Blind State of Polarisation Monitoring Using Variational AutoEncoders-inspired Adaptive Filter

Louis Tomczyk, Élie Awwad, Diane Prato, Cédric Ware

LTCI, Télécom Paris, Institut Polytechnique de Paris, France louis.tomczyk@telecom-paris.fr

Abstract *We demonstrate that probabilistically shaped modulations make the Constant Modulus Algorithm unfit for State of Polarisation monitoring in coherent optical communications. Accordingly, we study the potential of Variational AutoEncoders-inspired equalisers for this purpose. ©2024 The Authors*

Introduction

The monitoring of physical parameters in coherent optical networks have seen a growing interest over the last decade. Among the other physical parameters to monitor, the State of Polarisation (SoP) has been investigated for the last 20 years[1]–[5]. To the best of our knowledge, the SoP is currently extracted from the adaptive filter coefficients of the polarisation demultiplexing step of the Digital Signal Processing (DSP) chain[6]. However, the conventional Constant Modulus Algorithm (CMA) is known to face convergence issues when transmitting Probabilistic Constellation Shaped (PCS) modulations[7]. The latter is being used more and more due to its ability to approach Shannon's capacity[8],[9]. In such scenarios, SoP monitoring becomes a sensitive question. While data-aided channel estimation can solve these issues, it reduces the net data rate. Hence, the quest for new blind adaptive filters continued. A new kind of adaptive blind filters inspired by Variational AutoEncoders (VAEs) has been proposed by[10] and extended for PCS by[7]. We will refer to it as the VAE-Finite Impulse Response (FIR) filter. In those works, the new loss (function that has to be minimised by the equaliser), promises to be convergent for PCS-Quadrature Amplitude Modulation (QAM), allowing to decode the symbols as well as providing SoP estimations and other optical channel parameters. In[7], a preliminary study of SoP tracking using a linear SoP drift model and uniform dual-polarisation (DP) QAM was done. In this work, we investigate through simulations the capability of the VAE-FIR for SoP monitoring in scenarios that have not been explored yet. We use two different SoP drift models. The first one is based on a "polarisation linewidth" model suggested by authors in[3] at typical SoP change rates measured in field trials from dozens to dozens of thousand hertz[1]–[5]. The second one is a linear drift similar to the one in[7]. However, our study considers PCS DP-QAM and shows estimation errors of SoP.

We first shortly recall the fundamentals of the VAE. Then, we describe the simulation setup and the studied scenarios. Finally, we compare results between the CMA and the VAE-FIR showing the accuracy in SoP estimation and resilience to the latter of the VAE-FIR while the CMA fails to converge for PCS transmission.

Loss Function of Variational AutoEncoders for Communications

We model optical propagation as successive changes of SoP, quantified by angle shifts used in the corresponding Jones rotation matrices, noted $\Theta_h = \{\theta_k\}_{k\in\mathbb{N}}$ with Polarisation Mode Dispersion (PMD). Symbols at the channel input ($\boldsymbol{X}$) become samples detected at the receiver ($\boldsymbol{Y}$) then symbols at receiver output after linear equalisation and decision ($\widehat{\boldsymbol{X}}$). We note the similarity with the structure of a VAE, which is a couple of two successive neural networks: the first ("encoder") transforms random variables from its input space to a so-called latent space; and the second ("decoder") converts from latent space to output space. While we are not using a VAE per se, we can use the VAE's loss function to update the weights of the linear equaliser instead of using the losses traditionally based on the moduli of the signals.

The appropriate loss, $\mathcal{L}$, is then based on the Kullback-Leibler Divergence (D_{KL}) and a log-likelihood[11] parameterised by a demapper, with a probability density function denoted as $p_{\widehat{\boldsymbol{X}}|\boldsymbol{Y}}$. The loss is written as[7]:

$$\mathcal{L} = D_{KL}\left[p_{\widehat{\boldsymbol{X}}|\boldsymbol{Y}}||p_{\boldsymbol{X}}\right] - \mathbb{E}_{p_{\widehat{\boldsymbol{X}}|\boldsymbol{Y}}}\left[\ln(p_{\boldsymbol{Y}|\boldsymbol{X}})\right] \quad (1)$$

where $\mathbb{E}(\bullet)$ (resp. $p_\bullet$) is the expectation (resp. the probability density function) of the random variable $\bullet$ and $\ln(\bullet)$ is the natural logarithm. The divergence part of the loss aims to obtain the best estimation of the symbols, while the likelihood gives more weight to the estimation of the channel. For further details, we refer interested readers to references[7],[10].

Simulation setup

We simulate a single-wavelength dual-polarisation transmission over a linear optical channel with only random polarisation state changes and $10\ \mathrm{ps}$ PMD.

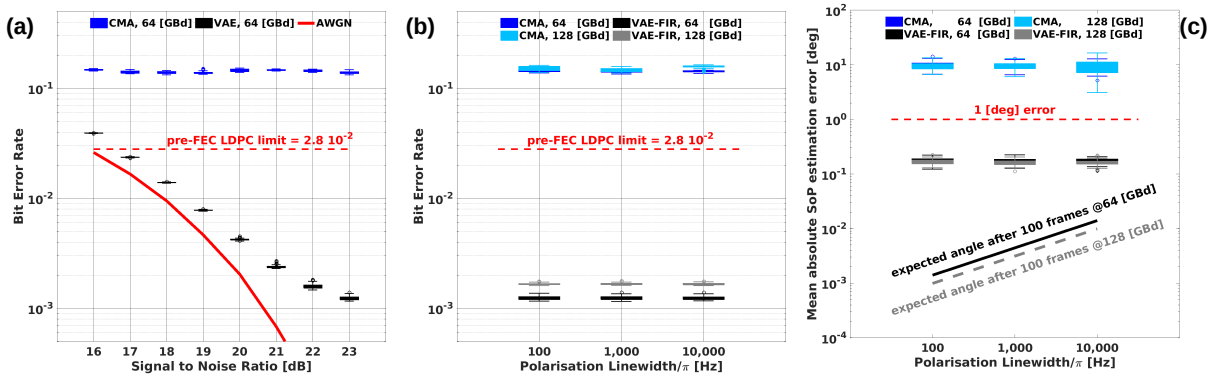

Fig. 1: (a) Evolution of the BER as a function of SNR for a polarisation linewidth $\mathrm{f_{pol}} = 100\pi$ Hz. (b) Evolution of the BER as a function of $\mathrm{f_{pol}}$ at a $\mathrm{SNR} = 23$ dB. (c) Evolution of the mean absolute error SoP estimation for the "polarisation linewidth" model after 100 frames. The whisker lines give the the minimum and maximum values of the distribution. The points are the outliers.

Additive White Circular Normal ($\mathcal{CN}$)/Gaussian Noise is added at both ends of the communication chain and as in[7], no laser phase noise is considered. At the transmitter side, we set, once for all, a total noise power at -50 dBm accounting for multiple noise sources (quantisation, amplifiers, etc.). The total signal power is also fixed at -5 dBm. The noise power at the receiver side is varied to target different Signal-to-Noise Ratios (SNRs). Consequently, $\boldsymbol{Y}|\boldsymbol{X}$ follows $\mathcal{CN}(\boldsymbol{h} \circledast \boldsymbol{x},\ \sigma^2 \cdot \boldsymbol{I})$ where $\circledast$ is a matrix convolution operator, $\boldsymbol{h}$ (resp. $\boldsymbol{I}$) is the channel (resp. identity) matrix.

We consider a Gray-mapped Maxwell-Boltzmann PCS-64QAM constellation with an entropy of $\mathcal{H} = 5.72$ bits corresponding to the case where authors in[7] showed the convergence of the CMA for a 3.16 ps PMD. Symbols are shaped with a Root-Raised Cosine (RRC) filter of $\beta_{\mathrm{roll}} = 0.1$ roll-off and up-sampled at 2 samples per symbol.

As in[7], we partition symbols into "frames", themselves sliced into "batches". Over each frame, the SoP is assumed to be constant. Each frame (resp. batch) contains $\mathrm{N_{Sb/f}}$ (resp. $\mathrm{N_{Sb/b}}$) symbols. We simulate transmissions using two symbol rates: $\{64; 128\}$ GBd. To enable the VAE-FIR convergence for these two symbol rates, we need to adjust the frame and batch sizes to: $(\mathrm{N_{Sb/f}};\ \mathrm{N_{Sb/b}}) = (12,800;\ 160)$ symbols for 64 GBd and to $(17,040;\ 213)$ symbols for 128 GBd. Similarly, it requires respectively $\{20;\ 40\text{or } 50\}$ frames to converge for these two rates. Both equalisers are initialised by a Dirac.

Unless otherwise specified, the learning rate, the number of filter taps and the number of non training frames are set constant for all the scenarios and for both equalisers to respectively $(\eta_{\mathrm{lr}};\ \mathrm{N_{taps}};\ \mathrm{N_f}) = (5 \cdot 10^{-4};\ 13;\ 100)$. All parameters of equalisers were set to: (1) account for the memory effect of the channel induced by PMD, (2) ensure an error rate with a confidence level $> 99\%$ and (3) meet a trade-off between computational time and performance for the VAE-FIR. Consequently, we will see that those parameters may not be the most suited for CMA or even too stringent for it to converge. We refer to[7] for an extensive study of the impact due to each parameter on the performance (number of frames, learning rate, etc.). We set a pre-Forward Error Correction (FEC) Bit Error Rate (BER) threshold up to which the errors can be corrected using Low-Density Parity-Check (LDPC) codes at $2.8 \cdot 10^{-2}$ as authors in[12] showed to be enough to guarantee error-free communication (post-FEC BER $< 10^{-15}$).

Let us recall that the objective is to compare the performance of both CMA and VAE-FIR over a time-varying channel. To do so, we use the following metrics: the BER, and the error in the estimation of the SoP drifts. Two main directions have been explored. In the first one, we study the penalty under SoP drift models that we detail in the next paragraph. We targeted $\mathrm{N_r} = 20$ realisations per SNR value in order to reduce estimation error. However, as both algorithms can sometimes fail to converge and because of the large computation time, we choose to rely only on the realisations that succeeded, which enable to estimate a failure rate for both algorithms: $> 25\%$ (resp. $< 5\%$) for the CMA (resp. VAE-FIR).

The second direction is the robustness to SoP changes and the ability to track them. We simulate two different models of SoP variations. One is a Wiener process with a "polarisation linewidth" $\mathrm{f_{pol}}$ as suggested in[3]. The variance of the process is given by $\sigma^2_{\mathrm{pol}} = 2\pi \cdot \mathrm{f_{pol}}/\mathrm{R_s}$ with $\mathrm{f_{pol}} \in \pi \cdot \{10^2;\ 10^3;\ 10^4\}$ Hz. The second model is linear such as: $\theta_k = \theta_{\mathrm{start}} + \mathrm{S} \cdot k$, $k \in [\![0, \mathrm{N_{f,max}}]\!]$. The variations include different ranges that contain the tricky rotation angle: 45 deg. Indeed, this corresponds to an equal power distribution in both received polarisation states leading to higher probability of failure for the CMA. The values for S and θ_{start} are respectively $\{0.5; 1\}$ deg/frame, $\{5;\ 10;\ 11;\ 12;\ 13;\ 14;\ 15;\ 20\}$ deg. One should note that θ_{start} is an angle offset accounting for a sudden SoP drift. At 64 GBd, we increased the learning rate to $1.1 \cdot 10^{-3}$ and at 128 GBd we used

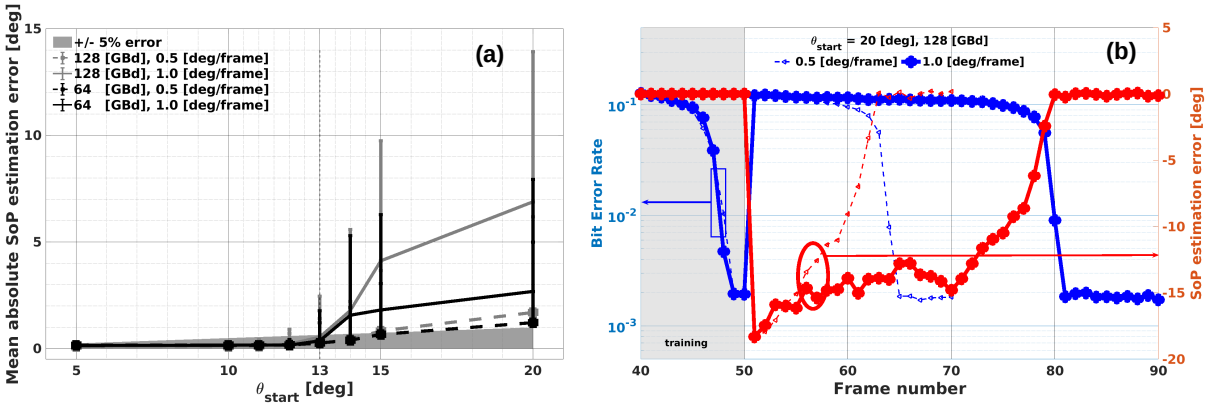

Fig. 2: Evolution of the mean absolute error of SoP estimations when the SoP drift is linear, at $\mathrm{SNR} = 23$ dB with 3 realisations, (a) for different values of θ_{start}; (b) for $\theta_{\mathrm{start}} = 20$ deg along with time. Large error bars are due to the transient regime.

50 training frames.

Results

Figure 1(a) (resp. 1(b)) shows the evolution of the BER versus SNR with a polarisation linewidth $\mathrm{f_{pol}} = 100\pi$ Hz (resp. the BER with respect to $\mathrm{f_{pol}}$ at $\mathrm{SNR} = 23$ dB). We notice that, contrary to[7], the CMA fails at converging as the SER values stay way beyond the pre-FEC limit for all the possible $(\mathrm{SNR}, \mathrm{f_{pol}})$ combinations. In[7], the authors simulated a linear evolution for the SoP drift and used twice as many filter taps along with longer batches. This leads to a better performance of the CMA at the cost of an increased complexity. On the contrary, the VAE-FIR stays below the threshold for all tested polarisation linewidth values as long as $\mathrm{SNR} > 17$ dB.

In figure 1(c), we show the errors of SoP drift estimations with both equalisers. We also show the expectation of the absolute value of SoP changes after 100 frames that are given by[13]: $\mathbb{E}(|\theta_{\mathrm{N_{f,max}}} - \theta_{\mathrm{start}}|) = 2 \cdot \sqrt{\mathrm{N_{f,max}} \cdot \mathrm{f_{pol}}/\mathrm{R_s}}$. We notice that tens of thousands of kilo-Hertz for the polarisation linewidth gives SoP drifts below $5 \cdot 10^{-2}$ deg. This may question the relevance of such a model to test the tracking of the polarisation. Authors in[1] measured a maximum of 25 deg shift after 1 min for an aerial cable, which for our setup and this model, requires $\mathrm{f_{pol}} \approx 40$ mHz at both rates. Besides, we clearly see that the CMA provides estimation errors of tens of degrees which clearly shows its difficulty to track such low-speed SoP changes. A similar comment can be made for the VAE-FIR, even though the mean estimation errors are always below 1 deg. Thus, we conclude that, for the tested set of parameters, the VAE-FIR does not provide accurate enough estimations for small values of SoP changes, even though it still out-performs the CMA by approximately 100 times. Our second SoP drift test will show its ability to track larger SoP variations. For this model, we did not convey simulations with the CMA as we have seen that the CMA fails even with $\theta_{\mathrm{start}} = 0$ deg.

Figure 2(a) shows the mean estimation errors and their 1σ-error bars (computed for all the frames at once) for both baud rates and both slopes. We superimposed a shaded area corresponding to an error of $\pm 5\%$. We notice that we have two regimes delimited by $\theta_{\mathrm{start}} \approx 13$ deg. Before this value, in all the scenarios, the VAE-FIR provides good enough SoP estimations. Beyond, it is not the case anymore. We conclude that while the SoP changes are small enough, the errors on their estimation is still acceptable. Finally, we explain the significant error bars for higher sudden SoP offsets and specifically for 20 deg. We show, in figure 2(b), the evolution of both BER (left axis) and the error of estimation of θ_k (right axis) at the end of the training phase (frame 50) and in the beginning of the tracking phase for different slopes. We display the worst case which is at 128 GBd. In the training phase, we clearly see the convergence of the VAE-FIR with the waterfall-like BER evolution. Yet, in the second phase, right at the sudden SoP change, we notice the drop of the performance. However, we see that for both slopes, the VAE-FIR can catch-up with the error of estimations after a transient regime whose duration is shorter for a smaller slope after a sudden large SoP change occurs. This last result shows the resilience of the VAE-FIR.

Conclusion

We investigated the ability of the VAE-FIR to track the evolution of the SoP in DP transmission over a time-varying channel, for scenarios triggering failures of the CMA. We observed that the VAE-FIR out-performs the CMA by almost 100 times when using PCS-64QAM with an entropy $\mathcal{H} = 5.72$ bits for slow SoP changes, using the "polarisation linewidth" model of[3]. We have also showed the ability of the VAE-FIR to track SoP changes even when crossing the 45 deg angle corresponding to the maximum polarisation cross-talk scenario, at different approaching speeds within a 1 deg-accuracy. Finally, we demonstrated the ability of the VAE to catch up with the SoP changes after a steep drift.

Acknowledgements

This work has received funding from BPI France (grant 0168463/00) within the CELTIC-NEXT European project AI-NET Antillas.

References

[1] K. Ogaki, M. Nakada, Y. Nagao, and K. Nishijima, "Fluctuation differences in the principal states of polarization in aerial and buried cables", in *OFC 2003 Optical Fiber Communications Conference, 2003.*, 2003, 14–15 vol.1. DOI: 10.1109/OFC.2003.1247454.

[2] P. Krummrich and K. Kotten, "Extremely fast (microsecond timescale) polarization changes in high speed long haul wdm transmission systems", in *Optical Fiber Communication Conference, 2004. OFC 2004*, vol. 2, 2004, 3 pp. vol.2-.

[3] C. B. Czegledi, M. Karlsson, E. Agrell, and J. Pontus, "Polarization drift channel model for coherent fibre-optic systems", *Sci Rep*, vol. 6, no. 21217, 2016. DOI: 10.1038/srep21217.

[4] J. E. Simsarian and P. J. Winzer, "Shake before break: Per-span fiber sensing with in-line polarization monitoring", in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.

[5] M. Mazur, L. Dallachiesa, R. Ryf, *et al.*, "Real-time monitoring of cable break in a live fiber network using a coherent transceiver prototype", in *ArXivs*, 3 July 2023.

[6] J. C. Geyer, F. N. Hauske, C. R. S. Fludger, *et al.*, "Channel parameter estimation for polarization diverse coherent receivers", *IEEE Photonics Technology Letters*, vol. 20, no. 10, pp. 776–778, 2008. DOI: 10.1109/LPT.2008.921104.

[7] V. Lauinger, F. Buchali, and L. Schmalen, "Blind equalization and channel estimation in coherent optical communications using variational autoencoders", *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2529–2539, 2022. DOI: 10.1109/JSAC.2022.3191346.

[8] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications", *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019. DOI: 10.1109/JLT.2019.2898855.

[9] H. Sun, M. Torbatian, M. Karimi, *et al.*, "800g dsp asic design using probabilistic shaping and digital sub-carrier multiplexing", *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4744–4756, 2020. DOI: 10.1109/JLT.2020.2996188.

[10] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders", in *Proceedings of Internationnal Conference on Communications*, 2018. DOI: 10.1109/ICCW.2018.8403666.

[11] D. Kingma and W. Max, *An introduction to variational autoencoders (foundations and trends in machine learning)*. page 18, eq. 2.11: Now publishers, 2019, ISBN: 978-1680836226.

[12] L. Zhang, K. Tao, W. Qian, *et al.*, "Real-time fpga investigation of interplay between probabilistic shaping and forward error correction", *Journal of Lightwave Technology*, vol. 40, no. 5, pp. 1339–1345, 2022. DOI: 10.1109/JLT.2021.3128490.

[13] M. Tsagris, C. Beneki, and H. Hassani, "On the folded normal distribution", *Mathematics*, vol. 2, no. 1, pp. 12–28, 2014, ISSN: 2227-7390. DOI: 10.3390/math2010012. [Online]. Available: https://www.mdpi.com/2227-7390/2/1/12.