



[12] 发明专利申请公开说明书

[21] 申请号 01813024.0

[43] 公开日 2005年3月2日

[11] 公开号 CN 1589428A

[22] 申请日 2001.6.11 [21] 申请号 01813024.0

[30] 优先权

[32] 2000.6.12 [33] US [31] 09/592,572

[86] 国际申请 PCT/US2001/040923 2001.6.11

[87] 国际公布 WO2001/097003 英 2001.12.20

[85] 进入国家阶段日期 2003.1.20

[71] 申请人 夸德研究所

地址 美国加利福尼亚州

[72] 发明人 R·德拉康纳

[74] 专利代理机构 北京纪凯知识产权代理有限公司

司

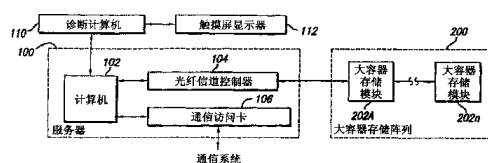
代理人 赵蓉民 彭益群

权利要求书6页 说明书13页 附图9页

[54] 发明名称 信息和应用服务器的高速信息处理和海量存储系统及方法

[57] 摘要

一高速、基于微计算机、兼容光纤通道以及容错信息处理和大容量存储系统，尤其适于信息服务器和应用服务器。一独特且通用的系统结构，包括双回路仲裁的、兼容光纤通道的、容多错的、热可交换的大容量存储磁盘阵列，允许服务器与大容量存储阵列结合以处理大量应用并配置成以备份、速度、处理能力、存储容量等为重要特征的系统。一用于连接系统组件的独特的后平面和/或中平面安排使得在大多数情况下易于在线区域升级和/或服务并同时提供对组件，尤其对那些产生大量热能的诸如磁盘驱动器，有效冷却。



- 1、 一高速大容量存储系统，当系统运行时易于扩展以增加存储容量，所述系统包含至少一个模块，此模块包括：
- (a) 至少一 CPU；
 - (b) 多个用于存储信息的插入式存储设备；
 - 5 (c) 一存储设备旁路电路板，与每个存储设备相连的，每个存储设备插入到旁路电路板上的连接器上；
 - (d) 一模块旁路电路板，其包括一用于将电信号转换为光信号从模块输出并将光信号转换为电信号输入到模块的光学输入/输出连接器；以及
 - 10 (e) 一控制器，通过与存储设备相关的存储设备旁路电路板和模块旁路电路板在所述 CPU 与每个存储设备之间建立通信路径。
- 2、 根据权利要求 1 所述的高速大容量存储系统，其中每个存储设备旁路电路板包括不管存储设备是否存在，可完成 CPU 与其他存储设备旁路电路板和与它们相关存储设备的连接的电路。
- 15 3、 根据权利要求 1 所述的高速大容量存储系统，其中当光输入/输出连接器接收到光信号时，模块旁路电路板从至少一个模块中经由所述光输入/输出连接器输出电信号。
- 4、 根据权利要求 1 所述的高速大容量存储系统，包括第一和第二模块，每个模块包括所述元素 (a) 至 (e)，其中模块的光输入/输出连接器通过个一光纤传输介质相连，这使得信号以光的形式在模块间通信。
- 20 5、 根据权利要求 4 所述的高速大容量存储系统，其中当所述光输入/输出连接器接收到来自第二模块的光信号时，第一模块的模块旁路电路板将电信号从第一模块经由光输入/输出连接器输入到第二模
- 25 块。
- 6、 根据权利要求 1 所述的高速大容量存储系统，其中至少一个

模块包括一用于每个存储设备旁路电路板的存储设备旁路板连接器，每个连接器之间带有一个开口以允许空气在连接器之间并沿旁路电路板和存储设备流动以达到冷却的目的。

7、 根据权利要求 1 所述的高速大容量存储系统，其中存储设备
5 是磁盘驱动器并且存储驱动旁路电路板是具有接收磁盘驱动器的连接器的磁盘驱动旁路电路板。

8、 根据权利要求 7 所述的高速大容量存储系统，其中至少一个
模块包括用于每个磁盘旁路电路板的一磁盘驱动旁路电路板连接器，
每个连接器之间带有一个开口以允许空气在连接器之间流动，并且其
10 中每个驱动旁路电路板是在相对边具有一连接器的相对平的电路板，
其中连接器之一是接收磁盘驱动器的连接器并且另一连接器连接到所
述驱动旁路电路板连接器，所述连接器、旁路电路板和驱动器如此安
排以使得当它们互连时存在一用于从模块外沿每个旁路电路板及其相
关磁盘驱动器而无任何后平面阻碍的空气流通的路径以达到冷却的目
15 的。

9、 根据权利要求 8 所述的高速大容量存储系统，其中模块被装
在一机壳中且安装至少一个风扇以使空气从所述外壳外通过所述旁路
板和驱动器间的空间。

10、 根据权利要求 1 所述的高速大容量存储系统，其中控制器以
20 光纤通道协议操作。

11、 根据权利要求 1 所述的高速大容量存储系统，其中控制器是
一仲裁双通道光纤通道控制器。

12、 根据权利要求 10 所述的高速大容量存储系统，其中每个存
储设备是一磁盘驱动器且其中每个存储设备旁路电路板包括一磁盘驱
25 动旁路电路板，此磁盘驱动旁路电路板包括不管磁盘驱动器是否存在，
可完成 CPU 与其他驱动旁路电路板和与它们相关的磁盘驱动器的连接
的电路。

13、 根据权利要求 12 所述的高速大容量存储系统，其中当所述光输入/输出连接器接收到光信号时，模块旁路电路板将从至少一个模块经由光输入/输出连接器输出电信号。

14、 根据权利要求 11 所述的高速大容量存储系统，其中包括第一和第二模块，每个模块包括所述元素 (a) 至 (e)，其中模块的光输入/输出连接器通过个一光纤传输介质相连，这使得信号以光的形式在模块间通信。

15、 根据权利要求 14 所述的高速大容量存储系统，其中当第二模块通过所述光输入/输出连接器接收光信号时，第一模块的模块旁路电路板从第一模块经由光输入/输出连接器输出电信号至第二模块。

16、 根据权利要求 11 所述的高速大容量存储系统，其中存储设备是磁盘驱动器且存储设备旁路电路板是磁盘驱动旁路电路板，而且其中至少一个模块包括用于每个磁盘驱动旁路电路板的一磁盘驱动旁路电路板连接器，每个连接器之间带有一个开口以允许空气在连接器之间并沿旁路电路板和磁盘驱动器流动以达到冷却的目的。

17、 根据权利要求 12 所述的高速大容量存储系统，其中至少一个模块包括用于每个磁盘旁路电路板的一磁盘驱动旁路电路板连接器，每个连接器之间带有一个开口以允许空气在连接器之间流动，并且其中每个驱动旁路电路板是在相对边具有一连接器的相对平的电路板，其中连接器之一是接收磁盘驱动器的连接器并且另一连接器连接到所述驱动旁路电路板连接器，所述连接器、旁路电路板和驱动器如此安排以使得当它们互连时存在一用于从模块外沿每个旁路电路板及其相关磁盘驱动器而无任何后平面阻碍的空气流通的路径以达到冷却的目的。

18、 根据权利要求 17 所述的高速大容量存储系统，其中模块被装在一机壳中且安装至少一个风扇以使空气从所述外壳外通过所述旁路板和驱动器间的空间。

19、 一高速大容量存储系统，当系统运行时易于扩展以增加存储

容量，所述系统包含至少一个模块，此模块包括：

- (a) 至少一 CPU；
- (b) 多个用于存储信息的插入式存储设备；
- (c) 一磁盘驱动旁路电路板，与每个磁盘驱动器相连，在其中一边缘包括一磁盘驱动连接器并在另一边缘包括一旁路板连接器，每个
5 磁盘驱动器插入到磁盘驱动旁路电路板上的所述连接器上；
- (d) 一模块旁路电路板包括一用于将电信号转换为光信号从模块输出并将光信号转换为电信号输入到模块的光学输入/输出连接器；
以及
- 10 (e) 一控制器，通过相关驱动旁路电路板和模块旁路电路板将 CPU 与每个磁盘驱动器相连，以使得无论磁盘驱动器是否插入到磁盘驱动连接器，在控制器的输出和输入之间形成一回路，每个磁盘驱动旁路电路板和模块旁路电路板在此回路中并完成此回路。

20、 根据权利要求 19 所述的高速大容量存储系统，其中包括一
15 经由光输入/输出连接器连接到所述至少一模块的第二模块，至少一模块的所述模块旁路电路板通过第二模块完成所述回路。

21、 根据权利要求 19 所述的高速大容量存储系统，其中当所述光输入/输出连接器接收光信号时，模块旁路电路板从至少一模块经由光输入/输出连接器输出电信号。

20 22、 根据权利要求 20 所述的高速大容量存储系统，其中每个所述模块包括所述元素 (a) 至 (e)，其中模块的光输入/输出连接器通过个一光纤传输介质相连，这使得信号以光的形式在模块间通信。

23、 根据权利要求 22 所述的高速大容量存储系统，其中当所述光输入/输出连接器接收来自第二模块的光信号时，至少一模块的模块
25 旁路电路板将电信号从至少一模块经由光输入/输出连接器输出到第二模块。

24、 根据权利要求 19 所述的高速大容量存储系统，其中至少一模块包括一用于每个驱动旁路电路板的驱动旁路板连接器，每个连接

器之间带有一个开口以允许空气在连接器之间并沿旁路电路板和磁盘驱动器流动以达到冷却的目的。

25、 根据权利要求 19 所述的高速大容量存储系统，其中至少一个模块包括用于每个驱动器旁路电路板的一驱动旁路电路板连接器，其中每个驱动旁路电路板是在相对边具有一连接器的相对平的电路板，其中连接器之一是接收磁盘驱动器的连接器并且另一连接器连接到所述驱动旁路电路板连接器，所述连接器、旁路电路板和驱动器如此安排以使得当它们互连时存在一用于从模块外沿每个旁路电路板及其相关磁盘驱动器而无任何后平面阻碍的空气流通的路径以达到冷却的目的。

26、 根据权利要求 25 所述的高速大容量存储系统，其中模块被装在一机壳中且安装至少一个风扇以使空气从所述外壳外通过所述旁路板和驱动器间的空间。

27、 根据权利要求 26 所述的高速大容量存储系统，其中每个驱动旁路电路板连接器安装在彼此间隔相同的平面上。

28、 根据权利要求 19 所述的高速大容量存储系统，其中控制器以光纤通道协议操作。

29、 根据权利要求 19 所述的高速大容量存储系统，其中控制器是一仲裁双通道光纤通道系统。

30、 根据权利要求 29 所述的高速大容量存储系统，其中每个驱动旁路电路板包括不管磁盘驱动器是否存在，可完成 CPU 与其他驱动旁路电路板和与它们相关的磁盘驱动器的连接的电路。

31、 根据权利要求 29 所述的高速大容量存储系统，其中当光输入/输出连接器接收到光信号时，模块旁路电路板从至少一模块经由光输入/输出连接器输出电信号。

32、 根据权利要求 29 所述的高速大容量存储系统，其中包括第一和第二模块，每个模块包括所述元素 (a) 至 (e)，其中模块的光输

入/输出连接器通过一光纤传输介质相连，这使得信号以光的形式在模块间通信。

33、 根据权利要求 32 所述的高速大容量存储系统，其中当所述光输入/输出连接器通过所述光输入/输出连接器接收到来自第二模块的光信号，第一模块的模块旁路电路板将光信号从第一模块经由光学
5 输入/输出连接器输出到第二模块。

34、 根据权利要求 29 所述的高速大容量存储系统，其中至少一模块包括一用于每个驱动旁路电路板的驱动旁路电路板连接器，每个连接器之间带有一个开口以允许空气在连接器之间并沿旁路电路板和
10 磁盘驱动器流动以达到冷却的目的。

35、 根据权利要求 29 所述的高速大容量存储系统，其中至少一个模块包括用于每个驱动旁路电路板的一驱动旁路电路板连接器，每个连接器间的开口用于连接器间空气流通，其中每个驱动旁路电路板是在相对边具有一连接器的相对平的电路板，其中连接器之一是接收
15 磁盘驱动器的连接器并且另一连接器连接到所述驱动旁路电路板连接器，所述连接器、旁路电路板和驱动器如此安排以使得当它们互连时存在一用于从模块外沿每个旁路电路板及其相关磁盘驱动器而无任何后平面阻碍的空气流通的路径以达到冷却的目的。

36、 根据权利要求 35 所述的高速大容量存储系统，其中模块被
20 装在一机壳中且安装至少一个风扇以使空气从所述外壳外通过所述旁路板和驱动器之间的空间。

信息和应用服务器的高速信息处理和海量存储系统及方法

相关申请：

本申请涉及基于序列号为 60/077643 的临时专利申请、申请日为
5 1998 年 3 月 10 日的申请号为 NO. PCT/US99/05231、申请人为 Richard
Dellacona、申请日为 1999 年 3 月 10 日的专利申请和申请序列号
09/071282、申请人为 Richard Dellacona、申请日为 1998 年 5 月 1 日的
美国专利申请，以上申请在此引入作为参考。

10

发明背景

技术领域

本发明涉及一高速、基于微计算机、兼容光纤通道以及容错信息
处理和海量存储（mass storage，即大容量存储）系统，尤其适于信息
服务器和应用服务器。特别是，本发明涉及具有独特且通用的系统结
15 构的用于信息处理和存储的方法和装置，其中包括双回路仲裁的、兼
容光纤通道的、容多错的、热可交换的大容量存储磁盘阵列和采用此
磁盘阵列提供组织宽信息或应用服务器系统的装置。

背景技术

20 过去，为提供在组织内能进行高速数据吞吐的信息传递、具有便
于改变的数据存储容量、易于采用已知的工业标准操作软件和硬件配
置附件的大容量存储文件服务器已作了许多努力。然而，由于装备不
兼容以及现有可用计算机和通信硬件性能不支持且组件设备服务失
败，此系统存在许多缺点和问题。这些缺点包括不能适应增长的存储
25 需求来增加存储设备或需要信息服务器完全断电才能替换有缺陷的存
储设备。一些兼容问题包括，例如，各供应商的装备组件间共享信息
瓶颈。以上供参考的 Dellacona 专利申请解决了这些中的一些和其他问
题并在其中描述和要求中提供了独特解决方案。

正如其他问题一样，本发明还解决了供参考的 Dellacona 专利申请

中讨论的一些问题。例如，本发明还解决了用于不同应用的信息处理和存储系统中规模可伸缩性和用户化的问题。在一些应用中对处理能力可能要比存储容量有更大的需求，而其他应用的需求可能恰恰相反。其他应用可能需要对现有信息处理系统进行存储扩容。本发明提供了
5 易于适应以上需要的结构。

而且，大容量存储系统会产生大量热量，尤其当它们基于磁盘驱动器时。如果所产生的热量未能有效转移，则会影响系统可靠性和寿命。通常，后平面和中平面的物理配置阻碍了空气流通导致热量难以转移。典型地，例如，大容量存储模块或阵列的所有磁盘驱动器插到
10 整个模块扩展的后平面或中平面表面上的连接器上。无论是在机壳或其他外壳还是无需外壳安装的机架所包围处，此种结构安排抑制了模块内的空气流动，产生了过多热量，尤其是在磁盘驱动器的附近。

而且，为适应对更多存储容量的需求，人们希望大容量存储模块的磁盘驱动器能单独进行热交换，但是对系统存储需求的增长要大于
15 模块存储容量，因此也希望系统无需断电便可增加模块容量。当然，当模块增加时，以上能力要在不干扰现有模块的操作并且信号降低最小的前提下提供。

发明内容

20 本发明通过提供信息处理和大容量存储方法和系统克服了一个或更多前述现有技术的问题和/或缺点，包括独特的大容量存储阵列，尤其适合信息服务器或应用服务器，采用允许在不中断或不严重降低系统操作时增加和替换存储设备并且是高容错和可靠的新系统结构。而且，本发明通过提供可将热量从包含热量产生组件的系统中有效转移
25 的新物理布局克服了一个或更多前述问题。

依据本发明的一个实施例，一信息处理和大容量系统，当系统运行时易于扩展以增加存储容量，包含（a）至少一个计算机，（b）多个诸如用于存储信息的磁盘驱动器的插入式存储设备，（c）与每个存储设备相连的一存储设备旁路（bypass）电路板，每个存储设备插入到
30 旁路电路板上的连接器上，（d）一模块旁路电路板包括一用于将电信号转换为光信号从模块输出并将光信号转换为电信号输入到模块的光学

输入/输出连接器以及(e)一控制器,通过与存储设备相连的存储设备旁路电路板和模块旁路电路板将计算机与每个存储设备相连。

依据本发明的某信息服务器配置包括一个或多个计算机,每个计算机连接到一光纤通道控制器并通过它与一大容量存储阵列通信,此
5 大容量存储阵列包括多个旁路电路板,至少其中一些与信息存储设备相连。在一个实施例中,控制器提供一包括两个通往每个旁路电路板和相连存储设备的完整通信路径的双回路通信信道。在另一实施例中,控制器提供一条两次遍历旁路电路板和任何相关存储设备的单回路。在一带有两个或更多计算机的配置中,与每个计算机相连的光纤通道
10 控制器通过光纤通道控制器旁路卡与大容量存储阵列通信。

在最佳实施例中,计算机最好为一适合的传统单板计算机。控制器最好为一传统的仲裁双信道光纤通道系统,计算机通过此系统与每个存储设备旁路电路板和模块旁路电路板通信。旁路电路板可为任何适合的电路,无论一磁盘驱动器是否插入到驱动旁路电路板,此电路
15 形成一用于光纤通道控制器的连续回路。回路连续通过连接到模块旁路电路的其他模块从而当保持单个信息处理和大容量存储系统时易于扩展。

依据本发明的另一实施例,一高速信息处理和大容量存储系统包括两个模块,每个模块包括多个在一热可交换磁盘驱动阵列中的磁盘
20 驱动器。每个磁盘驱动阵列与一包括一光学输入输出连接器,最好一光电收发器的模块旁路电路相连。模块的光输入/输出连接器通过光纤光学传输媒质互连使得模块间信号以光的形式进行通信。在此配置下,可在不中断彼此操作并不严重降低信号的情况下增加模块以增加存储容量。

依据本发明的还一实施例,一带有磁盘驱动器用于信息存储的高
25 速信息处理和/或大容量存储系统包括至少一个带有多个驱动旁路电路板的模块,每个驱动旁路电路板包括一驱动旁路电路板连接器。连接器间提供至少一个开口以使连接器间的空气流通。每个驱动旁路电路板是在不同边缘带有连接器的相对平的电路板,连接器之一是接收磁
30 盘驱动的连接器和另一连接器与上述驱动旁路电路板连接器相连。连接器、旁路电路板和驱动器安排成当它们相连时从模块外沿着每个

旁路电路板和相连的磁盘驱动器形成无任何后平面阻碍的用于冷却的空气流通路径。在包括大容量存储系统的外壳上安装了至少一个风扇以使空气从所述外壳外通过所述旁路板和驱动器间的空间，最好通过驱动旁路电路板连接器间的开口流动。

- 5 本发明提供了一种新的尤其适用于信息和应用服务器的高速信息处理和/或大容量存储系统，新的系统结构和物理布局使得此系统为规模可伸缩的、容错的且可靠的。本发明的其他特征和优点将在以下对优选实施例结合附图的详细描述中更为明显，其中实施例示例性地描述了发明原理。

10

附图说明

图 1 是一描述依据本发明采用本发明的高速大容量存储系统的一信息或应用服务系统的功能块图；

- 15 图 2A、2B 和 2C 是描述用于对处理和存储需求比率不同的应用的各种系统配置的图示；

图 3 是一更详细描述如图 1 的大容量存储模块的功能块图；

图 4A 是描述从如图 3 的光纤通道控制器开始并与之相连的双回路通信路径的功能块图，用以提供一遍历每个存储设备旁路电路板两次的逻辑光纤通道通信路径；

- 20 图 4B 是描述从如图 3 的光纤通道控制器开始并与之相连的双回路通信路径的功能块图，用以提供两个遍历每个存储设备旁路电路板一次从而可独立有效地与每个存储设备通信的逻辑光纤通道通信路径；

- 25 图 5 是一描述依据本发明的信息或应用服务器系统的功能块图，其中两个或更多服务器连接到一可包括一个或多个大容量存储模块的大容量存储阵列；

图 6 是一描述依据本发明的结构配置的网络服务器应用的实施例的功能块图；

图 7 是一描述依据本发明的结构配置的基本视频流应用的实施例的功能块图；

- 30 图 8 是一描述依据本发明的结构配置并包括一双索引服务器的另一视频流应用的实施例的功能块图；

图 9 是一描述依据本发明的结构配置并包括一分布式索引服务器的另一视频流应用的实施例的功能块图；

图 10 是一描述依据本发明的存储设备旁路电路板的一实施例的功能块图；

5 图 11 是一描述依据本发明的底板旁路电路板的一实施例的功能块图；

图 12A 和 12B 是描述依据本发明最佳连接器安排的一驱动旁路电路板和磁盘驱动器的物理布局的图示；以及

10 图 13A 和 13B 是依据本发明的用于将一单板计算机（SBC）连接到各输入/输出设备的连接器安排的图示。

具体实施方式

图 1 描述了依据本发明采用本发明的高速大容量存储系统的一信息或应用服务系统的实施例。如图 1，服务器，附图标记为 100，包括
15 一计算机 102、一控制器 104，最好一光纤通道控制器以及一通信接口或访问卡 106。服务器 100 通过光纤通道控制器 104 与包括一个或更多大容量存储模块 202A...202n 的大容量存储阵列 200 通信。作为例子，计算机 102 也可如前面提到的 Dellacona 申请所述与适合的诊断计算机 110 通信。

20 计算机 102 最好包括运行诸如微软公司认证的 Windows NT 的计算机工业标准操作系统软件程序的单板高速计算机。假如想要保留存储器或加快操作速度，诸如 Windows NT 的操作系统可拆开以删除那些不需要的程序元素。适合用于类似应用的传统驱动器必须支持要实现的特殊结构。

25 计算机可包括一诸如触摸屏的显示器、以及所需的各种存储和外围设备（未示出）。单板计算机可包括任意大量适合的设备，其中包括但不局限于 Ziatech 公司认证的型号为 ZT5510 的带有奔腾处理器的压缩 PCI CPU 板。为提高 ZT5510 性能的改进包括一用于永久存储 Windows NT 操作系统软件不可重构部分的板上 40MB 闪存卡以及
30 Sandisk 公司认证的可读写存储 Windows NT 操作系统软件可重构部分的板上、可拆卸的 PCMCIA40Mb 闪存卡“D2 FlashDisk”。

光纤通道控制器 104 可依据光纤通道协定设计为一分离板或合并到单板计算机中。通信接口或访问卡 106 可为依据已知 T-1 通信体系结构和/或兼容其他网络和通信体系结构、协议和拓扑结构的、包括但不局限于 T-3、DS-3、OC-3C、OC-12C、OC-192C、FDDI、SONET、SCSI、TCP/IP、HiPPI 和 ATM 的体系结构制造的任何适合的设备。而且，计算机 102 和 110 可通过网络相连并通过适合的以太网卡或其他网络技术与其他计算机相连。相应的制造商、光纤通道协定和详细描述每个前述组件操作能力和规格的 120 专用业务组参考设计数据表和材料共同作为参考。而且，更多包含可能的子系统和用于服务器的连接协议的更多信息在供参考的 Dellacona 申请中被描述。

图 2A、2B 和 2C 描述了能依据本发明的体系结构寻址的三种系统配置。每图给出了带有一电源供应区 302 和一诊断区 304 的底板 300。而且，底板包括诸如图 1 中的一个或更多服务器 100 的一处理区 306 和/或诸如图 1 中的一个或更多大容量存储模块 202 的一存储区 308。处理和存储区的大小按特殊应用来配置。

例如，图 2A 描述了用于诸如一应用服务器、一电影服务器（例如，用于多用户视频流的）或结合一载体类开关通信的大容量服务器的安排。处理区 306 包括 10 个插槽，每个最好代表一单板计算机。相反，存储区 308 只有 5 个存储插槽，每个代表诸如磁盘驱动器的一高速、大容量存储设备。因此，处理功能比存储功能强。另一方面，图 2B 描述了适用于网络服务器或网络主机的安排。此处，处理区 306 包括两个底板插槽而存储区 308 包括 15 个插槽。图 2C 描述了主要适用于存储扩展的底板安排。此处，无处理区且底板专用作存储器。

图 3 更详细描述了如图 1 的大容量存储阵列 200 配置为一单个计算机服务器 100 和一个或更多大容量存储模块 202 的实施例。参照图 1，大容量存储阵列 200 包括一个或更多模块 202A、202B...202n，每个模块最好结构相同，尽管它们可能包含不同数量存储设备和不同种类存储设备。每个模块 202 包括旁路电路板 210A、210B...210n 和至少一些与板 210 相连的存储设备 212A、212B...212n。而且，每个存储设备最好分别与读/写开关 214A、214B...214n 相连。

每个模块 202 在光纤通道控制器 104 的通信路径上具有一模块或

底板旁路电路板 216。模块 202A 的底板旁路电路板 216 带有用于从模块 202A 将电信号转换为光信号输出并将光信号作为电信号输入到模块 202A 的光输入/输出连接器 218。同样，模块 202B...202n 具有与输入/输出连接器 218 相连的底板旁路电路板 216（未示出）。可知，模块 202A 的底板旁路电路板的输入/输出连接器 218 经由诸如光纤 220 的光传输媒质与模块 202B 的底板旁路电路板的光输入/输出连接器 218（未示出）相连。

如前所述，控制器 104 最好是一以光纤通道协议操作的传统光纤通道控制器（FCC）且最好是一仲裁双通道光纤通道控制器。控制器 104 在每个模块 202 内计算机 102 和每个可操作的存储设备 212 之间提供了一双通道通信路径。如以下更详细地描述，旁路电路板 210 确保即使存储设备 212 不工作（即，如下更详细地描述的在一些最低水平或之上不工作）或已被从旁路电路板上的连接器上移开时通信路径也是完整的。

在那一点上，每个存储设备 212 最好是高速的、大容量、通常有效并以可拆卸的方式与相关的旁路电路板 210 相连的磁盘驱动器。最好，磁盘驱动器插入到旁路电路板 210 的连接器上使得它易于拆卸和替换或当需要扩展模块存储容量时可在旁路电路板空的位置增加驱动器。当插入并将磁盘驱动器传送到在某一最低水平可操作的旁路电路板时，每个旁路电路板 210 包括将控制器 104 连接到磁盘驱动器 212 的电路。另一方面，当未插入或未在最低水平或之上操作磁盘驱动器时，旁路电路板 210 绕过磁盘驱动器 212 将控制器 104 直接连接到下一旁路电路板。无论如何，运行自诊断并提供错误/无错输出信号的传统磁盘驱动器可用于此目的。

参照图 3，例如，当一可操作的存储设备 212A 与旁路电路板 210A 相连时，旁路电路板 210A 直接与相关存储设备 212A 通信然后再与旁路电路板 210B 通信。当存储设备 212A 不工作，即在一些最低满意水平不操作时，或当无存储设备连接到旁路电路板 210A，即存储设备被替换拆卸时，双通道通信路径不中断通过旁路电路板 210A，即，绕过存储设备连接器。在此后将详细描述旁路电路板完成前述连接和旁路功能的一个实施例，本领域普通技术人员可采用任何适合的常规方式

通过受来自存储设备的错误/无错信号控制的电子开关电路完成这些功能。

再参照图 3, 模块或底板旁路电路板 216 提供与存储或驱动旁路电路板 210 类似的功能, 因此假如无附加模块 202B 连接到模块 202A 则路由直接返回控制器或者假如有附加模块且信号被接收则路由与下一模块 202B 通信。因此, 在单个模块 202 的情况下, 计算机 102 通过模块内一双通道通信回路与每个磁盘驱动器 212 通信。另一方面, 由于底板旁路电路板 216, 当一个模块被连接时双通道通信回路扩展到在下一模块中的每个磁盘驱动器 212。尤其, 假如有附加模块 202B 与模块 202A 相连时, 模块 202A 的计算机 102 直接通过模块旁路电路板 216 经由连接器 218 和光纤 220 与下一模块 202B 通信, 这样模块 202B 包括在双光纤通道通信回路中并被遍历。

由以上所述体系结构可推知, 大容量存储阵列 200 的单个大容量存储设备模块 202 可通过增加磁盘驱动器和其他适合的存储设备来内部扩展, 而且坏磁盘驱动器在不影响其余模块或使用其的系统操作的情况下可替换。这提供了一通用热可交换、热扩展、大容量存储设备阵列。而且, 当系统命令超过单个模块容量时, 无需中断阵列或系统的其余部分便可再增加一附加模块。

如上所述, 光纤通道控制器提供了一通过大容量存储阵列 200 的每个模块的双路径。依据本发明, 系统可配置成双路径为如图 4A 和 4B 的遍历大容量存储阵列两次的一条单路径或两条独立路径。

现在参照图 4A 和 4B, 光纤通道控制器 (FCC) 104 包括两条输出路径 A 和 B。每条路径通过与每个现有可操作的存储设备 212 通信并返回 FCC 来遍历大容量存储模块 202。在图 4A 的实施例中, 返回 FCC 的 A 路径回送到 B 输出路径。因此, 一单连续路径遍历了大容量存储模块两次。在图 4B 的实施例中, A 返回路径未回送到 B 输出路径。因此, 两条路径遍历了大容量存储模块并可独立使用。之后的实施例将在一条路径失败时提供第二条路径。

本领域普通技术人员可知图 4A 和 4B 的实施例提供了备份和容错。图 4A 的实施例只需要一套芯片来提供单回路所需的单光纤通道, 因此实现更为简单。但是, 假如回路之一被破坏或遇到一些其他错误,

且另一回路仍然有效，则回路可能在控制器内被绕过。同样，要实现图 4B 的实施例将更复杂且花费更高，但它提供了两条容错和备份的独立寻址回路，还提供了更大的有效通信带宽。

图 5 表示本发明的系统结构可能制造的更详细的配置。在如图 5 的实施例中，两个服务器连接到一大容量存储阵列，其中一第一模块被表示出来。服务器 100A 和 100B 分别与光纤通道旁路板 222A 和 222B 相连。光纤通道旁路板，例如通过检测底板旁路板的信号流，来检测计算机 102A 和 102B 的操作，这使得计算机 102A 和 102B 可协同工作或在一错误的情况下分别工作。正如前面提到的，每个服务器具有闪存，而且在如图 5 的实施例中，每个计算机 102A 和 102B 具有自己的存储操作系统的闪存。这样，计算机可从与其相关的闪存而不是共享存储器或其他磁盘安排中独立启动。

如图 5 的系统配置提供两个服务器和因此增加的处理电源。而且，一服务器为另一服务器在失败的情况下提供备份，从而提供增加的容错和可靠性。而且，如图 5 的体系结构允许两个服务器彼此以光纤通道速度，比采用 IP 协议的 100BaseT 局域网所能达到的更高的速度，通信。

依据本发明的系统结构本身可进行多种配置以满足多种应用。图 6 表示一配置，其中两个服务器 A 和 B 通过与适合的局域网 (LAN) 相连来访问 Internet。服务器 A 和 B 均具有它自己的大容量存储阵列 200。网络连接的几个服务器和存储阵列可提供一功能非常强的尤其适用于网络服务器应用的信息或应用服务器系统。而且，此配置允许计算机/大容量存储系统均加倍而不是像如图 5 的实施例中一样共享一大容量存储系统。而且，如图 6 的实施例所描述的 LAN，典型地为 100BaseT LAN，假如需要 GB 范围的速率时可采用光纤通道局域网。这些特征和优点是用于诸如网络服务器的关键处理应用的理想选择。

图 7 表示尤其用于视频流应用的另一配置。在此实施例中，诊断计算机 110 也有用于控制访问内容服务器 1、2 和 3 的索引功能。内容服务器 100 提供对存储在与其相关的存储阵列 200 中视频的访问，可选地，如图 8 和 9 所示也尤其适合视频流应用。在图 8 的实施例中，复制的索引服务器与诊断计算机分离并具有自己的存储阵列。在此实

5 施例中，有两个索引服务器 1 和 2 且有三个内容服务器 1、2 和 3，每个均带有一相关的存储阵列。在图 9 的实施例中，索引服务器的功能被分布在内容服务器中这使得存在多个索引/内容服务器 1-5。本领域普通技术人员可知如图 8 的方法采用了索引服务器或与来自内容服务器内容流一致的服务器的体系结构，而如图 9 索引和内容服务器功能被分布在所有服务器中，提供了大范围的规模伸缩性。

10 图 10 和 11 是更详细地分别表示存储和底板旁路电路板 210 和 216 的功能块图。参照图 10，存储旁路电路板 210 包括一用于插入到附图标记为 232 的后平面上的连接器的旁路板后平面连接器 230。A 和 B 信号路径从前一旁路电路或直接从光纤通道控制器通过后平面经由后平面连接器 230 至旁路电路。同样，

15 从后平面连接器 230 开始的 A 信号路径连接到一适合的常规电子开关 234。B 信号路径类似地连接到一电子开关 236。A 和 B 信号路径从电子开关 234 和 236 连接到一旁路板存储卡或驱动连接器 238 并通到存储设备（例如，一磁盘驱动器）212。

20 从旁路板驱动连接器 238 返回的 A 信号路径连接到开关 234，而从旁路板驱动连接器 238 返回的 B 信号路径连接到开关 236。如上所述的存储设备产生的指示是否存在可操作性以及可操作性水平的错误信号控制电子开关 234 和 236。来自开关 234 和 236 的 A 和 B 返回路径连接到旁路板后平面连接器 230 并通过后平面 232 至下一旁路电路板或光纤通道控制器。

25 在操作过程中，A 信号路径进入旁路电路板并与开关 234 相连。假如不存在表示存储设备不存在或不可操作的错误信号（即，无错误且信号在一低或负信号状态），开关 234 返回 A 信号路径至旁路板后平面连接器 230 以绕开存储设备 212。假如错误信号为低，B 信号路径同样通过电子开关 236 循环回后平面连接器 230。另一方面，当错误信号为高或正表示存储设备存在且可操作，A 和 B 信号路径通过开关 234 和 236 连接到存储设备然后返回通过开关。

30 现在参照图 11，底板旁路电路板与存储旁路电路板基本相同，除了电子开关 234 和 236 所作的选择是作为一旁路还是将 A 和 B 连接到

I/O 连接器 238。从这一点上，I/O 连接器最好是一用于多模式光纤的双向通信应用，尤其用于多模式或单模式光纤通道应用的常规光纤收发器。例如，收发器可为芝加哥伊利诺斯州 Methode Electronic 公司认证的型号为 MLC-25-6-X-T 的光纤通道小系数 (SFF) 收发器。此收发器包括一光传送器和接收器即用于接收一工业标准光纤连接器的接收器。而且，接收器提供了一信号检测输出 (图 11 的错误信号) 指示接收器是否收到一光学信号。假如未收到，错误信号控制电子开关 234A 和 236A 将 A 和 B 路径返回至底板旁路板连接器 230A。假如，另一方面，错误信号指示收发器 218 已结束到一光信号，A 和 B 路径上的电信号则通过开关 234A 和 234B 到达收发器 218 并转换成光信号经由形成 A 和 B 路径的光纤传送至下一底板。同样，从下一底板返回的 A 和 B 路径上的光信号被收发器 218 转换为电信号并沿 A 和 B 路径通过开关 234A 和 236A 返回至连接器 230A 和下一旁路电路或光纤通道控制器。

典型地，如图 3 所示的大容量存储设备阵列的每个模块 202 均密封在一外壳内或假如安装在一支架内，则可采用其他构件和/或其他电路板包围。诸如计算机 102 和 110 的服务器组件、触摸屏显示器 112、光纤通道控制器以及诸如通信访问卡、以太网卡、LAN 组件等其他系统组件也可安装在相同外壳内或机架上。热量因此成为一个问题，尤其用在模块内的存储设备是由马达驱动的磁盘驱动器的情况下。因此，如图 12A 和 12B 的本发明最佳实施例包括不同旁路电路板和存储设备的物理结构和安排，适应于无最少障碍的整个模块内冷空气循环。

依据本发明此物理结构的一方面，每个驱动旁路电路板 210 是相对薄的电路板。然而电路板不像典型地用于接收一磁盘驱动或其他大容量存储设备的电路板。在此常规电路板内，接收插入式磁盘驱动器的连接器或插头被标准地定位以使得磁盘驱动器垂直于板平面扩展。例如，通常在它们之间在模块上扩展的后平面或中平面电路板上具有旁路电路和/或通信路径，这类似于扩展在计算机底板上的计算机主板，还具有用来接受各种插入卡或板的连接器。此安排使得诸如磁盘驱动器的存储设备产生的热量更难有效散去。

如图 12A 和 12B 所示，本发明未采用电路板中平面或是后平面结

构。而是，接受磁盘驱动器（例如，图 10 中的连接器 238）的驱动旁路电路板上的连接器在板边缘，这使得当插入驱动器时平行扩展了电路板。而且，驱动旁路电路板本身不插入一后平面电路板。而是，在板边缘最好与图 10 中磁盘驱动连接器 239 相反而提供一连接器，而且连接器插入一单个连接器中，此单个连接器安装在一框架或底板的其他结构成员上且是有线的或与其它磁盘旁路电路板的类似连接器连接。

图 12A 是磁盘驱动器 212 经由连接器 238 插入到驱动旁路电路板 210 的侧视图，其中磁盘旁路电路板插入安装在包含大容量存储模块和/或服务器及其相关元件的底板的结构成员或支架上的连接器 240。图 12B 是表示几个旁路电路板 210 依次通过螺丝钉或其他适合的方式与结构成员 242 相连从而插入连接器 240 的端视图。由于无通常用来连接相邻组件的后平面，在连接器 240 间提供的电子或光连接 246 提供如图 3 和 5 所示的所需通信。

在每个连接器 240 之间存在空隙 244 使得空气易于被风扇或其他必须的空气循环装置吹开通风。即使结构成员是包围组件的外壳的一部分，所提供的屏蔽或其他适合的开口使得在区域 244 有效空气流动。由于磁盘被插入到旁路电路板，它们的平面平行而不垂直，对空气流通无阻碍。

此安排尤其适合单元区域服务并易于更新。无需改变整个后平面或中平面板便可替换区域内连接器，且此修改只需花费很少的时间而且无需断电。而且，不局限于采用后平面或中平面时遇到的扩展插头数量和其他物理和电子特性，可扩展所述连接安排。

相同的物理安排可用于连接计算机和与其相关的组件以产生想要的服务器配置。例如，如图 13A 所述，之前描述的与服务器 100 有关的一单板计算机可通过一小型压缩 PCI (CPCI) 后平面与通信卡 106 相连。在此情况下，单板计算机 SBC 可与光纤通道控制器 FCC 结合。同样，在图 13B 中，SBC 可连接到 CPCI 后平面的左后部连接器的一输入/输出单元 I/O。在此情况下，I/O 可包括 FCC 功能。又有，这种简单的不在机壳或支架上扩展的后平面或中平面结构实现了前述优点。

以上所述的典型实施例尽可能全面但不对本发明作任何限制，因

此本领域普通技术人员依据本发明可在实现细节上做许多变形。所有变形和修改均应在本发明权利要求所规定的范围和精神内。

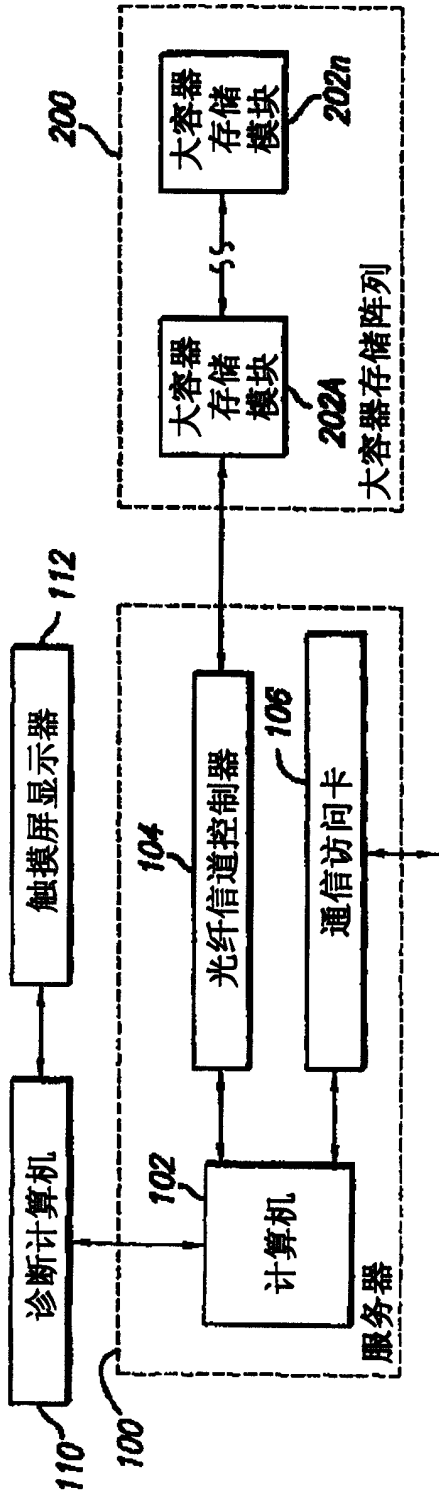


图1

通信系统

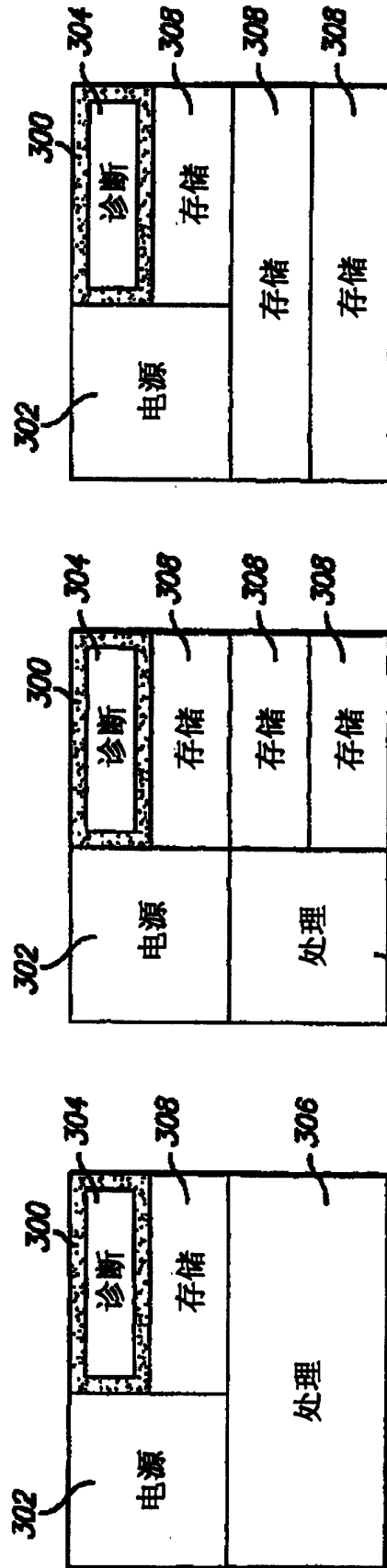


图2A

图2B

图2C

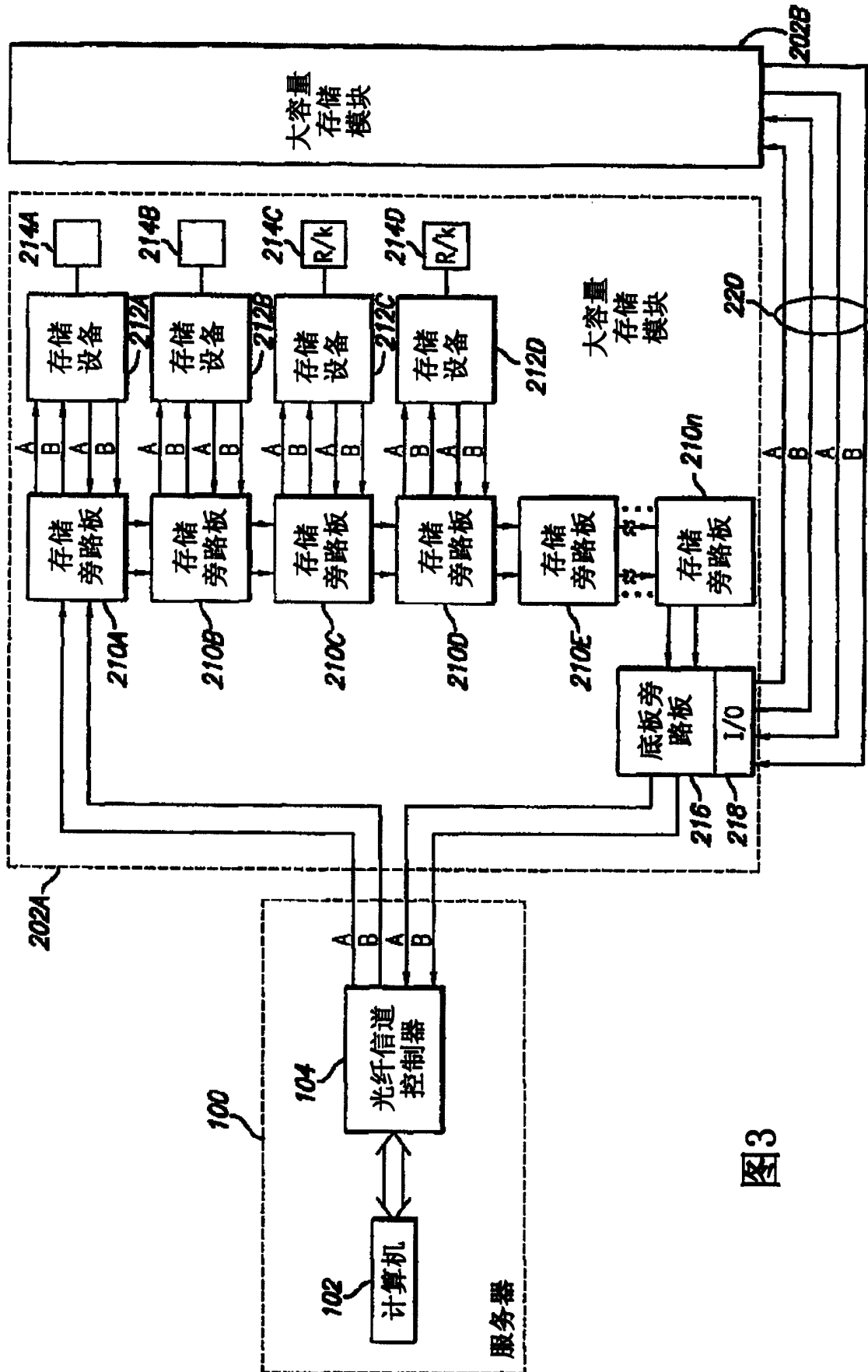


图3

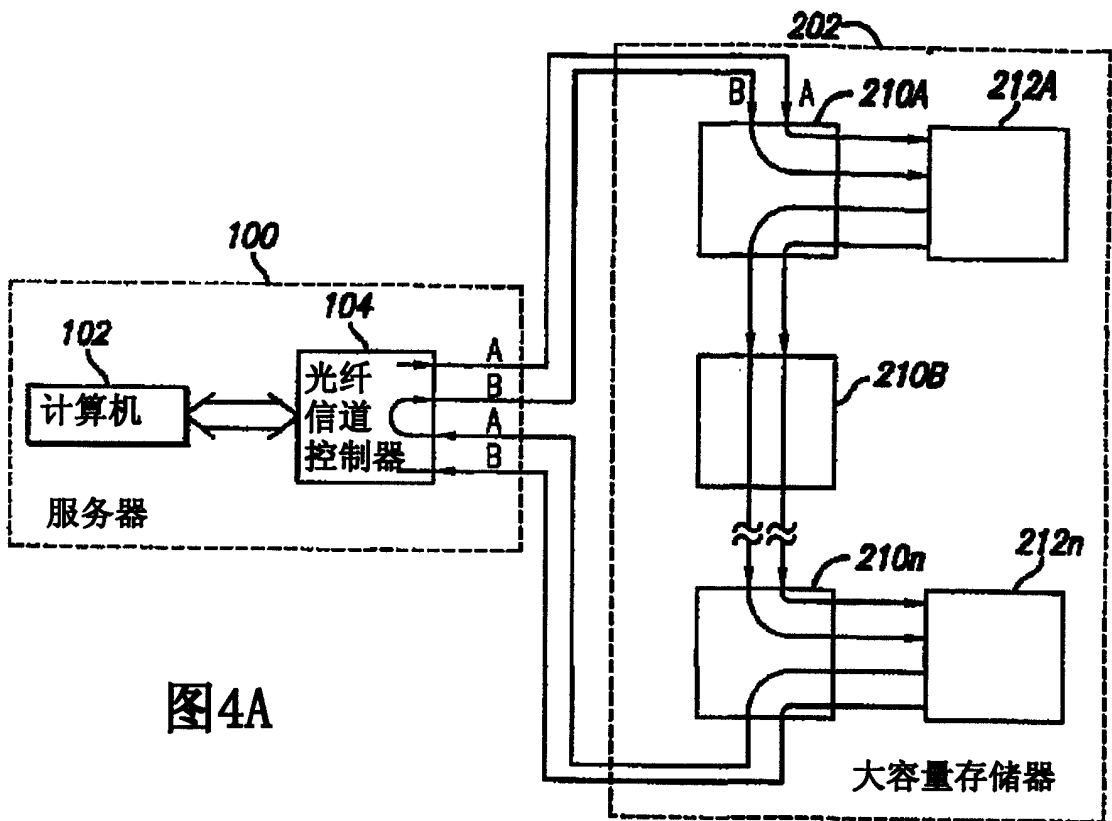


图4A

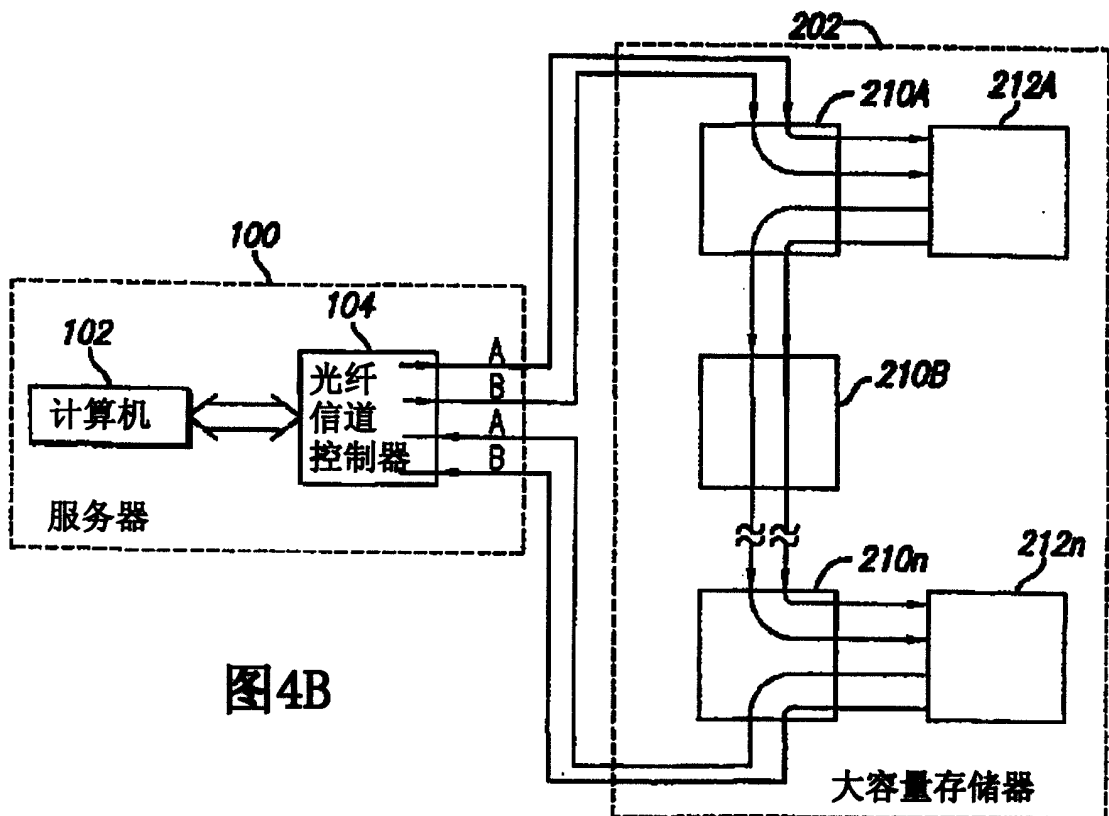
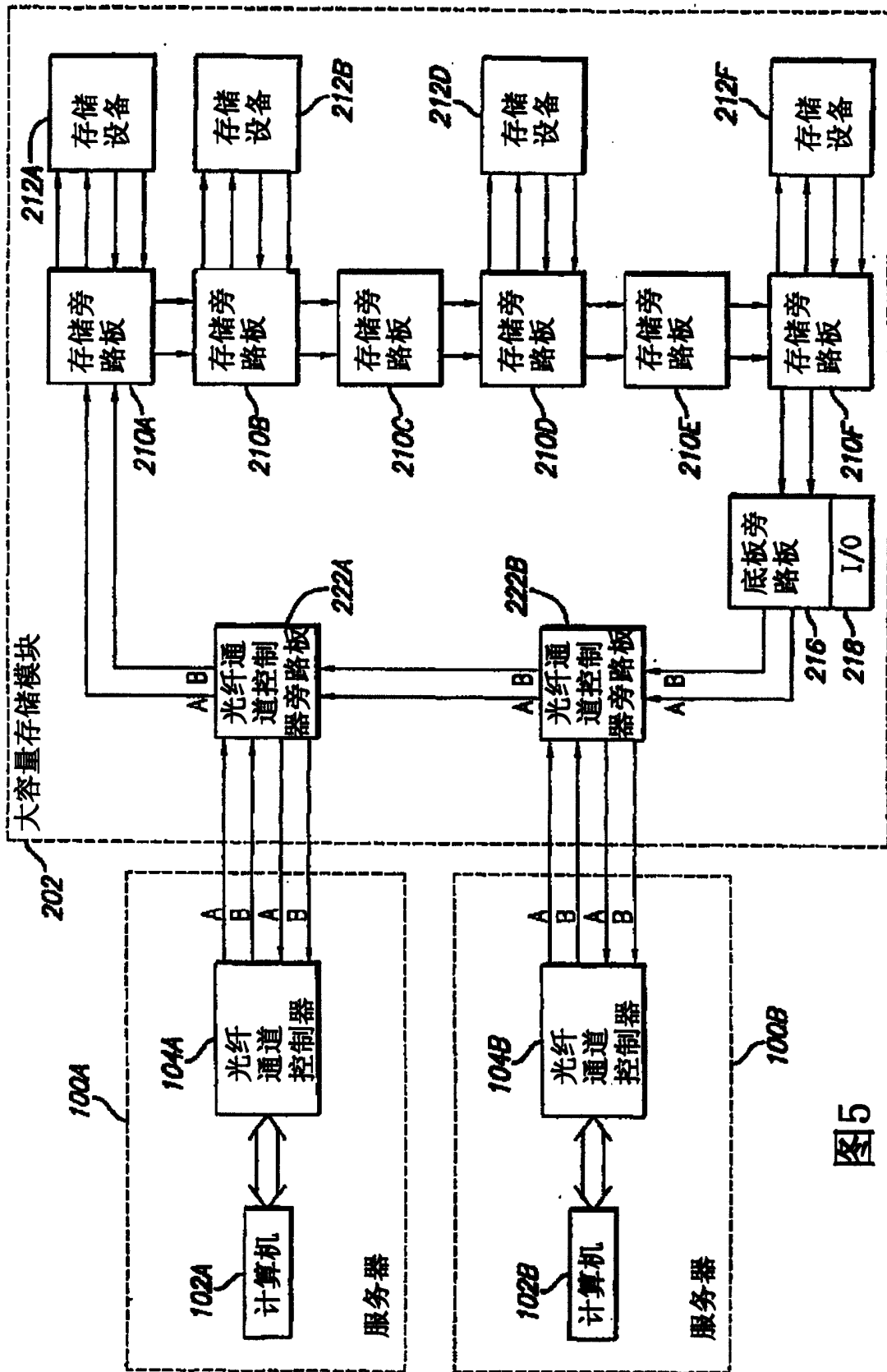


图4B



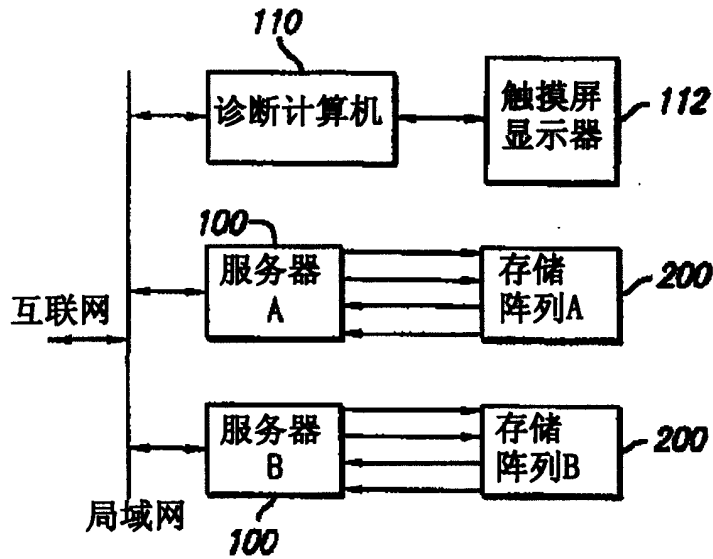


图6

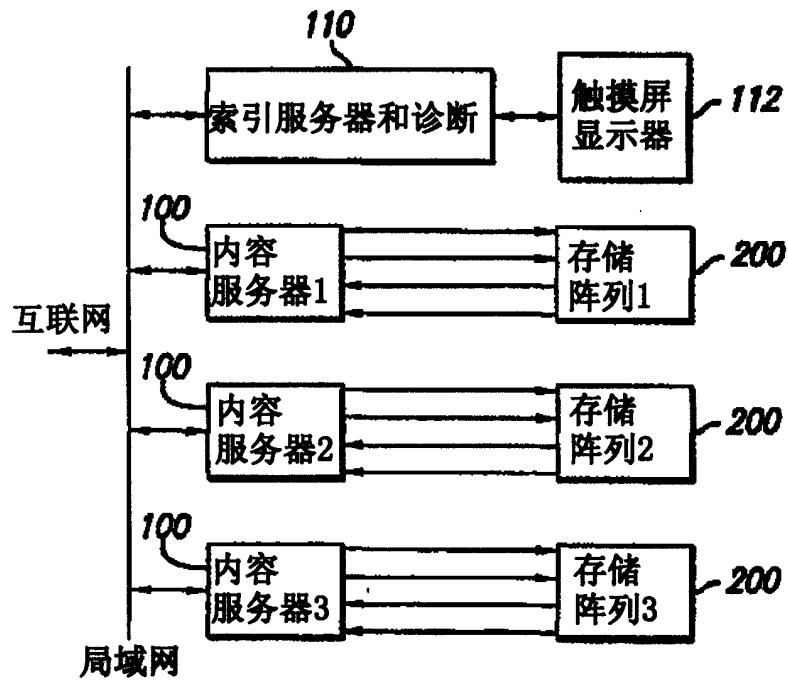


图7

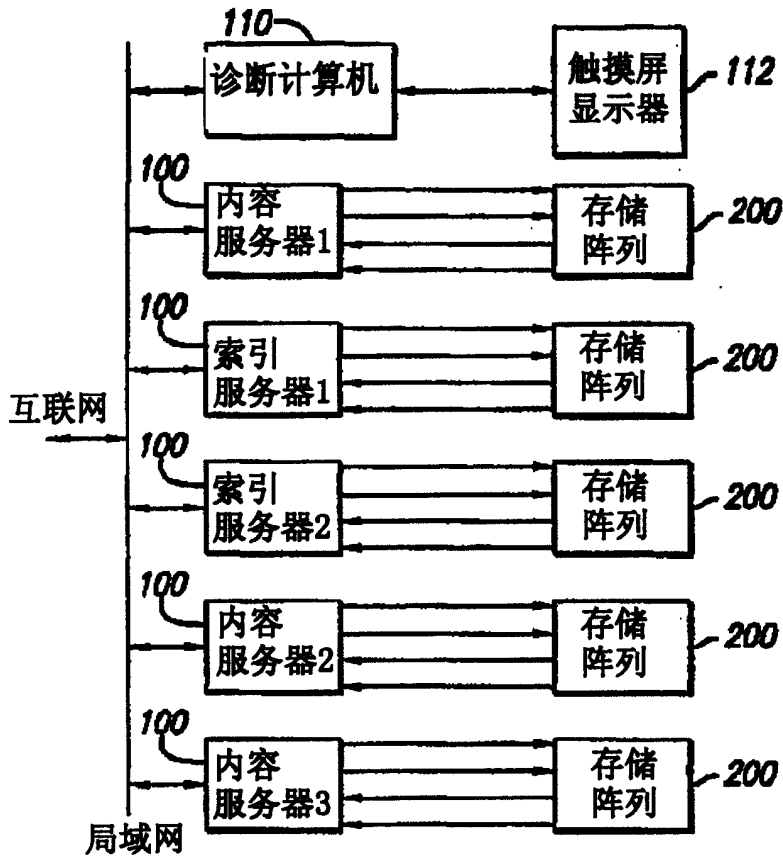


图8

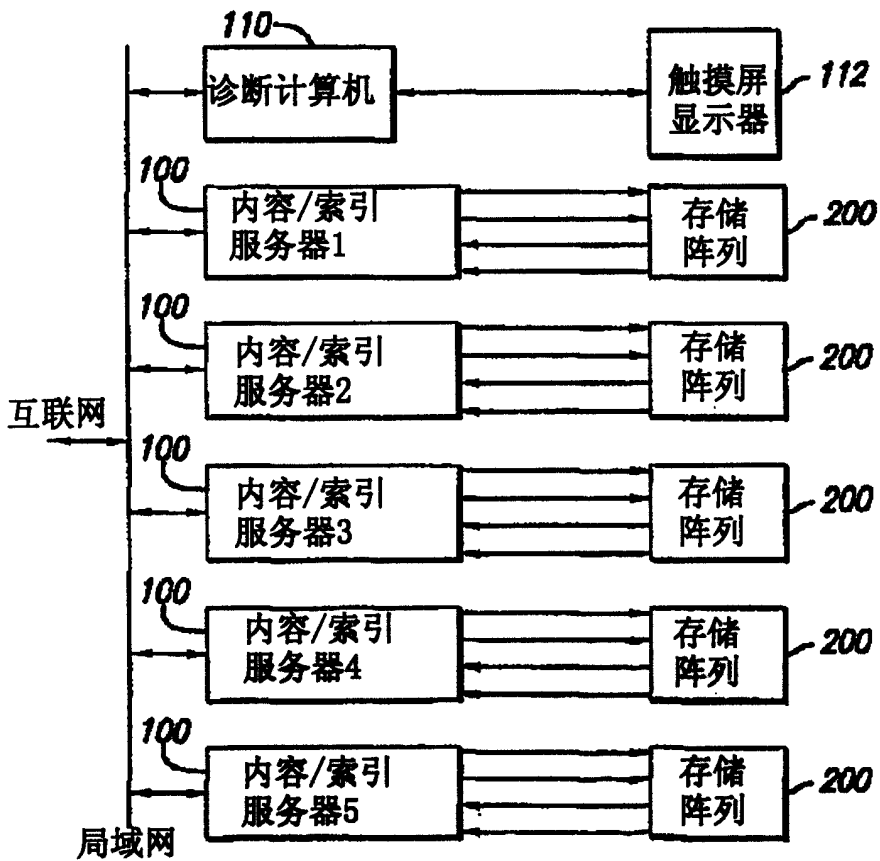


图9

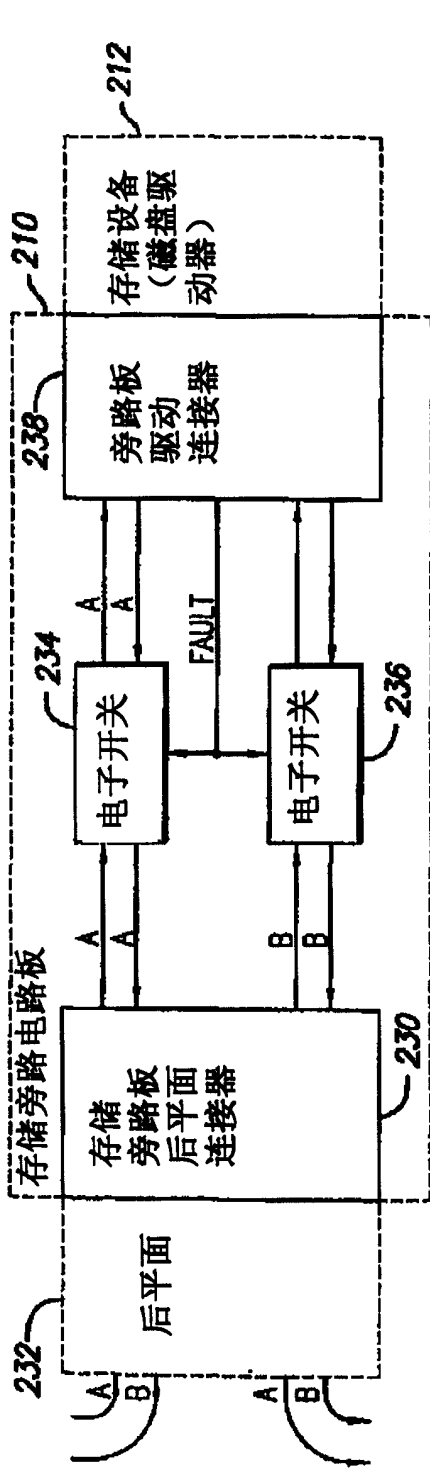


图10

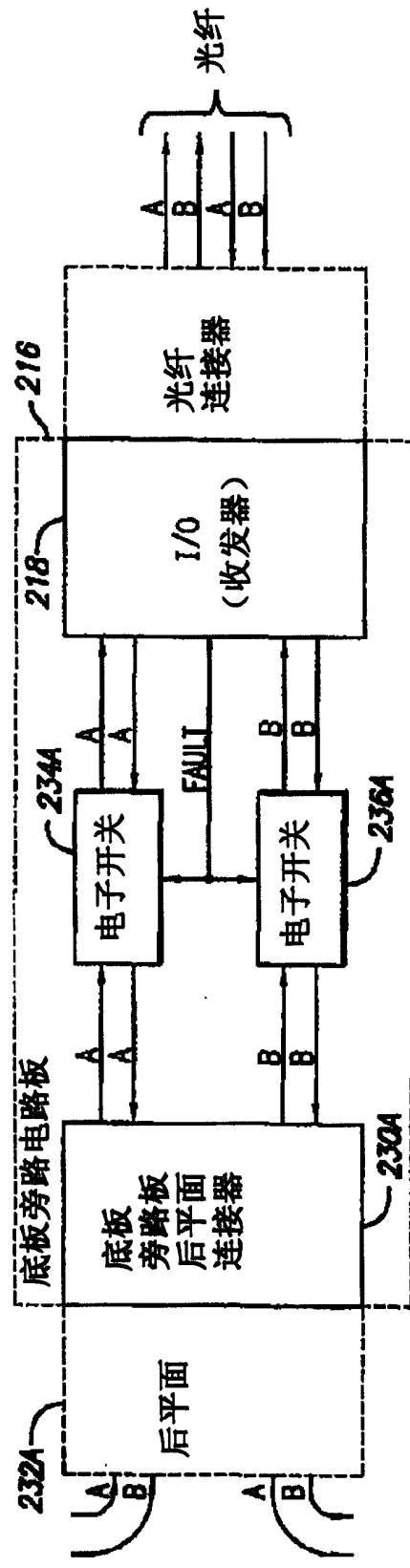


图11

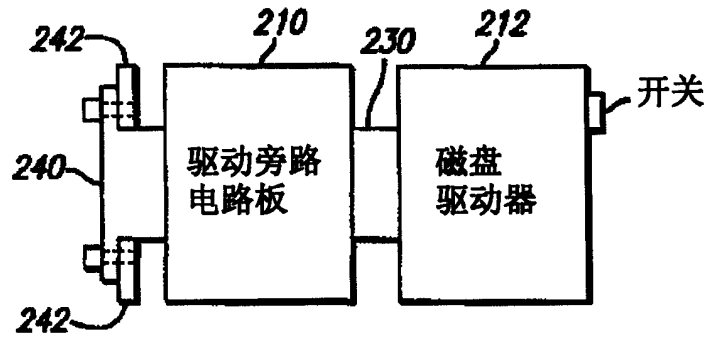


图12A

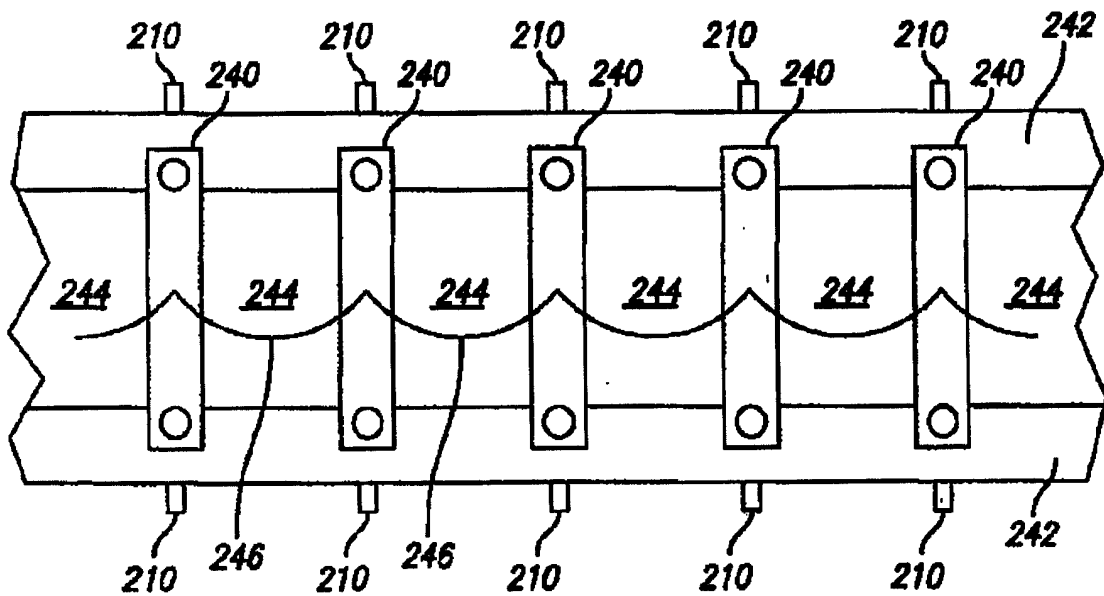


图12B

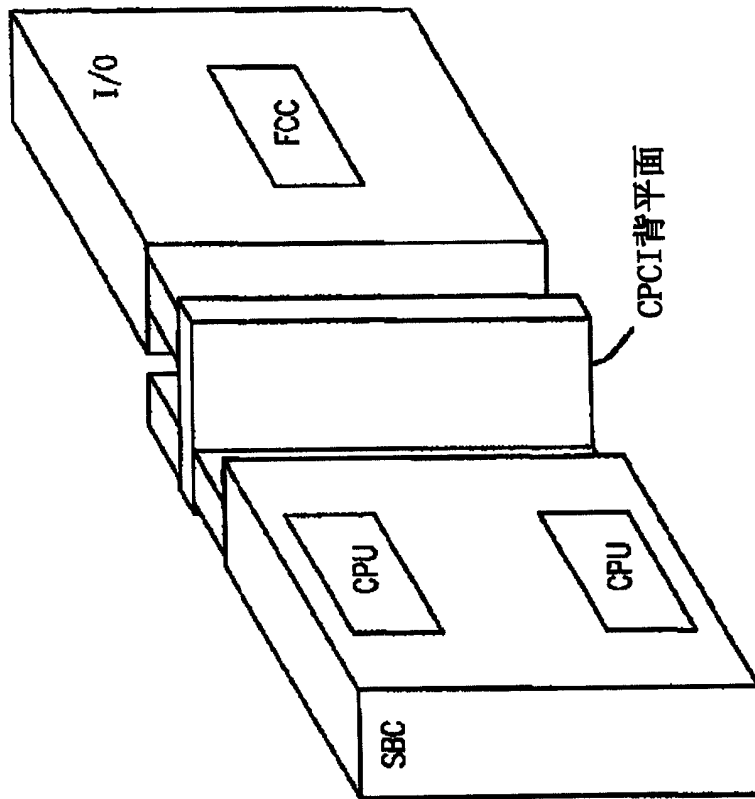


图13B

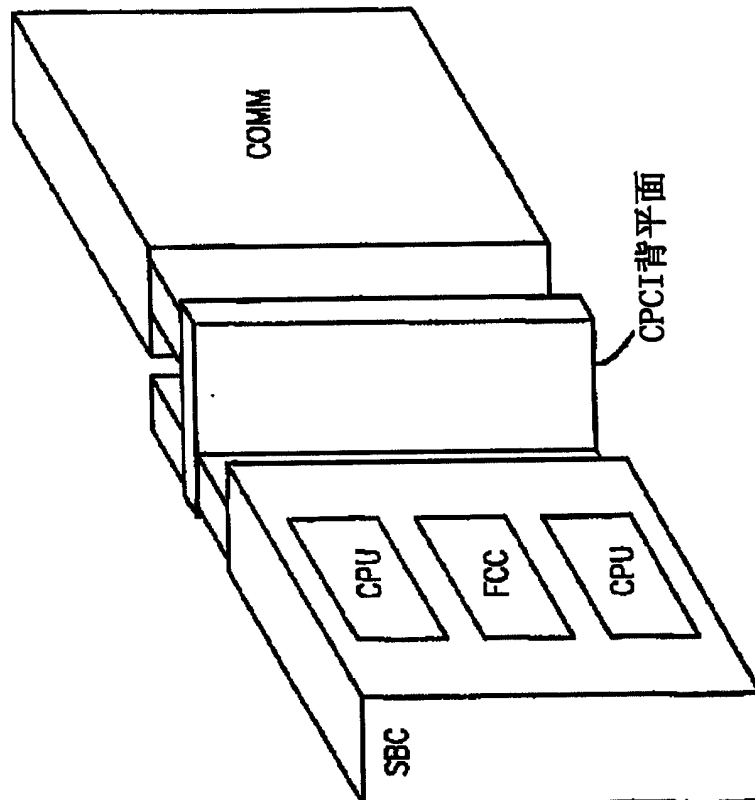


图13A