

# 一种用 HT 协议实现光纤传输的数据转换结构<sup>①</sup>

梁华岳<sup>②</sup> 胡伟武

(中国科学院计算技术研究所计算机体系结构国家重点实验室 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** 进行了将 Hyper Transport (HT) 协议用于处理器光纤互连系统的研究, 针对现有的处理器互连协议接口不能直接连接光纤传输系统的问题, 设计了一种可配置的数据转换结构。该结构通过特定的 3B/4B 编码模块和相应的并转串电路, 将原 HT 协议数据格式转换成适用于光纤互连的数据格式。该结构的特点是数据格式可配置, 编码简单, 电路可复用, 直流电流平衡性能高, 对传输协议透明和适于高速逻辑实现。该结构的提出解决了处理器光纤互连过程中遇到的并行总线多、单路速率低、电流不平衡等问题, 为处理器光纤互连提供了简捷的解决方案。采用此方案解决处理器光纤互连的数据格式问题尚属首次。

**关键词** 处理器, 光纤互连, 3B/4B 编码, 数据转换

## 0 引言

随着硅基光电子技术的发展, 处理器互连接口逐步采用光传输信号。处理器采用光纤互连有高带宽、低损耗、低延迟等优点。单路光信号传输数据速率远远高于电信号, 使用光纤互连可在提高带宽的同时减少输入输出接口。短距离光纤传输系统一般使用非源同步串行通信协议, 如 RapidIO、SERDES 等<sup>[1,2]</sup>, 而处理器互连系统一般使用源同步结构, 如 Hyper Transport (HT)<sup>[3]</sup>。HT 协议用于光纤传输面临很多困难, 如并行总线位宽过多、数据码率不适合、单通道协议速率过低等。现有的光互连研究集中在电路和光电器件上, 对这个问题, 没有可参考的解决方案, 因此本文提出了一种数据转换结构来解决这个问题。这种数据转换可将原 HT 数据转换成适于光纤系统的数据格式。这种方法可以不改变现有处理器逻辑结构, 部分替换原接口的物理结构。

## 1 结构特点

数据转换要完成两个任务, 一是调整并行数据通路个数, 二是保证生成的每个通道都保持 0/1 码率平衡。本结构采用并行实时编码, 根据编码调制发送时钟频率, 将 HT 数据实时发送。结构的两个主要模块是 3B/4B 编码模块和并转串电路。在高速传输系统中, 广泛使用的一种编码结构是 8B/10B 编码<sup>[4]</sup>, 它能够产生较好的平衡码, 并且能够区分指令和数据。在实现时, 它一般处于接口的低速逻辑电路中, 配合缓存和协议数据包, 实现数据连续发送。本文结构处于高速电路中, 面对的是连续的传输数据, 需要实时发送。在此情况下, HT 接口的 10 路数据很难配合 8B/10B 编码, 因此本文设计了 3B/4B 编码。新的编码结构针对 HT 接口特点设计, 逻辑简洁, 方便高速实现, 而且电流平衡性能与 8B/10B 相同, 还具有可以配置的特点, 可与并转串电路

<sup>①</sup> 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。

<sup>②</sup> 男, 1984 年生, 博士; 研究方向: 计算机体系结构; 通讯作者, E-mail: lianghuayue@ict.ac.cn  
(收稿日期: 2014-12-18)

结构配合实现数据实时传输。从协议结构看,本结构处于 HT 协议链路层,采用光信号传输与原电信号传输的链路没有区别。从电路结构看,本结构在接口的协议处理逻辑之后。它处理的数据一面是原 HT 接口传输的 10 路并行数据,另一面是光纤传输的数据。随着半导体光纤通信技术的发展,光通路与电通路的数据速率的比例会逐渐增加,因此本文的转换结构有三种配置方式,以适应不同的光纤传输性能需求。未来采用 HT 接口的处理器,在 HT 接口的数字逻辑模块之后增加这个转换电路结构,就可以生成适用于光纤传输的数据格式,能够直接连接光纤传输系统,实现处理器光纤互连。

## 2 数据转换需求

HT 协议要求接口单向通道 10 路一组,有 1 路时钟,1 路 CTL,8 路 CAD。图 1 是 HT 发送端物理层结构示意图。如果每路需要一路光信号,光信号发送端占用的接口数量没有减少,但光电器件面积比纯电驱动电路要大许多,浪费面积。另一方面,HT3.1 的单路 CAD 数据速率最高达到 3.2GHz,而 CMOS 光电器件的驱动电路工作频率超过 10GHz 甚至达到 20GHz<sup>[5,6]</sup>,光电器件本身频率甚至达到 40GHz<sup>[7]</sup>,极大浪费带宽。解决方法是将多路信号转换成单路信号,从而既提高了数据速率,又减少并行总线位宽。光纤系统中要求单个通道的数据 DC 平衡,一般是通过编码实现,HT3.1 协议带有这样的功能,但早期的版本 HT1.X 中没有这样的功能。如果没有经过 8B/10B 编码,每路信号本身的电流是不平衡的,转换后也不能保持平衡。即使经过编码的,转换后也达不到光纤传输的要求,因为原来单路数据中每 10 比特数据构成一组码变成现在几十比特数据构成一组码。以 4 转 1 为例,原数据采用 8B/10B 编码后,原每路数据中每 10 比特数据最多有 6 比特连续的 0 或 1,而生成的单路数据中会出现 24 比特连续的 0 或 1,这严重偏离 8B/10B 编码所要求的状态,造成接收电路严重偏离平衡状态,无法正确恢复数据。因此要实现多转一功能,就需要增加额外的编码模块。HT 协议的工作频率有多种

选择,光纤通路的结构设计也会影响光信号的传输速率,因此 HT 数据的转换方式也有多种,即可以通过多路光信号传输。

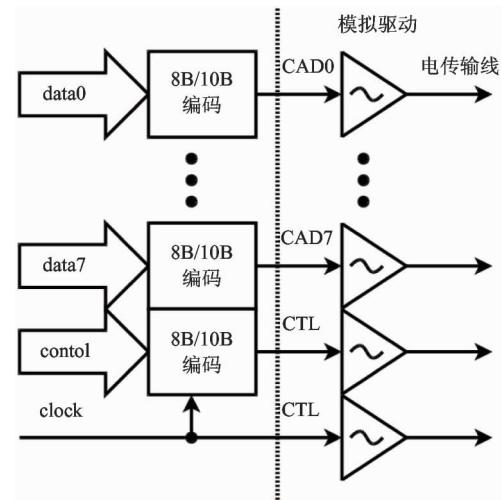


图 1 HT 输出端结构图

传统的 8B/10B 编码功能在同频时钟下完成,即编码前后数据的时钟频率相同。通路上每 8 拍的数据需要 10 拍才能发送。因为不能实时发送,所以在编码之前要有缓存电路。如果在本处数据转换中也采用同频编码,也需要缓存电路。HT 数据以数据包形式发送,如果转换电路不从中区分有效数据和无效数据,那缓存电路的规模需要无穷大;如果要区分,就需要解包模块和一定规模的缓存电路。如果这样设计就相当于用 HT 协议桥接了光传输协议进行处理器互连。这种方式电路开销大,传输延时长,是桥片的连接方式,不适用于处理器间直接互连。要实现转换电路对协议透明,一定需要非同频编码和实时数据传输。

HT 协议是并行源同步传输,而在光纤传输时,数据通常是非源同步的。这种差异不会带来额外的开销,反而有利于减少光纤传输信号的通路,因为采用光纤传输后,HT 的接收端不再需要恢复时钟并进行数据采样,而是由光接收端时钟数据恢复电路完成,HT 接收端逻辑处理的是恢复出来的数字信号。这样 HT 信号中的时钟信号可以不需要传输,从而会减少并行总线数量,降低并行信号对相对延时的要求。

### 3 转换结构设计

根据前文分析,HT 数据通道中有 9 路数据需要传输,光信号通路也可以是多路,数据转换结构是 9 转 M。不同的光纤通路设计就需要不同的数据转换结构。在平衡数据转换效率和电路复杂度的情况下,本文设计了可拆分组合的 3B/4B 编码电路,针对 M 的每个合适取值,提出统一的数据转换电路结构。本文设计数据转换方案的基本结构是 9B/12B 并行编码加分组并串转换,见图 2。编码通过 3 组 3B/4B 编码实现,并串转换根据光通路个数可以配置。每组 3B/4B 编码结构是一样的,但是状态计算

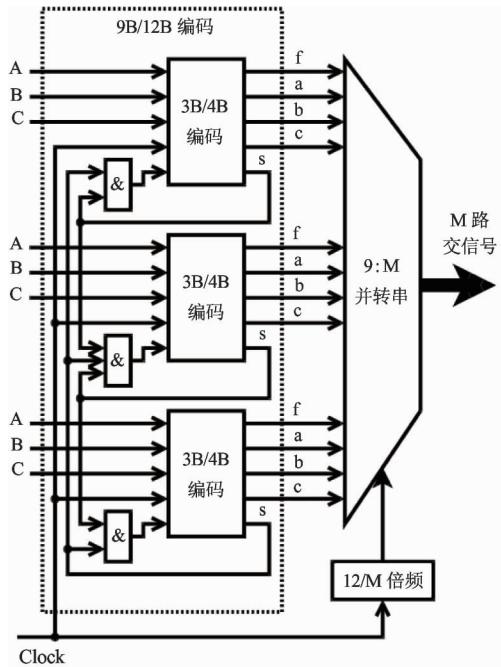


图 2 编码和并转串结构

不同。这种结构实现了编码电路复用,数据转换效率较高,电路结构最简,能实时数据传输。采用数据转换结构后,HT 发送端结构如图 3 所示。变化的部分是 HT 的驱动电路(图 1 中虚线往右的部分),对协议的物理层以上没有影响。原 10 路电信号采用模拟电路的方式通过铜导线传输,转换后生成 M 路电信号,通过数字驱动的方式发送到光电模块转成光信号,通过光纤传输。之所以可以采用数字信号发送,是因为光调制器与处理器通过硅上集成或封装集成的方式连接在一起,信号传输距离在 0.5cm

以内,可以忽略传输线效应。原 HT 数据传输距离要 7cm 以上,信号传输时要考虑传播延时,必须是模拟信号。本文设计的数据转换结构是经过各方面均衡后的选择,除此之外还有其它各种可行的方式,下文对各种可行结构的性能进行了比较,比较结果在第 5 节中列出。

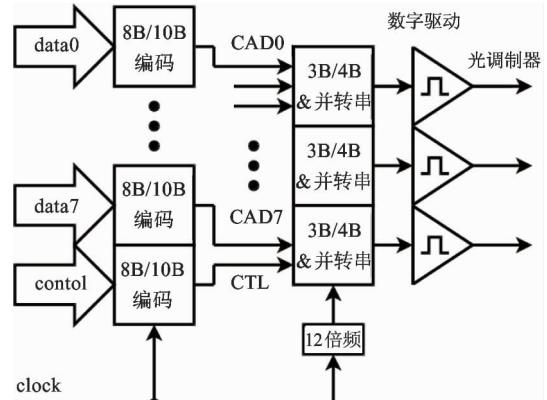


图 3 带转换结构的 HT 输出端

集成度最高的转换方式是将 9 路数据转成 1 路光信号。在这种转换方式下,最高效的编码方式是 9B/10B 编码,但是这种编码逻辑十分复杂,并且电流平衡能力不高,在高速逻辑下实现也是十分困难的。可行的一种编码方式是 9B/11B 编码。直接编码也是逻辑复杂,不利于高速实现。通过分组的方式分成 4B/5B 和 5B/6B 编码,在不考虑纠错和指令码的情况下,经过编码优化,可以较好地实现目标。缺点是需要设计两组配合使用的编码,而且只能在这种情况下使用。还可以再增加一位编码位,构成 9B/12B 编码。这种方式可以将原始数据分成 3 组,每组都是 3B/4B 编码。这种组合的优势是编码逻辑进一步得到简化,有利于高速实现。分 3 组数据还有一种结构是先进行 3 到 1 的并转串,然后进行 1 次 3B/4B 编码,最后通过 4 路输入的并转串电路,生成一路数据。这两者从电路结构上来看差别不大,数据转换效率也相同,本文选择前一种方式是为了与其它方式的结构保持一致。如果再增加编码位,编码效率更低,而且没有得到其它优势。

可选的一种转换方式是采用 2 路光信号传输 9 路数据。采用两路光信号,原始数据分成两组编码。如果两组编码不同,那么就需要两组不同的编码电

路、并转串电路,电路复杂度比较高。较好的编码方式是采用 4B/6B 和 5B/6B 编码。只要在 4 路输入数据的一组增加一路无效输入,这样的两组转换电路的结构就完全相同。因为电路的性能受限于 5B/6B 编码那一路,所以增加一路无效输入虽然浪费了一点效率,却降低一半电路复杂度。如果采取这种编码逻辑,需要专门设计一套编码,更简化的方式是能够复用 3B/4B 编码。从整体看,这样的编码也是 9B/12B 编码,与 1 路光信号传输在编码效率上相同。本文的结构是采用 3 组 3B/4B 编码生成 12 路数据,将生成的数据分成 2 组分别进行 6 转 1 的并串转换,最终生成两路数据。这种方式面临的一个问题是编码后的 12 路数据是 3 组编码,拆成两组数据有可能不是平衡码,因此本文设计了可拆分组合的 3B/4B 编码方式,以实现电路复用。

使用更多光通路的方式是采用 3 路光信号。这种方式自然将原始数据分成 3 组,3B/4B 编码是最优的选择,在电路结构上与之前两种相同。如果再增加光信号的通路,达到 4 路光信号,那么就无法复用 3B/4B 编码。因为如果复用编码,虽然 12 路可以分成 4 路传输,但每路数据都无法做到电流平衡。如果设计其它并行编码,只能是 2B/3B 或是 1B/2B 编码,这样的编码效率太低,浪费光通路带宽。如果不进行编码,那么数据通路中会出现 12 位连续 0 或 1,电流运行状态有较大偏移,接收端电路设计难度增加一倍,数据速率会有所降低。因此多于 3 路光信号的传输方式会带来效率降低、接收电路设计困难、传输速率下降等问题,不是合适的选择。

#### 4 3B/4B 编解码方法

一般的编码逻辑是完整的独立的一个模块,不可以拆分使用。本文设计的 3B/4B 编码可以独立作为一个编码模块,也可以用拆分成两组与另一个编码模块配合使用。可拆分再组合的目标是实现编码逻辑统一。表 1 是编码表。IN 列是输入数据,并行 3 路 ABC 为一组。根据编码不同,将输入数据组合分成 3 类 C1、C2、C3。选择状态列是根据通路上运行状态来决定生成的编码。选择状态有两位,前一位决定 fa 的编码,后一位决定 bc 的编码。fabc 列

表 1 数据编码表

IN	分类	选择状态(S0,S1)					
		*	,	*	,		
ABC	数据	fabc	NS	fabc	NS		
001	C1	1001	0,0	1001	0,0		
010	C1	1010	0,0	1010	0,0		
101	C1	0101	0,0	0101	0,0		
110	C1	0110	0,0	0110	0,0		
			*	,	-1		
					*	,	+1
100	C2	1011	0,+2	1000	0,-2		
011	C2	0111	0,+2	0100	0,-2		
			-1,*		+1,*		
000	C3	1110	+2,0	0010	-2,0		
111	C3	1101	+2,0	0001	-2,0		

表示并行 4 路数据,也是并转串后的顺序。NS 列代表经过这组编码后通路上的运行状态的变化,0 代表 01 平衡,状态不变,+2 代表状态增加两比特 1,-2 代表状态增加两比特 0。对 C1 类输入,不论什么状态,编码都一种,并且保持当前状态不变。其它两类输入编码时,需要根据两种选择状态才能确定编码。根据不同的光信号通路,选择不同分组的运行状态作为编码的选择状态。状态选择方法见表 2。第一列是光通路的个数。将转换电路的输入数据分成 3 组 T1、T2、T3,每组 3 路数据。ST\* 代表每路的当前状态是,ST\* 代表传输状态,是指光通路传输本小节数据后的运行状态,NS 列是小组的下一状态。图 4 是编码后数据序列图。状态选择表中每组的下一状态对应到光通路的运行状态。对于 3 路光通路,每路的编码是独立的,相互没有配合,电流独立平衡。双路光通路时,将 T2 组数据生成的编码拆分成前、后两小节,分别补充到 T1 和 T3 组后,构成 6 位一组的编码(T1 或 T3 组 4 比特编码是第一节,T2 组 2 比特编码是第二节)。光通路上的下一运行状态是传输了两节编码的状态,所以 T2 组前小节编码的选择状态是 T1 组的传输状态,后小节是 T3 组的传输状态,而 T1、T3 组的选择状态是 T1 组的当前状态,下一状态分别是 T2 组前、后小节的传输状态。T2 组的下一状态没有直接使用意义,因此不作计算和记录。单路的情况是状态循环,T1 的传输状态是 T2 的选择状态,T2 的传输状

态是 T3 的选择状态,T3 的传输状态是 T1 的下一状态,也是下一组数据中 T1 的选择状态。

表 2 状态选择表

转换方式	分组	状态选择		NS
		前	后	
三路	T1	ST1	ST1	ST1
	T2	ST2	ST2	ST2
	T3	ST3	ST3	ST3
双路	T1	ST1	ST1	ST20
	T2	ST1	ST3	
	T3	ST3	ST3	ST21
单路	T1	ST3	ST3	ST3
	T2	ST1	ST1	
	T3	ST2	ST2	

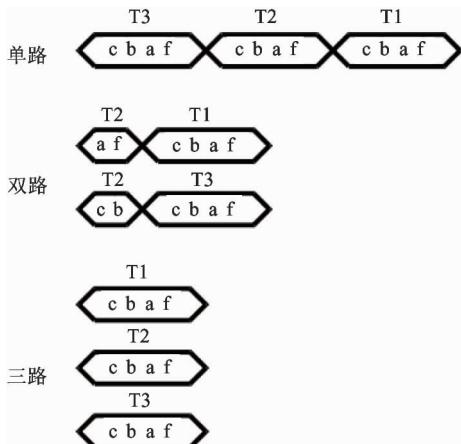


图 4 编码序列

## 5 性能分析

编码的一个重要目的是保持较小的电流偏移,本编码的运行状态图如图 5 所示。图 5(a)是单路和 3 路的状态,图 5(b)是双路的状态。横轴代表分组时间,0 是每组数据的起始点,每个点代表一比特数据。竖轴代表运行状态。横轴每单位时间间隔内,运行状态从低到高表示传输 1,从高到低表示传输 0。每组数据的起点或终点,运行状态都稳定在 +1 或 -1。在传输数据过程中,通路上运行状态为 2,表示在 +2 到 -2 之内,最长运行长度为 4。不管怎样组合,3 种方式的电流平衡能力相同。

表 3 列出前文提出的各种数据转换结构的各项指标。从表中看出,最高效的编码方式是 9B/11B

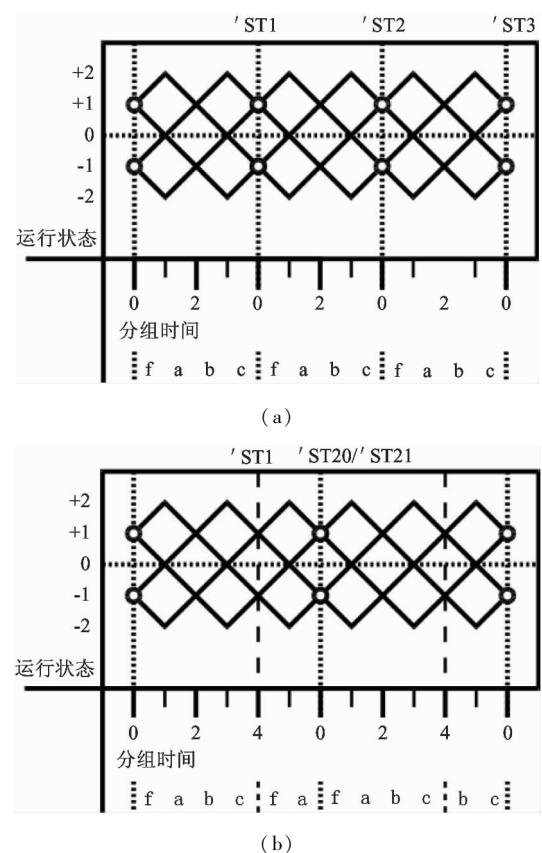


图 5 运行状态图

表 3 性能分析

结构 光信号 分组	编码 效率	时钟 倍频	运行 长度	运行 状态	电路 结构	复用
1 路 1 组	0.818	11	6	3	一般	否
1 路 3 组	0.75	12	4	2	简单	可
2 路 2 组	0.75	6	6	3	一般	否
2 路 3 组	0.75	6	4	2	简单	可
3 路 3 组	0.75	4	4	2	简单	可

编码效率:编码位数/原码位数;

时钟倍频:编码数据速率/原码数据速率;

运行长度:编码后数据中连续的 0 或 1 的个数;

运行状态:传输每组编码后电路的稳定状态

编码。如果只采用单路光信号传输数据,这是最好的选择。其它方式的编码效率都相同。运行长度和运行状态反映了电流平衡的能力,值越小说明电路状态越接近中间值,越有利于接收端恢复电路的设计。两组电流平衡能力较差的结构主要是受到 5B/6B 编码的限制。时钟倍频越低代表电路复杂度和设计难度越低,而电路规模越大(同样带宽下,通路增多)。电路结构的选择是多项指标平衡的结果,包

括系统结构、数据带宽、端口个数、电路复杂度、光电器件选择等因素。可复用的转换电路结构电路性能最好,兼顾编码效率,电路结构简单,降低电路设计成本,实现电路结构统一,是最优的选择。

## 6 结 论

针对用 HT 协议实现处理器光纤互连的设计要求,本文设计了可配置的数据转换结构,将原 HT 传输数据转换成适用于光通路传输的数据格式。本结构可配置成三种输出的数据转换结构,分别适用于单路、双路和三路三种光纤连接方式。本结构采用统一的可拆分配置的 3B/4B 编码方法,既能达到电流平衡能力最优,又能实现电路结构复用,降低设计复杂度,同时兼顾了编码效率。另外 3B/4B 编码逻辑简单,适用于在高速逻辑下实现。本结构用于替换处理器 HT 输出输入驱动模块,改变的是 HT 协议中的链路层,对处理器结构没有影响,可以直接使用。本文的工作为处理器光纤互连的逻辑设计提供了便利,为降低采用光纤互连的处理器的设计成本提供了技术和方法。

## 参考文献

- [1] 丁亮. 甚短距离光互连协议的研究及硬件实现:[硕士学位论文]. 成都:电子科技大学,2008
- [2] 千应庆,王晓锋,劳力等. 一种基于 RapidIO 协议的光纤总线硬件架构设计与分析. 兵工学报,2012,1(12): 1480-1484
- [3] Hu W, Zhang Y, Yang L, et al. Godson-3b1500: a 32nm 1.35ghz 40w 172.8gflops 8-core processor. In: Proceeding of 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers , San Francisco, USA, 2013, 54-55
- [4] Widmer A, Franaszek P. A DC-Balanced, Partitioned-block, 8b/10b Transmission Code. *IBM Journal of Research and Development*, 1983, 27(5): 440-451
- [5] Kim D W, Chi H K, Chun Y S, et al. 12.5-gb/s Analog Front-end of an Optical Transceiver in 0.13- $\mu$ m Cmos. In: Proceeding of 2013 IEEE International Symposium on Circuits and Systems, San Francisco, USA, 2013. 1115-1118
- [6] Proesel J, Schow C, Rylyakov A. 25gb/s 3.6pj/b and 15gb/s 1.37pj/b vcsel-based optical links in 90nm Cmos. In: Proceeding of 2012 IEEE International Solid-State Circuits Conference Digest of Technical Papers, San Francisco, USA, 2012. 418-420
- [7] Fedeli J, Ben bakir B, Vivien L, et al. Latest developments of 40g silicon photonics active devices. In: Proceeding of 2012 the National Fiber Optic Engineers Conference/Optical Fiber Communication Conference and Exposition, Los Angeles, USA, 2012. 1-3

## A data conversion structure for optical fiber transmission using hyper-transport protocol

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

The study of applying the hyper transport (HT) protocol to the processor systems using optical fiber interconnection was conducted, and a configurable data conversion structure was designed to solve the problem that the existing interfaces based on the processor interconnection protocol can not be connected directly to optical fiber transmission systems. This data conversion a structure converts the original HT protocol data to the format suitable for optical fiber transmission by using a special ad-hoc 3B/4B decoding module and the corresponding circuit for parallel-series conversion, with the features of data format configurable, coding simple, circuit reusable, well DC-Balance, transparent to the transmission protocol, and suitable for high frequency logic implementation. The design of the structure resolves the problems of huge amount of parallel buses, low data rate for single channel and unbalance of DC current during the optical fiber transmission, providing a state-of-the-art solution of data format conversion for the optical fiber interconnection of processor systems.

**Key words:** processor, optical fiber interconnection, 3B/4B coding, data conversion