

Nonlinear Component Equalization: A Comparison of Deep Neural Networks and Volterra Series

Maximilian Schädler⁽¹⁾, Georg Böcherer⁽¹⁾, Francesca Diedolo⁽²⁾, Stefano Calabrò⁽¹⁾

⁽¹⁾ Huawei Munich Research Center, Riesstr. 25, 80992 Munich, Germany ⁽²⁾ Technical University of Munich, Theresienstr. 90, 80333 Munich, Germany, maximilian.schaedler@huawei.com

Abstract Coherent optical transmission systems suffer from distortions induced by nonlinear components. As a countermeasure, Volterra equalizers and deep neural networks have attracted growing attention. In this paper, optimal objectives to maximize achievable rates as well as performance and complexity aspects are discussed. ©2022 The Author(s)

Introduction

In optical fiber communication, especially in short reach communication, optical and electrical components introduce non-linearities, which require effective compensation to attain highest data rates^[1]. Equalizers based on truncated Volterra series are a popular countermeasure for receiver-side equalization of nonlinear component impairments and their memory effects. However, Volterra nonlinear equalizer (VNLE) architectures are generally very complex.

This contribution investigates time delay deep neural network (TDNN) architectures as an alternative for nonlinear equalization and places special attention on hard and soft-demapping and on the corresponding loss functions. The paper makes use of the following publications^{[2]–[9]}.

Volterra Nonlinear Equalizer

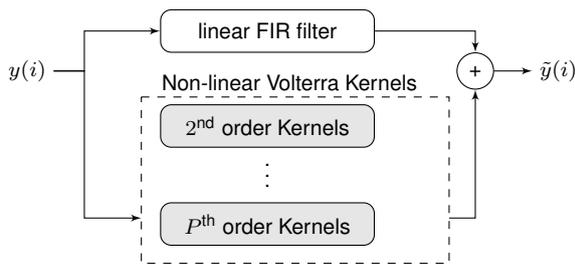


Fig. 1: Blockdiagram of a VNLE.

Fig. 1 illustrates the structure of a VNLE^{[10],[11]}. By combining linear convolution and non-linear power series, the VNLE is capable to describe time-invariant non-linear systems with finite fading memory. Let $y(i)$ and $\tilde{y}(i)$ represent a system with single input and single output, respectively, the p -th order causal discrete time Volterra series is given by^[12]

$$\tilde{y}(i) = \sum_{p=1}^P \sum_{s_1=0}^{M_1} \cdots \sum_{s_p=0}^{M_p} h_p(s_1, \dots, s_p) \prod_{k=1}^p y(i-s_k), \quad (1)$$

where $h_p(s_1, \dots, s_p)$ denotes the p -th order

Volterra kernel, M_1 the memory length for the linear terms and M_2 to M_p the memory lengths for the non-linear terms of second order and higher. The relationship between the memory length and the number of kernels of order p is given by^[3]

$$N_p = \frac{1}{p!} \prod_{i=0}^{p-1} (M_p + i). \quad (2)$$

Their number is directly connected to the required number of multipliers^[3]

$$N_{\text{mul-VNLE}} = N_1 + \underbrace{\sum_{i=2}^P N_i}_{\text{Kernels}} + \underbrace{M_i + 1}_{\text{Feature Matrix}^{[13]}}. \quad (3)$$

Eq. (3) is derived from the structure presented in^[13]. It considers terms that can be obtained by delaying other terms as well as the reuse of products of order k to compute products of order $k+1$.

Time Delay Deep Neural Network Equalizer

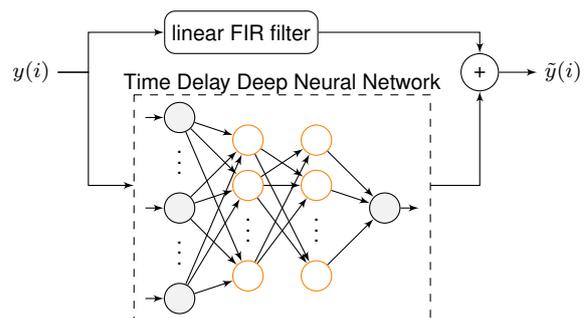


Fig. 3: Blockdiagram of a TDNNE.

Fig. 3 shows the block diagram of a TDNNE with multiple hidden layers. The memory is again considered by processing time delayed versions of the observed signal. This allows the TDNNE to have, similar to the VNLE, a finite dynamic response to time series input data and to describe causal time-invariant non-linear systems with finite fading memory. Compared to the VNLE, which represents the solutions of nonlinear differential systems based on its Volterra series and

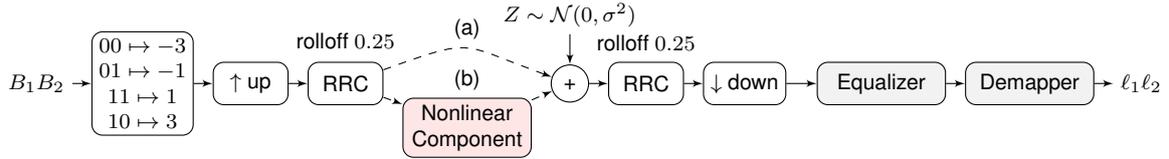


Fig. 2: Simulated link. (a) AWGN channel (b) Nonlinear Channel $y = \tanh(\text{filter}(x) + 0.2)$ with 13 taps.

hence models non-linearities with polynomials, the TDNNE computes the solutions to a large class of general nonlinear systems on basis of the nonlinear activation functions. The memory length M of the TDNNE step delay line is a single parameter, while the memory length of each Volterra order is individually adjustable. Let $\mathbf{y}(i) = [y(i-M), \dots, y(i), \dots, y(i+M)]$ denote the delayed symbol-level input vector, the TDNNE with L -layers is given by

$$\mathbf{a}^{[0]}(i) = \mathbf{y}(i), \quad (4)$$

$$\mathbf{a}^{[l]}(i) = g(\mathbf{W}^{[l]} \mathbf{a}^{[l-1]}(i) + \mathbf{b}^{[l]}), \quad l=1, \dots, L \quad (5)$$

$$\tilde{\mathbf{y}}(i) = \mathbf{a}^{[L]}(i) \quad (6)$$

where $\mathbf{a}^{[l]}$ denotes the output vector of the l -th layer, $\mathbf{W}^{[l]}$ the weight matrices and $\mathbf{b}^{[l]}$ the bias vectors. For the activation function g , we use the non-linear ReLU function for the hidden layers and the linear function for the output layer. The total required number of multipliers is defined as^[3]

$$N_{\text{mul-TDNNE}} = \underbrace{\sum_{i=1}^{L-1} s_i s_{i+1}}_{\text{Weights}} + \underbrace{\sum_{i=2}^{L-1} s_i}_{\text{Activation Function (if slope } \neq 1)}, \quad (7)$$

where $s = s_0 |s_2| \dots |s_L$ denotes the design, i.e., s_i is the number of neurons in the i -th layer.

Training of Linear and Non-linear Equalizer

Before operations, the parameters of the VNLE and the TDNNE have to be identified, i.e., configured upon training data, in order to match the non-linearities of interest. We consider the scenario of Fig. 4, where the nonlinear equalizer is followed by a soft-demapper, generating the soft-bits for the FEC decoder.

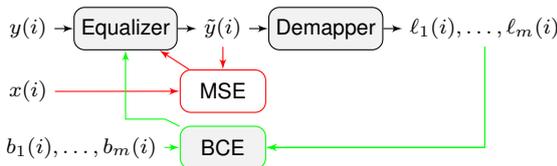


Fig. 4: Blockdiagram of the training process.

It has been shown that the training with respect to the mean square error (MSE) between the equalizer's output $\tilde{y}(i)$ and the target symbols $x(i)$ with 1 sample-per-symbol (SpS) each minimizes the pre-FEC SER, however, it does not maximize the achievable rate of a system with

SD-FEC. For maximizing the achievable rate, a more appropriate approach is to train an equalizer using the Binary-Cross-Equivocation (BCE) loss

$$\mathcal{L}(b, \ell) = \log_2[1 + \exp(-(1 - 2b)\ell)], \quad (8)$$

where b is the transmitted bit and where ℓ is the soft-demapper output. To optimize an equalizer w.r.t. the BCE loss the soft-demapper has to be differentiable in order to propagate the gradient backwards through the soft-demapper to the equalizer. In this paper a max-log approximation (MLA) soft demapper is used. The MLA uses piecewise differentiable linear approximations as shown in Fig. 5 for the case of Gray mapped 4ASK. During training the BCE loss uses $\pm\infty$ as targets for the soft-bits, corresponding to 0 and 1. This pushes the outer constellation points towards infinity. To counteract this effect we introduce a penalty function represented by the green line in Fig. 5. An alternative solution to this problem is shown in^[14], where an entropy-regularized MSE based loss function is proposed.

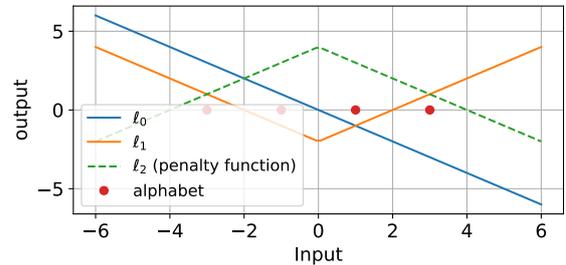


Fig. 5: Transfer function of the MLA soft-demapper (4ASK).

Numerical Study over an AWGN Channel

To highlight the advantages of the BCE loss, we first consider option (a) in Fig. 2, i.e., an AWGN channel without any impairments. Fig. 7 depicts the performance in terms of BER and information rate when a 5th-order VNLE or a TDNNE with 1 tap is applied. Both architectures are trained w.r.t to MSE or BCE. While all equalizers achieve optimal performance in terms of BER, regardless of which loss is used for training, only the equalizers trained w.r.t BCE achieve optimal performance in terms of information rate.

Fig. 6 shows the learned transfer functions (TF) and the corresponding histograms of the equalized signals. The MSE trained non-linear equalizers have the ability of concentrating the constellation points and to saturate the possible out-

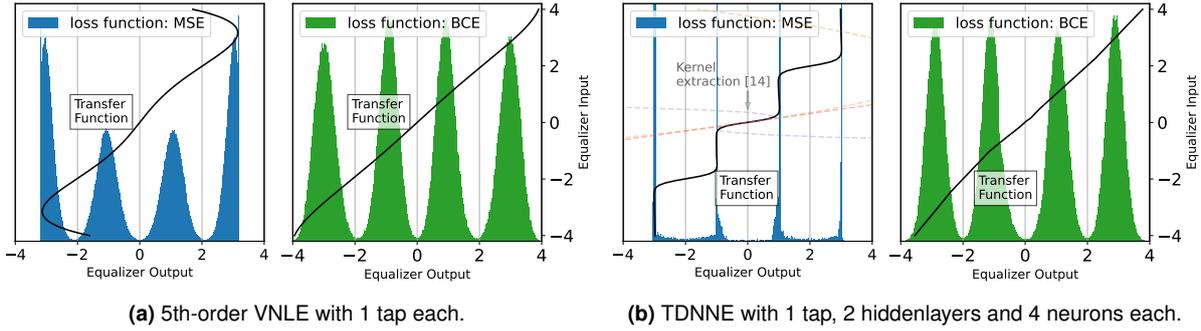


Fig. 6: Output histograms and TF of the non-linear equalizers trained w.r.t. MSE (left) and BCE (right). Also shown is the TF based on the extracted Volterra kernels, proposed by D.I. Soloway^[15], of order $k = 1, 2, 3$ at $x = 0$ from the MSE-trained TDNNE with \tanh instead of ReLU. According to Taylor's theorem, this approach is only valid in a limited area, hence it is non applicable.

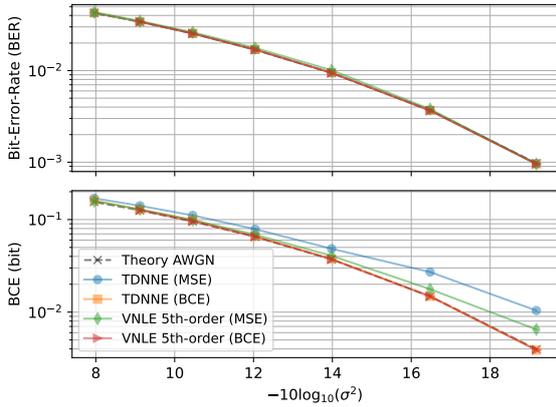


Fig. 7: Equalizers applied on an AWGN channel.

puts to the target points, in order to reach a low value of the cost function, i.e., to reach the optimal MSE for non-linear equalization^[16]. This saturation causes non-Gaussian distributed noise and induces soft information loss which impairs the evaluation of the soft-bits by the soft-demappers. This behavior is more pronounced for VNLEs with higher orders, for deeper TDNNEs, and for smaller constellation, respectively. In comparison, the output signals of the BCE trained equalizers exhibit Gaussian distributed noise.

Numerical Study over a Non-linear Channel

To compare the equalization performances and complexities in a non-linear scenario, we consider option (b) in Fig. 2. The non-linear component emulates a driver/MZM-modulator. It exhibits a bandwidth limitation and hence introduces intersymbol interference (ISI) as well as a non-linear distortion by a $\tanh(x + \text{offset})$. Fig. 8 shows the performance in terms of BCE by applying a linear equalizer, a VNLE and a TDNNE. All architectures are optimized regarding complexity, e.g., the VNLE architectures are optimized with respect to the number of taps. However, the architectures are not pruned, hence fully connected. A comparison of pruned VNLE and TDNNE can be found in^[3]. It can be observed that 1st, 2nd

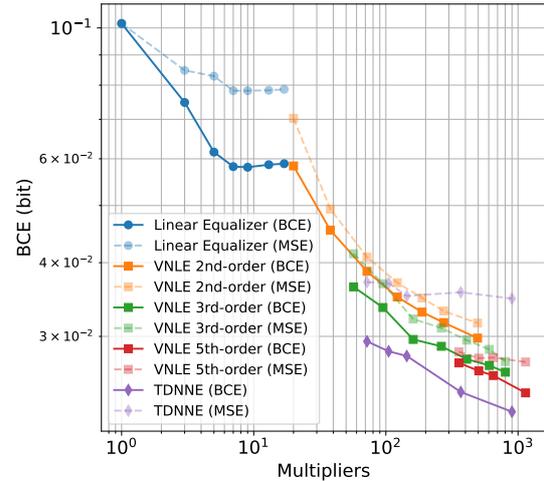


Fig. 8: Equalizers applied on a non-linear channel.

and 3rd order kernels yield the major benefit. 1st-order kernels compensate for ISI, 2nd-order kernels are pronounced due to the presence of an offset (e.g. bias voltage), while the 3rd-order kernels compensate for the \tanh behavior. However, the best performance and complexity trade-off can be achieved with the TDNNE architecture. All options outperform all VNLE architectures with equal complexity. This indicates that in this scenario, where typical optoelectronic components are considered, the ReLUs of the TDNNE approximate systematic non-linearities more efficiently than the polynomials of a 5th-order VNLE.

Conclusions

The growing demand for non-linear equalizers has motivated investigations in optimal objectives for soft-decision FEC scenarios. Applied to Volterra non-linear equalizers and time delay deep neural network equalizers, it is shown that deep neural networks reflect typical non-linearities more accurately and efficiently than Volterra series. This makes deep neural networks a promising candidate for mid-term deployment in non-linearity impaired optical systems.

References

- [1] C. Bluemm, M. Schaedler, S. Calabrò, *et al.*, "Equalizing nonlinearities with memory effects: Volterra series vs. deep neural networks", in *45th ECOC*, IET, 2019, pp. 1–4.
- [2] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, and S. Pachnicke, "Deep neural network equalization for optical short reach communication", *Applied Sciences*, vol. 9, no. 21, p. 4675, 2019.
- [3] M. Schaedler, G. Boecherer, and S. Pachnicke, "Soft-demapping for short reach optical communication: A comparison of deep neural networks and Volterra series", *J. Lightw. Technol.*, 2021.
- [4] M. Schädler, G. Böcherer, F. Pittalà, *et al.*, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions", *J. Lightwave Technol.*, vol. 39, no. 16, pp. 5278–5286, Aug. 2021.
- [5] M. Schaedler, F. Pittalà, G. Böcherer, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions", in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333204.
- [6] M. Schaedler, S. Calabrò, F. Pittalà, *et al.*, "Neural network assisted geometric shaping for 800Gbit/s and 1Tbit/s optical transmission", in *Proc. Opt. Fiber Commun. Conf.*, IEEE, 2020, pp. 1–3.
- [7] M. Schaedler, S. Calabrò, F. Pittalà, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Neural network-based soft-demapping for nonlinear channels", in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2020, pp. 1–3.
- [8] M. Schaedler, F. Pittalà, S. Calabrò, G. Böcherer, C. Bluemm, and S. Pachnicke, "Recurrent neural network soft demapping for mitigation of fiber nonlinearities and ISI", in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2021, pp. 1–3.
- [9] M. Schädler, "Machine learning in digital signal processing for optical transmission systems", Ph.D. dissertation, Christian-Albrechts-Universität of Kiel, 2022.
- [10] A. Rezaia, J. Cartledge, A. Bakhshali, and W.-Y. Chan, "Compensation schemes for transmitter-and receiver-based pattern-dependent distortion", *PTL*, vol. 28, no. 22, pp. 2641–2644, 2016.
- [11] J. C. Cartledge, "Volterra equalization for nonlinearities in optical fiber communications", in *SPPCom*, JOSA, 2017, SpTu2F–2.
- [12] L. Guan, *FPGA-based Digital Convolution for Wireless Applications*. Springer, 2017.
- [13] C. Tarver, A. Balatsoukas-Stimming, and J. R. Cavallaro, "Design and implementation of a neural network based predistorter for enhanced mobile broadband", in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, IEEE, 2019, pp. 296–301.
- [14] F. Diedolo, G. Böcherer, M. Schädler, and S. Calabro, *Nonlinear equalization for optical communications based on entropy-regularized mean square error*, 2022. DOI: 10.48550/ARXIV.2206.01004. [Online]. Available: <https://arxiv.org/abs/2206.01004>.
- [15] D. I. Soloway and J. T. Bialasiewicz, "Neural network modeling of nonlinear systems based on Volterra series extension of a linear model", in *Proc. IEEE International Symposium on Intelligent Control (ISIC)*, 1992, pp. 7–12.
- [16] G. Böcherer, *Lecture notes on machine learning for communications*, [online]. Available: <http://georg-boecherer.de/mlcomm.pdf>, accessed: 14-06-2022.