



(12)发明专利

(10)授权公告号 CN 105009091 B

(45)授权公告日 2018.01.02

(21)申请号 201380068657.5

(73)专利权人 西部数据技术公司

(22)申请日 2013.09.23

地址 美国加利福尼亚

(65)同一申请的已公布的文献号

(72)发明人 R·L·霍恩

申请公布号 CN 105009091 A

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245

(43)申请公布日 2015.10.28

代理人 赵蓉民

(30)优先权数据

(51)Int.Cl.

13/727,150 2012.12.26 US

G06F 12/00(2006.01)

(85)PCT国际申请进入国家阶段日

审查员 李中兴

2015.06.26

(86)PCT国际申请的申请数据

PCT/US2013/061242 2013.09.23

(87)PCT国际申请的公布数据

W02014/105228 EN 2014.07.03

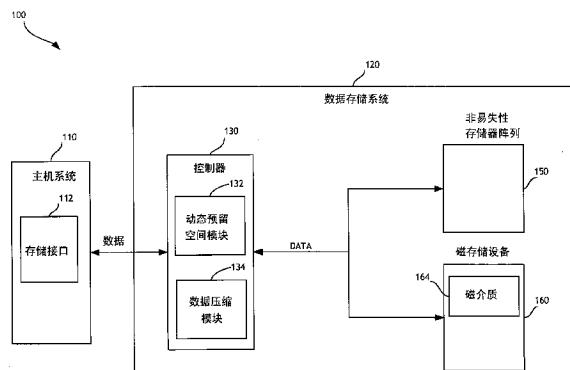
权利要求书3页 说明书5页 附图4页

(54)发明名称

一种数据存储系统及用于数据存储系统的动态预留空间方法

(57)摘要

公开的实施例针对用于数据存储系统的动态预留空间的系统和方法。在一个实施例中，数据存储系统可以为预留空间保留诸如非易失性固态存储器的存储器的部分。取决于各种预留空间因素，由于压缩用户数据而恢复的存储空间可以被分配用于存储用户数据和/或预留空间。利用公开的动态预留空间系统或方法可以导致更有效地使用缓存存储器，减小写入放大，增大缓存命中率等，从而可以获得改进的数据存储系统性能和增大的耐久性和寿命。



1. 一种数据存储系统,包括:

非易失性存储器阵列,其包括用户数据部分和预留空间部分,所述用户数据部分被配置为存储用户数据;

控制器,其被配置为借助以下步骤动态调整所述非易失性存储器阵列的所述预留空间部分的容量:

压缩被存储在所述用户数据部分中的至少一些用户数据;

确定由于所述压缩而恢复的存储容量的量;

计算一个或更多个预留空间参数,所述预留空间参数包括在所述非易失性存储器阵列中存储的非冗余数据的量的测量值,所述非冗余数据包括不是永久地存储在除所述非易失性存储器阵列之外的所述数据存储系统的任何存储器中的数据;以及

至少部分基于所述一个或更多个预留空间参数,将所述恢复的存储容量分配给以下项中的至少一个:所述用户数据部分的容量和所述预留空间部分的所述容量。

2. 根据权利要求1所述的数据存储系统,其中,所述控制器被进一步配置为至少部分基于所述一个或更多个预留空间参数中的两个或更多个的组合,分配所述恢复的存储容量。

3. 根据权利要求2所述的数据存储系统,其中,所述控制器被进一步配置为确定所述两个或更多个预留空间参数的加权平均值。

4. 根据权利要求1所述的数据存储系统,其中,所述一个或更多个预留空间参数包括以下项中的至少一个:

用户数据压缩率;

所述预留空间部分的所述容量;

所述非易失性存储器阵列的磨损级别;

不工作的非易失性存储器阵列单元的数量;

从所述非易失性存储器阵列读取的数据量与写入所述非易失性存储器阵列的数据量的比率;以及

写入所述非易失性存储器阵列的连续数据量与写入所述非易失性存储器阵列的非连续数据量的比率。

5. 根据权利要求4所述的数据存储系统,其中,所述控制器被进一步配置为响应于以下项中的至少一个,分配至少一些所述恢复的存储容量给所述预留空间部分的所述容量:

相比于磨损级别阈值的所述非易失性存储器阵列的所述磨损级别的增大;

相比于读/写阈值的从所述非易失性存储器阵列读取的所述数据量与写入所述非易失性存储器阵列的所述数据量的所述比率的增大;以及

相比于连续数据写阈值的写入所述非易失性存储器阵列的所述连续数据量与写入所述非易失性存储器阵列的所述非连续数据量的所述比率的减小。

6. 根据权利要求4所述的数据存储系统,其中,所述控制器被进一步配置为响应于以下项中的至少一个,分配至少一些所述恢复的存储容量给所述用户数据部分的所述容量:

相比于不工作单元阈值的不工作的非易失性存储器单元的所述数量的增大;

相比于数据压缩阈值的所述用户数据压缩率的减小;

相比于预留空间阈值的所述预留空间部分的所述容量的增大;以及

相比于连续数据写阈值的写入所述非易失性存储器阵列的所述连续数据量与写入所

述非易失性存储器阵列的所述非连续数据量的所述比率的增大。

7. 根据权利要求1所述的数据存储系统，其中，所述控制器被进一步配置为使用无损压缩来压缩存储在所述用户数据部分中的所述至少一些用户数据。

8. 根据权利要求1所述的数据存储系统，其中，所述数据存储系统进一步包括磁存储设备，并且其中，所述非易失性存储器阵列被配置为用于所述磁存储设备的缓存存储器。

9. 根据权利要求1所述的数据存储系统，其中，所述控制器被进一步配置为响应于相比于非冗余数据阈值的所述测量值的增大，分配至少一些所述恢复的存储容量给所述用户数据部分的所述容量，所述测量值包括在所述非易失性存储器阵列中存储的非冗余数据的百分比。

10. 根据权利要求1所述的数据存储系统，其中，所述非易失性存储器阵列被配置为针对远程数据存储设备的缓存存储器。

11. 一种在包括非易失性存储器阵列的数据存储系统中动态调整预留空间部分的容量的方法，所述非易失性存储器阵列包括用户数据部分和所述预留空间部分，所述用户数据部分被配置为存储用户数据，所述方法包括：

 压缩被存储在所述用户数据部分中的至少一些用户数据；

 确定由于所述压缩而恢复的存储容量的量；

 计算一个或更多个预留空间参数，所述一个或更多个预留空间参数包括在所述非易失性存储器阵列中存储的非冗余数据的量的测量值，所述非冗余数据包括不是永久地存储在除所述非易失性存储器阵列之外的所述数据存储系统的任何存储器中的数据；以及

 至少部分基于所述一个或更多个预留空间参数，将所述恢复的存储容量分配给以下项中的至少一个：所述用户数据部分的容量和所述预留空间部分的所述容量，

 其中，在控制器的控制下执行所述方法。

12. 根据权利要求11所述的方法，其中所述分配包括至少部分基于所述一个或更多个预留空间部分中的每个的组合分配所述恢复的存储容量。

13. 根据权利要求12所述的方法，其中，所述组合包括所述一个或更多个预留空间参数中的每个的加权平均值。

14. 根据权利要求11所述的方法，其中，所述一个或更多个预留空间参数包括以下项中的至少一个：

 用户数据压缩率；

 所述预留空间部分的所述容量；

 所述非易失性存储器阵列的磨损级别；

 不工作的非易失性存储器阵列单元的数量；以及

 从所述非易失性存储器阵列读取的数据量与写入所述非易失性存储器阵列的数据量的比率，写入所述非易失性存储器阵列的连续数据量与写入所述非易失性存储器阵列的非连续数据量的比率。

15. 根据权利要求14所述的方法，其中所述分配包括响应于以下项中的至少两个，分配至少一些所述恢复的存储容量给所述预留空间部分的所述容量：

 相比于磨损级别阈值的所述非易失性存储器阵列的所述磨损级别的增大；

 相比于读/写阈值的从所述非易失性存储器阵列读取的所述数据量与写入所述非易失

性存储器阵列的所述数据量的所述比率的增大;以及

相比于连续数据写阈值的写入所述非易失性存储器阵列的所述连续数据量与写入所述非易失性存储器阵列的所述非连续数据量的所述比率的减小。

16. 根据权利要求14所述的方法,进一步包括响应于以下项中的至少两个,分配至少一些所述恢复的存储容量给所述用户数据部分的所述容量:

相比于不工作单元阈值的不工作的非易失性存储器单元的所述数量的增大;

相比于数据压缩阈值的所述用户数据压缩率的减小;

相比于预留空间阈值的所述预留空间部分的所述容量的增大;以及

相比于连续数据写阈值的写入所述非易失性存储器阵列的所述连续数据量与写入所述非易失性存储器阵列的所述非连续数据量的所述比率的增大。

17. 根据权利要求11所述的方法,其中,所述压缩存储在所述用户数据部分中的所述至少一些用户数据包括使用无损压缩进行压缩。

18. 根据权利要求11所述的方法,进一步包括磁存储设备,并且其中,所述非易失性存储器阵列被配置为用于所述磁存储设备的缓存存储器。

19. 根据权利要求11所述的方法,其中,进一步包括响应于相比于非冗余数据阈值的测量值的增大,分配至少一些所述恢复的存储容量给所述用户数据部分的所述容量,所述测量值包括在所述非易失性存储器阵列中存储的非冗余数据的百分比。

20. 根据权利要求11所述的方法,其中,所述非易失性存储器阵列被配置为针对远程数据存储设备的缓存存储器。

一种数据存储系统及用于数据存储系统的动态预留空间方法

技术领域

[0001] 本公开涉及用于计算机系统的数据存储系统。具体而言，本公开涉及用于数据存储系统的动态预留空间 (overprovisioning)。

背景技术

[0002] 数据存储系统执行许多系统任务和管理操作，例如其正常操作过程中的垃圾收集、损耗均衡、坏块管理等。执行系统任务和管理操作涉及实质性的开销，例如在将非易失性固态存储器用于存储数据的情况下增大的写入放大。因此，希望提供更有效的机制以便执行管理操作。

附图说明

[0003] 将参考以下附图来说明体现本发明的各种特征的系统和方法，在附图中：

[0004] 图1示出了根据本发明一个实施例的实施动态预留空间的主机系统和数据存储系统的组合。

[0005] 图2示出了根据本发明一个实施例的预留空间参数。

[0006] 图3示出了根据本发明一个实施例的动态预留空间。

[0007] 图4示出了根据本发明一个实施例的动态预留空间过程的流程图。

具体实施方式

[0008] 尽管说明了特定实施例，但仅是示例性地呈现这些实施例，并非旨在限定保护的范围。实际上，本文所述的创新的方法和系统可以以各种其他形式来体现。而且，在不脱离保护范围的情况下，可以在本文所述的方法和系统的形式上做出各种省略、替换、以及变化。

[0009] 概述

[0010] 数据存储系统执行许多管理操作，例如其正常操作过程中的垃圾收集、损耗均衡、坏块管理等。执行管理操作涉及实质性的开销，例如在将非易失性固态存储器 (NVSM) 用于存储数据的情况下增大的写入放大。在某些情况下，为了改进数据存储系统的效率、寿命、以及性能，分配额外的存储器以执行系统任务和/或管理操作可能是有利的。但为系统和/或管理任务分配额外的存储器典型地以减少用于用户数据的存储容量为代价来进行。而数据存储系统典型地向主机系统报告给定存储容量，这个报告的存储容量通常不能在数据存储系统的操作期间被修改。

[0011] 本发明的实施例针对用于动态预留空间的系统和方法。数据存储系统可以为预留空间保留诸如NVSM缓存存储器的存储器的部分。预留空间部分可以用于有效地执行系统任务和/或管理操作。例如，预留空间部分可以用于减小例如与向NVSM缓存写入数据相关联的写入放大。数据存储系统可以通过压缩存储在NVSM缓存中的数据来恢复存储容量。取决于各种预留空间因素，恢复的存储容量的部分或全部量可以被分配用于预留空间或用于存储

用户数据。例如,当由于主机系统活动,应缓存在NVSM中的用户数据量增大时,部分或全部恢复的存储容量可以被用于存储用户数据。作为另一个示例,当NVSM的磨损级别超过阈值时,整个恢复的存储容量的部分可以被用于预留空间。恢复的存储容量的这种动态分配可以改进效率和性能。

[0012] 在一个实施例中,数据存储系统可以包括非易失性存储器阵列,其具有被配置为存储用户数据的用户数据部分。另外,可以保留预留空间部分。在压缩了用户数据后,可以确定恢复的存储容量的量。基于一个或多个预留空间参数,数据存储系统可以分配恢复的存储容量,用于存储用户数据和/或预留空间部分。

[0013] 系统概述

[0014] 图1示出了根据本发明一个实施例的实施基于优先级的垃圾收集的主机系统和数据存储系统的组合100。如所示的,数据存储系统120(例如混合盘驱动器)包括控制器130和非易失性存储器阵列150及磁存储设备160,其包括磁性介质164(例如传统的或叠瓦式(shingled))。非易失性存储器阵列150可以包括非易失性固态存储器(NVSM),例如闪存集成电路、硫属化合物RAM(C-RAM)、相变存储器(PC-RAM或PRAM)、可编程金属化单元RAM(PMC-RAM或PMCM)、Ovonic Unified Memory(OUM)、电阻RAM(RRAM)、NAND存储器(例如单级单元(SLC)存储器、多级单元(MLC)存储器、或其任意组合)、NOR存储器、EEPROM、铁电存储器(FeRAM)、磁阻RAM(MRAM)、其他分立NVM(非易失性存储器)芯片,或其任意组合。非易失性存储器阵列150可以包括一个或多个存储器区,例如块、页等。存储器区可以包括存储器单元。在一个实施例中,非易失性存储器阵列150可以充当用于磁存储设备160的缓存。数据存储系统120可以进一步包括其他类型的存储设备。在一个实施例中,磁存储设备160可以被配置为叠瓦式磁存储设备,非易失性存储器阵列150被配置为用作叠瓦式磁存储设备的介质缓存。

[0015] 控制器130可以被配置为从主机系统110的存储接口模块112(例如设备驱动器)接收数据和/或存储访问命令。由存储接口112传送的存储访问命令可以包括由主机系统110发出的写数据和读数据命令。读和写命令可以指定逻辑地址(例如逻辑块地址或LBA),其用于访问数据存储系统120。控制器130可以在非易失性存储器阵列150中执行接收的命令。

[0016] 数据存储系统120可以存储由主机系统110传送的数据。换句话说,数据存储系统120可以充当用于主机系统110的存储器存储设备。为了便于该功能,控制器130可以实施逻辑接口。逻辑接口可以将数据存储系统的存储器作为可以存储用户数据的逻辑地址集合(例如相连的地址)而呈现给主机系统110。在内部,控制器130可以将逻辑地址映射到非易失性存储器阵列150、磁存储设备160、和/或其他存储模块中的各种物理单元或地址。物理单元可以被配置为存储数据。控制器130包括动态预留空间模块132,其被配置为执行动态预留空间,以及数据压缩模块134,其被配置为压缩数据以便存储于非易失性存储器阵列150和/或磁存储设备160中。

[0017] 在其他实施例中,代替磁存储设备160,数据存储系统120可以包括另一类数据存储设备,例如第二非易失性存储器阵列。例如,非易失性存储器阵列150可以包括一类存储器,其提供比用于第二非易失性存储器阵列中的存储器类型更快的写/读性能。在一些实施例中,非易失性存储器阵列150可以充当到远端地点的数据存储设备的缓存,数据的同步可以通过一个或多个网络连接发生。

[0018] 动态预留空间

[0019] 图2示出了根据本发明一个实施例的预留空间参数200。如所示的，预留空间参数是：用户数据压缩率202、不工作的非易失性存储器阵列150单元的数量204、从非易失性存储器阵列150读取的数据量与写入非易失性存储器阵列150的数据量的比率206、写入非易失性存储器阵列150的连续数据量与写入非易失性存储器阵列150的非连续数据量的比率208、当前预留空间级别210、存储在非易失性存储器阵列150中的非冗余数据的百分比212、及非易失性存储器阵列150的磨损级别214。非冗余数据包括存储在非易失性存储器阵列150中、但没有与其他存储介质（非易失性存储器阵列150充当其缓存，例如磁存储设备160）同步的数据。可以使用额外的预留空间参数。预留空间参数可以由控制器130和/或动态预留空间模块132 和/或数据压缩模块134产生、跟踪、和/或更新。

[0020] 图3示出了根据本发明一个实施例的动态预留空间300。动态预留空间 300可以由控制器130和/或动态预留空间模块132和/或数据压缩模块134 执行。如所示的，非易失性存储器阵列150可以被分为用户数据部分154 和系统数据部分158。另外，存储器阵列150可以包括预留空间部分156。可以基于诸如图2的参数200的一个或多个预留空间参数，调整预留空间部分156的大小。在一个实施例中，可以基于诸如加权平均值的预留空间参数的组合，调整预留空间部分的大小。

[0021] 如图3所示的，在一个实施例中，未压缩的用户数据存储在用户数据部分154中。例如当由数据压缩模块134压缩数据时，数据存储系统120 恢复存储容量170的量。恢复的存储容量可以被分配用于存储用户数据和/ 或用于预留空间。

[0022] 图4是示出根据本发明一个实施例的动态预留空间的过程400的流程图。过程400可以由控制器130和/或动态预留空间模块132和/或数据压缩模块134执行。过程400在块402开始，在此它执行从主机系统110接收的一个或多个存储命令。例如，过程400可以执行包括用户数据的写入或程序命令。过程400转移到块404，在此，它压缩与存储命令相关联的用户数据。压缩度可以取决于用户数据的类型，较高压缩度可以表示将用户数据压缩到较大程度。例如，可以以高压缩度压缩未压缩的音频和/或视频数据。在一个实施例中，过程400使用无损压缩，例如Lempel-Ziv (LZ)。

[0023] 过程400转移到块406，在此它确定由于用户数据压缩而恢复的存储容量的量。在块408，过程400确定和/或更新预留空间参数，例如图2的参数200。过程400转移到块408，在此它确定如何将恢复的存储设备分配用于用户数据存储和/或预留空间。

[0024] 在一个实施例中，响应于相比于磨损级别阈值的非易失性存储器阵列 150的磨损级别的增大，过程400分配至少部分或全部恢复的存储容量用于预留空间。在此情况下，例如，磨损级别的增大指示非易失性存储器阵列 150受到磨损，分配存储容量用于预留空间可以通过减小写入放大而减少磨损率。另一方面，响应于相比于磨损级别阈值的非易失性存储器阵列150 的磨损级别的减小，过程400分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下，例如，由于非易失性存储器阵列150未受到磨损，因此希望给用户数据分配更多的存储空间以增强非易失性存储器缓存命中率。

[0025] 在一个实施例中，响应于相比于不工作单元阈值的不工作(或有故障) 非易失性存储器150单元数量的增大，过程400分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下，例如，分配可用存储器来存储用户数据可能是有利的。响应于非易失性存

储器阵列150中存储的非冗余数据相对于非冗余数据阈值的百分比的增大,过程400也分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下,例如,由于磁存储设备160不接受用于存储的数据(例如因为磁盘未旋转),主机系统110可以将非易失性存储器150用作数据缓存。稍后可以将至少部分缓存的用户数据转储清除(flush)(或同步)到例如磁存储设备160。响应于从非易失性存储器阵列150读取的数据量与写入非易失性存储器阵列的数据量的比率相对于读/写阈值的增大,过程400也分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下,例如,主机系统110可以执行更多的数据取回操作,执行这些操作导致小的或没有写入放大。另一方面,响应于从非易失性存储器阵列150读取的数据量与写入非易失性存储器阵列的数据量的比率相对于读/写阈值的减小,过程400分配至少部分或全部恢复的存储容量用于预留空间。在此情况下,例如,主机系统110可以执行与增大的写入放大相关联的更多数据程序操作。分配更多存储容量用于预留空间可以减小非易失性存储器阵列150的磨损。

[0026] 在一个实施例中,响应于相比于数据压缩阈值的数据压缩率的减小,过程400分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下,例如,用户数据可以被较少压缩,从而占用更多空间。可以分配更多空间用于存储用户数据。响应于相比于预留空间阈值的预留空间部分的大小的增大,过程400也分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下,例如,预留空间部分大小可能已经增长过大。响应于相比于连续数据写阈值的写入非易失性存储器阵列150的连续数据量与写入非易失性存储器阵列150的非连续数据量的比率的增大,过程400也分配至少部分或全部恢复的存储容量用于存储用户数据。在此情况下,例如,相比于写入非顺序数据,将连续或顺序数据写入非易失性存储器阵列150与较低的写入放大相关联。因而,需要较少的预留空间来实现希望的总写入放大。另一方面,响应于写入非易失性存储器阵列150的连续数据量与写入非易失性存储器阵列150的非连续数据量的比率相对于连续数据写阈值的减小,过程400分配至少部分或全部恢复的存储容量用于预留空间。在此情况下,例如,写入更多的非连续或随机数据,其与增大的写入放大相关联。因此,可以分配更多的存储容量用于预留空间。

[0027] 结论

[0028] 利用公开的动态预留空间系统和方法可以导致更有效地使用非易失性存储器,减小写入放大,增大缓存命中率等。可以获得改进的数据存储系统性能和增大的耐久性。

[0029] 其他变化

[0030] 本领域技术人员会意识到在一些实施例中可以使用另外的预留空间参数。另外,使用任意适合的线性和/或非线性方法可以组合预留空间参数。此外,公开的系统和方法可以由任何数据存储系统使用,其例如由于存储介质的限制而不能写随机存储器单元。这种数据存储系统也可以包括缓存存储器。此外,用户数据可以包括任何类型数据和/或数据类型的组合,例如由主机提供的数据、由数据存储系统产生的数据等。在公开的过程中实际进行的步骤,例如图4中所示的过程,可以与图中所示的不同。可以使用额外的系统部件,可以组合或省略公开的系统部件。取决于实施例,可以去除上述的某些步骤,可以添加其他步骤。因此,本公开的范围旨在仅参照所附权利要求书来限定。

[0031] 尽管说明了某些实施例,但这些实施例仅被示例性地呈现,并非旨在限定保护的范围。实际上,本文所述的创新的方法和系统可以以各种其他形式来体现。而且,在不脱离

保护的精神的情况下,可以在本文所述的方法和系统的形式上做出各种省略、替换和变化。所附权利要求书及其等效替代旨在覆盖这样的形式或修改,视其为落在保护的范围和精神内。例如,本文公开的系统和方法可以应用于硬盘驱动器、固态驱动器等。另外,可以额外地或可替换地使用其他形式的存储设备(例如DRAM或SRAM、配有电池的易失性DRAM或SRAM设备、EPROM、EEPROM存储器等)。作为另一个示例,图中所示的各种部件可以被实施为处理器上的软件和/或固件、ASIC/FPGA、或专用硬件。此外,以上公开的特定实施例的特征和属性可以以不同方式组合,以构成另外的实施例,其全都落在本公开的范围内。尽管本公开提供了某些优选实施例和应用,但对于本领域普通技术人员来说是显而易见的是其他实施例也在本公开的范围内,包括没有提供本文阐述的全部特征和优点的实施例。因此,本公开的范围旨在仅参照所附权利要求书来限定。

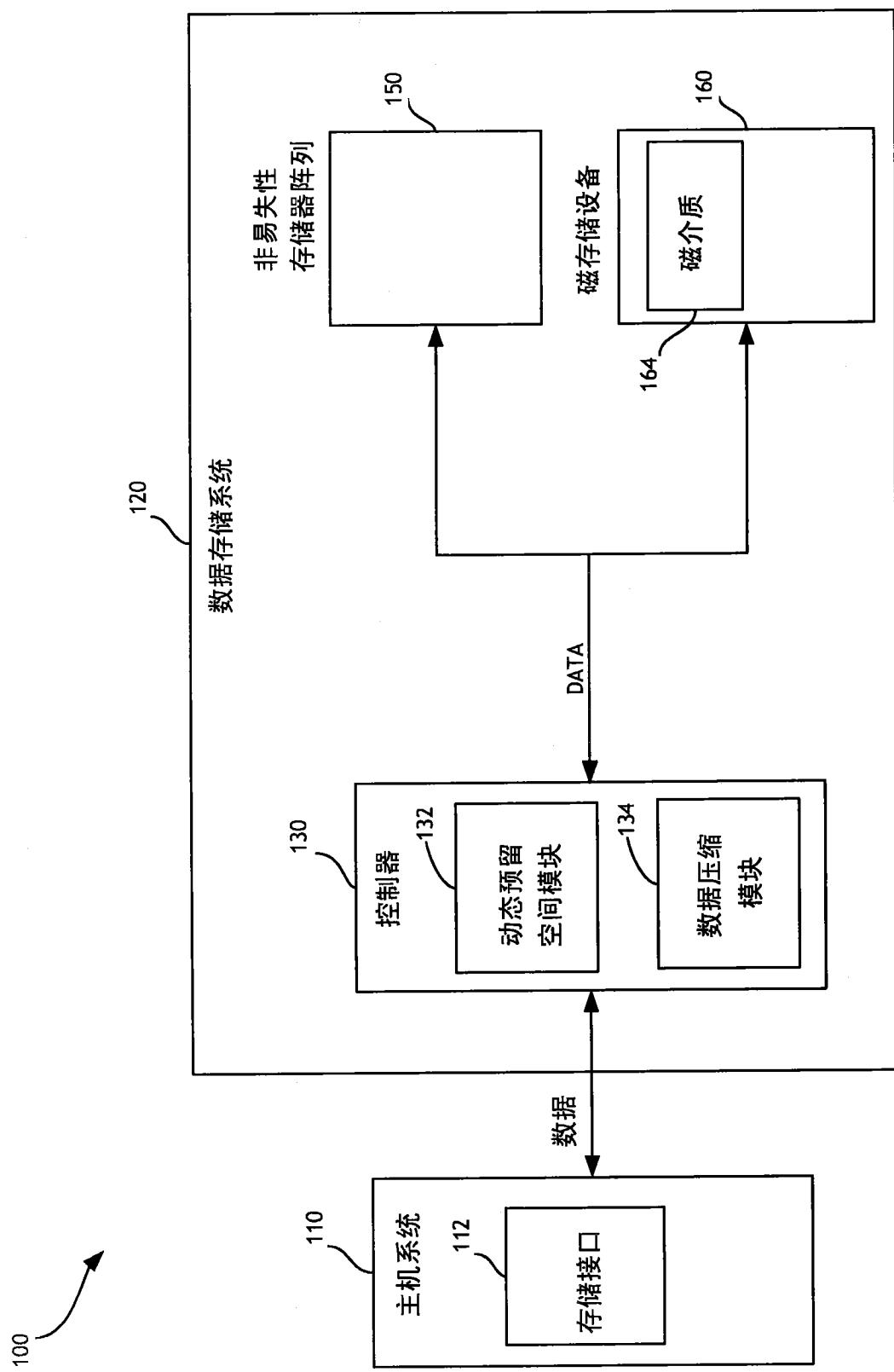


图 1

200
↓

202	数据压缩率
204	有故障的NVM阵列单元的数量
206	读/写比率
208	写入的顺序数据与非顺序数据的比率
210	当前预留空间级别
212	存储在NVM阵列中的非冗余数据的百分比
214	NVM阵列的磨损级别
...	...

图2

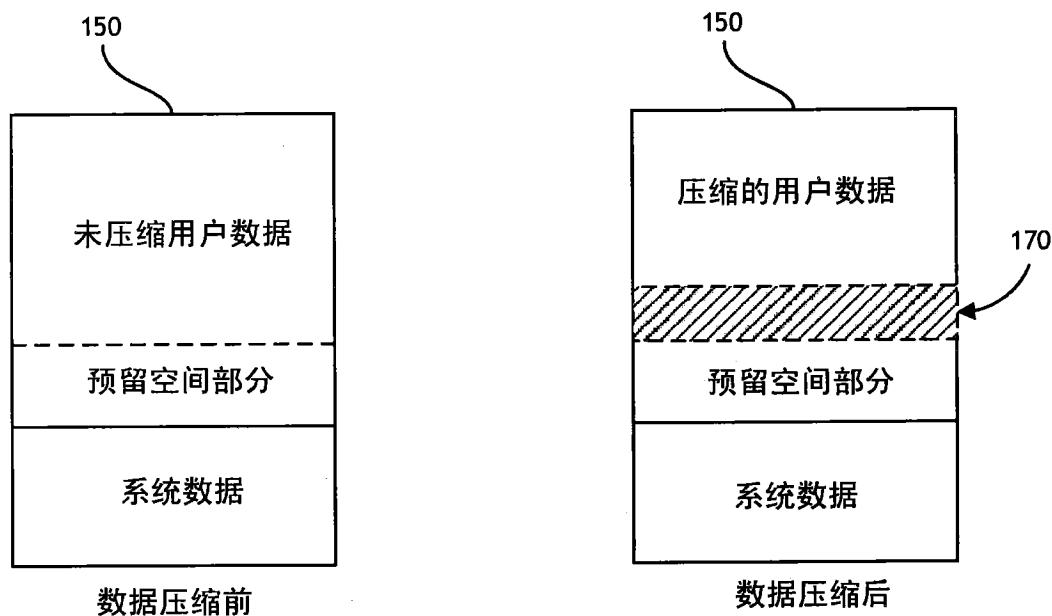
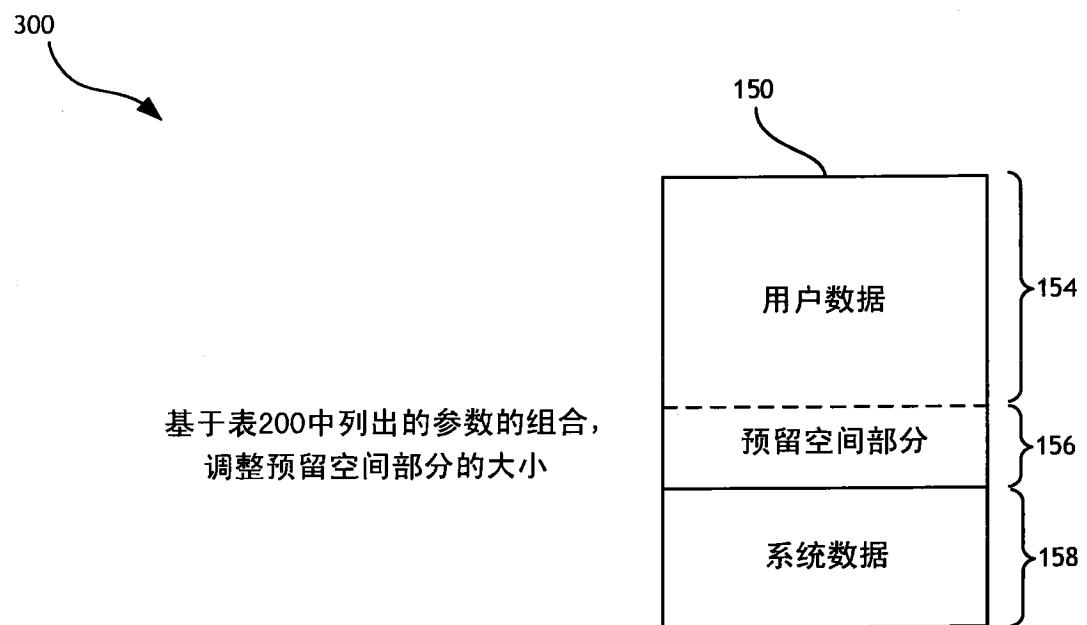


图3

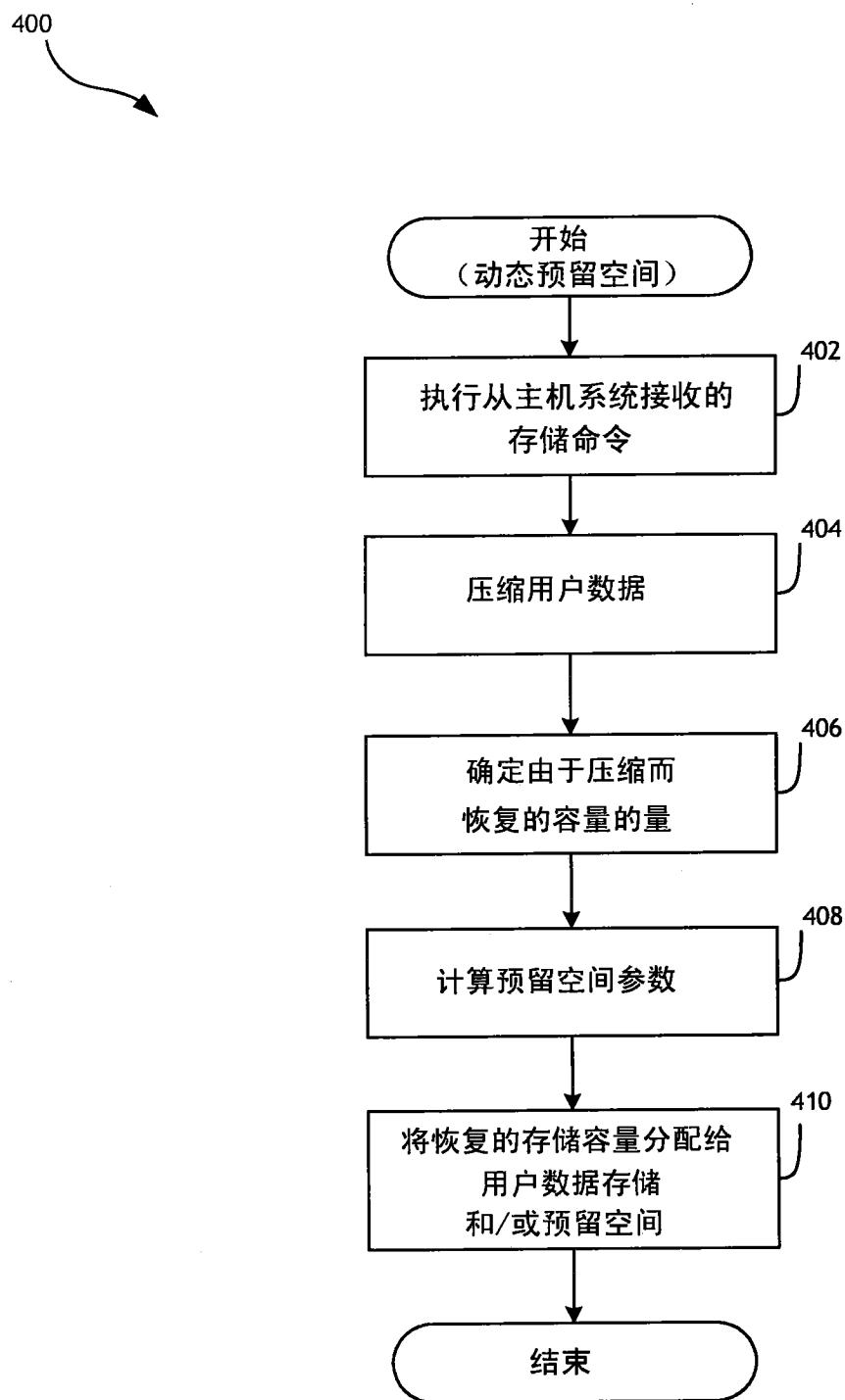


图4