Experimental Demonstration of Neural Network-based Soft Demapper for Long-haul Optical Transmission

Wenkai Fang[1], Bin Chen[1,*], Yi Lei[1], Can Zhao[1], Menno van den Hout[2], Sjoerd van der Heide[2], Chigo Okonkwo[2], Lin Sun[3] and Xuwei Xue[4], and Shanguo Huang[4]
[1]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China
[2]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[3]School of Electronic and Information Engineering, Soochow University, Suzhou, China
[4]IPOC, Beijing University of Posts and Telecommunications, Beijing, China
*Email: bin.chen@hfut.edu.cn

***Abstract*—Experimental validation for a proposed neural network-based soft demapper is demonstrated. A reach increase of 9.8% and a demapping complexity reduction of 28.6% for PM-64QAM with 11×450Gbps DWDM is achieved over the conventional demapper.**

***Index Terms*—soft demapper, neural network, optical fiber communication**

I. Introduction

High-order quadrature modulation (QAM) and forward error correction (FEC) are significant technologies for achieving high spectral efficiencies (SE) in modern communication systems [1]. While early optical communication systems relied on hard-decision (HD)-FEC, most current coherent optical communication systems use soft-decision (SD)-FEC. In these systems, bit-interleaved coded modulation (BICM) is typically employed, with bit-wise (BW) decoders that separate signal recovery and FEC. A crucial component of an SD-FEC BICM system is the soft demapper, which computes the log-likelihood ratios (LLRs) or soft information [2]. The quality of the LLRs directly affects the overall system performance [3], as they are the only inputs to the FEC decoder.

Accurate LLRs require knowledge of the conditional probability density function (PDF) of the channel output given the input, known as the log-maximum a-posteriori (MAP) algorithm. The log-MAP algorithm computes the logarithm of the ratio between the MAP probabilities of the two hypotheses for the latent bit given the observed symbol [4]. Classical approaches for computing LLRs assume channel impairments as independent identical Gaussian noise. However, in a practical optical communication system, the residual impairments from various hardware components and the channel cannot be completely compensated for. Consequently, computing LLRs under the AWGN assumption incurs a performance penalty [5], and the computational complexity of this method also scales with the size of the symbol constellation, making direct implementation impractical in realistic systems [3].

Machine learning (ML) methods have recently shown significant performance gains in various applications, including optical communication systems. Researchers have adopted ML techniques to solve a wide range of problems, including accurate LLR computation [3], [4], [6]–[9]. By learning the probability density function (PDF) of the real channel without any prior knowledge, well-trained neural networks (NNs) can accurately evaluate LLRs. In [4], an NN is trained to compute LLRs by imitating analytical computations for AWGN channels, thereby reducing complexity. In [8], [9], authors applied machine learning for L-value computation and demonstrated that deep neural networks (DNNs) can predict probabilities and learn soft demodulation schemes effectively. In [3] and [7], researchers proposed novel NN-based soft demappers that can model transition probabilities of memory channels and compensate for nonlinear effects in optical channels. However, as the complexity of the channel increases, the network can become more redundant, leading to higher computational complexity.

In this paper, we propose a novel approach for soft demapping by using NN that learns soft information. The proposed NN-based soft demapper is trained with experimental data via an optical transmission system. To balance the performance and complexity, the required number of hidden layer neurons and the amount of training data is analyzed. The NN-based demapper enabled us to experimentally demonstrate a 11 wavelength division multiplexing transmission upto 11×450Gbps and outperforms conventional demappers.

II. System Model and Proposed NN-based Soft Demapper

A. System Model

The system model we considered in this paper is shown on Fig. 1 (a), which includes the blocks of experimental

The work is supported by the National Natural Science Foundation of China (NSFC) under Grant (62171175, 62001151), Fundamental Research Funds for the Central Universities (No. JZ2022HGTB0262)) and open fund of IPOC (Beijing University of Posts and Telecommunications).

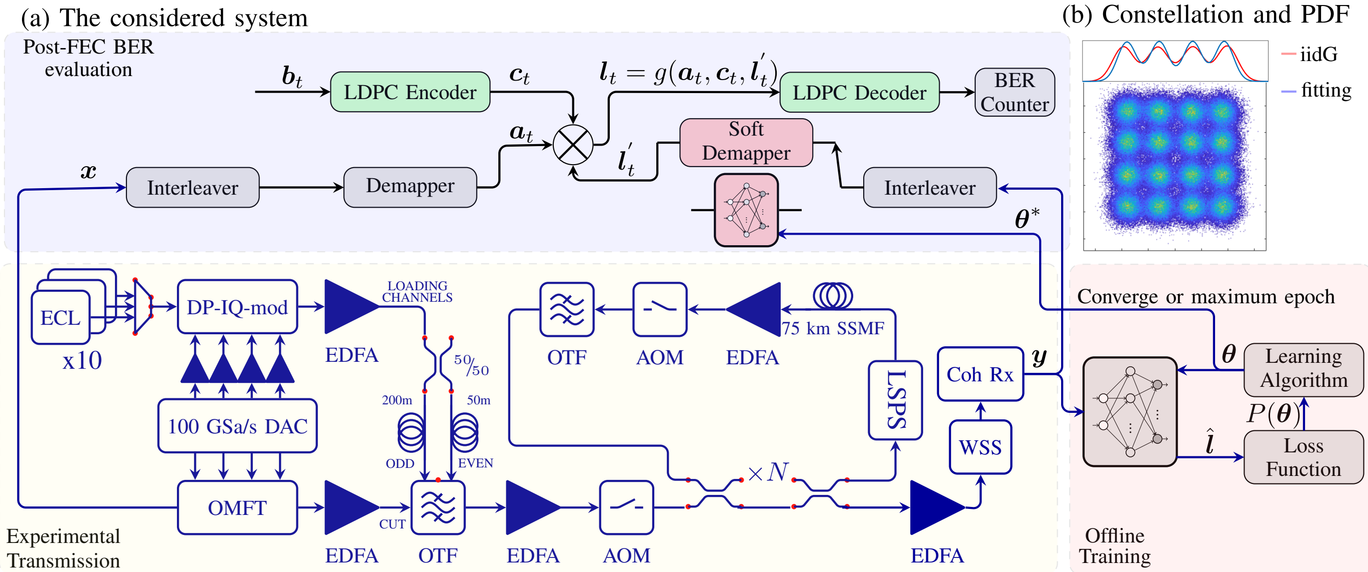

Fig. 1. (a) The considered system model in this paper, including experimental transmission block, the block of offline training for NN-based demapper and the block of post-LDPC BER performance evaluation; (b) Constellation diagram for 16QAM at 3000 km and PDF under the assumption of iidG and idd distribution.

transmission, offline training for NN-based demapper and post-FEC BER performance evaluation.

At the block of post-LDPC BER evaluation, a low-density parity-check code (LDPC) encoder and decoder is employed to encode information bits into codeword and decode the received symbols with the given LLRs, respectively. The channel noise is emulated with the experimental data of a long-haul optical transmission, which is explained in detail in Sec. III-A.

Using interleaver and demapper, the transmitted symbols $\boldsymbol{x}$ of optical-multi-format transmitter (OMFT) are demapped into bits $\boldsymbol{a}_t$. And the received symbols $\boldsymbol{y}$ after coherent receiver (Coh Rx) are soft demapped into $\boldsymbol{l}_t^{'}$. The $\otimes$ denotes a descrambler operation [10], which can be depicted by a function $\boldsymbol{l}_t = g(\boldsymbol{a}_t, \boldsymbol{c}_t, \boldsymbol{l}_t^{'}) = (-1)^{\boldsymbol{a}_t \boldsymbol{c}_t} \boldsymbol{l}_t^{'}$, where $\boldsymbol{c}_t$ is the codewords as the output of a LDPC encoder. A BER counter is applied to compute the post-LDPC BER. As shown in the block of post-FEC BER evaluation, a NN-based demapper can be employed to replace the conventional soft demapper, whose network parameter is trained by gradient descent algorithm and update after every training in the block of offline training.

B. Conventional Soft-Demapper

To formalize different methods for computing LLRs, we define the n-dimensional vector of information bits as $\boldsymbol{b}_t = [b_{1,t}, b_{2,t}, ..., b_{n,t}], t = 1, 2, ..., L$ (L is the block length), which is encoded to an m-dimensional bits sequence $\boldsymbol{c}_t = [c_{1,t}, c_{2,t}, ..., c_{m,t}]$ and the code rate is n/m. At the soft demapper, two conventional demappper are considered for computing LLRs. One is supposing the channel is an ideal AWGN channel and assuming each constellation point has the same noise distribution as independent and identically distributed Gaussian (iidG) [11]. In this case, the LLR of the i^{th} bit in $\boldsymbol{c}_t$ can be defined as

$$l_{i,t} = \log \frac{\sum_{x \in \chi_i^0} f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{\sum_{x \in \chi_i^1} f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}, \qquad i = 1, 2, ..., m \tag{1}$$

where X, Y denotes the vectors of transmitted symbols and received symbols, respectively. And χ_i^c denotes the set of indices of constellation points labeled by a bit 0 or 1 at bit position i.

As shown in Fig. 1 (b), a PDF difference exists under the iidG assumption and the fitting distribution fitted by scatter density for measured constellations for 16QAM at 3000 km. Thus, a more appropriate distribution is needed, and an improved way is to estimate the noise distribution for each constellation point as independent and differential distributed Gaussian (iddG) distribution. Under the assumption of iddG disbution, the LLR of the i^{th} bit in $\boldsymbol{c}_t$ can be expressed as

$$l_{i,t}^{*} = \log \frac{\sum_{x \in \chi_i^0} \frac{1}{\sigma_i} \exp(\frac{||y - x_i||^2}{\sigma_i^2})}{\sum_{x \in \chi_i^1} \frac{1}{\sigma_i} \exp(\frac{||y - x_i||^2}{\sigma_i^2})}, \qquad i = 1, 2, ..., m \tag{2}$$

where x_i and σ_i is the mean and variance of i^{th} constellation point, respectively.

C. Proposed NN-based Soft Demapper

From the discussion of Sec. II-B, we can see that the demappper performance of (2) should be obliviously superior to (1) for taking into the consideration of mean and variance values approximation per constellation point, but it will also lead to a higher complexity with more computation operations. Here, a NN-based demapper is proposed to trade-off performance and complexity, which is with two input neurons and one hidden layer. The input is the real part and imaginary part of the received symbols. The activation function is the tansig $\tau(x) = \frac{2}{1+e^{-2x}} - 1$, and the LLRs can be estimated as

$$\hat{l}_{i,t} = (\tau(\boldsymbol{y}^T\boldsymbol{\omega}_1 + \boldsymbol{b}_1))^T\boldsymbol{\omega}_2 + \boldsymbol{b}_2, \qquad i = 1, 2, ..., m \quad (3)$$

where $\boldsymbol{\omega}_1$, $\boldsymbol{b}_1$ and $\boldsymbol{\omega}_2$, $\boldsymbol{b}_2$ denote the weights, bias of input layer and hidden layer, respectively. $(\cdot)^T$ denotes the transpose operation.

In the training process, $\boldsymbol{\theta}$ is defined as the set of trainable parameters ω_1, b_1 and ω_2, b_2. The trainable parameters $\boldsymbol{\theta}$ of NN can be optimized base on a learning algorithm using gradient descent with the mean-squared error (MSE) loss function $P(\boldsymbol{\theta}) = \sum_{t=1}^{L} ||\boldsymbol{l}_t^* - \hat{\boldsymbol{l}}_t||^2$, where $\boldsymbol{l}_t^*$ and $\hat{\boldsymbol{l}}_t$ is the vector of LLRs calculated by (2) and (3), respectively. When the MSE is converged or the maximum training epoch reaches, the training is complete and the NN-based soft demapper with trained parameters $\boldsymbol{\theta}^*$ can be applied to evaluate the SD-FEC decoding performance.

III. Experimental Setup and Results

A. Experimental Setup

The experimental transmission setup employed is a optical fiber communication system with a recirculating loop testbed. In Fig. 1, the symbols are firstly filtered by a root-raised-cosine with 1% roll-off at 41.79 GBd, pre compensated for transmitter impairments, and uploaded to a 100 GSa/s digital-to-analog converter (DAC). An optical-multi-format transmitter modulates the channel under test (CUT), which chosen among the 11 tested C-band channels. The OMFT is comprised of an external cavity laser (ECL), a dual-polarization IQ-modulator (DP-IQM), an automatic bias controller and RF-amplifiers.

To generate 11 loading channels, the multiplexed outputs of ECLS are modulated. The multiplexed outputs of ECLs are modulated using a DP-IQM, amplified, split into even and odd, decorrelated by 10,200 symbols (50 m) and 40,800 symbols (200 m) with respect to the CUT, respectively. The CUT, odd, and even channels are combined onto the 50 GHz spaced DWDM grid. The signal is launched with total power of 9.5 dBm and enters the recirculating loop which consists of a loop-synchronous polarization scrambler (LSPS), a 75 km span of standard single mode fiber (SSMF), an Erbium-doped Optical Fiber Amplifier (EDFA), an acousto-optical modulator (AOM) and an optical tunable filter (OTF) used for gain equalization. After transmission, the signal is amplified, the CUT selected using a wavelength selective switch (WSS), detected using an intradyne coherent receiver. Receiver DSP consisting of front-end compensation, chromatic dispersion (CD) compensation, frequency-offset compensation, and decision-directed equalization with in-loop phase search is performed offline. More details of experimental setup can be also found in [12].

B. Results

To ascertain the structure of NN and the volume of training data, the influence of the number of hidden neurons and volume of training data to MSE is first analyzed. Taking for example of the analysis for 64QAM at 525 km, which is shown in Fig. 2. We found that the MSE value around 0.1 as the dashed line marks is acceptable for the training. Therefore, the highlighted black circle, which is 32 hidden neurons and 20% volume of training data, is chosen to trade-off performance and computational complexity. In other words, more hidden neurons can be chosen to achieve better performance but with more complex computation. By using the same analysis method, 20% experimental data and 8 hidden neurons are chosen for training 16QAM.

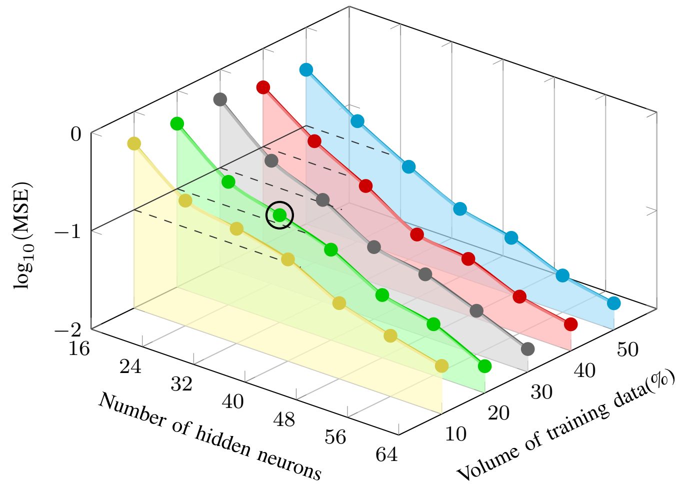

Fig. 2. MSE vs the numbers of hidden layer neurons and different amount of training data for 64QAM at 525 km. The dashed line denotes the value of MSE is 0.1 and the black circle highlights the chosen parameters with 32 hidden neurons and 20% training data for 64QAM.

By using the considered system model and the experimental data in Fig. 1, post-FEC BER versus transmission distances of 16QAM and 64QAM for three demappers under the assumption of iidG, iddG distribution and NN-based demapper is shown in Fig. 3, respectively.

BER performance after FEC for different soft demapper under the assumption of iddG (circles) and iidG (square) distribution are shown in Fig. 3. For the post-FEC BER evaluation, LDPC blocks are constructed by using the DVB-S2 LDPC code with different code rate $R \in \{4/5, 5/6, 8/9, 9/10\}$ and code length $n = 64800$. It can be observed that demapped by under iddG assumption can transmit longer distances, which comes from the fact that iddG distribution is more accurate with the actual channel distribution. Demapping under iddG assumption or NN-based demapper has a gain of 50 km (9.8%) and 55 km (6.4%) compared with demapping by iidG for 64QAM with code rate at 0.89 and 0.83, respectively. For 16QAM, the gain is 85 km (5.0%), 135 km (6.1%) and 255 km (9.6%) with code rate of 0.9, 0.83 and 0.8, respectively.

C. Complexity Analysis

The computational complexity of above three methods of soft demapping is depicted in Table. I, which shows the number of computation for addition and subtraction (Add&Sub), multiplication and division (Mul&Div), exponent and logarithm (Exp&Log) operations, respectively. For demappers

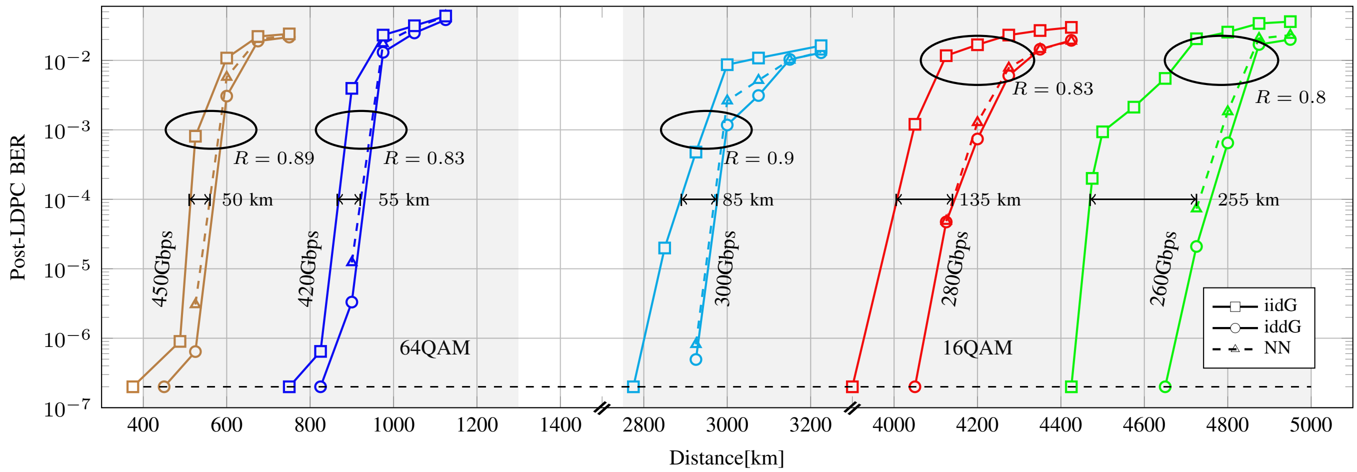

Fig. 3. Post-FEC BER versus transmission distance for three different soft demappers, including assuming each constellation point has the same noise distribution as iidG or different noise distribution as iddG and the NN-based demapper.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON OF THREE DEMAPPERS IN TERMS OF OPERATIONS PER RECEIVED SYMBOL

	Demapper	Add&Sub	Mul&Div	Exp&Log	Total
16QAM	iidG	104	52	20	176
	iddG	104	68	20	192
	NN	64	56	16	136
64QAM	iidG	564	198	70	832
	iddG	564	262	70	896
	NN	320	256	64	640

under iidG and iddG assumption, the latter has a higher complexity no matter for 16QAM or 64QAM of Mul&Div operations, but its performance is better as well in Fig. 3. When comparing to demapper under iidG assumption, the NN-based demapper has better post-LDPC-BER performance and less complex computation. Comparing to demapper under iddG assumption, NN-based demapper has similar performance but less complex computation. For the sum of three demapper methods, demapped by a NN-based demapper shows a complexity reduction of 29.2% and 28.6% compared with demapping under iddG assumption for 16QAM and 64QAM, respectively. Thus the NN-based demapper can be proven to be an attractive method to replace conventional demapper.

IV. CONCLUSION

In this paper, we presented a NN-based soft demapper for estimating LLRs, which trade-off the post-FEC BER and complexity. The analysis was performed by using the experimental optical fiber transmission data. One of the advantages of the analyzed NN-based soft demapper is its low implementation complexity as the less computational operations are required, compared to the conventional demappers. In this paper, we showed that by combining the demapper with LDPC, reach increases for a wide range of distances (from 400 km to 5000 km) with multiple net rates between 255 Gbps and 450 Gbps were reported. Even though the demapper we proposed is still under the assumption of Gaussian distribution, it can be extended to a non-Gaussian distribution approximation for exploring larger gains, which is left for future research.

REFERENCES

[1] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable information rates for fiber optics: Applications and computations," *J. Lightw. Technol.*, vol. 36, no. 2, pp. 424–439, Jan. 2018.

[2] M. Ivanov, C. Häger, F. Brännström, A. Graell i Amat, A. Alvarado, and E. Agrell, "On the information loss of the max-log approximation in bicm systems," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3011–3025, 2016.

[3] M. Schaedler, S. Calabrò, F. Pittalà, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Neural network-based soft-demapping for nonlinear channels," in *Opt. Fiber Commun. Conf.*, 2020, pp. 1–3.

[4] O. Shental and J. Hoydis, "Machine LLRning: Learning to softly demodulate," in *2019 IEEE Globecom Workshops*, 2019, pp. 1–7.

[5] A. Alvarado, L. Szczecinski, T. Fehenberger, M. Paskov, and P. Bayvel, "Improved soft-decision forward error correction via post-processing of mismatched log-likelihood ratios," in *Eur. Conf. Opt. Commun.*, 2016, pp. 1–3.

[6] M. Schädler and R. Olivier, "Machine learning in digital signal processing for optical transmission systems," Ph.D. dissertation, 2022.

[7] M. Schädler, G. Böcherer, F. Pittalà, S. Calabrò, N. Stojanovic, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Recurrent neural network soft-demapping for nonlinear isi in 800Gbit/s DWDM coherent optical transmissions," *J. Lightw. Technol.*, vol. 39, no. 16, pp. 5278–5286, 2021.

[8] T. Koike-Akino, D. S. Millar, K. Parsons, and K. Kojima, "Fiber nonlinearity equalization with multi-label deep learning scalable to high-order DP-QAM," in *Advanced Photonics*, 2018, p. SpM4G.1.

[9] M. Schädler, G. Böcherer, and S. Pachnicke, "Soft-demapping for short reach optical communication: A comparison of deep neural networks and volterra series," *J. Lightw. Technol.*, vol. 39, no. 10, pp. 3095–3105, 2021.

[10] N. Stojanovic, Y. Zhao, D. Chang, Z. Xiao, and F. Yu, "Reusing common uncoded experimental data in performance estimation of different fec codes," *IEEE Photon. Technol. Lett.*, vol. 25, no. 24, pp. 2494–2497, 2013.

[11] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2256–2266, 2016.

[12] S. van der Heide, B. Chen, M. van den Hout, G. Liga, T. Koonen, H. Hafermann, A. Alvarado, and C. Okonkwo, "11,700 km transmission at 4.8 bit/4D-sym via four-dimensional geometrically-shaped polarization-ring-switching modulation," in *OptoElectronics and Communications Conference (OECC) and International Conference on Photonics in Switching and Computing (PSC)*, 2019, pp. 1–3.