値の再構築に責任を持つ構成要素に送信することである、ということに注意して欲しい。好ましい実施形態では、再構築用読出し要求を送信するメモリ・コントローラ152aが、このような再構築に責任を持つ構成要素である。従って、現在の例では、メモリ・コントローラ152c~152eのそれぞれが、前記再構築用読出し要求の1つに応答してメモリ・コントローラ152aにそれぞれの再構築データ値を送信する。好ましくは、各メモリ・コントローラ152c~152eは、受け取った再構築用読出し要求に応答して、再構築データ値と共に、メモリ・システム識別子及び、その再構築データ値が以前送信された再構築用読出し要求に対するものであることを示すデータを送信することが好ましい。従って、メモリ・コントローラ152aが再構築データ値を受け取る時、メモリ・コントローラ152aは、(1)受け取ったデータ値は消失データ値の回復のために使われることと(2)受け取ったデータ値を送信したのはどのメモリ・システム139c~139eか、を決定できる。

【0068】

ここで、特定の記憶位置から取り出す再構築用読出し要求は、以前に受け取ったその特定の記憶位置に書き込む書込み要求のすべてがそれぞれの関連するチェックサム更新と共に完了するまでは、受け付けられまたは満足されるべきでない、ということに注意して欲しい。同様に、特定の記憶位置への後からの書込み要求は、再構築要求が完了してチェックサム・コントローラが(チェックサムは消失データ値の回復に使われているので)特定の記憶位置に関連したメモリ・コントローラからのチェックサム更新を禁止するまで、チェックサム更新を起こすべきではない。

【0069】

従って、再構築データ値のそれぞれについて、メモリ・コントローラ152aは、図11のステップ281~284に示すように、チェックサム標識161aを更新し(ステップ282)、再構築データ値とチェックサム位置に格納された値との排他的論理和をとる(ステップ283)。これに関して、メモリ・コントローラ152aは、ステップ282において、チェックサム位置に相関するビット列265fをチェックサム標識161a内で更新することによって、再構築データ値を送信したメモリ・コントローラ152cがビット列265fによってアクティブと表示されないようにする。好ましい実施形態では、これは、受け取った再構築データ値を送信したメモリ・システム139c~139eに対応して前記のビット列265f内でアサートすることによって達成される。

【 0 0 7 0 】

従って、メモリ・コントローラ152aは、メモリ・コントローラ152cから送信された再構築データ値を受け取ると、メモリ・システム139cに対応するビットを前記のビット列265f内でデアサートする。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152aは再構築データ値を送信したメモリ・コントローラ152cからのチェックサム更新要求を受け入れるべきであり、従って、受け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すべきである。一旦前記ビットがデアサートされると、図11のステップ283の実行において、メモリ・コントローラ152aは再構築データ値とチェックサム位置に現在格納されているデータ値との排他的論理和をとる。メモリ・コントローラ152aは、次に、この排他的論理和演算の結果をチェックサム位置に格納してステップ283を完了する。

10

【 0 0 7 1 】

さらに、メモリ・コントローラ152aは、メモリ・コントローラ152dから送信された再構築データ値を受け取ると、ステップ282でメモリ・システム139dに対応するビットを前記のビット列265f内でデアサートする。一旦このビットがデアサートされると、メモリ・コントローラ152aはステップ283で再構築データ値とチェックサム位置に現在格納されているデータ値との排他的論理和をとる。メモリ・コントローラ152aは、次に、この排他的論理和演算の結果をチェックサム位置に格納してステップ283を完了する。

【 0 0 7 2 】

さらに、メモリ・コントローラ152aは、メモリ・コントローラ152eから再構築データ値を受け取ると、ステップ282でメモリ・システム139eに対応するビットを前記のビット列265f内でデアサートする。一旦このビットがデアサートされると、メモリ・コントローラ152aはステップ283で再構築データ値とチェックサム位置に現在格納されているデータ値との排他的論理和をとる。メモリ・コントローラ152aは、次に、この排他的論理和演算の結果をチェックサム位置に格納してステップ283を完了する。

20

【 0 0 7 3 】

前述の各ステップを完了した後、チェックサム位置に格納されたデータ値は消失データ値であるはずである。従って、このデータ値を再構築要求に応答して回復されたデータ値としてマッパー146に送信しても良いし、必要に応じて他のことに用いても良い。好ましい実施形態では、回復されたデータ値はチェックサム位置に格納される。従って、データ回復過程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で効果的に置き換える。マッパー146は、以前消失データ値の位置に変換されたバス・アドレスが今度はチェックサム位置に変換されるように更新されるべきである。もはやチェックサム・セットに関連するチェックサムはないので、チェックサムからデータ値を回復しようとするすべての試みは禁止されるべきである。これは、チェックサム位置のメモリ・ユニット・オフセットに対応するビットをデータ回復標識171内でデアサートすることによって達成できる。システムマネージャ136からの指令は、前記の更新をマッピングとデータ回復標識171に対して実行するようマッパー146に指示してもよい。

30

【 0 0 7 4 】

ここで注意すべきことは、一旦消失データ値の回復が完了すると、各処理ユニット17a~17eは、チェックサム位置に格納された消失データ値より新しいバージョンの消失データ値を持っているか否か決定するよう問われてもよい、ということである。もし、処理ユニット17a~17eのどれかがより新しいバージョンを持っていれば、チェックサム位置の値をそのより新しいバージョンで上書きしてよい。

40

【 0 0 7 5 】

さらに注意すべきことは、図11のステップ282において、受け取った再構築データ値を送信したメモリ・コントローラ152a~152eのアクティブ状態を変更するステップは、図11に示したデータ回復過程中的操作可能なメモリ・システム139a、139c~139eへのデータ格納の発生を可能にするため重要である、ということである。これに関して、前記のステップはデータ回復過程中にチェックサム更新がエラーを起こすことを防ぐ。

【 0 0 7 6 】

50

特に、データ回復過程中的エラーを防ぐため、ある特定のメモリ・コントローラ152a~152eが再構築用読出し要求(すなわち、図11のステップ277で送信された要求)に回答する前に、図9のステップ219でその特定のメモリ・コントローラ152a~152eによって送信された各チェックサム更新要求は、消失データ値を再構築しているメモリ・コントローラ152a~152eによって満たされるべきである。このようなチェックサム更新要求を満たすことは、回復過程で使われるチェックサムと、データ回復中にその特定のメモリ・コントローラ152a~152eによって送信される再構築データ値との整合性を保証する。

【0077】

しかし、その特定のメモリ・コントローラ152a~152eが再構築用読出し要求(すなわち図11のステップ277で送信された要求)に反応した後に、図9のステップ219でその特定のメモリ・コントローラ152a~152eによって送信された各チェックサム更新要求は、消失データ値を再構築しているメモリ・コントローラ152a~152eによって満たされるべきでない。このような要求を満たすことは、おそらくデータ回復過程中的エラーを起こすだろう。さらに、図11のステップ282と図9のステップ231を実行することによって、このような要求を満たすことは防止され、従って、データ回復中にメモリ・システム139a~139eへのデータ格納が許されるときにデータ回復エラーの可能性が防止される。

10

【0078】

ステップ231と282にこのような要求をすべて確実に防止させるためには、メモリ・コントローラ152a~152eは、再構築データで回答する前に、再構築用読出し要求によって要求されたデータに対する以前のチェックサム更新要求がすべて完了しているよう保証すべきである。このデータに対する新たな要求を処理する前に、再構築用読出し要求に回答するメモリ・コントローラ152a~152eは、その再構築用読出し要求が完了していることと、その再構築データを受け取るチェックサム・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ためのチェックサム更新要求をそれ以上受け付けないことを保証すべきである。

20

【0079】

前述の過程によれば、特定のチェックサム・セット内にあって故障したメモリ・システム139bに格納されているデータ値は、そのチェックサム・セットのチェックサムとそのチェックサム・セットのその他の非チェックサムデータ値を用いることによって回復できる。好ましい実施形態では、回復されたデータ値は以前にチェックサム・セットのチェックサムを格納していたチェックサム位置に格納される。従って、前記の例では、回復されたデータ値は、メモリ・ユニット155b内の宛先位置と同じメモリ・ユニット・オフセットによって指定された位置(すなわち消失データ値の位置)でメモリ・ユニット155a内に、チェックサム・セットの非チェックサム値を格納しているメモリ・ユニット155c~155e内の位置に格納される。

30

【0080】

しかし、チェックサム・セットのチェックサムは回復されたデータ値と置き換えられるので、チェックサム・セットに関連するチェックサムはもうない。従って、もし他のメモリ・システム139a、139c~139eが故障すると、故障した他のメモリ・システム139a、139c~139eのチェックサム・セット内のデータ値は、最初の消失データ値の回復後にチェックサム・セットをバックアップする追加手段をとらない限り、回復できないだろう。以下は、好ましい実施形態において追加のメモリ・システム139f(図12)を用いてチェックサム・セットをバックアップする方法についての記述である。

40

【0081】

図7、図13で示すように、追加のメモリ・システム139fは他のメモリ・システム139a~139eと同様に構成される。これに関して、メモリ・システム139fは、データを格納するための複数の記憶位置を持つメモリ・ユニット155fを含む。各記憶位置は異なるメモリ・ユニット・オフセットによって識別される。さらに、メモリ・コントローラ152fは、メモリ・コントローラ152a~152eについて以前に記述したものと同一技法に従ってメモリ・ユニット155f内のデータにアクセス(すなわち格納と取出し)するよう構成される。従って、図

50

14で示すように、メモリ・システム139fは、チェックサム標識161a~161eと同様のチェックサム標識161fを含み、これによってメモリ・コントローラ152fは前述の技法に従ってチェックサム更新を開始できる。従って、好ましい実施形態では、チェックサム標識161fは複数のビット列265a~265fを含み、各ビット列はアクティブなメモリ・システム139a~139fのそれぞれに対して少なくとも1ビットを持っている。

【0082】

さらに、メモリ・システム139fは、故障したメモリ・システム139bからの消失データ値の回復の前及び/または最中にシステム115によってデータ格納に使われたり、システム115の内部に含まれたりしてもよい。代替的に、メモリ・システム139fは、メモリ・システム139bの故障後及び/または前述のデータ回復過程後にシステム115に追加されていてもよい。いずれにしても、メモリ・ユニット155fは、チェックサム・セットの他のデータ値を格納している位置と同じメモリ・ユニットオフセットによって指定され有効なデータ値の格納に使われていない記憶位置を持つべきである。さらに、メモリ・システム139bとメモリ・システム139fの両方を含む実施形態では、チェックサム標識161a~161f、171、172内のビット列のそれぞれは、追加のメモリ・システム139fに対応する追加のビットを含むべきであり、これによってメモリ・システム139fはメモリ・システム139a~139eについて述べたのと同じ技法に従って用いられることができる。代替的に、故障したメモリ・システム139bの代わりにメモリ・システム139fが実装されるとき、以前メモリ・システム139bに対応していたビット列161a、161c~161f、171、172のビットをメモリ・システム139fに対応するために使うことができ、これによって追加のビットの必要性がなくなる。

【0083】

好ましい実施形態では、前記の追加のメモリ・ユニット155fの記憶位置は、以前に回復されたデータ値を含むチェックサム・セットのチェックサムの構築と保持に使われる。前述のチェックサムを格納するためのメモリ・ユニット155fのこの記憶位置を以後「新しいチェックサム位置」と呼び、以前「チェックサム位置」と呼んでいた回復されたデータ値を格納しているメモリ・システム139aの記憶位置を「古いチェックサム位置」と呼ぶ。

【0084】

従って、一旦前述のデータ回復過程が完了すれば、古いチェックサム位置は回復されたデータ値を格納しているはずであり、その値はチェックサム・セットの非チェックサムデータ値の1つである。さらに、古いチェックサム位置と新しいチェックサム位置は同じメモリ・ユニット・オフセットを持つはずであり、それはチェックサム・セットの他の非チェックサムデータ値を格納しているメモリ・ユニット139c~139e内の記憶位置のそれぞれを指定する同じメモリ・ユニット・オフセットであるはずである。

【0085】

最初、図15のステップ305で、チェックサム中立情報が新しいチェックサム位置に格納される。システムマネージャ136は新しいチェックサム位置に値ゼロを格納させることによって好ましい実施形態において前記を達成することができる。さらに、メモリ・システム139fに対応するビットを除いて、新しいチェックサム位置に相関するチェックサム標識161fの各ビットはデアサートされており、図15のステップ308において、新しいチェックサム位置の更新を求めるチェックサム更新要求を受け取ってもメモリ・コントローラ152fはそれを無視すべきである、ということを示す。

【0086】

ステップ311で、システムマネージャ136は、チェックサム・セットからデータ値を回復しようとする試みが禁止されることを保証する。好ましい実施形態では、システムマネージャ136は、このような回復の試みを禁止する指令をマッパー146に送信することによって前述したことを達成する。これに回答して、マッパー146はデータ回復標識171においてチェックサム・セットに対応するビットをデアサートし、それによりチェックサム・セットのデータ値を回復する試みを禁止する。

【0087】

次に、ステップ315で、システムマネージャ136は、マッパー146のチェックサム標識172へ

10

20

30

40

50

の更新を起こすことによって、チェックサム標識172がメモリ・システム139fをチェックサム・セットのチェックサムを含むものとして識別するようにする。好ましい実施形態では、前述したことは、古いチェックサム位置と新しいチェックサム位置のメモリ・ユニットオフセットに相關したビット列をチェックサム標識172内で更新するようマップー146に指令することによって達成される。換言すれば、前述したことは、チェックサム・セットに相關したビット列をチェックサム標識172内で更新するようマップー146に指令することによって達成される。より具体的には、システムマネージャ136がマップー146に、新しいチェックサム位置のメモリ・システム139fに対応するビットはアサートされ残りのビットはデアサートされるように前記のビット列を更新するよう指示する。ステップ315を実施することによって、チェックサム・セットのデータ値の1つを回復しようとする将来の試みは、そのチェックサム・セットに対するチェックサムとして新しいチェックサム位置に格納されたデータ値を用いるべきである。

10

【0088】

ステップ321で、システムマネージャ136は操作可能なメモリ・コントローラ152a、152c～152e(新しいチェックサム・メモリ・コントローラ152f以外)のそれぞれに各自のチェックサム標識161a、161c～161eを更新するよう指令する。これに関して、各メモリ・コントローラ152a、152c～152eは、チェックサム標識161a、161c～161eがメモリ・システム139fをチェックサム・セットのチェックサムを格納するものとして識別するよう各チェックサム標識161a、161c～161eを更新する。特に、チェックサム・セットに相關する各チェックサム標識161a、161c～161e内のビット列265fは、メモリ・システム139fに対応するビットがアサートされ残りのビットがデアサートされるよう更新される。

20

【0089】

従って、ステップ321で更新されたメモリ・コントローラ152a、152c～152eの1つが、チェックサム・セットのデータ値を上書きする(すなわち、新しいチェックサム位置と同じメモリ・ユニットオフセットによって指定された記憶位置に格納されたデータ値を上書きする)ようメモリ・コントローラ152a、152c～152eに指示する書込み要求を受け取ると、その1つのメモリ・コントローラ152a、152c～152eは前述の技法に従ってメモリ・システム139fへのチェックサム更新要求を送信すべきである。このチェックサム更新要求は、その要求に応じて更新されるべき位置として新しいチェックサム位置を指定すべきであり、どのメモリ・コントローラ152a、152c～152eがその要求を送信したのが特定すべきである。しかし、チェックサム構築過程のこの時点では、このようなチェックサム更新要求はすべて無視されるべきである。なぜなら、チェックサム標識161fは、ステップ308を通じて、各メモリ・コントローラ152a、152c～152eは新しいチェックサム位置の更新を禁止されている、ということを示しているからである。

30

【0090】

ステップ321の後、システムマネージャ136はチェックサム・コントローラ152fに、新しいチェックサム位置の値を更新してその値がチェックサム・セットのチェックサムを表すようにするよう指示する。これに回答して、チェックサム・コントローラ152fは、ステップ324で、他の操作可能なメモリ・コントローラ152a、152c～152eのそれぞれに再構築用読出し要求を送信する。この再構築用読出し要求は並列に送信してもよいし、直列に送信してもよい。

40

【0091】

ステップ324で送信される各再構築用読出し要求は、図11のステップ277で送信される再構築用読出し要求と同じでよい。ただし、この要求が取出し中のメモリ・コントローラに指示することは、関連するチェックサム・セットの各非チェックサムデータ値と矛盾しないチェックサムの構築の役割を持つ構成要素に、取り出したデータ値を送信する点が異なる。この例では、メモリ・コントローラ152fがチェックサムの構築の役割を持つ。従って、メモリ・コントローラ152a、152c～152eのそれぞれが、前記の再構築用読出し要求に応じて、各自の再構築データ値をメモリ・コントローラ152fに送信する。

【0092】

50

ステップ277で送信された再構築用読出し要求と同様に、各メモリ・コントローラ152a、152c～152eは、受け取った再構築用読出し要求に応じて、各自のメモリ・システム識別子と共に再構築データ値(すなわち受け取った再構築用読出し要求に応じて取り出したデータ値)と、その再構築データ値が以前に送信された再構築用読出し要求に応えるものであることを示すデータと、を送信する。従って、メモリ・コントローラ152fが再構築データ値を受け取る時、メモリ・コントローラ152fは、(1)受け取ったデータ値はチェックサム・セットのチェックサムの再構築に使われるべきものであることと、(2)受け取ったデータを送信したのはどのメモリ・システム139a、139c～139eかということを決めることができる。

【0093】

ここで、特定の記憶位置から取り出すための再構築用読出し要求は、その特定の記憶位置への書込みについて以前に受け取った書込み要求がすべて完了して、そのような書込み要求に関連するチェックサム更新要求が新しいチェックサム位置に更新を起こさないことが保証されるまで、受け付けたりまたは満足されるべきでない、ということに注意して欲しい。前述したことは、以前の各書込み要求が完了され、関連したチェックサム更新要求がメモリ・コントローラ152fに受け取られ無視されるまで、前記の再構築用読出し要求を満たさないことによって達成できる。同様に、その特定の記憶位置への以後の書込み要求は、再構築用読出し要求が完了してチェックサム・コントローラ152fが新しいチェックサム位置についてその特定の記憶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を可能にするまで、チェックサム更新を起こすべきでない。

【0094】

受け取った再構築データ値に応じて、メモリ・コントローラ152fは、図14のステップ331～334に示すように、チェックサム標識161fを更新し(ステップ332)、その再構築データ値と新しいチェックサム位置に格納された値の排他的論理和をとる(ステップ333)。これに関して、メモリ・コントローラ152fは、ステップ332で、新しいチェックサム位置に相關するビット列265fをチェックサム標識161f内で更新することによって、再構築データ値を送信したメモリ・コントローラ152a、152c～152eがビット列265fによってアクティブと示されるようにする。好ましい実施形態では、これは受け取った再構築データ値を送信したメモリ・システム139a、139c～139eに対応するビットを前記のビット列265f内でアサートすることによって達成される。

【0095】

従って、メモリ・コントローラ152fがメモリ・コントローラ152aから送信された再構築データ値を受け取る時、メモリ・コントローラ152fは前記ビット列265f内でメモリ・システム139aに対応するビットをアサートする。このビットがアサートされるまで、チェックサム・コントローラ152fは再構築データ値を送信したメモリ・コントローラ152aからのチェックサム更新要求を無視すべきである。一旦前記ビットがアサートされると、メモリ・コントローラ152fは、図15のステップ333の実行において、再構築データ値と新しいチェックサム位置に現在格納されているデータ値との排他的論理和をとる。そしてメモリ・コントローラ152fは、この排他的論理和演算の結果を新しいチェックサム位置に格納してステップ333を完了する。

【0096】

さらに、メモリ・コントローラ152fがメモリ・コントローラ152cから送信された再構築データ値を受け取る時、メモリ・コントローラ152fは、ステップ332において、前記ビット列265f内で、メモリ・システム139cに対応するビットをアサートする。一旦このビットがアサートされると、メモリ・コントローラ152fは、ステップ333で、再構築データ値と新しいチェックサム位置に現在格納されているデータ値との排他的論理和をとる。そしてメモリ・コントローラ152fは、この排他的論理和演算の結果を新しいチェックサム位置に格納して、ステップ333を完了する。

【0097】

さらに、メモリ・コントローラ152fがメモリ・コントローラ152dから再構築データ値を受け取る時、メモリ・コントローラ152fは、ステップ332において、前記ビット列265f内

10

20

30

40

50

で、メモリ・システム139dに対応するビットをアサートする。一旦このビットがアサートされると、メモリ・コントローラ152fは、ステップ333で、再構築データ値と新しいチェックサム位置に現在格納されているデータ値との排他的論理和をとる。そしてメモリ・コントローラ152fは、この排他的論理和演算の結果を新しいチェックサム位置に格納して、ステップ333を完了する。

【0098】

最後に、メモリ・コントローラ152fがメモリ・コントローラ152eから再構築データ値を受け取るとき、メモリ・コントローラ152fは、ステップ332において、前記ビット列265f内で、メモリ・システム139eに対応するビットをアサートする。一旦このビットがアサートされると、メモリ・コントローラ152fは、ステップ333で、再構築データ値と新しいチェックサム位置に現在格納されているデータ値との排他的論理和をとる。そしてメモリ・コントローラ152fは、この排他的論理和演算の結果を新しいチェックサム位置に格納して、ステップ333を完了する。

10

【0099】

前述のステップのそれぞれを完了した後、新しいチェックサム位置に格納されたデータ値は、チェックサム・セットのチェックサムを表すはずである。換言すれば、新しいチェックサム位置に格納されたデータ値は、新しいチェックサム位置と同じメモリ・ユニットオフセットを持つ位置でメモリ・ユニット155a、155c～155eに格納された非チェックサムデータ値のチェックサムを表すはずである。さらに、メモリ・コントローラ152a、152c～152eのどれかがチェックサム・セットのデータ値の1つを上書きする書込み要求に回答してチェックサム更新要求を送信するとき、メモリ・コントローラ152fは、チェックサムが現在のチェックサム・セットの非チェックサムデータ値と整合するように、新しいチェックサム位置のチェックサムを更新する。従って、もしメモリ・システム139a、139c～139eのどれか1つが故障すると、チェックサム・セット内にある消失データ値は、消失データ値の回復について前述した技法に従って新しいチェックサム位置のチェックサムを用いることによって、回復することができる。

20

【0100】

従って、ステップ339で、チェックサム・セットのデータ値の1つを回復しようとする試みは可能とされる。これに関して、システムマネージャ136は、このようなデータ回復の試みを可能にするためデータ回復標識171を更新するようマップー146に指示する指令をマップー146に送信することができる。このような指令に回答して、マップー146はチェックサム・セットに対応するデータ回復標識171内のビットをアサートすべきである。

30

【0101】

上述した本発明の実施形態、特に「好ましい」実施形態は、単に可能な実施例であり、本発明の原理の明確な理解のために述べたものである。本発明の精神と原理から大きく離れることなく、本発明の上述の実施形態に多くの変形や変更を加えることができる。

【0102】

本発明には例として以下の実施形態が含まれる。

【0103】

(1) それぞれが複数の記憶位置を持つ複数のメモリ・ユニット(155a～155f)と、それぞれが前記メモリ・ユニット(155a～155f)のそれぞれ1つの内部の記憶位置にアクセスするよう構成されている複数のメモリ・コントローラ(152a～152f)であって、これらメモリ・コントローラの1つが前記記憶位置の1つにチェックサムを構築するよう構成されている、複数のメモリ・コントローラと、前記複数のメモリ・コントローラ(152a～152f)のうち、前記1つの記憶位置の更新が可能であるものを示すよう構成されている標識(161a～161f)と、を含み、前記1つのメモリ・コントローラ(152a～152f)が、前記チェックサム値の構築の際に、他の各メモリ・コントローラ(152a～152f)は前記1つの記憶位置の更新を禁止されていることを示すよう前記標識(161a～161f)を設定するステップと、

40

50

前記設定ステップに続いて、前記他の各メモリ・コントローラ(152a~152f)に再構築用読出し要求を送信するステップと、
前記再構築用読出し要求に応答して、前記メモリ・ユニット(155a~155f)から取り出された再構築値を受信するステップと、
前記1つの記憶位置を前記再構築値でそれぞれ更新するステップと、
前記各再構築値に対し前記各再構築値の受信に応答して、前記各再構築値を送信したメモリ・コントローラ(152a~152f)は前記1つの記憶位置の更新が可能であることを示すよう前記標識(161a~161f)を変更するステップと、
の各ステップを実行するよう構成されている、チェックサムを構築するためのシステム(115)。

10

【0104】

(2) 前記各再構築値は、前記1つの記憶位置と同じメモリ・ユニット・オフセットを持つ各記憶位置に格納される、上記(1)に記載のシステム(115)。

【0105】

(3) 前記メモリ・コントローラ(152a~152f)のうち別の1つは、書込み要求を受信して、前記書込み要求に関連したデータ値を前記書込み要求によって識別される記憶位置に書き込むよう構成されており、

前記他のメモリ・コントローラ(152a~152f)は、前記データ値と前記書込み要求によって識別される前記記憶位置に以前格納されたデータ値とを組み合わせることによって複合値を形成するようさらに構成されており、

20

前記他のメモリ・コントローラ(152a~152f)は、前記複合値を前記1つのメモリ・コントローラ(152a~152f)に送信するようさらに構成されており、

前記1つのメモリ・コントローラ(152a~152f)は、該1つのメモリ・コントローラが前記複合値を受け取る時に前記他のメモリ・コントローラ(152a~152f)が前記1つの記憶位置を更新可能であると前記標識が示す場合に限り、前記複合値に基づいてかつ前記複合値に応答して、前記1つの記憶位置を更新するよう構成されている、

上記(1)に記載のシステム(115)。

【0106】

(4) それぞれが複数の記憶位置を持つ複数のメモリ・ユニット(155a~155f)を提供するステップと、

30

複数のメモリ・コントローラ(152a~152f)を提供するステップと、

前記メモリ・コントローラ(152a~152f)のうち、前記メモリ・ユニット(155a~155f)の1つの中の前記記憶位置の1つの更新が可能であることを示す標識を維持するステップと、

前記各メモリ・コントローラ(152a~152f)が前記1つの記憶位置の更新を禁止されていることを示すよう前記標識(161a~161f)を設定するステップと、

前記設定ステップに続いて、前記各メモリ・コントローラ(152a~152f)に再構築用読出し要求を送信するステップと、

前記再構築用読出し要求に応答して、前記メモリ・コントローラ(152a~152f)によって前記メモリ・ユニット(155a~155f)から取り出された再構築値を受信するステップと、

前記1つの記憶位置を前記各再構築値で更新するステップと、

40

前記各再構築値に対し前記各再構築値の受信に応答して、前記各再構築値を送信したメモリ・コントローラ(152a~152f)は前記1つの記憶位置の更新が可能であることを示すよう前記標識(161a~161f)を変更するステップと、

を含む、チェックサムを構築する方法。

【0107】

(5) 前記再構築用読出し要求の1つに応答して記憶位置から前記再構築値の1つを取り出すステップと、

前記1つの記憶位置に格納されたデータ値を用いて、前記取出しステップによりアクセスした前記記憶位置内に格納されたデータ値を回復するステップと、

をさらに含む上記(4)に記載の方法。

50

【 0 1 0 8 】

(6) 前記再構築用読出し要求に応答して、前記 1 つの記憶位置と同じメモリ・ユニット・オフセットを持つ記憶位置から前記各再構築値を取り出すステップをさらに含む、上記 (4) に記載の方法。

【 0 1 0 9 】

(7) 前記メモリ・ユニット (155a ~ 155f) 内の前記他の位置であるもう 1 つの記憶位置にデータ値を格納するステップと、

前記データ値と前記他の記憶位置に以前格納されたデータ値とを組み合わせる複合値を形成するステップと、

前記複合値に基づいて、前記 1 つの記憶位置を更新する要求を前記メモリ・コントローラ (152a ~ 152f) の 1 つから送信するステップと、

前記要求に応答して前記標識 (161a ~ 161f) を解析するステップと、

前記 1 つのメモリコントローラ (152a ~ 152f) が前記 1 つの記憶位置を更新可能であると前記標識 (161a ~ 161f) が前記解析ステップの間に示す場合に限り、前記要求に応答して前記 1 つの記憶位置を更新するステップと、

を含む上記 (4) に記載の方法。

【 0 1 1 0 】

(8) もう 1 つの標識 (161a ~ 161f)、前記メモリ・コントローラ (152a ~ 152f) の 1 つを識別する前記他の標識 (161a ~ 161f) を維持するステップと、

前記設定ステップに続いて、前記メモリ・コントローラ (152a ~ 152f) のうち前記 1 つの記憶位置にアクセスできるものを識別するよう前記他の標識 (161a ~ 161f) を更新するステップと、

前記メモリ・コントローラ (152a ~ 152f) の他の 1 つで、データ値を格納する書込み要求を受信するステップと、

前記書込み要求受信ステップに応答して、前記他の標識 (161a ~ 161f) を解析するステップと、

前記解析ステップに基づいて、前記 1 つの記憶位置を更新する要求を送信するステップと、

を含む上記 (4) に記載の方法。

【 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 で示した処理ユニットをさらに詳しく示すブロック図である。

【 図 3 】 図 1 で示したメモリ・システムをさらに詳しく示すブロック図である。

【 図 4 】 本発明の原理を実施するのに使用できるコンピュータ・システムを示すブロック図である。

【 図 5 】 図 4 で示した処理システムをさらに詳しく示すブロック図である。

【 図 6 】 図 5 で示した処理ユニットをさらに詳しく示すブロック図である。

【 図 7 】 図 5 で示したメモリ・システムをさらに詳しく示すブロック図である。

【 図 8 】 図 7 で示したメモリ・ユニットをさらに詳しく示すブロック図である。

【 図 9 】 図 5 のメモリ・システムの 1 つにデータ値を格納する際の図 4 のコンピュータ・システムのアーキテクチャと機能を示す流れ図である。

【 図 10 】 図 7 で示したチェックサム標識をさらに詳しく示すブロック図である。

【 図 11 】 消失データ値を回復する際の図 4 のコンピュータ・システムのアーキテクチャと機能を示す流れ図である。

【 図 12 】 図 5 で示した処理システムのもう 1 つの実施形態を示すブロック図である。

【 図 13 】 図 12 で示したメモリ・システムをさらに詳しく示すブロック図である。

【 図 14 】 図 13 で示したチェックサム標識をさらに詳しく示すブロック図である。

【 図 15 】 チェックサムを構築する際の図 4 のコンピュータ・システムのアーキテクチャと機能を示す流れ図である。

【 符号の説明 】

10

20

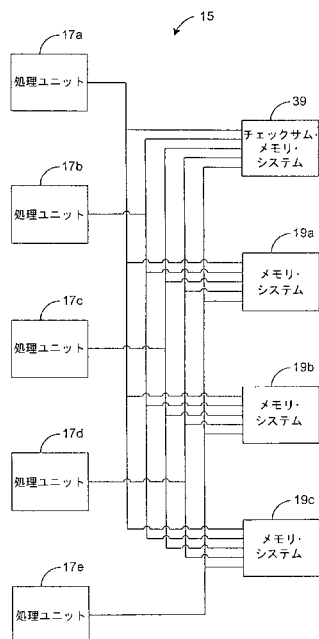
30

40

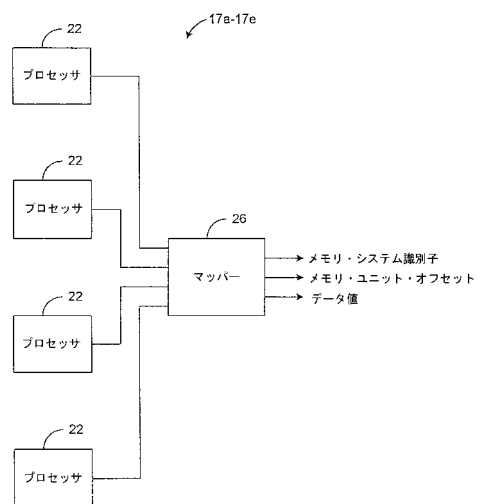
50

115	処理システム
137a ~ 137e	処理ユニット
139a ~ 139f	メモリ・システム
142	プロセッサ
146	マッパー
152a ~ 152f	メモリ・コントローラ
155a ~ 155f	メモリ・ユニット
161a ~ 161f	チェックサム 標識
171	データ回復 標識
172	チェックサム 標識
265a ~ 265f	ビット列

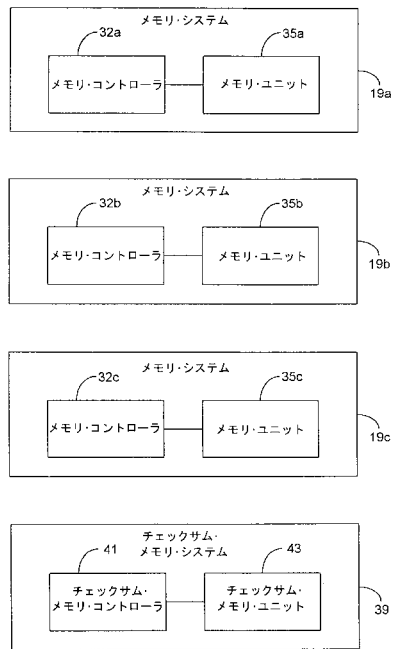
【図 1】



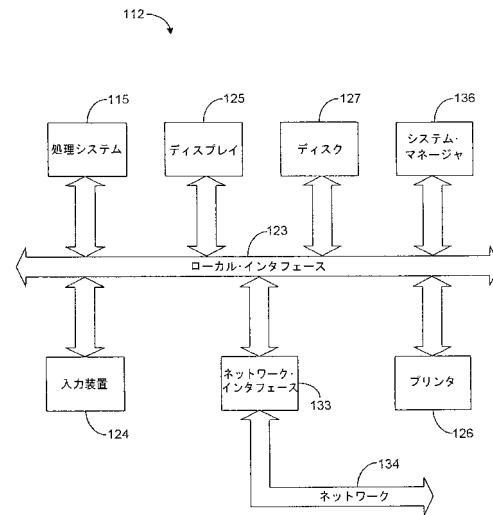
【図 2】



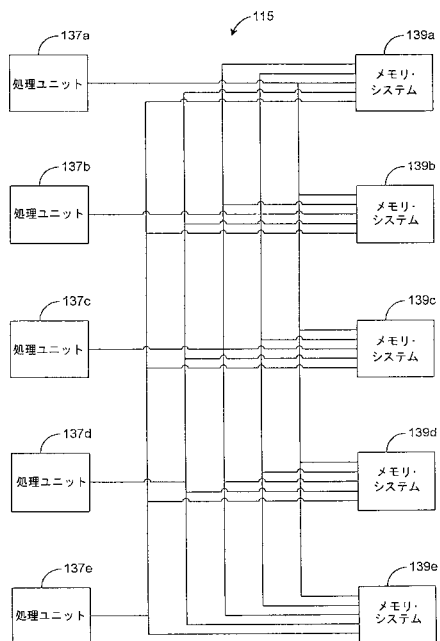
【図 3】



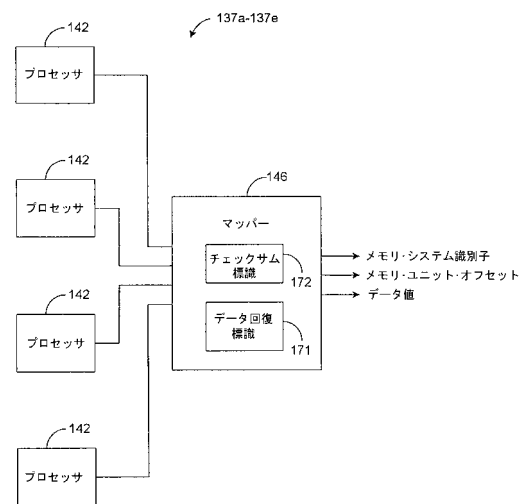
【図 4】



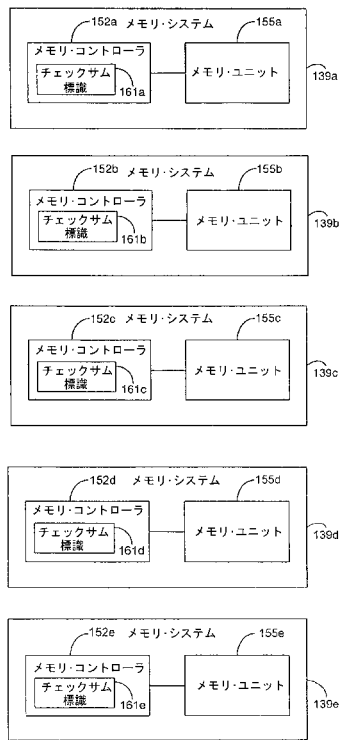
【図 5】



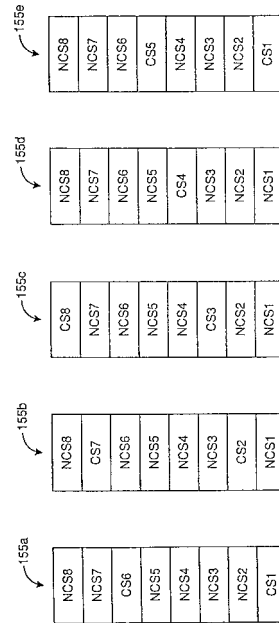
【図 6】



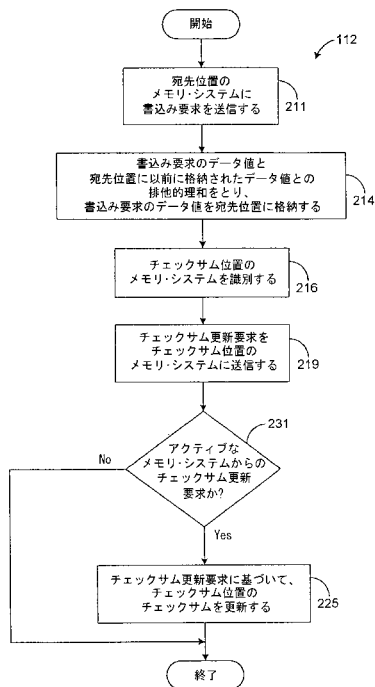
【図 7】



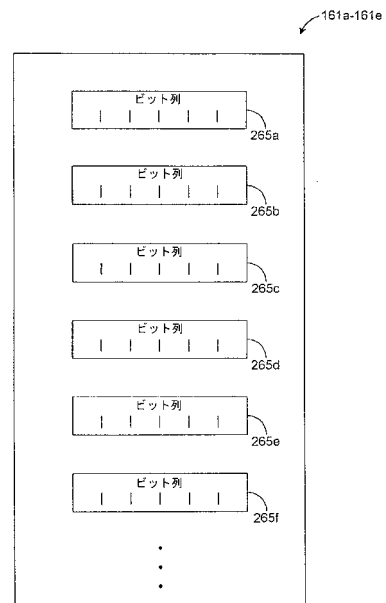
【図 8】



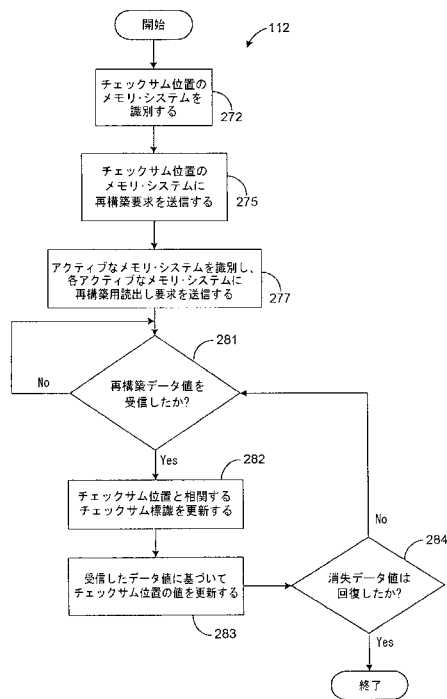
【図 9】



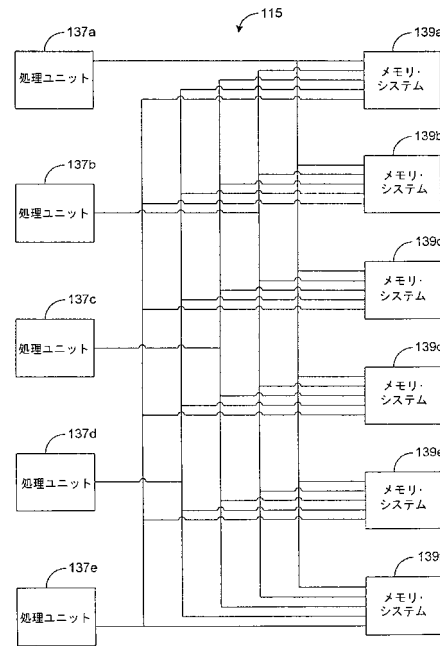
【図 10】



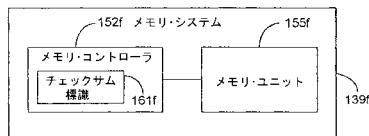
【図 1 1】



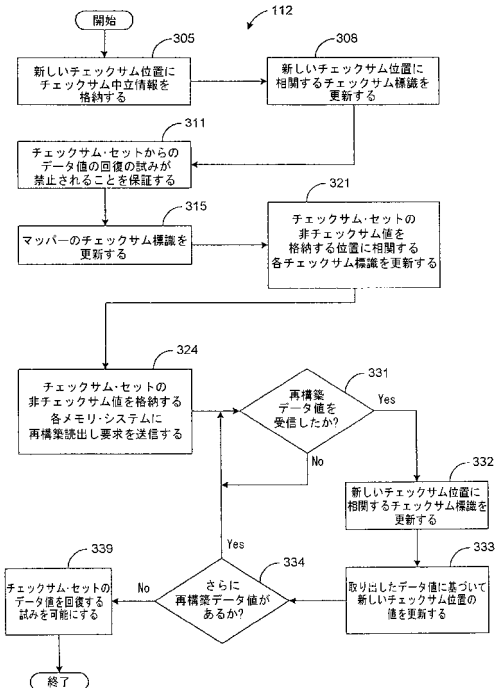
【図 1 2】



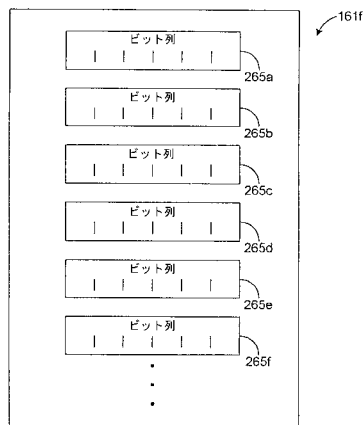
【図 1 3】



【図 1 5】



【図 1 4】



フロントページの続き

- (72)発明者 ブライアン・ホーナング
アメリカ合衆国75075テキサス州プラノ、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・パーマー
アメリカ合衆国75025テキサス州プラノ、メイソン・ドライブ 3012
- (72)発明者 ボール・エフ・ヴォーゲル
アメリカ合衆国75044テキサス州ガーランド、ブライアー・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