Signal Processing Techniques for Optical Transmission Based on Eigenvalue Communication

Jonas Koch, Ken Chan, Christian G. Schaeffer, Stephan Pachnicke, *Senior Member, IEEE*

(Invited Paper)

Abstract—A minimum mean squared error (MMSE) equalizer is a way to effectively increase transmission performance for nonlinear Fourier transform (NFT) based communication systems. Other equalization schemes, based on nonlinear equalizer approaches or neural networks, are interesting for NFT transmission due to their ability to deal with nonlinear correlations of the NFTs' eigenvalues and their coefficients. We experimentally investigated single- and dual-polarization long haul transmission with several modulation schemes and compared different equalization techniques including joint detection equalization and the use of neural networks. We observed that joint detection equalization provides range increases for shorter transmission distances while having low numeric complexity. We could further achieve bit error rates (BER) under HD-FEC for significant longer transmission distances in comparison to no equalization with different equalizers.

Index Terms—Nonlinear Fourier transform, eigenvalue communication, neural networks, nonlinear equalization.

I. INTRODUCTION

CURRENT long-haul optical fiber communication systems are impaired by Kerr nonlinearities, limiting the maximal data rates using available coherent detection schemes. A promising scheme to beneficially consider the nonlinear effects is the Nonlinear Fourier Transformation (NFT), where the mathematical foundations were established by Zakharov and Shabat [1]. The NFT linearizes a loss- and noiseless optical channel and enables the representation of a time domain signal in a continuous and a discrete spectrum in the nonlinear Fourier domain [2]. Vice versa the inverse NFT (INFT) can be used to generate a specific time domain signal at the transmitter since the aforementioned spectra can be combined to provide a high number of degrees of freedom for information encoding.

Since the nonlinearity of the fiber is an integral part of the NFT, its complexity is theoretically increasing less by e.g. extending the transmission distance or signal power. This is in contrast to several other extensively researched nonlinear compensation methods, such as the digital back-propagation [3] and Volterra equalizers [4,5]. For example, a digital back-propagation would be reduced to a one-step linear transmission function, if the NFT is used as shown by Wahls et al in [6].

Manuscript received August 31, 2020; revised November 2, 2020; Accepted xxx; Date of publication XXXX xx, 2020; Date of current version xxx. This work was partly supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program "Electronic-Photonic Integrated Systems for Ultrafast Signal Processing" by grant PA 1705/1-1.

J. Koch and S. Pachnicke are with Kiel University, Kiel, Germany (e-mail: Jonas.Koch@tf.uni-kiel.de; Stephan.Pachnicke@tf.uni-kiel.de).

K. Chan and C. G. Schaeffer are with Hemlut-Schmidt-University, Hamburg, Germany (e-mail: kchan@hsu-hh.de; cgs@hsu-hh.de).

Here, the NFT performs similarly well to a common digital back-propagation, which uses 4 km steps.

The use of both (continuous and discrete) spectra of the NFT has been experimentally demonstrated in several works, where also both polarizations have been utilized [7,8]. Such systems are commonly known as nonlinear frequency division multiplexing (NFDM) systems and allow transmission of e.g. OFDM signals and solitons together.

In this paper we use a subclass of the NFDM systems, proposed by Hasegawa and Nyu [9]. This approach uses only the discrete part of the nonlinear spectrum, which corresponds to solitons in time domain. Amplitude and phase information can be modulated onto the discrete spectrum $\tilde{q}(\lambda_k) = b(\lambda_k) \frac{da(\lambda_k)}{d\lambda} |_{\lambda_k}$, where λ_k is a specific frequency component or eigenvalue (EV), and a and b are the nonlinear Fourier coefficients (NFC). The Darboux Transformation shown in e.g. [10] for single- and [11] for dual-polarization is commonly used for soliton generation. This so-called eigenvalue communication has been experimentally demonstrated in single-polarization [12,13] and is crucial in order to use the full capacity of NFT-based transmission systems, even though its' individual transmission rates are not competitive to systems, which use the continuous spectrum or both spectra combined. By utilizing both polarizations the data rate can be theoretically doubled [14], which has been experimentally demonstrated in [15,16]. The performance can further be increased by using the differential coding scheme proposed by Span et al [17].

To further optimize the transmission performance, several techniques have been proposed and experimentally verified. It has been shown, that detecting only the b -coefficient for data modulation can increase the performance [18]–[21]. Furthermore, it has been observed, that using eigenvalues with real-parts leads to a position shift of the time domain signal in the symbol period while propagating in the fiber, which can be compensated by pre-shifting the soliton at the transmitter site [19]. It has been demonstrated, that the intentional use of this effect can increase transmission range by choosing suitable soliton parameters [22].

Additionally, it has been experimentally verified, that for a complex modulated $\tilde{q}(\lambda_k)$ a change of the received eigenvalue λ_k leads to partially deterministic phase and amplitude changes. This correlation can be used to equalize the received symbols [23]. This extends to correlations between eigenvalues in multi-eigenvalue transmissions, which can be exploited to maximize transmission performance [24]. Although $da(\lambda_k)/d\lambda|_{\lambda=\lambda_k}$ does not attribute to symbol demodulation, it is correlated with λ_k and $\tilde{q}(\lambda_k)$ and can therefore be used

to apply equalization schemes to further increase the detection performance, e.g. with a least minimum mean squared error (LMMSE) equalizer proposed in [18].

Since neural networks have gained increased interest for optical communication systems [25,26], they are also a promising candidate for equalizing NFT-based communication [27,28]. Eigenvalue communication systems, where the receiver is based on a neural network only, have been proposed in [29,30], which is a very effective alternative to the standard NFT approach.

In this paper we analyze different equalizer concepts to improve signal transmission with the NFT. It is organized as follows: Section II briefly introduces the NFT, and discusses the impact of noise. Section III contains the mathematical models of the used equalization schemes. After the experimental setup and parameters are presented, we discuss the performance of the different equalization schemes in Section IV. To summarize all results we then give a final conclusion in Section V.

II. NONLINEAR FOURIER TRANSFORM

In this section we briefly summarize the mathematical basics of the (inverse) Nonlinear Fourier Transform for both single-polarization (SP) and dual-polarization (DP) cases. As this paper focuses on experimental soliton transmissions, the reader is referred to [2,11,14], which contain a deeper analysis and derivation of the mathematical background.

A. Nonlinear Fourier Transform

The Nonlinear Fourier Transform (NFT) solves integrable, nonlinear and partial differential equations like the nonlinear Schrödinger equation (NLSE). The evolution of an optical pulse through an ideal loss- and noiseless single-mode fiber (SMF) is characterized by two coupled NLSEs, known as a Manakov system (MS). We want to note, that the modeling can also be done with two independent NLSEs for each polarizations [31]. In eigenvalue communication the MS is commonly used as model, and for convenience we introduce the normalized MS [17]

$$j \frac{\partial \mathbf{q}(t, z)}{\partial z} = \frac{\partial^2 \mathbf{q}(t, z)}{\partial t^2} + 2 \|\mathbf{q}(t, z)\| \mathbf{q}(t, z) \quad (1)$$

with $\mathbf{q}(t, z) = (q_1(t, z), q_2(t, z))^T$ as signal vector with the pulse's components in x- and y-polarization, respectively. The index $p = 1, 2$ indicates the polarization in the following equations. This notation will later be used for the variables of the NFT as well. The scaling of the physical units is given by

$$Q(T, Z) = \sqrt{P_0} q_p\left(\frac{T}{T_s}, \frac{Z}{Z_s}\right) \\ \text{with } P_0 T_s^2 = \frac{|\beta_2|}{8\gamma}, \quad L_0 = \frac{2T_0^2}{|\beta_2|}, \quad p = 1, 2,$$

where P_0 is the physical power, β_2 the dispersion, γ the nonlinear coefficient, T the retarded time [2], Z the position in the fiber and T_s the pulse duration, which will also be denoting the symbol duration in the following sections. Let

the transmitter be located at $Z = 0$ and the receiver at $Z = Z$. Note that we consider $\beta_2 < 0$ in this paper, since we operate in the anomalous dispersion regime in SMF.

The so called extended Zakharov-Shabat (ZS) system represents a closed form solution of (1) in the nonlinear spectrum with [17]

$$\frac{\partial}{\partial t} \mathbf{v} = \Lambda \mathbf{v} \quad (2) \\ \mathbf{v} = \begin{pmatrix} v_1(\lambda; t, z) \\ v_2(\lambda; t, z) \\ v_3(\lambda; t, z) \end{pmatrix}, \Lambda = \begin{pmatrix} -j\lambda & q_1(t, z) & q_2(t, z) \\ q_1^*(t, z) & j\lambda & 0 \\ q_2^*(t, z) & 0 & j\lambda \end{pmatrix},$$

where $*$ indicates the complex conjugation, v is a solution and λ is a spectral parameter or eigenvalue. With the vanishing boundary condition of the signal $\mathbf{q}(t, z) \rightarrow 0, t \rightarrow \infty$ we can find a set of canonical solutions [2,14]. Using the solution

$$\boldsymbol{\vartheta} \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} e^{-j\lambda t} \quad \text{for } t \rightarrow -\infty \quad (3)$$

we obtain the Jost or nonlinear Fourier coefficients (NFC)

$$a(\lambda; z) = \lim_{t \rightarrow \infty} v_1 e^{+j\lambda t} \quad (4a)$$

$$b_1(\lambda; z) = \lim_{t \rightarrow \infty} v_2 e^{-j\lambda t} \quad (4b)$$

$$b_2(\lambda; z) = \lim_{t \rightarrow \infty} v_3 e^{-j\lambda t}. \quad (4c)$$

When an optical pulse propagates through an ideal fiber in z , the NFCs are transformed as

$$a(\lambda; z_0 + z) = a(\lambda; z_0) = a(\lambda) \quad (5)$$

$$b_p(\lambda; z_0 + z) = b_p(\lambda; z_0) e^{-4j\lambda^2 z}, \quad (6)$$

where z_0 is the position of the transmitter, where the pulse is launched. (5) shows two important properties: for an ideal transmission the a -coefficient is invariant in z , leading to invariance in z of the eigenvalue λ as well. Therefore we can drop the variable z for the rest of the paper. The second property is that the b -coefficients are transformed linearly.

The NFT represents a signal $q(t)$ with its continuous and discrete spectrum and the corresponding spectral amplitudes via $a(\lambda)$ and $b(\lambda)$. The continuous spectrum lies in $\mathbb{R}$ and the discrete spectrum only contains spectral amplitudes for specific points called discrete eigenvalues λ_k , which are zeros of $a(\lambda)$ in $\mathbb{C}^+$. Let

$$\hat{q}_{1,2}(\lambda) = \frac{b_{1,2}(\lambda)}{a(\lambda)}, \quad \lambda \in \mathbb{R} \quad (7)$$

represent the continuous spectral amplitude and

$$\tilde{q}_{1,2}(\lambda_k) = \frac{b_{1,2}(\lambda_k)}{da(\lambda_k)/d\lambda|_{\lambda=\lambda_k}}, \quad a(\lambda_k) = 0, \lambda_k \in \mathbb{C}^+, k \in \mathbb{N}, \quad (8)$$

be the discrete spectral amplitude. Note that one can obtain the SP case by setting $q_2(t) = 0$, which is equal to dropping the third dimension in the equations and therefore obtaining only one b -coefficient.

B. Inverse Nonlinear Fourier Transform

The NFT can be computed analytically for many cases without continuous spectrum and also for simple waveforms like a rectangular pulse [2]. One can use numeric methods for the inverse nonlinear Fourier transform (INFT) to construct the time domain signal from a given NFT spectrum. In this section we shortly introduce the convenient and well-developed Darboux transform (DT), which is a suitable algorithm, if the information is only encoded in the discrete spectrum $\tilde{q}$. In this case, the resulting time domain signal $q(t)$ is a soliton. In this section we summarize the DT for both single- and dual-polarization cases, which are mainly based on the derivations in [10,11].

The DT provides the possibility to construct the time domain signal $q(t)$ by iteratively adding discrete eigenvalues λ_k to the NFT spectrum, while the corresponding $q(t)$ is updated in time domain. Let $(\lambda_k, \tilde{q}(\lambda_k))$ with $k = 1, \dots, L$ be the discrete spectrum of an L -th order soliton.

According to (2), v is a solution of the ZS system for a signal $q(t)$ and a specific discrete eigenvalue. A new solution $v^{(k)}$ is given by [11]

$$v^{(k)} = (\lambda_i I_3 - G)v^{(k-1)}, \quad (9)$$

where I_3 is $\text{diag}(1, 1, 1)$ and $G = \Theta M \Theta^{-1}$, with

$$\Theta = \begin{pmatrix} \bar{v}_1^{(k-1)} & \bar{v}_2^{(k-1)*} & \bar{v}_3^{(k-1)*} \\ \bar{v}_2^{(k-1)} & -\bar{v}_1^{(k-1)*} & 0 \\ \bar{v}_3^{(k-1)} & 0 & -\bar{v}_1^{(k-1)*} \end{pmatrix}, \quad (10)$$

where $\bar{v} = (\bar{v}_1, \bar{v}_2, \bar{v}_3)^T$ is a so called auxiliary solution of (2) for a seed signal $q_k^{(k)}$ and a discrete eigenvalue λ_i and $M = \text{diag}(\lambda_i, \lambda_i^*, \lambda_k^*)$ [14]. We start from the initial solution $q_k^{(0)}(t) = 0$ and generate an updated signal $q_k^{(k)}(t)$ by adding a new λ_i to the NFT spectrum at each iteration with

$$q_k^{(k)}(t) = q_k^{(k-1)}(t) + 2j(\lambda_i^* - \lambda_k) \frac{u_k^{(k-1)*}}{1 + \sum_{s=1}^2 |u_s^{(k-1)}|^2}, \quad (11)$$

$$\text{with } u_k = \frac{\bar{v}_{k+1}}{\bar{v}_1}$$

to obtain a time domain DP signal with a given NFT spectrum with an arbitrary number of discrete eigenvalues. The initial auxiliary solution $\bar{v}^{(0)}$ [14]

$$\bar{v}^{(0)} = \begin{pmatrix} 1 \cdot e^{-j\lambda_k t} \\ -b_1(\lambda_i)^{(k)} e^{-j\lambda_k t} \\ -b_2(\lambda_i)^{(k)} e^{-j\lambda_k t} \end{pmatrix}, \quad (12)$$

is updated as in (9). Based on that the time domain signal as in (11) is constructed until all L discrete eigenvalues are added to the spectrum. For the SP case one can drop the third dimension, similar to the NFT case. One can find the detailed algorithms and derivations for the SP and DP cases in [10] and [14], respectively. Since λ_k and $b_p(\lambda_k)$ are free parameters, we can use them for data modulation. $b_p(\lambda_i)$ represents the amplitude and phase for a desired data symbol, e.g. from a PSK-constellation for the corresponding λ_k . Note, that according to (12), $da(\lambda_k)/d\lambda$ is not used for encoding information in the DT.

C. Noise Impact

The NFT defined above is tailored to a system without any disturbances such as noise or low-pass characteristics. Disturbed systems show deviations of the eigenvalues on the complex half-plane, which can be partly deterministic and partly stochastic. These deviations in turn lead to different propagation properties and therefore to changes of the NFT coefficients. Hence, correlations between deviations of the NFT coefficients and the NFT eigenvalues can be observed [17]–[19,27,32]. Additionally, the diagonalization property of the NFT can be lost. This in turn leads to cross-correlations between different spectral components of the same symbol.

As an example the received eigenvalues, the received NFT coefficients $b(\lambda_k)$ and their respective deviations from the transmitted values are plotted in Fig. 1 after 905 km for 4ASK-8PSK, single eigenvalue transmission in one polarization. The experimental setup is described in Section IV-A. In Fig. 1c the strong dependencies between the deviations of the received amplitude of $b(\lambda_k)$ ($\Delta A(b(\lambda_k))$) from their respective transmitted values and the deviations of the imaginary value of λ_k can be seen. Furthermore, the sign of the correlation changes, depending on the transmitted symbols' amplitude. Fig. 1d shows a weak dependency when compared to Fig. 1c between the deviations of the phase and deviations of the eigenvalues' imaginary value.

The reason for the correlations shown above are two-fold. One reason lies in the calculations of the (I)NFT and another in fiber transmission effects i.e. influence of optical noise. In the case of a single eigenvalue transmission, the amplitude of $\tilde{q}(\lambda_k)$ is depending on the position of the soliton in the retarded time window and the eigenvalues' imaginary value, according to [19]

$$|\tilde{q}(\lambda_k)| = 2\text{Im}(\lambda_k)e^{2t_0\text{Im}(\lambda_k)}. \quad (13)$$

Here, t_0 is the time shift of the soliton in normalized units. Assuming the time center to be correct, deviations of $|q(\lambda_k)|$ (and hence $b(\lambda_k)$) are only due to deviations of $\text{Im}(\lambda_k)$, owing to noise or other effects, such as narrow filters.

Due to dispersion, the time center of the soliton moves, if the eigenvalue is shifted in real part. This is because the real-part is directly proportional to the center frequency of the soliton and hence the effect of group-velocity dispersion changes. Additionally, a change in imaginary part of the solitons leads to changes of the peak amplitudes of the solitons. This in turn leads to different phase deviations, caused by the Kerr nonlinearities.

To calculate the linear correlation between the deviations, the Pearson correlation can be used according to

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}. \quad (14)$$

Fig. 2 shows the amount of absolute correlation between eigenvalue and NFT-coefficient deviations depending on the transmission distance. Fig. 2a shows differences between the simulated and experimental results. This is due to low-pass characteristics of devices in the setup. Narrow filtering of solitons leads to decreasing average imaginary values of their eigenvalues, with only a small amount of eigenvalues higher

Fig. 1. a): Received eigenvalues after 905 km experimental single-polarization transmission (blue) and transmitted eigenvalue position (red). b): Constellation of received $b(\lambda)$ in polar coordinates. c): Deviation of the amplitude of $b(\lambda)$ over the deviation of the eigenvalues' imaginary part (color coded to the transmitted amplitude to show different correlations depending on the transmitted symbol) d): Deviation of the phase of $b(\lambda)$ over the deviation of the eigenvalues' imaginary part. The colors in Figs b) and c) are coded to the transmitted amplitudes of $b(\lambda)$ ($\exp(-1.5)$ (blue), $\exp(-0.5)$ (red), $\exp(0.5)$ (yellow) and $\exp(1.5)$ (purple)). Details of the experiment can be found in Table II.

Fig. 2. Pearson correlations between $\Delta\lambda$ and $\Delta b(\lambda)$ during transmission. Black lines refer to simulations and blue X to experimental results.

than the transmitted eigenvalue due to noise contribution. This can also be seen in Fig. 1a. Together with (13) this will result in decreasing amplitudes of $b(\lambda_k)$, if the average absolute of $a'(\lambda_k)$ does not increase proportionally to the eigenvalues. However, the amount of this correlation decreases during transmission. This is because the centre time of the solitons is further changing during transmission. One reason for this are deviations of the eigenvalues' real-parts due to noise leading to timing-jitters as described by the Gordon-Haus effect [33]. As a result, increasing correlations between $\text{Re}(\Delta\lambda)$ and the deviation of the amplitude of $b(\lambda_k)$, $\Delta A(b(\lambda))$, can be seen in Fig. 2c. Fig. 2b shows the increasing amount of correlation of the deviations of the eigenvalues' imaginary parts and the phase of $b(\lambda)$. This is a result of different impacts of self phase modulation on the peak amplitude of the soliton, if $\text{Im}(\lambda_k)$ changes. Mind, that the change of $\phi(\lambda)$ also depends on changes of the real part of λ , as described later in eq. (17). However, this correlation is comparatively small, when compared to the correlation between the phase of b and the imaginary value of λ . This has also been observed in e.g. [18].

III. SIGNAL PROCESSING TECHNIQUES

A. Joint Detection Equalization

Traditional joint detection (JD) equalization is based on parallel data transmission, where detecting all channels jointly

allows estimating and compensating crosstalk between channels [34]. We propose an equalization for single eigenvalue transmissions with both amplitude and phase modulation, where the detection performance of the phase information can be increased with the help of the information from the amplitude detection. The influence of noise on solitons during its propagation through a fiber has been investigated and modeled in several works [35]–[37]. For the proposed JD equalization we use a simplified channel model with complex additive noise, since the goal is to compensate for a basic relationship as shown later in this section. We assume ideal soliton transmission and no signal-noise interaction along the fiber in a SP [12]. The proposed scheme and derivations are based on [12]. Let $q(t)$ be a fundamental soliton with its discrete eigenvalues λ_k and the NFC $b(\lambda_k)$. The phase of $b(\lambda_k)$ corresponds to the phase of $q(t)$, while the amplitude $|b(\lambda_k)|$ corresponds to the position of $q(t)$ in a symbol period T_s . This means that a time-shifted soliton $q(t - t_0)$, where t_0 denotes the time shift in soliton units, corresponds to a scaled coefficient $b(\lambda_k, t_0) = b(\lambda_k, 0) \exp(-j2\lambda_k t_0)$, which can also be seen from the time shift theorem of the linear Fourier transform [2]. Assuming a purely imaginary eigenvalue λ_k , the amplitude modulation of $b(\lambda_k)$ is equal to a pulse-position modulation of the corresponding soliton. We want to simplify the analysis to the following assumption: The NFT of a noisy soliton results in a disturbed eigenvalue λ_{kN} and b -coefficient

$b_N(\lambda_k)$ as in

$$\lambda_{kN} = \lambda_k + n_{\lambda i} + jn_{\lambda q} \quad (15)$$

$$b_N(\lambda_k) = b(\lambda_k) + n_{bi} + jn_{bq}, \quad (16)$$

where $n_{\lambda i}$, $n_{\lambda q}$, n_{bi} and n_{bq} denote real-valued noise without any specific assumptions regarding statistical properties.

We now derive the impact of a time shift of the soliton (i.e. amplitude modulation of $b(\lambda_k)$) on the noise, which affects the received $b_N(\lambda_k, t_0)$. Let the noise values n_{bi} and n_{bq} denote the disturbance on $b_N(\lambda_k, 0)$, i.e. on a soliton centered within the symbol period T_s . Using the time shift theorem, we find

$$\begin{aligned} b_N(\lambda_k, t_0) &= b_N(\lambda_k, 0) \exp(-j2\lambda_{kN}t_0) \\ &= (b(\lambda_k, 0) + n_{bi} + jn_{bq}) \\ &\quad \times \exp(-j2(\lambda_k + n_{\lambda i} + jn_{\lambda q})t_0) \\ &= b(\lambda_k, t_0) \left(1 + \frac{n_{bi} + jn_{bq}}{b(\lambda_k, 0)}\right) \\ &\quad \times \exp(-j2n_{\lambda i}t_0) \exp(2n_{\lambda q}t_0). \end{aligned} \quad (17)$$

Equation (17) shows that $b(\lambda_k, t_0)$ suffers from amplitude and phase noise that increases with t_0 and vanishes for $t_0 = 0$. Moreover, assuming that we know $n_{\lambda i}$, $n_{\lambda q}$ and t_0 , we can easily compensate this noise. The strategy is to first compute $n_{\lambda q}$ and $n_{\lambda i}$ on the received λ_{kN} . With the expressions in (17) we can use $n_{\lambda q}$ to detect the likeliest transmitted $|b_{tx}(\lambda_k)|$. Based on that decision we can improve phase detection by using $n_{\lambda i}$ on the $\angle b_{tx}(\lambda_k)$.

B. Advanced Equalization Techniques

More advanced equalizers can be trained to use the correlations explained in Section II-C. The minimum mean squared error equalizer (MMSE) can use linear correlations between $\Delta\lambda$ and $\Delta b(\lambda)$ [18,39]. This equalizer can be extended to take nonlinear impacts into account (such as shown in (13)), using similarities to the nonlinear Volterra theory [38]. Another way to equalize nonlinearly are neural networks. Neural networks can be used either on the NFT eigenvalues and their coefficients [27] or the sampled pulses directly [29].

C. Minimum Mean Squared Error Equalizer

The approach to equalize NFT signals using an MMSE were introduced in [18,39]. Using $\mathbf{n} = [\Delta a'_R, \Delta a'_I, \Delta\lambda_R, \Delta\lambda_I]$ the errors $\Delta A(b)$ and $\Delta\phi(b)$ of $b(\lambda_k)$ can be reduced. Keep in mind, that in addition to the deviations of λ_k ($\Delta\lambda$) the deviations of the NFT coefficient $a'(\lambda_k)$ (Δa) can be used, if only $b(\lambda_k)$ is being used for data modulation. In this case $a'(\lambda_k)$ can be set to a fixed value. All equalizers used in this paper are real valued, hence all values are split into imaginary and real values, as denoted by the subscripts. With $\mathbf{n}$, vectors $\mathbf{c}, \mathbf{d}$ can be calculated using training symbols to reduce the amplitude and phase deviations:

$$\arg\min_{\mathbf{c}} \mathbf{E}[(\Delta A(b) - \mathbf{c}^T \mathbf{n})^2] \quad (18)$$

$$\arg\min_{\mathbf{d}} \mathbf{E}[(\Delta\phi(b) - \mathbf{d}^T \mathbf{n})^2]. \quad (19)$$

Here $\mathbf{E}$ denotes the expectation value [18]. $\mathbf{c}$ and $\mathbf{d}$ can be calculated using

$$\mathbf{c}^T = \mathbf{E}[\Delta A(b) \mathbf{n}^T] \cdot \text{cov}(\mathbf{n})^{-1} \quad (20)$$

$$\mathbf{d}^T = \mathbf{E}[\Delta\phi(b) \mathbf{n}^T] \cdot \text{cov}(\mathbf{n})^{-1}. \quad (21)$$

The MMSE can generally be expanded to multi-eigenvalue problems [18,40]. One advantage of the linear MMSE equalizer is the simple structure, since only four multiplications are needed after training for an online equalization.

D. Nonlinear Equalizer

To account for nonlinear impacts of the deviations summarized in $\mathbf{n}$, e.g. caused by the nonlinear term in Equation (13), the linear MMSE equalizer as described above gets nonlinearly expanded. Hence, $\mathbf{n}$ gets extended depending on the order of nonlinearity which is to be used. Again, two vectors $\mathbf{c}_{NL}$, $\mathbf{d}_{NL}$ are being calculated as stated in (21) and (20), but with modified $\mathbf{n}$. For this, two vectors $\mathbf{n}_A$ and $\mathbf{n}_\phi$ according to

$$\mathbf{n}_A = [|b(\lambda_k)|, \Delta a'_R, \Delta a'_I, \Delta\lambda_R, \Delta\lambda_I] \quad (22)$$

$$\mathbf{n}_\phi = [\angle b(\lambda_k), \Delta a'_R, \Delta a'_I, \Delta\lambda_R, \Delta\lambda_I] \quad (23)$$

are being used for initialization. Here, $\mathbf{n}_A$ is used for equalization of the amplitude, and $\mathbf{n}_\phi$ is used for equalization of the phase. Mind, that in the case of nonlinear equalization the vectors of origin $\mathbf{n}_A$ and $\mathbf{n}_\phi$ also include the amplitude and the phase of $b(\lambda_k)$ respectively. If a multi-soliton is being transmitted, these vectors can be expanded by $\Delta a'$ and $\Delta\lambda$ of other eigenvalues. The initialised vectors are then expanded according to Fig. 3 to create nonlinear extensions to $\mathbf{n}_A$ and $\mathbf{n}_\phi$. Using the nonlinearly extended vectors $\mathbf{n}_{A,NL}$ and $\mathbf{n}_{\phi,NL}$ in conjunction with (20), (21) $\mathbf{c}_{NL}$ and $\mathbf{d}_{NL}$ can be calculated, which can be used for nonlinear equalization. The amount of total coefficients to be computed for each $\mathbf{n}_{NL}$ can be calculated using

$$N_{\text{coeffs}}(O_{NL}) = \sum_{m=1}^{O_{NL}} \frac{1}{m!} \prod_{n=0}^{m-1} (L+n), \quad (24)$$

where O_{NL} is the order of nonlinearity to be used, and L refers to the length of the vectors $\mathbf{n}_A$ or $\mathbf{n}_\phi$. If this is compared to conventional Volterra filters for nonlinear time-series equalization the amount of coefficients can be kept small, since the length L is fixed [38]. For example, for a single eigenvalue transmission a second order equalizer requires 20 coefficients. Using a third order equalizer increases the amount of coefficients to 55. The amount of used coefficients then directly converts to the amount of needed multiplications for the equalization.

E. Neural Network Equalizer

Neural network (NN) equalizers have been shown to work within simulations for compensating the impact of eigenvalue deviations on the received NFT coefficients [27]. Using a feed forward NN complex nonlinear mappings can be approximated from the input data. For single eigenvalue transmission the input data consists of $\mathbf{n}_{NN} =$

Fig. 3. Schematic depiction of $\mathbf{n}_{A,NL}$ with nonlinearities of third order. Mind that for $\mathbf{n}_{\phi,NL}$ $|b(\lambda)|$ is replaced by the phase $\angle b(\lambda)$ of b . The first order stands for the initialization vector from (22). The second order is color coded to show the second order mixing terms of $\mathbf{n}_A$ and $|b(\lambda)|$ (black), $\Delta\lambda_I$ (blue), $\Delta\lambda_R$ (yellow), $\Delta a'_I$ (purple) and $\Delta a'_R$ (orange), respectively. Additionally the first parts of third order nonlinearity mixing terms are shown (dashed lines).

$[\log(|b(\lambda_k)|) \angle(b(\lambda_k)) \Delta a'_R \Delta a'_I \Delta\lambda_R \Delta\lambda_I]$. These inputs refer to the data carrying NFT coefficient $b(\lambda_k)$ and the known deviations of the NFT-coefficient $a'(\lambda_k)$ and the eigenvalue λ_k , which are correlated to the unknown deviations of $b(\lambda_k)$. The NN is able to equalize the amplitude and phase deviations jointly. This can be expanded for multi-eigenvalue or DP transmissions.

An NN consists of multiple layers, which are built out of nodes. These are connected to all nodes of the following layer by so-called edges. Each node represents an arbitrary activation function, which calculates an output depending on a weighted sum of the outputs of all previous layers' nodes. The parameters of the NN to be optimized are the weights of the edges $\mathbf{W} = w_{jp}$, which indicate the weight from the output of node p to node j . The number of input and output layer nodes are fixed to the length of $\mathbf{n}_{NN}$, and to 2^{N_b} , where N_b stands for the number of bits per transmitted soliton, respectively. Thus, each of the output nodes stands for one of the possibly transmitted symbols. The size and number of the hidden layers are arbitrary and can be chosen heuristically or by certain algorithms such as grid-search or Bayesian hyper parameter optimization [41].

In addition to the layers described above, drop-out layers can be used during the training of the neural network. If a drop-out layer is employed a fixed amount of randomly chosen edge weights of one layer are set to zero. This way an emphasis is set to the weights, which are not set to zero. This can lead to more robust training results and can also reduce the amount of strict memorization of the problem to be solved.

The activation function used in the input and hidden layers is the rectified linear unit (ReLU), which can be mathematically described by $f(x) = \max(0, x)$. The activation function of the output layer is the softmax function according to [41]

$$g(x_j) = \frac{\exp(x_j)}{\sum_{m=1}^M \exp(x_m)}, \quad (25)$$

where M is the amount of output nodes and j the index of the output. This leads to values between zero and one at the output layer, where the highest value refers to the predicted transmitted symbol. This way the NN not only equalizes $b(\lambda_k)$, but also accomplishes the task of classification (demodulation of $b_k(\lambda_k)$). In addition, if the highest value of the output layer is relatively low, an uncertainty is indicated, which could be

Fig. 4. Net layout with the input layer (left), one hidden layer and the output layer (right) represented as nodes and additional bias nodes. The grey box highlights the known parameters of the transmitter.

used for e.g. improved decoding of the bits.

To train the NNs a gradient descend algorithm has been used, where the gradient was optimized by an Adam algorithm [41]. During training, the cross-entropy loss [41] between the desired output vector $\mathbf{y}$ and the estimated vector $\hat{\mathbf{y}}$ is minimized as depicted in Fig. 4.

In summary the output of an NN with one hidden layer can be expressed as

$$\hat{y}_j(\mathbf{n}_{NN}, \mathbf{W}) = g \left(\sum_{k=1}^K w_{jk} \cdot f \left(\sum_{p=1}^P w_{kp} n_{NN,p} + w_{k0} \right) + w_{j0} \right). \quad (26)$$

Here P is the size of the input vector $\mathbf{n}_{NN}$, and K the size of the hidden layer. $n_{NN,p}$ is the p -th element of $\mathbf{n}_{NN}$ and w_{kp} is the weight of the edge from node p of the input layer to node k of the hidden layer, being stored in $\mathbf{W}$. The index 0 of the weights refers to the weight of an additional fixed bias. The computational complexity of a NN for equalization after the training period is dependent on the amount of nodes and the used activation functions. Since the ReLU activation function used in this work is linear only one multiplier and one comparator are needed for clipping. The total amount of

Fig. 5. Experimental loop setup for SP. An orange lines indicate SMF fibers, a blue line indicates polarization-maintaining fiber and a black line indicates RF cables or connection to devices of DSP. The lower power output of both couplers are shown in percent at their lines, respectively. The polarization scrambler in the dash lined box is only used for the DP transmissions.

multiplications required for online equalization is then

$$N_{\text{mult}} = \sum_{l=1}^{L-1} k_l k_{l+1} + \sum_{l=2}^{L-1} k_l, \quad (27)$$

where L is the number of layers and k_l is the amount of nodes in layer l .

IV. EXPERIMENTAL VERIFICATION

In this section the impact of different detection and equalization techniques on single-eigenvalue transmission and DP multi-eigenvalue transmissions as described in Table II is examined.

A. Experimental Setup

This section features the experimental setup and the chosen parameters, which are based on the experiments conducted in [12,16]. The theoretical requirements of a lossless and noiseless link as introduced in Section II-A for the NFT are not fulfilled in reality. To take non-ideal link properties into account we use the lossless path averaged (LPA) method and determine the theoretical optical input power for the soliton pulses [42]. The experimental setup for the SP case is shown in Fig. 5. The digital signal processing (DSP) is offline for both transmitter and receiver. For the Tx signal the soliton signals are generated via Darboux Transform [10,17] and then denormalized into physical units as shown in Section II-A. We also precompensate for the frequency and amplitude responses of the DAC and modulator. We use the differential precoding scheme for the DP case proposed in [17] to increase transmission performance. The main parameters for the experimental setup can be seen in Table I. In the SP case, we use 1 GBd symbol rate, whereas in the DP case we use 1 or 2 GBd, resulting in different T_s and t for each case. Choosing sufficient samples and taking system parameters, e.g. analog bandwidth of the DAC/ADC or linewidth of the carrier laser into consideration is necessary to optimize transmission performance but also limits the symbol rate [43]. The signal contains 2978 or 5957 symbols for 1 or 2 GBd respectively, where the last four symbols are used to recover the sequence at the receiver DSP. In the DP case, we additionally replace the

first 256 symbols with a reference sequence containing fundamental solitons to recover the polarization. We compensate the frequency response of the transmitter and the coherent receiver. The Tx-DAC feeds the I/Q-modulator, which imprints the data on a laser source. The coherent receiver obtains a local oscillator signal from a 10 dB tap from the transmitter laser source. In the SP case we use a variable optical attenuator (VOA) in front of the pre-amplifier EDFA to set the OSNR. After the EDFA we use an amplified spontaneous noise (ASE) filter and another VOA to adjust the power to the theoretical value for the generated soliton pulses. In the SP case, the signal is then split in a 20 dB coupler to measure the OSNR with an optical spectrum analyzer (OSA). The loop consists of two 50.3 km non-zero dispersion-shifted fibers (NZDSF) with counterpropagating Raman amplification, whose power profile is shown in the lower right-hand inset in Fig. 5. To compensate insertion loss (IL) of the first Raman amplifier the pumping in the first span is slightly higher than in the second span. Using three acousto-optic modulators (AOMs) has the advantage of compensating the initial frequency shift regardless of the round trip number. For the DP transmissions we use a polarization scrambler to randomize the polarization, which is indicated as dashed lined box in Fig. 5. To control the received power and to reduce noise, a VOA and an optical bandpass filter (OBF) are used right before a polarization controller (PC). The signal is then detected by the coherent receiver, converted to a digital signal by an analog-digital-converter (ADC) and then processed with offline DSP by applying phase recovery, normalization to soliton units, NFT and demodulation of the data symbols. To compute the NFT and detect the b -coefficients the forward-backward (FB) method proposed in [10,14] is used. Note that the sample rate at the receiver side is reduced from 80 to 40 GSa/s for the DP case, since all four channels of the scope are used. We further limit the bandwidth in the DSP to the theoretical spectrum of the generated solitons. The parameters of the NFT are shown in Table II.

B. Single Eigenvalue Transmission

In this section a single eigenvalue soliton was transmitted using one polarization, according to Table II.

Table I
PARAMETERS OF THE EXPERIMENTAL SETUP.

DAC sample rate	88 GSa/s
ADC sample rate	80 GSa/s or 40 GSa/s
Baud rate	1 GBd or 2 GBd
Carrier laser wavelength λ_c	1550.12 nm
Carrier laser linewidth $\delta\nu$	< 1 kHz
Symbol period T_s	1 ns or 0.5 ns
Normalized time $t = T_s/11$	91 ps or 45.45 ps
Span length	50.3 km
Fiber loop length	100.6 km
Effective nonlinearity γ_{eff}	0.96 W ⁻¹ km ⁻¹
Fiber attenuation α	0.2 dB/km
Chromatic dispersion β_2	-5.75 ps ² /km

Table II
PARAMETERS OF THE NFT.

	Single-polarization	Dual-polarization
Eigenvalues λ_k	0.5j	$\mp 0.15 + 0.3j$
Modulation format	4ASK-8PSK	QPSK-QPSK
Amplitudes $ b(\lambda_k) $	$\exp([-1.5 \ -0.5 \ 0.5 \ 1.5])$	[0.5 2.5]
LPA Launch power P_{in}	-3.85 dBm	-5 dBm

1) *Joint detection equalization:* Note that this section is based on the results from [12]. To verify the theoretical relationship between t_0 with noise shown in (17) we simulate the transmission of 2-ASK modulated fundamental solitons with $t_0 \in \{0, 1\}$ and $\lambda_k = 0.5j$. For the experimental verification the modulation scheme is changed to 2ASK-BPSK because a DC-free electrical driving signal is needed to avoid undesired modulator behavior. Fig. 6 shows the correlation between $\ln|b_N(\lambda_k, t_0)|$ and $\text{Re}\{\lambda_{kN}\}$ and $\text{Im}\{\lambda_{kN}\}$ for back-to-back (B2B) transmissions. The upper plots show the centered soliton pulses with $t_0 = 0$, where $\ln|b(\lambda_k)|$ is scattered around 0, while the lower plots show the case for $t_0 = 1$. In Fig. 6 d) one can see a correlation between $\text{Im}\{\lambda_k\}$ and $\ln|b(\lambda_k)|$, which confirms the relationship between b_N and the exponential term including $n_{\lambda_i} t_0$ as shown in (17). Note that Fig. 6 b) also shows a small correlation. This is caused by a slight shift of T_s at the receiver, so that the soliton is not centered in the symbol period anymore. Fig. 7 shows the standard deviation σ for $\ln|b_N|$ and $\angle b_N$ for $t_0 \in \{-1.5, -0.5, 0, 0.5, 1.5\}$. Without equalization, σ increases with higher t_0 , whereas equalization reduces σ to the level of a centered soliton, eliminating the noise caused by the NFT. As discussed before, the offset of T_s leads to a shift of the minimum of σ . Note, that so far t_0 is assumed to be known, while in the following transmission experiments t_0 is detected for each symbol with a maximum likelihood decision.

For BER analysis we use a fundamental soliton with $\lambda_k = 0.5j$ and a 4-ASK-8-PSK modulation scheme with $t_0 \in \{-1.5, -0.5, 0.5, 1.5\}$ since the two schemes individually have the same order of magnitude of the BER. We obtain a total of five bits per symbol for data transmission. The constellation mapping and absolute pulse shapes are shown in the upper right-hand inset in Fig. 5. To verify the simulated performance gain after JD equalization, we vary the OSNR for several link lengths. The measured BER over OSNR in dB is shown in Fig. 8 for B2B and 301.8 km. For the

Fig. 6. $\ln(|b_N(\lambda_k, t_0)|)$ over a) $\text{Re}\{\lambda_{kN}\}$ and b) $\text{Im}\{\lambda_{kN}\}$ for $t_0 = 0$ in c) and $t_0 = 1$ in d) respectively.

Fig. 7. Measured standard deviation σ of a) $\ln|b_N(\lambda_{kN}, t_0)|$ and b) $\angle b_N(\lambda_k, t_0)$ over t_0 with (orange) and without equalization (blue).

chosen modulation scheme, JD equalization decreases the OSNR requirements by 4.4 dB at a BER level of 10^{-3} . For 301.8 km we observe an increase to 3.7 dB, leading to an overall performance decrease for higher link lengths. To investigate the performance of JD equalization on long-haul transmissions, we use the configuration shown in Fig. 5. Fig. 15 shows the BER over the link length for JD equalization in purple squares. The transmission is error-free up to a link length of 503 km and reaches a BER of 10^{-1} at 3319 km. We can observe a slightly better performance and a reach increase from 1560 km to 1680 km at a BER of 10^{-2} by applying JD equalization. We also observe the decreasing performance gain for increasing link lengths due to the decreasing OSNR gain shown in Fig. 8. We attribute the limited improvement to the fact that the compensation method relies on a very

Fig. 8. Measured BER without (blue) and with (orange) equalization over OSNR for a) B2B and b) 301.8 km for JD equalization.

Fig. 9. Resulting BER after equalization using linear and nonlinear MMSE equalizers for 1911 km transmission reach depending on the number of training symbols and order of nonlinearity.

simplified model for which the validity decreases for longer transmission lengths. Due to its numerical simplicity the proposed scheme is an effective equalization for shorter link lengths.

2) *MMSE*: For the linear MMSE a point of interest is the amount of training needed. Fig. 9 shows the BER after linear MMSE equalization for different numbers of training symbols after 1911 km reach. Here one can see that approximately 1500 training symbols are sufficient to obtain good results. Using an MMSE the reach of this system can be improved by 200 km to around 1550 km, if HD-FEC is considered. This is depicted in Fig. 15.

3) *Nonlinear Equalizer*: Since the amount of coefficients to be calculated increases nonlinearly with the order of nonlinearity used in the equalizer, it can be assumed that the amount of training data needed to get satisfactory results rises as well. Therefore the amount of training symbols used for equalization has been varied for a transmission reach of 1911 km. The resulting BERs are plotted in Figs. 9 and 10. Here one can see that approximately 1600 training symbols are needed for nonlinear equalization of 2nd order to reach the required BER of 3.5×10^{-3} . Increasing the amount

Fig. 10. Zoomed in depiction of Fig. 9 showing asymptotic behaviour of the BER after employing nonlinear MMSE equalizers of different order depending on the amount of training symbols used.

Fig. 11. Resulting BER after varying transmission distances for different orders of nonlinearity for the nonlinear MMSE equalizer using 10,000 training symbols. 2nd order (cross) and 3rd order (circle). For readability the HD-FEC limit is included (dashed line).

of training symbols does not lead to improvements. If the order of nonlinearity is increased to 3, at least 7100 training symbols are needed to reach the lowest BER of roughly 1.5×10^{-3} . However, decreasing the number of training symbols to 3100 leads to satisfactory results at a BER of 1.85×10^{-3} and further decreasing the amount of training symbols to 1600 leads to BERs close to the ones of the 2nd order equalizer. Increasing the order of nonlinearity to the 4th order leads to improvements compared to the 3rd order equalizer, if more than 8100 training symbols are being used (resulting BER of 1.3×10^{-3}). However, a trade-off occurs, if less training symbols are being used, since the 4th order equalizer performs worse than the equalizer of 3rd order at below 6000 training symbols. If the order of nonlinearity is further increased to the 5th order, more than 15,000 training symbols are being needed to reach asymptotic behaviour in this transmission experiment. Since the gain of nonlinearities beyond the 3rd order is minimal, especially if the amount of training needed and the increase in computational complexity are considered, in the following filters up to the 3rd order are considered only. The results of nonlinear equalization for different distances are depicted in Fig. 11. Exemplary constellations are shown

Fig. 12. Received eigenvalues after 1307 km transmission. Orange markups represent eigenvalues, which lead to symbol errors.

Fig. 13. Received $b(\lambda)$ after 2515 km transmission before equalization a) and after equalization using a 3rd order nonlinear MMSE equalizer b).

in Fig. 13. Independent of the amount of implemented nonlinearity, an increase in transmission distance can be seen. If HD-FEC is assumed, the reach can be improved by 700 km using a 2nd order equalizer. This can be further improved, if higher order equalizers are employed with an increase in distance of 1350 km for 3rd order equalizers.

Fig. 12 shows the received eigenvalues after 1307 km transmission reach. Mind that due to the representation method of the scatter plot the mean deviation of the eigenvalues seems to be very high, which is not the case. Still, one can see that a small portion of the eigenvalues are strongly shifted to the lower parts of the complex plane. These eigenvalues are spurious eigenvalues which emerged due to perturbations during transmission. If an eigenvalue is part of the main group of eigenvalues, the nonlinear equalizer is able to almost fully equalize the received $b(\lambda_k)$ and hence leads to lower BERs. However, if a spurious eigenvalue is chosen the equalizer is not able to fully compensate this mistake. Furthermore, if the order of nonlinearity is high, this leads to error propagation, because $b(\lambda_k)$ can be in a region where a higher bit-error penalty occurs after equalization. Due to this, the relative improvement by using higher order equalizers for shorter distances is smaller when compared to equalizers of lower orders.

Fig. 14. BER after NN equalization depending on the number of training symbols for different transmission distances and single eigenvalue transmission. 2515 km (triangle), 1710 km (cross).

4) *Neural network (NN) equalizer:* The NN used for this transmission consisted of two layers comprising of 250 nodes each. To evaluate the amount of training symbols needed, different amounts have been used as depicted in Fig. 14. Overall, 8000 training symbols can be sufficient for good results. However, an increase in the amount of training solitons used can improve the performance especially for lower transmission distances. The resulting BER for different distances is depicted in Fig. 15. Using this NN as an equalizer, BERs below the HD-FEC limit can be achieved for transmissions up to 3300 km.

5) *Comparison of different equalization techniques:* To draw a conclusion for single eigenvalue transmission, the BERs for different distances and the various equalizers are depicted in Fig. 15. Here, one can see that with increasing complexity of the equalizers an increase in possible reach for certain FEC limits and a reduction of the BER from 14% to 4.1e-3 for 3319 km can be achieved. In addition to the total BER performance the complexity of the equalizers is compared in Fig. 16. Here the BER in dependence of the total amount of additional multiplications needed for online equalization for each equalizer included in Fig. 15 is depicted. Since the training is assumed to be taken offline, it was not considered in this depiction. For the neural network equalizer each ReLu activation function inside the nodes is considered to be one multiplication, since it is a linear function which comes nonlinear by clipping. The amount of additional training symbols and numbers of multiplications for equalization are summarized in Tab. III. Additionally, the achieved BER normalized to the BER which was obtained without additional detection or equalization techniques employed, is displayed.

C. Dual-polarization transmission

In this section a dual-polarization transmission with two eigenvalues is being examined, according to Table II.

1) *Single eigenvalue equalizers:* Initially, the equalizers explained in Sections IV-B2 and IV-B3 were applied to each $b_{1,2}(\lambda_k)$ individually. Here, each λ_k and $a'(\lambda_k)$ associated to

Fig. 15. Comparison of BERs after different equalization techniques as described above for single eigenvalue transmission. For readability the HD-FEC limit is included (dashed line).

Fig. 16. Comparison of additional multiplications needed for equalization of the different equalizers depicted in Fig. 15 for 3019 km reach.

Table III
COMPARISON OF EQUALIZER TECHNIQUES FOR SINGLE EIGENVALUE TRANSMISSION AT 3019 KM REACH

	N_{train}	N_{mult}	BER improvement
JD	0	5	-1.21 dB
MMSE	≈ 2000	8	-8.87 dB
3rd-order NL-MMSE	≈ 6000	110	-17.1 dB
NN	$\approx 30,000$	72500	-30 dB

Fig. 17. BER after multi eigenvalue equalization for dual polarization two-eigenvalue transmission for 3420 km 1 GBd transmission. A 2nd order (x) and a 3rd order MMSE equalizer (circle) were trained.

the $b_{1,2}(\lambda_k)$ to be equalized were used. Hence, four equalizers were trained for this transmission. For good results 2000 training symbols for the LMMSE and 8000 training symbols for the 3rd order nonlinear equalizer were used in this section. Fig. 19 shows the resulting BERs for both equalizers and two symbol rates. In general both equalizers can improve the transmission distance. However, the nonlinear equalization deteriorates the results for relatively short transmission distances when compared to the linear counterpart and only shows small improvements for longer transmissions especially for the 1 GBd case. This may be due to the low order of modulation used in this transmission.

2) *Cross-correlation equalizer:* To improve the performance of the mentioned equalizers, the nonlinear structure was expanded to multiple eigenvalues as explained in Section III-D. The amount of training symbols needed for this equalizer structure is shown in Fig. 17. Here, the 3rd order equalizer outperforms the 2nd order equalizer when more than 10,000 training symbols are being used. However, good results can be seen using only 5000 training symbols and a 2nd order equalizer. The resulting BER depending on the fiber length using two cross-correlation equalizers is depicted in Fig. 18. For the higher baud-rate of 2 GBd the BER improvement (if using a 3rd order) is more prevalent than for the 1 GBd transmission. In conclusion, range improvements at the HD-FEC limit of roughly 700 km for 2 GBd and of 1000 km for 1 GBd can be achieved.

3) *Cross-correlation neural network equalizer:* To account for all cross-correlations, we investigated a more complex NN. In this network all received $b_{1,2}(\lambda_k)$ and all $\Delta\lambda_I$ and $\Delta a'(\lambda_k)$ were used for the input layer (so 4 $b_{1,2}(\lambda_k)$, 2 $\Delta\lambda_I$ and 2 $\Delta a'(\lambda_k)$). The output layer consisted of 256 nodes, which were coded to all 8 bits modulated on each multi-soliton in both polarizations. Since the size of the input and output layers for this transmission is higher when compared to the one eigenvalue NN the amount of hyperparameters (i.e. the amount of layers and nodes per layer) have to be changed.

Fig. 18. BER after cross-correlation nonlinear equalization for dual polarization two-eigenvalue transmission. No equalization (diamond), 2nd-order equalization (square), 3rd-order equalization (circle). Filled markers refer to 2 GBd transmission and unfilled markers to 1 GBd symbol rate. For readability the HD-FEC limit is included (dashed line).

Fig. 19. BER depending on transmission distance for using different equalizers. Filled markers refer to 2 GBd transmission and unfilled markers to 1 GBd symbol rate. No equalization (diamond), LMMSE using 2000 training symbols (square), 3rd-order NLMMSE using 15,000 training symbols (triangle), 3rd-order cross-correlation NLMMSE using 15,000 training symbols (circle), and 132/596-layer NN using 45,000 training symbols (star). For readability the HD-FEC limit is included (dashed line).

In a first step the amount of nodes was chosen to be as high as 1024 for two hidden layers. Together with 45,000 training symbols, the resulting BERs for all transmission distances under test were satisfying as can be seen in Fig. 20. However, the computational complexity of such an NN is very high. Hence, by using a random-search algorithm a better set of hyperparameters was found with 132 nodes for the first hidden layer and 596 nodes for the second hidden layer. This reduces the amount of weights to be trained by a factor of 5.7, while keeping the BER in the same region. However, the needed amount of training symbols stayed the same (45,000). The resulting BER using this optimized NN for both symbol rates can be seen in Fig. 19 in comparison to the other equalization techniques. In this figure one can also see that the NN is not able to decrease the BER for short transmission distances in both scenarios. This may be due to overfitting issues caused by the hyperparameters of the artificial neural network being optimized for higher transmission distances. However, if the solitons are transmitted over longer distances, major improvements of the BER can be observed. As can be seen

BERs below the HD-FEC threshold can be achieved for all transmission distances used in this experiment. This way the transmission distance for 1 GBd can be increased by 2000 km when compared to a transmission without equalization. In the 2 GBd case, the reach can be increased by roughly 1250 km, if HD-FEC is considered.

4) *Comparison of different equalization techniques:* Fig. 19 shows a comparison of different equalization techniques for the DP two-eigenvalue transmission experiment. If a simple LMMSE equalizer is being used, the overall BER can be improved. This way the transmission distance can be increased by around 200 km for both symbol rates. Here, the implementation of nonlinear equalizers does not show benefits, especially when the increasing complexity is considered. However, if the cross-correlations between the eigenvalues are taken into account using a nonlinear equalizer the distance below the HD-FEC limit can be increased by another 200 km when compared to the LMMSE equalizer. This can be further improved by using a neural-network equalizer. This way transmissions below HD-FEC BER limits for 2 GBd up to 2500 km and for

Fig. 20. BER after NN equalization for dual polarization two-eigenvalue transmission for using different network hyperparameters 1 GBd transmission. Two layers with 1024 nodes each (square) and two layers consisting of 132 and 596 nodes respectively (star). For readability the HD-FEC limit is included (dashed line).

Fig. 21. Comparison of additional multiplications needed for online equalization of the different equalizers depicted in Figs. 18 and 19 for 2414 km transmission distance and 2 GBd symbol-rate. The cross-correlation nonlinear equalizers are marked with an X in the legend.

1 GBd up to 5000 km can be realized. The complexity of the equalizers is compared in Fig. 21. Here the BER in dependence of the total amount of additional multiplications needed for online equalization for all equalizers included in Figs. 18 and 19 is depicted. Additionally, the employed equalization techniques are compared in Tab. IV.

V. CONCLUSION

In this paper we presented possibilities to further increase the reach of soliton transmission systems based on the NFT. Joint detection equalization is a useful method to compensate noise on the received symbols when using both amplitude and phase modulation in the discrete spectrum of a single eigenvalue system. Using JD, the reach at HD-FEC can be increased by roughly 100 km in the investigated systems. Since JD is computationally simple and no training is needed, this is

Table IV
COMPARISON OF EQUALIZER TECHNIQUES FOR DUAL-POLARISATION TRANSMISSION AT 2414 KM REACH FOR 2 GBd TRANSMISSION

	N_{train}	N_{mult}	BER improvement
MMSE	≈ 2000	32	-11.2 dB
2nd-order cross-corr. NL-MMSE	≈ 5000	432	-12.4 dB
NN	$\approx 45,000$	220,171	-31.7 dB

a cost-effective way to extend the transmission distance. Using nonlinear equalizers the reach of a single eigenvalue system can be more than doubled and increased by more than 1500 km at the HD-FEC limit in our case. However, this comes at the cost of required training symbols (around 6000) and more complex equalizer structures. If a smaller order of nonlinearity is implemented and further research is done with regard to a sophisticated choice of coefficients, the complexity of the nonlinear equalizers could be decreased substantially without a strong loss in performance. This could be an efficient way to achieve large increases in possible reach. By employing neural networks, additional improvements can be seen regarding the BER. At the cost of a high amount of approximately 30,000 required training symbols, the reach could be improved by a factor of 2.6.

In addition to the single eigenvalue transmission, a transmission using two eigenvalues on both polarisations has been conducted. Here, nonlinear equalization using only one of the eigenvalues showed no improvements in comparison to linear MMSE equalization. However, if the cross correlations between the eigenvalues were taken into account, using nonlinear equalizers, the reach could be increased. This way the transmission distance could be improved by 600 km for 1 GBd and 2 GBd symbol rates at HD-FEC limit. The largest increase in transmission distance could be seen, if a neural network was employed. Using a two-layer deep NN, which took cross-correlations into account, reach improvements of up to 2000 km for 1 GBd and 1200 km for 2 GBd transmission can be observed.

ACKNOWLEDGMENT

We want to thank Alexander Geisler and Jochen Leibrich for their extensive support with the transmission experiments.

We also want to thank Sebastian Kühl for the productive discussions about neural network equalizers.

REFERENCES

- [1] V. E. Zakharov and A. B. Shabat, *Exact Theory of Two-dimensional Self-focusing and One-dimensional Self-modulation of Wave in Nonlinear Media*, Sov. Phys. JETP, Vol. 34, No. 1, pp. 62, Jan. 1972.
- [2] M. Yousefi and F. Kschischang, *Information Transmission Using the Nonlinear Fourier Transform, Part I-III*, IEEE Transactions on Information Theory, Vol. 60 No. 7, Aug. 2014.
- [3] E. Ip, J. M. Kahn, *Compensation of dispersion and nonlinear impairments using digital backpropagation*, J. Lightw. Technol., Vol. 26, No. 20, pp.3416-3425, Oct. 2008.
- [4] K. V. Peddanarappagari, M. Brand-Pearce, *Volterra Series Transfer Function of Single-Mode Fibers*, J. Lightw. Technol., Vol. 15, No. 12, Dec. 1997.

- [5] Y. Gao et al, *Intra-Channel nonlinearities mitigation in pseudo-linear coherent QPSK transmission systems via nonlinear electrical equalizer*, Opt. Comm., Vol. 282, No.12., pp. 221-425, 2009.
- [6] S. Wahls et al, *Digital Backpropagation in the Nonlinear Fourier Domain*, 16th IEEE Int. Workshop on Sig. Proc. Adv. in Wireless Comm., 2015.
- [7] S. T. Le et al. *Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit*, Nature Photonics 11 (9), pp. 570-576, 2017.
- [8] F. Da Ros et al. *Dual-Polarization NFDm Transmission With Continuous and Discrete Spectral Modulation*, Journal of Lightwave Technology Vol. 37 No. 10, pp. 2335-2343, 2019.
- [9] A. Hasegawa and T. Nyu, *Eigenvalue communication*, Journal of Lightwave Technology, Vol. 11, No. 3, pp. 395-399, Mar. 1993.10
- [10] V. Aref, *Control and Detection of Discrete Spectral Amplitudes in nonlinear Fourier spectrum*, arXiv:1605.06328v1, 2016.
- [11] O. C. Wright, *The Darboux Transform of some Manakov systems*, Applied Mathematics Letters, Vol. 16, 2003.
- [12] K. Chan, J. Leibrich et al., *Joint Detection Equalization on Nonlinear Fourier Transform based Optical Communication*, ECOC 2019; 45th European Conference on Optical Communication, Dublin, Ireland, 22-26 September, 2019.
- [13] A. Geisler, C. G. Schaeffer, *Experimental Nonlinear Frequency Division Multiplexed Transmission Using Eigenvalues with Symmetric Real Part*, ECOC 2016; 42nd European Conference on Optical Communication, Dusseldorf, Germany, 18-22 Sep. 2016.
- [14] S. Gaiarin, A. M. Perego et al., *Dual Polarization nonlinear Fourier transform-based optical communication system*, Optica 5, 263-270, Mar. 2018.
- [15] S. Gaiarin et al. *Experimental Demonstration of Dual Polarization Nonlinear Frequency Division Multiplexed Optical Transmission System*, ECOC 2017; 43rd European Conference on Optical Communication, Gothenburg, Sweden, 17-21 Sep. 2017.
- [16] K. Chan, A. Geisler, et al., *Experimental Demonstration of Differential Coding Gain for Dual Polarization Two-Eigenvalue Signals*, IEEE Photon. Technol. Lett., Vol. 31, No. 24, Dec. 2019.
- [17] A. Span, V. Aref et al., *Efficient Precoding Scheme for Dual-Polarization Multi-Soliton Spectral Amplitude Modulation*, IEEE Transactions on Communications, Vol. 67, No. 11, pp. 7604-7615, Nov. 2019.
- [18] T. Gui, T. H. Chan et al., *Alternative Decoding Methods for Optical Communications Based on Nonlinear Fourier Transform*, J. Lightw. Technol., Vol. 35, No. 9, May, 2017.
- [19] T. Gui, T. C. Lu et al., *High-order modulation on a single discrete eigenvalue for optical communications based on nonlinear Fourier transform*, Optics Express, Vol. 25, No. 17 pp. 20286-20297, Aug., 2017.
- [20] M. Shehadeh et al. *Reduced-Complexity Nonlinear Soliton Amplitude Estimators*, Photonics Technology Letters, Vol. 31, No. 24, pp. 1933-1935, Dec. 2019.
- [21] A. Vasylichenkova et al. *Multi-eigenvalue communication paired with b-modulation*, ECOC 2019; 45th European Conference on Optical Communication, Dublin, Ireland, 22-26 Sep. 2019.
- [22] G. Zhou et al. *Improving Soliton Transmission Systems Through Soliton Interactions*, Optics Express, Journal of Lightwave Technology Vol. 38 No.14, pp. 3563-357, 2020.
- [23] H. Buelow, V. Aref, K. Schuh, and W. Idler, *Experimental Nonlinear Frequency Domain Equalization of QPSK Modulated 2-Eigenvalue Soliton*, Optical Fiber Communication Conference, Anaheim, USA, 20-24 Mar. 2016.
- [24] W. Zhang, T. Gui, *Correlated Eigenvalues of Multi-Soliton Optical Communications*, Scientific Reports 9, Article number 6399, 2019.
- [25] F. Musumeci et al., *An Overview on Application of Machine Learning Techniques in Optical Networks*, IEEE Communications Surveys & Tutorials, Vol. 21, No. 2, pp. 1383-1408, 2019.
- [26] T. A. Eriksson et al., *Applying Neural Networks in Optical Communication Systems: Possible Pitfalls*, Photon. Technol. Lett., Vol. 29, No. 23, pp. 2091-2094, 2017.
- [27] J. Koch, R. Weixer, S. Pachnicke, *Equalization of Soliton Transmission Based on Nonlinear Fourier Transform using Neural Networks*, ECOC 2019; 45th European Conference on Optical Communication, Dublin, Ireland, 22-26 September, 2019.
- [28] O. Kotlyar et al, *Combining nonlinear Fourier transform and neural network-based processing in optical communications*, Optics Lett., Vol. 45, No. 13, 2020.
- [29] Y. Wu, L. Xi, X. Zhang, Z. Zheng, J. Wei, S. Du, W. Zhang, X. Zhang, *Robust neural network receiver for multiple-eigenvalue modulated nonlinear frequency division multiplexing system*, Opt. Express 28, 18304-18316, 2020.
- [30] R. T. Jones et al, *Time-Domain Neural Network Receiver for Nonlinear Frequency Division Multiplexed Systems*, Photon. Technol. Lett., Vol. 30, No. 12, pp. 1079-1082, 2018.
- [31] S. Civelli et al. *Polarization-multiplexed nonlinear inverse synthesis with standard and reduced-complexity NFT processing*, Optics Express, Vol. 26 No. 13, pp. 17360-17377, Jun. 2018.
- [32] F. J. García-Gómez, V. Aref, *Statistics of the Nonlinear Discrete Spektrum of a Noisy Pulse*, J. Lightw. Technol., Vol. 37, No.14, Jul. 2019.
- [33] J. Gordon, H. Haus, *Random walk of coherently amplified solitons in optical fiber transmission*, Optics Letters, Vol. 11 No. 10, Oct. 1986.
- [34] A. Grant, L. K. Rasmussen, *Iterative techniques*, in Honig, M.L. (Ed.): *dvances in multiuser detection*, Wiley, pp. 47-128, 2009.
- [35] S. A. Derevyanko et al. *Non-Gaussian statistics of an optical soliton in the presence of amplified spontaneous emission*, Optics letters Vol. 28 No. 21, pp. 2097-2099, 2003.
- [36] E. Meron et al. *On the achievable communication rates of generalized soliton transmission systems*, arXiv:1207.0297, Feb. 2017.
- [37] Q. Zhang et al. *Noise Models in the Nonlinear Spectral Domain for Optical Fibre Communications*, arXiv:1702.06226v1, 2012.
- [38] R. Weixer, J. Koch, P. Plany, S. Ohlendorf, S. Pachnicke, *Mitigation of Nonlinear Impairments by using Support Vector Machine and Nonlinear Volterra Equalizer*, J. Appl. Sci., Special Issue "Optics for AI and AI for Optics", Vol. 9, No. 18, 3800, Sep. 2019.
- [39] V. Aref, H. Buelow et al., *On Spectral Phase Estimation of Noisy Solitonic Transmission*, 2016 Optical Fiber Communications Conference and Exhibition (OFC), Anaheim, USA, 20-24 Mar. 2016.
- [40] J. Koch, A. Moscoso-Mártir, J. Mueller, F. Merget, S. Pachnicke, J. Witzens, *Silicon Photonics DWDM NLFT Soliton Transmitter*, arXiv preprint arXiv:2004.03183, 2020.
- [41] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2017.
- [42] S. T. Le, J. E. Prilepsky et al., *Nonlinear Inverse Synthesis for Optical Links With Distributed Raman Amplification*, J. Lightw. Technol. Vol 34, No. 8, Apr. 2016.
- [43] J. Koch et al, *Transmission of Higher Order Solitons Created by Optical Multiplexing*, J. Lightw. Technol., Vol. 37, No. 3, Feb. 2019.

Jonas Koch received the M.Sc. degree in electrical engineering, information technology and business management from Kiel University, Kiel, Germany, in 2017. He is currently working towards the Ph.D. degree with the Chair of Communications, where his research concentrates on the nonlinear Fourier transform. He is involved in a project funded by the DFG, which examines the possibilities of joint optical-electrical signal processing.

Ken Chan received the M.Sc. degree in electrical engineering from Kiel University, Kiel, Germany, in 2018. He is currently working towards the Ph.D. degree at the Chair of High Frequency Technology at the Helmut-Schmidt University in Hamburg, Germany, where his research concentrates on noise impact on nonlinear Fourier transform.

Christian Schaeffer (M'88) received the Dipl.-Ing. and Dr.-Ing. degrees from the Technical University Berlin, Berlin, Germany, in 1984 and 1989, respectively. From 1988 to 1992, he was with the R&D Center within the Telecommunication Department of DASA, where he was engaged in research on coherent optical generation of microwave signals and optical interconnects for phased array antennas. In 1992, he was with the Fachhochschule Lübeck, Germany, as a Professor with the Department of Electrical Engineering. Between 1999–2009, he was a Full Professor for RF and Photonics, at the Communications Laboratory, Dresden University of Technology, Germany. He is currently with the Helmut-Schmidt-University, Hamburg, Germany. His research interests include microwave photonics, optical frequency

synthesis, Fiber-Bragg gratings, dispersion compensation, Silicon photonics and coherent quantum communication systems.

Stephan Pachnicke (M'09–SM'12) received the M.Sc. degree in information engineering from City University, London, U.K., in 2001, the Dr.-Ing. degree in electrical engineering from the Technical University of Dortmund, Dortmund, Germany, in 2005, and the Dipl.-Wirt.-Ing. degree in business administration from Fern Universitat, Hagen, Germany, in 2005. He finished his habilitation on optical transmission networks in 2012. From 2011 to 2015, he was with ADVA Optical Networking SE in the Advanced Technology Group (CTO Office). Since 2016, he has been a Full Professor and heading the Chair of Communications, Kiel University, Kiel, Germany. He is author or co-author of more than 150 scientific publications, author of a book on Fiber-Optic Transmission Networks (Springer, 2011), and holds several patents. Prof. Pachnicke is serving on the technical program committee of the European Conference of Optical Communication.