Neural Network-Based Soft-Demapping for Nonlinear Channels

Maximilian Schaedler [1,2], Stefano Calabrò [1], Fabio Pittalà [1], Christian Bluemm [1], Maxim Kuschnerov [1], Stephan Pachnicke [2]

[1] Huawei Munich Research Center, Riesstr. 25, 80992 Munich, Germany
[2] Kiel University (CAU), Chair of Communications, Kaiserstr. 2, 24143 Kiel, Germany
maximilian.schaedler@huawei.com

Abstract: Conventional soft demappers designed for AWGN channels suffer from performance loss under realistic channels. We propose a neural network soft demapper and show a gain of 0.35dB in an 800Gb/s coherent transmission experiment using DP-32QAM.

OCIS codes: (060.1660) Coherent communication, (060.0060) Fiber optics and optical communication

1. Introduction

In modern communication systems, forward error correction (FEC) and high-order quadrature amplitude modulation (QAM) schemes are key technologies for realizing high spectral efficiencies (SE) [1]. Whereas the first optical communication systems relied on hard-decision (HD)-FEC, nowadays, modern systems use soft-decision (SD)-FEC. Current digital coherent systems typically employ bit-interleaved coded modulation (BICM) [2] and rely upon bit-wise (BW) decoders that separate signal recovery and FEC. A key component of a SD-FEC BICM system is the soft demapper, which computes the soft bits in the form of the so-called L-values, expressed as log-likelihood ratios (LLRs) [3]. The L-values feed the following error correction decoder and, thus, affect directly with their quality the overall system performance.

Classical approaches to evaluate the L-values, as the analytical computation of the a-posteriori LLR [4] or the less complex max-log approximation (MLA) [5], assume ideal channel compensation of the channel impairments and thus additive white Gaussian noise (AWGN). However, an optical communication system comprises even in back-to-back (BtB) configuration residual impairments from a large number of optical/electrical (O/E) components with nonlinear memory effects that, in practice, cannot be completely compensated. Therefore, computing the L-values under the AWGN assumption implies a performance penalty.

In recent years, machine learning (ML) methods have demonstrated amazing performance in various tasks and along with other disciplines, the optical communication community is adopting ML for a broad range of problems [6]. In a previous paper [7], where ML is applied for L-values computation, the focus was on imitating the analytical computation for AWGN channels with a neural network (NN) with reduced computational effort. In [8], the authors propose a highly complex soft-decision deep neural turbo equalizer with emphasis on long-haul scenarios. This contribution suggests a low-complexity NN application to the problem of soft-demapping on a non-AWGN channel. In particular, we propose the use of NN structures to learn the channel transition probabilities without any prior channel knowledge and hence to compute the optimal L-values. In order to maintain complexity low, we use neural structures with no more than one nonlinear hidden layer. The performance of the analytical approach and the MLA for AWGN channels is benchmarked against the NN-based soft demapper on the basis of captured samples in the case of BtB transmission of coherent dual polarization (DP) 32QAM. For the measurements, a DCI/DCN compatible net bit-rate of 800Gb/s with 15% FEC overhead has been chosen.

2. Principles of Neural Log-Likelihood Ratio (LLR) Demapping

A soft demapper provides the embedded soft bits of a received symbol. The soft bits, assessing the level of reliance in the quality of the symbols' demapping process, are usually expressed as log-likelihood ratios (LLRs). The demapper demaps the received complex symbol, $y \in \mathbb{C}$, into a vector of M-bit LLRs, $L \in \mathbb{R}^M$, corresponding to the probabilities of the transmitted bits of the binary label $\underline{c} = \{c_1, ..., c_M\}$. The a-posteriori LLR of the i-th bit is defined in [9] and given by equation (1). An analytic computation of the log-MAP expression (1) can only be achieved if the transition probabilities are known. Equation (2) represents the analytical solution of

$$L_i(y) = \log\left(\frac{\Pr(c_i=0|y)}{\Pr(c_i=1|y)}\right) \quad (1)$$

$$L_i(y)_{\text{MLA}} = \frac{1}{N_0}\left(\min_{x\in X(c_i=1)}(|y-x|^2) - \min_{x\in X(c_i=0)}(|y-x|^2)\right) \quad (3)$$

$$L_i(y)_{\text{AWGN}} = \log\left(\frac{\sum_{x\in X(c_i=0)} \exp\left(\frac{|y-x|^2}{N_0}\right)}{\sum_{x\in X(c_i=1)} \exp\left(\frac{|y-x|^2}{N_0}\right)}\right) \quad (2)$$

expression (1) under the AWGN assumption, while expression (3) represents the alternative low-complexity MLA. Both schemes, the analytical computation of the log-MAP expression and the MLA, assume an

AWGN channel. In systems with various impairments, e.g. optical systems, which are subject to phase noise and nonlinear effects in O/E components as well as in the fiber, the AWGN assumption leads to suboptimal LLR values. NNs are a convenient tool to predict probabilities and hence, to learn soft demodulation schemes, where a received symbol has to be classified to its transmitted bits. In the context of machine learning, the symbol-wise LLR computation can be formulated as a multi-label classification problem.

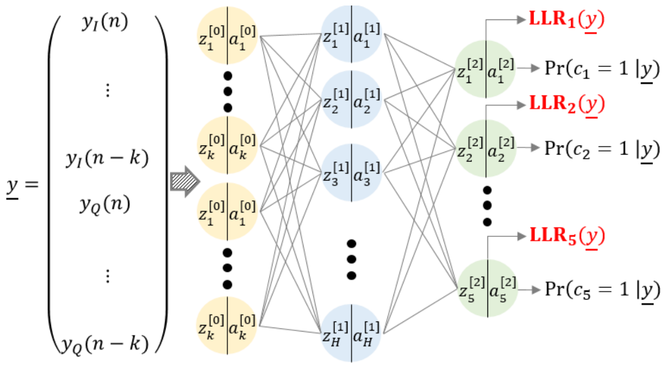


Fig. 1: Neural Network soft demapper

$$\underline{a}^{[0]} = \underline{y} \tag{4}$$

$$\underline{z}^{[l]} = \mathbf{W}^{[l]}\underline{a}^{[l-1]} + \underline{b}^{[l]} \tag{5}$$

$$\underline{a}^{[l]} = \sigma^{[l]}\left(\underline{z}^{[l]}\right) \tag{6}$$

$$\Pr(\underline{c} = 1|\underline{y}) = \sigma^{[2]}\left(\underline{z}^{[2]}\right) = \frac{1}{1+\exp(-\underline{z}^{[2]})} \tag{7}$$

$$\underline{L}(y)_{\text{NN}} = \log\left(\frac{1-\Pr(\underline{c}=1|\underline{y})}{\Pr(\underline{c}=1|\underline{y})}\right) = -\underline{z}^{[2]} \tag{8}$$

Fig. 1 shows the NN soft demapper as a multi-label classification network. We indicate the possibility to allocate extra input nodes in order to cope with channel memory-effects. Here k denotes the memory depth and H the number of neurons in the hidden layer. The notation is as follows [10] :

- $\underline{y}$: the received signal and its delayed versions, • $a_i^{[l]}$: the output of the i-th neuron in the l-th layer
- $\mathbf{W}^{[l]}$: weight matrix of the $(l-1)$-th layer to l-th layer, • $\sigma_i^{[l]}$: activation function of the i-th neuron in the l-th layer
- $z_i^{[l]}$: summed up input to the i-th neuron in the l-th layer, • $\Pr(c_i = 1|y)$: conditional probability of the i-th bit.

The corresponding mathematical description is given next to the figure. Equation (4) denotes the input layer and the second, as well as the third lines, equations (5) and (6), are executed iteratively for subsequent layers to obtain the output $\underline{a}^{[2]}$. In classification networks, the last layer contains a nonlinear function $f(a) = \log(\frac{a}{1-a})$: $[0,1] \rightarrow [-\infty, \infty]$ to restrict the outputs of the neurons to the interval $[0,1]$, which is suitable to represent the probabilities of each class. By inverting this function $f(\cdot)$, we obtain the sigmoid function and hence the activation function of the last layer. The sigmoid is applied to each neuron at the output layer to enable multi-classification. The NN models the conditional probability of $c_i = 1$ given $\underline{y}$ as the output of the NN with respect to the input $\underline{y}$ and the sigmoid activation function for each class or bit independently. The final vector of M-bit LLRs per received symbol is obtained by combining expression (1) and (7), shown by equation (8). It can be observed that the unknown optimal L-values are learned in an unsupervised manner inside the neurons in the output layer by using the sigmoid activation function and the cross-entropy [10] as cost function. After the training phase, the output layer activation functions can be replaced by linear functions to obtain the L-values instead of the conditional probabilities.

3. Measurement Setup and Results

A coherent single-carrier transmission system over a single mode fiber (SMF) is employed to evaluate experimentally the performance of the proposed scheme. The quality of the obtained L-values is evaluated by means of the generalized mutual information (GMI) [1]. The DP-32QAM 800Gb/s experimental setup and the offline DSP stack are shown in fig. 2. The measurements were performed in BtB configuration at a wavelength of 1550nm with ASE noise loading, in order to compare GMI at varying OSNR loads. For a net bit-rate of 800Gb/s, 920Gb/s of pseudo random data including FEC overhead are transmitted at 92 GBd. The photos show a 100GSa/s DAC (Micram) with 40GHz 3dB-bandwidth and 4.5 ENOB and a 160 GSa/s Keysight Infiniium oscilloscope with

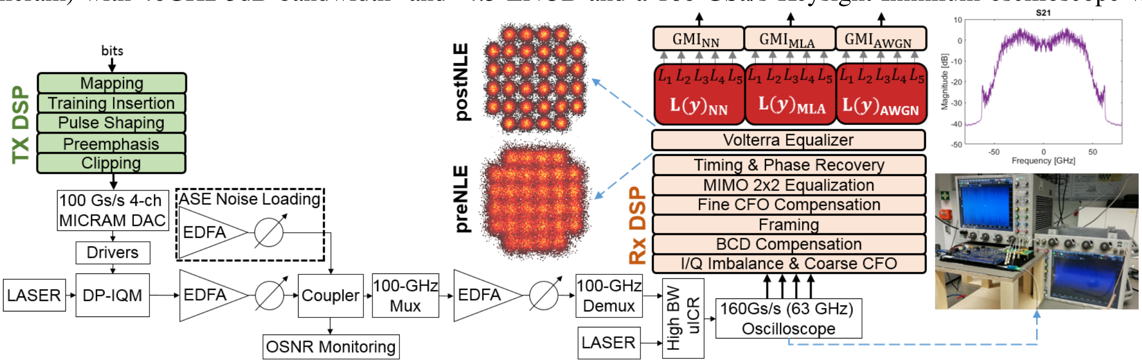
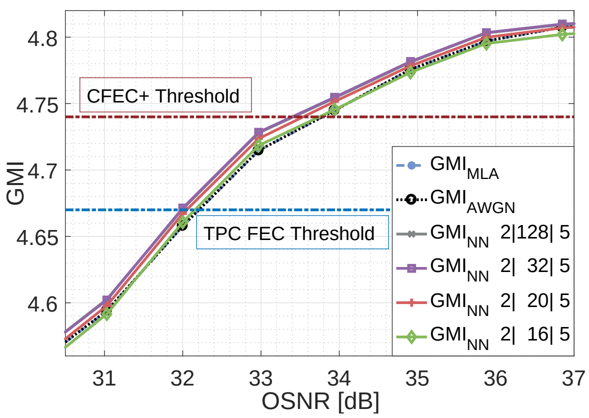
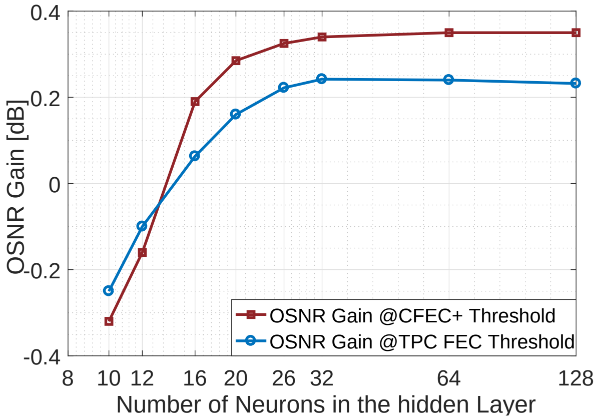


Fig. 2: Measurement setup including Tx and Rx DSP with different methods to compute the L-values

63 GHz bandwidth. The setup includes also an SHF 804A driver, a Fujitsu 64GBd DP-IQM and a NeoPhotonics 64GBd class-40 HB-μICR (40GHz). In this experiment we focus on the impairments, including the nonlinear distortion, introduced by the components. Longer fiber links would introduce additional impairments. The DSP stack includes classical signal recovery blocks. In order to compensate the nonlinearities and memory effects, a full kernel 2nd/3rd order Volterra nonlinear equalizer (NLE) with five memory taps is used. The different soft-demappers are operated after the nonlinear equalizer. They process the same signal and are executed in parallel to ensure fair comparison. Training of the NN is essential before deployment. In our experiment this is done upon the payload of the first captured frame for one of the OSNR values. Once trained, the NN provides the L-values for all other captures without further training.



(a) GMI of the L-values and the transmitted bits related to the OSNR.



(b) Gain in OSNR at different FEC thresholds related to the number of neurons in the hidden layer

Fig. 3: Optical back-to-back 800G/λ GMI measurements results

Fig. 3 shows the performance of the different soft-demappers for the optical BtB 800G measurements. On the left hand side, fig. 3a plots the GMIs related to the OSNR of the computed L-values and the transmitted bits. The analytical solution and the MLA (both computed under the AWGN assumption) provide nearly equal performance. In order to ensure fair comparison, the memory k of the NN is set to zero. Thus, the NN structures only differ in the number of neurons in the hidden layer. The notation, e.g. 2/32/5, indicates the number of input neurons (2 for I & Q with 0 memory taps) followed by the neurons in the hidden layer (32) and terminated by the number of L-value output nodes (5). Fig. 3b shows the improvements of the base-line curve. The gains in OSNR at two FEC thresholds are plotted against the number of neurons in the hidden layer. The design with 32 neurons outperforms the methods based on the AWGN assumption by up to 0.35 dB. The gain saturates and remains almost identical even with a much larger design with 128 neurons. On the contrary, if less than 32 neurons are used, the performance and the gain decrease immediately. Apparently the optimal number of neurons is directly connected with the number of constellation points, since for a NN it is easier to learn the desired task, if each particular neuron represents only one and not multiple possible combinations of the input.

4. Conclusion

In this paper, we studied the achievable rates on a non-Gaussian channel in case of 32QAM transmission for conventional (mismatched) soft demappers developed under the AWGN assumption and for a novel adaptive soft-demapper based on machine learning techniques. We showed that the AWGN assumption leads to suboptimal soft-information, also in the presence of powerful channel equalization. The proposed low-complexity neural network-based soft-demapper is able to model the transition probabilities of arbitrary channels without any prior information. Consequently, it can provide optimal soft information also in the presence of colored noise, phase noise, and nonlinear impairments and memory effects from optical and electrical components. The proposed scheme is experimentally assessed and shown to outperform the conventional soft demappers by up to 0.35 dB on a non-AWGN channel.

References

1. Alvarado, A., et al. "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems." JLT (2015)
2. Szczecinski, L., and Alvarado, A. "Bit-interleaved coded modulation: fundamentals, analysis and design." John Wiley and Sons, 2015.
3. Ivanov, M., et al. "On the information loss of the max-log approximation in BICM systems." IEEE Trans. Inf. Theorie (2016)
4. Erfanian, J., et al. "Reduced complexity symbol detectors with parallel structure for ISI channels." TCOM (1994)
5. Robertson, P., et al. "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain." ICC (1995)
6. Khan, F. N., et al. "An Optical Communication's Perspective on Machine Learning and Its Applications." JLT (2019)
7. Shental, O., et al. ""Machine LLRning": Learning to Softly Demodulate." arXiv preprint arXiv
8. Koike-Akino, Toshiaki, et al. "Neural Turbo Equalization to Mitigate Fiber Nonlinearity." ECOC (2019).
9. Hagenauer, J., et al. "Iterative decoding of binary block and convolutional codes." IEEE Trans. Inf. Theorie (1996)
10. Bishop, C. M. "Pattern recognition and machine learning." springer, 2006.
11. Hirbawi, Y. "Continuation & Results of FEC Proposals evaluation for ITU G.709.3 200-400G 450km Black Link", ITU-T SG15-CD11