Memory-controlled deep LSTM neural network post-equalizer used in high-speed PAM VLC system

XINGYU LU,[1,2] CHAO LU,[2] WEIXIANG YU, [1] LIANG QIAO, [1] SHANGYU LIANG,[1] ALAN PAK TAO LAU,[2] AND NAN CHI[1,*]

[1]*Shanghai Institute for Advanced Communication and Data Science, Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China*
[2]*Photonics Research Centre, The Hong Kong Polytechnic University, Hong Kong, China*
[*]nanchi@fudan.edu.cn

Abstract: Linear and nonlinear impairments severely limit the transmission performance of high-speed visible light communication systems. Neural network-based equalizers have been applied to optical communication systems, which enables significantly improved system performance, such as transmission data rate and distance. In this paper, a memory-controlled deep long short-term memory (LSTM) neural network post-equalizer is proposed to mitigate both linear and nonlinear impairments in pulse amplitude modulation (PAM) based visible light communication (VLC) systems. Both 1.15-Gbps PAM4 and 0.9Gbps PAM8 VLC systems are successfully demonstrated, based on a single red-LED with bit error ratio (BER) below the hard decision forward error correction (HD-FEC) limit of 3.8 x 10^{-3}. Compared with the traditional finite impulse response (FIR) based equalizer, the Q factor performance is improved by 1.2dB and the transmission distance is increased by one-third in the same experimental hardware setups. Compared with traditional nonlinear hybrid Volterra equalizers, the significant complexity and system performance advantages of using a LSTM-based equalizer is demonstrated. To the best of our knowledge, this is the first demonstration of using deep LSTM in VLC systems.

1. Introduction

Visible light communications (VLC) based on light emitting diodes (LEDs) has become an attractive and promising technology due to its cost effectiveness, immunity to electromagnetic interference, license-free and high security [1]. In recent years, transmission rate of visible light systems has been increasing with the use of` high-order modulation, such as *orthogonal frequency division multiplexing* (OFDM) [2], *carrier-less amplitude and phase modulation* (CAP) [3] and *pulse amplitude modulation* (PAM) [4].

Equalizer is one of the most critical parts of the VLC systems. As a typical communication system, the transmitted signal of VLC systems is distorted in amplitude and delay causing the inevitable *inter-symbol interference* (ISI). On the other hand, nonlinear distortion has gradually become a new bottleneck in high-speed transmission systems due to nonlinear V-I response of LED source [5] and other origins that have not been well-modeled such as non-linear distortion arising from the transmitter driving circuits and the electrical amplifier (EA). Recently, Adaptive *finite impulse response* (FIR) filter based linear equalizers in VLC systems have been widely studied, such as *scalar modified constant multi-modulus algorithm* (S-MCMMA) blind equalization algorithm [6], data-aided *recursive least square* (RLS) [7] and *least mean squares* (LMS) [8]. In order to compensate for both linear and nonlinear effects, hybrid equalization schemes have been shown to be effective approaches to improve the performance of VLC systems, such as adaptive FIR linear equalizer with

Volterra series scheme [9], *look-up table* (LUT) scheme [10] and *machine learning* (ML) scheme [11,12].

With the increase in system transmission rate, ISI is enhanced and more taps are needed when designing the equalizer. Due to the high computational complexity, the order of hybrid Volterra filters is generally limited to 2 to 3 orders with limited high-order taps [9] and LUT scheme generally uses extreme limited symbols [10,13]. An ideal nonlinear equalizer can effectively compensate for nonlinearities. However, the computational complexity increases as the number of taps increases. It is well known that the level of complexity that should normally be avoided is *non-deterministic polynomial-time hardness* (NP) and above. An LSTM neural cell-based equalization network is designed to solve this problem. Long-term memory parameters are used to store relatively slow-changing channels. Parameters in short-term memory are used to quickly process finite taps that affect a symbol. Benefit from the characteristics of neural networks, the complexity of the equalizer can be significantly reduced to meet the complexity requirements.

Compared to traditional adaptive equalization, ML techniques are effective at using limited training samples to create a probability-based model. More specifically, ML emphasizes generalization ability and *concept learning* (CL) [14], and quantifies learning ability through generalization errors and under/over-fitting [15]. In recent years, many powerful traditional ML models [16] and simple multilayer *neural network* (NN) models [17] have been successfully applied in the field of communication systems, but studies have pointed out that some wrong experimental methods in communication systems may make the performance of machine learning seriously overestimated [18]. With the rise of deep neural networks in other fields and breakthroughs in nonlinear problems [19,20], we explore the use of deep neural networks as a way to solve nonlinear compensation in VLC systems.

In this paper, we propose and experimentally demonstrate a post-equalizer employing a memory controlled deep 32 layers LSTM neural network and a softmax-function based probabilistic classification model for PAM4 and PAM8 VLC systems. The memory controlled LSTM NN can simultaneously compensate for linear and nonlinear distortions. Applying the LSTM-based equalization scheme can improve system performance compared with the original FIR-based and Volterra-based equalization schemes. A data rate of 1.15-Gb/s is successfully achieved over 1-m indoor free space transmission based on a single red-LED with bit error rate (BER) under the 7% hard decision forward error correction (HD-FEC) limit of 3.8×10^{-3}. Further, our experiments demonstrate that the LSTM neural network is suitable for truly random sequences. To the best of our knowledge, this is the first time that the LSTM of machine learning is successfully applied in VLC systems.

2. Principle

In VLC systems, it is easy to illustrate the nonlinear distortion by mapping the Tx and Rx signal amplitudes. Figure 1 shows the back-to-back PAM-8 data through the VLC channel. Impaired by nonlinearity, the data deviates from the linear straight line. As a trade-off in VLC systems, reducing the modulation depth will reduce nonlinearity while increasing the modulation depth improves the signal-to-noise ratio due to increased extinction ratio.
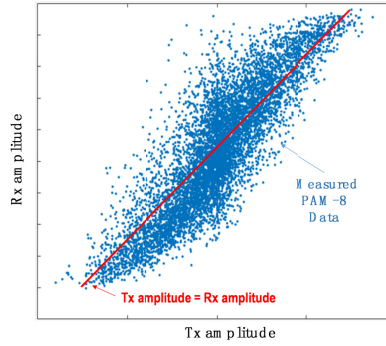
Fig. 1. Back-to-back transfer curve of PAM8 data of VLC system.

The most widely used linear FIR filter equalizer and Volterra based equalizer nonlinear are shown in Eq. (1) and Eq. (2):

$$y(n) = \sum_{i=0}^{N-1} w_i \cdot x(n-i) \tag{1}$$

where $x(n)$ is the input signal and $y(n)$ is the output signal, N is taps number of the FIR filter.

$$y(n) = \sum_{k_1=0}^{N_1-1} w_{k_1}(n) \cdot x(n-k_1) + \sum_{k_1=0}^{N_2-1}\sum_{k_2=k_1}^{N_2-1} w_{k_1 k_2}(n) \cdot x(n-k_1)x(n-k_2)$$
$$+ \sum_{k_1=0}^{N_3-1}\sum_{k_2=k_1}^{N_3-1}\sum_{k_3=k_2}^{N_3-1} w_{k_1 k_2 k_3}(n) \cdot x(n-k_1)x(n-k_2)x(n-k_3) + ... \tag{2}$$
$$+ \sum_{k_1=0}^{N_p-1} ... \sum_{k_p=k_{p-1}}^{N_p-1} w_{k_1...k_p}(n) \cdot x(n-k_1)...x(n-k_p)$$

where $x(n)$ is the input signal and $y(n)$ is the output signal, $N_1 ...N_p$ is taps numbers of the Pth order nonlinear equalizer, and w_{ki} is the weights of the equalizer. In practical applications, *mean square error* (MSE) and FIR based linear equalizer are utilized to adjust the weights adaptively and compensate for linear distortion. LSTM theoretically compensates for both linear and nonlinear distortions with complete channel equalization as a pattern classifier, as Eq. (3):

$$y(n) = f(X_n)$$
$$= g\{h[\sum_{i=0}^{N} w_i \cdot x(n-i) + b]\} \tag{3}$$

where b is the bias, $h(*)$ is a output function of the LSTM cells. $g(*)$ is a function designed for PAM systems, as shown in Eq. (6) and Eq. (7) below.
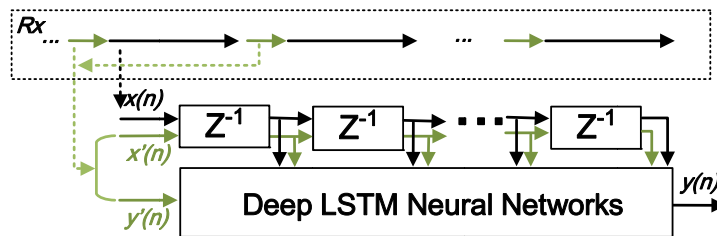

Fig. 2. Diagram of LSTM.

For traditional equalizer weights fitting is reset after each retraining, so there is no memory effect in it. Differently, the parameters in the neural network have long-term memory in multiple trainings to help the model compensate for nonlinearity more accurately. Specifically, the LSTM assigns the highest priority training data to the latest training sequence during training, but also retains some weights in the original channel model. As shown in Fig. 2, a certain length of $x'(n)$ and $y'(n)$ are used as the training sequence to converge parameters of the neural network. As can be seen from Eq. (2), $y(n)$ is related to the N taps vector X_n. For convenience, we express X_n as Eq. (4):

$$X_n = X^T = [x(n), x(n-1),...,x(0)]^T \tag{4}$$

where $x(n)$ is the n-th input symbol of N taps and T denotes transpose or the vector. Therefore, each of our training samples and equalization set of input and output can be expressed as Eq. (5):

$$S(n) = \{X_n, y(n)\} \tag{5}$$
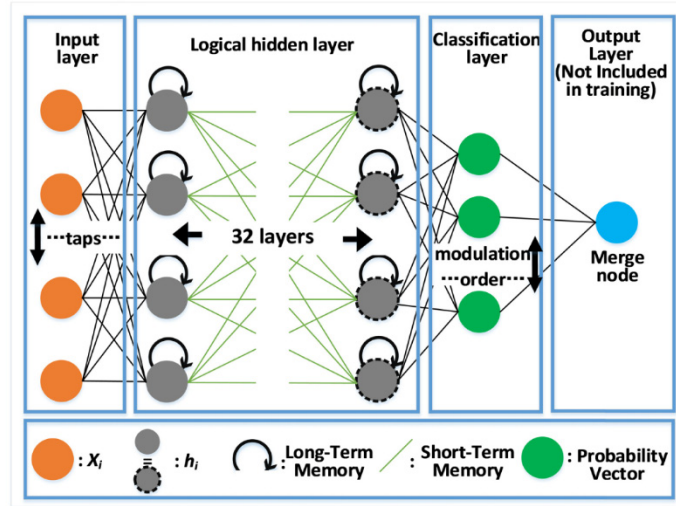
2.1 Deep LSTM neural network structure

Fig. 3. Structure of LSTM based equalizer.

LSTM is a probability-based model which cannot be used directly for equalization. Figure 3 shows the structure of the equalizer based on LSTM. The equalizer includes an input layer, a hidden layer, a classification layer, and an output layer. The standard LSTM cells is defined in [20]. We propose to transform the neural network output into a time-domain equalizer via a merge node according to the softmax function of Eq. (6) and Eq. (7):

$$P(\mathrm{y}{=}L_i){=}\frac{e^{X^T w_i}}{\sum_{k=1}^{K} e^{X^T w_k}} \tag{6}$$

Where $P(y = Li)$ represents the probability that y is equal to different amplitude levels Li, x is the Rx vector of tap numbers and w_k is the weights of the equalizer.

$$g(\cdot) = \overline{\sum_{i=0}^{NoL} P(\mathrm{y}{=}L_i) \cdot L_i} \tag{7}$$

Where *NoL* is the number of the amplitude levels (e.g. *NoL* of PAM-8 is 8, L1 = −7**Normalization Factor* (NF), L2 = −5*NF, L7 = 7*NF). For the long/short-term memory links in Fig. 3, we use the standard LSTM cell structure in [20].
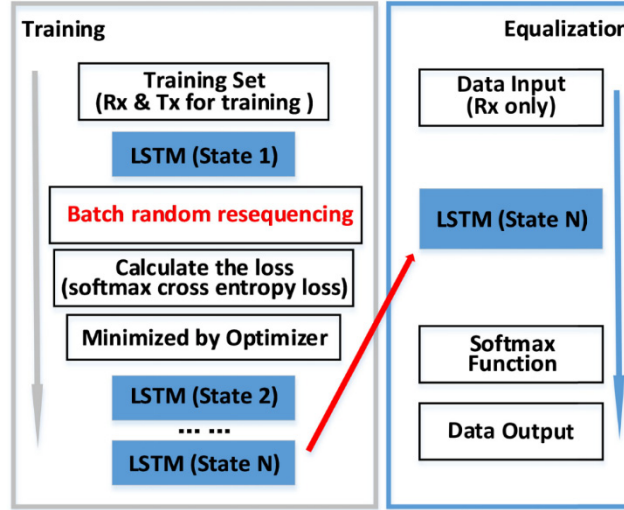
2.2 Training and transmission

Fig. 4. Flow chart of training and equalization.

Figure 4 shows the flow chart of training process and equalization. In order to train more precisely than using direct accuracy based error, such as Mean Square Error(MSE) between actual and predicted during training, a softmax based cross-entropy equation is carried out in the algorithm as Eq. (8):

$$H_{Y^{train}}(Y) = -\sum_i Y_i^{trian} \log(P_{X^{train}}(y = L_i))$$

$$= -\sum_i Y_i^{trian} \log(\frac{e^{X^T w_i}}{\sum_{k=1}^{K} e^{X^T w_k}}) \tag{8}$$

where Y_i is the i-th probability of the amplitude levels, X_{train} and Y_{train} is the Rx and Tx training set.

However, inside the NN optimizer, the sample is cut by the whole block called *batch* to calculate the training error, as Eq. (9):

$$batch_k = \{S_n, S_{n+1}, ..., S_{n+m}\} = \begin{bmatrix} \{ & X_n, y(n) & \} \\ & ... & \\ \{ & X_{n+m}, y(n+m) & \} \end{bmatrix} \tag{9}$$

where *m* is the batch size, and $\text{Complete sample} = \sum_i batch_i$, so If pseudo-random sequences are used for training and testing, there is a high probability situation exists:

$$P(y = L_i \mid y \in batch_k) > P_{X^{train}}(y = L_j) = \frac{e^{X^T w_i}}{\sum_{k=1}^{K} e^{X^T w_k}} \tag{10}$$

where $i \neq j$. This situation is very common in the communication field and cause the serial numbers to be remembered, which is known to be the memory of the neural network and will give a wrong estimation of system performance [18]. We used a simple but effective way to control this memory effect: randomly pick single sample of all sample sets to compose each batch as Eq. (11):

$$batch_k' = r(batch_k) = random\{S_n, S_{n+1}, ..., S_{n+m}\} \tag{11}$$

where $batch_k'$ is the new input to the optimizer. It is theoretically not possible to have the situation as in Eq. (9) after this method is used. In order to verify the feasibility of this method, we experimentally compared pseudo-random sequences with random sequences to verify that the LSTM equalizer avoids the memory trap described above.

Most neural networks have memory effects, whether it is *deep neural networks* (DNN), *convolutional neural networks* (CNN) or LSTM. It should be noted that when the neural network model is applied to the communication system, on the one hand, a part of the memory is beneficial, for example, remembering the previous parameters before the new training greatly reduces the amount of training required. On the other hand, part of the memory is harmful, such as remembering the sequence of training, because obviously the sequence to be transmitted should be unpredictable in communication system.

2.3 Complexity analysis

We estimated theoretical the complexity of LSTM network and the results show that it has an acceptable algorithm complexity. The level of training complexity is the same as that of the second-order Volterra equalizer, and it is much lower than the 3rd-order Volterra equalizer. In addition, high-order Volterra without simplification is not a practical algorithm with current computing power since its complexity is NP- hard.

Table 1. Complexity Comparison of Equalizers[a]

	LMS	DD LMS	Volterra Equalizer with data aided (General Volterra Series)	Volterra Equalizer with data aided (2-order only)	LSTM
Number of parameters	N	$N*O$	$\dfrac{(O+N-1)!}{O!(N-1)!}$	$\dfrac{(N+1)N}{2}$	$4*(N*H + H2 + H)$
Number of training operations	$4*N*L*I$	$6*N*L$	$\left\{\dfrac{4N(O+N-1)!}{O!(N-1)!}\right\}*L*I$	$\dfrac{4N(N+1)N}{2}$	$4*(N*H + H2 + H)*I$
Complexity of training	$O(n^2)$	$O(n^2)$	$O(n!)$	$O(n^3)$	$O(n^3)$
Number of equalizing operations	$2N$	$4N$	$\dfrac{2(O+N-1)!}{O!(N-1)!}$	$\dfrac{2(N+1)N}{2}$	$4*(N*H + H2 + H) + 24N$
Complexity of equalizing	$O(n)$	$O(n)$	$O(n^O)$	$O(n^2)$	$O(n)$

[a]N is the number of equalizer taps, L is the length of the training sequence, I is the number of iterations, H is the number of hidden layers.

As shown in Table 1, the complexity of training and equalizing determines the complexity of algorithm implementation. According to Eq. (2), the high-order multi-tap Volterra equalizer is theoretically impossible to be implemented since calculation of the parameters by solving higher-order Volterra series exceeds the NP hard level. In existing practice, the second-order Volterra algorithm with limited taps based on minimum MSE convergence is generally used.

3. Experimental setup and results

Figure 5 shows the experimental setup of the VLC system with PAM modulation employing LSTM based equalization.
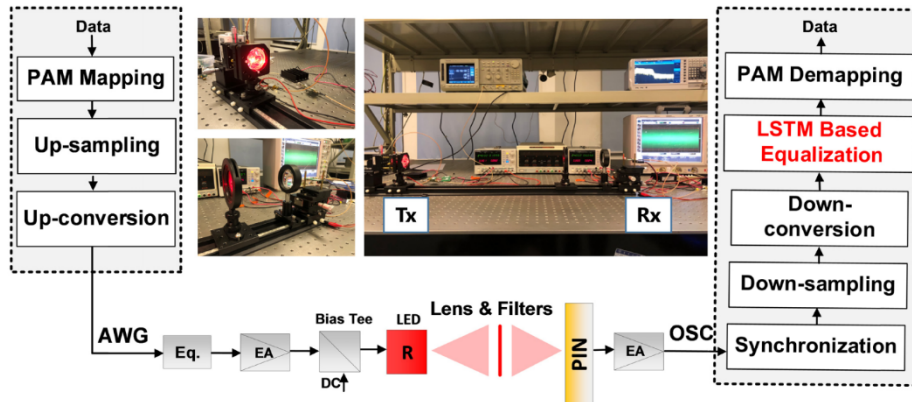

Fig. 5. Experimental setup.

At the transmitter, the original binary bits are converted to PAM symbols, and an up-conversion is performed after up-sampling. Then, the signal is generated by an AWG and passes through the basic hardware equalizer [21]. In order to reach the LED's switching current threshold, the signal amplified by an *electronic amplifier* (EA) is added with a *direct current* (DC) by a bias-T. The red illuminating chip in an ordinary commercial RGB LED (Engine LZ4-20MA00) was selected for transmitting signals to free space.

At the receiving end, the detected optical signal is converted into an electrical signal by a PIN photodiode (Hamamatsu 10784). The weak original signal is amplified by an EA and sampled by an oscilloscope (Agilent 54855A) with sampling rate of 2-GSample/s. Finally, the signal is stored in a laptop with a *graphics processing unit* (GPU Nvidia GTX1060) for offline digital signal processing through a network cable/GBIP interface.

In offline digital signal processing of PAM4 and PAM8 signals, after synchronization, the down-conversion and down-sampling corresponding to the transmitter are used first. After passing through the equalizer, the symbol is decoded into the original binary signal.
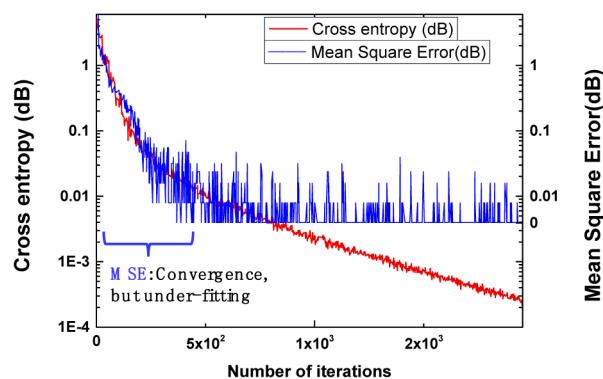

Fig. 6. Training convergence curve of LSTM.

As shown in Fig. 6, the cross-entropy continues to converge after the *mean square error* (MSE) almost close to 0 in the training process. If MSE is chosen as the error function in LSTM, as shown by the blue line in the figure, the MSE is 0 with 4×10^2 iterations, so the parameter convergence will end regardless of the convergence threshold. In particular, we

artificially set 0 points on the log coordinates of MSE to make the curve conform to the display rule. So, it will cause under-fitting problems and cannot accurately fit the parameters in the network. Actually, in general deep neural network experiments, cross entropy error is widely used in the training process of other applications, such as image identification and natural language processing.
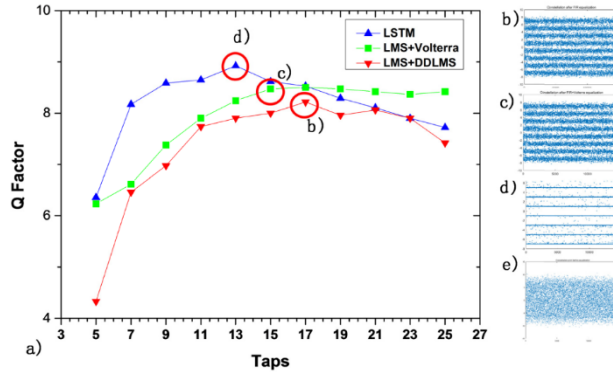

Fig. 7. a) Q factor (dB) comparison of LSTM and LMS + DDLMS. b) Constellation of the best working conditions of the LMS + DDLMS equalizer. b) Constellation of the best working conditions of the LMS + Volterra equalizer. d) Probability distribution of every symbols for the best working conditions of the LSTM equalizer and e) Constellations without equalizer.

For a clear comparison, better performing but more complex FIR equalizers, data-aid LMS equalizer with decision-directed least mean square (DD-LMS) adaptive equalizer and data-aid LMS with Volterra equalizer were used. The results of applying LMS equalizer, hybrid Volterra and LSTM equalizer to PAM8 with different taps are shown in Fig. 7. It can be found that the system performance of using LSTM-based equalization scheme is superior to the original LMS and DD-LMS based equalization scheme by approximately 1.2 dB in the case of optimal number of taps. PAM8 systems using only the LSTM equalizer reduces the error by about half comparing with systems using the LMS + Volterra equalizer, resulting in a Q increase of approximately 0.8. The BER performance of VLC PAM8 systems using only Volterra equalizer are unacceptable due to severe over-fitting. The Volterra equalizers pre-converged by LMS has approximately 0.5 better Q value than LMS equalizer in PAM8 systems. It is worth noting that Volterra equalizer in this comparison uses more tap to trade off complexity for optimal performance. When the number of taps is lower than 11, the parameter fitting of the equalizer obviously cannot follow the non-ideal response of the system. On the other hand, when the number of taps is too high, the performance of the system is reduced, because the *additive white Gaussian noise* (AWGN) accumulated in the system affects the accuracy of the optimizer in the equalizer.
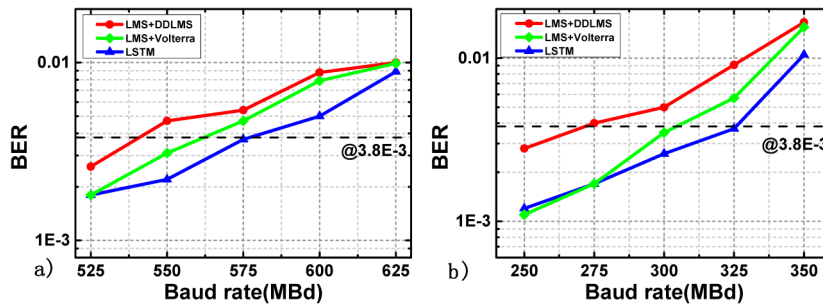

Fig. 8. BER comparison of LSTM and LMS + DDLMS of a) PAM4 and b) PAM8 VLC systems under different baud rate, respectively.

The performance of the LSTM equalizer in the PAM4 and PAM8 systems were studied separately. With the increase in baud rate, ISI also increases. Obviously, the number of taps need to be increased to cope with nonlinearity of the system, the complexity of the high-order Volterra equalizer makes its use no longer theoretically suitable. However, LSTM nonlinear equalizer can still be used since its complexity increases linearly with the number of taps. LSTM equalization scheme can make the BER performance of the system lower than the HD-FEC threshold when the transmission rate reaches 1.15-Gb/s in PAM4 and 0.975-Gb/s in PAM8 VLC systems. As shown in Fig. 8. It can be found that the system performance of using LSTM-based equalization scheme is significantly better than LMS + DDLMS and LMS + Volterra based equalization schemes. Due to the bandwidth limitation and filtering effect, system performance is seriously degraded by noise and cannot be compensated by DSP beyond certain baud rate. When the PAM8 system is operating at 325MBd, applying LSTM equalizer can achieve error-free transmission. However, at 275MBd, LMS + DDLMS scheme already failed to meet the BER requirements while the BER remains below the HD-FEC with the LSTM scheme.
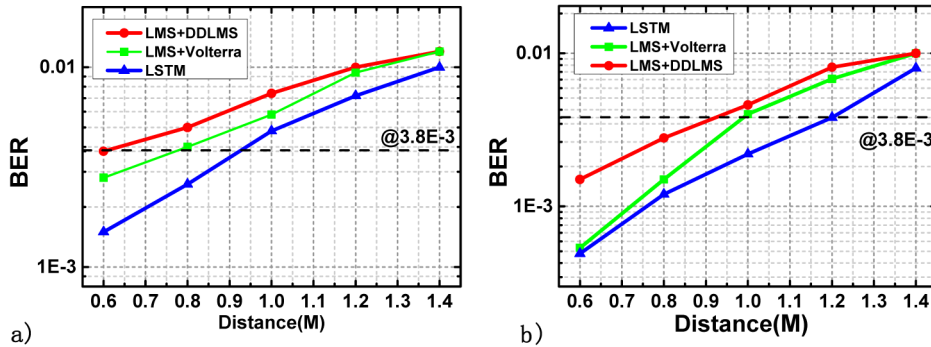

Fig. 9. Transmission distance comparison with LSTM and LMS + DDLMS in PAM8 VLC systems versus different bitrates: a) 750Mbit/s; b) 900Mbit/s.

At the optimal working conditions, the LSTM scheme improves the transmission distance performance of the PAM8 VLC systems compared with the traditional schemes. As shown in Fig. 9, BER performance improvements are more significant when the noise impact introduced by bandwidth is smaller. When the transmission bit rate is 750Mbit/s, the LSTM equalizer can make the system's transmission distance reach 1.2M and keep the bit error rate below the HD-FEC threshold. The solution using the LMS + DDLMS and LMS + Volterra equalizer can only achieve a transmission distance of 0.8 meters and 0.9 meters. When the transmission bit rate is 900 Mbit/s, the LSTM equalizer can extend the transmission distance of the system from 0.7 meters to 0.9 meters.
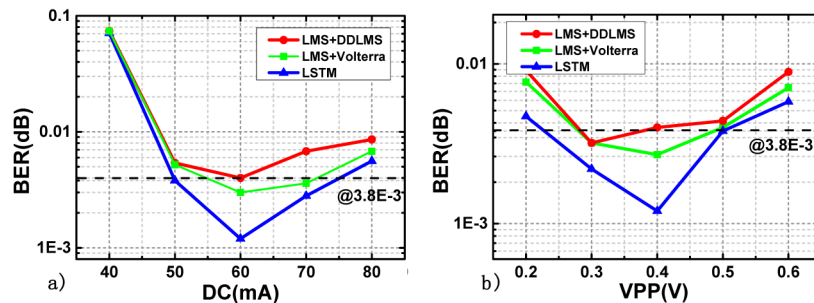

Fig. 10. Measured BER versus bias currents and input signal Vpp.

In addition, we studied the optimal operation of the LSTM equalizer at different bias *direct current* (DC) and input signal *peak-to-peak voltage* (VPP) conditions. As shown in Fig. 10, there are different working conditions with a transmission distance of 0.8m and a transmission baud rate of 750 Mbit/s, the LSTM equalizer extends the operating range under the FEC threshold limit. The PAM8 system has the best BER performance when the bias current is 60mA and VPP is 0.4 volts.
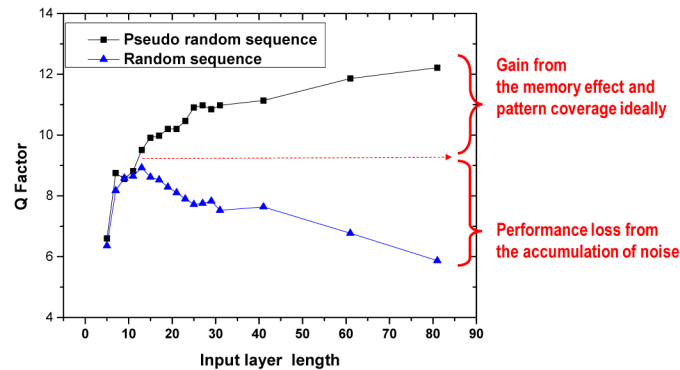

Fig. 11. Q factor performance curve of pseudo-random and random sequence.

In particular, the problem of over-estimation of the performance of LSTM when using pseudo-random sequence has been experimentally analyzed as described in section 2.2 of this paper. As shown in Fig. 11. Pseudo-random sequences give overestimation of system performance over random sequences from two aspects: Memory effects and unrealistic coverage of patterns.

Specifically, the memory effect is due to the incorrect use of the same training sequence and test sequence in the experiment, which causes the neural network to remember the serial number of the transmission sequence. Therefore, if pseudo-random sequence is used in the experiment, there will be overestimation of system performance. The reason for the unrealistic pattern coverage which is impossible to cover the taps with too long training sequences in practical applications. For example, the PAM4's 40 taps completely cover training sequence length to 4^{40} patterns. However, using some training methods in deep learning can make the performance closer to the theoretical limit—pseudo-random sequence performance in Fig. 11. One possible direction is to introduce some forgetting factors to ignore the unnecessary redundant training parts.
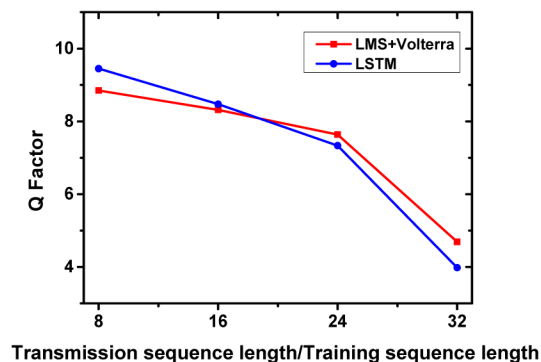

Fig. 12. Q factor performance curve of training sequence length.

Finally, the effect of the length of the training sequence on different nonlinear algorithms is experimentally measured. The Q factor after only one training is shown in the Fig. 12, in

general, LSTM and Volterra based equalizers are similar in the length of the training sequence. Specifically, when the length of the training sequence is less than 1/16, the parameters of the two equalizers do not converge well. In all of the above experiments, we only used one training sequence of the same length that is sufficiently convergent to compare the LSTM and Volterra equalizer algorithms. It is worth mentioning that in the case of well convergence, the traditional equalizer's multiple training cycles will not have a significant performance improvement, because each complete training cycle will update the parameters without memory. But differently, under the influence of long-term memory, systems with LSTM-based equalizer will get higher performance which has been trained more than one times.

4. Conclusions

We theoretically and experimentally investigate the time-domain memory controlled LSTM neural network based equalization scheme for a band-limited PAM-8 VLC system. A data rate of 1.15-Gb/s is successfully demonstrated over 0.8-m indoor free space transmission based on a single red-LED with BER lower than 3.8×10^{-3}. Besides, it is shown experimentally that LSTM-based equalizer can outperform original FIR based equalizer and Volterra based equalizer. Furthermore, we theoretically and experimentally verified that the scheme is applicable to random sequences and pointed out the causes of the performance overestimation problem of neural network found in previous studies.

References

1. N. Chi, H. Haas, M. Kavehrad, T. Little, and X. Huang, "Visible light communications: demand factors, benefits and opportunities," IEEE Wirel. Commun. **22**(2), 5–7 (2015).
2. Y. Wang, N. Chi, Y. Wang, R. Li, X. Huang, C. Yang, and Z. Zhang, "High-speed quasi-balanced detection OFDM in visible light communication," Opt. Express **21**(23), 27558–27564 (2013).
3. K. Wang, R. Liang, and P. Qiu, "Fluorescence Signal Generation Optimization by Optimal Filling of the High Numerical Aperture Objective Lens for High-Order Deep-Tissue Multiphoton Fluorescence Microscopy," IEEE Photonics J. **7**(6), 1–8 (2015).
4. M. Zhang, M. Shi, F. Wang, J. Zhao, Y. Zhou, Z. Wang, and N. Chi, "4.05-Gb/s RGB LED-based VLC system utilizing PS-Manchester coded Nyquist PAM-8 modulation and hybrid time-frequency domain equalization." Proc. OFC, W2A.42. LA (2017).
5. I. Stefan, H. Elgala, and H. Haas, "Study of dimming and LED nonlinearity for ACO-OFDM based VLC systems," in Wireless Communications and Networking Conference (WCNC). IEEE, 2012, pp. 990–994.
6. N. Chi, M. Zhang, Y. Zhou, and J. Zhao, "3.375-Gb/s RGB-LED based WDM visible light communication system employing PAM-8 modulation with phase shifted Manchester coding," Opt. Express **24**(19), 21663–21673 (2016).
7. Y. Wang, X. Huang, L. Tao, J. Shi, and N. Chi, "4.5-Gb/s RGB-LED based WDM visible light communication system employing CAP modulation and RLS based adaptive equalization," Opt. Express **23**(10), 13626–13633 (2015).
8. G. Stepniak, J. Siuzdak, and P. Zwierko, "Compensation of a VLC phosphorescent white LED nonlinearity by means of Volterra DFE," IEEE Photonics Technol. Lett. **25**(16), 1597–1600 (2013).

9. K. Wang, R. Liang, and P. Qiu, "Fluorescence Signal Generation Optimization by Optimal Filling of the High Numerical Aperture Objective Lens for High-Order Deep-Tissue Multiphoton Fluorescence Microscopy," IEEE Photonics J. **7**(6), 1–8 (2015).
10. S. Liang, Z. Jiang, L. Qiao, X. Lu, and N. Chi, "Faster-Than-Nyquist precoded CAP modulation visible light communication system based on nonlinear weighted look-up table predistortion," IEEE Photonics J. **10**(1), 1–9 (2018).
11. P. Haigh, Z. Ghassemlooy, S. Rajbhandari, and I. Papakonstantinou, "Visible light communications using organic light emitting diodes," IEEE Commun. Mag. **51**(8), 148–154 (2013).
12. X. Lu, K. Wang, L. Qiao, W. Zhou, Y. Wang, and N. Chi, "Non-linear compensation of multi-cap VLC system employing clustering algorithm-based perception decision," IEEE Photonics Journal, **9**(5), 1-9 (2017).
13. J. Zhang, J. Yu, and H. Chien, EML-based IM/DD 400G (4× 112.5-Gbit/s) PAM-4 over 80 km SSMF based on linear pre-equalization and nonlinear LUT pre-distortion for inter-DCI applications. In Optical Fiber Communications Conference and Exhibition (OFC), 2017 (pp. 1–3).
14. D. Angluin, "Queries and concept learning," Mach. Learn. **2**(4), 319–342 (1988).
15. P. Domingos, "A few useful things to know about machine learning," Commun. ACM **55**(10), 78–87 (2012).
16. J. Thrane, J. Wass, M. Piels, J. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," J. Lightwave Technol. **35**(4), 868–875 (2017).
17. F. Khan, C. Lu, and A. Lau, "Machine learning methods for optical communication systems." Signal Processing in Photonic Communications. OSA, 2017.
18. T. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: possible pitfalls," IEEE Photonics Technol. Lett. **29**(23), 2091–2094 (2017).
19. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," Adv. Neural Inf. Process. Syst. **1**, 1097–1105 (2012).
20. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2017).
21. X. Huang, J. Shi, J. Li, Y. Wang, Y. Wang, and N. Chi, "750 Mbit/s visible light communications employing 64 QAM-OFDM based on amplitude equalization circuit," presented at the Opt. Fiber Commun. Conf., 2015, Tu2G.1.