



(12) 发明专利

(10) 授权公告号 CN 102789429 B

(45) 授权公告日 2016. 06. 22

(21) 申请号 201210167652. 5

(22) 申请日 2008. 05. 30

(30) 优先权数据

11/757, 103 2007. 06. 01 US

(62) 分案原申请数据

200810100050. 1 2008. 05. 30

(73) 专利权人 英特尔公司

地址 美国加利福尼亚州

(72) 发明人 O·法里克 B-Z·弗莱德曼

J·D·多维科 E·维斯曼

J·B·克罗斯兰德

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 姬利永

(51) Int. Cl.

G06F 12/10(2016. 01)

(56) 对比文件

US 5918251 A, 1999. 06. 29,

US 5918251 A, 1999. 06. 29,

US 2005/0154854 A1, 2005. 07. 14,

CN 85106711 A, 1987. 02. 04,

US 2006/0206687 A1, 2006. 09. 14,

WO 2007/139529 A1, 2007. 12. 06,

CN 1784663 A, 2006. 06. 07,

审查员 徐菲

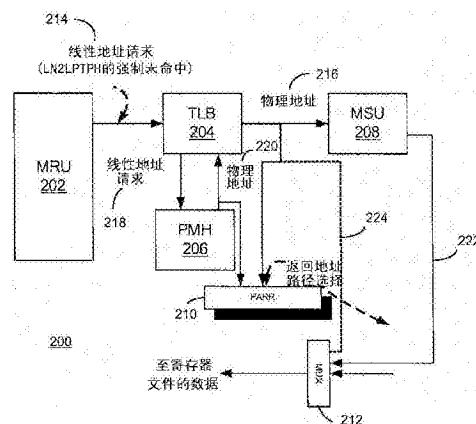
权利要求书3页 说明书7页 附图4页

(54) 发明名称

支持页属性的虚拟地址至物理地址的转换

(57) 摘要

一种支持页属性的虚拟地址至物理地址的转换。本发明的实施例一般涉及用于支持页属性的线性至物理地址转换的系统、方法和装置。在某些实施例中,系统接收一指令以将存储器指针转换成存储器位置的物理存储器地址。该系统可返回物理存储器地址和一个或多个页属性。描述并要求保护包括其它实施例。



1. 一种机器实现方法,包括:

接收指令以将虚拟存储器指针转换成存储器位置的物理存储器地址;

提供标志,所述标志标识转换备用缓冲器与哪个地址空间相关联,这使得能够在异步调用中使用相关联的转换备用缓冲器项;

基于页表信息将所述虚拟存储器指针转换成所述物理存储器地址,包括读取所述转换备用缓冲器项以获取所述物理存储器地址和一个或多个页属性;以及

返回所述物理存储器地址和所述一个或多个页属性而无需访问存储在所述物理存储器地址的数据。

2. 如权利要求1所述的方法,其特征在于,基于所述页表信息将所述虚拟存储器指针转换成所述物理存储器地址包括不基于页表信息,确定与所述虚拟存储器指针相对应的页是否在存储器中。

3. 如权利要求1或2所述的方法,其特征在于,所述接收指令以将所述虚拟存储器指针转换成所述存储器位置的物理存储器地址包括从虚拟机管理器接收所述指令;以及

其中所述返回所述物理存储器地址和所述一个或多个页属性而无需访问存储在所述物理存储器地址的所述数据包括返回主机物理存储器地址。

4. 如权利要求1或2所述的方法,其特征在于,所述接收指令以将所述虚拟存储器指针转换成所述存储器位置的物理存储器地址包括从虚拟机管理器接收所述指令;以及

其中所述返回所述物理存储器地址和所述一个或多个页属性而无需访问存储在所述物理存储器地址的所述数据包括返回客机物理存储器地址。

5. 如权利要求1或2所述的方法,其特征在于,基于所述页表信息将所述虚拟存储器指针转换成所述物理存储器地址包括:

如果转换备用缓冲器不包括所述物理存储器地址,读取至少一个页表项以获取所述物理存储器地址和所述一个或多个页属性。

6. 如权利要求1或2所述的方法,其特征在于,所述一个或多个属性包括以下至少一个:

存在位,

读/写位,

用户/管理员位,

访问位,

脏位,

全局标志,

执行禁用位,

页大小指示符,以及

转换表错误。

7. 如权利要求1或2所述的方法,其特征在于,所述返回所述物理存储器地址和所述一个或多个页属性而无需访问存储在所述物理存储器地址的所述数据包括:

在专用寄存器中存储所述物理存储器地址和所述一个或多个页属性;以及

访问所述专用寄存器。

8. 如权利要求1或2所述的方法,其特征在于,所述返回所述物理存储器地址和所述一个或多个页属性而无需访问存储在所述物理存储器地址的所述数据包括:

经由读取返回总线返回所述物理存储器地址和所述一个或多个页属性。

9. 如权利要求1或2所述的方法, 其特征在于, 所述接收指令以将所述虚拟存储器指针转换成所述存储器位置的物理存储器地址包括:

接收异步指令以将所述虚拟存储器指针转换成所述物理存储器地址, 所述异步指令包括用于定义所述异步指令的上下文的操作数。

10. 一种用于执行虚拟地址至物理地址转换的设备, 包括:

存储器请求单元, 被配置为用于发出存储器请求, 所述存储器请求包括由指令所指定的虚拟存储器指针, 所述指令用于将所述虚拟存储器指针转换成存储器位置的物理存储器地址;

转换备用缓冲器, 耦合至所述存储器请求单元, 被配置为接收所述存储器请求, 基于来自页表项的信息将所述虚拟存储器指针转换成所述物理存储器地址, 包括读取转换备用缓冲器项以获取所述物理存储器地址和一个或多个页属性, 其中标签用于标识所述转换备用缓冲器与哪个地址空间相关联, 以使得能够在异步调用中使用相关联的转换备用缓冲器项; 以及

寄存器, 耦合至所述转换备用缓冲器, 被配置为存储所述物理存储器地址和所述一个或多个页属性,

其中, 所述设备用于根据诸个页表末级中的状态来确定与所述虚拟存储器指针相对应的页是否在存储器中, 且不访问所述存储器中所述物理存储器地址上的数据。

11. 如权利要求10所述的设备, 其特征在于, 所述设备进一步用于不基于页表信息, 确定与所述虚拟存储器指针相对应的页是否在所述存储器中。

12. 如权利要求10或11所述的设备, 其特征在于, 所述指令将从虚拟机管理器接收的; 以及

其中所述物理存储器地址将包括主机物理存储器地址。

13. 如权利要求10或11所述的设备, 其特征在于, 所述指令将从虚拟机管理器接收的; 以及

其中所述物理存储器地址将包括客机物理存储器地址。

14. 如权利要求10或11所述的设备, 其特征在于, 还包括: 页未命中处理器, 用于如果转换备用缓冲器不包括所述物理存储器地址, 读取所述页表项以获取所述物理存储器地址和所述一个或多个页属性。

15. 如权利要求10或11所述的设备, 其特征在于, 所述一个或多个属性包括以下至少一个:

存在位,

读/写位,

用户/管理员位,

访问位,

脏位,

全局标志,

执行禁用位,

页大小指示符, 以及

转换表错误。

支持页属性的虚拟地址至物理地址的转换

[0001] 本申请是申请日为2008年5月30日、申请号为200810100050.1、题为“支持页属性的虚拟地址至物理地址的转换”的发明专利申请的分案申请。

技术领域

[0002] 本发明的实施例一般涉及计算系统领域,尤其涉及用于支持页属性的虚拟地址至物理地址转换的系统、方法和装置。

背景技术

[0003] 虚拟存储器允许程序员使用比处理器可用的实际物理存储器所提供的范围更大的存储器用于程序和数据。此外,虚拟存储器允许程序相互平行地加载,并具有与其它程序的存在及其加载的位置无关的存储器映射。计算系统利用地址转换硬件将程序的虚拟地址(在IA32体系结构中也称为线性地址——在本说明书中“线性地址”或LA与“虚拟地址”在使用时意思相同)映射到真实的硬件存储器地址(例如,物理存储器地址)。该硬件使用存储器中的表树作为地址转换的输入数据。该树的根由保持页表树中第一表的物理地址的寄存器所指向。这一寄存器的例子是IA32体系结构中的CR3。页表项(PTE)利用基址和索引来寻址。基址是从寄存器或先前的表存储器指针中取出的。索引利用线性地址的一部分。PTE或者包括页——如果存储器中不存在其余的子树,或者包括以下讨论的存储器指针和其它信息。存储器指针针对存储器中可以包括数据(属于应用程序或操作系统)或另一级页表的页。如果是后一种情况,则线性地址的另一部分用于索引至类似于所述内容的方案中的表。如果地址针对应用程序页,则物理地址通过将剩余的地址位(没有用于索引)添加至从页表项检索的页基址来构造。同样,某些实施例可例示出如寄存器中所述的几种转换方案(例如,不同的表树索引结构),为了简化本说明书,我们将考虑这些CR3寄存器中的信息部分,即使它可以保持在一个或多个其它寄存器中。

[0004] 除地址转换信息外,这些表包括诸如访问权读取、写入或执行、存储器中数据的存在、高速缓存策略指示、页修改状态等信息。在某些情形中,页表可包括不同大小的页,其中较大的页指向页表树的较低级(而不是指向另一指针页)。将所指向的页大小在页表树中存储为属性(一般在指向数据页的级中)。

[0005] 为了检索物理地址,以从根(IA32中的CR3)开始的递归方式读取页表项,并检索和合并页性质。IA32程序员参考手册(例如,卷3A)提供了一个检索物理地址的常规方法的例子。该进程需要若干存储器访问操作,并由页-表处理硬件或uCode(u码)序列来实现。

[0006] 有时,需要软件来检索物理存储器地址。在这种情形中,使用仿真全表遍历(emulated full table walk)或者基于操作系统施加的表设置限度的经缩短启发法。

[0007] 转换后备缓冲器(TLB)是保持先前转换结果的高速缓存,使得对一个地址(或地址范围)的连续访问可避免遍历数据结构,并可使用先前转换的结果。在很多情形中,地址转换还检查操作,以满足为存储器位置所设置的条件。常规的地址转换指令一般被提供作为操作数以返回针对线性地址的物理存储器地址,而无需提供任何附加信息。

[0008] 附图简述

[0009] 本发明的实施例在附图的各图中作为例子而不是限制示出,附图中相同的附图标记指示相似的元件。

[0010] 图1是示出根据本发明的一个实施例实现的计算系统的所选方面的高级框图。

[0011] 图2是根据本发明的一个实施例示出地址转换硬件的所选方面的框图。

[0012] 图3是根据本发明的一个实施例示出具有扩展页表的页树的所选方面的框图。

[0013] 图4是根据本发明的一个实施例示出页表项的所选方面的框图。

[0014] 图5是根据本发明的一个实施例示出支持页属性的线性至物理地址转换的方法的所选方面的流程图。

[0015] 详细描述

[0016] 本发明的实施例一般涉及支持页属性的虚拟(将被称为线性)至物理地址转换的系统、方法和装置。在某些实施例中,系统接收一指令,以将线性存储器指针转换成存储器位置的物理存储器地址。系统可返回物理存储器地址和一个或多个页属性。

[0017] 图1是示出根据本发明的一个实施例实现的计算系统的所选方面的高级框图。系统100包括应用程序102、内核104和硬件106。内核104管理硬件106并在应用程序102和硬件106之间提供通信。为了提供故障容错和安全性,内核104可以在高于应用程序102的特权级(例如,环0)操作。在某些实施例中,内核104包括LOAD PHYSICAL ADDRESS(加载物理地址)指令(LAP)108。LPA 108的输入参数几乎可以包括指向存储器的任何指针。LPA 108的输出参数包括与上述指针和一个或多个页属性对应的物理存储器地址(PA)。如以下参考图2-5进一步描述的,LPA 108可支持touch(创建)操作和/或虚拟化技术。

[0018] 表1示出根据本发明的某些实施例的LPA_{xx}指令的所选方面。在某些实施例中,LPA_{xx}指令取出线性地址(例如,m8)作为输入,并输出该存储器位置的物理地址以及一个或多个页属性。在32位模式中且在物理地址可能大于寄存器的情况下,指令可具有2个寄存器用于地址输出(以支持大于32位的物理地址空间)。在64位模式中,单个输出寄存器就足够了。在某些实施例中,指令的基本结构支持读语义。可支持写语义以直接检查页是否是写使能的(例如,参见以下的设计方案(trap)讨论)。

[0019] 表1

[0020]

指令	64-位模式	兼容性/支路 (Compatibility/Leg) 模式	操作
LPA	有效	有效	转换地址并返回属性
LPARD 32-位规则	有效	有效	转换地址, 返回属性并创建为读取(touch as read)
LPARD 64-位规则	有效	N.E.	

[0021]

LPAWR 32-位规则	有效	有效	转换地址, 返回属性并创建为写入
LPAWR 64-位规则	有效	N.E.	

[0022] N.E.-在该模式中指令未被编码

[0023] 在某些实施例中,当在旧版本模式中操作时,LPAxx指令将DS:(E)SI中的字节地址的地址移至EDX:EAX,且EBX保持访问许可和页大小。在64位模式中,LPAxx可将(R|E)SI中的字节地址的地址移至EDX:EAX(或r64寄存器),且EBX保持访问许可和页大小。在可选实施例中,可使用不同的寄存器,并可支持不同的模式。

[0024] 在某些实施例中,LPAxx指令是在例如内核级中运行的受保护的指令。在应用程序级运行该指令可导致违反用户应用程序虚拟化效应(例如,用户应用程序不知道其运行的实际地址)。在某些实施例中,LPAxx指令在同步操作系统调用中操作,其中内核代码与用户上下文(在IA32模式中这指示特定的CR3值)一起操作。在可选实施例中(如以下进一步描述的),指令在异步操作系统(OS)调用中操作。在某些实施例中,LPAxx的使用类似于操作系统中存在的等价SW操作,以确保所得到的物理地址与页表同步。

[0025] 输入参数可以是指向存储器的几乎任何有效方式。例如,m8值或r/m8值可以被选择性地用于在特定线性地址中给予指令更大的灵活性。地址转换硬件(例如,系统200,图2所示)可进行地址转换操作,好像该指令所要求的读(写)一样,并返回不同的结果。转换后备缓冲器(例如,图2所示的TLB 204)中的地址信息可被直接使用,而无需通过页遍历阶段。

[0026] 在某些实施例中,输出参数是物理地址(PA)(例如,长度达64位)和一个或多个页属性(例如,表2中所示属性中的一个或多个)。可将PA存储在单个64位寄存器中(例如,在IA32e模式中操作时)或隐含定义的寄存器对中(例如,在IA32模式或可选的IA32e兼容模式中操作时)。可将一个或多个页属性存储在隐含定义的寄存器中。在替换实施例中,使用了使得数据可用于稍后使用的软件的其它手段,诸如将信息存储在专用寄存器或存储器位置中。

[0027] 如表1所示,在某些实施例中,可存在两种(或更多种)类型的LPAxx指令。例如,某些实施例可包括LPA指令和带有touch指令的LPA。如上所述,LPA指令接收LA作为输入,并返回PA和一个或多个页属性。带有touch指令的LPA也返回PA和一个或多个页属性。此外,带有touch指令的LPA确定页是否存在于存储器中(根据页表末级中的状态)。如果最后页不存在,则带有touch指令的LPA产生错误,该错误与由RD或WR分别针对LPARD和LPAWR产生的错误相同。这允许标准错误处理程序维护错误情况(例如,在页错误的情况下把该页带入存储器,或针对写touch将只读页变换成写使能)。带有touch指令的LPA确定页状态而不访问存储器中的数据,这节省了对存储器分层结构的访问和在很多情况下的高速缓存未命中。相反,状态的确定则是基于页表项(PTE)的状态。

[0028] 图2是根据本发明的一个实施例示出地址转换硬件的所选方面的框图。地址转换硬件200包括存储器请求单元(MRU)202、转换备用缓冲器(TLB)204、页未命中处理程序

(PMH)206、存储器子系统(MSU)208、物理地址返回寄存器(PARR)210以及多路复用器(MUX)212。在替换实施例中,地址转换硬件200包括更多的元件、更少的元件和/或不同的元件。

[0029] MRU 202发出包括线性地址(例如,线性地址214)的存储器请求(例如,负载或存储)。TLB 204是保持物理地址和已被转换的线性地址的属性的高速缓存机构(其可包括多层TLB)。并非所有先前转换的地址都被保持在TLB中,且应用各种高速缓存启发法以确定哪些地址被保持。当页表被更新时,受影响的项由直接软件控制(软件管理的TLB相关性)或由硬件(硬件管理的TLB相关性)从TLB中清除。如果在TLB 204中发现LA 214的转换,则TLB直接输出物理存储器地址(PA)。PA(例如,PA 216)可用于直接从MSU 208获取数据。如果TLB 204不包括地址,则LA被发送(218)至PMH 206。PMH 206执行页遍历,且在页遍历成功完成时将PA(220)返回至TLB 204。当PMH遍历失败(例如,存储器中不存在针对页表任何级的数据)时,生成服从操作系统的一个异常。对于某些实施例中的LPA操作,操作用指示发生错误的状态终止而不生成异常,关于所返回信息的更多细节将在下文中提供。MSU 208表示存储器分层结构,它可由直接访问的存储器或用于减少存储器访问延迟的一级或多级高速缓存组成。MSU 208将存储在物理地址216中的数据经由返回总线返回至数据寄存器(222)。

[0030] 在某些实施例中,除(取决于具体情况从TLB 204或PMH 206)获取PA外,可响应于指令获取一个或多个页属性。PA和相关联的页属性作为LPA参数的一部分返回至处理器寄存器。在某些实施例中,PA和相关联的页属性被置于物理地址返回寄存器(例如,PARR 210)中,该寄存器是专用于保持PA和/或属性信息以及指示操作完成的状态的寄存器。可选地,处理器可使用微代码来从PARR读取数据并将其置于通用寄存器。在可选的实施例中,PARR可由其它处理器指令直接寻址。在另一个可选实施例中,PA和相关联的页属性可经由读取返回总线(例如,224)返回至处理器。在又一个实施例中,不同的技术可用于将PA和相关联的页属性返回至处理器。

[0031] 表2在本发明的某些实施例中提供可被返回的带PA的所选页属性的列表。在可选实施例中,可返回更多属性、更少属性和/或不同属性。

[0032] 表2

[0033]

页属性	简述
存在位(P-位)	指示页是否存在于存储器中。
读/写(R/W 位)	指示页是否是只读的或读/写访问是否可能。
用户/管理员位(U/S 位)	指示页的保护级。
访问位(A 位)	指示页是否被访问。
脏位(D 位)	指示页是否被修改。
全局标志(G 位)	指示页被全局使用(全局标志)。
执行禁用位(EXB 位)	指示页是数据页，且代码不能从其执行
页大小	根据页表结构指示页大小。例如，页的大小可以是 4KB、2MB、4MB 等。
转换表错误	指示在访问页表之一的同时地址转换的过程中发生页错误。该值指示从根开始的页表级(例如，正如由 CR3 寄存器或任何其它适当的寄存器所指向的)。页自身的存在由 P 位所指示。当页转换被

[0034]

	存储在 TLB 中时，页表树存在检验可有波动。
--	-------------------------

[0035] 对于大多数情况，异常以与对于正常的存储器访问(读或写)基本相同的方式来处理。然而，在某些实施例中，应用以下的异常。如果 LPA_{xx} 未能访问数据页，则指令返回等于 0 的地址，且 p 位被清零。如果地址转换未能访问任何页表，则“转换页表错误”状态被置位，以指示错误和错误的级别。关于带有 touch 指令的 LPA_{xx}，在对页访问失败时页错误被取得(例如，生成页错误异常)。在某些实施例中，在对利用“转换页表错误”状态标记的任何页表访问失败时异常未被取得。

[0036] 在某些实施例中，LPA_{xx} 指令支持带有增强页表(EPT)的虚拟化技术(VT)。当在 VT 环境中操作时，地址转换可部分地取决于请求者是客机 OS 还是主机 OS(或虚拟机管理器—VMM)，如处理器中的状态位所指示的。在某些实施例中，如果请求者是主机 OS，则 LPA_{xx} 指令返回主机 PA(HPA)。然而，如果请求者是客机 OS，则 LPA_{xx} 指令可返回客机 PA(GPA)。图 3 是根据本发明的一个实施例示出 VT 环境中地址转换的所选方面的框图。

[0037] 页表树 300 包括扩展页表(EPT) 302-306。EPT 302-306 类似于常规页表(例如，308 和 310)，不同之处在于 EPT 302-306 属于主机 OS。EPT 使用由 EPT 基址寄存器指向的表树来将 GPA 转换至 HPA。该操作针对需要包括对作为客机页表遍历一部分的客机页表的访问的转换的任何 GPA 而执行。客机 LA(GLA) 312 的一部分和 CR3(314) 引用 EPT 302 中的位置。EPT 302 又

指向页目录308中的位置。页目录308和GLA 312的另一部分可用于引用EPT 304中的位置。该过程可持续几乎任何级数直到到达GPA 316。EPT 306将GPA 316映射到HPA 318。在某些实施例中,当在VT环境中操作时,LPAxx指令返回存储在最后客机页表中的页自身的GPA(例如,GPA 316)以及一个或多个相关联的页属性。

[0038] 在某些实施例中,TLB仅保持HPA,这需要遍历页表用以返回GPA,就像在TLB未命中情况下所做的那样。在替换实施例中,TLB包括保持GPA和使能从TLB检索GPA的页属性的附加字段。

[0039] 图4是根据本发明的一个实施例示出页表项(PTE)的所选方面的框图。在某些实施例中,LPAxx指令的输出参数可从TLB(例如,TLB 204)中的PTE(例如,PTE 400)或TLB内的类似项中获取。例如,PA地址可从字段402中获取,且相关联的页属性可从字段404-420中获取。字段404-420中所示的页属性在以上的表2中进行了讨论。在可选实施例中,PTE 400可包括更多的字段、更少的字段和/或不同的字段。

[0040] 图5是根据本发明的一个实施例示出支持页属性的线性至物理地址转换的方法的所选方面的流程图。参考过程块502,地址转换逻辑(例如,图2所示的地址转换逻辑200)接收一指令(例如,LPAxx指令)以将存储器指针(例如,LA)转换至PA。地址转换逻辑将存储器指针转换至PA而不进行存储器访问。相反,或者从TLB项或者从PTE获取输出参数。参考过程块506,PA和一个或多个相关联的页属性被返回至处理器(例如,通过将数据置于PARR或经由读取返回总线返回数据)。短语“不进行存储器访问”指的是依赖页表信息来进行转换。页表信息可由页表项(例如,TLB中)或从存储器提供。

[0041] 在某些实施例中,LPAxx指令除同步调用外还支持异步调用。异步调用指的是从不与正用于转换的当前CR3对应的应用程序至LPAxx指令的调用。在某些实施例中,LPAxx指令包括告知页树页表遍历从哪开始的附加参数(例如,与进行异步调用的应用程序对应的CR3值)。在很多情况下,TLB中的数据在异步调用期间不能使用,因为该数据对应于不同的程序。于是在很多情况下,执行页表遍历以获取PA和一个或多个相关联的页属性。在其它情况下,提供标志以标识TLB项与哪个地址空间(CR3)相关联。这可实现于异步调用中使用TLB项。

[0042] 还可将本发明的实施例的各元素提供为用于存储机器可执行指令的机器可读介质。机器可读介质可包括但不限于闪存、光盘、紧致盘只读存储器(CD-ROM)、数字通用/视频盘(DVD)ROM、随机存取存储器(RAM)、可擦除可编程只读存储器(EPROM)、磁或光卡、传播介质或适用于存储电子指令的其它类型的机器可读介质。例如,本发明的实施例可作为从远程计算机(例如,服务器)经由通信链路(例如,调制解调器或网络连接)作为包括在载波或其它传播介质中的数据信号传输至请求计算机(例如,客户机)的计算机程序下载。

[0043] 应意识到,在本说明书中对“一个实施例”和“实施例”的应用表示结合包括在本发明的至少一个实施例中的实施例所描述的特定特征、结构或特性。因此,强调并应意识到,在本说明书的各部分中对“实施例”或“一个实施例”或“可选实施例”的两个或多个引用未必全指同一实施例。此外,特定特征、结构或特性可适当地结合在本发明的一个或多个实施例中。

[0044] 类似地,应意识到,在本发明的实施例的上述描述中,为了使公开简单化以帮助理解各发明方面中的一个或多个方面的目的,有时可将各特征组合在单个实施例、附图或其

描述中。然而,这种公开方法不应被解释为反映发明所要求保护的主体需要比各权利要求中所明确陈诉的特征要多。相反,如以下权利要求所反映的,发明方面少于单个上述公开实施例的全部特征。因此,详细描述之后的权利要求由此明确地结合在该详细描述中。

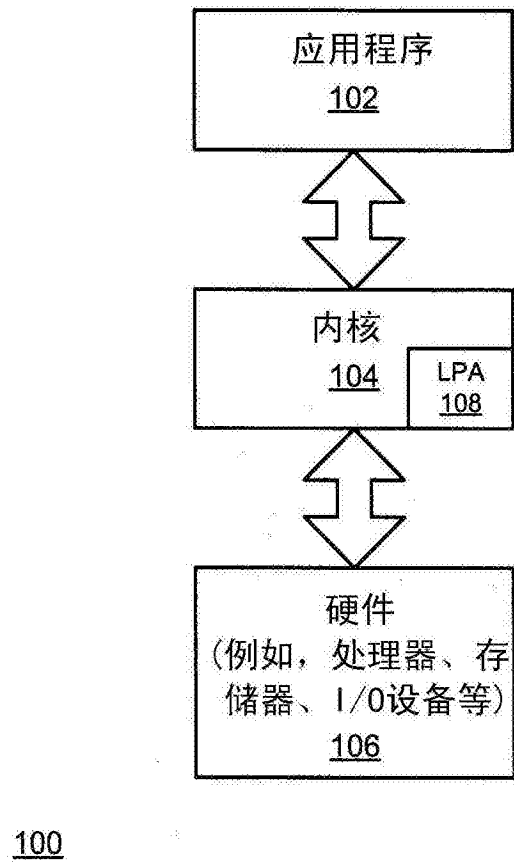


图1

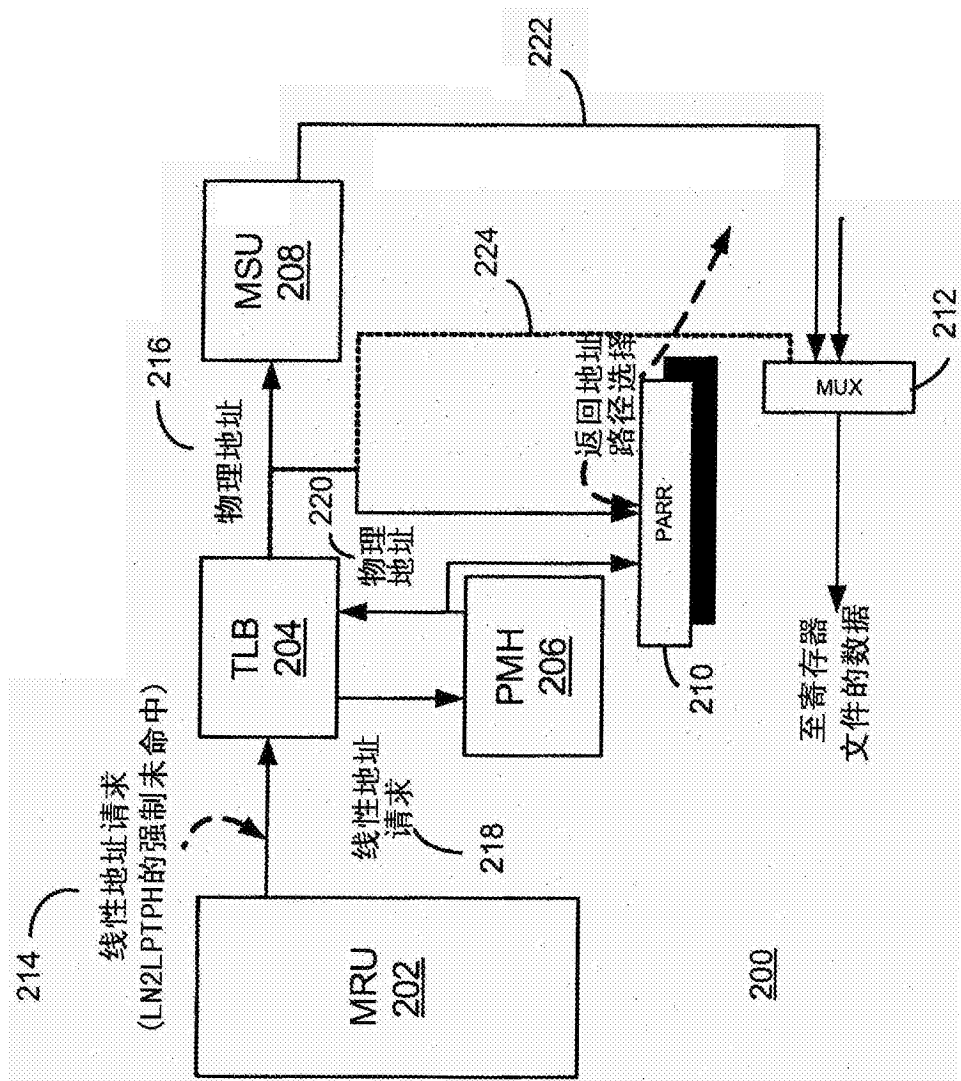


图2

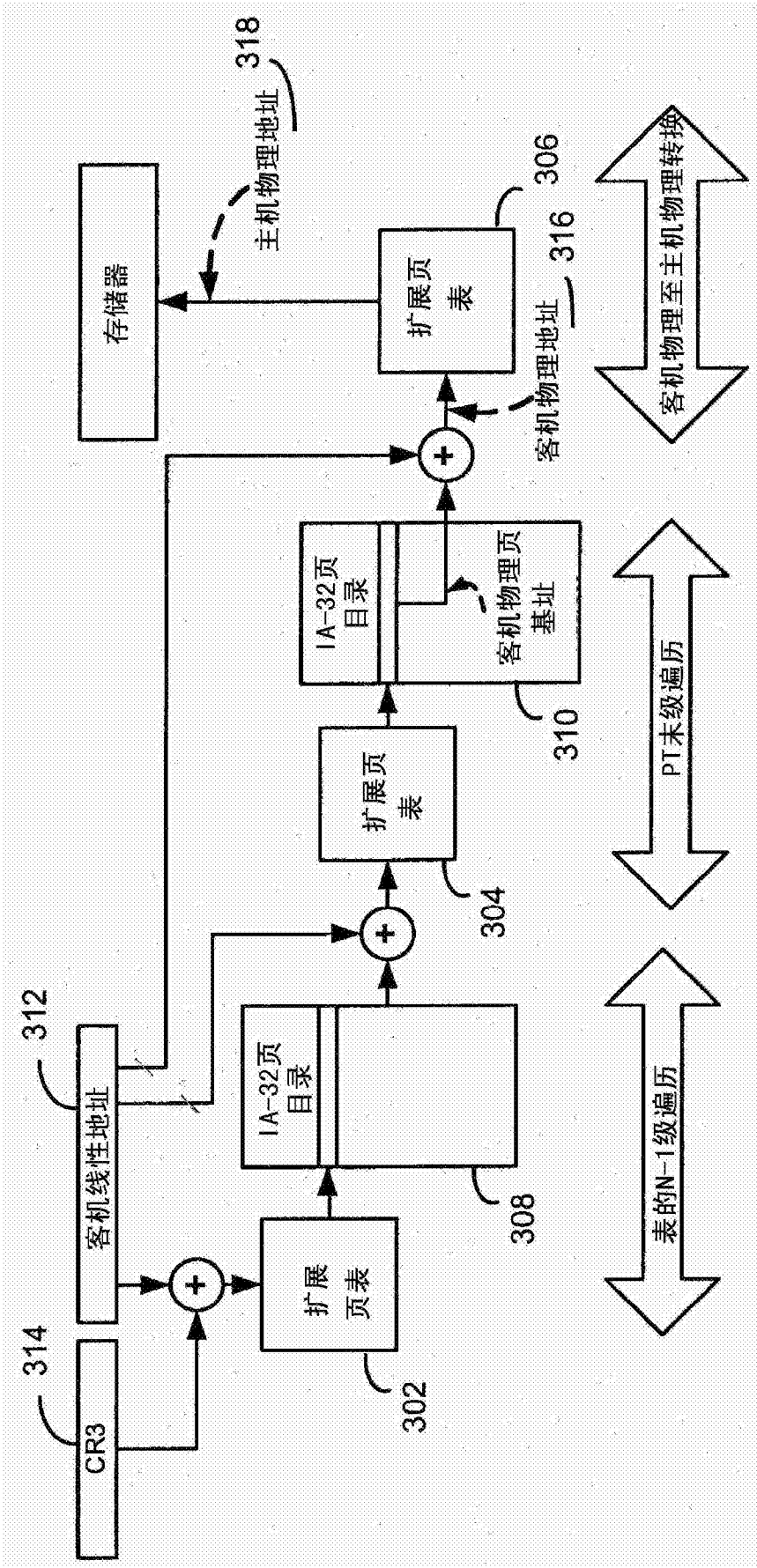


图3

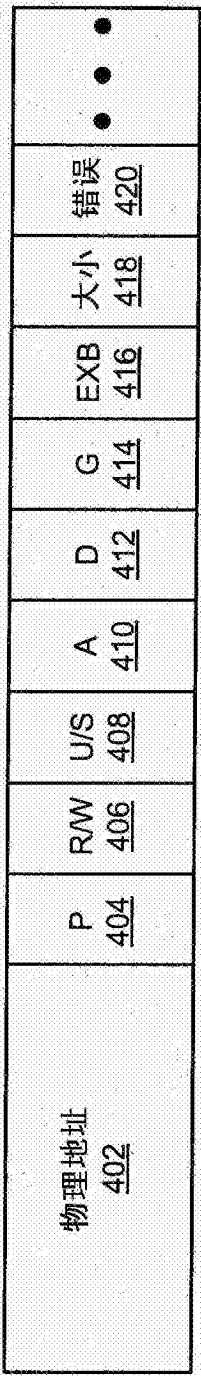


图4

400

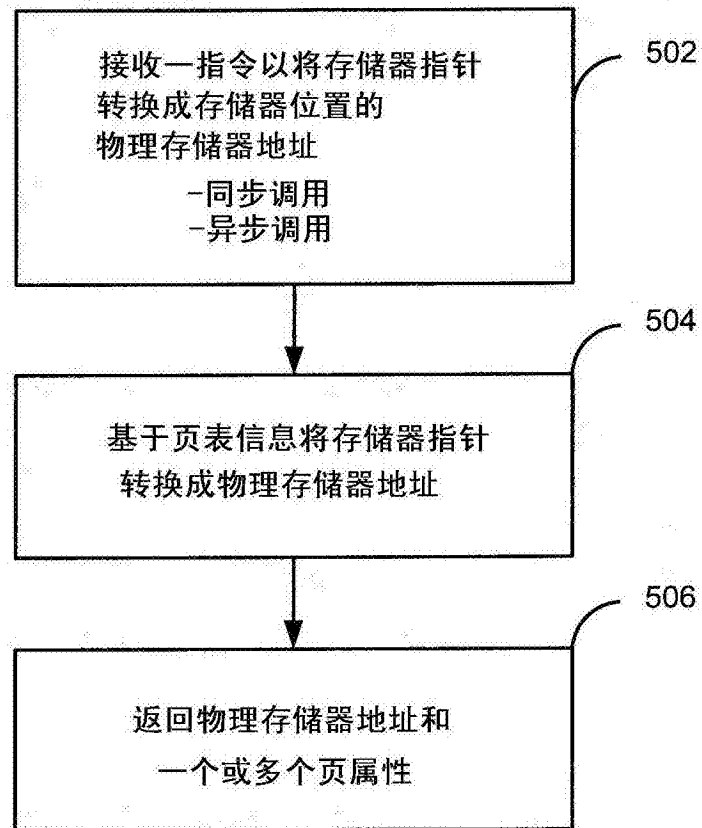


图5