(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115620778 A
(43)申请公布日 2023.01.17

(21)申请号 202210835484.6

(22)申请日 2022.07.15

(30)优先权数据
63/222,685 2021.07.16 US
17/481,897 2021.09.22 US

(71)申请人 三星电子株式会社
地址 韩国京畿道

(72)发明人 K.S.帕特瓦尔丹 N.拉玛克里什南

(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 刘虹

(51)Int.Cl.
G11C 16/04(2006.01)
G11C 16/10(2006.01)

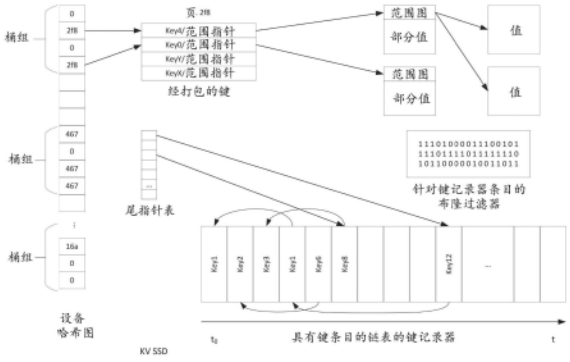
权利要求书3页 说明书11页 附图11页

(54)发明名称

用于闪存键值存储操作的键打包

(57)摘要

本文提供了键值(KV)存储、其方法和存储系统。KV存储可以包括键记录器;以及处理器,该处理器被配置为接收用于在KV存储中存储第一KV的第一命令,将第一KV的第一值写入第一NAND页,生成用于标识包括第一值的第一存储器页的范围图,将范围图写入第二存储器页,将用于存储第一KV的条目附加到键记录器,以及当在键记录器内达到阈值时,更新KV存储的设备哈希图以包括第一KV的第一键。



CN 115620778 A

1.一种键值KV存储,包括:

键记录器;以及

处理器,被配置为:

接收用于在KV存储中存储第一KV的第一命令,

将第一KV的第一值写入第一存储器页,

生成用于标识包括第一值的第一存储器页的范围图,

将范围图写入第二存储器页,

将用于存储第一KV的条目附加到键记录器,以及

当在键记录器内达到阈值时,更新KV存储的设备哈希图以包括第一KV的第一键。

2.根据权利要求1所述的KV存储,其中,所述第一KV的第一键与至少一个其他键一起存储在第三存储器页上,并且包括指向包括范围图的第二存储器页的范围指针。

3.根据权利要求2所述的KV存储,其中,用于存储第一KV的条目包括第一键、包括范围图的第二存储器页的指示、第一命令的操作码或尾条目指针中的至少一个。

4.根据权利要求2所述的KV存储,其中,所述处理器还被配置为通过以下方式更新设备哈希图:

收集属于设备哈希图中的一组相邻桶的键条目;

合并设备哈希图中的现有键与来自键记录器的条目,以形成所述一组相邻桶的经打包的键;

将经打包的键写入第三存储器页;以及

更新设备哈希图以指向包括经打包的键的第三存储器页。

5.根据权利要求1所述的KV存储,还包括布隆过滤器,

其中,所述处理器还被配置为更新布隆过滤器以指示对用于存储第一KV的第一命令的接收。

6.根据权利要求1所述的KV存储,还包括尾指针表,

其中,所述处理器还被配置为:

基于第一KV的第一键的哈希选择尾指针表中的槽,以及

将键记录器中的当前偏移写入所选择的槽。

7.根据权利要求1所述的KV存储,其中,所述处理器还被配置为清除键记录器。

8.根据权利要求1所述的KV存储,其中,所述处理器还被配置为:

接收用于从KV存储中删除第二KV的第二命令,

确定键记录器是否包括第二KV的条目、或者第二KV是否存在于设备哈希图中,以及

响应于确定键记录器包括第二KV的条目或者第二KV存在于设备哈希图中,将用于删除第二KV的条目附加到键记录器,并且当在键记录器内满足阈值时,从设备哈希图中删除第二KV的第二键。

9.根据权利要求8所述的KV存储,还包括布隆过滤器,

其中,所述处理器还被配置为在确定键记录器是否包括第二KV的条目之前,在布隆过滤器中检查第二KV的先前条目。

10.根据权利要求1所述的KV存储,其中,所述处理器还被配置为:

接收用于从KV存储检索第二KV的第三命令,

在键记录器中标识第二KV的条目，

识别用于键记录器中第二KV的条目的操作码是用于删除操作的，以及

基于用于键记录器中第二KV的条目的操作码是用于删除操作的，识别第三命令失败。

11.根据权利要求10所述的KV存储，还包括布隆过滤器，

其中，所述处理器还被配置为在标识出键记录器包括第二KV的条目之前，在布隆过滤器中检查第二KV的先前条目。

12.根据权利要求1所述的KV存储，其中，所述处理器还被配置为：

接收用于从KV存储检索第二KV的第三命令，以及

在设备哈希图中标识第二KV。

13.根据权利要求12所述的KV存储，其中，所述处理器还被配置为：

读取包括第二KV的第二键的第二范围图的第三存储器页，以及

基于所述第二范围图，从第四存储器页读取第二KV的第二值。

14.一种操作键值KV存储的方法，所述方法包括：

接收用于在KV存储中存储第一KV的第一命令；

将第一KV的第一值写入第一存储器页；

生成用于标识包括第一值的第一存储器页的范围图；

将范围图写入第二存储器页；

将用于存储第一KV的条目附加到KV存储的键记录器；以及

当在键记录器内达到阈值时，更新KV存储的设备哈希图以包括第一KV的第一键。

15.根据权利要求14所述的方法，其中，所述第一KV的第一键与至少一个其他键一起存储在第三存储器页上，并且包括指向包括范围图的第二存储器页的范围指针。

16.根据权利要求15所述的方法，其中，用于存储第一KV的条目包括第一键、包括范围图的第二存储器页的指示、第一命令的操作码或尾条目指针中的至少一个。

17.根据权利要求15所述的方法，其中，更新设备哈希图包括：

收集属于设备哈希图中一组相邻桶的键条目；

合并设备哈希图中的现有键与来自键记录器的条目，以形成所述一组相邻桶的经打包的键；

将经打包的键写入第三存储器页；以及

更新设备哈希图以指向包括经打包的键的第三存储器页。

18.根据权利要求14所述的方法，还包括更新键记录器的布隆过滤器以指示对用于存储第一KV的第一命令的接收。

19.根据权利要求14所述的方法，还包括：

基于第一KV的第一键的哈希选择尾指针表中的槽；以及

将键记录器中的当前偏移写入所选择的槽。

20.一种存储系统，包括：

键值KV存储；以及

处理器，被配置为：

接收用于在KV存储中存储第一KV的第一命令，

将第一KV的第一值写入第一存储器页，

生成用于标识包括第一值的第一存储器页的范围图，

将范围图写入第二存储器页，

将用于存储第一KV的条目附加到KV存储的键记录器，以及

当在键记录器内达到阈值时，更新KV存储的设备哈希图以包括第一KV的第一键。

用于闪存键值存储操作的键打包

技术领域

[0001] 本公开总体上涉及键值(key value,KV)存储操作,并且更具体地,涉及通过将多个键一起存储在单个NAND页中来改进KV存储操作。

背景技术

[0002] KV存储(KV store)(其可以被理解为“KV存储设备”)或KV数据库(KV database)是被设计用于存储、检索和管理关联数组的数据存储范例,以及通常被称为哈希表的数据结构。哈希表包含对象或记录的集合,这些对象或记录又包含多个不同的字段,每个字段都包含数据。这些记录可以使用唯一标识该记录的键来存储和检索,并且可以用于在数据库中查找数据。KV存储通常比关系数据库使用更少的内存。

[0003] 以上信息仅作为背景信息呈现,以帮助理解本公开。关于上述任何一个是否可以作为现有技术应用于本公开,没有做出确定,也没有做出断言。

发明内容

[0004] 本公开是为了解决上述缺点中的至少一些,并且提供至少下述优点。

[0005] 本公开的一个方面是提供一种系统和方法,其允许打包没有时间局部性地插入的多个键,使得它们可以被分组为具有空间局部性的一堆存储器页(例如NAND页)。由于密集的键打包,这提供了对存储所有键所需的NAND页的数量的减少,以及读取放大因子(RAF)/写入放大因子(WAF)减少。此外,NAND页可以将垃圾收集(GC)限制在页中的所有键都无效的情况。

[0006] 因此,本公开可以通过打包密度的因子减少GC开销显著,并且减少哈希图(hash-map)冲突链的长度,这改善了尾延迟,并且缩减了NAND页读取的数量,显著降低了查找延迟。

[0007] 根据一个实施例,可以提供KV存储,其包括键记录器(key logger);以及处理器,该处理器被配置为接收用于在KV存储中存储第一KV的第一命令,将第一KV的第一值写入第一存储器页,生成用于标识包括该值的第一存储器页的范围图(extent map),将该范围图写入第二存储器页,将用于存储第一KV的条目附加到键记录器,以及当在键记录器内达到阈值时,更新KV存储的设备哈希图以包括第一KV的第一键。

[0008] 根据一个实施例,可以提供一种操作KV存储的方法。该方法包括接收用于在KV存储中存储第一KV的第一命令;将第一KV的第一值写入第一存储器页;生成用于标识包括该值的第一存储器页的范围图;将该范围图写入第二NAND页;将用于存储第一KV的条目附加到KV存储的键记录器;以及当在键记录器内达到阈值时,更新KV存储的设备哈希图以包括第一KV的第一键。

[0009] 根据一个实施例,可以提供一种存储系统,其包括KV存储;以及处理器,该处理器被配置为接收用于在KV存储中存储第一KV的第一命令,将第一KV的第一值写入第一存储器页,生成用于标识包括该值的第一存储器页的范围图,将该范围图写入第二存储器页,将用

于存储第一KV的条目附加到KV存储的键记录器，并且当在键记录器内达到阈值时，更新KV存储的设备哈希图以包括第一KV的第一键。

附图说明

[0010] 从下面结合附图的详细描述中，本公开的特定实施例的上述和其他方面、特征和优点将变得更加清楚，其中：

[0011] 图1示出了KV固态驱动(SSD)中的哈希实现方式；

[0012] 图2示出了根据实施例的KV SSD中的哈希实现方式；

[0013] 图3示出了根据实施例的KV存储；

[0014] 图4A和图4B示出了根据实施例的KV存储中的推迟的键打包操作；

[0015] 图5是示出根据实施例的KV存储的键打包操作的流程图；

[0016] 图6是示出根据实施例的KV存储的放置(Put)操作的流程图；

[0017] 图7是示出根据实施例的KV存储的删除(Delete)操作的流程图；

[0018] 图8是示出根据实施例的KV存储的获取(Get)操作的流程图；

[0019] 图9示出了根据实施例的网络环境中的电子设备的框图；和

[0020] 图10示出了根据实施例的存储系统的图。

具体实施方式

[0021] 在下文中，参考附图详细描述本公开的实施例。应当注意，尽管在不同的附图中示出了相同的元素，但是相同的元素将由相同的附图标记表示。在以下描述中，诸如详细配置和组件的具体细节仅被提供来帮助对本公开的实施例的整体理解。因此，对于本领域技术人员来说清楚的是，在不脱离本公开的范围的情况下，可以对本文描述的实施例进行各种改变和修改。此外，为了清楚和简洁，省略了对公知功能和结构的描述。下面描述的术语是考虑到本公开中的功能而定义的术语，并且可以根据用户、用户的意图或习惯而不同。因此，术语的定义应该基于整个说明书的内容来确定。

[0022] 本公开可以具有各种修改和各种实施例，下面参考附图详细描述其中的实施例。然而，应当理解，本公开不限于这些实施例，而是包括本公开范围内的所有修改、等同物和替代物。

[0023] 尽管包括序数(诸如第一、第二等)的术语可以用于描述各种元件，但是结构元件不受这些术语的限制。这些术语仅用于区分一个元件和另一个元件。例如，在不脱离本公开的范围的情况下，第一结构元件可以被称为第二结构元件。类似地，第二结构元件也可以被称为第一结构元件。如本文所用，术语“和/或”包括一个或多个关联项目的任何和所有组合。

[0024] 本文使用的术语仅用于描述本公开的各种实施例，而不旨在限制本公开。单数形式旨在包括复数形式，除非上下文另有明确指示。在本公开中，应该理解，术语“包括”或“具有”表示特征、数字、步骤、操作、结构元件、部件或其组合的存在，并且不排除一个或多个其他特征、数字、步骤、操作、结构元件、部件或其组合的存在或添加的可能性。

[0025] 除非有不同的定义，否则本文使用的所有术语具有与本公开所属领域的技术人员所理解的相同的含义。诸如在通用词典中定义的那些术语将被解释为具有与相关技术领域

中的上下文含义相同的含义，并且除非在本公开中明确定义，否则不应被解释为具有理想的或过于正式的含义。

[0026] 根据一个实施例的电子设备可以是各种类型的电子设备之一。电子设备可以包括例如便携式通信设备(例如，智能电话)、计算机、便携式多媒体设备、便携式医疗设备、相机、可穿戴设备或家用电器。根据本公开的一个实施例，电子设备不限于上述那些。

[0027] 本公开中使用的术语不旨在限制本公开，而是旨在包括相应实施例的各种变化、等同物或替换物。关于附图的描述，相似的附图标记可以用于指代相似或相关的元素。与项目相对应的名词的单数形式可以包括一个或多个事物，除非相关上下文清楚地表明不是这样。如本文所使用的，诸如“A或B”、“A和B中的至少一个”、“A或B中的至少一个”、“A、B或C中的至少一个”、“A、B和C中的至少一个”、以及“A、B或C中的至少一个”的短语中的每一个都可以包括在短语中的相应一个中一起列举的项目的所有可能的组合。如本文所使用的，诸如“第1”、“第2”、“第一”和“第二”的术语可以用于将相应的组件与另一组件区分开，但不旨在在其他方面(例如，重要性或顺序)限制组件。意图是，如果元件(例如，第一元件)被称为“与另一元件(例如，第二元件)“耦合”、“耦合到”、“连接”或“连接到”，无论有无术语“可操作地”或“通信地”，它都指示该元件可以直接地(例如，有线地)、无线地或经由第三元件与该另一元件耦合。

[0028] 如本文所使用的，术语“模块”可以包括以硬件、软件或固件实现的单元，并且可以与其他术语(例如，“逻辑”、“逻辑块”、“部分”和“电路”)互换使用。模块可以是适于执行一个或多个功能的单个集成组件、或者是其最小单元或部分。例如，根据一个实施例，模块可以以专用集成电路(ASIC)的形式实现。

[0029] 在本文中，所描述的(即，在设备图中的)任何组件或组件的任何组合可以用于执行流程图的一个或多个操作。流程图中描述的操作是示例性操作，并且可以包括流程图中没有明确提供的各种附加步骤。流程图中描述的操作顺序是示例性的而非排他性的，因为该顺序可以取决于实现方式而变化。

[0030] 键值(KV)存储的目标可以是在尽可能快的时间内完成操作(例如，放置/获取/删除操作)，并且在典型的KV存储实现方式中，键可以被哈希化并插入到哈希图中。

[0031] 减少KV操作的尾延迟(tail latency)可能有利于多种工作负载，如人工智能/机器学习(AI/ML)、数据科学等。此外，KV技术将会适用于寿命更短的致密介质可能是有益的。

[0032] 高性能计算(High performance computing，HPC)存储装置可以被设计为与具有50-500个写入周期的介质兼容。因此，减少由应用或设备诱导的GC引起的WAF也是有益的。例如，减少设备WAF也可以降低KV存储中的RAF，从而可能提高整体存储性能。

[0033] 单个NAND页可以包含一个键，这可能导致在处理删除操作时更高的设备WAF以及在处理放置/获取/存在(Exist)操作时更高的RAF。

[0034] 此外，KV存储的最坏情况的尾延迟可以与哈希图的桶(bucket)中最长的冲突链(collision chain)的长度成比例。增加桶的数量以减少冲突链会导致需要额外的存储器空间来托管桶。

[0035] 此外，如果桶本身存储在NAND中，则当处理KV的放置时，设备WAF增加，因为对于几乎每个KV插入，托管桶的NAND页可能需要被替换。

[0036] 根据本公开，提供了允许将多个键一起打包在单个NAND页中的系统和方法，这可

以减少设备哈希图中最坏情况的冲突长度。设备哈希图中的键可以被分组在一起以具有空间局部性,即使这些键本身没有时间局部性。

[0037] 图1示出了KV SSD中的哈希实现方式。

[0038] 如图1所示,键被哈希化并插入到哈希图中。然而,单个NAND页可能只包含一个键。例如,哈希图中的桶2f8包括Key1、Key34和Key21的链。Key1、Key34和Key21中的每一个都存储在单独的NAND页上。如上所述,这可能导致在放置/获取/存在期间更高的RAF,以及在删除期间更高的设备WAF。此外,最坏情况的尾延迟可以与哈希图的桶(例如,图1中的桶467)中最长的冲突链的长度成比例。

[0039] 根据本公开的实施例,可以将多个键一起打包在单个NAND页中,以通过打包因子P(每页经打包的键的平均数量)减少哈希图中的最坏情况的冲突长度。

[0040] 图2示出了根据实施例的KV SSD中的哈希实现方式。

[0041] 参考图2,可以将多个键一起打包在单个NAND页中,其中每个键包含范围指针(extent pointer)。

[0042] 根据本公开实施例的方法包括聚集不具有时间局部性的相关键,并将它们分组在一起以具有空间局部性。也就是说,在哈希图的连续桶内被索引的键可以被分组到桶组中。

[0043] 此外,可以修改KV条目的格式,使得值可以被写入NAND页,并且可以被范围图跟踪。如果在写入范围图后NAND页中有空间可用,则范围图本身可以被写入NAND页,该NAND页也可以包含值的一部分。键包含一个指向NAND页的范围指针,该NAND页包含实际的范围图,该实际的范围图又定位包含该键的值的页。

[0044] 当以因子P来处理获取/放置/删除操作时,可以显著减少设备WAF/RAF。例如,如果平均来说,4个键可以被打包到一个页中,则冲突链长度减少4倍,同时相应的设备WAF/RAF减少4倍。

[0045] 如上所述,键打包可以在KV存储内消耗最少额外资源的情况下实现。

[0046] 图3示出了根据实施例的KV存储。

[0047] 参考图3,KV存储300包括可以用于帮助键打包的键记录器301(例如,自动操作记录器(Auto Operation Logger,AOL))、跟踪键记录器301中的键的布隆过滤器(Bloom filter)303,以及在键记录器301中创建键的反向链表的尾指针表305。

[0048] 键记录器301可以是设备范围的,并且可以仅具有到其的附加操作。因此,它可以在内部实现,就像分区命名空间一样。

[0049] 布隆过滤器303跟踪键记录器301中的键,因此可以是紧凑的。

[0050] 例如,如果键记录器301包含10K个条目,则布隆过滤器303可以包括少至3个页(每个页对应4K RAM、4K SRAM和4K DRAM)。

[0051] 假设与KV存储中的全部键相比,少于0.01%的键被记录在键记录器301中,则大多数获取/放置/删除操作将不包括遍历键记录器301中的条目的尾指针表305。

[0052] 尾指针表305可以被构造为使得如果布隆过滤器303表示命中,则需要在键记录器301中遍历的条目的数量可以被限制为仅一个或两个条目。例如,如果键记录器301具有10K个条目,则尾指针表305可以具有5K个条目(每个指针仅16比特)。

[0053] 尽管图3将KV存储300的组件示为分离的元件,但是本公开不限于此。例如,这些组件中的至少两个组件可以组合成单个元件,诸如处理器、集成电路(IC)、片上系统(SoC)等,

其可以执行相应的操作。

[0054] 图4A和图4B示出了根据实施例的KV存储中的推迟的键打包操作。更具体地，图4A示出了在KV存储中处理推迟的键打包之前的示例，而图4B示出了在处理来自键记录器的键的推迟的打包之后的示例。推迟的键打包操作可以基于对KV的存储(Store)(或放置)和删除操作首先仅被写入键记录器。也就是说，不是立即执行接收到的存储和删除操作，而是可以将条目写入键记录器中，直到达到特定阈值并且可以执行该操作。例如，阈值可以是键记录器中条目的百分比和/或以其填充键记录器的速率的函数。

[0055] 参考图4A，桶组包括8个桶B1至B8，并且指向逻辑NAND页6878和6784，在这些页上，多个键与范围指针(ExtentP)一起被打包。在图4A所示的示例中，每个逻辑NAND页可以包括最多6个键，并且包括指向包含桶组的键的下一页的指针。

[0056] 在逻辑页6878上，Key1包含指向存储在页7上的范围图的范围指针。存储在页7上的范围图标识了其上存储对应于Key1的值的页。类似地，Key12包含指向存储在页480上的范围图的范围指针，Key2包含指向存储在页21上的范围图的范围指针，KeyA包含指向存储在页78上的范围图的范围指针，而Key5包含指向存储在页790上的范围图的范围指针。范围图中的每一个都标识其上存储相应键的值的页。

[0057] 此外，图4A提供了具有两个条目的链的键记录器的示例。假设达到了特定阈值，则可以处理键记录器中的条目，即，可以执行推迟的键打包。

[0058] 参考图4B，在被处理的链的条目中，存在针对Key1/ExtentP 240的存储条目和删除条目，它们基本上彼此抵消。

[0059] 也存在针对Key1/ExtentP 250的存储条目。因此，逻辑页6878接收包含指向其上存储了范围图的页250的范围指针的Key1的新条目，该范围图标识其上存储了对应于Key1的值的页。

[0060] 此外，从包括旧Key1的范围图的页7和逻辑页6878中移除旧Key1(如图4A所示)，并且由旧Key1的范围图所标识的页被排队等待擦除。

[0061] 类似地，存在针对Key2/ExtentP 310的存储条目。因此，逻辑页6878接收包含指向其上存储了范围图的页310的范围指针的Key2的新条目，该范围图标识其上存储了对应于Key2的值的页。

[0062] 此外，从包括旧Key2的范围图的页21和逻辑页6878中移除旧Key2(如图4A所示)，并且由旧Key2的范围图标识的页被排队等待擦除。

[0063] 也存在针对Key3/ExtentP 560的存储条目。因此，逻辑页6878接收包含指向其上存储了范围图的页560的范围指针的Key3的新条目，该范围图标识其上存储了对应于Key3的值的页。

[0064] 还存在针对Key12/ExtenP 480的删除条目，这导致Key12从包括Key12的范围图的页480和逻辑页6878中被移除，并且由Key12的范围图标识的页被排队等待擦除。

[0065] 类似地，存在针对Key5/ExtenP 480的删除条目，这导致Key5从包括Key5的范围图的页790和逻辑页6878中被移除，并且由Key5的范围图标识的页被排队等待擦除。

[0066] 因为从逻辑页6878移除了两个键(Key12和Key5)并且向逻辑页6878添加了一个新的键(Key3)，所以逻辑页6878上只剩下5个键。因此，将KeyX从逻辑页6784(如图4A所示)移动到逻辑页6878(如图4B所示)，使得逻辑页6878包括最多6个键。

[0067] 图5是示出根据实施例的KV存储的键打包操作的流程图。

[0068] 参考图5,在步骤501中,键记录器记录用于存储和删除操作的键条目。

[0069] 在步骤503中,在键记录器中满足特定阈值之后,例如,键记录器被填充超过设定的阈值,在步骤505中,收集(collate)属于设备哈希图中的一组相邻桶的键条目。

[0070] 在步骤507中,设备哈希图中的现有键与来自键记录器的条目合并,以形成每组桶的新的经打包的键。

[0071] 在步骤509中,将经打包的键写入至少一个NAND页,并且更新设备哈希图以指向包含经打包的键的页。

[0072] 在步骤511中,清除键记录器及其相关联的布隆过滤器,使得键记录器可以记录新的键条目。

[0073] 对对应于一组桶的键进行分组可以是O(1)复杂度的操作。

[0074] 此外,当桶组的相邻桶仅包含一个或两个键时,它们可能最终共享经打包的键NAND页。

[0075] 为了在处理键记录器时的放置/删除/获取操作的连续性,根据实施例的设备可以实现两个或更多个独立的键记录器。

[0076] 图6是示出根据实施例的KV存储的放置操作的流程图。

[0077] 参考图6,在步骤601中,在接收到用于存储KV的放置命令时,KV存储仅将KV的值写入NAND页。

[0078] 在步骤603中,KV存储创建跟踪包含KV的值的各种NAND页的位置的范围图页。

[0079] 在步骤605中,更新键记录器的布隆过滤器。例如,可以更新布隆过滤器,使得预定数量的已知哈希函数可以用于设置布隆过滤器比特图中的比特。

[0080] 在步骤607中,基于KV的键的哈希在尾指针表中选择槽(slot)。例如,尾指针表可以包含K/2个条目,其中K=键记录器中的条目总数。

[0081] 在步骤609中,KV存储标记(note)尾指针表的所选槽的内容(即尾条目),并将键记录器中的当前偏移写入尾指针表的同一槽。基本上,尾指针表可以是键记录器中存在的条目的哈希图,使得哈希图的每个桶指针指向映射到桶的最新条目。尾指针表可以减少键记录器中要被遍历以找到由相关联的布隆过滤器指示可能存在于键记录器中的特定键的槽的数量。

[0082] 在步骤611中,KV存储创建键、包含范围图的NAND页、操作码和尾条目指针的四路元组,并将该元组附加到键记录器。尽管图6利用了键的四路元组,但是本公开不限于此,并且可以使用键的不同尺寸的元组。

[0083] 在步骤613中,在足够数量的键被记录到键记录器中之后,KV存储更新KV的哈希图。

[0084] 图7是示出根据实施例的KV存储的删除操作的流程图。

[0085] 参考图7,在步骤701中,在接收到用于从KV存储中删除KV的删除命令时,KV存储在跟踪键记录器中的键条目的布隆过滤器中检查要删除的键的存在。

[0086] 当在步骤703中键记录器的布隆过滤器建议针对该键的命中时,在步骤705中,KV存储遍历键记录器中的键的哈希图,以检查匹配的键名(keyname)的存在。例如,如果接收到的删除命令用于删除Key33的KV,则KV确定键记录器是否已经包括Key33的条目。

[0087] 当在步骤707在键记录器中存在匹配的键名时,操作进行到步骤711。

[0088] 然而,当在步骤703中键记录器的布隆过滤器不建议针对该键的命中时、或者当在步骤707中键记录器中没有匹配的键名时,在步骤709中,KV存储确定该键是否存在于设备哈希图中。

[0089] 当在步骤709中该键不存在于设备哈希图中时,操作结束,因为没有可删除的键。

[0090] 然而,当在步骤709中该键存在于设备哈希图中时,操作进行到步骤711。

[0091] 在步骤711中,KV存储创建删除键条目,并将其附加到键记录器。例如,删除键条目元组包含要删除的键的指示、包含键的范围图的NAND页的地址(例如,从键记录器中键的先前条目读取的)、删除操作码和尾指针表中的当前条目。

[0092] 在步骤713中,KV存储更新尾指针表中的键条目的地址,以反映新的删除键条目。

[0093] 在步骤715中,KV存储推迟从设备哈希图中对键的实际删除。也就是说,KV存储等待执行删除操作,直到在键记录器内达到特定阈值为止,如上面参考图4A和图4B所述的。

[0094] 图8是示出根据实施例的KV存储的获取操作的流程图。

[0095] 参考图8,在步骤801中,在接收到用于检索存储的KV的获取命令时,KV存储在跟踪键记录器中的键条目的布隆过滤器中检查要检索的键的存在。

[0096] 当在步骤803中键记录器的布隆过滤器建议针对该键的命中时,在步骤805中,KV存储遍历键记录器中的键的哈希图,以检查匹配的键名的存在。例如,如果接收到的获取命令用于检索Key77的KV,则KV存储确定键记录器是否已经包括Key77的条目。

[0097] 当在步骤807中在键记录器中存在匹配的键名时,在步骤815中,KV存储确定键记录器中的匹配的键名是否包括用于删除的操作码。

[0098] 当在步骤815中键记录器中的匹配的键名包括用于删除的操作码时,在步骤817中,获取操作失败。例如,如果获取命令用于检索Key88的KV,但是键记录器已经包括删除Key88的KV的命令条目,则检索操作失败。

[0099] 然而,当在步骤815中键记录器中的匹配的键名不包括用于删除的操作码时,例如,匹配的键名包括用于存储的操作码,操作进行到步骤811。

[0100] 当在步骤803中键记录器的布隆过滤器不建议针对该键的命中时、或者当在步骤807中键记录器中没有匹配的键名时,在步骤809中,KV存储确定该键是否存在于设备哈希图中。

[0101] 当在步骤809中该键不存在于设备哈希图中时,在步骤813中,获取操作失败,因为没有可用于检索的键。

[0102] 然而,当在步骤809中该键存在于设备哈希图中时,操作进行到步骤811。

[0103] 在步骤811中,KV存储读取包含该键的范围图的NAND页,然后从范围图所标识的一个或多个NAND页中读取该键的值。如上所述,因为设备哈希图存储可以在单个NAND页上存储多个键和范围指针,这些范围指针指向包含键的范围图的NAND页,所以KV存储能够快速地标识包含键的范围图的NAND页,然后从范围图所标识的一个或多个NAND页中读取键的值。

[0104] 图9示出了根据一个实施例的网络环境900中的电子设备901的框图。

[0105] 参考图9,网络环境900中的电子设备901可以经由第一网络998(例如,短距离无线通信网络)与电子设备902通信、或者经由第二网络999(例如,长距离无线通信网络)与电子

设备904或服务器908通信。电子设备901可以经由服务器908与电子设备904通信。电子设备901可以包括处理器920、存储器930、输入设备950、声音输出设备955、显示设备960、音频模块970、传感器模块976、接口977、触觉模块979、相机模块980、电力管理模块988、电池989、通信模块990、订户标识模块(SIM)996或天线模块997。在一个实施例中，可以从电子设备901中省略至少一个组件(例如，显示设备960或相机模块980)、或者可以向电子设备901添加一个或多个其他组件。在一个实施例中，组件中的一些组件可以被实现为单个集成电路(IC)。例如，传感器模块976(例如，指纹传感器、虹膜传感器或照度传感器)可以嵌入显示设备960(例如，显示器)中。

[0106] 处理器920可以执行例如软件(例如，程序940)来控制与处理器920耦合的电子设备901的至少一个其他组件(例如，硬件或软件组件)，并且可以执行各种数据处理或计算。作为数据处理或计算的至少一部分，处理器920可以将从另一组件(例如，传感器模块976或通信模块990)接收到的命令或数据加载到易失性存储器932中，处理存储在易失性存储器932中的命令或数据，并将所得数据存储在非易失性存储器934中。处理器920可以包括主处理器921(例如，中央处理单元(CPU)或应用处理器(AP))和辅助处理器923(例如，图形处理单元(GPU)、图像信号处理器(ISP)、传感器集线器处理器或通信处理器(CP))，该辅助处理器923可以独立于主处理器921或与主处理器921结合操作。附加地或替代地，辅助处理器923可以适于消耗比主处理器921更少的功率或者执行特定的功能。辅助处理器923可以被实现为独立于主处理器921或者是主处理器921的一部分。

[0107] 当主处理器921处于不活动(例如，睡眠)状态时，辅助处理器923可以代替主处理器921、或者当主处理器921处于活动状态(例如，执行应用)时，辅助处理器923可以与主处理器921一起，来控制与电子设备901的组件中的至少一个组件(例如，显示设备960、传感器模块976或通信模块990)相关的至少一些功能或状态。根据一个实施例，辅助处理器923(例如，图像信号处理器或通信处理器)可以被实现为功能上与辅助处理器923相关的另一组件(例如，相机模块980或通信模块990)的一部分。

[0108] 存储器930可以存储被电子设备901的至少一个组件(例如，处理器920或传感器模块976)使用的各种数据。各种数据可以包括例如软件(例如，程序940)和与其相关的命令的输入数据或输出数据。存储器930可以包括易失性存储器932或非易失性存储器934。

[0109] 程序940可以作为软件存储在存储器930中，并且可以包括例如操作系统(OS)942、中间件944或应用946。

[0110] 输入设备950可以从电子设备901的外部(例如，用户)接收将被电子设备901的其他组件(例如，处理器920)使用的命令或数据。输入设备950可以包括例如麦克风、鼠标或键盘。

[0111] 声音输出设备955可以向电子设备901的外部输出声音信号。声音输出设备955可以包括例如扬声器或接收器。扬声器可以用于一般目的，诸如播放多媒体或录音，而接收器可以用于接收来电。根据一个实施例，接收器可以被实现为与扬声器分离或者是扬声器的一部分。

[0112] 显示设备960可以向电子设备901的外部(例如，用户)可视地提供信息。显示设备960可以包括例如显示器、全息设备或投影仪以及控制显示器、全息设备和投影仪中相应的一个的控制电路。根据一个实施例，显示设备960可以包括适于检测触摸的触摸电路、或者

适于测量由触摸引起的力的强度的传感器电路(例如,压力传感器)。

[0113] 音频模块970可以将声音转换成电信号,反之亦然。根据一个实施例,音频模块970可以经由输入设备950获得声音、或者经由声音输出设备955或与电子设备901直接(例如,有线)或无线耦合的外部电子设备902的耳机输出声音。

[0114] 传感器模块976可以检测电子设备901的操作状态(例如,电力或温度)或电子设备901外部的环境状态(例如,用户的状态),然后生成与检测到的状态相对应的电信号或数据值。传感器模块976可以包括例如手势传感器、陀螺仪传感器、大气压力传感器、磁传感器、加速度传感器、抓握传感器、接近传感器、颜色传感器、红外(IR)传感器、生物测定传感器、温度传感器、湿度传感器或照度传感器。

[0115] 接口977可以支持将用于电子设备901直接(例如,有线)或无线地与外部电子设备902耦合的一个或多个指定协议。根据一个实施例,接口977可以包括例如高清晰度多媒体接口(HDMI)、通用串行总线(USB)接口、安全数字(SD)卡接口或音频接口。

[0116] 连接端子978可以包括电子设备901可以经由其与外部电子设备902物理连接的连接器。根据一个实施例,连接端子978可以包括例如HDMI连接器、USB连接器、SD卡连接器或音频连接器(例如,耳机连接器)。

[0117] 触觉模块979可以将电信号转换成可以由用户经由触觉或动觉来辨识的机械刺激(例如,振动或移动)或电刺激。根据一个实施例,触觉模块979可以包括例如马达、压电元件或电刺激器。

[0118] 相机模块980可以捕获静止图像或运动图像。根据一个实施例,相机模块980可以包括一个或多个镜头、图像传感器、图像信号处理器或闪光灯。

[0119] 电力管理模块988可以管理供应到电子设备901的电力。电力管理模块988可以被实现为例如电力管理集成电路(PMIC)的至少一部分。

[0120] 电池989可以向电子设备901的至少一个组件供电。根据一个实施例,电池989可以包括例如不可充电的原电池、可充电的二次电池或燃料电池。

[0121] 通信模块990可以支持在电子设备901和外部电子设备(例如,电子设备902、电子设备904或服务器908)之间建立直接(例如,有线)通信信道或无线通信信道,并且经由建立的通信信道执行通信。通信模块990可以包括一个或多个通信处理器,其可以独立于处理器920(例如,AP)进行操作并且支持直接(例如,有线)通信或无线通信。根据一个实施例,通信模块990可以包括无线通信模块992(例如,蜂窝通信模块、短程无线通信模块或全球导航卫星系统(GNSS)通信模块)或有线通信模块994(例如,局域网(LAN)通信模块或电力线通信(PLC)模块)。这些通信模块中相应的一个可以经由第一网络998(例如,短程通信网络,诸如Bluetooth™、无线保真(Wi-Fi)直连或红外数据协会(IrDA)的标准)或第二网络999(例如,远程通信网络,诸如蜂窝网络、互联网或计算机网络(例如,LAN或广域网(WAN))与外部电子设备通信。这些不同类型的通信模块可以被实现为单个组件(例如,单个IC)或者可以被实现为彼此分离的多个组件(例如,多个IC)。无线通信模块992可以使用存储在订户标识模块996中的订户信息(例如,国际移动订户标识(IMSI))来标识和认证诸如第一网络998或第二网络999的通信网络中的电子设备901。

[0122] 天线模块997可以向电子设备901的外部(例如,外部电子设备)发送信号或电力、或者从电子设备901的外部(例如,外部电子设备)接收信号或电力。根据一个实施例,天线

模块997可以包括一个或多个天线，并且，例如，可以由通信模块990(例如，无线通信模块992)从其中选择适合于在诸如第一网络998或第二网络999的通信网络中使用的通信方案的至少一个天线。然后可以经由所选择的至少一个天线在通信模块990和外部电子设备之间发送或接收信号或电力。

[0123] 上述组件中的至少一些可以相互耦合，并经由外围设备间通信方案(例如，总线、通用输入和输出(GPIO)、串行外围接口(SPI)或移动工业处理器接口(MIPI))在它们之间传送信号(例如，命令或数据)。

[0124] 根据一个实施例，可以经由与第二网络999耦合的服务器908在电子设备901和外部电子设备904之间发送或接收命令或数据。电子设备902和904中的每一个可以是与电子设备901相同类型或不同类型的设备。要在电子设备901处执行的所有或一些操作可以在一个或多个外部电子设备902、904或908处执行。例如，如果电子设备901应该自动地或者响应于来自用户或另一设备的请求来执行功能或服务，则代替执行功能或服务或者除了执行功能或服务之外，电子设备901可以请求一个或多个外部电子设备来执行功能或服务的至少一部分。接收请求的一个或多个外部电子设备可以执行所请求的功能或服务的至少一部分或者与请求相关的附加功能或附加服务，并将执行的结果传送到电子设备901。电子设备901可以在对结果进行或不进行进一步处理的情况下提供结果，作为对请求的答复的至少一部分。为此，例如，可以使用云计算、分布式计算或客户端-服务器计算技术。

[0125] 一个实施例可以被实现为软件(例如，程序940)，该软件包括存储在机器(例如，电子设备901)可读的存储介质(例如，内部存储器936或外部存储器938)中的一个或多个指令。例如，电子设备901的处理器可以在处理器的控制下调用存储在存储介质中的一个或多个指令中的至少一个，并且在使用或不使用一个或多个其它组件的情况下执行该一个或多个指令中的至少一个。因此，可以操作机器来根据所调用的至少一个指令执行至少一个功能。一个或多个指令可以包括由编译器生成的代码或解释器可执行的代码。可以以非暂时性存储介质的形式提供机器可读存储介质。术语“非暂时性”表示存储介质是有形设备，并且不包括信号(例如电磁波)，但是该术语不区分数据半永久性地存储在存储介质中的情况和数据临时存储在存储介质中的情况。

[0126] 根据一个实施例，可以在计算机程序产品中包括并提供本公开的方法。计算机程序产品可以作为产品在卖方和买方之间交易。计算机程序产品可以以机器可读存储介质(例如，光盘只读存储器(CD-ROM))的形式分发、或经由应用商店(例如，PlayStore™)在线分发(例如，下载或上传)或者直接在两个用户设备(例如，智能电话)之间分发。如果在线分发，则计算机程序产品的至少一部分可以临时生成或至少临时存储在机器可读存储介质(诸如制造商的服务器、应用存储的服务器或中继服务器的存储器)中。

[0127] 根据一个实施例，上述组件中的每个组件(例如，模块或程序)可以包括单个实体或多个实体。可以省略上述组件中的一个或多个、或者可以添加一个或多个其他组件。可替代地或附加地，多个组件(例如，模块或程序)可以集成到单个组件中。在这种情况下，集成的组件仍然可以以与集成之前由多个组件中相应的一个组件执行的相同或相似的方式来执行多个组件中的每个组件的一个或多个功能。由模块、程序或另一组件执行的操作可以顺序地、并行地、重复地或启发式地执行、或一个或多个操作可以以不同的顺序执行或省略或者可以添加一个或多个其他操作。

[0128] 图10示出了根据实施例的存储系统1000的图。

[0129] 存储系统1000包括主机1002和存储设备1004。尽管描述了一个主机和一个存储设备，但是存储系统1000可以包括多个主机和/或多个存储设备。存储设备1004可以是SSD、通用闪存(UFS)等。存储设备1004包括控制器1006和连接到控制器1006的存储介质1008。控制器1006可以是SSD控制器、UFS控制器等。存储介质1008可以包括易失性存储器、非易失性存储器或两者，并且可以包括一个或多个闪存芯片(或其他存储介质)。控制器1006可以包括一个或多个处理器、一个或多个纠错电路、一个或多个现场可编程门阵列(FPGA)、一个或多个主机接口、一个或多个闪存总线接口等或它们的组合。控制器1006可以被配置为便于主机1002和存储介质1008之间的数据/命令传送。主机1002向存储设备1004发送数据/命令，以由控制器1006接收并结合存储介质1008进行处理。如本文所述，方法、过程和算法可以在存储设备控制器(诸如控制器1006)上实现。本文描述的源和目的地可以对应于主机1002(即，处理器或应用)和存储介质1008的元件。

[0130] 根据上述实施例，提供了一种系统和方法，其允许将多个键一起打包在单个NAND页中，以通过打包因子P(即，每页经打包的键的平均数量)减少设备哈希图中的最坏情况的冲突长度。

[0131] 此外，不具有时间局部性的相关键可以被分组在一起以具有空间局部性。也就是说，映射到设备哈希图中相邻桶的键可以被分组在一起，即使这些键本身没有时间局部性。

[0132] 因此，由于密集的键打包，存储所有键所需的NAND页的数量可以减少，并且RAF/WAF可以减少。此外，NAND页可以减少当页中的所有键都无效时的GC。

[0133] 因此，本公开可以通过打包密度的因子减少GC开销显著，并且减少哈希图冲突链的长度，这改善了尾延迟，并且缩减了NAND页读取的数量，显著降低了查找延迟。

[0134] 尽管在本公开的详细描述中已经描述了本公开的特定实施例，但是在不脱离本公开的范围的情况下，可以以各种形式修改本公开。因此，本公开的范围不应仅基于所描述的实施例来确定，而是基于所附权利要求及其等同物来确定。

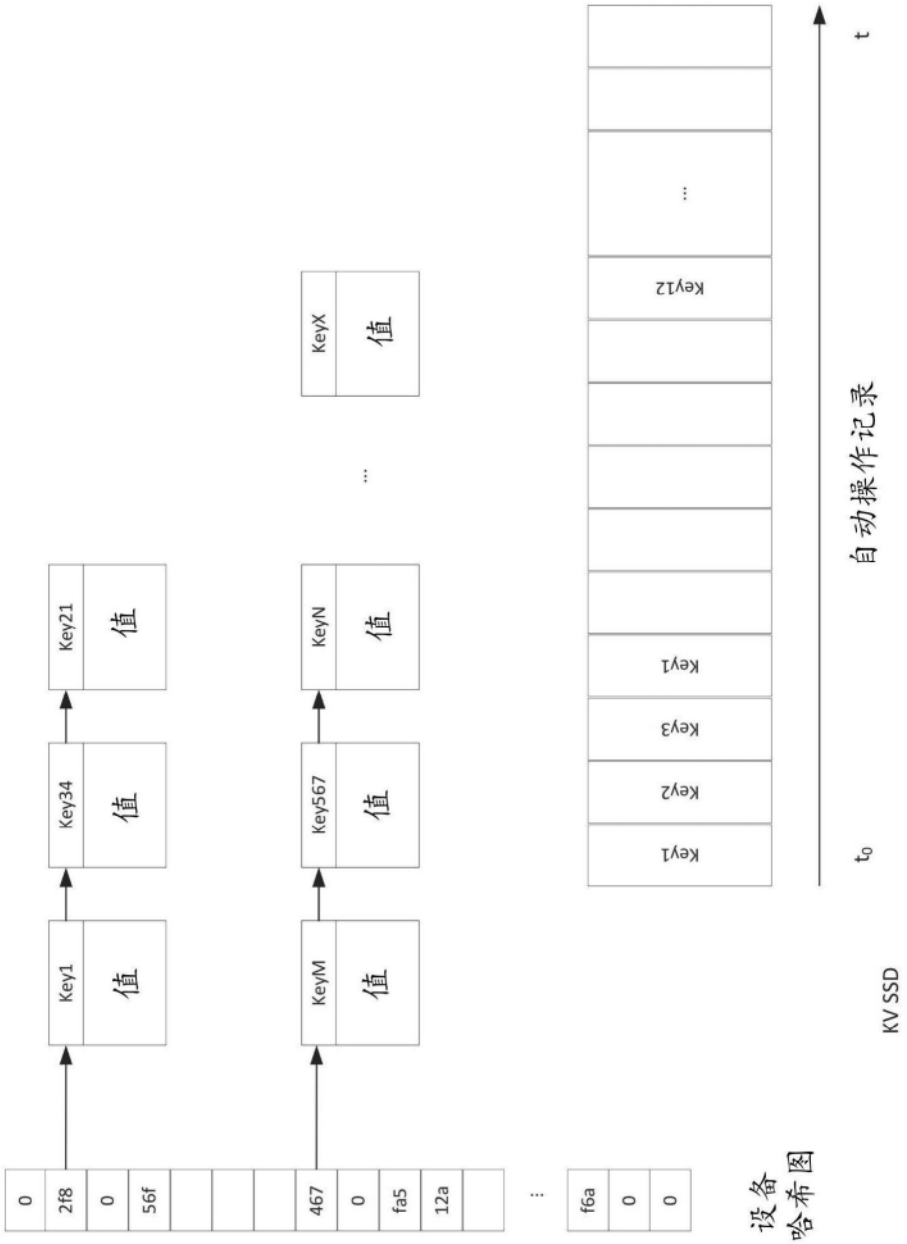

设备
哈希图
0
2f8
0
56f
467
0
fa5
12a
f6a
0
0
Key1
值
Key21
值
Key34
值
KeyM
值
KeyN
值
Key567
值
KeyX
值
KV SSD
自动操作记录
t0
t
key1
key2
key3
key1
key12

图1

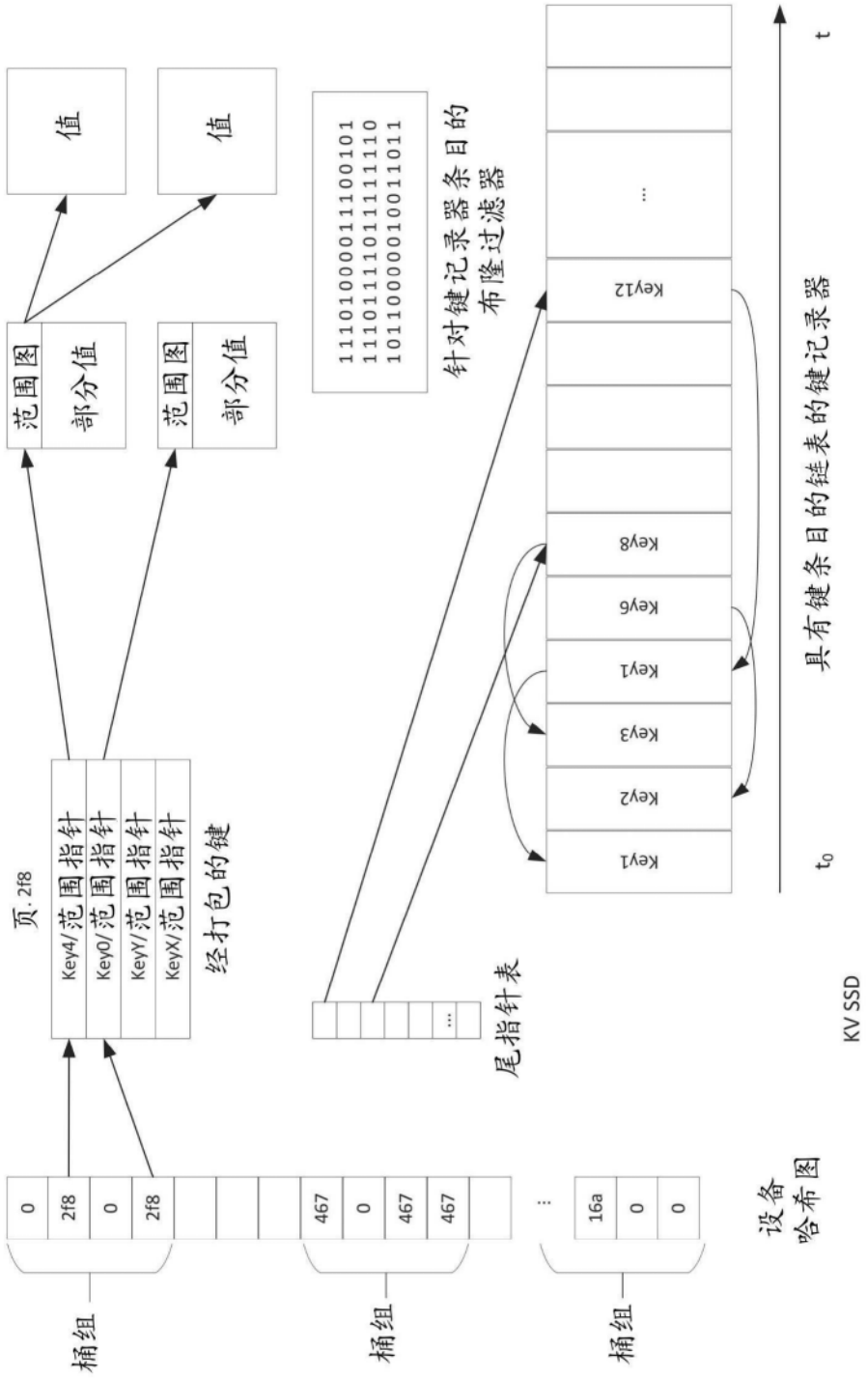

值
值
范围图
部分值
范围图
部分值
页. 2f8
Key4/范围指针
KeyO/范围指针
KeyY/范围指针
KeyX/范围指针
经打包的键
桶组
桶组
桶组
0
2f8
0
2f8
467
0
467
467
16a
0
0
1110100001110010 1
1110111101111111 0
1011000001001101 1
针对键记录器条目的
布隆过滤器
尾指针表
...
Key12
Key8
Key6
Key1
Key3
Key2
Key1
t_0
t
具有键条目的链表的键记录器
KV SSD
设备
哈希图

图2

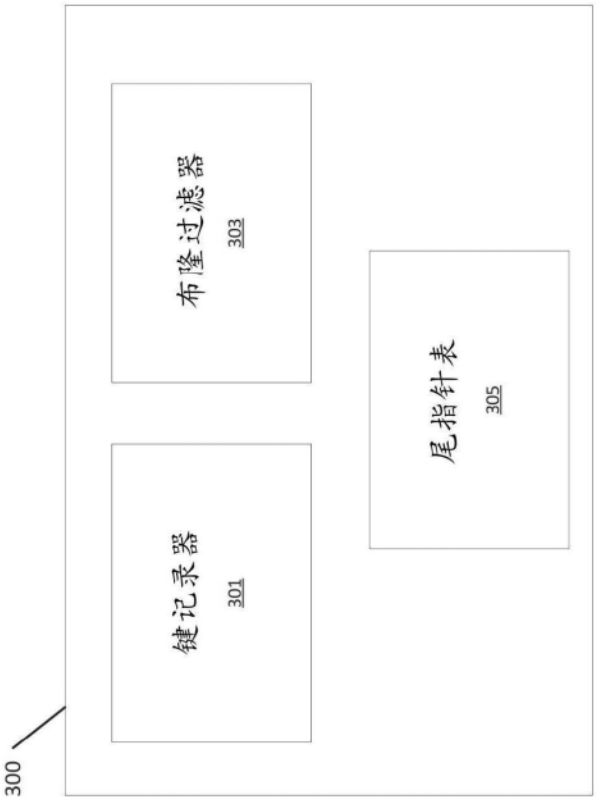


图3

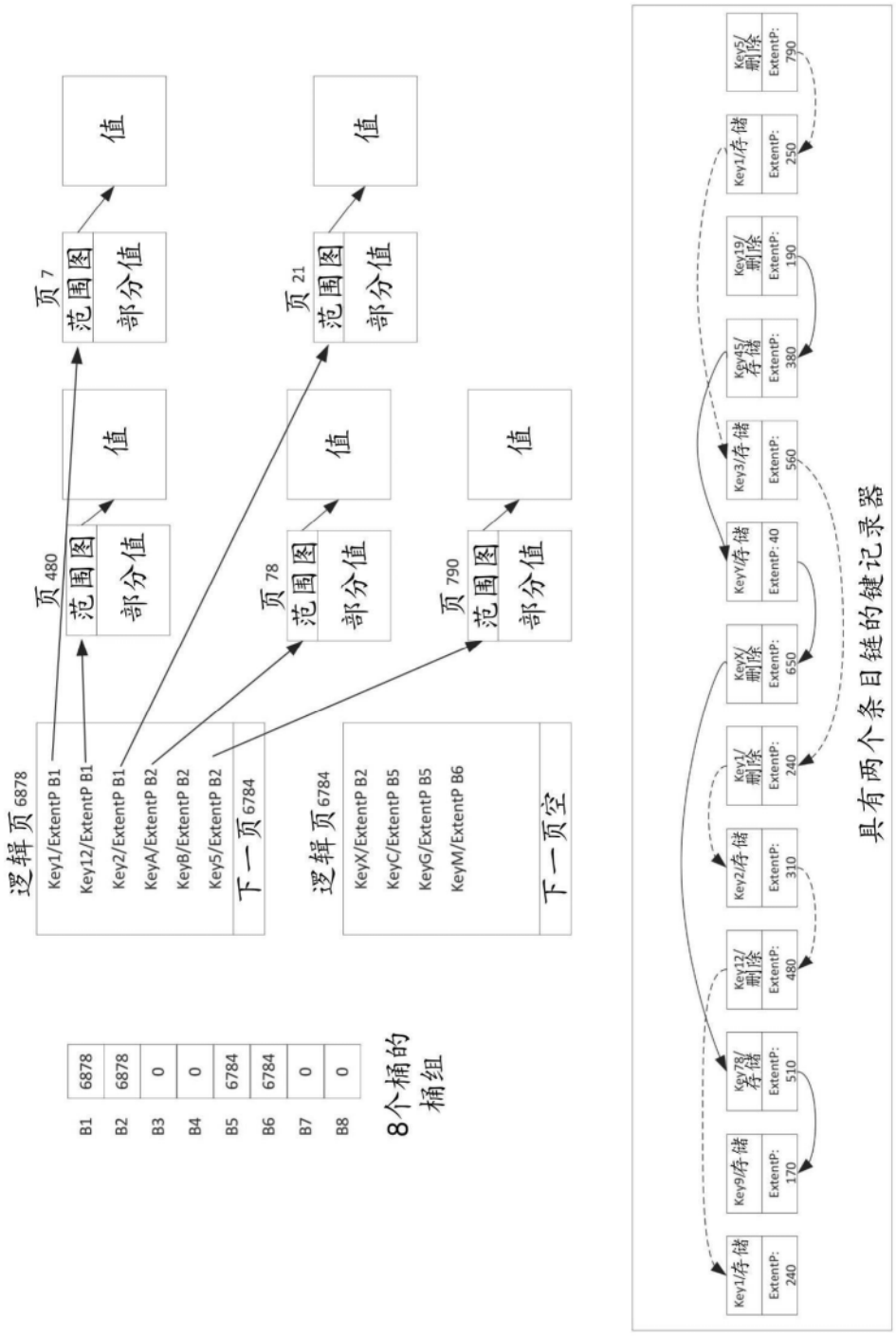

逻辑页 6878
Key1/ExtentP B1
Key12/ExtentP B1
Key2/ExtentP B1
KeyA/ExtentP B2
KeyB/ExtentP B2
Key5/ExtentP B2
下一页 6784
逻辑页 6784
KeyX/ExtentP B2
KeyC/ExtentP B5
KeyG/ExtentP B5
KeyM/ExtentP B6
下一页空
页 480
页 78
页 790
页 7
页 21
范围图
部分值
值
B1 6878
B2 6878
B3 0
B4 0
B5 6784
B6 6784
B7 0
B8 0
8个桶的桶组
具有两个条目链的键记录器
Key1/存储 ExtentP: 240
Key9/存储 ExtentP: 170
Key78/存储 ExtentP: 510
Key12/删除 ExtentP: 480
Key2/存储 ExtentP: 310
Key1/删除 ExtentP: 240
KeyX/删除 ExtentP: 650
KeyY/存储 ExtentP: 40
Key3/存储 ExtentP: 560
Key45/存储 ExtentP: 380
Key19/删除 ExtentP: 190
Key1/存储 ExtentP: 250
Key5/删除 ExtentP: 790

图4A

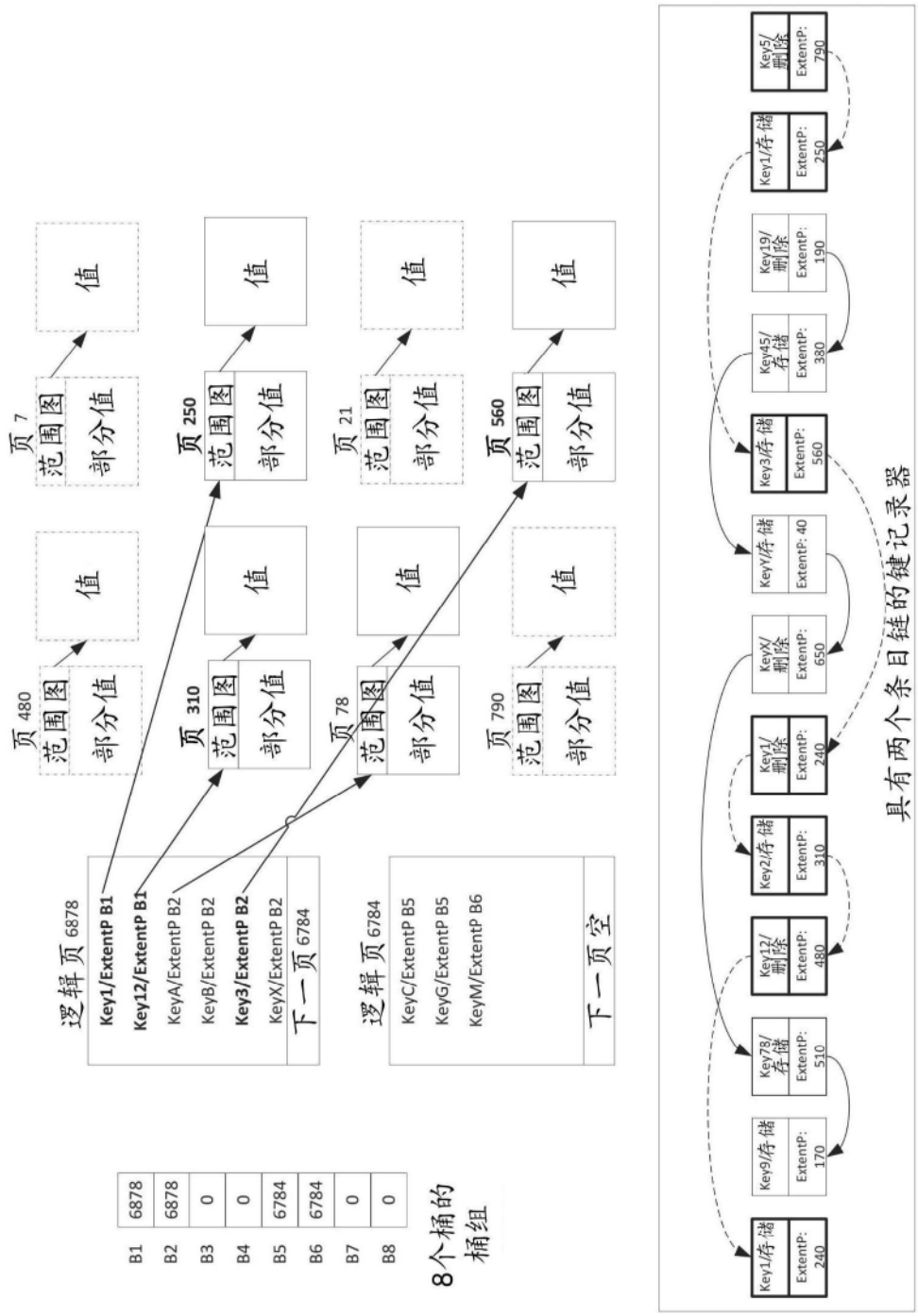

逻辑 页 6878
Key1/ExtentP B1
Key12/ExtentP B1
KeyA/ExtentP B2
KeyB/ExtentP B2
Key3/ExtentP B2
KeyX/ExtentP B2
下一 页 6784
逻辑 页 6784
KeyC/ExtentP B5
KeyG/ExtentP B5
KeyM/ExtentP B6
下一页空
页 480
范围图
部分值
值
页 310
页 78
页 790
页 7
页 250
页 21
页 560
B1 6878
B2 6878
B3 0
B4 0
B5 6784
B6 6784
B7 0
B8 0
8个桶的桶组
具有两个条目链的键记录器
Key1/存储 ExtentP: 240
Key9/存储 ExtentP: 170
Key73/存储 ExtentP: 510
Key12/删除 ExtentP: 480
Key2/存储 ExtentP: 310
Key1/删除 ExtentP: 240
KeyX/删除 ExtentP: 650
KeyY/存储 ExtentP: 40
Key3/存储 ExtentP: 560
Key45/存储 ExtentP: 380
Key19/删除 ExtentP: 190
Key1/存储 ExtentP: 250
Key5/删除 ExtentP: 790

图4B

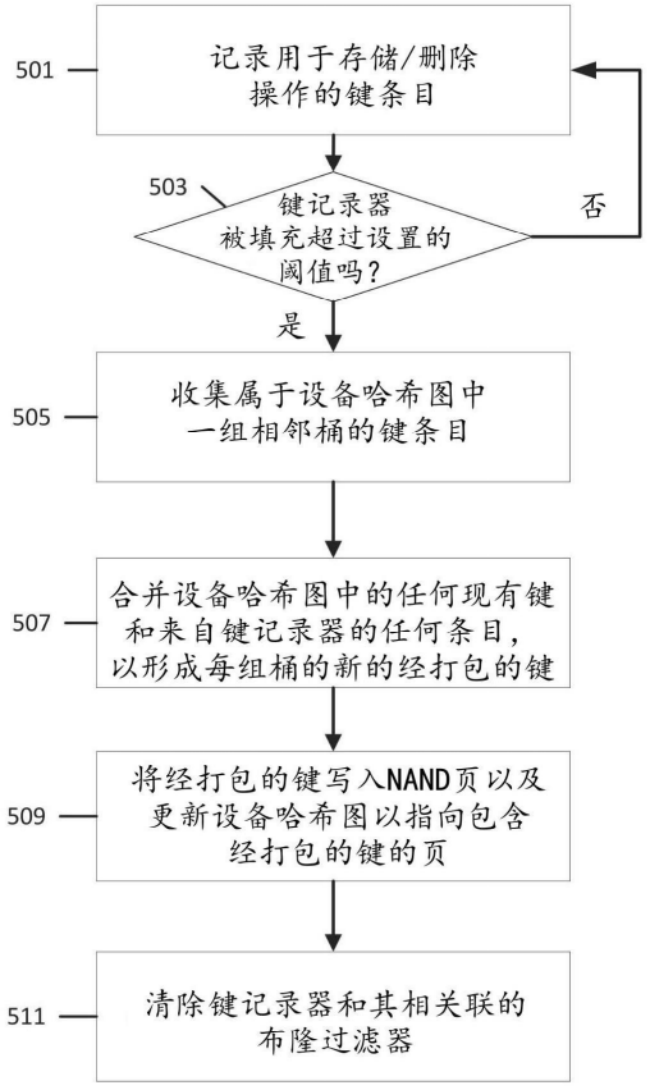

501
记录用于存储/删除操作的键条目
503
键记录器被填充超过设置的阈值吗?
否
是
505
收集属于设备哈希图中一组相邻桶的键条目
507
合并设备哈希图中的任何现有键和来自键记录器的任何条目,以形成每组桶的新的经打包的键
509
将经打包的键写入NAND页以及更新设备哈希图以指向包含经打包的键的页
511
清除键记录器和其相关联的布隆过滤器

图5

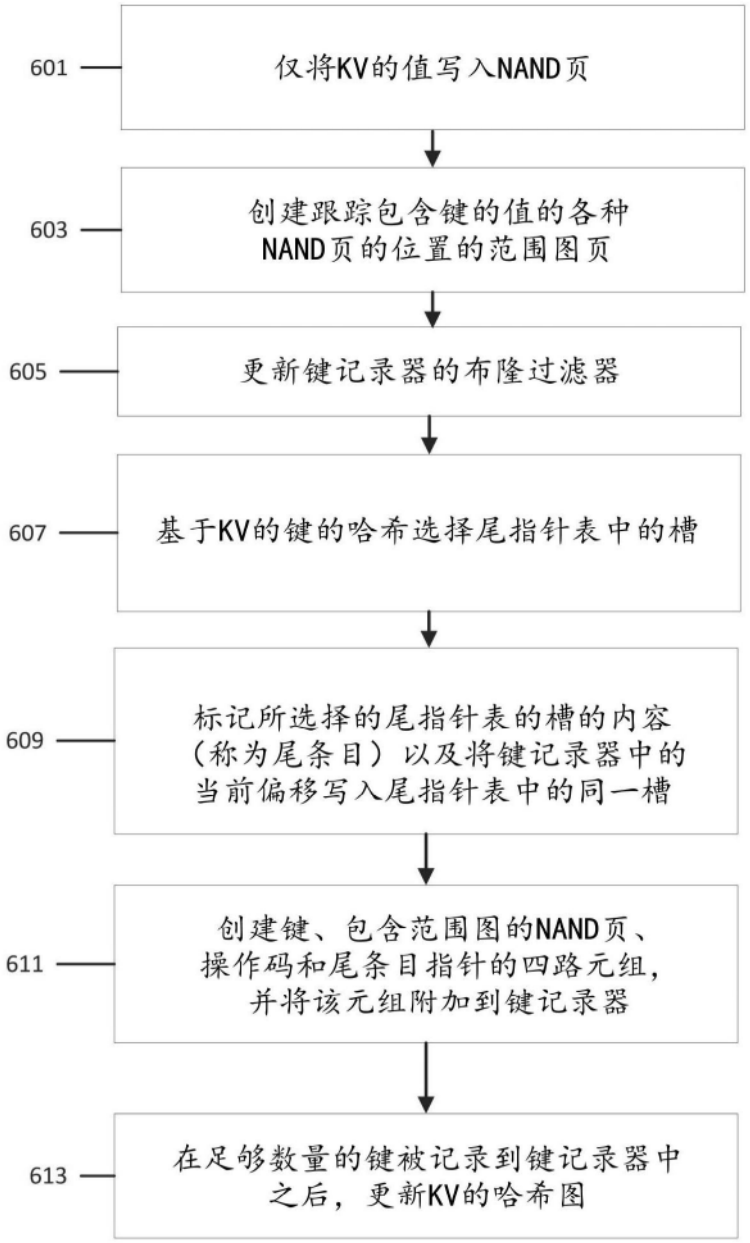

601
仅将KV的值写入NAND页
603
创建跟踪包含键的值的各种NAND页的位置的范围图页
605
更新键记录器的布隆过滤器
607
基于KV的键的哈希选择尾指针表中的槽
609
标记所选择的尾指针表的槽的内容（称为尾条目）以及将键记录器中的当前偏移写入尾指针表中的同一槽
611
创建键、包含范围图的NAND页、操作码和尾条目指针的四路元组，并将该元组附加到键记录器
613
在足够数量的键被记录到键记录器中之后，更新KV的哈希图

图6

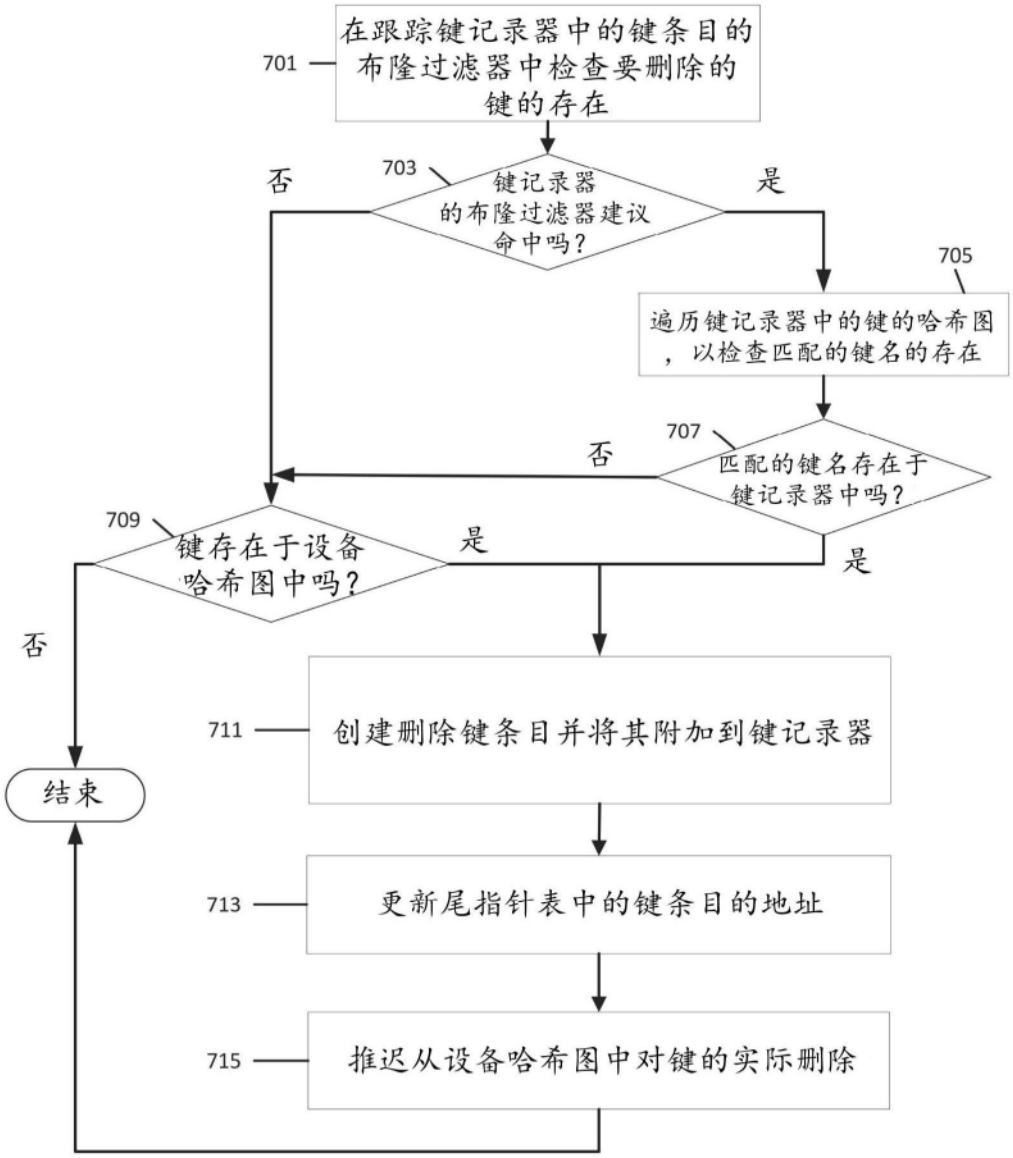

在跟踪键记录器中的键条目的布隆过滤器中检查要删除的键的存在
701
703
键记录器的布隆过滤器建议命中吗?
否
是
705
遍历键记录器中的键的哈希图，以检查匹配的键名的存在
707
匹配的键名存在于键记录器中吗?
否
是
709
键存在于设备哈希图中吗?
是
否
结束
711
创建删除键条目并将其附加到键记录器
713
更新尾指针表中的键条目的地址
715
推迟从设备哈希图中对键的实际删除

图7

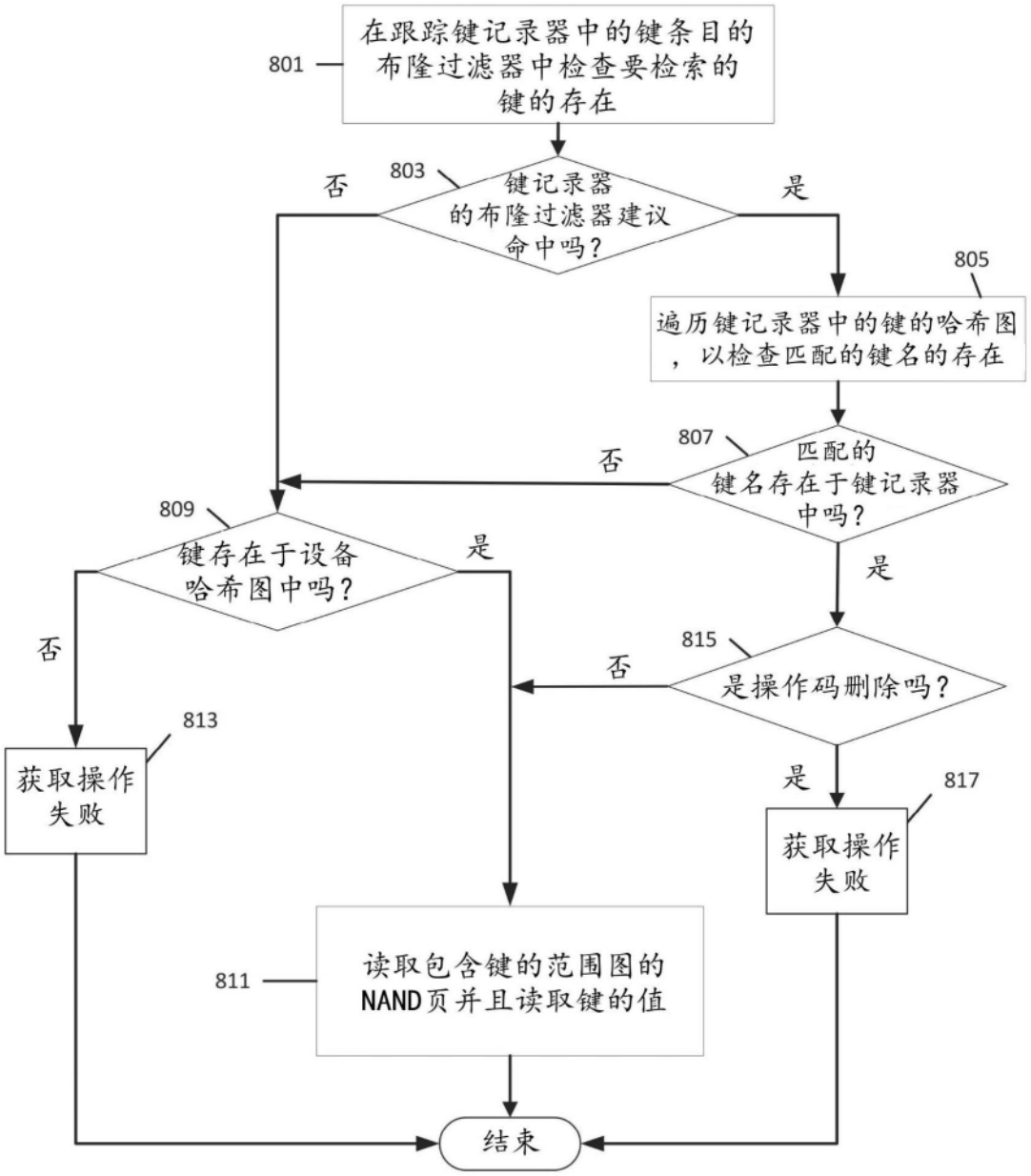

801
在跟踪键记录器中的键条目的布隆过滤器中检查要检索的键的存在
803
键记录器的布隆过滤器建议命中吗?
否
是
805
遍历键记录器中的键的哈希图，以检查匹配的键名的存在
807
匹配的键名存在于键记录器中吗?
否
是
809
键存在于设备哈希图中吗?
是
否
815
是操作码删除吗?
否
是
813
获取操作失败
817
获取操作失败
811
读取包含键的范围图的NAND页并且读取键的值
结束

图8

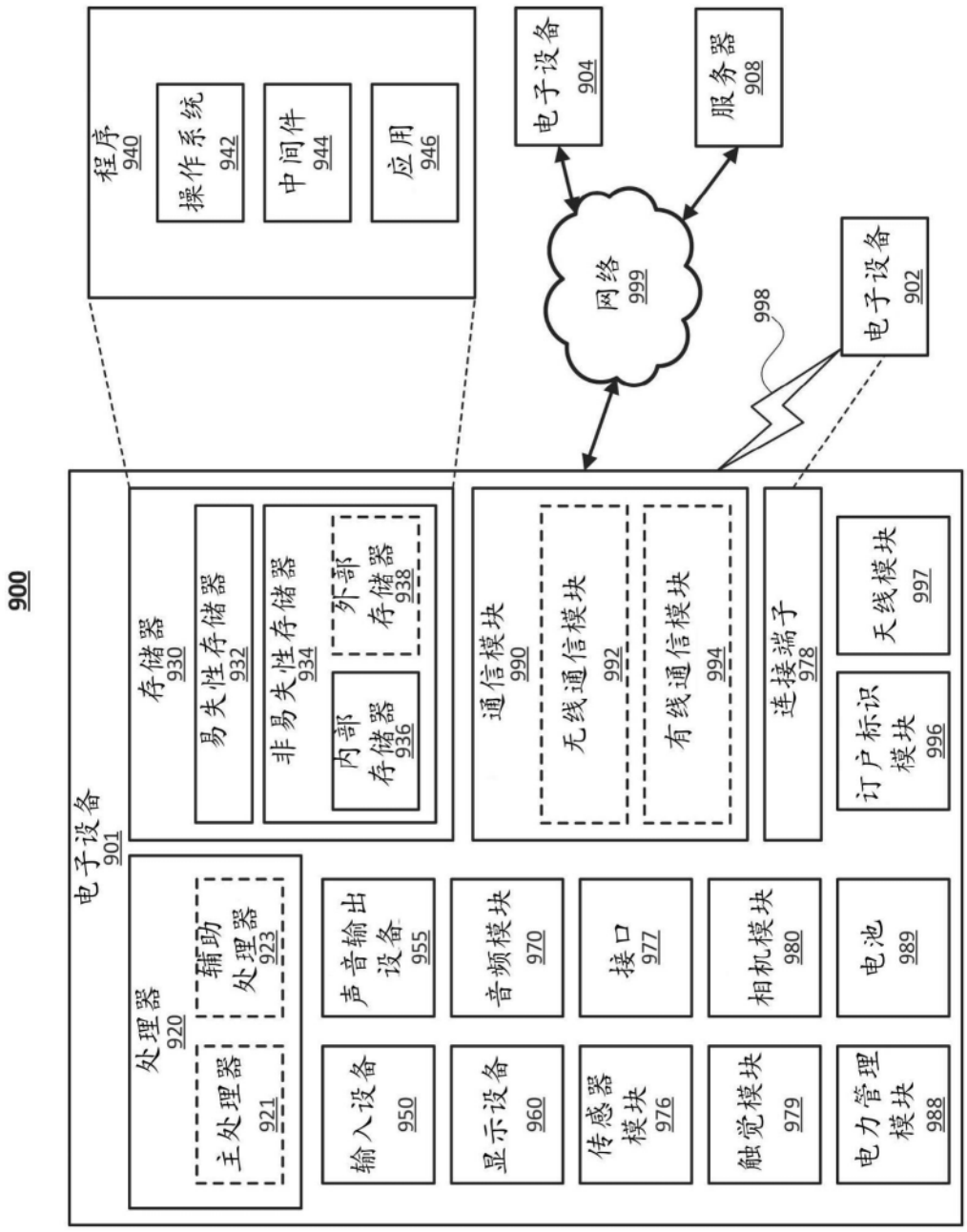

900
电子设备 901
处理器 920
主处理器 921
辅助处理器 923
存储器 930
易失性存储器 932
非易失性存储器 934
内部存储器 936
外部存储器 938
通信模块 990
无线通信模块 992
有线通信模块 994
连接端子 978
天线模块 997
订户标识模块 996
输入设备 950
声音输出设备 955
显示设备 960
音频模块 970
传感器模块 976
接口 977
触觉模块 979
相机模块 980
电力管理模块 988
电池 989
程序 940
操作系统 942
中间件 944
应用 946
电子设备 904
服务器 908
网络 999
998
电子设备 902

图9

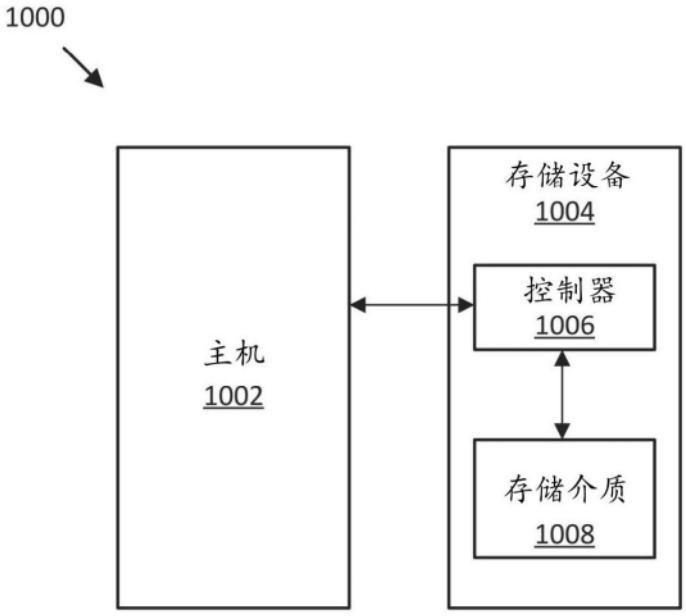


图10