Transfer-Learning Multi-Input Multi-Output (TL-MIMO) Equalizer for Mode-Division Multiplexing (MDM) Systems

Tianfeng Zhao (赵天烽)[1], Feng Wen (文 峰)[1]*, Mingming Tan[2], Baojian Wu (武保剑)[1], Bo Xu (许 渤)[1], and Kun Qiu (邱 昆)[1]

[1] Key Lab of Optical Fiber Sensing and Communication, Ministry of Education, University of Electronic Science and Technology of China, Chengdu 611731, China

[2] Aston Institute of Photonics Technologies, Aston University, Birmingham B4 7ET, UK
*Corresponding author: fengwen@uestc.edu.cn

Received Month X, XXXX; accepted Month X, XXXX; posted online Month X, XXXX

We propose a transfer-learning multi-input multi-output (TL-MIMO) scheme to significantly reduce the required training complexity for converging the equalizers in mode-division multiplexing (MDM) systems. Based on a built three-mode (LP_{01}, LP_{11a} and LP_{11b}) multiplexed experimental system, we thoughtfully investigate the TL-MIMO performances on the three-typed data, collecting from different sampling times, launched optical powers, and input optical signal-to-noise ratios (OSNRs). The dramatic reduction of 40%~83.33% on the required training complexity is achieved in all of three scenarios. Furthermore, the good stability of TL-MIMO in both the launched power and OSNR test bands has also been proved.

Keywords*:* mode division multiplexing (MDM); multi-input multi-output (MIMO); transfer learning (TL); training complexity
DOI: 10.3788/COLXXXXXX.XXXXXX.

1. Introduction

Spatial division multiplexing (SDM) technology has been widely studied to address the capacity bottleneck faced by the fiber-optic communication system, through the feature of considering different guide modes or fiber cores as independent channels [1-2]. In SDM system, the few-mode fiber (FMF) has been intensively investigated due to its high multiplexing density and easy implementation [3]. However, various mode-dependent distortions existed in FMF, such as the mode-coupling (MC) induced crosstalk, the differential mode group delay (DMGD), the mode-dependent loss (MDL) and the mode-dependent nonlinearity severely degrades the signal quality and limits the system transmission performance [4,5]. In order to improve the transmission performance of the SDM system, the digital signal processing (DSP)-based compensation scheme through multi-input multi-output (MIMO) equalizer is the crucial part at the optical receiver [6,7], which could mitigate the distortions across multi spatial-channels simultaneously. Nevertheless, because of the sensitivity of the MIMO equalizers on the channel status, the retraining operation for MIMO is required for the practical SDM systems when different operational conditions applied, e.g., launched powers, optical signal-to-noise ratios (OSNRs) and time-varying channel status, etc., resulting massive expected computation and thereby hinders the real-time application of the SDM technology.

To reduce the computational complexity, the current work mainly focuses on channel optimization, which involves carefully designing the optical fibers to achieve the MIMO complexity reduction. On one hand, in earlier years, people focused on DMGD management to reduce MIMO complexity. The low MIMO complexity mode-division multiplexing (MDM) transmission was demonstrated with the 10.5 km low-DMGD four-mode fiber [8]. The maximum DMGD over the C+L band was controlled below 50 ps/km. In addition, to achieve more flexible DMGD management, the nearly zero-, positive- and negative-DMGD two-mode fibers were designed and fabricated [9]. On the other hand, reducing the MIMO scale is also an effective solution. For example, through ignoring the inter-mode coupling induced crosstalk in the weak-coupling scenario, the simplified MIMO scheme is only responsible for the intra-mode equalization [10,11]. This method indeed reduces the MIMO scale, but in practical implementation, the accumulated fusion will damage the original weakly-coupling property of the optical fiber.

Transfer-learning (TL) has become a powerful tool for fast modeling of various algorithms with the training process [12]. When new tasks appear, TL exploit the similarity in feature distributions between new and existing tasks, thereby lowering the computational cost required for model reconstruction. In recent published literatures, the TL approach is first implemented in optical networks for optical performance monitoring (OPM), such as the quality of transmission (QoT) prediction in [13], the OSNR monitoring for the 10-span standard single mode fiber (SSMF) [14] and the fast modulation format identification (MFI) [15]. In addition, the TL has also been used in the nonlinearity mitigation algorithms [16-18], by transferring compensation algorithm parameters between the fiber channels with different nonlinear features caused by the transmission length, the launched power, and the optical signal-to-noise ratio (OSNR). Similar to nonlinear fiber channel, the mode-dependent channel state in more sensitive FMF channels is also affected by the above factors. In addition, time-varying nature of FMF also constitutes a crucial factor contributing to the dynamic changes in channel state.

With the increase of the spatial modes considered in MDM system, see approaching 1000 modes already discussed [19], the massive computation demand for data

recovering would be the major issue to this new multiplexing scheme in the real industry implementation. It is necessary to reduce the training complexity required for frequent MIMO remodeling in FMF, the currently most sensitive fiber channel. To the best of our knowledge, there is still no works about using TL in MIMO equalization process for MDM system to reduce the computational complexity. Therefore, we propose the TL aided MIMO (TL-MIMO) equalization scheme to significantly reduce the equalizers' training complexity for the FMF-based transmission system and evaluate its compensation performance in a 3-mode point-to-point experiment system. According to the experimental test, when transferring behavior happens between the data groups collected at different time moments, corresponding to different channel states, the full TL-MIMO method reduces the training complexity by 88.46% compared to training the equalizers from scratch. Moreover, we have also performed the TL-MIMO for different launched optical powers and OSNRs. Consequently, the training complexity is reduced by up to 83.33% with the good stability in the measured ranges. The experimental results indicate that the reduced training complexity, in terms of training times and training data length, improves the channel tracking efficiency of MDM system, which is expected to achieve the real-time equalization even with the large MIMO scale.

2. Transfer learning for MIMO equalization

2.1. Transfer Learning Aided MIMO Scheme

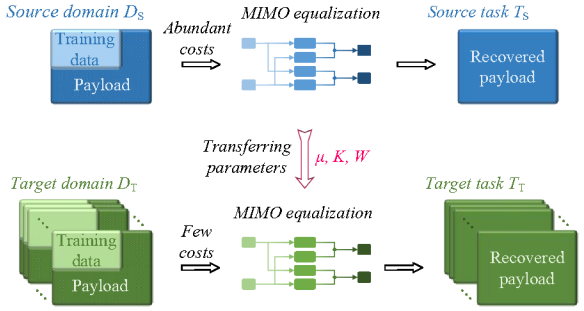

Fig. 1. Schematic design of the proposed TL-MIMO.

We propose the TL-MIMO scheme to reduce the training complexity in equalizer remodeling, the schematic design is depicted in Fig. 1. The data-aided algorithm, frequency-domain least mean square (FD-LMS), is adopted as the MIMO equalization algorithm due to its excellent stability and rapid computation speed. When employing FD-LMS, the received data of the MDM system comprises both training data and payload. The payload is recovered by the equalizers W trained with training data, the optimized step size μ and taps number K. It worth noting that the computation cost of MIMO equalization primarily involves training W and optimizing the μ-K combinations.

As illustrated in Fig. 1, we incorporate the MIMO training process into the TL framework. The upper part represents the source domain D_S and the source task T_S, where D_S comprises a group of received data, T_S is set to recover the payload in D_S, requiring iterative training of W with the training data from D_S to ensure convergence. Then, the lower part of Fig. 1 contains the target domain D_T and the target task T_T of TL-MIMO. Various groups of received data with distinct channel states are placed in D_T, and T_T focuses on recovering the payload in D_T. In the context of FD-LMS, μ, K, and W are the key important parameters, determining the equalizer convergence speed and the BER level [20]. In order to reduce the computation cost required to achieve T_T, by leveraging the similarity in channel states between the received data in D_S and D_T, TL-MIMO transfers these three well-trained parameters from D_S to D_T as initial values, accelerating the equalizers' convergence speed of D_T.

2.2. FD-LMS Algorithm in TL-MIMO

Figure 2 illustrates the structure of 2×2 FD-LMS-based MIMO equalization after transferring the equalization parameters (μ, K, W) from D_S. As indicated by three arrows in Fig. 2, the three parameters respectively affect the equalizers initialization, updating, and the data-block size in MIMO equalization. The steps affected by each parameter are marked with their corresponding colors. Upon algorithm initiation, the two-fold oversampled input sequences x_1 and x_2 undergo serial-to-parallel (S/P) conversion and are then divided into even and odd branches. The sequence in each branch is split into several data blocks of size K, denoted by $x_{1e}(n)$, $x_{1o}(n)$, $x_{2e}(n)$, and $x_{2o}(n)$, where n represents the block index. These blocks are combined into new blocks with 50% overlap-save method and then converted to frequency-domain with fast Fourie transform (FFT) expressed as $X_{im}(n) = FFT[x_{im}(n-1), x_{im}(n)]$, where i and m represent the input port index and even/odd branch, respectively.

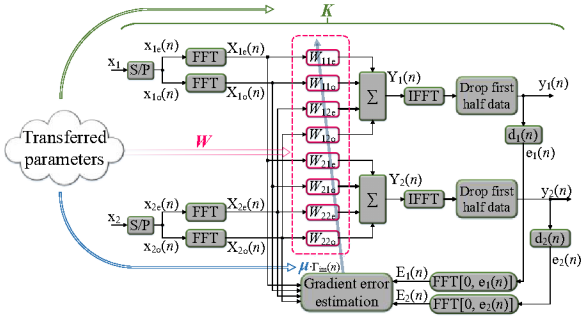

Fig. 2. 2×2 FD-LMS-based MIMO structure combined with TL.

Because the number of equalizers corresponding to each input port is twice the number of input ports, the system with P input ports has $2P^2$ equalizers. In the 2×2 MIMO module, the W matrix consists of 8 FD-equalizers, namely W_{11e}, W_{11o}, W_{12e}, W_{12o}, W_{21e}, W_{21o}, W_{22e}, and W_{22o}. These FD-equalizers, denoted as W_{ijm}, are transformed from their corresponding time-domain (TD)-equalizers w_{ijm}, that is $W_{ijm} = FFT[w_{ijm}, Z(K)]$, where $Z(K)$ is a zero matrix with a size of $1\times K$. Since the sizes of w_{ijm} and W_{ijm} are $1\times K$ and $1\times 2K$, respectively, the size of W should be $2P^2\times 2K$. In TL-MIMO, the transferred W from D_S serves as the initial values of the equalizers in D_T for subsequent training. The output frequency response $Y_i(n)$ is summed from the products of both branches data and the corresponding equalizers, as presented in Eq. (1). The output vector $y_i(n)$ consists of the last K data points from the inverse fast Fourier transform (IFFT) of $Y_i(n)$.

$$Y_i(n) = \sum_{j=1}^{P} X_{je}(n)\cdot W_{ije}(n) + X_{jo}(n)\cdot W_{ijo}(n) \quad (1)$$

The following is the training link. We define $d_i(n)$ as

the n^{th} desired output block of the i^{th} input port. The feedback error vector is then defined as $e_i(n) = d_i(n) - y_i(n)$. After prefixed K zeros, $e_i(n)$ is transformed to frequency-domain error vector $E_i(n)$ by FFT. During the n^{th} training time, the equalizers are updated according to the gradient error estimation method, as shown in Eq. (2):

$$W_{ijm}(n+1) = W_{ijm}(n) + \mu\Gamma_{im}(n) \quad (2)$$

where μ is the transferred step size factor, and $\Gamma_{im}(n)$ is calculated according to Eq. (3). $\tau_{im}(n)$ in Eq. (3) is defined as the first K data in the IFFT results of the product of $X_{im}(n)$ and $E_i(n)$.

$$\Gamma_{im}(n) = \mathrm{FFT}[\tau_{im}(n), Z(K)] \quad (3)$$

In order to save the proportion of training data in the transmission data, we consider the self-recycling training scheme to extend a string of training data several times to fully extract valuable information [21].

We quantify the algorithm complexity by the times of complex multiplication. For each training time in FD-LMS algorithm, each $W_{im}(n)$ undergoes $2K$ complex multiplications with the input signal $X_{im}(n)$, the same as to the length of themselves, to yield the equalized signal. In our work, we set the number of input ports is equal to the excited modes in FMF. For a system with P input ports, it requires $4PK$ multiplications per port. Similarly, in gradient error estimation, the generation of $\tau_{im}(n)$ per port involves an additional $4PK$ complex multiplications. In addition, during the equalization, the data of each port is successively processed 8 times by FFT/IFFT, resulting $8K\log_2(2K)$ times multiplication. Therefore, for a single training time, the number of complex multiplications C_{single} in FD-LMS-based MIMO equalization system can be expressed as Eq. (4).

$$\begin{aligned} C_{single} &= P\times[4PK + 4PK + 8K\log_2(2K)] \\ &= 8P^2K + 8PK\log_2(2K) \end{aligned} \quad (4)$$

Moreover, the total number of training times, T_{total}, can be expressed as $T_{total} = T_{recycling} \times L_{data} / K = C_{train} / K$, where $T_{recycling}$ and L_{data} respectively represent the self-recycling training times and the length of training data. Here, $C_{train} = T_{recycling} \times L_{data}$ is defined as the training complexity. Therefore, the overall training complexity of FD-LMS-based MIMO equalization system, denoted as C_{total}, is expressed in Eq. (5). This equation indicates that, while maintaining signal quality requirements, effective reduction of training complexity can be achieved through the optimization of training data length and training times.

$$\begin{aligned} C_{total} &= T_{total} \times C_{single} \\ &= C_{train} / K \times [8P^2K + 8PK\log_2(2K)] \\ &= T_{recycling} \times L_{data} \times [8P^2 + 8P\log_2(2K)] \end{aligned} \quad (5)$$

For the payload in D_T, let $T_{retrain}$ and $L_{retrain}$ denote the training times and training data length needed for training from scratch to achieve a specific BER, respectively. Under the same BER requirement, the necessary training times and training data length for using TL-MIMO are denoted as T_{TL} and L_{TL}, respectively. The saved training complexity C_{saved} by TL-MIMO is expressed as Eq. (6). In Section 4, we investigate the performance of TL-MIMO through evaluating the reduction of training complexity.

$$C_{saved} = (T_{retrain} \times L_{retrain} - T_{TL} \times L_{TL}) / (T_{retrain} \times L_{retrain}) \quad (6)$$

3. Experiment setup

To verify the performance of the proposed TL-MIMO scheme, we built up an MDM coherent transmission system based on a 12 km-length three-mode fiber, depicted in Fig. 3. Firstly, the two independent electrical non-return-to-zero (NRZ) signals generated by an arbitrary waveform generator (AWG) were used to drive an optical I/Q modulator (IQM). A continuous-wave (CW) light of 1552.52 nm as the optical carrier was also injected into the IQM. The data rate of the optical QPSK was 10 Gbit/s. An erbium-doped fiber amplifier ($EDFA_1$) as well as the variable optical attenuator (VOA_1) was responsible to adjust the power level of the optical signal. In order to achieve the OSNR control in the system, an amplified-spontaneous emission (ASE) source was used to generate the noise, which coupled with the signal through the first optical coupler (OC_1) with the coupling ratio of 99:1. And we monitored the input OSNR through the optical spectrum analyzer (OSA) received from the 1% port of OC_1. Then, the distorted signal was divided into three branches through three 3dB optical-couplers (OCs), i.e., OC_2, OC_3, and OC_4. In each path, the VOAs were used to ensure the same input power for each branch before injecting into the mode-selective photonic lantern ($MSPL_1$). Due the polarization sensitivity of the few-mode system [22], the single-mode polarization controllers (PCs) were employed to maintain the few-mode system in a favorable polarization state and ensure similar output power for signals in each mode after demultiplexing. In order to de-correlate the three channels, the different delay was applied through three pieces of the single-mode-fibers with different lengths placed at the input port of $MSPL_1$. Consequently, the data frames transmitted by the other two branches has different symbol offsets compared to the first branch. The mode conversion was happened at the $MSPL_1$, which also connected with the three-mode transmission fiber through its few-mode pigtail. The same three-mode graded-refractive-index fiber for the MSPL and the FMF channel was produced by YOFC [23]. The three-mode signals, LP_{01}, LP_{11a} and LP_{11b} propagated through the 12km-length fiber and injected into $MSPL_2$. Then, three single-mode signals were obtained after the $MSPL_2$ and detected by a coherent receiver (Co-Rx) separately. We used an optical switch (OS) to select the data each time for one channel. And we calculate the relative symbol offsets among the channels based on the length differences of the delay lines placed before $MSPL_1$. The header of each frame was defined by cropping the respective symbol offset from the original received frame of each branch. The real-time sampling was performed by an oscilloscope with the sampling rate of 50 GSa/s.

During the transmission in the FMF, besides the inherent distortion, such as the attenuation, the dispersion, that already existed in the SSMF, the major degradation coming from the mode-dependent distortions (e.g., MC and DMGD.). So as to mitigate the impact from the mode-related distortions, the MIMO equalizer is the important module in the DSP algorithms for MDM systems. In Fig. 3, the offline DSP processing flow is illustrated,

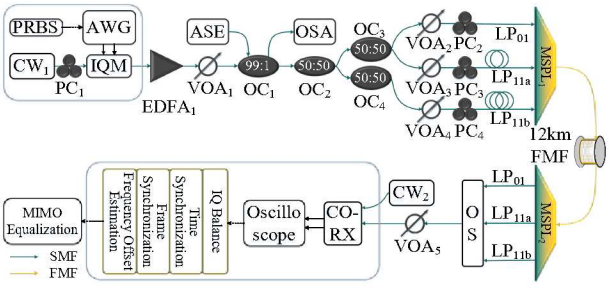

Fig. 3. Coherent three-mode multiplexed experimental setup.

wherein the signal successively undergoes IQ balancing, Gardner algorithm-based time synchronization [24], Schmidl algorithm-based frame synchronization [25], and quartic method-based frequency offset estimation [26] before being processed by the MIMO module.

4. Results and discussion

In order to investigate the performance of the proposed TL-MIMO scheme, we have measured the experimental results in three scenarios, i.e., different time moments, different launched power, and different OSNR, and investigated the complexity reduction for each case.

4.1. TL-MIMO between different time moments

Just like the conventional SSMF-based communication system, the system performance is variable due to the slowly-varying factors such as the state-of-polarization (SOP) evolution, which results into the equalizer retraining at the receiver. During parallel transmission in the FMF channel, time-dependent distortions, such as crosstalk induced by MC, can result in fluctuations in transmission performance [27,28]. Therefore, in the MDM system, equalizer remodeling is an important step to adapt the latest transmission conditions. And it is necessary to investigate the features of the TL-MIMO scheme in this most practical scenario.

In this scenario, we have calculated the BER results of the data collected at different time moments, the results with retraining and two TL-MIMO methods are depicted in Fig. 4. For this test, the data-length of each frame was 81920, the launched power and the input OSNR of the signals in both D_S and D_T were -4 dBm and 25 dB, respectively. Firstly, taking a set of data as an example, Fig. 4 (a) depicts the BER results with the number of the self-recycling times $T_{recycling}$ from 0 to 40 and the fixed length of training data L_{data} of 16384. The blue line represents the process of the training from scratch. The yellow line with the legend of TL w/o W (partial TL-MIMO method), means that the transferred parameters are the optimal μ-K combination of D_S, but the values of W in D_T are initialized as zeros. The red line corresponds to the full TL-MIMO method. In this method, besides the optimal μ-K combination, the values of the well-trained W in D_S are also transferred to the target training system. In addition, since the signals of three modes are generated by the same data source, the output three-channel signals are consistent after MIMO equalization. We can find that the BER trained from scratch requires at least 2 self-recycling times to meet the 7% HD-FEC threshold of 3.8×10^{-3}, but whether the partial or full TL-MIMO method, only one self-recycling time is sufficient to meet this requirement. Therefore, according to Eq. (6), TL-MIMO reduced the training times by 50% with both lower initial BER and faster convergence speed compared to training from scratch. According to Fig. 4 (a), the red dot-line initially decreased to its lowest point and then started increasing again with the increasing $T_{recycling}$, and the final BER was slightly higher than that of the retrain scenario. Two primary reasons account for this phenomenon: (1) During equalizer training, the overfitting of the equalizers led to degraded BER results; (2) The transferred μ-K combination is not optimal for the payload in D_T, thus the BER results with TL-MIMO after sufficient training were slightly higher. As mentioned in Section 2, locating the optimal μ-K combination for each group of data requires multi calls to the equalization algorithm, resulting in significant computational costs. TL-MIMO scheme can avoid such computational cost with slight and affordable BER degradation. In addition, we have also demonstrated the constellations for the three methods with the equalized BERs of around 1.7×10^{-4}. The similar constellation results indicate TL-MIMO has non-significant impact on the signal features.

The length of training data is another determining factor of the training complexity C_{train} as discussed in Section 2. The BER results with L_{data} are depicted in Fig. 4 (b). The measured L_{data} length was from 1024 to 16384 with a step of 1024 and each point in the figure was trained with 40 self-recycling times. In comparison to the necessary 4096 training data for the retraining process to meet the 7% FEC threshold, the shortest length of training data in TL-MIMO, which is 2048, results in a 50% reduction in training complexity.

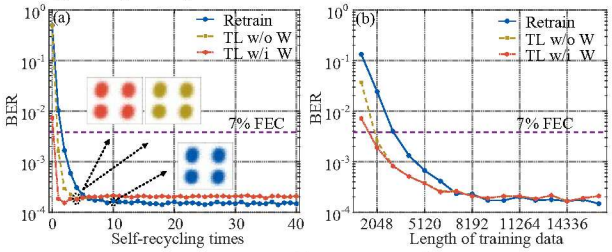

Fig. 4. BER results of D_T versus (a) $T_{recycling}$ and (b) L_{data} with TL-MIMO, while the data of D_T and D_S is collected at different moments.

In FMF, the evolution speed of channel state is one order of magnitude faster than that in SSMF [29]. To verify the time-varying characteristic of the experimental system built in Section 3, we demonstrate the BER distributions on the μ-K plane for the payload in D_S and D_T, both under the same equalization conditions, as shown in Figs. 5 (a) and 5 (b). The diversity in BER distribution proves the channel state differences between D_S and D_T.

To investigate the relationship between BER and the parameters $T_{recycling}$ and L_{data} both before and after applying TL-MIMO more accurately, we monitored the BER results with varying L_{data} and $T_{recycling}$. The results, both without and with TL-MIMO, are depicted in Figs. 6 (a) and 6 (b), respectively. In these figures, color highlights areas where the BER is less than 3.8×10^{-3}. As per Eq. (5), training complexity C_{train} is defined as the product of $T_{recycling}$ and L_{data}. Comparing Figs. 6 (a) and 6 (b), it's evident that the colored area extends toward lower C_{train} values when the TL-MIMO scheme is applied. Specifically, the lowest C_{train}

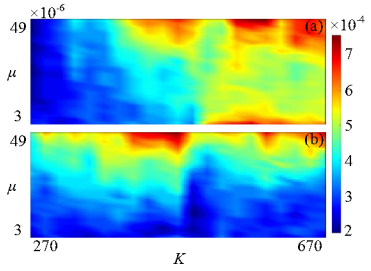

Fig. 5. BER versus μ-K combination in (a) D_S, and (b) D_T with the data collected at different moments.

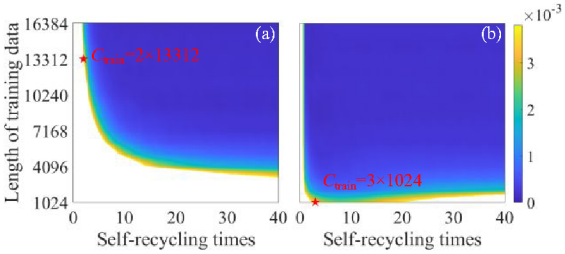

Fig. 6. BER vs. $T_{recycling} \times L_{data}$, illustrating (a) without and (b) with the full TL-MIMO method.

through training from scratch was 26624 ($T_{recycling}$=2 and L_{data}=13312), marked in Fig. 6 (a) with a red pentagram. However, the TL-MIMO scheme achieved a minimum C_{train} of only 3072, marked in Fig. 6 (b), corresponding to $T_{recycling}$=3 and L_{data}=1024. This represents an 88.46% reduction in C_{train} achieved through TL-MIMO.

We have also conducted a comparison between the TL-MIMO scheme and the variable step size FD-LMS (VSS-FD-LMS) algorithm [30], which is designed to accelerate equalizers convergence through optimizing step size during training. The statistical results of 30 sets of independently received data are presented in Fig. 7, with dot-lines representing average BER results and shadows covering the range of tested results. In Fig. 7 (a), compared to the VSS algorithm, the TL-MIMO scheme exhibits smaller initial BER values and converges to minimum values at a faster pace. Taking the average BER result as an example, the TL-MIMO scheme in Fig. 7 (b) achieves a 60% reduction in L_{train} compared to training from scratch and a 50% reduction compared to VSS algorithm. These findings indicate the effectiveness of TL-MIMO in reducing C_{train} in MIMO equalization.

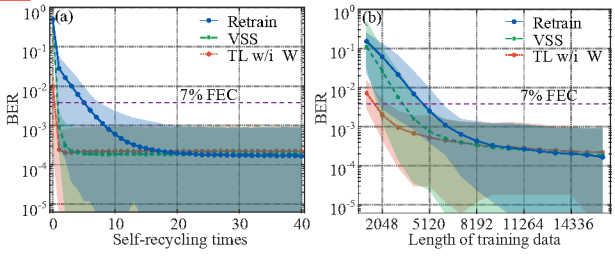

Fig. 7. Statistical BER results for 30 data sets in D_T comparing different MIMO schemes with respect to (a) $T_{recycling}$ and (b) L_{data}.

4.2. TL-MIMO between different launched powers

In this subsection, we measured the TL-MIMO results between different launched optical powers. In the experiment, the launched optical power was adjusted from -15 dBm to -4 dBm with the fixed input OSNR of 25 dB.

Figure 8 demonstrates the BER versus $T_{recycling}$ and L_{data} in two transferring cases between different launched powers, i.e., transferring the equalization parameters from the lower launched power of -10 dBm to the higher launched power of -4 dBm (L-H case), and from -10 dBm to -15 dBm (H-L case). The results in Fig. 8 are statistically counted from 30 groups of data. For the L-H case, as depicted in Figs. 8 (a) and 8 (b), $T_{recycling}$ and L_{data} were reduced by 83.33% and 60%, respectively, by the full TL-MIMO method. And for the H-L case, as depicted in Figs. 8 (c) and 8 (d), the $T_{recycling}$ and the L_{data} were also significantly reduced by 80% and 40%, respectively.

To investigate more detailed TL-MIMO performance for the scenario of the transferring parameters from different launched powers, we have transferred the equalization parameters from the well-trained equalizers with different launched powers, i.e., with different transfer powers P_t, to other power points of the test band, and depicted the results in Fig. 9. The green, red, and yellow markers respectively represent the results with P_t of -15, -10, and -4 dBm. In the test, the 16384 training-data and 40 self-recycling times were applied to all the datasets. Through using the full TL-MIMO equalization, all experimental results in the testing range of -15 dBm to -4 dBm were below the FEC threshold. The similar BER curves obtained by the three transferring cases further confirms the stability of the proposed TL-MIMO scheme. A small BER gap between the case of retraining and TL-MIMO came from the inherent optimal-value issue.

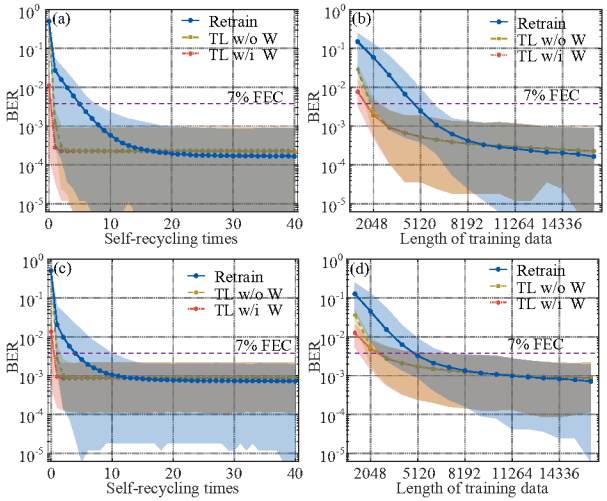

Fig. 8. In L-H case of launched power, BER results versus (a) $T_{recycling}$ and (b) L_{data}; in H-L case of launched power, BER results versus (c) $T_{recycling}$ and (d) L_{data} for 30 groups of data in D_T.

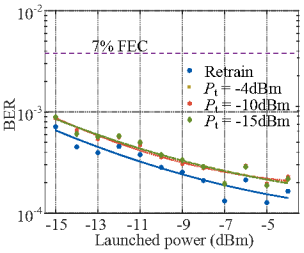

Fig. 9. BER vs. launched power in D_T with different P_t.

4.3. TL-MIMO between different OSNRs

In fiber-optic transmission system, OSNR is regarded as a significant parameter to directly judge the channel quality in the optical domain. Hence, it is necessary to analyze the compatibility of TL-MIMO scheme under

different input OSNRs. In the experimental setup of Section 3, we could adjust the OSNR of input signals through an ASE source and a power controllable device, i.e., VOA_1 in the experiment. The measured OSNR was in the range of 15 dB to 25 dB with the launched power of -4 dBm. Like in the previous subsection measured the results in two transferring cases, we transferred parameters here from the case of input OSNR of 20 dB to the case of 25 dB (L-H case), or to the case of 15 dB (H-L case), also representing the two-way transferring behavior.

The statistical BER results for the two OSNR transferring cases were depicted in Figs. 10 (a) to 10 (d). The results prove that TL-MIMO has a significant improvement on reducing training complexity. For the case that transferring from the OSNR of 20 dB to 25 dB, 83.33% of the $T_{recycling}$ and 60% of the L_{data} were saved by the full TL-MIMO. And for another case, the $T_{recycling}$ and the L_{data} were respectively save by 83.33% and 50%. Additionally, in the scenarios corresponding to Figs 8 and 10, TL-MIMO remains more efficient in our measurements compared to the VSS-FD-LMS algorithm.

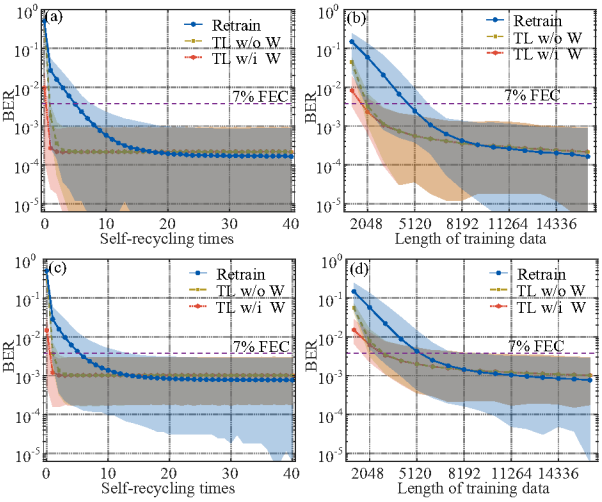

Fig. 10. BER results in L-H OSNR case, vs. (a) $T_{recycling}$ and (b) L_{data}; in H-L OSNR case, vs. (c) $T_{recycling}$ and (d) L_{data} for 30 groups of data in D_T.

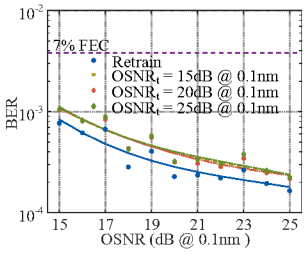

Fig. 11. BER vs. OSNR in D_T with different $OSNR_t$.

Figure 11 demonstrates the relationship between BER and OSNR in different cases where the parameters transferred from the OSNR of 15 dB, 20 dB and 25 dB to each OSNR-testing point, see the yellow, red and green markers. The blue circulars are the tested results from the retraining MIMO. For each marker, it represents the average value of 30 BER-data. Similar to Fig. 9, the TL-MIMO scheme for the three transferring cases shows the similar results, confirming the stability of the proposed method used in the OSNR scenario. As the other two scenarios, the small BER-penalty is also observed.

Conclusion

We proposed TL-MIMO scheme to reduce the equalizer training complexity in MDM transmission system. Based on a 3-mode multiplexed experimental system with 12 km few-mode fiber, we carried out the investigation in three scenarios, i.e., transferring parameters between different time moments, launched optical powers and input OSNRs to verify the performance of TL-MIMO scheme. Experimental results show that under the premise of achieving the 7% HD-FEC threshold, 88.46% training complexity was reduced with the full TL-MIMO method in the first scenario. In the latter two scenarios, the training complexity was reduced by up to 83.33% and at least 40% according to the average results of 30 groups of data. Additionally, TL-MIMO exhibited robust stability and compatibility across different launched power and OSNR testing bands.

Acknowledgement

This work has been supported by the National Key R&D Program of China (2018YFB1801001) and Royal Society International Exchange Grant (IEC\NSFC\211244).

References

1. J. Du, W. Shen, and J. Liu et al., Chin. Opt. Lett., **19** (2021).
2. B. J. Puttnam, G. Rademacher, and R. S. Luís, Optica, **8**, 1186-1203 (2021).
3. G. Rademacher, R. S. Luís, and B. J. Puttnam et al., OFC, paper W4C.3 (2018).
4. K.P. Ho and J. M. Kahn, J. Lightwave Technol., **32**, 614-628 (2014).
5. F. M. Ferreira, C. S. Costa, and S. Sygletos et al., J. Lightwave Technol., **35**, 4011-4022 (2017).
6. P. J. Winzer and G. J. Foschini, Opt. Express, **19**, 16680-16696 (2011).
7. K. Shibahara, T. Mizuno, and D. Lee et al., J. Lightwave Technol., **36**, 336-348 (2018).
8. T. Mori, T. Sakamoto, and M. Wada et al., OFC, paper OTh3K.1 (2013).
9. R. Maruyama, N. Kuwaki, and Shoichiro Matsuo et al., Opt. Express, **22**, 14311-14321 (2014).
10. D. Soma, S. Beppu, and Y. Wakayama et al., J. Lightwave Technol., **36**, 1375-1381 (2018).
11. S. Beppu, D. Soma, and S. Sumita, et al., " J. Lightwave Technol., **38**, 2835-2841 (2020).
12. S. J. Pan and Q. Yang, IEEE Transactions on Knowledge and Data Engineering, **22**, 1345-1359 (2010).
13. W. Mo, Y. Huang, and S. Zhang et al., OFC, paper W4F.3 (2018).
14. L. Xia, J. Zhang, and S. Hu et al., Opt. Express, **27**, 19398-19406 (2019).
15. Y. Cheng, W. Zhang, and S. Fu et al., Opt. Express, **28**, 7607-7617 (2020).
16. J. Zhang, L. Xia, and M. Zhu et al., Opt. Lett., **44**, 4243-4246 (2019).
17. W. Zhang, T. Jin, and T. Xu et al., ACP/IPOC, paper M4A.321 (2020).
18. P. J. Freire, D. Abode, and J. E. Prilepsky et al., J. Lightwave Technol., **39**, 6733-6745 (2021).
19. F. M. Ferreira and F. A. Barbosa. arXiv:2206.09855 (2022).
20. B. F. Boroujeny and K. S. Chan, IEEE Transactions on Signal Processing, **48**, 2332-2342 (2000).
21. T. Zhao, F. Wen, and F. Li et al., OSA OECC, paper. M3F.5 (2021).
22. M. Arikawa and T. Ito, Opt. Express, **26**, 28263-28276 (2018).
23. https://en.yofc.com/view/2351.html
24. F. Gardner, IEEE Trans. on Comm., **34**, 423-429 (1986).
25. T. M. Schmidl and D. C. Cox, IEEE Trans. on Comm., **45**, 1613-1621 (1997).
26. A. Leven, N. Kaneda, and U. -V. Koc *et al.*, IEEE Photonics Technology Letters, **19**, 366-368 (2007).
27. M. Nakazawa, M. Yoshida, and T. Hirooka, Opt. Express, **22**, 31299-31309 (2014).
28. T. Zhao, F. Wen, and K. Qiu, PIERS (2021).
29. X. Chen, J. He, and A. Li et al., IEEE Photon. Technol. Lett., **25**, 1819-1822 (2013).
30. J. Qin, and J. Ouyang. Journal of Data Acquisition & Processing, **12**, 171-194 (1997).

References with full name

1. J. Du, W. Shen, and J. Liu et al., "Mode division multiplexing: from photonic integration to optical fiber transmission," Chin. Opt. Lett., vol 19, 2021.
2. B. J. Puttnam, G. Rademacher, and R. S. Luís, "Space-division multiplexing for optical fiber communications," Optica, vol. 8, pp. 1186-1203, 2021.
3. G. Rademacher, R. S. Luís, and B. J. Puttnam et al., "93.34 Tbit/s/mode (280 Tbit/s) transmission in a 3-mode graded-index few-mode fiber," OFC, 2018, paper W4C.3.
4. K.P. Ho and J. M. Kahn, "Linear propagation effects in mode-division multiplexing systems," J. Lightwave Technol., vol. 32, pp. 614-628, 2014.
5. F. M. Ferreira, C. S. Costa, and S. Sygletos et al., "Semi-analytical modelling of linear mode coupling in few-mode fibers," J. Lightwave Technol., vol. 35, no. 18, pp. 4011-4022, 2017.
6. P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," Opt. Express, vol. 19, pp. 16680-16696, 2011.
7. K. Shibahara, T. Mizuno, and D. Lee et al., "Advanced MIMO signal processing techniques enabling long-haul dense SDM transmissions," J. Lightwave Technol., vol. 36, pp. 336-348, 2018.
8. T. Mori, T. Sakamoto, and M. Wada et al., "Low DMD four LP mode transmission fiber for wide-band WDM-MIMO system," OFC, 2013, paper OTh3K.1.
9. R. Maruyama, N. Kuwaki, and Shoichiro Matsuo et al., "Two mode optical fibers with low and flattened differential modal delay suitable for WDM-MIMO combined system," Opt. Express, vol. 22, pp. 14311-14321, 2014.
10. D. Soma, S. Beppu, and Y. Wakayama et al., "257-Tbit/s weakly coupled 10-mode C+L-band WDM transmission," J. Lightwave Technol., vol. 36, pp. 1375-1381, 2018.
11. S. Beppu, D. Soma, and S. Sumita, et al., "402.7-Tb/s MDM-WDM transmission over weakly coupled 10-mode fiber using rate-adaptive PS-16QAM signals," J. Lightwave Technol., vol. 38, pp. 2835-2841, 2020.
12. S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.
13. W. Mo, Y. Huang, and S. Zhang et al., "ANN-based transfer learning for QoT prediction in real-time mixed line-rate systems," OFC, 2018, paper W4F.3.
14. L. Xia, J. Zhang, and S. Hu et al., "Transfer learning assisted deep neural network for OSNR estimation," Opt. Express, vol. 27, pp. 19398-19406, 2019.
15. Y. Cheng, W. Zhang, and S. Fu et al., "Transfer learning simplified multi-task deep neural network for PDM-64QAM optical performance monitoring," Opt. Express, vol. 28, pp. 7607-7617, 2020.
16. J. Zhang, L. Xia, and M. Zhu et al., "Fast remodeling for nonlinear distortion mitigation based on transfer learning," Opt. Lett., vol. 44, pp. 4243-4246, 2019.
17. W. Zhang, T. Jin, and T. Xu et al., "Nonlinear mitigation with TL-NN-NLC in coherent optical fiber communications," ACP/IPOC, 2020, paper M4A.321.
18. P. J. Freire, D. Abode, and J. E. Prilepsky et al., "Transfer learning for neural networks-based equalizers in coherent optical systems," J. Lightwave Technol., vol. 39, pp. 6733-6745, 2021.
19. F. M. Ferreira and F. A. Barbosa. "Towards 1000-mode optical fibres," arXiv:2206.09855, 2022.
20. B. F. Boroujeny and K. S. Chan, "Analysis of the frequency-domain block LMS algorithm," IEEE Transactions on Signal Processing, vol. 48, no. 8, pp. 2332-2342, 2000.
21. T. Zhao, F. Wen, and F. Li et al., "Noise compensation for a nonlinear six-mode fiber channel through self-recycling training equalizer," OSA OECC, 2021, paper. M3F.5
22. M. Arikawa and T. Ito, "Performance of mode diversity reception of a polarization-division-multiplexed signal for free-space optical communication under atmospheric turbulence," *Opt. Express* 26, 28263-28276 (2018).
23. https://en.yofc.com/view/2351.html
24. F. Gardner, "A BPSK/QPSK Timing-Error Detector for Sampled Receivers," *IEEE Transactions on Communications*, vol. 34, no. 5, pp. 423-429, 1986.
25. T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613-1621, Dec. 1997.
26. A. Leven, N. Kaneda, U. -V. Koc and Y. -K. Chen, "Frequency Estimation in Intradyne Reception," *IEEE Photonics Technology Letters*, vol. 19, no. 6, pp. 366-368, March15, 2007.
27. M. Nakazawa, M. Yoshida, and T. Hirooka, "Measurement of mode coupling distribution along a few-mode fiber using a synchronous multi-channel OTDR," Opt. Express, vol. 22, no. 25, pp. 31299-31309, 2014.
28. T. Zhao, F. Wen, and K. Qiu, "Spectral features with the temporal and spatial mode-coupling dynamic in a few-mode system," PIERS, 2021.
29. X. Chen, J. He, and A. Li et al., "Characterization and analysis of few-mode fiber channel dynamics," IEEE Photon. Technol. Lett., vol. 25, no. 18, pp. 1819-1822, 2013.
30. J. Qin, and J. Ouyang. "A novel variable step size LMS adaptive filtering algorithm," Journal of Data Acquisition & Processing, vol.12, no.3, pp.171-194,1997.