

Machine-Learning Based Equalizers for Mitigating the Interference in Asynchronous MIMO OWC Systems

Yatian Li , Member, IEEE, Tianwen Geng , Ruotong Tian , and Shijie Gao 

Abstract—The error performance of the optical wireless communication (OWC) link suffers from the effects of atmospheric turbulence and pointing errors. The multi-input-multi-output (MIMO) system can combat the damage by transmitting diverse replicas of symbols to the receivers, i.e. the spatial diversity. However, different delays between the transceivers can introduce inter-symbol interference (ISI) which degrades the system performance. The delays during transmission are mainly caused by placing locations and optical path differences. This article proposes algorithms to mitigate the ISI based on machine learning techniques, including both neural networks and the genetic algorithm. In the case of multi-input-single-output (MISO) system, we propose an algorithm based on a bidirectional long short-term memory (LSTM) recurrent neural network (RNN), which works as an equalizer. In the case of single-input-multiple-output (SIMO) system, additional delayers are utilized to align the signals in different apertures. The problem of deducing the values of the delayers is considered as seeking the minimum value in a high-dimensional space. With the help of the genetic algorithm, optimal values of the delayers are maintained, which is named as GAD (genetic algorithm-based delayers). In a more general MIMO case, the GAD and LSTM equalizers are further combined to deal with the ISI issue in the asynchronous MIMO OWC systems, (i.e. GAD-LSTM). Both experimental and simulation results show the remarkable performance improvement of the proposed method over conventional methods.

Index Terms—Interference, MIMO systems, neural networks, optical communication.

I. INTRODUCTION

IN THE past decade, the demand for capacity in wireless links has grown in an explosive manner. This demand has triggered off the optical wireless communications (OWC) due to its potential for bandwidth-hungry applications. The importance

Manuscript received June 4, 2020; revised January 8, 2021 and February 1, 2021; accepted February 2, 2021. Date of publication February 5, 2021; date of current version May 2, 2021. This work was supported by the Research Project of Scientific Research Equipment of Chinese Academy of Sciences. (Corresponding author: Yatian Li.)

Yatian Li, Tianwen Geng, and Shijie Gao are with the Changchun Institute of Optics Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China (e-mail: yt_li@ciomp.ac.cn; gtw525@sina.com; gaoshijie@ciomp.ac.cn).

Ruotong Tian is with the Changchun Institute of Optics Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: tianruotong17@mails.ucas.ac.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2021.3057396>.

Digital Object Identifier 10.1109/JLT.2021.3057396

of OWC lies in the main advantages of low cost, high security, freedom from spectral licensing issues [1]–[4]. Compared with free space coherent optical communication, the intensity modulation and direct detection (IM/DD) system has more simplicity to implement, especially the on-off keying (OOK) system [5], [6]. Despite the bright future, some challenges still remain, especially the fading events. For atmospheric environments, turbulence will lead to power fluctuation and phase distortion in the receiving end, which degrades the link performance [7]–[9]. In addition, there will be inevitable impairments such as terminal-sway, aerosol scattering and pointing errors [10], [11]. As a result, the fading events happen frequently.

To combat these fading events, multi-input-multi-output (MIMO) OWC systems provide diverse replicas of transmitted symbols to the receivers with sufficient separation between each other so that the fading events for each receiver can be considered to be independent of others [12]–[14]. Unlike radio frequency (RF) systems, it has been proved that the repetition codes (RC) outperform orthogonal space-time block codes (OSTBC) in the OWC IM/DD link, although both schemes are able to extract full diversity [15], [16]. Owing to the multiple paths between these transceivers, optical path differences will result in an asynchronous superposition of received signals. That is to say, the inter-symbol interference (ISI) happens. For example, a delay of half symbol (0.5 ns) will occur when the optical path difference is only 15 cm with the data rate of 1 Gbps. This will also be the case in cooperative MIMO systems [17], [18].

To the best of the authors' knowledge, there are three kinds of techniques in the open literature to deal with the issue of ISI in MIMO systems (including both RF and OWC systems), which are frequency domain approaches, time domain solutions and the equalization techniques, respectively. These three kinds of techniques are illustrated as follows. Firstly, the delays in the time domain can be converted to the phases in the frequency domain. A filter with the conjugate frequency response of the channel can compensate the effect of different delays, assuming the delay values and channel gains were available to the receivers [19]. However, the filter with limited registers makes it gapped to the theoretical results. The second method is designing space-time block codes (STBCs), such as linear asynchronous STBCs (LA-STBCs) [20]. LA-STBCs are robust when the propagation delay differences are equal to several symbol durations. Unfortunately, it remains vulnerable to propagation delays that

are not equal to the integral multiples of the single symbol duration. The third solution is to utilize equalizers including linear equalizers and non-linear equalizers [21]–[23]. According to Ref. [24], non-linear equalizers (such as the DFE (Decision Feedback Equalizer), the MLSD (maximum likelihood symbol detection) assisted equalizer, and the MLSE (maximum likelihood sequence estimation) assisted equalizer) outperform the linear ones (such as ZF (Zero-Forcing) equalizers, MMSE (Minimum Mean Square Error) equalizers). Recently, machine learning (ML) based equalizers have been proposed for ISI suppression [24]–[28].

The ML-based equalizers, in principle, are defined by the possible output of the symbol alphabet and a proper nonlinear boundary found in a higher dimensional space. Typical ML methods include multilayer perceptrons (MLP) [25], radial basis functions (RBF) [26], recurrent neural networks (RNN) [27] and deep learning (DL) [28]. One major limitation of conventional RNNs in many practical applications is that they are not able to learn long-term dependencies in data (i.e., dependencies between events that are far apart). To overcome this, RNN with long short-term memory (LSTM) is designed which can model and learn temporal sequences and long-range dependencies more accurately through better storing and accessing of information [29]. Thus, the equalizers based on LSTM have better performance [30], [31].

Generally speaking, there are three motivations for us to employ ML-based equalizers, even when the ISI is modeled as a linear one. Firstly, it's known to all that the nonlinear equalizers (especially MLSE and MLSD) have better performance than linear equalizers, which implies that we may also use ML-based equalizers (also nonlinear equalizers) to have better performance. Secondly, the ML based equalizers can reduce the computational cost of traditional ones, where complex theoretical derivation can be also avoided. Thirdly, the loss function of ML can be set to MSE (mean-squared error) in solving regression problems, which is the same as DFE or MMSE linear equalizers.

Considering the benefits brought by the ML based equalizers, this paper proposes a bidirectional LSTM neural network with delayers based on the genetic algorithm to compensate the interference in asynchronous MIMO OWC links. According to the authors' knowledge, it's a novel attempt to apply LSTM networks in the MIMO OWC system, while most LSTM equalizers are analyzed in single-input-single-output (SISO) scheme in existing literatures. The circumstances of single receiver (MISO) and multiple receivers (single transmitter or multiple transmitters) are discussed in this manuscript respectively. Our main contributions are summarized as follows.

- For the MISO case, an equalizer is proposed based on a bidirectional LSTM recurrent neural network, such that the delays from different transmitting antennas are mitigated.
- For the SIMO (single-input-multiple-output) case, the GAD (genetic algorithm-based delayers) algorithm is also proposed to align the signals in different receiving apertures.
- For a more general MIMO circumstance, the signals in different receiving apertures are first aligned by the

GAD algorithm, then superimposed and transmitted to the LSTM module, which is the GAD-LSTM algorithm. The GAD-LSTM algorithm has a more satisfying performance, which also outperforms the traditional methods.

The structure of this paper is as follows. We formulate the problem of the asynchronous events in MIMO links in Section II, as well as the system model. The MISO case and SIMO/MIMO cases are discussed in Section III-A and Section III-B, where the LSTM and the GAD-LSTM algorithms are illustrated respectively. The indoor experiment and simulation results are given in Section IV. In the end, conclusions are drawn in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a MIMO OWC OOK system with M transmitting apertures and N receiving apertures. When RC scheme is used, the same signals are transmitted in each transmitter. Assume $x(t)$ is the transmitted signal in the t -th time interval. We may assume T to be a unit bit interval. To facilitate discussion, we normalize T to be 1. Then for the j -th receiver, the superimposed signals in the t -th interval $y_j(t)$ is modeled as

$$y_j(t) = \sum_{i=1}^M h_{ij} \cdot x(t - \tau_{ij}), \quad (1)$$

where h_{ij} denotes the channel gain between the i -th transmitter and the j -th receiver. τ_{ij} represents the transmitting delay between the i -th transmitter and the j -th receiver. According to Ref. [32], the channel consists of turbulence (Gamma-Gamma Model), misalignment (Rayleigh Pointing Error Model), and attenuation (constant h_i). The probability density function (PDF) $f_h(h_{ij})$ of the channel gain h_{ij} has the form of Meijer' G function, furnished in Eq. (2).

$$f_h(h_{ij}) = \frac{\alpha\beta\rho^2}{A_0 h_i \Gamma(\alpha)\Gamma(\beta)} \cdot \mathbf{G}_{1,3}^{3,0} \left(\frac{\alpha\beta h}{A_0 h_i} \middle| \rho^2 - 1, \alpha - 1, \beta - 1 \right), \quad (2)$$

where α and β represent the effective number of large and small scale turbulent eddies, respectively. $\Gamma(\bullet)$ is the Gamma function. A_0 denotes the maximum fraction of the collected power in the receiving lens. $\rho = w_{zeq} / 2\sigma_s$ represents the ratio between the equivalent beam radius w_{zeq} and standard deviation σ_s of the pointing errors.

The variables that we are focusing on are $\{\tau_{ij}\}$, $i \in [1, M], j \in [1, N]$, as shown in Fig. 1. τ_{ij} obeys Gaussian distribution. The mean value is equal to the displacements of different transceivers. The random items around the mean value are caused by the optical path differences, which leads from the turbulence. Owing to the quasi-static feature, τ_{ij} can be considered as a constant value during a large amount of frame periods. τ_{ij} can be separated into an integer part v_{ij} and a fractional part ε_{ij} , i.e. $\tau_{ij} = v_{ij} + \varepsilon_{ij}$, where $v_{ij} = \lfloor \tau_{ij} \rfloor$ denotes the round down operation of τ_{ij} .

Suppose that x_k represents the transmitted power for k -th time interval, i.e. $\int_{(k-1)T}^{kT} x(t)dt = x_k \in \{0, 2P_t/R_b\}$, where P_t and R_b stand for the average transmitting power and the data rate, respectively. The electrical output signal y_k^j of photodiode in the

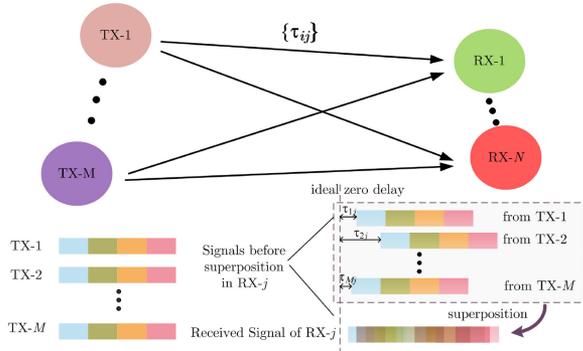


Fig. 1. Signals with different delays in MIMO OWC system.

j -th receiver is depicted as

$$y_k^j = \eta \sum_{i=1}^M h_{ij} [\varepsilon_{ij} \cdot x_{k-v_{ij}-1} + (1 - \varepsilon_{ij}) \cdot x_{k-v_{ij}}] + n_k^j \quad (3)$$

where η presents the responsivity. n_k^j denotes the AWGN (additive white gaussian noise) for the j -th receiver in k -th time slot.

For better performance, it's evident for us to choose the path with the largest gain as the signal part, while the other parts are considered as interference items. We may assume the signal from the l -th transmitter has the largest gain, such that this symbol is considered as a benchmark. In this way, relative delays $\tilde{\tau}_{ij}$ of other signals are equal to $\tau_{ij} - \tau_{lj}$ for an arbitrary integer $i \neq l$ ranging from 1 to M . Thanks to the Gaussian distribution of $\tilde{\tau}_{ij}$ ($i \neq l$), $\tau_{ij} - \tau_{lj}$ also has a Gaussian distribution, while $\tilde{\tau}_{lj}$ is equal to zero. Similar to above, \tilde{v}_{ij} and $\tilde{\varepsilon}_{ij}$ stand for the integer part and fractional part of $\tilde{\tau}_{ij}$. Then y_k^j can be viewed as the sum of the signal item, the ISI items and the noise term, as given in Eq. (4).

$$y_k^j = \underbrace{h_{lj} \cdot x_{k-\tilde{v}_{lj}}}_{\text{signal item}} + \underbrace{n_k^j}_{\text{noise item}} + \underbrace{\sum_{i=1, i \neq l}^M h_{ij} \cdot [\tilde{\varepsilon}_{ij} \cdot x_{k-\tilde{v}_{ij}-1} + (1 - \tilde{\varepsilon}_{ij}) \cdot x_{k-\tilde{v}_{ij}}]}_{\text{ISI items}} \quad (4)$$

From Eq. (4), it's apparent to obtain that the ISI can degrade the system performance in the MISO case. However, we need to face the additional interference in the MIMO case, which reduces the diversity gain.* For arbitrary j_1 -th and j_2 -th receiving apertures ($j_1 \neq j_2$), it's common that $\tau_{ij_1} \neq \tau_{ij_2}$, such that the signal items have a displacement of $|\tau_{ij_1} - \tau_{ij_2}|$ to each other. If the displacement item $|\tau_{ij_1} - \tau_{ij_2}|$ is an integer, the signal part $x_{k-\tilde{v}_{lj_1}}$ and $x_{k-\tilde{v}_{lj_2}}$ can still benefit from the diversity gain from multiple apertures by designing proper STBC, known as the "delay diversity" [20], although there are interference items with

*The SIMO scheme can be considered as a special case of MIMO scheme. Because the signals in an arbitrary receiver still have unequal delays to the signals in other receivers. In a SIMO scheme, we do not need to worry about the interference existing in multiple transmitters, illustrated as the $M = 1$ situation in Eq. (1).

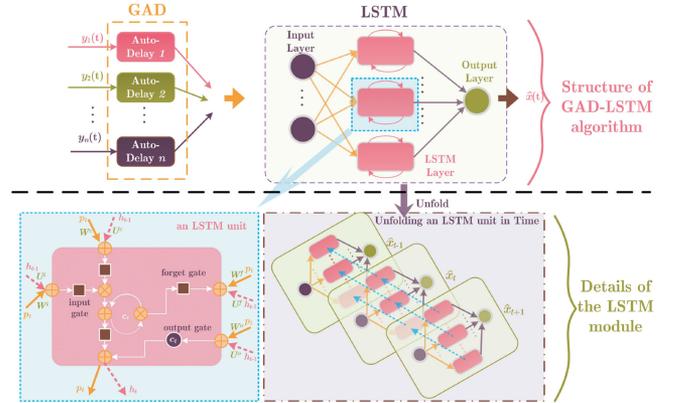


Fig. 2. The structure of the GAD-LSTM algorithm and the details of an LSTM module.

the forms of the second item in Eq. (4). However, the fractional part of $|\tau_{ij_1} - \tau_{ij_2}|$ results in an additional interference to each other.

Here, we give a brief summary of the above formulated problems. In the MISO case, we are facing the problem of the interference from different transmitters as given in Eq. (4). In terms of a system with multiple receivers, the signal parts in different receivers also have relative delays, which will introduce an extra interference. And we will treat the two issues in Section III.

III. GAD-LSTM ALGORITHM

In this section, we will illustrate our LSTM based equalizers to combat the interference of MISO case in Section III-A. The SIMO/MIMO cases are discussed in Section III-B, where both the GAD and GAD-LSTM algorithms are depicted.

A. MISO Case

For the sake of space limitations, we demonstrate the whole structure of the GAD-LSTM algorithm in Fig. 2, and we do not waste pages to depict the structure in the next subsection. For now, we only focus on the LSTM based equalizer, which is the northeast part of Fig. 2. There is no need to utilize the GAD algorithm in the MISO case, due to the fact that there is only one receiver. The southeast part of Fig. 2 shows how the LSTM network is unfolded in time, while the southwest part describes the detailed contents in an LSTM unit.

As seen from Fig. 2, an LSTM unit mainly consists of a forget gate, an input gate, an output gate, and a memory cell. The contents to be stored are determined by input gates into the cell. The forget gate decides whether to reset the cell or not, while the output gate determines the current state and the content to be sent to the next layer. Besides, the cell controls long memory, and the state controls short memory. Benefitting from the cell and these gates, the LSTM unit allows the network to fetch long temporal dependency in a time sequence. Thanks to this unique feature, this paper utilizes the LSTM based equalizers to combat the interference in the systems with multiple apertures.

As a whole, the LSTM network consists of three layers, which are an input layer (p neurons), an LSTM layer (g LSTM units) and an output layer (1 neuron), respectively. It's noted that there is a trade-off for selecting the optimal numbers for input neurons or LSTM units. More input neurons may also equalize the received signals. However, they put more pressure on the computational load. In addition, too small number of units will make it difficult to converge. As a result, we can set the input neurons to be the similar magnitude of M , which can be also concluded from various training results. The number of neurons in the LSTM units may be set to have the same order of magnitude of ones in the input layer, in order to make the network converge. We may assume that \mathbf{I}_k (a p -order column vector) and O_k (a variable) define the input and output of the LSTM network in the k -th time interval. \mathbf{H}_I (a $g \times p$ matrix) and \mathbf{H}_O (a g -order row vector) represent the weights for the input layer and output layer, while \mathbf{J}_I (a p -order column vector) and \mathbf{J}_O (a variable) denote the corresponding biases for the input and output layers. The input of the LSTM layer $\mathbf{p}_k = [p_k^1, p_k^2, \dots, p_k^g]$ (a g -order column vector) and the output of the LSTM network O_k are furnished as

$$\begin{aligned} \mathbf{p}_k &= \Psi(\mathbf{H}_I \cdot \mathbf{I}_k + \mathbf{J}_I) \\ O_k &= \mathcal{L}(\mathbf{H}_O \cdot \mathbf{h}_k + \mathbf{J}_O), \end{aligned} \quad (5)$$

where the subscript \bullet_k stands for an arbitrary k -th interval. $\Psi(\bullet)$ is the logistic sigmoid function, while $\mathcal{L}(\bullet)$ represents the linear function. \mathbf{h}_k (a g -order column vector) denotes the output of an LSTM layer, which is equal to $[h_k^1, h_k^2, \dots, h_k^g]$. It's defined that h_k^q , c_k^q are the state, and cell content of an arbitrary q -th ($1 \leq q \leq g$) individual LSTM unit. These variables are determined by both current input p_k^q and previous state h_{k-1}^q controlled by these three gates.

We also define f_k^q , i_k^q , and o_k^q as the outputs of the forget gate, input gate, and output gate for the q -th LSTM unit, respectively. Eq. (6) depicts how the outputs of the gates and content of the cell and state are calculated.

$$\begin{aligned} f_k^q &= \Psi(W_q^f p_k + V_q^f h_{k-1} + b_q^f) \\ i_k^q &= \Psi(W_q^i p_k + V_q^i h_{k-1} + b_q^i) \\ o_k^q &= \Psi(W_q^o p_k + V_q^o h_{k-1} + b_q^o) \\ c_k^q &= f_k^q \cdot c_{k-1}^q + i_k^q \cdot \tanh(W_q^c p_k + V_q^c h_{k-1} + b_q^c) \\ h_k^q &= o_k^q \cdot \tanh(c_k^q), \end{aligned} \quad (6)$$

where W_q , V_q are the weight values in the arbitrary q -th LSTM unit, and b_q denotes the bias value in the arbitrary q -th LSTM unit. The superscript \bullet^f , \bullet^i , \bullet^o stand for the corresponding parameters for the forget gate, input gate, output gate, respectively.

As a whole, the p inputs of LSTM network is $\mathbf{I}_k = [y_k, y_{k-1}, \dots, y_{k-p+1}]$ from Eq. (4). Their superscript \bullet^j is ignored due to a single receiver, while the output of the LSTM network is the decoded signal \hat{x}_k . As a supervised learning network, the LSTM system chooses the symbols in the training sequence as the target values. MSE is chosen as the loss function. When updating the weights' and biases' values, the Adam optimization is applied in network training process to adjust the learning rate

adaptively, which is a widely adopted extension of the classic stochastic gradient descent algorithm. The maximum training epoch is set to be 60. An early-stopping is declared when six consecutive no-improvement epochs, in order to prevent over-fitting occurring.

B. SIMO/MIMO Case

As mentioned above, we have to face the problem that signals in different receivers are not aligned. For two different j_1 -th and j_2 -th receiving apertures, signal items have a displacement of $|\tau_{ij_1} - \tau_{ij_2}|$ to each other for an arbitrary i -th transmitter. These non-integer displacements make the system suffer from additional ISI. We can align these signals by adjusting the corresponding delays. Due to the quasi-static channel, a natural solution is to employ ML estimation to obtain the values of delays. Let's discuss the ML estimation process for deducing the delay vector $\boldsymbol{\tau}^j = [\tau_{1j}, \tau_{2j}, \dots, \tau_{Mj}]^T$ in an arbitrary j -th receiver, where \bullet^T denotes the transpose operation. Assuming the length of a training sequence is equal to N_{tr} , the likelihood function $P_{\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j}(\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j, \mathbf{x}^{tr})$ is given in Eq. (7), where \mathbf{x}^{tr} denotes the training sequence known to the receivers.

$$\begin{aligned} P_{\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j}(\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j, \mathbf{x}^{tr}) &= \frac{1}{(2\pi\sigma^2)^{N_{tr}/2}} \\ &\cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=0}^{N_{tr}-1} \left[y^j(k) - \sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right]^2 \right\}, \end{aligned} \quad (7)$$

where $\mathbf{y}^j = [y^j(0), y^j(1), \dots, y^j(N_{tr} - 1)]^T$ stands for the received vector of training sequence in the j -th receiver.

The estimation of parameters is obtained by maximizing the log-likelihood function.

$$\begin{aligned} &\max_{\boldsymbol{\tau}^j, \mathbf{h}^j} \left[\log P_{\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j}(\mathbf{y}^j | \mathbf{h}^j, \boldsymbol{\tau}^j, \mathbf{x}^{tr}) \right] \\ &= \max_{\boldsymbol{\tau}^j, \mathbf{h}^j} \left\{ -\frac{N}{2} \log(2\pi\sigma^2) \right. \\ &\quad \left. - \frac{1}{2\sigma^2} \sum_{k=0}^{N_{tr}-1} \left[y^j(k) - \sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right]^2 \right\} \end{aligned} \quad (8)$$

By removing the terms independent of parameters \mathbf{h}^j and $\boldsymbol{\tau}^j$, the above can be further written as:

$$\begin{aligned} &\min_{\boldsymbol{\tau}^j, \mathbf{h}^j} \left\{ \sum_{k=0}^{N_{tr}-1} \left[\left(\sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right)^2 \right. \right. \\ &\quad \left. \left. - 2 \times y^j(k) \left(\sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right) \right] \right\}. \end{aligned} \quad (9)$$

For single transmitter case i.e. $M = 1$, the minima of the above expression can be easily obtained. However, it's beyond the authors' ability to obtain the closed-form solution under the circumstance of $M > 1$. Noting the task has changed into seeking minimum value in a multidimensional space, we utilize the genetic algorithm to achieve all delayers' values $\boldsymbol{\tau}^j$ as well as

the channel states, named as GAD algorithm. It needs to be mentioned that we will also choose the delay value from the transmitter whose path has the best channel condition, which ensures the larger SINR (signal-to-interference plus noise ratio). That is to say, the final delayer's value of the j -th ($j = 1, 2, \dots, N$) receiver is $\hat{G}_j = \hat{\tau}_{ij}$, satisfying $i = \max_i \hat{h}_{ij}$.

Remark 1: The GAD algorithm in solving Eq. (9) is equivalent to the correlation (COR) method in the case of $M = 1$, while the GAD approach outperforms the COR method in the case of $M \geq 2$.

Proof: For keeping a logic consistency, Remark 1 will be proved in the Appendix. ■

After illustrating the principle, we will introduce the GAD algorithm. In general, a genetic algorithm maintains a population of individuals (also called strings or chromosomes) and probabilistically modifies the population by some genetic operators such as selection, crossover and mutation, with the intent of seeking a near-optimal solution to the problem.

The first stage of a GAD is coding and initializing the population. In our scheme, we can initialize the population with M_{pop} individuals near the correlation results in order to achieve faster convergence. For each $\hat{\tau}_{ij}$ (or \hat{h}_{ij}), we choose the N_{cor} intervals ranging from $[-1, 1]$ around N_{cor} locations (or values) of the peaks. Each individual in a population is usually coded as a fixed-length binary string, corresponding to a potential $\hat{\tau}_{ij}$. The length of the string depends on the domain of the parameters and the required precision. For the requirement of six places after the decimal point, the length of the string N_p is supposed to be $N_p = \lceil \log_2(N - 1) + \log_2 N_{cor} + \log_2(1e6) + 1 \rceil$. In each generation for which the GAD is run, each individual in the population is evaluated with the fitness value. Due to the fact that Eq. (9) is seeking the minimum, we set the opposite number of Eq. (9) to be the fitness value \mathcal{F}_n for the n -th iteration, i.e. $\mathcal{F}_n = \sum_{k=0}^{N_{tr}-1} [2y^j(k) \cdot \sum_{i=1}^M h_{ij}x(k - \tau_{ij}) - (\sum_{i=1}^M h_{ij}x(k - \tau_{ij}))^2]$. Due to the mutative channels, it's not a wise choice to setting the threshold of fitness function as the judgment of iteration termination. Thus, we determine whether to terminate by the number of iterations.

In general, there are three genetic operators in the GAD algorithm, which are reproduction probability p_r , crossover probability p_c and mutation probability p_m , respectively. These three genetic operators are illustrated briefly as follows. As an individual is selected, reproduction operator only copy it from the current population into the new population (i.e., the new generation) without alternation. The crossover operator starts with two selected individuals and then the crossover point is selected randomly. The third genetic operator, the mutation, introduces random changes in structures in the population, and it may occasionally have beneficial results: escaping from a local optimum. For clear description of the GAD algorithm, the pseudo-code diagram is depicted in Table I.

As illustrated above, the GAD can be considered as an estimation process. Therefore, the Cramer-Rao bound (CRB) is introduced to depict whether the estimator is efficient. Considering the ISI model of Eq. (3), the CRB of each receiver is the same. The CRB for an arbitrary j -th receiver can be derived

TABLE I
PSEUDO-CODE OF GAD FOR THE j -TH RECEIVER

1	Initialization: Produce initial population with M_{pop} individuals near correlation results. Set the iteration identification flag = 0. Set crossover probability $p_c = 0.2$, mutation probability $p_m = 0.1$. Set the genetic iteration number $G = 0$.
2	while (flag == 0)
3	Reproduction: Calculate individual fitness function \mathcal{F}_n , total fitness values $\sum_{n=1}^{M_{pop}} \mathcal{F}_n$. Update the probability $p_r = \mathcal{F}_n / \sum_{n=1}^{M_{pop}} \mathcal{F}_n$. Choose two arbitrary individuals with the probabilities $\{p_r\}$.
4	Crossover: If rand (0,1) < p_c Swap the lowest 4 bits of the two individuals.
5	Mutation: If rand (0,1) < p_m Randomly flip three binary bits.
6	Updating the population with the two individuals.
7	if $G = 50$ update flag = 1; else update $G = G + 1$.
8	Decode the best individual sets as $\{\hat{\tau}_{ij}\}_{i=1, \dots, M}$ and $\{\hat{h}_{ij}\}_{i=1, \dots, M}$.
9	Output: $\hat{G}_j = \hat{\tau}_{ij}$, satisfying $i = \max_i \hat{h}_{ij}$.

from the inverse of the Fisher-Information Matrix (\mathcal{J}^j), which is known to satisfy the inequality $Cov(\{\tau^j\}) \leq \mathcal{J}^{j-1}$ [33]. The elements of Fisher-Information Matrix can be calculated by Eq. (10).

$$\mathcal{J}_{p,q}^j = \sum_{k=0}^{N_{tr}-1} \frac{\frac{\partial \mathcal{S}_{\tau\tau}(w_k)}{\partial \tau_{pj}} \frac{\partial \mathcal{S}_{\tau\tau}(w_k)}{\partial \tau_{qj}}}{\mathcal{S}_{\tau\tau}^2(w_k)} \quad p, q = 1, 2, \dots, M \quad (10)$$

where $\frac{\partial \mathcal{S}_{\tau\tau}(w)}{\partial \tau_{ij}} = -\sum_{m=1}^M h_{ij}h_{mj} \times \sin[w(\tau_{mj} - \tau_{ij})]$. The $\mathcal{S}_{\tau\tau}(w)$ is the power spectral density (PSD) of received signal.

After the GAD algorithm, the signals in SIMO case can be further processed with traditional combination method. That is to say, there is no need to utilize the LSTM algorithm in the SIMO case, because the interference caused by multiple transmitters does not exist.

In the MIMO case, the delays of signals are first adjusted by GAD, and they are further transported into the LSTM module afterwards, as one can see from Fig. 2. Recall that the LSTM equalizer for the MISO case has p inputs and a single output. The LSTM equalizer for the MIMO case should have $N \times p$ inputs and a single output. In the training process, the original bits of training sequence are chosen as the target value. The remaining details of the MIMO LSTM equalizer is nearly the same as the MISO case, which are no longer repeated here.

IV. NUMERICAL RESULTS

This section shows the numerical results of the GAD-LSTM algorithm. The MISO case and SIMO/MIMO case are discussed respectively.

A. The Numerical Results for MISO Case

As shown in in Fig. 3, our indoor experiment is carried out. The simultaneous channels of the arbitrary wave generator (AWG, Tektronix AWG70002 A) produce the baseband signals for two transmitters. After modulated by Mach-Zehnder

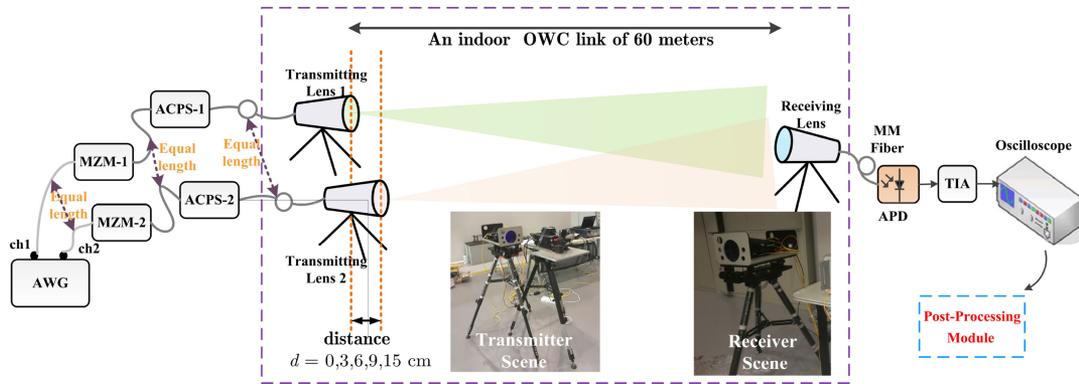


Fig. 3. Experimental setup for the 2×1 MISO case.

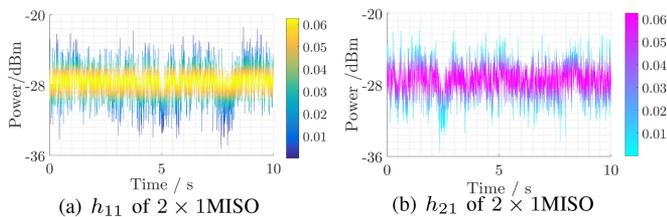


Fig. 4. The received power of two channels in the experiments (a) 1st Channel and (b) 2nd Channel.

Modulator (MZM, Photline MX-LN-10), the two signals are transmitted into the air. The turbulent channels are simulated by the ACPS (atmosphere scintillation playback system), which is elaborated by our own lab. The ACPS has the ability of adjusting the attenuation controlled by a host computer through the Ethernet interface. It can attenuate the input optical signal dynamically ranging from 0 dB to 30 dB, with the minimum time resolution of 1 μ s. Fig. 4(a) and 4(b) shows the sampled power by the receiver when 1st (or 2nd) transmitter is emitting data with the other transmitter powered off. The turbulent channels are referring to the power collected in a horizontal 8.9 km OWC link. Either Figs. 4(a) or 4(b) has a pseudo-color bar implying the similar distribution model. By altering the relative position distance d of the two transmitting apertures, kinds of ISI can be created. With the help of an avalanche photodiode (APD, Lightsensing LSIAPDT-2.5 G), the optical signal coupled into a multi-mode fiber can be changed into electric current. After amplified by the trans-impedance amplifier (TIA, EOCHIP EOC1088), the voltage signal is collected and stored by oscilloscope (Tektronix DPO73304D). The stored data are further transmitted and then processed by a computer with a graphics processing unit (GPU Nvidia GTX1060) for offline digital signal processing. Considering the oscilloscope sampling depth and rate, our signal rate is set to be 1 Gbps. Due to the data rate to be 1 Gbps, 2 ms data (2 M bits) will be obtained by 20 M samples (storage depth) with the sampling rate of 10 Gbps. We may call the action as a single trail. During any arbitrary trail, we further divide 60, 20, 20 percent into training, testing, and verification sets. The displacements d is set to be 3 cm, 6 cm, 9 cm, 15 cm, respectively, while the corresponding delay in the time domain is equal to 1/10, 2/10, 3/10 or 5/10 bit interval.

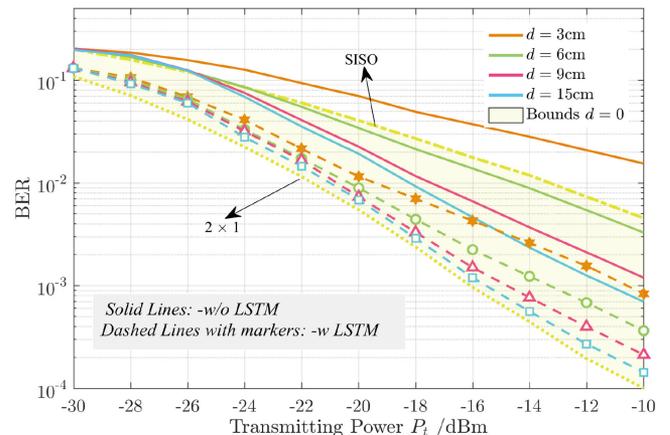


Fig. 5. Experimental BER results for the 2×1 MISO case versus different transmitting power by the LSTM algorithm.

Fig. 5 demonstrates the LSTM equalizer's performance in the 2×1 MISO case. As one can see from Fig. 5, the delays between transmitters are deleterious to the diversity gain achieved by multiple apertures, due to the decrease of the curves' slopes, where the orange, green, purple, and blue lines represent the cases of $d = 3$ cm, 6 cm, 9 cm, 15 cm, respectively. In addition, the bounds of shadowed area are referred to the ideal situation of $d = 0$ in SISO and 2×1 MISO cases, respectively. With regard to the situation of $d = 15$ cm (i.e. half the bit interval), the BER (bit error rate) result is even worse than the SISO case (the dash-dotted yellow line). Compared with the $d = 0$ scheme (dotted yellow line), the LSTM equalizer can behave so close to the border, even though the case of larger d performs a little worse than the ones with smaller d .

In order to compare the LSTM equalizer with other benchmarks, Fig. 6 furnishes the experimental performance of equalizers with LSTM, RNN, MLP, DFE methods for 2×1 MISO system. The number of input neurons is chosen to be 6 in each neural network. All these ML-based equalizers have one hidden layer. The DFE method is set to have 6 feedforward registers and 3 feedback registers, where the coefficients are updated by least-mean-square(LMS) principle. Fig. 6(a) shows the scheme for relative delay $\tau_{21} - \tau_{11} = 0.5T_s$, while Fig. 6(b) discusses the case that relative delay $\tau_{21} - \tau_{11}$ is equal to $6.5T_s$. The

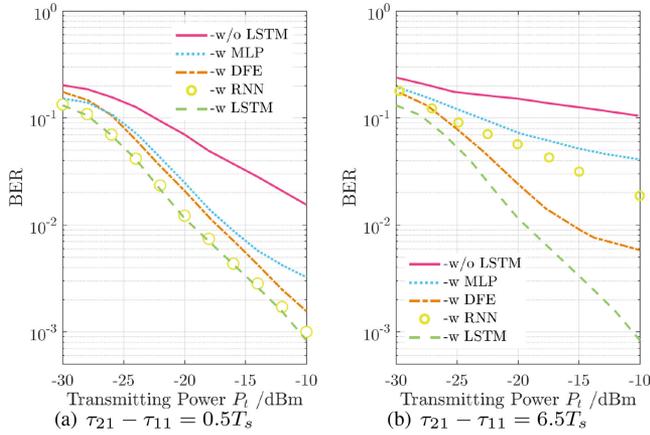


Fig. 6. Experimental results of the LSTM algorithm for a 2×1 MISO system (a) $\tau_{21} - \tau_{11} = 0.5T_s$ (b) $\tau_{21} - \tau_{11} = 6.5T_s$.

delay is simulated for two parts, where $0.5T_s$ is brought by a relative distance $d = 15$ cm. The delay of $6T_s$ is achieved by adjusting the delay of the 2^{nd} output channel of the AWG, because it's difficult to aim at the receiver in the experiment when moving the transmitter by a large amount. For all these two cases, the LSTM equalizer outperforms the other schemes. This difference is more obvious with the increasing P_t . As can be seen from Figs. 6(a) and 6(b), the RNN equalizer outperforms the MLP, which owes to the feedback structure. Besides, the RNN is inferior to the LSTM equalizer for the situation of a large relative delay $\tau_{21} - \tau_{11}$, while it's almost well-matched with the LSTM equalizer in the short delay. It gives the credit to the advantage of learning long-term dependencies in LSTM networks.

B. The Numerical Results for SIMO/MIMO Case

This subsection discusses the numerical results of multiple receivers, where the case of SIMO and MIMO cases are furnished respectively. The evaluation of the SIMO condition can be analyzed on the basis of the above experiment. In the 1×2 SIMO case, the 1^{st} channel of the AWG produces the source signal. After modulated by the MZM, the optical signal is transmitted through the transmitting aperture. Then the two receivers are placed with different relative distances d , which correspond different delays. In this case, the fading channels are simulated by employing two ACPSSs in two fibers of the two receivers. It needs to mention that, there are two types of ACPSSs in the MISO and SIMO experiments, whose interfaces are single mode fibers and multi-mode fibers.

Fig. 7 illustrates the experimental results of the GAD algorithm on estimating the delay $\tau_{12} - \tau_{11}$, where the results by COR method are also given as a contrast. Besides, the GAD performs almost the same as the COR method for the $M = 1$ case (i.e. SIMO), which validates Remark 1. However, the performance by GAD will be significantly improved than COR method with $M \geq 2$, which will be validated by Fig. 8.

As stated above, the GAD algorithm can be considered as an estimator on the delays. Fig. 8 compares the simulation MSE

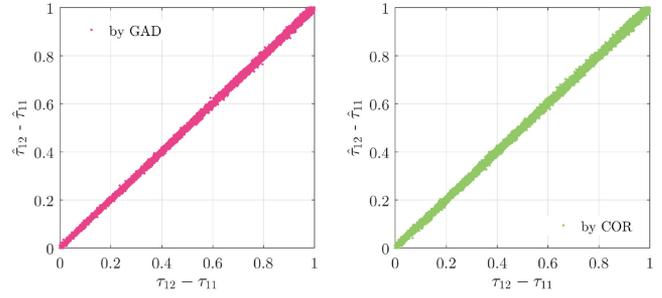


Fig. 7. The experimental results of the estimation on delay $\tau_{12} - \tau_{11}$ by GAD (left) against the COR method (right) in the 1×2 SIMO case.

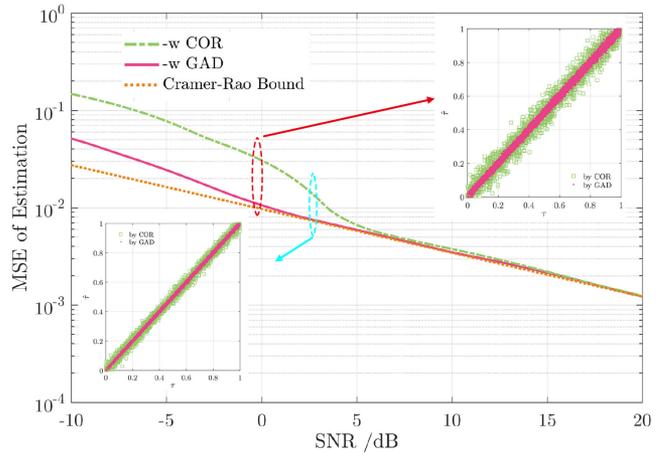


Fig. 8. Comparison of the estimation MSE by GAD or COR with the Cramer-Rao Bound.

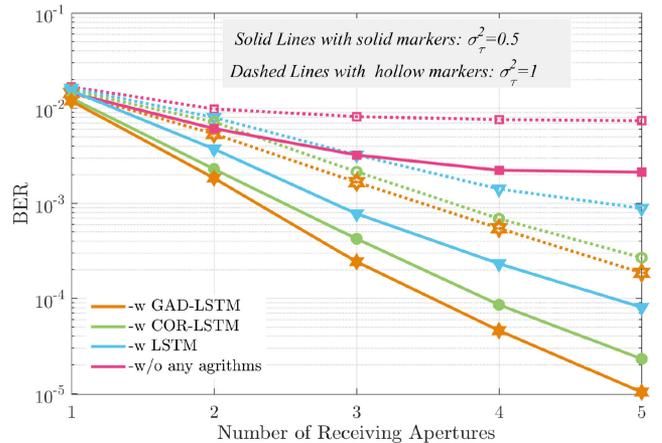


Fig. 9. The simulation performance of the GAD-LSTM algorithm in MIMO cases with different variances $\sigma_\tau^2 = 0.5$ or 1 .

results of GAD algorithm and COR method as well as the CRB, when the number of transmitters is set to be $M = 2$. We choose the PRBS-7 sequence as the training pilot, which has a sharp correlation function. The training sequence's length N_{tr} is set to be 127. Due to the independence of the receivers, we may focus on the delay for any arbitrary receiver. In other words, the experimental scene of 2×1 case can be viewed as any arbitrary receiver of the $2 \times N$ system. As can be seen from Fig. 8, the

GAD approaches to the CRB with SNR = 2.04 dB, while the COR performs almost the same as GAD with SNR > 5.6dB. The GAD outperforms COR method with SNR < 5.6dB, where the gap between the two curves gradually decreases with increasing SNR.

As can be seen from both Fig. 7 and Fig. 8, the signals after the GAD in different receivers can be nearly aligned, which reduces the pressure caused by ISI on LSTM modules. Thus, a more satisfying performance may be gotten in the SIMO and MIMO cases, as depicted below. Due to the limited experimental environment, the results of MIMO scheme are furnished by simulation rather than experiment. Fig. 9 demonstrates the superiority of the GAD-LSTM algorithm, where different variances σ_τ^2 are considered. The solid lines denote the results for $\sigma_\tau^2 = 0.5$ bit interval, while the dotted lines represent the ones for $\sigma_\tau^2 = 1$. It's evident that the different delays $\{\tau_{ij}\}$ deteriorate the BER performance. Additionally, the purple lines (without any algorithms) indicate that the asynchronous issue degrades the diversity gain, since multiple apertures doesn't make the curves drop sharply. What's more, unlike the MISO case, the LSTM equalizers (blue lines) can only provide limited progress for the formidable damage completely caused by τ_{ij} . That is to say, the delayer in each receiver is necessary. As also seen from Fig. 9, the GAD-LSTM outperforms the COR-LSTM algorithm, which verifies the effectiveness of the algorithm.

V. CONCLUSION

This paper focuses on the asynchronous MIMO-OWC channels with diversity mode, where the system suffers from the interference caused by the transceivers' placement and optical path differences. The GAD-LSTM algorithm is proposed to combat the interference, where the circumstances of single receiver and multiple receivers are analyzed respectively. In the case of a single receiving aperture, the LSTM based equalizer outperforms the DFE, MLP, RNN equalizers, whose advantages are more obvious with large delays. In the case of multiple receivers, the GAD algorithm is proposed to align the signals from different receiving apertures. The GAD outperforms the traditional correlation method, which also approaches to the Cramer-Rao bound. With the help of GAD, the LSTM can be further applied to the MIMO scheme, i.e. GAD-LSTM algorithm, which outperforms both the LSTM equalizer and correlation-LSTM equalizer. The experimental and simulation results verify its feasibility and effectiveness.

APPENDIX

In this appendix, we will furnish the proof on Remark 1. The case of $M \geq 2$ is discussed first. Recalling the expression of Eq. (9), the results of GAD are equivalent to that of correlation method only if the following approximation in Eq. (A1) is valid.

$$\sum_{k=1}^{N_{tr}} \left\{ \sum_{l=0}^M \sum_{i=1}^M h_{ij} h_{lj} x(k - \tau_{ij}) \cdot x(k - \tau_{lj}) \right\} \approx 0, \quad (\text{A1})$$

for $i \neq l$

The assumption of Eq. (A1) indicates that the delays τ^j are non-overlapping, which is not practical. However, if we utilize

the assumption of of Eq. (A1), the cross terms of Eq. (9) are ignored and Eq. (9) turns to Eq. (A2).

$$\min_{\tau^j, h^j} \left\{ \sum_{k=0}^{N_{tr}-1} \sum_{i=1}^M (h_{ij} x(k - \tau_{ij}))^2 - 2 \times y^j(k) \left(\sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right) \right\} \quad (\text{A2})$$

By denoting the energy of emitted signal $\mathbb{E}_x = \sum_{k=0}^{N_{tr}-1} (x(k - \tau_{ij}))^2$, the solution of Eq. (A2) can be derived in Eq. (A3).

$$\begin{aligned} & \min_{\tau^j, h^j} \left\{ \mathbb{E}_x \sum_{i=1}^M h_{ij}^2 - 2 \sum_{k=0}^{N_{tr}-1} y^j(k) \times \sum_{i=1}^M h_{ij} x(k - \tau_{ij}) \right\} \\ &= \min_{\tau^j, h^j} \left\{ \sum_{i=1}^M h_{ij}^2 - \frac{2}{\mathbb{E}_x} \sum_{i=1}^M h_{ij} \sum_{k=0}^{N_{tr}-1} x(k - \tau_{ij}) \times y^j(k) \right\} \\ &= \min_{\tau^j, h^j} \left\{ \sum_{i=1}^M h_{ij}^2 - \frac{2}{\mathbb{E}_x} \sum_{i=1}^M h_{ij} \cdot \hat{R}_{xy}(\tau_{ij}) \right\} \end{aligned} \quad (\text{A3})$$

where $\hat{R}_{xy}(\tau_{ij}) = \sum_{k=0}^{N_{tr}-1} x(k - \tau_{ij}) \times y^j(k)$ denotes the correlation function. The estimation of the delay can be obtained, which is $\hat{\tau}_{ij} = \max_{\tau_{ij}} \hat{R}_{xy}(\tau_{ij})$.

However, in most practical cases of asynchronous MIMO systems, the delayers for different transmitters cannot be non-overlapping. If we ignore the assumption in Eq. (A1), the correlation method is a suboptimal one, which performs worse than maximum likelihood estimation. Recalling that our GAD algorithm is the approximation to the maximum likelihood estimation, we may draw the conclusion that our GAD algorithm outperforms the correlation method. The statement can be validated by the estimation variances in Fig. 8.

Now, let's focus on the $M = 1$ case. It's apparent to see that the Eq. (A2) is also equivalent to (9) under the circumstance of $M = 1$, because $[\sum_{i=1}^M h_{ij} x(k - \tau_{ij})]^2$ turns to $h_{1j} x(k - \tau_{1j})^2$.

In summary, Remark 1 has been proved.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially. Also thank Dr. Shaoli Guo for supporting in research work.

REFERENCES

- [1] K. P. Peppas, G. C. Alexandropoulos, E. D. Xenos, and A. Maras, "The Fischer-Snedecor F-distribution model for turbulence-induced fading in free-space optical systems," *J. Lightw. Technol.*, vol. 38, no. 6, pp. 1286-1295, 2019.
- [2] A. Jaiswal and M. R. Bhatnagar, "Free-space optical communication: A diversity-multiplexing tradeoff perspective," *IEEE Trans. Inform. Theory*, vol. 65, no. 2, pp. 1113-1125, Feb. 2019.
- [3] H. Zhou *et al.*, "Orbital angular momentum complex spectrum analyzer for vortex light based on the rotational doppler effect," *Light Sci. Appl.*, vol. 6, no. 16251, pp. 1-8, 2016.
- [4] A. E. Elfiqui, H. S. Khallaf, S. F. Hegazy, A. Elsonbaty, H. M. H. Shalaby, and S. S. A. Obayya, "Chaotic polarization-assisted LDPSK-MPPM modulation for free-space optical communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4225-4237, Sep. 2019.
- [5] C. Gong, K. Wang, Z. Xu, and X. Wang, "On full-duplex relaying for optical wireless scattering communication with on-off keying modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2525-2538, Apr. 2018.

- [6] W. Shin, J. Choi, and S. Han, "Fixed threshold on-off keying differential detection for satellite optical communications," *Opt. Exp.*, vol. 27, no. 2, pp. 1590–1596, 2019.
- [7] M. Elamassie and M. Uysal, "Incremental diversity order for characterization of FSO communication systems over lognormal fading channels," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 825–829, Apr. 2020.
- [8] W. Gappmair and H. E. Nistazakis, "Subcarrier PSK performance in terrestrial FSO links impaired by gamma-gamma fading, pointing errors, and phase noise," *J. Lightw. Technol.*, vol. 35, no. 9, pp. 1624–1632, 2017.
- [9] R. Priyadarshani, M. R. Bhatnagar, Z. Ghassemlooy, and S. Zvanovec, "Outage analysis of a SIMO FSO system over an arbitrarily correlated M. -distributed channel," *IEEE Photon. Technol. Lett.*, vol. 30, no. 2, pp. 141–144, Jan. 2018.
- [10] Y. Li, T. Geng, S. Ma, S. Gao, and H. Gao, "Timing jitter's influence on the symbol error rate performance of the L-ary pulse position modulation free-space optical link in atmospheric turbulent channels with pointing errors," *Opt. Eng.*, vol. 56, no. 3, 2017, Art. no. 036116.
- [11] Y. Li, S. Guo, T. Geng, S. Ma, S. Gao and H. Gao, "Evaluation on the capacity and outage performance of the free space optical system impaired by timing jitters over an aggregate channel," *Opt. Eng.*, vol. 56, no. 7, 2017, Art. no. 076108.
- [12] A. Jaiswal, M. Abaza, M. R. Bhatnagar, and V. K. Jain, "An investigation of performance and diversity property of optical space shift keying-based FSO-MIMO system," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4028–4042, Sep. 2018.
- [13] L. Mroueh, "Extended golden light code for FSO-MIMO communications with time diversity," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 553–563, Jan. 2019.
- [14] Y. Zhang, P. Wang, L. Guo, W. Wang, and H. Tian, "Performance analysis of an OAM multiplexing-based MIMO FSO system over atmospheric turbulence using space-time coding with channel estimation," *Opt. Exp.*, vol. 25, no. 17, pp. 19995–20011, 2017.
- [15] M. Safari and M. Uysal, "Do we really need OSTBCs for free-space optical communication with direct detection?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4445–4448, Nov. 2008.
- [16] Y. Zhang, H. Yu, J. Zhang, Y. Zhu, J. Wang and X. Ji, "On the optimality of spatial repetition coding for MIMO optical wireless communications," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 846–849, May 2016.
- [17] B.-R. Ruben, G.-Z. Antonio, C.-V. Carmen, and C.-V. Beatriz, "Adaptive selective relaying in cooperative free-space optical systems over atmospheric turbulence and misalignment fading channels," *Opt. Exp.*, vol. 22, no. 13, pp. 16629–16644, 2014.
- [18] A. S. Ghazy, H. A. I. Selmy, and H. M. H. Shalaby, "Fair resource allocation schemes for cooperative dynamic free- space optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 11, pp. 822–833, 2016.
- [19] S. Sohaib and D. K. C. So, "Asynchronous polarized cooperative MIMO communication," in *Proc. IEEE Veh. Technol. Conf., VTC*, Barcelona, 2009, pp. 1–5.
- [20] Y. Shang and X. G. Xia, "Limited-shift-full-rank matrices with applications in asynchronous cooperative communications," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4119–4126, Nov. 2007.
- [21] A. Nuwanpriya, S. Ho, J. A. Zhang, A. J. Grant, and L. Luo, "PAM-SCFDE for optical wireless communications," *J. Lightw. Technol.*, vol. 33, no. 14, pp. 2938–2949, 2015.
- [22] J. Fickers, A. Ghazisaeidi, M. Salsi, G. Charlet, P. Emplit, and F. Horlin, "Decision-feedback equalization of bandwidth-constrained N-WDM coherent optical communication systems," *J. Lightw. Technol.*, vol. 31, no. 10, pp. 1529–1537, 2013.
- [23] J. Fan, Y. Ren, Y. Zhang, and X. Luo, "MLSE equalizer with channel shortening for faster-than-Nyquist signaling," *IEEE Photon. Technol. Lett.*, vol. 30, no. 9, 793–796, May 2018.
- [24] F.-L. Luo, *Machine Learning for Future Wireless Communications*. Hoboken, NJ, USA: Wiley, 2019.
- [25] S. Yang, C. Ho, and C. Lee, "HBP: Improvement in BP algorithm for an adaptive MLP decision feedback equalizer," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 53, no. 3, pp. 240–244, Mar. 2006.
- [26] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Computational complexity comparison of feedforward/radial basis function/recurrent neural network-based equalizer for a 50-Gb/s PAM4 direct-detection optical link," *Opt. Exp.* vol. 27, no. 25, pp. 36953–36964, 2019.
- [27] Y. Li, M. Chen, Y. Yang, M. Zhou and C. Wang, "Convolutional recurrent neural network-based channel equalization: An experimental study," in *Proc. IEEE Asia-Pacific Conf. Commun.*, 2017, pp. 1–6.
- [28] H. Ye, G. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [29] T. Ergen *et al.*, "Online training of LSTM networks in distributed systems for variable length data sequences," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 5159–5165, Oct. 2018.
- [30] X. X. Dai, M. Li, Q. Luo You, and S. Yu, "LSTM networks enabled nonlinear equalization in 50-Gb/s PAM-4 transmission links," *Appl. Opt.*, vol. 58, no. 22, pp. 6079–6084, 2019.
- [31] X. Lu *et al.*, "Memory-controlled deep LSTM neural network post-equalizer used in high-speed PAM VLC system," *Opt. Exp.*, vol. 27, no. 5, pp. 7822–7833, 2019.
- [32] M. Uysal, C. Capsoni, Z. Ghassemlooy and A. Boucouvalas, *Optical Wireless Communications—An Emerging Technology*. Switzerland, The Netherlands: Springer, 2017.
- [33] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer, 2010.

Yatian Li (Member, IEEE) received the B.S. and M.S. degrees (with Hons.) in communications engineering from the Harbin Institute of Technology, Harbin, China, and the Ph.D. degree in mechanical and electrical engineering with the University of Chinese Academy of Sciences, Beijing, China, in 2013, 2015, and 2019, respectively. He is a Research Assistant with the Laboratory of Laser Communication, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests include free space optics and wireless communications.

Tianwen Geng received the B.S. and M.S. degrees in electronic circuits and systems from Jilin University, Changchun, China, in 2004 and 2007, respectively. He is currently an Associate Researcher with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His current research interests include free space optics and fiber communication.

Ruotong Tian received the B.S. degree in electric information science and engineering from Jilin University, Changchun, China, in 2017. She is currently working toward the Ph.D. degree in circuits and systems with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, and also with the University of Chinese Academy of Sciences, Beijing, China. Her research interests include coherent detection and digital signal processing.

Shijie Gao received the B.S. degree in mechanical and electrical engineering from the Harbin University of Science and Technology, Harbin, China, the M.S. and Ph.D. degrees in electronic circuits and systems from the University of Chinese Academy of Sciences, Beijing, China, in 2002, 2006, and 2015, respectively. He is currently an Associate Researcher with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His current research interests include free space optics and laser technology.