



(12)发明专利申请

(10)申请公布号 CN 108463077 A

(43)申请公布日 2018.08.28

(21)申请号 201810296818.0

(22)申请日 2018.04.03

(71)申请人 郑州云海信息技术有限公司
地址 450018 河南省郑州市郑东新区心怡路278号16层1601室

(72)发明人 李纪伟

(51)Int. Cl.
H05K 7/10(2006.01)

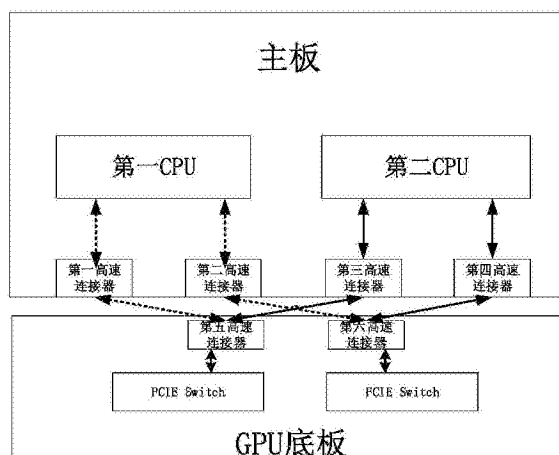
权利要求书1页 说明书4页 附图3页

(54)发明名称

一种互联板卡组合

(57)摘要

本申请提供了一种互联板卡组合,包括:主板,主板上包括第一CPU和第二CPU,第一CPU与主板上的第一高速连接线和第二高速连接器分别通过PCIE接口相连,所述第二CPU与主板上的第三高速连接线和第四高速连接器通过PCIE接口相连;GPU底板,GPU底板上包括第五高速连接器和第六高速连接器,第五高速连接器和第六高速连接器用于与主板上的第一CPU或第二CPU连接,所述GPU主板包括两个PCIE Switch,所述每个PCIE Switch至少包括4个下行端口,所述两个PCIE Switch用于通过第一提升板和第二提升板分别与8颗GPU相连;两个第一提升板和两个第二提升板,其中,所述第一提升板与所述第二提升板配合使用。



1. 一种互联板卡组合,其特征在于,包括:

主板,所述主板上包括第一CPU和第二CPU,所述第一CPU与所述主板上的第一高速连接线和第二高速连接器分别通过PCIE接口相连,所述第二CPU与所述主板上的第三高速连接线和第四高速连接器通过PCIE接口相连;

GPU底板,所述GPU底板上包括第五高速连接器和第六高速连接器,所述第五高速连接器和所述第六高速连接器用于与所述主板上的第一CPU或第二CPU连接,所述GPU主板包括两个PCIE Switch,所述每个PCIE Switch至少包括4个下行端口,所述两个PCIE Switch用于通过第一提升板和第二提升板分别与8颗GPU相连;

两个第一提升板和两个第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

2. 根据权利要求1所述的方法,其特征在于,所述第二提升板长度比所述第一提升板长度短13cm。

3. 根据权利要求1或2所述的方法,其特征在于,所述PCIE Switch为PEX9797芯片。

4. 根据权利要求1至3中任一项所述的方法,其特征在于,所述PCIE Switch为包括5个下行端口,其中,4个下行端口用于扩展GPU卡,1个下行端口用于扩展无限带宽IB卡。

5. 一种提升板组合,其特征在于,包括第一提升板和第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

一种互联板卡组合

技术领域

[0001] 本申请涉及计算机领域,并且更具体地,涉及一种互联板卡组合。

背景技术

[0002] 从计算机诞生之初,人工智能就是人类对计算机的终极追求。随着近几年计算机水平迅猛提升,语音识别、大数据运算、汽车自动驾驶功能、甚至围棋领域,人工智能都成绩斐然,这些应用都基于深度学习算法,而目前支持该算法的硬件平台大都基于图像处理器(英文:Graphics Processing Unit,简写:GPU)强大的并行运算能力。

[0003] 为满足超强运算能力的需求,当前的服务器设计通常采用多颗CPU且至少4U高度的机箱设计,将多个标准的GPU卡集成在机箱内部。具体设计为:CPU的高速串行计算机扩展总线标准(英文:peripheral component interconnect express,简写PCIE)资源直接与PCIE插槽(PCIE Slot)互联,并将GPU安装在PCIE Slot实现多GPU的集成。

[0004] 现有的多CPU方案,不仅增加硬件开发成本、也增加客户采购成本;而4U设计方案严重降低空间利用率,导致客户部署机房时需要更大的空间成本。基于上述两点,均不利于此类型服务器的市场推广。

[0005] Intel Purley平台的每颗CPU最大支持48个通道(Lane),对于传统服务器应用已经足够,但对于高密度服务器、大运算应用显然无法满足要求。一般来讲,高端显卡因3D运算吞吐量巨大通常需要X16通道,此类服务器平台对PCIE资源有特殊需求,PCIE交换芯片(PCIE Switch)即为应对通道数量限制开发。

[0006] 传统的开发设计中,PCIE Slot位于GPU BOX板卡、与机箱底部垂直,多颗GPU平行于机箱侧壁插入PCIE Slot,故机箱需要至少4U高度。

[0007] 因此,如何能够将8颗GPU集成在2U高度的机箱内,且能够放置在标准的服务器机架中,是目前亟待解决的问题。

发明内容

[0008] 本申请提供一种互联板卡组合,能够将8颗GPU集成在2U高度的机箱内。

[0009] 一方面,提供了一种互联板卡组合,包括:主板,所述主板上包括第一CPU和第二CPU,所述第一CPU与所述主板上的第一高速连接线和第二高速连接器分别通过PCIE接口相连,所述第二CPU与所述主板上的第三高速连接线和第四高速连接器通过PCIE接口相连;GPU底板,所述GPU底板上包括第五高速连接器和第六高速连接器,所述第五高速连接器和所述第六高速连接器用于与所述主板上的第一CPU或第二CPU连接,所述GPU主板包括两个PCIE Switch,所述每个PCIE Switch至少包括4个下行端口,所述两个PCIE Switch用于通过第一提升板和第二提升板分别与8颗GPU相连;两个第一提升板和两个第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配

合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

[0010] 结合第一方面,在第一方面的第一种可能的实现方式中,所述第二提升板长度比所述第一提升板长度短13cm。

[0011] 结合第一方面及其上述实现方式,在第一方面的第二种可能的实现方式中,所述PCIE Switch为PEX9797芯片。

[0012] 结合第一方面及其上述实现方式,在第一方面的第三种可能的实现方式中,所述PCIE Switch为包括5个下行端口,其中,4个下行端口用于扩展GPU卡,1个下行端口用于扩展无限带宽IB卡。

[0013] 第二方面,提供一种提升板组合,包括第一提升板和第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

[0014] 本申请通过更改硬件结构设计,能够将8颗GPU集成在2U高度的机箱内,且能够放置在标准的服务器机架中,方便客户安装、拆卸。

附图说明

[0015] 图1示出了本申请一个实施例的互联板卡。

[0016] 图2示出了本申请另一个实施例的互联板卡。

[0017] 图3示出了本申请一个实施例的提升卡。

[0018] 图4示出了本申请另一实施例的提升卡

具体实施方式

[0019] 下面将结合附图,对本申请中的技术方案进行描述。

[0020] 图1示出了一种互联板卡组合,包括:主板,所述主板上包括第一CPU和第二CPU,所述第一CPU与所述主板上的第一高速连接线和第二高速连接器分别通过PCIE接口相连,所述第二CPU与所述主板上的第三高速连接线和第四高速连接器通过PCIE接口相连;GPU底板,所述GPU底板上包括第五高速连接器和第六高速连接器,所述第五高速连接器和所述第六高速连接器用于与所述主板上的第一CPU或第二CPU连接,所述GPU主板包括两个PCIE Switch,所述每个PCIE Switch至少包括4个下行端口,所述两个PCIE Switch用于通过第一提升板和第二提升板分别与8颗GPU相连。

[0021] 图2示出了另一实施例的互联板卡组合,包括图1中示出的GPU主板,两个第一提升板和两个第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

[0022] 具体地,在图2中示出的GPU主板分别通过四个高速连接器分别与两个第一提升板和两个第二提升板相连,其中,GPU中的PCIE Switch可以通过32PCIE line与各个高速连接

器连接。

[0023] 可选地,作为本申请一个实施例,所述第二提升板长度比所述第一提升板长度短13cm。

[0024] 可选地,作为本申请一个实施例,所述PCIE Switch为PEX9797芯片。

[0025] 可选地,作为本申请一个实施例,所述PCIE Switch为包括5个下行端口,其中,4个下行端口用于扩展GPU卡,1个下行端口用于扩展无限带宽IB卡。

[0026] 具体地,首先,每颗CPU通过BIOS可配置为3组PCIE X16。该专利使用两颗CPU(均位于主板),通过高速线缆HS CONN(cable)与GPU底板互联;同一时刻只有一颗CPU的2组PCIE X16root port与GPU底板互联,或者第一CPU,或者第二CPU。

[0027] 具体地,例如GPU底板使用两颗PEX9797芯片用于PCIE资源扩展。PEX9797包含97个Gen 3的PCIE Lane资源、最多可配置6组X16PCIE,用户根据需求建立高性能、低延迟的应用。

[0028] 例如,在具体实施例中,每颗PEX9797共计5个downstream PCIE X16port,其中4个用于扩展GPU卡、1个用于扩展IB卡。考虑空间限制,方案设计选用高密度的高速连接器与提升板Riser板卡的互联,其中,每两组PCIE X16共用一组高速连接器。

[0029] 不同于4U机箱将所有GPU平行于机箱侧壁的机构设计,此专利采用2U标准机架式服务器尺寸。Riser板通过高速连接器与GPU底板互联、并平行于机箱侧壁,而PCIE Slot垂直于Riser板(或机箱侧壁)、平行于机箱底部。为满足空间需求,本专利设计两种提升板Riser板卡:Riser A,也就是上述实施例中的第一提升板,Riser B,也就是上述实施例中的第二提升板。

[0030] 可选地,作为本申请一个实施例,互联板卡结构包括第一提升板和第二提升板,其中,所述第一提升板中的两个PCIE卡槽与所述第一提升板上的高速连接器位于所述第一提升板的两侧,所述第二提升板中的两个PCIE卡槽与所述第二提升板的高速连接器位于所述第二提升板的两侧,所述第一提升板上包括挖空区域,所述第一提升板与所述第二提升板配合使用,所述第一提升板上的挖空区域用于放置所述第二提升板上的PCIE卡槽。

[0031] 如图3所示,示出了本申请一个实施例的第一提升板,图4示出了本申请一个实施例的第二提升板。

[0032] 其中,第一提升板,即图3中Riser A板卡中,高速连接器、PCIE Slot(图中虚线代表位于板卡另外一侧)分别位于板卡的两侧,高速连接器与GPU底板互联、PCIE Slot与GPU互联。Riser A板卡中间挖空,便于Riser B板卡的PCIE Slot从挖空区域穿过。

[0033] 第二提升板,即图4中的Riser B板卡长度短于Riser A板卡约为13cm,高速连接器、PCIE Slot也分别位于板卡两端(图中虚线代表位于板卡另外一侧),高速连接器与GPU底板互联、PCIE Slot与GPU互联。

[0034] Riser A、Riser B板卡通过螺丝固定在托架上并分别位于托架两侧,Riser B的PCIE Slot将穿过Riser A挖空区域。Riser A板卡上的GPU前端朝向机箱后端、Riser B板卡上的GPU前端朝向机箱前端,如此,两板卡上的GPU在二维空间上有重叠、但在三维空间上不存在任何机构干涉,最大限度的利用空间在2U机箱内集成8颗PCIE标准的GPU卡。

[0035] 因此,服务器的实际应用中,最多使用两颗CPU:只有一颗CPU时,两组PCIE X16均来自第一CPU;有两颗CPU时,两组PCIE X16均来自第二CPU,分别与两颗PCIE Switch互联,

而第一CPU用于扩展NVME硬盘或IB万兆网卡。从一定角度提高CPU利用率、同时降低硬件设计成本。

[0036] 两颗PCIE Switch芯片均配置为Base Mode模式、6组PCIE X16port,其upstream port与CPU root port互联、downstream port与GPU互联。为满足高密度集成8颗GPU,采用2U高度机架式通用服务器,硬件上采用GPU底板、两种GPU Riser板搭配解决空间冲突问题。这将极大增加该服务器的应用范围、降低研发成本、提高空间利用率。

[0037] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0038] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0039] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0040] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0041] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0042] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者第二设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0043] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应所述以权利要求的保护范围为准。

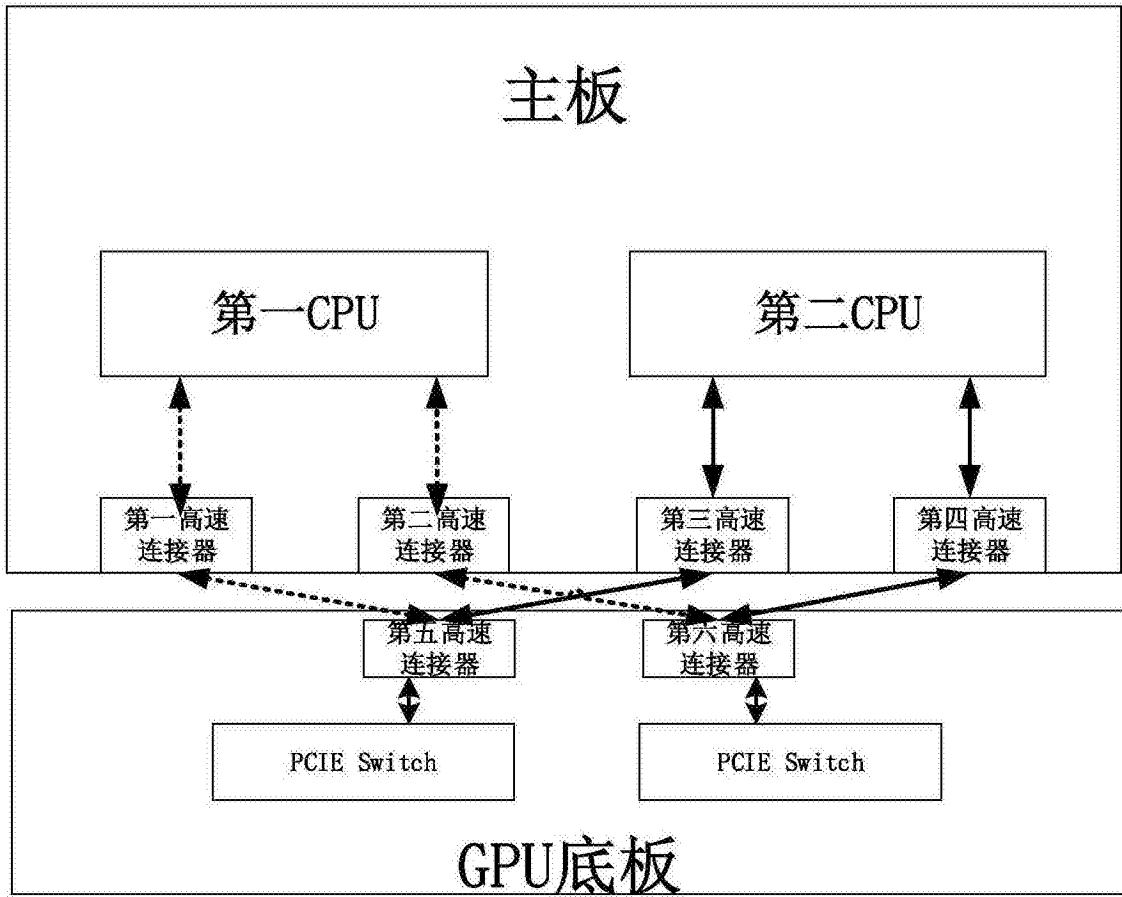


图1

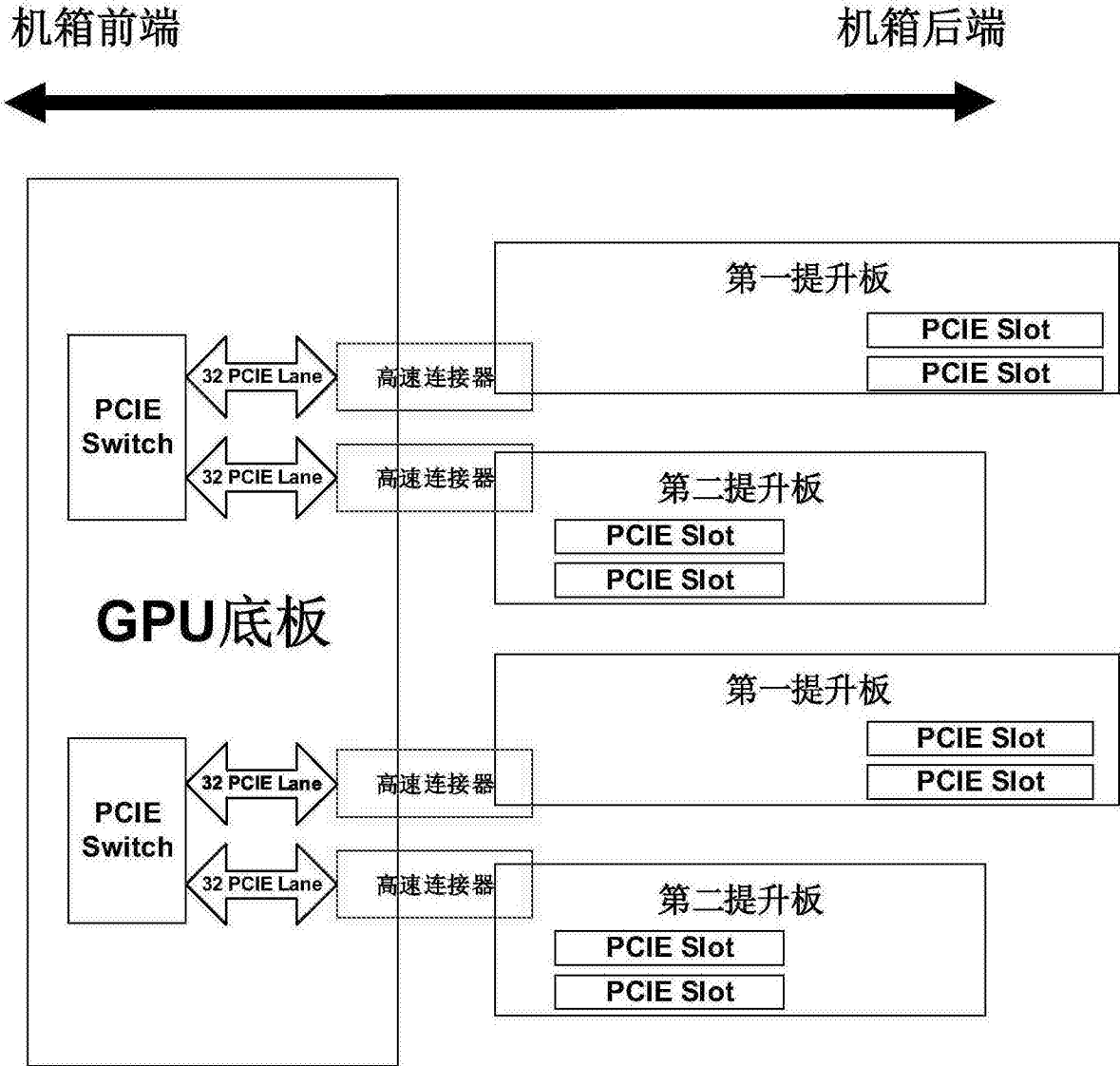


图2

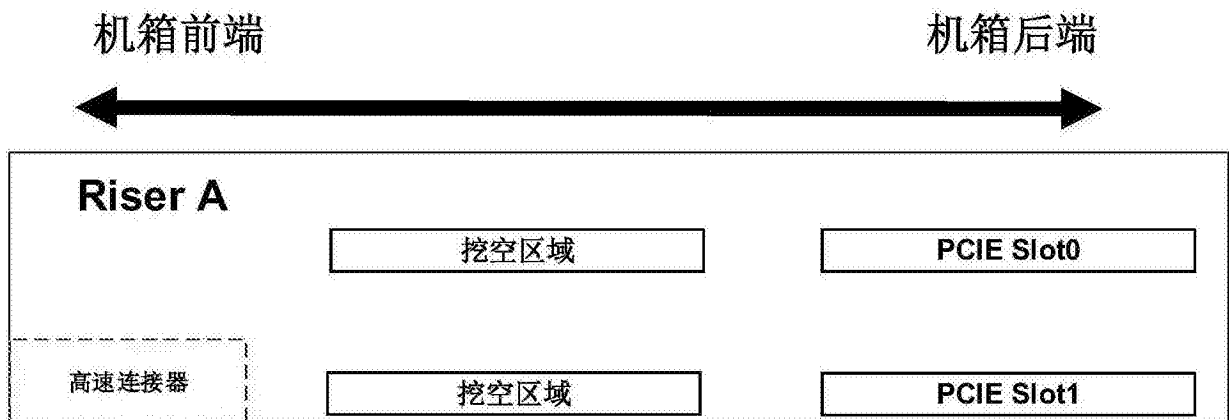


图3

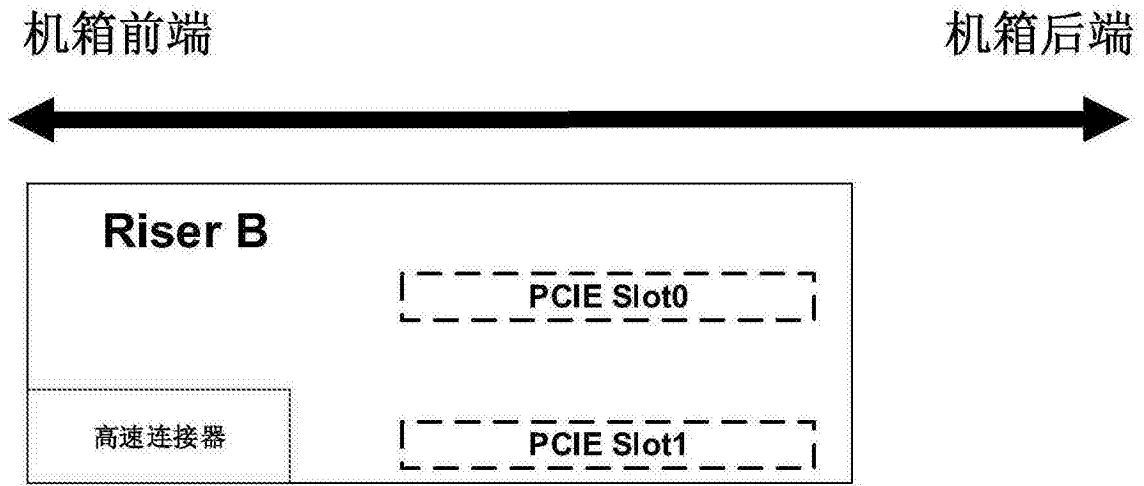


图4