



(12) 发明专利申请

(10) 申请公布号 CN 104932618 A

(43) 申请公布日 2015. 09. 23

(21) 申请号 201510333558. 6

(22) 申请日 2015. 06. 16

(71) 申请人 浪潮电子信息产业股份有限公司  
地址 250100 山东省济南市高新区浪潮路  
1036 号

(72) 发明人 罗嗣恒

(74) 专利代理机构 济南信达专利事务有限公  
司 37100

代理人 李世喆

(51) Int. Cl.

G06F 1/16(2006. 01)

G06F 1/26(2006. 01)

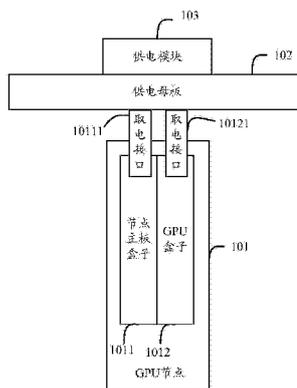
权利要求书1页 说明书5页 附图2页

(54) 发明名称

一种 GPU 服务器设备

(57) 摘要

本发明提供一种服务器设备,包括:至少一个 GPU 节点,供电母板,供电模块,其中,所述供电模块与所述供电母板相连,用于通过所述供电母板供电;每一个 GPU 节点中包括:节点主板盒子和 GPU 盒子,节点主板盒子与 GPU 盒子相连,节点主板盒子和 GPU 盒子上均分别包括取电接口,每一个 GPU 节点通过内部的节点主板盒子上的第一取电接口和 GPU 盒子上的第二取电接口连接所述供电母板。提高业务实现的灵活性。



1. 一种服务器设备,其特征在于,包括:  
至少一个 GPU 节点,供电母板,供电模块,其中,  
所述供电模块与所述供电母板相连,用于通过所述供电母板供电;  
每一个 GPU 节点中包括:节点主板盒子和 GPU 盒子,节点主板盒子与 GPU 盒子相连,节点主板盒子和 GPU 盒子上均分别包括取电接口,每一个 GPU 节点通过内部的节点主板盒子上的第一取电接口和 GPU 盒子上的第二取电接口连接所述供电母板。
2. 根据权利要求 1 所述的服务器设备,其特征在于,  
所述节点主板盒子中包括:第一取电接口、节点主板、节点系统硬盘和第一连接器接口;  
所述 GPU 盒子包括:至少 1 个 GPU 卡、第二取电接口、GPU 供电板、GPU 转接板和第二连接器接口,其中,所述 1 个 GPU 卡连接在所述 GPU 转接板上;  
所述第一连接器接口与所述第二连接器接口扣合。
3. 根据权利要求 2 所述的服务器设备,其特征在于,在所述节点主板盒子中,所述取电接口、节点系统硬盘和第一连接器接口均设置在所述节点主板上;  
和/或,  
在所述 GPU 盒子中,所述第二连接器接口设置在所述 GPU 转接板上,所述第二取电接口设置在所述 GPU 供电板上,所述 GPU 供电板通过供电控制信号线与所述 GPU 转接板相连,所述 GPU 供电板通过供电线与所述至少一个 GPU 卡相连。
4. 根据权利要求 3 所述的服务器设备,其特征在于,包括:在所述节点主板盒子中进一步包括:设置在所述节点主板上的第一热插拔保护单元;  
和/或,  
在所述 GPU 盒子中进一步包括:设置在所述 GPU 供电板上的第二热插拔保护单元。
5. 根据权利要求 1 至 4 中任一所述的服务器设备,其特征在于,  
所述供电模块包括:位于所述供电母板第一侧的 4 个 PSU;  
和/或,  
所述 GPU 节点位于所述供电母板的第二侧。
6. 根据权利要求 1 至 5 中任一所述的服务器设备,其特征在于,进一步包括:  
位于所述第一侧的至少一个风扇。
7. 根据权利要求 1 至 5 中任一所述的服务器设备,其特征在于,所述节点主板盒子上设置有推拉把手;  
所述 GPU 盒子上设置有推拉把手。
8. 根据权利要求 1 至 5 中任一所述的服务器设备,其特征在于,进一步包括:机箱,用于收容所述至少一个 GPU 节点,供电母板,供电模块。
9. 根据权利要求 8 所述的服务器设备,其特征在于,  
所述机箱高 4U,宽 8U;  
所述 GPU 节点的个数为 4。

## 一种 GPU 服务器设备

### 技术领域

[0001] 本发明涉及计算机技术领域,特别涉及一种 GPU 服务器设备。

### 背景技术

[0002] 在现代的计算机中,特别是家用系统和游戏的发烧友,图形的处理变得越来越重要,需要一个专门的图形的核心处理器,因此,图形处理器 (GPU, Graphic Processing Unit) 应运而生。随着 GPU 的发展, GPU 已经不再局限于图形处理了, GPU 通用计算技术发展已经引起业界不少的关注,事实也证明在浮点运算、并行计算等部分计算方面, GPU 可以提供数十倍乃至上百倍于 CPU 的性能。

[0003] 目前, GPU 服务器的结构通常包括: GPU 服务器中包括一块主板,在该主板上设置有 PCIE 插槽,至少一个的 GPU 卡插接在 PCIE 插槽上。

[0004] 可见,在 GPU 服务器中,虽然 GPU 卡是插接在 PCIE 插槽上的,但是 PCIE 插槽和 GPU 卡还是整体布设在主板上,仍然属于整体结构,业务实现的灵活性较差。

### 发明内容

[0005] 本发明提供一种 GPU 服务器设备,提高业务实现的灵活性。

[0006] 一种服务器设备,包括:

[0007] 至少一个 GPU 节点,供电母板,供电模块,其中,

[0008] 所述供电模块与所述供电母板相连,用于通过所述供电母板供电;

[0009] 每一个 GPU 节点中包括:节点主板盒子和 GPU 盒子,节点主板盒子与 GPU 盒子相连,节点主板盒子和 GPU 盒子上均分别包括取电接口,每一个 GPU 节点通过内部的节点主板盒子上的第一取电接口和 GPU 盒子上的第二取电接口连接所述供电母板。

[0010] 所述节点主板盒子中包括:第一取电接口、节点主板、节点系统硬盘和第一连接器接口;

[0011] 所述 GPU 盒子包括:至少 1 个 GPU 卡、第二取电接口、GPU 供电板、GPU 转接板和第二连接器接口,其中,所述 1 个 GPU 卡连接在所述 GPU 转接板上;

[0012] 所述第一连接器接口与所述第二连接器接口扣合。

[0013] 在所述节点主板盒子中,所述第一取电接口、节点系统硬盘和第一连接器接口均设置在所述节点主板上;

[0014] 和/或,

[0015] 在所述 GPU 盒子中,所述第二连接器接口设置在所述 GPU 转接板上,所述第二取电接口设置在所述 GPU 供电板上,所述 GPU 供电板通过供电控制信号线与所述 GPU 转接板相连,所述 GPU 供电板通过供电线与所述至少一个 GPU 卡相连。

[0016] 包括:在所述节点主板盒子中进一步包括:设置在所述节点主板上的第一热插拔保护单元;

[0017] 和/或,

- [0018] 在所述 GPU 盒子中进一步包括：设置在所述 GPU 供电板上的第二热插拔保护单元。
- [0019] 所述供电模块包括：位于所述供电母板第一侧的 4 个 PSU；
- [0020] 和 / 或，
- [0021] 所述 GPU 节点位于所述供电母板的第二侧。
- [0022] 进一步包括：
- [0023] 位于所述第一侧的至少一个风扇。
- [0024] 所述节点主板盒子上设置有推拉把手；
- [0025] 所述 GPU 盒子上设置有推拉把手。
- [0026] 进一步包括：机箱，用于收容所述至少一个 GPU 节点，供电母板，供电模块。
- [0027] 所述机箱高 4U，宽 8U；
- [0028] 所述 GPU 节点的个数为 4。
- [0029] 本发明实施例提供了一种服务器设备，以结构化的方式利用节点主板盒子和 GPU 盒子组成一个 GPU 节点，其中，节点主板盒子包括：第一取电接口、节点主板、节点系统硬盘和第一连接器接口；GPU 盒子包括：至少 1 个 GPU 卡、第二取电接口、GPU 供电板、GPU 转接板和第二连接器接口，其中，所述 1 个 GPU 卡连接在所述 GPU 转接板上；所述第一连接器接口与所述第二连接器接口扣合。这样，单个 GPU 节点既包括了主板又包括了 GPU 卡，则可以具有完成的数据处理功能，而不是服务器设备内部所有的 GPU 卡整体布设在一个主板上，而每一个 GPU 节点内部都具有节点主板，因此此种结构化的方式，增加了业务实现的灵活性。

#### 附图说明

- [0030] 图 1 是本发明一个实施例中服务器设备的结构示意图。
- [0031] 图 2 是本发明一个实施例中节点主板盒子内部的结构示意图。
- [0032] 图 3 是本发明一个实施例中 GPU 盒子内部的结构示意图。
- [0033] 图 4 是本发明一个实施例中服务器设备的组成示意图。

#### 具体实施方式

[0034] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

- [0035] 本发明一个实施例提出了一种服务器设备，参见图 1，包括：
- [0036] 至少一个 GPU 节点 101，供电母板 102，供电模块 103，其中，
- [0037] 所述供电模块 103 与所述供电母板 102 相连，用于通过所述供电母板 102 供电；
- [0038] 每一个 GPU 节点 101 中包括：节点主板盒子 1011 和 GPU 盒子 1012，节点主板盒子 1011 与 GPU 盒子 1012 相连，节点主板盒子 1011 和 GPU 盒子 1012 上均分别包括取电接口，每一个 GPU 节点 101 通过内部的节点主板盒子 1011 上的第一取电接口 10111 和 GPU 盒子 1012 上的第二取电接口 10121 连接所述供电母板 102。
- [0039] 参见图 1，在服务器设备内部，包括至少一个 GPU 节点，而单个 GPU 节点既包括了主板又包括了 GPU 卡，单个 GPU 节点内部的 GPU 卡只插接在内部的主板上，就可以具有完整的

数据处理功能,而不是服务器设备内部所有的 GPU 卡整体布设在一个主板上,因此此种结构化的方式,增加了业务实现的灵活性。比如,单个 GPU 节点内部的主板故障时,只需更换该 GPU 节点内部的主板,而不会影响服务器设备内部其他 GPU 节点的正常工作。再如,各个 GPU 节点内部的主板和 GPU 卡可以根据需要配置为完成不同的数据处理功能等。

[0040] 在本发明一个实施例中,参见图 2,节点主板盒子中可以包括:取电接口 201、节点主板 202、节点系统硬盘 203 和连接器接口 204。

[0041] 在本发明一个实施例中,参见图 3, GPU 盒子可以包括:取电接口 301、2 个 GPU 卡 302、GPU 供电板 303、连接器接口 304 和 GPU 转接板 305,其中,所有 GPU 卡 302 连接在所述 GPU 转接板 305 上;

[0042] 所述连接器接口 204 与所述连接器接口 304 扣合,从而实现 GPU 盒子与节点主板盒子的连接。

[0043] 可见,一个 GPU 节点是由两个功能盒子组成,节点主板盒子可以完成主板的相关功能, GPU 盒子可以完成 GPU 卡的相关功能,从而大大提高了数据处理性能。

[0044] 参见图 2,在本发明一个实施例中,在所述节点主板盒子中,取电接口 201、节点系统硬盘 203 和连接器接口 204 均设置在所述节点主板 202 上。

[0045] 参见图 3,在本发明一个实施例中,在所述 GPU 盒子中,连接器接口 304 设置在所述 GPU 转接板 305 上,所述取电接口 301 设置在所述 GPU 供电板 303 上,所述 GPU 供电板 303 通过供电控制信号线与所述 GPU 转接板 305 相连,所述 GPU 供电板 303 通过供电线与所述至少一个 GPU 卡 302 相连。

[0046] 图 3 中的 GPU 供电控制信号线的作用:保证系统工作时,某个 GPU 节点没工作时,将该节点上的 GPU 供电链路断开,防止主板没开机, GPU 盒子开始供电的情况发生。

[0047] 参见图 2,在所述节点主板盒子中进一步包括:设置在所述节点主板 202 上的第一热插拔保护单元 205。通过第一热插拔保护单元 205 可以对节点主板盒子通过其取电接口在供电母板上的热插拔进行保护。

[0048] 参见图 3,在所述 GPU 盒子中进一步包括:设置在所述 GPU 供电板 303 上的第二热插拔保护单元 306。通过第二热插拔保护单元 306 可以对 GPU 盒子通过其取电接口在供电母板上的热插拔进行保护。

[0049] 需要说明的是,图 2 只是一种可能的节点主板盒子的结构,比如包括上述的各个功能模块。在实际的业务实现中,节点主板盒子也可以只包括图 2 中的部分功能模块,或者,包括其他未示出的功能模块。

[0050] 需要说明的是,图 3 只是一种可能的 GPU 盒子的结构,比如包括上述的各个功能模块。在实际的业务实现中, GPU 盒子也可以只包括图 3 中的部分功能模块,或者,包括其他未示出的功能模块。

[0051] 参见图 4,在本发明一个实施例中,所述供电模块包括:位于所述供电母板第一侧的 4 个 PSU;

[0052] GPU 节点位于所述供电母板的第二侧。

[0053] 参见图 4,在本发明一个实施例中,在图 1 所示基础上,服务器设备进一步包括:位于所述第一侧的至少一个风扇,比如图中所示 3 个风扇。

[0054] 参见图 4,服务器设备可以包括 4 个 GPU 节点。

[0055] 在本发明一个实施例中,为了方便装配,所述节点主板盒子上设置有推拉把手;所述 GPU 盒子上设置有推拉把手。

[0056] 在本发明一个实施例中,服务器设备中进一步包括:机箱,用于收容所述至少一个 GPU 节点,供电母板,供电模块。

[0057] 在本发明一个实施例中,所述机箱高 4U,宽 8U;所述 GPU 节点的个数为 4。

[0058] 在本发明一个实施例中,装配服务器设备的过程可以包括:

[0059] 1)、根据节点主板满配的电流值,为节点主板盒子上,选择合适的热拔插线路及取电端口连接器的规格;

[0060] 2)、根据 GPU 的信号定义、GPU 供电控制信号及盒子的结构尺寸来确定主板和 GPU 转接板上高密连接器的规格(如:引脚数量、结构尺寸);

[0061] 3)、依据 GPU 的额定功率,为 GPU 盒子上的 GPU 供电转接板选择合适的热拔插线路及取电端口连接器的规格;

[0062] 4)、根据 GPU 的额定功率,确定 GPU 转接线的规格;

[0063] 5)、将 GPU 转接板、GPU 供电转接板固定在 GPU 盒子托盘上,GPU0\GPU1 插接在 GPU 转接板上,连接好供电线及信号线,完成 GPU 盒子的组装;

[0064] 6)、然后,将节点主板盒子与 GPU 盒子通过高密连接器扣合在一起,并用螺丝将两个盒子锁在一起。完成单个 GPU 服务器节点的搭建;

[0065] 7)、最后,按照 1-6 步骤,一次搭建好剩下的 3 个 GPU 服务器节点。将 GPU 服务器节点依次插入机箱内。

[0066] 这样,1 台 4U 的 GPU 服务器系统即可搭建完成。从而,可在高 4U、宽 8U 的机箱内,部署 8 颗 CPU、8 个 GPU(节点主板为 2 路服务器主板),很大程度的提高了服务器系统单位空间的数据处理能力。

[0067] 本发明各实施例至少具有如下的有益效果:

[0068] 1、以结构化的方式利用节点主板盒子和 GPU 盒子组成一个 GPU 节点,其中,节点主板盒子包括:第一取电接口、节点主板、节点系统硬盘和第一连接器接口;GPU 盒子包括:至少 1 个 GPU 卡、第二取电接口、GPU 供电板、GPU 转接板和第二连接器接口,其中,所述 1 个 GPU 卡连接在所述 GPU 转接板上;所述第一连接器接口与所述第二连接器接口扣合。这样,单个 GPU 节点既包括了主板又包括了 GPU 卡,则可以具有完整的数据处理功能,而不是服务器设备内部所有的 GPU 卡整体布设在一个主板上,而每一个 GPU 节点内部都具有节点主板,因此此种结构化的方式,增加了业务实现的灵活性。

[0069] 2、在服务器设备内部,包括至少一个 GPU 节点,而单个 GPU 节点既包括了主板又包括了 GPU 卡,单个 GPU 节点内部的 GPU 卡只插接在内部的主板上,就可以具有完整的数据处理功能,而不是服务器设备内部所有的 GPU 卡整体布设在一个主板上,因此此种结构化的方式,增加了业务实现的灵活性。比如,单个 GPU 节点内部的主板故障时,只需更换该 GPU 节点内部的主板,而不会影响服务器设备内部其他 GPU 节点的正常工作。再如,各个 GPU 节点内部的主板和 GPU 卡可以根据需要配置为完成不同的数据处理功能等。

[0070] 3、所有 GPU 节点中的 GPU 盒子和节点主板盒子都通过自身的取电接口连接到服务器设备内部的供电母板,也就是说,进行的是集中供电,因此,提高了系统整体供电效率,降低整体功耗;并且,可以方便地实现 GPU 节点的热插拔。

[0071] 4、服务器设备中可以包含多个 GPU 节点,并以 GPU 节点作为基本单元部署在同一个机箱框体内,具备超强的数据处理能力。

[0072] 需要说明的是,在本文中,诸如第一和第二之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同因素。

[0073] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

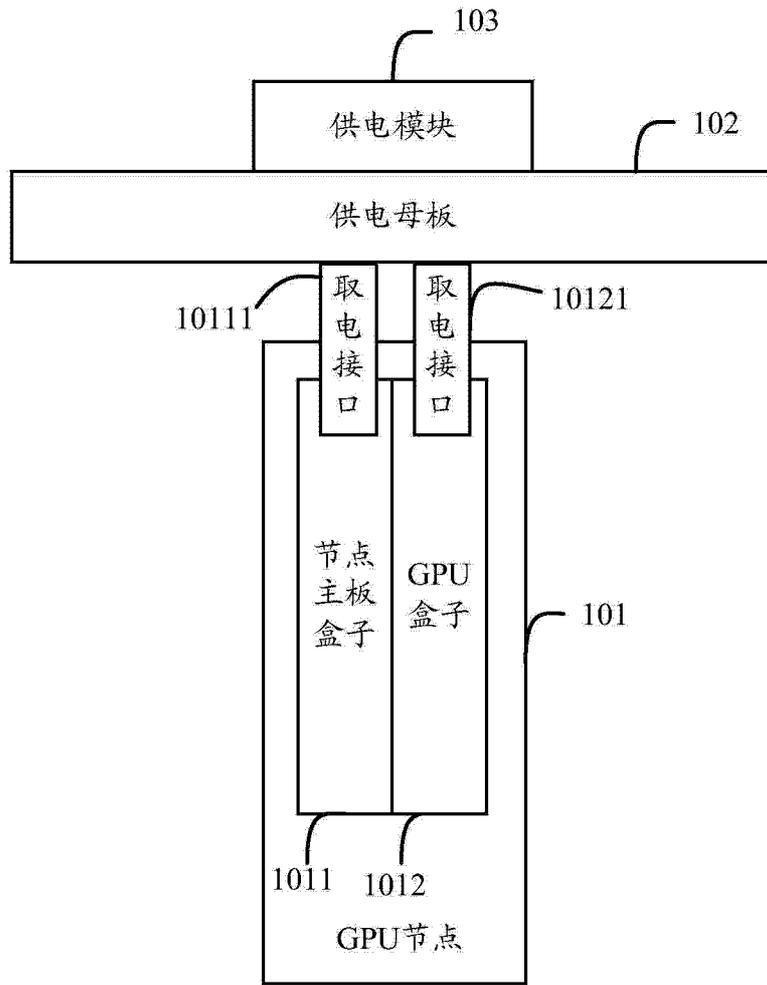


图 1

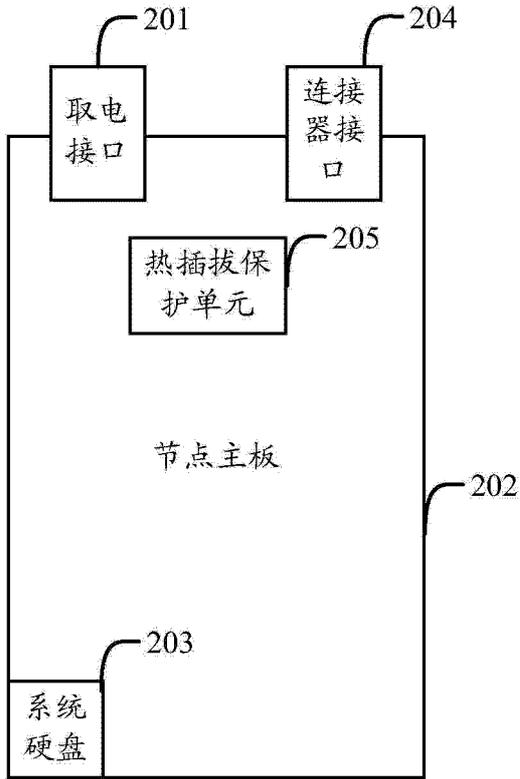


图 2

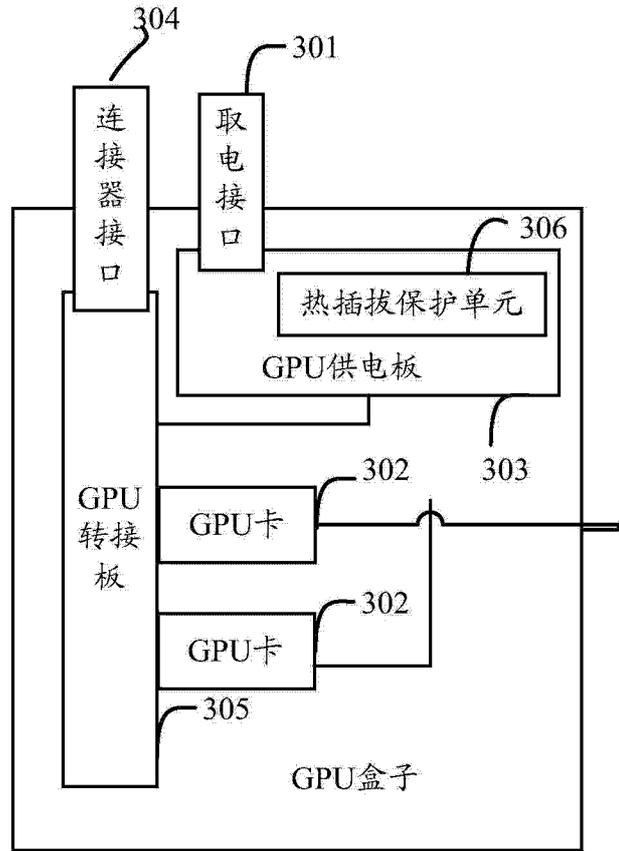


图 3

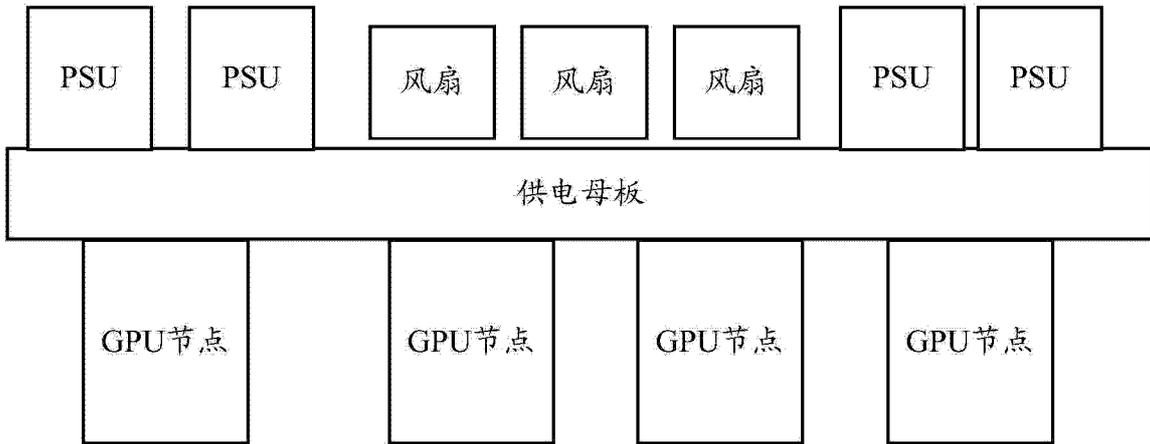


图 4