

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 3/00 (2006.01)

G06F 11/00 (2006.01)



[12] 发明专利说明书

专利号 ZL 03107743.9

[45] 授权公告日 2006 年 2 月 8 日

[11] 授权公告号 CN 1241098C

[22] 申请日 2003.3.28 [21] 申请号 03107743.9
[30] 优先权
[32] 2002. 3.28 [33] JP [31] 92685/02
[71] 专利权人 株式会社日立制作所
地址 日本东京都
[72] 发明人 大桥和伸 佐藤孝夫 占部喜一郎
中野俊夫 横畑静生
审查员 王 丹

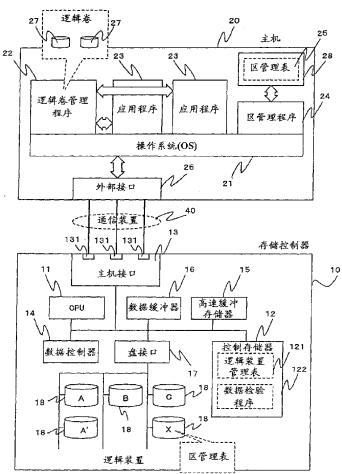
[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 杨 凯 张志醒

权利要求书 4 页 说明书 15 页 附图 8 页

[54] 发明名称
存储系统中检验数据的方法及设备

[57] 摘要

公开一种在存储系统中检验数据的方法。主机向存储控制器传送区管理数据。区管理数据指定要由具有检验数据适用性的机制的应用程序使用的存储装置中的存储区的范围。接收到主机发出的输入/输出请求时，存储控制器执行通常由应用程序执行的、对要根据数据输入/输出请求来处理以及要输入到根据接收的区管理数据所指定的存储区或从中输出的数据的检验。



1. 一种用于在存储系统中检验数据的方法，所述存储系统包括主机和存储控制器，所述存储控制器与所述主机连接以便与其进行通信，能够接收从所述主机发出的输入/输出请求，并且能够根据所述请求向存储装置执行数据输入/输出处理，所述方法包括以下步骤：

所述主机执行应用程序，以便在所述存储控制器中根据预定义的规范来组织和管理待存储的数据；

所述主机向所述存储控制器传送区管理数据，其中所述区管理数据用于指定所述存储装置中提供的、要由所述应用程序使用的存储区的范围；

所述存储控制器接收所述区管理数据；以及

所述存储控制器

接收所述应用程序执行的处理所导致的、从所述主机发出的数据输入/输出请求，以及

检验要根据所述接收的数据输入/输出请求来处理以及要输入按照所述区管理数据所指定的所述应用程序所用的所述存储区或从中输出的数据是否是按照所述预定义的规范来组织的，其中

参考从主机上接收过来的区管理数据表，并获取关于设在所接收的命令中的控制信息偏移量和大小及逻辑装置的容量等信息，然后再根据这些信息指定要检验的候选区，并将输入/输出命令所指定的地址与这个候选区的地址进行比较，看这个候选区的地址是否包括了输入/输出命令所指定的地址，如果不是，则结束检验，如果是，则启动一个与该应用程序所特定的算法来检验这些地址所对应的数据是否是按照预定义的规范来组织的。

2. 如权利要求1所述的用于在存储系统中检验数据的方法，其特征在于还包括以下步骤：

所述存储控制器存储多个算法，其中每个所述算法用于执行对运

行于所述主机上的多个应用程序中每一个的所述检验;

所述存储控制器根据从所述主机传送的所述区管理数据来指定每个所述应用程序使用的存储区;

所述存储控制器

- 5 接收所述应用程序之一执行的处理所导致的、从所述主机发出的数据输入/输出请求, 以及

根据所述区管理数据指定与所述接收的数据输入/输出请求有关的应用程序; 以及

- 所述存储控制器采用提供给所述指定应用程序的所述算法来检验
10 要根据所述接收的数据输入/输出请求来处理以及要输入按照所述区管理数据指定的所述存储区或从中输出的数据是否是根据所述预定义的规范来组织的。

3. 如权利要求 2 所述的用于在存储系统中检验数据的方法, 其特征在于还包括以下步骤:

- 15 所述主机将所述算法从所述主机发送给所述存储控制器; 以及
所述存储控制器接收和存储所述算法。

4. 如权利要求 1 所述的用于在存储系统中检验数据的方法, 其特征在于还包括以下步骤:

- 所述主机在所述存储区的所述范围发生变化时, 将反映所述变化的
20 的所述区管理数据传送到所述存储控制器。

5. 如权利要求 1 所述的用于在存储系统中检验数据的方法, 其特征在于, 所述应用程序所用的所述存储装置中的所述存储区是一个不包括由运行于所述主机上的所述应用程序之外的软件所添加的控制信息所用区域的区域。

- 25 6. 如权利要求 1 所述的用于在存储系统中检验数据的方法, 其特征在于, 所述区管理数据包括用于将提供给所述应用程序的所述存储区的范围指定为采用所述存储装置的所述存储区在逻辑上组织的卷的

数据。

7. 如权利要求 6 所述的用于在存储系统中检验数据的方法，其特征在于，通过根据独立磁盘冗余阵列方法管理所述存储装置的存储区来组织所述卷。

5 8. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，所述存储装置与所述存储控制器结合成整体来提供。

9. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于还包括以下步骤：

向所述主机发送表示所述检验结果的数据。

10 10. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，所述存储区作为由至少一个物理存储装置所提供的物理存储区中组织的逻辑存储装置来提供。

11. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，所述存储控制器连接到所述主机，以便通过符合 SCSI 标准的通信线路与其通信。
15

12. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，所述存储控制器连接到所述主机，以便通过存储区网络与其通信。

13. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，所述存储控制器连接到所述主机，以便通过 LAN 与其通信。
20

14. 如权利要求 1 所述的用于在存储系统中检验数据的方法，其特征在于，通过除用于连接所述存储控制器和所述主机的通信路径之外的通信路径来执行从所述主机到所述存储控制器的所述区管理数据的所述传送。
25

15. 一种存储控制器，它连接到主机以便与其通信，能够接收所述主机发出的输入/输出请求，并且能够根据所述请求来执行对存储装

置的数据输入/输出处理，所述存储控制器包括：

接收区管理数据的装置，其中所述区管理数据用于指定要由应用程序使用的所述存储装置中提供的存储区的范围，并且是从所述主机发送的，以及所述应用程序在所述主机中提供，用于在所述存储控制

5 器中根据预定义的规范来组织和管理要存储的数据；

接收由所述应用程序执行的处理所导致的、从所述主机发出的数据输入/输出请求的装置，以及

检验要根据所述接收的数据输入/输出请求来处理以及要输入所述应用程序使用的所述存储区或从中输出的数据是否是根据所述预定

10 义的规范来组织的装置。

存储系统中检验数据的方法及设备

5 相关申请的交叉引用

本申请要求基于 2002 年 3 月 28 日提交的日本特许公报 No.2002-092685 的优先权，现将其通过引用结合于此。

技术领域

10 本发明涉及用于在存储系统中检验数据的方法、存储系统中的主机以及存储控制器。

背景技术

一种存储系统引起极大的关注，在该系统中，主机和存储控制器、
15 如盘阵列装置通过 SAN(存储区网络)或其它网络互连，用于密集的存储操作或其它类似目的。如上述配置的存储系统目前正在 IDC(因特网数据中心)及许多其它地方运行。最近，焦点集中于使用称作 NAS(网络附属存储器)的存储控制器的系统，它配备了文件系统并且采用 LAN(局域网)作为网络。

20 经常要求运行于如上述配置和工作的存储系统中的主机上的数据库软件或其它应用程序所用的数据符合某些规范(这种特性以下称作“适用性”)。在这些情况下，某些应用程序配备了用于检验待处理数据的适用性的算法。

即使应用程序配备了上述算法，但是，数据适用性可能随着将主机
25 连接到存储控制器的 SAN 或其它通信路径的情况或者因存储控制器所执行的内部处理而丢失。此外，存储控制器通常由多个主机共享。在这种情况下，如果例如运行于一个主机上的应用程序无意中访问了运行于另一主机上的应用程序的数据，则可能损害数据适用性。

例如，如果在未涉及应用程序的情况下丢失了待写入存储控制器的数据的适用性，则存储控制器无法识别数据适用性的这种丢失，并自动存储不适用的数据。在这种情况下，该应用程序在后来读取该数据之前不会知道数据适用性的丢失。如果在写入数据的时刻和读取数据5 的时刻之间存在长的时滞，则难以实现数据恢复。

即使在存储控制器备份了数据的情况下，如果备份数据已丢失适用性，则不能实现数据恢复。此外，如果在未涉及应用程序的情况下丢失了数据适用性，如上所述，通常难以指明问题的原因，使随后的故障恢复及其它类似操作难以完成。

10 如果存储控制器被配置成检验数据，则能够解决上述问题。在存储系统中，通过应用程序存储在存储控制器中的数据一般配备了附加信息、如由运行在主机上的操作系统(以下称作“OS”)添加的信息以及有关由主机逻辑组织的卷的信息。此外，例如由于 OS、卷管理程序等原因，由应用程序存储在存储控制器中的数据的安排方式不是将主机上的数据安排为由应用程序调用。然而，存储控制器无法获得上述15 信息和安排。也就是说，只要使用传统的存储系统方案，应用程序所执行的数据检验处理就无法由存储控制器执行。如果要允许存储控制器完成这种数据检验，需要建立一种方案，使存储控制器能够从应用程序、OS 以及卷管理程序中获取相关信息。

20 例如，作为简化增强数据可靠性、即 RAS(可靠性、有效性和可服务性)的检验程序并且在应用程序从磁盘读取数据或向其中写入数据时执行的一种方案，日本公开特许公报 No.8-263223 公开了一种机制，其中，CPU 在应用程序执行 I/O 操作时生成 RAS 数据，并将 RAS 数据附加到待写入磁盘的数据中，以便使磁盘输入/输出信道及磁盘控制器能够进行数据检验。25

但是，在上述特许公报中公开的方案中，RAS 数据由 CPU 独立生成并附加到待写入的数据中。此外，RAS 数据未涉及应用程序所要求的数据适用性。所公开的方案未被配置成识别这种信息或配置，也未

被配置成检验运行在主机上的各应用程序所需的适用性。

发明内容

5 本发明是为了解决上述及其它问题，本发明的一个目的是提供数据检验方法，以提高的准确性来检验和确保存储系统中的数据适用性，更具体地说，本发明的目的是提供一种在存储系统中进行应用程序所要求的数据检验的方法。

本发明另一目的是提供实现上述功能的主机和存储控制器。

10 根据实现上述及其它目的的本发明的一个方面，提供一种在存储系统中检验数据的方法，所述存储系统包括主机和存储控制器，存储控制器与主机连接以便与其进行通信，能够接收从主机发送的输入/输出请求，并且能够根据请求执行对存储装置的数据输入/输出处理，该方法包括以下步骤：主机执行应用程序，以便在存储控制器中根据预定义的规范来组织和管理要存储的数据；主机向存储控制器发送区管理数据，其中，区管理数据用于指定存储装置中提供的要由应用程序使用的存储区的范围；存储控制器接收区管理数据；以及存储控制器接收应用程序执行的处理所导致的、从主机发送的数据输入/输出请求，并检验要根据接收的数据输入/输出请求处理以及要向根据区管理数据指定的应用程序所用的存储区输入或从中输出的数据是否是按照
15 20 预定义的规范组织的。

根据本发明的这个方面，能够提供数据检验方法、主机以及存储控制器，用于以提高的准确性来检验和确保存储系统中的数据适用性。

25 通过阅读以下结合附图对本发明进行的说明，本发明上述之外的其它特征和优点会非常明显。

附图说明

为了更透彻地理解本发明及其优点，参考以下结合附图来进行的

说明，图中：

图 1 是说明根据本发明一个实施例的存储系统的配置的示意图；

图 2 是说明根据本发明一个实施例的区管理表的示意图；

图 3 是流程图，说明在主机向存储控制器 10 发送区管理表时本发
5 明的一个实施例中执行的处理步骤。

图 4 是说明根据本发明一个实施例的逻辑装置管理表的示意图；

图 5A 是示意图，说明当应用程序对逻辑卷执行数据输入/输出操
作时本发明一个实施例中使用的数据单元；

图 5B 是示意图，说明在向存储控制器执行输入/输出操作时，逻
辑卷管理程序如何划分图 5A 所示的数据单元；
10

图 5C 是示意图，说明划分的数据如何存储在逻辑装置中；

图 6A 是示意图，说明图 5A 所示的 SCSI 数据格式的数据单元；

图 6B 是示意图，说明图 5B 所示的 SCSI 数据格式的数据；

图 7 是流程图，说明根据本发明的一个实施例、由存储控制器所
15 执行的数据检验处理步骤；以及

图 8 是说明主机的示例配置的框图。

具体实施方式

通过本说明书中的阐述以及附图的描述，至少会理解以下事项。

20 概述的要点

根据本发明一个实施例的检验数据的方法是用于在存储系统中检
验数据的方法，该存储系统包括主机和存储控制器，存储控制器与主
机连接以便与其进行通信，能够接收主机发送的输入/输出请求，并且
能够根据该请求执行对存储装置的数据输入/输出处理，该方法包括以
25 下步骤：主机执行应用程序，用于在存储控制器中根据预定义的规范
来组织和管理要存储的数据；主机将区管理数据发送到存储控制器，
其中，区管理数据用于指定由应用程序使用的存储装置中提供的存储
区的范围；存储控制器接收区管理数据；以及存储控制器接收由应用

程序所执行的处理导致的、从主机发出的数据输入/输出请求，并且检验要根据所接收的数据输入/输出请求处理以及要向根据区管理数据指定的应用程序所用的存储区输入或从中输出的数据是否是按照预定义的规范组织的。

- 5 例如，“从主机发出的输入/输出请求”是对诸如包含在存储控制器中或者外部连接到存储控制器的盘驱动器之类的上述存储装置的数据写请求或数据读请求。“区管理数据”是在区管理表中登记的数据，稍后说明。根据本发明的数据检验方法还在存储控制器中检验是否按照预定义的规范来组织应用程序所处理的数据。这确保了数据适用性
- 10 能够在提高的准确性下得到检验和保证。

 上述数据检验方法可包括以下步骤：存储控制器存储多个算法，其中各算法用于对运行于主机上的多个应用程序中的每一个执行检验；存储控制器根据从主机发送的区管理数据指定各应用程序所用的存储区；存储控制器接收应用程序之一执行的处理所导致的、从主机

15 发出的数据输入/输出请求，并且根据区管理数据指定与所接收的数据输入/输出请求相关的应用程序；以及存储控制器采用提供给指定应用程序的算法来检验要根据所接收的数据输入/输出请求来处理以及要输入到根据区管理数据指定的存储区或从中输出的数据是否是按照预定义的规范组织的。

- 20 因此，即使在多个不同应用程序运行于主机上时，也能够存储在存储控制器这一侧检验不同应用程序所处理的数据。

 上述数据检验方法可包括以下步骤：主机将算法从主机传送到存储控制器；以及存储控制器接收并存储算法。

- 上述数据检验方法可包括以下步骤：在存储区的范围发生变化
- 25 时，主机向存储控制器传送反映这种变化的区管理数据。因此，还能够确保存储控制器获得与应用程序所用的存储装置的存储区有关的最新信息。

 应用程序所用的存储装置中的存储区可以是除由运行于主机上的

应用程序之外的软件所添加的控制信息所用区域以外的区域。

区管理数据可包括用于将提供给应用程序的存储区的范围指定为采用存储装置的存储区在逻辑上组织的卷的数据。

可通过根据 RAID 方法管理存储装置的存储区来组织卷。

- 5 存储装置可与存储控制器结合成整体来提供。此外，表示检验结果的数据可传送到主机。例如，存储区可作为由至少一个物理存储装置所提供的物理存储区中组织的逻辑存储装置来提供。例如，预定义的规范表示例如指定的数据插入该数据的指定位置。存储控制器可与主机连接，以便通过符合 SCSI 标准的通信线路经 SAN 或 LAN 与其通信。
- 10

此外，可使用另一种配置，使得从主机到存储控制器的区管理数据的传送通过与连接存储控制器和主机的通信通路不同的通信通路来执行。这种替代配置的使用可实现例如通信负荷分配。

实施例

- 15 图 1 说明根据本发明一个实施例的存储系统的配置。

例如，图中所示的存储控制器 10 是盘阵列装置，主机 20 是使用存储控制器 10 作为存储资源的大型计算机或个人计算机。存储控制器 10 经通信装置 40 连接到主机 20。例如，通信装置 40 是符合 SCSI 接口标准的通信线路、LAN(局域网)或 SAN(存储区网络)。

- 20 存储控制器 10 可包括以下装置：CPU 11，用于控制存储控制器 10 中的各种部件和功能，并且执行和控制各种处理程序；控制存储器 12，用于存储各种信息；物理装置(未标出)、如盘驱动器，用作“存储装置”；主机接口 13，可用作“接收区管理数据的装置”和/或“接收数据输入/输出请求的装置”，用于连接到主机 20；数据控制器 14，
- 25 用于根据从主机 20 接收的数据输入/输出请求来控制对物理装置的数据输入/输出；高速缓冲存储器 15；数据缓冲器 16，用于临时存储要登记在高速缓冲存储器 15 中的数据；以及盘接口 17，用于根据来自数据控制器 14 的指令控制物理装置。

主机接口 13 配备一个或多个连接端口 131。例如，主机 20 的外部接口 26 是主机总线适配器，用作信道接口，并且可用作“发送区管理数据的装置”。

5 在存储控制器 10 中，在物理装置所提供的物理存储区中组织一个或多个逻辑装置 18。对每个组织的逻辑装置分配唯一的逻辑装置 ID，它可在主机 20 指定存储控制器 10 的存储区时使用。在本实施例中，假定在存储控制器 10 中组织了五个逻辑装置 18，由逻辑装置 ID 指定的各逻辑装置为 A、A'、B、C 以及 X。

OS 21 在主机 20 上运行。逻辑卷管理程序 22、应用程序 23 以及
10 区管理程序 24 运行于 OS 21 上。

图 8 是说明主机 20 的配置的框图。主机 20 包括至少一个 CPU(中央处理器)201、存储器 28、存储装置 202 以及上述外部接口 26。

CPU 201 负责对主机 20 的整个控制，并且通过执行存储在存储器
28 中的各种程序来实现根据本例的各种功能。存储装置 202 可存储各
15 种程序和数据。例如，硬盘驱动器等可用于存储装置 202。在本例中，存储装置 202 存储 OS 21、逻辑卷管理程序 22、应用程序 23 以及区管理程序 24，稍后进行详细说明。

逻辑卷管理程序 22 在存储控制器 10 的逻辑装置 18 的逻辑存储区
中组织一个或多个逻辑卷 27，并向应用程序 23 提供由逻辑卷 27 指定
20 的存储区。逻辑卷管理程序 22 的一个实例可以是所谓的 SoftRAID 程序，它根据基于软件的 RAID(独立磁盘冗余阵列)方法来管理存储控制器 10 的逻辑装置所提供的存储区，并将最终组织的逻辑卷 27 提供给应用程序 23。以下说明假定 SoftRAID 程序用作逻辑卷管理程序 22。

例如，各应用程序 23 是数据库软件，它配备一种方案，该方案用
25 于根据预定义的规范、即用于检验数据适用性的机制来组织和管理要存储在存储控制器 10 中的数据。各应用程序 23 还配备一种算法，该算法用于检查要处理的数据是否是按照预定义的规范来组织的，即用于检验数据是否适用。

区管理程序 24 管理存储在主机 20 的存储器 28 中的区管理表 25。下面说明关于区管理表 25 的详细情况。区管理程序 24 能够在适当时间将区管理表 25 发送给存储控制器 10。稍后进行详细说明。

区管理表

5 图 2 说明区管理表 25 的一个实例。区管理表 25 是为逻辑卷管理程序 22 所组织的各个逻辑卷 27 生成的。例如，当逻辑卷管理程序 22 改变逻辑装置 18 和逻辑卷 27 之间的关系时，区管理程序 24 根据需要按照从逻辑卷管理程序 22、应用程序 23 等所获取的信息来更新区管理表 25。

10 图 2 中，逻辑卷 ID 210 是分配给各逻辑卷 27 的唯一标识符。逻辑卷 ID 表示区管理表 25 对应于哪个逻辑卷 27。OS 控制信息偏移量 211 和 OS 控制信息大小 212 指示逻辑装置 18 中的存储单元，用于存储 OS 21 所提供的 OS 控制信息，以便管理逻辑装置 18 和逻辑卷 27。OS 控制信息的一个实例可以是文件控制信息。

15 逻辑卷控制信息偏移量 213 和逻辑卷控制信息大小 214 指示逻辑装置 18 中的存储单元，用于存储逻辑卷管理程序 22 生成的逻辑卷控制信息，以便管理逻辑卷 27。逻辑卷控制信息是逻辑卷管理程序 22 用来管理逻辑装置 18 的信息。逻辑卷控制信息的一个实例可以是逻辑卷管理程序 22 根据 RAID 方法操作逻辑装置 18 时所需的条配置信息。

20

逻辑装置 ID 215 是组成逻辑卷 27 的逻辑装置 18 的 ID，以及应用 ID 216 是使用逻辑卷 27 的应用程序 23 的 ID。例如，用户可通过操作主机 20 的用户界面来登记应用 ID。

应用处理数据大小 217 是使用逻辑卷 27 的应用程序 23 向逻辑卷 27 执行数据输入/输出操作时所处理的数据大小。条大小 218 是逻辑卷管理程序 22 通过根据 RAID 方法对数据分条而在逻辑装置 18 上写数据时所采用的数据大小。完整数据大小 219 是逻辑卷 27 的总存储容量。

25

区管理表传送

区管理表 25 在适当时间由区管理程序 24 从主机 20 传送到存储控制器 10。这种传送的“适当时间”可以是例如在主机 20 启动时或者在区管理表 25 被更新时，或者在已经设置为区管理程序 24 的调度功能时。

从主机 20 接收到区管理表 25 时，存储控制器 10 将该表存储在逻辑装置 18 中。在本实施例中，虽然假定区管理表 25 存储在具有逻辑装置 ID “X” 的逻辑装置 18 中，但区管理表 25 也可存储在控制存储器 12 等中。

图 3 是流程图，说明紧接主机向存储控制器 10 传送区管理表 25 之后的处理步骤。这些处理步骤可大致分为两种处理：区管理程序 24 获取表示最新位置的区管理表 25 的处理；以及程序传送所获取的区管理表 25 的处理；

首先，区管理程序 24 对用于处理过程中的标志及其它项的内容进行初始化(S310)，请求 OS 21 递交最新的 OS 控制信息偏移量 211 和 OS 控制信息大小 212，并获取这类信息(S311)。随后，区管理程序 24 检查主机 20 的存储器 28 中是否存在区管理表 25(S312)。

如果在存储器 28 中没有找到区管理表 25，区管理程序 24 在存储器 28 中生成区管理表 25(S313)。另一方面，如果在存储器 28 中找到了区管理表 25，区管理程序 24 则将区管理表 25 中的 OS 控制信息偏移量 211 及 OS 控制信息大小 212 与所获取的 OS 控制信息偏移量 211 及 OS 控制信息大小 212 进行比较(S314)。如果所获取的信息的内容不同于区管理表 25 中的内容，区管理程序 24 则更新表 25 以登记所获取的信息(S315)。在这种更新之后，区管理程序 24 启用更新标志(S316)。

随后，区管理程序 24 还从 OS 21 获取逻辑卷控制信息偏移量 213 及逻辑卷控制信息大小 214，并执行类似于对 OS 控制信息所进行的比较处理。也就是说，所获取的信息与存储器中的区管理表 25 的内容进行比较(S317、318)，如果发现所获取信息的内容不同于区管理表 25

中的内容，区管理程序 24 则更新区管理表 25 以便反映所获取的信息(S319)，并在这种更新之后启用更新标志(S320)。

根据上述程序，在存储器 28 中生成表示最新状态的区管理表 25。

随后，区管理程序 24 检查更新标志(S321)。当更新标志启用时，
5 区管理程序 24 将存储在存储器 28 中的区管理表 25 的内容传送到存储
控制器 10(S322)。在接收到区管理表 25 时，存储控制器 10 将该表存
储在逻辑装置 X 中。

例如，如果组成逻辑卷 27 的逻辑装置 18 的数量发生变化，则在
例如 OS 控制信息或逻辑卷控制信息中出现变化的情况下或者在逻辑
10 卷 27 的配置已经改变的情况下启动 OS 21 时，根据需要执行将区管理
表 25 从主机 20 传送到存储控制器 10 的上述处理。这确保表示最新状
态的区管理表 25 始终设在存储控制器 10 中。

逻辑装置管理表

在存储控制器 10 的控制存储器 12 中存储逻辑装置管理表 121。
15 逻辑装置管理表的一个实例如图 4 所示。关于逻辑装置 ID 411，该表
管理例如以下各项：对于各逻辑装置 ID 411 为唯一的 LUN(逻辑单元
号)412；各逻辑装置 18 的存储容量 413；以及端口 ID 414，它是各逻
辑装置 18 所连接的主机接口 13 的端口 131 的 ID。逻辑装置管理表 121
的内容保持为最新。例如，它们由操作员通过连接存储控制器 10 的管
20 理终端(未标出)进行手动更新，或者根据例如存储在存储控制器 10 中
或从主机 20 传送的信息自动更新。

数据状态

下面说明从运行于主机 20 上的应用程序 23 输出的数据如何存储
在存储控制器 10 的逻辑装置 18 中，注意数据配置。

25 图 5A 说明当应用程序 23 向逻辑卷 27 执行数据输入/输出操作时
的数据单元。数据单元 51 的数据大小等于图 2 所示应用程序的处理数
据大小。

图 5B 说明在向存储控制器 10 执行输入/输出操作时、逻辑卷管理

程序 22 如何划分图 5A 所示的数据单元 51。在该图所示的实例中，数据单元 51 分为三个部分：数据 52、数据 53 以及数据 54。

图 5C 说明数据 52、数据 53 以及数据 54 如何存储在逻辑装置 18 中。该图说明一个实例，其中，具有应用程序 23 指定为“001h”的逻辑卷 ID 210 的逻辑卷 27 包括逻辑装置 ID 215 为“A”的逻辑装置 18 以及逻辑装置 ID 215 为“A”的逻辑装置 18。划分的数据 52 和 53 存储在逻辑装置 ID 215 为“A”的逻辑装置 18 中；划分的数据 54 存储在逻辑装置 ID 215 为“A”的逻辑装置 18 中。

这些逻辑装置 18 的固定区存储上述 OS 控制信息 55、56 以及逻辑卷控制信息 57、58。

图 6A 说明如图 5A 所示的数据单元，它的数据格式符合 SCSI 标准。图 6B 说明图 5B 所示的 SCSI 数据格式的数据。

如图 6B 所示，对于命令帧中的各操作码字段 611 至 613，设置了表示所请求处理的类型的命令。在该图的实例中，设置了写命令。对于各个 LUN 字段 621 至 623，设置了将成为处理目标的逻辑装置 ID 215 或 LUN(逻辑单元号)。在这些实例中，“0001h”设置在 LUN 字段 621、622 中，以及“0002h”设置在 LUN 字段 623 中。对于逻辑地址字段 631 至 633，设置了与数据 52 至 54 将写入其中的逻辑装置 18 中的存储起始单元(即开始存储的单元)相对应的地址。对于数据长度字段 641 至 643，设置了与相应命令帧中的写数据 52 至 54 相对应的数据长度。

以上说明涉及从应用程序 23 输出数据写请求的情况。当输出数据读请求时，读命令将设置到上述各操作码字段 611 至 613，指定逻辑装置 18 中的数据读起始单元(即开始读取数据的单元)的地址将设置到逻辑地址字段 631 至 633，以及待读取数据的大小将设置到各数据长度字段。

数据检验处理

下面说明在主机 20 向存储控制器 10 传送 SCSI 数据时、由存储控制 10 执行的数据检验处理。

通过执行存储控制器 10 存储在其控制存储器 12 中的数据检验程序 122(算法)来执行数据检验处理。因此,该程序可用作“检验装置”。数据检验程序 122 是为主机 20 上运行的各应用程序 23 准备的。数据检验程序 122 包括一种功能,用于在执行目标数据(即待处理数据)的处理之前、检验目标数据是否符合预定义的规范,该功能执行等效于或者
5 优于在主机 20 上运行的应用程序 23 所包含的上述算法的处理。

数据检验程序 122 例如通过从主机 20 的传送或操作员采用管理终端(未标出)对存储控制器 10 的手动操作而存储在控制存储器 12 中。对于待存储的各数据检验程序 122 分配相关应用程序 23 的 ID。

10 数据检验处理用于检查目标数据是否符合应用程序 23 所要求的预定义规范。为此,在已经存储于逻辑装置 18 的存储区中的整个数据中,只有存储从应用程序 23 输入或者输出到应用程序 23 的数据的区域必须受到数据检验处理。因此,在数据检验处理之前,执行排除这种区域的处理,稍后说明。

15 下面利用图 7 所示的流程图,结合从运行于主机 20 上的应用程序 23 传送图 6A 所示的写入数据 51 的数据写请求以及对应于数据写请求的图 6B 所示写命令帧被传送到存储控制器 10 的情况,对存储控制器 10 中执行的数据检验处理进行说明。例如通过存储在存储控制器 10 的控制存储器 12 中的微程序来执行这种处理。

20 从主机 20 接收到带有写命令的命令帧时(S711),存储控制器 10 将命令帧存储在数据缓冲器 16 中(S712)。然后,存储控制器 10 参考存储在逻辑装置 ID 为“X”的逻辑装置 18 中的区管理表 25,获取关于设在所接收数据写请求中的逻辑装置 ID 的 OS 控制信息偏移量 211、OS 控制信息大小 212、逻辑卷控制信息偏移量 213 以及逻辑卷控制信息大小 214(S713)。此外,存储控制器 10 参考逻辑装置管理表
25 121 以获取对应于上述逻辑装置 ID 的逻辑装置 18 的存储容量 413(S714)。

随后,存储控制器 10 存储用于在对应于上述逻辑装置 ID 的逻辑

装置 18 的整个存储区中指定除上述 OS 控制信息偏移量 211、OS 控制信息大小 212、逻辑卷控制信息偏移量 213 以及逻辑卷控制信息大小 214 指定区域之外的区域(S715)。以下将上述地址所指定的存储区称作用于检验的候选区。

5 随后,存储控制器 10 将这些地址所指定的用于检验的上述候选区与设置在所接收命令帧的逻辑地址字段 631 至 633 中的地址进行比较(S716)。如果设置在逻辑地址字段中的地址不在用于检验的候选区中,则结束数据检验处理(S717)。另一方面,如果该地址包含在用于检验的候选区中,则存储控制器 10 还检查由设置在逻辑地址字段中的地址以及设置在所接收命令帧的数据长度字段 641 至 643 中的数据长度所定义的、并被指定为写入写数据的写目的地址的整个存储区是否处于用于检验的候选区中(S718)。

10 如果所指定的整个存储区在用于检验的候选区中,存储控制器 10 则将已经指定为写目的地址的整个存储区设置为用于数据检验的目标区,然后存储用于定义这个区的地址(例如定义该区的起始和结束地址)(S719)。

15 另一方面,如果整个存储区不在用于检验的候选区中,则存储控制器 10 将从包含在命令帧中的上述逻辑地址开始、直到用于检验的候选区的结束地址为止的区域设置为用于数据检验的目标区,然后再存储用于定义这个区的地址(例如定义该区的起始和结束地址)(S720)。

20 按照上述设置用于数据检验的目标区之后,存储控制器 10 获得与待写入逻辑装置 18 相关的应用程序 ID。应当指出,存储控制器 10 存储表示提供给各应用程序 23 的应用 ID 和数据检验程序 122 之间关系的对应关系。因此,存储控制器 10 启动与所获得的应用程序 ID 相关的数据检验程序 122(S721)。

25 启动数据检验程序 122 时,存储控制器 10 向数据检验程序 122 提供指定用于数据检验的目标区的上述地址。然后,数据检验程序 122 开始执行与上述地址所指定的逻辑装置 18 的存储区相关的数据检验

处理。

如果在数据检验程序 122 所执行的数据检验处理中检测到差错 (S723), 则存储控制器 10 将表示该差错的消息(例如表示非法请求或写错误的消息)传送到主机 20(S724)。如果没有检测到任何差错, 存储控制器 10 将目前存储在数据缓冲器 16 中的命令帧中的写数据传送到高速缓冲存储器 15, 并将数据写入逻辑装置 18(S725)。

以上说明涉及存储控制器 10 从主机 20 接收数据写请求的情况。但是, 当接收数据读请求时, 不是检验命令帧中的数据, 而是从存储装置中读出待读取的数据, 存储在数据缓冲器 16 中, 并通过执行与写命令相同的处理来检验。也就是说, 即使接收到数据读请求, 也会执行数据检验处理, 并且如果检测到任何差错, 相关的出错消息会发送给主机 20。因此, 例如, 在读数据递交给应用程序 23 之前, 可通知应用程序 23 数据不适用(即未确定数据的“适用性”)。因此能够防止任何数据丢失或数据中的损坏, 以及防止应用程序 23 执行错误操作。

如上所述, 根据本发明, 存储控制器 10 还能够获取数据检验所需的信息。因此, 通常由应用程序 23 执行的数据检验处理也能够由存储控制器 10 来执行。由于存储控制器 10 还检验应用程序 23 所要求的数据适用性, 因此能够进一步提高数据管理准确性。

其它

应当指出, 在上述说明中, 逻辑卷管理程序 22 不一定是必要的构件。

此外, 区管理表 25 的整个内容不一定要从主机 20 传送到存储控制器 10; 它可这样配置, 以便仅传送表示因数据变化而产生的差别的差异数据。

当区管理程序 24 首次执行时, 没有任何数据写入主机 20 的区管理表 25 中。在这种情况下, 它可经过配置, 以便在假定例如某些变化已加至区管理表 25 的情况下, 将区管理表 25 传送到存储控制器 10。

此外, 通过将主机 20 通过 LAN 之类的网络与存储控制器 10 连接

以便经 LAN 传送区管理表 25，也可以高速执行从主机 20 到存储控制器 10 的区管理表传送。

虽然详细说明了本发明的最佳实施例，但应当理解，只要不背离所附权利要求所定义的本发明的精神和范围，可以进行各种变更、替换和更改。

5

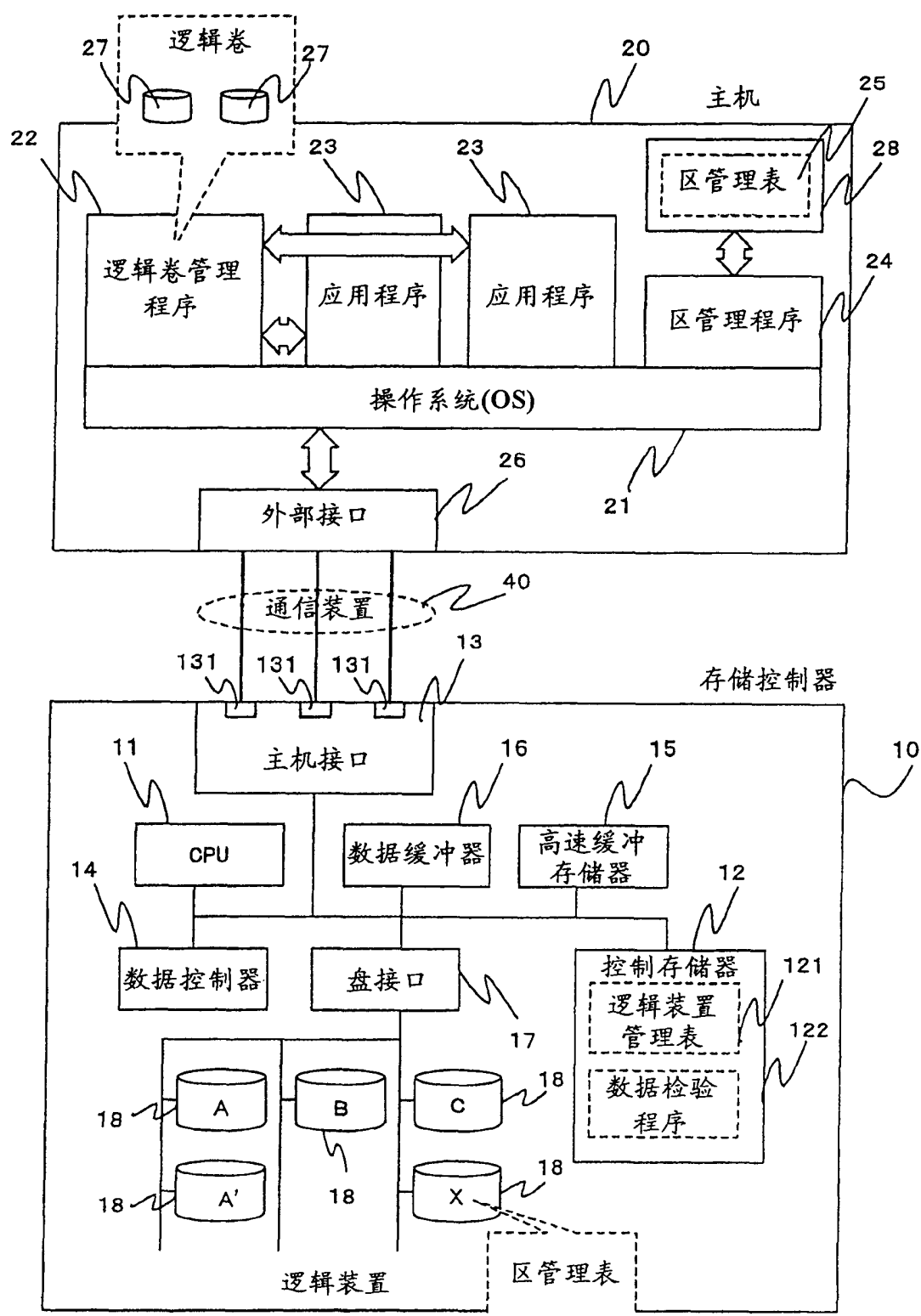


图 1

逻辑卷 ID	001h	210
OS 控制信息偏移量	0000h	211
OS 控制信息大小	1024 字节	212
逻辑卷控制信息偏移量	1024 字节	213
逻辑卷控制信息大小	512 字节	214
逻辑装置 ID	A, A'	215
应用 ID	α	216
应用处理数据大小	1024	217
条大小	4096	218
总数据大小	100GB	219

图 2

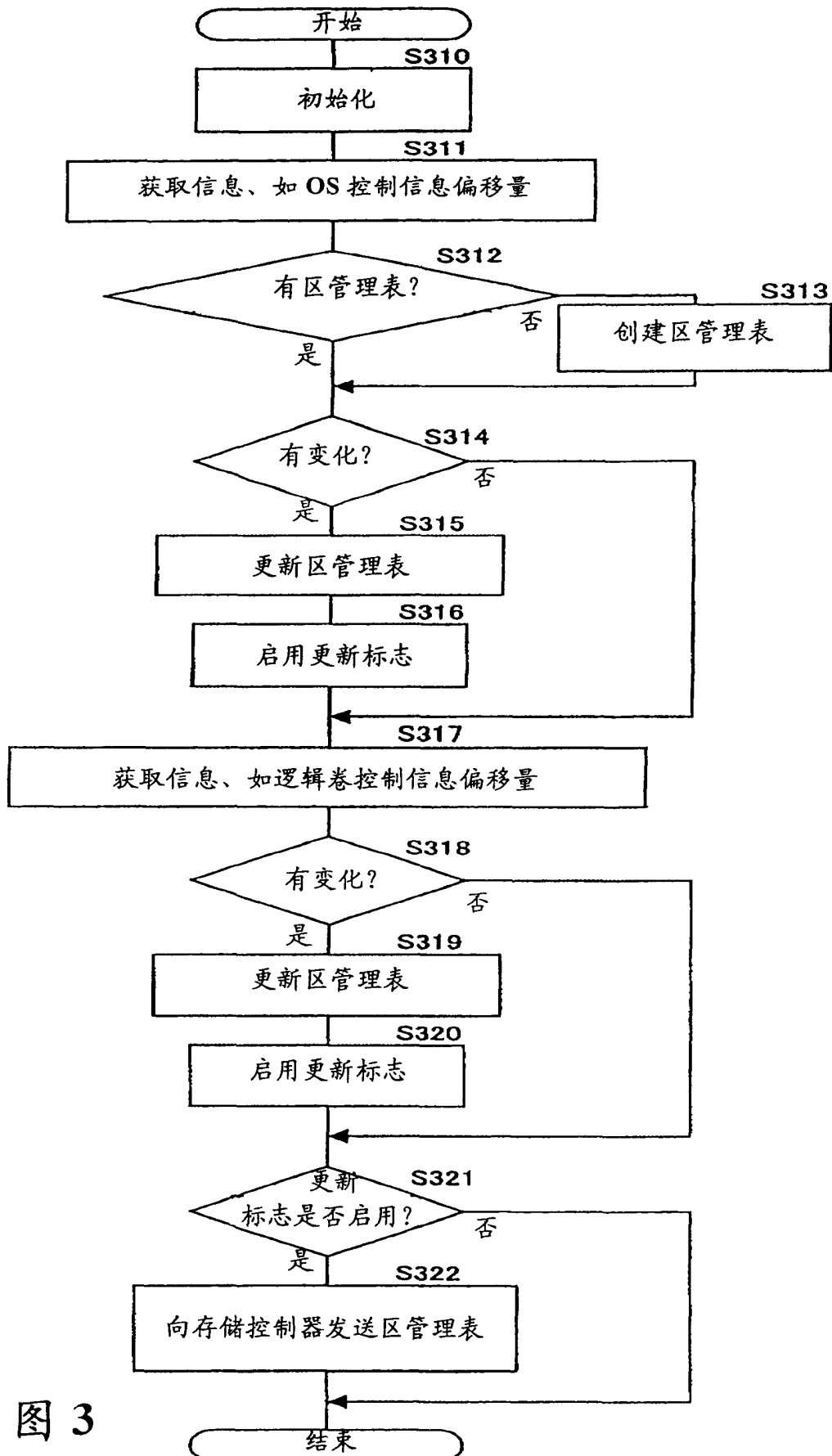
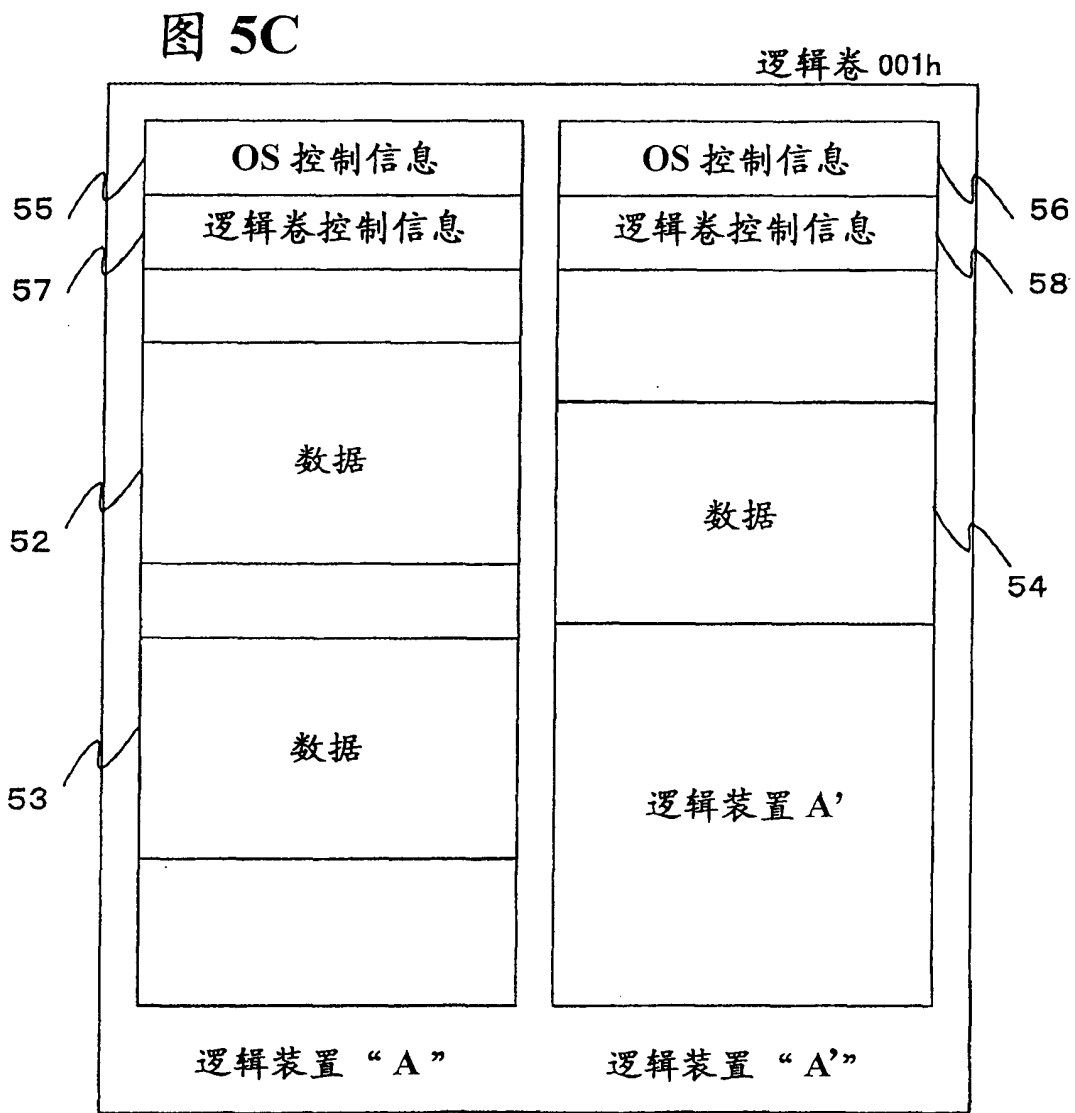
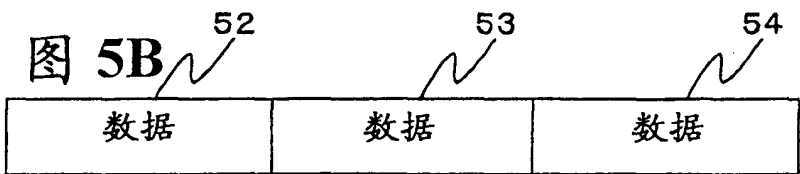
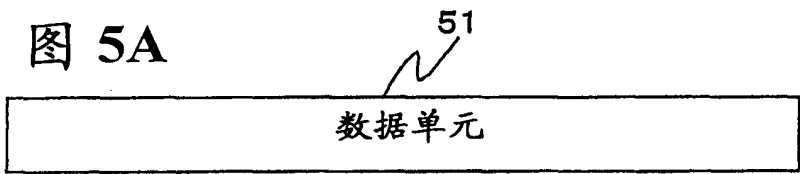


图 3

⁴¹¹ 逻辑装置 ID	⁴¹² LUN	⁴¹³ 存储容量	⁴¹⁴ 端口 ID	.
A	0001h	500Gb	01h	.
A'	0002h	300Gb	01h	.
B	0003h	1Tb	02h	.
C	0004h	1Tb	03h	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

图 4



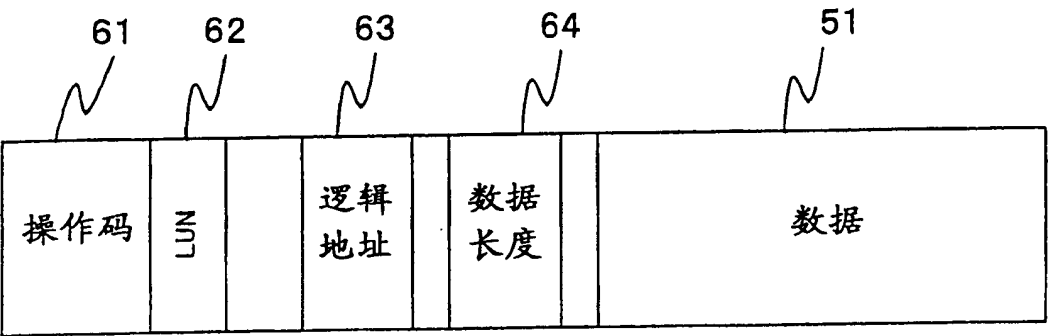


图 6A

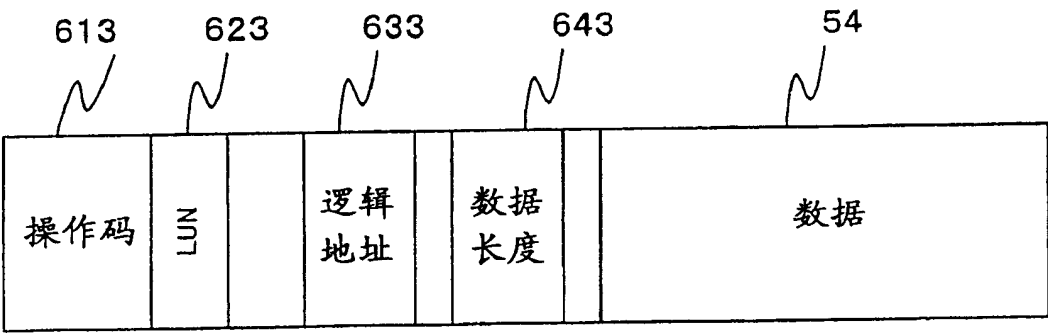
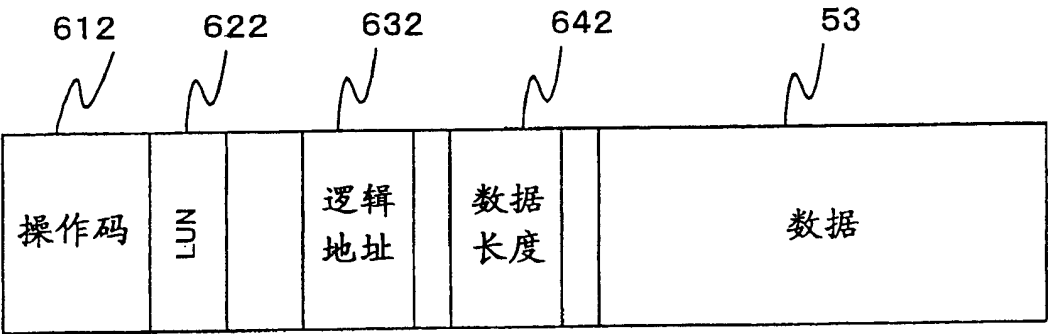
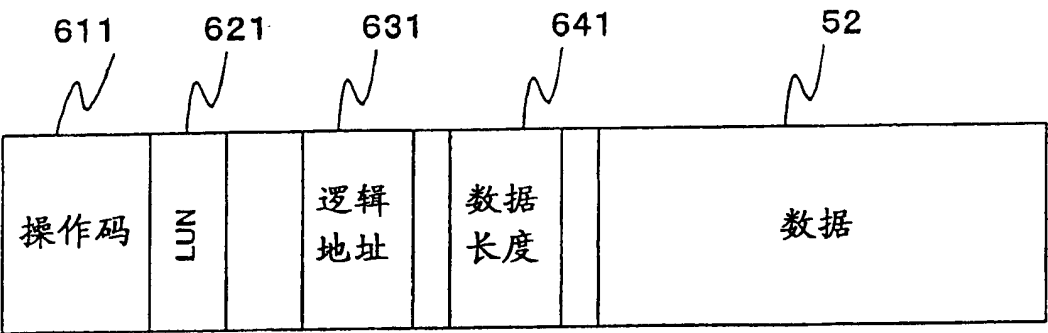


图 6B

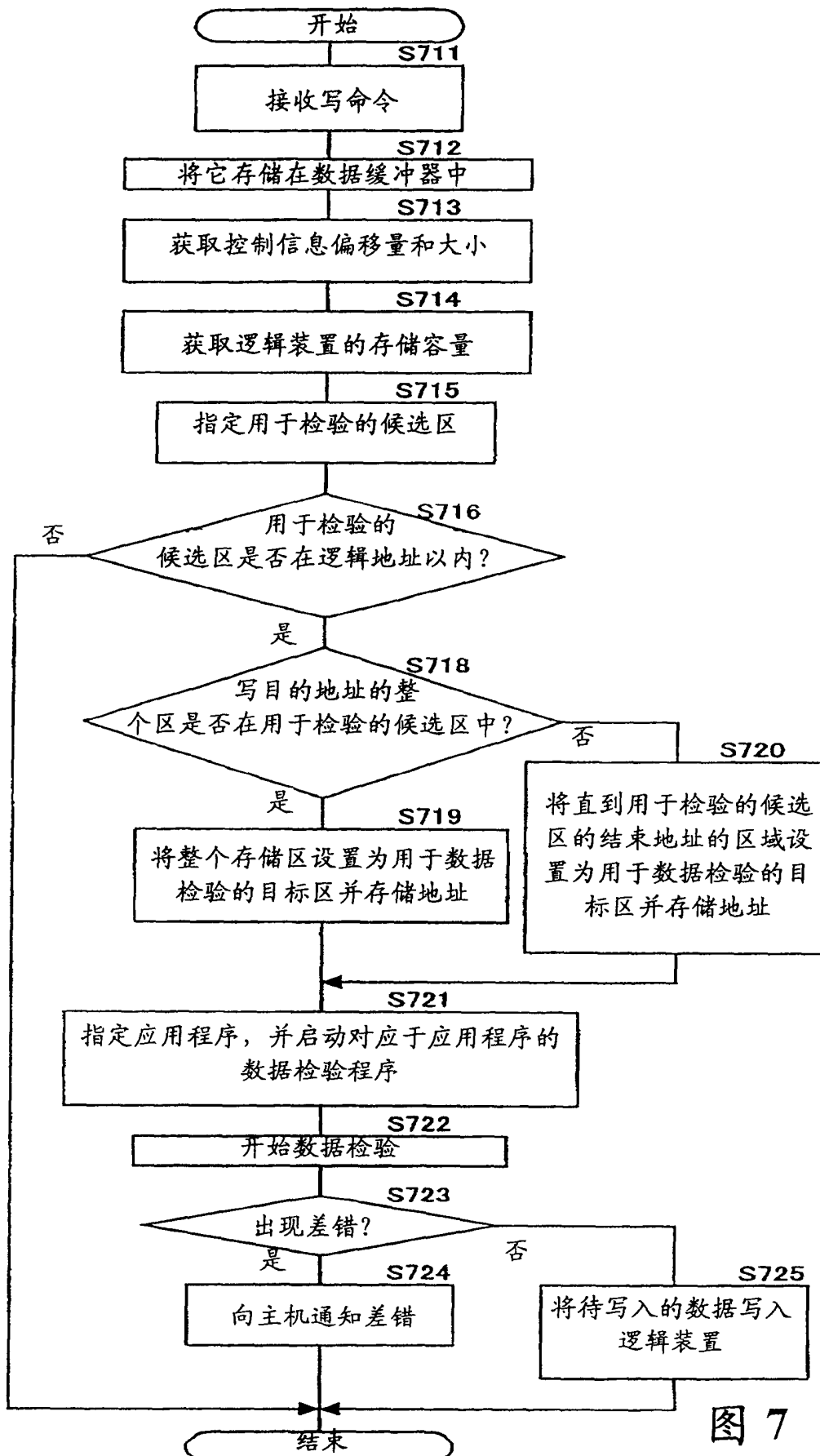


图 7

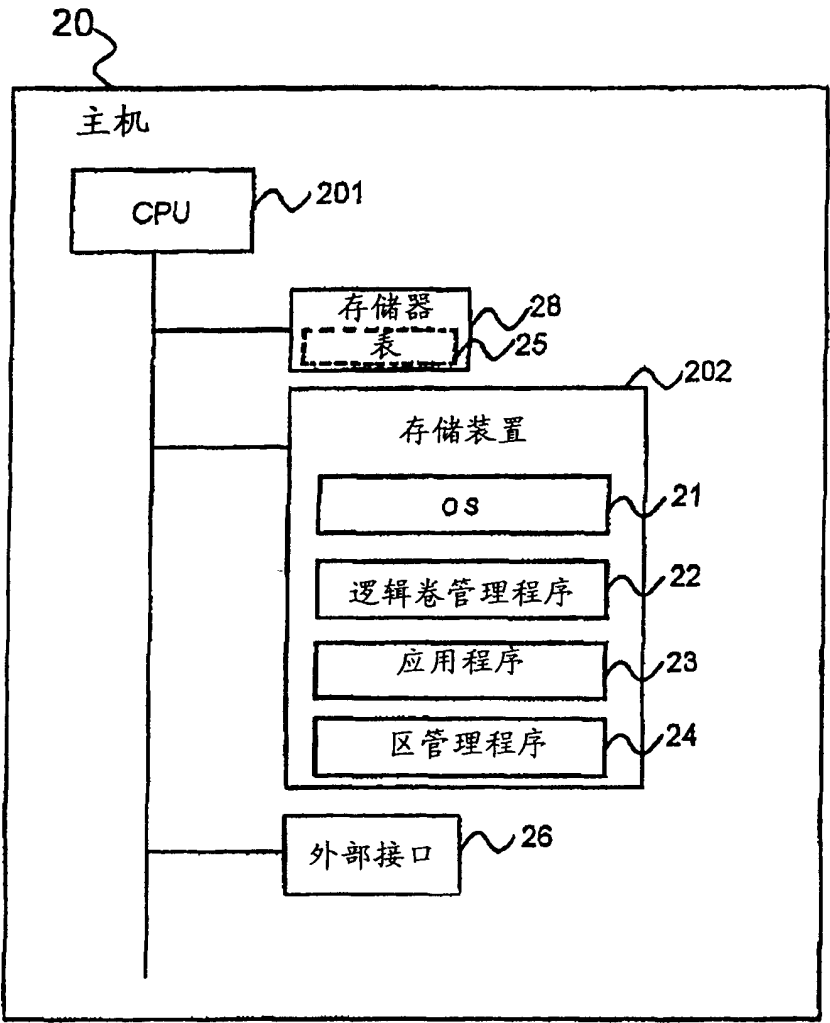


图 8