

(12) 发明专利

(10) 授权公告号 CN 101946242 B

(45) 授权公告日 2012.09.19

(21) 申请号 200980105228.4

代理人 于静 杨晓光

(22) 申请日 2009.02.10

(51) Int. Cl.

(30) 优先权数据

G06F 13/10(2006.01)

12/031,038 2008.02.14 US

(56) 对比文件

(85) PCT申请进入国家阶段日

US 5584039 A, 1996.12.10,

2010.08.13

EP 0374074 A2, 1990.06.20,

(86) PCT申请的申请数据

JP 63145562 A, 1988.06.17,

PCT/EP2009/051485 2009.02.10

US 5461721 A, 1995.10.24,

(87) PCT申请的公布数据

审查员 石志昕

W02009/101067 EN 2009.08.20

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 D·卡斯帕 J·弗拉纳根

H·尤登弗兰德 M·凯洛斯

M·班迪克 G·西格曼三世 C·黄

D·莱迪 U·恩尤库

(74) 专利代理机构 北京市中咨律师事务所

11247

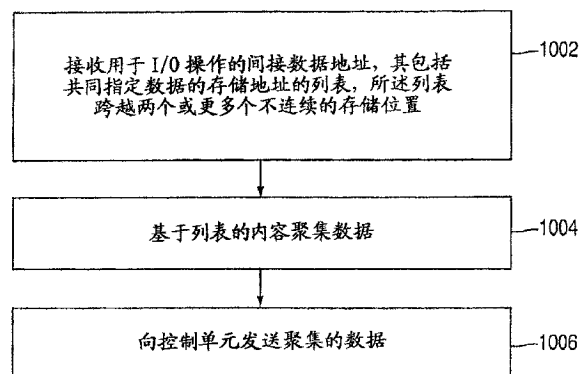
权利要求书 2 页 说明书 10 页 附图 10 页

(54) 发明名称

一种提供间接数据寻址的方法和系统

(57) 摘要

用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统、方法和计算机程序产品。所述计算机程序产品包括处理电路可读的有形存储介质,和用于处理电路执行的存储指令,用于执行一种方法。所述方法包括:接收用于 I/O 操作的控制字。所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址。所述间接数据地址包括共同指定数据的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置。响应于所述列表聚集数据。向所述 I/O 处理系统中的控制单元发送聚集的数据。



1. 一种用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,所述方法包括:

在所述 I/O 子系统处接收用于 I/O 操作的控制字,所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址,所述间接数据地址包括共同指定数据的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置;

响应于所述列表聚集数据,向所述 I/O 处理系统中的控制单元发送聚集的数据;

其中所述列表还包括与所述列表中的每个存储地址关联的标志;所述标志具有第一值,其指示关联的存储地址指向数据的一部分;以及所述标志具有第二值,其指示关联的存储地址指向所述列表中的附加存储地址所位于的存储位置;以及

其中对于所述列表上的每个存储地址,所述响应于所述列表聚集数据,向所述 I/O 处理系统中的控制单元发送聚集的数据包括:

访问所述存储地址以及与所述存储地址关联的标志;

响应于所述标志具有第一值,向数据增加来自所述存储地址的内容;

响应于所述标志具有第二值,访问第二存储地址以及位于所述存储地址的第二标志,以及

响应于所述第二标志具有第一值,向数据增加来自所述第二存储地址的内容。

2. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中当所述标志具有第一值时,所述列表中的第二存储地址定位于与所述列表中的第一存储地址连续;以及当所述标志具有第二值时,所述第二存储地址定位于与所述第一存储地址不连续。

3. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述列表还包括与所述列表中的每个存储地址关联的计数字段,所述计数字段指示从所述存储地址读取的字节数,以及所述聚集步骤响应于所述计数字段和所述存储地址。

4. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述数据包括控制数据。

5. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述数据包括客户数据。

6. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述控制字是传送控制字,存储地址的列表是传送间接数据地址列表,以及每个存储地址是传送间接数据地址字。

7. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述 I/O 子系统是通道子系统。

8. 如权利要求 1 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法,其中所述控制字包括用以指示其包含间接数据地址的标志。

9. 一种用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,所述系统包括:

用于在所述 I/O 子系统处接收用于 I/O 操作的控制字的装置,所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址,所述间接数据地址包括共同指定数据的存储地

址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置;

用于响应于所述列表聚集数据,向所述 I/O 处理系统中的控制单元发送聚集的数据的装置;

其中所述列表还包括与所述列表中的每个存储地址关联的标志;所述标志具有第一值,其指示关联的存储地址指向数据的一部分;以及所述标志具有第二值,其指示关联的存储地址指向所述列表中的附加存储地址所位于的存储位置;以及

其中对于所述列表上的每个存储地址,所述用于响应于所述列表聚集数据,向所述 I/O 处理系统中的控制单元发送聚集的数据的装置包括:

用于访问所述存储地址以及与所述存储地址关联的标志的装置;

用于响应于所述标志具有第一值,向数据增加来自所述存储地址的内容的装置;

用于响应于所述标志具有第二值,访问第二存储地址以及位于所述存储地址的第二标志的装置,以及

用于响应于所述第二标志具有第一值,向数据增加来自所述第二存储地址的内容的装置。

10. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中当所述标志具有第一值时,所述列表中的第二存储地址定位于与所述列表中的第一存储地址连续;以及当所述标志具有第二值时,所述第二存储地址定位于与所述第一存储地址不连续。

11. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述列表还包括与所述列表中的每个存储地址关联的计数字段,所述计数字段指示从所述存储地址读取的字节数,以及所述用于响应于所述列表聚集数据,向所述 I/O 处理系统中的控制单元发送聚集的数据的装置响应于所述计数字段和所述存储地址。

12. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述数据包括控制数据。

13. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述数据包括客户数据。

14. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述控制字是传送控制字,存储地址的列表是传送间接数据地址列表,以及每个存储地址是传送间接数据地址字。

15. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述 I/O 子系统是通道子系统。

16. 如权利要求 9 所述的用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的系统,其中所述控制字包括用以指示其包含间接数据地址的标志。

## 一种提供间接数据寻址的方法和系统

### 技术领域

[0001] 概括地说,本发明涉及输入 / 输出处理,更具体地,涉及在 I/O 处理系统的 I/O 子系统处提供不连续的间接数据寻址列表。

[0002] 背景技术

[0003] 输入 / 输出 (I/O) 操作用于在 I/O 处理系统的存储器和 I/O 设备之间传送数据。具体地,通过执行 I/O 操作从存储器向一个或多个 I/O 设备写入数据,以及从一个或多个 I/O 设备向存储器读取数据。

[0004] 为了便于 I/O 操作的处理,采用 I/O 处理系统的 I/O 子系统。I/O 子系统耦合至 I/O 处理系统的主存储器和 I/O 设备,以及指引在存储器和 I/O 设备之间的信息流。I/O 子系统的实例是通道子系统。通道子系统将通道路径用作通信介质。每个通道路径包括耦合至控制单元的通道,控制单元进一步耦合至一个或多个 I/O 设备。

[0005] 通道子系统可采用通道命令字 (CCW) 在 I/O 设备和存储器之间传送数据。CCW 指定要执行的命令。对于启动某些 I/O 操作的命令,CCW 指明与操作关联的存储器区域、每当完成向或从该区域的传送时要采取的行为、和其他选项。

[0006] 在 I/O 处理期间,通过通道从存储器提取 CCW 列表。通道解析来自 CCW 列表的每条命令,并将多条命令 (在其自身实体中的每条命令) 转发至与通道耦合的控制单元。然后,控制单元处理命令。通道跟踪每条命令的状态,并控制何时将下一命令集发送至控制单元用于处理。通道确保将每条命令发送至其自身实体中的控制单元。此外,通道推断与处理来自控制单元的对于每条命令的响应关联的某些信息。

[0007] 由于通道解析 CCW,跟踪状态信息,对来自控制单元的响应作出反应,所以对于通道子系统,基于每个 CCW 执行 I/O 处理可涉及大量处理负载。因此,有益地,将与解释和管理 CCW 和状态信息关联的大部分处理负担从通道子系统转移至控制单元。由于简化 I/O 处理系统中的控制单元和操作系统之间通信的通道角色执行更少的握手,所以这样可增加通信吞吐量。然而,改变命令序列以及通道子系统和控制单元的角色使得在一个 I/O 操作内传送的数据量达到多于 1 兆字节。当系统页面大小为 4k 字节时,上述字节是可利用一个传送间接数据地址的连续列表传送的最大数据量。目前,由于现有的通道命令字 (CCW) 中的 2 个字节计数字段的限制,所以 CCW 不能够在一个 I/O 操作中支持多于 64k 字节的数据传送。传送控制字 (TCW) 通过将字节计数增加至 TCW 中的 4 字节来解决这个问题,但是随后遇到下一个 1 兆字节的限制,因为传送间接数据地址列表 (TIDAL) 必须包含在为仅允许 256 个地址列表项的 4k 字节的一个页面中。

### 发明内容

[0008] 本发明的实施例包括一种用于在输入 / 输出 (I/O) 处理系统的 I/O 子系统处提供间接数据寻址的计算机程序产品。所述计算机程序产品包括:处理电路可读的有形存储介质,和用于处理电路执行的存储指令,用于执行一种方法。所述方法包括:接收用于 I/O 操作的控制字。所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址。所述间接

数据地址包括共同指定数据的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置。响应于所述列表聚集数据。向所述 I/O 处理系统中的控制单元发送聚集的数据。

[0009] 附加实施例包括一种用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的装置。所述装置包括: I/O 子系统,适于与所述 I/O 处理系统中的控制单元通信。所述 I/O 子系统执行一种方法,包括:接收用于 I/O 操作的控制字。所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址。所述间接数据地址包括共同指定数据的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置。响应于所述列表聚集数据。向所述控制单元发送聚集的数据。

[0010] 其他实施例包括一种用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的方法。所述方法包括:接收用于 I/O 操作的控制字。所述控制字包括用于与所述 I/O 操作关联的数据的间接数据地址。所述间接数据地址包括共同指定数据的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置。响应于所述列表聚集数据。向所述 I/O 处理系统中的控制单元发送聚集的数据。

[0011] 另外实施例包括一种用于在 I/O 处理系统的 I/O 子系统处提供间接数据寻址的计算机程序产品。所述计算机程序产品包括:处理电路可读的有形存储介质,和用于处理电路执行的存储指令,用于执行一种方法。所述方法包括:接收用于 I/O 操作的控制字。所述控制字包括用于来自所述 I/O 操作的客户数据输出的间接数据地址。所述间接数据地址包括共同指定客户数据输出的存储位置的存储地址的列表的开始位置,所述列表跨越两个或更多个不连续的存储位置。响应于所述 I/O 操作在所述 I/O 处理系统中的控制单元处执行,从所述控制单元接收所述客户数据输出。响应于所述列表,将所述客户数据输出存储在所述 I/O 子系统处的存储位置处。

[0012] 在阅览了以下附图和具体实施方式之后,对于本领域普通技术人员而言,根据实施例的制品、装置和 / 或方法的其他物品将变得清楚。目的在于,制品、装置和 / 或方法的所有这些附加物品都应该包括在本说明书中,在本发明的范围内,以及受到所附权利要求的保护。

## 附图说明

[0013] 特别地指出作为本发明的主题,并且在说明书中包含的权利要求书中不同地主张。根据集合附图进行的以下具体实施方式,本发明的以上和其他目的、特征、和优点变得清楚,其中:

[0014] 图 1 示出结合和使用本发明的一个或多个方面的 I/O 处理系统的一个实施例;

[0015] 图 2A 示出现有技术的通道命令字的一个实例;

[0016] 图 2B 示出现有技术的通道命令字通道程序的一个实例;

[0017] 图 3 示出在通道和控制单元之间的通信中使用的执行图 2B 的通道命令字通道程序的现有技术链路协议的一个实施例;

[0018] 图 4 示出根据本发明一方面的传送控制子通道程序的一个实施例;

[0019] 图 5 示出根据本发明一方面的用于在通道和控制单元之间通信的执行图 4 的传送控制字通道程序的链路协议的一个实施例;

[0020] 图 6 示出用于在通道和控制单元之间通信的以执行通道命令字通道程序的 4 个写入命令的现有技术链路协议的一个实施例；

[0021] 图 7 示出根据本发明一方面的用于在通道和控制单元之间通信的以处理传送控制字通道程序的 4 个写入命令的链路协议的一个实施例；

[0022] 图 8 示出根据本发明一方面的控制单元和通道子系统的实施例；

[0023] 图 9 示出根据本发明一方面的传送模式间接数据地址字 (TIDAW) 的一个实施例；

[0024] 图 10 示出用于在 I/O 子系统处提供不连续间接数据寻址的处理的一个实施例；以及

[0025] 图 11 示出结合本发明的一方面或多个方面的制品的一个实施例。

[0026] 具体实施方式参照附图借助于实例说明本发明的优选实施例、以及优点和特点。

### 具体实施方式

[0027] 根据本发明的一方面,便于输入 / 输出 (I/O) 处理。在一个实例中,通过减少用于执行 I/O 处理的 I/O 处理系统的组件之间的通信来便于 I/O 处理。例如,减少在 I/O 通信适配器 (例如通道) 和控制单元之间交换和序列的数目。这可通过从 I/O 通信适配器向作为单独实体的控制单元发送多条命令用于控制单元的执行,以及通过控制单元作为单独实体发送从命令得到的数据 (如果存在) 来实现。

[0028] 多条命令包括在这里称为传送命令控制块 (TCCB) 的块中,其地址在传送控制字 (TCW) 中指定。将 TCW 从操作系统 (OS) 或其他应用发送至 I/O 通信适配器,后者随后将命令消息中的 TCCB 转发至控制单元用于处理。控制单元通过 I/O 通信适配器处理每条命令,其不具有相对于那些个体命令的状态的跟踪。多条命令还称为通道程序,其在控制单元而非 I/O 通信适配器上被解析和执行。

[0029] 在示例性实施例中,用于 I/O 操作的 TCW 包括用于指示控制数据 (例如 TCCB) 的位置的指针以及与 I/O 操作关联的客户数据。在示例性实施例中,指针指向间接数据地址,这里称为传送模式间接数据地址列表 (TIDAL)。TIDAL 包括数据所在的地址的列表;这些地址在这里称为传送模式间接数据地址字 (TIDAW)。在具有 4K 页面的当前系统设计中,TIDAL 不能够大于 4 千字节 (4K),由此将可包括在单一 TIDAL 中的连续 TIDAW 的数目限制在 256 (每个 TIDAL 为 16 字节)。假设 4K 的页面,每个 TIDAL 为 16 字节,并且 TIDAL 必须包含在一个 4K 页面中,则这将可由一个 I/O 操作传送的总数据量限制在 1 兆字节。在本发明的示例性实施例中,通过允许 TIDAW 地址指示在另一存储位置 (可能在不同页面中) 处的下一 TIDAW 的开始地址来消除这个限制。利用 TIDAW 中的标志来指示 TIDAW 的地址包括数据地址还是 TIDAW 列表的连续的地址。这样,构成单一 TIDAL 的 TIDAW 可位于不连续的存储位置中,因此多于 256 个 TIDAW 可包含在一个 TIDAL 列表中。这允许更多的数据在一个 I/O 操作中传送。例如,在示例性实施例中,如果 TCW 中的计数字段为 4 字节,则可在单一 I/O 操作中传送大约 4 兆字节 (4 兆字节减 1 字节)。

[0030] 参照图 1 描述结合和使用本发明一个或多个方面的 I/O 处理系统的一个实例。I/O 处理系统 100 包括主机系统 101,其进一步包含例如主存储器 102、一个或多个中央处理单元 (CPU) 104、存储控制元件 106、和通道子系统 108。主机系统 101 可以是大型计算系统,例如大型机或服务器。I/O 处理系统 100 还包括一个或多个控制单元 110 和一个或多个 I/O

设备 112,其每个如下所述。

[0031] 主存储器 102 存储可从 I/O 设备 112 输入的数据和程序。例如,主存储器 102 可包括由 CPU 104 中的一个或多个执行的一个或多个操作系统 (OS) 103。例如,作为不同的虚拟机实例,一个 CPU 104 可执行 Linux<sup>®</sup>操作系统 103 和 z/OS<sup>®</sup>操作系统 103。主存储器 102 是可直接寻址的,以及提供由 CPU 104 和通道子系统 108 进行的数据的高速处理。

[0032] CPU 104 是 I/O 处理系统 100 的控制中心。其包含用于指令执行、中断行为、定时功能、初始程序加载、和其他机器相关功能的定序和处理装备。CPU 104 经由连接 114(例如双向或单向总线)耦合至存储控制元件 106。

[0033] 存储控制元件 106 经由连接 116(例如总线)耦合至主存储器 102;经由连接 114 耦合至 CPU 104;以及经由连接 118 耦合至通道子系统 108。存储控制元件 106 控制例如由 CPU 014 和通道子系统 108 发出的请求的排队和执行。

[0034] 在示例性实施例中,通道子系统 108 提供主机系统 101 和控制单元 110 之间的通信接口。通道子系统 108 耦合至存储控制元件 106,如上所述;以及经由连接 120(例如串行链路)耦合至每个控制单元 110。连接 120 可实现为光链路,在光纤通道结构中采用单模或多模波导。通道子系统 108 指引 I/O 设备 112 和主存储器 102 之间的信息流。其免除了 CPU 104 直接与 I/O 设备 112 通信的任务,并允许数据处理与 I/O 处理同时进行。在管理到或从 I/O 设备 112 的信息流时,通道子系统 108 将一个或多个通道路径 122 用作通信链路。作为 I/O 处理的一部分,通道子系统 108 还执行测试通道路径可用性、选择可用通道路径 122、和通过 I/O 设备启动操作的执行的路径管理功能。

[0035] 每个通道路径 122 包括通道 124(在一个实例中,通道 124 位于通道子系统 108 中,如图 1 所示)、一个或多个控制单元 110 和一个或多个连接 120。在另一实例中,其还可以具有一个或多个动态开关(未示出),作为通道路径 122 的一部分。动态开关耦合至通道 124 和控制单元 110,并提供对于附连至开关的任意两个链路物理互连的能力。在另一实例中,其还可具有多个系统,因此具有附连至控制单元 110 的多个通道子系统(未示出)。

[0036] 同样位于通道子系统 108 中的有子通道(未示出)。提供一个子通道用于和专用于可通过通道子系统 108 访问程序的每个 I/O 设备 112。子通道(例如数据结构,如表)向程序提供设备的逻辑表示。每个子通道提供关于关联的 I/O 设备 112 的信息及其到通道子系统 108 的附连。子通道还提供关于 I/O 操作的信息以及涉及关联的 I/O 设备 112 的其他功能。子通道是通道子系统 108 向 CPU 104 提供关于关联的 I/O 设备 112 的信息所借助的手段,所述 CPU 104 通过执行 I/O 指令获得该信息。

[0037] 通道子系统 108 耦合至一个或多个控制单元 110。每个控制单元 110 提供操作和控制一个或多个 I/O 设备 112 以及通过使用共同装备使得每个 I/O 设备 112 的特征适配于通道 124 提供的链路接口的逻辑。共同装备提供了 I/O 操作的执行、关于 I/O 设备 112 和控制单元 110 的状态的指示、在通道路径 112 上的数据传送的定时的控制、以及 I/O 设备 112 的某些等级的控制。

[0038] 每个控制单元 110 经由连接 126(例如总线)附连至一个或多个 I/O 设备 112。I/O 设备 112 接收信息,或将信息存储在主存储器 102 和/或其他存储器中。I/O 设备 112 的实例包括例如读卡器和打孔器、磁带单元、直接访问存储设备、显示器、键盘、打印机、定点设备、远程处理设备、通信控制器和基于传感器的装置。

[0039] I/O 处理系统 100 的上述组件中的一个或多个进一步描述在以下文献中：“IBM® z/Architecture Principles of Operation”，公开号 No. SA22-7832-05，第 6 版，2007 年 4 月；美国专利 No. 5,461,721，题为“System For Transferring Data Between I/O Devices And Main Or Expanded Storage Under Dynamic Control Of Independent Indirect Address Words (IDAWS)” Cormier 等人，发布于 1995 年 10 月 24 日；和美国专利 No. 5,526,484，题为“Method And System For Pipelining The Processing Of Channel Command Words” Casper 等人，发布于 1996 年 6 月 11 日，其中每个的全部内容通过引用合并于此。IBM 是美国纽约阿蒙克的国际商业机器公司的注册商标。这里使用的其他名称可以是国际商业机器公司或其他公司的注册商标、商标或产品名称。

[0040] 在一个实施例中，为了在 I/O 设备 112 和存储器 102 之间传送数据，使用通道命令字 (CCW)。CCW 指定要执行的命令，并包括控制处理的其他字段。参照图 2A 描述 CCW 的一个实例。CCW 200 包括例如，命令码 202，指定要执行（例如，读取、反向读取、控制、感测和写入）的命令；多个标志 204，用于针对指定数据传送的命令控制 I/O 操作；计数字段 206，指定在要传送的 CCW 所指明的存储区域中的字节数；以及数据地址 208，当采用直接寻址时其指向包括数据的主存储器中的位置，或当采用修改间接数据寻址时其指向要处理的修改间接数据地址字 (MIDAW) 的列表（例如连续列表）。修改间接寻址进一步在以下文件中描述：美国专利序号 11/464,613，题为“Flexibly Controlling The Transfer Of Data Between Input/Output Devices And Memory” Brice 等人，2006 年 8 月 15 日递交，其全部内容通过引用合并于此。

[0041] 被设置用于按序执行的一个或多个 CCW 形成通道程序，这里还称为 CCW 通道程序。CCW 通道程序通过例如操作系统或其他软件来设置。软件设置 CCW，并获得分配给通道程序的存储器地址。参照图 2B 描述 CCW 通道程序的实例。CCW 通道程序 210 包括例如界定范围 CCW 212，其具有指针 214，指向要通过界定范围命令使用的界定范围数据 216 的存储器中的位置。在该实例中，通道传送 (TIC) 218 遵循界定范围命令，其使得通道程序指向包括一个或多个其他 CCW（例如，定位记录 217，其具有指针 219 以定位记录数据 220；和一个或多个写入 CCW 221）的存储器中的另一区域（例如应用区域）。每个写入 CCW 220 具有指向数据区域 224 的指针 222。数据区域包括直接访问数据的地址或间接访问数据的数据地址字的列表（例如 MIDAW 或 IDAW）。此外，CCW 通道程序 210 包括由设备地址定义的通道子系统内的预定区域，称为用于从 CCW 通道程序的执行得到的状态 226 的子通道。

[0042] 参照图 3 以及参照图 2B 描述 CCW 通道程序的处理。具体地，图 3 示出当 CCW 通道程序正在执行时发生在通道和控制单元之间的各种交换和序列的实例。在这个实例中，用于通信的链路协议为 FICON（光纤连接）。关于 FICON 的信息在以下文件中描述：“Fibre Channel Single Byte Command Code Sets-3 Mapping Protocol (FC-SB-3)”，T11/Project1357-D/Rev. 1.6，INCITS (2003 年 3 月)，其全部内容通过引用合并于此。

[0043] 参照图 3，通道 300 打开与控制单元 302 的交换，并向控制单元 302 发送界定范围命令和与其关联的数据 304。从界定范围 CCW 212 提取命令（图 2B），以及从界定范围数据区域 216 获得数据。通道 300 使用 TIC 218 来定位所述定位记录 CCW 和写入 CCW。其从定位记录 CCW 217（图 2B）和定位记录数据 220 提取定位记录命令和数据 306（图 3）。从写入 CCW 221 和数据区域 224（图 2B）提取写入命令和数据 308（图 3）。将每个发送至控制单元



302。

[0044] 响应于通道 300 的打开的交换,控制单元 302 打开与通道 300 的交换 310。这可发生在定位命令和数据 306 和 / 或写入命令和数据 308 之前或之后。除了打开的交换之外,向通道 300 转发响应 (CMR)。CMR 向通道 300 提供控制单元 302 为活动和正在运行的指示。

[0045] 控制单元 302 向通道 300 提供状态,并关闭交换 312。响应于此,通道 300 存储数据,检查状态和关闭交换 314,其向控制单元 302 指示状态被接收。

[0046] 上述 CCW 通道程序的写入 4k 数据的处理需要两个要打开和关闭的交换和六个序列。通过将通道程序的多个命令放置于 TCCB 中来减少在通道和控制单元之间的交换和序列的总数。通道 (例如图 1 的通道 124) 使用 TCW 来识别 TCCB 的位置,以及用于访问和存储与执行通道程序关联的状态和数据的位置。TCW 通过通道来解释,并且不通过控制单元发送或看到。

[0047] 参照图 4 描述通道程序的写入 4k 数据 (如图 2B,但包括 TCCB,并非单独的个体 CCW) 的一个实例。如图所示,通道程序 400 (这里称为 TCW 通道程序) 包括 TCW 402,其指定 TCCB 404 的存储器中的位置,以及数据区域 406 的存储器中的位置或指向数据区域 406 的 TIDAL 410 (即,传送模式间接数据地址字 (TIDAW) 的列表,类似于 MIDAW)、和状态区域 408。TIDAW 零 412、TIDAW 一 414、和 TIDAW 二 416 (共同地,TIDAW 412-416) 可指向用于获取或存储数据的数据区域 406 中的不同位置。TIDAW 412-416 可指向数据的不连续块或数据的连续块。TIDAL 410 中的 TIDAW 412-416 可在存储器中顺序地定位,或相对于彼此不连续地定位。尽管在 TIDAL 410 中仅示出 3 个 TIDAW 412-416,但是应理解,可在 TIDAL 410 中包括任意数目个 TIDAW。

[0048] 参照图 5 描述 TCW 通道程序的处理。用于这些通信的链路协议为例如光纤通道协议 (FCP)。具体地,使用 FCP 链路协议的 3 个阶段,从而允许使用主机总线适配器,其支持 FCP 执行可由 CCW 控制的数据传送。FCP 及其阶段进一步在以下文献中描述:“Information Technology-Fibre Channel Protocol for SCSI, Third Version (FCP-3)” T10Project 1560-D, 版本 4, 2005 年 9 月 13 日,其全部内容通过引用合并于此。

[0049] 参照图 5,通道 500 打开与控制单元 502 的交换,并向控制单元 502 发送 TCCB 504。在一个实例中,在 FCP 命令 (称为 FCP CMND 信息单元 (IU) 或传送命令 IU) 中将 TCCB 504 和序列初步操作 (initiative) 传送至控制单元 502。当通道 500 准备接收用于 TCCB 504 中接收的写入命令的数据时,控制单元 502 将传送准备 (XFER\_RDY) IU 510 发送至通道 500。响应于接收 XFER\_RDY IU 510,通道 500 例如经由 FCP\_Data IU 将数据 506 传送至控制单元 502。

[0050] 控制单元 502 执行 TCCB 504 的多个命令 (例如,作为设备控制字 (DCW) 的界定范围命令、定位记录命令、写入命令),并写入从通道 500 接收的数据 506。控制单元 502 还提供状态并关闭交换 508。作为一个实例,在例如 FCP\_RSP IU (还称为传送响应 IU) 的有效载荷的字节 10 或 11 中具有比特活动的 FCP 状态帧中发送最终状态。除了附加状态信息之外,FCP\_RSP IU 有效载荷可用于传送 FICON 结束状态。

[0051] 当启用 XFER\_RDY 时,利用图 5 中所示的链路协议的实施例。在图 5 中所示的实施例中,通道 500 不能够向控制单元 502 发送数据 506,直到由控制单元 502 经由 XFER\_RDY IU 510 对其进行请求。在备选示例性实施例中,禁用 XFER\_RDY,并且控制单元不向通道 500 发

送 XFER\_RDYIU 510。因此,通道 500 不必在发送数据 506 之前等待控制单元 502 请求数据 506。当通道 500 和控制单元 502 定位于地理上彼此远离(例如大于 2 万米,大于 5 万米)时,可利用这个备选实施例(其中 XFER\_RDY 被禁用)来提高性能。除非特别指定,这里的讨论假设启用 XFER\_RDY。

[0052] 在另一实例中,为了写入 4K 客户数据,通道 500 使用 FCP 链路协议阶段如下:

[0053] 1. 向控制单元 502 传送 FCP\_CMND IU 中的 TCCB 和序列初步操作。

[0054] 2. 等待 XFER\_RDY IU,其指示控制单元准备接收数据。

[0055] 3. 向控制单元 502 传送数据的 IU,和序列初步操作。

[0056] 4. 在例如 FCP\_RSP IU 有效载荷的字节 10 或 11 中具有比特活动的 FCP 状态帧中发送最终状态。除了附加状态信息之外,FCP\_RSP\_INFO 字段或感测字段用于传送 FICON 结束状态。

[0057] 通过执行图 4 的 TCW 通道程序,仅存在 1 个打开和关闭的交换(也可见图 5),并非图 2B 的 CCW 通道程序的 2 个交换(也可见图 3)。此外,相比于针对 CCW 通道程序的 6 个序列(见图 2B-3),对于 TCW 通道程序,存在 4 个通信序列(见图 4-5)。

[0058] 对于 TCW 通道程序,即使向该程序增加附加命令,交换和序列的数目保持相同。例如,将图 6 的 CCW 通道程序的通信与图 7 的 TCW 通道程序的通信相比较。在图 6 的 CCW 通道程序中,在从通道 610 到控制单元 612 的单独序列中发送每条命令(例如界定范围命令和数据 600、定位记录命令和数据 601、写入命令和数据 602、写入命令和数据 604、定位记录命令和数据 606、写入命令和数据 608、和写入命令和数据 620)。这个 CCW 通道程序需要 2 个打开和关闭的交换(例如打开交换 622、624,和关闭交换 626、628),和 10 个通信序列。这相比于图 7 的 TCW 通道程序的 4 个序列和 1 个交换,实现了与图 6 的 CCW 通道程序相同的任务。

[0059] 如图 7 所示,通道 700 打开与控制单元 702 的交换,并向控制单元 702 发送 TCCB 704。TCCB 704 包括 DCW 中的界定范围命令、两个定位记录命令、和四个写入命令,如上所述。类似于图 5 中所示的实例,假如 XFER\_RDY 支持没有禁用,控制单元 702 可使用 XFER\_RDY IU 710 来向通道 700 通知其准备接收数据。在接收到 XFER\_RDY IU 710 时,通道 700 在单一序列中向控制单元 702 发送 16k 数据 706。通道 700 在序列中的 16k 数据 706 的每 4K 插入 CRC。每 4K 的 CRC 的插入允许控制单元 702 递增地验证 16K 数据,而并非在完成 TCCB 704 中的写入命令之前缓冲整个 16K 用于验证。此外,控制单元 702 向通道 700 提供状态,并关闭交换 708。因此,图 7 的 TCW 通道程序需要更少的通信来传送与图 6 的 CCW 通道程序相同的数据量,同时经由在来自通道 700 的输出数据流中的多 CRC 插入来支持递增的数据验证。

[0060] 现在转到图 8,更详细地示出了支持 TCW 通道程序执行的在通道子系统 108 中的通道 124 和图 1 的控制单元 110 的一个实施例。控制单元 110 包括 CU 控制逻辑 802,用以解析和处理含有 TCCB(例如图 7 的 TCCB704)的消息,以及经由连接 120 从通道 124 接收的数据。CU 控制逻辑 802 可从在控制单元 110 处接收的 TCCB 提取 DCW 和控制数据,以经由连接 126 控制例如 I/O 设备 112 的设备。CU 控制逻辑 802 向 I/O 设备 112 发送设备命令和数据,以及从 I/O 设备 112 接收状态信息和其他反馈。当 CU 控制逻辑 802 接收数据(例如,图 7 的 16K 字节数据 706 的第一检查块边界)时,CU 控制逻辑 802 将接收的这个数据写入数据

缓冲器 804,用于临时存储,直到接收到用于检查块边界的 CRC,然后可将该数据发送至 I/O 设备 112。对于每个检查块边界,这个操作继续,直到完成 I/O 操作。

[0061] 控制单元 110 可进一步包括其他队列或存储器元件(未示出),用于存储与通道 124 和 I/O 设备 112 之间的通信关联的附加消息或状态信息。

[0062] 通道子系统 108 中的通道 124 包括支持与控制单元 110 通信的元件。例如,通道 124 可包括 CHN 控制逻辑 806,其与聚集数据逻辑 812 对接。以下参照图 10 来描述聚集数据逻辑 812。在示例性实施例中,CHN 控制逻辑 806 控制通道子系统 108 和控制单元 110 之间的通信。CHN 控制逻辑 806 可经由连接 120 直接对接至 CU 控制逻辑 802,以发送命令和接收响应,例如传送命令和响应 IU。备选地,消息接口和/或附加缓冲器(未示出)可放置于 CHN 控制逻辑 806 和 CU 控制逻辑 802 之间。CHN 子系统寄存器 814 可包括提供配置和状态信息以及动态状态信息的固定值,其在传送命令和接收响应时更新。CHN 子系统寄存器 814 可以是专用的硬件寄存器和/或使用存储器映射建立的虚拟寄存器。

[0063] 在一个实施例中,CHN 子系统寄存器 814 包括图 4 的 TIDAL 410 和 TIDAW 412-416,作为存储映射的寄存器。

[0064] 图 9 中示出 TIDAW 900 的一个实例。TIDAW 900 对于 TCW 通道程序中使用的数据提供间接寻址,例如图 4 中的 TIDAW 412-416。TIDAW900 包括标志 902、计数 904、和地址 906。将 TIDAW 格式 900 的每个字段(即标志 902、计数 904、和地址 906)分配给特定字节地址,以支持字段的解析。尽管图 9 中示出 TIDAW 900 中的字段的一个安排,但是应理解,字段的顺序可以重新安排以改变排序。

[0065] 在示例性实施例中,除了其他标志之外,标志 902 包括最后 TIDAW 标志和传输-通道传送(T-TIC)标志。根据 MIDAW 的定义,最后 TIDAW 标志指示关联的 TIDAW 是 TIDAL 中的最后 TIDAW。当通过最后 TIDAW 标志的设置使得计数 904 到达零时,完成对于关联的 I/O 操作的数据传送。T-TIC 标志指示地址 906 的内容包括数据还是 TIDAL 中的下一 TIDAW 的地址。在示例性实施例中,当 T-TIC 标志被设置时,随后 TIDAW 中的地址 906 是 TIDAL 中的下一 TIDAW 的地址。以这个方式,可利用地址 906 从当前 TIDAW 访问在不连续存储位置处的 TIDAW。因此,TIDAW 列表可包含多于 256 个项,从而超过当前方案中可容许的间接数据地址字的最大数。在备选实施例中,当 T-TIC 标志被设置时,地址 906 必须具有设置为零的四个低阶比特,因为 16 字节 TIDAW 必须在 16 字节地址边界上。当 T-TIC 标志没有被设置时,随后,TIDAW 中的地址 906 是构成为 I/O 操作正在被聚集的数据的一部分数据的地址。在计数字段 904 中指示出数据的大小。如果最后 TIDAW 标志和 T-TIC 标志没有被设置,则下一 TIDAW 定位于下一存储位置(例如,他与当前 TIDAW 连续)。

[0066] 现在转到图 10,现在将根据示例性实施例并引用图 1 的 I/O 处理系统 100 来描述为 I/O 操作聚集数据的处理 1000。在框 1002,通道子系统 108 接收用于 I/O 操作的控制字(例如 TCW)。该控制字包括间接数据地址,其指向构成由 I/O 操作利用的数据的存储地址列表(例如具有多个 TIDAW 的 TIDAL)的开始地址。在示例性实施例中,数据是客户数据(例如从 I/O 操作输入或输出的数据)。在备选示例性实施例中,数据是控制数据(例如 TCCB)。在示例性实施例中,存储地址列表跨越两个或更多个不连续的存储位置。

[0067] 在框 1004,通过定位于通道子系统 108 的指令聚集数据。聚集是基于列表的内容。在示例性实施例中,列表中的每个项(例如每个 TIDAW)包括存储地址和 T-TIC 标志,以指

示存储地址是否为数据的一部分的地址,或存储地址是否指向含有更多存储地址的列表的另一部分的位置。这样,单一 TIDAL 可包含在多个页面中,以排除在 TIDAL 的长度方面的任何限制。当 T-TIC 标志指示存储地址为数据的一部分的位置(例如 T-TIC 标志没有被设置)时,在该存储位置的数据被访问并增加至数据。新数据的这种增加可通过本领域已知的任意方式执行,例如与已经聚集的数据合并,附加至已经聚集的数据等。

[0068] 当 T-TIC 标志指示当前存储地址为列表的另一部分的位置时,则处理通过访问位于特定存储位置的新 TIDAW 来继续。访问在新 TIDAW 中指定的存储位置处的数据,并将其增加至数据。在示例性实施例中,TIDAW 包括计数 904,以指定要从每个存储地址读取(或写入)多少数据。当前设置了具有最后 TIDAW 标志的 TIDAW 时,完成聚集。否则,访问下一 TIDAW,并且数据聚集继续。

[0069] 在框 1006,通过通道子系统 108 向控制单元 110 发送数据。

[0070] 示例性实施例的技术效果包括在 I/O 子系统处提供不连续间接数据寻址。通过允许多个 4k 字节页面包含存储地址的列表(TIDAL)使得 TIDAW 的数目不受限制,因此更多数据可与单一 I/O 操作关联。数据的大量传送可通过避免附加握手和与多条更小的消息关联的其他延迟来减少通信负担。

[0071] 以下实例描述了在其中可利用示例性实施例来执行不仅仅移动较大数据块的方式。

[0072] 示例性实施例可用于通过操作系统来辅助通道程序加前缀。例如,当通道程序通过操作系统时,操作系统可修改通道程序,以增加、替换、或修改命令 CCW 或 DCW。在示例性实施例中,这暗示了将创建 TCCBTIDAL,其中第一 TIDAW 指向含有修改的 TCCB 的第一部分的存储区域。第二 TIDAW 指向未修改的 TCCB 的剩余部分。

[0073] 示例性实施例也可用于通过管理程序来辅助通道程序加前缀。如果管理程序需要修改通道程序,则其将需要创建具有比访客传递的 TIDAL 多一项的 TIDAL。如果访客传递的 TIDAL 已经是最大的大小(256TIDAW),则管理程序可通过创建具有 2 个 TIDAW(其中一个是对访客传递的 TIDAL 的第一 TIDAW 的 T-TIC)的 TCCB TIDAL 来作出其自身的修改。

[0074] TIDAL T-TIC 也可改善操作系统的存储需求。如果含有 TCCBTIDAL 的通道程序需要被修改,则相比于调用者的 TCCB TIDAL 加又一个 TIDAW 的分配的存储空间,其采用更少的存储空间来创建 TIDAW 和 T-TIC。

[0075] 如上所述,实施例可通过用于实践这些处理的计算机实现的处理和装置的形式来实现。在示例性实施例中,本发明在一个或多个网络元件执行的计算机程序代码中实现。实施例包括如图 11 所示的在计算机可用介质 1102 上的计算机程序产品 1100,其具有计算机程序代码逻辑 1104,包含在有形介质中作为制品实现的指令。用于计算机可用介质 1102 的示例性制品可包括软盘、CD-ROM、硬盘驱动器、通用串行总线(USB) 闪存驱动器、或任意其他计算机可读存储介质,其中,当计算机程序代码逻辑 1104 加载至计算机中并由其执行时,计算机成为实践本发明的装置。

[0076] 实施例包括计算机程序代码逻辑 1104,例如可存储在存储介质中,加载至计算机中和/或由其执行,或通过某些传输介质(例如在电线或电缆上,通过光纤,或经由电磁辐射)来发送,其中,当计算机程序代码逻辑 1104 加载至计算机中并由其执行时,计算机成为实践本发明的装置。当在通用微处理器上实现时,计算机程序代码逻辑 1104 分段配置微处

理器,以创建特定的逻辑电路。

[0077] 尽管参照示例性实施例描述了本发明,但是本领域普通技术人员应理解,在不脱离本发明的范围的情况下,可作出各种改变,以及等价物可替换其元素。此外,可作出许多修改,以在不脱离本发明实质范围的情况下,可针对本发明的教导采用特定情形或材料。因此,目的在于本发明不限于为了执行本发明而设想的最优模式而公开的特定实施例,但是本发明将需要落入所附权利要求的范围内的所有实施例。此外,术语第一、第二等的使用没有指定任意顺序或重要性,相反术语第一、第二等用于一个元件与另一个元件区分。此外,术语一、一个的使用没有指定数量的限制,相反指定存在至少一个所引用物体。

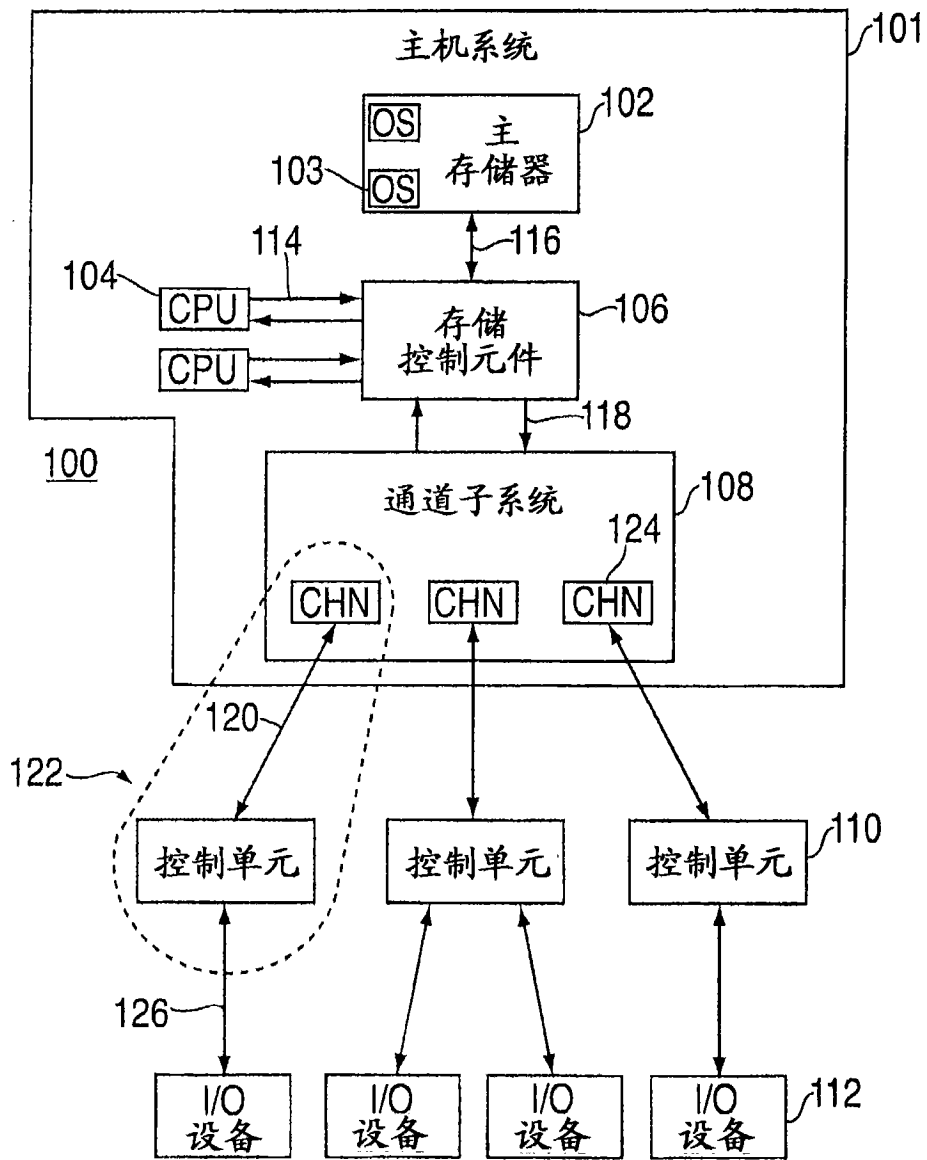


图 1

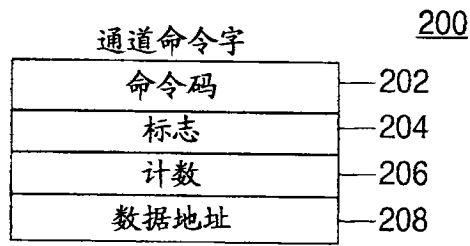


图 2A  
现有技术

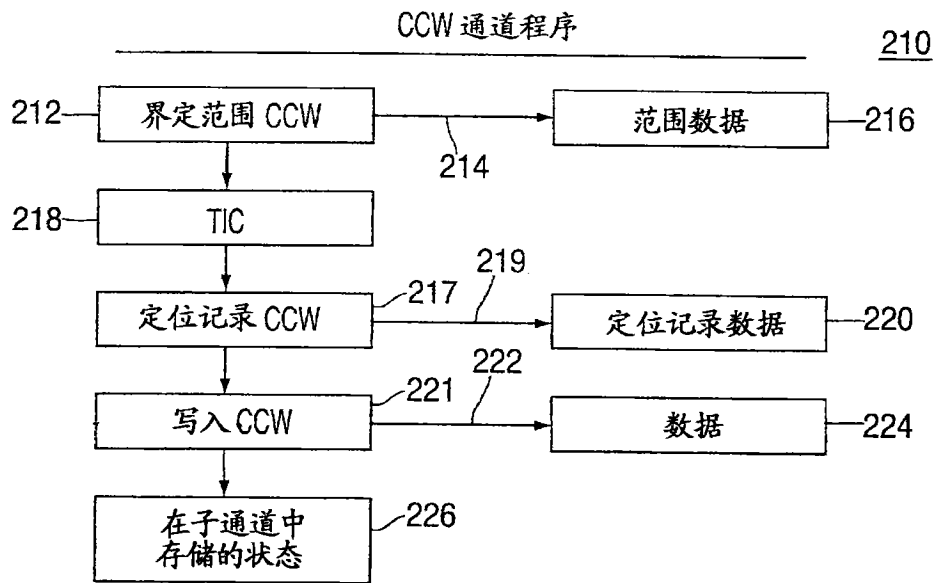


图 2B  
现有技术

图 2A

现有技术  
图 2B  
现有技术

对于4K写入CCW通道  
程序的链路协议

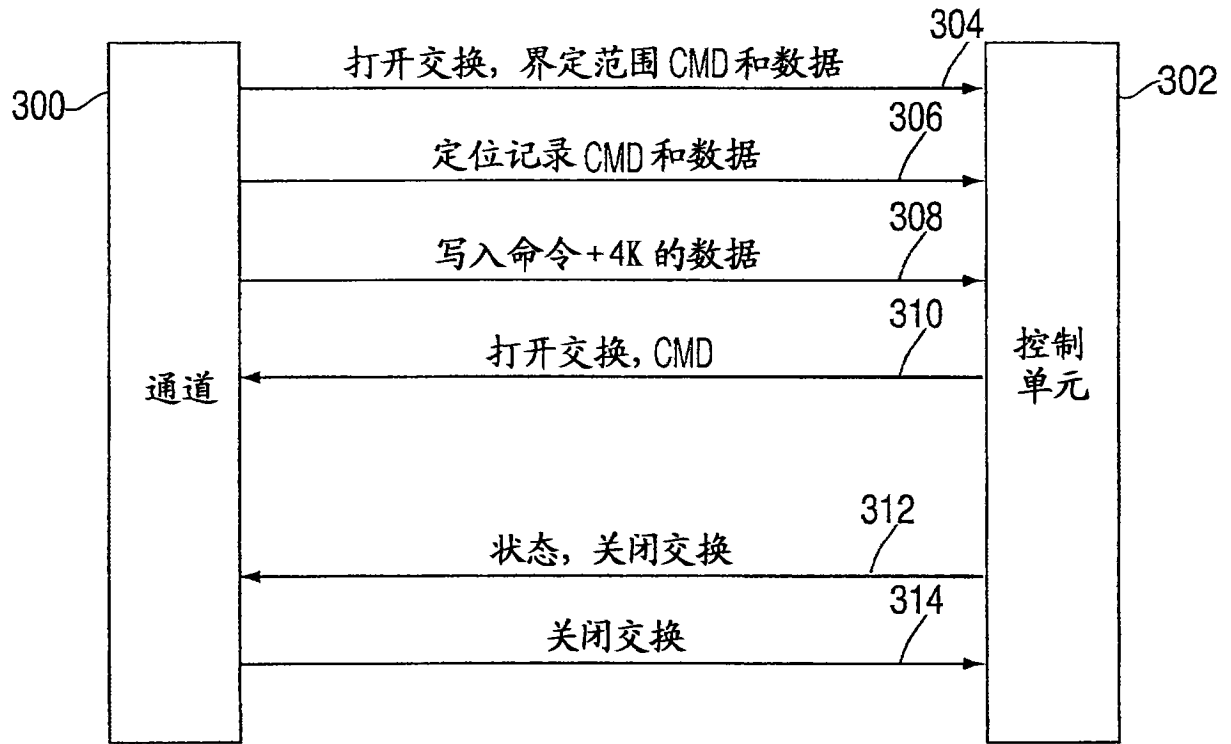


图 3

现有技术



TCW 通道程序

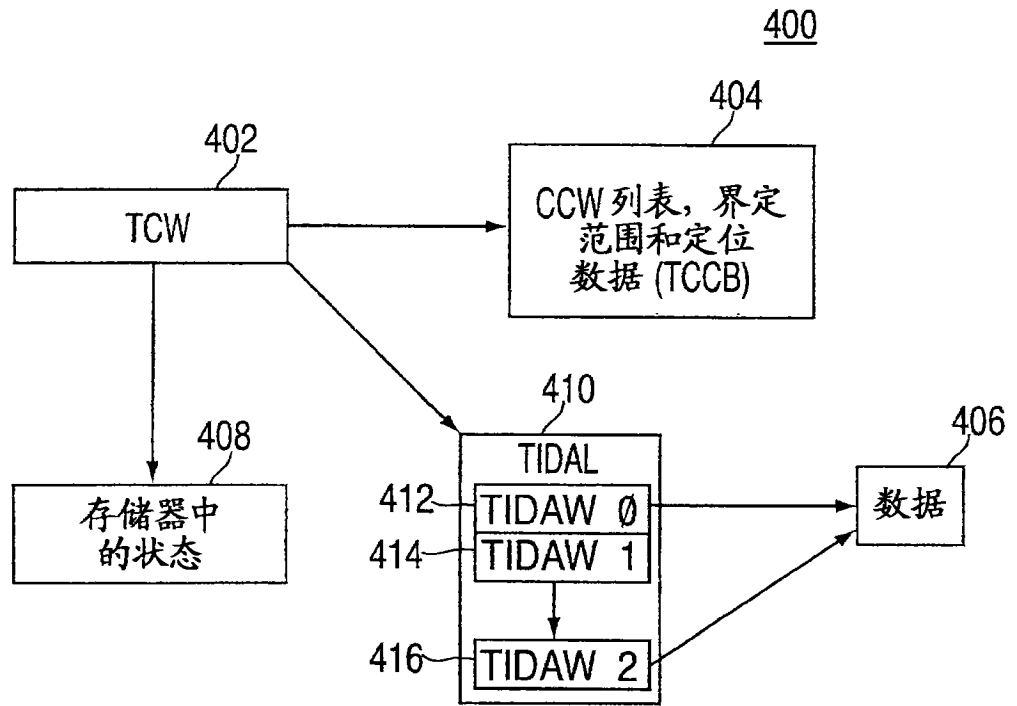


图 4

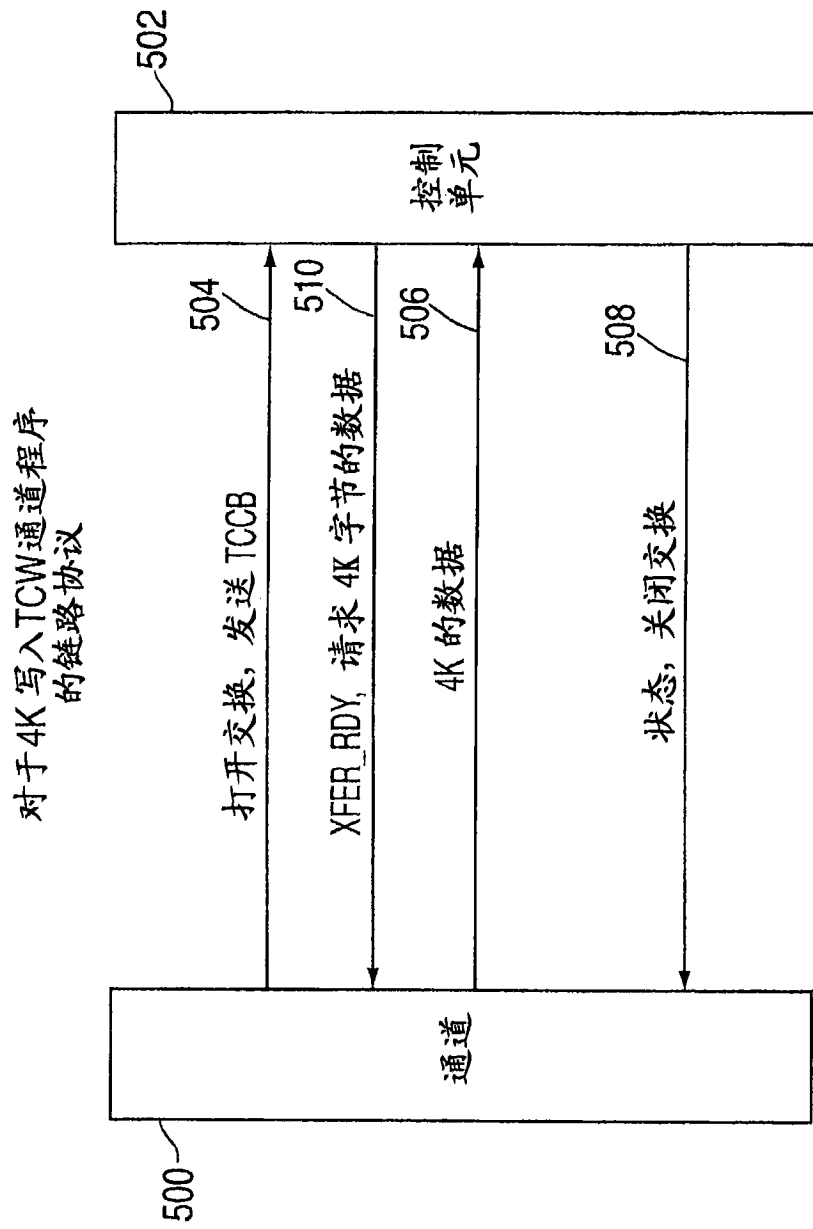


图 5

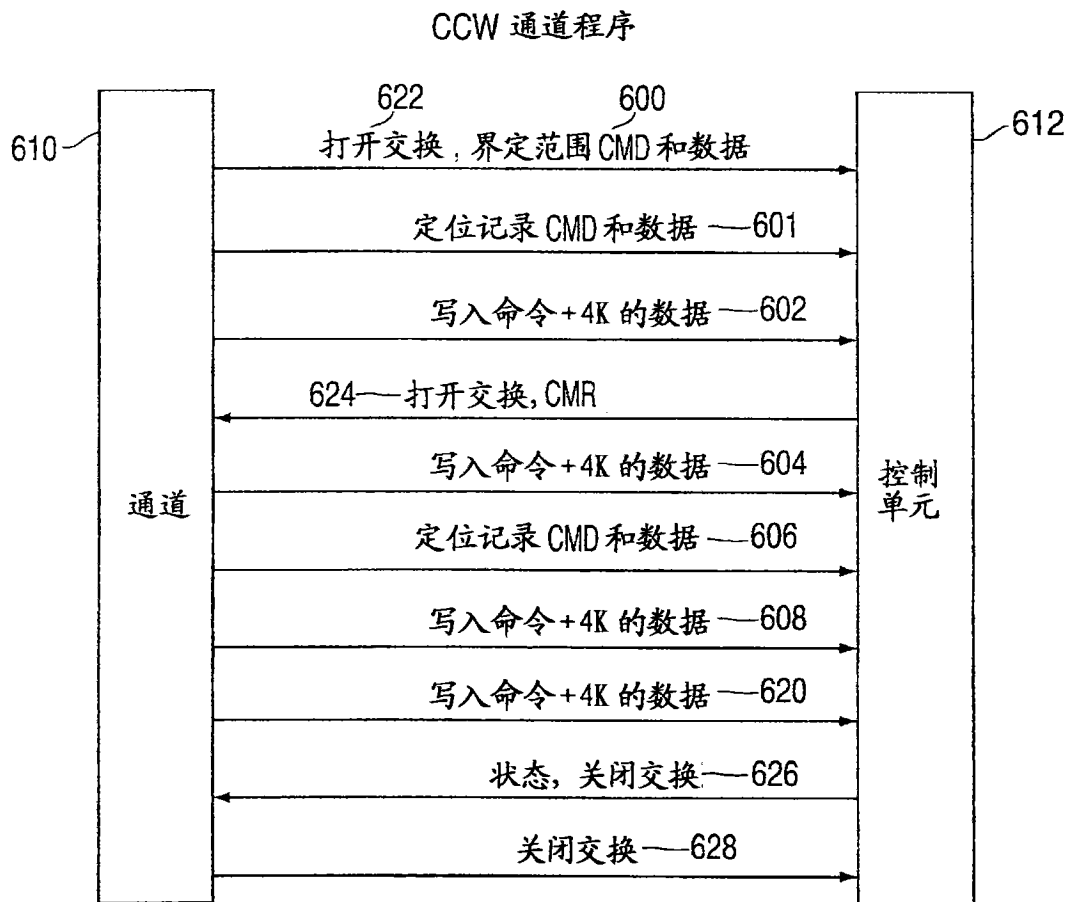


图 6

现有技术

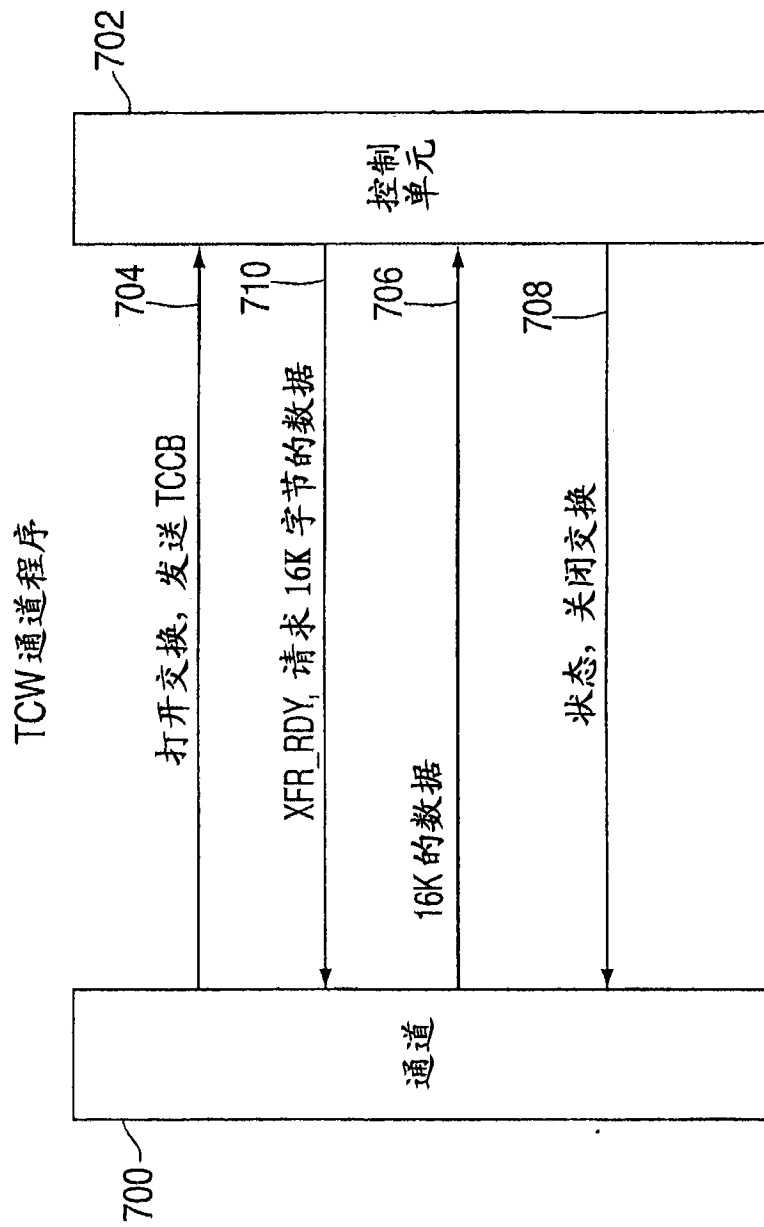


图 7

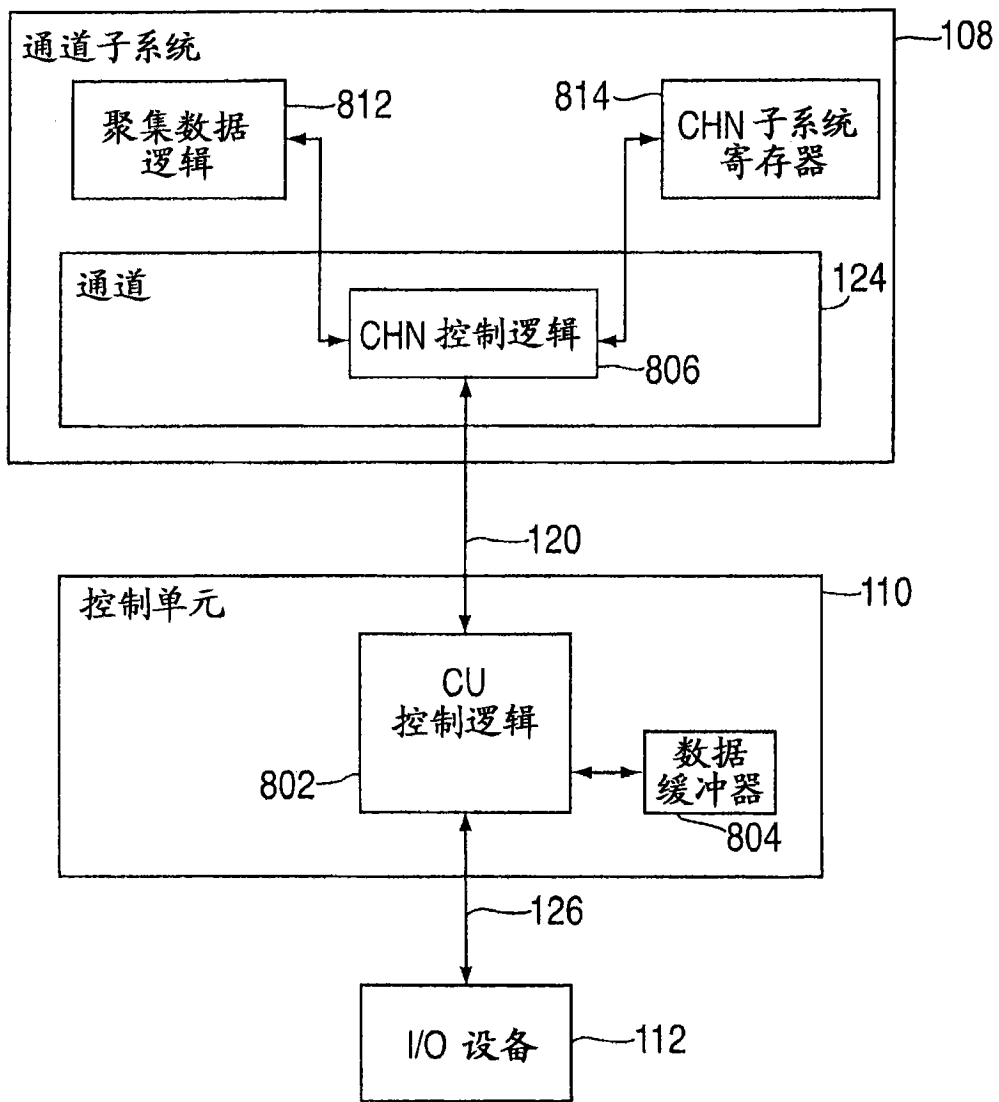


图 8

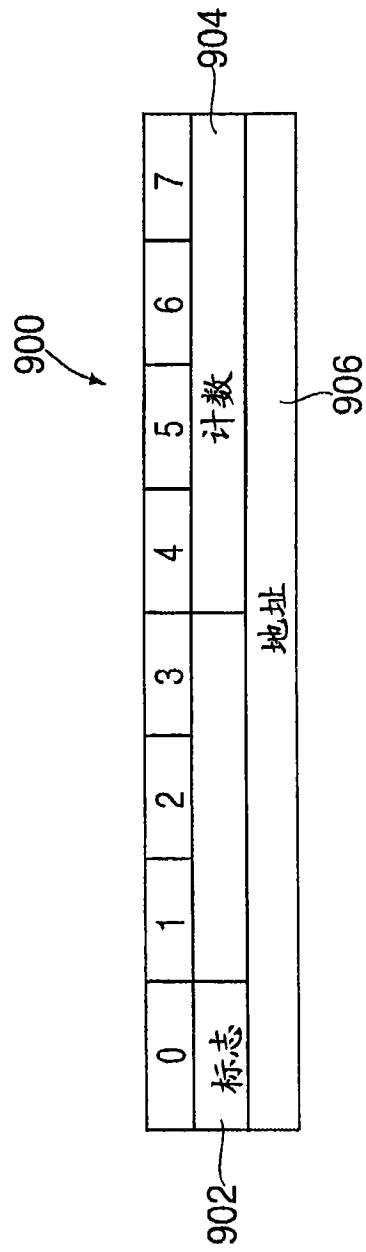


图 9

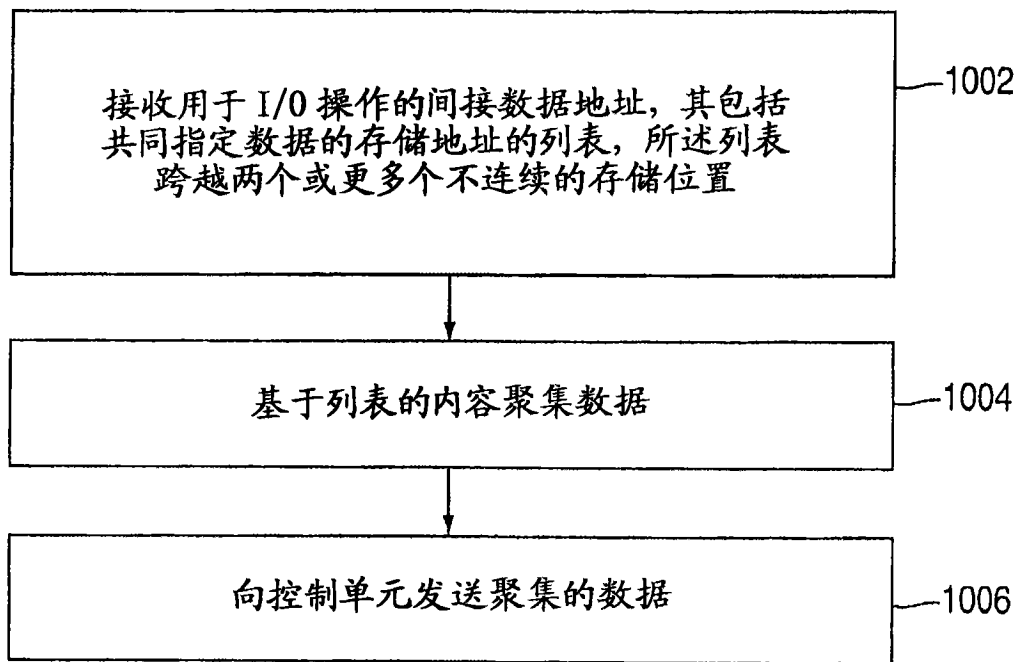


图 10

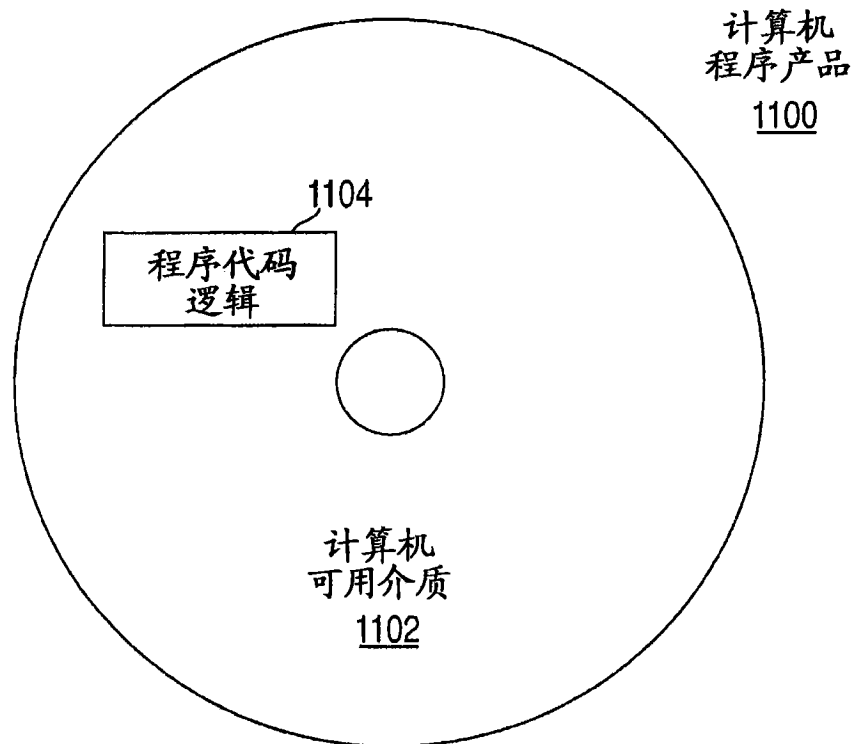


图 11