



(12) 发明专利

(10) 授权公告号 CN 101681322 B

(45) 授权公告日 2012. 09. 05

(21) 申请号 200880015506. 2

(51) Int. Cl.

(22) 申请日 2008. 06. 25

G06F 13/16 (2006. 01)

(30) 优先权数据

11/769, 001 2007. 06. 27 US

(56) 对比文件

US 2006/0245226 A1, 2006. 11. 02, 全文.

US 2002/0084458 A1, 2002. 07. 04, 全文.

US 2004/0256638 A1, 2004. 12. 23, 全文.

(85) PCT申请进入国家阶段日

2009. 11. 10

审查员 赵婷

(86) PCT申请的申请数据

PCT/EP2008/058082 2008. 06. 25

(87) PCT申请的公布数据

W02009/000857 EN 2008. 12. 31

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 G·K·巴特利 J·M·博肯哈根

P·R·格尔曼

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 于静 李峥

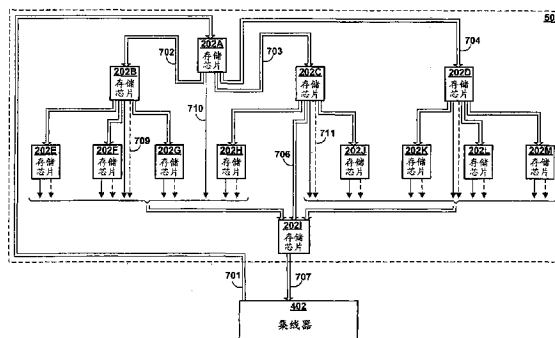
权利要求书 2 页 说明书 20 页 附图 9 页

(54) 发明名称

支持命令数据复制的高容量存储子系统的存储芯片

(57) 摘要

一种存储器模块, 包含用于接收数据访问命令的第一接口和用于将数据访问命令重传到其他存储器模块的第二接口, 所述第二接口将所接收的数据访问命令的多个副本传播到多个其他存储器模块。所述存储器模块优选地在组织为其中交织数据访问的树配置的高容量存储子系统中使用。优选地, 所述存储器模块具有多模式操作, 其中一种模式支持多次命令复制, 而另一模式支持传统的菊链。



1. 一种第一存储器模块,包括:

多个用于存储数据的可寻址存储单元;

第一接口,其用于接收访问所述可寻址存储单元中存储的数据的数据访问命令;

控制逻辑单元,其支持多个操作模式,所述多个操作模式包括第一操作模式和第二操作模式,在所述第一操作模式中,所述第一存储器模块重传在所述第一接口中接收的至少一些所述数据访问命令的多个副本,使用多个第二接口中的相应接口传输每个副本,每个副本被传输到不同于所述第一存储器模块的一个或多个相应存储器模块,其中所述第一存储器模块包括所述多个第二接口;

并且,在所述第二操作模式中,所述控制逻辑单元判定由所述第一接口接收的每个所述数据访问命令是否寻址所述第一存储器模块中的数据存储单元,并且 (a) 如果所述数据访问命令寻址所述第一存储器模块中的数据存储单元,则访问所述第一存储器模块中的对应数据存储单元,以及 (b) 如果所述数据访问命令未寻址所述第一存储器模块中的数据存储单元,则将所述数据访问命令的单个副本重传到另一存储器模块,

其中,在所述第一操作模式中,所述第一存储器模块的第一组组件以第一时钟频率运行,并且其中,在所述第二操作模式中,所述第一存储器模块的所述第一组组件以大于所述第一时钟频率的第二时钟频率运行。

2. 如权利要求 1 中所述的第一存储器模块,其中在所述第一操作模式中,所述第一存储器模块重传在所述第一接口中接收的每个所述数据访问命令的多个副本,使用多个所述第二接口中的相应接口传输每个副本,每个副本被传输到不同于所述第一存储器模块的一个或多个相应存储器模块。

3. 如权利要求 1 中所述的第一存储器模块,其中在所述第一操作模式中,由所述第一接口接收的每个所述数据访问命令访问一组存储器模块中的数据单元,所述组包含多个存储器模块,所述组包括所述第一存储器模块以及从所述第二接口向其重传所述数据访问命令的副本的每个存储器模块。。

4. 如权利要求 1 中所述的第一存储器模块,其中所述控制逻辑单元支持多个操作模式,所述多个操作模式包括:第一操作模式,其中所述第一存储器模块重传在所述第一接口中接收的至少一些所述数据访问命令的多个副本;以及第二操作模式,其中所述第一存储器模块不重传任何所述数据访问命令的多个副本。

5. 如权利要求 1 中所述的第一存储器模块,

其中,在所述第一操作模式中,所述第一接口用于耦合到将所述第一存储器模块与其接收所述数据访问命令的设备相连的点到点通信链路;以及

其中,在所述第一操作模式中,每个所述第二接口用于耦合到将所述第一存储器模块与另一相应存储器模块相连的相应点到点通信链路。

6. 如权利要求 1 中所述的第一存储器模块,

其中,在所述第一操作模式中,所述第一接口在 N 个连续总线周期的每个周期中接收数据访问命令的相应部分,其中 N 大于 1;以及

其中,在所述第一操作模式中,每个所述第二接口在 M 个总线周期延迟之后在 N 个连续总线周期的每个周期中重传所述数据访问命令的所述相应部分,其中 $1 \leq M < N$ 。

7. 如权利要求 1 中所述的第一存储器模块,其中,在所述第二操作模式中,所述第一存

存储器模块用于在存储器模块的菊链式配置中使用,其中第一组 I/O 端口用于从控制对存储器模块菊链的访问的访问控制模块之一接收出站数据,所述菊链包括所述第一存储器模块以及所述存储器模块菊链中的上一存储器模块,第二组 I/O 端口用于将在所述第一组 I/O 端口中接收的出站数据重传到所述存储器模块菊链中的下一存储器模块,第三组 I/O 端口用于接收来自所述存储器模块菊链中的所述下一存储器模块的进站数据,以及第四组 I/O 端口用于重传在所述第一组 I/O 端口中接收的入 站数据并用于将所述第一存储器模块中发起的进站数据传输到所述访问控制模块以及所述存储器模块菊链中的所述上一存储器模块中的一个,所述第一组 I/O 端口的端口数与所述第二组 I/O 端口的端口数相同,所述第三组 I/O 端口的端口数与所述第四组 I/O 端口的端口数相同。

8. 如权利要求 1 中所述的第一存储器模块,其中所述第二时钟频率是所述第一时钟频率的 M 倍,其中 M 是整数。

9. 如权利要求 1 中所述的第一存储器模块,其中从外部时钟信号得到所述第一和第二时钟频率,将所述外部时钟信号除以整数 N 以获得所述第一时钟频率并且将所述外部时钟信号除以整数 M 以获得所述第二时钟频率,其中 M 大于或等于 1,并且 N 大于 M。

10. 如权利要求 1 中所述的第一存储器模块,其中所述第一组组件包括至少一个 I/O 端口。

11. 如权利要求 10 中所述的第一存储器模块,其中所述第一组组件的所述至少一个 I/O 端口用于响应于所述数据访问命令,将从所述第一存储器模块中的存储单元读取的数据传输到访问控制模块。

12. 如权利要求 1 中所述的第一存储器模块,

其中,在所述第一操作模式中,第一组 I/O 端口用于响应于所述数据访问命令,将从所述第一存储器模块中的存储单元读取的数据传输到访问控制模块;以及

其中,在所述第二操作模式中,第二组 I/O 端口用于响应于所述数据访问命令,将从所述第一存储器模块中的存储单元读取的数据传输到访问控制模块,所述第一组 I/O 端口包含至少一个 I/O 端口,所述第二组 I/O 端口包含的 I/O 端口数大于所述第一组 I/O 端口包含的 I/O 端口数。

13. 如权利要求 1 中所述的第一存储器模块,

其中,在所述第一操作模式中,第一组 I/O 端口用于响应于所述数据访问命令,将从所述第一存储器模块中的存储单元读取的数据传输到访问控制模块,所述第一组 I/O 端口响应于所述数据访问命令,需要 N 个总线周期来传输一个可访问的数据单位;以及

其中,在所述第二操作模式中,第二组 I/O 端口用于响应于所述数据访问命令,将从所述第一存储器模块中的存储单元读取的数据传输到访问控制模块,所述第二组 I/O 端口响应于所述数据访问命令,需要 M 个总线周期来传输一个可访问的数据单位,其中 N 大于或等于 1,并且其中 M 大于 N。

支持命令数据复制的高容量存储子系统的存储芯片

技术领域

[0001] 本发明涉及数字数据处理硬件,更具体地说,涉及数字数据处理系统中的存储系统和存储器互连的设计和操作。

背景技术

[0002] 在二十世纪的后半叶,出现了一种称为信息革命的现象。虽然信息革命是一种在范围上较任何一个事件或机械革命都广泛的历史发展,但是没有一种单独的设备比数字电子计算机更能代表信息革命。计算机系统的发展无疑也是一场革命。每年,计算机系统都迅猛发展、存储更多的数据,并为用户提供更多的应用。

[0003] 现代计算机系统通常包括一个或多个中央处理单元(CPU),并支持存储、检索和传输信息所需的硬件,例如通信总线和存储器。它还包括与外界通信所需的硬件,例如输入/输出控制器或存储控制器,以及与所述计算机系统相连的设备,例如键盘、监视器、磁带驱动器、盘驱动器、与网络相连的通信线路等。CPU是系统的核心。它执行包括计算机程序的指令并引导其他系统组件的操作。

[0004] 从计算机硬件的角度来看,多数系统的运行方式基本相同。处理器能执行一组有限的非常简单的操作,例如算术运算、逻辑比较以及将数据从一个位置移到另一位置。但是每种操作的执行速度都非常快。引导计算机执行大量此类简单操作的程序会造成计算机正在执行复杂操作的错觉。用户所认为的新的或改进的计算机系统能力基本上是通过执行同一组非常简单的操作来实现的,但是执行操作的速度要快得多。因此,计算机系统的不断改进要求这些系统速度更快。

[0005] 计算机的CPU对计算机的可寻址主存储器中存储的数据进行操作。存储器同时存储在处理器中执行的指令以及由这些指令操纵的数据。在操作中,处理器不断访问存储器中的指令和其他数据,不执行此操作便无法执行有用的工作。存储子系统及其运行速度的设计是所有计算机系统的整体性能中的关键问题。

[0006] 存储器通常包含在一组集成电路模块中。访问存储器所需的时间不只是存储器模块自身运算速度的函数,而且还是处理器与存储器间的路径的速度的函数。随着计算机变得更加复杂,此路径会占用更多的访问时间份额。早期的计算机只有一个处理器以及相对较小的存储器,从而使得处理器与存储器间的路径相对直接。现代大型系统通常包含多个处理器,多级高速缓存、复杂的寻址机制以及非常大的主存储器来支持系统的数据需求。在这些系统中,从每个处理器到每个存储器模块之间不可能再有直接的路径。复杂的总线结构支持数据在各种系统组件间的移动。通常,数据必须穿过处理器与实际存储器模块间的若干结构。随着处理器数量和存储器大小的增加,这些问题变得更加突出。

[0007] 为了获得规模生产效益并降低计算成本,集成电路存储器模块已变成一种商品,具有标准化的外部接口、存储器容量以及其他参数。诸如存储控制器、总线、中继器之类的访问存储器的其他计算机系统组件被设计为与这些标准化的存储芯片协同工作。标准化要求外部接口设计的某些方面在一段时间内固定,尽管可能对内部设计做出改进。在设计保

持固定期间,计算机行业的技术能力及生产需求将继续改进。在某种程度上,这种能力与期望的演进证明根据标准设计的新一代存储芯片更适合当前技术水平与行业要求。

[0008] 标准化存储器模块的设计可以针对多种计算机体系结构和使用中的任意一种进行优化。预计未来的需求很大程度上由通常称为低端系统的诸如单用户台式机和膝上型计算机之类的高容量、小型、通用计算机系统,以及诸如游戏系统、视频系统等专用设备来推动。将继续制造被称为“大型机”的计算机系统和其他大型系统,但是这些系统将占总体存储芯片需求中相对较小的部分。因此有理由认为未来的存储器模块标准将由低端、高容量市场来推动,并针对在该市场的典型设备中的使用进行优化。

[0009] 如果针对低端系统来优化未来标准化存储器模块的设计,则当这些模块在具有不同设计参数的大型系统中使用时将缺乏效率。可以针对在大型系统中的使用设计一组单独的存储器模块,但是这样会失去使用高容量、标准化存储器模块的规模效益,并在实质上增加大型系统的成本。

[0010] 需要一种改进的存储子系统设计技术来实现在广泛的系统中使用标准化存储器模块,同时在不过分丧失效率的情况下满足不同类型系统的操作需求。

发明内容

[0011] 一种存储器模块,包含用于接收数据访问命令的第一接口和用于将数据访问命令重传到其他存储器模块的第二接口,所述第二接口将所接收的数据访问命令的多个副本传播到多个其他存储器模块。

[0012] 在优选实施例中,所述存储器模块用于在高容量存储子系统中使用,所述存储子系统包括具有至少一个存储芯片总线的存储控制器,所述存储芯片总线以全频和全部总线宽度运行并且与多个采用菊链式配置的集线器再驱动芯片耦合,每个集线器再驱动芯片都使用全频和全部总线宽度与所述链中的下一集线器再驱动芯片通信。每个集线器再驱动芯片都支持存储交织数据的缓冲存储芯片的至少一个相应群集,所述群集排列在至少一个树中。命令和写入数据沿着树向下传播,在树的每个接连的级别处芯片数量增加。所述存储芯片优选地设计为以全部总线宽度、全总线频率模式(低端系统的典型模式)在菊链式配置中操作,以及设计为在具有减少的总线宽度和/或减少的总线频率的至少一个单独模式中操作,以便用于在优选实施例的交织树配置中使用。

[0013] 优选地,所述存储器模块具有各种体系结构特性并在此处所述以及上面的各种相关申请交叉引用中所要求保护的高容量交织树配置中使用。但是应理解,本发明并非一定限于那些采用相关申请中要求保护的特性的实施方式,备选地,可以构造与本发明一致且不使用相关申请中所要求保护的部分或全部特性的存储器模块。

[0014] 通过根据优选实施例配置可用于交织树配置中的菊链的存储芯片,在不使用定制存储芯片的情况下,可以将更多存储器配置到存储控制器所支持的有限数量的总线。此外,所述交织配置可以通过降低许多(尽管不一定是全部)总线的总线操作频率,以及通过减少实际使用的 I/O 端口数来实现显著的功率节约。

[0015] 通过参考附图,可以最佳地了解本发明的结构与操作的细节,在所述附图中,相同的标号代表相同的部件,其中:

附图说明

[0016] 图 1 是根据本发明优选实施例的使用具有缓冲存储芯片的存储子系统的计算机系统的主要硬件组件的高级方块图；

[0017] 图 2 是其中使用菊链式配置的缓冲存储芯片来配置存储子系统的计算机系统的典型存储子系统的主要硬件组件的方块图；

[0018] 图 3 是根据优选实施例的缓冲存储芯片的特定主要内部组件的示意图；

[0019] 图 4 是根据本发明优选实施例的使用集线器和群集配置的存储子系统的高级方块图；

[0020] 图 5 是更详细地示出根据优选实施例的特定变型的集线器与关联存储芯片群集间的链路的方块图；

[0021] 图 6-8 更详细地示出分别根据优选实施例的第一、第二和第三变型的子群集的存储芯片与关联集线器间的数据路径的各种配置；

[0022] 图 9 是示出根据优选实施例的支持一个或多个存储芯片群集的集线器的特定主要内部组件的方块图。

具体实施方式

[0023] 未来的存储芯片概述

[0024] 低端系统通常需要大量存储器带宽,即在给定时间内读取和写入大量存储器的能力,但是不一定需要大量存储器容量。为了满足这些要求,本发明构想了缓冲存储芯片和支持存储芯片链的存储芯片体系结构。此类体系结构在下面共同受让的共同未决专利申请中描述,每个申请都在此引入作为参考:2006 年 7 月 26 提交的标题为“Daisy Chained Memory System”的美国申请第 11/459,956 号;2006 年 7 月 26 提交的标题为“Memory System Having Self Timed Daisy Chained Memory Chips”的美国申请第 11/459,957 号;2006 年 7 月 26 提交的标题为“Carrier Having Daisy Chained Memory Chips”的美国申请第 11/459,969 号;2006 年 7 月 26 提交的标题为“Carrier Having Daisy Chain of Self Timed Memory Chips”的美国申请第 11/459,983 号;2006 年 7 月 26 提交的标题为“Daisy Chainable Memory Chips”的美国申请第 11/459,994 号;2006 年 7 月 26 提交的标题为“DaisyChainable Self Timed Memory Chip”的美国申请第 11/459,997 号;2006 年 7 月 26 提交的标题为“Computer System Having Daisy Chained Memory Chips”的美国申请第 11/459,974 号;2006 年 7 月 26 提交的标题为“Computer System Having Daisy Chained Self Timed Memory Chips”的美国申请第 11/459,968 号;2006 年 7 月 26 提交的标题为“Memory Controller for Daisy Chained Memory Chips”的美国申请第 11/459,966 号;2006 年 7 月 26 提交的标题为“Memory Controller for Daisy Chained SelfTimed Memory Chips”的美国申请第 11/459,961 号;2006 年 7 月 26 提交的标题为“Memory Chip Having an Apportionable Data Bus”的美国申请第 11/459,943 号;2006 年 7 月 26 提交的标题为“Self Timed Memory Chip Having an Apportionable Data Bus”的美国申请第 11/459,947 号;2006 年 7 月 26 提交的标题为“Computer System Having an Apportionable Data Bus”的美国申请第 11/459,955 号;以及 2006 年 7 月 26 提交的标题为“Memory System Having an Apportionable Data Bus and Daisy Chained Memory Chips”的美国申请第 11/459,959

号。

[0025] 如其中所述,缓冲存储芯片被设计为在菊链式配置中使用。所述存储芯片具有两组高频通信接口。这些接口旨在通过相应点到点通信链路(总线)连接到其他存储芯片或存储控制器。一个点到点链路将芯片与菊链上的下一上游设备相连,后者可以是另一芯片,也可以是存储控制器。其他点到点链路将芯片与下一下游存储芯片(如果其存在)相连。点到点链路的菊花链接无需存储控制器与存储芯片间的传统缓冲芯片,确保所有链路都是点到点的,因此便于总线以更高的频率运行。

[0026] 尽管每个链路具有多个数据线路,但是链路按顺序方式运行,使用这种方式需要多个总线周期来传输单个命令或数据字。每个芯片中的缓冲器临时存储在总线上传输的数据字和命令的各部分。

[0027] 菊链设计对用于在存储芯片中存储数据的内部存储技术没有任何限制。虽然也可以使用静态 RAM,但通常最多是使用动态随机存取存储器(DRAM)。此外,任何未来对内存存储技术的改进或新技术一般都适合此处所述的缓冲存储芯片配置的基本框架。

[0028] 详细说明

[0029] 现在参考附图,其中在多个图中,相同的标号表示相同的部分,图 1 是根据优选实施例的具有使用缓冲存储芯片的存储子系统的计算机系统 100 的主要硬件组件的高级表示。计算机系统 100 的主要组件包括一个或多个中央处理单元(CPU)101A-101D,主存储器系统 102、高速缓冲存储器 106、终端接口 111、存储器接口 112、I/O 设备接口 113,以及通信/网络接口 114,上述所有组件都经由总线 103、104 以及总线接口 105 进行耦合以实现组件间通信。

[0030] 系统 100 包含一个或多个通用可编程中央处理单元(CPU)101A-101D,此处统称为部件 101。在优选实施例中,系统 100 包含多个处理器,这在相对较大的系统中是典型的;但是,系统 100 可以备选地为单 CPU 系统。每个处理器 101 都执行存储器 102 中存储的指令。指令和其他数据被从主存储器 102 加载到高速缓冲存储器 106 中进行处理。主存储器 102 是用于存储包括程序的数据的随机存取半导体存储器。尽管主存储器 102 和高速缓存 106 在图 1 中从概念上被表示为单个实体,但是可以理解这些组件实际上更加复杂,并且高速缓存可以存在于多个不同的级别,如本领域中的技术人员公知的。具体而言,主存储器系统 102 包括多个模块和通信组件,如将更详细地说明的。

[0031] 总线 103-105 提供各个系统组件之间的通信路径。处理器/存储器总线 103(此处称为前端总线)提供用于在 CPU 101、高速缓存 106、主存储器 102 以及 I/O 总线接口单元 105 之间传输数据的数据通信路径。I/O 总线接口 105 进一步耦合到系统 I/O 总线 104 以便向/从各个 I/O 单元传输数据。I/O 总线接口 105 通过系统 I/O 总线 104 与多个 I/O 接口单元 111-114 通信,这些接口单元也称为 I/O 处理器(IOP)或 I/O 适配器(IOA)。系统 I/O 总线可以例如是行业标准 PCI 总线或任何其他适当的总线技术。

[0032] I/O 接口单元 111-114 支持与各种存储设备和 I/O 设备通信。例如,终端接口单元 111 支持与一个或多个用户终端 121-124 相连。存储接口单元 112 支持与一个或多个直接存取存储设备(DASD)125-127(它们通常为旋转磁盘驱动存储设备,但是也可以备选地为其他设备,包括被配置为对于主机而言是单个大型存储设备的盘驱动器阵列)相连。I/O 设备和其他设备接口 113 提供到任意其他各种输入/输出设备或其他类型的设备的接口。图

1 的示例性实施例中示出了两个此类设备：打印机 128 和传真机 129，可以理解，可存在许多其他此类设备，它们可以具有不同的类型。网络接口 114 提供从系统 100 到其他数字设备和计算机系统的一个或多个通信路径；此类路径可以例如包括诸如因特网、局域网之类的一个或多个网络 130 或其他网络，也可以包括远程设备通信线路、无线连接等。

[0033] 应理解的是，图 1 旨在在较高级别上示出系统 100 的代表性主要组件，个别组件可以比图 1 中所示的更加复杂，可以存在图 1 所示的组件以外的组件或存在替代图 1 所示的组件的组件，此类组件的数量、类型和配置可以变化。可以进一步理解在特定计算机系统中，不一定存在图 1 所示的所有组件。本文披露了具有此类附加复杂性或附加变型的若干特定实例，应该理解它们仅作为实例，并非一定只有此类变型。

[0034] 尽管前端总线 103 在图 1 中被示为提供高速缓存 106、主存储器 102 和 I/O 总线接口 105 间的直接通信路径的相对简单的单总线结构，但实际上前端总线 103 可以包括多个不同的总线或通信路径，后者可以使用任何不同的形式进行排列，例如具有层次结构的点到点链路、星型或网状配置，多层次结构总线、平行和冗余路径等。此外，尽管 I/O 总线接口 105 和 I/O 总线 104 被示为单个独立的单元，但是系统 100 实际上可包含多个 I/O 总线接口单元 105 和 / 或多个 I/O 总线 104。虽然示出的多个 I/O 接口单元将系统 I/O 总线 104 与通往各个 I/O 设备的各个通信路径隔开，但是可以备选地将部分或全部 I/O 设备直接连接到一个或多个系统 I/O 总线。

[0035] 主存储器 102 在图 1 中被示为单块实体，但是应该理解主存储器可具有更复杂的结构。例如，主存储器可以是分布式的并与不同的 CPU 或 CPU 组关联，如各种所谓的非一致性存储器存取 (NUMA) 计算机体系结构的任意一种体系结构中公知的，也可以被分为由共同构成前端总线 103 的单独总线访问的分离子集，或者可以形成其他某种体系结构。类似地，尽管高速缓存被示为单个实体，但是可以存在多个层级的高速缓存，其中一些由 CPU 101A-101D 中的全部或部分共享，一些专用于相应单个 CPU。此外，可以按功能划分高速缓存，以使一个高速缓存保存指令，而另一高速缓存保存由一个或多个处理器使用的非指令数据。如此处使用的，“存储子系统”是存储器或高速缓存或它们的任意部分。存储子系统可以包括全部主存储器 102，或部分主存储器 102，或者全部或部分高速缓存 106。尽管本发明不限于在主存储器中使用并且也适于在某些高速缓冲存储器中使用，但是存储子系统特别优选地为全部或部分主存储器 102，因为高速缓存通常要求更快的存取速度。

[0036] 图 1 中示出的计算机系统 100 具有多个连接终端 121-124，例如多用户“大型机”计算机系统的典型终端。通常在这种情况下，连接设备的实际数量大于图 1 中所示。尽管预期此处所述的存储子系统配置最适于在相对较大的多用户系统中使用，但是本发明不限于任何特定大小的系统。计算机系统 100 可以备选地为单用户系统，通常只包含一个用户显示器和键盘输入，或者可以为没有或只有很少直接用户接口，但是从其他计算机系统（客户机）接收请求的服务器或类似设备。

[0037] 尽管在较高级别上描述和示出了各种系统组件，但是应该理解典型的计算机系统包含许多其他未示出的组件，其并不是理解本发明所必需的。

[0038] 图 2 是其中使用菊链式配置的缓冲存储芯片来配置存储子系统的计算机系统的典型存储子系统 102 的主要硬件组件的方块图。如上所述，在优选实施例中，使用针对菊链式配置设计的缓冲存储芯片。给出图 2 来作为背景以解释存储芯片所要采用的典型菊链式

配置,尽管在优选实施例中实际并非使用该配置。

[0039] 现在参考图 2,存储控制器 201 与系统前端总线 103 耦合以便与一个或多个处理器通信。存储控制器又支持一个或(典型地)多个存储芯片总线,后者被配置为相应多组菊花链接的缓冲存储芯片 202A-L(此处统称为部件 202)。尽管图 2 中示出了一个存储控制器和连接的存储芯片,但是主存储器可以包含多个存储控制器,每个控制器支持一组相应的连接存储芯片。

[0040] 每组菊花链接 203 的存储芯片 202 都包括从存储控制器到链中的第一缓冲存储芯片的点到点链路 204A,以及接下来的从链中的每个缓冲存储芯片到链中下一缓冲存储芯片,直到到达链的末尾的点到点链路 204B-D(点到点链路在此通常称为部件 204)。存储控制器 201 典型地支持多个点到点存储芯片链路 204,每个链路都能够支持相应一组菊花链接的缓冲存储芯片。尽管图 2 示出了三组菊链,但是数量可以更改,最少可以为一组,但通常多于一组。

[0041] 每个点到点链路 204 都包括出站数据部分 205 和进站数据部分 206。出站部分 205 用于沿着出站方向传输数据,即沿着远离存储控制器 201 的方向。在优选实施例中,出站部分包括一组命令/地址线路和一组单独的写入数据线路。命令/地址线路沿着出站方向将地址和命令传输到芯片。写入数据线路沿着出站方向传输写入数据,即数据来自存储控制器 201(其从前端总线 103 获取数据)并要被写入到在命令/地址线路上指定的存储地址。尽管在优选实施例中,将一组单独的专用线路用于命令/地址数据和用于写入数据,但是可以备选地使用一组共享的线路,后者使用时间多路复用的方式传输命令/地址数据和写入数据两者。进站数据部分 206 沿着进站方向传输读取的数据,即响应于先前在出站部分 205 的命令/地址线路上发送的读取命令而从存储地址读取的数据。

[0042] 在操作中,存储控制器 201 通过前端总线 103 接收命令,并且确定应用命令的菊链组 203,即,适合的地址所在的组。在本文说明中,为了简化,假设命令为到特定地址的读取命令或写入命令,尽管某些体系结构可支持其他命令类型。控制器 201 然后在适合菊链组中第一链路 204 的出站部分 205 的命令/地址线路上传输每个命令和地址(并且,对于写入命令,在出站部分 205 的写入数据线路上传输适合的写入数据)。该数据由第一存储器模块 202 接收,然后根据需要在后续的总线周期内被重传到菊链组中的下一模块。重传将持续进行,直到命令/地址数据(以及可选的写入数据)到达其适合的模块 202。此时,所述模块将数据写入指定的地址,或读取指定地址处的数据。如果是读取数据,则所述模块在进站数据部分 206 上向存储控制器传输读取的数据。该读取的数据在后续的总线周期中由菊链组中的每个中间模块进行重传,直到它到达存储控制器 201,所述数据经由前端总线 103 被从存储控制器传输到发出请求的实体。

[0043] 在优选实施例中,每个点到点链路 204 都为高速宽带串行链路,即,它包括多个并行信号线路,但是线路数在单个总线周期内不足以传输完整命令/地址或可访问的数据单位(即,在单个数据访问中传输的数据量)。需要多个总线周期来在出站部分 205 上传输单个命令/地址(以及可选的可访问的写入数据单位),或者在进站部分 206 上传输可访问的读取数据单位。在优选实施例中,一个可访问的数据单位为 8 个字节。出站总线部分 205 包括五个命令/地址数据信号线路,以及十三个写入数据信号线路。进站总线部分 206 包括十三个读取数据信号线路。需要 6 个速度为 6 千兆周/秒(6GT/s)的总线周期来传输

单个命令 / 地址 (最大为 30 位)、写入数据 (最大为 78 位, 其中包括 8 字节的可寻址数据以及最多 14 个辅助位), 以及读取数据 (最大为 78 位, 其中包括 8 字节的可寻址数据以及最多 14 个辅助位), 以使总线能够以 1 千兆 - 命令 / 秒 (giga-command/sec) 的速度传输命令。辅助位位置可用于纠错代码 (ECC) 或其他数据。在本文更详细地描述的数据交织的优选实施例中, 可使用一对或多对辅助位位置来支持对应的备用存储芯片, 后者可用于替代失效的存储芯片。为了进行说明, 本文的通篇说明中都使用这些参数。但是应该理解, 这些参数只是设计参数组的一种可能代表, 并且线路数、每个命令或数据所需的串行总线周期数以及总线频率等可以更改。应进一步理解, 存储芯片 202 可具有分配总线宽度能力, 其中可以可变地配置不同总线部分的宽度, 如上面引用的相关专利申请中所描述的。

[0044] 当数据在菊链组 203 上向内或向外移动时, 会被每个后续模块 202 接收, 在模块中进行缓冲, 并在后续总线周期内被从模块重传。因此, 存储器访问操作的延迟时间取决于对应存储器模块 202 在菊链组中的物理位置, 访问位于菊链末尾的模块需要的总线周期较多。这一事实对菊链组的长度具有实际限制。

[0045] 图 3 是根据优选实施例的缓冲存储芯片 202 的特定主要内部组件的示意图。缓冲存储芯片包括用于在多个存储器地址处存储数据的随机存取数据存储阵列 301、用于对收到的命令信息进行解码并控制芯片 202 的操作的控制逻辑、用于产生总线计时信号的锁相环 (PLL) 时钟电路 303, 以及用于存储芯片 202 的操作模式的模式控制寄存器 304。芯片 202 还包括各种用于通过与外部源耦合的相应 I/O 线路 311 和 312 接收数据的接收缓冲器 306 和 308, 以及用于通过与外部源耦合的相应 I/O 线路 310 和 313 驱动 (传输) 数据的驱动器 305 和 309。

[0046] 根据本发明的优选实施例, 芯片 202 可使用多个不同模式中的任一模式执行操作, 并且响应于以不同方式实现的相同输入而具有不同的行为, 具体取决于操作模式。模式控制寄存器 304 存储一个或多个识别芯片 202 的操作模式的位。通常, 一旦在诸如放置多个存储芯片的印刷电路卡之类的计算机系统组件中配置了芯片 202, 便不希望更改操作模式。模式控制寄存器 304 表示任何可以被配置为存储模式值 (无论该值在制造之后是否可以重新写入) 的实体。实际上, 对于许多技术而言, 操作模式的值可以在将芯片组装到电路卡中时被永久地写入芯片。模式控制寄存器的目的是无需多种单独的芯片设计, 便可在多种不同的模式中使用 (从而在多种不同的存储配置下使用) 公共芯片设计。

[0047] 根据优选实施例, 某些接收器或驱动器线路的功能可以根据存储芯片 202 的操作模式而变化, 具体由模式控制寄存器 304 确定。在基本操作模式中, 为了在图 2 的典型菊链配置中使用, 进站数据驱动器 305 用于在进站数据总线部分 206 上将数据向内驱动到存储控制器 201; 并且出站数据驱动器 309 用于在出站数据总线部分 205 上将命令 / 地址数据 (以及可选的写入数据) 向外驱动到菊链 203 中的下一芯片 202。类似地, 在基本操作模式中, 进站数据接收缓冲器 308 在进站数据总线部分 206 上从菊链中的下一芯片接收入站数据; 并且出站数据接收缓冲器 306 在出站数据总线部分 205 上从上一芯片或存储控制器接收入站命令 / 地址数据 (以及可选的写入数据)。因此, 在基本模式中, I/O 线路 310 和 312 对应于相应菊链链路 204 的出站数据总线部分 205; 并且 I/O 线路 311 和 313 对应于相应菊链链路 204 的进站数据总线部分 206。在特定备选操作模式中, 由特定接收器接收的数据和 / 或由特定驱动器驱动的数据可以不同于基本操作模式中的数据, 并且 I/O 线路 311-313

具有不同的功能,如本文将详细说明的。

[0048] PLL 电路 303 从外部源接收总线时钟信号并将所述信号再次驱动到另一外部源,例如,它从菊链中的上一芯片(或存储控制器)接收总线时钟并将其再次驱动到链中的下一芯片。该信号用于提供数据和命令/地址总线部分上的传输计时。在优选实施例中,芯片 202 支持不同的总线频率,具体取决于操作模式,在本文的优选实施例的某些变型中,芯片 202 同时支持多个总线频率,即,总线的不同部分以不同的频率运行。PLL 根据需要增大入站频率以将具有最高所需频率的总线时钟信号提供给控制逻辑 302,控制逻辑 302 通过对较高频率的周期计数来生成任何所需的低频总线时钟信号。一般而言,控制逻辑 302 解释收到的命令/地址数据,判定数据是针对芯片 202 还是另一芯片,或同时针对这两个芯片,访问数据存储阵列 301 以响应针对芯片 202 的命令,将读取的数据返回到存储控制器,以及根据需要将数据再次驱动到其他一个或多个芯片。控制逻辑 302 在多种模式中操作,具体取决于模式寄存器 304 的设置。对入站数据的解释将取决于操作模式。如上所述,在基本模式中,命令和地址数据在出站数据接收缓冲器 306 的五个专用 I/O 端口处接收;但是,在特定备选操作模式中,至少一些命令/地址数据在其他端口或缓冲器处接收。

[0049] 存储芯片 202 被设计为在图 2 的用于典型低端系统的配置中使用,所述低端系统即单用户计算机、可携式数字设备、游戏控制器、所谓的“机顶盒”视频控制器等。如果将相同的芯片改造为在大型计算机系统中使用,则图 2 的配置具有一些缺陷。大型计算机系统所需的主存储器容量通常远大于低端系统。可以通过采用多个存储控制器 201 来在一定程度上增加存储器容量,但是每个附加的存储控制器都会给前端总线 103 的设计带来很大的负担。还可以进一步通过增加菊链的长度来几乎无限地增大存储器容量。但是,增加长度会增加访问的延迟。对于大容量存储子系统而言,延迟将变得大到无法使用系统。此外,较长的菊链相对耗电。每个模块都必须缓冲传输并向/从所述链中的更多后续模块重传传输,从而只是在将数据传出芯片时就会消耗大量电力。

[0050] 根据本发明的优选实施例,可以通过以菊链式集线器所支持的芯片群集排列缓冲存储芯片 202 来减少这些缺点。图 4 是根据本发明优选实施例的使用集线器和群集配置的存储子系统的高级方块图。

[0051] 参考图 4,存储控制器 401 与前端总线 103 耦合以便与一个或多个处理器通信。存储控制器 401 又支持一个或(典型地)多个被配置为相应菊链式集线器 402A-D(此处统称为部件 402)的存储芯片总线,每个集线器支持至少一个(以及优选地多个)存储芯片 403A-O(此处统称为部件 403)群集。主存储器 102 可包含多个存储控制器 401,每个存储控制器支持一组相应的连接集线器和芯片群集。

[0052] 在图 2 的菊链式芯片的情况下,优选实施例的集线器 402 通过点到点链路 405A-D(此处统称为部件 405)连接为链 404,后者从存储控制器 401 到第一集线器,然后从每个集线器到所述链中的下一集线器,直到到达所述链的末尾。存储控制器 401 通常支持多个点到点集线器链路 405,每个链路都能支持一组相应菊花链接 404 的集线器和群集。尽管图 4 示出了两组菊链式集线器,但是数量可以更改,并且通常多余两组。

[0053] 每个点到点链路 405 都包括出站数据部分 407 和入站数据部分 408。优选地,点到点链路 405 的运行方式基本上与图 2 的配置中的菊链式缓冲存储芯片 202 间的点到点链路 204 的运行方式相同。具体地说,在优选实施例中,总线链路 405 的总线频率与总线链路

204 的总线频率相同,并且每个数据访问事务(读取或写入)都需要多个(例如六个)周期进行传输。入站数据部分 408 优选地与入站数据部分 206 的总线宽度(线路数)相同(例如都为 13 个线路)。出站数据部分 407 也使用 13 个线路来传输出站写入数据,加上一些数量的线路来传输命令/地址数据。出站数据部分 407 优选地支持与存储芯片菊链 203 中的链路 204 的出站数据部分 205 的地址和命令协议类似的地址和命令协议,但是可能需要一个或多个附加地址线路来适合更大的潜在地址空间,因为优选实施例排列的一个目的是在存储控制器所支持的每个存储芯片总线端口上支持更多数量的芯片(更大的存储器地址范围)。

[0054] 在操作中,存储控制器 401 通过前端总线 103 接收命令,并且确定应用命令的菊链式集线器/群集组 404,即,适当的地址的所在的组。控制器 401 然后在出站数据部分 407 上将每个命令和地址(与写入数据一起,如果命令为写入命令)传输到适当的菊链式集线器/群集组中的第一集线器,所述第一集线器接收数据并根据需要在后续总线周期中将数据重传到下一集线器,直到到达为存储适当的数据地址的群集 403 提供服务的集线器。所述集线器然后将命令/地址(以及写入数据,如果适用)传输到适当的群集 403 的存储器模块,如本文将详细说明的。当命令/地址被重传到群集时,优选地删除指定集线器和群集时所需的额外地址位以减少所需的地址宽度。数据被优选地在群集的多个模块之间交织,如本文将更详细说明的。所述群集的所有模块将所有写入数据写入指定的地址,或读取由读取命令指定的数据,将读取的数据返回集线器。所述集线器然后沿着链向上在入站数据部分 408 上将数据传输到存储控制器,所述链中的每个中间集线器都在相应后续总线周期内重传所述数据,直到该数据到达存储控制器 401,然后所述数据从存储控制器 401 经由前端总线 103 被传输到发出请求的实体。

[0055] 根据本发明的优选实施例,存储芯片的“群集”403 为一组存储数据的多个存储芯片,其被配置以便在针对群集的每个事务中访问群集的所有芯片。换言之,所述群集中的数据将在群集的所有芯片之间交织。可以按多个位、按字节或字在逐位的基础上或在某些其他基础上进行交织。除了交织数据,在优选实施例中,到芯片的通信链路的至少一部分以低于集线器链 404 的链路 405 的总线频率运行。除了较低的频率,使用交织允许芯片以等同于集线器链 404 的数据速率的速率来传输或接收数据。因此,集线器和群集的使用增加了与存储控制器(与图 2 所示的芯片菊链相对)相连的每个总线链 404 上所支持的芯片数,同时通过降低总线频率减少了每个芯片消耗的平均功率。存储芯片群集 403 可以按照各种配置中的任意配置来实现。下文将描述若干此类代表性配置,应该理解此处说明的示例性配置并非穷举。

[0056] 图 5 是比图 4 中的表示更详细地示出根据优选实施例的特定变型的集线器与关联存储芯片群集间的链路的方块图。如图 5 所示,群集 403 包括 39 个芯片 202,所述芯片排列在三个子群集 501A-C(此处统称为部件 501)中。对于每个子群集 501,集线器 402 与单独的子群集通信链路 502A-C(此处统称为部件 502)耦合。每个子群集通信链路都包括用于传输命令/地址数据(以及可选的写入数据,如果是到子群集的芯片的写入命令)的相应命令/地址/写入数据部分 504,以及用于接收从群集的芯片读取,然后被重传到存储控制器的数据的相应读取数据部分 505。优选地,命令/地址/写入数据部分包括用于命令/地址数据和写入数据的单独线路组,所需的线路数取决于配置,下面将介绍一些配置实例;但

是,命令 / 地址数据可以备选地与写入数据一起进行时间多路复用以便共享一组线路。为了清晰,图 5 中仅示出一个与集线器相连的群集,尽管将理解,集线器可以并且通常将为多个群集提供服务。

[0057] 集线器 402 实质上是具有累积 / 按顺序排列以不同频率运行的总线间的数据的能力的复杂交换器。图 9 是示出根据优选实施例的集线器的某些主要内部组件的方块图。集线器 402 包括用于对收到的地址进行解码并控制集线器的操作的解码器 / 控制器 901 ;用于生成总线计时信号的锁相环 (PLL) 时钟电路 902 ;用于对进站数据进行缓冲的进站数据缓冲器 903 ;用于在集线器 402 所属的集线器链的链路 405 上接收和驱动 (传输) 数据的接收缓冲器 905 和 906 以及驱动器 904 和 907 ,以及至少一个用于与存储芯片群集 403 连接的群集接口 910 。

[0058] 集线器 402 可选地包含与解码器 / 控制器 901 耦合的模式寄存器 908 ,后者识别操作模式。集线器 402 可选地被设计为支持多个不同的群集配置,将在下文和图 6-8 中介绍多个此类配置实例,以及 / 或者可以被设计为支持其他操作变型。例如,不同的配置可能需要具备下面任意一项 :不同数量的命令 / 地址线路以及至少一些线路具有不同的总线频率 ;群集中不同数量的子群集 ;不同的数据交织粒度等。

[0059] 群集接口 910 包括写入累加器 911、读取定序器 912、命令 / 地址线路输出驱动器 913、写入数据输出驱动器 914 以及进站读取数据接收缓冲器 915。尽管为了图示清晰,图 9 中只示出了一个接口 910,但是将理解对于集线器 402 所支持的每个群集 403 都存在单独的接口。

[0060] 写入累加器 911 累积通过总线 405 从存储控制器接收并且要以低频率传输到群集的数据。即,通过总线 405 在多个总线周期内接收的数据将在更大的寄存器内累积,直到该数据的宽度足以被在低频、较宽宽度的出站链路 (多个)504 上传输到群集。优选地,至少一些数据以较低的总线速度发出。具体而言,在本文所披露的所有示例性配置中,用于写入芯片的写入数据以降低的总线速度在链路 (多个)504 上传输。在一个示例性配置中,地址 / 命令数据也以较低的总线速度传输。在通过写入累加器来根据需要调整宽度之后,旨在用于群集接口 910 所支持的群集的所接收数据在命令 / 地址线路输出驱动器 913 (用于命令 / 地址数据) 和写入数据输出驱动器 914 (用于写入数据) 上被向外传输到群集。

[0061] 读取定序器 912 将通过链路 (多个)505 接收的较宽宽度进站读取数据按顺序排列到读取缓冲器 915 中,以在多个周期中在链路 405 的较窄宽度进站部分 408 上传输到存储控制器。定序后的数据被放置在进站缓冲器 903 中以在进站驱动器 904 上被重传到存储控制器。

[0062] 在操作中,解码器 / 控制器 901 对收到的命令 / 地址的地址进行解码以判定所述命令是否旨在用于集线器 402 所支持的群集。如果是,则进站数据被路由到对应的群集接口 910。如果否,则进站数据被路由到出站驱动器 907,使其通过集线器 402 到达菊链 404 中的下一集线器。(某些体系结构可支持被路由到群集接口并沿着菊链向下传送的广播命令)。

[0063] PLL 电路 902 从外部源接收总线时钟信号并将所述信号再次驱动到另一外部源,例如,它从菊链中的上一集线器 (或存储控制器) 接收总线时钟信号并将该信号再次驱动到所述链中的下一集线器。可选地,每个群集接口进一步包括用于将总线时钟信号再次驱

动到群集的一个或多个芯片的再驱动器 916, 尽管可由单独装置生成这些芯片的时钟。从 PLL 902 得到的时钟信号还被提供给解码器 / 控制器 901 以控制集线器 402 的内部操作。

[0064] 如上所述, 在示例性实施例中, 存储访问操作存储或读取 78 位数据 (包括辅助位)。群集 403 内的数据在群集的 39 个芯片间交织, 这样寻找群集的芯片的任何存储器访问操作都分布在全部 39 个芯片之间, 即, 每个芯片都存储或读取访问群集的每个存储器访问操作中的恰好两位。由于需要群集中的每个芯片在每个存储器访问操作中仅接收写入数据中的两位或者仅传输读取数据中的两位, 因此承载此数据的线路可以低于构成集线器链 404 的通信链路 405 的线路的频率运行。例如, 在示例性实施例中, 链路 405 以 6GT/s 的总线频率运行, 需要 6 个总线周期来完成存储器访问操作。由于群集中的每个芯片可在 2 个总线周期内在单个线路上接收或传输 2 位数据, 因此它可以在 2 个总线周期内接收或传输其在存储器访问数据中的相应部分, 从而能够在 2GT/s 的总线频率 (在链路 502 上) 上达到充足的数据速率。

[0065] 可以看到每个芯片 202 包含总共 31 个用于在 31 个 I/O 线路上接收数据的接收器以及 31 个用于在 31 个 I/O 线路上传输数据的驱动器。在基本操作模式中, 对于每个接收器 I/O 和驱动器 I/O, 它们作为 13 个出站数据、13 个进站数据以及 5 个命令 / 地址运行。所支持的物理接收器和驱动器 I/O 线路数具有显著限制。由于每个 I/O 线路都是显著成本负担, 因此不太可能使用任何多于最低需求的 I/O 来设计通用芯片。因此, 希望存储子系统的任何备选配置使用的 I/O 线路都不多于图 2 的基本配置 (即, 菊花链接) 中所需的数量。但是, 在控制逻辑 302 中提供根据操作模式以不同方式使用不同 I/O 线路的额外功能所需的成本相对较低, 因此, 可以接受在为某些 I/O 线路指定不同角色的备选配置中配置存储子系统。

[0066] 图 6-8 更详细地示出根据优选实施例的某些变型的子群集的存储芯片与关联集线器之间的数据路径的各种备选配置。在图 6-8 的变型中, 每个群集 403 都包含三个子群集 501, 每个子群集都包含 13 个存储芯片, 如图 5 所示以及上文所述。单个群集 403 的全部三个子群集 501 由同一集线器 402 提供服务, 尽管为了图示清晰从图 6-8 中省略了其他子群集。

[0067] 参考图 6, 其代表子群集配置 501 的第一变型, 出站链路 601 从集线器 402 到存储芯片 202A, 出站链路 601 包括分别用于将写入数据和命令 / 地址数据传输到存储芯片 202A 的 13 个数据线路和 5 个命令线路。13 个数据线路以 2GT/s 的总线频率运行, 即, 以集线器链的通信链路 405 的总线频率的 1/3 运行。5 个命令 / 地址线路以 6GT/s 的总线频率运行。

[0068] 存储芯片 202A 又分别将三个单独的出站链路 602、603、604 驱动到存储芯片 202B、202C 和 202D。来自存储芯片 202A 的每个出站链路 602、603、604 都包括 4 个以 2GT/s 的总线频率运行的数据线路, 以及 5 个以 6GT/s 的总线频率运行的命令 / 地址线路。

[0069] 存储芯片 202B、202C 和 202D 中的每个又将三个单独的出站链路驱动到由三个附加存储芯片构成的相应组。例如, 存储芯片 202B 分别将三个单独的出站链路 605、606、607 驱动到存储芯片 202E、202F 和 202G。每个出站链路 605、605 和 607 都包括 1 个以 2GT/s 的总线频率运行的数据线路, 以及 5 个以 6GT/s 的总线频率运行的命令地址线路。

[0070] 每个存储芯片 202A-202M 都将单个以 2GT/s 运行的进站数据线路驱动到集线器 402。例如, 存储芯片 202E 驱动进站数据线路 608。

[0071] 因此,可以看到子群集被配置为以芯片 202A 作为根的存储芯片树,其中命令 / 地址数据和写入数据沿着树向下传播。但是,存在用于将读取数据从每个芯片传输到集线器的直接连接,也就是说,不需要沿着树向上传播读取数据。

[0072] 尽管在集线器总线链路 405 上以 6GT/s 的速度传输数据,但是按照图 6 所示配置的群集能够跟上所述传输速率,因为数据在三个子群集间交织。即,对于每个传输 78 位可寻址数据以及辅助位的总线操作而言,三个子群集中的每个子群集中存储 26 位,并且每个存储芯片恰好存储两位。每个芯片只需要两个总线拍来传输两位(在单个入站位线路 608 上),而集线器需要六个总线拍来沿着集线器链 404 将其 78 位数据向上传输到存储控制器 401。因此,群集的各个芯片可通过以 2GT/s 的速度将数据传输到集线器来跟上集线器的数据速率。从集线器传输到芯片的出站写入数据也是如此。

[0073] 但是,在出站命令 / 地址数据的情况下,有必要将命令 / 地址数据复制到群集的所有芯片,因此交织不会减少到达每个芯片的命令 / 地址数据量。在图 6 的示例性配置中,通过按照与集线器链链路 405 相同的总线频率(即,6GT/s)运行出站链路的命令 / 数据部分而解决了此问题。即,在该实施例中,某些总线线路运行的频率高于其他线路。与图 2 所示的菊链式配置相比,借助使某些线路以较低的频率运行(而在菊链式配置中,所有线路都以 6GT/s 运行),以及借助使某些芯片使用少于全部线路数的线路,实现了每个芯片功率的降低。

[0074] 在操作中,集线器 402 在连续的总线周期内从链 404 接收连续的命令 / 地址数据(以及可选的写入数据)部分。优选地,识别群集的地址数据包括在第一总线周期(多个)内以缩短延迟。如果收到的信息足以确定目的地群集,则集线器 402 根据与集线器相连的每个群集判定命令是否寻址所述群集。如果否,则集线器忽略该命令,但是该命令被沿着链 404 向下转发。如果是,则针对每个群集在链路 601 的 5 个命令 / 地址线路上以 6GT/s 的频率将该命令重传到位于树根部的全部三个存储芯片 202A(它们位于命令所寻址的群集的相应子群集内)。如果命令为写入命令,则集线器还在其 13 个出站数据输入线路上从存储控制器接收写入数据。该写入数据被以交织的方式重传到三个子群集,每个子群集的根存储芯片 202A 在链路 601 的 13 个线路数据部分上以 2GT/s 的频率接收写入数据。

[0075] 存储芯片 202A 然后分别在链路 602、603 和 604 的三个单独的 5 线路 6GT/s 部分上将三个单独的命令 / 地址数据副本分别重传到芯片 202B、202C 和 202D。但是,如果命令为写入命令,则不需要芯片 202A 重传所有写入数据。链路 601 的 13 位部分针对存储芯片 202A-202M 中的每个芯片携带一位。因此,仅需在链路 602 上将四个数据位重传到存储芯片 202B(以 2GTS 的频率),在链路 603 上将四个位重传到存储芯片 202C(以 2GT/s 的频率),在链路 604 上将四个位重传到存储芯片 202D(以 2GT/s 的频率)。第十三个位用于存储芯片 202A 本身,不进行重传。

[0076] 每个存储芯片 202A、202B 和 202C 然后以 6GT/s 的频率将 5 位命令 / 地址数据的 3 个更多副本重传到由三个芯片构成的相应组。即,芯片 202B 将数据重传到芯片 202E-202G,以此类推,如图 6 所示。在此重传期间,只需以 2GT/s 的频率将单个位的写入数据传输到每个芯片(除了五位命令 / 地址数据以外)。

[0077] 如果命令为读取命令,则每个芯片在相应 1 位线路 608 上将来自芯片的对应读取数据以 2GT/s 的频率直接传输到集线器 402。在每个总线操作中,每个芯片仅存储 78 位读

取数据中的两位,因此通过以 2GT/s 的频率运行,每个芯片可以在完成总线操作的时间内传输所需的数据。由于命令 / 地址数据被传播到树配置中的各芯片,因此各芯片不会全部在同一时刻收到数据。优选地,集线器 402 被设计为缓冲来自各芯片的进站数据以考虑该延迟。备选地,在将读取数据传输到集线器 402 之前,在接收命令的那些芯片中引入一个或两个延迟,从而使集线器 402 同时收到所有数据。

[0078] 构想了存储芯片 202 根据低端系统的需要进行设计,并被设计为使用图 2 的菊链式配置作为基本操作模式(尽管这不一定是本发明的要求)来进行配置。为了能够使用旨在用于低端、菊链式配置的同一通用芯片 202,特别希望本文所披露的任何备选配置尽量少添加芯片电路,具体而言,特别希望任何备选配置中的芯片所需的 I/O 引脚都不多于标准菊链式配置中所用的引脚。每个附加的 I/O 引脚都会显著增加芯片的设计和制造费用,芯片设计者不可能通过添加附加引脚来支持只在小范围的安装环境中使用的备选配置。如上所述,示例性实施例的芯片 202 具有 62 个 I/O 引脚,其中 18 个引脚用于接收来自控制器的出站数据(13 个数据和 5 个命令 / 地址),18 个引脚用于将出站数据再次驱动到菊链中的下一芯片(13 个数据和 5 个命令 / 地址),13 个引脚用于接收来自菊链中的下一芯片的进站数据,以及 13 个引脚用于将进站数据再次驱动到控制器。在 62 个 I/O 引脚中,31 个引脚为用于在外部线路上传输的驱动器以及 31 个引脚为用于从外部线路进行接收的接收器。应该理解这些仅作为示例性设计参数提供,根据本发明使用的存储芯片可以具有不同数量的 I/O 引脚。

[0079] 在本文所披露的所有配置中,在配置的每个芯片中使用的传输引脚和接收引脚的数量被分别限于 31 和 31,尽管引脚的功能(即,在引脚上驱动或接收的数据类型)并不一定与菊链式配置中的功能相同。需要记录操作模式的模式寄存器 304 以及少量响应于模式寄存器中保存的操作模式的附加内部芯片逻辑 302 来相应地对输入线路进行解码或传输到输出线路。此内部芯片逻辑远比附加的 I/O 引脚便宜,因此可有效支持通用芯片设计,即使在相对较少的安装芯片中使用也是如此。

[0080] 适当地为各引脚分配功能可以最大程度上减少支持备选配置所需的内部逻辑。下表示出了在基本操作模式(旨在在图 2 的菊链式配置中使用),以及图 6 的备选配置模式中对 31 个用于接收数据的接收器引脚(从 R1 到 R31)和 31 个用于传输数据的驱动器引脚(从 T1 到 T31)的示例性引脚功能分配。在备选模式中,读取数据 1 和写入数据 1 指在芯片中存储的位,而写入数据 2:13 是以交织的方式在同一群集的其他芯片上存储的数据(该数据因此必须被传播到其他芯片)。如将观察到的,大多数引脚在备选模式中具有相同的功能或者并未在备选模式中使用(因此不需要特殊逻辑)。在备选模式中,内部逻辑仅为芯片使用数据位 1(读取或写入),数据 2:13 被视为用于其他芯片的通过数据。只有引脚 T20:T29 具有不同的功能,被用于在备选模式中驱动命令数据的附加副本。此外,响应于操作模式的控制逻辑 302 将在备选模式中以不同的时钟速率操作数据引脚,即,引脚 R6:R18 和 T7:T19 以 1/3 的总线时钟速率(例如,2GT/s,相对于 6GT/s)运行。由于降低了时钟速率,并且在备选模式中不使用某些引脚,所以将节省功率。

[0081] 表 1:基本模式和备选模式(图 6)的引脚分配

[0082]

	基本模式	备选模式 (图 6)
R1:R5	命令 1:5	命令 1:5
R6	出站写入数据 1	出站写入数据 1(芯片), 1/3 时钟
R7:R18	出站写入数据 2:13	出站写入数据 2:13, 1/3 时钟
R19:R31	入站读取数据 1:13	未使用
T1:5	命令 1:5	命令 1:5
T6	出站写入数据 1	未使用
T7:T18	出站写入数据 2:13	出站写入数据 2:13, 1/3 时钟
T19	入站读取数据 1	入站读取数据 1(芯片), 1/3 时钟
T20:24	入站读取数据 2:6	命令 1:5(副本 B)
T25:29	入站读取数据 7:11	命令 1:5(副本 C)
T30:31	入站读取数据 12:13	未使用

[0083] 上表的一组备选模式引脚分配可用于图 6 的备选配置的所有芯片, 尽管只有芯片 202A 实际使用所列的所有引脚分配。其他芯片使用的引脚较少。例如, 芯片 202B 只接收 4 位写入数据, 因此它只需要 R6:R9 来接收出站写入数据, 使引脚 R10:R18 保留未用。类似地, 它只将 3 位写入数据传输到其他芯片, 因此它只需要 T7:T9 来传输出站写入数据, 使引脚 T10:T18 保留未用。

[0084] 应该理解, 上表的引脚分配只是一个可能的分配组, 还可以使用其他分配。可以针对此处所述的各种其他配置做出类似的分配, 只要接收引脚和传输引脚的总数不超过可用引脚数即可。

[0085] 图 7 表示子群集配置 501 的第二变型。图 7 的配置类似于图 6 的配置, 并且以类似的方式运行, 只是来自存储芯片 202A-202H 以及 202J-202M 的读取数据被传输到芯片 202I, 而不是直接传输到集线器 402。芯片 202I 又将此读取数据与其自己的读取数据一起传输到集线器 402。

[0086] 参考图 7, 出站链路 701 从集线器 402 到存储芯片 202A, 出站链路 701 包括 13 个数据线路和 5 个命令线路来分别将写入数据和命令 / 地址数据传输到存储芯片 202A。所述 13 个数据线路以 2GT/s 的总线频率运行, 而所述 5 个命令 / 地址线路以 6GT/s 的总线频率运行。

[0087] 存储芯片 202A 又分别将三个单独的出站链路 702、703、704 驱动到存储芯片 202B、202C 和 202D。来自存储芯片 202A 的每个出站链路 702、703、704 包括 4 个以 2GT/s 的总线频率运行的数据线路, 以及 5 个以 6GT/s 的总线频率运行的命令 / 地址线路。每个存储芯

片 202B、202C 和 202D 又将三个单独的出站链路驱动到三个附加存储芯片构成的相应组，每个此类链路都包括 1 个以 2GT/s 运行的数据线路，以及 5 个以 6GT/s 运行的命令 / 地址线路，如图 6 中的配置。

[0088] 每个存储芯片 202A-H 以及 202J-202M 都将以 2GT/s 运行的单个入站数据线路驱动到存储芯片 202I。例如，存储芯片 202E 驱动入站数据线路 708。存储芯片 202I 接收该数据，然后在入站链路 707 上将该数据连同它自己的数据一起转发到集线器 402。入站链路 707 是以 2GT/s 运行的 13- 线路链路，并且包含来自每个芯片 202A-202M 的数据。

[0089] 尽管与图 6 的配置相比，图 7 的配置必需额外的延迟周期才能读取数据，但是它具有一些潜在的优势。所有到集线器 402 的读取线路都来自单个芯片，即，芯片 202I，这可以简化计时或物理布局问题。图 7 的配置的另一优势是存在充足的端口来从存储芯片 202A-202H 和 202J-202M 中的每个芯片提供冗余入站线路。这些冗余入站线路在图 7 中示为虚线 709 和 711。在任何主入站端口或线路发生故障时，冗余线路能承担原有端口的功能。因此，存储芯片 202I 具有 12 个被分配给主入站线路 708 的接收器引脚以及 12 个被分配给冗余入站线路 709 和 711 的接收器引脚。由于芯片 202I 只需 6 个附加的接收器引脚来在链路 706 上从芯片 202C 接收 1 位写入数据和 5 位命令 / 地址数据，因此总引脚要求为 30（使一个引脚保留未用）。当需要冗余时，链路 707 优选地包含一个附加冗余线路，后者可以被指定为从芯片 202I 到集线器 402 的 13 个入站读取线路中的任一线路的备用线路。也可以在图 6 的配置中提供冗余，但是需要集线器中具有 13 个额外的接收端口，这将显著地增加成本。

[0090] 如果需要，还可以在出站链路中提供冗余。即，除了一个例外之外，存在足够的未使用端口来为每个出站链路 701-704、706 提供额外的冗余线路。这个例外是芯片 202A，它具有 3 个出站链路 702、703、704，每个链路具有 9 个线路。如果向链路 702、703 和 704 中的每个链路添加第 10 个冗余线路，则没有剩余的未使用线路，使到芯片 202I 的数据线路 710 没有冗余。解决这一问题的方法是共享链路 703 的冗余线路。即，如果线路 710 发生故障，则在到达芯片 202C 的链路 703 上传输读取数据，然后在冗余线路 711 上传输芯片 202I 的数据。冗余将涉及内部逻辑中额外的复杂性，但是不需要额外的 I/O 端口。

[0091] 下表 2 示出了采用上述冗余的图 7 的配置的示例性引脚分配。假设基本操作模式的引脚分配与上面表 1 中的引脚分配相同。线路 R6:18、R20:31 和 T7:19 以 1/3 的时钟速度运行；某些其他线路也可以 1/3 的时钟速度运行，具体取决于配置或使用。即，在被使用时，冗余线路将以它所替代的线路的时钟速度运行。

[0092] 表 2：第二备选模式（图 7）的引脚分配

[0093]

	芯片 202A	芯片 202I	所有其他芯片
R1:R5	命令 1:5	命令 1:5	命令 1:5
R6	出站写入 1	出站写入 1	出站写入 1
R7:9	出站写入 2:4	冗余入站读取 2:4	出站写入 2:4

R10:18	出站写入 5:13	冗余入站读取 5:13	未使用
R19	冗余链路 701	冗余链路 706	冗余命令 / 数据输入
R20:31	未使用	入站读取 2:13	未使用
T1:5	命令 1:5(副本 A)	未使用	命令 1:5(副本 A)
T6	冗余链路 A	未使用	冗余链路 A
T7:9	出站写入 2:4	未使用	出站写入 2:4
T10:17	出站写入 5:12	未使用	未使用
T18	出站写入 13	冗余入站读取 1:13	冗余入站读取 1
T19	入站读取 1	入站读取 1	入站读取 1
T20:24	命令 1:5(副本 B)	入站读取 2:6	命令 1:5(副本 B)
T25:29	命令 1:5(副本 C)	入站读取 7:11	命令 1:5(副本 C)
T30	冗余链路 B / 入站读取 1	入站读取 12	冗余链路 B
T31	冗余链路 C	入站读取 13	冗余链路 C

[0094] 图 8 示出了子群集配置 501 的第三变型。在图 8 的配置中,所有线路(即,命令 / 地址以及存储的数据)都以相同的时钟频率运行,具体而言,以 $1/3$ 的时钟频率 $2GT/s$ 运行。为了使用低频线路传输所有命令 / 地址数据,图 8 的配置针对每个链路使用更多数量的命令 / 地址线路。可以通过放宽所有总线链路均为点到点链路的默认限制,即,通过使用多点线路来支持所述更多数量的线路。在这种情况下,出站命令 / 地址数据在多点线路上同时被传输到两个存储芯片。优选地,单个存储芯片在物理排列上相互靠近并且与集线器 402 靠近。这种排列连同降低的总线时钟频率的使用使得能够在不修改芯片 I/O 驱动器 / 接收器硬件的情况下可靠地支持多点配置,尽管存储芯片 202 旨在用于点到点通信链路。

[0095] 参考图 8,出站链路 801 从集线器 402 到存储芯片 202A 和 202B。出站链路 801 包括 13 个用于传输写入数据的点到点数据线路,在这 13 个线路中,6 个到存储芯片 202A 以及 7 个到存储芯片 202B。出站链路 801 还包括 15 个用于传输命令 / 地址数据的命令 / 地址线路,它们为多点线路并同时到芯片 202A 和 202B。在备选方案中,如果出站链路 801 太长或者由于其他原因而不支持多点,则可以重复 15 个命令 / 地址线路,一组到存储芯片 202A 以及另一组到存储芯片 202B,从而使两个单独的点到点链路从集线器 402 到芯片 202A 和 202B。链路 801 的所有线路都以 $2GT/s$ 的总线频率运行,即,以 $1/3$ 的总线频率运行。由于命令地址数据在 15 个线路上传输,因此可以较低的频率传输相同的数据量。

[0096] 存储芯片 202A 和 202B 又将各个出站多点链路 802 和 803 驱动到存储芯片 202C

和 202D(对于链路 802),以及驱动到 202E 和 202F(对于链路 803)。来自存储芯片 202A 的出站链路 802 包括 5 个数据线路和 15 个命令 / 地址线路。来自存储芯片 202B 的出站链路 803 包括 6 个数据线路和 15 个命令 / 地址线路。链路 802 和 803 上的所有线路都以 2GT/s 的总线频率运行。

[0097] 每个存储芯片 202C、202D、202E 和 202F 又将各个出站链路驱动到相应存储芯片对(或对于芯片 202D,驱动到单个存储芯片 202I)。例如,存储芯片 202C 将出站链路 804 驱动到芯片 202G、202H,链路 804 包括 15 个多点命令 / 地址线路(它们到达两个芯片)以及 2 个写入数据线路,它们分别到达芯片 202G 和 202H。所有这些线路都以 2GT/s 的总线频率运行。

[0098] 每个存储芯片 202A-202M 都将单个相应入站读取数据线路直接驱动到集线器 402,所述线路也以 2GT/s 的总线频率运行。

[0099] 在操作中,集线器 402 累积命令 / 地址的前三个总线周期(总共 15 位命令 / 地址数据),其优选地包含足够的信息来判定所述命令是否寻址到目标群集。如果是,则针对子群集在链路 801 的 15 个命令 / 地址线路上以 2GT/s 的总线频率同时将所述命令重传到所述群集的命令所寻址的三个子群集中的每个子群集。由于这 15 个命令 / 地址线路为多点线路,将同时在芯片 202A 和芯片 202B 处接收到命令 / 地址数据。如果命令为写入命令,则集线器还在其 13 个出站数据输入线路上从存储控制器接收写入数据。该写入数据被以交织的方式重传到三个子群集,每个子群集在链路 801 的 13- 线路数据部分上以 2GT/s 的频率接收写入数据。

[0100] 存储芯片 202A 和 202B 然后分别在链路 802 和 803 的 15- 线路命令 / 地址部分上将命令 / 地址数据重传到芯片 202C、202D、202E 和 202F。如果命令为写入命令,则芯片 202A 为存储芯片 202C、202D、202G、202H 以及 202I 在链路 802 上重传 5 位写入数据,同时芯片 202B 以类似的方式为存储芯片 202E、202F 和 202J-M 在链路 803 上重传 6 位写入数据。每个存储芯片 202C、202D、202E 和 202F 然后将命令 / 地址数据以及适当的写入数据重传到对应芯片对(或单个芯片)。即,芯片 202C 将数据重传到芯片 202G 和 202H,以此类推,如图 8 所示。

[0101] 如果命令为读取命令,则每个芯片在其相应 1- 位线路上以 2GT/s 的频率将来自芯片的对应读取数据直接传输到集线器 402。该数据以交织的方式存储,如图 6 和图 7 中的配置所示。所述读取数据根据需要由集线器 402 进行缓冲,然后在链 404 上被重传到存储控制器。

[0102] 表 3 示出了图 8 的配置的代表性引脚分配组,与基本模式(菊链式)中的分配进行了比较。所有芯片 202A-202M 都可以使用相同的引脚分配模式配置,尽管某些芯片不使用图 8 中列出的所有引脚。例如,芯片 202G 不为另一芯片重传命令 / 地址数据或接收写入数据,因此不使用引脚 R7:18、T1:5、T7:18 和 T20:29。可以看到图 8 的配置中的引脚分配几乎与菊链模式中的基本引脚分配相同,并且在未使用的入站读取线路 2:11 上接收和转发命令 6:15。当然,内部芯片逻辑有必要将线路 R20:29 上的切换信号识别为命令 / 地址数据(而非入站读取数据),以便对命令和地址进行解码。

[0103] 表 3:基本模式和第三备选模式(图 8)的引脚分配

[0104]

	基本操作模式	备选模式 (图 8)
R1:R5	命令 1:5	命令 1:5
R6	出站写入数据 1	出站写入数据 1 (芯片)
R7:R18	出站写入数据 2:13	出站写入数据 2:13
R19:R31	入站读取数据 1	未使用
R20:29	入站读取数据 2:11	命令 6:15
R30:31	入站读取数据 12:13	未使用
T1:5	命令 1:5	命令 1:5
T6	出站写入数据 1	未使用
T7:T18	出站写入数据 2:13	出站写入数据 2:13
T19	入站读取数据 1	入站读取数据 1 (芯片)
T20:29	入站读取数据 2:11	命令 6:15
T30:31	入站读取数据 12:13	未使用

[0105] 如上所示,在优选实施例的每个备选配置中,每个存储器访问内的最多 78 位可寻址数据和辅助位在三个子群集之间交织,每个子群集包含 13 个芯片,以使每个芯片包含每个存储器访问中的 2 位。在标准菊链式配置中,存储器访问的所有数据(8 个字节加上所有辅助位)都位于一个芯片上。对于大小相同的芯片,与指定 8 个字节(如在基本操作模式中)相比,另外需要 5 位地址才能指定 2 位(如在优选实施例中)。似乎这将需要附加的地址线路。但是,基本操作模式还使用一些地址位来指定寻址菊链中的哪个芯片。根据优选实施例访问群集中的芯片将不需要这些芯片选择地址位,因为集线器对从存储控制器接收的完整地址进行解码,并且只有在命令旨在用于特定群集时才将命令转发到该群集。本文假设在群集配置中指定芯片时不需要的芯片选择地址位足以提供指定芯片内的 2 位交织数据(与芯片内的 8 个字节相对)所需的附加地址数据。

[0106] 尽管优选实施例的配置最多支持 14 个辅助位,但是不必使用全部辅助位,或实际上不使用任何辅助位位置。如果希望节省附加芯片的成本,备选地可不使用对应芯片填充某些或全部辅助位位置。

[0107] 尽管可针对本文中所披露的最多 14 个辅助位做出各种使用来作为优选实施例,但是一个特定应用是支持冗余存储芯片,也称为芯片销毁(kill)。通过在每个群集中指定 39 个芯片中的一个或多个作为冗余备用芯片来支持冗余。如果在群集中的任何芯片内检测到芯片故障,则被分配给故障芯片的数据可以随后被分配给备用芯片。如果需要,先前被写入故障芯片的数据可以使用 ECC 来重新构建并被重新写入备用芯片。此类芯片存储能力的

再映射可以在存储控制器 401 或集线器 402 中执行。

[0108] 此处所述的分层交织设计所支持的一种可能存储器体系结构变型是：每个命令在存储控制器-集线器总线链路 405 和集线器-芯片总线链路 505 上传输的数据量明显较多。在上述各种实施例中，每个读取命令或写入命令传输 8 字节的读取数据或写入数据，加上辅助位（最多共 78 位），并在存储控制器-集线器总线链路 405 上需要 6 个总线周期。对于每个数据访问而言，此周期数有时称为突发速率，对于传统菊链式配置而言，典型的突发速率为 4 或 8。菊链式配置中的突发速率通常受纠错代码 (ECC)、支持芯片销毁的愿望，以及其他因素限制。但是，此处所述的分层交织设计具有实现更高突发速率的固有冗余。实际上，在单个数据访问中传输的数据量可以与高速缓冲线大小一样，例如 64 或 128 字节。

[0109] 在每个数据访问中传输更多数据可以消除重复命令的需要，从而减少所传输的命令/地址数据量。这种减少使得减少用于命令/地址数据的线路数和/或降低这些线路的频率成为可能。在这种情况下，可以优选地针对命令/地址数据和写入数据使用共享线路，而不是使用上述的专用线路。

[0110] 例如，如果在每个数据访问中传输 64 字节的读取数据或写入数据，则上述总线链路 405 上的 13 个数据线路将需要 48 个周期来传输数据。但是由于只需传输 30 位的命令/地址，因此单个命令/地址线路便足够，因为该线路有 48 个周期来传输 30 位数据。（实际上，位数可以减少，因为指定 64 字节高速缓存线所需的地址线路可以再减少三个，假设该高速缓存线在 64 字节边界上对齐）。但是，在这种情况下，可能不希望针对命令/地址使用单个专用线路，因为接收设备必须等待很多周期才能获知所访问数据的地址。优选地将共享所有线路，以便首先使用所有线路传输命令/地址数据，然后传输写入数据（如果适用）。事实是所需的总线路数仍会减少，因为对于等量被访问的数据而言，所需传输的总线数据总量减少了。

[0111] 该实例可以进一步被应用于集线器-芯片总线链路 502。例如，如果图 6 中的配置在每个访问期间传输 64 字节数据的存储体系结构中使用，则链路 601 优选地包含多个针对命令/地址和数据共享的线路，所有线路以 2GT/s 运行，即，以存储器总线-集线器链路 405 的时钟频率的 1/3 运行。首先传输命令/地址，接着传输数据。由于对于每个数据访问操作，存储器总线-集线器链路 405 上需要 48 个周期，因此，以 1/3 频率运行的链路 601 将完成 16 个周期，并且必须传输近 30 位命令/地址和 208 位数据以及传输辅助位。链路 601 上只需 15 个线路。

[0112] 类似地，每个链路 602、603 和 604 必须传输相同的近 30 位命令/地址和 64 位数据以及辅助位。由于只有 16 个周期可用，因此对于每个链路 602、603 和 604，最少需要 6 个线路。但是，可能希望使用更多线路（例如 9 或 10 个线路），因为使用 6 个线路时，需要 5 个周期才能传输所有命令/地址，这会增加延迟。线路数应限于 10 以处于每个芯片上 31 个可用输出端口的总数内。类似的分析也适用于链路 605、606 和 607。

[0113] 当然，在此类配置中，因为需要支持不同的线路使用、缓冲命令/地址信息等，所以芯片的内部逻辑可能更复杂。与本文先前所述的各种配置相比，如上所述在每个存储器访问命令中提供更多数据可能会增加延迟，如果需要更多周期（或更缓慢的周期）来传输命令/地址数据的话，但是可减少所需的线路数，从而使存储控制器和/或集线器能够更容易支持大型存储器配置，并且还可以通过降低某些线路的总线频率来减少功耗。

[0114] 在上面根据图 6-8 所述的各种备选配置中,出站命令 / 地址和写入数据沿着存储芯片树的多个层级向下传播。例如,在图 6 的配置中,出站命令 / 地址和写入数据首先从集线器传输到芯片 202A(在第一层级),然后传输到芯片 202B、202C 和 202D(在第二层级),然后在第三层级传输到其余的芯片。每个后续的层级都在传播存储器访问命令中引入附加延迟。可以使用不同的层级数来配置群集中的芯片。例如,可以将所有数据提供给单个芯片,然后将数据重新传播给单个群集的后续层级,而不是将群集分为三个子群集,然后将单独的命令 / 地址数据同时驱动到全部三个子群集。这种方法可减少集线器中所需的线路数,但是会增加延迟和功耗。

[0115] 在上面根据图 6-8 所述的各种配置中,假设写入数据使用与命令 / 地址数据相同的方式沿着树依次向下传播。但是,由于写入数据的每个位只有一个目的地,所以可备选地针对写入数据在存储器模块与集线器间提供直接链路,并且只沿着树向下传播命令 / 地址。这种变型可能增加集线器或存储芯片或两者的内部逻辑以及缓冲的复杂性。它不一定会减少集线器中的输出线路数,但是会减少沿着树向下重新传播写入数据时,芯片内所需的 I/O 线路数,因此会降低功耗并可能提供附加的配置灵活性。但是可以看到,当在同一共享线路上传输写入数据和命令 / 地址数据时,这种变型将变得不实用。

[0116] 尽管图 6-8 示出了体现本发明的通用原理的特定配置,但是可以理解可根据本发明使用多种备选的存储芯片配置。通过举例而非限制,除了本文其他位置所披露的任何变型以外,可以在本发明的范围内更改下面任何参数:群集可以包含或不包含子群集,以及子群集的数量可以更改;群集或子群集中的芯片数可以更改;命令 / 地址线路或数据线路的数量可以更改;每个存储器访问的总线频率和 / 或总线周期数可以更改;每个存储器访问的数据位数和 / 或命令 / 地址位数可以更改;将信号传播到群集的芯片树中的层级数可以更改;数据交织的粒度可以更改;存储芯片中的端口数量和功能可以更改等。

[0117] 在此处所述的优选实施例中,在菊链中相连的多个集线器再驱动芯片用于访问多个存储芯片群集。该配置被用于支持每个存储控制器总线端口上的大量存储芯片。但是,可备选地在不使用集线器再驱动芯片的情况下,将存储芯片群集或子群集直接连接到存储控制器。一般地,此类备选配置支持的存储芯片数少于优选实施例的配置所支持的存储芯片数。

[0118] 在此处所述的优选实施例中,使用多模式存储芯片,其中在至少一个操作模式中,芯片可以在传统的菊链式配置中使用,并且在至少另一操作模式中,芯片可以在此处所述的分层存储器配置中使用。但是,可备选地使用专门为此类分层配置设计的单模式存储芯片,或使用不支持所述菊链式配置的存储芯片。

[0119] 尽管披露了本发明的特定实施例以及某些备选实施例,但是本领域中的技术人员将理解,可以在下面的权利要求的范围内做出更多形式与细节上的变型。

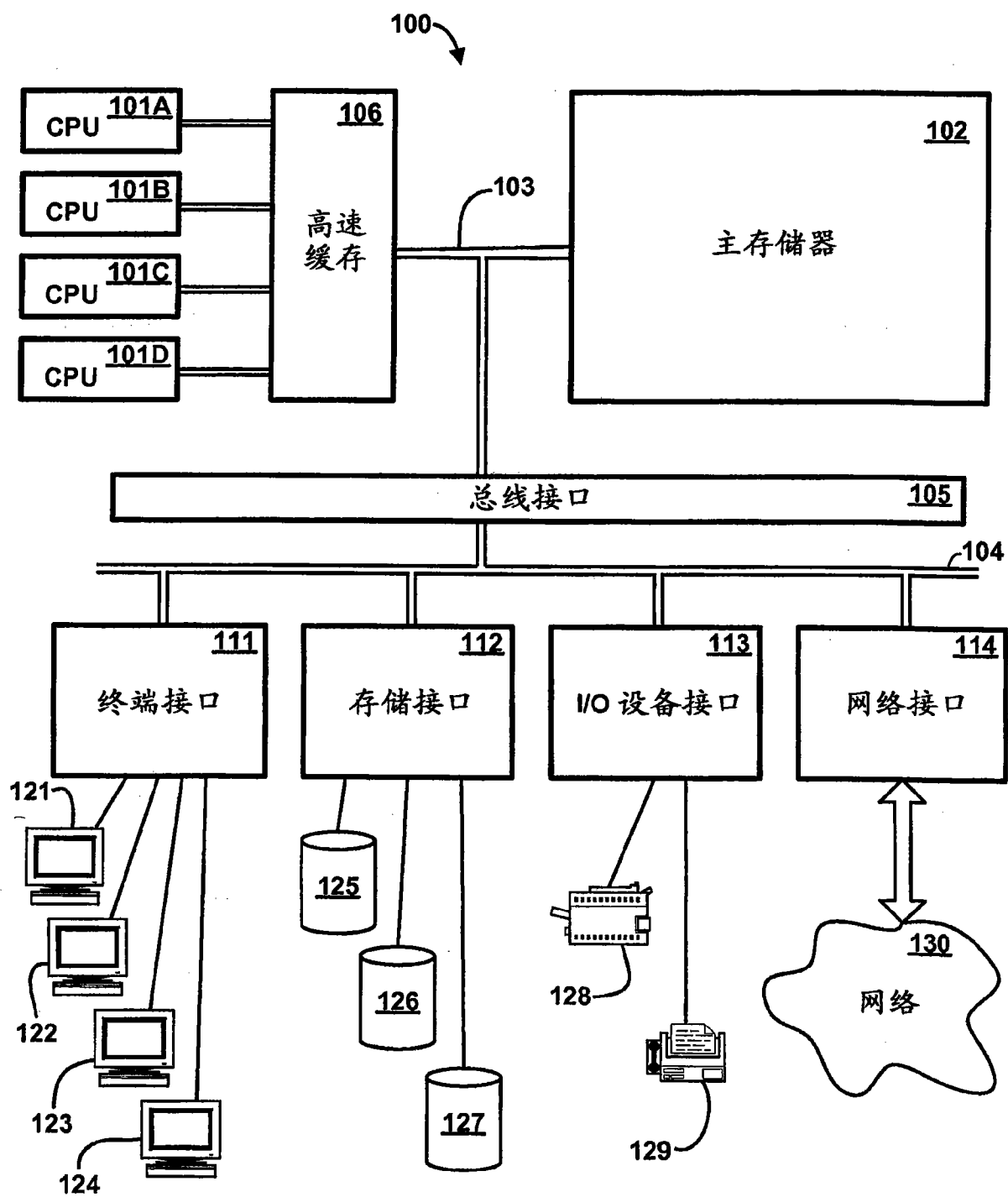


图 1

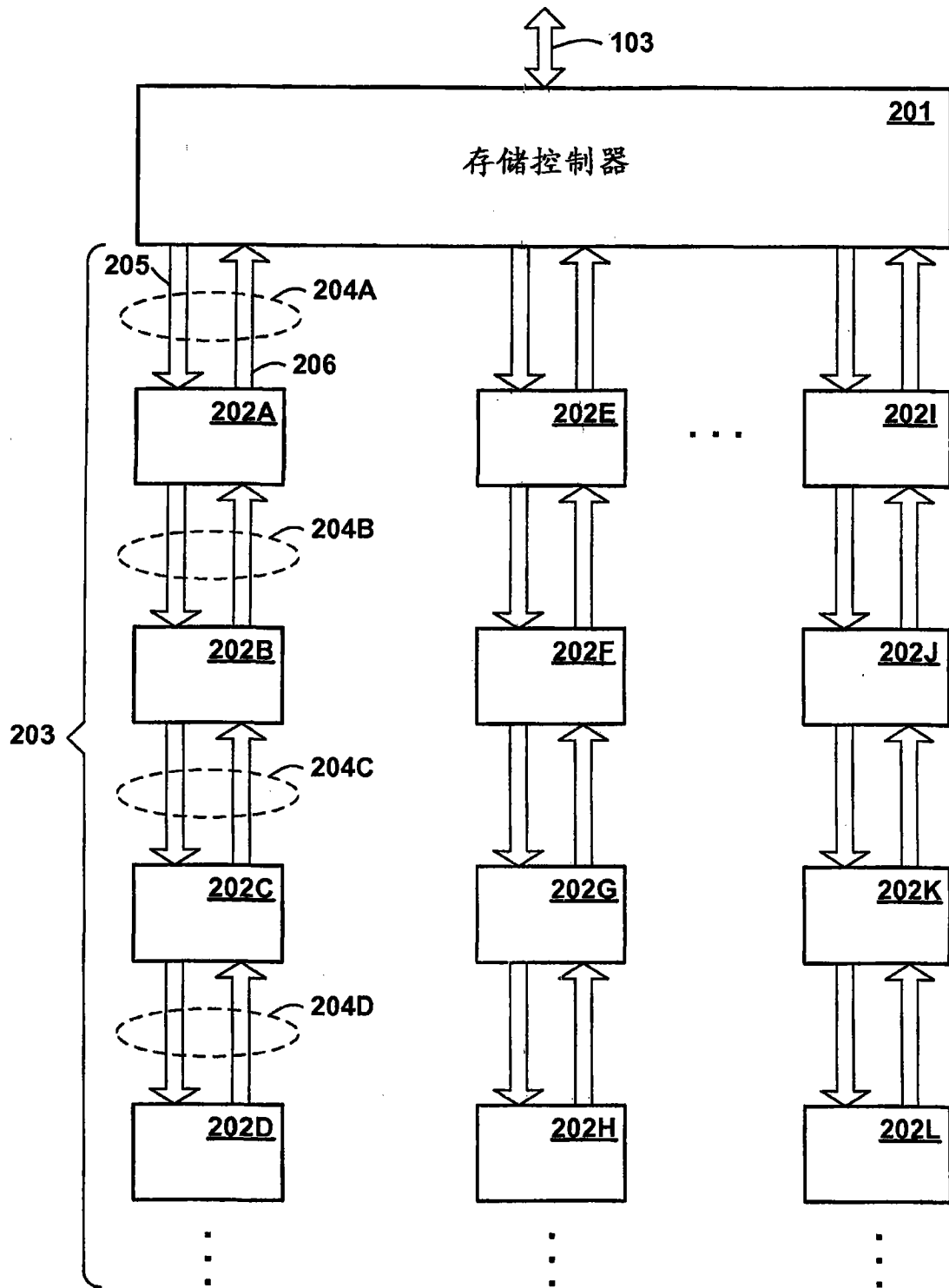


图 2

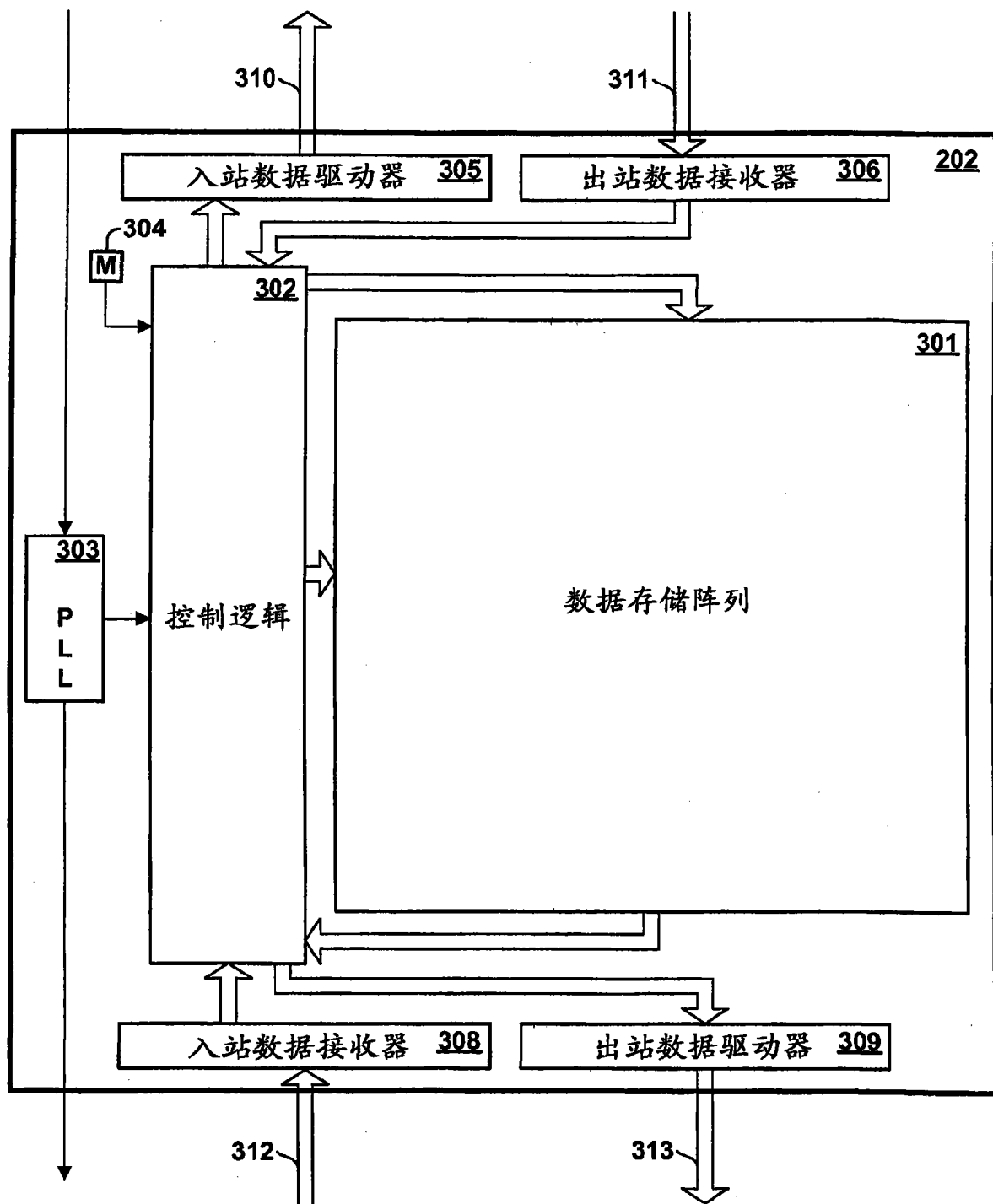


图 3

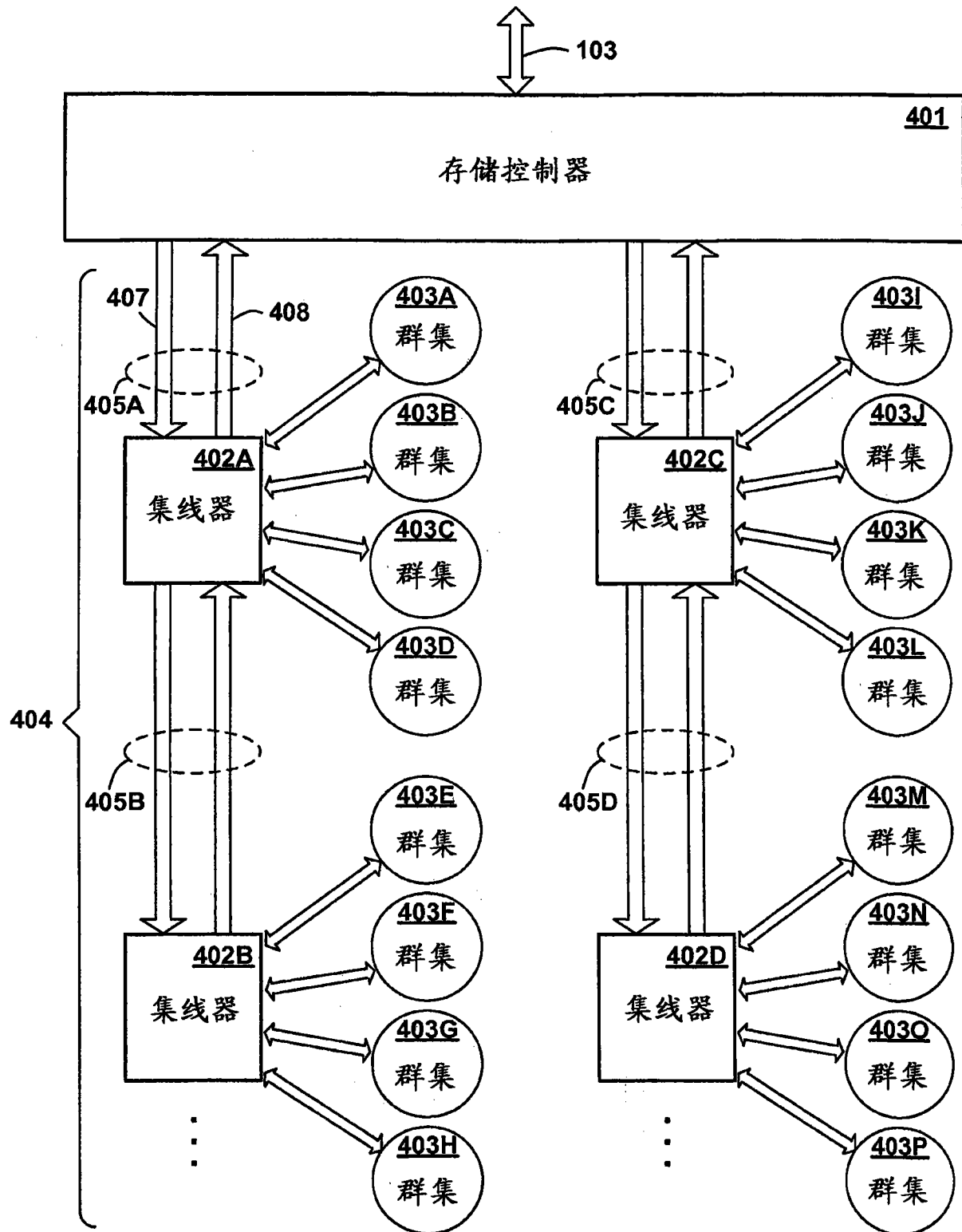


图 4

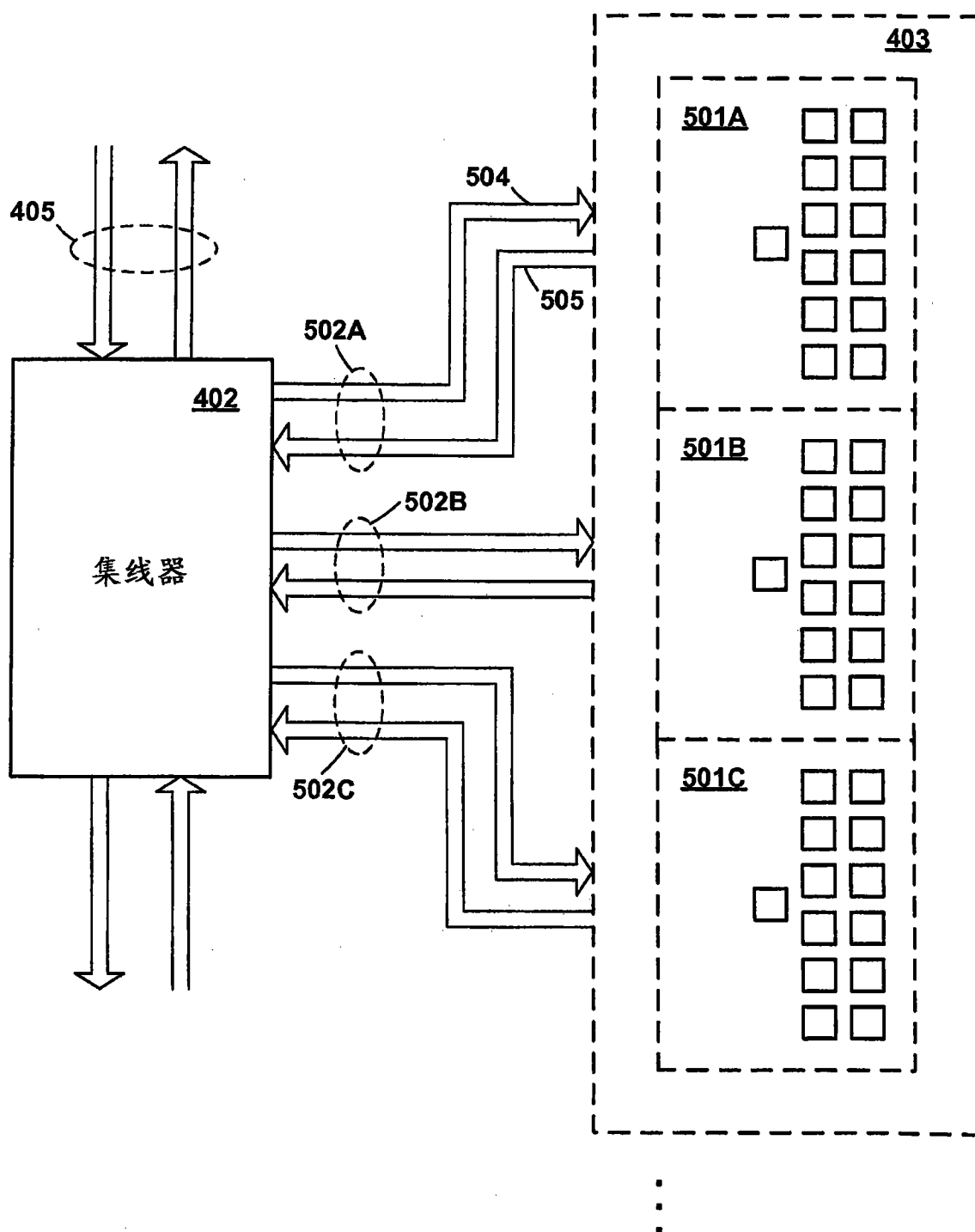


图 5

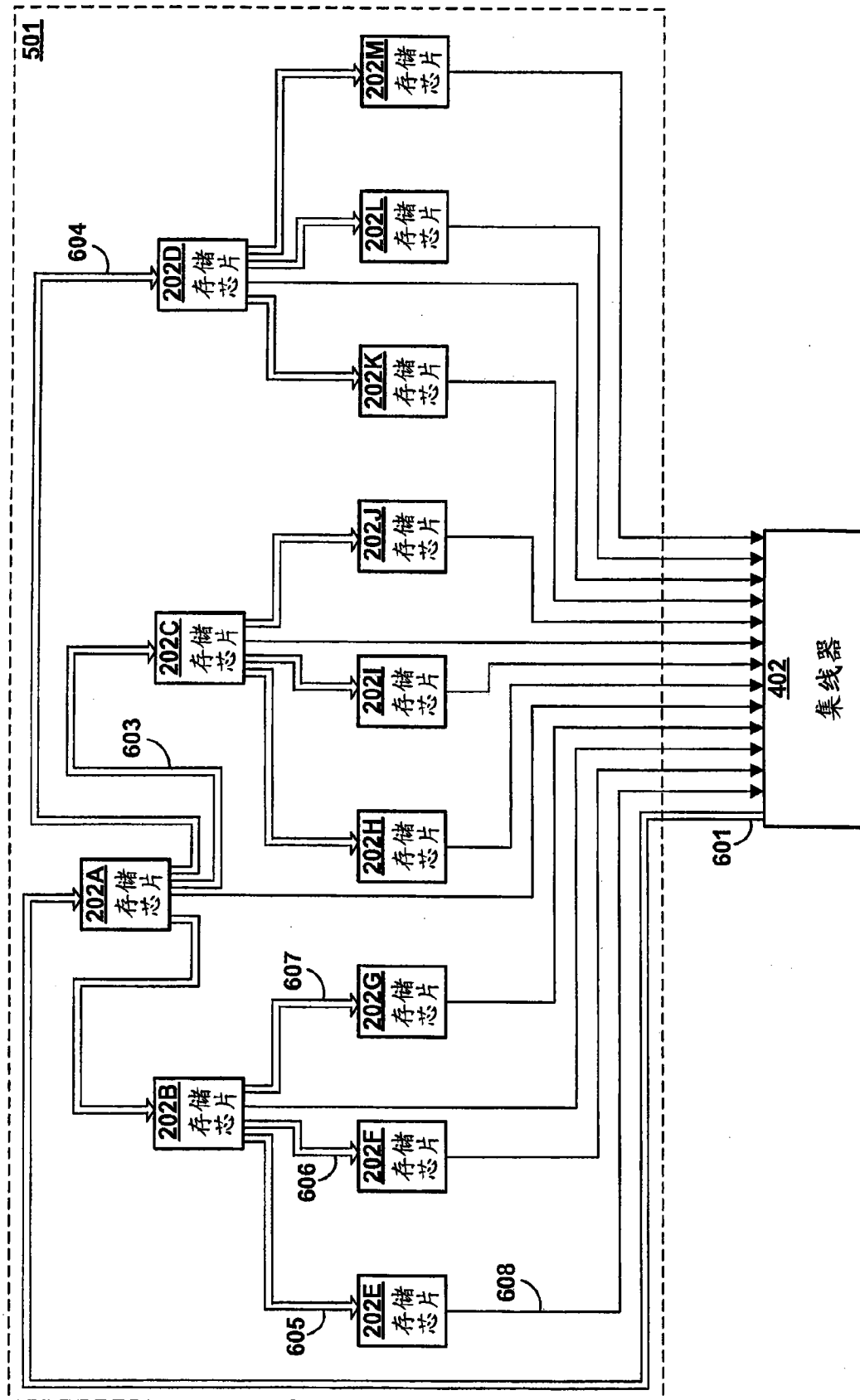


图 6

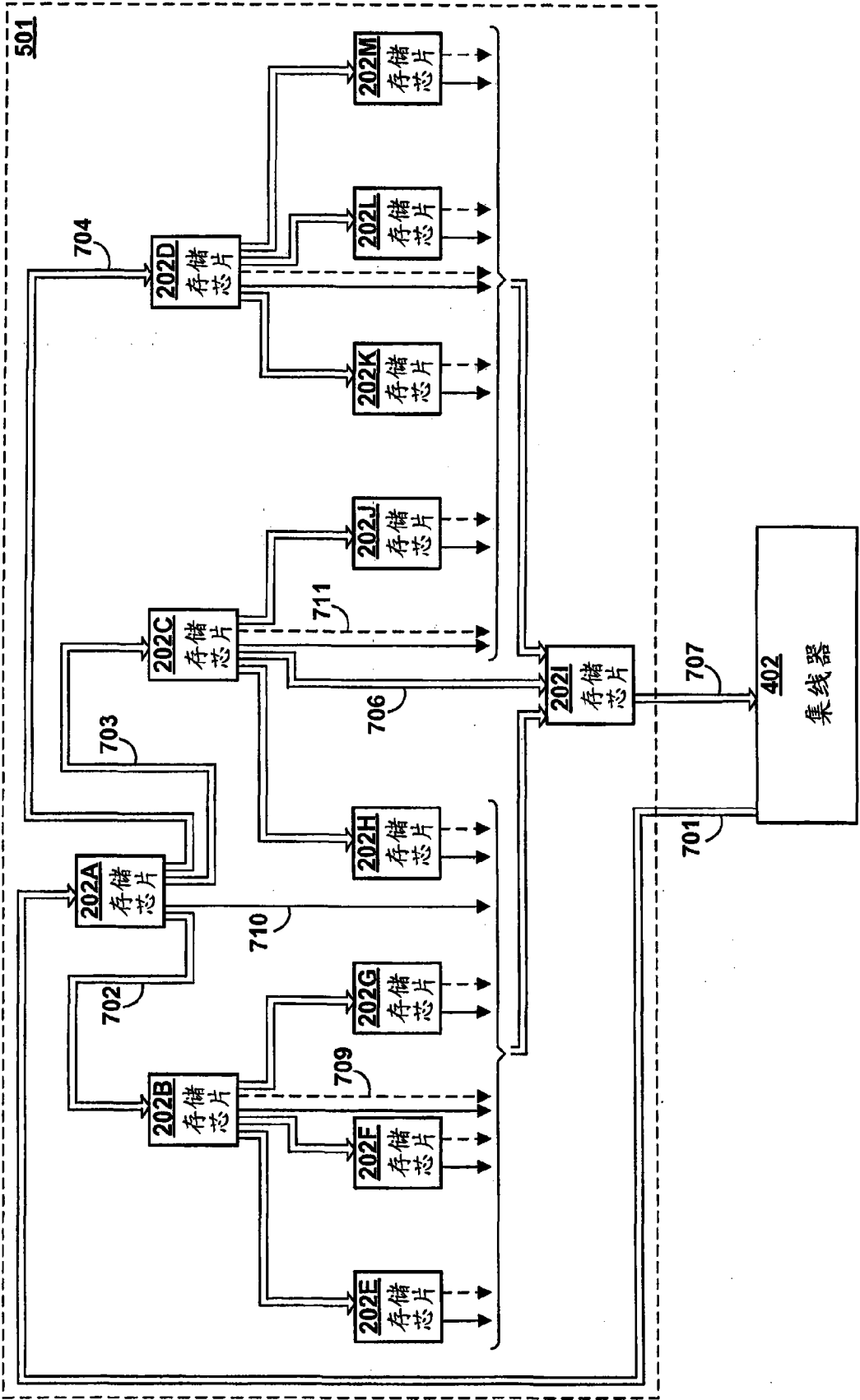


图 7

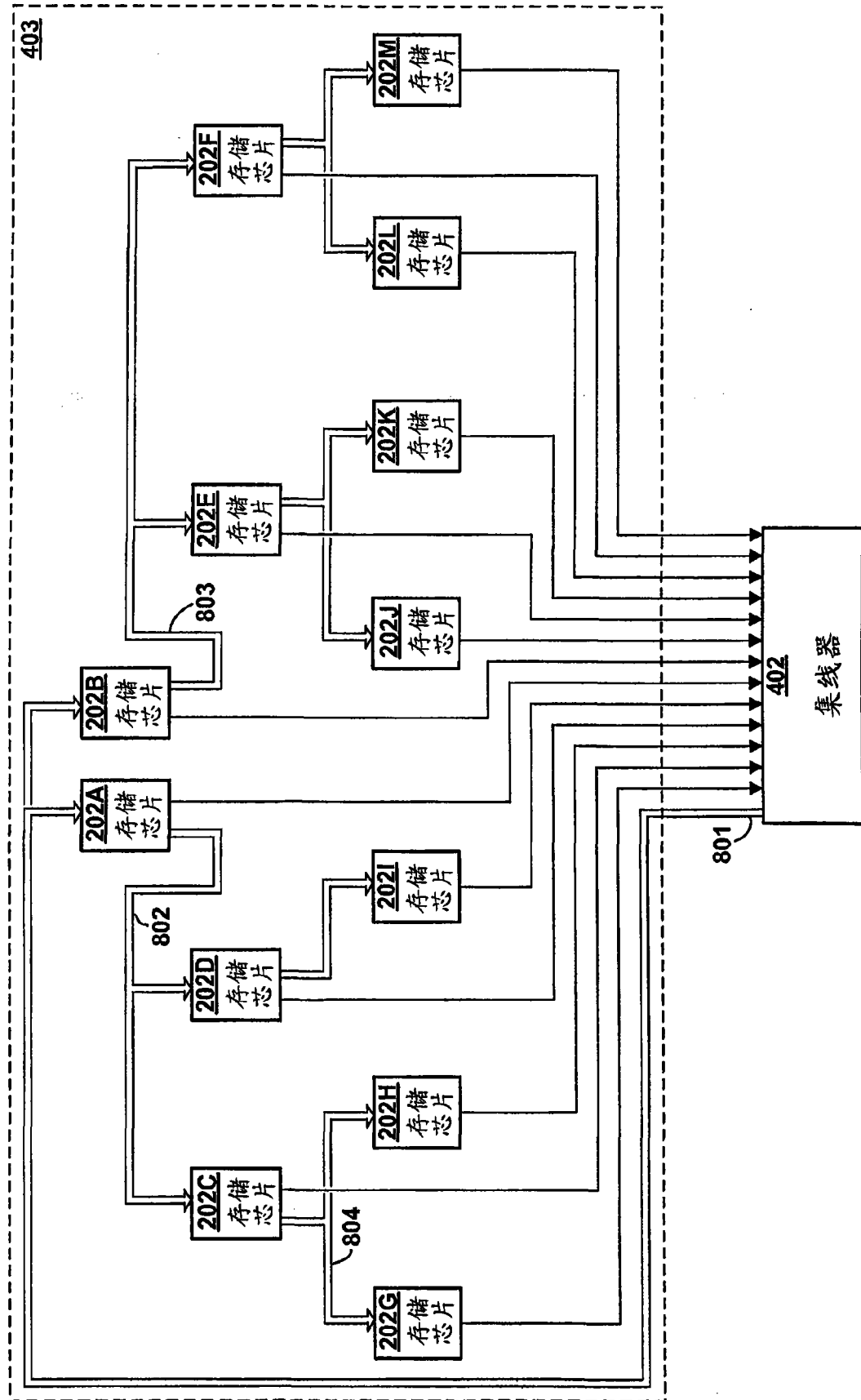


图 8

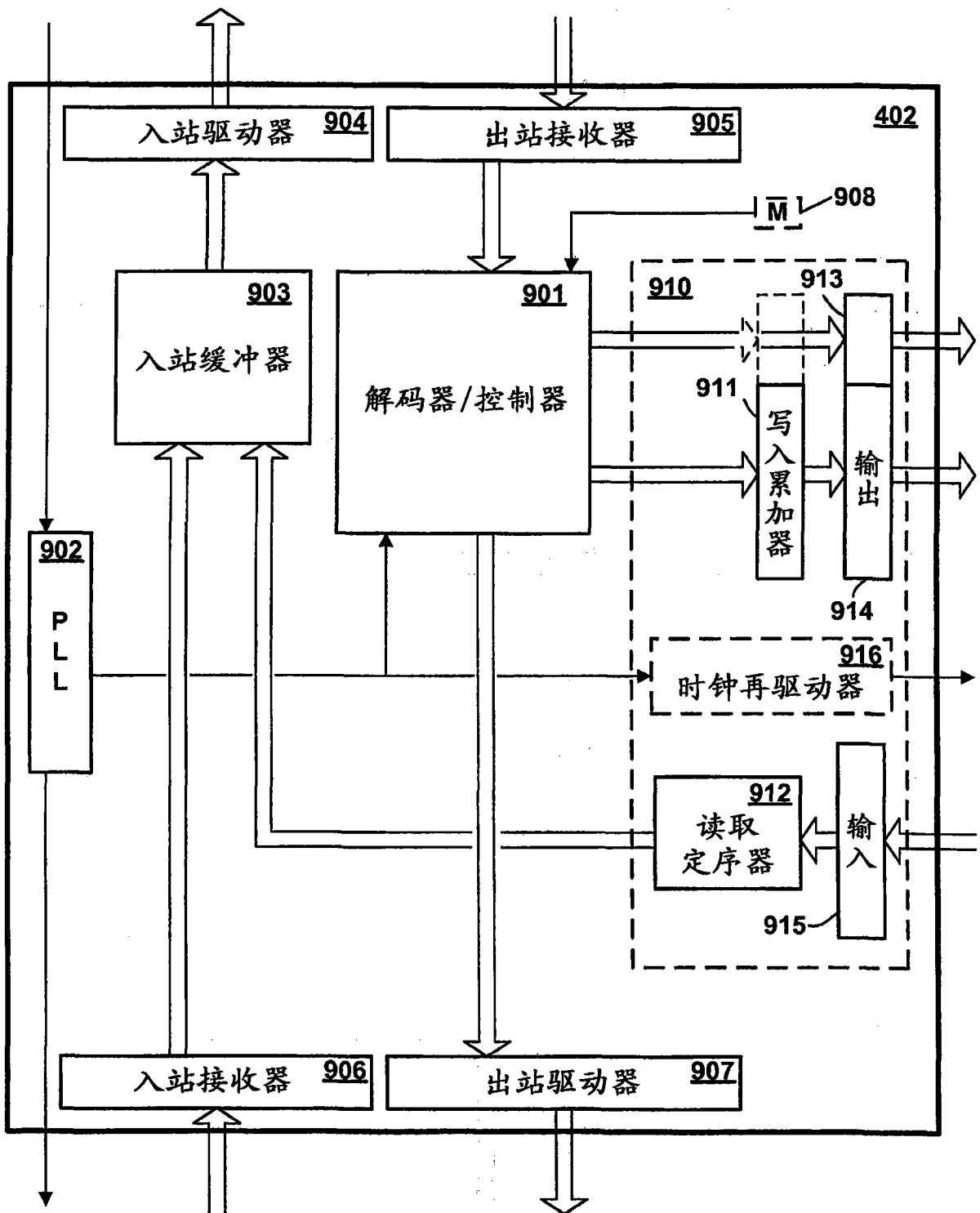


图 9