800Gb/s PAM4 Transmission over 10km SSMF Enabled by Low-complex Duobinary Neural Network Equalization

Tu3C.4

Christian Bluemm⁽¹⁾, Bo Liu⁽²⁾, Bing Li⁽²⁾, Talha Rahman⁽¹⁾, Md Sabbir-Bin Hossain⁽¹⁾, Maximilian Schaedler⁽¹⁾, Ulf Schlichtmann⁽²⁾, Maxim Kuschnerov⁽¹⁾, Stefano Calabrò⁽¹⁾

⁽¹⁾ Huawei Technologies Duesseldorf GmbH, Munich, Germany, christian.bluemm@huawei.com ⁽²⁾ Technical University Munich, Chair for Electronic Design Automation, Munich, Germany

Abstract On 10km 200Gb/s per lane IM/DD PAM4 CWDM4 O-band measurements, neural network equalization meets Volterra equalization performance with 30% less hardware multiplier complexity. Key enabler against strong CD penalties at these reaches/rates is duobinary training. ©2022 The Authors

Introduction

For optical short reach use cases like data centre interconnects (DCI) or data centre networks (DCN), transceivers must balance objectives of tight footprints, power and cost against rising capacity needs. One of the most discussed topics in this field remains the question about the better system choice: Intensity-modulated with direct detection (IM/DD) or coherent. Although coherent systems allow for much higher robustness against channel penalties, global sales of PAM4 chipsets should surpass sales of coherent chipsets in 2022 [1]. Further growth, however, is more and more hindered by IM/DD's inherent low tolerance to polarization mode dispersion (PMD) and even more critical, chromatic dispersion (CD). CD describes quasi-static, linear phase distortions with large memory, arising from varying group velocities of different optical frequency components. Square law photo-detection turns this linear problem into a nonlinear one [2]. CD becomes an increasingly severe bottleneck with higher reach and rate.

In the past, multi-lane coarse wavelength division multiplexing (CWDM) proved to be an effective remedy to keep baud rates low with rising payload rates. However, scaling beyond 4 parallel lanes, (CWDM4) is widely considered ineffective in terms of complexity [3].

This paper discusses one of the most urgent nextgeneration IM/DD use cases, 800Gb/s PAM4 O-band transmission. Instead of scaling beyond CWDM4, we focus on low-complexity digital signal processing (DSP) at the receiver (Rx) to enable a higher bitrate per lane. Interestingly, already for 100Gb/s per lane (400GBASE-LR4), experts predict that CD dispersion enforces a reduction of classical Ethernet "LR" reach from 10 km to 6 km [4].

In this paper, 200Gb/s per lane is realized via a line rate of 112GBd PAM4, accounting for typical forward error correction (FEC) overheads. Our approach to compensate for CD distortions is a combination of feed forward equalization (FFE) in form of nonlinear equalization (NLE) and maximum likelihood sequence estimation (MLSE). While low complex MLSE variants can tackle CD still effectively enough in combination with NLEs, classical Volterra-based NLEs (V-NLEs) become quickly too complex, even without exceeding 3^{rd} order Volterra components.

We show that neural network NLEs (NN-NLEs) can be a very effective alternative to match performance in terms of pre-FEC bit error rate (BER), while saving more than 30% of complexity vs. V-NLEs in terms of number of multipliers. This is due to NN's vast and flexible parameter space, which may compensate channel impairments very accurately. Unlike other fields, long training times of NNs is no issue here, since NNs can be pre-trained for quasi-static distortions like CD before deployment in an ASIC. A second typical challenge of NNs, a lack of training data, does not apply here, either, as abundance of optical transmit data can be generated in very short time. With these advantages, NN-enabled IM/DD has been studied in different flavours for a while [5-9] and lately even for 200Gb/s per lane PAM4 transmission for 1km [10] and 2km [11]. In our 10km case, CD becomes by far the most dominant distortion. Our key enabler is the introduction of duobinary (DB) training targets. DB signalling intentionally introduces controlled inter-symbol interference (ISI) in such a way that the resulting signal requires less bandwidth (BW) [12]. In combination with FFE, DB can significantly reduce BW restrictions and CD. To the best of our knowledge, DB training targets have not been considered for NN-NLEs before.

Volterra Nonlinear Equalization

With x(n) and y(n) representing system input and output, respectively, the *Pth* order discrete time Volterra series with M_p memory taps for order p is [13]:

$$y(n) = \sum_{p=1}^{P} \sum_{m_1=0}^{M_1} \cdots \sum_{m_p=0}^{M_p} h_p(m_1, \cdots, m_p) \prod_{k=1}^{p} x(n - m_k)$$
(1)

Fig. 1 illustrates a corresponding equalizer configuration, which maps current input samples x(n) and



historic samples *x(n-m)* as a linear combination of nonlinear functions (kernels). Implemented as such, each *pth*-order Volterra ker-

Tu3C.4

Fig. 1: Volterra nonlinear equalizer

nel $h_p(m_1,...,m_p)$ describes all possible combinations of a product of p time shifts of the input signal up to the memory m_p . Pruned versions with reduced kernel sets are not considered here.

The V-NLEs in this paper identify optimal kernels in terms of the least squares (LS) error criterion upon training data, which are received sequences with their known transmitted counterparts [14].

According to [13], the number of multipliers for V-NLE equals the sum of unique kernels across all orders, with C_{M+1}^1 unique kernels for first, $C_{M+1}^2 + C_{M+1}^1$ for second and $C_{M+1}^3 + 2C_{M+1}^2 + C_{M+1}^1$ for third order,

with
$$C_m^p = \frac{m!}{(m-p)!p!}$$
 (2)

Neural Network Nonlinear Equalization



Fig. 2 shows the basic processing unit of any NN, the artificial neuron. Weights $w_1,...,w_K$ scale the input signals $x_1,...,x_K$ before feeding their sum to a nonlinear activation function. The bias *b* off-

Fig. 2: Artificial neural network with tanh and H-tanh activation

sets the output. For our NN-NLEs, *tanh* serves as activation for initial training and is then replaced by its low-cost variant *H*-*tanh* for further training as described in [15]. *H*-*tanh* requires only one hardware multiplier and two comparators for clipping.

Fig. 3 shows a corresponding equalizer configuration, which maps memory at the input with a tapped delay line and then nonlinearities with hidden layers (HL) and an output layer, which is chosen as a single linear neuron for optimal performance in our case. Our HLs are all fully connected. As with V-NLEs, pruned subsets are out of scope.

For identification of our adjustable NN-NLEs parameters (weights, biases) against CD distortions, we apply minibatch backpropagation with ADAM optimization [16, 17] within numerous iterations (epochs)



about 50k to 100k until convergence. The total required number of multipliers for a NN-NLE with H-tanh activation function is defined as:

$$mul_{NN-NLE} = \sum_{i=1}^{d-1} s_i s_{i+1} + \sum_{i=2}^{d-1} s_i$$
(3)

where *d* is the number of layers including input and output layer and where $s = s_1/s_2/.../s_d$ describes the NN design with s_i neurons in the *i*-th layer [15]. The first sum relates to the number of weights and the second sum to the number of H-tanh activations.

Duobinary Training Target

Classical equalizers model inverse linear/nonlinear channel characteristics for compensation. In contrast, our training target is not the transmitted data sequence directly, but its DB version, which reduces the overall signal BW [18]. The advantage is less noise enhancement in those high frequency components, which suffer the most from channel-induced BW limitations and CD, i.e. those with lowest signal to noise ratio. Fig. 4 exemplifies this benefit with actual 112Gbd measurement data and their corresponding power spectral densities (PSD).

In the top plots of Fig. 4, the green Rx signal spectra are input for both PAM4 equalizers – with classical training targets (left) and DB target (right). The blue PSDs after equalization reveal much lower relative noise enhancements with the DB target (bottom right) compared to without (bottom left). While this example is based on a V-NLE, it applies equally to NN-NLE equalization. In practice, changing the training target from original transmitted sequences c_n to their DB counterpart \tilde{c}_n means to apply a simple DB filter:

$$\tilde{c}_n = c_n + c_{n-1} \tag{4}$$

For PAM4, this DB filter turns four modulation levels into seven. The DB filter is only required for training. For the deployed system, the simplest operations to recover PAM-4 symbols from the equalizer output is modulo 4 reduction (Mod4) [19].

Mod4 can be implemented in ASICs without additional hardware costs, by truncating all bit levels but the two least significant bits.

A more complex, but more effective alternative to Mod4 is classical MLSE, which recovers the enforced ISI with theoretically best possible performance.



Fig. 4: Measured power spectral densities of Rx signals for PAM4 (left top) and DB-PAM4 (right top) before and after V-NLE equalization with noise enhancements (bottom)



Tu3C.4

Fig. 5: Experimental 10km setup with DSP. Inset (a) shows classical CWDM4 O-band wavelengths, (b) optical spectra before and after optical filtering and (c) DB-targeted PAM4 equalization input and output with its seven modulation levels

Experimental setup

Fig. 5 depicts the IM/DD measurement setup and offline DSP of our 10 km PAM4 transmission with 112 Gbd per lane. 3 dB BWs are mentioned below each electro/optical component.

At the transmitter (Tx), pseudorandom binary sequences (PRBS) are Gray-mapped to PAM-4 symbols and pulse-shaped with a raised cosine filter with a roll off factor of 0.14. After resampling, the sequences match the 120 GS/s arbitrary waveform generator (AWG), which converts to analog signals. A 60 GHz driver amplifier (DA) amplifies towards O-band Mach Zehnder modulation (MZM). While the focus are standard O-band CWDM4 wavelengths 1270nm, 1290m, 1310nm and 1330nm [20] as illustrated in Fig. 5 (a), further captures at in-between wavelengths allow for better insights of the performance/wavelength relationship.

After 10 km transmission over standard single mode fiber (SSMF), a variable optical attenuator (VOA) controls the received optical power (ROP) at the input of a Praseodymium-doped fiber amplifier (PDFA). An optical filter suppresses the broadband noise of the PDFA. Fig. 5 (b) shows good noise suppression without cutting the optical signal spectrum.

The filtered optical signal is fed to a photo diode (PD) and its electrical output is digitized at 256 GS/s by a real time digital oscilloscope. The ROPs at the PDFA input were tuned to yield approximately 7 dBm optical power at the PD for optimal performance.

The offline Rx DSP starts with resampling and timing recovery. Equalization is done under one sample per symbol signalling. Fig. 5 (c) shows histograms of the received PRBS data before and after DB-targeted equalization. MLSE with Euclidian distance metric and complexity-optimized memory length of only 1 or Mod 4 DB recovery is applied to the equalizer outputs before BERs are counted.

Results

Fig. 6 presents pre-FEC BERs vs. wavelengths. Accumulated CD values for 10 km, as defined in [20], are added to the x-axis. Data sets for BER estimation were

strictly separated from those for training of parameters (kernel coefficients for V-NLEs and weights/biases for NN-NLEs).

Fig. 6 presents three DSP configurations, linear equalization (LE) in green color, V-NLE with [1st,2nd,3rd] order memory taps in blue and NN-NLE in red. Space restrictions allow only to show a subset of vast sweep studies regarding memory taps for LE and V-NLE and number of hidden layers and neurons per layer for NN-NLEs. All results included in Fig. 6 optimize performance vs. complexity, meaning more complex variants improve results only insignificantly.

Dotted lines represent classical training for LE and V-NLE, dashed lines DB training with Mod4 and solid lines with added MLSE. Massive DB gains are obvious. Depending on the FEC in use, MLSE might be optional. For 800Gb/s IM/DD at 10 km, many FEC options are still under discussion [21]. The NN-NLE architecture 21|11|7|1 matches the pre-FEC BER performance of V-NLE [21/9/7], but requires only 334 hardware multipliers instead of 486 for V-NLE, thus >30% less.



Fig. 6: Linear equalization (LE) vs. V-NLE vs. NN-NLE with and without DB target, with and without MLSE

Conclusion

This paper presents Rx DSP options for 10km IM/DD CWDM4 transmission with 200Gb/s per lane. With a duobinary training target, NN-NLEs match V-NLE performance with more than 30% complexity reduction in hardware multipliers.

- Lightcounting, "Emerging Market for PAM4 and Coherent DSPs", 3rd ed., 2021
- [2] M. Chagnon, "Direct-Detection Technologies for Intraand Inter-Data Center Optical Links," in *Proc. OFC*, 2019
- [3] J. Wei et al., "Experimental comparison of modulation formats for 200 G/λ IMDD data centre networks," in *Proc.* ECOC, 2019
- [4] Lewis, D., "Ethernet PMDs and Chromatic Dispersion" ITU-T, 2020
- [5] S. Gaiarin *et al.*, "High Speed PAM-8 Optical Interconnects with Digital Equalization Based on Neural Network," in *Proc. ACP*, 2016
- [6] G. Reza and J. K. Rhee, "Nonlinear Equalizer Based on Neural Networks for PAM-4 Signal Transmission Using DML," *IEEE Photonics Tech. Lett.*, vol. 30, no. 15, pp. 1416–1419, 2018
- [7] P. Li *et al.*, "100Gbps IM/DD Transmission over 25km SSMF using 20G-class DML and PIN Enabled by Machine Learning," in *Proc. OFC*, 2018
- [8] F. Da Ros, S. M. Ranzini, R. Dischler, A. Cem, V. Aref, H. Bülow, D. Zibar, "Machine-learning-based equalization for short-reach transmission: neural networks and reservoir computing," Proc. SPIE, Metro and Data Center Optical Networks and Short-Reach Links IV, 2021
- [9] Z. Xu, C. Sun, T. Ji, J. H. Manton and W. Shieh, "Feedforward and Recurrent Neural Network-Based Transfer Learning for Nonlinear Equalization in Short-Reach Optical Links," in *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 475-480, 2021
- [10] B. Sang et al., "Multi-Symbol Output Long Short-Term Memory Neural Network Equalizer For 200+ Gbps IM/DD System," in Proc. ECOC, 2021
- [11] Hiroki Taniguchi, Shuto Yamamoto, Akira Masuda, Yoshiaki Kisaka, Shigeru Kanazawa, "800-Gbps PAM-4 2km Transmission using 4-λ LAN-WDM TOSA with MLSE based on Deep Neural Network", in *Proc.* OFC, 2022
- [12] J. Proakis and M. Salehi, "Digital Communications", New York, NY, USA:McGraw-Hill, pp. 610-619, 2001
- [13] L. Guan, "FPGA-Based Digital Convolution for Wireless Applications". Berlin, Germany: Springer, 2017
- [14] Kim, J. and Konstantinou, K. "Digital predistortion of wideband signals based on power amplifier model with memory", Electronics Letters, 37, (23), pp. 1417–1418, 2001
- [15] M. Schädler, G. Böcherer and S. Pachnicke, "Soft-Demapping for Short Reach Optical Communication: A Comparison of Deep Neural Networks and Volterra Series," in *Journal of Lightwave Technology*, vol. 39, no. 10, pp. 3095-3105, 2021
- [16] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in Neural Networks for Perception, Academic Press: Cambridge, MA, USA, pp. 65–93, 1992
- [17] D. P. Kingma and L. J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. ICLR, 2015
- [18] Gutiérrez-Castrejón, R. Saber, M.G.; Alam, M.S.; Xing, Z.; El-Fiky, E.; Ceballos-Herrera, D.E.; Cavaliere, F.; Vall-Llosera, G.; Giorgi, L.; Lessard, S.; Brunner, R.; Plant, D.V. "Systematic Performance Comparison of (Duobinary)-PAM-2,4 Signaling under Light and Strong Opto-Electronic Bandwidth Conditions", MDPI Photonics, 8, 81, 2021

- [19] Prodaniuc, C., "Advanced Signal Processing for Pulse-Amplitude Modulation Optical Transmission Systems", 2019
- [20] ITU-T Rec. G.652 "Characteristics of a single-mode optical fibre and cable", Series G: Transmission Systems and Media, Digital Systems and Networks – Optical fibre cables, 2016
- [21] He, X. "FEC Architecture of B400GbE to Support BER Objective" IEEE 802.3 Beyond 400G Study Group, 2021