

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2012年8月16日 (16.08.2012)



(10) 国际公布号
WO 2012/106929 A1

- (51) 国际专利分类号:
G06F 1/12 (2006.01)
- (21) 国际申请号: PCT/CN2011/077625
- (22) 国际申请日: 2011年7月26日 (26.07.2011)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人 (对除美国外的所有指定国): **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): **吴登奔 (WU, Dengben)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **张羽 (ZHANG, Yu)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **俞柏峰 (YU, Baifeng)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: **北京同立钧成知识产权代理有限公司 (LEADER PATENT & TRADEMARK FIRM);** 中国

北京市海淀区西直门北大街 32 号枫蓝国际 A 座 8F-6, Beijing 100082 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

[见续页]

(54) Title: COMPUTER SYSTEM AND CLOCK CONFIGURATION METHOD THEREOF

(54) 发明名称: 计算机系统及其配置时钟的方法

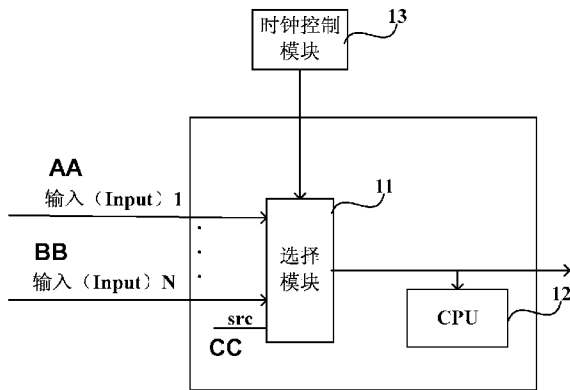
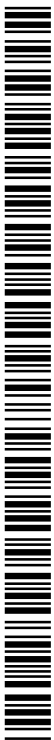


图 1 / Fig.1

- 11 SELECTION MODULE
- 12 CPU
- 13 CLOCK CONTROL MODULE
- AA INPUT 1
- BB INPUT N
- CC SRC

(57) Abstract: The present invention relates to a computer system and a clock configuration method; the computer system comprises two or more nodes, the two or more nodes all comprise a selection module and a CPU, the input of the selection module is a clock of the node and a clock outputted by another node, and the output end is connected to the CPU and a selection module of another node; the computer system further comprises a clock control module, the output end of the clock control module connects to the control end of the selection module to control the clocks of the two or more nodes to be the same clock. When the clocks of a plurality of nodes malfunction, as long as the computer system still has a normal clock, the computer system is still capable of normal operations.

[见续页]



WO 2012/106929 A1



-
- 在修改权利要求的期限届满之前进行，在收到该修改后将重新公布(细则 48.2(h))。 — 根据申请人的请求，在条约第 21 条(2)(a)所规定的期限届满之前进行。

(57) 摘要:

本发明涉及计算机系统及配置时钟方法，计算机系统包括至少两个节点，其中，所述至少两个节点均包括选择模块及 CPU，所述选择模块的输入为本节点的时钟及其他节点输出的时钟，输出端与所述 CPU 及其他节点的选择模块的输入端相连；所述计算机系统还包括时钟控制模块，所述时钟控制模块的输出端与所述选择模块的控制端相连，用于控制所述至少两个节点的时钟为同一个时钟。当多个节点的时钟异常时，只要该计算机系统还存在一个正常的时钟，该计算机系统仍能正常工作。

计算机系统及其配置时钟的方法

技术领域

5 本发明涉及计算机系统技术，尤其涉及一种计算机系统及用于实现计算机系统同一分区系统的节点时钟同步的时钟配置方法。

背景技术

通常，计算机系统，如对计算以及容错性能要求较高的小型机，由多个
10 不同的节点构成，其中，节点为硬分区的最小单元，每个节点可以单独组成一个分区，也可以与其他节点互联组成一个分区，每个分区上都可以运行独立的操作系统具体地，计算机系统中几个独立的节点通过 CPU 互联技术形成一个整体即分区系统，在该整体下可以运行一个操作系统，该操作系统可以访问任何该分区所属节点上的有效设备。

15 一个分区系统中的每个节点不仅需要时钟，而且每个节点需要相同来源的时钟。这是因为时钟在数字电路中非常重要，一旦没有了时钟，数字逻辑完全无法正常工作，而分区系统中的各个节点若不使用相同来源的时钟，则整个分区也无法正常工作。

现有技术中，分区系统采用单一式时钟同步方案实现分区系统内的时钟
20 同步。即一个分区系统内只有一个时钟，作为该分区系统内各节点的时钟源，该时钟放在该分区系统的某一节点上，或者与各节点分开，单独设置。

现有技术存在的缺陷在于：一个分区系统只有一个时钟可作为时钟源，当该时钟异常时，该分区系统无时钟源可用，无法实现该分区系统的各节点的时钟同步。

25

发明内容

本发明实施例的目的在于提出一种计算机系统及用于实现计算机系统同一分区系统的节点时钟同步的时钟配置方法，以在分区系统内的时钟源故障时，仍然能够实现计算机系统中同一分区系统的各节点时钟同步。

本发明实施例提供了一种计算机系统，包括至少两个节点，其中，所述至少两个节点均包括选择模块及 CPU，所述选择模块的输入为本节点的时钟及其他节点输出的时钟，输出端与所述 CPU 及其他节点的选择模块的输入端相连；

所述计算机系统还包括时钟控制模块，所述时钟控制模块的输出端与所述选择模块的控制端相连，用于控制所述至少两个节点的时钟为同一个时钟。

本发明实施例提供的计算机系统中，每个节点的选择模块的输入端有本节点的时钟源，也有其他节点的时钟输出，通过第三方如上述的时钟控制模块来控制选择模块选择一个时钟输入作为该节点的时钟输入，并且可以为其他的节点提供时钟，以保证每个节点使用同一个时钟源，当多个节点的时钟异常时，只要该计算机系统还存在一个正常的时钟，该计算机系统仍能正常工作。

本发明实施例还提供了一种用于实现上述计算机系统中同一分区系统的节点时钟同步的时钟配置方法，包括：

选择一节点的时钟作为所属分区系统的时钟源的选择步骤；

开通所述节点作为起点，与所述节点之间相连的节点作为终点的连接通道作为所述节点及与之直接相连的节点之间的时钟通道的开通步骤；

判断作为终点的节点是否与所述分区系统中剩余的节点相连，若是，则执行所述开通步骤；否则，完成所述分区系统的时钟配置。

本发明实施例提供的用于实现计算机系统同一分区系统的节点时钟同步的时钟配置方法中，利用计算机系统的同一分区系统中每个节点都有多个时钟可选的结构特点，通过为同一分区系统的相连两个节点选择一条连接通道保证同一分区系统内的各节点的时钟源统一，从而实现同一分区系统内的各

节点的时钟同步，并且在时钟通道异常，可选择其他连接通道，仍然保证了同一分区系统内的各节点的时钟源统一，实现了同一分区系统内的各节点的时钟同步。

5 附图说明

图 1 为本发明实施例提供的计算机系统中节点的结构示意图；

图 2 为本发明实施例提供的计算机系统中两个节点的时钟连接示意图；

图 3A 为本发明实施例提供的计算机系统中四个节点的时钟连接示意图；

图 3B 为图 3A 的框架示意图；

10 图 4A 为本发明实施例提供的计算机系统中 8 个节点的时钟连接框架示意图；

图 4B 为本发明实施例提供的计算机系统中 16 个节点的时钟连接框架示意图；

15 图 5 为本发明实施例提供的计算机系统中 32 个节点的时钟连接框架示意图；

图 6 为本发明实施例提供的 8 节点计算机系统的结构示意图；

图 7 为本发明实施例提供的用于实现上述计算机系统中同一分区系统的节点时钟同步的时钟配置方法流程图；

图 8A 为图 7 配置完时钟后的时钟路径图；

20 图 8B 为图 7 配置过程中未被配置的剩余节点示意图；

图 8C 为图 6 所示计算机系统中节点 0 与节点 1 之间时钟通道异常后的替换时钟路径图；

图 8D 为图 8C 形成过程中未被配置的剩余节点示意图；

图 9 为图 6 所示计算机系统中 8 个节点的另外一种时钟连接示意图。

25

具体实施方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明作进一步地详细描述。

本发明实施例提供的计算机系统，包括至少两个节点，其中，如图 1 所示，所述至少两个节点均包括选择模块 11 及 CPU 12，所述选择模块 11 的输入 Input 1、...、Input N 为本节点的时钟 src 及其他节点输出的时钟，输出端 (output) 与所述 CPU 12 及其他节点的选择模块的输入端相连；

所述计算机系统还包括时钟控制模块 13，所述时钟控制模块 13 的输出端与所述选择模块的控制端相连，用于控制所述至少两个节点的选择模块输出端输出的时钟为同一个时钟。

上述技术方案中，每个节点的选择模块的输入端有本节点的时钟源，也有其他节点的时钟输出，通过第三方如上述的时钟控制模块来控制选择模块选择一个时钟输入作为该节点的时钟输入，并且可以为其他的节点提供时钟，以保证每个节点使用同一个时钟源，当多个节点的时钟异常时，只要该计算机系统还存在一个正常的时钟，该计算机系统仍能正常工作。

当计算机系统中有两个节点时，两节点直接相连。如图 2 所示，节点 21 中选择模块 211 的输出端，与节点 22 中选择模块 221 的输入端直接连接，节点 22 中选择模块 221 的输出端与节点 21 中选择模块 211 的输入端相连。并且，节点 21 及节点 22 中，选择模块的另一个输入为本节点的时钟 src。节点 21 中选择模块 211 的输出端与本节点的 CPU 212 相连，节点 22 中选择模块 221 的输出端与本节点的 CPU 222 相连。节点 21 与节点 22 为双向连接，使用时可关闭一条连接，当开启的连接异常时，启用关闭的连接。

选择模块 211 及选择模块 221 的控制端均由时钟控制模块 23 控制输出。换句话说，每个节点中选择模块的输入为本节点的时钟与其他节点的时钟，在时钟控制模块 23 的控制下，将一个节点的时钟 src 作为共同的时钟源，从而保证每个节点的时钟同步。

当计算机系统中有四个节点时，连接方式与图 2 类似，所述四个节点通

过选择模块连接为环形，每个节点的选择模块的输入为本节点的时钟及相连的两个节点的选择模块的输出。如图 3A、图 3B 所示，计算机系统中共有节点 31、节点 32、节点 33 及节点 34 共四个节点，四个节点的选择模块依次相连为一个环形，如图 3B 所示。具体地，节点 31 中，选择模块 311 的输出端
5 与节点 32 中选择模块 321 及节点 34 中选择模块 341 的输入端相连，并与本节点的 CPU 312 相连；节点 32 中，选择模块 321 的输出端与节点 31 中选择模块 311 及节点 33 中选择模块 331 的输入端相连，并与本节点的 CPU 332 相连；节点 33 中，选择模块 331 的输出端与节点 32 中选择模块 321 及节点 34 中选择模块 341 的输入端相连，并与本节点的 CPU 322 相连；节点 34 中，
10 选择模块 341 的输出端与节点 31 中选择模块 311 及节点 33 中选择模块 331 的输入端相连，并与本节点的 CPU 342 相连。且所有选择模块的控制端均与时钟控制模块 35 相连，在时钟控制模块 35 的控制下，所有节点选择同一个节点的时钟源 src 作为共同的时钟源。同样，图 3A 中，任意两个节点之间的连接均为双向连接，即相连的两个节点之间有两个连接通道也即时钟通道，
15 使用时需关闭未启用的连接通道，。当使用的时钟通路异常时，则根据时钟通道正常情况重新选择时钟通道。

当计算机系统中有 8 个节点时，8 个节点位于一个长方体的各个顶点处，所述长方体各个表面内的 4 个节点通过选择模块交叉连接或通过选择模块两两相邻连接，且每个节点与三个节点通过选择模块连接。如图 4A 所示，长方体的各个表面的 4 个节点通过选择模块顺次连接为一个环形，其中，环形的连接与图 3A 类似，每个节点均与三个节点通过选择模块相连，任意两个节点之间的连接均为双向连接即相连的两个节点之间有两条连接通道也即时钟通道，使用时可关闭其中一条连接通道，使用另一条连接通道。当使用的连接通道异常时，启用关闭的连接通道。得到的长方体连接方式与图 3A 的
20 不同之处在于，每个节点与三个节点相连，具体地，每个节点的选择模块的输入为本节点的时钟及相连的三个节点的选择模块的输出。

当计算机系统中有 $n \times 8$ 个节点时，每 8 个节点位于一个长方体的各个顶点处，所述长方体各个表面内的 4 个节点通过选择模块交叉连接或通过选择模块顺次连接，且每个节点与 $3+n$ 个节点通过选择模块连接，长方体的连接与图 4A 类似，每个所述长方体相同位置上的节点通过选择模块相连，每个节点的选择模块的输入为本节点的时钟与相连的 $3+n$ 个节点的选择模块的输出，其中， n 为自然数。同样，任意两个节点之间的连接均为双向连接即相连的两个节点之间有两条连接通道也即时钟通道，使用时可关闭其中一条连接通道，使用另一条连接通道。当使用的连接异常时，启用关闭的连接通道。

例如，当计算机系统中有 16 个节点时，连接后得到如图 4B 所示的结构，当计算机系统中有 32 个节点时，连接后得到如图 5 所示的结构。64 节点、128 节点等等计算机系统的节点连接类似。8 个节点位于一个长方体的各个顶点处，每个节点与相邻的 3 个节点相联。当系统中有 8 的整数倍个节点时，以上述方式连接的各个 8 个节点整体又可以分别看成一个点，再利用上述连接方式将各个点连接。如，16 个节点位于两个长方体的各个顶点处，两个长方体的同一个位置点再相联，以此类推，实现 32 个节点、64 个节点的联接，这里不一一列举。

对上述实施例提供的计算机系统分区时，为了保证分区系统容错率达到最高，应尽可能选择连接最多的几个节点组成一个分区，如将直接相连的两个节点创建为一个分区系统，或将连接为一个环形的 4 个节点创建为一个分区系统。如对于 $n \times 8$ 节点计算机系统还可将连接为一个长方体的 8 个节点创建为一个分区系统。这是因为时钟同步仅仅是对同一分区系统有要求，将连接最多的几个节点组成一个分区系统，可以最大程度上保证在该分区系统内某些节点异常的情况下，找到替换路径，且不影响其他分区系统。

并且，当节点时钟存在异常，创建分区系统时应保证分区内的节点时钟至少有一个正常，否则该分区系统只能使用其他分区系统的时钟。计算机系统中创建分区后，所述计算机系统中属于同一个分区系统的节点相互之间直

接连接或间接连接，所述间接连接的中间节点与所述间接连接的两端节点位于同一个分区系统内。例如对于 $n \times 8$ 节点计算机系统中，连接为一个长方体的 8 个节点创建为一个分区系统。

上述实施例提供的计算机系统进行分区后，某一分区系统内若时钟源异常，只要该分区系统内存在正常时钟，则该分区系统的运行不受影响。若该分区系统的某一时钟通道异常，该时钟通道的下游节点可以正常工作，则只要存在其他正常通道，该分区系统仍然可以正常工作。若该分区系统内某一节点功能异常，剔除后，只要剩余节点间的时钟可以路由互联，则该分区系统还可以正常工作。这样，若该分区系统内节点异常个数过多，造成剩余节点间时钟通道无法路由互联，则剩余节点无法组建一个分区，但可以根据分区情况重新组建几个独立的分区。

下面以 8 节点小机型为例做进一步详细说明。

如图 6 所示，节点 0、节点 1、...、节点 7 通过选择模块连接，表现为一个长方体，或者说，节点 0~节点 7 位于立方体的各个顶点处，立方体的前后表面及下表面上，4 节点各顺次连接为一个环形，上表面的 4 个节点交叉相连，作为备份时钟同步路由，如：前后表面顺次联接，上表面交叉联接，这是因为当出现多个节点错误时，直接相联的几个节点同时出现的概率会相对较大，因此时钟连接中采用一些交叉联接的方式，总体上每个节点的相联个数都不会变。通过图 6 可以直观的说明某个节点时钟异常时，如何找到相应的时钟通道，从而保证时钟链正常。

根据图 6 所示的连接关系，生成如下表 1 所示的时钟关系路由表。

表 1 时钟关系连接表

起点	终点
0	1、2、6

1	0、3、7
2	0、3、5
3	1、2、4
4	3、5、6
5	2、4、7
6	0、4、7
7	1、5、6

当任意有限个时钟及时钟通道异常时，都可以从尝试从时钟关系路由表
中找到其他替代的时钟通道，用找到的替代时钟通道替代异常的时钟通道。
另外时钟选择时，不仅需要选择是否使用该节点的时钟，还需要选择启用的
时钟通路，保证所有节点的时钟来源唯一。

- 5 由于不是任意两个节点之间都存在时钟连接，部分时钟通道需要通过其
他节点路由形成通道，为了减少出错概率及提高设置速度，在创建分区时，
按照上述分区方式进行分区。

假设创建 2P 分区系统，则选择直接联接的 2 个节点组成一个分区系统。
一般情况下，分区系统中各节点时钟都由分区内的节点提供，以免分区系统
10 之间相互影响。若某分区系统内的所有节点时钟都异常，此时可以由相联分
区系统的一个节点为该分区系统的所有节点提供时钟，但该分区系统时钟是
否正常工作，受控于另一分区。这种情况下，可以将这两节点分开，进行重
新分区，与其他相联的节点分别组成两个独立的 2P 分区。

假设创建 4P 分区系统，正常情况下，任意选择上下或前后相互连接构成
15 环形的 4 个节点组建一个分区系统，此时容错概率最高。但若创建分区时，
多个节点时钟已存在异常，则应该根据上述分区原则合理创建分区，保证所

有分区系统可以正常工作。

假设创建 6P 分区系统，则按照创建 4P 分区的方式先选择 4 个节点，然后再任意选择相邻的可以组成 2P 的两个节点，即直接相连的两个节点一起创建为 6P 分区系统。

- 5 假设创建 8P 分区系统，则该计算机系统中的所有节点可以组成一个 8P 分区系统。

下面以 8P 分区系统为例对分区系统的时钟配置进行说明。

如图 7 所示，配置流程包括：

- 步骤 701、根据时钟联接方式创建时钟关系连接表，如表 1 所示，表 1 的
- 10 的每一行都即一层，每层均有起点和终点。起点与终点之间的连接为双向连接通道，既有起点到终点的连接通道，也有终点到起点的连接通道，如在表 1 的第 2 行节点 0 为起点，节点 1 为终点，第 3 行节点 1 为起点，节点 0 为
- 15 终点。后续的步骤就是对直接或间接的多次连接的节点简化为单向且不重复地连接，且对直接相连的节点之间的双向连接通道进行选择，选择一条连接通道作为直接相连的节点之间的时钟通道，另一条连接通道作为备份，以在
- 时钟异常、通道异常或节点异常的情况下重新对分区系统进行时钟配置。

步骤 702、检测各节点时钟是否正常；

- 步骤 703、选择并开启时钟源。具体地，按照主节点时钟→本分区节点时
- 20 钟→其他分区节点时钟的顺序，从主节点时钟、本分区节点时钟、其他分区节点时钟中选择一个正常的节点时钟作为时钟源，并开启该时钟。

- 一般选择主节点的时钟为时钟源，若没有异常，则选择主节点为时钟，若出现异常，则选择同一分区与主节点直接相联的节点时钟作为时钟源。假设图 6 中，若创建 8P 分区，节点 0 为主节点，但节点 0 的时钟异常，则选择与节点 0 相连的节点即节点 1、节点 2 或节点 6 的时钟的任意一个正常时钟
- 25 作为时钟源。

步骤 704、将时钟源所在节点添加到时钟关系路由表中，设置该节点为当

前层起点。时钟关系路由表与表 1 类似，表项包含起点和终点，不同之处在于时钟关系路由表开始时，起点列和终点列均为空，通过执行后续步骤逐渐添加。以图 6 所示计算机系统为例，若节点 0 的时钟为时钟源，则将节点 0 添加到时钟关系路由表中第一层的起点项。

5 步骤 705、判断该分区系统中所有节点是否都已添加到时钟关系路由表中。若所有节点已添加到时钟关系路由表中，完成时钟配置，否则，执行步骤 706。

 步骤 706、判断该分区系统中与当前层起点直接相联的节点是否都已在时钟关系路由表中。若已在时钟关系路由表中，则执行步骤 709，否则，执行
10 步骤 707。

 步骤 707、从与当前层起点直接相连且时钟通道正常，并未被添加到时钟关系路由表的节点中，选择一个与当前层起点直接相连的节点；

 步骤 708、将步骤 707 中选择的节点作为当前层的终点，添加到时钟关系路由表的当前层的终点项中，开启当前层节点到步骤 707 中选择的节点方向
15 上的时钟通道，并将选择的节点作为下一层起点添加到时钟关系路由表的新一行表项中。然后，再次执行步骤 706。

 步骤 709、判断是否还存在与当前层起点同级的节点如与时钟源之间均为 N 跳的节点，若存在与当前层起点同级的节点未加入到时钟关系路由表，且该节点作为上一层终点时与该上一层起点之间的时钟通道正常，则执行步骤
20 712，否则，执行步骤 710。

 步骤 710、判断是否存在下一层起点，若存在下一层起点，则执行步骤 711，否则完成时钟配置。

 步骤 711、将存在的第一个下一层起点设置为当前层起点，继续执行步骤 705。

25 步骤 712、选择该节点为当前层起点，继续执行步骤 705。

 以图 6 所示计算机系统为例，假设计算机系统中节点 0 为主节点，且节

点 0 的时钟正常。

则选择节点 0 的时钟为时钟源，将节点 0 添加到时钟关系路由表中。如表 2 所示。

表 2 时钟关系路由表

起点	终点
节点 0	

5

系统中未被配置的剩余节点如图 8B 中的第一个竖框所示。

然后判断与节点 0 直接相连的节点 1、节点 2 及节点 6 之间的连接通道是否正常，若节点 0 与节点 1、节点 2 及节点 6 之间的连接通道均正常，则开通或选择节点 0 到节点 1、节点 2 及节点 6 方向上的连接通道作为节点 0 与节点 1、节点 2 及节点 6 之间的时钟通道，并将节点 1、节点 2 及节点 6 作为第一层的终点添加到时钟关系路由表中。

10

表 3

起点	终点
节点 0	节点 1、节点 2、节点 6

系统中未被配置的剩余节点如图 8B 中的第二个竖框所示。

进一步地，将节点 1、节点 2 及节点 6 作为下一层起点，判断与节点 1 直接连接的节点 3、节点 7 之间的连接通道是否正常，若节点 1 与节点 3、节点 7 之间的连接通道均正常，则开通或选择节点 1 到节点 3、节点 7 方向上的连接通道作为节点 1 与节点 3、节点 7 之间的时钟通道，并将节点 1 及节点 3、节点 7 作为一层即以节点 1 为一层的起点，节点 3、节点 7 为该层的终点，添加到时钟关系路由表中；如表 4 所示。

20

表 4

起点	终点

节点 0	节点 1、节点 2、节点 6
节点 1	节点 3、节点 7

判断剩余的与节点 2 直接连接的节点 5 之间的连接通道是否正常，若节点 2 与节点 5 之间的连接通道正常，则将节点 2 到节点 5 方向上的连接通道作为节点 2 与节点 5 之间的时钟通道，并将节 2 及节点 5 作为一层即以节点 2 为一层的起点，节点 5 为一层的终点，添加到时钟关系路由表中；如表 5 所示。

表 5

起点	终点
节点 0	节点 1、节点 2、节点 6
节点 1	节点 3、节点 7
节点 2	节点 5

判断剩余的于节点 6 直接连接的节点 4 之间的连接通道是否正常，若节点 6 与节点 4 之间的连接通道正常，则将节点 6 到节点 4 方向上的连接通道作为节点 6 与节点 4 之间的时钟通道，并将节点 6 及节点 4 作为一层即以节点 6 为一层的起点，节点 4 为该层的终点，添加到时钟关系路由表中，如表 6 所示。

表 6

起点	终点
节点 0	节点 1、节点 2、节点 6
节点 1	节点 3、节点 7
节点 2	节点 5
节点 6	节点 4

相应的时钟路由如图 8A 所示。

当配置的节点 0 与节点 1 之间的时钟通道发生异常，则节点 1、节点 3 及节点 7 将无法得到时钟源提供的时钟，此时，可根据时钟关系连接表表 1，获知节点 3 与节点 2 之间还备份有连接通道，而节点 1 与节点 3 直接相连，因此可开通或选择节点 2 到节点 3 方向上的连接通道作为节点 2 与节点 3 之间的时钟通道，并关闭节点 1 到节点 3 方向上的连接通道，开通节点 3 到节点 1 上的连接通道作为节点 3 与节点 1 之间的时钟通道，如图 8C 所示，从而节点 1、节点 3 仍可得到时钟源提供的时钟。配置过程中未被配置的剩余节点如图 8D 所示。并且，根据时钟连接表表 1 还可获知节点 7 与节点 6 之间还备份有连接通道，开通节点 6 到节点 7 方向上的连接通道作为节点 6 与节点 7 之间的时钟通道，如图 8C 所示，这样，节点 7 仍可得到时钟源提供的时钟。

图 9 为图 6 所示计算机系统中 8 个节点的另外一种时钟连接示意图。其表现也是一个长方体，上下前后左右都为顺序环，不交叉，其时钟关系连接表如表 7 所示。

表 7

起点	终点
0	1、3、4
1	0、2、5
2	1、3、6
3	0、2、7
4	0、5、7
5	1、4、6
6	2、5、7

7	3、4、6
---	-------

上述装置及方法实施例通过节点之间的时钟连接通道及时钟配置方法，保证了分区系统内各节点的时钟同步；并且，当部分节点时钟异常时，不会对系统有所影响，即系统还没起来时，若部分节点时钟异常时，也可以组建一个分区，正常的把系统运行起来；当部分节点时钟通道异常，最大程度地保证了分区系统不受影响；当部分节点异常时，最大程度地保证了剩余节点所组成的分区系统不受影响；当节点异常数过多时无法组建一个分区时，还可以将剩余节点组建几个独立的分区系统。

上述装置及方法实施例中的节点之间的时钟连接方式也可应用于分区的快速通道互联（Quick Path Interconnect, QPI）联接，及类似地其他系统中节点之间的联接。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

权利要求书

1. 一种计算机系统，包括至少两个节点，其特征在于，所述至少两个节点均包括选择模块及 CPU，所述选择模块的输入为本节点的时钟及其他节点输出的时钟，输出端与所述 CPU 及其他节点的选择模块的输入端相连；

5 所述计算机系统还包括时钟控制模块，所述时钟控制模块的输出端与所述选择模块的控制端相连，用于控制所述至少两个节点的时钟为同一个时钟。

2、根据权利要求 1 所述的计算机系统，其特征在于，所述至少两个节点为四个节点，所述四个节点通过选择模块连接为环形，每个节点的选择模块的输入为本节点的时钟及相连的两个节点的选择模块的输出。

10 3、根据权利要求 1 所述的计算机系统，其特征在于，所述至少两个节点为 8 个节点，位于一个长方体的各个顶点处，所述长方体各个表面内的 4 个节点通过选择模块交叉连接或通过选择模块两两相邻连接，且每个节点与三个节点通过选择模块连接，每一节点的选择模块的输入为该节点的时钟及相连的三个节点的选择模块的输出。

15 4、根据权利要求 1 所述的计算机系统，其特征在于，所述至少两个节点为 $n \times 8$ 个节点，其中，每 8 个节点位于一个长方体的各个顶点处，所述长方体各个表面内的 4 个节点通过选择模块交叉连接或通过选择模块顺次连接，且每个节点与 $3+n$ 个节点通过选择模块连接，每个所述长方体相同位置上的节点通过选择模块相连，每个节点的选择模块的输入为本节点的时钟与相连
20 的 $3+n$ 个节点的选择模块的输出，其中， n 为大于 0 的自然数。

5、根据权利要求 2-4 任一项所述的计算机系统，其特征在于，所述计算机系统中属于同一个分区系统的节点相互之间直接连接或间接连接，所述间接连接的中间节点与所述间接连接的两端节点位于同一个分区系统内。

6、根据权利要求 1-4 任一项所述的计算机系统，其特征在于，所述选择
25 模块的输出端与其他节点的选择模块的输入端的连接通过 CPU 互联线缆实现。

7、一种用于实现权利要求 1-6 任一项所述的计算机系统中同一分区系统的节点时钟同步的时钟配置方法，其特征在于，包括：

选择一节点的时钟作为所属分区系统的时钟源的选择步骤；

5 开通所述节点作为起点，与所述节点之间相连的节点作为终点的连接通道作为所述节点及与之直接相连的节点之间的时钟通道的开通步骤；

判断作为终点的节点是否与所述分区系统中剩余的节点相连，若是，则执行所述开通步骤；否则，完成所述分区系统的时钟配置。

8、根据权利要求 7 所述的时钟配置方法，其特征在于，还包括：根据所述计算机系统中节点之间的时钟连接关系建立时钟关系连接表；

10 所述开通步骤根据所述时钟关系连接表得到节点之间的连接关系。

9、根据权利要求 7 或 8 所述的时钟配置方法，其特征在于，执行所述开通步骤时或之后，还包括：

建立时钟关系路由表，将所述起点及终点添加到所述时钟关系路由表中。

15 10、根据权利要求 9 所述的时钟配置方法，其特征在于，根据所述时钟关系路由表，重新开通由于时钟异常、时钟通道异常或节点异常导致时钟通道断开的节点与时钟源所在节点之间的时钟通道。

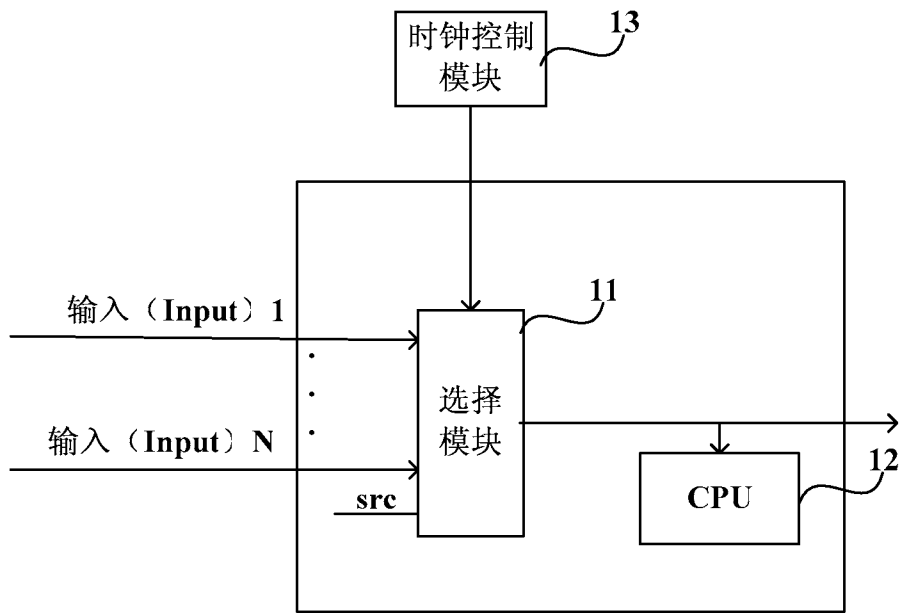


图 1

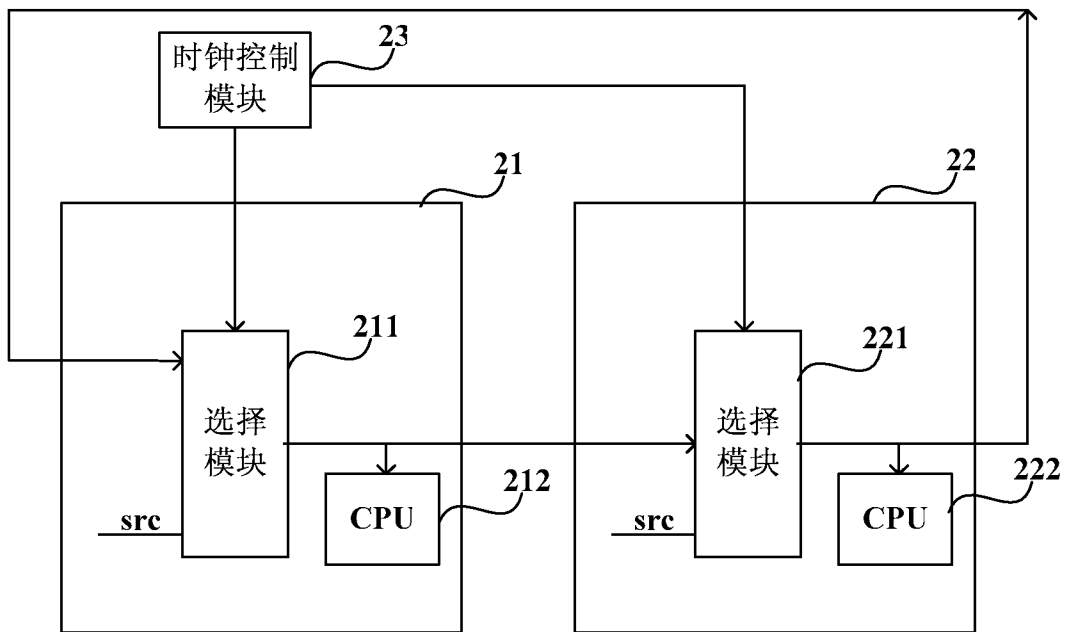


图 2

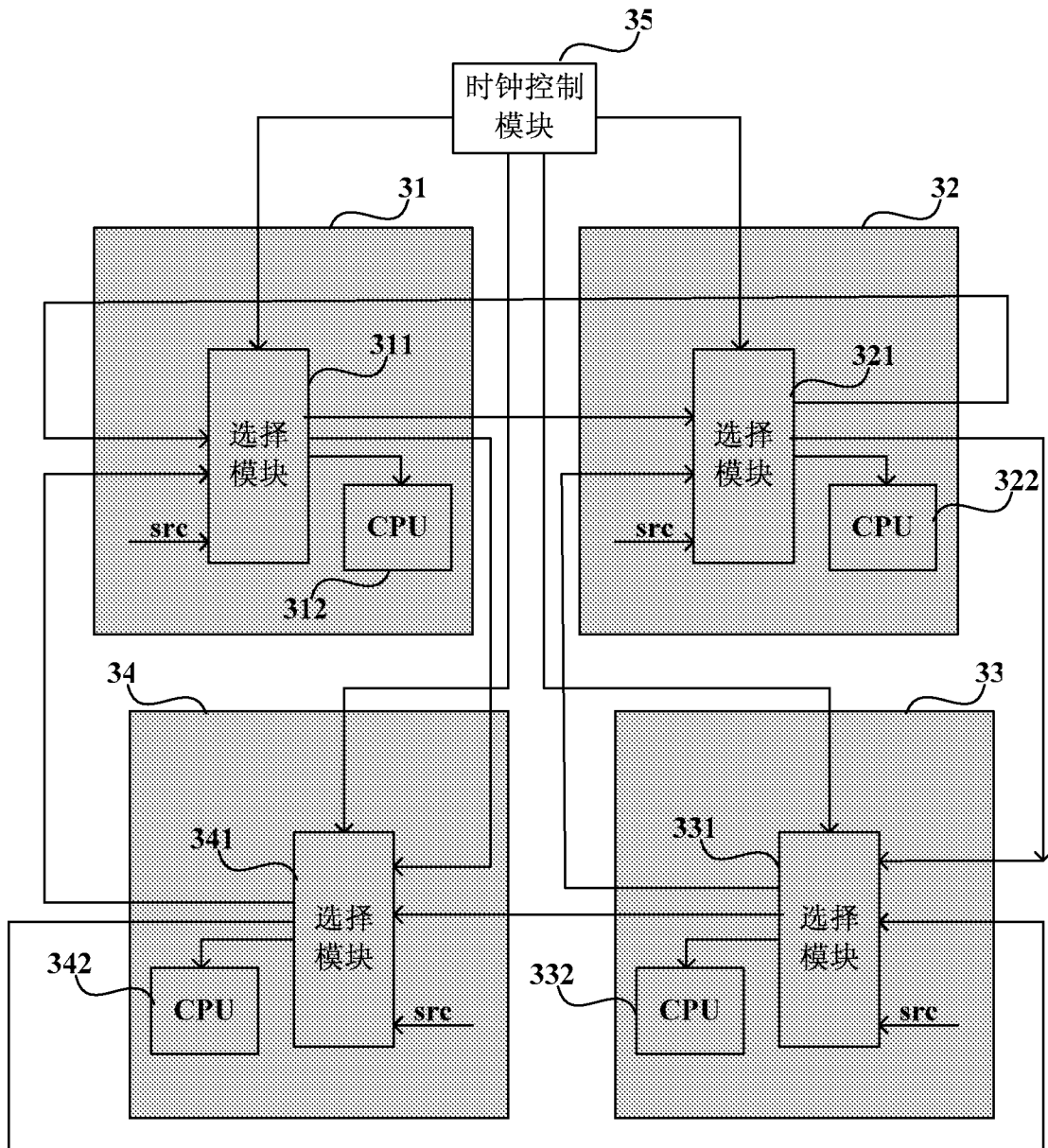


图 3A



图 3B

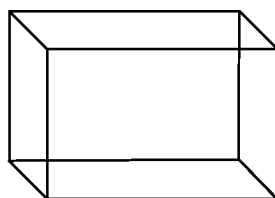


图 4A

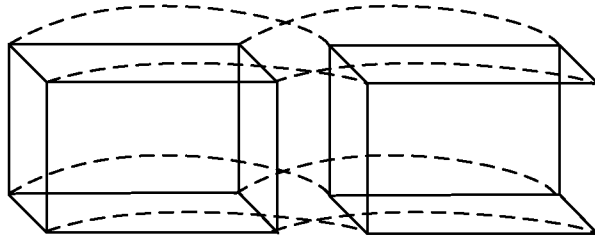


图 4B

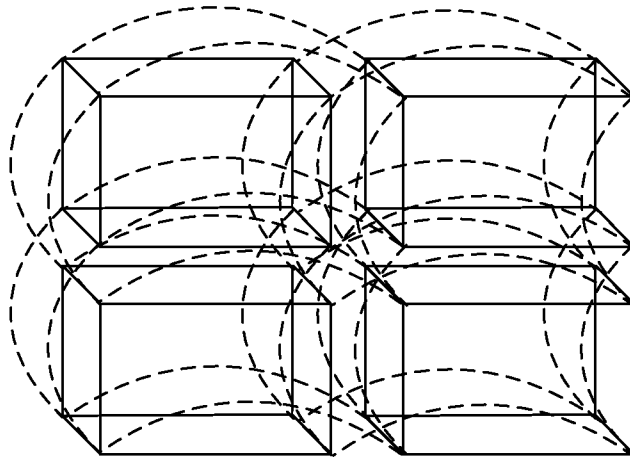


图 5

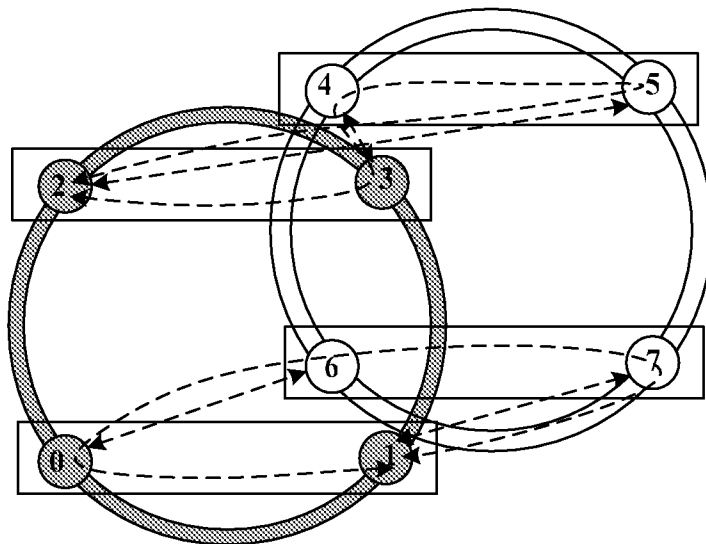


图 6

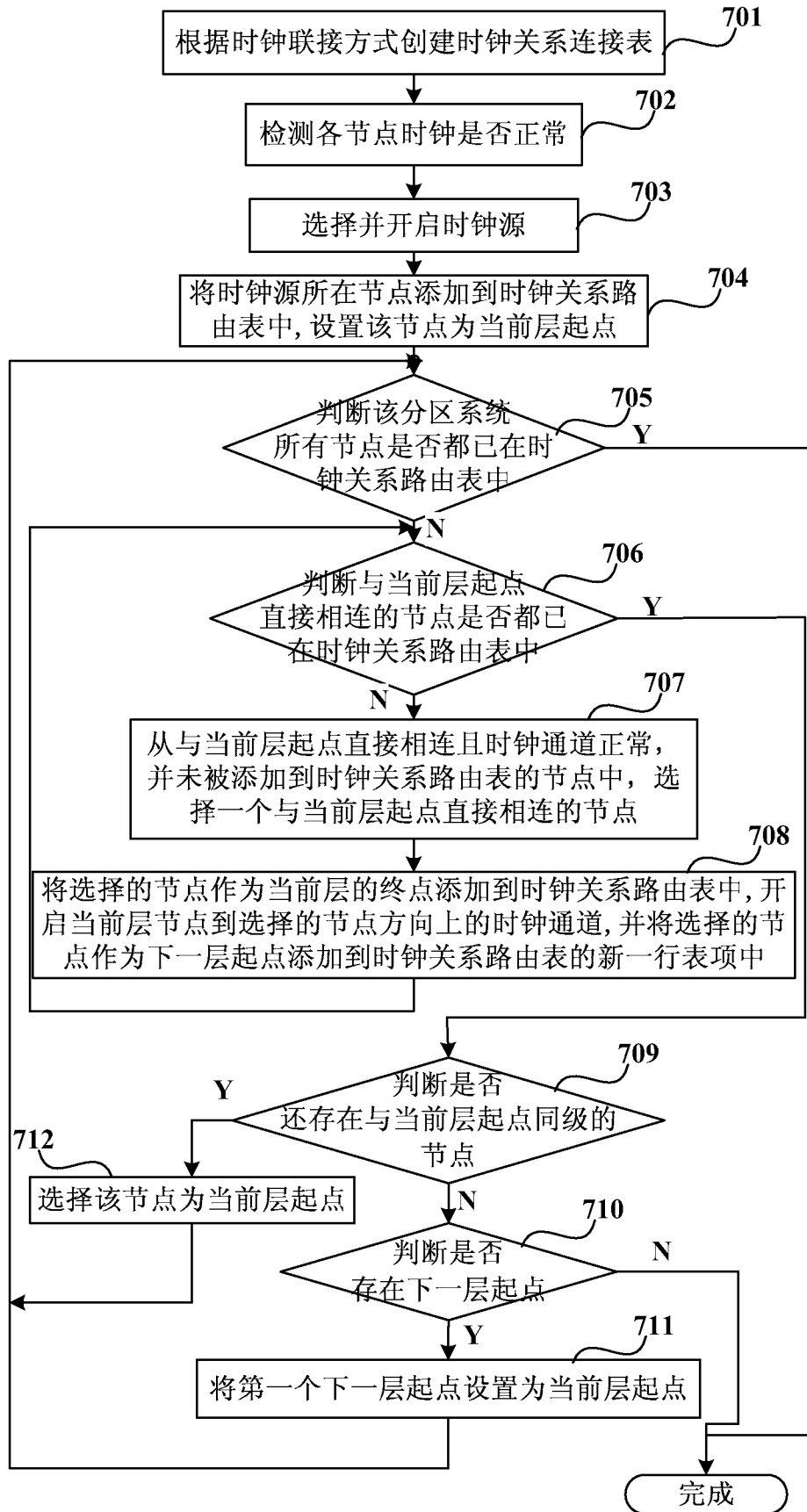


图 7

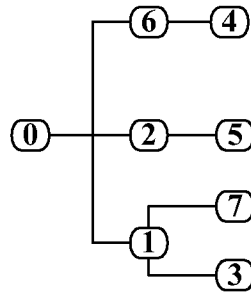


图 8A

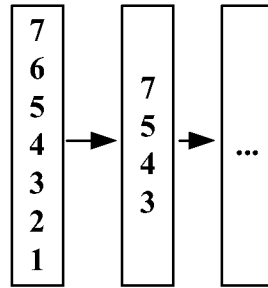


图 8B

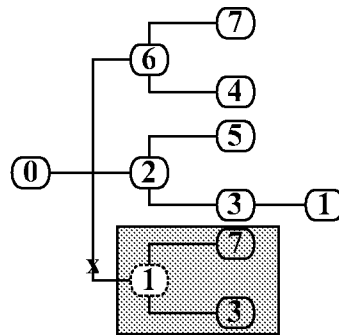


图 8C

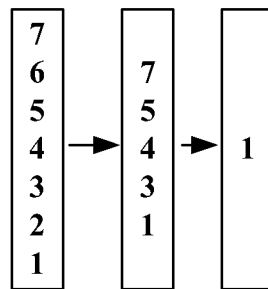


图 8D

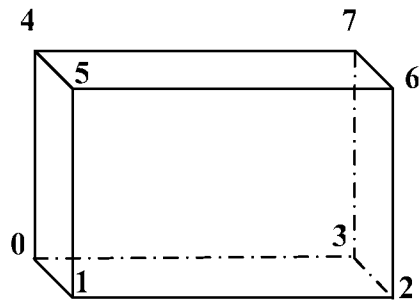


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/077625

A. CLASSIFICATION OF SUBJECT MATTER

G06F 1/12 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F, H04B, H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNABS, CNTXT, VEN: clock, clocking, node, vertex, select, elect, choose, switch, switching, source, synchronization, synchronic, synchronical, synchronically

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	CN 102317885 A (HUAWEI TECHNOLOGIES CO., LTD.), 11 January 2012 (11.01.2012), see claims 1-10	1-10
Y	CN 101192913 A (ZTE CORP.), 04 June 2008 (04.06.2008), see description, page 2, paragraph 5, page 9, paragraphs 5-7, page 10, paragraph 5, and figures 1-4	1-2, 5-6
Y	CN 1394004 A (SHANGHAI NO.2 RESEARCH INSTITUTE, ZTE CORPORATION), 29 January 2003 (29.01.2003), see description, page 1, paragraphs 1-2, page 7, paragraph 1, and figures 2-3	1-2, 5-6
A	CN 101090341 A (SUPCON GROUP CO., LTD. et al.), 19 December 2007 (19.12.2007), see the whole document	1-10
A	JP 11103312 A (MITSUBISHI ELECTRIC CORP.), 13 April 1999 (13.04.1999), see the whole document	1-10

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search
07 March 2012 (07.03.2012)

Date of mailing of the international search report
12 April 2012 (12.04.2012)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer
LIANG, Tao
Telephone No.: (86-10) **62411847**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2011/077625

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102317885 A	11.01.2012	None	
CN 101192913 A	04.06.2008	CN 101192913 B	08.12.2010
CN 1394004 A	29.01.2003	CN 1281005 C	18.10.2006
CN 101090341 A	19.12.2007	CN 100525212 C	05.08.2009
JP 11103312 A	13.04.1999	None	

A. 主题的分类		
G06F 1/12(2006.01)i		
按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类		
B. 检索领域		
检索的最低限度文献(标明分类系统和分类号)		
IPC: G06F, H04B, H04L		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))		
CPRSABS, CNABS, CNTXT, VEN: 时钟,节点,顶点,选择,切换,源,同步; Clock, clocking, Node, vertex, Select, elect, choose, switch, switching, Source, Synchronization, Synchronic, Synchronical, Synchronically		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
E	CN102317885A (华为技术有限公司) 11. 1 月 2012 (11.01.2012) 参见权利要求 1-10	1-10
Y	CN101192913A (中兴通讯股份有限公司) 04. 6 月 2008 (04.06.2008) 参见说明书第 2 页第 5 段, 第 9 页第 5-7 段, 第 10 页第 5 段, 附图 1-4	1-2, 5-6
Y	CN1394004A (深圳市中兴通讯股份有限公司上海第二研究所) 29. 1 月 2003 (29.01.2003) 参见说明书第 1 页第 1-2 段, 第 7 页第 1 段, 附图 2-3	1-2, 5-6
A	CN101090341A (中控科技集团有限公司等) 19. 12 月 2007(19.12.2007) 参见全文	1-10
A	JP11103312A (MITSUBISHI ELECTRIC CORP) 13. 4 月 1999 (13.04.1999) 参见全文	1-10
<input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。		
* 引用文件的具体类型:		“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件
“A” 认为不特别相关的表示了现有技术一般状态的文件		“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性
“E” 在国际申请日的当天或之后公布的在先申请或专利		“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性
“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)		“&” 同族专利的文件
“O” 涉及口头公开、使用、展览或其他方式公开的文件		
“P” 公布日先于国际申请日但迟于所要求的优先权日的文件		
国际检索实际完成的日期 07.3 月 2012 (07.03.2012)	国际检索报告邮寄日期 12.4 月 2012 (12.04.2012)	
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号: (86-10)62019451	受权官员 梁韬 电话号码: (86-10) 62411847	

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2011/077625

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 102317885 A	11.01.2012	无	
CN 101192913 A	04.06.2008	CN 101192913 B	08.12.2010
CN 1394004 A	29.01.2003	CN 1281005 C	18.10.2006
CN 101090341 A	19.12.2007	CN 100525212 C	05.08.2009
JP 11103312 A	13.04.1999	无	