



# [12] 发明专利申请公开说明书

[21] 申请号 200410042397.7

[43] 公开日 2005年8月17日

[11] 公开号 CN 1655111A

[22] 申请日 2004.5.28

[21] 申请号 200410042397.7

[30] 优先权

[32] 2004.2.10 [33] JP [31] 2004-032810

[71] 申请人 株式会社日立制作所

地址 日本东京都

[72] 发明人 藤本和久 井上靖雄 细谷睦

岛田健太郎 渡边直企

[74] 专利代理机构 北京银龙知识产权代理有限公司

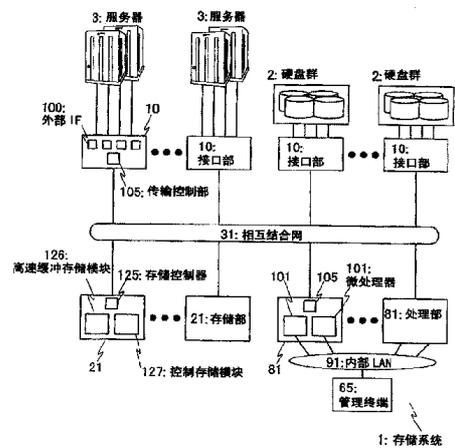
代理人 郝庆芬

权利要求书5页 说明书24页 附图23页

[54] 发明名称 存储系统

[57] 摘要

由以下这样的接口部 10、存储部 21、处理部 81 以及相互结合网 31 构成的存储系统，其中，接口部 10，具有与服务器 3 或硬盘群 2 的接口；存储部 21，具有用于存储与服务器 3 或硬盘群 2 之间的读/写数据的高速缓冲存储模块 126、以及用于存储系统控制信息的控制存储模块 127；具有微处理器的处理部 81，其中微处理器用于控制服务器和硬盘群 2 之间的数据的读/写；接口部 10、存储部 21 和处理部 81 之间通过相互结合网 31 而相互连接，如此构成存储系统。



I S S N 1 0 0 8 - 4 2 7 4

1. 一种存储系统，包含：

接口部，具有与计算机或盘装置相连的连接部；

存储部；

处理部；以及

盘装置，

其中，所述接口部、所述存储部以及所述处理部通过相互结合网而相互连接。

2. 如权利要求 1 所述的存储系统，其中，所述存储部具有高速缓冲存储器以及控制存储器，高速缓冲存储器用于存储在所述计算机或所述盘装置之间读出或写入的数据，控制存储器用于存储控制信息；

其中，所述处理器部，具有多个微处理器，用于控制所述计算机和所述盘装置之间的数据在该存储系统内的传输。

3. 如权利要求 2 所述的存储系统，其中，所述多个微处理器，在控制该存储系统内的数据传输时，通过所述相互结合网，将所述控制信息传输给成为控制对象的所述接口部或所述存储部。

4. 如权利要求 3 所述的存储系统，其中，所述相互结合网，具有传输数据的相互结合网，以及传输所述控制信息的相互结合网。

5. 如权利要求 4 所述的存储系统，其中，所述相互结合网具有多个开关部。

6. 如权利要求 5 所述的存储系统，其中，所述多个微处理器中的任何一个，都只执行所述接口部和所述存储部之间的数据传输的控制。

7. 如权利要求 6 所述的存储系统，其中，所述多个微处理器中的第 1 微处理器，只执行连接在所述计算机上的接口部和所述存储部之间的数据传输的控制，所述多个微处理器中的第 2 微处理器，仅执行连接在所述盘装置上的接口部和所述存储部之间的数据传输控制。

8. 一种存储系统，包括多个群，

其中，所述群的每一个还包含：

具有与计算机或盘装置的连接部之接口部；

存储部，具有高速缓冲存储器以及控制存储器，其中，高速缓冲存储器用于存储在所述计算机或所述盘装置之间收发的数据，控制存储器用于存储控制信息；

处理部，具有控制在所述计算机和所述盘装置之间的数据传输之微处理器；以及

盘装置；

其中，所述多个群的每一个所具有的所述接口部、所述存储部以及所述处理部，通过相互结合网，连接到所述多个群中的另一个群所具有的所述接口部、所述存储部以及所述处理部上。

9. 如权利要求 8 所述的存储系统，其中，所述多个群的每一个都具有开关部，

其中，所述多个群的每一个所具有的所述接口部、所述存储部以及所述处理部，使用所述开关部，而在所述群内相互连接；

其中，所述多个群通过在所述开关部之间连接，而相互连接。

10. 如权利要求 9 所述的存储系统，其中，在所述开关部间，使用另一个开关进行连接。

11. 如权利要求 10 所述的存储系统，其中，所述计算机请求的数据，被存储在与所述多个群中的、连接所述计算机的第 1 群不同的第 2 群所具有的盘装置内。

12. 如权利要求 11 所述的存储系统，其中，在所述计算机请求的数据被存储在与所述多个群中的、连接所述计算机的第 1 群不同的第 2 群所具有的盘装置内的情况下，所述第 1 群的所述处理部，通过所述开关部，向所述第 2 群的所述接口部发送数据传输指令。

13. 如权利要求 5 所述的存储系统，其中，所述接口部安装于第 1 电路基板上，所述存储部安装于第 2 电路基板上，所述处理部安装于第 3 电路基板上，所述开关部安装于第 4 电路基板上，

其中，还具有一个背板，其上印刷有连接在所述第 1、第 2、第 3 以及第 4 电路基板间的信号线，并具有用于将所述第 1、第 2、第 3 和第 4

电路板连接到印刷的所述信号线上的第 1 连接器；

其中，所述第 1、第 2、第 3 以及第 4 电路板具有用于连接到所述背板的所述第 1 连接器上的第 2 连接器。

14. 如权利要求 13 所述的存储系统，其中，能够连接于所述背板上的所述电路板的总数为  $n$ ，预定所述第 4 电路板的数目和连接位置，在所述第 1、第 2、第 3 和第 4 的电路板的总数不超过  $n$  的范围内，能够自由选择连接到所述背板上的所述第 1、第 2 以及第 3 电路板上各自的数目。

15. 如权利要求 9 所述的存储系统，其中，所述群的每一个，具有安装了所述接口部的第 1 电路板、安装了所述存储部的第 2 电路板、安装了所述处理部的第 3 电路板、安装了所述开关部的第 4 电路板、以及一个背板；其中，所述背板上印刷了连接在所述第 1、第 2、第 3 以及第 4 电路板间的信号线，并具有用于将所述第 1、第 2、第 3 以及第 4 电路板连接到印刷的所述信号线上的第 1 连接器；

其中，所述第 1、第 2、第 3 以及第 4 电路板具有用于连接在所述背板的所述第 1 连接器上的第 2 连接器。

16. 如权利要求 15 所述的存储系统，其中，所述多个群的数目与所述背板的数目相等。

17. 如权利要求 16 所述的存储系统，其中，所述第 4 电路板具有用于连接电缆的第 3 连接器，且连接所述第 3 连接器和所述开关部信号线被配布在基板上，

其中，所述多个群之间，通过利用所述电缆连接在第 3 连接器之间，从而被相互连接。

18. 如权利要求 5 所述的存储系统，其中，所述接口部安装于第 1 电路板上，

其中，所述存储部、所述处理部和所述开关部安装于第 5 电路板上；

其中，还具有一个背板，其上印刷由连接在所述第 1 和所述第 5 电路板间的信号线，该背板包含用于将所述第 1 和所述第 5 电路板连

接到印刷的所述信号线上的第 4 连接器，

其中，所述第 1 和所述第 5 电路基板具有第 5 连接器，用于连接所述背板的所述第 4 连接器。

19. 如权利要求 5 所述的存储系统，其中，所述接口部、所述存储部、所述处理部以及所述开关部，安装于第 6 电路基板上。

20. 一种存储系统，包含：

接口部，具有与连接计算机或盘装置相连的连接部；

存储部；

处理部；以及

盘装置部；

其中，所述接口部、所述存储部和所述处理部分之间通过相互结合网而相互连接；

其中，接收了来自所述计算机的数据读出指令之所述接口，将所述接收的指令传送给所述处理部；

其中，所述处理部分析所述指令，指定所述指令所请求的数据的存储地点，对所述存储部进行访问，并确认所述指令所请求的数据是否存储在所述存储部内；

其中，在所述存储部内存储了所述指令所请求的数据的情况下，所述处理部，通过所述相互结合网，指示所述接口部从所述存储部分读出所述请求的数据，

其中，所述接口部按照所述处理部的指示，通过所述相互结合网，从所述存储部读出所述被请求的数据，并将其传送给所述计算机；

其中，在所述存储部内没有存储所述指令所请求的数据的情况下，所述处理部，通过所述相互结合网，指示给连接有存储了所述被请求数据的所述盘装置之所述接口部，使从所述盘装置中读出所述被请求的数据，并将其存储到所述存储部，

其中，连接了所述盘装置的所述接口部，基于来自所述处理部的指示，从所述盘装置读出所述被请求的数据，并通过所述相互结合网，将其传送到所述存储部，并将传输结束传送给所述处理部；

其中，所述处理部，在接收了所述传输的结束后，通过所述相互结合网，指示给连接所述计算机的所述接口部，从所述存储部读出所述被请求的数据，将其传送到所述计算机，并

其中，连接了所述计算机的所述接口部，基于所述处理部的指示，通过所述相互结合网，从所述存储部中读出所述被请求的数据，并将其传送给所述计算机。

## 存储系统

### 技术领域

本发明涉及可从小规模到大规模的、可升级扩展结构的存储系统。

### 背景技术

近来，保存由信息处理系统处理的数据的存储系统，担负着信息系统的中心任务。在存储系统中，从小规模结构到大规模存在多种系统。

例如，在美国专利第 6385681 号中揭示了图 20 所示结构的存储系统。在该存储系统中，具有：执行与计算机(以下称为“服务器”)3 之间的数据传输的多个信道接口(以下称为“IF”)部 11、执行与硬盘群 2 之间的数据传输的多个盘 IF 部 16、对存储于硬盘群 2 内的数据临时进行存储的高速缓冲存储部 14、存储有关存储系统 8 的控制信息(例如是有关存储系统 8 内的数据传输控制的信息、存储于硬盘群 2 内的数据的管理信息等)的控制存储部 15 以及硬盘群 2。于是，信道 IF 部 11、盘 IF 部 16 以及高速缓冲存储部 14 之间，通过相互结合网 41 而连接。信道 IF 部 11、盘 IF 部 16 以及控制存储部 15 之间，通过相互结合网 42 而连接。相互结合网 41 和相互结合网 42 由共用的通路和开关构成。

在美国专利第 6385681 号记载的存储系统中，利用上述结构，在一个存储系统 8 内，构成了可从所有信道 IF 部 11 和盘 IF 部 16 来访问高速缓冲存储部 14 和控制存储部 15 的结构。

在美国专利第 6542961 号中揭示的已有技术中，如图 21 所示，众多的盘阵列装置 4 通过盘阵列开关 5 而连接到多个服务器 3 上，利用连接于盘阵列开关 5 以及各盘阵列装置 4 上的系统结构管理单元 60，将多个盘阵列装置 4 作为一个存储系统 9 来进行管理。

### 发明内容

企业存在控制对于信息处理系统的初期投资，并根据商业规模的扩展来扩展信息处理系统的倾向。由此，对于存储系统要求：初期规模小，且与事业规模相一致，具有以合理的投资来扩展规模的成本以及性能的可

扩缩性(*scalability*)。这里,对已有技术的性能的可扩缩性以及成本进行研讨。

存储系统中所要求的性能(每单位时间的数据的输入输出次数或每单位时间的数据的传输量)年年上升。因此,为了对应于未来的性能提高,还需要提高专利文献1的存储系统所具有的信道IF部11和盘IF部16的数据传输处理性能。

但是,在美国专利第6385681号的技术中,所有的信道IF部11和所有的盘IF部16,都通过高速缓冲存储部14和控制存储部15,来控制信道IF部11和盘IF部16间的数据传输。因此,如果提高信道IF部11和盘IF部16的数据传输处理性能,则要增大对高速缓冲存储器14或控制存储部的访问负担。如此,这种访问负荷会成为瓶颈,使得将来难以提高存储系统8的性能,亦即不能确保性能的可扩缩性。

另一方面,在美国专利第6542961号技术中,通过增加盘阵列开关5的端口数、或通过多级连接多个盘阵列开关5,能够增加可连接的盘阵列装置4以及服务器3的数量。即,能够确保性能的可扩缩性。

但是,在美国专利第6542961号的技术中,服务器3通过盘阵列开关5来访问盘阵列装置4。因此,产生了两次协议变换处理,即所谓的:在盘阵列开关5所具有的与服务器3的接口部中,将服务器和盘阵列开关之间的协议转换为盘阵列开关内的协议之协议转换处理;还有,在盘阵列开关5所具有的与盘阵列装置4的接口部中,将盘阵列开关内的协议转换为盘阵列开关与盘阵列装置之间的协议之协议转换处理。因此,与不通过盘阵列开关而可直接访问盘阵列装置的情况相比,应答性能差。

如果不考虑成本,则在美国专利第6385681号中,使高速缓冲存储部14或控制存储部大规模化,可以提高可允许的访问性能。但是,为了可以从所有的信道IF部11和盘IF部16来访问高速缓冲存储部14和控制存储部15,必须将高速缓冲存储部14和控制存储部15各作为相应的一个共用存储空间来进行管理。由此,如果使高速缓冲存储部14或控制存储部15大规模化,则小规模结构中的存储系统的低成本化很难,难以以低价格来提供小规模结构的存储系统。

为了解决上述问题，本发明的一个实施例具有以下结构。具体而言，本发明是这样一种存储系统，它具有：包含与计算机或盘装置的连接部的接口部、存储在计算机或盘装置之间收发的数据或控制信息的存储部、具有用于控制在计算机和盘装置之间的数据传输之微处理器的处理部、以及盘部。该存储系统在接口部、存储部、处理部之间是通过相互结合网而互相连接的。

于是，在本发明的存储系统中，由于处理部在接口部和存储部之间互送控制信息，因此，处理部根据对有关计算机请求的数据读出或数据写入而指示数据的传输。

另外，也可以将相互结合网的一部分或全部分离为传输数据的相互结合网以及传输控制信息的相互结合网来构成。再者，相互结合网也可以由多个开关部构成。

作为本发明的另一个实施例，存在以下结构。具体而言，多个群(cluster)是通过通信网而被连接的存储系统。这里，各个群包含：具有与计算机或盘装置的连接部的接口部、存储与计算机或盘装置之间的读/写数据或系统的控制信息的存储部、具有用于控制与计算机和盘装置之间的数据的读/写之微处理器的处理部、以及盘部。于是，各群内的接口部、存储部以及处理部通过通信网而与其他群内的各部相连。

各群内的接口部、存储部、以及处理部也可以这样构成：在群内通过至少一个开关部而被连接，在各群的开关部间通过连接通路而相互连接。

另外，也可以通过以另外的开关为媒介来在各群具有的开关部间连接，由此而在各群之间进行连接。

作为另一个实施例，上述实施例中的接口部也可以是具有协议处理用的处理器来构成。这种情况下，也可以这样构成：在接口部中执行协议处理，在处理部中控制存储系统内的数据传输。

此外，将利用发明的实施例以及附图来进一步说明本申请所揭示的问题以及其解决方法。

附图说明

- 图 1 图示了系统 1 的结构例；  
图 2 图示了存储系统 1 的相互结合网的详细结构例子；  
图 3 图示了存储系统 1 的另一个结构例；  
图 4 图示了图 3 所示的相互结合网的详细结构例；  
图 5 图示了存储系统的结构例；  
图 6 图示了存储系统的相互结合网的详细结构例；  
图 7 图示了存储系统的相互结合网的另一个详细结构例；  
图 8 图示了接口部的结构例；  
图 9 图示了处理部的结构例；  
图 10 图示了存储部的结构例；  
图 11 图示了开关部的结构例；  
图 12 图示了包格式的一个例子；  
图 13 图示了应用控制部的结构例；  
图 14 图示了安装到存储系统外壳上的安装例；  
图 15 图示了封装以及背板(back plane)的结构例；  
图 16 图示了相互结合网的其他详细结构例；  
图 17 图示了接口部和外部装置的连接结构例；  
图 18 图示了接口部和外部装置的其他连接结构例；  
图 19 图示了安装到存储系统外壳上的其他安装例；  
图 20 图示了已有的存储系统的结构例；  
图 21 图示了已有的存储系统的其他结构例子；  
图 22 图示了存储系统 1 的读取操作流程；  
图 23 图示了存储系统 1 的写入操作的流程。

### 具体实施方式

以下，将使用附图对本发明的实施例进行说明。

图 1 图示了第 1 实施例的存储系统的结构例。存储系统 1 具有执行与服务器 2 或硬盘群 2 的数据收发的接口部 10、处理部 81、存储部 21 以及硬盘群 2。接口部 10、处理部 81 以及存储部 21 之间通过相互结合网 31 而被连接。

在图 2 中显示了相互结合网 31 的具体结构的一个例子。

相互结合网 31 具有 2 个开关部 51。接口部 10、处理部 81 以及存储部 21，各通过一条通信通路而分别与 2 个开关部 51 相连。这里，所谓通信通路是由用于传输数据或控制信息的 1 条或多条信号线构成的传输通路。由此，在接口部 10、处理部 81 以及存储部 21 的彼此之间确保由 2 条通信通路，从而可以提高可靠性。这里，上述个数和条数只不过是一个实施例，并没有将个数限制到上述情况。这种情况也适用于以下说明的所有实施例。

相互结合网尽管是以利用了开关的情况为例来进行的说明，但如果是相互连接、传输控制信息或数据良好的网络则也可以，例如也可以通过总线来构成。

如图 3 所示，也可以将相互结合网 31 分离为传输数据的相互结合网 41 以及传输控制信息的相互结合网 42。于是，与利用 1 个通信通路来传输数据和控制信息的情况(图 1)相比，数据和控制信息的传输没有相互干扰。由此，能够提高数据和控制信息的传输性能。

图 4 图示了相互结合网 41、42 的具体结构的一个例子。相互结合网 41、42 分别有 2 个开关部 52、56。接口部 10、处理部 81 以及存储部 21，各通过一条通信通路而分别与 2 个开关部 52 以及 2 个开关部 56 相连。由此，分别确保：在接口部 10、处理部 81、以及存储部 21 的彼此之间有 2 条数据用通路 91，有 2 条控制信息用通路 92，可以提高可靠度。

图 8 图示了接口部 10 的结构的具体例子。

接口部 10 具有：与服务器 3 或硬盘群 2 相连的 4 个 IF(外部 IF) 100、控制与处理部 81 或存储部 21 之间的数据/控制信息的传输的传输控制部 105、以及执行数据的缓冲或控制信息的存储之存储模块 123。

外部 IF 100 与传输控制部 105 相连。存储模块 123 连接到传输控制部 105。传输控制部 105，即使作为用于控制对于存储模块 123 的数据/控制信息的读/写的存储控制器，也能执行操作。

这里，外部 IF 100 或存储模块 123 和传输控制部 105 间的连接结构只不过是一个实施例，其结构并没有被限制为上述内容。至少，也可以

是这样一种结构：可以经由传输控制部 105，从外部 IF 100 向处理部 81、存储部 21 传输数/控制信息。

在分离出图 4 所示的数据用通路 91 和控制信息用通路 92 情况下的接口部 10 中，传输控制部 105 上连接有 2 条数据用通路 91，以及 2 条控制信息用通路 92。

图 9 图示了处理部 81 的结构的具体例子。

处理部 21 具有：2 个微处理器 101、用于控制与接口部 10 或存储部 21 之间的数据/控制信息的传输之传输控制部 105、以及存储模块 123。存储模块 123 连接到传输控制部 105。传输控制部 105 即使作为用于控制对于存储模块 123 的数据/控制信息的读/写的存储控制器，也能执行操作。存储模块 123 作为 2 个微处理器 101 的主存储而被共用，用于存储数据或控制信息。处理部 21，也可以不使用为 2 个微处理器 101 所共用的存储模块 123，而代之以有微处理器的数目那么多的各微处理器 101 专用的存储模块。

微处理器 101 连接在传输控制部 105 上。微处理器 101，基于存储部 21 的控制存储模块 127 内存储的控制信息，来控制对于存储部 21 所具有的高速缓冲存储器的数据读/写、目录管理、接口部 10 和存储部 21 之间的数据传输。

具体而言，例如，接口部 10 内的外部 IF 100，将表示数据的读或写的访问请求之控制信息写入处理部 81 内的存储模块 123 内。之后，微处理器 101 读出写入的控制信息，并对其解释，将表示从外部 IF 100 向哪个存储部 21 进行数据传输的控制信息及在该数据传输中必需的参数写入接口部 10 内的存储模块 123。外部 IF 100 按照该控制信息和参数来执行向存储部 21 的数据传输。

微处理器 101 执行向连接在接口部 10 上的硬盘群 2 写入的数据之冗余处理，即所谓的 RAID 处理。该 RAID 处理，即便是在接口部 10 或存储部 21 中执行也没有问题。再有，微处理器 101 也执行存储系统 1 中的存储区域的管理(逻辑变换等)。

这里，只不过是微处理器 101、传输控制部 105 以及存储模块 123

之间的连接结构的一个例子，其结构并不只限定为上述例子。也可以是能够至少在微处理器 101、传输控制部 105 及存储模块 123 之间相互传输数据的结构。

如图 4 所示，在分离了数据用通路 91 和控制信息用通路 92 的情况下，处理部 81 的传输控制部 196 上连接数据用通路 91(这里为 2 条)和控制信息用通路 92(这里为 2 条)。

图 10 图示了存储部 21 的结构的具体例子。

存储部 21 具有高速缓冲存储模块 126、控制存储模块 127 以及存储控制器 125。在高速缓冲存储模块 126 中，临时存储有写入硬盘群 2 的数据或从硬盘群 2 读出的数据(以下称为高速缓冲)。在控制存储模块 127 内，存储有：高速缓冲存储模块 126 的目录信息(有关存储高速缓冲存储器上的数据之逻辑区段的信息)；用于控制接口部 10、处理部 81 以及存储部 21 间的数据传输的信息；存储系统 1 的管理信息以及结构信息等。存储控制器 125 独立控制对于高速缓冲存储模块 126 以及控制存储模块 127 的数据的读/写处理。

存储控制器 125 控制与接口部 10、处理部 81 以及其他存储部 21 之间的数据/控制信息的传输。

这里，也可以将高速缓冲存储模块 126 和控制存储模块 127 在物理上统一为一个，而将高速缓冲存储区和控制存储区分割为在一个存储空间上的逻辑上不同的区域。由此，能够减少存储模块数目，从而能够削减部件成本。

也可以将存储控制器 125 分离为高速缓冲存储模块控制用和控制存储模块控制用两部分。

这里，在存储系统 1 具有多个存储部 21 的情况下，也可以将多个存储部 21 分为 2 组，并将存储到该组间的高速缓冲存储模块和控制存储模块的数据或控制信息做成双份。由此，在一组高速缓冲存储模块或控制存储模块中产生障碍的情况下，可以利用另一组高速缓冲存储模块或控制存储模块中存储的数据等继续执行操作，从而提高存储系统 1 的可靠性。

如图 4 所示, 在分离了数据用通路 91 和控制信息用通路 92 的情况下, 存储控制器 125 上连接了数据用通路 91(这里是 2 条)以及控制信息用通路 92(这里是 2 条)。

图 11 图示了开关部 51 的结构的具体例子。

开关部 51 具有开关 LSI 58。开关 LSI 58 具有 4 个通路 IF 130、头解析部 131、仲裁器(arbiter)132、十字开关 133、8 个缓冲器 134 及 4 个通路 IF 135。

通路 IF 130 是连接与接口部 10 相连的通信通路的 IF。接口部 10 和通路 IF 130 是一对一连接的。通路 IF 135 是连接与处理部 81 或存储部 21 相连的通信通路的 IF。处理部 81 或存储部 21 与通路 IF 135 是一对一连接的。在缓冲器 134 中, 临时存储(缓冲)有在接口部 10、处理部 81 和存储部 21 之间传输的数据包。

图 12 图示了在接口部 10、处理部 81 和存储部 21 之间传输的包之格式的一个例子。所谓包, 是在各部分之间进行数据传输时使用的协议中之数据传输的单位。包 200 具有头 210、有效负荷 220 以及纠错码 230。头 210 中, 至少存储有表示包的发送者和发送接受者的信息。在有效负荷 220 中, 存储有指令、地址、数据、状态等信息。纠错码 230 是为了检测出包传输时包内产生的错误而使用的码。

在通路 IF 130 或 135 接收了包后, 开关 LSI 58 将接收的包的头 210 送给头解析部 131。头解析部 131 基于头 210 中包含的包的发送接受者的信息, 推导出各通路 IF 间的连接请求。具体而言, 头解析部 131 推导出与由头 210 所指示的包发送接受者的装置(存储部)相连的通路 IF, 并产生接受包的通路 IF 与推导出的通路 IF 之间的连接请求。

之后, 头解析部 131 将产生的连接请求送到仲裁器 132。仲裁器 132 以推导出的各通路 IF 的连接请求为基础, 执行各通路 IF 间的调停(仲裁)。基于其结果, 仲裁器 132 对于十字开关 133 输出表示连接切换的信号。接收了信号的十字开关 133 基于信号内容来切换十字开关 133 内的连接, 来实现所希望的通路 IF 间的连接。

这里, 在本实施例中, 尽管是在各通路 IF 中一对一地持有缓冲器的

结构，但是，也可以这样构成：开关 LSI 58 持有 1 个大的缓冲器，从中，为各通路 IF 分配包存储区域。开关 LSI 58 具有存储开关部 51 内的障碍信息的存储器。

图 16 图示了相互结合网 31 的其他结构例子。

在图 16 中，将开关部 51 的通路 IF 的数目增加到 10，且开关部 51 的数目增加到 4。其结果，接口部 10、处理部 81 以及存储部 21 的数目成为图 2 结构的 2 倍。另外，在图 16 中，是这样一种结构：接口部 10 只能和一部分开关部 51 相连，处理部 81 和存储部 21 与所有的开关部 51 相连。这样，可以从所有的接口部 10 对于所有的存储部 21 以及所有的处理部 81 进行访问。

相反，也可以是接口部 10 各个与所有的开关部 51 相连接，处理部 81 和存储部 21 分别与一部分开关 51 相连接的结构。例如，假设为以下结构：将处理部 81 和存储部 21 分为 2 组，1 组与两个开关部 51 相连，另一组与剩余的 2 个开关部 51 相连。这样，也能够从所有的接口部 10 访问所有的存储部 21 和所有的处理部 81。

接下来，将描述在从服务器 3 读出记录在存储系统 1 的硬盘群 2 上的数据的情况下的处理步骤例子。在以下说明中，在使用开关 51 的数据传输中，使用了所有的数据包。又，在处理部 81 和接口部 10 的通信中，接口部 10 存储从处理部 81 发送出的控制信息(数据传输等中必要的信息)的位置是预先决定的。

图 22 是一张流程图，表示从服务器 3 读出存储系统 1 的硬盘群 2 中记录的数据之情况下的处理步骤例子。

首先，服务器 3 对于存储系统 1 发出数据的读出指令。在接口部 10 内的外部 IF 100 接收了指令(742)之后，处于指令等待(741)情况下的外部 IF 100，通过传输控制部 105 和相互结合网 31(这里取开关部 51)，将接收到的指令传送到处理部 81 内的传输控制部分 105。接收了指令的传输控制部 105 将所接收的指令写入存储模块 123 内。

处理部 81 的微处理器 101，通过轮询存储模块 123，或者通过插入表示来自传输控制部 105 的写入，来检测已将指令写入存储模块 123 这

件事。检测到指令写入的微处理器 101 从存储模块 123 读出该指令，并执行指令分析(743)。微处理器 101 推导出表示记录了分析结果、服务器 3 请求之数据的存储区域的信息(744)。

微处理器 101 根据利用指令分析所得到的存储区域的信息，以及处理部 81 内的存储模块 123 或者是存储部 21 内的控制存储模块 127 内存储的高速缓冲存储模块的目录信息，来确认在存储部 21 内的高速缓冲存储模块 126 内，是否记录由指令所请求的数据(以下称为“请求数据”)(745)。

在高速缓冲存储模块 126 中有请求数据的情况下(以下称为“高速缓冲命中(cache hit)”)(746)，微处理器 101 把为了从高速缓冲存储模块 126 向接口部 10 内的外部 IF 100 传输请求数据所需要的信息，通过处理部 81 内的传输控制部 105、开关部 51 和接口部 10 内的传输控制部 105，传送到接口部 10 内的存储模块 123，上述所需要的信息，具体而言，是存储请求数据的高速缓冲存储模块 126 内的地址和成为传送接受者的接口部 10 所具有的存储模块 123 内的地址的信息。

之后，微处理器 101 对外部 IF 100 指示从存储部 21 中读出数据(752)。

接受了指示的接口部 10 内的外部 IF 100，首先，从自接口部 10 内的存储模块 123 的规定位置读出请求数据的传输中必要的信息。以该信息为基础，接口部 10 内的外部 IF 100 访问存储部 21 的存储控制器 125，并请求从高速缓冲存储模块 126 中读出请求数据。接受请求的存储控制器 125 从高速缓冲存储模块 126 中读出请求数据，并将该请求数据传送到接收请求的接口部 10(753)。接收请求数据后的接口部 10，将所接收的请求数据传送到服务器 3 (754)。

另一方面，在高速缓冲存储模块 126 内没有请求数据的情况下(以下称为“高速缓冲未命中(cache miss)”)(746)，首先，微处理器 101 访问存储部 21 内的控制存储模块 127，在高速缓冲存储模块的目录信息内，登录了用于确保在存储部 21 内的高速缓冲存储模块 126 内存储请求数据的区域的信息、具体而言是记录指定空的高速缓冲位置的信息(以下称为“高速缓冲区域确保”)(747)。在高速缓冲区域确保后，微处理器 101，访问

存储部 21 内的控制存储模块 127，并根据控制存储模块 127 内存储的存储区域的管理信息，来推导出连接由存储了请求数据的硬盘群 2 的接口部 10(以下称为“目的接口部 10”)(748)。

此后，微处理器 101 把为了从目的接口部 10 内的外部 IF 100 向高速缓冲存储模块 126 传输请求数据所需要的信息，通过处理部 81 内的传输控制部 105、开关部 51 和目的接口部 10 内的传输控制部 105，传送到目的接口部 10 内的存储模块 123。于是，微处理器 101，为了从硬盘群 2 读出请求数据并将请求数据写入存储部 21，而向目的接口部 10 内的外部 IF 100 进行指示。

接受指示的目的接口部 10 内的外部 IF 100，基于指示，从自接口部 10 内的存储模块 123 的规定位置读出请求数据的传输中所必需的信息。以该信息为基础，目的接口部 10 内的外部 IF 100 从硬盘群 2 中读出请求数据(749)，并将所读出的数据传送到存储部 21 内的存储控制器 125。存储控制器 125，将接收的请求数据写入高速缓冲存储模块 126 内(750)。如果请求数据的写入结束，则存储控制器 125 将该结束通知给处理器 101。

检测出对于高速缓冲存储模块 126 的写入结束的微处理器 101，访问存储部 21 内的控制存储器模块 127，并更新高速缓冲存储模块的目录信息。具体而言，微处理器 101，在目录信息中登录高速缓冲存储模块的内容被更新这件事(751)。另外，微处理器 101 发送指示，要求接受数据读出的请求指令之接口部 10 从存储部 21 中读出请求数据。

接受了指示的接口部 10，与高速缓冲命中时刻的处理步骤相同，从高速缓冲存储模块 126 中读出请求数据，并传送到服务器 3。如上所述，存储系统 1，对于来自服务器 3 的数据读出请求，从高速缓冲存储模块或硬盘群 2 中读出数据，并发送给服务器 3。

接着，叙述将数据从服务器 3 写入存储系统 1 内情况下的处理步骤例子。图 23 是一张流程图，表示将数据从服务器 3 写入存储系统 1 的情况下之处理步骤的例子。

首先，服务器 3 对于存储系统 1 发行数据写入指令。在本实施例中，以在写入指令中含有应当写入的数据(以下称为“更新数据”)的情况进行

说明。但是，还存在在写入指令中不包含更新数据的情况。在这种情况下，在根据一条写入指令确认了存储系统 1 的状态后，服务器 3 发送更新数据。

在接口部 10 内的外部 IF 100 接收了指令(762)后，存在于指令等待状态(761)下的外部 IF 100，通过传输控制部 105 和开关部 51，将接收的指令传送给处理部 81 内的传输控制部 105。传输控制部 105，将所接收的指令写入处理部的存储模块 123。另外，更新数据被临时保存在接口部 10 的存储模块 123 内。

处理部 81 的微处理器 101，通过轮询存储模块 123，或者通过插入表示来自传输控制部 105 的写入等，来检测已将指令写入存储模块 123 这件事。检测出指令写入的微处理器 101，从存储模块 123 中读出该指令，并执行指令分析(763)。微处理器 101，根据指令分析的结果，推导出表示记录了服务器 3 请求写入的更新数据之存储区的信息(764)。微处理器 101，根据表示写入更新数据的存储区的信息，以及处理部 21 内的存储模块 123 或者存储部 21 内的控制模块 127 内存储的高速缓冲存储模块的目录信息，判断在存储部 21 内的高速缓冲存储模块 126 中，是否记录了成为写入请求的对象，即成为更新对象的数据(以下称为“更新对象数据”)(765)。

在高速缓冲存储模块 126 中存在更新对象数据的情况下(以下称为“轻命中 (light hit)”)(766)，微处理器 101 把为了将更新数据从接口部 10 内的外部 IF 100 传送到高速缓冲存储模块 126 所需要的信息，通过处理部 81 内的传输控制部 105、开关部 51 以及接口部 10 内的传输控制部 105，传送给接口部 10 内的存储模块 123。于是，微处理器 101，为了将从服务器 3 传送的更新数据写入存储部 21 内的高速缓冲存储模块 126 而指示外部 IF 100 (768)。

接受了指示的接口部 10 内的外部 IF 100，从自接口部 10 内的存储模块 123 的规定位置读出更新数据传输中必需的信息。以读出的信息为基础，接口部 10 内的外部 IF 100，通过传输控制部 105 和开关 51，向存储部 21 内的存储控制器 125 传输更新数据。接收了更新数据的存储控制

器 125, 在请求数据上写上高速缓冲存储模块 126 中存储的根新对象数据 (769)。在写入结束后, 存储控制器 125, 向发送指示的微处理器 101 通知更新数据的写入结束。

在检测到对于高速缓冲存储模块 126 之更新数据的写入结束了的微处理器 101, 访问存储部 21 内的控制存储模块 127, 并更新高速缓冲存储器的目录信息(770)。具体而言, 微处理器 101, 在目录信息中登录了高速缓冲存储模块的内容得以更新这件事。与此同时, 微处理器 101 为了向服务器 3 送出写入完毕通知, 而对接受了来自服务器 3 的写入请求的外部 IF 100 进行指示(771)。接受了该指示的外部 IF 100 将写入完毕通知送出给服务器 3 (772)。

在高速缓冲存储模块 126 内没有更新对象数据的情况下(以下称为“轻失中 (light miss)”) (766), 微处理器 101 访问存储部 21 内的控制存储模块 127, 并在高速缓冲存储模块的目录信息中, 登录用于确保在存储部 21 内的高速缓冲存储模块 126 内存储更新数据的区域之信息, 基体而言是指示空的高速缓冲位置的信息(高速缓冲区域确保)(767)。在高速缓冲区域确保后, 存储系统 1 执行与轻命中时相同的控制。但是, 在轻失中的情况下, 由于高速缓冲存储模块 126 内不存在更新对象数据, 因此, 存储控制器 125 将更新数据存储作为存储更新数据的地点之确保的存储区域内。

之后, 微处理器 101 判断高速缓冲存储模块 126 的空闲容量等(781), 并与来自服务器 3 写入请求不同步地, 执行将写入存储部 21 内的高速缓冲存储模块 126 内之更新数据记录到硬盘群 2 内的处理。具体而言, 微处理器 101 访问存储部 21 内的控制存储模块 127, 并根据存储区域的管理信息, 推导出连接存储更新数据的硬盘群 2 的接口部 10(以下称为“更新目的接口部 10”) (782)。之后, 微处理器 101 把为了从高速缓冲存储模块 126 向更新目的接口部 10 内的外部 IF 100 传输更新数据所需要的信息, 通过处理部 81 内的传输控制部 105、开关部 51 和接口部 10 内的传输控制部 105, 传送到更新目的接口部 10 内的存储模块 123。

此后, 微处理器 101, 为了从高速缓冲存储模块 126 中读出更新数

据，并将其传送给更新目的接口部 10 的外部 IF 100，而向更新目的接口部 10 进行指示。接受了指示的更新目的接口部 10 内的外部 IF 100，从自接口部 10 内的存储模块 123 的规定位置读出更新数据的传输中必需的信息。以读出的信息为基础，更新目的接口部 10 内的外部 IF 100 对存储部 21 内的存储控制器 125 进行指示，使从高速缓冲存储模块 126 中读出更新数据，并通过更新目的接口部 10 内的传输控制部 105，将该更新数据从存储控制器 125 传送到外部 IF 100。

接受了指示的存储控制器 125，将更新数据传送给更新目的接口部 10 的外部 IF 100 (783)。接收了更新数据的外部 IF 100 将更新数据写入硬盘群 2(784)。如上所述，对于来自服务器 3 的数据写入请求，存储系统 1 将数据写入高速缓冲存储模块，并将数据写入硬盘群 2。

在本实施例中表示的存储系统 1 中，管理终端 65 连接在存储系统 1 内，由管理终端 65 来执行系统结构信息的设置、系统的开始/停止之控制、系统内各部分的利用率、运行状况、障碍信息的收集、产生障碍时的障碍部分的闭塞/交换处理、控制程序的更新等。这里，系统的结构信息、利用率、运行状况、障碍信息存储于存储部 21 的控制存储模块 127 内。在存储系统 1 内设置了内部 LAN (局域网) 91。各处理部 81 具有 LAN 接口，管理终端 65 和各处理部 81 通过内部 LAN 91 而连接。管理终端 65 经由内部 LAN 91 来访问各处理部 81，并执行上述各种处理。

图 14 和 15 图示了将本实施例所示结构的存储系统 1 安装到外壳上之情况下的结构例子。

构成存储系统 1 的框架的外壳，具有电源单元底座 823、控制单元底座 821、以及盘单元底座 822。在这些底座上装填有上述各部分。在控制单元底座 821 的一面上，设置了印刷有连接在接口部 10、开关部 51、处理部 81 以及存储部 21 之间的信号线之背板 831(图 15)。背板 831 由在各层上印刷了信号线的多层基板构成。背板 831 具有连接了 IF 封装 801、SW 封装 802、存储封装 803、或处理器封装 804 的连接器的连接器 911。为了连接到连接各封装的连接器的连接器 911 内的规定端子上，印刷了封装 831 上的信号线。另外，用于供给各封装的电源的电源用信号线印刷在背板 831 上。

IF 封装 801 由在各层上印刷信号线的多层电路板构成。IF 封装 801 具有用于连接到背板 831 上的连接器 912。在 IF 封装 801 的电路板上，印刷有图 8 所示的接口部 10 的结构中之外部 IF 100 和传输控制部 105 间的信号线、将把存储模块 123 和传输控制部 105 间的信号线以及传输控制部 105 连接到开关 51 上之信号线连接至连接器 912 的信号线。另外，在 IF 封装 801 的电路板上，按照电路板上的布线来安装完成外部 IF 100 的作用之外部 IF-LSI 901、完成传输控制部 105 之作用的传输控制 LSI 902、以及构成存储模块 123 的多个存储器 LSI 903。

另外，在 IF 封装 801 的电路板上，还印刷了用于驱动外部 IF-LSI 901、传输控制 LSI 902 以及存储器 LSI 903 的电源以及时钟用的信号线。IF 封装 801，具有用于将电缆 920 连接到 IF 封装 801 上的连接器 913；其中，电缆 920 用于连接服务器 3 或硬盘群 2 和外部 IF-LSI 901。在电路板上印刷有连接器 913 和外部 IF-LSI 901 间的信号线。

SW 封装 802、存储封装 803 以及处理器封装 804 也是基本上与 IF 封装 801 相同的结构。即，具体而言，在电路板上安装实现上述各部分效果的 LSI，并在电路板上印刷连接其间的信号线。但是，其他封装，不具有 IF 封装 801 具有的连接器和用于与其相连的信号线。

控制单元底座 821 上设置了盘单元底座 822，用于装填安装了硬盘驱动器的硬盘单元 811。盘单元底座 822 具有背板 832，用于连接硬盘单元 811 和硬盘单元底座。盘单元 811 和背板 832 具有用于连接两者的连接器。与背板 831 相同，背板 832 由各层印刷了信号线的多层基板构成。另外，背板 832 具有连接器，用于连接连接至 IF 封装 801 上的电缆 920。在背板 832 上，印刷有连接该连接器和盘单元 811 的连接器之间的信号线以及供电用的信号线。

也可以设置连接电缆 920 的专用封装，并将该封装连接到设置在背板 832 上的连接器上。

在控制单元底座 821 下，设置了电源单元底座 823，其中收纳了向存储系统 1 的全体供电的电源单元和电池单元。

于是，将这些底座收纳于 19 英寸的(图中未示的)架子内。另外，底

座的配置关系并不仅仅限制在图示的例子，例如，也可以在外壳的顶上装填电源单元底座。

存储系统 1 也可能是没有硬盘群 2 的结构。这种情况，通过 IF 封装 801 中设置的连接电缆 920，来连接存在于与存储系统 1 不同位置上的硬盘群 2 或其他存储系统 1 和存储系统 1。在这种情况下，将硬盘群 2 收纳于盘单元底座 822 内，将盘单元底座 822 收纳于盘单元底座专用的 19 英寸的架子中。另外，存储系统 1 具有硬盘群 2，还进一步存在与其他存储系统 1 相连的情况。这种情况下，存储系统 1 和其他存储系统 1 也是通过 IF 封装 801 内设置的连接电缆 920 而相互连接。

在上述内容中，尽管是对将接口部 10、处理部 81、存储部 21 以及开关部分分别安装于各个封装内的情况进行的说明，但是，例如，也可以汇集开关部 51、处理部 81 以及存储部 21，将其安装于 1 个封装内。也可以汇集所有的接口部 10、开关部 51、处理部 81 以及存储部 21，并将它们安装于 1 个封装内。这样的情况下，改变了封装的大小，与此相应，还必需改变图 8 所示的控制单元底座 821 的宽度、高度。在图 14 中，尽管以使封装与底面垂直的形式，安装到控制单元底座 821 上，但也可以是使封装与底面平行的方式来安装到控制单元底座 821 上。可以任意决定是否将上述的接口部 10、处理部 81、存储部 21 以及开关部 51 内的任何组合安装到 1 个封装内，上述安装组合仅是一个例子。

控制单元底座 821 内可安装的封装数是由控制单元底座 821 的宽度和各封装的厚度物理决定的。另一方面，从图 2 所示的结构中可以看出，由于存储系统 1 是通过开关部 51 而使得接口部 10、处理部 81 和存储部 21 相互连接的结构，因此，可以自由设置与请求的系统规模、服务器连接数目、硬盘连接数目、以及性能相应的各部分的数目。因此，共用与图 14 所示的 IF 封装 801、存储器封装 803 以及处理器封装 804 中设置的背板 831 的连接器的，另外，通过预定搭载的 SW 封装 802 的个数、以及连接 SW 封装 802 的背板 831 上连接器，以从控制单元底座 821 上可装载的封装个数中扣除了搭载的 SW 封装的个数之数作为上限，可以自由选择安装的 IF 封装 801、存储封装 803 以及处理封装 804 的个数。通过

这样，可以按照用户请求的系统规模、服务器连接数目、硬盘连接数目、以及性能，来结构灵活地构成存储系统 1。

在本实施例中，其特征在于，从图 20 所示的已有技术的信道 IF 部 11 和盘 IF 部 16 中，分离出微处理器 103 作为处理部 81 独立。如此，可以提供这样一种存储系统，其可以与服务器 3 或是硬盘群 2 的连接接口数目的增减无关地增减微处理器数，并可以对应于所谓服务器 3 或硬盘群 2 的连接数或系统性能的用户请求的灵活的结构。

另外，在本实施例中，在数据读取或写入时，由图 1 所示的处理部 81 内的 1 个微处理器 101 统一处理由信道 IF 部 11 内的微处理器 103 执行的处理，以及由盘 IF 部 16 内的微处理器 103 执行的处理。由此，可以削减曾在已有技术中必需的、接替信道 IF 部和盘 IF 部的各个微处理器 103 之间的处理的费用。

也可以利用处理部 81 的 2 个微处理器 101、或从不同的各个处理部 81 中各选择一个选出的 2 个微处理器 101，其中一方的微处理器 101 执行与服务器 3 的接口部 10 侧的处理，另一方执行与硬盘群 2 的接口部 10 侧的处理。

在与服务器 3 的接口侧的处理负载要比与硬盘群 2 的接口侧的处理负载大的情况下，能够对前者的处理分配更多的微处理器 101 的处理量(例如处理器数、一个处理器的占有率等)。在负载的大小相反的情况下，可以对后者的处理分配更多的微处理器 101 的处理量。因此，根据存储系统内的各处理的负载的大小，能够灵活地分配微处理器的处理量(资源)。

图 5 图示了第 2 实施例的结构例。

存储系统 1 具有通过相互结合网 31 而使多个群 70-1~70-n 相互连接的结构。一个群 70 汇总具有若干连接服务器 3 或硬盘群 2 的接口部 10、存储部 21 以及处理部 81 和相互结合网 31 的一部分。一个群 70 具有的各部件的数目任意。各群 70 的接口部 10、存储部 21 和处理部 81 连接在相互结合网 31 上。因此，各群 70 的各部分，能够通过相互结合网 31 而执行与其他群 70 的各部分互送数据包。另外，各群 70 也可以具有硬盘

群 2。因此，在一个存储系统 1 中，也有包含硬盘群 2 的群 70 和不包含硬盘群 2 的群 70 混杂在一起的情况。另外，也有所有的群 70 都有硬盘群 2 的情况。

图 6 图示了相互结合网 31 的具体的结构例子。

相互结合网 31 具有 4 个开关部 51 和与其相连的通信通路。这些开关 51 分别设置在各个群 70 内。存储系统 1 具有 2 个群 70。1 个群 70 具有 4 个接口部 10、2 个处理部 81 以及存储部 21。如上所述，1 个群 70 中，包含作为相互结合网 31 的开关 51 中的 2 个。

接口部 10、处理部 81 以及存储部 21，通过每条通信通路而连接有包含各部的群 70 内的 2 个开关部 51。由此，在接口部 10、处理部 81 以及存储部 21 之间，确保 2 条通信通路，能够提高可靠度。

为了连接群 70-1 和群 70-2，1 个群 70 内的 1 个开关部 51 分别通过 1 条通信通路而连接另一个群 70 内的 2 个开关。由此，即便在一个开关部 51 出现故障或开关部 51 间的通信通路出现故障时，也能实现跨群的访问，能够提高可靠性。

图 7 图示了存储系统 1 内的群间连接的不同形式的例子。如图 7 所示，利用群间连接专用的开关部 55 来连接在各群 70 间。在这种情况下，群 70-1~3 的各开关部 51 分别通过 1 条通信通路而连接到 2 个开关部 55 上。由此，即便在一个开关部 55 出现故障、或开关部分 51-开关部 55 间的通信通路出现故障时，也能实现跨群访问，能够提高可靠性。

在这种情况下，与图 6 的结构相比，能够增加群的连接数目。即，开关部 51 上可连接的通信通路的数目在物理上有上限。但是，通过将专用开关部 55 用于群间连接，与图 6 的结构相比，能够增大群的连接数。

在本实施例的结构中，其特征也在于，在图 20 所示的已有技术中，从信道 IF 部 11 以及盘 IF 部 16 中分离出微处理器 103，使其独立于处理部 81 内。通过这样做，能够提供具有这样一种灵活结构的存储系统：可与服务器 3 或硬盘群 2 的连接接口数目的增减无关地增减微处理器的数目，并可灵活地响应服务器 3 或硬盘群 2 的连接数或系统性能这样的用户请求。

在本实施例中，也执行与第1实施例相同的数据的读和写处理。因此，在本实施例中，在数据的读或写时，由图1所示的处理部81内的1个微处理器101统一处理由信道IF部11内的微处理器103执行的处理，和由盘IF部16内的微处理器103执行的处理。通过这样做，能够削减曾在已有技术中为必需的接替信道IF部和盘IF部各个微处理器103之间的处理的费用。

另外，在本实施例中，在执行数据的读或写的情况下，存在执行从连接在一个群70上的服务器3向其他群70具有的硬盘群2(或者连接在其他群70上的存储系统)的数据读或写的情况。即便在这种情况下，也可以执行在第1实施例中说明过的读和写处理。这种情况下，将各个群70所具有的存储部21的存储空间作为存储系统1整体中的一个逻辑存储空间，由此，一个群的处理部81等能够得到用于访问其他群70的存储部21等的信息。另外，一个群的处理部81能够对于其他群具有的接口10指示数据的传输。

存储系统1为了使由各群上连接的硬盘群2构成的卷(volume)为所有处理部所共用，而由1个存储空间进行管理。

即便在本实施例中，也与第1实施例相同，将管理终端65连接在存储系统1内，从管理终端65执行系统的结构信息的设置、系统的开始/停止的控制、系统内各部分的利用率、运行状况、故障信息的收集、出现故障时的故障部位的闭塞/交换处理、控制程序的更新等。这里，系统的结构信息、利用率、运行状况、故障信息都存储在存储部21的控制存储模块127内。在本实施例的情况下，由于利用了多个群70来构成存储系统1，因此，为每个群70设置了具有辅助处理器的板(辅助处理部85)。辅助处理部85将管理终端65的指示传给各处理部81，收集来自各处理部85的信息，并完成将其传送给管理终端65的任务。通过内部LAN92来连接管理终端65和辅助处理部85。于是，在群70内设置内部LAN91，各处理部81具有LAN接口，辅助处理部85和各处理部81通过内部LAN91相连。管理终端65通过辅助处理部85而对各个处理部81进行访问，以执行上述各种处理。另外，也可以不通过服务处理器，而通过LAN等

直接将处理部 81 和管理终端 65 连接在一起。

图 17 是存储系统 1 的本实施例的再一个变形例。如图 17 所示，在连接服务器 3 或硬盘群 2 的接口部 10 上，连接有其他存储系统 4。这种情况下，存储系统 1，在连接其他存储系统 4 的接口部 10 所属的群 70 内的控制存储模块 127 以及高速缓冲存储模块 126 中，存储了其他存储系统 4 提供的存储区域(以下称为“卷”)之信息以及其他存储系统 4 内存储的(或读出的)数据。

连接其他存储系统 4 的群 70 内的微处理器 101，基于控制存储模块 127 内存储的信息，来管理其他存储系统 4 提供的卷。例如，微处理器 101，将其他存储系统 4 提供的卷作为存储系统 1 提供的卷，分配给服务器 3。由此，服务器 3 可通过存储系统 1，访问其他存储服务器 4 的卷。

这种情况下，存储系统 1 统一管理由自己具有的硬盘群 2 构成的卷和其他存储系统 4 提供的卷。

在图 17 中，存储系统 1，将表示哪个接口部 10 上连接了哪个服务器 3 的表存储于存储部 21 内的控制存储模块 127 内。于是，同一群 70 内的微处理器 101 管理该表。具体而言，在追加变更了服务器 3 和主 IF 100 的连接关系等情况下，微处理器 101 改变上述表的内容(更新、追加或删除)。由此，可以进行存储系统 1 上连接的多个服务器 3 之间的、以存储系统 1 为媒介的通信以及数据传输。这一点，在第一实施例中也可以同样实现。

另外，在图 17 中，接口部 10 上连接的服务器 3 执行与存储系统 4 之间的数据传输时，存储系统 1 通过相互结合网 31，在服务器 3 连接的接口部 10 和存储系统 4 连接的接口部 10 之间，执行数据传输。此时，存储系统 1 也可以将所传输的数据高速缓冲存储在存储部 21 内的高速缓冲存储模块 126 内。由此，提高服务器 3 和存储系统 4 之间的数据传输性能。

另外，在本实施例中，也考虑了这样一种结构：如图 18 所示，通过开关 65，在存储系统 1 和服务器 3 和其他存储系统 4 之间的进行连接的结构。这种情况下，服务器 3，通过接口部 10 内的外部 IF 100 以及开关

65, 对服务器 3 和其他存储系统 4 进行访问。通过这样做, 可以从连接在存储系统 1 上的服务器 3, 对连接在由开关 65 和多个开关 65 构成的网络上的服务器 3 或其他存储系统 4 进行访问。

图 19 图示了将图 6 所示结构的存储系统 1 安装于外壳内的情况下之结构例。

安装的结构基本上与图 14 的安装结构相同。即, 将接口部 10、处理部 81、存储部 21 以及开关部 51 安装于封装内, 并连接在控制单元底座 821 内的背板 831 上。

在图 6 的结构中, 将接口部 10、处理部 81、存储部 21 以及开关部 51 作为群 70 而被分组。因此, 为每个群 70 准备 1 个控制单元底座 821。一个群 70 内的各部被安装在一个控制单元底座 821 上。即, 将不同的群 70 的封装安装于不同的控制单元底座 821 上。另外, 为了群 70 间的连接, 如图 19 所示, 在安装于不同的控制单元底座的 SW 封装 802 之间, 通过电缆 921 来连接。这种情况下, 如图 19 所示的 IF 封装 801 相同, 在 SW 封装 802 上安装了电缆 921 连接用的连接器。

安装于 1 个控制单元底座 821 上的群的数目, 也可以不是 1 个。例如, 安装于 1 个控制单元底座 821 上的群数目也可以是 2。

在实施例 1 和 2 的结构的存储系统 1 中, 由接口部 10 接收的指令的分析是由处理部 81 执行的。但是, 按照在服务器 3 和存储系统 1 之间互送的指令之协议有多种多样, 利用普通的协议来执行所有的协议分析处理是不现实的。这里, 所谓协议, 有例如是使用文件名的文件 I/O(输入/输出)协议、iSCSI(互联网小型计算机系统接口)协议、使用大型计算机(主机)作为服务器时的协议(信道指令字: CCW)等。

因此, 在本实施例中, 将高速处理这些协议的专用处理器追加到实施例 1 和 2 的所有或部分接口部 10 上。图 13 是一张图, 它表示在传输控制部 105 上连接的微处理器 102 的接口部 10(以下将该接口部 10 称为“应用控制部 19”)的一个例子。

本实施例的存储系统 1, 具有应用控制部 19, 以替代实施例 1 和 2 的存储系统 1 所具有的所有或部分接口部 10。应用控制部 19 与相互结

合网 31 相连。这里，应用控制部 19 具有的外部 IF 100，成为专用于接收遵循应用控制部 19 的微处理器 102 所处理之协议的指令之外部 IF。但是，也可以是利用 1 个外部 IF 100，来接收按照不同协议的多个指令的结构。

微处理器 102 与外部 IF 100 联动地执行协议转换处理。具体而言，在应用控制部 19 接受了来自于服务器 3 的访问请求的情况下，微处理器 102 执行将外部 IF 接收的指令的协议转换成内部数据传输用协议的转换处理。

也可以不准备专用的应用控制部 19，而考虑使用这样一种结构：原样不动地使用接口部 10，将处理部 81 内的微处理器 101 中的一个作为协议处理专用的处理器。

本实施例中的数据的读和写处理，与第 1 实施例相同地执行。但是，在第 1 实施例中，接收指令的接口部 10 不分析指令，而将该指令传送给处理部 81，但是在本实施例中，在应用控制部 19 中，执行指令的分析处理。然后，应用控制部 19 将其分析结果(指令内容、数据的接受者等)传送给处理部 81。处理部 81 基于分析信息，执行存储系统 1 内的数据传输控制。

另外，作为本发明的其他实施方式，还考虑了以下结构。具体而言，具有：多个接口部，具有与计算机或盘装置的接口；多个存储部，具有高速缓冲存储器和控制存储器，其中，高速缓冲存储器用于存储与计算机或盘装置之间的读/写数据，而控制存储器用于存储系统的控制信息；以及，多个处理部，具有控制与计算机和盘装置之间的数据读/写的微处理器。多个接口部、多个存储部以及多个处理部通过由至少一个开关部构成的相互结合网而彼此连接，是通过相互结合网，在多个接口部、多个存储部、以及多个处理部之间执行数据或控制信息的收发之存储系统。

于是，在本结构中，接口部分、存储部、以及处理部，具有控制数据或控制信息的收发之传输控制部。在本结构中，接口部安装于第 1 电路基板上，存储部安装于第 2 电路基板上，处理部安装于第 3 电路基板上，至少一个开关部安装于第 4 电路基板上。另外，在本结构中，印刷

有连接在第 1-4 的电路基板间的信号线,并具有包含用于将所述第 1-4 的电路基板连接到印刷的信号线上之第 1 连接器的至少 1 个背板。另外,在本结构中,第 1-4 电路基板具有用于连接到所述背板的第 1 连接器上的第 2 连接器。

此外,在上述实施例中,设能够连接到背板上的电路基板的总数为  $n$ ,预定第 4 电路基板的数目以及连接位置,在 1-4 的电路基板总数没有超过  $n$  的情况下,也可以自由选择连接到背板上的所述第 1、第 2、以及第 3 电路基板上各自的数目。

作为本发明的另一个实施例,还考虑了以下结构。具体而言,提供了一种具有多个群的存储系统,这些群包含:多个接口部,包含与计算机或盘装置的接口;包含高速缓冲存储器和控制存储器的多个存储部,其中,高速缓冲存储器用于存储与计算机或盘装置之间的读/写数据,控制存储器用于存储系统的控制信息;以及多个处理部,包含具有用于控制计算机和盘装置之间的数据读/写的微处理器。

在本结构中,各群具有的多个接口部、多个存储部以及多个处理部之间,通过由多个开关部构成的相互结合网跨多个群而相互连接。由此,通过相互结合网,在各群之间,在多个接口部、多个存储部以及多个处理部之间执行数据或控制信息的收发。另外,在本结构中,接口部、存储部以及处理部,分别具有用于控制与各个开关相连的、数据或控制信息的收发的传输控制部。

另外,在本结构中,接口部安装于第 1 电路基板上,存储部安装于第 2 电路基板上,处理部安装于第 3 电路基板上,至少一个开关部安装于第 4 电路基板上。于是,在本结构中,印刷有连接第 1-4 电路基板间的信号线,并具有多个背板,该背板包含用来将第 1-4 电路基板连接在印刷的信号线上的第 1 连接器;还具有第 2 连接器,用于将第 1-4 电路基板连接到所述背板的第 1 连接器上。在本结构中,群由连接第 1-4 电路基板的背板构成。另外,也可以是群数和背板数目相等的结构。

再有,在本结构中,第 4 电路基板具有用来连接电缆的第 3 连接器,连接第 3 连接器和开关部的信号线配置在第 4 基板上。通过这样做,群

间是通过电缆连接第3连接器间，从而得以连接。

再有，作为本发明的另一实施例，还考虑了以下结构。具体而言，本实施例是一种存储系统，它具有：接口部，包含与计算机或盘装置的接口；包含高速缓冲存储器和控制存储器的存储部，其中，高速缓冲存储器用于存储与计算机或盘装置之间的读/写数据，控制存储器用于存储系统的控制信息；以及处理部，包含具有用于控制计算机和盘装置之间的数据读/写的微处理器，接口部、存储部和处理部之间通过由至少一个开关部构成的相互结合网而相互连接，通过相互结合网，在多个接口部、多个存储部以及多个处理部之间执行数据或控制信息的收发。在本结构中，通过相互结合网，在接口部、存储部以及处理部之间，执行数据或控制信息的收发。

在本结构中，接口部安装于第1电路基板上，存储部、处理部和开关部安装于第5电路基板上。于是，在本结构中，印刷有连接在第1和第5电路基板间的信号线，本结构具有至少一个背板，它具有用于将第1和第5电路基板连接到印刷信号线上的第4连接器；本结构还具有第5连接器，用于将第1和第5电路基板连接到背板的第4连接器上。

再者，作为本发明的又一个实施例，考虑了以下结构。具体而言，本实施例是一种存储系统，包含：接口部，包含与计算机或盘装置的接口；包含高速缓冲存储器和控制存储器的存储部，其中，高速缓冲存储器用于存储与计算机或盘装置之间的读/写数据，控制存储器用于存储系统的控制信息；以及处理部，包含具有用于控制计算机和盘装置之间的数据读/写的微处理器，接口部、存储部和处理部之间通过由至少一个开关部构成的相互结合网而相互连接。在本结构中，接口部、存储部、处理部以及开关部安装于第6电路基板上。

根据本发明，能够提供一种存储系统，它具有可以灵活响应针对服务器连接数、硬盘连接数、系统性能的用户请求的结构。在解除了存储系统的共有存储器瓶颈的同时，还能提供这样一种存储系统：谋求小规模结构的低成本化，可实现从小规模到大规模结构下成本和性能的可扩展性。

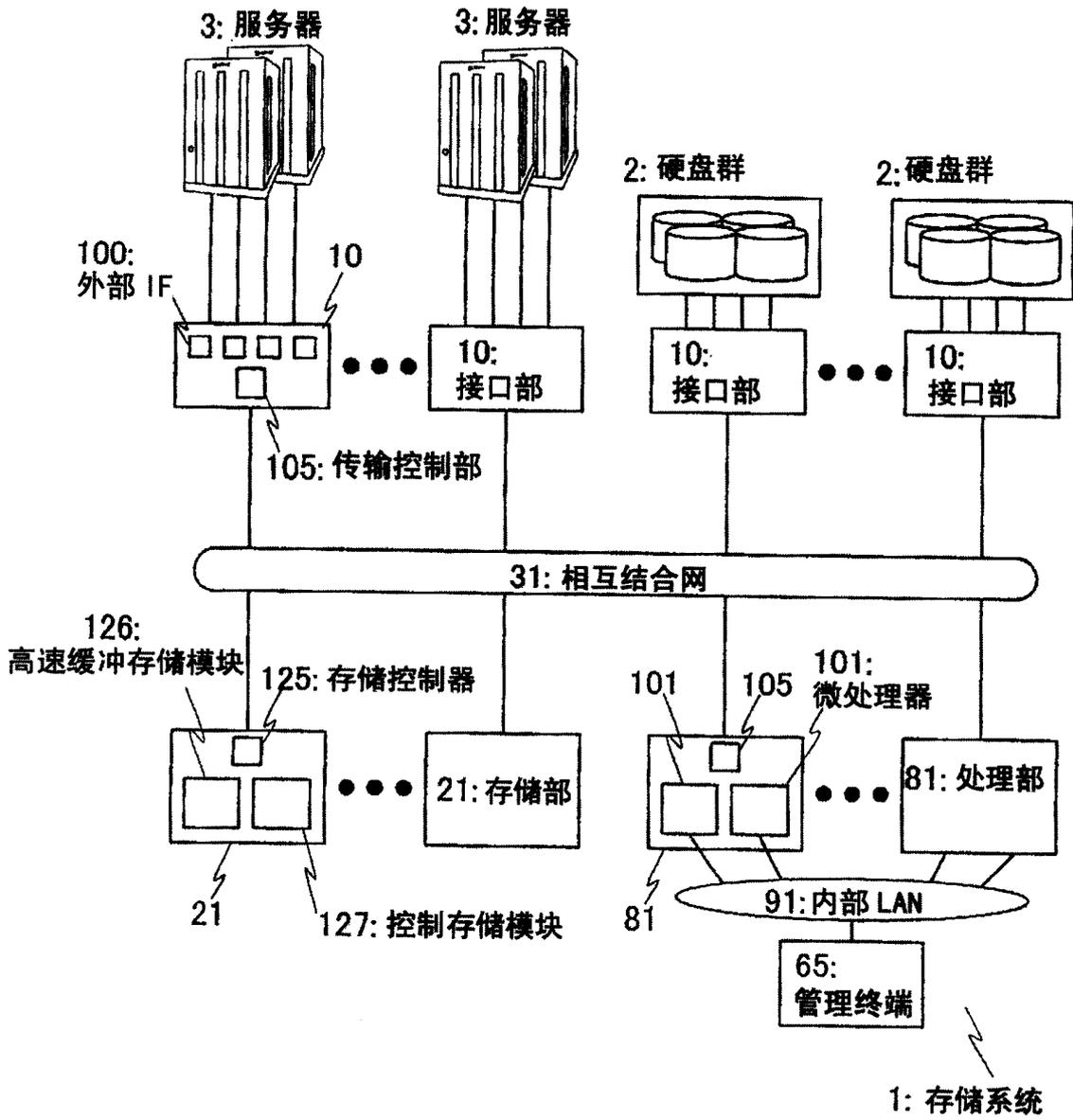


图 1

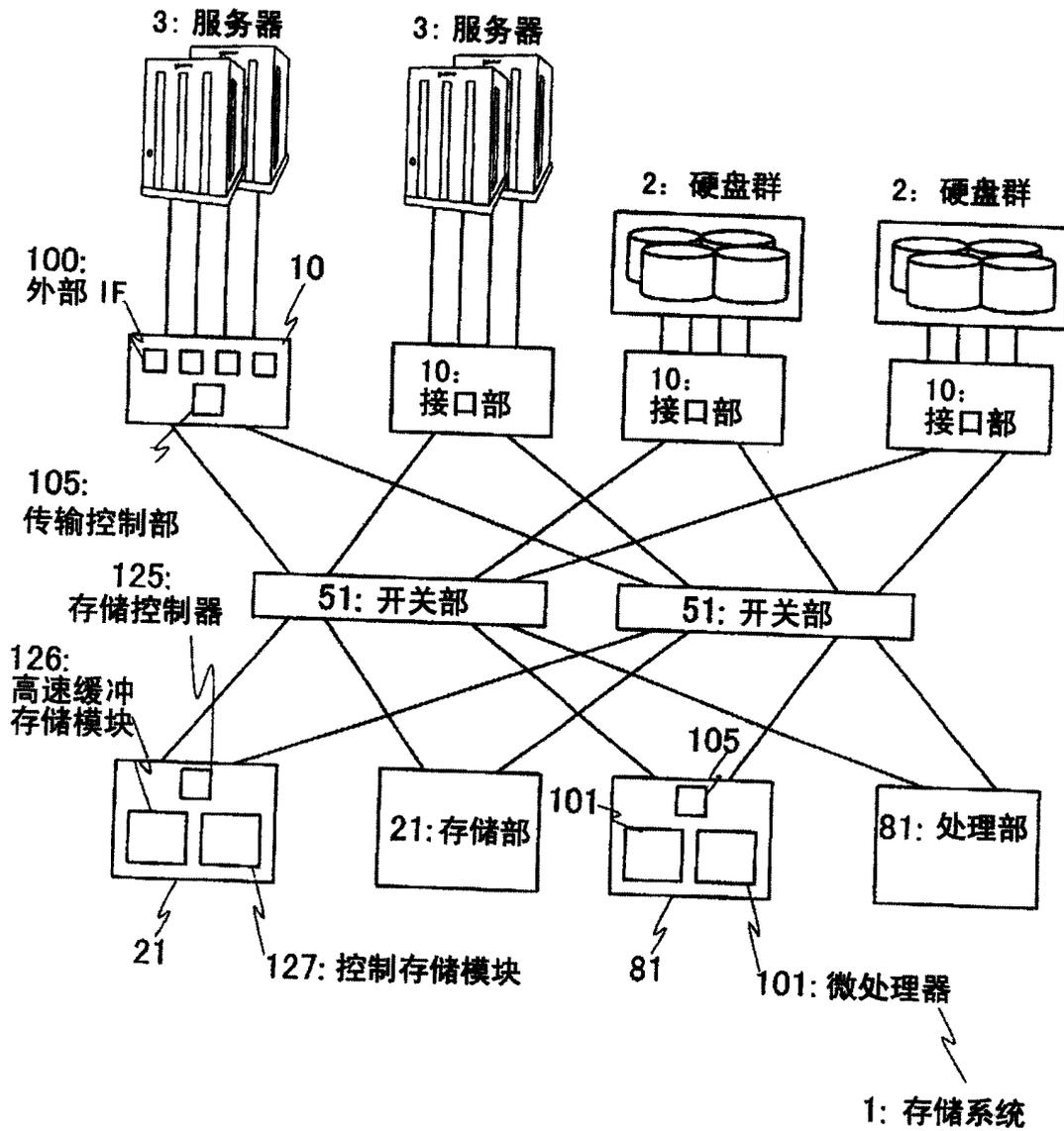


图 2

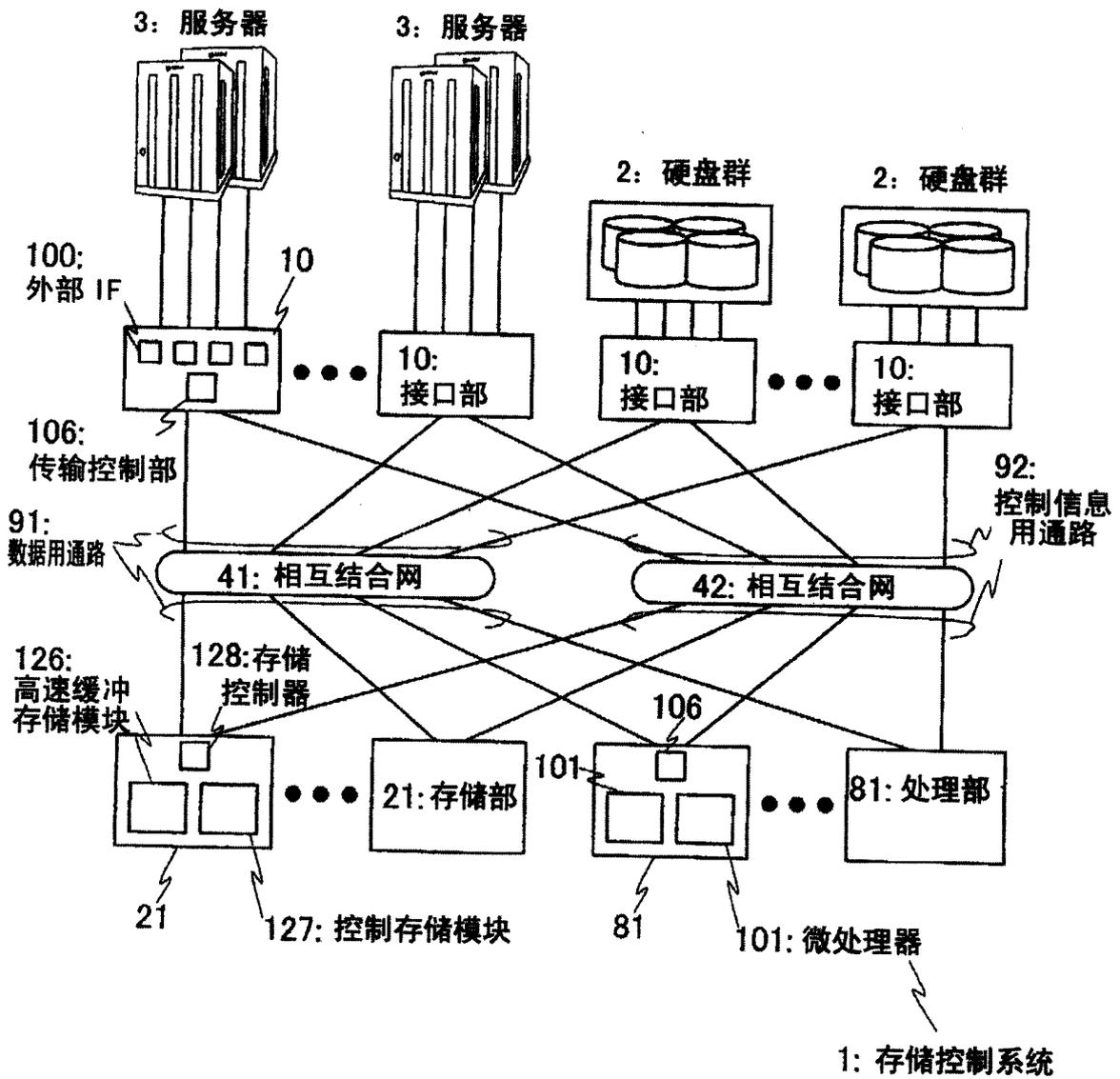


图 3

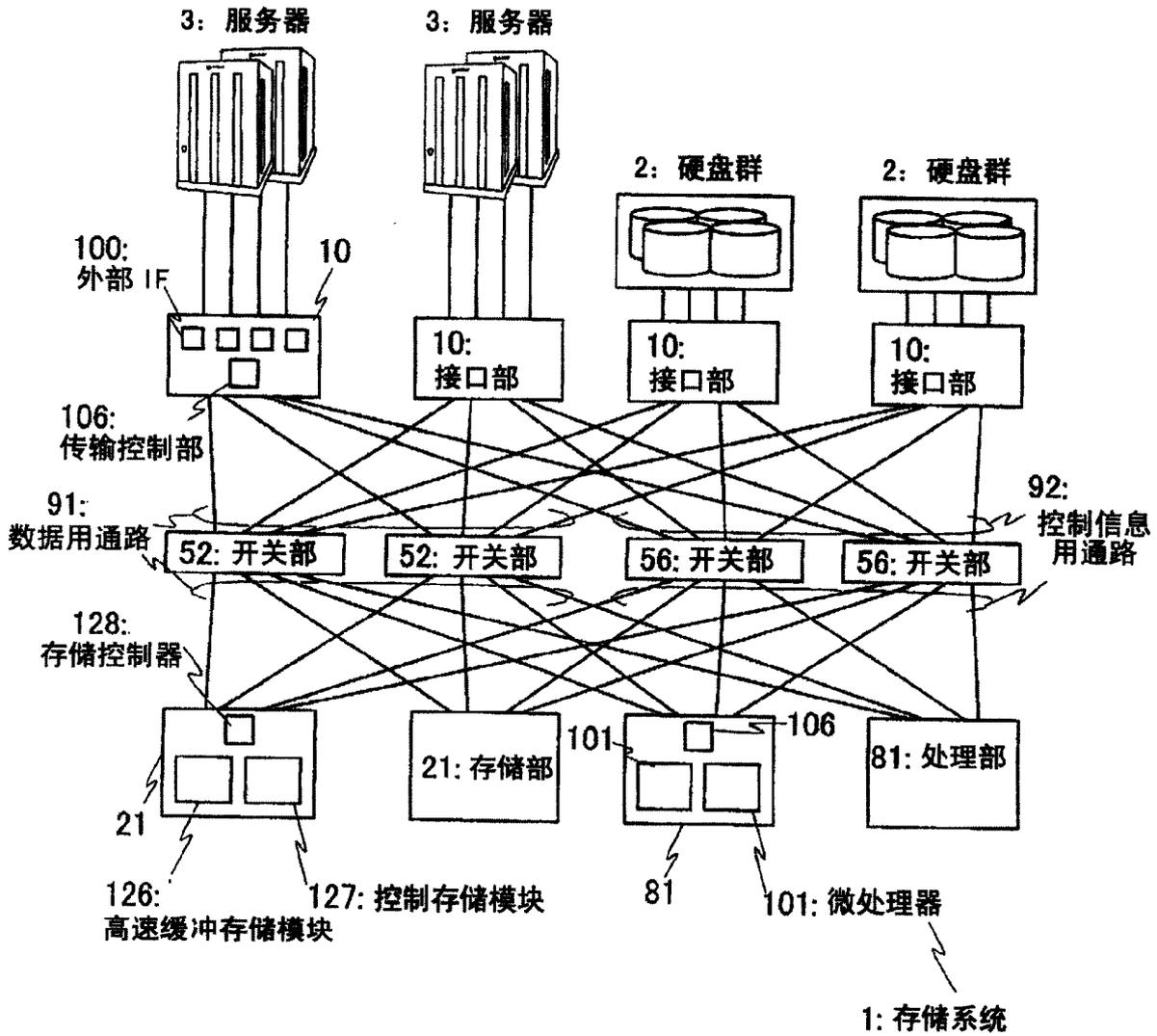


图 4

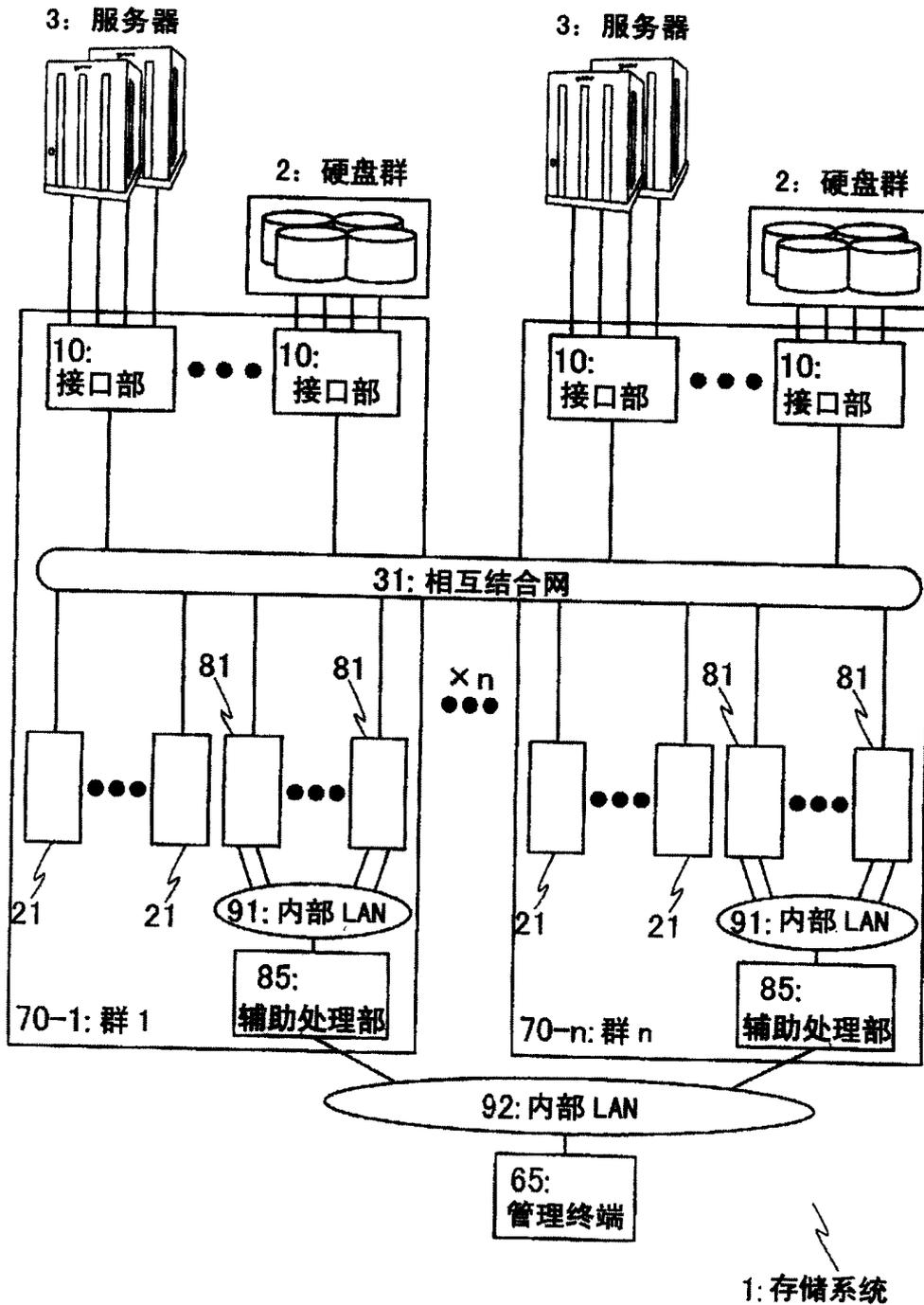


图 5

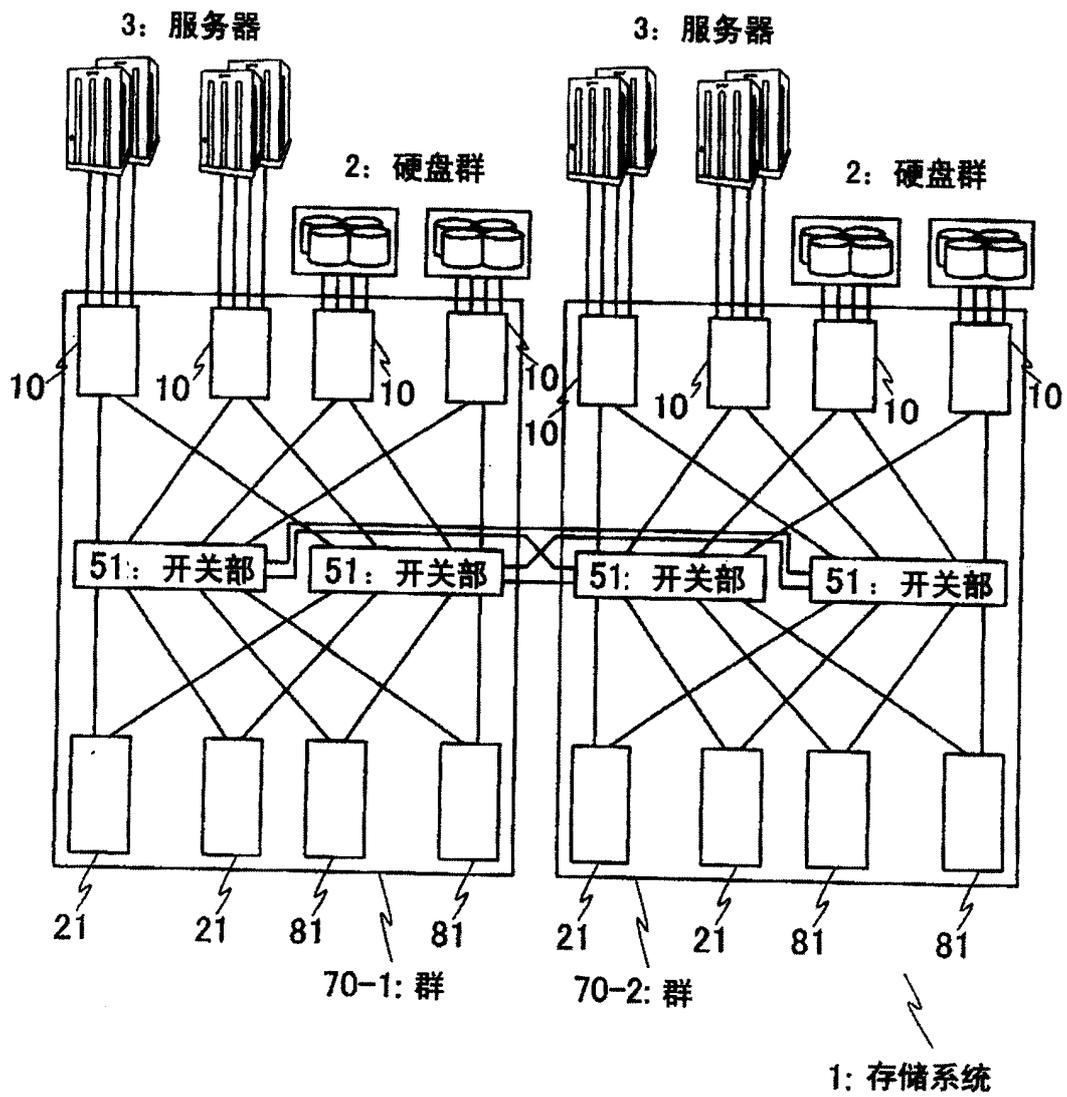


图 6

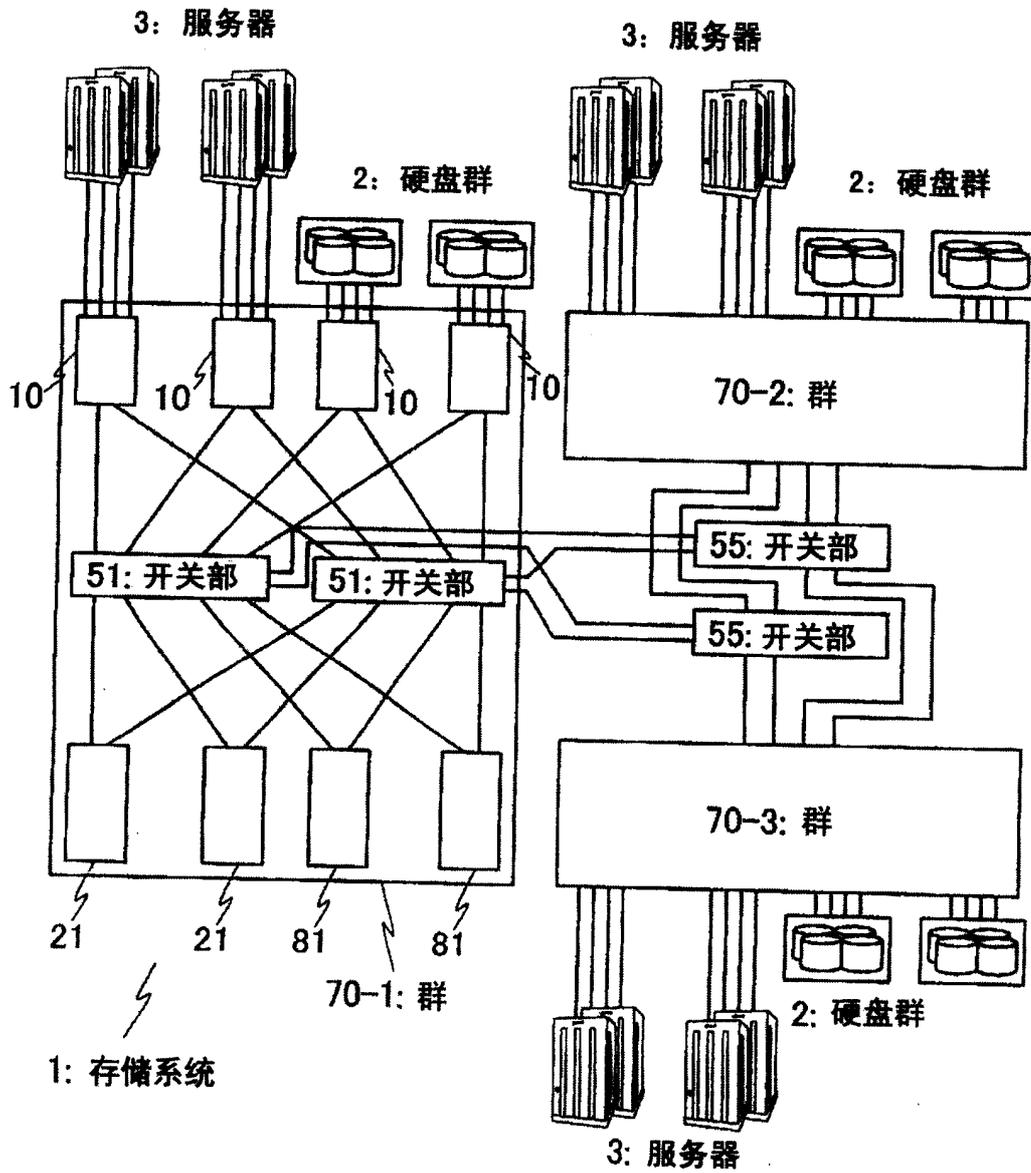


图 7

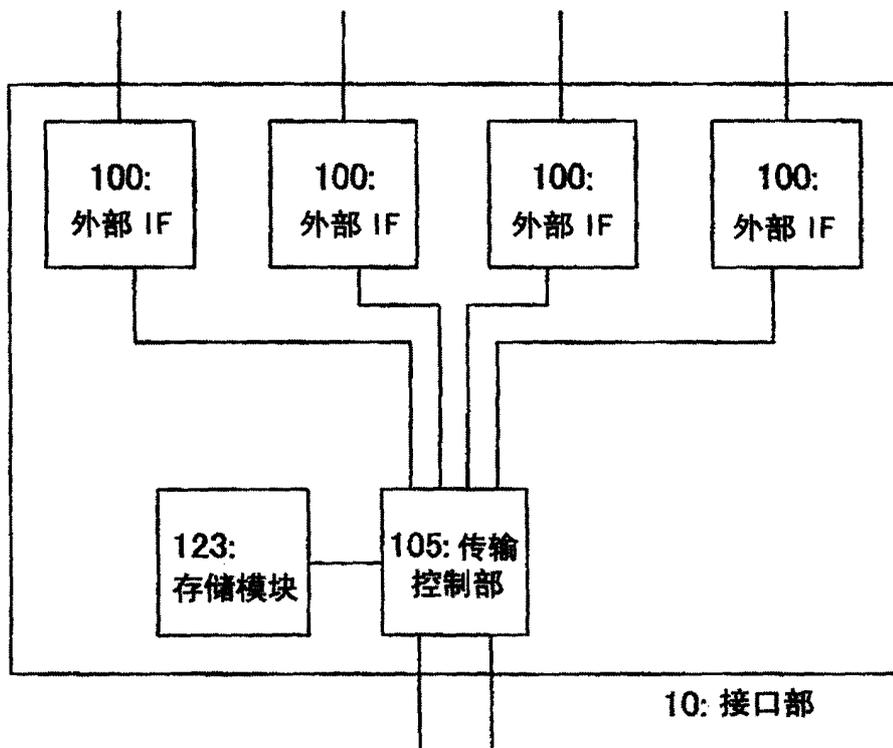


图 8

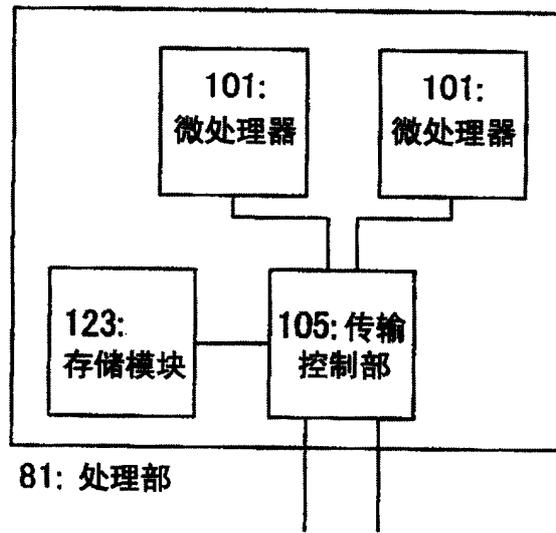


图 9

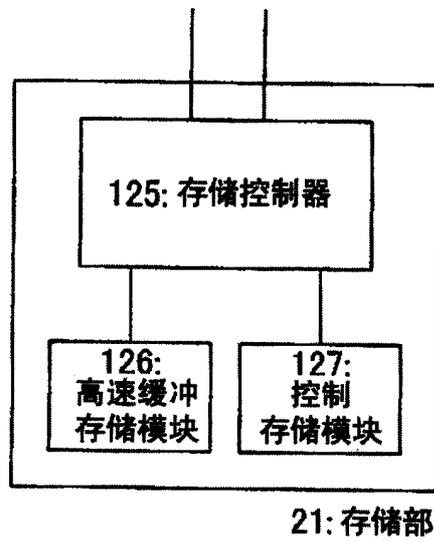


图 10

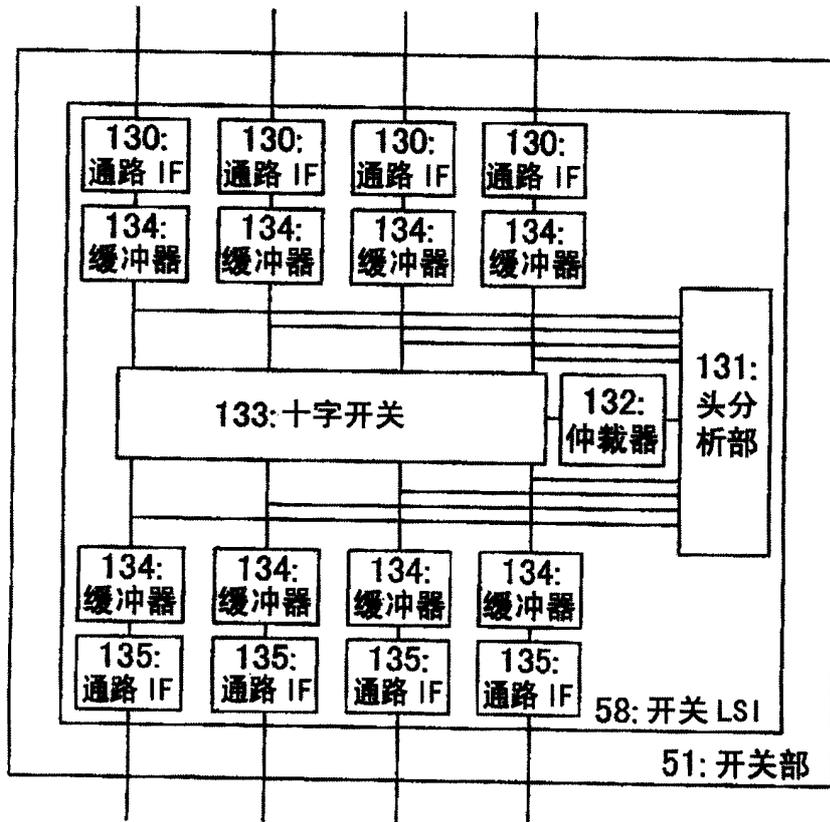


图 11

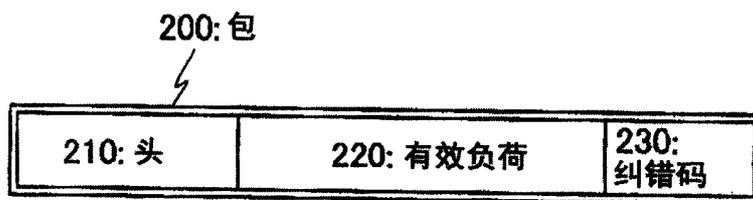


图 12

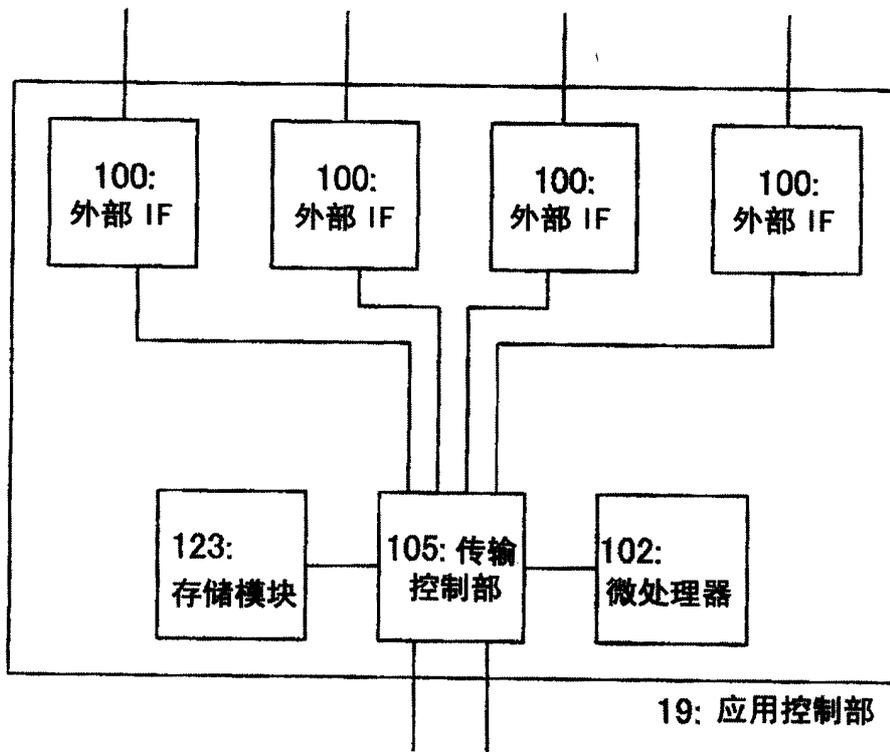


图 13

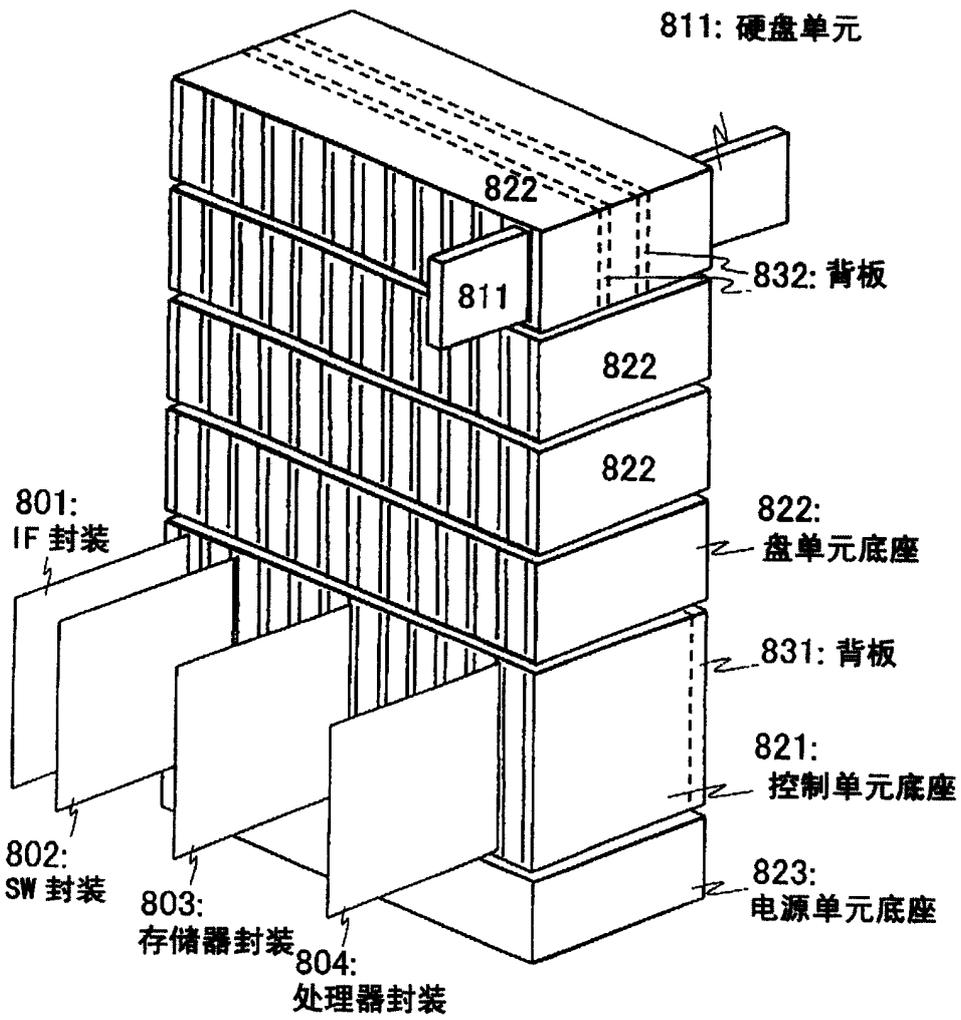


图 14

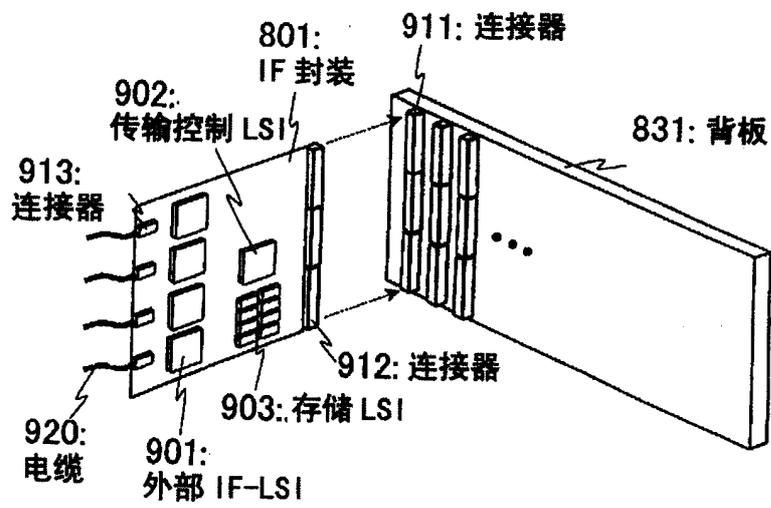


图 15

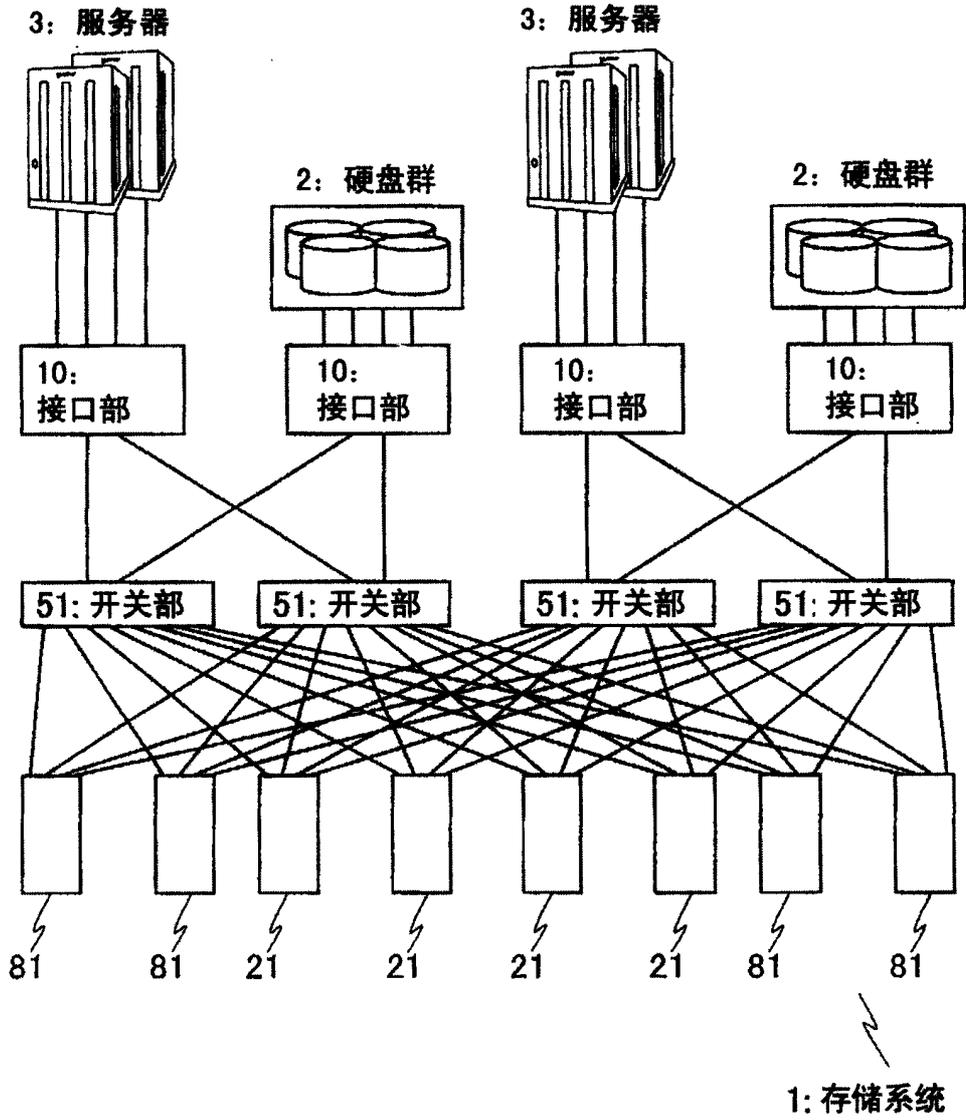


图 16

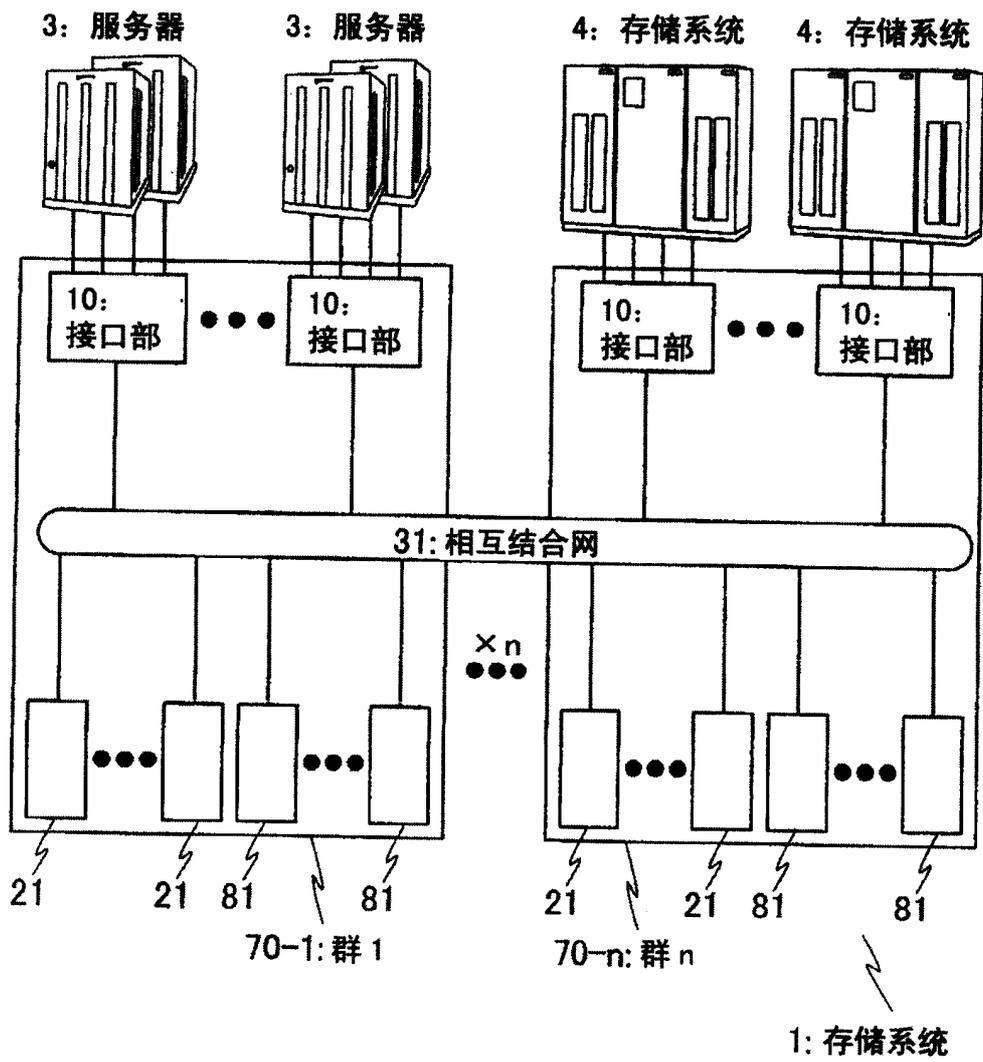


图 17

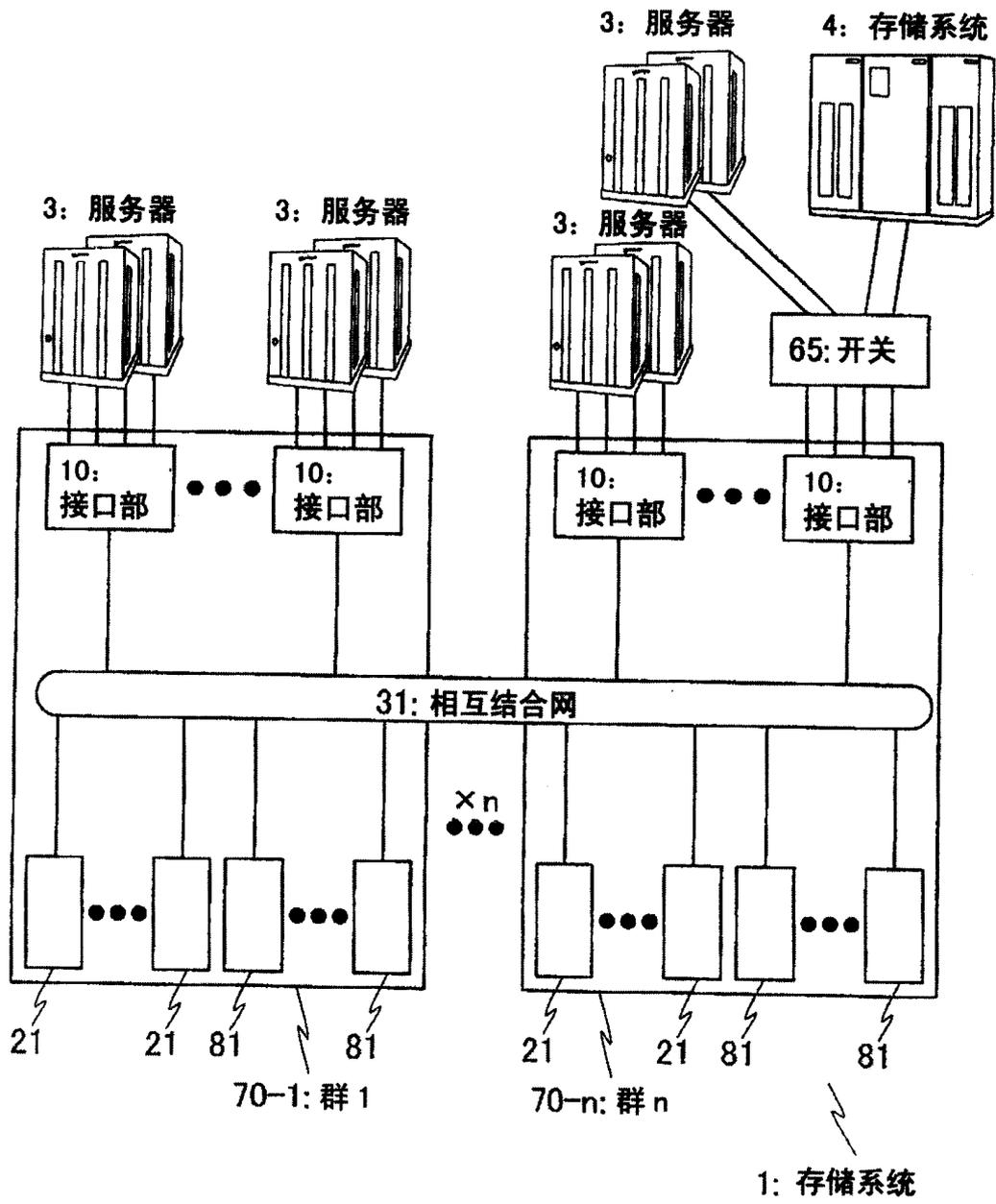


图 18

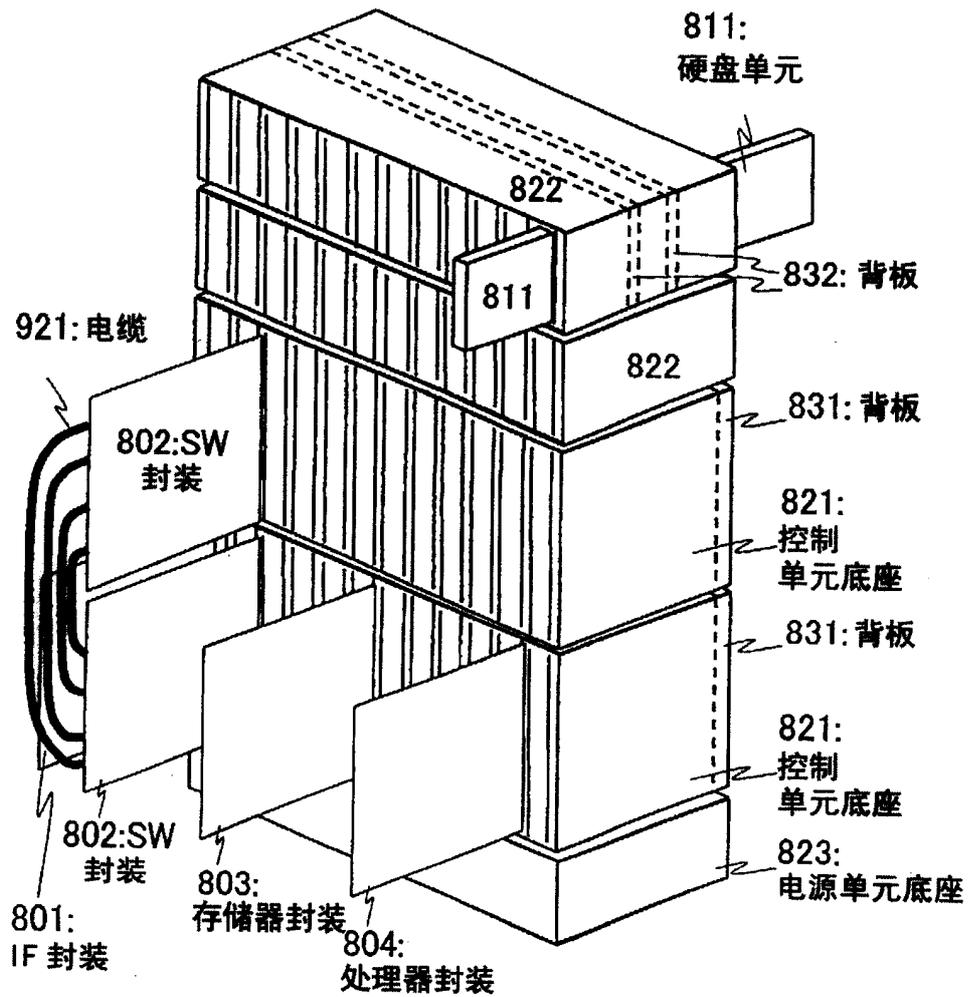


图 19

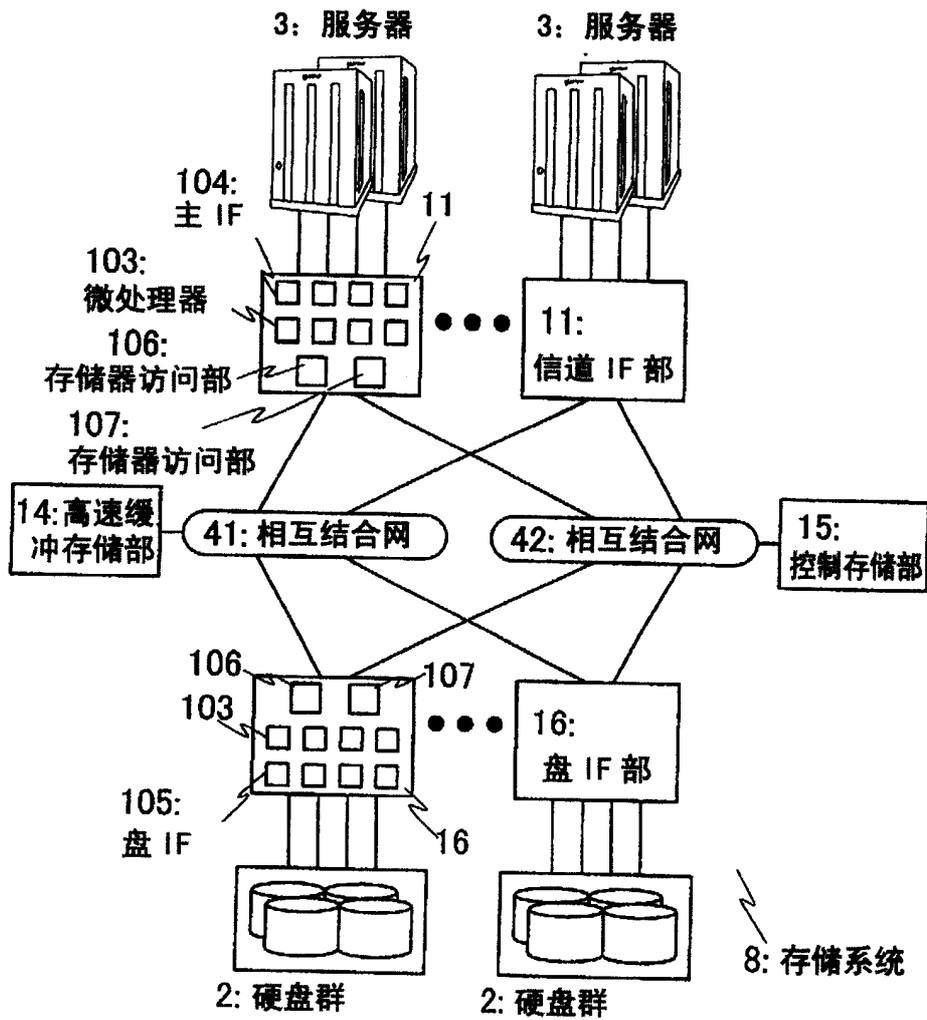


图 20

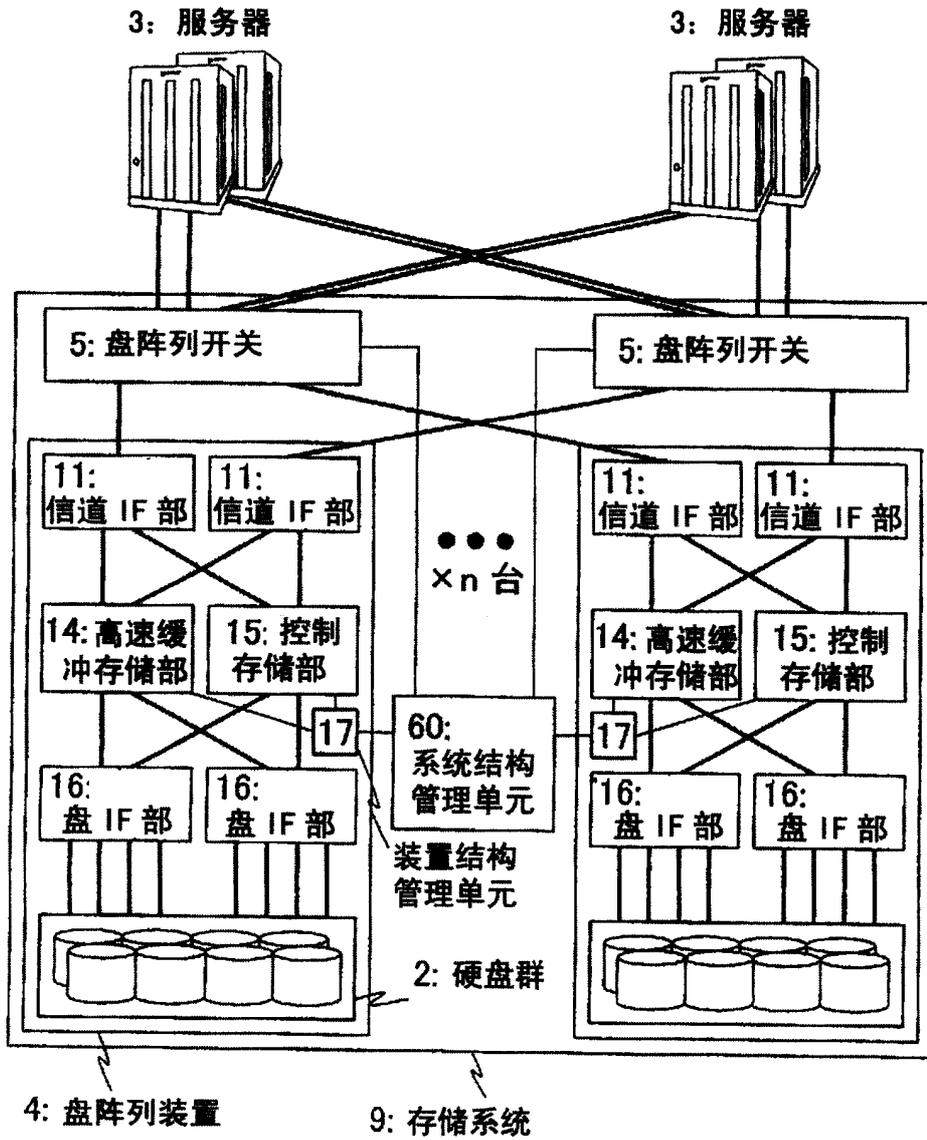


图 21

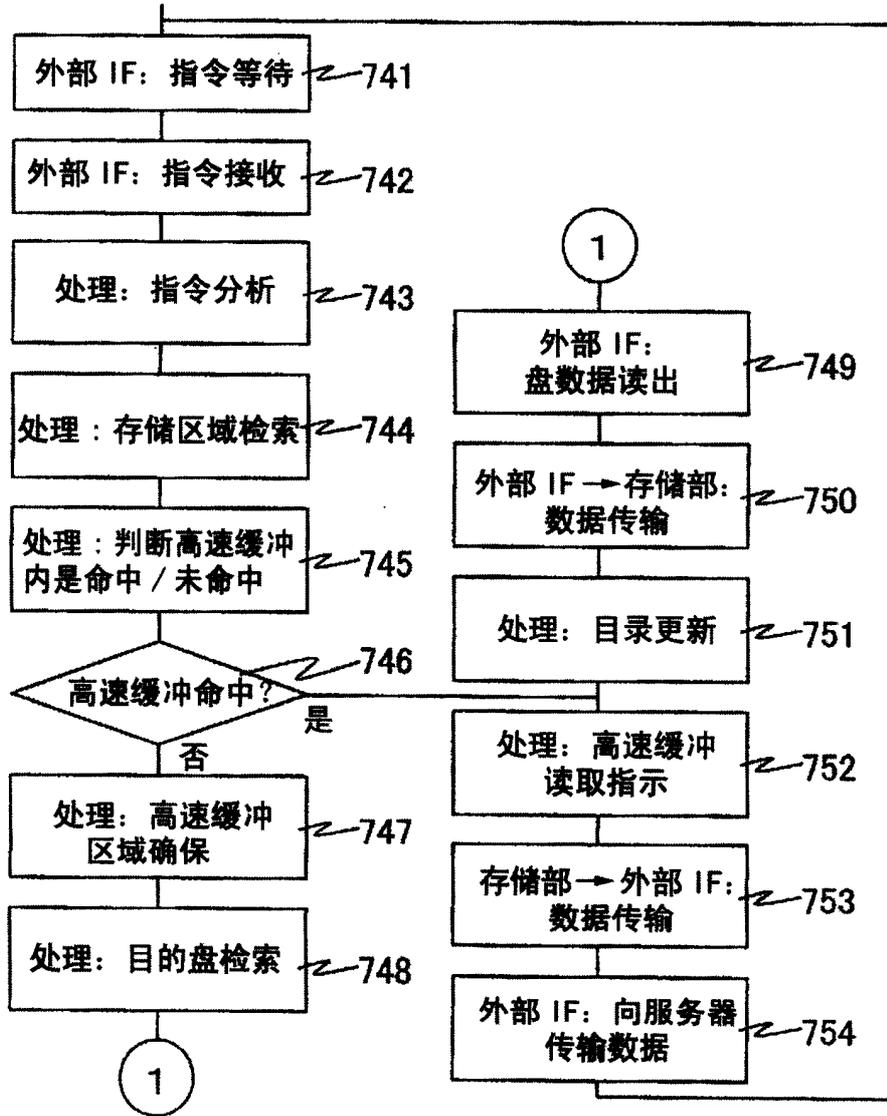


图 22

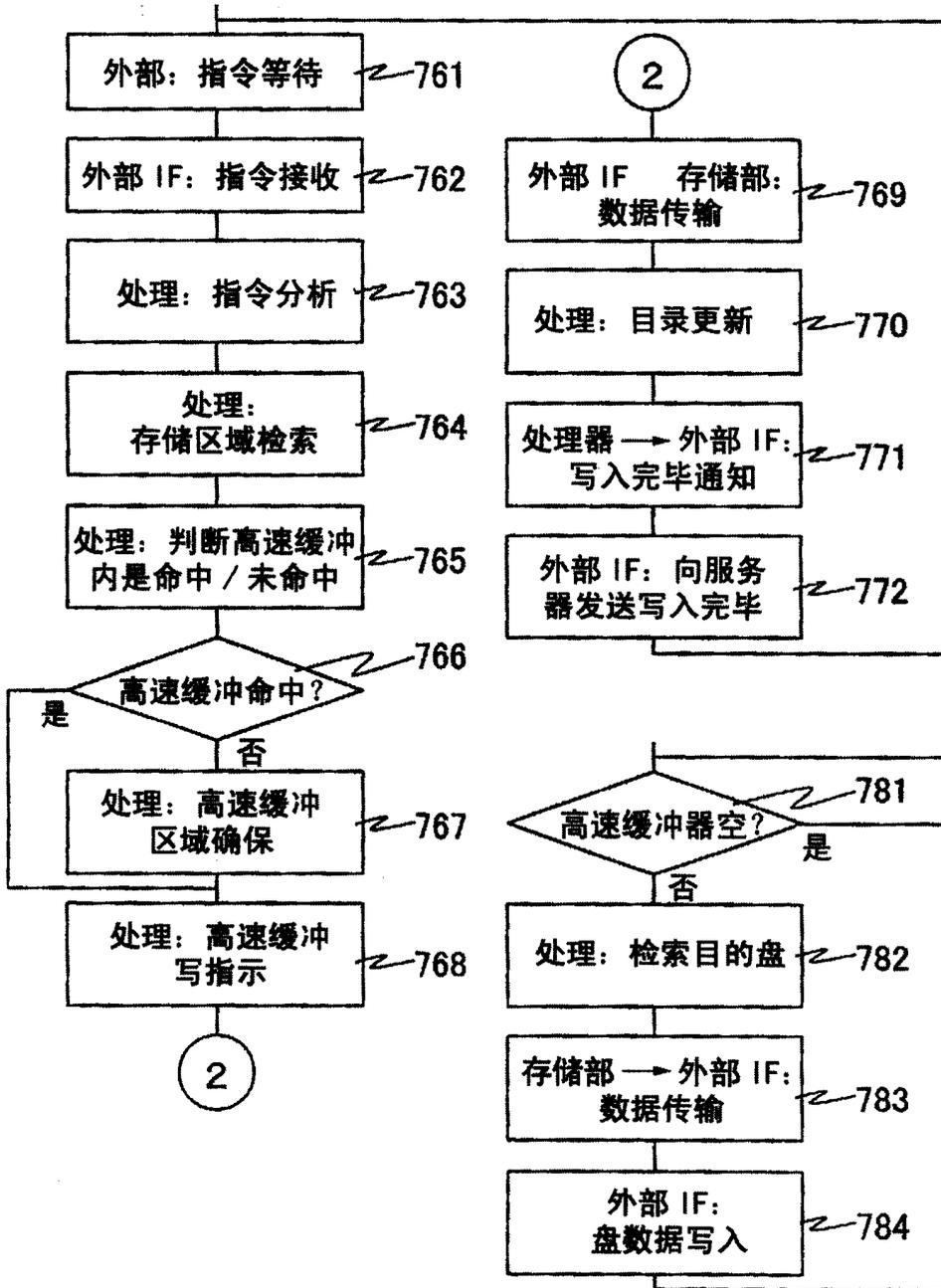


图 23