



(12)发明专利

(10)授权公告号 CN 103020009 B

(45)授权公告日 2017.06.09

(21)申请号 201210392549.0

(22)申请日 2012.08.23

(65)同一申请的已公布的文献号
申请公布号 CN 103020009 A

(43)申请公布日 2013.04.03

(30)优先权数据
1157471 2011.08.23 FR

(73)专利权人 卡雷公司
地址 法国奥尔赛

(72)发明人 M·哈兰德

(74)专利代理机构 北京市柳沈律师事务所
11105
代理人 吕晓章

(51)Int.Cl.

G06F 15/78(2006.01)

G06F 13/40(2006.01)

(56)对比文件

CN 1542604 A,2004.11.03,全文.

WO 8903564 A1,1989.04.20,全文.

US 7598958 B1,2009.10.06,全文.

CN 102014050 A,2011.04.13,全文.

EP 2320331 A1,2011.05.11,全文.

审查员 张滔

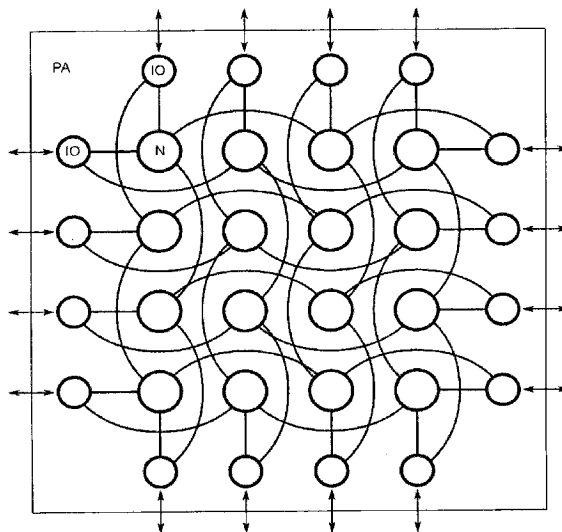
权利要求书2页 说明书5页 附图3页

(54)发明名称

可扩展的片上网络

(57)摘要

一种集成电路包括排列成阵列的计算节点;与计算节点互连的环面拓扑片上网络;以及在阵列的每一行或列的每个末端处并被插入到两个计算节点之间的网络链接的网络扩展单元。该扩展单元具有建立两个相应计算节点间的网络链接的连续性的普通模式,和将网络链接分成可从集成电路外部访问的两个独立片段的扩展模式。



1. 一种集成电路,包括:
排列成阵列的计算节点;
通过并行总线链接使得计算节点互连的环面拓扑片上网络;
在所述阵列的每一行或列的每个末端处插入到两个计算节点之间的总线中的网络扩展单元,所述网络扩展单元具有建立所述两个计算节点之间的总线的连续性的普通模式和将所述总线分成两个独立的总线片段的扩展模式;
一组并行/串行转换器,每个形成流出串行信道以在所述集成电路的第一外部端子上串行传输在总线片段上并行存在的数据;
一组串行/并行转换器,每个形成流入串行信道以在总线片段上并行传输所述集成电路的第二外部端子上串行到达的数据;以及
所述阵列同一边缘的所述网络扩展单元共有的负载均衡器,被配置为在进行向外传输的总线片段之间分配可用的流出串行信道。
2. 如权利要求1所述的集成电路,其中在普通模式下,行或列末端的计算节点连接到位于所述行或列的末端处的计算节点之间的链接上的输入/输出接口。
3. 如权利要求1所述的集成电路,其中所述负载均衡器被配置为在每个流出串行传送的首标中插入来源总线片段的标识符。
4. 如权利要求3所述的集成电路,其中所述负载均衡器被配置为解析每个流入的串行传送首标,并将相应的串行信道切换到所述首标中所识别的总线片段。
5. 如权利要求1所述的集成电路,其中所述串行信道通过分组传送数据,并且包括用于分组等待传送的队列,所述负载均衡器被配置为将分组路由到具有最不满的队列的串行信道。
6. 一种集成电路,包括:
排列成阵列的计算节点;
使得所述计算节点互连的环面拓扑片上网络;以及
在所述阵列的每一行或列的每个末端处插入到两个计算节点之间的网络链接中的网络扩展单元,所述网络扩展单元具有:
建立所述两个计算节点之间的所述网络链接的连续性的普通模式,和
将所述网络链接分成可从所述集成电路外部访问的两个独立片段的扩展模式。
7. 如权利要求6所述的集成电路,其中所述网络链接包括并行总线并且所述网络扩展单元包括,用于片段的:
并行/串行转换器,形成流出串行信道,用于在所述集成电路的第一外部端子上串行传送在所述片段上并行存在的数据;以及
串行/并行转换器,形成流入串行信道,用于在所述片段上并行传输所述集成电路的第二外部端子上串行到达的数据。
8. 如权利要求6所述的集成电路,包括位于行或列的末端处的计算节点之间的链接中的输入/输出接口,并被配置为通过输入/输出端子与所述集成电路的外部进行通信,其中在扩展模式中所述网络扩展单元被配置为将所述输入/输出端子连接到所述片段。
9. 如权利要求7所述的集成电路,包括所述阵列同一边缘的所述网络扩展单元共有的负载均衡器,其被配置为在进行流出传送的所述片段之间分配可用的流出串行信道。

10. 如权利要求9所述的集成电路,其中所述负载均衡器被配置为在每个流出串行传送的首标中插入来源片段的标识。

11. 如权利要求10所述的集成电路,其中所述负载均衡器被配置为解析每个流入的串行传送的首标,并将相应的串行信道切换到所述首标中识别的所述片段。

12. 如权利要求9所述的集成电路,其中所述串行信道在分组中传送数据,并且包括用于存储分组等待传送的队列,所述负载均衡器被配置为将分组路由到具有最不满的队列的串行信道。

可扩展的片上网络

技术领域

[0001] 本发明涉及一种集成处理器阵列,其中处理器通过片上网络(NoCs)进行互连。本发明更特别地涉及一种处理器阵列结构,其具有开发工具可以在程序设计员的最小的帮助下适应阵列的处理器数量的这种规律。

背景技术

[0002] 图1示意地示出一种在折叠环面拓扑的片上网络中包括排列成 4×4 的计算节点N的处理器阵列PA,如在美国专利申请第20110058569号中描述的。在阵列拓扑中,每个节点都通过点对点的双向链接连接到在同一行中的两个其它的节点和在同一列中的两个其它的节点。在环面拓扑中,阵列的节点在每一行和每一列中还以环形方式进行连接,以至于包括位于阵列边缘的节点的所有节点就它们的互连而言都具有相同的物理结构。在图1中描述的折叠拓扑中,每个节点(除非其置于阵列的边缘)都连接到行和列中相同距离的两个其它的节点,因此节点间的链接基本上具有相同的长度。

[0003] 每个节点N包括五路路由器,其管理与行和列中与下一节点的四个链接,其称作北、南、东、西链接,以及与处理单元的链接,例如通过共享总线互连的一群处理器。

[0004] 处理器阵列PA被制作为单独的集成电路。为了与外界通信,其包括插入到阵列边缘处的片上网络中的输入/输出IO单元。正如所示,这种IO单元可以被设置在每行或每列的两个末端处。更特别地,每个单元都插入到连接同一行或同一列的两个极端节点N的链接中。

[0005] 每个IO单元都具有管理两个与节点N的链接和一个与输入/输出接口的链接的三路路由器。输入/输出接口允许通过集成电路的金属垫与电路外部进行通信,旨在放入与印刷电路板或其它基板的传导轨道的接触。

[0006] 为了使对这样的处理器阵列的编程更容易,所有的计算节点N都具有相似的特性,允许开发工具在任何节点上以自动模式映射任务(map tasks)。为了达到这一目的,IO单元被设计为对片上网络的内部通信是透明的。美国专利申请第20110058569号也描述了一种对于内部通信而言通过IO单元的路由器减少等待时间的解决方案。

[0007] 为了在销售集成电路中标准化的目的,将在相对窄的范围内提供处理器阵列的尺寸。因此,由该范围内的最大阵列提供的计算能力对于更苛刻的应用来说可能是不够的。

发明内容

[0008] 因此有必要提供比在范围内最大处理器阵列更多的可用的计算能力。一个必然的需要是在不改变已有的用于处理器阵列的开发工具的情况下增加计算能力。

[0009] 这些需要通过集成电路解决,该集成电路包括:排列成阵列的计算节点;使得计算节点互连的环面拓扑片上网络;以及在阵列的每一行或列的每一末端上插入到两个计算节点之间的网络链接的网络扩展单元。扩展单元具有建立两个相应计算节点之间的网络链接连续性的普通模式,以及将网络链接分成可从集成电路外部访问的两个独立片段的扩展模

式。

[0010] 根据实施例,网络链接包括并行总线,扩展单元包括用于片段的并行/串行转换器,该并行/串行转换器形成流出串行信道,用于在片段上并行存在的电路数据的第一外部端子上串行传送;和串行/并行转换器,该串行/并行转换器形成流入串行信道,用于在集成电路的第二外部端子上在片段数据串行到达后并行传送。

[0011] 根据实施例,集成电路包括位于行或列末端处的计算节点之间的链接中的输入/输出接口,并被配置为通过输入/输出端子与集成电路外部进行通信,其中扩展单元被配置为在扩展模式中,将所述输入/输出端子连接到所述片段。

[0012] 根据实施例,集成电路包括阵列同一边缘的扩展单元共有的负载均衡器,其被配置为在流出传送进行中的片段之间分配可用的流出串行信道。

[0013] 根据实施例,负载均衡器被配置为在每个流出串行传送的首标中插入来源片段的标识。

[0014] 根据实施例,负载均衡器被配置成解析每个流入串行传送的首标,并将相应的串行信道切换到首标中识别的片段。

[0015] 根据实施例,串行信道在分组中传送数据,并且包括用于存储分组等待传送的分组的队列,负载均衡器被配置为路由分组给具有最不满的队列的串行信道分组。

附图说明

[0016] 下面通过描述本发明的特定实施例,本发明其它的优点和特征将更加明显,这些特定实施例仅为了示范目的并展示在附图中。

[0017] 图1是先前已描述的表示通过折叠环面拓扑的片上网络互连的处理器阵列。

[0018] 图2图示了由多个处理器阵列形成的巨型阵列(macro-array)。

[0019] 图3图示了巨型阵列的两个相邻阵列之间的理想互连,在保留拓扑的同时能够扩展网络。

[0020] 图4示出了网络扩展单元的一个实施例。

[0021] 图5示出了网络扩展单元的另一个实施例。

具体实施方式

[0022] 图2示出了一种用于当由单个处理器阵列以标准集成电路形式所提供的计算能力不够时增加可利用的计算能力的可能的解决方案。正如所示,将若干处理器阵列PA1、PA2.....在基板上,如印刷电路板,组装成足够尺寸的巨型阵列以达到所需的计算能力。

[0023] 每个PA阵列都可以独立地进行编程和使用,但这涉及程序设计员在计算能力方面将任务分割成独立且平衡的子任务的努力的部分。操作系统也需要在阵列外部运行以在阵列之间分配子任务,而阵列通常被设计为运行它自己的操作系统,从而是自主的。

[0024] 为了避免这种复杂性,从开发工具的观点来看,将巨型阵列考虑为一个且唯一一个处理器阵列是理想的。为达到这个目的,优选的是所有PA阵列的计算节点在一起形成一个且唯一一个网络。

[0025] 这个的一个可能的解决方案是通过其输入/输出接口将PA阵列相互连接,同时模仿两个相邻阵列接口之间的双向网络连接。然而,这样的模仿涉及额外的取决于形成巨型

阵列的阵列尺寸和数量的软件复杂性。

[0026] 另外,这个解决方案可能要求输入/输出接口完全相同,还要求所有行和列的末端都适合这样的接口。在实践中,标准处理器阵列可能具有有限数量的输入/输出接口,并且这些接口可能是不相同的。

[0027] 图3示出了在折叠环面拓扑阵列的情景下的相同拓扑的两个相邻阵列PA1和PA2之间理想的连接类型,这种连接类型允许两个阵列的片上网络形成相同拓扑的单个网络。所示出的例子与通过阵列的行的网络扩展对应——应当注意的是同样的原理也适用于列。

[0028] 在阵列PA1的每一行中,最后两个节点N与它们的输入/输出单元IO的连接是断开的(如果在这个位置上没有IO单元,那么断开的是最后两个节点之间的连接)。相似地,在阵列PA2与之相应的行中,最开始两个节点N与它们的输入/输出单元IO的连接是断开的(如果在这个位置上没有IO单元,那么断开的是最开始两个节点之间的连接)。由此断开的内部连接(虚线所示)被外部链接Le1和Le2所代替,从而确保阵列PA1的行与阵列PA2相应的行的接合,同时形成与内部行的拓扑相同的扩展行。为达到这个目的,链接Le1将阵列PA1的行的倒数第二个节点连接到阵列PA2的行的第一个节点上,并且链接Le2将阵列PA1的行的最后一个节点连接到阵列PA2的行的第二个节点上。

[0029] 在实际的实施中,每个由此被外部链接“替代”的内部链接都被分成被使得能独立地从外部访问的两个片段。因此,在某些情况下,跨过输入/输出单元IO的、行的两个最末端节点之间的内部链接被分成分别通过外部链接Le1和Le2与相邻电路的相应片段连接的两个片段。

[0030] 注意到折叠环面拓扑尤其适合这种扩展。当然,阵列的每一行中被外部链接影响的两个节点恰巧是离边缘最近的节点。

[0031] 还注意到阵列PA1和PA2外面边缘的IO单元将不再使用。这与建立具有与单独阵列相同的拓扑的巨型阵列的希望是一致的,其中IO单元处于外围。

[0032] 因此,在扩展的列和行具有与单独的PA电路的行和列相同的折叠环面拓扑的配置中,将行和列扩展为跨过若干相邻的PA电路是有可能的。

[0033] 可以使用与传统的PA阵列的相同的开发工具对由此形成的巨型阵列编程。当然,考虑到传统阵列的规律性和节点N的可互换性,开发工具仅需要被配置为具有该阵列的尺寸以便以自动的方式将任务映射到不同的节点上,并通过片上网络建立节点之间的通信图。在巨型阵列具有完全的传统阵列拓扑的情况下,现有的开发工具仅需要在计算节点方面被配置为具有巨型阵列的新尺寸。

[0034] 图4示出了用于在两个相邻阵列PA1和PA2的两行之间建立外部链接Le1和Le2的结构的具体实施例。通常节点N之间的内部连接是具有许多传导线的总线。使用具有同样多的线的外部链接Le1和Le2扩展这些总线实际上是不可行的,这是因为包括这些阵列的集成电路在大多数情况下不能布置足够的外部接触端子。为了避免这种复杂性,每个外部连接Le1和Le2都被提供为快速串行连接的形式。事实上,如图所示,由于内部链接是双向的,因此每个外部链接Le1和Le2都包括两个相反方向的串行链接。那么每个外部链接Le1,Le2在每个集成电路PA上仅需要两个接触端子40。这些端子可以从不使用的输入/输出接口IO得到,如对于链接Le2所示。

[0035] 通过足够多地放置端子40,即,使得在两个相邻电路PA之间相互连接的端子面对

面,电路可以被放置得离彼此很近以缩短用于电路间串行链接的传导轨道。通过这样地缩短轨道(达到毫米级),并且由于串行接口不需要遵循标准,因此串行信号可以达到非常高的10Gb/s级的传送速率。

[0036] 阵列PA的行和列的每个末端都被配备具有扩展单元42。单元42包括用于每个外部链接Le1、Le2的串行/并行/串行转换器(SERDES)——它将内部并行数据转换为流出串行链接上的串行流,并将串行流入数据转换为并行的内部数据流。并行流通过分别与外部链接Le1,Le2相关联的开关S1、S2。开关S1和S2通过网络扩展信号EXT控制。

[0037] 当信号EXT无效时,单元42处于正常模式。在阵列PA的常规独立配置中,开关S1和S2将一对节点N的末端连接到它们的输入/输出单元I0。在没有单元I0处,存在开关S1和S2之间的直接连接。

[0038] 当信号EXT有效时,单元42处于“网络扩展”模式。开关S1和S2将这对节点连接到它们各自的SERDES转换器上,将电路PA放置在图3的配置中。

[0039] 信号EXT优选地被电路PA的相同边缘的所有的扩展单元42共有。这样,基于巨型阵列中的电路PA的位置,四个信号EXT被提供给每一个电路PA以独立地控制在电路每一边缘上的扩展单元42。信号EXT的状态存储在,例如可编程配置寄存器中。

[0040] 尽管可以在两个相邻PA电路之间实现快速串行连接,但是在某些情况下,它们不会达到内部并行连接的流量速率。然后,扩展的网络在两个PA电路之间的前沿处可能有带宽限制,从而导致通过使用巨型阵列获得的性能可能与PA电路的数量不成正比。

[0041] 图5示出了用于在两个PA电路之间的前沿处增加平均带宽的实施例。在此图中,单元42被展示为“网络扩展”模式——为了清楚,未示出包括单元I0在内的普通模式中使用的元件。该实施例旨在优化外部连接的使用,并且它基于有用带宽通常在链接之间特别是在流出链接之间不均匀分配的假设。流出串行信道的带宽在有效流出传送之间动态分配。对于每行(或列),在电路的相同边缘处都存在两个流出信道,分别地与外部串行链接Le1和Le2相关联。如果PA电路具有M个行(或列),那么在电路的一个边缘上存在2M个流出串行信道。

[0042] 阵列边缘的所有扩展单元42的开关S1和S2都被负载均衡器LB代替,所述负载均衡器LB负责取决于流出串行信道的可用性将流出并行流切换到一个或更多个SERDES转换器上。

[0043] 在图5的例子中,通过链接Le2退出第一行的传送并行地借用链接Le1的可用流出信道。通过例如分组达到负载平衡:来自电路PA1的顶端右节点的一些分组采用链接Le1而其它的采用链接Le2。

[0044] 这个附图同时也示出了通过第四行的链接Le2退出的传送并行地借用第二和第三行的链接Le2的流出信道。

[0045] 串行传送通常被分组化。每个串行信道都具有待传送的分组被堆栈于其中的传送队列。确定能够被分配以达到负载平衡的串行信道可以通过使用例如信道的队列填充等级获得:外传分组一到达负载均衡器就将会被按路线传送给最不满的队列。

[0046] 刚刚描述了由传送PA电路(PA1)执行的一部分负载平衡功能。功能剩余的部分由接收电路(PA2)的负载均衡器LB执行。传送电路的负载均衡器,即分配流出串行信道的电路(PA1),识别正在进行的传送和它们的原始内部连接。接收电路(PA2)的负载均衡器检索识

别信息并使流入串行信道重新定向到识别的内部链接。

[0047] 识别信息可以被插入到包含在串行传送中的首标中,与标准的串行传送协议一致,例如因特拉肯协议(Interlaken protocol)。

[0048] 如果电路PA2有数据要传送给电路PA1,那么传送是通过反转描述的电路PA1和PA2的角色来达到的。一个方向上的传送并且其它的借用分开的串行信道,借此两种传送就能同时独立地执行。

[0049] 正如刚刚说明的,通过使用动态工作的负载均衡器LB,提供比内部连接更少的双向串行信道是可能的。在某些应用中,例如,为两个或四个内部连接提供一个双向串行信道可能就足够了。这减少了电路外部端口的数量,尤其SERDES转换器占有的表面面积。负载均衡器将以与上面描述的相同的方式工作;它将只有更少的串行信道库来分配。

[0050] 在本说明书中,在获得处理器阵列的无限制可扩展性的同时保持与为单个电路设计的现有开发工具兼容的情景下,展示了可从外部扩展的片上网络的实施例。这些开发工具仅需要使用扩展阵列的尺寸进行配置。

[0051] 不排除开发工具可能发展到考虑电路之间外部连接的特点。在这种情况下,并非使用负载均衡器动态地路由流出分组,取而代之的是,在编程时,使用位于分组首标中的路由选择信息静态地分配串行信道分组。负载均衡器由基于首标中的信息指示包裹到串行信道的路由器替代。

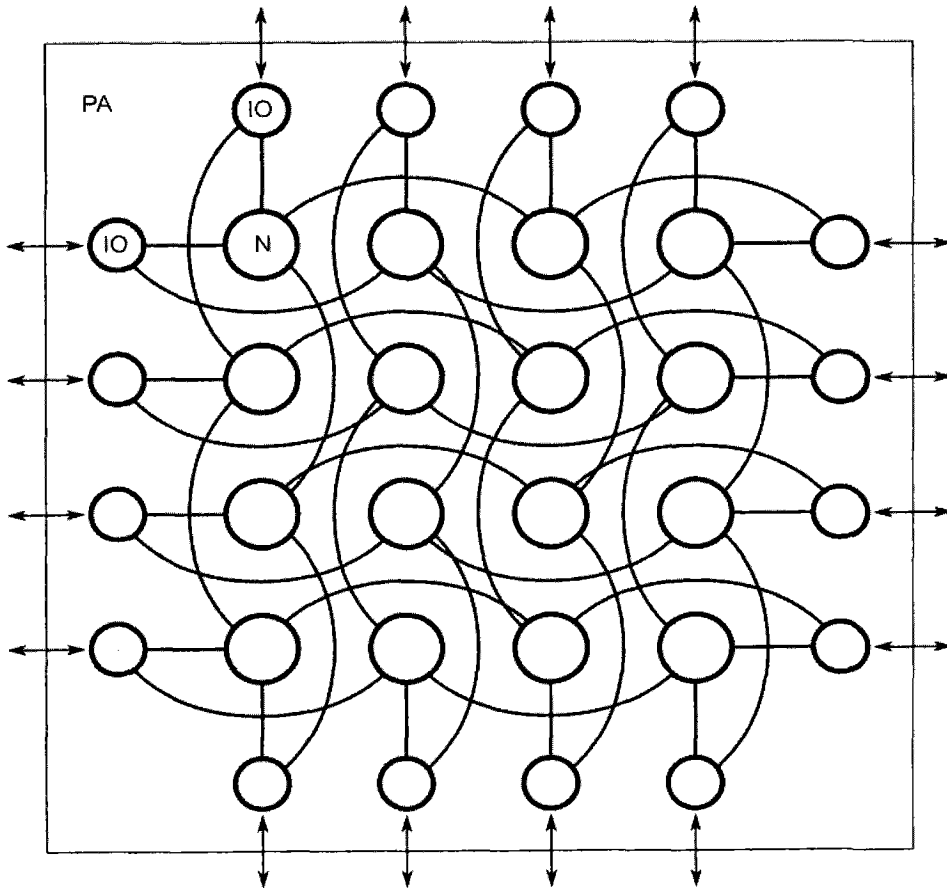


图1

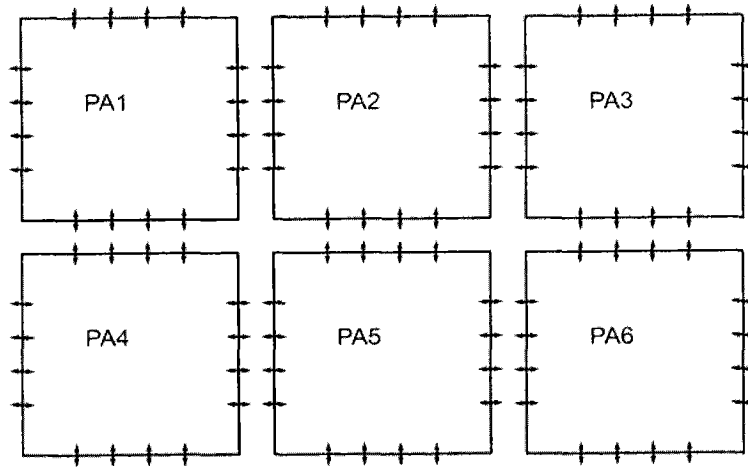


图2

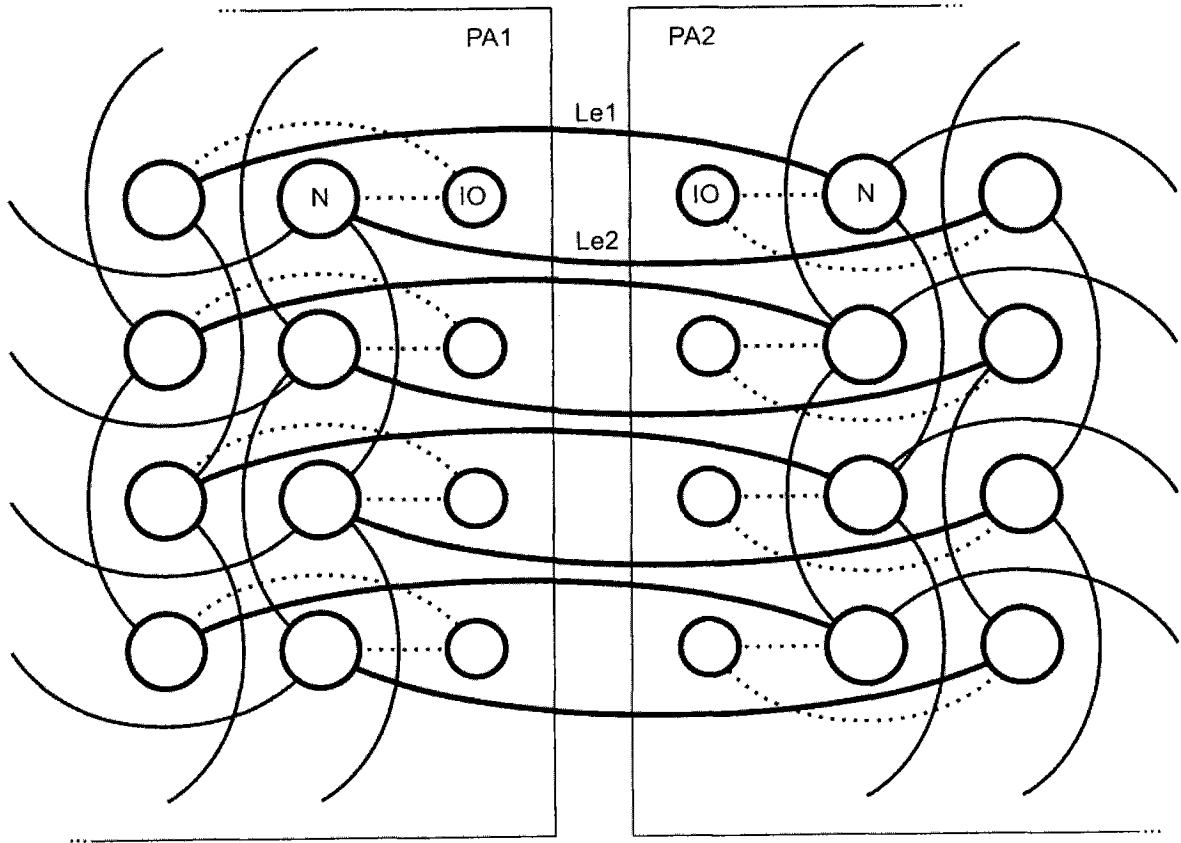


图3

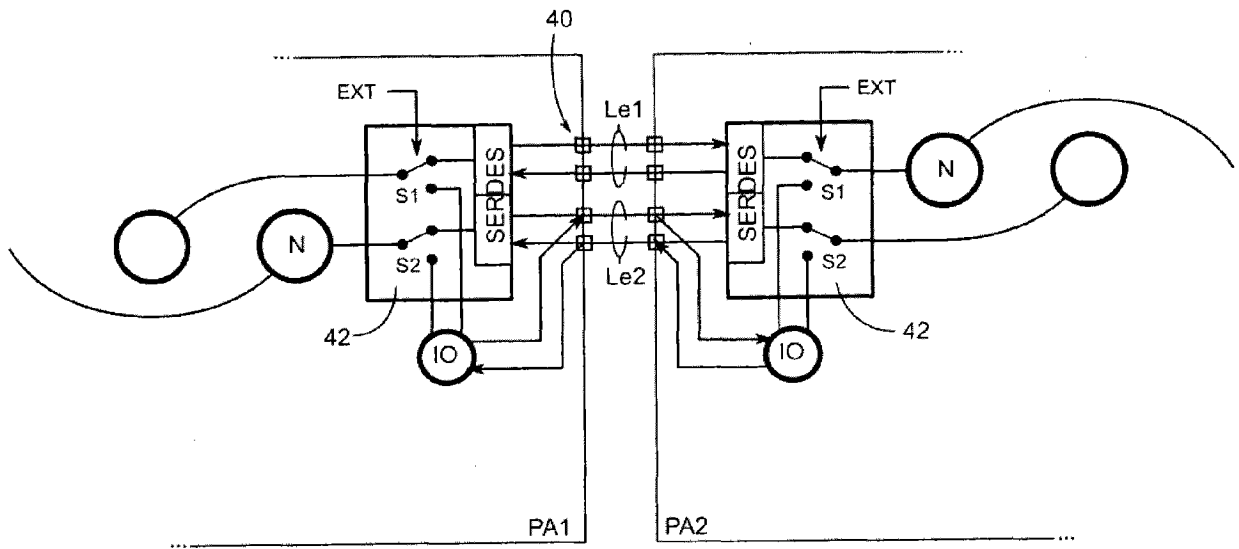


图4

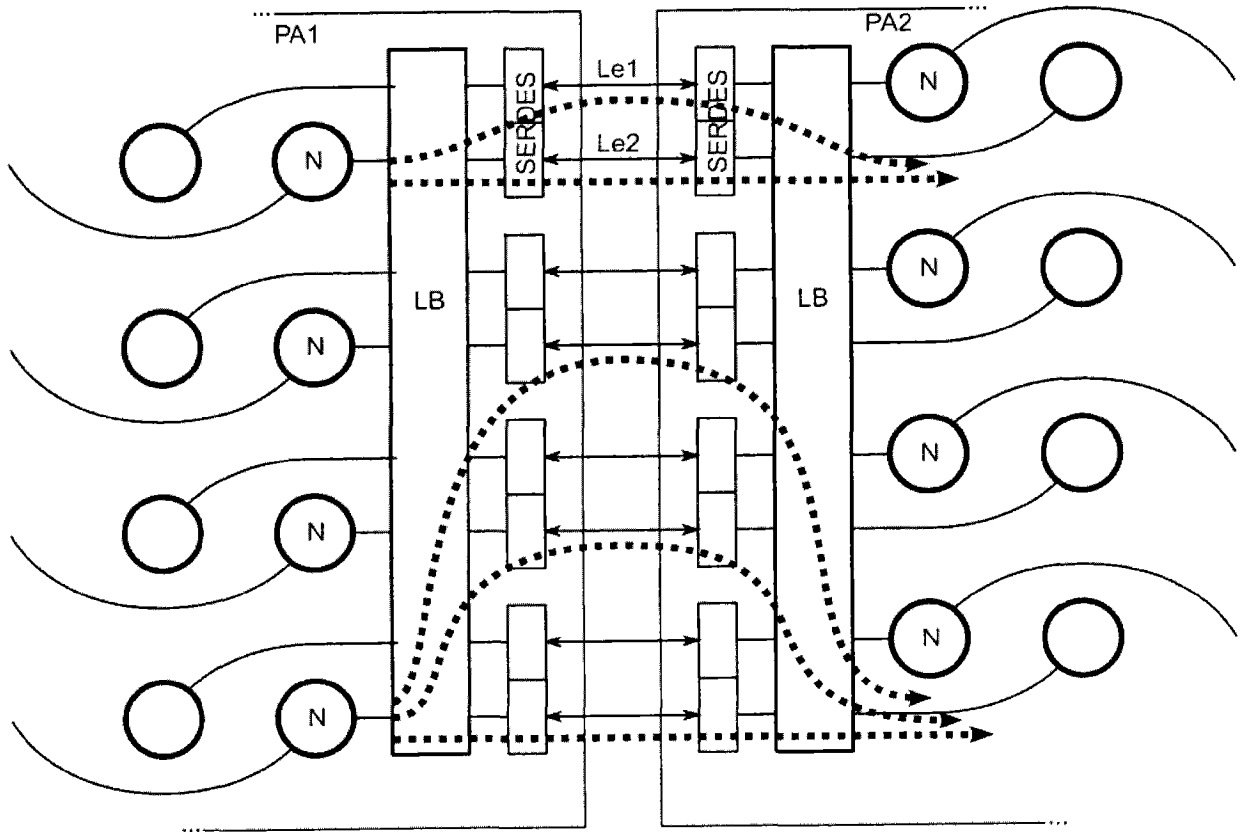


图5