Nonlinear Phase Noise Compensation in Experimental WDM Systems with 256QAM

Yankov, Metodi Plamenov; Da Ros, Francesco; Porto da Silva, Edson; Fehenberger, Tobias; Barletta, Luca; Zibar, Darko; Oxenløwe, Leif Katsuo; Galili, Michael; Forchhammer, Søren

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/JLT.2016.2644860

Publication date:
2017

Document Version
Peer reviewed version

Citation (APA):
Nonlinear Phase Noise Compensation in Experimental WDM Systems with 256QAM

Metodi P. Yankov\textsuperscript{1}, Member, IEEE, Francesco Da Ros\textsuperscript{1} Member, IEEE, OSA, Edson P. da Silva\textsuperscript{1} Student Member, IEEE, OSA, Tobias Fehenberger\textsuperscript{2} Student Member, IEEE, Luca Barletta\textsuperscript{3} Member, IEEE, Darko Zibar\textsuperscript{1} Member, IEEE, Leif K. Oxenløwe\textsuperscript{1} Member, OSA, Michael Galili\textsuperscript{1} Member, IEEE, and Søren Forchhammer\textsuperscript{1} Member, IEEE

Abstract—Nonlinear phase noise (NLPN) is studied in an experimental wavelength division multiplexed (WDM) system operating at 256QAM. Extremely narrow linewidth lasers (<1 kHz) at the transmitter and the receiver allow for extracting the phase part of the nonlinear noise in a Raman amplified link. Based on the experimental data, the autocorrelation function of the NLPN is estimated and it matches the theoretical predictions. Several algorithms are examined as candidates for tracking and compensating the NLPN. It is shown that algorithms which exploit the distribution of the NLPN achieve higher gains than standard methods, which only exploit the correlation properties. Up to 300 km reach increase is achieved for a 5x10 GBaud WDM system with base distance of up to 1600 km. The gains are comparable to the gains of single channel digital back-propagation, with even further improvements from the combination of both techniques.

Index Terms—Nonlinear phase noise, phase noise distribution, nonlinear compensation, high-order QAM, WDM

I. INTRODUCTION

Nonlinear interference noise (NLIN) is currently among the major limiting factors to improve the distance and spectral efficiency of wavelength division multiplexed (WDM) coherent optical communication systems. Such systems are currently employing modulation formats as high as 16 quadrature amplitude modulation (QAM) with prospects for increased modulation order for short and metro network distances. Single channel systems can potentially be operated at high spectral efficiency with proper digital post- or pre-compensation, e.g. digital back-propagation (DBP). In case of WDM systems, compensation of cross-channel nonlinearities requires DBP of the corresponding interfering channels, which significantly increases the complexity of the receiver. Furthermore, in a WDM network, knowledge of the interfering channels is usually unavailable, making multi-channel DBP infeasible.

It has been previously shown that the NLIN is not completely random [1] and its properties can be used for partial mitigation. Particularly, the phase and polarization rotation noise (PPRN) part of the NLIN exhibits long temporal and spectral correlations [2], which can be exploited for its compensation [3]. The PPRN part is significant mostly in short distances and for high-order QAM, and its compensation can lead to increased distances especially for fully loaded systems [4] and e.g. 64QAM/256QAM. As the distance increases, the NLIN can be described by circular symmetric Gaussian noise, and is therefore more difficult to compensate.

Previous works on nonlinear phase noise (NLPN) compensation rely on standard phase noise (PN) recovery techniques, which mostly exploit the temporal/spectral correlations of the process. The gains of such methods are therefore limited to the above mentioned scenarios, where the autocorrelation function (ACF) of the NLIN is long enough to allow for such tracking. Frequency domain equalization is used in [5], while time-domain sliding window averaging is used in [6]. A more sophisticated trellis-based method was used in [7], which allows for increasing the optimal launch power and thereby the achievable information rate (AIR) by around 5-10%.

Similar studies in experimental environment are difficult due to the non-ideal lasers used at the transmitter and receiver. The non-zero laser linewidth results in PN, which is usually dominating over the NLPN. As discussed in [8], carrier phase estimation (CPE) techniques, which are mandatory in experimental environment, mask the NLIN. The NLIN is experimentally characterized for QPSK/16QAM signals in [9]. The NLIN non-circular statistics are demonstrated by data-aided approaches, assuming the laser PN was ideally removed by a blind phase search (BPS) algorithm. It is shown that with increased distance, the NLIN becomes circular, confirming the predictions made in [2]. The largest deviation from circular statistics was found to be at around 200 km of transmission, which is where the highest gains can be expected for the corresponding setup. In [10], a 9x32 GBaud WDM system is studied with 16QAM, and it is argued that the narrow linewidth lasers used (10 kHz) ensure purely nonlinear PN. However, no further experimental justification is provided for the distinction between laser PN and NLIN.

In this paper, the NLPN is studied experimentally in a 5x10 GBaud WDM system with 256QAM for up to 1600 km. The system is studied in back-to-back and also in the linear regime of transmission, where it is shown that PN compensation provides no gain. The gain is increased with launch power and also with distance for the studied distances, demonstrating...
ing NLPN compensation experimentally. In addition to the previous work from [11], the NLPN properties are studied in detail in Section V, where further evidence is provided into the benefits of soft NLPN tracking and compensation.

II. NONLINEAR PHASE NOISE MODEL AND RECOVERY ALGORITHMS

Laser PN is usually modeled as a first order Wiener process [12]

\[ \theta_k = \theta_{k-1} + \Delta \cdot v_k, \]  

(1)

where \( \theta_k \) is the PN at time \( k \), \( v_k \)'s are independent, identically distributed (i.i.d.) standard Gaussian variables, and \( \Delta \) is the process noise variance. The parameter \( \Delta \) can be found as \( \Delta^2 = 2 \cdot \pi \cdot \delta_f \cdot T_s \), where \( \delta_f \) is the laser linewidth and \( T_s \) is the symbol period. It has been shown that this model can also be used with reasonable accuracy for the NLPN in WDM systems [13], [14]. Since the NLPN and laser PN can be considered independent, their variances add-up and the system process noise variance is

\[ \Sigma_x = \Delta^2 + \sigma^2_z. \]

\( \theta \)

\( \Delta \)

\( \sigma \)

\( \sigma_z \)

\( x \)

\( n \)

where \( \theta \) is the symbol period. It has been shown that this model

\[ \mathbf{y}_k = \mathbf{e}^{j\theta_k} \cdot \mathbf{x}_k + \mathbf{n}_k, \]

(2)

where \( \mathbf{y}_k \) are the received symbols, \( \mathbf{x}_k \) are the transmitted symbols, \( \mathbf{n}_k \) is additive white Gaussian noise (AWGN) and the PN contribution of the transmitter and receiver lasers are added to the NLPN contribution to form the phase \( \theta_k \).

Under this model, well-developed techniques can be used to compensate the PN. In this work, we focus on the probabilistic method proposed in [13], and also the more popular BPS on a sliding window technique [15].

Several receiver implementations are studied in this work. The first is PN unaware, which only assumes AWGN channel with circular symmetric statistics, estimated for each constellation symbol from training data. The statistics which are extracted are the mean \( \mu_x \) and variance \( \sigma^2_x \) of the likelihood

\[ p(\mathbf{y}_k | \mathbf{x} = \mathbf{x}) = \mathcal{N}(\mathbf{y}_k; \mu_x, \sigma^2_x) \]

of each symbol from the alphabet \( \mathbf{x} \in \mathcal{X} \), where \( \mathcal{N}(\mathbf{y}; \mu_\mathbf{y}, \Sigma_\mathbf{y}) \) is the Gaussian distribution in \( \mathbf{y} \). The likelihoods are then used for performance evaluation as explained below. In addition, the AWGN receiver is studied with multi-dimensional statistics, in which case the \( 2 \times 2 \) covariance matrix \( \Sigma_x \) is extracted instead of the variance \( \sigma^2_x \). The symbols \( \mathbf{y}_k \) are then processed as a two-dimensional, real-valued vector signal with elements \( \tilde{\mathbf{y}}_k = [\Re(\mathbf{y}_k), \ln|\mathbf{y}_k|]^T \) and the likelihoods become

\[ p(\mathbf{Y} | \mathbf{X} = \mathbf{x}) = \mathcal{N}(\tilde{\mathbf{y}}_k; \mu_{\mathbf{y}}, \Sigma_{\mathbf{y}}) \]

[16]. This receiver allows for capturing the non-circularity of the NLIN through the non-diagonal covariance matrix. However, it does not capture the temporal correlations of the NLPN.

Next we study the BPS algorithm [15], which assumes the PN is constant in a window of \( L + 1 \) symbols. The algorithm makes decisions

\[ \hat{\phi} = \arg \min_{\phi} \sum_{l=k-L/2}^{k+L/2} |y_l - \hat{\mathbf{x}}_l(\phi) \cdot e^{j\phi}|^2 \]

for a certain number of candidate phases \( \phi \) and estimates the PN as

\[ \hat{\theta}_k = \arg \min_{\phi} \sum_{l=k-L/2}^{k+L/2} |y_l - \hat{\mathbf{x}}_l(\phi) \cdot e^{j\phi}|^2. \]

(3)

In order to mitigate the effect of the AWGN on the quality of the decisions, the sliding window needs to be long. This is particularly problematic for rich constellations at low received SNR (relative to the modulation format size), since most of the decisions are wrong and the phase is thereby incorrectly estimated. On the other hand, long windows might exceed the correlation length of the PN process and fail to capture fast variations. In order to put an upper bound on the performance of such algorithms, we also study a genie phase noise removal (GPNR) technique, which uses the true transmit symbols \( \mathbf{x}_k \) when estimating the phase. The window length can therefore be kept short, thus exploiting the autocorrelation function of the PN process (however short) completely.

The last receiver we study is the algorithm from [13], which assumes a Wiener process for the NLPN and is tailored to estimate PN in such environments. The algorithm requires the values of \( \Delta^2 \) and the AWGN statistics \( \sigma^2_z \) and \( \mu_x \). These are estimated from training data using the GPNR receiver as in Algorithm 1.

Algorithm 1 Estimation of Wiener model parameters

1: Estimate \( \hat{\theta}_k = \frac{1}{L} \sum_{l=k-L/2}^{k+L/2} y_l \cdot x_l^{*} \)

2: Back-rotate \( \hat{y}_k = \hat{y}_k \cdot e^{-j\hat{\theta}_k} \)

3: Estimate \( \Delta^2 = \mathbb{E}_{k_x} (|\hat{\theta}_k - \theta_{k-1}|^2) \)

4: Estimate \( \mu_x = \mathbb{E}_{k_x} [y_{k_x}] \), where \( k_x \) are the symbol indices for which \( x_k = x \)

5: Estimate \( \sigma^2_x = \mathbb{E}_{k_x} [|y_{k_x} - \mu_x|^2] \)

The belief propagation algorithm is then applied, which is a popular method for estimating densities on a factor graph by message passing techniques (further details are given in [13]). Since the messages in the graph are modeled as mixtures of Tikhonov distributions, this receiver is referred to as the Tikhonov mixture model (TMM).

In addition, the performance is studied for each algorithm individually and combining it with single-channel DBP. An overview of the receivers is given in Fig. 1. The figure of merit used in this paper is the generalized mutual information (GMI) [17]. The GMI gives an upper bound on the AIR after forward error correction (FEC) decoding, and is estimated in bits/QAM symbol. In order to estimate the GMI, the posterior
probabilities of all symbols \( p(x_k|y_k^K) \) are estimated, where \( K \) is the total sequence length. The first four receivers assume the PN is ideally removed and the processing is therefore memoryless, i.e. \( p(x_k|y_k^K) = p(x_k|y_k) \propto p(y_k|x_k) \). The memoryless posteriors are then estimated using standard Gaussian receiver techniques. The TMM receiver estimates the posteriors with optimal process for computing the posterior probability of \( x_k \).

Note that the GPNR does not give an upper bound to the GMI performance of the TMM, because it uses a sub-optimal process for computing the posterior probability of \( x_k \).

### III. Experimental setup

The experimental setup is shown in Fig. 2. At the transmitter, 256QAM data symbols are interleaved with QPSK pilots at a pilot rate of 10%. A square root raised cosine pulse shaping is then applied with roll-off factor of 0.5. Five channels on a 25 GHz grid are modulated at 10 Gbaud by this signal with 25 GHz spacing.

The central channel, which is the channel under test, uses a sub-kHz linewidth fiber laser (Koheras BasIQ E-15). Such narrow linewidth virtually ensures nonexistent laser PN in the system, even for such rather low symbol rates. Since this property of the system is of crucial importance to NLPN estimation, the effect of laser PN is studied in back-to-back transmission and transmission at low launch power, as will be seen in Section IV.

The four co-propagating channels use standard, external cavity lasers (ECL) with 100 kHz linewidth. The interfering channels are decorrelated by a wavelength selective switch (WSS) and delay lines, combined with the channel of interest and a delay-and-add polarization emulator provides the dual-polarization signal. The delay lines ensure more than 40 symbols of inter-channel delay, thus preventing nonlinear artifacts due to correlations between the WDM channels [18].

The recirculating loop consists of 100 km of standard, single mode fiber (SSMF) using distributed Raman amplification with backward pumping every 50 km. In order to compensate for the power losses of the acusto-optic modulators (AOMs) used as switches, an erbium doped fiber amplifier (EDFA) is inserted in the loop.

The signal is detected by an 80 GSa/s coherent receiver with a fiber laser as local oscillator (LO). The LO has a similar to the transmitter laser linewidth of < 1 kHz (model Koheras BasiK E-15). Offline processing is performed consisting of (in order) low-pass filtering, down sampling, chromatic dispersion (CD) compensation in frequency domain, frequency offset estimation based on the pilots, time-domain equalization and carrier phase recovery. As mentioned, CD compensation is replaced by DBP for part of the study. DBP is performed with 40 steps per span, which we found to be sufficient for maximum performance improvement. The power profile of the signal as a function of distance during DBP was optimized based on the analytical models for backward pumping Raman amplification [19]. Equalization is performed with the constant modulus algorithm (CMA) with 101 taps. The equalizer taps are only updated at the pilot positions, and after convergence, they are linearly interpolated in order to equalize the entire received sequence. The sequence after equalization in each polarization is denoted \( y_k^K \). In this work, \( K = 6600 \) symbols per block, and 12 blocks are received and processed for each polarization. The average GMI over all blocks is reported.

The receivers from Fig. 1 are then applied. We note that the BPS and the TMM receiver take advantage of the QPSK pilot symbols: the former by using the actual transmitted symbol when estimating the cost function in Eq. (3), and the latter by exploiting the fact that the prior distribution \( p(x_k) = 1 \) for the true symbol and \( p(x_k) = 0 \) for all other symbols at the pilot instances (details are given in [13]).

### IV. Results

We studied 256QAM input in optical back-to-back and for distances between 800 km and 1600 km. The sliding window length is set to \( L = 50 \) symbols. The GMI average of both polarizations after 1400 km is given in Fig. 3. The BPS algorithm could not achieve reasonable performance for this window length and distance, and is thus not shown (discussion on the window length is provided in the next section). The solid lines are obtained by replacing the CD compensation with single-channel DBP. In the linear region of transmission, PN tracking is not beneficial. We can therefore conclude that laser PN is insignificant and any PN in the system originates in nonlinear...
V. NONLINEAR PHASE NOISE PROPERTIES

As shown in Fig. 1, the parameters $\sigma_x^2$, $\mu_x$ and $\Delta^2$ depend on the PN estimates of the GPNR receiver, which depend strongly on the window length $L + 1$ used for averaging. It is therefore of interest to optimize the window length for best performance of each receiver. Such performance results are given in Fig. 6 for a transmission of 1400 km at the optimal total launch power, $-3$ dBm. As seen, the BPS requires a certain number of symbols (>100) to average the effect of the AWGN. For shorter $L$, it performs even worse than the AWGN receiver, which assumes no PN in the system. This penalty is due to more than 50% wrong decisions, which results in estimation of non-existent PN. The GPNR receiver achieves a gain of approximately 0.1 bits/symbol w.r.t. the AWGN receiver, which is only slightly dependent of the window length. The gain is increased to 0.2 bits/symbol with the TMM, which operates optimally for $L = 50$ in the GPNR parameter estimation. We see that the TMM is also almost independent of $L$, even though different values results in significantly different estimated $\Delta^2$. This can be explained by Fig. 7, where the estimated values of $\Delta^2$ are given as a function of $L$, together with the variance of the estimates $\hat{\theta}_k | y_1^L$. For short $L$, $\Delta^2$ is found to be large, which is the result

1We expect that for the channel spacing and the low channel count considered here the inter-channel NLIN is dominated by cross-phase modulation rather than four-wave mixing.
of the GPNR receiver’s fast adaptation to PN. On the other hand, for such $L$ the uncertainty around the PN estimates of the TMM receiver is also large, which penalizes over-fitting to the estimate. As mentioned, the GPNR assumes a Delta function around this estimate and thus has worse performance. For long windows, $\Delta^2$ is estimated to a smaller value since small-scale variations are not captured by the GPNR. The uncertainty of the estimates $\theta_k|y^K_1$ also diminishes. In the limit of $L = K$, the estimates $\Delta^2 = 0$, $\theta_k|y^K_1 = 0$ for all $k$, $\text{Var} \{\hat{\theta}_k|y^K_1\} = 0$ and the performance of all receivers converges to the performance of the AWGN receiver.

Finally, in Fig. 8 the ACF of the estimated PN is given for several $L$. The theoretical ACF as calculated in [1] is also given. The ACF of the TMM algorithm is calculated by first estimating the values $\hat{\theta}_k = \arg\max p(\theta_k|y^K_1)$, and then calculating the ACF of the process $\Theta$. For $L = 10$, the ACF of the GPNR estimates approaches the theoretical ACF (at least for the small arguments of the ACF of $<20$ symbols), and as we saw in Fig. 6, it produces the best performance. We conclude that the correlation properties of the NLPN are completely captured by this receiver and this $L$. When $L$ is increased, the estimated ACF becomes longer. This is partly due to the filtering properties of the sliding window phase estimation, which effectively introduces correlations of up to $L$ symbols itself. This filtering property is also the reason for the overestimation of the ACF for arguments $>20$ symbols. Since the correlation properties are not correctly estimated, the GPNR receiver is penalized. On the other hand, the TMM operates at a different trade-off between correlation and estimation uncertainty. As seen, the ACF of the TMM estimates $\hat{\theta}_k = \arg\max p(\theta_k|y^K_1)$ is always overestimated w.r.t. the ACF of the GPNR estimates. However, the broader distribution of $\hat{\theta}_k|y^K_1$ allows for improved performance. As we saw in Fig. 6, the TMM performance is optimized for $L = 50$ during the Wiener process parameter estimation. This suggests that when GMI is the figure of merit, the ACF of the NLPN is not the only statistic of interest for describing the process and analyzing the performance gains from NLPN compensation.

VI. DISCUSSION AND FUTURE WORK

In systems employing lasers with broader linewidth than the fiber lasers used in this work, the contributions of the NLPN and laser PN are generally independent, and the more sophisticated methods for PN tracking (such as the TMM) will provide additional gains from mitigating the NLPN part more effectively.

It has been shown that the NLPN becomes a more significant part of the total NLIN with increased number of WDM channels [20]. We can therefore expect that the gains from NLPN compensation in that case exceed the gains in the five-channel system demonstrated here.

The sliding window approaches considered in this work employ a flat window, i.e. all values in the window have the same weight. We have also examined triangular and exponential windows, which produced negligible difference in the GMI.

In this work, a first-order Wiener process was assumed for the PN component of the NLIN. More complex models (e.g.
of higher order) may provide even higher gains by capturing longer memory of the process. Finally, it is noted that the studied implementation of the TMM algorithm assumes circular Gaussian noise. Modifications of the algorithm which allow for 2D processing with correlations between dimensions are also of interest for potential further improvement of the AIRs.

VII. CONCLUSION

Several algorithms for nonlinear phase noise (NLPN) compensation were studied experimentally in a 5x10 Gb/s WDM system. It was demonstrated that probabilistic methods which exploit the distribution of the phase noise are beneficial w.r.t. simply exploiting the NLPN correlation properties. Such methods are necessary particularly in metro-to-long haul distance systems, where the NLPN correlations are not strong enough and do not allow for simple phase estimation. In the studied system, up to 300 km of gain was achieved by NLPN compensation for distances of up to 1600 km, which is comparable to the gains offered by single channel digital backpropagation. It was also demonstrated that the gains increase with distance, which cannot be achieved with simple methods even in scenarios with strong temporal/spatial correlations of the NLPN.

VIII. ACKNOWLEDGMENTS

This work was supported by the DNRF Research Centre of Excellence, SPOC, ref. DNRF123. NKT Photonics A/S is acknowledged for providing the narrow linewidth fiber lasers used in the experiment.

REFERENCES