On Probabilistic Shaping of Quadrature Amplitude Modulation for the Nonlinear Fiber Channel

Tobias Fehenberger, *Student Member, IEEE*, Alex Alvarado, *Senior Member, IEEE*, Georg Böcherer, *Member, IEEE*, and Norbert Hanik, *Senior Member, IEEE*

***Abstract*—Different aspects of probabilistic shaping for a multi-span optical communication system are studied. First, a numerical analysis of the additive white Gaussian noise (AWGN) channel investigates the effect of using a small number of input probability mass functions (PMFs) for a range of signal-to-noise ratios (SNRs), instead of optimizing the constellation shaping for every SNR. It is shown that if a small penalty of at most 0.1 dB SNR to the full shaping gain is acceptable, only two shaped PMFs per quadrature amplitude modulation (QAM) are required over a large SNR range. For a multi-span wavelength division multiplexing (WDM) optical fiber system with 64QAM input, it is shown that only one PMF is required to achieve large gains over uniform input for distances from 1400 km to 3000 km. Using a Gaussian noise (GN) model and full-field split-step simulations, we illustrate the ramifications of probabilistic shaping on the effective SNR after fiber propagation. Our results show that a shaping gain is obtained for the noise contributions from fiber amplifiers and modulation-independent nonlinear interference (NLI), while shaping simultaneously causes a penalty as it leads to an increased NLI. This nonlinear shaping loss, however, is found to have a minor impact, and optimizing the shaped PMF with a modulation-dependent GN model confirms that the PMF found for AWGN is also a good choice for a multi-span fiber system.**

***Index Terms*—Achievable Information Rates, Bit-Wise Decoders, Gaussian Noise Models, Mutual Information, Nonlinear Fiber Channel, Probabilistic Shaping, Wavelength Division Multiplexing.**

I. Introduction

THROUGH a series of revolutionary technology advances, optical transmission systems have been enabling the growth of Internet traffic for decades [1]. Most of the huge bandwidth of fiber systems has been used [2] and the capacities in the optical core network cannot keep up with the traffic growth [3].

The usable bandwidth of an optical communication system with legacy standard single-mode fiber (SMF) is effectively limited by the loss profile of the fiber and the Erbium-doped fiber amplifiers (EDFAs) placed between every span. It is thus of high practical importance to increase the spectral efficiency (SE) in optical fiber systems. Even with new fibers, the transceiver will eventually become a limiting factor in the pursue of higher SE as the practically achievable signal-to-noise ratio (SNR) can be limited by transceiver electronics [4]. Digital signal processing (DSP) techniques that are robust against fiber nonlinearities and also offer sensitivity and SE improvements in the linear transmission regime are thus of great interest.

A technique that fulfills these requirements and that has been very popular in recent years is *signal shaping*. There are two types of shaping, geometric shaping and probabilistic shaping. In geometric shaping, a nonuniformly spaced constellation with equiprobable symbols is used, while in probabilistic shaping, the constellation is on a uniform grid with differing probabilities per constellation point. Both techniques offer an SNR gain up to the ultimate shaping gain of 1.53 dB for the AWGN channel [5, Sec. IV-B], [6, Sec. VIII-A]. Geometric shaping has been used in fiber optics to demonstrate increased SE [7]–[11]. Probabilistic shaping has attracted a lot of attention in fiber optics [12]–[20]. In particular, [15], [16], [18], [19] use the probabilistic amplitude shaping scheme of [21] which allows to largely separate forward-error correction (FEC) from shaping by concatenating a distribution matcher [22] and an off-the-shelf systematic FEC encoder.

Probabilistic shaping offers several advantages over geometric shaping. Using the scheme in [21], the labeling of the quadrature amplitude modulation (QAM) symbols can remain an off-the-shelf binary reflected Gray code, which gives large achievable information rates (AIRs) for bit-wise decoders and makes exhaustive numerical searches for an optimal labeling obsolete. A further feature of probabilistic shaping that, in fiber-optics, has only been considered in [16], [19] is that it can yield rate adaptivity, i.e., the overall coding overhead can be changed without modifying the actual FEC. Probabilistic shaping also gives larger shaping gains than purely geometric shaping [23, Fig. 4.8 (bottom)] for a constellation with a fixed number of points. Given these advantages, we restrict our analysis in this work to probabilistic shaping on a symbol-by-symbol basis. Shaping over several time slots has been studied theoretically in [24] and is out of the scope for this work.

In this paper, we extend our previous work on probabilistic shaping for optical back-to-back systems [18] and investigate the impact of shaping for QAM formats on the nonlinear interference (NLI) of an optical fiber channel with wavelength division multiplexing (WDM). For the analysis, we use a modulation-dependent Gaussian noise (GN) model [25] in addition to full-field split-step Fourier method (SSFM) simula-

Tobias Fehenberger, Georg Böcherer, and Norbert Hanik are with the Institute for Communications Engineering, Technical University of Munich (TUM), 80333 Munich, Germany (Emails: tobias.fehenberger@tum.de, georg.boecherer@tum.de, norbert.hanik@tum.de).

Alex Alvarado is with the Optical Networks Group, University College London (UCL), London, WC1E 7JE, UK (Email: alex.alvarado@ieee.org).

Alex Alvarado's research is supported by the Engineering and Physical Sciences Research Council (EPSRC) project UNLOC (EP/J017582/1), United Kingdom. Georg Böcherer's research is supported by the German Federal Ministry of Education and Research in the framework of an Alexander von Humboldt Professorship.

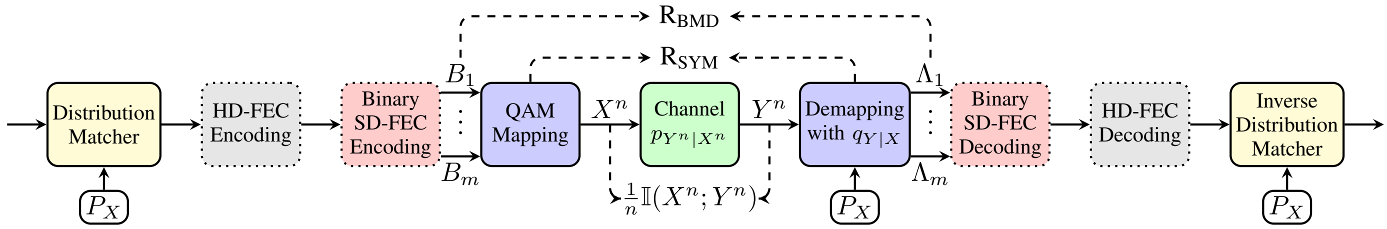

Fig. 1. Block diagram of a coded modulation transmitter with probabilistic shaping, a bit-wise demapper that uses the auxiliary channel $q_{Y|X}$, and a binary decoder. The dotted FEC blocks are omitted in the fiber simulations as we focus on AIRs. The transceiver block that use the nonuniform distribution are marked with P_X.

tions. This GN model includes the impact of the channel input on the NLI by taking into account higher-order standardized moments of the modulation, which allows us to study the impact of probabilistic shaping on the NLI from a theoretical point of view.

The contributions of this paper are twofold. First, we show that one shaped QAM input, optimized for the AWGN channel, gives large shaping gains also for a multi-span fiber system. This potentially allows for a simplified implementation of probabilistic shaping as one input PMF can be used for different fiber parameters. Secondly, no significant additional shaping gain is obtained for such a multi-span system with 64QAM when the PMF is optimized to the optical fiber channel using a GN model. The relevance of this result is that numerical optimizations of the channel input are shown to be obsolete for many practical long-haul fiber systems.

II. Fundamentals of Probabilistic Shaping

In the following, we review the basic principles of probabilistic shaping. The focus is on AIRs rather than bit-error ratios after FEC. Both symbol-wise AIRs as well as AIRs for bit-wise decoding are discussed. For a more detailed comparison, we refer the reader to [19, Sec. III], [23, Ch. 4], [26], [27].

A. Achievable Information Rates

Consider an independent and identically distributed (iid) discrete channel input $X^n = X_1, X_2, \ldots, X_n$ and the corresponding continuous outputs $Y^n = Y_1, Y_2, \ldots, Y_n$. The channel is described by the channel transition probability density $p_{Y^n|X^n}$, as shown in the center of Fig. 1. The symbolwise inputs X are complex QAM symbols that take on values in $\mathcal{X} = \{x_1, \ldots, x_M\}$ according to the probability mass function (PMF) P_X on $\mathcal{X}$. Without loss of generality, the channel input is normalized to unit energy, i.e., $\mathbb{E}[|X|^2] = 1$. The constellation size $|\mathcal{X}|$ is the modulation order and denoted by M. Unless otherwise stated, we consider QAM input that can be decomposed into its constituent 1D pulse amplitude modulation (PAM) constellation without loss of information. This means that every QAM symbol can be considered as two consecutive PAM symbols that represent real and imaginary part of the QAM symbol, and the probability of each two-dimensional (2D) QAM constellation is the product of the respective one-dimensional (1D) PAM probabilities, denoted by P_{1D}. The analysis in this work is carried out with 2D QAM symbols, and it is explicitly stated when a 1D input is considered, which is done mainly for the ease of notation and graphical representation.

The mutual information (MI) between the channel input and output sequences, normalized by the sequence length, is defined as

$$\frac{1}{n}\mathbb{I}(X^n;Y^n) = \frac{1}{n}\mathbb{E}\left[\log_2 \frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}\right], \tag{1}$$

where $\mathbb{E}[\cdot]$ denotes expectation and p_{Y^n} is the marginal distribution of Y^n. The MI in (1) is an AIR for a decoder that uses soft metrics based on $p_{Y^n|X^n}$.

Since the optical channel is not known in closed form, we cannot directly evaluate (1). A technique called mismatched decoding [28], [29] is used in this paper, which gives an AIR for a decoder that operates with the auxiliary channel $q_{Y^n|X^n}$ instead of the true $p_{Y^n|X^n}$. In this paper we consider memoryless auxiliary channels of the form

$$q_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} q_{Y_i|X_i}(y_i|x_i), \tag{2}$$

which means that, in the context of fiber-optics, all correlations over polarization and time are neglected at the decoder. We assume a fixed auxiliary channel, i.e., $q_{Y_i|X_i} = q_{Y|X}\ \forall i$, and restrict the analysis in this paper to 2D Gaussian distributions

$$q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|y-x|^2}{2\sigma^2}}, \tag{3}$$

where σ^2 is the noise variance of the auxiliary channel, $x \in \mathcal{X}$, and y is complex. For details on the impact of higher-dimensional Gaussian auxiliary channels, see [30], [31]. Irrespective of the particular choice of the auxiliary channel, we get a lower bound to $\mathbb{I}(X;Y)$ by using $q_{Y|X}$ instead of $p_{Y^n|X^n}$ [32, Sec. VI],

$$\frac{1}{n}\mathbb{I}(X^n;Y^n) \geq \mathbb{E}\left[\log_2 \frac{q_{Y|X}(Y|X)}{q_Y(Y)}\right] \triangleq \text{R}_{\text{SYM}}, \tag{4}$$

where the expectation is taken with respect to p_{XY}, and $q_Y(Y) = \sum_{x' \in \mathcal{X}} q_{Y|X}(Y|x') P_X(x')$. R_{SYM} can be estimated from Monte Carlo simulations of N input-output pairs (x_k, y_k) of the channel as

$$\text{R}_{\text{SYM}} \approx \frac{1}{N}\sum_{k=1}^{N} \log_2 \frac{q_{Y|X}(y_k|x_k)}{q_Y(y_k)}. \tag{5}$$

The symbolwise AIR $\mathrm{R}_{\mathrm{SYM}}$ is achievable for a decoder that assumes $q_{Y|X}$. For the practical bit-interleaved coded modulation schemes [33], which are also used in fiber-optics [26], a bit-wise demapper is followed by a binary decoder, as shown in Fig. 1. In this setup, the symbolwise input X is considered to consist of m bit levels $\boldsymbol{B} = B_1, \dots, B_m$[1] that can be stochastically dependent, and the decoder operates on bit-wise metrics. An AIR for this bit-metric decoding (BMD) scheme is the BMD rate [27]

$$\mathrm{R}_{\mathrm{BMD}} \triangleq \left[\sum_{i=1}^{m} \mathbb{I}(B_i;Y) - \underbrace{\left[\sum_{i=1}^{m} \mathbb{H}(B_i) - \mathbb{H}(\boldsymbol{B})\right]}_{(\star)}\right]^{+} \tag{6}$$

$$= \left[\mathbb{H}(\boldsymbol{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y)\right]^{+}, \tag{7}$$

which is the AIR considered for the simulations in this work. In (6), the index i indicates the bit level, $\mathbb{H}[\cdot]$ denotes entropy and $[\ \cdot\ ]^{+}$ is $\max(\cdot\ ,0)$. Note that $\mathrm{R}_{\mathrm{BMD}}$ is upper-bounded by the symbolwise MI, $\mathbb{I}(X;Y) \geq \mathrm{R}_{\mathrm{BMD}}$ [27]. The first term of (6) is the sum of MIs of m parallel bit-wise channels. Subtracting $(\star)$ in (6) corrects for a rate overestimate due to dependent bit levels. For independent bit levels, i.e., $P_{\boldsymbol{B}} = \prod_{i=1}^{m} P_{B_i}$, the term $(\star)$ equals 0 and $\mathrm{R}_{\mathrm{BMD}}$ becomes the well-known generalized mutual information calculated with soft metrics that are matched to the channel. We calculate $\mathrm{R}_{\mathrm{BMD}}$, which is an instance of (4), in Monte Carlo simulations of N samples as

$$\mathrm{R}_{\mathrm{BMD}} \approx \frac{1}{N}\sum_{k=1}^{N}\left[-\log_2 P_X(x_k)\right] - \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{m}\left[\log_2\left(1+e^{(-1)^{b_{k,i}}\Lambda_{k,i}}\right)\right], \tag{8}$$

where $b_{k,i}$ are the sent bits. The AIR $\mathrm{R}_{\mathrm{BMD}}$ is a function of the soft bit-wise demapper output $\Lambda_{k,i}$. These log-likelihood ratios (LLRs) are computed with the auxiliary channel as

$$\Lambda_{k,i} = \log \frac{\sum_{x\in\mathcal{X}_1^i} q_{Y|X}(y_k|x)P_X(x)}{\sum_{x\in\mathcal{X}_0^i} q_{Y|X}(y_k|x)P_X(x)} \tag{9}$$

$$= \log \frac{q_{Y|B_i}(y_k|1)}{q_{Y|B_i}(y_k|0)} + \log \frac{P_{B_i}(1)}{P_{B_i}(0)}, \tag{10}$$

where $\mathcal{X}_1^i$ and $\mathcal{X}_0^i$ denote the set of constellation points whose i^{th} bit is 1 and 0, respectively. The first term of (10) is the LLR from the channel and the second term the a-priori information. For uniformly distributed input, the priors equal zero. Using the 2D Gaussian auxiliary channel of (3), we have

$$\Lambda_{k,i} = \log \frac{\sum_{x\in\mathcal{X}_1^i} e^{-\frac{|y_k-x|^2}{2\sigma^2}} P_X(x)}{\sum_{x\in\mathcal{X}_0^i} e^{-\frac{|y_k-x|^2}{2\sigma^2}} P_X(x)}. \tag{11}$$

These LLRs can equivalently be computed in 1D if a symmetric auxiliary channel is chosen, a product labeling is used [23, Sec. 2.5.2] and $\mathcal{X}$ is generated from the product of 1D constellations.

[1] A binary-reflected Gray code is used as labeling rule because it gives large BMD rates and offers the symmetry that is required for the shaping coded modulation scheme [21, Sec. IV-A].

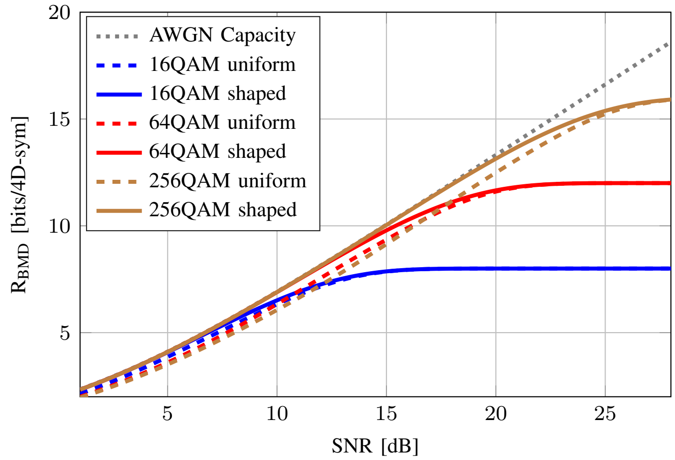

Fig. 2. $\mathrm{R}_{\mathrm{BMD}}$ in bits/4D-sym for uniform QAM input (dashed lines) and QAM with the shaped, SNR-dependent MB PMF of (14) (solid lines). The AWGN capacity (dotted line) is shown as a reference.

B. Probabilistic Shaping with Maxwell-Boltzmann PMFs

We search for the input distribution P_X^* that maximizes $\mathrm{R}_{\mathrm{BMD}}$ of (6),

$$P_X^* = \max_{P_X:\ \mathbb{E}[|X|^2]\leq 1} \mathrm{R}_{\mathrm{BMD}}, \tag{12}$$

where the underlying channel is AWGN. Probabilistic shaping for the nonlinear fiber channel is discussed in Sec. IV. As the AWGN channel is symmetric, the 1D PMFs are also symmetric around the origin, i.e.,

$$P_{\mathrm{1D}}^*(x) = P_{\mathrm{1D}}^*(-x), \tag{13}$$

which corresponds in 2D to a fixed probability per QAM ring. A common optimized input for (12) is to use shaped input distributions from the family of Maxwell-Boltzmann (MB) distributions [34, Sec. IV], [6, Sec. VIII-A]. The method to find an optimized input for a particular SNR is discussed in detail in [21, Sec. III-C] and briefly reviewed in the following.

Let the positive scalar ρ denote a constellation scaling of X with a fixed constellation $\mathcal{X}$. Further, let the PMF of the input be

$$P_X^*(x_i) = \frac{1}{\sum_{j'=1}^{M} e^{-\nu|x_{j'}|^2}} e^{-\nu|x_i|^2}, \tag{14}$$

where ν is another scaling factor. For every choice of ν, there exists a scaling ρ that fulfills the average-power constraint $\mathbb{E}[|\rho X|^2] = 1$. We optimize the scalings ρ and ν such that $\mathrm{R}_{\mathrm{BMD}}$ is maximized while using a distribution from (14) and operating at the channel SNR that is defined as

$$\mathrm{SNR} \triangleq \frac{\mathbb{E}[|\rho X|^2]}{\sigma^2} = \frac{E_{\mathrm{s}}}{N_0} = \frac{1}{N_0}, \tag{15}$$

where the 1D signal power E_{s} is normalized to 1 due to the average-power constraint and $N_0 = \sigma^2$ is the noise variance per dimension. This optimization can be carried out with efficient algorithms, see [21, Sec. III-C].

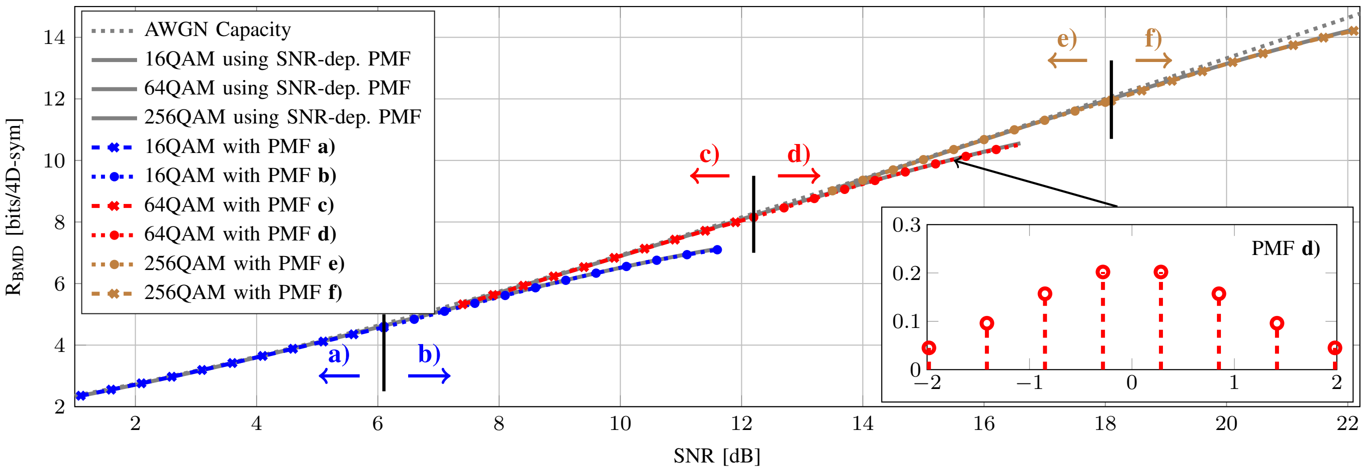

Fig. 3. $\mathrm{R_{BMD}}$ in bits/4D-sym for 16QAM, 64QAM, and 256QAM over the AWGN channel. The QAM formats shaped with SNR-dependent PMFs are shown as reference (gray solid lines), and shaping with fixed PMFs **a)** to **f)** is shown as colored lines with markers, with the respective SNR intervals indicated by vertical black lines. The inset shows the mismatched PMF **d)** that is used for SNRs from 12.2 dB to 16.6 dB. Details on the fixed distributions are given in Table I.

In Fig. 2, $\mathrm{R_{BMD}}$ in bits per four-dimensional symbol (bit/4D-sym) is shown vs. the SNR of an AWGN channel. We choose to plot $\mathrm{R_{BMD}}$ per 4D symbol to have values that are consistent with the dual-polarization AIRs of Sec IV. Further, only $\mathrm{R_{BMD}}$ is shown as it virtually achieves $\mathrm{R_{SYM}}$ [21, Table 3], [27, Fig. 1], and $\mathrm{R_{BMD}}$ is the more practical AIR compared to $\mathrm{R_{SYM}}$. The dotted curves represent uniformly distributed input and the solid lines show the $\mathrm{R_{BMD}}$ for QAM with optimized MB input. The AWGN capacity $2 \cdot \log_2(1 + \mathrm{SNR})$ is given as reference. Significant gains from probabilistic shaping over uniform input are found, with sensitivity improvement of up to 0.43 dB for 16QAM, 0.8 dB for 64QAM and more than 1 dB for 256QAM.

C. Shaping with Fixed PMFs

In order to find the optimized MB input, the SNR of the channel over which we transmit, denoted *channel* SNR, must be known or estimated *a priori* at the transmitter. This transmitter-side estimate of the SNR is referred to as *shaping* SNR. In a realistic communication system, it can be difficult to know the channel SNR at the transmitter because of varying channel conditions such as the number and properties of co-propagating signals, DSP convergence behavior, aging of components etc. Hence, shaping without knowledge of the channel SNR could simplify the implementation of probabilistic shaping. We will see in the following that an offset from the shaping SNR to the channel SNR has a minor effect on $\mathrm{R_{BMD}}$ for the AWGN channel if a suitable combination of QAM format and shaping SNR is used in the proper SNR regime.

To realize large shaping gains, we choose to operate each QAM format within 0.1 dB of the AWGN capacity. The maximum SNRs under this constraint are found to be 6.1 dB for 16QAM, 12.2 dB for 64QAM, and 18.1 dB for 256QAM and depicted in Fig. 3 as black vertical lines. We numerically search for the MB PMFs (obtained for a particular shaping SNR) that have at most 0.1 dB SNR loss compared to the full gain obtained when channel SNR and shaping SNR are identical. There are many distributions that fulfill this requirement, and we use the PMF that covers the largest SNR range while not exceeding the 0.1 dB penalty limit. The resulting PMFs of this numerical optimization are given as **a)**, **c)**, and **e)** in Table I[2]. We observe that a large SNR range is covered by a single PMF per QAM. These intervals, however, are disconnected, and an additional PMF per QAM is required to cover the entire SNR range without gaps. The shaped input PMFs **b)**, **d)**, and **f)** are given in Table I. They also have a penalty of at most 0.1 dB SNR to the full shaping gain, yet their operating range begins at a larger SNR, i.e., the upper end of the SNR range of **a)**, **c)**, and **e)**. This, however, comes at the expense at operating away from capacity. $\mathrm{R_{BMD}}$ for all PMFs of Table I are shown in Fig. 3. We see that only

TABLE I
FIXED DISTRIBUTIONS OF M-QAM (WITH $E_s = 1$) THAT LEAD TO AT MOST 0.1 dB SNR LOSS COMPARED TO THE FULL SHAPING GAIN

	a)	b)	c)	d)	e)	f)
M-QAM	16	16	64	64	256	256
Channel SNR range [dB]	1.1 to 6.1	6.1 to 11.6	7.4 to 12.2	12.2 to 16.6	13.5 to 18.1	18.1 to 22.1
Shaping SNR [dB]	1.2	9.9	9.3	15.0	15.5	20.6
One-sided 1D PMF $\begin{bmatrix} P_{\mathrm{1D}}(x_{\sqrt{M}/2+1}) \\ \vdots \\ P_{\mathrm{1D}}(x_{\sqrt{M}}) \end{bmatrix}$	0.432 0.068	0.332 0.168	0.277 0.16 0.053 0.01	0.200 0.157 0.096 0.046	0.152 0.131 0.097 0.062 0.034 0.016 0.006 0.002	0.109 0.101 0.088 0.071 0.054 0.038 0.025 0.015
One-sided 1D const. $\begin{bmatrix} x_{\sqrt{M}/2+1} \\ \vdots \\ x_{\sqrt{M}} \end{bmatrix}$	0.691 2.072	0.521 1.562	0.377 1.131 1.884 2.638	0.282 0.845 1.409 1.972	0.196 0.588 0.981 1.373 1.765 2.157 2.55 2.942	0.147 0.441 0.734 1.028 1.322 1.616 1.909 2.203

[2] To reduce the size of Table I, we use the symmetry property of (13) and show the 1D PMF P_{1D} for positive PAM constellations points $x_{\sqrt{M}/2+1}, \dots, x_{\sqrt{M}}$ only.

two input distributions per modulation format are required to obtain a large shaping gain. We will investigate the impact of probabilistic shaping on the fiber nonlinearities in the remaining of the paper.

III. SPM-XPM Model for the Nonlinear Fiber Channel

The effective SNR $\mathrm{SNR}_{\mathrm{eff}}$ of a signal after propagation over an optical fiber channel and receiver DSP can be approximated by a Gaussian noise (GN) model [35, Sec. VI],

$$\mathrm{SNR}_{\mathrm{eff}} = \frac{P_{\mathrm{tx}}}{\sigma^2_{\mathrm{eff}}} = \frac{P_{\mathrm{tx}}}{\sigma^2_{\mathrm{ASE}} + \sigma^2_{\mathrm{NLI}}}, \tag{16}$$

where P_{tx} is the optical launch power, the noise term σ^2_{ASE} represents the amplified spontaneous emission (ASE) noise from the optical amplifiers and σ^2_{NLI} is NLI which includes both intra-channel and inter-channel distortions. In the classic GN model of [35], the nonlinearities are modeled as additive memoryless noise that follows a circularly symmetric (c.s.) Gaussian distribution. In particular, in this model, the choice of channel input X does not have an impact on σ^2_{NLI} in [35, Sec. VI], later shown in [36] to be an inaccurate simplification. As a consequence, refined GN models have been presented in [25], [37], which now include properties of the channel input in the modeling of σ^2_{NLI}, resulting in a more accurate representation of modulation-dependent nonlinear effects. These models allow us to study probabilistic shaping for the fiber-optic channel without computationally expensive SSFM simulations.

In this work, we use the frequency-domain model by Dar *et al.* [25, Sec. III] [38] with both intra-channel, i.e., self-phase modulation (SPM), and inter-channel effects, i.e., cross-phase modulation (XPM) and refer to it as *SPM-XPM model*. Four-wave mixing has numerically been found to give a negligible contribution to the total NLI for the considered multi-span fiber setup and is thus omitted in the analysis. In the following, we give an overview over the model and refer the reader to [25, Sec. III] for details and derivations.

A. SPM-XPM Model

By rearranging the results in [25, Sec. III] and [39, App.], the NLI variance σ^2_{NLI} in (16) can be expressed as

$$\begin{aligned}\sigma^2_{\mathrm{NLI}} = P^3_{\mathrm{tx}} [\, \chi_0 + (\hat{\mu}_4 - 2)\cdot \chi_4 \\ + (\hat{\mu}_4 - 2)^2 \cdot \chi_4' + \hat{\mu}_6 \cdot \chi_6 \,],\end{aligned} \tag{17}$$

where $\hat{\mu}_4$ and $\hat{\mu}_6$ are standardized moments of the input X and are discussed in Sec. III-B, and χ_0, χ_4, χ_4', and χ_6 represent real coefficients that represent the contributions of the fiber nonlinearities. Combining (16) and (17), the total noise variance σ^2_{eff} is

$$\begin{aligned}\sigma^2_{\mathrm{eff}} = \underbrace{\sigma^2_{\mathrm{ASE}} + P^3_{\mathrm{tx}}\cdot\chi_0}_{\text{modulation-independent}} \\ + \underbrace{P^3_{\mathrm{tx}}\left[(\hat{\mu}_4-2)\cdot\chi_4 + (\hat{\mu}_4-2)^2\cdot\chi_4' + \hat{\mu}_6\cdot\chi_6\right]}_{\text{modulation-dependent}},\end{aligned} \tag{18}$$

where we have split the overall noise into two terms. A modulation-independent noise contribution, given in the first line of (18), models ASE and partly NLI, and it is based solely on the system and fiber parameters, but not on the channel input. These two noise contributions are included in the classic GN model [35, Sec. VI]. The expression in the second line of (18) is a function of $\hat{\mu}_4$ and $\hat{\mu}_6$, which are functions of the channel input, and thus models the modulation-dependency of σ^2_{eff}.

TABLE II

Overview of $\hat{\mu}_4$ and $\hat{\mu}_6$ of (19) for different complex (2D) modulation formats and distributions

Modulation	P_X	$\hat{\mu}_4$	$\hat{\mu}_6$
M-PSK	uniform	1	1
16QAM	uniform	1.32	1.96
64QAM	uniform	1.381	2.226
256QAM	uniform	1.395	2.292
Continuous 2D	uniform	1.4	2.316
16QAM	**b)** of Table I	1.525	2.76
64QAM	**d)** of Table I	1.664	3.518
256QAM	**f)** of Table I	1.713	3.808
Continuous 2D	Gaussian	2	6

B. Standardized Moments

We have seen that σ^2_{eff} in (18) depends on the standardized moments $\hat{\mu}_4$ and $\hat{\mu}_6$. In general, the k^{th} standardized moment $\hat{\mu}_k$ of the channel input X is defined as

$$\hat{\mu}_k = \frac{\mathbb{E}[|X-\mathbb{E}[X]|^k]}{(\mathbb{E}[|X-\mathbb{E}[X]|^2])^{\frac{k}{2}}} = \mathbb{E}[|X|^k], \tag{19}$$

where the last equality in (19) holds because the channel input X is symmetric around the origin (see (13)), which gives $\mathbb{E}[X] = 0$, and X is normalized to unit energy, i.e., $\mathbb{E}[|X|^2] = 1$.

Table II shows the moments $\hat{\mu}_4$ and $\hat{\mu}_6$ for different modulation formats and PMFs. Constant-modulus modulation such as phase-shift keying (PSK) minimizes both moments. For uniform QAM, $\hat{\mu}_4$ and $\hat{\mu}_6$ increase with modulation order. The limit for complex uniform input is given by uniform QAM with infinitely many signal points, which corresponds to a continuous uniform input in 2D. When the input is shaped with an MB PMF that fulfills (12), e.g. with the mismatched distributions in Table I, $\hat{\mu}_4$ and $\hat{\mu}_6$ are larger than for a uniform input distribution. The respective maxima for these moments are again given by a complex continuous Gaussian density.

C. NLI Increase due to Shaping

The modulation-dependent coefficients χ_4, χ_4', and χ_6 in (18) together with the results in Table II give us an insight into how the choice of a particular modulation affects σ^2_{NLI}. Considering the first modulation-dependent term in (18), a small $\hat{\mu}_4$ corresponds to a decrease in $(\hat{\mu}_4-2)$ and thus, to less NLI. PSK formats, for example, minimize $\hat{\mu}_4$ and $\hat{\mu}_6$ and thus induce a minimum amount of NLI, which is why these formats can have superior performance to QAM in highly nonlinear channels, e.g., with inline dispersion management [40, Fig. 4]. On the other hand, distributions that are well-suited for the AWGN channel, such as the shaped PMFs in Table II, have

TABLE III
SYSTEM AND SIMULATION PARAMETERS

Parameter	Value
Modulation	64QAM
Input PMF	uniform and shaped
Polarization	dual-polarization
Symbol rate	28 GBaud
Pulse shape	root-raised cosine (RRC)
RRC roll-off	0.01
WDM channels	9
WDM spacing	30 GHz
Nonlinear coefficient γ	1.3 1/W/km
Dispersion	17 ps/nm/km
Attenuation α	0.2 dB/km
Length per span	100 km
Amplification	EDFA
EDFA noise figure	4 dB
Dispersion compensation	digital
Demapper statistics	c.s. Gaussian
SSFM step size	0.1 km
Oversampling factor	32
QAM symbols (SSFM sim.)	500,000
Samples (SPM-XPM model)	1,000,000

increased moments $\hat{\mu}_4$ and $\hat{\mu}_6$, which results in stronger NLI than uniform input. The interpretation of probabilistic shaping for the nonlinear fiber channel is then that a shaping gain can be obtained from the channel portion described by the linear noise contribution σ^2_{ASE} and the moment-independent term χ_0. Simultaneously, an increase in NLI is introduced by shaping as the modulation-dependent term of (18) becomes larger, and the optimal trade-off is not obvious. We will study this in detail in Sec. IV-E.

IV. Probabilistic Shaping of 64QAM for a Multi-Span Fiber System

In the following, we numerically evaluate the AIRs for a multi-span fiber link with uniform and shaped 64QAM input. The presented analysis focuses on the effect that shaping has on the fiber nonlinearities and thus, on the effective SNR. Transceiver impairments and further effects that require advanced DSP are not included in this work.

A. Fiber Simulations

A multi-span optical fiber system is simulated, with the main parameters given in Table III. 64QAM symbols are generated with the constant-composition distribution matcher [22] and shaped according to the specified input distribution. Pulse shaping is done digitally and the resulting signal is ideally transfered into the optical domain. The center WDM channel is the channel of interest and all WDM channels have the same PMF, yet with decorrelated symbol sequences. The propagation of the signal over each span fiber is simulated using the SSFM. After every span, the signal is amplified by an EDFA and ASE noise is added. At the receiver, the center WDM channel is filtered and ideally transfered into the digital domain. Chromatic dispersion is digitally compensated, a matched filter is applied, and the signal is downsampled. The channel SNR is computed as the average over both polarizations, and R_{BMD} per polarization is calculated as stated in (8) and we sum over both polarizations. For the BMD rate

TABLE IV
ASE NOISE AND NLI TERMS IN (18)

Noise term	Value
σ^2_{ASE}	17.85×10^{-6} W
χ_0	3.09×10^{4} W^{-2}
χ_4	1.05×10^{4} W^{-2}
χ'_4	-1.22×10^{2} W^{-2}
χ_6	1.29×10^{2} W^{-2}

estimation, 2D c.s. Gaussian statistics as given in (3) [41] with static mean values[3] are used.

B. Numerical Evaluation of SPM-XPM Model

The nonlinear terms χ_0, χ_4, χ'_4, and χ_6 in (18) are calculated via Monte Carlo simulations for which a ready-to-use web interface [42] and Matlab code [39, App.] is available. Note that (18) includes inter-channel and intra-channel effects as well as additional intra-channel terms that occur at very dense WDM spacings, as discussed in [37], [42]. We also note that virtually identical results are obtained for the sinc pulse shape described in [39, App.] and the narrow RRC filtering in this work.

For the parameters given in Table III and a transmission distance of 2000 km, the amplifier noise power σ^2_{ASE} and the nonlinear coefficients of (18) are given in Table IV. We observe that χ_0 and χ_4 are the dominant NLI contributions. The values of all NLI terms are used to compute the effective SNR_{eff} of (16). The BMD rate in (6) is computed by numerical integration. Note that χ'_4 can be negative, as we will see in Sec. IV-F, but due to its dependence on $(\hat{\mu}_4 - 2)^2$, its overall contribution to σ^2_{NLI} is also that an increased $\hat{\mu}_4$ leads to more NLI.

C. Reach Increase from Shaping

We compare R_{BMD} for 64QAM with uniform input, with a MB input PMF that is dependent on the transmission distance (and thus on the channel SNR, see Sec. II-B), and with the fixed PMF **d)** of Table I. Figure 4 shows R_{BMD} in bit/4D-sym for transmission distances from 1000 km to 3000 km in steps of 100 km. Results for SSFM simulations (markers) and for the SPM-XPM model (dashed lines) are shown, and we observe a good agreement between them. For every transmission distance, the launch power is varied with a granularity of 0.5 dB and the optimal power is used, which is -1.5 dBm or -1 dBm per WDM channel for all distances and input PMFs.

The channel SNR for uniform 64QAM is between 17.2 dB SNR for 1000 km and 12.35 dB SNR for 3000 km, and the SNR for each distance is used as shaping SNR for the SNR-dependent input PMF. Using this shaped input gives an AIR gain over uniform input for a fixed transmission distance or, equivalently, an increase in transmission distance for a fixed AIR. For example, shaping gives a 300 km reach improvement, from 2000 km to 2300 km, at an AIR of 8.86 bit/4D-sym.

[3]For static mean values, the centroids of the Gaussian distributions are identical to the sent constellation points x, as stated in (3). In contrast, using adaptive mean values [31] means that the centroids are calculated from the received symbols y_k.

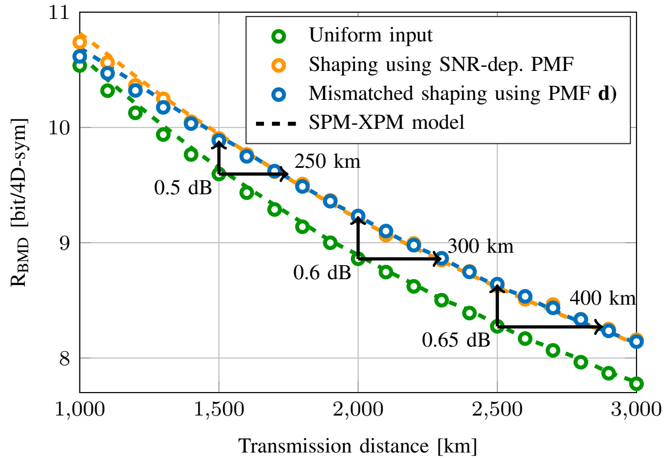

Fig. 4. R_{BMD} in bit/4D-sym vs. transmission distance in km for the SPM-XPM model of Sec. III (dashed lines) with 64QAM input and SSFM simulations (markers). The colors represent the different PMFs: uniform input (green), SNR-dependent shaped PMFs (orange), and the SNR-independent PMF **d)** of Table I (blue).

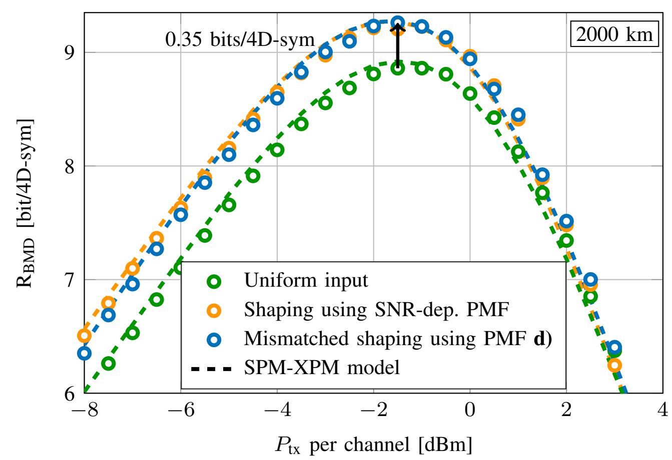

Fig. 5. R_{BMD} in bit/4D-sym vs. P_{tx} per channel in dBm for 64QAM. The SPM-XPM model results (dashed lines) for uniform input, SNR-dependent shaping and shaping with a fixed PMF are shown. For the latter, the distribution **d)** of Table I is used for all launch powers. SSFM simulation results (markers) validate the SPM-XPM model.

Similar gains are observed for all link lengths and in agreement with previous shaping simulations of a WDM system [15] [41, Sec. 3.5]. The AIR gains from shaping translate to sensitivity improvements of up to 0.65 dB, which is slightly below the maximum shaping gain of 0.8 dB seen for 64QAM in back-to-back experiments [18, Fig. 2] where no NLI is present, and larger gains are possible with higher-order modulation. For distances between 1400 km and 3000 km, the PMF **d)** gives identical gains as the shaped input that is matched to the SNR at every transmission distance. For smaller distances, a gap between mismatched and SNR-dependent shaping exists because the system is operated in the high-SNR regime beyond the channel SNR range of PMF **d)**, and switching to 256QAM is advisable.

D. AIR Gain of Shaped 64QAM at 2000 km Distance

The effects of shaping in the presence of fiber nonlinearities are investigated for a transmission distance of 2000 km (all other parameters as given Sec. IV-A). In Fig. 5, R_{BMD} in bit/4D-sym is shown vs. P_{tx} per channel in dBm. A good match between simulation results (markers) and the SPM-XPM model (dashed lines) is again observed. At the optimum launch power, a shaped input distribution gives an AIR improvement of 0.35 bit/4D-sym over uniform input. For all relevant launch powers, SNR-dependent shaping and shaping with the fixed PMF **d)** give identical gains. This shows again that it is sufficient to use one input distribution to realize the shaping gain for various transmit powers. We see from the SPM-XPM model that in the highly nonlinear regime, the shaping gain is significantly reduced and disappears for very high launch powers, which is due to the adverse ramifications of shaping. The sensitivity of these effects on $\mathrm{SNR_{eff}}$ and the AIRs is investigated next.

E. Sensitivity Analysis of Probabilistic Shaping

In the following, the sensitivity of NLI on probabilistic shaping is studied. The SNR mismatch between shaping SNR and channel SNR is chosen as a figure of merit for this analysis as it describes how strongly a QAM input is shaped with one single number that parametrizes a MB PMF. The SNR mismatch is denoted by Δ and calculated for each simulation run as the shaping SNR that is used at the transmitter minus the channel SNR that is estimated after the DSP. The chosen definition of Δ means that a large Δ, i.e., a large shaping SNR, corresponds to a distribution that is closer to uniform, while a small Δ represents a PMF that is strongly shaped. The mismatch Δ was varied by diverting from the channel SNR of uniform 64QAM at 2000 km, which is 14.17 dB, in steps of 0.1 dB. These values are used as shaping SNRs. In total, 100 full SSFM simulation runs with a transmission length of 2000 km were performed to gather sufficient statistics for Δ in the range of -4 dB to 6 dB.

1) Shaping Decreases SNR: Figure 6 shows the dependence of the channel SNR on Δ, with blue markers representing simulation results for shaped 64QAM over 2000 km and the solid curve being a linear fit to the simulations. In the considered range of Δ, we observe a good match of the simulation data to a linear fit. The results of the SPM-XPM model, shown as dashed curve, are within 0.05 dB of the fit and are an accurate approximation of the simulation results. Hence, the SPM-XPM model correctly predicts σ^2_{NLI} to grow, and thus the effective SNR to decrease, with increasing moments $\hat{\mu}_4$ and $\hat{\mu}_6$ that result from a decrease in Δ. The AWGN reference (dotted curve), which also corresponds to constant channel SNR of uniform 64QAM input in SSFM simulations, confirms that for a linear channel without NLI, the channel SNR does not depend on the input distribution. It is, however, important to realize that the increase in SNR that is observed for increasing Δ does not imply a gain in AIR, as we will show next.

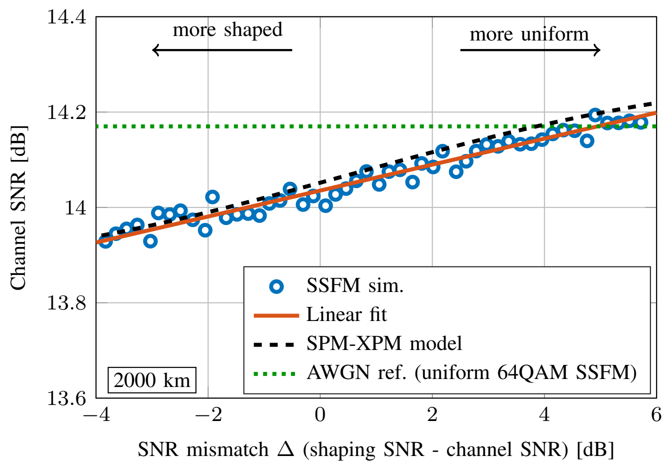

Fig. 6. The channel SNR measured after DSP vs. the SNR mismatch Δ, both in dB. The SPM-XPM model (dashed line) shows the dependence of the channel SNR on the input. This behavior would not be present in a linear AWGN channel (dotted line), which also represents the channel SNR for SSFM simulations with uniform 64QAM. Further SSFM simulations of shaped 64QAM over 2000 km at the optimum launch power are also shown (circles), where each marker represent a simulation run. The solid orange line is a linear fit of the simulation results.

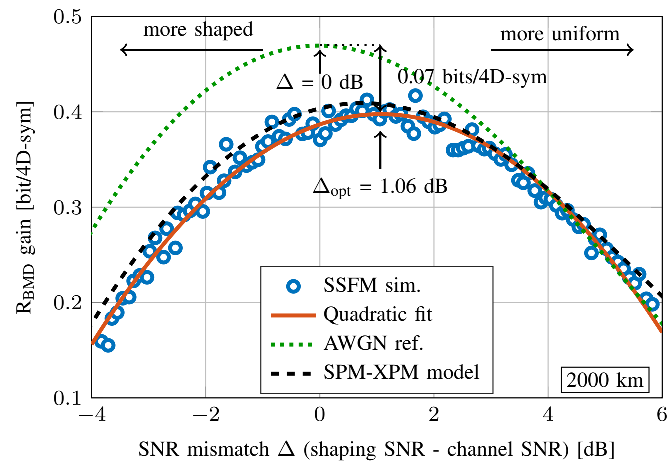

Fig. 7. $\mathrm{R_{BMD}}$ gain in bit/4D-sym of shaped 64QAM over uniform 64QAM vs. SNR mismatch Δ in dB. The solid line is a quadratic fit to the simulations and has its maximum at Δ_{opt} = 1.06 dB. The SPM-XPM model (dashed line) correctly predicts the trend of the simulations. The dotted line shows as a reference the BMD rate gain for an AWGN channel with 14.17 dB SNR where the maximum shaping gain is obtained in the case of no SNR mismatch, i.e., Δ = 0 dB.

2) Shaping Increases AIR: In Fig. 7, the sensitivity of the shaping gain is investigated by plotting the $\mathrm{R_{BMD}}$ gain over uniform 64QAM (with a $\mathrm{R_{BMD}}$ of 8.86 bit/4D-sym) as a function of the SNR mismatch Δ. Blue markers indicate SSFM simulations and a quadratic fit to the simulation results is given as solid line. The fitted parabola is relatively flat around its peak of 0.4 bit/4D-sym, and significant shaping gains of more than 0.3 bit/4D-sym are obtained for Δ from approximately -2 dB to 4 dB. Hence, only small penalties in shaping gain in comparison to matched shaping are to be expected if the SNR mismatch is within a range of several dB. This is again supported by the SPM-XPM model (dashed curve) that accurately predicts this behavior.

Another interesting aspect of Fig. 7 is that the maximum shaping gain, according to the quadratic fit of the simulation data, is found at Δ_{opt} = 1.06 dB. In the absence of fiber nonlinearities, i.e., for an AWGN channel, no mismatch, i.e., Δ = 0 dB, is expected to be optimal, as confirmed by the AWGN reference (dotted curve). As explained by the SPM-XPM model in Sec. III-A and also in the context of Fig. 6, a shaped input causes stronger NLI than uniform input. Thus, Δ_{opt} is expected to be larger than 0 dB since a positive Δ indicates a more uniform-like input that introduces less NLI. The magnitude of this effect, however, is very small and significant shaping gains are observed around the optimum SNR mismatch. The NLI increase due to shaping is also the reason for the difference in $\mathrm{R_{BMD}}$ gain of the optical fiber simulations and the AWGN channel. This gap of approx. 0.07 bit/4D-sym at the optimum Δ disappears for large positive Δ's because in this range, a uniform input is approached and increased NLI due to shaping does no longer occur. The effect that causes the gap between AWGN and fiber simulations can be considered as a shaping penalty due to increased NLI, and we observe its magnitude to be small.

F. Optimized Shaping for the Nonlinear Fiber Channel

So far, we have restricted our analysis of probabilistic shaping to distributions of the MB family, see (14), and considered only 1D PMFs that were then extended to 2D. We have shown that these inputs are an excellent choice for the AWGN channel and large shaping gains are also obtained for the optical channel. It is, however, not clear whether we can find a better shaped input PMF for the nonlinear fiber channel as there might be input distributions that have a better trade-off between the shaping gain and the shaping penalty due an NLI increase. In the following, we use the SPM-XPM model of Sec. III-A as channel for the optimization problem (12) and numerically search for the shaped inputs that give the largest $\mathrm{R_{BMD}}$. We consider inputs in 1D (which are then extended to 2D), and also optimize PMFs directly in 2D. This 2D approach gives us more degrees of freedom in the optimization problem and allows us to consider any probabilistically shaped input in 2D, including multi-ring constellations [43, Sec. IV-B]. The benefit of 1D PMFs, however, is that the receiver can operate in 1D, without loss of information for a symmetric channel, which leads to a reduced demapper complexity compared to the 2D case.

The system described in Sec. IV-A with a distance of 2000 km is considered for the optimization. In Fig. 8, two 1D PMFs are shown that are the respective result of the optimization problem in the linear regime at $P_{\mathrm{tx}} = -8$ dBm and in the nonlinear regime at P_{tx} =3 dBm. Details on the PMFs (and for completeness the PMFs at the optimum P_{tx}) are given in Table V. Despite the two different launch powers, their effective SNRs $\mathrm{SNR_{eff}}$ are virtually identical, and using MB PMFs that are based solely on the channel SNR would result in the same PMF for both launch powers. In the considered optimization problem, this restriction is lifted, and we observe from Fig. 8 that the PMF for 3 dBm is less shaped than the one for -8 dBm. This illustrates that strong shaping

TABLE V
INPUT PMFS OPTIMIZED WITH THE SPM-XPM MODEL

P_{tx} per ch.	One-sided 1D PMF	$\hat{\mu}_4$	$\hat{\mu}_6$	SNR_{eff}
−8 dBm	[0.2725, 0.16, 0.055, 0.0125]	1.918	5.145	9.46 dB
3 dBm	[0.1625, 0.16, 0.13, 0.0475]	1.511	2.822	9.49 dB
−1.5 dBm	[0.195, 0.1625, 0.1, 0.0425]	1.642	3.432	14.09 dB
−1.5 dBm	2D PMF: see Fig. 9 (inset)	1.599	3.197	14.12 dB

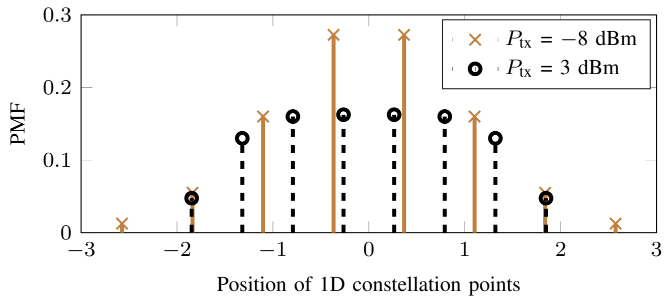

Fig. 8. Optimized 1D PAM PMFs for $P_{tx} = -8$ dBm (solid, crosses) and $P_{tx} = 3$ dBm (dashed, circles). Although these two power levels have approximately the same SNR_{eff}, the PMF for 3 dBm is less shaped to avoid increased NLI.

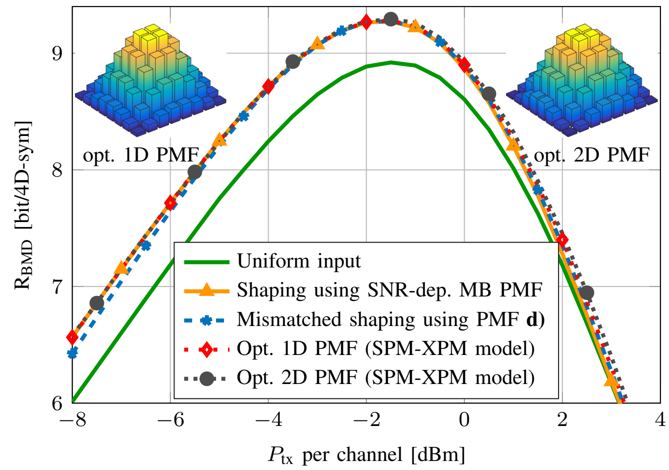

Fig. 9. R_{BMD} in bit/4D-sym for 64QAM vs. P_{tx} per channel in dBm for the SPM-XPM model. The AIRs for the 1D-optimized input (red dotted), for the 2D-optimized PMF (gray dotted) and for all other shaped inputs lie on top of each other over a wide range of launch powers. Inset: The optimized 1D PMF in its 2D representation and the 2D PMF, each for P_{tx} of −1.5 dBm.

is avoided at high power levels when a PMF optimization with the SPM-XPM model is carried out.

In Fig. 9, R_{BMD} is shown vs. P_{tx} per channel for different input distributions of 64QAM. All results are obtained from the SPM-XPM model. The dotted curves show R_{BMD} for a 1D PMF (red) and a 2D PMF (gray), both optimized with the SPM-XPM model, and the AIRs for uniform, MB-shaped input, PMF **d)** are also included. The two optimized PMFs give identical R_{BMD}, and their shape is very similar as the insets in Fig. 9 show. We conclude that for the considered system, there is virtually no benefit from using the optimized 2D input. Additionally, the MB shaped input gives identical gains to the 1D-optimized input at low powers and around the optimal power. It is only in the high-power regime that slightly increased AIRs are obtained with the optimized input. This indicates that, also for a multi-span fiber channel, the shaping gain is very insensitive to variations in the input distribution, and an optimized input does not give larger shaping gains than a MB PMF. It is in fact sufficient for the considered system to simply use the fixed input distribution **d)** from Table I to virtually obtain the maximum shaping gain.

V. CONCLUSIONS

In this work, we have studied probabilistic shaping for long-haul optical fiber systems both via numerical simulations and via a GN model. We based our analysis on AWGN results that show that only two input PMFs from the family of Maxwell-Boltzmann distributions are sufficient per QAM format to realize large shaping gains over a wide range of SNRs. We have found that these mismatched shaped distributions are also an excellent choice for applying shaping to a multi-span fiber system. Using one input distribution for 64QAM, large shaping gains are reported from transmission distances between 1400 km to 3000 km. For a fixed distance of 2000 km, we have studied the impact of probabilistic shaping with Maxwell-Boltzmann distributions and other PMFs. The adverse effects of shaping in the presence of modulation-dependent nonlinear effects of a WDM system have been shown to be present. An NLI penalty from shaping is found to be minor around the optimal launch power in a multi-span system. This means that for the considered system, one input PMF for 64QAM virtually gives the maximum shaping gain and an optimization for the fiber channel is not necessary. This could greatly simplify the implementation and design of probabilistic shaping in practical optical fiber systems. We expect similar results for other QAM formats such as 16QAM or 256QAM when they are used in fiber systems that are comparable to the ones in this work. We have also found that the GN model is in excellent agreement with the SSFM results, confirming its accuracy for shaped QAM input.

For highly nonlinear fiber links, e.g., with in-line dispersion management or single-span links with high power, further optimizations of the shaping scheme can be beneficial to give a large shaping gain but at the same time incur low NLI. Additionally, instead of shaping on a per-symbol basis, constellation shaping over several time slots to exploit the temporal correlations by XPM is an interesting future step to increase SE. Also, optimizing distributions in 4D could be beneficial for highly nonlinear polarization-multiplexed fiber links.

VI. ACKNOWLEDGMENTS

The authors would like to thank Prof. Frank Kschischang (University of Toronto) for encouraging us to use the SPM-XPM model to study probabilistic shaping for the nonlinear fiber channel.

REFERENCES

[1] D. J. Richardson, “Filling the light pipe,” *Science*, vol. 330, no. 6002, pp. 327–328, Oct. 2010.

[2] R. J. Essiambre and R. W. Tkach, “Capacity trends and limits of optical communication networks,” *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1035–1055, May 2012.

[3] P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey, "Maximizing the optical network capacity," *Philosophical Transactions of the Royal Society of London A*, vol. 374, no. 2062, Jan. 2016.

[4] R. Maher, A. Alvarado, D. Lavery, and P. Bayvel, "Modulation order and code rate optimisation for digital coherent transceivers using generalised mutual information," in *Proc. European Conference and Exhibition on Optical Communication (ECOC)*. Valencia, Spain: Paper Mo.3.3.4, Sep. 2015.

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

[6] U. Wachsmann, R. F. H. Fisher, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[7] I. B. Djordjevic, H. G. Batshon, L. Xu, and T. Wang, "Coded polarization-multiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission," in *Proc. Optical Fiber Communication Conference (OFC)*. San Diego, CA, USA: Paper OMK2, Mar. 2010.

[8] H. G. Batshon, I. B. Djordjevic, L. Xu, and T. Wang, "Iterative polar quantization based modulation to achieve channel capacity in ultra-high-speed optical communication systems," *IEEE Photonics Journal*, vol. 2, no. 4, pp. 593–599, Aug. 2010.

[9] T. Liu and I. B. Djordjevic, "Multidimensional optimal signal constellation sets and symbol mappings for block-interleaved coded-modulation enabling ultrahigh-speed optical transport," *IEEE Photonics Journal*, vol. 6, no. 4, pp. 1–14, Aug. 2014.

[10] T. H. Lotz, X. Liu, S. Chandrasekhar, P. J. Winzer, H. Haunstein, S. Randel, S. Corteselli, B. Zhu, and D. W. Peckham, "Coded PDM-OFDM transmission with shaped 256-iterative-polar-modulation achieving 11.15-b/s/Hz intrachannel spectral efficiency and 800-km reach," *Journal of Lightwave Technology*, vol. 31, no. 4, pp. 538–545, Feb. 2013.

[11] J. Estaran, D. Zibar, A. Caballero, C. Peucheret, and I. T. Monroy, "Experimental demonstration of capacity-achieving phase-shifted superposition modulation," in *Proc. European Conference on Optical Communications (ECOC)*. London, UK: Paper We.4.D.5, Sep. 2013.

[12] B. P. Smith and F. R. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," *Journal of Lightwave Technology*, vol. 30, no. 13, pp. 2047–2053, Jul. 2012.

[13] L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson, "Rate-adaptive coded modulation for fiber-optic communications," *Journal of Lightwave Technology*, vol. 32, no. 2, pp. 333–343, Jan. 2014.

[14] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photonics Technology Letters*, vol. 26, no. 23, pp. 2407–2410, Dec. 2014.

[15] T. Fehenberger, G. Böcherer, A. Alvarado, and N. Hanik, "LDPC coded modulation with probabilistic shaping for optical fiber systems," in *Proc. Optical Fiber Communication Conference (OFC)*. Los Angeles, CA, USA: Paper Th.2.A.23, Mar. 2015.

[16] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," in *Proc. European Conference and Exhibition on Optical Communication (ECOC)*. Valencia,Spain: Paper PDP.3.4, Sep. 2015.

[17] C. Diniz, J. H. Junior, A. Souza, T. Lima, R. Lopes, S. Rossi, M. Garrich, J. D. Reis, D. Arantes, J. Oliveira, and D. A. Mello, "Network cost savings enabled by probabilistic shaping in DP-16QAM 200-Gb/s systems," in *Proc. Optical Fiber Communication Conference (OFC)*. Anaheim, CA, USA: Paper Tu3F.7, Mar. 2016.

[18] T. Fehenberger, D. Lavery, R. Maher, A. Alvarado, P. Bayvel, and N. Hanik, "Sensitivity gains by mismatched probabilistic shaping for optical communication systems," *IEEE Photonics Technology Letters*, vol. 28, no. 7, pp. 786–789, Apr. 2016.

[19] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[20] M. P. Yankov, F. Da Ros, E. P. da Silva, S. Forchhammer, K. J. Larsen, L. K. Oxenløwe, M. Galili, and D. Zibar, "Constellation shaping for WDM systems using 256QAM/1024QAM with probabilistic optimization," *arXiv preprint arXiv:1603.07327*, Mar. 2016.

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

[22] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[23] L. Szczecinski and A. Alvarado, *Bit-interleaved coded modulation: fundamentals, analysis and design*. John Wiley & Sons, 2015.

[24] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014.

[25] ——, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Optics Express*, vol. 21, no. 22, pp. 25 685–25 699, Oct. 2013.

[26] A. Alvarado and E. Agrell, "Four-dimensional coded modulation with bit-wise decoders for future optical communications," *Journal of Lightwave Technology*, vol. 33, no. 10, pp. 1993–2003, May 2015.

[27] G. Böcherer, "Achievable rates for shaped bit-metric decoding," May 2016. [Online]. Available: http://arxiv.org/abs/1410.8075

[28] A. Ganti, A. Lapidoth, and L. Emre Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[29] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[30] T. Fehenberger, T. A. Eriksson, A. Alvarado, M. Karlsson, E. Agrell, and N. Hanik, "Improved achievable information rates by optimized four-dimensional demappers in optical transmission experiments," in *Proc. Optical Fiber Communication Conference (OFC)*. Anaheim, CA, USA: Paper W1I.4, Mar. 2016.

[31] T. A. Eriksson, T. Fehenberger, P. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *Journal of Lightwave Technology*, vol. 34, no. 9, pp. 2256–2266, May 2016.

[32] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[33] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[34] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[35] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694–721, Feb. 2014.

[36] A. Mecozzi and R. J. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *Journal of Lightwave Technology*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.

[37] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Optics Express*, vol. 22, no. 13, pp. 16 335–16 362, Jun. 2014.

[38] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.

[39] ——, "Accumulation of nonlinear interference noise in fiber-optic systems," *Optics Express*, vol. 22, no. 12, pp. 14 199–14 211, Jun. 2014.

[40] T. Koike-Akino, K. Kojima, D. S. Millar, K. Parsons, T. Yoshida, and T. Sugihara, "Pareto-efficient set of modulation and coding based on RGMI in nonlinear fiber transmissions."

[41] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Optics Express*, vol. 23, no. 7, pp. 9183–9191, Apr. 2015.

[42] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "NLIN Wizard," accessed April 4, 2016. [Online]. Available: http://nlinwizard.eng.tau.ac.il

[43] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, Feb. 2010.