

光学学报

基于 NRZ + Manchester 信号和偏振复用的无源光互连数据中心

邓鸿胜, 卢阳*, 曹露芳, 许蒙蒙, 毕美华

杭州电子科技大学通信工程学院, 浙江 杭州 310000

摘要 无源光互连数据中心由于其低成本、低功耗和高速率受到了广泛关注。在无源光互连数据中心中,同时可用的信道数目不能超过可用波长数,波长数量也限制了总吞吐量。为了减少光互连数据中心的波长需求、提高光互连数据中心的波长利用率与可扩展性,将 NRZ+Manchester 信号结合偏振复用应用于光互连数据中心,实现机架内 4 个服务器共享一条波长,从而将吞吐量提升 4 倍。通过仿真验证了该方案在 10 Gbit/s 速率下能够只使用 16 个波长实现机架内 64 个服务器的光互连,在 25 Gbit/s 速率下只使用 8 个波长就能够实现 32 个服务器的光互连。该方案能够节省 75% 的波长数量,降低设备成本,提高光互连数据中心的吞吐量。

关键词 光通信; 光互连数据中心; 偏振复用; 叠加信号; 无源光互连

中图分类号 O436

文献标志码 A

doi: 10.3788/AOS202141.1506001

Passive Optical Interconnection Data Center Based on NRZ + Manchester Signal and Polarization Multiplexing

Deng Hongsheng, Lu Yang*, Cao Lufang, Xu Mengmeng, Bi Meihua

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang 310000, China

Abstract Due to their low costs, low power consumption, and high speed, passive optical interconnection data centers have been widely concerned. In a passive optical interconnection data center, the channel number available at the same time cannot exceed the available wavelength number, and the wavelength number limits the total throughput. To reduce the wavelength requirements for optical interconnection data centers and improve the wavelength utilization and scalability of the optical interconnection data centers, we apply NRZ+Manchester signals combined with polarization multiplexing to the optical interconnection data centers in this paper. Four servers share one wavelength, which increases the throughput by four times. Simulation results show that the proposed scheme can realize the optical interconnection of 64 servers in the rack at the rate of 10 Gbit/s, where only 16 wavelengths are needed. At the rate of 25 Gbit/s, only eight wavelengths are required for the optical interconnection of 32 servers. This scheme can save 75% wavelengths, reduce equipment costs, and improve the throughput of optical interconnection data centers.

Key words optical communications; optical interconnection data center; polarization multiplexing; superimposed signal; passive optical interconnection

OCIS codes 060.4510; 060.4230; 060.1155; 060.4264

收稿日期: 2021-01-04; 修回日期: 2021-02-03; 录用日期: 2021-03-05

基金项目: 浙江省属高校基本科研业务费专项资金(GK209907299001-018)、杭州电子科技大学研究生科研创新基金(CXJJ202003)

通信作者: *luyang@hdu.edu.cn

1 引言

科技的发展带动了互联网的发展,如今,网络已成为生活中不可缺少的一部分。随着云计算等新型应用的飞速发展,近几年人们对于通信网络带宽和容量的需求快速增长^[1]。几乎所有的网络服务都在数据中心运行,快速的流量增长已经逐步逼近电带宽的上限^[2]。相较于电互连,光互连能耗更低,速率更高,短距离传输优势明显^[3],例如超大规模云服务数据中心的有源光纤速率能达到 400 Gbit/s^[4]。然而,有源光学元件成本高、重构速度慢,成为实现全光互连的障碍^[5]。无源光互连(POI)因其成本低、功耗小、速率快的特点,受到了广泛关注。

数据中心流量主要有机架内流量、机架间流量和数据中心间流量^[6]。其中机架内流量和机架间流量约占数据中心总流量的 77%^[7]。而机架内流量与内部总流量的比值在 20%~80%之间变化^[8],机架顶部交换机功率约占数据中心网络中所有类型交换机总功率的 90%以上^[9]。机架内流量、功耗不可忽视,如何减少机架内部功耗、提高传输效率成为了数据中心光互连中的重要课题。为了降低数据中心机架内部成本与功耗,文献[10]提出三种基于阵列波导光栅(AWG)或耦合器(coupler)的机架内 POI 架构并对这三种结构的可靠性与能耗进行了研究。文献[11]提出了一种机架顶部 POI 体系结构,以替代传统的电子交换机。

这些 POI 方案采用波分复用来增加数据中心的用户数以及吞吐量。但是,波长数量越多,这些 POI 方案对于每个服务器中的可调收发机的波长调节范围或者波长调节精度的要求也越高,这将大大提高服务器中收发器的成本。而且,可用的波长数目是有限的,例如,波长间隔为 0.8 nm(100 GHz grid)时,C 波段光源最多能划分为 43 组波长,这限制了机架内服务器的数量,因此文献[12]提出的数据中心光互连结构中单个机架内服务器仅有 16 个。当数据中心规模扩大、所需波长数目增加时,数据中心的成本也急剧增加,而且可用的波段资源也会很快被消耗殆尽。为了防止波长资源枯竭、降低波长过多带来的高成本,文献[13]利用物理层网络编码来提高数据中心 POI 网络的吞吐量。通过使用单波长信道,使一对节点之间以全双工方式同时进行相互通信,这节省了波长,频谱效率提高一倍。但是这种方式在两个节点速率不匹配时解调困难^[14],且对信号的同步要求较高^[15]。

本文提出一种新的数据中心机架内 POI 方案,将之前用于光标签交换的 NRZ + Manchester 信号^[16]应用于光互连数据中心,结合偏振复用,实现 4 个节点信号的同波长传输。不同于光标签方案中低速的 NRZ 标签信号,本文中 NRZ 与 Manchester 信号的有效比特率相同,通过延迟相加替代低通滤波解调 NRZ 信号,这种方法将信道数增加到波长数的 4 倍,大大提高了数据中心光互连的扩展性与吞吐量。

2 基本原理

NRZ + Manchester 信号的产生与接收原理如图 1 所示, NRZ 前后半比特电平相同[图 1(a)],而 Manchester 前后半比特电平相反[图 1(b)]。利用这一特性,在相同有效比特率的情况下,当对 NRZ + Manchester 信号进行延迟半比特相加时,由于 Manchester 前后半比特相加恒定为 1,而 NRZ 可能为 0 或 2,因此该操作能提取出 NRZ 信号;同理,当对 NRZ + Manchester 信号进行延迟半比特相减时,由于 Manchester 前后半比特相减为 1 或者 -1,而 NRZ 恒为 0,这样就能获得 Manchester 信号。

提出的数据中心机架内光互连结构如图 2 所示,将 n 个互连的服务器每 2 个为一对,划分为 $n/2$ 对。每对服务器中,一个服务器发射 NRZ 信号,另一个服务器发送 Manchester 信号(图 2 中的服务器 1 和服务器 2),这两个信号在电域叠加后被调制到一个光载波上,每对服务器共享一个光源。每两个服务器对(共 4 个服务器)通过偏振复用共用一个波长,组成一个服务器组(图 2 中的服务器对 1-2 和服务器对 3-4),每个服务器组共用一个波长。这些光信号进入耦合器并广播至其他服务器。在接收端,每个服务器先根据待接收信号所在的服务器组并通过滤波选出所需波长,然后根据待接收信号所在的服务器对调整偏振分束器,选出该信号所在的偏振态,最后根据所需信号的编码形式选择相应的接收方式(延迟半比特相加或者相减)。

在服务器内部,发射机发射 NRZ 格式或 Manchester 格式的电信号(图 2 中以 NRZ 为例),接收端配有可调滤波器、偏振分束器,以及延迟半比特相加和延迟半比特相减两种解调模式的接收机。当某个服务器需要接收来自服务器 1 的信号(图 2 中的服务器 3)时,先用可调滤波器进行滤波,选择波长 λ_1 ,再用偏振分束器进行偏振选择,选择偏振态 1,最后进行解调模式选择,选择延迟半比特相加以解调出服务器 1 的 NRZ 信号。当需要接收服务

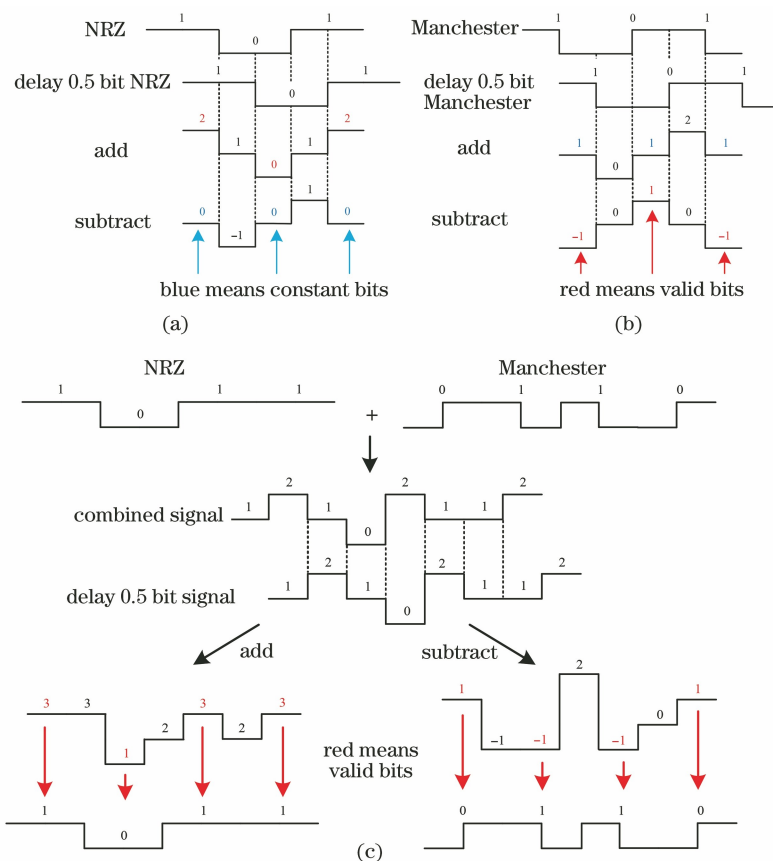
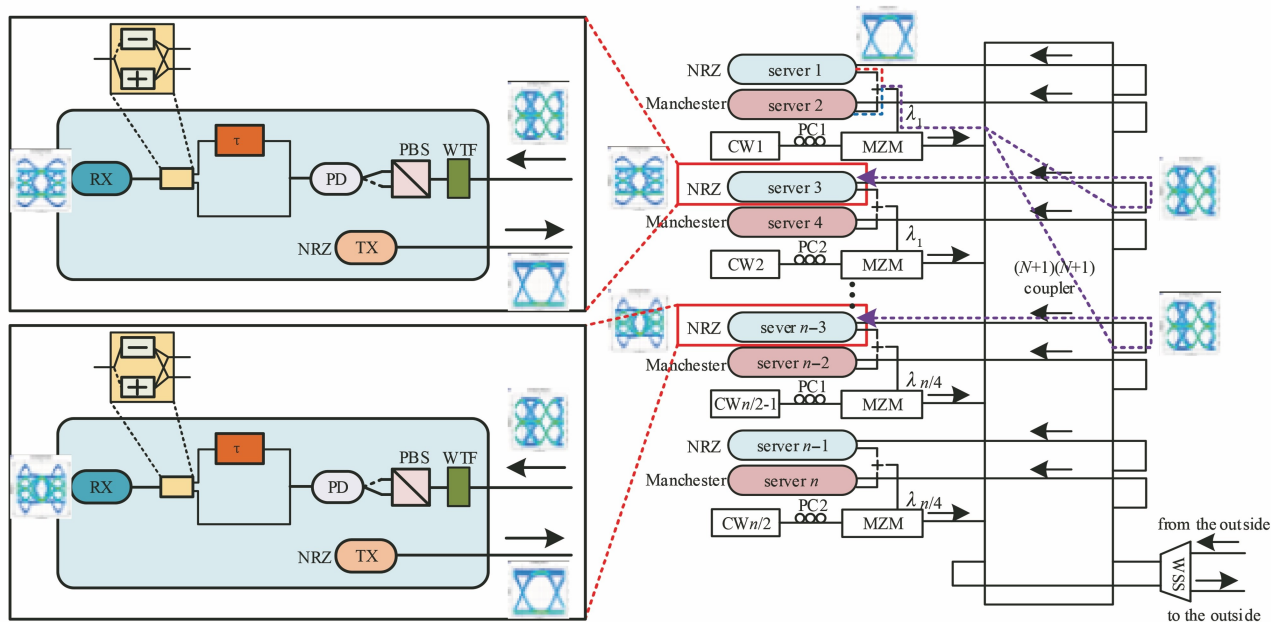


图 1 各信号码型图。(a) NRZ 延迟半比特操作；(b) Manchester 延迟半比特操作；(c) NRZ+Manchester 信号操作
 Fig. 1 Diagrams of signal patterns. (a) NRZ delayed 0.5 bit operation; (b) Manchester delayed 0.5 bit operation; (c) NRZ+Manchester signal operation



TX: transmitter; RX: receiver; WTF: wavelength tunable filter; PBS: polarizing beam splitter; PD: photodiode; τ : delayer; CW: continuous wave optical signal generator; PC: polarization controller; MZM: Mach-Zehnder modulator; WSS: wavelength selection switch

图 2 数据中心内光互连结构
 Fig. 2 Optical interconnection structure in data center

器 2 的信号(图 2 中的服务器 $n-3$)时,将波长选择为 λ_1 ,偏振态选择为偏振态 1,解调模式选择延迟半比特相减,从而解调出 Manchester 信号。

3 测试结果及分析讨论

使用 OptiSystem 软件对两个服务器组进行了仿真。目前数据中心比特率多为 1 Gbit/s 或 10 Gbit/s 并有向 25 Gbit/s 变化的趋势^[17],比特率设置为 10 Gbit/s。部分仿真设置如表 1、图 2 所示,

表 1 仿真参数

Table 1 Simulation parameters

Optical component	Extinction ratio of Mach-Zehnder modulator /dB	CW laser wavelength /THz	Azimuth of polarization controller /($^{\circ}$)	Optical filter frequency	Polarization filter
Server 1	14	193.1	0	Null	Null
Server 2	14	193.1	0	Null	Null
Server 3	14	193.1	90	193.1 THz	Polarization y
Server 4	14	193.1	90	Null	Null
Server $n-3$	14	193.3	0	193.1 THz	Polarization y
Server $n-2$	14	193.3	0	Null	Null
Server $n-1$	14	193.3	90	Null	Null
Server n	14	193.3	90	Null </td <td>Null</td>	Null

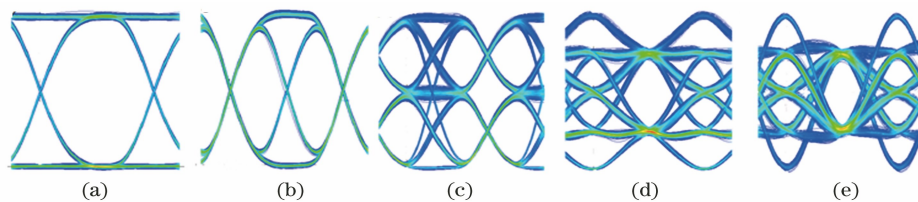


图 3 数据中心比特率为 10 Gbit/s 时的眼图。(a)服务器 1 发射的 NRZ 信号;(b)服务器 2 发射的 Manchester 信号;(c)同一服务器对叠加的 NRZ+Manchester 信号;(d)服务器 3 接收端解调的 NRZ 信号;(e)服务器 $n-3$ 接收端解调的 Manchester 信号

Fig. 3 Eye diagrams for data center bit rate of 10 Gbit/s. (a) NRZ signal from server 1; (b) Manchester signal from server 2; (c) superimposed NRZ+Manchester signal of the same server pair; (d) demodulated NRZ signal from receiving terminal of server 3; (e) demodulated Manchester signal from receiving terminal of server $n-3$

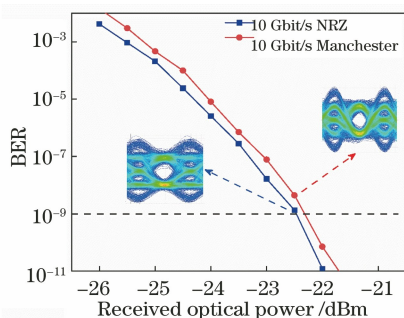


图 4 比特率为 10 Gbit/s 信号的接收光功率与误码率之间的关系(无偏振角度误差)

Fig. 4 Relationship between received optical power and bit error rate for signal with bit rate of 10 Gbit/s (without polarization angle error)

其中服务器 1-4 来自同一个服务器组,服务器 $(n-3)-n$ 为其他波长服务器组的干扰项。光发出后经过偏振控制,由 MZM 调制服务器 1,2 的 NRZ 与 Manchester 叠加信号,再与其他信号耦合。在服务器 3, $n-3$ 的接收端对耦合信号进行波长滤波、偏振控制和延迟半比特相加或相减,以解调出各自的信号。同一服务器对分别发射的信号、合路信号和不同解调方式的接收信号的眼图和解调信号误码率曲线如图 3、图 4 所示。

本文所采用的光互连数据中心总链路损耗可表示为

$$L = L_M + L_P + L_C + L_W, \quad (1)$$

式中: L_M 为调制器的插入损耗; L_P 为偏振器的插入损耗; L_C 为耦合器的插入损耗; L_W 为可调滤波器的插入损耗。结合数据中心光互连中元器件的插入损耗(表 2),发射机功率为 6 dBm 的情况下,根据(1)式计算得到的无偏振角度误差时耦合器损耗为 20.1 dB,耦合器端口数能够大于 64。耦合器端口数量决定了机架内服务器的数量,该方案在数据率为 10 Gbit/s 时能够支持 64 个服务器互连,保证了系统有足够的功率预算和稳定性。传统方案中需要使用 64 个波长,0.8 nm 波长间隔下波长范围超过

整个 C 波段光谱,同时所有服务器中的收发机需要具备 64 个波长的调节能力,技术挑战很大、成本很高。所提方案可以将光互连波长数减少为 16 个,波长数量大大减小,对服务器中可调收发机的调节范

围要求降低了,波长间隔可以更大,并且降低滤波要求。在使用相同波长的情况下,该方案可以增加服务器数量(4 倍),大大提高了可扩展性。

表 2 光元器件典型的插入损耗

Table 2 Typical insertion loss of optical components

Optical component	Mach-Zehnder modulator (MX-LN)	Polarization component (MeadowLark PBS)	Coupler	Wavelength tunable filter (FFP-TF2)
Insertion loss /dB	4	<2.5	3lb N	1.5

光在光纤中传输时会产生一定的偏振角度误差,为了测试该方案对偏振角度误差的忍受程度,首先测试了 10 Gbit/s 情况时不同偏振误差角度下接收光功率与误码率的关系。测试结果如图 5、图 6 所示。从图 5 中可以看出,在相同的接收光功率下,随着偏振角度误差的增大,延迟相加解调的 NRZ 信号眼图张开程度减小,而延迟相减解调的 Manchester 信号眼图变化不大。从图 6 的误码率

曲线中可以看出,延迟相加解调的 NRZ 信号对角度误差较为敏感,当角度误差为 2° 时,无误码传输的接收光功率约为 -21.5 dBm,耦合器损耗为 19.5 dB;当角度误差为 3° 时,该信号难以达到无误码传输;延迟相减解调的 Manchester 信号则不同,微小的偏振误差造成的误码率变化不大。经计算可得该方案在 2° 角度误差的损耗下仍支持 64 个服务器互连。偏振角度偏转主要由偏振光的两个正交分

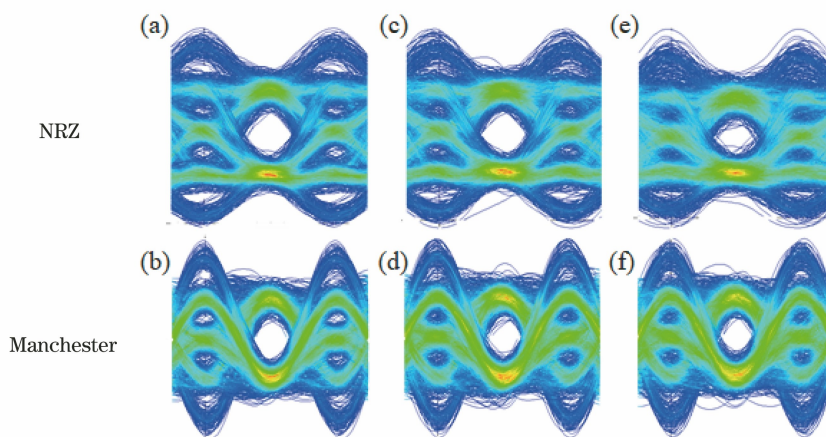


图 5 接收光功率为 -22 dBm 时不同偏振角度误差下解调的 NRZ 与 Manchester 信号眼图。

(a)(b)偏振角度误差为 1°;(c)(d)偏振角度误差为 2°;(e)(f)偏振角度误差为 3°

Fig. 5 Eye diagrams of demodulated NRZ and Manchester signals with different polarization angle errors at -22 dBm received optical power. (a)(b) Polarization angle error is 1°; (c)(d) polarization angle error is 2°; (e)(f) polarization angle error is 3°

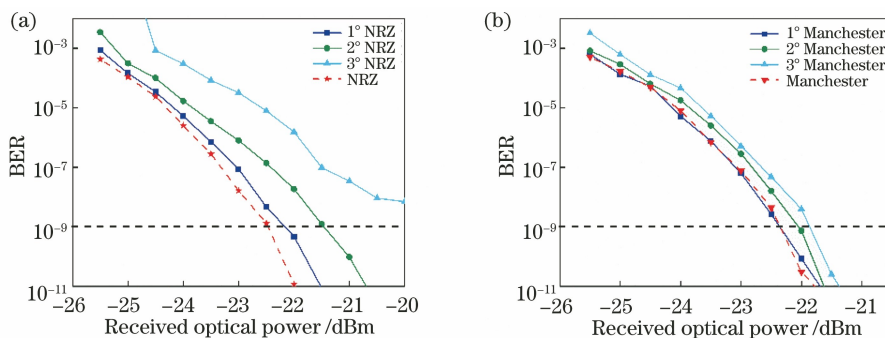


图 6 偏振角度误差对 10 Gbit/s 比特率下解调信号误码率的影响。(a)解调的 NRZ 信号;(b)解调的 Manchester 信号

Fig. 6 Influence of polarization angle error on BER of demodulated signals for 10 Gbit/s bit rate. (a) Demodulated NRZ signal; (b) demodulated Manchester signal

量改变引起的光纤双折射效应造成^[18],长距离传输中信号在光纤中随机双折射产生的偏振角度变化范围为 $2^{\circ}\sim 10^{\circ}$ ^[19],而数据中心机架内光纤长度一般只有几米到几十米,光纤产生的偏振误差角度较小,远小于 2° 。同时,传输距离较短时对光纤的损耗和成本要求不高,可以直接采用保偏光纤。因此,该方案在比特率为 10 Gbit/s 时能够忍受机架内光纤传输引起的偏振角度误差。

进一步测量了比特率为 25 Gbit/s 时的误码率,结果如图 7、图 8 所示。从图 7 中可以看出,比特率为 25 Gbit/s 信号的接收灵敏度提高到了 -19.6 dBm,并且当存在 2° 以内偏振角度误差时,两种解调信号的接收灵敏度都在 -19.1 dBm 以下(图 8)。经计算可得, 2° 偏振角度误差内该方案能够支持 32 个服务器的互连,所提方案仅需要 8 组波长就能完成光互连的波分复用结构,极大地

节省了波长资源,降低了对可调光源或者可调滤波器的调节范围或者波长精度的要求,且降低了设备成本。

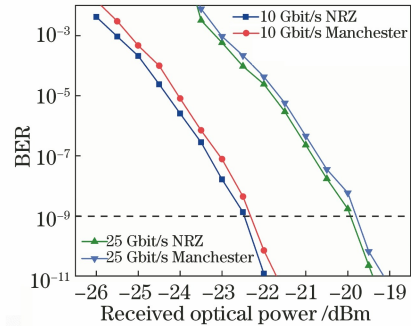


图 7 不同比特率下接收光功率与误码率的关系
(无偏振角度误差)

Fig. 7 Relationship between received optical power and BER for different bit rate (without polarization angle error)

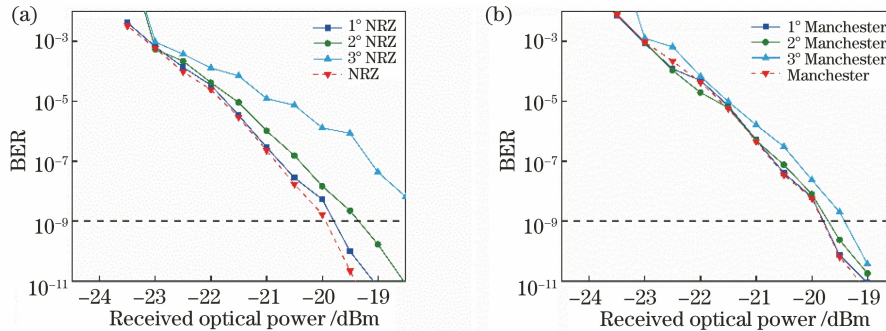


图 8 偏振角度误差对 25 Gbit/s 比特率下解调信号误码率的影响。(a)解调的 NRZ 信号;(b)解调的 Manchester 信号

Fig. 8 Influence of polarization angle error on BER of demodulated signals for 25 Gbit/s bit rate.

(a) Demodulated NRZ signal; (b) demodulated Manchester signal

4 结 论

提出一种 NRZ+Manchester 信号结合偏振复用的方案以解决光互连数据中心机架内波长数量少的问题。该方案利用 NRZ 和 Manchester 叠加信号延迟半比特相加减的结果不同以及偏振复用技术,实现 4 个信号同波长传输,以节省波长资源,并结合两种解调方式设计了一种透明、可切换的系统接收端。通过仿真测试了比特率为 10 Gbit/s 和 25 Gbit/s 以及不同偏振角度误差情况下的信号误码率。测试结果表明所提方案在 10 Gbit/s 比特率时能够支持 64 个服务器机架内的光互连并且波长仅需 16 个,在 25 Gbit/s 情况下能够支持 32 个服务器的光互连,且仅需 8 个波长。该方案极大地减少了波长数量,降低了设备成本,同时也可以增加光互连数据中心的用户数以及吞吐量。未来本课题组计

划通过实验验证该方案,并测试该方案接收端在较大负载下的数据解调延迟。

参 考 文 献

- [1] Li W P, Kong M, Yu J J. Generation of PDM-16QAM radio frequency signal based on a polarization multiplexing optical modulator [J]. Acta Optica Sinica, 2020, 40(23): 2306002.
李韦萍, 孔森, 余建军. 基于偏振复用光调制器产生 PDM-16QAM 射频信号 [J]. 光学学报, 2020, 40(23): 2306002.
- [2] Glesk I, Buis A, Davidson A. How photonic networking can help data centres [C] // 2014 16th International Conference on Transparent Optical Networks (ICTON), July 6-10, 2014, Graz, Austria. New York: IEEE Press, 2014: 1-4.
- [3] Li Q, Qiao Y J, Ji Y F. Comparison of parallel computing performance between the optical and

- electrical interconnects[J]. *Acta Optica Sinica*, 2010, 30(2): 498-502.
- 李倩, 乔耀军, 纪越峰. 光互连与电互连的并行运算性能比较[J]. *光学学报*, 2010, 30(2): 498-502.
- [4] Pi D, Shan Z H, Wu X K. Nanostructured antireflection micro-optics in the optical fiber communication band[J]. *Acta Optica Sinica*, 2020, 40(6): 0622002.
- 皮顿, 单子豪, 吴兴坤. 光纤通信波段微光学件的抗反射纳米结构[J]. *光学学报*, 2020, 40(6): 0622002.
- [5] Xu M T, Liu C, Subramaniam S. PODCA: a passive optical data center architecture [C] // 2016 IEEE International Conference on Communications (ICC), May 22-27, 2016, Kuala Lumpur, Malaysia. New York: IEEE Press, 2016: 1-6.
- [6] Gong Y, Lu Y, Hong X Z, et al. Passive optical interconnects at top of the rack for data center networks [C] // 2014 International Conference on Optical Network Design and Modeling, May 19-22, 2014, Stockholm, Sweden. New York: IEEE Press, 2014: 78-83.
- [7] Zhang F, Zhu Y X. High-speed transmission technologies for data center optical interconnection [J]. *ZTE Technology Journal*, 2019, 25(5): 17-24.
- 张帆, 朱逸萧. 面向数据中心光互连的高速光传输技术[J]. *中兴通讯技术*, 2019, 25(5): 17-24.
- [8] Kachris C, Tomkos I. A survey on optical interconnects for data centers [J]. *IEEE Communications Surveys & Tutorials*, 2012, 14(4): 1021-1036.
- [9] Pries R, Jarschel M, Schlosser D, et al. Power consumption analysis of data center architectures[M]//RodriguesLiang J J P C, Zhou L, Chen M, et al. Green communications and networking. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering. Berlin: Springer, 2012, 51: 114-124.
- [10] Chen J J, Gong Y, Fiorani M, et al. Optical interconnects at the top of the rack for energy-efficient data centers [J]. *IEEE Communications Magazine*, 2015, 53(8): 140-148.
- [11] Cheng Y X, Fiorani M, Lin R, et al. POTORI: a passive optical top-of-rack interconnect architecture for data centers[J]. *IEEE/OSA Journal of Optical Communications and Networking*, 2017, 9(5): 401-411.
- [12] Yuang M, Tien P L, Ruan W Z, et al. OPTUNS: optical intra-data center network architecture and prototype testbed for a 5G edge cloud [J]. *IEEE/OSA Journal of Optical Communications and Networking*, 2020, 12(1): A28-A37.
- [13] Lin R, Lu Y, Pang X D, et al. First experimental demonstration of physical-layer network coding in PAM4 system for passive optical interconnects[C]// 2017 European Conference on Optical Communication (ECOC), September 17-21, 2017, Gothenburg, Sweden. New York: IEEE Press, 2017: 1-3.
- [14] Feng N, Sun X. Implementation of network-coding approach for improving the BER performance in non-orthogonal multiple access (NOMA)-PON [J]. *Optics Communications*, 2020, 462: 125301.
- [15] Zhao M F, Zhou Y J, Yuan Q, et al. Research survey on physical layer network coding[J]. *Journal of Computer Applications*, 2011, 31(8): 2015-2020, 2028.
- 赵明峰, 周亚建, 原泉, 等. 物理层网络编码研究进展[J]. *计算机应用*, 2011, 31(8): 2015-2020, 2028.
- [16] Lu Y, Wang Y Y, Hu L H, et al. Optical label switching based on Manchester code + NRZ modulation[J]. *Optical Fiber Technology*, 2021, 61: 102387.
- [17] Yin X, van Kerrebrouck J, Coudyzer G, et al. Multi-level high speed burst-mode receivers[C]// 2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS), July 3-7, 2016, Niigata, Japan. New York: IEEE Press, 2016: 1-3.
- [18] Huang Y, Zhao J Y, Wang J D, et al. A real-time polarization compensation system based on wavelength-division multiplexing for optical fiber communication systems [J]. *Acta Optica Sinica*, 2020, 40(14): 1406003.
- 黄媛, 赵家钰, 王金东, 等. 一种基于波分复用的实时光纤信道偏振补偿系统[J]. *光学学报*, 2020, 40(14): 1406003.
- [19] Xu L, Chi N, Oxenlowe L K, et al. A new orthogonal labeling scheme based on a 40-Gb/s DPSK payload and a 2.5-Gb/s PolSK label [J]. *IEEE Photonics Technology Letters*, 2005, 17(12): 2772-2774.