



(12) 发明专利

(10) 授权公告号 CN 106776379 B

(45) 授权公告日 2021.09.07

(21) 申请号 201611013967.9

(22) 申请日 2012.07.24

(65) 同一申请的已公布的文献号
申请公布号 CN 106776379 A

(43) 申请公布日 2017.05.31

(30) 优先权数据
13/191,327 2011.07.26 US

(62) 分案原申请数据
201280031876.1 2012.07.24

(73) 专利权人 英特尔公司
地址 美国加利福尼亚州

(72) 发明人 R·M·桑卡兰 A·考克 P·兰兹
A·K·马里克 J·B·克罗斯兰
A·纳瓦勒 G·尼格
A·V·安德森

(74) 专利代理机构 上海专利商标事务所有限公
司 31100

代理人 何焜

(51) Int.Cl.
G06F 12/1027 (2016.01)

(56) 对比文件
US 2004215898 A1, 2004.10.28
US 2002078308 A1, 2002.06.20
CN 101149707 A, 2008.03.26
CN 1617111 A, 2005.05.18
Moon-Seok Chang, Kern Koh. Lazy TLB
Consistency for Large-Scale
Multiprocessors.《IEEE》.1999, 第308-315页.

审查员 郑艳梅

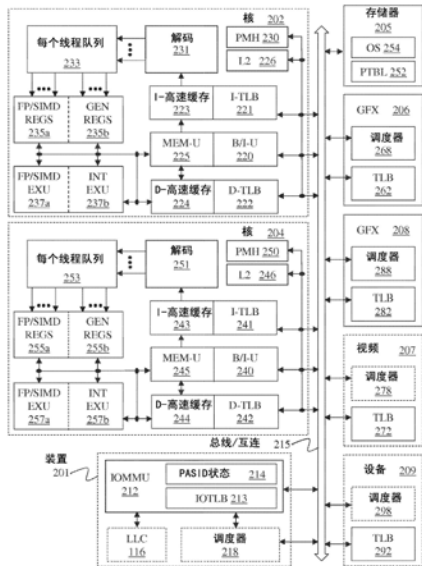
权利要求书1页 说明书10页 附图6页

(54) 发明名称

用于共享虚拟存储器的系统

(57) 摘要

本申请公开了用于在支持共享虚拟存储器的异构计算系统中的TLB关闭的方法和装置。公开了用于在多核系统中的共享虚拟存储器的异构设备的有效的TLB(转换后备缓冲器)关闭的方法和装置。用于有效的TLB关闭的装置的实施例可以包括:TLB,用于存储虚拟地址转换条目;以及与所述TLB耦合的存储器管理单元,用于保持与所述虚拟地址转换条目相对应的PASID(进程地址空间标识符)状态条目。PASID状态条目可以包括活跃参考状态和懒惰无效状态。该存储器管理单元可以响应于接收来自多核系统中的设备的PASID状态更新请求而执行PASID状态条目的原子修改,并读取PASID状态条目的懒惰无效状态。该存储器管理单元可以在响应各自的懒惰无效状态的激活之前将PASID状态更新响应发送到该设备以同步TLB条目。



1. 一种用于共享虚拟存储器的系统,包括:
异构设备,用于向存储器管理单元发出用于激活线程的状态更新请求;
所述存储器管理单元,用于:
响应于接收来自所述异构设备的所述状态更新请求而执行对与所述线程对应的地址空间条目的修改,以读取懒惰无效状态;
对于所述对应的地址空间条目状态执行懒惰无效校验,以确定所读取的懒惰无效状态的状态;以及
响应于所述懒惰无效状态被预先设置,为所述状态更新请求设置设备同步标记以向所述异构设备指示所述异构设备需要在激活之前同步所述异构设备的TLB条目。
2. 如权利要求1所述的系统,其特征在于,所述存储器管理单元是IOMMU。
3. 如权利要求1所述的系统,其特征在于,所述存储器管理单元用于通过状态更新响应来回应所述状态更新请求。
4. 如权利要求1-3中的任一项所述的系统,其特征在于,所述异构设备是非中央处理单元设备。
5. 如权利要求1-3中的任一项所述的系统,其特征在于,所述地址空间是利用进程地址空间标识符PASID来标识的。
6. 如权利要求1-3中的任一项所述的系统,其特征在于,所述地址空间条目对应于IOTLB中的条目。
7. 如权利要求3所述的系统,其特征在于,所述异构设备用于在所述线程激活之前并且响应于所述状态更新响应来同步TLB条目。
8. 一种用于共享虚拟存储器的系统,包括:
异构设备,用于向存储器管理单元发出用于停用线程的状态更新请求;
所述存储器管理单元,用于:
响应于接收来自所述异构设备的所述状态更新请求而执行对与所述线程对应的地址空间条目的修改,以读取懒惰无效状态;
对于所述对应的地址空间条目状态执行懒惰无效校验,以确定所读取的懒惰无效状态的状态;以及
响应于所述懒惰无效状态未被预先设置,为所述状态更新请求清除设备同步标记以向所述异构设备指示所述异构设备不需要在激活之前同步所述异构设备的TLB条目。
9. 如权利要求8所述的系统,其特征在于,所述存储器管理单元是IOMMU。
10. 如权利要求8所述的系统,其特征在于,所述存储器管理单元用于通过状态更新响应来回应所述状态更新请求。
11. 如权利要求8-10中的任一项所述的系统,其特征在于,所述异构设备是非中央处理单元设备。
12. 如权利要求8-10中的任一项所述的系统,其特征在于,所述地址空间是利用进程地址空间标识符PASID来标识的。
13. 如权利要求8-10中的任一项所述的系统,其特征在于,所述地址空间条目对应于IOTLB中的条目。

用于共享虚拟存储器的系统

[0001] 本发明专利申请是国际申请号为PCT/US2012/047991,国际申请日为2012年7月24日,进入中国国家阶段的申请号为201280031876.1,名称为“用于在支持共享虚拟存储器的异构计算系统中的TLB关闭的方法和装置”的发明专利申请的分案申请。

技术领域

[0002] 本公开总体上涉及计算和/或通信领域。特别地,本公开涉及用于在多核系统中共享虚拟存储器的异构设备的有效的TLB(转换后备缓冲器)关闭(shoot-down)。

背景技术

[0003] 现代的处理器的通常包括高速缓存,以通过在具有更短的存取延迟的较小的速度更快的存储器中提供指令和/或数据的副本,以改善存取外部存储器的性能。此外,通过使用设备(例如转换后备缓冲器(TLB))以高速缓存虚拟到物理地址的转换,例如来自分页的虚拟存储系统中的系统页表的虚拟到物理地址的转换,这样的高速缓存可以为快速虚拟到物理地址的转换提供支持。当TLB将虚拟地址与TLB中存储的转换之一相匹配时,我们可以将这样的事件称为TLB命中,并且所检索的物理地址可以被用来更快地访问高速缓存或主存储器中的存储器。当TLB未将虚拟地址与TLB中存储的转换之一相匹配时,我们可以将这样的事件称为TLB未命中(miss)或页面未命中,并且转换通过查找存储器中的相应页表而继续。

[0004] 典型的多核处理器或多核处理系统可以具有各种同构或异构的处理设备——例如,一定数目的同构处理器核连同—个或多个异构图形处理器或视频处理器等。这些同构或异构处理设备可以共享虚拟存储器地址空间。通常,同构处理器核可以具有高速缓存层次结构和上述的TLB,以高速缓存来自分页的虚拟存储器系统中的系统页表的虚拟到物理地址的转换。在现代处理器中,同构处理器核还可以包括硬件中的页未命中处理器(PMH),用于更快地通过系统页表执行页面行走。当这些同构处理器核中的一个试图访问未映射到物理存储器中任何地址的虚拟存储器中的页面时,所述存储器访问将会导致页面错误,并且将调用操作系统来“修复”该问题。

[0005] 异构处理设备(例如图形处理器)还可以具有例如TLB的机构,用于执行虚拟到物理地址的转换,但是,通常存在一些软件,所述一些软件运行于同构处理器核上,并可能与用于异构处理设备的驱动器通信以建立页表等,以实现与异构处理设备的对分页虚拟存储器系统的一些有限的共享。但是,不像在同构处理器核中,这些机构可能具有有限的能力,例如缺乏处理页面故障的能力。因此,由于需求决定,与这些异构处理设备共享的页面可能需要在存储器中锁定,从而防止操作系统从磁盘中将它们分页,以及将它们向外分页到磁盘。因此,可被支持的应用程序的数量可能是有限的,或者所需的物理存储器的量可能大于以其他方式所需的物理存储器量。此外,运行在同构处理器上以建立页表等、用于用异构处理设备共享分页的虚拟存储器的软件引起性能开销,这可能减少使用异构处理设备的有效性。

[0006] 当同构处理器核上运行的软件(例如操作系统)更改页表时,整个系统的各种TLB中高速缓存的相应的转换必须被无效。这被称为TLB关闭。一个简单的方法是将无效消息广播到该系统中所有的设备,但是这可能会导致不需要的通信拥塞。此外,为异构处理设备建立的页表和由同构处理器核所使用的页表可能不相同,并且在异构处理设备中使用的机构用来使它们的TLB条目(entries)无效、并用来独自恢复新的页表信息的能力有限。

[0007] 为了避免不需要的拥塞,操作系统已经使用了懒惰无效方案的形式。由于操作系统运行于同构处理器核上并且有权访问各种同构处理器核上的活跃的进程、应用程序或线程,它可以在必要时“关闭”或者触发TLB条目的无效,换言之,仅在活跃运行使用该TLB条目的处理的核的TLB中和/或在可能要激活将使用该TLB条目的处理的核的TLB中使TLB条目无效。因此,可避免对TLB关闭的不必要的活动和开销。但由于操作系统不在异构处理设备上运行,它无法跟踪哪些进程是活跃的,以及哪些TLB条目可能需要被无效,即使我们假定异构处理设备能够恢复新的页表信息和/或生成页面错误。因此,将需要发送广播到可能共享进程虚拟地址空间的所有的异构处理设备。

[0008] 随着处理核和其它异构处理设备访问高速缓存或使用虚拟存储器的数量增加,可能完全有理由来预期例如附加存储器拥塞的问题以及当虚拟存储器空间由更多的这些设备所共享时的冲突。迄今为止,对于这样合理预期的问题的潜在解决方案还没有得到充分的研究。

附图说明

[0009] 本发明是通过举例的方式示出,并不局限于附图中的图。

[0010] 图1示出了用于在具有异构设备的多核系统中共享虚拟存储器的现有技术。

[0011] 图2示出了用于在具有异构设备的多核系统中共享虚拟存储器的装置的一个实施例,该装置支持有效的TLB关闭。

[0012] 图3示出了用于支持有效的TLB关闭以在具有异构设备的多核系统中共享虚拟存储器的过程的一个实施例的流程图。

[0013] 图4示出了用于支持有效的TLB关闭以在具有异构设备的多核系统中共享虚拟存储器的过程的替代实施例的流程图。

[0014] 图5示出了使用支持有效的TLB关闭的过程和装置以在具有异构设备的多核系统中共享虚拟存储器的处理器的一个实施例。

[0015] 图6示出了使用支持有效的TLB关闭的过程和装置以在具有异构设备的多核系统中共享虚拟存储器的系统的一个实施例。

具体实施方式

[0016] 公开了用于在多核系统中共享虚拟存储器的异构设备的有效TLB(转换后备缓冲器)关闭的方法和装置。用于有效的TLB关闭的装置的实施例可以包括TLB以存储虚拟地址转换条目,以及与所述TLB耦合的存储器管理单元,以保持与所述虚拟地址转换条目相对应的PASID(进程地址空间标识符)状态条目。PASID状态条目可以包括活跃参考状态和懒惰无效状态。该存储器管理单元可以响应于接收来自多核系统中的设备的PASID状态更新请求而执行PASID状态条目的原子修改,并读取PASID状态条目的懒惰无效状态。该存储器管理

单元可以在响应各自的懒惰无效状态的激活之前将PASID状态更新响应发送到该设备以同步TLB条目。

[0017] 这些和其它实施例可以根据以下教导实现,并且应该明白的是:在以下的教导中可做出各种修改和变化,而不脱离本发明的更宽的精神和范围。相应地,说明书和附图将被认为是示例性的而非限制性的理解,并且仅根据权利要求及其等同物来衡量本发明。

[0018] 图1示出了用于在具有异构设备的多核系统中共享虚拟存储器的现有技术。多核处理器或多核处理系统可以具有各种同构或异构处理设备。例如,图1的系统具有一些数量的同构处理器核,其中两个示为核102和核104,连同一个或多个异构图形处理器(如GFX 106)或视频处理器(例如说,设备109)。核102和核104可以是处理多个进程的线程的多线程核,用于分别经解码131和解码151、每个线程队列133和153、浮点/单指令多数据寄存器FP/SIMD REGS 135a和FP/SIMD REGS 155a、通用寄存器GEN REGS 135b和GEN REGS 155b、浮点/单指令多数据执行单元FP/SIMD EXU 137a和FP/SIMD EXU 157a、以及整数执行单元INT EXU 137b和INT EXU 157b而执行。核102和核104也可以分别经总线/互连115和存储器单元MEM-U 125和MEM-U 145、通过总线/互连单元B/I-U 120和B/I-U 140与物理存储器105耦合。核102和核104也可以经物理存储器105和总线/互连115以及可选的末级高速缓存(LLC 116)与异构处理设备GFX 106和设备109耦合。

[0019] 这些同构或异构处理设备可以经物理存储器105以及选择性地通过末级高速缓存(LLC 116)共享虚拟存储器地址空间。通常,同构处理器核可能具有高速缓存层次结构,例如分别是:I-高速缓存123、D-高速缓存124、L2 126和I-高速缓存143、D-高速缓存144、L2 166;以及TLB,例如分别是:I-TLB 121、D-TLB 122和I-TLB 141、D-TLB 142,用于在分页的虚拟存储器系统中高速缓存来自所述系统(OS 154)、页表(PTBL 152)的虚拟到物理地址的转换。

[0020] 同构处理器核102和104也可以包括硬件中的页面未命中处理程序,PMH 130和PMH 150,以通过系统页表(PTBL 152)执行页面行走(page walk)。当这些同构处理器核102和104中的一个尝试访问未映射到物理存储器105中任何地址的虚拟存储器中的页面时,该存储器访问将导致页面错误,并将调用操作系统(OS 154)以“修复”该问题。

[0021] 异构处理设备,例如图形处理器(GFX 106)以及设备109还可以具有例如TLB的机构,用于执行虚拟到物理地址的转换,但是,通常存在一些软件,这些软件运行于同构处理器核102和104上,并可能与用于异构处理设备GFX 106和设备109的驱动器通信以建立页表等,以实现与异构处理设备GFX 106和设备109的对分页虚拟存储器系统的一些有限的共享。但是,不像在同构处理器核102和104中,这些机构可能具有有限的能力,例如缺乏处理页面故障的能力。因此,由于需求决定,与这些异构处理设备共享的页面可能需要在物理存储器105中锁定,从而防止操作系统(OS 154)从磁盘中将它们分页,以及将它们向外分页到磁盘。因此,可被支持的应用程序的数量可能是有限的,或者所需的物理存储器的量可能大于以其他方式所需的物理存储器量。此外,运行在同构处理器102或104上以建立页表等、用于与所述异构处理设备GFX 106和设备109共享分页的虚拟存储器的软件引起了性能开销,这可能减少使用异构处理设备GFX 106和设备109的有效性。

[0022] 当同构处理器核102和104上运行的软件(例如操作系统,OS 154)更改页表PTBL 152时,分别被高速缓存入各种TLB(例如I-TLB 121、D-TLB 122和I-TLB 141、D-TLB 142)以

及GFX 106和/或设备109中的任何TLB中的对应转换必须被无效。将无效消息广播至系统中的所有设备可能会导致不需要的通信拥塞。为了避免不需要的拥塞,操作系统(OS 152)可以使用懒惰无效方案。由于操作系统(OS 152)运行于同构处理器核102和104上并有权访问各种同构处理器核上的活跃的进程、应用程序或线程,它可以在必要时“关闭”TLB条目或者触发TLB条目的无效,换句话说,仅在活跃运行使用该TLB条目的处理的核的TLB(分别是I-TLB 121、D-TLB 122和I-TLB 141、D-TLB 142)中和/或在可能想要激活将使用该TLB条目的处理的核的TLB中使TLB条目无效。因此,可避免对TLB关闭的不必要的活动和开销,但只用于同构处理器核102和104。由于操作系统(OS 152)不在异构处理设备GFX 106和设备109上运行,它无法跟踪哪些进程是活跃的以及哪些TLB条目可能需要被无效,即使我们假定异构处理设备(GFX 106和设备109)能够恢复的新的页表信息并生成页面错误。因此,将需要向可能共享进程虚拟地址空间的所有的异构处理设备(GFX 106和设备109)发送广播。

[0023] 将理解的是:随着访问高速缓存或使用虚拟存储器的处理核和其它异构处理设备的数量增加,可能完全有理由来预期例如附加存储器拥塞的问题,以及当虚拟存储器空间由更多的这些设备所共享时的冲突或竞争状况。

[0024] 图2示出了用于在具有异构设备的多核系统中共享虚拟存储器的装置201的一个实施例,该装置支持有效的TLB关闭。装置201包括IOTLB 213以存储虚拟地址转换条目。存储器管理单元IOMMU 212与该IOTLB 213和PASID状态存储器214耦合,以保持对应于IOTLB 213的虚拟地址转换条目的至少一部分的PASID状态条目。PASID状态条目的实施例可各自包含活跃参考状态和懒惰无效状态,这将在下面更详细地描述。为了说明,IOTLB 213和PASID状态存储器214被示为包括在IOMMU 212中,但将理解的是,IOTLB 213和PASID状态存储器214中的一者或两者可以被实现为单独的硬件或软件数据结构,并驻留在IOMMU 212外部,例如在存储器205中。

[0025] 装置201可选地还包括末级高速缓存(LLC 116)和调度器218,LLC 116和调度器218被示为包括在IOMMU 212中,但可以被实现为单独的硬件或软件,并可以驻留在IOMMU 212外部。

[0026] 装置201与总线/互连215操作地耦合,用于与具有异构设备的多核处理器或多核处理系统通信,并用于在具有异构设备的多核系统中经由存储器205共享虚拟存储器。该系统可以包括多个同构核(其中两个被示为核202和核204)以及异构设备(例如图形设备,其中两个被示为GFX 206和GFX 208)、以及可选的其它异构设备(例如视频设备207和设备209)。

[0027] 多个同构核202和核204可以是处理多个进程的线程的多线程核,用于分别经解码231和解码251、每个线程队列233和253、浮点/单指令多数据寄存器FP/SIMD REGS 235a和FP/SIMD REGS 255a、通用寄存器GEN REGS 235b和GEN REGS 255b、浮点/单指令多数据执行单元FP/SIMD EXU 237a和FP/SIMD EXU 257a、以及整数执行单元INT EXU 237b和INT EXU 257b的执行。核202和核204也可以分别经总线/互连215与存储器205耦合并通过总线/互连单元B/I-U 220和B/I-U 240与存储器单元MEM-U 225和MEM-U 245耦合。核202和核204也可以经存储器205和总线/互连215以及可选地经末级高速缓存(LLC 116)与异构处理设备GFX 206和GFX 208耦合,以及可选地与其他异构设备(例如视频设备207和设备209)耦合。

[0028] 这些同构或异构处理设备可以经存储器205以及选择性地通过末级高速缓存(LLC 116)共享虚拟存储器地址空间。通常,同构处理器核可能具有高速缓存层次结构,例如分别是:I-高速缓存223、D-高速缓存224、L2226和I-高速缓存243、D-高速缓存244、L2 266;以及TLB,例如分别是:I-TLB 221、D-TLB 222和I-TLB 241、D-TLB 242,用于在分页的虚拟存储器系统中高速缓存来自所述系统(OS 254)、页表(PTBL 252)的虚拟到物理地址的转换。

[0029] 同构处理器核202和204也可以包括硬件中的页面未命中处理程序(PMH 230和PMH 250),以通过系统页表(PTBL 252)执行页面行走。当这些同构处理器核202和204中的一个尝试访问未映射到物理存储器205中任何地址的虚拟存储器中的页面时,该存储器访问将导致页面错误,并将调用操作系统(OS 254)以“修复”该问题。

[0030] 异构处理设备(例如图形处理器GFX 206和GFX 208)以及可选的其他异构设备(例如视频设备207和设备209)还可能具有例如TLB(例如分别是TLB 262、TLB 282、TLB272和TLB 292)的机构,用于执行虚拟到物理地址的转换。TLB 262、TLB 282、TLB272和TLB 292的各种实施例分别可能具有或可能不具有相同的能力,或堪比同构处理器核202和204的能力。异构处理设备GFX 206、GFX 208以及可选的视频设备207和设备209也可以具有调度器,例如SCHED 268、SCHED 288、SCHED 278和SCHED 298,分别用于调度线程以在异构处理设备GFX 206、GFX 208、视频设备207和设备209上执行。所述调度器SCHED 268、SCHED 288、SCHED 278和SCHED 298(并且可选地包括中央调度器SCHED 218)可以分别发出来自异构处理设备的PASID状态更新请求,以便激活或停用该设备上的进程线程。

[0031] 存储器管理单元IOMMU 212可能接收来自多核系统中的多个异构设备的一个或多个各自的请求设备的一个或多个PASID状态更新请求。当处理来自一个请求设备的PASID状态更新请求时,IOMMU 212可以执行对应于此PASID状态更新请求的PASID状态的原子修改。这样的原子修改的一个实施例可以是原子的读-修改-写操作的形式,其中IOMMU 212可以读取PASID状态条目的懒惰无效状态,并为相应的PASID状态执行懒惰无效校验,例如,以查明是否需要在由请求设备激活之前将用于进程地址空间的虚拟地址转换条目与系统页表同步。该IOMMU 212然后可将PASID状态更新响应发送到对应于该PASID状态更新请求的请求设备,例如,在响应于懒惰无效状态被设置而激活之前,告诉该请求设备同步该设备的TLB条目。

[0032] 还将在下文中更详细地、尤其是参考图3-4来描述用于支持有效TLB关闭以在具有异构设备的多核系统中共享虚拟存储器的一个或多个过程的实施例。

[0033] 图3示出了用于支持有效的TLB关闭以在具有异构设备的多核系统中共享虚拟存储器的过程301的一个实施例的流程图。本文所公开的过程301和其他过程由处理块来执行,所述处理块可以包括可由通用机器或由专用机器或由两者的组合执行的专用硬件或软件或固件操作代码。

[0034] 在过程301的处理块310中,从多核系统中多个异构设备的一个或多个各自的请求设备接收一个或多个PASID状态更新请求。在处理块320中,与那些PASID状态更新请求之一对应地执行PASID状态的原子修改,并且对处理块330中相应的PASID状态执行懒惰无效校验。所述原子修改可以是原子读-修改-写操作的形式,其中IOMMU可以读取PASID状态条目的懒惰无效状态,并为相应的PASID状态执行懒惰无效校验,例如,查明是否需要在由请求设备激活之前使用于进程地址空间的虚拟地址转换条目与系统页表同步。然后在处理块

350,将PASID状态更新响应发送到与该特定PASID状态更新请求相对应的请求设备,例如,在响应于懒惰无效状态被设置而激活之前,指示该请求设备需要同步设备TLB条目。

[0035] 然后在处理块360中出现了来自多核系统的操作系统的TLB关闭的处理。一旦IOTLB条目被激活,TLB关闭的处理就可能以会使IOTLB条目无效开始。这样的处理也可以包括:例如,如果PASID状态条目中的活跃参考计数不为零,则使一个或多个异构设备的设备TLB条目无效。可替代地,处理TLB关闭可能包括:如果PASID状态条目中的活跃参考计数为零,即当前没有设备使用该TLB条目但在一些设备TLB中它仍可能被激活时,读取PASID状态条目并在PASID状态条目中设置懒惰无效字段。

[0036] 将理解的是,为了说明,在过程301的一些实施例中,导致处理发生的事件可能以其他顺序发生,或随其他事件发生而同时发生。例如,可在任何时间从操作系统接收TLB关闭,与此同时IOMMU可能正在处理PASID状态更新请求。因此,防止竞争状况以及当心对于PASID状态和TLB条目的更新的排序保留那些结构中的虚拟地址转换关于系统页表变化的正确性是重要的。

[0037] 图4示出了用于支持有效的TLB关闭以在具有异构设备的多核系统中共享虚拟存储器的过程的替代实施例的流程图。

[0038] 在过程401的处理块410中,从多核系统中的多个异构设备中的请求设备接收PASID状态更新请求。在处理块415中确定该更新请求是否将激活或停用用于该请求设备的PASID状态。如果该请求将激活用于该请求设备的PASID状态,则在处理块420中,与该PASID状态更新请求相对地执行PASID状态的原子修改,其中懒惰无效状态被读取,活跃的参考计数RCOUNT递增,而懒惰无效状态被清除。该原子修改可以是原子的读-修改-写(read-modify-write)操作的形式。否则,如果该请求将停用用于该请求设备的PASID状态,则在处理块425中,递减该活跃参考计数RCOUNT。

[0039] 在处理块430中执行对于相应PASID状态的懒惰无效校验,由于在原子修改中,该IOMMU已读取了该PASID状态条目的懒惰无效状态,所以对于相应PASID状态的懒惰无效校验可包括校验所读取的状态,以查明它是否被预先设置,例如,用于确定用于进程地址空间的虚拟地址转换条目是否需要在由请求设备激活之前与系统页表进行同步。

[0040] 如果是这样,则处理块440以使IOTLB条目无效开始,并进一步包括设置同步标记以使一个或多个需要与系统页表同步的异构设备的设备TLB条目无效,因为PASID状态条目中的活跃参考计数现在是非零,并且懒惰无效状态先前曾被设置。然后在处理块450中,将PASID状态更新响应发送到与该特定PASID状态更新请求相对应的请求设备,并将同步标记置位,向该请求设备指示它需要在激活之前同步该设备TLB条目。

[0041] 否则,处理进行到处理块435,如同在停用请求的情况下,其中所述同步标记被清除。然后在处理块445中,将PASID状态更新响应发送到与该特定PASID状态更新请求相对应的请求设备,并将同步标记清除,向该请求设备指示不需要在激活之前同步该设备TLB条目。

[0042] 在处理块455中,请求设备可能需要在激活之前基于所述PASID状态更新响应中同步标记的设置来同步该设备TLB条目。将理解的是,每当PASID状态条目中活跃的参考计数从零增加到一并且懒惰无效状态先前曾被设置时,就将以这种方式设置同步标记。也可使用基于懒惰无效状态并用于多个设备的同步的其他实施例。

[0043] 然后在处理块460中开始的过程402中出现了来自多核系统的操作系统的TLB关闭的处理,其中确定全局页面是否已被修改。如果是,则TLB关闭的处理可以包括将IOTLB条目和处理块475中异构设备的所有相应的设备TLB条目无效。否则,处理可以前进到处理块465,其中读取对应的PASID状态条目。然后在处理块470中,确定PASID状态条目中的活跃参考计数是否是零。如果不是,则该TLB关闭的处理再次包括在处理块475中使IOTLB条目和异构设备的所有相应的设备TLB条目无效。但是,如果PASID状态条目中的活跃参考计数为零,则设置PASID状态条目中的懒惰无效字段,以触发如上所述的关于处理块420、430和440的懒惰无效。该过程然后重申以过程401的处理块410开始。

[0044] 将理解的是,为了说明,在过程401和402的一些实施例中,导致过程401和402的处理发生的事件可能以其他顺序或同时发生。例如,可在任何时间从操作系统接收TLB关闭并导致过程402的处理发生,而在过程401中,IOMMU可能处理PASID状态更新请求。因此,防止竞争状况以及当心对PASID状态和TLB条目的更新的排序保留那些结构中的虚拟地址转换关于系统页表的变化正确性是重要的。本领域技术人员可以在各种实施例中使用原子操作,例如读-修改-写和测试-和-设置,或用于处理临界区的其他技术(例如信号灯)来实现这些目标。

[0045] 图5示出了使用支持有效的TLB关闭的过程(例如过程301)和装置501以在具有异构设备506-509的多核系统中共享虚拟存储器的处理器503的一个实施例。处理器503的装置501包括IOTLB 513以存储虚拟地址转换条目。存储器管理单元(MMU) 512与该IOTLB 513和PASID状态存储器514耦合,以保持对应于IOTLB 513的虚拟地址转换条目的至少一部分的PASID状态条目。该PASID状态条目的实施例可各包括活跃参考状态和懒惰无效状态。为了说明,IOTLB 513和PASID状态存储器514被示为包括在MMU 512中,但将理解的是,IOTLB 513和PASID状态存储器514中的一者或两者可以被实现为单独的硬件或软件数据结构,并驻留在MMU 512外部,例如在存储器505中。

[0046] 处理器503的装置501可选地还包括末级高速缓存(LLC 516)和调度器518,LLC 516和调度器518被示为包括在MMU 512中,但可以被实现为单独的硬件或软件,并可以驻留在MMU 512外部。

[0047] 处理器503的装置501可操作地与总线/互连515和551耦合,以与具有异构设备的多核处理器或多核处理系统通信,并用于经由存储控制器553、通过存储器505而在具有异构设备的多核系统中共享虚拟存储器。该系统可以包括多个同构核(其中两个被示为核502和核504)以及异构设备(例如图形设备,其中两个被示为GFX 506和GFX 508)、以及可选的其它异构设备(例如视频设备507和设备509)。

[0048] 例如关于图2,多个同构核502和504可以是用于所述执行的多个进程线程的多线程核。核502和核504可以经总线/互连515与各种设备耦合,例如与I/O扩展设备537、NAND控制器557、传输处理器558、安全处理器559、视频显示逻辑527、音频/视频I/O 548、音频解码逻辑549以及可选的单指令多数据(SIMD)协处理器591耦合。核502和核504也可以经总线/互连551和存储控制器553与存储器505耦合。核502和核504也可以经存储器205和总线/互连515和551以及可选地经末级高速缓存(LLC 516)与异构处理设备GFX 506和GFX 508耦合,以及可选地与其他异构设备(例如视频设备507和设备509)耦合。

[0049] 这些同构或异构处理设备可以经存储器505以及选择性地通过末级高速缓存(LLC

516) 共享虚拟存储器地址空间。通常,同构处理器核可能具有:高速缓存层次结构,例如分别包括L2 526和L2 546;以及TLB,例如分别是TLB 522和TLB 542,用于在分页的虚拟存储器系统中高速缓存来自所述系统(OS 554)、页表(PTBL 552)的虚拟到物理地址的转换。

[0050] 同构处理器核502和504也可以包括硬件中的页面未命中处理程序(未示出),以通过系统页表(PTBL 552)执行页面行走。当这些同构处理器核502和504中的一个尝试访问未映射到物理存储器505中任何地址的虚拟存储器中的页面时,该存储器访问将导致页面错误,并将调用操作系统(OS 554)以“修复”该问题。

[0051] 异构处理设备(例如图形处理器GFX 506和GFX 508)以及可选的其他异构设备(例如视频设备507和设备509)还可能具有例如TLB(分别例如是TLB 562、TLB 582、TLB 572和TLB 592)的机构,用于执行虚拟到物理地址的转换。TLB 562、TLB 582、TLB 572和TLB 592的各种实施例分别可能具有或可能不具有相同的能力,或堪比同构处理器核502和504的能力。异构处理设备GFX 506、GFX 508以及可选的视频设备507和设备509也可以具有调度器,例如SCH 568、SCH 588、SCH 578和SCH 598,分别用于调度进程线程以在异构处理设备GFX 506、GFX 508、视频设备507和设备509上执行。所述调度器SCH 568、SCH 588、SCH 578和SCH 598(并可选地包括中央调度器SCH 518)分别可以发出来自异构处理设备的PASID状态更新请求,以便激活或停用该设备上的进程线程。

[0052] 存储器管理单元512可能接收来自多核系统中的多个异构设备的一个或多个各自的请求设备的一个或多个PASID状态更新请求。当处理来自一个请求设备的PASID状态更新请求时,MMU 512可以执行对应于此PASID状态更新请求的PASID状态的原子修改。这样的原子修改的一个实施例可以是原子的读-修改-写操作的形式,其中MMU 512可以读取PASID状态条目的懒惰无效状态,并对相应的PASID状态执行懒惰无效校验,例如,以查明是否需要在由请求设备激活之前将用于进程地址空间的虚拟地址转换条目与系统页表同步。该MMU 512然后可将PASID状态更新响应发送到与该PASID状态更新请求相对应的请求设备,例如,告诉该请求设备,响应于懒惰无效状态被设置,在激活之前同步该设备的TLB条目。

[0053] 图6示出了利用支持有效的TLB关闭的过程(例如过程301)和装置601以在具有异构设备606-609的多核系统中共享虚拟存储器的系统610的一个实施例。

[0054] 系统610包括处理器603的装置601,其包括IOTLB 613以存储虚拟地址转换条目。存储器管理单元612与该IOTLB 613和PASID状态存储器614耦合,以保持与IOTLB 613的虚拟地址转换条目的至少一部分相对应的PASID状态条目。该PASID状态条目的实施例可各包括活跃参考状态和懒惰无效状态。为了说明,IOTLB 613和PASID状态存储器614被示为包括在MMU 612中,但将理解的是,IOTLB 613和PASID状态存储器614中的一者或两者可以被实现为单独的硬件或软件数据结构,并驻留在MMU 612外部,例如在存储器605中。

[0055] 处理器603的装置601可选地还包括末级高速缓存(LLC 616)和调度器618,LLC 616和调度器618被示为包括在MMU 612中,但可以被实现为单独的硬件或软件,并可以驻留在MMU 612外部。

[0056] 处理器603的装置601可操作地耦接总线/互连615和651,用于与具有异构设备的多核处理器或多核处理系统通信,并用于经由存储控制器653、通过存储器605而在具有异构设备的多核系统中共享虚拟存储器。系统610的实施例可以使用标准的或非标准的或专有的技术、接口、总线或互连615和651(例如用于与多核处理器或多核处理系统通信的(外

围组件互连)PCI或PCI Express或(串行高级技术附件)SATA)来实现。

[0057] 系统610的其它实施例可使用标准的或非标准的或专有的技术、接口、总线或互连来实现,例如:SPI(串行外围接口)总线;ISA(工业标准结构)总线、PC/104、PC/104+和扩展ISA、USB(通用串行总线)AVC(音频视频类);AMBA(高级微控制器总线架构)(高级外设总线)APB;火线(IEEE标准1394a-2000高性能串行总线-修订1,ISBN 0-7381-1958-X;IEEE标准1394b-2002高性能串行总线-修订2,ISBN 0-7381-3253-5;IEEE标准1394c-2006,2007-06-08,ISBN 0-7381-5237-4)、HDMI(高清晰度多媒体接口)、VESA(视频电子标准协会)的DisplayPort和Mini DisplayPort;MIPI[®](移动行业处理器接口)联盟的SLIMbus[®](串行低功耗芯片间媒体总线)、LLI(低延迟接口)、CSI(照相机串行接口)DSI(显示屏串行接口)等。

[0058] 系统610可以包括多个同构核(其中两个被示为核602和核604)、以及异构设备(例如图形设备,其中两个被示为GFX 606和GFX 608)、以及任选的其它异构设备(例如视频设备607和设备609)。

[0059] 多个同构核602和604可以是用于执行的处理多个进程线程的多线程核。同构处理器核602和核604可以经由总线/互连615耦接各种设备,例如桥接器30、无线连接设备20、调制解调器设备26和音频输入/输出设备28。系统610的一些实施例可以实现为芯片上系统,例如,在平板电脑或智能电话中使用。在这种实施例中,无线连接设备20可提供无线LAN(局域网)链路,调制解调器设备26可提供4G(第四代)、5G(第五代)或更高版本的电话链路,而音频I/O设备28可以提供一组音频人机接口设备,例如:耳机、音箱、手持设备话筒、音频输入和输出通道以及放大器。

[0060] 同构处理器核602和604与总线/互连615、桥接器30和总线/互连18耦合,用于与各种其它系统设备通信,所述各种其它系统设备可以包括但不限于:无线连接设备20、调制解调器设备26和音频I/O设备28、照相机接口21、快速IrDA(红外数据协会)端口23、HD(高清晰度)多媒体接口24、USB 25、显示控制器27以及备用主接口29。同构处理器核602和604还与总线/互连615、桥接器30和总线/互连11耦合,用于与各种其它系统设备通信,所述各种其它系统设备可以包括但不限于:闪速存储器13、SD(安全数字)存储器16、MMC(多媒体卡)17和SSD(固态驱动器)19。同构处理器核602和604与总线/互连615、桥接器30和总线/互连18耦合,用于与各种其它系统设备通信,所述各种其它系统设备可以包括但不限于:UART(通用异步接收机/发射机)31、照相机控制器32、可选地包括符合Wi-Fi 802.11a/b/g的收发机和/或GPS(全球定位系统)接收机的蓝牙UART 33、键盘34、电池控制器35、I/O扩展37和触摸屏控制器39。

[0061] 同构处理器核602和核604也可经由总线/互连651和存储控制器653耦接存储器605。同构处理器核602和核604也可经由存储器605和总线/互连615和651以及可选地经由末级高速缓存(LLC 616)与异构处理设备GFX 606和GFX 608耦合,以及可选地与其它异构设备(例如视频设备607和设备609)耦合。存储器605和系统610的其他有形的存储介质可记录功能描述性材料,所述功能描述性材料包括OS 654、MMU 612、SCH 618、668、678、688和698以及其他设备和/或驱动器(未示出)的可执行指令,所述可执行指令可以由同构处理器核602和604执行以支持有效的TLB关闭,用于在具有异构设备606-609的多核系统中共享虚拟存储器。

[0062] 系统610的一些实施例可以遵循行业标准,其允许在一台计算机内同时运行多个操作系统以本地共享设备,所述设备比如单根I/O虚拟化(SRIOV),其在PCI Express拓扑结构中提供本地I/O虚拟化,或者比如多根I/O虚拟化(MRIOV),其在多根复合体共享PCI Express层次结构的拓扑结构中提供本地I/O虚拟化。系统610的一些优选实施例可以包括标准的或非标准的或专有的技术、接口、总线或互连,例如:SPI总线、USB、AMBA APB;火线、HDMI、Mini DisplayPort、MIPI SLIMbus、MIPI LLI、MIPI CSI、MIPI DSI等。

[0063] 这些同构或异构处理设备可以经存储器605以及选择性地通过末级高速缓存(LLC 616)共享虚拟存储器地址空间。通常,同构处理器核可能具有:高速缓存层次结构,例如分别是L2 626和L2 646;以及TLB,例如分别是TLB 622和TLB 642,用于在分页的虚拟存储器系统中高速缓存来自主机或客户系统(OS 654)、页表(PTBL 652)的虚拟到物理地址的转换。

[0064] 同构处理器核602和604也可以包括硬件中的页面未命中处理程序(未示出)以通过系统页表(PTBL 652)执行页面行走。当这些同构处理器核602和604中的一个尝试访问未映射到物理存储器605中任何地址的虚拟存储器中的页面时,该存储器访问将导致页面错误,并将调用操作系统(OS 654)以“修复”该问题。

[0065] 异构处理设备(例如图形处理器GFX 606和GFX 608)以及可选的其他异构设备(例如视频设备607和设备609)还可能具有例如TLB(例如分别是TLB 662、TLB 682、TLB 672和TLB 692)的机构,用于执行虚拟到物理地址的转换。TLB 662、TLB 682、TLB 672和TLB 692的各种实施例可能分别具有或可能不具有相同的能力,或堪比同构处理器核602和604的能力。异构处理设备GFX 606、GFX 608以及可选的视频设备607和设备609也可以具有调度器,例如SCH 668、SCH 688、SCH 678和SCH 698,分别用于调度进程线程以在异构处理设备GFX 606、GFX 608、视频设备607和设备609上执行。所述调度器SCH 668、SCH 688、SCH 678和SCH 698(以及可选地包括中央调度器SCH 618)分别可以发出来自异构处理设备的PASID状态更新请求,以便激活或停用该设备上的进程线程。

[0066] 存储器管理单元612可能接收来自多核系统中的多个异构设备中的一个或多个相应的请求设备的一个或多个PASID状态更新请求。当处理来自一个请求设备的PASID状态更新请求时,MMU 612可以执行与此PASID状态更新请求相对应的PASID状态的原子修改。这样的原子修改的一个实施例可以是原子的读-修改-写操作的形式,其中MMU 612可以读取该PASID状态条目的懒惰无效状态,并对相应的PASID状态执行懒惰无效校验,例如,以查明是否需要在由请求设备激活之前将用于进程地址空间的虚拟地址转换条目与系统页表同步。该MMU 612然后可将PASID状态更新响应发送到与该PASID状态更新请求相对应的请求设备,例如,告诉该请求设备,响应于懒惰无效状态被设置,在激活之前同步该设备的TLB条目。

[0067] 上面的描述旨在说明本发明的优选实施例。从上面的讨论也应明白,尤其在这种发展速度快而进一步发展又不容易预见的这类技术领域,可以由本领域技术人员对本发明的配置和细节进行修改,而不背离在所附权利要求及其等同物的范围内的本发明的原理。

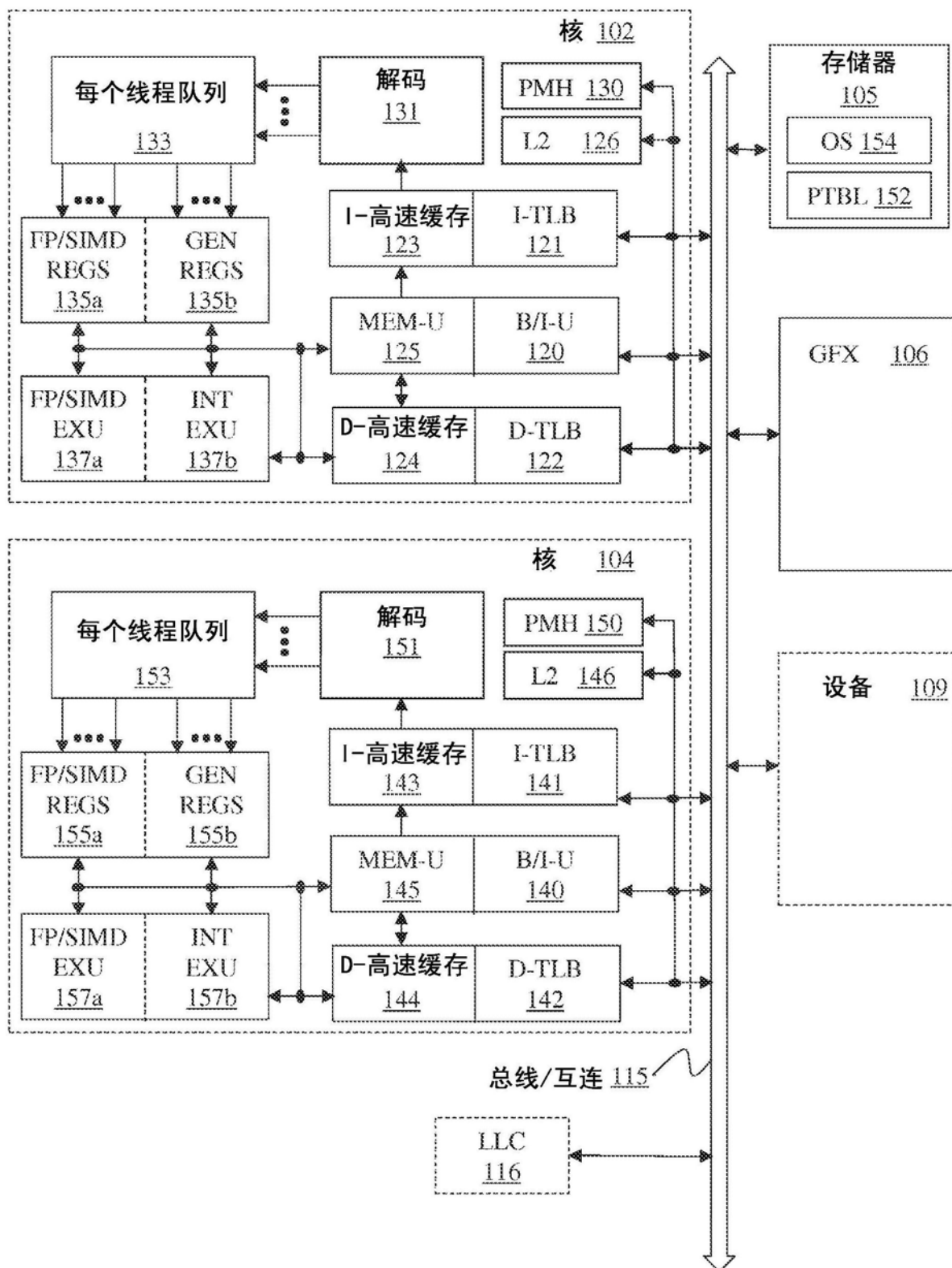


图1现有技术

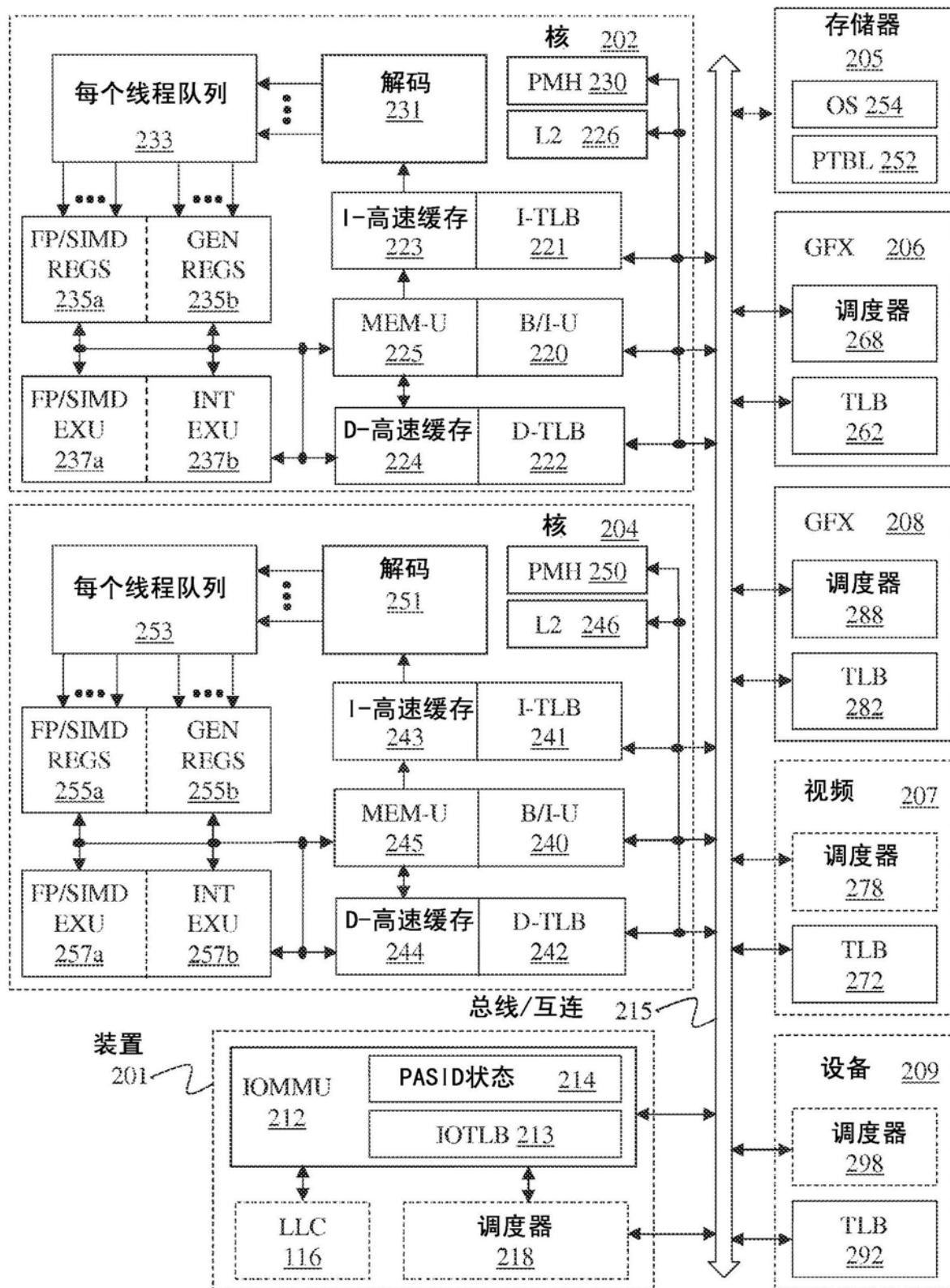


图2

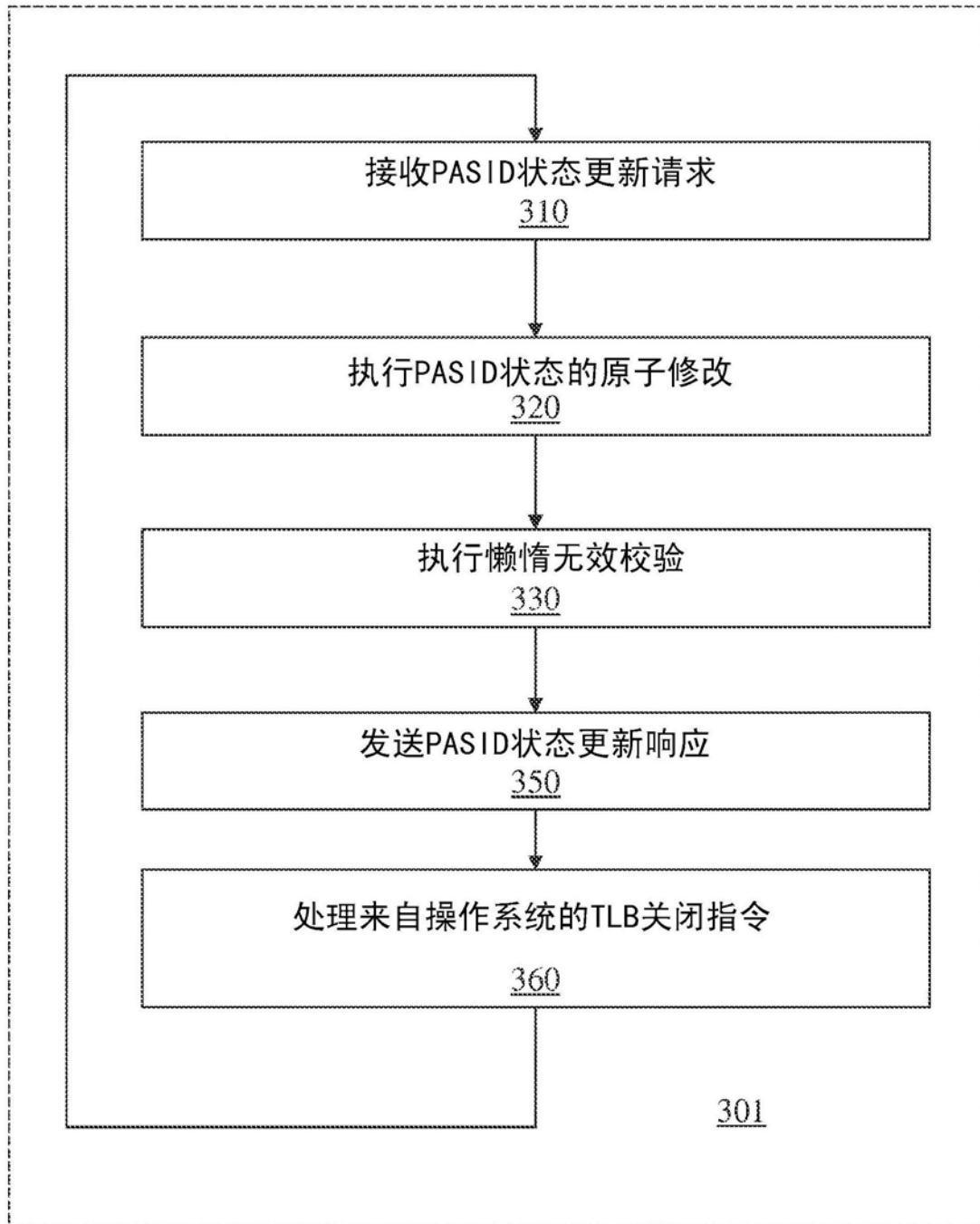


图3

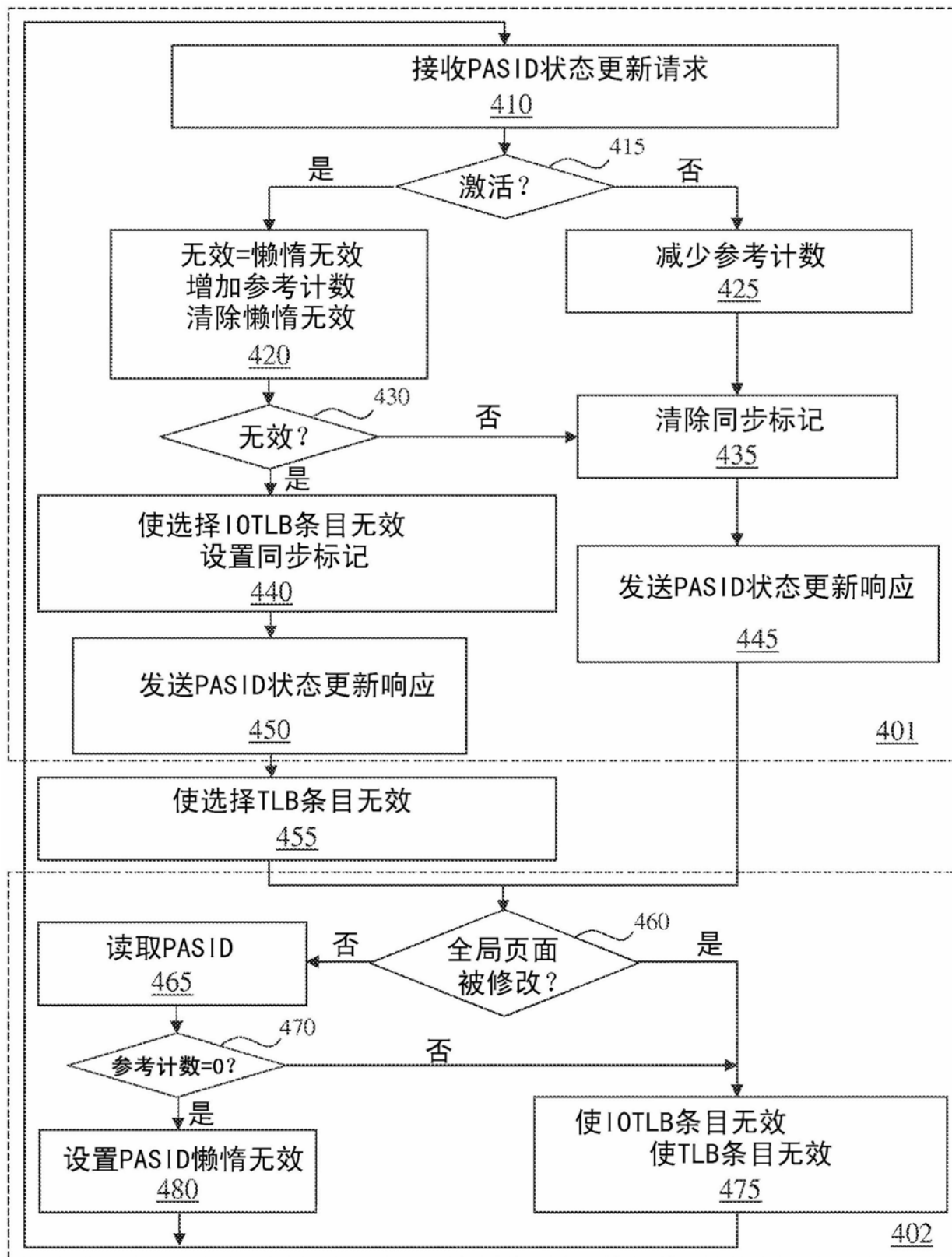


图4

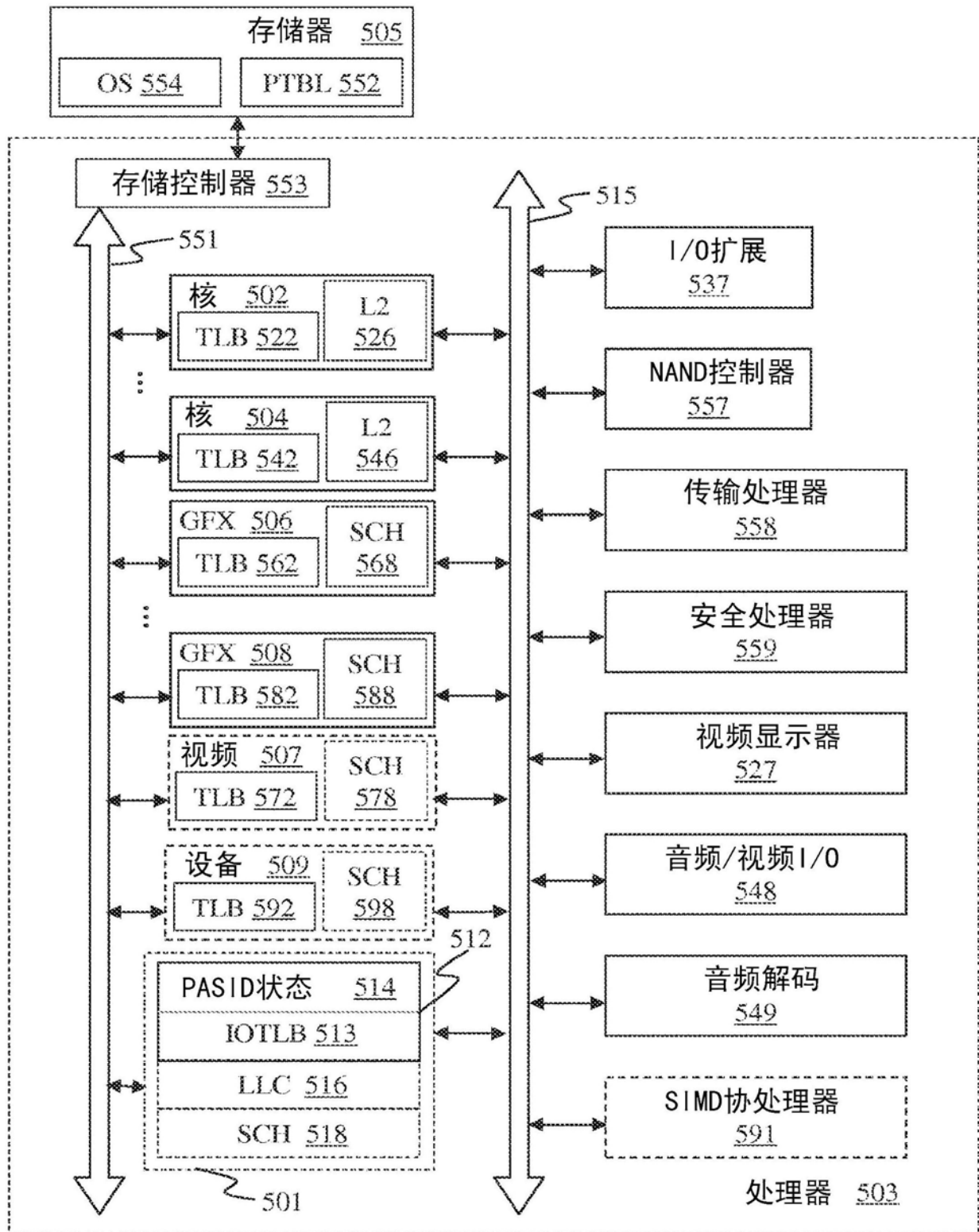


图5

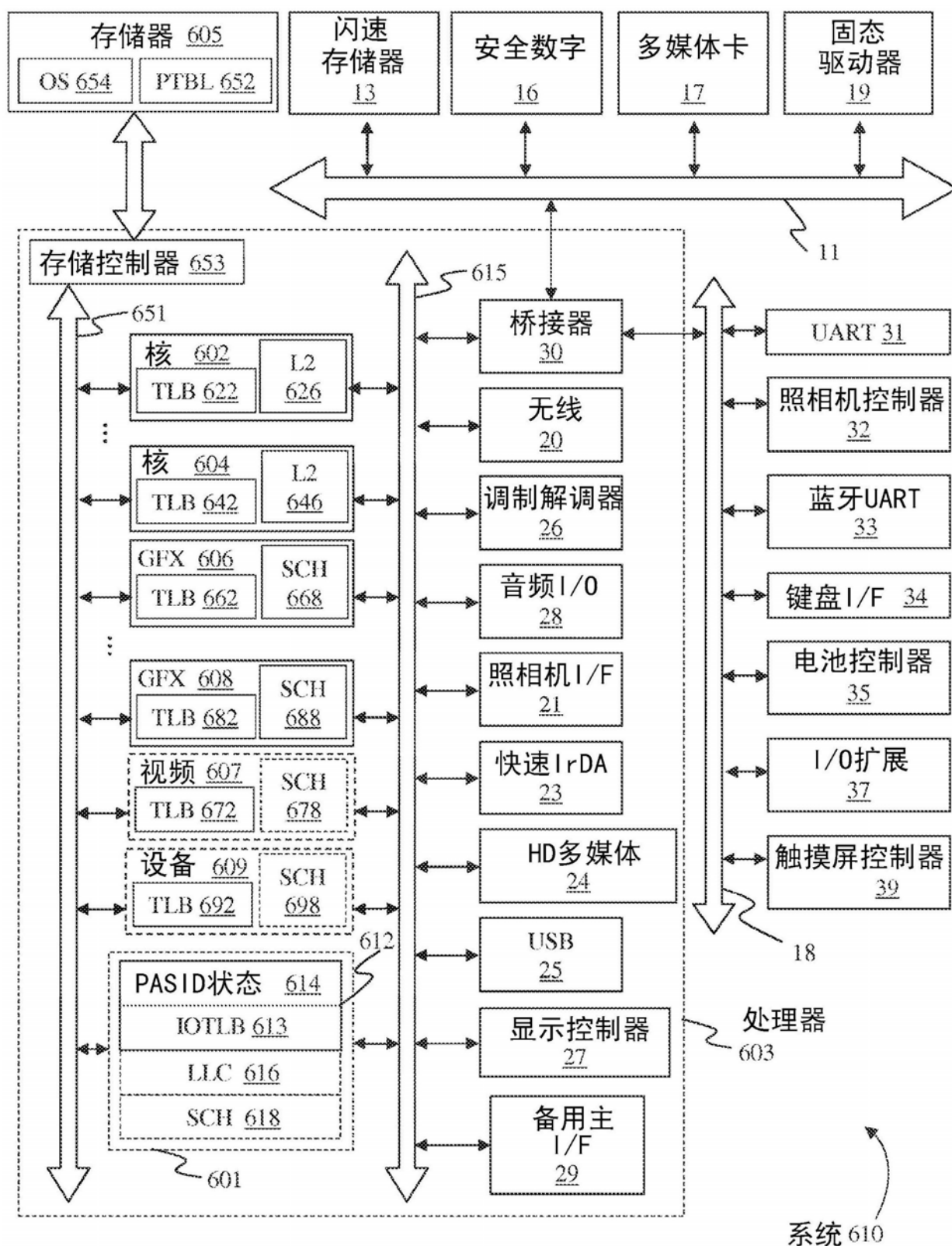


图6