Design of Time-Domain Learned Volterra Equalisers for WDM Systems

Nelson Castro
AiPT, Aston University
Birmingham B4 7ET, UK
cast1901@aston.ac.uk

Sonia Boscolo
AiPT, Aston University
Birmingham B4 7ET, UK
s.a.boscolo@aston.ac.uk

Andrew D. Ellis
AiPT, Aston University
Birmingham B4 7ET, UK
andrew.ellis@aston.ac.uk

Stylianos Sygletos
AiPT, Aston University
Birmingham B4 7ET, UK
s.sygletos@aston.ac.uk

***Abstract*—We examine various design aspects of a learned time-domain multiple-input multiple-output (MIMO) Volterra-based equaliser and reveal their impact on the convergence and performance of the model. We show that appropriate parameter initialisation is vital for the model's convergence and scalability to a higher number of channels. This design optimisation enables the first demonstration of a 7×7 operation of the MIMO algorithm at one step per span, achieving 1.5 dB effective signal-to-noise ratio improvement over single-channel nonlinear equalisation, hence significantly enhancing the transmission performance in a wavelength-division multiplexing scenario.**

***Index Terms*—nonlinearity equalisation, machine learning, Volterra series, optical fibre systems**

I. Introduction

Digital equalisation of Kerr-effect induced nonlinear impairments is a promising approach to enhance the capacity of optical fibre links [1]. In practical wavelength division multiplexing (WDM) systems, inter-channel nonlinearities are the dominant degradation factors of transmission performance. Effective nonlinearity compensation requires sophisticated digital signal processing (DSP) developed from an in-depth knowledge of the physical layer model and capable of simultaneously equalising multiple co-propagating channels. Initial approaches to multichannel equalisation were based primarily on digital backpropagation (DBP) techniques. In particular, broadband full-field DBP [2], an approach that processes WDM channels as a unified field, demonstrated remarkable efficacy in counteracting four-wave mixing (FWM) and cross-phase modulation (XPM) impairments. Despite its potential, the practical deployment of this approach encountered significant limitations, notably, the necessity for a large number of processing steps and high digital oversampling rates [3]. Conversely, multiple-input multiple-output (MIMO) DBP approaches, which were designed to solve coupled nonlinear Schrödinger equations and targeted the mitigation of XPM impairments only, featured less computational complexity at the cost of reduced performance improvements. Subsequent advancements made MIMO DBP more efficient. For example, integrating frequency-domain filtered nonlinear steps reduced the number of steps needed for multichannel operation significantly [4]. Nevertheless, such approaches are still impractical for real-world applications due to their computational cost.

This work was partly supported by the UK EPSRC grants TRANSNET (EP/R035342/1), (EP/X019241/1), and EEMC (EP/S016171/1).

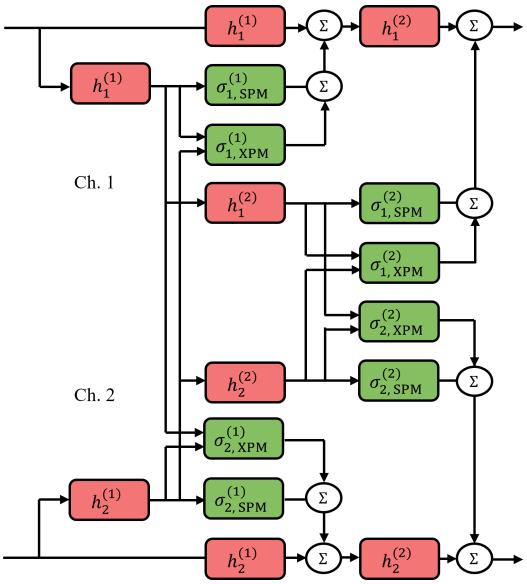

Fig. 1. Block diagram of a 2×2 time-domain MIMO L-simIVSTF for two computational steps.

The integration of machine learning (ML) into nonlinear equalisation (NLE) has completely revolutionised the approach to improving system performance. Substantial research has focused on applying ML to refine traditional equalisation techniques. Using ML in single-channel DBP has reduced complexity and enhanced equalisation capability [5]. ML-based optimisation techniques have also advanced the operational efficiency of MIMO-DBP so as to require only one step per span, for both frequency- and time-domain implementations [6], [7]. Furthermore, ML has enabled the efficient operation of inverse Volterra series transfer function (IVSTF) based schemes in MIMO configurations. In our prior work [8], we introduced a new time-domain learned (L) MIMO model and demonstrated its multichannel operation at two steps per span. This model built on our earlier simplified IVSTF (simIVSTF) architecture [9], which achieved significant complexity reduction through efficient finite impulse response (FIR) filter re-use in its linear steps and incorporated enhanced filtering in its nonlinear steps [10]. Although ML had a crucial role in the optimisation of L-Volterra-based MIMO equalisation, key design elements

remain still unexplored.

This paper delves into various design aspects of our time-domain MIMO L-simIVSTF model [8]. We examine how different MIMO configurations affect performance and identify dimensions that drive an optimum balance between computational cost and performance. This design optimisation enables the first demonstration of 7×7 L-Volterra-based MIMO equalisation at only one step per span, featuring significant performance improvement compared to single-channel operation in a WDM transmission scenario.

II. Principle, Simulation Setup and Results

The architecture of our equaliser is depicted in Fig. 1 for a 2×2 MIMO implementation with two computational steps. Each linear step s was implemented similarly to [11], by convoluting the complex signal $y_n^{(s-1)}(t)$ of the nth channel with the impulse response $h_n^{(s)}(t)$ of the FIR filter compensating the chromatic dispersion (CD) of the corresponding fibre length. Before each linear step, an appropriate delay element for each channel addressed the walk-off effect. The self-phase modulation (SPM) and XPM activation functions at the sth stage were expressed as

$$\begin{aligned}\sigma_{n,\mathrm{SPM}}^{(s)}(t) &= -i\gamma L_{\mathrm{eff}} y_n^{(s)}(t) \sum_{c=-l}^{l} \alpha_c^{(s)} |y_n^{(s)}(t+cT_s)|^2, \\ \sigma_{n,\mathrm{XPM}}^{(s)}(t) &= -2i\gamma L_{\mathrm{eff}} y_n^{(s)}(t) \sum_{k\neq n}^{N_{\mathrm{ch}}} \sum_{c=-m}^{m} \beta_{c,k}^{(s)} |y_k^{(s)}(t+cT_s)|^2,\end{aligned} \tag{1}$$

where γ is the nonlinear coefficient of the fibre, L_{eff} is the effective fibre length, $\alpha_c^{(s)}$ and $\beta_{c,k}^{(s)}$ are the filter's coefficients, T_s is the sampling interval, and N_{ch} is the dimension of the MIMO algorithm. The filtering operations were implemented using the one-dimensional convolutional layers available in popular ML frameworks but suitably adapted to perform circular convolution on mini-batches of the received signal by fulfilling the boundary conditions.

To assess the performance of our equaliser, we considered an 11-channel single-polarisation WDM transmission over a 6×100-km standard single-mode fibre link (dispersion parameter $D = 17\,\mathrm{ps/(nm \cdot km)}$, $\gamma = 1.3\,(\mathrm{W \cdot km})^{-1}$, loss coefficient $\alpha = 0.2\,\mathrm{dB/km}$). Each channel carried a stream of root-raised cosine pulses of 0.1 roll-off, modulated by 64 quadrature-amplitude modulation symbols at a rate of 32 Gbaud. The channel spacing was 40 GHz. Data transmission was simulated using the split-step Fourier method in batches of 2^{18} symbols, with an up-sampling factor of 32. At the receiver, the WDM channels were brought to baseband and down-sampled to 2 samples per symbol. The channels of interest were then passed to the MIMO NLE for processing. The receiver's DSP blocks were implemented as a differentiable computation graph in TensorFlow. After the NLE stage, each channel underwent matched filtering and further down-sampling to 1 sample per symbol. During the training phase, the outputs of the MIMO algorithm were linked to a single mean-squared-error (MSE) function for computing the gradients of the model's trainable parameters,

$$L_{\mathrm{MSE}} = \frac{1}{N_{ch} \cdot M} \sum_{n=1}^{N_{ch}} \sum_{c=1}^{M} \left| x_n^{(c)} - \hat{x}_n^{(c)} \right|^2 \tag{2}$$

where $\hat{x}_n^{(c)}$ and $x_n^{(c)}$ are the reference and recovered symbols, respectively, and M is the total number of symbols within a batch. During the testing phase, the recovered symbols from each channel were used to compute the bit error rate, which was then mapped to an effective signal-to-noise ratio (SNR). Datasets for a given launch power included 2^{19} symbols for training and 2^{18} symbols for validation and testing. The model was trained using the Adam optimiser, with a learning rate of 0.001 and a batch size of 40. These parameters were manually adjusted to ensure convergence stability.

Our model's trainable parameters included the coefficients $h_n^{(s)}$ of the CD FIR filters and the coefficients $\alpha_c^{(s)}$, $\beta_{c,k}^{(s)}$ of the SPM and XPM FIR filters, respectively. These were treated as independent parameters across each equalisation layer and channel path. The CD filters were initialised as in [12]. The initialisation of the SPM and XPM filters was a subject of study. We assumed zero-valued initial conditions for all their taps but the central ones, α_0 and $\beta_{0,k}$, which were set to ξ_{SPM} and ξ_{XPM}, respectively, and optimised. Separate training was performed for each MIMO implementation (3×3, 5×5, and 7×7) and launch power level. We also optimised the model's hyper-parameters to maximise performance and reduce complexity [8]. The hyper-parameter set included the lengths of the FIR filters of the linear and nonlinear steps, assumed to be the same throughout the model. Following the same procedure as in [8], we identified the optimum lengths to be 43 taps, 7 taps, and 41 taps for the CD, SPM, and XPM filters, respectively.

The study on the initialisation factors ξ_{SPM} and ξ_{XPM} highlighted their critical role in the model's convergence. The various MIMO configurations were trained for up to 1500 epochs, beyond which no performance improvement was observed. The validation curves from the training process of the 5×5, one-step-per-span implementation are depicted in panels (a) and (b) of Fig. 2. These curves represent the evolution of the MSE over the number of training epochs for the validation dataset. We can observe from Fig. 2(a) that the initialisation factor for the SPM filters has minimal influence on the convergence speed. By contrast, suitable initialisation of the XPM filters leads to a significantly faster convergence (Fig. 2(b)). This is attributed to the fact that XPM is the dominant effect responsible for transmission performance degradation. Repeating the same exercise for two-, three-, and four-steps-per-span configurations showed similar results. It is worth noting here that optimising the initial conditions for the SPM and XPM filters facilitated the convergence of the model, thereby enabling the implementation of 5×5 and 7×7 MIMO configurations at only one step per span, which had not been achieved in [8]. Figure 2(c) shows the effective SNR performance across the channels for different MIMO sizes.

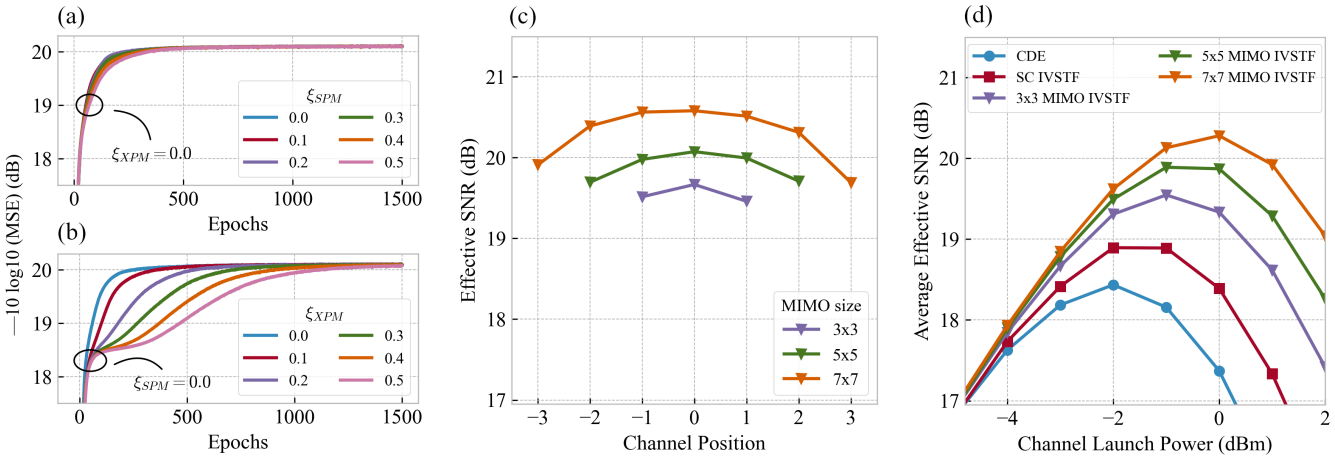

Fig. 2. MSE evolution over number of training epochs for varying (a) ξ_{SPM} and (b) ξ_{XPM} values. (c) SNR performance of individual channels for 3×3 (violet), 5×5 (green) and 7×7 (orange) MIMO schemes operating at one step per span. (d) Average SNR versus channel launch power for linear equalisation (blue), and single-channel (red), 3×3 (violet), 5×5 (green) and 7×7 (orange) MIMO models operating at one step per span.

We can see that up to 4 channels can be equalised with a performance difference of less than 0.5 dB relative to the central channel. The outermost channels in the 7×7 MIMO implementation feature a performance degradation of ~ 0.7 dB compared to the central channel. For all implementations, the edge channels show the lowest performance because they were affected by the nonlinearity of the channels outside the equalisation band.

Furthermore, we assessed the average effective SNR performance as a function of the channel launch power for different MIMO sizes and operation at one step per span. The results are summarised in Fig. 2(d). We can see that single-channel equalisation brings about a peak-to-peak SNR improvement of only 0.5 dB over linear equalisation, which is not sufficient for practical use. Increasing the number of equalised channels with a 3×3 MIMO model contributes a further ~ 0.6 dB improvement. Similarly, the other two MIMO sizes bring about a further ~ 0.3 dB improvement each. Contrasting with the recurrent neural network-based multichannel equaliser in [13], which offers negligible performance benefit when increasing its size from 3×3 to 5×5, our scheme scales to higher MIMO sizes with improved performance.

III. Conclusions

We have examined several design aspects of a learned time-domain MIMO Volterra-based equaliser and their impact on the model's performance and convergence. Our results have shown that suitable initialisation of the MIMO model's parameters is crucial to its convergence and scalability. We have demonstrated for the first time a 7×7 MIMO L-simIVSTF equaliser at only one step per span, achieving 1.8 dB effective SNR improvement over linear equalisation, and 1.5 dB over the nonlinear equalisation provided by its single-channel counterpart for the same number of steps.

References

[1] A. Ellis, A. Ali, M. Tan, N. Salgado, and S. Sygletos, "Mitigation of nonlinear effects in optical communications using digital and optical techniques," in *Optica Advanced Photonics Congress 2022*. Optica Publishing Group, 2022, p. NeTu3D.4.

[2] R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Spectrally Shaped DP-16QAM Super-Channel Transmission with Multi-Channel Digital Back-Propagation," *Sci Rep*, vol. 5, no. 1, p. 8214, Jul. 2015.

[3] G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," *Opt. Express*, vol. 22, no. 24, pp. 30053–30062, Dec 2014.

[4] E. F. Mateo, F. Yaman, and G. Li, "Efficient compensation of inter-channel nonlinear effects via digital backward propagation in wdm optical transmission," *Opt. Express*, vol. 18, no. 14, pp. 15144–15154, Jul 2010.

[5] C. Häger and H. D. Pfister, "Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications," in *2018 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2018, pp. 1590–1594, iSSN: 2157-8117.

[6] T. Inoue, R. Matsumoto, and S. Namiki, "Learning-based digital back propagation to compensate for fiber nonlinearity considering self-phase and cross-phase modulation for wavelength-division multiplexed systems," *Opt. Express*, vol. 30, no. 9, pp. 14851–14872, Apr 2022.

[7] O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul wdm transmission systems," *J. Lightwave Technol.*, vol. 39, no. 8, pp. 2397–2406, Apr 2021.

[8] N. Castro and S. Sygletos, "Learned volterra equalization for wdm systems," in *2023 Asia Communications and Photonics Conference/2023 International Photonics and Optoelectronics Meetings (ACP/POEM)*, 2023, pp. 1–4.

[9] N. Castro and S. Sygletos, "A novel learned volterra-based scheme for time-domain nonlinear equalization," in *Conference on Lasers and Electro-Optics*. Optica Publishing Group, 2022, p. SF3M.1.

[10] M. Secondini, D. Marsella, and E. Forestieri, "Enhanced split-step fourier method for digital backpropagation," in *2014 The European Conference on Optical Communication (ECOC)*, 2014, pp. 1–3.

[11] C. Fougstedt, M. Mazur, L. Svensson, H. Eliasson, M. Karlsson, and P. Larsson-Edefors, "Time-domain digital back propagation: Algorithm and finite-precision implementation aspects," in *Optical Fiber Communication Conference*. Optica Publishing Group, 2017, p. W1G.4.

[12] A. Sheikh, C. Fougstedt, A. G. i. Amat, P. Johannisson, P. Larsson-Edefors, and M. Karlsson, "Dispersion compensation fir filter with improved robustness to coefficient quantization errors," *Journal of Lightwave Technology*, vol. 34, no. 22, pp. 5110–5117, 2016.

[13] S. Deligiannidis, K. R. H. Bottrill, K. Sozos, C. Mesaritakis, P. Petropoulos, and A. Bogris, "Multichannel nonlinear equalization in coherent wdm systems based on bi-directional recurrent neural networks," *Journal of Lightwave Technology*, vol. 42, no. 2, pp. 541–549, 2024.