



(12)发明专利申请

(10)申请公布号 CN 108628772 A

(43)申请公布日 2018.10.09

(21)申请号 201810255148.8

(22)申请日 2018.03.26

(30)优先权数据

62/476,178 2017.03.24 US

15/605,442 2017.05.25 US

(71)申请人 希捷科技有限公司

地址 美国加利福尼亚州

(72)发明人 杰弗里·蒙希尔 杰克逊·埃利斯

瑞安·J·戈斯

(74)专利代理机构 北京律盟知识产权代理有限

责任公司 11287

代理人 章蕾

(51)Int.Cl.

G06F 12/128(2016.01)

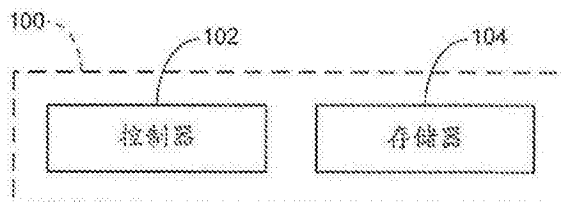
权利要求书4页 说明书9页 附图6页

(54)发明名称

用于管理数据存储设备中的映射数据的装置和方法

(57)摘要

本发明题为“用于管理数据存储设备中的映射数据的装置和方法”。可编程处理器发出查找命令以定位映射结构的所请求的映射页面并将所请求的映射页面放置到第一高速缓存中以服务接收到的主机命令。不可编程的硬件电路搜索转发表以确定所请求的映射页面是否位于第二高速缓存中,若是,将所述映射页面加载到所述第一高速缓存。否则,所述硬件电路从后端处理器请求所请求的映射页面,所述后端处理器从诸如闪存存储器阵列的非易失性存储器(NVM)中检索所请求的映射页面。所述硬件电路搜索倒排表和所述第一高速缓存以在所述第二高速缓存中为从所述NVM检索的所请求的映射页面选择候选位置,并引导所请求的映射页面的副本在所述候选位置处的存储。



1. 一种装置,包括:

映射结构,所述映射结构作为多个映射页面被存储在非易失性存储器(NVM)中,所述多个映射页面将用户数据块的逻辑地址与所述NVM中存储所述用户块的物理地址相关联;

第一高速缓存,所述第一高速缓存存储所述NVM中存储的所述映射页面的第一子组的副本;

第二高速缓存,所述第二高速缓存存储所述NVM中存储的所述映射页面的所述第一子组的副本以及所述映射页面的第二子组的副本;

可编程处理器,所述可编程处理器具有存储在相关联的存储器中的编程,所述编程在被执行时发出查找命令以定位与接收到的主机命令相关联的所请求的映射页面并将所请求的映射页面放置到所述第一高速缓存中;和

不可编程的硬件电路,所述不可编程的硬件电路被配置为响应于所述查找命令,从所述NVM引导所请求的映射页面的检索,访问存储器中的倒排表以识别存储在所述第二高速缓存中的选择的物理地址处的潜在替换映射页面,并且响应于由所述不可编程的硬件电路确定所述潜在替换映射页面不在所述第一高速缓存中的所述映射页面的所述第一子组中,引导所请求的映射页面的副本到所述第二高速缓存中的所述选择的物理地址的写入。

2. 根据权利要求1所述的装置,还包括所述第二高速缓存中的物理地址的先入先出(FIFO)列表,所述物理地址的先入先出(FIFO)列表作为存储器中的数据结构,所述不可编程的硬件电路被进一步配置为从所述FIFO列表中选择第一物理地址作为所述选择的物理地址,对所述倒排表编索引以将所述潜在替换映射页面识别为存储在所述第一物理地址处的第一映射页面,以及搜索所述第一高速缓存以确定所述第一映射页面不是存储在所述第一高速缓存中的映射页面的所述第一子组的成员。

3. 根据权利要求2所述的装置,其中响应于由所述不可编程的硬件电路确定所述潜在替换映射页面是存储在所述第一高速缓存中的映射页面的所述第一子组的成员,所述不可编程的硬件电路被进一步配置为从所述FIFO列表选择第二物理地址作为所述选择的物理地址,对所述倒排表编索引以将所述潜在替换映射页面识别为存储在所述第二物理地址处的第二映射页面,以及搜索所述第一高速缓存以确定所述第二映射页面不是存储在所述第一高速缓存中的映射页面的所述第一子组的成员。

4. 根据权利要求1所述的装置,其中所述不可编程的硬件电路通过执行所请求的映射页面的副本到所述第二高速缓存中的所述选择的物理地址的所述写入来引导所请求的映射页面的所述副本到所述第二高速缓存中的所述选择的物理地址的所述写入。

5. 根据权利要求1所述的装置,其中所述不可编程的硬件电路通过向所述可编程处理器提供所述选择的物理地址的通知,然后通过由所述可编程处理器将所请求的映射页面的所述副本写入所述选择的物理地址,来引导所请求的映射页面的副本到所述第二高速缓存中的所述选择的物理地址的所述写入。

6. 根据权利要求1所述的装置,其中所述倒排表是通过所述第二高速缓存中的物理地址与逻辑地址相关联的转换表,使得通过将所述第二高速缓存的所述选择的物理地址输入所述倒排表中,所述倒排表指示存储在所述选择的物理地址处的对应的映射页面。

7. 根据权利要求1所述的装置,其中所述不可编程的硬件电路被进一步配置为搜索存储器中被存储为数据结构的转发表,所述数据结构被布置为将所述第二高速缓存中的物理

地址关联到与所述映射页面相关联的逻辑地址的转换表,使得通过将选择的逻辑地址输入到所述转发表中,所述转发表指示所述第二高速缓存中存储与所述选择的逻辑地址相关联的所述映射页面的对应的所述选择的物理地址。

8. 根据权利要求1所述的装置,其中所述不可编程的硬件电路被进一步配置为响应于所述查找命令并响应于所述第一高速缓存上的高速缓存命中,搜索所述第一高速缓存并向所述可编程处理器通知存储所请求的映射页面的所述第一高速缓存的物理地址。

9. 根据权利要求1所述的装置,其中所述可编程处理器被进一步配置为结合所述查找命令向所述不可编程的硬件电路识别所述第一高速缓存中的第二潜在替换映射页面,其中所述不可编程的硬件电路被进一步配置为利用所请求的映射页面重写所述第一高速缓存中的所述第二潜在替换映射页面。

10. 根据权利要求1所述的装置,其中所述可编程处理器是为数据存储设备提供整体处理控制的核心可编程处理器,并且所述装置还包括具有编程的后端可编程处理器,所述编程存储在配置为引导编程和读取操作以在所述NVM与缓冲存储器之间传输数据的相关联的存储器中,其中所述不可编程的硬件电路被进一步配置为响应于所请求的映射页面在所述第二高速缓存上的高速缓存未命中,向所述后端可编程处理器发出读取请求以从所述NVM检索所请求的映射页面。

11. 根据权利要求1所述的装置,其中响应于指示所请求的映射页面位于所述第二高速缓存中的转发表,在所请求的映射页面位于所述第二高速缓存中时,所述不可编程的硬件电路被进一步配置为执行以下中的至少选择的一个:将所请求的映射页面从所述第二高速缓存复制到所述第一高速缓存,或者向所述可编程存储器提供指针,所述指针识别所请求的映射页面在所述第二高速缓存中的所述物理位置。

12. 一种数据存储设备,包括:

非易失性存储器(NVM),所述非易失性存储器存储被布置为映射页面的总数的映射结构,所述映射页面将用户数据块的逻辑地址与存储所述用户数据块的所述NVM的物理地址相关联;

第一高速缓存,所述第一高速缓存存储小于所述总数的第一数目的所述映射页面;

第二高速缓存,所述第二高速缓存存储小于所述总数的第二数目的映射页面,所述第二数目的映射页面包括所述第一数目的映射页面中的每一个映射页面;

转发表,所述转发表在存储器中存储为数据结构,所述转发表识别与存储在所述第二高速缓存中的所述第二数目的映射页面中的每一个映射页面相关联的逻辑地址;

倒排表,所述倒排表在存储器中存储为数据结构,所述倒排表识别存储所述第二数目的映射页面中的每一个映射页面的所述第二高速缓存的物理地址;

可编程处理器,所述可编程处理器具有存储在相关联的存储器中的编程,所述编程被配置为响应于从主机接收到主机访问命令而发出查找命令以查找所请求的映射页面;和

硬件电路,所述硬件电路被表征为基于非处理器的映射硬件辅助管理器电路,所述硬件电路被配置为:

响应于所述查找命令,搜索所述第一高速缓存并向所述可编程处理器通知存储所请求的映射页面的所述第一高速缓存的物理地址;

响应于所述第一高速缓存上的高速缓存未命中,搜索所述转发表,并响应于指示所请求

的映射页面位于所述第二高速缓存中的所述转发表,将所请求的映射页面从所述第二高速缓存复制到所述第一高速缓存;以及

响应于所述第二高速缓存上的高速缓存未中,搜索所述倒排表以识别存储在所述选择的物理地址处的第二映射页面,搜索所述第一高速缓存以确定所述第二映射页面未存储在所述第一高速缓存中,并向所述可编程处理器识别所述潜在替换映射页面作为从所述第二高速缓存逐出的候选项。

13. 根据权利要求12所述的数据存储设备,其中所述可编程处理器被进一步配置为结合所述查找命令向所述不可编程的硬件电路识别所述第一高速缓存中的潜在替换映射页面,其中所述硬件电路被进一步配置为利用所请求的映射页面重写所述第一高速缓存中的所述潜在替换映射页面。

14. 根据权利要求12所述的数据存储设备,其中所述可编程处理器是为数据存储设备提供整体处理控制的核心可编程处理器,并且所述数据存储设备还包括具有编程的后端可编程处理器,所述编程存储在被配置为引导编程和读取操作以在所述主存储器与缓冲存储器之间传输数据的相关联的存储器中,其中所述硬件电路被进一步配置为响应于所请求的映射页面在所述第二高速缓存上的高速缓存未中,访问所述映射结构以识别所请求的映射页面所位于的所述主存储器中的主存储器物理地址,以及结合对所述后端可编程处理器的从所述主存储器检索所请求的映射页面的读取请求,将所述主存储器物理地址提供给所述后端处理器。

15. 根据权利要求12所述的数据存储设备,还包括作为本地存储器中的数据结构的先入先出(FIFO)列表,所述本地存储器由所述硬件电路访问以将所述第二高速缓存中所述选择的物理地址识别为倒排表的输入,所述FIFO列表以最近最少使用(LRU)为基础布置所述第二高速缓存中的位置。

16. 根据权利要求12所述的数据存储设备,其中所述映射结构包括具有第一级映射和第二级映射的多级映射,所述第二级映射存储在所述主存储器中并且包括多个映射页面,所述多个映射页面包括所请求的映射页面,所述第一级映射包括识别所述多个映射页面存储在所述主存储器中的物理地址的条目,并且其中所述第二级映射的第一数目的所述映射页面存储在所述第一高速缓存中,所述第二级的更大的第二数目的所述映射页面存储在所述第二高速缓存中,并且存储在所述第一高速缓存中的每个映射页面也存储在所述第二高速缓存中。

17. 根据权利要求12所述的数据存储设备,所述数据存储设备表征为固态驱动器(SSD),其中所述主存储器包括闪存存储器阵列。

18. 一种方法,包括:

通过具有存储在相关联的存储器中的编程的可编程处理器发出查找命令以定位映射结构的所请求的映射页面并将所请求的映射页面放置到第一高速缓存中以服务接收到的主机命令,所请求的映射页面将用户数据块的逻辑地址与非易失性存储器(NVM)中存储所述用户数据块的物理地址相关联;

使用不可编程的硬件电路以:

响应于指示所请求的映射页面位于所述第二高速缓存中的所述转发表并响应于所述第一高速缓存上的高速缓存未中,搜索转发表并将所请求的映射页面从第二高速缓存复制

到所述第一高速缓存,所述转发表在存储器中存储为数据结构,所述转发表识别与存储在所述第二高速缓存中的所述映射结构的多个映射页面中的每一个映射页面相关联的逻辑地址;以及

识别所述第二高速缓存中的选择的物理地址,搜索倒排表以识别存储在所述选择的物理地址处的潜在替换映射页面,搜索所述第一高速缓存以确定所述潜在替换映射页面未存储在所述第一高速缓存中,以及引导所请求的映射页面到所述第二高速缓存中的所述选择的物理地址的重写。

19. 根据权利要求18所述的方法,还包括响应于所述查找命令并响应于所述第一高速缓存上的高速缓存命中,使用所述不可编程的硬件电路来搜索所述第一高速缓存并向所述可编程处理器通知存储所请求的映射页面的所述第一高速缓存的物理地址。

20. 根据权利要求18所述的方法,还包括响应于所请求的映射页面在所述第二高速缓存上的高速缓存未命中,操作所述不可编程的硬件电路来访问所述映射结构以识别所述NVM中所请求的映射页面所在的物理地址,并且将所述物理地址和读取请求提供给后端处理器以从所述NVM检索所请求的映射页面,所述后端处理器包括第二可编程处理器,所述第二可编程处理器具有存储在存储器中的用于编程和从所述NVM读取数据的相关联的编程。

用于管理数据存储设备中的映射数据的装置和方法

[0001] 相关专利申请

[0002] 本专利申请根据35U.S.C.119(e)要求于2017年3月24日提交的美国临时专利申请62/476,178的国内优先权,该专利申请的内容以引用方式并入本文。

技术领域

[0003] 本主题公开涉及用于管理数据存储设备中的映射数据的装置和方法。

背景技术

发明内容

[0004] 本发明的各种实施方案通常涉及数据存储设备中的数据管理。

[0005] 在一些实施方案中,装置具有作为多个映射页面存储在非易失性存储器(NVM)中的映射结构,所述多个映射页面将用户数据块的逻辑地址与NVM中存储用户块的物理地址相关联。第一高速缓存存储NVM中存储的映射页面的第一子组的副本,第二高速缓存存储NVM中存储的映射页面的第一子组的副本以及映射页面的第二子组的副本。

[0006] 可编程处理器具有存储在相关联的存储器中的编程,所述编程在被执行时发出查找命令以定位与接收到的主机命令相关联的所请求的映射页面并将所请求的映射页面放置到第一高速缓存中。不可编程的硬件电路被配置为响应于查找命令,从NVM引导所请求的映射页面的检索,访问存储器中的倒排表以识别存储在第二高速缓存中选择的物理地址处的潜在替换映射页面,以及响应于由不可编程的硬件电路确定该潜在替换映射页面不在第一高速缓存中的映射页面的第一子组中,引导所请求的映射页面的副本到第二高速缓存中的选择的物理地址的写入。

[0007] 根据以下详细论述和附图,可以理解表征本公开的各种实施方案的这些和其他特征和优点。

附图说明

[0008] 图1提供了根据本公开的各种实施方案的示例性数据存储设备的功能块表示。

[0009] 图2示出了根据一些实施方案的被配置为固态驱动器(SSD)的图1的数据存储设备的各个方面。

[0010] 图3是图2的示例性SSD的核心控制器的功能块表示。

[0011] 图4示出了一些实施方案中的多级映射结构的示例性格式。

[0012] 图5示出了映射结构的第二级映射(SLM)的示例性格式。

[0013] 图6示出了映射结构的第一级映射(FLM)的示例性格式。

[0014] 图7图示了一些实施方案中的SSD的各种存储器和它们的内容。

[0015] 图8是一些实施方案中的图3的核心控制器的核心中央处理单元(CPU)可编程处理器与基于非处理器的卸载映射硬件辅助(MHA)管理器之间的通信的功能块表示。

[0016] 图9A和图9B提供了用于说明根据本公开的各种实施方案执行的步骤的数据处理例程的流程图。

具体实施方式

[0017] 本公开整体涉及数据存储,并且更具体地涉及用于使用硬件电路管理数据存储设备中的映射数据以从可编程处理器卸载高速缓存管理任务的方法和装置。

[0018] 数据存储设备用于以快速和有效的方式存储并检索用户数据。映射结构通常用于跟踪存储在存储设备的主存储器中的用户数据的物理位置,以使设备能够定位并检索先前存储的数据。此类映射结构可将主机接收到的数据块的逻辑地址与该介质的物理地址以及与数据相关联的其他状态信息相关联。

[0019] 映射结构的管理可以为存储设备控制器在服务来自主机设备的访问命令(例如,读取命令、写入命令、状态命令等)时提供显著的处理瓶颈。闪存存储器和其他形式的可擦除介质尤其如此。在诸如SSD的基于闪存存储器的存储设备中,给定数据块的每个后续所接收的版本将被存储到闪存存储器内的不同位置。

[0020] 本公开的各种实施方案通常涉及数据存储设备中的数据管理。如下文所解释的那样,在一些实施方案中,数据存储设备设置有控制器电路和被表征为非易失性存储器或NVM的主存储器。NVM可采用闪存存储器阵列的形式。

[0021] 控制器电路提供了最高级的控制器功能以引导用户数据块在NVM与主机设备之间的传输。存储在NVM中的用户数据块由具有多个映射页面的映射结构描述,所述多个映射页面描述了由主机设备使用的逻辑地址与NVM的物理地址之间的关系。映射结构可以是多级映射结构或者可采用其他形式。

[0022] 控制器电路包括可编程处理器,其使用存储在存储器位置中的编程(例如,固件)来处理主机访问命令。在服务命令之前,与每个接收到的访问命令相关联的数据的一个或多个映射页面需要被加载到多高速缓存结构的第一高速缓存。

[0023] 多高速缓存结构包括第一高速缓存和至少第二高速缓存。还可以使用附加的高速缓存级别。第一高速缓存中保留了映射页面的较小(第一)子组。第二高速缓存存储映射页面的较大子组,使得第一高速缓存中的每个映射页面也保留在第二高速缓存中。换言之,第一高速缓存存储映射页面的第一子组,第二高速缓存存储映射页面的第一子组和第二子组。所有的映射页面(例如,整个映射结构)也都保留在NVM中。

[0024] 可编程处理器将某些操作卸载到基于非处理器的硬件电路,以定位映射页面并将其加载到第一高速缓存。响应于来自处理器的查找命令,硬件电路首先执行搜索(例如,访问操作)以确定所请求的(第一)映射页面是否已经驻留在第一高速缓存中。如果是这样,则硬件电路向处理器通知第一高速缓存内的相关联的地址(高速缓存行)。

[0025] 在第一高速缓存上的高速缓存未中的情况下(例如,所请求的映射页面未驻留在第一高速缓存中),硬件电路在存储器中搜索转发表,该转发表识别与存储在第二高速缓存中的每个映射页面相关联的逻辑地址。如果转发表指示所请求的映射页面存在于第二高速缓存中,则硬件电路将所请求的映射页面从第二高速缓存复制到第一高速缓存并通知处理器。由于第一高速缓存通常已满,因此可能需要从第一高速缓存中逐出至少一个现有的映射页面,从而为所请求的映射页面腾出空间。如有必要,处理器可识别第一高速缓存中的牺

牲(替换)映射页面以及将由所请求的映射页面替换的查找命令。

[0026] 在第二高速缓存上的高速缓存未中的情况下(例如,所请求的映射页面不位于第二高速缓存中),硬件电路引导两个并发操作:从NVM(例如,闪存存储器)检索所请求的映射页面,以及在第二高速缓存中选择将由所请求的映射页面替换的候选(第二)映射页面。

[0027] 为了从NVM检索所请求的映射页面,硬件电路访问映射结构的第一级映射以定位所请求的映射页面的闪存地址,并且向控制器的后端处理器发出请求以将所请求的映射页面从NVM检索到本地存储器。

[0028] 为了识别第二高速缓存中将由所请求的映射页面替换的潜在替换映射页面,硬件电路识别第二高速缓存中选择的物理地址,并在本地存储器中搜索倒排表。倒排表识别每个映射页面存储在第二高速缓存中的物理地址。因此,倒排表识别存储在选择的物理地址处的潜在替换映射页面。硬件电路搜索第一高速缓存以确定潜在替换映射页面是否也驻留在第一高速缓存中。如果是这样,则选择第二高速缓存中的新物理地址并重复该过程,直到将潜在替换映射页面识别为替换候选项。可以使用先进先出(FIFO)列表或类似机制来跟踪和选择候选物理地址以从第二高速缓存逐出其内容(例如,利用新页面在该位置重写)。潜在替换页面可从由可编程处理器管理的本地存储器中的先进先出(FIFO)列表提供。

[0029] FIFO机制可以采取多种形式。在一些情况下,可以使用从第一位置0绕到最后位置N-1,然后返回0的计数器。使用计数器值,使得系统替换位置0,然后是1,然后是2,依此类推直到N-1,之后计数再次返回0。确定位于第一高速缓存中的任何位置都将被跳过。如果给定的操作在第一计数值处结束(例如,位置3被更新),则下一个评估循环从序列中的下一个计数值处开始(例如,位置4)。

[0030] 上述操作是闭锁且不可分割的。硬件电路可将映射页面物理写入相应的第一高速缓存和第二高速缓存,或者可建立可由可编程处理器用于定位和使用所请求的映射页面的指针。这样,各种映射定位功能可以通过独立的硬件来执行,而无需占用可编程处理器,使得处理器能够执行更高优先级的任务并且加速存储设备的数据传输速率I/O性能。

[0031] 本公开的各种实施方案的这些和其他特征和优点可以从对图1的查看开始理解,图1提供了数据存储设备100的简化功能块表示。

[0032] 设备100包括控制器102和存储器模块104。控制器102为设备100提供了最高级别控制,并且可被配置为多个电路元件,这些电路元件包括在本地存储器中具有相关联的编程的可编程处理器以及一个或多个基于非处理器的硬件电路。

[0033] 存储器模块104可被布置为诸如可旋转记录介质(盘)和固态存储器阵列的一个或多个非易失性存储器元件。虽然在图1中示出了独立的控制器102,但这不是必须的,因为另选实施方案可将任何必要的控制器功能直接结合到存储器模块中。

[0034] 存储器模块104作为来自主机设备的用户数据的主存储器。主机设备可以是与存储设备100进行通信的任何设备。例如但不作为限制,存储设备可物理地结合到主机设备中,或者主机设备可使用任何合适的协议经由网络与主机设备进行通信。在一些实施方案中,存储设备100被配置为在诸如分布式对象存储系统、基于云的计算环境、RAID(独立冗余磁盘阵列)系统等大容量存储环境中形成多设备存储壳体的一部分。

[0035] 图2是根据一些实施方案的数据存储设备110的功能块表示。设备110通常对应于设备100并且被表征为使用二维(2D)或三维(3D)NAND闪存存储器作为主存储器存储的固态

驱动器 (SSD)。这仅仅是为了说明的目的,而不是进行限制。其他电路和部件可根据需要结合到SSD 110中,但是为了清楚起见,从图2中省略了这些电路和部件。图2的电路可被结合到诸如片上系统 (SOC) 设备的单个集成电路 (IC) 中,或者可涉及多个连接的IC设备。

[0036] 来自图1的控制器功能由主机接口 (I/F) 控制器电路112、核心控制器电路114和设备I/F控制器电路116执行。主机I/F控制器电路112有时可称为前端控制器或处理器,并且设备I/F控制器电路116可称为后端控制器或处理器。每个控制器112、114和116包括在合适的存储器位置中具有相关联的编程 (例如,固件、FW) 的独立可编程处理器以及各种硬件元件,以执行数据管理和传输功能。这仅仅是对一个实施方案的说明;在其他实施方案中,单个可编程处理器 (或少于三个可编程处理器) 可被配置为使用合适的存储器位置中的相关联的FW执行前端、核心和后端处理中的每一个。

[0037] 前端控制器112处理与主机设备的主机通信 (未单独示出)。后端控制器116利用闪存存储器118管理数据读取/写入/擦除 (R/W/E) 功能。闪存存储器118,有时也称为非易失性存储器 (NVM) 或主存储器,可由多个闪存芯片组成,以促进并行数据操作。核心控制器114,也称为主控制器或中央控制器,执行设备110的主要数据管理和控制。

[0038] 图3示出了一些实施方案中的来自图2的核心控制器114。可使用其他配置,因此这仅仅是说明性的而非限制性的。SRAM存储器120是专用作缓冲存储空间的易失性存储器,用于在利用闪存118进行数据传输操作期间暂时存储用户数据。DDR (DRAM/SDRAM) 存储器122是易失性存储器,其也可作为缓冲存储器,以及存储由系统使用的其他数据。相应的存储器120、122可被实现为单个集成电路 (IC),或者可被分布在多个物理存储器设备上,这些物理存储器设备在组合时提供总体可用的存储空间。

[0039] 核心处理器 (中央处理单元,CPU) 124是为核心控制器114提供主处理引擎的可编程处理器。映射硬件辅助 (MHA) 管理器电路126是基于非处理器的卸载硬件电路,其执行由核心处理器124引导的卸载功能,如下所述。

[0040] 本地存储器通常用128表示。该存储器被设想为包括一个或多个分立的本地存储器,其可用于存储由核心控制器使用的各种数据结构,包括由核心处理器126使用的固件 (FW) 编程130、系统映射132和各种映射表134。

[0041] 此时,区分术语“处理器”和诸如“基于非处理器”、“不可编程”和“硬件”等术语是有帮助的。如本文所用,术语处理器是指执行指令 (例如,FW) 以执行各种功能的CPU或类似的可编程设备。术语非处理器、基于非处理器、不可编程、硬件等由MHA管理器126例示,并且是指不利用存储在存储器中的编程,而是通过各种硬件电路元件 (逻辑门、FPGA等) 进行配置来操作的电路。MHA管理器126用作状态机或其他硬连线设备。管理器具有各种操作能力和功能,诸如直接内存存取 (DMA)、搜索、加载、比较等。

[0042] 映射132在图4中更全面地示出为具有第一级映射138和第二级映射140的多级映射。尽管SSD110采用两级映射,但是可以容易地使用其他映射结构,包括单级映射或具有多于两级的多级映射。

[0043] 图5示出了第二级映射 (SLM) 140的示例布置。SLM 140由多个第二级映射条目142 (SLME或条目) 组成,这些条目描述了驻留在闪存118中或可写入闪存118的用户数据的各个块。在本示例中,块 (也称为映射单元 (MU)) 的长度被设置为4KB (4096字节),但也可以使用其他大小。即使某些逻辑地址尚未使用或当前未使用,SLM 140也描述了SSD 110可容纳的

块的逻辑地址的全部可能范围。

[0044] 每个条目142包括多个字段,包括逻辑块地址(LBA)字段144、物理块地址字段146、偏移字段148和状态字段150。可使用其他格式。LBA值从最小值到最大值是连续的(例如,从LBA 0到LBA N,其中N是由SSD的总数据容量确定的一些较大数目)。可以使用其他逻辑寻址方案,诸如键值、虚拟块地址等。虽然LBA值可形成条目的一部分,但是在其他实施方案中,LBA可替代地用作到相关联的数据结构中的索引以定位各种条目。

[0045] 在典型的闪存阵列中,数据块被布置为在特定擦除块中沿着闪存存储器单元的行写入的页面。PBA可以用阵列、芯片、垃圾收集单元(GCU)、擦除块、页面等表示。偏移值可以是沿着选择的存储器页面的位偏移。状态值可指示相关联的块的状态(例如,有效、无效、空等)。

[0046] 条目组142被布置成较大的数据集,本文称为映射页面152。在每个映射页面中提供了一些选定数量的条目(由变量A表示)。在这种情况下,每个映射页面144总共具有100个条目。条目的其他分组可以在每个页面中进行,包括2的幂数。

[0047] 第二级映射(SLM)140构成了系统中所有映射页面152的布置。可以设想,为了描述SSD的总存储容量,将需要一些大总数的映射页面B。每个映射页面具有相关联的映射ID值,该值可以是0到B的连续数字。尽管SLM将可能在各种芯片的不同组上写入(而不是在闪存中的中央位置),但SLM 140被存储在NVM(闪存118)中。

[0048] 图6示出了来自图4的第一级映射(FLM)138的布置。FLM 138使得SSD 110能够定位存储在闪存中的各种映射页面152。为此,FLM 138由第一级映射条目162(FLME或条目)组成。每个FLME 162具有映射页面ID字段164、PBA字段166、偏移字段168和状态字段170。上文讨论了映射ID。字段166中的PBA描述了相关联的映射页面的位置。偏移值像之前一样作为位偏移沿着特定页面或其他位置操作。状态值可与第二级映射中的相同,或者可根据需要与映射页面本身的状态相关。如前所述,尽管图6中的格式示出了形成第一级映射中的每个条目的一部分的映射ID。但是在其他实施方案中,映射ID可替代地用作到数据结构的索引以定位相关联的条目。

[0049] 第一级映射(FLM)138构成了从条目0到条目C的所有条目162的布置。在一些情况下,B将等于C,尽管这些值可能不同。访问FLM 138允许通过映射ID在闪存存储器118内搜索期望的映射页面的位置。从闪存中检索期望的映射页面将在该映射页面中提供第二级映射条目,然后可基于相关联的第二级条目中的PBA信息来识别并检索各个LBA。

[0050] 图7示出了一些实施方案中由SSD 110利用的各种存储器位置的布置。这些包括上文在图2至图3中所讨论的闪存存储器118和本地核心存储器128,以及第一级高速缓存(FLC)180和第二级高速缓存(SLC)182。

[0051] 第一级高速缓存180,也称为第一高速缓存和第1层高速缓存,被设想为独立的存储器位置,诸如核心控制器的板上存储器。如上文所讨论的那样,将用于服务未决主机访问命令的映射页面被加载到第一高速缓存。图7示出了驻留在第一高速缓存180中的总数为D的映射页面。可以设想,D将是相对小的数字,诸如D=128,尽管可以使用其他数字。第一高速缓存的大小是固定的。

[0052] 第二级高速缓存182,也称为第二高速缓存和第二层高速缓存,被设想为构成DDR存储器122(见图2)的至少一部分。可以使用其他存储器位置。第二高速缓存的大小可以是

可变的或固定的。第二高速缓存存储多达最大数量的映射页面E,其中E是显著大于D ($E > D$) 的某个数字。如上所述,第一高速缓存中的每个D映射页面也存储在第二高速缓存中。

[0053] 闪存118主要用于存储由映射结构132描述的用户数据块,但此类存储未在图7中表示。图7清楚地示出了第一级映射 (FLM) 138的一个或多个备份副本以及第二级映射 (SLM) 140的完整副本被存储在闪存存储器中。SLM 140的备份副本也可被存储在闪存中以实现冗余,但是在可以直接访问此类冗余副本之前需要重新配置FLM 138。如上所述,FLM 138指向闪存中SLM 140的映射页面的主副本的位置。

[0054] 本地核心存储器128包括FLM138的主动副本,该主动副本根据需求由HMA管理器 (硬件电路) 126访问,从而根据需求以下文所讨论的方式从闪存中检索映射页面。存储器128还存储来自图3的映射表134,该映射表在图7中被布置为转发表和倒排表。

[0055] 转发表,也称为第一表格,是识别与存储在第二高速缓存182中的每个映射页面相关联的逻辑地址的数据结构。倒排表,也称为第二表格,识别每个映射页面存储在第二高速缓存中的物理地址。

[0056] 转发表通常可被视为LBA到DDR的转换表。通过输入选择的LBA (或者与期望的逻辑地址相关联的其他输入值),可定位该条目在第二高速缓存 (在这种情况下为DDR存储器) 中的相关联的位置。倒排表通常可被视为DDR到LBA的转换表。通过在第二高速缓存 (DDR存储器) 内输入选择的物理地址,可定位相关联的LBA (或与期望的逻辑地址相关联的其他值)。

[0057] 图8是示出来自图3的核心CPU (处理器) 124与MHA管理器 (硬件电路) 126之间的交互作用的功能表示。可以设想,响应于接收到选择的主机访问命令 (诸如在传输到主机之前从闪存存储器118检索某些用户数据块 (例如,LBA) 到SRAM缓冲器122 (图2) 的读取命令),图8的操作被执行。

[0058] 为了服务主机访问命令,需要将所请求的用户数据块的一个或多个映射页面加载到第一高速缓存180。通过向硬件电路发出查找命令以查找选择的 (第一) 映射页面并确保所请求的映射页面位于第一高速缓存中,处理器124将该操作卸载到硬件电路126。

[0059] 在一些情况下,处理器124可识别第一高速缓存中的替换映射页面,该替换映射页面可根据硬件电路126的需要重写以便将所请求的映射页面放入第一高速缓存中。可以设想,处理器将保持对第一高速缓存中的映射页面的逐出的最终控制,因为在映射页面可从第一高速缓存安全地释放之前,映射页面可能是脏的并且需要清洁 (例如,基于用户数据位置的改变对映射结构进行更新等)。

[0060] 响应于查找命令,硬件电路126可访问包括第一级高速缓存 (FLC) 180、第二级高速缓存 (SLC) 182、第一级映射 (FLM) 138、闪存存储器118、转发表 (在图8中以184表示)、倒排表 (以186表示)、先进先出 (FIFO) 列表188以及后端处理器190的各种存储器位置、数据结构和处理器。

[0061] FIFO列表188可以是基于使用的第二高速缓存地址的列表,诸如最近最少使用 (LRU) 方案。后端处理器190包括后端控制器116中的可编程处理器,其使用相关联的编程 (FW) 来执行利用闪存的编程和读取操作。硬件电路126将直接读取对后端处理器190的请求,使得处理器可以执行必要的的数据调整 (例如,解码、解压缩、错误校正等) 以从闪存访问选择的映射页面,而不是直接访问闪存以获取所请求的映射页面。

[0062] 核心处理器124保持第一级高速缓存180的最高级控制,因为这些条目表示与正在

进行的和未决数据传输相关的最有价值的映射页面。由核心处理器124决定第一级高速缓存中的哪些映射页面保留在高速缓存中以及哪些映射页面可用于逐出。直到核心处理器完成与映射页面相关联的所有剩余任务后,才会发生此类逐出。如果核心处理器124将第一级高速缓存中的替换映射页面识别到硬件电路126,则在一些情况下,硬件电路可转到利用第一级高速缓存中新请求的页面重写替换页面,使得可编程核心处理器不需要执行此操作。

[0063] 根据核心处理器授权的程度,硬件电路126具有对第二级高速缓存182的内容的一定程度的控制。应当记得,第一级高速缓存180中的每个映射页面的副本也存储在第二级高速缓存182中,但反之亦然。将新映射页面提升到第一级高速缓存(来自第二级高速缓存或来自闪存)通常需要替换第一级高速缓存中现有的映射页面,因为出于效率原因,第一级高速缓存通常(或总是)已满。新映射页面到第一级高速缓存的重写由核心处理器控制;即使硬件电路是实际将新映射页面写入第一级高速缓存的实体,核心处理器也会通知硬件电路将新映射页面放置在第一级高速缓存中的位置。

[0064] 因此,从闪存中检索新映射页面需要将新映射页面写入第一级高速缓存和第二级高速缓存。涉及第一级高速缓存的副本被提供给由核心处理器识别的地址。涉及第二级高速缓存的副本被写入最终由硬件电路基于图8中的FIFO列表188的内容确定的地址。

[0065] FIFO列表188是由第二级高速缓存182中的候选位置(地址)的核心控制器124保持的数据结构的示例,所述位置可用作逐出的候选项。该列表可以以任意种合适的方式生成,诸如基于最近最少使用(LRU)。硬件电路126从FIFO列表中识别选择的地址,使用倒排表186来识别存储在选择的地址处的映射页面,搜索第一级高速缓存180以查看位于选择的地址处的映射页面是否驻留在第一级高速缓存中。如果是,则识别第二级高速缓存中的不同候选地址并重复该过程。如果不是,则硬件电路将选择的地址识别为用于存储新映射页面的位置,由此引导新映射页面被写入由硬件电路或核心控制器执行的位置。

[0066] 图9A和图9B提供了根据一些实施方案的说明由图8的布置执行的步骤的数据处理例程200。该例程仅仅是示例性的并且不是限制性的。因为可附加、省略、以不同的顺序或通过其他模块等来执行各种步骤。一般来讲,例程中的每个步骤都具有“FW”或“HW”标识符。FW是指通过执行固件来由可编程处理器执行的步骤,并且为了方便起见被描述为由固件执行。HW是指由硬件电路126执行的步骤,因此被描述为由硬件执行。

[0067] 例程在固件接收到主机访问命令时开始于步骤202。在步骤204,作为响应,向硬件发出第一命令以定位与该命令相关联的选择的映射页面(“所请求的映射页面”)并根据需要将其加载到第一级高速缓存180中。当前驻留在第一级高速缓存中的替换映射页面也可根据需要被识别为用于逐出/重写。

[0068] 在步骤206,硬件继续在第一级高速缓存(FLC或第一级高速缓存)中搜索所请求的映射页面。判定步骤208确定是否获得高速缓存命中(例如,所请求的映射页面实际上驻留在第一级高速缓存中)。如果是,则流程转到步骤210,其中硬件向固件通知第一级高速缓存中的物理地址。此后。在步骤212,固件继续使用映射页面来服务命令,并且流程在步骤214结束。

[0069] 如果在步骤208确定高速缓存未命中(例如,选择的映射页面未驻留在第一级高速缓存中),则在步骤216,硬件继续访问转发表184以确定选择的映射页面是否位于第二级高速缓存(SLC)中。判定步骤218确定在第二级高速缓存上是否获得高速缓存命中。如果是,则流程转到步骤220,其中硬件将选择的映射页面复制到第一级高速缓存(FLC),重写由固件识别的

替换映射页面。此后执行步骤210和212,如前所述。

[0070] 如果在第二高速缓存上确定高速缓存未中(例如,选择的映射页面未驻留在第二高速缓存中),则流程遵循标记“A”到图9B,其中两个并行路径被硬件占用。

[0071] 第一路径开始于步骤222,其中硬件访问第一级映射 (FLM) 138以定位选择的映射页面的闪存地址(见图6)。在步骤224,硬件将闪存地址以及从闪存地址位置检索选择的映射页面的请求提供给后端处理器190。在步骤226,接收检索到的映射页面。

[0072] 第二路径开始于步骤228,其中硬件选择第二级高速缓存 (SLC) 182中的潜在替换位置(物理地址)以容纳从闪存检索的映射页面。如上所述,这可以使用FIFO列表188或类似的数据结构来执行。在步骤230,硬件搜索倒排表186以识别哪个映射页面位于替换位置,并且搜索第一级高速缓存 (FLC) 180以确定该映射页面是否位于第一高速缓存中。如果是,则在判定步骤234中,选择新的潜在替换位置并重复该过程,直到找到合适的候选替换位置。

[0073] 此后,在步骤236,两条路径汇合,硬件将从闪存存储器中检索到的选择的映射页面复制到潜在替换映射页面中的候选替换位置,并且例程遵循标记“B”回到图9A以执行步骤220、210和212。

[0074] 需注意,如果硬件被配置为在第二高速缓存中实际执行逐出,则第二高速缓存的内容的管理可被视为处于硬件的控制之下。另选地,第一高速缓存和第二高速缓存两者的管理可由处理器(固件)来保持,使得固件做出最终确定。在这两种情况下,硬件都会通知第二高速缓存中候选替换映射页面的固件。

[0075] 可以看出,由于后端处理器不需要涉及从闪存中检索映射页面,因此在第二高速缓存中保持部分映射可显著增强映射页面到第一高速缓存的加载。即使固件最终决定哪些映射页面停留在第二高速缓存中,通过硬件使用倒排表进行评估可以显著减少固件进行该确定时的工作量。

[0076] 上面的讨论已经将可编程处理器设想为将查找命令直接传递到硬件电路,但这不是必需的。其他实施方案考虑使用附加的不可编程硬件的命令处理,使得尽管在较高等级的固件的控制下,在该过程期间硬件也与硬件对话。

[0077] 如上所述,硬件电路可被配置为直接从闪存中检索所请求的映射页面,而不是向后端处理器(其可以是与核心处理器相同的处理器)发出请求。

[0078] 虽然用户数据已被视为查找命令的主题,但也可检索其他形式的数据,诸如描述系统状态的元数据,诸如但不限于关于系统中的用户数据的信息。

[0079] 硬件电路的一个方面是能够从第二高速缓存中选择适用于自动从第二高速缓存逐出的候选条目。倒排表可以采用任何合适的格式来识别此类候选项,并且可以包括链接列表或其他可以提供所需信息的简单结构。

[0080] 因此,对倒排表的引用不一定限于实际查找表,而是适用于如上所述通过第二高速缓存中的地址指示映射页面任何形式的元数据。在其他实施方案中,用于逐出的候选项可以由硬件电路来识别,并且硬件电路可以实际上自动地管理逐出,而不必通知可编程处理器等等。

[0081] 现在应当理解,本文所公开的各种实施方案可以提供许多有益效果。使用卸载硬件电路来管理多级高速缓存中的映射数据可显著加速映射加载功能并提高处理器效率。尽管各种实施方案已经设想了具有闪存存储器的固态驱动器(SSD)的环境,但是可以容易地

使用其他配置,包括不同形式的存储器,不同数量的高速缓存等。

[0082] 应当理解,尽管在前面的描述中已经阐述了本公开的各种实施方案的诸多特征和优点以及本公开的各种实施方案的结构和功能的细节,但是该具体实施方式仅仅是说明性的,并且可以在由所附权利要求所表达的术语的广泛一般含义所指示的全部范围内对细节进行修改,尤其是在本公开的原理内的部件的结构和布置方面的改变。

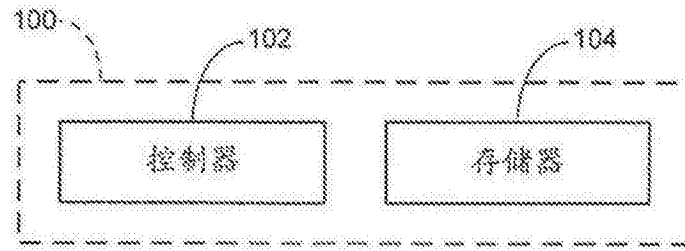


图1

固态硬盘 (SSD)

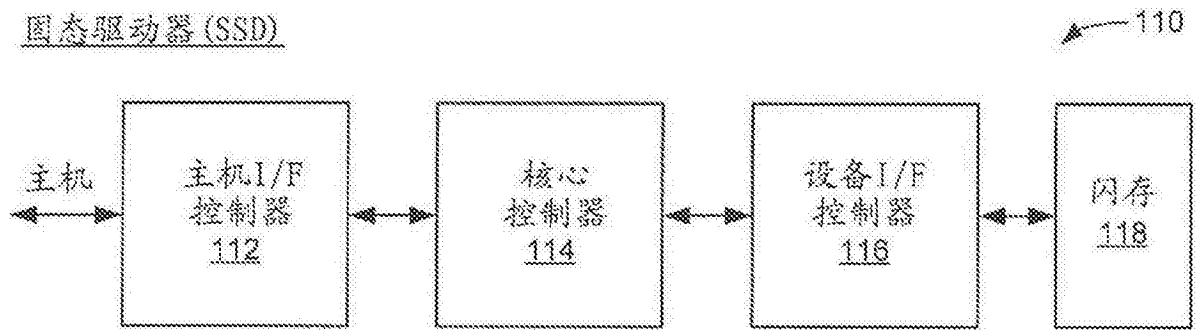


图2

核心控制器

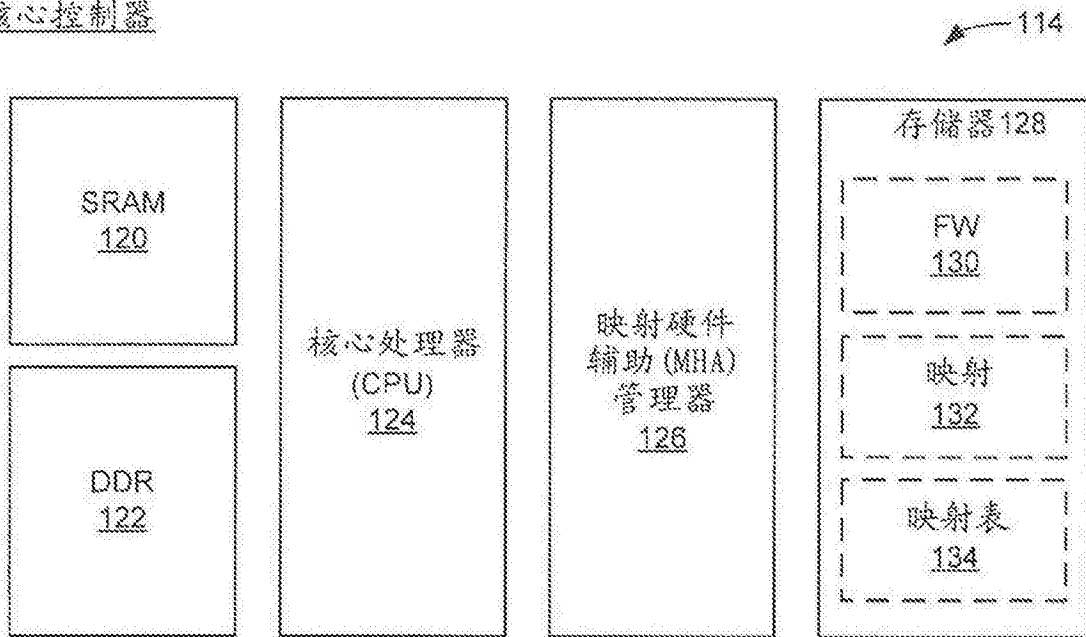


图3

多级映射

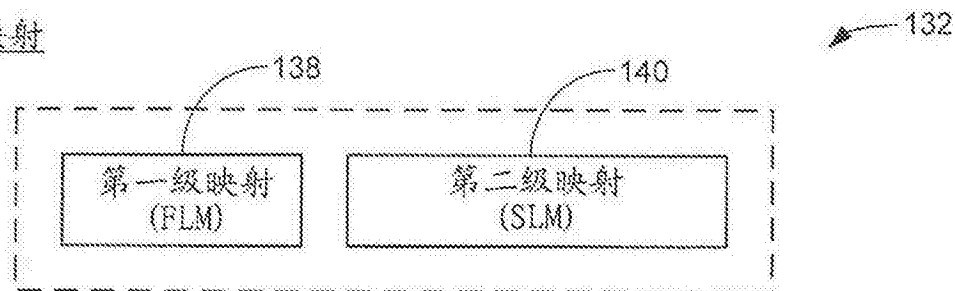


图4

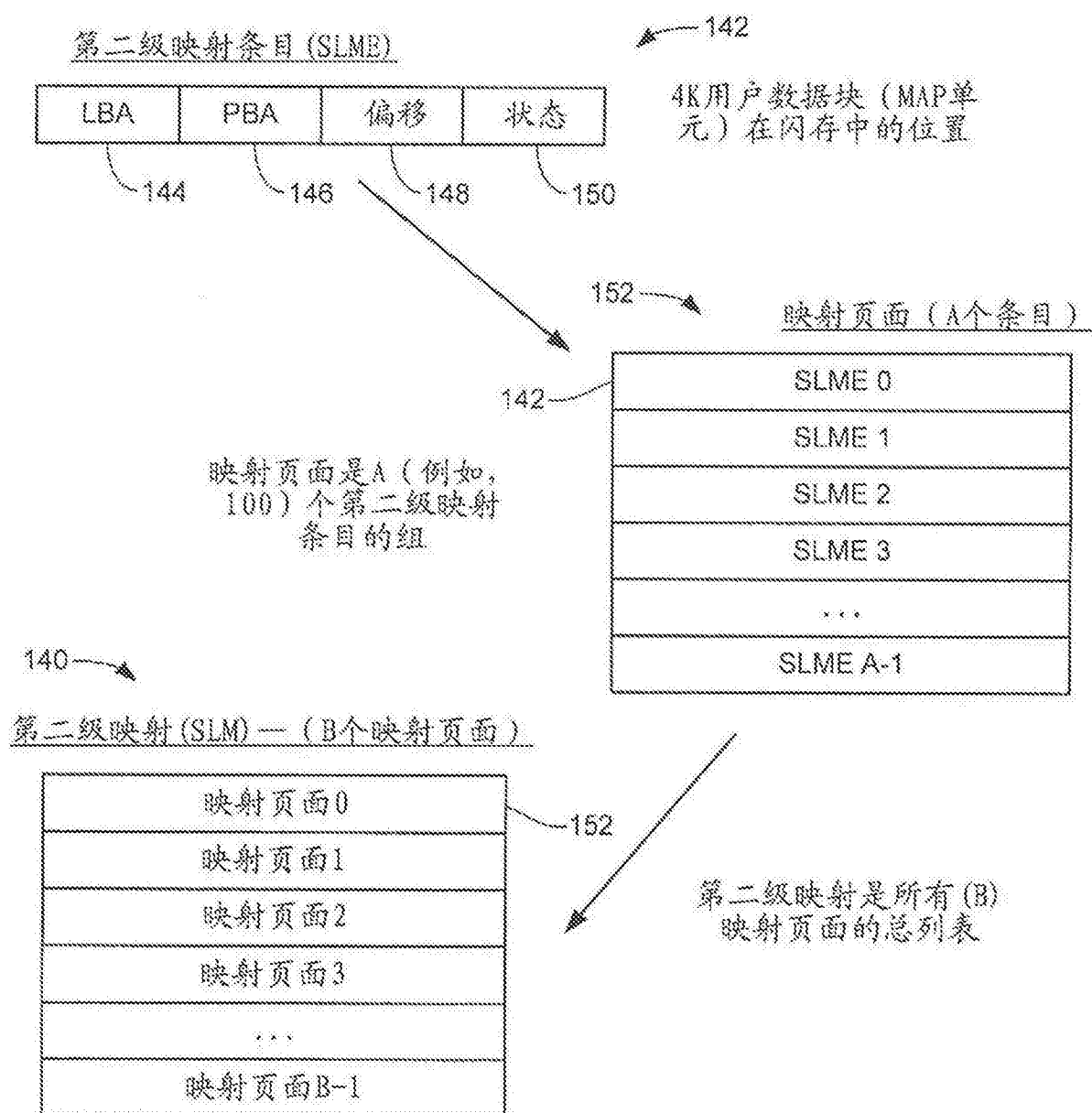


图5

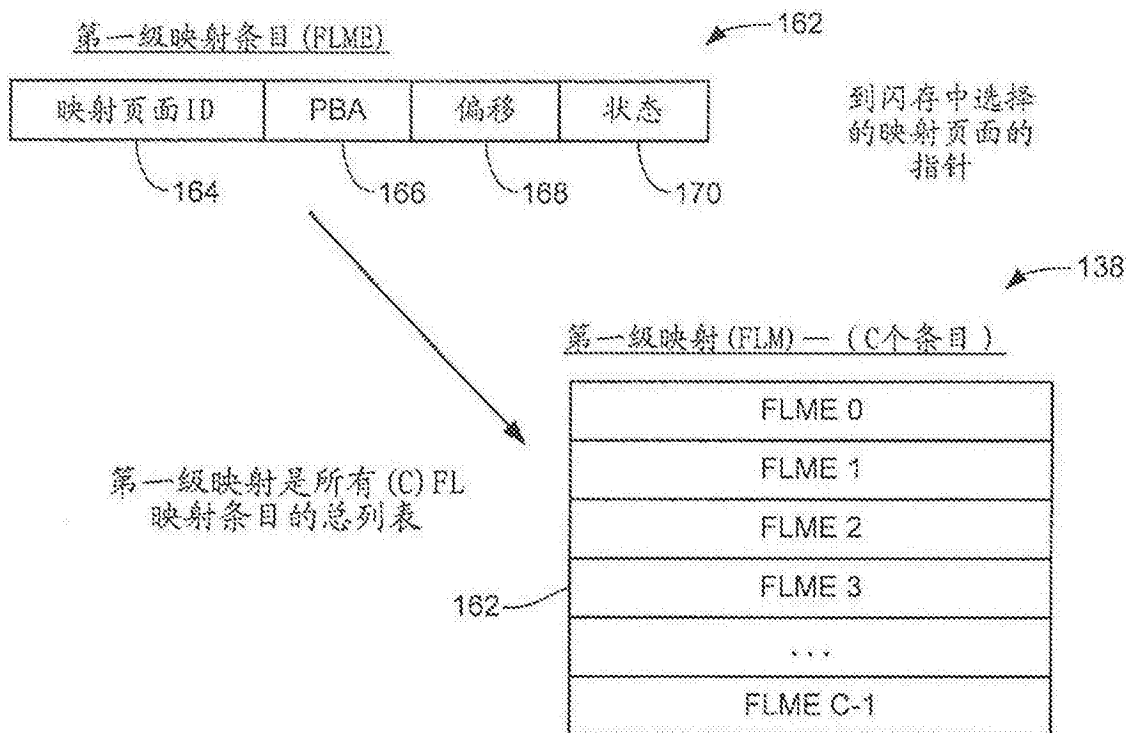


图6

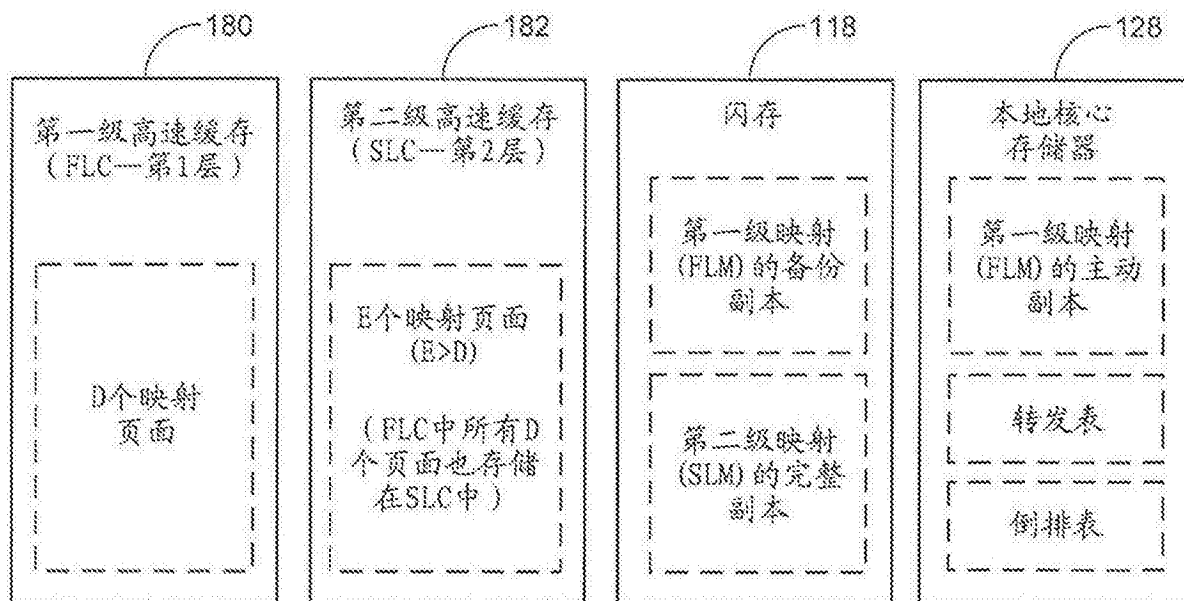


图7

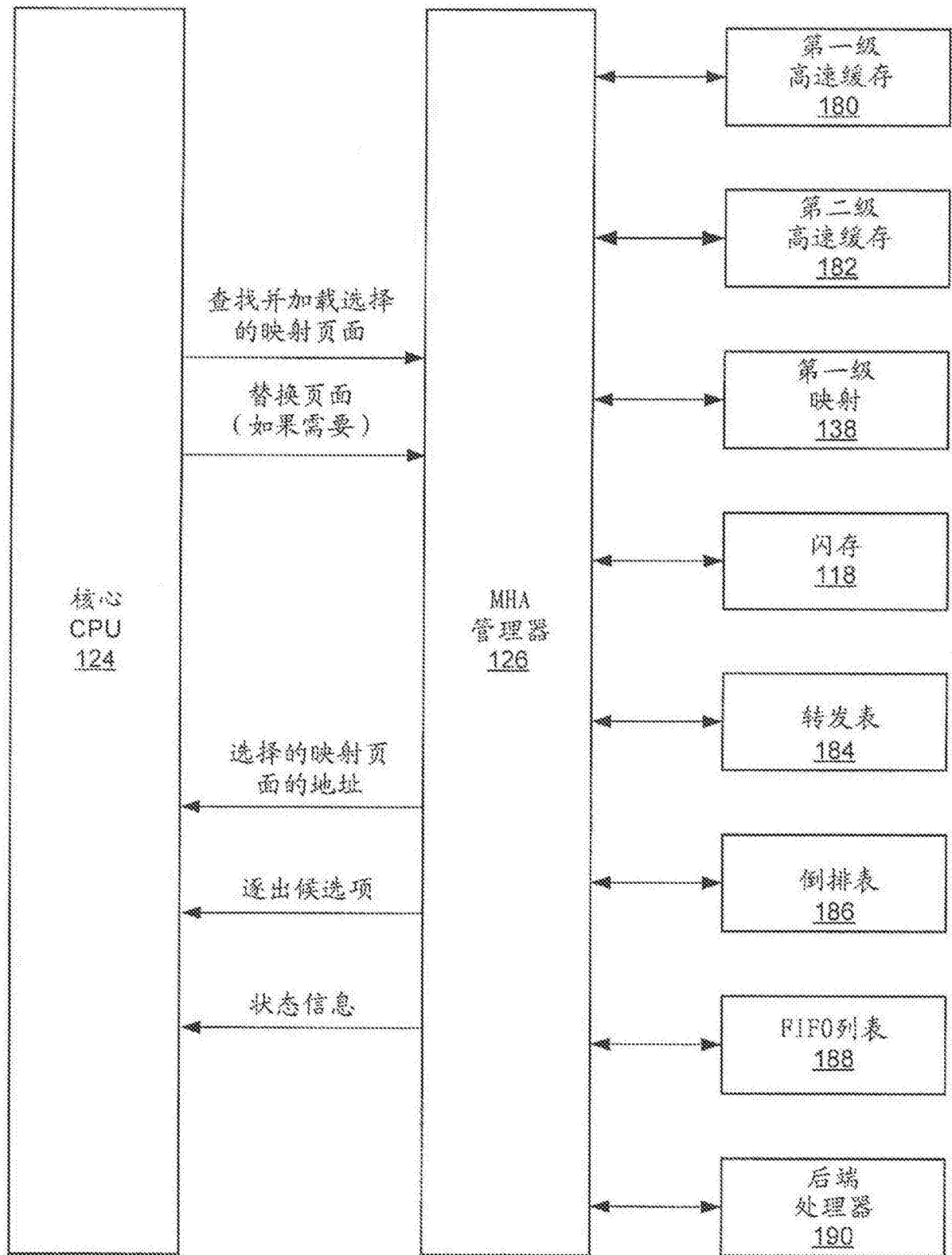


图8

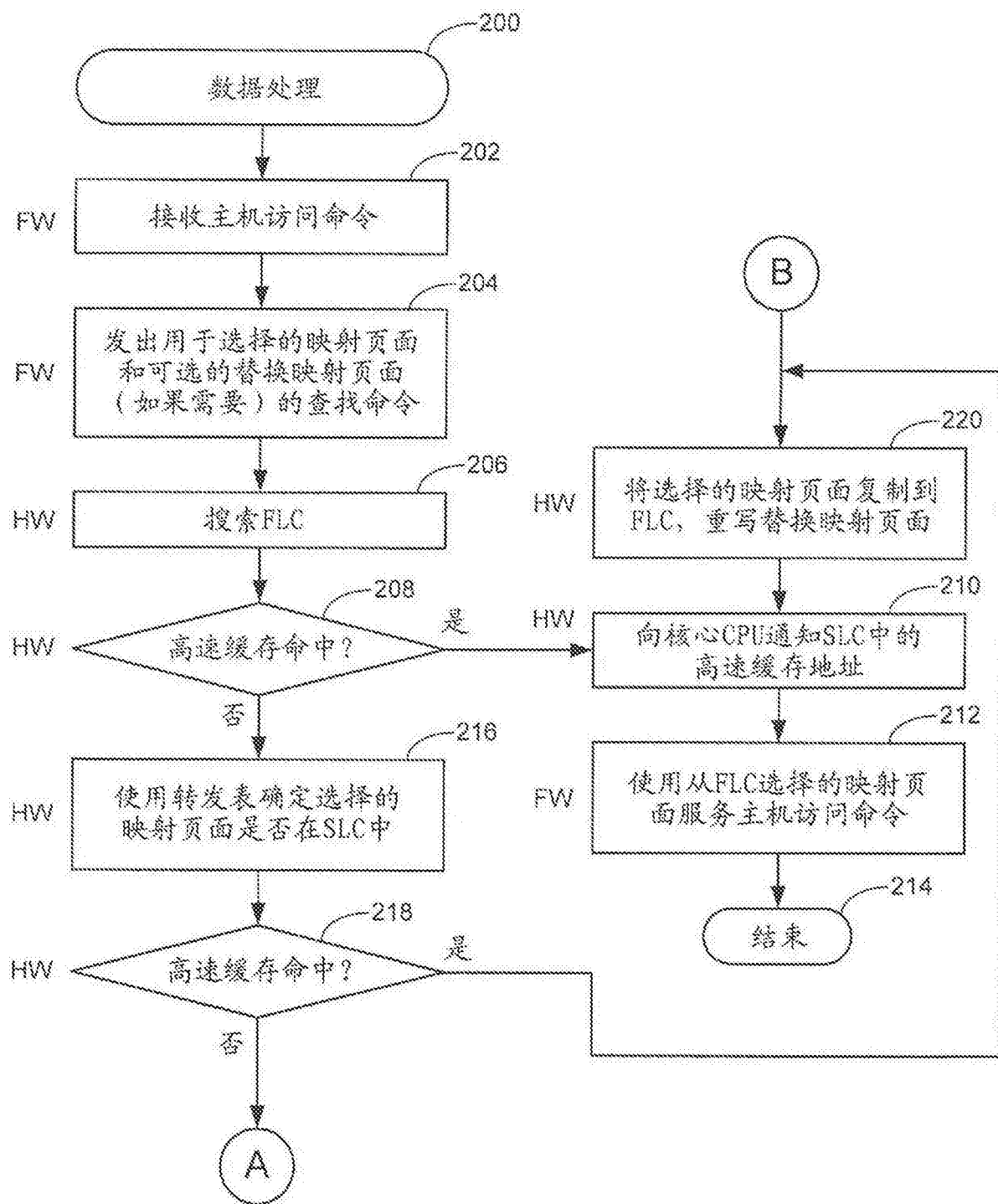


图9A

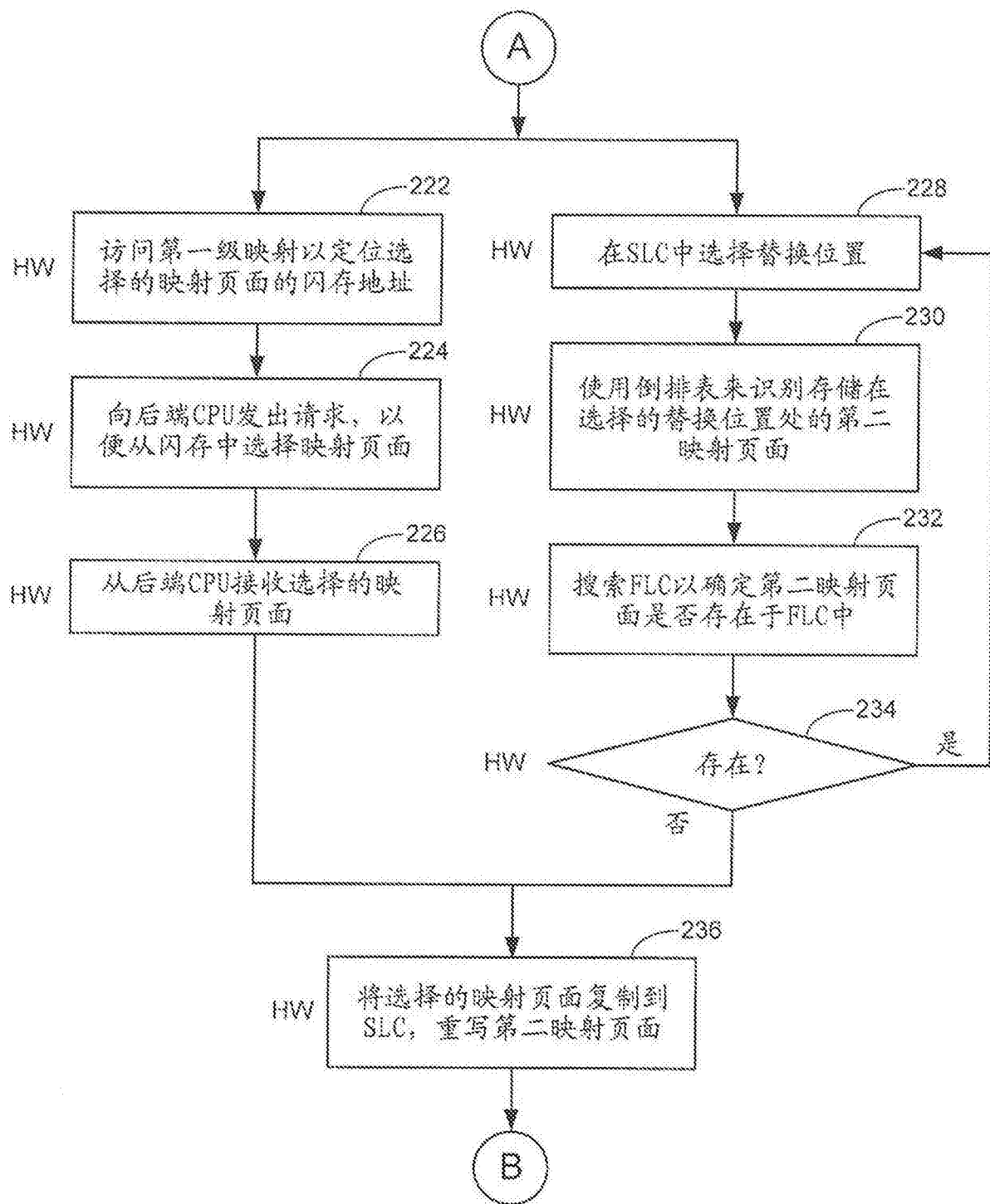


图9B