



(12) 发明专利

(10) 授权公告号 CN 103049387 B

(45) 授权公告日 2015. 12. 09

(21) 申请号 201210519322. 8

(22) 申请日 2012. 12. 06

(73) 专利权人 深圳市江波龙电子有限公司
地址 518000 广东省深圳市南山区科发路 8 号金融服务技术创新基地 1 栋 8 楼 A、B、C、D、E、F1

(72) 发明人 郭丹 梁小庆

(74) 专利代理机构 深圳中一专利商标事务所
44237

代理人 张全文

(51) Int. Cl.

G06F 12/02(2006. 01)

G06F 11/10(2006. 01)

(56) 对比文件

CN 101719099 A, 2010. 06. 02, 说明书第 [0008]-[0012]、[0033]-[0042]、[0048] 段, 权

利要求 1, 图 2-3.

CN 101645043 A, 2010. 02. 10,

CN 102722455 A, 2012. 10. 10, 说明书第 [0036]-[0043]、[0064]-[0073], 权利要求 1、4-6, 图 2、4.

CN 101930345 A, 2010. 12. 29,

US 6374266 B1, 2002. 04. 16,

审查员 周丹丹

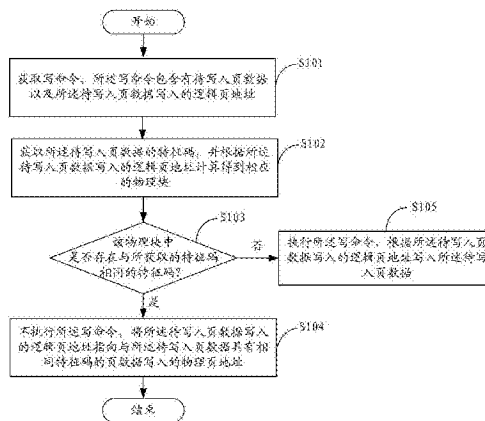
权利要求书3页 说明书9页 附图1页

(54) 发明名称

一种分页存储器件的压缩管理方法及装置

(57) 摘要

本发明适用于存储器技术领域, 提供了一种分页存储器件的压缩管理方法及装置, 所述方法包括: 获取写命令, 所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址; 获取所述待写入页数据的特征码, 并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块; 判断该物理块中是否存在与所获取的特征码相同的特征码; 若是, 不执行所述写命令, 将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址; 若否, 执行所述写命令, 根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据。通过本发明可有效解决以页为最小存储单元的存储器件的管理问题。



1. 一种分页存储器件的压缩管理方法,其特征在于,所述方法包括:

步骤 A、获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址;

步骤 B、获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块;

步骤 C、判断该物理块中是否存在与所获取的特征码相同的特征码;

步骤 D、若是,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;

步骤 E、若否,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据;

在步骤 C 之前,还包括:

创建信息表,所述信息表中记录有该物理块内已写入页数据的特征码以及所述已写入页数据写入的物理页地址和 / 或逻辑页地址;

所述步骤 C 具体包括:

判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码;

所述步骤 D 还包括:

获取所述待写入页数据写入的逻辑页地址与物理页地址的相对偏移值,并将所获取的相对偏移值存储至所述信息表。

2. 如权利要求 1 所述的方法,其特征在于,所述步骤 C 具体包括:

判断该物理块中前 N 个物理页地址中是否存在与所获取的特征码相同的特征码,所述 N 为大于零小于物理块中页数的整数。

3. 如权利要求 1 所述的方法,其特征在于,所述步骤 E 还包括:

将所述待写入页数据的特征码以及所述待写入页数据写入的物理页地址和 / 或逻辑页地址存储至所述信息表。

4. 如权利要求 1 所述的方法,其特征在于,所述获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块,判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码,若是,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;若否,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据具体包括:

在所述分页存储器件支持多种哈希算法时,对所述待写入页数据进行多次消息摘录,获取每次消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次消息摘录后的特征码都相同的特征码,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将所述写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次消息摘录后的特征码;

或者,在所述分页存储器件支持一种哈希算法时,对所述待写入页数据进行多次不同偏移的消息摘录,获取每次不同偏移消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次不同偏移消息摘录后的特征码都相同的特征码,若否,执行当前写命

令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次不同偏移消息摘录后的特征码;

或者,在所述分页存储器件支持错误检查和纠正 ECC 码时,获取所述待写入页数据的 ECC 码和消息摘录后的特征码,判断该物理块的所述信息表中是否同时存在该 ECC 码和消息摘录后的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据的 ECC 码和消息摘录后的特征码;

或者,获取所述待写入页数据的特征码,在该物理块的所述信息表中存在与所获取的特征码相同的特征码时,判断所述待写入页数据中的 N 个字节与该物理块中相同特征码对应的页数据中的 N 个字节是否全部相同,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述 N 为大于零且小于或者等于所述分页存储器件中页的大小的整数。

5. 一种分页存储器件的压缩管理装置,其特征在于,所述装置包括:

第一获取单元,用于获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址;

第二获取单元,用于获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块;

判断单元,用于判断该物理块中是否存在与所获取的特征码相同的特征码;

指向单元,用于在所述判断单元判断结果为是时,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;

写入单元,用于在所述判断单元判断结果为否时,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据;

所述装置还包括:

信息表创建单元,用于创建信息表,所述信息表中记录有该物理块内已写入页数据的特征码以及所述已写入页数据写入的物理页地址和 / 或逻辑页地址;

所述判断单元具体用于,判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码;

所述指向单元还用于,获取所述待写入页数据写入的逻辑页地址与物理页地址的相对偏移值,并将所获取相对偏移值存储至所述信息表。

6. 如权利要求 5 所述的装置,其特征在于,所述判断单元具体用于,判断该物理块中前 N 个物理页地址中是否存在与所获取的特征码相同的特征码,所述 N 为大于零小于物理块中页数的整数。

7. 如权利要求 5 所述的装置,其特征在于,所述写入单元还用于,将所述待写入页数据

的特征码以及所述待写入页数据写入的物理页地址和 / 或逻辑页地址存储至所述信息表。

8. 如权利要求 5 所述的装置,其特征不在于,所述装置还包括:

处理单元,用于在所述分页存储器件支持多种哈希算法时,对所述待写入页数据进行多次消息摘录,获取每次消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次消息摘录后的特征码都相同的特征码,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将所述写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次消息摘录后的特征码;

或者,在所述分页存储器件支持一种哈希算法时,对所述待写入页数据进行多次不同偏移的消息摘录,获取每次不同偏移消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次不同偏移消息摘录后的特征码都相同的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次不同偏移消息摘录后的特征码;

或者,在所述分页存储器件支持错误检查和纠正 ECC 码时,获取所述待写入页数据的 ECC 码和消息摘录后的特征码,判断该物理块的所述信息表中是否同时存在该 ECC 码和消息摘录后的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据的 ECC 码和消息摘录后的特征码;

或者,获取所述待写入页数据的特征码,在该物理块的所述信息表中存在与所获取的特征码相同的特征码时,判断所述待写入页数据中的 N 个字节与该物理块中相同特征码对应的页数据中的 N 个字节是否全部相同,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述 N 为大于零且小于或者等于所述分页存储器件中页的大小的整数。

一种分页存储器件的压缩管理方法及装置

技术领域

[0001] 本发明属于存储器技术领域,尤其涉及一种分页存储器件的压缩管理方法及装置。

背景技术

[0002] 当前通用的数据压缩方法包括无损数据压缩和有损数据压缩,如基于统计模型的压缩技术和基于样本库模型的压缩技术为无损数据压缩技术;多媒体数据压缩包括音频压缩、图像压缩、视频压缩等则大多为有损数据压缩技术;对存储设备而言,必须使用无损的数据压缩方式。

[0003] 现有的无损数据压缩方式包括:

[0004] 一、样本库型数据压缩方法,如表 1 所示:

[0005]

Index	String
1	Prince
2	Love
3	Princess
4	Forever

[0006] 表 1

[0007] 例如需要压缩 Yes ! Prince love Princess forever 则根据表 1 提供的样本库得到压缩结果为:

[0008] Yes ! &1&2&3&4 (其中 & 是特殊符号,如果需要输出 & 则压缩数据为 &&)

[0009] 缺点:样本库的大小和好坏直接影响压缩结果。大的样本库开销较大,小的样本库无法满足压缩要求,例如通过表 1 提供的样本库压缩 I am happy today! 则完全无法压缩。

[0010] 二、行程编码型数据压缩方法

[0011] 行程编码的基本原理是:通过一个符号值或串长代替具有相同值的连续符号(连续符号构成了一段连续的“行程”),使符号长度少于原始数据的长度。在各行或者各列数据的代码发生变化时,一次记录该代码及相同代码重复的个数,从而实现数据的压缩。

[0012] 例如:55555577777733322221111111

[0013] 行程编码为:(5,6)(7,5)(3,3)(2,4)(1,7)。可见,行程编码的位数远远少于原始字符串的位数。

[0014] 然而,并不是所有的行程编码都远远少于原始字符串的位数,例如:555555 是 6 个字符,而(5,6)是 5 个字符,这也存在压缩量的问题。另外,行程编码针对随机数据的压缩出来的编码可能比没压缩还大。

[0015] 三、哈夫曼型数据压缩方法

[0016] 现有的哈夫曼型数据压缩方法在进行数据压缩时需要构建哈夫曼树,空间开销过大。而且,压缩后的长度是一个变化值,不便于以页为最小存储单元的存储器件,例如对 8k 的数据进行压缩时,压缩完可能是 7.5k,对于以页为最小存储单元的存储器件来说,如果其页的大小为 8k,则压缩后剩余的 0.5k 基本无法使用。

[0017] 综上所述,现有的数据压缩方法存在以下问题:(1)对运算开销过大或者对存储空间要求过高;(2)压缩的效果无法保证;(3)压缩结果不便于以页为最小存储单元的存储器件管理。

发明内容

[0018] 本发明实施例的目的在于提供一种分页存储器件的压缩管理方法,以解决现有的压缩管理方法对运算开销过大或者对存储空间要求过高、压缩效果无法保证、且压缩结果不便于以页为最小存储单元的存储器件管理的问题。

[0019] 本发明实施例是这样实现的,一种分页存储器件的压缩管理方法,所述方法包括:

[0020] 步骤 A、获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址;

[0021] 步骤 B、获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块;

[0022] 步骤 C、判断该物理块中是否存在与所获取的特征码相同的特征码;

[0023] 步骤 D、若是,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;

[0024] 步骤 E、若否,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据。

[0025] 本发明实施例的另一目的在于提供一种分页存储器件的压缩管理装置,所述装置包括:

[0026] 第一获取单元,用于获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址;

[0027] 第二获取单元,用于获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块;

[0028] 判断单元,用于判断该物理块中是否存在与所获取的特征码相同的特征码;

[0029] 指向单元,用于在所述判断单元判断结果为是时,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;

[0030] 写入单元,用于在所述判断单元判断结果为否时,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据。

[0031] 从上述技术方案可以看出,本发明实施例中数据压缩的最小单位是页,页内的数据不进行压缩和解压缩,数据的写入和读出也是以页为单位,因此本发明实施例提供的压缩管理方法非常适合采用运算能力不足或者成本低的分页存储器件。而且本发明实施例提

供的压缩管理方法,是通过判断分页存储器件的物理块中是否存在与所获取的特征码相同的特征码,在存在时,则不写入数据。由于可以根据所述待写入页数据写入的逻辑页地址快速定位到分页存储器件中相应的物理块,因而能有效提高特征码比较的效率,进而提高分页存储器件的数据写入效率以及数据压缩的效率。

附图说明

[0032] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0033] 图 1 是本发明实施例一提供的分页存储器件压缩管理方法的实现流程图;

[0034] 图 2 是本发明实施例二提供的分页存储器件压缩管理装置的组成结构图。

具体实施方式

[0035] 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。

[0036] 为了说明本发明所述的技术方案,下面通过具体实施例来进行说明。

[0037] 实施例一:

[0038] 图 1 示出了本发明实施例一提供的分页存储器件压缩管理方法的实现流程,该方法过程详述如下:

[0039] 在步骤 S101 中,获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址等信息。

[0040] 在本实施例中,分页存储器件为以页为最小存储单元的存储器件,即按块擦除、按页读写的存储器件,例如闪存存储器件等。所述页数据为页内存储的数据。

[0041] 在步骤 S102 中,获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块。

[0042] 在步骤 S103 中,判断该物理块中是否存在与所获取的特征码相同的特征码,若判断结果为“是”,则执行步骤 S104,若判断结果为“否”,则执行步骤 S105。

[0043] 由于一个物理块可能包含很多页,为了提高特征码比较的效率,判断该物理块中是否存在与所获取的特征码相同的特征码具体可以为:判断该物理块中前 N 个物理页地址中是否存在与所获取的特征码相同的特征码,若是,则表示当前写命令中的待写入页数据已写入,执行步骤 S104;否则执行步骤 S105。其中,所述 N 为大于零小于物理块中页数的整数。

[0044] 优选的是,本实施例在判断该物理块中是否存在与所获取的特征码相同的特征码之前,还包括:

[0045] 创建信息表,所述信息表中记录有该物理块内已写入页数据的特征码以及所述已写入页数据写入的物理页地址和/或逻辑页地址;

[0046] 所述判断该物理块中是否存在与所获取的特征码相同的特征码具体包括:

[0047] 判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码。

[0048] 本实施例中所述待写入页数据的特征码以及所述信息表中已写入页数据的特征码都可采用相同的算法获取。优选的是,通过哈希算法获取。当然也可以采用其他算法获取,在此并不用以限制本发明。

[0049] 哈希(Hash)算法属于对一个大的数据进行消息摘录,得到该段数据的特征码,主要用于密码验证和数据完整性验证,同时也可以用于快速判断两段数据是否相同。

[0050] 以 MD5 哈希算法为例,例如表 2 中如下的几段数据:

[0051]

页数据	MD5 值
1234567890abcdefg	7206ddfa511b6ab05734b603c1b88be6
2234567890abcdefg	628cb0947d31075fa305699029ddfecce
0234567890abcdefg	df9e94f848b6608dd2af317b693744ba

[0052]

1234567890abcdefg	ad91a28a12e83ce4999b23260f7785c3
1234567890abcdef	996ce17f6abc9fe126b57aa5f1d8c92c
1234567890abcdefg	7206ddfa511b6ab05734b603c1b88be6

[0053] 表 1

[0054] 从表 1 可以看到,即使页数据只改变一点点,经哈希算法计算后的特征码都会产生大幅度变动,因此采用哈希算法计算特征码,能够降低重码率。

[0055] 在步骤 S104 中,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址。

[0056] 具体的是,在所述物理块中存在与所获取的特征码相同的特征码时,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,同时获取所述待写入页数据写入的逻辑页地址与物理页地址的相对偏移值,并将所获取的相对偏移值存储至所述样本库。其中,所述相对偏移值为所述写命令中逻辑页地址与实际写入的物理页地址的差。例如,写入的逻辑页地址为逻辑页 5,实际写入的物理页地址为物理页 2,则相对偏移值为 3。

[0057] 在本实施例中,在所述写命令中的待写入页数据之前已写入过,即已存在于分页存储器件中时,不执行所述写指令,只将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,并提醒用户该数据已写入过,因此能有效提高分页存储器件的数据写入效率,减少数据的写入次数,进而减少分页存储器件的磨损。

[0058] 在步骤 S105 中,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据。

[0059] 具体的是,在所述物理块中不存在与所获取的特征码相同的特征码时,执行所述

写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据,并将所述待写入页数据的特征码以及所述待写入页数据写入的物理页地址和 / 或逻辑页地址等信息存储至所述信息表。

[0060] 需要说明的是,本实施例所述信息表中记录的已写入页数据地址既可以为物理页地址,也可以为逻辑页地址。如果为逻辑页地址,在写命令中的待写入页数据之前已写入过时,只需将当前写命令中待写入页数据的写入逻辑页地址指向具有相同特征码的原始数据写入的逻辑页地址即可。另外,当记录的是逻辑页地址时,在读出数据时需要将逻辑页地址转换为物理页地址。

[0061] 举例说明上述过程:

[0062]

逻辑页地址	物理页地址	相对偏移值	页数据	特征码
0	0	0	55555555	f638f4354ff089323d1a5f78fd8f63ca
1	1	0	Aaaaaaaa	99472c386d5ac7c7d02202ce0a23908a
2	2	0	Hahahahah a	387854bba282485b859648fca07b442b
3	0	3	55555555	f638f4354ff089323d1a5f78fd8f63ca
4	3	1	Asdqwdqas	cda4eb815003672a029713f05ba3e765
5	2	3	Hahahahah a	387854bba282485b859648fca07b442b
...				
127				

[0063] 表 2

[0064] 如表 2 所示,逻辑页地址 0 (LogPage0): 该物理块的信息表中未匹配到与待写入页数据的特征码相同的特征码,写入物理页地址 0 (PhyPage0),其特征码记入信息表,相对偏移值为 0。

[0065] 逻辑页地址 1 (LogPage1): 该物理块的信息表中未匹配到与待写入页数据的特征码相同的特征码,写入物理页地址 1 (PhyPage1),其特征码记入信息表,相对偏移值为 0。

[0066] 逻辑页地址 2 (LogPage2): 该物理块的信息表中未匹配到与待写入页数据的特征码相同的特征码,写入物理页地址 2 (PhyPage2),其特征码记入信息表,相对偏移值为 0。

[0067] 逻辑页地址 3 (LogPage3): 该物理块的信息表中匹配到与待写入页数据的特征码相同的特征码,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址 PhyPage0,相对偏移值为 3。

[0068] 逻辑页地址 4 (LogPage4): 该物理块的信息表中未匹配到与待写入页数据的特征码相同的特征码,写入物理页地址 3 (PhyPage3),其特征码记入信息表,相对偏移值为 1。

[0069] 逻辑页地址 5 (LogPage5): 该物理块的信息表中匹配到与待写入页数据的特征码相同的特征码,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址 PhyPage2,相对偏移值为 3。

[0070] 作为本发明的一优选实施例,为了降低重码率,所述获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块,判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码,若是,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;若否,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据具体包括:

[0071] 在所述分页存储器件支持多种哈希算法时,对所述待写入页数据进行多次消息摘录,获取每次消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次消息摘录后的特征码都相同的特征码,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将所述写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次消息摘录后的特征码;

[0072] 或者,在所述分页存储器件支持一种哈希算法时,对所述待写入页数据进行多次不同偏移的消息摘录,获取每次不同偏移消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次不同偏移消息摘录后的特征码都相同的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次不同偏移消息摘录后的特征码;

[0073] 或者,在所述分页存储器件支持错误检查和纠正 ECC 码时,获取所述待写入页数据的 ECC 码和消息摘录后的特征码,判断该物理块的所述信息表中是否同时存在该 ECC 码和消息摘录后的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据的 ECC 码和消息摘录后的特征码;

[0074] 或者,获取所述待写入页数据的特征码,在该物理块的所述信息表中存在与所获取的特征码相同的特征码时,判断所述待写入页数据中的 N 个字节与该物理块中相同特征码对应的页数据中的 N 个字节是否全部相同,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述 N 为大于零且小于或者等于所述分页存储器中页的大小的整数。

[0075] 本实施例中数据压缩的最小单位是页,页内的数据不进行压缩和解压缩,数据的写入和读出也是以页为单位,因此本发明实施例提供的压缩管理方法非常适合采用运算能力不足或者成本低廉的分页存储器件。而且本发明实施例提供的压缩管理方法,是通过判断分页存储器件的物理块中是否存在与所获取的特征码相同的特征码,在存在时,则不写入数据。由于可以根据所述待写入页数据写入的逻辑页地址快速定位到分页存储器件中相应的物理块,因而能有效提高特征码比较的效率,进而提高分页存储器件的数据写入效率以及数据压缩的效率。本发明实施例提供的压缩管理方法可用于压缩出现重复次数较少的数

据,通过记录数据与相对偏移值(例如“吃葡萄不吐葡萄皮,不吃葡萄倒吐葡萄皮”,“葡萄”第一次出现以后,后面再出现“葡萄”描述为(2,4)表示内容出现在当前位置往前偏移4个字),即可实现数据的压缩,不需要花费多余的空间来保存字典,而且方便于搜索和解压缩。

[0076] **实施例二:**

[0077] 图2示出了本发明实施例二提供的分页存储器件压缩管理装置的组成结构,为了便于说明,仅示出了与本发明实施例相关的部分。

[0078] 该分页存储器件压缩管理装置可以应用于存储设备中,可以是运行于存储设备内的软件单元、硬件单元或者软硬件相结合的单元,也可以作为独立的挂件集成到存储设备中或者运行于存储设备的应用系统中。

[0079] 该分页存储器件压缩管理装置包括第一获取单元21、第二获取单元22、判断单元23、指向单元24以及写入单元25。其中,各单元的具体功能如下:

[0080] 第一获取单元21,用于获取写命令,所述写命令包含有待写入页数据以及所述待写入页数据写入的逻辑页地址;

[0081] 第二获取单元22,用于获取所述待写入页数据的特征码,并根据所述待写入页数据写入的逻辑页地址计算得到相应的物理块;

[0082] 判断单元23,用于判断该物理块中是否存在与所获取的特征码相同的特征码;

[0083] 指向单元24,用于在所述判断单元23判断结果为是时,不执行所述写命令,将所述待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址;

[0084] 写入单元25,用于在所述判断单元23判断结果为否时,执行所述写命令,根据所述待写入页数据写入的逻辑页地址写入所述待写入页数据。

[0085] 进一步的,所述判断单元23具体用于,判断该物理块中前N个物理页地址中是否存在与所获取的特征码相同的特征码,所述N为大于零小于物理块中页数的整数。

[0086] 进一步的,所述装置还包括:

[0087] 信息表创建单元26,用于创建信息表,所述信息表中记录有该物理块内已写入页数据对应的特征码以及所述已写入页数据写入的物理页地址和/或逻辑页地址;

[0088] 所述判断单元23具体用于,判断该物理块的所述信息表中是否存在与所获取的特征码相同的特征码。

[0089] 进一步的,所述指向单元24还用于,获取所述待写入页数据写入的逻辑页地址与物理页地址的相对偏移值,并将所获取相对偏移值存储至所述信息表。

[0090] 进一步的,所述写入单元25还用于,将所述待写入页数据的特征码以及所述待写入页数据写入的物理页地址和/或逻辑页地址存储至所述信息表。

[0091] 进一步的,所述装置还包括:

[0092] 处理单元,用于在所述分页存储器件支持多种哈希算法时,对所述待写入页数据进行多次消息摘录,获取每次消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次消息摘录后的特征码都相同的特征码,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将所述写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次消息摘录后的特

征码；

[0093] 或者,在所述分页存储器件支持一种哈希算法时,对所述待写入页数据进行多次不同偏移的消息摘录,获取每次不同偏移消息摘录后的特征码,判断该物理块的所述信息表中是否存在与所述每次不同偏移消息摘录后的特征码都相同的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据多次不同偏移消息摘录后的特征码;

[0094] 或者,在所述分页存储器件支持错误检查和纠正 ECC 码时,获取所述待写入页数据的 ECC 码和消息摘录后的特征码,判断该物理块的所述信息表中是否同时存在该 ECC 码和消息摘录后的特征码,若否,执行当前写命令,根据当前写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述信息表中记录有已写入页数据的 ECC 码和消息摘录后的特征码;

[0095] 或者,获取所述待写入页数据的特征码,在该物理块的所述信息表中存在与所获取的特征码相同的特征码时,判断所述待写入页数据中的 N 个字节与该物理块中相同特征码对应的页数据中的 N 个字节是否全部相同,若否,执行所述写命令,根据所述写命令中待写入页数据写入的逻辑页地址写入所述待写入页数据;若是,不执行所述写命令,将当前写命令中待写入页数据写入的逻辑页地址指向与所述待写入页数据具有相同特征码的页数据写入的物理页地址,其中所述 N 为大于零且小于或者等于所述分页存储器中页的大小的整数。

[0096] 需要说明的是,实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,例如第二获取单元 22、判断单元 23、指向单元 24 以及写入单元 25 可以集成到处理单元中。

[0097] 本实施例提供的分页存储器件压缩管理装置可以使用在前述对应的分页存储器件压缩管理方法,详情参见上述分页存储器件压缩管理方法实施例一的相关描述,在此不再赘述。

[0098] 本领域普通技术人员可以理解为上述实施例二所包括的各个单元只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0099] 综上所述,本实施例中数据压缩的最小单位是页,页内的数据不进行压缩和解压缩,数据的写入和读出也是以页为单位,因此本发明实施例提供的压缩管理方法非常适合采用运算能力不足或者成本低廉的分页存储器件。而且本发明实施例提供的压缩管理方法,是通过判断分页存储器件的物理块中是否存在与所获取的特征码相同的特征码,在存在时,则不写入数据。由于可以根据所述待写入页数据写入的逻辑页地址快速定位到分页存储器件中相应的物理块,因而能有效提高特征码比较的效率,进而提高分页存储器件的数据写入效率以及数据压缩的效率。本发明实施例提供的压缩管理方法可用于压缩出现重复次数较少的数据,通过记录数据与相对偏移值(例如“吃葡萄不吐葡萄皮,不吃葡萄倒吐葡萄皮”,“葡萄”第一次出现以后,后面再出现“葡萄”描述为(2,4)表示内容出现在当前位置

往前偏移 4 个字),即可实现数据的压缩,不需要花费多余的空间来保存字典,而且便于搜索和解压缩。另外,本发明实施例通过多种比较方式,可有效降低比较过程中的重码率,具有较强的易用性和实用性。

[0100] 以上内容是结合具体的优选实施方式对本发明所作的进一步详细说明,不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说,在不脱离本发明构思的前提下做出若干等同替代或明显变型,而且性能或用途相同,都应当视为属于本发明由所提交的权利要求书确定的专利保护范围。

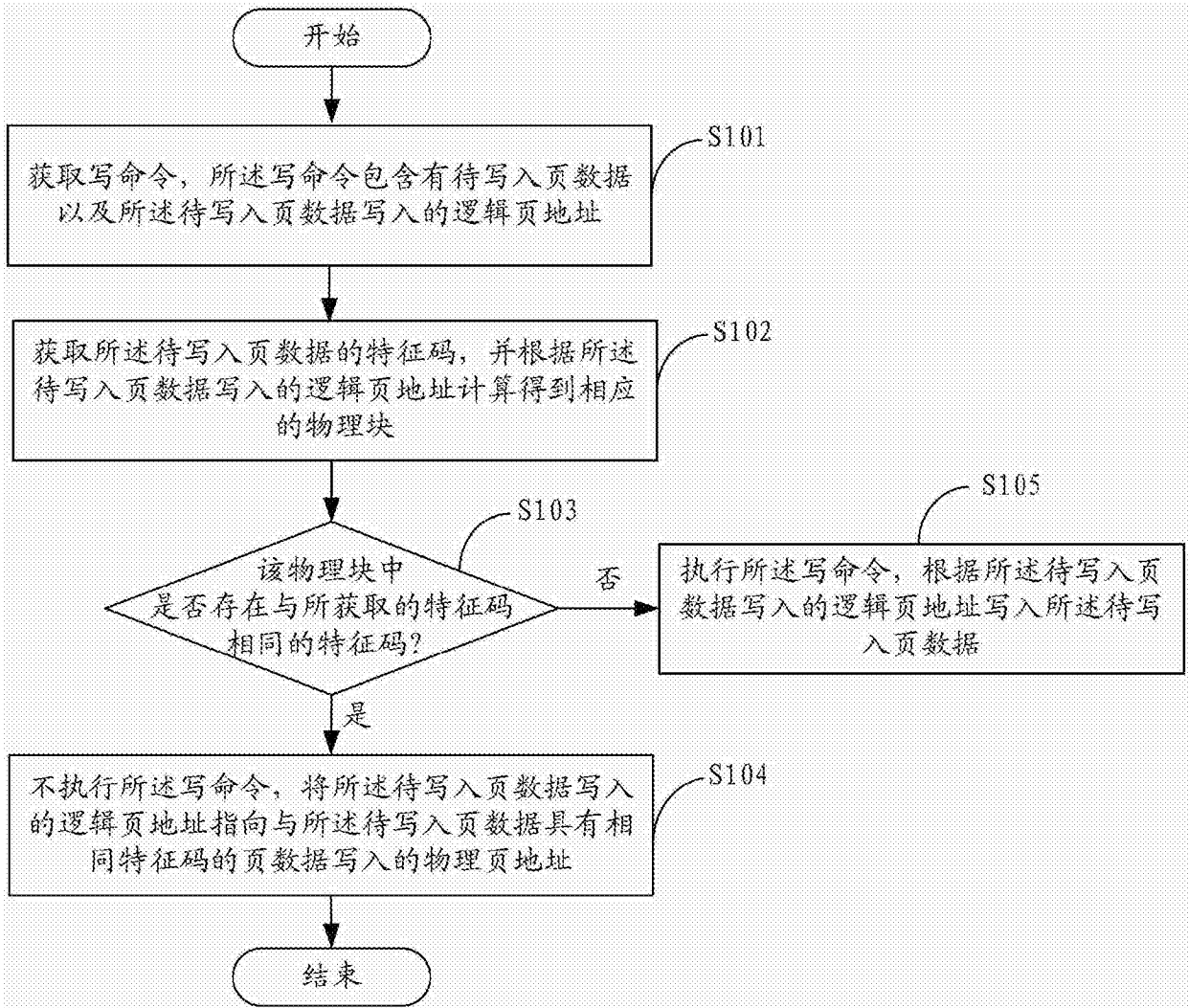


图 1

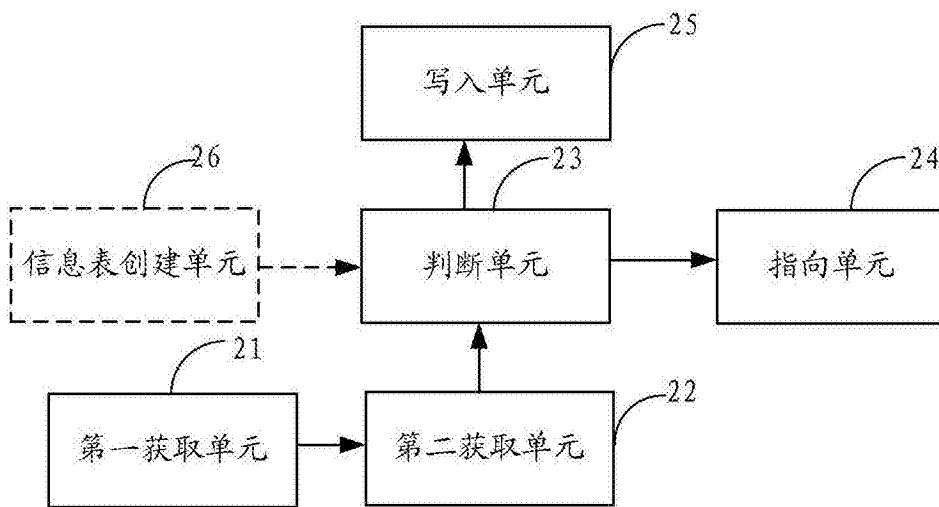


图 2