

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5142995号  
(P5142995)

(45) 発行日 平成25年2月13日(2013.2.13)

(24) 登録日 平成24年11月30日(2012.11.30)

(51) Int.Cl. F I  
**G06F 12/12 (2006.01)** G06F 12/12 553C  
 G06F 12/12 505

請求項の数 47 (全 17 頁)

<p>(21) 出願番号 特願2008-519413 (P2008-519413)                  (86) (22) 出願日 平成18年6月22日 (2006.6.22)                  (65) 公表番号 特表2008-544422 (P2008-544422A)                  (43) 公表日 平成20年12月4日 (2008.12.4)                  (86) 国際出願番号 PCT/US2006/024291                  (87) 国際公開番号 W02007/002282                  (87) 国際公開日 平成19年1月4日 (2007.1.4)                  審査請求日 平成21年6月4日 (2009.6.4)                  (31) 優先権主張番号 11/169,246                  (32) 優先日 平成17年6月27日 (2005.6.27)                  (33) 優先権主張国 米国 (US)</p>	<p>(73) 特許権者 509123208                  アビニシオ テクノロジー エルエルシー                  アメリカ合衆国 02421 マサチュー                  セッツ州 レキシントン スプリング ス                  トリート 201                  (74) 代理人 100079108                  弁理士 稲葉 良幸                  (74) 代理人 100109346                  弁理士 大貫 敏史                  (72) 発明者 スタンフィル クレイグ ダブリュ                  アメリカ合衆国 01773 マサチュー                  セッツ州 リンカーン ハックルベリー                  ヒル ロード 43                  審査官 中野 裕二</p>
---	---

最終頁に続く

(54) 【発明の名称】 メモリページ管理

(57) 【特許請求の範囲】

【請求項1】

一つ以上の待ち行列のために用いられるメモリをページングするための方法であって、前記待ち行列の少なくとも一部を格納する少なくとも幾つかのメモリページを含むメモリページを識別するとともに、前記識別されるメモリページが再使用される順序を表すリサイクルリストデータ構造を提供するステップと、

前記待ち行列の少なくとも一つの一部を格納するメモリページへのアクセス履歴に基づいて、前記リサイクルリストデータ構造を維持して、前記順序における前記メモリページの位置を、前記メモリページからのシーケンシャルアクセスの移行に依存させるステップを含む方法。

【請求項2】

前記メモリページからのシーケンシャルアクセスの前記移行は、前記メモリページからのポインタの移行に対応する請求項1の方法。

【請求項3】

前記順序における前記メモリページの前記位置は、シーケンシャルアクセスの前記移行が、前記メモリページからの書き込みアクセスの移行に対応するかどうかによって依存する請求項1の方法。

【請求項4】

前記待ち行列の一つと関係付けられた、メモリページへのアクセス履歴に基づいて、前記順序を表す前記リサイクルリストデータ構造を維持するステップは、

前記メモリページからの書き込みアクセスの移行に応じて、前記リサイクルリストデータ構造の第1エンドに、前記メモリページのための識別子を配置するステップと、

閉じている前記メモリページと関係付けられる前記待ち行列に応じて、前記リサイクルリストデータ構造の第2エンドに、前記メモリページのための識別子を配置するステップとを含む請求項1の方法。

【請求項5】

前記リサイクルリストデータ構造の前記第1エンドからのメモリページを再使用するステップをさらに含む請求項4の方法。

【請求項6】

空きメモリページリストが空であることを判定した後、前記リサイクルリストデータ構造の前記第1エンドから前記メモリページを再使用するステップをさらに含む請求項5の方法。

10

【請求項7】

前記メモリページを再使用するステップは、前記メモリページの現在のコンテンツを不揮発性メモリ格納へ書き込むステップを含む請求項5の方法。

【請求項8】

前記メモリページを再使用するステップは、前記不揮発性メモリ格納からのメモリページを前記メモリページへ書き込むステップをさらに含む請求項7の方法。

【請求項9】

一つ以上の待ち行列のために用いられるメモリをページングするための、コンピュータ読取り可能な記録媒体上に格納されるソフトウェアであって、

20

前記待ち行列の一部を格納する少なくとも幾つかのメモリページを含むメモリページを識別するとともに、前記識別されるメモリページが再使用される順序を表すリサイクルリストデータ構造を提供する命令と、

前記待ち行列の少なくとも一つの一部を格納するメモリページへのアクセス履歴に基づいて、前記リサイクルリストデータ構造を維持して、前記順序における前記メモリページの位置を、前記メモリページからのシーケンシャルアクセスの移行に依存させる命令とをコンピュータシステムに実行させるソフトウェア。

【請求項10】

一つ以上の待ち行列のために用いられるメモリをページングするためのシステムであって、

30

前記待ち行列の一部を格納する少なくとも幾つかのメモリページを含むメモリページを識別するとともに、前記識別されるメモリページが再使用される順序を表すリサイクルリストデータ構造を提供するための手段と、

前記待ち行列の少なくとも一つの一部を格納するメモリページへのアクセス履歴に基づいて、前記リサイクルリストデータ構造を維持して、前記順序における前記メモリページの位置を、前記メモリページからのシーケンシャルアクセスの移行に依存させる手段とを備えたシステム。

【請求項11】

一つ以上の待ち行列のために用いられるメモリをページングするための方法であって、

40

前記待ち行列の少なくとも一部を格納するメモリページを識別するとともに、前記識別されるメモリページが再使用される順序を表すリサイクルリストデータ構造を提供するステップと、

第1のケースでは最新使用方式に基づいて、前記リサイクルリストデータ構造の第1エンドで識別されたメモリページを再使用し、第2のケースでは最長時間未使用方式に基づいて、前記リサイクルリストデータ構造の第2エンドで識別されたメモリページを再使用するステップを含む方法。

【請求項12】

前記第1のケースは、前記メモリページからの書き込みポインタの移行に応じて発生する請求項11の方法。

50

## 【請求項 13】

前記第2のケースは、アクセスエンティティが前記待ち行列を閉じることに応じて発生する請求項11の方法。

## 【請求項 14】

前記第2のケースは、前記メモリページからの読み出しポインタの移行に応じて発生する請求項11の方法。

## 【請求項 15】

一つ以上の待ち行列のために用いられるメモリをページングするための、コンピュータ読取り可能な記録媒体上に格納されるソフトウェアであって、

前記待ち行列の少なくとも一部を格納するメモリページを識別するとともに、前記識別されるメモリページが再使用される順序を表すリサイクルリストデータ構造を提供する命令と、

第1のケースでは最新使用方式に基づいて、前記リサイクルリストデータ構造の第1エンドで識別されたメモリページを再使用し、第2のケースでは最長時間未使用方式に基づいて、前記リサイクルリストデータ構造の第2エンドで識別されたメモリページを再使用させるための命令を、コンピュータシステムに実行させるソフトウェア。

## 【請求項 16】

前記リサイクルリストデータ構造によって識別されたメモリページを再使用して、前記待ち行列の一つの一部を格納するステップをさらに含む請求項1の方法。

## 【請求項 17】

前記リサイクルリストデータ構造によって識別された前記再使用されるメモリページは、再使用される前に前記待ち行列の一つの別の一部を格納していた請求項16の方法。

## 【請求項 18】

前記再使用されるメモリページに格納されている、前記待ち行列の一つの前記一部は、待ち行列に書き込まれている待ち行列メッセージを含む請求項16の方法。

## 【請求項 19】

前記待ち行列を格納する前記メモリページは、待ち行列メッセージと、前記待ち行列メッセージの管理に関連する記録とを格納する請求項1の方法。

## 【請求項 20】

前記第1のケースでは、前記メモリページは、待ち行列の第1の部分を格納し、前記待ち行列の第2の部分を格納するために再使用される請求項11の方法。

## 【請求項 21】

書き込みアクセスエンティティは、書き込みポインタで前記第2の部分を前記メモリページに書き込み、読み出しアクセスエンティティは、前記書き込みポインタから遅延する読み出しポインタで前記待ち行列からデータを読み出す請求項20の方法。

## 【請求項 22】

前記メモリページは、他のメモリページを使用する前に、一つ以上の最長時間未使用メモリページを前記読み出しアクセスエンティティのために利用可能としたまま、前記書き込みアクセスエンティティによって前記待ち行列に書き込むために繰り返し再使用される請求項21の方法。

## 【請求項 23】

前記第2のケースでは、前記メモリページは、第1の待ち行列の一部を格納し、第2の待ち行列の一部を格納するために再使用される請求項11の方法。

## 【請求項 24】

前記第2の待ち行列の前記一部は、アクセスエンティティによって前記第1の待ち行列が閉じられた後に、前記メモリページに格納される請求項23の方法。

## 【請求項 25】

前記第2のケースでは、前記メモリページは、待ち行列の第1の部分を格納し、前記待ち行列の第2の部分を格納するために再使用される請求項11の方法。

## 【請求項 26】

10

20

30

40

50

書き込みアクセスエンティティは、前記第 1 の部分を読み出す読み出しアクセスエンティティの読み出しポインタが前記メモリページから移行した後に、前記第 2 の部分を前記メモリページに書き込む請求項 25 の方法。

【請求項 27】

前記待ち行列に格納される論理的に順序付けられたエレメントは、複数のシーケンシャル待ち行列メッセージを含む請求項 1 の方法。

【請求項 28】

前記論理的に順序付けられたエレメントは、前記シーケンシャル待ち行列メッセージに関連する少なくとも 1 つの記録を含む請求項 27 の方法。

【請求項 29】

前記待ち行列に格納される論理的に順序付けられたエレメントは、書き込みアクセスエンティティによって書き込まれ、少なくとも一つの読み出しアクセスエンティティによって読み出される請求項 1 の方法。

【請求項 30】

前記書き込みアクセスエンティティと前記読み出しアクセスエンティティとは異なる請求項 29 の方法。

【請求項 31】

第 1 のメモリページのコンテンツを揮発性メモリ格納から不揮発性メモリ格納に書き込んだ後に第 1 の待ち行列の一部を前記第 1 のメモリページに格納するステップを含む、第 1 のアクセスエンティティからの要求に応じて前記第 1 の待ち行列を開くステップと、

前記第 1 のキューのための書き込みポインタを前記第 1 のアクセスエンティティに提供するステップとをさらに含む請求項 1 の方法。

【請求項 32】

前記第 1 の待ち行列のための一つ以上の読み出しポインタを一つ以上の追加のアクセスエンティティに提供するステップをさらに含む請求項 31 の方法。

【請求項 33】

前記第 1 の待ち行列の一部を格納する第 2 のメモリページを、前記書き込みポインタまたは前記読み出しポインタの何れも前記第 2 のメモリページを指し示していないことを判定した後に、前記リサイクルリストデータ構造に移動させるステップをさらに含む請求項 32 の方法。

【請求項 34】

前記第 1 のケースでは、前記メモリページは、待ち行列の第 1 の部分を格納し、前記待ち行列の第 2 の部分を格納するために再使用される請求項 15 のソフトウェア。

【請求項 35】

書き込みアクセスエンティティは、書き込みポインタで前記第 2 の部分を前記メモリページに書き込み、読み出しアクセスエンティティは、前記書き込みポインタから遅延する読み出しポインタで前記待ち行列からデータを読み出す請求項 34 のソフトウェア。

【請求項 36】

前記メモリページは、他のメモリページを使用する前に、一つ以上の最長時間未使用メモリページを前記読み出しアクセスエンティティのために利用可能としたまま、前記書き込みアクセスエンティティによって前記待ち行列に書き込むために繰り返し再使用される請求項 35 のソフトウェア。

【請求項 37】

前記第 2 のケースでは、前記メモリページは、第 1 の待ち行列の一部を格納し、第 2 の待ち行列の一部を格納するために再使用される請求項 15 のソフトウェア。

【請求項 38】

前記第 2 の待ち行列の前記一部は、アクセスエンティティによって前記第 1 の待ち行列が閉じられた後に、前記メモリページに格納される請求項 37 のソフトウェア。

【請求項 39】

前記第 2 のケースでは、前記メモリページは、待ち行列の第 1 の部分を格納し、前記待

10

20

30

40

50

ち行列の第 2 の部分を格納するために再使用される請求項 15 のソフトウェア。

【請求項 40】

書き込みアクセスエンティティは、前記第 1 の部分を読み出す読み出しアクセスエンティティの読み出しポイントが前記メモリページから移行した後に、前記第 2 の部分を前記メモリページに書き込む請求項 39 のソフトウェア。

【請求項 41】

前記メモリページは、待ち行列の第 1 の部分を格納し、前記待ち行列の第 2 の部分を格納するために再使用される請求項 10 のシステム。

【請求項 42】

書き込みアクセスエンティティは、書き込みポイントで前記第 2 の部分を前記メモリページに書き込み、読み出しアクセスエンティティは、前記書き込みポイントから遅延する読み出しポイントで前記待ち行列からデータを読み出す請求項 41 のシステム。

10

【請求項 43】

前記メモリページは、他のメモリページを使用する前に、一つ以上の最長時間未使用メモリページを前記読み出しアクセスエンティティのために利用可能としたまま、前記書き込みアクセスエンティティによって前記待ち行列に書き込むために繰り返し再使用される請求項 42 のシステム。

【請求項 44】

前記メモリページは、第 1 の待ち行列の一部を格納し、第 2 の待ち行列の一部を格納するために再使用される請求項 10 のシステム。

20

【請求項 45】

前記第 2 の待ち行列の前記一部は、アクセスエンティティによって前記第 1 の待ち行列が閉じられた後に、前記メモリページに格納される請求項 44 のシステム。

【請求項 46】

前記メモリページは、待ち行列の第 1 の部分を格納し、前記待ち行列の第 2 の部分を格納するために再使用される請求項 10 のシステム。

【請求項 47】

書き込みアクセスエンティティは、前記第 1 の部分を読み出す読み出しアクセスエンティティの読み出しポイントが前記メモリページから移行した後に、前記第 2 の部分を前記メモリページに書き込む請求項 46 のシステム。

30

【発明の詳細な説明】

【発明の詳細な説明】

【0001】

発明の背景

本発明はメモリページの管理に関する。

【0002】

揮発性メモリ格納（例えば、ランダムアクセスメモリ（RAM））を用いて、不揮発性メモリ格納（例えば、一つ以上のディスクドライブ）にも格納されているデータと連携をとることは有用である。例えば、揮発性メモリ格納は、普通は不揮発性メモリ格納よりアクセスおよび読み出し時間が高速である。揮発性メモリ格納および不揮発性メモリ格納のアドレス空間は、所定数の連続するアドレス配置にまたがるページに編成することができる。コンピュータシステムは、ページング技法を用いて、例えば、プログラミングまたはデータ構造が使用する揮発性メモリからメモリページを割り当てる。不揮発性メモリ格納は、普通は揮発性メモリ格納よりずっと規模が大きい。仮想メモリ方式を用いて、使用していない（例えば、現在、書き込みも読み出しもしていない）揮発性メモリ格納内のページを不揮発性メモリ格納へ書き込んでおいて、後で必要な時に、不揮発性メモリ格納から再度読み出すことができる。したがって、揮発性メモリ格納内のページは、各種のプログラミングおよび/またはデータ構造による使用のために「リサイクル（再使用）」される。ページを前後しての移動と関係付けられるオーバーヘッドを低減するために、技法によっては、最長時間未使用（LRU）方式に基づいてページをリサイクルする。LRU方式

40

50

では、不揮発性メモリ格納へ書き込まれたページが揮発性メモリ格納への移動対象であるにもかかわらず、揮発性メモリ格納内に空きページがない場合、揮発性メモリ格納から不揮発性メモリ格納へリサイクル（つまり、スワッピングまたはページアウト）するように選択されるページは、未使用時間が最長の（例えば、最も長い時間、読み出しまたは書き込みがなされていない）ページである。

【0003】

#### 発明の概要

一般的な態様において、本発明は、シーケンシャルアクセスされる一つ以上のデータ構造のために用いられるメモリをページングするための方法、ならびにそれに対応するソフトウェアおよびシステムを特徴とする。本方法は、メモリページの再使用順序を表すデータ構造を提供するステップ、およびシーケンシャルアクセスされるデータ構造の一つと関係付けられた、メモリページへのアクセス履歴に基づいて、データ構造を維持するステップを含み、よって、その順序におけるメモリページの位置を、メモリページからのシーケンシャルアクセスの移行に依存させる。

10

【0004】

この態様は、以下の特徴の一つ以上を含む：

メモリページからのシーケンシャルアクセスの移行は、メモリページからのポインタの移行に対応する。

その順序におけるメモリページの位置は、シーケンシャルアクセスの移行が、メモリページからの書き込みアクセスの移行に対応するかどうかに依存する。

20

シーケンシャルアクセスされるデータ構造の一つと関係付けられた、メモリページへのアクセス履歴に基づくデータ構造を維持するステップが、メモリページからの書き込みアクセスの移行に応じて、順序を表すデータ構造の第1エンドに、メモリページ用の識別子を配置するステップと、閉じられているメモリページと関係付けられる、シーケンシャルアクセスされるデータ構造に応じて、順序を表すデータ構造の第2エンドに、メモリページ用の識別子を配置するステップとを含む。

【0005】

本方法は、順序を表すデータ構造の第1エンドからのメモリページを再使用するステップをさらに含む。

【0006】

本方法は、空きメモリページリストが空であることを判定した後、データ構造の第1エンドからの前記メモリページを再使用するステップをさらに含む。

30

【0007】

メモリページを再使用するステップは、メモリページの現在のコンテンツを不揮発性メモリ格納へ書き込むステップを含む。

【0008】

メモリページを再使用するステップは、不揮発性メモリ格納からのメモリページをメモリページへ書き込むステップをさらに含む。

【0009】

メモリページと関係付けられる、シーケンシャルアクセスされるデータ構造は、待ち行列を含む。

40

【0010】

一般的な態様において、本発明は、シーケンシャルアクセスされる一つ以上のデータ構造のために用いられるメモリをページングするための方法、ならびにそれに対応するソフトウェアおよびシステムを特徴とし、第1のケースでは最長時間使用方式に基づき、第2のケースでは最長時間未使用方式に基づいて、シーケンシャルアクセスされるデータ構造に関係付けられるメモリページを再使用するステップを含む。

【0011】

本態様は以下の特徴の一つ以上を含むことができる：

第1のケースは、メモリページからの書き込みポインタの移行を含む。

50

第2のケースは、シーケンシャルアクセスされるデータ構造を閉じるアクセスエンティティを含む。

第2のケースは、メモリページからの読み出しポインタの移行を含む。

【0012】

本発明の態様は、以下の利点の一つ以上を含むことができる：

再使用のためのページを順序付けるステップは、待ち行列または他のシーケンシャルI/Oデータ構造を格納するメモリシステムに対して、どのページがすぐに再使用される可能性があるのか、より良好な予測を提供できる。書き込まれたばかりの（すなわち、書き込みポインタがページを去ったばかりの）ページは、読み出しポインタが書き込みポインタに「追いつかなければ」ならないので、すぐに再使用される可能性が少ない。したがって、これらのページは、最初にリサイクルされる。さらに、閉じたばかりの待ち行列に対するページは、すぐに再オープンされて、再使用されるかもしれないので、後までリサイクルされない。しばらく閉じたままのページはその間にどこかにリサイクルされる。

10

【0013】

待ち行列特有のページングアルゴリズムの使用は、待ち行列構造またはアクセスパターンを利用しないページングアルゴリズムの使用を超える高い効率を提供できる。データ格納待ち行列のページングは、どのページが再使用されるかを決定する際に、待ち行列アクセス履歴を利用することにより改良できる。例えば、アクセス履歴に基づいて、幾つかのページは他のページより早く再アクセスされる、と予測することができる。この改良により、作業している記憶装置内に必要なページが存在する可能性がより高くなるので、作業中の記憶装置へデータを転送して戻すことに起因する遅延を低減できる。

20

【0014】

本発明の他の特徴および利点は、以下の説明および特許請求の範囲から明らかになる。

【0015】

詳細な説明

図1を参照すると、待ち行列システム100により、アクセスエンティティ102A~102Pは、1セット以上の待ち行列QUEUE\_1~QUEUE\_Nへメッセージを追加し、待ち行列からメッセージを読み出すことができる。アクセスエンティティは待ち行列マネージャ102を介してシステム100と相互作用する。待ち行列マネージャ102はメモリ管理モジュール(MMM)106を用いて、データ読み書きのアクセスが相対的に高速な一時的作業記憶装置である揮発性メモリ格納108と、データの永続性が相対的に高い恒久的記憶装置であって揮発性格納より読み書きのアクセスが相対的に低速な不揮発性メモリ110との間で、ページのスワッピングを支援する。MMM106は、ソフトウェアまたはハードウェアまたはその両方で実装できる。MMMは、揮発性メモリアドレス空間と不揮発性メモリアドレス空間との間のページアドレスを翻訳する。MMMは揮発性メモリへのページ参照の監視も行い、利用不可のページ（例えば、揮発性メモリページが不揮発性メモリへ書き込まれた後、新規使用のためにリサイクルされたページ）への参照を検出できる。MMMは、この「ページフォールト」について、待ち行列マネージャ102に警告することにより応答する。次いで、待ち行列マネージャ102は、参照されたページを不揮発性メモリから読み出し、揮発性メモリの空きページ内へそのページをコピーする。揮発性メモリに空きページがない場合、待ち行列マネージャ102は、後で詳細に説明するように、「リサイクルリスト」に基づいてページをリサイクルする。待ち行列マネージャ102およびMMM106は、任意の種類オペレーティングシステムおよびコンピュータ環境で実装できる。待ち行列マネージャおよび/またはMMM106は、オペレーティングシステム内へ統合できる。代替として、オペレーティングシステムは、待ち行列マネージャ102へ割り当てられないページを管理するための、別の仮想メモリシステムを有してもよい。

30

40

【0016】

所与の待ち行列（例えばQUEUE\_2）は一連の記録を格納でき、一セットの「待ち行列ペ

50

ージ」内に、シーケンシャル待ち行列メッセージおよびこれらのメッセージの管理に関連する記録を含む。メッセージおよび待ち行列ページの相対的なサイズに応じて、待ち行列ページは一つ以上のメッセージ、または一つのメッセージの一部を保持できる。待ち行列マネージャ102は、一セットの待ち行列ページに待ち行列を割り当て、待ち行列が成長すると、そのセットへページを追加できる。図示の実施の形態では、セット内の待ち行列ページは全て不揮発性メモリ格納110内に格納され、アクティブな待ち行列ページのサブセットは揮発性メモリ格納108内に格納される。

【0017】

待ち行列（または他のシーケンシャルアクセスされるデータ構造）の「シーケンシャル」状況は、メッセージ（またはデータ構造の他の要素）の論理的順序を参照する。メッセージの格納配置は、必ずしも、システム100の物理的または仮想的なアドレス空間内のシーケンシャルアドレスにある必要はない。

10

【0018】

揮発性メモリ格納108内の待ち行列ページのコンテンツが変更された場合、その待ち行列ページは「ダーティ」とマーキングされる。ダーティ待ち行列ページが揮発性メモリ格納108から削除された場合、不揮発性メモリ格納110内の待ち行列ページの古いコピーは、その変更により更新される（すなわち、待ち行列ページは、不揮発性メモリ格納110へ「同期」される）。ダーティ待ち行列ページは、待ち行列ページが揮発性メモリ格納108から削除される前に、不揮発性メモリ格納110に同期させることもできる。待ち行列マネージャは、待ち行列が待ち行列ページをもはや必要としなくなった後（例えば、待ち行列が削除された後）、待ち行列に対する幾つかのまたは全ての待ち行列ページを回収する。

20

【0019】

図2Aを参照すると、リサイクルリスト200は、リサイクルされる可能性がある揮発性メモリ内のページを識別する。リサイクルリスト200は、第1の「ヘッド」エンド202および第2の「テール」エンド204を有する。ページは、後述するように、テールまたはヘッドエンドのどちらかに追加することができ、ヘッドエンド202から削除される。リサイクルリストは、例えば、二重リンクされたリストとして実装できる。

【0020】

一実施の形態では、リサイクルリスト200は、ページテーブル206内の待ち行列ページを識別する「仮想アドレス」を格納する（例えば、QUEUE\_1へ割り当てられた待ち行列ページ「5」を識別する「Q1/P5」）。ページテーブル206は、待ち行列システム100内の待ち行列へ割り当てられる待ち行列ページそれぞれに対するエントリを含む。図では、ページテーブル206は、それぞれのエントリに対して、不揮発性メモリ空間208内に不揮発性メモリページアドレス「N-PAGE」を含み、揮発性メモリ空間210内に揮発性メモリページアドレス「V-PAGE」を含む。揮発性メモリページアドレスは、揮発性メモリ格納108内に格納されない待ち行列ページに対しては空（例えば、未使用アドレス値と一致）であってもよい。ページテーブル206は、待ち行列ページがリサイクルリスト200内にあるかどうかを指示する「リサイクルビット」を含む。ページテーブルは、幾つのアクセスエンティティが待ち行列ページにアクセスしているかを示す「参照カウント」を含む。

30

40

【0021】

各待ち行列は、不揮発性メモリ格納110内に格納される割り当て待ち行列ページセットを有する（例えば、図3Bで、QUEUE\_2は、アドレスN-PAGE\_10～N-PAGE\_16に格納されるページQ2/P1～Q2/P7を有する）。待ち行列ページによっては、揮発性および不揮発性メモリの両方にコピーを持つことができる（例えば、図3Aおよび3BのページQ2/P3～Q2/P7）。揮発性メモリ格納108内に格納される待ち行列ページに対しては、ページテーブル206は、「ダーティビット」を用いて、まだ不揮発性メモリ格納110へ戻して同期していない、待ち行列ページのコンテンツへ変更があるかどうかを示す。

【0022】

50

例えば、待ち行列へ新規に追加されたメッセージは、単に揮発性メモリ内に格納されているだけとしてもよい。関係する待ち行列メッセージを格納する待ち行列ページは、待ち行列ページが揮発性メモリから不揮発性メモリへコミット操作で同期されるので、揮発性および不揮発性メモリの両方に格納されていてもよい。(幾つかの方式の何れかを用いて、関係するメッセージを同期させることでメッセージが失われないようにすることができる。メッセージを確実にする一手法は、引用して本明細書に組み込まれる同時係属/同時出願の「メッセージ待ち行列管理」に記載されている)。待ち行列ページを格納する揮発性メモリのページは、不揮発性メモリ内のコピーだけを残して、リサイクルしてしまってもよい(例えば、別の待ち行列が使用するために)。

#### 【0023】

一つ以上のアクセスエンティティは、読み出しおよび/または書き込みのために待ち行列を開くことができる。アクセスエンティティは、待ち行列を開いて、待ち行列の最後に一つ以上のメッセージを追加することができる。アクセスエンティティは、待ち行列を開いて、待ち行列の前から、または待ち行列内の特定の場所(例えば、以前に待ち行列から読み出した時に到達した場所)から、関係するメッセージを読み出すことができる。アクセスエンティティによるアクセスに対して待ち行列が開かれると、待ち行列システム100は、アクセスエンティティに対して読み出しおよび/または書き込みポインタを設定する等の、開くための作業を実行する。アクセスエンティティが待ち行列へのアクセス終了後、アクセスエンティティは待ち行列を閉じ、待ち行列システム100は、閉じるための適切な作業を実行する。多数のアクセスエンティティが、同一待ち行列へ同時にアクセス

#### 【0024】

アクセスエンティティが待ち行列を開くと、アクセスエンティティは、アクセスされている待ち行列ページに対する参照カウントをインクリメントする。アクセス対象である待ち行列ページが既に揮発性メモリ格納108内にはない場合、その待ち行列ページは、不揮発性メモリ格納110から、揮発性メモリ格納108へコピーされる。アクセスエンティティが待ち行列ページへのアクセスを終了した後(例えば、待ち行列を閉じた後または待ち行列ページの最後を越えて書き込んだ後)、アクセスエンティティは、その待ち行列ページに対する参照カウントをデクリメントする。

#### 【0025】

待ち行列システム100は、アクセスエンティティが読み出しポインタでデータを読み出している場所、およびアクセスエンティティが書き込みポインタでデータを追加(すなわち書き込み)している場所のトラックを保持する。アクセスエンティティが待ち行列から読み出し、またはそれへ書き込んでいる間、アクセスされている待ち行列ページは、揮発性メモリ格納108内にある。待ち行列システム100は、空きリスト212により利用可能な揮発性メモリページのトラックを保持する。新規待ち行列ページが不揮発性メモリ格納110から揮発性メモリ格納108へコピーされる対象であるのに、空きページがない場合(すなわち、空きリストが空の場合)、待ち行列システム100は、リサイクルリスト200に基づいて揮発性メモリページをリサイクルする。

#### 【0026】

揮発性メモリページをリサイクルするには、待ち行列システム100が、リサイクルリスト200のヘッド202で識別されるページをチェックすることにより、リサイクルするページを探す。識別されたページが、所定の基準によりリサイクル可能なページとして適格である場合(例えば、読み出しまたは書き込みポインタが現在参照していないことを示す参照カウントゼロをそのページが有する)、そのページの現在のコンテンツは、(必要なら)不揮発性メモリ格納110へ書き込まれ、そのページは割り当てが自由になる(次いで、リサイクルリスト200から削除される)。そのページの現在のコンテンツが、既に不揮発性メモリに同期している場合、待ち行列システム100は、不揮発性メモリへの書き込みをせずにページを開放する。識別されたページが、リサイクル可能なページと

10

20

30

40

50

して不適格である場合、そのページはリサイクルリストから削除され、リサイクルリストの新規ヘッド 202 が、リサイクル可能ページに対するチェックを受ける。

【0027】

－実施では、待ち行列システム 100 が、リサイクルリスト 200 へ揮発性メモリページを追加する（すなわち、そのページは、開放される可能性のある候補となる）のには、2つのケースがある。

ケース 1：書き込みポインタがページから移動し、他の読み出しまたは書き込みポインタがどれもそのページを指し示さず、そのページは既にリサイクルリスト 200 上にない。

ケース 2：ページを使用している待ち行列が閉じられ、他の読み出しまたは書き込みポインタがどれもそのページを指し示さず、そのページは既にリサイクルリスト 200 上にない。

10

【0028】

ケース 1 は、例えば、待ち行列に一連のメッセージを追加しているアクセスエンティティが、メッセージで待ち行列ページを埋める場合に生じる可能性がある。ケース 1 では、待ち行列システム 100 は、リサイクルリスト 200 のヘッド 202 へそのページを追加する。このケースは、より新しく書き込まれたページが、それより前に書き込まれたページの前にリサイクルされるので、最長時間使用（MRU）方式に対応する。

【0029】

ケース 2 では、待ち行列システム 100 は、リサイクルリスト 200 のテール 204 へそのページを追加する。このケースは、ページがリサイクルリストの先頭から用いられ、それにより、古い閉じられたページが、より新しい閉じられたページの前にリサイクルされるので、最長時間未使用（LRU）方式に対応する。

20

【0030】

他の実施では、ページをリサイクルリスト 200 へ追加するために他のケースを含めることができる。例えば、ページをリサイクルリスト 200 のテール 204 へ追加することができる別のケースは：

ケース 3：読み出しポインタがページから移動し、他の読み出しまたは書き込みポインタがどれもそのページを指し示さず、そのページは既にリサイクルリスト 200 上にない。

【0031】

他の実施は、各ページへのアクセス履歴に基づいて、リサイクルリスト内でページ順序を維持するための多様な規則のどれかを含むことができる。メモリページからのシーケンシャルアクセスの移行（例えば、ページから移動する書き込みポインタ）に依存するよう、メモリページの少なくとも幾つかの位置を順序よく編成することにより、本システムは、シーケンシャルアクセスされるデータ構造の統計的特性を利用して、ページングの効率を高めることができる。

30

【0032】

空きリスト 212 内のページは、もはや必要ではない、したがって、待ち行列ページと関係付けられないページに対応する。待ち行列システム 100 がページを空きリスト 212 へ追加するケースは様々である。あるケースでは、システム 100 は、リーダー（読み出し者）として待ち行列へ申し込んでいる、全てのアクセスエンティティにより読み出された（すなわち、全ての申し込み者の読み出しポインタが、ページ上の各メッセージを処理してしまい、そのページから去ってしまった）待ち行列メッセージを保持するページを開放する。待ち行列システム 100 は、何らかの様々な他の理由でページを開放できる。例えば、ページ上のメッセージ全てが、削除されるかまたは「片づけられた」場合、ページが開放される。

40

【0033】

実施によっては、ページは時間制限に基づいて開放される。例えば、ページは、ページと関係付けられる待ち行列が、所定時間（例えば 2 分間）を越えて閉じていた後に、開放される。別のケースでは、ページは、関係する待ち行列が開いていても、所定時間（例えば 10 分間）を越えてアクセスされなかった後は開放される。

50

## 【 0 0 3 4 】

さらに、待ち行列マネージャ 1 0 2 は、他のプログラムまたはデータ構造が使用する揮発性メモリページを開放して、待ち行列システム 1 0 0 が待ち行列ページを割り当てるために使用できる。待ち行列システム 1 0 0 は、全ての揮発性メモリページを開放するために用いることもできるグローバルフリーリスト 2 1 2 を用いることができ、さもなければ代替として、待ち行列ページに用いる揮発性メモリページに対してセパレートフリーリスト 2 1 2 を用いることができる。

## 【 0 0 3 5 】

図 2 B に示すのは、仮想アドレスに基づいてシーケンシャルに並べた QUEUE\_1 に対するページ Q1/P1 ~ Q1/P8 のセットの実施例である。本実施例では、一つの書き込みアクセスエンティティが、一連のメッセージを待ち行列に書き込んでいて、二つの読み出しアクセスエンティティが、利用可能になった時にメッセージを読み出している。書き込みアクセスエンティティは、シーケンシャルに増加する仮想アドレスで、ページ上にメッセージを書き込む書き込みポインタ WRITER\_1 を有する。書き込みポインタ WRITER\_1 がページ Q1/P7 からページ Q1/P8 へ移動した後、ページ Q1/P7 は、ケース 1 のもとでリサイクルリスト 2 0 0 のヘッド 2 0 2 上に配置可能となる。読み出しポインタ READER\_2 がページ Q1/P2 からページ Q1/P3 へ移動した後、ページ Q1/P2 は、ケース 3 のもとでリサイクルリスト 2 0 0 のテール 2 0 4 上に配置可能となる。

## 【 0 0 3 6 】

第 1 の読み出しアクセスエンティティは読み出しポインタ READER\_1 を有し、このポインタが、書き込みポインタ WRITER\_1 を何ページかの待ち行列ページ分だけ遅延させる。この遅延は、そのアクセスエンティティが待ち行列システム 1 0 0 と相互作用する時に短くなったり長くなったりする。第 2 の読み出しアクセスエンティティは、同様に、書き込みポインタ WRITER\_1 を遅延させる読み出しポインタ READER\_2 を有する。空きページがライターにとって利用可能なままである場合、リーダーは、揮発性メモリ格納 1 0 8 から、書き込まれたページにアクセスできる。空きページを利用できない場合（例えば、揮発性メモリ空間が、リーダーによる大きな遅延が原因で一杯になっている場合）、書き込まれたページの幾つかをリサイクルできるかもしれない。例えば、このページ管理方式のもとでは、最新使用ページが繰り返しリサイクルされることになり、揮発性メモリ内には、遅れているリーダーが利用可能な最長時間未使用ページのシーケンスが残される。

## 【 0 0 3 7 】

QUEUE\_1 を閉じた後、リサイクルリスト 2 0 0 上に既がない待ち行列ページはどれも、リサイクルリスト 2 0 0 のテール 2 0 4 上に配置可能となる。テール 2 0 4 上に配置される最新の閉じた待ち行列に対するこのような待ち行列ページはどれも、テール 2 0 4 からヘッド 2 0 2 まで、そのページがリサイクルリスト 2 0 0 を上方「スクロール」するまで、別のアクセスエンティティによるアクセスに対して揮発性メモリ格納 1 0 8 内に利用可能なまま残る。

## 【 0 0 3 8 】

図 3 A は、揮発性メモリ格納 1 0 8 に対するメモリアドレス空間 2 1 0 内の待ち行列ページの実施例を示す。ページアドレス V-PAGE\_1 ~ V-PAGE\_2 は空なので、図 3 D に示す空きリスト 2 1 2 内に含まれる。アドレス V-PAGE\_3 ~ V-PAGE\_7 を用いて、QUEUE\_2 の待ち行列ページを格納する。読み出しアクセスエンティティは、待ち行列ページ Q2/P3 内のメッセージを読み出すための、V-PAGE\_2 内に配置された読み出しポインタ READER\_1 を有する（本実施例では待ち行列ページは複数のメッセージを格納する）。書き込みアクセスエンティティは、待ち行列ページ Q2/P7 内にメッセージを書き込むための、V-PAGE\_7 内に配置された書き込みポインタ WRITER\_1 を有する。アドレス V-PAGE\_8 ~ V-PAGE\_10 を用いて、閉じられる QUEUE\_1 の待ち行列ページを格納する。アドレス V-PAGE\_11 ~ V-PAGE\_12 を用いて他のデータを格納する。

## 【 0 0 3 9 】

図 3 B に示すのは、不揮発性メモリ格納 1 1 0 のメモリアドレス空間 2 0 8 の実施例で

10

20

30

40

50

ある。不揮発性メモリ格納 110 は、同期されている（例えばコミット操作による同期）QUEUE\_1 および QUEUE\_2 の待ち行列ページを格納し、その内の幾つかは揮発性メモリ格納 108 にも新しくロードされている。本実施例では、揮発性メモリ内の待ち行列ページ Q2/P7 は、不揮発性メモリとまだ同期していない新規メッセージで書き込まれている。

#### 【0040】

図 3C は、揮発性メモリ格納 108 内の活動履歴に対応するリサイクルリスト 200 を示す。例えば、QUEUE\_1 は閉じたばかりなので、待ち行列システム 100 は、LRU 方式に基づいてリサイクルするために、リサイクルリスト 200 のテールエンド 204 上にこれらの仮想アドレスを配置することにより、待ち行列ページ Q1/P3 ~ Q1/P5 を記録したところである。書き込みポインタ WRITER\_1 は、待ち行列ページ Q2/P4、Q2/P5、および Q2/P6 (この順序で) から移動したばかりである。したがって、待ち行列システム 100 は、これらの仮想アドレスを、リサイクルリスト 200 の、テールエンド 204 ではなくヘッドエンド 202 上に配置したところである。

10

#### 【0041】

図 4A ~ 4C は、続いて起こる活動の後の、揮発性アドレス空間 210、不揮発性アドレス空間 208、およびリサイクルリスト 200 の状態をそれぞれ示す。書き込みポインタ WRITER\_1 は、空きリスト 212 から割り当てられるアドレス V-PAGE\_2 で新規待ち行列ページ Q2/P8 上に移動したところである。第 2 の読み出しアクセスエンティティは、V-PAGE\_1 内に配置された読み出しポインタ READER\_2 を有し、空きリスト 212 から割り当てられた待ち行列ページ Q1/P1 内のメッセージを読み出す。これらのページ割り当て後、空きリスト 212 は空になる。

20

#### 【0042】

図 5A ~ 5C は、リサイクルリスト 200 からのページ割り当て後の、揮発性アドレス空間 210、不揮発性アドレス空間 208、およびリサイクルリスト 200 の状態をそれぞれ示す。書き込みポインタ WRITER\_1 は、待ち行列ページ Q2/P8 上に残る。読み出しポインタ READER\_1 は、V-PAGE\_4 で待ち行列ページ Q2/P4 上に移動したところである。読み出しポインタ READER\_2 は、リサイクルリスト 200 (図 4C) のヘッド 202 から割り当てられた待ち行列ページ Q1/P2 上に移動したところである。このページ割り当て後、リサイクルリスト 200 の新規ヘッド 202 は、Q2/P6 となる (図 5C)。待ち行列ページ QUEUE\_1A および QUEUE\_1C はもはや必要でなくなるので、アドレス V-PAGE\_1 および V-PAGE\_3 は、空きリスト 204 (図 5D) 上に配置される。

30

#### 【0043】

この実施例では、単一ライターおよび単一リーダーが所与の待ち行列にアクセスする。その他の実施例では、多数のアクセスエンティティ (例えば、加入者および/またはライター) が同一の待ち行列と相互作用できる (それぞれが自身の読み出しまたは書き込みポインタを有する)。多数のアクセスエンティティが一つの待ち行列と相互作用する場合、メモリページをリサイクルするために上記説明の同一ケースを用いることができる。

#### 【0044】

待ち行列システム 100 は、本明細書で説明した技法と組み合わせた他の技法を用いてシステム性能を高めることができる。例えば、待ち行列マネージャ 102 は、ページが要求されようとしている時を予測でき、多数のページを同時にリサイクルできる。待ち行列マネージャ 102 は、例えば、リサイクルリスト内のページが追加された後にそれを並べ替えることによりリサイクルリストを修正することもできる。

40

#### 【0045】

上記のページ管理方式は、コンピュータ上で実行するためのソフトウェアを用いて実施できる。例えば、このソフトウェアは、プログラムされるか、またはプログラム可能な一台以上のコンピュータシステム (分散型、クライアント/サーバ型、またはグリッド型のような各種アーキテクチャで構成できる) 上で実行される、一つ以上のコンピュータプログラム内の手順を形成し、上記コンピュータシステムはそれぞれ、少なくとも一つのプロセッサ、少なくとも一つのデータ記憶装置システム (揮発性および不揮発性メモリおよび

50

／または記憶素子を含む)、少なくとも一つの入力装置または入力ポート、および少なくとも一つ出力装置または出力ポートを含む。このソフトウェアは、例えば、計算グラフの設計および構成に関連する他のサービスを提供する大規模なプログラムの一つ以上のモジュールを形成してもよい。グラフのノードおよび要素は、コンピュータ読取り可能な媒体内に格納されるデータ構造、またはデータ収納庫内に格納されるデータモデルに準拠した他の体系化されたデータとして実装できる。

【0046】

このソフトウェアは、汎用または専用のプログラマブルコンピュータが読みとれるCD-ROMのような、または実行している場所のコンピュータにネットワークを介して配信(伝送信号に符号化されて)可能な媒体上で提供できる。機能は全て、専用コンピュータ上で、またはコプロセッサのような専用ハードウェアを用いて実行できる。本ソフトウェアは、ソフトウェアが規定する、計算の異なる部分を異なるコンピュータで実行する分散方式で実施することができる。このようなコンピュータプログラムはそれぞれ、コンピュータシステムが記憶媒体または装置を読みとって本明細書に記載する手順を実行する場合、コンピュータを構成し、動作させるために、汎用または専用プログラマブルコンピュータが読みとれる記憶媒体または装置(例えば、固体のメモリもしくは媒体または磁気もしくは光の媒体)に格納するか、ダウンロードするのが好ましい。本発明のシステムは、コンピュータプログラムとともに構成されるコンピュータ読取り可能な記憶媒体として、実装されるべきであると考えられることもでき、そのように構成された記憶媒体は、コンピュータシステムを特定のおよび所定の方法で動作させ、本明細書に記載の機能を実行させる。

【0047】

言うまでもなく、上記記載は説明用であって、本発明の範囲を制限するものではなく、その範囲は、付帯する特許請求の範囲により定義される。例えば、上記の幾つかの機能ステップは、処理全体へ実質的に影響を及ぼさずに異なる順序で実行できる。他の実施の形態は、以下の特許請求の範囲内にある。

【図面の簡単な説明】

【0048】

【図1】図1は、待ち行列システムの図である。

【図2A】図2Aは、揮発性および不揮発性メモリ空間へリサイクルリストをリンクするページテーブルの図である。

【図2B】図2Bは、メモリページの図である。

【図3A】図3Aは、揮発性および不揮発性メモリ空間内のメモリページの図である。

【図3B】図3Bは、揮発性および不揮発性メモリ空間内のメモリページの図である。

【図3C】図3Cは、図3Aの揮発性メモリ空間に対応するリサイクルリストおよび空きリストの図である。

【図3D】図3Dは、図3Aの揮発性メモリ空間に対応するリサイクルリストおよび空きリストの図である。

【図4A】図4Aは、揮発性および不揮発性メモリ空間内のメモリページの図である。

【図4B】図4Bは、揮発性および不揮発性メモリ空間内のメモリページの図である。

【図4C】図4Cは、図4Aの揮発性メモリ空間に対応するリサイクルリストの図である。

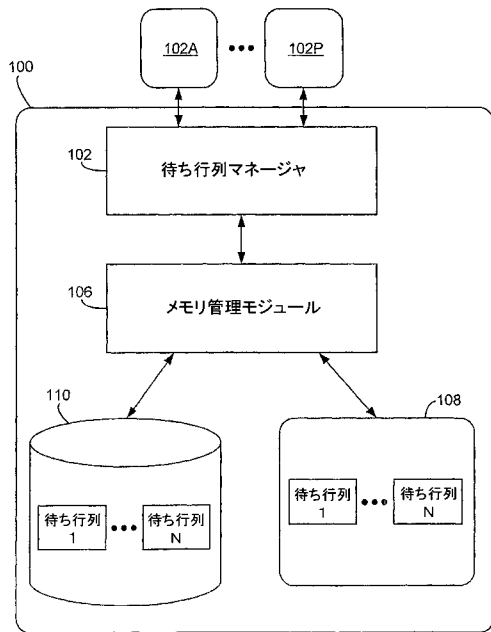
【図5A】図5Aは、揮発性および不揮発性メモリ空間内のメモリページの図である。

【図5B】図5Bは、揮発性および不揮発性メモリ空間内のメモリページの図である。

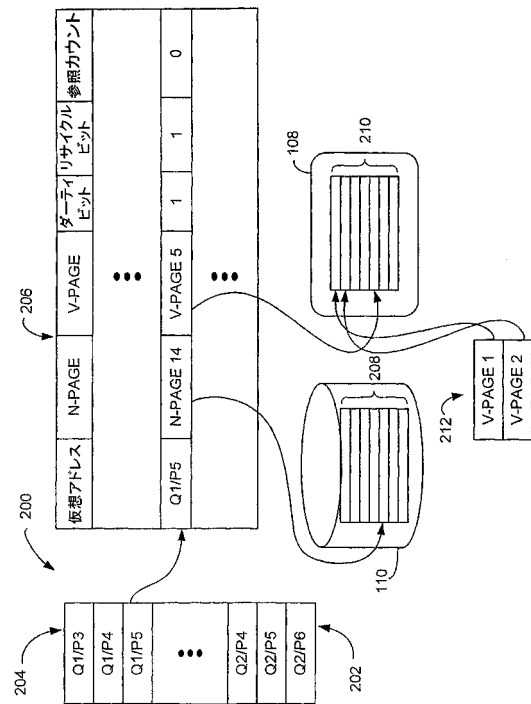
【図5C】図5Dは、図5Aの揮発性メモリ空間に対応するリサイクルリストおよび空きリストの図である。

【図5D】図5Dは、図5Aの揮発性メモリ空間に対応するリサイクルリストおよび空きリストの図である。

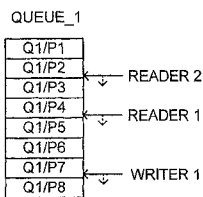
【図 1】



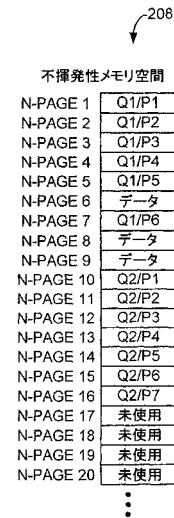
【図 2 A】



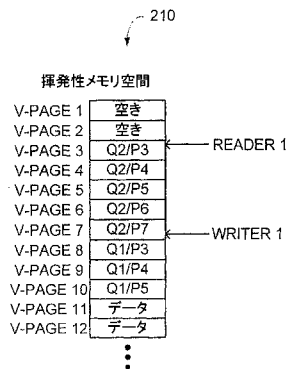
【図 2 B】



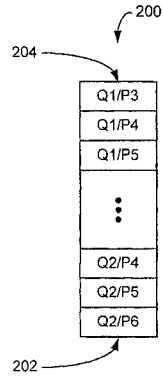
【図 3 B】



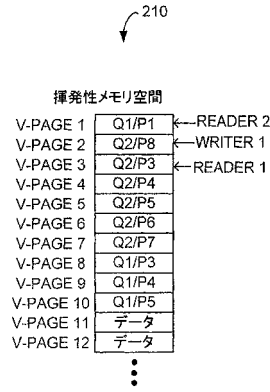
【図 3 A】



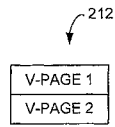
【図3C】



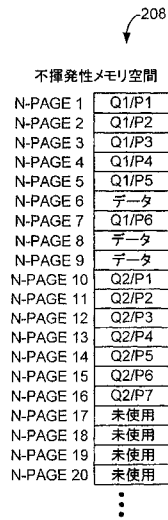
【図4A】



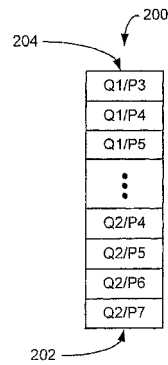
【図3D】



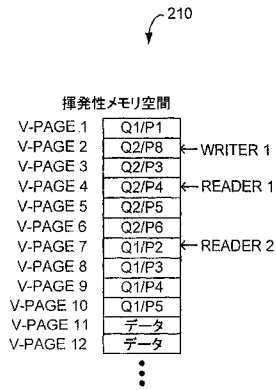
【図4B】



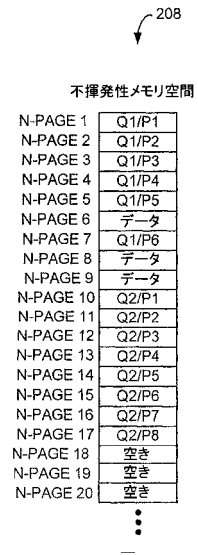
【図4C】



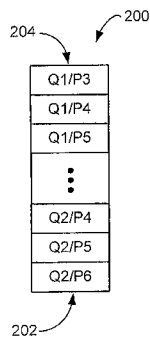
【図 5 A】



【図 5 B】



【図 5 C】



【図 5 D】

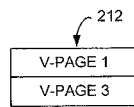


FIG. 5D

---

フロントページの続き

- (56)参考文献 特開昭55-097081(JP,A)  
特開昭63-008851(JP,A)  
特表2002-540502(JP,A)  
特開平07-168762(JP,A)  
特開2002-183074(JP,A)  
特開平05-046668(JP,A)  
特開昭60-086638(JP,A)  
特開2004-152173(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/08-12/12