



(12) 发明专利

(10) 授权公告号 CN 107436855 B

(45) 授权公告日 2022.07.01

(21) 申请号 201710084938.X

(22) 申请日 2017.02.16

(65) 同一申请的已公布的文献号
申请公布号 CN 107436855 A

(43) 申请公布日 2017.12.05

(30) 优先权数据
62/341,602 2016.05.25 US
15/227,959 2016.08.03 US

(73) 专利权人 三星电子株式会社
地址 韩国京畿道

(72) 发明人 朴熙权 奇亮奭

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
专利代理师 李敬文

(51) Int.Cl.
G06F 13/42 (2006.01)

(56) 对比文件

- CN 1233022 A, 1999.10.27
- CN 103281794 A, 2013.09.04
- US 2014173221 A1, 2014.06.19
- CN 103207842 A, 2013.07.17
- CN 1729675 A, 2006.02.01
- CN 104871508 A, 2015.08.26
- US 6412027 B1, 2002.06.25
- US 5634076 A, 1997.05.27
- US 2009276571 A1, 2009.11.05
- US 2007156968 A1, 2007.07.05
- CN 102007479 A, 2011.04.06
- US 2008063004 A1, 2008.03.13
- CN 104335196 A, 2015.02.04
- CN 1218227 A, 1999.06.02
- CN 104011656 A, 2014.08.27
- US 2010064080 A1, 2010.03.11
- CN 1233022 A, 1999.10.27

审查员 姜繁

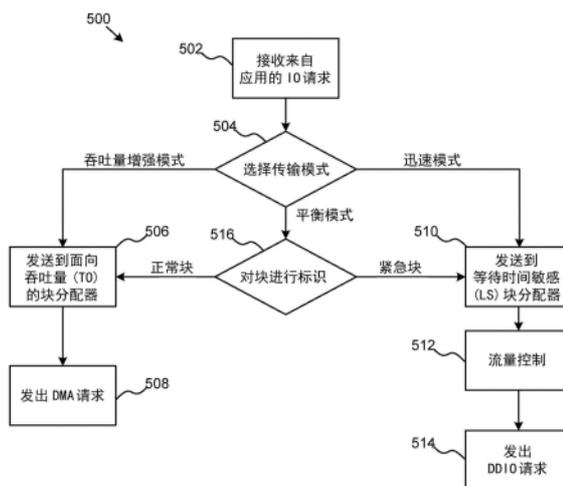
权利要求书3页 说明书8页 附图12页

(54) 发明名称

用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理

(57) 摘要

描述了一种用于具有可重配置多端口的外围组件互连高速 (PCIe) 存储系统的服务质量 (QoS) 认知输入/输出 (IO) 管理的方法和系统。在实施例中,一种方法包括在接收器接口中接收来自软件应用的IO请求,所述IO请求包括用于通过数据访问端口传送数据的信息。该方法还可以包括在传输模式选择器中确定用于通过数据访问端口传送数据的传输模式,传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式。另外,该方法可以包括响应于IO请求并根据数据传输模式,由块分配单元通过数据访问端口传送数据。



1. 一种用于具有可重配置多端口的的外围组件互连高速PCIe存储系统的服务质量QoS认知输入/输出I/O管理的方法,包括:

在接收器接口中接收来自软件应用的输入/输出I/O请求,所述I/O请求包括用于通过数据访问端口传送数据的信息,所述信息包括针对要传送的数据所定义的数据特性,所述数据特性指示数据是紧急数据还是非紧急数据;

在传输模式选择器中根据所述数据特性选择用于通过数据访问端口传送数据的传输模式,所述传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式;以及

响应于I/O请求并且根据所选的传输模式,由块分配单元通过数据访问端口传送数据。

2. 根据权利要求1所述的方法,其中所述面向吞吐量的模式配置为根据直接存储器访问DMA协议,通过数据访问端口与面向吞吐量TO的块分配器传送数据。

3. 根据权利要求1所述的方法,其中所述等待时间敏感模式配置为根据数据直接I/O“DDIO”协议,通过数据访问端口与等待时间敏感LS的块分配器传送数据。

4. 根据权利要求1所述的方法,其中所述平衡模式配置为可选择地根据第一协议或第二协议,通过数据访问端口传送数据。

5. 根据权利要求4所述的方法,其中第一协议是直接存储器访问DMA协议,第二协议是数据直接I/O“DDIO”协议。

6. 根据权利要求5所述的方法,还包括:响应于所述数据特性,利用块标识器确定是根据DMA协议还是根据DDIO协议通过数据访问端口传送数据。

7. 根据权利要求1所述的方法,其中所述数据特性包括与数据相关联的紧急标识符。

8. 根据权利要求1所述的方法,还包括:响应于被配置为根据数据直接I/O“DDIO”协议接收数据的高速缓存的状态,确定是根据直接存储器访问DMA协议还是根据DDIO协议通过数据访问端口传送数据,其中,响应于确定高速缓存处于溢出保护状态,根据DMA协议传送数据。

9. 根据权利要求1所述的方法,还包括:

配置用于根据直接存储器访问DMA协议传送数据的第一组数据访问端口;以及

配置用于根据数据直接I/O“DDIO”协议传送数据的第二组数据访问端口。

10. 根据权利要求9所述的方法,还包括:根据由I/O请求定义的传输模式,通过第一组数据访问端口和第二组数据访问端口中的至少一组对数据的通信进行多路复用。

11. 一种用于服务质量QoS认知输入/输出I/O管理的系统,包括:

数据通信控制器,包括:

接收器接口,配置为接收来自软件应用的输入/输出I/O请求,所述I/O请求包括用于通过数据访问端口传送数据的信息,所述信息包括针对要传送的数据所定义的数据特性,所述数据特性指示数据是紧急数据还是非紧急数据;

传输模式选择器,耦合到接收器接口,并配置为根据所述数据特性选择用于通过数据访问端口传送数据的传输模式,所述传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式;和

块分配单元,配置为响应于I/O请求并根据所选的传输模式,通过数据访问端口传送数据;以及

数据访问端口,耦合到数据通信控制器,所述数据访问端口配置为通过外围组件互连

总线将数据传送到外围设备。

12. 根据权利要求11所述的系统,其中所述面向吞吐量的模式配置为根据直接存储器访问DMA协议,通过数据访问端口传送数据。

13. 根据权利要求11所述的系统,其中所述等待时间敏感模式配置为根据数据直接I/O“DDIO”协议,通过数据访问端口传送数据。

14. 根据权利要求11所述的系统,其中所述平衡模式配置为可选择地根据第一协议或第二协议,通过数据访问端口传送数据。

15. 根据权利要求14所述的系统,其中第一协议是直接存储器访问DMA协议,第二协议是数据直接I/O“DDIO”协议。

16. 根据权利要求15所述的系统,其中所述数据通信控制器还配置为响应于所述数据特性,利用块标识器确定是根据DMA协议还是根据DDIO协议通过数据访问端口传送数据。

17. 根据权利要求11所述的系统,其中所述数据特性包括与数据相关联的紧急标识符。

18. 根据权利要求11所述的系统,其中所述数据通信控制器还配置为响应于被配置为根据数据直接I/O“DDIO”协议接收数据的高速缓存的状态,确定是根据直接存储器访问DMA协议还是根据DDIO协议通过数据访问端口传送数据,其中,响应于确定高速缓存处于溢出保护状态,根据DMA协议传送数据。

19. 根据权利要求11所述的系统,其中所述数据通信控制器还配置为:

配置用于根据直接存储器访问DMA协议传送数据的第一组数据访问端口;以及

配置用于根据数据直接I/O“DDIO”协议传送数据的第二组数据访问端口。

20. 根据权利要求19所述的系统,其中所述数据通信控制器还配置为根据IO请求所定义的传输模式,通过第一组数据访问端口和第二组数据访问端口中的至少一组对数据的通信进行多路复用。

21. 一种用于具有可重配置多端口的外围组件互连高速PCIe存储系统的服务质量QoS认知输入/输出IO管理的装置,包括:

接收器接口,配置为接收来自软件应用的输入/输出IO请求,所述IO请求包括用于通过数据访问端口传送数据的信息,所述信息包括针对要传送的数据所定义的数据特性,所述数据特性指示数据是紧急数据还是非紧急数据;

传输模式选择器,耦合到接收器接口,并配置为根据所述数据特性选择用于通过数据访问端口传送数据的传输模式,所述传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式;以及

块分配单元,配置为响应于IO请求并根据所选的传输模式,通过数据访问端口传送数据。

22. 根据权利要求21所述的装置,其中:

所述面向吞吐量的模式配置为根据直接存储器访问DMA协议通过数据访问端口传送数据;

所述等待时间敏感模式配置为根据数据直接I/O“DDIO”协议通过数据访问端口传送数据;以及

所述平衡模式配置为可选择地根据第一协议或第二协议通过数据访问端口传送数据。

23. 根据权利要求22所述的装置,其中第一协议是DMA协议,第二协议是DDIO协议。

24. 一种非暂时性机器可读介质,包括指令,当由一个或多个处理器执行时,所述指令使得所述一个或多个处理器执行以下操作:

接收来自软件应用的输入/输出I/O请求,所述I/O请求包括用于通过数据访问端口传送数据的信息,所述信息包括针对要传送的数据所定义的数据特性,所述数据特性指示数据是紧急数据还是非紧急数据;

根据所述数据特性选择用于通过数据访问端口传送数据的传输模式,所述传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式;以及

响应于I/O请求并且根据所选的传输模式,通过数据访问端口传送数据。

用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理

[0001] 相关申请数据

[0002] 本申请根据35U.S.C.§119(e)要求于2016年5月25日提交的美国专利申请序列号62/341,602以及2016年8月3日提交的美国专利申请序列号15/227,959的优先权,其通过引用并入本文。

技术领域

[0003] 本公开总体涉及数据处理系统,并且更具体地,涉及用于具有可重配置多端口的外围组件互连高速(PCIe)存储系统的服务质量(QoS)认知((QoS)-aware)输入/输出(IO)管理。

背景技术

[0004] 诸如计算机系统的数据处理系统通常包括具有若干核心组件的主机系统,包括中央处理单元(CPU)、存储器设备、IO设备以及用于在主机系统的组件之间传输信息的设备互连或根联合体(root complex)。互连协议的一个示例是PCIe,其定义了主机系统组件之间的数据传输模式。

[0005] 直接存储器访问(DMA)、直接高速缓存访问(DCA)和数据直接I/O(DDIO)是用于在主机系统(存储器或高速缓存)和IO设备之间传输数据的现有技术。DMA在IO设备和主机系统存储器之间传输数据。DDIO在IO设备和主机CPU的最末级高速缓存(LLC)之间传输数据。DCA就像DMA一样在IO设备和主机系统存储器之间传输数据,但是在数据传输完成后,DCA从存储器将数据预取到CPU高速缓存中。

[0006] 主机系统通常包括多个数据处理端口,每一个数据处理端口可以配置为按照DMA、DCA或DDIO协议之一来传输数据。这样的端口在本文中被称为“多端口”。公共系统可以具有被配置用于DMA的一些多端口、被配置用于DDIO的一些多端口以及可能被配置用于DCA的一些多端口。现有系统软件使用这些技术(DMA、DCA或DDIO)之一向设备发送IO请求。然而,当前没有用来组合DMA、DCA或DDIO协议的功能方面或者针对多路复用、QoS控制或高速缓存流量控制来在协议之间动态切换的方法。

发明内容

[0007] 描述了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法和系统的实施例。在实施例中,一种方法包括在接收器接口中接收来自软件应用的IO请求,所述IO请求包括用于通过数据访问端口传送数据的信息。该方法还可以包括在传输模式选择器中确定用于通过数据访问端口传送数据的传输模式,传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式。另外,该方法可以包括响应于IO请求并根据数据传输模式,由块分配单元通过数据访问端口传送数据。

[0008] 系统的实施例可以包括数据通信控制器。数据通信控制器可以包括接收器接口,配置为接收来自软件应用的输入/输出(IO)请求,所述IO请求包括用于通过数据访问端口

传送数据的信息。另外,数据通信控制器可以包括传输模式选择器,其耦合到接收器接口并且配置为确定用于通过数据访问端口传送数据的传输模式,传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式。数据通信控制器还可以包括块分配单元,配置为响应于IO请求并根据数据传输模式通过数据访问端口传送数据。该系统还可以包括耦合到数据通信控制器的数据访问端口,数据访问端口被配置为通过外围组件互连总线将数据传送到外围设备。

[0009] 装置的实施例可以包括接收器接口,配置为接收来自软件应用的输入/输出(IO)请求,所述IO请求包括用于通过数据访问端口传送数据的信息。装置还可以包括传输模式选择器,其耦合到接收器接口并且配置为确定用于通过数据访问端口传送数据的传输模式,传输模式选自面向吞吐量的模式、等待时间敏感模式或平衡模式。另外,装置可以包括块分配单元,配置为响应于IO请求并根据数据传输模式通过数据访问端口传送数据。

附图说明

[0010] 本发明通过示例来说明,并且不受附图限制,其中类似的附图标记指示相似的元件。附图中的元件是为了简单和清楚而示出的,并且不一定按比例绘制。

[0011] 图1是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统的实施例的示意性框图。

[0012] 图2是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统的实施例的示意性框图。

[0013] 图3是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的实施例的示意性流程图。

[0014] 图4是示出了专门配置用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统的实施例的逻辑图。

[0015] 图5是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的实施例的示意性流程图。

[0016] 图6是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统中的数据多路复用方法的实施例的逻辑图。

[0017] 图7是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的实施例的示意性流程图。

[0018] 图8是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统中的数据多路复用的方法的实施例的逻辑图。

[0019] 图9是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的实施例的示意性流程图。

[0020] 图10是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的性能的逻辑图。

[0021] 图11是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法的实施例的示意性流程图。

[0022] 图12是根据本实施例专门配置的计算系统的框图。

具体实施方式

[0023] 描述了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法和系统的实施例。所描述的实施例可以提高某些数据处理系统的性能。数据处理系统可以包括但不限于个人计算设备,诸如计算机工作站、膝上型计算机、平板计算机、个人数据助理、智能电话等。其他实施例可以与企业数据处理系统(诸如服务器、存储控制器、云服务器农场和数据农场以及其他企业数据中心系统和组件)结合使用。

[0024] 所描述的实施例为与主机系统相关联地操作的应用提供逻辑设备接口。在实施例中,逻辑设备将具有可选数据传输能力和QoS的多个物理端口组合成多个数据路径。可以根据多个数据传输模式之一来配置数据路径,数据传输模式包括“吞吐量增强模式”、“迅速模式”和“平衡模式”。在实施例中,吞吐量增强模式在配置用于DMA数据传输的端口上传输所有数据。在实施例中,迅速模式在配置用于DDIO数据传输的端口上传输所有数据。在实施例中,平衡模式根据所检查的数据特性在DMA端口和DDIO端口之间切换。另外,本实施例描述了在平衡模式中基于QoS要求的数据路径的多路复用,其中取决于QoS要求和数据特性,根据一组DMA或DDIO端口对IO块进行分组和传送。此外,本实施例包括用于在使用DDIO端口时防止CPU高速缓存溢出的流量控制的方法。

[0025] 普通技术人员将认识到本实施例的各种优点和益处。具体地,与现有系统相比,可以认识到QoS益处,包括系统等待时间减少。另外,根据一些实施例可以提高系统吞吐量。此外,与先前的DDIO或DCA技术相比时,在某些实施例中可以防止高速缓存溢出。尽管参考PCIe具体描述了本实施例,但是普通技术人员将认识到本实施例可以适用于的附加或备选系统,只要选择的数据传输协议提供在数据端口、存储器和处理器高速缓存之间的访问即可。

[0026] 图1是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统100的实施例的示意性框图。在实施例中,系统100包括主机系统102,以及可选地,包括一个或多个外围端点(endpoint)114。主机系统102可以是例如数据处理设备,诸如计算机。外围端点可以是被配置为通过PCIe或类似数据总线与主机系统102通信的外围设备。示例可以包括但不限于PCIe扩展设备、数据存储设备、刀片服务器等。

[0027] 在实施例中,主机系统102可以包括根联合体104,所述根联合体104形成用于在CPU 106、存储器108和一个或多个内部端点112或外围端点114之间传送数据的数据通信结构,内部端点112和外围端点114经由物理接口110耦合到根联合体104。根联合体104可以根据数据通信协议操作,所述数据通信协议诸如是但不限于PCIe。在实施例中,根联合体104和/或物理接口110可以包括如图4所示的多个数据通信端口。各种端点112、114可以经由根联合体104与CPU 106和/或存储器108传送数据。在实施例中,CPU 106可以包括或可以外部耦合到高速缓存116。高速缓存116的一个实施例是LLC。

[0028] 图2是示出了系统200的另一实施例的示意性框图。在一个实施例中,CPU 106可以包括多个CPU核202。系统200可类似地包括用于在CPU核202、存储器108和一个或多个端点之间传送数据的根联合体104。术语“端点”可以包括图1中所示的内部端点112和外围端点114。端点的示例可以包括图形处理单元204、一个或多个网络接口206、盘控制器208和其他PCIe端点210a。另外,系统200可以包括PCI交换机/到PCI/PCI-x的PCIe桥212。在一些备选实施例中,PCI交换机可以与PCIe桥分离。PCI交换机可以向多个PCIe端点210b-c提供交换

式通信。PCIe桥可以提供与根据传统PCI或PCI-x协议操作的传统端点214的连接性。

[0029] 图3是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的方法300的实施例的示意性流程图。在实施例中,方法300开始于接收来自软件应用的IO请求,其中IO请求包括用于通过数据访问端口传送数据的信息或指令,如框302所示。在框304,方法300包括确定用于通过数据访问端口传送数据的传输模式。传输模式可以从一组传输模式中选择,包括:面向吞吐量的模式、等待时间敏感模式和平衡模式。在框306,方法300包括响应于IO请求并根据数据传输模式,通过数据访问端口传送数据。

[0030] 图4是示出了专门配置用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统400的实施例的逻辑图。在实施例中,系统400的各方面可以被编程为根联合体104的组件或其他PCIe控制组件。系统软件或固件模块可以提供由应用使用的逻辑设备402。在实施例中,系统400可以包括数据通信控制器401,其可以包括具有传输模式选择器404的接收器接口403。另外,数据通信控制器401可以包括块分配单元405,其具有用于处理DMA请求的T0块分配器414、用于处理DDIO请求的LS块分配器416、以及用于根据数据特性的值在T0块分配器414和LS块分配器416之间切换块的块标识器412。数据通信控制器401还可以包括被配置为操作多个DMA端口420和/或多个DDIO端口422的多端口设备驱动器407。

[0031] 在实施例中,连接到存储端点(例如多端口存储设备428)的端口可以提供由应用使用的逻辑存储组件。在实施例中,系统400包括用于接收来自应用的IO请求的接收器接口403。接收器接口403可以包括传输模式选择器404。传输模式选择器404可以在多个数据传输模式之间进行选择。数据传输模式的示例包括吞吐量增强模式406、迅速模式410和平衡模式408。

[0032] 在实施例中,吞吐量增强模式406可以允许应用向存储器设备108传输数据。这样的实施例可以提供增强的数据吞吐量。在实施例中,迅速模式410可以通过将数据传送到与一个或多个CPU核202相关联的高速缓存116(例如LLC或其他高速缓存),来向应用提供与CPU 106的快速数据通信。平衡模式408可以基于从应用传送的数据特性,可选择地将数据传送到存储器108或高速缓存116。

[0033] 在这样的实施例中,在来自应用的IO请求中接收数据块。传输模式选择器404标识要应用的传输模式。在实施例中,传输模式选择器404可以响应于IO请求中的信息(例如数据特性)来选择传输模式。在另一实施例中,传输模式选择器404可以响应于确定是DMA端口可用还是DDIO端口可用来选择传输模式。在又一个实施例中,传输模式选择器404可以响应于预定的QoS协议、规则或负载均衡序列来确定传输模式。普通技术人员可以认识到,传输模式选择器404可以使用多种标准来确定要用于具体IO请求的传输模式。在又一个实施例中,传输模式选择器404可以根据时序或数据包计数规则来确定传输模式。

[0034] 传输模式选择器404可以将数据块传输给块分配单元405。在实施例中,块分配单元405可以包括面向吞吐量(T0)的块分配器414、等待时间敏感(LS)块分配器416、块标识器412和流量控制器418。

[0035] 如果选择吞吐量增强模式406,则数据块可以被传递到T0块分配器414。T0块分配器414可以将请求块划分为多个子块,每一个子块被传递到配置为通过DMA协议将数据传输到存储器108的一个或多个DMA端口420。在一些实施例中,可以在系统启动时,例如通过多端口设备驱动器或通过系统基本输入/输出系统(BIOS)来定义一个或多个DMA端口420。类

似地,可以在系统启动时定义一个或多个DDIO端口422。在备选实施例中,可以在操作期间并且响应于IO请求,将一个或多个DMA端口420动态地重配置为DDIO端口422。类似地,可以响应于IO请求,动态地将一个或多个DDIO端口422重配置为DMA端口420。

[0036] 在迅速模式410中,数据块可以传送到LS块分配器416,LS块分配器416然后将数据块或子块传递到DDIO端口422。在另一实施例中,流量控制器418可以对数据块或子块到DDIO端口的流量进行计量,以防止与CPU 106相关联的高速缓存116的溢出或污染。下面关于图10-12描述了流量控制方法的实施例。在实施例中,多端口存储设备428可以包括多个DMA端口424和/或多个DDIO端口426,被配置为将数据传递到多端口存储设备428的组件。

[0037] 在实施例中,平衡模式408可以可选择地将数据块传递到用于通过DMA端口420通信的T0块分配器414或用于通过DDIO端口422通信的LS块分配器416。在这样的实施例中,块标识器412可以检查数据块的数据特性,并且根据数据特性确定是将数据块传递到T0块分配器414还是将数据块传递到LS块分配器416。在实施例中,数据特性可以由应用设置的紧急标识符。例如,数据特性将数据块表征为紧急或非紧急。在这样的实施例中,紧急数据块可以传递到LS块分配器416,非紧急数据块可以传递到T0块分配器414。

[0038] 在实施例中,数据特性可以是标签或与数据块相关联的元数据。备选地,数据特性可以是在数据块内设置的比特。在又一个实施例中,数据特性可以是响应于用户或提供数据块的应用或者响应于所提供的数据块的类型而固有地导出的。在另一实施例中,一些用户或应用可以生成紧急数据块,而其他用户或应用可以生成非紧急数据块。在实施例中,用户请求的块可能是紧急的,因为用户可以在块准备好访问时就立即访问该块,而预取块可能是非紧急的,因为该块可能或可能不会在近期被访问。普通技术人员将认识到附加的或替代的数据特性,其可由块标识器412用来确定是将数据块发送到T0块分配器414还是发送到LS块分配器416。

[0039] 图5示出了描述由图4的结构执行的方法500的流程图。在这样的实施例中,方法500包括接收来自应用的IO请求,如框502所示。在框504处,传输模式选择器404可以基于数据特性选择传输模式,如框504所示。如果所有端口被设置为DMA端口420,则选择吞吐量增强模式406,并且将所有数据块发送到T0块分配器414。如果所有端口被设置为DDIO端口422,则选择迅速模式410,并且所有数据块被引导到LS块分配器416。如果一部分端口被设置为DMA端口420并且一部分端口被设置为DDIO端口422,则可以在框504处选择平衡模式。在另一实施例中,不需要将所有端口设置为一种类型或另一种类型。吞吐量请求可以仅向作为DMA的端口子集发送请求,而不使用DDIO端口。在这样的实施例中,在框516处,数据块可以被传递到块标识器412以对块进行标识。如框506所示,“正常”或“非紧急”块可被传递到T0块分配器414。紧急块可以被传递到LS块分配器416,如框510所示。在框512处,可以执行流量控制过程,并且可以在框514处发出DDIO请求。备选地,可在框508处发出DMA请求。

[0040] 图6是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理系统中的数据多路复用的过程600的实施例的逻辑图。在实施例中,图6的过程可以由T0块分配器414或DMA传输路径中的关联组件来执行。在实施例中,可以接收和检查所请求的数据块602。数据块602可以划分为多个子块604(例如,八(8)个或更多个子块)。在框606处,可以关于子块大小执行确定。例如,如果数据块602中的子块604的数量小于或等于可用于传送子块604的DMA端口420的数量,则可以维持每一个子块604的当前大小。然而,如果在框606处

确定子块604的数量大于可用于传输子块604的DMA端口420的数量,则可以调整每一个子块604的大小,并且可以减少子块604的总数。例如,对应于子块604数量小于或等于DMA端口420数量的情况的第一结果610可以具有与初始请求的数据块602相同的子块配置。第二结果608可以具有更少的子块,对于通过DMA端口420的通信,每一个子块的大小更大。在框612处,按照循环序列向端口420分配子块。

[0041] 图7是示出了图6所描述过程的实施例的示意性流程图。在图7的实施例中,方法700包括在702处接收数据块。在704处,方法700包括计算在接收的数据块中包括的子块的数量。子块的数量可以被计算为接收的数据块的大小除以最小数据传输单元。该计算的结果可以是对图6所示的初始数据块602中的子块604的标识。

[0042] 在图7的方法中,在706处,确定子块的数量是否小于或等于可用DMA端口420的数量。如果是,则子块的大小等于接收的数据块的大小除以子块的数量,如在708处所示。如果不是,则在710处调整子块的大小。在这样的实施例中,子块的大小可以增加至所请求的块的大小除以可用DMA端口420的数量。在图6的示例中,可以在710处生成第二结果608。

[0043] 一旦在708或710处确定了子块的大小,则将DMA请求与子块一起发送到设备驱动器,如框712所示。如果在714处确定已经到达数据块的结尾,则过程结束。如果不是,则子块被递增,如在716处所示,并且在712处发送下一个DMA请求。该循环重复,直到数据块中的所有子块都已经通过DMA端口420发送。

[0044] 图8是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知IO管理的系统中数据多路复用的过程800的实施例的逻辑图。图8的实施例示出了迅速模式的示例,其中数据块通过DDIO端口422发送。过程800可以至少部分地由LS块分配器416和流量控制器418执行。如同图6的实施例,过程800开始于接收要传送的数据块602。数据块602可以包括若干个子块604。在802处,确定子块604的数量是否小于或等于可用于传送子块604的DDIO端口422的数量。如果是,则第一结果804包括初始大小的子块。如果不是,则第二结果806包括调整大小的子块,其中块大小被修改以适应可用于传送子块的DDIO端口的数量。在808处,附加步骤包括确定子块的大小是否大于DDIO高速缓存大小。如果是,则调整大小以适应DDIO高速缓存大小要求。在812处,执行流量控制操作以防止高速缓存超限或污染。参考图12描述了流量控制过程的示例。在810处,按照循环方式将子块分配到DDIO端口422。

[0045] 图9是示出了图8中描述的过程的实施例的示意性流程图。在图9的实施例中,方法900包括在902处接收数据块。在904处,方法900包括计算在接收的数据块中包括的子块的数量。子块的数量可以被计算为接收的数据块的大小除以最小数据传输单元。这种计算的结果可以是对初始数据块602中的子块604的标识,如图8所示。

[0046] 在图9的方法中,在906处,确定子块的数量是否小于或等于可用DDIO端口422的数量。如果是,则子块的大小等于接收的数据块的大小除以子块的数量,如在908处所示。如果不是,则在910处调整子块的大小。在这样的实施例中,子块的大小可以被增加至所请求的块的大小除以可用的DDIO端口422的数量。在图8的示例中,可以在910处生成第二结果806。

[0047] 在实施例中,在912处,确定在910处调整的子块的大小是否小于LLC中DDIO的预留大小。如果是,则在912处执行流量控制。如果不是,则对子块进行大小调整,调整为小于或等于LLC中的DDIO的预留大小。并且在912处执行流量控制。

[0048] 一旦确定了子块的大小并执行了流量控制,则如在916处所示,将DDIO请求与子块

一起发送到设备驱动器。如果在918处确定已经到达数据块的结尾,则过程结束。如果不是,则子块被递增,如在920处所示,并且执行512处的流量控制,并在916处发送下一个DDIO请求。该循环重复,直到数据块中的所有子块都已经通过DDIO端口422发送。

[0049] 图10是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知I/O管理的方法的性能的逻辑图。图10示出了与CPU核202相关联的高速缓存116,诸如LLC。在所描述的实施例中,LLC包括多个高速缓存行(CL)。被占用的高速缓存行是包含由CPU核202检索的数据的高速缓存行。剩余的高速缓存行是未被占用的或可用的高速缓存行,其中可以存储新接收的子块。被占用的高速缓存行和剩余的高速缓存行的组合包括LLC的总共预留的高速缓存行。

[0050] 图10示出了三种高速缓存状态的示例。在第一状态1002中,三个高速缓存行被接收的数据占用。在实施例中,高速缓存行大小为64字节。在一个实施例中,LLC的访问等待时间可以是100ns,并且请求间(inter-request)到达时间可以是200ns。因此,对于每个请求到达,可以通过CPU核202检索高速缓存行来确保两个附加高速缓存行。因此,在第二状态1004中,仅一个高速缓存行被占用。如果接收到具有256字节数据的子块的第二请求,则对于总共五个被占用的高速缓存行,可以占用四个附加高速缓存行,如第三状态1006所示。

[0051] 图11是示出了用于具有可重配置多端口的PCIe存储系统的QoS认知I/O管理的方法的实施例的示意性流程图。用于流量控制的方法1100的实施例可以至少部分地由流量控制器418执行。流量控制器418可以检查高速缓存116的状态,并通过触发溢出保护状态来防止高速缓存超限。如本文所使用的,术语“溢出保护状态”是指如下逻辑或操作状态:其中通过节制或计量到高速缓存116的数据流来防止高速缓存超限。在实施例中,在步骤1102处接收数据块。所访问的高速缓存行的数量计算为请求间到达时间除以平均高速缓存行访问时间,如在1104处所示。在1106处,通过将所访问的高速缓存行的数量与先前剩余的高速缓存行的数量相加来计算剩余的高速缓存行。在图10的第二状态1004中看到这样的示例。

[0052] 可以触发高速缓存保护状态以防止高速缓存的超限。不是将整个数据块写入高速缓存,而是基于剩余高速缓存行的数量,将该块可以分解成用于传送至高速缓存的合适大小的子块。如果需要,高速缓存保护状态可以使子块被顺序地写入或者间隔开以防止高速缓存超限。在另一实施例中,高速缓存保护状态可以使数据块被写入存储器108,而不是高速缓存116。因此,在1108处,确定剩余的高速缓存行是否大于总共预留的高速缓存行。如果是,则剩余的高速缓存行等于总共预留的高速缓存行,如1110处所示。如果不是,则在1112处确定高速缓存行中子块的数量是否小于剩余的高速缓存行。如果是,则剩余的高速缓存行被计算为剩余的高速缓存行的初始数量减去高速缓存行中子块的数量,如在1114处所示。如果不是,则循环地执行计算,直到高速缓存行中子块的数量小于剩余的高速缓存行。

[0053] 图12是包括PCIe数据通信控制器401的计算系统1200的框图,PCIe数据通信控制器401包括图4的接收器接口403、块分配单元405和多端口设备驱动器407。计算系统1200可以包括时钟1204、诸如随机存取存储器(RAM)之类的存储器108、用户接口1208、诸如基带芯片组之类的调制解调器1210、固态驱动/盘(SSD)1214和/或CPU106,这些组件中任何一个或全部可以电耦合到系统总线1202。计算系统1200可以包括PCIe数据通信控制器401,包括图4的接收器接口403、块分配单元405和多端口设备驱动器407,其也可以电或逻辑耦合到系统总线1202。

[0054] 本实施例可以包括仲裁或看门狗定时器,以防止等待时间敏感的IO事务(例如视频流)压倒IO队列。因此,性能增强IO请求将有机会向前推进。另外,实施例也可以容纳随机(小随机读/写)IO事务。

[0055] 以下讨论旨在提供其中可以实现本发明构思的某些方面的合适的一个或多个机器的简要、通用的描述。通常,一个或多个机器包括系统总线,附着到其的有处理器、存储器(例如随机存取存储器(RAM))、只读存储器(ROM)或其他状态保存介质、存储设备、视频接口和输入/输出接口端口。一个或多个机器可以至少部分地通过来自常规输入设备(诸如键盘,鼠标等)的输入以及通过从另一机器接收的指示、与虚拟现实(VR)环境的交互、生物测定反馈或其他输入信号来进行控制。如本文所使用的,术语“机器”旨在广义地涵盖单个机器、虚拟机或者通信耦合的机器、虚拟机或一起操作的设备的系统。示例性机器包括诸如个人计算机、工作站、服务器、便携式计算机、手持设备、电话、平板电脑等的计算设备,以及诸如私人或公共运输(例如汽车、火车、出租车等)之类的运输设备。

[0056] 一个或多个机器可以包括嵌入式控制器,诸如可编程或不可编程逻辑器件或阵列、专用集成电路(ASIC)、嵌入式计算机、智能卡等。一个或多个机器可以利用到一个或多个远程机器的一个或多个连接,例如通过网络接口、调制解调器或其他通信耦合。机器可以通过物理和/或逻辑网络(例如内联网、互联网、局域网、广域网等)的方式进行互连。本领域技术人员将理解,网络通信可以利用各种有线和/或无线短距离或长距离载波和协议,包括射频(RF)、卫星、微波、电气和电子工程师协会(IEEE) 545.11、蓝牙®、光学、红外、电缆、激光等。

[0057] 本发明构思的实施例可以通过参考或结合包括功能、过程、数据结构、应用程序等的关联数据来描述,当由机器访问时,其导致机器执行任务或定义抽象数据类型或低级硬件上下文。关联数据可以被存储在例如易失性和/或非易失性存储器(例如RAM、ROM等)中或存储在其他存储设备及其相关联的存储介质中,包括硬盘驱动、软盘、光学存储器件、磁带、闪存、记忆棒、数字视频盘、生物存储器件等。关联数据可以以数据包、串行数据、并行数据,传播信号等的形式通过传输环境(包括物理和/或逻辑网络)传递,并且可以以压缩或加密格式使用。关联数据可以在分布式环境中使用,并且本地和/或远程地存储以用于机器访问。

[0058] 已经参考所示实施例描述和说明了本发明构思的原理,将认识到,所示实施例可以在布置和细节上修改,而不脱离这些原理,并且可以以任何期望的方式组合。并且尽管前述讨论集中在具体实施例上,但是设想了其他配置。具体地,尽管在本文使用了诸如“根据本发明构思的实施例”等表达,但是这些短语意指一般地参考实施例可能性,并且不旨在将本发明构思限于具体实施例配置。如本文所使用的,这些术语可以提及可组合到其他实施例中的相同或不同的实施例。

[0059] 本发明构思的实施例可以包括包含可由一个或多个处理器执行的指令的非瞬时机器可读介质,所述指令包括用于执行如本文所述的发明构思的要素的指令。

[0060] 前述说明性实施例不应被解释为限制其发明构思。尽管描述了一些实施例,然而本领域技术人员应容易理解,在不实质上脱离本公开的新颖教义和优点的前提下,可以对那些实施例进行多种修改。因此,所有这种修改旨在被包括在如在权利要求中限定的本发明构思的范围内。

100

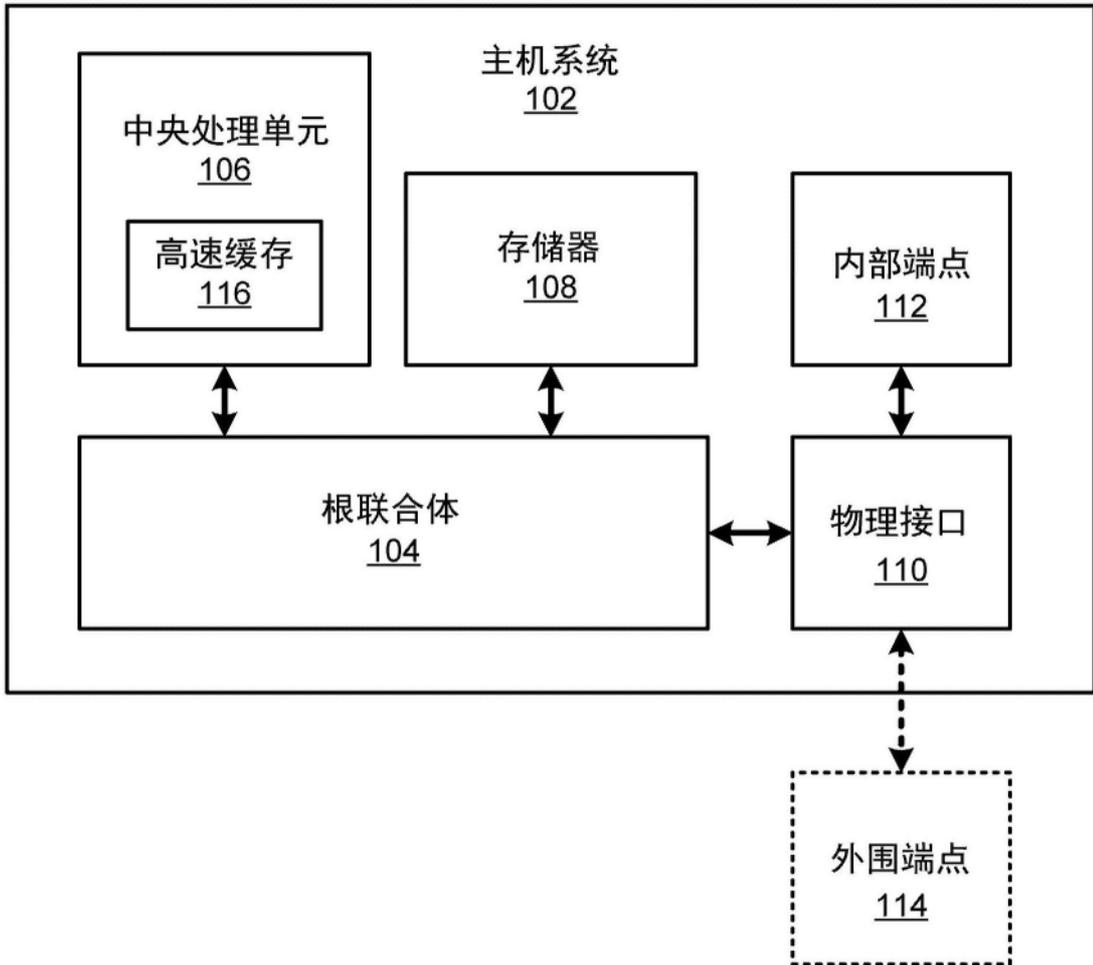


图1

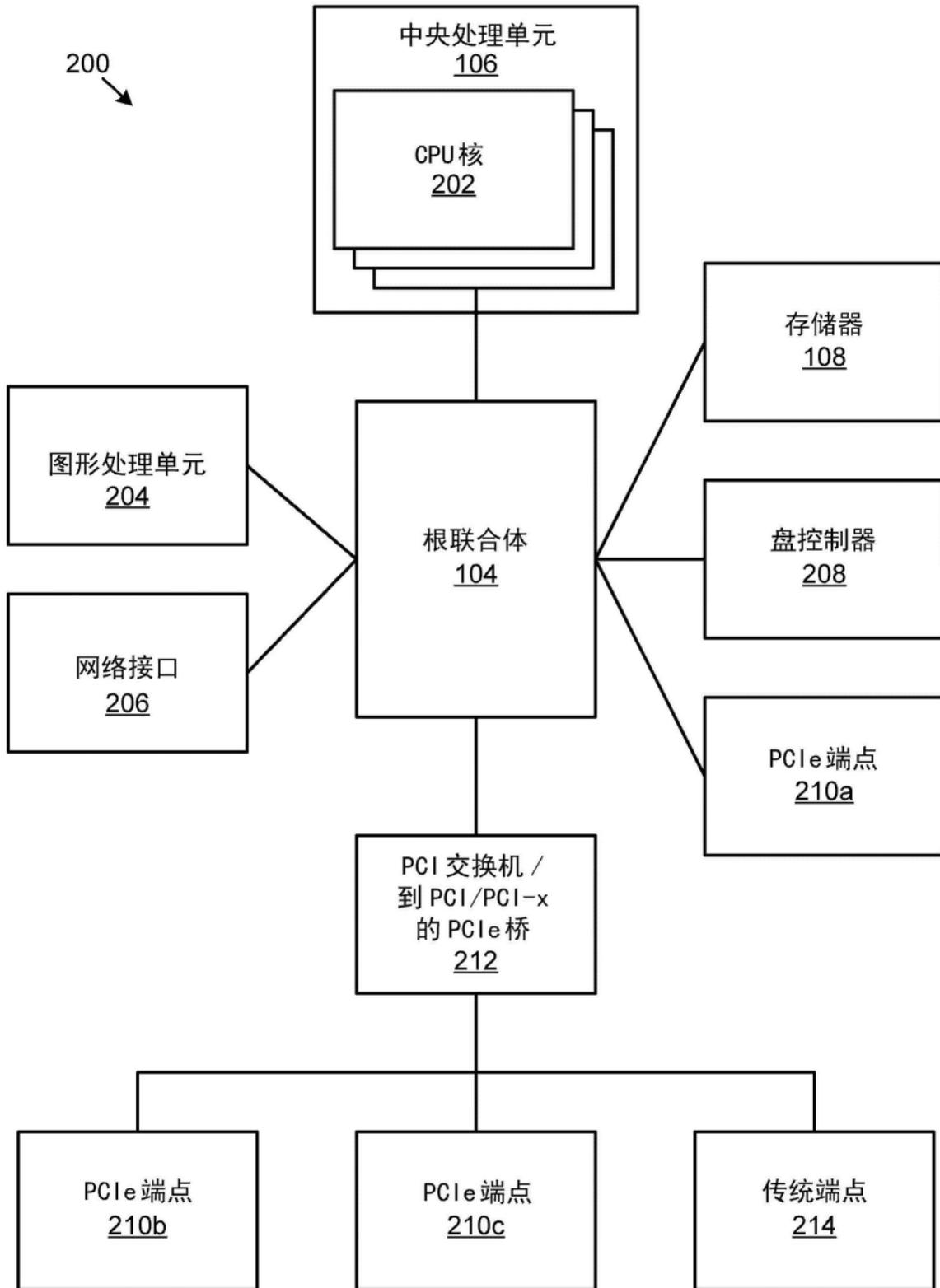


图2

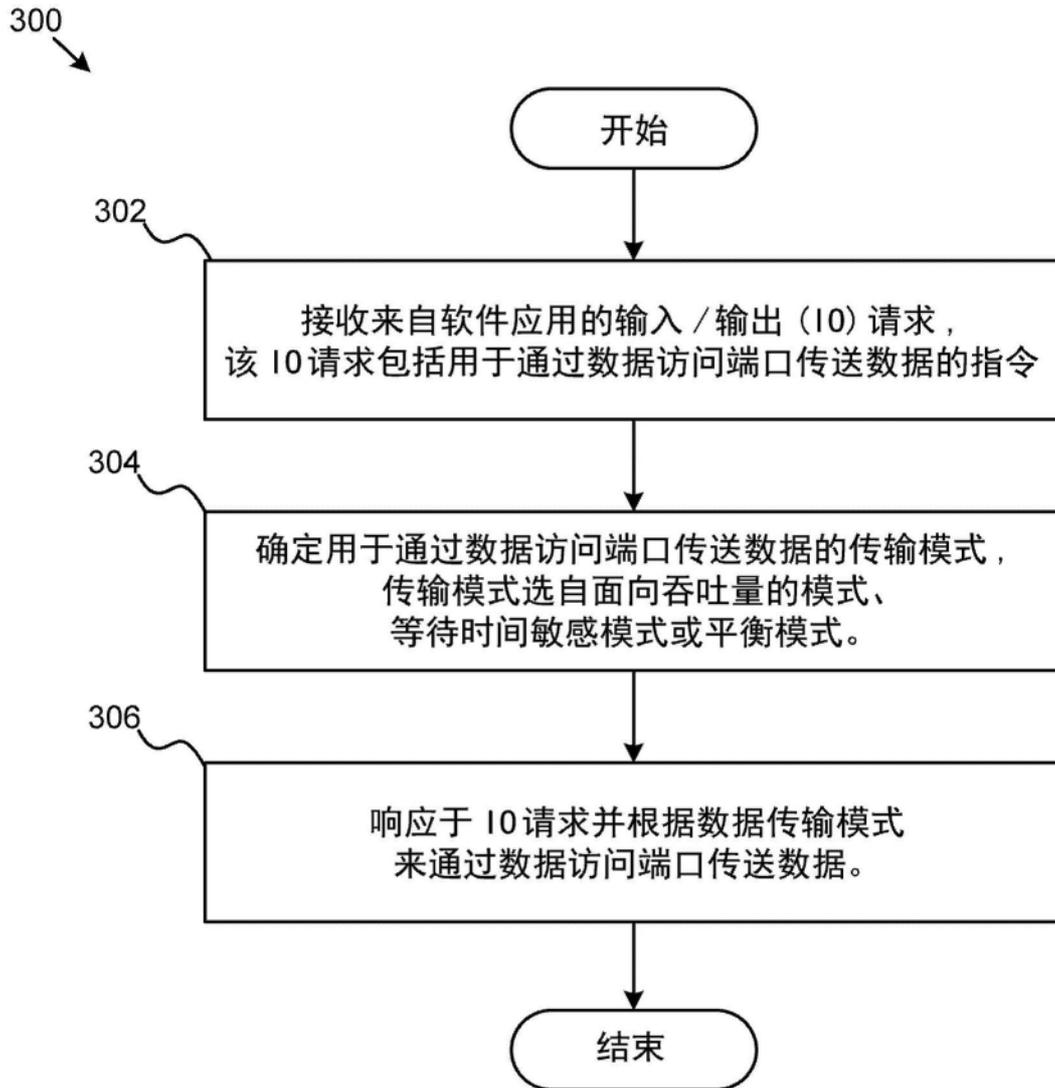


图3

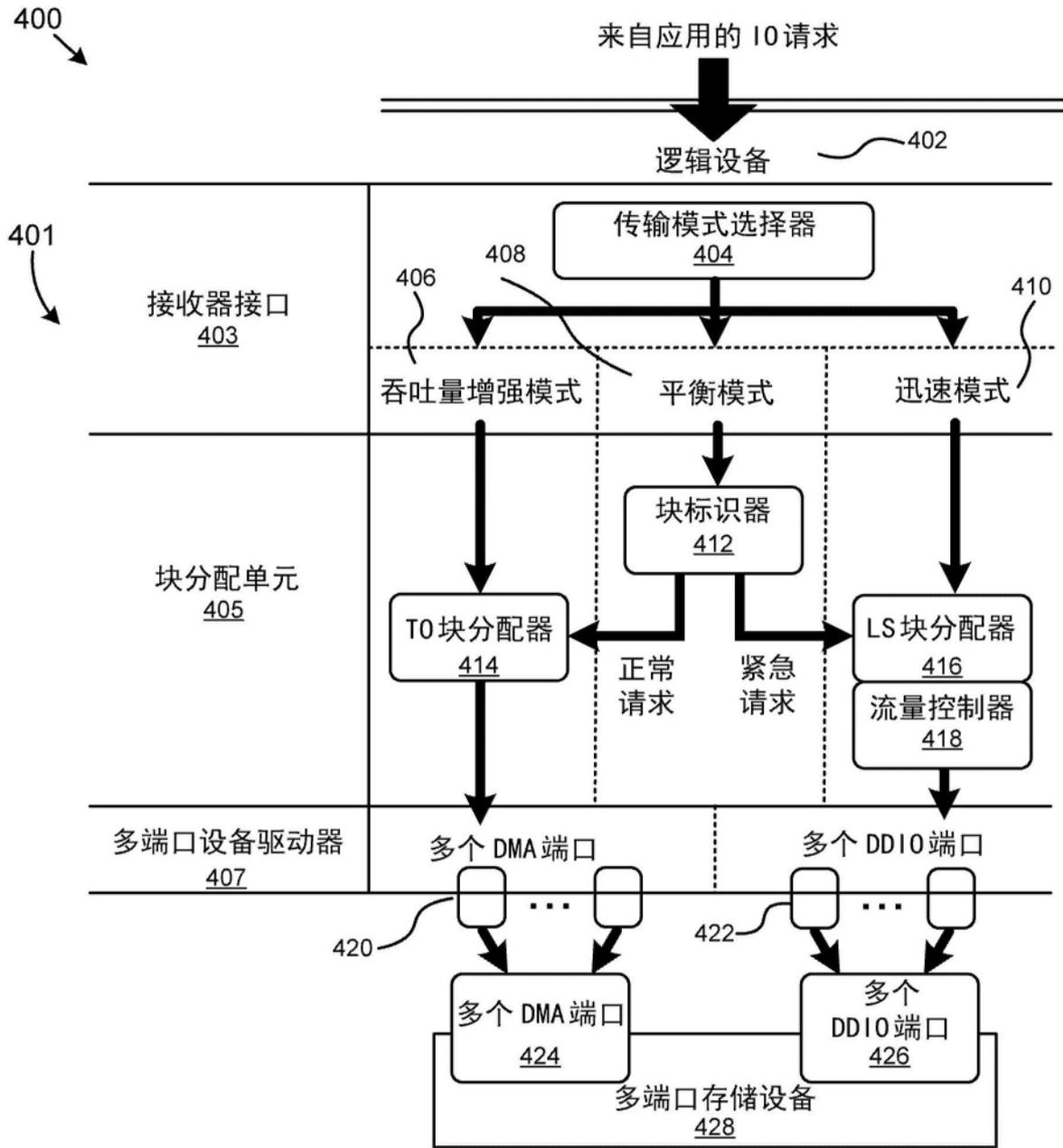


图4

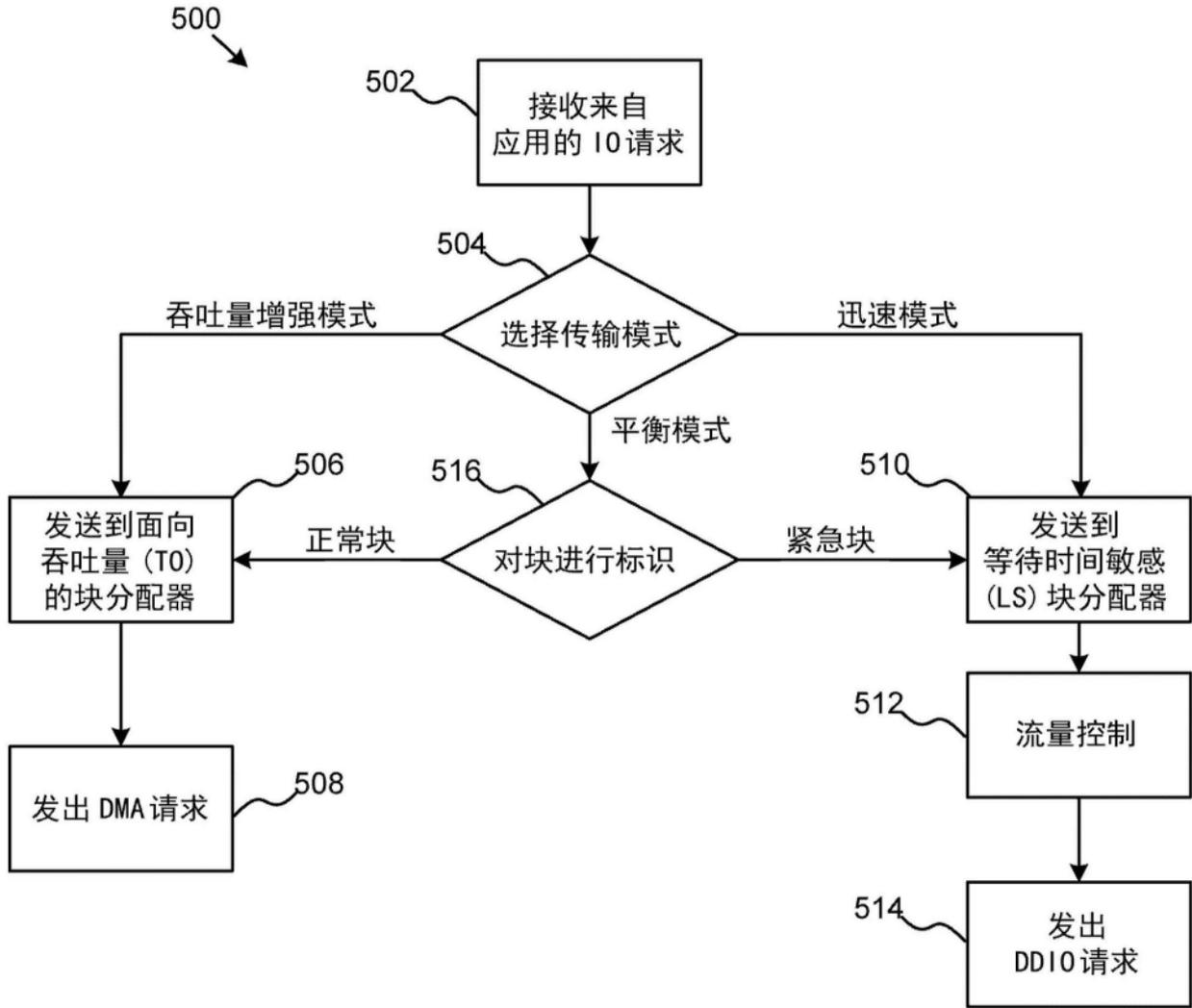


图5

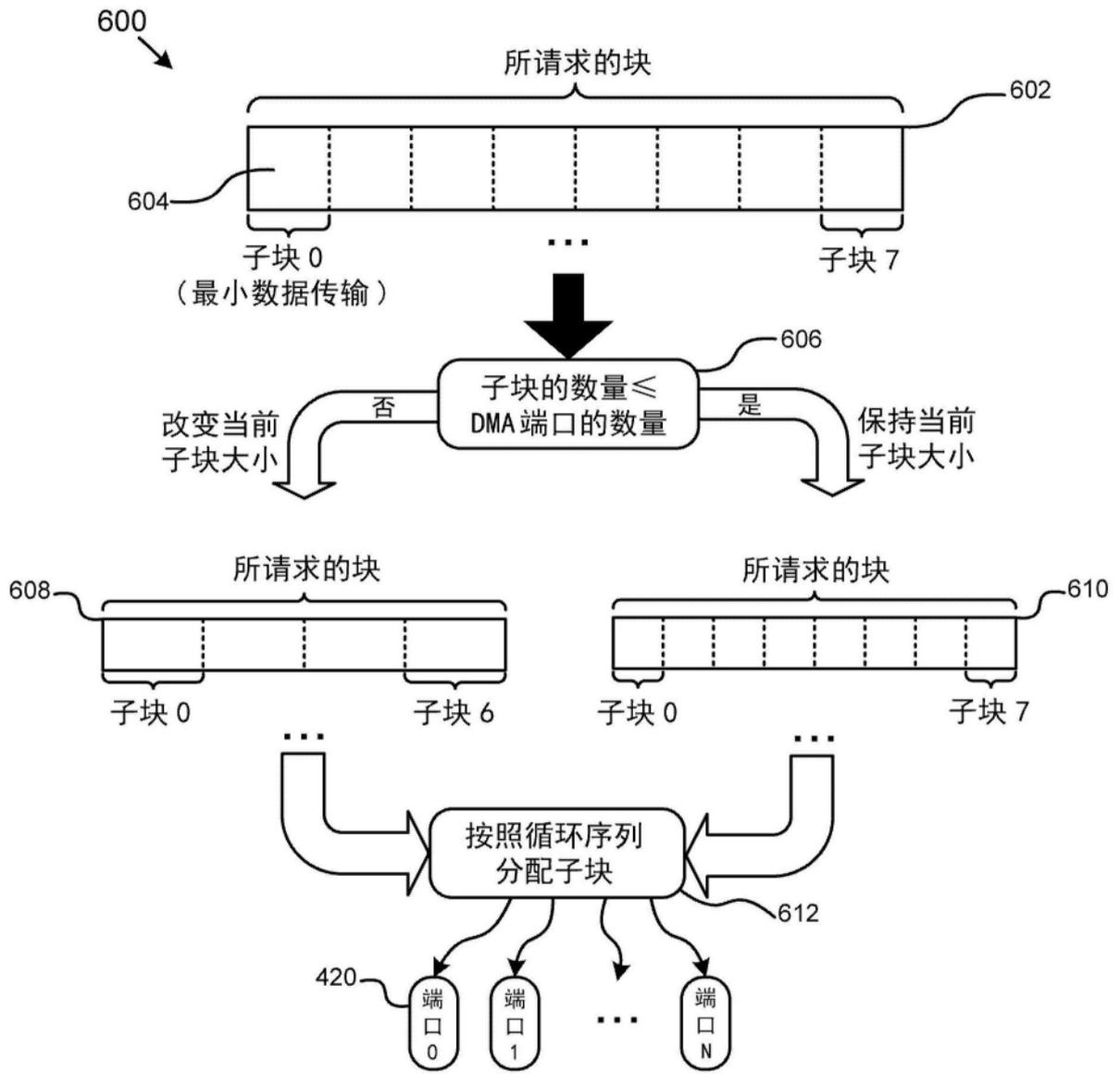


图6

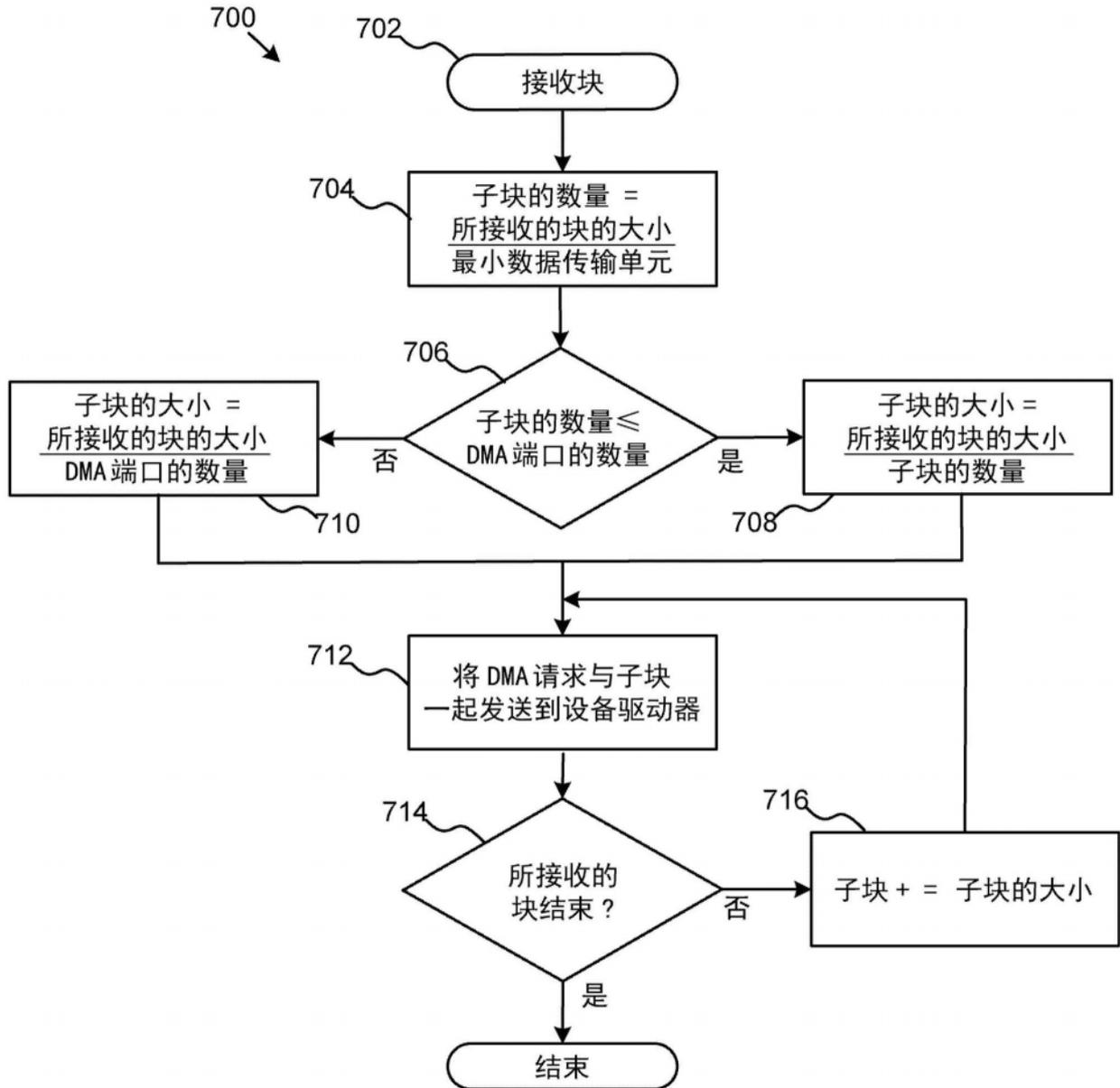


图7

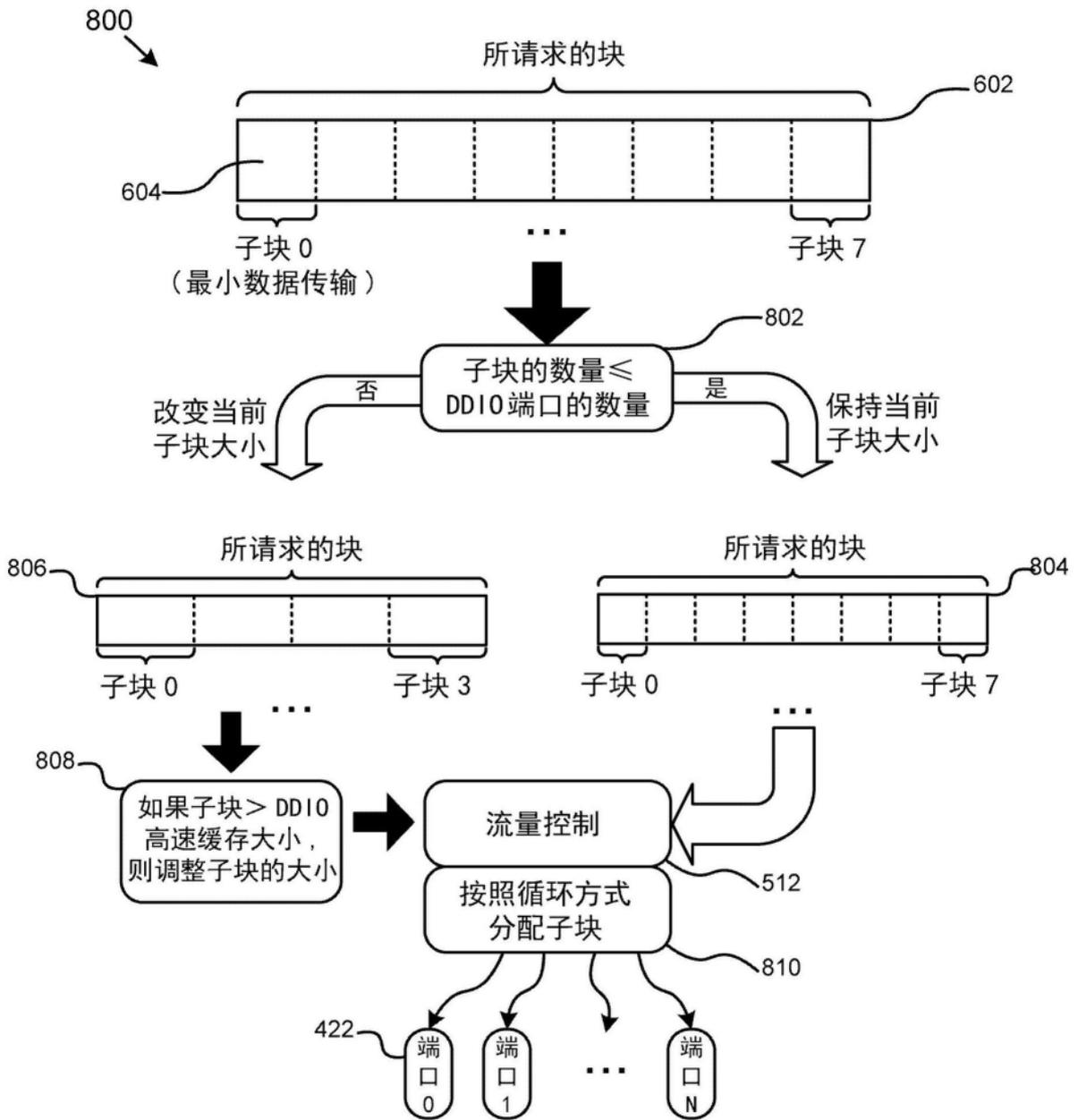


图8

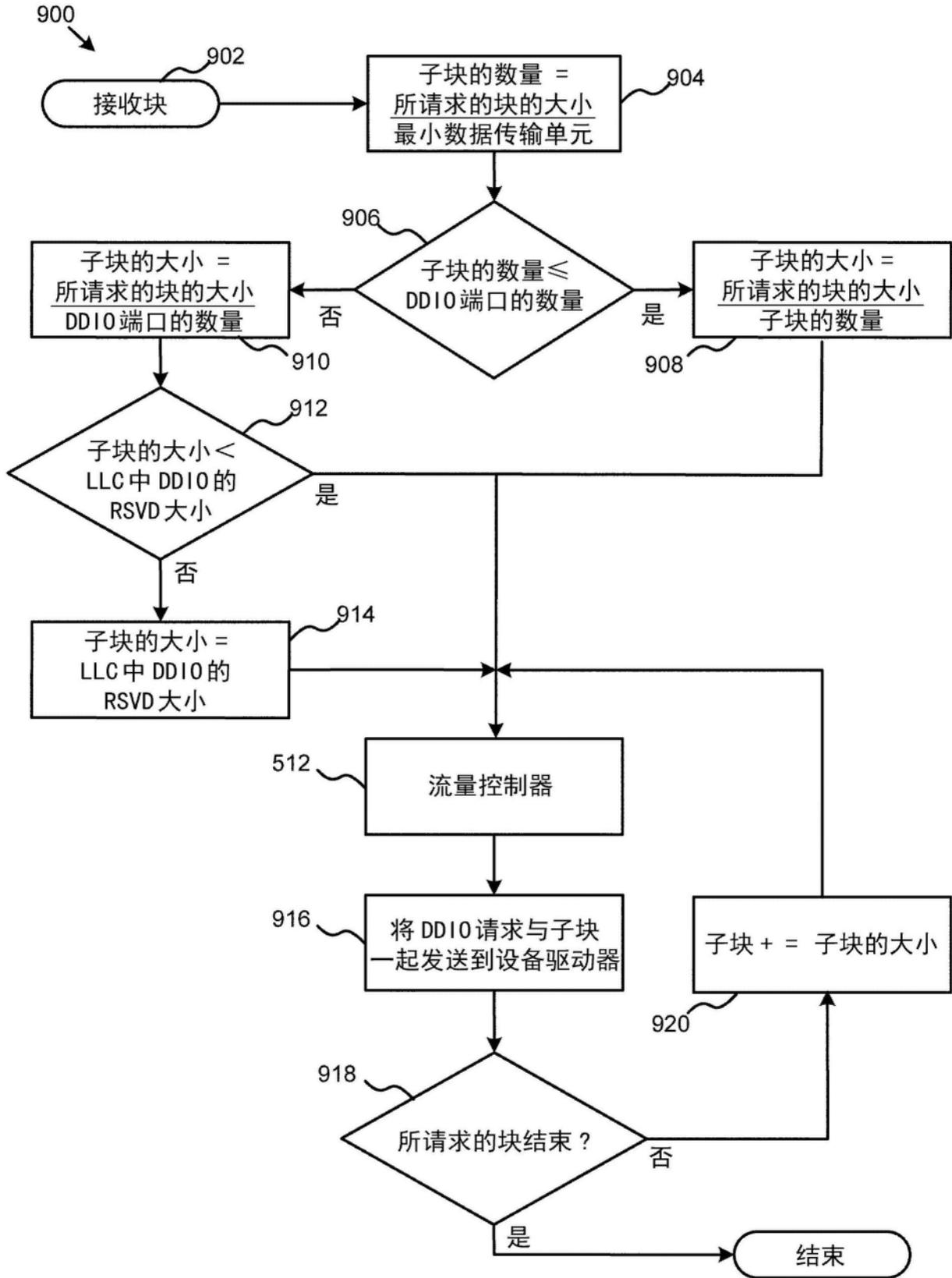


图9

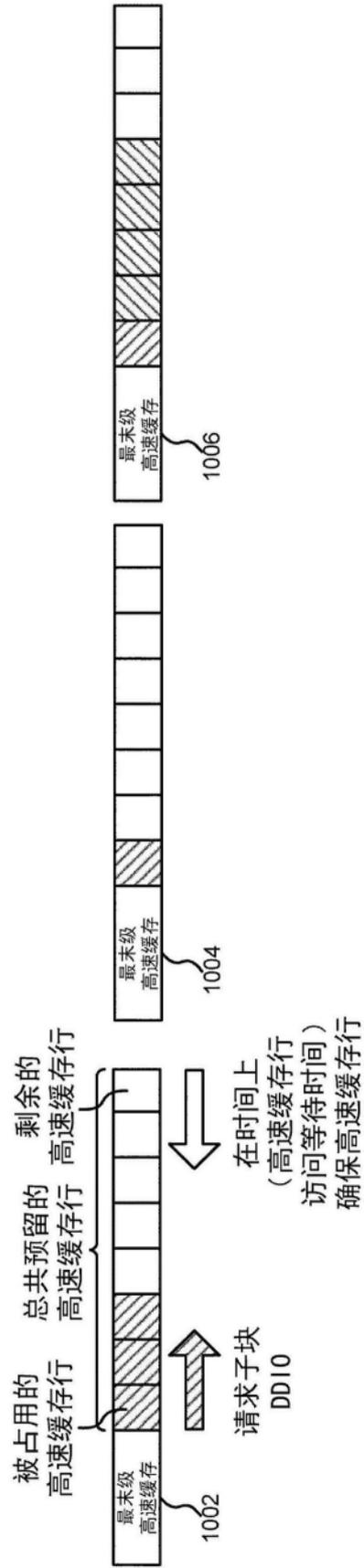


图10

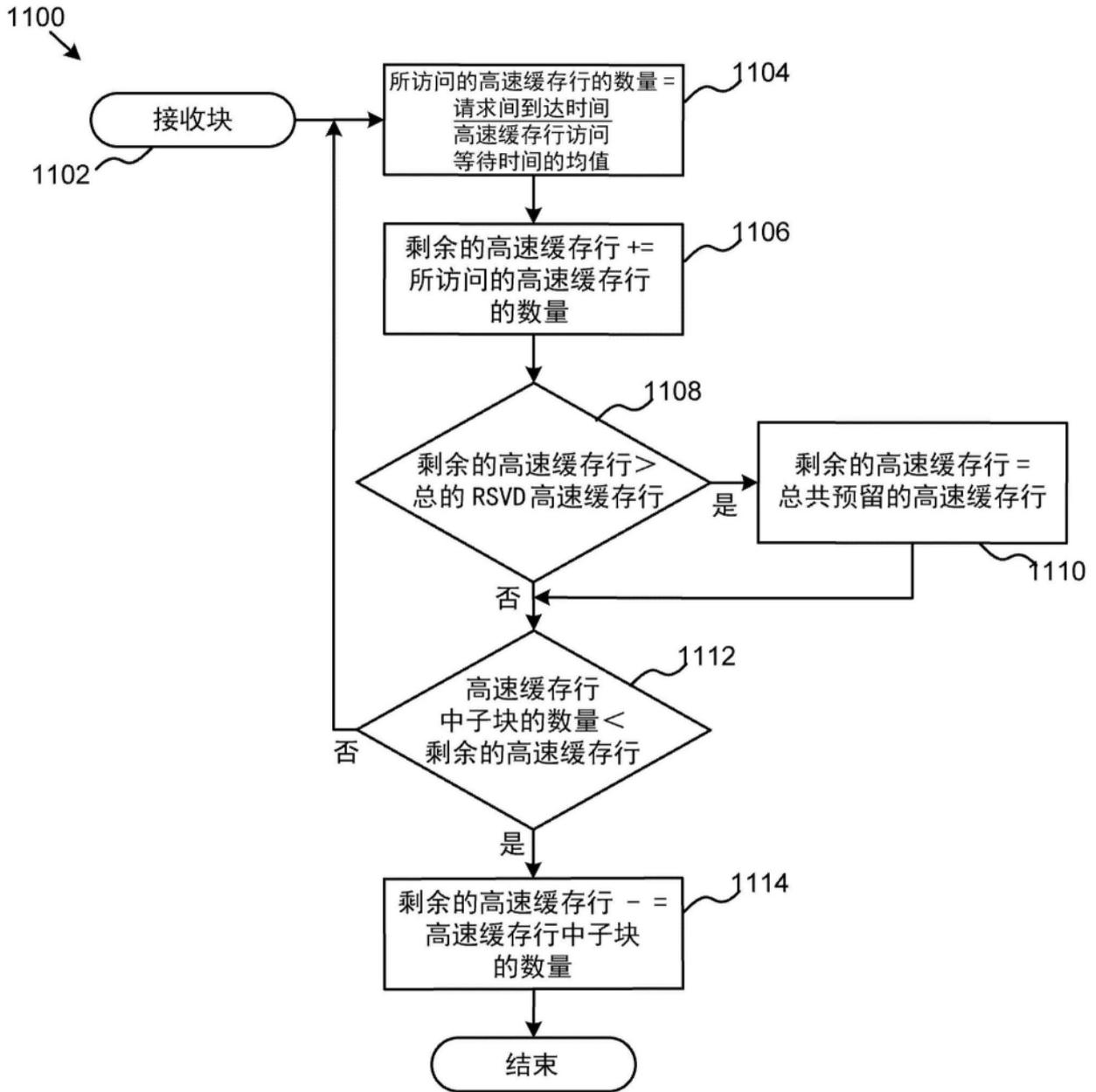


图11

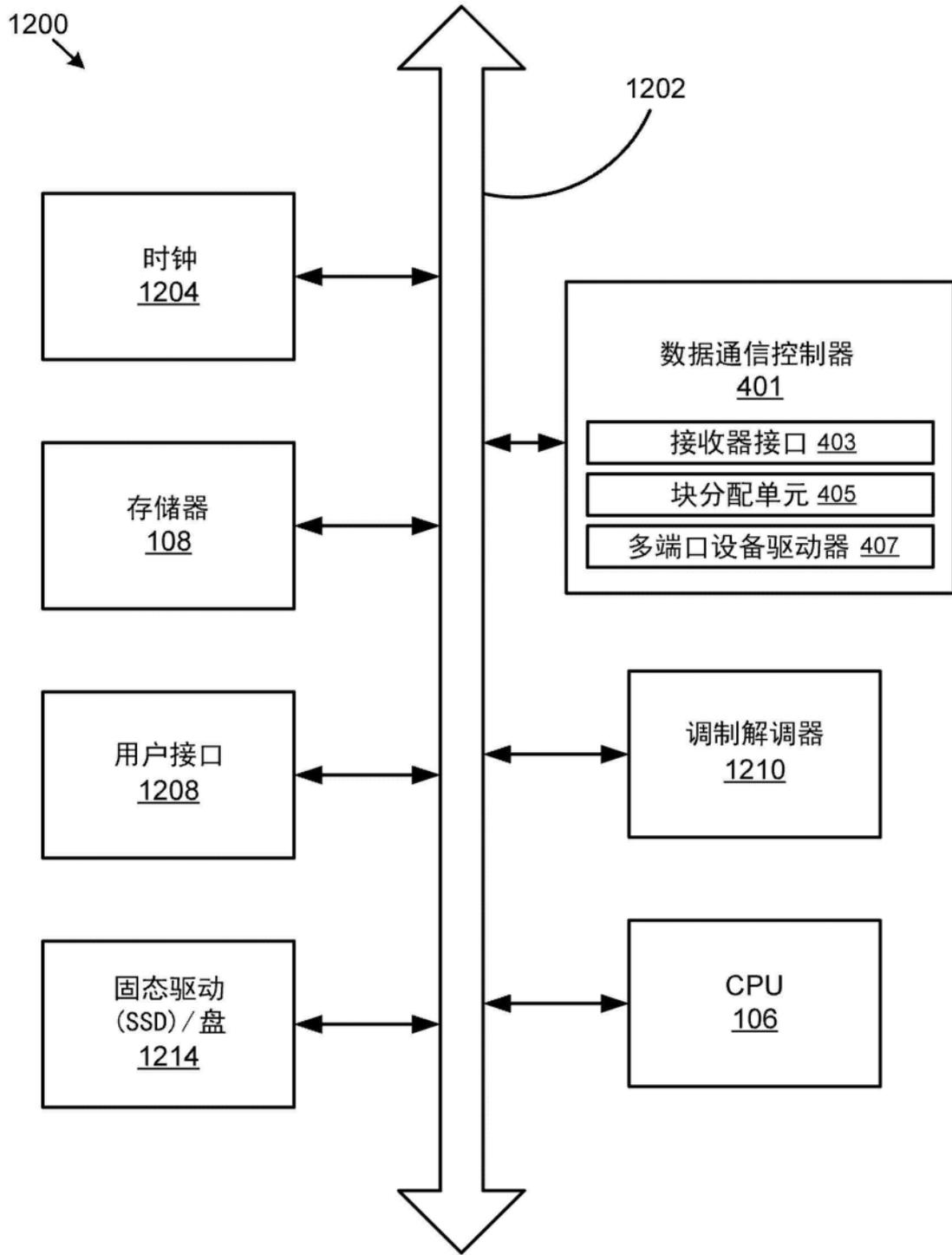


图12