



[12] 发明专利申请公开说明书

[21] 申请号 03127740.3

[43] 公开日 2004年4月14日

[11] 公开号 CN 1489059A

[22] 申请日 2003.8.13 [21] 申请号 03127740.3
 [30] 优先权
 [32] 2002.10.10 [33] US [31] 10/268,743
 [71] 申请人 国际商业机器公司
 地址 美国纽约州
 [72] 发明人 拉万·K·阿里米利
 约翰·S·多德森 桑吉夫·加伊
 肯尼思·L·赖特

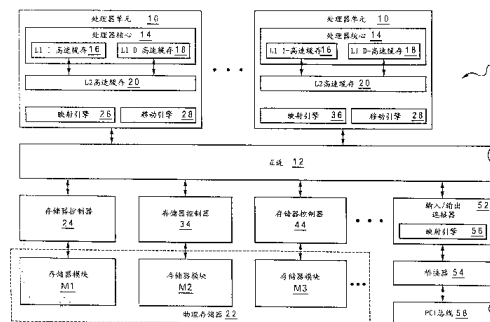
[74] 专利代理机构 北京市柳沈律师事务所
 代理人 邸万奎 黄小临

权利要求书 4 页 说明书 11 页 附图 4 页

[54] 发明名称 管理多处理器系统中虚拟化的物理存储器的方法和系统

[57] 摘要

处理器包含移动引擎和映射引擎，其透明地重新配置物理存储器，以完成存储器模块的添加、减去或替换。映射引擎寄存器存储来自和去向真实地址，其使得这些引擎能够虚拟化被重新配置的存储器模块的物理地址，并通过使用硬件功能而非软件来提供实时重新配置。使用来自和去向真实地址来选择来源和目标，移动引擎将要移走或重新配置的存储器模块中的内容复制到剩余的或插入的存储器模块中。然后，将与已重新配置的存储器模块关联的真实地址重新分配给接收所复制内容的存储器模块，从而创建从操作系统利用的可寻址真实地址空间向虚拟物理地址空间的虚拟化物理映射。



1. 一种计算系统，其耦接到具有多个存储器模块的物理存储系统，该多个存储器模块作为多个存储块存储数据，每个存储块包括物理存储器的相邻字节，并进一步耦接到至少一个存储器控制器，其中至少一个存储器控制器中的每一个存储器控制器与多个存储器模块中的一个或多个存储器模块耦接，其中至少一个存储器控制器中的每一个存储器控制器通过写入或读取在与其耦接的一个或多个存储器模块内的存储块，响应存储器存取，该计算系统包括：
- 5
- 10 处理器设备，用于产生包含与物理存储系统的存储位置关联的真实地址的存储器存取，以对其读和写数据；
- 寄存器，位于处理器设备内，具有：第一字段，用于存储对应于与其耦接的多个存储器模块中的第一存储器模块的来自（FROM）真实地址，其中耦接到第一存储器模块的第一存储器控制器被编程为响应寻址到来自真实地址的存储器请求；
- 15 和第二字段，用于存储对应于与其耦接的多个存储器模块中的第二存储器模块的去向（TO）真实地址，其中耦接到第二个存储器模块的第二存储器控制器被编程为响应寻址到去向真实地址的存储器请求；
- 移动引擎，位于处理器设备内，响应于有关修改第一和第二存储器模块的配置的通知，基于来自真实地址和去向真实地址，将多个存储块从第一存储器模块复制到第二存储器模块；以及
- 20 映射引擎，位于处理器设备内，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块期间，响应于处理器设备发出寻址到第一字段或第二字段之一中存储的真实地址的写入存储器请求，映射引擎发出寻址到来自真实地址和去向真实地址的写入存储器请求，并在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之后，将第二存储器控制器编程为响应寻址到来自真实地址的存储器请求。
- 25
2. 如权利要求 1 所述的计算系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块期间，响应于处理器设备请求寻址到来自真实地址的写入存储器请求，映射引擎发出既寻址到来自真实地址又寻址到去向真实地址的写入存储器请求。
- 30
3. 如权利要求 1 所述的计算系统，其特征在于在移动引擎将多个存储

块从第一存储器模块复制到第二存储器模块，而后映射引擎使得第二存储器控制器能够响应寻址到来自真实地址的存储器存取之后，从物理存储系统中移走第一存储器模块。

4. 如权利要求 3 所述的计算系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间中。

5. 如权利要求 1 所述的计算系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，在物理存储系统中插入第二存储器模块。

6. 如权利要求 5 所述的计算系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间之外。

7. 一种数据处理系统内的数据处理方法，其中数据处理系统包括处理器设备，用于产生包含与物理存储系统的存储位置关联的物理地址的存储器存取，以对其读和写数据；物理存储系统包括多个存储器控制器，通过对与其耦接的一个或多个存储器模块内存储的存储块写入和读取，每个存储器控制器响应由处理器设备请求的存储器存取，该方法包括：

在处理器设备中设置寄存器，以指示分别对应于第一和第二存储器模块的来自真实地址和去向真实地址，第一和第二存储器模块分别与第一和第二存储器控制器耦接；

基于来自真实地址和去向真实地址，将多个存储块从第一存储器模块复制到第二存储器模块；

在完成复制步骤之前，响应于处理器设备发出寻址到第一字段或第二字段之一中存储的真实地址的写入存储器请求，发出寻址到来自真实地址和去向真实地址的写入存储器请求；以及

在完成复制步骤之后，配置第一和第二存储器控制器，以便只响应寻址到新的真实地址的存储器存取。

8. 如权利要求 7 所述的方法，其特征在于，在复制步骤期间，响应于处理器设备发出寻址到当前真实地址的写入存储器请求，发出寻址到来自真实地址和去向真实地址的写入存储器请求。

9. 如权利要求 7 所述的方法，在配置步骤之后，还包括从物理存储系

统中移走第一存储器模块的步骤，其中来自真实地址在物理存储系统的当前可寻址空间内。

10. 如权利要求 7 所述的方法，在复制步骤之前，还包括在存储系统中插入第一存储器模块的步骤，其中去向真实地址在物理存储器的当前可寻址空间之外。

11. 如权利要求 7 所述的方法，所述方法还包括：向处理器通知多个存储器模块的配置正在被修改，其中响应于该通知而进行设置步骤。

12. 一种数据处理系统，包括：

物理存储系统，其具有多个用于以多个存储块存储数据的存储器模块，
10 每个存储块包括物理存储器的相邻字节；

多个存储器控制器，其中多个存储器控制器中的每一个存储器控制器与多个存储器模块中的一个或多个存储器模块耦接，其中多个存储器控制器中的每一个存储器控制器，通过写入或读取在与其耦接的一个或多个存储器模块内存储的存储块，响应存储器存取；以及

15 多个处理器设备，用于产生包含与物理存储系统的存储位置关联的真实地址的存储器存取，以对其读和写数据，其中多个处理器设备中的每一个处理器设备包括：

寄存器，具有：第一字段，用于存储对应于与其耦接的多个存储器模块中的第一存储器模块的来自真实地址，其中耦接到第一存储器模块的第一存储器控制器被编程为响应寻址到来自真实地址的存储器请求；
20 和第二字段，用于存储对应于与其耦接的多个存储器模块中的第二存储器模块的去向真实地址，其中耦接到第二存储器模块的第二存储器控制器被编程为响应寻址到去向真实地址的存储器请求；

25 移动引擎，响应于有关正在修改第一和第二存储器模块的配置的通知，基于来自真实地址和去向真实地址，将多个存储块从第一存储器模块复制到第二存储器模块；以及

映射引擎，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块期间，响应于处理器设备发出寻址到第一字段或第二字段之一中存储的真实地址的写入存储器请求，映射引擎发出寻址到来自真实地址和去向真实地址的写入存储器请求，并在移动引擎将多个存储块从
30 第一存储器模块复制到第二存储器模块之后，将第二存储器控制器编程

为响应寻址到来自真实地址的存储器请求。

13. 如权利要求 12 所述的系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块期间，响应于处理设备请求寻址到来自真实地址的写入存储器请求，映射引擎发出既寻址到来自真实地址又
5 寻址到去向真实地址的写入存储器请求。

14. 如权利要求 12 所述的系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块，而后映射引擎使得第二存储器控制器能够响应寻址到去向真实地址的存储器存取之后，从物理存储系统中移走第一存储器模块。

- 10 15. 如权利要求 14 所述的系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间中。

16. 如权利要求 12 所述的系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，在物理存储系统中插入第二
15 存储器模块。

17. 如权利要求 16 所述的系统，其特征在于在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间之外。

20

管理多处理器系统中虚拟化的
物理存储器的方法和系统

5

技术领域

本发明一般地涉及数据处理，特别涉及管理数据处理系统中的物理存储器。更特别地，本发明涉及通过操作系统控制自主地管理数据处理系统中的物理存储器的方法和系统。

10

背景技术

在计算机系统中，习惯上，在处理器产生的存储地址和系统的物理存储器中的特定区域之间具有一一对应的关系。这将操作系统和应用程序限制在由系统中安装的实际物理存储器确定的地址空间上。此外，许多现代计算机系统运行多个并行任务或处理，每个各有其自己的地址空间。将存储器的全部补码专用于每个任务和操作系统是昂贵的，尤其因为许多处理在任何给定时间都只使用其地址空间的一小部分。现代计算机系统已通过使用虚拟存储器而克服了这一限制，其实现了将程序地址(或虚拟地址)映射到实际存储地址的变换表。虚拟存储器使得程序能够在看上去很大而且连续的、全部专用于该程序的物理存储地址空间上运行。然而，实际上，多个程序或处理之间共享虚拟存储系统中的可用物理存储器。通过结合计算机硬件和软件而使处理中使用的虚拟地址转换为物理存储器的实际地址。这一处理被称为存储器映射或地址转换。

在虚拟存储系统中，最一般地，由操作系统软件(OS)进行存储器的分配。操作系统的功能之一是确保程序当前使用的数据和代码在主存储器中，并且转换表可以正确地将虚拟地址映射到实际地址。这需要指令序列的中断，以便使特权内核代码可以将物理存储器分配给被访问的区域，从而正常程序流可以无错误地继续。这一中断和分配物理存储器的内核处理需要大量处理时间，并且扰乱了通过CPU的指令的正常流水线操作。

当在计算机系统运转期间重新配置物理存储器时，管理物理存储器的操作系统的负担增加了。在物理存储器的大小增加或减小的情况下，或者当在

系统操作期间替换存储器模块时(例如, 当在需要替换的存储器模块中发生错误时), OS 需要临时中断正在处理的任

5 务, 修改转换表中的系统存储器配置信息, 并使用已改变的物理地址来将数据从坏存储设备存储到盘上, 然后重新配置剩余的存储设备。如果存储设备已被移走, OS 必须使移走的设备的物理地址空间无效, 并保持无效的地址空间, 以使其不能被使用, 实质上, 在存储系统的可寻址空间内留下了不可用的空间块。然后, 操作系统必须将逻辑地址映射到物理地址上, 以避免具有坏存储位置的页。这一问题增加了 OS 的开销, 并且使存储器的控制变得复杂。需要一种能够解决现有技术的上述问题, 并且能够迅速而又有效地根据需

10 要实现物理存储器的动态重新配置的物理存储器控制方法和系统。

发明内容

根据优选实施例, 提供了一种管理多处理器系统中虚拟化的物理存储器的方法和系统。多处理器系统耦接到具有多个用于以多个存储块存储数据的

15 存储器模块的物理存储系统, 每个存储块包括物理存储器的相邻字节, 并进一步耦接到至少一个存储器控制器, 其中至少一个存储器控制器中的每一个存储器与多个存储器模块中的一个或多个存储器模块耦接, 其中至少一个存储器控制器中的每一个存储器控制器通过写入和读取在与其耦接的一个或多个存储器模块内存储的存储块, 响应存储器存取。多处理器系统包括至少一个

20 处理器设备, 用于产生包含与物理存储系统的存储位置关联的真实地址的存储器存取, 以对其读和写数据。处理器设备内的寄存器具有: 第一字段, 用于存储对应于与其耦接的多个存储器模块中的第一存储器模块的来自真实地址, 其中耦接到第一存储器模块的第一存储器控制器被编程为响应寻址到来自真实地址的存储器请求; 和第二字段, 用于存储对应于与其耦接的多个

25 存储器模块中的第二存储器模块的去向真实地址, 其中耦接到第二存储器模块的第二存储器控制器被编程为响应寻址到去向真实地址的存储器请求。处理器设备内的移动引擎响应于有关正在修改第一和第二存储器模块的配置的通知, 基于来自真实地址和去向真实地址, 将多个存储块从第一存储器模块复制到第二存储器模块。在移动引擎将多个存储块从第一存储器模块复制到

30 第二存储器模块期间, 响应于处理器设备发出的寻址到第一字段或第二字段之一中存储的真实地址的写入存储器请求, 处理器设备内的映射引擎发出寻

址到来自真实地址和去向真实地址的写入存储器请求，并在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之后，将第二存储器控制器编程为响应寻址到来自真实地址的存储器请求。

5 在替换实施例中，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块期间，响应于处理器设备请求寻址到来自真实地址的写入存储器请求，映射引擎发出既寻址到来自真实地址又寻址到去向真实地址的写入存储器请求。

10 在另一个替换实施例中，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块，而后映射引擎使得第二存储器控制器能够响应寻址到来自真实地址的存储器存取之后，从物理存储系统中移走第一存储器模块。

在另一个替换实施例中，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间中。

15 在另一个替换实施例中，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，在物理存储系统中插入第二存储器模块。

在另一个替换实施例中，在移动引擎将多个存储块从第一存储器模块复制到第二存储器模块之前，来自真实地址位于物理存储系统的当前可寻址空间之外。

20 附图说明

在所附权利要求中阐明了本发明的确信新颖的特征。然而，通过参考附图详细描述本发明的示范实施例，本发明自身以及其优选使用模式、其它目的和优点将变得更加清楚，其中：

图 1 图解了根据本发明的优选实施例实现的数据处理系统；

25 图 2 展示了根据本发明的优选实施例，对如图 1 所示的数据处理系统的存储系统进行的存储地址转换处理的方框图表示；

图 3 展示了根据本发明的优选实施例，在将存储器模块从物理存储系统中移走的情况下，图 1 所示的数据处理系统的简化图；

30 图 4 展示了根据本发明的优选实施例，在将存储器模块从物理存储系统中移走的情况下，图 1 所示的数据处理系统的简化图。

具体实施方式

下面参考附图，特别参考图 1，其中图解了根据本发明的一个实施例，支持对虚拟化的物理存储器的存储器管理的多处理器(MP)数据处理系统的高级方框图。如图所示，数据处理系统 8 包括大量(例如 64 个)处理单元 10，通过系统互连 12 耦接以进行通信。每个处理单元 10 是包括一个或多个处理器核心 14 的集成电路。除了寄存器、指令流逻辑以及用于执行程序指令的执行单元，每个处理器核心 14 包括关联第一级(L1)指令和数据高速缓冲存储器 16 和 18，其分别临时缓冲指令和操作数(operand)数据，很可能由关联处理器核心 14 访问。

图 1 中还图解，数据处理器系统 8 的存储器分层体系还包括：物理存储器 22，物理存储器 22 包括一个或多个存储器模块(如存储器模块 M1、M2 以及 M3 所示)，其形成了存储器分层体系中最低级的易失性数据存储器；以及一个或多个较低级高速缓冲存储器，例如片上(on-chip)第二级(L2)高速缓冲存储器 20，用于将指令和操作数据从物理存储器 22 分阶段传递到处理器核心 14。本领域一般技术人员应该明白，与较高级相比，存储器分层体系中每个随后的较低级通常能存储更大量的数据，但是需要更多的存取等待时间。如图所示，通过存储器控制器 24、34 以及 44 与互连 12 接口的物理存储器 22 可以存储操作数数据、部分或更多操作系统以及一个或多个应用程序。存储器控制器 24、34 以及 44 分别耦接并控制对应的存储器模块 M1、M2 以及 M3 (尽管图中只展示了每个存储器控制器与单个存储器模块耦接，但是应该明白，存储器控制器可以控制存储系统 22 中多个存储器模块)。包括所有或部分存储器模块 M1、M2 以及 M3 的集合组成机器的操作系统和应用程序的一组物理存储器资源。

系统互连 12 可以包括一个或多个总线、转换器结构或其它连接架构，其用于在耦接到系统互连 12 上的设备(例如处理单元 10、存储器控制器 24、34、44 等)中通信的渠道。系统互连 12 上的典型事务以请求开始，请求可以包括指示事务类型的事务字段、一个或多个指示事务来源和/或计划的接收者的标记、以及地址和/或数据。每个连接到系统互连 12 的设备最好探查系统互连 12 上的所有事务，并且如果合适的话，用探查响应来响应请求。这些行为可以包括确定系统互连 12 上的数据的来源、存储通过请求探查器提供的数据、使高速缓冲存储器的数据无效等。输入/输出连接器 52 也耦接到互连 12，并

通过桥接器 54 提供在耦接到互连 12 的其它设备和耦接到 PCI 总线 58 的外部设备之间通信的渠道。

数据处理系统 8 利用虚拟存储系统，其实现了将程序地址(或有效地址)映射到实际存储地址的转换表。虚拟存储系统使得可用物理存储器能够在多个程序或处理之间共享。通过提供将有效地址(EA)转换为物理地址(PA)的地址转换机制，处理单元 10 使得处理单元的地址空间(“逻辑地址空间”)能够具有与可用物理存储器 22 不同的大小，这里，物理地址指实际物理存储器中的位置。该虚拟存储系统还允许多个程序能够同时驻留在系统存储器中，而各自不需要了解其物理基地址的位置。相反地，这多个程序只需要知道其逻辑基地址。此外，不需要尝试保持对每个可能的有效地址转换或映射，虚拟存储系统将有效和物理存储器划分为多个块。在许多系统中，这些块大小固定，并称作节(section)或页(page)。单个页中的地址都具有相同的最高位。因此，存储地址由对应于地址较高位的页号和对应于地址较低位的页偏移串联而成。

通常，数据结构保持在物理存储器中，以从有效页号转换为真实页地址。这些数据结构经常采用转换表的形式，通常称作段表(segment table)和页表。段表通过有效页地址或号码来索引，并且通常具有大量对应于有效地址空间中的页的条目。每个条目都是特定页号或有效页地址向虚拟页地址的映射。页表通过虚拟页地址或号码来索引，并且通常具有大量对应于虚拟地址空间中的页的条目。每个条目都是特定页号或虚拟页地址向真实页地址的映射。

在处理器单元 10 中使用专门的硬件高速缓冲存储器(未示出)来进行有效到真实地址转换，该硬件高速缓冲存储器被称作地址转换高速缓冲存储器或转换后备缓冲器(TLB)，其专用于转换。TLB 是快速而又小型的静态存储器，用于存储页表中最常被引用的条目。其通常具有固定数量的条目。当处理存储器请求时，计算机首先尝试在 TLB 中发现合适的地址转换。如果未发现这样的地址转换，则自动访问页表，以检索适当的转换。

本领域一般技术人员将明白，数据处理系统 8 可以包括许多附加的未示出的组件，例如 I/O 适配器、互连桥接器、非易失性存储器、用于连接到网络或附属设备的端口等。因为这样的附加组件对理解本发明没有必要，所以图 1 中没有示出，这里也没有讨论。然而，应该明白，本发明提供的增强可应用于任何架构的数据处理系统，并且决不限于图 1 所示一般化的 MP 架构。

现在参考图 2，其中展示了根据本发明的优选实施例，存储管理系统为数据处理系统 8 的虚拟存储系统进行的存储器映射。有效地址空间(Effective Address Space)表示其中有多处理在彼此独立操作的处理器单元 10 上操作的地址空间。图中展示了三个处理，处理(Process) 1 (P1)、处理 2(P2)以及处理 3(P3)，各有其自己的逻辑地址空间。对于每个处理，由有效地址对其可寻址空间中的页或块寻址。如图 2 所示，对三个执行处理，物理地址空间(Physical Address Space)中当前载入页在有效地址空间内的有效地址分别被示为 P1-EA、P2-EA 以及 P3-EA。每个有效地址是通过段表转换(Segment Table Translation)204 而转换的 64 位地址。由处理器核心 14 内的段后备缓冲器(SLB)或段后备寄存器(SLR)进行段表转换 204。由 SLB 或 SLR 将应用程序地址空间 P1、P2 以及 P3 内的每个可寻址空间转换为虚拟地址空间(Virtual Address Space)206 内的 80 位虚拟地址。从而，P1-EA、P2-EA 以及 P3-EA 通过段表转换而分别转换为 P2-VA、P1-VA 以及 P3-VA。然后，使用处理器核心 14 内的转换后备缓冲器(TLB)，每个虚拟地址空间 P2-VA、P2-VA 以及 P3-VA 通过页表转换(Page Table Translation)208 而转换到真实地址空间(Real Address Space)210 中，从而将每个 80 位虚拟地址 P1-VA、P2-VA 以及 P3-VA 转换为其代表系统存储器内的真实地址的 64 位真实地址 P1-RA、P2-RA 以及 P3-RA。真实地址 RA 通常由两个字段组成：将物理页号编码为数字“i”个位的字段，例如高阶位；和将从物理页开始起的偏移编码为数字“j”个位的字段，例如低阶位。二的数字 j 次幂表示页的大小，例如数字 j 等于十二表示页大小为四千字节。二的数字 i 次幂表示以页数表示的物理存储器的大小，例如数字 i 等于二十表示一个物理百万页的数量，或四吉字节(gigabytes)的物理存储器。

如图 1 所示，存储器控制器 24、34 以及 44 进行对真实地址的物理映射，以访问存储器模块 M1、M2 以及 M3 中所请求的页。因此，如果在 L1 高速缓冲存储器 16、18 以及 L2 高速缓冲存储器 20 中没有发现真实地址空间 210 中已寻址的页，在互连 12 上请求存储器存取。一旦在互连 12 中检测到存储器存取请求，每个存储器控制器 24、34、44 核对由存储器存取寻址的真实地址空间。由存储器控制器 24、34 以及 44 将来自处理器单元的真实地址的高阶位解码，存储器控制器 24、34 以及 44 识别每个存储器控制器的相应存储器模块 M1、M2、M3 的可寻址真实空间。每个存储器控制器 24、34 以及 44 响应已寻址到其对应的存储器模块的存储器存取。

如图 2 所示, 这样由存储器控制器 24、34、44 进行物理映射 212。物理映射 212 为地址页 P1-RA、P2-RA 以及 P3-RA 转换真实地址, 并将其分别映射到对应的物理地址 P1-PA、P2-PA 以及 P3-PA, 而物理地址 P1-PA、P2-PA 以及 P3-PA 表示所请求的那些页在对应的存储器模块 M1、M2 以及 M3 中的物理地址。物理地址指示存储寻址信息的存储器模块内的特定存储位置。例如, P2-PA 指定特定的行和列地址, 以唯一地识别存储器模块 2 中已寻址的页。这一物理映射机制对操作系统 OS 是不可见的, 操作系统 OS 借助于真实地址查看所有物理存储器资源, 而不将这些资源的位置推理地辨别为特定的系统存储器 22 内的存储器模块 M1、M2、M3。

返回参考图 1, 在每个处理器单元 10 中包含有映射引擎 36 和移动引擎 28。根据本发明, 映射引擎 36 和移动引擎 28 提供对物理存储器的虚拟化功能, 以使得能够有效地重新配置物理存储器 22。当物理存储器 22 被重新配置之后, 例如当在系统中插入、移走或替换存储器模块 M1、M2 和 M3 中的一个时, 移动引擎 28 在物理存储器 22 的存储器模块之间进行数据传输, 而映射引擎 36 控制存储器模块 M1、M2、M3 的从属存储器模块的真实寻址, 以使得能够添加、减去或替换特定的存储器模块。这一存储器管理在硬件/固件级有效地完成, 只需要少量的操作系统资源来完成物理存储器的重新配置。在操作中, 每个映射引擎 36 为正在被重新配置的所选择的存储器模块提供真实地址空间(具体说来, 高阶真实地址位)的可配置分配, 这有效地改变了该存储设备的基地址。

图 3 在数据处理系统 8 的简化图中图解了从物理存储器移走存储器模块的实施例。下面将说明, 处理器的移动引擎结合关联的映射引擎工作, 以在将关联的存储器模块物理移走之前使其脱机。一般地, 移动引擎将要移走的存储器模块的内容复制到物理存储器中剩余的存储器模块中。然后, 将该存储器模块的真实地址重新分配给接收所复制内容的存储器模块。

在此例子中, 将存储器模块 M2 从数据处理系统 8 中移走。第一步, 处理器单元 10 向操作系统报告其总可用物理存储器现在减少了一个存储器模块。例如, 如果每个存储器模块 M1、M2、M3 都是 64 吉位(GB)的存储设备, 则向操作系统通知现在其可用物理存储器为 128 GB。相应地, 操作系统立即开始调出页, 以减少所存储数据的量。处理器单元 10 向所有移动引擎 28 和映射引擎 36 通知, 存储器模块 M2 正在从物理存储器 22 中移走。移动引擎

28 立即选择将用于存储包含在存储器模块 M2 中的数据的剩余的一个或多个模块。

5 处理器单元 10 内的每个映射引擎 36 包含寄存器 305，用于存储移走的存储器模块的“当前”真实地址和移走的存储器模块的“新”真实地址(在这里使用的真实地址指整个真实地址或需要唯一地识别用于存储由存储器的已索引块寻址的数据的关联存储器模块的那一部分(例如高阶位))。每个处理单元 10 按需要载入其各自的寄存器 305，以进行给定的存储器重新配置。

10 如图 3 所示，映射引擎 36 包含寄存器 305，其具有：字段 306，其包含存储器模块 M2 的 FROM(来自)真实地址；和字段 308，其包含保留了移走的模块的存储内容的相应存储器模块(在此例中是存储器模块 M3)的 TO(去向)真实地址。寄存器 305 包含：字段 306，其显示存储器模块 M2 的来自真实地址为 RA2；和字段 308，其包含存储器模块 M2 的去向真实地址，即 RA3。

15 在图 3 所示的例子中，移动引擎 28 选择存储器模块 M3 接收存储器模块 M2 中存储的数据。存储器模块 M1 保持在线，并且不从存储器模块 M2 接收任何数据。在载入寄存器 305 之后，移动引擎 28 开始“移动处理”，以将来自真实地址(RA2)处的存储器地址空间的内容复制到去向真实地址(RA3)处的存储器地址空间。相应地，移动引擎 28 通过在互连 12 上向存储器控制器发出存储器存取请求，而开始将存储器模块 M2 中的内容复制到存储器模块 M3 中，以将存储器模块 M2 的每个存储单元复制到真实地址 RA3 处的存储器地址空间。从而，移动引擎 28 将由真实地址 RA2 寻址的存储器模块 M2 中的所有存储内容复制到由真实地址 RA3 寻址的存储器模块 M3 中，如路径 325 所示。在替换实施例 20 中，移动引擎 28 将存储器模块 M2 的部分内容复制到存储器模块 M3 中，而将剩余部分复制到存储系统 22 的其它存储器模块中(例如存储器模块 M1)。

25 在移动存储器存储内容的处理期间，存储器控制器 24、34 以及 44 继续响应互连 12 上的存储器存取请求。根据优选实施例，映射引擎 36 通过提供向存储器模块 M2 和/或 M3 的物理地址的映射来执行这样指向当前真实地址空间的存储器存取，而使其对应的处理器能发出存储器请求，例如被寻址到当前真实地址处存储的存储器的“Read(读)”和“Write(写)”，如来自真实地址字段 306 所指示。在图 3 的例子中，存储器模块 M2 将继续响应寻址到真实地址 RA2 的 Read，如路径 327 所示。根据优选实施例，映射引擎 36 将对

30

当前真实地址空间的 Write 请求从其关联的处理器映射到来自真实地址和新真实地址，如来自和去向真实地址字段 306、308 所分别指示。相应地，如图 3 中路径 329 所示，因为存储器模块 M2 的来自真实地址为 RA2(如字段 306 所示)，而其去向真实地址为 RA3(如字段 308 所示)，所以对真实地址 RA2 的 Write 指向存储器控制器 34 和 44 两者。因为对 RA2 的存储器写入继续存储在存储器模块 M2 和 M3 两者中，所以存储系统确保在整个移动处理中具有一致性。

一旦完成从存储器模块 M2 向存储器模块 M3 的数据传输，移动引擎 28 就通过将去向真实地址复制到来自真实地址字段 306 中、或通过重置映射引擎 36，来更新寄存器 305。然后存储器控制器 44 被重新编程，以响应先前由存储器控制器 34 管理的真实地址 RA2。现在，存储器模块 M1 由真实地址 RA1 寻址，而存储器模块 M3 由真实地址 RA2 寻址。现在，存储器控制器 44 进行 RA2 空间中的真实地址向存储器模块 M3 的直接物理映射，从而创建从操作系统使用的可寻址真实地址空间向虚拟物理地址空间的虚拟化物理映射。

下面参考图 4，其中展示了数据处理系统 8 的简化方框图，其描绘了插入物理存储器 22 中的存储器模块 M2。每个处理器单元 10 向其对应的移动引擎 28 和映射引擎 36 通知：存储器模块正在被添加到物理存储器 22 中，并且要插入与存储器控制器 34 关联并由其控制的存储器槽中。在这一点上，数据处理系统 8 的当前真实地址空间由存储器模块 M1 和 M3 的物理存储器组成，并且可以由真实地址空间 RA1-RA2 寻址。

根据优选实施例，每个处理器单元 10 内的每个寄存器 305 用对应的当前和新真实地址来编程。分配给存储器模块 M2 的去向真实地址用物理存储器 22 的当前真实地址空间之外的真实地址(此例中为 RA4)来编程。来自真实地址是在可实寻址空间插入和重新编程之后，将被分配有插入的存储器模块的地址。字段 306 用来自真实地址 RA2 来编程，而字段 308 用去向真实地址 RA4 来编程。对与插入的存储器模块关联的存储器控制器(此例中为存储器控制器 34)编程以响应真实地址 RA4。应该明白，为每个字段选择的真实地址取决于插入的存储器模块 M2 的大小和现有存储器模块 M3 的大小。在此例中，假设其大小相同。在其大小不同的情况下，将存储器模块 M3 的真实地址选择落在存储器模块 M2 的存储器边界上，以提供邻接的真实存储空间。

在将存储器模块 M2 物理地插入物理存储器 22 之后，并且在载入寄存器 305 之后，移动引擎 28 开始“移动处理”，以将来自真实地址(RA2)处的存储地址空间的内容复制到去向真实地址(RA4)处的存储地址空间。相应地，通过在互连 12 上向存储器控制器 44 发出存储器存取请求，移动引擎 28 开始将存储器模块 M3 的内容复制到存储器模块 M2，以将存储器模块 M3 的每个存储单元复制到真实地址 RA4 处的存储器地址空间。存储器控制器 44 从存储器模块 M3 的每个存储单元接收写入，并将其映射到存储器模块 M2 的物理地址空间。这在图 4 中由路径 425 表示，其展示了真实地址 RA2 处的存储单元向真实地址 RA4 处的存储单元的复制。

10 在移动处理期间，将映射引擎 36 编程为响应被寻址到来自真实地址的处理器单元 10 进行的存储器存取，而来自真实地址被编程到其寄存器 305 字段 306 处。如图 4 所示，由被编程为响应真实地址空间 RA2 的存储控制器 44 为从处理器单元 10 指向真实地址 RA2 的 Read 存储器存取继续服务。从处理器单元 10 指向真实地址 RA2 的 Write 存储器存取将引起映射引擎 36 向两个
15 分别响应对寄存器 305 中的来自和去向真实地址的存储器存取的存储器控制器(此例中为存储器控制器 34 和 44)都发出 Write 存储器请求。这些存储器控制器通过物理地映射到对应的行与列物理地址，将所接收到的数据写入到与其耦接的存储器模块。如图 4 所示，路径 429 描绘了从处理器单元 10 向真实地址 RA2 的写入请求，存储器控制器 34 和 44 都接收该请求，引起存储器模
20 块 M2 和 M3 中的每个都被写入的数据更新。应该明白，这一机制允许移动处理期间的存储一致性。

一旦完成了存储器模块移动处理，移动引擎 28 指引存储器控制器 34 响应从互连 12 对字段 306 中的来自真实地址处的存储器存取，并指引存储器控制器 44 响应从互连 12 对已重新配置的真实存储器空间内另一个真实地址处的
25 的存储器存取。在此例中，将存储器控制器 44 重新编程为真实地址 RA3，以提供邻近的真实可寻址存储器 RA1-RA3。然后操作系统被通知：系统的真实地址空间增加了等于存储器模块 M2 的可寻址空间的量。然后操作系统开始跨越存储器模块 M1、M2 以及 M3 的真实地址空间 RA1-RA3 存储和存取存储页。

30 应该明白，有些情况下，插入的存储器模块与现有存储器的存储器边界相匹配，而真实地址空间不必重新配置。因此，在替换实施例中，不进行存

储移动处理，而映射引擎 36 立即对与插入的存储器模块关联的存储器控制器编程，以响应与新添加的真实地址空间关联的去向真实地址。例如，如果将图 4 所示的存储器模块 M2 添加到物理存储器 22，那么将用真实地址 RA3 对字段 306 和 308 编程。在这种情况下，不将存储器模块 M3 中存储的数据复制

5 复制到存储器模块 M2，而保留在存储器模块 M3 中。一旦在物理存储器 22 的存储器插槽中添加了存储器模块 M2，就向操作系统通知，其真实地址空间增加了等于存储器模块 M2 的存储器存储量的量，而存储器控制器 34 立即开始响应对真实地址 RA3 的存储器存取。

在替换物理存储器 22 中的存储器模块的情况下，根据结合图 3 描述的处理，从物理存储器 22 中移走存储器模块，然后，根据结合图 4 描述的处理，向物理存储器插回新的存储器模块。同样，现在返回参考图 1，输入/输出连接器 52 内的映射引擎 56 以与映射引擎 36 相同的方式操作。当进行对已根据

10 优选实施例重新配置的存储器模块进行直接存储器存取操作时，输入/输出连接器 52 以与处理器单元 10 相似地方式操作。应该明白，系统可以具有附加的输入/输出连接器，与输入/输出连接器 52 相同，连接到互连 12。随着在 PCI 总线 58(或其它类似的从属总线)上添加或移走诸如盘驱动器和视频监视器的各种输入/输出设备，映射引擎 56 连同移动引擎 28 操作，以准许对该输入/输出设备的物理地址虚拟化，其操作方式与上面描述的在物理存储器 22 中移走或添加存储器模块的方式相同。

应该明白，优选实施例为要在物理存储器 22 中插入、移走或替换的存储器模块做准备，以完成该物理存储器改变，而操作系统不必指引或控制物理存储器的重新配置。在优选实施例中，引动引擎 28 和映射引擎 36 一同工作，以透明地重新配置物理存储器，以完成在物理存储器中特定存储器模块的添加、减去或替换。每个映射引擎寄存器存储当前或新的真实地址，其使得移

15 动和映射引擎能够虚拟化被重新配置的存储器模块的物理地址，并通过使用硬件功能性而不使用软件，实时提供对所存储的数据的重新配置和操纵。

尽管参考本发明的优选实施例具体示出和描述了本发明，但是本领域一般技术人员应该明白，在不脱离本发明的精神和范围的情况下，可以对其进行形式和细节上的各种修改。

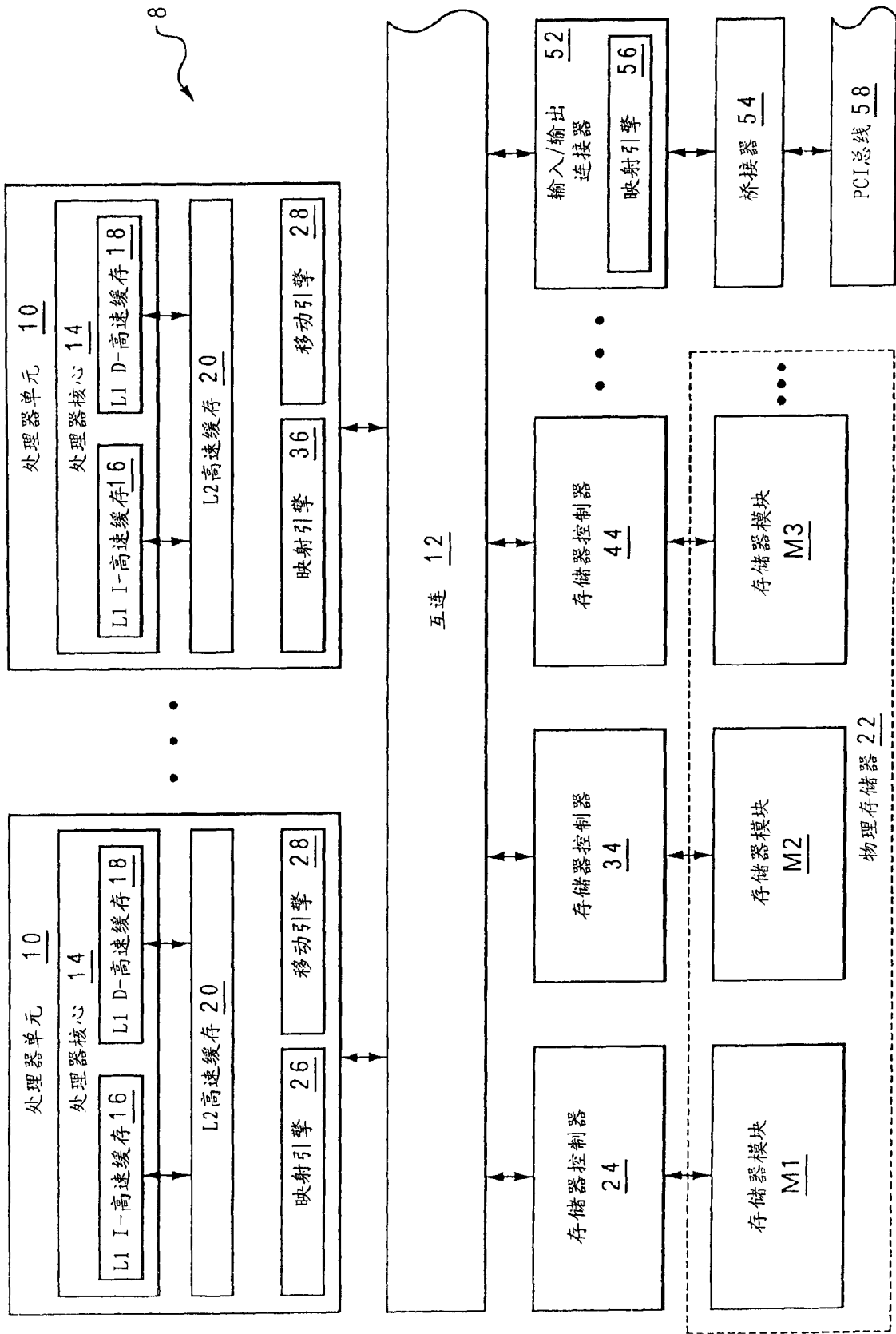


图 1

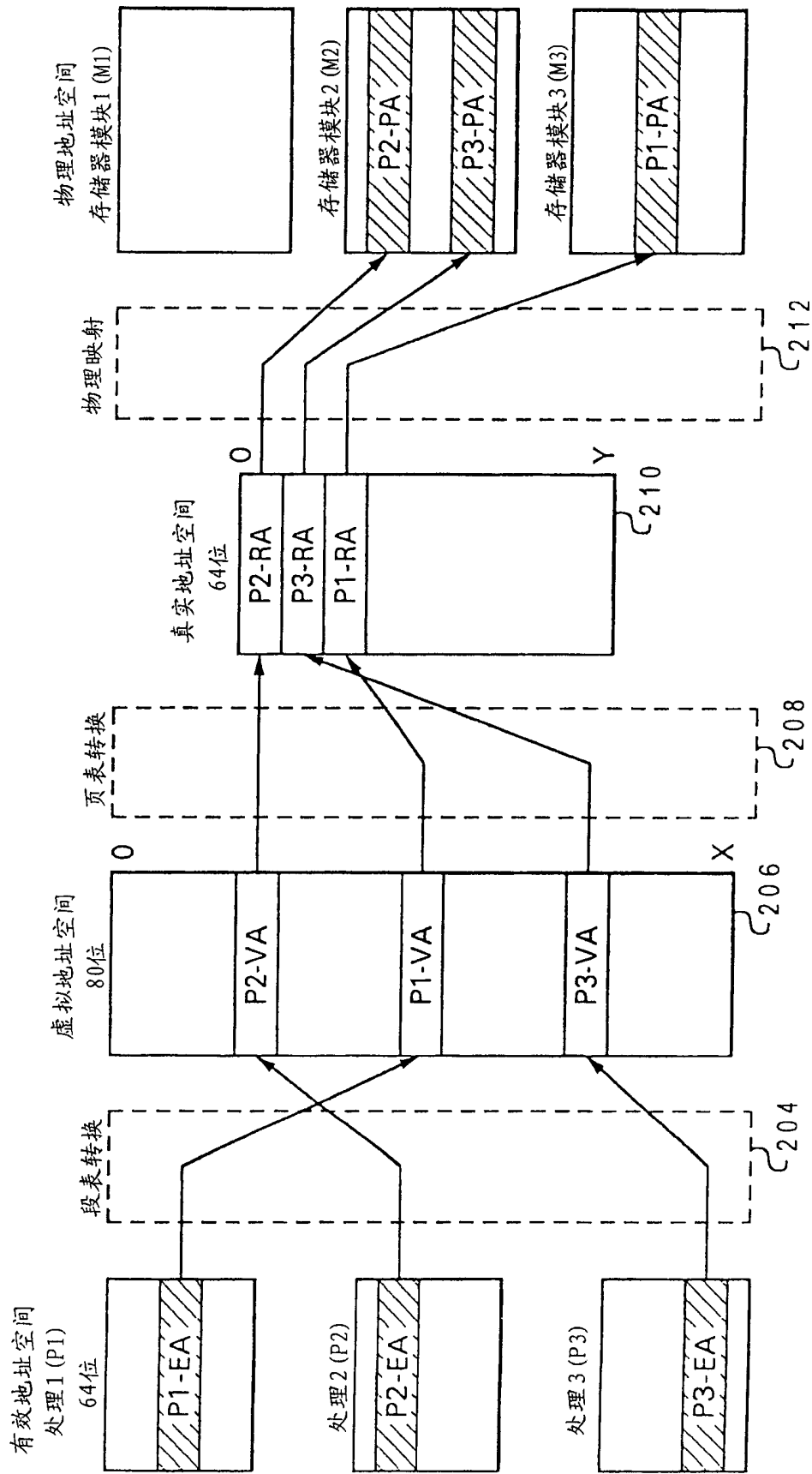


图 2

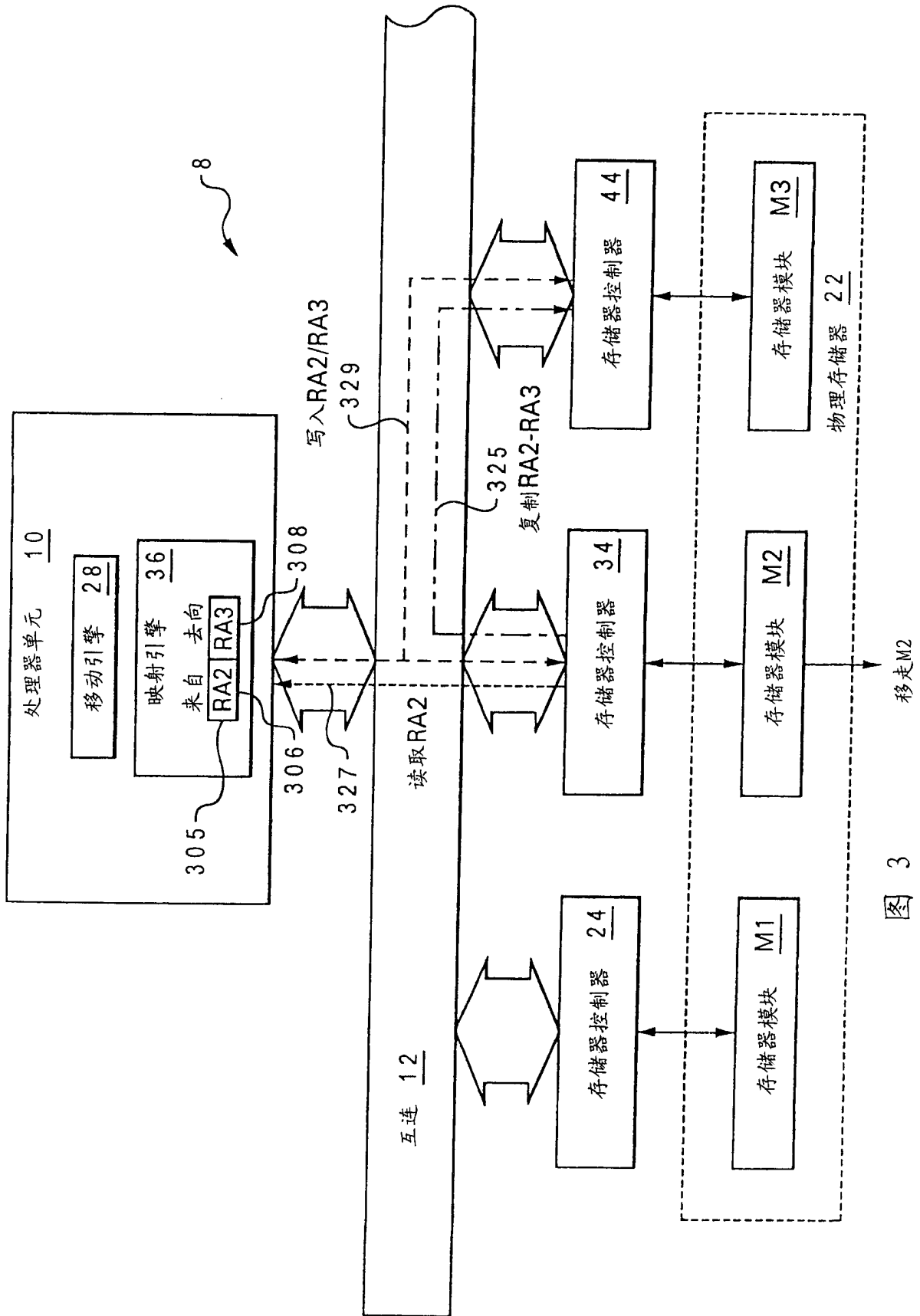


图 3

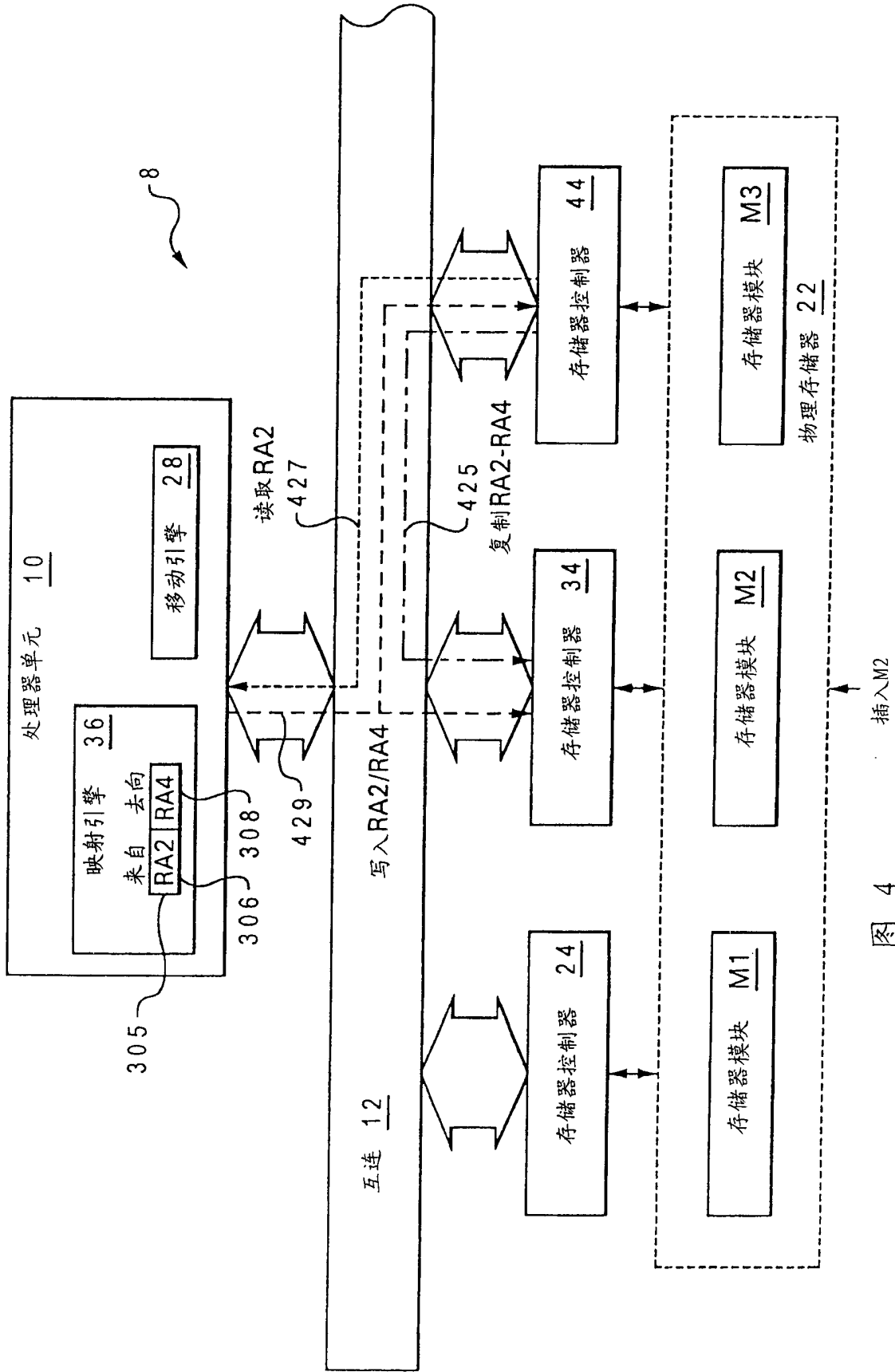


图 4