

Accelerating LMS-Based Equalization With Correlated Training Sequence in Bandlimited IM/DD Systems

Shuang Yao¹, Xizi Tang, Rui Zhang¹, Qi Zhou¹, Shang-Jen Su², Chin-Wei Hsu, Yahya Alfadhli, Xiaoli Ma¹, *Fellow, IEEE*, and Gee-Kung Chang², *Fellow, IEEE*

Abstract—As higher symbol rates are utilized in the intensity modulation and direct detection (IM/DD) scheme to meet the unrelenting growth of data traffic, overcoming the inter-symbol interference (ISI) induced by the limited bandwidth has become increasingly crucial. Channel equalization based on digital signal processing (DSP) is an effective solution, where least-mean squares (LMS) algorithm is adopted to adjust tap coefficients. However, the LMS algorithm usually has slow rate of convergence and requires lots of training symbols. This work proposes a novel training sequence to accelerate the LMS-based equalization. A first-order Markov chain (MC) is employed for sequence generation, which introduces correlation between samples and shapes signal spectrum. Compared with the conventional training sequence that consists of independent and identically distributed (i.i.d.) samples and has a white spectrum, the MC sequence enables faster convergence of tap coefficients and mean-squared error (MSE). Moreover, an experimental demonstration of a 43 Gbaud PAM-4 signal shows that the proposed sequence can achieve a lower pre-forward-error-correction (pre-FEC) bit error rate (BER) than that of the i.i.d. sequence with the same length. When the PAM-4 signal is transmitted over a 5-km standard single mode fiber (SSMF) with 6-dB system bandwidth of 10 GHz, more than 70% training sequence length reduction can be attained. When the fiber length is increased to 10 km and the signal suffers from severe power fading, more than 48% reduction can be achieved.

Index Terms—Channel equalization, digital signal processing, intensity modulation and direct detection, least-mean squares.

I. INTRODUCTION

THE exponential growth of data traffic fueled by data-hungry services including cloud computing and video streaming drives the development of high-speed transmissions in short-reach, metro as well as long-haul optical networks [1]. Considering the requirements of cost-effectiveness, low

power consumption and small footprint of short-reach optical networks, the intensity modulation and direct detection (IM/DD) scheme is an attractive solution [2]. To keep pace with the required capacity, multilevel pulse amplitude modulation (PAM) has been utilized, and a steady increase of symbol rate has been observed, from 10 Gbaud (e.g., IEEE 802.3ae [3]) to 25 Gbaud (e.g., IEEE 802.3bs [4]) and 50 Gbaud (e.g., IEEE 802.3cu [5]), with higher than 100 Gbaud being studied for beyond 400 Gb/s operation [6].

One of the challenges with high symbol rate is the limited bandwidth of the deployed devices, such as digital-to-analog and analog-to-digital converters (DACs/ADCs), modulators and photodetectors (PDs). It induces inter-symbol interference (ISI) and degrades receiver sensitivity. Channel equalization via digital signal processing (DSP) can effectively mitigate the penalty of ISI. Feedforward equalizers (FFE) are commonly utilized linear equalizers that correct both postcursor and precursor ISI. They can be used alone [7], [8], or along with decision feedback equalizers (DFEs) and/or maximum likelihood sequence estimation (MLSE) [9]–[14]. The tap coefficients can be adjusted with least-mean squares (LMS) algorithm, where stochastic gradient descent (SGD) is conducted to minimize the mean-squared error (MSE) between training symbols and equalized symbols. LMS algorithm has attracted a lot of attention, mainly due to its computational simplicity. However, its convergence rate depends on the autocorrelation of the equalizer input signal, which is strongly influenced by the channel spectral characteristics. The convergence rate is slow when the channel has severe distortion [15], [16], which is often the case with bandlimited IM/DD systems. A large number of training symbols are thus required [9], [11], resulting in high transmission overhead, especially for burst data transmission.

Recursive least squares (RLS) algorithm is an alternative algorithm to adjust tap coefficients. The cost function is the weighted sum of squared error calculated using all the past received symbols. For the tap coefficients update, a matrix is used in place of the scalar step size in the conventional LMS algorithm, which greatly accelerates convergence [15], [16]. The convergence rate can be improved by one order of magnitude [13]. Nevertheless, the calculation of this matrix also gives rise to higher computational complexity. Another solution is to apply variable step size in tap coefficients adjustment. Rather than being constant, the step size is initialized to a larger value

Manuscript received January 22, 2022; revised March 22, 2022; accepted March 29, 2022. Date of publication March 31, 2022; date of current version July 2, 2022. This work was supported in part by a grant from the Industry/University Cooperative Research program of National Science Foundation for Center of Fiber Wireless Integration and Networking (FiWIN) for Heterogeneous Mobile Communications under Grant 1821819. (Corresponding author: Shuang Yao.)

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: syao65@gatech.edu; txizi3@gatech.edu; ruizhangce@gatech.edu; qi.zhou@gatech.edu; taiwanjen@gatech.edu; chsu95@gatech.edu; yalfadhli@gatech.edu; xiaoli@gatech.edu; gkchang@ece.gatech.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2022.3164031>.

Digital Object Identifier 10.1109/JLT.2022.3164031

to speed up initial convergence, and then reduced to smaller ones as the equalizer output error decreases. Several algorithms have been proposed, such as a two-step procedure in [17], and varying step size at each iteration with values determined by the number of sign changes of an estimated gradient of MSE [18] or the squared instantaneous error [19]. The learning rate decay techniques developed for the SGD including exponential decay and staircase decay can also be applied to adjust step size, achieving similar performance with the aforementioned algorithms [8].

In this paper, the LMS-based equalization is accelerated by exploiting a novel training sequence. Unlike the traditional training sequence that is made up of independent and identically distributed (i.i.d.) samples, correlated samples generated from a first-order Markov chain (MC) are utilized. It alters the autocorrelation of the equalizer input signal and enables rapid convergence. The performance of the proposed technique is compared with the conventional i.i.d. sequence, where they are both employed for training. Faster convergence is observed in terms of tap coefficient adaptation and MSE. And lower pre-forward-error-correction (pre-FEC) bit error rate (BER) can be attained at a given training sequence length. An experimental demonstration of a 43 Gbaud four-level PAM (PAM-4) signal is conducted, where more than 70% training sequence length reduction can be achieved for 5-km standard single mode fiber (SSMF) transmission, and more than 48% reduction can be achieved for 10-km SSMF transmission.

II. PRINCIPLES OF OPERATION

A. LMS Algorithm

For an N -tap equalizer adapted with LMS algorithm, the equalizer input signal $y(k)$ is often arranged into vectors of length N :

$$\mathbf{y}(k) = [y(k - (N - 1)/2), \dots, y(k + (N - 1)/2)]^T \quad (1)$$

Here N is assumed to be an odd number. Let the equalizer tap coefficients at instant k be $\mathbf{w}(k)$, and then the equalizer output signal is

$$z(k) = \mathbf{w}(k)^T \mathbf{y}(k) \quad (2)$$

where the superscript T denotes transpose. The tap coefficients are updated according to

$$\mathbf{w}(k + 1) = \mathbf{w}(k) + \mu e(k) \mathbf{y}(k) \quad (3)$$

where μ is the step size and $e(k)$ is the equalizer error calculated using $e(k) = x(k) - z(k)$, with $x(k)$ being the training signal. The optimal tap coefficients that achieve the minimum MSE (MMSE) is

$$\mathbf{w}_{opt} = \mathbf{A}^{-1} \mathbf{b} \quad (4)$$

where $\mathbf{A} = E\{\mathbf{y}(k)\mathbf{y}(k)^T\}$ is an $N \times N$ autocorrelation matrix, and $\mathbf{b} = E\{\mathbf{y}(k)x(k)\}$ is an $N \times 1$ cross-correlation vector.

Since $\mathbf{A}$ is a symmetric matrix, it can be expressed as $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is an orthonormal matrix whose column vectors are the eigenvectors of $\mathbf{A}$, and $\mathbf{\Lambda}$ is a diagonal matrix with

elements equal to the eigenvalues of $\mathbf{A}$. Define a transformed tap coefficient error vector $\mathbf{v}(k)$ as

$$\mathbf{v}(k) = \mathbf{U}^T (\mathbf{w}(k) - \mathbf{w}_{opt}) \quad (5)$$

Then it can be proven that [15]–[17]:

$$E\{\mathbf{v}(k + 1)\} = (\mathbf{I} - \mu\mathbf{\Lambda}) E\{\mathbf{v}(k)\} \quad (6)$$

As iteration increases, $\mathbf{v}(k)$ decreases exponentially, and decay rate of the i -th element is determined by the i -th eigenvalue λ_i and the step size. To have $E\{\mathbf{v}(k)\}$ converge, the following condition has to be met

$$|1 - \mu\lambda_i| < 1, i = 1, 2, \dots, N \quad (7)$$

Consequently,

$$0 < \mu < \frac{2}{\lambda_{max}} \quad (8)$$

with λ_{max} being the maximum eigenvalue. When there is a large difference between λ_{max} and λ_{min} (the minimum eigenvalue), having μ satisfy (8) slows down the convergence of the elements with small eigenvalues. Therefore, eigenvalue spread, defined as $\chi = \lambda_{max}/\lambda_{min}$, determines the convergence rate of LMS algorithm. χ is generally difficult to calculate, but it is bounded by the dynamic range of the power spectrum $P(\omega)$ of $y(k)$ [20], i.e.,

$$\chi \leq \frac{\max_{-\pi \leq \omega < \pi} P(\omega)}{\min_{-\pi \leq \omega < \pi} P(\omega)} \quad (9)$$

The equality holds as $N \rightarrow \infty$.

B. Correlated Training Sequence to Accelerate LMS-Based Equalization

Both RLS algorithm and LMS adaptation with variable step size focus on amending (3), either by replacing μ with an estimate of $\mathbf{A}^{-1}$ or by having μ take different values at different stages. The proposed method, on the other hand, aims at changing the eigenvalue spread of $\mathbf{A}$. As (9) shows, smaller χ can be obtained if the dynamic range of $P(\omega)$ reduces. Furthermore, $P(\omega) = |H(\omega)|^2 |X(\omega)|^2$, with $H(\omega)$ and $X(\omega)$ being the channel response and the spectrum of the training signal. While $H(\omega)$ is fixed, it is still possible to modify $P(\omega)$ by changing $X(\omega)$. Since most of the bandlimited IM/DD systems behave as low-pass filters, the training sequence should have more power distributed at higher frequencies to “flatten” $P(\omega)$. A first-order MC is employed to generate such a sequence and the transition probability is

$$P(x(k + 1) = a_n | x(k) = a_m) = \frac{1}{Z_m} e^{\beta(a_n - a_m)^2} \quad (10)$$

where a_m and a_n are symbols from the signal alphabet, β is a hyperparameter to tune the spectrum of $x(k)$, and Z_m is a normalization factor to ensure that the probabilities sum to 1. When $\beta = 0$, it becomes the conventional i.i.d. process. When $\beta > 0$, symbols with larger Euclidean distances from the current symbol are transmitted at the next instant with higher probabilities.

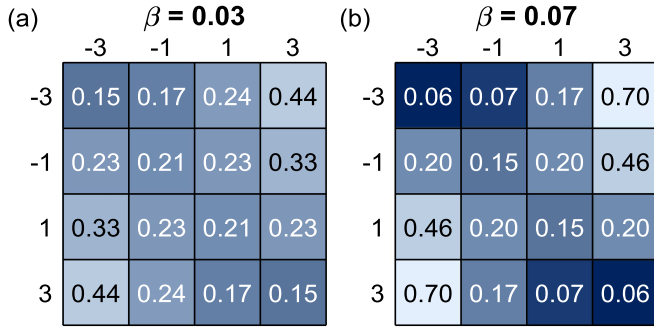


Fig. 1. Transition matrices of the MC with (a) $\beta = 0.03$; (b) $\beta = 0.07$.

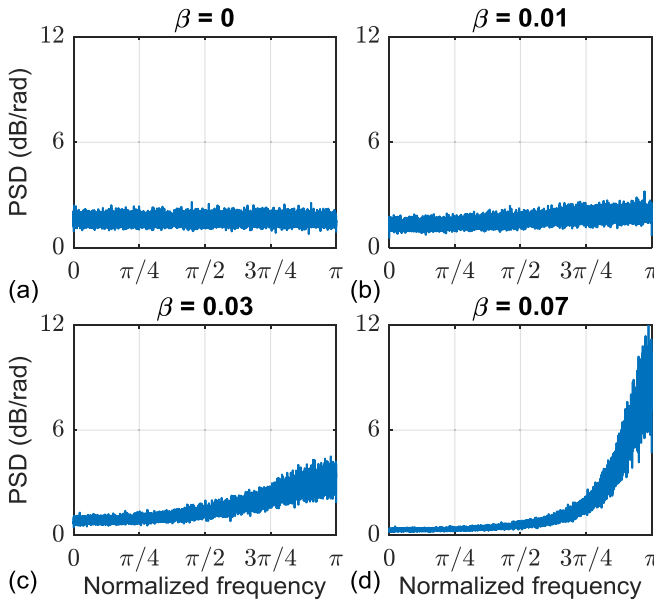


Fig. 2. PSDs of the sequences generated by the first-order MC when (a) $\beta = 0$; (b) $\beta = 0.01$; (c) $\beta = 0.03$; (d) $\beta = 0.07$.

Fig. 1 shows the transition matrices for PAM-4 sequences with $\beta = 0.03$ and $\beta = 0.07$. In both cases, all the symbols have the lowest probability to remain unchanged at the next instant. Thus, the diagonal element achieves the minimum in each row. For the symbols with negative amplitudes, the most probable symbol at the next instant is 3, while it becomes -3 for the symbols with positive amplitudes. Although the exponential term $e^{\beta(a_n - a_m)^2}$ is the same for the transitions from a_m to a_n and from a_n to a_m , Z_m are not necessarily equal as the symbol at the current instant varies, and thus the transition matrix is not symmetric. As β increases, the symbol “furthest” from the current symbol gains higher probability to appear in the MC sequence at the following instant. Conversely, the current symbol itself has lower probability to be retransmitted. Fig. 2(a)–(d) plot the power spectral densities (PSDs) of the MC sequences when β are 0, 0.01, 0.03 and 0.07, respectively. The correlation between symbols enhances the power at high frequencies, and such spectral shaping becomes stronger with larger β .

TABLE I
EIGENVALUE SPREAD OF THE AUTOCORRELATION MATRIX

Type	$H(\omega)$	i.i.d. sequence	MC sequence $\beta = 0.01$	MC sequence $\beta = 0.03$	MC sequence $\beta = 0.07$
	3-dB bandwidth				
Fourth-order Bessel filter	$0.4 BW_{Nyq}$	19.252	12.935	5.470	2.125
	$0.5 BW_{Nyq}$	4.685	3.137	1.324	4.914
	$0.6 BW_{Nyq}$	1.879	1.265	1.910	12.307
Gaussian filter	$0.4 BW_{Nyq}$	17.251	11.585	4.895	2.161
	$0.5 BW_{Nyq}$	3.928	2.628	1.176	5.934
	$0.6 BW_{Nyq}$	1.840	1.237	1.954	12.593

C. Simulations

To illustrate the effect of the proposed MC sequence in the LMS-based equalization, simulations are conducted where $H(\omega)$ is assumed to be fourth-order Bessel filters or Gaussian filters with varying bandwidths. Additive white Gaussian noise (AWGN) is added after the channel. A 15-tap T -spaced FFE is employed to compensate the ISI. Table I lists χ of i.i.d. sequence and MC sequence under various $H(\omega)$. 3-dB bandwidth of $H(\omega)$ is normalized by the Nyquist bandwidth BW_{Nyq} (half the symbol rate). The element at the i -th row and j -th column of A is estimated by the Monte Carlo estimation:

$$A(i, j) = \frac{1}{L - |j - i|} \sum_{k=1}^{L - |j - i|} y(k) y(k + |j - i|) \quad (11)$$

where L is the length of the training sequence and is fixed at 10^5 in the simulations. Each kind of sequence is generated independently for 1000 times and χ reported in Table I is the mean value of these 1000 sequences. For both kinds of filters, χ of the i.i.d. sequence increases with the decrease of the 3-dB bandwidth as a result of stronger bandwidth limitation. For the MC sequence, χ depends on both β and the 3-dB bandwidth. For the MC sequence with $\beta = 0.01$, the relationship between χ and the 3-dB bandwidth is similar to that of the i.i.d. sequence. On the other hand, for the MC sequence with $\beta = 0.07$, χ increases as the 3-dB bandwidth becomes larger. In all the cases considered in Table I, there exists at least one MC sequence whose χ is smaller than that of the i.i.d. sequence and the reduction becomes greater when the bandwidth limitation becomes more severe.

All the tap coefficients of the 15-tap FFE is initialized to 0, except for the center tap, which is initialized to 1. Fig. 3(a) and (b) respectively plot the value of the center tap and the MSE versus the number of iterations when $H(\omega)$ is a fourth-order Bessel filter with 3-dB bandwidth of $0.4 BW_{Nyq}$. The training sequence is an i.i.d. sequence or an MC sequence with $\beta = 0.07$. It can be clearly seen that the smaller χ of the MC sequence results in faster convergence. It takes the center tap fewer iterations to converge and the MSE decreases at a faster speed. Furthermore, as shown in Fig. 4, there is negligible difference between the converged tap coefficients of these two kinds of training sequences. Thus, the tap coefficients trained by the MC

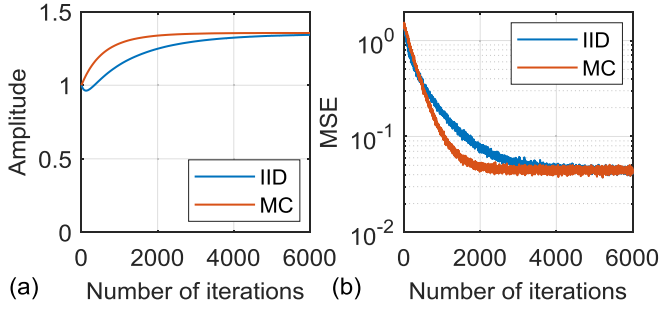


Fig. 3. (a) Amplitude of the center tap versus the number of iterations. (b) MSE versus the number of iterations.

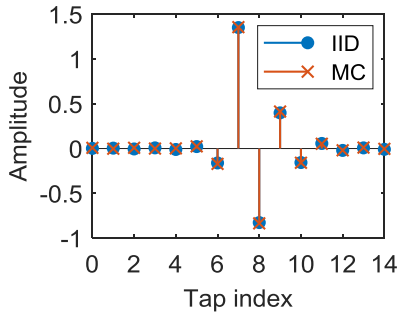


Fig. 4. Tap coefficients after reaching convergence.

sequence can be applied to channel equalization when i.i.d. data are transmitted.

It is worth mentioning that by exploiting the MC, the probability distribution of the symbols is also shaped. In the case of PAM-4 sequence, -3 and 3 have higher probabilities than -1 and 1 . However, the probabilistic shaping itself does not contribute to a faster convergence since the spectrum remains unchanged. For the example considered in Table I, if the MC sequence is randomly shuffled before transmission, which produces a sequence with a white spectrum and the same probability distribution with the original MC sequence, similar χ to that of the i.i.d. sequence is observed. Consequently, the shuffled MC sequence has a similar convergence rate with the i.i.d. sequence and cannot accelerate the equalizer training as the unshuffled MC sequence does.

III. EXPERIMENTAL SETUP

Fig. 5(a) shows the experimental setup. A 43 Gbaud PAM-4 signal is generated offline and pulse shaped by a root-raised-cosine (RRC) filter with a roll-off factor of 0.1. The digital signal is converted to an analog waveform by an arbitrary waveform generator (AWG, Keysight M8195A) at 64 GSa/s with an analog bandwidth of 25 GHz. The generated signal is amplified by a 35-GHz driver amplifier and modulated onto a 1550-nm optical carrier via a 40-GHz Mach-Zehnder modulator (MZM) biased at the quadrature point. The modulated optical signal is transmitted over a 5-km or 10-km SSMF. A variable optical attenuator (VOA) adjusts the received optical power (ROP) before a 11-GHz photodetector (PD) with integrated amplifiers. The

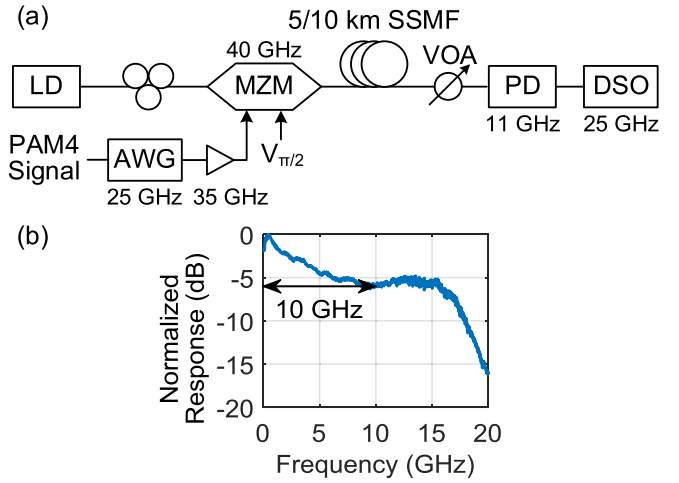


Fig. 5. (a) Experimental setup. (b) Frequency response of the transmission system with a 5-km SSMF (LD: Laser diode).

detected signal is captured for offline DSP by a digital storage scope (DSO, Keysight DSOZ254A) with an analog bandwidth of 25 GHz at 80 GSa/s. The frequency response of the transmission system with a 5-km SSMF is shown in Fig. 5(b). The 6-dB bandwidth is about 10 GHz.

The transmitted sequence consists of an i.i.d. training sequence or an MC training sequence followed by a data sequence with 1.5×10^5 i.i.d. PAM-4 samples with uniform distribution. As mentioned in Section II.C, the probability distribution of the PAM-4 symbols is not uniform in the MC sequence, resulting in higher average power. For fair comparison, the MC sequence is normalized to have the same average power with the i.i.d. sequence before transmission. In the receiver-side DSP, the received signal is first resampled to 2 samples per symbol, and then equalized by a $T/2$ -spaced FFE for 5-km SSMF transmission or a $T/2$ -spaced FFE + T -spaced DFE for 10-km SSMF transmission. Throughout the experiments, the center-tap initialization that has been used in the simulation is adopted for the FFE. For the DFE, the first tap is initialized to -1 and the other taps are initialized to 0. BER calculation is conducted after the channel equalization using only the data sequence.

IV. RESULTS AND DISCUSSION

A. 5-km SSMF Transmission

Hyperparameter β is first fixed at 0.07 and the FFE has 61 taps. Fig. 6 shows the PSDs of the received training sequences. After transmitting over the link, the spectrum of the MC sequence has a smaller dynamic range due to the spectral shaping at the transmitter. Fig. 7 plots the MSE versus the number of iterations. Again, the MC training sequence has a faster speed of convergence, achieving lower MSE after the same number of iterations. Such improvement in MSE can be translated into improvement of BER. Fig. 8(a) and (b) show the pre-FEC BER versus the length of the training sequence at ROPs of -2.7 dBm and -6.7 dBm, respectively. For both kinds of training sequences, pre-FEC BER decreases as there are more training symbols

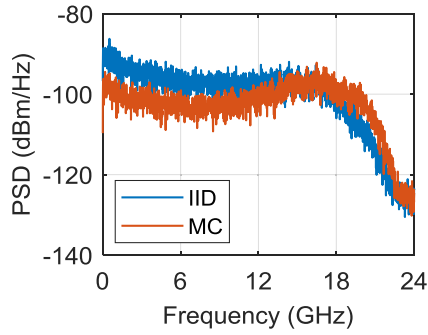


Fig. 6. PSDs of the received i.i.d. sequence and the received MC sequence after 5-km SSMF.

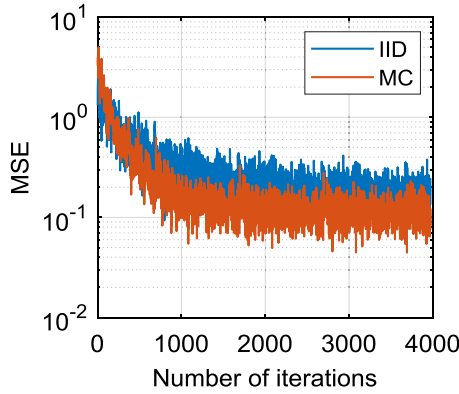


Fig. 7. MSE versus the number of iterations.

involved in the tap coefficients adaptation. At a fixed training sequence length, the MC sequence has a lower pre-FEC BER. Fig. 8(c) and (d) depict the pre-FEC BER versus ROP when the training sequence lengths are 150 and 4000, respectively. If the i.i.d. training sequence has only 150 symbols, the tap coefficients are still far from the optimal solution and there is severe ISI in the equalized signal, as can be seen from the eye diagram shown in Fig. 8(e) and the histogram shown in Fig. 8(f). The performance is dominated by the residual ISI and increasing ROP cannot bring about improvement. Contrarily, the FFE acts more effectively to eliminate ISI when it is trained by the MC sequence with 150 symbols. Consequently, pre-FEC BER decreases with higher ROP. The eye diagram and the histogram at an ROP of -2.7 dBm are illustrated in Fig. 8(g) and (h), respectively. When the length of the training sequence is increased to 4000, the residual ISI for both i.i.d. sequence and MC sequence becomes smaller and the noise dominates the performance. Still, the pre-FEC BER of the MC sequence is lower at all the measured ROPs, with slightly higher performance gain at higher ROP.

Table II summarizes the reductions in training sequence length of the MC sequence compared with the i.i.d. sequence at various FEC thresholds. At a fixed ROP, the improvement of adopting the MC training sequence generally becomes larger as the considered FEC threshold becomes lower. For instance, at an ROP of -2.7 dBm, the length of the training sequence can be reduced by 72.63% at the FEC threshold of 1.25×10^{-2} , while it increases

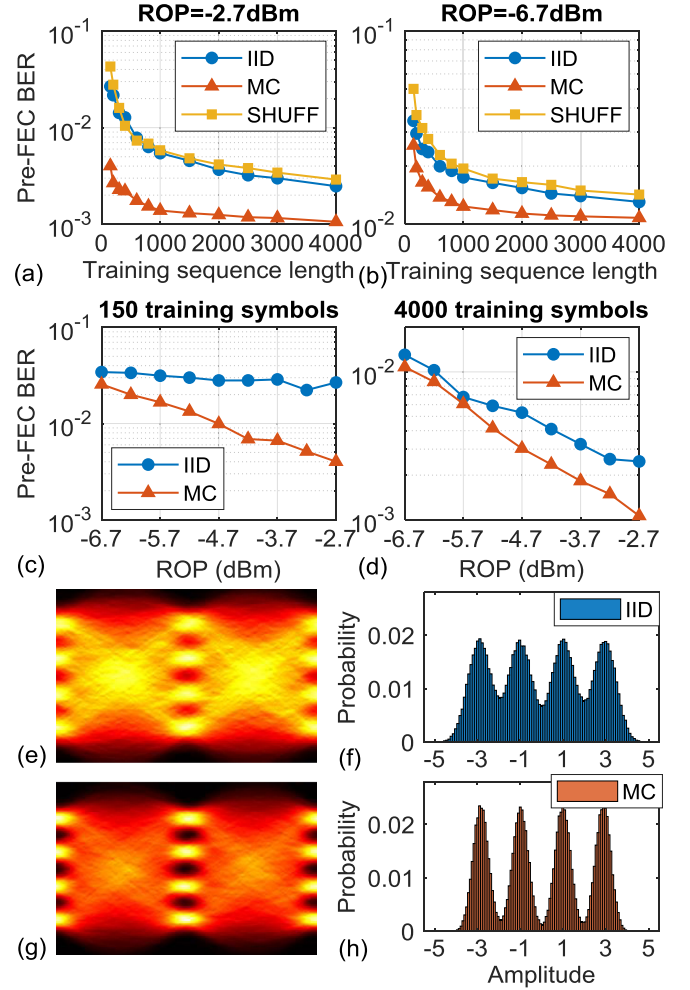


Fig. 8. Pre-FEC BER versus training sequence length at ROPs of (a) -2.7 dBm and (b) -6.7 dBm; pre-FEC BER versus ROP when the training sequence lengths are (c) 150 and (d) 4000; (e) and (f) Eye diagram and histogram of the equalized data sequence at an ROP of -2.7 dBm with a 150-symbol i.i.d. training sequence; (g) and (h) Eye diagram and histogram of the equalized data sequence at an ROP of -2.7 dBm with a 150-symbol MC training sequence.

TABLE II
TRAINING SEQUENCE LENGTH REDUCTION FOR 5-KM SSMF TRANSMISSION

ROP (dBm)	FEC threshold		
	3.15×10^{-3} [21]	4.5×10^{-3} [22]	1.25×10^{-2} [23]
-2.7	93.20%	90.51%	72.63%
-3.7	89.65%	85.10%	72.11%
-4.7	90.31%	92.64%	70.62%
-5.7	/	/	75.52%
-6.7	/	/	80.67%

to 93.20% at the FEC threshold of 3.15×10^{-3} . For all the cases listed in Table II, more than 70% reduction can be achieved with the MC training sequence.

To further verify that the change in the probability distribution does not lead to improvement, the shuffled MC sequence described in Section II.C is also utilized for training. The pre-FEC

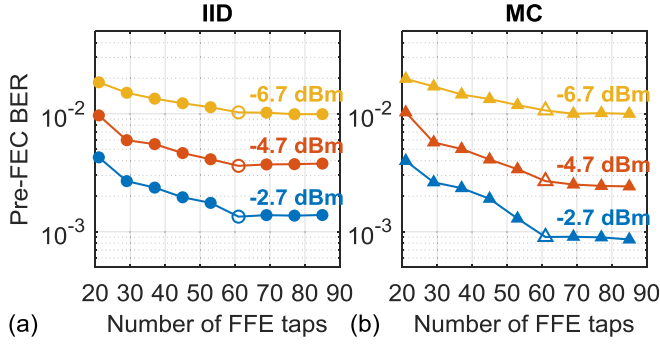


Fig. 9. Pre-FEC BER versus the number of FFE taps for (a) i.i.d. training sequence and (b) MC training sequence.

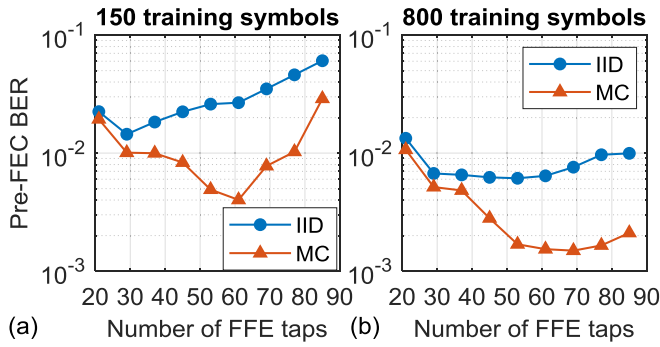


Fig. 10. Pre-FEC BER versus the number of FFE taps when the training sequence lengths are (a) 150 and (b) 800.

BER versus the length of the training sequence is plotted in Fig. 8(a) and (b), where it is labelled as “SHUFF”. The shuffled MC sequence achieves similar BER with the i.i.d. PAM-4 sequence under a given training sequence length at an ROP of -2.7 dBm, and slightly worse BER at a lower ROP of -6.7 dBm.

The performance of the LMS-based equalization also depends on the number of FFE taps. If the training sequence is adequately long, the BER can be improved by increasing the number of FFE taps. Fig. 9 shows the pre-FEC BER versus the number of FFE taps in such a scenario, where there are 2×10^4 samples in the training sequence. Fig. 9(a) and (b) are the results for i.i.d. sequence and MC sequence, respectively. The BERs of a 61-tap FFE that have been shown before are illustrated with empty circles. The relationships between BER and the number of FFE taps are similar at ROPs of -2.7 dBm, -4.7 dBm and -6.7 dBm, where the BER decreases as the number of FFE taps becomes larger. However, this relationship would change if the length of the training sequence is limited. Fig. 10(a) and (b) depict the pre-FEC BER versus the number of FFE taps at an ROP of -2.7 dBm, when the training sequences have 150 and 800 symbols, respectively. The BER first decreases and then increases as the number of FFE taps increases from 21 to 85. This is because although increasing the number of FFE taps can improve the capability to mitigate channel distortion, it would slow the rate of convergence at the same time. No matter which

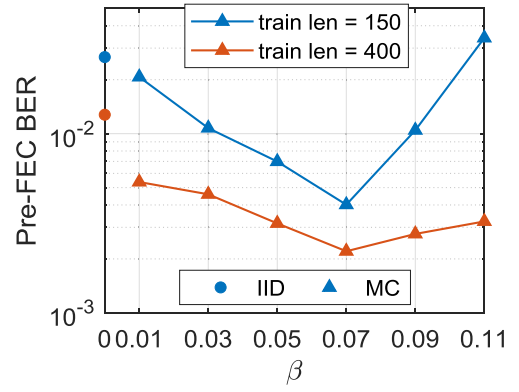


Fig. 11. Pre-FEC BER versus hyperparameter β .

FFE is selected, the MC sequence offers a better choice for training since it achieves lower BER in all the cases.

Another factor that affects the performance of the MC sequence is the hyperparameter β . The pre-FEC BERs at different β are shown in Fig. 11, in which the ROP is -2.7 dBm and the number of FFE taps is 61. Training sequences with 150 and 400 symbols are considered. In both cases, lower BER compared with the i.i.d. sequence can already be observed at $\beta = 0.01$. And it gets further reduced as β becomes larger until 0.07. After 0.07, the BER gets worse with higher β . This is mainly because the eigenvalue spread, on the other hand, will be increased if the spectral shaping is too strong. It can also be seen from Table I, where the MC sequences with β of 0.03 and 0.07 have higher eigenvalue spreads than that of the i.i.d. sequence when the 3-dB bandwidth of the channel is $0.6 BW_{Nyq}$. The performance discrepancy between MC sequences with different β gets smaller as the length of the training sequence increases. In real bandlimited IM/DD systems, β can be selected according to the electrical bandwidth of the deployed components. A larger β should be chosen as the bandwidth limitation becomes stronger.

B. 10-km SSF Transmission

Fig. 12 depicts the PSDs of the received signals after 10-km SSF transmission. The spectral null at around 19.6 GHz is due to the power fading induced by chromatic dispersion. β is 0.07 in the MC sequence and its spectrum again has a smaller dynamic range. The channel equalization is fulfilled with a 77-tap FFE + 12-tap DFE. To mitigate the error propagation of the DFE, joint symbol decision proposed in [14] is adopted. When the Euclidean distance between the output of the equalizer and the output of a decision device is larger than 0.5, the symbol-wise decision is regarded as unreliable and 6 following symbols are utilized for joint decision. To calculate the equalizer output for the following symbols, all possible combinations of PAM-4 symbols are used in the DFE part. The combination that yields the smallest sum of squared distance between the equalizer outputs and the assumed symbol decisions is selected. Fig. 13(a) and (b) illustrate the PSDs of the equalized i.i.d. sequence and the equalized MC sequence when the sequences

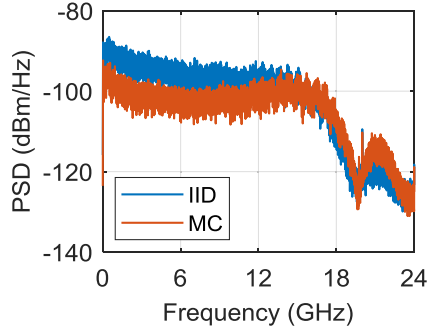


Fig. 12. PSDs of the received i.i.d. sequence and the received MC sequence after 10-km SSMF.

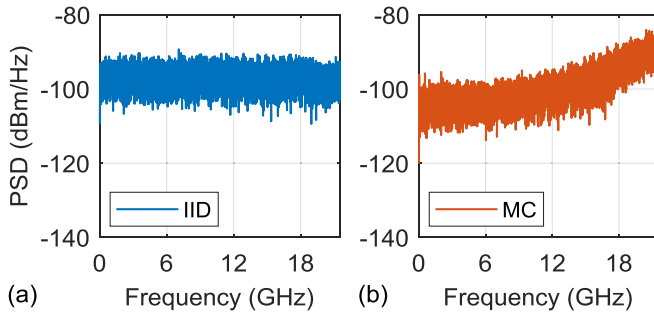


Fig. 13. PSDs of the equalized (a) i.i.d. sequence; (b) MC sequence.

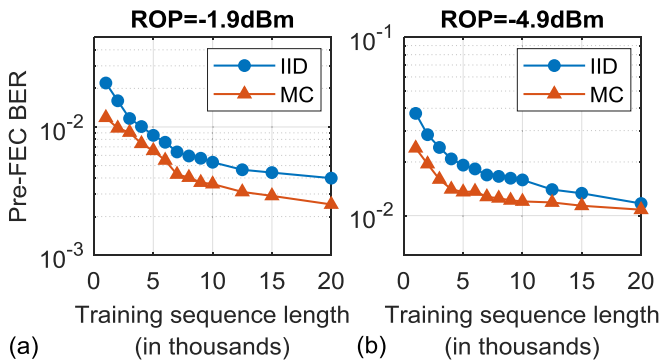


Fig. 14. Pre-FEC BER versus training sequence length at ROPs of (a) -1.9 dBm and (b) -4.9 dBm.

are sufficiently long. The spectral null is successfully removed through channel equalization.

The pre-FEC BER versus the length of the training sequence at ROPs of -1.9 dBm and -4.9 dBm are plotted in Fig. 14(a) and (b), respectively. Compared to the case of 5-km SSMF, more training symbols are required since the equalizer exploits more taps to recover the severe distortion. The MC sequence achieves lower BER at both ROPs and the reductions of training sequence length are summarized in Table III. At the FEC threshold of 1.25×10^{-2} , the reductions are generally larger at higher ROPs. 54.02% fewer training symbols are required when the ROP is -4.9 dBm, while it becomes 72.19% at an ROP of -1.9 dBm.

Fig. 15 plots the pre-FEC BER versus β when the ROP is -1.9 dBm and the training sequence has 1000 symbols. The

TABLE III
TRAINING SEQUENCE LENGTH REDUCTION FOR 10-KM SSMF TRANSMISSION

ROP (dBm)	FEC threshold		
	3.15×10^{-3} [21]	4.5×10^{-3} [22]	1.25×10^{-2} [23]
-1.9	67.11%	51.42%	72.19%
-2.9	/	57.43%	56.70%
-3.9	/	48.20%	69.11%
-4.9	/	/	54.02%

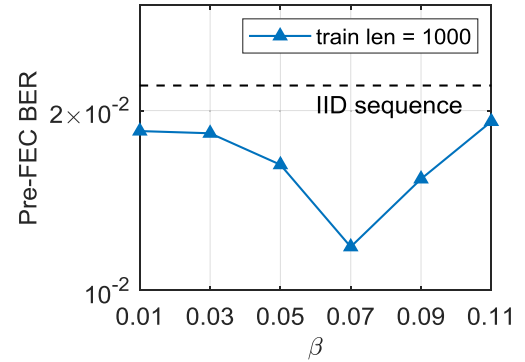


Fig. 15. Pre-FEC BER versus hyperparameter β .

BER of the i.i.d. sequence is shown with dashed line. The MC sequence has better BER than i.i.d. sequence as β varies from 0.01 to 0.11. And the MC sequence with β of 0.07 still achieves the best performance.

V. CONCLUSION

In this paper, a training sequence is proposed to accelerate the LMS-based equalization in bandlimited IM/DD systems. The sequence is generated with a first-order MC that introduces correlation between samples and enhances the power at high frequencies. Compared with the traditional i.i.d. sequence with a white spectrum, the MC sequence can reach the convergence with fewer iterations and achieve lower pre-FEC BER with a fixed training sequence length. Experimental results show that to transmit a 43 Gbaud PAM-4 signal over a system of 5-km SSMF and 6-dB bandwidth about 10 GHz, the length of the training sequence can be reduced by more than 70%. When a 10-km SSMF is utilized and the signal suffers from severe power fading induced by chromatic dispersion, the training sequence length can be reduced by more than 48%.

REFERENCES

- [1] X. Liu and F. Effenberger, "Emerging optical access network technologies for 5G wireless," *J. Opt. Commun. Netw.*, vol. 8, no. 12, pp. B70–B79, Dec. 2016.
- [2] Q. Hu, M. Chagnon, K. Schuh, F. Buchali, and H. Bülow, "IM/DD beyond bandwidth limitation for data center optical interconnects," *J. Lightw. Technol.*, vol. 37, no. 19, pp. 4940–4946, Oct. 2019.
- [3] *IEEE Standard for Information technology - Local and metropolitan area networks - Part 3: CSMA/CD Access Method and Physical Layer Specifications - Media Access Control (MAC) Parameters, Physical Layer, and Management Parameters for 10 Gb/s Operation*, IEEE Standard 802.3ae-2002, Aug. 2002.

- [4] *IEEE Standard for Ethernet - Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation*, IEEE Standard 802.3bs-2017, Dec. 2017.
- [5] *IEEE Standard for Ethernet - Amendment 11: Physical Layers and Management Parameters for 100 Gb/s and 400 Gb/s Operation over Single-Mode Fiber at 100 Gb/s per Wavelength*, IEEE Standard 802.3cu-2021, Feb. 2021.
- [6] T. Zhang, S. Zhang, and Y. Zhuang, "Considerations on the '10km @ 800Gb/s' objective," in *contribution to IEEE 802.3 Beyond 400 Gb/s Ethernet Study Group, Interim Meeting*, Jul. 2021. [Online]. Available: https://www.ieee802.org/3/B400G/public/21_07/zhang_b400g_01a_210720.pdf
- [7] M. Chagnon, M. Morsy-Osman, M. Poulin, C. Paquet, S. Lessard, and D. V. Plant, "Experimental parametric study of a silicon photonic modulator enabled 112-Gb/s PAM transmission system with a DAC and ADC," *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1380–1387, Apr. 2015.
- [8] Q. Zhang and C. Shu, "Performance investigation of learning rate decay in LMS-based equalization," *IEEE Photon. Technol. Lett.*, vol. 33, no. 2, pp. 109–112, Jan. 2021.
- [9] P. Ossieur *et al.*, "Burst-mode electronic equalization for 10-Gb/s passive optical networks," *IEEE Photon. Technol. Lett.*, vol. 20, no. 20, pp. 1706–1708, Oct. 2008.
- [10] K. Zhong *et al.*, "140-Gb/s 20-km transmission of PAM-4 signal at 1.3 μ m for short reach communications," *IEEE Photon. Technol. Lett.*, vol. 27, no. 16, pp. 1757–1760, Aug. 2015.
- [11] J. Chen *et al.*, "Adaptive equalization enabled 25Gb/s NRZ modulation based on 10-G class optics for upstream burst-mode transmission," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, Mar. 2018, pp. 1–3.
- [12] W. Wang *et al.*, "Advanced digital signal processing for reach extension and performance enhancement of 112 Gbps and beyond direct detected DML-based transmission," *J. Lightw. Technol.*, vol. 37, no. 1, pp. 163–169, Jan. 2019.
- [13] G. Coudyzer *et al.*, "Study of burst-mode adaptive equalization for >25G PON applications," *J. Opt. Commun. Netw.*, vol. 12, no. 1, pp. A104–A112, Jan. 2020.
- [14] X. Tang, Y. Qiao, Y.-W. Chen, Y. Lu, and G.-K. Chang, "Digital pre- and post-equalization for C-band 112-Gb/s PAM4 short-reach transport systems," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4683–4690, Sep. 2020.
- [15] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.
- [16] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008.
- [17] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Develop.*, vol. 16, no. 6, pp. 546–555, Nov. 1972.
- [18] R. Harris, D. Chabries, and F. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 309–316, Apr. 1986.
- [19] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.
- [20] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Technol. J.*, vol. 48, no. 1, pp. 55–70, Jan. 1969.
- [21] ITU-T Recommendations G.975.1, "Forward error correction for high bit-rate DWDM submarine systems," Feb. 2004.
- [22] ITU-T Recommendations G.709.2/Y.1331.2, "OTU4 long-reach interface," Jul. 2018.
- [23] OIF, "Implementation Agreement 400ZR," Mar. 2020. [Online]. Available: https://www.oiforum.com/wp-content/uploads/OIF-400ZR-01.0_reduced2.pdf