Review

Equalization Methods for Out-of-Band Nonlinearity Mitigation in Fiber-Optic Communications

Ori Golani * , Meir Feder and Mark Shtaif

School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel; meir@eng.tau.ac.il (M.F.); shtaif@tauex.tau.ac.il (M.S.)

* Correspondence: origolani@post.tau.ac.il

Received: 14 December 2018; Accepted: 28 January 2019; Published: 2 February 2019

Abstract: In recent years, it has been established that the adverse effects of nonlinear interference noise (NLIN) can be mitigated using adaptive equalization methods. As such, a wide variety of adaptive equalization methods have been used to treat nonlinearity, in different transmission scenarios. This paper reviews the principles of out-of-band nonlinearity mitigation using adaptive equalization. Statistical properties of NLIN that can be exploited for mitigation are discussed, as well as the cost and benefit of various types of equalizers. In particular we describe the equivalence between the NLIN and time-dependent inter-symbol-interference (ISI) and discuss ways in which the ISI coefficients can be characterized theoretically and experimentally. We further discuss the effectiveness of existing ISI mitigation algorithms, and explain the need for designing customized algorithms that take advantage of the various correlation properties characterizing the ISI coefficients. This paper is intended to be a practical reference for researchers who want to apply equalization algorithms or design new methods for nonlinearity mitigation.

Keywords: optical fiber communication; nonlinear Interference; time varying inter symbol interference; nonlinearity mitigation

1. Introduction

The challenge of mitigating nonlinear interference in wavelength division multiplexed (WDM) networks has attracted significant efforts. Receiver-side digital nonlinearity compensation is of particular interest, as it requires no additional optical hardware and can be easily adapted to many cases. Nonlinear interference caused by Kerr nonlinearity is a deterministic process—given the input waveform and fiber link structure, one can predict the effect of nonlinearity at the fiber's output with an accuracy limited only by the nonlinear noise-signal and noise-noise interference that kick in only at very high signal powers and low signal to noise ratios. Furthermore, neglecting nonlinearity involving the noise, the nonlinear distortions can be canceled by digitally back-propagating the signal so as to invert the nonlinearity accumulated along the fiber. The problem is that in realistic WDM systems, the receiver has access only to a small fraction of the overall signal bandwidth. Namely, the nonlinear interference that is experienced by the detected signal and which was induced by signals that are outside the bandwidth of the receiver, cannot be inverted by such means. Broadly, the sources of nonlinearity can be divided into two groups, as illustrated in Figure 1: in-band nonlinearity, which involves interactions of frequency components that are within the receiver's bandwidth, and out-of-band nonlinearity, which is produced by interaction with WDM channels that are outside the received bandwidth. The treatment of these two nonlinearity sources is very different; in-band nonlinearity can be viewed as a deterministic signal-dependent effect, and hence it can be effectively reduced by means of digital back propagation [1], sequence detection [2], or Volterra-series equalization [3]. In contrast, out-of-band nonlinearity must be considered as noise, commonly referred to as “nonlinear

interference noise" (NLIN). In most systems, the dominant source of out-of-band NLIN is two-channel interference, namely cases where one out-of-band WDM channel interacts with one in-band channel. Interactions involving three or four channels are also present, but their effects in standard fibers are usually negligible compared to two-channel interactions (Three and four channels effects can be neglected in systems using standard single-mode fiber, when the baud rate is larger than 10 GBaud. In dispersion shifted fibers, or if the symbol rate is significantly lower and the number of channels is higher, these effects may become the dominant noise source [4]. As such systems are not common, we do not consider this regime in the paper). It has been well established that two-channel interference is manifested predominantly in the form of linear time-varying inter-symbol interference (ISI) [5,6]. Therefore, out-of-band NLIN can be mitigated using adaptive equalization techniques.

Figure 1. Schematic classification of NLIN types, and methods of mitigating them.

Nonlinearity mitigation using adaptive equalization has been demonstrated in many reported studies. A wide variety of equalization algorithms has been used in different transmission scenarios. The vast majority of these efforts have been directed towards phase and polarization-rotation noise (PPRN) mitigation [6–16], which is the most significant NLIN contribution, as we elaborate in what follows. Yet, as discussed below, PPRN is only one component of the interference, and it is accounted for by the zeroth-order term in the time-varying ISI model. The contribution of higher order terms is non-negligible and, as we will show, its mitigation has the potential of notably improving performance. Various aspects of this issue have been dealt with in [17–21].

In this paper, we review the recent advances in the application of equalization techniques for out-of-band NLIN mitigation, discuss the differences and similarities between different approaches, and assess the performance of several equalization schemes. The paper reviews the concepts needed to design new equalization algorithms for nonlinearity mitigation, and discusses the properties of several such methods that have been previously reported in the literature. Section 2 introduces the time-varying ISI model of NLIN, and discusses briefly its origin and properties. Methods of characterizing the statistical properties of the ISI process are also reviewed, as they are of importance for designing equalization algorithms. The different types of NLIN correlations are discussed in Section 3. These correlations are the key feature that enables NLIN mitigations, and so exploiting as many types of correlations as possible can dramatically improve an equalizer's performance. Section 4 reviews the main types of algorithms that have been used in the literature, comparing their performance and computational cost in an exemplary link. In particular, in Section 4.3 we introduce the concept of Turbo-equalization, which shows a high potential for NLIN mitigation. Section 5 is devoted to conclusions.

2. The Time-Varying ISI Model of NLIN

We begin the discussion from the “Time-varying ISI” model of NLIN, which describes the effect of two-channel nonlinear interference [5,8,22]. Our goal is to provide an intuition for the consequences of this model, without repeating the rigorous derivations given in [8]. We label the two interacting channels, co-propagating along the fiber, by A and B , where channel A is regarded as the channel of interest (COI) and channel B is an interfering channel (IC). At the link’s output, the channels are de-multiplexed, detected, sampled, and dispersion compensation is applied to them. It is assumed that the receivers are optimized for the linear channel, i.e., that, without the effects of nonlinearity, the received symbols are corrupted only by amplified spontaneous emission (ASE) noise. The two channels carry a series of pulses, modulated by the transmitted data. For the sake of simplicity, we consider only single-polarization transmission at this point, whereas dual polarization transmission will be discussed later. We denote the symbols at time slot n by a_n and b_n , for channels A and B , respectively.

Suppose that we want to evaluate the effect of NLIN on a certain data symbol, a_n . As shown in [23], under first-order perturbation approximation the nonlinear interactions induced by two-channel interference take the form

$$\Delta a_n = \sum_{h,k,l} i X_{h,k,l} b_{n-h} b_{n-k}^* a_{n-l}. \quad (1)$$

where $X_{h,k,l}$ is an interaction coefficient. Each term in the summation can be viewed as representing a collision between pulses [24], and it accounts for the changes imposed by the interaction of these particular pulses on the data symbol a_n . The interaction coefficients $X_{h,k,l}$ are determined by many parameters, such as the link length, fiber properties, frequency difference between the channels, and their spectral shapes. The derivation of the expression in Equation (1) and the exact calculation of these interaction coefficients are described in [23]. In the case of a WDM system with more than two channels this process is performed separately for each of the ICs, and their contributions must be summed in order to obtain the overall effect.

A simpler description can be obtained by characterizing the collective effect of the nonlinear interactions, rather than the contributions of each individual pulse collision. By rearranging the indices in Equation (1), one obtains

$$\Delta a_n = i \sum_l \left(\sum_{h,k} X_{h,k,l} b_{n-h} b_{n-k}^* \right) a_{n-l} = i \sum_l R_l^{(n)} a_{n-l}. \quad (2)$$

The idea behind this formulation is to separate between the IC’s data symbols, which are not available to the receiver, and those of the COI, which need to be estimated. The coefficients $R_l^{(n)}$ encapsulate the effect of the interfering channel. The fact that the IC carries random data causes the coefficients $R_l^{(n)}$ to change over time, as is indicated by the superscript (n) . The key strength of this approach is that it is sufficient to characterize the behavior of the $R_l^{(n)}$ coefficients in order to fully describe the effect of nonlinearity. The full derivation of this concept is described in [7,8,25].

Using Equation (2), we can express the received signal after dispersion compensation, matched filtering, and standard carrier-phase and frequency recovery, as

$$s_n = a_n + \Delta a_n + w_n = a_n + i \sum_l R_l^{(n)} a_{n-l} + w_n, \quad (3)$$

where w_n is a noise term containing the contribution of ASE, as well as all of the nonlinear effects that are not treated by the model (such as three or four-channel interactions). This equation describes a *linear*, time-varying ISI channel, which has been extensively investigated in the field of RF communication, and is commonly referred to as a “doubly-selective channel” [26,27].

In the case of polarization multiplexed transmission the transmitted data symbols become two-element vectors, $\bar{a}_n$ and $\bar{b}_n$. The derivation is identical to the single-polarization case, and the time-varying ISI model is expressed as

$$\bar{s}_n = \bar{a}_n + i \sum_l \mathbf{R}_l^{(n)} \bar{a}_{n-l} + \bar{w}_n. \quad (4)$$

In this case, the ISI coefficients, $\mathbf{R}_l^{(n)}$, are 2×2 matrices, where the off-diagonal elements describe interactions between the two polarizations.

2.1. Phase and Polarization Rotation Noise (PPRN)

Separating the 0-th ISI term from the summation in Equation (4), we obtain

$$\bar{s}_n = \bar{a}_n + i \mathbf{R}_0^{(n)} \bar{a}_n + i \sum_{l \neq 0} \mathbf{R}_l^{(n)} \bar{a}_{n-l} + \bar{w}_n, \quad (5)$$

where, as shown in [8,23], the ISI matrix $\mathbf{R}_0^{(n)}$ is Hermitian. Since the entire perturbation analysis is only meaningful to first order in the nonlinearity, the above can be rewritten as

$$\bar{s}_n = e^{i \mathbf{R}_0^{(n)}} \bar{a}_n + i \sum_{l \neq 0} \mathbf{R}_l^{(n)} \bar{a}_{n-l} + \bar{w}_n. \quad (6)$$

The Hermiticity of $\mathbf{R}_0^{(n)}$ implies that $e^{i \mathbf{R}_0^{(n)}}$ is unitary, and hence it can only represent phase and polarization rotation. Naturally, these rotations are time dependent because $\mathbf{R}_0^{(n)}$ depends on the IC's data, and hence this phenomenon can be referred to as phase and polarization-rotation noise, or PPRN. The PPRN element is the most significant component in the time-varying ISI model, as it has the largest variance and longest correlation length [28]. For this reason, most of the initial efforts in the field of nonlinearity equalization have been directed towards PPRN mitigation. Moreover, it must be mentioned that since all coherent systems are inevitably equipped by phase-recovery and polarization demultiplexing algorithms, some of the PPRN contribution is compensated intrinsically, even when no dedicated NLIN mitigation algorithms are introduced.

2.2. Characterization of ISI Statistics

In order to design a successful equalization algorithm, one must characterize the statistical behavior of the ISI coefficients, $\mathbf{R}_l^{(n)}$. This behavior is determined by many factors, such as the modulation format, the baud rate of the ICs, the fiber link's structure, and more. To date, most efforts in characterizing these properties have been directed towards second-order moments, namely the variance and covariance of the ISI coefficients. These were obtained analytically by determining the interaction coefficients ($X_{h,k,l}$ in Equation (1)) [8], or extracted from experiments or split-step simulations [4,28,29]. The extraction from experiments and simulations relies on estimating each ISI coefficient from Equation (3) using

$$\tilde{R}_l^{(n)} = -i \frac{s_n - a_n}{a_{n-l}} = R_l^{(n)} + v_n, \quad (7)$$

where s_n and a_n are taken directly from the simulation or experiment, and where v_n is estimation noise, which can be modeled as white noise provided that the transmitted symbols are uncorrelated to each other. Thus, by assessing various statistical properties of $\tilde{R}_l^{(n)}$, we can infer those of $R_l^{(n)}$. For example, the whiteness of v_n implies that $\langle \tilde{R}_l^{(n)} \tilde{R}_l^{(n+\Delta n)} \rangle = \langle R_l^{(n)} R_l^{(n+\Delta n)} \rangle$ for $\Delta n \neq 0$, which means that the autocorrelation functions of $\tilde{R}_l^{(n)}$ and $R_l^{(n)}$ are identical (with the exception of the point $\Delta n = 0$). Similarly, the cross-correlation between $\tilde{R}_{l_1}^{(n)}$ and $\tilde{R}_{l_2}^{(n)}$ is identical to that of $R_{l_1}^{(n)}$ and $R_{l_2}^{(n)}$. Note that the method of evaluating the amount of nonlinear phase noise [4,29,30] is obtained by setting $l = 0$ in Equation (7). For more details, the reader is referred to [28].

2.3. Estimating System Performance under the Time-Varying ISI Model

While the distribution of the NLIN may be fairly close to Gaussian, the various correlations that characterize it make NLIN substantially different from standard additive white Gaussian noise (AWGN), such as ASE. As a result, the system's BER performance cannot be accurately predicted by using the standard noise formulas that were constructed for AWGN [31], particularly when accounting for the presence of equalization mechanisms (designed for NLIN mitigation, or even of the kinds that are standardly included in optical communications receivers). As a result, one is forced to rely on simulations that are based on the split-step Fourier method (SSFM) which are almost unacceptably time consuming, especially when considering large channel-count WDM systems. Moreover, as implied by the results of [32], attempts to avoid this difficulty by simulating only a small fraction of the interfering channels will also suffer from significant inaccuracies.

In order to overcome this difficulty the Virtual Lab tool was introduced in Ref. [8]. The idea behind the Virtual Lab is that it is sufficient to characterize the properties of the ISI coefficients in order to accurately describe out-of-band NLIN. Therefore, if one artificially generates ISI coefficients with the correct statistical properties, it is possible to reproduce the relevant characteristics of NLIN without using SSFM. The Virtual Lab tool has been shown to provide accurate estimation of system performance (namely, the SNR and BER) in the presence of adaptive equalization [7,8,18].

In following sections we show system performance estimates that are based both on SSFM and on the virtual Lab, allowing the reader to appreciate the accuracy of the tool.

3. Making Sense of Correlations

Two properties of NLIN that make the time-dependent ISI model particularly useful, are the correlation times that characterize the individual ISI coefficients and the correlations between different ISI coefficients. These two properties enable the receiver to evaluate the nonlinearity and attempt to mitigate its effect. Naturally, the more types of correlations the equalization algorithm can exploit, the better it will perform. We briefly discuss the main types and origins of these correlations.

1. Temporal correlations

As stated previously the ISI coefficients are time-varying, because of the random data carried by the ICs. However, adjacent symbols of the COI 'feel' the interaction with the IC in a similar way, which implies that the nonlinear perturbations that are imposed upon them are correlated [23]. The correlation times are best evaluated in terms of the number of symbols, and they increase with the fiber dispersion, the baud rate and the channel spacing. In typical WDM systems implemented over standard single-mode fiber (SMF) and using baud-rates of the order of 30 Gbaud, the correlation times are of the order of tens to hundreds of symbols in metro and long haul applications, which in principle allows their tracking and mitigation. Generally, the higher the order of the ISI coefficient, the shorter its correlation time, and the more difficult is its mitigation. Indeed, the longest correlation time is that of the 0-th order ISI element, which accounts for PPRN [28,30].

2. Cross-polarization correlations

As shown in Equation (4), for dual-polarization transmission, the ISI coefficients become 2×2 random matrices. The elements of the matrices are, however, strongly correlated to each other. For each of the ISI matrices, $\mathbf{R}_l^{(n)}$, it can be shown that the two on-diagonal elements are almost identical, whereas the two off-diagonal elements are complex-conjugates of one another. A detailed explanation of this fact can be found in [8]. The fact that the various terms in the ISI matrices are strongly correlated with each other implies that there is a great advantage in joint-processing of the two polarizations.

3. Correlations between different ISI orders

It has been found in [28] that measurable cross-correlations exist only in the case of symmetric ISI

orders, namely $R_l^{(n)}$ and $R_{-l}^{(n)}$. The correlations are fairly strong for small values of l and reduce as the value of l increases. Other pairs of ISI matrices are uncorrelated for all practical purposes.

4. Correlations in multi-subcarrier systems

Another relevant type of correlations relates to the scenario where multiple data-channels are jointly detected by the same receiver. This characterizes for example the cases of OFDM, or sub-carrier multiplexing. In such cases, out of band NLIN (i.e., NLIN generated by channels that are outside of the receiver's bandwidth) affects the different tributaries in a correlated manner. These correlations follow from the fact the frequency separation between the tributaries is much smaller than the separation between them and the out of band ICs, and hence they are affected by them in a similar way. This type of correlations dominates in the very low baud rate regime (as in OFDM or in symbol-rate optimization schemes [33,34]), where the temporal correlations diminish.

An example for these various correlation types is shown in Figure 2. Multi-subcarrier systems are not discussed further in this review, and remain a subject for future study. The results of Figure 2 were obtained for the 1000 km link discussed in Section 4, and calculated using the method described in [28]. The results for Figure 2d are taken from [34].

Figure 2. Examples of different auto-correlation and cross-correlations characterizing NLIN. (a) the auto-correlation functions of the first 4 ISI components. (b) cross-correlation functions between different elements of the ISI matrix $\mathbf{R}_0^{(n)}$, and (c) cross-correlation between different ISI orders. The superscripts Re and Im denotes the real and imaginary parts of the coefficients, whereas the subscripts xx , xy , yx , and yy denote the four components of a 2×2 ISI matrix. For example, the curve marked by $R_{0,xy}^{Re}, R_{0,yx}^{Re}$ denotes the cross-correlation between the real part of the off-diagonal elements of $\mathbf{R}_0^{(n)}$. (d) normalized (Pearson) correlation of the phase-noise component in a multi-subcarrier system. The curves show the correlation between the phase noise affecting two tributaries within the COI, as a function of the frequency separation between them. Results are shown for 64-QAM modulation, and probabilistically-shaped (PS) 64-QAM. All of the results were obtained for a 10×100 km link of SMF. The auto-correlation and cross-correlation functions of figs. (a–c) were measured for a system carrying 15 WDM channels, whereas in fig (d) 21 WDM channels were used.

4. Equalization Algorithms

Two distinct features that characterize the ISI generated by NLIN must be kept in mind when devising methods for its mitigation. On the one hand, the typical correlation times (tens to hundreds of symbols) are significantly shorter than in the case of most other ISI systems. This fact makes the design of an adaptive equalization algorithm fairly challenging, as it hinders the seamless adoption of readily available adaptive equalization methods. On the other hand the distortion that is generated by NLIN induced ISI usually does not prevent the extraction of the transmitted data even without mitigation, albeit with an unacceptably higher error-rate (This is in contrast to the ISI produced by other effects, such as chromatic dispersion, which make the signal completely unintelligible). This feature facilitates the process of equalization in ways that will be elaborated in what follows. An adaptive equalization algorithm generally consists of two blocks: estimating the ISI coefficients (channel estimation), and attempting to remove their adverse effect (equalizing). All of the algorithms reviewed here are decision-directed equalizers, namely they contain an additional block that performs an initial estimation of the symbols which is subsequently improved by the equalization process (see Figure 3). The performance relies on the quality of the initial symbol estimation and therefore it is expected to degrade considerably in low SNR regimes. We discuss several equalization algorithms in what follows.

Figure 3. A block-diagram description of a general equalizer. All equalizers discussed in this section follow the same structure, with the differences being in the operation of the channel estimation and equalization blocks.

4.1. Standard Equalization Algorithms

There are several important aspects that require consideration when choosing an equalization algorithm. The method should be able to track the fast variations of the ISI coefficient, be robust to symbol estimation errors, and be computationally efficient. We will now discuss a few of these algorithms. The figures of merit to describe the quality of channel estimation is the improvement in post-equalization SNR, and the computational cost of the algorithm. For the reader's convenience, we also include a short description of the algorithms themselves. For the clarity of presentation, the algorithms are described for the scalar case, keeping in mind that the extension to polarization-multiplexed transmission is usually straightforward. In the following, $\hat{s}_n$ denotes the equalized signals, $\tilde{H}_n$ is a vector containing the equalizers' taps, and $D(s_n)$ represents the hard decision regaining the value of the n -th data symbol (i.e., in the absence of errors $D(s_n) = a_n$).

1. Least mean squares (LMS) algorithms

The LMS algorithm is essentially a gradient descent method, where the evaluation of the channel filter improves over time, until it converges. The LMS algorithm has been used in [6,7] to mitigate

PPRN. The benefit of using dual-polarization decision to improve its performance was shown in [20,21,35]. The algorithm can be found in textbooks [36], and is summarized as

$$\hat{s}_n := \bar{H}_n^\dagger \bar{s}_n \quad (8)$$

$$\epsilon_n := D(s_n) - \hat{s}_n \quad (9)$$

$$\bar{H}_{n+1} := \bar{H}_n + \mu \frac{\epsilon_n^* \bar{s}_n}{\bar{s}_n^\dagger \bar{s}_n}. \quad (10)$$

Here $\bar{s}_n$ is a vector containing the received signal samples s_n at times $n \dots n + N - 1$, where N is the equalizer's number of taps. The parameter μ controls the speed of convergence, and it is chosen empirically so as to produce the best equalization results.

The main advantage of this algorithm is its computational efficiency. However, its tracking speed is fairly slow. In practice, the LMS algorithm has been shown to provide reasonable mitigation of PPRN, but not of higher order ISI contributions.

2. Recursive Least squares (RLS) algorithms

The RLS algorithm is usually characterized by a higher convergence speed than LMS, which generally implies improved tracking speed of fast channel variations. It has been used to mitigate PPRN [6,8] as well as higher order NLIN induced ISI contributions, although the latter was only demonstrated in idealized (error-free) scenarios [17]. The algorithm can be summarized as

$$\hat{r}_n := \mathbf{P}_n \bar{s}_n \quad (11)$$

$$\mathbf{P}_{n+1} := \frac{1}{\lambda} \left(\mathbf{P}_n - \frac{\hat{r}_n \hat{r}_n^\dagger}{\lambda + \hat{r}_n^\dagger \hat{r}_n} \right) \quad (12)$$

$$\bar{H}_{n+1} := \bar{H}_n + \frac{\hat{r}_n \epsilon_n^*}{\lambda + \hat{r}_n^\dagger \hat{r}_n}, \quad (13)$$

where the error metric ϵ_n , the signal vector $\bar{s}_n$, and the equalized signal $\hat{s}_n$ are defined as in Equations (8) and (9). The parameter λ is called the algorithms 'forget factor', and it determines the algorithm's tracking accuracy.

In general, RLS is considered to be more computationally expensive than LMS, because it involves matrix multiplications. In the case of NLIN mitigation, the equalizer's number of taps is fairly small, corresponding to small matrices. In practice, we found that in relevant scenarios, the computational costs of RLS and LMS are similar (see Figure 4b).

3. Window-averaging algorithms

Window-averaging algorithms use the fact that the NLIN is temporally correlated whereas other noise sources are not. Therefore, when averaged over a finite time window (possibly weighted averaging), the effects of ASE noise will tend to diminish and an estimation of the ISI coefficients can be obtained. This principle has been used for phase-noise mitigation [9–12], polarization rotation cancellation [13,16], as well as for higher order nonlinear ISI mitigation [20,21].

We implemented the algorithm described in [21]. For the single polarization case, it can be summarized by the following procedure.

$$\epsilon_n := D(s_n) - s_n \quad (14)$$

$$H_l := \alpha_l \frac{\sum_k W_{l,k} \epsilon_{n+k} s_{n+k+l}^*}{\sum_k W_{l,k} |s_{n+k+l}|^2} \quad (15)$$

$$\hat{s}_n := s_n - \sum_l H_l s_{n-l}. \quad (16)$$

Here $W_{l,k}$ is a weighting factor (for example, in [21] exponential weighting was used), and α_l is a constant weight function assigned to each filter tap. The parameters α_l and the window weighting functions $W_{l,k}$ need to be optimized for each specific transmission scenario.

One of the main advantages of this algorithm is its relative versatility. While LMS and RLS each have a single parameter to be optimized, the $W_{l,k}$ and α_l can be different for each filter tap, and so they can be adapted to account for the different properties of the ISI coefficients. However, this algorithm tends to be more computationally expensive than LMS or RLS, as the required window sizes can be quite large.

4. Kalman filters

Kalman filters are the most complex equalizers reviewed in this paper. They rely on an auto-regressive model description of the NLIN process. The advantage of this description is that it enables to account for a detailed statistical model of NLIN, utilizing all of the correlations that it entails. The properties of the auto-regressive model of NLIN can either be found heuristically [14], or by using an estimation of the NLIN's covariance matrix and autocorrelation [19,37]. The description of this algorithm is fairly involved and hence the reader is referred to [19], for the details of the algorithm that we implemented in this paper.

Of the methods reviewed in this paper, equalizers based on Kalman filters offer the highest NLIN mitigation capability, but are also the most computationally expensive.

4.2. Numerical Estimation of Equalization Performance

We turn now to a quantitative comparison of the algorithms using a set of simulations. The simulations were performed using both the split-step method and the Virtual Lab tool [8], under the same test conditions. The test system consisted of 10×100 km spans of single mode fiber ($D = 17$ ps/nm/km, $\gamma = 1.3$ W⁻¹km⁻¹, $\alpha = 0.2$ dB/km), separated by lumped amplifiers characterized by a noise figure of 4 dB. The transmitted signal consisted of 15 WDM channels, carrying a polarization multiplexed 32-QAM modulated constellation, operating at a baud-rate of 32 Gbaud with root-raised-cosine pulses having a roll-off coefficient of 0.02, and with channel spacing of 50 GHz. Each trace contained 2^{16} dual-polarization symbols.

All of the above mentioned algorithms were applied to the received signal. The algorithms' parameters (convergence parameter μ for LMS, forget factor λ for RLS, window weights W_l and a_l for the window averaging algorithm) were optimized so as to obtain the best performance. The Kalman filter's parameters were found using a measurement of the statistics of NLIN, as described in [37]. Figure 4 shows the effect of each of the algorithms on the output SNR, defined as

$$\text{SNR}_{\text{equ}} = \frac{\text{Var}(a_n)}{\text{Var}(\hat{s}_n - a_n)}, \quad (17)$$

where $\hat{s}_n$ is the equalized signal for each method. Figure 4a shows an SNR vs. power plot for each of the cases, where the black curve represents unequalized transmission, and the colored curves show different types of equalizers.

Figure 4b shows the SNR improvement of each equalizer (i.e., the difference between the peaks of colored curves to that of the black curve in Figure 4a), vs. their computational cost. All of the algorithms described in this paper rely only on matrix multiplications and vector scalar product. The sizes of these vectors and matrices is typically fairly small, which means that an asymptotic analysis of the algorithm's complexity is not useful. For example, the RLS algorithm requires matrix multiplication, whereas the LMS algorithm requires only vector multiplications, which would suggest that the RLS algorithm is more expensive. However, because of the small sizes of these vectors and matrices, we found that the cost of these two algorithms was similar. The complexity analysis is particularly difficult for the case of the Kalman filter algorithm, which uses multiplication of sparse matrices [19].

To circumvent this issue, we quantify the computational cost in terms of the computation time required to process a full trace of 2^{12} symbols. The computation times are normalized by that of the LMS equalizer, which is used as a reference. Naturally, this metric depends on the specific implementation and type of computer used for processing, and the absolute values are likely to vary significantly between different computers. Nevertheless, it gives a useful insight as to the relative cost of different equalization algorithms. All of the methods in this paper were implemented using Matlab programming language, and executed on a desktop computer. The LMS and RLS algorithms provided similar performance, both in terms of cost and SNR gain, which was approximately 0.2 dB. The window-averaging algorithm performed significantly better, with an 0.5 dB and 0.8 dB SNR gain for 1-tap and 3-tap equalizers, respectively. In terms of computation time, the 1-tap window equalizer was twice slower than LMS, and the 3-tap equalizer was 6 times slower. The highest gain was provided by a 3-tap Kalman filter, which yielded a 1.6 dB SNR improvement, at the cost of being 50 times slower than the LMS equalizer. With such a high computational cost, real-time implementation of these algorithm may be challenging. Nevertheless, dedicated hardware (e.g., using an field programmable gate array (FPGA)), even the Kalman filter equalizer may be implemented in real time.

Figure 4. SNR improvement of different equalization algorithms. (a) SNR vs. power for each algorithm. The curves represent results obtained using the Virtual Lab tool [8], whereas the dots correspond to split-step simulations. (b) Peak SNR gain vs. computational cost, obtained by measuring the time that was needed to process a single set of symbols for each algorithm. The SNR gain is defined as the difference between the peak SNR of the unequalized case to those of the different algorithms.

4.3. Turbo Equalization

As mentioned earlier, all of the equalization algorithms that we employed are decision-directed, meaning that they rely on the estimation of the transmitted symbols. In most of the works dealing with equalizing NLIN, this estimation was based on a simple minimum-distance decision regarding symbol values. A significant improvement to the algorithm performance can be achieved by combining the process of equalization with forward error correction (FEC) coding—a procedure that is usually referred to as *turbo-equalization* [38,39]. Indeed, including a decoder in the equalization loop results in more accurate symbol estimation, and thus better equalization performance. For example, in [20,21,35], a maximum-a posteriori (MAP) was used to provide improved symbol estimation for an LMS equalizer and a window-averaging equalizer, resulting in improved detection. A low-density parity check (LDPC) decoder was used to improve nonlinear phase noise mitigation in [15], an LDPC code and an RLS equalizer were used in [40], while [19] used an LDPC code together with a Kalman equalizer. The structure of a generic Turbo-equalizer is shown in Figure 5. The main idea is that the symbol estimation is performed using the post-FEC decoded bits. These include much fewer errors than the pre-FEC bits, and so the symbol estimation is much better. The way in which the symbols are estimated

varies between different decoding schemes. For example, one may want to use soft symbol estimates instead of hard decisions, as done in [15,19], or include bit interleaving. The decoding-equalization loop is usually iterated several times, until the algorithm converges.

Figure 5. A block-diagram description of a Turbo equalizer.

In order to compare the performance of the different equalizer, we employed a Turbo-equalization scheme based on soft-decision LDPC encoder, as done in [19]. The code rate was $R = 0.85$, corresponding to a FEC overhead of 15%, while all other system parameters were kept as in the previous section. The rate was chosen so that the post-FEC BER of the unequalized signal was significant (5×10^{-3}), which enabled to observe the differences between equalizers. The post-FEC BER is shown in Figure 6, as a function of Turbo-equalization iterations, using the same equalizers discussed above. It is evident that each equalizer converges to a different error-floor. As in the case of SNR gain estimation, the best performance is achieved by the 3-tap Kalman filter, where the BER was 10 times lower than that of the unequalized case.

For the reader's convenience, Table 1 summarizes the gain and cost of the various equalization schemes. The results shown are for the conditions studied at Sections 4.2 and 4.3.

Table 1. Summary of equalization algorithms performance. Note that the minimal BER for the unequalized case is 4.5×10^{-3} .

Title 1	Normalized Cost	SNR Gain	Minimal BER (with Turbo equ.)
1 tap LMS	1	0.2 dB	3.4×10^{-3}
1 tap RLS	0.88	0.19 dB	3.5×10^{-3}
1 tap Window averaging	1.88	0.45 dB	2.4×10^{-3}
3 tap Window averaging	5.5	0.9 dB	1×10^{-3}
1 tap Kalman filter	22	0.7 dB	1.4×10^{-3}
3 tap Kalman filter	45	1.6 dB	2.6×10^{-4}

Figure 6. BER as a function of the number of Turbo-equalization iterations, using different equalizer types. The curves represent results obtained using the Virtual Lab tool [8], whereas the dots correspond to split-step simulations. Note that the case of zero iterations (leftmost point in the graph) corresponds to the case of no adaptive equalization.

5. Conclusions

In this paper, we discussed the prospects of out-of-band NLIN mitigation using equalization. At the heart of this concept is the time-varying ISI model, and the many types of correlations that it entails. There are several points to consider when one attempts to choose an equalization algorithm; we have shown that the more complex algorithms (e.g., Kalman filtering) provides superior NLIN cancellation capability. However, this comes with a significant increase in computational cost. While designing an adaptive equalization algorithm, one must be aware of the statistical properties of NLIN in the system. For example, very short systems tend to be dominated by the 0-th order ISI term, corresponding to PPRN. They are also characterized by relatively short correlation times. In longer systems, higher order ISI terms become more significant, and the correlation lengths increase. Thus, a higher number of equalizer taps will be needed as the link length increases. In cases of polarization-multiplexed transmission, it is highly advantageous to treat the two polarization channels jointly. This allows both to compensate for cross-polarization effects, and to utilize the correlation between the NLIN components affecting the two polarizations.

Another important aspect that was discussed is the type of symbol estimation method that is used. Incorporating the decoder with the equalizer in a turbo-equalization (or similar) setup have been shown to significantly improve performance. Naturally, the choice of code-type and decoding method can significantly affect the equalizer's performance.

The equalization algorithms reviewed in this paper represent the main types discussed in the literature, although many variants of these algorithms exist. We have showed that significant gain can be achieved by using these equalization methods, and that adaptive equalization is a viable method for NLIN mitigation. Still, there is much work to be done on this subject, such as designing new and efficient algorithms tailored for NLIN mitigation, as well as real-time implementation of the current algorithms.

Author Contributions: O.G. is responsible for the gathering of materials, study of the various equalization schemes, the evaluation of their performance, and preparation of the original draft. M.F. contributed to the professional guidance and editing of the final version. M.S. provided professional guidance and academic supervision, and participated in the editing of the final version of the paper.

Funding: This work was supported by the Israel Science Foundation under Grant 1401/16.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASE	Amplified Spontaneous Emission
BER	Bit Error Rate
COI	Channel of Interest
DBP	Digital Back-Propagation
FEC	Forward Error Correction
IC	Interfering Channel
ISI	Inter-Symbol Interference
LDPC	Low density parity check
LMS	Least Mean Squares
MAP	Maximum a-posteriori
NLIN	Nonlinear Interference Noise
OFDM	Orthogonal Frequency Division Multiplexing
PPRN	Phase and Polarization Rotation Noise
RLS	Recursive Least Squares
SMF	Single mode fiber
SNR	Signal to Noise Ratio
SSFM	Split-step Fourier method
WDM	Wavelength Division Multiplexing

References

1. Ip, E.; Kahn, J.M. Compensation of dispersion and nonlinear impairments using digital backpropagation. *J. Lightw. Technol.* **2008**, *26*, 3416–3425. [\[CrossRef\]](#)
2. Marsella, D.; Secondini, M.; Forestieri, E. Maximum likelihood sequence detection for mitigating nonlinear effects. *J. Lightw. Technol.* **2014**, *32*, 908–916. [\[CrossRef\]](#)
3. Guimaraes, F.P.; Reis, J.D.; Teixeira, A.L.; Pinto, A.N. Mitigation of intra-channel nonlinearities using a frequency-domain Volterra series equalizer. *Opt. Express* **2012**, *20*, 1360–1369. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Poggiolini, P.; Jiang, Y. Recent Advances in the Modeling of the Impact of Nonlinear Fiber Propagation Effects on Uncompensated Coherent Transmission Systems. *J. Lightw. Technol.* **2017**, *35*, 458–480. [\[CrossRef\]](#)
5. Secondini, M.; Forestieri, E. Analytical fiber-optic channel model in the presence of cross-phase modulation. *Photonics Technol. Lett.* **2012**, *24*, 2016–2019. [\[CrossRef\]](#)
6. Dar, R.; Feder, M.; Mecozzi, A.; Shtaif, M. Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation. *J. Lightw. Technol.* **2015**, *33*, 1044–1053. [\[CrossRef\]](#)
7. Golani, O.; Feder, M.; Mecozzi, A.; Shtaif, M. Correlations and phase noise in NLIN- modelling and system implications. In Proceedings of the 2016 Optical Fiber Communications Conference and Exhibition (OFC) 2016, Anaheim, CA, USA, 20–24 March 2016. [\[CrossRef\]](#)
8. Golani, O.; Dar, R.; Feder, M.; Mecozzi, A.; Shtaif, M. Modeling the bit-error-rate performance of nonlinear fiber-optic systems. *J. Lightw. Technol.* **2016**, *34*, 3482–3489. [\[CrossRef\]](#)
9. Fehenberger, T.; Yankov, M.P.; Barletta, L.; Hanik, N. Compensation of XPM interference by blind tracking of the nonlinear phase in WDM systems with QAM input. In Proceedings of the 2015 IEEE European Conference on Optical Communication (ECOC), Valencia, Spain, 27 September–1 October 2015; pp. 1–3.
10. Yankov, M.P.; Fehenberger, T.; Barletta, L.; Hanik, N. Low-Complexity Tracking of Laser and Nonlinear Phase Noise in WDM Optical Fiber Systems. *J. Lightw. Technol.* **2015**, *33*, 4975–4984. [\[CrossRef\]](#)
11. Yankov, M.; Da Ros, F.; da Silva, E.P.; Fehenberger, T.; Barletta, L.; Zibar, D.; Oxenloewe, L.; Galili, M.; Forchhammer, S. Nonlinear Phase Noise Compensation in Experimental WDM Systems with 256QAM. *J. Lightw. Technol.* **2017**, *35*, 1438–1443. [\[CrossRef\]](#)
12. Zhang, S.; Li, X.; Kam, P.Y.; Yu, C.; Chen, J. Pilot-assisted decision-aided maximum-likelihood phase estimation in coherent optical phase-modulated systems with nonlinear phase noise. *Photonics Technol. Lett.* **2010**, *22*, 380–382. [\[CrossRef\]](#)
13. Li, L.; Tao, Z.; Liu, L.; Yan, W.; Oda, S.; Hoshida, T.; Rasmussen, J.C. Nonlinear polarization crosstalk canceller for dual-polarization digital coherent receivers. In Proceedings of the Optical Fiber Communication Conference, San Diego, CA, USA, 21–25 March 2010; Optical Society of America: Washington, DC, USA, 2010. [\[CrossRef\]](#)
14. Zibar, D.; Piels, M.; Jones, R.; Schaeffer, C.G. Machine learning techniques in optical communication. *J. Lightw. Technol.* **2016**, *34*, 1442–1452. [\[CrossRef\]](#)
15. Pan, C.; Bülow, H.; Idler, W.; Schmalen, L.; Kschischang, F.R. Optical Nonlinear Phase Noise Compensation for 9×32 -Gbaud PolDM-16 QAM Transmission Using a Code-Aided Expectation-Maximization Algorithm. *J. Lightw. Technol.* **2015**, *33*, 3679–3686. [\[CrossRef\]](#)
16. Ghazisaeidi, A.; Salsi, M.; Renaudier, J.; Bertran-Pardo, O.; Mardoyan, H.; Charlet, G. Performance analysis of decision-aided nonlinear cross-polarization mitigation algorithm. In Proceedings of the 2012 38th European Conference and Exhibition on Optical Communications, Amsterdam, The Netherlands, 16–20 September 2012. [\[CrossRef\]](#)
17. Secondini, M.; Forestieri, E. On XPM mitigation in WDM fiber-optic systems. *Photonics Technol. Lett.* **2014**, *26*, 2252–2255. [\[CrossRef\]](#)
18. Golani, O.; Feder, M.; Shtaif, M. Kalman-MLSE equalization of nonlinear noise. In Proceedings of the 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, USA, 19–23 March 2017; Optical Society of America: Washington, DC, USA, 2017. [\[CrossRef\]](#)
19. Golani, O.; Feder, M.; Shtaif, M. NLIN Mitigation using Turbo Equalization and an Extended Kalman Smoother. *J. Lightw. Technol.* **2019**. [\[CrossRef\]](#)

20. Cai, J.X.; Batshon, H.; Mazurczyk, M.; Sinkin, O.; Wang, D.; Paskov, M.; Patterson, W.; Davidson, C.; Corbett, P.; Wolter, G.; et al. 70.4 Tb/s capacity over 7,600 km in C+ L band using coded modulation with hybrid constellation shaping and nonlinearity compensation. In Proceedings of the 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, USA, 19–23 March 2017; Optical Society of America: Washington, DC, USA, 2017. [\[CrossRef\]](#)
21. Cai, J.X.; Batshon, H.G.; Mazurczyk, M.V.; Sinkin, O.V.; Wang, D.; Paskov, M.; Patterson, W.W.; Davidson, C.R.; Corbett, P.C.; Wolter, G.M.; et al. 70.46 Tb/s over 7,600 km and 71.65 Tb/s over 6,970 km transmission in c+ l band using coded modulation with hybrid constellation shaping and nonlinearity compensation. *J. Lightw. Technol.* **2018**, *36*, 114–121. [\[CrossRef\]](#)
22. Dar, R.; Feder, M.; Mecozzi, A.; Shtai, M. Time varying ISI model for nonlinear interference noise. In Proceedings of the Optical Fiber Communication Conference, San Francisco, CA, USA, 9–13 March 2014; Optical Society of America: Washington, DC, USA, 2014. [\[CrossRef\]](#)
23. Dar, R.; Feder, M.; Mecozzi, A.; Shtai, M. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Opt. Express* **2013**, *21*, 25685–25699. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Dar, R.; Feder, M.; Mecozzi, A.; Shtai, M. Pulse collision picture of inter-channel nonlinear interference in fiber-optic communications. *J. Lightw. Technol.* **2015**. [\[CrossRef\]](#)
25. Dar, R.; Feder, M.; Mecozzi, A.; Shtai, M. Accumulation of nonlinear interference noise in fiber-optic systems. *Opt. Express* **2014**, *22*, 14199–14211. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Barhumi, I.; Leus, G.; Moonen, M. Time-varying FIR equalization for doubly selective channels. *IEEE Trans. Wirel. Commun.* **2005**, *4*, 202–214. [\[CrossRef\]](#)
27. Choi, J.W.; Riedl, T.J.; Kim, K.; Singer, A.C.; Preisig, J.C. Adaptive linear turbo equalization over doubly selective channels. *IEEE J. Ocean. Eng.* **2011**, *36*, 473–489. [\[CrossRef\]](#)
28. Golani, O.; Elson, D.; Lavery, D.; Galdino, L.; Killey, R.; Bayvel, P.; Shtai, M. Experimental characterization of nonlinear interference noise as a process of intersymbol interference. *Opt. Lett.* **2018**, *43*, 1123–1126. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Jiang, Y.; Carena, A.; Poggiolini, P.; Forghieri, F. On the impact of non-linear phase-noise on the assessment of long-haul uncompensated coherent systems performance. In Proceedings of the 2014 European Conference on IEEE Optical Communication (ECOC), Cannes, France, 21–25 September 2014; pp. 1–3.
30. Fehenberger, T.; Mazur, M.; Eriksson, T.A.; Karlsson, M.; Hanik, N. Experimental analysis of correlations in the nonlinear phase noise in optical fiber systems. In Proceedings of the 42nd European Conference on Optical Communication ECOC 2016, Dusseldorf, Germany, 18–22 September 2016; pp. 1–3.
31. Poggiolini, P.; Carena, A.; Curri, V.; Bosco, G.; Forghieri, F. Analytical modeling of nonlinear propagation in uncompensated optical transmission links. *Photonics Technol. Lett.* **2011**, *23*, 742–744. [\[CrossRef\]](#)
32. Dar, R.; Winzer, P.J. Nonlinear interference mitigation: Methods and potential gain. *J. Lightw. Technol.* **2017**, *35*, 903–930. [\[CrossRef\]](#)
33. Piori, D.; Forghieri, F.; Bosco, G. Residual Non-Linear Phase Noise in Probabilistically Shaped 64-QAM Optical Links. In Proceedings of the 2018 IEEE Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, USA, 11–15 March 2018; pp. 1–3.
34. Golani, O.; Piori, D.; Bosco, G.; Shtai, M. Correlated Non-Linear Phase Noise in Multi-Subcarrier Systems. *arXiv* **2018**, arXiv:1812.07502.
35. Zhang, H.; Davidson, C.; Batshon, H.; Mazurczyk, M.; Bolshtyansky, M.A.; Foursa, D.; Pilipetskii, A. DP-16QAM based coded modulation transmission in C+ L band system at transoceanic distance. In Proceedings of the Optical Fiber Communication Conference, Anaheim, CA, USA, 20–24 March 2016; Optical Society of America: Washington, DC, USA, 2016. [\[CrossRef\]](#)
36. Madhow, U. *Fundamentals of Digital Communication*; Cambridge University Press: Cambridge, UK, 2008.
37. Golani, O.; Feder, M.; Shtai, M. Kalman-MLSE Equalization for NLIN Mitigation. *J. Lightw. Technol.* **2018**, *36*, 2541–2550. [\[CrossRef\]](#)
38. Djordjevic, I.B.; Minkov, L.L.; Batshon, H.G. Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPC-coded turbo equalization. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 73–83. [\[CrossRef\]](#)

39. Djordjevic, I.B.; Minkov, L.L.; Xu, L.; Wang, T. Suppression of fiber nonlinearities and PMD in coded-modulation schemes with coherent detection by using turbo equalization. *J. Opt. Commun. Netw.* **2009**, *1*, 555–564. [\[CrossRef\]](#)
40. da Silva, E.P.; Yankov, M.P.; Da Ros, F.; Morioka, T.; Oxenloewe, L.K. Perturbation-based FEC-assisted Iterative Nonlinearity Compensation for WDM Systems. *J. Lightw. Technol.* **2018**. [\[CrossRef\]](#)

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).