(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105027095 B
(45)授权公告日 2019.01.22

(21)申请号 201480011808.8
(22)申请日 2014.03.04
(65)同一申请的已公布的文献号
申请公布号 CN 105027095 A
(43)申请公布日 2015.11.04
(30)优先权数据
13/785,877 2013.03.05 US
(85)PCT国际申请进入国家阶段日
2015.09.01
(86)PCT国际申请的申请数据
PCT/US2014/020101 2014.03.04
(87)PCT国际申请的公布数据
WO2014/137970 EN 2014.09.12
(73)专利权人 高通股份有限公司
地址 美国加利福尼亚
(72)发明人 T·曾 A·托兹尼 T·R·曾 P·J·博斯特利
(74)专利代理机构 永新专利商标代理有限公司 72002
代理人 张立达 王英
(51)Int.Cl.
G06F 12/10(2016.01)
(56)对比文件
CN 101558388 A,2009.10.14,
US 2006075285 A1,2006.04.06,
CN 102722451 A,2012.10.10,
TW Barr, AL Cox, S Rixner.SpecTLB: a mechanism for speculative address translation.《International Symposium on Computer Architecture》.2011,307-317.
审查员 刘瑞

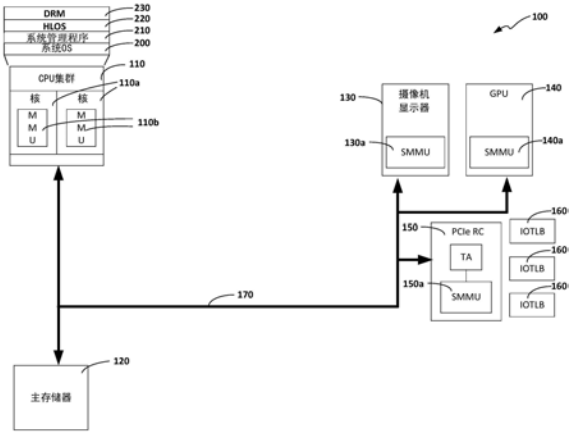
权利要求书3页 说明书7页 附图6页

(54)发明名称

降低执行硬件表移动所需资源量的方法、系统和可读介质

(57)摘要

提供了一种计算机系统和一种方法，其在发生转换旁路缓冲器(TLB)错过的情况下，降低执行硬件表移动(HWTW)需要的时间和计算资源的量。如果在执行阶段2(S2)HWTW以在存储阶段1(S1)的页表的位置处查找物理地址(PA)时发生TLB未中，则MMU使用中间物理地址(IPA)来预测相应的PA，由此避免执行S2表查找中的任一个的需求。这极大地降低了当执行这些类型的HWTW读事物时需要被执行的查找的数量，其极大地降低了与执行这些类型的事物相关联的处理开销和性能代偿。



CN 105027095 B

1.一种降低与执行硬件表移动HWTW相关联的处理开销的计算机系统,所述系统包括:

至少一个中央处理单元CPU,所述CPU运行主机操作系统OS和系统管理程序,所述系统管理程序控制所述CPU上的至少第一客户OS的执行,所述系统管理程序运行与所述第一客户OS相关联的至少第一虚拟机VM;

与所述CPU通信的物理内存,所述物理内存具有可由物理地址PA寻址的物理内存位置,其中,将至少一个页表存储在所述物理内存的物理内存位置处,所述页表包括与用于将中间物理地址IPA映射成所述物理内存的真实PA的映射相对应的页表条目;

至少一个转换旁路缓冲器TLB,其存储所述页表条目的子集;以及

至少一个内存管理单元MMU,所述MMU与所述CPU、与所述物理内存以及与所述TLB相通信,其中,所述MMU确定与IPA相关联的页表条目是否被存储在所述TLB中,其中,如果与所述IPA相关联的页表条目未被存储在所述TLB中,则已经发生TLB未中,并且其中,如果发生TLB未中,则所述MMU预测存储与所述IPA相关联的数据处的所述物理内存的PA;

其中,所述MMU将所述PA预测为所述IPA的函数f:PA=f(IPA);

其中,所述函数f选自多个函数,并且其中,所述多个函数中的每一个函数提供所述IPA与所预测的PA之间的一一对应的映射;以及

其中,所述系统管理程序运行与数字版权管理器DRM计算机程序相关联的至少第二VM,并且其中,所述函数f是IPA+Offset_function(VMID),其中VMID是跨越第一VM和所述第二VM、标识与所述TLB未中相关联的VM的唯一标识符,并且其中,Offset_function是具有基于与所述第一VM或所述第二VM的VMID相关联的特定偏移值来选择的输出的函数,其中当发生所述TLB未中时,所述第一VM或所述第二VM使用所述IPA来存取内存,并且其中,所预测的PA被预测为:PA=IPA+Offset_function(VMID)。

2.根据权利要求1所述的计算机系统,其中,所述函数f是多项式。

3.根据权利要求1所述的计算机系统,其中,所述函数f是一致函数,使得PA=IPA。

4.根据权利要求1所述的计算机系统,其中,所述跨越所述第一VM和所述第二VM的唯一标识符标识在发生所述TLB未中时使用所述IPA来存取内存的VM。

5.根据权利要求1所述的计算机系统,其中,所述函数f是IPA XOR Extended_VMID,其中XOR表示异或操作,并且Extended_VMID是扩展的VMID,并且其中,所预测的PA被预测为:

PA=IPA XOR Extended_VMID。

6.根据权利要求1所述的计算机系统,其中,所述计算机系统是移动设备的一部分。

7.根据权利要求6所述的计算机系统,其中,所述移动设备是移动电话。

8.根据权利要求7所述的计算机系统,其中,所述移动电话是智能电话。

9.一种降低与执行硬件表移动HWTW相关联的处理开销的方法,所述方法包括:

提供至少一个中央处理单元CPU、至少一个物理内存、至少一个转换旁路缓冲器TLB以及至少一个内存管理单元MMU,所述CPU、所述物理内存、所述TLB和所述MMU相互通信,所述CPU运行主机操作系统OS和系统管理程序,所述系统管理程序控制所述CPU上的至少第一客户OS的执行,所述系统管理程序运行与所述第一客户OS相关联的至少第一虚拟机VM,所述物理内存具有可由物理地址PA寻址的物理内存位置,其中,将至少一个页表存储在所述物理内存的物理内存位置处,所述页表包括与用于将中间物理地址IPA映射成所述物理内存的真实PA的映射相对应的页表条目,所述TLB存储所述页表条目的子集;以及

在所述MMU中：

确定与IPA相关联的页表条目是否被存储在所述TLB中，

如果作出与所述IPA相关联的页表条目未被存储在所述TLB中的确定，则判定已经发生TLB未中，并且

如果作出已经发生TLB未中的判定，则预测存储与所述IPA相关联的数据处的所述物理内存的PA；

其中，所述MMU将所述PA预测为所述IPA的函数f：PA＝f(IPA)；

其中，所述函数f选自多个函数，并且其中，所述多个函数中的每一个函数提供所述IPA与所预测的PA之间的一一对应的映射；以及

其中，所述系统管理程序运行与数字版权管理器DRM计算机程序相关联的至少第二VM，并且其中，所述函数f是IPA+Offset_function(VMID)，其中VMID是跨越第一VM和所述第二VM、标识与所述TLB未中相关联的VM的唯一标识符，并且其中，Offset_function是具有基于与所述第一VM或所述第二VM的VMID相关联的特定偏移值来选择的输出的函数，其中当发生所述TLB未中时，所述第一VM或所述第二VM使用所述IPA来存取内存，并且其中，所预测的PA被预测为：PA＝IPA+Offset_function(VMID)。

10.根据权利要求9所述的方法，其中，所述函数f是多项式。

11.根据权利要求9所述的方法，其中所述函数f是一致函数，使得PA＝IPA。

12.根据权利要求9所述的方法，其中，所述跨越所述第一VM和所述第二VM的唯一标识符标识在发生所述TLB未中时使用所述IPA来存取内存的VM。

13.根据权利要求9所述的方法，其中，所述函数f是IPA XOR Extended_VMID，其中XOR表示异或操作，并且Extended_VMID是扩展的VMID，并且其中，所预测的PA被预测为：

PA＝IPA XOR Extended_VMID。

14.根据权利要求9所述的方法，其中，所述系统管理程序控制所述CPU上的至少第一客户OS和第二客户OS的执行，并且其中，所述系统管理程序还运行与所述第二客户OS相关联的至少第二VM，并且其中，被所述MMU用来预测PA的所述函数f针对与所述第一VM相关联的未中，预测在第一PA范围中的PA，并且针对与所述第二VM相关联的未中，预测在第二PA范围中的PA，并且其中，所述第一PA范围和所述第二PA范围不相同。

15.根据权利要求9所述的方法，其中，所述方法是由移动设备的所述计算机系统执行的。

16.根据权利要求15所述的方法，其中，所述移动设备是移动电话。

17.根据权利要求16所述的方法，其中，所述移动电话是智能电话。

18.一种非暂时性计算机可读介质CRM，其具有存储于其上由一个或多个处理器执行以降低与执行硬件表移动HWTW相关联的处理开销的计算机代码，当所述计算机代码被运行主机操作系统OS和管理程序的至少一个中央处理单元CPU执行时，使用至少一个物理内存、至少一个转换旁路缓冲器TLB、以及至少一个内存管理单元MMU，以使得系统管理程序能够控制所述CPU上的至少第一客户OS的执行，所述系统管理程序运行与所述第一客户OS相关联的至少第一虚拟机VM，所述物理内存具有可由物理地址PA寻址的物理内存位置，其中，将至少一个页表存储在所述至少一个物理内存的物理内存位置处，所述至少一个页表包括与用于将中间物理地址IPA映射成所述至少一个物理内存的真实PA的映射相对应的页表条目，

所述至少一个TLB存储所述页表条目的子集,所述计算机代码包括:

第一代码部分,当被所述至少一个CPU执行时,确定与中间物理地址IPA相关联的页表条目是否被存储在所述至少一个TLB中,其中,如果作出与所述IPA相关联的页表条目未被存储在所述至少一个TLB中的确定,则当所述第一代码部分被所述至少一个CPU执行时,确定已经发生TLB未中;以及

第二代码部分,当被所述CPU执行时,在当已经发生TLB未中时预测存储要被读取的页表数据处的物理内存的物理地址PA,当所述第二代码部分被所述CPU执行时,应用在IPA的范围上确保IPA与PA之间的一一对应的映射的函数;

其中,所述第二代码部分将所述PA预测为所述IPA的函数f:PA=f(IPA);

其中,所述第二代码部分从多个函数选择所述函数f,并且其中,所述多个函数中的每一个函数提供所述IPA与所预测的PA之间的一一对应的映射;以及

其中,所述函数f是IPA+Offset_function(VMID),其中VMID是跨越第一VM和第二VM、将所述第一VM和所述第二VM之一标识为与所述TLB未中相关联的VM的唯一标识符,并且其中,Offset_function是具有基于与所述第一VM或所述第二VM的VMID相关联的特定偏移值来选择的输出的函数,其中当发生所述TLB未中时,所述第一VM或所述第二VM使用所述IPA来存取内存,并且其中,所预测的PA被预测为:PA=IPA+Offset_function(VMID)。

19.根据权利要求18所述的非暂时性计算机可读介质CRM,其中,所述函数f是多项式。

20.根据权利要求18所述的非暂时性计算机可读介质CRM,其中,所述函数f是一致函数,使得PA=IPA。

21.根据权利要求18所述的非暂时性计算机可读介质CRM,其中,所述唯一标识符标识在发生所述TLB未中时使用所述IPA来存取内存的VM。

22.根据权利要求18所述的非暂时性计算机可读介质CRM,其中,所述函数f是IPA XOR Extended_VMID,其中,XOR表示异或操作,并且Extended_VMID是扩展的VMID,并且其中,所预测的PA被预测为:

PA=IPA XOR Extended_VMID。

降低执行硬件表移动所需资源量的方法、系统和可读介质

技术领域

[0001] 本发明涉及计算机系统，而更具体地说，涉及计算机系统和供计算机系统中使用以降低执行硬件表移动(HWTW)需要的时间和计算资源的量的方法。

背景技术

[0002] 现代计算机系统使用内存管理单元(MMU)来管理向一个或多个物理存储设备(诸如例如固态存储设备)写数据以及从一个或多个物理存储设备读取数据。计算机系统的MMU向计算机系统的中央处理单元(CPU)提供虚拟内存，其允许CPU在其自己专用的、连续的虚拟内存地址空间上运行每一个应用程序，而不是使所有的应用程序共享物理内存地址空间，所述物理内存地址空间经常是成碎片的或非连续的。MMU的目的是针对CPU将虚拟内存地址(VA)转化成物理内存地址(PA)。CPU通过直接地对MMU读和写VA来间接地读和写PA，所述MMU将VA转化成PA，并且然后写或读PA。

[0003] 为了执行该转化，MMU存取被存储在系统主存储器中的页表。该页表由页表的条目组成。该页表的条目是由MMU使用来将VA映射成PA的信息。MMU通常包括转换旁路缓冲器(TLB)，其是被用于缓存最近使用的映射的高速缓冲存储器单元。当MMU需要将VA转化成PA时，MMU首先检查TLB以确定是否存在针对该VA的匹配。如果有，则MMU使用在TLB中找到的映射来计算PA，并且然后存取PA(即，读或写PA)。这被称为TLB“命中”。如果MMU没有在TLB中找到匹配，则这被称为TLB“未中”。

[0004] 在TLB未中事件中，MMU执行被称为硬件表移动(HWTW)的操作。HWTW是耗时的且计算上开销很高的过程，所述过程涉及执行“表移动”，以在MMU中查找相应的页表，并且在页表中读取多个位置，以查找相应的VA-至-PA的地址映射。然后，MMU使用该映射来计算相应的PA，并且将该映射写回到TLB。

[0005] 在实现操作系统(OS)虚拟化的计算机系统中，将虚拟内存监视器(VMM)(通常还被称为系统管理程序)插入到计算机系统的硬件与计算机系统的系统OS之间。系统管理程序在特权模式下执行，并且能够主管一个或多个客户高级OS。在这样的系统中，运行在OS上的应用程序使用虚拟内存的第一层的VA来对内存寻址，以及运行在系统管理程序上的OS使用虚拟内存的第二层的中间物理地址(IPA)来对内存寻址。在MMU中，执行阶段1(S1)转化，以将每一个VA转化成IPA，以及执行阶段2(S2)转化，以将每一个IPA转化成PA。

[0006] 如果在执行这样的转化时发生TLB未中，则执行多级的、二维(2-D)的HWTW，以获得计算相应的IPA和PA需要的表条目。执行这些多级的、2-D的HWTW，能够引起MMU的大量计算开销，其通常导致性能代偿。

[0007] 图1是当执行读事务时发生TLB未中时执行已知的、三级的、2-D的HWTW的插图。图1中示出的HWTW表示针对三级的、2-D的HWTW的最坏情况的场景，需要十五次表查找的执行来在数据被存储在物理内存中的地方获得PA。针对这个示例，计算机系统的MMU正在运行主管至少一个客户高级OS(HLOS)的系统管理程序，继而，其正运行至少一个应用程序。在这样的配置中，正被客户HLOS分配的内存不是系统的真实的物理内存，而是上述的中间物理内存。

系统管理程序分配真实的物理内存。因此，将每一个VA转化成IPA，然后，将IPA转化成读取的数据实际被存储的地方的真实物理内存的PA。

[0008] 该过程开始于MMU接收S1页全局目录(PGD)IPA 2。针对这个最坏情况场景的示例，将假设的是，在MMU针对匹配检查TLB时发生TLB未中。因为该未中，MMU必须执行HWTW。HWTW涉及执行三次S2表查找3、4和5，以获得将IPA 2转变成PA所需的映射，以及一次额外的查找6来读取PA。表查找3、4和5分别涉及读取S2 PGD、页中间目录(PMD)和页表条目(PTE)。在查找6处读取PA为MMU提供S1 PMD IPA7。针对这个最坏情况场景的示例，将假设的是，当MMU针对匹配利用S1 PMD IPA 7来检查TLB时发生TLB未中。因为该文中，所以MMU必须执行另一HWTW。该HWTW涉及执行三次S2表查找8、9和11，以获得将S1 PMD IPA 7转变成PA所需的映射，以及一次额外的查找12来读取PA。表查找8、9和11分别涉及读取S2 PGD、PMD和PTE。在查找12处读取PA为MMU提供S1 PET IPA13。

[0009] 针对这个最坏情况场景的示例，将假设的是，当MMU针对匹配利用S1 PTE IPA 13来检查TLB时发生TLB未中。因为该未中，MMU必须执行另一HWTW。该HWTW涉及执行三次S2表查找14、15和16，以获得将S1 PTE IPA 13转变成PA所需的映射，以及一次额外的查找17来读取PA。表查找14、15和16分别涉及读取S2 PGW、PMD和PTE。在查找17处读取PA为MMU提供真实的IPA 18。针对这个最坏情况场景的示例，将假设的是，当MMU针对匹配利用真实的IPA 18来检查TLB时发生TLB未中。因为该未中，MMU必须执行另一HWTW。该HWTW涉及执行三次S2表查找19、21和22，以获得将真实的IPA 18转变成PA所需的映射。表查找19、21和22分别涉及读取S2 PGD、PMD和PTE。然后，读取PA来获得相应的读数据。在查找18处读取PA为MMU提供S1 PTE IPA 13。

[0010] 因此，能够看出的是，在针对三级的、2-D的HWTW的最坏情况场景下，执行了12次S2表查找和三次S1表查找，其是消耗大量时间并导致性能代偿的大量计算开销。已经使用各种各样的技术和架构来降低在执行HWTW中涉及的时间和处理开销的量，包括，例如，增加TLB的大小，使用多个TLB，使用平面嵌套的页表，使用影子分页或推测的影子分页，以及使用页移动高速缓冲存储器。虽然所有的这些技术和架构能够降低与执行HWTW相关联的处理开销，但是，它们经常导致计算机系统中别的地方的处理开销的增加。

[0011] 因此，存在针对降低执行HWTW需要的时间和计算资源的量的计算机系统和方法的需求。

发明内容

[0012] 本发明针对一种计算机系统和一种供在计算机系统中使用的方法，用于降低执行HWTW需要的时间和计算资源的量。该计算机系统包括至少一个中央处理单元(CPU)、至少一个物理内存、至少一个TLB和至少一个MMU。CPU运行主机OS和系统管理程序。系统管理程序控制CPU上的至少第一客户OS的执行。系统管理程序运行与第一客户OS相关联的至少第一VM。物理内存具有可由PA寻址的物理内存位置。将至少一个页表存储在物理内存的物理内存位置处。页表包括与用于将IPA映射成物理内存的真实PA的映射相对应的页表条目。TLB存储页表条目的子集。当执行内存访问时，MMU确定与IPA相关联的页表条目是否被存储在TLB中。如果与IPA相关联的页表条目未被存储在TLB中，则已经发生TLB未中。如果发生TLB未中，则预测存储与IPA相关联的数据处的物理内存的PA，由此避免通过执行HWTW来计算PA

的需求。

[0013] 该方法包括：

[0014] 在MMU中：

[0015] 确定与IPA相关联的页表条目是否被存储在TLB中；

[0016] 如果作出与IPA相关联的页表条目未被存储在TLB中的确定，则判定已经发生TLB未中；以及

[0017] 如果作出已经发生TLB未中的判定，则预测存储与所述IPA相关联的数据处的物理内存的PA。

[0018] 本发明还提供了一种计算机可读介质(CRM)，所述计算机可读介质存储用于由一个或多个处理器执行以降低与执行HWTW相关联的处理开销的计算机代码。该计算机代码包括第一代码部分和第二代码部分。第一代码部分确定与IPA相关联的页表条目是否被存储在TLB中。如果作出与IPA相关联的页表条目未被存储在TLB中的确定，则第一代码部分判定已经发生TLB未中。如果第一代码部分判定已经发生TLB未中，则第二代码部分预测存储与所述IPA相关联的数据处的物理内存的PA。

[0019] 根据下面的描述、附图和权利要求，这些和其它特征和优点将变得显而易见。

附图说明

[0020] 图1是根据本发明的说明性的实施例的计算机系统的框图。

[0021] 图2示出了根据被配置为执行用于降低执行HWTW需要的时间和计算资源的量的方法的说明性的、或示例性的实施例的计算机系统的框图。

[0022] 图3是根据说明性的实施例来表示被图2中示出的系统管理程序执行的、以降低执行HWTW读事务需要的时间和处理开销的量的方法的流程图。

[0023] 图4是根据说明性的实施例来展示在其中使用由图3中示出的流程图表示的方法来执行HWTW读事务的方式的示意图。

[0024] 图5是根据说明性的实施例执行由图3中示出的流程图表示的方法的硬件预测器的框图。

[0025] 图6示出在其中并入了图2中示出的计算机系统的移动智能手机的框图。

具体实施方式

[0026] 根据本文描述的说明性实施例，提供了一种计算机系统和一种供在计算机系统中使用的方法，用于降低执行HWTW需要的时间和计算资源的量。根据本文描述的实施例，当执行S2 HWTW以在存储S1页表的位置处查找PA时发生TLB未中时，MMU使用IPA来预测相应的PA，由此避免执行S2表查找中的任一个的需求。这极大地降低了当执行这些类型的HWTW读事务时需要被执行的查找的数量，其极大地降低了与执行这些类型的事务相关联的处理开销和性能代偿。

[0027] 图2根据被配置为执行用于降低执行S2 HWTW以在存储S1页表的位置处查找PA需要的时间和计算资源的量的方法的说明性的、或示例性的实施例示出了计算机系统100的框图。图2中示出的计算机系统100的示例包括CPU集群110、主存储器120、摄像机显示器130、图形处理单元(GPU)140、外围连接接口快速(PCIe)输入/输出(I/O)设备150、多个IO

TLB(IOTLB)160以及系统总线170。计算机集群110具有多个CPU核110a，其中的每一个CPU核110a具有MMU 110b。每一个CPU核110a可以是微处理器或任意其它合适的处理器。摄像机显示器130具有系统MMU(SMMU)130a。GPU 140具有其自己的SMMU 140a。同样地，PCIe IO设备150具有其自己的SMMU 150a。

[0028] 处理器核110a的MMU 110b被配置为执行将VA转化成IPA以及将IPA转化成PA的任务。将页表存储在主存储器120中。MMU 110b和SMMU 130a、140a和150a中的每一个具有其自己的TLB(出于清楚的目的，未示出)，所述TLB存储被存储在主存储器120中的页表的子集。根据这个说明性的实施例，在发生TLB未中之后，MMU 110b执行处理IPA以预测PA的预测算法。在数学上，可以将该预测算法表示为：

[0029] PA＝f(IPA)， (等式1)

[0030] 其中，f表示数学函数。下文关于图5详细地描述可以被用于这个目的的函数f。短语“以预测”，如该短语在本文被使用的，表示“以确定”的意思，并且尽管随机的或概率的确定不必然地被排除在本发明的范围之外，但是，该短语并不意味随机的或概率的确定。由预测算法进行的预测通常是确定性的，但是，不必然地是确定性的。

[0031] CPU集群110运行系统OS 200和虚拟机监视器(VMM)或系统管理程序210。系统管理程序210管理转化任务，所述转化任务除去执行转化之外包括更新被存储在MMU 110b和SMMU 130a、140a和150a中的页表。系统管理程序210还运行客户HLOS 220和/或客户数字版权管理器(DRM)230。可以将HLOS 220与摄像机显示器130相关联，以及可以将DRM 230与GPU 140相关联。系统管理程序210管理HLOS 220和DRM 230。

[0032] 在发生TLB未中之后，系统管理程序210配置MMU 110b和SMMU 130a、140a和150a，以执行预测算法来将IPA转变成PA。在这样的情况下，以在其中S1转化正常开始的典型方式，从CPU集群110的硬件基址寄存器(出于清楚的目的，未示出)获得针对与TLB未中相关联的VA的起始IPA。然后，如下文将更详细地描述的，预测算法根据等式1预测PA。为了管理和更新SMMU 130a、140a和150a，CPU MMU 110b在总线170上向SMMU 130a、140a和150a发送分布式虚拟内存(DVM)消息。MMU 110b和SMMU 130a、140a和150a访问主存储器120，以执行HWTW。

[0033] 根据说明性的实施例，CPU MMU 110b将MMU业务分成三个事务类别，即：(1)S2HWTW读事务，以查找存储S1页表的PA；(2)客户端事务；以及(3)地址故障(AF)/脏标志写事务。根据这个说明性的实施例，对于类别1事务(即，HWTW读事务)，预测算法针仅将IPA转变成PA。针对所有其它的事务类别，根据这个说明性的实施例，MMU 110b和SMMU 130a、140a和150a以典型的方式执行所有其它的转化(例如，S1和客户端事务S2转化)。

[0034] 图3是根据说明性的实施例来表示由CPU MMU 110b执行以降低执行HWTW读事务需要的时间和处理开销的量的方法的流程图。框301表示方法开始，通常在CPU集群110启动并开始运行系统OS 200和系统管理程序210时发生。如由框302指示的，MMU 110b将业务分成上述的事务类别(1)、(2)和(3)。分类过程可以将事务分成比这三个类别要多或要少的类别，但是，分类中的至少一个分类将是类别(1)事务，即，S2 HWTW读事务，以查找存储S1页表的PA。在由框303表示的步骤处，对在执行类别(1)的事务时是否已经发生TLB未中进行确定。如果未发生，则该方法进行到框306，在框306处，MMU 110b或SMMU 130a、140a或150a以正常的方式执行HWTW。

[0035] 如果在由框303表示的步骤处CPU MMU 110b确定当执行类别(1)事务时发生未中，则该方法进行到由框305表示的步骤。在由框305表示的步骤处，执行上述预测算法，以将IPA转变或转化成PA。

[0036] 图4是根据说明性的实施例来展示在其中执行HWTW读事务的方式的示意图。针对这个说明性的实施例，出于示例性的目的，假设页表是三级页表并且HWTW是2-D的HWTW。该示例还假设TLB未中的最坏情况场景。该过程开始于MMU接收VA，以及然后从控制寄存器(出于清楚的目的，未示出)检索S1 PGD IPA 401。然后，MMU针对匹配利用S1 PGD IPA 401来检查TLB。针对这个最坏情况场景的示例，将假设的是，当MMU针对匹配来检查TLB时发生TLB未中。因为这个未中，MMU执行预测算法，以将S1 PGD IPA 401转变成存储S1 PMD IPA 403的PA 402。因此，使用单个查找来将S1 PGD IPA 401转变成PA 402。

[0037] 针对该最坏情况场景的示例，将假设的是，当MMU针对匹配利用S1 PMD IPA 403来检查TLB时发生TLB未中。因为该未中，MMU执行预测算法，以将S1 PMD IPA 403转变成存储S1 PTE IPA 405的PA 404。因此，使用单个查找来将S1 PMD IPA 403转变成PA 404。针对这个最坏情况场景的示例，将假设的是，当MMU针对匹配利用S1 PTE IPA 405来检查TLB时发生TLB未中。因为该未中，MMU执行预测算法，以将S1 PTE IPA 405转变成存储IPA1 407的PA 406。一旦已经获得IPA1 407，就执行三次查找408、409和411，以获得存储要读取的数据的最终PA 412。

[0038] 因此，根据这个实施例，能够看出的是，已经将查找的总次数从十五(图1)降低到了六，其表示处理开销降低了60％。当然，本发明不限于具有特定级数或特定数量的HWTW维度的MMU配置。本领域技术人员将理解的是，应用本发明的概念和原理，而不管页表的配置。同样地，尽管本文关于IPA-至-PA转变描述了方法和系统，但是，它们同样可应用于不使用IPA的系统中的直接的VA-至-PA的转变。

[0039] 图5是执行预测算法的预测器500的说明性实施例的框图。通常在MMU 110b中和SMMU 130a、140a和150a中实现预测器500。如上文指示的，根据说明性的实施例，仅当执行类别1的读事务时执行预测算法。图5中示出的预测器500的配置是允许针对类别1的事务启用预测器500和针对所有其它的事物类别(包括类别2和3的事务)禁用预测器500的一个配置的示例。

[0040] 图5中示出的预测器500的配置还允许预测器500选择被用在上文等式1中的函数f，以基于IPA来计算PA。每一个虚拟机(VM)可以使用不同的函数f的集合，所以，重要的是，被使用的函数的集合确保在IPA的范围上的IPA与PA之间存在一一对应的映射。系统管理程序210可以管理多个HLOS或DRM，所述多个HLOS或DRM中的每一个将具有运行在系统管理程序210中的相应的VM。被使用的函数的集合确保预测的PA不与被分配给另一VM的预测的PA重叠。

[0041] 函数f的示例是：

[0042] PA＝IPA；

[0043] PA＝IPA+Offset_function(VMID)，其中，VMID是跨越所有VM标识与HWTW读事务相关联的VM的唯一标识符，Offset_function是具有基于与VMID相关联的特定的偏移值选择的输出的函数；以及

[0044] PA＝IPA XOR Extended_VMID，其中，XOR表示异或操作，以及Extended_VMID是扩

展的VMID。系统管理程序210选择函数f，从而避免VM之间的冲突。

[0045] 在图5中，假设函数f是多项式，并且系统管理程序210从多个多项式选择要被用作函数f的多项式。可以基于例如针对其执行HWTW读事物的VM的VMID来选择多项式。预测器500的配置寄存器510存储一个或多个预测使能比特510a和一个或多个多项式选择比特510b。预测器500的多项式计算硬件520包括基于从寄存器510接收的多项式选择比特510b的值来选择多项式函数的硬件。多项式计算硬件520还接收IPA-至-PA的转化请求，并根据选择的多项式函数来处理该请求，以生成预测的PA。

[0046] 在与门530的输入处接收预测使能比特510a和类别1使能比特。当在执行类别1的读事物时已经发生未中时，断言类别1的使能比特。预测器500的复用器(MUX)540在MUX 540的选择器端口处接收与门530的输出，并且接收以正常方式获得的预测的PA和IPA-至PA的转化结果。当预测使能比特510a和类别1的使能比特均被断言时，S2WALK控制逻辑与状态机550被禁用，并且MUX 540选择要从MUX 540输出的预测的PA。

[0047] 当预测使能比特510a和/或类别1的使能比特被解除认定时，S2移动控制逻辑与状态机550被启用。当S2移动控制逻辑与状态机550被启用时，可以由S2移动控制逻辑与状态机550在主存储器120中执行其它类型的S2移动(例如，类别2和类别3)。因此，当S2移动控制逻辑与状态机550被启用时，MUX 540输出IPA-至-PA的转化结果，所述IPA-至-PA的转化结果是根据S2移动控制逻辑与状态机550输出的。

[0048] 应当注意到的是，预测器500可以具有多个不同的配置。图5中示出的预测器500的配置仅是用于执行预测算法的多个合适的配置中的一个配置。本领域技术人员将理解的是，可以使用除了图5中示出的之外的多个配置来执行预测算法。

[0049] 可以在执行内存虚拟化的任意类型的系统(包括例如，台式计算机、服务器和移动智能手机)中实现图2中示出的计算机系统100。图6示出并入了计算机系统100的移动智能手机600的框图。除了智能手机600必须能够执行本文描述的方法之外，智能手机600不限于任意特定类型的智能手机或具有任意特定的配置。同样地，图6中示出的智能手机600旨在是具有情境感知和用于执行本文描述的方法的处理能力的蜂窝电话的简化示例。具有本领域普通技术的人员将理解，已经省略了智能手机的操作和结构，以及由此的实现方式的细节。

[0050] 根据这个说明性的实施例，智能手机600包括在系统总线612上被连接在一起的基带子系统610和射频(RF)子系统620。系统总线612通常包括将上述元件耦合在一起并启用它们的互操作性的物理的和逻辑的连接。RF子系统620可以是无线收发机。尽管为了清楚未示出细节，但是，RF子系统620通常包括具有用于为发送准备基带信息信号的调制、上变频和放大电路的发送(Tx)模块630，包括具有用于接收RF信号并将RF信号下变频到基带信息信号以恢复数据的放大、滤波和下变频电路的接收(Rx)模块640，以及包括前端模块(FEM)650，所述前端模块(FEM)650包括天线共用器电路、双工器电路或如对本领域技术人员来说是已知的能够将发送信号与接收信号分离的任意其它电路。将天线660连接至FEM 650。

[0051] 基带子系统610通常包括经经由系统总线612电耦合在一起的计算机系统100、模拟电路元件616和数字电路元件618。系统总线612通常包括物理连接和逻辑连接，以将上述元件耦合在一起并启用它们的互操作性。

[0052] 经由连接624将输入/输出(I/O)元件621连接至基带子系统610。I/O元件621通常

包括，例如，麦克风、小键盘、扬声器、定点设备、用户接口控制元件和允许用户提供输入命令并从智能手机600接收输出的任意其它设备或系统。经由连接629将存储器628连接至基带子系统610。存储器628可以是任意类型的易失性或非易失性存储器。可以将存储器628永久地安装在智能手机600中，或可以是诸如可移动存储卡的可移动存储元件。

[0053] 模拟电路616和数字电路618包括信号处理、信号转变和将由I/O元件621提供的输入信号转变成要被发送的信息信号的逻辑单元。类似地，模拟电路616和数字电路618包括被用于生成包含从接收的信号恢复的信息的信息信号的信号处理元件。数字电路618可以包括，例如，数字信号处理器(DSP)、现场可编程门阵列(FPGA)或任意其它处理设备。因为基带子系统610包括模拟元件和数字元件，所以其可以被称为混合信号设备(MSD)。

[0054] 智能手机600可以包括各种各样的传感器中的一个或多个传感器，例如照相机661、麦克风662、全球定位系统(GPS)传感器663、加速计665、陀螺仪667和数字罗盘668。这些传感器经由总线612与基带子系统610相通信。

[0055] 使计算机系统100嵌入到智能手机600中允许多个OS和多个各自的VM运行在智能手机600上。在这个环境下，计算机系统100的系统管理程序210(图2)在智能手机600的硬件与被VM执行的应用软件之间提供安全隔离。

[0056] 可以单独地在硬件中、或在硬件与软件的组合中、或在硬件与固件的组合中实现上文关于图3描述的方法。同样地，可以单独地在硬件中、或在硬件与软件或固件的组合中实现图2中示出的计算机系统100的部件中的多个。例如，可以单独地在硬件中、或在硬件与软件或固件的组合中实现系统管理程序210。在软件或固件中实现计算机系统100的方法或部件的情况下，将相应的代码存储在计算机可读介质的主存储器120中(图2)。主存储器120通常是固态计算机可读介质，例如，非易失性随机存取存储器(RAM)、只读存储器(ROM)设备、可编程ROM(PROM)、可擦除PROM(EPROM)等。然而，可以使用其它类型的计算机可读介质(诸如例如，磁性存储设备和光学存储设备)来存储代码。

[0057] 还应当注意到的是，在不偏离本发明的保护范围的情况下，可以对上文关于图2-6描述的方法进行多种变型。例如，如将被本领域技术人员理解的，可以以多个方式修改图2中示出的计算机系统100的配置。同样地，图6中示出的智能手机600仅是具有用于执行该方法的合适的配置和功能的移动设备的一个示例。鉴于本文提供的描述，本领域技术人员将理解的是，在不偏离本发明的保护范围的情况下，可以对图6中示出的智能手机600进行多种变型。这些变型和其它变型在本发明的保护范围内。如将被本领域技术人员理解的是，本文描述的说明性实施例旨在展示本发明的原理和概念，但是，本发明不限于这些实施例。

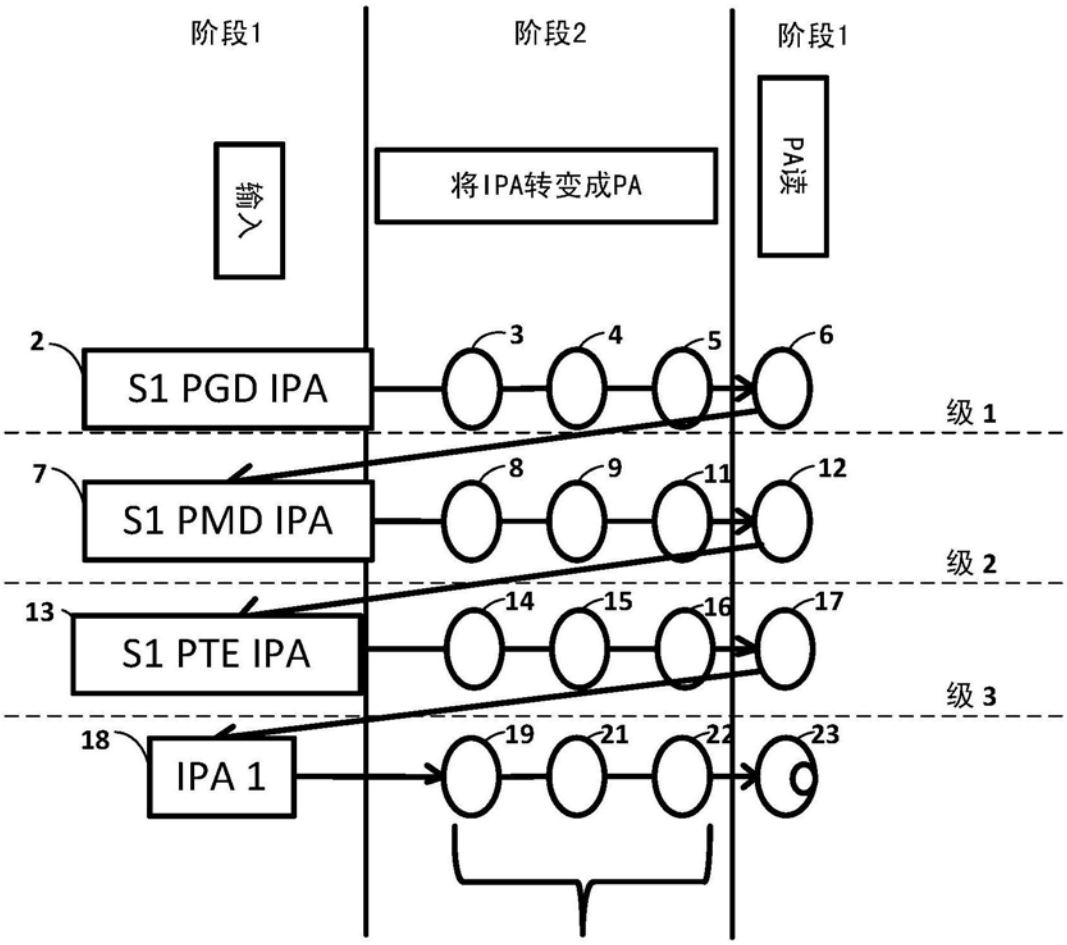

阶段1
阶段2
阶段1
输入
将IPA转变成PA
PA读
2
S1 PGD IPA
3
4
5
6
级 1
7
S1 PMD IPA
8
9
11
12
级 2
13
S1 PTE IPA
14
15
16
17
级 3
18
IPA 1
19
21
22
23

图1(现有技术)

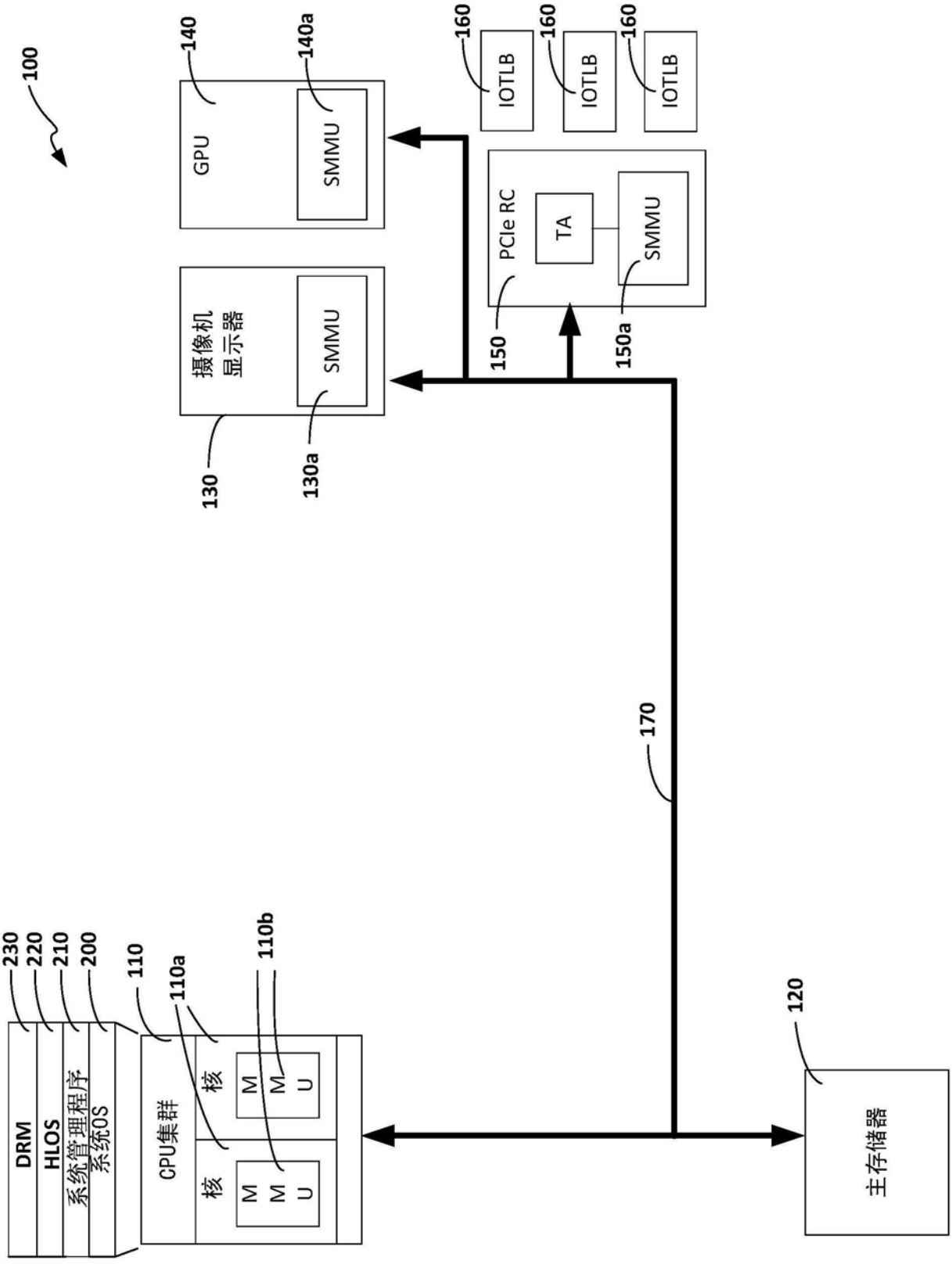

100
GPU
140
SMMU
140a
IOTLB
160
IOTLB
160
IOTLB
160
PCIe RC
TA
SMMU
150
150a
摄像机
显示器
SMMU
130
130a
170
DRM
HLOS
系统管理程序
系统OS
230
220
210
200
110
CPU集群
核
核
MMU
MMU
110a
110b
120
主存储器

图2

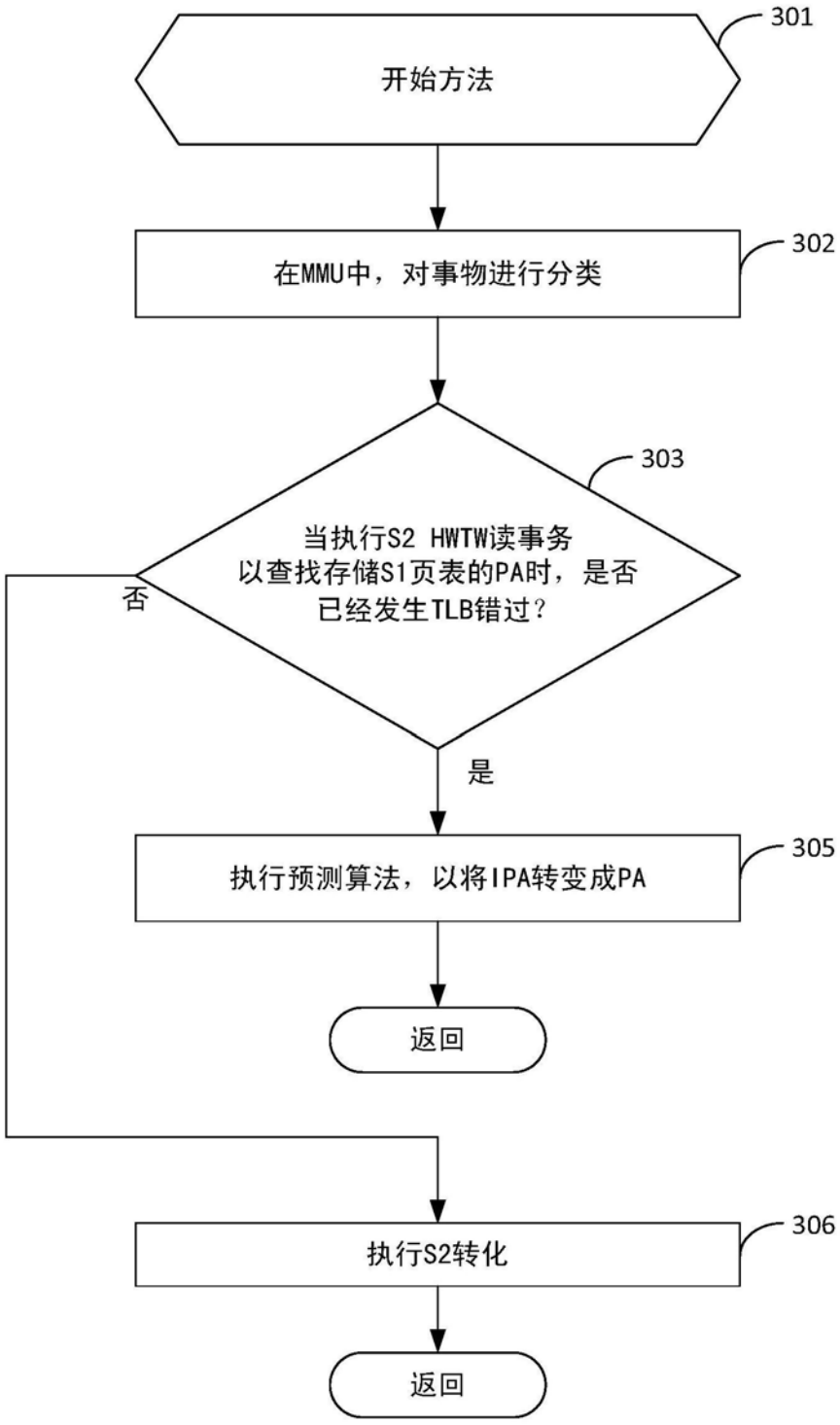

301
开始方法
302
在MMU中，对事物进行分类
303
当执行S2 HWTW读事务
以查找存储S1页表的PA时，是否
已经发生TLB错过？
否
是
305
执行预测算法，以将IPA转变成PA
返回
306
执行S2转化
返回

图3

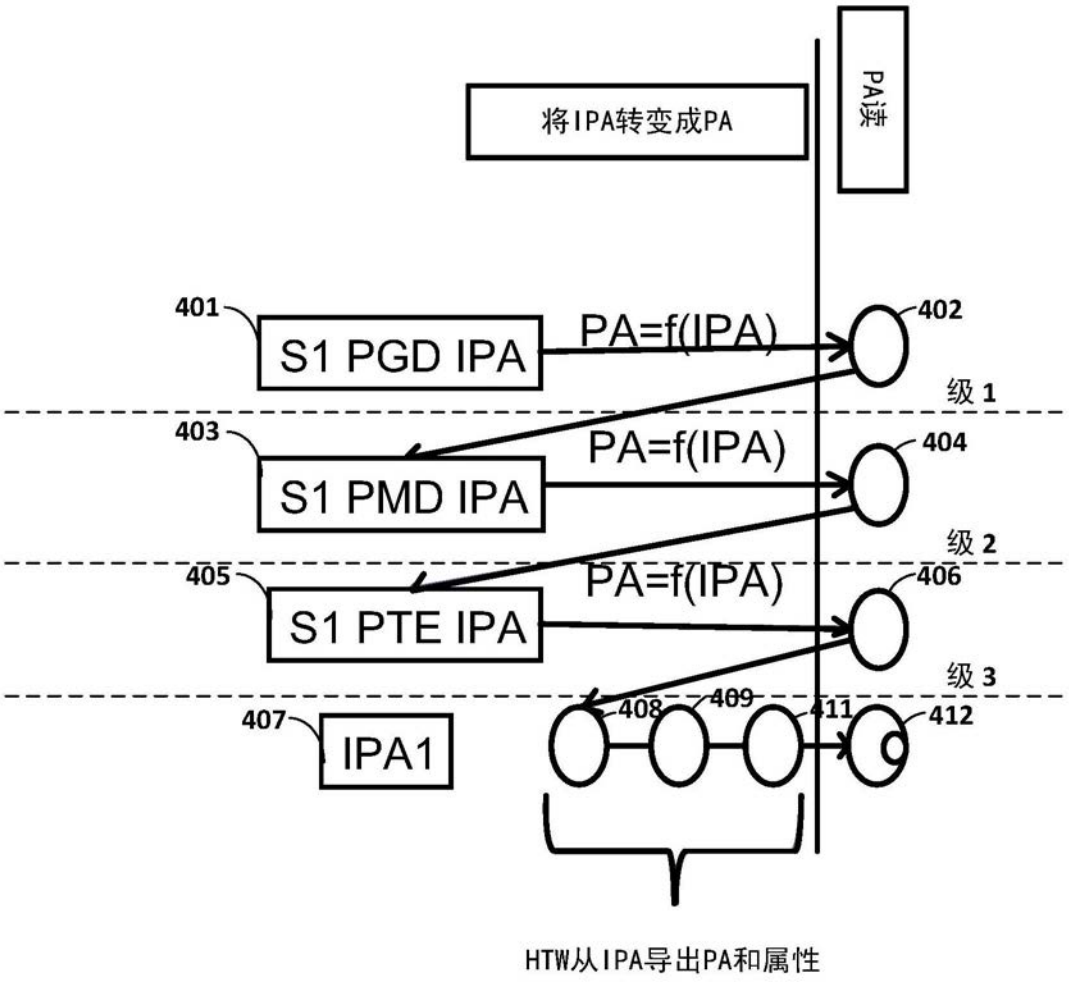

PA读
将IPA转变成PA
401
S1 PGD IPA
PA=f(IPA)
402
级 1
403
S1 PMD IPA
PA=f(IPA)
404
级 2
405
S1 PTE IPA
PA=f(IPA)
406
级 3
407
IPA1
408
409
411
412
HTW从IPA导出PA和属性

图4

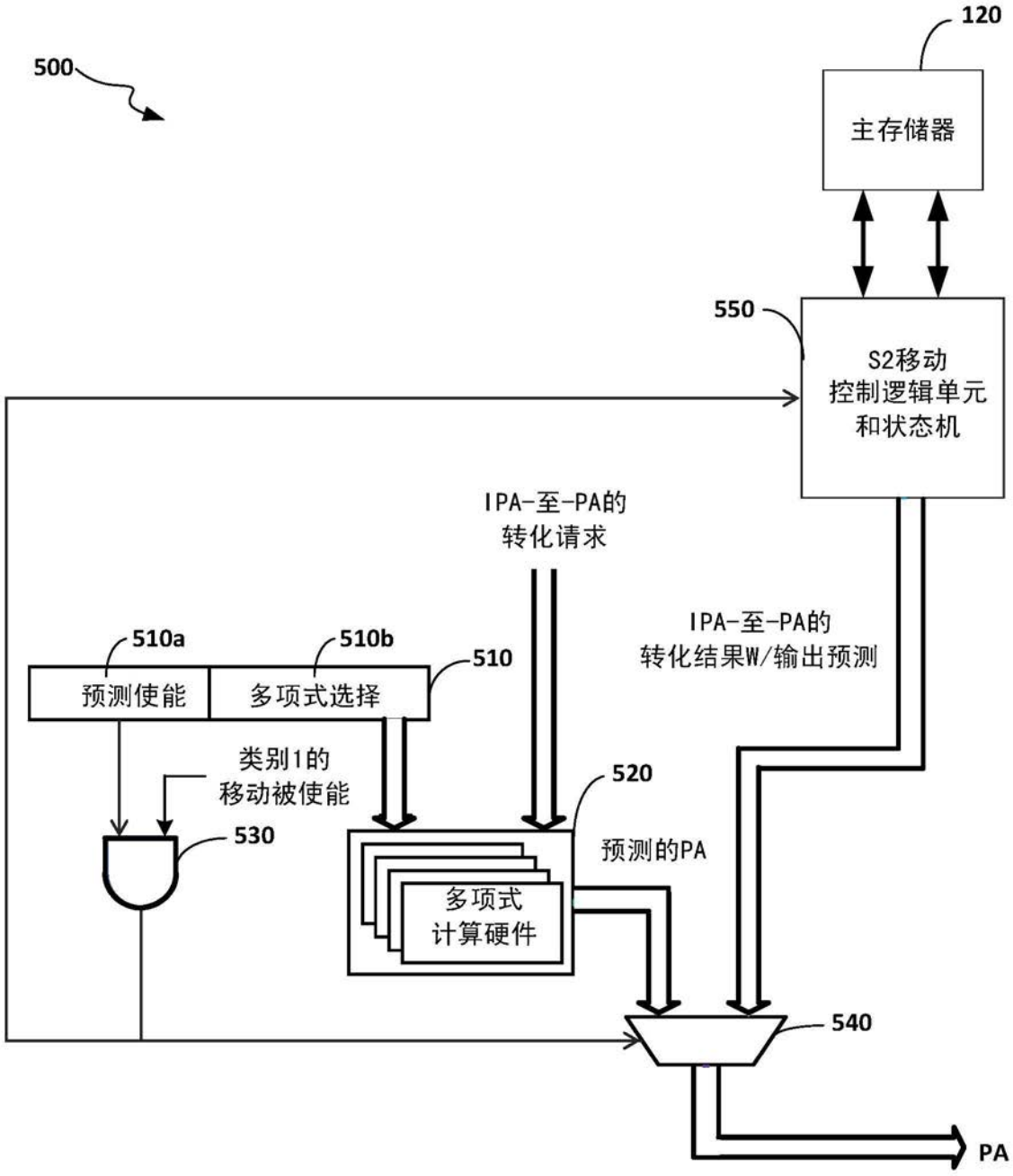

500
120
主存储器
550
S2移动
控制逻辑单元
和状态机
IPA-至-PA的
转化请求
IPA-至-PA的
转化结果W/输出预测
510a
510b
510
预测使能
多项式选择
类别1的
移动被使能
530
520
预测的PA
多项式
计算硬件
540
PA

图5

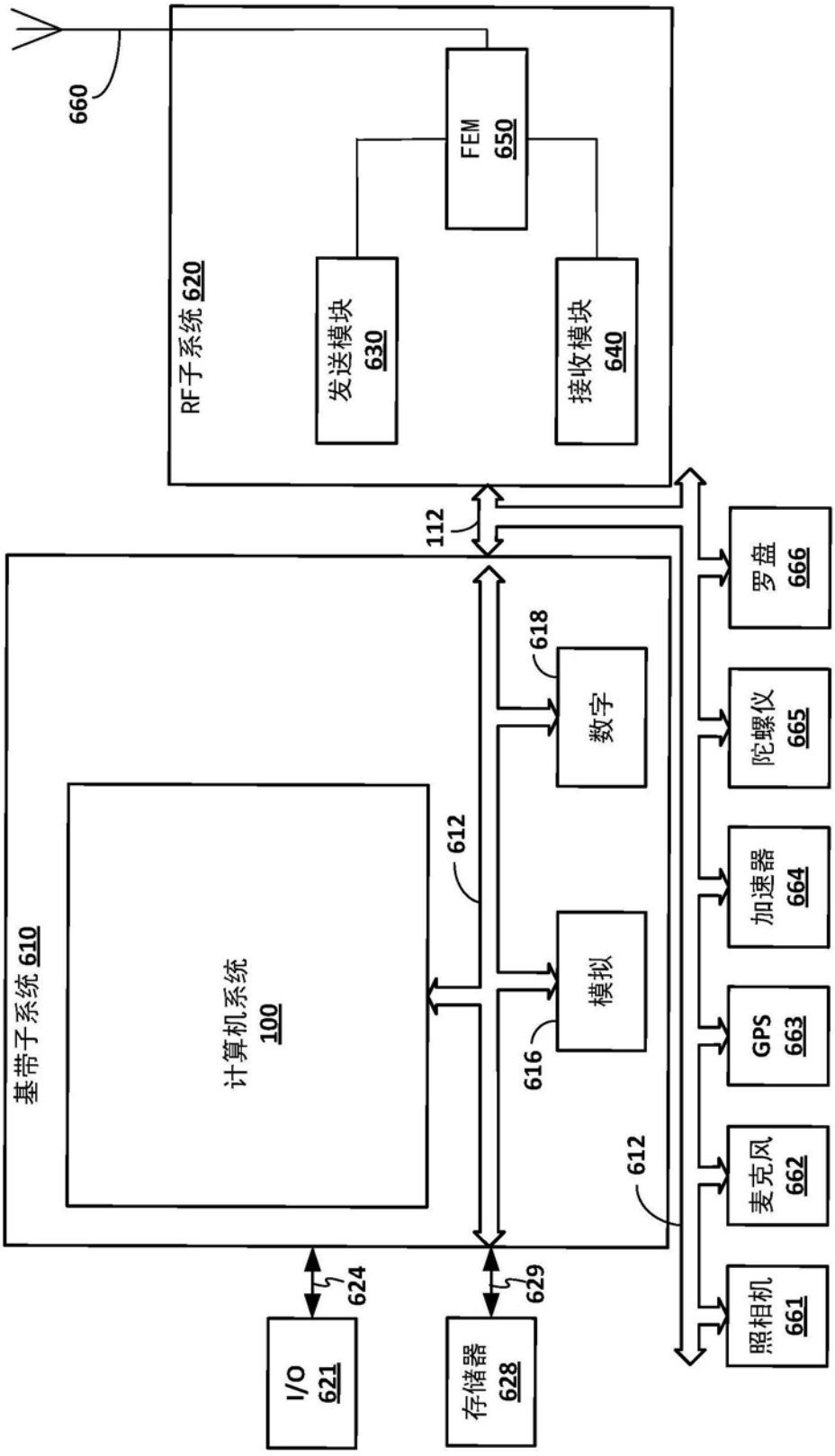

660
RF子系统 620
FEM 650
发送模块 630
接收模块 640
112
基带子系统 610
计算机系统 100
618
数字
612
模拟
616
612
罗盘 666
陀螺仪 665
加速器 664
GPS 663
麦克风 662
照相机 661
624
629
I/O 621
存储器 628

图6