Convolutional Neural Network Based Blind Estimation of Generalized Mutual Information for Optical Communication

Johannes Ackermann(1), Maximilian Schädler (1) and Christian Bluemm(1)

(1) Huawei Munich Research Center, Riesstr. 25, 80992 Munich, Germany,
maximilian.schaedler@huawei.com

Abstract *We propose blind GMI estimation from received constellation diagrams, based on a convolutional neural network. For coherent optical transmissions between 600Gb/s and 960Gb/s with varying modulation schemes, mean errors of 1.4% without and 0.8% with channel specific fine-tuning are achieved, significantly outperforming previous methods.*

Introduction

In modern optical communication networks, connections can be dynamically rerouted in order to adapt to changing network conditions. In-system performance monitoring allows to scope which links are able to offer the requested capacity. The ultimate, thus desired, monitoring metric would be the post-FEC BER. However, it is infeasible to reliably estimate it due to the target error rates lower than 10^{-15}. Hence, more common metrics for in-system performance monitoring are pre-FEC BER or Q-Factor, in combination with FEC limits.

Recently it has been shown that the generalized mutual information (GMI) is a better predictor for post-FEC BER than the pre-FEC BER, when using a soft decision (SD) FEC[1]. To determine GMI, however, knowledge of the transmitted bits is required, which is not available in a live system, or comes at the cost of added overhead when relying on dedicated preambles. We thus want to estimate GMI blindly, without knowledge of the transmitted bits.

While blind estimation approaches are widely known for OSNR[2], modulation schemes[3], baudrate[4], or BER[5], blind GMI estimation has not been studied yet to the best of our knowledge. What comes closest is an approach to blindly estimate the related asymmetric information (ASI)[6]. However, this approach severely lacks of accuracy in the experimentally assessed use cases of this paper: Coherent 88Gbaud DP-16QAM 600Gb/s, 92Gbaud DP-32QAM 800Gb/s and 92Gbaud DP-64QAM 960Gb/s optical back-to-back (BtB) measurements. In comparison, our approach of blindly estimating GMI from buffered received symbols with convolutional neural networks can achieve mean errors below 0.8%.

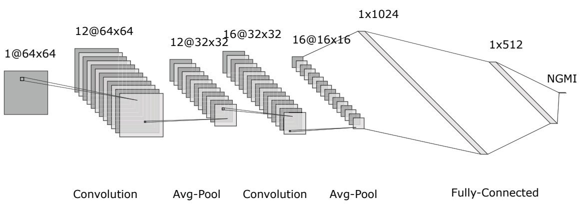

Fig. 1: Structure of the Convolutional Neural Network used to blindly predict NGMI.

Background

The logarithmic likelihood ratios, also called L values, are defined as[1]

$$L_k = \log \frac{\sum_{x \in \mathcal{X}_k^0} P_{X|C_k}(x|1) f_{Y|X}(y|x)}{\sum_{x \in \mathcal{X}_k^1} P_{X|C_k}(x|0) f_{Y|X}(y|x)}, \tag{1}$$

where X represents the transmitted symbols, Y the received complex symbols, and k denotes the bit which the likelihood ratio corresponds to. $f_{Y|X}(y|x)$ represents the likelihood of receiving symbol y when transmitting x over the channel, and $\mathcal{X}_k^b$ is the subset of $\mathcal{X}$ with bit b at position k. From the L values we can evaluate the GMI, an achievable information rate recommended in[1] to predict the the post SD-FEC BER:

$$\text{GMI} = \max_{s \geq 0} \sum_{k=1}^{m} \mathbb{E}\left[\log_2 \frac{f_{Y|B}(Y|B_k)^s}{\sum_{b \in \mathbb{B}} P_{B_k}(b) f_{Y|B_K}(Y|b)^s}\right] \tag{2}$$

The normalized GMI (NGMI)[7] is defined as $\text{NGMI} = 1 - \frac{1}{m}(\mathbb{H}(B) - \text{GMI})$. For mn_s transmitted bits $b_{k,l}$ the GMI is then

$$\text{GMI} \approx \tag{3}$$

$$m - \frac{1}{n} \min_{s \geq 0} \sum_{k=1}^{m} \sum_{l=1}^{n_s} \log_2\left(1 + \exp(s(-1)^{b_{k,l}} L_{k,l}\right).$$

We compare our method with a method that blindly estimates the ASI which is equivalent to the GMI, expressed in (2), with $s = 1$[8].

Convolutional Neural Network Based Blind Estimation of Generalized Mutual Information

Convolutional Neural Networks (CNNs)[9] are a subset of artificial neural networks, initially pro-

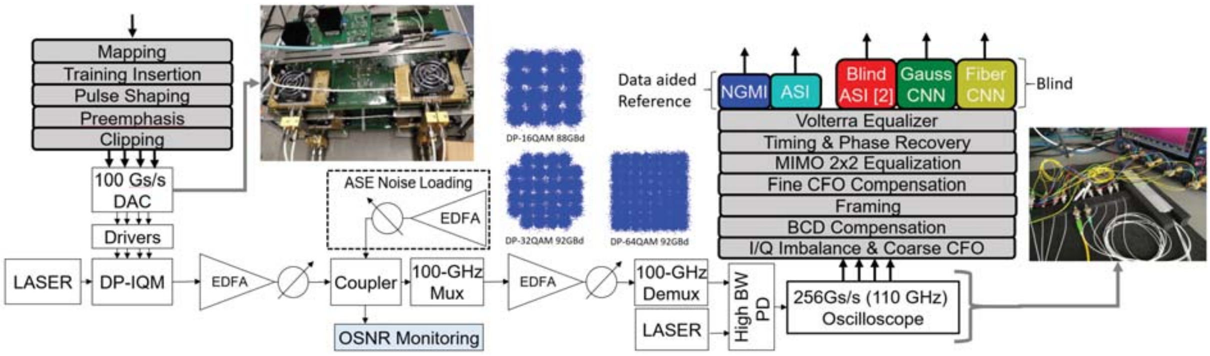

Fig. 2: Measurement setup including TX and Rx DSP with Blind GMI Monitoring

posed for image classification. In comparison to fully connected neural networks, CNNs comprise convolutional and pooling layers. The input of a convolution layer is convoluted by three-dimensional filters which are defined by the corresponding branch weights. Subsequently, the signal is fed into a pooling layer that reduces the dimension by compressing it with either max or average pooling. Our proposed network structure is shown in Fig. 1. It consists of a convolutional layer with 12 $5 \times 5 \times 1$ filters, followed by a 2×2 average-pooling layer, a convolutional layer with 16 $3 \times 3 \times 12$ filters and again 2×2 average-pooling. The output of the last average-pooling layer is flattened and used as input for two fully-connected layers with 1024 and 512 units, followed by a single output node. The input for our network is a normalized 2D histogram of received symbols, which can be regarded as an approximation of $f_Y(y)$. As activation we use ReLU functions[10] in all hidden layers and a sigmoid function on the output.

During the training phase, known transmitted sequences are used to evaluate the target NGMI and to generate the corresponding input 2D-histograms h for the CNNs. The histograms are normalized as follows $h_{i,j} = \frac{h_{i,j}}{\sum_{i,j \in [0, n_b in]} h_{i,j}}$. We generate our training set with 50000 histograms and corresponding target labels and use them to train the CNN using the Adam optimizer[11] with learning rate 0.001 for 3000 epochs[12]. The training is rerun for each modulation scheme separately. Each histogram is generated with 10000 symbols, and $[64 \times 64]$ uniformly distributed bins. Moderate variations of input data size and network structure did show no significant impact on performance. For improved training accuracy with a limited training set across the whole range of expected signal-to-noise (SNR) ratio or optical SNR (OSNR), the captured data is scrambled for CNN training.

Coherent Optical Measurement Setup

A coherent single-carrier dual-polarization (DP) transmission system over a single mode fiber (SMF) is employed to experimentally evaluate the performance of the proposed CNN based blind estimation of NGMI[13]. The setup and the offline DSP stack is shown in Fig. 2. The measurements were performed in a BtB configuration at 1550 nm with ASE noise loading, in order to compare NGMI at varying OSNR values. The electrical signals are generated by a 100 GSa/s DAC (Micram). Four SHF S804A amplifiers with 60 GHz bandwidth drive the RF signals. In the optical domain, an external cavity laser (ECL) source (1 kHz linewidth) generates a continuous wave signal which is modulated by a LiNbO3 DP-IQ modulator with 32 GHz bandwidth (Fujitsu-FTM7992HM). At the receiver side, the optical signal is combined with amplified spontaneous emission (ASE) noise generated by an EDFA and then amplified. After a stage of four 70 GHz balanced photodiodes, the electrical signals are captured by a 110 GHz bandwidth real-time oscilloscope operating at 256 GSa/s.

Performance and Accuracy Evaluation

We evaluate two versions of our approach: The "G-CNN", which is trained on samples generated by a simulated Gaussian channel, and the "F-CNN", which is trained on data from the real optical fiber channel. The methods are compared to the blind ASI estimation method proposed by Yoshida et al.[6], which we denote as "Blind ASI". Part of their approach compares the histogram of L-values to a set of candidate Gaussian functions. As our method is computationally more complex, we quadruple the number of these candidate functions to 32400 and use histograms with 256 bins, which slightly improves performance. Further increases did not yield no additional gain. Finally, the ASI is calculated from the discretized L-values, as in[6].

	Simulated Gaussian Channel			Measured Optical Channel		
	16-QAM	32-QAM	64-QAM	16-QAM	32-QAM	64-QAM
Proposed Blind F-CNN	1.52%	1.24%	0.74%	0.18%	0.76%	0.55%
Proposed Blind G-CNN	0.15%	0.21%	0.18%	1.36%	1.12%	0.76%
Blind ASI[6]	3.69%	3.47%	3.37%	2.26%	3.27%	5.44%

Tab. 1: Mean relative error in the estimation of NGMI, for Fiber CNN (F-CNN) and Gauss CNN (G-CNN), and mean relative error in the estimation of ASI, for blind ASI.

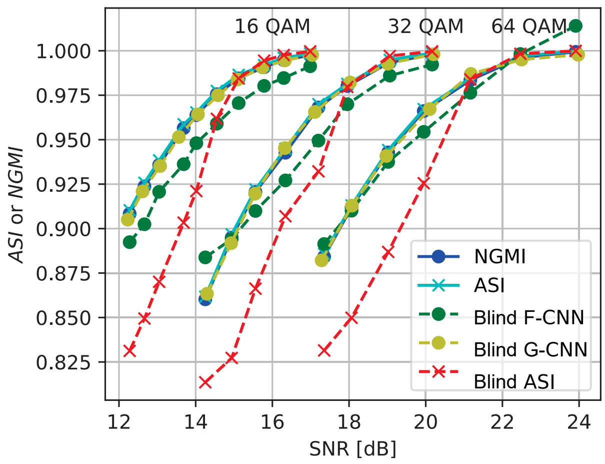

Fig. 3: Simulations on a Gaussian channel.

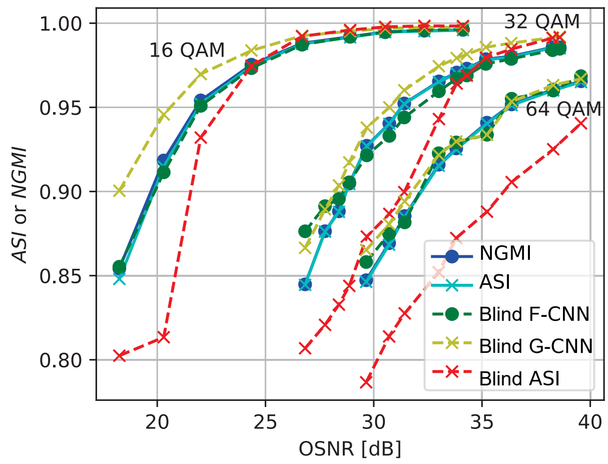

Fig. 4: Measurements on an optical back-to-back channel

We first evaluate our approach in simulations upon a Gaussian channel model on basis of 16QAM, 32QAM and 64QAM. The training and test data are generated independently. The results of the simulations for different SNRs are shown in Fig. 3 and Tab. 1. The Blind ASI method yields same accuracies as the reported accuracies by the authors in[6]. However, the G-CNN significantly outperforms Blind ASI for both, lower and higher SNR values.

The performance of a real optical fiber system is shown in Fig. 4 and Tab. 1 on basis of 88Gbd DP-16QAM 600Gb/s, 92Gbd DP-32QAM 800Gb/s and 92Gbd DP-64QAM 960Gb/s BtB measurements. The F-CNN is trained and hence fine-tuned on the first received frame and remains static for the following frames. This ensures strict separation of the training and testing data. It can be observed, that the F-CNN performs significantly better than the Blind ASI. The Blind ASI performs well on high NGMIs, however, loses accuracy on lower values, while our approach remains accurate. An advantage of the Blind ASI over the F-CNN is that it does not require any training data from the intended channel. Therefore, to provide a comparison in a setting where additional training on the intended channel is not possible, the performance of the G-CNN, which is trained only on the simulated Gaussian channel, is evaluated on the optical channel. The G-CNN performs significantly better than the blind ASI, despite not having additional information about the optical channel either. This indicates that G-CNN was able to learn a network that can be applied to realistic channels, without requiring access to specific training data. Additionally, to incomplete all combinations, the F-CNN, which is trained and fine-tuned on the optical channel, is applied on the simulated Gaussian channel. The corresponding performance is shown in Fig. 3. While less accurate is observed on very high NGMI values, the F-CNN overall outperforms the Blind ASI. This indicates that the F-CNN learned a general enough model, which performs well on other channels too, despite only having access to one channel behaviour during training.

Conclusion

We propose a method to blindly estimate the generalized mutual information based on a convolutional neural network. We evaluate two versions, one trained only on a simulated Gaussian channel and one trained and hence fine-tuned on a real optical fiber channel. We show significant accuracy improvements compared to previously proposed blind estimation methods in both, the simulated Gaussian as well as the real coherent optical channel. A mean error rates of 0.21% in simulation and 1.4% without and 0.8% with fine-tuning in measurements are achieved and verified accross a broad set of modulation schemes and high data rates.

References

[1] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, “Replacing the soft-decision FEC limit paradigm in the design of optical communication systems”, *JLT*, vol. 33, no. 20, pp. 4338–4352, 2015.

[2] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, “Convolutional neural network-based optical performance monitoring for optical transport networks”, *JOCN*, vol. 11, no. 1, A52–A59, 2019.

[3] J. Langston, S. Varughese, R. DeSalvo, and S. E. Ralph, “Blind cyclostationarity-based format classification for coherent links”, pp. 1–3, 2018.

[4] Y. Cheng, S. Fu, M. Tang, and D. Liu, “Multi-task deep neural network (mt-dnn) enabled optical performance monitoring from directly detected PDM-QAM signals”, *OE*, vol. 27, no. 13, pp. 19 062–19 074, 2019.

[5] Q. Zhang, Y. Yang, C. Guo, X. Zhou, Y. Yao, A. P. T. Lau, and C. Lu, “Accurate BER estimation scheme based on K-means clustering assisted Gaussian approach for arbitrary modulation format”, *JLT*, vol. 38, no. 8, pp. 2152–2157, 2020.

[6] T. Yoshida, M. Mazur, J. Schröder, M. Karlsson, and E. Agrell, “Performance monitoring for live systems with soft FEC and multilevel modulation”, *JLT*, vol. 38, no. 11, pp. 2912–2921, 2020.

[7] J. Cho, L. Schmalen, and P. J. Winzer, “Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped qam”, in *ECOC*, IEEE, 2017, pp. 1–3.

[8] T. Yoshida, M. Karlsson, and E. Agrell, “Performance metrics for systems with soft-decision fec and probabilistic shaping”, *PTL*, vol. 29, no. 23, pp. 2111–2114, 2017.

[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition”, *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[10] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines”, in *ICML-10*, 2010, pp. 807–814.

[11] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014.

[12] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[13] M. Schaedler, S. Calabro, F. Pittala, C. Bluemm, M. Kuschnerov, and S. Pachnicke, “Neural network-based soft-demapping for nonlinear channels”, in *OFC*, 2020, W3D–2.