On Optimally Shaped Signals for Nonlinear Frequency Division Multiplexed Fiber Systems

Citation for published version:
Chen, Y, Baniasadi, A & Safari, M 2023, 'On Optimally Shaped Signals for Nonlinear Frequency Division Multiplexed Fiber Systems', *IEEE Transactions on Communications*, vol. 71, no. 9, pp. 5379-5391. https://doi.org/10.1109/TCOMM.2023.3285750

Digital Object Identifier (DOI):
10.1109/TCOMM.2023.3285750

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IEEE Transactions on Communications

On Optimally Shaped Signals for Nonlinear Frequency Division Multiplexed Fiber Systems

Yu Chen, Mohammadamin Baniasadi, and Majid Safari

***Abstract*—An approximated channel model is proposed for direct signaling on the continuous spectrum of a nonlinear frequency division multiplexed (NFDM) communication system, describing the effect of noise and nonlinearity at the receiver. The optimal input distribution that maximizes the mutual information of the proposed approximated channel under the peak amplitude constraint is then studied. We present that, considering the input-dependency of the noise, the conventional amplitude-constrained constellation designs can be geometrically shaped to provide significant mutual information gains. However, it is observed that further probabilistic shaping and constellation size optimization can provide only limited additional gains beyond the best geometrically shaped benchmark scheme, i.e., 64 Amplitude Phase Shift Keying. Then, an approximated channel model that neglects the correlation between subcarriers is proposed for the matched filtered signaling system, based on which the input constellation is geometrically shaped. We demonstrate that although the inter-subcarrier interference in the filtered case is neglected in the channel model, the shaping of the matched filtered case can provide promising gains in mismatch capacity over the unfiltered scenario.**

***Index Terms*—Nonlinear Frequency Division Multiplexing, Continuous Spectrum, Geometric Shaping, Probabilistic Shaping**

I. Introduction

IT is shown that the achievable data rate of a long-haul fiber communication system will reach saturation beyond a launch power limit due to the signal dependent Kerr nonlinearity [2]. To overcome this limitation and meet the increasing data demand, various techniques such as digital backpropagation, Volterra series nonlinear equalizers are proposed to mitigate or compensate the nonlinearity as reviewed in [3]. Recently, the novel application of nonlinear Fourier transform (NFT), or inverse scattering, which can be described as a nonlinear extension of the conventional Fourier transform (FT) is suggested to linearize the evolution of the signal in fiber channels described by nonlinear Schrödinger equation (NLSE) [4]. Using NFT, the temporal degrees of freedom of an arbitrary optical signal are mapped into two distinct spectra in the so-called nonlinear frequency spectrum, namely, the discrete spectrum (DS) and the continuous spectrum (CS) [4]. If the nonlinear frequency spectrum is employed to modulate the signal, the system is commonly known as a nonlinear frequency division multiplexed (NFDM) system [4].

The authors are with the Institute for Digital Communication, University of Edinburgh, Edinburgh, UK, contact e-mail: yu.chen@ed.ac.uk. Research supported by China Scholarship Council, China and a Leverhulme Trust Research Project Grant. Part of this work was presented in International Symposium on Information Theory (ISIT) 2022, Espoo, Finland [1].

The DS domain corresponds to the solitonic part of the time domain signal, which possesses invariant property against the pulse broadening dispersion effect and the pulse compressing nonlinearity. Such pulse invariant property can only be sustained without noise, which is unavoidable in long distance transmission with optical amplifiers. Considering a signal dependent non-Gaussian noise model [5], the capacity estimation for the system is studied in various works [6]–[8]. Most of the mentioned works neglected the inter-soliton interaction [9], which further limits the performance of the soliton transmission as reported in [8].

In contrast, the CS domain, which corresponds to non-solitonic radiation, has a signal space similar to that of orthogonal frequency division multiplexing (OFDM), i.e., linear frequency domain. This allows the direct adoption of conventional modulation techniques originally designed for linear systems to be used for modulating data on CS. Thus, the CS domain is the focus of this work. Two aspects of the CS domain can be used for modulation, namely, CS spectrum modulation (also known as ρ-modulation) and b-modulation, where $\rho = b/a$. In some works, b-modulation is favored because of its compact support, and better spectrum efficiency [10], [11]. As one of the main NFT-based modulation techniques in the literature, in this paper, CS spectrum modulation is considered. This avoids the energy barrier issue caused by the unity constraint of $|a|^2 + |b|^2 = 1$ for the b-modulation [12], which could lead to singularity in the channel modeling. Though the energy barrier could be avoided using either one-to-one nonlinear mapping before modulation [13] or tailored pulse shaping [12]. It is concluded with a dual-polarization system that the nonlinear mapping reduces the benefit of spectral efficiency [13].

In the absence of noise, the propagation of CS signal over the fiber is well defined. However, when noise is introduced, the CS transmission channel model still remains an open problem despite the recent progress in [14]–[16]. Among all relevant works, it is pointed out that the channel noise of the CS domain shows strong dependence on the input signal. The statistics of the CS noise is described in both [14] and [15] up to the second moment, but the closed-form expression of the conditional probability density function (PDF) of the noisy signal that defines the channel is not available. To estimate the channel capacity, the Pinsker formula is used to lower bound the capacity of a nonlinear inverse synthesis (NIS) CS modulation system [14]. The application of the Pinsker formula can be a useful capacity lower bounding tool when the exact channel model is not known [14]. But it should be emphasized that the resulting bound is not necessarily a tight one [14], [17]. In [15], the real and imaginary channels of the complex CS spectrum are assumed to be two independent real channels, allowing variance normalizing transform (VNT) to be performed on each channel [18]. After VNT, each individual channel is approximated with an additive white Gaussian noise (AWGN) model, hence simplifying the capacity estimation under peak power constraint [19]. In [20], the mentioned VNT technique is employed to generate a nonuniform spacing constellation, then the gain in signaling over CS is compared with the other signaling methods like NIS and CS domain matched filtering.

In this work, we explore the optimal achievable mutual information (MI) of the CS modulated system with and without matched filter in the CS domain. At first, an approximated channel model is proposed combining the channel statistics based on perturbation theory [14], VNT, and also some insights from numerical simulations for both systems. Then, the MI of the approximated channel model is maximized under peak amplitude constraint to obtain the optimal discrete input distribution with a finite number of mass points. For the system without the CS domain filter, two types of input distribution shaping are considered in this work, namely geometric shaping (GS) and full optimization (FO). The GS shaping attempts to maximize the MI by optimizing the positions of the mass points of the input distribution, assuming an equiprobable distribution with a given distribution size. The FO scheme maximizes the MI by performing probabilistic geometric hybrid shaping for a number of distribution sizes. Additionally, the Pinsker formula of the channel model and the mismatch capacity employing the optimized constellation are also produced. The results suggest that for the underlying signal dependent channel considered in this work, the additional probabilistic shaping gain of FO is insignificant when compared with the best geometrically shaped scheme, i.e., 64 Amplitude Phase Shift Keying (APSK). Relying on this result, only geometric shaping is performed to maximize the MI for the approximated filtered channel model. Note that the developed signal shaping methods are based on the proposed approximated channel model rather than the actual NLSE model. Therefore, mismatch capacities are estimated for both systems with the Monte Carlo method to provide capacity lower bounds. Although introducing matched filtering in nonlinear frequency domain could result in extra interference, the gain in capacity lower bounds implies that filtering is still beneficial in terms of removing excessive noise.

II. Channel Model

The pulse evolution in an ideal distributed Raman amplified standard single mode fiber is governed by the stochastic NLSE,

$$jq_z(t,z) = q_{tt}(t,z) + 2|q(t,z)|^2 q(t,z) + n(t,z), \quad (1)$$

where $q(t,z)$ is the normalized complex envelope of the optical field, $n(t,z)$ represents the normalized amplifier spontaneous emission (ASE) noise (both normalized with $1/\gamma L_{\rm D}$), while t and z stand for the unitless time (normalized with T_0) and the propagation distance (normalized with the dispersion length $L_{\rm D} = 2T_0^2/|\beta_2|$). The autocorrelation of the normalized ASE noise is $\mathsf{E}[n(t,z)n^*(t',z')] = \sigma^2\delta(t-t')\delta(z-z')$, where $\sigma^2 = \alpha h\nu_0 K_{\rm T}\frac{2\gamma T_0^3}{|\beta_2|^2}$. The coefficients α, β_2 and γ describe the fiber loss, the group velocity dispersion and the Kerr nonlinearity, respectively. In addition, the $h\nu_0$ denotes the photon energy, and $K_{\rm T}$ is the phonon occupancy factor. Performing NFT on the pulse $q(t,z)$ will decompose the signal into CS $\rho(\lambda,z)$ and DS $\{\lambda_m(z)_{m=1}^M, C_m(z)_{m=1}^M\}$. In a noise-free propagation, the evolution of the CS is described by a linear phasor, i.e. $\rho(\lambda, z=l) = e^{4j\lambda^2 l}\rho(\lambda, z=0)$ [4]. Taking advantage of this property, the transmitted information symbols are first encoded directly on the CS shape $\rho(\lambda,0)$ via certain pulse shaping and inverse NFT (INFT) is taken to convert it into the time domain pulse $q(t,0)$. The pulse is then transmitted through a standard single mode fiber, where NFT is performed at the receiver to acquire the CS $\rho(\lambda,l)$ of the received pulse $q(t,l)$. After equalization with $e^{-4j\lambda^2 l}$, the received information symbol can be extracted by down-sampling the equalized CS.

A. Channel Statistics and Approximated Model

In the proposed system, the information symbols are directly modulated at different nonlinear frequencies, allowing them to be detected by down-sampling the equalized CS. In [14], the continuous channel statistics is discussed using perturbative theory. If one denotes the equalized output as Y_λ and the input as X_λ, the input-output relationship of the system is given as [14]

$$Y_\lambda = e^{-4j\lambda^2 l}\rho(\lambda,l) = X_\lambda + N_\lambda, \quad (2)$$

where the conditional statistics of the zero mean noise N_λ given X_λ are as [14]

$$\begin{aligned}\mathsf{E}\left[(Y_\lambda - X_\lambda)(Y_{\lambda'} - X_{\lambda'})^*\right] &= \mathsf{E}\left[N_\lambda N_{\lambda'}^*\right] \\ &= \sigma^2\delta(\lambda-\lambda')(1+|X_\lambda|^2+|X_\lambda|^4), \quad (3)\\ \mathsf{E}\left[(Y_\lambda - X_\lambda)(Y_{\lambda'} - X_{\lambda'})\right] &= \mathsf{E}\left[N_\lambda N_{\lambda'}\right] \\ &= \sigma^2\delta(\lambda-\lambda')X_\lambda^2, \quad (4)\end{aligned}$$

where σ^2 denotes the unitless spectral density per propagation length as in (1), $(\cdot)^*$ denotes complex conjugation, and $\delta(\cdot)$ denotes the Dirac-Delta function. Note that the condition is omitted for simplicity, the X_λ should be considered given unless specified otherwise. The δ function in the correlation indicates that the noise at one λ is uncorrelated to that at another λ'. It is assumed that when sufficiently sampled, the discretized CS preserves its continuous nature well, hence approximately, the inter-subcarrier interference (ISI) could be neglected. Detailed discussion about this approximation is given in the Appendix A. Employing such approximation, the statistics of the noise conditional on the input X_λ are written as

$$\mathsf{E}\left[|Y_\lambda - X_\lambda|^2\right] = \mathsf{E}\left[|N_\lambda|^2\right] = \sigma_{\rm N}^2(1+|X_\lambda|^2+|X_\lambda|^4), \quad (5)$$

$$\mathsf{E}\left[(Y_\lambda - X_\lambda)^2\right] = \mathsf{E}\left[N_\lambda^2\right] = \sigma_{\rm N}^2 X_\lambda^2, \quad (6)$$

where the $\sigma_{\rm N}^2 = \sigma^2 l t_{\rm w}$ corresponds to the received noise power when the transmitted symbol $X_\lambda = 0$, l denotes the

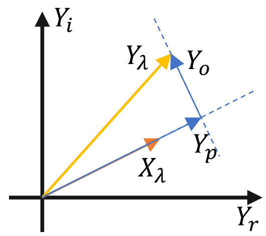


Fig. 1. Decomposition operation of the received symbol Y_λ into the parallel and orthogonal direction of the transmitted symbol X_λ.

unitless propagation distance and t_{w} indicates the unitless time domain pulse width. The σ_{N}^2 can be calculated using Parseval's theorem and the FT approximation of NFT when the L_1 norm of the signal is small without solitonic signal [4].

Within the reviewed literature, although the statistics of the channel noise is available up to the second moment [14], [15], an accurate channel model with closed-form expression for the CS NFDM system remains unknown. To allow further analysis, approximations should be made to derive the closed-form channel law. Inspired by the Pinsker formula, a noncircular Gaussian (NCG) model with the available first and second order statistics can provide a closed-form expression, a more accurate channel law can be possibly found. In this section, a more accurate channel model compared to the benchmark NCG model is proposed by employing VNT after a change of basis. The advantages of our proposed model over NCG are demonstrated through numerical simulations later.

In [15], [20], the correlation between the real and imaginary parts of the channel is reported with numerical simulation results. Rewriting the noise statistics (5) and (6) with the real and imaginary parts as

$$\mathsf{E}[N_{\mathrm{r}}^2] = \frac{\sigma_{\mathrm{N}}^2}{2}\left(1 + 2X_{\mathrm{r}}^2 + 2X_{\mathrm{r}}^2 X_{\mathrm{i}}^2 + X_{\mathrm{r}}^4 + X_{\mathrm{i}}^4\right), \tag{7}$$

$$\mathsf{E}[N_{\mathrm{i}}^2] = \frac{\sigma_{\mathrm{N}}^2}{2}\left(1 + 2X_{\mathrm{i}}^2 + 2X_{\mathrm{r}}^2 X_{\mathrm{i}}^2 + X_{\mathrm{r}}^4 + X_{\mathrm{i}}^4\right), \tag{8}$$

the correlation between them is then given as

$$\mathsf{E}[N_{\mathrm{r}} N_{\mathrm{i}}] = \sigma_{\mathrm{N}}^2 X_{\mathrm{r}} X_{\mathrm{i}}, \tag{9}$$

where $Y_\lambda = Y_{\mathrm{r}} + jY_{\mathrm{i}} = X_\lambda + N_\lambda = (X_{\mathrm{r}} + N_{\mathrm{r}}) + j(X_{\mathrm{i}} + N_{\mathrm{i}})$. Despite the non-zero signal dependent correlation (9), it is pointed out in [20] that the statistics of the phase noise is dependent on the transmitted symbol amplitude rather than its phase. The insight reveals a set of signal dependent decomposition where the decomposed components are uncorrelated to each other. One could decompose the received symbol Y_λ into the parallel and orthogonal directions of X_λ as

$$Y_\lambda = Y_{\mathrm{p}} \cdot \frac{X_{\mathrm{r}} + jX_{\mathrm{i}}}{|X_\lambda|} + Y_{\mathrm{o}} \cdot \frac{-X_{\mathrm{i}} + jX_{\mathrm{r}}}{|X_\lambda|}, \tag{10}$$

where $Y_{\mathrm{p}} = \frac{Y_{\mathrm{r}} X_{\mathrm{r}} + Y_{\mathrm{i}} X_{\mathrm{i}}}{|X_\lambda|}$ denotes the component on the parallel direction and $Y_{\mathrm{o}} = \frac{-Y_{\mathrm{r}} X_{\mathrm{i}} + Y_{\mathrm{i}} X_{\mathrm{r}}}{|X_\lambda|}$ denotes the component on the orthogonal direction as shown in the Fig. 1.

Taking the conditional expectation given X_λ over the parallel component Y_{p} allows the statistics to be derived as

$$\mathsf{E}[Y_{\mathrm{p}}] = |X_\lambda|, \quad \mathrm{var}[Y_{\mathrm{p}}] = \frac{\sigma_{\mathrm{N}}^2}{2}\left(1 + |X_\lambda|^2\right)^2. \tag{11}$$

Similarly, the conditional statistics of the orthogonal component Y_{o} are obtained as

$$\mathsf{E}[Y_{\mathrm{o}}] = 0, \quad \mathrm{var}[Y_{\mathrm{o}}] = \frac{\sigma_{\mathrm{N}}^2}{2}\left(1 + |X_\lambda|^4\right). \tag{12}$$

Note that the condition on the input X_λ is omitted for simplicity. In terms of the correlation between the decomposed components, it is shown that the parallel component is uncorrelated to the orthogonal component as

$$\mathsf{E}[Y_{\mathrm{p}} Y_{\mathrm{o}}] = \frac{X_{\mathrm{r}} X_{\mathrm{i}}}{|X_\lambda^2|}\mathsf{E}[N_{\mathrm{r}}^2 - N_{\mathrm{i}}^2] - \frac{X_{\mathrm{r}}^2 - X_{\mathrm{i}}^2}{|X_\lambda^2|}\mathsf{E}[N_{\mathrm{r}} N_{\mathrm{i}}] = 0. \tag{13}$$

If the uncorrelatedness is extended to assume that the two components are independent, the complex channel condition on a given input can be decomposed into two independent channels. The joint distribution of the decomposed channels is hence given as the product of their corresponding marginal distributions as

$$P(Y_{\mathrm{p}}, Y_{\mathrm{o}} | X_\lambda) = P(Y_{\mathrm{p}} | X_\lambda) P(Y_{\mathrm{o}} | X_\lambda). \tag{14}$$

As emphasized previously, even though the conditional statistics of the decomposed components are obtained using the perturbative theory results [14], the accurate models with closed-form expression for the decomposed components are still unavailable. In order to further the capacity analysis by performing MI maximization, additional approximations for the channel model should be made. For the parallel component Y_{p}, the dependence between the mean $\mathsf{E}[Y_{\mathrm{p}}]$ and the variance $\mathrm{var}[Y_{\mathrm{p}}]$ reveals the potential of employing VNT [15].

Lemma 1: Given a real random variable Y with mean μ_Y and mean dependent variance $\sigma_Y^2 = f(\mu_Y)$, the VNT transformed random variable $R = T(Y)$ should have mean $\mu_R \approx T(\mu_Y)$, and variance $\sigma_R^2 \approx 1$, where VNT $T(\cdot)$ is defined as

$$T(u) = \int \frac{1}{\sqrt{f(u)}} du. \tag{15}$$

Proof: Consider a transformed random variable $R = T(Y)$ and applied Taylor expansion around $Y = \mu_Y$, then R can be approximated with

$$R = T(Y) \approx T(\mu_Y) + T'(\mu_Y)(Y - \mu_Y), \tag{16}$$

where the higher order terms are omitted. The mean and variance of R are then derived as

$$\mathsf{E}[R] \approx \mathsf{E}[T(\mu_Y) + T'(\mu_Y)(Y - \mu_Y)] = T(\mu_Y), \tag{17}$$

$$\mathrm{var}[R] \approx {T'}^2(\mu_Y)\sigma_Y^2 = {T'}^2(\mu_Y) f(\mu_Y). \tag{18}$$

It is clear that the transformation $T(\cdot)$ normalizes the mean dependent variance to 1. Note that the variance could be normalized to any other convenient constant according to the use case by scaling $T(\cdot)$ with the square root of the constant. ■

Using the statistics of Y_p, the VNT transform $T(\cdot)$ is derived according to Lemma 1 as

$$T(u)=\int\sqrt{\frac{2}{\sigma_\mathrm{N}^2}\frac{1}{(1+u^2)^2}}du=\sqrt{\frac{2}{\sigma_\mathrm{N}^2}}\mathrm{atan}(u). \tag{19}$$

Moreover, the distribution of the VNT transformed random variable can be approximately described with a Gaussian distribution. The Gaussian approximation is accurate for a variety of mean-variance dependent non-Gaussian distributions, such as Poisson [21], Gamma [22] and noncentral-chi distributions [8]. In this work, the Gaussian approximation for the VNT transformed random variable is employed for the parallel component to allow further investigation of the optimal constellation design. Thus, it is assumed that the conditional PDF of $R_\mathrm{p}=T(Y_\mathrm{p})$ is an unit variance Gaussian with the mean at $T(|X_\lambda|)$. The probability transformation rule is then employed to derive the marginal conditional distribution $P(Y_\mathrm{p}|X_\lambda)=P(T^{-1}(R_\mathrm{p})|X_\lambda)$ as

$$P(Y_\mathrm{p}|X_\lambda)=\frac{\cos^2\left(\mathrm{atan}(Y_\mathrm{p})\right)}{\sqrt{\pi\sigma_\mathrm{N}^2}}e^{-\frac{(T(Y_\mathrm{p})-T(|X_\lambda|))^2}{2}}. \tag{20}$$

Note that the functionality of the VNT transformation could introduce a peak amplitude constraint on the R_p as reported in [15]. Due to the atan in $T(\cdot)$, the R_p is constrained within $[-\sqrt{\frac{2}{\sigma_\mathrm{N}^2}}\frac{\pi}{2},\sqrt{\frac{2}{\sigma_\mathrm{N}^2}}\frac{\pi}{2}]$ to ensure the bijectivity of the transformation. Consequently, for the Gaussian approximation made for R_p to be effective, the constraint on R_p further limits the regime of $T(|X_\lambda|)$ to $[-\sqrt{\frac{2}{\sigma_\mathrm{N}^2}}\frac{\pi}{2}+4,\sqrt{\frac{2}{\sigma_\mathrm{N}^2}}\frac{\pi}{2}-4]$ to ensure almost 100% of the Gaussian PDF $P(R_\mathrm{p}|X_\lambda)$ is contained within the valid range.

In [20], it is pointed out that the distribution of the phase angle of the received symbol is dominantly dependent on the input symbol amplitude rather than on the phase. Combining this insight, the component statistics (12) and also some numerical simulation results, the Y_o is approximated with a zero mean symmetric distribution. In this work, a signal dependent Gaussian distribution as

$$P(Y_\mathrm{o}|X_\lambda)=\frac{1}{\sqrt{\pi\sigma_\mathrm{N}^2\left(1+|X_\lambda|^4\right)}}e^{-\frac{Y_\mathrm{o}^2}{\sigma_\mathrm{N}^2(1+|X_\lambda|^4)}} \tag{21}$$

is selected to approximate the distribution of the orthogonal component Y_o for its simplicity. Note that Y_o does not possess mean-dependent variance, thus, VNT cannot be derived using Lemma 1 for this component.

So far, the approximated marginal conditional PDFs of the parallel channel $P(Y_\mathrm{p}|X_\lambda)$ and the orthogonal channel $P(Y_\mathrm{o}|X_\lambda)$ have been proposed. Recall the assumed independence between the two channels when $|X_\lambda|$ is given, the conditional channel is then approximated with

$$\begin{aligned}P(Y_\mathrm{p},Y_\mathrm{o}|X_\lambda)=&\frac{\cos^2\left(\mathrm{atan}(Y_\mathrm{p})\right)}{\pi\sigma_\mathrm{N}^2\sqrt{1+|X_\lambda|^4}}\\&\times e^{\left(-\frac{(T(Y_\mathrm{p})-T(|X_\lambda|))^2}{2}-\frac{Y_\mathrm{o}^2}{\sigma_\mathrm{N}^2(1+|X_\lambda|^4)}\right)}.\end{aligned} \tag{22}$$

Note that the MI cannot be derived within the decomposed domain as the parallel and orthogonal directions will vary depending on different input symbol phase $\angle X_\lambda$. To calculate the MI, one will have to convert the distribution (22) back to the original real and imaginary domain. Since the transformation $[Y_\mathrm{p},Y_\mathrm{o}]\rightarrow[Y_\mathrm{r},Y_\mathrm{i}]$ has the Jacobian of unity, the conversion is given as

$$\begin{aligned}&P(Y_\mathrm{r},Y_\mathrm{i}|X_\lambda)=\\&\quad P\left(Y_\mathrm{p}=\frac{Y_\mathrm{r}X_\mathrm{r}+Y_\mathrm{i}X_\mathrm{i}}{|X_\lambda|},Y_\mathrm{o}=\frac{-Y_\mathrm{r}X_\mathrm{i}+Y_\mathrm{i}X_\mathrm{r}}{|X_\lambda|}|X_\lambda\right),\end{aligned} \tag{23}$$

where the relationships between $[Y_\mathrm{p},Y_\mathrm{o}]$ and $[Y_\mathrm{r},Y_\mathrm{i}]$ are substituted in equation (22).

B. Numerical Analysis of the Approximated Channel Model

With the approximated channel model being established, the accuracy of the approximated channel model should be investigated with numerical simulations of transmitting CS modulated signals. In order to achieve reasonable accuracy in terms of describing the CS of a time domain pulse, the sampling frequency of the signal should be sufficiently high. Furthermore, as pointed out in Section II-A, sufficient samples should also be taken in the CS domain such that the interference between nonlinear frequency subcarriers can be neglected. In this set of numerical experiments, the transmissions of an NFDM data frame are considered. The NFDM data frame is constructed as an analogy of an OFDM signal, $M=128$ subcarriers are considered within the nonlinear frequency window $[-4,4]$ with sinc pulse shaping as

$$\rho(\lambda,0)=\sum_{m=1}^{M}x_m\mathrm{sinc}\left(\frac{M}{\Lambda}\lambda+\frac{M-(2m-1)}{2}\right), \tag{24}$$

where x_m denotes the information symbol encoded on the m-th subcarrier and $\Lambda=8$ indicates the nonlinear frequency width of the NFDM symbol. The corresponding time domain pulses $q(t,0)$ at the transmitter should be obtained using INFT, then transmitted through a split step Fourier simulated fiber. In this work, we considered a long haul fiber of 2000 km with loss of $\alpha=0.02$ dB/km, group velocity dispersion factor $\beta_2=-2.2\times10^{-26}$ s^2/m, and Kerr nonlinear factor $\gamma=1.27\times10^{-3}$ /W/m. The time domain pulse width is selected to be 20 ns, including total guarding interval of 10 ns to avoid interference between neighbor pulses caused by dispersion induced pulse broadening. As for the normalizing factors, the normalization time is selected to be $T_0=0.1$ ns, the dispersion length and normalization power are $L_\mathrm{D}=455$ km and $1/\gamma L_\mathrm{D}=2.39$ dBm accordingly. Additionally, the wavelength employed for the signaling is 1.55 μm with the phonon occupancy factor $K_\mathrm{T}=1.13$.

In this work, the NFT/INFT algorithm provided in [23] is employed to efficiently travel between the CS and time domains. As highlighted previously, high resolution in both time and CS domains is essential for two reasons. Firstly, to reduce the computation error of NFT, the time domain signal should be sampled sufficiently higher than Nyquist sampling frequency. Secondly, the CS domain should have high resolution such that the discretization error can be neglected. In this work, 7.8 times Nyquist sampling is employed in the time domain sampling and 2 times of time domain samples are used to ensure sufficient sampling in both the time and CS domain.

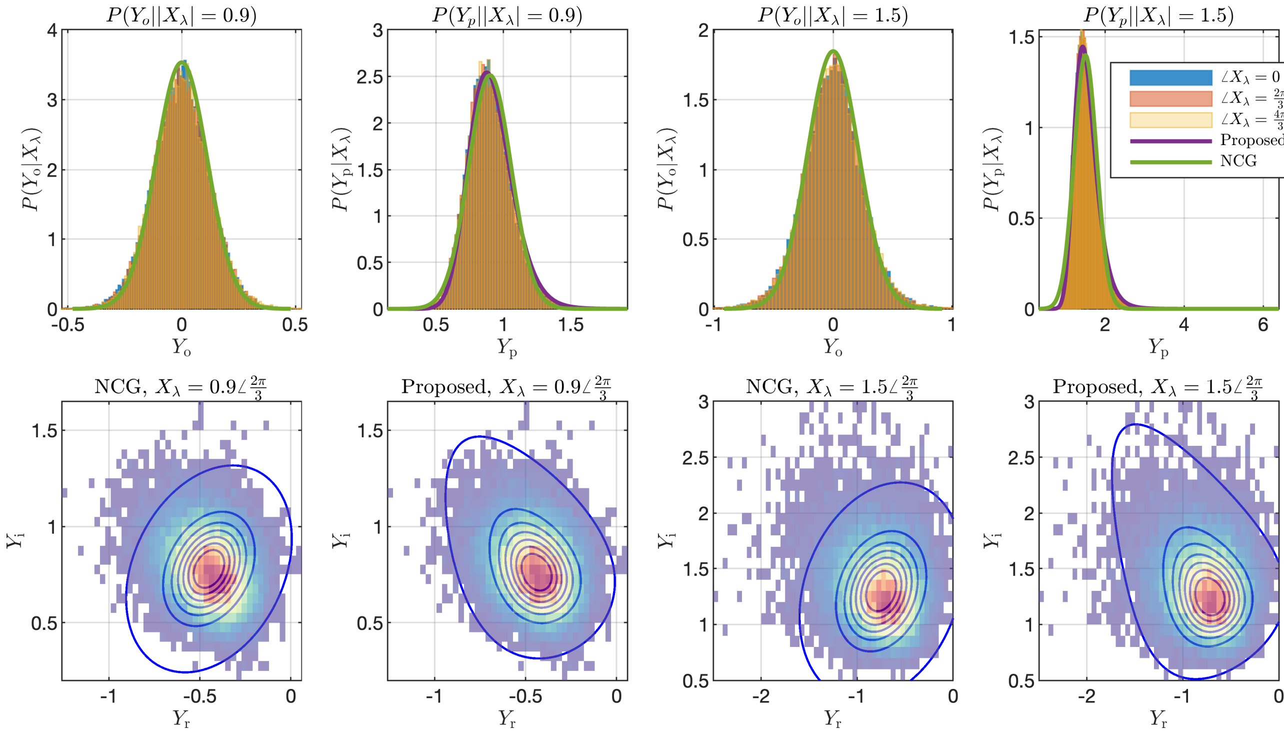


Fig. 2. The histograms (first row) of the decomposed parallel and orthogonal components Y_p and Y_o of the received symbol Y_λ given the transmitted symbol X_λ with different amplitudes being 0.9 and 1.5 and phases being 0, $2\pi/3$ and $4\pi/3$. The distributions (second row) of the received symbol given the transmitted symbol the transmitted symbol X_λ with different amplitudes being 0.9 and 1.5 and phases being $2\pi/3$. The corresponding non-circular Gaussian (NCG) model and the proposed model are also included as solid curves and contour lines.

To obtain the signal dependent channel statistics, 200 transmissions of the NFDM data frame are performed, which include 25600 random realizations for each symbol considered. Recall the peak amplitude constraint introduced by the functionality of the VNT (19), $|X_\lambda|$ for the given fiber parameters should be limited within $[0, 2.7]$. Hence, the symbol amplitudes considered here are 0.9 and 1.5 while phases selected are 0, $2\pi/3$ and $4\pi/3$. In the first row of Fig. 2, the histograms with the same amplitude $|X_\lambda|$ are presented in the same plot, the overlap between histograms given different phases $\angle X_\lambda$ are shown in both Y_p and Y_o components. This verifies the insight that the channel model is dominantly dependent on the amplitude of the transmitted symbol rather than the phase. The marginalized PDFs of the non-circular Gaussian (NCG) model with given statistics (3) and (4) and the proposed model are also plotted. Furthermore, using $\angle X_\lambda = 2\pi/3$ as an example, the NCG model and the proposed model (23) are sketched in two-dimensional contours over color-scaled histogram of the received symbol. A clear preference for the proposed model could be observed for describing the peak of the histogram. In addition to the advantage in capturing the peak, the proposed model contours also show good agreement with the simulated histogram, while the NCG model contours fail to capture its elliptical direction.

III. Capacity Analysis

In the previous section, the approximated channel model has been proposed based on the perturbation theory channel statistics and the VNT transformation, which will be used here to inspire the optimal constellation designs for signaling on CS. The model provides a full analytical description of the channel statistics which is matched up to the second moment to the statistics derived from perturbation theory. Therefore, the channel model can be used for shaping the input constellation to adapt to the signal dependence of the noise. However, due to the assumptions made in the approximation, the shaped input constellation is not necessarily capacity-achieving, but the optimized constellation would lead to a lower bound on the channel capacity. To further study the capacity of the CS channel, other than the Pinsker lower bound [14], [17], mismatch capacity [8], [21] with optimized input constellation will also be included in the next sections.

A. Input Constellation Shaping

Using the proposed model, the MI between the complex input X_λ and output Y_λ is written as

$$I(Y_\lambda; X_\lambda) = I(Y_\mathrm{r}, Y_\mathrm{i}; X_\lambda) = H(Y_\mathrm{r}, Y_\mathrm{i}) - H(Y_\mathrm{r}, Y_\mathrm{i}|X_\lambda), \tag{25}$$

where $H(Y_\mathrm{r}, Y_\mathrm{i})$ denotes the output entropy calculated by

$$H(Y_\mathrm{r}, Y_\mathrm{i}) = \int_{Y_\mathrm{r}}\int_{Y_\mathrm{i}} P(Y_\mathrm{r}, Y_\mathrm{i})\log_2\frac{1}{P(Y_\mathrm{r}, Y_\mathrm{i})}\mathrm{d}Y_\mathrm{r}\mathrm{d}Y_\mathrm{i}, \tag{26}$$

and $H(Y_\mathrm{r}, Y_\mathrm{i}|X_\lambda)$ denotes the conditional entropy given as

$$H(Y_\mathrm{r}, Y_\mathrm{i}|X_\lambda) = \sum_{x\in X_\lambda} P(X_\lambda = x)H(Y_\mathrm{r}, Y_\mathrm{i}|X_\lambda = x). \tag{27}$$

Note that discrete constellation is assumed to be employed in this work, considering a closer to practice constellation. Because of the discreteness assumed, besides the conditional entropy (27), the output distribution $P(Y_\mathrm{r}, Y_\mathrm{i})$ is also written with summation as

$$P(Y_\mathrm{r}, Y_\mathrm{i}) = \sum_{x \in X_\lambda} P(X_\lambda = x) P(Y_\mathrm{r}, Y_\mathrm{i} | X_\lambda = x). \tag{28}$$

Moreover, recall the unity Jacobian of the transformation and the assumed independence between the two components, when the input symbol is given the conditional entropy $H(Y_\mathrm{r}, Y_\mathrm{i} | X_\lambda = x)$ is then rewritten into

$$\begin{aligned} H(Y_\mathrm{r}, Y_\mathrm{i} | X_\lambda = x) &= H(Y_\mathrm{p}, Y_\mathrm{o} | X_\lambda = x) \\ &= H(Y_\mathrm{p} | X_\lambda = x) + \log_2 \sqrt{\pi e \sigma_\mathrm{N}^2 (1 + |x|^4)}, \end{aligned} \tag{29}$$

where the entropy $H(Y_\mathrm{p} | X_\lambda = x)$ is expressed as

$$\begin{aligned} &H(Y_\mathrm{p} | X_\lambda = x) \\ &= \int_{Y_\mathrm{p}} P(Y_\mathrm{p} | X_\lambda = x) \log_2 \frac{1}{P(Y_\mathrm{p} | X_\lambda = x)} \mathrm{d}Y_\mathrm{p}. \end{aligned} \tag{30}$$

Using the MI as the objective function, the constellation shaping problem is formulated as

$$\max_{K} \max_{[\mathbf{X}, \mathbf{P}(\mathbf{X})]} I(Y_\lambda; X_\lambda), \tag{31}$$

$$\text{s.t.} \sum \mathbf{P}(\mathbf{X}) = 1, \quad |X_1|, ..., |X_K| \leq X_{\max}, \tag{32}$$

where $\mathbf{X} = [X_1, X_2, ..., X_K]$ denotes the vector whose elements indicate the input alphabet, while the $\mathbf{P}(\mathbf{X}) = [P(X_1), P(X_2), ..., P(X_K)]$ indicates the probabilities correspond to the symbols in the alphabet. The later constraint in (32) is interpreted as a peak amplitude constraint, which is inherently imposed by VNT [15].

Solving the optimization problem defined in the equation (31) subject to the constraints (32) is equivalent to the optimal constellation design denoted as FO in this paper. In some scenarios, FO is not favored due to its complexity in implementation. Since the symbols of the constellation are originally equiprobable, a probability matcher is required to be implemented such that an information source with optimized probabilities can be established.

To avoid the additional probability matcher, GS is preferred in some scenarios by constraining probability further to $\mathbf{P}(\mathbf{X}) = \frac{1}{K}\mathbf{1}$, where the bold $\mathbf{1}$ denotes a K-length vector of ones. Furthermore, only K that is positive power of 2 is considered in GS to simplify the source coding further. GS optimizes the positions of the input symbols X_λ with the prior that they are equiprobable. Besides the simplified system structure, the GS optimization also consumes less time to converge as the dimension of the feasible region is smaller than that of the probabilistic geometric hybrid shaping in the FO scheme for the same K.

B. Pinsker Lower Bound

In the previous section, the MI of the approximated channel model is maximized by performing two constellation shaping optimizations, while the channel model is proposed under assumptions of sufficient sampling, independence between parallel and orthogonal components, and Gaussian PDF of the appropriate stage. If only the sufficient sampling approximation is made, a capacity lower bound can also be derived using the Pinsker formula and the channel statistics (5) and (6). The Pinsker formula provides a capacity lower bound for a given channel with known first and second order statistics while the exact model is unknown [14], [17], [24]. The lower bounding is provided in two aspects. On the one hand, the input distribution is assumed to be a zero-mean circular Gaussian with given constellation variance [24]. On the other hand, the channel is assumed to be a real vector Gaussian channel with the given statistics (hence only first and second order statistics are required) [24]. Using the channel statistics (7), (8) and (9), the Pinsker lower bound can be written considering the input being zero-mean circular Gaussian with variance of σ_S^2 as

$$C_\mathrm{pin} = \log_2 \left(1 + \frac{\sigma_\mathrm{S}^2}{\sigma_\mathrm{N}^2 (1 + \sigma_\mathrm{S}^2 + 2\sigma_\mathrm{S}^4)} \right). \tag{33}$$

The detailed derivation of (33) is provided in the Appendix B. Note that the Gaussian input assumed in this lower bound does not take into account the peak amplitude constraint discussed in this work. If identical constellation variance as the optimized constellation is employed, the Pinsker formula will provide a lower bound with less constraint, regarding the peak amplitude constraint as an additional constraint in the constellation shaping schemes.

C. Mismatch Capacity Lower Bound

So far, the capacity analysis of the CS channel has been performed under some approximations which allow the analytical result to be driven. In this section, the mismatch capacity will be discussed based on a set of realistic channel realizations, including transmissions of CS pulses in a split step Fourier simulated fiber with ASE noise added at each step. Furthermore, a practical and efficient NFT/INFT [23] algorithm is also employed to compute the CS domain pulse shaped, which includes the effect of computational error in practice. The mismatch capacity is given as

$$\begin{aligned} C_\mathrm{M} = &\sum_{k=1}^{K} \int_{Y_\mathrm{r}} \int_{Y_\mathrm{i}} P(X_k) P_\mathrm{T}(Y_\mathrm{r}, Y_\mathrm{i} | X_k) \\ &\times \log_2 \frac{P(Y_\mathrm{r}, Y_\mathrm{i} | X_k)}{\sum_{n=1}^{K} P(X_n) P(Y_\mathrm{r}, Y_\mathrm{i} | X_n)} \mathrm{d}Y_\mathrm{r} \mathrm{d}Y_\mathrm{i}, \\ = &\sum_{k=1}^{K} P(X_k) \mathsf{E}_\mathrm{T} \left[\log_2 \frac{P(Y_\mathrm{r}, Y_\mathrm{i} | X_k)}{\sum_{n=1}^{K} P(X_n) P(Y_\mathrm{r}, Y_\mathrm{i} | X_n)} \right], \end{aligned} \tag{34}$$

where the $P_\mathrm{T}(Y_\mathrm{r}, Y_\mathrm{i} | X_k)$ denotes the true channel model given symbol X_k is transmitted, and the $\mathsf{E}_\mathrm{T}(\cdot)$ denotes the expectation over the true channel, which is approximated with the Monte Carlo method by numerically averaging over large number of realizations. Mismatch capacity quantifies the amount of information that could be reliably transmitted in the true physical channel using the decoding rules optimally designed for the auxiliary channel without specifying the exact decoder [8], [21], [25]. This is also known as the achievable information rate (AIR), where the achievability relies on the assumed auxiliary channel based on which the optimal decoding rule is already known. The input distributions employed here are obtained from the constellation shaping discussed in the previous section, while the approximated channel $P(Y_\mathrm{r}, Y_\mathrm{i}|X_k)$ in (23) and (22) is selected to be the auxiliary mismatch channel. Such an estimation not only provides a proven capacity lower bound, but also gives an estimate of the impact of the approximations made in deriving the channel model (23) and (22), considering the equivalence between the C_M and the MI achieved by the same input $P(X)$ when $P(Y_\mathrm{r}, Y_\mathrm{i}|X_k) = P_\mathrm{T}(Y_\mathrm{r}, Y_\mathrm{i}|X_k)$.

D. Numerical Results and Discussion

In this work, both FO and GS schemes are implemented to estimate their corresponding optimal MI under the peak amplitude constraint in the CS domain. By performing optimizations for both schemes, the shaping gain will be estimated to evaluate the necessity to perform the additional probabilistic shaping and constellation size optimization. Note that the optimal shaping in the FO scheme and the GS shaping corresponding to the inner optimization in (31) are solved with an interior-point algorithm with random initialization to enhance the algorithm convergence. The optimization on the constellation size K in FO scheme is done by identifying the non-increasing trend of the optimized MI for different K's. The channel parameter σ_N^2 is selected to be 0.0154, which corresponds to the same long-haul fiber employed in Section II-B.

Implementing the optimization with the parameters described above, the optimized MI under the peak amplitude constraint $X_{\max} \in [0.3, 2.4]$ is shown in Fig. 3a. The MI for unshaped 16, 32, 64 APSK are calculated to represent the performance of unshaped constellation as benchmarks[1]. The APSKs are chosen over Quadrature Amplitude Modulations (QAM) which are more common in optical fiber communication as the optimally shaped constellations also converge to an APSK multi-ring constellation. The range of interest considered here roughly corresponds to the time domain power regime of -30 to -10 dBm. Recall that the conversion between the symbol in CS domain and the time domain power is not straightforward, the power regime mentioned above is estimated under the assumptions of sinc pulse shaping and 128 subcarriers as in (24).

Overall, the MIs for both GS and FO shaping techniques increase at higher $X_{\max}$, which corresponds to a higher signal power. In addition, both schemes outperform the best unshaped APSK constellation. The gain over unshaped APSK schemes at lower values of $X_{\max}$ is not as significant. For example, at $X_{\max} = 0.6$, an MI of 3.59 bits per channel use can be achieved with unshaped 32-APSK, while FO can only provide an improvement of 0.07 bits per channel use. Conversely, more significant gains are observed at higher values of $X_{\max}$. For instance, at $X_{\max} = 2.4$, unshaped 64-APSK achieves an MI of 4.51 bits per channel use, and FO can provide a gain of 0.41 bits per channel use. Moreover, it is also worth pointing out that the unshaped APSK will result in a decreasing MI for the high $X_{\max}$ because of the unshaped APSK not being able to adaptively change the spacing between the constellation rings to account for the increased nonlinear effects.

When comparing between the two shaping techniques, the additional probabilistic shaping and constellation size optimization in FO result in less than 0.1% MI gain over their GS counterparts for all the $X_{\max}$. Moreover, it is worth noticing that the rate of increasing will start to reduce, showing a trend towards saturation. This could be the consequence of the noise power increasing faster than the signal power with the increasing signal amplitude.

Comparing the optimized distributions of both schemes shown in Fig. 4, it is observed that under the peak amplitude constraint, both schemes converge to multi-ring constellations. The FO scheme will allocate the probabilities of the mass points on the ring adaptively, for example, in the Fig. 4b, the mass points on the outer ring are assigned with higher probabilities than those on the central rings. The GS shaping, on the other hand, will also arrange the spacing between the constellation rings with respect to the signal dependence similar to FO scheme. In both Fig. 4c and Fig. 4f, it should be noticed that the spacing between the rings is increased along with the signal amplitude to adapt to the signal dependency of the noise.

Employing the shaped input constellations discussed above, the mismatch capacity is estimated as a proven lower bound for the true capacity with $1000 \times M = 128000$ realizations of the channel. Additionally, the Pinsker lower bound [14] is also calculated as a lower bound subjected to the identical variance of the input as the FO shaped constellation. Recall this variance constraint is a weaker constraint compared to the peak amplitude constraint when identical variance is considered. The mismatch capacity and Pinsker lower bounds are sketched in Fig. 3b. As expected from the previous discussion of the additional shaping gain from performing the extra constellation size optimization and probabilistic shaping, almost identical mismatch lower bounds are achieved by the FO shaped constellation and the GS shaped 64 APSK respectively. It is worth pointing out that the MI gap of 0.02 is observed at $X_{\max} = 0.5$ for the FO while the gap of 0.15 is produced at $X_{\max} = 2.1$. This implies the gap between the optimized MI, which is calculated from the approximated channel model (23), and the mismatch capacity will become larger at higher $X_{\max}$.

The trend of the gap matches with the discussion in Section II-A. The perturbative channel statistics (3) and (4) [14] correspond to the continuous signal. However, the signal will be discretized before further digital signal processing such as numerical NFT [23] in the practical system and also in the simulations considered within this work. The discretized channel (5) and (6) are obtained based on the infinitely high sampling rate approximation as discussed in Appendix A. To capture the deviation between the continuous perturbative model and the discretized model with 400 GHz sampling rate, i.e. 7.8 times Nyquist rate, 25600 random realizations of

[1]The APSKs employed here correspond to 4+12APSK, 4+12+16APSK, and 8+16+20+20APSK respectively. The radius and phase angle of the points are selected according to the highest coding spectra efficiency in [26].

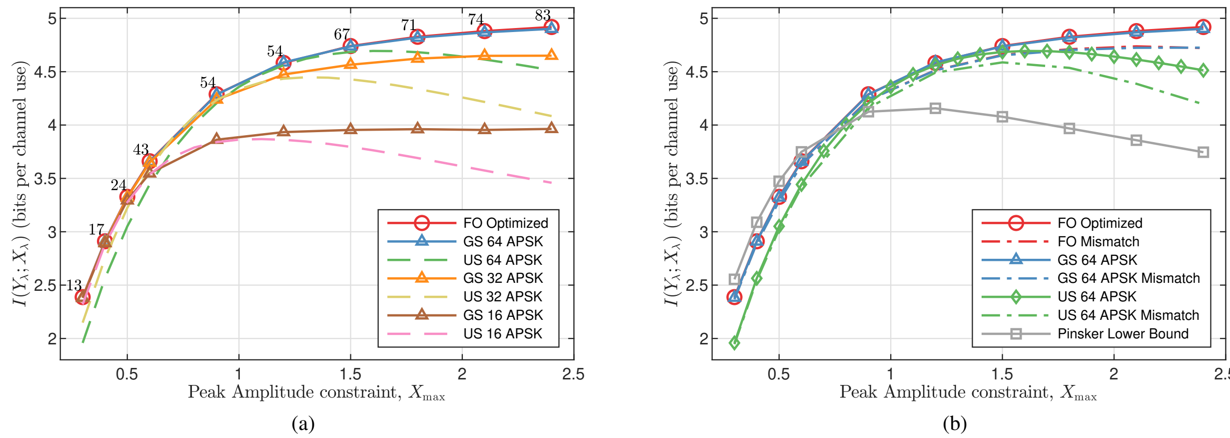


Fig. 3. (a) MI achieved by the FO optimized (optimal sizes specified), GS optimized constellations and unshaped (US) 16, 32, 64 APSK. (b) Mismatch capacity lower bound employing FO optimized, GS optimized, US 64 APSK, and also Pinsker Capacity lower bound when the input variance is equal to the input variance of FO optimized constellation.

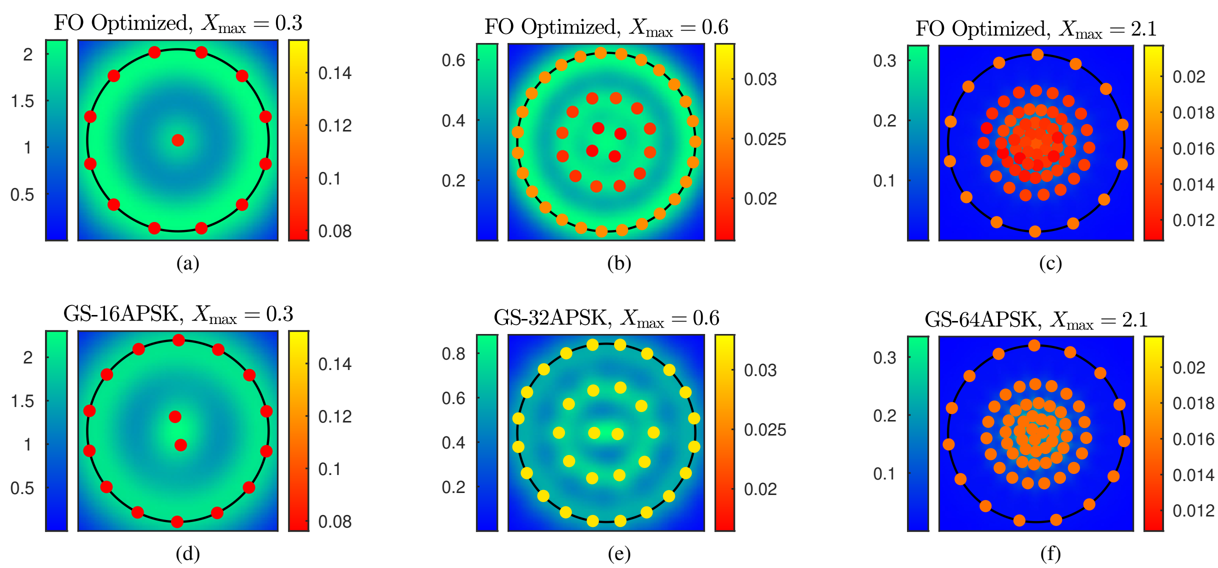


Fig. 4. FO Optimized input distribution $P(X_{\mathrm{r}}, X_{\mathrm{i}})$ (right color bar) and corresponding output distribution (left color bar) $P(Y_{\mathrm{r}}, Y_{\mathrm{i}})$ with the peak amplitude constraint $X_{\max}$ = (a) 0.3, (b) 0.6, (c) 2.1; GS Optimized APSK input and output distribution with the peak amplitude constraint $X_{\max}$ = (d) 0.3, 16 APSK (e) 0.6, 32 APSK, (f) 2.1, 64 APSK; The solid lines denote the peak amplitude constraint $X_{\max}$.

transmitting randomly selected symbols from a constellation with continuous uniform distributed phase angles and the same symbol amplitudes from 0 to 2.4. The variance of the received symbol and its decomposed components are shown in Fig. 5. When the transmitted symbol amplitude is lower than 1.2, the well match between the numerical and analytical estimated statistics implies the discretized channel model provides an accurate estimation for the statistics of the noise in this regime. Hence, the gap between the optimized MI and mismatch capacity should also be small. However, in the higher signal regime, it is observed that the gap between the numerical and analytical statistics is increasing. As explained in Appendix A, the error is signal dependent and it is introduced by discretizing a signal dependent continuous random process. Referred as processing noise, such signal dependent deviation was also reported in the literature due to its increasing dominance in higher power regime [16].

Despite the deviation from the analytical MI, the shaping constellation can provide gain over the unshaped constellation as expected, considering the unshaped 64 APSK only achieves a maximal of 4.59 bits per channel use at $X_{\max} = 1.5$ while the maximal of 4.72 can be achieved by the GS 64 APSK at $X_{\max} = 2.1$. When comparing the Pinsker lower bound, a better lower bound is generated by the Pinsker lower bound when the $X_{\max}$ is small due to the additional peak amplitude constraint on the shaping scheme. However, the shaped constellations achieve higher lower bounds even with the extra constraint at higher $X_{\max}$, this reveals the potential of performing constellation shaping even using an approximated signal dependent channel model.

IV. Signaling With Matched Filter

If sinc pulse shaping is used in frequency domain, it is commonly known that the corresponding inverse FT (IFT) is

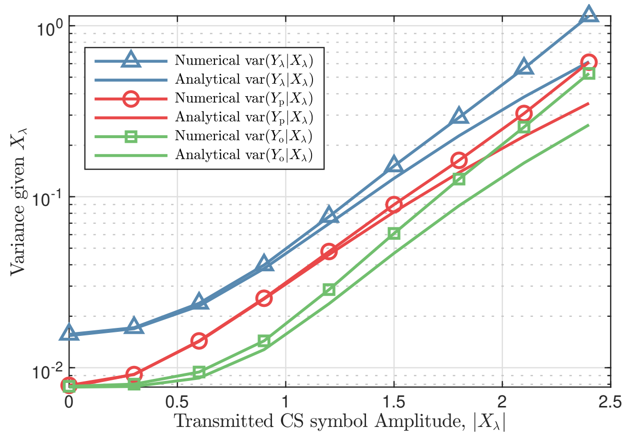


Fig. 5. The variance of the received symbol Y_λ, and its parallel components Y_p, and orthogonal components Y_o given the amplitude of the transmitted symbols $|X_\lambda|$ with uniformly distributed phase $\angle X_\lambda$.

a rectangle function which is band-limited in time domain. Exploiting this property of the sinc pulse shaping (24), the signal dependent noise in the CS modulated system could be potentially reduced with a matched filter implemented in the CS domain [20]. However, the signal dependence of the CS noise will induce ISI at the matched filter output. Convolving the received symbol Y_λ with the matched filter, the input-output relationship is then given as

$$Y_{\mathrm{f}\lambda} = h_\mathrm{f}(\lambda) * Y_\lambda = X_\lambda + h_\mathrm{f}(\lambda) * N_\lambda = X_\lambda + N_{\mathrm{f}\lambda}, \quad (35)$$

where $h_\mathrm{f}(\lambda)$ denotes the impulse response of the matched filter in CS domain given as

$$h_\mathrm{f}(\lambda) = \frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}\lambda\right). \quad (36)$$

The convolution is equivalent to multiplying a rectangular brick wall filter with the IFT domain of the received CS pulses. Due to the band-limitedness of the sinc pulse in its IFT domain, the signal X_λ will remain unchanged after matched filtering. Using the input-output relationship (35), autocorrelation (3) and (4), it is clear that the filtered noise will remain zero mean, while the second order statistics are given as

$$\mathsf{E}[N_{\mathrm{f}\lambda}N^*_{\mathrm{f}\lambda'}] = \int \sigma^2 h_\mathrm{f}(\lambda-\xi)h_\mathrm{f}(\lambda'-\xi) \times \left(1+|X_\xi|^2+|X_\xi|^4\right)\mathrm{d}\xi, \quad (37)$$

$$E\left[N_{\mathrm{f}\lambda}N_{\mathrm{f}\lambda'}\right] = \int \sigma^2 h_\mathrm{f}(\lambda-\xi)h_\mathrm{f}(\lambda'-\xi)X_\xi^2\mathrm{d}\xi, \quad (38)$$

where ξ is the auxiliary variable introduced due to the convolution and $\frac{M}{\Lambda}$ corresponds to the separation between the subcarriers. It can be observed from (37) and (38) that although the subcarriers are allocated on the zero crossing points of the $\mathrm{sinc}(\cdot)$ functions of their neighbors, the signal dependence will still introduce ISI. In the previous sections, when filtering is not employed, the discretized channel statistics (5) and (6) are obtained assuming sampling in the CS domain is sufficiently high, the ISI can be neglected as explained in Appendix A. However, for the filtering to be effective, the passband of the filter should be comparable to the bandwidth of the signal and this will make the no ISI approximation not accurate. More detailed discussion and comparison with a signal independent noise case is also discussed in Appendix A.

In order to implement constellation shaping, the closed-form expression of the channel law is required for the formulation of the MI, which would be the objective function of the shaping optimization. When ISI is considered, such a closed-form expression would be intractable, hence, making it infeasible to optimize the constellation. To simplify the implementation of the constellation shaping, only the signal dependence part of the channel noise will be considered despite the existence of the ISI. The noise at each subcarrier is considered as a white signal dependent noise that only depends on the signal on the same subcarrier. Using similar approaches as Section II-A, the CS noise statistics after filtering are then derived as

$$\mathsf{E}\left[|Y_{\mathrm{f}\lambda}-X_\lambda|^2\right] = \mathsf{E}\left[|N_{\mathrm{f}\lambda}|^2\right] \approx \sigma^2_{\mathrm{fil}}(1+|X_\lambda|^2+|X_\lambda|^4), \quad (39)$$

$$\mathsf{E}\left[(Y_{\mathrm{f}\lambda}-X_\lambda)^2\right] = \mathsf{E}\left[N^2_{\mathrm{f}\lambda}\right] \approx \sigma^2_{\mathrm{fil}}X^2_\lambda, \quad (40)$$

where the $\sigma^2_{\mathrm{fil}} = \sigma^2 l\pi M/\Lambda$ corresponds to the received noise power when transmitted symbol $X_\lambda = 0$, the $\pi M/\Lambda$ denotes the unitless passband width of the filter (36) when the small signal asymptotic IFT is performed [4]. Employing the same methodology as the unfiltered channel, an identical approximated channel model with attenuated noise power can be obtained by substituting the approximated filtered received noise power σ^2_{fil}, the filtered received symbol Y_f, and its corresponding decomposed components into equations (23) and (22) as

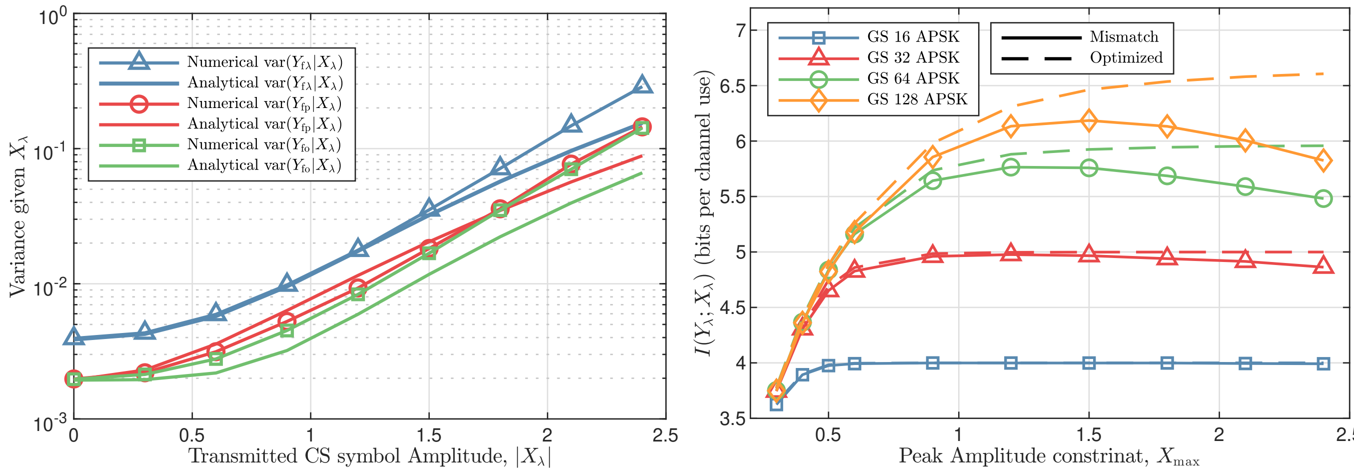
$$P(Y_{\mathrm{fr}}, Y_{\mathrm{fi}}|X_\lambda) = P(Y_r = Y_{\mathrm{fr}}, Y_i = Y_{\mathrm{fi}}|X_\lambda, \sigma^2_\mathrm{N} = \sigma^2_{\mathrm{fil}}),$$

where

$$P(Y_{\mathrm{fp}}, Y_{\mathrm{fo}}|X_\lambda) = P(Y_\mathrm{p} = Y_{\mathrm{fp}}, Y_\mathrm{o} = Y_{\mathrm{fo}}|X_\lambda, \sigma^2_\mathrm{N} = \sigma^2_{\mathrm{fil}}).$$

Similar to the previous section, numerical simulations are performed to estimate the channel statistics to evaluate the influence of the neglected ISI. Since the conditional channel statistics are not dependent on the phase of the input, they are estimated with the same approach as in the previous section. This approach allows signal dependence of the noise and the ISI to be decoupled to some extent, and the numerical estimation of the filtered noise statistics is shown in Fig. 6a. When the signal is small, the error of the NFT can be neglected, it is observed that the variance of the filtered noise is well described by the model (39), while the decomposed components show otherwise. The decomposed components show a slightly larger deviation from the model when compared to the unfiltered noise shown in Fig. 5, which implies the effect of ISI when the signal is small. As the signal becomes larger, the error introduced by the discretized NFT becomes more significant, the model becomes more inaccurate due to the combined effect of the NFT error and the ISI.

Recall in Section III-D, it is pointed out that the additional constellation size optimization and probabilistic shaping do not provide significant gain for the signal dependent noise



(a) (b)

Fig. 6. (a) The variance of the matched filtered received symbol $Y_{\mathrm{f}\lambda}$, and its parallel components Y_{fp}, and orthogonal components Y_{fo} given the amplitude of the transmitted symbols $|X_\lambda|$ with uniformly distributed phase $\angle X_\lambda$. (b) Mismatch capacity employing GS 16, 32, 64 and 128 APSKs optimized constellations and the optimized MI corresponding to the no ISI approximated filtered channel.

considered in this work. Furthermore, the filtered channel (35) with no ISI approximation has the same functionality as the unfiltered channel (2) considered previously. Thus, for more practical implementation of the optimizer, only the GS-APSK will be considered for the filtered system. The optimization problem is hence formulated as

$$\max_{\mathbf{X}} I(Y_{\mathrm{f}\lambda}; X_\lambda), \tag{41}$$

subject to the same constraint (32), where probabilities $\mathbf{P_X} = \frac{1}{K}\mathbf{1}$ and $K = 16, 32, 64, 128$ are considered. Note that $K = 128$ is taken into account based on the fact that the noise power is reduced by the matched filter, suggesting that more constellation points could be supported. Using the identical optimizer and system parameters as in the previous section, the optimal discrete constellation for the channel $P(Y_{\mathrm{fr}}, Y_{\mathrm{fi}}|X_\lambda)$ can be obtained. It is worth emphasizing that the ISI is not considered in the channel model, and the numerical simulation shows that the ISI is non-negligible. For the optimized constellation to produce an effective capacity lower bound of the system, mismatch capacity is necessary. Considering the channel model $P(Y_{\mathrm{fr}}, Y_{\mathrm{fi}}|X_\lambda)$ as an auxiliary channel, the mismatch capacity for the filtered system is given as

$$C_{\mathrm{Mf}} = \sum_{k=1}^{K} P(X_k)\mathsf{E}_{\mathrm{Tf}}\left[\log_2 \frac{P(Y_{\mathrm{fr}}, Y_{\mathrm{fi}}|X_k)}{\sum_{n=1}^{K} P(X_n)P(Y_{\mathrm{fr}}, Y_{\mathrm{fi}}|X_n)}\right], \tag{42}$$

where the $\mathsf{E}_{\mathrm{Tf}}(\cdot)$ denotes the expectation over the true filtered channel, which is approximated with the Monte Carlo method by numerically averaging over $1000M$ realizations of the filtered channel, where the ISI will be included. In Fig. 6b, the mismatch capacity of the filtered system and the MI for the no ISI approximated filtered channel model for GS 16, 32, 64 and 128 - APSKs are shown. The optimized MIs will keep increasing along with the relaxation of the peak amplitude constraint till saturation. The saturation is due to the capacity of the approximated channel is higher than the source entropy, as seen in the figure that 4 and 5 bits per channel use are provided by 16 and 32-APSK at $X_{\max} = 2.4$.

The capacity lower bound of the true system that taken in account the ISI, however, shows a different trend. At low $X_{\max}$, relaxing the constraint will improve the mismatch capacity as the ISI is small when the signal is of low amplitude. At higher $X_{\max}$, the mismatch capacity will decrease because of the increasing ISI along with the increased signal. For instance, the maximum capacity lower bound of 6.19 bits per channel use is achieved by GS 128 APSK at $X_{\max} = 1.5$, while only 5.82 can be achieved at $X_{\max} = 2.4$. Considering the pulse width being 20ns, the maximum data rate of 2.41 Mbits/s/subcarrier is achieved, which is significantly higher than the maximum data rate of 1.85 Mbits/s/subcarrier achieved in the direct signaling system.

When comparing the optimized MI and the mismatch capacity, the alignment at low $X_{\max}$ are all reasonably good, which highlights the effectiveness of the approximation when ISI is weak. When ISI becomes more dominant with the increasing $X_{\max}$, the gap between the optimized MI and the mismatch capacity becomes larger because the no ISI approximation is no longer accurate as well as the signal dependent discretization error becomes larger. The additional error power introduced by the ISI will limit the mismatch capacity achieved. For example, 6.4 bits per channel use MI are predicted by no ISI optimization for GS 128 APSK at $X_{\max} = 1.5$, while only 6.19 can be achieved by mismatch capacity.

Furthermore, it is also worth noticing that the size of the constellation affects the alignment of the optimized MI and mismatch capacity. For a smaller constellation size like 16-APSK, the distances between the constellation points are sufficiently large to sustain the MI under the combined effect of ISI and signal dependent noise. Hence the corresponding mismatch capacity and optimized MI will both saturate at 4 bits per channel use. On the contrary, increasing the constellation size will reduce the distance between the constellation points, the influence of the ISI will be more significant. As an example, the gaps between optimized MI and mismatch capacity are 0.14, 0.48 and 0.79 for GS 32, 64 and 128 APSKs, respectively.

V. Conclusion

In this work, both direct signaling and matched filter signaling of the CS modulated NFDM system are discussed. We first propose an approximated channel model for the direct signaling complex CS modulated NFDM system, based on which, the mutual information gain from shaping the distribution of the input is discussed. Promising MI gains are observed for both FO and GS constellations over the unshaped ring-based APSK constellations. However, the gain from additional probabilistic shaping is not as significant when the number of constellation points in each ring is optimally determined. Conventionally, the probabilistic shaped QAM constellation is preferred to avoid redesigning the analogue digital converter and optical signal processing algorithm, in addition to the effective shaping gain provided by Maxwell-Boltzmann shaping [27]. The GS technique has attracted certain attention from the development of end-to-end learning and autoencoders, which allow the constellation to be shaped geometrically according to the learned features of the channel [28]. However, the autoencoder is considered to be complex due to the large amount of training data required and the complicated structure in the implementation. Comparing to the learned GS, the GS constellation demonstrated in this work is more practical to implement, as the shaped constellation can be practically obtained using practical numerical optimizers based on the proposed channel model. Moreover, the capacity lower bound is also estimated with the mismatch capacity and the Pinsker Formula. Neglecting the ISI, the approximated channel model for the complex CS modulated NFDM system with matched filtering is employed to geometrically shape the constellation adapting only to the signal dependent noise. The estimated mismatch capacity lower bound outperforms that of the unfiltered system although the ISI is neglected in the constellation shaping.

In future works, one could further investigate the detailed channel model of the ISI in the system with matched filter. If such a model can be developed and input constellation could be shaped accordingly, a tighter capacity lower bound could be expected.

Appendix A
Inter Subcarrier Symbol Interference Approximation Analysis

In this work, the inter subcarrier interference is neglected when the discretized CS channel statistics is proposed assuming that the sampling in the CS domain is sufficiently high. In this section, the validity of this assumption shall be explained. For simplicity, consider a general continuous random process with zero mean and autocorrelation as

$$\mathsf{E}\left[N_t N_{t'}^*\right] = \sigma^2\delta(t-t')(1+|X_t|^2+|X_t|^4), \qquad (43)$$

$$\mathsf{E}\left[N_t N_{t'}\right] = \sigma^2\delta(t-t')X_t^2, \qquad (44)$$

where the σ^2 is a factor that corresponds to power spectrum density of this infinite bandwidth random process that is dependent on a series of time variant signal X_t. If τ is introduced as an auxiliary variable in convolution, the random process $N_{\mathrm{h}t}$ after filtering with an arbitrary filter $h(t)$ would possess autocorrelation as

$$\begin{aligned}\mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'}^* | h(t)\right] &= \int \sigma^2 h(t-\tau)h^*(t'-\tau) \\ &\quad\times\left(1+|X_\tau|^2+|X_\tau|^4\right)\mathrm{d}\tau, \qquad (45)\\ \mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'} | h(t)\right] &= \int \sigma^2 h(t-\tau)h(t'-\tau)\left(X_\tau^2\right)\mathrm{d}\tau. \qquad (46)\end{aligned}$$

In practice, the random process will be sampled. Sampling of an unlimited bandwidth signal would correspond to multiplying a brick wall filter $\tilde{h_{\mathrm{s}}}(f) = \mathrm{rect}(f/f_{\mathrm{s}})$ whose bandwidth is equal to the sampling frequency f_{s} in the frequency domain, which will then corresponds to convolving the corresponding time domain impulse response $h_{\mathrm{s}}(t) = f_{\mathrm{s}}\mathrm{sinc}(f_{\mathrm{s}}t)$. When f_{s} is sufficiently large, the approximation $\lim_{f_{\mathrm{s}}\to\infty} h_{\mathrm{s}}(t) = \delta(t)$ can be employed to show that $\mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'}^* | h_{\mathrm{s}}(t)\right] = \mathsf{E}\left[N_t N_{t'}^*\right]$, and $\mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'} | h_{\mathrm{s}}(t)\right] = \mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'}\right]$. The equivalence implies that the ISI can be neglected if sampled at sufficiently high sampling rate.

However, as briefly mentioned in the main body of this work, such an approximation cannot be made for the filtered signal dependent noise. As an analogy to the CS system discussed in this work, the input signal X_t is defined as

$$X_t = \sum_{m=1}^{M} x_m \mathrm{sinc}\left(\frac{M}{\Lambda}t + \frac{M-(2m-1)}{2}\right), \qquad (47)$$

where $\frac{\Lambda}{M}$ denotes the separation between the time domain subcarriers, and x_m denotes the information symbol on the m-th subcarrier at $t = \frac{(2m-1)-M}{2}\frac{\Lambda}{M}$. Matching to the pulse shaping function, the impulse response of the matched filter is given as $h_{\mathrm{f}}(t) = \frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}t\right)$.

For a conventional system where the noise is modeled with i.i.d AWGN noise η with zero mean and σ^2 power spectrum density, it is trivial to find out that the power of the $h_{\mathrm{f}}(t)$ filtered noise η_{f} is given as $\frac{M}{\Lambda}\sigma^2$. On the contrary, if the noise is signal dependent with the correlation (43) and (44), the noise is no longer i.i.d. The correlation of the filtered noise is given as $\mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'}^* | h_{\mathrm{f}}(t)\right]$ and $\mathsf{E}\left[N_{\mathrm{h}t} N_{\mathrm{h}t'} | h_{\mathrm{f}}(t)\right]$. Additionally, the passband of the filter $h_{\mathrm{f}}(t)$ is equal to the bandwidth of the signal X_t, making the $\delta(t)$ approximation employed previously not applicable. Hence, the noise statistics of each subcarrier will become dependent on not only the signal on the subcarrier, but also on the signals on the neighbor subcarriers. As an example, the autocorrelation of the filtered noise is written as

$$
\begin{aligned}
&\mathsf{E}[N_{\mathrm{ht}}N^{*}_{\mathrm{ht}'}|h_{\mathrm{f}}(t)]\\
&\quad=\int \sigma^2\frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}(t-\tau)\right)\frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}(t'-\tau)\right)\\
&\qquad\quad\times\left(1+|X_\tau|^2+|X_\tau|^4\right)\mathrm{d}\tau,\\
&\quad=\frac{M}{\Lambda}\sigma^2\mathrm{sinc}\left(\frac{M}{\Lambda}(t-t')\right)\\
&\qquad+\int \sigma^2\frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}(t-\tau)\right)\frac{M}{\Lambda}\mathrm{sinc}\left(\frac{M}{\Lambda}(t'-\tau)\right)\\
&\qquad\quad\times\left(|X_\tau|^2+|X_\tau|^4\right)\mathrm{d}\tau,
\end{aligned}\tag{48}
$$

by substituting signal (47) into (45). Now considering $t=\frac{(2m-1)-M}{2}\frac{\Lambda}{M}$, $m=1,2,...,M$ and $t'=\frac{(2m'-1)-M}{2}\frac{\Lambda}{M}$, $m'=1,2,...,M$ at subcarriers, in the absence of signal dependency (i.e., the terms $|X_\tau|^2$ and $|X_\tau|^4$ are dropped), the integral term in the right hand side (RHS) of (48) vanishes. In addition, the sinc term in the RHS of (48) equals $\frac{M}{\Lambda}\sigma^2$ at $m=m'$ and equals zero at $m\neq m'$, suggesting no ISI in the absence of signal dependency. On the contrary, the integral term in the RHS of the (48) shows nonzero autocorrelation in both $m=m'$ and $m\neq m'$ cases. Hence, it is clear that ISI will be introduced in the filtered system by the signal dependence of the noise.

Appendix B
Pinsker Formula Lower Bound Derivation

In order to derive a capacity lower bound using the Pinsker formula, one will first have to rewrite the complex system with a vector real systems as [14], [24]

$$\mathbf{Y}=\mathbf{X}+\mathbf{N},\tag{49}$$

where the conditional statistics of the noise vector are described by (7), (8) and (9). Without knowing the exact channel model (i.e. the conditional PDF), the capacity of such a system could be lower bounded by the MI assuming input is a complex circular Gaussian distribution and the channel is a Gaussian distribution with specified correlations. In this work, we refer to this lower bound as the Pinsker lower bound C_{pin}, and it is defined by substituting the Gaussian distributions into the random vector MI equation as

$$C_{\mathrm{pin}}=\frac{1}{2}\log_2\frac{\det\mathbf{\Sigma_X}\det\mathbf{\Sigma_Y}}{\det\mathbf{\Sigma_U}}\tag{50}$$

where the vector $\mathbf{U}=[X_{\mathrm{r}},X_{\mathrm{i}},Y_{\mathrm{r}},Y_{\mathrm{i}}]^{\mathrm{T}}$ is an auxiliary random vector introduced for the convenience in the notation, $\mathbf{\Sigma_X}$, $\mathbf{\Sigma_Y}$ and $\mathbf{\Sigma_U}$ are the input, output, and input-output joint covariance matrices respectively. The complex input is considered to be a complex circular Gaussian with zero mean and σ_{S}^2 variance, the covariance matrix of the rewritten input vector $\mathbf{X}$ is given as $\mathbf{\Sigma_X}=\frac{\sigma_{\mathrm{S}}^2}{2}\mathbf{I_2}$. The zero mean signal dependent noise is characterized by the conditional statistics (7), (8) and (9), hence the covariance matrix of the noise vector $\mathbf{N}$ is rearranged as (51). Using the covariance matrix of the input $\mathbf{\Sigma_X}$ and the noise $\mathbf{\Sigma_{N|X}}$, the covariance matrix of the output $\mathbf{Y}$ can be acquired by marginalizing the conditional output correlation matrix as

$$\mathbf{\Sigma_Y}=\left[\frac{\sigma_{\mathrm{N}}^2}{2}(1+\sigma_{\mathrm{S}}^2+2\sigma_{\mathrm{S}}^4)+\frac{\sigma_{\mathrm{S}}^2}{2}\right]\mathbf{I_2}.\tag{52}$$

Recall the circular complex Gaussian assumed for the input, the real and imaginary parts of the input are independent to each other. The independence leads to zero marginal correlation of the real and imaginary channels, i.e. $\mathsf{E}[Y_{\mathrm{r}}Y_{\mathrm{i}}]=0$. Furthermore, this property also simplifies the computation of the marginalization of the second moment of each output channel, resulting in equal second moment, $\mathsf{E}[Y_{\mathrm{r}}Y_{\mathrm{r}}]=\mathsf{E}[Y_{\mathrm{i}}Y_{\mathrm{i}}]$.

The final component required for the Pinsker lower bound (50) is the input-output joint covariance matrix $\mathbf{\Sigma_U}$. The joint covariance matrix consists of three sub-matrices as

$$\mathbf{\Sigma_U}=\begin{bmatrix}\mathbf{\Sigma_X} & \mathbf{\Sigma_{XY}}\\ \mathbf{\Sigma_{XY}}^{\mathrm{T}} & \mathbf{\Sigma_Y}\end{bmatrix}.\tag{53}$$

where $\mathbf{\Sigma_{XY}}$ denotes the input-output covariance matrix. Due to the assumed circular symmetric input, it is easy to show that the input-output correlation matrix is a diagonal matrix as

$$\mathbf{\Sigma_{XY}}=\frac{\sigma_{\mathrm{S}}^2}{2}\mathbf{I_2}\tag{54}$$

Finally, substitute the covariance matrices derived above into (50) using the same techniques as in [14], the Pinsker lower bound C_{pin} is obtained as

$$C_{\mathrm{pin}}=\log_2\left(1+\frac{\sigma_{\mathrm{S}}^2}{\sigma_{\mathrm{N}}^2(1+\sigma_{\mathrm{S}}^2+2\sigma_{\mathrm{S}}^4)}\right).\tag{55}$$

References

[1] Y. Chen, M. Baniasadi, and M. Safari, "Optimal shaping gains for continuous spectrum nonlinear frequency signalling over long fibers," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 1927–1932.

[2] 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, 2010.

[3] J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities," *Optics express*, vol. 25, no. 3, pp. 1916–1936, 2017.

[4] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, part i: Mathematical tools," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4312–4328, 2014.

[5] S. A. Derevyanko, S. Turitsyn, and D. Yakushev, "Non-gaussian statistics of an optical soliton in the presence of amplified spontaneous emission," *Optics letters*, vol. 28, no. 21, pp. 2097–2099, 2003.

[6] 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 Transactions on Communications*, vol. 66, no. 7, pp. 2978–2993, 2018.

[7] A. Buchberger, A. G. i Amat, V. Aref, and L. Schmalen, "Probabilistic eigenvalue shaping for nonlinear fourier transform transmission," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4799–4807, 2018.

[8] Y. Chen, I. Tavakkolnia, A. Alvarado, and M. Safari, "On the capacity of amplitude modulated soliton communication over long haul fibers," *Entropy*, vol. 22, no. 8, p. 899, 2020.

[9] J. Gordon, "Interaction forces among solitons in optical fibers," *Optics letters*, vol. 8, no. 11, pp. 596–598, 1983.

[10] S. Derevyanko, M. Balogun, O. Aluf, D. Shepelsky, and J. E. Prilepsky, "Channel model and the achievable information rates of the optical nonlinear frequency division-multiplexed systems employing continuous b-modulation," *Optics Express*, vol. 29, no. 5, pp. 6384–6406, 2021.

[11] M. Baniasadi, Y. Chen, and M. Safari, "On the capacity of b-modulated nonlinear frequency division multiplexing," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 1933–1938.

$$\mathbf{\Sigma}_{\mathbf{N}|\mathbf{X}} = \frac{\sigma_{\mathrm{N}}^2}{2}\begin{bmatrix} 1+2X_{\mathrm{r}}^2+2X_{\mathrm{r}}^2X_{\mathrm{i}}^2+X_{\mathrm{r}}^4+X_{\mathrm{i}}^4 & 2X_{\mathrm{r}}X_{\mathrm{i}} \\ 2X_{\mathrm{i}}X_{\mathrm{r}} & 1+2X_{\mathrm{i}}^2+2X_{\mathrm{r}}^2X_{\mathrm{i}}^2+X_{\mathrm{r}}^4+X_{\mathrm{i}}^4 \end{bmatrix} \tag{51}$$

[12] T. Gui, G. Zhou, C. Lu, A. P. T. Lau, and S. Wahls, "Nonlinear frequency division multiplexing with b-modulation: shifting the energy barrier," *Optics express*, vol. 26, no. 21, pp. 27 978–27 990, 2018.

[13] X. Yangzhang, V. Aref, S. T. Le, H. Buelow, D. Lavery, and P. Bayvel, "Dual-polarization non-linear frequency-division multiplexed transmission with b-modulation," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1570–1578, 2019.

[14] S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, "Capacity estimates for optical transmission based on the nonlinear fourier transform," *Nature communications*, vol. 7, no. 1, pp. 1–9, 2016.

[15] I. Tavakkolnia and M. Safari, "Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers," *Journal of Lightwave Technology*, vol. 35, no. 11, pp. 2086–2097, 2017.

[16] M. Pankratova, A. Vasylchenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, "Signal-noise interaction in optical-fiber communication systems employing nonlinear frequency-division multiplexing," *Physical Review Applied*, vol. 13, no. 5, p. 054021, 2020.

[17] K. Turitsyn, S. Derevyanko, I. Yurkevich, and S. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Physical review letters*, vol. 91, no. 20, p. 203901, 2003.

[18] M. Safari, "Efficient optical wireless communication in the presence of signal-dependent noise," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015, pp. 1387–1391.

[19] J. G. Smith, "The information capacity of amplitude-and variance-constrained sclar gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[20] I. Tavakkolnia and M. Safari, "Signaling on the continuous spectrum of nonlinear optical fiber," *Optics express*, vol. 25, no. 16, pp. 18 685–18 702, 2017.

[21] A. Tsiatmas, F. M. J. Willems, and C. P. M. J. Baggen, "Square root approximation to the poisson channel," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 1695–1699.

[22] P. R. Prucnal and B. E. Saleh, "Transformation of image-signal-dependent noise into image-signal-independent noise," *Optics letters*, vol. 6, no. 7, pp. 316–318, 1981.

[23] S. Wahls, S. Chimmalgi, and P. Prins, "FNFT: A Software Library for Computing Nonlinear Fourier Transforms," *The Journal of Open Source Software*, vol. 3, p. 597, 2018. [Online]. Available: https://doi.org/10.21105/joss.00597

[24] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, 2001.

[25] A. Ganti, A. Lapidoth, and I. E. 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, 2000.

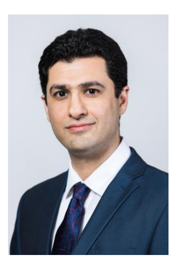
[26] E. ETSI, "302 307-2 v1. 3.1 (2021-07) digital video broadcasting (dvb)," *Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications*, 2021.

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

[28] S. Gaiarin, F. Da Ros, R. T. Jones, and D. Zibar, "End-to-end optimization of coherent optical communications over the split-step fourier method guided by the nonlinear fourier transform theory," *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 418–428, 2020.

Yu Chen Yu Chen is a PhD candidate at the University of Edinburgh, specializing in nonlinear fibre optics communication, free-space optics communication, and information theory. He received a BEng and an MSc from the same institution. His research has been published in journals such as Entropy, where his paper was selected as a featured article, and TCOM, as well as at conferences including IWCMC and ISIT. His work at IWCMC was recognized with the Best Project Prize during his MSc studies.

Mohammadamin Baniasadi Mohammadamin (Amin) Baniasadi (Member, IEEE) received his B.S. and M.S. degrees in Electrical Engineering from the University of Tehran, Iran, in 2012 and 2015, and his Ph.D. degree in Electrical and Computer Engineering from the University of California, Riverside in 2020. Since November 2020, he is a research associate in Nonlinear Communication Theory at The University of Edinburgh, UK. His research interests include Information Theory, Communication Theory, Optical Communications, and Game Theory.



Majid Safari Dr. Majid Safari received PhD degree in Electrical and Computer Engineering from the University of Waterloo, Canada. He is currently a Reader (associate professor) at the Institute for Digital Communication, University of Edinburgh, UK. His research interests include the application of information theory and advanced signal processing in optical wireless, optical fibre, and quantum communications. Dr Safari has published more than 100 papers and was named an "Emerging Leader" in Optics by the editorial board of the Journal of Optics in 2018. Some of his current research works include designing 6G Optical wireless networks as part of the EPSRC program grant TOWS, developing single-photon avalanche diode based receivers for classical communication, and the design of novel communication schemes for nonlinear long-haul fibre-optic channels. Dr. Safari is an associate editor of IEEE Transactions on Communication and was the TPC co-chair of the 4th International Workshop on Optical Wireless Communication in 2015.