[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 12/16 (2006.01)
G06F 12/08 (2006.01)
G06F 1/30 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200710194621.8

[43] 公开日 2008 年 8 月 13 日

[11] 公开号 CN 101241477A

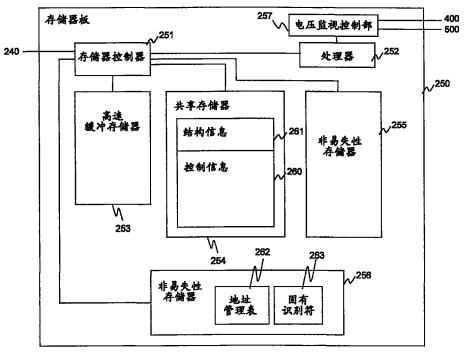
[22] 申请日 2007.11.27
[21] 申请号 200710194621.8
[30] 优先权
[32] 2007. 2. 7 [33] JP [31] 2007 - 027620
[71] 申请人 株式会社日立制作所
地址 日本东京都
[72] 发明人 饭田纯一 姜小明
[74] 专利代理机构 北京银龙知识产权代理有限公司
代理人 许 静
权利要求书 5 页 说明书 23 页 附图 15 页

[54] 发明名称

存储控制装置和数据管理方法

[57] 摘要

本发明提供存储控制装置和数据管理方法，可以实现抑制非易失性存储器的容量、并且可以使高速缓冲存储器中存储的数据妥当地转移的技术。I/O 处理器(220)判定高速缓冲存储器(253)上的脏数据量是否超过阈值，当判定为超过时，将高速缓冲存储器(253)的脏数据的一部分写入存储设备(310)，当通过电压监视控制部(257)检测出供给电力的电压异常时，电压监视控制部(257)使用来自电池(500)的电力维持电力供给，处理器(252)接受来自电池(500)的电力供给，使高速缓冲存储器(253)中存储的脏数据转移至非易失性存储器(255)。



1. 一种存储控制装置，从外部装置接收写入访问请求，进行将写入访问请求对象数据写入存储装置的控制，其中，

具有:

进行来自电源的电力供给的电力供给部;

以可以供给电力的方式积蓄电力的电池;

作为接受电力供给可以存储数据的易失性存储器的高速缓冲存储器;

即使不接受电力供给也可以继续存储数据的非易失性存储器;

接受所述电力供给，接收写入访问请求的请求接收部;

第一数据存储部，将所述写入访问请求的对象数据作为高速缓冲存储数据，存储在所述高速缓冲存储器中;

判定部，判定所述高速缓冲存储器的所述高速缓冲存储数据中未被反映到所述存储装置的脏数据的数据量是否超过预定的阈值;

第二数据存储部，当判定为超过所述阈值时，将所述高速缓冲存储器的所述脏数据中至少一部分向所述存储装置存储;

电源监视部，检测从所述电力供给部供给的电力的电压异常;

转移存储部，当通过所述电源监视部检测出所述电压异常时，接受来自所述电池的所述电力供给，使所述高速缓冲存储器中存储的脏数据向所述非易失性存储器转移；以及

电力供给控制部，当通过所述电源监视部检测出所述电压异常时，使用来自所述电池的电力，维持向所述高速缓冲存储器以及所述转移存储部的电力供给。

2. 根据权利要求 1 所述的存储控制装置，其中，

根据所述非易失性存储器的容量来决定所述阈值。

3. 根据权利要求 2 所述的存储控制装置，其中，

具有根据所述非易失性存储器的容量来决定所述阈值的阈值决定部。

4. 根据权利要求 1 至 3 中任意一项所述的存储控制装置，其中，

所述转移存储部将所述脏数据加密，并使其转移至所述非易失性存储器。

5. 根据权利要求 4 所述的存储控制装置，其中，

所述转移存储部执行不使所述脏数据的数据量发生变化的加密。

6. 根据权利要求 1 至 5 中任意一项所述的存储控制装置，其中，

所述转移存储部压缩所述脏数据，并使其转移至所述非易使性存储器。

7. 根据权利要求 1 至 6 中任意一项所述的存储控制装置，其中，

还具备存储目的地信息存储部，使所述脏数据的所述高速缓冲存储器中的存储目的地信息转移至所述非易失性存储器。

8. 根据权利要求 1 至 7 中任意一项所述的存储控制装置，其中，

还具备数据恢复部，使转移至所述非易失性存储器的脏数据在所述高速缓冲存储器中恢复。

9. 根据权利要求 8 所述的存储控制装置，其中，

可装卸保持所述非易失性存储器的存储器板，

具有：

安装板信息保存部，保存已安装的所述存储器板的第一识别信息；

安装检测部，对新安装了保持非易失性存储器的存储器板进行检测；

板信息取得部，取得所述新安装的存储器板的第二识别信息；

安装判定部，判定所取得的所述第二识别信息与所述安装板信息保存部所保存的所述第一识别信息是否一致；以及

初始化部，当判定为不一致时，执行所述新安装的存储器板的非易失性存储器的数据初始化，而不进行基于所述数据恢复部的脏数据的恢复。

10. 根据权利要求 9 所述的存储控制装置，其中，

具备可以装卸保持所述非易失性存储器的存储器板的多个插槽，

所述安装板信息保存部，将所述第一识别信息与安装了所述第一识别信息的存储器板的插槽的第一插槽识别信息对应起来存储，

所述板信息取得部取得所述第二识别信息和安装了该第二识别信息的存储器板的插槽的第二插槽识别信息，

所述安装判定部，判定所述第一识别信息与所述第一插槽识别信息、以及所述第二识别信息与所述第二插槽识别信息是否一致，

所述初始化部，当所述第一识别信息与所述第一插槽识别信息、以及所述

第二识别信息与所述第二插槽识别信息不一致时，执行所述新安装的存储器板的非易失性存储器的数据初始化，而不进行基于所述数据恢复部的脏数据的恢复。

11. 根据权利要求 1 至 10 中任意一项所述的存储控制装置，其中，

所述非易失性存储器具有多个非易失性存储器件，

所述转移存储部向由所述多个非易失性存储器件构成的 RAID 组存储所述脏数据。

12. 根据权利要求 11 所述的存储控制装置，其中，

所述转移存储部将所述脏数据分割成预定大小的多个数据单位，使其分散地存储于所述 RAID 组的多个所述非易失性存储器件中，并且将根据预定数量的数据单位的数据而生成的奇偶校验位存储在所述 RAID 组的所述非易失性存储器件中。

13. 根据权利要求 1 至 12 中任意一项所述的存储控制装置，其中，

所述电力供给控制部，在所述脏数据向所述非易失性存储器的转移结束后，切断从所述电池向所述易失性存储器的电力供给。

14. 根据权利要求 1 至 13 中任意一项所述的存储控制装置，其中，

所述高速缓冲存储器由多个易失性存储器件构成，

所述电力供给控制部，从由所述转移存储部结束所述脏数据的转移的各所述易失性存储器件开始依次切断电力供给。

15. 根据权利要求 1 至 14 中任意一项所述的存储控制装置，其中，

所述电力供给控制部，当通过所述电源监视部检测出所述电压异常时，不向所述请求接收部供给电力，而向所述高速缓冲存储器以及所述转移存储部供给电力。

16. 根据权利要求 15 所述的存储控制装置，其中，

在同一板上具备所述高速缓冲存储器、所述非易失性存储器、所述转移存储部以及所述电力供给控制部，

所述电池可以对所述板供给电力。

17. 根据权利要求 1 至 16 中任意一项所述的存储控制装置，其中，

所述转移存储部与所述第二数据存储部由不同的设备构成，

所述电力供给控制部，当通过所述电源监视部检测出所述电压异常时，不向所述第二数据存储部供给电力，而向所述转移存储部供给电力。

18. 根据权利要求 1 至 17 中任意一项所述的存储控制装置，其中，

具有多个所述高速缓冲存储器以及所述非易失性存储器的组，

所述第一数据存储部将所述写入访问请求对象数据存储在各组的所述高速缓冲存储器的每一个中，

所述转移存储部，当通过所述电源监视部检测出所述电压异常时，接受来自所述电池的所述电力供给，从所述多个高速缓冲存储器的某一个高速缓冲存储器中读出脏数据，使其分散地转移至所述多个非易失性存储器中。

19. 一种存储控制装置的数据管理方法，所述存储控制装置从外部装置接收写入访问请求，进行将写入访问请求对象数据写入存储装置的控制，其中，

所述存储控制装置具备：进行来自电源的电力供给的电力供给部；以可以供给电力的方式积蓄电力的电池；作为接受电力供给可以存储数据的易失性存储器的高速缓冲存储器；即使不接受电力供给也可以继续存储数据的非易失性存储器；以及检测从所述电力供给部供给的电力的电压异常的电源监视部，

接收写入访问请求；将所述写入访问请求的对象数据作为高速缓冲存储数据存储在所述高速缓冲存储器中；判定所述高速缓冲存储器的所述高速缓冲存储数据中未被反映到所述存储装置的脏数据的数据量是否超过预定的阈值；当判定为超过所述阈值时，将所述高速缓冲存储器的所述脏数据中至少一部分向所述存储装置存储；当通过所述电源监视部检测出所述电压异常时，使用来自所述电池的电力，维持向所述高速缓冲存储器的电力供给；接受来自所述电池的所述电力供给，使存储在所述高速缓冲存储器中的脏数据向所述非易失性存储器转移。

20. 一种存储控制装置，从外部装置接收写入访问请求，进行将写入访问请求对象数据写入存储装置的控制，其中，

具备：

进行来自电源的电力供给的电源电路；

以可以供给电力的方式积蓄电力的电池；

作为接受电力供给可以存储数据的易失性存储器的高速缓冲存储器；

即使不接受电力供给也可以继续存储数据的非易失性存储器;

接收来自所述外部装置的写入访问请求的接口;

与所述接口连接,并且可以进行与所述高速缓冲存储器的数据输入输出的第一处理器;

在所述高速缓冲存储器与所述非易失性存储器之间可以进行数据输入输出的第二处理器;以及

检测来自所述电源电路的电压异常的电源监视控制部,

所述第一处理器,受理所述接口接收到的写入访问请求,将所述写入访问请求的对象数据作为高速缓冲存储数据存储在所述高速缓冲存储器中,判定所述高速缓冲存储器的所述高速缓冲存储数据中未被反映到所述存储装置的脏数据的数据量是否超过预定的阈值,当判定为超过所述阈值时,将所述高速缓冲存储器的所述脏数据中至少一部分向所述存储装置存储,

所述第二处理器,当通过所述电源监视控制部检测出所述电压异常时,接受来自所述电池的所述电力供给,使所述高速缓冲存储器中存储的脏数据向所述非易失性存储器转移,

所述电源监视控制部,当检测出所述电压异常时,使用来自所述电池的电力,维持向所述高速缓冲存储器以及所述第二处理器的电力供给。

存储控制装置和数据管理方法

技术领域

本发明涉及，例如在发生了电源故障时将易失性存储器中存储的数据转移至非易失性存储器的存储控制装置和数据管理方法。

背景技术

在存储控制装置上连接了多个例如硬盘驱动器那样的存储设备。存储控制装置从主计算机接收写入命令（Write Command），向多个存储装置中的至少一个存储装置写入数据，还接收来自主计算机的读取命令（Read Command），从多个存储装置中的至少一个存储装置读出数据并发送至主计算机。

为了暂时存储按照写入命令而写入存储装置的数据，或者为了暂时存储按照读取命令而从存储装置读出的数据，在上述的存储控制装置中具备高速缓冲存储器（Cache Memory）。

作为该高速缓冲存储器，一般使用通过供给电力可以存储数据的易失性存储器。

在具备易失性存储器作为高速缓冲存储器的存储控制装置中，例如，当发生外部电源的故障等导致不向高速缓冲存储器供给电力时，高速缓冲存储器中存储的数据会丢失。

因此，为了应对这种外部电源的故障等，在存储控制装置中具备可以供给电力的电池，当发生外部电源的故障时，通过从电池对高速缓冲存储器供给电力，对高速缓冲存储器中存储的数据进行保持。

然而，直到故障解除为止，需要维持向高速缓冲存储器供给电力，因此需要较大的电池容量。从而产生存储控制装置的制造成本增加的问题。

针对此问题，公开了如下技术，通过使高速缓冲存储器的数据转移至非易失性存储器，即使通过电池向高速缓冲存储器的电力供给没有维持到故障解除，也可以保全数据（例如，专利文献1）。

【专利文献1】特开2004－21811号公报

发明内容

例如，设想使高速缓冲存储器中存储的数据转移至非易失性存储器的情况下，在使高速缓冲存储器的全部数据妥当地转移时，需要准备具有与高速缓冲存储器的容量同等容量的非易失性存储器。在这种情况下，存储控制装置的制造成本会增加。

另一方面，在为了抑制制造成本而准备容量比高速缓冲存储器小的非易失性存储器来使数据转移的情况下，无法使高速缓冲存储器的数据妥当地转移至非易失性存储器，可能会发生需要的数据消失的情况。

因此，鉴于上述问题而提出本发明，其目的在于提供一种能够抑制非易失性存储器的容量、并且可以使高速缓冲存储器中存储的数据妥当地转移的技术。

为了解决上述问题，依据本发明的一个方式的存储控制装置着眼于，在高速缓冲存储器中存储的数据中，有被反映到存储装置的数据（干净数据）、和未被反映到存储装置的数据（脏数据（Dirty Data））。即，依据本发明的一个方式的存储控制装置，根据高速缓冲存储器中存储的脏数据的数据量来决定是否将脏数据存储在存储装置中，当电压异常时，将高速缓冲存储器的脏数据转移至非易失性存储器。

具体而言，依据本发明的一个方式的存储控制装置，作为从外部装置接收写入访问请求，进行将写入访问请求对象数据写入存储装置的控制的存储控制装置，具有以下各部：进行来自电源的电力供给的电力供给部；以可以供给电力的方式积蓄电力的电池；作为接受电力供给可以存储数据的易失性存储器的高速缓冲存储器；即使不接受电力供给也可以继续存储数据的非易失性存储器；接受所述电力供给，接收写入访问请求的请求接收部；第一数据存储部，将所述写入访问请求的对象数据作为高速缓冲存储数据，存储在所述高速缓冲存储器中；判定部，判定所述高速缓冲存储器的所述高速缓冲存储数据中未被反映到所述存储装置的脏数据的数据量是否超过预定的阈值；第二数据存储部，当判定为超过所述阈值时，将所述高速缓冲存储器的所述脏数据中至少一部分向所述存储装置存储；电源监视部，检测从所述电力供给部供给的电力的电压异常；转移存储部，当通过所述电源监视部检测出所述电压异常时，接受

来自所述电池的所述电力供给，使所述高速缓冲存储器中存储的脏数据向所述非易失性存储器转移；以及电力供给控制部，当通过所述电源监视部检测出所述电压异常时，使用来自所述电池的电力，维持向所述高速缓冲存储器以及所述转移存储部的电力供给。

附图说明

图 1 是本发明的第一实施方式的计算机系统的结构图。

图 2 是本发明的第一实施方式的存储器板的结构图。

图 3A 是表示本发明的第一实施方式的控制信息以及结构信息的一例的图。

图 3B 是表示本发明的第一实施方式的控制信息以及结构信息的一例的图。

图 4A 是表示关于本发明的第一实施方式的控制信息的地址管理表的一例的图。

图 4B 是表示关于本发明的第一实施方式的结构信息的地址管理表的一例的图。

图 4C 是表示关于本发明的第一实施方式的高速缓冲存储数据的地址管理表的一例的图。

图 5A 是本发明的第一实施方式的写入访问请求时处理的流程图。

图 5B 是本发明的第一实施方式的读出访问请求时处理的流程图。

图 6 是说明本发明的第一实施方式的盘子系统中的升级以及降级的图。

图 7 是本发明的第一实施方式的数据转移处理的流程图。

图 8 是说明本发明的第一实施方式的数据转移的图。

图 9 是本发明的第一实施方式的数据恢复处理的流程图。

图 10 是本发明的第一实施方式的数据恢复判定处理的流程图。

图 11 是本发明的变形例的计算机系统的结构图。

图 12 是本发明的第二实施方式的计算机系统的结构图。

图 13 是详细说明本发明的第二实施方式的存储控制装置的一部分的图。

图 14 是本发明的第二实施方式的数据转移处理的流程图。

图 15 是本发明的第二实施方式的数据恢复处理的流程图。

符号说明

10 主机装置、20 网络、100 盘子系统、200 盘控制装置、210 通道适配器、220 I/O 处理器、230 控制单元、240 连接部、250 存储器板、251 存储器控制器、252 处理器、253 高速缓冲存储器、254 共享存储器、255 非易失性存储器、256 非易失性存储器、257 电压监视控制部、270 盘适配器、300 存储装置、310 存储设备、400 电源电路、500 电池

具体实施方式

参照附图，说明本发明的实施方式。此外，以下说明的实施方式不将该发明限定于专利请求的范围内，另外，实施方式中所说明的特征的组合的全部，在发明的解决手段中未必是必需的。

<第一实施方式>

图 1 是本发明的第一实施方式的计算机系统的结构图。

计算机系统具备一台以上的主机装置 10、和一台以上的盘子系统（Disk Subsystem）100。主机装置 10 和盘子系统 100 通过网络 20 相连。作为网络，可以是 SAN（Storage Area Network）、LAN（Local Area Network）、因特网、专用线路、公共线路等，只要是能够进行数据通信的网络即可。另外，作为网络 20 中的协议，可以是光纤通道协议、TCP/IP 协议，只要是可以在主机装置 10 与盘子系统 100 之间进行数据交换的协议，任何协议均可。此外，可以代替网络 20 而通过电缆将主机装置 10 和盘子系统 100 直接连接。

主机装置 10 具备未图示的 CPU（Central Processing Unit）、未图示的存储器、键盘等输入装置、显示器等。主机装置 10 例如可以由通用计算机（个人计算机）构成。主机装置 10 中具备应用程序 11。另外，主机装置 10 中具备可以与网络 20 进行连接的端口（PORT）12。

主机装置 10 的 CPU 执行应用程序 11，由此可以对盘子系统 100 进行数据的写入访问（Write Access）、或数据的读出访问（Read Access）。

盘子系统 100 具有：作为存储控制装置的一例的盘控制装置 200、存储装置 300、多个电源电路 400 和多个电池 500。

存储装置 300 包含多个存储设备 310。存储设备 310 例如是硬盘驱动器（HDD）。在盘子系统 100 中，基于多个存储设备 310 的存储空间，可以提供

1个或多个逻辑卷。另外，在盘子系统100中也可以通过多个存储设备310内的两个以上存储设备310构成RAID（Redundant Array of Independent Disks）组，提供RAID组的存储空间作为逻辑卷。

电源电路400例如将从外部工频电源供给的电力提供给盘控制装置200的各部。在本实施方式中，附图右侧的电源电路400向盘控制装置200以虚线划分的右侧各部供给电力，附图左侧的电源电路400向盘控制装置200以虚线划分的左侧各部供给电力。

电池500积蓄了电力，可以向盘控制装置200的预定部位供给电力。在本实施方式中，附图右侧的电池500可以向附图右侧的存储器板250上的各部供给电力，附图左侧的电池500可以向附图左侧的存储器板250上的各部供给电力。

盘控制装置200具有：多个通道适配器210、多个I/O处理器220、控制单元230、连接部240、多个存储器板250和多个盘适配器270。通道适配器210、I/O处理器220、控制单元230、存储器板250和盘适配器270通过连接部240分别连接。

连接部240可以使通道适配器210、I/O处理器220、控制单元230、存储器板250、盘适配器270之间相互通信。连接部240也可以是例如通过开关动作进行数据传输的交叉开关（Crossbar Switch）。

通道适配器210具有用于与网络20连接的端口211。通道适配器210执行与通过端口211连接的主机装置10之间的通信。在本实施方式中，执行与主机装置10之间的数据读出（数据读）以及数据的写入（数据写）中的各种信息的收发。

控制单元230通过连接部240可以对通道适配器210、I/O处理器220、控制单元230、存储器板250、盘适配器270进行访问，用于管理者对所述各部进行维护管理。控制单元230可以具备例如管理者进行输入的键盘、鼠标等输入装置、CPU、ROM、RAM、硬盘驱动器、用于显示输出信息的显示器等。在本实施方式中，控制单元230从存储器板250取得在盘控制装置200的存储器板用插槽中安装的存储器板250的固有识别符，同时取得所安装的插槽的识别号码（插槽号码），将它们对应起来存储。

盘适配器270具有用于与存储装置300的各存储设备310连接的端口271。盘适配器270在与存储设备310之间进行数据收发。

I/O处理器220通过执行被读出到存储器板250上的共享存储器254（参照图2）中的程序，来执行各种控制处理。I/O处理器220对通道适配器210、存储器板250、盘适配器270之间的数据交换进行控制。例如，进行使通道适配器210接收到的数据存储到存储器板250上的高速缓冲存储器253（参照图2）中的控制。另外，I/O处理器220进行将高速缓冲存储器253中存储的数据移送至盘适配器270、或者移送至通道适配器210的控制。另外，I/O处理器220进行使盘适配器270从存储设备310取得的数据存储在高速缓冲存储器253中的控制。另外，I/O处理器220进行用于使转移存储到非易失性存储器255中的数据，在高速缓冲存储器253中恢复的处理。

存储器板250相对于盘控制装置200的存储器板用插槽可以装卸。

图2是本发明的第一实施方式的存储器板的结构图。

存储器板250具有：存储器控制器251、处理器（processor）252、高速缓冲存储器253、共享存储器254、非易失性存储器255、非易失性存储器256和电压监视控制部257。

存储器控制器251与连接部240相连，同时与处理器252、高速缓冲存储器253、共享存储器254、非易失性存储器255、非易失性存储器256相连。

存储器控制器251通过I/O处理器220、控制单元230或处理器252的控制，执行将从连接部240发送的数据存储在存储器板250内的存储器（高速缓冲存储器253、共享存储器254、非易失性存储器255或非易失性存储器256）中的处理、将存储器板250内的存储器中存储的数据向连接部240发送的处理、存储器板250内的存储器之间的数据交换处理。另外，存储器控制器251进行对在非易失性存储器255中转移存储的数据加密的处理。在本实施方式中，存储器控制器251进行数据的数据量不发生变化的加密、例如使用Caesar密码进行加密。

电压监视控制部257监视从电源电路400向存储器板250供给的电力的电压，判定电压中是否存在异常、例如是否在预定电压以下，当判定为存在异常时，向处理器252通知该情况，同时进行控制以便向存储器板250的预定部位

（例如处理器 252、存储器控制器 251、高速缓冲存储器 253、共享存储器 254 以及非易失性存储器 255、256）供给来自电池 500 的电力。另外，电压监视控制部 257 进行在后述的数据转移处理中切断来自电池 500 的电力供给的控制。

高速缓冲存储器 253 是易失性存储器，例如是 DRAM（Dynamic Random Access Memory）。高速缓冲存储器 253 暂时存储通过通道适配器 210 接收到的数据、或通过盘适配器 270 从存储设备 310 取得的数据。高速缓冲存储器 253 例如由可以独立进行输入输出动作的多个高速缓冲存储器器件构成。

处理器 252 通过执行被读出到共享存储器 254 中的程序，执行各种控制处理。例如，处理器 252 执行将高速缓冲存储器 253 中存储的高速缓冲存储数据转移存储到非易失性存储器 255 中的处理。

共享存储器 254 是易失性存储器，存储各种信息。作为存储的信息，存在例如与主机装置进行交换的数据相关的结构信息 261（例如表示数据被存储在存储设备 310 的何处的信息）以及控制信息 260（例如表示数据被存储在高速缓冲存储器 253 的何处的信息）。

非易失性存储器 255 和 256 例如是闪速存储器、MRAM（Magnetoresistive Random Access Memory）、PARM（Phase change RAM）等即使不提供电源也能够存储数据的存储器。

非易失性存储器 255，例如用于对存储在高速缓冲存储器 253 或共享存储器 254 中的数据进行转移存储。在本实施方式中，在非易失性存储器 255 中存储脏数据，因此，在该非易失性存储器 255 中仅需要具有可以对高速缓冲存储器 253 中存储的脏数据进行存储的容量。这意味着，为了可靠地进行数据转移，根据非易失性存储器 255 的容量，决定可以存储在高速缓冲存储器 253 中的脏数据的量。另外，在本实施方式中，在每个存储器板 250 上，可以使高速缓冲存储器 253 的数据转移存储到该存储器板 250 内的非易失性存储器 255 中，因此可以在每个存储器板 250 上可靠地转移存储数据。非易失性存储器 256 存储：用于将转移存储到非易失性存储器 255 中的数据恢复至原始状态的地址管理表 262、和唯一识别存储器板 250 的固有识别符 263（例如，存储器板 250 的生产号码）。

接下来，参照附图，说明共享存储器 254 中存储的结构信息以及控制信息的一例。

图 3 是表示本发明的第一实施方式的控制信息以及结构信息的一例的图。图 3A 表示本发明的第一实施方式的控制信息的一例，图 3B 表示本发明的第一实施方式的结构信息的一例。

控制信息 260 如图 3A 所示，包含将逻辑地址 2601、高速缓冲存储器地址 2602、升级 bit2603、脏 bit2604 对应起来的记录。

在逻辑地址 2601 中，对用于确定数据的逻辑上的地址（逻辑地址）进行存储。作为逻辑地址，存在例如从主机装置 10 发送的访问请求中的 LUN（Logical Unit Number）和 LBA（Logical Block Address）的组合。在本实施方式中以逻辑地址为单位进行管理，因此各记录管理的数据的数据量，成为与逻辑地址相对应的预定数据量。

在高速缓冲存储器地址 2602 中，对存储了对应的数据的高速缓冲存储器 253 的地址进行存储。

在升级 bit2603 中，对表示对应的数据与存储设备 310 中存储的数据是否一致的位进行存储。例如，在升级 bit2603 中，当对应的数据与存储设备 310 中存储的数据一致时存储“1”，当不一致时存储“0”。

在脏 bit2604 中存储，表示对应的数据是被反映到存储设备 310 中的数据（干净数据（Clean Data））还是未被反映的数据（脏数据）的位。例如，在脏 bit2604 中，当对应的数据是干净数据时存储“0”，当是脏数据时存储“1”。在脏 bit2604 中存储了“0”的数据、即干净数据存在于存储设备 310 中。因此，即使停止向高速缓冲存储器 253 的电力供给，将其从高速缓冲存储器 253 中删除，也可以从存储设备 310 取出。因此，即使停止电力供给，数据也不会从盘子系统 100 消失。另一方面，在脏 bit2604 中存储了“1”的数据、即脏数据存在于高速缓冲存储器 253 上，但未被反映到存储设备 310 中。因此，当停止向高速缓冲存储器 253 的电力供给时，数据消失，并从盘子系统 100 完全消失。因此，在盘控制装置 200 中，当发生了电压异常时，将脏数据从高速缓冲存储器 253 转移存储至非易失性存储器 255 中。这样，由于使脏数据转移到非易失性存储器 255 中，可以防止数据从盘子系统 100 完全消失。在本实施方式中，

以脏数据作为转移对象，不以干净数据作为转移对象，因此可以减小数据转移所需的非易失性存储器的容量，并且可以迅速地进行数据转移处理。

结构信息 261 如图 3B 所示，包含将逻辑地址 2611、物理地址 2612 对应起来的记录。

在逻辑地址 2611 中存储用于确定数据的逻辑地址。作为逻辑地址，存在例如从主机装置 10 发送的访问命令中的 LUN（Logical Unit Number）和 LBA（Logical Block Address）的组合。在物理地址 2612 中存储，表示存储了对应的逻辑地址的数据的存储设备 310 以及该存储设备 310 中的存储区域的物理的地址（物理地址）。

接下来，参照附图说明非易失性存储器 256 中存储的地址管理表的一例。

图 4 是表示本发明的第一实施方式的地址管理表的一例的图。图 4A 表示关于本发明的第一实施方式的控制信息的地址管理表的一例，图 4B 表示关于本发明的第一实施方式的结构信息的地址管理表的一例，图 4C 表示关于本发明的第一实施方式的高速缓冲存储数据的地址管理表的一例。

地址管理表 262 包含：用于对非易失性存储器 255 中存储的控制信息的地址进行管理的控制信息地址管理表 262A、用于对非易失性存储器 255 中存储的结构信息的地址进行管理的地址管理表 262B、以及用于对非易失性存储器 255 中存储的高速缓冲存储数据的地址进行管理的地址管理表 262C。

控制信息的地址管理表 262A 包含将非易失性存储器地址 2621、共享存储器地址 2622 和数据长度 2623 对应起来的记录。

在非易失性存储器地址 2621 中，对可以分配给控制信息的存储的非易失性存储器 255 上的地址（非易失性存储器地址）进行存储。在共享存储器地址 2622 中，对控制信息在共享存储器 254 上存储的地址（共享存储器地址）进行存储，所述控制信息被分配了从对应的非易失性存储器地址开始的存储区域。在数据长度 2623 中存储对应的控制信息在非易失性存储器 255 上的数据长度。

结构信息的地址管理表 262B 包含将非易失性存储器地址 2624、共享存储器地址 2625、和数据长度 2626 对应起来的记录。

在非易失性存储器地址 2624 中，对可以分配给结构信息的存储的非易失

性存储器255上的地址（非易失性存储器地址）进行存储。在共享存储器地址2625中，对结构信息在共享存储器254上存储的地址（共享存储器地址）进行存储，所述结构信息被分配了从对应的非易失性存储器地址开始的存储区域。在数据长度2626中存储对应的结构信息在非易失性存储器255上的数据长度。

高速缓冲存储数据的地址管理表262C包含将非易失性存储器地址2627、高速缓冲存储器地址2628、和数据长度2629对应起来的记录。

在非易失性存储器地址2627中，对可以分配给高速缓冲存储数据的存储的非易失性存储器255上的地址（非易失性存储器地址）进行存储。在高速缓冲存储器地址2628中，对高速缓冲存储数据在高速缓冲存储器253上存储的地址（高速缓冲存储器地址）进行存储，所述高速缓冲存储数据被分配了从对应的非易失性存储器地址开始的存储区域。在数据长度2629中存储对应的高速缓冲存储数据在非易失性存储器255上的数据长度。

接下来，对本发明的第一实施方式的盘控制装置的处理动作进行说明。

图5A是本发明的第一实施方式的写入访问请求时处理的流程图。

当盘子系统100的通道适配器210通过端口211接收从主机装置10发送的写入请求，I/O处理器220取得该写入访问请求时，开始执行写入访问请求时处理。

首先，当I/O处理器220从通道适配器210取得写入访问请求时（步骤S11），I/O处理器220从通道适配器210取得写入访问请求的对象数据（写入数据），将该写入数据写入高速缓冲存储器253（步骤S12）。然后，I/O处理器220更新共享存储器254的结构信息261中与该写入数据所对应的记录（步骤S13）。即，I/O处理器220，在结构信息261中与写入数据所对应的记录的高速缓冲存储器地址2602中，对存储了写入数据的高速缓冲存储器253的高速缓冲存储器地址进行存储，而且在脏bit2604中存储表示是脏数据的“1”。

接着，I/O处理器220检测高速缓冲存储器253中存储的脏数据的数据量，判定是否超过预先设定的阈值（写入高速缓冲存储阈值）（步骤S14）。在此，可以参照共享存储器254的结构信息261，根据结构信息261的脏bit2604中存储了“1”的地址数，检测高速缓冲存储器253中存储的脏数据的数据量。

另外，写入高速缓冲存储阈值表示，若高速缓冲存储器253的脏数据的数据量在该阈值以下，则可以将该脏数据向非易失性存储器255可靠地转移存储。该写入高速缓冲存储阈值，例如可以根据管理者输入的指示由控制单元 230 设定，另外也可以由控制单元230根据非易失性存储器255的数据容量设定为一个阈值，另外，也可以由控制单元230根据盘子系统100的动作状态和非易失性存储器255的数据容量来动态地设定阈值。作为写入高速缓冲存储阈值，可以设定为非易失性存储器255的容量的例如50％～80％之间的任意容量。

当步骤 S14 的判定结果为脏数据的数据量超过写入高速缓冲存储阈值时（步骤S14，是），I/O处理器220使至少一部分脏数据降级（步骤S15）。即，I/O处理器220使高速缓冲存储器253的脏数据的至少一部分存储在存储设备310中。由此，该数据是高速缓冲存储器253的内容被反映到存储设备310中的数据。此外，作为降级的脏数据，例如可以将访问频率较少的脏数据作为对象，另外也可以将从上一次访问起经过了最长时间的脏数据作为对象。

接着，I/O处理器220更新共享存储器254的结构信息261中的降级的数据所对应的记录（步骤S16）。即，I/O处理器220在结构信息261中的降级的数据所对应的记录的脏bit2604中存储表示是干净数据的“0”，结束写入访问请求时处理。由此，在后述的数据转移处理中，可以将高速缓冲存储器 253 内的脏数据可靠地转移存储在非易失性存储器255中。

另一方面，当步骤S14的判定的结果为脏数据的数据量未超过写入高速缓冲存储阈值时（步骤 S14，否），表示高速缓冲存储器 253 内的脏数据可以可靠地转移存储在非易失性存储器255中，因此I/O处理器220结束写入访问请求时处理。此外，将高速缓冲存储器253的脏数据存储在存储设备310中的降级处理（与步骤S15和步骤S16相同的处理），不仅在写入访问请求时处理中，例如在I/O处理器220的处理负荷较轻等情况下，也适宜由I/O处理器220执行。

图5B是本发明的第一实施方式的读出访问请求处理时的流程图。

当盘子系统100的通道适配器210通过端口 211 接收从主机装置10发送的读出访问请求， I/O处理器220取得该读出访问请求时，开始执行读出访问请求时处理。

首先，当 I/O 处理器 220 从通道适配器 210 取得读出访问请求时（步骤 S21），I/O 处理器 220 判定读出访问请求的对象数据（读取数据（Read Data））是否存储在高速缓冲存储器 253 中（步骤 S22）。在此，例如通过确认在共享存储器 254 中的控制信息 260 中是否存储了读出访问请求内包含的逻辑地址所对应的记录，可以判定读取数据是否存储在高速缓冲存储器 253 中。

当步骤 S22 中的判定的结果为存储在高速缓冲存储器 253 中时（步骤 S22，是），I/O 处理器 220 从高速缓冲存储器 253 读出对应的高速缓冲存储数据，通过通道适配器 210，对读出到作为请求源的主机装置 10 的数据进行发送（步骤 S23），结束读出访问请求时处理。

另一方面，当步骤 S22 中的判定的结果为未存储在高速缓冲存储器 253 中时（步骤 S22，否），I/O 处理器 220 使对应的数据升级（步骤 S24）。即，I/O 处理器 220 从存储着对应的数据的存储设备 310 中读出该数据，存储在高速缓冲存储器 253 中。然后，I/O 处理器 220 更新共享存储器 254 的结构信息 261 中的与读出的数据对应的记录（步骤 S25）。即，I/O 处理器 220 在共享存储器 254 的控制信息 260 中追加与读出的数据对应的记录，在该记录的高速缓冲存储器地址 2602 中，对存储了读出的数据的高速缓冲存储器 253 的高速缓冲存储器地址进行存储，而且在脏 bit2604 中存储表示是干净数据的“0”。然后，I/O 处理器 220 将读出到该高速缓冲存储器 253 中的数据，通过通道适配器 210 发送到作为请求源的主机装置 10，结束读出访问请求时处理。

图 6 是说明本发明的第一实施方式的盘子系统中的升级以及降级的图。

所谓升级（Stage），如图 6 所示，是指将存储在存储设备 310 中的数据存储在高速缓冲存储器 253 中；所谓降级（Destage），如图 6 所示，是指使存储在高速缓冲存储器 253 中的高速缓冲存储数据反映到存储设备 310 中。

接下来，说明本发明的第一实施方式的盘子系统 100 中的数据转移处理。

图 7 是本发明的第一实施方式的数据转移处理的流程图。

通过电压监视控制部 257 检测出电源故障、例如从电源电路 400 供给的电压表示异常值，来开始数据转移处理（步骤 S31）。电压监视控制部 257 向存储器板 250 的处理器 252 通知发生了电压异常，同时将向存储器板 250 的各部供给的电力，从由电源电路 400 供给的电力切换为由电池 500 供给的电力（步

骤 S32)。由此，存储器板 250 的各部可以通过由电池 500 供给的电力而继续动作。因此，高速缓冲存储器 253 和共享存储器 254 可以维持数据的存储。此外，在以下的处理中，仅向存储器板 250 供给电池 500 的电力即可。因此，可以减少在电池 500 中应积蓄的电量。

处理器 252 参照共享存储器 254 的控制信息 260(步骤 S33)，以高速缓冲存储器 253 中的一个高速缓冲存储器器件作为处理对象，判定是否存在未转移的脏数据(步骤 S34)。

当步骤 S34 的判定的结果为，在高速缓冲存储器器件中存在未转移的脏数据时(步骤 S34，是)，处理器 252 从该高速缓冲存储器器件读出脏数据(步骤 S35)，根据非易失性存储器 256 的地址管理表 262，决定存储该脏数据的非易失性存储器 255 的地址(非易失性存储器地址)，在该非易失性存储器地址所对应的记录的高速缓冲存储器地址 2628 中，对存储了该脏数据的高速缓冲存储器 253 的地址进行存储，同时在数据长度 2629 中存储该脏数据的数据长度(步骤 S36)。

然后，处理器 252 将脏数据与存储该脏数据的非易失性存储器地址一起移送至存储器控制器 251。存储器控制器 251 将从处理器 252 移送的脏数据加密(步骤 S37)，存储在非易失性存储器 255 中的指定的非易失性存储器地址(步骤 S38)。这样，脏数据被加密并被存储在非易失性存储器 255 中，因此，即使从非易失性存储器 255 中读取数据，也无法根据该数据容易地掌握原来的数据，因此可以妥当地防止信息泄漏。

然后，处理器 252 以同一高速缓冲存储器器件为对象，重复从上述步骤 S33 开始的处理。通过如此重复处理，可以使同一高速缓冲存储器器件中存储的全部脏数据转移至非易失性存储器 255 中。

另一方面，当步骤 S34 的判定的结果为，高速缓冲存储器器件中没有未转移的脏数据时(步骤 S34，否)，表示作为对象的高速缓冲存储器器件中不存在脏数据，或者已使该高速缓冲存储器器件的全部脏数据转移，因此处理器 252 通过电压监视控制部 257 切断向该高速缓冲存储器器件的电力供给(步骤 S39)，判定是否存在成为转移脏数据处理的对象的其它高速缓冲存储器器件(步骤 S40)。

当步骤S40的结果为，存在作为转移脏数据处理的对象的其它高速缓冲存储器器件时（步骤 S40，是），针对其它高速缓冲存储器器件执行与上述相同的从步骤S33开始的处理。

另一方面，当步骤S40的结果为，不存在作为转移脏数据处理的对象的其它高速缓冲存储器器件时（步骤S40，否），意味着高速缓冲存储器253的全部脏数据的转移已完成，因此，处理器252从共享存储器254读出结构信息261、以及控制信息260中与脏数据相关的控制信息（步骤S41），按照非易失性存储器 256 的控制信息的地址管理表 262A 以及结构信息的地址管理表262B，决定存储结构信息以及控制信息的非易失性存储器 255 的地址（非易失性存储器地址），在该非易失性存储器地址所对应的记录的共享存储器地址2622、2625 中，对存储了该结构信息或控制信息的共享存储器 254 的地址进行存储，同时在数据长度2623、2626中存储该结构信息或控制信息的数据长度（步骤S42）。

接着，处理器252将结构信息和控制信息、以及存储该结构信息和控制信息的非易失性存储器地址移送至存储器控制器251。存储器控制器251将从处理器 252 移送来的结构信息和控制信息加密（步骤 S43），存储在非易失性存储器 255 的指定的非易失性存储器地址（步骤S44）。然后，处理器252通过电压监视控制部257切断向该存储器板250的电力供给（步骤S45）。

在本实施方式中，在同一存储器板250上具有与上述数据转移处理相关的存储器控制器 251、高速缓冲存储器 253、共享存储器 254、非易失性存储器255、256以及处理器252，因此可以迅速地进行数据转移处理。

图8是说明本发明的第一实施方式的数据转移的图。

当执行上述图7所示的数据转移处理时，作为高速缓冲存储器253中存储的脏数据的数据 d2 被转移至非易失性存储器255。另外，共享存储器254的控制信息261也被转移至非易失性存储器255中。另外，共享存储器254的控制信息260内的、数据d2的控制信息也被转移至非易失性存储器255中。此时，在非易失性存储器256中存储地址管理表262，该地址管理表262表示被转移至非易失性存储器 255 的数据 d2、结构信息以及控制信息的原来的存储目的地。

图 9 是本发明的第一实施方式的数据恢复处理的流程图。

当盘控制装置 200 的电源恢复时，开始数据恢复处理（步骤 S51），首先，I/O 处理器 220 参照非易失性存储器 256 内的地址管理表 262（步骤 S52），判定是否存储了应该恢复的数据等（高速缓冲存储数据、结构信息、控制信息）（步骤 S53）。此外，可以根据在地址管理表 262 的共享存储器地址 2622、2625、或高速缓冲存储器地址 2628 中是否存储了地址，来判定是否存储了数据等。

当该判定的结果为存储了应该恢复的数据等时（步骤 S53，是），I/O 处理器 220 按照地址管理表 262，从非易失性存储器 255 的对应的地址将数据等读出到存储器控制器 251 中，而且将该读出的数据等解密，将相应的数据等的地址变换为易失性存储器（高速缓冲存储器 253 或共享存储器 254）用的地址（步骤 S56）。即，从地址管理表 262 取得对应的数据等的共享存储器地址 2622、2625、或高速缓冲存储器地址 2628。

接着，I/O 处理器 220，通过存储器控制器 251，根据变换而得的地址，将数据等写入共享存储器 254 或高速缓冲存储器 253（步骤 S57），判定是否存在应该恢复的其它数据等（步骤 S58），当存在应该恢复的其它数据时（步骤 S58，是），通过重复执行从上述步骤 S54 开始的处理，将数据转移前的脏数据以及与脏数据相关的结构信息和控制信息恢复为原始状态。由此，可以与数据转移前同样地在各种处理中利用脏数据。

另一方面，当未存储应该恢复的数据等时（步骤 S53，否）、或者应该恢复的全部数据的恢复结束时（步骤 S58，否），转移至通常的 I/O 处理（步骤 S59）。

本实施方式中的存储器板 250，如上所述，相对于盘控制装置 200 可以装卸，另外，高速缓冲存储数据被转移至存储器板 250 的非易失性存储器 255 中。因此，当将转移了高速缓冲存储数据的存储器板 250 从盘控制装置 200 拆除，安装于其它盘控制装置 200 时，高速缓冲存储数据的内容可能会泄漏。因此，在本实施方式中执行以下的数据恢复判定处理，转移到存储器板 250 中的数据不会泄漏。

图 10 是本发明的第一实施方式的数据恢复判定处理的流程图。

I/O 处理器 220，当检测出在盘控制装置 200 的存储器板插槽中插入了存

储器板 250 时（步骤 S61），从所安装的存储器板 250 的非易失性存储器 256 取得存储器板 250 的固有识别符 263，根据该固有识别符 263 和安装了该存储器板 250 的插槽的号码，判定数据恢复的必要性（步骤 S62）。在本实施方式中，I/O 处理器 220 从控制单元 230 取得以前所安装的存储器板 250 的固有识别符和插槽号码，根据是否与新安装的存储器板 250 的固有识别符 263 和插槽号码一致来判定恢复的必要性。即，当固有识别符 263 和插槽号码一致时，意味着暂时拆除存储器板 250 后，再次在同一插槽中插入了同一存储器板 250，因此进行数据恢复，当固有识别符 263 不同时，存储器板不是以前安装在该盘控制装置 200 中的存储器板，因此，为防止该存储器板的数据泄漏而不进行数据恢复，另外，即使固有识别符相同但插槽号码不同时，意味着进行与数据转移时不同的连接，因此不进行数据恢复。

当上述判定的结果为不需要数据恢复时（步骤 S63，否），为了可靠地防止数据的泄漏，I/O 处理器 220 通过存储器控制器 251 对非易失性存储器 255 的数据进行初始化，例如在全部存储区域中写入“0”（步骤 S64），转移至通常 I/O 处理（步骤 S66）。

另一方面，当判定为需要数据恢复时（步骤 S63，是），执行数据恢复处理（步骤 S65：与图 9 的步骤 S52 以后的处理相同），转移至通常 I/O 的处理（步骤 S66）。

接下来，对上述第一实施方式的计算机系统的变形例进行说明。

图 11 是本发明的变形例的计算机系统的结构图。此外，对与第一实施方式相同的功能部分赋予相同号码，省略说明。

变形例的盘子系统 101 的盘控制装置 201 如下构成，代替第一实施方式的盘控制装置 200 中的通道适配器 210 而具备通道适配器 212，代替盘适配器 270 而具备盘适配器 272，在与存储器板 250 不同的共享存储器板 265 上具备存储器板 250 的共享存储器 254，拆除了 I/O 处理器 220。

通道适配器 212 相对于通道适配器 210 还具备处理器 213。盘适配器 272 在盘适配器 270 上还具备处理器 273。处理器 213 和处理器 273 分散执行通过 I/O 处理器 220 所执行的处理。

在这种盘控制装置 201 中也可以执行与上述图 7、图 9、图 10 相同的处理，

可以取得同样的效果。在这种情况下，在图 7、图 9 中处理器 252 所执行的处理，例如由处理器 213、273 的某个执行即可，另外，图 10 中的 I/O 处理器 220 的处理，例如由处理器 213、273 的某个执行即可。

<第二实施方式>

图 12 是本发明的第二实施方式的计算机系统的结构图。此外，对于与第一实施方式相同的功能部分赋予相同的号码。

盘控制装置 202 具备结构相同的多个集群（Cluster）203。各集群 203 例如由一个子系统控制板构成，具有：通道适配器 210、I/O 处理器 280、子系统控制器 281、易失性存储器 282、非易失性存储器 283、盘适配器 270 和电压监视控制部 257。

电源电路 400 例如将从外部工频电源供给的电力提供给盘控制装置 202 的各部。在本实施方式中，电源电路 400 没有被多重化，向多个集群 203 的各部供给电力。此外，也可以具备多个电源电路 400，分别向各集群 203 供给电力。

电池 500 积蓄了电力，可以向盘控制装置 202 的预定部位供给电力。在本实施方式中，电池 500 没有被多重化，向多个集群 203 的预定部位供给电力。此外，也可以具备多个电池 500，分别向各集群 203 的预定部位供给电力。

I/O 处理器 280 通过执行被读出至易失性存储器 282 中的程序，控制集群 203 的整体的动作。I/O 处理器 280 对通道适配器 210、易失性存储器 282、非易失性存储器 283、盘适配器 270 之间的子系统控制器 281 的数据交换进行控制。例如，进行使通道适配器 210 接收到的数据存储在易失性存储器 282 中的控制。另外，I/O 处理器 280 进行将存储在易失性存储器 282 中的数据向盘适配器 270 移送、或者向通道适配器 270 移送的控制。另外，I/O 处理器 280 进行使盘适配器 270 从存储设备 310 取得的数据存储在易失性存储器 282 中的控制。另外，I/O 处理器 280 进行用于在易失性存储器 282 中，对在非易失性存储器 283 中转移存储的数据进行恢复的处理。

子系统控制器 281 与通道适配器 210、盘适配器 270、易失性存储器 282、非易失性存储器 283、处理器 280 以及其它集群 203 的子系统控制器 281 相连，对各部间交换的数据进行中继。通过此结构，子系统控制器 281 通过 I//O 处理器 280 的控制，将通过通道适配器 210 从主机装置 10 接收到的写入数据存

储在易失性存储器 282 中，同时将写入数据向其它集群 203 的子系统控制器 281 发送，可以使写入数据存储在其它集群 203 侧的易失性存储器 282 中。另外，在本实施方式中，子系统控制器 281 通过 I/O 处理器 280 的控制，将数据向其它集群 203 的子系统控制器 281 发送，可以使其存储在其它集群 203 的非易失性存储器 283 中，或者可以从其它集群 203 的非易失性存储器 283 中读出数据。

另外，子系统控制器 281 执行向由非易失性存储器 283 的多个非易失性存储器器件 2831（参照图 13）构成的 RAID 组的数据存储处理。例如，子系统控制器 281，当在 RAID 组中进行存储时，将存储对象数据分割为预定的数据单位，与此同时，针对多个（例如 3 个）数据单位中的每一个，通过奇偶校验位（Parity）生成电路 2811 生成所述每个数据单位的奇偶校验位，将所述多个数据单位和生成的奇偶校验位存储在 RAID 组内的不同非易失性存储器器件 2831 中。在本实施方式中，子系统控制器 281 对数据单位和奇偶校验位进行加密，并存储在非易失性存储器器件 2831 中。

接下来，详细说明易失性存储器 282 和非易失性存储器 283。

图 13 是详细说明本发明的第二实施方式的存储控制装置的一部分的图。

易失性存储器 282 存储与第一实施方式的共享存储器 254 和高速缓冲存储器 253 相同的各种信息。作为存储的信息，存在例如与主机装置 10 进行交换的数据相关的结构信息 261（例如表示被存储在存储设备 310 的何处的信息）以及控制信息 260（例如表示被存储在易失性存储器 282 的何处的信息）。另外，易失性存储器 282 暂时存储通过通道适配器 210 接收的数据、和通过盘适配器 270 从存储设备 310 取得的数据。易失性存储器 282，例如由可以独立地进行输入输出动作的多个易失性存储器器件构成。

非易失性存储器 283 是例如闪速存储器、MRAM（Magnetoresistive Random Access Memory）、PRAM（Phase change RAM）等即使不供给电源也可以存储数据的存储器。非易失性存储器 283 由多个非易失性存储器器件 2831 构成。非易失性存储器 283 例如用于转移存储在易失性存储器 282 中存储的数据（高速缓冲存储数据、结构信息、控制信息）。在本实施方式中，向由多个集群 203 的非易失性存储器 283 的多个非易失性存储器器件 2831 构成的 RAID 组，存

储高速缓冲存储数据、结构信息、以及控制信息。当在 RAID 组中进行存储时，例如，可以是 RAID 级别 2～5 中的任意一种。若是这些 RAID 级别，则可以抑制非易失性存储器 283 所需的容量，并可以提高数据的可靠性。另外，非易失性存储器 283 存储地址管理表 262，该地址管理表 262 用于将转移存储在非易失性存储器 283 中的数据恢复到原始状态。此外，地址管理表 262 中的非易失性存储器地址 2621、2624、2627，在第二实施方式中不是非易失性存储器 283 的物理地址，而成为 RAID 组的逻辑存储区域中的地址（逻辑地址）。子系统控制器 281 可以根据该逻辑地址确定物理地址（即，是哪个非易失性存储器器件 2831（也包含其它集群 203 的非易失性存储器器件 2831）的哪个地址）。另外，在地址管理表 262 的共享存储器地址 2622、共享存储器地址 2625、以及高速缓冲存储器地址 2628 中存储易失性存储器 282 中的地址。

图 14 是本发明的第二实施方式的数据转移处理的流程图。

通过电压监视控制部 257 检测出电源故障、例如从电源电路 400 供给的电压表示异常值，由此开始数据转移处理（步骤 S71）。电压监视控制部 257 向 I/O 处理器 280 通知发生了电压异常，同时将向集群 203（子系统控制板）的各部供给的电力，从由电源电路 400 供给的电力切换为由电池 500 供给的电力（步骤 S72）。由此，子系统控制板的各部可以通过由电池 500 供给的电力继续动作。因此，易失性存储器 282 可以维持数据的存储。

I/O 处理器 280 参照易失性存储器 282 的控制信息 260（步骤 S73），以易失性存储器 282 中的一个易失性存储器器件作为处理对象，判定是否存在未转移的脏数据（步骤 S74）。

当步骤 S74 的判定的结果为，在易失性存储器器件中存在未转移的脏数据时（步骤 S74，是），I/O 处理器 280 从该易失性存储器器件中读出脏数据（步骤 S75），按照非易失性存储器 283 的地址管理表 262，决定存储该脏数据的非易失性存储器 283 的逻辑地址，在该非易失性存储器 283 的逻辑地址所对应的记录的高速缓冲存储器地址 2628 中，对存储了该脏数据的易失性存储器 282 的地址进行存储，另外在数据长度 2629 中存储该脏数据的数据长度（步骤 S76）。

接着，I/O 处理器 280 将脏数据与存储该脏数据的非易失性存储器 283 的

逻辑地址一起移送至子系统控制器281。子系统控制器281将从I/O处理器280移送来的脏数据分割为预定大小（数据量）的数据单位，针对多个数据单位中的每一个，生成所述数据单位所对应的奇偶校验位（步骤 S77），将多个数据单位以及生成的奇偶校验位加密（步骤 S78）。接着，子系统控制器 281 根据所指定的非易失性存储器283的逻辑地址，确定存储每个数据单位以及奇偶校验位的物理地址，在对应的物理地址所表示的非易失性存储器器件2831中存储各数据单位以及奇偶校验位（步骤 S79）。在本实施方式中，也存储在其它集群203的非易失性存储器器件2831中。由此，数据和这些数据所对应的奇偶校验位通过多个非易失性存储器器件2831而分散存储。从而，即使某一个非易失性存储器器件2831中发生了故障，也可以恢复原来的数据。另外，脏数据被加密地存储在非易失性存储器 283 中，因此，即使从非易失性存储器283读取数据，也无法根据该数据容易地掌握原来的数据，因此可以妥当地防止信息泄漏。

接着，I/O处理器280以同一易失性存储器器件为对象，重复上述从步骤S73开始的处理。通过如此重复处理，可以使存储在同一易失性存储器器件中的全部脏数据转移至非易失性存储器283中。

另一方面，当步骤S74的判定的结果为，在易失性存储器器件中没有未转移的脏数据时（步骤 S74，否），表示在作为对象的易失性存储器器件中不存在脏数据、或已将该易失性存储器器件的全部脏数据转移，因此I/O处理器280通过电压监视控制部257切断向该易失性存储器器件的电力供给（步骤S80），判定是否存在成为转移脏数据处理的对象的其它易失性存储器器件（步骤S81）。

当步骤S81的结果为，存在成为转移脏数据处理的对象的其它易失性存储器器件时（步骤 S81，是），对其它易失性存储器器件执行上述同样的从步骤S73开始的处理。

另一方面，当步骤S81的结果为，不存在作为转移脏数据处理的对象的其它易失性存储器器件时（步骤 S81，否），意味着易失性存储器 282 的全部脏数据的转移已完成，因此 I/O 处理器 280 从易失性存储器 282 读出结构信息261、和控制信息260中与脏数据相关的控制信息（步骤S82），按照非易失性

存储器 283 的地址管理表 262，决定存储该结构信息 261 和控制信息 260 的非易失性存储器 283 的逻辑地址，在该非易失性存储器的逻辑地址所对应的记录的共享存储器地址 2622 或 2625 中，对存储了该结构信息 261 或控制信息 260 的易失性存储器 282 的地址进行存储，另外，在数据长度 2623、2626 中存储该结构信息 261 或控制信息 260 的数据长度（步骤 S83）。

接着，I/O 处理器 280 将结构信息 261、控制信息 260 与存储的非易失性存储器 283 的逻辑地址一起移送至子系统控制器 281。子系统控制器 281 将从 I/O 处理器 280 移送来的结构信息 261 和控制信息 260 分割为预定大小（数据量）的数据单位，针对预定数量的数据单位中的每一个，生成所述数据单位所对应的奇偶校验位（步骤 S84），将多个数据单位和生成的奇偶校验位加密（步骤 S85）。接着，子系统控制器 281 根据所指定的非易失性存储器 283 的逻辑地址，确定存储每一个数据单位和奇偶校验位的物理地址，在对应的物理地址表示的非易失性存储器器件 2831 中存储各数据单位和奇偶校验位（步骤 S86）。由此，通过多个非易失性存储器器件 2831 分散存储数据和这些数据所对应的奇偶校验位。从而，即使某一个非易失性存储器器件 2831 中发生故障，也可以恢复原来的数据。

接着，I/O 处理器 280 通过电压监视控制部 257 切断向集群 203 的所有部位的电力供给（步骤 S87）。

图 15 是本发明的第二实施方式的数据恢复处理的流程图。

在盘控制装置 202 的电源恢复了的情况下，开始数据恢复处理（步骤 S91），首先，I/O 处理器 280 参照非易失性存储器 283 内的地址管理表 262（步骤 S92），判定是否存储了应该恢复的数据等（高速缓冲存储数据、结构信息以及控制信息）（步骤 S93）。此外，可以根据是否在地址管理表 262 的共享存储器地址 2622、2625 或高速缓冲存储器地址 2628 中存储了地址，来判定是否存储了数据等。

当该判定的结果为存储了应该恢复的数据等时（步骤 S93，是），I/O 处理器 280 向子系统控制器 281 移送从地址管理表 262 取得的非易失性存储器 283 的逻辑地址。子系统控制器 281 取得该逻辑地址所对应的物理地址，从该物理地址表示的非易失性存储器 283 读取数据等，而且将该读取的数据等解密（步

骤 S95），进行各数据等的奇偶校验（步骤 S96）。由此，当预定数量的数据单位与其所对应的奇偶校验位有预定的关系时，由于意味着数据正确，因此直接进行下面的处理，另一方面，当没有预定的关系时，再生数据来进行下面的处理。

接着，子系统控制器 281 通过将多个数据单位按原样排列而得到原来的数据，并移送至 I/O 处理器 280。I/O 处理器 280 从地址管理表 262 的共享存储器地址 2622、共享存储器地址 2625 或高速缓冲存储器地址 2628 中，取得数据在转移时被存储的易失性存储器 282 的地址（步骤 S97）。然后，I/O 处理器 280 通过子系统控制器 281，在取得的易失性存储器 282 的地址存储从非易失性存储器 283 取得的数据（步骤 S98）。此外，在本实施方式中，也向其它集群的子系统控制器 281，同样地也向其它集群 203 的易失性存储器 282 存储数据。由此，可以使多个集群的易失性存储器 282 的状态相同。

接着，I/O 处理器 280 判定是否存在应该恢复的其它数据等（步骤 S99），当存在应该恢复的其它数据时（步骤 S99，是），通过重复执行从上述步骤 S94 开始的处理，将数据转移前的脏数据以及与脏数据有关的结构信息和控制信息恢复为原来的状态。由此，可以与数据转移前同样地在各种处理中利用脏数据。

另一方面，当未存储应该恢复的数据等时（步骤 S93，否），或应该恢复的全部数据的恢复已结束时（步骤 S99，否），转移至通常的 I/O 处理（步骤 S100）。

以上，根据一个实施方式说明了本发明，但本发明不限于上述实施方式，可以应用于其它各种形态。

例如，在上述各实施方式中表示了作为存储设备 310 而使用了硬盘驱动器（HDD）的例子，但本发明不限于此，例如可以将硬盘驱动器的至少一部分或全部替换为 DVD 驱动器、磁带驱动器、闪速存储器设备等可以存储数据的其它存储设备。

另外，在上述第一实施方式中以共享存储器 254 作为易失性存储器进行了说明，但本发明不限于此，例如可以作为非易失性存储器。在以共享存储器 254 作为非易失性存储器的情况下，当进行数据转移时也可以不进行控制信息 260、结构信息 261 的转移处理。

另外，在上述第一实施方式中说明了将高速缓冲存储器253和共享存储器254在物理上分离的结构，但不限于此，也可以将高速缓冲存储器253和共享存储器254构成为一个集合体。

另外，在上述第一实施方式中说明了将非易失性存储器255和非易失性存储器256在物理上分离的结构，但不限于此，也可以将非易失性存储器构成为一个集合体。

在上述各实施方式中使用了数据量不发生变化的加密，但本发明不限于此，也可以使用例如数据量变化的加密。此外，在这种情况下，需要使地址管理表262中存储的数据长度成为加密后的数据长度。

另外，在上述各实施方式中，将高速缓冲存储器253的脏数据以本来的数据长度存储在非易失性存储器255中，但本发明不限于此，例如可以将高速缓冲存储器253的脏数据压缩后存储在非易失性存储器255中。这样一来，可以提高非易失性存储器255中的存储效率，并且可以缩短向非易失性存储器255的写入处理所需的时间。

另外，在上述第二实施方式中，各集群203中同样地具备非易失性存储器283，向由多个集群203的非易失性存储器283构成的RAID组转移脏数据，但本发明不限于此，例如可以仅在一方的集群203中具备非易失性存储器283，用于脏数据的转移。

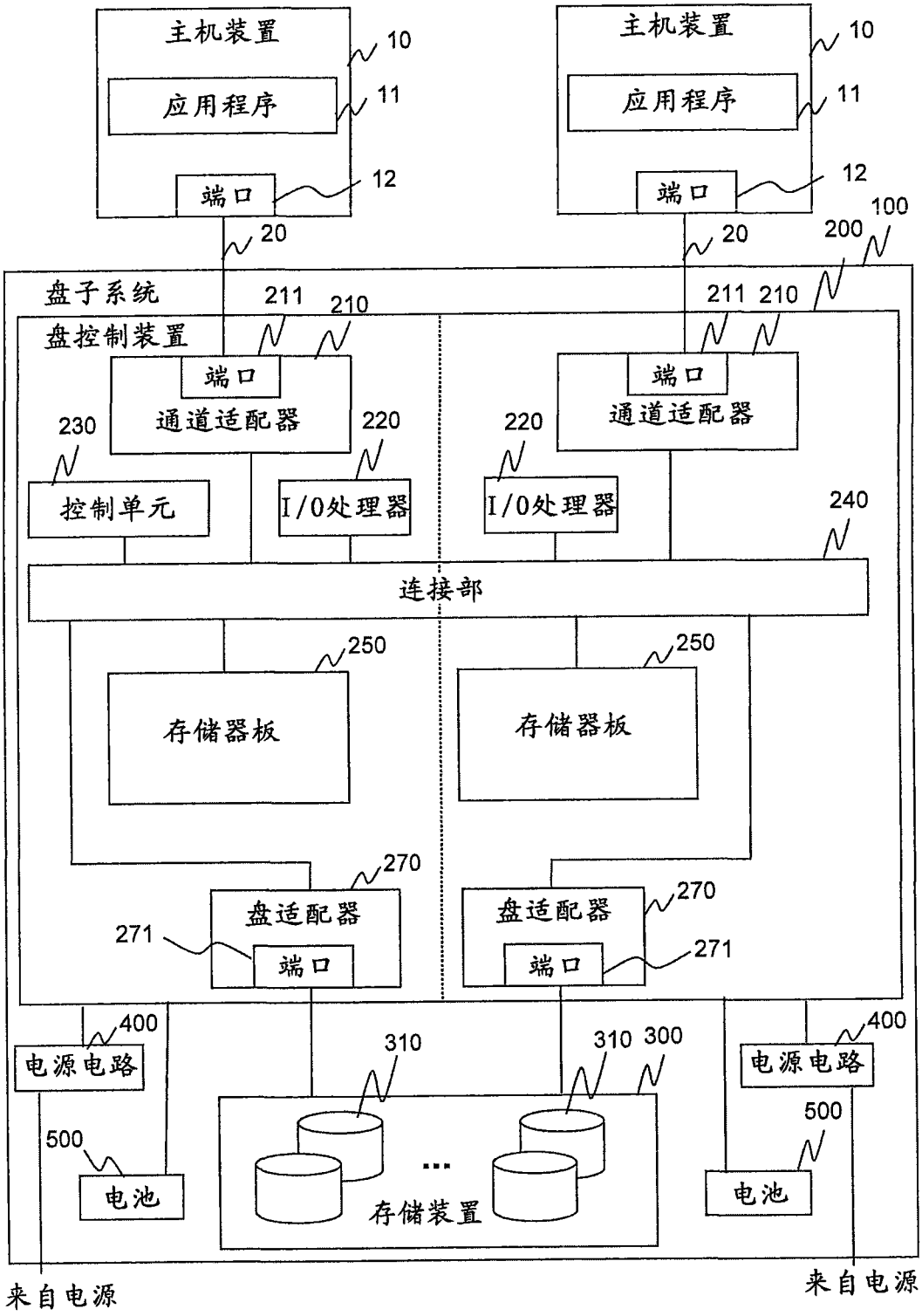

主机装置
10
应用程序
11
端口
12
主机装置
10
应用程序
11
端口
12
20
20
100
200
盘子系统
盘控制装置
211
210
211
210
端口
通道适配器
端口
通道适配器
230
220
220
控制单元
I/O处理器
I/O处理器
240
连接部
250
250
存储器板
存储器板
270
270
盘适配器
271
端口
盘适配器
端口
271
400
400
电源电路
310
310
300
电源电路
500
500
电池
...
存储装置
电池
来自电源
来自电源

图 1

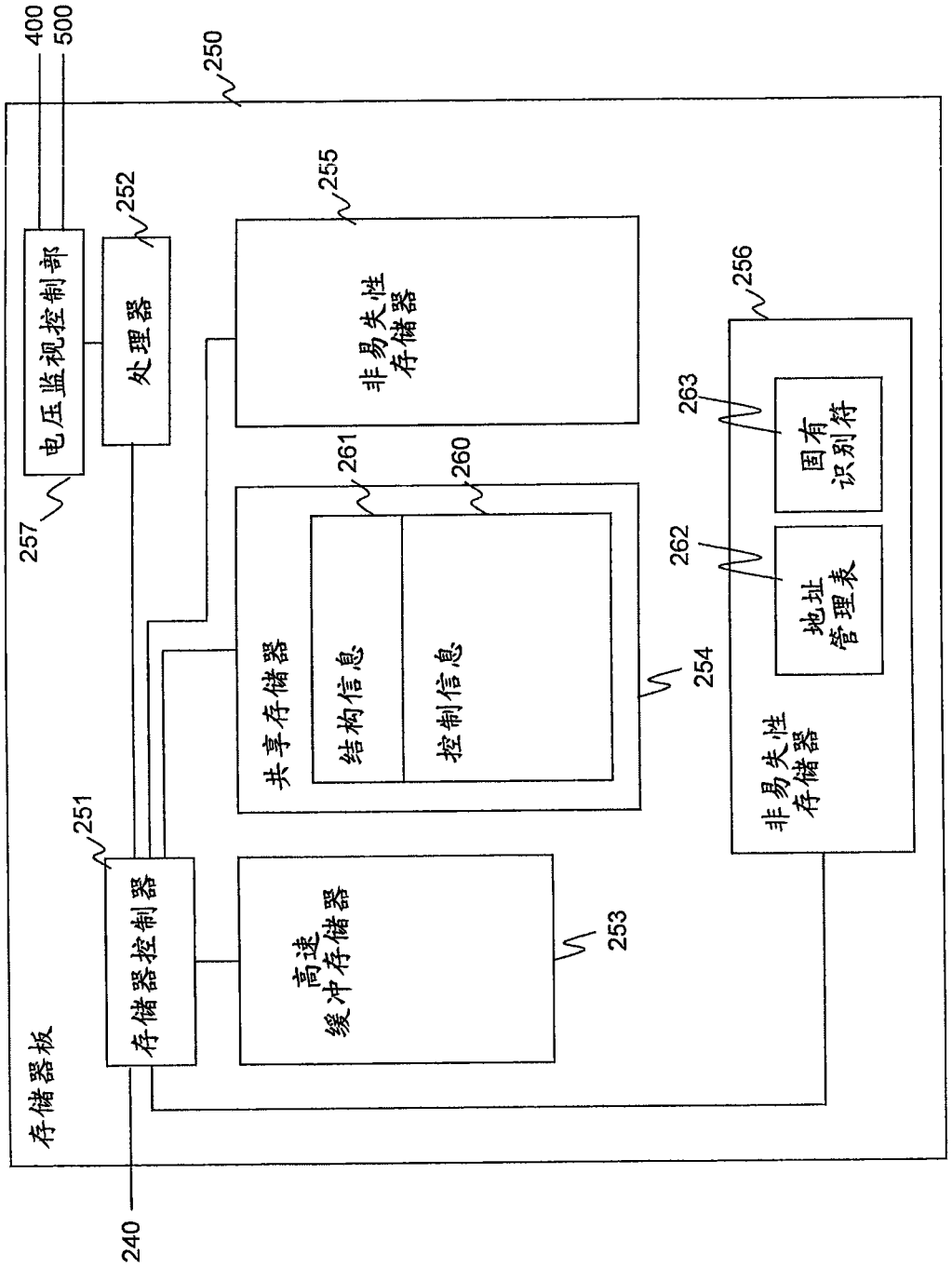

400
500
250
电压监视控制部
257
处理器
252
非易失性存储器
255
256
263
固有识别符
261
260
262
地址管理表
共享存储器
结构信息
控制信息
254
非易失性存储器
251
存储器控制器
高速缓冲存储器
253
存储器板
240

图 2

控制信息

2601 逻辑地址	2602 高速缓冲存储器地址	2603 升级bit	2604 脏bit
逻辑地址la1	高速缓冲存储器地址ca1	升级bitsb1	脏bitdb1
逻辑地址la2	高速缓冲存储器地址ca2	升级bitsb2	脏bitdb2
...	...	...	...
逻辑地址laXX	高速缓冲存储器地址caXX	登台bitsbXX	脏bitdbXX

260

图 3A

结构信息

2611 逻辑地址	2612 物理地址
逻辑地址la1	物理地址pa1
逻辑地址la2	物理地址pa2
...	...
逻辑地址laXX	物理地址paXX

261

图 3B

地址管理表（控制信息）

非易失性存储器地址 (2621)	共享存储器地址（控制信息）(2622)	数据长度 (2623)
非易失性存储器地址na1	共享存储器地址（控制信息）sac1	数据长度d11
非易失性存储器地址na2	共享存储器地址（控制信息）sac2	数据长度d12
非易失性存储器地址naXX	共享存储器地址（控制信息）sacXX	数据长度d1XX

262A

图 4A

地址管理表（结构信息）

非易失性存储器地址 (2624)	共享存储器地址（结构信息）(2625)	数据长度 (2626)
非易失性存储器地址na1	共享存储器地址（结构信息）sac1	数据长度d11
非易失性存储器地址na2	共享存储器地址（结构信息）sac2	数据长度d12
非易失性存储器地址naYY	共享存储器地址（结构信息）sacYY	数据长度d1YY

262B

图 4B

地址管理表（高速缓冲存储数据）

非易失性存储器地址 (2627)	高速缓冲存储器地址 (2628)	数据长度 (2629)
非易失性存储器地址na21	高速缓冲存储器地址caa1	数据长度d121
非易失性存储器地址na22	高速缓冲存储器地址caa2	数据长度d122
非易失性存储器地址naZZ	高速缓冲存储器地址caaZZ	数据长度d1ZZ

262C

图 4C

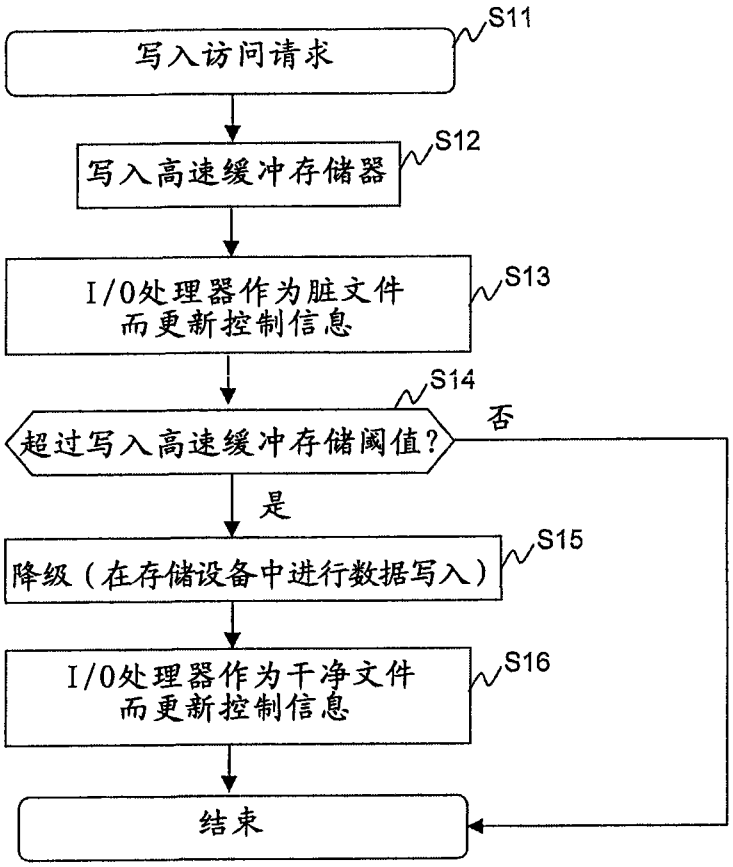

写入访问请求
S11
写入高速缓冲存储器
S12
I/O处理器作为脏文件
而更新控制信息
S13
超过写入高速缓冲存储阈值?
S14
否
是
降级（在存储设备中进行数据写入）
S15
I/O处理器作为干净文件
而更新控制信息
S16
结束

图 5A

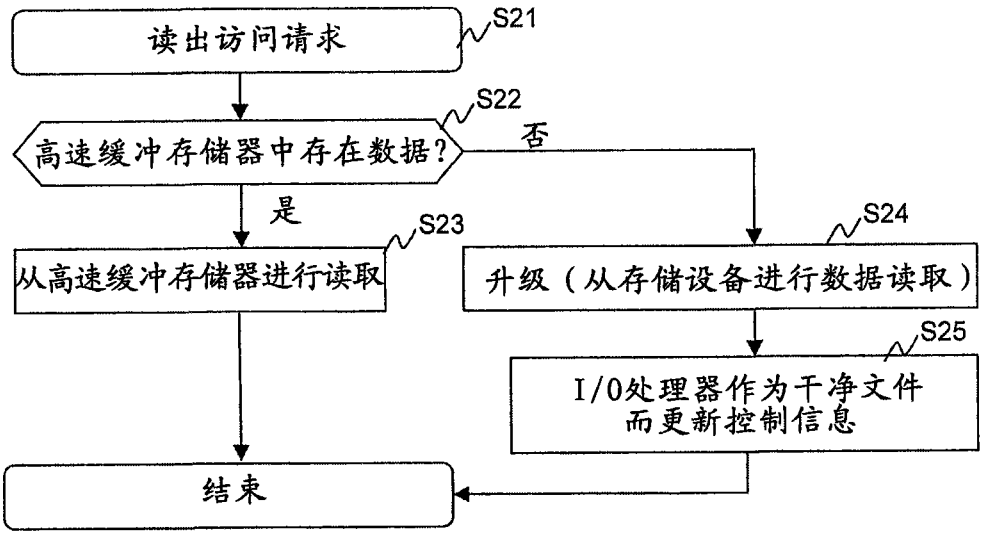

读出访问请求
S21
高速缓冲存储器中存在数据?
S22
否
是
从高速缓冲存储器进行读取
S23
升级（从存储设备进行数据读取）
S24
I/O处理器作为干净文件
而更新控制信息
S25
结束

图 5B

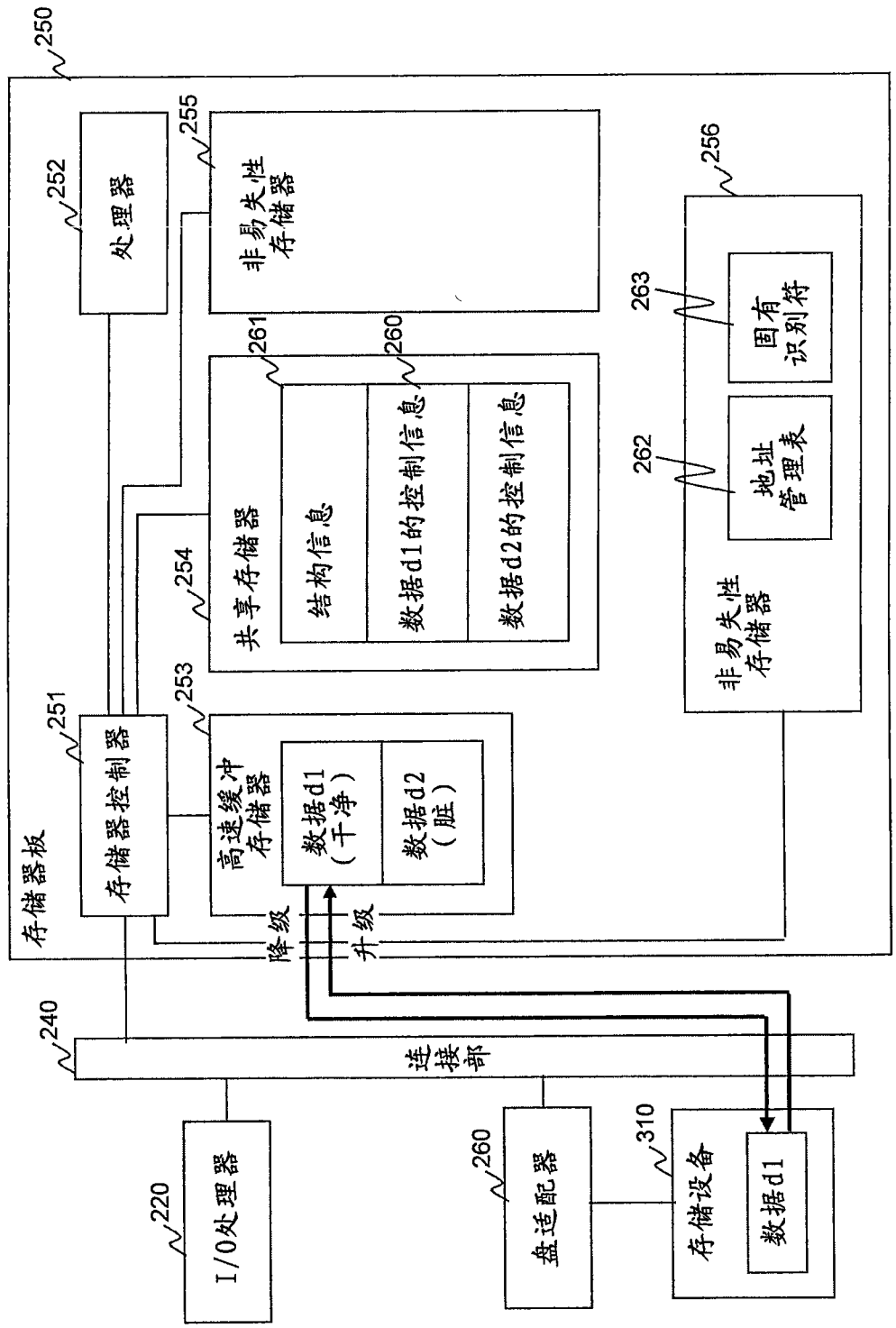

250
存储器板
252
处理器
255
非易失性存储器
256
263
固有识别符
261
260
结构信息
数据d1的控制信息
数据d2的控制信息
262
地址管理表
254
共享存储器
非易失性存储器
253
高速缓冲存储器
数据d1（干净）
数据d2（脏）
251
存储器控制器
降级
升级
240
连接部
220
I/O处理器
260
盘适配器
310
存储设备
数据d1

图 6

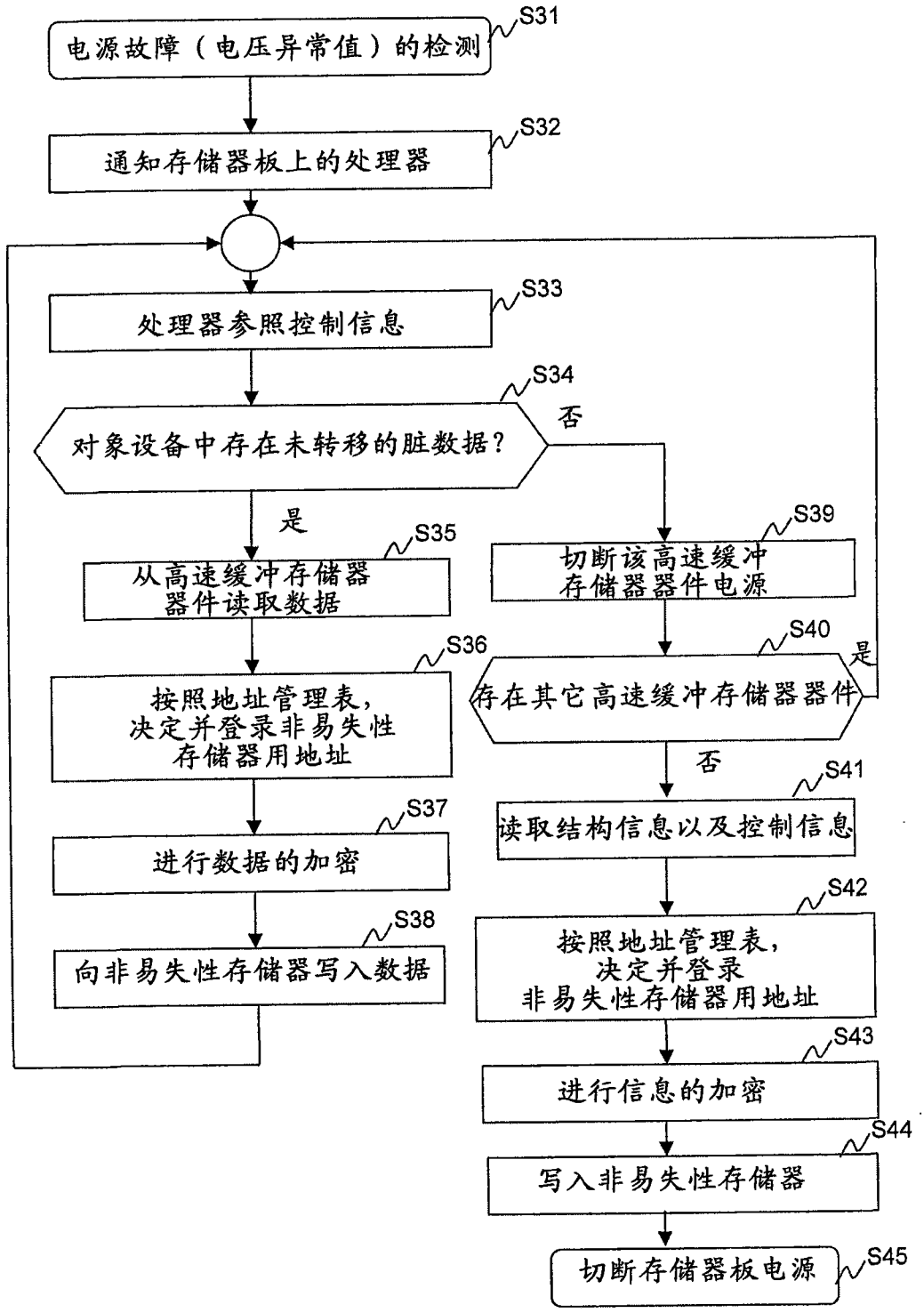

电源故障（电压异常值）的检测
S31
通知存储器板上的处理器
S32
处理器参照控制信息
S33
对象设备中存在未转移的脏数据？
S34
否
是
从高速缓冲存储器
器件读取数据
S35
按照地址管理表，
决定并登录非易失性
存储器用地址
S36
进行数据的加密
S37
向非易失性存储器写入数据
S38
切断该高速缓冲
存储器器件电源
S39
存在其它高速缓冲存储器器件
S40
是
否
读取结构信息以及控制信息
S41
按照地址管理表，
决定并登录
非易失性存储器用地址
S42
进行信息的加密
S43
写入非易失性存储器
S44
切断存储器板电源
S45

图 7

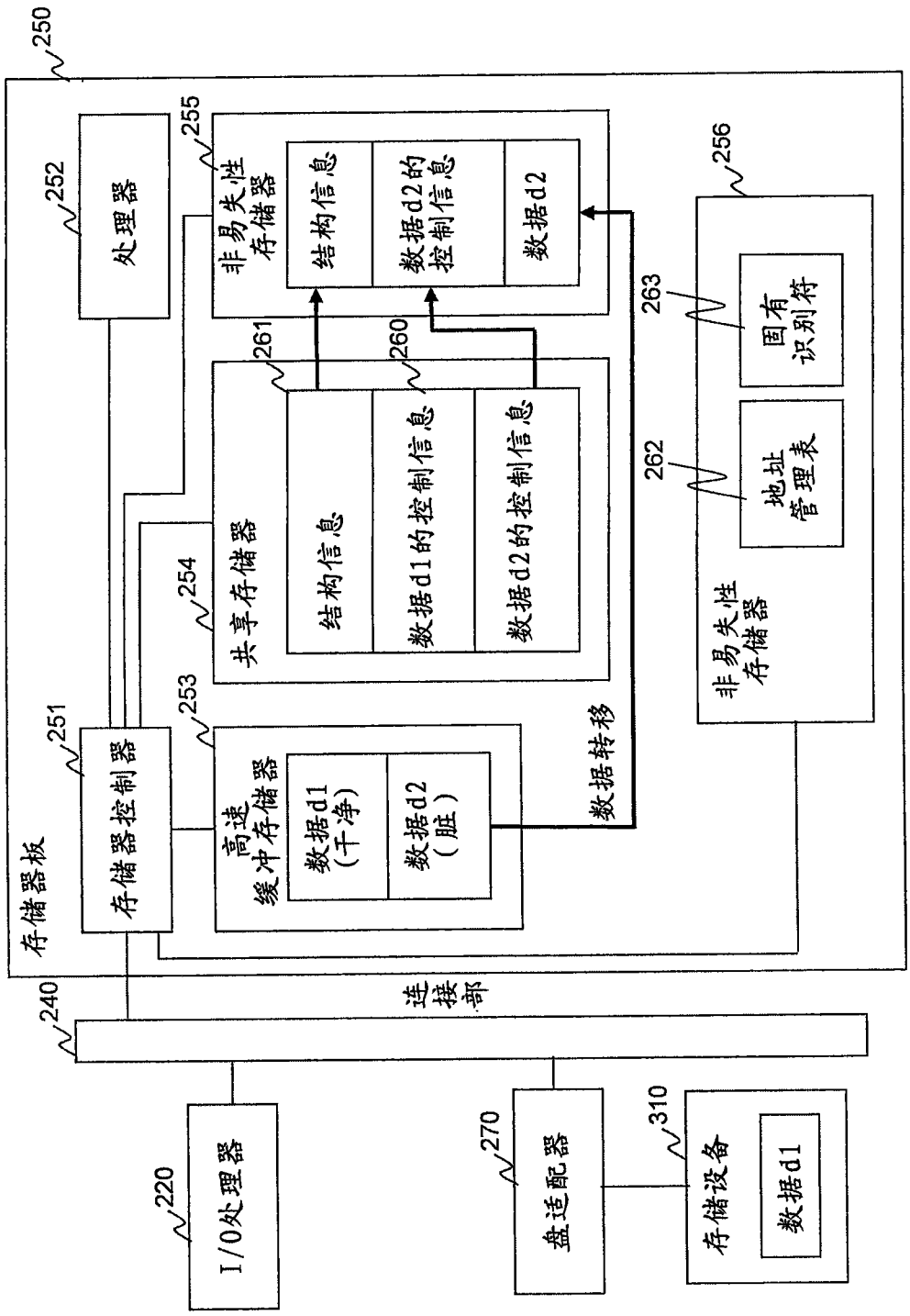

250
存储器板
252
处理器
255
非易失性存储器
结构信息
数据d2的控制信息
数据d2
256
263
固有识别符
261
260
262
地址管理表
共享存储器
结构信息
数据d1的控制信息
数据d2的控制信息
254
非易失性存储器
253
251
存储器控制器
高速缓冲存储器
数据d1（干净）
数据d2（脏）
数据转移
240
连接部
220
I/O处理器
270
盘适配器
310
存储设备
数据d1

图 8

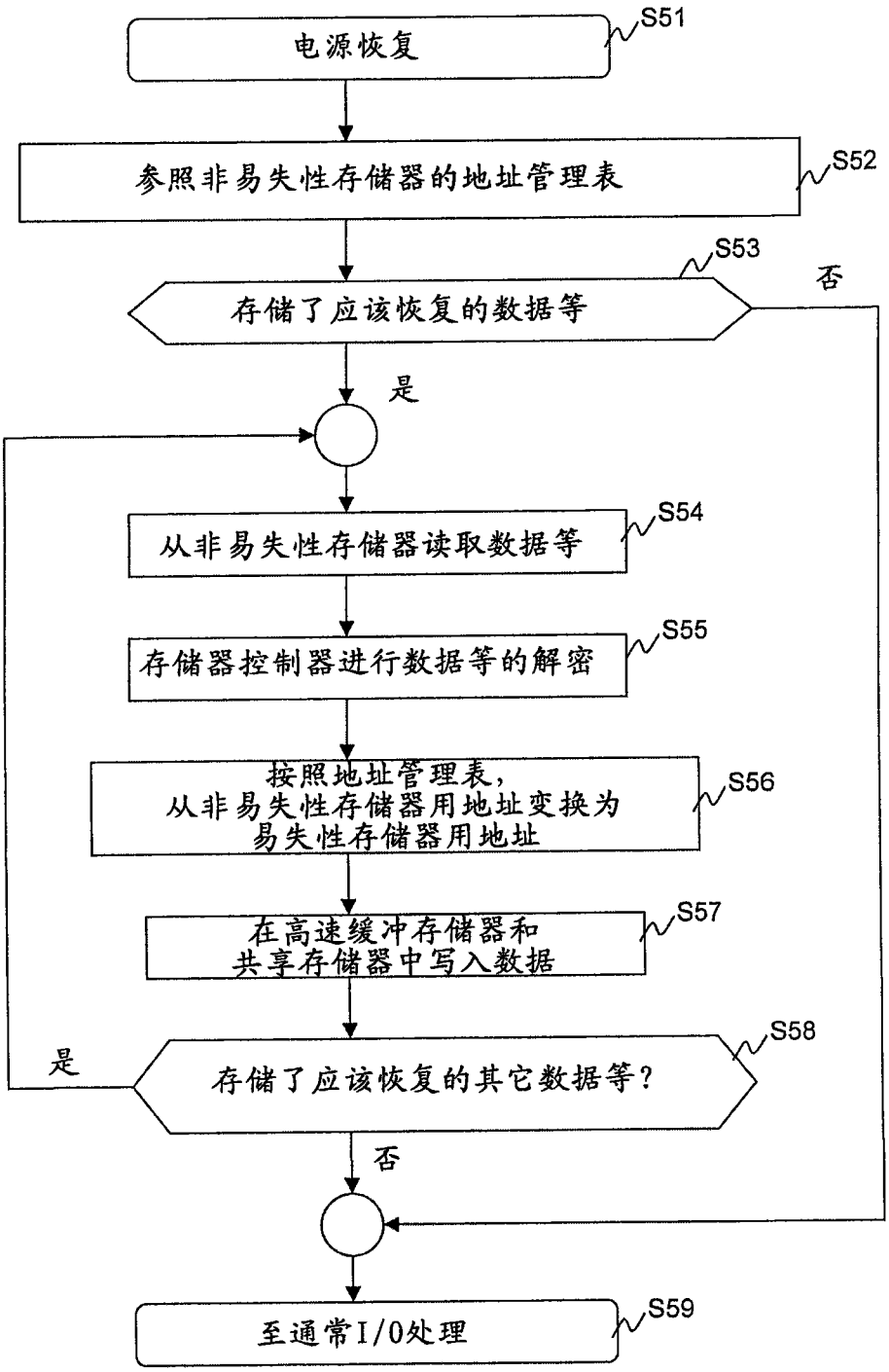

电源恢复
S51
参照非易失性存储器的地址管理表
S52
存储了应该恢复的数据等
S53
否
是
从非易失性存储器读取数据等
S54
存储器控制器进行数据等的解密
S55
按照地址管理表，
从非易失性存储器用地址变换为
易失性存储器用地址
S56
在高速缓冲存储器和
共享存储器中写入数据
S57
存储了应该恢复的其它数据等？
S58
是
否
至通常I/O处理
S59

图 9

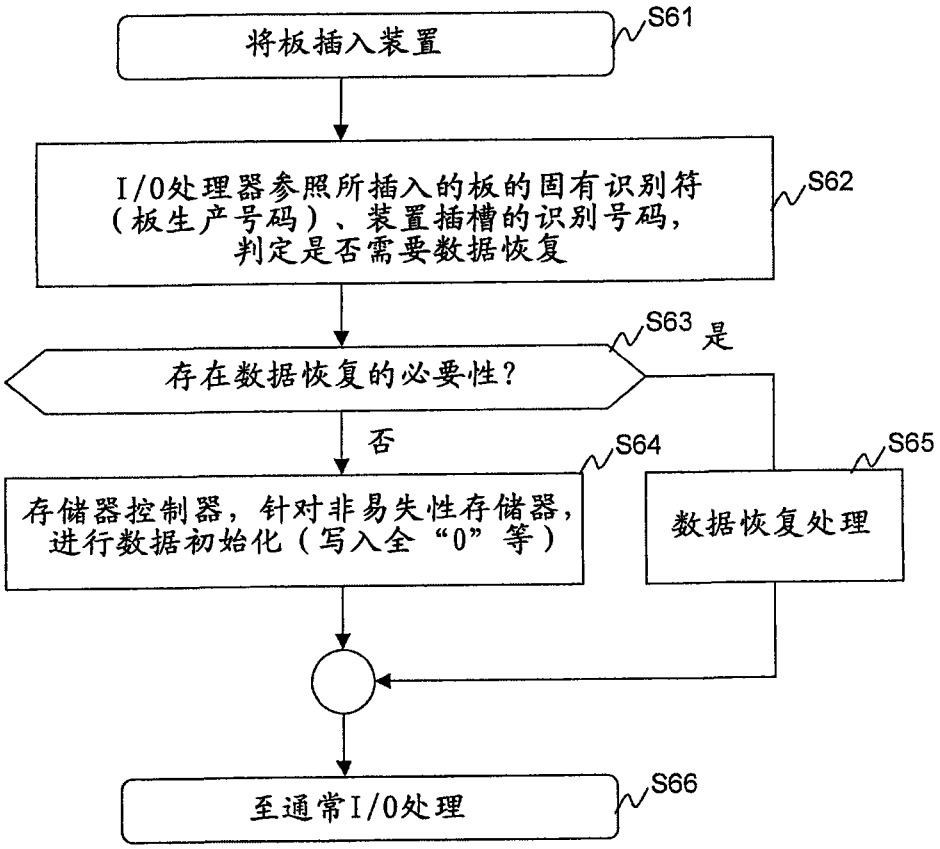

将板插入装置
S61
I/O处理器参照所插入的板的固有识别符（板生产号码）、装置插槽的识别号码，判定是否需要数据恢复
S62
存在数据恢复的必要性？
S63
是
否
存储器控制器，针对非易失性存储器，进行数据初始化（写入全“0”等）
S64
数据恢复处理
S65
至通常I/O处理
S66

图 10

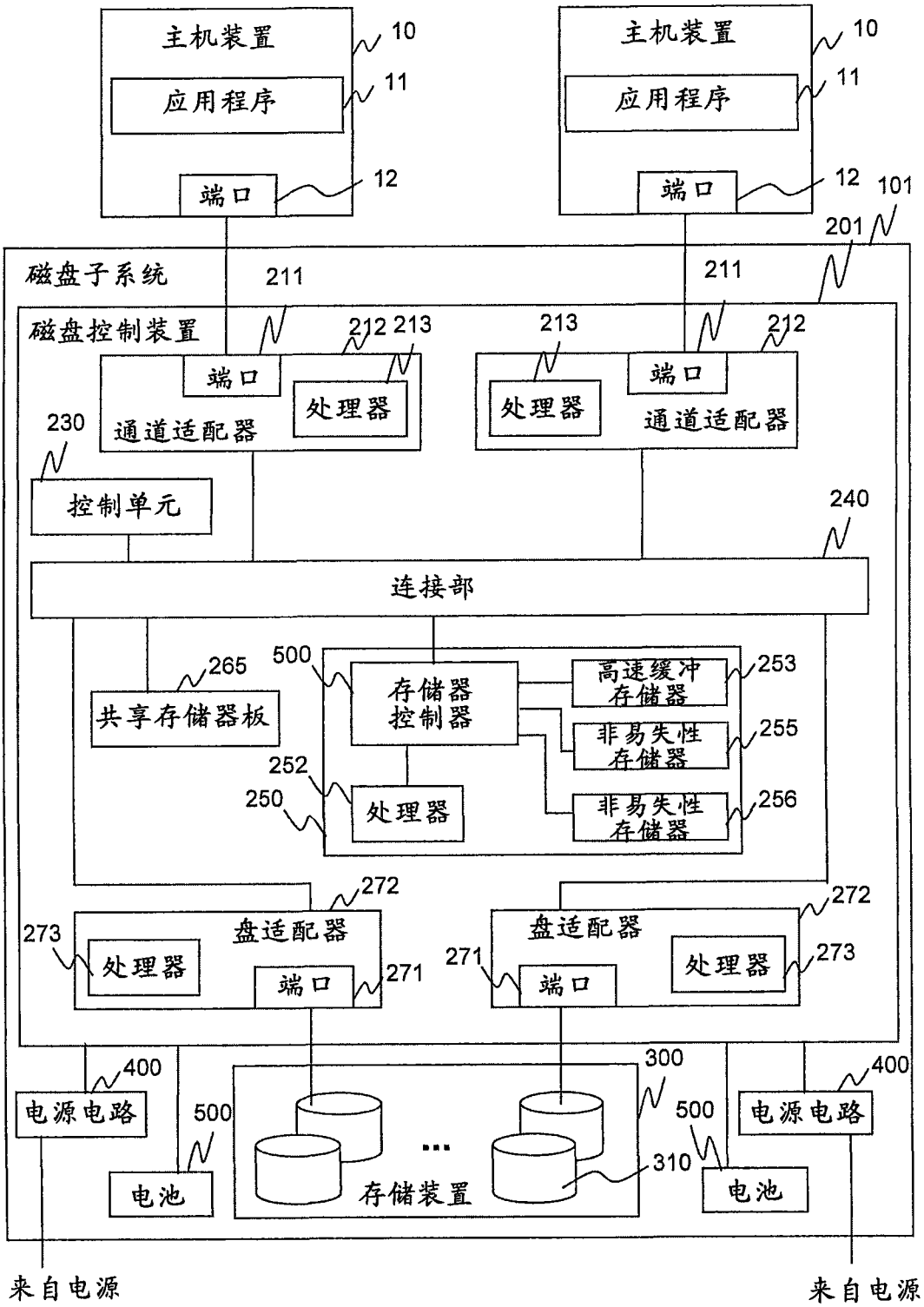

主机装置
10
应用程序
11
端口
12
主机装置
10
应用程序
11
端口
12
101
201
磁盘子系统
211
211
磁盘控制装置
212
213
213
212
端口
处理器
处理器
端口
通道适配器
通道适配器
230
控制单元
240
连接部
265
共享存储器板
500
存储器
控制器
高速缓冲
存储器
253
非易失性
存储器
255
252
处理器
250
非易失性
存储器
256
272
盘适配器
272
盘适配器
273
处理器
端口
271
271
端口
处理器
273
400
电源电路
500
300
500
400
电源电路
电池
存储装置
310
电池
来自电源
来自电源

图 11

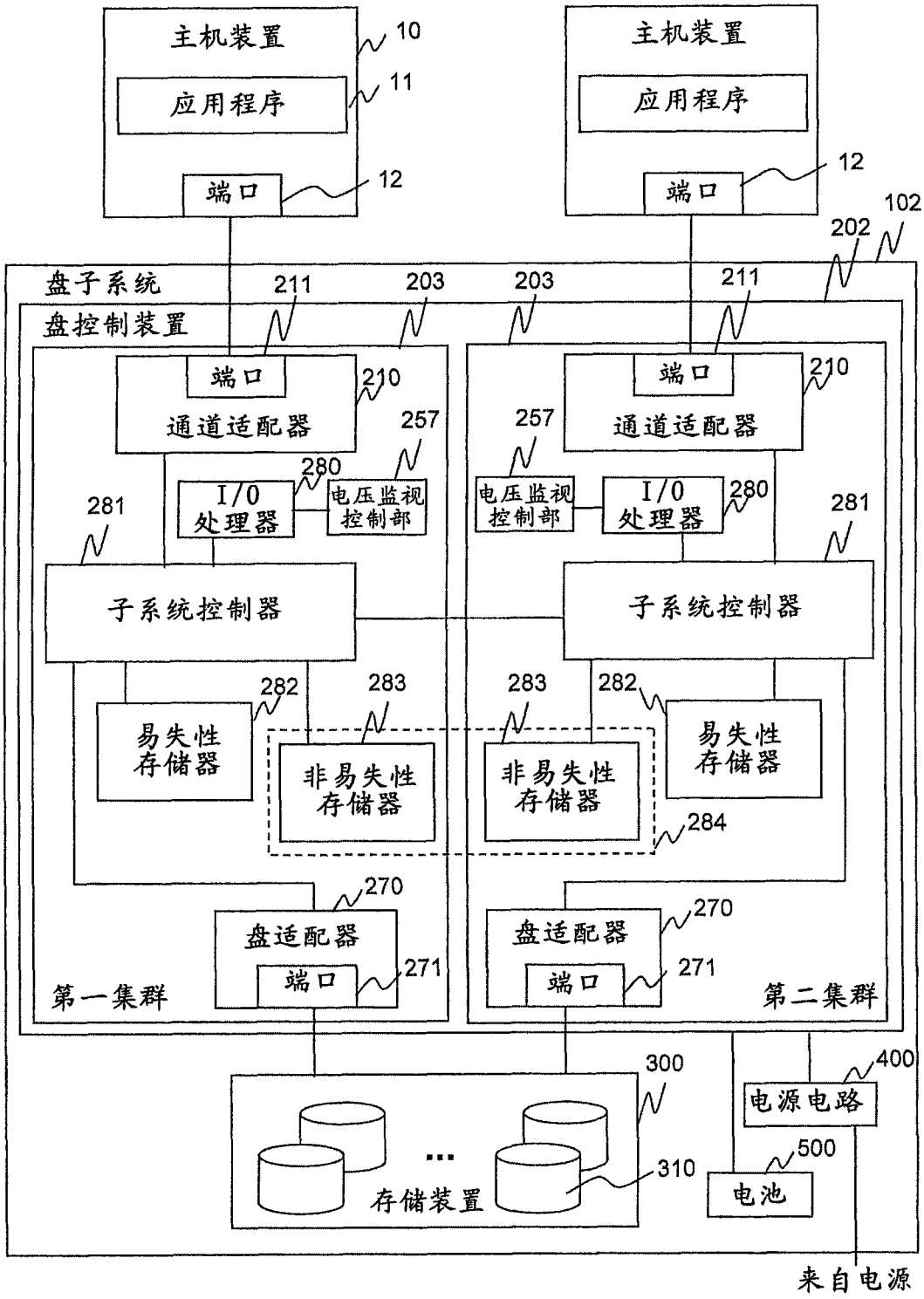

主机装置
10
应用程序
11
端口
12
主机装置
应用程序
端口
12
102
202
盘子系统
211
203
203
211
盘控制装置
端口
210
通道适配器
257
280
I/O
处理器
电压监视
控制部
281
子系统控制器
282
283
易失性
存储器
非易失性
存储器
284
270
盘适配器
端口
271
第一集群
端口
210
通道适配器
257
电压监视
控制部
I/O
处理器
280
281
子系统控制器
283
282
易失性
存储器
非易失性
存储器
270
盘适配器
端口
271
第二集群
300
存储装置
310
400
电源电路
500
电池
来自电源

图 12

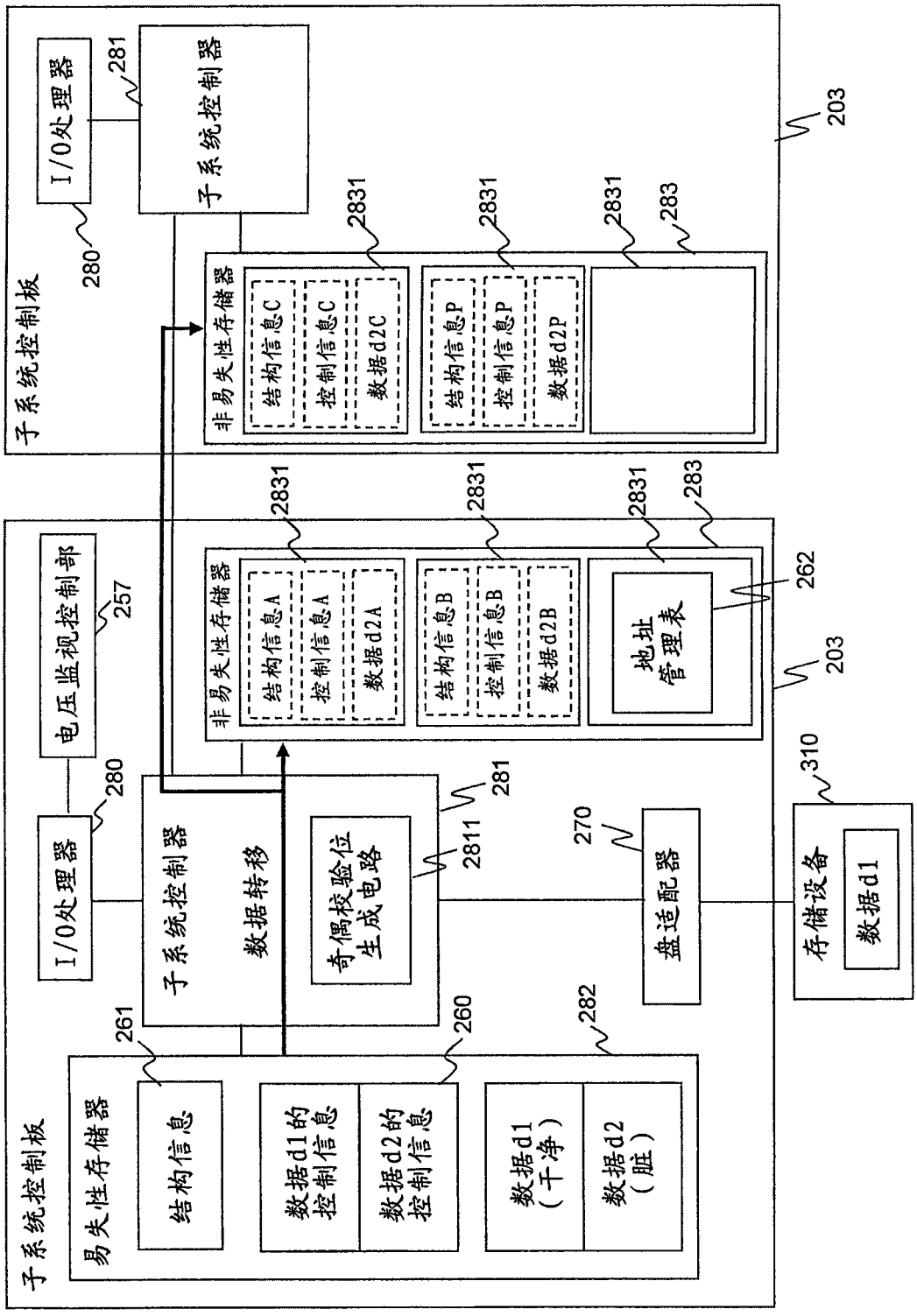

子系统控制板
I/O处理器
280
子系统控制器
281
203
非易失性存储器
结构信息C
控制信息C
数据d2C
2831
结构信息P
控制信息P
数据d2P
2831
2831
283
电压监视控制部
257
结构信息A
控制信息A
数据d2A
2831
结构信息B
控制信息B
数据d2B
2831
地址管理表
283
262
203
I/O处理器
280
子系统控制器
数据转移
奇偶校验位生成电路
2811
281
270
盘适配器
310
存储设备
数据d1
易失性存储器
261
结构信息
数据d1的控制信息
数据d2的控制信息
260
数据d1（干净）
数据d2（脏）
282

图 13

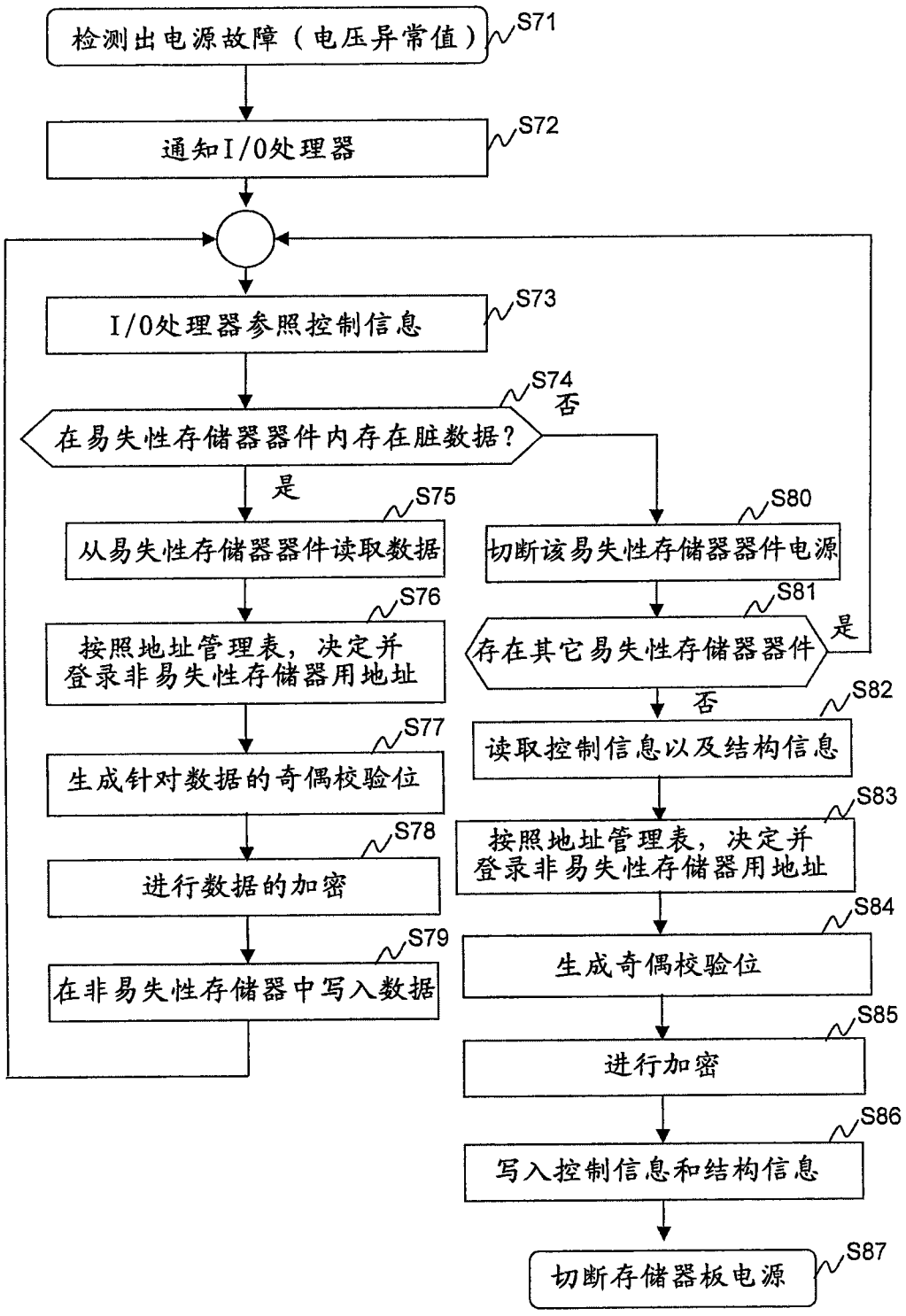

检测出电源故障（电压异常值）
S71
通知I/O处理器
S72
I/O处理器参照控制信息
S73
在易失性存储器器件内存在脏数据？
S74
否
是
从易失性存储器器件读取数据
S75
按照地址管理表，决定并登录非易失性存储器用地址
S76
生成针对数据的奇偶校验位
S77
进行数据的加密
S78
在非易失性存储器中写入数据
S79
切断该易失性存储器器件电源
S80
存在其它易失性存储器器件
S81
是
否
读取控制信息以及结构信息
S82
按照地址管理表，决定并登录非易失性存储器用地址
S83
生成奇偶校验位
S84
进行加密
S85
写入控制信息和结构信息
S86
切断存储器板电源
S87

图 14

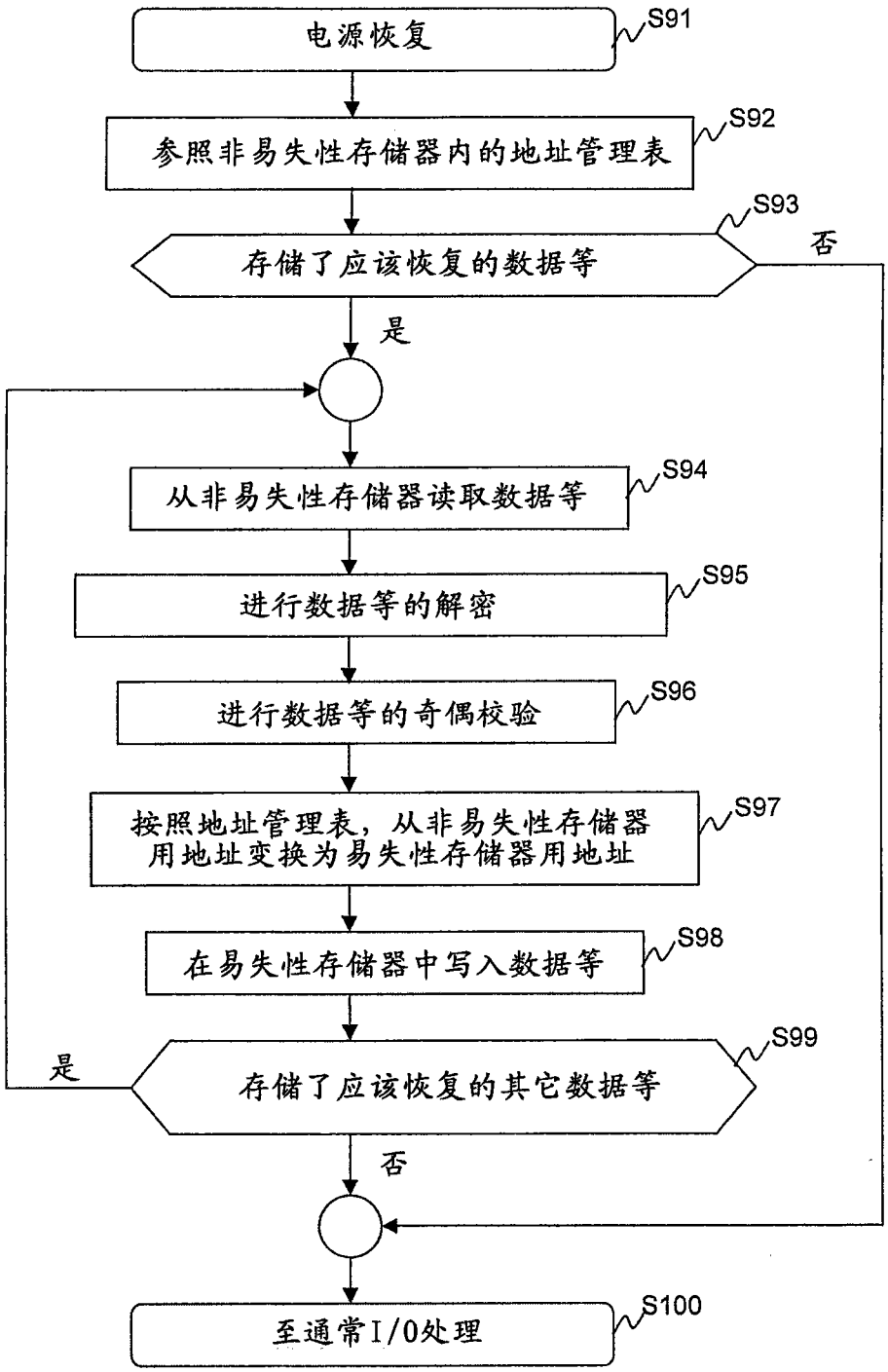

电源恢复
S91
参照非易失性存储器内的地址管理表
S92
存储了应该恢复的数据等
S93
否
是
从非易失性存储器读取数据等
S94
进行数据等的解密
S95
进行数据等的奇偶校验
S96
按照地址管理表，从非易失性存储器用地址变换为易失性存储器用地址
S97
在易失性存储器中写入数据等
S98
存储了应该恢复的其它数据等
S99
是
否
至通常I/O处理
S100

图 15