Research Article | Vol. 27, No. 22 / 28 October 2019 / Optics Express 32210
Optics EXPRESS

Adaptive moment estimation for polynomial nonlinear equalizer in PAM8-based optical interconnects

JI ZHOU,[1,8] HAIDE WANG,[1] JINLONG WEI,[2] LONG LIU,[1] XINCHENG HUANG,[3] SHECHENG GAO,[1] WEIPING LIU,[1] JIANPING LI,[1] CHANGYUAN YU,[4] AND ZHAOHUI LI[5,6,7]

[1]*Department of Electronic Engineering, College of Information Science and Technology, Jinan University, Guangzhou 510632, China*
[2]*Huawei Technologies Duesseldorf GmbH, European Research Center, Germany*
[3]*Department of electronic and information, Huzhou University, Huzhou 313000, China*
[4]*Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hongkong, China*
[5]*State Key Laboratory of Optoelectronic Materials and Technologies, School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China*
[6]*Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China*
[7]*lzhh88@mail.sysu.edu.cn*
[8]*zhouji@jnu.edu.cn*

Abstract: Adaptive moment estimation (Adam) is a popular optimization method to estimate large-scale parameters in neural networks. This paper proposes the first use of Adam algorithm to fast and stably converge large-scale tap coefficients of polynomial nonlinear equalizer (PNLE) for 129-Gbit/s PAM8-based optical interconnects. PNLE is one of simplified Volterra nonlinear equalizer for making a trade-off between complexity and performance. Different from serial least-mean square (LMS) adaptive algorithm, Adam algorithm is a parallel processing algorithm, which can obtain globally optimal tap coefficients without being trapped in locally optimal tap coefficients. Timing error is one of the main obstacles to the PAM systems with high baud rate and high modulation order. Owing to parallel processing and global optimization, Adam algorithm has much better performance on resisting the timing error, which can achieve faster, more-stable and lower-MSE convergence compared to LMS adaptive algorithm. In conclusion, Adam algorithm shows great potential for converging the tap coefficients of PNLE in PAM8-based optical interconnects.

© 2019 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

In recent years, development of data center drives demand of optical interconnects with data rate up to 400 Gbit/s. By 2020, the optical interconnects are expected to reach data rate of 800 Gbit/s or 1 Tbit/s [1–3]. To transmit high-capacity signal on limited bandwidth, multi-level modulations have been widely investigated in research and commercial fields. 4-level pulse-amplitude modulation (PAM4) has been standardized by IEEE P802.3bs 400-Gb/s Ethernet Task Force for short-reach optical interconnects [4–6]. 4×100-Gbit/s optical PAM4 system is a favored option to achieve 400-Gbit/s interface for short-reach optical interconnects. To further improve spectral efficiency, 8-level pulse-amplitude modulation (PAM8) is gradually adopted [7,8]. However, compared to optical PAM4 system, optical PAM8 system requires higher signal-to-noise ratio and is more sensitive to inter-symbol interference (ISI) and nonlinear distortions.

In PAM-based optical interconnects, some simplified Volterra nonlinear equalizers (VNLEs) have been widely applied to compensate nonlinear distortions, mainly including the polynomial nonlinear equalizer (PNLE) and sparse VNLE [9,10]. Least-mean square (LMS) adaptive

#376472 https://doi.org/10.1364/OE.27.032210
Journal © 2019 Received 10 Sep 2019; revised 8 Oct 2019; accepted 8 Oct 2019; published 21 Oct 2019

Optics EXPRESS

algorithm is one of the most popular algorithms to converge the tap coefficients of VNLE. However, with increase of data rate and modulation order, LMS-based VNLE requires large number of taps and training symbols. Meanwhile, the convergence of LMS-based VNLE is slow, unstable and with high mean square error (MSE) [11,12]. Nowadays, neural network is one of the most popular spots for academic researches and industry applications. In optical transmissions, neural networks have been applied to compensate some complicated distortions [13–15]. In short-reach optical interconnects, the distortion models are almost certain. Neural networks have relatively high computational complexity for compensating the distortions with certain models, which is not applicable in short-reach optical interconnects due to the stringent requirement on power. However, the optimization algorithms in neural networks are remarkably efficient for minimizing or maximizing objective function, which has the potential to effectively minimize the error function of traditional equalizer.

In this paper, for the first time, we propose the first use of adaptive moment estimation (Adam), a popular optimization algorithm used in neural networks, in the PNLE for realizing fast and stable convergence. Compared to the neural networks, the Adam-based PNLE has simpler structure and lower computational complexity. Different from serial LMS adaptive algorithm, Adam algorithm is a parallel processing algorithm, which can obtain globally optimal tap coefficients without being trapped in locally optimal tap coefficients. 129-Gbit/s PAM8-based optical interconnects have been experimentally demonstrated for verify the feasibility of Adam-based PNLE. Owing to parallel processing and global optimization, Adam algorithm has much better performance on resisting the timing error, which can achieve faster, more-stable and lower-MSE convergence compared to LMS adaptive algorithm.

2. Principle of Adam-based PNLE

In this section, the principle of Adam-based PNLE is introduced. The output of three-order VNLE can be expressed as

$$s_n = \sum_{k=0}^{N-1} h_k^{(1)} x_{n-k} + \sum_{k=0}^{N-1}\sum_{l=0}^{k} h_{k,l}^{(2)} x_{n-k}x_{n-l} + \sum_{k=0}^{N-1}\sum_{l=0}^{k}\sum_{m=0}^{l} h_{k,l,m}^{(3)} x_{n-k}x_{n-l}x_{n-m} \tag{1}$$

where $h_k^{(1)}$, $h_{k,l}^{(2)}$ and $h_{k,l,m}^{(3)}$ are the tap coefficients of 1st, 2nd and 3rd terms in VNLE, respectively. As a simplified version of VNLE, three-order PNLE has been proposed to compensate the nonlinear distortions in order to make a trade-off between computational complexity and performance. The output of the three-order PNLE is expressed as

$$s_n = \sum_{k=0}^{N-1} h_k^{(1)} x_{n-k} + h_k^{(2)} x_{n-k}^2 + h_k^{(3)} x_{n-k}^3 \tag{2}$$

where N is the tap number of each term in the PNLE. Linear feed-forward equalizer (LFFE) is a special case of PNLE with only linear term.

Adam-based PNLE requires M training samples to update the tap coefficients in the training process. At training processing of Adam-based VNLE, the training samples $\boldsymbol{x}$ are received and stored to construct a training matrix $\boldsymbol{R}$ for parallel processing. The structure of the training matrix $\boldsymbol{R}$ can be expressed as

$$\boldsymbol{R} = \begin{bmatrix} x_{N-1} & x_{N-2} & \cdots & x_0 \\ x_N & x_{N-1} & \cdots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ x_{M-1} & x_{M-2} & \cdots & x_{M-N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{N-1} \\ \boldsymbol{r}_N \\ \vdots \\ \boldsymbol{r}_{M-1} \end{bmatrix}. \tag{3}$$

Optics EXPRESS

Obviously, the dimension of $\boldsymbol{R}$ is $(M - N + 1)$-by-N where M is the length of training samples, which should be larger than N. Therefore, the elementwise second-order training matrix $\boldsymbol{R}^2$ and the third-order training matrix $\boldsymbol{R}^3$ can be derived from $\boldsymbol{R}$. The desired training vector is

$$\boldsymbol{Y} = [\ y_{N-1} \quad y_N \quad \dots \quad y_{M-1}\ ]^T \tag{4}$$

where $(\cdot)^T$ denotes matrix transpose. The error function of Adam-based PNLE can be calculated by

$$J(\boldsymbol{H}) = \frac{1}{M-N+1}(\boldsymbol{R}\boldsymbol{H}^{(1)} + \boldsymbol{R}^2\boldsymbol{H}^{(2)} + \boldsymbol{R}^3\boldsymbol{H}^{(3)} - \boldsymbol{Y})^T(\boldsymbol{R}\boldsymbol{H}^{(1)} + \boldsymbol{R}^2\boldsymbol{H}^{(2)} + \boldsymbol{R}^3\boldsymbol{H}^{(3)} - \boldsymbol{Y}) \tag{5}$$

where $\boldsymbol{H}^{(i)} = [\ h_0^{(i)} \quad h_1^{(i)} \quad \dots \quad h_{N-1}^{(i)}\ ]^T$ is the tap coefficient vector of i-th term in the PNLE. Gradient $\boldsymbol{G}^{(i)}$ is the partial derivative of $J(\boldsymbol{H})$ with respect to $\boldsymbol{H}^{(i)}$, which can be calculated as

$$\boldsymbol{G}^{(i)} = \frac{2}{M-N+1}(\boldsymbol{R}^i)^T(\boldsymbol{R}\boldsymbol{H}^{(1)} + \boldsymbol{R}^2\boldsymbol{H}^{(2)} + \boldsymbol{R}^3\boldsymbol{H}^{(3)} - \boldsymbol{Y}). \tag{6}$$

Conventional LMS adaptive algorithm for PNLE updates $\boldsymbol{H}^{(i)}$ using one training sample every iteration, which can be expressed as

$$\boldsymbol{H}_t^{(i)} = \boldsymbol{H}_{t-1}^{(i)} - \theta^{(i)} \times (\boldsymbol{r}_{t-1}^i)^T(s_{t-1} - y_{t-1}) \tag{7}$$

where $\theta^{(i)}$ are fixed step sizes ranging from 0 to 1 and subscript t denotes t-th iteration. Different step sizes are necessary for the polynomial coefficients with different terms to obtain fast and stable convergence. Generally speaking, when the step size is too large, the LMS adaptive algorithm may fail to converge or even diverge; but it requires a large number of iterations when the step size is too small. Different from LMS adaptive algorithm, Adam algorithm is much less sensitive to the step size θ for the reason that it computes adaptive step sizes from estimates of biased first and second moments of gradients [16]. The biased first and second moment estimates $\mathbf{m}_t$ and $\mathbf{v}_t$ of $\boldsymbol{G}_t$ are initialized as zeros vector, which can be expressed as

$$\mathbf{m}_t^{(i)} = \beta_1 \times \mathbf{m}_{t-1}^{(i)} + (1 - \beta_1) \times \boldsymbol{G}_t^{(i)}, \tag{8}$$

$$\mathbf{v}_t^{(i)} = \beta_2 \times \mathbf{v}_{t-1}^{(i)} + (1 - \beta_2) \times (\boldsymbol{G}_t^{(i)})^2 \tag{9}$$

where β_1 and β_2 are set to 0.9 and 0.999, respectively. The bias-corrected operations keep the biased first and second moment estimates from moving towards zeros at the beginning of iterations, which can be expressed as

$$\hat{\mathbf{m}}_t^{(i)} = \mathbf{m}_t^{(i)} / \left(1 - \beta_1^t\right), \tag{10}$$

$$\hat{\mathbf{v}}_t^{(i)} = \mathbf{v}_t^{(i)} / \left(1 - \beta_2^t\right). \tag{11}$$

A relative small value ϵ is used to prevent zero-division error and the tap coefficients can be updated as

$$\boldsymbol{H}_t^{(i)} = \boldsymbol{H}_{t-1}^{(i)} - \theta^{(i)} \times \hat{\mathbf{m}}_t / (\sqrt{\hat{\mathbf{v}}_t^{(i)}} + \epsilon). \tag{12}$$

It is worth noting that Adam algorithm is employed only at the convergence stage in Adam-based PNLE. Then, it will be switched into decision-directed LMS-based PNLE algorithm soon after convergence. Therefore, it harnesses both fast and stable convergence and low computational complexity, which is ideal for short-reach optical interconnects.

Optics EXPRESS

3. Experimental setups and results

Figure 1(a) shows block diagram of 129-Gbit/s optical PAM8 system using Adam-based PNLE for short-reach optical interconnects. Only Gray coding is used to generate the digital PAM8 frame by using MATLAB. The generated digital PAM8 frame is resampled to 2 samples per symbol and then uploaded into digital-to-analog converter (DAC) with 8-bit resolution, 86-GSa/s sampling rate and 16-GHz 3-dB bandwidth. Thus, the generated electrical PAM8 signal has a baud rate of 43 Gbaud. The length of training sequence is set to 1000 and the length of payload is set to 81240. Therefore, by considering a hard-decision FEC with 7% overhead, net rate of the generated electrical PAM8 signal is approximately 119 Gbit/s (43 GSa/s $\times$ 3 bit/Sa $\times$ 81240/82240/(1 + 7%) $\approx$ 119 Gbit/s). After a linear electrical amplifier, a 40-Gbit/s electro-absorption integrated laser modulator (Opnext LE7B60) is used to modulate the amplified electrical signal on a continuous wave optical carrier at 1550 nm. The generated optical PAM8 signal is fed into 2-km standard single-mode fiber (SSMF).

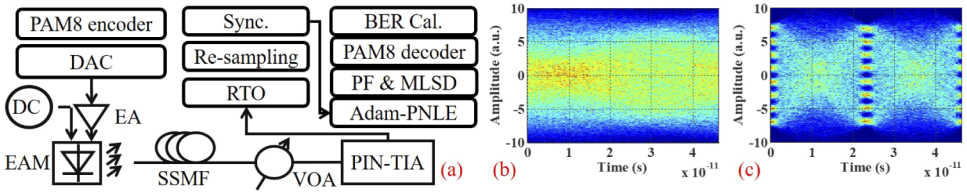

Fig. 1. (a) Block diagram of 129-Gbit/s optical PAM8 system using Adam-based PNLE for short-reach optical interconnects, (b) Eye diagrams of PAM8 signal before Adam-based PNLE, (c) Eye diagrams of PAM8 signal after Adam-based PNLE.

At the receiver, a variable optical attenuator (VOA) is used to adjust received optical power (ROP). The optical signal is converted into an electrical signal by a 20-GHz PIN-TIA (DSC-R401HG). The electrical signal is fed into 80-GSa/s real-time oscilloscope (RTO) with 3-dB bandwidth of 36 GHz to implement analog-to-digital conversion. The digital PAM8 signal is decoded by offline processing, including re-sampling, synchronization, Adam-based PNLE, post filter (PF), maximum likelihood sequence detection (MLSD), PAM8 symbol-to-bit de-mapper and BER calculation. It is worth noting that Adam algorithm is employed for only converging the tap coefficients at the training processing and decision-directed LMS adaptive algorithm is used to update the tap coefficients after the training processing. Figures 1(b) and 1(c) reveal eye diagrams of PAM8 signal before and after Adam-based PNLE, respectively. The eye diagrams depict that Adam-based PNLE can effectively compensate serious distortions of the received PAM8 signal. The post filter is used to eliminate the enhanced in-band noise, which is defined as $z(k) = r(k) + \alpha \times r(k-1)$. MLSD is employed to recover the transmitted signal $x(k)$ from the $z(k)$, which is equivalent to minimize $\sum_k \{z(k) - [x(k) + \alpha \times x(k-1)]\}^2$.

Figure 2(a) shows BER versus ROP for 129-Gbit/s optical PAM8 system using T-spaced Adam-based PNLE and LFFE after optical BTB transmission and 2-km SSMF transmission. The tap numbers of linear, square and cubic terms are set to (97, 97, 97) in PNLE, respectively. The tap number of LFFE is set to 291, which is the same with the total tap number of PNLE. The length of training symbols is set to 1000. When T-spaced Adam-based PNLE is employed, the required ROP for the BER of 10^{-3} after 2-km SSMF transmission is approximately 1 dB higher than that after optical BTB transmission. Meanwhile, compared to T-spaced Adam-based LFFE, T-spaced Adam-based PNLE can achieve approximately 1-dB improvement of receiver sensitivity.

Figure 2(b) reveals BER versus ROP for 129-Gbit/s optical PAM8 system after 2-km SSMF transmission using T-spaced Adam-based PNLE, T-spaced LMS-based PNLE and $T/2$-spaced LMS-based PNLE. The length of training samples is set to 1000. The tap numbers of linear,

Optics EXPRESS

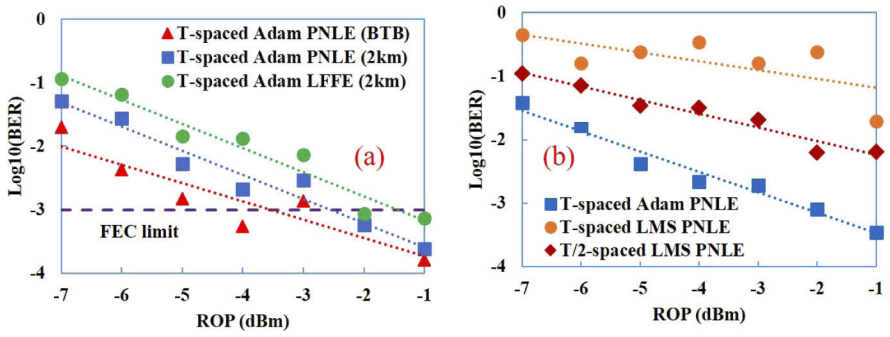

Fig. 2. (a) BER versus ROP for 129-Gbit/s optical PAM8 system using T-spaced Adam-based PNLE and LFFE after optical BTB transmission and 2-km SSMF transmission and (b) BER versus ROP for 129-Gbit/s optical PAM8 system after 2-km SSMF transmission using T-spaced Adam-based PNLE, T-spaced LMS-based PNLE and $T/2$-spaced LMS-based PNLE when the length of training samples is set to 1000.

square and cubic terms are set to (97, 97, 97) in PNLE, respectively. Due to the timing error, T-spaced LMS-based PNLE has poor performance and cannot achieve the FEC limit. In general, $T/2$-spaced structure should be employed in LMS-based PNLE to resist the timing error. Therefore, $T/2$-spaced LMS-based PNLE has better performance than the T-spaced one. Owing to globally optimization of Adam algorithm, T-spaced Adam-based PNLE has an ability to resist the timing error. The experimental results show that T-spaced Adam-based PNLE achieves much better performance than $T/2$-spaced LMS-based PNLE.

Figure 3 reveals MSE against iterations for (a) T-spaced Adam-based PNLE, (b) $T/2$-spaced LMS-based PNLE and (c) T-spaced LMS-based PNLE. The ROP is set to -1 dBm after 2-km SSMF transmission. In Adam-based PNLE, the β_1, β_2, ϵ, $\theta^{(1)}$, $\theta^{(2)}$, and $\theta^{(3)}$ are set to 0.9, 0.999, 10^{-8}, 0.06, 0.0005, and 0.00001, respectively. The length of training samples is set to 1000 and tap number of linear, square and cubic terms is set to (97, 97, 97) in PNLE. The taps of T-spaced PNLE can be converged by Adam algorithm after only 100 iterations. T-spaced LMS-based PNLE cannot be converged due to the influence of timing error. For resisting timing error, $T/2$-spaced structure is usually employed in LMS-based PNLE. Compared to $T/2$-spaced LMS-based PNLE, the convergence of T-spaced Adam-based PNLE is faster and more stable. Figure 3 proves the experimental results shown in Fig. 2.

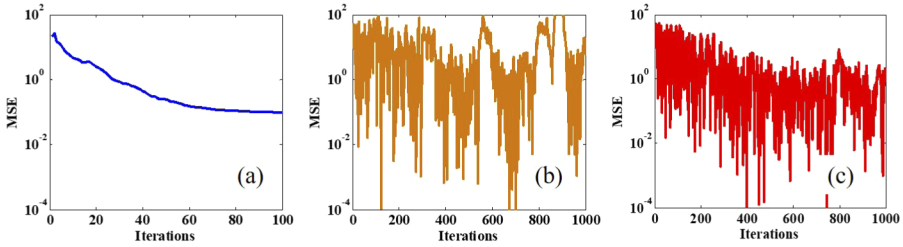

Fig. 3. MSE against iterations for (a) T-spaced Adam-based PNLE, (b) T-spaced LMS-based PNLE and (c) $T/2$-spaced LMS-based PNLE. The ROP is set to -1 dBm after 2-km SSMF transmission.

Figure 4 shows the $\mathrm{Log}_{10}BER$ contour plot versus length of training symbols and tap number for 129-Gbit/s optical PAM8 system using (a) T-spaced Adam-based PNLE and (b) $T/2$-spaced LMS-based PNLE. The ROP is set to -1 dBm after 2-km SSMF transmission. When the length of

Optics EXPRESS

training symbols is set to 2000 and tap number is set to (117, 117, 117) in T-spaced Adam-based PNLE, the $\mathrm{Log}_{10}BER$ can reach to less than −3.8. However, under the same number of taps and training symbols, $\mathrm{Log}_{10}BER$ can reach to higher than −2.8 when T-spaced LMS-based PNLE. Therefore, the BER of 129-Gbit/s PAM8-based optical interconnects using T-spaced Adam-based PNLE is one-level lower than that using $T/2$-spaced LMS-based PNLE. For achieving the FEC limit (i.e., $\mathrm{Log}_{10}BER$ of −3), the length of training symbols is set to 5000 or 3000 when the tap number is set to (57, 57, 57) or (77, 77, 77) in $T/2$-spaced LMS-based PNLE. In T-spaced Adam-based PNLE, the length of training symbols is set to 500 when the tap number is set to (37, 37, 37). Thus, the required training symbols and taps for T-spaced Adam-based PNLE are much less than that for $T/2$-spaced LMS-based PNLE.

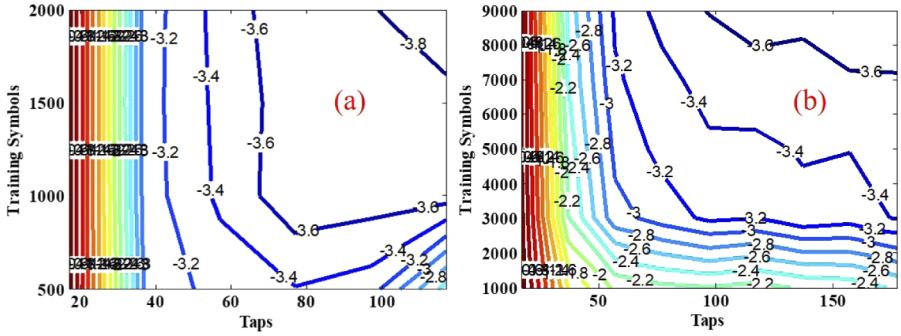

Fig. 4. $\mathrm{Log}_{10}BER$ contour plot versus length of training symbols and tap number for 129-Gbit/s optical PAM8 system using (a) T-spaced Adam-based PNLE and (b) $T/2$-spaced LMS-based PNLE.

The computational complexity of PNLE can be calculated as

$$\mathrm{C}_{\mathrm{Adam}} = [2(M_T - N + 1) + 9] \times 3N \times I + 3M_P \times (2N + 1), \tag{13}$$

$$\mathrm{C}_{\mathrm{LMS}} = 3M \times (2N + 1) \tag{14}$$

where M_T and M_P are the length of training sequences and payload in T-spaced Adam-based PNLE, respectively. N is the number of taps and I is the iteration number of Adam algorithm. M is equal to M_T plus M_P. When the length of training symbols is set to 5000 and the tap number is set to (57, 57, 57), the computational complexity of $T/2$-spaced LMS-based PNLE is approximately 3×10^7. When the length of training symbols is set to 500, I is set to 100 and the tap number is set to (37, 37, 37), the computational complexity of T-spaced Adam-based PNLE is approximately 3×10^7. Therefore, T-spaced Adam-based PNLE has almost the same computational complexity with the $T/2$-spaced LMS-based PNLE for the same performance.

4. Conclusion

In this paper, for the first time, Adam algorithm is used to efficiently calculate the large-scale tap coefficients of PNLE for 129-Gbit/s PAM8-based optical interconnects. Compared to Adam-based LFFE, Adam-based PNLE can achieve approximately 1-dB improvement of receiver sensitivity. Different from serial LMS adaptive algorithm, Adam algorithm is a parallel processing algorithm, which can obtain globally optimal tap coefficients without being trapped in locally optimal tap coefficients. Therefore, T-spaced Adam-based PNLE has an ability to resist the timing error. However, $T/2$-spaced structure is usually employed in LMS-based PNLE. Under the same number of training symbols and taps, T-spaced Adam-based PNLE has better performance than $T/2$-spaced LMS-based PNLE. For reaching the same BER, the required training symbols and taps

Optics EXPRESS

for T-spaced Adam-based PNLE is less than that for $T/2$-spaced LMS-based PNLE. Meanwhile, T-spaced Adam-based PNLE has almost the same computational complexity with $T/2$-spaced LMS-based PNLE. In conclusion, Adam algorithm shows great potential for converging the tap coefficients of PNLE in PAM-based optical interconnects.

Funding

National Key R&D Program of China (2018YFB1801704); Science and Technology Planning Project of Guangdong Province (2017B010123005, 2018B010114002); Local Innovation and Research Teams Project of Guangdong Pearl River Talents Program (2017BT01X121); National Natural Science Foundation of China (61525502, 61975242); Fundamental Research Funds for the Central Universities (21619309).

Disclosures

The authors declare no conflicts of interest.

References

1. J. Wei, L. Zhang, C. Prodaniuc, N. Stojanović, and C. Xie, "Linear Pre-Equalization Techniques for Short Reach Single Lambda 225 Gb/s PAM IMDD Systems," in *European Conference on Optical Communication* (IEEE, 2018), pp. 1–3.
2. X. Pang, O. Ozolins, L. Zhang, A. Udalcovs, R. Lin, R. Schatz, U. Westergren, S. Xiao, W. Hu, G. Jacobsen, S. Popov, and J. Chen, "Beyond 200 Gbps per Lane Intensity Modulation Direct Detection (IM/DD) Transmissions for Optical Interconnects: Challenges and Recent Developments," in *Optical Fiber Communication Conference* (Optical Society of America, 2019), paper W4I.7.
3. Y. Zhou, K. Smith, P. Weir, M. Gilson, J. Chen, W. Pan, Y. Chang, S. Wu, and S. Wu, "Field Trial of 400G Single-Carrier Ultra-Efficient 1.2Tb/s Superchannel Over 250 km," IEEE Photonics Technol. Lett. **29**(17), 1451–1454 (2017).
4. E. E. Fiky, M. Chagnon, M. Sowailem, A. Samani, M. M. Osman, and D. V. Plant, "168-Gb/s Single Carrier PAM4 Transmission for Intra-Data Center Optical Interconnects," IEEE Photonics Technol. Lett. **29**(3), 314–317 (2017).
5. F. Li, D. Zou, L. Ding, Y. Sun, J. Li, Q. Sui, L. Li, X. Yi, and Z. Li, "100 Gbit/s PAM4 signal transmission and reception for 2-km interconnect with adaptive notch filter for narrowband interference," Opt. Express **26**(18), 24066–24074 (2018).
6. N. Eiselt, D. Muench, A. Dochhan, H. Griesser, M. Eiselt, J. Olmos, I. Monroy, and J. Elbers, "Performance Comparison of 112-Gb/s DMT, Nyquist PAM4, and Partial-Response PAM4 for Future 5G Ethernet-Based Fronthaul Architecture," J. Lightwave Technol. **36**(10), 1807–1814 (2018).
7. J. Zhang, P. Gou, M. Kong, K. Fang, J. Xiao, Q. Zhang, X. Xin, and J. Yu, "PAM-8 IM/DD Transmission Based on Modified Lookup Table Nonlinear Predistortion," IEEE Photonics J. **10**(3), 7903709 (2018).
8. J. Shi, J. Zhang, N. Chi, and J. Yu, "Comparison of 100G PAM-8, CAP-64 and DFT-S OFDM with a bandwidth-limited direct-detection receiver," Opt. Express **25**(26), 32254–32262 (2017).
9. Z. Wan, J. Li, L. Shu, S. Fu, Y. Fan, F. Yin, Y. Zhou, Y. Dai, and K. Xu, "64-Gb/s SSB-PAM4 Transmission Over 120-km Dispersion-Uncompensated SSMF With Blind Nonlinear Equalization, Adaptive Noise-Whitening Postfilter and MLSD," J. Lightwave Technol. **35**(23), 5193–5200 (2017).
10. N. Diamantopoulos, H. Nishi, W. Kobayashi, K. Takeda, T. Kakitsuka, and S. Matsuo, "On the Complexity Reduction of the Second-Order Volterra Nonlinear Equalizer for IM/DD Systems," J. Lightwave Technol. **37**(4), 1214–1224 (2019).
11. K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, and A. P. T. Lau, "Digital Signal Processing for Short-Reach Optical Communications: A Review of Current Technologies and Future Trends," J. Lightwave Technol. **36**(2), 377–400 (2018).
12. J. Zhou, Y. Qiao, X. Huang, C. Yu, Q. Cheng, X. Tang, M. Guo, W. Liu, and Z. Li, "Joint FDE and MLSD Algorithm for 56-Gbit/s Optical FTN-PAM4 System Using 10G-Class Optics," J. Lightwave Technol. **37**(13), 3343–3350 (2019).
13. C. Häger and H. Pfister, "Nonlinear Interference Mitigation via Deep Neural Networks," in *Optical Fiber Communication Conference* (Optical Society of America, 2018), paper W3A.4.
14. E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind nonlinearity equalization by machine-learning-based clustering for single-and multichannel coherent optical OFDM," J. Lightwave Technol. **36**(3), 721–727 (2018).
15. L. Shu, J. Li, Z. Wan, W. Zhang, S. Fu, and K. Xu, "Overestimation Trap of Artificial Neural Network: Learning the Rule of PRBS," in *European Conference on Optical Communication* (IEEE, 2018), pp. 1–3.
16. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations* (2015), pp. 1–15.

158.132.161.52 {ts '2020-01-22 02:35:05'}