Probabilistic Eigenvalue Shaping for Nonlinear Fourier Transform Transmission

Andreas Buchberger, Alexandre Graell i Amat, *Senior Member, IEEE,*
Vahid Aref, *Member, IEEE,* and Laurent Schmalen, *Senior Member, IEEE*

Abstract—We consider a nonlinear Fourier transform (NFT)-based transmission scheme, where data is embedded into the imaginary part of the nonlinear discrete spectrum. Inspired by probabilistic amplitude shaping, we propose a probabilistic eigenvalue shaping (PES) scheme as a means to increase the data rate of the system. We exploit the fact that for an NFT-based transmission scheme, the pulses in the time domain are of unequal duration by transmitting them with a dynamic symbol interval and find a capacity-achieving distribution. The PES scheme shapes the information symbols according to the capacity-achieving distribution and transmits them together with the parity symbols at the output of a low-density parity-check encoder, suitably modulated, via time-sharing. We furthermore derive an achievable rate for the proposed PES scheme. We verify our results with simulations of the discrete-time model as well as with split-step Fourier simulations.

Index Terms—Discrete spectrum, nonlinear Fourier transform (NFT), probabilistic shaping, soliton communication.

I. INTRODUCTION

PULSE propagation in optical fibers is severely impaired by nonlinear effects that should be either compensated or utilized for the design of the communication system. The nonlinear Fourier transform (NFT) [1] provides a method to transform a signal from the time domain into a nonlinear frequency domain (spectrum), where the channel acts as a multiplicative filter on the signal. The nonlinear spectrum consists of a continuous and a discrete part. Both parts can be used to transmit information, either separately or jointly, and several schemes have been presented in theory and practice [1]–[6]. However, very little is known so far about the probability density function (PDF) of the received signal in the nonlinear spectral domain when it is contaminated by channel noise.

In [7], a simplified communication system modulating only the imaginary part of the eigenvalues in the discrete nonlinear spectrum was presented. For this scheme, an approximation for the conditional PDF of the channel can be obtained in closed form. In general, for a given channel, the capacity-achieving distribution is not known and is often different from the conventional distribution with equispaced signal points and uniform signaling. Hence, some form of shaping is required [8]. Two popular methods of shaping are probabilistic shaping and geometric shaping. In geometric shaping, the capacity-achieving distribution is mimicked by optimizing the position of the constellation points for equiprobable signaling [9] whereas probabilistic shaping uses uniformly spaced constellation points and approximates the capacity-achieving distribution by assigning different probabilities to different constellation points [8].

The main drawback of probabilistic shaping is its practical implementation. An abundance of probabilistic shaping schemes have been presented, most suffering from high decoding complexity, low flexibility in adapting the spectral efficiency, or error propagation. For a literature review on probabilistic shaping, we refer the reader to [10, Section II].

Recently, a new scheme called probabilistic amplitude shaping (PAS) has been proposed in [10]. Compared to other shaping schemes, PAS yields high flexibility and close-to-capacity performance over a wide range of spectral efficiencies for the additive white Gaussian noise (AWGN) channel while still allowing bit-metric decoding. Although originally introduced for the AWGN channel, PAS can be applied to other channels with a symmetric capacity-achieving input distribution assuming a sufficiently high spectral efficiency.

In this paper, we consider a similar NFT-based transmission scheme to the one presented in [7], where data is embedded into the imaginary part of the nonlinear discrete spectrum. As a means to increase the data rate, we demonstrate that the concept of PAS can be adapted to this NFT-based transmission system. In particular, we propose a probabilistic eigenvalue shaping (PES) scheme, enabling similar low complexity and bit-metric decoding as PAS. We take advantage of the dependence of the pulse length on the data for the NFT-based transmission system and transmit each pulse as soon as the previous one has been transmitted rather than with a fixed interval as in [7], yielding increased data rate. Accordingly, we find the capacity-achieving input distribution, maximizing the time-scaled mutual information (MI). For ease of notation, we refer to the maximized MI as capacity noting that it is in fact the constrained capacity of a system transmitting first-order solitons. The PES scheme then shapes the information symbols according to the capacity-achieving distribution by a distribution matcher (DM). The information symbols are also encoded by a low-density parity-check (LDPC) encoder and the parity symbols at the output of the encoder are suitably modulated. The resulting sequence of modulated symbols and

the sequence at the output of the DM are transmitted via time-sharing. We further derive an achievable rate for such a PES scheme. We demonstrate via discrete-time Monte-Carlo and split-step Fourier (SSF) simulations, that PES performs at around $2\,\mathrm{dB}$ from capacity using off-the-shelf LDPC codes. The proposed PES scheme yields a significant improvement of up to twice the data rate compared to an unshaped system as in [7].

It is important to note that although first-order solitons do not outperform conventional coherent systems due to their spectrally inefficient pulse shape compared to a Nyquist pulse shape, they have some other advantages. For instance, the first-order soliton transmission does not require chromatic dispersion (CD) compensation or digital backpropagation (DBP) as dispersion and nonlinearity are balanced and hence compensated. This work attempts to approach the limits of current NFT-based systems. To improve the spectral efficiency further, one should use higher-order solitons as well as the continuous part of the nonlinear spectrum together [4]. However, the channel equalization will not be as easy as the one for the first-order solitons and the channel model is not yet fully known.

The remainder of the paper is organized as follows. In Section II, we describe pulse propagation in an optical fiber and the NFT-based transmission scheme. In Section III, we optimize the input distribution and in Section IV, we introduce and describe the proposed PES scheme and derive an achievable rate. In Section V, we present numerical results for PES, both from Monte-Carlo simulation and SSF simulation, and in Section VI we draw some conclusions.

Notation: The following notation is used throughout the paper. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and the imaginary part of a complex number, respectively, and $\jmath = \sqrt{-1}$ denotes the imaginary unit. Vectors are typeset in bold, e.g., $\boldsymbol{x}$, random variables (RVs) are capitalized, e.g., X, and hence vectors of RVs are capitalized bold, e.g., $\boldsymbol{X}$. The PDF of an RV X is written as $p_X(x)$ and its expectation as $\mathbb{E}_X\{x\}$. The conditional PDF of Y given X is denoted as $p_{Y|X}(y|x)$. The probability mass function (PMF) of an RV X is denoted by $P_X(x)$. The transpose of a vector or matrix is given as $(\cdot)^\mathsf{T}$. A set is denoted by a capitalized Greek letter, e.g., Λ, and its cardinality by $|\Lambda|$. We write $\log_a(\cdot)$ for the logarithm of base a and $\ln(\cdot)$ for the natural logarithm.

II. NONLINEAR FOURIER TRANSFORM-BASED TRANSMISSION SYSTEM

A. Pulse Propagation and the Nonlinear Fourier Transform

Pulse propagation in optical fibers is governed by a partial differential equation, the stochastic nonlinear Schrödinger equation (NLSE),

$$\jmath\frac{\partial u(\tau,\ell)}{\partial \ell} + \jmath\frac{\alpha}{2}u(\tau,\ell) - \frac{\beta_2}{2}\frac{\partial^2 u(\tau,\ell)}{\partial \tau^2} + \gamma u(\tau,\ell)|u(\tau,\ell)|^2$$
$$= n(\tau,\ell),$$

where $u(\tau,\ell)$ denotes the envelope of the electrical field as a function of the position ℓ along the fiber and time τ, α the attenuation, β_2 the second order dispersion, γ the nonlinearity parameter, and $n(\tau,\ell)$ is a white Gaussian process in time and

in space with spectral density σ_0^2. The spectral density depends on the system and for distributed Raman amplification is given as $\sigma_0^2 = \alpha K_T h\nu_0$, where K_T is the temperature-dependent phonon occupancy factor, and $h\nu_0$ is the average photon energy [7]. A general closed-form solution of the stochastic NLSE does not exist. In some special cases, e.g., for noisefree and lossless fibers, special solutions like, e.g., solitons, exist. Furthermore, we consider the NLSE in normalized form in the focusing regime, i.e., $\beta_2 < 0$, under the assumption of ideal distributed Raman amplification, i.e., $\alpha = 0$,

$$\jmath\frac{\partial q(t,z)}{\partial z} + \frac{\partial^2 q(t,z)}{\partial t^2} + 2q(t,z)|q(t,z)|^2 = 0, \qquad (1)$$

where $t = \tau/\sqrt{|\beta_2|L/2}$, $z = \ell/L$, $q = u\sqrt{\gamma L}/\sqrt{2}$, and L is the length of the fiber. In this case, the NLSE is an integrable partial differential equation for which a pair of operators, called Lax pair, can be found. The eigenvalues of such an operator remain invariant during noiseless propagation and the Lax pair can be used to solve the partial differential equation. Solutions of (1) can be uniquely represented in terms of its eigenvalues via the so-called NFT. For a given position z, the NFT of a signal $q(t)$ (we drop the position z for simplicity of presentation) with support on the time interval $t \in [t_1, t_2]$, is calculated by solving the partial differential equation

$$\frac{\partial \boldsymbol{v}(t,\lambda)}{\partial t} = \begin{pmatrix} -\jmath\lambda & q(t) \\ -q(t)^* & \jmath\lambda \end{pmatrix}\boldsymbol{v}(t,\lambda), \qquad (2)$$

where $\boldsymbol{v}(t,\lambda) = \begin{pmatrix} v_1(t,\lambda) & v_2(t,\lambda) \end{pmatrix}$ is the eigenvector of the auxiliary operator, with boundary conditions

$$\boldsymbol{v}^{(1)}(t,\lambda) \rightarrow \begin{pmatrix} 0 & 1 \end{pmatrix}^\mathsf{T} \mathrm{e}^{\jmath\lambda t}, \qquad \text{as } t \rightarrow t_2$$
$$\boldsymbol{v}^{(2)}(t,\lambda) \rightarrow \begin{pmatrix} 1 & 0 \end{pmatrix}^\mathsf{T} \mathrm{e}^{-\jmath\lambda t}, \qquad \text{as } t \rightarrow t_1,$$

and λ is the spectral component. Solving (2) gives rise to the continuous and discrete nonlinear spectrum

$$\hat{q}(\lambda) = \frac{b(\lambda)}{a(\lambda)}, \lambda \in \mathbb{R} \quad \tilde{q}(\lambda_i) = \frac{b(\lambda_i)}{\mathrm{d}a(\lambda)/\mathrm{d}\lambda|_{\lambda=\lambda_i}}, \lambda_i \in \mathbb{C}^+,$$

respectively, where $a(\lambda) = \lim_{t\rightarrow t_2} v_1^{(2)}(t,\lambda)\,\mathrm{e}^{\jmath\lambda t}$, $b(\lambda) = \lim_{t\rightarrow t_2} v_2^{(2)}(t,\lambda)\,\mathrm{e}^{-\jmath\lambda t}$, and λ_i are the zeros of $a(\lambda)$, $\lambda_i \in \mathbb{C}^+$, a finite set of isolated complex zeros, referred to as eigenvalues. Hence, the NFT represents the signal in the nonlinear spectral domain, where the influence of the channel on the signal is a multiplicative filter.

As a counterpart to the NFT that transforms a signal from the time domain to the nonlinear spectral domain, the inverse nonlinear Fourier transform (INFT) transforms a signal from the nonlinear spectral domain to the time domain. For an in-depth mathematical description of the INFT, we refer the interested reader to [1].

B. Soliton Transmission

As in [7], we embed information in the imaginary part of the discrete spectrum, also referred to as eigenvalues. Hence, the input of the channel is an RV $X \in \Lambda = \{\lambda_1, \ldots, \lambda_M\}$, where Λ is the set of eigenvalues, λ_i is the ith eigenvalue, and M is the order of the modulation. The eigenvalues $\{\lambda_i\}$ are

Fig. 1. Block diagram of the NFT-based system.

assumed to be ordered in ascending order by their imaginary parts. Furthermore, the output of the channel is an RV $Y \in \Psi$, where $\Psi = \{y \in \mathbb{C} : \Re\{y\} = 0, \Im\{y\} \geq 0\}$. A block diagram is depicted in Fig. 1. The information embedded in a single eigenvalue $\lambda \in \Lambda$ is transformed to a time-domain signal $q(t,0)$ via the INFT where the transmitter is located at position $z = 0$ along the fiber. At position $z = 1$, the receiver calculates the discrete spectrum $\psi \in \Psi$ from the received signal $q(t,1)$ via the NFT. The time-domain signal corresponds to first order solitons, i.e.,

$$q(t,0) = 2\Im\{\lambda\}\operatorname{sech}(2\Im\{\lambda\}t).$$

For the NFT to be valid, the signal must have finite support, i.e., before transmitting the next pulse, the previous one must have returned to zero. As the pulses in general have infinite tails, we truncate them when they fall below a threshold close to zero. We define the pulse over the smallest support containing a fraction $(1 - \delta)$ of the energy of the pulse and hence, we can formally define the pulse width as follows.

Definition 1. *The pulse width of λ is defined as the smallest support containing a fraction $(1-\delta)$ of the energy of the pulse,*

$$T(\lambda, \delta) \triangleq \frac{1}{2\Im\{\lambda\}} \ln\left(\frac{2}{\delta} - 1\right),$$

where $0 < \delta < 1$.

The value of the cutoff parameter δ must be chosen in a way such that soliton-soliton interactions are negligible. For longer transmission distances, δ decreases, i.e., the pulses must be spaced further apart. Furthermore, the condition

$$\mathrm{e}^{-2\Im\{\lambda\}\Delta(\lambda,\delta)} = \mathrm{e}^{-\ln\left(\frac{2}{\delta}-1\right)} \ll 1 \tag{3}$$

must be fulfilled [7].

At this point, it is important to comment on the memorylessness of the system emanating from the absence of soliton-soliton interactions. A pulse train of well-separated first order solitons was investigated in [7] for launch powers of $-1.5\,\mathrm{dBm}$ and $1.45\,\mathrm{dBm}$ and transmission over $500\,\mathrm{km}$ and $2000\,\mathrm{km}$. It was shown via SSF simulations that the correlation between the symbols at the receiver is essentially zero, concluding that the channel is indeed memoryless in the transmission range of $500\,\mathrm{km}$ to $2000\,\mathrm{km}$ and transmit power range of $-1.5\,\mathrm{dBm}$ to $1.45\,\mathrm{dBm}$ for which the model (4) is applicable. While this approach is not a rigorous proof, the results indicate that memorylessness is a valid assumption. Although the transmission scheme is different in [7], the underlying condition that any two pulses need to be sufficiently separated is the same. Hence, we can treat the NFT-based transmission system in this work as a memoryless channel.

In a practical system, we assume distributed Raman amplification and amplifier-induced spontaneous emission (ASE)
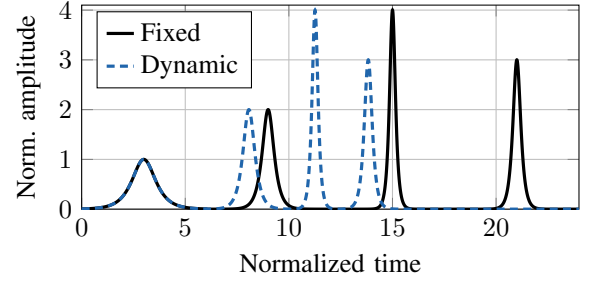

Fig. 2. Comparison of a pulse sequence with static symbol intervals and dynamic symbol intervals.

noise with received power spectral density σ^2 to compensate for the lossy fiber and be able to use the NFT to relate the input and the output. The conditional PDF of such a system has been derived via a perturbative approach and the Fokker-Planck equation method [11] and is used to design a communication system in [7]. It is given by

$$p_{Y|X}(\psi|\lambda) = \frac{2}{\sigma^2}\sqrt{\frac{\Im\{\psi\}}{\Im\{\lambda\}}}\, \mathrm{e}^{-2\frac{\Im\{\lambda\}+\Im\{\psi\}}{\sigma^2}}\, \mathrm{I}_1\left(\frac{4\sqrt{\Im\{\lambda\}\Im\{\psi\}}}{\sigma^2}\right), \tag{4}$$

where ψ is the received symbol as in Fig. 1, and $\mathrm{I}_1(\cdot)$ is the modified Bessel function of the first kind of order one. The power spectral density of the received ASE noise σ^2 is normalized and relates to real world units as $\sigma^2 = \gamma\sqrt{L^3}\sigma_0^2/\left(\sqrt{2|\beta_2|}\right)$. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} \triangleq 4\mathbb{E}_X\{\Im\{\lambda\}\}/\sigma^2$. It is important to note that the model (4) assumes the noise intensity to be small such that it can be treated as a perturbation to the soliton. Hence, the model is only applicable if the signal energy is not the same order as that of the noise. Furthermore, for very high signal powers, (4) is no longer valid either since the impact of the inelastic scattering effects (i.e., stimulated Raman or Brillouin scattering) is not considered within the 1st-order perturbation approach. For a detailed derivation of the model, we refer the reader to [11].

In [7], the shortest possible symbol interval is defined by the pulse duration of λ_1, i.e., the longest pulse. However, this tends to be inefficient since especially for short pulses, the guard interval between two consecutive pulses is longer than necessary and thereby limits the data rate. Here, we exploit the effect of varying pulse lengths and transmit each pulse as soon as the previous one has returned to zero. This concept is depicted in Fig. 2, where pulse sequences with fixed and varying symbol interval are compared. The figure clearly shows the advantage of a varying pulse interval and also demonstrates the aforementioned inefficiencies. The data rate of a system with varying symbol intervals depends on the distribution of the data. Thus, we define the average symbol interval as follows.

Definition 2. *The average symbol interval is*

$$\bar{T}(X) \triangleq \sum_{k=1}^{M} p_X(\lambda_k)T(\lambda_k) = \mathbb{E}_X\{T(\lambda)\}.$$

In [7], only eigenvalues with an imaginary part larger than zero are used. We extend this by allowing $\Im\{\lambda\} = 0$. In the time domain, this results in a pulse with amplitude zero, i.e., we do not transmit anything. We define its corresponding duration as the same as the duration of the shortest pulse, $T(\lambda = 0) \triangleq T(\lambda_M)$.

As any practical system can handle only a maximum peak power and a maximum bandwidth, we enforce a peak power constraint which relates to a maximum eigenvalue constraint. Especially in systems with lumped amplification and erbium-doped fiber amplifiers (EDFAs), such a constraint is required as eigenvalues fluctuate depending on their amplitude, which decreases the performance [12].

We note that the varying symbol interval introduces additional challenges on detection. In particular, an erroneously detected symbol may lead to error propagation, insertion errors (detection of symbols when none was transmitted), deletion errors (not detecting a transmitted symbol), or the loss of synchronization. To calculate the capacity, however, we neglect these effects. Hence, the results can be seen as an upper bound on the performance.

III. CAPACITY ACHIEVING DISTRIBUTION

From Fig. 2, it is intuitive that pulses with short duration should be transmitted more frequently than pulses with long duration. However, shorter pulses are more perturbed by noise than longer pulses. Hence, the optimal input distribution to the channel as described by the conditional PDF (4) is not the conventional uniform distribution. The channel capacity is obtained by maximizing the MI,

$$\mathbb{I}(X;Y) \triangleq \mathbb{E}_{X,Y}\left\{\log_2\left(\frac{p_{Y|X}(Y|X)}{\sum_{\tilde{\lambda}\in\Lambda} p_{Y|X}(Y|\tilde{\lambda})p_X(\tilde{\lambda})}\right)\right\}$$

over all possible input distributions $p_X(\lambda)$. Here, due to the variable transmission duration, we need to consider the MI under a variable cost constraint $\bar{T}(\cdot)$ [13],
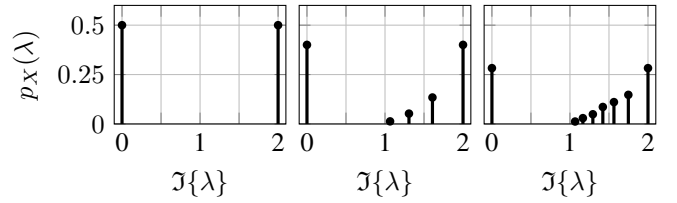
$$\mathsf{I}(X;Y) \triangleq \frac{\mathbb{I}(X;Y)}{\bar{T}(X)}. \tag{5}$$

To emphasize that the cost of a symbol is its corresponding pulse duration, we refer to the MI in the form of (5) as time-scaled MI. We can therefore define the capacity as

$$\mathsf{C} \triangleq \sup_{p_X(\lambda)} \mathsf{I}(X;Y) \tag{6}$$

where we set the supremum to zero if the set of distributions therein is empty. The capacity-achieving distribution, denoted by $p_X^*(\lambda)$, is in the set for which the supremum is non-zero.

As the MI $\mathbb{I}(X;Y)$ is concave in $p_X(\lambda)$ and $\bar{T}(X)$ is linear in $p_X(\lambda)$ and positive, the time-scaled MI $\mathsf{I}(X;Y)$ is quasiconcave [14, Table 2.5.2]. We can solve (6) and obtain the corresponding capacity-achieving distribution numerically. Exemplary results of the capacity-achieving distribution are shown in Fig. 3. We note that the lowest and highest amplitudes are always used with equal and high probability. For low SNRs, only these are used, i.e., on-off keying (OOK) is optimal. Furthermore, the capacity-achieving distribution is

(a) SNR = 12.14 dB (b) SNR = 19.71 dB (c) SNR = 29.51 dB

Fig. 3. Optimal distribution for different SNRs.

discrete and is of exponential-like shape with the exception of a point mass at zero as it can be seen in Fig. 3.

Note that C assumes memorylessness, which does not necessarily hold due to the variable symbol interval. Hence, C is, in fact, the constraint capacity under the assumption of a memoryless channel and the constraint of transmitting only first-order solitons. However, for notational simplicity, we refer to it simply as capacity with its corresponding capacity-achieving distribution.

In the case of a noiseless channel, it is possible to derive a closed form solution to (6) under the assumption of a finite discretization.

Lemma 1. *Let $\lambda_1, \lambda_2, \ldots, \lambda_M$ be $M \geq 2$ eigenvalues with $0 \leq \Im\{\lambda_1\} < \Im\{\lambda_2\} < \ldots < \Im\{\lambda_M\}$ and let $T(\lambda_k)$ be the time of transmitting a pulse with eigenvalue λ_k. Let r be the unique real positive root of the polynomial $\sum_{k=1}^{M} x^{-T(\lambda_k)} - 1$. Then, in the noiseless case, the capacity is obtained as*

$$\mathsf{C} = \log_2(r)$$

and the capacity-achieving distribution is given by

$$P_X^\diamond(\lambda_k) = \mathrm{e}^{-\ln(r)T(\lambda_k)}, \qquad k = 1, \ldots, M. \tag{7}$$

Proof. Suppose that the k-th eigenvalue is transmitted with probability P_k. For any fixed average symbol interval $\bar{T}(X) = \sum_k P_k T(\lambda_k)$, where $T(\lambda_M) \leq \bar{T}(X) \leq T(\lambda_1)$, we are interested in the distribution that maximizes the entropy while leading to the average symbol duration $\bar{T}(X)$. It is known that this distribution takes the form [15, Ch. 12]

$$P_k = \frac{\mathrm{e}^{-\theta T(\lambda_k)}}{\xi(\theta)} \tag{8}$$

where $\xi(\theta) = \sum_i \mathrm{e}^{-\theta T(\lambda_i)}$ ensures that $\sum_k P_k = 1$ and θ has to be selected such that $\sum_k P_k T(\lambda_k) = \bar{T}$. In the noiseless case, the MI is given by $\mathbb{I}(X;Y) = \mathbb{H}(X)$. The entropy $\mathbb{H}(X)$ then is

$$\mathbb{H}(X) =: \mathbb{H}(\theta) = -\sum_{k=1}^{M} P_k \log_2\left(\frac{\mathrm{e}^{-\theta T(\lambda_k)}}{\xi(\theta)}\right)$$

$$= \frac{1}{\ln(2)} \sum_{k=1}^{M} P_k(\theta T(\lambda_k) + \ln(\xi(\theta)))$$

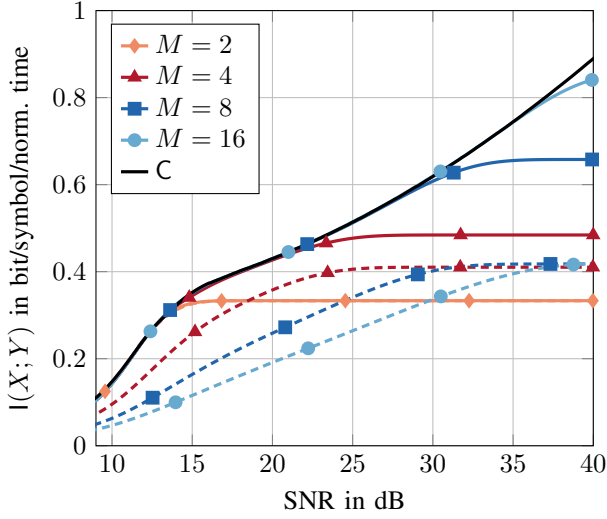$$= \frac{\theta \bar{T}(X)}{\ln(2)} + \log_2(\xi(\theta)).$$

Fig. 4. Time-scaled MI of the optimal distribution for linearly spaced constellations with M points (colored with markers solid), and of a system as in [7] (with markers dotted). As a reference, the capacity C is plotted as well (black solid without markers). For the cutoff parameter, $\delta = 0.005$ was used.

The time-scaled MI hence takes the form

$$
\begin{aligned}
I(X;Y) &= \frac{\theta}{\ln(2)} + \frac{\log_2(\xi(\theta))}{\bar{T}(X)} \\
&= \frac{\theta}{\ln(2)} + \frac{\log_2(\xi(\theta))}{\sum_k P_k T(\lambda_k)} \\
&= \frac{\theta}{\ln(2)} + \frac{\log_2(\sum_k e^{-\theta T(\lambda_k)}) \sum_k e^{-\theta T(\lambda_k)}}{\sum_k e^{-\theta T(\lambda_k)} T(\lambda_k)}.
\end{aligned}
$$

In order to maximize $I(X;Y)$, we find the optimal parameter θ by setting $\xi(\theta) = 1$. This can be seen by setting the derivative of $I(X;Y)$ to zero, with

$$
\frac{\partial}{\partial \theta} I(X;Y) =
$$
$$
\log_2\left(\sum_k e^{-\theta T(\lambda_k)}\right)\left(\frac{\sum_k e^{-\theta T(\lambda_k)}}{\sum_k T(\lambda_k) e^{-\theta T(\lambda_k)}}\right)^2 \operatorname{var}(T(\lambda)),
$$

where $\operatorname{var}(T(\lambda))$ denotes the variance of $T(\lambda_k)$ for the given θ. By assumption, as all $T(\lambda_k)$ are different, the middle part of this expression is strictly positive and $\operatorname{var}(T(\lambda)) > 0$. Hence, it is easy to see that this derivative can only be zero if $\sum_k e^{-\theta T(\lambda_k)} = 1$. The optimal θ is hence found by setting $\xi(\theta) = 1$. Consider the polynomial

$$
f(x) = \sum_{k=1}^{M} x^{-T(\lambda_k)} - 1.
$$

As this polynomial is monotonically decreasing for positive x, with $\lim_{x \to 0^+} f(x) = +\infty$ and $\lim_{x \to +\infty} f(x) = -1$, $f(x)$ has exactly one positive real root. Let r be the unique positive real root of $f(x)$. Then $\hat{\theta} = \ln(r)$. Inserting $\hat{\theta}$ into $I(X;Y)$ and (8) proves the lemma. $\qquad\square$

We clearly see that (7) is of exponential shape with an additional point mass at zero. Furthermore, we note that the shape of the distribution is mostly caused by the variable pulse
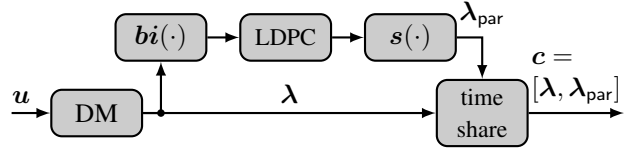
duration. The noise then determines the optimal location and optimal number of constellation points.

For a transmission system, the MI is an upper bound on the achievable rate. In Fig. 4 we evaluate the time-scaled MI for various input distributions for a cutoff parameter $\delta = 0.005$. The capacity is depicted with a black solid line. To reduce the complexity of implementation, we constrain the constellation Λ to M linearly spaced points from $\lambda_1 = 0$ to λ_M, i.e.,

$$
\lambda_i = (i-1)\frac{\lambda_M}{M-1} \text{ for } i = 1, \dots, M,
$$

and plot the corresponding time-scaled MI in colored solid lines with markers. We note that the time-scaled MI is very close to the capacity curve until it saturates. Increasing the modulation order M shows significant increase in the time-scaled MI. For comparison purposes, we also plot the time-scaled MI for a system with fixed symbol duration and conventional uniform distribution on a linearly spaced constellation as in [7]. We observe that the rate saturates at very low values and that increasing the modulation order M shows only slight improvement.

IV. Probabilistic Eigenvalue Shaping

In the previous section, we observed a significant gap between the time-scaled MI of the system in [7] and the capacity. This gap is referred to as shaping gap. In order to close it, we propose a PES system as shown in Fig. 5, inspired by PAS [10].

In the PAS scheme, the sequence of uniformly distributed data bits is mapped to a sequence of positive amplitudes distributed half Gaussian by a DM. The binary image of this sequence is encoded by a systematic forward error correction (FEC) code, resulting in uniformly distributed parity bits, which are then used to map the sequence of half Gaussian distributed symbols to a stream of Gaussian distributed symbols.

As the capacity-achieving distribution $p_X^*(\lambda)$ is not symmetric, PAS cannot be directly applied here. However, in order to keep the benefits of PAS, we wish to apply the DM before the FEC. We describe PES in the following with reference to Fig. 5. The binary data sequence u of length k_s bits is mapped by the DM to a sequence of eigenvalues $\lambda \in \Lambda^{n_s}$ of length n_s distributed according to $p_X^*(\lambda)$. The constant composition distribution matcher (CCDM) can be used for that purpose [16]. It is asymptotically optimal as its rate R_s approaches the entropy of the desired channel input X,

$$
R_s = \frac{k_s}{n_s} \to \mathbb{H}(X) \text{ as } n_s \to \infty.
$$

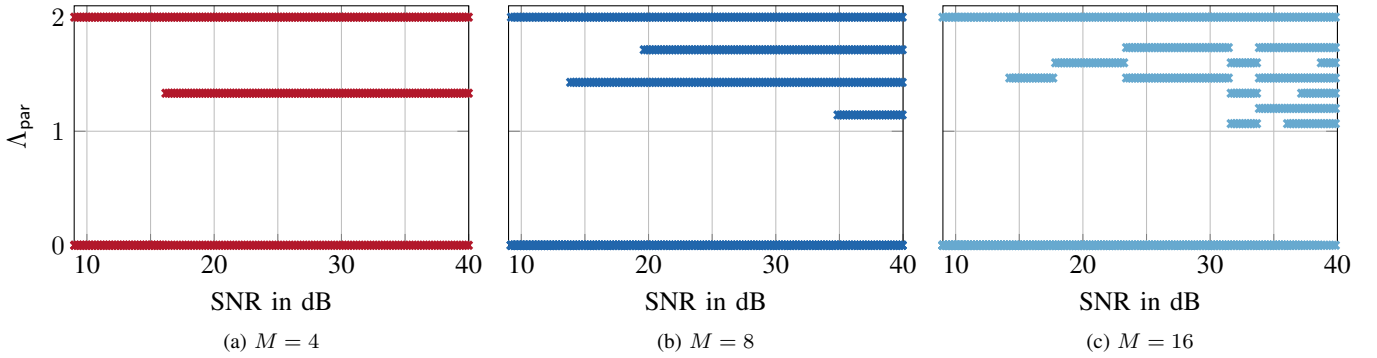Fig. 5. Block diagram of the PES scheme.

Fig. 6. Resulting constellations for the parity symbols for different SNRs. Note that the highest and the lowest eigenvalue is always occupied for every modulation order.

For large block sizes, the gap between R_s and $\mathbb{H}(X)$ is sufficiently small and can be neglected. Note that some of the possible eigenvalues may occur with probability zero.

We consider the modulation order M to be a power of two such that we can define its binary image. The binary image of $\boldsymbol{\lambda}$, $\boldsymbol{bi}(\boldsymbol{\lambda})$, is then encoded by a systematic encoder with information block length k_c, code length n_c, and rate $R_c = \frac{k_c}{n_c}$. The code is denoted by $\mathcal{C}$, with $|\mathcal{C}| = 2^{k_c}$. The parity bits at the output of the encoder are mapped to a sequence of eigenvalues $\boldsymbol{\lambda}_{\text{par}} \in \Lambda_{\text{par}}$ with modulation order $M_{\text{par}} = |\Lambda_{\text{par}}|$ and $\Lambda_{\text{par}} \subseteq \Lambda$ by the block $\boldsymbol{s}(\cdot)$ in Fig. 5 such that they are uniformly distributed.

Assuming that a high code rate R_c is used, we accept a small penalty with respect to the optimal channel input distribution and transmit $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{\text{par}}$ via time-sharing. The major difference of PES compared to PAS is the fact that the channel input distribution is not the optimal distribution due to the time-sharing with the sequence $\boldsymbol{\lambda}_{\text{par}}$. Consequently, this causes a performance degradation. However, PES is highly flexible as the spectral efficiency can be adapted by the DM and the code rate R_c, and a single code can be used. Note that every eigenvalue is protected by the code as FEC is performed after the DM and decoding and demapping can be performed independently. Thus, PES shares these advantages with PAS.

We wish for a high code rate R_c to keep the performance degradation due to the time-sharing low. More precisely, we wish to maximize the number of symbols distributed according to $p_X^*(\lambda)$. The ratio between information symbols and coded symbols, denoted by R_{ts}, is an indication for the expected performance degradation,

$$R_{\text{ts}} = \frac{\frac{n_c R_c}{\log_2(M)}}{\frac{n_c R_c}{\log_2(M)} + \frac{n_c(1-R_c)}{\log_2(M_{\text{par}})}}$$

$$= \frac{R_c \log_2(M_{\text{par}})}{\log_2(M)(1-R_c) + R_c \log_2(M_{\text{par}})}. \quad (9)$$

A. Parity symbols

The parity symbols at the output of the FEC code encoder are uniformly distributed. In Fig. 4, we observed that OOK with uniform signaling, i.e., $\Lambda_{\text{par}} = \{\lambda_1, \lambda_M\}$ and $M_{\text{par}} = 2$, is optimal for low SNR as it achieves capacity and performs

reasonably well for high SNR. However, we note from Fig. 4 that for a higher order modulation, even with uniform signaling, higher rates are possible. Hence, here we consider a scenario where $M_{\text{par}} > 2$. We further increase the rate by only using a subset of Λ and by picking the eigenvalues such that they are not uniformly spaced.

Example 1. *Consider the information symbol alphabet* $\Lambda = \{\lambda_1, \ldots, \lambda_8\}$ *with* $M = 8$. *For the* Λ_{par}, *we could pick* $\Lambda_{\text{par}} = \{\lambda_1, \lambda_6, \lambda_7, \lambda_8\}$ *with* $p_X(\lambda) = \{0.25, 0.25, 0.25, 0.25\}$ *and* $M_{\text{par}} = 4$.

To find the function $\boldsymbol{s}(\cdot)$ that maps the parity symbols onto $\lambda \in \Lambda_{\text{par}}$, we use a greedy algorithm as described in Algorithm 1. It starts with OOK, i.e., $\Lambda_{\text{par}} = \{\lambda_1, \lambda_M\}$. For each of the remaining symbols $\lambda \in \Lambda \setminus \Lambda_{\text{par}}$, it calculates the time-scaled MI of $\lambda \cup \Lambda_{\text{par}}$, finds the symbol λ for which the time-scaled MI of $\lambda \in \Lambda \setminus \Lambda_{\text{par}}$ is maximized, and adds it to Λ_{par}. All symbols with a greater or equal imaginary part than λ are removed, i.e., the eigenvalues $\{\lambda' \in \Lambda : \Im\{\lambda'\} \geq \Im\{\lambda\}\}$ are removed. This process is repeated until there are no symbols left. We then choose the set of symbols that gives the highest time-scaled MI as Λ_{par}. We note that this procedure does not guarantee an optimal solution. However, for $M = \{4, 8\}$ an exhaustive search gives the same result as that of Algorithm 1.

In Fig. 6, we show Λ_{par} for different modulation orders and SNRs. For $M = 4$, we note that for low SNR OOK gives the best result. Increasing the SNR results in a third level being added. The same behavior is observed for $M = 8$. Compared to $M = 4$, the third level is introduced at a slightly lower SNR. This results from the fact that for $M = 8$, different constellation points are available. For $M = 16$, we note that again a third level appears when increasing the SNR. When further increasing it, this third level moves to an eigenvalue with larger imaginary part and consequently a fourth level at an eigenvalue with lower imaginary part appears. This behavior can be observed repeatedly. To map the binary parity bits to the constellation points, we require M_{par} to be a power of two. As this is not always the case (see Fig. 6), we pick the largest power of two that is smaller or equal than the number of constellation points given by Algorithm 1.

Algorithm 1 Algorithm to calculate the signal points for the parity symbols. With a slight abuse of notation, we denote the time-scaled MI of a set Λ_{par} by $\mathsf{I}(\Lambda_{\text{par}})$. We assume the symbols in the set to be uniformly distributed.

Input: Constellation Λ
Output: Constellation Λ_{par}
1: $\Lambda_{\text{placed}} = \{\lambda_1, \lambda_M\}$
2: $\Lambda_{\text{par}} = \{\lambda_1, \lambda_M\}$
3: $\Lambda_{\text{not placed}} = \Lambda \setminus \Lambda_{\text{par}}$
4: **while** $\Lambda_{\text{not placed}} \neq \emptyset$ **do**
5:    **for all** $\lambda_i \in \Lambda_{\text{not placed}}$ **do**
6:       Calculate $\mathsf{I}(\Lambda_{\text{placed}} \cup \lambda_i)$
7:    **end for**
8:    $\lambda_{\text{max}} := \arg\max \mathsf{I}(\cdot)$
9:    $\Lambda_{\text{placed}} = \Lambda_{\text{placed}} \cup \lambda_{\text{max}}$
10:    **if** $\mathsf{I}(\Lambda_{\text{placed}}) > \mathsf{I}(\Lambda_{\text{par}})$ **then**
11:       $\Lambda_{\text{par}} = \Lambda_{\text{placed}}$
12:    **end if**
13:    $\Lambda_{\text{not placed}} = \Lambda_{\text{not placed}} \setminus \{\lambda : \lambda \in \Lambda_{\text{not placed}}, \Im\{\lambda\} \geq \Im\{\lambda_{\text{max}}\}\}$
14: **end while**
15: **return** Λ_{par}

B. Achievable Rate of Probabilistic Eigenvalue Shaping

To characterize the performance of PES, we derive the achievable rate of PES, denoted by R_{ps}. We assume that the channel is memoryless and that the decoder performs bit-metric decoding.

Theorem 1. *The achievable rate of PES is*

$$
\mathsf{R}_{\text{ps}} = R_{\text{ts}}\left(\mathbb{H}(X) - \sum_{i=1}^{m}\mathbb{H}(X_i^{\mathsf{B}}|Y_i^{\mathsf{B}})\right)
$$
$$
+ (1 - R_{\text{ts}})\left(m_{\text{par}} - \sum_{i=1}^{m_{\text{par}}}\mathbb{H}(X_{\text{par},i}^{\mathsf{B}}|Y_{\text{par},i}^{\mathsf{B}})\right). \quad (10)
$$

Proof. The achievable rate for PAS has been derived in [17]. For a system employing time-sharing, the resulting achievable rate is the average of the achievable rate of the two transmission schemes. □

In Fig. 7, we plot the capacity and the achievable rate (10) for different code rates $R_c = \{1/4,\ 1/3,\ 2/5,\ 1/2,\ 3/5,\ 2/3,\ 3/4,\ 4/5,\ 5/6,\ 8/9,\ 9/10\}$ and modulation orders for a cutoff parameter $\delta = 0.005$. Λ_{par} and hence M_{par} are chosen according to the results of Algorithm 1. For each modulation order, we notice that the curves cross at a certain SNR. For SNRs below this point, the lowest code rate (corresponding to the highest curve) gives the best performance whereas for SNRs above this point, the highest code rate (corresponding to the highest curve) gives the best performance. We note the influence of time-sharing, which results in a gap between the achievable rate and capacity. The gap increases for lower code rates R_c as the channel input distribution deviates more from the optimal one.
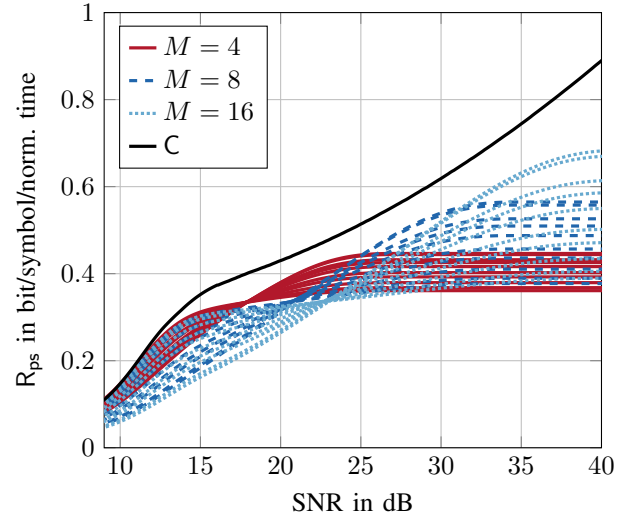

Fig. 7. Achievable rates for different code rates with Λ_{par} according to Algorithm 1 for a cutoff parameter $\delta = 0.005$.

V. NUMERICAL EVALUATION

In this section, we evaluate the performance of the PES scheme via discrete-time Monte-Carlo and SSF simulations. For the mapping $\boldsymbol{bi}(\cdot)$ (see Fig. 5), we use Gray labeling. Also, for the FEC, we use the binary LDPC codes of the DVB-S2 standard with code length $n_c = 64800$ and code rates $R_c = \{1/4,\ 1/3,\ 2/5,\ 1/2,\ 3/5,\ 2/3,\ 3/4,\ 4/5,\ 5/6,\ 8/9,\ 9/10\}$. For the parity symbols, we use the constellation arising from Algorithm 1, depicted in Fig. 6.

A. Detection

For the SSF simulation, we simulate a continuous signal and hence, we require a detector. We use the following method to deal with the variable pulse durations: We set a threshold θ sufficiently higher than the noise. Once the magnitude of the signal rises above θ, we save the time as t_{start} and when the magnitude of the signal falls below θ, we save the time as t_{end}. We then extend the interval bounded by t_{start} and t_{end}, i.e., $\tilde{t}_{\text{start}} = t_{\text{start}} - \delta_t$ and $\tilde{t}_{\text{end}} = t_{\text{end}} + \delta_t$. Calculating the NFT over the interval $[\tilde{t}_{\text{start}}, \tilde{t}_{\text{end}}]$ using the spectral method [1, Part II, Section IV] and only considering the imaginary part of the discrete eigenvalue gives the received symbol y. This approach requires that the SNR is sufficiently high. As the model has the same requirement due to the perturbation approach, this requirement is fulfilled.

It may happen that due to noise, a received pulse never rises above the threshold θ. In this case, the shortest duration is assumed (i.e., the duration of the pulse with amplitude zero). This scenario can be avoided by choosing the threshold sufficiently lower than the lowest amplitude. Furthermore, due to the shape of the capacity-achieving distribution, lower amplitudes are less likely, hence preventing this scenario.

To find the best threshold, we tested the performance for different values of θ and found that the performance of a threshold at 75% of the lowest non-zero amplitude of the constellation works best. We observed that small deviations

TABLE I
SIMULATION PARAMETERS.

Span length	l_{span}	$80\,\text{km}$
Second order dispersion	β_2	$-21.137\,\text{ps}^2\,\text{km}^{-1}$
Nonlinearity parameter	γ	$1.4\,\text{W}^{-1}\,\text{km}^{-1}$
Attenuation	α	$0.2\,\text{dB}\,\text{km}^{-1}$
Shortest pulse	T_{short}	$0.83\,\text{ns}$
Longest pulse	T_{long}	$1.3\,\text{ns}$
Bandwidth	B	$1.2\,\text{GHz}$
Avg. transmit power	P	$-4.86\,\text{dBm}$
Cutoff parameter	δ	0.005

of the threshold do not affect the performance significantly whereas setting the threshold too high (missing symbols with low amplitude) or to low (detecting a symbol where there is none) leads to performance degradation. Furthermore, we assume synchronization sequences spread sufficiently far apart in order not to impact the rate. We assume synchronization to be ideal such that it is guaranteed that error propagation is limited.

B. Numerical Results

We perform Monte-Carlo simulations of the discrete-time model (4) and show the results in Fig. 8, where we plot the transmission rate at a bit error rate (BER) of 10^{-5} for $M = 4, 8$ and 16. The highest transmission rate for each modulation order corresponds to the highest code rate R_c. We notice that the gap to capacity for $M = 4$ is smaller than for $M = 8$ and $M = 16$. If we consider $\Delta M = M - M_{\text{par}}$, i.e., the difference of the modulation order of Λ and Λ_{par}, we note that for a low M, ΔM is low was well. For example, for $M = 4$, $\Delta M \leq 2$. Hence, the rate loss due to time-sharing is small. For $M = 16$, the gap to capacity is smaller than for $M = 8$. Considering the relevant SNR range, we note that ΔM is smaller for $M = 16$ than for $M = 8$ and thus explaining the smaller rate loss.

We also simulated the transmission over a fiber using SSF simulations transmitting a train of solitons. We consider a single mode fiber (SMF) with parameters as in Table I and two different amplification schemes, distributed Raman amplification and lumped amplification using EDFAs. For both schemes, the peak power constraint is chosen such that the effect of the EDFAs can be neglected, i.e., $\lambda_{\text{max}} = 2\jmath$. We employ the detection schemes as described in V-A and choose the cutoff-parameter $\delta = 0.005$, i.e., 99.5% of the energy is contained in the pulse, for which the condition (3) is fulfilled. This then leads to a similar cutoff parameter as in [7]. For each modulation order $M = \{4, 8, 16\}$, we determine the furthest distance over which we achieve a BER of less than 10^{-5} and consider the rate gain compared to an unshaped system as in [7]. This results for $M = \{4, 8, 16\}$ in transmission over $3200\,\text{km}$, $3040\,\text{km}$, and $2960\,\text{km}$ at a rate gain of 20%, 26%, and 95%, respectively. The results do not differ for distributed and lumped amplification as this is ensured by the peak power constraint.
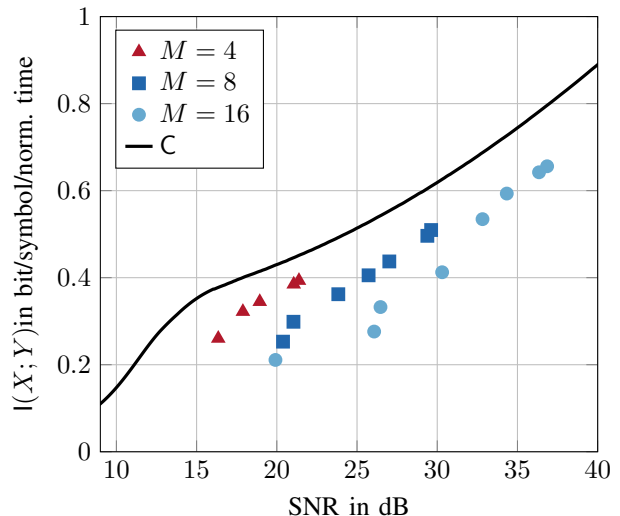

Fig. 8. Performance of time sharing with parity symbols according to Algorithm 1. The rate points correspond to a performance at BER $= 10^{-5}$. The highest transmission rate for each modulation order corresponds to the highest code rate R_c.

VI. CONCLUSION

In this paper, we presented a probabilistic shaping scheme for an NFT-based transmission system embedding information in the imaginary part of the discrete spectrum. It shapes the information symbols according to the capacity-achieving distribution and transmits them via time-sharing together with the uniformly distributed, suitably modulated parity symbols. We exploited the fact that the pulses of the signal in the time domain are of unequal length to improve the data rate compared to [7]. We used the time-scaled MI and derived the capacity-achieving distribution in closed form for the noiseless case and numerically in the general case. We showed that probabilistic eigenvalue shaping significantly improves the performance of an NFT-based transmission scheme, and can almost double the data rate. As a possible extension of our work, the continuous spectrum can be used to increase the spectral efficiency [4].

REFERENCES

[1] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part I-III," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4312–4369, Jul. 2014.

[2] Z. Dong, S. Hari, T. Gui, K. Zhong, M. I. Yousefi, C. Lu, P. K. A. Wai, F. R. Kschischang, and A. P. T. Lau, "Nonlinear frequency division multiplexed transmissions based on NFT," *IEEE Photon. Technol. Lett.*, vol. 27, no. 15, pp. 1621–1623, Aug. 2015.

[3] V. Aref, H. Bülow, K. Schuh, and W. Idler, "Experimental demonstration of nonlinear frequency division multiplexed transmission," in *Proc. 41st Eur. Conf. Opt. Commun. (ECOC)*, Valencia, Spain, Sep. 2015, pp. 1–3.

[4] V. Aref, S. T. Le, and H. Bülow, "Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime," in *Proc. 42nd Eur. Conf. Opt. Commun. (ECOC)*, Düsseldorf, Germany, Sep. 2016, pp. 1–3.

[5] A. Geisler and C. Schaeffer, "Experimental nonlinear frequency division multiplexed transmission using eigenvalues with symmetric real part," in *Proc. 42nd Eur. Conf. Opt. Commun. (ECOC)*, Düsseldorf, Germany, Sep. 2016, pp. 1–3.

[6] S. Hari, M. I. Yousefi, and F. R. Kschischang, "Multieigenvalue communication," *J. Lightw. Technol.*, vol. 34, no. 13, pp. 3110–3117, Jul. 2016.

[7] N. A. Shevchenko, S. A. Derevyanko, J. E. Prilepsky, A. Alvarado, P. Bayvel, and S. K. Turitsyn, "Capacity lower bounds of the noncentral chi-channel with applications to soliton amplitude modulation," *IEEE Trans. Commun.*, to appear.

[8] G. D. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[9] F.-W. Sun and H. C. A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1714–1716, Sep. 1993.

[10] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[11] S. A. Derevyanko, S. K. Turitsyn, and D. A. Yakushev, "Fokker-planck equation approach to the description of soliton statistics in optical fiber transmission systems," *J. Opt. Soc. Am. B*, vol. 22, no. 4, pp. 743–752, Apr. 2005.

[12] M. Zafruullah, M. Waris, and M. K. Islam, "Simulation and design of EDFAs for long-haul soliton based communication systems," in *Proc. Asia-Pacific Conf. Commun. (APCC)*, Penang, Malaysia, Sep. 2003.

[13] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[14] I. M. Stancu-Minasian, *Fractional Programming*, 1st ed. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1997.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[16] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[17] G. Böcherer, "Achievable rates for probabilistic shaping," *ArXiv e-prints*, Jul. 2017. [Online]. Available: https://arxiv.org/abs/1707.01134