Machine Learning-based Pre-Equalizers for Maximum Likelihood Sequence Estimation in High-Speed PONs

Wouter Lanneer
Fixed Networks Department
Nokia Bell Labs
Antwerp, Belgium
wouter.lanneer@nokia-bell-labs.com

Yannick Lefevre
Fixed Networks Department
Nokia Bell Labs
Antwerp, Belgium
yannick.lefevre@nokia-bell-labs.com

***Abstract*—High-speed passive optical networks (PONs) use advanced signal processing techniques like inter-symbol interference (ISI) equalization. While equalizers based on maximum likelihood sequence estimation (MLSE) via the Viterbi algorithm achieve excellent performance, they suffer from excessive implementation complexity except for very short channel responses. In this work, we employ a pre-equalizer for joint "channel shortening" and branch metric computation needed for the Viterbi algorithm. We then propose an optimization method for iteratively updating the pre-equalizer towards optimal end-to-end MLSE performance, by minimizing the multi-class cross-entropy loss based upon the path metrics. Numerical evaluations demonstrate that our proposed solution for MLSE with a small number of taps achieves significant ISI equalization improvements w.r.t. prior art approaches, and a performance close to MLSE with a high number of taps.**

I. Introduction

Passive optical networks (PONs) provide a widely-used technology for delivering multi-gigabit broadband access like e.g. fiber-to-the-home services. Research and standardization efforts in PON are currently driven towards increasing the signaling rate per wavelength, while keeping costs low by employing bandwidth-limited opto-electronic receivers in combination with advanced digital signal processing (DSP) techniques [1]. For instance, the latest PON standard developed by ITU-T, called G.9804, defines 50G binary signaling whereas typically 25G class receivers will be used in deployment. Hence, such PONs experience high levels of inter-symbol interference (ISI), arising from strong chromatic dispersion and bandwidth-limited reception.

The feed-forward equalizer (FFE) and decision-feedback equalizer (DFE) are popular choices for ISI equalization due to their low implementation complexity and robust performance. In this work, ISI equalizers based on maximum likelihood sequence estimation (MLSE) [2] are considered, since they may provide significant performance gains over FFE/DFE. A well-known efficient, programming solution for MLSE is obtained by the Viterbi Algorithm [3]. The classic MLSE provides a hard-output (binary) sequence, whereas soft-output variants exist as well, like for instance the soft-output Viterbi algorithm (SOVA) or the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm.

A key challenge of Viterbi algorithms is the notoriously high implementation complexity. The data path structure (i.e., the number of Viterbi states) scales exponentially with the number of ISI channel taps. Except for very short channels, the hardware implementation may require excessive power and area consumption. Other issues involve performance degradation and unpredictable results due to non-ideal noise assumptions. In general, the absence of noise correlation between different received samples is assumed, whereas such noise correlation is often present in practical systems due to for instance bandwidth limitations.

A. Related work

Already in the 1970s the combination of linear equalization with Viterbi algorithms has been investigated to reduce the overall complexity [4]–[7]. The purpose of the linear equalizer is to preceed the Viterbi algorithm and to reduce the number of consecutive transmit symbols that affect any received sample, i.e., to "shorten" the channel impulse response to an acceptable target length. These early works propose different strategies for choosing the target channel impulse response that is perceived by the Viterbi algorithm after pre-equalizing the original channel. Once the target channel is determined, the optimal pre-equalizer may be computed via a closed-form expression (or using adaptive gradient-based schemes).

More recently, in the context of (coherent) optical systems [8] and 50G-PON [9], the combination of an FFE with a digital postfilter has been proposed as pre-equalizer. The FFE removes all ISI (up to some residual error) at the cost of strong increased noise coloring by filtering the received samples. The postfilter then acts as a whitening filter by re-introducing some ISI taps in a controlled manner equal to the target number of Viterbi taps L. The noise whitening filter can be considered here the target channel impulse response.

This work was supported by the Flemish Government funding agency VLAIO, through the SPIC (HBC.2020.2197) and SPARTA (HBC.2022.0428) projects. The authors would also like to thank Vincent Houtsma and Dora van Veen for providing 50Gb/s PON measurement traces.

Alternatively, so-called deep neural networks (DNN) together with machine learning (ML) techniques have also been studied for ISI equalization in optical systems. In addition to full DNN-based equalizers [10]–[12], also hybrid DNN-Viterbi solutions are proposed in [13]. Notice in [13] the presence of channel shortening equalizers is assumed if necessary. DNN-based equalizers have several advantages like data-driven learning without any signal model assumptions, advanced data filter structures that are demonstrated in many cases to achieve excellent performance. However, they require elaborated offline training procedures on large data-sets via advanced optimizers, whereas re-training during run-time is limited and very costly. In addition, large networks are used in combination with pruning and/or multiple initialization strategies to avoid training convergence to poor performing solutions. The run-time complexity may be excessive high due to a large number of weights and layers.

B. Main contributions

To enable the use of a small number of Viterbi states, channel shortening solutions are adopted in [4]–[9] by employing pre-equalizers. Although they may work reasonably well in some practical settings, these solutions suffer from ISI leakage and inherently introduce additional noise coloring by filtering the received samples. They remain heuristic in nature and are not guaranteed to work optimally under all conditions. In this work, we propose to train a pre-equalizer via ML-based techniques in order to optimize the end-to-end MLSE performance. The novelty exists in using the internal Viterbi path metrics within the learning loss function. Hence, the pre-equalizer will learn to directly compute the Viterbi input metrics while taking any channel shortening requirements into account. This model-free or data-driven learning, without any channel and/or noise assumptions, yields high performance in case of strong non-ideal transmission impairments. Further, as pre-equalizer we propose a fully linear structure by means of a filter-bank comprising a set of parallel finite impulse response (FIR) filters. Combined with a small, efficient Viterbi algorithm, this results in cheap(er) run-time complexity compared to a DNN structure with many layers and weights [9], [13], without sacrificing any performance as will be demonstrated via simulations (see Section IV).

II. Viterbi Algorithm for ISI Equalization

Consider digital transmission over a PON communication link as illustrated in Fig. 1. Without loss of generality, non-return to zero on-off keying (NRZ-OOK) modulation is used, such that the digital transmit signal x_k at time index k is represented by the binary constellation $\{+1, -1\}$. The digital received samples at time index k are denoted by y_k and obtained by the ADC sampler. The dispersive and noisy end-to-end channel includes fiber propagation as well as the hardware-induced distortion effects. It is assumed to be static, causal and with finite memory of length M. That is, any received sample y_k is affected by maximum M consecutive transmit symbols.

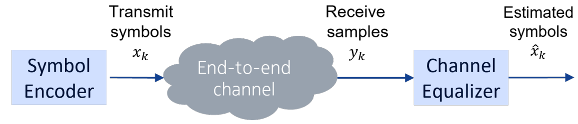

Fig. 1: Digital representation of next-generation PON communication link.

Standard MLSE for ISI equalization boils down to selecting the most likely (i.e. the one with maximum likelihood) symbol sequence from all possible transmit sequences given the received samples [2]:

$$\begin{aligned}\mathbf{x}^* &= \arg\max_{\mathbf{x}} P(\mathbf{y}|x_1, \dots, x_K)\\ &= \arg\min_{\mathbf{x}} -\log(P(\mathbf{y}|x_1, \dots, x_K))\end{aligned} \tag{1}$$

where $P(y|x_1, \dots, x_K)$ is the probability that $\mathbf{y} = [y_1, \dots, y_K]$ is observed given that the $\mathbf{x} = [x_1, \dots, x_K]$ is transmitted.

A well-known efficient MLSE implementation is the Viterbi algorithm. While a detailed explanation of the Viterbi algorithm is out of scope, a few key concepts will be introduced that are relevant for understanding the pre-equalizer. Note that the pre-equalizer works for any MLSE variant by using the appropriate path metrics.

A. Viterbi Algorithm

Generally, an L-tap MLSE equalizer assumes $L \geq M$ is the expected number of consecutive transmit symbols that affect any received sample. Consequently, it can be written that

$$P(\mathbf{y}|x_1, \dots, x_K) = \prod_{k=1}^{K} P(y_k|x_{k-L+1}, \dots, x_k). \tag{2}$$

such that the problem (1) reduces to

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{k=1}^{K} -\log(P(y_k|\mathbf{x}_k)), \tag{3}$$

where $\mathbf{x}_k = [x_{k-L+1}, \dots, x_k]$ is defining the transmit symbol sub-sequence of length L.

Viterbi-type algorithms solve problem (3) by finding the optimal path through a 2^{L-1}-state trellis. Each state $s \in [1, 2^{L-1}]$ corresponds to one particular symbol sub-sequence $[x_{k-L+2}, \dots, x_k]$ being transmitted at time step k from 2^{L-1} feasible sub-sequence combinations. The transition from a previous state s' (corresponding to the sequence $x_{k-L+1}, \dots, x_{k-1}$) to the current state s (corresponding to the sequence $x_{k-L+2}, \dots, x_k$) is associated with a likelihood $c_k(s, s')$.

$$c_k(s, s') = -\log(P(y_k|x_{k-L+1}, \dots, x_k)). \tag{4}$$

These are also called the branch metric (BMs). Further, let $v_k(s)$ denote the likelihood of state s at time k in the trellis

$$v_k(s) = -\log(P(y_1, \dots, y_k|x_{k-L+2}, \dots, x_k)). \tag{5}$$

These are the so-called path metrics (PMs). Key is that the optimal path is computed incrementally using the previously computed path and branch metrics, by sequentially processing all received samples and pruning unlikely paths at every step. That is, the PM $v_k(s)$ can be computed recursively from the two feasible predecessor states s' as

$$v_k(s) = \min_{s'}[v_{k-1}(s') + c_k(s,s')], \quad \forall k > 1. \tag{6}$$

The standard trace-back approach is to accumulate PMs for up to a depth of $D = 5L$ before making a final symbol decision [14] with negligible performance loss.

For ease of notation in Section III, we also define the extended PMs as:

$$\tilde{v}_k(s,s') = v_{k-1}(s') + c_k(s,s') \tag{7}$$

These extended PMs, denoted as $\tilde{v}_k(s,s')$, correspond to the log of the joint probability on state s and s', and are equivalent to the log probabilities before the minimization to calculate the regular $v_k(s)$ as $\min_{s'}[\tilde{v}_k(s,s')]$.

B. BM Computation $c_k(s,s')$

To implement performant ISI equalization with the Viterbi algorithm, accurate BM computation is required by estimating the log-likelihood function $c_k(s,s')$ in (4). Hence, the probability density functions (PDF) of the received samples of the communication channel need to be learned or estimated. In general, the received samples corresponding to every transmit sequence $\mathbf{x}_k$ are modelled by a white Gaussian PDF with a mean and noise variance. The means may then be estimated via tracking a linear FIR channel [14]. Alternatively, 2^L separate means and noise variances can be directly estimated as well, allowing for nonlinear channel and signal dependent Gaussian noise modelling. More advanced BM models compute the transition probabilities via histogram metrics measured at the receiver or employ DNNs trained offline [13].

Important to consider is the case where the number of Viterbi taps is smaller than the channel ISI length ($L < M$). Such a small number of Viterbi taps is often used regardless of the channel ISI length in order to limit the implementation complexity. Consequently, the estimation accuracy of the BM $c_k(s,s')$ will degrade and the assumption in (2) will no longer be valid. To prevent performance degradation of the Viterbi algorithm, a (linear) pre-equalizer is typically employed to "shorten" the ISI channel to L taps as outlined in Section I-A. In general, any other non-ideal channel and/or noise assumptions may degrade the Viterbi algorithm if not properly taken into account.

III. Novel ML-based Pre-Equalizer for Viterbi Algorithms

We specifically consider ISI equalization with the Viterbi algorithm where the number of Viterbi taps is smaller than the channel ISI length ($L < M$). To overcome in such scenarios the limitations of the heuristic approaches outlined in Section I-A, we propose a novel ML-based pre-equalizer for the Viterbi algorithm as illustrated in Fig. 2. Key is that the pre-equalizer is trained using a learning loss function based on the Viterbi extended PMs, in order to directly optimize the end-to-end performance. In fact, the pre-equalizer effectively combines the roles of "channel shortening" and BM computation given the desired number of Viterbi taps L.

Although any data filter structure can be used for the pre-equalizer (like DNNs for instance), we propose a single bank of 2^L FIR filters, yielding efficient run-time and training complexity via adaptive gradient-based schemes. Each filter estimates the optimal BM $\hat{c}_k(s,s')$ for one specific extended state (s,s'):

$$\hat{c}_k(s,s') = \mathbf{w}(s,s')^T \mathbf{y}_k + b(s,s'), \tag{8}$$

where $\mathbf{y}_k = [y_{k-N+1+\Delta}, \ldots, y_{k+\Delta}]$ is the current observation of N received samples with delay parameter Δ; and $\mathbf{w}(s,s') = [w^1_{(s,s')}, \ldots, w^N_{(s,s')}]$ and $b(s,s')$ are the filter taps and bias tap respectively corresponding to state (s,s'). Note that the bias tap can be considered an FIR tap to which the input sample is a unit-valued constant.

In the remainder of this section, we derive an optimal loss function and propose an iterative training procedure that minimizes the optimal loss function.

A. Optimal loss function

Observe the Viterbi algorithm may be seen as a probabilistic state classifier of new received samples, in which the states correspond to the consecutive (overlapping) binary transmit sub-sequences of length $L-1$. Every time step k, the current state is classified by selecting the minimum extended PM of all pairs of s and previous state s' in (6). Hence, the pre-equalizer can be trained by using supervised ML techniques based upon the so-called multi-nomial or multi-class cross-entropy loss function of the extended PMs $\tilde{v}_k(s,s')$, averaged over all training examples k:

$$\mathcal{L} = -\frac{1}{K}\sum_k \sum_{s,s'} \delta(s = s^*_k, s' = s^*_{k-1}) \log\left(\frac{\exp\left(-\tilde{v}_k(s,s')\right)}{\sum_{p,p'} \exp\left(-\tilde{v}_k(p,p')\right)}\right) \tag{9}$$

Here s^*_k and s^*_{k-1} are the true realized states at time steps k and $k-1$, corresponding to the actual transmitted bits. The Dirac delta function $\delta(s = s^*_k, s' = s^*_{k-1}) = 1$ if the extended state (s,s') corresponds to the true realized states (s^*_k, s^*_{k-1}), and is zero otherwise. The factor $\sum_{p,p'} \exp\left(-\tilde{v}_k(p,p')\right)$ normalizes the probabilities of the different states at time k to the range $[0,1]$.

Using such a loss function minimizes the Kullback-Leibler divergence between the true states and the classified states corresponding to the extended PMs inside the Viterbi algorithm. In other words, minimizing this loss function will drive the pre-equalizer towards computing BMs $\hat{c}_k(s,s')$ that in combination with the Viterbi algorithm yield estimated probabilities $\tilde{v}_k(s,s')$ close to one for the transmitted states and close to zero for the non-transmitted states. The Viterbi algorithm will hence effectively solve problem (3) without assumption (2) being valid, enabling end-to-end optimization of the MLSE-based equalizer. Notice the pre-equalizer will not explicitly aim to compute the exact BM corresponding to the true probability in (4), but will merely provide a parametric BM estimate that optimizes the proposed loss function.

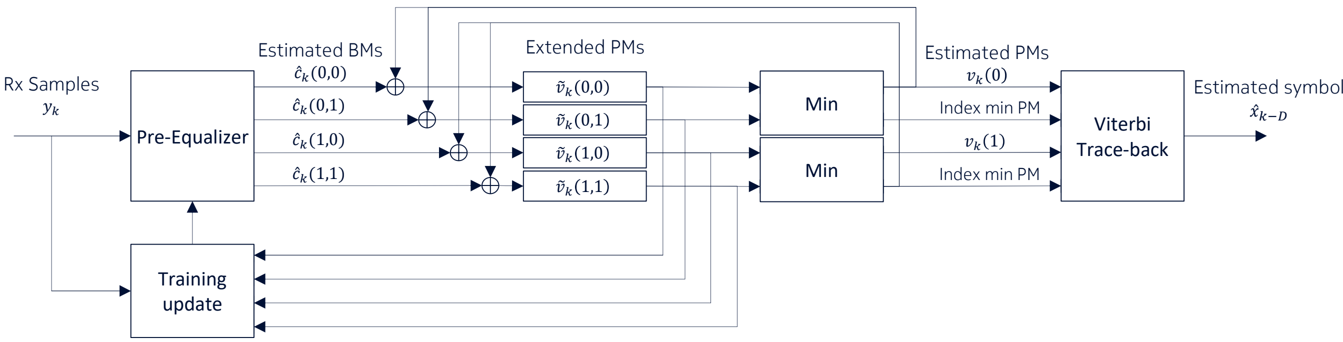

Fig. 2: Block diagram of the training method for the proposed ML-based pre-equalizer in combination with the Viterbi algorithm for ISI equalization. The case for binary transmission is depicted with $L = 2$ Viterbi taps and D denoting the trace-back depth.

It is stressed that using the BMs $\hat{c}_k(s, s')$ within the cross-entropy loss function, e.g., as proposed in [13], is not working well for the case $L < M$. Using the BMs would directly optimize the estimation of $P(y_k|x_{k-L+1}, \ldots, x_k)$ in (4) based on the current window of received samples being inputted to the pre-equalizer, without taking into account the information all previous received samples required for optimizing the objective in (3). This approach does not cope well with for instance noise coloring or channel impulse responses longer than L as a consequence.

B. Optimization Algorithm

The so-called stochastic gradient descent (SGD) algorithm is the workhorse of ML techniques, and may be used here as well for adapting the coefficients of the pre-equalizer structure towards minimizing the proposed loss function. The derivative of the loss function w.r.t. pre-equalizer $\mathbf{w}_{(s,s')}$ for one training example k is calculated as follows:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_{(s,s')}} = \left(\delta(s = s_k^*, s' = s_{k-1}^*) - \frac{\exp\left(-\tilde{\nu}_k(s,s')\right)}{\sum_{p,p'} exp(-\tilde{\nu}_k(p,p'))}\right) \cdot \left(\mathbf{y}_k + \frac{\partial \nu_{k-1}(s')}{\partial \mathbf{w}_{(s,s')}}\right), \tag{10}$$

where the derivative of $\nu_{k-1}(s')$ takes the contribution of all previously selected BMs into account. We have observed in simulations that ignoring the derivative of $\nu_{k-1}(s')$ during training yields negligible loss after convergence. In principle, any number of previous BMs could be taken into account for the derivatives and the update rules. The derivative for $b(s, s')$ can be derived similarly. Mini-batch training may be used for faster convergence, together with possibly random update orders. The optimization method for the ML-based pre-equalizer is listed as Algorithm 1 with μ denoting the learning rate. Since the proposed loss function is convex for affine input structures, the training algorithm is guaranteed to converge to the optimal solution for linear pre-equalizers.

Algorithm 1: Proposed SGD algorithm for the ML-based pre-equalizer with MLSE

repeat
  for *every training example* k **do**
    Compute $\hat{c}_k(s,s') = \mathbf{w}(s,s')^T \mathbf{y}_k + b(s,s')$
    Compute $\tilde{\nu}_k(s,s') = \nu_{k-1}(s') + \hat{c}_k(s,s')$
    Update $\mathbf{w}(s,s') = \mathbf{w}(s,s') - \mu \nabla_{\mathbf{w}(s,s')} \mathcal{L}$
    Update $b(s,s') = b(s,s') - \mu \nabla_{b(s,s')} \mathcal{L}$
    Update $\nu_k(s) = \min_{s'} [\tilde{\nu}_k(s,s')]$
  end
until *convergence*

IV. Simulations

We numerically evaluate the proposed pre-equalizer with MLSE first for an ideal 5-tap FIR channel with additive white Gaussian noise (AWGN) according to

$$y_k = \sum_{m=1}^{M} h_m x_{k-m} + n_k$$

where the variance of noise signal n_k corresponds to the inverse SNR. The simulated bit error rate (BER) results are shown in Fig. 3 for a symmetric ISI channel $[0.1, 0.3, 0.9, 0.3, 0.1]$. The proposed pre-equalizer consists of four (parallel) 16-tap filters which are trained with the optimal PM-based loss function. Here the 5-tap MLSE with exact BM calculations is the theoretically-optimal ISI equalizer. The proposed ML-based pre-equalizer with MLSE is able to approach this theoretic optimum with $L = 3$ or $L = 5$ taps. Yet for higher SNR levels there is a small performance loss. Keep in mind that the filter-bank-based pre-equalizer is only computing parametric estimates of the BMs. More advanced pre-equalizer structures might provide small gains at high SNR levels. However, typically the FEC BER threshold is located at low SNR and high BER levels.

Second, we numerically evaluate different ISI equalizers for experimental 50Gb/s PON traces, which have been obtained according to the experimental setup described in [12], [15]. A Mach-Zehnder modulator has been used with the laser wavelength set to 1342 nm and 30km single-mode fiber, in order to achieve a high level of fiber dispersion around 83 ps/nm. The photoreceiver consists of a 25 Gb/s class avalanche

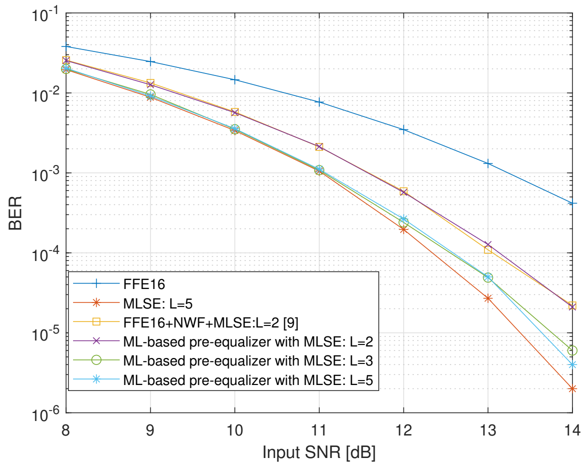

Fig. 3: The simulated BER versus the input SNR for 5-tap FIR channel with AWGN.

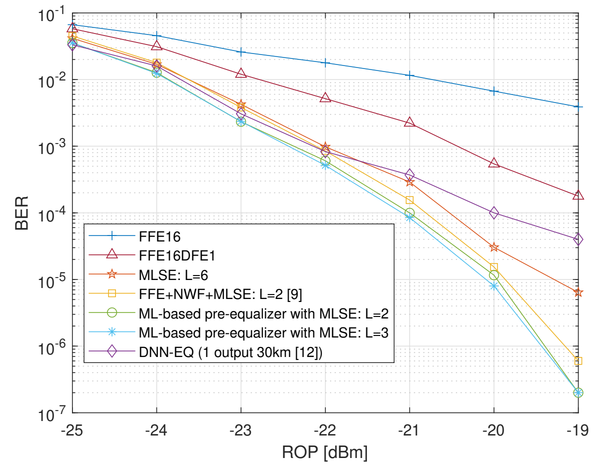

Fig. 4: The simulated BER versus the receive optical power for the 50Gb/s PON measurement traces.

photo diode integrated with transimpedance amplifier, and is captured with a real-time sampling scope. The captured data is then pre-processed offline by a baud-rate Mueller-Muller CDR and then used for training and BER evaluation.

The evaluated BER versus the received optical power (ROP) is shown in Fig. 4. Notice that here the proposed MLSE with $L = 2$ already operates close to $L = 3$ taps, due to the smaller amount of ISI in this measured trace than in the FIR channel example. The standalone MLSE with high number taps is suffering from performance loss due to the presence of colored noise and distortion in this scenario, and therefore cannot be used as performance upper bound. Further, the proposed ML-based pre-equalizer with MLSE outperforms the state-of-the-art solution in [9] consisting of a 2-tap MLSE preceeded by a 16-tap FFE and 2-tap noise whitening filter (NWF). The BER numbers obtained from [12] are shown as well, corresponding to a DNN equalizer (the version with 1 output symbol in particular) that is trained and evaluated on the same measurements, except that a different resampling script has been used to obtain baud-rate receive samples.

V. Conclusion

We specifically consider MLSE for ISI equalization where the number of Viterbi taps is smaller than the ISI channel length (i.e., $L < M$). This keeps the implementation complexity of Viterbi algorithms limited. To improve in such scenarios the heuristic prior-art approaches, we propose a novel ML-based training method for Viterbi algorithms in combination with a pre-equalizer. Key is that the pre-equalizer is trained using a learning loss function based on the Viterbi extended PMs, such that the end-to-end MLSE performance is directly optimized. We have demonstrated via numerical evaluations that our proposed solution for Viterbi equalizers with a small number of taps achieves significant performance improvements w.r.t. prior art approaches.

References

[1] L. A. Neto, J. Maes, P. Larsson-Edefors, J. Nakagawa, K. Onohara and S. J. Trowbridge, "Considerations on the Use of Digital Signal Processing in Future Optical Access Networks," in *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 598-607, 1 Feb.1, 2020.

[2] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, Chapter 3, Cambridge, 2005.

[3] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". *IEEE Trans. Inform. Theory*, vol. 13, no. 2, Apr. 1967, pp. 260–269.

[4] S. Qureshi and E. Newhall, "An Adaptive Receiver for Data Transmission over Time-Dispersive Channels", in *IEEE Trans. on Inf. Theory*, vol. 19, pp. 448–457, Jul. 1973

[5] D. D. Falconer, and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation." *The Bell System Technical Journal*, vol. 52, no. 9, pp. 1541–1562, Nov., 1973.

[6] N. Al-Dhahir and J. M. Cioffi,"Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: a unified approach", in *IEEE Trans. Inf. Theory*, vol. 42, pp. 903–915, Apr. 1996.

[7] C. Beare, "The Choice of the Desired Impulse Response in Combined Linear–Viterbi Algorithm Equalizers", in *IEEE Trans. on Commun.*,vol. 26, no. 8, pp. 1301–1307, Aug. 1978

[8] J. Li, E. Tipsuwannakul, T. Eriksson, M. Karlsson and P. A. Andrekson, "Approaching Nyquist Limit in WDM Systems by Low-Complexity Receiver-Side Duobinary Shaping," in J*ournal of Lightwave Technology*, vol. 30, no. 11, pp. 1664-1676, June1, 2012.

[9] YBorui Li, Kuo Zhang, Dechao Zhang, Jiale He, Xiaolong Dong, Qian Liu, and Shengping Li, "DSP enabled next generation 50G TDM-PON," *J. Opt. Commun. Netw.*, Dec. 2020.

[10] J. Estaran et al., "Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IM/DD systems," in *Proc. Eur. Conf. Opt. Commun.* (ECOC), Düsseldorf, Germany, 2016

[11] M. Chagnon, J. Siirtola, T. Rissa, A. Verma, "Quantized deep neural network empowering an IM-DD link running in real-time on a field programmable gate array," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Dublin, Ireland, 2019.

[12] N. Kaneda et al., "FPGA Implementation of Deep Neural Network Based Equalizers for High-Speed PON," *Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1-3.

[13] N. Shlezinger, Y. C. Eldar, N. Farsad, and A. J. Goldsmith, "Viterbinet: Symbol detection using a deep learning based viterbi algorithm," *IEEE Int Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, pp. 1–5, 2019.

[14] Proakis, *Digital Communications 5th Edition*, McGraw Hill, 2007.

[15] D. van Veen and V. Houtsma, "Transceiver Technologies for Next-Generation PON (Tutorial)," in *Optical Fiber Communications Conference and Exhibition* (OFC), San Diego, CA, USA, 2020, pp. 1-28.