

基于 PCIe 2.0 协议的 PCS 层弹性缓冲器设计

武桂林, 黄 鲁

(中国科学技术大学 电子科学与技术系, 安徽 合肥 230027)

摘 要: 弹性缓冲器能够补偿时钟偏差, 解决了不同时钟域下的数据传输问题, 被广泛应用于高速数据流通信中. 结合 PCIe 2.0 协议, 采用常半满方式对弹性缓冲器进行了设计. 在读写指针的对比过程中, 相对于传统的同步指针的实现方法, 采用常半满的检查方式和半满同步方法, 简化了逻辑结构. 并且在充分考虑协议的基础上, 不仅实现了时钟补偿的功能, 而且能很好地与接收端的其他模块完成接口对接. 经 modelsim 仿真验证, 所设计的缓冲器支持 250 MHz 的读写时钟频率, 输入输出数据为 20 bit, 适用于 5G 全速率(250 MHz)和 2.5G 半速率(125 MHz)两种模式.

关键词: 弹性缓冲器; 半满检查; 常半满方式; PCIe 2.0 协议

中图分类号: TN492

文献标识码: A

文章编号: 1000-7180(2016)09-0051-04

Design of Elastic Buffer in Physical Coding Sublayer Based on PCIe 2.0

WU Gui-lin, HUANG Lu

(Department of Electronic Science and Technology, University of Science and
Technology of China, Hefei 230027, China)

Abstract: Elastic buffer can compensate the clock deviation to solve the problem of data transmission under different clock domains, so it is widely used in high speed communication system. Based on PCIe 2.0 protocol, this paper designs an Elastic buffer with normal half-full method. Comparing with the traditional method to synchronize the pointer, it checks the reading pointer when the writing pointer is on the half position in elastic buffer. Thus, it simplified the logical structure. In addition, it not only compensate the clock deviation, but also connect with interfaces of other module compatibly at the receiver of PCS. After verifying on modelsim, the writing and reading clock can achieve 250 MHz, the width of input and output data is 20 bit. So it can support two modes, 5 GT/s(full-rate) and 2.5 GT/s(half-rate).

Key words: elastic buffer; SKP ordered-sets; half-full method; PCIe 2.0 protocol

1 引言

PCIe 总线是新一代总线结构, 与传统的 PCI, ISA 和 AGP 并行总线不同, 它采用点对点的串行方式传输数据, 支持 1x 至 32x 的通道连接, 具有较高的数据传输率. PCIe 2.0 协议要求发射端发送的数据采用 IBM 的 8B/10B 编码方式^[1], 这一编码方式保证了数据流中有足够的 0-1 和 1-0 的跳变, 有利于接收端的 CDR 从数据流中恢复出时钟. 但是由于

CDR 恢复出来的时钟与接收端本地的时钟来自不同的时钟域, 这就需要弹性缓冲器(Elastic Buffer, EB)对时钟的偏差进行补偿^[2].

EB 其实就是一个异步 FIFO^[3], 由于我们可以根据需要在 FIFO 中插入或者删除数据来补偿读写时钟的偏差, 所以称为弹性缓冲器. 弹性缓冲器的概念可以追溯到上世纪 60 年代. 在 1963 年, Maurice Karnaugh 获得这一技术的专利, 只不过当时是应用在 PCM 电话网络上. 现在 EB 的作用越发明显, 在

USB、InfiniBand 架构、光纤通道、千兆以太网中都有应用^[4],这里只针 PCIe 2.0 协议. PCIe 2.0 协议要求的最大运行速率为 5 GT/s(全速率),因此采用半满方式设计的 EB 最大运行时钟频率为 250 MHz,且在 125 MHz 的半速率模式下同样适用.

2 弹性缓冲器的作用

PCIe 2.0 协议支持 5 GT/s(全速率)和 2.5 GT/s(半速率)两种速率模式. 以 5 GT/s 的全速率模式为例,时钟允许的误差范围为 $\pm 300 \times 10^{-6}$ [1],即总线允许的频率范围为 4.998 5~5.0015 GHz. 分频后产生的接收端本地读时钟与 CDR 恢复的写时钟之间的最大符号周期偏差为 1.2 ps. 当经过 1 666 个符号周期后,就可能移位了一个时钟周期(2 ns). 如果 CDR 恢复的时钟频率大于接收端本地的时钟频率,就有可能导致缓冲区出现被读空的状态;反之,就有可能导致缓冲区溢出的情况. 这两种情况都会造成数据的丢失或者出错. EB 通过在合适的时机插入或者删除 SKP 字符的方式来避免这些情况的发生.

SKP 有序集是为了实现时钟补偿而规范的一组特殊有序集,由 1 个 COM 字符和 3 个 SKP 字符组成,如图 1 所示. 在两个时钟偏差最大的情况下,每隔 1 666 符号周期,就会出现一个周期的偏差,因此 SKP 有序集的发送间隔不能大于 1 666.

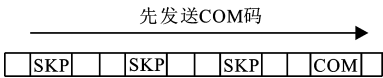


图 1 SKP 有序集

3 弹性缓冲器的位置

EB 对 SKP 有序集中的 SKP 字符的插入和删除操作是在符号的层次上来进行的,因此它位于解串器的后面. 其次由于 PCIe 协议规范了一种回环的操作模式,即将接收端接收的未解码的字符回环到接收端,这也需要先进行时钟补偿,因此 EB 位于 8B/10B 解码之前. 下面讨论 EB 与解偏电路之间的位置关系.

解偏电路主要是用来去除由于工艺或者布线导致的通道与通道之间数据偏移问题(1x 模式不存在这种状况). 以 4x 模式为例,发送端在发送 SKP 有序集时,4 个通道是同时发送的,如图 2 所示. 由于通道之间的偏移,数据经过链路被接收端接收时就会变成图 3 下的状态. 如果 EB 位于解偏电路的后

面,则解偏电路需要进行处理的情况就如图 3 所示. 反之,如果 EB 位于解偏电路前面,从 EB 读取出来给解偏电路的数据状态如图 4 所示,经过对比可以发现,EB 具有一定的去偏移功能,因此将 EB 放在解偏电路的前面可以减少一部分解偏电路的工作量,这样的安排显得更加合理.

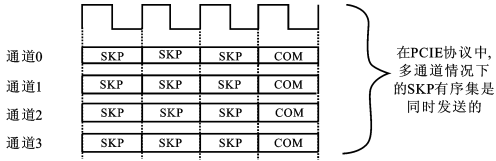


图 2 发射端同时发送 SKP 有序集

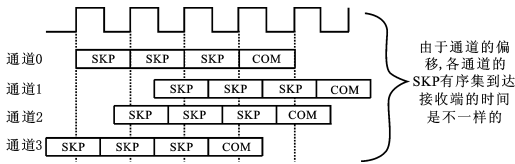


图 3 未经过 EB 之前的 SKP 有序集在各通道的情况

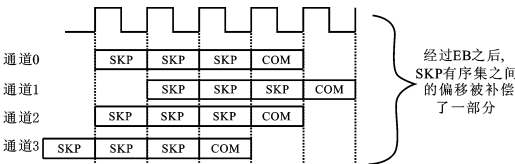


图 4 经过 EB 后 SKP 有序集在各通道里面的情况

4 弹性缓冲器的深度

PCIe 2.0 协议规定,发送端每 1 180~1 538 符号周期,就发送一个 SKP 有序集,但如果发送器正好在 1 538 个符号的时间处发送一个最大尺寸的 TLP 包,就要推迟发送 SKP 有序集. 因此,EB 必须具有处理最大间距的 SKP 有序集的能力. 根据设备支持的最大数据包的尺寸,其最大可能产生的符号移位个数由下式表示:

$$\text{Max_Symbol_Shifted} = \left(\frac{\text{Max_Payload_Size} + \text{TLP_Overhead}}{\text{Link_Width}} \right) + 1\ 538$$

1 666

(1)

式中,Max_Payload_Size 是一个变量,它为器件允许的最大包的长度,TLP_overhead 是一个值为 28 的常量,是一个包的附加长度,如包头,帧字符(STP,SDP 或者 END),Link_Width 是 1,而不是链路支持的最大链路宽度. 数据最大移位个数与最大数据载荷之间的关系如表 1 所示.

表 1 最大移位个数与数据载荷之间的关系

Max_Payload_Size	128	256	512	1 024	2 048	4 096
Max _ Sym-bols-Shifted	1. 02	1. 09	1. 25	1. 55	2. 17	3. 40

对于缓冲器来说,它存在上溢和下溢两种情况,因此它的深度为它的最大移位字符数目的 2 倍. 例如,数据的最大载荷是 4 096 的话,它的最小深度就为 8. 但在实际应用中,为了可靠地避免上溢和下溢情况的发生,EB 的深度会被适当地加大.

5 弹性缓冲器的工作原理

EB 的实现有两种方式,一种是基于半满的实现方式,另一种是基于流程控制的方式,这里只介绍基于半满的实现方式.

正常情况下,EB 中存储的数据数目为 EB 深度的一半,EB 在进行时钟补偿时分为两种情况:一种是写时钟速率大于读时钟速率,为防止数据溢出,EB 删除 SKP 字符使自己回到半满的状态,如图 5 所示;另一种是写时钟速率小于读时钟速率,为防止数据下溢,EB 插入 SKP 字符同样使自己回到半满的状态,如图 6 所示.

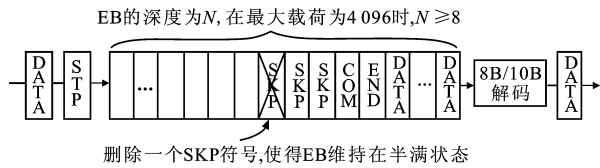


图 5 SKP 字符删除操作

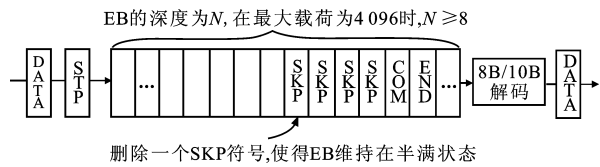


图 6 SKP 字符插入操作

6 弹性缓冲器的实现

PCIe 2.0 协议允许的最大数据包长度为 4 096^[2],由于读写指针在同步检测过程中需要时间,因此将 EB 的深度扩大到了 16,其宽度为 11(最高位用于表示数据是否有效,低 10 位用于存储数据),读写指针宽度为 4. 在全速率模式下,如果没有时钟偏差,读写时钟频率都等于 250 MHz,一个时钟周期传输 20 bit 的数据,EB 的读写指针以加 2 的形式递增.

本文设计的弹性缓冲器的结构如图 7 所示,它包括输入输出控制模块,读写指针控制模块,写指针半满状态检查模块,半满状态信号同步模块,读指针状态检查模块,SKP 字符检查模块和存储器模块.

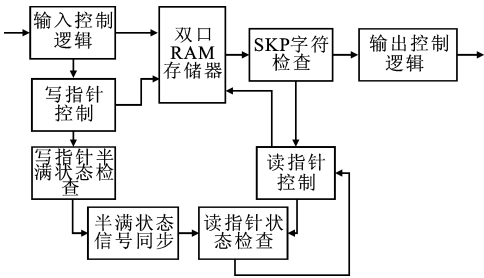


图 7 弹性缓冲器结构

6.1 输入输出控制模块

输入输出控制模块主要实现与其他模块接口的连接. EB 的运作受到 PCS 层的接收模块和 CDR 的双重控制,在 EB 的输入端,它除了拥有各自时钟域下的复位信号 rst_wr_clk 和 rst_rd_clk 外,还有一个全局的复位信号 $global_reset$ (高电平有效),它由 PCS 层和 CDR 共同控制,当 PCS 层允许数据输出且 CDR 运行时,数据才被允许写入存储器中. 输入控制逻辑就是用来确定数据是否写入 EB. 在 EB 的输出端,输出控制逻辑不仅要输出数据,还要输出 EB 当前的操作状态. 此外,为了方便后面的 8B/10B 解码模块,SKP 有序集检测模块中产生的 COM 码检测信息 $com_det[1:0]$ 也会随数据一起传递出去.

6.2 写指针半满检查和半满状态同步模块

读写指针是在各自时钟域下进行递增操作的. 由于读写时钟存在偏差,直接进行指针的对比容易产生错误,因此需要将指针进行同步后再进行对比. 传统同步模块同步的是读写指针,指针的同步和判断是一直进行的,同时为了避免毛刺的产生会使用格雷码转换单元,这就导致较多资源的占用,逻辑结构也会比较复杂. 为解决这一问题,所设计的 EB 采用半满状态的检查方式,即写指针只在 EB 半满处时才会触发读指针的检查操作,这就减少了资源的占用. 此外,EB 的写操作是一直领先于读操作,因此只需要进行写指针域到读指针域的同步操作. 在半满状态下时, $half_flg$ 有效,此时只需要将一位的 $half_flg$ 信号同步到读时钟域下,判断当前读指针的位置就可以确定当前 EB 的状态. 相对于同步指针的实现方法来说,这就减少了同步的位数,且由于不需要格雷码转换单元,逻辑结构也被简化.

6.3 双口 RAM 存储器

该弹性缓冲器的深度为 16,数据位宽扩展到 11

位,其中低 10 位存储数据信息,最高位存储信号 wr_data_valid 的值. 信号 wr_data_valid 在 CDR 实现了比特锁定和符号锁定后为 1^[5],PCS 层通过读取 EB 内部的值就可以判断读出的值是否有效.

6.4 读写指针控制和 SKP 检查模块

6.4.1 SKP 字符插入操作检测

SKP 字符的插入操作发生在 rd_clk 的频率大于 wr_clk 的情况下. 当写指针 wr_ptr 位于半满位置时, $half_flg$ 会被同步到 rd_clk 时钟域触发一次读指针检查. 如果 rd_ptr 指针位于图 8 中所示的位置,插入 SKP 字符的使能信号会有效,一旦检测到 SKP 有序集(在进行读操作时,SKP 有序集的检测是一直进行的), rd_ptr 指针进行加 1 操作(正常情况是加 2),这会导致 EB 中有一个 SKP 字符被多读了一次,由此向存储单元中插入了一个 SKP 字符. 如果插入的 SKP 字符位于 20 bit 数据的高 10 位, $add_skp[1:0] = 2'b10$; 反之, $add_skp[1:0] = 2'b01$. 如果 rd_ptr 的指针位于图中 7 或 8 的位置,则表示 EB 将被读空, $empty = 1$,此时不管有没有检测到 SKP 字符, rd_ptr 都将保持一个时钟周期的原值,避免 EB 一直处于被读空的状态.

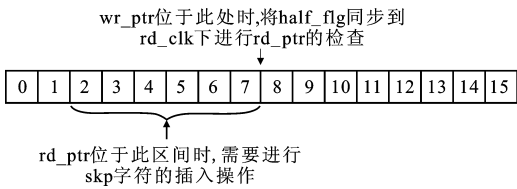


图 8 进行 SKP 字符插入操作下的指针状态

6.4.2 SKP 字符删除操作检测

SKP 字符的删除操作发生在 rd_clk 的频率小于 wr_clk 的情况下. 在进行读指针检查时,如果 rd_ptr 位于图 9 中所示区域,删除 SKP 字符的使能信号保持有效,一旦检测到 SKP 有序集(在进行读操作时,SKP 有序集的检测是一直进行的), rd_ptr 就会加 3,这就导致 EB 中有一个 SKP 字符没有被读取,从而达到删除 SKP 字符的目的. 如果删除的 SKP 字符位于 20 bit 数据的高位, $del_skp[1:0] = 2'b10$; 反之, $del_skp[1:0] = 2'b01$. 如果 rd_ptr 位于 14 或者 15 的位置时,则表示 EB 将被写满, $full = 1$,此时不管有没有检测到 SKP 字符, rd_ptr 指针将加 4,来避免 EB 持续溢出.

7 仿真结果与分析

本文采用 Verilog HDL 对 EB 进行设计,并采用

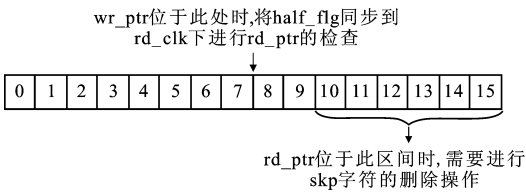


图 9 进行 SKP 字符删除操作下的指针状态

modelsim 软件做功能仿真验证,仿真结果如图 10~12 所示. 在图 10 中,当 EB 处于半满时,释放读指针下的指针复位信号,开始读操作(读写指针由于信号同步逻辑的存在,导致偏差大于 8);在图 11 中,由于信号同步逻辑大概需要两个时钟周期,因此 $half_flg$ 在 wr_ptr 为 6 和 8 时有效(延长信号有效时间,保证可靠触发). 触发读指针检查时,发现需要进行 SKP 字符的删除,因此在检测到 SKP 有序集时,删除了一个 SKP 字符,读指针加 3,同时 $del_skp = 2'b01$,表示删除的 SKP 字符位于 20 比特字符的低 10 位;在图 12 中,需要插入一个 SKP 字符,当检测到 SKP 字符时,读指针只加 1,同时 $add_skp = 2'b10$,表示删除的 SKP 字符位于 20 比特字符的高 10 位.

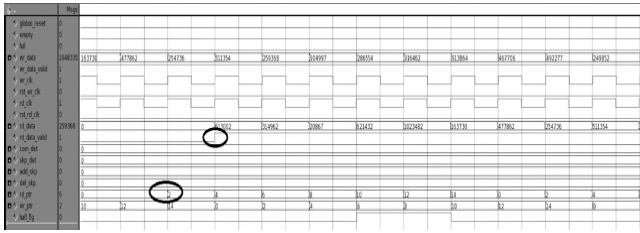


图 10 EB 进入半满状态下释放读指针

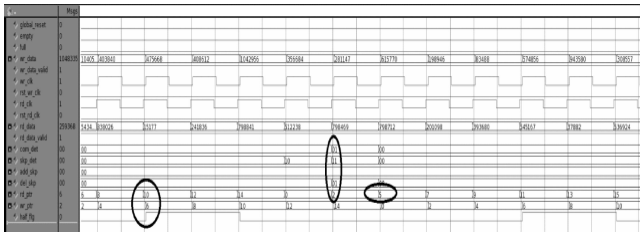


图 11 SKP 符号的删除操作

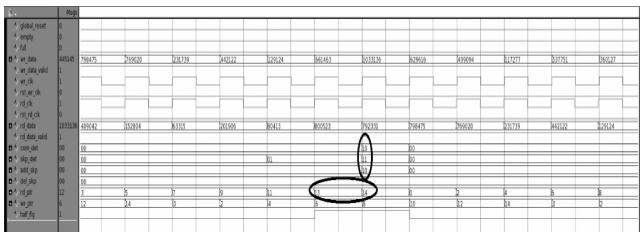


图 12 SKP 符号的插入操作

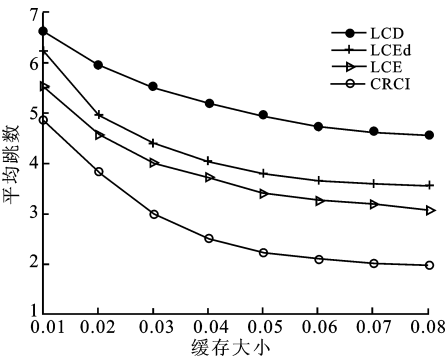


图 6 平均跳数 VS 缓存大小

制到边缘的速度很慢,边缘用户的请求跳数较大.本文提出的 CRCI 可以大大减少兴趣包转发的跳数.收到兴趣包后,如果存在匹配的 CIT 表项,则根据 CIT 转发兴趣包,找到最近的副本放置节点,这样就减少了转发兴趣包的跳数,大大减少了用户获取内容的延迟.

5 结束语

本文从网络内置缓存的可用性出发,提出了一种基于缓存索引的缓存路由机制.通过改进传统 CCN 路由器的转发引擎模块,添加 CIT 模块,详细地介绍了 CIT 表项的创建以及管理流程,并结合 CIT 提出一种基于缓存索引的缓存路由算法 CRCI.仿真实验证明,提出的基于缓存索引的缓存路由机制,可以减少服务器的命中率,减少兴趣包转发的平均跳数.

参考文献:

[1] Koponen T, Chawla M, Chun B-G, et al. A data-ori-

ented (and beyond) network architecture [C]//Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM). New York, NY, USA: ACM, 2007: 181-192.

[2] Jacobson V, Smetters D K, Thornton J D, et al. Networking named content [C]//Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies (CoNEXT). New York, NY, USA: ACM, 2009: 1-12.

[3] Sung-Won Lee, Dabin Kim, Young-Bae Ko, et al. Cache capacity-aware CCN: selective caching and cache-aware routing [C]// Global Communications Conference (GLOBECOM). Atlanta, GA: IEEE, 2013: 2114-2119.

[4] Psaras I, Chai WK, Pavlou G. Probabilistic in-network caching for information-centric networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(11): 2920-2931.

[5] L Zhang, D Estrin, J burke, et al, Named data networking (NDN) project, NDN0001 [R]. PARC, 2010.

[6] Munyoung Lee, Kideok Cho, Kunwoo Park, et al. SCAN: Scalable content routing for content-aware networking [C]//IEEE International Conference on Communications (ICC). Kyoto: IEEE, 2011: 1-5.

作者简介:

王好齐 男,(1991-),硕士研究生.研究方向为未来网络.
E-mail: hyhqwang@mail.ustc.edu.cn.

(上接第 54 页)

8 结束语

根据 PCIe 2.0 协议要求设计的 EB,读写时钟可以达到 250 MHz,完全符合协议的全速率和半速率要求.由此可见,基于指针的 SKP 字符的插入和删除操作同样可以完成时钟补偿的操作.

参考文献:

[1] PCI-SIG. PCI Express 2.0 Base Specification Revision 0.9. [EB/OL]. [2015-11-20]. www.pci-sig.com, 2011.

[2] 朱小明,王小利,程曾. USB3.0 物理层中弹性缓冲的设计

与实现[J]. 微电子学与计算机, 2012, 29(6): 117-118.

[3] Cummings CE. Simulation and synthesis techniques for asynchronous FIFO design [R]. SNUG, 2002.

[4] Joe Winkles. Elastic Buffer Implementations in PCI Express Devices [M]. US: MindShare Inc, 2003.

[5] Intel Corporation. PHY Interface for the PCI Express Architecture Version 2.0 [P]. 2007.

作者简介:

武桂林 女,(1990-),硕士研究生.研究方向为数字集成电路设计. E-mail: wgl157@mail.ustc.edu.cn.

黄鲁 男,(1961-),副教授,硕士生导师.研究方向为混合信号 IC 芯片设计.