



(12) 发明专利申请

(10) 申请公布号 CN 103473266 A

(43) 申请公布日 2013. 12. 25

(21) 申请号 201310345298. 5

(22) 申请日 2013. 08. 09

(71) 申请人 记忆科技(深圳)有限公司

地址 518000 广东省深圳市南山区蛇口后海
大道东角头东南工贸大厦 5 楼

(72) 发明人 雷伟

(74) 专利代理机构 北京律诚同业知识产权代理
有限公司 11006

代理人 刘健 黄韧敏

(51) Int. Cl.

G06F 17/30(2006. 01)

G06F 3/06(2006. 01)

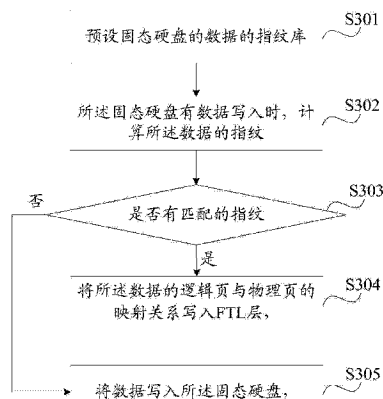
权利要求书2页 说明书5页 附图3页

(54) 发明名称

固态硬盘及其删除重复数据的方法

(57) 摘要

本发明适用于存储技术领域,提供了一种固态硬盘删除重复数据的方法,所述方法包括:预设所述固态硬盘的数据的指纹库,所述指纹库储存有所述固态硬盘数据的指纹;所述固态硬盘有数据写入时,计算所述数据的指纹;将所述数据的指纹与所述指纹库匹配处理;若所述指纹库中有匹配的指纹,则将所述数据的逻辑页与物理页的映射关系写入 FTL 层;若所述指纹库中没有匹配的指纹,则将所述数据写入所述固态硬盘。本发明还提供一种实现上述方法的固态硬盘。借此,本发明能延长固态硬盘的使用寿命,且能大大提高固态硬盘的空间利用率及速率。



1. 一种固态硬盘删除重复数据的方法,其特征在于,所述方法包括:
预设所述固态硬盘的数据的指纹库,所述指纹库储存有所述固态硬盘数据的指纹;
所述固态硬盘有数据写入时,计算所述数据的指纹;
将所述数据的指纹与所述指纹库匹配处理;
若所述指纹库中有匹配的指纹,则将所述数据的逻辑页与物理页的映射关系写入FTL层;
若所述指纹库中没有匹配的指纹,则将所述数据写入所述固态硬盘,并将所述数据的指纹写入所述指纹库。
2. 根据权利要求1所述的固态硬盘删除重复数据的方法,其特征在于,所述固态硬盘通过LRU算法处理所述指纹库中的指纹数据。
3. 根据权利要求1所述的固态硬盘删除重复数据的方法,其特征在于,所述预设所述固态硬盘的数据的指纹库步骤包括:
设置所述固态硬盘读写的最小数据单位;
计算并保存每个最小数据单位的指纹。
4. 根据权利要求3所述的固态硬盘删除重复数据的方法,其特征在于,所述固态硬盘有数据写入时,计算所述数据的指纹步骤包括:
将写入数据划分为若干所述最小数据单位,计算每个所述最小数据单位的指纹;
所述将所述数据的指纹与所述指纹库匹配处理的步骤包括:
将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理。
5. 根据权利要求3或4所述的固态硬盘删除重复数据的方法,其特征在于,所述最小数据单元为数据页;
所述数据的指纹通过hash、校验或加密算法对所述最小数据单位计算获取。
6. 一种固态硬盘,其特征在于,包括:
指纹库设置模块,用于预设所述固态硬盘的数据的指纹库,所述指纹库储存有所述固态硬盘数据的指纹;
指纹计算模块,用于所述固态硬盘有数据写入时,计算所述数据的指纹;
指纹匹配模块,用于将所述数据的指纹与所述指纹库匹配处理;
数据处理模块,用于所述指纹库中有匹配的指纹时,将所述数据的逻辑页与物理页的映射关系写入FTL层;和/或用于所述指纹库中没有匹配的指纹时,将所述数据写入所述固态硬盘,并将所述数据的指纹写入所述指纹库。
7. 根据权利要求6所述的固态硬盘,其特征在于,所述固态硬盘还包括算法处理模块,用于通过LRU算法处理所述指纹库中的指纹。
8. 根据权利要求6所述的固态硬盘,其特征在于,所述指纹库设置模块包括:
设置子模块,用于设置所述固态硬盘读写的最小数据单位;
计算子模块,用于计算并保存每个最小数据单位的指纹。
9. 根据权利要求8所述的固态硬盘,其特征在于,所述指纹计算模块进一步用于将写入数据划分为若干所述最小数据单位,并计算每个所述最小数据单位的指纹;
所述指纹匹配模块进一步用于将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理。

10. 根据权利要求 8 或 9 所述的固态硬盘,其特征在于,所述最小数据单元为数据页;所述数据的指纹通过 hash、校验或加密算法对所述最小数据单位计算获取。

固态硬盘及其删除重复数据的方法

技术领域

[0001] 本发明涉及存储技术领域,尤其涉及一种固态硬盘及其删除重复数据的方法。

背景技术

[0002] 目前,SSD(Solid State Disk,固态硬盘)产品在市场已经趋于成熟,其高速的性能,尤其是随机性能也广为人知,但目前的固态硬盘的寿命问题还有待提高。影响 SSD 寿命的因素有多种,但其根本原因在于其 NAND Flash 编程/擦除次数,即对 NAND Flash 编程/擦除次数越多,损耗越大,寿命越短。

[0003] 综上所述,现有固态硬盘及其删除重复数据的方法,在实际使用上显然存在不便与缺陷,所以有必要加以改进。

发明内容

[0004] 针对上述的缺陷,本发明的目的在于提供一种固态硬盘及其删除重复数据的方法,其可以延长固态硬盘的使用寿命,且能大大提高固态硬盘的空间利用率及速率。

[0005] 为了实现上述目的,本发明提供一种固态硬盘删除重复数据的方法,所述方法包括:

[0006] 预设所述固态硬盘的数据的指纹库,所述指纹库储存有所述固态硬盘数据的指纹;

[0007] 所述固态硬盘有数据写入时,计算所述数据的指纹;

[0008] 将所述数据的指纹与所述指纹库匹配处理;

[0009] 若所述指纹库中有匹配的指纹,则将所述数据的逻辑页与物理页的映射关系写入 FTL 层;

[0010] 若所述指纹库中没有匹配的指纹,则将所述数据写入所述固态硬盘。

[0011] 根据本发明的固态硬盘删除重复数据的方法,所述固态硬盘通过 LRU 算法处理所述指纹库中的指纹数据。

[0012] 根据本发明的固态硬盘删除重复数据的方法,所述预设所述固态硬盘的数据的指纹库步骤包括:

[0013] 设置所述固态硬盘读写的最小数据单位;

[0014] 计算并保存每个最小数据单位的指纹。

[0015] 根据本发明的固态硬盘删除重复数据的方法,所述固态硬盘有数据写入时,计算所述数据的指纹步骤包括:

[0016] 将写入数据划分为若干所述最小数据单位,计算每个所述最小数据单位的指纹;

[0017] 所述将所述数据的指纹与所述指纹库匹配处理的步骤包括:

[0018] 将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理。

[0019] 根据本发明的固态硬盘删除重复数据的方法,所述最小数据单元为数据页;

[0020] 所述数据的指纹通过 hash、校验或加密算法对所述最小数据单位计算获取。

- [0021] 本发明还提供一种固态硬盘,包括:
- [0022] 指纹库设置模块,用于预设所述固态硬盘的数据的指纹库,所述指纹库储存有所述固态硬盘数据的指纹;
- [0023] 指纹计算模块,用于所述固态硬盘有数据写入时,计算所述数据的指纹;
- [0024] 指纹匹配模块,用于将所述数据的指纹与所述指纹库匹配处理;
- [0025] 数据处理模块,用于所述指纹库中有匹配的指纹时,将所述数据的逻辑页与物理页的映射关系写入FTL层;和/或用于所述指纹库中没有匹配的指纹时,将所述数据写入所述固态硬盘。
- [0026] 根据本发明的固态硬盘,所述固态硬盘还包括算法处理模块,用于通过LRU算法处理所述指纹库中的指纹。
- [0027] 根据本发明的固态硬盘,所述指纹库设置模块包括:
- [0028] 设置子模块,用于设置所述固态硬盘读写的最小数据单位;
- [0029] 计算子模块,用于计算并保存每个最小数据单位的指纹。
- [0030] 根据本发明的固态硬盘,所述指纹计算模块进一步用于将写入数据划分为若干所述最小数据单位,并计算每个所述最小数据单位的指纹;
- [0031] 所述指纹匹配模块进一步用于将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理。
- [0032] 根据本发明的固态硬盘,所述最小数据单元为数据页;
- [0033] 所述数据的指纹通过hash、校验或加密算法对所述最小数据单位计算获取。
- [0034] 本发明通过预设所述固态硬盘的数据的指纹库,该指纹库储存有固态硬盘数据的指纹,当固态硬盘有数据写入时,计算所述数据的指纹,并判断所述指纹库中是否有与所述数据匹配的指纹,若所述指纹库中有匹配的指纹,则将所述数据的逻辑页与物理页的映射关系写入FTL层;若所述指纹库中没有匹配的指纹,则将所述数据写入所述固态硬盘。借此,本发明能延长固态硬盘的使用寿命,且能大大提高固态硬盘的空间利用率及速率。

附图说明

- [0035] 图1是本发明一实施例的固态硬盘的结构示意图;
- [0036] 图2是本发明另一实施例的固态硬盘的结构示意图;
- [0037] 图3是本发明一实施例的固态硬盘删除重复数据的方法;
- [0038] 图4是本发明另一实施例的固态硬盘删除重复数据的方法。

具体实施方式

[0039] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0040] 参见图1,本发明提供了一种固态硬盘,该固态硬盘100包括指纹库设置模块10、指纹计算模块20、指纹匹配模块30以及数据处理模块40,其中:

[0041] 指纹库设置模块10,用于预设所述固态硬盘100的数据的指纹库,该指纹库储存有所述固态硬盘100数据的指纹。固态硬盘100的NAND Flash中储存有其应用的数据,各

数据以预设的单位进行储存,比如以页为单位进行储存,或以块为单位进行储存。指纹库设置模块 10 可以对这些单位数据进行指纹计算,不同的单位数据对应不同的指纹。所述指纹库可以保存在固态硬盘 100 的固定保留空间内,空间大小由指纹算法和固态硬盘 100 的容量大小计算获取,且由固件预分配,对用户不可见。

[0042] 指纹计算模块 20,用于所述固态硬盘 100 有数据写入时,计算所述数据的指纹。本发明的指纹是采用特定的指纹算法,比如 Hash、加密及检验等算法对数据单位进行计算获取,且该计算的是在固态硬盘 100 将数据交由 FTL 层之前完成的。

[0043] 指纹匹配模块 30,用于将所述数据的指纹与所述指纹库匹配处理。数据在交由 FTL 层处理之前,查找指纹库中的指纹,是否有与所述数据相同的指纹,若指纹相同,则说明固态硬盘 100 中储存有与写入的数据相同的数据,并交由数据处理模块 40 处理。

[0044] 数据处理模块 40,用于所述指纹库中有匹配的指纹时,将所述数据的逻辑页与物理页的映射关系写入 FTL 层,即,只需将该数据的映射关系写入 FTL 层进行保存,在实际应用中,若用到该数据,直接通过该映射关系调用相关数据,而不用分配新的物理区域储存该数据。若所述指纹库中没有匹配的指纹,数据处理模块 40 将数据写入所述固态硬盘,即按照正常 IO 的方式,交付给 FTL 层进行处理,包括分配物理页、记录逻辑页与地址页的映射等操作,并将所述数据的指纹写入指纹库。

[0045] 参见图 2,本发明另一实施例提供的固态硬盘,该实施例中,固态硬盘 100 还设有算法处理模块 50,用于通过 LRU 算法处理所述指纹库中的指纹。

[0046] 在固态硬盘的实际使用中,指纹库中的指纹会逐渐增多,也将带来指纹库空间占用较多,和指纹匹配过程变慢的弊端。针对此问题,结合应用中重复数据二八划分的原则,即仅有 20% 的数据会有重复的几率,本发明通过算法处理模块 50 对指纹库中的指纹进行替换处理,将最易重复数据对应的指纹留在指纹库中,对不容易重复数据对应的指纹替换出指纹库,并且保证仅占用少量的 SSD 保留空间。

[0047] 进一步的,所述指纹库设置模块 10 包括:

[0048] 设置子模块 11,用于设置所述固态硬盘 100 读写的最小数据单位。设置子模块 11 将数据以某个粒度划分(本实例以 NAND Flash page 为单位,如 4K),假设应用层写入的均以页为单位,如实际应用中非页为单位,则可以将数据以页为单位拆分,即一个数据页为最小数据单元。

[0049] 计算子模块 12,用于计算并保存每个最小数据单位的指纹,所述数据的指纹则通过 hash、校验或加密算法对其数据页计算获取。

[0050] 优选的,所述指纹计算模块 20 进一步用于将写入数据划分为若干所述最小数据单位,即固态硬盘 100 有数据写入时,按页写入,并计算每个所述最小数据单位的指纹。

[0051] 所述指纹匹配模块 30 进一步用于将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理。

[0052] 本发明的一具体实施例中,对写入数据 D1,计算其指纹为 F1,此时指纹库中无指纹,则匹配失败,通知 FTL 层以正常 IO 方式处理,在 FTL 完成写入后,将其指纹 F1 和对应物理页 PP1 写入指纹库。当写入数据 D2 时,计算其指纹为 f1,查询指纹库,匹配成功,则通知 FTL 仅写入此逻辑页与物理页 PP1 的映射关系即可,借此可节省储存空间,减少其 Flash 的擦除次数,提高其使用寿命。

[0053] 参见图 3, 本发明提供了一种固态硬盘删除重复数据的方法, 其可以通过如图 1 所示的固态硬盘 100 实现, 该方法包括:

[0054] 步骤 S301, 预设所述固态硬盘 100 的数据的指纹库, 该指纹库储存有所述固态硬盘 100 数据的指纹。固态硬盘 100 的 NAND Flash 中储存有其应用的数据, 各数据以预设的单位进行储存, 比如以页为单位进行储存, 或以块为单位进行储存。指纹库设置模块 10 可以对这些单位数据进行指纹计算, 不同的单位数据对应不同的指纹。所述指纹库可以保存在固态硬盘 100 的固定保留空间内, 空间大小由指纹算法和固态硬盘 100 的容量大小计算获取, 且由固件预分配, 对用户不可见。

[0055] 步骤 S302, 所述固态硬盘 100 有数据写入时, 计算所述数据的指纹。本发明的指纹是采用特定的指纹算法, 比如 Hash、加密及检验等算法对数据单位进行计算获取, 且该计算的是在固态硬盘 100 将数据交由 FTL 层之前完成的。

[0056] 步骤 S303, 将所述数据的指纹与所述指纹库匹配处理。数据在交由 FTL 层处理之前, 查找指纹库中的指纹, 是否有与所述数据相同的指纹, 若指纹相同, 则说明固态硬盘 100 中储存有与写入的数据相同的数据, 并交由数据处理模块 40 处理。

[0057] 步骤 S304, 所述指纹库中有匹配的指纹时, 将所述数据的逻辑页与物理页的映射关系写入 FTL 层, 即, 只需将该数据的映射关系写入 FTL 层进行保存, 在实际应用中, 若用到该数据, 直接通过该映射关系调用相关数据, 而不用分配新的物理区域储存该数据。

[0058] 步骤 S305, 所述指纹库中没有匹配的指纹时, 数据处理模块 40 将数据写入所述固态硬盘 100, 即按照正常 IO 的方式, 交付给 FTL 层进行处理, 包括分配物理页、记录逻辑页与地址页的映射等操作, 并将数据的指纹写入所述指纹库。

[0059] 再参见图 4, 本发明的另一实施例, 其可以通过如图 2 所示的固态硬盘 100 实现, 该方法包括:

[0060] 步骤 S401, 设置所述固态硬盘读写的最小数据单位。本发明将数据以某个粒度划分 (本实例以 NAND Flash page 为单位, 如 4K), 假设应用层写入的均以页为单位, 如实际应用中非页为单位, 则可以将数据以页为单位拆分, 即一个数据页为最小数据单元。

[0061] 步骤 S402, 计算并保存每个最小数据单位的指纹。所述数据的指纹则通过 hash、校验或加密算法对其数据页计算获取。

[0062] 步骤 S403, 将写入数据划分为若干所述最小数据单位, 计算每个所述最小数据单位的指纹。

[0063] 步骤 S404, 将所述数据的最小数据单位的指纹与所述指纹库中的最小数据单位匹配处理, 若有匹配的指纹则执行步骤 S405, 否则执行步骤 S406。

[0064] 步骤 S405, 将所述数据的逻辑页与物理页的映射关系写入 FTL 层。在实际应用中, 若用到该数据, 直接通过该映射关系调用相关数据, 而不用分配新的物理区域储存该数据。

[0065] 步骤 S406, 将所述数据写入所述固态硬盘。若所述指纹库中没有匹配的指纹, 数据处理模块 40 将数据写入所述固态硬盘, 即按照正常 IO 的方式, 交付给 FTL 层进行处理, 包括分配物理页、记录逻辑页与地址页的映射等操作, 并将所述数据的指纹写入指纹库。

[0066] 步骤 S407, 通过 LRU 算法处理所述指纹库中的指纹数据。该步骤可以保持指纹库的容量, 使其不会占用很大空间。

[0067] 综上所述, 本发明通过预设所述固态硬盘的数据的指纹库, 该指纹库储存有固态

硬盘数据的指纹,当固态硬盘有数据写入时,计算所述数据的指纹,并判断所述指纹库中是否有与所述数据匹配的指纹,若所述指纹库中有匹配的指纹,则将所述数据的逻辑页与物理页的映射关系写入 FTL 层;若所述指纹库中没有匹配的指纹,则将所述数据写入所述固态硬盘。借此,本发明能延长固态硬盘的使用寿命,且能大大提高固态硬盘的空间利用率及速率。

[0068] 当然,本发明还可有其它多种实施例,在不背离本发明精神及其实质的情况下,熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形,但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

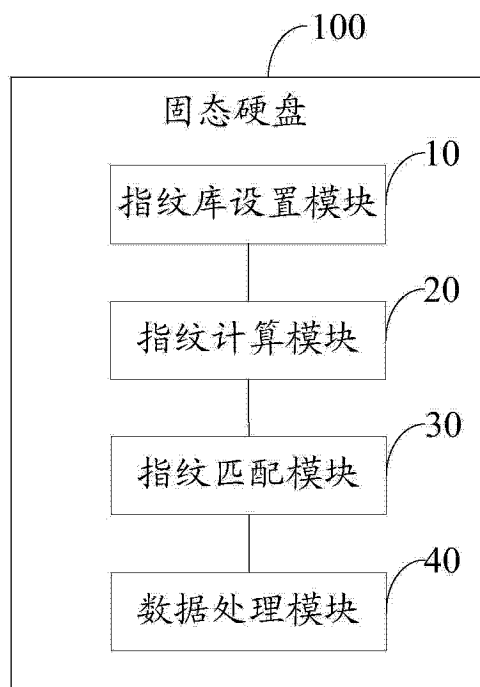


图 1

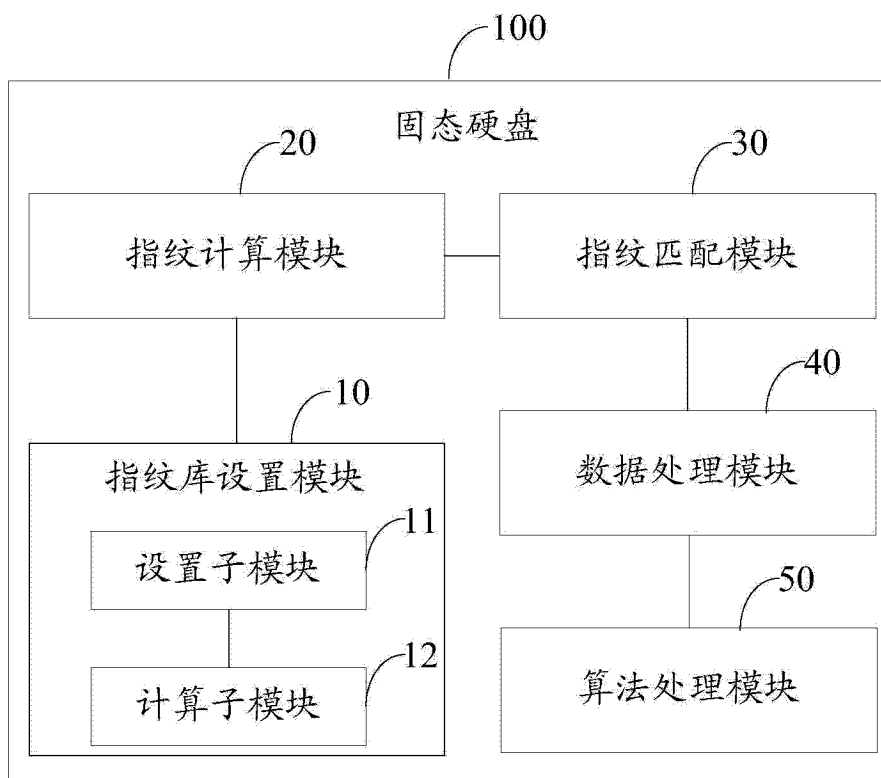


图 2

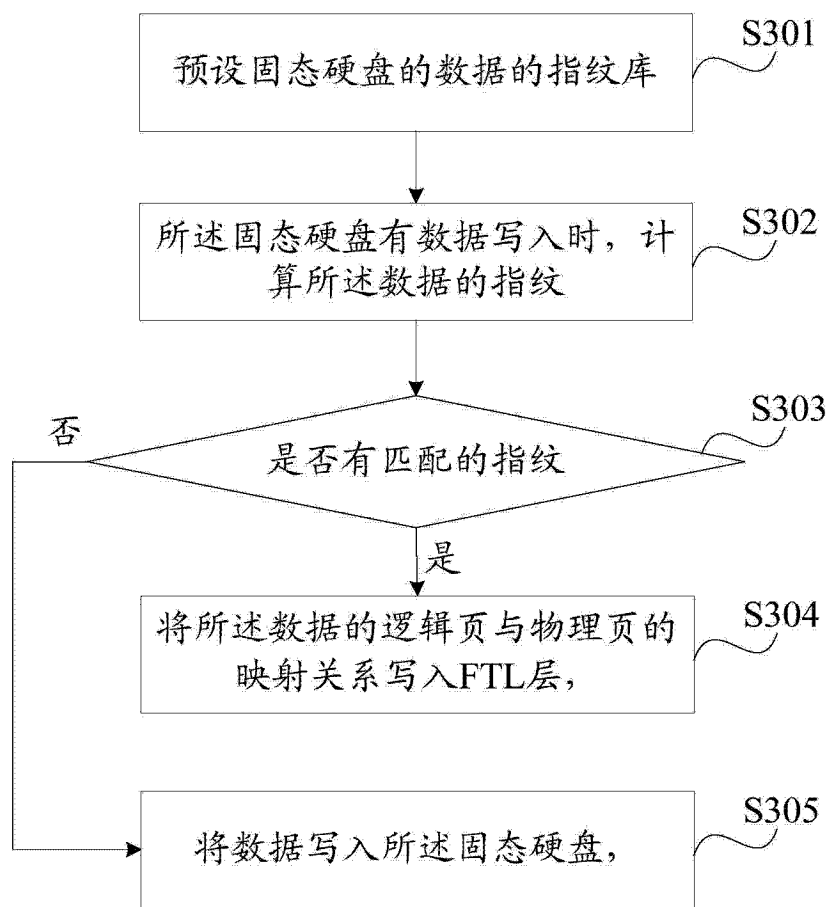


图 3

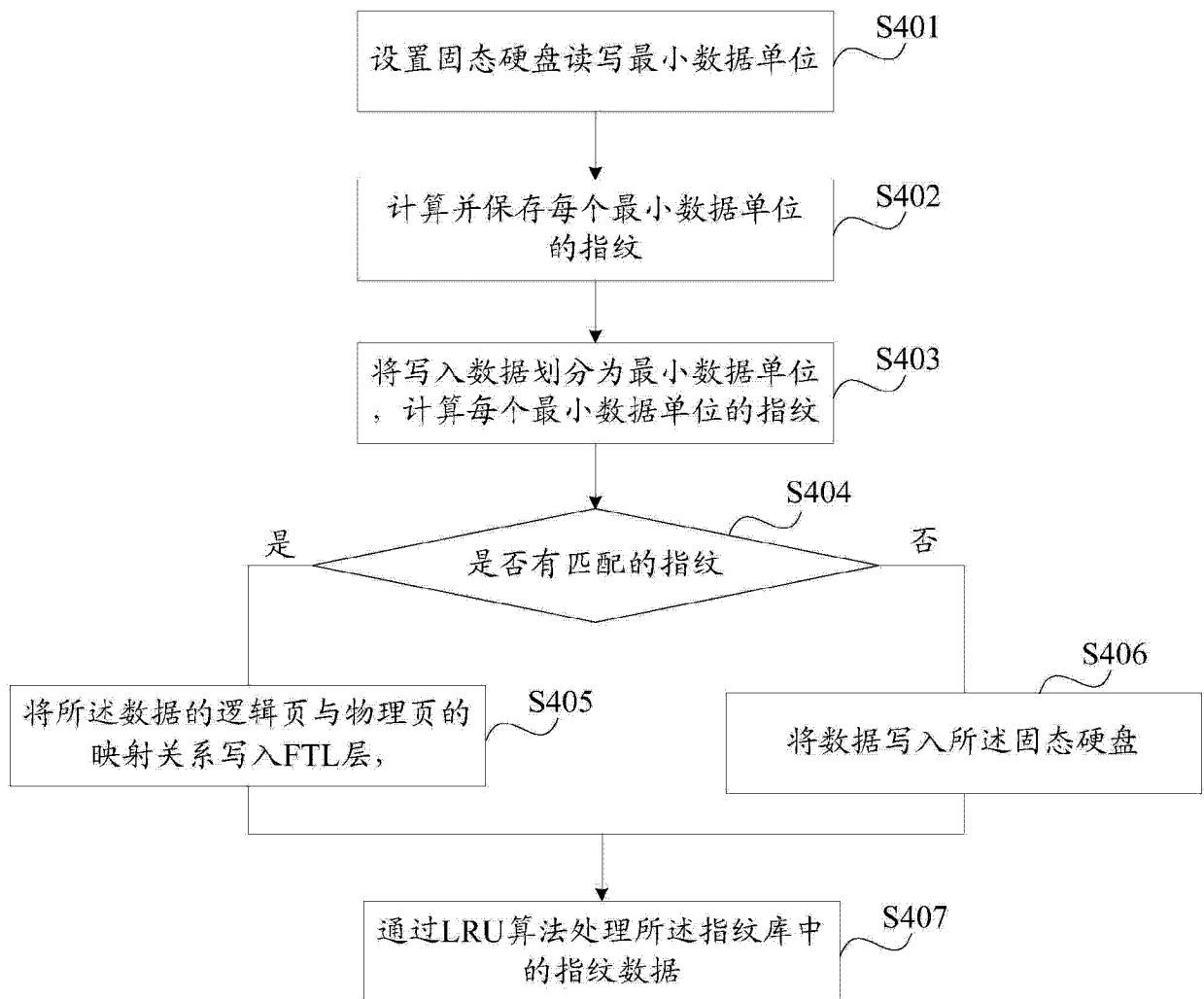


图 4