

(12) 发明专利申请

(10) 申请公布号 CN 102301427 A

(43) 申请公布日 2011. 12. 28

(21) 申请号 201080005997. X

(51) Int. Cl.

(22) 申请日 2010. 03. 25

G11C 16/34(2006. 01)

(30) 优先权数据

G11C 29/52(2006. 01)

12/420, 628 2009. 04. 08 US

G06F 11/00(2006. 01)

(85) PCT申请进入国家阶段日

2011. 07. 29

(86) PCT申请的申请数据

PCT/EP2010/053943 2010. 03. 25

(87) PCT申请的公布数据

W02010/115726 EN 2010. 10. 14

(71) 申请人 国际商业机器公司

地址 美国纽约阿芒克

(72) 发明人 W. J. 卡贝拉克 S. R. 赫兹勒

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 周少杰

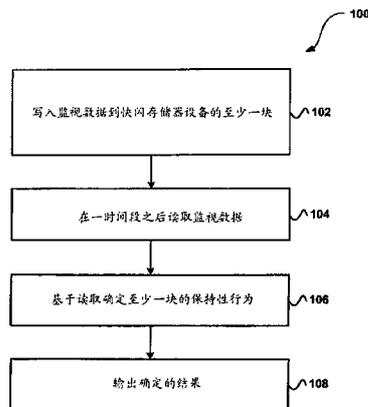
权利要求书 1 页 说明书 11 页 附图 11 页

(54) 发明名称

分析来自具有有限耐久性和 / 或保持性的存储器设备的监视数据信息

(57) 摘要

根据一个实施例的方法包括 : 从具有有限耐久性和 / 或保持性的多个存储器设备收集监视数据信息, 所述监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据 ; 分析监视数据信息 ; 以及基于分析采取与设备中的至少一个相关的动作。还公开了另外的系统、方法和计算机程序产品。



1. 一种方法,包括:

从具有有限耐久性和 / 或保持性的多个存储器设备收集监视数据信息,所述监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据;

分析监视数据信息;以及

基于分析采取与设备中的至少一个相关的动作。

2. 如权利要求 1 所述的方法,还包括生成监视数据。

3. 如权利要求 1 所述的方法,其中监视数据信息包括与监视数据的数据年龄和写入循环计数的至少一个相关的错误率信息。

4. 如权利要求 1 所述的方法,还包括执行收集和分析步骤至少两次,并且基于其生成统计趋势。

5. 如权利要求 1 所述的方法,还包括执行收集和分析步骤至少两次,并且跟踪监视数据的时间依赖性。

6. 如权利要求 1 所述的方法,其中采取动作包括为设备中的至少一个设置行为目标。

7. 如权利要求 6 所述的方法,其中行为目标是对于数据写入循环计数的限制。

8. 如权利要求 6 所述的方法,其中行为目标是对于数据年龄的限制。

9. 如权利要求 1 所述的方法,其中监视数据不经历磨损均衡。

10. 如权利要求 6 所述的方法,还包括在已经过去一段时间之后为设备的至少一个更新行为目标。

11. 如权利要求 1 所述的方法,其中存储器设备是 NAND 设备。

12. 如权利要求 1 所述的方法,其中监视数据包括小于大约 1% 的每个存储器设备的存储器容量。

13. 如权利要求 1 所述的方法,其中每个存储器设备中的监视数据写入到多组块,其中对每个组执行的写入循环的数目不相同。

14. 一种计算机程序产品,包括具有与其体现的计算机可用程序代码的计算机可用介质,当由计算机执行所述计算机可用程序代码时,所述计算机可用程序代码使得计算机执行权利要求 1 到 13 的步骤。

15. 一种系统,包括:

具有有限耐久性和 / 或保持性的多个存储器设备,每个存储器设备具有:

多个存储器块,块的至少一个具有写入其中的监视数据,其中在写入监视数据之前,至少一块已经写入多次;以及

用于寻址块的电路;

与存储器设备通信的处理器;以及

时钟,所述处理器与所述时钟通信。

16. 如权利要求 15 所述的系统,还包括存储监视数据的副本的次级存储器。

17. 如权利要求 15 所述的系统,还包括如权利要求 14 所述的计算机可用介质。

18. 如权利要求 15 到 17 的任一所述的系统,其中存储器设备是 NAND 设备。

分析来自具有有限耐久性和 / 或保持性的存储器设备的监视数据信息

技术领域

[0001] 本发明一般涉及存储器设备,更具体地涉及分析来自具有有限耐久性(endurance)和 / 或保持性(retention)的存储器设备的监视数据信息。

背景技术

[0002] 如 NAND 快闪存储器设备的一些存储器设备拥有有限的写入 / 擦除耐久性和 / 或有限的数据保持性。通常希望跟踪写入 / 擦除耐久性和数据保持性,以识别这样存储器的当前有用性和可靠性。然而,迄今为止,提出的解决方案未能为与这样的存储器设备关联的这些和 / 或其它问题提供可接受的解决方案。

发明内容

[0003] 根据一个实施例的方法包括:从具有有限耐久性和 / 或保持性的多个存储器设备收集监视数据信息,所述监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据;分析监视数据信息;以及基于分析采取与设备中的至少一个相关的动作。

[0004] 根据一个实施例的计算机程序产品包括:具有与其体现的计算机可用程序代码的计算机可用介质。计算机可用程序代码包括:计算机可用程序代码,配置为从具有有限耐久性和 / 或保持性的多个存储器设备收集监视数据信息,所述监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据;计算机可用程序代码,配置为分析监视数据信息;以及计算机可用程序代码,配置为基于分析采取与设备中的至少一个相关的动作。

[0005] 根据另一实施例的方法包括:从具有有限耐久性和 / 或保持性的多个存储器设备收集监视数据信息,所述监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据;分析监视数据信息;以及基于分析为设备中的至少一个设置行为目标。

[0006] 根据另一实施例的系统包括具有有限耐久性和 / 或保持性的多个存储器设备,每个存储器设备具有:多个存储器块,块中的至少一个具有写入其中的监视数据,其中在写入监视数据之前,至少一块已经被写入多次;以及用于寻址块的电路。所述系统还包括:与存储器设备通信的处理器;以及时钟,处理器与所述时钟通信。

[0007] 从以下结合附图以示例的方式说明本发明的原理的详细描述,本发明的其它方面和优点将变得明显。

附图说明

[0008] 接下来,将参照附图以示例的方式描述本发明的实施例,附图中:

[0009] 图 1 图示根据一个实施例的用于确定快闪存储器设备的至少一块的保持性行为的方法。

[0010] 图 2 图示根据一个实施例的示出闪存错误率面(flash error rate surface)的形状的代表性图。

[0011] 图 3 示出根据一个实施例的示例性累积分布函数 (CDF), 该累积分布函数图示由于在某一数据年龄 (data age) 的磨损, 给定块将超过扇区级纠错码 (CEC) 的校正能力的概率。

[0012] 图 4 图示根据一个实施例的在某一数据年龄的存取分布和 CDF 之间的重叠的表示。

[0013] 图 5 示出根据一个实施例的在某一时间点的用户和监视数据的代表性循环计数直方图。

[0014] 图 6 示出来自监视数据系统的可能的一组测量。

[0015] 图 7 图示根据一个实施例的用于分析来自多个快闪存储器设备的监视数据信息的方法。

[0016] 图 8 图示根据一个实施例的代表性存储系统。

[0017] 图 9 示出根据一个示例的结果的曲线图。

[0018] 图 10 图示根据一个实施例的网络架构。

[0019] 图 11 示出根据一个实施例的可以与图 10 的服务器和 / 或客户端相关联的代表性硬件环境。

具体实施方式

[0020] 在此描述的方法和系统可以结合任何类型的具有有限耐久性和 / 或保持性的存储器设备实现, 或者包括任何类型的具有有限耐久性和 / 或保持性的存储器设备。说明性类型的具有有限写入 / 擦除耐久性或有有限数据保持性的存储器设备包括: 包含快闪存储器、相变存储器、铁电存储器等的设备。

[0021] 为了易于理解和将在此呈现的教导置于上下文中, 按照闪存设备来阐述大量当前描述, 要理解描述仅仅以示例的方式进行, 并且不解释为限制。

[0022] NAND 闪存设备拥有有限的写入 / 擦除耐久性和有限的数据保持性。当前用于检测磨损和保持性的技术具有不足以用于信息技术 (IT)、及一些消费者应用的换码率 (escape rate)。换码的结果可能是严重的。例如, 数据可能由于换码而被破坏。

[0023] 由于闪存中的错误率依赖于写入循环计数和数据的年龄, 所以其是 3D 面。参见, 例如以下图 2 及其讨论。因为单元可能驱动为完全故障, 所以错误率可能超过任何扇区级纠错码 (ECC) 的检测能力。这可以导致破坏的数据作为有效在随后的读取操作上传递。

[0024] 在一些情况下, NAND 存储设备可以实施一些形式的磨损均衡 (leveling), 以增加首次磨损事件之前的时间。因为许多块可能同时接近磨损, 所以这样方法 (approach) 的副作用是进一步提高磨损群集。使用扇区 ECC 来监视位错误率可能遭受混叠效应 (aliasing effect), 并且可能限制检测磨损开始的能力。

[0025] 用于检测磨损的一些其它技术类似地缺少精确性。这些包括对于成功擦除和成功编程的测试。这些技术不能检测以下情况, 其中数据保持时间长于 100ns 的操作校验时间, 但是显著短于作为 15 个数量级扩展的 3×10^{17} ns (~ 10 年) 的标称耐久性。至少一些在此提供的实施例, 通过提供用于确定快闪存储器设备的保持性行为的技术, 处理了这些和其他限制。

[0026] 图 1 图示根据一个实施例的用于确定快闪存储器设备的至少一块的保持性行为

的方法 100。如所示,监视数据写入到快闪存储器设备的至少一块。见操作 102。

[0027] 快闪存储器设备可以包括具有存储器块的任何快闪存储器设备。例如,在一个实施例中,快闪存储器设备可以是 NAND 设备。

[0028] 在本说明的上下文中,由监视数据使用的数据模式可以参照已知的任何数据或数据模式。例如,在一个实施例中,监视数据可以包括其整个值对于存储器系统已知的数据模式,系统包括存储器或其它主机系统。在此情况下,可以根据数目、位置和模式表征错误。

[0029] 此外,在一个实施例中,可以生成监视数据。在此情况下,监视数据可以不作为用户数据传递,而是可以通过主机系统内部生成。监视数据可以写入到物理块,绕过任何磨损均衡。以此方式,监视数据可以不经磨损均衡。

[0030] 此外,可以选择快闪存储器设备的多个块用于存储监视数据,其中至少两个块不是相互物理邻近,例如,以提供设备不同区域的行为的采样。在任何情况下,监视数据可以写入到多组块,其中对于每组执行的写入循环的数目是不同的。

[0031] 应该注意,至少一块在写入监视数据之前已经多次被写入。例如,相同数据可以连续重写,不同数据可以在每个重写写入,以上两者的组合等。因此,最终存储在设备上的监视数据可能并非必定是在写入的每个循环期间使用的监视数据。

[0032] 在一段时间之后,读取监视数据。见操作 104。该时间段可以包括写入监视数据的点之后的任何长度的时间(例如,几秒、几分钟、几小时、几天等)。在一个实施例中,该时间段可以包括用户或系统定义的时间段。作为选项,该时间段可以是可配置的。在任何方法中,运行时时钟、实时时钟、日历等的输出可以用于测量数据年龄、时间段等。使用实时时钟是有利的,因为数据年龄可以容易地确定,而不管断电、功率损耗的时段等。

[0033] 一旦读取监视数据,就基于读取确定至少一块的保持性行为。见操作 106。在监视数据对于主机系统已知的情况下,保持性行为的确定可以包括比较读取的监视数据与其存储的版本。

[0034] 此外,监视数据的读取和至少一块的保持性行为的确定可以以周期性间隔发生。在此情况下,周期性间隔可以包括由用户定义或由系统自动定义的间隔。

[0035] 一旦确定保持性行为,就输出确定的结果。见操作 108。在一个实施例中,可以基于至少一块的保持性行为的确定结果,确定快闪存储器设备或其部分的操作极限。作为选项,当满足操作极限时,可以不再允许写入到快闪存储器设备或其一部分。

[0036] 优选地,通过本领域已知的任何方式,例如,通过将监视数据从可用数据空间的映射忽略写保护监视数据。应注意到,在一个实施例中,可以写保护具有监视数据的至少一块的至少一部分。例如,可以写保护具有监视数据的至少一块的至少一部分,以避免用用户数据盖写。

[0037] 在一个方法中,可以通过在存储器控制器中将块标记为不可写来完成写保护。在另一方法中,可以将受保护块或部分受保护块的指示存储在快闪存储器设备上。此外,在一个实施例中,监视数据可以包括小于大约 1% 的快闪存储器设备的存储器容量,并且更优选地小于大约 0.5%。在其它实施例中,小于 0.1% 的存储器容量可以用于存储监视数据。

[0038] 使用方法 100,可以通过添加监视数据到快闪存储器设备,测量闪存磨损和保持性。监视数据可以包括所有值对于主机系统已知的数据模式。因此,可以根据数目、位置和模式表征错误。

[0039] 在任何用户数据达到这些循环计数之前,一组块可以位于高循环计数。可以使用该数据进行错误率面的采样。一旦位于高循环计数,就可以按间隔检查监视数据以测量保持性行为。作为选项,可以经由自身监视、分析和报告技术(S. M. A. R. T.)来报告测量的磨损和数据保持性极限。

[0040] 如上面所述的,NAND 闪存设备可能遭受擦除 / 写入循环的有限耐久性和有限数据保持性。这是驱动电荷通过绝缘体的数据存储处理的不良可逆性的结果。结果,闪存中的位错误率可以表征为 3D 面。

[0041] 图 2 图示根据一个实施例的示出闪存错误率面的形状的代表性图示 200。如所示,X 轴以秒为单位示出从 1 毫秒到 10 年的数据年龄。深度轴示出从 1 到 1×10^6 循环的写入循环计数。垂直轴是错误率乘数。大多数闪存设备的错误率面具有如图 2 所示的一般特性。

[0042] 如果写入单元足够高的循环计数,则它们完全出故障(例如,1/2 的错误率)。因此,在图 2 所示的面的左后角的错误率在某一循环计数处增大为故障。此外,当数据的年龄增加时,在较低的循环计数处出现故障。因此,在右后角的错误率甚至更快地进入故障。设计 ECC 或加上循环冗余校验(CRC)的 ECC 取决于知道错误率或面的形状。

[0043] 许多 NAND 设备具有可能影响数据完整性的许多问题。这些可能包括电荷脱阱(detraping)、感生应力漏电流(SILC)、读取干扰和写入干扰、和 / 或其它问题。

[0044] 对于闪存的制造测试(例如,JEDEC、JESD47F 等)对于 IT 数据完整性是不足的。SILC 测试仅使用每批 38 个设备 500 小时。零错误在这样的测试中指示高达 6% 到 90% 置信度的换码率。此外,500 小时测试允许最大 10% 的设备测试到耐久性极限(假设剩下的设备完全没有测试),但是处于仅 100ns 的数据年龄。

[0045] 对于 IT 系统的数据完整性要求可能非常严格。返回破坏的数据通常不可接受。可以设置每 10M 单元年一个事件的宽松目标。以 10kIOPS 速率,这给出对于典型的 4kB I/O 传送的每位 1×10^{-23} 的未校正错误率。因此,使用标称的耐久性规范设计 IT 闪存系统可能承担重大风险。

[0046] 可以利用磨损均衡来增加时间,直到不存在耐久性循环计数之下的自由块。例如,可以通过备件(spare)或者甚至通过移动静态数据实现自由块。静态磨损均衡算法的一个目的是最大化循环计数,直到不存在耐久性极限之下的自由或可自由块。

[0047] 理想的静态磨损均衡算法将在任何时间点产生最窄的分布。这将为设备给出最大总循环计数。然而,在此情况下,磨损均衡假设存在用于确定块已经磨损的健壮技术。也就是说,磨损均衡假设用于磨损的累积分布函数(CDF)具有突然的和以已知循环计数的开始。此外,假设所有块具有相同的磨损特性。此外,忽略保持性的影响。

[0048] 图 3 示出根据一个实施例的示例性 CDF 300,该 CDF 300 图示由于在某一数据年龄的磨损,给定块将超过扇区级 ECC 的校正能力的概率。扇区级 ECC 可能在闪存中用于针对软错误(soft error)和其它影响进行保护。然而,ECC 的能力存在有限极限,并且该极限可能由于块磨损在某一点被超过。当其出现时,可能返回破坏的数据。如 CRC 的额外扇区检查移动其出现的点,但是可能不能消除该问题。

[0049] 图 3 中的曲线是根据一个实施例的实际现场行为处于短数据年龄可以是什么的代表。假设具有 10^5 循环的耐久性规范的设备,给定 JEDEC 测试结果,这可能分配到 6% 概率。在该示例中,假设在 10^8 循环 94% 的块出故障。在更长的数据年龄,曲线将移动到左边。

[0050] 可以在图 4 中看到可靠性影响。图 4 图示根据一个实施例的在某一数据年龄的存取分布和 CDF 之间的重叠的表示 400, 示出了关注的区域。

[0051] 使用图 4, 可以看到存取分布和 CDF 之间的重叠。存取分布越窄, 大百分比的块在通知磨损之前越可能处于磨损区。因此, 更可能将超过扇区 ECC 的能力, 并且将返回破坏的数据。

[0052] 为了处理这些和其它问题, 可以采样错误率面以辅助闪存设备的耐久性和保持性极限的确定。在一个实施例中, 这可以通过保留许多没有经过磨损均衡的物理块来实现。这些块可以包括对主机系统已知的数据, 使得值可以完全验证, 避免在使用从 ECC 确定的错误率中固有的任何混叠问题。

[0053] 此外, 可以在超过 ECC 能力的错误率使用这样的块, 而不危害数据完整性。还可以按间隔验证监视数据以测量错误率对保持性。

[0054] 在一个实施例中, 位于接近 CDF 的预期上升的一组完全擦除块, 可以用来辅助在磨损导致数据破坏之前确定磨损。例如, 闪存设备可以具有 16000 个擦除块。使用 15 个块作为监视数据将占用小于 0.1% 的设备容量。

[0055] 作为选项, 可以利用在 3 个一组的 5 组中安排的 15 块。每组可以具有相同的循环计数。在此情况下, 在一组中具有多个块可以改进测量的质量。可以选择每组的位置 (根据循环计数), 以接近预期 CDF 开始具有可能影响数据完整性的的大小的区域。

[0056] 作为示例, 表 1 示出根据一个实施例的闪存设备的特性代表。

[0057] 表 1

[0058]

容量	8GB	页面写入	800 μ s
页面	4kB	块擦除	1500 μ s
擦除块	128 页	页面读取	60 μ s
块	16384	块读取	21ms
耐久性	10 ⁵	块写入	116ms

[0059] 在该示例中, 用于擦除 / 写入 / 读取块的时间是 138ms。因此, 可以选择 5 个监视组的监视数据目标布局。表 2 示出根据一个实施例的 5 个监视组的监视数据目标布局的代表。

[0060] 表 2

[0061]

目标循环 (K)	20	36	63	112	200	总计
部署时间 (H)	2.3	4.1	7.3	12.9	23.0	50

[0062] 在此情况下, 假设磨损具有对数正态行为, 选择循环计数为幂律分布。在各种实施例中, 可以选择其它分布来最大化监视数据的效率。在该示例中, 将花费总计 50 小时来循环所有监视数据到最后计数。

[0063] 图 5 示出根据一个实施例的在某一时间点的用户和监视数据的代表性循环计数直方图 500。在此情况下, 用户数据和监视数据的直方图示出为写入循环计数的函数。此外, 数据包括磨损均衡。

[0064] 如所示的, 监视数据位于接近预期的磨损循环计数。在一些情况下, 循环所有监视数据块需要的时间可能长于出货之前希望执行的。在这样的情况下, 监视数据可以在出货之前部分循环, 循环的剩余部分出现在现场。

[0065] 一旦监视数据处于最后循环计数, 就可以按间隔读取块, 并且读取数据可以与已知数据比较。这允许错误的完全识别, 包括维持改变率的能力。

[0066] 图 6 示出来自监视数据系统的可能的一组测量 600。每个监视组示出为列组, 并且该组中的每个代表不同的数据年龄。每列的高度是每扇区的错误数目, 并且标注为“ECC 极限”的较低区域代表可能的 ECC 校正极限。如所示的, 一些测量可能超过 ECC 能力, 因此对于数据完整性造成风险。

[0067] 在一个实施例中, 可以从该数据得到操作极限。例如, 表 3 示出根据一个实施例, 可以为图 6 中的示例性数据生成的操作极限。

[0068] 表 3

[0069]

最小循环计数	最大循环计数	最大数据年龄
< 40k		4
40k	60k	3
60k	80k	2
80k	100k	2
100k	120k	1
>120k		不允许

[0070] 还可以选择监视数据模式。在各种实施例中, 这可以是静态模式或者为特定块计算。作为选项, 可以利用强调耐久性和 / 或保持性的模式。在任一情况下, 可以配置系统使得可以知道每个监视块的预期内容, 而不从闪存设备读取任何内容。当到达极限之一时, 可以通过如不再接受写入 (例如, 返回关于写入的错误等) S. M. A. R. T. 标志或其它技术通知系统。

[0071] 虽然大量上述描述已经将监视数据单元描述为包括整个闪存块, 但是所讨论的技术不限于使用整个闪存块。例如, 在一个实施例中, 监视数据单元可以与扇区大小一样小。

[0072] 尽管, 使用扇区大小监视数据单元将保持监视特性, 但是在一些情况下, 这可能导致暴露在 (exposure to) 硬错误 (hard error) 下。例如, 如果利用页面大小监视, 用户数据在块的剩余部分, 则如果监视页表现出接近 ECC 极限的错误率, 用户数据就可能被标记为丢失 (即, 硬错误)。这是因为扇区 ECC 在这种情况下可能不可靠。然而, 使用该技术, 破坏事件的概率将仍以较低的开销减小。

[0073] 图 7 图示根据一个实施例的用于分析来自多个快闪存储器设备的监视数据信息的方法 700。如所示的, 从多个快闪存储器设备收集监视数据信息, 监视数据是存储在已知写入循环计数的专用存储器单元中的已知内容的数据。见操作 702。

[0074] 监视数据可以包括已知的任何数据或数据模式。在一个实施例中, 监视数据信息可以包括与监视数据的数据年龄和写入循环计数中的至少一个相关的错误率信息。此外,

可以使用多种技术收集监视数据和 / 或监视数据信息。在一个实施例中,可以生成监视数据。

[0075] 一旦收集监视数据信息,就分析监视数据信息。见操作 704。此外,基于分析采取与设备中的至少一个相关的动作。见操作 706。

[0076] 动作可以包括多种动作。在一个实施例中,采取动作可以包括为设备中的至少一个设置行为目标。在这种情况下,行为目标可以是对于数据写入循环计数的限制。作为另一选项,行为目标可以是对于数据年龄的限制。

[0077] 在另一实施例中,采取动作可以包括为至少一个设备设置行为目标。在此情况下,行为目标可以是对于数据写入循环计数的限制。作为另一选项,行为目标可以是对于数据年龄的限制。

[0078] 应该注意,监视数据信息的收集和信息的分析可以发生任何次数。例如,收集和分析步骤可以执行至少两次,并且可以基于此生成统计趋势。此外,收集和分析步骤可以执行至少两次,并且可以跟踪监视数据的时间依赖性 (time-dependence)。

[0079] 通过实施该技术,可以通过从一组设备收集监视数据信息、分析数据和为设备设置行为目标,提供用于增加基于闪存的存储系统的可靠性的系统和方法。监视数据可以包括根据每个设备的循环计数和数据年龄的错误率信息。此外,行为目标可以包括对循环计数和数据年龄设置设备等级极限。集合来自多个设备的数据允许增加采样大小,并且可以帮助识别统计趋势。此外,集合还可以跟踪数据的时间依赖性。

[0080] 可以在一时间段之后更新行为目标。在一个方法中,可以再次收集和分析监视数据,以便在从设置之前的行为目标时已经过去一时间段之后,生成更新的行为目标。这将因此考虑监视数据的较长年龄的影响,潜在地改进行为目标。在另一方法中,可以至少部分基于其它因素(如存储器设备的故障事件、因为设置目标所以变得可用的额外信息或参数等)更新行为目标。

[0081] 在一个实施例中,可以利用监视数据方法来提供存储系统中的一组闪存设备中的错误率面的测量。此外,可以为给定闪存设备提供错误率对循环计数和错误率对数据年龄的测量。

[0082] 图 8 图示根据一个实施例的代表性存储系统 800。如所示的,存储系统 800 可以包括存储控制器 802 和多个闪存设备 804。可以响应于监视数据信息调节系统 800 的行为。控制器 802 可以与远程数据储存库 806 通信,远程数据储存库 806 从这种存储系统集合现场数据。

[0083] 控制器 802 可以基于该数据,确定该组本地闪存设备 804 的可能的初始循环计数目标,或确定为工厂预置。这些值可以传递到闪存设备 804,闪存设备 804 将循环它们的监视数据以将其置于预定的循环计数。存储控制器 802 然后可以收集在不同次数的监视数据的错误特性。

[0084] 在一个实施例中,存储控制器 802 可以包括或访问实时时钟。该时钟可以用于记录监视数据的写入次数(在监视数据中或在另一保留区域中),并且确定监视数据的年龄。

[0085] 由监视数据报告的错误特性可以通过接口收集到闪存设备,如经由 S. M. A. R. T.。存储系统 800 可以从闪存设备 804 集合收集的数据以改进统计,并且识别离群值(outlier)。

[0086] 例如,存储系统 800 可以包括 16 个闪存单元 (例如,SSD),每个闪存单元可以包括 16 个闪存设备。这提供 256 个闪存设备,因此改进采样的质量 16 倍 (假设泊松统计)。如果每个闪存设备具有每个数据组中的 3 块,则系统数据组大小将是每组 768 块采样。因此,可能获得好的统计并且执行离群值检测。

[0087] 监视数据可以报告每个监视数据组的错误配置。每组可以具有不同的循环计数。因此,存储控制器 802 可以使用它们来估计总体的错误率曲线。错误率曲线可以用于选择耐久性和保持性极限标准。

[0088] 作为可以如何在系统 800 中使用数据的示例,可以执行蒙特卡洛模拟以生成每个设备上的监视数据的一组错误计数测量。在此情况下,假设监视数据至少报告每个监视数据组的平均错误计数。合并每个循环计数的错误计数值以计算平均和标准偏差。

[0089] 假设高斯扩展,可以计算换码极限目标,外部 ECC 将能够超过该换码极限目标来校正。还假设换码的目标概率是 1×10^{-6} ,其应该足以避免设备到设备校正压制 (overwhelming) 外部 ECC。这对于该示例计算出 4.8 标准偏差 (假设高斯统计)。表 4 示出用于该示例的蒙特卡洛模拟的结果。

[0090] 表 4

[0091]

循环计数 (k)	20	36	63	112	200
平均错误/扇区	1.4	2.4	4.2	7.4	13.2
Σ 错误/扇区	0.2	0.3	0.4	0.6	1.1
Σ 修剪曲线	2.2	3.7	6.1	10.4	18.4
离群值数	1	1	1	1	1

[0092] 在此情况下,以 5 个分开的循环计数采样监视数据。“平均”行列出了每扇区的错误的平均数,并且“ Σ ”行列出相应的标准偏差。“ Σ 修剪 (clip) 曲线”行是在每个循环计数的 4.8 标准偏差值。最后一行是具有超过 4.8 标准偏差的错误计数的设备数。

[0093] 在该示例中,仅存在单个设备超过修剪,因此假设这是离群值。此外,假设选择每扇区 8 个错误的错误极限。这将由扇区级错误校正的能力确定。

[0094] 给出这样的值,可以通过插值 Σ 修剪曲线为非离群值估计循环计数极限。对于该示例,极限值是 85k 写入循环。类似地,在该示例中离群值极限可以计算为 61k 循环。还可以用离群值利用更保守的方法。例如,可以选择较低循环计数,而不是简单插值。类似地,可以使用估计的其它技术,如曲线拟合。

[0095] 图 9 示出根据该示例的结果的曲线图 900。曲线图 900 图示离群值完全超过分布的边缘,因此应用正常修剪极限将可能导致错误暴露。一旦已经计算循环计数极限,它们就可以反馈回闪存设备或负责耐久性极限的控制器。这可以在不同的次数重复,因此反映不同的数据年龄。如此,可以为存储系统生成一组耐久性和保持性目标。

[0096] 由于每个设备受益于总体统计,因此该方法对允许每个设备设置它自己的极限有改进,并且可以从采样噪声更精确地区分离群值。每个监视数据组可以包括多个擦除块,每个擦除块包括许多数据扇区。例如,典型的 NAND 设备可以具有 528kB 擦除块,包括 1024 扇区,每个扇区具有 528 字节。假设每个监视组使用 3 个数据块,对于整个 256 个设备的组,

每组有 3.3×10^9 位。因此,因为避免了过度杀伤 (overkill),所以可靠性增强,并且成本降低。

[0097] 应该注意,本发明可以采用整体硬件实施例、整体软件实施例或包含硬件或软件元素的实施例的形式。例如,在一个实施例中,系统可以包括具有多个存储器块的快闪存储器设备,其中块中的至少一个具有写入其中的监视数据,并且至少一块在写入监视数据之前已经写入多次。系统还可以包括用于寻址块的电路和存储监视数据的副本的次级存储器。此外,系统可以包括存储控制器和多个快闪存储器设备。

[0098] 如本领域的技术人员将理解的,本发明可以体现为系统、方法或计算机程序产品。因此,本发明可以采用以下的形式:完全硬件实施例、完全软件实施例(包括固件、驻留软件、微代码等),或通常可以在此称为“电路”、“模块”或“系统”的组合软件和硬件方面的实施例。此外,本发明可以采用体现在任何有形表达介质中的计算机程序产品的形式,该有形表达介质具有在介质中体现的计算机可用程序代码。

[0099] 可以利用一个或多个计算机可用或计算机可读介质的任何组合。计算机可用或计算机可读介质例如可以是但不限于电的、磁的、光的、电磁的、红外线的或半导体的系统、装置、器件或传播介质。计算机可读介质的更具体的示例(非穷举列表)将包括以下:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EEPROM或快闪存储器)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光学存储器件、传输介质(如支持因特网或内联网的传输介质)、或磁存储器件。注意,计算机可用或计算机可读介质可能甚至是上面印有程序的纸张或其它合适的介质,由于可以例如经由光学扫描纸张或其它介质,以电子方式捕获程序,然后以合适的方式编译、解释或其它的处理,如果必要,然后存储在计算机存储器中。在本文的语境中,计算机可用或计算机可读介质可以是任何可含有、存储、传达、传播、或传输供指令执行系统、装置或器件使用的、或与指令执行系统、装置或器件相联系的程序的介质。计算机可用的介质可包括在基带中或者作为载波一部分传播的、由其体现计算机可用的程序码的数据信号。计算机可用程序代码可以使用任何适当的介质传输,包括但不限于无线、有线、光缆、RF等。

[0100] 用于执行本发明的操作的计算机程序代码,可以以一个或多个编程语言的任何组合书写,编程语言包括面向对象的编程语言(如Java、Smalltalk、C++等)和常规的过程编程语言(如“C”编程语言或类似编程语言)。程序代码可以完全地在用户的计算机上执行、部分地在用户的计算机上执行、作为一个独立的软件包执行、部分在用户的计算机上和部分在远程计算机上或完全地在远程计算机或服务器上执行。在后一情形中,远程计算机可以通过任何类型的网络(包括局域网(LAN)或广域网(WAN))连接到用户的计算机,或者可以(例如,利用因特网服务提供商通过因特网)连接到外部计算机。

[0101] 在此参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图图示和/或框图描述了本发明。将理解,流程图图示和/或框图中的每一框、以及流程图图示和/或框图中的各方框的组合,可以通过计算机程序指令实现。这些计算机程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,以产生一种机器,使得通过计算机或其它可编程数据处理装置的处理器执行的指令,产生实现流程图和/或框图框或各方框中规定的功能/动作的装置(means)。

[0102] 这些计算机程序指令还可以存储在计算机可读介质中,其能指示计算机或其它可

编程数据处理装置,以特定方式起作用,使得存储在计算机可读介质中的指令产生一种包括指令装置(instruction means)的制品,该指令装置实现流程图和/或框图的一个或多个方框中规定的功能/操作。

[0103] 计算机程序指令还可以加载到计算机或其它可编程数据处理装置上,使得在计算机或其它可编程装置上执行一系列操作步骤,以产生计算机实现的过程,从而在计算机或其它可编程装置上执行的指令,提供用于实现在流程图和/或框图的一个或多个方框中规定的功能/动作的过程。

[0104] 附图中的流程图和框图,图示了根据本发明的各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表模块、程序段、或代码的一部分,所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。还应该注意,在一些替代实施例中,方框中所标注的功能也可以以脱离附图中所标注的顺序发生。例如,依所涉及的功能,两个接连地表示的方框实际上可以基本并行地执行,或者该方框有时可以按相反的顺序执行。还将注意,框图和/或流程图图示中的每个方框、以及框图和/或流程图图示中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统、或者专用硬件与计算机指令的组合来实现。

[0105] 适于存储和/或执行程序代码的数据处理系统将包括至少一个处理器,该处理器通过系统总线直接或间接耦合到存储器元件。存储器元件可以包括在程序代码的实际执行期间采用的本地存储器、大容量存储器和高速缓冲存储器,该高速缓冲存储器提供至少一些程序代码的临时存储,以便减少在执行期间必须从大容量存储器取回代码的次数。

[0106] 输入/输出或 I/O 设备(包括但不限于键盘、显示器、指向设备等)可以直接或通过居间 I/O 控制器耦合到系统。

[0107] 网络适配器也可以耦合到系统,以使数据处理系统能够变得通过居间专用或公用网络耦合到其它数据处理系统、或远程打印机、或存储设备。调制解调器、线缆调制解调器和以太网卡正好是几个当前可用类型的网络适配器。

[0108] 图 10 图示根据一个实施例的网络架构 1000。作为选项,图 1-9 的功能和架构可以在网络架构 1000 的环境中实现。然而,图 1-9 的功能和架构可以在任何希望环境中执行。应该注意,前述限定可以在本描述期间应用。

[0109] 如图 10 所示,提供多个远程网络 1002,包括第一远程网络 1004 和第二远程网络 1006。网关 1007 可以耦合在远程网络 1002 和最近网络 1008 之间。在本网络架构 1000 的环境中,网络 1004、1006 每个可以采取任何形式,包括但不限于 LAN、如因特网、PSTN、内部电话网的 WAN 等。

[0110] 在使用中,网关 1007 用作从远程网络 1002 到最近网络 1008 的进入点。如此,网关 1007 可以用作路由器和交换机,路由器能够引导到达网关 1007 的给定数据分组,交换机提供用于给定分组的进出网关 1007 的实际路径。

[0111] 还包括耦合到最近网络 1008 的至少一个数据服务器 1014,并且可经由网关 1007 从远程网络 1002 访问数据服务器 1014。应该注意到,一个(或多个)数据服务器 1014 可以包括任何类型的计算设备/群件。多个用户设备 1016 耦合到每个数据服务器 1014。这样的用户设备 1016 可以包括桌面型计算机、膝上型计算机、手持计算机、打印机或任何其它类型的逻辑电路。应该注意,在一个实施例中,用户设备 1017 还可以直接耦合到任何网

络。

[0112] 传真机 1020 或一系列传真机 1020 可以耦合到一个或多个网络 1004、1006、1008。应该注意,数据库和 / 或额外组件可以与耦合到网络 1004、1006、1008 的任何类型的网络元件一起利用,或者集成到耦合到网络 1004、1006、1008 的任何类型的网络元件中。在本描述的上下文中,网络元件可以指网络的任何组件。

[0113] 图 11 示出根据一个实施例的与图 10 的用户设备 1016 相关联的代表性硬件环境。这样的图图示工作站的典型硬件配置,该工作站具有如微处理器的中央处理单元 1110、以及经由系统总线 1112 互连的许多其它单元。

[0114] 图 11 所示的工作站包括随机存取存储器 (RAM) 1114、只读存储器 (ROM) 1116、用于将如磁盘存储单元 1120 的外围设备连接到总线 1112 的 I/O 适配器 1118、用于连接键盘 1124、鼠标 1126、扬声器 1128、麦克风 1132 和 / 或其它用户接口设备 (如触摸屏和数字相机 (未示出)) 到总线 1112 的用户接口适配器 1122、用于将工作站连接到通信网络 1135 (例如,数据处理网络) 的通信适配器 1134、以及用于将总线 1112 连接到显示设备 1138 的显示适配器 1136。

[0115] 工作站可以具有驻留其上的操作系统,如 Microsoft Windows® 操作系统 (OS)、MAC OS 或 UNIX 操作系统。将理解,优选实施例还可以在不同于上面提到的那些的平台和操作系统上实现。优选实施例可以使用 JAVA、XML、C 和 / 或 C++ 语言、或者其它编程语言连同面向对象的编程方法一起编写。可以使用变得越来越多地用于开发复杂应用的面向对象的编程 (OOP)。

[0116] 虽然上面已经描述了各种实施例,但是应该理解的是它们仅仅通过示例而不是限制的方式呈现。因此,优选实施例的宽度和范围不应由上述任何示例性实施例限制,而是应该根据权利要求和其等价来限定。

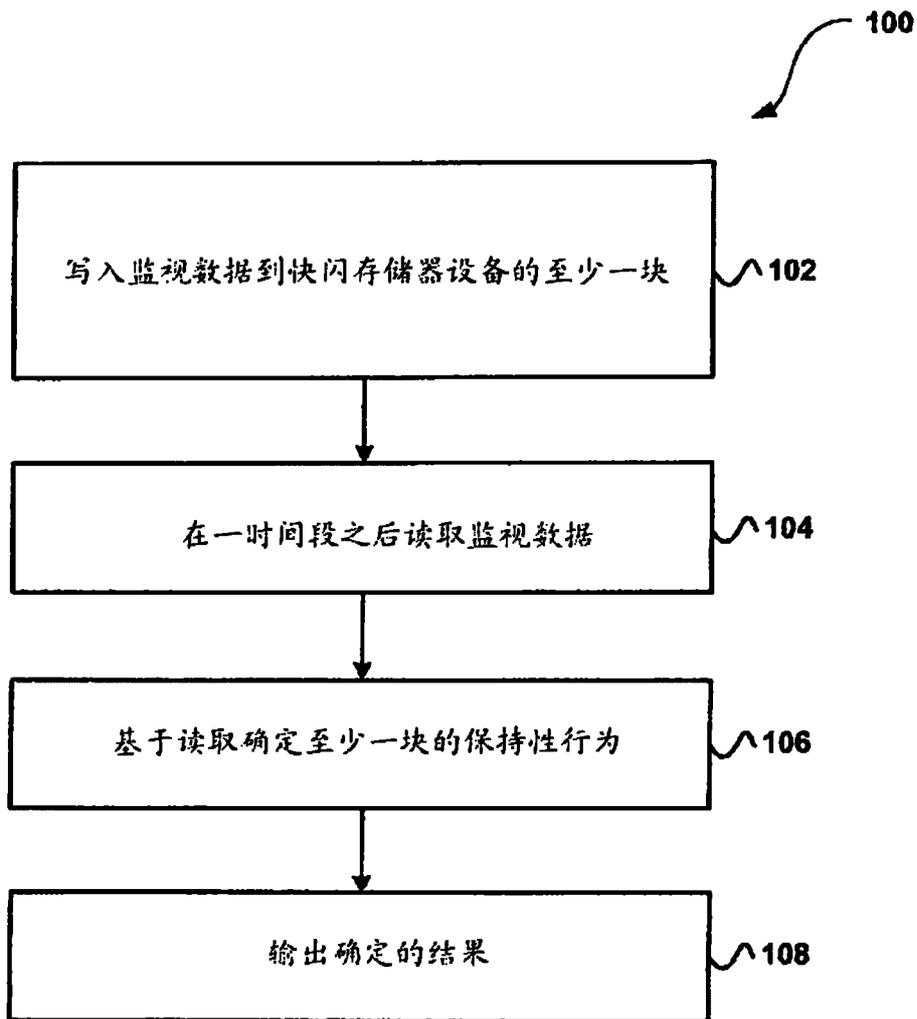


图 1

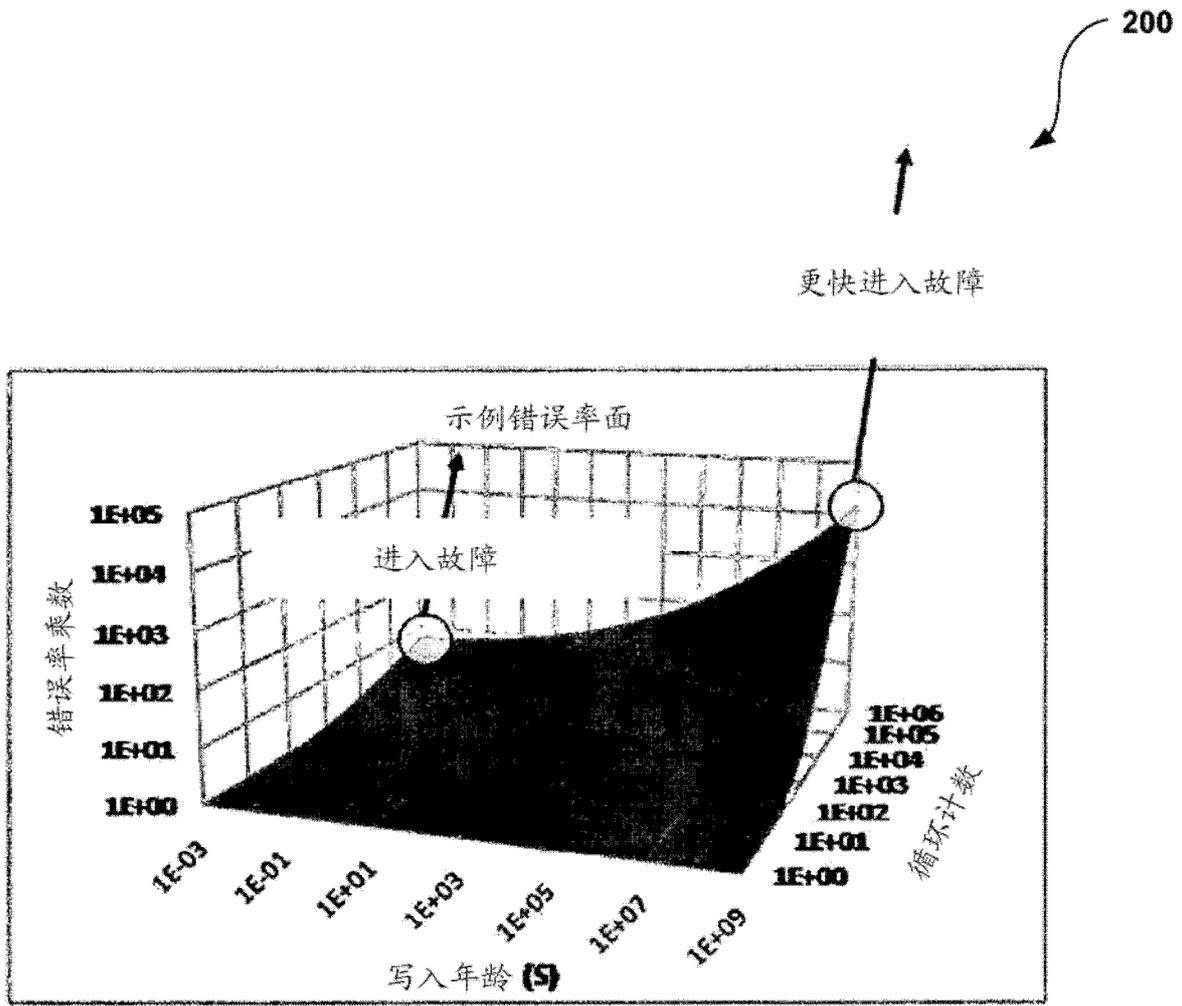


图 2

300

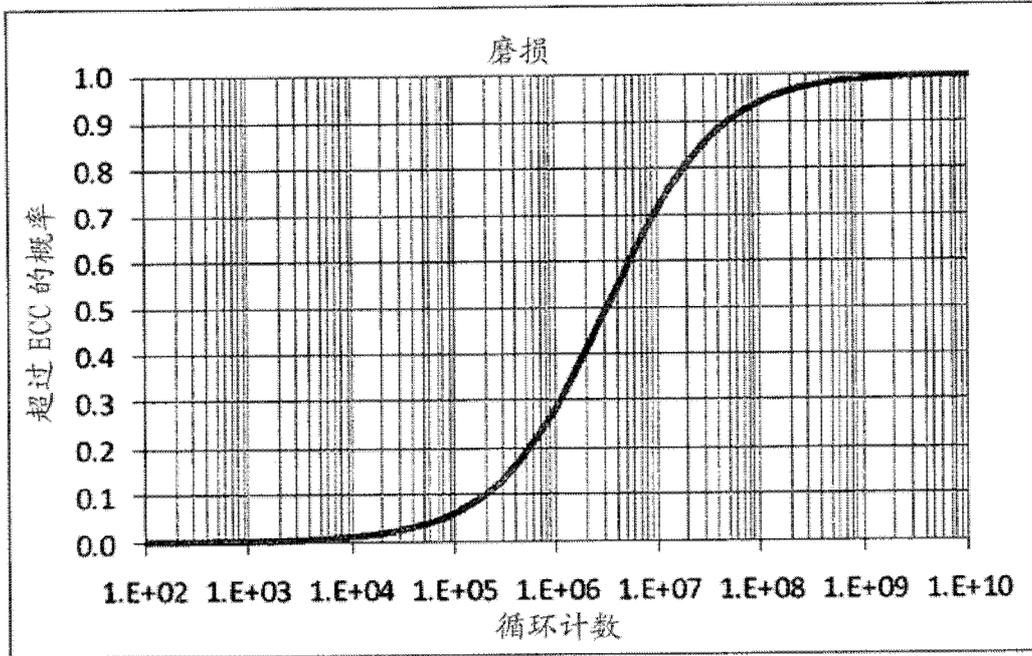


图 3

400

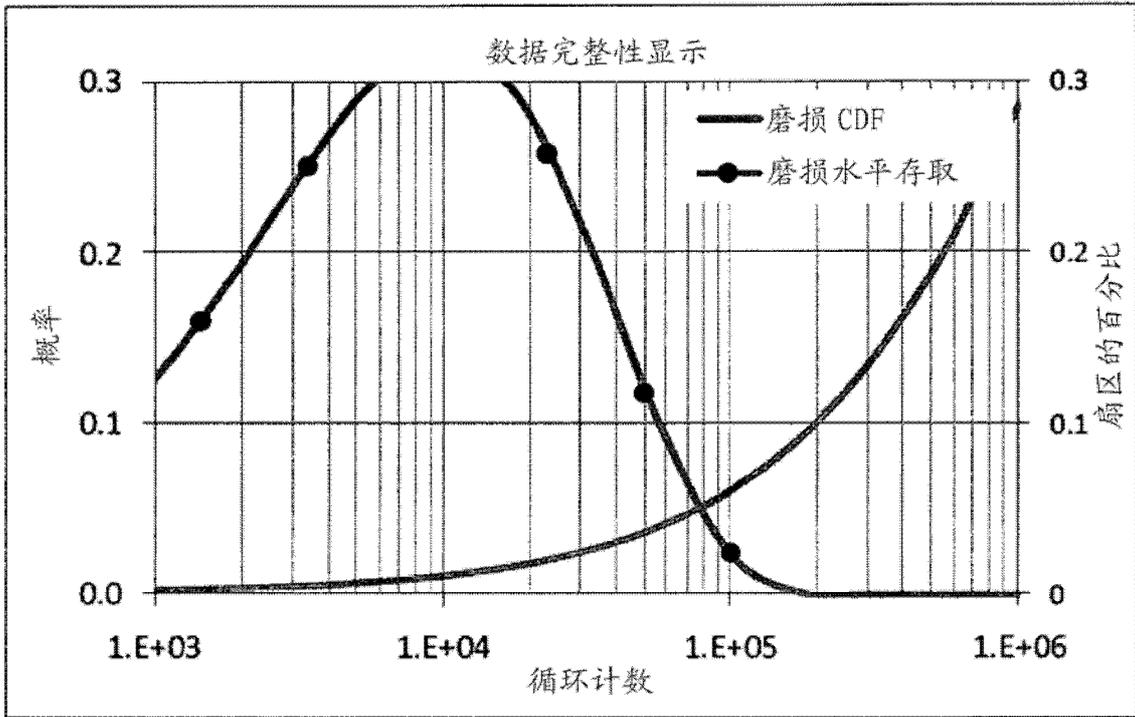


图 4

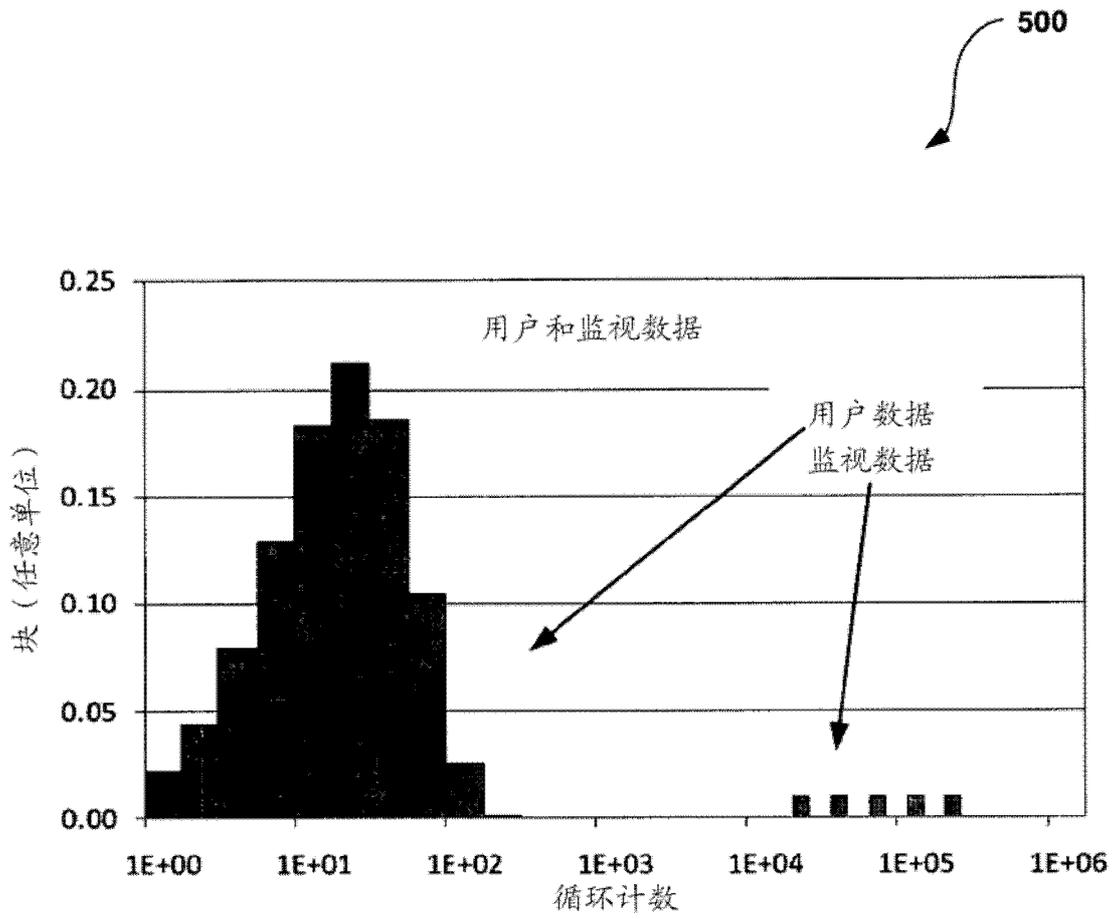


图 5

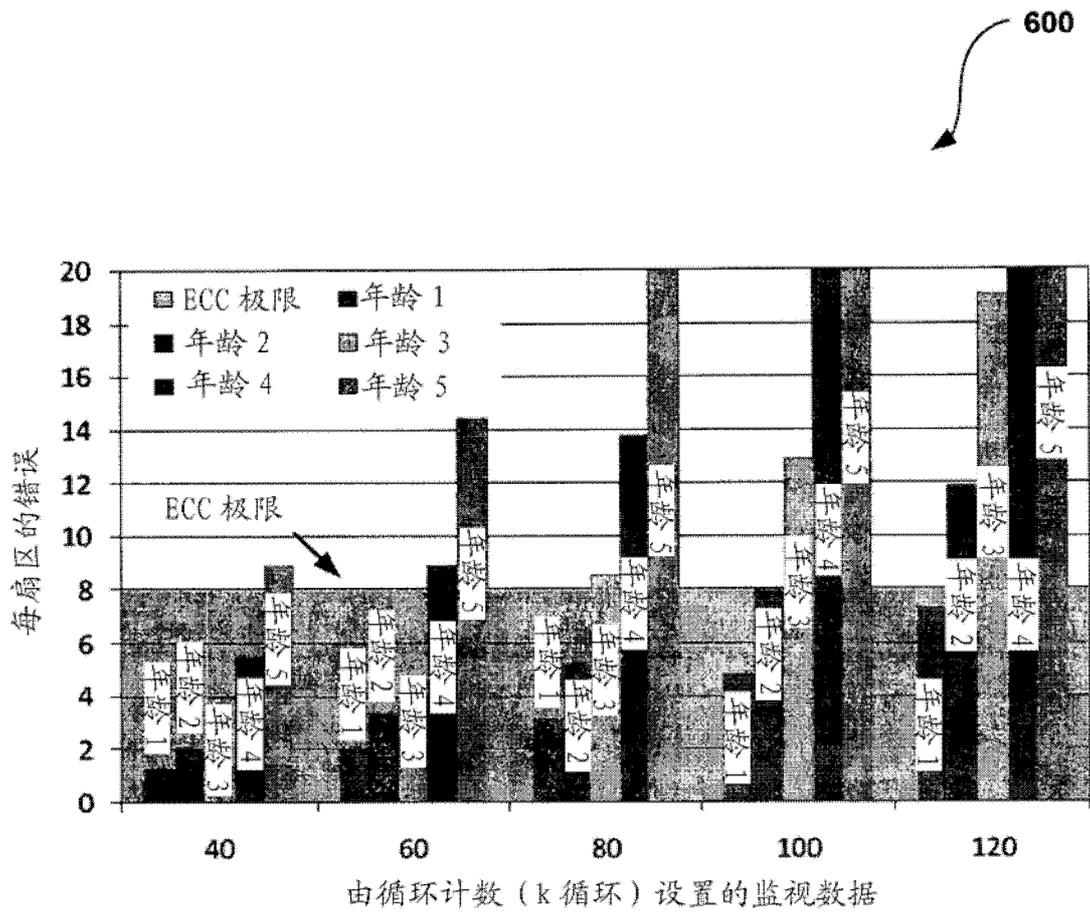


图 6

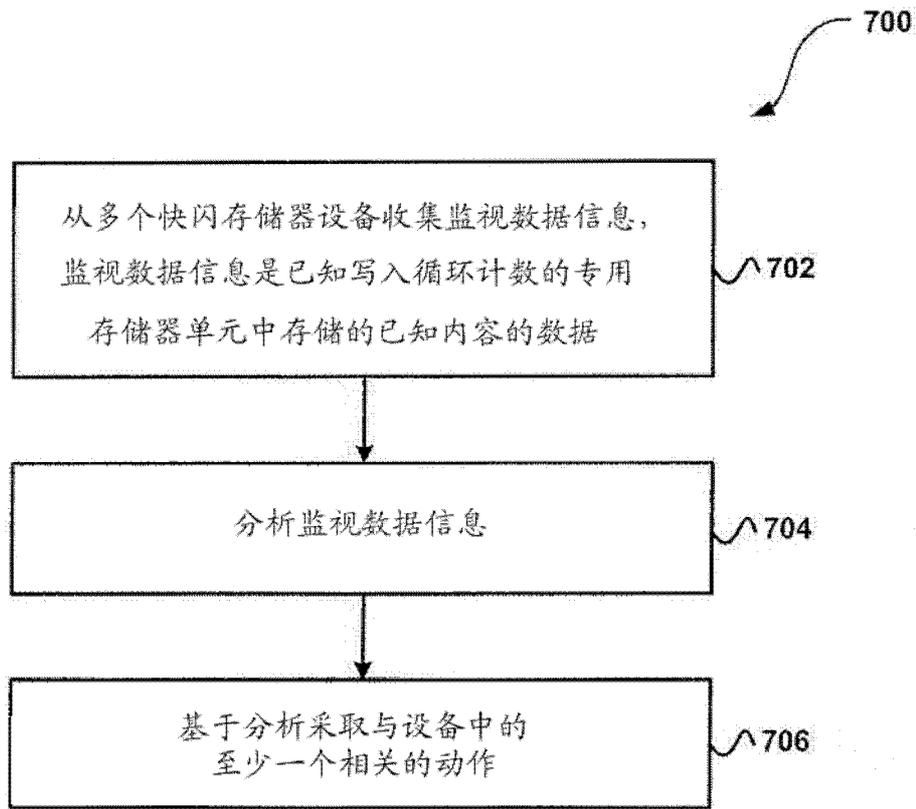


图 7

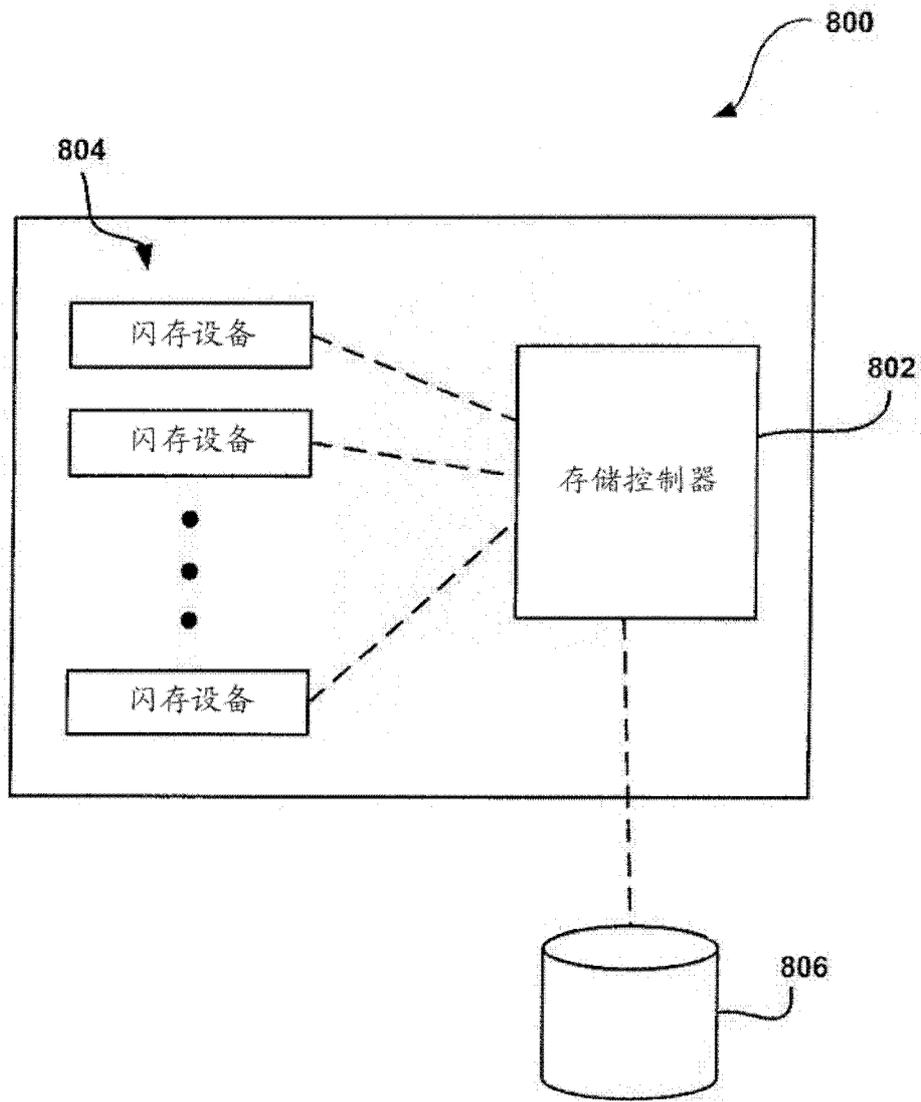
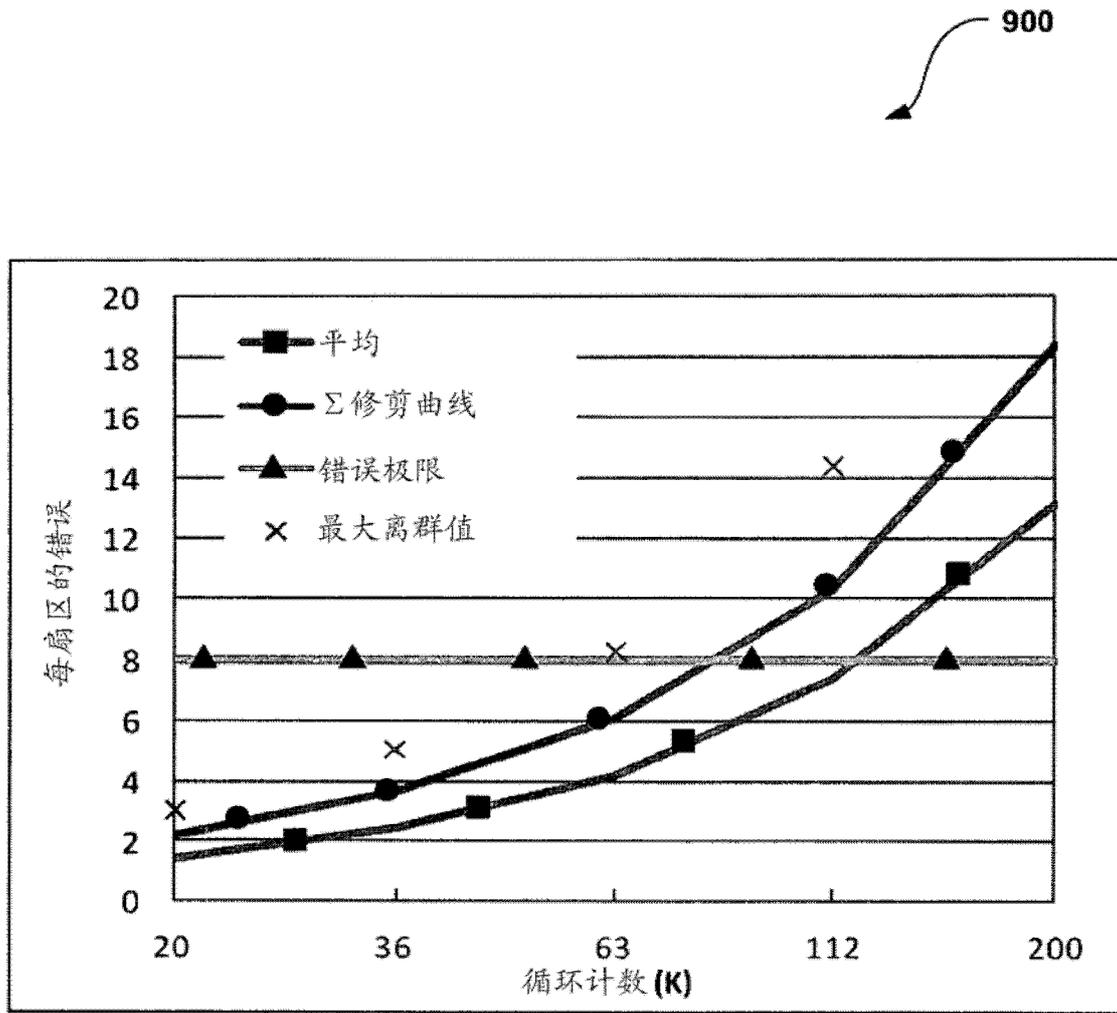


图 8



900

图 9

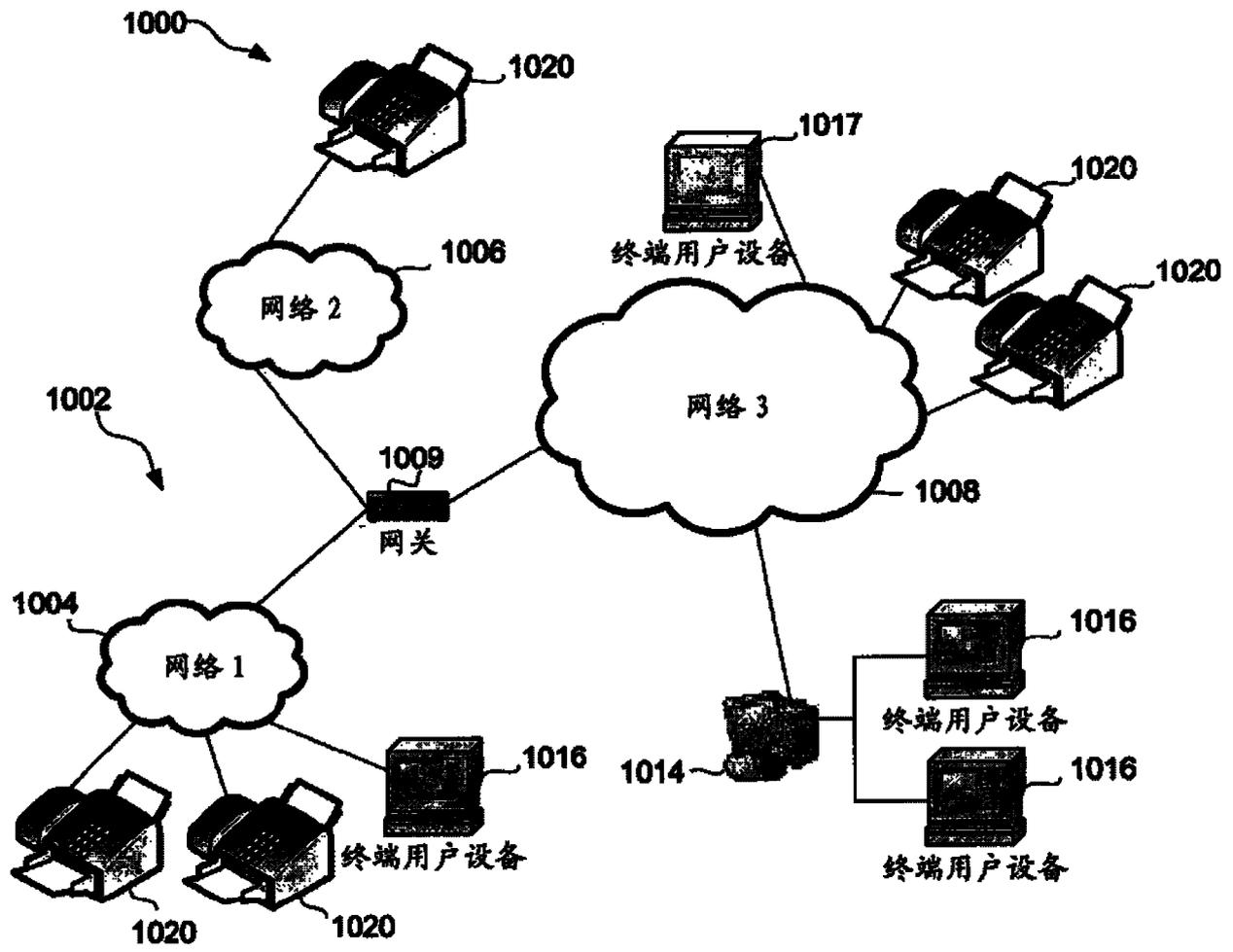


图 10

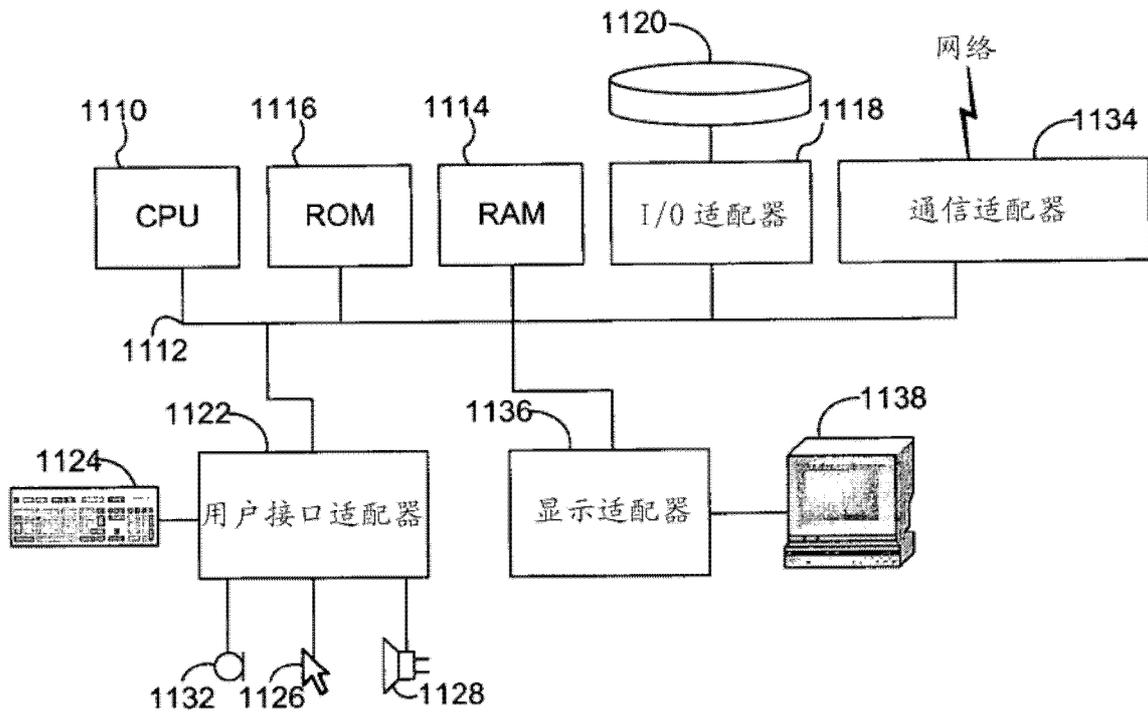


图 11