Supplemental Document

Optics Letters

Enhanced mutual information neural estimators for optical fiber communication: supplement

ZEKUN NIU,[1] CHENHAO DAI,[1] HANG YANG,[1] CHUYAN ZENG,[1] ZHIXUE HE,[2] WEISHENG HU,[1] AND LILIN YI[1,*]

[1] *State Key Lab of Advanced Optical Communication Systems and Networks, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[2] *Peng Cheng Laboratory, Shenzhen 518055, China*
[*] *lilinyi@sjtu.edu.cn*

Supplement DOI: https://doi.org/10.6084/m9.figshare.26325142

Parent Article DOI: https://doi.org/10.1364/OL.534025

THE ENHANCED MUTUAL INFORMATION NEURAL ESTIMATORS FOR OPTICAL FIBER COMMUNICATION: SUPPLEMENTAL DOCUMENT

1. Hyperparameters for MINE, E-MINE, and NFEM.

Below, we provide the information for MINE, E-MINE, and NFEM as shown in Table 1. MINE and E-MINE have the same structure, each consisting of a fully connected NN (FCNN) comprising four layers: an input layer, an output layer, and two hidden layers. Each hidden layer is composed of 128 neurons. The input layer consists of 4 neurons for 2D symbols and 8 neurons for 4D symbols. ELU serves as the activation function for the hidden layers, while the outputs are activated using the linear function to calculate the loss. The Adam optimizer is utilized with a cosine decay learning rate that begins at 0.01.

Table S1

	MINE	E-MINE	NFEM
Input size	4 for 2D 8 for 4D	4 for 2D 8 for 4D	42 for 2D 84 for 4D
Hidden layers	2	2	3
Hidden layer neurons	128	128	128
Output size	1	1	2 for 2D 4 for 4D
Activation function	ELU	ELU	ELU
Learning rate	0.01	0.01	0.01

Another priority is the choice of batch size N and M. Although the true MI does not depend on the batch size in theory, the batch size affects the accuracy and stability of the MI estimation, which is reflected in Eq. (4). The large batch size can help to reduce the estimation bias in gradients and stabilize the results. In addition, the work shown in [1] proves that the MI estimation cannot be larger than O (log N), where N is the batch size. To achieve an accurate measure while mindful of computational limitations, it is essential to utilize as large a batch size as feasibly possible. In terms of communications, MI is typically less than 10 bits/symbol, corresponding to 1024 quadrature amplitude modulation (QAM). We recommend setting the batch size M to 1000 and N to 10000, as log2(1000) = 9.96, which is suitable for most cases.

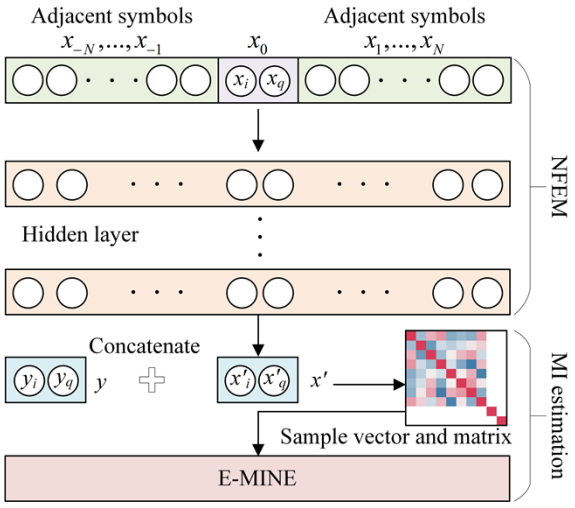

Fig. S1. The structure of NFEM.

The NFEM is a FCNN, depicted in Fig. S1, where the inputs consider the adjacent symbols $x_{-N}, \ldots, x_{-1}$ and $x_1, \ldots, x_N$ around the processed symbol x_0. The parameters with 21 symbol inputs are shown in Table S1. The outputs are limited to the same size as the processed symbol. Then, along with the Tx symbols, we can construct the sample vector and matrix for MI estimation. The learning rate and optimizer are the same as with E-MINE.

2. Details of digital signal processing in the experiments.

Figure S2 illustrates the digital signal processing (DSP) configuration used in our experiments [2]. Initially, the signals undergo IQ equalization, which includes IQ balancing using the Gram-Schmidt orthogonalization procedure and IQ skew compensation in the frequency domain. Subsequently, frequency domain compensation is applied for chromatic dispersion (CD) effects. Next, a 2x2 time-domain multi-input-multi-output (MIMO) adaptive filter equalizes the polarization mode dispersion (PMD) effects and residual CD effects. The MIMO filter tap is set at 81, and the least mean square (LMS) method updates the MIMO filter coefficients. To compensate for the frequency offset between the transmitter and receiver lasers, we employ a 4th power operation for coarse compensation of the 16 QAM signal. Then, a blind phase search algorithm compensates for laser phase noise and residual frequency offset, with the test phase set at 8.

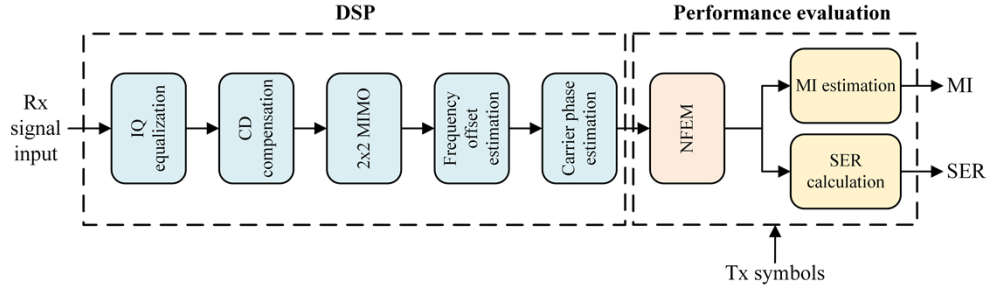

Fig. S2. The digital signal processing in the experiments.

The DSP compensates for most linear effects, while residual effects such as nonlinear distortions are addressed by the NFEM. Then the output symbols are used for performance evaluation including MI estimation and symbol error rate (SER) calculation.

3. Details of Monte-Carlo estimator based on Gaussian assumption.

In the following, we show the principles and equations for Monte-Carlo (MC) estimator. MI can be estimated via MC integration, which can be used to approximate an integral using a finite sum, as:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{R}}[f(\boldsymbol{r})] &= \int_{\Omega} P_{\boldsymbol{R}}(\boldsymbol{r}) f(\boldsymbol{r}) d\boldsymbol{r} \\ &= \frac{1}{N}\sum_{k=1}^{N} f(\boldsymbol{r}^k)\end{aligned} \tag{S1}$$

where f: $\Omega \to \mathbb{R}$ is a real-valued function and Ω is the domain of s-dimensional integration. The vectors $\boldsymbol{r} = (\boldsymbol{r}^1, \ldots, \boldsymbol{r}^N)$ are random samples from the probability distribution $P_{\boldsymbol{R}}(\boldsymbol{r})$.

Typically, let $\boldsymbol{X} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^k)$ be the transmitted symbol and its probability distribution is $P_{\boldsymbol{X}}(\boldsymbol{x})$, $\boldsymbol{Y} = (\boldsymbol{y}^1, \ldots, \mathrm{y}^k)$ is the corresponding channel outputs. Then the MI of the input-output pairs $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ using MC integration can be expressed by:

$$I(\boldsymbol{X};\boldsymbol{Y}) = \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})})] = \frac{1}{N}\sum_{k=1}^{N}\log(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}^k\,|\,\boldsymbol{x}^k)}{P_{\boldsymbol{Y}}(\boldsymbol{y}^k)}) \tag{S2}$$

Therefore, to accurately calculate MI between channel inputs and outputs in a communication system, it is essential to have an analytical model that characterizes the channel's behavior through the conditional probability distribution function (PDF) $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$, which we refer to as the 'channel law'. As for the AWGN channel, it is commonly understood that the channel law follows a Gaussian distribution. However, in optical fiber communications, the unknown channel law of the optical fiber channel limits the computation of MI and results in an unavailable channel capacity. The most used methods for MI estimation are based on the auxiliary channel, which assumes a known conditional PDF for the channel law, denoting by $Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. Thus, the auxiliary channel's output distribution can be written as:

$$Q_p(\boldsymbol{y}) = \sum_i P_{\boldsymbol{X}}(\boldsymbol{x}_i) Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i) \tag{S3}$$

Under the assumption of selecting an appropriate auxiliary channel $Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ such that $Q_p(\boldsymbol{y}) > 0$ whenever $P(y) > 0$, we have:

$$\begin{aligned} & I(\boldsymbol{X};\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log(\frac{Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{Q_p(\boldsymbol{y})})] \\ &= \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log(\frac{P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x})P_{\boldsymbol{Y}}(\boldsymbol{y})}) - \log(\frac{Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{Q_p(\boldsymbol{y})})] \\ &= \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log(\frac{P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x})P_{\boldsymbol{Y}}(\boldsymbol{y})Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})/Q_p(\boldsymbol{y})})] \\ &= D(P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) \,\|\, P_{\boldsymbol{Y}}(\boldsymbol{y})R_p(\boldsymbol{x}|\boldsymbol{y})) \\ &\geq 0 \end{aligned} \tag{S4}$$

where $R_p(\boldsymbol{x}|\boldsymbol{y})$ is defined as:

$$R_p(\boldsymbol{x}|\boldsymbol{y}) \triangleq \frac{P_{\boldsymbol{X}}(\boldsymbol{x})Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{Q_p(\boldsymbol{y})} = \frac{P_{\boldsymbol{X}}(\boldsymbol{x})Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{\sum_i P_{\boldsymbol{X}}(\boldsymbol{x}_i)Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i)} \tag{S5}$$

Since, we can get the estimation of MI as [3]:

$$I(\boldsymbol{X};\boldsymbol{Y}) \geq \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log\frac{Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{\sum_i Q_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i)P_{\boldsymbol{X}}(\boldsymbol{x}_i)}] \tag{S6}$$

In such cases, we can obtain a mismatched MI, and the accuracy of the estimation is dependent on the degree of approximation to the optical fiber channel. Numerous studies have demonstrated that the Gaussian auxiliary channel provides a good approximation for long-haul fiber transmission. Hence, a straightforward approach to estimate MI is to consider the channel model as a circular-symmetric Gaussian channel for all transmitted symbols. Nonetheless, assuming the same circular-symmetric Gaussian channel for all transmitted symbols ignores the different nonlinear noises on each modulation symbol, resulting in inaccuracies. To obtain a more precise estimation, one can use different circular-symmetric Gaussian probability density functions for transmitted symbols, in which every candidate symbol $\boldsymbol{x}_i$ in the constellation set has a different noise variance σ_i, i.e.,

$$Q_{\boldsymbol{Y}|\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{x}_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp(-\frac{\|\boldsymbol{y}-\boldsymbol{x}_i\|^2}{2\sigma_i^2}) \tag{S7}$$

By substituting Eq.(S7) into Eq.(S6), we can obtain an estimation of MI for the optical fiber channel based on the Gaussian assumption with MC method.

4. Numerical results of E-MINE.

In the additive white Gaussian noise (AWGN) channel, the progression of MINE and E-MINE during training is showcased in Fig. S3. The baseline is based on Gauss-Hermite estimation. The test signals are uniform distribution 256 QAM. Fig. S3 depictS how MI estimations ascend with training epochs until nearing the actual values, highlighting E-MINE's enhanced stability over MINE. These results highlight the potential of E-MINE for accurate MI estimation in the AWGN channel.

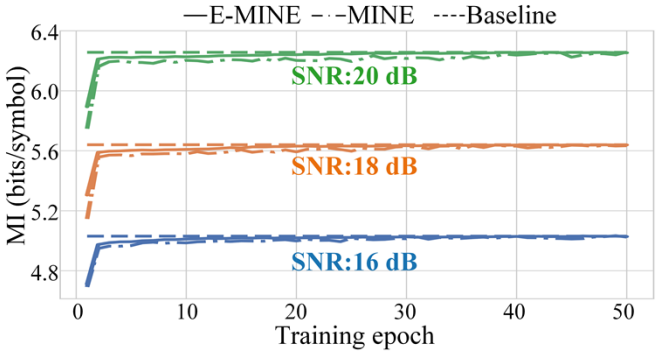

Fig. S3. MI estimation as the function of the training epoch in the AWGN channel.

We also evaluate the E-MINE and GH method for non-uniform probability distribution of the input signal using the Maxwell-Boltzmann (MB) distribution. As shown in Fig. S4, the E-MINE demonstrate accurate MI estimation for a wide range of channel conditions and signal distributions. The uniform and MB distribution markers demonstrate similar accuracy for MI estimation under different SNR conditions. These results highlight the potential of E-MINE for accurate MI estimation in the AWGN channel.

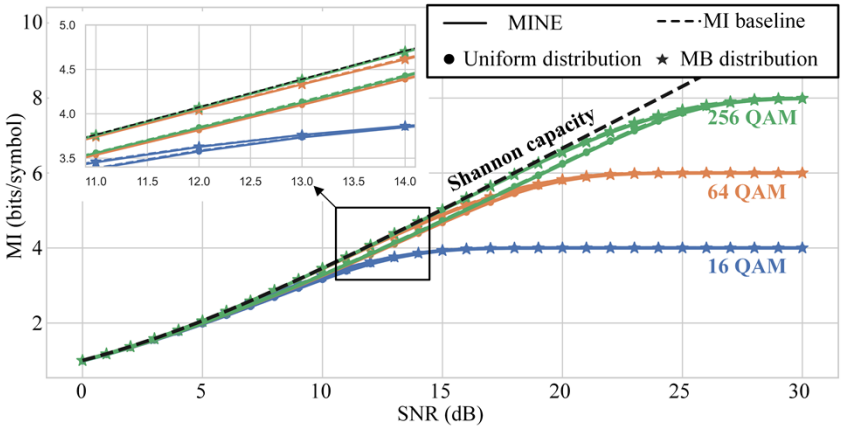

Fig. S4. MI estimation performance in the AWGN channel.

Figure S4 shows the maximum MI difference between E-MINE and MC estimation based on Gaussian distribution in the numerical optical fiber channel using split-step Fourier method. Indeed, there is a maximum MI difference with nonlinearity increase. To the extreme, the MI estimations of both methods tend to be zero because no information can be decoded through severe nonlinearity. The maximum MI difference is typically observed under 6 dBm to 15 dBm, where the power decreases with distance. As distance increases, the MI difference decreases,

indicating that the noise distribution tends to be Gaussian for long-haul transmissions. This conclusion aligns with the GN model prediction. However, in short-distance transmission, nonlinearity will differ from Gaussian noise, resulting in a higher difference.

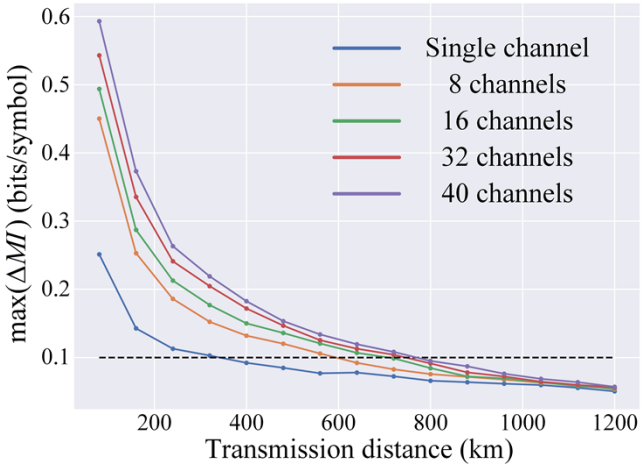

Fig. S5. The maximum MI differences as the function of transmission distance.

References

1. D. McAllester and K. Stratos, "Formal Limitations on the Measurement of Mutual Information," in Proc. Int. Conf. Artif. Intell. Stats., (2020), pp. 875-884.
2. M. S. Faruk and S. J. Savory, "Digital Signal Processing for Coherent Transceivers Employing Multilevel Formats," J. Lightwave Technol. **35**, 5 (2017).
3. D. Arnold, H.-A. Loeliger, P. Vontobel, *et al.*, "Simulation-based computation of information rates for channels with memory," IEEE Trans. Inf. Theory, **52**, 8 (2006).