(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111107020 A
(43)申请公布日 2020.05.05

(21)申请号 201911421236.1

(22)申请日 2019.12.31

(71)申请人 盛科网络(苏州)有限公司
地址 215000 江苏省苏州市工业园区星汉街5号B幢4楼13/16单元

(72)发明人 蒋震 方沛昱 夏杰 龚海东

(74)专利代理机构 苏州三英知识产权代理有限公司 32412
代理人 潘时伟

(51)Int.Cl.
H04L 12/931(2013.01)
H04L 12/933(2013.01)

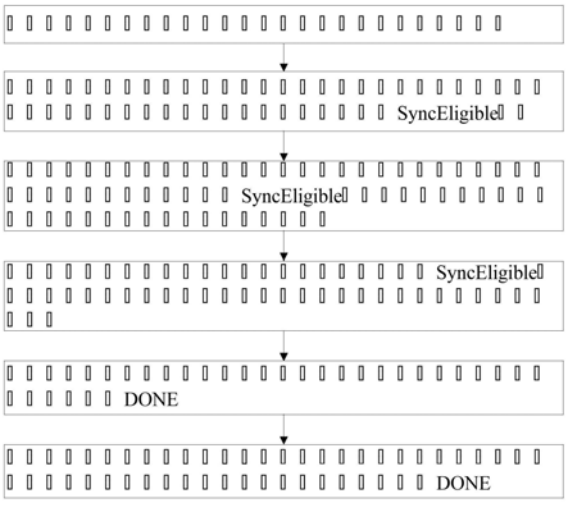
权利要求书1页 说明书4页 附图1页

(54)发明名称

一种多核心以太网交换芯片时间同步的方法

(57)摘要

本发明揭示了一种多核心以太网交换芯片时间同步的方法,通过在两个核心的自由的计数器上加入一套自同步机制,完成两个核心之间的校准,解决两个核心在一个封装上的互操作问题,使得使用该方案设计的交换芯片在支撑时间同步协议的时候与单核心芯片的行为没有差异。



CN 111107020 A

1.一种多核心以太网交换芯片时间同步的方法,其特征在于,所述方法包括:

S1,当第一核心释放复位信号时,其计数器在参考时钟的驱动下自增,同时第一核心在每一个周期向第二核心发送可同步消息;

S2,当第二核心释放复位信号时,所述第二核心的计数器保持不变,直至第一次收到所述第一核心发送的所述可同步消息时,开始自增,并在每个周期向第一核心回复同步请求消息;

S3,当第一核心收到第一个所述同步请求消息后,停止发送可同步消息并记录当前第一计数器数值,同时发送一个同步结束消息给第二核心;

S4,第二核心收到同步结束消息后,回复同步结束消息给第一核心,并置位自身为完成;

S5,第一核心收到同步结束消息时,记录当前第二计数器数值,并对所述第二计数器数值进行计算调整,调整后置位自身为完成。

2.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,在所述S1之前,将第一核心和第二核心分别配置为主工作模式和从工作模式。

3.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,S1中,第一核心通过两个核心间的通信接口向第二核心发送可同步消息。

4.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,S5中,第一核心对第二计数器数值进行计算调整包括:

S51,第一核心将S5中收到同步结束消息时的第一时间,与S3中收到第一同步请求消息时的第二时间相减,记录计算得到的时间差;

S52,将第二计数器数值减去所述第二时间与所述时间差/2之和。

5.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,第一核心和第二核心使用相同的参考时钟进行驱动。

6.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,所述方法还包括:第一核心和第二核心各自在自身完成信号置位的时候,用自身的计数器数值与CPU配置的偏移值相加,用计算得到的时间对时间同步协议报文进行打戳处理。

7.根据权利要求6所述的多核心以太网交换芯片时间同步的方法,其特征在于,CPU同时配置第一核心和第二核心中的所述偏移值。

8.根据权利要求1所述的多核心以太网交换芯片时间同步的方法,其特征在于,所述可同步消息、同步请求消息和同步结束消息的位数为2比特。

9.根据权利要求6所述的多核心以太网交换芯片时间同步的方法,其特征在于,所述时间同步协议报文为PTP报文。

一种多核心以太网交换芯片时间同步的方法

技术领域

[0001] 本发明属于多核心以太网交换芯片时间同步技术领域，具体涉及一种多核心以太网交换芯片时间同步的方法。

背景技术

[0002] 随着超大规模云网络，存储网络及HPC(高性能计算)等场景的发展，网络上的数据交换量越来越大，最高单芯片处理能力不断的上升，从Gbps发展到Tbps数量级。但是当前芯片生产工艺14nm/12nm或7nm/6nm，其IP core(Intellectual Property core，知识产权核)可以运行的时钟频率最高分别在1.05GHz或1.7GHz。在单一流水线核心的前提下，无法支撑高达25.6Tbps的处理能力。

[0003] 从工程角度来看，为了应对迅速上升的报文处理带宽，在单一核心频率受限的情况下，多核心设计成为必选的方向。从应用角度，芯片工作时所展现出来的系统行为不应当感知到芯片架构是单核心还是多核心设计。因此，多个核心间的状态同步是必须的。另一方面，随着芯片生产工艺的进步，流片(Tape out)费用越来越高。为了丰富产品线，高带宽和超高带宽的交换芯片都需要有设计。因此，通过使用D2D(Die-to-Die，裸晶互联)技术，可以在双核设计的前提下，实现一次流片覆盖多个产品线:超高带宽单芯片使用两个Die(裸晶)连接封装;高带宽芯片使用一个Die封装。

[0004] 支持双Die(裸晶)互联封装的超高带宽以太网交换芯片，两个Die是对称的。即，如果只有一个Die进行封装，可以独立作为高带宽以太网交换芯片工作。针对时间同步特性，每一个Die上都有一个自由的计数器表达Die上电后经过的Tick数。如果这一个自由的计数器的参考时钟是500MHz，那么计数器每一个Tick自增2，则1秒钟可以得到1,000,000,000的计数。换句话说，计数器每1,000,000,000代表现实经过了1秒。两个Die各自独立初始化，初始化完成后，由于两个Die都有计数器，释放计数器复位信号的时间并不一致，因此，即便两个计数器可以使用同一个参考时钟，两个Die上的计数器数值还是会存在差异。

[0005] 因此，针对上述技术问题，有必要提供一种多核心以太网交换芯片时间同步的方法及装置。

发明内容

[0006] 有鉴于此，本发明的目的在于提供一种多核心以太网交换芯片时间同步的方法。

[0007] 为了实现上述目的，本发明一实施例提供的技术方案如下：

[0008] 一种多核心以太网交换芯片时间同步的方法，所述方法包括：

[0009] S1，当第一核心释放复位信号时，其计数器在参考时钟的驱动下自增，同时第一核心在每一个周期向第二核心发送可同步消息；

[0010] S2，当第二核心释放复位信号时，所述第二核心的计数器保持不变，直至第一次收到所述第一核心发送的所述可同步消息时，开始自增，并在每个周期向第一核心回复同步请求消息；

[0011] S3,当第一核心收到第一个所述同步请求消息后,停止发送可同步消息并记录当前第一计数器数值,同时发送一个同步结束消息给第二核心;

[0012] S4,第二核心收到同步结束消息后,回复同步结束消息给第一核心,并置位自身为完成;

[0013] S5,第一核心收到同步结束消息时,记录当前第二计数器数值,并对所述第二计数器数值进行计算调整,调整后置位自身为完成。

[0014] 一实施例中,在所述S1之前,将第一核心和第二核心分别配置为主工作模式和从工作模式。

[0015] 一实施例中,S1中,第一核心通过两个核心间的通信接口向第二核心发送可同步消息。

[0016] 一实施例中,S5中,第一核心对第二计数器数值进行计算调整包括:

[0017] S51,第一核心将S5中收到同步结束消息时的第一时间,与S3中收到第一同步请求消息时的第二时间相减,记录计算得到的时间差;

[0018] S52,将第二计数器数值减去所述第二时间与所述时间差/2之和。

[0019] 一实施例中,第一核心和第二核心使用相同的参考时钟进行驱动。

[0020] 一实施例中,所述方法还包括:第一核心和第二核心各自在自身完成信号置位的时候,用自身的计数器数值与CPU配置的偏移值相加,用计算得到的时间对时间同步协议报文进行打戳处理。

[0021] 一实施例中,CPU同时配置第一核心和第二核心中的所述偏移值。

[0022] 一实施例中,所述可同步消息、同步请求消息和同步结束消息的位数为2比特。

[0023] 一实施例中,所述时间同步协议报文为PTP报文。

[0024] 本发明具有以下有益效果:通过在两个核心的自由的计数器上加入一套自同步机制,完成两个核心之间的校准,解决两个核心在一个封装上的互操作问题,使得使用该方案设计的交换芯片在支撑时间同步协议的时候与单核心芯片的行为没有差异。

附图说明

[0025] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0026] 图1为本发明的流程示意图。

具体实施方式

[0027] 为了使本技术领域的人员更好地理解本发明中的技术方案,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,都应当属于本发明保护的范围。

[0028] 本发明所揭示的一种多核心以太网交换芯片时间同步的方法,通过在两个核心的

自由的计数器上加入一套自同步机制,实现两个核心之间的同步。

[0029] 如图1所示,本发明所揭示的一种多核心以太网交换芯片时间同步的方法,包括:

[0030] 第一步,将第一核心和第二核心分别配置为主工作模式和从工作模式。

[0031] 具体地,为了方便描述,定义第一核心为核心0,第二核心为核心1,即,本实施例中,将核心0配置为主工作模式,核心1配置为从工作模式,当然,在其他实施例中,可替换为将核心1配置为主工作模式,核心0配置为从工作模式。

[0032] 第二步,当第一核心释放复位信号时,其计数器在参考时钟的驱动下自增,同时第一核心在每一个周期向第二核心发送可同步消息。

[0033] 具体地,本实施例中,当核心0释放复位信号时,核心0的计数器开始在参考时钟的驱动下自增。同时核心0在参考时钟驱动下每一个周期通过核心间通信接口向核心1发送可同步(SyncEligible)消息,本实施例中,可同步消息只需要2bit,设置值为0x1即可。

[0034] 第三步,当第二核心释放复位信号时,所述第二核心的计数器保持不变,直至第一次收到所述第一核心发送的所述可同步消息时,开始自增,并在每个周期向第一核心回复同步请求消息。

[0035] 具体地,本实施例中,当核心1释放复位信号的时候,核心1的计数器保持不变,直到其第一次收到核心0发送的可同步消息的时候,开始自增。并在每一个周期向核心0回复同步请求(SyncAck)消息,本实施例中,SyncAck消息同样只需要2bit,设置值为0x2。

[0036] 第四步,当第一核心收到第一个所述同步请求消息后,停止发送可同步消息并记录当前第一计数器数值,同时发送一个同步结束消息给第二核心。

[0037] 具体地,本实施例中,当核心0收到第一个SyncAck消息后,停止发送可同步消息并记录当前计数器的数值,同时发送一个同步结束(SyncFin)消息给核心1。本实施例中,SyncFin消息同样只需要2bit,设置值为0x3。

[0038] 第五步,第二核心收到同步结束消息后,回复同步结束消息给第一核心,并置位自身为完成。

[0039] 具体地,本实施例中,核心1收到SyncFin消息后,直接回复SyncFin消息给核心0,并置位自己为完成(DONE)。

[0040] 第六步,第一核心收到同步结束消息时,记录当前第二计数器数值,并对所述第二计数器数值进行计算调整,调整后置位自身为完成。

[0041] 具体地,本实施例中,核心0收到SyncFin消息的时候,记录当前计数器数值。并将收到SyncFin消息的第一时间T1和收到第一个SyncAck消息的第二时间T2相减,得到的deltaT记录下来,并将计数器的值减去(T2+deltaT/2),调整后置位自己为完成。此时两个核心同步完成。

[0042] 优选地,上述两个核心(即第一核心和第二核心)使用相同的参考时钟进行驱动,因此两侧的自由时间值(即计数器值)不会再次产生偏差。

[0043] 进一步地,每一个核心在自身完成信号置位的时候,用自己的自由计数器值与CPU配置的偏移(Offset)值相加,得到的值即为当前的校准后时间,使用这个时间对时间同步协议报文(如PTP报文,其中,PTP英文全称为Precision Time Protocol,中文解释为精确时间协议报文)进行打戳处理。由于CPU会同时配置两个核心中的Offset,因此虽然两个核心不再交互,但是两处的时间值是保持一致的。此时两个核心对外的行为是一致的、无差别的

单一系统时间。

[0044] 由以上技术方案可以看出，本发明具有以下优点：解决了两个核心在一个封装上的互操作问题，使得使用该方法设计的交换芯片在支撑时间同步协议的时候与单核心芯片的行为没有差异。

[0045] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。

[0046] 为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本说明书一个或多个实施例时可以把各模块的功能在同一个或多个软件和/或硬件中实现。

[0047] 还需要说明的是，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、商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、商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、商品或者设备中还存在另外的相同要素。

[0048] 本领域技术人员应明白，本说明书一个或多个实施例的实施例可提供为方法、系统或计算机程序产品。因此，本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本说明书一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0049] 本说明书一个或多个实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书一个或多个实施例，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0050] 对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化囊括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。

[0051] 此外，应当理解，虽然本说明书按照实施方式加以描述，但并非每个实施方式仅包含一个独立的技术方案，说明书的这种叙述方式仅仅是为清楚起见，本领域技术人员应当将说明书作为一个整体，各实施例中的技术方案也可以经适当组合，形成本领域技术人员可以理解的其他实施方式。

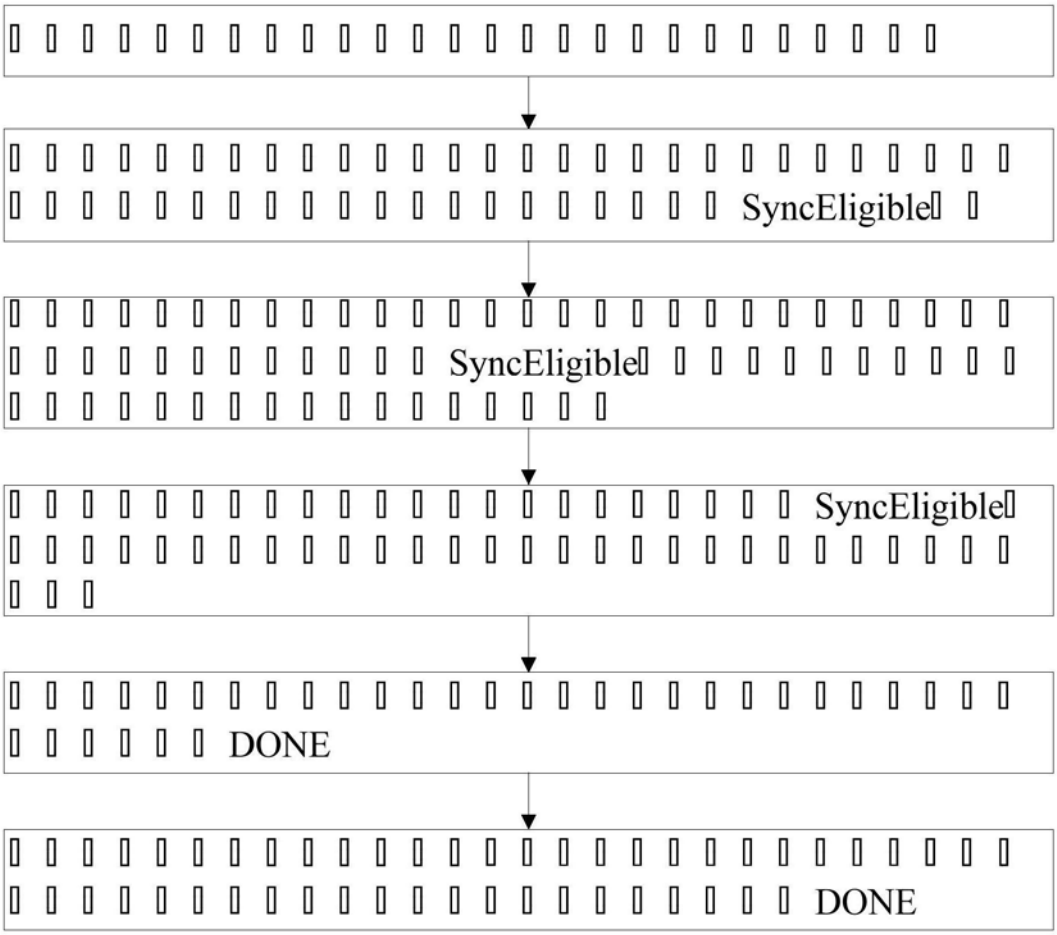


图1