



(12)发明专利

(10)授权公告号 CN 105009085 B

(45)授权公告日 2018.06.15

(21)申请号 201380074108.9

(22)申请日 2013.08.29

(65)同一申请的已公布的文献号

申请公布号 CN 105009085 A

(43)申请公布日 2015.10.28

(30)优先权数据

2013-055548 2013.03.18 JP

(85)PCT国际申请进入国家阶段日

2015.08.31

(86)PCT国际申请的申请数据

PCT/JP2013/073859 2013.08.29

(87)PCT国际申请的公布数据

W02014/147865 EN 2014.09.25

(73)专利权人 东芝存储器株式会社

地址 日本东京都

(72)发明人 桥本大辅 永井宏一 渡部孝纪

(74)专利代理机构 北京市中咨律师事务所

11247

代理人 李峥 刘薇

(51)Int.Cl.

G06F 11/00(2006.01)

G06F 12/02(2006.01)

G11C 16/34(2006.01)

G06F 11/20(2006.01)

(56)对比文件

US 2010/0306577 A1,2010.12.02,

CN 102221981 A,2011.10.19,

JP 2005190187 A,2005.07.14,

CN 1848067 A,2006.10.18,

US 2010/0125751 A1,2010.05.20,

CN 101796494 A,2010.08.04,

审查员 王敏

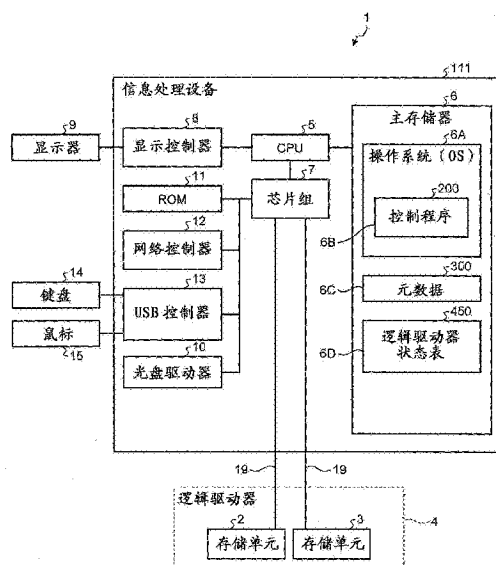
权利要求书11页 说明书66页 附图74页

(54)发明名称

信息处理系统、控制程序以及信息处理设备

(57)摘要

根据一个实施例,在第一存储单元的存储状态被识别为受保护状态时,控制单元将数据写入第二存储单元。在读取目标地址记录在数据迁移日志区域中时,所述控制单元从所述第二存储单元读取数据。在所述读取目标地址未记录在所述数据迁移日志区域中时,所述控制单元从所述第一存储单元读取数据。



1. 一种信息处理设备,其能够连接第一存储单元和第二存储单元,所述第一存储单元包括可读取和可写入的第一存储器,所述第二存储单元存储有数据迁移日志区域,包括可读取和可写入的第二存储器,

所述信息处理设备包括控制单元,所述控制单元被配置为执行:

第一处理,其用于基于从所述第一存储单元获得的可靠性信息,确定所述第一存储单元的存储状态;

第二处理,其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时,仅对所述第一和第二存储单元中的所述第二存储单元执行写入,在读取目标地址记录在所述数据迁移日志区域中时,执行从所述第二存储单元读取数据,在所述读取目标地址未记录在所述数据迁移日志区域中时,执行从所述第一存储单元读取数据;

第三处理,其用于根据由所述第二处理执行的写入,将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中;以及

第四处理,其用于在所述第一处理未将所述第一存储单元的存储状态识别为受保护状态时,执行从所述第一存储单元读取数据而不从所述数据迁移日志区域读取。

2. 根据权利要求1所述的信息处理设备,其中,

所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为所述受保护状态。

3. 根据权利要求1所述的信息处理设备,其中,

所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态识别为所述低可靠性状态时,识别所述第二存储单元处于存储所述第一存储单元中的数据的数据迁移目的地状态,并且用于将所述第一存储单元的存储状态更改为所述受保护状态。

4. 根据权利要求1所述的信息处理设备,其中,所述信息处理设备可连接到一个或多个第三存储单元,所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

所述控制单元,

管理所述第一和第二存储单元中的存储的数据的地址,以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时,将所述第一存储单元和所述第二存储单元识别为同一逻辑单元;以及

在将所述第一存储单元的存储状态识别为受保护状态时,控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

5. 根据权利要求1所述的信息处理设备,其中,

所述控制单元,

在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数

据,并且用于在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时,确定结束所述第一存储单元的受保护状态。

6. 根据权利要求1所述的信息处理设备,其中,
所述控制单元,

在所述第三处理中,在将第一地址记录在所述数据迁移日志区域中时,当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时,将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中,并且使所述第二地址无效。

7. 根据权利要求5所述的信息处理设备,其中,所述控制单元,在所述第一存储单元的受保护状态结束时,删除所述数据迁移日志区域。

8. 一种信息处理设备,其能够连接第一存储单元、第二存储单元以及数据迁移日志区域,所述第一存储单元包括可读取和可写入的作为非易失性半导体存储器的第一存储器,所述第二存储单元存储有数据迁移日志区域,包括可读取和可写入的作为非易失性半导体存储器的第二存储器,

所述信息处理设备包括控制单元,所述控制单元被配置为执行:

第一处理,其用于基于从所述第一存储单元获得的可靠性信息,确定所述第一存储单元的存储状态;

第二处理,其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时,仅对所述第一和第二存储单元中的所述第二存储单元执行写入,在读取目标地址记录在所述数据迁移日志区域中时,执行从所述第二存储单元读取数据,在读取目标地址未记录在所述数据迁移日志区域中时,执行从所述第一存储单元读取数据;

第三处理,其用于根据由所述第二处理执行的写入,将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中;以及

第四处理,其用于在所述第一处理未将所述第一存储单元的存储状态识别为受保护状态时,执行从所述第一存储单元读取数据而不从所述数据迁移日志区域读取。

9. 根据权利要求8所述的信息处理设备,其中,
所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为所述受保护状态。

10. 根据权利要求8所述的信息处理设备,其中,
所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的所述存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态识别为所述低可靠性状态时,识别所述第二存储单元处于存储在所述第一存储单元中的数据的迁移目的地状态,并且用于将所述第一存储单元的存储状态更改为所述受保护状态。

11. 根据权利要求8所述的信息处理设备,其中,所述信息处理设备可连接到一个或多

个第三存储单元,所述一个或多个第三存储单元包括可读取和可写入的第三存储器,
所述控制单元,

管理所述第一和第二存储单元中的存储的数据的地址,以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时,将所述第一存储单元和所述第二存储单元识别为同一逻辑单元;以及

在将所述第一存储单元的存储状态识别为受保护状态时,控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

12. 根据权利要求8所述的信息处理设备,其中,
所述控制单元,

在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数据,并且用于在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时,确定结束所述第一存储单元的受保护状态。

13. 根据权利要求8所述的信息处理设备,其中,
所述控制单元,

在所述第三处理中,在将第一地址记录在所述数据迁移日志区域中,当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时,将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中并且使所述第二地址无效。

14. 根据权利要求12所述的信息处理设备,其中,所述控制单元,在所述第一存储单元的受保护状态结束时,删除所述数据迁移日志区域。

15. 一种信息处理设备,其能够连接第一存储单元和第二存储单元,所述第一存储单元包括可读取和可写入的第一存储器,所述第二存储单元包括可读取和可写入的第二存储器,

所述信息处理设备包括控制单元,所述控制单元被配置为执行:

第一处理,其用于基于从所述第一存储单元获得的可靠性信息,确定所述第一存储单元的存储状态;

第二处理,其用于在所述第一处理将所述第一存储单元的存储状态确定为正常状态时,仅对所述第一存储单元执行写入,在所述第一处理将所述第一存储单元的存储状态识别为是受保护状态时,仅对所述第二存储单元执行写入;以及

第三处理,其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态后发生了针对所述第一存储单元的读取时,从所述第一存储单元读取数据,并将读取的所述数据写入所述第二存储单元。

16. 根据权利要求15所述的信息处理设备,其中,
所述控制单元,

在所述第三处理中,并行地执行从所述第一存储单元读取数据的处理和将读取的所述数据写入所述第二存储单元中的处理。

17. 根据权利要求15所述的信息处理设备,其中,
所述信息处理设备连接到存储元数据的元数据存储单元,

所述控制单元，

在所述第一处理确定所述第一存储单元的存储状态为所述正常状态时，在所述元数据中记录写入所述第一存储单元中的数据的地址和所述第一存储单元的识别信息的映射，并且在所述第一处理确定所述第一存储单元的存储状态为所述受保护状态时，在所述元数据中记录写入所述第二存储单元中的数据的地址和所述第二存储单元的识别信息的映射，

在读取处理中，在读取目标地址和所述第一存储单元的识别信息的映射被记录在所述元数据中时，从所述第一存储单元读取数据，并且在读取目标地址和所述第二存储单元的映射被记录在所述元数据中时，从所述第二存储单元读取数据。

18. 根据权利要求15所述的信息处理设备，其中，

所述信息处理设备与数据迁移日志区域连接，

所述控制单元，

在所述第三处理中，

在所述第一处理将所述第一存储单元的存储状态确定为受保护状态时，将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中，

在读取处理中，在所述数据迁移日志区域中存储有读取目标地址时，从所述第二存储单元读取数据，在所述数据迁移日志区域中未存储有读取目标地址时，从所述第一存储单元读取数据。

19. 根据权利要求15所述的信息处理设备，其中，

所述控制单元，

在所述第一处理中执行所述可靠性信息与阈值的比较，并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时，确定所述第一存储单元的存储状态为所述受保护状态。

20. 根据权利要求15所述的信息处理设备，其中，

所述控制单元，

在所述第一处理中执行所述可靠性信息与阈值的比较，并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时，确定所述第一存储单元的存储状态为低可靠性状态，

在所述第一处理将所述第一存储单元的存储状态确定为所述低可靠性状态时，识别所述第二存储单元处于存储所述第一存储单元中的数据的数据迁移目的地状态，并且用于将所述第一存储单元的存储状态更改为所述受保护状态。

21. 根据权利要求15所述的信息处理设备，其中，

所述信息处理设备可连接到一个或多个第三存储单元，所述一个或多个第三存储单元包括可读取和可写入的第三存储器，

所述控制单元，

管理所述第一和第二存储单元中的存储的数据的地址，以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内；

在将所述第一存储单元的存储状态识别为受保护状态时，将所述第一存储单元和所述第二存储单元识别为同一逻辑单元；以及

在将所述第一存储单元的存储状态识别为受保护状态时，控制所述一个或多个第三存

储单元和所述逻辑单元以便构成RAID。

22. 一种信息处理设备, 其能够连接第一存储单元、第二存储单元以及数据迁移日志区域, 所述第一存储单元包括可读取和可写入的第一存储器, 所述第二存储单元包括可读取和可写入的第二存储器,

所述信息处理设备包括控制单元, 所述控制单元被配置为执行:

第一处理, 其用于基于从所述第一存储单元获得的可靠性信息, 确定所述第一存储单元的存储状态;

第二处理, 其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时, 仅对所述第一和第二存储单元中的所述第二存储单元执行写入, 在读取目标地址记录在所述数据迁移日志区域中时, 执行从所述第二存储单元读取数据, 在读取目标地址未记录在所述数据迁移日志区域中时, 执行从所述第一存储单元读取数据;

第三处理, 其用于根据由所述第二处理执行的写入, 将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中; 以及

第四处理, 其用于在所述第一处理确定所述第一存储单元的存储状态为受保护状态之后, 在接收到针对存储在所述第一存储单元中的数据的删除处理命令时, 将所述删除处理命令所针对的删除目标数据的地址记录在所述数据迁移日志区域中。

23. 根据权利要求22所述的信息处理设备, 其中,

所述控制单元,

在所述第四处理中,

在所述第一处理确定所述第一存储单元的存储状态为受保护状态之后, 接收到存储在所述第一或第二存储单元中的数据的删除处理命令时, 在所述数据迁移日志区域中存储有所述删除处理命令所针对的删除目标数据的地址时, 将接收到的所述删除处理命令向所述第二存储单元发送, 而不在所述数据迁移日志区域中存储所述删除处理命令所针对的删除目标数据的地址,

在所述数据迁移日志区域中未存储有所述删除处理命令所针对的删除目标数据的地址时, 将所述删除处理命令所针对的删除目标数据的地址存储于所述数据迁移日志区域中。

24. 根据权利要求22所述的信息处理设备, 其中,

所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较, 并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时, 确定所述第一存储单元的存储状态为所述受保护状态。

25. 根据权利要求22所述的信息处理设备, 其中,

所述控制单元,

在所述第一处理中执行所述可靠性信息与阈值的比较, 并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时, 确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态确定为所述低可靠性状态时, 识别所述第二存储单元处于存储在所述第一存储单元中的数据的迁移目的地状态, 并且用于

将所述第一存储单元的存储状态更改为所述受保护状态。

26. 根据权利要求22所述的信息处理设备, 其中,

所述信息处理设备可连接到一个或多个第三存储单元, 所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

所述控制单元,

管理所述第一和第二存储单元中的存储的数据的地址, 以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时, 将所述第一存储单元和所述第二存储单元识别为同一逻辑单元; 以及

在将所述第一存储单元的存储状态识别为受保护状态时, 控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

27. 根据权利要求22所述的信息处理设备, 其中,

所述控制单元,

在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数据, 并且用于在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时, 确定结束所述第一存储单元的受保护状态。

28. 根据权利要求22所述的信息处理设备, 其中,

所述控制单元,

在所述第三处理中, 在将第一地址记录在所述数据迁移日志区域中, 当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时, 将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中并且使所述第二地址无效。

29. 一种信息处理设备的控制方法, 所述信息处理设备能够连接第一存储单元和第二存储单元, 所述第一存储单元包括可读取和可写入的第一存储器, 所述第二存储单元存储有数据迁移日志区域, 包括可读取和可写入的第二存储器,

所述信息处理设备的控制方法包括:

第一处理, 其用于基于从所述第一存储单元获得的可靠性信息, 确定所述第一存储单元的存储状态;

第二处理, 其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时, 仅对所述第一和第二存储单元中的所述第二存储单元执行写入, 在读取目标地址记录在所述数据迁移日志区域中时, 执行从所述第二存储单元读取数据, 在所述读取目标地址未记录在所述数据迁移日志区域中时, 执行从所述第一存储单元读取数据;

第三处理, 其用于根据由所述第二处理执行的写入, 将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中; 以及

第四处理, 其用于在所述第一处理未将所述第一存储单元的存储状态识别为受保护状态时, 执行从所述第一存储单元读取数据而不从所述数据迁移日志区域读取。

30. 根据权利要求29所述的控制方法, 其中, 在所述第一处理中执行所述可靠性信息与阈值的比较, 并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时, 确定所述第一存储单元的存储状态为所述受保护状态。

31. 根据权利要求29所述的控制方法, 其中, 在所述第一处理中执行所述可靠性信息与阈值的比较, 并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时, 确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态识别为所述低可靠性状态时, 识别所述第二存储单元处于存储在所述第一存储单元中的数据的数据迁移目的地状态, 并且用于将所述第一存储单元的存储状态更改为所述受保护状态。

32. 根据权利要求29所述的控制方法, 其中, 所述信息处理设备可连接到一个或多个第三存储单元, 所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

在所述控制方法中,

管理所述第一和第二存储单元中的存储的数据的地址, 以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时, 将所述第一存储单元和所述第二存储单元识别为同一逻辑单元; 以及

在将所述第一存储单元的存储状态识别为受保护状态时, 控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

33. 根据权利要求29所述的控制方法, 其中, 在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数据, 并且在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时, 确定结束所述第一存储单元的受保护状态。

34. 根据权利要求29所述的控制方法, 其中,

在所述第三处理中,

在将第一地址记录在所述数据迁移日志区域中时, 当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时, 将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中, 并且使所述第二地址无效。

35. 根据权利要求33所述的控制方法, 其中, 在所述第一存储单元的受保护状态结束时, 删除所述数据迁移日志区域。

36. 一种信息处理设备的控制方法, 所述信息处理设备能够连接第一存储单元、第二存储单元以及数据迁移日志区域, 所述第一存储单元包括可读取和可写入的作为非易失性半导体存储器的第一存储器, 所述第二存储单元存储有数据迁移日志区域, 包括可读取和可写入的作为非易失性半导体存储器的第二存储器,

所述控制方法包括:

第一处理, 其用于基于从所述第一存储单元获得的可靠性信息, 确定所述第一存储单元的存储状态;

第二处理, 其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时, 仅对所述第一和第二存储单元中的所述第二存储单元执行写入, 在读取目标地址记录在所述数据迁移日志区域中时, 执行从所述第二存储单元读取数据, 在读取目标地址未记录在所述数据迁移日志区域中时, 执行从所述第一存储单元读取数据;

第三处理, 其用于根据由所述第二处理执行的写入, 将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中; 以及

第四处理,其用于在所述第一处理未将所述第一存储单元的存储状态识别为受保护状态时,执行从所述第一存储单元读取数据而不从所述数据迁移日志区域读取。

37.根据权利要求36所述的控制方法,其中,在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为所述受保护状态。

38.根据权利要求36所述的控制方法,其中,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态识别为所述低可靠性状态时,识别所述第二存储单元处于存储在所述第一存储单元中的数据的数据迁移目的地状态,并且将所述第一存储单元的存储状态更改为所述受保护状态。

39.根据权利要求36所述的控制方法,其中,所述信息处理设备可连接到一个或多个第三存储单元,所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

在所述控制方法中,

管理所述第一和第二存储单元中的存储的数据的地址,以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时,将所述第一存储单元和所述第二存储单元识别为同一逻辑单元;以及

在将所述第一存储单元的存储状态识别为受保护状态时,控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

40.根据权利要求36所述的控制方法,其中,在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数据,并且在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时,确定结束所述第一存储单元的受保护状态。

41.根据权利要求36所述的控制方法,其中,在所述第三处理中,在将第一地址记录在所述数据迁移日志区域中,当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时,将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中并且使所述第二地址无效。

42.根据权利要求40所述的控制方法,其中,在所述第一存储单元的受保护状态结束时,删除所述数据迁移日志区域。

43.一种信息处理设备的控制方法,所述信息处理设备能够连接第一存储单元和第二存储单元,所述第一存储单元包括可读取和可写入的第一存储器,所述第二存储单元包括可读取和可写入的第二存储器,

所述控制方法包括:

第一处理,其用于基于从所述第一存储单元获得的可靠性信息,确定所述第一存储单元的存储状态;

第二处理,其用于在所述第一处理将所述第一存储单元的存储状态确定为正常状态时,仅对所述第一存储单元执行写入,在所述第一处理将所述第一存储单元的存储状态识别为是受保护状态时,仅对所述第二存储单元执行写入;以及

第三处理,其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态后发生了针对所述第一存储单元的读取时,从所述第一存储单元读取数据,并将读取的所述数据写入所述第二存储单元。

44.根据权利要求43所述的控制方法,其中,在所述第三处理中,并行地执行从所述第一存储单元读取数据的处理和将读取的所述数据写入所述第二存储单元中的处理。

45.根据权利要求43所述的控制方法,其中,

所述信息处理设备连接到存储元数据的元数据存储单元,

在所述控制方法中,

在所述第一处理确定所述第一存储单元的存储状态为所述正常状态时,在所述元数据中记录写入所述第一存储单元中的数据的地址和所述第一存储单元的识别信息的映射,并且在所述第一处理确定所述第一存储单元的存储状态为所述受保护状态时,在所述元数据中记录写入所述第二存储单元中的数据的地址和所述第二存储单元的识别信息的映射,

在读取处理中,在读取目标地址和所述第一存储单元的识别信息的映射被记录在所述元数据中时,从所述第一存储单元读取数据,并且在读取目标地址和所述第二存储单元的映射被记录在所述元数据中时,从所述第二存储单元读取数据。

46.根据权利要求43所述的控制方法,其中,

所述信息处理设备与数据迁移日志区域连接,

在所述第三处理中,

在所述第一处理将所述第一存储单元的存储状态确定为受保护状态时,将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中,

在读取处理中,在所述数据迁移日志区域中存储有读取目标地址时,从所述第二存储单元读取数据,在所述数据迁移日志区域中未存储有读取目标地址时,从所述第一存储单元读取数据。

47.根据权利要求43所述的控制方法,其中,在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为所述受保护状态。

48.根据权利要求43所述的控制方法,其中,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态确定为所述低可靠性状态时,识别所述第二存储单元处于存储在所述第一存储单元中的数据的数据迁移目的地状态,并且将所述第一存储单元的存储状态更改为所述受保护状态。

49.根据权利要求43所述的控制方法,其中,

所述信息处理设备可连接到一个或多个第三存储单元,所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

在所述控制方法中,

管理所述第一和第二存储单元中的存储的数据的地址,以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时,将所述第一存储单元和所述第二存储单元识别为同一逻辑单元;以及

在将所述第一存储单元的存储状态识别为受保护状态时,控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

50.一种信息处理设备的控制方法,所述信息处理设备能够连接第一存储单元、第二存储单元以及数据迁移日志区域,所述第一存储单元包括可读取和可写入的第一存储器,所述第二存储单元包括可读取和可写入的第二存储器,

所述控制方法包括:

第一处理,其用于基于从所述第一存储单元获得的可靠性信息,确定所述第一存储单元的存储状态;

第二处理,其用于在所述第一处理将所述第一存储单元的存储状态识别为受保护状态时,仅对所述第一和第二存储单元中的所述第二存储单元执行写入,在读取目标地址记录在所述数据迁移日志区域中时,执行从所述第二存储单元读取数据,在读取目标地址未记录在所述数据迁移日志区域中时,执行从所述第一存储单元读取数据;

第三处理,其用于根据由所述第二处理执行的写入,将写入所述第二存储单元的数据的地址记录在所述数据迁移日志区域中;以及

第四处理,其用于在所述第一处理确定所述第一存储单元的存储状态为受保护状态之后,在接收到针对存储在所述第一存储单元中的数据的删除处理命令时,将所述删除处理命令所针对的删除目标数据的地址记录在所述数据迁移日志区域中。

51.根据权利要求50所述的控制方法,其中,

在所述第四处理中,

在所述第一处理确定所述第一存储单元的存储状态为受保护状态之后,接收到存储在所述第一或第二存储单元中的数据的删除处理命令时,在所述数据迁移日志区域中存储有所述删除处理命令所针对的删除目标数据的地址时,将接收到的所述删除处理命令向所述第二存储单元发送,而不在所述数据迁移日志区域中存储所述删除处理命令所针对的删除目标数据的地址,

在所述数据迁移日志区域中未存储有所述删除处理命令所针对的删除目标数据的地址时,将所述删除处理命令所针对的删除目标数据的地址存储于所述数据迁移日志区域中。

52.根据权利要求50所述的控制方法,其中,在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为所述受保护状态。

53.根据权利要求50所述的控制方法,其中,

在所述第一处理中执行所述可靠性信息与阈值的比较,并且在基于所述比较确定满足所述第一存储单元的寿命结束条件时,确定所述第一存储单元的存储状态为低可靠性状态,

在所述第一处理将所述第一存储单元的存储状态确定为所述低可靠性状态时,识别所述第二存储单元处于存储在所述第一存储单元中的数据的数据迁移目的地状态,并且将所述第一存储单元的存储状态更改为所述受保护状态。

54. 根据权利要求50所述的控制方法, 其中,

所述信息处理设备可连接到一个或多个第三存储单元, 所述一个或多个第三存储单元包括可读取和可写入的第三存储器,

所述控制方法包括:

管理所述第一和第二存储单元中的存储的数据的地址, 以便具有相同地址的数据存在于所述第一和第二存储单元中的一个内;

在将所述第一存储单元的存储状态识别为受保护状态时, 将所述第一存储单元和所述第二存储单元识别为同一逻辑单元; 以及

在将所述第一存储单元的存储状态识别为受保护状态时, 控制所述一个或多个第三存储单元和所述逻辑单元以便构成RAID。

55. 根据权利要求50所述的控制方法, 其中, 在写入所述第一存储单元中的数据中选择作为向所述第二存储单元迁移的目标的数据, 并且在与所述迁移的目标的数据对应的所有地址被存储在所述数据迁移日志区域中时, 确定结束所述第一存储单元的受保护状态。

56. 根据权利要求50所述的控制方法, 其中, 在所述第三处理中, 在将第一地址记录在所述数据迁移日志区域中, 当与所述第一地址重叠或与所述第一地址连续的第二地址已经存储在所述数据迁移日志区域中时, 将通过组合所述第一地址和所述第二地址而获得的第三地址记录在所述数据迁移日志区域中并且使所述第二地址无效。

信息处理系统、控制程序以及信息处理设备

[0001] 相关申请的交叉引用

[0002] 本申请基于要求2013年3月18日提交的第2013-055548号日本专利申请的优先权，此日本专利申请的完整内容在此全部引入作为参考。

技术领域

[0003] 在此描述的实施例涉及用于将第一存储单元的数据迁移到第二存储单元的信息处理系统、控制程序以及信息处理设备。

背景技术

[0004] 当使用闪存配置用于个人计算机或服务器的辅助存储设备时，在某种情况下，由于例如经常出现错误，存在不能用作存储区域的有缺陷块或不能读取的有缺陷区域。当有缺陷块的数量或有缺陷区域的数量超过上限值时，因为不能登记新的有缺陷块或有缺陷区域，所以不能保证请求写入的数据在闪存中的写入。因此，当有缺陷块的数量或有缺陷区域的数量超过上限值时，即使闪存中存在空闲区域，也会突然禁用数据的写入。

[0005] 因此，存在这样一种方法：其监视存储设备的劣化，并且在存储设备的寿命结束之前限制将来自信息处理设备的数据写入到存储设备。使用此方法，可在另一个存储设备中备份用户数据，并且在存储设备的寿命结束之前迁移用户数据，防止由于存储设备的寿命结束而导致的数据丢失。

[0006] 但是，此方法是复杂的，因为必须执行用于在另一个存储设备中备份用户数据的工作。此外，因为原始存储设备中用于备份的数据写入受信息处理设备限制，所以很可能的是：载入信息处理设备中的各种应用程序的操作受限制并且应用程序的处理速度下降。

[0007] 存在如下面列出的关于相关技术的文献。

[0008] 专利文献1第2012/0284453号美国专利申请公开

[0009] 专利文献2第2012/0246388号美国专利申请公开

[0010] 专利文献3第2009/0222617号美国专利申请公开

[0011] 专利文献4第8,230,164号美国专利

[0012] 专利文献5第7,861,122号美国专利

[0013] 专利文献6第7,840,836号美国专利

[0014] 专利文献7第2007/0214187号美国专利申请公开

[0015] 专利文献8第7,606,946号美国专利

[0016] 专利文献9第6,529,992号美国专利

[0017] 专利文献10第7,222,259号美国专利

[0018] 专利文献11第7,424,639号美国专利

[0019] 专利文献12第7,739,544号美国专利

[0020] 专利文献13第7,525,749号美国专利

[0021] 专利文献14第8,037,380号美国专利

- [0022] 专利文献15第2010/0325199号美国专利申请公开
- [0023] 专利文献16第2011/0239083号美国专利申请公开
- [0024] 专利文献17第5052376号日本专利
- [0025] 专利文献18第2010-225021号日本专利申请特许公开
- [0026] 专利文献19第3,565,687号日本专利
- [0027] 专利文献20第8,176,367号美国专利
- [0028] 专利文献21第2011/0197045号美国专利申请公开
- [0029] 专利文献22第4,643,456号日本专利
- [0030] 专利文献23第4,764,490号日本专利
- [0031] 专利文献24第2011-209879号日本专利申请特许公开
- [0032] 专利文献25W02013/027642
- [0033] 非专利文献1信息技术ATA/ATAPI命令集-3 (ACS-3), d2161r4-ATAATAPI_Command_Set_-_3.pdf, 2012年9月4日, <http://www.t13.org/>
- [0034] 非专利文献2快速NVM修订1.1, 2012年10月11日, NVM_Express_1_1.pdf, <http://www.nvmexpress.org/>
- [0035] 非专利文献3串行ATA国际组织: 串行ATA修订3.1金修订, 2011年7月18日, <http://www.serialata.org/>
- [0036] 非专利文献4快速PCI基本规范修订3.0, 2010年11月10日, <http://www.pcisig.com/>
- [0037] 非专利文献5串行ATA技术提案: SATA31_TPR_C108标题: 设备睡眠, <http://www.serialata.org/>
- [0038] 非专利文献6SCSI初级命令-4 (SPC-4), INCITS T10/1731-D, 修订36e, <http://www.t10.org/>
- [0039] 非专利文献7Masaru Kitsuregawa, “Yokuwaku Storage Networking”, 第一版本, Ohm, Ltd., 2011年5月20日, 第6至9页以及第67至93页

附图说明

- [0040] 图1是第一实施例中的信息处理系统的功能配置实例的框图;
- [0041] 图2是在将控制程序存储在SSD中的信息处理系统的功能配置实例的框图;
- [0042] 图3是在将控制程序存储在另一个外部存储设备中的信息处理系统的功能配置实例的框图;
- [0043] 图4是在从WEB安装控制程序中的信息处理系统的功能配置实例的框图;
- [0044] 图5是在从光盘驱动器安装控制程序中的信息处理系统的功能配置实例的框图;
- [0045] 图6是在从USB存储器安装控制程序中的信息处理系统的功能配置实例的框图;
- [0046] 图7是信息处理设备的分层功能配置实例的框图;
- [0047] 图8是信息处理系统的外部配置的图;
- [0048] 图9是信息处理系统的另一种外部配置的图;
- [0049] 图10是NAND存储芯片的功能配置实例的框图;
- [0050] 图11是包括在NAND存储芯片中的一个平面的配置实例的电路图;

- [0051] 图12是四级数据存储方法中的阈值电压分布的图；
- [0052] 图13是SSD的功能配置实例的框图；
- [0053] 图14是SSD的管理信息的图；
- [0054] 图15是SSD的LBA与管理单元之间的关系的图；
- [0055] 图16是用于解释用于从LBA指定物理地址的程序的流程图；
- [0056] 图17是用于解释从SSD进行的读取操作的实例的流程图；
- [0057] 图18是用于解释从SSD进行的读取操作的实例的流程图；
- [0058] 图19是用于解释从SSD进行的读取操作的实例的流程图；
- [0059] 图20A和图20B是用于解释从SSD进行的读取操作的实例的流程图；
- [0060] 图21是用于解释在执行NAND GC中的操作程序的流程图；
- [0061] 图22是用于解释在接收删除通知时执行的SSD的操作实例的流程图；
- [0062] 图23是用于解释在出现错误时执行的SSD的操作实例的流程图；
- [0063] 图24是用于解释统计信息获取处理的流程图；
- [0064] 图25是坏逻辑扇区表的图；
- [0065] 图26是坏集群表的图；
- [0066] 图27A和图27B是用作统计信息的实例的SMART信息的图；
- [0067] 图28是统计信息的原始值与SSD的退件率(reject ratio)之间的关系的图；
- [0068] 图29是用于解释错误命令响应处理过程的流程图；
- [0069] 图30是用于解释另一个错误命令响应处理过程的流程图；
- [0070] 图31是在执行寿命结束处理之前由信息处理设备管理的数据的配置实例的图；
- [0071] 图32是引导加载器(boot loader)的内部数据的图；
- [0072] 图33是元数据的配置实例的图；
- [0073] 图34是在应用程序将对逻辑驱动器的存取请求传输到OS时执行的程序的流程图；
- [0074] 图35是用于解释例如在信息处理设备启动时执行的控制程序的处理程序的流程图；
- [0075] 图36是用于解释存储单元的寿命周期的状态转变图；
- [0076] 图37是用于解释存储单元的寿命周期的状态转变图；
- [0077] 图38是在其中数据迁移源存储单元和数据迁移目的地存储单元连接的状态下由信息处理设备管理的数据的配置实例的图；
- [0078] 图39是存储单元的寿命周期的状态转变图；
- [0079] 图40是用于解释由控制程序在寿命结束处理中执行的处理的流程图；
- [0080] 图41是逻辑驱动器状态表的图；
- [0081] 图42是用于解释在将写入请求从应用程序传输到OS时执行的OS的处理程序的流程图；
- [0082] 图43是用于解释在将删除请求从应用程序传输到OS时执行的OS的处理程序的流程图；
- [0083] 图44是用于解释在将读取请求从应用程序传输到OS时执行的OS的处理程序的流程图；
- [0084] 图45是用于解释后台备份的处理程序的流程图；

- [0085] 图46是用于解释在完成数据迁移时执行的控制程序的操作程序的流程图；
- [0086] 图47是在数据迁移状态下从逻辑驱动器进行的读取的概念图；
- [0087] 图48是在数据迁移状态下在逻辑驱动器中进行的写入的概念图；
- [0088] 图49是第二实施例中的信息处理系统的功能配置实例的框图；
- [0089] 图50是数据迁移日志的图；
- [0090] 图51是数据迁移日志中的日志的写入实例的图；
- [0091] 图52是用于解释寿命结束处理的流程图；
- [0092] 图53A和图53B是用于解释从逻辑驱动器进行的读取的操作程序的流程图；
- [0093] 图54是用于解释在将LBA数据删除请求从应用程序传输到OS时执行的OS的处理程序的流程图；
- [0094] 图55是用于解释在逻辑驱动器中进行的写入的操作程序的流程图；
- [0095] 图56是用于解释使用数据迁移日志执行的数据迁移状态的监视程序的流程图；
- [0096] 图57是在存储单元达到寿命结束时发生的逻辑驱动器的状态转变的图；
- [0097] 图58是第三实施例中的信息处理系统的功能配置实例的框图；
- [0098] 图59是用于解释阵列LBA (ALBA) 和存储单元LBA (SLBA) 在正常状态下的变换方法的图；
- [0099] 图60是用于解释紧接在迁移状态转变之后的阵列LBA (ALBA) 和存储单元LBA (SLBA) 的变换方法的图；
- [0100] 图61是数据迁移日志的图；
- [0101] 图62是用于解释寿命结束处理的流程图；
- [0102] 图63是用于解释例如在RAID控制器启动时控制单元执行的处理程序的流程图；
- [0103] 图64是用于解释RAID控制器在执行读取请求中的处理程序的流程图；
- [0104] 图65是用于解释RAID控制器在执行读取请求中的另一个处理程序的流程图；
- [0105] 图66是用于解释RAID控制器在执行写入请求中的处理程序的流程图；
- [0106] 图67是写入处理的特定实例的图；
- [0107] 图68是用于解释使用数据迁移日志执行的数据迁移状态的监视程序的流程图；
- [0108] 图69是存储单元的状态转变的图；
- [0109] 图70是其中一个存储单元的数据迁移状态和另一个存储单元的故障同时出现的实例的图；
- [0110] 图71是其中两个存储单元的数据迁移状态和另一个存储单元的故障同时出现的实例的图；
- [0111] 图72是其中在一个存储单元的数据迁移状态下另一个存储单元中出现不可纠正的ECC错误的实例的图；
- [0112] 图73是第四实施例中的信息处理系统的功能配置实例的框图；
- [0113] 图74是第四实施例中的信息处理系统的另一个功能配置实例的框图；
- [0114] 图75是状态表的图；
- [0115] 图76是第五实施例中的信息处理系统的功能配置实例的框图；
- [0116] 图77是第六实施例中的信息处理系统的功能配置实例的框图；
- [0117] 图78是第六实施例中的信息处理系统的另一个功能配置实例的框图；

- [0118] 图79是用于解释控制单元从正常状态到低可靠性状态的转变操作的流程图；
- [0119] 图80是用于解释在控制单元从CPU接收用于存储单元的存储单元信息请求时执行的处理程序的实例的流程图；
- [0120] 图81是用于解释在控制单元从CPU接收用于存储单元的存储单元信息请求时执行的另一个处理程序的实例的流程图；
- [0121] 图82是信息处理设备中的驱动器显示屏幕实例的图；
- [0122] 图83是信息处理设备中的另一个驱动器显示屏幕实例的图；
- [0123] 图84是信息处理设备中的另一个显示屏幕实例的图；以及
- [0124] 图85是第六实施例中的信息处理系统的另一个功能配置实例的框图。

具体实施方式

[0125] 根据一个实施例，一种信息处理系统包括：第一存储单元，其包括可读取和可写入的第一存储器；第二存储单元，其包括可读取和可写入的第二存储器；以及信息处理设备，其可连接到所述第一存储单元和所述第二存储单元。所述信息处理设备包括控制单元，所述控制单元被配置为：基于从所述第一存储单元获得的可靠性信息，执行用于确定所述第一存储单元的存储状态的第一处理，以及在所述第一处理将所述第一存储单元的所述存储状态识别为受保护状态时，用于对所述第二存储单元执行写入并且对所述第一和第二存储单元执行读取的第二处理，以及用于根据由所述第二处理执行的写入，更新所述第一和第二存储单元中的存储的数据的地址的第三处理。

[0126] 下面参考附图详细地解释根据实施例的信息处理系统、控制程序以及信息处理设备。本发明不受这些实施例限制。

[0127] (第一实施例)

[0128] 系统的配置

[0129] 图1中示出第一实施例中的计算机系统的配置，该计算机系统是信息处理系统的实例。信息处理系统1包括信息处理设备111、一个或多个存储单元，以及被配置为连接信息处理设备111和存储单元的接口19。在此实施例中，将属于非易失性存储设备的SSD(固态驱动器)用作存储单元。但是，存储单元可以是其它存储设备，例如硬盘驱动器(HDD)、混合驱动器、SD卡、USB存储器、NAND闪存芯片、磁带、嵌入式多媒体卡(eMMC)，以及第2012/0117354号美国专利申请公开中描述的存储设备和存储节点。包括在一个存储设备中的多个存储区域可以分别用作单独的存储单元。在此实施例中的以下解释中，将SATA(串行高级技术附件)接口用作接口19。但是，可以使用快速PCI(快速外围组件互连，PCIe)、USB(通用串行总线)、SAS(串行连接SCSI)、雷电(注册商标)、以太网(注册商标)、光纤通道等。CPU(控制电路)5是信息处理设备111中的中央处理单元。信息处理设备111中的各种计算和控制都由CPU 5执行。CPU 5和芯片组7由诸如DMI(直接媒体接口)之类的接口连接。CPU 5经由芯片组7执行对存储单元2和光盘驱动器10(例如DVD驱动器)的控制。CPU 5执行对主存储器6的控制。作为主存储器6，例如可以采用DRAM(动态随机存取存储器)、MRAM(磁阻随机存取存储器)、ReRAM(电阻随机存取存储器)和FeRAM(铁电随机存取存储器)。

[0130] 用户通过诸如键盘14或鼠标15之类的输入设备，执行对信息处理设备111的控制。CPU 5例如经由USB(通用串行总线)控制器13和芯片组7，处理来自键盘14或鼠标15的信号。

CPU 5经由显示控制器8,将图像数据和文本数据发送到显示器(显示设备)9。用户可以经由显示器9,在视觉上识别从信息处理设备111发送的图像数据和文本数据。

[0131] CPU 5是用于控制信息处理设备111的操作的处理器。CPU 5例如执行从存储单元2加载到主存储器6的操作系统(OS) 100。此外,当光盘驱动器10能够对插入的光盘执行读取处理和写入处理中的至少一个时,CPU 5执行该处理。CPU 5执行存储在ROM 11中的UEFI(统一可扩展固件接口)固件、系统BIOS(基本输入/输出系统)等。UEFI固件和系统BIOS是用于信息处理设备111中的硬件控制的程序。此外,CPU 5经由芯片组7控制网络控制器12。作为网络控制器12,例如存在LAN(局域网)控制器和无线LAN控制器。

[0132] 主存储器6临时存储程序和数据,并且充当CPU 5的工作存储器。主存储器6包括用于存储OS 100的存储区域6A以及用于存储控制程序200的存储区域6B。如通常例如被称为Windows(注册商标)那样,OS 100是用于管理整个信息处理设备111的程序,例如管理信息处理设备111的输入和输出设备,管理存储单元和存储器,以及执行控制以便使软件能够使用信息处理设备111的硬件和存储设备。在此实施例中解释的实例中,控制程序200是OS 100的一部分。但是,控制程序200和OS 100可以是独立的程序。在此实施例中,控制程序200被解释为经由接口19读取到主存储器6并且执行的程序。但是,控制程序200可以存储在ROM 11中,并且作为UEFI固件或系统BIOS的一部分而操作。控制程序200可以作为硬件在信息处理设备111中实现。在主存储器6中,存储元数据300和逻辑驱动器状态表450。

[0133] 显示控制器8是视频再现控制器,其被配置为控制信息处理设备111的显示器9。芯片组7是连接到CPU 5的局域总线的网桥设备。芯片组7控制存储单元2和存储单元3,这些存储单元是被配置为经由接口19存储各种软件 and 数据的存储设备。存储单元2和存储单元3能够通过芯片组7连接到CPU 5,或者可以直接连接到CPU 5。

[0134] 信息处理设备111以逻辑扇区为单位来存取存储单元2和存储单元3。经由接口19,将写入命令(写入请求)、读取命令(读取请求)、闪存命令等输入到存储单元2和存储单元3。

[0135] 芯片组7还具有用于使以下各项经受存取控制的功能:ROM 11、光盘驱动器10、网络控制器12和USB控制器13。键盘14和鼠标15连接到USB控制器13。

[0136] 控制程序的形成

[0137] 在此实施例中,例如如图2中所示,当信息处理设备111断开时,将控制程序200存储在存储单元2的NAND闪存(NAND存储器) 16的区域16B中。如图2中所示,当信息处理设备111启动或程序启动时,将控制程序200从NAND存储器16的区域16B加载到主存储器6上的区域6B。另一方面,当多个存储单元连接到信息处理设备111时,例如当独立于存储单元2和存储单元3的存储单元20连接到信息处理设备111时,如图3中所示,能够将控制程序200存储在存储单元20的区域20B中。然后,当信息处理设备111启动或程序启动时,将控制程序200从区域20B加载到主存储器6上的区域6B。具体地说,当存储单元20用作被配置为存储OS的系统驱动器,并且存储单元2用作被配置为存储用户数据(例如文档、静态图像数据和移动图像数据)的数据驱动器时,希望将控制程序200存储在存储单元20中,就清楚地区分存储单元2和存储单元20的作用而言,存储单元20充当系统驱动器。例如,将充当系统驱动器的存储单元20用作被配置为主要存储OS和应用程序的存储驱动器,并且将充当数据驱动器的存储单元2用作被配置为存储用户数据的存储驱动器。

[0138] 就节省用户设置控制程序200的劳力而言,例如如图2和图3中所示,希望在其中控

制程序200存储在存储单元2或存储单元20中的状态下,从制造商那里运送信息处理系统1,在商店中展示信息处理系统1,并且将信息处理系统1交付给用户。另一方面,就使用户能够选择控制程序200的安装适当性而言,并且就能够将最新控制程序提供给用户而言,希望能够通过从WEB进行的下载或者从外部存储介质(例如DVD-ROM或USB存储器)进行安装,将控制程序200存储在存储单元2或存储单元20中。

[0139] 图4是从WEB进行的下载的实例。控制程序200存储在WEB服务器21中的存储介质22的区域22B中。在存储单元的NAND存储器16上的区域16B中,例如经由诸如因特网、局域网或无线LAN之类的网络,通过网络控制器12下载(或安装)控制程序200。作为存储介质22,例如使用SSD、HDD、混合驱动器或磁带。在图3的情况下,将控制程序200下载到或安装在存储单元20上的区域20B中。

[0140] 图5是从光学介质进行的安装的实例。控制程序200存储在诸如DVD-ROM、CD-ROM、DVD-RW或蓝光(注册商标)光盘之类的光学介质23中。当在光盘驱动器10上设置光学介质23时,经由光盘驱动器10将控制程序200安装在存储单元2的NAND存储器16上的区域16B(或区域20B)中。

[0141] 图6是从USB存储器进行的安装的实例。控制程序200存储在USB存储器24的区域24B中。当USB存储器24连接到USB控制器13时,经由USB控制器13将控制程序200安装在存储单元2的NAND存储器16上的区域16B(或区域20B)中。自然地,可以使用诸如SD卡之类的其它外部存储器代替USB存储器24。就用户获取的容易性而言,希望在运送信息处理系统1或存储单元2时,将光学介质23和USB存储器24作为附件与信息处理系统1或存储单元2包装在一起。另一方面,光学介质23和USB存储器24可以作为软件商品独立销售,或者可以作为杂志或书籍的补充而附带。

[0142] 软件的分层结构

[0143] 图7中示出信息处理设备111的软件级别的分层结构。通常,加载到主存储器6上的应用程序400并不直接与诸如存储单元2和存储单元3之类的存储单元通信,而通过加载到主存储器6的OS 100与存储单元通信。OS 100通过UEFI固件或系统BIOS与存储单元通信。在下面解释的寿命结束处理(存储单元的寿命结束时的处理或达到寿命的处理)之前,OS 100将存储单元2识别为充当逻辑单元的逻辑驱动器4(参见图38),并且向应用程序400通知作为可存取存储驱动器的逻辑驱动器4。当应用程序400需要将诸如读取请求或写入请求之类的命令传输到逻辑驱动器4时,应用程序400将用于存取文件单元中的逻辑驱动器的请求传输到OS 100。OS 100参考存储在主存储器6中的元数据300、指定对应于请求存取的文件存储单元2的逻辑地址(LBA:逻辑块地址),并且经由接口19将命令、LBA和数据传输到存储单元2。在接收来自存储单元2的响应时,OS 100将响应传输到应用程序400。当逻辑驱动器4更改为下面解释的迁移状态时,OS 100将存储单元2和存储单元3识别为逻辑驱动器4。

[0144] 信息处理设备的配置

[0145] 解释信息处理系统1的配置实例。图8是充当信息处理系统1的台式计算机的示意图。台式计算机包括计算机主体31、显示器9、键盘14和鼠标15。计算机主体31包括主板30,其装载有主要硬件、存储单元2和电源设备32。存储单元2经由SATA电缆物理连接到主板30,并且经由在主板30上装载的芯片组7电连接到也在主板上装载的CPU 5。电源设备32产生用于台式计算机中的各种电力,并且经由电源电缆将电力提供给主板30、存储单元2等。存储

单元3可以经由SATA电缆物理连接到主板30。因此,存储单元3经由在主板30上装载的芯片组7电连接到也在主板上装载的CPU 5。

[0146] 图9是充当信息处理系统1的便携式计算机的示意图。便携式计算机包括计算机主体34和显示单元35。在显示单元35中,结合了例如通过LCD(液晶显示器)配置的显示设备9。显示单元35附接到计算机主体34,以便能够在计算机主体34的上表面被暴露的打开位置和计算机主体34的上表面被覆盖的闭合位置之间枢转。计算机主体34包括薄盒状外壳。电源开关36、键盘14、触摸板33等布置在计算机主体34的上表面上。如在台式计算机中,计算机主体34包括存储单元2、主板和电源设备。

[0147] 除了上述情形之外,随本发明应用的信息系统处理1还可以是诸如静物摄像机或视频摄像机之类的图像拾取装置,或者可以是平板计算机、智能电话、游戏装置、汽车导航系统、打印机装置、扫描仪装置、服务器系统等。

[0148] 存储设备(1)的配置

[0149] 解释存储单元2的配置。在此实施例中的解释中,存储单元2的主要组件是NAND存储器16。在此实施例中的解释中,存储单元2和存储单元3具有相同的配置。另一方面,当存储单元3具有的配置不同于存储单元2的配置时,也可以应用本发明。例如,存储单元2是包括作为主要组件的NAND存储器16的SSD,并且存储单元3是包括作为主要组件的磁盘的硬盘驱动器。希望存储单元2和存储单元3在其中存储统计信息65。但是,当存储单元2在其中存储统计信息65并且存储单元3未在其中存储统计信息65时,也可以应用本发明。希望作为数据迁移目的地的存储单元3的存储容量等于或大于作为数据迁移源的存储单元2的存储容量。但是,当存储单元3的存储容量小于存储单元2的存储容量时,也可以应用本发明。

[0150] NAND存储器的配置

[0151] 在图10中,示出在此实施例中用作存储单元2和存储单元3的组件的NAND存储器16中包括的NAND存储芯片80的内部配置实例。NAND存储器16包括一个或多个NAND存储芯片80。

[0152] NAND存储芯片80包括存储单元阵列,其中以矩阵形状排列多个存储单元。存储单元阵列中包括的存储单元晶体管通过MOSFET(金属氧化物半导体场效应晶体管)配置,此MOSFET具有在半导体衬底上形成的层叠式栅极结构。此层叠式栅极结构包括经由栅极绝缘膜在半导体衬底上形成的电荷存储层(浮置栅极电极),以及经由栅极间绝缘膜在浮置栅极电极上形成的控制栅极电极。存储单元晶体管的阈值电压根据在浮置栅极电极中累积的电子数量而更改。存储单元晶体管根据阈值电压之差存储数据。

[0153] 在此实施例中的解释中,相应存储单元采用用于使用上部页面和下部页面写入数据的2位/单元的四级存储方法的写入系统。但是,当存储单元采用用于使用单个页面写入数据的1位/单元的二级存储方法的写入系统、用于使用上部页面、中间页面和下部页面写入数据的3位/单元的八级存储方法的写入系统,或者4位/单元或4位/单元以上的多级存储方法的写入系统的写入系统时,本发明的本质相同。存储单元晶体管并不限于包括浮置栅极电极的结构,并且能够是诸如MONOS(金属-氧化物-氮化物-氧化物-硅)类型的结构,MONOS类型的结构能够通过充当电荷存储层的氮化物界面上捕获电子而调整阈值电压。同样,MONOS类型的存储单元晶体管可以被配置为存储1位,或者可以被配置为存储多级。作为非易失性存储介质,存储单元晶体管可以是半导体存储介质,其中存储单元进行三维布

置,如在第8,189,391号美国专利、第2010/0207195号美国专利申请公开和第2010/0254191号美国专利申请公开中所描述的那样。

[0154] 如图10中所示,NAND存储芯片80包括存储单元阵列82,其通过以矩阵形状排列存储单元以便存储数据来配置。存储单元阵列82包括多个位线、多个字线和一个公共源级线。电可重写数据存储单元以矩阵形状排列在位线和字线的交叉点处。用于控制位线的位线控制电路83和用于控制字线的字线控制电路85连接到存储单元阵列82。即,位线控制电路83经由位线读取存储单元阵列82中的存储单元的数据,并且经由位线将写入控制电压施加到存储单元阵列82中的存储单元并将数据写入存储单元中。

[0155] 列解码器84、数据输入/输出缓冲器89和数据输入/输出端子88连接到位线控制电路83。经由位线控制电路83和数据输入/输出缓冲器89,将从存储单元阵列82读取的存储单元的数据从数据输入/输出端子88输出到外部。列解码器84经由数据输入/输出缓冲器89,将从外部输入到数据输入/输出端子88的写入数据输入到位线控制电路83,并且将此写入数据写入指定的存储单元中。

[0156] 存储单元阵列82、位线控制电路83、列解码器84、数据输入/输出缓冲器89和字线控制电路85连接到控制电路86。控制电路86根据输入到控制信号输入端子87的控制信号,生成用于控制存储单元阵列82、位线控制电路83、列解码器84、数据输入/输出缓冲器89和字线控制电路85的控制信号和控制电压。NAND存储芯片80中的存储单元阵列82之外的电路部分被称为NAND控制器(NANDC)81。

[0157] 图11示出图10中所示的存储单元阵列82的配置。存储单元阵列82是NAND单元型存储单元阵列,并且被配置为包括多个NAND单元。一个NAND单元包括由串联的存储单元形成的存储串MS,以及连接到存储串MS的两端的选择栅极S1和S2。选择栅极S1连接到位线BL,并且选择栅极S2连接到源级线SRC。布置在同一行中的存储单元MC的控制栅极共同连接到字线WL0至WLm-1。第一选择栅极S1共同连接到选择线SGD,并且第二选择栅极S2共同连接到选择线SGS。

[0158] 存储单元阵列82包括一个或多个平面。所述平面包括多个块。每个块包括多个NAND单元。以块为单位来擦除数据。

[0159] 连接到一个字线的多个存储单元配置一个物理扇区。针对每个物理扇区写入和读取数据。物理扇区与下面解释的LBA的逻辑扇区无关。在一个物理扇区中,例如在2位/单元写入系统(四级)的情况下,存储相当于两个物理页面(两个页面)的数据。另一方面,在1位/单元写入系统(二级)的情况下,例如将相当于一个物理页面(一个页面)的数据存储在一个物理扇区中。在3位/单元写入系统(八级)的情况下,例如将相当于三个物理页面(三个页面)的数据存储在一个物理扇区中。

[0160] 在读取操作、程序检验操作和程序操作期间,选择一个字线,并且根据物理地址(例如从下面解释的SSDC 41接收的行地址)选择一个物理扇区。根据物理地址执行物理扇区中的页面切换。在此实施例中,NAND存储器16采用2位/单元写入系统,在假设将两个页面(即,上部页面和下部页面)分配给物理扇区作为物理页面的情况下,SSDC 41处理物理扇区。将物理地址分配给所有页面。

[0161] 配置2位/单元的四级NAND存储器,以便一个存储单元中的阈值电压可以具有四种分布。图12示出存储在四级NAND单元型闪存的存储单元中的2位四级数据(数据“11”、“01”、

“10”和“00”)之间的关系以及所述存储单元的阈值电压分布。在图12中,VA1表示在关于物理扇区读取两个数据时施加到选定字线的电压,此物理扇区的下部页面已经写入并且其上部页面尚未写入。VA1V表示在执行在A1中的写入时,为了检查写入是否完成而施加的检验电压。

[0162] VA2、VB2和VC2表示在关于物理扇区读取四个数据时施加到选定字线的电压,此物理扇区的下部页面和上部页面已经写入。VA2V、VB2V和VC2V表示在执行阈值电压分布中的写入时,为了检查写入是否完成而施加的检验电压。Vread1和Vread2表示在执行数据的读取时,施加到NAND单元中未选择的存储单元并且使未选择的存储单元传导而不管未选择的存储单元的保留数据的读取电压。进一步,Vev1和Vev2表示在擦除存储单元的数据时,为了检查擦除是否完成而施加到存储单元的擦除检验电压。Vev1和Vev2具有负值。考虑邻近存储单元的干扰的影响来确定Vev1和Vev2的量值。电压的量值关系如下:

[0163] $V_{ev1} < VA1 < VA1V < V_{read1}$

[0164] $V_{ev2} < VA2 < VA2V < VB2 < VB2V < VC2 < VC2V < V_{read2}$

[0165] 擦除检验电压Vev1、Vev2和Vev3是负值,如上面所解释。但是,在擦除检验操作中实际施加到存储单元MC的控制栅极的电压并非负值而为0或正值。即,在实际擦除检验操作中,将正电压提供给存储单元MC的背栅极,并且将具有0或小于背栅极电压的正值的电压施加到存储单元MC的控制栅极。换言之,擦除检验电压Vev1、Vev2和Vev3是相当于具有负值的电压。

[0166] 在块擦除之后的存储单元的阈值电压分布ER的上限值也是负值。将数据“11”分配给存储单元。在下部页面和上部页面写入的状态下,数据“11”、“01”、“10”和“00”的存储单元分别具有正阈值电压分布ER2、A2、B2和C2(A2、B2和C2的下限值也是正值)。数据“01”的阈值电压分布A2具有最低电压值,数据“00”的阈值电压分布C2具有最高电压值,并且各种阈值电压分布的电压值具有关系: $A2 < B2 < C2$ 。在下部页面写入并且上部页面未写入的状态下,数据“10”的存储单元具有正阈值电压分布A(A1的下限值也是正值)。图12中所示的阈值电压分布仅是实例。本发明并不限于此。例如,在参考图12的解释中,所有阈值电压分布A2、B2和C2都是正阈值电压分布。但是,当阈值电压分布A2是负电压分布并且阈值电压分布B2和C2是正电压分布时,这些分布也包括在本发明的范围内。即使阈值电压分布ER1和ER2是正值,本发明也并不限于此。在此实施例中,ER2、A2、B2和C2的数据中的对应关系是“11”、“01”、“10”和“00”。但是,对应关系可以是另一种对应关系,例如“11”、“01”、“00”和“10”。

[0167] 一个存储单元的2位数据包括下部页面数据和上部页面数据。根据单独的写入操作(即,两个写入操作),将下部页面数据和上部页面数据写入存储单元中。当将数据表示为“*@”时,*表示上部页面数据,并且@表示下部页面数据。

[0168] 首先,参考图12的第一和第二部分解释下部页面数据的写入。假设所有存储单元在擦除状态下具有阈值电压分布ER并且存储数据“11”。如图12中所示,当执行下部页面数据的写入时,根据下部页面数据的值(“1”或“0”)将存储单元的阈值电压分布ER分成两个阈值电压分布(ER1和A1)。当下部页面数据的值是“1”时, $ER1 = ER$,因为维持了在擦除状态下的阈值电压分布ER。但是,阈值电压分布ER和ER1之间的关系能够是 $ER1 > ER$ 。

[0169] 另一方面,当下部页面数据的值是“0”时,将高电场施加到存储单元的隧道氧化物膜、将电子注入浮置栅极电极中,并且使存储单元的阈值电压 V_{th} 增加预定值。具体地说,设

置检验电压 $VA1V$ 并且重复写入操作,直至阈值电压增加到等于或高于检验电压 $VA1V$ 。因此,存储单元更改为写入状态(数据“10”)。当即使将写入操作重复预定次数,存储单元也不会达到阈值电压时(或者未达到阈值电压的存储单元数量等于或大于阈值时),物理页面中的写入是“程序错误”(程序失败或写入错误)。

[0170] 参考图12的第二和第三部分解释上部页面数据的写入。基于从芯片外部输入的写入数据(上部页面数据)和已经写入存储单元中的下部页面数据,执行上部页面数据的写入。

[0171] 即,如图12的第二和第三部分中所示,当更新数据的值是“1”时,未将高电场施加到存储单元的隧道氧化物膜以便防止存储单元的阈值电压 V_{th} 的升高。因此,具有数据“11”(擦除状态的阈值电压分布 $ER1$)的存储单元维持数据“11”(ER2),并且具有数据“10”(阈值电压分布 $A1$)的存储单元维持数据“10”(B2)。但是,就保证这些分布之间的电压裕度而言,希望使用大于检验电压 $VA1V$ 的正检验电压 $VB2V$ 调整阈值电压分布的下限值,从而形成通过使阈值电压分布的宽度变窄而获得的阈值电压分布B2。当即使将下限值调整重复预定次数,存储单元也不会达到阈值电压时(或者未达到阈值电压的存储单元数量等于或大于预定值时),物理页面中的写入是“程序错误”。

[0172] 另一方面,当上部页面数据的值是“0”时,将高电场施加到存储单元的隧道氧化物膜、将电子注入浮置栅极电极中,并且使存储单元的阈值电压 V_{th} 增加预定值。具体地说,设置检验电压 $VA2V$ 和 $VC2V$ 并且重复写入操作,直至阈值电压增加到等于或高于检验电压 $VA2V$ 和 $VC2V$ 。因此,具有数据“11”(擦除状态下的阈值电压分布 $ER1$)的存储单元更改为具有阈值电压分布 $A2$ 的数据“01”,并且具有数据“10”(A1)的存储单元更改为具有阈值电压分布 $C2$ 的数据“00”。此时,使用检验电压 $VA2V$ 和 $VC2V$,并且调整阈值电压分布 $A2$ 和 $C2$ 的下限值。当即使将写入操作重复预定次数,存储单元也不会达到阈值电压时(或者未达到阈值电压的存储单元数量等于或大于阈值时),物理页面中的写入是“程序错误”。

[0173] 另一方面,在擦除操作中,设置擦除检验电压 V_{ev} 并且重复擦除操作,直至阈值电压减小到等于或低于检验电压 V_{ev} 。因此,存储单元更改为写入状态(数据“00”)。当即使将擦除操作重复预定次数,存储单元也不会达到阈值电压时(或者未达到阈值电压的存储单元数量等于或大于预定值时),物理页面的擦除是“擦除错误”(擦除失败)。

[0174] 上面解释了一般四级存储方法中的数据写入系统的实例。在3位/单元或3位/单元以上的多位存储系统中,根据更高阶页面数据,仅将用于将阈值电压分布分成八种或八种以上的操作添加到上面解释的操作。因此,基本操作相同。

[0175] 存储设备(2)的配置

[0176] 解释存储单元2和存储单元3的配置实例。在此实施例中,如图13中所示,充当SSD的存储单元2和存储单元3包括:NAND闪存(以下缩写为NAND存储器)16,其充当非易失性半导体存储器;接口控制器(IFC)42,其被配置为经由接口19执行去往和来自信息处理设备111的信号传输和接收;RAM(随机存取存储器)40,其充当半导体存储器,包括充当IFC 42与NAND存储器16之间的中间缓冲器的高速缓冲存储器(CM)46;SSD控制器(SSDC)41,其被配置为实施对NAND存储器16和RAM 40的管理和控制以及对接口控制器42的控制;以及总线43,其被配置为连接这些组件。

[0177] 作为RAM 40,例如可以采用易失性RAM,例如DRAM(动态随机存取存储器)或SRAM

(静态随机存取存储器);或者非易失性RAM,例如FeRAM(铁电随机存取存储器)、MRAM(磁阻随机存取存储器)、PRAM(相变随机存取存储器)或ReRAM(电阻随机存取存储器)。RAM 40可以包括在SSDC 41中。

[0178] NAND存储器16包括多个NAND存储芯片80。NAND存储器16存储由信息处理设备111指定的用户数据,存储用于管理用户数据的管理表,并且存储由RAM 40管理的管理信息以便备份。NAND存储器16包括存储单元阵列82,其中以矩阵形状排列多个存储单元。相应存储单元可以使用上部页面和下部页面执行多级存储。NAND存储器16包括多个存储芯片。每个存储芯片通过排列作为数据擦除单位的多个块来配置。在NAND存储器16中,针对每个页面执行数据的写入和数据的读取。块包括多个页面。

[0179] RAM 40包括充当高速缓存的高速缓冲存储器(CM) 46,其用于信息处理设备111与NAND存储器16之间的数据传输。RAM 40充当用于管理信息存储的存储器以及用于工作区域的存储器。RAM 40的区域40A中管理的管理表是存储在NAND存储器16的区域40M中并且在启动存储单元2和存储单元3时加载的各种管理表。定期地,当接收待机命令时,当接收闪存命令时,或者当中断电源时,在NAND存储器16的区域40M中备份管理数据。

[0180] SSDC 41的功能通过被配置为执行存储在NAND存储器16中的系统程序(固件)的处理器、各种硬件电路等实现。响应于来自信息处理设备111的诸如写入请求、高速缓存闪存请求和读取请求之类的各种命令,SSDC 41执行信息处理设备111与NAND存储器16之间的数据传输控制;存储在RAM 40和NAND存储器16中的各种管理表的更新和管理;从NAND存储器16读取的数据的EDC解密等。

[0181] 当信息处理设备111将读取请求或写入请求传输到存储单元2时,信息处理设备111经由接口19输入用作逻辑地址的LBA。LBA是其中将序列号提供给从0开始的逻辑扇区(大小:例如512字节)的逻辑地址。当信息处理设备111向存储单元2发出读取请求或写入请求时,信息处理设备111连同LBA一起输入发出的读取请求或写入请求所针对的逻辑扇区的大小。

[0182] IFC 42具有用于执行以下操作的功能:接收读取请求、写入请求、其它请求和来自信息处理设备111的数据,将接收的请求和数据传输到SSDC 41,并且根据SSDC 41的控制将数据传输到RAM 40。

[0183] 图14中示出用于存储单元2和存储单元3中的管理信息44的配置实例。如上面所解释,管理信息44以非易失性方式存储在NAND存储器16的区域40M中。将存储在区域40M中的管理信息加载到RAM 40的区域40A,并且在启动存储单元2时使用此管理信息。定期地或者当电源中断时,将区域40A中的管理信息44备份在区域40M中。当RAM 40是诸如MRAM或FeRAM之类的非易失性RAM时,能够将管理信息44仅存储在RAM 40中。然后,不将管理信息44存储在NAND存储器16中。为了减少NAND存储器16中的写入量,希望存储在管理信息44中的数据是通过压缩存储在RAM 40的区域40A中的数据而获得的数据。为了减少NAND存储器16中的写入频率,希望附加地写入存储在RAM 40的区域40A中的管理信息44的更新信息(差异信息)。

[0184] 如图14中所示,管理信息包括空闲块表(FBT) 60、坏块表(BT) 61、活动块表(ABT) 62、磁道表(磁道单元中的逻辑到物理变换表) 63、集群表(集群单元中的逻辑到物理变换表) 64和统计信息65。

[0185] 如图15中所示,LBA是其中将序列号提供给从0开始的逻辑扇区(大小:例如512字节)的逻辑地址。在此实施例中,作为存储单元2的逻辑地址(LBA)的管理单位,定义通过从LBA的低阶第(s+1)位起的高阶位行配置的集群地址,以及通过LBA的低阶第(s+t+1)位起的高阶位行配置的磁道地址。即,逻辑扇区是来自信息处理设备111的最小存取单位。集群是用于管理SSD中的“小数据”的管理单位。集群大小被定义为逻辑扇区大小的自然数倍。磁道是用于管理SSD中的“大数据”的管理单位。磁道大小被定义为集群大小的两倍或更大自然数倍。因此,通过将LBA除以磁道大小来获得磁道地址。磁道中的地址是通过将LBA除以磁道大小而获得的余数。通过将LBA除以集群大小来获得集群地址。集群中的地址是通过将LBA除以集群大小而获得的余数。在以下解释中,为了方便起见,磁道的大小等于可记录在一个物理块中的数据的大小(当由SSDC 41执行的ECC处理的冗余位包括在物理块中时,通过移除冗余位而获得的大小)。集群的大小等于可记录在一个物理页面中的数据的大小(当由SSDC 41执行的ECC处理的冗余位包括在物理页面中时,通过移除冗余位而获得的大小)。

[0186] 空闲块表(FBT)60管理用途未分配的物理块(空闲块:FB)的块地址(物理块ID),此用途未分配的物理块可以在执行NAND存储器16中的写入时被重新分配以便写入。空闲块表(FBT)60管理每个物理块ID的擦除计数。当物理块被擦除时,空闲块表(FBT)60递增此块的擦除计数。

[0187] 坏块表(BBT)61管理坏块(BB)的块ID,坏块(BB)用作例如因为经常出现错误而不能用作存储装置的物理块。如在FBT 60中,可以针对每个物理块ID管理擦除计数。

[0188] 活动块表(ABT)62管理活动块(AB),活动块(AB)是为其分配了用途的物理块。活动块表(ABT)62管理每个物理块ID的擦除计数。当物理块被擦除时,活动块表(ABT)62递增此块的擦除次数。

[0189] 磁道表63管理磁道地址与物理块的物理块ID之间的对应关系,对应于磁道地址的磁道数据存储在物理块ID中。

[0190] 集群表64管理以下各项中的对应关系:集群地址、物理块的物理块ID(对应于集群地址的集群数据存储在物理块ID中),以及物理块中的页面地址(对应于集群地址的集群数据存储在物理块ID中的页面地址中)。

[0191] SSDC 41将与存储单元2的可靠性相关的各种参数(X01至X32)作为可靠性信息存储在统计信息65中(参见图27)。

[0192] 用作可靠性信息的实例的统计信息65的值(原始值)包括总的坏块计数(统计信息X01)、总的坏逻辑扇区计数(统计信息X02)、总的擦除计数(统计信息X03)、平均擦除计数(统计信息X04)、NAND存储器的总的程序错误计数(统计信息X05)、NAND存储器的总的擦除错误计数(统计信息X06)、总的读取逻辑扇区计数(统计信息X07)、总的写入逻辑扇区计数(统计信息X08)、总的不可纠正的ECC错误计数(统计信息X09)、总的重试读取计数(统计信息X10)、纠正的n位-m位ECC事件计数(统计信息X11)、接口19的数据损坏错误计数(统计信息X12)、接口19的链路速度减速计数(统计信息X13)、接口19的通道计数减小计数(统计信息X14)、接口19的错误计数(统计信息X15)、RAM 40的错误计数(统计信息X16)、存储单元2的通电时间(统计信息X17)、电源循环计数(统计信息X18)、意外断电计数(统计信息X19)、温度超过推荐的操作温度的最大值时的累计时间(统计信息X20)、温度低于推荐的操作温度的最小值时的累计时间(统计信息X21)、命令的响应时间的最大值(统计信息X22)、命令

的响应时间的平均值 (统计信息X23)、NAND存储器的响应时间的最大值 (统计信息X24)、NAND存储器的响应时间的平均值 (统计信息X25)、当前温度 (统计信息X26)、最高温度 (统计信息X27)、最低温度 (统计信息X28)、系统数据冗余性 (统计信息X29)、RAM 40中总的写入数据量 (统计信息X30)、统计信息增加比率 (统计信息X31), 以及NAND GC错误标志 (统计信息X32)。

[0193] 解释总的坏块计数 (统计信息X01)。每次将存储单元2中的NAND存储器16的一个物理块添加到坏块时,SSDC 41将统计信息X01递增1。希望在存储单元2的制造期间 (例如,在测试过程之前),SSDC 41将统计信息X01重置为0。希望当在测试过程期间在块中出现错误时,或者当检测到具有阈值电压分布的的小的分布间裕度的块时,SSDC 41提前将此块添加到坏块。SSDC 41能够直接从BBT 61计算统计信息X01,而不将统计信息X01存储在统计信息65中。统计信息X01的较大值指示可靠性进一步恶化。

[0194] 解释总的坏逻辑扇区计数 (统计信息X02)。当在从NAND闪存16读取期间从信息处理设备111接收读取命令和LBA并且不能使读取的数据经受ECC纠正时,SSDC 41能够将LBA作为坏逻辑扇区登记在管理信息44中的坏逻辑扇区表中 (参见图25)。SSDC 41将登记在坏逻辑扇区表中的LBA的数量作为总的坏逻辑扇区计数 (统计信息X02) 存储在统计信息65中。当从信息处理设备111接收读取命令时,SSDC 41读取RAM 40上的坏逻辑扇区表并且在坏逻辑扇区表中搜索接收的LBA。当在坏逻辑扇区表中找到LBA时,SSDC 41向信息处理设备111通知读取错误,而不从NAND闪存16读取数据。当从信息处理设备111接收关于坏逻辑扇区的LBA的写入命令并且执行写入处理时,SSDC 41将写入的LBA从坏逻辑扇区表中删除。当从信息处理设备111接收关于坏逻辑扇区的LBA的删除通知并且执行删除通知处理时,SSDC 41将经受删除通知处理的LBA从坏逻辑扇区表中删除。当从信息处理设备111接收用于存储单元2的擦除命令 (安全擦除命令) 时,SSDC 41擦除坏逻辑扇区表。作为用于存储单元2的擦除命令,例如可以使用ACS-3的“F4h安全擦除单元”命令或快速NVM修订1.1的“80h格式NVM”命令。代替管理LBA单元 (逻辑扇区单元) 中的坏逻辑扇区表,如图26中所示,SSDC 41能够将集群单元中的坏逻辑扇区表作为坏集群表进行管理。SSDC 41将登记在坏逻辑扇区表中的LBA的数量或者登记在坏集群表中的集群地址的数量作为统计信息X02进行管理。SSDC 41能够直接从坏逻辑扇区表和坏集群表计算统计信息X02,而不将统计信息X02存储在统计信息65中。统计信息X02的较大值指示可靠性进一步恶化。

[0195] 解释总的擦除计数 (统计信息X03)。统计信息X03指示存储单元2中的NAND存储器16的所有块的擦除计数的累计值。每当擦除存储单元2中的NAND存储器16的一个物理块时,SSDC 41将统计信息X03递增1。希望在存储单元2的制造期间 (例如,在测试过程之前),将统计信息X03重置为0。SSDC 41能够直接从FBT 60、BBT 61和ABT 62计算统计信息X03,而不将统计信息X03存储在统计信息65中。统计信息X03的较大值指示可靠性进一步恶化。

[0196] 解释平均擦除计数 (统计信息X04)。SSDC 41计算关于NAND存储器16的所有块的每个块的平均擦除计数,并且将平均擦除计数作为统计信息X04存储在统计信息65中。SSDC 41能够将一部分块 (例如其中存储管理信息44的块) 排除在统计信息X04的总体化目标之外。希望在存储单元2的制造期间 (例如,在测试过程之前),将统计信息X04重置为0。SSDC 41能够直接从FBT 60、BBT 61和ABT 62计算统计信息X04,而不将统计信息X04存储在统计信息65中。SSDC 41能够使用擦除计数的最大值或擦除计数的最小值代替平均擦除计数。统

计信息X04的较大值指示可靠性进一步劣化。

[0197] 解释NAND存储器的总的程序错误计数(统计信息X05)。每当在存储单元2中的NAND存储器16中的一个写入单元中出现程序错误时,SSDC 41将统计信息X05递增1(或者能够以块单位递增)。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X05重置为0。统计信息X05的较大值指示可靠性进一步劣化。

[0198] 解释NAND存储器的总的擦除错误计数(统计信息X06)。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X06重置为0。每当在存储单元2中的NAND存储器16中的一个块中出现擦除错误时,SSDC 41将统计信息X06递增1。SSDC 41可以将多个块共同地设置为擦除单元,并且每当一个擦除单元中出现擦除错误时,将统计信息X06递增1。统计信息X06的较大值指示可靠性进一步劣化。

[0199] 解释总的读取逻辑扇区计数(统计信息X07)。SSDC 41将由IFC 42作为读取数据传输到信息处理设备111的数据的逻辑扇区数量的累计数量作为统计信息X07存储在统计信息65中。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X07重置为0。统计信息X07的较大值指示可靠性进一步劣化。

[0200] 解释总的写入逻辑扇区计数(统计信息X08)。SSDC 41将由IFC 42作为写入数据从信息处理设备111接收的数据的逻辑扇区总数作为统计信息X08存储在统计信息65中。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X08重置为0。统计信息X08的较大值指示可靠性进一步劣化。

[0201] 解释总的不可纠正的ECC错误计数(统计信息X09)。当不能通过ECC纠正来纠正错误位时,SSDC 41在每个读取单元中将统计信息X09递增1。SSDC 41能够添加不能纠正的错误位的数量的估计值,或者能够添加不能纠正的错误块的数量。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X09重置为0。统计信息X09的较大值指示可靠性进一步劣化。

[0202] 解释总的重试读取计数(统计信息X10)。当错误位的数量较大并且在数据读取期间错误纠正不可能(ECC错误)时,希望SSDC 41再次使用ECC执行错误纠正。具体地说,当SSDC 41将图12中所示的读取级别VA1、VA2、VB2和VC2从默认值移位并且执行读取时,有时能够对不能被错误纠正的数据进行错误纠正。SSDC 41可以将总的重试读取计数作为统计信息X10存储在统计信息X09中,并且使用总的重试读取计数进行寿命结束预测和寿命结束确定。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X10重置为0。统计信息X10的较大值指示可靠性进一步劣化。

[0203] 解释纠正的n位-m位ECC事件计数(统计信息X11)。在纠正的n位-m位ECC事件计数中,n和m是自然数并且 $0 \leq n \leq m \leq$ 可纠正位的最大数量。当针对ECC纠正单元(例如,物理页面)执行ECC纠正时,如果所有错误位正常地被纠正并且纠正的错误位的数量等于或大于n且等于或小于m,则SSDC 41针对一个ECC纠正单元将纠正的n位-m位ECC事件计数递增1。如果每个纠正单元可以通过ECC纠正而纠正最大64位,则例如SSDC 41保留八个参数“纠正的1位-8位ECC事件计数”、“纠正的9位-16位ECC事件计数”、“纠正的17位-24位ECC事件计数”、“纠正的25位-32位ECC事件计数”、“纠正的33位-40位ECC事件计数”、“纠正的41位-48位ECC事件计数”、“纠正的49位-56位ECC事件计数”和“纠正的57位-64位ECC事件计数”。如果ECC纠正正常地执行,则每当执行一个ECC纠正单元中的ECC纠正时,SSDC 41将此八个参数中的

任意一个递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X11重置为0。统计信息X11的较大值指示可靠性进一步劣化。

[0204] 解释接口19的数据损坏错误计数(统计信息X12)。每当在接口19上检测到信号的数据损坏时,SSDC 41将统计信息X12递增1。使在接口19上传输和接收的数据经受由SSDC 41、IFC 42和芯片组7例如使用循环冗余校验(CRC)码、博斯-查德胡里-霍昆格母(BCH)码、里德-所罗门(RS)码、低密度奇偶校验(LDPC)码等进行的错误检测和错误纠正。当检测到错误时或者当不能执行错误纠正时,SSDC 41将统计信息X12递增1。例如,当接口19基于SATA标准时,每当出现一次SATA标准中的R错误(接收错误,R_ERR),SSDC 41将统计信息X12递增1。作为统计信息X12,可以采用SATA标准的物理事件计数器中的任意一个。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X12重置为0。统计信息X12的较大值指示可靠性进一步劣化。

[0205] 解释接口19的链路速度减速计数(统计信息X13)。当SSDC 41、IFC 42和芯片组7检测到接口19的通信速度减小到小于设计值时,SSDC 41将统计信息X13递增1。例如,尽管在最大6Gbps的SATA通信速度下设计接口19、IFC 42和SSDC 41,但当检测到实际上在接口19、IFC 42和SSDC 41与存储单元2和信息处理设备111之间建立的通信速度为诸如3Gbps之类的较低通信速度时,SSDC 41将此情况视为SATA通信中的错误并且将统计信息X13递增1。例如,尽管在最大8Gbps的快速通信速度下设计接口19、IFC 42和SSDC 41,但当检测到实际上在接口19、IFC 42和SSDC 41与存储单元2和信息处理设备111之间建立的通信速度为诸如5Gbps之类的较低通信速度时,SSDC 41将此情况视为快速PCI通信中的错误并且将统计信息X13递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X13重置为0。统计信息X13的较大值指示可靠性进一步劣化。

[0206] 解释接口19的通道计数减小计数(统计信息X14)。当SSDC 41、IFC 42和芯片组7检测到接口19的活动传输线的数量减小到小于设计值时,SSDC 41将统计信息X14递增1。例如,尽管在最大8个通道的快速PCI传输线的数量(通道的数量)下设计接口19、IFC 42和SSDC 41,但当检测到实际上在接口19、IFC 42和SSDC 41与存储单元2和信息处理设备111之间建立的传输线的数量为诸如4个通道之类的较小数量的传输线时,SSDC 41将此情况视为快速PCI通信中的错误并且将统计信息X14递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X14重置为0。统计信息X14的较大值指示可靠性进一步劣化。

[0207] 解释接口19的错误计数(统计信息X15)。每当SSDC 41、IFC 42和芯片组7检测到一次在接口19中的其它异常(X12除外),SSDC 41将统计信息X15递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X15重置为0。统计信息X15的较大值指示可靠性进一步劣化。

[0208] 解释RAM 40的错误计数(统计信息X16)。当SSDC 41将数据写入RAM 40中时,SSDC 41或RAM 40的ECC编码单元或错误检测码创建单元对此数据编码并且将此数据写入RAM 40中。当SSDC 41从RAM 40读取数据时,SSDC 41或RAM 40的ECC解码单元或错误检测单元使数据经受错误纠正或错误检测并且从RAM 40读取数据。如果不能执行错误纠正或者如果在SSDC 41从RAM 40读取数据时检测到错误,则SSDC 41将统计信息X16递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X16重置为0。统计信息X16的较大值指示可靠性进一步劣化。

[0209] 解释存储单元2的通电时间(统计信息X17)。当用于存储单元2的电源接通时,当SSDC 41对时钟计数或者从内部计时电路接收时间信息时,SSDC 41递增作为经过的时间的统计信息X17。备选地,SSDC 41能够定期地从信息处理设备111接收信息处理设备111的时间信息,并且递增时间信息之差。作为通电时间的实例,例如存在通电数小时和通电数秒。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X17重置为0。统计信息X17的较大值指示可靠性进一步劣化。

[0210] 解释电源循环计数(统计信息X18)。每当将电源提供给存储单元2并且启动存储单元2时,SSDC 41将统计信息X18递增1。在电源提供和启动期间,在某种情况下,针对NAND闪存16发生读取操作并且发生写入操作。因此,统计信息18的较大值指示可靠性进一步劣化。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X18重置为0。

[0211] 解释意外断电计数(统计信息X19)。通常,当用于存储单元2的电源关闭时,例如信息处理设备111向存储单元2例如发出信息技术ATA/ATAPI命令集-3(ACS-3)中描述的“E0h立即待机”命令,或者将快速NVM修订1.1中描述的“关机通知(CC.SHN)”设置为01b。这样,信息处理设备111将存储单元2转变为其中可以中断电源的状态,并且然后中断用于存储单元2的电源。另一方面,在某种情况下,当存储单元2不处于其中能够中断电源的状态时,无意间出现电源中断。此情况被称为意外断电(无礼貌的断电、不安全的关机和故意断电)。当存储单元2在不适当的电源中断之后第一次启动时,SSDC 41将统计信息X19递增1。不适当的电源中断还导致存储单元2的可靠性劣化,因为用户数据损坏或者在不适当的电源中断中发生来自和去往NAND存储器16的大量读取和写入操作。因此,统计信息X19的较大值指示可靠性进一步劣化。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X19重置为0。

[0212] 解释温度超过推荐的操作温度的最大值时的累计时间(统计信息X20)。当在存储单元2中安装温度计时,例如安装在存储单元2的衬底上、SSDC 41中或NAND存储器16中,SSDC 41定期地从温度计接收温度信息。当接收的温度超过推荐的操作温度(例如,100℃)时,SSDC 41基于从时钟、内部时钟或信息处理设备111获得的时间信息,递增等于或高于推荐的操作温度的温度下的操作小时数。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X20重置为0。统计信息X20的较大值指示可靠性进一步劣化。

[0213] 解释温度低于推荐的操作温度的最小值时的累计时间(统计信息X21)。当在存储单元2中安装温度计时,SSDC 41定期地从温度计接收温度信息。当接收的温度低于推荐的操作温度(例如,-40℃)时,SSDC 41基于从时钟、内部时钟或信息处理设备111获得的时间信息,递增等于或高于推荐的操作温度的温度下的操作小时数。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X21重置为0。统计信息X21的较大值指示可靠性进一步劣化。

[0214] 解释命令的响应时间最大值(统计信息X22)。SSDC 41测量从接收来自信息处理设备111的命令时起直至将响应传输到信息处理设备111(或者完成对命令的执行)为止所需的时间(或时钟的数量),并且将时间的最大值作为统计信息X22存储在统计信息65中。当出现超过统计信息X22的响应时间时,SSDC 41使用此响应时间盖写统计信息X22。SSDC 41能够针对每个命令存储统计信息X22。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X22重置为0。

[0215] 解释命令的响应时间平均值(统计信息X23)。SSDC 41测量从接收来自信息处理设备111的命令时起直至将响应传输到信息处理设备111(或者完成对命令的执行)为止所需的时间(或时钟的数量),并且将时间的平均值作为统计信息X23存储在统计信息65中。例如,SSDC 41将固定数量的响应时间列表存储在RAM 40中,并且计算这些响应时间列表的平均值以便从而计算统计信息X23。SSDC 41能够针对每个命令存储统计信息X23。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X23重置为0。

[0216] 解释NAND存储器的响应时间最大值(统计信息X24)。SSDC 41测量从由SSDC 41向NAND存储器16发出命令时起直至接收响应(或者接收命令执行完成通知)为止所需的时间(或时钟的数量),并且将时间的最大值作为统计信息X24存储在统计信息65中。当出现超过统计信息X24的响应时间时,SSDC 41使用此响应时间盖写统计信息X24。SSDC 41能够针对每个命令存储统计信息X24。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X24重置为0。

[0217] 解释NAND存储器的响应时间平均值(统计信息X25)。SSDC 41测量从向NAND存储器16发出命令时起直至接收响应(或者接收命令执行完成通知)为止所需的时间(或时钟的数量),并且将时间的平均值作为统计信息X25存储在统计信息65中。例如,SSDC 41将固定数量的响应时间列表存储在RAM 40中,并且计算这些响应时间列表的平均值以便从而获得统计信息X25。SSDC 41能够针对每个命令存储统计信息X25。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X25重置为0。

[0218] 解释当前温度(统计信息X26)。当在存储单元2中安装温度计时,SSDC 41定期地从温度计接收温度信息。SSDC 41将最后从温度计接收的温度作为当前温度存储在统计信息X26中。在当前温度极高(例如,等于或高于85℃)时,SSDC 41确定存储单元2的可靠性受到不利影响。当温度极低(例如,等于或低于-10℃)时,SSDC 41确定存储单元2的可靠性受到不利影响。

[0219] 解释最高温度(统计信息X27)。SSDC 41将当前温度X26的最大值作为最高温度存储在统计信息X27中。当最高温度极高(例如,等于或高于85℃)时,存储单元2的可靠性受到不利影响。当SSDC 41从温度计接收高于统计信息X27的当前温度时,SSDC 41使用当前温度重写统计信息X27。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X27重置为与存储单元2的操作温度相比足够低的温度(例如,-40℃)。

[0220] 解释最低温度(统计信息X28)。SSDC 41将当前温度X26的最小值作为最低温度存储在统计信息X28中。当最低温度极低(例如,等于或低于-40℃)时,SSDC 41确定存储单元2的可靠性受到不利影响。当SSDC 41从温度计接收低于统计信息X28的当前温度时,SSDC 41使用当前温度重写统计信息X28。希望在存储单元2的制造期间(例如,在测试过程之前)或者在存储单元2的运送期间,将统计信息X28重置为与存储单元2的操作温度相比足够高的温度(例如,120℃)。

[0221] 解释系统数据冗余性(统计信息X29)。当系统数据(例如NAND存储器16的管理信息区域40M中的数据或者存储在NAND存储器16中的系统程序(固件))受到损坏并且不能读取时,很可能存储单元2无法执行正常操作。为了改进存储单元2的可靠性,希望SSDC 41使用

RAID1、RAID5或RAID6使得系统数据跨越多个物理块或多个通道而冗余,并且将系统数据存储在区域40M中。SSDC 41将系统数据的冗余性转换为数值,并且将该数值作为系统数据冗余性(统计信息X29)存储在统计信息65中。当冗余性X29=R时,可以恢复高达最大值(R-1)个块的数据丢失。例如,当SSDC 41使用RAID1管理四个块上的管理信息45时,将管理信息45作为副本分别存储在块A、块B、块C和块D中。然后,因为管理信息45保留总共四个副本,所以管理信息45的冗余性X29为4。例如,当块A的数据受到损坏并且不能被读取时,SSDC 41可以通过从块B、块C或块D执行数据读取而读取管理信息45。然后,因为管理信息45保留总共三个副本,所以管理信息45的冗余性X29为3。例如,当SSDC 41使用RAID5管理四个块上的管理信息45时,分别通过四个RAID5将管理信息45例如存储在块A、块B、块C和块D中。然后,即使最大一个块的数据丢失,但因为能够恢复此数据,所以管理信息的冗余性X29为2。在其中一个块的数据丢失的状态下,冗余性X29为1。当冗余性X29减小时,系统数据更可能不能被恢复并且存储单元2的故障率增加。冗余性X29的较小值指示可靠性进一步劣化。当冗余性X29减小时,希望SSDC 41通过重写其中数据丢失的块中的恢复数据而恢复冗余性。

[0222] 解释RAM 40中总的写入数据量(统计信息X30)。SSDC 41将在存储单元2的RAM 40中写入的数据写入量的累计值作为统计信息X30存储在统计信息65中。每次在RAM 40中写入一个页面的数据时,SSDC 41便将统计信息X30递增1。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X30重置为0。统计信息X30的较大值指示可靠性进一步劣化。

[0223] 解释统计信息增加比率(统计信息X31)。SSDC 41将统计信息X01至X25的非最新信息(例如,固定时间之前的值、存储单元2通电时的值,以及存储单元2上次断电时的值)分别存储在管理信息44中。SSDC 41例如根据以下公式中的任意一个计算统计信息X31:

[0224] 统计信息增加比率 = (最新统计信息) - (旧信息)

[0225] 统计信息增加比率 = ((最新统计信息) - (旧信息)) / (获得旧信息之后的经过的时间)

[0226] 统计信息增加比率 = ((最新统计信息) - (旧信息)) / (获得旧信息之后的NAND存取的次数)

[0227] 希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X31重置为0。统计信息X31的较大值指示可靠性进一步劣化。

[0228] 解释NAND GC错误标志(统计信息X32)。当统计信息X32为1时,不能获得足够用于操作的多个空闲块,甚至在存储在NAND存储器16中的数据的垃圾收集(GC)(以下被称为NAND GC)的情况下也是如此。希望在存储单元2的制造期间(例如,在测试过程之前),将统计信息X32重置为0。统计信息X32的较大值指示可靠性进一步劣化。

[0229] SSDC 41能够存储上面解释的所有参数,或者能够将这些参数中的任意一个的一部分存储在统计信息65中。希望SSDC 41将统计信息65的最新信息存储在RAM 40上的区域40A中,并且定期地将最新信息作为备份数据备份在NAND存储器16上的区域40A中。另一方面,SSDC 41能够将最新信息存储在RAM 40和NAND存储器16中的一个内,并且能够将统计信息传输到信息处理设备111并将统计信息存储在信息处理设备111或连接到信息处理设备111的存储设备中。

[0230] 前向LBA查找变换

[0231] 参考图16解释其中SSDC 41指定来自LBA的物理地址的程序(前向LBA查找变换)。当指定LBA时,SSDC 41从LBA计算磁道地址、集群地址和集群内地址。

[0232] 首先,SSDC 41搜索磁道表63并且指定对应于计算的磁道地址的物理块ID(步骤S100和S101)。SSDC 41判定指定的物理块ID是否有效(步骤S102)。当物理块ID非空并且是有效值时(步骤S102处的是),SSDC 41搜索ABT 62并且判定是否在ABT 62中输入物理块ID(步骤S103)。当在ABT 62中输入物理块ID时(步骤S104处的是),从由物理块ID指定的物理块的头位置移位了磁道内地址的位置是对应于指定的LBA的NAND存储器16上的物理位置(步骤S105)。在这种情况下,对于指定对应于LBA的NAND存储器16上的物理位置而言,集群表64是不必要的。此LBA被称为“以磁道单元管理的LBA”。当在步骤S104处未在ABT 62中输入物理块ID时(步骤S104处的否),指定的LBA没有对应于其的物理地址。此状态被称为“未写入状态”(步骤S106)。

[0233] 当在步骤S102处对应于指定的磁道地址的物理地址为空并且是无效值时(步骤S102处的否),SSDC 41从LBA计算集群地址,搜索集群表64,并且从集群表64获得对应于计算的集群地址的物理块ID和对应于其的物理块内地址(步骤S107)。从由物理块ID和物理块内页面地址指定的物理页面的头位置移位了集群内地址的位置是对应于指定的LBA的NAND存储器16上的物理位置。在这种情况下,不能仅从磁道表63指定对应于LBA的NAND存储器16上的物理位置,并且需要参考集群表64。此LBA被称为“以集群单元管理的LBA”(步骤S108)。

[0234] 读取操作

[0235] 参考图17和18解释由信息处理设备111进行的从存储单元2和存储单元3的读取操作。在此实施例中解释的读取操作中,将信息技术ATA/ATAPI命令集-3(ACS-3)中描述的“60h读取排队的FPDMA(60h READ FPDMA QUEUED)”用作读取命令。但是,可以采用诸如“25h读取DMA EXT(25h READ DMA EXT)”之类的其它读取命令。读取命令的类型差异并不影响本发明的本质。例如,可以将快速NVM修订1.1中描述的“02h读取(02h READ)”用作读取命令。当存储单元2从信息处理设备111接收读取命令时(步骤S110),SSDC 41将读取命令添加到RAM 40上的读取命令队列(步骤S111),并且将读取命令的接收通知返回到信息处理设备111。

[0236] 另一方面,当读取命令存在于RAM 40上的读取命令队列中时,SSDC 41判定是否可以执行读取处理(步骤S120)。在确定能够执行读取处理时,SSDC 41根据图16中所示的前向LBA查找变换程序,从接收自信息处理设备111的LBA指定数据的物理位置(步骤S121)。SSDC 41从指定的位置中的物理页面读取数据(步骤S123),使用读取的数据中的ECC冗余位使读取的数据经受ECC解密(步骤S124),经由IFC 42将解密后的数据传输到信息处理设备111(步骤S125),并且更新统计信息65。SSDC 41能够将从NAND存储器16读取的数据一次写入RAM 40中,对写入RAM 40中的数据解密,并且将解密后的数据传输到信息处理设备111,或者能够将解密后的数据一次写入RAM 40中并将写入RAM 40中的数据传输到信息处理设备111。

[0237] 在步骤S124处,SSDC 41尝试经由ECC进行解密。但是,当不能执行解密时,SSDC 41将包括不能解密的页面的物理块从ABT 62中删除,将此物理块登记在BBT 61中,并且递增统计信息65的总的不可纠正的ECC错误计数(统计信息X09)。然后,希望SSDC 41将该块的数据从FBT 60复制到分配的空闲块,将空闲块的物理块ID登记在ABT 62中,并且将磁道表63

和集群表64的物理块从复制源物理块ID重写为复制目的地物理块ID。

[0238] 写入操作

[0239] 参考图19、图20A和20B解释由信息处理设备111进行的存储单元2和存储单元3中的写入操作。在此实施例中解释的写入操作中,将信息技术ATA/ATAPI命令集-3 (ACS-3) 中描述的“61h写入排队的FPDMA (61h WRITE FPDMA QUEUED)”用作写入命令。但是,可以采用诸如“35h写入DMA EXT (35h WRITE DMA EXT)”之类的其它写入命令。写入命令的类型差异并不影响本发明的本质。例如,可以将快速NVM修订1.1中描述的“01h写入 (01h WRITE)”用作写入命令。例如,当存储单元2从信息处理设备111接收写入命令时 (步骤S130), SSDC 41将写入命令添加到RAM 40上的写入命令队列 (步骤S131), 并且将写入命令的接收通知返回到信息处理设备111。

[0240] 另一方面,当写入命令存在于RAM 40上的写入命令队列中时,SSDC 41判定是否能够执行写入处理 (步骤S140)。在确定能够执行写入处理时,SSDC 41向信息处理设备111通知写入是可能的,从信息处理设备111接收写入数据,使接收的数据经受ECC编码,并且将编码后的数据存储存储在RAM 40的高速缓冲存储器46中。SSDC 41能够将未编码的数据存储存储在高速缓冲存储器46中,并且在将数据写入NAND存储器16中时对此数据编码。

[0241] 随后,SSDC 41读取FBT 60 (步骤S141), 并且从FBT 60获得空闲块的物理块ID。当不存在空闲块时 (步骤S142处的否), SSDC 41执行下面解释的NAND存储器16的NAND GC (步骤S143)。在NAND GC之后,SSDC 41读取FBT 60 (步骤S144), 并且从FBT 60获得空闲块的物理块ID。SSDC 41将擦除操作应用于空闲块,获得此空闲块的物理块ID。当出现擦除错误时,SSDC 41将物理块ID添加到BBT 61,将物理块ID从FBT 60中删除,并且再次从S141执行处理以便再次获得空闲块。即使在其中出现一次擦除错误的物理块中,如果再次执行擦除操作,则有时能够正常地擦除物理块而不会导致擦除错误。因此,就防止坏块数量的不必要增加而言,希望SSDC 41针对每个块保留用作FBT 60和ABT 62中的统计信息X06的每个块的擦除错误计数的项目,在出现块的擦除错误时递增此项目,并且在每个块的擦除错误计数增加到等于或大于预定值时,将此块登记在BBT 61中。更期望地,为了仅将其中连续出现擦除错误的物理块设置为坏块,SSDC 41提供项目“每个块的擦除计数连续错误”代替“每个块的擦除错误计数”,在出现块的擦除错误时递增此项目,在能够无错误地执行擦除时将此项目重置为0,并且在“每个块的擦除计数连续错误”增加到等于或大于预定值时,将此块登记在BBT 61中。

[0242] 随后,为了发现写入命令中指定的LBA是否处于未写入状态,SSDC 41根据图16中所示的前向查找变换程序判定对应于LBA的有效数据是否已经存储在NAND存储器16中 (步骤S145和S146)。

[0243] 当LBA处于未写入状态时 (步骤S146处的是), SSDC 41将存储在高速缓冲存储器46中的接收数据写入空闲块中 (步骤S147), 将写入的空闲块 (新物理块) 的ID和空闲块的擦除计数登记在ABT 62中,并且将写入的物理块的ID和空闲块的擦除计数从FBT 60中删除 (步骤S151)。然后,SSDC 41通过磁道单元中的部分 (磁道部分) 划分接收数据的LBA,并且判定磁道部分是否填充有数据以便判定以磁道单元管理LBA还是以集群单元管理LBA (步骤S152)。即,当磁道部分填充有数据时,以磁道单元管理LBA,并且当磁道部分未填充有数据时,以集群单元管理LBA。当以集群单元管理LBA时,SSDC 41重写集群表64,将新的物理块ID

与LBA关联(步骤S153),重写磁道表63,并且将无效物理块ID(例如,空值)与LBA关联。当以磁道单元管理LBA时,SSDC 41重写磁道表并且将新的物理块ID与LBA关联(步骤S154)。

[0244] 另一方面,当LBA在步骤S146中未处于未写入状态时,SSDC 41基于通过前向查找变换获得的物理块ID,从NAND存储器16中读取对应于物理块ID的物理块中的所有数据,并且将数据写入RAM 40中(步骤S148)。然后,SSDC 41在RAM 40中使用存储在高速缓冲存储器46中的数据和从NAND存储器16读取并写入RAM 40中的数据来盖写数据(步骤S149),并且将组合后的数据写入空闲块中(步骤S150)。

[0245] 当在步骤S150处出现程序错误时,SSDC 41将物理块ID添加到BBT 61,将物理块ID从FBT 60中删除,并且再次从步骤S141执行处理以便再次获得空闲块。即使在其中出现一次程序错误的物理块中,如果再次执行写入操作,则有时能够正常地写入物理块而不会导致程序错误。因此,就防止坏块数量的不必要增加而言,希望SSDC 41针对每个块保留用作FBT 60和ABT 62中的统计信息X05的每个块的程序错误出现次数的项目,在出现块的程序错误时递增此项目,并且在每个块的程序错误计数增加到等于或大于预定值时,将此块登记在BBT 61中。更期望地,为了仅将其中连续出现程序错误的物理块设置为坏块,SSDC 41提供项目“每个块的写入连续错误的次数”代替“每个块的程序错误计数”,在出现块的程序错误时递增此项目,在能够无错误地执行写入时将此项目重置为0,并且在“每个块的写入连续错误的次数”增加到等于或大于预定值时,将此块登记在BBT 61中。

[0246] SSDC 41将写入的空闲块(新物理块)的ID和擦除计数登记在ABT 62中,并且将写入的物理块的ID从FBT 60中删除(步骤S151)。当以集群单元管理LBA时,SSDC 41将集群表64的旧物理块ID重写为新物理块ID(步骤S152和S153)。当以磁道单元管理LBA时,SSDC 41将磁道表的旧物理块ID重写为新物理块ID(步骤S152和S154)。进一步,SSDC 41将旧物理块ID和旧物理块ID的擦除计数添加到FBT 60,并且将旧物理块ID和旧物理块ID的擦除次数从ABT 62中删除(步骤S155)。SSDC 41基于上面解释的写入处理而更新统计信息65。

[0247] NAND GC

[0248] 通常,存储单元2的总的LBA容量(总的逻辑扇区计数)被设计为小于存储单元2的NAND存储器16的总容量(过度供应)。因此,只要继续以磁道单元执行写入操作,便不会用尽空闲块。另一方面,当针对未写入的LBA发生大量次数的以集群单元进行的写入时,将具有大于集群的容量的物理块分配给以集群单元进行的一个写入。因此,大于要写入的数据容量的NAND存储器16的物理块是有必要的。因此,很可能用尽空闲块。当用尽空闲块时,可以通过下面解释的NAND存储器16的布置重新获得空闲块。

[0249] 参考图21解释由SSDC 41进行的NAND GC。存储在物理块中的所有集群并非始终是有有效集群。不等同于有效集群的无效集群不被映射到LBA。有效集群是其中存储最新数据的集群。无效集群是其中同一LBA的数据被写入另一个位置中并且未被参考的集群。在物理块中,由于无效集群的空间而存在用于数据的空闲空间。可以通过以下操作获得空闲块:执行NAND GC以便收集有效集群的数据,并且将数据重写在不同块中。

[0250] 首先,SSDC 41将选定的物理块ID= i 设置为0,并且将空闲空间累计量 S 设置为0(步骤S160)。SSDC 41判定是否在磁道表63中输入具有ID $i=0$ 的物理块(步骤S161)。当在磁道表中输入物理块时,SSDC 41将 i 递增1(步骤S162),并且执行关于具有下一个ID编号的物理块的相同判定(步骤S161)。即,当物理块ID包括在磁道表63中时,因为物理块的数据以

磁道单元管理,所以数据并非NAND GC目标。

[0251] 当具有ID=i的物理块未以磁道单元管理时(步骤S161处的否),SSDC 41参考集群表64并且获得具有ID=i的物理块中包括的有效集群的所有地址(步骤S163)。SSDC 41计算获得的有效集群的总容量的大小v(步骤S164)。当大小v小于物理块大小时(步骤S165处的是),SSDC 41将当前物理块的ID添加到NAND GC目标块列表(步骤S166)。进一步,SSDC 41将获得的当前物理块的集群容量v添加到获得的集群累计量S,并且更新获得的集群累计量S(步骤S167)。

[0252] 当大小v不小于物理块大小或者获得的集群累计量S未达到物理块大小时(在步骤S168处),SSDC 41将i递增1(步骤S162),并且以关于具有下一个ID编号的物理块的相同方式执行步骤S161至S167处的程序。SSDC 41重复步骤S161至S167处的程序,直至获得的集群累计量S在步骤S168处达到物理块大小。

[0253] 当获得的集群累计量S在步骤S168处达到物理块大小时,SSDC 41读取关于NAND GC目标块列表上的所有物理块的所有有效集群的数据并且将数据写入RAM 40中(步骤S169),擦除NAND GC目标块列表上的所有物理块(步骤S170),并且将擦除的所有物理块从ABT 62中删除并将物理块添加到FBT 60(步骤S171)。然后,SSDC 41递增擦除计数。在步骤S170处执行的擦除处理的目标能够限于在步骤S172处将数据写入其中的块。就抑制块的擦除计数而言,此情况是所期望的。

[0254] 当出现擦除错误时,SSDC 41将物理块ID添加到BBT 61,并且将物理块ID从FBT 60中删除。即使在其中出现一次擦除错误的物理块中,如果再次执行擦除,则有时能够正常地擦除物理块而不会导致擦除错误。因此,为了防止坏块数量的不必要增加,希望SSDC 41针对每个块保留FBT 60和ABT 62中的项目“每个块的擦除错误计数”,在出现块的擦除错误时递增此项目,并且在每个块的擦除错误计数增加到等于或大于预定值时,将此块登记在BBT 61中。更期望地,为了仅将其中连续出现擦除错误的物理块设置为坏块,SSDC 41保留项目“每个块的擦除计数连续错误”代替“每个块的擦除错误计数”,在出现块的擦除错误时递增此项目,在能够无错误地执行擦除时将此项目重置为0,并且在“每个块的擦除计数连续错误”增加到等于或大于预定值时,将此块登记在BBT 61中。

[0255] SSDC 41从FBT 60获得新的空闲块,将写入RAM 40中的数据写入获得的空闲块中(步骤S172),将其中写入此数据的空闲块的物理块ID和块的擦除计数添加到ABT 62,并且将其中写入此数据的块的块ID从FBT 60中删除(步骤S173)。进一步,SSDC 41更新集群表64中的集群地址、物理块ID和物理块内页面地址以便对应于此时的NAND GC(步骤S174)。SSDC 41基于NAND GC的处理内容而更新统计信息65。

[0256] 当在步骤S172处出现程序错误时,SSDC 41将物理块ID添加到BBT 61,将物理块ID从FBT 60中删除,并且再次获得空闲块。即使在其中出现一次程序错误的物理块中,如果再次执行写入操作,则有时能够正常地写入物理块而不会导致程序错误。因此,为了防止坏块数量的不必要增加,希望SSDC 41针对每个块保留FBT 60和ABT 62中的项目“每个块的程序错误计数”,在出现块的程序错误时递增此项目,并且在“每个块的程序错误计数”增加到等于或大于预定值时,将此块登记在BBT 61中。更期望地,为了仅将其中连续出现程序错误的物理块设置为坏块,SSDC 41保留项目“每个块的写入连续错误的次数”代替“每个块的程序错误计数”,在出现块的程序错误时递增此项目,在能够无错误地执行写入时将此项目重置

为0,并且在“每个块的写入连续错误的次数”增加到等于或大于预定值时,将此块登记在BBT 61中。

[0257] 在图21中所示的程序中,执行用于优先地在空闲块中填充数据的NAND GC。但是,用于优先地获得空闲块的NAND GC能够通过以下操作执行:在步骤S164处通过从物理块大小减去获得的集群的容量而计算v,在步骤S165处判定v是否小于0,在v小于0时转到步骤S168,并且在v不小于0时转到步骤S163。

[0258] 删除通知

[0259] 参考图22解释由SSDC 41进行的删除通知处理。删除通知是在由信息处理设备111上的OS 100执行数据的删除时,从信息处理设备111传输到存储单元2和存储单元3的命令。用于删除通知处理的命令通常被称为修剪(trim)命令。修剪命令的实例包括信息技术ATA/ATAPI命令集-3 (ACS-3) 中描述的“06h数据集管理”命令和快速NVM修订1.1中描述的“09h数据集管理”命令的解除分配。删除通知处理是这样的系统:其中,当在OS 100上或在文件系统中删除数据时,将其中存在删除的数据的逻辑地址区域(LBA区域)作为包括LBA集合和扇区数量的LBA范围表项通知给存储设备,由此也可以将此区域视为存储单元2和存储单元3上的空闲区域。SSDC 41能够根据删除通知重新获得空闲块。修剪命令的功能不仅能够通过此命令实现,而且还能够例如通过以下命令实现:信息技术ATA/ATAPI命令集-3 (ACS-3) 中描述的SCT命令传输、快速NVM修订1.1中描述的“08h写入0”命令以及诸如供应商独特命令之类的其它命令。

[0260] 当存储单元2和存储单元3从信息处理设备111接收删除通知时(步骤S180),SSDC 41根据图16中所示的程序使删除通知中指定的LBA经受前向LBA查找变换。当删除通知中包括的LBA以磁道单元管理时(步骤S181处的是),SSDC 41将物理块ID添加到FBT 60并且将物理块ID从ABT 62中删除(步骤S184)。另一方面,当删除通知中包括的LBA以集群单元管理时(步骤S181处的否),SSDC 41将对应于物理块的所有集群从集群表64中删除(步骤S182),将适当有效值(例如,FFFF)写入对应于磁道表63中对应LBA的磁道的物理块ID中(步骤S183),并且将物理块ID添加到FBT 60并将物理块ID从ABT 62中删除(步骤S184)。SSDC 41能够通过NAND GC之外的删除通知处理获得空闲块。

[0261] 通常,能够通过NAND GC获得足够用于写入的空闲块数。当甚至通过NAND GC也不能获得足够用于写入的空闲块数时,希望SSDC 41将统计信息65的NAND GC错误标志设置为1,并且向信息处理设备111通知不能通过信息处理设备111进行的对统计信息65的获取而获得空闲块。例如,为了对从NAND GC错误标志更改为1时的时间直至存储单元2实际上停止操作为止的时间提供裕度,希望在满足条件(执行NAND GC之后的空闲块的数量)<(写入所需的空闲块的数量)+(裕度)时,将NAND GC故障标志设置为1,并且将空闲块的获取中的故障作为可靠性信息通知给信息处理设备111。

[0262] 不仅能够从信息处理设备111接收写入请求时执行NAND GC,而且例如还能够在最后一次从信息处理设备接收命令之后经过预定时间时或者在从信息处理设备111接收用于转到待机、闲置或睡眠状态的命令时执行NAND GC,或者例如能够在SSDC 41通过ACS-3中描述的SCT命令传输、其它供应商命令等从信息处理设备111接收用于启动NAND GC的命令时执行NAND GC。

[0263] 错误处理

[0264] 参考图23解释由SSDC 41进行的关于NAND存储器16的错误处理。通常如上面所解释来执行各种处理,例如用于来自信息处理设备111的写入请求的处理和NAND GC处理。但是,在某种情况下,例如在NAND存储器16的写入操作(程序操作)中出现程序错误,在NAND存储器16的擦除操作中出现擦除错误,或者在NAND存储器16的读取操作中出现ECC错误(错误纠正处理中的故障)。在这种情况下,用于错误的异常处理是必要的。

[0265] 当出现这些错误中的任意一个时(步骤S190),SSDC 41将其中出现错误的物理块添加到BBT 61(步骤S191),并且将其中出现错误的物理块从ABT 62和FBT 60中删除(步骤S192)以使得此后不可能存取其中出现错误的物理块。然后,SSDC 41能够将其中出现错误的物理块的数据复制到另一个物理块。SSDC 41基于错误处理而更新统计信息65。

[0266] 上面关于读取处理、写入处理和NAND GC处理解释了错误处理的实例。但是,错误处理不限于这些实例,并且可以应用于NAND存储器16的所有类型的读取处理、写入处理和擦除处理。

[0267] 用于确定寿命结束的处理

[0268] 当信息处理设备111正在使用存储单元2时,存储在统计信息65中的值劣化并且存储单元2达到寿命结束。例如,当信息处理设备111正在使用存储单元2时,NAND存储器16的块的可靠性劣化,坏块的数量增加,并且空闲块的数量和活动块的数量总和减小。进一步,当信息处理设备111正在使用存储单元2时,即使SSDC 41执行NAND GC,也不能获得足够用于执行写入处理的空闲块数。这是存储单元2的寿命结束的实例。在以下解释中,解释由控制程序200在存储单元2的寿命结束时执行的处理。

[0269] 当启动时,控制程序200驻留在主存储器6中,并且监视存储单元2的诸如统计信息65之类的可靠性信息。为了始终监视存储单元2的统计信息65,希望在将OS 100从区域16B(或20B)读取到区域6A时或者紧接在将OS 100从区域16B(或20B)读取到区域6A之后,将控制程序200从区域16B(或区域20B)读取到区域6B(参见图2和图3)。

[0270] 例如,如图24中所示,控制程序200每隔固定时间(例如,每隔一分钟)或者在每固定次数的处理(例如,在对存储单元2的每100次存取下或者在由存储单元2接收的每10GB数据下)中,从存储单元2获得诸如统计信息65之类的可靠性信息。可通过在存储单元2更接近寿命结束时增加可靠性信息的获取频率,执行对用户数据的更可靠保护。作为获得统计信息的方法,例如可以使用INCITS ACS-3中描述的用作存储器的自诊断功能的“B0h/D0h SMART读取数据(B0h/D0h SMART READ DATA)”命令或“B0h/D5h SMART读取日志(B0h/D5h SMART READ LOG)”命令(其是S.M.A.R.T(自监视分析和报告技术)的命令)。可以使用快速NVM修订1.1中描述的“02h获得日志页面(02h Get Log Page)”命令。可以使用ACS-3中描述的SCT命令传输或其它供应商独特命令。可以使用SCSI初级命令-4(SPC-4)(INCITS T10/1731-D和修订36e(<http://www.t10.org/>))中描述的“4Dh日志感测(4h LOG SENSE)”命令。

[0271] 在图27中所示的实例中,SSDC 41基于统计信息65生成表数据,或者直接将表数据存储存储在统计信息65中并将表数据作为可靠性数据传输到信息处理设备111。当使用S.M.A.R.T作为统计信息65时,如图27中所示,SSDC 41将属性ID分配给统计信息65的每个分量。SSDC 41能够将属性ID仅分配给统计信息65的分量的一部分。关于分量,当可靠性进一步劣化时,分量的值增加,在统计信息65的分量中,SSDC 41例如计算属性值,属性值是如下面所解释的标准化后的值。

[0272] 属性值 = $SMAL + SMAB \times (1 - AMALR) \times (RMAX - \text{原始值}) / RMAX$

[0273] 原始值是自身存储在统计信息65中的值。RMAX是用于实现可靠性保证的原始值的上限值。SMAB是提前设置为属性值的初始值的参数。例如,将100设置为SMAB。SMAL(=属性阈值)是在原始值等于上限值RMAX时属性值应该达到的值,并且是提前设置的参数。例如,将30设置为SMAL。AMALR是从关系 $AMALR = SMAL / SMAB$ 导出的参数,并且等于或大于0并小于1。这样,SSDC 41计算SMART信息的属性值(图27中的“值”),并且将属性值传输到控制程序200。属性阈值是图27中的“阈值”。原始值是图27中的“原始数据”。

[0274] 例如,当SSDC 41使用平均擦除计数(统计信息X04)作为原始值时,如果当前平均擦除计数是1000,则原始数据=1000。如果假设用于实现NAND存储器16的可靠性保证的最大擦除计数是10000,则 $RMAX = 10000$ 。如果SSDC 41被设计为在擦除计数=0的初始状态下将属性值的初始值设置为100,则 $SMAB = 100$ 。当擦除计数达到 $RMAX = 10000$ 时,属性值达到SMAL。

[0275] 关于分量,当可靠性进一步劣化时,分量的值减小,在统计信息65的分量中,SSDC 41例如计算属性值,属性值是如下面所解释的标准化后的值。

[0276] 属性值 = $SMAL + SMAB \times (1 - AMALR) \times (\text{原始值} - RMIN) / (RINIT - RMIN)$

[0277] RMIN是用于实现可靠性保证的原始值的下限值。RINIT是原始值的初始值。

[0278] 能够分别采用不同值作为X01至X32的RMAX、AMALR和SMAB。当在 $SMAB = 100$ 下采用 $AMALR = 0.3$ 时,关于要采用的统计信息,属性值的最佳值是100(例如,紧接在运送之后的100)。当属性值随着可靠性进一步劣化而逐渐减小并且不能保证存储单元2的可靠性(统计信息的原始值等于或大于RMAX)或者存储单元2将要达到寿命结束时,属性值达到30或较小值。控制程序200可以使用“B0h/DAh SMART返回状态(B0h/DAh SMART RETURN STATUS)”(其是ACS-3中描述的命令)作为用于检测属性值是否超过阈值的手段,并且根据属性值是否超过阈值从此命令的输出确定寿命结束。

[0279] 希望存储单元2的制造商导出统计信息的原始值与开发阶段中的存储单元2的不良率之间的关系(如图28中所示),并且采用不良率超过可允许值时的原始值作为RMAX。例如,制造商仅需要在存储单元2的开发阶段中执行磨损测试,以便检验在针对大量(例如,一百个)存储单元2的群组重复高温下的写入操作以进行测试时,写入的数据是否在固定时间或更多时间内继续得以正确地存储,同时继续监视统计信息,并且采用不良率达到固定比率时的统计信息的原始值作为RMAX。制造商仅需要在某个时间或更多时间内使磨损的存储单元2在高温状态下不被接触,此后降低存储单元2的温度,针对存储单元2执行读取操作,并且在不能使读取的数据经受ECC纠正(或者存在不能经受ECC纠正的固定数量或更多的数据)时,将此情况定义为存储单元2的故障,并且采用通过将故障数量除以针对其执行等效测试的存储单元2的数量获得的值作为不良率。制造商仅需要采用不良率在统计上显著低于可允许的不良率的原始值作为RMAX。制造商可以为RMAX提供某种程度的裕度,并且采用 $RMAX' = RMAX - \text{裕度}$ 作为RMAX。

[0280] 制造商能够采用图27中的“最差”作为由控制程序200对存储单元2的寿命结束的诊断的索引。由SSDC 41将“最差”计算为属性值的最差值。例如,最差是属性值的最小值,例如在运送存储单元2之后(或者在制造之后)。备选地,制造商能够采用在过去的固定时间范围内的属性值的最小值作为最差,或者能够采用在从过去执行通信或处理达某个固定次数

(达固定数据量)时至当前的周期内的最小值作为最差值。

[0281] 制造商能够采用图27中的“原始数据”(原始值)作为由控制程序200对存储单元2的寿命结束的诊断的规范。将统计信息的原始值作为原始数据从存储单元2传输到控制程序200。然后,控制程序200已经将RMAX保留在控制程序200中,单独从存储单元2读取RMAX,或者从另一个存储设备读取RMAX以便获得RMAX,比较RMAX与原始数据,并且在原始数据>RMAX或原始数据 $\geq$ RMAX时,确定存储单元2已经达到寿命结束。例如,在NAND GC故障标志的情况下,当NAND GC故障标志为1时,控制程序200确定存储单元2已经达到寿命结束。例如,在总的坏块计数的情况下,当总的坏块计数超过预定值时,控制程序200确定存储单元2已经达到寿命结束。作为从存储单元2传输到信息处理设备111的原始数据,不会始终需要输出统计信息的原始值。例如,SSDC 41能够将通过使统计信息的原始值经受四个算术运算而获得的值作为原始数据传输到控制程序200。控制程序200能够将原始数据与通过使RMAX经受四个算术运算而获得的值相比较,以便判定存储单元2是否已经达到寿命结束。SSDC 41能够将例如通过对统计信息的原始值加密而获得的散列数据作为原始数据传输到控制程序200。控制程序200能够对原始数据解密,并且将原始数据与解密之后的数据的RMAX相比较以便判定存储单元2是否已经达到寿命结束。

[0282] 如上面所解释,控制程序200判定存储单元2是否已经达到寿命结束(存储单元2是否处于异常状态)。当控制程序200确定存储单元2已经达到寿命结束时(当控制程序200确定存储单元2处于异常状态时),控制程序200转到下面解释的寿命结束处理(步骤S205)。统计信息65能够采用统计信息X01至X32之外的各种形式。本发明也能够应用于这些形式。不仅能够在统计信息与不良率之间的关系中存在正相关性时应用本发明,而且还能够在统计信息与不良率之间的关系中存在负相关性时应用本发明。统计信息例如是存储单元2在运送之后经历的最低温度。然后,控制程序200仅需要采用用于实现可靠性保证的下限值RMIN代替RMAX,并且在统计信息低于RMIN时,确定存储单元2已经达到寿命结束。

[0283] 在此实施例中,控制程序200使用S.M.A.R.T每隔固定时间(例如,每隔一分钟)获得统计信息,如图24中所示(步骤S200处的是)。控制程序200将ACS-3中描述的“B0h/D0h SMART读取数据(B0h/D0h SMART READ DATA)”(其是统计信息获取命令)传输到存储单元2(步骤S201),从存储单元2接收包括统计信息的数据(步骤S202),并且诊断所接收的数据(步骤S203)。诊断方法如上面所解释。当控制程序200在步骤S204处确定存储单元2已经达到寿命结束时或者当控制程序200确定存储单元2将要达到寿命结束时(步骤S204处的是),控制程序200转到寿命结束处理(步骤S205)。即使存储单元2未达到寿命结束,例如当统计信息超过提前设置的RMAX或者指示异常值时(此情况在正常操作中不可能),也希望控制程序200转到步骤S205处的处理。

[0284] 除了SMART之外,控制程序200还能够使用ACS-3中描述的能够从存储单元2获得的固态设备统计信息(日志地址04h,日志页面07h)执行寿命结束确定。例如,当控制程序200确定偏移8-15的使用百分比的耐久性指标的值超过100%时,控制程序200能够转到寿命结束处理。

[0285] 除了SMART之外,控制程序200还能够使用能够根据ACS-3中描述的“ECh标识设备(ECh IDENTITY DEVICE)”命令获得的标识设备数据来执行寿命结束确定。例如,当控制程序200确定在标识设备数据中的特定位中设置了标志时,控制程序200能够转到寿命结束处

理。具体地说,如果存储单元2是采用专利文献3的发明的SSD,则当SSDC 41转到只读模式时,SSDC 41能够在标识设备数据中的特定位置中设置标志。因此,控制程序200能够获得标识设备数据以便识别存储单元2转变为只读模式,并且能够转到寿命结束处理。

[0286] 控制程序200能够使用专利文献22和专利文献23中公开的寿命预测技术针对存储单元2执行寿命预测,并且当确定存储单元2的寿命将要在固定周期中结束时,转到寿命结束处理。

[0287] 控制程序200能够使用统计信息65之外的可靠性信息转到寿命结束处理。例如,如图29中所示,控制程序200从OS 100获得(监视)由OS 100从存储单元2接收的响应信息(参见图7),并且使用响应信息作为可靠性信息(步骤S210)。当响应信息是错误响应时(步骤S211),控制程序200确定存储单元2达到异常状态。控制程序200转到寿命结束处理(步骤S205)。要监视的响应能够是对任何命令的响应。但是,就CPU 5上的负载的减小而言,希望仅监视对到存储单元2的写入命令的响应,写入命令例如ACS-3中描述的“61h写入排队的FPDMA(61h WRITE FPDMA QUEUED)”和“35h写入DMA EXT(35h WRITE DMA EXT)”以及快速NVM修订1.1中描述的“01h写入(01h Write)”命令。具体地说,如果存储单元2是采用专利文献3的发明的SSD,则当存储单元2已经达到寿命结束时,返回对到存储单元2的写入命令的响应作为错误。因此,可确定寿命结束而不需要获得统计信息。自然地,当存储单元2并非采用专利文献3的发明的SSD时,也可以应用本发明。

[0288] 如果存储单元2是采用专利文献3的发明的存储单元2,则在其中存储单元2响应于写入命令而返回错误的状态下,在下面解释的存储单元2的引导加载器区域的重写中,希望SSDC 41被配置为在专利文献3的只读模式状态下响应于特殊写入命令(例如,ACS-3中描述的SCT命令传输和其它供应商独特命令)而不返回错误,并且在存储单元2中使用特殊写入命令执行写入。特殊写入命令不必用于存储单元2之外的存储设备中的写入。备选地,如果OS 100是仅使用某个写入命令(例如,61h写入排队的FPDMA(61h WRITE FPDMA QUEUED))作为写入命令的OS,则SSDC 41能够被配置为在SSDC 41达到专利文献3的只读时,响应于写入命令(例如,61h写入排队的FPDMA)3而返回错误,响应于另一个写入命令(例如,30h写入扇区(30h WRITE SECTOR(S)))而不返回错误,并且使用另一个写入命令(例如,30h写入扇区)在存储单元2中执行引导加载器区域的写入。

[0289] 自然地,要监视的命令能够是写入命令之外的命令。例如,作为命令响应,能够监视对ACS-3中描述的“B0h/D4h SMART立即离线执行(B0h/D4h SMART EXECUTE OFF-LINE IMMEDIATE)”命令的响应(输出)或报告,或者能够监视对“90h执行设备诊断(90h EXECUTE DEVICE DIAGNOSTIC)”的响应。例如,控制程序200能够使用从存储单元2获得的SMART的自测试结果转到寿命结束处理。控制程序200将ACS-3中描述的“B0h/D4h SMART立即离线执行”命令传输到存储单元2,由此SSDC 41执行自测试。控制程序200将ACS-3中描述的“B0h/D0h SMART读取数据(B0h/D0h SMART READ DATA)”和“B0h/D5h SMART读取日志(B0h/D5h SMART Read Log)”命令传输到存储单元2,以便获得自测试的结果作为可靠性信息。例如,当获得的自测试结果中包括错误时,控制程序200确定存储单元2已经达到寿命结束。

[0290] 即使某个命令响应是错误响应,但如果再次传输此命令,则可能命令响应并非错误。然后,因为可能存储单元2未达到寿命结束,所以就仅在出现具有再现性的命令错误时才执行寿命结束处理而言,希望在命令错误出现多次时执行寿命结束处理。进一步,就严格

地确定错误再现性而言,希望在命令错误连续出现多次时执行寿命结束处理。备选地,如图30中所示,当在监视到存储单元2的命令期间接收作为对命令的响应的错误响应时(步骤S220和步骤S221处的是),控制程序200或OS 100能够再次将此命令传输到存储单元2(命令重试)(步骤S222),并且在接收作为重试命令的响应的错误响应时(步骤S223处的是),执行寿命结束处理(步骤S205)。

[0291] 控制程序200能够使用从信息处理设备111获得的可靠性信息转到寿命结束处理。例如,当在信息处理设备111中设置了温度计时,控制程序200能够监视从温度计输出的温度,并且在温度超过上限值或者低于下限值时,执行作为可靠性劣化时间处理的寿命结束处理。

[0292] 正常状态下的数据配置

[0293] 图31是在执行寿命结束处理S205之前由信息处理设备111管理的数据的配置实例。如上面所解释,信息处理设备111将指定LBA(其是逻辑地址)的数据读取或写入命令传输到存储单元2,而不是向存储单元2请求直接指定NAND存储器16的物理地址的读取或写入数据。存储单元2中的SSDC 41基于存储在管理信息44中的映射信息,动态地映射LBA和NAND存储器16的物理地址。这样,根据LBA管理能够直接由信息处理设备111管理的数据。作为能够由信息处理设备111管理的地址空间,将LBA区域2001映射到存储单元2。LBA区域2001包括引导加载器区域2002、元数据区域2003和用户数据区域2004。可以将这些区域的一部分分配给LBA区域之外的区域,例如可由ACS-3的“SMART读取日志”命令或“读取日志”命令存取的日志页面地址。将存储单元2分配给逻辑驱动器4。

[0294] 引导加载器区域2002是在启动信息处理设备111期间读取的区域。在此实施例中,将引导加载器区域2002分配给LBA的固定区域。但是,信息处理设备111能够动态地分配引导加载器区域2002。作为引导加载器区域的实例,例如存在主引导记录(MBR)。在MBR中,例如将LBA=0x000的总共一个逻辑扇区(总共512字节)的区域分配为固定引导加载器区域。作为引导加载器区域的实例,例如存在GUID分区表(GPT)。在此实施例中,如图32中所示,引导加载器区域2002包括:元数据指针区域2005,其中存储元数据区域2003的头LBA;状态存储区域2006,其中存储了存储状态;以及地址区域2007(数据迁移目标管理区域2007),其中存储数据迁移目标存储设备地址。例如,当存储单元2是数据迁移源并且存储单元3是数据迁移目的地时,存储单元3的存储单元标识名称存储在存储单元2的数据迁移目标管理区域2007中,并且存储单元2的存储单元标识名称存储在存储单元3的数据迁移目标管理区域3007中。在此实施例中,存储在状态存储区域2006中的数据能够采用值0至5。相应值指示存储目的地的存储单元的状态,如下所示:

[0295] 0:初始存储状态

[0296] 1:正常状态

[0297] 2:低可靠性状态

[0298] 3:数据迁移源状态(受保护状态)

[0299] 4:数据迁移目的地状态

[0300] 5:丢弃目标状态

[0301] 信息处理设备111在启动时读取指针2005,指定元数据区域2003的LBA,并且将元数据300从LBA区域2001的元数据区域2003读取到主存储器6的元数据区域6C。当发生文件

的重写时,OS 100重写主存储器6的元数据区域6C的元数据300,定期地将元数据区域6C的元数据300备份在存储单元2的元数据区域2003中,并且按顺序将元数据300的日记记录在存储单元2的元数据区域2003中。

[0302] 图33是元数据300的配置实例。文件ID是数据的地址或文件名称,应用程序400使用它标识数据。逻辑驱动器地址是分配给每个逻辑驱动器的地址,它用于指定逻辑驱动器4(参见图31和图38)。存储单元标识名称是分配给物理存储设备的地址,它用于指定诸如存储单元2或存储单元3之类的物理存储设备。在此实施例中,WWN(全球名称)用作存储单元标识名称。因为将不同值分别分配给存储设备作为WWN,所以WWN能够用于区分物理存储设备。WWN例如存储在通过ACS-3中描述的“Ech标识设备”命令读取的数据的字108-111中。备选地,作为存储单元标识名称,例如能够使用分配给通过ACS-3中描述的“Ech标识设备”命令读取的数据的字10-19的序列号,能够使用分配给通过快速NVM修订1.1中描述的“06h标识”命令读取的数据的字节23:04的序列号(SN),或者能够使用网络协议中的MAC(介质访问控制)地址或IP(网际协议)地址。能够在存储设备的制造期间提前将存储单元标识名称写入存储设备中,或者可以在存储设备连接到信息处理设备111时由信息处理设备111重新分配存储单元标识名称。通过元数据300映射的用户数据区域2004的LBA地址存储在元数据300的LBA中。用户数据区域2004之外的区域的LBA(例如元数据区域2003或引导加载器区域2002的LBA)能够存储在元数据300的LBA中。扇区计数指示数据长度。

[0303] 元数据300由OS 100用于从文件ID和逻辑驱动器地址前向查找存储单元标识名称、LBA和扇区计数,或者由OS 100用于从存储单元标识名称、LBA和扇区计数反向查找逻辑驱动器地址和文件ID。通常,应用程序400直接指定存储单元2和LBA,并且不执行从存储单元和存储设备的读取以及存储单元和存储设备中的写入。OS 100将存储单元2识别(安装)为逻辑驱动器4。当应用程序400将逻辑驱动器地址和文件ID传输到OS 100时,OS 100读取元数据300,指定对应于逻辑驱动器和文件ID的存储单元和LBA,并且将命令和LBA传输到存储单元。

[0304] 由图7中所示的OS 100和控制程序200将逻辑驱动器分配给一个或多个物理驱动器(其是更低阶层),或者分配给物理驱动器的一部分中的LBA区域。因此,作为更高阶层的应用程序400虚拟地将逻辑驱动器识别为一个驱动器。在此实施例中,在寿命结束处理S205之前的状态下,将逻辑驱动器4分配给作为物理存储设备的存储单元2。甚至在寿命结束处理S205之前的状态下,也能够使用多个物理存储设备配置具有廉价磁盘冗余阵列(RAID)的存储阵列,例如RAID0或RAID5的存储阵列。能够将存储阵列识别为一个逻辑驱动器4。也能够在此情况下应用本发明。

[0305] 图34是在应用程序400将对逻辑驱动器4的存取请求传输到OS 100时执行的程序的流程图。当应用程序400将存取请求(例如,文件读取请求或文件写入请求)、逻辑驱动器地址和文件ID传输到OS 100时(步骤S300),OS 100从元数据区域6C读取元数据300(步骤S301)。OS 100将逻辑驱动器地址和文件ID前向查找变换成存储单元标识名称和LBA(步骤S302),并且将命令和LBA传输到对应于存储单元标识名称的存储单元(步骤S303)。存储单元根据所述命令执行用户数据区域2004中的写入操作和从用户数据区域2004的读取操作,并且将响应或数据传输到OS 100(步骤S304)。OS 100从存储单元接收响应和数据并将响应和数据传输到应用程序400,并且处理结束(步骤S305)。

[0306] 在此实施例中,在执行关于存储单元2的寿命结束处理S205之前的状态下,因为仅从存储单元2配置逻辑驱动器4,所以元数据300中对应于逻辑驱动器4的所有存储单元标识名称是存储单元2的WWN。另一方面,如在逻辑驱动器中的RAID阵列的构建中,能够在执行寿命结束处理S205之前,在元数据300中将多个存储单元标识名称分配给一个逻辑驱动器地址。

[0307] 在存储状态区域2006中,存储关于OS 100的存储单元的信息。图35示出在信息处理设备111启动时以及在存储单元连接到接口19时执行的控制程序200的处理程序。控制程序200通过接口19读取存储单元的引导加载器区域2002的存储状态区域2006,并且根据读取值更改向OS 100通知的存储单元的状态。当存储状态为0时(步骤S311),控制程序200向OS 100通知存储单元是初始存储单元。OS 100将存储单元识别为初始存储单元(步骤S312)。在存储单元的出厂运送期间或者在信息处理设备111通过使用ACS-3的“F4h安全擦除单元(F4h Security Erase Unit)”命令、快速NVM修订1.1的“80h格式化NVM(80h Format NVM)”命令等擦除存储单元时,将存储单元的存储状态区域2006的值更改为存储状态=0。当信息处理设备111格式化存储单元时,将存储单元的存储状态区域2006的值更改为存储状态=0。

[0308] 当存储状态为1时(步骤S313),控制程序200向OS 100通知存储单元处于正常状态。OS 100将存储单元识别为处于正常状态(步骤S314)。在寿命结束处理S205之前的存储单元2的存储状态是存储状态=1。

[0309] 当存储状态为2时(步骤S315),控制程序200向OS 100通知存储单元处于低可靠性状态。控制程序200将存储单元识别为处于低可靠性状态(步骤S316),并且执行寿命结束处理S205。

[0310] 当存储状态为3时(步骤S317),控制程序200向OS 100通知存储单元正在充当数据迁移源执行数据迁移工作。OS 100将存储单元识别为处于受保护状态(步骤S318)。

[0311] 当存储状态为4时(步骤S319),控制程序200向OS 100通知存储单元正在充当数据迁移目的地执行数据迁移工作。OS 100将存储单元识别为处于数据迁移目的地状态(步骤S320)。

[0312] 当存储状态为5时(步骤S321),控制程序200向OS 100通知存储单元处于丢弃目标状态。OS 100将存储单元识别为处于丢弃目标状态(步骤S322)。当存储状态为0至5之外的值时,希望OS 100将存储单元视为未授权的存储单元并且不安装存储单元(步骤S323)。

[0313] 图36和37是用于解释存储单元的寿命周期的状态转变图。紧接在制造存储单元之后的、紧接在运送存储单元之后的、紧接在擦除存储单元之后的以及紧接在格式化存储单元之后的存储单元的存储状态区域2006的值是指示初始存储状态的存储状态=0(步骤S330)。此后,存储状态更改为指示正常状态的存储状态=1,由此OS 100将存储单元用作正常存储单元(步骤S331)。随着存储单元的可靠性进一步劣化,存储单元转变为指示低可靠性状态的存储状态=2(步骤S332)。在转变为指示受保护状态的存储状态=3(步骤S333)之后,存储单元更改为指示丢弃目标状态的存储状态=5(步骤S334)。于是,最后由信息处理系统1的操作员或管理员丢弃存储单元。

[0314] 图37示出信息处理设备111上识别为处于数据迁移目的地状态的存储单元的寿命周期。然后,在步骤S330处的初始存储状态(存储状态=0)之后,在转变为数据迁移目的地

状态(存储状态=4)(步骤S330b)之后,分离处于数据迁移源状态的存储单元,由此在存储状态=1的正常状态下使用此存储单元(步骤S331)。

[0315] 图38示出其中在将存储单元2识别为处于低可靠性状态时连接不同于存储单元2的存储单元3的状态。LBA区域3001被分配给连接的源单元3作为可以由信息处理设备111管理的地址空间。LBA区域3001包括引导加载器区域3002、元数据区域3003和用户数据区域3004。

[0316] 当将存储单元2识别为处于低可靠性状态时,如通过图39的虚线部分指示,能够将已经连接到信息处理设备111的处于正常状态的存储单元3识别为处于数据迁移目的地状态,而不是在处于初始存储状态的存储单元连接到信息处理设备111时,将此存储单元识别为数据迁移目的地状态。

[0317] 寿命结束处理(存储单元的寿命结束时的处理)

[0318] 图40中示出由控制程序200在寿命结束处理S205中执行的处理的流程图。当连接的存储单元2达到寿命结束或将要到达寿命结束并且开始寿命结束处理时,控制程序200判定引导加载器区域2002的存储状态区域2006是否处于存储状态=2(低可靠性状态)(步骤S340)。当判定的结果为否定时,控制程序200将存储状态区域2006重写到存储状态=2(步骤S341),并且将程序转到步骤S342。当步骤S340处的判定结果为肯定时,控制程序200将程序转到步骤S342。

[0319] 当在步骤S342处时,希望控制程序200在显示器9上显示用于促使连接新存储单元的消息,例如“请将新存储单元连接到接口19”。控制程序200判定处于存储状态=0(初始存储状态)的存储单元是否被连接(步骤S343)。

[0320] 当连接了新存储单元时(当连接了存储单元3时),控制程序200将存储单元2的引导加载器区域2002复制到处于存储状态=0(初始存储状态)的连接的存储单元3的引导加载器区域3002(步骤S344,参见图32)。控制程序200将存储单元3的引导加载器3002的存储状态区域3006重写到存储状态=4(数据迁移目的地状态)(步骤S345)。

[0321] 当处于存储状态=0(初始存储状态)的存储单元已经连接时,或者当能够分配为数据迁移目的地的处于存储状态=1(正常状态)的存储单元在步骤S340处的“是”或者在步骤S341处已经连接时,控制程序200能够将此存储单元设置为存储单元3,将存储单元2的引导加载器区域2002复制到存储单元3的引导加载器区域3002,并且将存储单元3的引导加载器区域3002的存储状态区域3006重写到存储状态=4。

[0322] 控制程序200将存储单元2的存储单元标识名称写入存储单元3的引导加载器区域3002的数据迁移目标管理区域3007中(步骤S346)。控制程序200将存储单元2的引导加载器区域2002的存储状态区域2006重写到存储状态=3(受保护状态)(步骤S347)。控制程序200将存储单元3的存储单元标识名称写入存储单元2的引导加载器区域2002的数据迁移目标管理区域2007中(步骤S346)。查看存储单元2的数据迁移目标管理区域2007和存储单元3的数据迁移目标管理区域3007,用户能够识别存储单元2和存储单元3是用于数据迁移处理的一对存储单元。

[0323] 控制程序200从主存储器6或存储单元2的元数据存储区域2003读取最新元数据300,将读取的最新元数据300写入存储单元3的元数据存储区域3003中,并且复制最新元数据300(步骤S349)。控制程序200导致OS 100将存储单元2和存储单元3识别为一个逻辑驱动

器4(步骤S350)。控制程序200更新主存储器6上的区域6D中的逻辑驱动器状态表450,以便状态从“正常状态”更改为“数据迁移状态”(步骤S351)。

[0324] 图41示出存储在主存储器6的区域6D中的逻辑驱动器状态表450。在逻辑驱动器状态表450中,管理多个逻辑驱动器与多个状态之间的对应关系。控制程序200在任何时间基于逻辑驱动器的状态(正常状态或数据迁移状态)而更新逻辑驱动器状态表450。

[0325] 在此实施例中,如图38中所示,根据寿命结束处理S205,将存储单元2和存储单元3识别为充当单一逻辑驱动器的逻辑驱动器4。在连接新存储单元3之后直至使用存储单元3构建逻辑驱动器4为止的数据的读取和写入最多以相当于元数据300的数据量的程度发生。因此,与直至在RAID存储阵列中替换存储单元并且将存储单元安装为逻辑驱动器的时间相比,存储单元3的安装以极高速度执行。

[0326] 在寿命结束处理S205之前,由OS 100定期地将主存储器6的元数据区域6C中的元数据300和元数据300的日记备份在数据迁移源存储单元2的元数据区域2003中。但是,在寿命结束处理S205之后,由OS 100定期地将主存储器6的元数据区域6C中的元数据300和元数据300的日记备份在数据迁移目的地存储单元3的元数据区域3003中。因此,将接近于最新元数据的元数据存储存在元数据区域3003中。将寿命结束处理S205之前的旧元数据存储存在元数据区域2003中。

[0327] 在上面的解释中,定义存储状态=2的低可靠性状态。但是,当控制程序200由于可靠性信息与阈值的比较而确定存储单元2已经达到寿命结束时,控制程序200能够立即将存储单元2的存储状态转到存储状态=3的数据迁移源状态(受保护状态),而不将存储状态转到存储状态=2的低可靠性状态。即,当启动寿命结束处理时,在图40中的步骤S340处,控制程序200判定存储单元2的引导加载器区域2002的存储状态区域2006是否为存储状态=3(数据迁移源状态)。当判定的结果为否定时,在图40中的步骤S341处,控制程序200将存储状态区域2006重写到存储状态=3并且将程序转到步骤S342。除了删除步骤S347之外,后续程序与图40中所示的程序相同。

[0328] 逻辑驱动器中的写入

[0329] 图42示出在将文件数据写入请求从应用程序400传输到OS 100时执行的OS 100的处理程序。OS 100从应用程序400接收写入请求、逻辑驱动器地址、文件ID和数据(步骤S360)。OS 100从主存储器6读取逻辑驱动器状态表450(步骤S361),从主存储器6读取元数据300(步骤S362),并且参考元数据300针对数据写入分配LBA(步骤S363)。

[0330] OS 100判定由写入命令基于逻辑驱动器状态表450指定的逻辑驱动器处于正常状态还是数据迁移状态(步骤S364)。当逻辑驱动器处于正常状态时,OS 100将写入命令、LBA和写入数据传输到存储单元2(步骤S365)。OS 100从存储单元2接收响应(步骤S366)。OS 100更新主存储器6上的元数据,并且将写入文件ID映射到存储单元2、LBA和扇区计数(步骤S367)。OS 100将响应传输到应用程序400(步骤S371)。

[0331] 当逻辑驱动器处于数据迁移状态时,OS 100将写入命令、LBA和写入数据传输到数据迁移目的地存储单元3(步骤S368)。OS 100从存储单元3接收响应(步骤S369)。OS 100重写主存储器6上的元数据,并且将写入文件ID映射到存储单元3、LBA和扇区计数(步骤S370)。OS 100将响应传输到应用程序400(步骤S371)。即,当逻辑驱动器处于数据迁移状态时,根据数据迁移目的地存储单元3中的写入,OS 100更新主存储器6上的元数据,以便更改

存储单元2和存储单元3的所存储的数据的地址。用于写入中涉及的已存储数据的地址的元数据的更新处理能够与存储单元3中的写入处理同时和并行地执行,能够在存储单元3中的写入处理中执行,能够在执行存储单元3中的写入处理之前执行,以及能够在执行存储单元3中的写入处理之后执行。

[0332] 对逻辑驱动器的文件删除请求

[0333] 图43示出在将文件删除请求从应用程序400传输到OS 100时执行的OS 100的处理程序。OS 100从应用程序400接收删除命令、逻辑驱动器地址和文件ID(步骤S900)。OS 100从主存储器6读取逻辑驱动器状态表450(步骤S901),从主存储器6读取元数据300(步骤S902),并且将逻辑驱动器地址和文件ID前向查找变换成参考元数据300的存储单元标识名称和LBA(步骤S903)。OS 100将其中包括删除目标文件的文件ID的行从存储器6中的元数据中删除,或者使用无效ID重写主存储器6中的元数据上的删除目标文件的文件ID以便将删除目标文件ID从元数据300中删除(步骤S904)。

[0334] OS 100判定由删除命令基于逻辑驱动器状态表450指定的逻辑驱动器处于正常状态还是数据迁移状态(步骤S905)。当逻辑驱动器处于正常状态时,OS 100将删除通知和LBA传输到存储单元2(步骤S906)。OS 100从存储单元2接收响应。OS 100将响应传输到应用程序400(步骤S910)。

[0335] 当逻辑驱动器处于数据迁移状态时,OS 100判定在前向查找变换之后的存储单元标识名称是数据迁移源存储单元2还是数据迁移目的地存储单元3(步骤S907)。当在前向查找变换之后的存储单元标识名称是存储单元2时,OS 100将删除通知和LBA传输到存储单元2(步骤S908),从存储单元2接收响应,并且将响应传输到应用程序400(步骤S910)。当在前向查找变换之后的存储单元标识名称是存储单元3时,OS 100将删除通知和LBA传输到存储单元3(步骤S909),从存储单元3接收响应,并且将响应传输到应用程序400(步骤S910)。

[0336] 从逻辑驱动器进行的数据读取

[0337] 图44示出在将对文件数据的读取请求从应用程序400传输到OS 100时执行的OS 100的处理程序。OS 100从应用程序400接收读取请求、逻辑驱动器地址和文件ID(步骤S380)。OS 100从主存储器6读取逻辑驱动器状态表450(步骤S381),从主存储器6读取元数据300(步骤S382),并且将逻辑驱动器地址和文件ID前向查找变换成参考元数据300的读取的数据的存储单元标识名称、LBA和扇区计数(步骤S383)。

[0338] 当在前向查找变换之后的存储单元标识名称指定存储单元2时(步骤S384),OS 100将读取命令、LBA和扇区计数传输到存储单元2(步骤S385)。OS 100从存储单元2接收响应和读取的数据(步骤S386)。OS 100将读取的数据和响应传输到应用程序400(步骤S389)。

[0339] 当在前向查找变换之后的存储单元标识名称指定存储单元3时(步骤S384),OS 100将读取命令、LBA和扇区计数传输到存储单元3(步骤S387)。OS 100从存储单元3接收响应和读取的数据(步骤S388)。OS 100将读取的数据和响应传输到应用程序400(步骤S389)。例如,当LBA=0指示数据迁移状态并且LBA=1指示数据未迁移状态时,如果在LBA=0和扇区计数=1的情况下执行读取,则从存储单元2和存储单元3中的存储单元3执行读取,如果在LBA=1和扇区计数=1的情况下执行读取,则从存储单元2和存储单元3中的存储单元2执行读取,并且如果在LBA=0和扇区计数=2的情况下执行读取,则从存储单元2和存储单元3两者执行读取。

[0340] 这样,禁止了数据迁移源存储设备中的写入,并且使用数据迁移目的地存储单元3中的写入实现了从数据迁移源存储设备到数据迁移目的地存储设备的数据迁移。因此,由信息处理系统1的管理员、操作员或用户进行的备份工作是不必要的。在数据迁移中,不执行用户数据2004的复制。使用用户数据2004的新写入处理执行数据迁移。因此,甚至在数据迁移期间,应用程序400的写入处理性能也不劣化。在寿命结束处理S205之后,在存储单元2中发生的写入处理最多仅限于存储状态区域2006中的写入处理。因此,存储单元2中的写入处理几乎不发生。这样,甚至在针对存储单元2的寿命结束处理S205之后,也将逻辑驱动器4本身识别为应用程序400的可读取和可写入的驱动器。但是,实际上,对于信息处理设备111,如同存储单元2是只读设备那样来对待存储单元2。

[0341] 回写式备份

[0342] 当将数据迁移源存储单元2中的数据读取到主存储器6中的高速缓冲存储器区域时,能够更新元数据300以便将读取到高速缓冲存储器区域的数据写入数据迁移目的地存储单元3中(回写),并且将数据的文件ID映射到写入目的地LBA。下面参考图45解释此情况。

[0343] OS 100从应用程序400接收读取请求、逻辑驱动器地址和文件ID(步骤S400)。OS 100从主存储器6读取逻辑驱动器状态表450(步骤S401),从主存储器6读取元数据300(步骤S402),并且将逻辑驱动器地址和文件ID前向查找变换成参考元数据300的读取的数据的存储单元标识名称、LBA和扇区计数(步骤S403)。

[0344] 当前向查找变换之后的存储单元标识名称指定存储单元3时(步骤S404),OS 100将读取命令、LBA和扇区计数传输到存储单元3(步骤S409)。OS 100从存储单元3接收响应和读取的数据(步骤S410)。OS 100将从存储单元3读取的读取数据和响应传输到应用程序400(步骤S411)。

[0345] 当前向查找变换之后的存储单元标识名称指定存储单元2时(步骤S404),OS 100将读取命令、LBA和扇区计数传输到存储单元2(步骤S405)。OS 100将写入命令、LBA和扇区计数传输到存储单元3(步骤S406)。OS 100从存储单元2接收响应和读取的数据(步骤S407)。OS 100将从存储单元2读取的数据传输到存储单元3,以便从而执行后台写入以将从存储单元2读取的数据写入存储单元3中(步骤S408)。OS 100将从存储单元2接收的数据和响应传输到应用程序400(步骤S412)。OS 100更新主存储器6上的元数据以便将写入文件ID映射到存储单元3、LBA和扇区计数(步骤S413)。

[0346] 这样,在从逻辑驱动器4读取到信息处理设备111的数据的后台,可以执行到存储单元3的数据迁移。其中应该执行解释的后台备份的LBA区域的大小得以减小。从数据迁移状态的开始至完成的周期得以进一步减小。具体地说,在逻辑驱动器4的读取操作中,并行地执行从存储单元2的数据读取以及到存储单元3的数据回写,由此可以以较高速度执行数据迁移。

[0347] 后台备份

[0348] 逻辑驱动器4处于逻辑驱动器状态表450中的数据迁移状态,当从应用程序400和OS 100向逻辑驱动器4的存取几乎不发生时(在闲置期间),在后台执行从数据迁移源存储单元2到数据迁移目的地存储单元3的备份(后台备份)。控制程序200从主存储器6读取元数据300,并且搜索映射到存储单元2的文件ID。如果存在映射到存储单元2的文件,则控制程序200将读取命令传输到存储单元2以便执行从文件的LBA的读取,并且读取数据。控制程序

200将写入命令和读取的数据传输到存储单元3的LBA,执行写入,重写主存储器6上的元数据300,并且将文件ID映射到存储单元3。

[0349] 数据迁移完成时间

[0350] 图46示出在数据迁移完成时的控制程序的操作程序。当逻辑驱动器4的状态是逻辑驱动器状态表450中的“数据迁移状态”时(步骤S420),控制程序200定期地读取主存储器6上的元数据300(步骤S421),并且定期地检查是否存在映射到存储单元2的迁移目标文件ID(步骤S422)。例如,控制程序200定期地检查存储在逻辑驱动器4中的所有文件的文件ID中是否存在映射到存储单元2的迁移目标文件ID。当存在迁移目标文件ID时,因为数据迁移尚未完成,所以控制程序200继续数据迁移状态的状态。

[0351] 另一方面,当不存在迁移目标文件ID时,控制程序200将数据迁移目的地存储单元3的存储状态区域3006重写到存储状态=1(正常状态)(步骤S423),并且将数据迁移源存储单元2的区域2006重写到存储状态=5(丢弃目标状态)(步骤S424)。控制程序200将存储单元2与逻辑驱动器4分离,仅将存储单元2和存储单元3中的存储单元3识别(安装)为逻辑驱动器4(步骤S425),并且在逻辑驱动器状态表450中将逻辑驱动器4的状态从“数据迁移状态”重写到“正常状态”(步骤S426)。

[0352] 因此,可以物理地(机械地)分离存储单元2,并且可以在任何时间丢弃存储单元2。存储单元3在寿命结束处理S205之前起着存储单元2的作用。此后,可以将存储单元3视为存储单元2。信息处理设备111的数据配置返回到图31中所示的状态,图31是寿命结束处理S205之前的数据配置。

[0353] 为了安全地将存储单元2与信息处理设备111分离,在将存储单元2与逻辑驱动器4分离之后,希望将信息技术ATA/ATAPI命令集-3(ACS-3)中描述的“E0h立即待机(E0h Standby Immediate)”命令等传输到存储单元2,或者例如通过将快速NVM修订1.1中描述的“关机通知(Shutdown Notification)(CC.SHN)”设置为01b而更改存储单元2的登记值,以便将存储单元2转变为其中电源中断是可能的状态。

[0354] 为了减少能够被丢弃的存储单元2的功耗,在将存储单元2与逻辑驱动器4分离之后,能够将诸如信息技术ATA/ATAPI命令集-3(ACS-3)中描述的“E0h立即待机”命令和“E6h睡眠(E6h SLEEP)”命令之类的状态转变命令传输到存储单元2,能够中断到存储单元2的电源,能够将存储单元2转变为串行ATA修订3.1金修订中描述的“部分”状态和“睡眠”状态,能够激活“串行ATA技术提案:SATA31_TPR_C108标题:设备睡眠”中描述的DEVSLP信号以便将存储单元2转变为DevSleep状态,或者能够将存储单元2转变为快速PCI基本规范修订3.0中描述的D1状态、D2状态或D3状态,或者能够转变为快速PCI基本规范修订3.0中描述的L1状态、L2状态或L3状态。

[0355] 图47是在其中存储单元2处于存储状态=3(受保护状态)并且存储单元3处于存储状态=4(数据迁移目的地状态)的状态下,由信息处理设备111进行的从逻辑驱动器4的读取状态的概念图。在存储装置2和3中,映射源处的LBA并不重叠。在此状态下,从存储单元2和3中的至少一个读取数据。

[0356] 图48是在其中存储单元2处于存储状态=3(受保护状态)并且存储单元3处于存储状态=4(数据迁移目的地状态)的状态下,由信息处理设备111进行的在逻辑驱动器4中的写入状态的概念图。在此状态下,仅针对存储单元2和存储单元3中的存储单元3执行写入。

即,存储单元2如同存储单元2是只读设备那样工作。当将数据写入存储单元3中时,使到存储单元2的旧映射无效。

[0357] 如上面所解释,根据此实施例,当存储单元2已经达到寿命结束或将要达到寿命结束时,当重写存储单元2时,仅发生存储状态区域2006的重写。因此,几乎不执行重写处理,并且如同存储单元2是只读设备那样来对待存储单元2。另一方面,逻辑驱动器表现为可读取和可写入的驱动器。因此,对于应用程序400,逻辑驱动器4表现得与寿命结束之前的情形相同。当从应用程序400或OS 100请求逻辑驱动器4中的写入时,发生数据从存储单元2到存储单元3的数据迁移。数据迁移通过以下操作进行的逻辑数据转换的形式来执行:从应用程序400或OS 100到存储单元3的数据写入处理和元数据重写,而不是将实体数据从存储单元2复制到存储单元3。因此,能够在从应用程序400或OS 100到存储单元2的正常数据写入的后台执行数据从存储单元2到存储单元3的数据迁移。在逻辑数据传输中,与复制实体数据相比,NAND闪存16中的读取处理量和写入处理量明显较小并且接口19的频带使用率明显较小。如上面所解释,与比较实例(其中用于从存储单元2读取数据以及将数据写入存储单元3中的处理独立于从应用程序400或OS 100向逻辑驱动器4的存取而执行)中的备份处理相比,此实施例中的数据迁移处理以明显高的速度执行。即,对于其中发生从应用程序400或OS 100的写入的LBA,数据迁移时间基本上为0。

[0358] 对于其中不发生从应用程序400或OS 100的写入的LBA,单独需要备份处理。但是,不同于其中必须在安装新存储设备之前复制数据的比较实例中的RAID阵列的备份处理和重建,可以在安装存储单元2和存储单元3之后的闲置时间内在后台执行备份处理。因此,可抑制应用程序400中的性能劣化。不同于比较实例(其中在安装新存储设备和由RAID重建逻辑驱动器之前,用户数据的复制是必要的)或类似实例(其中在安装新存储设备之前,用户数据和奇偶校验数据的重建是必要的)中的备份处理,在根据此实施例的连接数据迁移目的地存储设备中涉及的逻辑驱动器重建中,如图40中所示,仅存储状态区域和存储单元标识名称区域的重写以及元数据区域的复制是必要的。因此,可以以极高速度执行逻辑驱动器重建。

[0359] (第二实施例)

[0360] 在第一实施例中解释的实例中,存储在主存储器6中的元数据300用作用于搜索迁移到存储单元3的数据的信息。因此,例如当应用程序400向OS 100请求读取具有指定的文件ID的逻辑驱动器4的数据时,通过读取元数据300,OS 100可以获得关于应从存储单元2和存储单元3中的哪一个读取数据的信息以及关于应该从哪一个读取LBA数据的信息。在第二实施例中解释的实例中,存储在存储单元3中的数据迁移日志区域550用作用于搜索迁移到存储单元3的数据的信息。例如,当应用程序400命令OS 100读取逻辑驱动器4的数据时,通过读取数据迁移日志区域550,OS 100可以获得关于应从存储单元2和存储单元3中的哪一个读取数据的信息。在此实施例中,应用程序400将直接指定LBA的读取请求和写入请求传输到OS 100。当如在第一实施例中,应用程序400将指定文件ID的读取命令和写入命令传输到OS 100时,也能够应用此实施例的发明。然后,控制程序200或OS 100可以通过读取元数据300将文件ID变换为LBA。以与此实施例相同的方式执行关于转变后的LBA的读取和写入处理。

[0361] 在图49中,示出第二实施例中的信息处理系统1的配置。信息处理系统1的基本配

置与第一实施例中相同。逻辑驱动器是OS 100能够识别的逻辑上构建的驱动器。将逻辑驱动器ID(驱动器名称、卷号、逻辑单元号等)分配给逻辑驱动器。OS 100将充当物理设备的一个或多个存储单元识别为逻辑驱动器。将逻辑驱动器分成逻辑扇区(逻辑块)。将LBA分配给相应逻辑扇区。由图7中所示的OS 100和控制程序200将逻辑驱动器分配给一个或多个物理驱动器(其是更低阶层),以及分配给物理驱动器的一部分中的LBA区域。OS 100对逻辑驱动器的LBA和物理驱动器的LBA进行彼此变换。作为更高阶层的应用程序400虚拟地将逻辑驱动器识别为一个驱动器。在此实施例中,在寿命结束处理S205之前的状态下,将逻辑驱动器4分配给充当单个物理存储单元的存储单元2。在这种情况下,逻辑驱动器的LBA和物理驱动器的LBA具有相同值。甚至在寿命结束处理S205之前的状态下,也能够使用多个物理存储设备配置具有廉价磁盘冗余阵列(RAID)的存储阵列(例如,具有RAID0或RAID5的存储阵列),并且将此存储阵列识别为一个逻辑驱动器4。甚至在这种情况下,也可以应用此实施例。应用程序400能够通过为OS 100提供包括逻辑驱动器ID以及由LBA形成的逻辑地址的命令,存取特定逻辑驱动器中的特定逻辑扇区。能够将逻辑驱动器ID分配给存储单元的LBA区域的一部分,而不是整个LBA区域。因此,可以将存储单元2和存储单元3分成多个逻辑驱动器并且进行管理。可以将单独的逻辑驱动器ID分配给相应逻辑驱动器。

[0362] 在此实施例中,作为实例,第一实施例中描述的充当存储单元2的SSD用作存储单元2,并且第一实施例中描述的充当存储单元3的SSD用作存储单元3。为了在可靠性劣化之后丢弃存储单元2以便减少设置空间并且减少整个系统1的功耗,希望存储单元2能够物理地、可分离地附接到信息处理设备111。

[0363] 希望数据迁移目的地存储单元3的存储容量等于或大于数据迁移源存储单元2的存储容量。但是,甚至当存储单元3的存储容量小于存储单元2的存储容量时,也能够应用本发明。

[0364] 在此实施例中,存储单元3是在确定存储单元2已经达到寿命结束或将要达到寿命结束之后重新连接到信息处理设备111的存储单元。当在确定存储单元2已经达到寿命结束或将要达到寿命结束之后,已经连接到信息处理设备111的处于正常状态的存储单元3用作迁移目的地而无需新连接时,也能够应用本发明。为了减少在连接存储单元3之前的设置空间并且减少整个系统1的功耗,以及为了在存储单元3的可靠性劣化之后丢弃存储单元3,减少设置空间并且减少整个系统1的功耗,希望存储单元3可以物理地、可分离地附接到信息处理设备111。

[0365] 存储在主存储器6中的控制程序200执行对存储单元2和存储单元3、相应存储单元的状态管理区域510、逻辑驱动器ID管理区域520和数据迁移日志区域550的统计信息的控制和管理,并且基于此统计信息而执行寿命结束处理、数据迁移处理等。

[0366] 存储单元2和存储单元3分别包括状态管理区域510和逻辑驱动器ID管理区域520。存储单元3包括数据迁移日志区域550。

[0367] 在此实施例中,存储在状态管理区域510中的数据能够采用值0至5。相应值指示对应于状态管理区域510的存储单元的状态,如下所示:

[0368] 0:初始存储状态

[0369] 1:正常状态

[0370] 2:低可靠性状态

[0371] 3:数据迁移源状态(受保护状态)

[0372] 4:数据迁移目的地状态

[0373] 5:丢弃目标状态

[0374] 能够配置信息处理系统1,以便状态管理区域510、逻辑驱动器ID管理区域520和数据迁移日志区域550存储在主存储器6中,而不是存储在相应存储单元中。相同逻辑驱动器ID存储在数据迁移源存储单元2和数据迁移目的地存储单元3的逻辑驱动器ID管理区域520中。

[0375] 数据迁移日志区域550包括在数据迁移目的地存储单元3中。当信息处理设备111执行存储单元3的LBA中的写入或者信息处理设备111将删除通知传输到逻辑驱动器4以使存储单元2的LBA中的数据无效时,控制程序200将写入目标LBA和写入目标扇区大小作为逻辑驱动器4的更新日志(数据迁移日志)存储在数据迁移日志区域550中(如图50中所示)。控制程序200在逻辑驱动器4中的数据写入期间、在删除通知期间以及在存储单元3的闲置期间,在任何时间执行数据迁移日志区域550的垃圾收集和优化。例如,在将具有扇区大小X的LBA区域 $LBA = cLBA \sim cLBA + X - 1$ (以下描述为 $(cLBA, X)$)存储在数据迁移日志区域550中,当与 $(cLBA, X)$ 重叠的LBA已经存储在数据迁移日志区域550中时,或者当与 $(cLBA, X)$ 连续的LBA区域已经记录在数据迁移日志区域550中时,希望控制程序200将通过组合(合并)所存储的LBA区域和LBA $(cLBA, X)$ 而获得的LBA区域作为新日志记录在数据迁移日志区域550中,并且删除组合源LBA区域的日志。例如,当具有 $A = (cLBA, X)$ 的LBA区域A重新记录在数据迁移日志区域550中并且具有 $B = (cLBA - a, a)$ 的LBA区域B已经存储在数据迁移日志区域550中时,LBA区域A和LBA区域B是连续LBA区域。然后,控制程序200能够通过使用日志数据 $(cLBA - a, a + X)$ (其是LBA区域A+B)盖写其中 $B = (cLBA - a, a)$ 存储在数据迁移日志区域550中的区域来更新数据迁移日志区域550,而不增加数据迁移日志区域550的日志数据量。当完成下面解释的数据迁移时,控制程序200能够删除或解除分配数据迁移日志区域550,并且将数据迁移日志区域550分配用作诸如用户数据存储之类的其它用途。

[0376] 图51是数据迁移日志区域550中的日志的写入实例。将在寿命结束之前的逻辑驱动器4的 $LBA = cLBA$ 中的数据 $D(cLBA)$ 存储在存储单元2的 $LBA = cLBA$ 中。当存储单元2已经达到寿命结束时,例如如果执行在逻辑驱动器4的 $LBA = cLBA$ 中的具有一个扇区大小的数据 $D(cLBA)$ 的写入,则控制程序200控制OS 100以便执行存储单元3的 $LBA = cLBA$ 中的写入。控制程序200将 $LBA = LBA$ 和扇区计数=1作为日志数据存储在数据迁移日志区域550中。在存储单元2已经达到寿命结束之后,当OS 100将具有扇区大小=X的数据 $D(cLBA)$ 、 $D(cLBA + 1)$,...,以及 $D(cLBA + X - 1)$ 写入逻辑驱动器4中的 $LBA = cLBA$ 中时,控制程序200控制OS 100以便将数据写入 $LBA = cLBA, cLBA + 1, \dots, \text{以及} cLBA + X - 1$ 中。控制程序200将 $LBA = cLBA$ 和扇区计数=X作为日志数据存储在数据迁移日志区域550中。

[0377] 在图51中,在存储单元3的 $LBA = 1, LBA = 3$ 和 $LBA = 4$ 中执行 $LBA = 1, LBA = 3$ 和 $LBA = 4$ 中的写入。将 $LBA = 1$ 和扇区计数=1的日志记录为数据迁移日志。将 $LBA = 3$ 和扇区计数=1的日志以及 $LBA = 4$ 和扇区计数=1的日志与 $LBA = 3$ 和扇区计数=2的日志相集成并且加以记录。

[0378] 控制程序200能够将LBA区域分配为状态管理区域510、逻辑驱动器ID管理区域520和数据迁移日志区域550。备选地,控制程序200能够将并非LBA区域的逻辑地址区域(例如,

日志页面地址区域)分配为状态管理区域510、逻辑驱动器ID管理区域520和数据迁移日志区域550。当分配日志页面地址区域时,例如根据非专利文献1的ACS-3中描述的“2Fh读取日志Ext”执行日志页面区域的读取,并且根据非专利文献1的ACS-3中描述的“3Fh写入日志Ext”执行日志页面地址区域中的写入。

[0379] 以与第一实施例中的图24和29中所示相同的方式,控制程序200判定连接到CPU 5的相应存储单元是已经达到寿命结束、将要达到寿命结束还是将要出故障。当存储单元已经达到寿命结束、将要达到寿命结束或将要出故障时,控制程序200执行存储单元的寿命结束处理。如在第一实施例中,每隔固定时间、在每固定数量类型的处理下或者在每个固定数据传输和接收下(图24中所示),或者在从存储单元接收的命令响应是错误响应(如图29和30中所示)时,执行寿命结束确定。

[0380] 寿命结束处理

[0381] 图52示出此实施例中在控制程序200确定存储单元2已经达到寿命结束时执行的存储单元2的寿命结束处理。当控制程序200确定存储单元2已经达到寿命结束时(步骤S430),控制程序200将存储单元2的状态从正常状态1重写到低可靠性状态2(步骤S431)。希望控制程序200通过设置在用于将存储单元3(其是新存储单元)连接到接口19的空闲端口的端口附近的显示设备或LED等,通知信息处理系统的管理员、操作员或用户(步骤S432)。备选地,当自动地执行存储单元2或存储单元3与接口19的物理附接和分离的机械装置作为存储加载/卸载装置(图中未示出)安装在信息处理系统1上时,控制程序能够控制所述存储加载/卸载装置以便将存储单元3(其是新存储单元)连接到接口19。

[0382] 当存储单元3作为新存储单元连接时(步骤S433),控制程序200将存储单元3的状态510重写到数据迁移目的地状态4(步骤S434),并且将存储单元2的逻辑驱动器ID管理区域520的数据复制到存储单元3的逻辑驱动器ID管理区域520以便匹配存储单元2和3的逻辑驱动器ID(步骤S435)。在本实例中,因为将存储单元2分配为逻辑驱动器4(如图49中所示),所以将逻辑驱动器4的ID写入存储单元3的逻辑驱动器ID管理区域520中。控制程序200将存储单元2的状态510重写到数据受保护状态(数据迁移源状态)(步骤S436),并且导致OS 100将存储单元2和存储单元3识别为逻辑驱动器4,逻辑驱动器4是同一逻辑驱动器(步骤S437)。在寿命结束处理之后,存储单元2的状态为3并且存储单元3的状态为4。逻辑驱动器4、存储单元2和存储单元3转变为数据迁移状态。

[0383] 在此实施例中,相应存储单元的状态以非易失性方式存储在存储单元的状态管理区域510中。如图35中所示,控制程序200通过每当OS 100启动时读取状态管理区域510而识别存储单元的状态。控制程序200通过识别存储单元的状态以及从逻辑驱动器ID管理区域520读取存储单元的逻辑驱动器ID,识别逻辑驱动器4是否处于数据迁移状态。

[0384] 从逻辑驱动器的读取

[0385] 控制程序200响应于来自应用程序400的读取请求而读取数据,如图53A和53B中所示。控制程序200从应用程序400接收读取请求、读取目标逻辑驱动器ID、读取目标LBA和扇区计数(步骤S440)。控制程序200检索其中逻辑驱动器ID管理区域520的数据等于读取目标逻辑驱动器ID的所有存储单元,并且指定存储单元2和存储单元3(步骤S441)。控制程序200读取检索的存储单元的状态管理区域510的值,并且确定存储单元的状态以便指定检索的每个存储单元是存储单元2和存储单元3中的哪一个(步骤S442)。为了抑制状态管理区域

510的读取处理中涉及的信息处理系统1的性能劣化,希望控制程序200在启动信息处理设备111期间将存储单元2和存储单元3的状态管理区域510的数据作为高速缓存数据加载在主存储器6上,并且此后从主存储器6读取状态管理区域510的数据。

[0386] 当属于目标逻辑驱动器的存储单元的状态为1时,逻辑驱动器的状态为正常状态(步骤S443处的是)。控制程序200控制OS 100以便将读取命令、读取目标LBA和扇区计数传输到存储单元2(步骤S444)。控制程序200从存储单元2接收响应和读取的数据(步骤S445)。控制程序200将读取的数据和响应传输到应用程序400(步骤S446)。

[0387] 当属于目标逻辑驱动器的存储单元的状态并非1时,逻辑驱动器的状态为数据迁移状态(步骤S443处的否)。控制程序200读取存储单元3的数据迁移日志区域550(步骤S447),并且判定读取目标LBA是否包括在数据迁移日志中(步骤S448)。当读取目标LBA包括在数据迁移日志中时(步骤S448处的是),控制程序200将读取命令、读取目标LBA和扇区计数传输到存储单元3(步骤S452)。控制程序200从存储单元3接收响应和读取的数据(步骤S453)。控制程序200将读取的数据和响应传输到应用程序400(步骤S455)。

[0388] 当读取目标LBA未包括在数据迁移日志中时(步骤S448处的否),控制程序200将读取命令、读取目标LBA和扇区计数传输到存储单元2(步骤S449)。控制程序200从存储单元2接收响应和读取的数据(步骤S450)。控制程序200将读取的数据和响应传输到应用程序400(步骤S455)。

[0389] 当包括在数据迁移日志中的LBA区域和不包括在数据迁移日志中的LBA区域在读取目标LBA区域中被混合时,控制程序200将读取目标LBA区域分成包括在数据迁移日志中的LBA区域和不包括在数据迁移日志中的LBA区域,并且执行上面针对相应区域解释的处理。

[0390] 回写式备份

[0391] 例如,在图53A和53B中,可以执行或者不必执行步骤S451处的回写式备份。在步骤S451处,当将处于数据迁移下的数据迁移源存储单元2的数据读取到主存储器6中的高速缓冲存储器区域时,将读取到高速缓冲存储器区域的数据写入处于数据迁移下的数据迁移目的地存储单元3中,并且将写入目的地LBA和扇区计数写入数据迁移日志区域550中。因此,可在将数据从逻辑驱动器4读取到信息处理设备111的后台执行到存储单元3的数据转变。应该在后台备份的LBA区域的大小得以减小,并且从数据迁移状态的开始至完成的周期得以进一步减小。具体地说,在到逻辑驱动器4的读取操作中,并行地执行从存储单元2的数据读取以及在存储单元3中的数据回写,由此能够以高速度执行数据迁移。

[0392] 对逻辑驱动器的数据删除请求

[0393] 图54示出在将LBA数据删除请求从应用程序400传输到OS 100时执行的OS 100的处理程序。OS 100从应用程序400接收数据删除请求、逻辑驱动器地址和删除目标LBA(步骤S920)。控制程序200检索其中逻辑驱动器ID管理区域520的数据等于LBA数据删除目标的逻辑驱动器ID的所有存储单元,并且指定存储单元2和存储单元3(步骤S921)。控制程序200读取检索的存储单元的状态管理区域510的值(步骤S922),并且确定状态以便指定每个检索存储单元是存储单元2和存储单元3中的哪一个。

[0394] 当属于目标逻辑驱动器的存储单元的状态为1时(步骤S923处的是),逻辑驱动器的状态为正常状态。OS 100将删除通知和LBA传输到存储单元2(步骤S924)。OS 100从存储

单元2接收响应。OS 100将响应传输到应用程序400 (步骤S930)。

[0395] 当属于目标逻辑驱动器的存储单元的状态并非1时 (步骤S923处的否), 逻辑驱动器的状态为数据迁移状态。控制程序200读取存储单元3的数据迁移日志区域550 (步骤S925), 并且判定数据删除目标LBA是否包括在数据迁移日志中 (步骤S926)。当删除目标LBA包括在数据迁移日志中时, 将删除目标数据存储在存储单元3中。控制程序200将删除通知和LBA传输到存储单元3 (步骤S927)。存储单元3使删除通知目标LBA的数据无效, 从存储单元3接收响应, 并且将响应传输到应用程序400 (步骤S930)。

[0396] 当删除目标LBA不包括在数据迁移日志中时 (步骤S926), 将删除目标数据存储在存储单元2中。控制程序200将删除通知和LBA传输到存储单元2 (步骤S928), 并且从存储单元2接收响应。控制程序200不必将删除通知传输到存储单元2。设置为来自应用程序400的删除命令的目标的LBA是不必在未来供应用程序400和OS 100读取的数据, 并且是不必迁移到存储单元3的数据。因此, 控制程序200将删除目标LBA和扇区计数记录在数据迁移日志区域550中, 以便从而使从删除目标LBA到存储单元2的映射无效 (步骤S929)。控制程序200将响应传输到应用程序400 (步骤S930)。

[0397] 当包括在数据迁移日志中的LBA区域和不包括在数据迁移日志中的LBA区域在删除目标LBA区域中被混合时, 控制程序200将删除目标LBA区域分成包括在数据迁移日志中的LBA区域和不包括在数据迁移日志中的LBA区域, 并且执行上面针对相应区域解释的处理。

[0398] 这样, 在数据删除请求的处理中, 控制程序200更新数据迁移日志区域550。因此, 逻辑上将数据从存储单元2迁移到存储单元3。数据删除也用作数据迁移操作。

[0399] 逻辑驱动器中的写入

[0400] 控制程序200响应于来自应用程序400的写入命令而写入数据, 如图55中所示。控制程序200从应用程序400接收写入请求、写入目标逻辑驱动器ID、写入目标LBA和扇区计数 (步骤S460)。控制程序200检索其中逻辑驱动器ID管理区域520的数据等于写入目标逻辑驱动器ID的所有存储单元, 并且指定存储单元2和存储单元3 (步骤S461)。控制程序200读取检索的存储单元的状态管理区域510的值, 并且确定存储单元的状态以便指定每个检索的存储单元是存储单元2和存储单元3中的哪一个 (步骤S462)。

[0401] 当属于目标逻辑驱动器的存储单元的状态为正常状态时, 逻辑驱动器的状态为正常状态 (步骤S463处的是)。控制程序200将写入命令、写入目标LBA和扇区计数传输到存储单元2 (步骤S464)。控制程序200将从应用接收的写入数据传输到存储单元2 (步骤S465)。

[0402] 当属于目标逻辑驱动器的存储单元的状态为数据迁移状态时, 逻辑驱动器的状态为数据迁移状态 (步骤S463处的否)。控制程序200将写入命令、写入目标LBA和扇区计数传输到数据迁移目的地存储单元3 (步骤S466)。控制程序200将从应用接收的写入数据传输到存储单元3 (步骤S467)。控制程序200读取存储单元3的数据迁移日志区域550 (步骤S468), 并且判定写入目标LBA是否包括在数据迁移日志中 (步骤S469)。当写入目标LBA包括在数据迁移日志中时, 写入目标LBA已经经受数据迁移。因此, 控制程序200并不更新数据迁移日志区域550。当写入目标LBA不包括在数据迁移日志中时, 写入目标LBA是重新完成迁移的LBA。因此, 控制程序200将写入目标LBA和扇区计数记录在数据迁移日志区域550中 (步骤S470)。存储单元3将写入数据写入在写入目标LBA中。

[0403] 这样,在数据迁移期间,控制OS 100以便不将写入请求传输到数据迁移源存储单元2,而将写入请求传输到数据迁移目的地存储单元3。将数据迁移日志记录在存储单元3的数据迁移日志区域550中。每当逻辑驱动器4从应用程序400接收写入请求时,逐渐地将存储在存储单元2中的有效数据迁移到存储单元3。新数据写入也用作数据迁移操作。

[0404] 如果假设以均匀概率分布将来自应用程序400的数据写入请求传输到存储单元2的所有LBA,则当总共写入足够大量的数据时,将存储单元2的几乎所有有效数据传输到存储单元3。有效数据几乎未遗留在存储单元2中。将存储单元2的逻辑扇区总数(其是存储容量)定义为 C_2 ,将存储单元3的逻辑扇区总数定义为 C_3 ,并且例如 $C_2 = C_3 = C$ 。如果假设所有LBA的写入分布是均匀概率分布(作为模型案例),则通过某一写入请求而写入 $LBA = cLBA$ 的概率是 $1/C$ 。当处理 n 个写入请求时,根本不写入 $LBA = cLBA$ 的概率是 $(1 - (1/C)^n)$ 。 n 是第 n 次幂。因此,在处理写入请求 n 次之后完成写入的逻辑扇区的数量的期望值是 $C - C \times (1 - (1/C)^n)$ 。

[0405] 如果根据一个写入请求执行一个逻辑扇区的写入,则当写入具有与存储单元2和存储单元3的存储容量的 N 倍一般大的容量的数据时,因为处理的写入命令的数量是 $n = NC$,所以其中不执行写入的逻辑扇区的数量的期望值 E 是 $E = C \times (1 - (1/C)^{NC})$ 。例如,当存储单元2的存储容量 G (以GB为单位)是 $G = 512GB (= 476.9GiB)$ (基于IDEMA(国际磁盘驱动器设备和材料协会)标准)时,因为 $C = 97,696,368 + 1,953,504 \times (G = 512 - 50) = 1,000,215,216$,并且通常 C 是足够大的整数,所以期望值 E 能够近似为 $E = C \times e^{-(N)}$ (e 是自然对数的底数)。因此,期望值 E 针对 N 的增加呈指数减小。例如,当针对具有容量 $G = 512GB (= 476.9GiB)$ 的逻辑驱动器4发生相当于逻辑驱动器4的一轮的 $476.9GiB$ 的数据写入时,逻辑驱动器4的大约63.2%的LBA中的写入完成。能够认为逻辑驱动器4的一半或更多数据从存储单元2到存储单元3的传输完成。例如,当针对具有容量 $G = 512GB (= 476.9GiB)$ 的逻辑驱动器4发生相当于逻辑驱动器4的4.6轮的13TB的数据写入时,逻辑驱动器4的大约99%的LBA中的写入完成。能够认为逻辑驱动器4的基本上所有数据从存储单元2到存储单元3的传输完成。 $K_i = 1024, M_i = 1024 \times K_i, G_i = 1024 \times M_i$,并且 $T_i = 1024 \times G_i$ 。

[0406] 数据迁移状态的监视

[0407] 控制程序200读取数据迁移日志区域550以便监视在数据迁移状态下的逻辑驱动器4的数据迁移状态。图56示出用于使用数据迁移日志监视数据迁移状态的监视程序。例如,每当经过预定时间时,控制程序200读取数据迁移日志区域550以便监视数据迁移状态(步骤S480和S481)。当所有迁移目标LBA包括在数据迁移日志区域550中时,控制程序200确定数据迁移完成。例如,数据迁移源存储单元2的所有LBA包括在数据迁移日志区域550中,控制程序200确定数据迁移完成(步骤S482)。备选地,作为数据迁移完成的确定,例如控制程序200能够判定存储单元3的所有LBA是否包括在数据迁移日志区域550中。

[0408] 当控制程序200确定数据迁移完成时,控制程序200将数据迁移源存储单元2的状态更改为丢弃目标状态5并且将数据迁移目的地存储单元3的状态更改为正常状态1(步骤S483),以便结束逻辑驱动器4的数据迁移状态并且结束用于逻辑驱动器4的数据迁移状态监视。为了减少存储单元2的功耗,希望控制程序200将到低功耗模式的转变请求传输到存储单元2。在数据迁移状态结束之后,希望控制程序200通过设置在端口附近的显示器9或LED通知信息处理系统的管理员、操作员或用户将存储单元2与接口19分离,或者控制存储

加载/卸载装置以便将存储单元2与接口19分离。

[0409] 整体状态的转变

[0410] 图57示出在存储单元2已经达到寿命结束时执行的逻辑驱动器4的状态转变。当存储单元2处于正常状态时,存储单元2的状态为1并且存储单元3未连接到CPU 5(步骤1)。当控制程序200确定存储单元2已经达到寿命结束时,控制程序200将存储单元2的状态更改为2(步骤2)。基于控制程序200的通知或控制,将存储单元3作为新存储单元连接到接口19(步骤3)。当连接存储单元3时,控制程序200将存储单元2的状态更改为3并且将存储单元3的状态更改为4,以便完成到数据迁移状态的转变(步骤4)。当控制程序200基于数据迁移日志区域550的信息而确定将存储单元2的所有有效数据迁移到存储单元3时,控制程序200将存储单元2的状态更改为5并且将存储单元3的状态更改为1。此后,存储单元3表现为如同存储单元3是原始存储单元2那样(处理返回到步骤1)。此后,当进一步确定存储单元3(即,新存储单元)已经达到寿命结束时,重复相同步骤2至5。这样,甚至在信息处理设备111的任意一个存储单元已经达到寿命结束时,将要达到寿命结束时,或者将要出故障时,可很容易地根据此实施例将存储单元2的数据迁移到新存储单元。

[0411] 后台备份

[0412] 例如,在闲置期间,当OS 100在固定时间或更多时间内未从应用程序400接收请求时,当OS 100从应用程序400接收待机模式转变请求时,或者信息处理设备111和OS 100关闭时,希望控制程序200针对其中到存储单元3的数据迁移未完成的LBA,执行后台备份以执行备份操作以便自动地从存储单元2读取数据并且自动地将数据写入存储单元3中。例如,后台备份操作由控制程序200通过以下操作执行:读取存储单元3的数据迁移日志区域550,针对未包括在数据迁移日志区域550中的cLBA执行从存储单元2的数据读取,并且将cLBA和写入数据的扇区大小作为日志数据存储在数据迁移日志区域550中。在将cLBA存储在数据迁移日志区域550中,当存在连续LBA区域时或者当存在重叠LBA区域时,希望将通过合并LBA区域而获得的LBA区域存储在数据迁移日志区域550中,并且将合并之前的LBA区域从数据迁移日志区域550中删除。

[0413] 为了减小数据迁移日志区域550的数据大小并且为了以高速度执行下面解释的数据迁移结束确定,希望针对cLBA区域中除了登记在数据迁移日志区域550中的cLBA区域之外的未迁移的碎片化cLBA区域优先地执行后台备份。控制程序200优先地备份未迁移的碎片化LBA区域,并且使未迁移的碎片化LBA区域经受数据迁移以便从而将新近迁移的LBA区域存储在数据迁移日志区域550中,作为通过将新近迁移的LBA区域与连续地位于新近迁移的LBA区域之前和之后的迁移的LBA区域相合并而形成的LBA区域。例如,当区域LBA=0((LBA=0,扇区大小=1))和区域LBA=2((LBA=2,扇区大小=1))的数据已经备份并且登记在数据迁移日志中时,控制程序200能够将具有连续LBA0至LBA2的LBA区域(LBA=0,扇区大小=3)更改为迁移完成状态。因此,数据迁移日志区域550的数据量得以减少。

[0414] 例如,如上面所解释,当应用程序400将总共476.9GiB的数据写入512GB(=476.9GiB)的SSD中时,控制程序200针对175.5GiB(其是存储容量的36.8%的容量)的区域执行后台备份操作,由此针对存储单元2的所有LBA区域完成数据迁移。存储单元2(其是SSD)和存储单元3(其是SSD)的典型读取速度和典型写入速度例如是大约400MiB/秒。476.9GiB的LBA区域中的从存储单元2的读取在大约449秒内完成,并且存储单元3中的写入

在大约449秒内完成。因此,在这种情况下,后台备份在最多大约15分钟内完成。当并行地执行从存储单元2的读取和存储单元3中的写入时,读取和写入在大约8分钟内完成。

[0415] 进一步,例如,如上面所解释,当应用程序400将总共13TiB的数据写入512GB(=476.9GiB)的SSD中时,控制程序200针对4.8GiB(其是存储容量的1%的容量)的区域执行后台备份操作,由此针对存储单元2的所有LBA区域完成数据迁移。存储单元2(其是SSD)和存储单元3(其是SSD)的典型读取速度和典型写入速度例如是大约400MiB/秒。4.8GiB的LBA区域中的从存储单元2的读取在大约12秒内完成,并且存储单元3中的写入在大约12秒内完成。因此,在这种情况下,后台备份在最多大约24分钟内完成。当并行地执行从存储单元2的读取和存储单元3中的写入时,读取和写入在大约12分钟内完成。

[0416] 另一方面,在不应用此实施例的情况下,用于通过从具有512GB容量的存储单元2读取所有数据并且将数据写入存储单元3中来备份数据的比较实例的数据备份需要20分钟至41分钟的时间。即,与比较实例相比,根据此实施例的应用,在相当于存储容量 $\times 1$ 的数据写入之后,实质备份所需的时间减少63%,并且在相当于存储容量 $\times 4.6$ 的数据写入之后,实质备份所需的时间减少99%。

[0417] 这样,根据此实施例的应用,由用户进行的数据备份工作是不必要的,并且显著减少了根据后台备份对信息处理设备111进行的处理的负载。应用程序400能够使用逻辑驱动器4,同时几乎不受存储单元3中来自存储单元2的数据备份的影响。此实施例中的大部分数据迁移处理是逻辑数据传输。与复制实体数据相比,NAND闪存16中的读取处理量和写入处理量明显较小并且接口19的频带使用率明显较小。仅在可靠性劣化的存储单元2中发生相当于状态管理区域510的大小的数量的重写,并且可减小由于存储单元2中的进一步数据写入而导致的存储单元2的故障率。当将新数据写入逻辑驱动器4中时,将数据写入具有高可靠性的存储单元3中而不是存储单元2中。因此,可防止写入数据的丢失。甚至在存储单元2已经达到寿命结束并且阻止进一步数据写入时,逻辑驱动器4(其是存储单元的上层)也表现为能够执行读取和写入两者的驱动器。因此,诸如应用程序之类的上部软件层能够等同地对待逻辑驱动器4,而不管逻辑驱动器4处于寿命结束模式还是正常状态。因此,不需要用于引入此实施例的应用程序修改,并且很容易转到采用此实施例的系统。

[0418] (第三实施例)

[0419] 在第三实施例中解释的实例中,将本发明应用于包括存储阵列的信息处理系统1。图58示出根据第三实施例的信息处理系统1。信息处理系统1包括:存储阵列设备1003;存储单元2A至2D;存储单元3;接口19,其被配置为连接存储阵列设备1003、存储单元2A至2D和存储单元3;客户机1002;以及存储网络1000,其被配置为连接客户机1002和存储阵列设备1003。在信息处理系统1中,存储单元2A至2D连接到存储阵列设备1003,并且分别识别为充当逻辑单元的逻辑插槽。RAID(廉价磁盘冗余阵列)阵列使用逻辑插槽构建。充当数据迁移目的地的存储单元3能够进一步连接到存储阵列设备1003。在此实施例中,在寿命结束处理之前,四个存储单元配置RAID阵列。但是,可以使用任意两个至多个存储单元构建RAID阵列。在此实施例中,RAID5用作RAID阵列。但是,当使用诸如RAID0、RAID2、RAID3、RAID4、RAID6和RAID Z之类的其它RAID技术和其它存储阵列实现形式构建存储阵列时,也能够应用此实施例。

[0420] 网络1000是用于存储存取的存储网络。例如,使用光纤通道或以太网(注册商标)。

具体地说,作为存储网络1000,例如使用SAN(存储区域网络)或NAS(网络附加存储)。作为SAN,例如使用FC-SAN(光纤通道存储区域网络)或IP-SAN(网际协议区域网络)。作为SAN的上层协议,例如使用SCSI(小型计算机系统接口)。在此实施例中解释的实例中,采用IP-SAN作为存储网络1000。作为IP-SAN的上层协议,使用iSCSI(因特网小型计算机系统接口)。存储网络1000包括网络交换机10001和集线器(图中未示出)。

[0421] 客户机1002是连接到存储网络1000的计算机,并且被配置为执行所需处理。通常,客户机1002包括诸如处理器、主存储器、通信接口和本地输入/输出设备之类的硬件资源。客户机1002包括诸如设备驱动程序、操作系统(OS)和应用程序(图中未示出)之类的软件资源。因此,客户机1002在控制器的控制下执行各种程序,并且与硬件资源协作实现处理。例如,客户机1002在控制器的控制下执行商业应用程序,以便从而通过存储网络1000对存储阵列设备1003进行I/O存取并且实现所需商业系统。客户机1002能够是其中数据库管理系统(DBMS)正在工作的数据库服务器(DB服务器)。然后,在通过存储网络1000或另一个网络(图中未示出)从连接到DB服务器的客户机(图中未示出)接收数据读取请求时,客户机1002从存储阵列设备1003读取数据并且将读取的数据传输到客户机。在从客户机接收数据写入请求时,客户机1002从此客户机接收写入数据并且将数据写入存储阵列设备1003中。

[0422] 存储阵列设备1003使用逻辑插槽0至3作为RAID的配置单元。逻辑插槽对应于第二实施例中的逻辑设备。在存储单元2A至2D中的任意一个已经达到寿命结束之前的正常状态下,存储单元2A至2D通过接口19连接到存储阵列设备1003。将存储单元2A分配给逻辑插槽0,将存储单元2B分配给逻辑插槽1,将存储单元2C分配给逻辑插槽2,并且将存储单元2D分配给逻辑插槽3。因此,存储阵列设备1003通知客户机1002将对应于四个存储单元2A至2D的四个逻辑插槽作为使用RAID5的一个虚拟逻辑设备。客户机1002传输用于存取存储阵列设备的LBA(以下被称为“阵列LBA”或“ALBA”)。RAID控制器1005中的控制单元200将阵列LBA变换为逻辑插槽号和用于存取存储单元2A至2D的LBA(以下被称为“存储单元LBA”或“SLBA”)。控制单元200将存取命令传输到通过逻辑插槽号指定的存储单元2A至2D中的至少一个存储单元的SLBA。

[0423] 存储阵列设备1003单独能够为客户机1002提供数据存储服务。备选地,通过存储阵列设备1003和未示出的另一个存储阵列设备虚拟地配置的一个虚拟存储设备能够为客户机1002提供数据存储服务。在存储阵列设备1003中,要提供给客户机1002的一个或多个逻辑设备(LDEV)在存储阵列设备1003中形成。

[0424] 逻辑设备是能够由客户机1002识别的逻辑存储设备。将逻辑单元(LU)分配给逻辑设备。客户机1002将在物理设备上形成的逻辑设备识别为逻辑单元。为逻辑单元提供逻辑单元号(LUN)。将逻辑单元分成逻辑扇区(逻辑块)。将阵列LBA分配给逻辑扇区。客户机1002能够通过向存储阵列设备1003给出包括由逻辑单元号和阵列LBA形成的逻辑地址的命令,存取特定逻辑单元中的特定逻辑扇区。在其中使用iSCSI的此实施例中,客户机1002和存储阵列设备1003分别充当发起者和目标,它们是被分配有iSCSI名称的iSCSI节点。因此,客户机1002和存储阵列设备1003经由通过IP地址和TCP端口号的组合指定的网络门户,传输和接收iSCSI PDU。因此,客户机1002指定iSCSI名称、IP地址和TCP端口号以便从而在网络1000上识别存储阵列设备1003,并且存取存储阵列设备1003的逻辑单元中的逻辑扇区。

[0425] 存储单元2A至2D是通过接口19连接到存储阵列设备1003的存储单元。作为存储单

元2A至2D,例如能够分别使用相当于第一实施例中解释的存储单元2的存储单元。在此实施例中,作为实例,作为存储单元2A至2D,使用第一实施例中解释的SSD。就在可靠性劣化之后丢弃存储单元2A至2D而言,减少设置空间,并且减少整个信息处理系统1的功耗,希望存储单元2A至2D物理地、可分离地附接到存储阵列设备1003。

[0426] 存储单元3是在确定存储单元2A至2D中的任意一个已经达到寿命结束或将要达到寿命结束之后重新连接到存储阵列设备1003的存储单元。例如,可以使用相当于第一实施例中描述的存储单元3的存储单元。在此实施例中,作为存储单元3,使用第一实施例中描述的SSD。为了减少在连接存储单元3之前的设置空间,为了减少整个信息处理系统1的功耗,并且为了在存储单元3的可靠性劣化之后丢弃存储单元3以便减少设置空间并且减少整个信息处理系统1的功耗,希望存储单元3物理地、可分离地附接到存储阵列设备1003。

[0427] RAID控制器1005控制连接到存储接口1007的存储单元的RAID阵列的构建和管理,并且包括控制单元200。控制单元200采用各种实现形式,例如存储在RAID控制器1005中的存储器中的固件和软件或RAID控制器1005中的硬件。在通过网络接口1004从客户机1002、网络交换机6009等接收命令时,控制单元200通过存储接口1007将读取命令、写入命令、其它命令和数据传输到存储单元,从存储单元接收响应和数据,并且通过网络接口1004将响应和数据传输到客户机1002。控制单元200执行对存储单元2A至2D和存储单元3的统计信息、状态管理区域510、插槽号管理区域530和数据迁移日志区域550的控制和管理,并且基于统计信息而执行寿命结束处理、数据迁移处理等。状态管理区域510、插槽号管理区域530和数据迁移日志区域550能够存储在信息处理系统1中的存储区域中,例如RAID控制器1005中的存储区域(图中未示出),而不是存储在存储单元中。在此实施例中,存储在状态管理区域510中的数据能够采用值0至5。相应值指示对应于其的存储单元的状态,如下所示:

[0428] 0:初始存储状态

[0429] 1:正常状态

[0430] 2:低可靠性状态

[0431] 3:数据迁移源状态(受保护状态)

[0432] 4:数据迁移目的地状态

[0433] 5:丢弃目标状态

[0434] 在图59中,示出在此实施例中的处于正常状态的阵列LBA (ALBA) 和存储单元LBA (SLBA) 的变换方法。采用RAID5的控制单元200使用三个连续逻辑扇区 $ALBA=3q$ 、 $3q+1$ 和 $3q+2$ (q 是等于或大于0的任意整数) 作为一个集合,针对阵列数据 $D(ALBA=3q)$ 、 $D(ALBA=3q+1)$ 和 $D(ALBA=3q+2)$ (其是相应逻辑扇区的数据) 生成相当于一个逻辑扇区的奇偶校验数据 $P(3q, 3q+2)$ 。

[0435] 奇偶校验数据 $P(3q, 3q+2)$ 通过按如下方式对 $D(ALBA=3q)$ 、 $D(ALBA=3q+1)$ 和 $D(ALBA=3q+2)$ 中的逻辑扇区中具有相同偏移的相应位进行异或来计算: $P(3q, 3q+2) = (D(ALBA=3q) \text{ XOR } D(ALBA=3q+1) \text{ XOR } D(ALBA=3q+2))$ 。例如,奇偶校验数据 $P(0, 2)$ 通过控制单元200从 $D(0)$ (其是 $ALBA=0$ 的数据)、 $D(1)$ (其是 $ALBA=1$ 的数据) 和 $D(2)$ (其是 $ALBA=2$ 的数据) 来计算。在正常状态下,如图59中所示,在存储单元2A至2D中分散地管理阵列数据 $D(ALBA=3q)$ 、 $D(ALBA=3q+1)$ 和 $D(ALBA=3q+2)$ 以及奇偶校验数据 $P(3q, 3q+2)$ 。例如,在 $ALBA=1$ 的情况下,对应于其的数据 $D(1)$ 存储在分配给逻辑插槽1的存储单元2B的 $LBA=SLBA=0$

中。在从客户机1002接收ALBA=1时,控制单元200指定逻辑插槽号=1和SLBA=0。

[0436] 例如,在从客户机1002接收用于ALBA=1的读取命令时,控制单元200指定逻辑插槽号=1和SLBA=0(它们是D(1)的存储目的地),将用于SLBA=0的读取命令传输到与逻辑插槽编号=1连接的存储单元2B,接收读取的数据,并且将接收的读取数据传输到客户机1002。当未从存储单元2B接收响应时或者当从存储单元2B接收错误响应时,控制单元200将用于SLBA=0的读取命令传输到与逻辑插槽1之外的逻辑插槽0以及逻辑插槽2和3连接的存储单元2A、存储单元2C和存储单元2D,通过异或从接收的数据D(0)、D(2)和P(0,2)恢复D(1),并且将恢复后的D(1)传输到客户机1002。能够在D(1)的读取期间并行地执行用于D(0)、D(2)和P(0,2)的数据读取。

[0437] 例如,在从客户机1002接收用于ALBA=1的写入命令和写入数据时,控制单元200指定逻辑插槽号=1和SLBA=0(它们是D(1)的存储目的地),将用于SLBA=0的写入命令传输到具有逻辑插槽号=1的存储单元2B,写入要写入的数据,从存储单元2A和存储单元2C(它们是连接到逻辑插槽号1之外的插槽并且没有存储在SLBA=0中的奇偶校验数据的存储单元)的SLBA=1读取数据D(0)和D(2),从D(0)、D(1)和D(2)计算奇偶校验数据P(0,2),并且将奇偶校验数据P(0,2)写入存储单元2D的SLBA=0中。

[0438] 在图60中,示出紧接在逻辑插槽1的迁移状态转变之后的阵列LBA和存储单元LBA的变换方法。紧接在迁移状态转变之后,阵列LBA和存储单元LBA的变换方法与图59中所示的正常状态下的变换方法相同。

[0439] 如图58中所示,存储单元2A至2D和存储单元3分别包括状态管理区域510和逻辑插槽号管理区域530。存储单元3包括数据迁移日志区域550。当完成下面解释的数据迁移时,控制单元200能够删除或解除分配数据迁移日志区域550,并且将数据迁移日志区域550分配用作诸如用户数据存储之类的另一种用途。控制单元200能够将LBA区域分配给状态管理区域510、逻辑插槽号管理区域530和数据迁移日志区域550。备选地,控制单元200能够将并非LBA区域的区域(例如,被分配有日志页面地址的日志区域)分配给状态管理区域510、逻辑插槽号管理区域530和数据迁移日志区域550。在并非LBA区域的区域中,例如根据非专利文献1的ACS-3中描述的“2Fh读取日志Ext”执行读取,并且根据非专利文献1的ACS-3中描述的“3Fh写入日志Ext”执行写入。下面解释的数据迁移状态的日志存储在数据迁移日志区域550中。

[0440] 作为记录在数据迁移日志区域550中的日志数据,在此实施例中,使用图61中所示的表数据。当在数据迁移状态下将扇区计数X的数据写入存储单元3的LBA(=SLBA)中时,控制单元200另外将(SLBA,X)写入数据迁移日志区域550中。当与LBA=SLBA~SLBA+X-1的区域重叠或连续的LBA已经记录在数据迁移日志区域550中时,希望将通过组合LBA而获得的日志记录在数据迁移日志区域550中,并且将组合后的日志从数据迁移日志区域550中删除。例如,当重新将(SLBA,X)的LBA区域A写入数据迁移日志区域550中时以及当(SLBA-a,a)的LBA区域B已经记录在数据迁移日志区域550中时,因为LBA区域A和LBA区域B是连续的,所以可通过使用指示区域A+B的数据(SLBA-a,a+X)盖写日志(SLBA-a,a)来更新数据迁移日志区域550,而不增加日志数据量。

[0441] 以与第一实施例中的图24和29中所示相同的方式,控制单元200判定连接到存储接口1007的相应存储单元是已经达到寿命结束、将要达到寿命结束还是将要出故障。当存

储单元已经达到寿命结束、将要达到寿命结束或将要出故障时,控制单元200执行寿命结束处理。如在第一实施例中,每隔固定时间、在每固定数量类型的处理下或者在每个固定数据传输和接收下(图24中所示),或者在从存储单元接收的命令响应是错误响应(如图29和30中所示)时,执行寿命结束确定。

[0442] 图62示出例如在控制单元200确定存储单元2B已经达到寿命结束时执行的存储单元2B的寿命结束处理。当控制单元200确定存储单元2B已经达到寿命结束时(步骤S500),控制单元200将存储单元2B的状态从1重写到2(步骤S501)。希望控制单元200通过显示设备、LED等通知网络管理员将新存储单元连接到存储接口1007的空闲插槽(步骤S502)。备选地,当自动地执行存储单元2A至2D或存储单元3与存储接口1007的物理附和分离的机械装置作为存储加载/卸载装置(图中未示出)安装在信息处理系统1上时,控制单元200能够控制所述存储加载/卸载装置以便将新存储单元连接到存储接口1007(步骤S502)。

[0443] 当存储单元3作为新存储单元连接时(步骤S503),控制程序200将存储单元3的状态510重写到4(步骤S504),并且将存储单元2B的逻辑插槽号管理区域530的数据复制到存储单元3的逻辑插槽号管理区域530(步骤S505)。在本实例中,因为将存储单元2B分配为逻辑插槽1(如图57中所示),所以将1写入存储单元3的逻辑插槽号管理区域530中。控制程序200将存储单元2B的状态510重写到3(步骤S506),并且导致RAID控制器1005将存储单元2B和存储单元3识别为逻辑插槽1,逻辑插槽1是相同逻辑插槽(步骤S507)。在寿命结束处理之后,存储单元2B的状态为3并且存储单元3的状态为4。逻辑插槽1转变为数据迁移状态。

[0444] 在此实施例中,相应存储单元的状态以非易失性方式存储在存储单元的状态管理区域510中。图63示出在RAID控制器1005启动时以及在存储单元连接到存储接口1007时控制单元200执行的处理程序。每当RAID控制器1005启动时,控制单元200读取状态管理区域510以便识别存储单元的状态。控制单元200识别存储单元的状态,并且从逻辑插槽号管理区域530读取存储单元的逻辑插槽号以便判定逻辑插槽0至3中的每一个是否处于数据迁移状态。

[0445] 即,当存储状态=0时(步骤S511),控制单元200将存储单元识别为处于初始存储状态(步骤S512)。当存储状态=1时(步骤S513),控制单元200将存储单元识别为处于正常状态(步骤S514)。当存储状态=2时(步骤S515),控制单元200将存储单元识别为处于低可靠性状态(步骤S516)。当存储状态=3时(步骤S517),控制单元200将存储单元识别为在数据迁移工作期间处于数据迁移源状态(受保护状态)(步骤S518)。当存储状态=4时(步骤S519),控制单元200将存储单元识别为在数据迁移工作期间处于数据迁移目的地状态(步骤S520)。当存储状态=5时(步骤S521),控制单元200将存储单元识别为处于丢弃目标状态(步骤S522)。当存储状态为0至5之外的值时,控制单元200将存储单元视为未授权的存储单元(步骤S523)。

[0446] 从逻辑驱动器(1)的读取

[0447] 图64示出在将读取请求从客户机1002传输到存储阵列设备1003时控制单元200执行的处理程序。控制单元200从客户机1002接收用于ALBA=cALBA(其是读取目标阵列LBA)的读取命令(步骤S530)。控制单元200从cALBA计算逻辑插槽号cSLOT和SLBA=cSLBA(其是读取目标存储单元LBA)(步骤S531)。控制单元200判定cSLOT的逻辑单元的存储单元是否出故障(步骤S532)。当cSLOT的逻辑插槽的存储单元并非出故障时,随后控制单元200判定

cSLOT的逻辑插槽的存储单元是否处于数据迁移状态(步骤S533)。

[0448] 当cSLOT的逻辑插槽的存储单元处于数据迁移状态时(步骤S533),控制单元200从cSLOT的插槽之外的插槽读取包括奇偶校验的数据,使用此数据恢复cSLOT的cSLBA的数据,并且将恢复后的数据传输到客户机1002(步骤S534)。控制单元200将cSLOT的cSLBA的恢复后的数据写回到数据迁移目的地存储单元3,并且将数据迁移日志记录在数据迁移日志区域550中(步骤S535)。当cSLOT的逻辑插槽的存储单元并未执行数据迁移时(步骤S533),控制单元200从cSLOT的存储单元读取数据D(cSLBA)并且将读取的数据传输到客户机1002。这样,当读取目标逻辑插槽处于数据迁移状态时,从读取目标之外的插槽恢复读取目标数据。因此,即使处于数据迁移状态的存储单元的数据丢失,也可弥补此损失。进一步,减少了从处于数据迁移状态的存储单元的读取,以便抑制读取干扰(其中存储的数据中出现错误的现象,因为非常小的电荷存储在从中读取数据的存储单元的同一块中包括的未选择的存储单元的浮置栅极中)。

[0449] 当cSLOT的逻辑插槽的存储单元出故障时,控制单元200判定是否存在处于数据迁移状态的插槽(步骤S537)。当不存在处于数据迁移状态的插槽时,控制单元200从cSLOT的插槽之外的插槽读取数据,使用此数据恢复cSLOT的cSLBA的数据,并且将恢复后的数据传输到客户机1002(步骤S538)。当读取目标插槽出故障并且这样其它插槽中不存在处于数据迁移状态的插槽时,控制单元200从读取目标之外的插槽恢复读取目标数据。

[0450] 当cSLOT的逻辑插槽的存储单元出故障并且存在处于数据迁移状态的插槽时(步骤S537),控制单元200读取数据迁移日志区域550并且判定读取目标SLBA的数据存在于数据迁移源和数据迁移目的地中的哪一个中(步骤S539)。控制单元200从读取自数据迁移目的地存储单元、数据迁移源存储单元和正常状态存储单元的数据来恢复cSLOT的cSLBA的数据,并且将恢复后的数据传输到客户机1002(步骤S540)。当使用数据迁移源的数据时,控制单元200将所使用的数据迁移源数据写回到数据迁移目的地存储单元3,并且将数据迁移日志记录在数据迁移目的地数据迁移日志区域550中(步骤S541)。

[0451] 从逻辑驱动器(2)的读取

[0452] 图65示出在将读取请求从客户机1002传输到存储阵列设备1003时控制单元200执行的另一个处理程序。在图65中所示的处理程序中,将在图64中的步骤S533处的判定是肯定时执行的从步骤S534和S535更改为步骤S550至S554。在图65中,删除图64中的步骤S541处的处理。当cSLOT的逻辑插槽的存储单元处于数据迁移状态时(步骤S533),控制单元200从数据迁移目的地存储单元3读取数据迁移日志并且判定cSLBA的数据是否包括在数据迁移日志中(步骤S551)。当cSLBA的数据包括在数据迁移日志中时,控制单元200从数据迁移目的地存储单元3读取cSLBA的数据,并且将读取的数据传输到客户机1002(步骤S554)。当cSLBA的数据未包括在数据迁移日志中时,控制单元200从数据迁移源存储单元读取cSLBA的数据,并且将读取的数据传输到客户机1002(步骤S552)。控制单元200将从数据迁移源读取的数据写回到数据迁移目的地存储单元3,并且更新数据迁移目的地存储单元3的日志550(步骤S553)。

[0453] 逻辑驱动器中的写入

[0454] 图66是用于解释用于此实施例中来客户机1002的写入命令的处理的流程图。在从客户机1002接收包括cLBA(其是阵列LBA)的写入命令和扇区长度时(步骤S560),控制单

元200从客户机1002接收写入数据(步骤S561)。控制单元200从cALBA计算其中应该写入数据的逻辑插槽号cSLOT,从cALBA检索cSLBA(它是其中应该写入数据的存储单元LBA),并且从cALBA计算cSLBA的奇偶校验数据存储目的地的逻辑插槽号cPSLOT(步骤S562)。控制单元200并行地读取所有插槽的cSLBA的数据。当存在故障插槽时,控制单元200从故障插槽之外的插槽读取数据,并且恢复故障插槽的数据(步骤S563)。此后,控制单元200并行地处理主体数据写入任务和奇偶校验数据写入任务。

[0455] 如下面所解释来执行主体数据写入任务。控制单元200判定cSLOT是否出故障(步骤S564)。当cSLOT出故障时,控制单元200结束任务而不写入数据。当cSLOT并未出故障时,控制单元200判定cSLOT是否处于数据迁移状态(步骤S565)。当cSLOT未处于数据迁移状态时,控制单元200将来自客户机1002的接收数据写入插槽cSLOT的cSLBA中(步骤S566)。当cSLOT处于数据迁移状态时,控制单元200将接收数据写入分配给插槽cSLOT的数据迁移源存储单元和数据迁移目的地存储单元中的数据迁移目的地存储单元的cSLBA中(步骤S567)。控制单元200将cSLBA和扇区大小作为数据迁移日志记录在数据迁移目的地存储单元的数据迁移日志区域550中(步骤S568)。当在cSLBA之前和之后的连续LBA和重叠LBA的数据迁移日志存在于数据迁移日志区域550中时,控制单元200将通过组合这些日志而获得的数据迁移日志写入数据迁移日志550中,并且删除组合源数据迁移日志(使组合源数据迁移日志无效)(步骤S568)。

[0456] 数据迁移日志从数据迁移日志550中的删除只需是逻辑删除。数据迁移日志不必通过NAND存储器16的块擦除来物理地擦除。例如,将用于使日志无效的标志写入数据迁移日志550中,或者将更新之后的数据迁移日志550存储在不同于更新之前的数据迁移日志550的区域中,并且更新指示数据迁移日志550的存储位置的指针,由此在逻辑上将数据迁移日志从数据迁移日志550中删除。

[0457] 如下面所解释来执行奇偶校验数据写入任务。控制单元200使用从客户机1002接收的写入数据盖写从所有逻辑插槽读取到RAID控制器1005中的存储器的数据,并且重新计算奇偶校验数据(步骤S570)。控制单元200判定cPSLOT是否出故障(步骤S571)。当cPSLOT出故障时,控制单元200结束任务而不写入奇偶校验数据。当cPSLOT并未出故障时,控制单元200判定cPSLOT是否处于数据迁移状态(步骤S572)。当cPSLOT未处于数据迁移状态时,控制单元200将奇偶校验数据写入插槽cPSLOT的cSLBA中(步骤S573)。当cPSLOT处于数据迁移状态时,控制单元200将奇偶校验数据写入分配给插槽cPSLOT的数据迁移源存储单元和数据迁移目的地源单元中的数据迁移目的地存储单元的cSLBA中(步骤S574)。控制单元200将cSLBA和扇区大小作为数据迁移日志记录在数据迁移目的地存储单元的数据迁移日志区域550中。当在cSLBA之前和之后的连续LBA和重叠LBA的数据迁移日志存在于数据迁移日志550中时,控制单元200将通过组合这些数据迁移日志而获得的数据迁移日志写入数据迁移日志区域550中,并且将组合源数据迁移日志从数据迁移日志550中删除(步骤S575)。

[0458] 这样,控制单元200响应于来自客户机的写入请求而将SLBA的写入命令传输到分配给未处于数据迁移状态的逻辑插槽0、2和3的存储单元2A、2C和2D。另一方面,针对处于数据迁移下的逻辑插槽1,控制单元200并不将写入命令传输到数据迁移源存储单元2B,将写入命令传输到数据迁移目的地存储单元3,并且将数据迁移日志记录在存储单元3的数据迁移日志区域550中。

[0459] 更详细的实例在图67中示出。在图67中所示的实例中,当逻辑插槽1处于数据迁移状态时,将存储单元2B分配为数据迁移源,将存储单元3分配为数据迁移目的地,并且逻辑插槽0、2和3处于正常状态,从客户机1002传输用于将D(1)写入存储阵列设备1003的ALBA=1中、将D(7)写入ALBA=7中以及将D(16)写入ALBA=16中的请求。

[0460] 实例1

[0461] 在从客户机1002接收用于ALBA=1中的数据D(1)_新的写入命令时,控制单元200从ALBA=1指定SLBA=0和逻辑插槽1。当逻辑插槽1未处于数据迁移状态时,将D(1)_新写入存储单元2B的SLBA=0中。但是,在本实例中,逻辑插槽1处于数据迁移状态。因此,控制单元200将D(1)_新写入存储单元3中。控制单元200基于上面解释的读取操作从逻辑插槽0的存储单元2A读取D(0)并从逻辑插槽2的存储单元2C读取D(2),通过异或从D(0)、D(1)_新和D(2)计算新奇偶校验数据P(0,2)_新,并且将P(0,2)_新存储在逻辑插槽3的存储单元2D的SLBA=0中。为了改进处理速度,希望并行地从控制单元200将用于D(1)_新的写入命令传输到存储单元3,将用于D(0)的读取命令传输到存储单元2A,并且将用于D(2)的读取命令传输到存储单元2C。进一步,希望在完成用于D(0)和D(2)的读取时,计算P(0,2)_新并且将写入命令传输到存储单元2D,而不等待对D(1)的写入的完成。控制单元200将(SLBA,扇区计数)记录在存储单元3的数据迁移日志区域550中,此(SLBA,扇区计数)指示其中在存储单元3中执行写入的SLBA和扇区计数。例如,在此实例中,在存储单元3的SLBA=0的一个扇区的SLBA区域中执行写入。因此,附加地将(0,1)写入数据迁移日志区域550中。如上面所解释,当连续LBA区域和重叠LBA区域已经记录在数据迁移日志区域550中时,将通过合并这些LBA区域而获得的SLBA区域记录在数据迁移日志区域550中。当在未处于数据迁移状态的逻辑插槽中执行写入时,不执行数据迁移日志550的更新。

[0462] 实例2

[0463] 在从客户机1002接收用于ALBA=7的新数据D(7)_新的写入命令时,控制单元200从ALBA=7指定SLBA=2和逻辑插槽2。因为逻辑插槽2未处于数据迁移状态,所以将D(7)_新写入存储单元2B的SLBA=0中。控制单元200基于读取操作执行从逻辑插槽0和逻辑插槽3读取D(6)和D(8),计算奇偶校验数据P(6,8)_新,并且将奇偶校验数据P(6,8)_新存储在逻辑插槽1中。因为逻辑插槽1处于数据迁移状态,所以控制单元200将P(6,8)_新写入存储单元3中而不是存储单元2B中,并且将(SLBA,扇区计数)=(2,1)(其指示其中在存储单元3中执行写入的SLBA和扇区计数)记录在存储单元3的数据迁移日志区域550中。

[0464] 实例3

[0465] 在从客户机1002接收用于ALBA=16的新数据D(16)_新的写入命令时,控制单元200从ALBA=16指定SLBA=5和逻辑插槽1。在本实例中,因为逻辑插槽1处于数据迁移状态,所以控制单元200将D(16)_新写入存储单元3中。控制单元200基于读取操作从逻辑插槽0的存储单元2A读取D(15)并从逻辑插槽2的存储单元2D读取D(17),通过异或从D(15)、D(16)_新和D(17)计算新奇偶校验数据P(15,17)_新,并且将P(15,17)_新存储在逻辑插槽2的存储单元2C的SLBA=5中。控制单元200将(SLBA,扇区计数)记录在存储单元3的数据迁移日志区域550中,此(SLBA,扇区计数)指示其中在存储单元3中执行写入的SLBA和扇区计数。例如,在此实例中,在存储单元3的SLBA=5的一个扇区的LBA区域中执行写入。因此,附加地将(5,1)写入数据迁移日志区域550中。如上面所解释,当连续SLBA区域和重叠SLBA区域已经记录在数据迁

移日志区域550中时,将通过合并这些SLBA区域而获得的SLBA区域记录在数据迁移日志区域550中。当在未处于数据迁移状态的插槽中执行写入时,不执行数据迁移日志550的更新。这样,如在第二实施例中,每当从客户机1002接收写入请求时,将对应于写入请求的SLBA的数据写入数据迁移目的地存储单元3中而不是数据迁移源存储单元2B中,并且记录数据迁移日志。因此,可同时执行新数据写入和数据迁移操作。

[0466] 后台备份

[0467] 在闲置期间,例如当在固定时间或更多时间内未从客户机1002接收命令时,或者当从客户机1002接收待机模式转变请求时,控制单元200执行从存储单元2到存储单元3的后台备份。控制单元200读取存储单元3的数据迁移日志区域550,执行从存储单元2B到未记录在数据迁移日志区域550中的SLBA的读取,将数据写入存储单元3中,并且将写入数据的SLBA和扇区大小存储在存储单元3的数据迁移日志区域550中以便执行后台备份。在将SLBA存储在数据迁移日志550中,当存在连续SLBA区域时或者当存在重叠SLBA区域时,希望将通过合并这些SLBA区域而获得的SLBA区域存储在数据迁移日志550中,并且将合并之前的SLBA区域从数据迁移日志区域550中删除。

[0468] 为了减小存储在数据迁移日志区域550中的数据迁移日志的数据大小,并且为了以高速度执行下面解释的数据迁移结束确定,希望针对SLBA区域中除了登记在数据迁移日志550中的SLBA区域之外的未迁移的碎片化SLBA区域优先地执行后台备份。控制程序200备份未迁移的碎片化SLBA区域,并且使未迁移的碎片化SLBA区域经受数据迁移以便从而将新近迁移的SLBA区域存储在数据迁移日志550中,作为通过将新近迁移的SLBA区域与连续地位于新近迁移的SLBA区域之前和之后的迁移的SLBA区域相合并而形成的SLBA区域。

[0469] 数据迁移状态的监视

[0470] 控制单元200读取数据迁移日志区域550以便监视处于数据迁移下的逻辑插槽的数据迁移状态。图68示出用于使用数据迁移日志监视数据迁移状态的监视程序。例如,每当经过预定时间时,控制单元200从数据迁移日志区域550读取数据迁移日志(步骤S600和S601)。当所有迁移目标SLBA包括在数据迁移日志550中时,控制单元200确定数据迁移完成。例如,当数据迁移源存储单元2B的所有SLBA被包括在数据迁移日志550中时,控制单元200确定数据迁移完成(步骤S602)。作为数据迁移完成的确定,例如控制单元200能够判定数据迁移目的地存储单元3的所有SLBA是否包括在数据迁移日志550中。

[0471] 当控制单元200确定数据迁移完成时,控制单元200将数据迁移源存储单元2的状态更改为丢弃目标状态5并且将数据迁移目的地存储单元3的状态更改为正常状态1(步骤S603),以便结束逻辑插槽1的数据迁移状态并且结束针对逻辑插槽1的数据迁移状态监视。就减少存储单元2B的功耗而言,希望控制单元200将到低功耗模式的转变请求传输到存储单元2。在数据迁移状态结束之后,希望控制单元200通过设置在端口附近的显示器9或LED通知用户或管理员将存储单元2B与接口19分离,或者控制存储加载/卸载装置以便将存储单元2B与接口19分离。

[0472] 整体状态的转变

[0473] 图69示出在存储单元2B已经达到寿命结束时出现的存储单元的状态转变。当所有存储单元2A至2D处于正常状态时,所有存储单元2A至2D的状态为1并且存储单元3未连接到存储接口1007(步骤1)。当控制单元200确定存储单元2B已经达到寿命结束时,控制单元200

将存储单元2B的状态更改为2(步骤2)。基于来自控制单元200的通知或命令,将存储单元3(其是新存储单元)连接到存储接口1007(步骤3)。当连接了存储单元3时,控制单元200将数据迁移源存储单元2B的状态更改为3并且将数据迁移目的地存储单元3的状态更改为4,以便将逻辑插槽1的状态转到数据迁移状态(步骤4)。当控制单元200基于从数据迁移日志区域550读取的信息而确定已将存储单元2B的所有有效数据迁移到存储单元3时,控制单元200将存储单元2B的状态更改为5并且将存储单元3的状态更改为1。此后,存储单元2A、存储单元3、存储单元2C和存储单元2D表现为如同存储单元2A、存储单元3、存储单元2C和存储单元2D是原始存储单元2A、原始存储单元2B、原始存储单元2C和原始存储单元2D那样(处理返回到步骤1)。此后,当进一步确定存储单元2A、存储单元3、存储单元2C和存储单元2D中的任意一个已经达到寿命结束时,重复相同步骤2至5。这样,甚至在存储阵列设备1003的存储单元中的任意一个已经达到寿命结束时,将要达到寿命结束时,或者将要出故障时,也可很容易地根据此实施例将存储单元的数据迁移到新存储单元。在此实施例中,因为数据迁移操作在后台执行,所以可抑制由于备份操作而导致的性能劣化。

[0474] 一个存储单元的数据迁移状态和另一个存储单元的故障同时出现

[0475] 在图70所示的实例中,一个存储单元的数据迁移状态和另一个存储单元的故障同时出现。在图70所示的实例中,当逻辑插槽1处于数据迁移状态时,将存储单元2B分配为数据迁移源,将存储单元3分配为数据迁移目的地,并且逻辑插槽0、2和3处于正常状态,在逻辑插槽3的存储单元2D中出现故障。

[0476] 在从SLBA=0读取的情况下,对于逻辑插槽0的D(0)和逻辑插槽2的D(2)的读取而言,数据恢复是不必要的。对逻辑插槽1的D(1)的读取可以通过从数据迁移目的地存储单元3读取作为最新数据的D(1)_新实现。在从SLBA=2读取的情况下,对于逻辑插槽0的D(6)和逻辑插槽2的D(7)的读取而言,数据恢复是不必要的。对逻辑插槽3的D(8)的读取可以通过以下操作实现:通过对来自数据迁移目的地存储单元3的D(6)和D(7)以及P(6,8)_新进行异或而恢复数据。在从SLBA=5读取的情况下,对于逻辑插槽0的D(15)的读取而言,数据恢复是不必要的。对逻辑插槽1的D(16)的读取可以通过从数据迁移目的地存储单元3读取作为最新数据的D(16)_新实现。对逻辑插槽3的D(17)的读取可以通过以下操作实现:通过对数据迁移目的地存储单元3的D(15)、D(16)_新和逻辑插槽2的P(15,17)_新进行异或而恢复数据。

[0477] 两个存储单元的数据迁移状态和一个存储单元的故障同时出现

[0478] 在图71中所示的实例中,两个存储单元的数据迁移状态和另一个存储单元的故障同时出现。在图71中所示的实例中,当逻辑插槽1和逻辑插槽2处于数据迁移状态时,将存储单元2B和2C分配为数据迁移源,将存储单元3B和3C分配为数据迁移目的地,并且逻辑插槽0和3处于正常状态,在逻辑插槽3的存储单元2D中出现故障。

[0479] 在从SLBA=0读取的情况下,对于逻辑插槽0的D(0)和逻辑插槽2的D(2)的读取而言,数据恢复是不必要的,对逻辑插槽1的D(1)的读取可以通过从数据迁移目的地存储单元3读取D(1)_新实现。在从SLBA=2读取的情况下,对于逻辑插槽0的D(6)的读取而言,数据恢复是不必要的。对逻辑插槽2的D(7)的读取可以通过从数据迁移目的地存储单元3C读取作为最新数据的D(7)_新实现。对逻辑插槽3的D(8)的读取可以通过以下操作实现:通过对D(6)、D(7)_新和P(6,8)_新进行异或而恢复数据。在从SLBA=5读取的情况下,对于逻辑插槽0的D(15)的读取而言,数据恢复是不必要的。对逻辑插槽1的D(16)的读取可以通过从数据迁移目的

地存储单元3B读取作为最新数据的D (16)_新实现。对逻辑插槽3的D (17) 的读取可以通过以下操作实现:通过对数据迁移目的地存储单元3B的D (15)、D (16)_新和逻辑插槽2的数据迁移目的地存储单元3C的P (15,17)_新进行异或而恢复数据。

[0480] 一个存储单元的数据迁移状态和无法进行数据读取错误同时出现

[0481] 在图72中所示的实例中,在数据迁移期间在一个存储单元中出现诸如不可纠正的ECC错误 (UECC错误) 之类的读取错误。

[0482] 实例1

[0483] 在接收用于ALBA=2的数据D (2) 的读取命令时,控制单元200从ALBA=2指定SLBA=0和逻辑单元2。当在从逻辑插槽2的存储单元2C读取数据D (2) 中出现UECC错误时,可以通过对数据迁移目的地存储单元3的D (1)_新和逻辑插槽3的P (0,2)_新进行异或而恢复D (2) 。

[0484] 实例2

[0485] 在从客户机1002接收用于ALBA=8的新数据D (8) 的读取命令时,控制单元200从ALBA=8指定SLBA=2和逻辑插槽3。当在从逻辑插槽3的存储单元2D读取数据D (8) 中出现UECC错误时,可以通过对存储单元3的D (6)、D (7)_新和P (6,8)_新进行异或而恢复D (8) 。

[0486] 实例3

[0487] 在从客户机1002接收用于ALBA=9的新数据D (9) 的读取命令时,控制单元200从ALBA=9指定SLBA=3和逻辑插槽1。当在从逻辑插槽1的存储单元2B读取数据D (9) 中出现UECC错误时,可以通过对D (10)、D (11) 和P (9,10) 进行异或而恢复D (9) 。

[0488] (第四实施例)

[0489] 在第四实施例中,将本发明应用于包括存储网络的信息处理系统1。在此实施例中,如图73中所示那样配置信息处理系统1。信息处理系统1能够如图74中所示那样配置。如图73中所示,此实施例中的信息处理系统1包括:存储单元2;存储单元3;一个或多个其它存储单元6004;一个或多个元数据服务器6003;服务器6001;存储网络1000,其被配置为将存储单元2、存储单元3、存储单元6004、元数据服务器6003和服务器6001彼此连接;一个或多个客户机6002;以及网络6000,其被配置为将服务器6001和客户机6002彼此连接。在图74的情况下,存储网络1000经由通道延伸器和长距离网络 (例如WAN) 连接到存储网络6001b。存储单元3连接到存储网络6001b。

[0490] 存储单元2和3包括状态管理区域510。存储单元2和3能够包括与第一实施例中相同的数据迁移目标管理区域2007和3007 (参见图32),或者能够包括与第二实施例中相同的逻辑驱动器ID管理区域520。如上面所解释,状态管理区域510分别指示对应于其的存储单元处于以下状态:

[0491] 0:初始存储状态

[0492] 1:正常状态

[0493] 2:低可靠性状态

[0494] 3:数据迁移源状态 (受保护状态)

[0495] 4:数据迁移目的地状态

[0496] 5:丢弃目标状态

[0497] 元数据服务器6003包括:主存储器6,其被配置为存储OS 100、控制单元200、元数据300和逻辑驱动器状态表450 (它们与第一实施例中相同);以及CPU 5,其是被配置为执行

OS 100和控制单元200的算术单元。元数据服务器6003起到相当于第一实施例中的信息处理设备111的OS 100和控制单元200的功能。存储在元数据服务器6003的主存储器6中的元数据300例如具有图33中所示的结构。主存储器6上的元数据和元数据的日记备份在元数据服务器6003中的非易失性存储单元、存储网络1000中的非易失性存储单元、存储单元2、存储单元3等中。逻辑驱动器状态表450例如具有图40中所示的结构。逻辑驱动器状态表450指示逻辑驱动器处于正常状态或数据迁移状态。例如,在从服务器6001接收文件ID时,元数据服务器6003从元数据区域300中检索文件ID,并且指定逻辑驱动器地址、存储单元标识名称(其是存储单元地址)、LBA、扇区计数等。作为存储单元标识名称(其是存储单元地址),例如能够使用IP(网际协议)地址、MAC(介质访问控制)地址或WWN(全球名称)。状态表510存储连接到存储网络1000的存储单元2和3的状态。

[0498] 逻辑驱动器状态表450能够如图41中所示那样构造。图75中所示的状态表650能够存储在元数据服务器6003的主存储器6中。逻辑驱动器状态表450不必存储在主存储器6中。在状态表650中,管理逻辑驱动器地址、存储单元标识名称、逻辑驱动器状态和存储状态。使用状态表650代替逻辑驱动器状态表450和状态管理区域510。然后,存储单元2和3中的状态管理区域510是不必要的。在图75中,逻辑驱动器地址B包括两个存储单元,它们的存储单元标识名称是b1和b2。这些存储单元处于数据迁移状态。具有存储单元标识名称b1的存储单元是数据迁移源。具有存储单元标识名称b2的存储单元是数据迁移目的地。

[0499] 存储网络1000是用于存储存取的网络。例如,使用光纤通道或以太网(注册商标)。具体地说,作为存储网络1000,例如使用SAN(存储区域网络)或NAS(网络附加存储)。作为SAN,例如使用FC-SAN(光纤通道存储区域网络)或IP-SAN(网际协议存储区域网络)。作为SAN的上层协议,例如使用SCSI(小型计算机系统接口)。在此实施例中解释的实例中,iSCSI用作上层协议。存储网络1000包括网络交换机6009和集线器(图中未示出)。

[0500] 网络6000是用于允许客户机存取服务器6001并且使用各种服务的网络。例如,使用光纤通道或以太网(注册商标)作为网络6000。例如,作为网络6000,存在WAN、LAN等。网络6000包括网络交换机(图中未示出)和集线器(图中未示出)。

[0501] 客户机1002是连接到网络6000的计算机,并且被配置为执行所需处理。客户机1002通常包括诸如处理器、主存储器、通信接口和本地输入/输出设备之类的硬件资源。进一步,客户机1002包括诸如设备驱动程序、操作系统(OS)和应用程序400(图中未示出)之类的软件资源。因此,客户机1002在处理器的控制下执行各种程序,并且与硬件资源协作实现处理。例如,客户机1002在处理器的控制下执行商业应用程序,以便从而通过网络6000对服务器6001进行I/O存取并且实现所需商业系统。

[0502] 服务器6001是连接到存储网络1000和网络6000的计算机,并且被配置为根据来自客户机6002的请求执行所需处理。服务器6001通常包括诸如处理器、主存储器、通信接口和本地输入/输出设备之类的硬件资源。进一步,服务器6001包括诸如设备驱动程序、操作系统(OS)和应用程序(图中未示出)之类的软件资源。因此,服务器6001在处理器的控制下执行各种程序,并且与硬件资源协作实现处理。例如,服务器6001响应于来自客户机6002的请求而在处理器的控制下执行应用服务程序,以便从而通过存储网络6000对存储单元2、存储单元3或存储单元6004进行I/O存取并且实现所需应用服务程序。

[0503] 例如,在从客户机6002接收文件数据读取请求和文件ID时,服务器6001将文件ID

传输到元数据服务器6003,接收其中存储来自元数据服务器6003的文件的存储单元的存储单元地址(例如IP地址、MAC地址或WWN)和LBA,将指定存储单元地址的分组传输到存储网络1000以便将读取命令传输到存储单元,从存储单元接收读取的数据,并且将读取的数据传输到客户机6002。例如,在从客户机6002接收文件数据写入请求和文件ID时,服务器6001将文件ID传输到元数据服务器6003,接收其中应该存储来自元数据服务器6003的文件的存储单元的存储单元地址和LBA,从客户机6002接收写入数据,并且将此数据传输到存储单元以便写入此数据。服务器6001能够是数据库服务器(DB服务器),它是数据库管理系统(DBMS)在其上操作的服务器。

[0504] 用于确定寿命结束的处理

[0505] 当启动时,控制单元200监视关于存储单元2的可靠性信息,例如统计信息65。例如,如图24中所示,控制单元200每隔固定时间(例如,每隔1分钟)或者在每固定次数的处理中,从存储单元2获得统计信息65。如在第一实施例中,控制单元200基于获得的可靠性信息而判定存储单元2是否已经达到寿命结束。当控制单元200确定存储单元2已经达到寿命结束时,控制单元200执行下面解释的寿命结束处理。

[0506] 寿命结束处理

[0507] 当连接的存储单元2已经达到寿命结束或将要达到寿命结束并且开始寿命结束处理时,在将存储单元2的状态510更改为存储状态=2(低可靠性状态)之后,控制单元200在元数据服务器6003、服务器6001和客户机6002的显示器9等上显示用于促使连接新存储单元的消息。当连接了新存储单元3时,控制单元200将状态表450中的存储单元3的状态510重写到存储状态=4(数据迁移目的地状态),并且进一步将存储单元2的状态510重写到存储状态=3(受保护状态)。控制单元200导致OS 100将存储单元2和存储单元3识别为一个逻辑驱动器4。控制单元200将存储在主存储器6上的逻辑驱动器状态表450或状态表650中的逻辑驱动器4的状态从“正常状态”重写到“数据迁移状态”。

[0508] 从处于数据迁移状态的逻辑驱动器的读取

[0509] 在从客户机6002接收读取请求和文件ID时,服务器6001将文件ID传输到元数据服务器6003。元数据服务器6003指定对应于来自元数据300的文件ID的逻辑驱动器4,从主存储器6读取逻辑驱动器状态表450和状态表650,并且识别所指定的逻辑驱动器4的状态是数据迁移状态。元数据服务器6003从元数据300获得其中存储由文件ID指定的文件的存储单元2或3的存储单元地址和LBA,并且将获得的存储单元地址和获得的LBA传输到服务器6001。服务器6001将指定所接收的存储单元地址和所接收的LBA的分组传输到存储网络1000以便将读取命令传输到存储单元2或3,从存储单元接收读取的数据,并且将读取的数据传输到客户机6002。

[0510] 逻辑驱动器中的写入

[0511] 例如,在从客户机6002接收文件数据写入请求和文件ID时,服务器6001将文件ID传输到元数据服务器6003。元数据服务器6003从逻辑驱动器状态表450或状态表650确定逻辑驱动器4的状态,并且将逻辑驱动器4识别为处于正常状态。然后,元数据服务器6003从主存储器6读取元数据300,并且参考元数据300分配用于数据写入的LBA。元数据服务器6003将此LBA以及存储单元2的存储单元地址和LBA传输到服务器6001。服务器6001将指定所接收的存储单元地址和所接收的LBA的分组传输到存储网络1000以便将写入命令传输到存储

单元2,并且将写入数据存储在存储单元2中。控制单元200重写元数据300,并且将写入数据的LBA和扇区计数映射到存储单元2和写入文件ID。

[0512] 当控制单元200将逻辑驱动器4识别为处于数据迁移状态时,控制单元200从主存储器6读取元数据300并且参考元数据300分配用于数据写入的LBA。元数据服务器6003将此LBA以及存储单元3的存储单元地址和LBA传输到服务器6001。服务器6001将指定存储单元3的所接收的存储单元地址和所接收的LBA的写入请求分组传输到存储网络1000以便将写入命令传输到存储单元3,并且将写入数据存储在存储单元3中。控制单元200更新元数据300,使从文件ID到存储单元2的映射无效,并且将文件ID映射到存储单元3、写入的LBA和扇区计数以便使用存储单元3中的写入实现从存储单元2到存储单元3的数据迁移。

[0513] 后台备份

[0514] 当逻辑驱动器4处于逻辑驱动器状态表450中的数据迁移状态的状态时,当几乎不发生由客户机6002进行的对逻辑驱动器4的存取时(在闲置期间),控制单元200能够执行从数据迁移源存储单元2到数据迁移目的地存储单元3的后台备份。控制单元200从主存储器6读取元数据300,并且搜索映射到存储单元2的文件ID。如果存在映射到存储单元2的文件,则控制单元200经由服务器6001将读取命令传输到存储单元2以便执行从文件的LBA的读取,并且接收读取的数据。控制单元200将写入命令和读取的数据传输到存储单元3的LBA,执行写入,重写主存储器6上的元数据300,使从文件ID到存储单元2的映射无效,并且将文件ID映射到存储单元3。

[0515] 能够采用无服务器备份(Server-free backup或server-less backup)作为控制单元200的后台备份操作。在这种情况下,例如能够使用扩展复制命令。作为扩展复制命令,例如能够使用SCSI初级命令-4 (SPC-4) (INCITS T10/1731-D,修订36e (<http://www.t10.org/>))中描述的“83h扩展复制(83h EXTENDED COPY)”命令。控制单元200将包括存储单元3的备份目标LBA和地址的扩展复制命令传输到存储单元2。然后,存储单元2从此LBA读取数据,并且将读取的数据传输到存储单元3。存储单元3将接收数据写入LBA中。

[0516] 数据迁移完成时间

[0517] 在逻辑驱动器状态表450中,当逻辑驱动器4的状态为“数据迁移状态”时,控制单元200定期地读取主存储器6上的元数据300,并且定期地检查是否存在映射到存储单元2的迁移目标文件ID。例如,控制单元200定期地检查存储在逻辑驱动器4中的所有文件的文件ID中是否存在映射到存储单元2的迁移目标文件ID。当不存在迁移目标文件ID时,控制单元200将数据迁移目的地存储单元3的状态510重写到存储状态=1(正常状态),并且将数据迁移源存储单元2的状态510重写到存储状态=5(丢弃目标状态)。控制单元200将存储单元2与逻辑驱动器4分离,将存储单元3识别为逻辑驱动器4,并且将逻辑驱动器状态表450或状态表650中的逻辑驱动器4的状态从“数据迁移状态”重写到“正常状态”。

[0518] 如上面所解释,在数据迁移状态下,在逻辑驱动器4中的数据写入中,不将写入命令传输到数据迁移源存储单元2而将写入命令传输到数据迁移目的地存储单元3。从存储单元2或3执行从逻辑驱动器4的数据读取。在逻辑驱动器4中,每当从客户机6002接收写入请求时,逐渐地将存储在存储单元2中的有效数据迁移到存储单元3。新数据写入操作也用作数据迁移操作。

[0519] (第五实施例)

[0520] 在第五实施例中,将本发明应用于包括多个数据中心(DC)和一个长距离网络(其被配置为连接数据中心(DC))的信息处理系统1。在此实施例中,如图76中所示那样配置信息处理系统1。在此实施例中,信息处理系统1包括:数据迁移源数据中心3002(存储单元3002);数据迁移目的地数据中心3003(存储单元3003);数据中心3005,其是另一个数据中心;服务器3006;数据中心管理服务器3007;以及长距离网络3000,其被配置为连接这些设备。在此实施例中,数据中心管理服务器3007将数据中心3002识别为充当逻辑单元的逻辑数据中心3004。当数据中心3002的可靠性劣化时,数据中心管理服务器3007将数据中心3002和数据中心3003识别为逻辑数据中心3004。

[0521] 数据中心3002和3003包括状态管理区域510。存储单元3002和3003能够包括数据迁移目标管理区域2007和3007,它们对应于第一实施例中的数据迁移目标管理区域2007和3007(参见图32)。存储单元3002和3003能够包括逻辑数据中心ID管理区域520,它们对应于第二实施例中的逻辑驱动器ID管理区域520。如上面所解释,状态管理区域510指示对应于其的存储单元处于以下状态:

[0522] 0:初始存储状态

[0523] 1:正常状态

[0524] 2:低可靠性状态

[0525] 3:数据迁移源状态(受保护状态)

[0526] 4:数据迁移目的地状态

[0527] 5:丢弃目标状态

[0528] 数据中心管理服务器3007包括:主存储器6,其存储OS 100、控制单元200、元数据300和逻辑数据中心状态表450(它们与第一实施例中相同);以及CPU 5,其是被配置为执行OS 100和控制单元200的算术单元。数据中心管理服务器3007起到相当于第一实施例中的信息处理设备111的OS 100和控制单元200的功能。存储在数据中心管理服务器3007的主存储器6中的元数据300例如具有图33中所示的结构。主存储器6上的元数据和元数据的日记被备份在长距离网络3000中的数据中心管理服务器3007以及数据中心3002、3003和3005中的非易失性存储单元中。逻辑数据中心状态表450例如具有图41中所示的结构。逻辑数据中心状态表450指示逻辑数据中心3004处于正常状态或数据迁移状态。例如,在从服务器3006接收文件ID时,数据中心管理服务器3007从元数据区域300中检索文件ID,并且指定对应于文件ID的逻辑数据中心ID、数据中心ID、LBA、扇区计数等。状态表510存储数据中心3002和3003的状态。

[0529] 除了以图41中所示的结构配置逻辑数据中心状态表450之外,还能够采用图75中所示的状态表650作为数据中心管理服务器3007。在状态表650中,管理逻辑数据中心ID、数据中心ID、逻辑数据中心状态和数据中心状态。在这种情况下,存储单元2和3中的状态管理区域510是不必要的。控制单元200能够使用如第二实施例中的数据迁移日志存储区域(代替元数据区域300)来管理数据从数据中心3002到数据中心3003的迁移。

[0530] 用于确定寿命结束的处理

[0531] 当启动时,控制单元200监视关于数据中心3002的可靠性信息。例如,如图24中所示,控制单元200每隔固定时间(例如,每隔1分钟)或者在每固定次数的处理中,从数据中心3002获得统计信息65。如在第一实施例中那样,控制单元200基于获得的统计信息65而判定

数据中心3002是否已经达到寿命结束。当控制单元200确定数据中心3002已经达到寿命结束时,控制单元200执行下面解释的寿命结束处理。

[0532] 寿命结束处理

[0533] 当连接的数据中心3002已经达到寿命结束或将要达到寿命结束并且开始寿命结束处理时,在将数据中心3002的状态510更改为数据中心状态=2 (低可靠性状态) 之后,控制单元200在数据中心管理服务器3007的显示器等上显示用于促使连接新数据中心的消息。当连接了新数据中心3003时,控制单元200将状态表450中的数据中心3003的状态510重写到数据中心状态=4 (数据迁移目的地状态),并且进一步将数据中心3002的状态510重写到数据中心状态=3 (受保护状态)。控制单元200导致OS 100将数据中心3002和数据中心3003识别为一个逻辑数据中心3004。控制单元200将主存储器6上的逻辑数据中心状态表450的状态从“正常状态”重写到“数据迁移状态”。

[0534] (从处于数据迁移状态的逻辑数据中心的读取)

[0535] 在从服务器3006接收读取请求和文件ID时,数据中心管理服务器3007指定对应于来自元数据300的文件ID的逻辑数据中心3004,从主存储器6读取逻辑数据中心状态表450,并且识别指定的逻辑数据中心3004的状态是数据迁移状态。数据中心管理服务器3007从元数据300获得其中存储由文件ID指定的文件的数据中心3002或3003的数据中心地址和LBA,并且将所获得的数据中心地址和所获得的LBA传输到服务器3006。服务器3006将指定所接收的数据中心地址和所接收的LBA的分组传输到网络3000以便将读取命令传输到数据中心3002或3003,并且从数据中心3002或3003接收读取的数据。

[0536] 逻辑数据中心中的写入

[0537] 在从服务器3006接收写入请求和文件ID时,数据中心管理服务器3007从逻辑数据中心状态表450确定逻辑数据中心3004的状态,并且将逻辑数据中心3004识别为处于正常状态。然后,数据中心管理服务器3007从主存储器6读取元数据300,并且参考元数据300分配用于数据写入的LBA。数据中心管理服务器3007将此LBA以及数据中心3002的数据中心ID和LBA传输到服务器3006。服务器3006将指定所接收的数据中心ID和所接收的LBA的分组传输到网络3000以便将写入命令传输到数据中心3002,并且将写入数据存储在数据中心3002中。

[0538] 当数据中心管理服务器3007将逻辑数据中心3004识别为处于数据迁移状态时,数据中心管理服务器3007从主存储器6读取元数据300,并且参考元数据300分配用于数据写入的LBA。数据中心管理服务器3007将此LBA以及数据中心3003的数据中心ID和LBA传输到服务器3006。服务器3006将指定数据中心3003的所接收的数据中心ID和所接收的LBA的分组传输到网络3000以便将写入命令传输到数据中心3003,并且将写入数据存储在数据中心3003中。

[0539] 后台备份

[0540] 当逻辑数据中心3004处于逻辑驱动器状态表450中的数据迁移状态的状态时,当几乎不发生由服务器3006进行的对逻辑数据中心3004的存取时(在闲置期间),控制单元200能够执行从数据迁移源数据中心3002到数据迁移目的地数据中心3003的后台备份。控制单元200从主存储器6读取元数据300,并且搜索映射到数据中心3002的文件ID。如果存在映射到数据中心3002的文件,则控制单元200经由服务器3006将读取命令传输到数据中心

3002以便执行从文件的LBA的读取,并且从数据中心3002接收读取的数据。控制单元200将写入命令和读取的数据传输到数据中心3002的LBA,执行写入,重写主存储器6上的元数据300,并且将文件ID映射到数据中心3002。

[0541] 能够采用无服务器备份作为控制单元200的后台备份操作。在这种情况下,例如能够使用扩展复制命令。作为扩展复制命令,例如能够使用SCSI初级命令-4 (SPC-4) (INCITS T10/1731-D,修订36e (<http://www.t10.org/>)) 中描述的“83h扩展复制”命令。控制单元200将包括数据中心3003的备份目标LBA和ID的扩展复制命令传输到数据中心3002。然后,数据中心3002从此LBA读取数据,并且将读取的数据传输到数据中心3003。数据中心3003将接收数据写入LBA中。

[0542] 数据迁移完成时间

[0543] 在逻辑驱动器状态表450中,当逻辑数据中心3004的状态为“数据迁移状态”时,控制单元200定期地读取主存储器6上的元数据300,并且定期地检查是否存在映射到数据中心3002的迁移目标文件ID。例如,控制单元200定期地检查存储在逻辑数据中心3004中的所有文件的文件ID中是否存在映射到数据中心3002的迁移目标文件ID。当不存在迁移目标文件ID时,控制单元200将数据迁移目的地数据中心3003的状态510重写到数据中心状态=1 (正常状态),并且将数据迁移源数据中心3002的状态510重写到数据中心状态=5 (丢弃目标状态)。控制单元200将数据中心3002与逻辑数据中心3004分离,将数据中心3003识别为逻辑数据中心3004,并且在逻辑驱动器状态表450中将逻辑数据中心3004的状态从“数据迁移状态”重写到“正常状态”。

[0544] 如上面所解释,在数据迁移期间,在逻辑数据中心3004中的数据写入中,不将写入命令传输到数据迁移源数据中心3002而将写入命令传输到数据迁移目的地数据中心3003。从数据中心3002或3003执行从逻辑数据中心3004的数据读取。在逻辑数据中心3004中,每当从服务器3006接收写入请求时,逐渐地将存储在数据中心3002中的有效数据迁移到数据中心3003。新数据写入操作也用作数据迁移操作。

[0545] (第六实施例)

[0546] 中继单元

[0547] 在第一和第二实施例中,当控制单元200确定存储单元2已经达到寿命结束、将要达到寿命结束或将要出故障时,或者确定存储单元2的可靠性劣化时,控制单元200处理写入请求以便防止写入存储单元2。因此,可抑制存储单元2的可靠性的进一步劣化,并且防止新写入数据的损失。为了安全地限制存储单元2中的数据写入,希望防止信息处理设备、客户机和服务器自发地执行存储单元中的写入。在此实施例中,信息处理系统包括中继单元5000。中继单元5000中的控制单元200监视诸如存储单元2的统计信息之类的可靠性信息。当中继单元5000确定存储单元2已经达到寿命结束、将要达到寿命结束或将要出故障或者存储单元2的可靠性劣化时,控制单元200向信息处理设备111、客户机和服务器通知指示存储单元2是只读设备的信息,以便从而防止信息处理设备111、客户机和服务器自发地执行存储单元2中的写入。可以独立地执行此实施例。另一方面,当将此实施例与第一实施例、第二实施例等组合时,可更可靠地抑制可靠性劣化的存储单元2中的写入。例如,当将此实施例与第一实施例、第二实施例等组合时,中继单元5000传输以下存储单元信息:此信息指示存储单元2是仅支持读取和写入中的读取的存储单元。CPU 5和存储在主存储器6中的控制

单元200将存储单元2识别为只读存储单元,并且将存储单元2识别为处于受保护状态(数据迁移源状态)。

[0548] 图77是其中将中继单元5000安装在诸如台式个人计算机或笔记本个人计算机之类的信息处理系统1上的实例。中继单元5000能够安装在信息处理设备111内部,或者能够安装在信息处理设备111外部。中继单元5000经由接口19连接到存储单元2,并且经由接口5001和芯片组7连接到CPU 5。中继单元5000能够不经由芯片组7而直接连接到CPU 5。中继单元5000能够嵌入在芯片组7中。

[0549] 中继单元5000包括控制单元200。如图77中所示,整个控制单元200包括在中继单元5000中。如图78中所示,控制单元200的一部分能够包括在中继单元5000中,并且控制单元200的一部分能够包括在主存储器6中。控制单元200的一部分能够包括在中继单元5000中,并且控制单元200的一部分能够包括在信息处理设备111中的另一个存储部分(例如ROM 11)中。控制单元200能够以固件或软件的形式实现,或者能够以硬件的形式实现。

[0550] 作为接口19和接口5001,例如能够使用SATA(串行高级技术附件)、快速PCI(快速外围组件互连,PCIe)、USB(通用串行总线)、SAS(串行连接SCSI)、雷电(注册商标)、以太网(注册商标)和光纤通道。接口19和接口5001能够是相同标准的接口,或者能够是不同标准的接口。在此实施例中,接口19和接口5000是SATA接口。

[0551] 以与第一实施例中的图24和29中所示相同的方式,控制单元200判定存储单元2已经达到寿命结束、将要达到寿命结束还是将要出故障。当控制单元200确定存储单元2已经达到寿命结束、将要达到寿命结束或将要出故障时,如图79中所示,控制单元200从正常状态转变为低可靠性状态以便执行寿命结束处理(步骤S800)。正常状态和低可靠性状态是控制单元200的模式。当存储单元2正常时,控制单元200在正常状态下操作。当控制单元200判定存储单元2已经达到寿命结束、将要达到寿命结束还是将要出故障时,或者存储单元2的可靠性劣化时,控制单元200在低可靠性状态下操作。如在第一实施例中,每隔固定时间、在每固定次数的处理下或者在每个固定数据的传输和接收下(如图24中所示),或者在从存储单元2接收的命令响应是错误响应(如图29和30中所示)时,执行寿命结束确定。

[0552] 控制单元200将经由接口5001从CPU 5接收的命令和数据通过接口19传输到存储单元2。控制单元200将通过接口19从存储单元2接收的响应和数据通过接口5001传输到CPU 5。当接口5001和接口19的协议不同时,在执行协议转换之后,控制单元200将转换之后的命令、响应和数据传输到CPU 5和存储单元2。当将存储单元2的存储单元信息传输到CPU 5时,控制单元200根据控制单元200是处于正常状态还是低可靠性状态而切换存储单元信息。即,在正常状态下,控制单元200将指示存储单元2是可读取和可写入的存储单元的存储单元信息传输到CPU 5。在低可靠性状态下,控制单元200将指示存储单元2是仅可读取的存储单元的存储单元信息传输到CPU 5。

[0553] 在从CPU 5接收用于存储单元2的存储单元信息请求时,在正常状态下,控制单元200将指示存储单元2是可读取和可写入的存储单元的存储单元信息作为对存储单元信息请求的响应传输到CPU 5。在低可靠性状态下,控制单元200将指示存储单元2是仅可读取的存储单元的存储单元信息传输到CPU 5。作为存储单元信息请求的实例,使用ACS-3中描述的“ECh标识设备(ECh IDENTIFY DEVICE)”命令、SCSI初级命令-4 (SPC-4)中描述的“A3h报告标识信息(A3h REPORT IDENTIFYING INFORMATION)”命令、信息技术多媒体命令-6 (MMC-

6) (<http://www.t10.org/>) 中描述的“46h获得配置 (46h GET CONFIGURATION)”命令和快速NVM修订1.1中描述的“06h标识 (06h Identify)”命令。

[0554] 图80示出在控制单元200从CPU 5接收用于存储单元2的存储单元信息请求时执行的处理程序的实例。当控制单元200从CPU 5接收用于存储单元2的存储单元信息请求时(步骤S801),控制单元200基于控制单元200处于正常状态还是低可靠性状态而判定存储单元2是否处于低可靠性状态(步骤S802)。当存储单元2处于正常状态时,控制单元200将存储单元信息请求传输到存储单元2(步骤S803),从存储单元2接收存储单元信息,并且将接收的存储单元信息传输到CPU 5(步骤S804)。当存储单元2处于低可靠性状态时,控制单元200将存储单元信息请求传输到存储单元2(步骤S805),从存储单元2接收存储单元信息,重写接收数据以更改存储单元信息以便指示似乎存储单元是ROM设备,并且将更改后的存储单元信息传输到CPU 5(步骤S806)。控制单元200能够自发地执行S802至S806处的处理,而不在步骤S801处从CPU 5接收存储单元信息请求。控制单元200能够在步骤S801和步骤S802之间将存储单元信息请求传输到存储单元2,并且不必在步骤S803和步骤S805将存储单元信息请求传输到存储单元2。

[0555] 如图81中所示,在低可靠性状态下,控制单元200能够生成指示似乎存储单元2是ROM设备的存储单元信息,并且将存储单元信息传输到CPU 5,而不将存储单元信息请求传输到存储单元2。即,在从CPU 5接收用于存储单元2的存储单元信息请求时(步骤S810),控制单元200基于控制单元200处于正常状态还是低可靠性状态而判定存储单元2是否处于低可靠性状态(步骤S811)。当控制单元200处于正常状态时,控制单元200将存储单元信息请求传输到存储单元2(步骤S812),并且将接收的存储单元信息传输到CPU 5(步骤S813)。当控制单元200处于低可靠性状态时,控制单元200生成用于向CPU 5指示似乎存储单元2是ROM设备的存储单元信息,并且将生成的存储单元信息传输到CPU 5(步骤S814),而不将存储单元信息请求传输到存储单元2。

[0556] 当在正常状态下操作时,因为存储单元信息指示存储单元2是可读取和可写入的存储单元,所以例如希望控制单元200明确地向CPU 5通知存储单元2是ATA设备。例如,可通过以下操作向CPU 5通知存储单元2是ATA设备:在ATA/ATAPI命令集-3 (ACS-3) 中描述的设备签名中,取代LBA (7:0) 中的01h、取代LBA (15:8) 中的00h并取代LBA (23:16) 中的00h,并且将LBA (7:0)、LBA (15:8) 和LBA (23:16) 输出到CPU 5。

[0557] 当在低可靠性状态下操作时,因为存储单元信息指示存储单元2是仅可读取的存储单元,所以例如控制单元200明确地向CPU 5通知存储单元2是ATAPI设备。例如,可通过以下操作向CPU 5通知存储单元2是ATAPI设备:在ACS-3中描述的设备签名中,取代LBA (7:0) 中的01h、取代LBA (15:8) 中的14h并取代LBA (23:16) 中的EBh,并且将LBA (7:0)、LBA (15:8) 和LBA (23:16) 输出到CPU 5。进一步,当在低可靠性状态下操作时,例如在从CPU 5接收INCITS多媒体命令-6 (MMC-6) 中描述的“46h获得配置”命令时,因为存储单元信息指示存储单元2是仅可读取的存储单元,所以控制单元200向CPU 5通知在以下特性中不支持所有写入功能:例如可随机写入(特性编号=0020h)、可递增流写入(特性编号=0021h) 和一次性写入(特性编号=0025h)。因此,甚至在OS 100是Windows (注册商标) 等时,也可导致OS 100将存储单元2识别为可读取设备。对于OS 100和高于OS 100的层中的应用程序400而言,将存储单元2视为如同存储单元2是只读设备那样。因此,可防止OS 100和应用程序400错误地

将写入命令传输到存储单元2。

[0558] 备选地,当在低可靠性状态下操作时,因为存储单元信息指示存储单元2是仅可读取的存储单元,所以例如控制单元200能够明确地向CPU 5通知存储单元2是ATA设备。在从CPU 5接收ACS-3中描述的“ECh标识设备”命令时,控制单元200能够向CPU 5通知指示不支持所有写入功能的信息。

[0559] 除了上面解释的形式之外,用于通知存储单元2是否是仅可读取的存储单元的方法还能够采用各种形式。当CPU 5接收指示存储单元2是仅可读取的存储单元的信息时,OS 100应用用于只读存储单元(例如,ATAPI只读存储单元)的驱动程序作为应用于存储单元2的驱动程序软件。在OS 100上,将存储单元2识别为诸如CD-ROM、DVD-ROM或BD-ROM之类的只读存储单元。例如,如图82和83中所示,希望OS 100通过显示设备9使用图标图形等明确地向信息处理设备111的管理员、操作员和用户通知:存储单元2是诸如CD-ROM、DVD-ROM或BD-ROM之类的只读设备。当用户尝试将文件等写入存储单元2中时,如图84中所示,更希望OS 100通过显示设备9通知存储单元2受写入保护。

[0560] 甚至在CPU 5和OS 100将存储单元2识别为仅可读取的存储单元时,例如当OS 100将用于只读存储单元的驱动程序应用于存储单元2时,CPU 5能够将读取命令传输到存储单元2。当CPU 5将读取命令传输到存储单元2时,控制单元200将读取命令传输到存储单元2,控制单元200从存储单元2接收读取的数据,并且控制单元200将读取的数据传输到CPU 5。这样,CPU 5能够从存储单元2读取数据,而不管控制单元200是处于正常状态还是低可靠性状态。

[0561] 就防止由于存储单元2的数据破坏或故障而导致的用户数据丢失而言,希望处于低可靠性状态的控制单元200被配置为根本不将写入命令传输到存储单元2。但是,例如当有必要将诸如OS 100的系统信息之类的数据的一部分写入存储单元2中时,作为例外,控制单元200能够允许将数据写入存储单元2中。但是,希望写入数据的数据量足够小(与存储单元2的容量相比)。为了防止用户错误地传输写入命令并且将数据写入存储单元2中,更希望:即使从CPU 5接收到达存储单元2的正常写入命令,控制单元200也不会将写入命令传输到存储单元2。作为例外,仅在有必要将数据写入存储单元2中时,例如仅在从CPU 5接收根据特殊命令的写入命令时,控制单元200才将写入命令传输到存储单元2。例如,当诸如ACS-3中描述的“35h写入DMA EXT”或“61h写入排队的FPDMA”之类的写入命令用作CPU 5的到达存储单元2的正常写入命令时,在从CPU 5接收“35h写入DMA EXT”命令或“61h写入排队的FPDMA”命令作为正常写入命令时,处于正常状态的控制单元200将此命令传输到存储单元2。在从CPU 5接收“35h写入DMA EXT”命令或“61h写入排队的FPDMA”命令作为正常写入命令时,处于低可靠性状态的控制单元200不将此写入命令传输到存储单元2。另一方面,在从CPU 5接收INCITS ACS-3中描述的“30h写入扇区”命令、“3Fh写入日志Ext”命令或SCT命令传输、其它供应商独特命令等作为特殊写入命令时,处于低可靠性状态的控制单元200将此命令传输到存储单元2。

[0562] 在上面的解释中,信息处理系统1是诸如台式个人计算机或笔记本个人计算机之类的计算机系统。但是,信息处理系统1例如能够是图85中所示的包括存储网络的信息处理系统。在图85中,存储网络1000用作接口19和接口5001,并且网络交换机6009用作中继单元5000。

[0563] 例如,使用SAN(存储区域网络)或NAS(网络附加存储)作为存储网络1000。作为SAN,例如使用FC-SAN(光纤通道存储区域网络)或IP-SAN(网际协议存储区域网络)。作为SAN的上层协议,例如使用SCSI(小型计算机系统接口)。例如,能够采用IP-SAN作为存储网络1000。能够采用iSCSI(因特网小型计算机系统接口)作为IP-SAN的上层协议。

[0564] 网络交换机6009是被配置为连接存储网络1000上的多个网络装置(例如客户机、服务器和存储单元)的网络设备。在从网络装置接收分组时,网络交换机6009基于接收分组的目的地地址,将接收分组传输到目的地处的网络装置。

[0565] 网络交换机6009包括控制单元200。控制单元200能够完全地包括在网络交换机6009中,如图85中所示。备选地,控制单元200的一部分能够包括在网络交换机6009中,并且控制单元200的另一部分能够包括在客户机7000A的主存储器6中,或者控制单元200的一部分能够包括在网络交换机6009中,并且控制单元200的另一部分能够包括在系统1中的另一个部分(例如,客户机7000A的ROM)中。控制单元200能够以固件或软件的形式实现,或者可以以硬件的形式实现。

[0566] 以与第一实施例中的图24和29中所示相同的方式,控制单元200判定连接到存储网络1000的一个或多个存储单元是已经达到寿命结束、将要达到寿命结束还是将要出故障。当存储单元已经达到寿命结束、将要达到寿命结束或将要出故障时,控制单元200仅针对要经受寿命结束处理的存储单元而从正常状态转变为低可靠性状态。正常状态和低可靠性状态是呈现给对应于连接到存储网络1000的一个或多个存储单元中的每一个的控制单元200的模式。例如,当存储单元2A正常时,控制单元200在正常状态(即,存储单元2A的模式)下操作。例如,当控制单元200确定存储单元2A已经达到寿命结束、将要达到寿命结束或将要出故障时,或者确定存储单元2A的可靠性劣化时,控制单元200在低可靠性状态(即,存储单元2A的模式)下操作。甚至在控制单元200针对要经受寿命结束处理的存储单元而从正常状态转变为低可靠性状态时,控制单元200对于不经受寿命结束处理的存储单元在正常状态下操作。如在第一实施例中,每隔固定时间、在每固定数量类型的处理下或者在每个固定数据传输和接收下(图24中所示),或者在从存储单元接收的命令响应是错误响应(如图29和30中所示)时,执行寿命结束确定。

[0567] 在此实施例中,作为实例,两个客户机7000A和7000B以及两个存储单元2A和2B连接到存储网络1000。但是,能够采用其它任意配置作为系统1的网络装置的配置。能够使用服务器7000A代替客户机7000A。能够使用服务器7000B代替客户机7000B。将地址分别分配给存储网络上的客户机7000A、客户机7000B、存储单元2A和存储单元2B。作为地址,例如能够使用IP地址或MAC地址。例如,当客户机7000A的CPU 5将仅指定存储单元2A的地址的命令和数据传输到存储网络1000时,网络交换机6009在配置此命令和此数据的分组中从此地址指定存储单元2A。网络交换机6009将分组仅传输到存储单元2A。例如,当存储单元2A将仅指定客户机7000A的地址的响应和数据传输到存储网络1000时,网络交换机6009在配置此响应和此数据的分组中从此地址指定客户机7000A。网络交换机6009将分组仅传输到客户机7000A。在地址指定中,不仅能够指定单个网络装置,而且还能够指定多个网络装置。

[0568] 当例如将存储单元2A的存储单元信息例如传输到客户机7000A时,控制单元200根据控制单元200的存储单元2A的模式处于正常状态还是低可靠性状态,切换传输到客户机7000A的存储单元2A的存储单元信息。即,当控制单元200的存储单元2A的模式为正常状态

时,控制单元200将指示存储单元2A是可读取和可写入的存储单元的存储单元信息传输到客户机7000A。当控制单元200的存储单元2A的模式为低可靠性状态时,控制单元200将指示存储单元2A是仅可读取的存储单元的存储单元信息传输到客户机7000A。也以相同方式将存储单元2A的存储单元信息传输到客户机7000B。也以相同方式将存储单元2B的存储单元信息传输到客户机7000A和客户机7000B。

[0569] 如上面所解释,在此实施例中,当控制单元200确定存储单元2已经达到寿命结束、将要达到寿命结束或将要出故障,或者确定存储单元2的可靠性劣化时,控制单元200处理或生成传输到CPU 5的存储单元2的存储单元信息,以便从而导致CPU 5和OS 100将存储单元识别为仅可读取的存储单元。因此,可防止存储单元2中的数据写入。在诸如OS 100之类的下部软件层中,执行用于识别存储单元2是仅可读取的存储单元的处理。因此,诸如应用程序400之类的上部软件层和用户不需要执行对存储单元2的特殊控制。

[0570] 所属技术领域的技术人员将很容易想到其它优点和修改。因此,本发明在其更广泛的方面并不限于此处示出和描述的特定细节和代表性实施例。因此,可以做出各种修改而不偏离如所附权利要求及其等同物限定的总体发明概念的精神或范围。

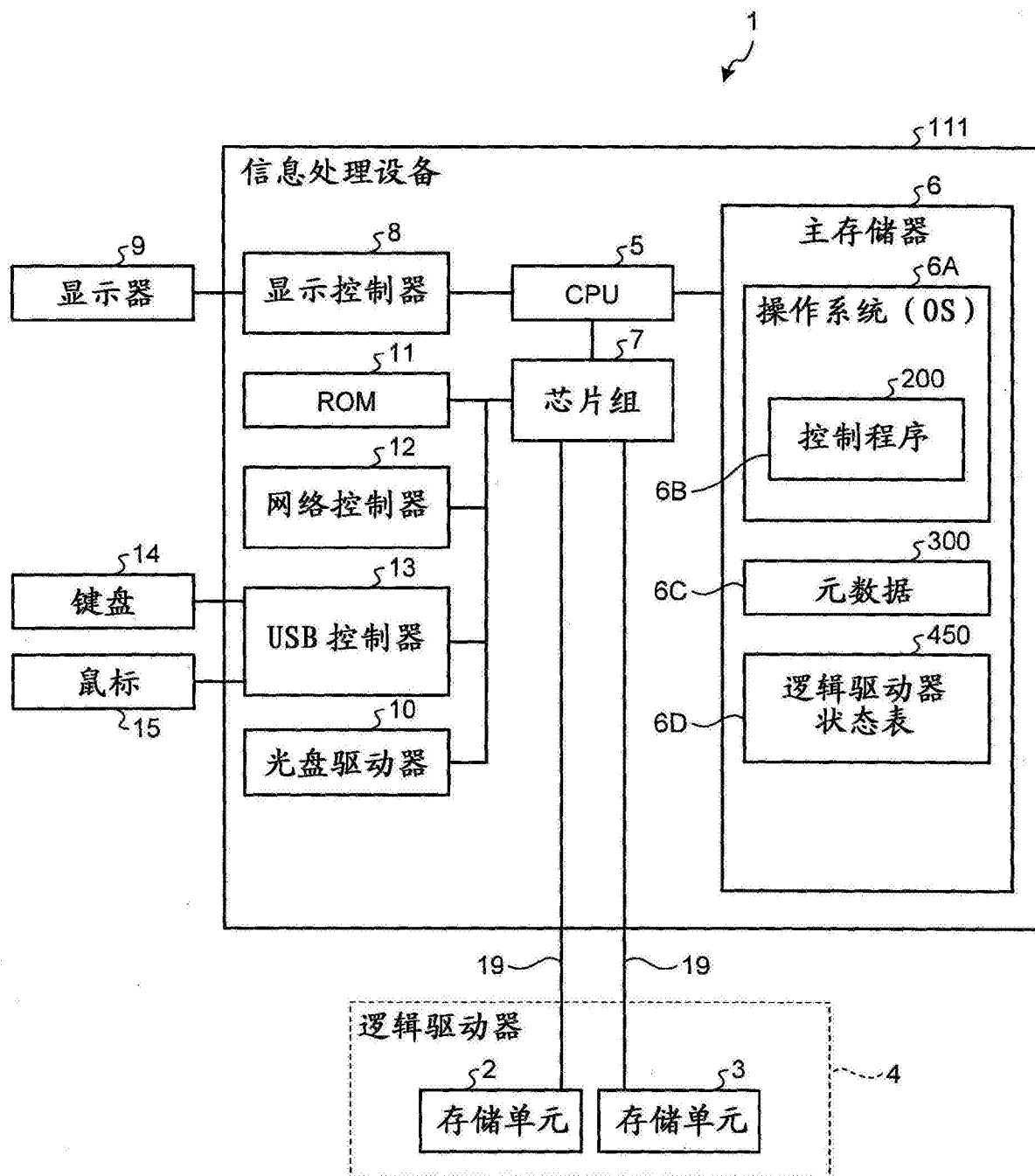


图1

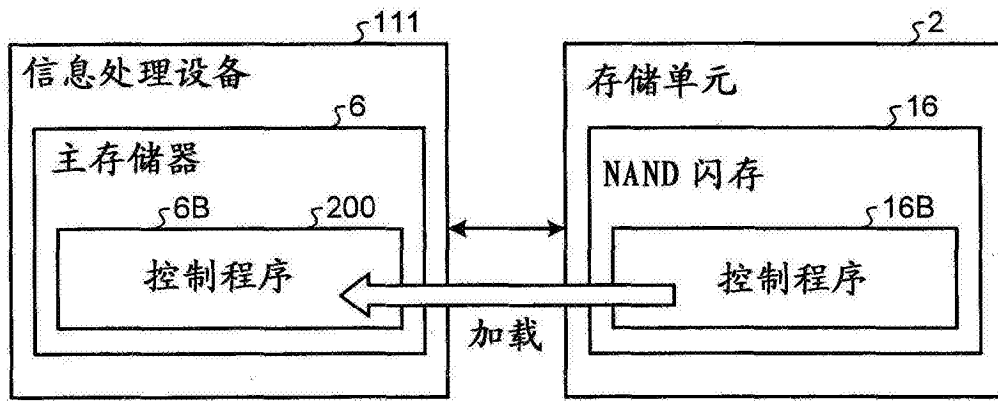


图2

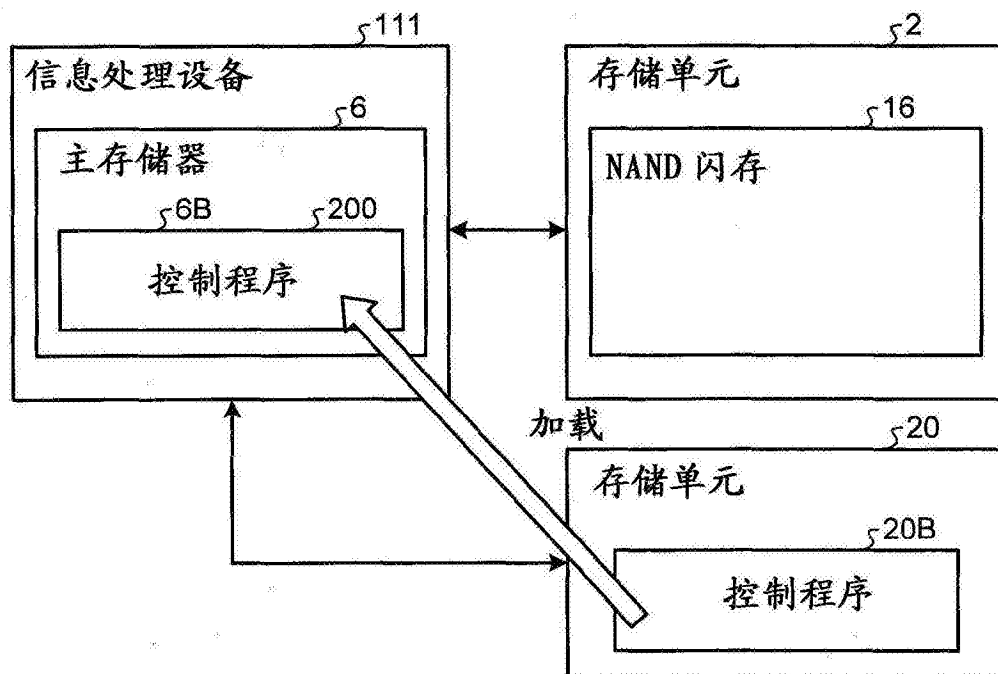


图3

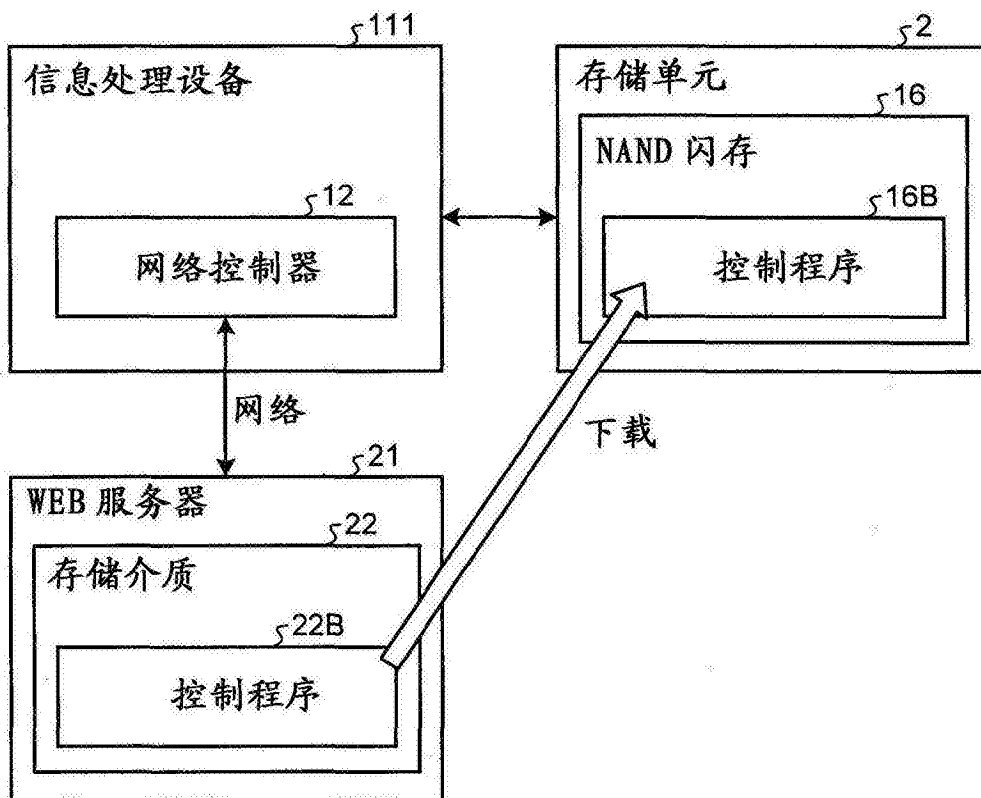


图4

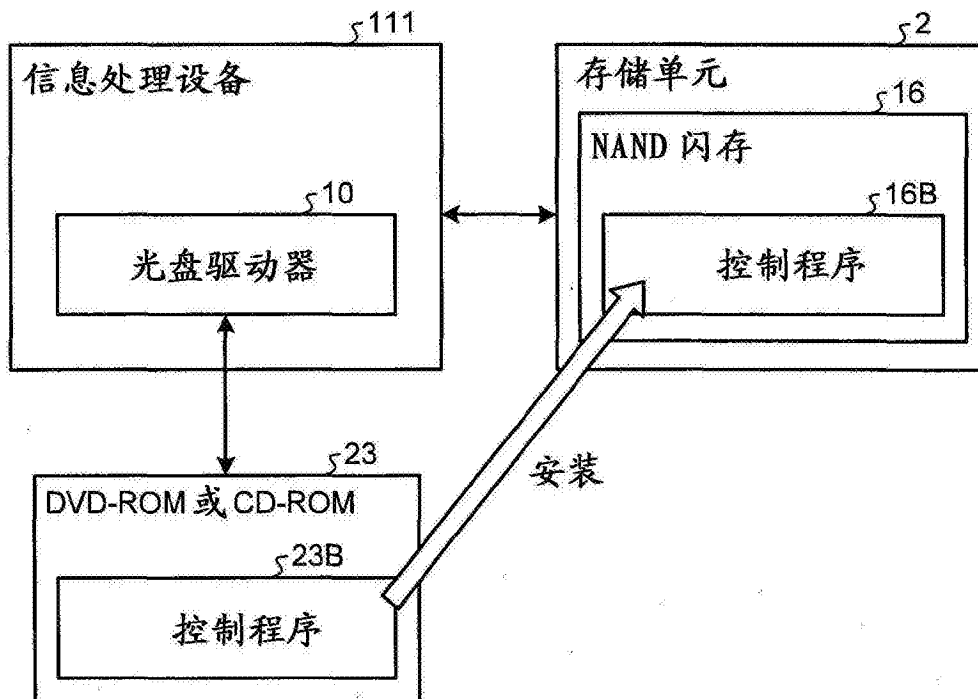


图5

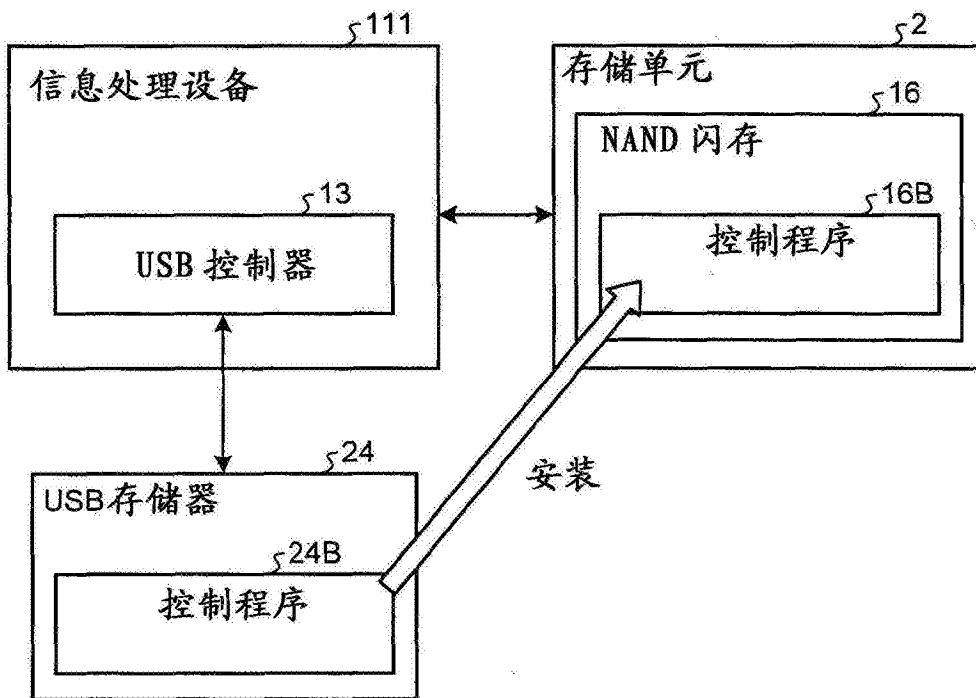


图6

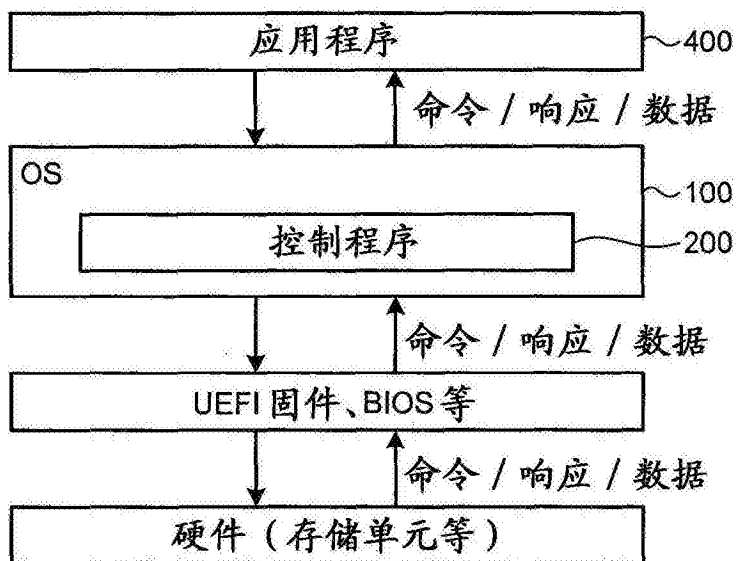


图7

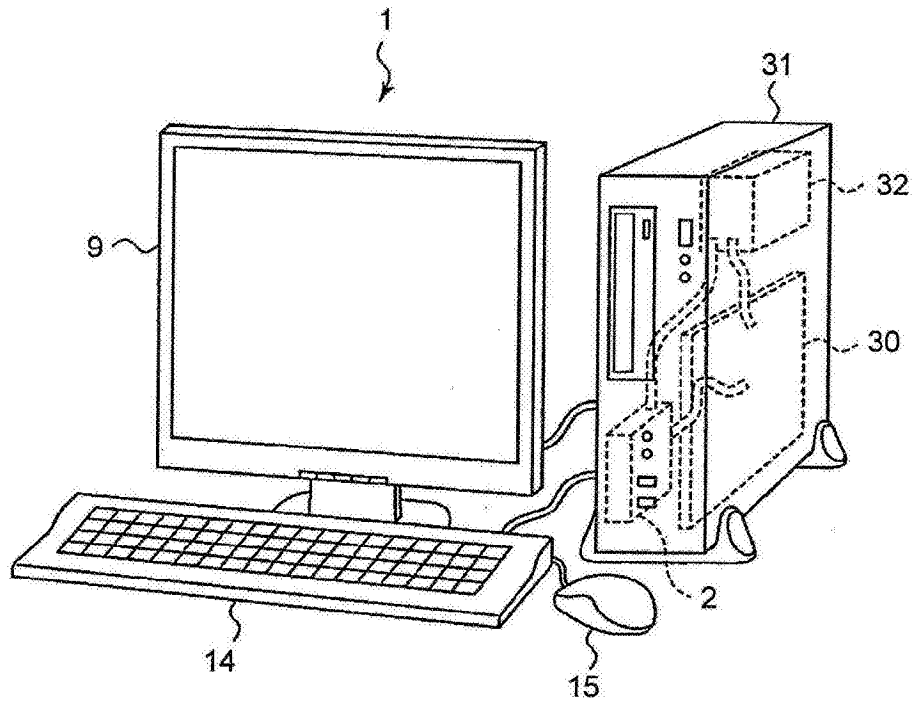


图8

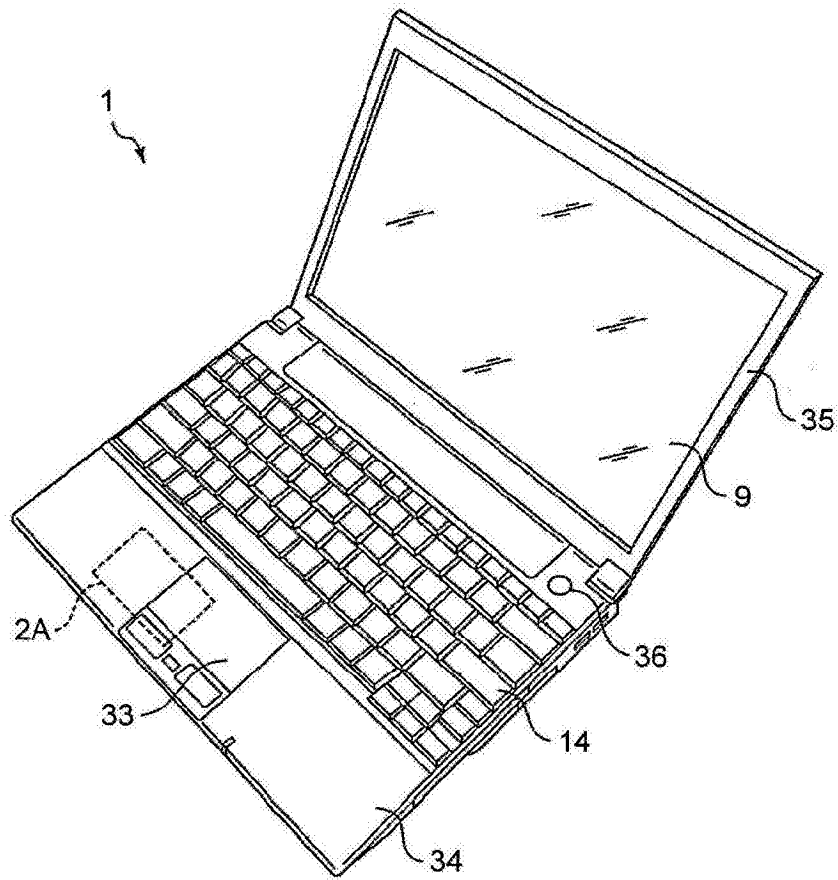


图9

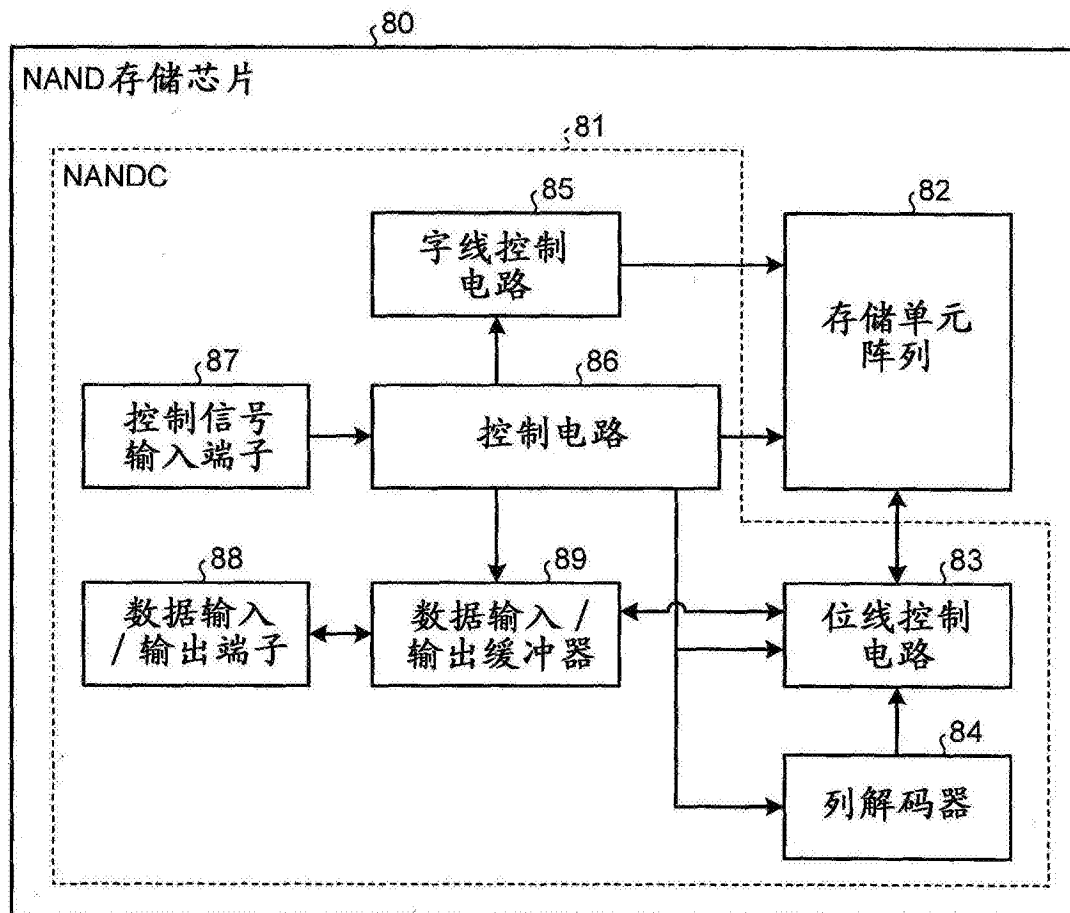


图10

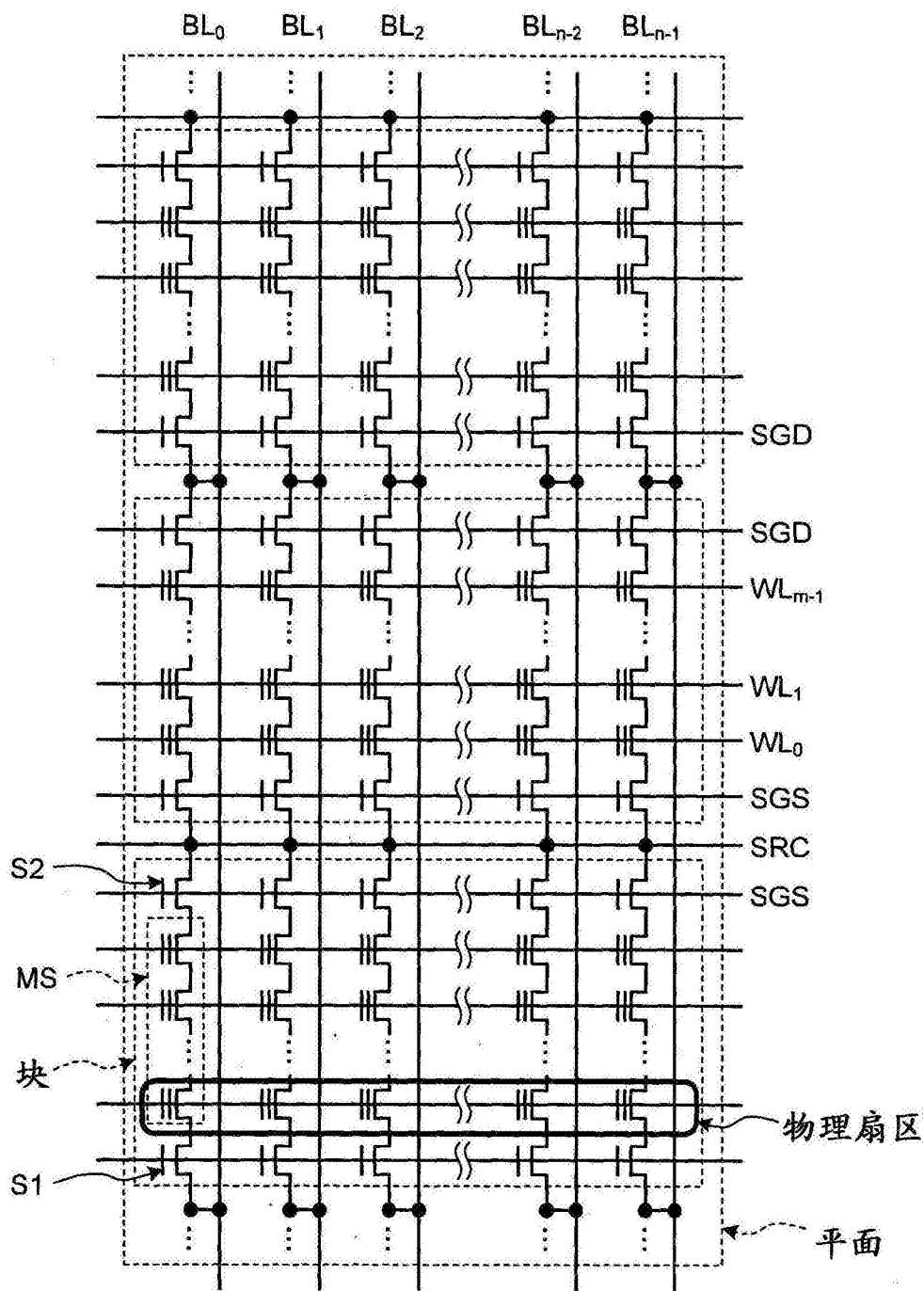


图11

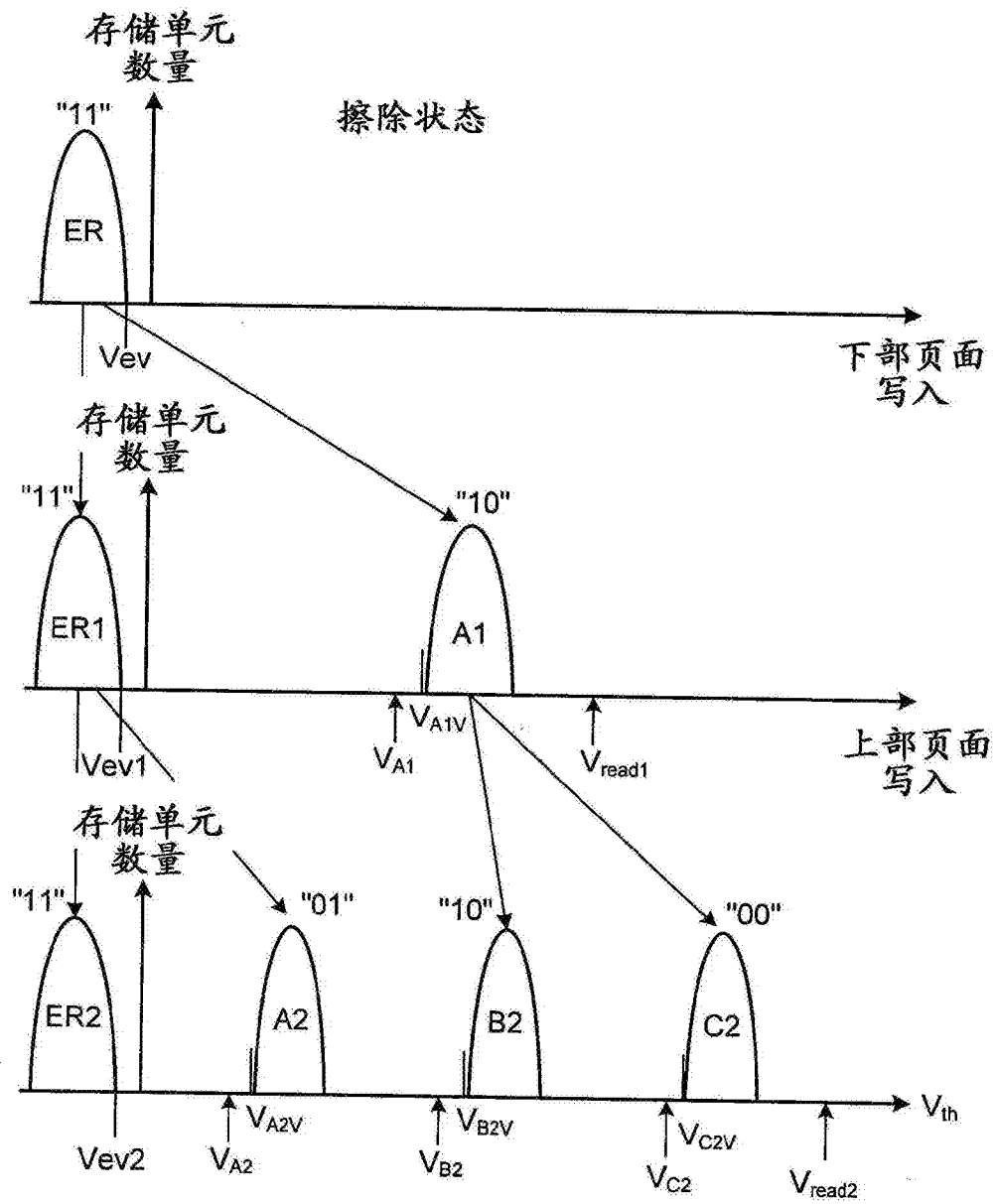


图12

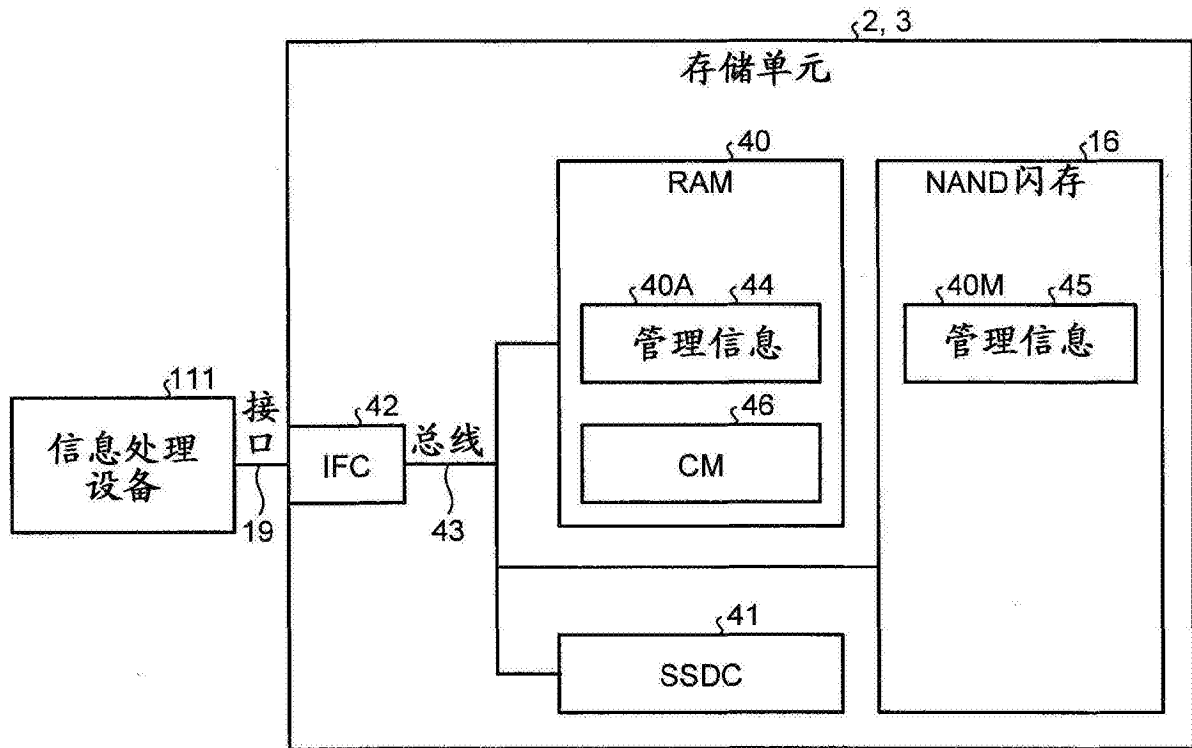


图13

管理信息

空闲块表 (FBT): 60

物理块ID	擦除计数
...	...
...	...
...	...
...	...

~44

坏块表 (BBT): 61

物理块ID
...
...
...
...

活动块表 (ABT): 62

物理块ID	擦除计数
...	...
...	...
...	...
...	...

磁道表: 63

磁道地址	物理块ID
...	...
...	...
...	...
...	...

集群表: 64

集群地址	物理块ID	物理块内页面地址
...	...	...
...	...	...
...	...	...
...	...	...

统计信息

~65

图14

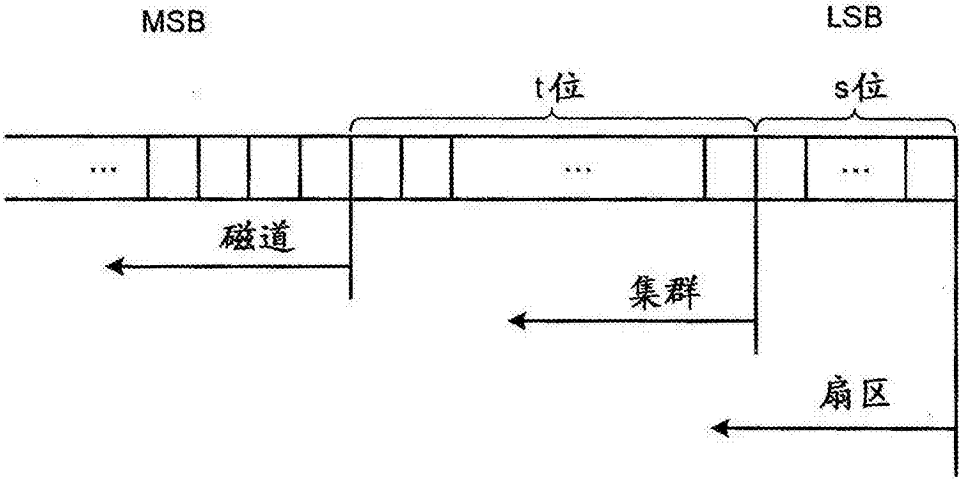


图15

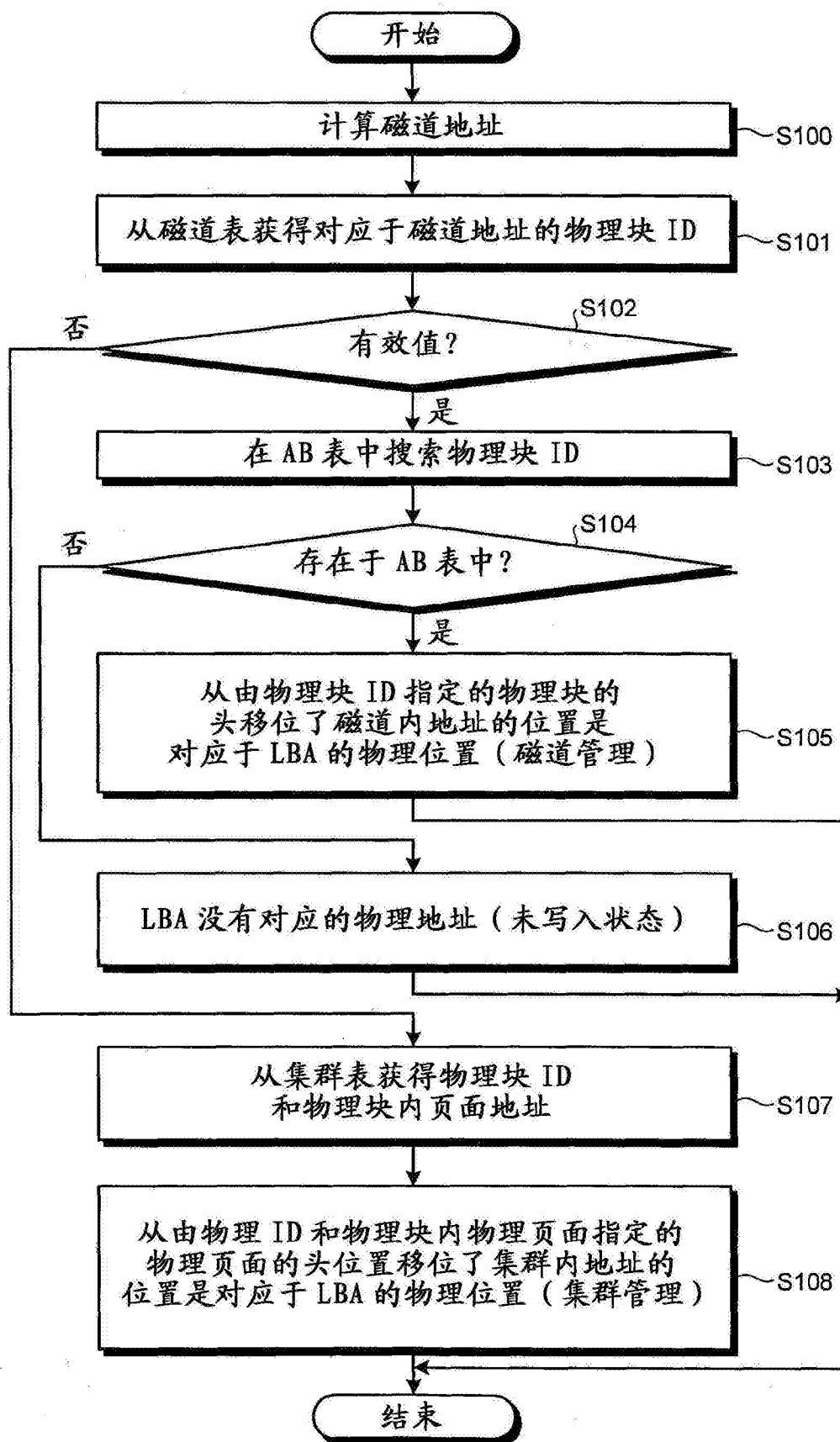


图16

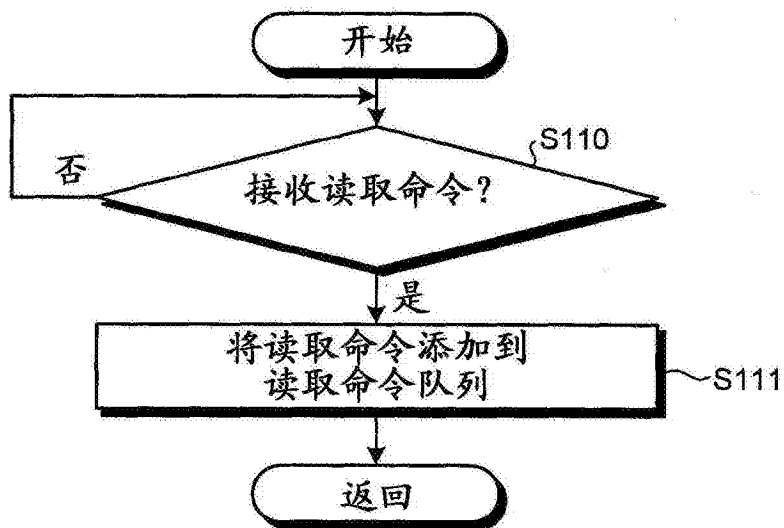


图17

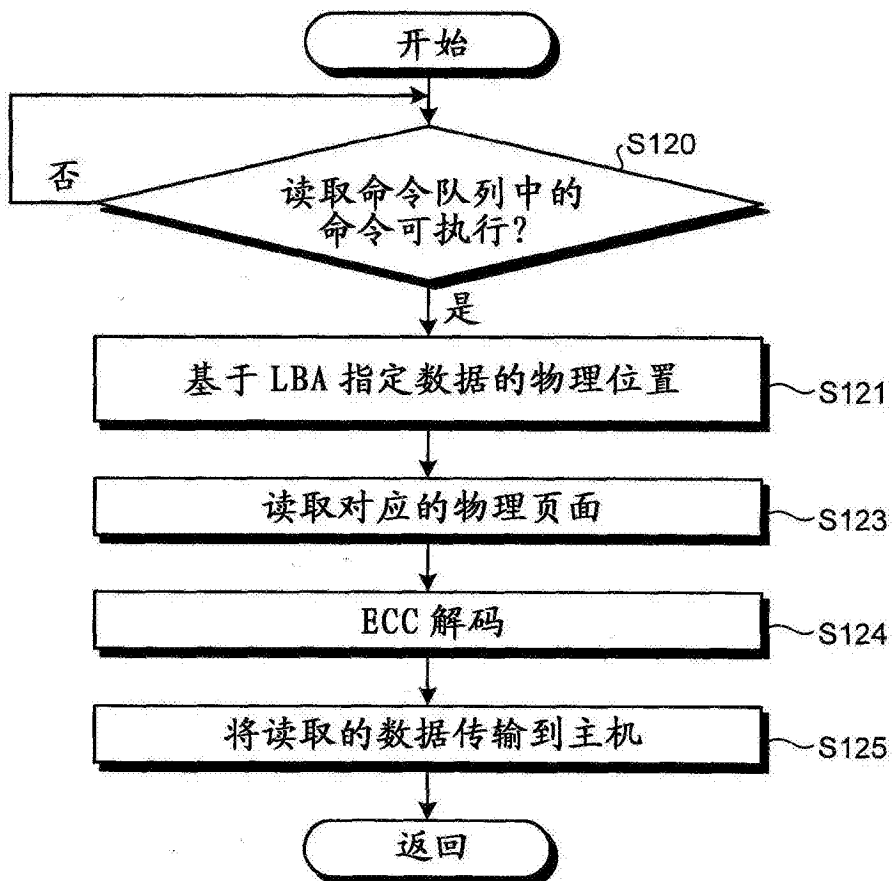


图18

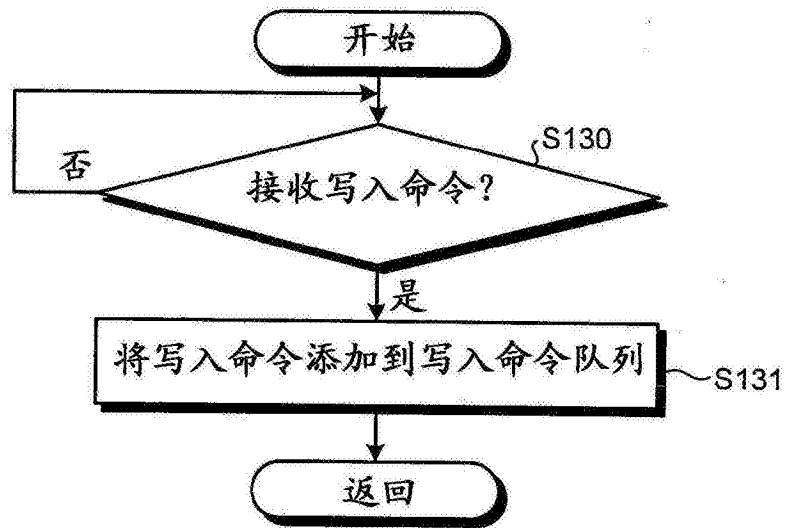


图19

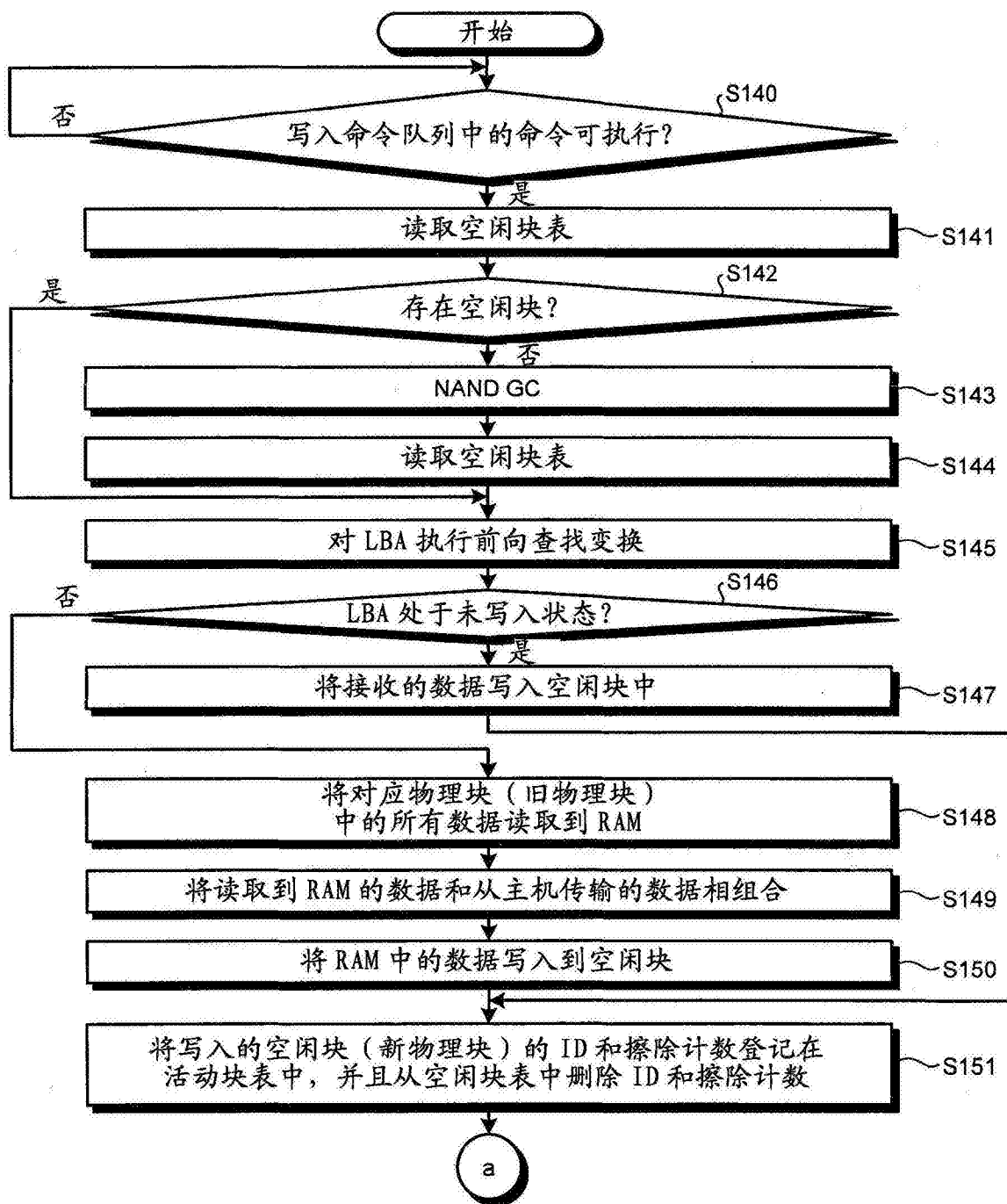


图20A

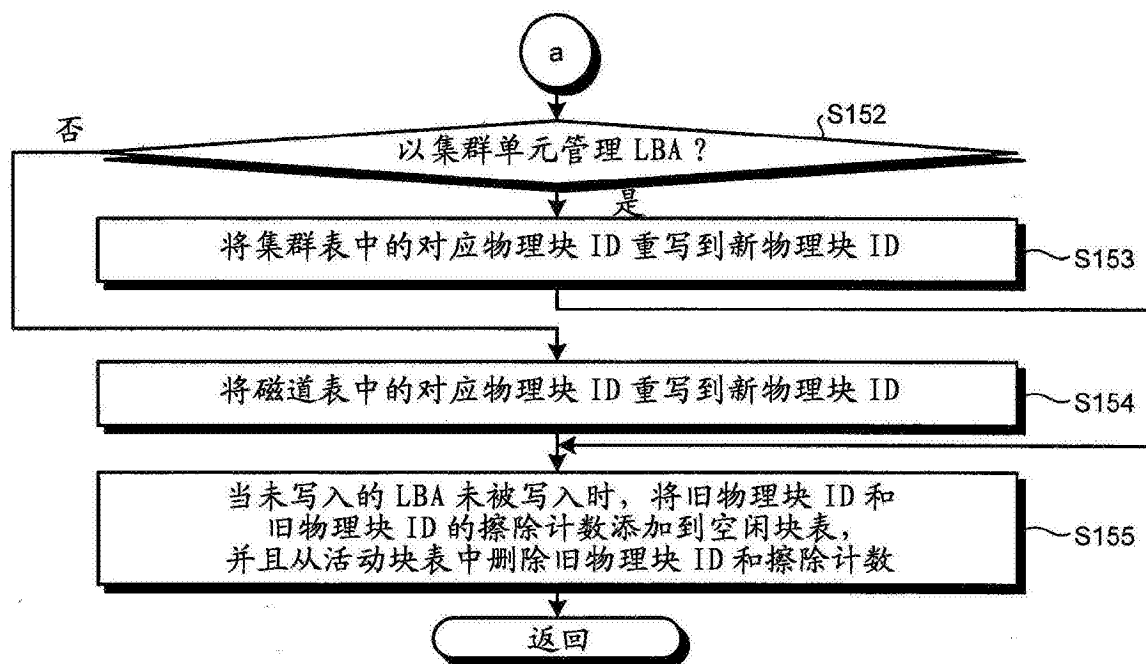


图20B

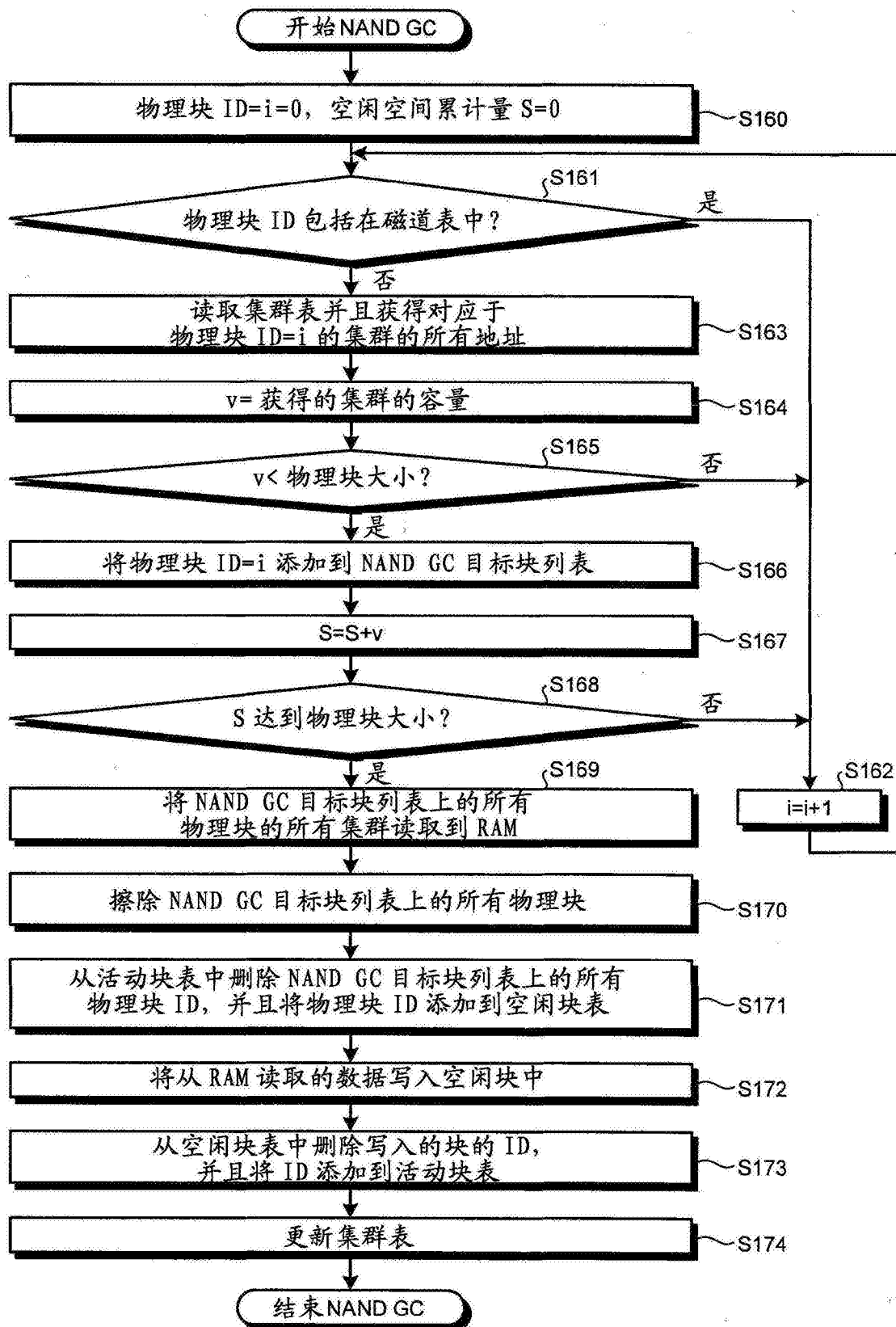


图21

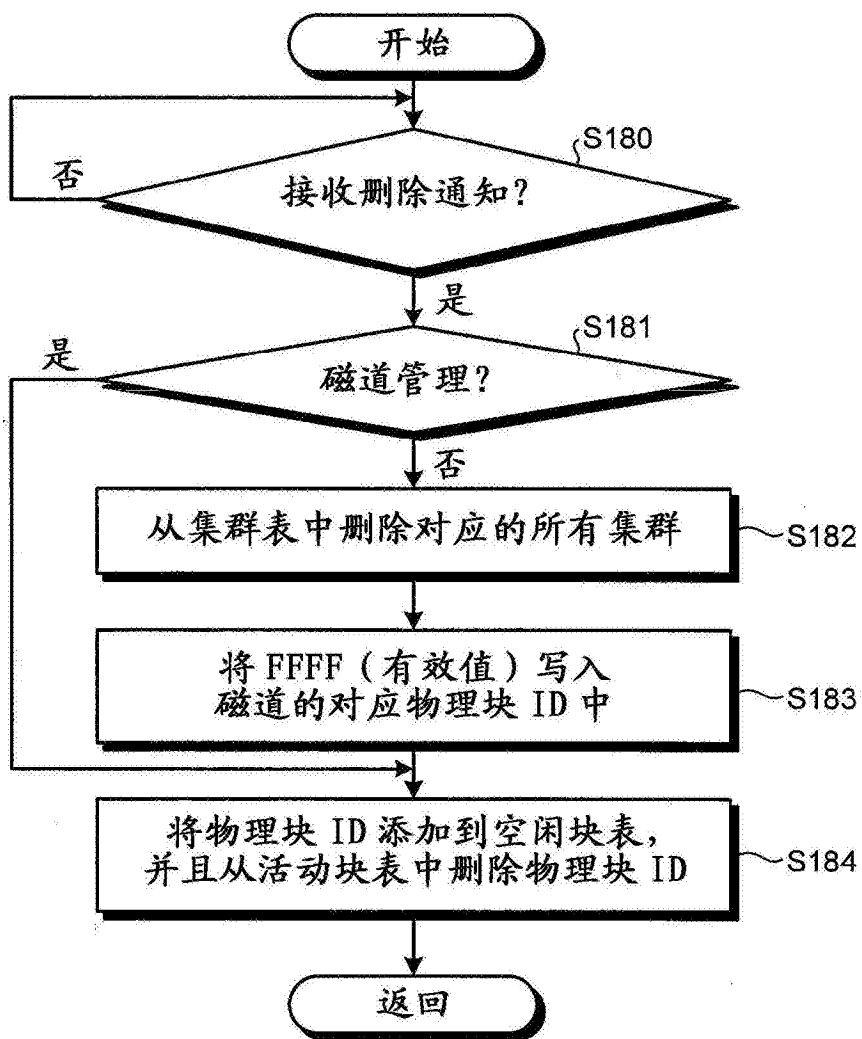


图22

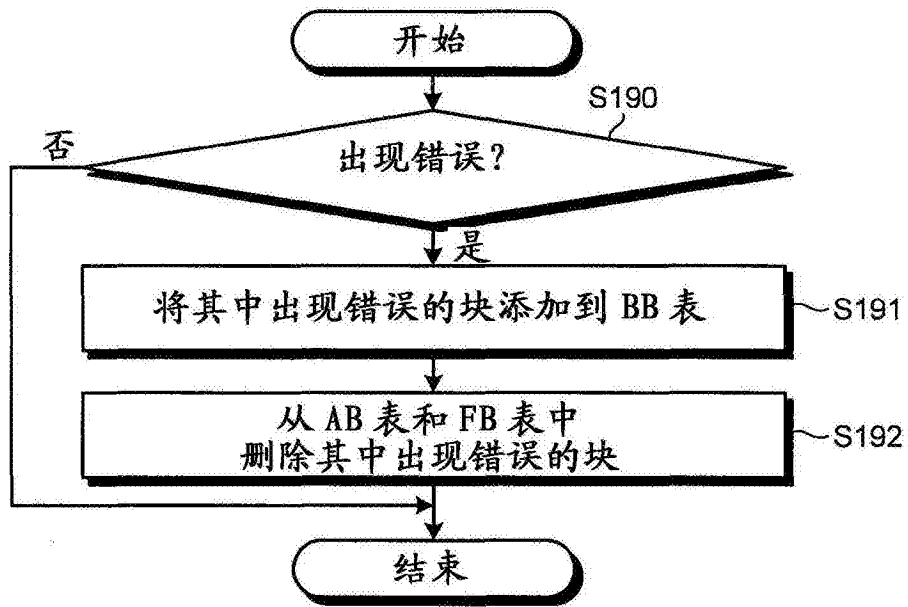


图23

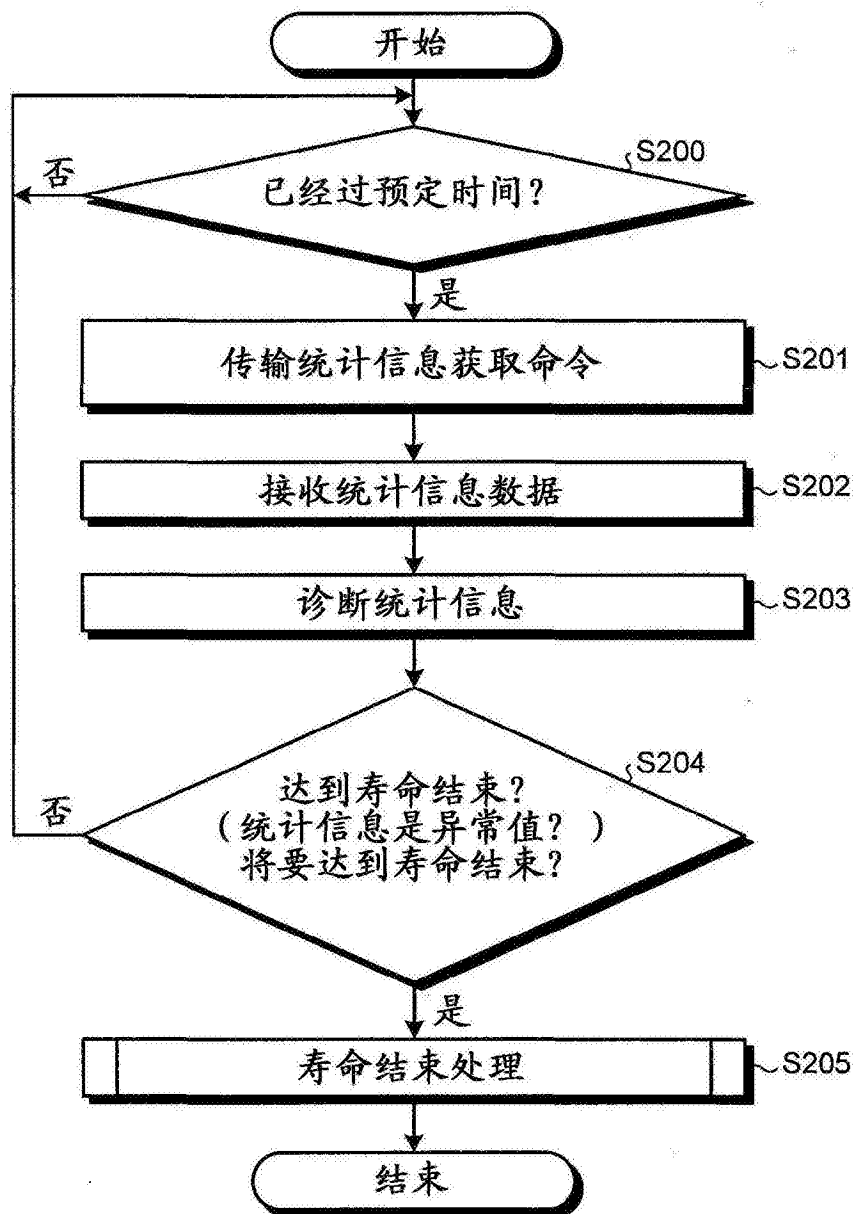


图24

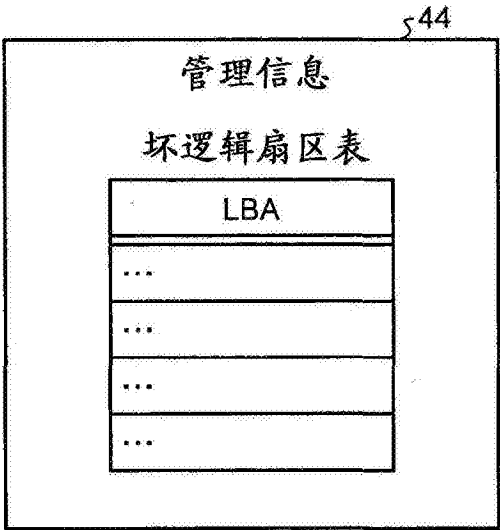


图25

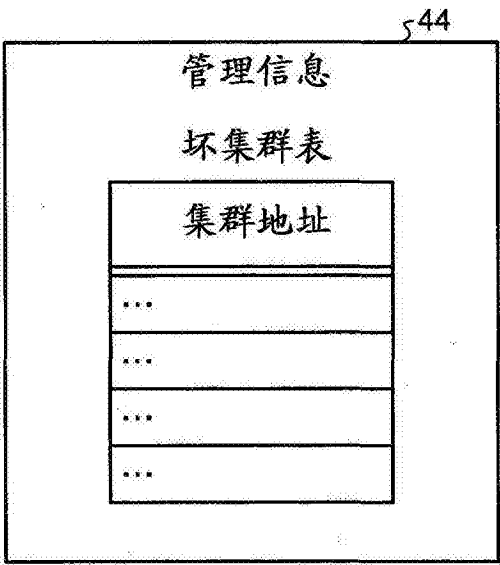


图26

属性 ID	属性名称	值	阈值	最差	原始数据
...	坏块总数 (统计信息 X01)	...	...	...	...
...	坏逻辑扇区总数 (统计信息 X02)	...	...	...	...
...	总的擦除计数 (统计信息 X03)	...	...	...	...
...	擦除计数的平均值 (统计信息 X04)	...	...	...	...
...	NAND 存储器的写入错误出现次数的累计值 (统计信息 X05)	...	...	...	...
...	NAND 存储器的擦除错误出现次数的累计值 (统计信息 X06)	...	...	...	...
...	读取逻辑扇区总数 (统计信息 X07)	...	...	...	...
...	写入逻辑扇区总数 (统计信息 X08)	...	...	...	...
...	无法进行 ECC 纠正的总次数 (统计信息 X09)	...	...	...	...
...	重试读取出现的次数 (统计信息 X10)	...	...	...	...
...	n 位至 m 位 ECC 纠正单元的总数 (统计信息 X11)	...	...	...	...
...	接口 19 的数据损坏错误次数 (统计信息 X12)	...	...	...	...
...	接口 19 的通信速度减慢次数 (统计信息 X13)	...	...	...	...
...	接口 19 的通道数量降低次数 (统计信息 X14)	...	...	...	...
...	接口 19 的错误出现次数 (统计信息 X15)	...	...	...	...
...	RAM 40 的错误出现次数 (统计信息 X16)	...	...	...	...
...	存储单元 2 的总使用时间 (统计信息 X17)	...	...	...	...
...	启动次数 (统计信息 X18)	...	...	...	...

图 27A

...	意外断电出现次数 (统计信息 X19)	...	...	...	...
...	温度超过推荐的操作温度的最大值时的累计时间 (统计信息 X20)	...	...	...	...
...	温度低于推荐的操作温度的最小值时的累计时间 (统计信息 X21)	...	...	...	...
...	命令的响应时间最大值 (统计信息 X22)	...	...	...	...
...	命令的响应时间平均值 (统计信息 X23)	...	...	...	...
...	NAND 存储器的响应时间最大值 (统计信息 X24)	...	...	...	...
...	NAND 存储器的响应时间平均值 (统计信息 X25)	...	...	...	...
...	当前温度 (统计信息 X26)	...	...	...	...
...	最高温度 (统计信息 X27)	...	...	...	...
...	最低温度 (统计信息 X28)	...	...	...	...
...	管理信息冗余 (统计信息 X29)	...	...	...	...
...	RAM 40 中总的写入数据量 (统计信息 X30)	...	...	...	...
...	统计信息增加比率 (统计信息 X31)	...	...	...	...
...	NAND GC 故障标志 (统计信息 X32)	...	...	...	...

图27B

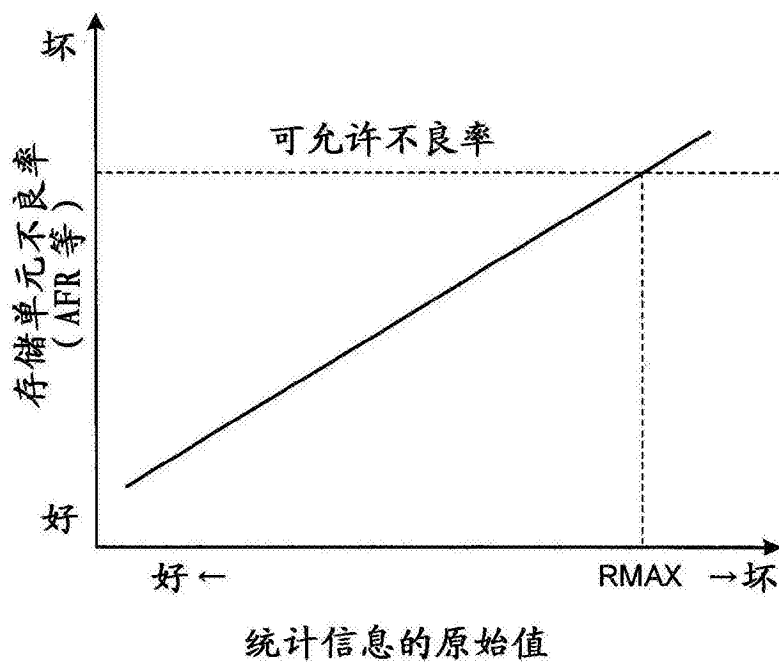


图28

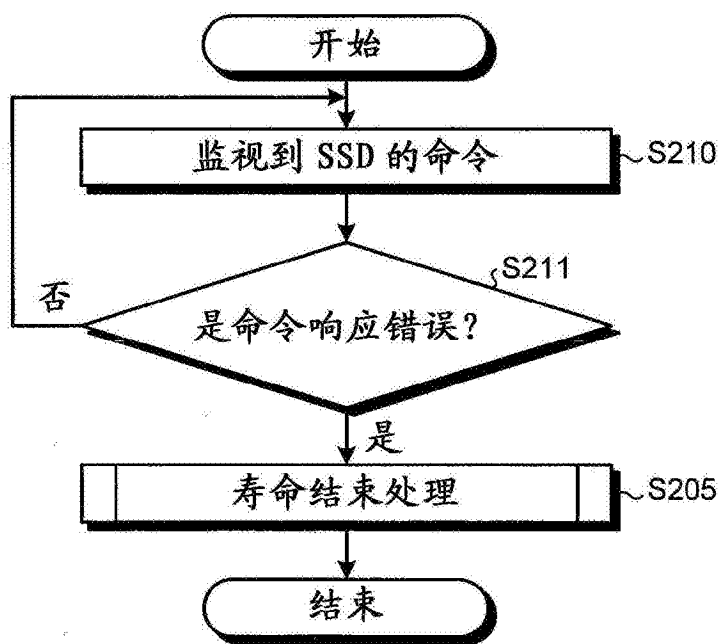


图29

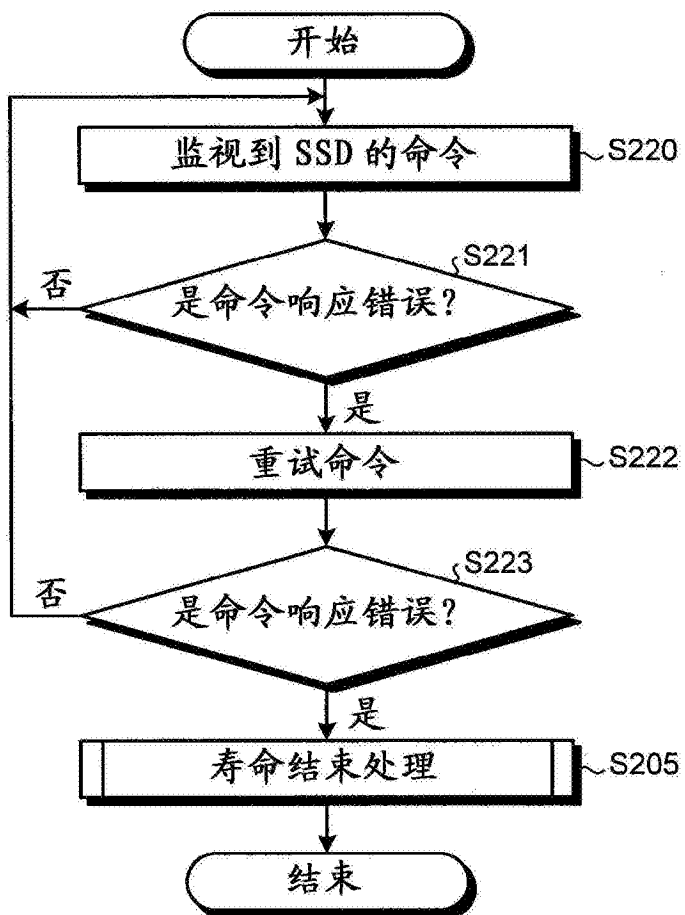


图30

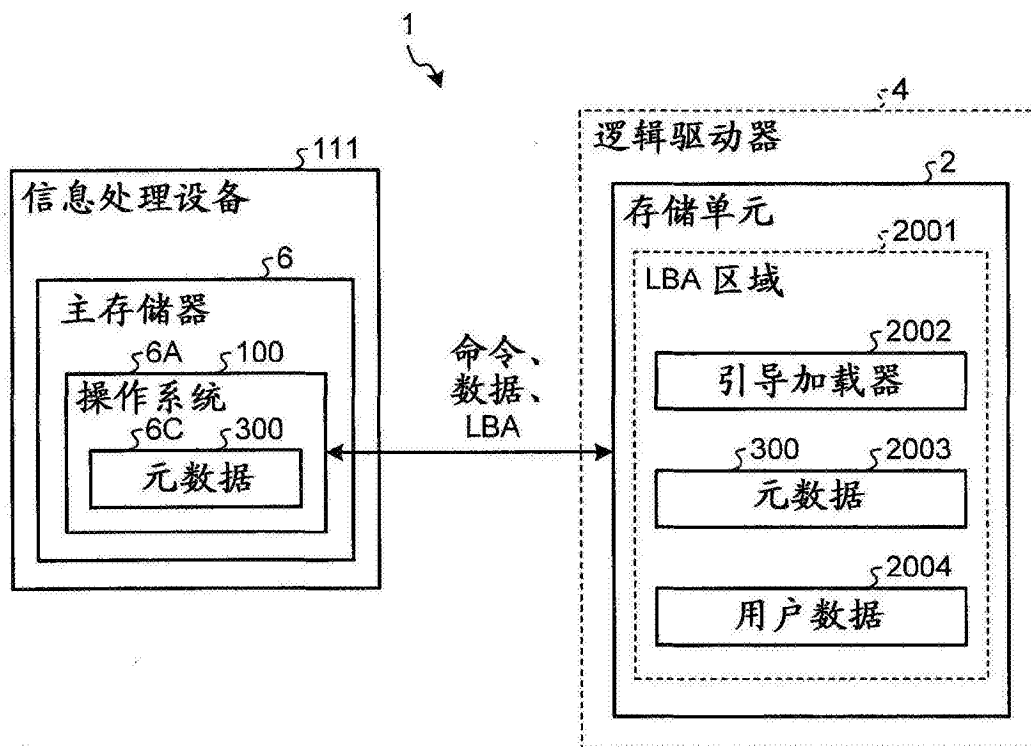


图31



图32

§6C §300

元数据				
文件 ID	逻辑驱动器地址	磁盘 ID	LBA	扇区计数
...	逻辑驱动器 4	存储单元 2	...	...
...	逻辑驱动器 4	存储单元 2	...	...
...	逻辑驱动器 4	存储单元 2	...	...
...	逻辑驱动器 4	存储单元 2	...	...
...	逻辑驱动器 4	存储单元 2	...	...

图33

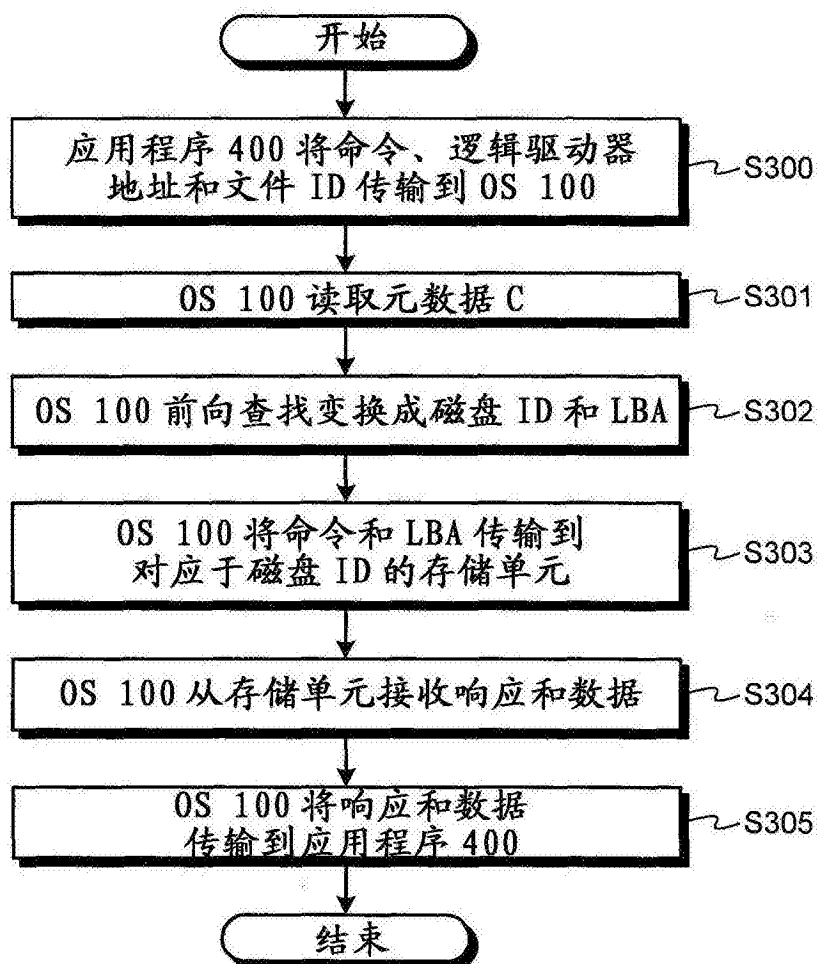


图34

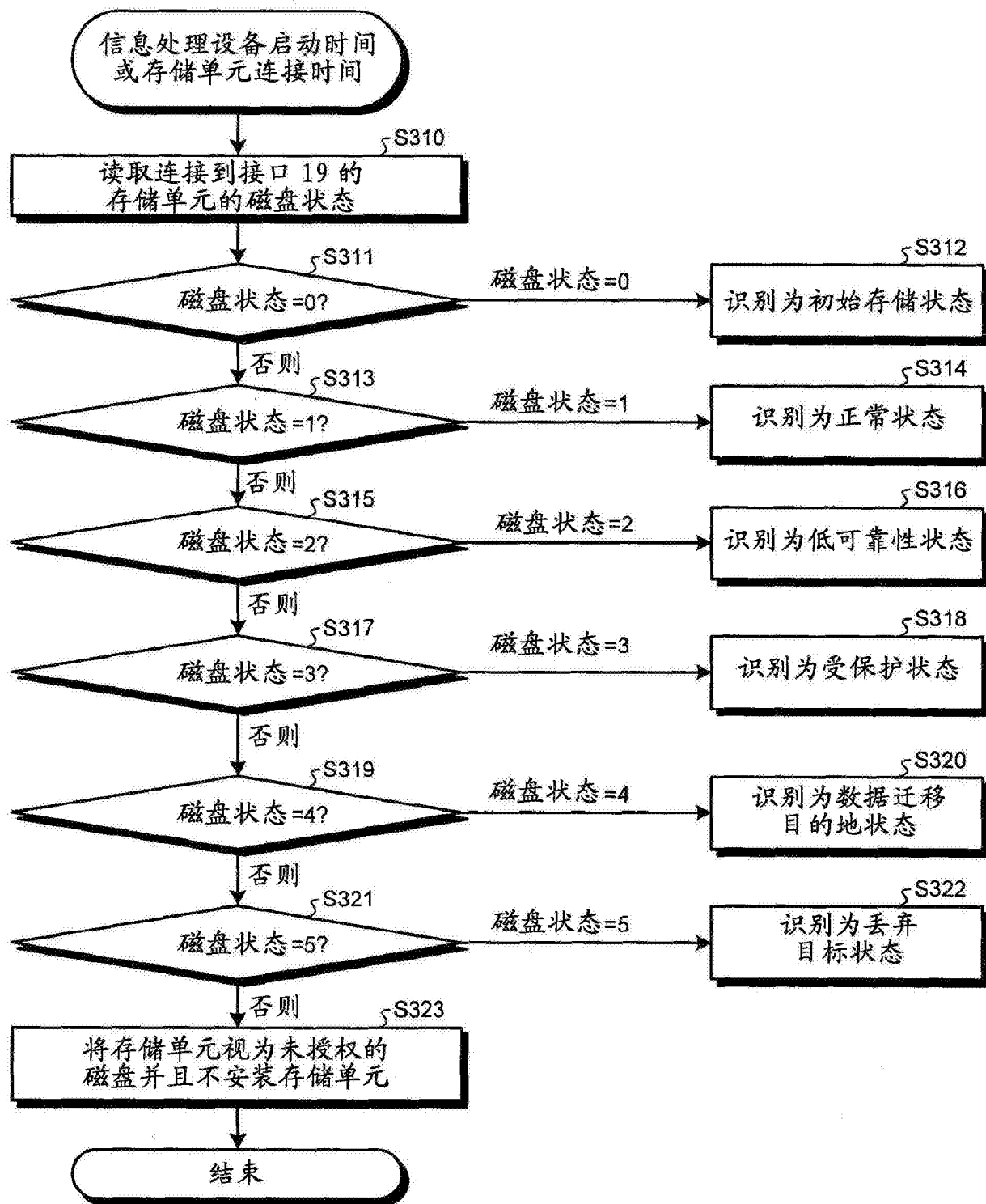


图35

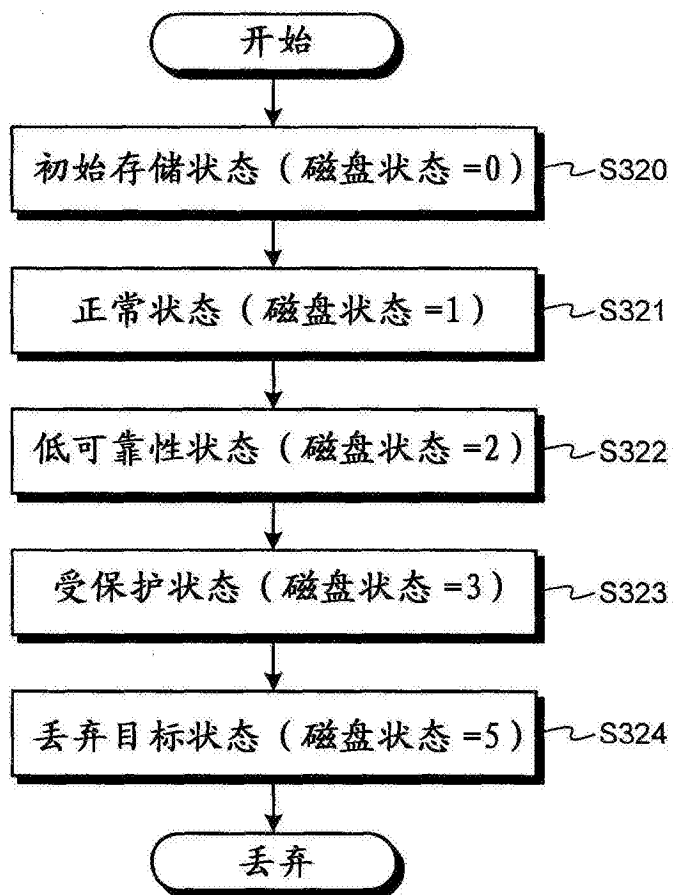


图36

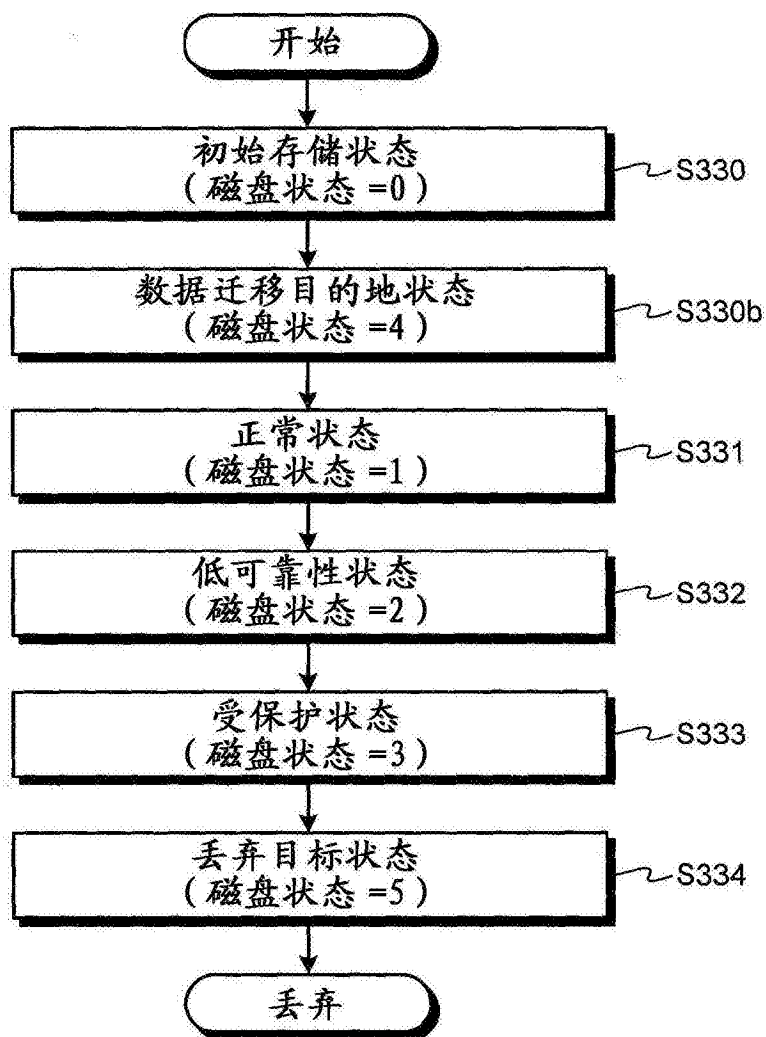


图37

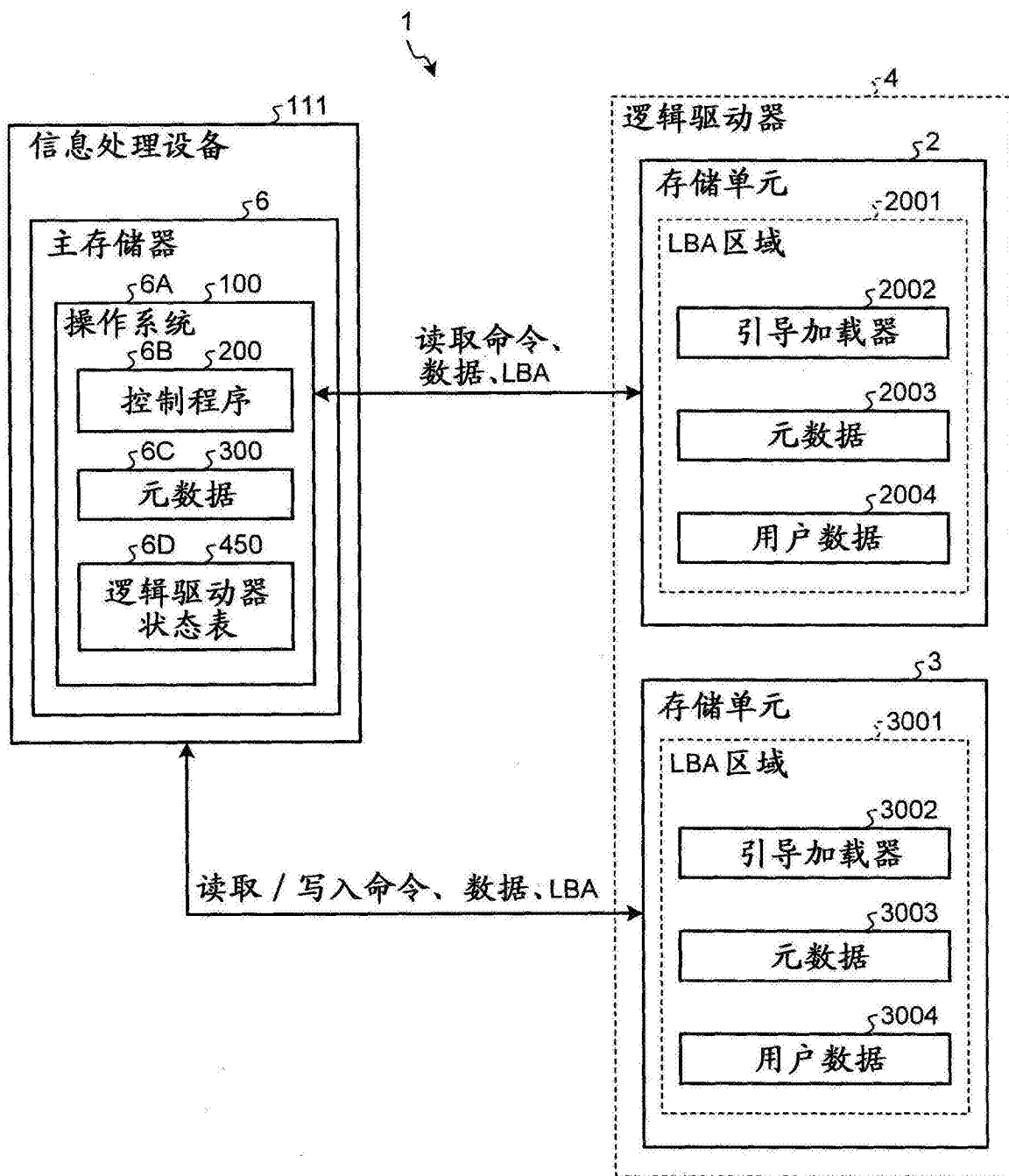


图38

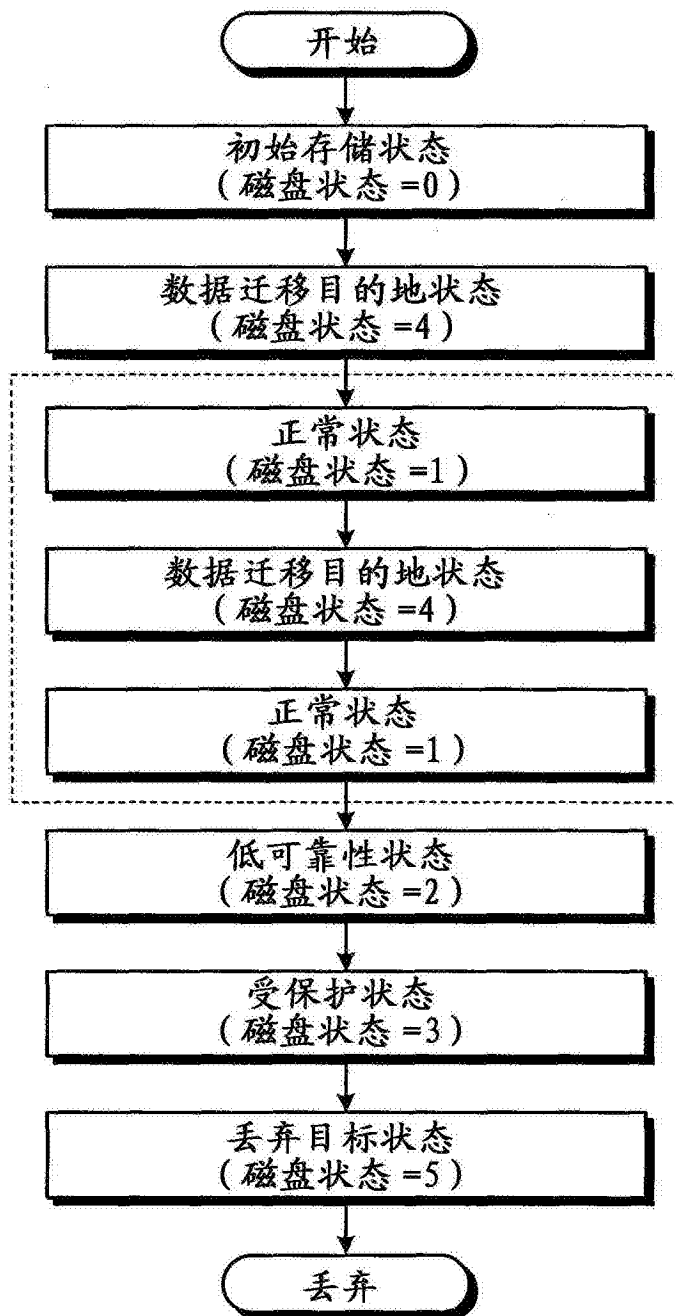


图39

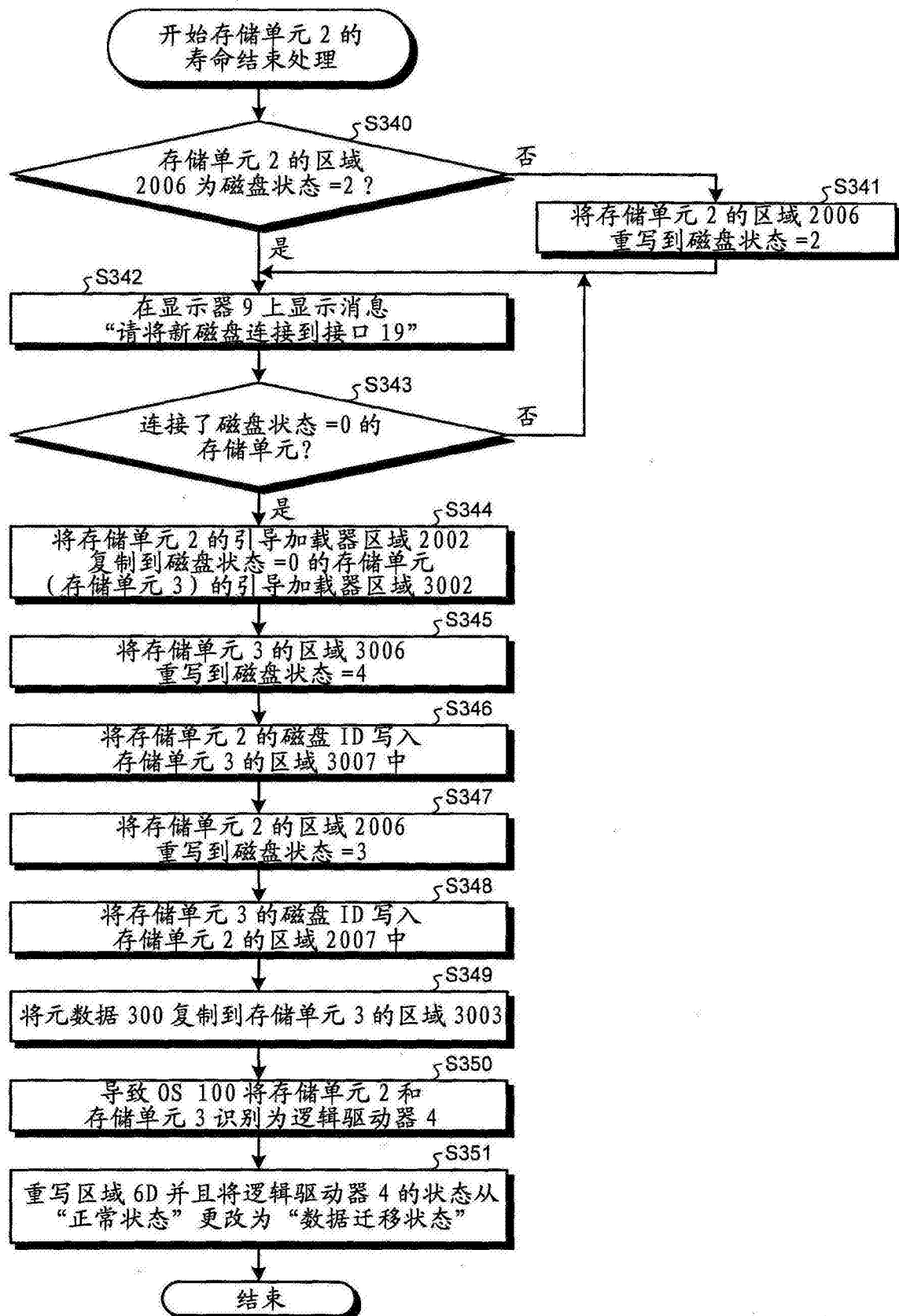


图40

§6D

§450

逻辑驱动器状态表

逻辑驱动器地址	状态
...	正常状态
...	数据迁移状态
...	正常状态

图41

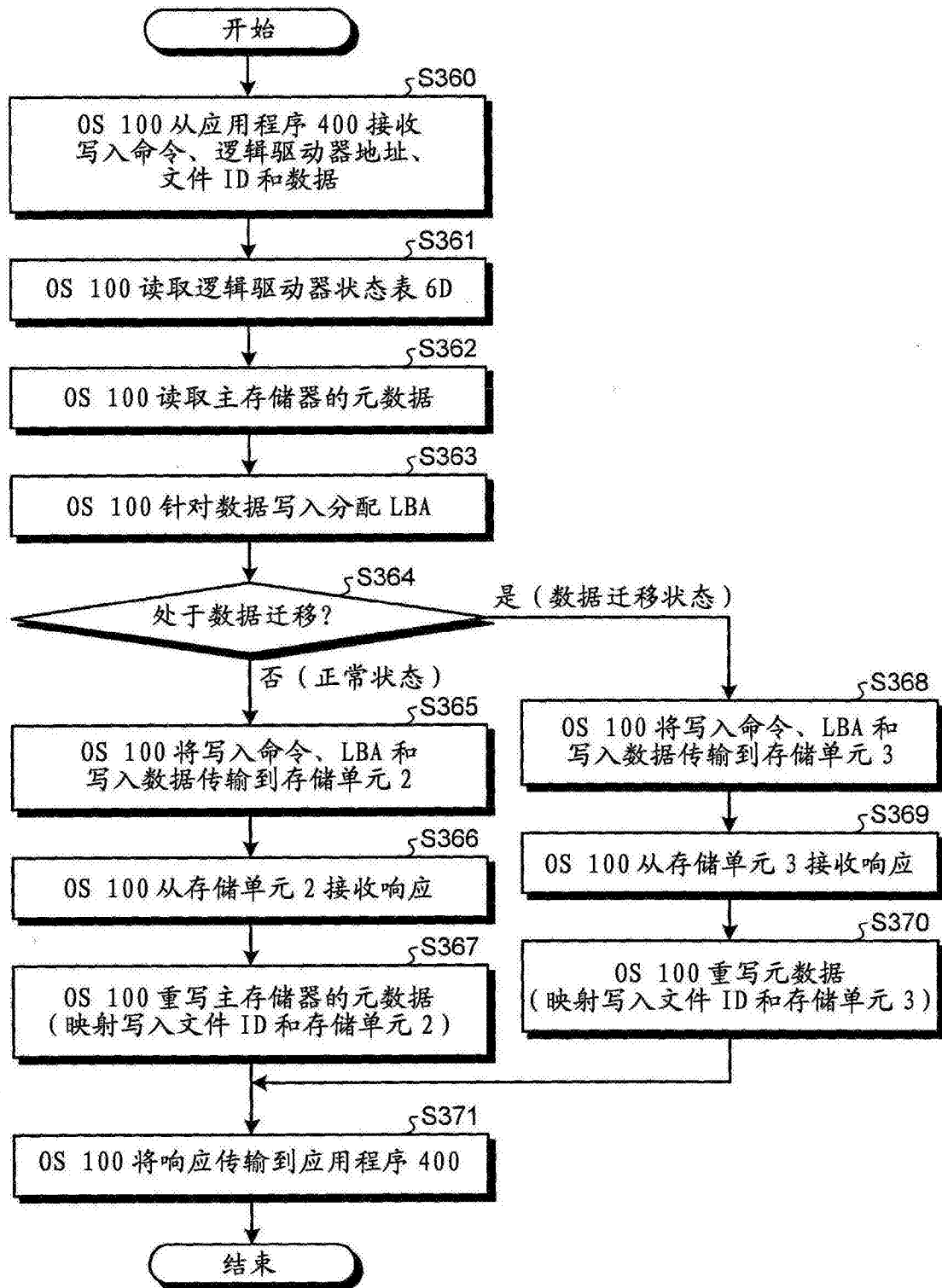


图42

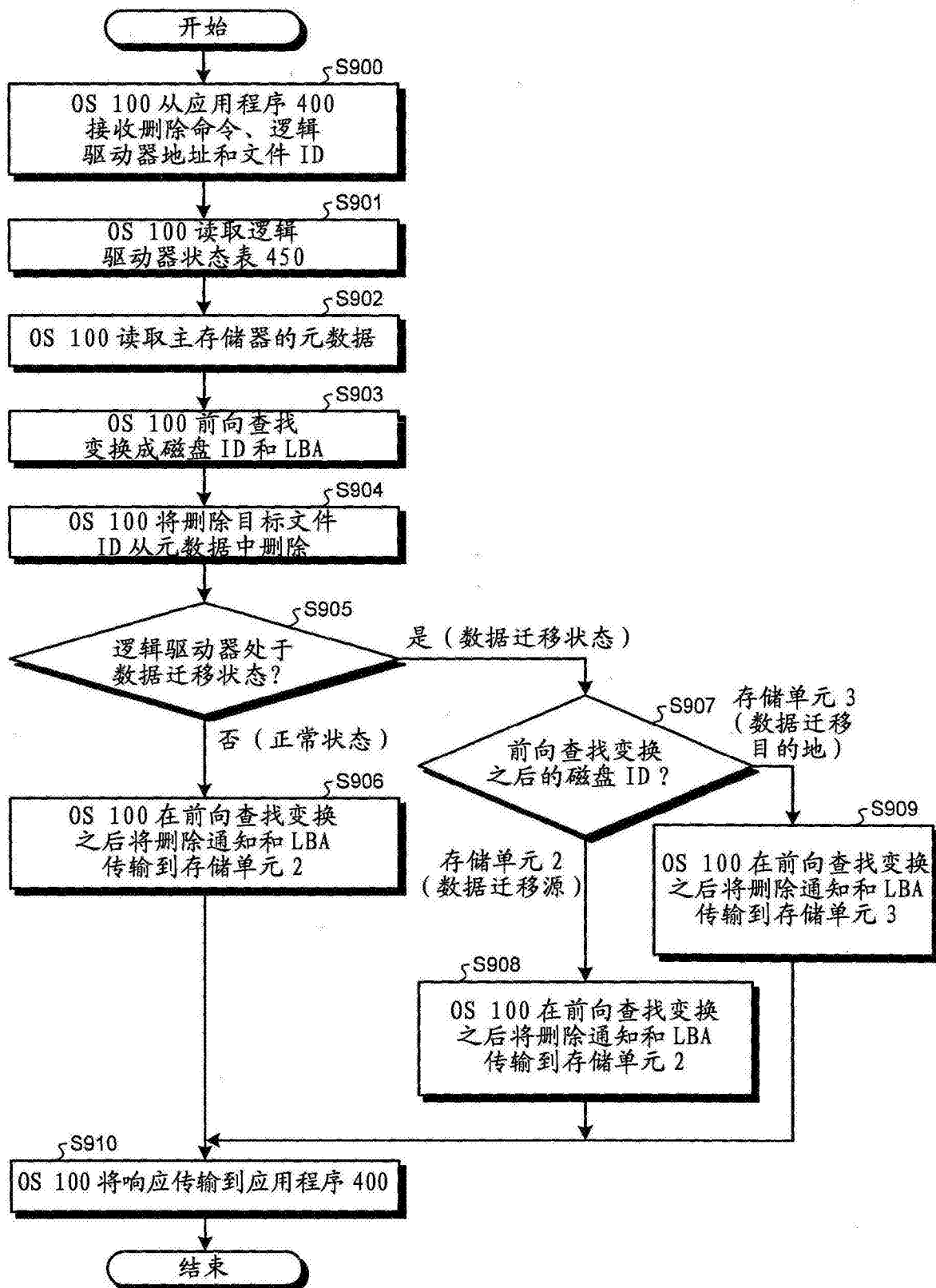


图43

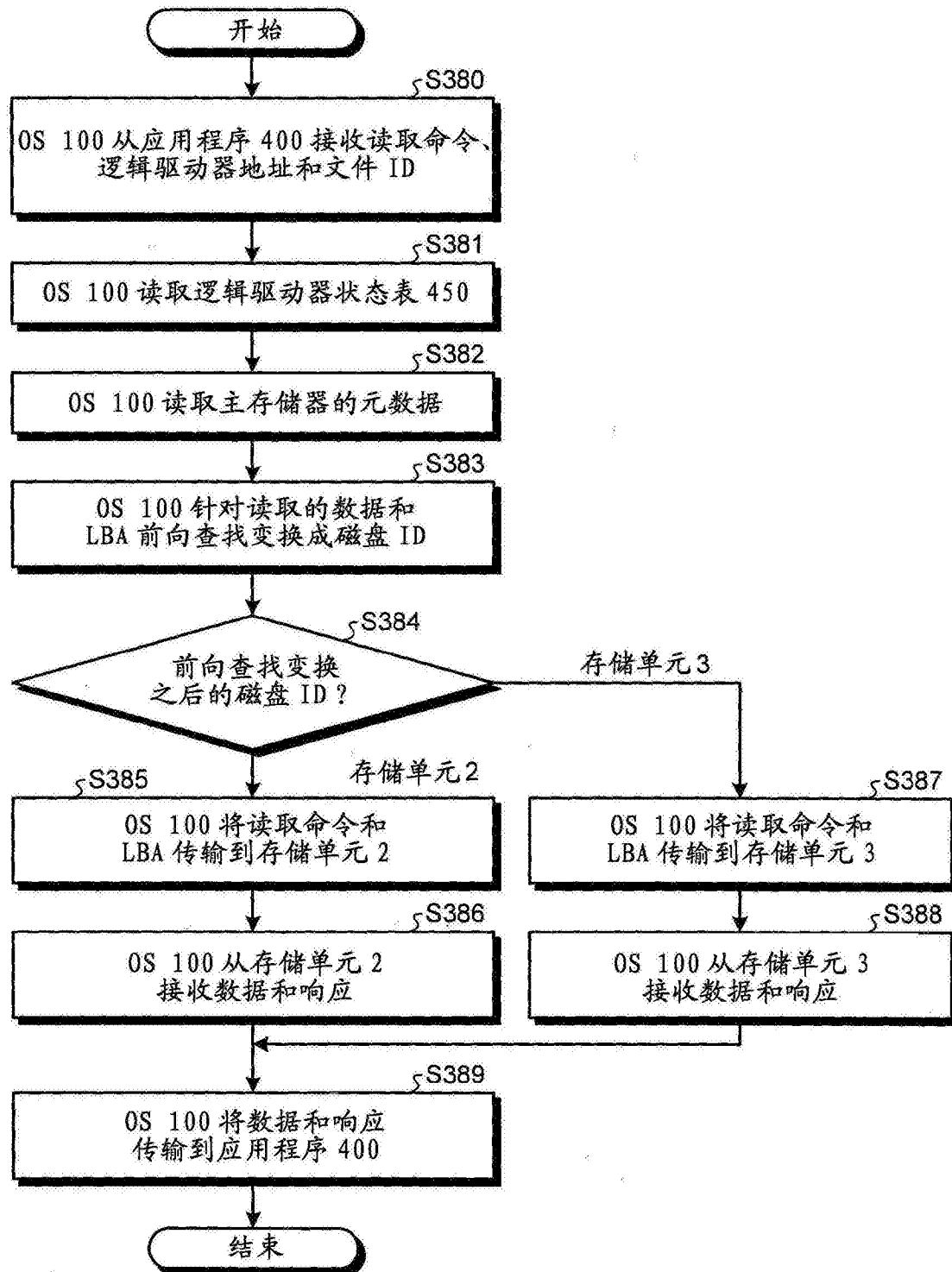


图44

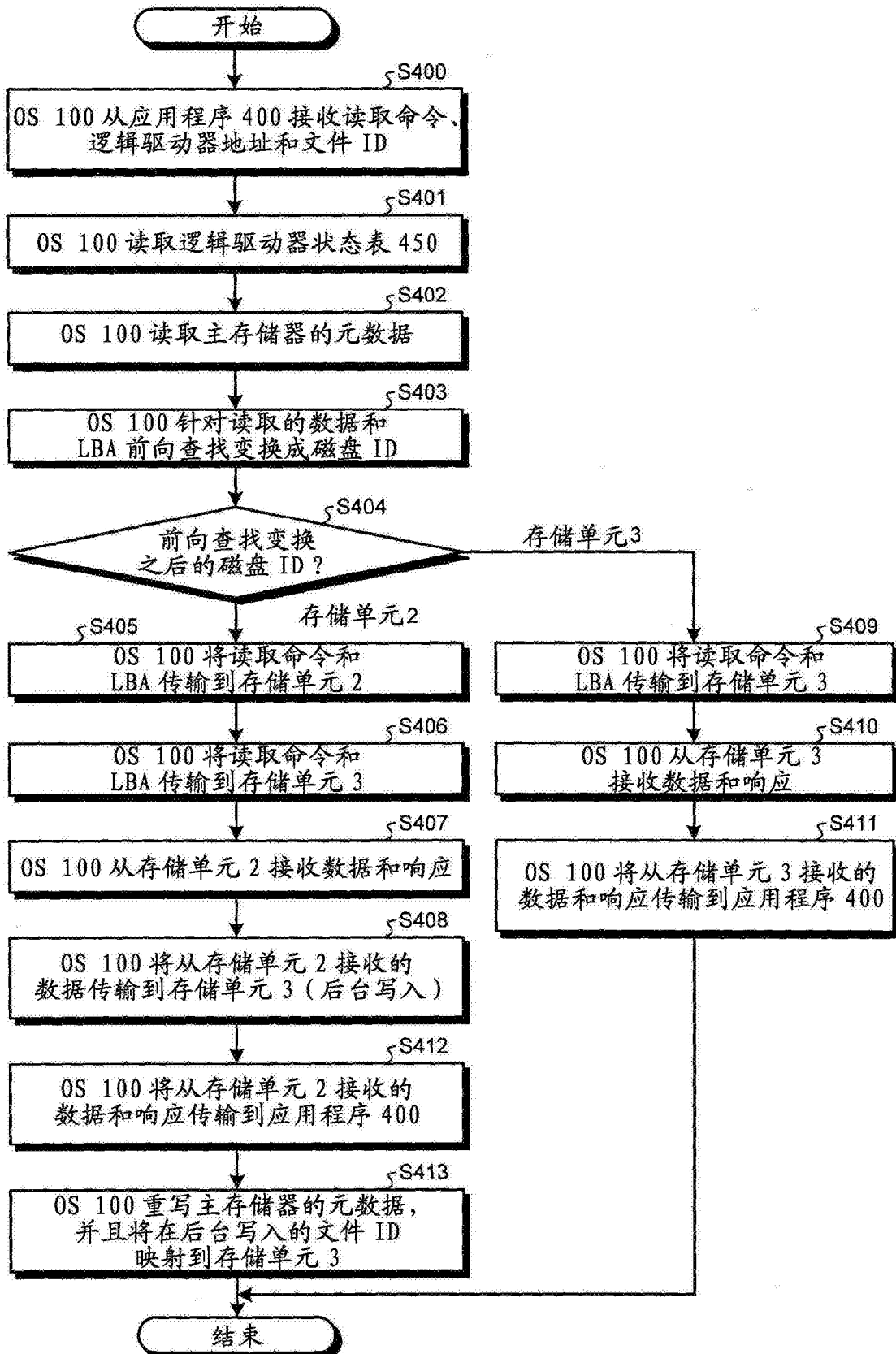


图45

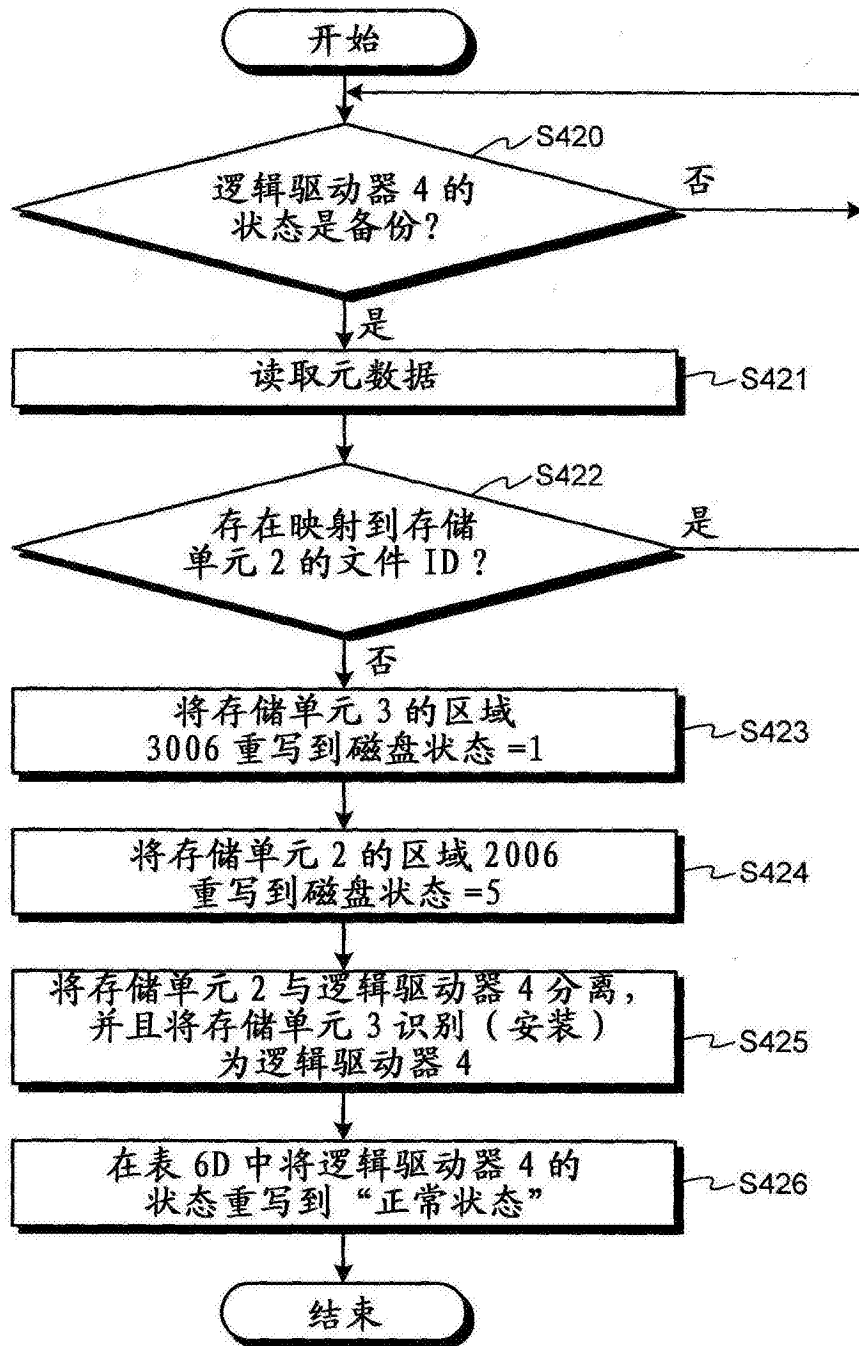


图46

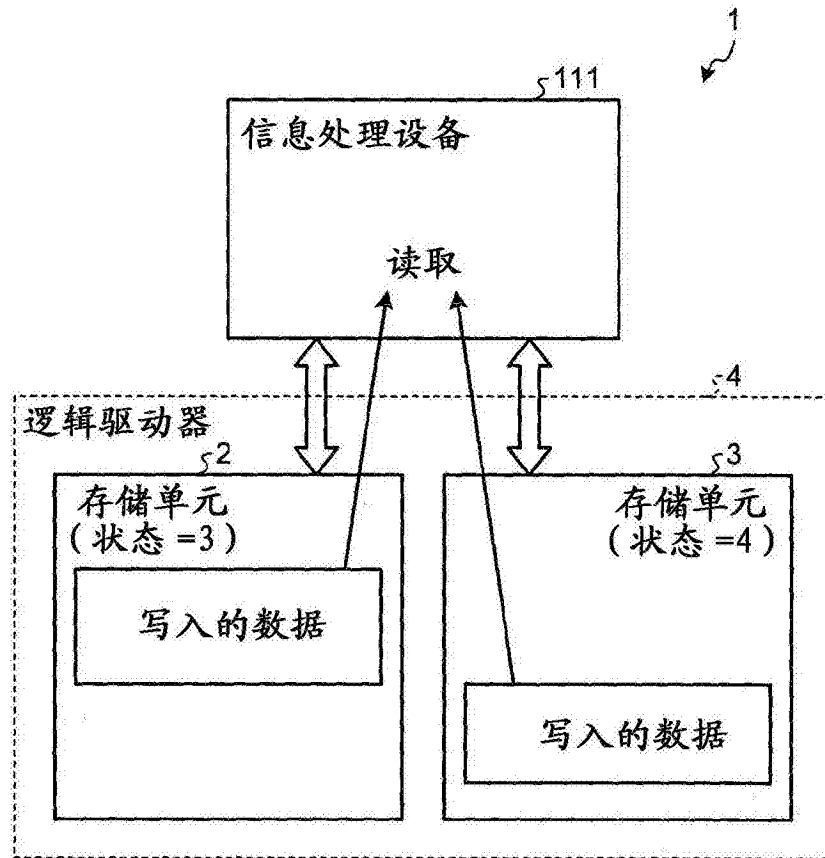


图47

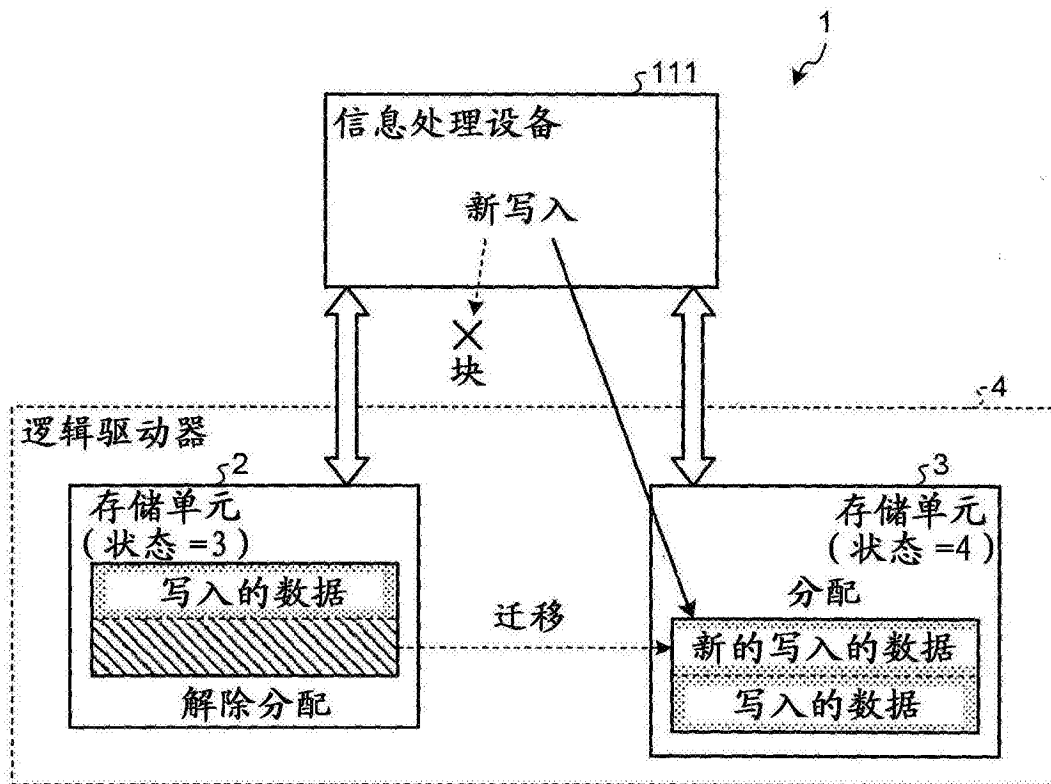


图48

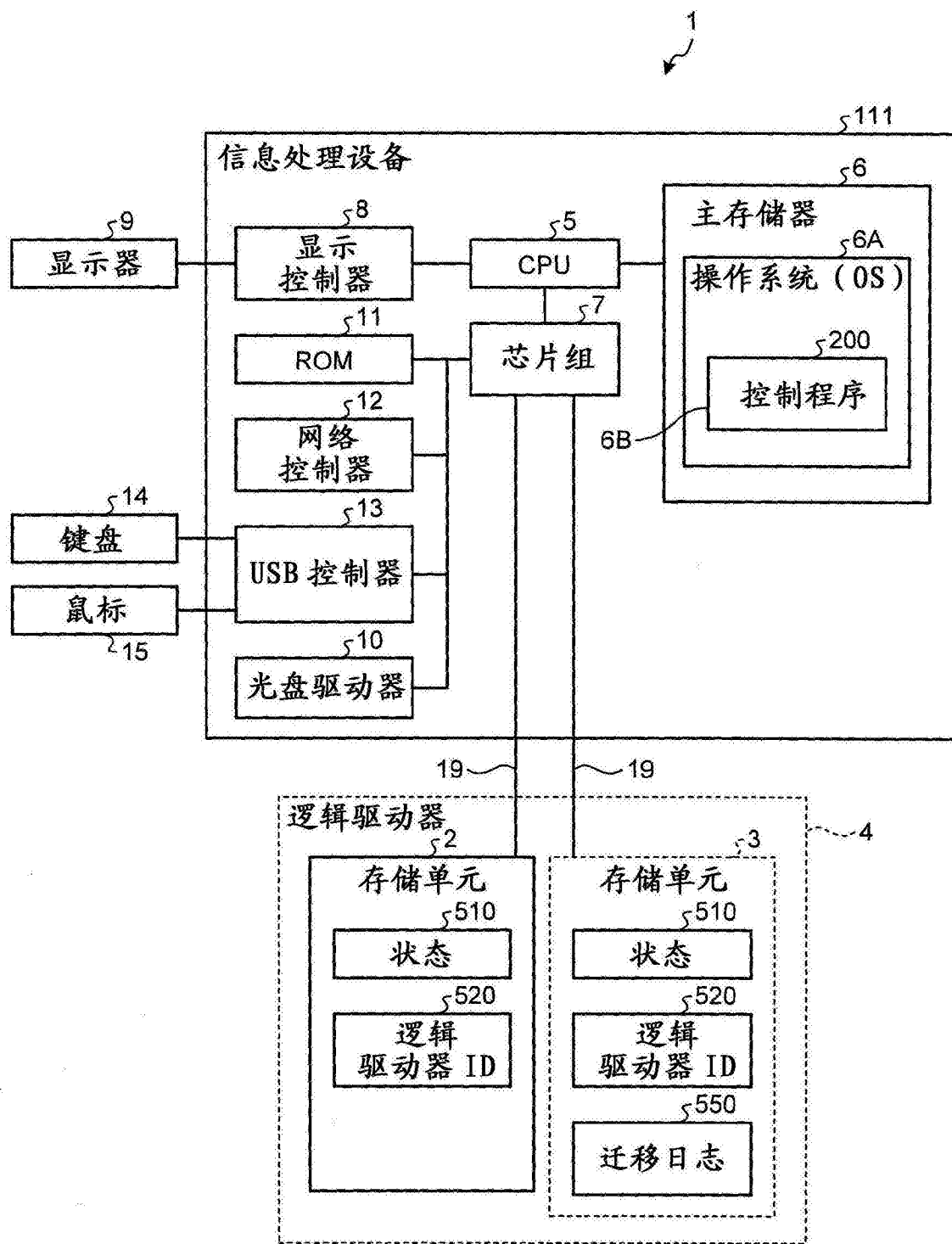


图49

550

迁移日志	
LBA	扇区计数
...	...
...	...
...	...
...	...

图50

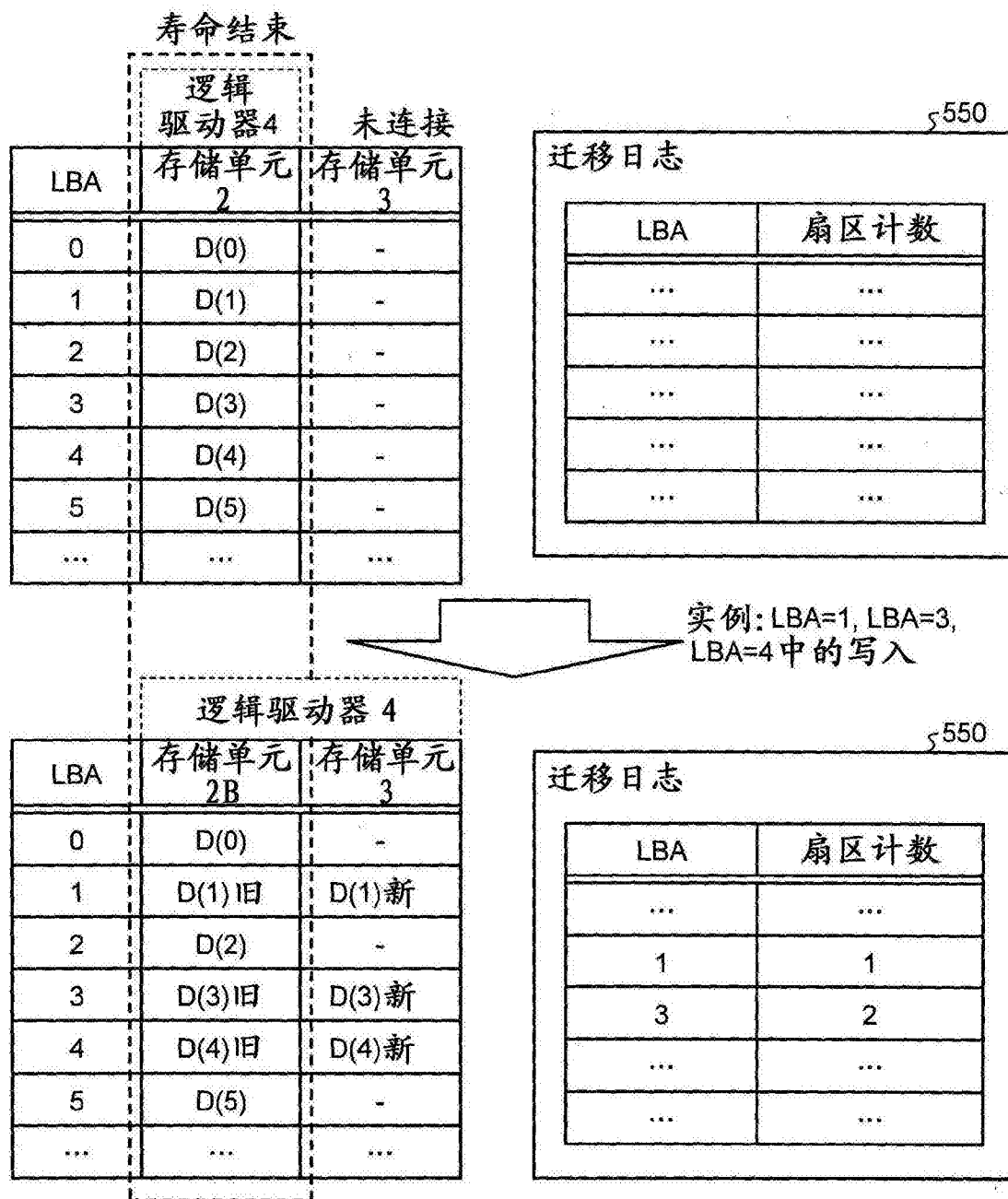


图51

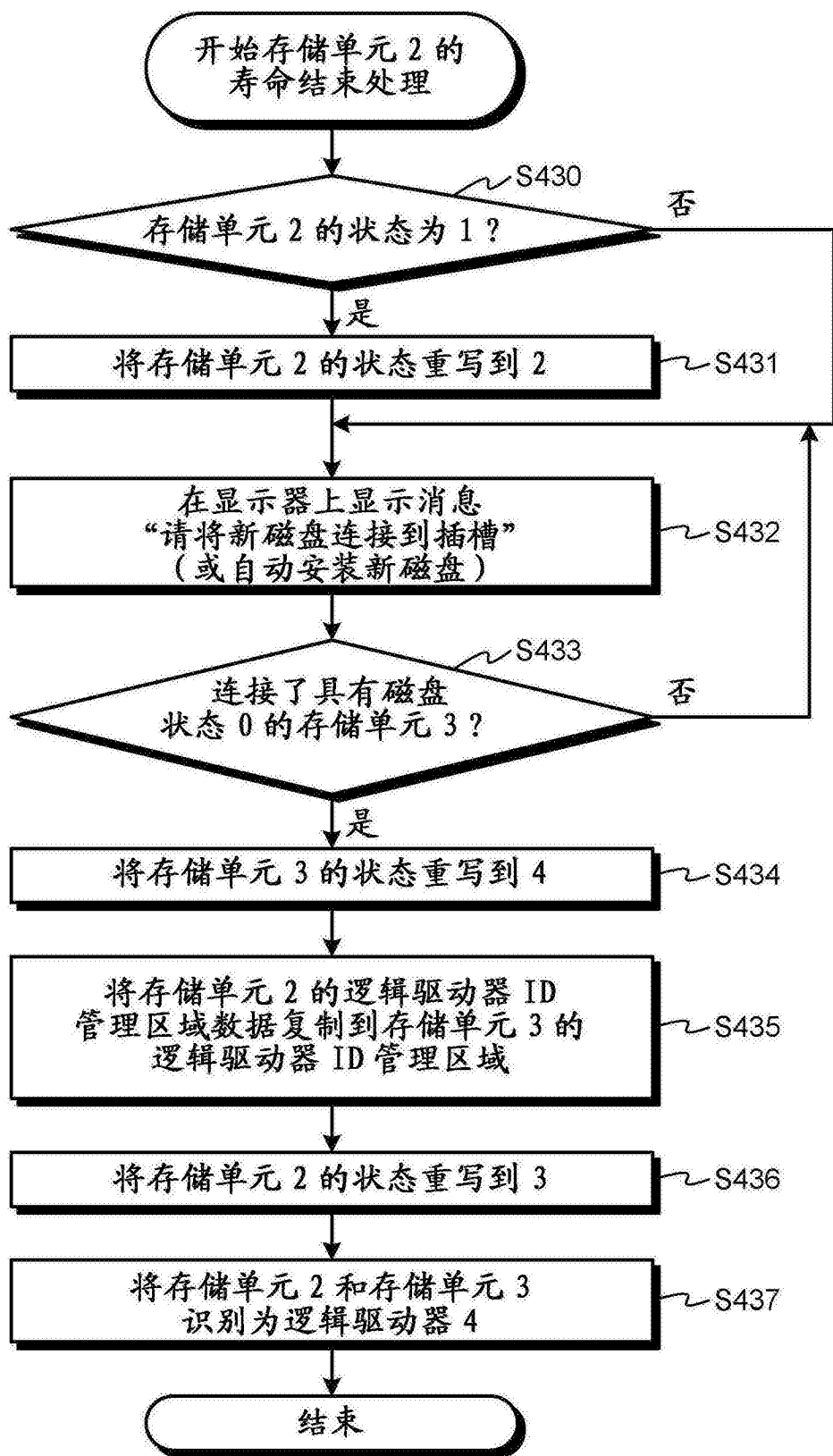


图52

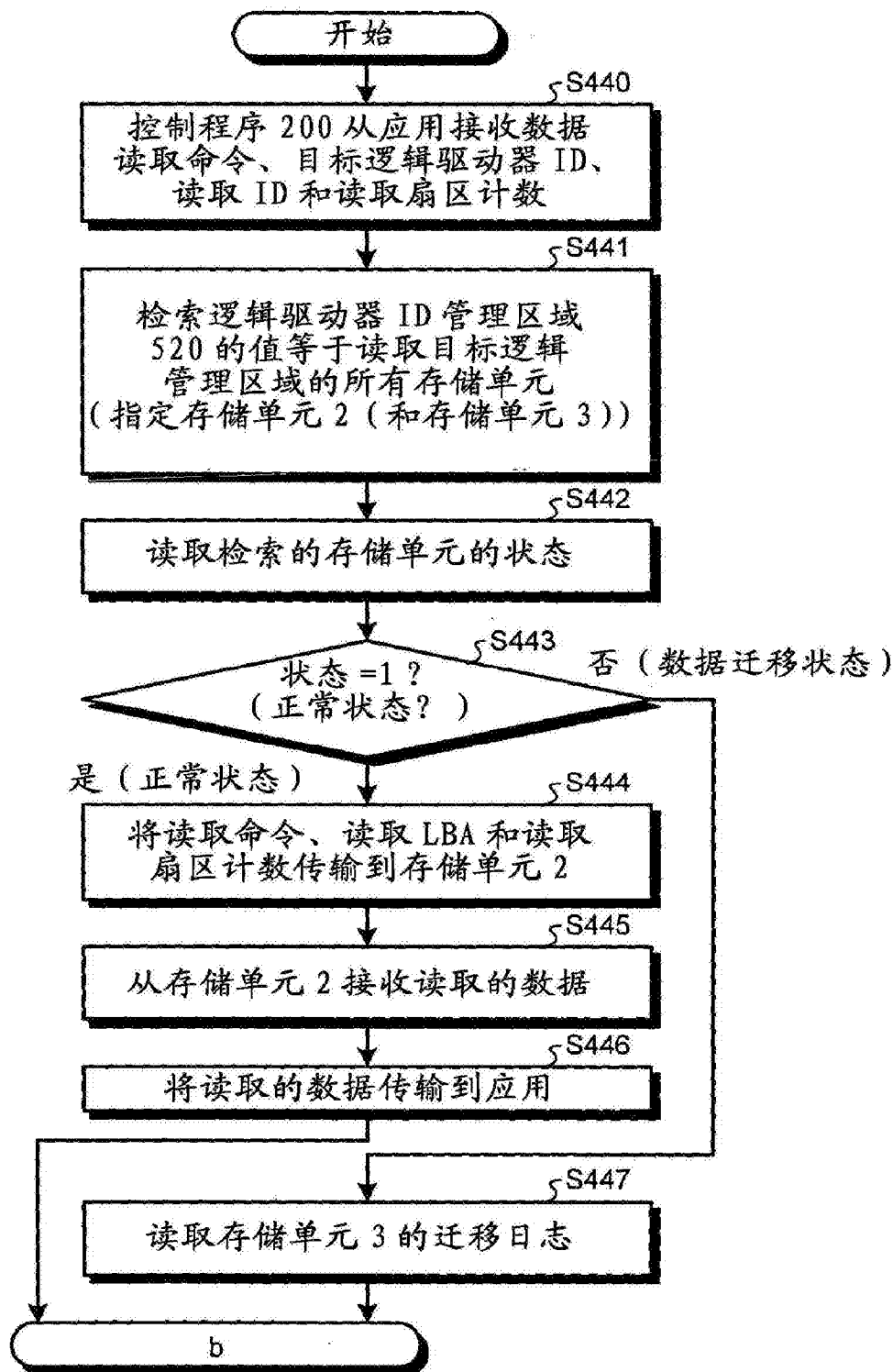


图53A

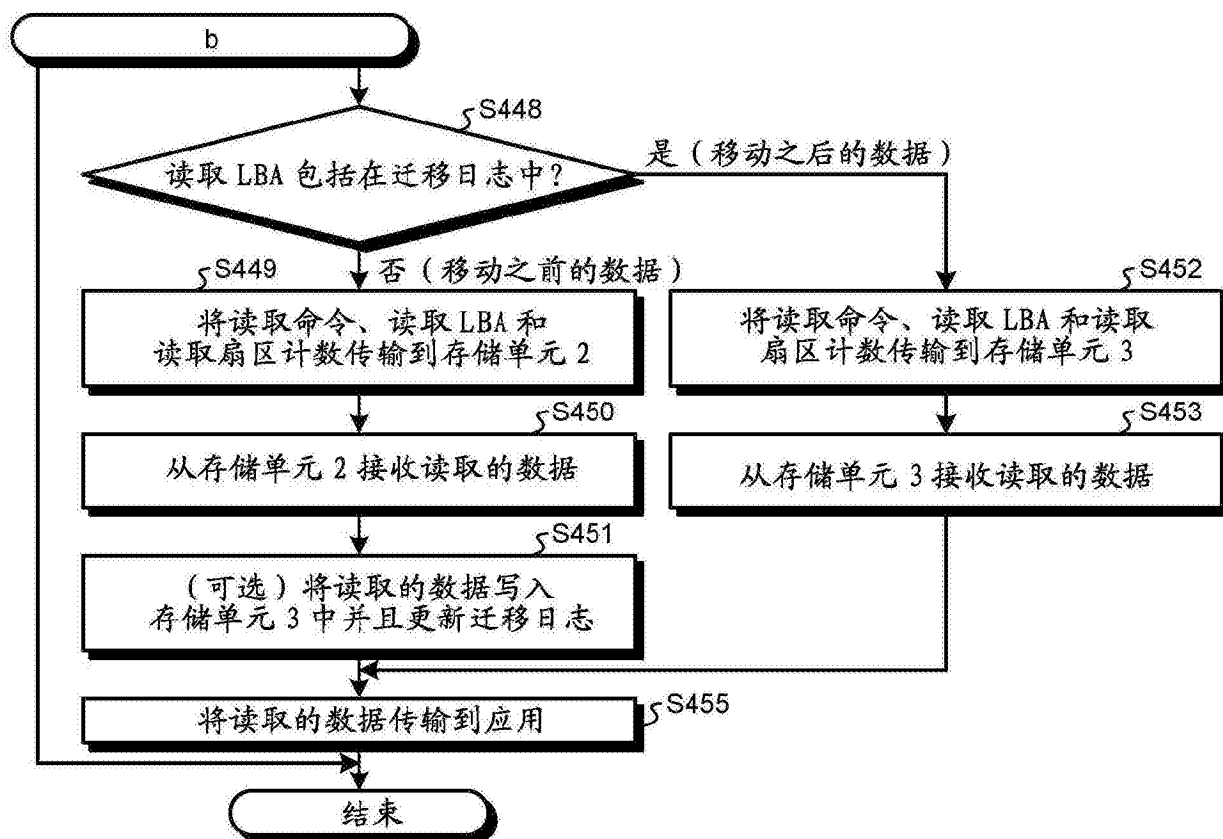


图53B

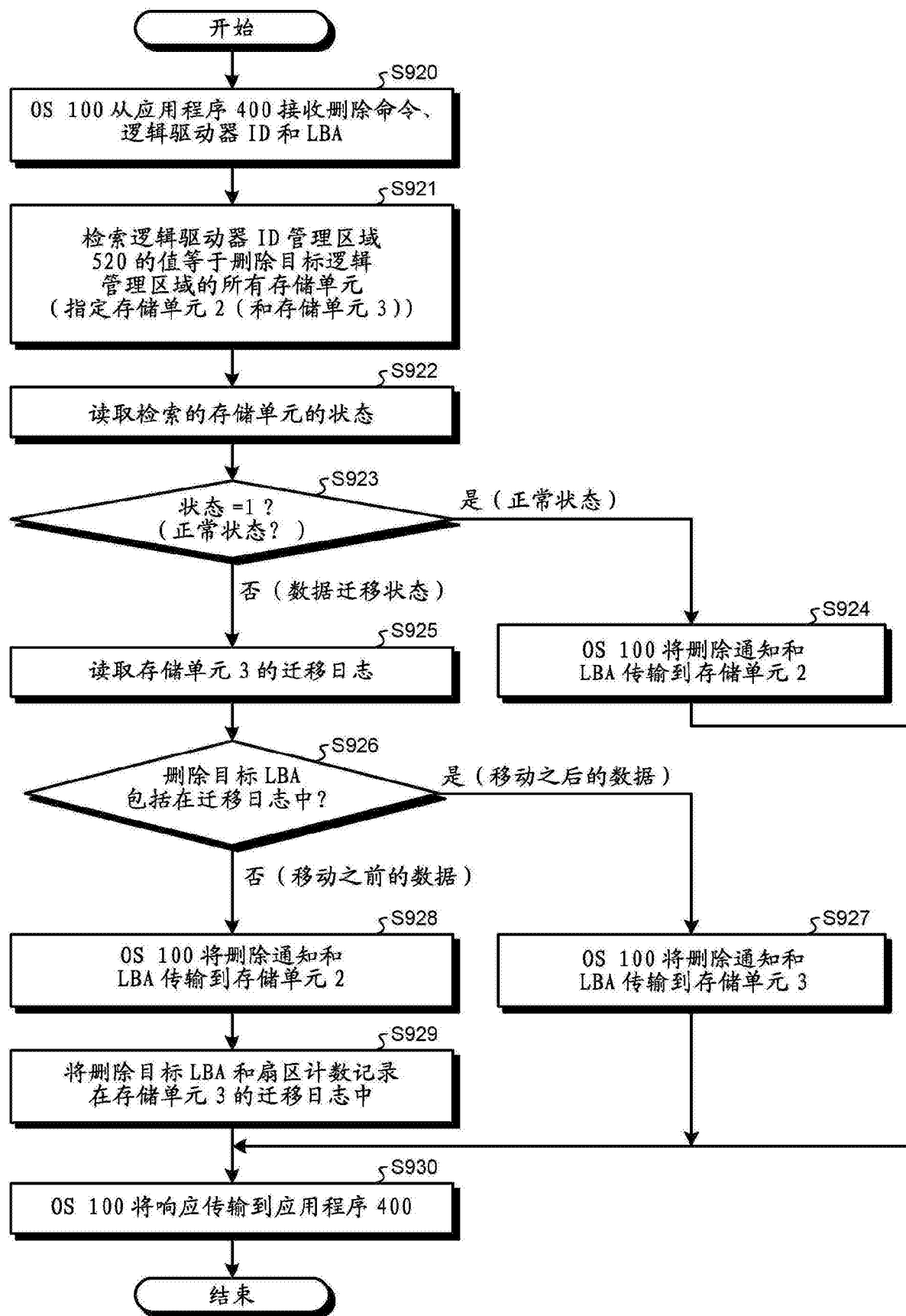


图54

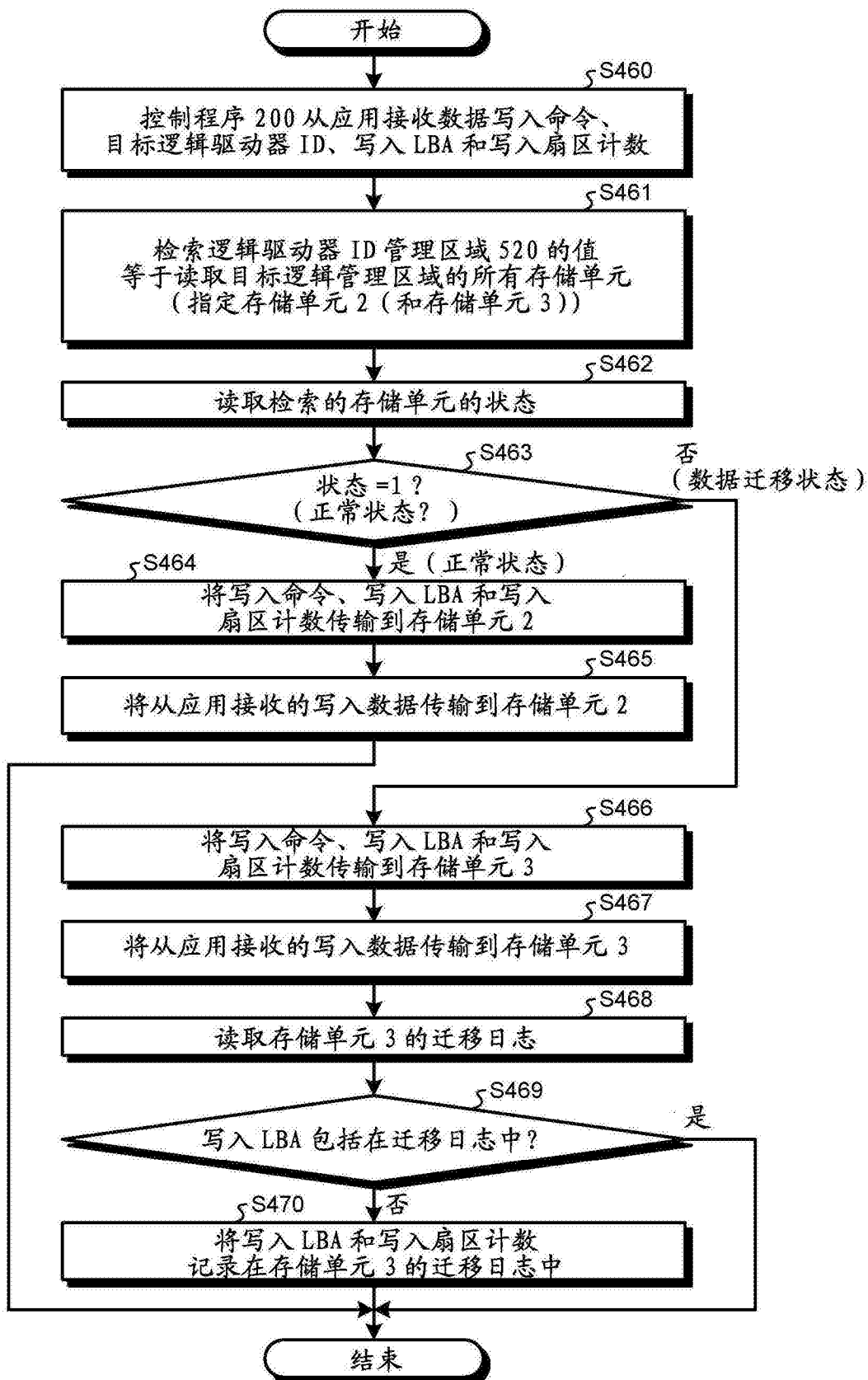


图55

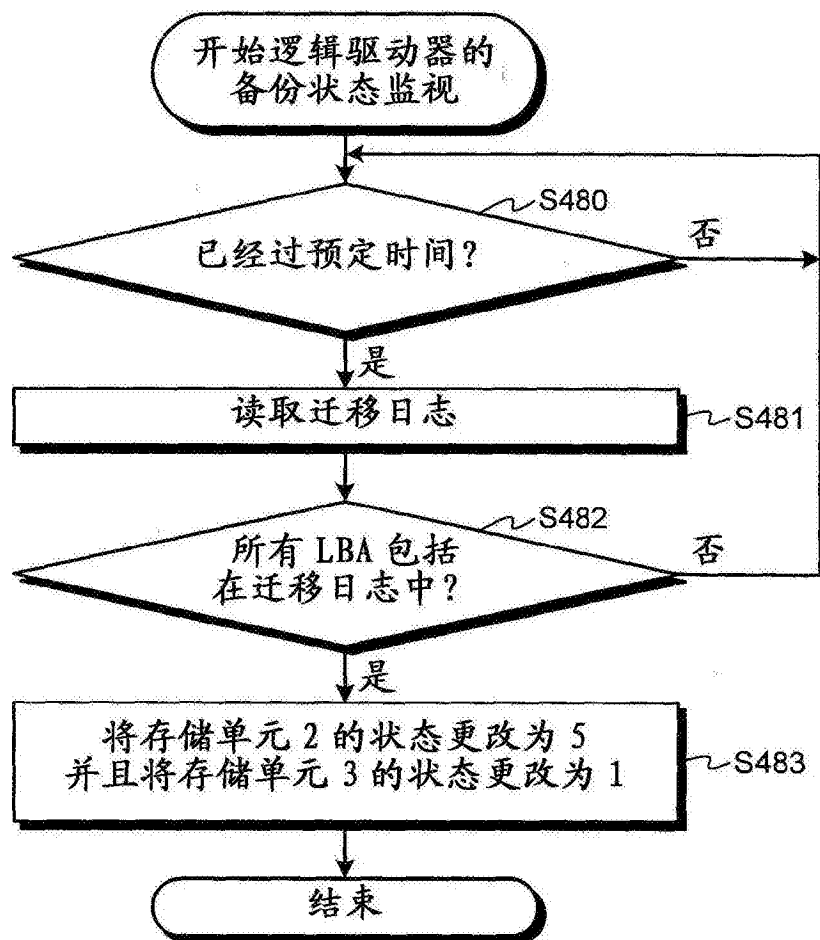


图56

步骤	逻辑驱动器 4		状态
	存储单元 2	存储单元 3	
1	1	未连接	在可靠性劣化之前
2	2	未连接	请求连接存储单元 3
3	2	0	刚好在连接存储单元 3 之后
4	3	4	处于数据迁移
5	5	1	数据迁移完成

图57

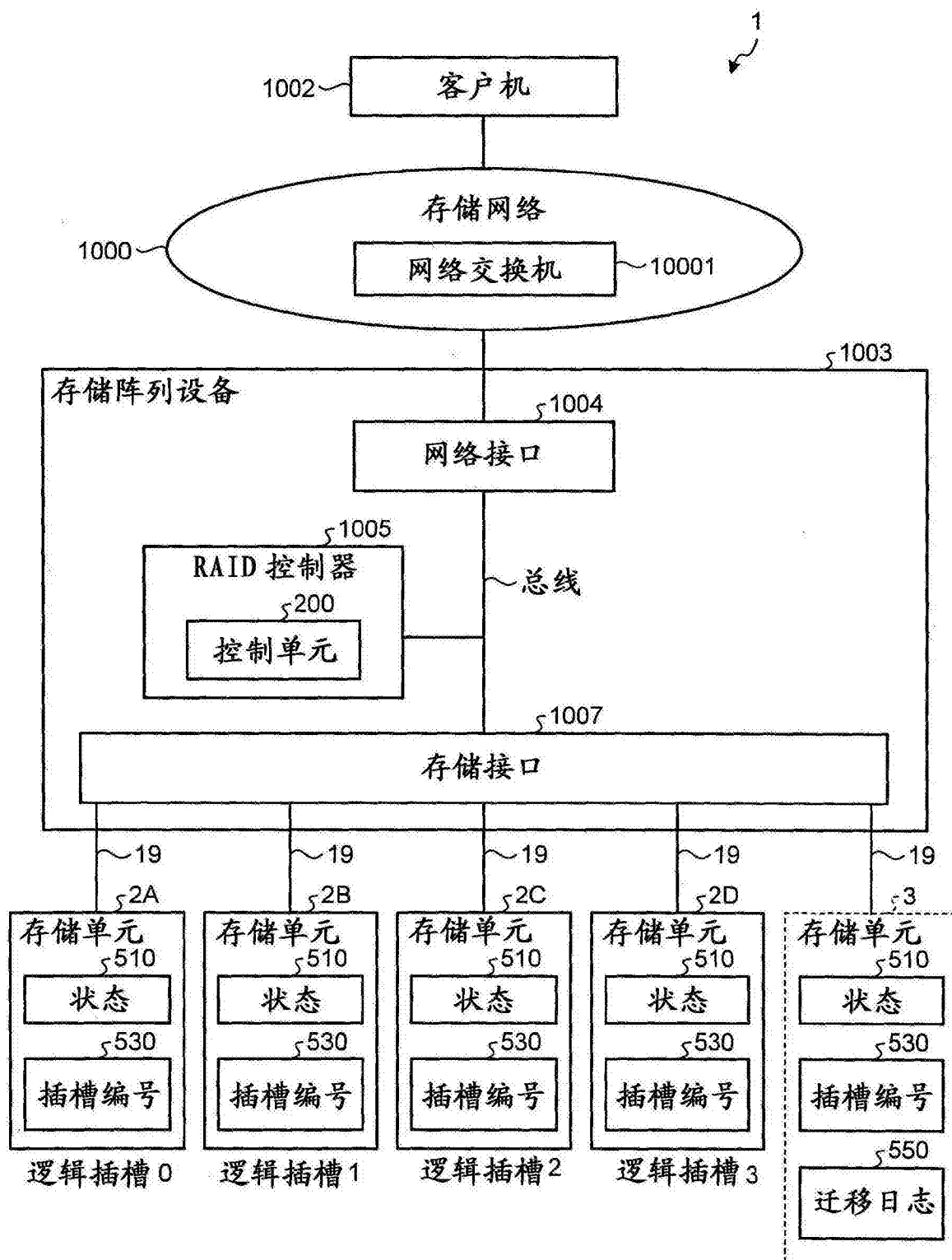


图58

逻辑插槽 逻辑插槽 逻辑插槽 逻辑插槽				
0 1 2 3				
存储单元 LBA	存储单元 2A	存储单元 2B	存储单元 2C	存储单元 2D
0	D(0)	D(1)	D(2)	P(0, 2)
1	D(3)	D(4)	P(3, 5)	D(5)
2	D(6)	P(6, 8)	D(7)	D(8)
3	P(9, 11)	D(9)	D(10)	D(11)
4	D(12)	D(13)	D(14)	P(12, 14)
5	D(15)	D(16)	P(15, 17)	D(17)
...	...	...	...	...

图59

逻辑插槽 逻辑插槽 逻辑插槽 逻辑插槽 逻辑插槽					
0 1 2 3 1					
存储单元 LBA	存储单元 2A	存储单元 2B	存储单元 2C	存储单元 2D	存储单元 3
0	D(0)	D(1)	D(2)	P(0, 2)	-
1	D(3)	D(4)	P(3, 5)	D(5)	-
2	D(6)	P(6, 8)	D(7)	D(8)	-
3	P(9, 11)	D(9)	D(10)	D(11)	-
4	D(12)	D(13)	D(14)	P(12, 14)	-
...	...	...	...	...	...

图60

550

迁移日志	
存储单元 LBA	扇区计数
...	...
...	...
...	...
...	...

图61

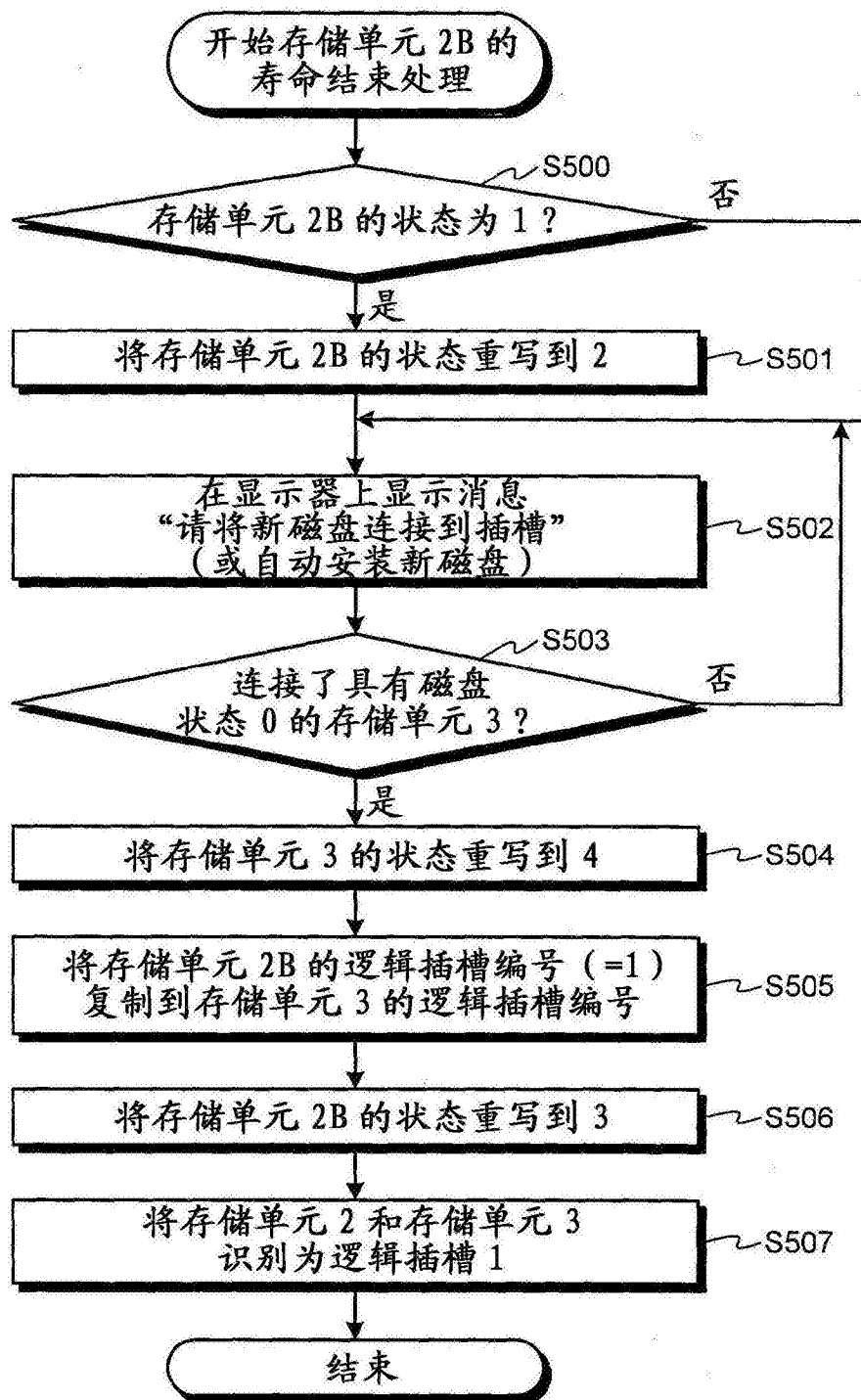


图62

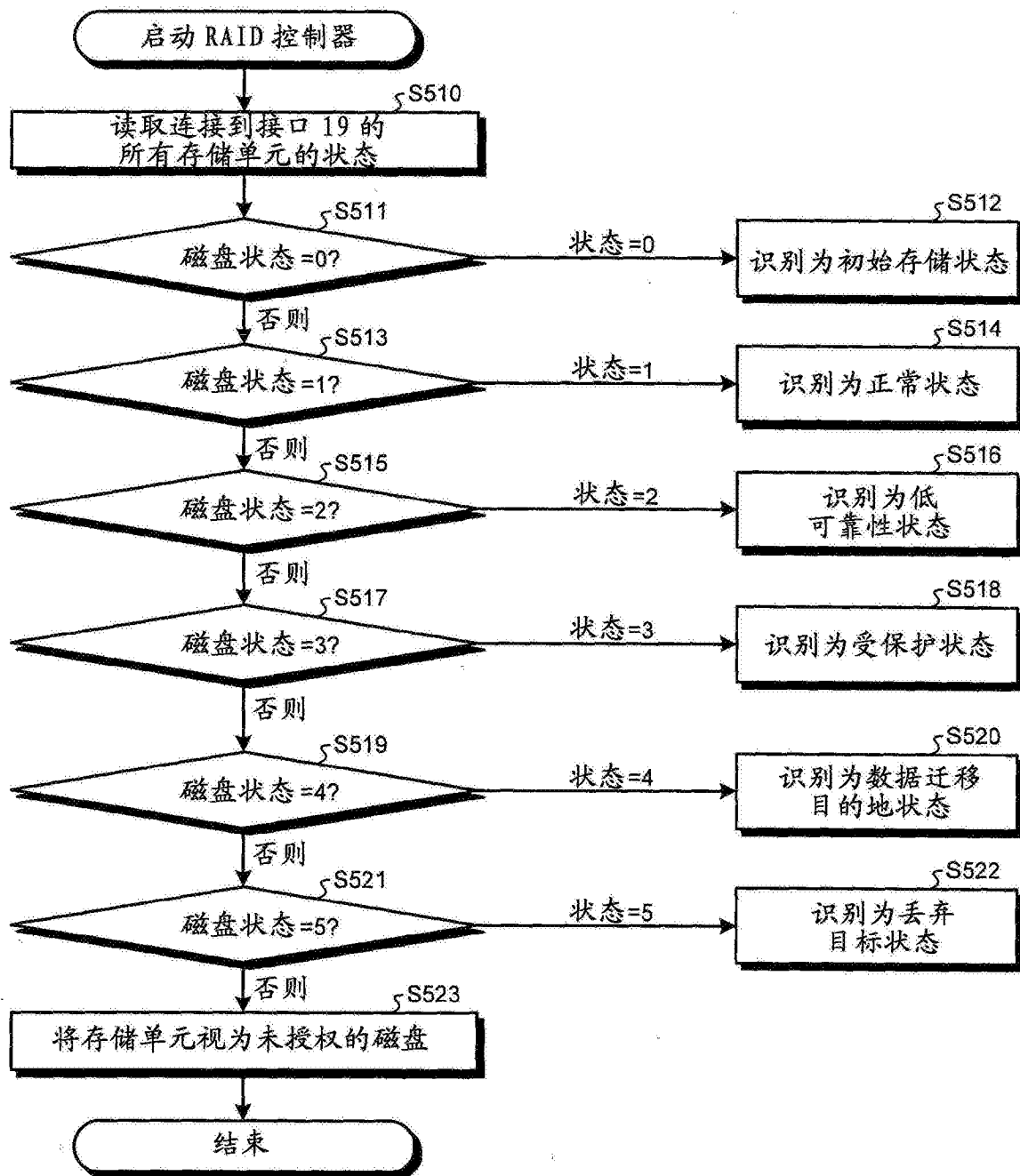


图63

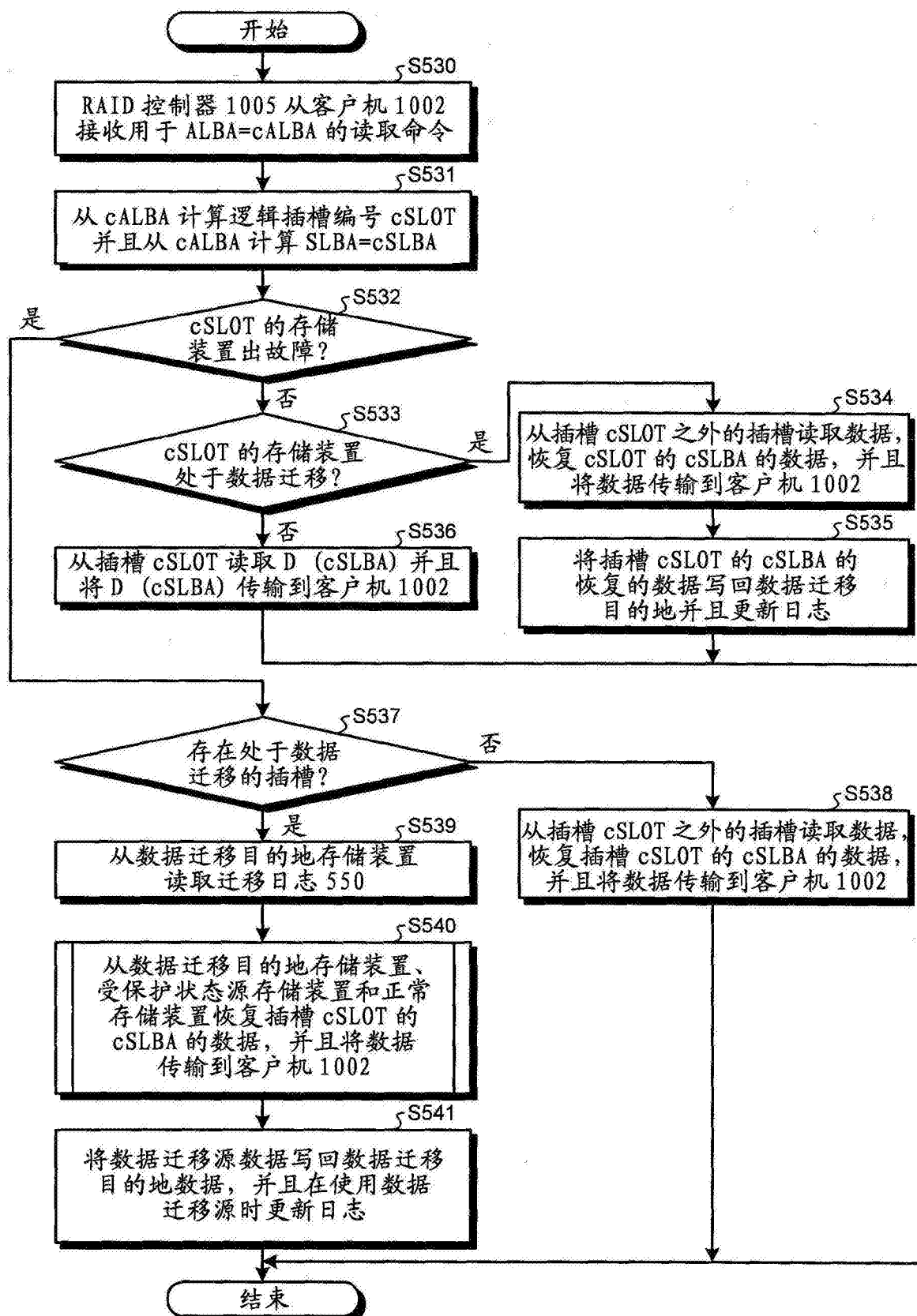


图64

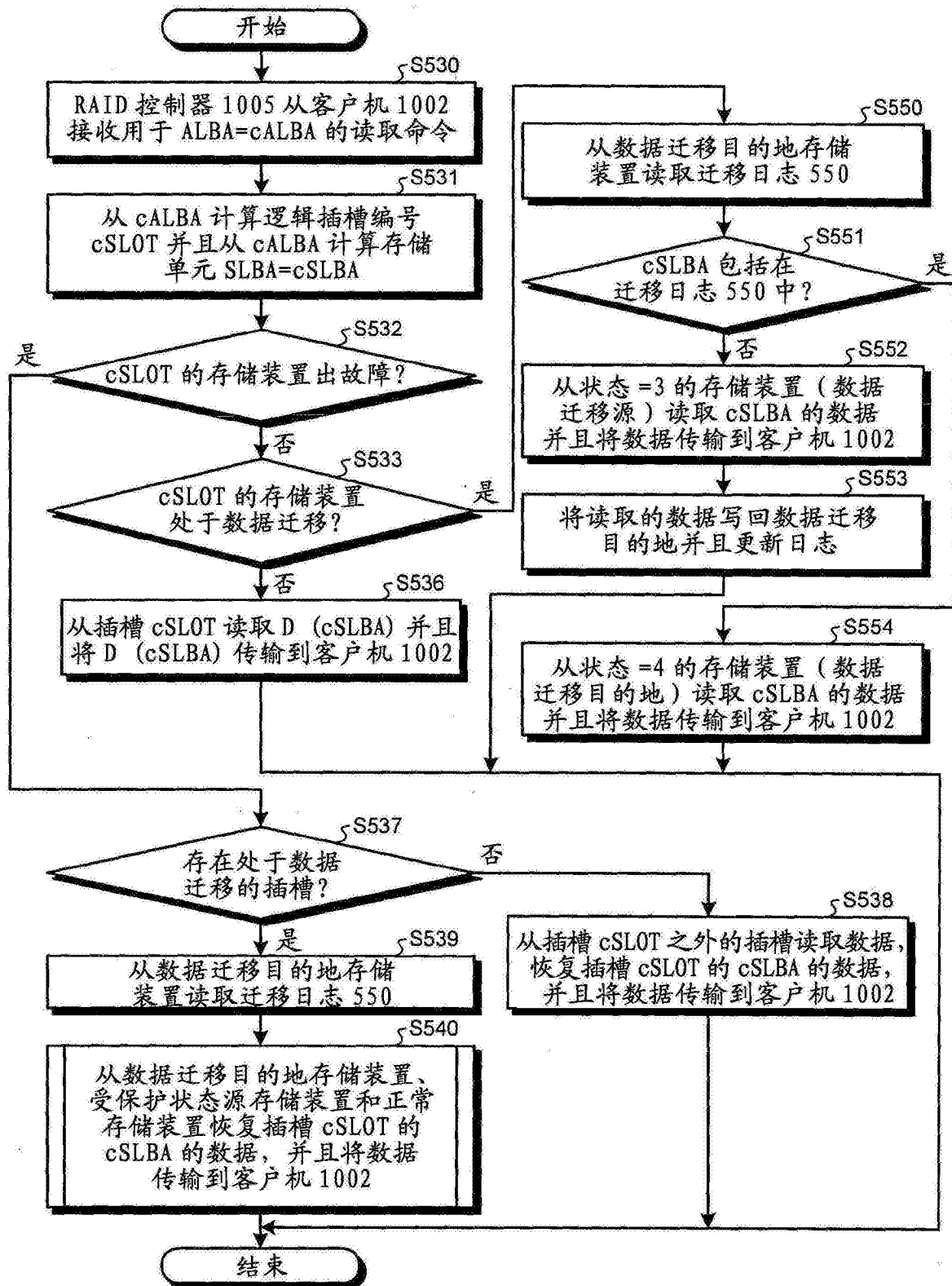


图65

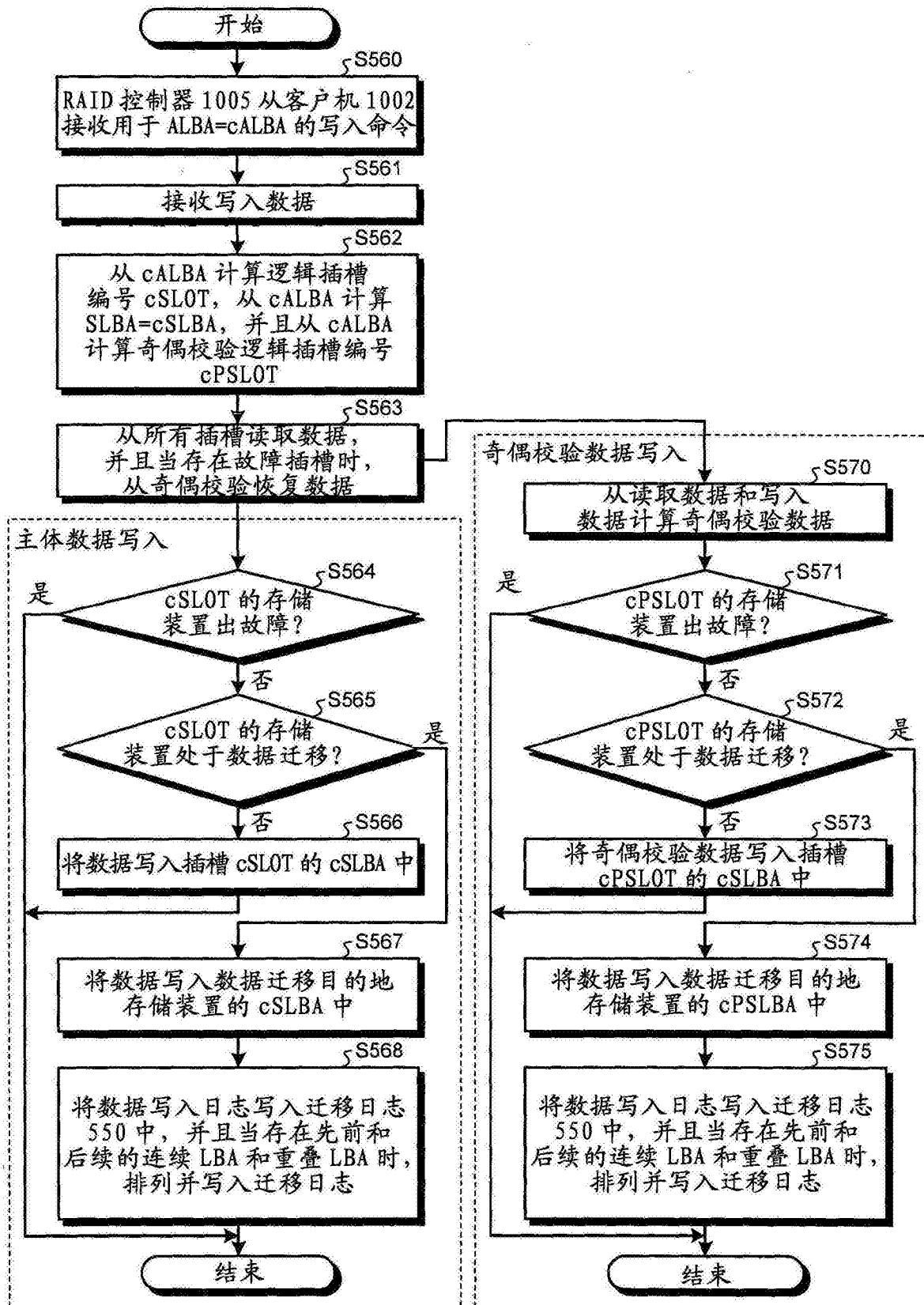


图66

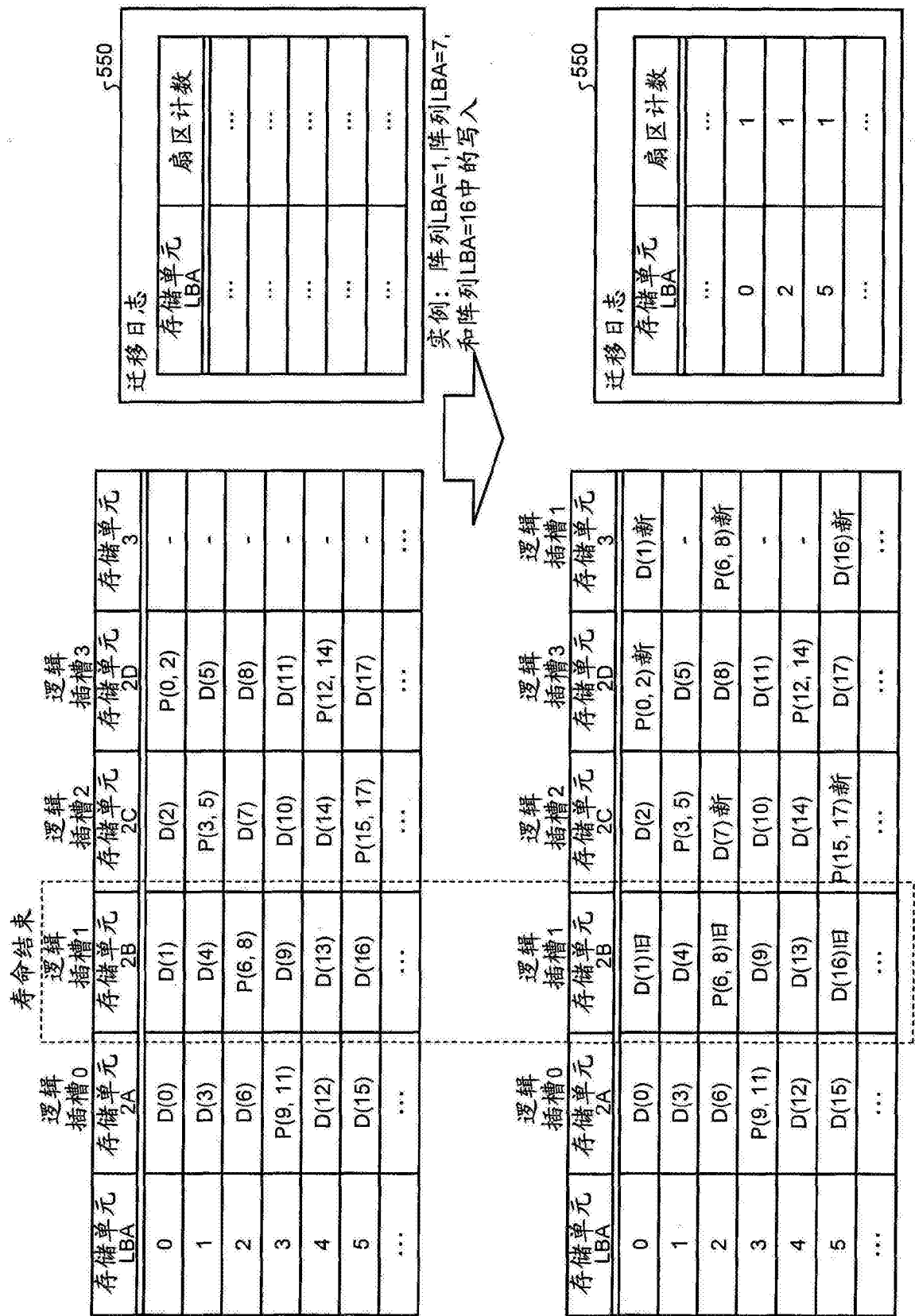


图67

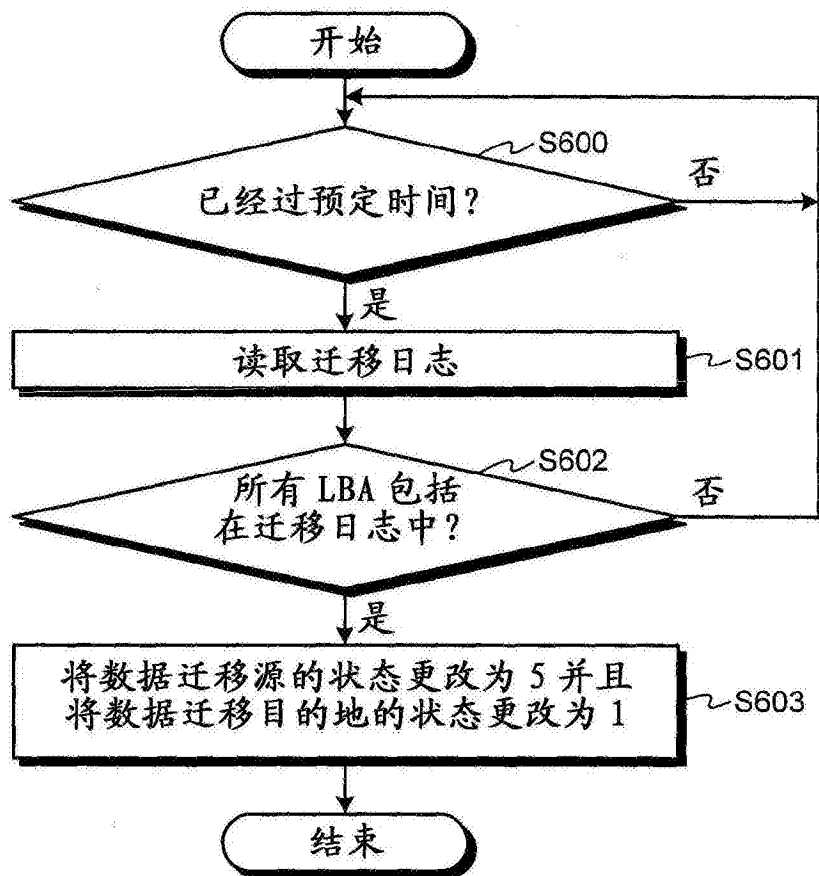


图68

步骤	逻辑 插槽0 存储单元 2A	逻辑 插槽1 存储单元 2B	逻辑 插槽2 存储单元 2C	逻辑 插槽3 存储单元 2D	逻辑 插槽1 存储单元 3	状态
1	1	1	1	1	未连接	在可靠劣化之前
2	1	2	1	1	未连接	请求连接存储单元 3
3	1	2	1	1	0	刚好在连接存储单元 3 之后
4	1	3	1	1	4	处于数据迁移 优先地执行从存储单元 2A/2C/2D 的 读取, 并且在适当时从奇偶校验恢复 存储单元 2B 的 LBA 数据
5	1	5	1	1	1	数据迁移完成

图69

可靠性劣化的数据迁移源			新磁盘数据迁移目的地址逻辑插槽1			读取期间的行为
逻辑插槽0	存储单元2A	存储单元2B	逻辑插槽2	存储单元2D (故障)	逻辑插槽3	
存储单元1BA					存储单元3	
0	D(0)	D(1)旧	D(2)	P(0, 2)新 (故障)	D(1)新	D(0)、D(2) 不需要恢复 从数据迁移目的地存储单元3读取D(1)
1	D(3)	D(4)	P(3, 5)	D(5) (故障)	-	D(3)、D(4) 不需要恢复 通过对D(3)、D(4)和P(3, 5) 进行异或而恢复D(5)
2	D(6)	P(6, 8)旧	D(7)新	D(8) (故障)	P(6, 8)新	D(6)、D(7) 不需要恢复 通过对D(6)、D(7)新和P(6, 8) 新进行异或而恢复D(8)
3	P(9, 11)	D(9)	D(10)	D(11) (故障)	-	D(9)、D(10) 不需要恢复 通过对D(9)、D(10)和P(9, 11) 进行异或而恢复D(11)
4	D(12)	D(13)	D(14)	P(12, 14) (故障)	-	不需要恢复
5	D(15)	D(16)旧	P(15, 17)新	D(17) (故障)	D(16)新	D(15) 不需要恢复 从数据迁移目的地存储单元3B读取D(16) 通过对D(15)、D(16)新和P(15, 17) 新进行异或而恢复D(17)
...	...	...	...	...	...	

图70

存储 单元LBA	可靠性 劣化的 备份源			可靠性 劣化的 备份源			新磁盘 数据迁移 目的地			读取期间的行为
	逻辑 插槽0	存储 单元2A	存储 单元2B	逻辑 插槽1	存储 单元2C	存储单元2D (故障)	逻辑 插槽2	存储 单元3B	存储 单元3C	
0		D(0)	D(1)旧		D(2)	P(0, 2)新 (故障)		D(1)新	-	D(0)、D(2) 不需要恢复 从数据迁移目的地3B读取D(1)
1		D(3)	D(4)		P(3, 5)	D(5) (故障)		-	-	D(3)、D(4) 不需要恢复 通过对D(3)、D(4)和P(3, 5) 进行异或而恢复D(5)
2		D(6)	P(6, 8)旧		D(7)旧	D(8) (故障)		P(6, 8)新	D(7)新	D(6) 不需要恢复 从数据迁移目的地3C读取D(7) 通过对D(6)、D(7)新和P(6, 8) 新进行异或而恢复D(8)
3		P(9, 11)	D(9)		D(10)	D(11) (故障)		-	-	D(9)、D(10) 不需要恢复 通过对D(9)、D(10)和P(9, 11) 进行异或而恢复D(11)
4		D(12)	D(13)		D(14)	P(12, 14) (故障)		-	-	不需要恢复
5		D(15)	D(16)旧		P(15, 17)旧	D(17) (故障)		D(16)新	P(15, 17) 新	D(15) 不需要恢复 从数据迁移目的地存储单元3B读取D(16) 通过对D(15)、D(16)新和P(15, 17) 新进行异或而恢复D(17)
...		...	...		...	...		...	...	

图71

存储 单元LBA	逻辑 插槽0			数据迁移源 逻辑插槽1			逻辑 插槽2			数据迁移目的地 逻辑插槽3			读取期间的行为
	存储 单元2A	存储 单元2B	存储 单元2C	存储 单元2D	存储 单元2E	存储 单元2F	存储 单元2G	存储 单元2H	存储 单元2I	存储 单元2J	存储 单元2K	存储 单元2L	
0	D(0)	D(1)旧	D(2)UECC	P(0, 2)新	D(1)新	在读取 ALBA=2 中无法进行 ECC 纠正的错误 → 通过对 D(0)、D(1) 新和 P(0, 2) 新进行异或而恢复							
1	D(3)	D(4)	P(3, 5)	D(5)	-								
2	D(6)	P(6, 8)旧	D(7)新	D(8)UECC	P(6, 8)新	在读取 ALBA=7 中无法进行 ECC 纠正的错误 → 通过对 D(6)、D(7) 新和 P(6, 8) 新进行异或而恢复							
3	P(9, 11)	D(9)UECC	D(10)	D(11)	-	在读取 ALBA=9 中无法进行 ECC 纠正的错误 → 通过对 D(10)、D(11) 和 P(9, 11) 进行异或而恢复							
4	D(12)	D(13)	D(14)	P(12, 14)	-								
5	D(15)	D(16)旧	P(15, 17)新	D(17)	D(16)新								
...	...	...	...	...	...								

图72

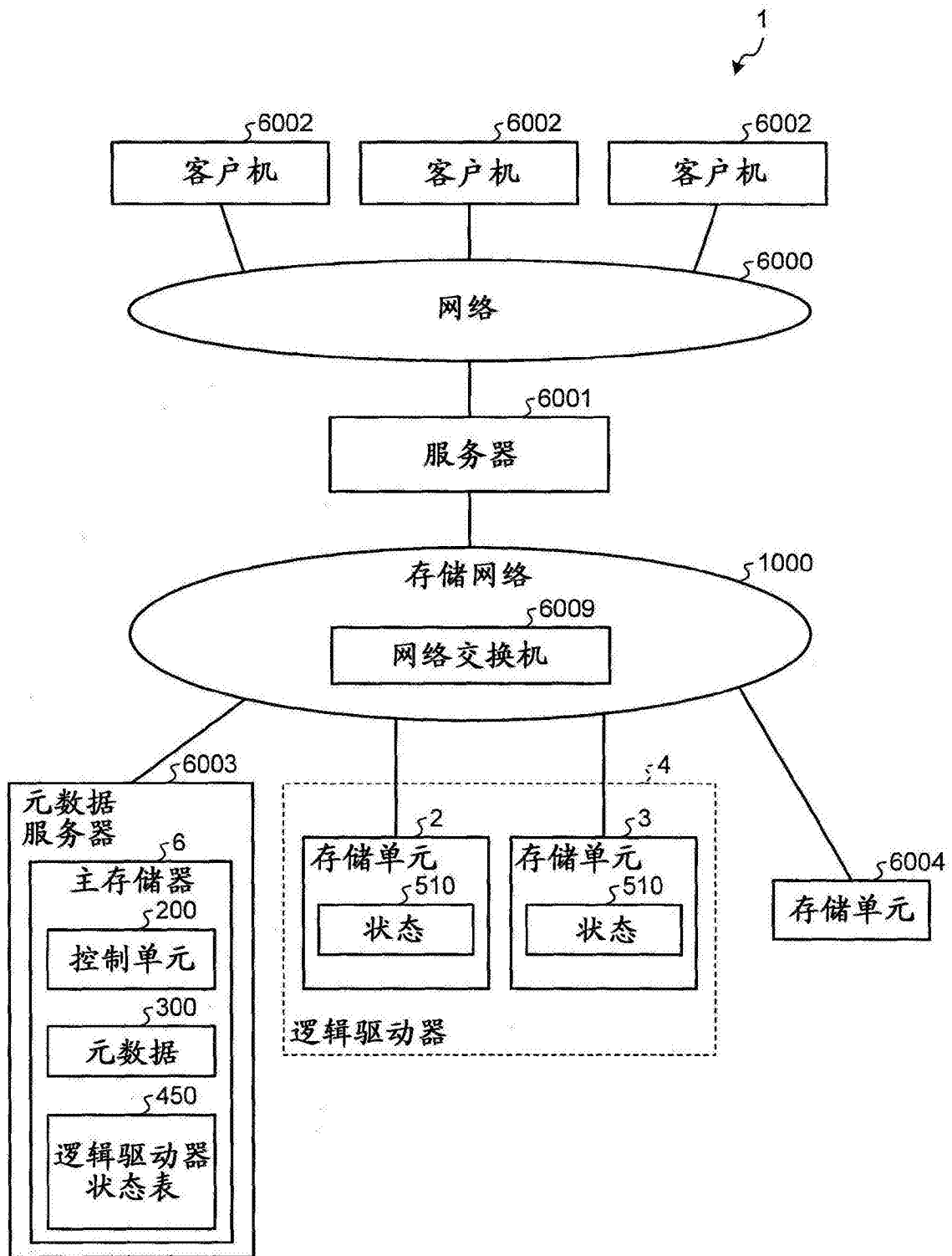


图73

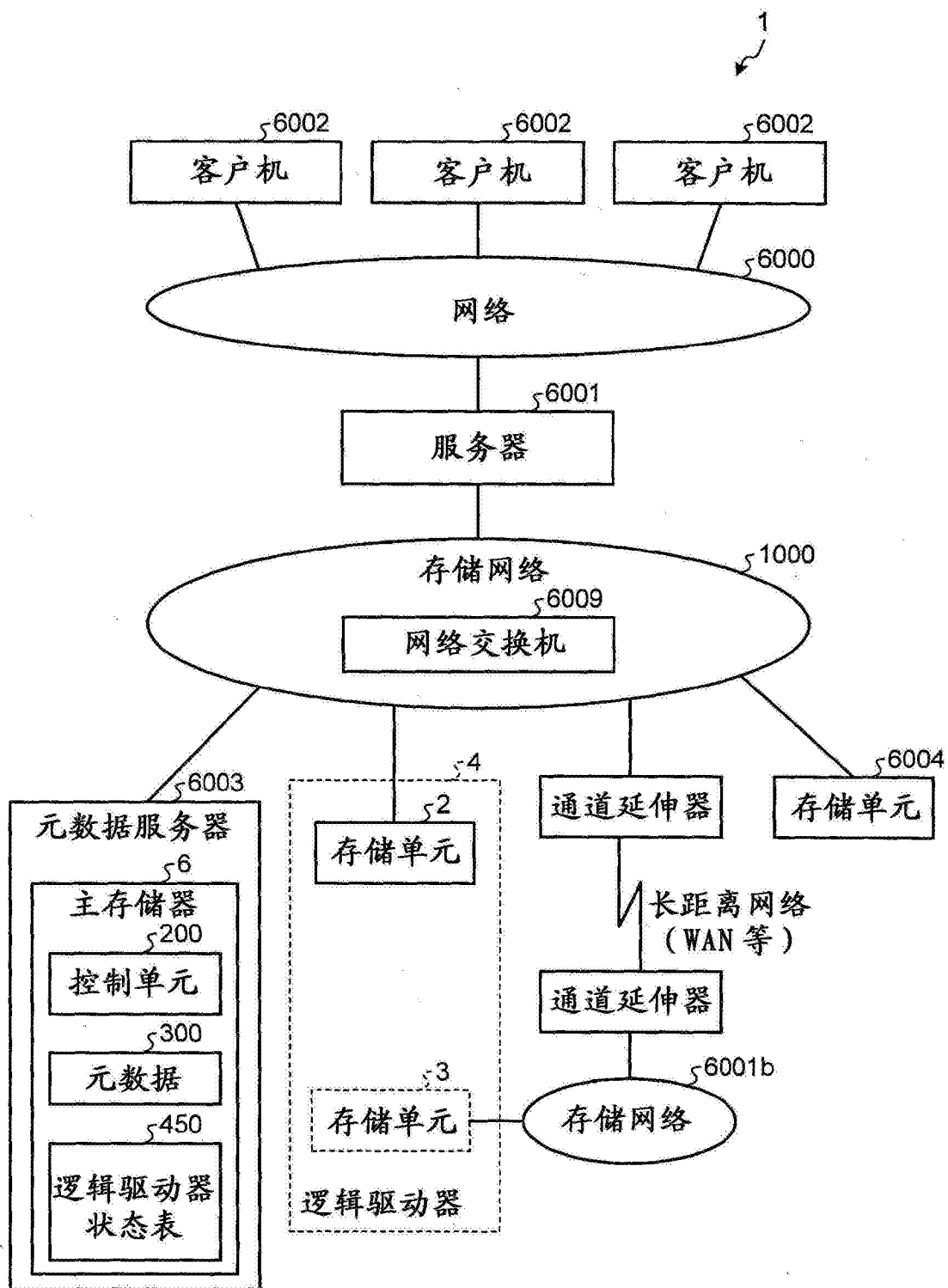


图74

5650

状态表			
逻辑驱动器地址	磁盘ID	逻辑驱动器状态	磁盘状态
A	a	正常状态	1(正常状态)
B	b1	处于数据迁移	3(数据迁移源)
B	b2	处于数据迁移	4(数据迁移目的地)
C	c	正常状态	1(正常状态)

图75

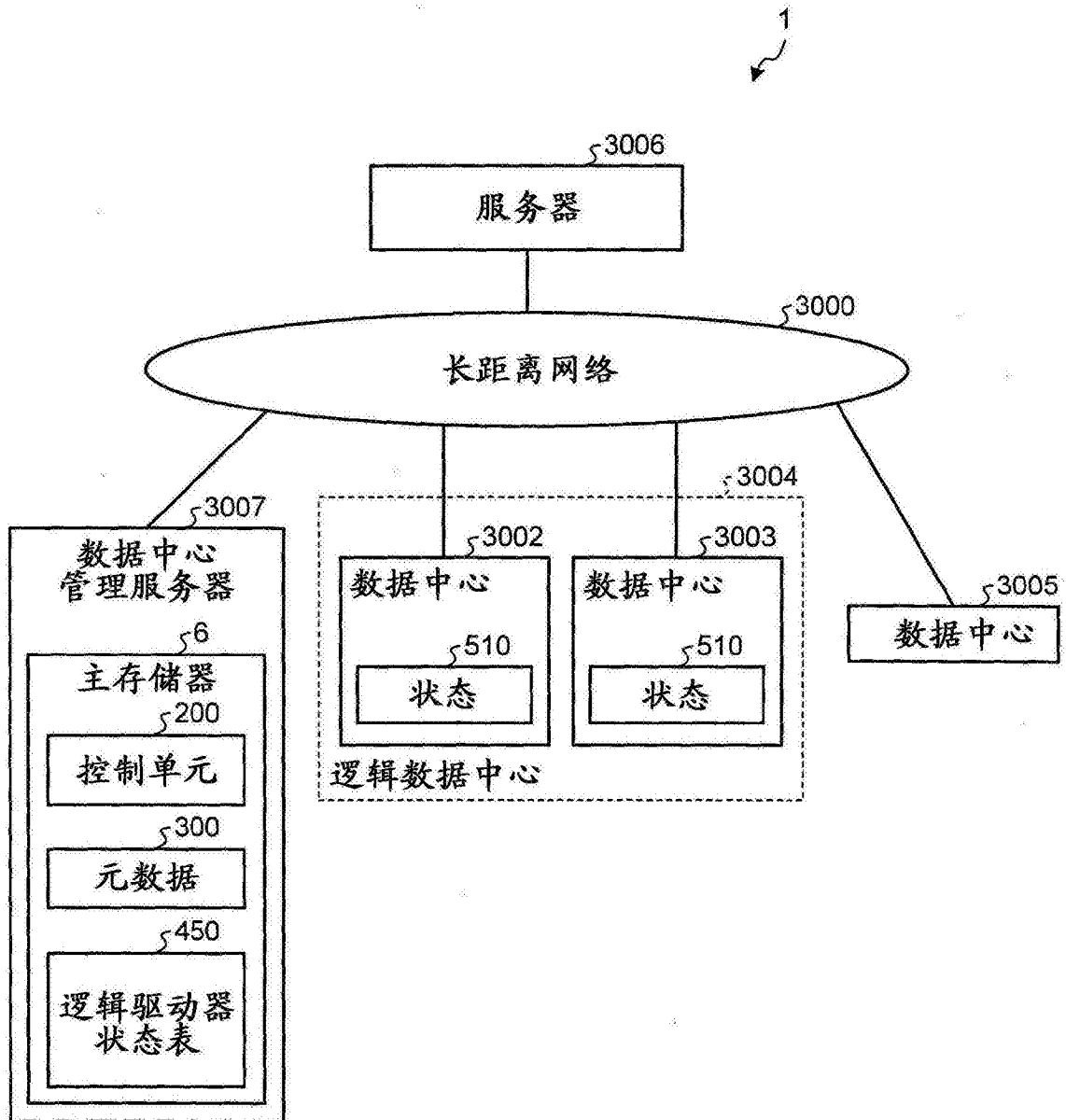


图76

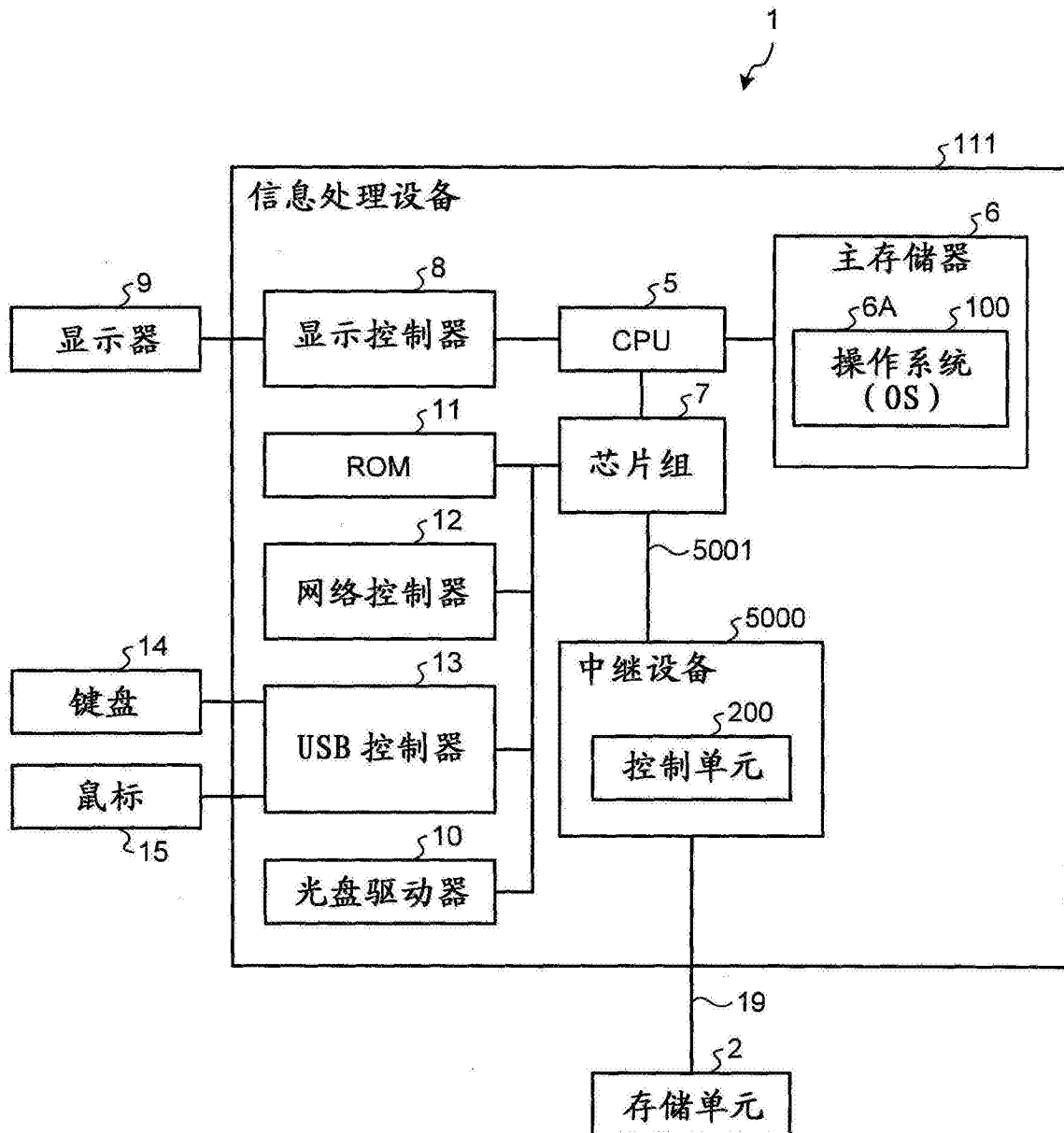


图77

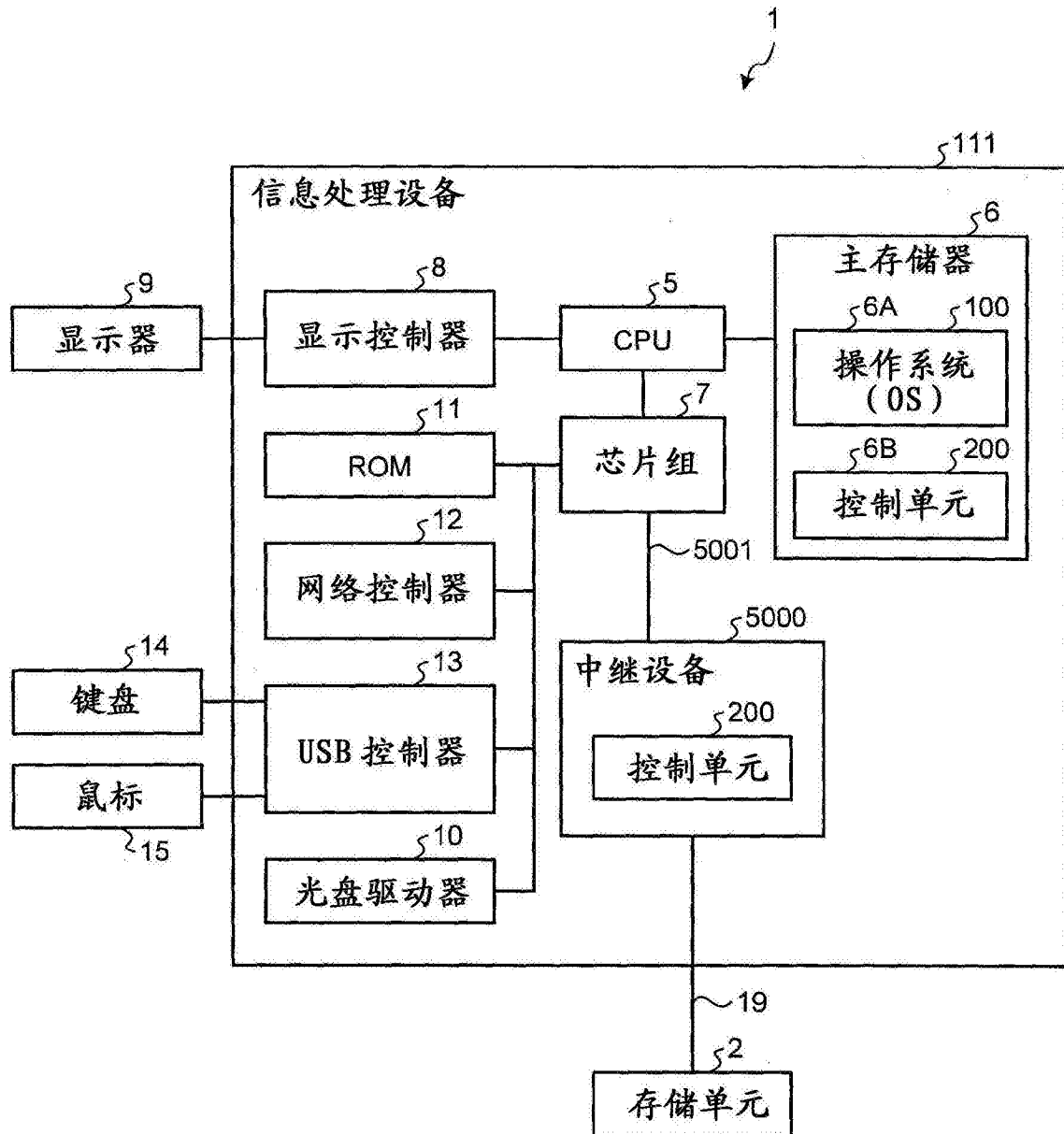


图78

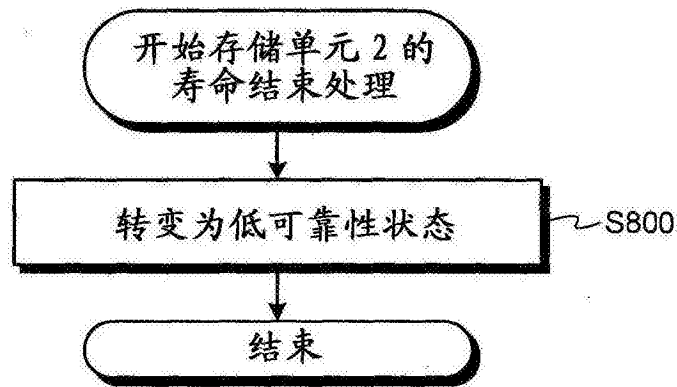


图79

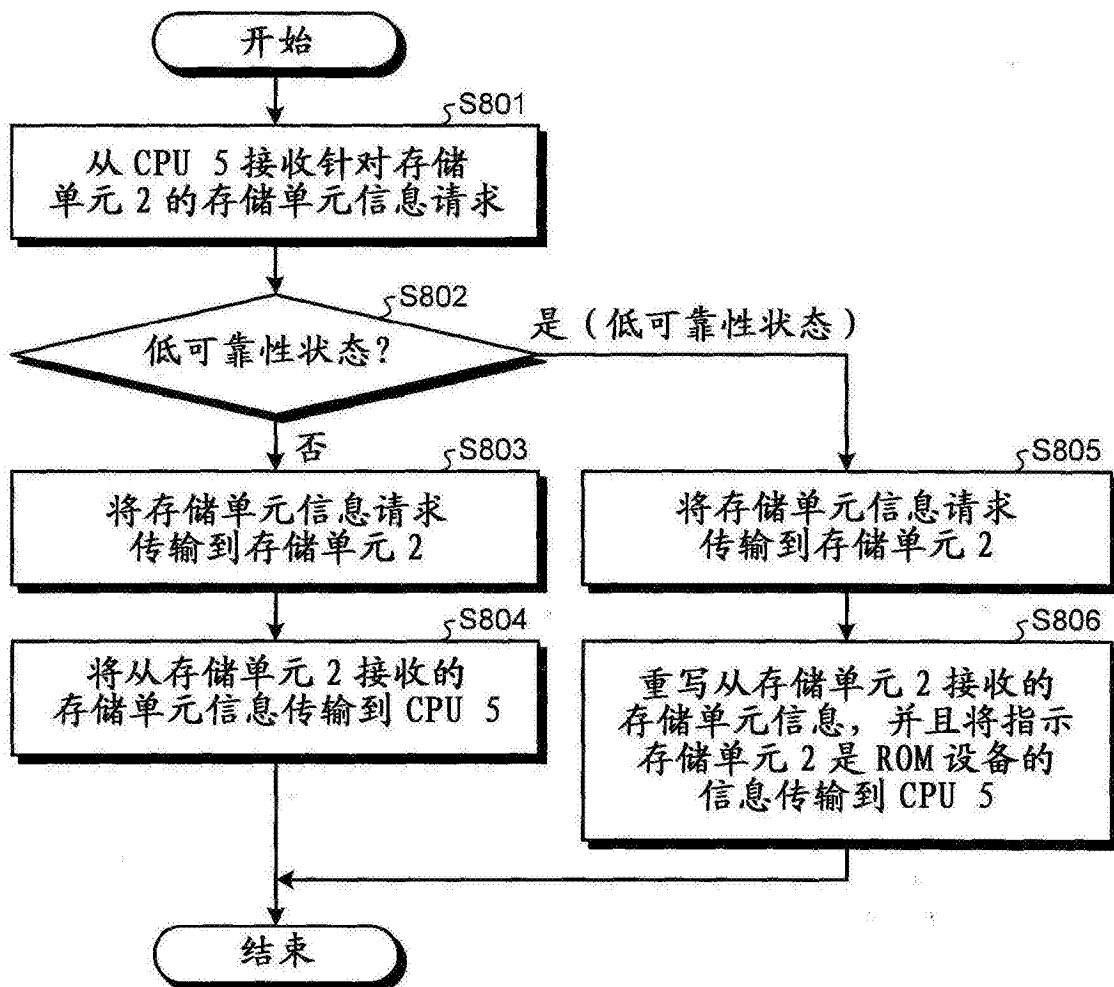


图80

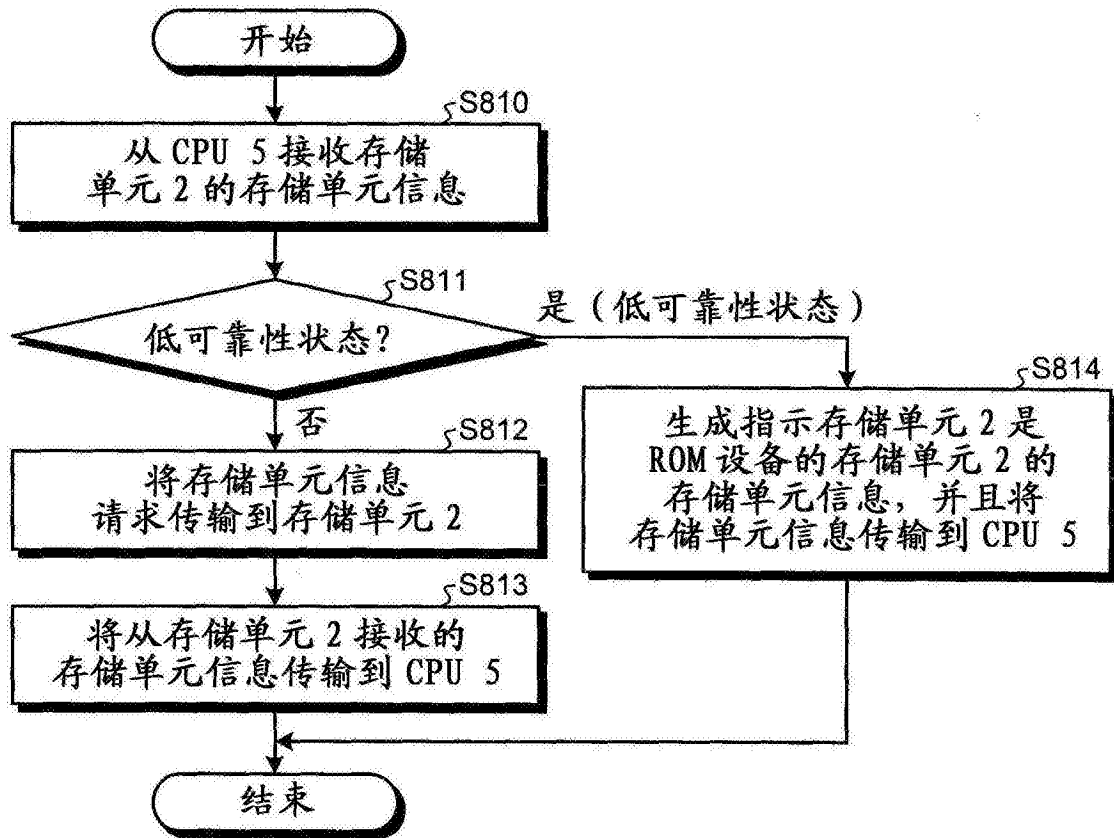


图81

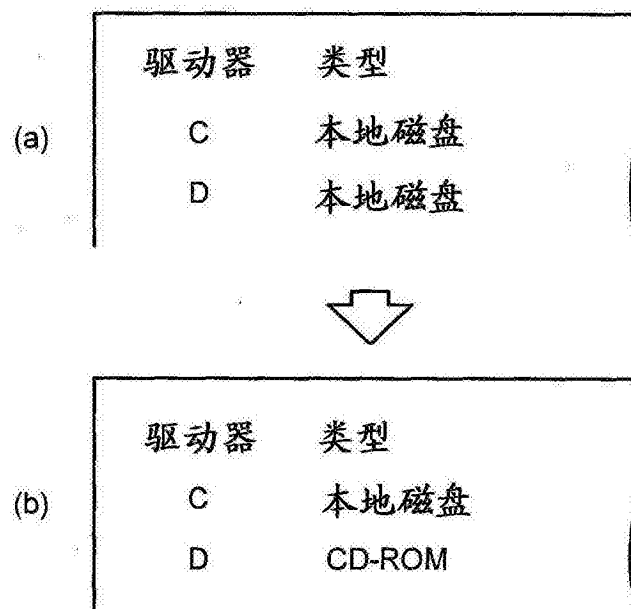


图82

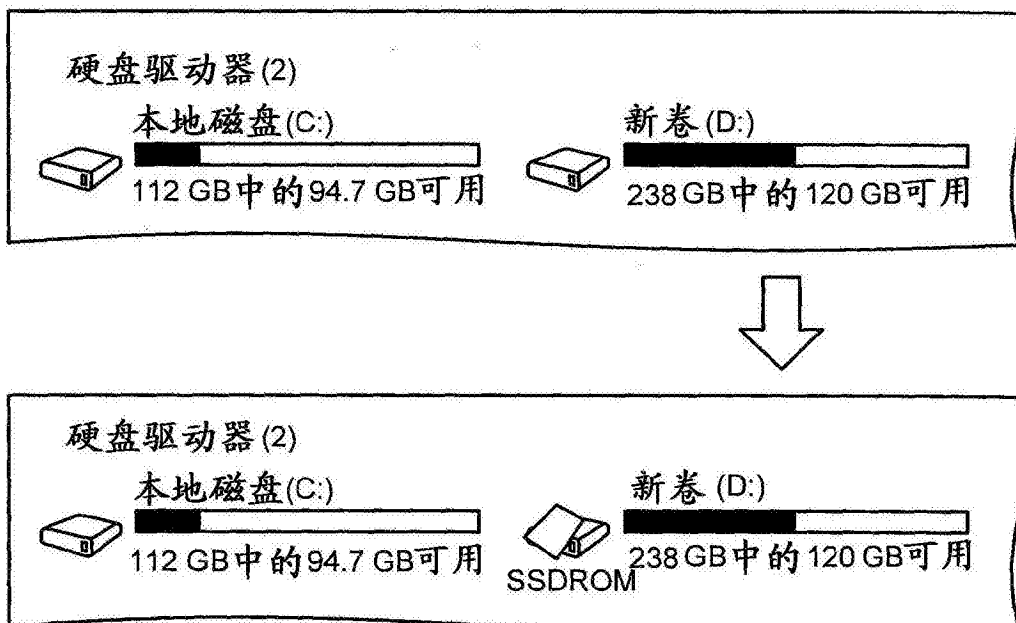


图83

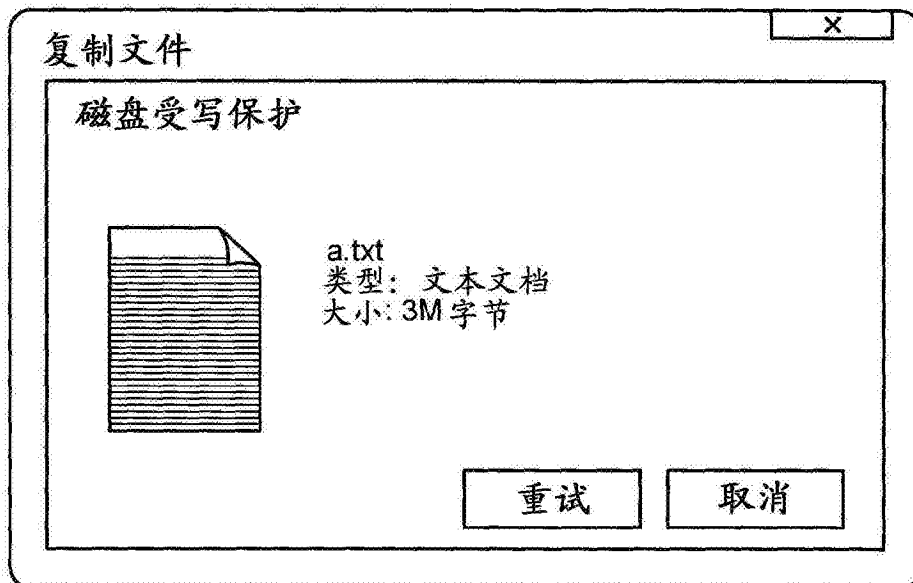


图84

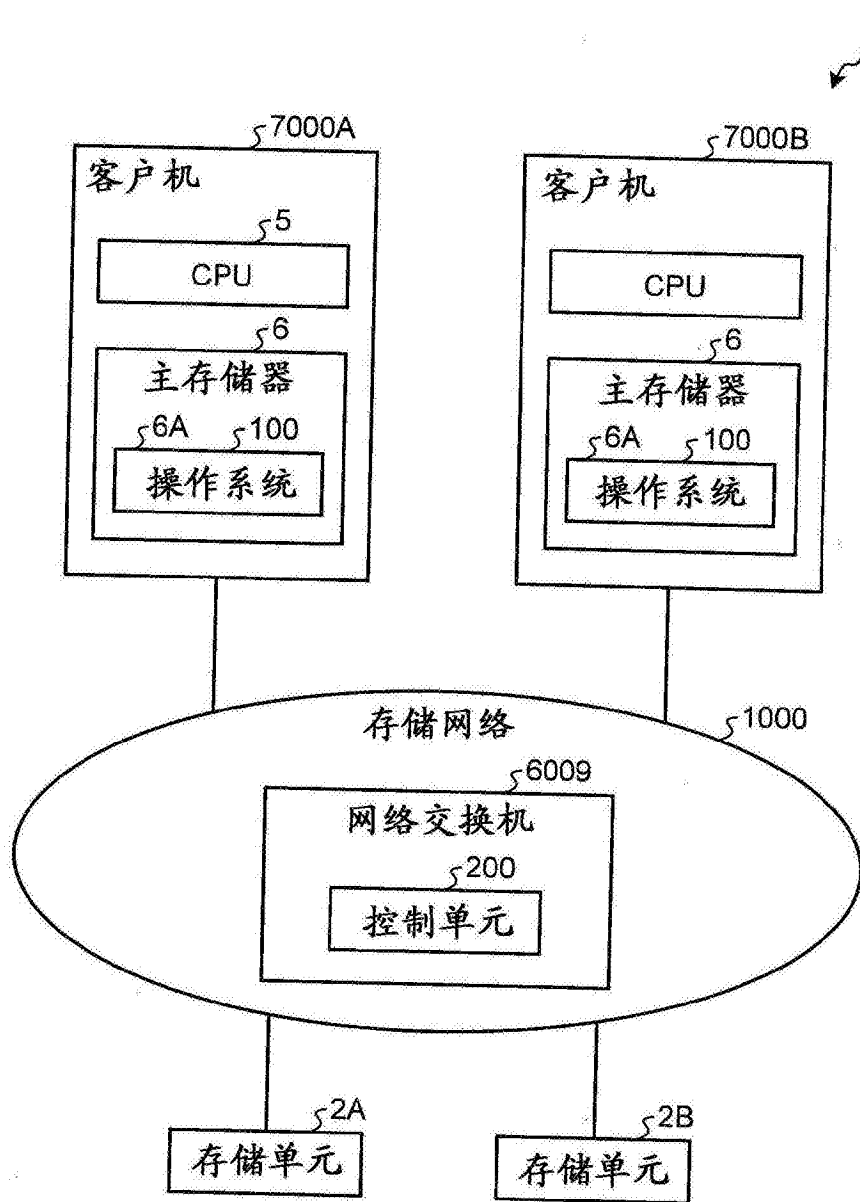


图85