(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105718213 A
(43)申请公布日 2016.06.29

(21)申请号 201510666998.3

(22)申请日 2015.10.15

(30)优先权数据
62/069,241 2014.10.27 US
14/677,662 2015.04.02 US

(71)申请人 桑迪士克科技股份有限公司
地址 美国得克萨斯州

(72)发明人 S. 斯普劳斯 S.B. 瓦萨德瓦
R. 布里特纳

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 邱军

(51)Int.Cl.
G06F 3/06(2006.01)

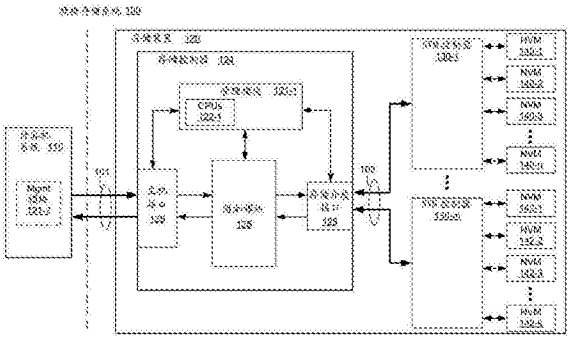
权利要求书2页 说明书12页 附图4页

(54)发明名称

改进在低队列深度工作负载中的混合随机性能的方法

(57)摘要

一种用于能够改进在存储装置（例如，包括多个非易失性存储器单位，诸如一个或多个闪存装置）中的低队列深度工作负载中的混合随机性能的系统、方法和/或装置。在一个方面，方法包含(1)保持相应于来自主机的写入命令的写入缓存，(2)根据来自主机的命令确定工作负载，(3)根据确定工作负载是不合格的工作负载，安排写入缓存的定期冲洗，以及(4)根据确定工作负载是合格的工作负载，安排写入缓存的优化的冲洗。



CN 105718213 A

1.一种管理存储系统的方法,方法包括:

保持相应于来自主机的写入命令的写入缓存,存储系统的存储装置被可操作地耦合到主机,所述存储装置包含多个非易失性存储器单位,其中写入缓存包含来自主机的、但尚未保存到非易失性存储器单位的写入命令,所述写入命令被映射用于多个非易失性存储器单位中的存储;

根据来自主机的命令确定工作负载,所述存储装置被可操作地耦合到主机;

根据确定工作负载是不合格的工作负载,安排写入缓存的定期冲洗,其中所述定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位;以及

根据确定工作负载是合格的工作负载,安排写入缓存的优化的冲洗,其中优化的冲洗包含冲洗第一倍数的预定数量的数据到相应数目的多个非易失性存储器单位,其中第一倍数是大于1的整数。

2.如权利要求1的方法,其中定期冲洗包含从写入命令累积足够的数据以将预定数量的数据写入到多个非易失性存储器单位的一个非易失性存储器单位,并且优化的冲洗包含从写入命令累积足够的数据以将预定数量的数据的第一倍数写入到相应数目的多个非易失性存储器单位。

3.如权利要求1的方法,还包括,在确定工作负载已经从合格的工作负载改变为不合格的工作负载之后,执行一个或多个定期冲洗,每个定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位。

4.如权利要求3的方法,其中执行一个或多个定期冲洗包含并行执行多个定期冲洗。

5.如权利要求1-4的任一项的方法,还包括,根据确定主机已经空闲至少预定数量的时间,冲洗写入缓存到多个非易失性存储器单位的一个或多个。

6.如权利要求1-4的任一项的方法,还包括,根据确定主机已经空闲至少预定数量的时间,并且根据确定写入缓存含有至少非零整数倍的预定数量的数据,冲洗非零整数倍的预定数量的数据到多个非易失性存储器单位的一个或多个。

7.如权利要求1-4的任一项的方法,其中存储装置包括一个或多个闪存装置。

8.一种存储装置,包括:

用于保持相应于来自主机的写入命令的写入缓存的部件,存储系统的存储装置被可操作地耦合到主机,所述存储装置包含多个非易失性存储器单位,其中写入缓存包含来自主机的、但尚未保存到非易失性存储器单位的写入命令,所述写入命令被映射用于在多个非易失性存储器单位中的存储;

用于根据来自主机的命令确定工作负载的部件,所述存储装置被可操作地耦合到主机;

用于根据确定工作负载是不合格的工作负载来安排写入缓存的定期冲洗的部件,其中所述定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位;以及

用于根据确定工作负载是合格的工作负载来安排写入缓存的优化的冲洗的部件,其中优化的冲洗包含冲洗第一倍数的预定数量的数据到相应数目的多个非易失性存储器单位,其中第一倍数是大于1的整数。

9.如权利要求8的存储装置，包含用于执行定期冲洗的部件，包含从写入命令累积足够的数据以将预定数量的数据写入到多个非易失性存储器单位的一个非易失性存储器单位，以及用于执行优化的冲洗的部件，包含从写入命令累积足够的数据以将第一倍数的预定数量的数据写入到相应数目的多个非易失性存储器单位。

10.如权利要求8的存储装置，还包含在确定工作负载已经从合格的工作负载改变为不合格的工作负载之后执行一个或多个定期冲洗的部件，每个定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位。

11.如权利要求10的存储装置，其中执行一个或多个定期冲洗的部件包含用于并行执行多个定期冲洗的装置。

12.如权利要求8-11的任一项的存储装置，还包括，根据确定主机已经空闲至少预定数量的时间，冲洗写入缓存到多个非易失性存储器单位的一个或多个的部件。

13.如权利要求8-11的任一项的存储装置，还包括，根据确定主机已经空闲至少预定数量的时间并且根据确定写入缓存含有至少非零整数倍的预定数量的数据，冲洗非零整数倍的预定数量的数据到多个非易失性存储器单位的一个或多个的部件。

14.如权利要求8-11的任一项的存储装置，其中存储装置包含一个或多个闪存装置。

改进在低队列深度工作负载中的混合随机性能的方法

技术领域

[0001] 公开的实施例一般涉及存储器系统，尤其涉及，改进在存储装置(例如，包括一个或多个闪存装置)中的低队列深度工作负载中的混合随机性能。

背景技术

[0002] 包含闪存的半导体存储器装置通常利用存储器单元存储数据为电值(诸如电荷或电压之类)。闪存单元，例如，包含具有浮置栅极的单个晶体管，该浮置栅极用于存储代表数据值的电荷。闪存是可以被电擦除和重编程的非易失性数据存储装置。更一般地，与易失性存储器相反，非易失性存储器(例如，闪存，以及使用各种各样的技术的任何一种实现的其它类型的非易失性存储器)即使不供电也保留存储的信息，这需要电力以保持存储的信息。

[0003] 由于一些存储器命令(例如，读取和/或写入命令)需要在预定数量的时间内完成(例如，为了满足资格要求)，优化存储器管理过程以提高存储器装置的性能是很重要的。

发明内容

[0004] 所附权利要求的范围内的各种实施例的系统、方法和装置每个具有几个方面，几个方面中没有单个的一个独自负责这里描述的属性。不限制所附权利要求的范围，在考虑本公开之后，特别是在考虑部分标题为“具体实施方式”，人们将理解各种实施例的各方面是如何用于能够改进在低队列深度工作负载中的混合随机性能。在一个方面，根据确定工作负载是不合格的工作负载，写入缓存的定期冲洗被安排，并且根据确定工作负载是合格的工作负载，写入缓存的优化的冲洗被安排。

附图说明

[0005] 因此本公开可以更详细地理解，通过参考各种实施例的特征可能具有更特别的描述，其中一些在附图中说明。附图，然而，仅仅说明了本公开的相关特征，并且因此不应被认为是限制性的，因为该描述可能允许其它有效的特征。

[0006] 图1示出了根据一些实施例的数据存储系统的实现方式的框图。

[0007] 图2示出了根据一些实施例的管理模块的实现方式的框图。

[0008] 图3A-3B示出了根据一些实施例的管理存储系统的方法的流程图表示。

[0009] 根据惯例，附图中说明的各种特征可能不按比例绘制。因此，各种特征的尺寸为了清楚可能任意扩大或缩小。此外，附图的一些可能未描述给定的系统、方法或装置的所有组件。最后，相同的附图标记可能用于表示贯穿说明书和附图的相同的特征。

具体实施方式

[0010] 这里所描述的各种实施例包含用于能够改进在低队列深度工作负载中的混合随机性能的系统、方法和/或装置。包含系统、方法和/或装置的一些实施例以根据确定工作负载是不合格的工作负载来安排写入缓存的定期冲洗，并且根据确定工作负载是合格的工作

负载来安排写入缓存的优化的冲洗。

[0011] (A1)更具体地，一些实施例包含一种管理存储系统的方法。在一些实施例中，方法包括：(1)保持相应于来自主机的写入命令的写入缓存，存储系统的存储装置被可操作地耦合到主机，存储装置包含多个非易失性存储器单位，其中写入缓存包含来自主机的、但尚未保存到非易失性存储器单位的写入命令，写入命令被映射用于在多个非易失性存储器单位中的存储，(2)根据来自主机的命令确定工作负载，存储装置被可操作地耦合到主机，(3)根据确定工作负载是不合格的工作负载，安排写入缓存的定期冲洗，其中定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位，以及(4)根据确定工作负载是合格的工作负载，安排写入缓存的优化的冲洗，其中优化的冲洗包含冲洗第一倍数的预定数量的数据到相应数目的多个非易失性存储器单位，其中第一倍数是大于1的整数。

[0012] (A2)在A1的的方法的一些实施例中，定期冲洗包含从写入命令累积足够的数据以将预定数量的数据写入到多个非易失性存储器单位的一个非易失性存储器单位，并且优化的冲洗包含从写入命令累积足够的数据以将第一倍数的预定数量的数据写入到相应数目的多个非易失性存储器单位。

[0013] (A3)在A1和A2的任一个的方法的一些实施例中，方法还包含，随后确定工作负载已经从合格的工作负载改变为不合格的工作负载，执行一个或多个定期冲洗，每个定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位。

[0014] (A4)在A3的的方法的一些实施例中，执行一个或多个定期冲洗包含并行执行多个定期冲洗。

[0015] (A5)在A1至A4的任一个的方法的一些实施例中，方法还包含，根据确定主机已经空闲至少预定数量的时间，冲洗写入缓存到多个非易失性存储器单位的一个或多个。

[0016] (A6)在A1至A4的任一个的方法的一些实施例中，方法还包含，根据确定主机已经空闲至少预定数量的时间，并且根据确定写入缓存含有至少非零整数倍的预定数量的数据，冲洗非零整数倍的预定数量的数据到多个非易失性存储器单位的一个或多个。

[0017] (A7)在A1至A6的任一个的方法的一些实施例中，存储装置包括一个或多个闪存装置。

[0018] (A8)在另一个方面，存储装置包含(1)非易失性存储器(例如，包括一个或多个非易失性存储装置，诸如闪存装置)，(2)一个或多个处理器，以及(3)存储一个或多个程序的控制器存储器(例如，在控制器中或耦合到控制器的非易失性存储器或易失性存储器)，当该一个或多个程序被一个或多个处理器执行时引起存储装置执行或控制这里描述的方法A1至A7的任一个的执行。

[0019] (A10)在另一个方面，如上的方法A1至A7的任一个由存储装置来执行，存储装置包含用于执行这里描述的方法的任一个的装置。

[0020] (A12)在另一个方面，存储系统包含(1)存储介质(例如，包括一个或多个非易失性存储装置，诸如闪存装置)，(2)一个或多个处理器，以及(3)存储一个或多个程序的存储器(例如，在存储系统中的非易失性存储器或易失性存储器)，当该一个或多个程序被一个或多个处理器执行时引起存储系统执行或控制这里描述的方法A1至A7的任一个的执行。

[0021] (A13)在另一个方面，一些实施例包含非临时性计算机可读存储介质，其存储配置成通过存储装置的一个或多个处理器执行的一个或多个程序，一个或多个程序包含用于执行这里描述的方法的任一个的指令。

[0022] 这里描述了很多的细节以便于提供在附图中示出的示例的实现例的完全理解。然而，一些实施例可以在没有很多具体的细节的情况下实施，并且权利要求的保护范围仅由在权利要求中明确列举的那些特征和方面来限制。此外，已知的方法、组件和电路没有以详尽的细节描述以便不必要地模糊这里描述的实现方式的更多相关的方面。

[0023] 包含下面描述的那些的数据存储系统使用多种技术以避免由各种故障机制引起的数据丢失，该故障机制包含存储介质故障、通信故障、和在系统和子系统级上的故障。这些机制的共同特征是使用数据冗余以保护数据，补偿实际的和潜在的数据错误(例如，介质错误、丢失数据、传输错误、不可访问数据等)。一类冗余机制被称为错误校正编码(error correction codes，ECC)。许多类型的错误校正编码是已知的(例如，BCH、LDPC、Reed-Solomon等)，作为用于存储具有被保护的数据的它们或与被保护的数据相结合的它们的许多方案。另一类的冗余机制是擦除编码(例如，角锥体(pyramid)、喷泉(fountain)、部分MDS、本地可修复的(locally repairable)、简单的再生等)

[0024] 另一种类型或级别的冗余机制典型地被称为RAID(独立磁盘冗余阵列)，即使当存储介质不是传统意义上的“盘”。有多种形式的RAID、或RAID方案，提供不同级别的数据保护(例如，RAID-1、RAID-5、RAID-6、RAID-10等)。典型地，在使用RAID的系统中，存储在多个不同的存储位置的数据的“条带”被当作一组，并且存储有本将已经丢失的条带中的任何数据的足够的冗余数据，在存储位置的任一个的部分或完全故障中，使用条带中的其他数据来恢复，可能包含冗余数据。

[0025] 图1示出了根据一些实施例的数据存储系统100的框图。虽然一些示例的特征被示出，为了简洁的目的，各种其他特征没有被示出，以免模糊这里公开的示例的实施例的相关方面。为此，作为非限制性的示例，数据存储系统100包含存储装置120(有时也被称为信息存储装置，或数据存储装置，或者存储器装置)，其包含存储控制器124、诸如闪存控制器的一个或多个非易失性存储器(non-volatile memory，NVM)控制器130、以及非易失性存储器(例如，一个或多个NVM装置140、142，诸如一个或多个闪存装置)，并且连同计算机系统110使用。在一些实施例中，存储装置120包含单个NVM装置，而在其他实施例中，存储装置120包含多个NVM装置。在一些实施例中，NVM装置140、142包含NAND类型闪存或者NOR类型闪存。此外，在一些实施例中，NVM控制器130是固态驱动器(solid-state drive，SSD)控制器。然而，根据多种实施例的各方面(例如，PCRAM、ReRAM、STT-RAM等)，可能包含其他类型的存储介质。在一些实施例中，闪存装置包含一个或多个闪存裸片、一个或多个闪存封装、一个或多个闪存通道等等。在一些实施例中，数据存储系统100可以含有一个或多个存储装置120。

[0026] 计算机系统110通过数据连接101耦合至存储控制器124。然而，在一些实施例中，计算机系统110包含存储控制器124、或者存储控制器124的部分，作为组件和/或子系统。例如，在一些实施例中，存储控制器124的一些或所有功能通过在计算机系统110上执行的软件来实现。计算机系统110可能是任何合适的计算机装置，例如计算机、笔记本电脑、平板设备、上网本、网络信息站、个人信息站、手机、智能电话、游戏设备、计算机服务器、或任何其他计算设备。计算机系统110有时称为主机、主机系统、客户端、或客户端系统。在一些实施

例中，计算机系统110是服务器系统，诸如在数据中心的服务器系统。在一些实施例中，计算机系统110包含一个或多个处理器，一种或多种类型的存储器，显示器和/或其他用户界面组件，例如，键盘、触摸屏显示器、鼠标、触控板、数字照相机和/或任何数量的补充设备以添加功能。在一些实施例中，计算机系统110不具有显示器和其他用户接口组件。

[0027] 一个或多个NVM控制器130通过连接103与存储控制器124耦合。连接103有时称为数据连接，但是通常传达除了数据之外的命令，并且可选地传达元数据、错误校正信息和/或除了要存储在NVM装置140、142中的数据值和从NVM装置140、142读取的数据值之外的其他信息。在一些实施例中，然而，存储控制器124、一个或多个NVM控制器130、和NVM装置140、142被包含在相同的装置中(例如，集成装置)作为其组件。此外，在一些实施例中，存储控制器124、一个或多个NVM控制器130、和NVM装置140、142被嵌入在主机装置中(例如，计算机系统110)，诸如移动装置、平板、其它计算机或计算机控制装置，并且这里描述的方法被执行，至少部分地，由嵌入的存储控制器。

[0028] 在一些实施例中，存储装置120包含NVM装置140、142，诸如闪存装置(例如，NVM装置140-1到140-n，以及NVM装置142-1到142-k)和NVM控制器130(例如，NVM控制器130-1到130-m)。从另一个方面看，存储装置120包含m个存储器通道，每一个存储器通道具有NVM控制器130和耦合到NVM控制器130的一组NVM装置140或142，其中m是大于1的整数。然而，在一些实施例中，两个或更多的存储器通道共享NVM控制器130。在任一示例中，每个存储器通道具有它自己的不同组的NVM装置140或142。在非限制性示例中，在典型的存储装置中的存储器通道的数目是8、16或32。在另一个非限制性示例中，每个存储器信道的NVM装置140或142的数目典型地是8、16、32或64。此外，在一些实施例中，NVM装置140/142的数目在不同的存储器通道中是不同的。

[0029] 在一些实施例中，NVM控制器130的每个NVM控制器包含配置成在一个或多个程序中执行指令(例如，在NVM控制器130中)的一个或多个处理单位(有时也称为CPU或处理器或微处理器或微控制器)。在一些实施例中，一个或多个处理器由内部的一个或多个组件共享，并且在一些情况下，超出NVM控制器130的功能。NVM装置140、142通过连接被耦合到NVM控制器130，该连接通常传达除了数据之外的命令，并且可选地传达元数据、错误校正信息和/或除了要存储在NVM装置140、142中的数据值和从NVM装置140、142读取的数据值之外的其他信息。NVM装置140、142可能包含任何数量(即，一个或多个)的存储器装置，存储器装置包含，但不限于，非易失性半导体存储器装置，例如闪存装置。

[0030] 例如，闪存装置(例如，NVM装置140、142)可以配置为适于诸如云计算的应用的企业存储、数据库应用、主要存储和/或辅助存储、或者用于存储在(或要存储在)辅助存储器中的超高速缓存数据，例如硬盘驱动器。此外和/或可选择，闪存装置(例如，NVM装置140、142)也可以配置为相对较小规模的应用，例如，个人闪存驱动器或替代个人、笔记本电脑和平板电脑的硬盘。虽然闪存装置和闪存控制器在这里作为示例使用，在一些实施例中存储装置120包含其他非易失性存储器装置和相应的非易失性存储控制器。

[0031] 在一些实施例中，NVM装置140、142被分成若干可寻址和可单独选择的块。在一些实施例中，单独可选择的块是闪存装置中最小尺寸的可擦除单元。换句话说，每个块包含可被同时擦除的存储器单元的最小数量。每个块通常被进一步分成多个页面和/或字线，其中每个页面或字线通常是在块中的最小单独可访问的(可读取的)的部分的一个实例。然而，

在一些实施例中(例如,使用一些类型的闪速存储器),数据组的最小单独可访问的单元是扇区,其是一个页面的子单元。即,块包含多个页面,每个页面包含多个扇区,并且每个扇区是数据的最小单元,用于从闪存装置读取数据。

[0032] 如上,当非易失性半导体存储器装置的数据存储密度通常增加时,增加存储密度的缺点是存储的数据是更易于错误地存储和/或读取。在一些实施例中,可以利用差错控制编码以限制由电波动引入的不可校正的差错的数量、存储介质中的缺陷、操作条件、装置历史、写入-读取电路,等,或者这些和各种其它因素的组合。

[0033] 在一些实施例中,存储控制器124包含管理模块121-1、主机接口129、存储介质(I/O)接口128和附加模块125。存储控制器124可能包含各种附加的特征,为了简洁的目的没有示出附件的特征,以免模糊这里所公开的示例的实施例的相关的特征,并且特征的不同排列是可能的。

[0034] 主机接口129通过数据连接101为计算机系统110提供接口。类似地,存储介质接口128通过连接103为NVM控制器130提供接口。在一些实施例中,存储介质接口128包含读取和写入电路,其中包含能够将读取信号提供给NVM控制器130(例如,为NAND类型闪存读取阈值电压)的电路。在一些实施例中,连接101和连接103作为通信介质使用诸如DDR3、SCSI、SATA、SAS等的协议被实现,在通信介质上命令和数据被通信。在一些实施例中,存储控制器124包含配置成在一个或多个程序中执行指令(例如,在存储控制器124中)的一个或多个处理单位(有时也称为CPU或处理器或微处理器或微控制器)。在一些实施例中,一个或多个处理器由内部(并且在一些情况下,超出存储控制器124的功能)的一个或多个组件共享。

[0035] 在一些实施例中,管理模块121-1包含一个或多个处理单元(CPU,有时也称为处理器或微处理器或微控制器)122,该处理单元122配置为在一个或多个程序中(例如,在管理模块121-1中)执行指令。在一些实施例中,一个或多个CPU 122由内部的一个或多个组件共享,并且在一些情况下,超出存储控制器124的功能。为了协调这些组件的操作,管理模块121-1被耦合到主机接口129、附加模块125和存储介质接口128。在一些实施例中,管理模块121-1的一个或多个模块在计算机系统110的管理模块的121-2中实现。在一些实施例中,计算机系统110(未示出)的一个或多个处理器被配置成在一个或多个程序中(例如,在管理模块121-2中)执行指令。为了管理存储装置120的操作,管理模块121-2被耦合到存储装置120。

[0036] 附加模块125被耦合到存储介质接口128、主机接口129、和管理模块121-1。作为示例,附加模块125可能包含错误控制模块,以限制在写入存储器和/或从存储器读取的期间无意中引入数据的不可校正的错误的数目。在一些实施例中,附加模块125中通过管理模块121-1的一个或多个CPU 122在软件中被执行,并且,在其它实施例中,附加模块125使用专用电路实现全部或部分(例如,执行编码和解码的功能)。在一些实施例中,附加模块125通过在计算机系统110上执行的软件实现全部或部分。

[0037] 在一些实施例中,包含在附加模块125中的错误控制模块包含编码器和解码器。在一些实施例中,编码器通过施加错误控制编码(ECC)来编码数据以产生码字,该码字随后被存储在NVM装置140、142中。当从NVM装置140、142读取编码数据(例如,一个或多个码字)时,解码器为编码数据施加解码处理以恢复数据,并且在错误控制编码的错误校正能力之内校正恢复的数据中的错误。本领域的技术人员将理解,各种错误控制编码具有不同的错误检

测和校正能力，以及出于超出本公开的范围的理由，特定编码被选择用于各种应用。正因如此，错误控制编码的各种类型的详尽回顾在这里不提供。此外，本领域的技术人员将理解，每种类型或每个族的错误控制编码可能具有编码和解码的算法，特别是错误控制编码的类型或族。另一方面，一些算法可能在若干不同类型或族的错误控制编码的解码中被利用至少某种程度。正因如此，为了简洁起见，本领域的技术人员一般可得的和已知的各种类型的编码和解码算法的详尽描述在这里不提供。

[0038] 在一些实施例中，在写入操作期间，主机接口129从计算机系统110接收要存储在NVM装置140、142中的数据。由主机接口129接收的数据使编码器可得的(例如，在附加模块125中)，该编码器编码数据以产生一个或多个码字。使得一个或多个码字对于存储介质接口128可用，存储介质接口128以依赖于所利用的存储介质的类型的方式将一个或多个码字传输给NVM装置140、142(例如，通过NVM控制器130)。

[0039] 在一些实施例中，当计算机系统(主机)110将一个或多个主机读取命令(例如，经由数据连接101，或者可选择地单独的控制线或总线)发送到从NVM装置140、142请求数据的存储控制器124时，读取操作启动。存储控制器124经由存储介质接口128将一个或多个读取访问命令发送到NVM装置140、142(例如，通过NVM控制器130)，以根据由一个或多个主机读取命令规定的存储器位置(地址)获得原始读取数。存储介质接口128将原始读取数据(例如，包含一个或多个码字)提供给解码器(例如，在附加模块125中)。如果解码成功，解码数据被提供给主机接口129，其中解码数据对计算机系统110是可得的。在一些实施例中，如果解码不成功，存储控制器124可能诉诸于若干补救的行动或提供无法解决的错误条件的指示。

[0040] 如上，存储介质(例如，NVM装置140、142)被分成若干可寻址的和单独的选择块，且每个块可选地(但典型地)被进一步分成多个页面和/或字线和/或扇区。虽然存储介质的擦除是以块为基础来执行的，在很多实施例中，存储介质的读取和编程在块的较小的亚单位上执行(例如，以页面、字线、或扇区为基础)。在一些实施例中，块的较小的亚单位由多个存储器单元(例如，单级单元或多级单元)组成。在一些实施例中，编程在整个页面上执行。在一些实施例中，多级单元(multi-level cell，MLC)NAND闪存每单元典型地具有四种可能的状态，每单元产生两位信息。此外，在一些实施例中，MLC NAND具有两个页面类型：(1)较低的页面(有时称为快页面)，以及(2)上部页面(有时称为慢页面)。在一些实施例中，三级单元(triple-level cell，TLC)NAND闪存每单元具有八种可能的状态，每单元产生三位信息。虽然这里的描述使用TLC、MLC和SLC作为示例，本领域的技术人员将理解，这里描述的实施例可能被扩展到存储器单元每单元具有多于八个可能的状态，每单元产生多于三位的信息。在一些实施例中，存储介质的编码格式(例如，TLC、MLC或SLC和/或选择的数据冗余机制)是当数据被实际写入到存储介质时做出的选择。

[0041] 作为示例，如果数据以页面被写入到存储介质中，但是存储介质在块中被擦除，在存储介质中的页面可能含有无效(例如，过时)的数据，但那些页面不能被覆盖直到含有那些页面的整个块被擦除。为了写入具有无效数据的页面，在该块中的具有有效数据的页面(如果有的话)被读取且重新写入到新的块，并且旧的块被擦除(或放在用于擦除的队列上)。该过程被称为垃圾收集。在垃圾收集后，新的块含有具有有效数据的页面，并且可能具有空闲页面，该空闲页面对要被写入的新的数据可用，且旧的块可以被擦除，以便对要被写

入的新的数据可用。由于闪存只可以被编程和擦除有限数目的次数，用于挑选下一个块以重新写入和擦除的算法的效率对寿命和基于闪存的存储系统的可靠性具有显著的影响。

[0042] 写入放大是一种现象，其中写入到存储介质(例如，在存储装置120中的NVM装置140、142)的物理数据的实际数量是由主机(例如，计算机系统110，有时称为主机)写入到存储介质的数据的逻辑数量的倍数。如上，当存储介质的块在它可以被重新写入之前必须被擦除时，执行这些操作的垃圾收集过程导致重新写入数据一次或多次。该倍增的效果增加了需要的写入的数目，其超出了存储介质的寿命，从而缩短了它可以可靠地操作的时间。计算存储系统的写入放大的公式由以下方程式给出：

[0043] $$\frac{\text{写入存储介质的数据的数量}}{\text{由主机写入的数据的数量}}$$

[0044] 基于数据储存系统架构的任何闪存的目标之一是尽可能减少写入放大，使得可得的耐力用于满足存储介质的可靠性和担保规格。由于存储系统可能需要更少的过度配置，更高的系统的耐力还导致更低的成本。通过减少写入放大，存储介质的耐力增加，并且存储系统的总成本降低。一般地，垃圾收集在具有用于最佳性能的有效页面的最少数目和最佳写入放大的擦除块上被执行。

[0045] 闪存装置利用存储器单元以存储数据为电值(诸如电荷或电压)。每个闪存单元典型地包含具有用于存储电荷的浮置栅极的单个晶体管，电荷修改了晶体管的阈值电压(即，打开晶体管所需要的电压)。电荷的幅度，和电荷创建的相应的阈值电压，用于表示一个或多个数据值。在一些实施例中，在读取操作期间，读取阈值电压被施加到晶体管的控制栅极，并且产生的感测电流或电压被映射到数据值。

[0046] 在闪存单元的背景下，词语“单元电压”和“存储器单元电压”意味着存储器单元的阈值电压，该阈值电压是为了使晶体管传导电流而需要施加给存储器单元的晶体管的栅极的最小电压。类似地，施加给闪存单元的读取阈值电压(有时也称为读取信号和读取电压)是施加给闪存单元的栅极的栅极电压以确定存储器单元在该栅极电压下是否传导电流的栅极电压。在一些实施例中，当闪存单元的晶体管在给定的读取阈值电压下传导电流时，指示单元电压小于读取阈值电压，该读取操作的原始数据值是“1”，否则原始数据值是“0”。

[0047] 图2示出了根据一些实施例的管理模块121-1的框图，如图1所示。管理模块121-1典型地包含用于执行存储在存储器206中的模块、程序和/或指令从而执行处理操作的一个或多个处理单位(有时称为CPU或处理器)122-1、存储器206(有时称为控制器存储器)、以及用于互连这些组件的一个或多个通信总线208。一个或多个通信总线208可选地包含互连并且控制系统组件之间的通信的电路(有时称为芯片组)。管理模块121-1通过一个或多个通信总线208被耦合到主机接口129、附加模块125和存储介质I/O 128。存储器206包含高速随机存取存储器，诸如DRAM、SRAM、DDR RAM或其他随机存取固态存储器装置，并且可能包含非易失存储器，诸如一个或多个磁盘存储装置、光盘存储装置、闪存装置，或其它非易失性固态存储装置。存储器206可选地包含位于CPU 122-1远程的一个或多个存储装置。存储器206，或作为选择地在存储器206内的非易失性存储器装置，包括非临时性计算机可读存储介质。在一些实施例中，存储器206，或存储器206的非临时性计算机可读存储介质存储下列程序、模块和数据结构，或者其子集或超集：

[0048] 转译表212，其用于将逻辑地址映射到物理地址；

[0049] 数据读取模块214，其用于从存储介质中(例如，NVM装置140、142，图1)的一个或多个码字、页面或块读取数据；

[0050] 数据写入模块216，其用于向存储介质中(例如，NVM装置140、142，图1)的一个或多个码字、页面或块写入数据；

[0051] 数据擦除模块218，其用于从存储介质中(例如，NVM装置140、142，图1)的一个或多个块擦除数据；

[0052] 垃圾收集模块220，其用于存储介质中(例如，NVM装置140、142，图1)的一个或多个块的垃圾收集；

[0053] 命令接收模块222，其用于从主机接收多个命令(例如，未映射的命令和I/O命令，诸如写入请求和/或读取请求)；

[0054] 工作负载模块224，其用于根据来自主机(例如，计算机系统110，图1)的命令确定工作负载(或工作负载中的变化)；

[0055] 写入缓存模块236，其用于保持、冲洗和/或安排写入缓存(例如，写入缓存238)的冲洗(例如，定期冲洗和/或优化的冲洗)；以及

[0056] 写入缓存238，其包含与来自主机的写入命令相应的数据结构的集合。

[0057] 上面标识的元件的每个可能被存储在先前提到的存储器装置的一个或多个中，并且相应于用于执行上述功能的一组指令。上面标识的模块或程序(即，指令组)不必被实现为分开的软件程序、进程或模块，并且因此这些模块中的各种子集可能被组合或以其他方式重新布置在各个实施例中。在一些实施例中，存储器206可能存储上面标识的模块和数据结构的子集。此外，存储器206可能存储上面未描述的附加模块和数据结构。在一些实施例中，存储在存储器206中的程序、模块和数据结构，或者存储器206的非临时性计算机可读存储介质，提供用于实现下面描述的一些方法的指令。在一些实施例中，一些或全部模块可能用归入部分或者全部的模块功能性的专用硬件电路来实现。

[0058] 尽管图2示出了根据一些实施例的管理模块121-1，图2比作为这里描述的实施例的结构原理图，更意在作为可能存在于管理模块121-1中的各种特征的功能性描述。实际上，由于被本领域的技术人员意识到，分别示出的程序、模块和数据结构可以被组合，并且一些程序、模块和数据结构可以被分开。

[0059] 在一些实施例中，对于具有低队列深度(例如，4个队列深度)的混合的读取/写入随机工作负载(例如，30％的写入和70％的读取)，来自主机的写入命令可以阻塞来自主机的读取命令，因为写入操作比读取操作更慢(例如，采取显著地更长的时间来完成)，引起主机在等待写入完成时连续失速。这里所描述的各种实施例包含用于能够改进在低队列深度工作负载(有时称为排位工作负载)中的混合随机性能的系统、方法和/或装置。在一些实施例中，存储器装置包括工作在第一模式(例如，读取模式)或者第二模式(例如，写入模式)的一个或多个非易失性存储器装置(例如，NVM装置140、142，图1)。在一些实施例中，在读取模式期间，所有的写入操作(例如，主机写入和垃圾收集写入)被缓冲在支持电容器的RAM(例如，DRAM或SRAM，有时被称为写入缓存)中，直到预定的标准被满足(例如，直到缓冲区满，直到有足够的数据已经累计完成RAID条等)。在这个时间期间，读取命令被迅速服务，因为没有阻塞的写入操作。在一些实施例中，当满足预定的标准时(例如，当缓冲器满时，当足够的数据已产生以完成RAID条等)，存储装置切换到写入模式，考虑到任何系统级功耗限制，在

此期间存储装置发出尽可能多的并行写入。当存储装置处于写入模式时发出的任何读取将被阻塞，直到写入完成。

[0060] 图3A-3B示出了根据一些实施例的管理存储系统的方法300的流程图表示。至少在一些实施例中，方法300由存储装置(例如，存储装置120，图1)或者存储装置的一个或多个组件(例如，存储控制器124，NVM控制器130和/或NVM装置140、142，图1)来执行，其中存储装置可操作地与主机系统(例如，计算机系统110，图1)耦合。在一些实施例中，方法300由指令支配，该指令存储在非临时性计算机可读存储介质中且由装置的一个或多个处理器来执行，诸如管理模块121-1的一个或多个处理单位(CPU)122-1，在图1和图2中示出。在一些实施例中，方法300由存储系统(例如，数据存储系统100，图1)或存储系统的一个或多个组件(例如，计算机系统执行110和/或存储装置120，图1)来执行。在一些实施例中，方法300的一些操作中在主机处执行(例如，计算机系统110，图1)，并且信息被传送到存储装置(例如，存储装置120，图1)。在一些实施例中，方法300被指令支配，至少部分地，该指令存储在非临时性计算机可读存储介质中且由主机的一个或多个处理器执行(图1中未示出)。为了易于说明，下面将方法300描述为由存储装置(例如，存储装置120，图1)来执行。然而，本领域的技术人员将理解，在其他实施例中，在方法300中描述的一个或多个操作由主机(例如，计算机系统110，图1)来执行。

[0061] 存储系统(例如，数据存储系统100，图1)的存储装置(例如，存储装置120，图1)保持(302)相应于来自主机(例如，计算机系统110，图1)的写入命令(例如，写入缓存228，图2)的写入缓存，存储装置被可操作地耦合到主机，该存储装置包含多个非易失性存储器单位(例如，NVM装置140、142，图1)的存储装置，其中写入缓存包含来自主机的、但尚未保存到非易失性存储器单位的写入命令，该写入命令被映射用于在多个非易失性存储器单位中的存储。在一些实施例中，多个非易失性存储器单位的非易失性存储器单位包含裸片(例如，闪存裸片)。在一些实施例中，多个非易失性存储器单位的非易失性存储器单位包含芯片(例如，具有两个或多个闪存裸片的闪存芯片)。在一些实施例中，多个非易失性存储器单位的非易失性存储器单位包含可以与其他存储器单位(例如，位面)并行编程的存储器单位。在一些实施例中，写入缓存模块(例如，写入缓存模块228，图2)用于保持相应于来自主机的写入命令的写入缓存，存储系统的存储装置被可操作地耦合到该主机，该存储装置包含多个非易失性存储器单位的存储装置，其中写入缓存包含来自主机的但尚未保存到非易失性存储器单位的写入命令，写入命令被映射用于在多个非易失性存储器单位中的存储，如上面关于图2的描述。

[0062] 在一些实施例中，存储装置包含(304)一个或多个闪存装置。在一些实施例中，存储装置包含存储介质(例如，NVM装置140、142，图1)，并且存储介质包含一个或多个非易失性存储装置，诸如闪存装置。在一些实施例中，存储介质是单个闪存装置，而在其他实施例中，存储介质包含多个闪存装置。例如，在一些实施例中，存储介质包含在并行存储器通道中组织的数十或数百的闪存装置，诸如每一个存储器通道16、32或64个闪存装置，以及8、16或32个并行存储器通道。在一些实施例中，非易失性存储介质(例如，NVM装置140、142，图1)包含NAND类型闪存或者NOR类型闪存。在其他实施例中，存储介质包括一种或多种其它类型的非易失性存储装置。

[0063] 存储装置根据来自主机的命令确定(306)工作负载，该存储装置被可操作地耦合

到主机。在一些实施例中，工作负载是根据来自主机的未解决的命令的队列深度来确定的。在一些实施例中，工作负载是根据来自主机的写入命令的百分比相对于来自主机的读取命令的百分比来确定的。在一些实施例中，工作负载模块(例如，工作负载模块224，图2)被用来根据来自主机的命令确定工作负载，存储装置被可操作地耦合到该主机，如关于图2的上面的描述。

[0064] 根据确定工作负载是不合格的工作负载，存储装置安排(308)写入缓存的定期冲洗，其中定期冲洗包含从写入缓存冲洗预定数量的数据(例如，数据的块)到多个非易失性存储器单位(NVM装置140、142，图1)的一个非易失性存储器单位(例如，NVM装置140-1，图1)。在一些实施例中，不合格的工作负载是当队列深度大于预定的深度阈值。例如，在一些实施例中，不合格的工作负载是当队列深度大于32。在一些实施例中，不合格的工作负载是当来自主机的写入命令的百分比相对于读取命令的百分比大于预定阈值。在一些实施例中，写入缓存模块(例如，写入缓存模块226，图2)被用来根据确定工作负载是不合格的工作负载安排写入缓存的定期冲洗，其中定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位，如关于图2的上面的描述。

[0065] 根据确定工作负载是合格的工作负载，存储装置安排(310)写入缓存的优化冲洗，其中优化的冲洗包含冲洗第一倍数的预定数量的数据到相应数目的多个非易失性存储器单位，其中第一倍数是大于1的整数。例如，在一些实施例中，写入缓存的优化的冲洗包含冲洗数据的n个块到n个非易失性存储器单位(例如，冲洗数据的一个块到NVM装置140-1、NVM装置140-2、...NVM装置140-n的每一个)。在一些实施例中，第一倍数是大于7的整数(即，8或更大)，且典型地具有介于8和32之间的值。在一些实施例中，合格的工作负载是当队列深度小于或等于预定的深度阈值。例如，在一些实施例中，合格的工作负载是当队列深度是“低”的(例如，1到32的队列深度)。在一些实施例中，写入缓存模块(例如，写入缓存模块226，图2)被用来根据确定工作负载是合格的工作负载安排写入缓存的优化的冲洗，其中优化的冲洗包含冲洗预定数量的数据的第一倍数到相应数目的多个非易失性存储器单位，其中第一倍数是大于1的整数，如关于图2的上面的描述。

[0066] 在一些实施例中，定期冲洗包含(312)从写入命令累积足够的数据以将预定数量的数据写入到多个非易失性存储器单位的一个非易失性存储器单位，并且优化的冲洗包含从写入命令累积足够的数据以将第一倍数的预定数量的数据写入到相应数目的多个非易失性存储器单位。例如，在一些实施例中，定期冲洗包含从写入命令累积足够的数据以将数据的块写入到一个非易失性存储器单位，并且优化的冲洗包含从写入命令累积足够的数据以将数据的n个块写入到n个非易失性存储器单位，其中n是大于1的整数。在一些实施例中，相应数目等于第一倍数，或这相应数目等于由整数功率除以2的第一倍数(例如，如果非易失性存储器单位的两个或多个部分可以被并行编程)。

[0067] 在一些实施方式中，在确定工作负载已经从合格的工作负载改变为不合格的工作负载之后，存储装置执行(314)一个或多个定期冲洗，每个定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位。在一些实施例中，写入缓存模块(例如，写入缓存模块226，图2)被用来在确定工作负载已经从合格的工作负载改变为不合格的工作负载之后执行一个或多个定期冲洗，每个定期冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位，如关于图2

的上面的描述。

[0068] 在一些实施例中，执行一个或多个定期冲洗(314)包含并行(316)执行(316)多个定期冲洗。在一些实施例中或在一些情形下，在操作314中执行的所有的定期冲洗被并行执行。在一些实施例中，“并行”地执行多个写入缓存冲洗被定义为指在至少部分重叠的时间期间内执行多个写入缓存。

[0069] 可选地，在一些实施例中，存储装置根据确定主机已经空闲至少预定数量的时间冲洗(318)写入缓存到多个非易失性存储器单位的一个或多个。在一些实施例中，根据确定主机已经空闲至少预定数量的时间，存储装置执行定期冲洗。在一些实施例中，或在一些情形下(例如，写入缓存存储足够的数据以执行优化的冲洗)，根据确定主机已经空闲至少预定数量的时间，存储装置执行优化的冲洗。在一些实施例中，根据确定主机已经空闲至少预定数量的时间，存储装置冲洗全部写入缓存到多个非易失性存储器单位的一个或多个。在一些实施例中，写入缓存模块(例如，写入缓存模块226，图2)被用来冲洗，根据确定主机已经空闲至少预定数量的时间，写入缓存到多个非易失性存储器单位的一个或多个，如关于图2的上面的描述。

[0070] 可选地，在一些实施例中，存储装置根据确定主机已经空闲至少预定数量的时间并且根据确定写入缓存含有至少非零整数倍的预定数量的数据，冲洗(320)非零整数倍的预定数量的数据到多个非易失性存储器单位的一个或多个。例如，在一些实施例中，如果写入缓存含有两个半的块的价值的数据，存储装置冲洗数据的两个块到多个非易失性存储器单位的一个或多个。在一些实施例中，写入缓存模块(例如，写入缓存模块226，图2)被用来根据确定主机已经空闲至少预定数量的时间并且根据确定写入缓存含有至少非零整数倍的预定数量的数据，冲洗非零整数倍的预定数量的数据到多个非易失性存储器单位的一个或多个，如关于图2的上面的描述。

[0071] 应该理解的是，虽然词语“第一”，“第二”等在这里可以用于描述各种元件，但是这些元件不应该由这些词语限制。这些词语仅用于彼此区分元件。例如，第一区域可以叫做第二区域，并且，类似地，第二区域可以叫做第一区域，没有改变描述的意义，只要“第一区域”的所有出现一致地改名，并且“第二区域”的所有出现一致地改名。第一区域和第二区域都是区域，但它们不是相同的区域。

[0072] 这里使用的术语只是为了描述特定的实施例的目的，不意在限制权利要求。如实施例和所附权利要求的描述中使用的，单数形式“一个”和“一”意在同样包含复数形式，否则除非上下文清楚地指示。还可以理解的是，这里使用的词语“和/或”指的是和包含一个或多个相关联的列出的项目的任何或所有可能的组合。还可以理解的是，当在本说明书中使用词语“包括”和/或“包含”时，说明存在陈述的特征、整数、步骤、操作、元件、和/或组件，但是不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件、和/或它们的群组。

[0073] 如这里使用的，短语“A、B和C的至少一个”被解释为要求所列项目的一个或多个，并且该阶段单独在A的单个实例上、单独在B的单个实例上、或单独在C的单个实例上读取，同时还包含的所列项目的组合，诸如“A的一个或多个以及B的一个或多个B没有任何的C，”等。

[0074] 如这里所使用的，依赖于上下文，词语“如果”可能解释成意为“当”或者“根据”或

者“响应于确定”或者“根据确定”或者“响应于检测”，陈述的先前条件是真实的。类似地，依赖于上下文，短语“如果确定[陈述的先前条件是真实的]”或者“如果[陈述的先前条件是真实的]”或者“当[陈述的先前条件是真实的]”可能解释成意为“根据确定”或者“响应于确定”或者“根据检测”或者“响应于检测”陈述的先前条件是真实的。

[0075] 上述的描述，为了说明的目的，已经参考具体的实施例被描述。然而，上述的说明不旨在穷尽或限制权利要求为公开的精确形式。根据上述教导的许多变型和变化是可能的。选择和描述实施例以便最好的说明操作原则和实际应用，从而使本领域的其他人员精通。

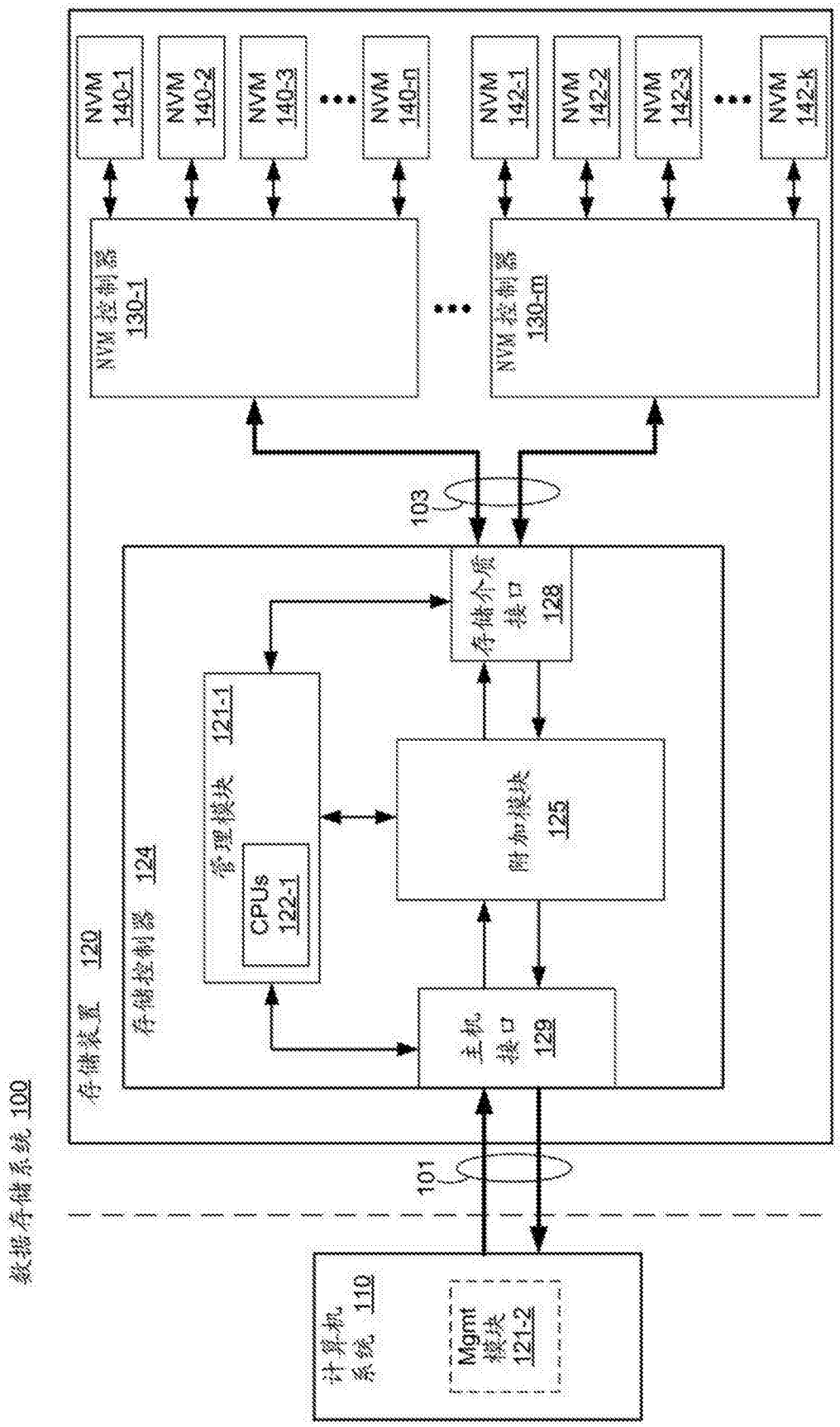

数据存储系统 100
计算机系统 110
Mgmt 模块 121-2
101
存储装置 120
主机接口 129
存储控制器 124
管理模块 121-1
CPUs 122-1
附加模块 125
存储介质接口 128
103
NVM 控制器 130-1
NVM 控制器 130-m
NVM 140-1
NVM 140-2
NVM 140-3
NVM 140-n
NVM 142-1
NVM 142-2
NVM 142-3
NVM 142-k

图1

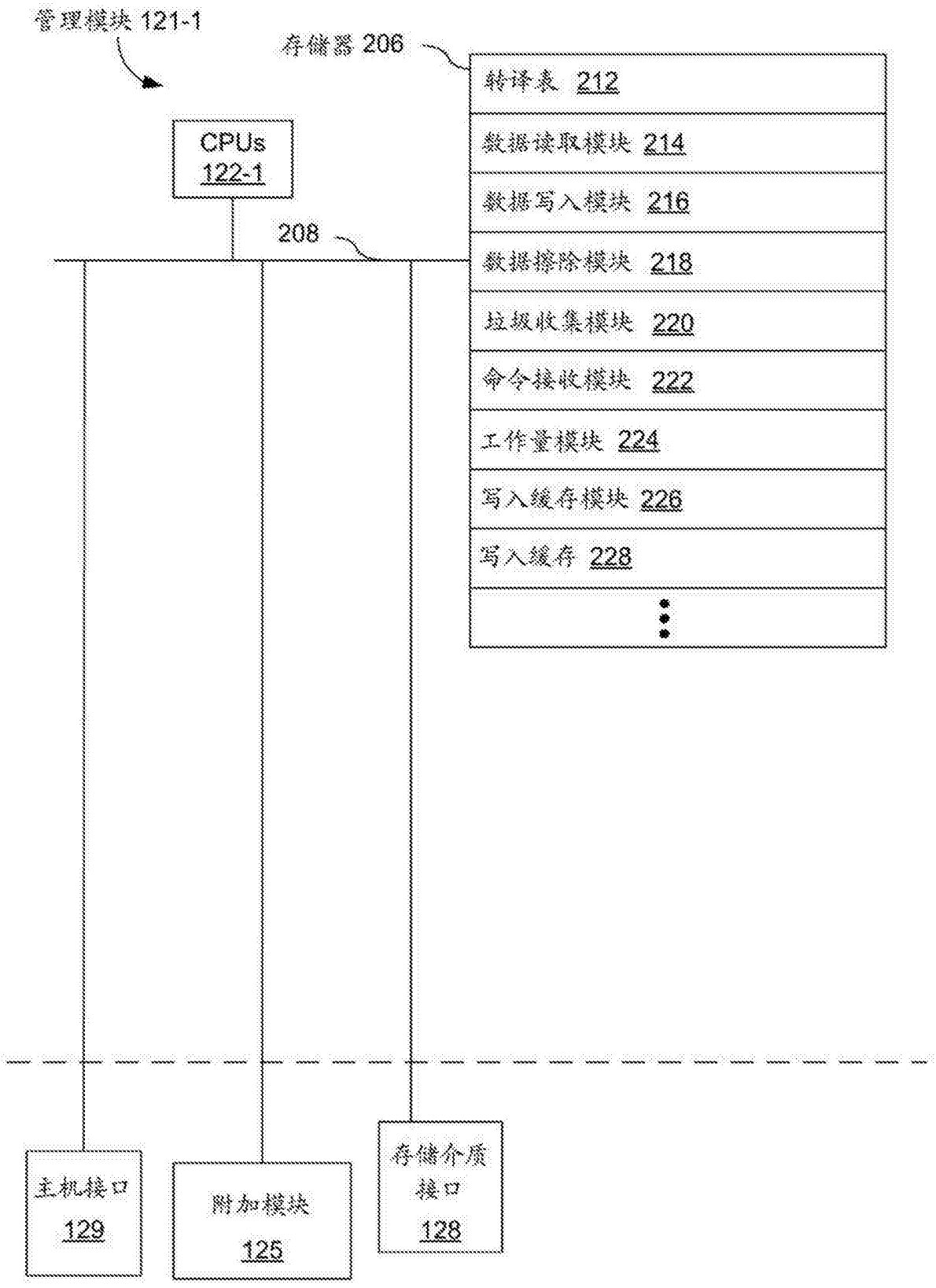

管理模块 121-1
存储器 206
转译表 212
数据读取模块 214
数据写入模块 216
数据擦除模块 218
垃圾收集模块 220
命令接收模块 222
工作量模块 224
写入缓存模块 226
写入缓存 228
CPUs 122-1
208
主机接口 129
附加模块 125
存储介质接口 128

图2

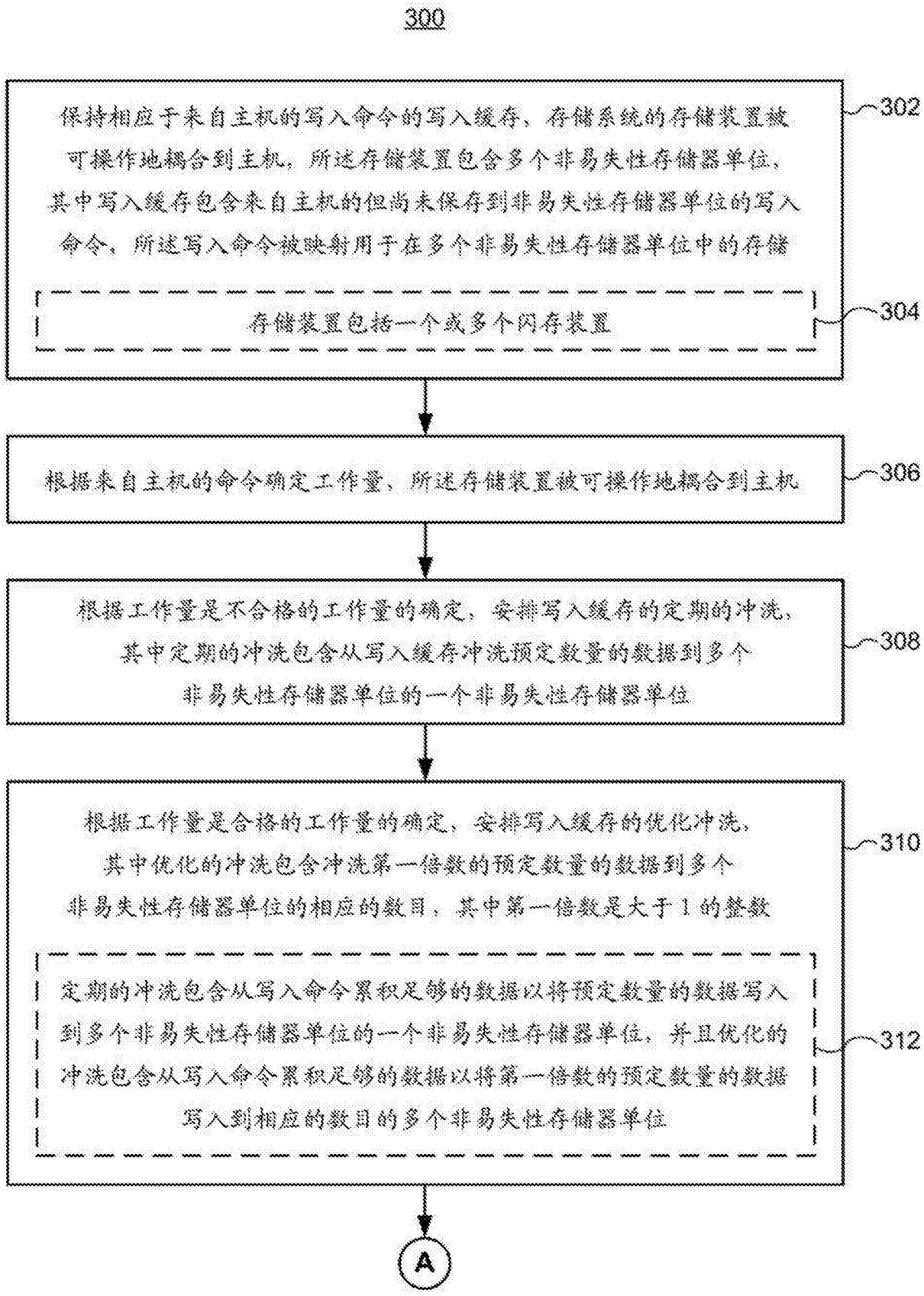

300
保持相应于来自主机的写入命令的写入缓存，存储系统的存储装置被可操作地耦合到主机，所述存储装置包含多个非易失性存储器单位，其中写入缓存包含来自主机的但尚未保存到非易失性存储器单位的写入命令，所述写入命令被映射用于在多个非易失性存储器单位中的存储
302
存储装置包括一个或多个闪存装置
304
根据来自主机的命令确定工作量，所述存储装置被可操作地耦合到主机
306
根据工作量是不合格的工作量的确定，安排写入缓存的定期的冲洗，其中定期的冲洗包含从写入缓存冲洗预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位
308
根据工作量是合格的工作量的确定，安排写入缓存的优化冲洗，其中优化的冲洗包含冲洗第一倍数的预定数量的数据到多个非易失性存储器单位的相应的数目，其中第一倍数是大于1的整数
310
定期的冲洗包含从写入命令累积足够的数据以将预定数量的数据写入到多个非易失性存储器单位的一个非易失性存储器单位，并且优化的冲洗包含从写入命令累积足够的数据以将第一倍数的预定数量的数据写入到相应的数目的多个非易失性存储器单位
312
A

图3A

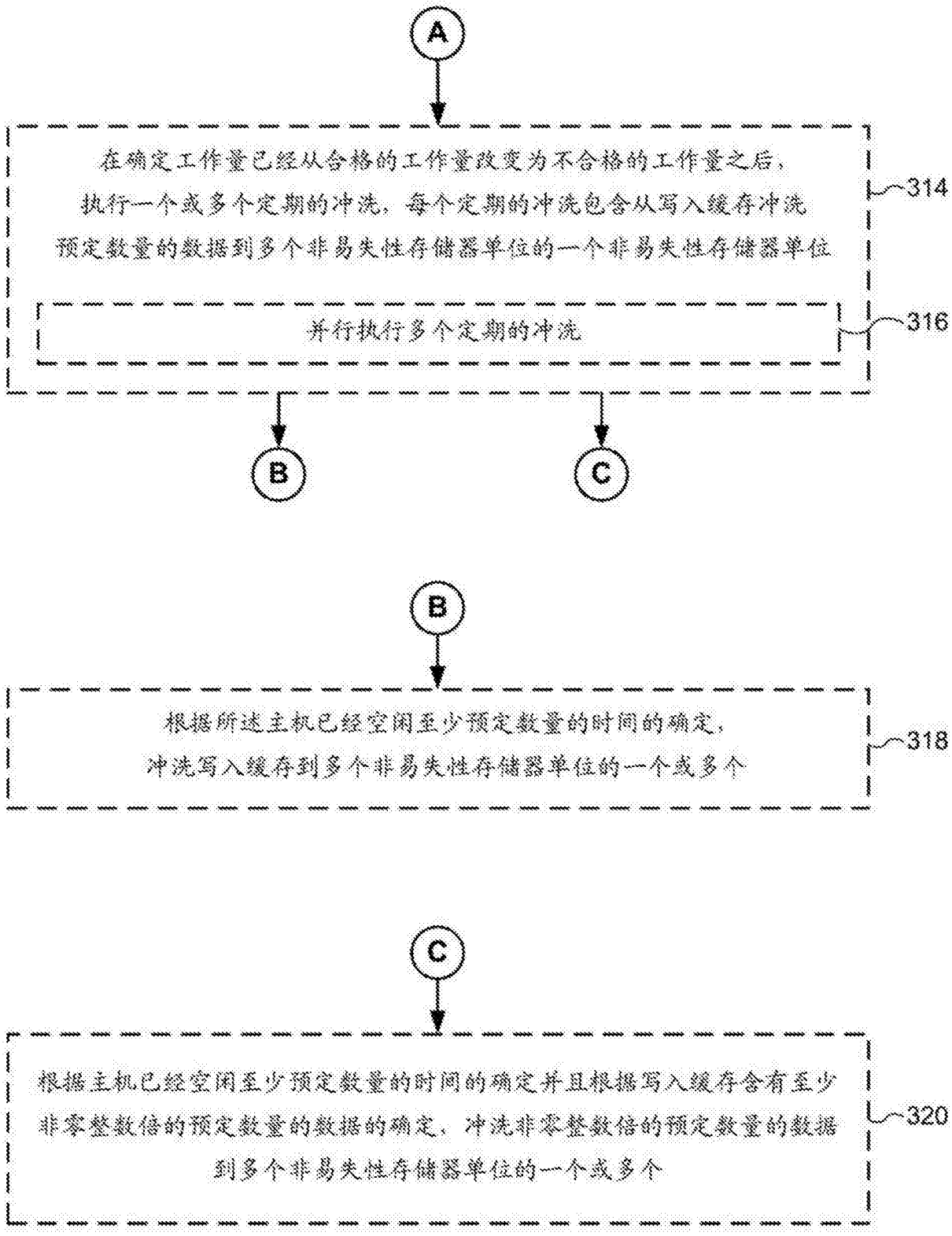

A
在确定工作量已经从合格的工作量改变为不合格的工作量之后，
执行一个或多个定期的冲洗，每个定期的冲洗包含从写入缓存冲洗
预定数量的数据到多个非易失性存储器单位的一个非易失性存储器单位
314
并行执行多个定期的冲洗
316
B
C
B
根据所述主机已经空闲至少预定数量的时间的确定，
冲洗写入缓存到多个非易失性存储器单位的一个或多个
318
C
根据主机已经空闲至少预定数量的时间的确定并且根据写入缓存含有至少
非零整数倍的预定数量的数据的确定，冲洗非零整数倍的预定数量的数据
到多个非易失性存储器单位的一个或多个
320

图3B