Transfer Learning of Decision Feedback Neural Network Equalizers for Faster-than-Nyquist Signals Transmitted over MCF

*Jiawang Xiao, Lin Sun, Caoyang Liu, Meng Mao and Gordon Ning Liu**
School of Electronic and Information Engineering
Soochow University
Suzhou, China
*gordonnliu@suda.edu.cn.

***Abstract*—We investigate a transfer learning (TL) of decision feedback neural network equalizers (DFNNE) for Faster-than-Nyquist signals transmitted over multi-core fibers (MCF). In the simulation scenarios, the convergence rate of the TL-aid DFNNE with a source link of 17.5GHz is the fastest, achieving 72.2% faster compared to the traditional DFNNE.**

Keywords—space division multiplexing, neural networks equalizer, transfer learning

I. INTRODUCTION

The capacity of wavelength division multiplexing (WDM) systems based on standard single-mode fiber (SSMF) is already close to the nonlinear Shannon limit, making it difficult to increase the capacity. The utilization of space division multiplexing (SDM) techniques on few-mode fibers (FMF) and multi-core fibers (MCF) is widely regarded as a cost-effective solution. The utilization of multiple cores and spatial modes in a single fiber, as done in FMF or MCF, enhances the number of transmission channels and hence the transmission capacity. However, simultaneous transmission of multiple cores or modes can lead to signal interference, thereby compromising the transmission quality. Currently, the most effective method for addressing the signal impairment caused by inter-core crosstalk (IC-XT) or mode coupling is the utilization of multiple-input and multiple-output digital signal processing (MIMO-DSP) to model the coupling matrix [1,2]. In addition, the spatial mode dispersion (SMD) also brings great challenges to signal equalization.

The recent advancements in machine learning (ML) have led to the application of neural network equalizers (NNE) into SDM systems in fiber optic communication. For instance, accelerating the convergence of adaptive MIMO equalizers in SDM transmission through optimizers in the field of neural network [3]. And feedforward neural networks are used to mitigate the effect of IC-XT on 256 Gb/s short-reach systems employing weakly coupled MCF and Kramers–Kronig (KK) receivers [4]. However, NNE typically requires large amounts of prior known symbols and extensive training, making the equalizer parameter initialization very time-consuming. Additionally, if system parameters change, such as the optical launch power or fiber distance, the NNE may no longer be effective due to changes in the nonlinear noise distribution [5]. This necessitates retraining the network with new data sets, which is both inefficient and costly.

This work is financially supported by the National Natural Science Foundation of China (62105273, 62171175).

In this work, we also explore the application of the Faster-than-Nyquist (FTN) technique in SDM links to enhance the spectrum efficiency (SE) per spatial dimension. However, we observed that the interplay of the IC-XT and the inter-symbol interference (ISI) induced by the FTN shaping exhibits a kind of nonlinear distortion, which makes a general equalizer hard to equalize and decouple the individual SDM tributaries. To address this issue, we use a decision feedback neural network equalizer (DFNNE) to process coupled signals in parallel. The presence of the decision feedback module brings a new dimension of information to the equalizer, enabling the effective mitigation of nonlinear distortion [6]. Additionally, to reduce the long training time of the NNE, we employ transfer learning (TL) to accelerate its convergence. When processing 4 cores modulated by 50Gbaud DP-16QAM FTN signals on MCF transmission, our testing results of the proposed TL-aided DFNNE show that its convergence speed is up to 72.2% faster than the one of classical DFNNE.

II. PRINCIPLE

A. MCF transmission model based on coupled nonlinear Schrodinger equations

The nonlinearity, attenuation, and dispersion of MCF are similar to those of single-mode fibers, with the distinction being that MCF enable the transmission of multiple cores, leading to IC-XT. The IC-XT can be modeled through distributed Fourier methods as follows:

$$\frac{d\boldsymbol{E}_A}{dz} = -\frac{\alpha}{2}\boldsymbol{E}_A + \beta_1\frac{d\boldsymbol{E}_A}{dt} + \frac{\beta_2}{2}\frac{d^2\boldsymbol{E}_A}{dt^2} + \frac{\beta_3}{6}\frac{d^3\boldsymbol{E}_A}{dt^3} + j\frac{8}{9}\gamma\boldsymbol{E}_A^3 - j\frac{\exp(-j\Delta\beta_{AB})\,\boldsymbol{E}_B}{L_C}, \tag{1}$$

where $\boldsymbol{E}_A$ and $\boldsymbol{E}_B$ are the energies of cores A and B , α is the attenuation coefficient, β_1, β_2, β_3 represent the propagation constant, the group velocity dispersion and the dispersion slope respectively. γ represents nonlinearity coefficient and L_C is coupling length.

To ensure the validity of the 4-core MCF transmission model, we conducted a self-check as [7] on IC-XT and SMD. For IC-XT, the optical pulse at the fiber reception can be estimated by statistical methods. An optical pulse of power P is fed into core 1. The other cores have no pulses. Then the output power P_i of i^{th} core is measured, and the measured IC-XT is calculated according to $(\sum_{i=2}^{4} P_i)/P$. Figure 1 (a) shows the error curve between the set IC-XT and the measured value after 4000 measurements, which indicates that the

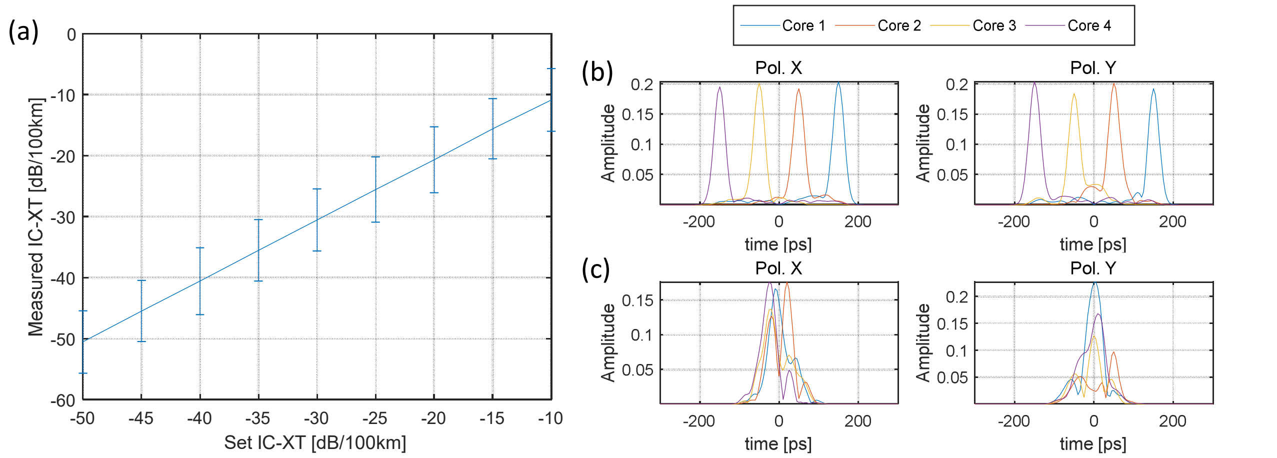

Fig. 1. Self-consistency examinations on the model of 100-km 4-core MCF links: (a) Error bar of IC-XT measurements. (b) Received pulses with coupling length at 1000km, (c) Received pulses with coupling length at 1km.

measurement values are well consistent with the set values. For SMD, the fiber length is set to 100 km, and the SMD is set to 1 ps/km. In order to better observe the delay between the optical pulses at the reception, the coupling length is set to be very long. A time-domain graph of the received optical pulses obtained under a coupling length of 1000 km is plotted in Fig. 1 (b), which shows that the delay between each pulse is 100 ps, verifying the correctness of the model. The impact of strong coupling and SMD on the optical pulse is shown in Fig. 1 (c), under a coupling length of 1 km. The strong coupling effectively reduces the delay between pulses by suppressing spatial mode dispersion [8].

B. Transfer learning-aided DFNNE

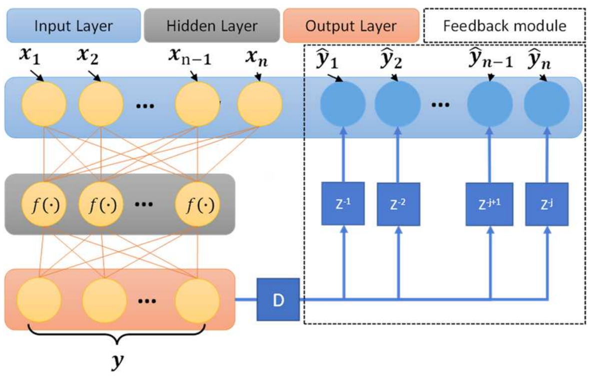

Fig. 2. Structure of decision feedback neural network equalizer.

To mitigate nonlinear distortion resulting from the interaction of ISI and IC-XT, DFNNE with inputs containing the 8 polarizations and the decision symbols is employed to recover all polarization tributaries. DFNNE with only one hidden layer is utilized in this work for the complexity consideration of the SDM system, whose structure is given in Fig. 2. In the figure, symbol D is the decision and Z is the delay module. Connection weights are trained through 8 tributaries of prior-known symbols. The IC-XT is expected to be mitigated by training the weights of DFNNE through the prior-known sequences. At the same time, the ISI due to narrowly filtering can be alleviated by using the decision feedback structure.

Assuming there are n tributaries of data in the SDM system that need to be received in parallel, let $\boldsymbol{x}_i$ be a vector composed of the i^{th} tributary data input into the equalizer, and $\hat{\boldsymbol{y}}$ be a vector composed of the i^{th} polarization feedback decision signal, then the vector composed of the n outputs of the DFNNE is represented as

$$\boldsymbol{y} = \boldsymbol{W}_2 f(\boldsymbol{W}_1[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n, \hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \ldots \hat{\boldsymbol{y}}_n] + \boldsymbol{b}_1) + \boldsymbol{b}_2, \quad (2)$$

where $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ and $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ are the weighted and bias matrixes respectively, and the activation function is denoted by $f(\cdot)$.

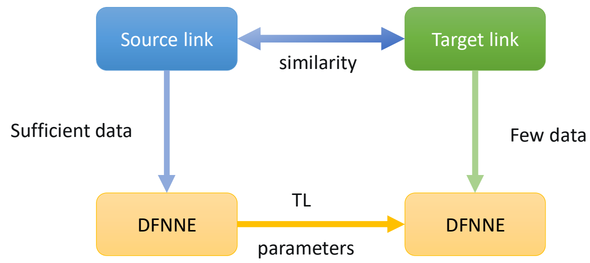

Fig. 3. Schematic diagram of transfer learning assisted DFNNE.

Transfer learning approach leverages pre-learning knowledge from a source link to facilitate the training of the NNE on a target link. Due to the similarity between the two links, pre-learning knowledge learned from the source link can be utilized in the target link. Thereby, TL reducing the time and amount of data required for training. The specific migration process is shown in Fig. 3. In particular, to use transfer learning effectively, there must be some similarity between the source and target links.

III. RESULTS

For evaluating the proposed method, we conducted a simulation using MATLAB and Pytorch. Figure 4 depicts the setup of 4-core 100-km transmission and reception of FTN signals. Baud rate is set at 50Gbaud, thus the total capacity is 1.6Tbps over 4 cores of MCF by using DP-16QAM formats. Coefficients of fiber attenuation, dispersion, nonlinearity and spatial mode dispersion are 0.16dB/km, 16ps/nm/km, 0.81W/m and 27ps/km respectively. OSNR contributed by ASE noise is set at 20dB when the launching power is 0dBm. Coupling length of MCF is 100m, indicating a strongly coupling regime.

The launch power is set to the optimal value of 7dBm for this scenario. For the equalizer parameters, we set the feedforward memory length of each core in the MCF to 15 and the feedback delay length to 11. To accommodate the 16QAM

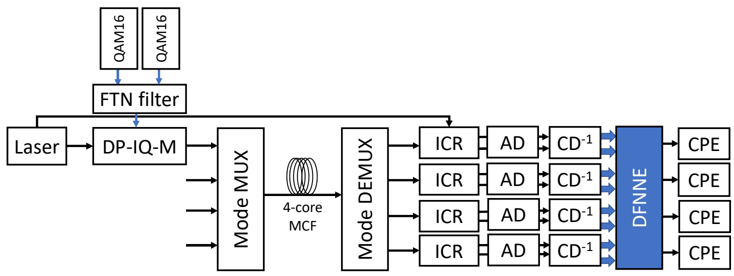

Fig. 4. Setup of 4-core MCF Faster-than-Nyquist transmission and reception.

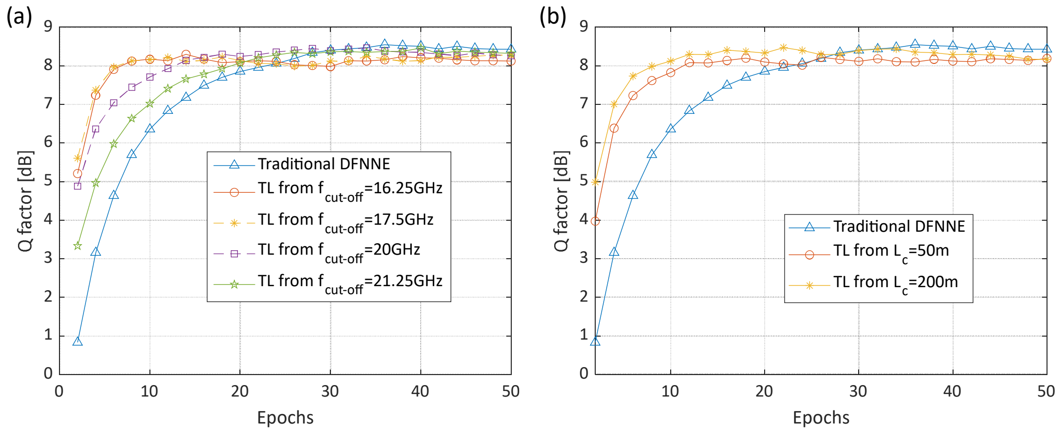

Fig. 5. Q factor vs. epochs for the target link with cut-off frequency at 18.75GHz and coupling length at 100m migrated from the source links (a) with cut-off frequency at 16.25GHz, 17.5GHz, 20GHz and 21.25GHz; (b) with coupling length at 50m and 200m.

transmission, the real and imaginary parts of the signal were separated and input into the equalizer, resulting in a DFNNE input layer of 416 neurons (calculated as (15+11)*2*4*2). The number of neurons in the hidden layer is 64, and the number of neurons in the output layer is 16, twice of the core numbers. The Tanh function is used as the activation function and Stochastic Gradient Descent (SGD) is selected as the optimization algorithm. This combination of parameters was determined to be optimal. During the training process, the DFNNE parameters are iterated with sequences of a prior known length of 20K and 50 epochs to make them converge completely.

The changes in the optical link are simulated by adjusting the cut-off frequency $f_{\mathrm{cut-off}}$ of the FTN filters and the inter-mode coupling length L_C. The performance of the DFNNE is found to be satisfactory when the FTN filter cut-off frequency is set to 18.75GHz and the coupling length is 100m. Therefore, this setting serves as the parameter configuration for the target link. Considering the requirement of correlation between the two links, we selected source links by changing either the cut-off frequency (16.25GHz, 17.5GHz, 20GHz and 21.25GHz) or coupling length (50m and 200m) separately. Figure 5(a) demonstrates that compared one with source link of 21.25GHz cut-off frequency, DFNNE with source links of 20GHz converges faster. Similarly, the DFNNE with the source link of 17.5Ghz converges faster than the one with source link of 16.25GHz. This indicates that the correlation between the two source links has a significant impact on the performance of the target link's DFNNE. Additionally, the convergence rate of the DFNNE is faster when migrating from a lower cut-off frequency to a higher one. This is because removing a large amount of information from the DFNNE is easier than starting from scratch and adding new knowledge. In addition, the DFNNE has a faster convergence rate when migrating from a larger to a smaller coupling length. This is due to the influence of fiber nonlinearity and SMD, as longer coupling lengths result in poorer DFNNE performance. Migration between links with different coupling lengths as shown in Fig. 5(b). Therefore, when selecting the source link, it is important to choose one that has strong correlation with the target link.

IV. Conclusion

We propose TL-aided DFNNE for parallel processing of Faster-than-Nyquist signals transmitted over MCF at different cut-off frequencies and coupling lengths. In this scenario, the convergence rate of the TL-aid DFNNE with a source link of 17.5GHz is the fastest, achieving 72.2% faster compared to the traditional DFNNE.

References

[1] G. Rademacher, B. J. Puttnam, R. S. Lus, J. Sakaguchi, W. Klaus, T. Eriksson, and et al, “10.66 peta-bit/s transmission over a 38-core-three-mode fiber,” Optical Fiber Communication Conference (2020), paper Th3H–1.

[2] X. Pan, B. Liu, L. Li, T. Qinghua. “Low complexity MIMO method based on matrix transformation for few-mode multi-core optical transmission system,” Optics Communications, 371, 238-242 (2016).

[3] M. Arikawa and H. Noguchi, "Fast Convergence by Machine Learning Optimizer for Adaptive MIMO Equalizer Used in SDM Transmission over Coupled-Core 4-Core Fiber and 4-Core EDFA," in European Conference on Optical Communications (2020), pp. 1-4.

[4] D. Piedade, T. Alves, T. Brandão, “Short-Reach MCF-Based Systems Employing KK Receivers and Feedforward Neural Networks for ICXT Mitigation,”Photonics, 9(5), 286 (2022).

[5] J. Zhang, L. Xia, M. Zhu, S. Hu, B. Xu, and K. Qiu, “Fast remodeling for nonlinear distortion mitigation based on transfer learning,” Optics Letters, 44(17), 4243–4246 (2019).

[6] C. Ye, D. Zhang, X. Hu, X. Huang, H. Feng and K. Zhang, "Recurrent Neural Network (RNN) Based End-to-End Nonlinear Management for Symmetrical 50Gbps NRZ PON with 29dB+ Loss Budget," in European Conference on Optical Communication (2018), pp. 1-3.

[7] L. Sun, J. Zhang, G. Zhou, B. Chen, G. N. Liu, Y. Cai, and et al, "Theoretical investigations of weakly- and strongly-coupled multi-core fibers for the applications of optical submarine communications under power and fiber count limits," Optics Express, 31, 4615-4629 (2022).

[8] T. Hayashi, Y. Tamura, T. Hasegawa and T. Taru, "Record-Low Spatial Mode Dispersion and Ultra-Low Loss Coupled Multi-Core Fiber for Ultra-Long-Haul Transmission," Journal of Lightwave Technology, 35(3), 450-457 (2017).