(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 104040508 B
(45)授权公告日 2017.03.29

(21)申请号 201280066402.0
(22)申请日 2012.12.10
(65)同一申请的已公布的文献号
申请公布号 CN 104040508 A
(43)申请公布日 2014.09.10
(30)优先权数据
13/352,230 2012.01.17 US
(85)PCT国际申请进入国家阶段日
2014.07.08
(86)PCT国际申请的申请数据
PCT/IB2012/057140 2012.12.10
(87)PCT国际申请的公布数据
WO2013/108097 EN 2013.07.25
(73)专利权人 国际商业机器公司
地址 美国纽约阿芒克
(72)发明人 L·M·古普塔 M·J·卡洛斯 M·T·本哈斯 K·A·尼尔森 K·J·阿什
(74)专利代理机构 北京市金杜律师事务所 11256
代理人 酆迅
(51)Int.Cl.
G06F 12/0868(2016.01)
G06F 12/128(2016.01)
(56)对比文件
TW 200410216 A,2004.06.16,
US 2003/0070042 A1,2003.04.10,
US 2011/0202732 A1,2011.08.18,
审查员 徐菲

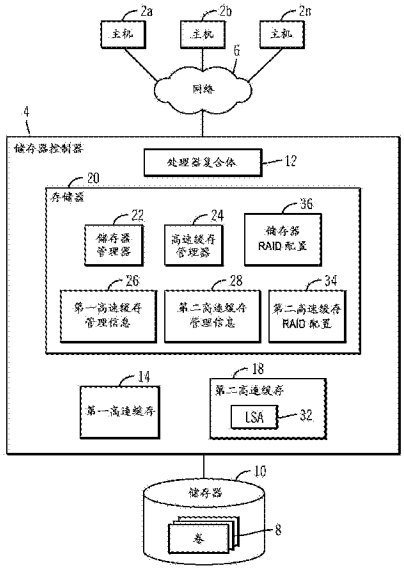
权利要求书3页 说明书10页 附图9页

(54)发明名称

用于在高速缓存系统中管理数据的方法和系统

(57)摘要

提供了用于在包括第一高速缓存、第二高速缓存以及存储系统的高速缓存系统中管理数据的计算机程序产品、系统以及方法。进行存储在存储系统中的轨道将从第一高速缓存降级的确定。形成包括要降级的确定轨道的第一步幅。进行其中将包括第一步幅中的轨道的第二高速缓存中的第二步幅的确定。将来自第一步幅的轨道添加到第二高速缓存中的第二步幅。进行将从第二高速缓存降级的第二高速缓存中的步幅中的轨道的确定。将被确定为从第二高速缓存降级的轨道降级。



CN 104040508 B

1.一种与存储系统通信的系统,包括:

处理器;

所述处理器可访问的第一高速缓存;

所述处理器可访问的第二高速缓存;

计算机可读存储介质,具有在其中体现的由所述处理器执行以进行操作的计算机可读程序代码,所述操作包括:

响应于第一高速缓存已满,将已修改最近最少使用(LRU)列表中指示的已修改非顺序轨道从第一高速缓存降级至存储系统;

将从第一高速缓存降级至存储系统的已修改非顺序轨道指示为第一高速缓存中的未修改非顺序轨道;

在第一未修改LRU列表中指示第一高速缓存中的所述未修改非顺序轨道;

确定在第一未修改LRU列表中指示的、要从所述第一高速缓存降级的存储在所述存储系统中的轨道;

根据第二高速缓存的配置,形成包括要降级的所确定轨道的第一步幅;

确定其中将包括所述第一步幅中的所述轨道的所述第二高速缓存中的第二步幅;

将来自所述第一步幅的所述轨道添加到所述第二高速缓存中的所述第二步幅;

在第二未修改LRU列表中指示被添加到所述第二高速缓存中的轨道;

从第二未修改LRU列表中确定要从所述第二高速缓存降级的所述第二高速缓存中的步幅中的轨道;以及

将被确定为要从所述第二高速缓存降级的轨道降级。

2.权利要求1的系统,其中所述第一高速缓存是比所述第二高速缓存更快的存取器件,并且其中所述第二高速缓存是比所述存储系统更快的存取器件。

3.权利要求1的系统,其中被确定为要从所述第二高速缓存降级的轨道来自所述第二高速缓存中的不同步幅。

4.权利要求1的系统,其中所述操作还包括:

接收对所述第一高速缓存中的轨道的写入;

确定在所述第二高速缓存中是否包括接收所述写入的所述轨道;以及

响应于确定所述第一高速缓存中的被写入的所述轨道包括在所述第二高速缓存中而使在所述第一高速缓存中更新的所述第二高速缓存中的所述轨道无效。

5.权利要求1的系统,其中所述操作还包括:

确定是否将所述第二高速缓存中的步幅合并;

响应于确定将步幅合并,执行:

确定不具有轨道的一个可用步幅;

确定具有有效轨道和无效轨道两者的至少两个步幅;

将来自至少两个步幅的所述有效轨道组合成所确定可用步幅,其中所述至少两个步幅可用于存储来自从所述第一高速缓存降级的步幅的轨道。

6.权利要求1的系统,其中形成轨道的所述第一步幅包括基于针对所述第二高速缓存被定义为具有n个器件的独立磁盘冗余阵列(RAID)配置而形成用于RAID配置的步幅,所述n个器件包括用于存储数据轨道的m个器件和存储根据用于所述m个器件的数据轨道计算的

奇偶数据的至少一个奇偶器件。

7.权利要求1的系统，其中所述操作还包括：

将第一未修改LRU列表用于所述第一高速缓存中的轨道，其中从所述第一未修改LRU列表确定要降级的轨道；以及

将第二未修改LRU列表用于所述第二高速缓存中的轨道以确定要从所述第二高速缓存降级的轨道。

8.权利要求7的系统，其中所述第一高速缓存存储包括已修改或未修改数据的轨道，并且其中要晋级到所述第二高速缓存的所述第一高速缓存中的分步幅形成的轨道包括未修改数据，其中所述操作还包括：

将来自所述第一高速缓存的已修改轨道降级到所述存储系统；以及

将已降级已修改轨道指示为未修改轨道，其中被指示为所述未修改轨道的已降级已修改轨道适合于被包括在被晋级至所述第二高速缓存的所述步幅中。

9.权利要求1的系统，其中所述第一高速缓存包括动态随机存取存储器(DRAM)，所述第二高速缓存包括多个闪速器件，并且所述存储系统包括比所述闪速器件缓慢的多个存取器件。

10.权利要求1的系统，其中所述操作还包括：

将被确定为要从所述第二高速缓存降级的轨道指示为无效，其中所述第二高速缓存中的步幅包括有效轨道和无效轨道。

11.权利要求5的系统，其中响应于在将从所述第一高速缓存降级的所述步幅中的所述轨道写入所述第二高速缓存中的一个可用步幅之后确定仅一个可用步幅具有所有空闲轨道而执行是否将步幅合并的确定。

12.权利要求6的系统，其中所述第一高速缓存包括动态随机存取存储器(DRAM)，并且其中所述第二高速缓存包括n个固态存储器件，其中跨所述n个固态存储器件使来自所述第一高速缓存的轨道的第一步幅条带化，以形成所述第二高速缓存中的所述第二步幅。

13.一种用于在高速缓存系统中管理数据的方法，其中，高速缓存系统包含第一高速缓存、第二高速缓存和存储系统，该方法包括：

响应于第一高速缓存已满，将已修改最近最少使用(LRU)列表中指示的已修改非顺序轨道从第一高速缓存降级至存储系统；

将从第一高速缓存降级至存储系统的已修改非顺序轨道指示为第一高速缓存中的未修改非顺序轨道；

在第一未修改LRU列表中指示第一高速缓存中的所述未修改非顺序轨道；

确定在第一未修改LRU列表中指示的、要从第一高速缓存降级的存储在存储系统中的轨道；

根据第二高速缓存的配置，形成包括要降级的所确定轨道的第一步幅；

确定其中将包括所述第一步幅中的所述轨道的第二高速缓存中的第二步幅；

将来自所述第一步幅的轨道添加到所述第二高速缓存中的所述第二步幅；

在第二未修改LRU列表中指示被添加到所述第二高速缓存中的轨道；

从第二未修改LRU列表中确定要从所述第二高速缓存降级的所述第二高速缓存中的步幅中的轨道；以及

将被确定为要从所述第二高速缓存降级的轨道降级。

14.权利要求13的方法，其中所述第一高速缓存是比所述第二高速缓存更快的存取器件，并且其中所述第二高速缓存是比所述存储系统更快的存取器件。

15.权利要求13的方法，其中被确定为要从所述第二高速缓存降级的轨道来自所述第二高速缓存中的不同步幅。

16.权利要求13的方法，还包括：

接收对所述第一高速缓存中的轨道的写入；

确定在所述第二高速缓存中是否包括接收所述写入的轨道；以及

响应于确定所述第一高速缓存中的被写入的所述轨道包括在所述第二高速缓存中而使在所述第一高速缓存中更新的所述第二高速缓存中的轨道无效。

17.权利要求13的方法，还包括：

确定是否将所述第二高速缓存中的步幅合并；

响应于确定将步幅合并，执行：

确定不具有轨道的一个可用步幅；

确定具有有效轨道和无效轨道两者的至少两个步幅；

将来自至少两个步幅的有效轨道组合成所确定可用步幅，其中所述至少两个步幅可用于存储来自从所述第一高速缓存降级的步幅的轨道。

18.权利要求13的方法，其中形成轨道的第一步幅包括基于针对所述第二高速缓存被定义为具有n个器件的独立磁盘冗余阵列(RAID)配置而形成用于RAID配置的步幅，所述n个器件包括用于存储数据轨道的m个器件和存储根据用于所述m个器件的数据轨道计算的奇偶数据的至少一个奇偶器件。

19.权利要求13的方法，还包括：

将第一未修改LRU列表用于所述第一高速缓存中的轨道，其中从所述第一未修改LRU列表确定要降级的轨道；以及

将第二未修改LRU列表用于所述第二高速缓存中的轨道以确定要从所述第二高速缓存降级的轨道。

20.权利要求19的方法，其中所述第一高速缓存存储包括已修改或未修改数据的轨道，并且其中要晋级到所述第二高速缓存的所述第一高速缓存中的分步幅形成的轨道包括未修改数据，还包括：

将来自所述第一高速缓存的已修改轨道降级到所述存储系统；以及

将已降级已修改轨道指示为未修改轨道，其中被指示为未修改轨道的已降级已修改轨道适合于被包括在被晋级至所述第二高速缓存的所述步幅中。

用于在高速缓存系统中管理数据的方法和系统

技术领域

[0001] 本发明涉及用于填充来自第一高速缓存的轨道的第一步幅(stride)以向第二高速缓存中的第二步幅写入的计算机程序产品、系统以及方法。

背景技术

[0002] 高速缓存管理系统将由于读和写操作而最近被存取的存储器件中的轨道缓存在诸如存储器之类的比存储所请求轨道的存储器件更快的存取存储器件中。对较快存取高速缓存存储器中的轨道的后续读请求被以与从较慢存取储存器返回所请求轨道相比更快的速率返回，因此减少读等待时间。高速缓存管理系统还可在指向存储器件的已修改轨道被写入高速缓存存储器时和已修改轨道被写出到存储器件、诸如硬盘驱动器之前向写请求返回完成。到存储器件的写等待时间通常明显长于向高速缓存存储器写入的等待时间。因此，使用高速缓存也减少写等待时间。

[0003] 高速缓存管理系统可保持链接列表，其具有用于存储在高速缓存中的每个轨道的一个条目，其可包括在向存储器件写入或读取数据之前缓存在高速缓存中的写数据。在一般使用的最近最少使用(LRU)高速缓存技术中，如果高速缓存中的轨道被访问，即高速缓存“命中”，则用于被存取轨道的LRU列表中的条目被移动至列表的最近最多使用(MRU)结尾。如果所请求轨道不在高速缓存中，即高速缓存未命中，则可去除高速缓存中的其条目在列表的LRU结尾处的轨道(或降级回到储存器)，并且向LRU列表的MRU结尾添加用于从储存器升级到高速缓存中的轨道数据的条目。用这种LRU高速缓存技术，被更频繁地存取的轨道很可能仍在高速缓存中，而较少频繁地被存取的数据将很可能被从列表的LRU结尾去除以在高速缓存中为新存取的轨道让出空间。

[0004] 在本领域中需要用于在存储系统中使用高速缓存的改进技术。

发明内容

[0005] 提供了用于在包括第一高速缓存、第二高速缓存以及存储系统的高速缓存系统中管理数据的计算机程序产品、系统以及方法。进行存储在存储系统中的轨道将从第一高速缓存降级的确定。形成包括被确定为降级的轨道的第一步幅。进行其中将包括第一步幅中的轨道的第二高速缓存中的第二步幅的确定。将来自第一步幅的轨道添加到第二高速缓存中的第二步幅。进行将从第二高速缓存降级的第二高速缓存中的分步幅的轨道的确定。将被确定为从第二高速缓存降级的轨道降级。

附图说明

[0006] 图1图示出计算环境的实施例。

[0007] 图2图示出第一高速缓存管理信息的实施例。

[0008] 图3图示出第二高速缓存管理信息的实施例。

[0009] 图4图示出第一高速缓存控制块的实施例。

[0010] 图5图示出第二高速缓存控制块的实施例。

[0011] 图6图示出步幅信息的实施例。

[0012] 图7图示出第二高速缓存RAID配置的实施例。

[0013] 图8图示出储存器RAID配置的实施例。

[0014] 图9图示出用以将未修改非顺序轨道从第一高速缓存降级以晋级到第二高速缓存的操作实施例。

[0015] 图10图示出用以向第一高速缓存添加轨道的操作实施例。

[0016] 图11图示出用以将轨道从第一步幅添加到第二步幅的操作实施例。

[0017] 图12图示出用以释放第二高速缓存中的空间的操作实施例。

[0018] 图13图示出用以释放第二高速缓存中的步幅的操作实施例。

[0019] 图14图示出用以处理用于轨道的请求以返回至读请求的操作实施例。

具体实施方式

[0020] 所述实施例提供了用于分步幅地将轨道从第一高速缓存晋级、使得可将该轨道作为全步幅写而写入第二高速缓存中的各步幅以改善高速缓存晋级操作的效率的技术。此外，在正在将轨道作为步幅从第一高速缓存14晋级至第二高速缓存18的同时，根据诸如LRU算法之类的高速缓存降级算法而将轨道基于轨道从第二高速缓存18降级。此外，可将部分已满、即具有有效和无效轨道的第二高速缓存中的步幅组合成一个步幅以释放第二高速缓存中的步幅以从第一高速缓存接收其他轨道步幅，使得第二高速缓存保持可用于由来自第一高速缓存的轨道形成的步幅的空闲步幅。

[0021] 图1图示出计算环境的实施例。多个主机2a、2b…2n可通过网络6向存储控制器4提交输入/输出(I/O)请求以在储存器10中的卷8(例如，逻辑单元号、逻辑器件、逻辑子系统等)处存取数据。存储控制器4包括处理器复合体12，包括具有单个或多个核的一个或多个处理器、第一高速缓存14和第二高速缓存18。第一高速缓存14和第二高速缓存18高速缓存在主机2a、2b…2n与储存器10之间传输的数据。

[0022] 存储控制器4具有存储器20，其包括在第一高速缓存14以及第二高速缓存18中的用于管理在主机2a、2b…2n与储存器10之间传输的轨道传输的存储管理器22以及管理在主机2a、2b…2n与储存器10之间传输的数据的高速缓存管理器24。轨道可包括在储存器10中配置的任何数据单元，诸如轨道、逻辑块地址(LBA)等，其是较大的一组轨道的一部分，诸如卷、逻辑器件等。高速缓存管理器24保持第一高速缓存管理信息26和第二高速缓存管理信息28以管理第一高速缓存14和第二高速缓存18中的读(未修改)和写(已修改)轨道。

[0023] 存储管理器22和高速缓存管理器24在图1中被示为被加载到存储器20中并由处理器复合体12执行的程序代码。备选地，可在存储控制器4中用硬件器件、诸如用专用集成电路(ASIC)来实现某些或所有功能。

[0024] 第二高速缓存18可在日志结构化阵列(LSA)32中存储轨道，其中，按照接收到的连续顺序对轨道进行写入，因此提供被写入第二高速缓存18的轨道的时间排序。在LSA中，在LSA 32的结尾处写入已存在于LSA中的轨道的较晚版本。在备选实施例中，第二高速缓存18可以用除LSA之外的格式来存储数据。

[0025] 存储器20还包括第二高速缓存RAID配置信息34，其提供关于用来确定如何形成要

存储在第二高速缓存18中的轨道步幅的RAID配置的信息。在一个实施例中，第二高速缓存18可包括多个存储器件，诸如单独固态存储器件(SSD)，使得跨形成第二高速缓存18的单独存储器件、诸如闪速存储器而使由来自第一高速缓存14的轨道形成的步幅形条带化。在另一实施例中，第二高速缓存18可包括单个存储器件，诸如一个闪速存储器，使得轨道被按照第二高速缓存RAID配置34的定义分步幅地分成组，但是轨道被作为步幅写入单个器件，诸如一个闪速存储器，实现第二高速缓存18。可将针对第二高速缓存RAID配置34而配置的步幅的轨道写入第二高速缓存18器件中的LSA 32。第二高速缓存RAID配置34可指定不同的RAID水平，例如5、10级等。

[0026] 存储器20还包括存储RAID配置信息36，其提供关于用来确定如何从第一高速缓存14或第二高速缓存18(如果第二高速缓存18应存储已修改数据)向存储系统10写入轨道的RAID配置的信息，其中，跨存储系统10中的存储器件、诸如磁盘驱动器而使降级步幅中的轨道形条带化。

[0027] 在一个实施例中，第一高速缓存14可包括随机存取存储器(RAM)，诸如动态随机存取存储器(DRAM)，并且第二高速缓存18可包括闪速存储器，诸如固态器件，并且储存器10包括一个或多个顺序存取存储器件，诸如硬盘驱动器和磁带。储存器10可包括单个顺序存取存储器件，或者可包括存储器件阵列，诸如磁盘簇(JBOD)、直接存取存储器(DASD)、独立磁盘冗余阵列(RAID)阵列、虚拟化设备等。在一个实施例中，第一高速缓存14是与第二高速缓存18相比的较快的存取器件，并且第二高速缓存18是与储存器10相比的较快的存储器件。此外，第一高速缓存14可具有比第二高速缓存18更大的每存储单位成本，并且第二高速缓存18可具有比储存器10中的存储器件更大的每存储单位成本。

[0028] 第一高速缓存14可以是存储器20的一部分或者在单个存储器件中实现，诸如DRAM。

[0029] 网络6可包括存储区域网(SAN)、局域网(LAN)、广域网(WAN)、因特网以及内部网等。

[0030] 图2图示出第一高速缓存管理信息26的实施例，其包括提供第一高速缓存14中的轨道的索引以控制控制块目录52中的块的轨道索引50；提供第一高速缓存14中的未修改顺序轨道的时间排序的未修改顺序LRU列表54；提供第一高速缓存14中的已修改顺序和非顺序轨道的时间排序的已修改LRU列表56；提供第一高速缓存14中的未修改非顺序轨道的时间排序的未修改非顺序LRU列表58；以及提供关于由第一高速缓存14中的未修改非顺序轨道形成的步幅的信息以作为全步幅写而写入第二高速缓存18的步幅信息60。

[0031] 在某些实施例中，在确定第一高速缓存18已满时，已修改LRU列表56用来将已修改轨道从第一高速缓存14降级至储存器10，使得拷贝第一高速缓存18中的那些降级的轨道。

[0032] 一旦已修改非顺序轨道被从第一高速缓存14降级至储存器10，则高速缓存管理器24可将该降级轨道指定为第一高速缓存14中的未修改非顺序轨道，并向未修改非顺序LRU列表58添加新指定的未修改轨道的指示，从中其有资格被晋级至第二高速缓存14。可通过更新第一高速缓存控制块104以在字段106中将已降级已修改非顺序轨道指定为未修改来改变已降级已修改轨道的状态。因此，第一高速缓存14中的未修改非顺序轨道可包括读数据或已修改非顺序轨道，其根据已修改LRU列表56而被降级至储存器10。因此，可将变成LRU列表58中的未修改轨道的已降级已修改轨道晋级至第二高速缓存14以可用于后续读请求。

在这些实施例中，第二高速缓存14包括只读高速缓存以高速缓存未修改非顺序轨道。

[0033] 图3图示出第二高速缓存管理信息28的实施例，其包括提供第二高速缓存18中的轨道的索引以控制控制块目录72中的块的磁带索引70；提供第二高速缓存18中的未修改轨道的时间排序的未修改列表74；以及提供关于被写入第二高速缓存18的轨道步幅的信息的步幅信息78。在一个实施例中，第二高速缓存18仅存储未修改、非顺序轨道。在其他实施例中，第二高速缓存18还可存储已修改和/或顺序轨道。

[0034] 所有LRU列表54、56、58和74可包括根据最后存取所识别轨道的时间而被排序的第一高速缓存14和第二高速缓存18中的轨道的轨道ID。LRU列表54、56、58和74具有指示最近被存取轨道的最近使用(MRU)结尾和指示最近最少使用或存取轨道的LRU结尾。被添加到高速缓存14和高速缓存18的轨道的轨道ID被添加到LRU列表的MRU结尾，并且从LRU结尾对从高速缓存14和高速缓存18降级的轨道进行存取。轨道索引50和轨道索引70可包括离散索引表(SIT)。备选类型数据结构可用来提供高速缓存14和高速缓存18中的轨道的时间排序。

[0035] 非顺序轨道可包括在线交易处理(OLTP)轨道，其常常包括并非完全随机且具有一定参考局部性、即具有被反复地存取的概率的小块。

[0036] 图4图示出控制块目录52中的第一高速缓存控制块100条目的实施例，包括控制块标识符(ID)102、第一高速缓存14中的轨道的物理位置的第一高速缓存位置104、指示轨道已修改或未修改的信息106、指示轨道是顺序还是非顺序存取的信息108以及指示用于轨道的降级状态，诸如未降级、即将降级以及降级完成的信息110。

[0037] 图5图示出第二高速缓存控制块目录72中的第二高速缓存控制块120条目的实施例，包括控制块标识符(ID)122；LSA位置124，其中，轨道位于LSA 32中；指示轨道已修改或未修改的已修改/未修改信息126；以及指示轨道有效或无效的有效/无效标志128。如果轨道在第一高速缓存14中被更新或者如果轨道被从第二高速缓存18降级，则第二高速缓存18中的轨道被指示为无效。

[0038] 一旦已修改非顺序轨道被从第一高速缓存14降级至储存器10，则高速缓存管理器24可将该降级轨道指定为第一高速缓存14中的未修改非顺序轨道，并向未修改非顺序LRU列表58添加新指定的未修改轨道的指示，从中其有资格被晋级至第二高速缓存14。可通过更新第一高速缓存控制块100以在字段106中将已降级已修改非顺序轨道指示为未修改来改变已降级已修改轨道的状态。因此，第一高速缓存14中的未修改非顺序轨道可包括读数据或已修改非顺序轨道，其根据已修改LRU列表56而被降级至储存器10。因此，可将变成LRU列表58中的未修改轨道的已降级已修改轨道晋级至第二高速缓存14以可用于后续读请求。在这些实施例中，第二高速缓存14包括只读高速缓存以高速缓存未修改非顺序轨道。

[0039] 图6图示出用于将在第二高速缓存18中形成的一个步幅的步幅信息60、78的实例130，包括步幅标识符(ID)132、包括在步幅132中的储存器10的轨道134以及指示轨道总数的步幅中的有效轨道的数目的占用136，其中，无效的步幅中的轨道适合于无用单元收集操作。

[0040] 图7图示出被保持以确定如何由第一高速缓存14中的轨道形成第二高速缓存18中的轨道的步幅的第二高速缓存RAID配置34的实施例。RAID水平140指示要使用的RAID配置，例如RAID 1、RAID 5、RAID 6、RAID 10等、存储用户数据的轨道的数据磁盘数目(m)142以及存储从数据磁盘142计算的奇偶性的奇偶磁盘数目(p)144，其中，p可以是一个或多个，指示

用于存储所计算奇偶块的磁盘数目。未修改奇偶可选标志148指示是否应针对被晋级至第二高速缓存18的第一高速缓存14中的未修改非顺序轨道计算奇偶性。此可选标志148允许仅包括步幅中的未修改非顺序轨道以用仅未修改非顺序轨道来填充步幅。可在LSA 32中指示第一高速缓存14中的未修改非顺序轨道的步幅，其中，跨形成第二高速缓存18的m加p个存储器件而使步幅的轨道形条带化。备选地，第二高速缓存18可包括少于n个器件。

[0041] 图8图示出被保持以确定如何形成第二高速缓存18中的已修改轨道的步幅以跨储存器10的磁盘形条带化的储存器RAID配置36的实施例。RAID水平150指示要使用的RAID配置、存储用户数据的轨道的数据磁盘数目(m)152、以及存储从数据磁盘152计算的奇偶性的奇偶磁盘数目(p)154，其中，p可以是一个或多个，指示用于存储所计算奇偶块的磁盘数目。可跨存储系统10中的磁盘使来自第二高速缓存18的轨道步幅条带化。

[0042] 在一个实施例中，第二高速缓存34和储存器36RAID配置可提供不同的参数或具有相同参数，诸如不同的RAID水平、数据磁盘、奇偶磁盘等。

[0043] 图9图示出由高速缓存管理器24执行以使未修改非顺序轨道从第一高速缓存14降级以晋级至第二高速缓存18的操作实施例，其中，可在需要空间时从未修改非顺序LRU列表58的LRU结尾选择未修改非顺序轨道。在发起(在方框200处)用以将所选未修改非顺序轨道降级的操作时，将被选择为降级的未修改非顺序轨道的降级状态110(图4)设置(在方框202处)为“就绪”。高速缓存管理器24使用(在方框204处)第二高速缓存RAID配置信息34而形成来自第一高速缓存114的轨道的第一步幅以晋级至第二高速缓存18中的步幅。例如，形成轨道的第一步幅可包括基于针对第二高速缓存被定义34为具有n个器件的RAID配置来形成用于RAID配置的步幅，所述n个器件包括用于存储数据轨道的m个器件和用以存储根据用于m个器件的数据轨道计算的奇偶数据的至少一个奇偶器件p。此外，在实施例中可在没有奇偶性的情况下跨n个固态存储器件使轨道的第一步幅条带化以形成第二步幅，其中，第二高速缓存包括至少n个固态存储器件。

[0044] 高速缓存管理器24处理(在方框206处)未修改非顺序LRU 58列表以确定在其控制块100中具有就绪的降级状态110的未修改非顺序轨道的数目。如果高速缓存管理器24确定(在方框208处)未修改非顺序轨道的数目足以形成步幅，则高速缓存管理器24填充(在方框210处)具有就绪的降级状态110的未修改非顺序轨道的第一步幅。在一个实施例中，可从未修改非顺序LRU列表58的LRU结尾开始填充第一步幅，并以步幅为单位对数据磁盘使用足够的轨道。如果(在方框212处)RAID配置指定奇偶磁盘，则高速缓存管理器24计算(在方框212处)用于包括在步幅中的未修改非顺序轨道的奇偶性并在步幅中包括奇偶数据(用于p个奇偶磁盘)。如果(在方框208处)在第一高速缓存14中不存在足以填充第一步幅的未修改非顺序轨道，则控制结束，直至存在具有可用于填充第一步幅的降级就绪状态的足够数目的未修改非顺序轨道为止。

[0045] 在填充第一步幅(在方框210和方框212)之后，高速缓存管理器14确定(在方框214处)其中将包括来自第一步幅的轨道的第二高速缓存18中的空闲第二步幅。来自第一步幅的轨道被作为全步幅写而写入或条带化(在方框216处)为跨形成第二高速缓存18的器件的第二步幅。在用来自第一步幅的轨道来填充第二高速缓存18中的第二步幅时，高速缓存管理器14将用于第二步幅的步幅信息130的占用136指定(在方框218处)为已满。高速缓存管理器24将用于包括在步幅中的未修改非顺序轨道的降级状态110更新(在方框220处)为降

级“完成”。

[0046] 虽然将图9的操作描述为将未修改非顺序轨道从第一高速缓存14降级以晋级至第二高速缓存18，但在备选实施例中，该操作可适用于将不同类型的轨道降级，诸如已修改、顺序等。

[0047] 用所述实施例，将来自第一高速缓存14的未修改轨道聚集并作为步幅写入第二高速缓存18，使得使用一个输入/输出(I/O)操作来转移多个轨道。

[0048] 图10图示出由高速缓存管理器24执行以向第一高速缓存14添加(即，晋级)轨道的操作实施例，该轨道可包括来自主机2a、2b、...2n的写或已修改轨道、经受读请求且作为结果被移动至第一高速缓存14的第二高速缓存18中的非顺序轨道或者在高速缓存14或18中未找到并从储存器10检索的被读请求的数据。在(在方框250处)接收到要添加到第一高速缓存14的轨道时，如果(在方框252处)轨道的拷贝已被包括在第一高速缓存14中，即接收到的轨道是写入，则高速缓存管理器24更新(在方框254处)第一高速缓存14中的轨道。如果(在方框252处)轨道未在高速缓存中，则高速缓存管理器24创建(在方框256处)用于要添加的轨道的控制块100(图4)，指示在第一高速缓存14中的位置104以及轨道是已修改/未修改106和顺序/非顺序108。此控制块100被添加到第一高速缓存14的控制块目录52。高速缓存管理器24向第一高速缓存轨道索引50添加(在方框258处)条目，其具有要添加的轨道的轨道ID和到控制块目录52中的创建高速缓存控制块100的索引。向要添加轨道的轨道类型的LRU列表54、56或58的MRU结尾添加(在方框260处)条目。如果(在方框262处)要添加的轨道是已修改非顺序轨道，并且如果(在方框264处)根据第二高速缓存轨道索引70的确定要添加的轨道的拷贝在第二高速缓存18中，则使第二高速高速缓存18中的轨道的拷贝无效(在方框266处)，诸如通过将用于第二高速缓存18中的轨道的高速缓存控制块120的有效/无效标志128设置成无效。如果(在方框306处)要添加的轨道是未修改顺序的，则控制结束。

[0049] 图11图示出由高速缓存管理器24执行以从来自第一高速缓存14的第一步幅向第二高速缓存18中的第二步幅添加轨道的操作实施例。高速缓存管理器24创建(在方框302处)用于第二步幅的步幅信息130(图6)，将来自被添加并指示占用136的第一步幅的轨道134指示为已满。针对被添加的第一步幅中的每个轨道，在方框304至318处执行操作循环。高速缓存管理器24添加(在方框302处)被晋级至第二高速缓存18中的LSA 32的轨道的指示，诸如轨道ID。如果(在方框308处)正在添加的轨道已经在第二高速缓存18中，则高速缓存管理器24更新(在方框310处)用于轨道的高速缓存控制块120，其指示LSA 32中的位置124、数据是未修改的126以及轨道是有效的128。如果(在方框308处)轨道不在第二高速缓存18中，则高速缓存管理器24创建(在方框312处)用于要添加的轨道的控制块120(图5)，其指示LSA 32中的轨道位置124以及轨道是已修改/未修改的126。向具有已晋级轨道的轨道ID和对用于第二高速缓存18的控制块目录72中的所创建高速缓存控制块120的索引的第二高速缓存轨道索引70添加条目(在方框314处)。从方框310至316，高速缓存管理器24指示(在方框316处)未修改LRU列表74的MRU结尾处的已晋级轨道，诸如通过向MRU结尾添加轨道ID。

[0050] 图12图示出由高速缓存管理器24执行以释放第二高速缓存18中的空间以用于要添加到第二高速缓存18的新轨道、即正在从第一高速缓存14降级的轨道的操作实施例。在发起此操作(在方框350处)时，高速缓存管理器24从未修改LRU列表74的LRU结尾确定(在方

框352处)第二高速缓存18中的未修改轨道,并使所确定未修改轨道(在方框354处)无效而不使无效未修改轨道降级至储存器10,并且还从未修改LRU列表74去除无效未修改轨道并在用于该轨道的高速缓存控制块120中将该轨道指示为无效128。第二高速缓存18中的未修改轨道可包括被添加到第一高速缓存14的读轨道或从第一高速缓存14降级的已修改轨道。此外,由高速缓存管理器24选择用于从第二高速缓存18降级的轨道可来自在第二高速缓存18中形成的不同步幅。此外,第二高速缓存中的步幅可包括有效轨道和无效轨道两者,其中,通过从第二高速缓存18降级或者通过在第一高速缓存18中更新轨道来使轨道无效。

[0051] 在某些实施例中,高速缓存管理器24使用不同的轨道降级算法以通过分别地对第一高速缓存14和第二高速缓存18使用单独的LRU列表58和74来确定将从第一高速缓存14和第二高速缓存18降级的轨道,以确定要降级的轨道。用来在第一高速缓存14和第二高速缓存18中选择用于降级的轨道的算法可考虑第一高速缓存14和第二高速缓存18中的轨道的特性以确定将首先降级的轨道。

[0052] 图13图示出由高速缓存管理器24执行以释放第二高速缓存18中的步幅以使得可用于第一高速缓存14中的轨道步幅的操作实施例。在发起(在方框370处)用以释放第二高速缓存18中的步幅时,高速缓存管理器确定(在方框372处)空闲步幅、即具有零占用136的步幅的数目是否小于空闲步幅阈值。例如,高速缓存管理器24可确保始终存在至少两个或某个其他数目的步幅可用于由第一高速缓存14轨道形成的步幅。如果空闲步幅的数目不在阈值以下,则控制结束。否则,如果(在方框372处)空闲步幅的数目小于阈值,则高速缓存管理器24确定(在方框374处)具有零占用136的可用步幅且确定(在方框376处)部分已满、即具有有效轨道和无效轨道的至少两个步幅,其有效轨道能够适合于空闲步幅。高速缓存管理器24(在方框378处)将来自所确定的至少两个部分已满步幅的有效轨道组合成所确定可用步幅。高速缓存管理器24然后(在方框380处)将所述至少两个步幅指示为具有零占用136,来自该至少两个步幅的轨道被合并,因此其可用于从来自第一高速缓存14的步幅接收轨道。

[0053] 图14图示出由高速缓存管理器24执行以从高速缓存14和高速缓存18及储存器10检索用于读请求的被请求轨道的操作实施例。处理读请求的存储管理器22可向高速缓存管理器24提交用于被请求轨道的请求。在接收到(在方框450处)用于轨道的请求时,高速缓存管理器24使用(在方框254处)第一高速缓存轨道索引50来确定是否所有被请求轨道都在第一高速缓存14中。如果(在方框454处)并非所有被请求轨道都在第一高速缓存14中,则高速缓存管理器24使用(在方框456处)第二高速缓存轨道索引70来确定不在第一高速缓存14中的第二高速缓存18中的任何被请求轨道。如果(在方框458处)存在第一高速缓存14和第二高速缓存18中未找到的任何被请求轨道,则高速缓存管理器24确定(在方框460处)来自第二高速缓存轨道索引70的储存器10中的任何被请求轨道不在第一高速缓存14和第二高速缓存18中。高速缓存管理器24然后将第二高速缓存18和储存器10中的任何所确定轨道(在方框462处)晋级至第一高速缓存14。高速缓存管理器24使用(在方框264处)第一高速缓存轨道索引50来从第一高速缓存14检索被请求轨道以返回至读请求。用于所检索轨道的条目被移动(在方框466处)至LRU列表54、56、58的MRU结尾,包括用于所检索轨道的条目。

[0054] 用图13的操作,高速缓存管理器24在前进至储存器10之前首先从最高水平高速缓存14、然后是第二高速缓存检索被请求轨道,因为高速缓存14和高速缓存18将具有被请求

轨道的最新修改版本。首先在第一高速缓存14中找到最新版本，如果不在第一高速缓存14中的话然后是第二高速缓存18，并且如果高速缓存14和高速缓存18中均不在的话然后是储存器10。

[0055] 所述实施例提供了用以按根据用于第二高速缓存的RAID配置定义的步幅将第一高速缓存中的轨道分组、使得能够按步幅将第一高速缓存中的轨道分组到第二高速缓存中的技术。然后可将高速缓存在第二高速缓存中的轨道分组成根据用于储存器的RAID配置定义的步幅，并且然后写入存储系统。

[0056] 所述实施例提供了用于分步幅地将轨道从第一高速缓存晋级、使得可将该轨道作为全步幅写而写入第二高速缓存中的各步幅以改善高速缓存晋级操作的效率的技术。所述实施例允许使用全步幅写来将第一高速缓存中的已降级轨道晋级到第二高速缓存，以便通过将整个步幅晋级至第二高速缓存作为单个I/O操作来节省资源。

[0057] 此外，在正在将轨道作为步幅从第一高速缓存14晋级至第二高速缓存18的同时，根据诸如LRU算法之类的高速缓存降级算法而逐个轨道地将轨道从第二高速缓存18降级。

[0058] 可使用标准编程和/或工程技术将所述操作实现为方法、装置或计算机程序产品以产生软件、固件、硬件或其任何组合。所属技术领域的技术人员知道，本实施例的各个方面可以实现为系统、方法或计算机程序产品。因此，本发明的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式(包括固件、驻留软件、微代码等)，或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。此外，在一些实施例中，本实施例的各个方面还可以实现为在一个或多个计算机可读介质中的计算机程序产品的形式，该计算机可读介质中包含计算机可读的程序代码。

[0059] 可以采用一个或多个计算机可读介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0060] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括——但不限于——电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0061] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括——但不限于——无线、有线、光缆、RF等等，或者上述的任意合适的组合。

[0062] 可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、

部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0063] 下面将参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述本发明。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机程序指令实现。这些计算机程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器，从而生产出一种机器，使得这些计算机程序指令在通过计算机或其它可编程数据处理装置的处理器执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。

[0064] 也可以把这些计算机程序指令存储在计算机可读介质中，这些指令使得计算机、其它可编程数据处理装置、或其他设备以特定方式工作，从而，存储在计算机可读介质中的指令就产生出包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的指令的制造品(article of manufacture)。

[0065] 也可以把计算机程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机或其它可编程装置上执行的指令能够提供实现流程图和/或框图中的方框中规定的功能/操作的过程。

[0066] 术语“一实施例”、“实施例”、“多个实施例”、“该实施例”、“一个或多个实施例”、“某些实施例”以及“一个实施例”意指“本发明的一个或多个(但并非全部)实施例”，除非另外明确地指明。

[0067] 术语“包括”、“包含”、“具有”及其变体意指“包括但不限于”，除非另外明确地指定。

[0068] 项目的枚举列表并不意味着任何或所有项目是相互排他性的，除非另外明确地指明。

[0069] 术语“一”、“一个”和“该”意指“一个或多个”，除非另外明确地指明。

[0070] 相互通信的设备不需要相互连续地通信，除非另外明确地指明。另外，相互通信的设备可以通过一个或多个中介来直接地或间接地进行通信。

[0071] 具有相互通信的多个部件的实施例的描述并不意味着要求所有此类部件。相反，描述多种可选部件是为了举例说明本发明的可能实施例的宽泛种类而描述的。

[0072] 此外，虽然可以按照相继顺序来描述过程步骤、方法步骤、算法等，但此类过程、方法和算法可以被配置成按照交替顺序来工作。换言之，可以描述的步骤的任何序列或顺序不一定指示按照该顺序来执行步骤的要求。可以按照任何实际的顺序来执行本文所述的过程步骤。此外，可以同时地执行某些步骤。

[0073] 当在本文中描述单个设备/物品(无论其是否合作)时，将很容易显而易见的是可以使用不止一个设备或物品来代替单个设备/物品。同样地，在本文中描述超过一个设备或物品(无论其是否合作)的情况下，将显而易见的是可使用单个设备/物品来代替超过一个设备或物品，或者可使用不同数目的设备/物品而不是所示数目的设备或程序。可以备选地用一个或多个其他设备来体现设备的功能和/或特征，其未被明确地描述为具有此类功能/

特征。因此，本发明的其他实施例不需要包括设备本身。

[0074] 图的所示操作示出了按照某个顺序发生的某些事件。在替换实施例中国，某些操作可按照不同顺序执行、修改或去除。此外，可向上述逻辑添加步骤且其仍符合所述实施例。此外，本文所述的操作可连续地发生，或者可并行地执行某些操作。此外，可由单个处理单元或由分布式处理单元来执行操作。

[0075] 本发明的各种实施例的先前描述是出于举例说明和描述的目的提出的。其并不意图是排他性的或使本发明局限于所公开的精确形式。鉴于上述教导，可以有许多修改和变更。意图在于本发明的范围不受此详细描述的限制，而是由所附权利要求来限制。以上说明书、示例和数据提供了本发明的组合物的制造和使用的完整的叙述。由于在不脱离本发明的精神和范围的情况下可实现本发明的许多实施例，所以本发明还存在于随后的所附权利要求中。

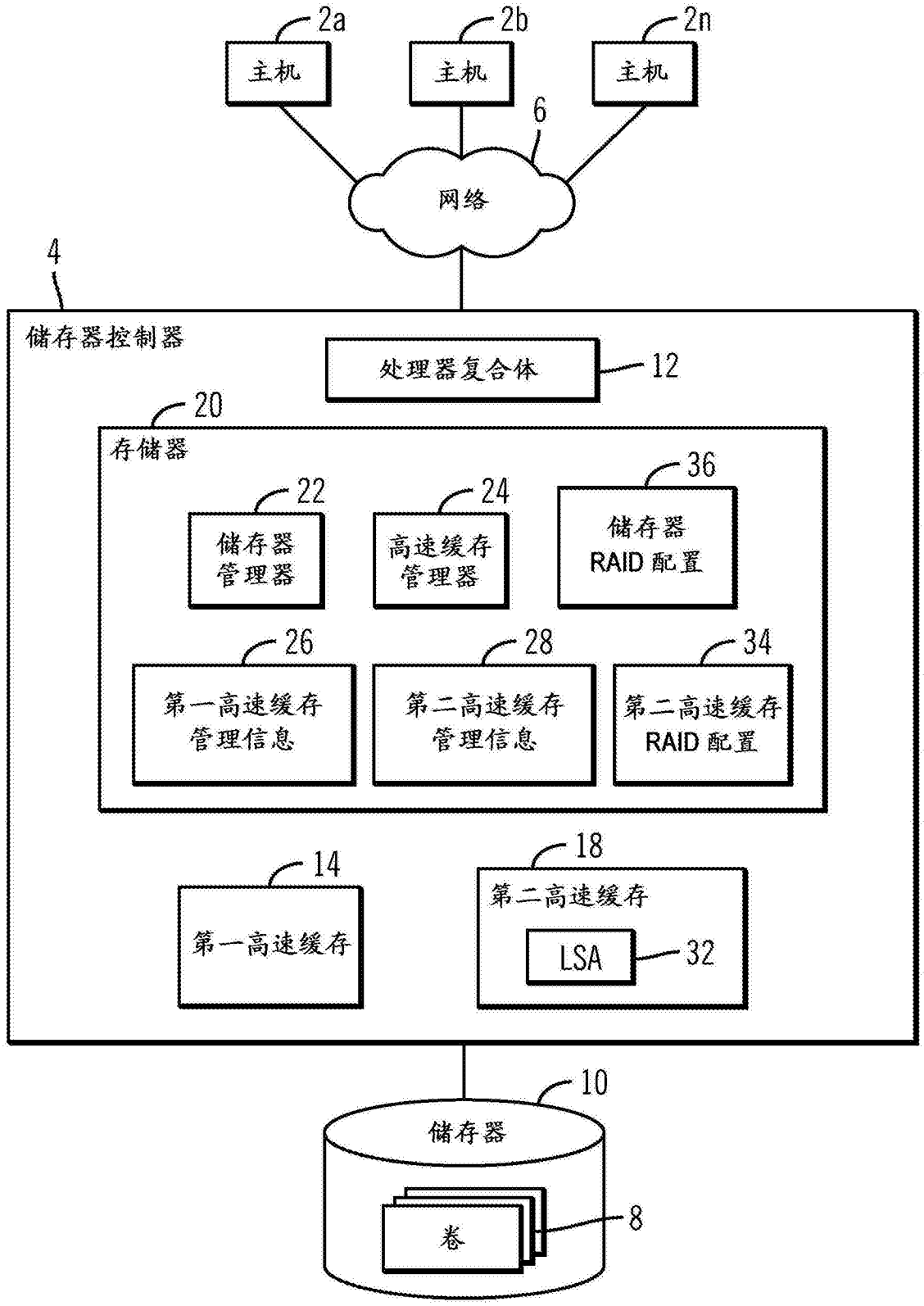

2a
主机
2b
主机
2n
主机
6
网络
4
储存器控制器
处理器复合体
12
20
存储器
22
储存器
管理器
24
高速缓存
管理器
36
储存器
RAID 配置
26
第一高速缓存
管理信息
28
第二高速缓存
管理信息
34
第二高速缓存
RAID 配置
14
第一高速缓存
18
第二高速缓存
LSA
32
10
储存器
卷
8

图1

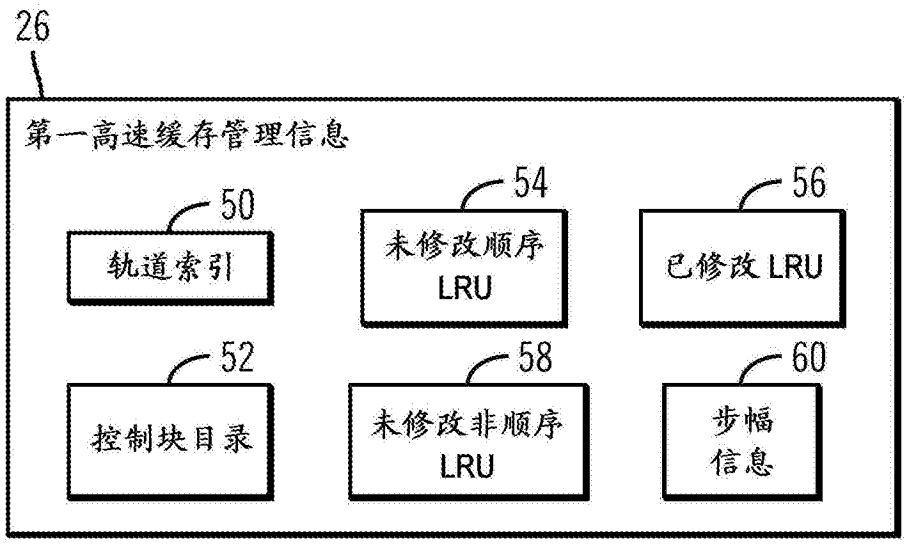

26
第一高速缓存管理信息
50
轨道索引
54
未修改顺序 LRU
56
已修改 LRU
52
控制块目录
58
未修改非顺序 LRU
60
步幅信息

图2

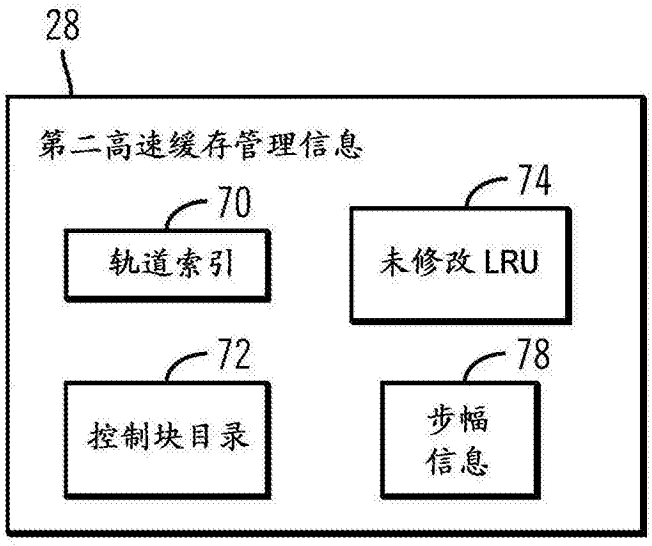

28
第二高速缓存管理信息
70
轨道索引
74
未修改 LRU
72
控制块目录
78
步幅信息

图3

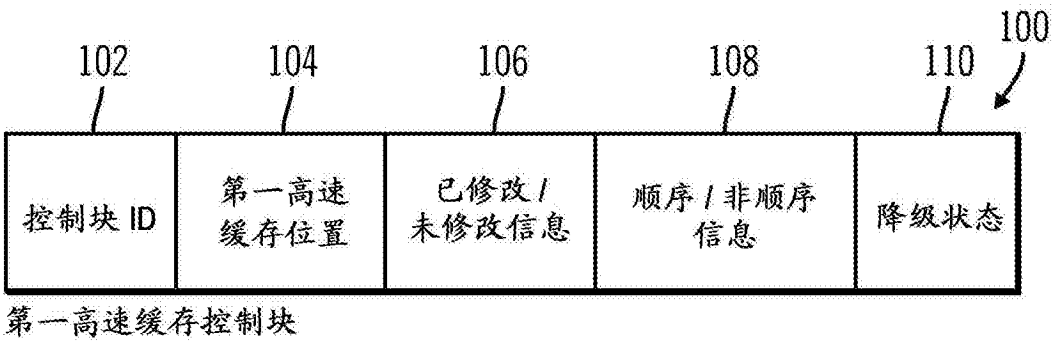

100
102
104
106
108
110
控制块 ID
第一高速缓存位置
已修改 / 未修改信息
顺序 / 非顺序信息
降级状态
第一高速缓存控制块

图4

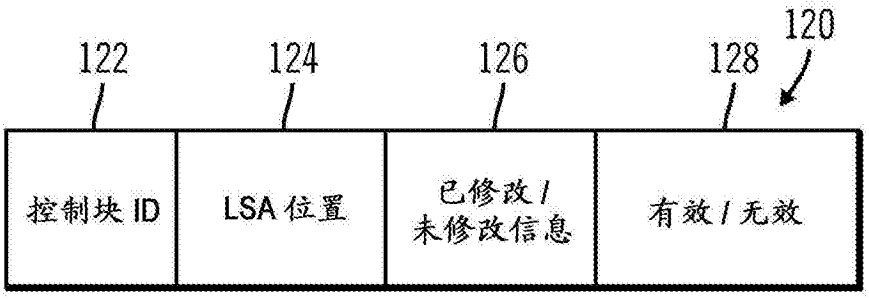

122
124
126
128
120
控制块 ID
LSA 位置
已修改 / 未修改信息
有效 / 无效
第二高速缓存控制块

图5

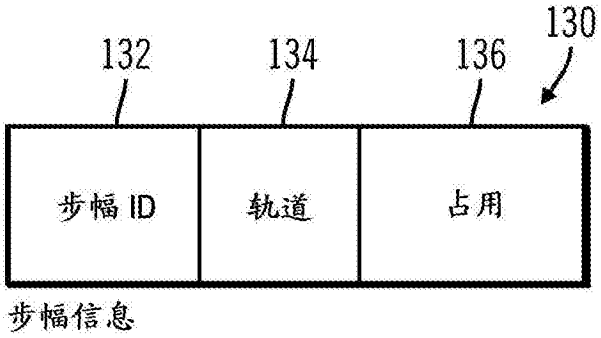

132
134
136
130
步幅 ID
轨道
占用
步幅信息

图6

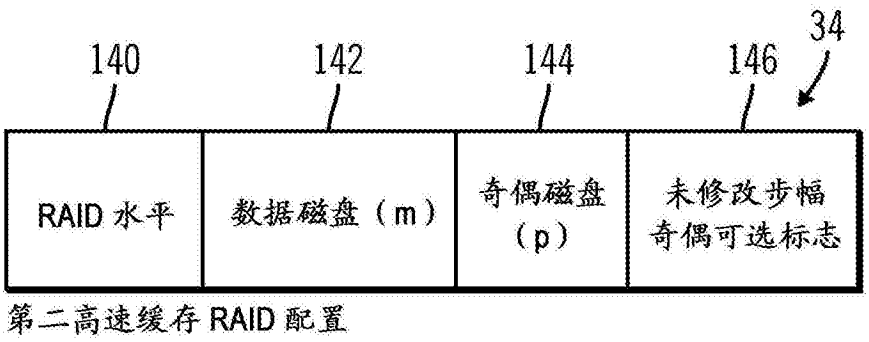

140
142
144
146
34
RAID 水平
数据磁盘（m）
奇偶磁盘（p）
未修改步幅奇偶可选标志
第二高速缓存 RAID 配置

图7

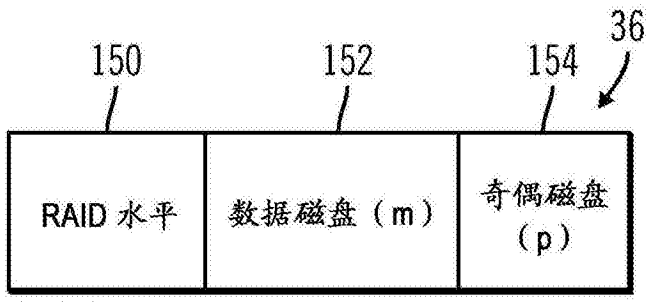


图8

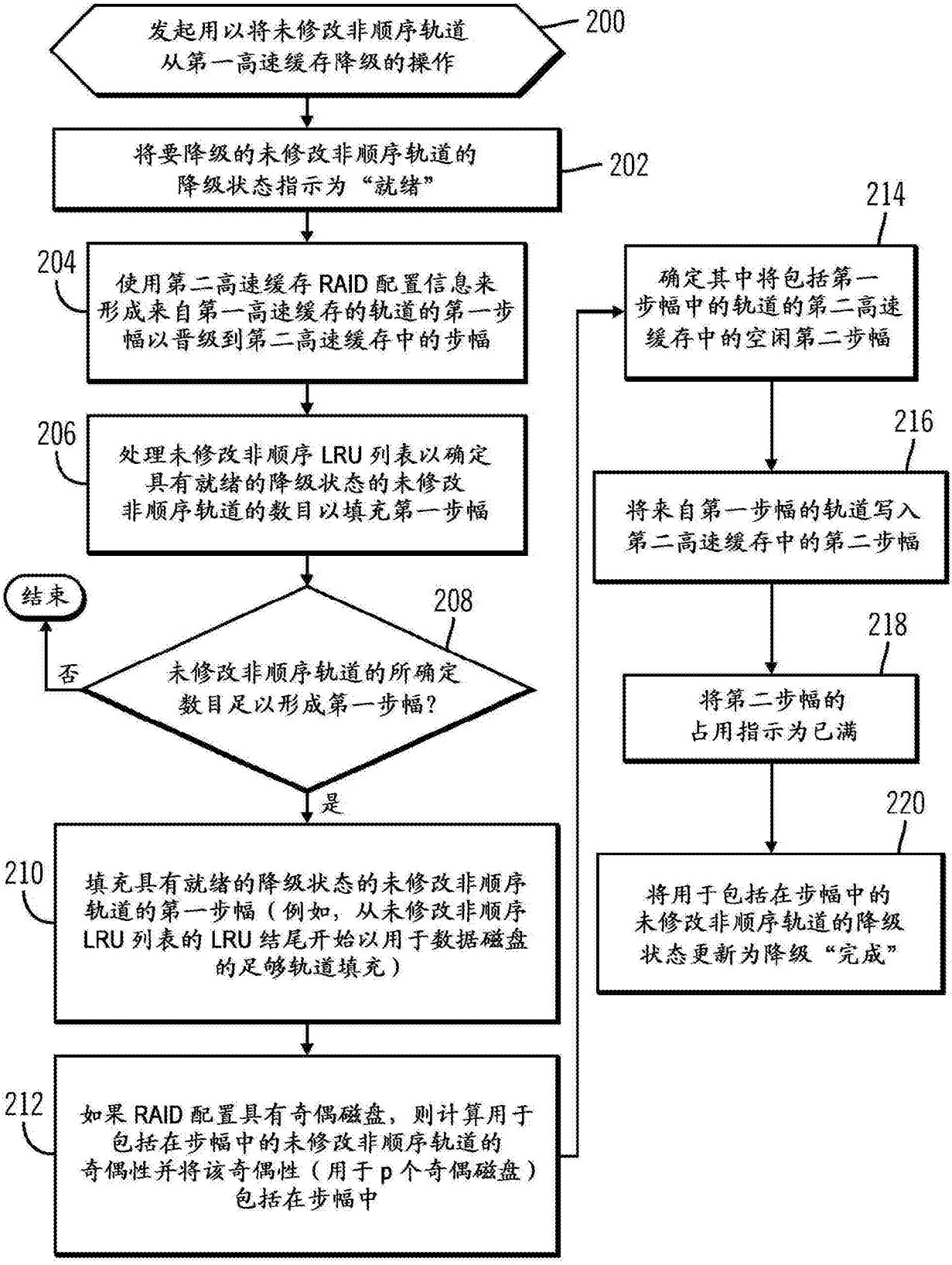

发起用以将未修改非顺序轨道从第一高速缓存降级的操作
200
将要降级的未修改非顺序轨道的降级状态指示为“就绪”
202
204
使用第二高速缓存 RAID 配置信息来形成来自第一高速缓存的轨道的第一步幅以晋级到第二高速缓存中的步幅
206
处理未修改非顺序 LRU 列表以确定具有就绪的降级状态的未修改非顺序轨道的数目以填充第一步幅
结束
否
208
未修改非顺序轨道的所确定数目足以形成第一步幅?
是
210
填充具有就绪的降级状态的未修改非顺序轨道的第一步幅（例如，从未修改非顺序 LRU 列表的 LRU 结尾开始以用于数据磁盘的足够轨道填充）
212
如果 RAID 配置具有奇偶磁盘，则计算用于包括在步幅中的未修改非顺序轨道的奇偶性并将该奇偶性（用于 p 个奇偶磁盘）包括在步幅中
214
确定其中将包括第一步幅中的轨道的第二高速缓存中的空闲第二步幅
216
将来自第一步幅的轨道写入第二高速缓存中的第二步幅
218
将第二步幅的占用指示为已满
220
将用于包括在步幅中的未修改非顺序轨道的降级状态更新为降级“完成”

图9

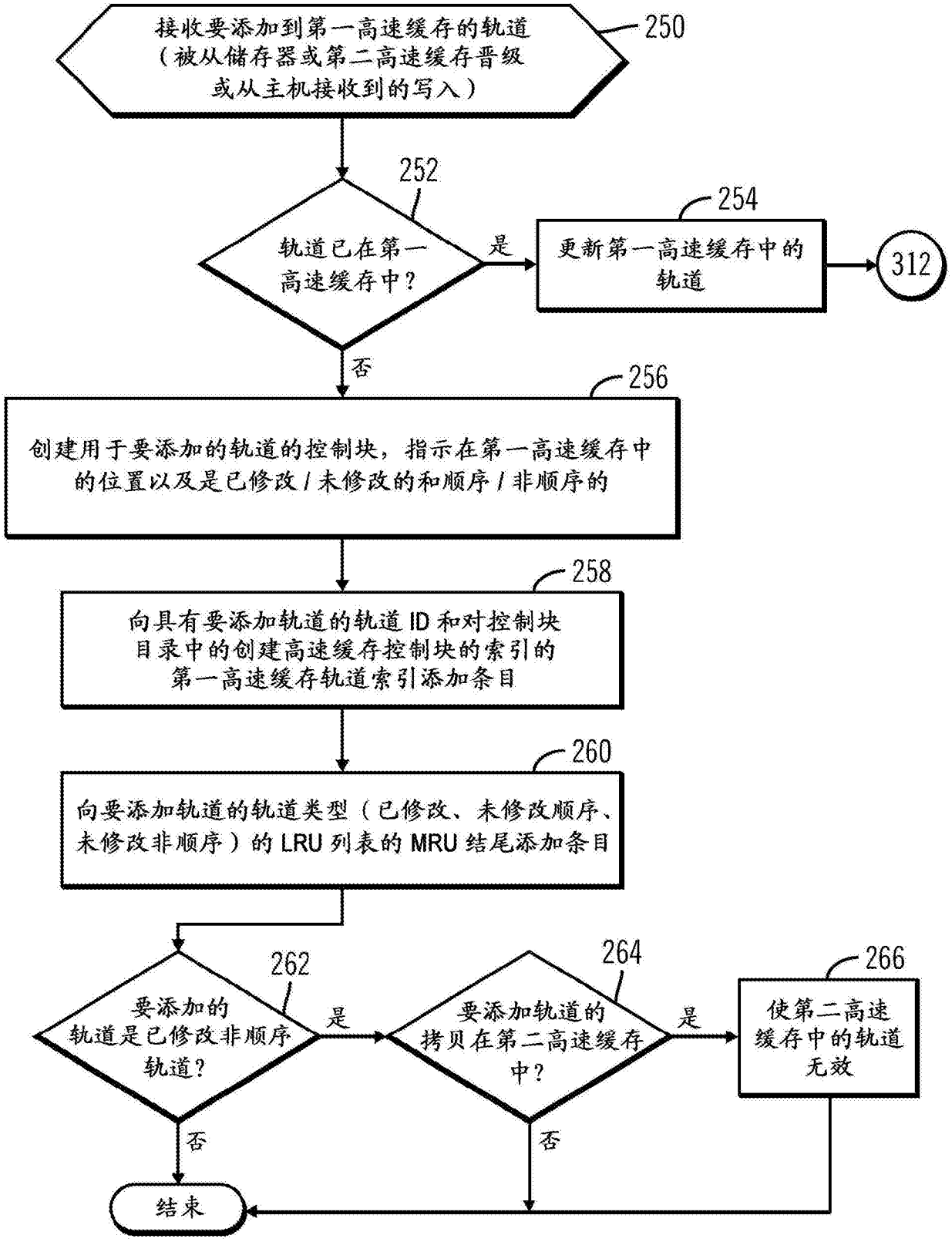

接收要添加到第一高速缓存的轨道（被从储存器或第二高速缓存晋级或从主机接收到的写入）
250
252
轨道已在第一高速缓存中？
是
254
更新第一高速缓存中的轨道
312
否
256
创建用于要添加的轨道的控制块，指示在第一高速缓存中的位置以及是已修改 / 未修改的和顺序 / 非顺序的
258
向具有要添加轨道的轨道 ID 和对控制块目录中的创建高速缓存控制块的索引的第一高速缓存轨道索引添加条目
260
向要添加轨道的轨道类型（已修改、未修改顺序、未修改非顺序）的 LRU 列表的 MRU 结尾添加条目
262
要添加的轨道是已修改非顺序轨道？
是
264
要添加轨道的拷贝在第二高速缓存中？
是
266
使第二高速缓存中的轨道无效
否
否
结束

图10

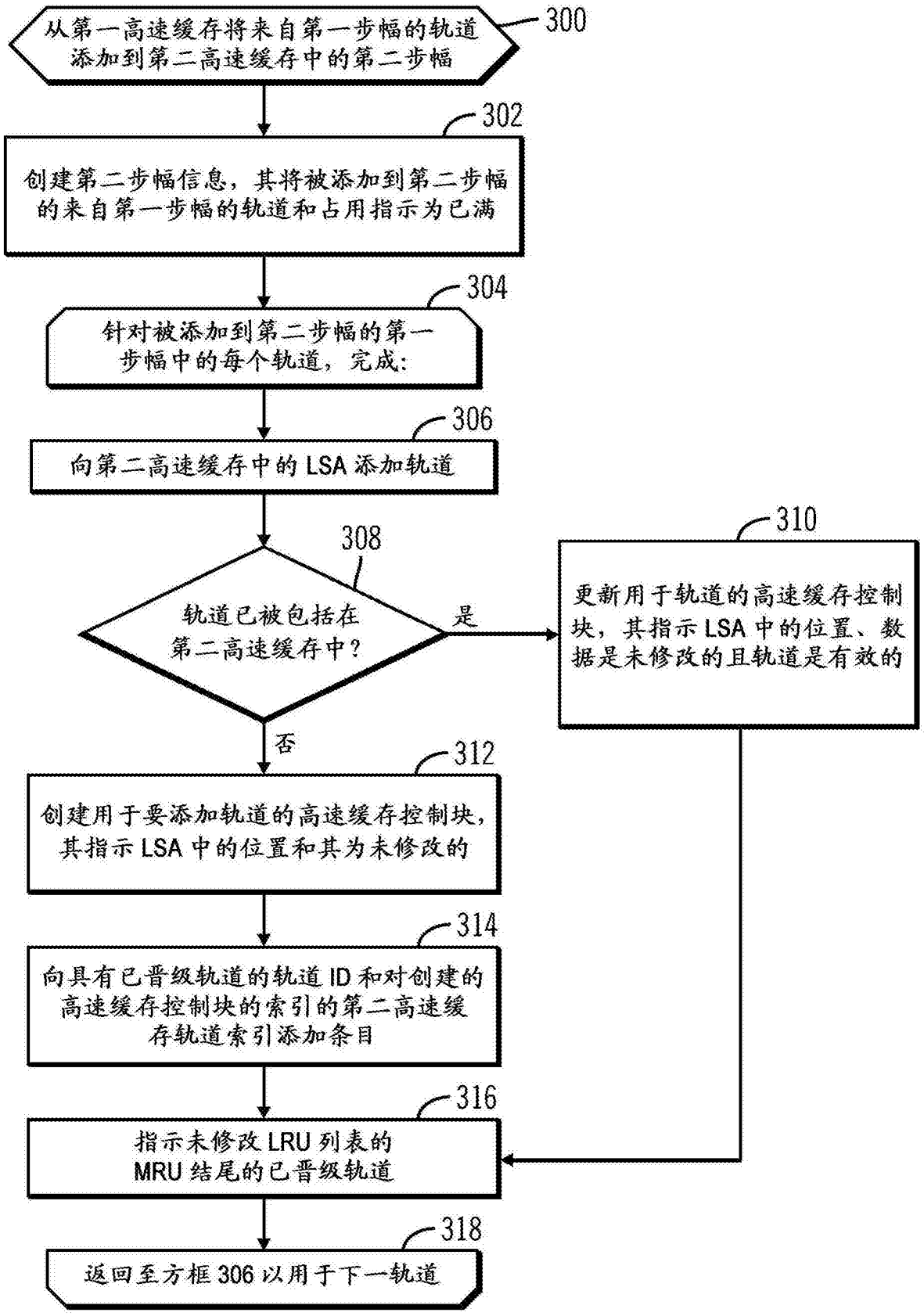

从第一高速缓存将来自第一步幅的轨道添加到第二高速缓存中的第二步幅
300
创建第二步幅信息，其将被添加到第二步幅的来自第一步幅的轨道和占用指示为已满
302
针对被添加到第二步幅的第一步幅中的每个轨道，完成：
304
向第二高速缓存中的 LSA 添加轨道
306
轨道已被包括在第二高速缓存中？
308
是
更新用于轨道的高速缓存控制块，其指示 LSA 中的位置、数据是未修改的且轨道是有效的
310
否
创建用于要添加轨道的高速缓存控制块，其指示 LSA 中的位置和其为未修改的
312
向具有已晋级轨道的轨道 ID 和对创建的高速缓存控制块的索引的第二高速缓存轨道索引添加条目
314
指示未修改 LRU 列表的 MRU 结尾的已晋级轨道
316
返回至方框 306 以用于下一轨道
318

图11

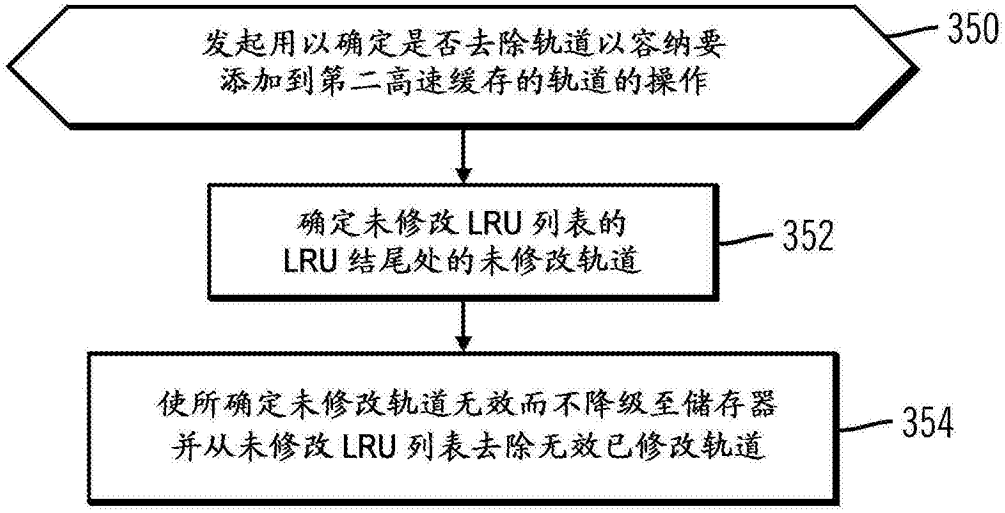

发起用以确定是否去除轨道以容纳要添加到第二高速缓存的轨道的操作
350
确定未修改 LRU 列表的 LRU 结尾处的未修改轨道
352
使所确定未修改轨道无效而不降级至储存器并从未修改 LRU 列表去除无效已修改轨道
354

图12

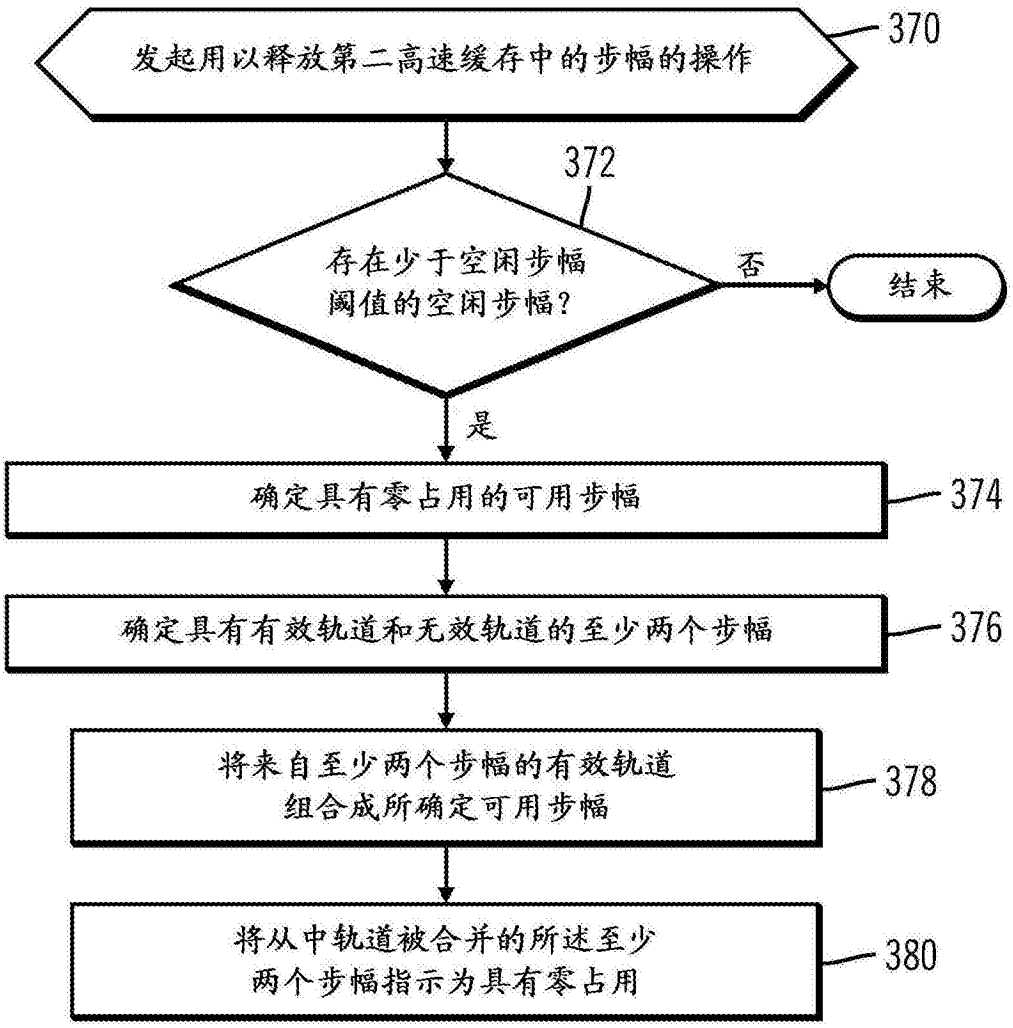

发起用以释放第二高速缓存中的步幅的操作
370
372
存在少于空闲步幅阈值的空闲步幅?
否
结束
是
确定具有零占用的可用步幅
374
确定具有有效轨道和无效轨道的至少两个步幅
376
将来自至少两个步幅的有效轨道组合成所确定可用步幅
378
将从中轨道被合并的所述至少两个步幅指示为具有零占用
380

图13

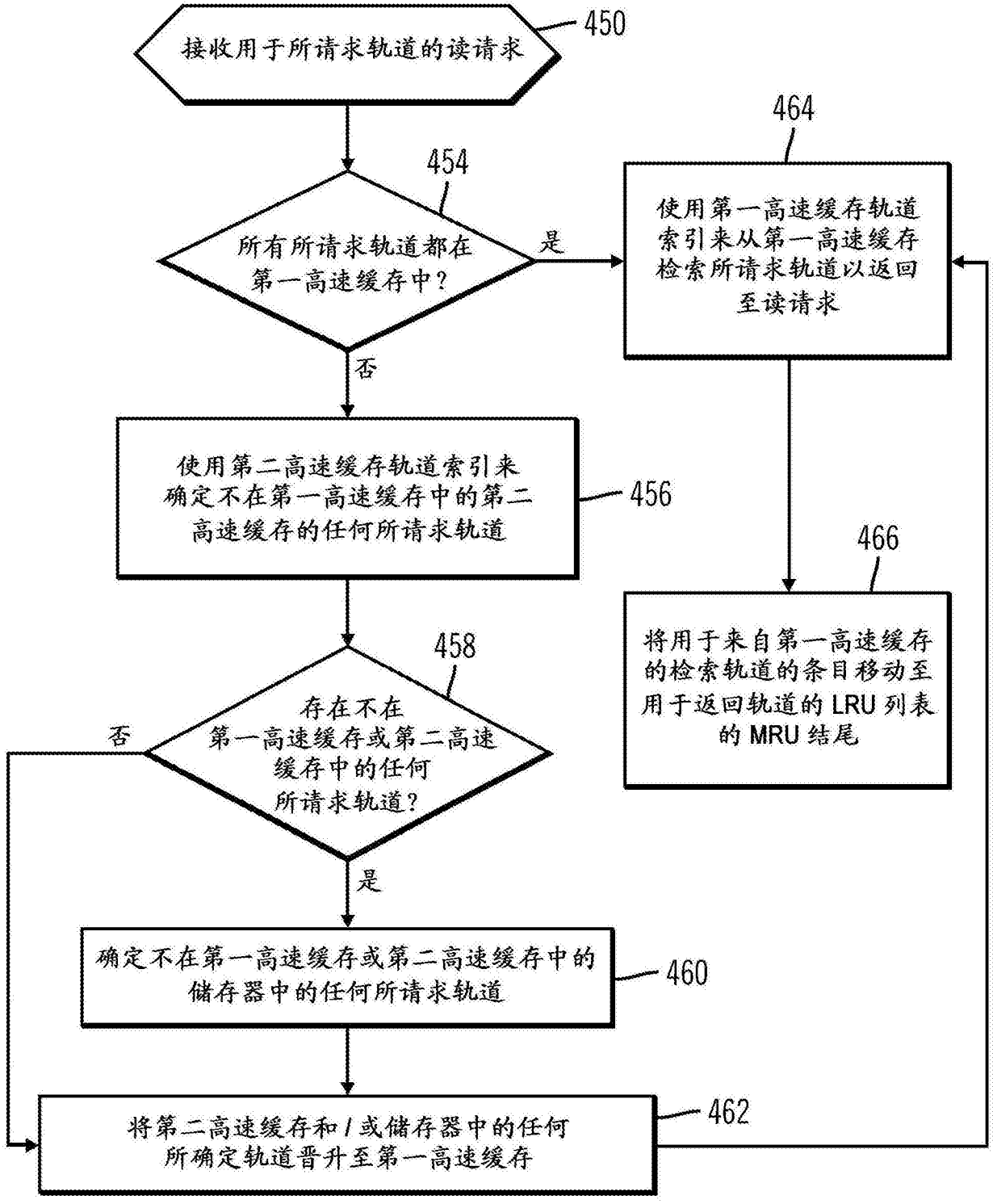

接收用于所请求轨道的读请求
450
所有所请求轨道都在第一高速缓存中？
454
是
否
使用第一高速缓存轨道索引来从第一高速缓存检索所请求轨道以返回至读请求
464
使用第二高速缓存轨道索引来确定不在第一高速缓存中的第二高速缓存的任何所请求轨道
456
将用于来自第一高速缓存的检索轨道的条目移动至用于返回轨道的 LRU 列表的 MRU 结尾
466
存在不在第一高速缓存或第二高速缓存中的任何所请求轨道？
458
否
是
确定不在第一高速缓存或第二高速缓存中的储存器中的任何所请求轨道
460
将第二高速缓存和 / 或储存器中的任何所确定轨道晋升至第一高速缓存
462

图14