



(12) 发明专利

(10) 授权公告号 CN 102713866 B

(45) 授权公告日 2015. 09. 30

(21) 申请号 201080057092. 7

(51) Int. Cl.

(22) 申请日 2010. 12. 14

G06F 12/02(2006. 01)

(30) 优先权数据

(56) 对比文件

12/637, 897 2009. 12. 15 US

US 2008/0250270 A1, 2008. 10. 09, 第 [0017]-[0101] 段及附图 1-9.

(85) PCT国际申请进入国家阶段日

US 2008/0250270 A1, 2008. 10. 09, 第 [0017]-[0101] 段及附图 1-9.

2012. 06. 15

(86) PCT国际申请的申请数据

US 6069827 A, 2000. 05. 30, 说明书第 3 栏第 19 行 - 第 7 栏第 8 行及附图 1-3.

PCT/IB2010/055802 2010. 12. 14

(87) PCT国际申请的公布数据

CN 1930635 A, 2007. 03. 14, 全文.

W02011/073902 EN 2011. 06. 23

审查员 徐菲

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 E·S·埃列夫特里奥 R·哈斯

P·米勒 R·A·普莱卡

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 高青

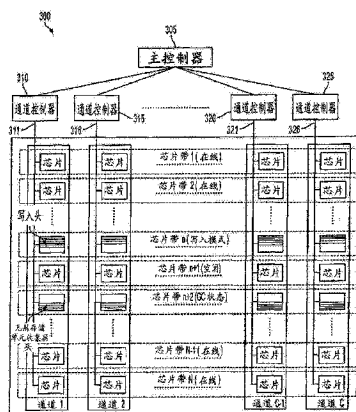
权利要求书3页 说明书7页 附图6页

(54) 发明名称

减少基于闪存的存储系统中的存取争用

(57) 摘要

示例性实施例包括一种用于减少基于闪存的存储系统中的存取争用的方法,该方法包括:从具有多个通道和多个存储块的存储设备中选择空闲状态的芯片带,其中,该芯片带包括多个页面;将芯片带设置为写入状态;设置多个通道中的每个通道中的写队列头;对于闪存带中的多个通道的每个通道,将写队列头设置到来自芯片带的属于该通道的芯片中的第一空闲页面;根据写入分配调度器,在通道间分配写入请求;生成页面写入,以及响应于页面写入,递增写队列头;以及当芯片带已满时,将芯片带设置为在线状态。



1. 一种用于减少存储芯片中的存取争用的方法,所述方法包括:  
从具有多个通道和多个存储块的存储设备中选择空闲状态的芯片带,其中,所述芯片带包括多个页面;  
将所述芯片带设置为写入状态,所述芯片带的写入状态与擦除状态相分离;  
设置所述多个通道中的每个通道中的写队列头;  
对于所述芯片带中的所述多个通道中的每个通道,  
将写队列头设置到来自所述芯片带的属于该通道的芯片中的第一空闲页面;  
根据写入分配调度器,在所述通道间分配写入请求;  
生成页面写入;  
响应于所述页面写入,递增所述写队列头;以及  
当所述芯片带已满时,将所述芯片带设置为在线状态。
2. 如权利要求 1 所述的方法,进一步包括触发无用存储单元收集。
3. 如权利要求 1 所述的方法,其中,所述多个存储块以与所述存储芯片上的所述多个存储块的相应地址相同的顺序,被保持在所述存储芯片上的空闲块队列中。
4. 如权利要求 2 所述的方法,其中,响应于所述芯片带是被改变为所述写入状态的最后一个空闲芯片带,触发无用存储单元收集。
5. 如权利要求 2 所述的方法,其中,响应于达到所述存储芯片上的空闲状态芯片带的预定阈值,触发无用存储单元收集。
6. 如权利要求 2 所述的方法,其中,响应于所述存储块变满,触发无用存储单元收集。
7. 如权利要求 2 所述的方法,其中,无用存储单元收集包括:  
将所述芯片带设置为清理状态;  
对于所述多个页面中的每个页面:  
读取页面元数据和 LBA 到 PBA 映射信息;  
响应于所述页面有效,  
读取完整的页面数据;  
获取所述页面的目标位置;  
将页面数据写入到新位置;  
将所述芯片带设置为擦除状态;  
擦除所述芯片带中的所有块;  
将所述芯片带设置为空闲状态。
8. 一种用于减少存储芯片中的存取争用的系统,所述系统包括:  
配置为从具有多个通道和多个存储块的存储设备中选择空闲状态的芯片带,其中,所述芯片带包括多个页面;  
配置为将所述芯片带设置为写入状态的模块,所述芯片带的写入状态与擦除状态相分离;  
配置为设置所述多个通道中的每个通道中的写队列头的模块;  
配置为对于所述芯片带中的所述多个通道中的每个通道,  
将写队列头设置到来自所述芯片带的属于该通道的芯片中的第一空闲页面的模块;  
配置为根据写入分配调度器,在所述通道间分配写入请求的模块;

配置为生成页面写入的模块；

配置为响应于所述页面写入，递增所述写队列头的模块；以及

配置为当所述芯片带已满时，将它设置为在线状态的模块。

9. 如权利要求 8 所述的系统，其中，所述配置为递增所述写队列头的模块包括：配置为将所述写队列头设置到来自所述多个存储块的存储块中的下一个空闲页面的模块。

10. 如权利要求 8 所述的系统，其中，所述系统进一步包括配置为触发无用存储单元收集的模块。

11. 如权利要求 8 所述的系统，其中，所述多个存储块以与所述存储芯片上的所述多个存储块的相应地址相同的顺序，被保持在所述存储器芯片上的空闲块队列中。

12. 如权利要求 10 所述的系统，其中，响应于所述芯片带是被改变为所述写入状态的最后一个空闲芯片带，触发所述无用存储单元收集模块。

13. 如权利要求 10 所述的系统，其中，响应于达到所述存储器设备上的空闲状态芯片带的预定阈值，触发所述无用存储单元收集模块。

14. 如权利要求 10 所述的系统，其中，响应于所述存储块变满，触发所述无用存储单元收集模块。

15. 如权利要求 10 所述的系统，其中，所述配置为无用存储单元收集的模块包括：

配置为将所述芯片带设置为清理状态的模块；

配置为对于所述多个页面中的每个页面的模块；

配置为读取页面元数据和 LBA 到 PBA 映射信息的模块；

配置为响应于所述页面有效的模块，

配置为读取完整的页面数据的模块；

配置为获取所述页面的目标位置的模块；

配置为将页面数据写入到新位置的模块；

配置为将所述芯片带设置为擦除状态的模块；

配置为擦除所述芯片带中的所有块的模块；

配置为将所述芯片带设置为空闲状态的模块。

16. 一种存储系统，包括：

具有写入分配调度器的多个通道；以及

可通信地耦接到所述多个通道中的每个通道的至少一个芯片带，所述至少一个芯片带具有空闲状态、写入状态、在线状态和擦除状态，

其中，所述至少一个芯片带被配置成被从所述空闲状态设置为所述写入状态以分配写入请求，并进一步被配置成当所述芯片带已满时被从所述写入状态设置为所述在线状态，所述芯片带的写入状态与擦除状态相分离。

17. 如权利要求 16 所述的存储系统，进一步包括位于所述多个通道中的每个通道上的写队列头，

其中，对于所述芯片带中的多个通道中的每个通道，所述系统被配置成：

将所述写队列头设置到来自所述芯片带的属于相应通道的芯片中的第一空闲页面；

根据所述写入分配调度器，在所述多个通道间分配写入请求；

生成页面写入；以及

响应于所述页面写入,递增所述写队列头;以及  
当所述芯片带已满时,将它设置为在线状态。

18. 如权利要求 17 所述的存储系统,其中,所述系统进一步被配置成:响应于所述芯片带是被改变为所述写入状态的最后一个空闲芯片带以及达到所述系统中的空闲状态芯片带的预定阈值这两项中的至少一项,触发无用存储单元收集。

19. 如权利要求 18 所述的存储系统,其中,当存储块已满并且达到所述存储系统上的空闲块的预定阈值时,触发无用存储单元收集。

20. 如权利要求 18 所述的存储系统,其中,无用存储单元收集包括:

将所述芯片带设置为清理状态;

对于所述多个页面中的每个页面:

读取页面元数据和 LBA 到 PBA 映射信息;

响应于所述页面有效,

读取完整的页面数据;

获取所述页面的目标位置;

将页面数据写入到新位置;

将所述芯片带设置为所述擦除状态;

擦除所述芯片带中的所有块;

将所述芯片带设置为所述空闲状态。

## 减少基于闪存的存储系统中的存取争用

### 技术领域

[0001] 本发明涉及基于闪存的存储系统,具体而言,涉及用于减少基于闪存的存储系统中的或与基于闪存的存储系统具有类似属性的其它存储系统中的存取争用的方法和系统。

### 背景技术

[0002] 闪存是能够被电擦除和重新编程的非易失性计算机存储器。诸如固态驱动器(SSD)之类的基于闪存的存储设备具有以下硬件特征:其中,通常以页面大小的字节片来执行读和写,该字节片通常大小为2到4KB。擦除通常以完整的块执行,其中,一个块通常包括64到128个页面。闪存包括NOR类型和NAND类型。一般而言,存在两种不同类型的NAND闪存芯片:基于单电平单元(SLC)的类型存储一个比特,以及基于多电平单元的类型保持多个电压电平以便存储多于一个的比特。基于SLC的闪存芯片中的4KB页面分别具有典型的25和600 $\mu$ s的读取和写入时间。擦除一个完整的块要花费显著较长的时间量,并且在企业级闪存芯片中要花费7ms。这些读取/写入/擦除特征不管工作负荷如何都是有效的。相比之下,在基于硬盘驱动器(HDD)的存储系统中,寻道时间限制了随机存取性能。然而,在基于闪存的存储设备中,在单元能够被编程(例如,写入)之前必须首先擦除它们。因此,用来隐藏块擦除等待时间的常见技术是始终错位地写入数据,并推迟对块的擦除,直到启动无用存储单元收集。当发出擦除命令时,芯片忙碌直到操作完成,并且没有办法在此时间期间对此芯片读取或写入,这被称为“阻止擦除”。错位写入策略要求被叫做闪存转换层(FTL, Flash Translation Layer)的一个特殊层,该层保持闪存中的逻辑块地址(LBA, logical block address)和实际物理页面/块地址(PBA, physical page/block address)之间的映射。

[0003] 对基于闪存的存储设备的存取时间仍能够受到延迟变化。图1示出了模拟的现有技术系统的概率密度函数相对于等待时间的图100。图2示出了模拟的现有技术系统的累积分布函数相对于等待时间的图200。图1和2示出了高负荷情况下的闪存模拟器。在示例中,当一个块正被擦除时,对同一个芯片的后续读取请求可能必须等待7ms才能被服务。类似地,读取请求可被正在进行的写入请求延迟高达600 $\mu$ s。这样显著的延迟在某些环境下是不可接受的。另外,某些国家(例如,日本)对于特定IT应用甚至强制最大延迟范围(~5-10ms)。因此,潜在的总延迟可能会超出所要求的延迟范围。此外,传统的DRAM存储技术没有与闪存相同的限制,因此,提供显著的、更加均匀的存取延迟。闪存高速缓存解决方案试图用闪存来替换昂贵的有后备电池的DRAM存储器(被叫做NVRAM)。因此,这样的闪存高速缓存的特征类似于那些存储技术的特征。因此,利用基于闪存的存储技术来提供这样的特征是有益的,但不普通。

### 发明内容

[0004] 示例性实施例包括一种用于减少基于闪存的存储系统中的或与基于闪存的存储系统具有类似属性的存储系统中的存取争用的方法,该方法包括:从具有多个通道和多个

存储块的存储设备中选择空闲状态的芯片带,其中,所述芯片带包括多个页面;将所述芯片带设置为写入状态;设置所述多个通道的每个通道中的写队列头;对于闪存带中的多个通道的每个通道,将写队列头设置到来自所述芯片带的属于该通道的芯片中的第一空闲页面;根据写入分配调度器,在通道间分配写入请求;生成页面写入;以及响应于所述页面写入,递增所述写队列头。

[0005] 另外的示范性实施例包括一种用于减少基于闪存的存储系统中的存取争用的计算机程序产品,该计算机程序产品包括用于使计算机实现一种方法的指令,该方法包括:从具有多个通道和多个存储块的存储设备中选择空闲状态的芯片带,其中,所述芯片带包括多个块,以及块包括多个页面;将芯片带设置为写入状态;设置多个通道的每个通道中的写队列头;对于闪存带中的多个通道的每个通道,将写队列头设置到来自所述芯片带的属于该通道的芯片中的第一空闲页面;根据写入分配调度器,在通道间分配写入请求;生成页面写入;以及响应于所述页面写入,递增所述写队列头。

[0006] 进一步的示范性实施例包括一种存储设备,包括具有写入分配调度器的多个通道、可通信地耦接到多个通道的每个通道的至少一个芯片带,所述至少一个芯片带具有空闲状态、写入状态、在线状态和擦除状态,其中,所述至少一个芯片带被配置成被从所述空闲状态设置为所述写入状态以分配写入请求,并进一步被配置成被从所述写入状态设置为所述在线状态以服务读取请求。

[0007] 另外的功能和优点通过本发明的技术来实现。此处将详细描述本发明的其它实施例和方面,它们被视为所要求保护的发明的一部分。为了更好地理解具有这些优点和功能的本发明,请参阅说明和附图。

## 附图说明

[0008] 在本说明书结尾时在权利要求书中特别指出了并明确声明了被视为本发明的主题。通过下面的结合附图对本发明进行的详细说明,可以更好地理解本发明的前述和其它特征和优点,其中:

[0009] 图 1 示出了模拟的现有技术系统的概率密度函数相对于等待时间的图;

[0010] 图 2 示出了模拟的现有技术系统的累积分布函数相对于等待时间的图;

[0011] 图 3 示出了其中能够实现示范性存取争用减少方法的示范性闪存体系结构;

[0012] 图 4 示出了显示这里所述的示范性四种状态和相应的转变的示范性状态图;

[0013] 图 5 示出了根据示范性实施例的用于减少基于闪存的存储系统中的存取争用的方法的流程图;

[0014] 图 6 示出了示范性无用存储单元收集方法的流程图;以及

[0015] 图 7 示出了显示图 6 的进一步细节的示范性无用存储单元收集方法的流程图。

## 具体实施方式

[0016] 在各示范性实施例中,这里所描述的系统和方法减少/消除由基于闪存的存储系统的擦除和写入延迟引起的存取争用。这里所描述的示范性实施例作为说明性示例讨论了基于闪存的存储系统。应当理解,在其它示范性实施例中,这里所描述的方法可被应用于在存取时间和阻止写入或擦除操作方面与基于闪存的存储系统具有类似属性的任何存储系

统。在各示例性实施例中,基于闪存的存储设备体系结构能够提供针对读取和写入请求的严格延迟范围。另外,该体系结构提供有效的无用存储单元收集和磨损平衡,特别是对于闪存高速缓存解决方案。在各示例性实施例中,此处所描述的系统和方法通过分离读取、写入以及擦除操作,来完全隐藏来自对同一芯片的在先操作的附加延迟。为了使得对芯片的读取和写入免于由对同一个芯片的在先擦除操作引起的附加延迟,读取和写入操作只对没有发生擦除而仍提供对存储的所有信息的存取的芯片执行(即,设备的完全 LBA 空间)。类似地,能够实现此处所描述的方法,以保护读取操作免受对同一个芯片的在先请求的擦除和写入延迟的不利影响。在各示例性实施例中,使读取操作免受不利影响,实现了其中闪存芯片被附接到通道总线并被分组到通道之中的芯片带的体系结构。此体系结构与利用写入页面分配器和无用存储单元收集算法的基于每通道空闲块队列的写入策略组合。

[0017] 诸如闪存高速缓存或 SSD 之类的基于闪存的存储设备包括闪存芯片被附接到的若干通道。每个通道具有管理对芯片的所有请求的低复杂度通道控制器。通道被专用通道控制器控制,所述通道控制器对芯片执行读取、写入以及擦除操作,并可能执行诸如添加或验证纠错信息的其它操作。通道控制器由主控制器控制,该主控制器执行诸如 LBA 到 PBA 映射、无用存储单元收集、空闲块和坏块管理之类的主要任务。图 3 示出了其中可以实现示例性存取争用减少方法的示例性闪存体系结构 300。如此处所描述的,该体系结构 300 包括用于多个较低复杂度通道控制器 310、315、320、325 的主控制器 305。在各示例性实施例中,主控制器 305 为每个通道保持一个空闲块队列和写队列头。所述空闲块队列保留能够被用于写入的所有块。所述写队列头指向当前正被写入数据的块中的第一空闲页面。如此处进一步所描述的,预先已从通道的空闲块队列中获取所述空闲块。在各示例性实施例中,每个通道控制器 310、315、320、325 都包括芯片的相应通道。如图所示,作为示例,控制器 310 控制通道 1,控制器 315 控制通道 2,控制器 320 控制通道 C-1,而控制器 325 控制通道 C。另外,每个通道控制器 310、315、320、325 都通过通道总线 311、316、321、326 可通信地耦接到其相应通道。每个通道还包括芯片,这些芯片被概括地示为“芯片”。在各示例性实施例中,写入模式带(例如,芯片带 n)可包括写入头芯片。另外,GC 带(例如,芯片带 n+2)可包括无用存储单元收集器头芯片。

[0018] 除通道结构之外,芯片被虚拟地分组到所谓的芯片带中,被示为芯片带 1、芯片带 2、芯片带 n、芯片带 n+1、芯片带 n+2、芯片带 N-1、芯片带 N。如此,每个通道中的第一芯片都属于第一芯片带,依此类推。在各示例性实施例中,芯片带处于四种状态之一:在线、无用存储单元收集(GC, garbage collection)、空闲或写入状态。图 4 示出了示例性状态图 400,示出了如此处所描述的示例性四种状态以及相应的转变。在图 3 所示出的示例中,芯片带 1、芯片带 2、芯片带 N-1 以及芯片带 N 是为读取请求实现的“在线”,芯片带 n 是写入模式带,芯片带 n+1 是空闲,而芯片带 n+2 处于 GC 状态。这些状态是说明性的,并且在其它示例性实施例中,可以预期其它状态。

[0019] 在各示例性实施例中,当芯片带处于在线状态时,能够从所有其芯片中读取数据。空闲状态的芯片带保留具有所有新被擦除并准备好被写入的块的芯片。处于写入状态的芯片带包括当前正在对其写入数据的芯片。在优选实施例中,在任何时间,只有一个芯片带处于写入状态。然而,在某些情况下,如此处进一步描述的,让多于一个的芯片带处于写入状态是有意义的。

[0020] 在各示例性实施例中,处于无用存储单元收集状态(GC 状态)的芯片带包括当前正在被清理以便准备它们用于写入的芯片。优选地,没有多于一个的芯片带处于 GC 状态。因而,在任何时间,GC 能够在每个通道中的单个芯片中发生。另外,可以有多于一个的芯片带处于 GC 状态,如果加速 GC,这是期望的。处于 GC 状态的芯片带的数量也可以是动态的(即,自适应于工作负荷)。大量的存储空间被预留在 GC 状态,因此,在 GC 期间无法被用来存储实际数据。在各示例性实施例中,GC 状态可以具有子状态,诸如“清理”和“擦除”状态。清理状态表示仍然有效的页面被移动到写入带中的新位置的状态,以及擦除状态表示在所有有效页面都被移动到新位置之后块正在被实际擦除的状态。在清理状态下,用户读取请求仍能够从芯片带服务。一旦芯片带改变为擦除状态,该芯片带上的所有数据都已被移出,以及必须从新位置供应。期望防止从处于擦除状态的带服务用户读取请求。因而,可从新位置服务读取请求。在各示例性实施例中,可在单个芯片上中断擦除,但在整个带上继续。

[0021] 图 5 示出了根据示例性实施例的用于减少基于闪存的存储系统中的存取争用的方法 500 的流程图。在各示例性实施例中,当方法 500 开始时,所有块都是空闲的,因此,在框 505,所有芯片带都处于空闲状态。在框 510,在数据能够被写入设备之前,选择空闲芯片带,并设置为写入模式。例如,在框 510,可以选择闪存芯片上的第一空闲芯片带,但是可以选择任何其它空闲芯片带。在框 515,对于每个通道,从相应的空闲块队列中选择空闲块。在各示例性实施例中,选择每个通道的给定芯片带中的第一空闲块,但是,也可以选择给定芯片带和通道中的任何其它空闲块。在框 520,对于每个通道,写队列头被设置到来自处于写入模式的芯片带的属于该通道的芯片中的第一空闲页面。在框 525,方法 500 检查设备是否已经停止。倘若如此,则结束方法 500。如果设备仍在运行,则在框 530,根据写入分配调度器,在各通道间分配写入请求。在各示例性实施例中,可以实现循环调度器,该调度器在所有可用通道间均匀地分配写入工作负荷。根据期望的属性,在其它示例性实施例中,可以预见其它调度器。在框 535,一旦写入了一个页面,就递增写队列头,以便将它设置到块中的下一个空闲页面。在框 540,方法 500 检查芯片带是否已满。在芯片带已满的情况下,框 545 根据预先配置的值来确定是否有足够的空闲带可用。如果在框 545 有足够的剩余芯片带,那么,方法继续到框 555。如果在框 545 没有足够的剩余芯片带,那么,在框 550 触发无用存储单元收集方法 600(参见图 6),以便从先前已被写入的块中的失效页面回收空间,并且方法 500 在框 555 继续。请注意,在优选实施例中,无用存储单元收集方法 600 独立地并与方法 500 并行地运行。在框 555,已知当前带已满并且它被设置为在线状态。然后,在框 510 中,方法 500 选择空闲的下一个较高的芯片带。在其它示例性实施例中,对下一个写入芯片带的选择可以根据某一策略来执行。在芯片带还没有满的情况下,框 560 检查当前块是否已满。在块还没有满的情况下,方法 500 在框 525 继续。当块已满时,在框 565,在没有足够的空闲块的情况下,在框 570 中,根据预定义的值触发无用存储单元收集,以及在框 575,从与此通道相关联的空闲块队列中获取下一个空闲块。在各示例性实施例中,能够通过以与芯片中的块的地址相同的顺序在空闲块队列中保持块(同时去除所有的坏块),来保持空闲块队列。

[0022] 在各示例性实施例中,可在最后一个空闲芯片带被改变为写入状态时触发 GC,或更早地基于可配置的阈值(即,基于空闲块的总数)、系统中应当可用的空闲芯片带的数量来触发 GC。在各示例性实施例中,阈值越大,储存备用的预留存储空间越高,因此,减小了可



用于读取和写入的总容量。阈值可被设置为大于零,以更好地适应写入脉冲串。类似地,可在有足够的空闲页面可用时停止 GC。在各示例性实施例中,可根据需要重新启动方法 500。

[0023] 图 6 示出了示例性无用存储单元收集方法 600 的流程图。由方法 500 在框 550 和 570 (参见上面的图 5) 中触发方法 600。它首先检查 GC 是否已经运行。在 GC 已经运行的情况下,已经运行的 GC 方法 600 释放足够的空间,以及没有必要开始方法 600 的另一个实例。在 GC 没有运行的情况下,无用存储单元收集在框 610 开始,以及在框 615 停止。在方法 700 中描述了详细的 GC 方法(参见此处的图 7)。在框 620,在方法 700 结束之后,方法 600 检查是否有足够的空闲块和芯片带可用。如果情况不是这样的,则方法 600 触发另一个 GC 循环,方法 600 停止。

[0024] 图 7 详细示出了示例性无用存储单元收集方法 700 的流程图。在各示例性实施例中,GC 方法 700 在启动时选择最旧的芯片带。在框 705,方法 700 将所选芯片带设置为清理状态。然后,方法 700 对于所选芯片带中的所有页面执行一系列框。在框 710,方法 700 读取页面元数据和 LBA 到 PBA 映射信息。在框 715,方法 700 确定页面是否仍有效。如果页面元数据和 LBA 到 PBA 映射信息表示页面仍有效,那么页面被写入到新位置。如此,如果在框 715 中页面仍有效,那么在框 720,方法 700 读取完整的页面数据。在框 725,方法 700 从写入分配调度器获取页面的目标位置。在各示例性实施例中,写入分配调度器可以分配同一个通道或另一个通道中的新位置。此外,如果通道控制器能够直接在通道中移动数据,那么优选通道控制器移动数据,因为它不消耗其它总线的资源。在框 730,方法 700 将页面写入到新位置。如果在框 715 中页面不再有效,则在方法 700 中,不采取动作,并且稍后自动回收无效页面所占用的空间,因为无效页面不被移动,并因而稍后被擦除。在框 735,该方法确定是否有剩余的页面。如果在框 735 有剩余的页面,那么方法 700 在框 710 继续。如果在框 735 没有剩余的页面,那么在框 740,方法 700 将芯片带设置为擦除状态。在框 745,方法 700 擦除芯片带中的所有块。在各示例性实施例中,通道控制器能够并行地执行块的擦除操作。另外,每个通道控制器都能够在后台对芯片执行块擦除,使得当块正在被擦除的同时相应的通道不被阻止。在框 750,方法 700 将芯片带的状态改变为空闲状态。在框 755,方法 700 确定下次将对哪个芯片带运行 GC。

[0025] 在各示例性实施例中,根据给定通道上芯片的数量,芯片带可被拆分成子带。另外,选择带的大小(即,同一个通道中属于同一个芯片带的芯片的数量)以影响通道带宽。当通道控制器支持对多个芯片的读取和写入操作的管线化时,特别期望带宽管理。因此,同一个通道上的多于一个的芯片可以与同一个写入带相关联。此外,对于每个芯片带状态,能够支持不同的带大小。

[0026] 在各示例性实施例中,能够使得用户读取请求免于由正在进行的对同一个芯片的擦除和写入操作引起的延迟。正在被写入到写入芯片带中的用户数据被保留在主控制器的写入高速缓存中,直到完整的芯片带已被写入。然后,从写入高速缓存服务对此新写入的数据的用户读取请求,因此,消除了由于在先写入导致的潜在延迟。然而,主控制器可能需要能够保持完整芯片带的巨大的写入高速缓存,以便执行这些操作。在各示例性实施例中,为了规避巨大的写入高速缓存,可以实现两个交替的写入带。当写入高速缓存转出数据(de-stage)到闪存时,数据首先被写入到第一写入带,然后,此写入带被临时切换到在线状态,从这里,数据能够被存取。在专用映射块中保持 LBA 到 PBA 映射的更新。然后,相同数

据被写入到第二写入带,以及创建包含该变化的第二 LBA 到 PBA 映射块。在写入完成之后,第二写入带被切换到在线状态(相应的 LBA 到 PBA 映射块被设置为活动)。第一芯片带被切换回写入状态,以及数据可被从写入高速缓存中转出等等。一旦写入芯片带已满,它们中的一个就被保持为在线状态,相应的 LBA 到 PBA 映射表也是如此。其它芯片带可被擦除并被设置为空闲状态或被用于冗余或并行读取(与 RAID 1 同理)。这种类型的操作使闪存写入的总数翻倍。

[0027] 在一些配置中,单个通道上可以只有几个芯片。保持用于 GC 的完整芯片带在不能被用来保持用户数据的预留存储容量方面变得昂贵。在这样的配置中,将多个通道分组为通道组并将芯片带拆分为子带是有益的。因此,一个子带只包括来自单个通道组的芯片。然后,可以对子带而不是对完整的芯片带应用此处所描述的相同算法。此方法降低了擦除速度,因为更少的芯片能够被并行地擦除。然而,如果写入工作负荷不高,则可以实现此方法。因而,能够根据观察到的写入工作负荷来动态地调整子带大小。

[0028] 在各示例性实施例中,如此处所描述的通道可以是物理通道,也可以是虚拟通道。若干物理通道可被分组到一个虚拟通道中。因此,此处所描述的系统和方法适用于物理和虚拟通道两者。

[0029] 在各示例性实施例中,GC、空闲、在线、以及写入状态按顺序移动通过所有芯片带,这可被称为环形缓冲器。环形缓冲器体系结构具有在芯片粒度上均匀地分配块的磨损的优点,因此,由于不要求额外的磨损平衡方法而降低了方案的复杂性。在此处所描述的示例中,用于分配要被写入的页面的循环策略是能够在环形缓冲器中实现的解决方案。尽管如此,由于坏块,写队列头可能在不同通道之间变得不同步。在这样的情况下,可在循环调度器中插入额外的循环,其中仅在最前面的写队列头的头后面的队列中分配页面。

[0030] 在各示例性实施例中,在可被分离成静态和动态数据的工作负荷的情况下,环形缓冲器体系结构可能导致数据在 GC 期间被不必要地来回移动。因而,可以实现的策略的另一个示例是基于窗口的贪婪回收策略。可以通过动态地将芯片分组为芯片带以及根据由 GC 保持的信息来动态地修改芯片带的大小,来调适此处所描述的方法以适合这样的基于窗口的贪婪回收策略。

[0031] 技术效果包括通过分离读取、写入、以及擦除操作,来完全隐藏来自在先操作的附加延迟。此外,使得对一个芯片的读取和写入操作免于由对同一个芯片的在先擦除操作引起的附加延迟。技术效果还包括确保读取和写入操作只对没有发生擦除而仍提供对存储的所有信息的存取的芯片执行。另外,还防止读取操作受到对同一个芯片的在先请求的擦除和写入延迟的不利影响。

[0032] 此处所使用的术语只是为了描述特定实施例,并不旨在对本发明作出限制。如此处所使用的,单数形式“一个”、“一”和“该”也包括复数形式,除非上下文明确地指出。还应当进一步理解,本说明中所使用的术语“包括”和/或“包含”表示陈述的特征、整数、步骤、操作、元件和/或组件的存在,但是不排除一个或多个其它特征、整数、步骤、操作、元件、组件或其组合的存在或添加。

[0033] 下面的权利要求中的所有装置或步骤加功能元件的相应结构、材料、动作以及等效物旨在包括用于与特别要求保护的其它要求保护的元件相结合地执行所述功能的任何结构、材料或动作。对本发明的描述是为了示出和描述的目的而呈现的,而不意欲是排外的

或限于所公开的形式。本领域的技术人员将认识到,在不偏离本发明的范围和精神的情况下,可以进行各种修改和变化。所选择和所描述的实施例只是为了最好地说明本发明的原理和实际应用,并使本领域技术人员能够理解具有适合于预期的具体用途的各种修改的各实施例的发明。

[0034] 此处所描述的流程图只是一个示例。在不偏离本发明的精神的情况下,对此处所描述的这些图或步骤(或操作)可以有许多更改。例如,可以以不同顺序执行各个步骤,也可以添加、删除或修改步骤。所有这些变化都被视为所要求保护的本发明的一部分。

[0035] 尽管已经描述了本发明的优选实施例,但是,可以理解,本领域的技术人员现在和将来都可以作出在随后的权利要求书的范围内的各种改进和增强。这些权利要求应该被解释为维持对于首先描述的本发明的适当的保护。

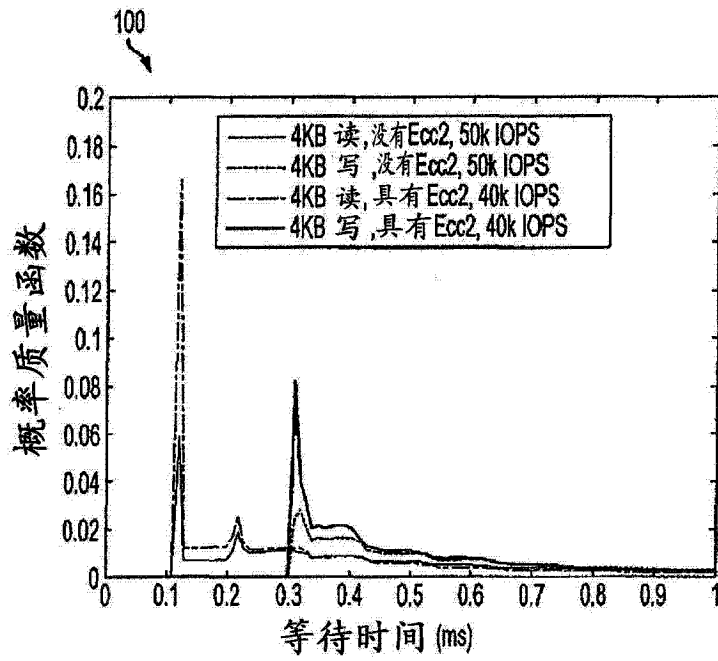


图 1

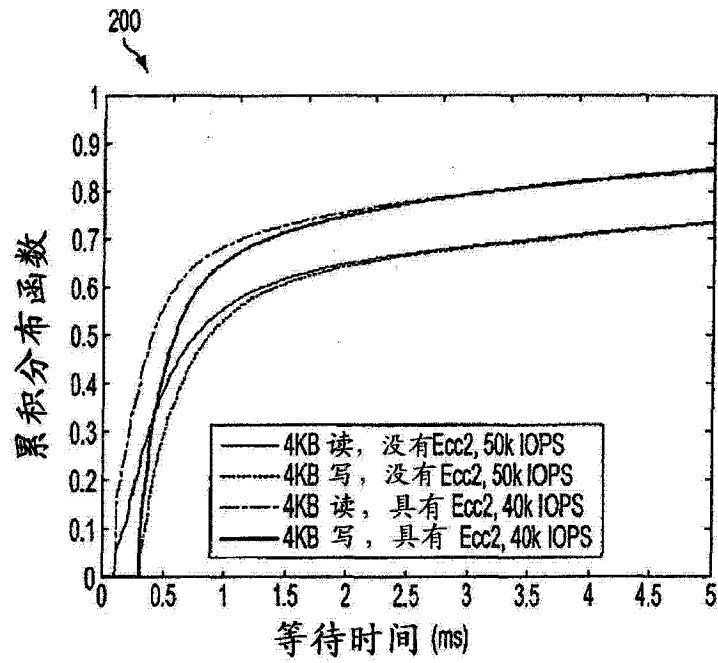


图 2

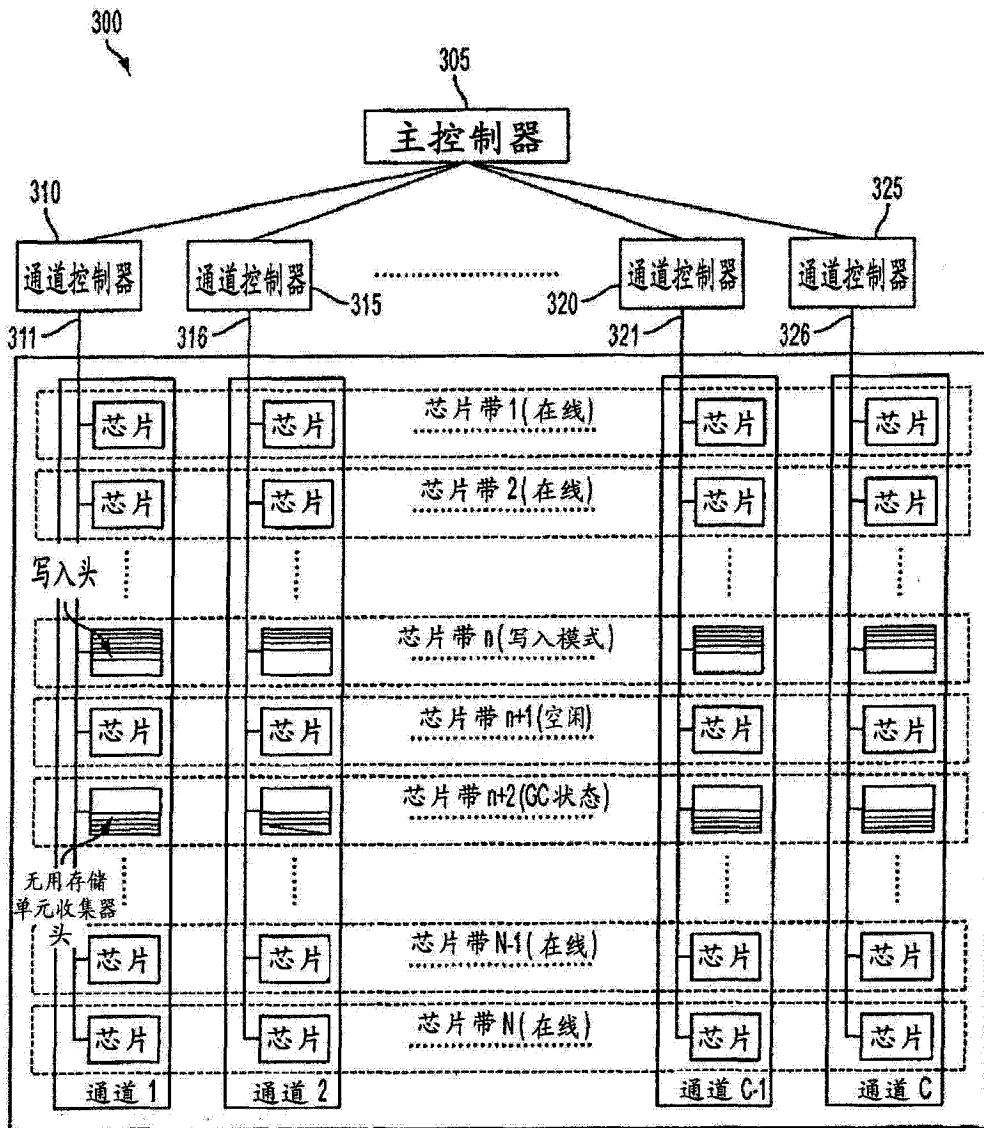


图 3

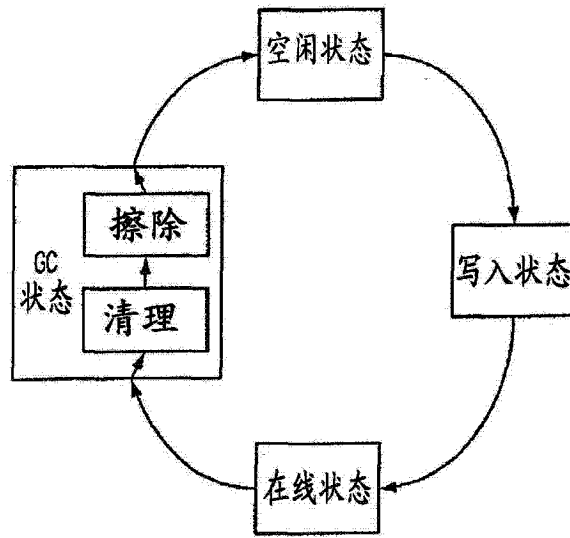


图 4

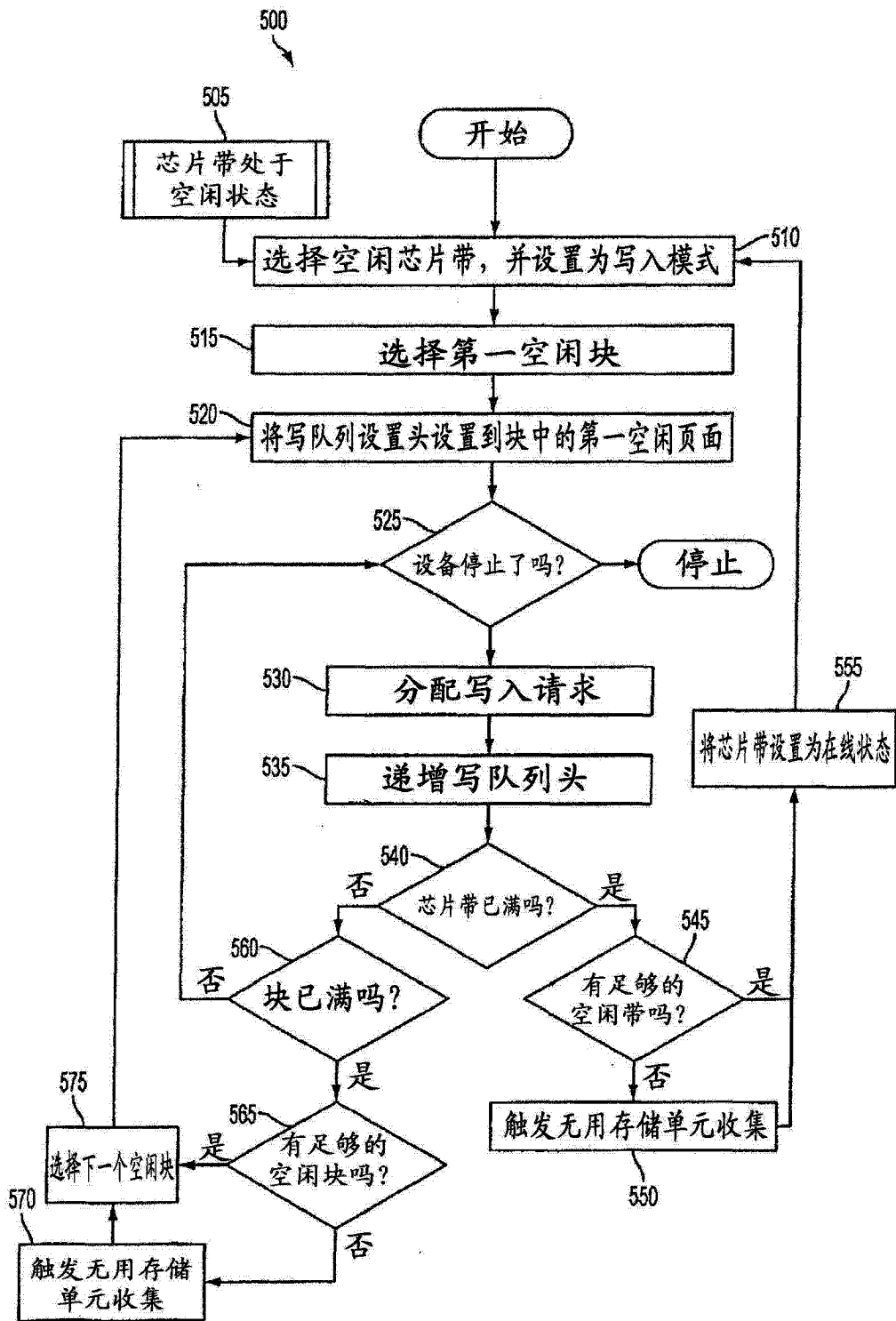


图 5

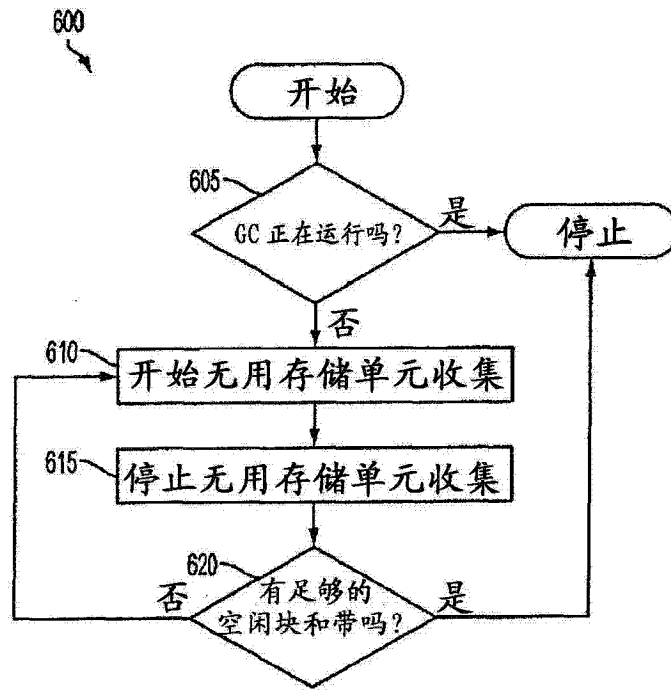


图 6



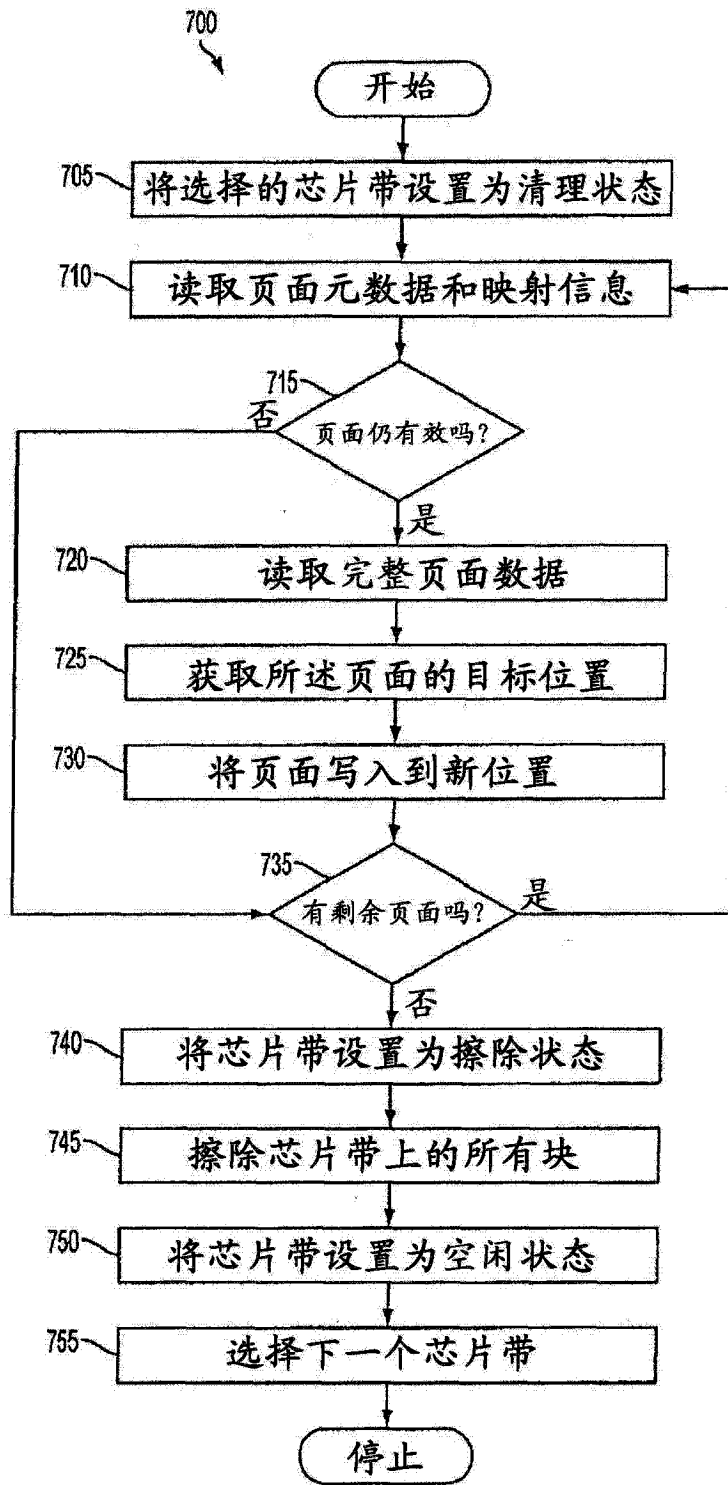


图 7