[19]中华人民共和国专利局

[11] 公开号 CN 1081005A

[12] 发明专利申请公开说明书

[21]申请号 93107031.7

[51]Int.Cl[5]

[43]公开日 1994 年 1 月 19 日

G06F 3/06

[22]申请日 93.5.10

[30]优先权

[32]92.5.12 [33]JP[31]119226/92

[71]申请人 国际商业机器公司

地址 美国纽约

[72]发明人 岩佐博之 浅野秀夫 清水丰

[74]专利代理机构 中国国际贸易促进委员会专利代理部

代理人 姜 华

说明书页数: 附图页数:

[54]发明名称 用于构成冗余磁盘存贮系统的适配器

[57]摘要

连接到主机和磁盘存贮装置之间的一适配器提供用于连接主机和磁盘存贮装置的接口，该磁盘存贮装置具有相同的接口设计。该适配器自身包括控制装置，用于建立一冗余磁盘存贮系统。此外，该适配器自身包括用于检测和指示一失效磁盘存贮装置的装置，用替换该失效磁盘存贮装置的装置和在磁盘存贮装置替换之后重建一冗余磁盘存贮系统的装置。一命令使该主机能够存取磁盘存贮装置的每一个以用于维修的目的。

权 利 要 求 书

1、连接于一主计算机和磁盘存贮装置之间构成一冗余磁盘存贮系统的适配器，其特征是：

用于连接该主计算机的一主接口；

用于分别连接到多个初级磁盘存贮装置和多个次级磁盘存贮装置的并且每一个都具有如上述主接口相同的接口设计的一初级装置接口和一次级装置接口，所述的磁盘存贮装置中的每一个在所述初级和次级装置接口中的每一个中由一异常鉴别码鉴别；

基于来自所述装置的信息用于检测所述初级和次级磁盘存贮装置中的每一个的故障的装置；和

用于对在所述初级和次级磁盘存贮装置接口中的每一个都具有相同鉴别码的一对所述 初级 和次级磁盘存贮装置的控制装置，在正常操作中，所述初级和次级 磁盘存贮装置的所述对从所述主计算机写信息到所述初级和次级磁盘存贮装置时，和从所述对中的所述初级磁盘存贮装置读出信息并送该信息到所述主计算机，和，当所述的初级和次级磁盘存贮装置对中的一个产生故障时，仅仅或者从其它正常在磁盘存贮装置中读或写信息。

2、根据权利要求 1 的该适配器，其特征是：

所述装置还包括用于当检测故障的装置在所述磁盘存贮装置对中 的一个中检测到一故障时，指示该发生故障的磁盘存贮装置的装置，独立于主计算机并意味着独立于所述主计算机的运行并能够替换该失效的磁盘存贮装置及从所述磁盘存贮装置

对中利用该另一个正常的磁盘存储装置重建一新磁盘存贮装置进入该冗余磁盘存贮系统。

3、根据权利要求 1 的该适配器，其特征是：

所述装置包括用于利用来自所述主计算机的一命令使得该主计算机对所 述磁盘存贮装置对中的每一个进行存取并给出一异常逻辑单元数 到所述该磁盘存贮装置对中的每一个的装置。

4、根据权利要求 1 的该适配器，其特征是：

所述装置还包括用于从具有正常数据的所述磁盘存贮装置对中的一个拷贝到不具有正常数据的所述磁盘存贮装置对中的另一个。

5、根据权利要求 1 的该适配器，其特征是：

用于检测故障的所述装置包括用于在所述主计算机到与信息一致的监视器的操作期间从所述磁盘存贮装置对比较相互信息的装置。

6、根据权利要求 1 的该适配器，其特征是：

所述主接口，所 述初级装置接口和所述次级装置接口是 SCSI 接口。

说　明　书

用于构成冗余磁盘存贮系统的适配器

本发明涉及将磁盘存贮装置连接到主计算机的一种适配器，特别涉及一种构成一冗余和备用磁盘存贮系统并能通过每一个都具有相同接口设计的若干接口连接到一主计算机和每一个磁盘存贮器装置的适配器。

用于对计算机系统存储大规模数据和程序的磁盘存储装置是一必不可少的装置，在磁盘存储装置失效的情况下，数据式程序会被阻止读出或存入磁盘存贮器装置，并且使用磁盘存贮装置的整个计算机系统会被中止，在计算机系统中，磁盘存贮装置包括比较容易失效的可移动机械部件，为予防失效，通常熟知的方法是冗余地构成双磁盘存贮装置，即具有两个磁盘存贮装置，其中的每个都贮相同的数据和程序，并在一个磁盘存贮装置失效的事故中，另一个正常的磁盘存贮装置代替该失效装置，这样冗余或备用系统对于需要高于可靠性的操作，例如在银行，保险公司，等等的操作是必不可少的。

例如，日本公开的未审查的专利申请（PUPAS），申请号为61—240320，61—249132 和 62—139172 披露了具有双结构的冗余或备用磁盘控制器，上述 PUPAS 所披露的磁盘控制器连接于一大中央处理单元和磁盘存贮装置之间，特别设计成在两个磁盘存贮装置中的每一个都存贮有相同的数据以予防故障，在一大系统中，无论如何，该中央处理单元通过磁盘控制器连接

于该磁盘存贮装置，在这种情况中，用于连接该磁盘控制器到该中 央处理单元的一主接口与用于连接该磁盘控制器到该磁盘存贮装置的一装置接口是不同的，那是由于称之为通道的该主接口能在高速率下传输数据，而并不限定在一确定的外围 单元；换言之，因为该磁盘控制器控制一确定的磁盘存贮装置，对该磁盘存贮装置，在它们之间的该装置接口被独特地限定，那就是这些在先技术并未提出在一主体和冗余磁盘存贮装置之间的适配器的结构，在这种结构中，主体和装置接口没有相同的接口设计。

在新信息处理设备中的降处理过程，由于它们对高可靠性应用的需要，利用冗余结构对用于小计算机（个人计算机，工作台，等等）的磁盘存贮装置中的故障进行预防，在这种情况，一系统利用用于小计算机的标准接口而不用改变主计算机和磁盘存贮装置这两部分就能建立起对故障的预防，这是便利的，由于这一理由，该系统采用更容易购置的标准计算机和一磁盘存贮装置就能构成对故障的预防，进而，如果冗余结构对一主计算机是透明的，那么在该主计算机上的一操作系统或应用程序就能运用该冗余或备用磁盘存贮装置预防在一单个磁盘存贮装置中相同的故障，就使迄今使用一操作系统或应用程序能进行高可靠的信息处理，这样的接口是，例如 SCSI（小计算机系统接口，美国国家标准协会（ANSI）ANSI X3.131—1986）。

连接磁盘存贮装置到一主计算机的用于建立相应本发明的一冗余磁盘存贮系统的一适配器，综连接在该主计算机和一对利用相同接口，例如 SCSI 的该磁盘存贮装置之间，该磁盘存贮装置对中的每一个在写操作时写入相同数据，并且，在读操作时从该磁盘存贮装置的一个读出数据，进而，该适配器包括控

制装置，用于当该磁盘存贮装置对中的一个发生故障时，从正常的一个中分离出有故障的一个，并对该磁盘存贮装置对中正常的另一个完成读或写操作。

详细说明用于连接该磁盘存贮器装置到该主计算机的相应于本发明的该适配器包括用于连接该主计算机的一主接口，一初级装置接口，和次级装置接口，每一个都具有如同该主接口的相同的接口设计，分别连接到该初级和次级装置接口是初级磁盘存贮装置（称有效存贮装置）和次级磁盘存贮装置（称备用存贮装置）对。

该主计算机对该磁盘存贮装置的一写操作期间，在该初级和次级磁盘存贮装置这二者被存入相同数据，因而，该初级和次级磁盘存贮装置中的每一个总是保持有相同数据，由该主计算机对该磁盘存贮器装置的一读操作期间，从该初级磁盘存贮装置读出的数据被传输到该主计算机，如果该初级和次级磁盘存贮装置中的任何一个发生故障，该读写操作由一个正常磁盘存贮装置完成，然后，当取代已失效磁盘存贮装置或磁盘存贮装置中的一个丢失了部分数据时，来自该正常磁盘存贮装置的读出数据被拷贝到一新磁盘存贮装置或丢失数据的该磁盘存贮装置。

不用该主计算机有任何介入或中断数据传 输就完成了以上操作，即该操作对主计算机是透明的，该主计算机完成该读或写操作仅是对一单个磁盘存贮装置，相应于本发明的该适配器自动进行冗余或转换一失效装置到一正常装置的一写操作，从而在该主计算机这一方面不需要任何软件或硬件的改变，和在该磁 盘存贮装置方面，当一接口被连在其中时，不需要任何改变就能构成预防故障的一冗余或备用系统，进而，相应于本

发明的用于冗余或备用磁盘存贮系统的该适配器包括独立于该主计算机的用于检测和指示在该磁盘存贮装置中的故障的装置，和用于将一个新的替换一失效的磁盘存贮装置并利用磁盘存贮装置对中的一正常的磁盘存贮装置重新建立备用系统而又不停止主机运行的装置，相应于本发明的用于冗余或备用磁盘存贮系统的该适配器进一步还包括利用来自主计算机的一命令使得该主计算机具有能分别存取初级和次级磁盘存贮装置对中的每一个的能力的装置。

参照附图在下面将描述本发明的一实施例。

图 1 是利用相应于本发明的一实施例的适配器所建立的具有冗余或备用磁盘存贮装置的一整个计算机系统的方框图，101 表示相应于该实施例的该适配器，它用于连接该磁盘存贮装置到一主计算机；该适配器 101 通过一主 SCSI 总线 102 与一主 SCSI 适配器 103 相连通，该主 SCSI 适配器 103 被连接到在主计算机 104 内部的一未画出的系统总线，在该主 SCSI 总线 102 上，该主 SCSI 适配器 103 和该适配器 101 被分别赋预一最大 ID（识别码）=7 和 ID=n（其中 n 是 0 至 6 的整数），在该 SCSI 总线 102 上，最大到 6 个 SCSI 装置具有正是由 n 所能赋给的 SCSI 接口 ID0 至 6。

图 2 是示明该适配器 101 的方框图，该适配器 101 具有连接到该主 SCSI 总线 102 的一主 SCSI 接口 201，该适配器 101 包括一初级装置 SCSI 接口 202，它连接到一初级 SCSI 总线 105 和一次级装置 SCSI 接口 203，它联到次级 SCSI 总线 106，每一个都具有相同该主 SCSI 接口 201 的相同接口，该初级装置 SCSI 接口 202 和该次级装置 SCSI 接口 203 仅在一初使模式送出一命令时运行，而该主 SCSI 接口 201 仅在一目标模式中接收一个

命令时运行。

在图 1 中示明的该初级和次级 SCSI 总线 105 和 106 能连接到直接分别具有相同的 SCSI 接口设计的 7SCSI 磁盘存贮装置（DASDs）107 和 108，被称为初级磁盘存贮装置的该磁盘存贮装置（DASDs）107 连接到该初级 SCSI 总线 105，而被称为次级磁盘存贮装置的磁盘存贮装置（DASDs）108 连接到该次级 SCSI 总线 106。

在该 SCSI 接口具有从 0 至 6 的 ID_s 的该初级和次级 SCSI 总线 105 和 106 及该初级和次级 DASDs107 和 108 上，该适配器 101 在该 SCSI 接口中具有最大的 ID（=7），在该相同的 SCSI 总线，装置不能有相同的 ID，分别连接到该初级和次级 SCSI 总线 105 和 106 的该初级和次级 DASDs107 和 108 并具有彼此相同的 ID 对并总保持相同的数据，以便能被此互补，通过下述描述会更加明显，该 $DASD_s$107 和 108 中的任何一个可以是磁盘存贮装置或光盘存贮装置。

如图 2 所示，连接到该主 SCSI 总线 102 的该适配器 101 的主 SCSI 接口 201 被连接到一初级缓冲存贮器 205 和一次级缓冲存贮器 206，该缓冲存贮器 205 和 206 每个都具有 32k 存贮容量，该初级和次级缓冲存贮器 205 和 206 分别被连接到该初级和次级装置 SCSI 接口 202 和 203，在图 2 中，在适配器 101 中的数据通道 208 由粗线表明，该适配器 101 包括用于表明操作装置的面板 109，该面板 109 通过一控制逻辑 204，一微处理器 MPU209、一只读存贮器 ROM201、一可重写只读存贮器 EEPROM211，一易失随机存取存贮器 RAM212 和一控制逻辑 207 连接到该 MPU209，这些是通过一局部总线 214 相连接，存贮在该非易失 ROM210 和 EEPROM211 的是由该适配器 101 用于完

成控制功能所需要的程序和参数。

图 3 表明在该实施例中的该适配器 101 中的数据和控制信息流向一方框图，310A，B…E 是表示双向门由来该控制逻辑 204 和该微处理器 209 的控制信息控制数据流向，该门 301A 被用于控制数据在该初级 DASDs107 和该次级 DASDs108 对之间传输；该门 301B 和 C 被用于分别控制数据在该主 SCSI 接口 201 和该初级缓冲存贮器 205 之间传输；该门 301D 和 310E 被用于分别控制数据在该主 SCSI 接口 201 和该次级缓冲器 206 之间及在该次级 DASDs107 和该次级缓冲存贮器 206 之间传输。

该实施例的适配器 101 能检测到该初级和次级 DASDs107 和 108 对中的任何故障，故障能被检测是基于从该初级或次级 DASDs107 或 108 给出表明异常或缺席或一SCSI 命令的装置信息表示没有准备好等等的信号；在该适配器 101 包括一比较器 302，该比较器 302 在读操作期间将来该初级和次级 DASDs107 和 108 的数据传输给该初级和次级缓冲存贮器 205 和 206，或在响应并传送给该初级和次级缓冲存贮器 205 和 206 的一 SCSI 命令的该装置信息期间，来自该初级和次级 DASDs107 和 108 的数据被比较，当不一致产生时，该比较器的误差信息被传输给主计算机 104 或该面板 109。

图 4 示明了在该适配器 101 中，对该初级和次级 DASDs017 和 108 在一写操作时来自主计算机 104 的数据流向，图 5 示明了该写操作的步，在该写操作中，该主计算机仅送出用于写入该 DASDs 中的一个的一命令。

来自该主计算机 104 的命令由该主 SCSI 接口 201 (步 501) 接收，判定被写入的来自主机 104 的数据的规模是否大于 32k

(步 502)，如果这样，该数据被分割送入 32k 位，然后从该主机 104 传输到该适配器 101(步 503)，传输到该主 SCSI 接口 201 的数据被暂存在该初级和次级缓冲存贮器 205，和 206 中，然后，通过该初级和次级装置 SCSI 接口 202 和 203 写入到构成磁盘存贮装置的该初级和次级 DASDs107 和 108 对中的每一个，即写入命令的目标 (步 504)，在该 DASDs 对中，相同的数据总是被存在相同的地址，然后，写入数据的总量由 32k 位缩短 (步 505) 从它的初始处被启动以便接收来自该主机 104 的下一个 32k 位数据，以形成一循环。

如果来自该主机 104 的被传输的数据量少于 32k 位，则来自该主机 的被传送的数据全部由该适配器接收 (步 506)，被传输该主 SCSI 接口 201 的该数据被暂存在该初级 和次级缓冲存贮器 205 和 206 中，然后，通过该初级和次级装置 SCSI 接口 202 和 203 写入构成磁盘存贮装置的该初级和次级 DASDs107 和 108 中的每一个，即写命令 (步 507 的一目标，然后，当数据总量写完时，该状态信息和一通知被从该初级 DASDs107 送到该主机 104 (步 508)。如果该主机 104 允许该适配器 101 与该主 SCSI 总线 102 分离，该适配器 101 可以分离该主 SCSI 总线 102，和当中必须通过数据时，在该适配器 101 操作期间可以与该主 SCSI 总线 102 再次连接(步 509 和 510)，这些将会被鉴别。

由于上述该适配器 101 的操作是由该适配器 101 自身独立完成的，所以对主计算机 104 这一方来说它们是透明的，这意味着对于该主计算机 104 的操作系统必须要一单一的写操作并且一直用程序在该操作系统下运行，即在该主计算机和该 DASDs 方面不用改变任何硬件和软件就能建立用于预防故障的一冗余 DASDs 系统。

图 6 示明在该适配器 101 中，当从该级和次级 DASDs107 和 108 读数据并传输给主计算机 104 时的数据流向，图 7 示明了在该读操作期间一读操作的步，对于该 DASDs 中的一个该主计算机 104 仅送出一读命令。

来自该主计算机 104 的该命令由该主 SCSI 接口 201 接收（步 701），判断来自 DASDs 的读出数据是否大于 32k 位（步 702），如果是这样，该数据被分割送入每个 32k 位，并从该适配器 101 传输到该主机（步 703），来自该初级和次级 DASDs107 和 108 的读出数据，作为该读出命令的一目标，通过该初级和次级装置 SCSI 接口 202 和 203 被暂存在该初级和次级缓冲存贮器 205 和 206 中，然后，具有来自该级 DASDs107 的数据通过主 SCSI 接口 201（步 704）被传输到该主机 104。也就是，总是从该 DASDs107 和 108 对中读出数据，但只有从该初级 DASDs107 读出的数据被传输到该主机，读出数据的总量由 32k 位缩短（步 705）和然后一环路从它的初始处被启动，以便从该初级和次级 DASDs 107 和 108 接收下一个 32k 位数据。

如果读出的数据总量总小于 32k 位，那么，来自该初级和次级 DASDs107 和 108（步 706）的整个数据被接收，然后接收来自该初级和次级 DASDs107 和 108 的状态信息和一通知（步 707），该数据，状态信息和通知被暂时存在该初级和次级缓冲缓冲存贮器 205 和 206，然后具有来自该初级缓冲存贮器 205 的数据通过该主 SCSI 接口 201 被传输到该主机 104（步 708），来自该初级缓冲存贮器 205 的所有数据被传输之后，具有来该初级 DASDs107 的状态信息和通知被传送给该主机 104（步 711），如果该主机 104 允许该适配器 101 与该主 SCSI 总线 102 相分离，那么该适配器 101 将可以与该主 SCSI 总线 102 分离，

具有如果需要通过数据，在该适配器 101 操作期间，该适配器 101 可以与该主 SCSI 总线 102 再次相连接。

由于该适配器 101 的上述操作是由适配器 101 自身独立完成的，所以对主计算机 104 而言是透明的，这意味着对于该主计算机 104 的操作系统只有一单一的读操作是必须的并且一应用程序在该操作系统下运行。那就是在该主计算机和该 DASDs 方面，不用改变任何硬件和软件就能容易地建立用于预防故障的一 DASDs 系统。来自该初级和次级 DASDs107 和 108 的数据和状态信息可以由比较器 302 进行比较进行签别，当这种比较被传送到该初级和次级缓冲存贮器 205 和 206 表明有任何不协调时，传送给该主机 104 或该 MPU209。

参照图 8 和图 9，在下述中描述了由该主计算机 104 控制的一写操作期间，当该初级和次级 DASDs107 和 108 对中的一个发生故障时，该适配器 101 所进行的操作。首先，当该初级 DASDs107 发生故障时，它送出“检 验状况”给该适配器 101，该适配器 101 送出“请求读出”给该初级 DASDs107 并从该初级 DASDs107 记录读出数据，该适配器 101 记录该初级 DASDs107 的一最终选取的逻辑块地址（LAB），然后，如图 8 所示，该适配器 101 关闭到该初级缓冲存贮器 205 和该初级 DASDs107 的一数据通道，然后进行对该次级缓冲存贮器 206 和该次级 DASDs108 的一写操作，在对该次级 DASDs108 的写操作完成时，该适配器 101 通知该主计算机 104 读写操作被成功地完成了，然后该适配器 101 试行由利用一重新指定块命令来恢复该初级 DASDs107 的该错误的逻辑块地址（LBA）。上述由该适配器 101 进行的操作对该主计算机 104 是透明的，那该主计算机 104 仅知道对该磁盘存贮装置的写操作已被正常完成了。

参照图 9，现在描述的操作是该初级 DASDs107 功能正常和该次级 DASDs108 发生故障的情况，在这种情况，参照图 8，除了通向该次级 DASDs108 的数据通道被关闭以外，与初级 DASDs107 失效时一样具有相同的操作，并且通常的写操作是对该初级 DASDs107 进行。

如果 DASDs107 和 108 两个都失效，该适配器 101 送出两者中任何一个的较后发生故障的状态信息给该主计算机 104，那就是，该主机对故障识别是基于两者中任何一个较后发生故障的一个的状态。

在下述中，参照图 10 和图 11 描述的是由该主计算机 104 控制下在一读操作期间，当该初级和次级 DASDs107 和 108 对中的任何一个发生故障时，该适配器 101 完成的操作。首先，当该初级 DASD107 失效时，它送出一“检 验状态”给该适配器 101，该适配器 101 送出“请求读出”给该初级 DASDs107 并记录来自该初级 DASDs107 的读出数据，该适配器 101 记录一最终选取的该初级 DASDs107 的逻辑地址（LBA），在对该次级 DASDs108 的一读操作完成之后，该适配器 101 建立从该次级缓冲存贮器 206 到该主机 104 的一数据通道，并关闭该初级缓冲存贮器 205 和该初级 DASDs107 的一通道，如图 10 所示，在对该次级 DASDs108 的一读命令完成的时间，该适配器 101 通知该主机 104 读操作顺利完成了。然后，该适配器 10 再试行从该初级 DASDs107 读数据。上述操作对该主机 104 是透明的，即该主机 104 仅知对该磁盘存贮装置的读操作正常完成了。

如图 11 所示，是该初级 DASD107 功能正常和该初级 DASDs108 失效的情况，除了来自次级 DASDs108 的通道被关闭以外，与如图 10 所示的操作相 同，仅对该初级 DASDs107 进行一

读操作。

如果 DASDs107 和 108 二者都失效，该 适配器 101 再试行对它们中任何一个发生故障较后的一个加以恢复，如果该恢复不成功，则将发生故障较后的一个 DASDs 的 状态和从该 DASDs 读出的数据传给该主机 104，该主机对故障的识别是基于该 DASDs 的 状态。如果对故障的恢复在试行中，该读操作继续进行并试图恢复第一个发生故障的另一个 DASDs。

该适配器 101 在下述条件下对该初级和次级 DASDs 对进行一恢复操作，初级和次级 DASDs 对中的任何一个的逻辑块地址需要恢复，该初级和次级 DASDs 这二者是利用的并且一个 DASDs 包含有另一个 DASDs 中需要恢复的该逻辑块地址 (LBA) 的有效数据，由主机而不是由该适配器 101 执行的请求命令。

如在图 12 中由实线和点线指明的该恢复复操作期间，对应于失效的 DASDs 构成的从初级 DASDs 到次级 DASDs 或从次级 DASDs 到初级 DASDs 的一数据通道，然后，有效数据从包含有该数据的一个 DASDs 直接传送给需要恢复的另一个 DASDs。在该恢复操作期间，该适配器 101 从该主机 104 接收一个命令，但该命令是在该恢复操作完成之后执行。该恢复操作由该适配器 101 自动完成的并对该主机 104 透明。

如果恢复是不可能的，例如，该 DASDs 中的一个由于它功能的故障不能被恢复，该适配器 101 由面版 109 指示通知操作人员必须替换该失效的 DASDs，在操作人员用一新的 DASDs 替换该失效的 DASDs 之后，该适配器 101 自动地重建该 DASDs 进入一冗余 DASDs 系统，即该适配器 101 自动地格式化该新的 DASDs 并从 DASDs 对中正常的一个拷贝数据。该适配器 101 拷

贝数据到该新的一个 DASDs 并不影响在主机 104 和该正常的 DASDs 之间的读写操作，这样该失效 DASDs 的替换和新 DASDs 的重建并不影响该主机和该正常 DASDs 之间的操作。

图 13 示明了用于该实施例的面板 109，该面板 109 有 14 个发光三级管（LED）801，指示对应于分别连接到该 SCSI 接口的该初级和次级装置 SCSI 总线 105 和 106 中的每一个的 ID_s0 至 6 的总共 14 个 DASDs 的每种状态。进而，该面板 109 包括一用于转换适配器从正常操作模式到维修模式的模式选择开关 802，发光二级管 803 用于指示该转换状态。进而该面板 109 还包括用于转换该 DASDs 中的一个进入可替换状态的一个 DASDs 选择开关 804，发光二级管 805 用于指示该状态。图 14 示明了 LED_s801，803 和 804 的每一种状态的相互关系。

当替换该失效 DASDs 时，操作人员在该面板 109 上将 DASDs 选择开关 804 置位并等待直到该 LED805 指示，在该 LED805 指示之后，该失效的 DASDs 被卸下，用一个新的 DASDs 替换该 DASDs，然后新的 DASDs 被按装上，最后，当该 DASDs 选择开关 804 被关掉时，该适配器 101 重建该 DASDs 进入一冗余 DASDs 系统，即，该适配器 101 比较被存贮在 EEPROM211 中的该 DASDs107 和 108 的系列数从明确哪一个 DASDs 被替换，并由送出的一磁盘的格多化命令格式化该新 的DASDs，从与该替换的 DASDs 配对的该正常的 DASDs 中取数据并拷贝该数据到该新的 DASDs。如上所述，独立于主计算机 104，能移检测该 DASDs 其中一个的故障，能移替换一失效 DASDs 和能够重建一冗余 DASDs 系统，也就是不用中止该主机 104 的运行。

图 15 示明了用于该实施例，该主机 104 和该 DASDs107 和 108 的该适配器 101 的 SCSI 命令，该主机 104 把由该适配器

101 控制的该 DASDs107 和 108 作为逻辑单元处理，为使该 DASDs107 和 108 对中的每一个都持有相同内容，从该主机 104 送出的所有命令首先由该适配器 101 鉴别，对应于一种命令类型，该适配器 101 处理这些命令彼此是不同的，这些命令通常的处理如下述：首先，该适配器 101 送出一命令给 DASDs107、108 时，如果一有效的 DASDs（该初级 DASDs 或一正常 DASDs，如果该 DASDs 对中只有一个正常的 DASDs 的话）回答“好”或“较好”的话，该适配器 101 对该主机 104 回答一个来自该 DASDs107、108 的回答。

如果该有效的 DASDs 用“忙”回答，该适配器 101 在一确定的时间间隔之后 再次送出同样的命令给该有效的 DASDs，如果该有效 DASDs 用“检验 状态”或某些其它未期望状态回答，该适配器 101 假设该回答是一错误，那么就转换备用 DASDs 为有效（通常是与该初级 DASDs107 能对的该次级 DASD108），并送出相同命令给 DASDs，如果是这样，那么来自该 DASDs 的表明它已成为有效的回答送给主机 104；假如该备用 DASDs 不存在，如果这样，来自该第 1 有效 DASDs 的回答送给该主机 104。

图 15 中在右侧由 A 表明了有关命令，即读出命令，该适配器 101 送该命令给 DASDs107 和 108 对中的两个。A 数据通道通常被转换，以便数据能够从该初级 107 传输给该主机 104，如果在一读操作期间在该初级 DASDs107 中发生一错误，那么数据通道被转换到该次级 DASDs108 一侧，在这里一错误无须通知该主机 104。在读操作完成之后进行恢复该初级 DASDs107。如果在该次级 DASDs108 中也产生一错误，该命令被再试行给该次级 DASDs108，如果该错误持续下去，则停止执行该命令，并把该次级 DASDs108 的错误状态通知该主机 104，如果只有一

个 DASDs 是适用的情况下，在读命令执行期间产生一错误，则该适配器 101 再试行该命令，如果该错误持续下去，停止命令的并将仅仅一个的 DASDs 的状态通知该主机。

图 15 中右侧由 B 指明了有关命令，读出命令，该适配器 101 进行如在由 A 表明的读命令中那样多的相同操作，但并不再试行它们，那该适配器 101 对 DASDs107 和 108 对中的两个试行一写命令，如果在一个 DASDs 中产生一错误，不用通知该主机而对另一个 DASDs 执行该命令，在执行完成之后，该适配器 101 进行错误恢复，如果在该另一个 DASDs 中也产生一错误，停止写命令并将该另一个 DASDs 状态通知该主机 104；如果在只有一个 DASDs 是适用的情况下，在执行一等命令期间产生一错误，则停止执行该命令并只将仅一个的 DASDs 状态通知该主机 104。

在图 15，在右侧由 B 指明一个命令，这是一个写命令，和由 A 指出的读命令一样，适配器 101 执行相同的操作，但并不重试它们，这就是说，适配器 101 对成对的 DASDs107 和 108 两个都试写合作命令，如果在一个 DASDs 内出现写错误，不通知主机，而对另一个 DASDs 执行命令，在执行完成后，适配器 101 进行错误恢复，如果错误发生在另一个 DASDs，停止执行写命令，另一个 DASDs 的状态送到主机 104，如果在执行写命令期间发生错误，这时仅一个 DASDs 是可以利用的，停止执行命令和仅仅一个 DASDs 的状态送到主机 104。

在图 15 中右侧由 C 指明的一个命令，即一维修转换命令（OZN）被称之为一贷主异常命令（Vender unique command）也即 SCSI 命令中的一个，它在设计者设计时任意给出特殊意义。该维修转换命令是本发明的一个特征，参照附图 16 作出如下详

细描述。该适配器101进行如下涉及该维修转换命令的操作，如果一个XFER位是O和MA—MODE是XOO（X可以是任何位置），那么该DASDs107和108对通常的操作模式中进行一冗余DASDs系统功能；如果MA—MODE不是XOO的某些其它值，那么对该DASDs107和108对的冗余DASDs系统功能并且该主机104能存取该DASDs107和108对中的每一个；如果XFER位是1，那么一恢复表和存贮在该适配器101的该非易失存贮器EEPROM211中的一恢复表和一错误记录能被利用.

如果该XFER位是0，MA—MADE确定该DASD,107和108的分组；如果如上所述MA—MADE是XOO，如图17所示，该冗余DASD系统功能被进行，被分别连接到该初级和次级SCSI总线105和106并且该SCSI总线105和106在该SCSI装置接口的每一个都具相同鉴别模式的该初级和次级DASD,107和108彼此配对，那么在该主机上它们具有DASD,对功能。这是一通常的操作模式和重建立后的一缺席模式。

如果在面板109上的该模式选择开关802被置位，那么用于MA—MODE的一值能被从XOO改变为另一值，即，该适配器101能够从该通常的操作模式（在该模式下能进行该冗余DASD系统功能）改变为维修模式（在该模式下该主机104能存取DASD,107和108对中的每一个）；如图18所示，当该XFER位是0和MA—MODE是X01则脱开该冗余DASD系统功能，和在该DASD,107和108中具有ID=4，5，6的DASDs被分别连接到该初级和次级SCSI总线105和106并被给定唯一逻辑单元数，并由该主机104单独存取。如图19所示，当该XFER位是O和MA—MODE是XX10时，该冗余功能被脱开，和在该DASDs107和108中具有ID=1，2，3的DASDs被分别连接到该

初级和次级 SCSI 总线 105 和 106 并被定明一逻辑单元数值并由该主机 104 单独存取。如图 20 所示，当该 XFER 位是 0 和 MA—MOOE 是 X11 时，该冗余功能被脱开和在该 DASDs107 和 108 中具有 ID＝0 的 DASDs 被分别连接到该初级和次级 SCSI 总线 105 和 106，并被给定唯一逻辑单元数并由该主机 104 单独存取。在图中由 N/A 指明的 DASDs 不能由主机 104 存取。

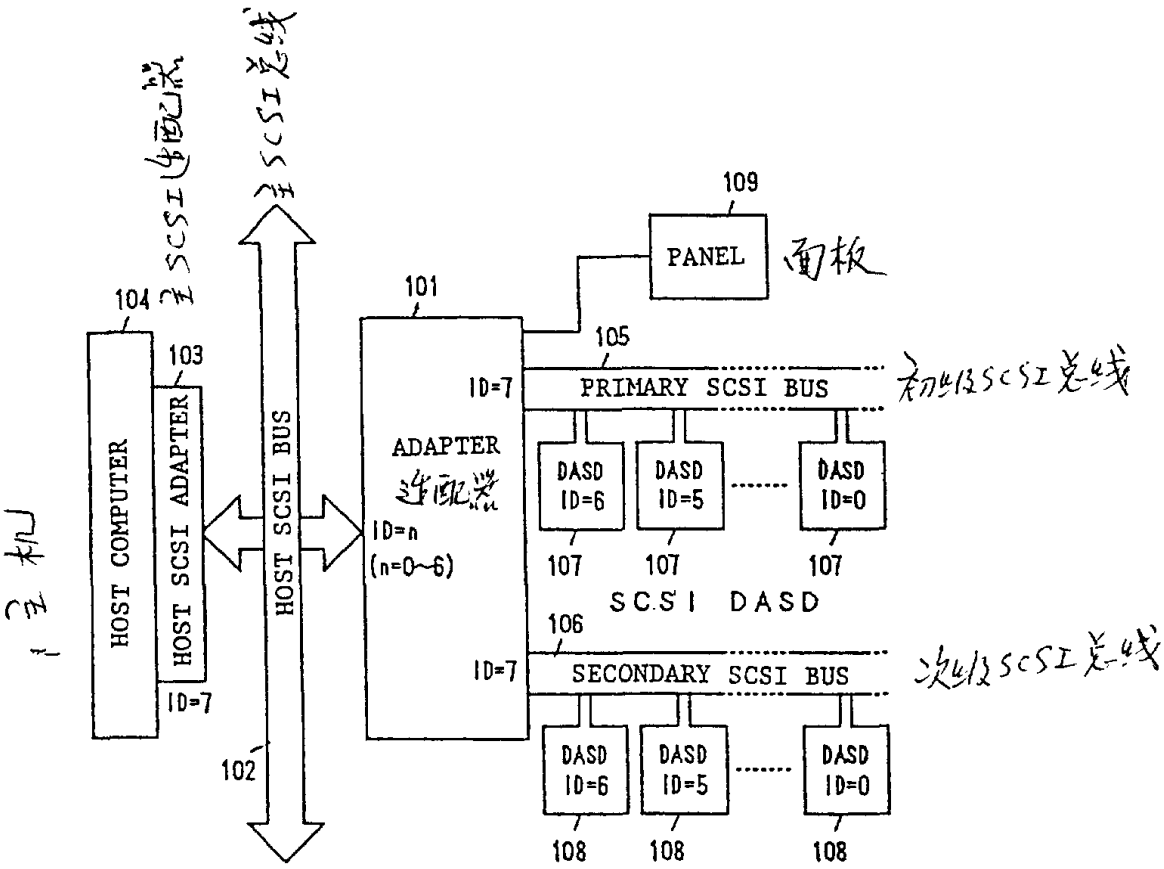
如上所述，被连接到该适配器 101 的该 DASDs107 和 108 的每一个都能由该主机 104 存取，以用于试验等的目的。在该面板 109 上的该模式选择开关 802 上的代替手动置位的软件，例如存贮在一介质的程序被称之为一基准塑料磁盒可以在该主机 104 上运行以脱开该冗余 DASDs 系统，并存取该 DASDs107 和 108 的每一个，如上所述，主机 104 仅基于这些操作，无论如何，在这种情况中，在该 DASDs107 和 108 中在未予料的情况下必须注意数据的完整。

相应于本发明的用于建立一冗余 DASDs 系统的一适配器，由于该适配器自身进行一冗余 DASDs 系统功能和一用于一主机的接口，并且用于 DASDs 的接口具有相同的接口设计，该适配器的实体对运行在该主机的一操作系统式一应用程序是透明的和该 DASDs 连接到适配器，从而在主机和 DASDs 一侧的操作系统和应用程序不需要改变。相应于本发明，用于预防故障的一冗余 DASDs 系统能够容易建立而不需要对运行在该主机或该 DASDs 上的该主机，操作系统和应用程序作任何改变，进而，相应于本发明的该适配器，由于该适配器自身具有在 DASDs 中检测故障的功能，替换失效的 DASDs、和重建一冗余 DASDs 系统，该失效 DASDs 的替换，对新 DASDs 格式化，使得不妨碍主机即不用停止主机的工作而建立冗余 DASDs 系统成为可

能。对于实验等等的目的，DASDs 中的每一个都能被存取，如果需要，对每个 DASDs 维修，这是它的优点。

说 明 书 附 图

图 1



Block Diagram Showing Whole Configuration

轮廓方框图

建议图1为文摘图

图 2

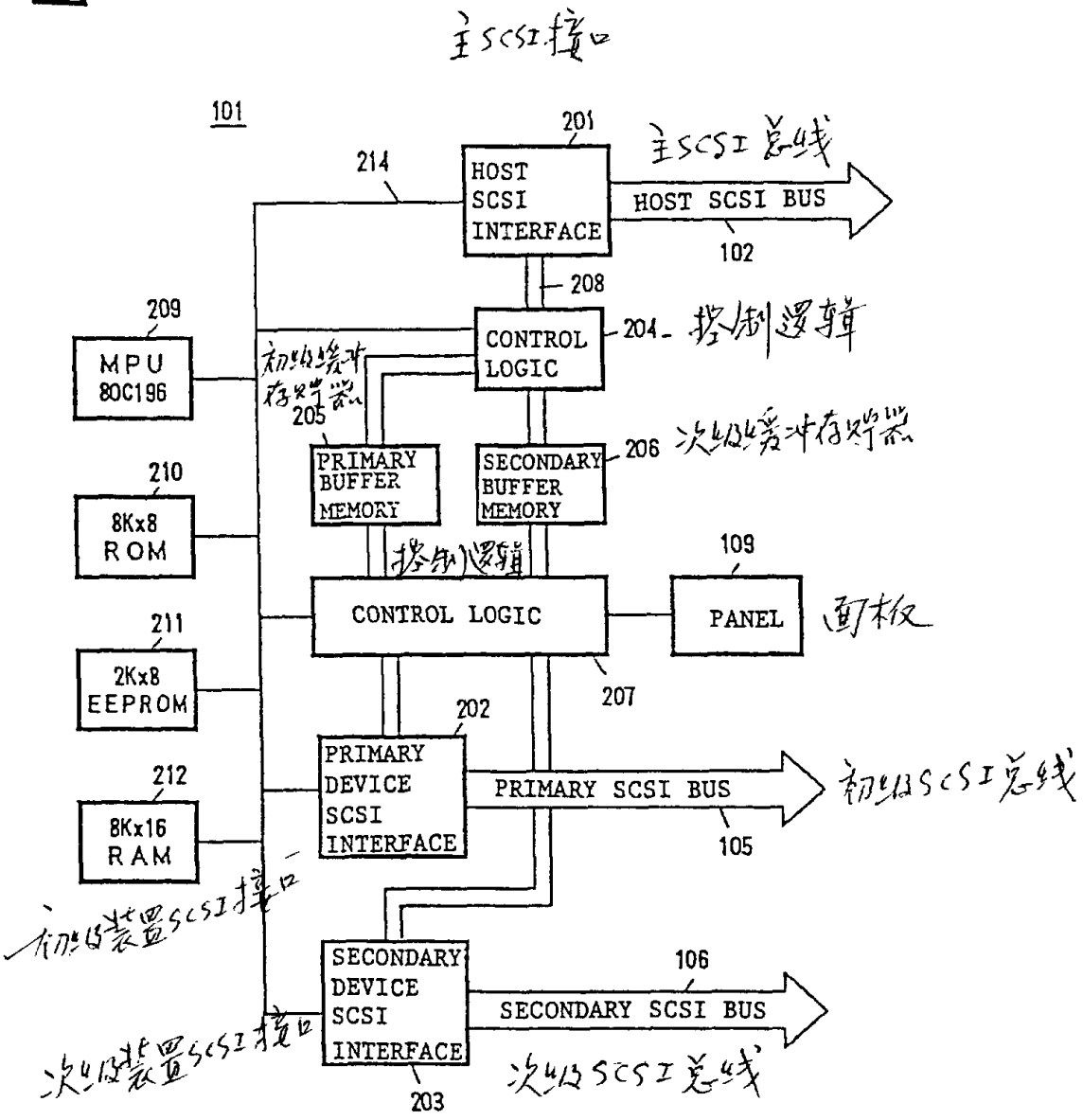
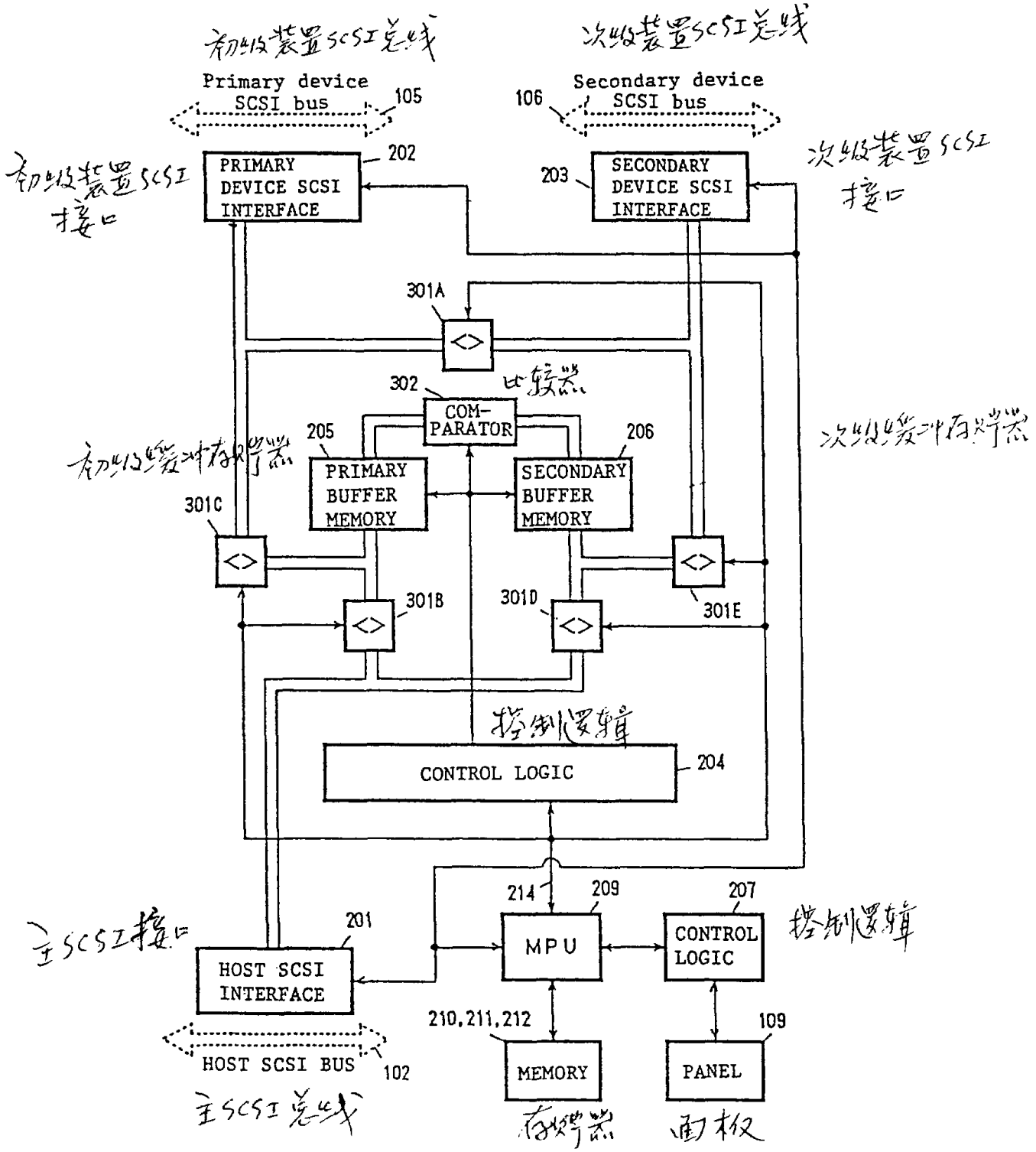

主SCSI接口
101
201
主SCSI总线
214
HOST SCSI INTERFACE
HOST SCSI BUS
102
208
209
MPU 80C196
CONTROL LOGIC
204
控制逻辑
初级缓冲存贮器
205
PRIMARY BUFFER MEMORY
SECONDARY BUFFER MEMORY
206
次级缓冲存贮器
210
8Kx8 ROM
控制逻辑
109
211
2Kx8 EEPROM
CONTROL LOGIC
PANEL
面板
207
202
212
PRIMARY DEVICE SCSI INTERFACE
PRIMARY SCSI BUS
初级SCSI总线
8Kx16 RAM
105
初级装置SCSI接口
SECONDARY DEVICE SCSI INTERFACE
106
SECONDARY SCSI BUS
次级装置SCSI接口
次级SCSI总线
203

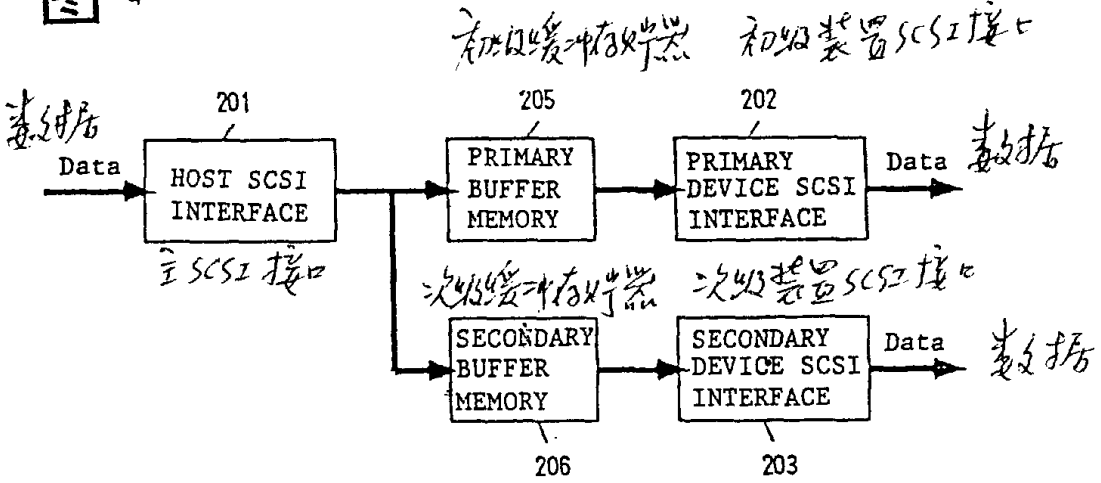
图 3



Adapter

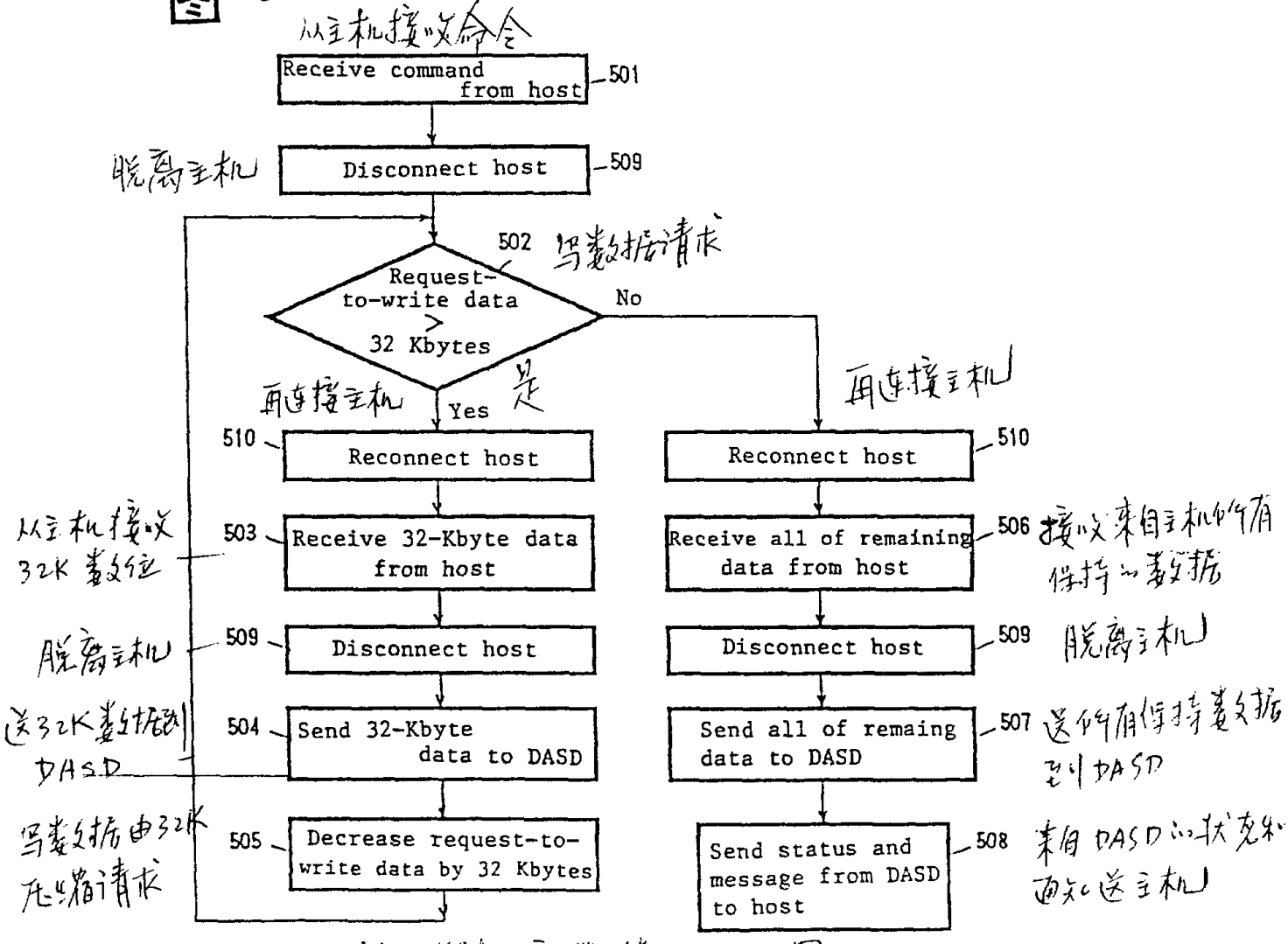
适配器

图 4



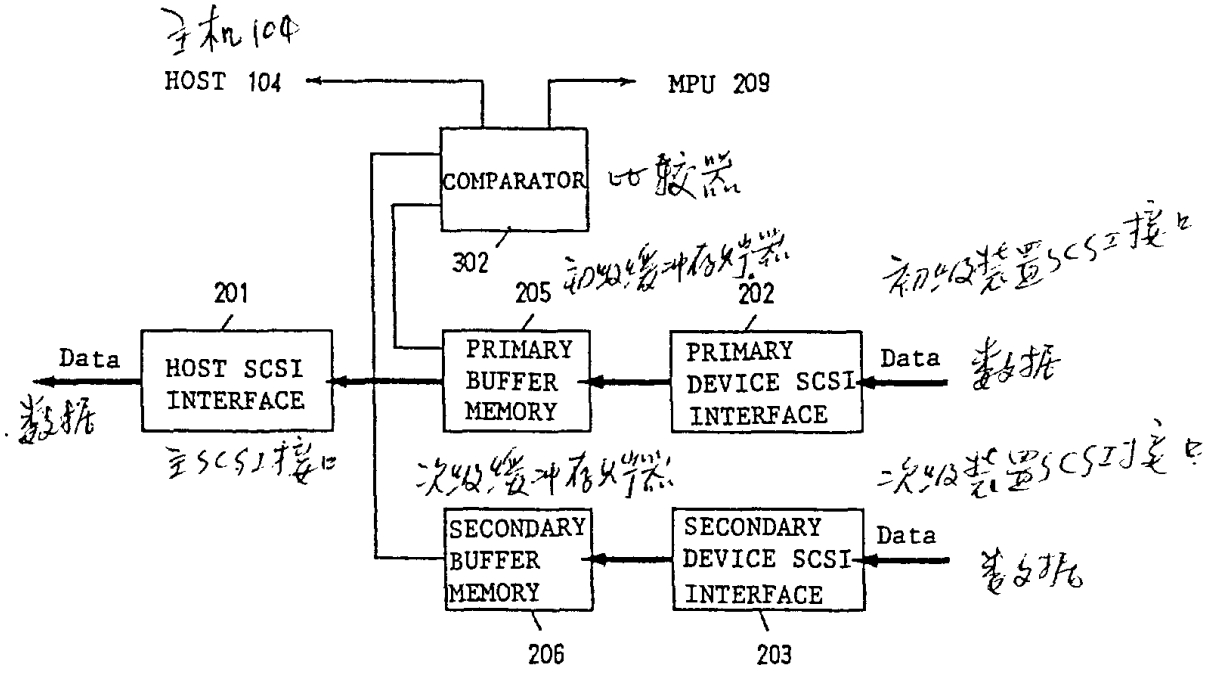
Data Flows in Adapter During Writing Operation

图 5



Flowchart Showing Operations in Adapter During Writing Operation

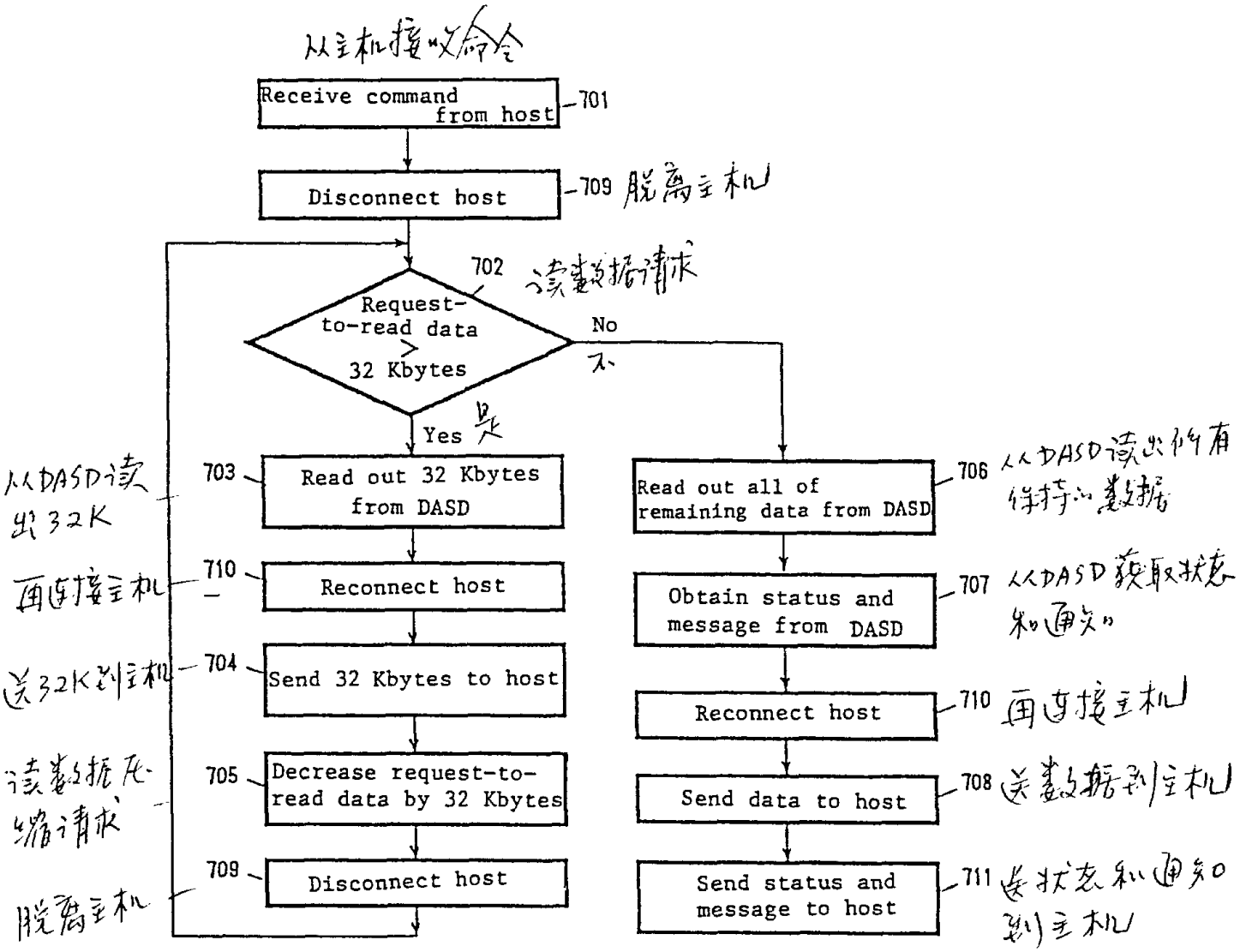
图 6



Data Flows in Adapter During Reading Operation

读操作期间适配器数据流向

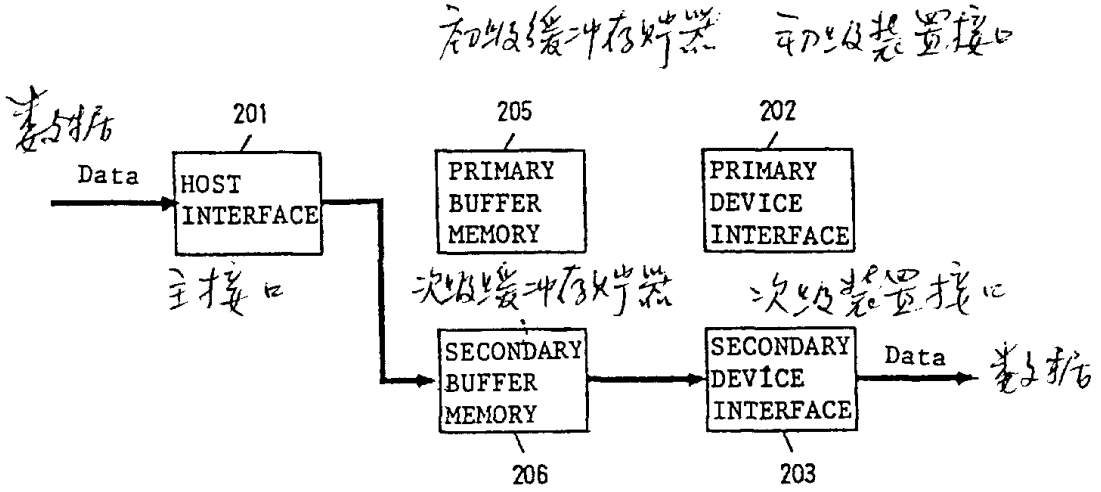
图 7



Flowchart Showing Operations in Adapter During Reading Operation

在读操作期间适配器操作的流程图

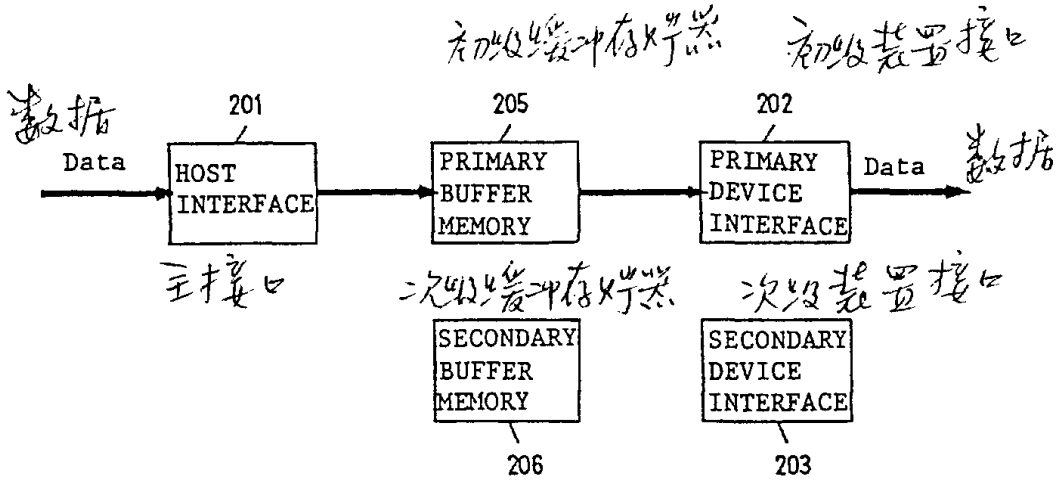
图 8



Flow of Write Data in the Case of Failure in Primary DASD

初级DASD发生故障时写数据流程

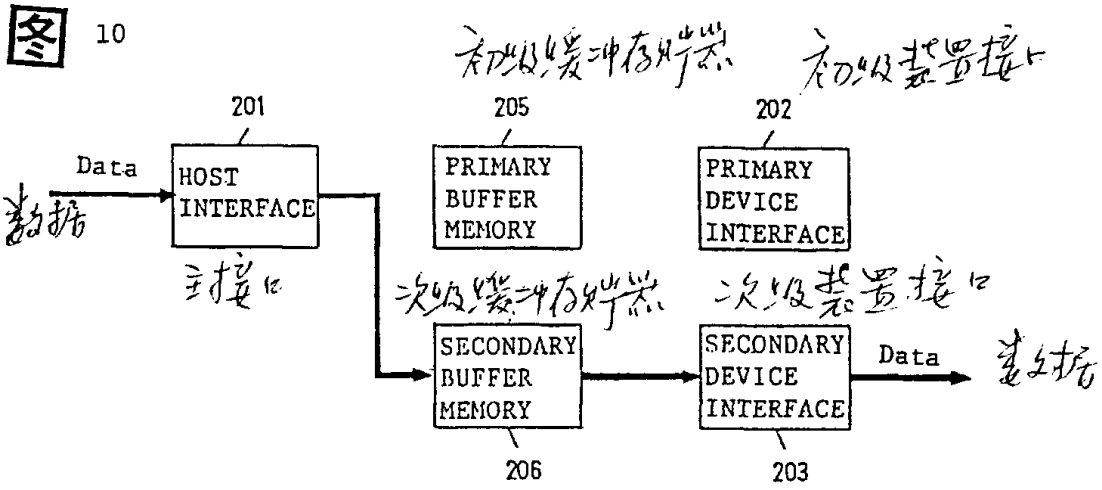
图 9



Flow of Write Data in the Case of Failure in Secondary DASD

次级DASD发生故障时写数据流程

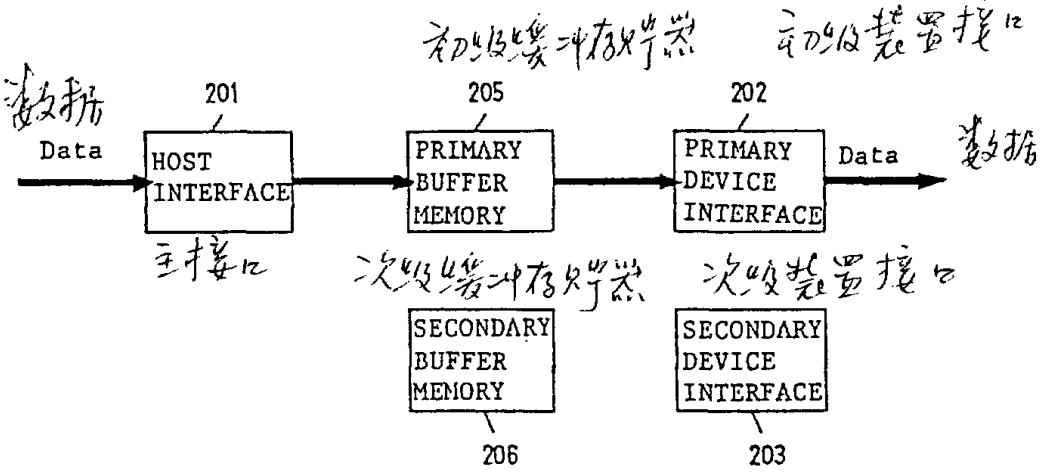
图 10



Flow of Read Data in the Case of Failure in Primary DASD

初级DASD发生故障时读数据流程

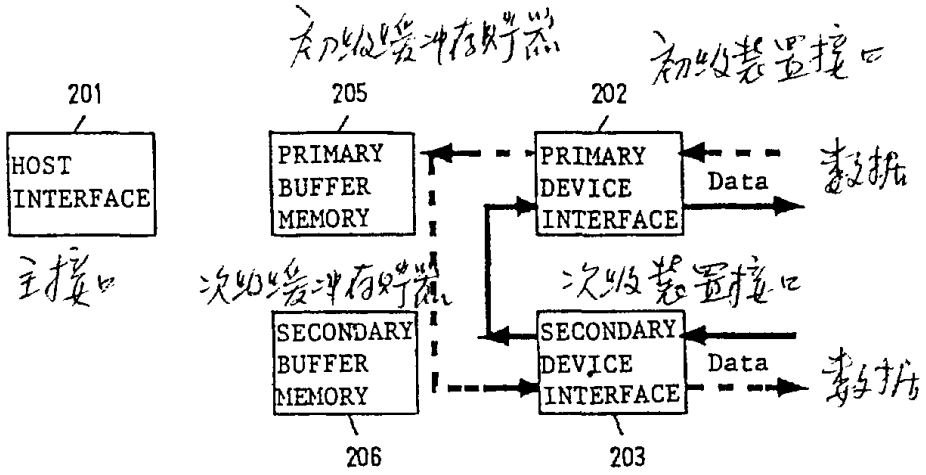
图 11



Flow of Read Data in the Case of Failure in Secondary DASD

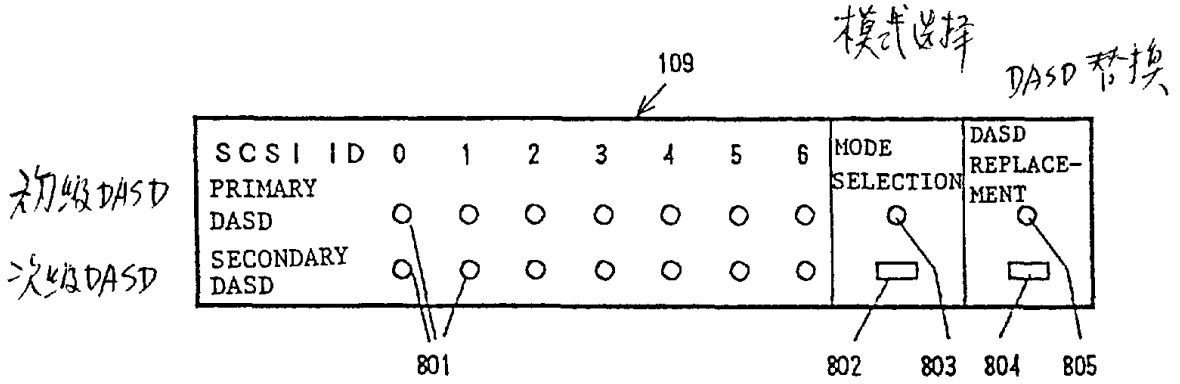
次级DASD发生故障时读数据流程

图 12



Data Flows Between Primary and Secondary DASDs

图 13



Panel

图 14

LED name	LED state	Meaning
Primary and secondary SCSI DASDs (801)	on	Corresponding one of DASDs 107 and 108 available
	off	Corresponding one of DASDs 107 and 108 does not exist or not available
	Blinks(fast)	Corresponding one of DASDs 107 and 108 in failure
	Comes on every two minutes	Corresponding one of DASDs 107 and 108 formatting
Mode selection (803)	on	Maintenance mode
	off	Operating mode of usual redundant DASD system
	Blinks(slow)	Power on self test error
DASD replace-ment (805)	on	DASD replaceable
	off	DASD connecting online to host
	Blinks(slow)	Acknowlegement for request to replace

Indication States of LED and Their Meaning

图 16

	7	6	5	4	3	2	1	0
Byte 0	Command code = 02h							
Byte 1	MA_MODE			0				XFER
Byte 2	0							
Byte 3	0							
Byte 4	ALLOCATION LENGTH							
Byte 5	UV = 0			0			FLAG	LINK

Maintenance Switch Command

图 15

16进制

CODE (hexadecimal)	SCSI Comands 命令		
00h	Test Unit Ready		试验单元准备
01h	Rezero Unit		归零单元
02h	Maintenance Switch	C	修复转换
03h	Request Sense		请求读出
04h	Format Unit		格式单元
07h	Reassign Block		指定块
08h	Read Out	A	读出
0Ah	Write	B	写
0Bh	Seek		探寻
12h	Inquiry		询问
15h	Mode Select		模式选择
16h	Reserve		保留
17h	Release		脱开
1Ah	Mode Sense		模式读
1Bh	Start/Stop Unit		启/停单元
1Dh	Send Diagnostic		送诊断行
25h	Read Capacity		读能力
28h	Read Out	A	读出
2Ah	Write	B	写
2Bh	Seek		探寻
2Eh	Write and Verify	B	写与检验
2Fh	Verify		检验
38h	Data Buffer Write		数据缓冲器写
3Ch	Data Buffer Read		数据缓冲器读
3Eh	Read Out Long	A	长读出
3Fh	Write Long	B	长写

SCSI Commands

SCSI 命令

图 17

模式选择开关，关

Mode Selector Switch	-OFF XFER-0 MA_MODE-x00b	
Logical unit number	Primary DASD SCSI ID	Secondary DASD SCSI ID
0	6	6
1	5	5
2	4	4
3	3	3
4	2	2
5	1	1
6	0	0

The Configuration of DASDs During Execution of Redundant DASD System Function

执行冗余DASD系统功能期间 DASDs 结构

图 18

模式选择开关，开

Mode Selector Switch	-ON XFER-0 MA_MODE-x01b	
Logical unit number	Primary DASD SCSI ID	Secondary DASD SCSI ID
0	6	N/A
1	N/A	6
2	5	N/A
3	N/A	5
4	4	N/A
5	N/A	4

The Configuration of DASDs in Maintenance Mode

修复模式中 DASDs 结构

图 19

模式选择开关 开

Mode Selector Switch	-ON XFER-0	MA_MODE-x10b
Logical unit number	Primary DASD SCSI ID	Secondary DASD SCSI ID
0	3	N/A
1	N/A	3
2	2	N/A
3	N/A	2
4	1	N/A
5	N/A	1

The Configuration of DASDs in Maintenance Mode

维修模式中 DASDs 结构

图 20

模式选择开关 开

Mode Selector Switch	-ON XFER-0	MA_MODE-x11b
Logical unit number	Primary DASD SCSI ID	Secondary DASD SCSI ID
0	0	N/A
1	N/A	0

The Configuration of DASDs in Maintenance Mode

维修模式中 DASDs 结构