

基于 HT 的光互连接口的设计^①

梁华岳^② 胡伟武

(计算机体系结构国家重点实验室中国科学院计算技术研究所 北京 100190)

(中国科学院大学 北京 100049)

(龙芯中科技术有限公司 北京 100190)

摘要 设计了一种应用于处理器光互连的光接口。该接口基于超传输(HT)协议设计,用光通路替代电通路实现处理器光互连。光接口用原 HT 接口逻辑结构,替换原来的输入输出电路的模拟电路部分,增加新的电路结构和光互连通路。该接口为互连接口是源同步结构的处理器应用光互连提供了一种通用的解决方案。整个互连结构兼顾电协议和光信号传输的特点,最大限度保留原接口的工作特点,同时发挥光信号的优势,以最小的代价和较高的效率实现处理器光互连。光接口处于互连的物理层,不需要修改 HT 接口的应用层,对处理器结构没有影响。

关键词 HT 光接口, 光互连, 片外互连, 数据转换

0 引言

提升处理器带宽是电路工程师一直努力的方向。提升带宽的方法通常有两种,一是提高单通道数据频率,二是增加并行通道数量。当前处理器互连电路带宽几乎达到电通路的极限,要进一步提升带宽比较困难:当前高速接口单通道数据频率达到 10Gbps^[1],而更高频率的电路设计十分复杂,总功耗巨大,并且传输距离越来越短;单个处理器面积达到 10cm²,引脚数超过 1000^[2],增加引脚数量也日益困难。面对这个问题,研究者一直在努力寻找新的解决方案,其中最引人注目的就是光互连(optical interconnection)方案。光通路具有高带宽、低损耗、抗干扰等优点,最适合作为高速信号的传输通道。随着近些年硅基光电子半导体技术的发展,光电器件面积已经接近集成电路的面积^[3],使得在处理器上集成光电器件成为可能。得益于半导体光电技术

的发展,利用绝缘层上硅(silicon on insulator, SOI)工艺可以实现驱动放大电路与光电器件集成在同一个硅衬底上^[4],这样就缩短了信号传输距离,解决了传统分离器件的传输线效应,简化了电路设计,极大提高了信号传输频率。

处理器应用光互连并不是简单地将光电器件和驱动电路连接到数据信号处理器上就可以实现的。处理器互连有专门的接口协议和电路结构,它们与处理器结构密切配合,实现整体性能最优。如果仅仅是在现有的接口上连接光通路,无法发挥光通路的优点。要充分发挥光互连的优势,一种可行的方案是专门为光互连设计一套互连接口协议,配合处理器结构和光通路,例如 Intel 为未来光互连设计了 ThunderBolt 接口。光互连专用接口与处理器体系结构密切配合,效率最优但系统设计十分复杂。在设计上需要考虑配合处理器结构和互连应用,还要考虑光通路的结构,综合平衡各方面因素。现有处理器有多种结构,应用的互连接口也不尽相同,其中

^① 国家科技重大专项(2009ZX01028-002-003, 2014ZX01020201, 2014ZX01030101)和国家自然科学基金(61133004, 61173001, 61232009, 61222204, 61432016)资助项目。

^② 男,1984 年生,博士;研究方向:计算机体系结构;联系人,E-mail: lianghuayue@ict.ac.cn
(收稿日期:2015-06-08)

一种高性能互连接口是超传输(hyper transport, HT)接口。它是一种高带宽、低延迟、点对点的传输技术,在处理器对处理器互连应用中性能尤其突出,是处理器互连的常用接口之一^[5]。本文基于 HT 接口协议设计了一种 HT 光互连接口。该接口技术是处理器光互连接口,支持处理器使用原有的接口协议,电路结构中部分替换原来的接口电路结构,对处理器结构没有影响,并且可以充分发挥光通路的性能优势。

1 HT 光互连接口结构

HT 接口的最新版本是 2009 年发布的 HT3.1。它的最高时钟频率是 3.2GHz,一组链路(link)最高带宽达到 51.2GB/s,工作模式分 AC 和 DC 两种,每组链路可拆分成两组半数链路。它具备支持热插拔,动态时钟/位宽调节,不同位宽链路、异步链路操作,时钟前驱(没有串并转换的延迟)等特点。HT 接口实现的目标是低延迟、高带宽的动态传输通路。处理器通过 HT 互连的结构如图 1 所示。光通路的一个主要特点就是高带宽,基于互补型金属氧化物半导体(CMOS)的光电器件工作频率一般在 10GHz 以上,III-V 族器件则更高^[6],因此单路光通路的数据频率是电通路的数倍以上。处理器应用 HT 接口与光接口互连结构如图 2 所示。从图中可以看出,光接口代替电接口中链路层,对于协议相关的物理层和逻辑层没有影响。

光通路带宽是电通路的数倍,需要将多路电信号通过一路光通路传输才能充分利用光通路的带宽。在光通讯应用中,为了保证传输信号在一定周期内是 DC 平衡的,光通路传输的信号都是经过编

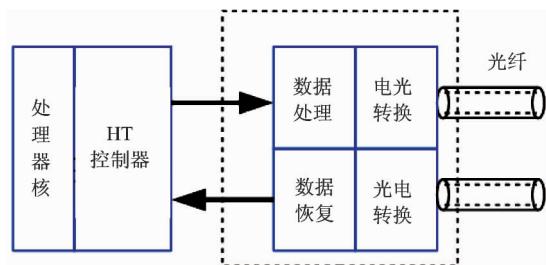


图 2 处理器 HT 光接口示意图

码的,原始的 HT 单通路数据也经过了编码,但并转串之后,数据编码被破坏。HT 使用的编码方式是 8b/10b 编码^[7],在 8b/10b 编码电路中,数据最长运行长度(在连续周期内数据不发生变化的数据位数)是 6,转换后(以四转一为例)是 24,平衡周期变长 4 倍,对接收端电路设计带来极大困难,因此需要重新编码。当 HT 接口数据位超过 8 位时,以 8 路数据位、1 路控制位和 1 路时钟构成一组通道,32 位数据用 4 组通道组合构成。HT 在工作时,4 组通道可以相互独立。当数据位少于 8 位时,每组通道可以只包含 2 位数据或 4 位数据。在片间互连应用中,数据交互量很大,带宽要求高,基本不会出现后两种配置模式,因此本文设计的光接口结构是以一组通道为单位对应一组光接口,实现时可以选择整组通道通过一路光通路或是 2、3 路光通路传输,传输结构配置成非源同步结构。

一组光接口结构如图 3 所示,与它交互的模块是 HT 接口中控制器的数字逻辑部分,输入输出的信号是数字信号,它替换的是原通路中的传输线及其发送端和接收端的驱动部分(图 1 中虚框部分)。这种结构完整保留了 HT 应用层的特性,改变的是传输层的电气特性。从图 3 中看出,光接口发送端有 4 个部分:编码模块、并转串模块、光驱动模块和电光转换器件。接收端有 5 个模块:光电转换器件、前置放大器、时钟数据恢复(clock and data recovery, CDR)、串转并模块和解码器。信号传输从发送到接收的过程是:多路数据经编码和并转串后转成单路 DC 平衡数据,单路数据需要经驱动电路放大才能被光调制器接收,从而生成相应的光信号;片间互连时,光信号通过 PCB 上的光波导进行传输;到达接收端,光信号通过光电二极管转换成小电流信号,然

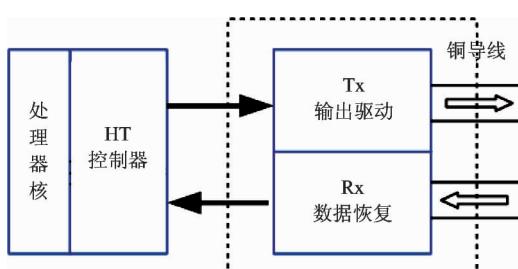


图 1 处理器 HT 接口示意图

后由前置放大器进行放大,放大后的信号由时钟数据恢复模块采样成数字信号,经过串转并,解码出原始数据。

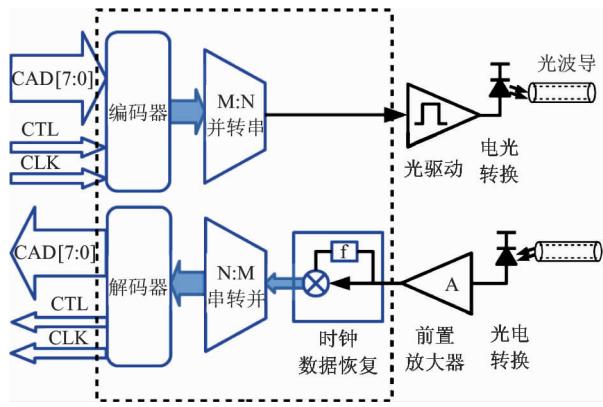


图 3 HT 光接口结构示意图

在高速片间互连时,传输线长度较短,一般在 15cm 以内,HT 的总线传输结构一般选择并行总线,时钟是源同步结构。用光通路传输时,本文不采用这种方式。光通路有完整的发送接收结构,不受限于 HT 物理层结构,可以不用保持相同的传输方式。源同步和非源同步的方法各有优缺点:采用非源同步的结构,当系统更改工作频率时,光接口调节响应要慢于原 HT 结构,并且需要系统给出控制信号;源同步结构能够保证原系统的响应时间,但需要单独一路光通路传输时钟,单路光通路带宽很高,这样严重浪费带宽和通路资源。在高速片间互连系统中,数据带宽需求很高,而初始化或动态调节的应用需求相对较低,同时经过串并转换后传输的数据频率是时钟频率的 4 倍以上,单独传输低速时钟严重浪费带宽,另一方面,HT 工作频率固定在几个频点附近,通过合理配置锁相环 (phase locked loop, PLL) 的分频器,可以保证时钟恢复电路能够在不同频点锁定,因此本文选择非源同步的时钟结构。

HT 接口特点之一是低延迟,它没有额外的数据转换的时间开销。为了保持这一特点,光接口电路没有数据缓存,只用两拍触发器同步。如果用触发器,每一级增加一个周期的延时,通常是电路传播延时的 10 倍以上。光电接口中,发送端数据编码后有

触发器同步一拍,接收端数据恢复后有高速触发器采样一拍,其它数字电路中数据通路上没有触发器(有状态触发器,不在数据通路上),保证整个通路的总延时最小。

本文设计的光接口单向可以是单通路也可以是双通路或三通路模式。采用单通路模式,8 路数据和 1 路控制位转成 1 路信号传输,传输频率至少是 HT 原数据率的 9 倍。采用三通路模式,传输频率至少是原数据率的 3 倍。HT3.1 最高数据率是 6.4Gbps,因此光通路数据率至少是 19.2Gbps。现在 CMOS 器件构成的光通路工作频率可以达到 10GHz 左右,正好匹配三通路模式,而且 CMOS 器件成本低,易于大规模电路集成。III-V 族器件的光通路则可以达到 40GHz 的工作频率,匹配单通路模式,但成本高昂,器件面积大。本文设计的多种通路模式可以匹配不同频率、成本和面积的光通路,方便系统结构选择。

2 接口电路设计

接口电路结构如图 4 和图 5 所示。通过前文讨论,一组 HT 光接口要传输 9 路数据,因此需要一种 9 到 n ($n >= 9$) 的并串转换方式。传统的 8b/10b 编码是应用在串行数据中,以每 8 位为一组。如果应用到这里,因为数据位数不是 8 的整数倍,需要缓存原始数据,增加数据编码延时,这样会破坏 HT 接口的低延迟特点,而且应用中串行数据频率很高,8b/10b 编码逻辑很难实现。如果不增加缓存,那就需要高频编码。编码时钟频率是原数据时钟的 1.125 倍,需要增加异步接口电路,同样增加延时。另外,8b/10b 编码生成的 10 位数据只能是通过 1 路通路传输,不具有灵活性。本文选用的编码方式是 3b/4b 编码。这种编码输入是 3 位,采用并行编码方式,正好三组构成 9 路并行数据,整体构成 9b/12b 编码。采用这种结构,编码不再需要缓存,编码时钟与原始数据时钟同步。3b/4b 编码还可以组合使用,既可以应用于三路独立通路,也可以应用于一路数据,十分灵活。

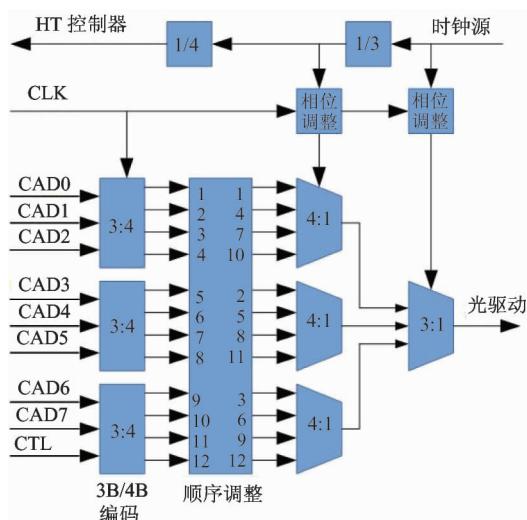


图 4 HT 光接口发送电路结构图

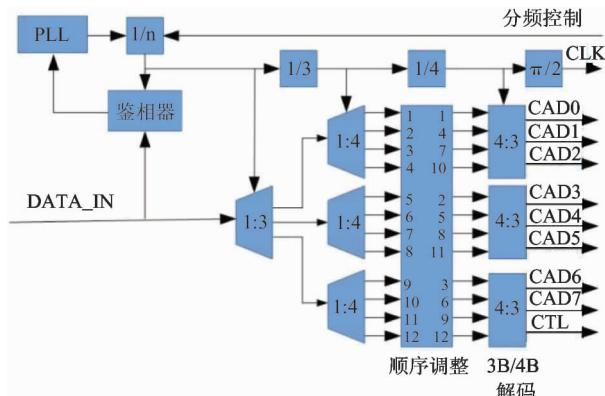


图 5 HT 光接口接收电路结构图

3b/4b 单独编码时, 时钟频率是原始数据的 4 倍, 使用树形结构 4 转 1 并转串电路可以满足要求。应用于一路数据时, 不能简单的用 4 转 1 加 3 转 1 实现, 而是有编码顺序的要求, 因为 3b/4b 编码单独一组 4 位码不是独立平衡的, 在配合使用时有顺序要求, 必须与编码时状态传递顺序一致。第一级 4 转 1 电路频率较低, 电路采用双沿采样的四路相位差时钟结构; 第二级 3 转 1 电路频率较高, 需要高速并转串电路。HT 数据传输采用双倍数据率结构, 时钟频率是数据频率的一半。这种结构降低系统对时钟的要求, 也有利于时钟传输。传统的高速并转串电路用两个时钟的相位差来选择数据^[8], 因为是双倍数据率, 用这样的结构需要 6 路相位差是 $1/3\pi$ 的时钟, 或者将时钟频率提高一倍, 与最终数据频率一致。这两种方法都极大增加了时钟电路设计难度, 并且它们的电路输出级是 6 路信号直接相连, 输出负载很大。结合 3 转 1 的特点, 本文设计的接口电路(图 6)仍然保持双倍数据频率特点, 利用 3 路双沿采样相位差时钟实现转换。它的特点是在两个时钟相位差生成的一个选择脉冲中输出两比特数据。图 7 是它的时序关系。三个时钟相位差是 $2\pi/3$, 而 C3 的下降沿与 C1、C2 的上升沿正好是 $\pi/3$ 的相位关系, 因此采样脉冲是 CK1 上升沿到 CK2 上升沿, 而数据选择利用 CK3 下降沿。电路结构如图 6 所示, 单元 1 的时序图见图 7, 高速并转串电路的输入数据通过 CK3 下降沿选择的两路数据, 输出级是 3 组单元相连。

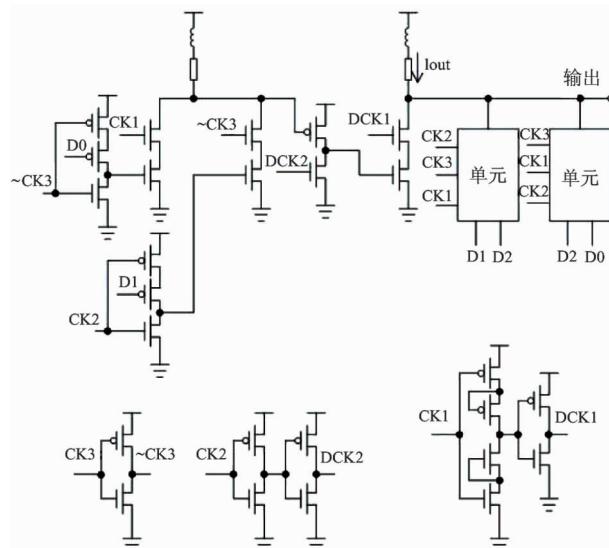


图 6 双沿采样 3 转 1 电路

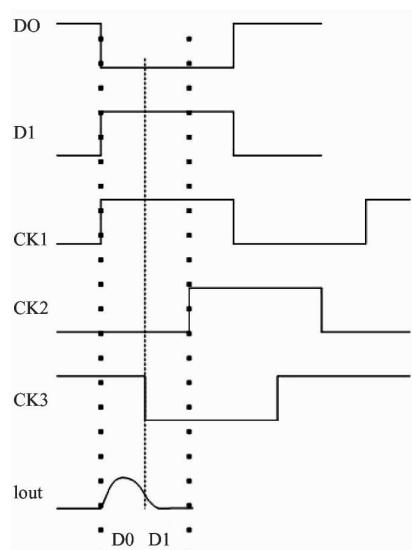


图 7 3 转 1 时序关系图

发送端原 HT 协议输出时钟和数据是 90 度相位差,以便于接收端数据采样。本文光接口电路中,在编码处所用时钟可以不需要相位差,只要保证 setup 和 hold,则原 HT 时钟调整电路就不再需要,直接用控制器的时钟做为编码时钟。从时钟源来的时钟经过控制器后会产生一定的延时,因此在发送端所用的高频时钟需要根据控制器时钟进行调整(图 4 中相位调整模块)。

经过并串转换之后的数据需要进行驱动放大才能进行电光转换。采用 SOI 工艺的光电技术中,驱动电路与光调制器是集成在一个硅衬底上,两者最长间距在毫米量级,传输延迟接近信号变化斜率,因此驱动输出级不需要考虑传输线效应,可以不用做阻抗匹配。驱动电路的负载是与电光器件 AC 耦合的电容。光电器件有多种选择,选择匹配带宽要求的即可。光信号通过光波导到达接收端。在接收端,光电二极管将光信号转换成电信号。光信号的转换、传输与接收的光电器件需要是采用 CMOS SOI 工艺实现集成,光波导则是在光 PCB 上实现(具体工艺问题不在本文讨论范围内)。因为前置放大器与光电二极管也是集成在同一硅衬底上,所以 TIA/LA 的设计也不用考虑输入静电释放(ESD)电容和阻抗匹配,直接使用薄栅晶体管,增加带宽增益积。先进制作工艺给光电集成和相应的电路设计带来的极大的便利,推动光互连应用的发展。

接收端时钟数据恢复(CDR)电路采用锁相环(PLL)结构,恢复时钟的频率是数据频率的一半。CDR 需要根据控制器接收端反馈的频率信号来快速动态调节时钟频率。HT 特点之一是在工作过程中会根据系统带宽要求动态改变工作频率,达到节省功耗的目的。要保持这一特点,时钟恢复电路要有很宽的工作频率范围。在初始化时,HT 从基频握手,握手成功后,可以设置更高的工作频率。HT 并不是工作在任意频率,而是在 200MHz 到 3.2GHz 之间取部分成 2 的整数倍的频率点。在光互连系统中,通过应用层软件设置,可以限定 HT 工作频率为 3.2GHz、1.6GHz、800MHz、200MHz 这几个固定频率上,这样 CDR 中的 PLL 通过几个固定的分频比就能满足频率要求,压控振荡器(VCO)的频率在一个较

小的范围内。HT 工作频率在调整前通过协议数据包发送相关指令到 HT 接收端,因而分频器的控制信号需要控制器接收端解析数据包后得到。

高速串转并电路一般用高速 latch 采样数据。因为是双倍数据频率,时钟上下沿都要采样数据。与前文并转串电路结构类似,在 1 转 3 的结构中利用时钟 $\pi/3$ 的相位差可以实现半速时钟采样。数据采样进行数据解码,将原码与时钟一起传输到原 HT 接收端。需要注意的是,数据与时钟需要有 90 度相位差,与 HT 协议要求保持一致。至此,HT 接收端能够正确接收到另一芯片上 HT 发送端发出的数据。

3 结 论

本文设计的光电接口适用于高性能处理器使用 HT 接口协议互连的应用情况。此接口结构设计充分考虑 SOI 工艺的研究趋势,面向光电器件与处理器集成的发展方向,是未来处理器 HT 接口光互连的一种选择。此接口可以用少于 3 路的光通路传输原来一组 10 路差分的 HT 通路,节省 IO 面积,增加互连带宽,并且接口电路延时较低。应用此接口,处理器不需要改变原来接口的逻辑电路结构,只需要增加转换电路,同时本接口不处理协议,是纯物理通路,处理器 HT 协议应用层完全不需要做修改,方便快捷。

参 考 文 献

- [1] Infiniband TA. FDR _ fact _ sheet. <https://ew.infinibandta.org/document/dl/7260>; Infiniband Trade Association, 2011
- [2] Li J, Cao L, Liu S, et al. Electrical design of high density low cost package for a switch asic. In: Electronic Packaging Technology High Density Packaging (icept-hdp), Xi'an, China, 2010. 548-551
- [3] Uemura T, Ishikawa Y, Nekado Y, et al. 1060-nm 10-gb/s × 12-channel parallel-optical modules for optical interconnects. In: Proceedings of the 2010 IEEE CPMT Symposium Japan, Tokyo, Japan, 2010. 1-4
- [4] 余金中. 硅基光电子学研究进展与趋势. 世界科技研

究与发展, 2007, 29(5): 50-56

- [5] Hu W, Zhang Y, Yang L, et al. Godson-3b1500: a 32nm 1.35ghz 40w 172.8gflops 8-core Processor. In: Solid-state Circuits Conference Digest of Technical Papers (isscc), San Francisco, California, USA, 2013. 54-55
- [6] Fedeli J, Ben bakir B, Vivien L, et al. Latest developments of 40g silicon photonics active devices. In: Optical Fiber Communication Conference and Exposition (ofc/

nfoec), Los Angeles, California, USA, 2012. 1-3

- [7] Widmer A, Franaszek P. A Dc-balanced, partitioned-block, 8b/10b transmission code. *IBM Journal of Research and Development*, 1983, 27(5): 440-451
- [8] Chen M, Yang C. A 50-64 Gb/s Serializing transmitter with a 4-tap, Lc-ladder-filter based Ffe in 65-nm CMOS. In: Custom Integrated Circuits Conference (cicc), San Jose, California, USA, 2014. 1-4

Design of an optical interconnection interface based on HT

Liang Huayue, Hu Weiwu

(State Key Laboratory of Computer Architecture, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100190)

(University of Chinese Academy of Sciences, Beijing 100049)
(Loongson Technology Corporation Limited, Beijing 100190)

Abstract

An optical interface, designed based on the Hyper Transport (HT) protocol for processor optical interconnection, is proposed. It connects processors with optical links instead of original electrical links, and uses the original HT interface logic structure to replace the analogy parts of the IO drivers with new digital circuits and optical links, thus offering a general solution for optical interconnection of source synchronized interface processors. The whole interface architecture considers both the electric protocol and optical signal features, and maximizes the operation features of the original interfaces and the advantages of typical optic signals to achieve processor optical interconnection with the least cost and the higher efficiency. This optical interface is in the physical layer of interconnection, and it need not to modify the application layer of the original HT interface. This does not bring any influence on the processor architecture.

Key words: HT optical interface, optical interconnection, off-chip interconnection, data transform