On the Capacity of Amplitude Modulated Soliton Communication over Long Haul Fibers

Yu Chen [1],* , **Iman Tavakkolnia** [1] , **Alex Alvarado** [2]   and **Majid Safari** [1],*

[1]   Institute for Digital Communication, School of Engineering, University of Edinburgh, Edinburgh EH9 3FD, UK; I.Tavakkolnia@ed.ac.uk

[2]   Information and Communication Theory Lab, Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; a.alvarado@tue.nl

*   Correspondence: yu.chen@ed.ac.uk (Y.C.); majid.safari@ed.ac.uk (M.S.)

Abstract: The capacity limits of fiber-optic communication systems in the nonlinear regime are not yet well understood. In this paper, we study the capacity of amplitude modulated first-order soliton transmission, defined as the maximum of the so-called time-scaled mutual information. Such definition allows us to directly incorporate the dependence of soliton pulse width to its amplitude into capacity formulation. The commonly used memoryless channel model based on noncentral chi-squared distribution is initially considered. Applying a variance normalizing transform, this channel is approximated by a unit-variance additive white Gaussian noise (AWGN) model. Based on a numerical capacity analysis of the approximated AWGN channel, a general form of capacity-approaching input distributions is determined. These optimal distributions are discrete comprising a mass point at zero (off symbol) and a finite number of mass points almost uniformly distributed away from zero. Using this general form of input distributions, a novel closed-form approximation of the capacity is determined showing a good match to numerical results. Finally, mismatch capacity bounds are developed based on split-step simulations of the nonlinear Schrödinger equation considering both single soliton and soliton sequence transmissions. This relaxes the initial assumption of memoryless channel to show the impact of both inter-soliton interaction and Gordon–Haus effects. Our results show that the inter-soliton interaction effect becomes increasingly significant at higher soliton amplitudes and would be the dominant impairment compared to the timing jitter induced by the Gordon–Haus effect.

Keywords: achievable information rate; channel capacity; nonlinear fiber optics; soliton communication; solitonic interaction; variance normalizing transform

1. Introduction

It is predicted that the capacity of data transfer network, mainly consists of optical fibers, will fall behind the data traffic demands in the near future [1]. The prediction implies the need for exploiting current optical fiber infrastructure to their limits before migrating to the next generation of optical fiber systems. However, the fundamental information transmission capacity of the most basic optical fiber link (i.e., standard single-mode fiber) is not fully known in the nonlinear regime. Different approaches have been used to tackle this problem in the literature including the recent application of nonlinear Fourier transform (NFT) to approach the limits of the nonlinear optical fiber [2,3]. Using NFT, the nonlinear dispersive fiber channel, defined by the nonlinear Schrödinger equation (NLSE), is transformed to linear channels in nonlinear spectral domain, redefining the capacity problem formulation for nonlinear optical fibers.

By applying NFT, the available degrees of freedom in temporal domain are transformed to two types of spectra in the nonlinear spectral domain, namely the discrete and continuous spectra. Therefore, NFT is regarded as a base for development of new techniques of data transmission, and different communication system designs have been proposed using NFT [4–13]. The performance of such NFT-employed system for long-haul communication is investigated by simulation and experiment [14,15]. However, it has been observed that the noise behavior is not trivial in these systems [16–18], and the performance largely depends on the design. Moreover, the application of NFT in estimating the capacity of nonlinear optical fibers is not straightforward since the NFT and inverse NFT (INFT) must be performed numerically and are computationally complex [19,20].

An estimation of the capacity of the nonlinear optical fiber by only signaling on its continuous spectrum defined by NFT is provided in [21,22]. Achievable rates have been predicted, but it has been shown that due to the signal dependency of the noise, the capacity will be saturated at high power. Moreover, several works in the literature have been focused on estimating the achievable information rates (AIR) of the fiber when the discrete spectrum (i.e., soliton transmission) is used as the signal space. In [23], a capacity lower bound for amplitude modulated first-order soliton communication system is estimated using a half-Gaussian input distribution. In [24], an achievable rate is estimated taken into account the Gordon–Haus effect that leads to timing jitter at the receiver. In [18], AIR is estimated for a more complicated system that modulates both the eigenvalue and the norming constant in the discrete spectrum. Assuming a receiver capable of detecting variable pulse duration, in [25], the time-scaled mutual information (MI) is numerically optimized considering the memoryless channel model for soliton communication.

In this paper, we investigate the capacity of the optical fiber channel when only a single discrete spectrum point is encoded and the data is mapped on the imaginary part of the corresponding eigenvalue. This is essentially equivalent to the amplitude modulated soliton communication in [26]. As mentioned above, a number of capacity bounds for such channel has been derived previously [18,23,24], and AIR in bits per second were also discussed in [25]. However, some intrinsic limitations, such as dependence of bandwidth on soliton amplitude and the interaction between neighboring soliton pulses have been ignored. Compared to the state-of-art works in the literature (e.g., [23,25]), we investigate the effect of channel memory induced by solitonic interaction, which is mostly ignored in the literature. In order to incorporate the time-bandwith degrees of freedom into the capacity problem formulation, we study the maximization of time-scaled MI similar to [25] but by assuming a more practical communication system that uses a fixed symbol duration (i.e., soliton pulse width). A general form of capacity-approaching input distributions are proposed through the optimization of an approximated normalized channel model, providing important insights into the optimal design of soliton communication systems. In addition, an analytical estimation of the capacity of amplitude modulated soliton transmission is provided.

This paper is structured as follows: In Section 2, we initially consider a commonly used memoryless non-Gaussian channel model for the imaginary part of the eigenvalue [16]. By applying the variance normalizing transform (VNT) [22,27], the original channel is transformed into an equivalent channel with normalized noise power, which is then approximated by a unit-variance additive white Gaussian noise (AWGN) model in Section 3. Taking into account a peak amplitude constraint imposed by bandwidth limitations, the capacity in bits/normalized time and its corresponding input distribution are estimated using the proposed AWGN model and also an approximate analytical approach. Next, in the Section 4, we consider the effect of channel memory by developing the mismatch capacity bounds based on the split-step simulation of single soliton and soliton sequence transmissions over the NLSE. Based on the mismatch capacity results, the impact of inter-soliton interaction and Gordon–Haus effects on the capacity of soliton communication systems is studied.

2. Channel Model

At a low launch power, the optical fiber channel can be modeled as a linear dispersive channel impaired by AWGN noise. However, the Kerr nonlinearity becomes significant when the signal power increases to allow transmission over long haul fibers. The propagation of the complex envelope of a narrowband optical field in a standard single-mode fiber can be described by the stochastic nonlinear Schrödinger equation (NLSE), as discussed in ([28], Chapter 4). Assuming the fiber loss to be perfectly compensated by an ideal distributed Raman amplification, the NLSE is given as

$$\frac{\partial Q(T,Z)}{\partial Z} = -j\frac{\beta_2}{2}\frac{\partial^2 Q(T,Z)}{\partial T^2} + j\gamma|Q(T,Z)|^2 Q(T,Z) + N(T,Z), \tag{1}$$

where $Q(T,Z)$ denotes the complex envelope of the optical field, $N(T,Z)$ represents the amplifier spontaneous emission (ASE) noise term, T and Z are time and propagation distance, and β_2 and γ indicate group velocity dispersion and Kerr nonlinearity respectively. Note that the fiber loss term α here is omitted since ideal distributed Raman amplification is assumed. The ASE noise is modeled by a zero mean white Gaussian noise with autocorrelation $\mathsf{E}[N(T,Z)N^*(T',Z')] = N_{\text{ASE}}\delta(T-T')\delta(Z-Z')$. The spectral density of the noise in $[\text{W}/(\text{km}\cdot\text{Hz})]$ is $N_{\text{ASE}} = \alpha h\nu_0 K_{\text{T}}$ for the ideal distributed Raman amplification assumed in this work, where $h\nu_0$ denotes the photon energy and K_{T} denotes the phonon occupancy factor. The NLSE could be normalized into the form

$$j\frac{\partial q(t,z)}{\partial z} = \frac{\partial^2 q(t,z)}{\partial t^2} + 2|q(t,z)|^2 q(t,z) + n(t,z), \tag{2}$$

with the corresponding normalized parameters as

$$q = \sqrt{\gamma L_D}Q, \quad z = Z/2L_D, \quad t = T/T_0, \tag{3}$$

where dispersion length is defined as $L_D = T_0^2/|\beta_2|$, and normalizing time T_0 can be selected independent of other parameters. Consequently, the autocorrelation of the normalized noise is,

$$\mathsf{E}[n(t,z)n^*(t',z')] = \sigma^2\delta(t-t')\delta(z-z'), \tag{4}$$

where $\sigma^2 = N_{\text{ASE}}\frac{2\gamma L_D^2}{T_0}$ according to the normalization (3).

Using the inverse scattering method, NFT transforms the time domain optical signal into scattering data, consisting of continuous spectrum $\rho(\lambda,z)$, eigenvalues $\lambda_m(z)_{m=1}^M$ and corresponding norming constants $C_m(z)_{m=1}^M$ which evolve linearly along the fiber in nonlinear spectral domain. It can be shown that, in a noise-free and interaction-free scenario, the eigenvalues λ_m are preserved during the evolution along the fiber [29]. If only one eigenvalue exists at $z = 0$ and $\rho(\lambda,0) = 0$, the solution of NLSE is a first-order soliton, which can be described analytically as

$$q(t,z) = 2\eta e^{-2i\zeta t + 4i(\zeta^2-\eta^2)z - i(\psi+\pi/2)}\text{sech}(2\eta t - 8\eta\zeta z - 2\epsilon), \tag{5}$$

where the only eigenvalue is $\lambda_1 = \zeta + i\eta$ ($\eta > 0$). Also, $e^{2\epsilon} = \frac{C_1}{2\eta}$ and $\psi = \arg C_1(z)$ where C_1 denotes the norming constant corresponding to eigenvalue λ_1.

The Energy of the soliton in (5) is equal to 4η, where the temporal width and bandwidth are proportional to $1/\eta$ and η respectively. Note that within this work, only the imaginary part of the eigenvalue is modulated and the real part is set to zero, i.e., $\eta = A$, $\zeta = 0$. Thus, at $z = 0$, the input pulse can be expressed as

$$q(t,z=0) = 2A\text{sech}(2At). \tag{6}$$

The propagation of the soliton pulse over the fiber is described by NLSE, and at the receiver side, the eigenvalue can be detected by NFT or pulse energy estimation. If the detected eigenvalue is

denoted as R, the channel model for this amplitude modulated first-order soliton transmission system can be described by a conditional PDF $P_{R|A}(r|a)$, which is non-Gaussian with a variance dependent on its mean [16,30]. Ignoring inter-soliton interactions, a memoryless channel model can be defined for the amplitude modulated soliton system based on a noncentral chi-squared distribution (NCX2) with 4 degrees of freedom as [16,23]

$$P_{R|A}(r|a) = \frac{2}{\sigma_N^2}\sqrt{\frac{r}{a}}\exp\left(-\frac{2a+2r}{\sigma_N^2}\right)I_1\left(\frac{4\sqrt{ar}}{\sigma_N^2}\right), \tag{7}$$

where $I_1(\cdot)$ denotes the modified first order Bessel function of the first kind. The mean and variance of this distribution for large a are $\mu_{\text{NCX2}}(a) = \sigma_N^2 + a$ and $\sigma_{\text{NCX2}}^2(a) = \frac{1}{2}\sigma_N^4 + a\sigma_N^2$ respectively, where $\sigma_N^2 = \frac{1}{2}\sigma^2\frac{L}{2L_D}$ at distance $Z = L$ and σ^2 is the power spectral density of the normalized ASE noise as defined in Equation (4). It can be seen that the channel model (7) for the imaginary part of the eigenvalue (soliton amplitude, or soliton energy) is non-Gaussian with signal dependent variance. In the next section, we develop different approaches to estimate the capacity of the channel described by (7).

3. Capacity Formulation for Memoryless Soliton Communication Channel

Here, the capacity problem for the channel defined by the conditional PDF (7) is formulated considering a peak amplitude constraint since the bandwidth occupied by soliton pulses is directly related to their amplitudes. That is, the modulating data on higher amplitudes requires larger bandwidth while the maximum signal bandwidth is restricted by physical limitations. Moreover, in practical scenarios, peak power is also constrained due to device limitations. Another important issue that needs to be considered for soliton communications systems is that soliton pulses defined as in (6) are not time-limited, and thus, they should be truncated for practical implementations.

We define the practical width of a soliton pulse (denoted by t_s) as the temporal width that contains $1 - \delta$ of the soliton energy. Recalling the energy of the normalized soliton (6) is equal to $4A$, this practical width can be obtained by solving the equation below for t_s

$$\int_{-t_s/2}^{+t_s/2}\left|2A\text{sech}(2At)\right|^2dt = (1-\delta)4A, \tag{8}$$

which is given by

$$t_s(A,\delta) = \frac{1}{2A}\ln\left(\frac{2}{\delta}-1\right), \tag{9}$$

where the fixed value δ should be sufficiently small to make the truncation error negligible compared to noise. For example, assuming that the soliton pulse width is defined based on containing 99.9% of its energy ($\delta = 0.001$), we have $t_s = 3.8/A$. Noting that the temporal width of soliton pulses is inversely related to their amplitudes, we can also introduce a minimum amplitude constraint to limit the utilization of the temporal resources. Based on the constraints mentioned above, the capacity problem can be formulated as

$$C_{\text{bpcu}} = \sup_{P_A(a):A\in\{0\}\cup[A_{\text{lb}},A_{\text{ub}}]}I(A;R), \tag{10}$$

where C_{bpcu} denotes the capacity in bits per symbol per channel use, $I(A;R)$ represents the MI. Denoting the transmitted and received eigenvalues with random variables A and R respectively, A_{ub} is the maximum amplitude constraint determined by maximum bandwidth or peak power and A_{lb} is the minimum amplitude constraint determined by the maximum allowed symbol duration. Note that we also consider the possibility of transmitting no soliton over a symbol duration (i.e., off symbol) with probability p_0, which is denoted by $A = 0$ here.

Noting that the signal space and the temporal resources are inter-related in the underlying soliton communication system, we will use an alternative capacity formulation that maximizes time-scaled MI [25] to get better insights into AIRs of the system in bits per second. Unlike [25], we assume a fixed symbol duration for all transmitted solitons to facilitate practical implementation. Since the pulse width is inversely related to the amplitude of the soliton, the minimum nonzero soliton amplitude $A_{\min} \geq A_{\text{lb}}$ (i.e., maximum pulse width) in a given input distribution determines the symbol duration. Note that $A_{\min}$ is not necessarily equal to the minimum amplitude constraint A_{lb} and $P(A < A_{\min}) = p_0$. The time-scaled MI (MI) is thus defined as

$$\mathbb{R}(A; R) = \frac{I(A; R)}{t_{\text{s}}(A_{\min}, \delta)}, \tag{11}$$

where MI is divided by the normalized symbol duration, resulting in a unit of [bits/normalized time]. The data rate in [bits/second] can be estimated by dividing the time-scale MI (11) with the normalizing time T_0 in (3). The corresponding time-scaled capacity formulation is then given by

$$C = \sup_{P_A(a): A \in \{0\} \cup [A_{\text{lb}}, A_{\text{ub}}]} \mathbb{R}(A; R). \tag{12}$$

Note that the minimum amplitude constraint A_{lb} can be also relaxed, since it is already inherently imposed by the modified objective function, i.e., the time-scaled MI. This is because the optimal solution would not include the small soliton amplitudes that consume the available temporal resources inefficiently due to their very large pulse width. Hence the capacity problem can be also written as

$$C = \sup_{P_A(a): A \in [0, A_{\text{ub}}]} \mathbb{R}(A; R). \tag{13}$$

In Section 3.2, it is shown that a minimum nonzero soliton amplitude $A_{\min}$ naturally appears in the optimal distribution of the capacity problem in (13).

3.1. Equivalent Channel Model Based on VNT

To simplify the capacity analysis, similar to the method used in [22,30–32], variance normalizing transform (VNT) is applied here to transform the original signal-dependent noise channel to a channel with a fixed noise power at sufficiently large signal-to-noise ratios. In general, the VNT can be applied to any random variable R where its variance σ_R^2 is related to the mean μ_R as $\sigma_R^2 = f^2(\mu_R)$. Then the variance of the transformed random variable, $Y = T(R)$, is normalized to one (i.e., mean independent) at sufficiently large values of μ_R. The general form of VNT can be written based on [33] as

$$T(u) = \int \frac{1}{f(u)} du. \tag{14}$$

Therefore the normalized random variable $Y = T(R)$ has the moments $\sigma_Y^2 \approx 1$ and $\mu_Y = \mathsf{E}[y] \approx T(\mu_R)$ for sufficiently large value of μ_R. Substituting the statistics of the NCX2 channel $\mu_{\text{NCX2}}(a) = \sigma_N^2 + a$ and $\sigma_{\text{NCX2}}^2(a) = \frac{1}{2}\sigma_N^4 + a\sigma_N^2 = \sigma_N^2(\frac{1}{2}\sigma_N^2 + a) = \sigma_N^2(\mu_{\text{NCX2}}(a) - \frac{1}{2}\sigma_N^2)$ considered in this work, the VNT will be given as

$$T(u) = \int \frac{1}{\sqrt{\sigma_N^2(u - \frac{1}{2}\sigma_N^2)}} du = 2\sqrt{\frac{u}{\sigma_N^2} - \frac{1}{2}} \approx 2\sqrt{\frac{u}{\sigma_N^2}}, \tag{15}$$

where the approximation is made for mathematical simplicity and due to the fact that the variance normalization itself defined by VNT is only precise at large values of u/σ_N^2 where the adopted approximation is also precise [22,27,31,32].

As shown in Figure 1, an equivalent soliton communication system can be defined based on the VNT approach where the noise power is signal-independent at large signal levels. Note that, in order to perform the coding and decoding at the same signal space, it is convenient to include both VNT and inverse VNT (IVNT) meaning that the soliton amplitude, A, is determined from the original input data $X = T(A)$ as

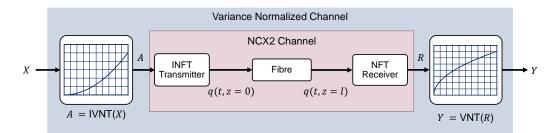$$A = T^{-1}(X) = \sigma_N^2 \frac{X^2}{4}. \tag{16}$$

Figure 1. Block diagram of an amplitude modulated soliton communication system with inverse variance normalizing transform (IVNT) and VNT, A and R denote the transmitted and received soliton amplitude, X and Y denote the transformed input and output signals, and q denotes the time domain signal.

Noting the square root form of the VNT defined in (15) and considering that the NCX2 model in (7) defines the channel between the soliton eigenvalues A and R in Figure 1, the equivalent channel model between the transformed random variables X and Y is described by a noncentral chi (NCX) conditional PDF as

$$P_{Y|X}(y|x) = \frac{y^2}{x} \exp\left(-\frac{y^2 + x^2}{2}\right) I_1(xy), \tag{17}$$

where $X = T(A) = 2\sqrt{A}/\sigma_N$ and $Y = T(R) = 2\sqrt{R}/\sigma_N$.

The capacity in bit per symbol of the system in (10) can then be rewritten based on the random variables X and Y as

$$C_{\text{bpcu}} = \sup_{P_X(x):X\in\{0\}\cup[X_{\text{lb}},X_{\text{ub}}]} I(X;Y), \tag{18}$$

where $X_{\text{lb}} = T(A_{\text{lb}})$ and $X_{\text{ub}} = T(A_{\text{ub}})$. Moreover, the corresponding time-scaled capacity formulation is given by

$$C = \sup_{P_X(x):X\in\{0\}\cup[X_{\text{lb}},X_{\text{ub}}]} \mathbb{R}(X;Y), \tag{19}$$

or based on the relaxed constraint as

$$C = \sup_{P_X(x):X\in[0,X_{\text{ub}}]} \mathbb{R}(X;Y), \tag{20}$$

where the time-scaled MI can be written as

$$\mathbb{R}(X;Y) = \frac{I(X;Y)}{t_s(A_{\min},\delta)} = \frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta - 1)} I(X;Y), \tag{21}$$

and $X_{\min}$ denotes the minimum nonzero symbol amplitude, i.e., $A_{\min} = T^{-1}(X_{\min}) = \sigma_N^2 X_{\min}^2/4$. It is important to notice that the VNT transformation does not affect the MI between input and output, i.e., $I(A;R) = I(X;Y)$, since the VNT function (15) is a monotonic and invertible function within the interested domain (See Lemma in [22]). Hence, the capacity formulations in (12) and (19) are equivalent.

3.2. Approximate AWGN Channel Model

It has been shown that the probability distribution of the normalized random variable after VNT tends to Gaussian distribution for a family of originally non-Gaussian probability distributions [22,31]. In this section, we first show that this is also true for the NCX distribution (17) in a Kullback–Leibler (KL) divergence sense. This inspires us to propose an approximate AWGN channel model to describe the amplitude modulated soliton communication system after VNT transformation as

$$Y = X + \Gamma, \tag{22}$$

where the additive noise Γ is Gaussian with zero mean and unit variance.

Proposition 1. *The KL divergence between the NCX distribution, $P_{Y|X}(y|x)$, given in (17) and a Gaussian distribution $Q_{Y|X}(y|x)$ with mean x and unit variance tends to zero for a sufficient large x, that is*

$$\lim_{x \to +\infty} D_{\mathrm{KL}}(P, Q|x) = 0, \tag{23}$$

where KL divergence, $D_{\mathrm{KL}}(P, Q|x)$, is defined as

$$D_{\mathrm{KL}}(P, Q|x) = \int_{-\infty}^{+\infty} P_{Y|X}(y|x) \ln \frac{P_{Y|X}(y|x)}{Q_{Y|X}(y|x)} dy, \tag{24}$$

Proof of Proposition 1. The detailed proof of Proposition 1 is shown in Appendix A □

Proposition 1 indicates that the NCX channel model (17) behaves similar to the approximate AWGN channel for a sufficiently large x. For example, The KL divergence D_{KL} is estimated as small as 1.77×10^{-12} for $x = 86.67$. This is by assuming that the pulse width contain 99.9% of the energy ($\delta = 0.001$) and some typical fiber parameters as in Table 1. Next, we will show that the proposed approximate AWGN channel converges to the original NCX channel at sufficiently large large X_{lb}.

Theorem 1. *Given the input $X \in \{0 \cup [X_{\mathrm{lb}}, X_{\mathrm{ub}}]\}$ at a sufficiently large X_{lb}, the mismatch capacity of the NCX channel with the approximate AWGN channel defined by (22) as auxiliary channel converges to the actual capacity of the NCX channel.*

Proof of Theorem 1. The detailed proof of Theorem 1 is shown in Appendix B. □

In [34,35], it is shown for the AWGN channel with amplitude constraints that the capacity-achieving distribution is discrete with a finite number of mass points for such channels. An upper bound is proposed in [36] for the number of mass points. However, these works focus on the MI-based capacity formulation. In the next Proposition, we extend the result in [34] to show the discreteness of the optimal solution to the time-scaled MI maximization problem for the proposed approximate AWGN channel.

Proposition 2. *Given an AWGN channel with the input amplitude constraint of $X \in \{0 \cup [X_{\mathrm{lb}}, X_{\mathrm{ub}}]\}$ and $X_{\mathrm{lb}} \to \infty$, the optimal input distribution for the capacity formulation in (19) is discrete with a finite number of mass points.*

Proof of Proposition 2. The detailed proof of Proposition 2 is shown in Appendix C. □

Now, approximating the channel in (19) with an AWGN model based on Theorem 1 and considering the conclusion of Proposition 2 on the discreteness of the optimal input distribution asymptotically, the MI between X and Y can be expressed as

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(\Gamma)$$

$$= \sum_{k=0}^{M} \int p_X(x_k) Q_{Y|X}(y|x_k) \log_2 \frac{1}{\sum_{j=0}^{M} p_X(x_j) Q_{Y|X}(y|x_j)} dy - \log_2 \sqrt{2\pi e}, \tag{25}$$

where $h(Y)$ denotes the output differential entropy, $h(\Gamma)$ denotes the differential entropy of the unit variance AWGN noise, x_k and $p_X(x_k)$ denote the input symbols and their corresponding probabilities within the input source alphabet, M denote the size of the nonzero alphabet, $x_0 = 0$ and $p_X(x_0) = p_0$ denotes the corresponding probability. Hence, the problem in (19) can be rewritten as

$$C = \max_{M} \left[\max_{[\mathbf{x}, \mathbf{p_X}]: x_k \in \{0\} \cup [X_{lb}, X_{ub}]} \mathbb{R}(X;Y) \right], \tag{26}$$

where the time-scaled MI function $\mathbb{R}(X;Y)$ is a function of two $(M+1)$-length vectors $\mathbf{x}$ and $\mathbf{p_X}$ which denote the mass points and their probabilities. As mentioned in the previous sections, the minimum amplitude constraint can be also relaxed yielding

$$C = \max_{M} \left[\max_{[\mathbf{x}, \mathbf{p_X}]: x_k \in [0, X_{ub}]} \mathbb{R}(X;Y) \right]. \tag{27}$$

Since the input distribution is discrete, the vector $[\mathbf{x}, \mathbf{p_X}]$ is sufficient to describe the input random variable X. The discreteness of the capacity-achieving input distribution allows for numerical evaluation of the capacity expression using similar algorithms as in [30,34]. In this work, the optimization over $[\mathbf{x}, \mathbf{p_X}]$ is performed using an interior-point optimizer in MATLAB given the number of nonzero mass point is fixed at M. The optimization on M is then performed based on an exhaustive search approach which will keep increasing M until additional mass points can no longer improve the optimized time-scaled MI.

Table 1. Fiber Parameter.

length L	2000 km
Loss α	0.2 dB/km
Group velocity dispersion factor β_2	-2.1×10^{-26} s^2/m
Kerr nonlinearity factor γ	1.27×10^{-3} /W/m
Phonon occupancy K_T	1.13
Signal wavelength ν_0	1.55 μm
Normalizing time T_0	0.1 ns

Figure 2 shows the capacity-achieving distributions obtained by solving (26) and the corresponding capacity estimation using the optimized input distribution. For these results, we assume an ideal distributed Raman amplified 2000 km fiber with the parameters detailed in Table 1. Using the constraint from $X_{ub} = 200$ to $X_{ub} = 500$. This range of peak amplitude constraint corresponds to the range of maximum eigenvalue from $A_{ub} = 0.4$ to $A_{ub} = 2.5$, which represent the peak optical power -5 dBm and $+10$ dBm, respectively.
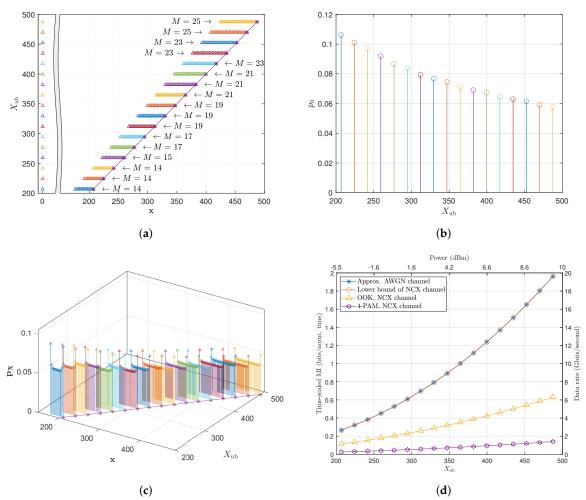
(a)

(b)

(c)

(d)

Figure 2. The optimal input distribution and the corresponding optimized time-scaled mutual information (MI) obtained as the numerical solution of (26) subject to the peak amplitude constraint X_{ub} assuming $\delta = 0.001$. (**a**) The location of the optimal mass points (the peak amplitude is shown as the purple solid line with star) (**b**) The optimal probability of the mass point at zero (i.e., off symbol) (**c**) The optimal probabilities of the nonzero mass points, (**d**) The maximum Time-scaled MI given based on the solution of (26) and the lower bounds on the time-scaled capacity of the original noncentral chi-squared distribution (NCX) channel achieved by using different input distributions, including, on-off keying (OOK), 4 pulse amplitude modulation (4-PAM) and the input distribution given in (**a**) to (**c**). Note that the additional power axis denotes the power level of the solitons corresponds to the peak amplitude X_{ub} assuming $\delta = 0.001$.

In Figure 2a–c, the optimal distributions are shown for various peak amplitude constraints X_{ub}. The figures show that the optimal distributions consist of an isolated mass point at zero (off symbol), and a uniform-like distribution starting from a minimum nonzero symbol (denoted by X_{min}) to the maximum symbol amplitude (denoted by $X_{max} = X_{ub}$). It is also important to point out that the probabilities at X_{min} and X_{max} getting closer to the probabilities of the mass points in between as X_{ub} increases, showing a convergence towards a uniform distribution. Note that the results in [25] shows a nonuniform distribution of optimal mass points since the pulse width is assumed to be variable.

Figure 2d presents the capacity of the approximate AWGN channel based on the solution of (26) as well as some lower bounds on the capacity of the original NCX channel (17). The best lower bound is obtained by applying the optimal distribution of the approximate AWGN channel as in Figure 2a–c to the time-scaled MI of the NCX channel. This lower bound precisely overlaps with the capacity of the approximate AWGN channel, further confirming the result of Theorem 1, in a MI sense, i.e., that the

AWGN channel is a very good approximation of the NCX channel within the range of consideration. Figure 2d also includes the time-scaled MI estimated for the transmission of conventional on-off keying (OOK) and 4 pulse amplitude modulation (4-PAM) signals over the original NCX channel. As expected, both conventional modulations show lower time-scaled MI comparing to the optimized input distribution. However, the conventional 4-PAM signal achieves even lower time-scaled MI than OOK. This is due to the fact that the fixed symbol duration is inversely related to the amplitude of the minimum nonzero amplitude X_{min}, which is $X_{min} = X_{ub}/3$ for 4-PAM but $X_{min} = X_{ub}$ for OOK. In general, for a K-PAM modulation scheme, the time-scaled MI can be upper bounded by the time-scaled source entropy, $\frac{H(X)}{t_s(A_{min},\delta)} = \frac{\sigma_N^2 X_{min}^2}{2\ln(2/\delta-1)}\log_2(K)$, where the $X_{min} = X_{ub}/(K-1)$. It can be then shown that the time-scaled source entropy for K-PAM will always decrease with respect to K for $K \geq 2$. This suggests that K-PAM with higher K cannot achieve better time-scaled MI than OOK. It is also worth noting that some of the sub-optimal distributions proposed in the literature (e.g., the half-Gaussian bound proposed in [23]) is not included here as the half-Gaussian input source would give a zero time-scaled MI when a fixed symbol duration is considered as in this paper.

3.3. Analytical Capacity Approximation

Inspired by the optimal input distributions obtained in the last section as presented in Figure 2, in this section, we focus on developing an analytical approach for time-scaled capacity estimation of the soliton communication system. Assuming that the peak amplitude constraint X_{ub} is sufficiently large, Figure 2 shows that the capacity-achieving input distribution obtained by solving (26) is discrete with a finite number of mass points including an almost uniform distribution within $[X_{min}, X_{max} = X_{ub}]$, and an additional mass points at zero, where the optimal X_{min} needs to be found by solving the optimization problem. We therefore consider a general form of discrete input distribution with a mass point at zero with probability p_0 and a discrete uniform distribution within $[X_{min}, X_{max}]$ to find an analytical estimation of the solution to the capacity problem given in (26). Note that the upper boundaries of the distribution is denoted by $X_{max} \leq X_{ub}$ rather than $X_{max} = X_{ub}$ to keep it inline with the peak amplitude constraint introduced earlier.

To write the corresponding MI based on (25), we first need to define the statistics of the channel output given the input signal parameters, $P_Y(y|p_0, X_{min}, X_{max})$. In order to make the capacity analysis tractable, we make an approximation that the distribution of the noisy output signal Y given the transmission of nonzero mass points, i.e., $P_Y(y|X \in [X_{min}, X_{max}])$ is approximated by a continuous uniform distribution within the range $[X_{min}, X_{max}]$. This approximation is reasonable when the number of mass points M are large and the noise variance is small compared to the signal level. Based on this approximation and also considering the Gaussian noise added to the zero mass point, we can write

$$P_Y(y|p_0, X_{min}, X_{max}) \approx p_0 f_G(y) + \frac{1-p_0}{X_{max} - X_{min}} u(y|X_{min}, X_{max}), \tag{28}$$

where the $f_G(\cdot)$ denotes the PDF of a zero mean, unit variance Gaussian distribution and $u(y|X_{min}, X_{max})$ denotes the step function that is equal to 1 when y is within $[X_{min}, X_{max}]$ and 0 otherwise.

Considering the approximate PDF in (28), we now calculate the differential entropy of the received signal as

$$
\begin{aligned}
h(Y) &= \int_{-\infty}^{+\infty} P_Y(y) \log_2 \frac{1}{P_Y} dy \\
&\overset{a}{\approx} \int_{-\infty}^{X_{\min}} p_0 f_G(y) \log_2 \frac{1}{P_Y} dy + \int_{X_{\min}}^{+\infty} P_Y(y) \log_2 \frac{1}{P_Y} dy \\
&\overset{b}{\approx} \int_{-\infty}^{X_{\min}} p_0 f_G(y) \log_2 \frac{1}{p_0 f_G(y)} dy + \int_{X_{\min}}^{X_{\max}} \frac{1 - p_0}{X_{\max} - X_{\min}} \log_2 \frac{X_{\max} - X_{\min}}{1 - p_0} dy \\
&= p_0 \log_2 \frac{1}{p_0} + p_0 \log_2 \sqrt{2\pi e} + (1 - p_0) \log_2 \frac{X_{\max} - X_{\min}}{1 - p_0},
\end{aligned} \tag{29}
$$

where the approximation a leads from applying the approximate output distribution in (28), and the approximation b is valid under the assumption that $X_{\min} \gg 0$, i.e., $f_G(y \geq X_{\min}) \approx 0$. Substituting (29) into the Equation (25), the approximated MI is then given as a function of p_0, $X_{\min}$ and $X_{\max}$ as

$$
I_{\text{app}}(X;Y) = p_0 \log_2 \frac{1}{p_0} + (1 - p_0) \log_2 \frac{X_{\max} - X_{\min}}{1 - p_0} - (1 - p_0) \log_2 \sqrt{2\pi e}. \tag{30}
$$

Noting that the scaling time (9) is a function of the minimum mass point $X_{\min}$, the approximate time-scaled MI function $R_{\text{app}}(X;Y)$ is then given as

$$
\mathbb{R}_{\text{app}}(X;Y) = \frac{\sigma_N^2 X_{\min}^2}{2 \ln(2/\delta - 1)} \left[p_0 \log_2 \frac{1}{p_0} + (1 - p_0) \log_2 \frac{X_{\max} - X_{\min}}{1 - p_0} - (1 - p_0) \log_2(\sqrt{2\pi e}) \right]. \tag{31}
$$

Theorem 2. *Given the approximated time-scaled MI function in (31), the solution to the capacity problem given in (26), is obtained as*

$$
C_{\text{app}} = R_{\text{app}}(X;Y)|_{p_0^*, X_{\min}^*, X_{\max}^*}, \tag{32}
$$

where the optimal parameters of the input distribution are given as

$$
X_{\max}^* = X_{\text{ub}}, \tag{33}
$$

$$
X_{\min}^* = (X_{\text{ub}} + \sqrt{2\pi e}) \left[1 - \frac{1}{2W\left(\frac{X_{\text{ub}}}{2\sqrt{2\pi}} + \frac{\sqrt{e}}{2}\right)} \right], \tag{34}
$$

$$
p_0^* = \frac{2\sqrt{2\pi e} W\left(\frac{X_{\text{ub}}}{2\sqrt{2\pi}} + \frac{\sqrt{e}}{2}\right)}{X_{\text{ub}} + \sqrt{2\pi e}}, \tag{35}
$$

where $W(\cdot)$ denotes the Lambert W function.

Proof of Theorem 2. The detailed proof of Theorem 2 is shown in Appendix D. □

Using Theorem 2, the approximate solution to the capacity problem in (26) can be calculated analytically. As it can be observed in Figure 3, this approximate capacity result demonstrates a close match to the exact capacity results obtained numerically.
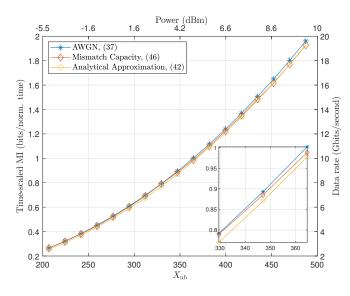
Figure 3. Time-scaled MI estimated from the additive white Gaussian noise (AWGN) model optimization in (26), the analytical approximation in (32), and the corresponding mismatch capacity bound in (36) for a 2000 km long fiber, assuming $\delta = 0.001$. The subplot shows the zoomed figure of $X_{\text{ub}} \in [330, 380]$.

4. Mismatch Capacity for Soliton Communication over the NLSE Channel

So far, we have focused on the capacity estimation of the first-order soliton transmission based on the commonly used memoryless channel model defined by the noncentral chi-squared distribution in (7). In this section, we study the capacity limits of the soliton transmission over a more realistic description of the fibre-optic channel defined by the NLSE. Hence, both the Gordon–Haus effect and the nonlinear interactions between adjacent soliton pulses can be incorporated into the capacity analysis. For this purpose, we use the numerical evaluation of mismatch capacity bounds based on split-step simulation of the NLSE. The mismatch capacity approach is commonly used to provide a lower bound on the capacity of a communication system, by assuming a mismatch distribution for decoding the received signal [32,37]. If the mismatch distribution is denoted by $Q_{Y|X}(y|x)$ and the real channel statistics is denoted by $P_{Y|X}(y|x)$, the time-scaled mismatch capacity bound for a discrete input signal is expressed as

$$
\begin{aligned}
C_{\text{Mismatch}} &= \frac{1}{t_s(A_{\min}, \delta)} \sum_{k=0}^{M} \int_{-\infty}^{+\infty} p_X(x_k) P_{Y|X}(y|x_k) \log \frac{Q_{Y|X}(y|x_k)}{\sum_{j=0}^{M} p_X(x_j) Q_{Y|X}(y|x_j)} dy \\
&= \frac{1}{t_s(A_{\min}, \delta)} \sum_{k=0}^{M} p_X(x_k) \mathsf{E}_{P_{Y|X}(y|x_k)} \left[\log \frac{Q_{Y|X}(y|x_k)}{\sum_{j=0}^{M} p_X(x_j) Q_{Y|X}(y|x_j)} \right],
\end{aligned}
\tag{36}
$$

where $p_X(x_j)$ denotes the input probability of symbol x_j taken from optimization (26), and $\mathsf{E}_{P_{Y|X}(y|x_k)}[\cdot]$ denotes an expectation operation over the channel model $P_{Y|X}(y|x_k)$. Recall from Section 3.1 that the unit-variance Gaussian distribution and the NCX distribution are well matched for the interested range of interest. Thus, a unit-variance Gaussian distribution $Q_{Y|X}(y|x)$ is a reasonable mismatch distribution to be employed in the calculation of the mismatch capacity.

To take into account the impairments introduced by ASE noise, such as Gordon–Haus timing jitter, as well as intersoliton interaction effects, we use the split-step method to simulate the propagation of single soliton or soliton sequence transmission over the fiber. Hence many realizations of the fiber-optic channel can be generated based on the simulation of NLSE to establish the statistics of the realistic channel given the capacity-approaching input distribution obtained in Section 3.2, (i.e., $P(y|x_k)$). The generated channel statistics can then be used to numerically estimate the mismatch capacity in (36)

through a Monte Carlo approach. Noting that the input distribution applied here is not necessarily the optimal distribution for the realistic channel, our results, C_{Mismatch}, provide a lower bound on the mismatch capacity, which in turn gives a lower bound on the capacity of the realistic soliton communication system. The simulation of the channel realization required for the Monte Carlo estimation of mismatch capacity is generated following each function block of the proposed system as in Figure 1. The pulses correspond to the input alphabet will be transmitted into a simulated fiber perturbed by ASE noise via split step Fourier method based on NLSE (1). The output pulse from the simulated fiber will then be put through an NFT detector, which extracts the eigenvalue R from the detected pulse. The received eigenvalue R will then be VNT transformed into the transformed domain for decoding the information. Unless otherwise mentioned, $\delta = 0.001$ is assumed to calculate the soliton duration, i.e., 99.9% soliton energy pulse-width.

4.1. Mismatch Capacity for Single Soliton Transmission

We first focus on single soliton transmission over the NLSE which takes into account the Gordon–Haus effect while ignoring the inter-soliton interaction effects. Using identical fiber parameters as in Table 1, Figure 3 compares the time-scaled mismatch capacity calculated based on 1000 realizations per possible symbol for $X_{\text{ub}} \in [200, 500]$ with the time-scaled capacity of AWGN model obtained in Section 3.2 and the analytical approximation derived in Section 3.3. From Figure 3, it can be observed that the time-scaled MI increases as the peak amplitude constraint increases. It is also observed all the curves provide a well-matched estimations of the capacity, confirming that the Gordon–Haus effect is not so significant within the range of interest here. Nevertheless, we can see that, for larger X_{ub}, the gap between mismatch and AWGN curves increases, which can be due to the stronger Gordon–Haus effect, that will be experienced by larger amplitude soliton pulses. Note that the timing jitter introduced by the Gordon–Haus effect can shift the soliton beyond the limited timing window over which the NFT is applied, which leads to energy loss and possible errors in eigenvalue detection.

4.2. Mismatch Capacity for Soliton Sequence Transmission

The memoryless channel model of soliton communication considered in Section 3 and in most of the literature is only valid when there is no intersoliton interactions, limiting the accuracy of the model to the cases where the sequence of soliton pulses are well separated. In this section, we use the mismatch capacity approach introduced above to provide some insights on the impact of inter-soliton interaction effects on the capacity of soliton communication systems. In the previous section, the performance of the system is discussed based on simulating the transmission of a single soliton pulse through a long haul fiber-optic channel, which neglects the inter-solitonic interactions. In this section, the transmission of a sequence of three soliton pulses is considered, where the middle soliton is considered to be the target soliton for detection. Meanwhile, the neighboring solitons (i.e., the first and the third solitons) are assumed to be independently and randomly selected based on the statistics of the input signal distribution taken from the solution of the AWGN capacity formulation in (26). Note that the pulse width of a soliton is a function of δ and X_{min} in the input signal distribution. The simulation is performed based on the same split step Fourier method employed in Section 4.1, while the NFT-based detection is only performed over the pulse width of the middle soliton.

It has been shown in [38] that, even in the absence of any noise, solitons can exert attracting or repelling forces on each other when they are not place far enough, and this leads to inter-soliton interaction effects. Thus, before implementing the soliton sequence transmission in the presence of the ASE noise, we intend to estimate the mean squared error (MSE) induced by the noiseless inter-soliton interaction to evaluate the significance of this effect for different soliton separations. Recall that the ASE noise power after VNT is normalized to 1. Hence, the inter-soliton interaction effect would be negligible relative to noise, if the inter-soliton interaction MSE is much less than 1, i.e.,

$$MSE = \mathsf{E}[(Y_{\text{nl}} - X)^2] \ll 1, \tag{37}$$

where $\mathsf{E}[\cdot]$ denotes expectation over all possible combination of the three-soliton sequences, Y_{nl} denotes the received VNT transformed eigenvalue in a noiseless scenario. The noiseless simulation is based on the identical simulation parameters as in Table 1 but in the absence of ASE noise (i.e., assuming noiseless ideal distributed Raman amplification) and using the input soliton amplitudes taken from the capacity-approaching distribution given in Section 3.2. In this section, the signaling of the solitons are based on four different δ parameters and their corresponding pulse width. Note that a smaller δ leads to a longer symbol duration as defined by (9), which results in more separation between solitons an thus less inter-soliton interaction.

Figure 4 shows the inter-soliton interaction MSE estimated by simulating the transmission of all possible three-soliton sequences following the input distribution given in Section 3.2 assuming different values of δ. The overall trend of the MSE is increasing as the peak amplitude constraint X_{ub} is increasing. Moreover, as expected, decreasing the δ parameter reduces the MSE. In fact, reducing δ corresponds to the decreasing the fraction of energy truncation that essentially extends the soliton temporal separation. The additional temporal separation will reduce the force between the solitons [38], thus, the inter-soliton interaction is mitigated. Note that, for $\delta = 10^{-3}$, the MSE goes beyond unity for $X_{ub} > 300$ as shown in Figure 4, meaning that the inter-soliton interaction effect becomes comparable to noise beyond that point, hence, the δ parameter needs to be reduced to maintain a low interaction effect. Similarly, it is observed that the MSE becomes comparable to noise for $\delta = 3 \times 10^{-4}$ beyond $X_{ub} = 400$.
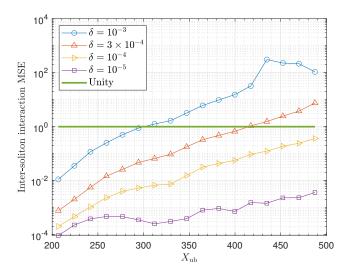

Figure 4. Inter-soliton interaction mean squared error (MSE) for different soliton pulse width determined by different values of δ and based on the link parameters stated in Table 1.

In order to evaluate the impact of intersoliton interaction effect on the capacity of the system, Figure 5 shows the time-scaled capacity results and the corresponding MI calculated based on different proposed methods including the AWGN model and mismatch decoding with or without inter-soliton interaction effects for different values of δ. Figure 5a shows the significant impact of intersoliton interaction effects on the time-scaled capacity at higher peak amplitudes. For example, for $\delta = 10^{-3}$, the time-scaled MI gradually drops beyond $X_{ub} = 300$ and tends to zero before $X_{ub} = 400$. It is also observed that when δ decreases, the longer symbol duration scales down the time-scaled MI in the whole range of X_{ub} but the efficiency of the communication system in combating intersoliton interaction effects improves (i.e., capacity drop shifts to higher soliton amplitudes). This indicates that there is a trade-off in selecting the parameter δ. On the one hand, a smaller δ mitigates more effectively both inter-soliton interaction and Gordon–Haus effects, and on the other hand, it reduces how efficiently the temporal resources are being used. Hence, in future work, δ also needs to be included in the capacity problem formulation. Nevertheless, Figure 5a gives an estimation of sensitivity of the time-scaled

capacity with respect to δ by providing the mismatch results at different values of this parameter. Therefore, by taking the supremum of the curves with different δ values in different parts of the dynamic range, we can obtain a good estimation of the capacity lower bound in the presence of soliton interaction. For example, based on the available results, the capacity result at $\delta = 10^{-3}$ is best up to $X_{ub} = 300$ while the capacity results for $\delta = 3 \times 10^{-4}$ and $\delta = 10^{-4}$ are best in ranges $X_{ub} \in [300, 400]$ and $X_{ub} > 400$, respectively.

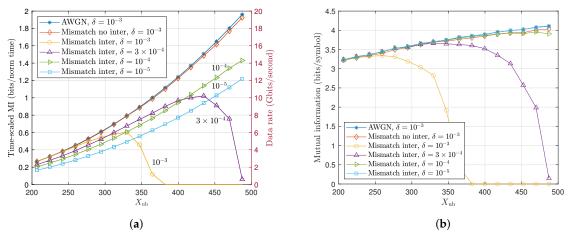(**a**)   (**b**)

Figure 5. The capacity estimation of the soliton communication based on the AWGN model optimization in (26), and the mismatch capacity bounds in the presence (mismatch inter) or absence (mismatch no inter) of inter-soliton interaction effects in terms of (**a**) time-scaled MI and (**b**) MI, for different values of δ and the link parameters stated in Table 1.

The MI results presented in Figure 5b is produced from scaling back the optimized time-scaled MI results in Figure 5a. It therefore focuses on how efficiently each soliton is decoded rather than how efficiently the temporal resources are being used. The figure shows that, for $\delta = 10^{-3}$, the inter-soliton interaction effect strongly degrades the mismatch capacity beyond $X_{ub} = 300$ as expected from Figures 4 and 5a. By reducing δ, it is observed that the inter-soliton interaction effect decreases and it almost matches the mismatch capacity results with no interaction at $\delta = 10^{-4}$. This is also expected from Figure 4, as $\delta = 10^{-4}$ shows $MSE \ll 1$ for most of the range of interest. In addition, the mismatch capacity at $\delta = 10^{-5}$ even outperforms the mismatch capacity with no interaction and almost matches the AWGN result. This is because the mismatch with interaction at $\delta = 10^{-5}$ corresponds to the transmission of a soliton sequence with longer symbol duration. The longer duration essentially eliminates both the Gordon–Haus effect as well as the interaction effects while this is not the case in the mismatch results with no interaction where we still assume shorter pulse width with $\delta = 10^{-3}$. This also verifies the accuracy of the proposed AWGN approximation model compared to the realistic simulated channel when both the Gordon–Haus and inter-soliton interaction effects are negligible.

5. Conclusions

In this paper, we proposed a number of new approaches for estimating the capacity of the amplitude-modulated soliton communication systems. We provided insights into the AIRs of such systems when effects such as Gordon–Haus and inter-soliton interaction are present. The non-central Chi squared channel model that is commonly used in the literature was initially considered and was then approximated by a unit-variance AWGN channel by applying VNT. Using the approximated channel model and subject to a peak amplitude constraint, optimal input distribution and the corresponding capacity were obtained numerically. The optimized distributions are discrete with a mass point at zero corresponding to no soliton transmission as well as an almost uniform distribution of mass points spread in a range away from zero up to the peak amplitude constraint. Using this general form of the optimal distribution based on the approximate AWGN model and applying

some mathematical simplifications, we developed an analytical expression to estimate the capacity of the soliton communication system. Despite the additional approximations, the analytical approach provides a close match to the results obtained numerically based on the AWGN model. The optimal input distribution based on AWGN model were also used to calculate the mismatch capacity of the soliton communication system using the split-step simulation of the realistic channel defined by the NLSE. The results show that the effect of inter-soliton interaction caused by limiting the soliton pulse width is stronger than the Gordon–Haus effect for long haul fibers operating in a range of launch powers up to 10 dBm. They also show the trade-off between extending the pulse width to avoid inter-soliton interaction and compressing the pulse width to improve the temporal efficiency.

In future works, the soliton pulse truncation factor δ can be included in the capacity problem formulation as an additional variable. This allows for a more comprehensive analysis of the soliton interaction effects. Moreover, the capacity problem based on the assumption of variable pulse width can be considered in the presence of soliton interaction effects. Another interesting problem related to this work is the capacity analysis of higher-order soliton transmissions.

Author Contributions: Conceptualization, M.S., I.T. and A.A.; Data curation, Y.C.; Formal analysis, Y.C. and M.S.; Funding acquisition, M.S., Y.C. and A.A.; Investigation, Y.C., I.T. and M.S.; Methodology, I.T. and M.S.; Resources, M.S.; Software, Y.C. and I.T.; Supervision, M.S.; Validation, Y.C.; Writing—original draft, Y.C.; Writing—review & editing, I.T., A.A. and M.S. All authors have read and agreed to the published version of the manuscript.

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

Notations

j: $\sqrt{-1}$; Euler's number: e; Absolute value: $|\cdot|$; Expectation: $\mathrm{E}[\cdot]$; Modified first order Bessel function of the first kind: $I_1(\cdot)$; Probability density function (PDF) of y: $P_Y(y)$; Conditional PDF of y given x: $P_{Y|X}(y|x)$; MI between input X and output Y: $I(X;Y)$; Time-scaled MI between input X and output Y: $\mathbb{R}(X;Y)$ [25]; KL divergence between distribution P and distribution Q given parameter x: $D_{\mathrm{KL}}(P,Q|x)$; Lower constraint: $(\cdot)_{\mathrm{lb}}$; Upper constraint: $(\cdot)_{\mathrm{ub}}$; Minimum nonzero value: $(\cdot)_{\mathrm{min}}$; Maximum nonzero value: $(\cdot)_{\mathrm{max}}$; Optimal value: $(\cdot)^*$; Lambert W function: $W(\cdot)$.

Appendix A

Proof of Proposition 1. Following a similar method as in [32], a non-negative term KL divergence is employed to evaluate difference between two distributions,

$$D_{\mathrm{KL}}(P,Q|x) = \int_{-\infty}^{+\infty} P_{Y|X}(y|x) \ln \frac{P_{Y|X}(y|x)}{Q_{Y|X}(y|x)} dy, \tag{A1}$$

where P and Q denote the distributions, x indicates the given parameter(s) of the two distribution P and Q. Within this proof, $P_{Y|X}(y|x)$ is considered to be a noncentral chi distribution as

$$P_{Y|X}(y|x) = \frac{y^2}{x} \exp\left(-\frac{y^2 + x^2}{2}\right) I_1(xy), \tag{A2}$$

where $I_1(\cdot)$ denotes the modified Bessel function of the first kind, and the mean and variance of $P_{Y|X}$ are denoted with $\mu_{NCX}(x)$ and $\sigma^2_{NCX}(x)$ respectively. $Q_{Y|X}(y|x)$ is considered as a Gaussian distribution with identical mean $\mu_{NCX}(x)$ and variance $\sigma^2_{NCX}(x)$, i.e.,

$$Q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2_{NCX}(x)}} \exp\left(-\frac{(y-\mu_{NCX}(x))^2}{2\sigma^2_{NCX}(x)}\right). \tag{A3}$$

To prove the convergence of the NCX distribution to a Gaussian distribution with mean x and unit variance at a sufficiently large x, we first verify the convergence of the moments of of the NCX distribution and then show its tendency to Gaussian distribution at large x. Taking the limits of the first and second moments at large values of x, we obtain

$$\lim_{x\to\infty} \mu_{NCX}(x) = \lim_{x\to\infty} \sqrt{4 + x^2 - \sigma^2_{NCX}(x)} = \lim_{x\to\infty} \sqrt{3 + x^2} = x, \tag{A4}$$

$$\lim_{x\to\infty} \sigma^2_{NCX}(x) = \lim_{x\to\infty} 4 + x^2 - \mu^2_{NCX}(x) = \lim_{x\to\infty} 4 + x^2 - \frac{\pi}{2}\left[L^{(1)}_{1/2}(-\frac{x^2}{2})\right]^2 = 1, \tag{A5}$$

which verifies the convergence of moments to the corresponding values in the theorem statement. Now substituting the NCX distribution (A2) and its corresponding Gaussian distribution (A3) into (A1), the KL divergence can be expressed as

$$D_{KL}(P,Q|x) = \int_{-\infty}^{+\infty} P_{Y|X}(y|x) \ln(P_{Y|X}(y|x))dy - \int_{-\infty}^{+\infty} P_{Y|X}(y|x) \ln(Q_{Y|X}(y|x))dy$$
$$= -h_{NCX}(x) - E_{NCX}[\ln(Q_{Y|X}(y|x))], \tag{A6}$$

where $h_{NCX}(x)$ denotes the differential entropy of the NCX distribution (A2) given parameter x, and $E_{NCX}(\cdot)$ denotes the expectation over the NCX distribution (A2). The first term can be expressed as

$$h_{NCX}(x) = -E_{NCX}[\ln(P_{Y|X}(y|x))]$$
$$= -E_{NCX}\left[\ln\left(\frac{y^2}{x}\exp\left(-\frac{y^2+x^2}{2}\right)I_1(xy)\right)\right]$$
$$= -2E_{NCX}[\ln(y)] - E_{NCX}[\ln(I_1(xy))] + x^2 + \ln(x) + 2, \tag{A7}$$

while the second term can be written as

$$E_{NCX}\left[\ln\left(Q_{Y|X}(y|x)\right)\right] = E_{NCX}\left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2_{NCX}(x)}}\exp\left(-\frac{(y-\mu_{NCX}(x))^2}{2\sigma^2_{NCX}(x)}\right)\right)\right]$$
$$= \ln\frac{1}{\sqrt{2\pi\sigma^2_{NCX}(x)}} - \frac{1}{2}. \tag{A8}$$

Since function $f(y) = \ln(I_1(xy))$ where x is a given non-negative constant and function $g(y) = \ln(y)$ are concave functions [39], Jensen's inequality is applied to obtain an upperbound on the KL divergence as

$$D_{KL}(P,Q|x) = 2E_{NCX}[\ln(y)] + E_{NCX}[\ln(I_1(xy))] - x^2 - \ln\frac{x}{\sqrt{2\pi\sigma^2_{NCX}(x)}} - \frac{3}{2}$$

$$\leq \ln\frac{(\mu_{NCX}(x))^2 I_1(x\mu_{NCX}(x))\sqrt{2\pi\sigma^2_{NCX}(x)}}{xe^{x^2+3/2}} = D_{ub}(x). \tag{A9}$$

Next, we find the limit of the upper bound of KL divergence $D_{\text{ub}}(x)$ using the limits of mean μ_{NCX} and variance $\sigma^2_{\text{NCX}}(x)$ already calculated in (A4) and (A5), that is

$$\lim_{x \to +\infty} D_{\text{ub}}(x) = \lim_{x \to +\infty} \ln \frac{(3+x^2)I_1(x\sqrt{3+x^2})\sqrt{2\pi}}{xe^{x^2+3/2}} = 0. \tag{A10}$$

At last, using the non-negativity of KL divergence, we have $0 \le \lim_{x\to+\infty} D_{\text{KL}} \le \lim_{x\to+\infty} D_{\text{ub}} = 0$, i.e., $\lim_{x\to+\infty} D_{\text{KL}} = 0$. Therefore, we can conclude that the KL divergence between (A2) and (A3) goes to zero when x is sufficiently large and this concludes the proof.　□

Appendix B

Proof of Theorem 1. In this proof, we show that the gap between the NCX channel capacity and the mismatch capacity of the NCX channel given the approximate AWGN channel as auxiliary channel tends to zero as $X_{\text{lb}} \to \infty$. Consider that the input random variable $X \in \{0 \cup [X_{\text{lb}}, X_{\text{ub}}]\}$ is separated into zero and nonzero sets. Then the PDF of X can be written as

$$P_X(x) = p_0\delta(x) + (1 - p_0)P_{\hat{X}}(x), \tag{A11}$$

where $\delta(x)$ denotes the Dirac delta function, and $P_{\hat{X}}(x)$ denotes the PDF of the nonzero input $\hat{X}$. Similarly, the output random variable Y can also be separated in a similar manners as

$$P_Y(y) = p_0 P_{Y|X}(y|0) + (1 - p_0)P_{\hat{Y}}(y), \tag{A12}$$

where the $P_{\hat{Y}}(y) = \int_{\hat{X}} P_{Y|X}P_{\hat{X}}(x)dx$ denotes the PDF of the output corresponding to the nonzero input. The MI between input X and output Y is then given as

$$I(X;Y) = h(Y) - h(Y|X) = \int_X \int_Y P_X(x)P_{Y|X}(y|x)\log_2 \frac{1}{P_Y(y)}dydx - \int_X \int_Y P_X(x)P_{Y|X}(y|x)\log_2\frac{1}{P_{Y|X}}dydx. \tag{A13}$$

Substituting Equation (A12) in the output differential entropy $h(Y)$, it is then rewritten as

$$\begin{aligned}
h(Y) &= \int_{X_{\text{lb}}}^{X_{\text{ub}}}\int_{X_{\text{lb}}/2}^{+\infty} P_X(x)P_{Y|X}(y|x)\log_2\frac{1}{P_Y(y)}dydx + \int_{-\infty}^{X_{\text{lb}}/2} p_0 P_{Y|X}(y|0)\log_2\frac{1}{P_Y(y)}dy \\
&= \int_{X_{\text{lb}}}^{X_{\text{ub}}}\int_{-X_{\text{lb}}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x)\log_2\frac{1}{p_0 P_{Y'|X}(y'|0) + (1-p_0)P'_{\hat{Y}'}(y')}dy'dx \\
&\quad + \int_{-\infty}^{X_{\text{lb}}/2} p_0 P_{Y|X}(y|0)\log_2\frac{1}{p_0 P_{Y|X}(y|0) + (1-p_0)P_{\hat{Y}}(y)}dy, \tag{A14}
\end{aligned}$$

where we changed the variable of integral in the first term of the last equality as $y' = y - X_{\text{lb}}$. Taking the Taylor expansion of the logarithmic functions inside the two integrals of the right hand side of (A14) at $y' = 0$ and $y = 0$, respectively, we obtain

$$\begin{aligned}
h(Y) &= \int_{X_{\text{lb}}}^{X_{\text{ub}}}\int_{-X_{\text{lb}}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x)\left[\log_2\frac{1}{(1-p_0)P_{\hat{Y}'}(y')} + \sum_{k=1}^{\infty}\frac{(-1)^k}{k}\left(\frac{p_0 P_{Y'|X}(y'|0)}{(1-p_0)P_{\hat{Y}'}(y')}\right)^k\right]dy'dx \\
&\quad + \int_{-\infty}^{X_{\text{lb}}/2} p_0 P_{Y|X}(y|0)\left[\log_2\frac{1}{p_0 P_{Y|X}(y|0)} + \sum_{k=1}^{\infty}\frac{(-1)^k}{k}\left(\frac{(1-p_0)P_{\hat{Y}}(y)}{p_0 P_{Y|X}(y|0)}\right)^k\right]dy \\
&= \int_{X_{\text{lb}}}^{X_{\text{ub}}}\int_{-X_{\text{lb}}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x)\log_2\frac{1}{(1-p_0)P_{\hat{Y}'}(y')}dy'dx + \hat{\Delta} \\
&\quad + \int_{-\infty}^{X_{\text{lb}}/2} p_0 P_{Y|X}(y|0)\log_2\frac{1}{p_0 P_{Y|X}(y|0)}dy + \Delta_0, \tag{A15}
\end{aligned}$$

where the $\hat{\Delta}$ and Δ_0 are higher order terms for nonzero and zero input, respectively. At $X_{lb} \to \infty$,

$\hat{\Delta} = E\left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{p_0 P'(y'|0)}{(1-p_0)P_{\hat{Y}'}(y')}\right]^k\right]$ and $\Delta_0 = E\left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{(1-p_0)P_{\hat{Y}}(y)}{p_0 P_{Y|X}(y|0)}\right]^k\right]$ can be written in the

form of expectations. They will therefore vanish since, for NCX distribution, $\lim_{X_{lb}\to\infty} P_{Y'|X}(y' \leq -X_{lb}/2|x \geq X_{lb}) = 0$ and $\lim_{X_{lb}\to\infty} P_{Y|X}(y \geq X_{lb}/2|0) = 0$. Inserting (A15) in (A13), the MI is given as

$$I(X;Y) = \int_{X_{lb}}^{-X_{ub}} \int_{X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \log_2 \frac{P_{Y'|X}(y'|x)}{(1-p_0)P'_{\hat{Y}'}(y')} dy' dx + \hat{\Delta}$$
$$+ \int_{-\infty}^{X_{lb}/2} P_0 P_{Y|X}(y|0) \log_2 \frac{1}{p_0} dy + \Delta_0. \tag{A16}$$

The mismatch capacity is a proven lower bound of the capacity. Assuming mismatch decoder design based on the Gaussian distribution, $Q_{Y|X}(y|x)$, the mismatch capacity I_{LB} is defined as

$$I_{LB} = \int_X \int_Y P_X(x)P_{Y|X}(y|x) \log_2 \frac{Q_{Y|X}(y|x)}{Q_Y(y)} dy dx, \tag{A17}$$

where the mismatch output distribution $Q_Y(y)$ can be written in similar manner as (A12) as

$$Q_Y(y) = p_0 Q_{Y|X}(y|0) + (1-p_0)Q_{\hat{Y}}(y), \tag{A18}$$

where the $Q_{\hat{Y}}(y) = \int_{\hat{X}} Q_{Y|X}(y|x)P_{\hat{X}}(x)dx$ denotes the PDF of the output corresponding to the nonzero input. The mismatch capacity at X_{lb} can be obtained via similar approach as before as

$$I_{LB}(X;Y) = \int_{X_{lb}}^{X_{ub}} \int_{-X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \left[\log_2 \frac{Q_{Y'|X}(y'|x)}{(1-p_0)Q_{\hat{Y}'}(y')} + \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left(\frac{p_0 Q_{Y'|X}(y'|0)}{(1-p_0)Q_{\hat{Y}'}(y')}\right)^k\right] dy' dx$$
$$+ \int_{-\infty}^{X_{lb}/2} P_0 P'(y'|0) \left[\log_2 \frac{Q_{Y|X}(y|0)}{p_0 Q_{Y|X}(y|0)} + \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left(\frac{(1-p_0)Q_{\hat{Y}}(y)}{p_0 Q_{Y|X}(y|0)}\right)^k\right] dy$$
$$= \int_{X_{lb}}^{X_{ub}} \int_{-X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \log_2 \frac{Q'(y|x)}{(1-p_0)Q_{\hat{Y}'}(y')} dy' dx + \hat{\nabla}$$
$$+ \int_{-\infty}^{X_{lb}/2} P_0 P_{Y|X}(y|0) \log_2 \frac{1}{p_0} dy + \nabla_0, \tag{A19}$$

where the $\hat{\nabla}$ and ∇_0 are higher order terms of the Taylor expansion for nonzero and zero inputs. Similarly, at $X_{lb} \to \infty$, $\hat{\nabla} = E\left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{p_0 Q_{Y'|X}(y'|0)}{(1-p_0)Q_{\hat{Y}'}(y')}\right]^k\right]$ and $\nabla_0 = E\left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{(1-p_0)Q_{\hat{Y}}(y)}{p_0 Q_{Y|X}(y|0)}\right]^k\right]$

can be written in the form of expectations. They will therefore vanish, for AWGN channel, since $\lim_{X_{lb}\to\infty} Q_{Y'|X}(y' \leq -X_{lb}/2|x \geq X_{lb}) = 0$ and $\lim_{X_{lb}\to\infty} Q_{Y|X}(y \geq X_{lb}/2|0) = 0$ for the Gaussian distribution. The gap between the MI $I(X;Y)$ and its lower bound $I_{LB}(X;Y)$ is then defined as

$$I_{gap} = I(X;Y) - I_{LB}(X;Y)$$
$$= \int_{X_{lb}}^{X_{ub}} \int_{-X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \log_2 \frac{P_{Y'|X}(y'|x)}{(1-p_0)P_{\hat{Y}'}(y')} \frac{(1-p_0)Q'_{\hat{Y}'}(y')}{Q_{Y'|X}(y'|x)} dy' dx + \hat{\Delta} - \hat{\nabla} + \Delta_0 - \nabla_0$$
$$= \hat{\Delta} - \hat{\nabla} + \Delta_0 - \nabla_0 + \int_{X_{lb}}^{X_{ub}} \int_{-X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \log_2 \frac{P_{Y'|X}(y'|x)}{Q_{Y'|X}(y'|x)} dy' dx$$
$$- \int_{X_{lb}}^{X_{ub}} \int_{-X_{lb}/2}^{+\infty} P_X(x)P_{Y'|X}(y'|x) \log_2 \frac{P_{\hat{Y}'}(y')}{Q_{\hat{Y}'}(y')} dy' dx. \tag{A20}$$

At $X_{\text{lb}} \to \infty$, the vanishing terms $\hat{\Delta}$, $\hat{\nabla}$, Δ_0 and ∇_0 tends to 0, Hence, the limit of I_{gap} is given by

$$
\begin{aligned}
\lim_{X_{\text{lb}} \to \infty} I_{\text{gap}} &= \int_{X_{\text{lb}}}^{X_{\text{ub}}} P_X(x) \left(\int_{-\infty}^{+\infty} P_{Y'|X}(y'|x) \log_2 \frac{P_{Y'|X}(y'|x)}{Q_{Y'|X}(y'|x)} dy \right) dx \\
&\quad - \int_{-\infty}^{+\infty} \left(\int_{X_{\text{lb}}}^{X_{\text{ub}}} P_X(x) P_{Y'|X}(y'|x) dx \right) \log_2 \frac{P'_{\hat{Y}'}(y')}{Q'_{\hat{Y}'}(y')} dy' \\
&= \int_{X_{\text{lb}}}^{X_{\text{ub}}} P_X(x) D_{\text{KL}}(P', Q'|x) dx - (1-p_0) D_{\text{KL}}(P'_{\hat{Y}'}, Q'_{\hat{Y}'}),
\end{aligned} \tag{A21}
$$

where the second term in (A20) is a nonnegative KL divergence term, hence, the I_{gap} at $X_{\text{lb}} \to \infty$ is bounded by

$$
0 \leq \lim_{X_{\text{lb}} \to \infty} I_{\text{gap}} \leq \int_{X_{\text{lb}}}^{X_{\text{ub}}} P_X(x) D_{\text{KL}}(P', Q'|x) dx = \int_{X_{\text{lb}}}^{X_{\text{ub}}} P_X(x) D_{\text{KL}}(P, Q|x) dx, \tag{A22}
$$

which is the expectation of the KL divergence over the nonzero range of X. According to Proposition 1, $\lim_{x \to \infty} D_{\text{KL}}(P, Q|x \in [X_{\text{lb}}, X_{\text{ub}}])$ for NCX PDF $P_{Y|X}(y|x)$ and Gaussian PDF $Q_{Y|X}(y|x)$ tends to 0, therefore, the upper bound of the I_{gap} also tends to 0. This completes the proof. $\square$

Appendix C

Proof of Proposition 2. Consider the input random variable $X \in \{0 \cup [X_{\text{lb}}, X_{\text{ub}}]\}$ is separated into zero and nonzero sets. Then the probability density function of X can be

$$
P_X(x) = p_0 \delta(x) + (1-p_0) P_{\hat{X}}(x), \tag{A23}
$$

where $\delta(x)$ denotes the Dirac delta function, and $P_{\hat{X}}(x)$ denotes the PDF of the nonzero input $\hat{X}$. Similarly, the output random variable Y can also be separated with similar manners as

$$
P_Y(y) = p_0 P_{Y|X}(y|0) + (1-p_0) P_{\hat{Y}}(y), \tag{A24}
$$

where the $P_{\hat{Y}}(y) = \int_{\hat{X}} P_{Y|X}(y|x) P_{\hat{X}}(x) dx$ denotes the PDF of the output corresponding to the nonzero input. Using the Taylor expansion as in Equation (A15), Appendix B, and considering the AWGN channel model, defined by $P_{Y|X}(y|x)$, the MI is given as

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Y|X) = h(Y) - h(\Gamma) \\
&= \int_{-\infty}^{+\infty} P_Y(y) \log_2 \frac{1}{P_Y(y)} dy - \log_2 \sqrt{2\pi e} \\
&= \int_{X_{\text{lb}}}^{X_{\text{ub}}} \int_{-X_{\text{lb}}/2}^{+\infty} P_X(x) P_{Y'|X}(y'|x) \log_2 \frac{1}{(1-p_0) P'_{\hat{Y}'}(y')} dy' dx + \hat{\Delta} \\
&\quad + \int_{-\infty}^{X_{\text{lb}}/2} P_0 P_{Y|X}(y|0) \log_2 \frac{1}{p_0 P_{Y|X}(y|0)} dy + \Delta_0 - \log_2 \sqrt{2\pi e},
\end{aligned} \tag{A25}
$$

where we changed the variable of integral in the first term of the last equality as $y' = y - X_{\text{lb}}$. At $X_{\text{lb}} \to \infty$, $\hat{\Delta} = E_{\hat{X} \times Y'} \left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{p_0 P'(y'|0)}{(1-p_0) P'_{\hat{Y}'}(y')} \right]^k \right]$ and $\Delta_0 = E_{X_0 \times Y} \left[\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\frac{(1-p_0) P_{\hat{Y}}(y)}{p_0 P_{Y|X}(y|0)} \right]^k \right]$ can

be written in the form of expectations. They will therefore vanish since, for AWGN channel, $\lim_{X_{lb}\to\infty} P_{Y'|X}(y' \leq X_{lb}/2|x \geq X_{lb}) = 0$ and $\lim_{X_{lb}\to\infty} P_{Y|X}(y \geq X_{lb}/2|0) = 0$. Hence, we will have

$$
\begin{aligned}
I_\infty(X;Y) &= \lim_{X_{lb}\to\infty} I(X;Y) \\
&= \int_{X_{lb}}^{X_{ub}} \int_{-\infty}^{+\infty} P_X(x) P_{Y|X}(y|x) \log_2 \frac{1}{(1-p_0)P_{\hat{Y}}(y)} dy dx \\
&\quad + \int_{-\infty}^{+\infty} P_0 P_{Y|X}(y|0) \log_2 \frac{1}{p_0 P_{Y|X}(y|0)} dy - \log_2 \sqrt{2\pi e}, \\
&= p_0 \log_2 \frac{1}{p_0} + (1-p_0) \log_2 \frac{1}{1-p_0} + (1-p_0)\left(h(\hat{Y}) - \log_2 \sqrt{2\pi e}\right) \\
&= h_0 + (1-p_0) I(\hat{X};\hat{Y}),
\end{aligned}
\tag{A26}
$$

where $h_0 = p_0 \log_2 \frac{1}{p_0} + (1-p_0) \log_2 \frac{1}{1-p_0}$, and $I(\hat{X};\hat{Y})$ denotes the MI between the nonzero input $\hat{X}$ and its corresponding output $\hat{Y}$. The time- scaled capacity formulation in (19) is then given as

$$
\begin{aligned}
C &= \sup_{P_X(x):X\in\{0\}\cup[X_{lb},X_{ub}]} \mathbb{R}_\infty(X;Y) = \sup_{P_X(x):X\in\{0\}\cup[X_{lb},X_{ub}]} \frac{I_\infty}{t_s(A_{min},\delta)} \\
&= \sup_{p_0,P_{\hat{X}}(x):x\in[X_{min},X_{ub}]} \frac{\sigma_N^2 X_{min}^2}{2\ln(2/\delta - 1)} \left(h_0 + (1-p_0) I(\hat{X};\hat{Y})\right).
\end{aligned}
\tag{A27}
$$

Let $P_X^*(x)$ denote the capacity achieving distribution for the problem in (A27), which is fully defined by p_0^*, X_{min}^* and the PDF of nonzero input, $P_{\hat{X}}^*$. Next, we will show that this distribution is discrete with a finite number of mass points, which in turn implies the statement of this proposition. We first show that the capacity achieving distribution $P_{\hat{X}}^*$ is also the solution of the following optimization problem

$$
\sup_{P_{\hat{X}}(x):x\in[X_{min}^*,X_{ub}]} I(\hat{X};\hat{Y}).
\tag{A28}
$$

Let $P_{\hat{X}}^\circ$ be any arbitrary distribution within the feasible set of the problem in (A28), implying that $X_{min}^\circ \geq X_{min}^*$. Since $P_{\hat{X}}^*$ defines the capacity achieving distribution for the problem in (A27), it yields a time-scaled MI larger than that of any arbitrary distribution such as $P_{\hat{X}}^\circ$. Therefore we can write

$$
\frac{\sigma_N^2 X_{min}^{*2}}{2\ln(2/\delta - 1)} \left(h_0|_{p_0^*} + (1-p_0^*) I^*(\hat{X};\hat{Y})\right) \geq \frac{\sigma_N^2 X_{min}^{\circ2}}{2\ln(2/\delta - 1)} \left(h_0|_{p_0^*} + (1-p_0^*) I^\circ(\hat{X};\hat{Y})\right),
\tag{A29}
$$

where $I^*(\hat{X};\hat{Y})$ and $I^\circ(\hat{X};\hat{Y})$ are the MI given by $P_{\hat{X}}^*$ and $P_{\hat{X}}^\circ$, respectively. Simplifying (A29) and using the fact that $X_{min}^\circ \geq X_{min}^*$, we have

$$
h_0|_{p_0^*} + (1-p_0^*) I^*(\hat{X};\hat{Y}) \geq \frac{X_{min}^{\circ2}}{X_{min}^{*2}} \left(h_0|_{p_0^*} + (1-p_0^*) I^\circ(\hat{X};\hat{Y})\right) \geq h_0|_{p_0^*} + (1-p_0^*) I^\circ(\hat{X};\hat{Y}),
\tag{A30}
$$

which implies that $I^*(\hat{X};\hat{Y}) \geq I^\circ(\hat{X};\hat{Y})$. Since this is true for any arbitrary $P_{\hat{X}}^\circ$ within the feasible set of the problem in (A28), we can conclude that $P_{\hat{X}}^*$ is also the optimal distribution for the problem in (A28). Note that the problem in (A28) is equivalent to the amplitude constrained AWGN channel capacity problem presented in [34]. In [34], Smith proved that the capacity achieving distribution for such channels are discrete with a finite number of mass points. Thus, the optimal $P_{\hat{X}}^*$ and thereby P_X^* should be discrete with a finite number of mass points as well, which concludes this proof. □

Appendix D

Proof of Theorem 2. Recalling the approximate time-scaled MI function is given as

$$\mathbb{R}_{\text{app}}(X;Y) = \frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta - 1)}\left[p_0 \log_2 \frac{1}{p_0} + (1 - p_0)\log_2 \frac{X_{\max} - X_{\min}}{1 - p_0} - (1 - p_0)\log_2(\sqrt{2\pi e})\right]. \qquad (A31)$$

In order to find the maximum of the function (A31) analytically, its first order partial derivatives with respect to p_0, $X_{\min}$ and $X_{\max}$, are first derived. These first order partial derivatives are given as

$$\frac{\partial \mathbb{R}_{\text{app}}}{\partial X_{\max}} = \frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta + 1)} \frac{1 - p_0}{X_{\max} - X_{\min}}, \qquad (A32)$$

$$\frac{\partial \mathbb{R}_{\text{app}}}{\partial X_{\min}} = \frac{2\sigma_N^2 X_{\min}}{2\ln(2/\delta + 1)}\left(p_0 \ln \frac{1}{p_0} + (1 - p_0)\ln \frac{X_{\max} - X_{\min}}{1 - p_0} - (1 - p_0)\ln \sqrt{2\pi e}\right) -$$
$$\frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta + 1)}\left(\frac{1 - p_0}{X_{\max} - X_{\min}}\right), \qquad (A33)$$

$$\frac{\partial \mathbb{R}_{\text{app}}}{\partial p_0} = \frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta + 1)}\left(\ln \frac{1}{p_0} - \ln \frac{X_{\max} - X_{\min}}{1 - p_0} + \ln \sqrt{2\pi e}\right), \qquad (A34)$$

Notice that (A32) is positive because $X_{\max} > X_{\min}$ and $p_0 < 1$. This implies that the approximate time-scaled MI, $\mathbb{R}_{\text{app}}(\cdot)$, monotonically increases with respect to $X_{\max}$, thus, $\mathbb{R}_{\text{app}}(\cdot)$ maximizes at the boundary as

$$X_{\max}^* = X_{\text{ub}}. \qquad (A35)$$

Now setting the partial derivative in (A33) to zero and using the boundary condition above, we obtain the following nonlinear equation that needs to be solved to obtain the possible optimal value of $X_{\min}$ denoted by $X_{\min}^*$.

$$2\ln \frac{X_{\text{ub}} - X_{\min}^* + \sqrt{2\pi e}}{\sqrt{2\pi e}} = \frac{X_{\min}^*}{X_{\text{ub}} - X_{\min}^* + \sqrt{2\pi e}}. \qquad (A36)$$

Note that the solution to this nonlinear equation can be written based on the Lambert W function $W(\cdot)$ as

$$X_{\min}^* = (X_{\text{ub}} + \sqrt{2\pi e})\left[1 - \frac{1}{2W\left(\frac{X_{\text{ub}}}{2\sqrt{2\pi}} + \frac{\sqrt{e}}{2}\right)}\right]. \qquad (A37)$$

Then, the corresponding probability of the zero mass point can be derived by setting (A34) to zero and using the results above as

$$p_0^* = \frac{\sqrt{2\pi e}}{X_{\text{ub}} - X_{\min}^* + \sqrt{2\pi e}} = \frac{2\sqrt{2\pi e}W\left(\frac{X_{\text{ub}}}{2\sqrt{2\pi}} + \frac{\sqrt{e}}{2}\right)}{X_{\text{ub}} + \sqrt{2\pi e}}. \qquad (A38)$$

Now, in order to show the optimally of p_0^* and $X_{\min}^*$ derived above, the second order partial derivative test should be performed. The second order partial derivative with respect to p_0 and $X_{\min}$ is taken first and are given as

$$\frac{\partial^2 \mathbb{R}_{\text{app}}}{\partial p_0^2} = -\frac{\sigma_N^2 X_{\min}^2}{2\ln(2/\delta + 1)}\left(\frac{1}{p_0} + \frac{1}{1 - p_0}\right), \qquad (A39)$$

$$\frac{\partial^2 \mathbb{R}_{\text{app}}}{\partial X_{\text{min}}^2} = \frac{\sigma_{\text{N}}^2}{2\ln(2/\delta+1)}\left[2\left(p_0\ln\frac{1}{p_0} + (1-p_0)\ln\frac{X_{\text{ub}}-X_{\text{min}}}{1-p_0} - (1-p_0)\ln\sqrt{2\pi e}\right)-\right.$$
$$\left. 4X_{\text{min}}\frac{1-p_0}{X_{\text{ub}}-X_{\text{min}}} - X_{\text{min}}^2\frac{1-p_0}{(X_{\text{ub}}-X_{\text{min}})^2}\right]. \tag{A40}$$

The mixed second order partial derivatives are also required, which are given as

$$\frac{\partial^2 \mathbb{R}_{\text{app}}}{\partial p_0 \partial X_{\text{min}}} = \frac{\partial^2 \mathbb{R}_{\text{app}}}{\partial X_{\text{min}}\partial p_0} = \frac{\sigma_{\text{N}}^2}{2\ln(2/\delta+1)}\left[2X_{\text{min}}\ln\frac{\sqrt{2\pi e}(1-p_0)}{p_0(X_{\text{ub}}-X_{\text{min}})} + \frac{X_{\text{min}}^2}{X_{\text{ub}}-X_{\text{min}}}\right]. \tag{A41}$$

By inspecting Equations (A39) and (A40), one may find that the second order partial derivatives are less than zero at $p_0 = p_0^*$ and $X_{\text{min}} = X_{\text{min}}^*$, while the determinant of the Hessian matrix, $\frac{\partial^2\mathbb{R}_{\text{app}}}{\partial p_0^2}\frac{\partial^2\mathbb{R}_{\text{app}}}{\partial X_{\text{min}}^2} - \frac{\partial^2\mathbb{R}_{\text{app}}}{\partial p_0\partial X_{\text{min}}}\frac{\partial^2\mathbb{R}_{\text{app}}}{\partial X_{\text{min}}\partial p_0}$, is larger than zero. Hence, the maximum of the time-scaled MI in (A31) is obtained at the optimal points $X_{\text{max}}^* = X_{\text{ub}}$, p_0^* and X_{min}^* defined above as

$$C_{\text{app}} = R_{\text{app}}(X;Y)|_{p_0^*, X_{\text{min}}^*, X_{\text{max}}^*}. \tag{A42}$$

□

References

1. Winzer, P.J.; Neilson, D.T. From scaling disparities to integrated parallelism: A decathlon for a decade. *J. Light. Technol.* **2017**, *35*, 1099–1115. [CrossRef]
2. Yousefi, M.I.; Kschischang, F.R. Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools. *IEEE Trans. Inf. Theory* **2014**, *60*, 4312–4328. [CrossRef]
3. Cartledge, J.C.; Guiomar, F.P.; Kschischang, F.R.; Liga, G.; Yankov, M.P. Digital signal processing for fiber nonlinearities. *Opt. Express* **2017**, *25*, 1916–1936. [CrossRef] [PubMed]
4. Bülow, H. Experimental demonstration of optical signal detection using nonlinear Fourier transform. *J. Light. Technol.* **2015**, *33*, 1433–1439. [CrossRef]
5. Gui, T.; Chan, T.H.; Lu, C.; Lau, A.P.T.; Wai, P.K.A. Alternative decoding methods for optical communications based on nonlinear Fourier transform. *J. Light. Technol.* **2017**, *35*, 1542–1550. [CrossRef]
6. Aref, V.; Le, S.T.; Buelow, H. Demonstration of Fully Nonlinear Spectrum Modulated System in the Highly Nonlinear Optical Transmission Regime. In Proceedings of the ECOC 2016—42nd European Conference on Optical Communication, Dusseldorf, Germany, 18–22 September 2016; pp. 1–3.
7. Le, S.T.; Philips, I.D.; Prilepsky, J.E.; Harper, P.; Ellis, A.D.; Turitsyn, S.K. Demonstration of nonlinear inverse synthesis transmission over transoceanic distances. *J. Light. Technol.* **2016**, *34*, 2459–2466. [CrossRef]
8. Tavakkolnia, I.; Safari, M. Signalling over nonlinear fibre-optic channels by utilizing both solitonic and radiative spectra. In Proceedings of the 2015 European Conference on Networks and Communications (EuCNC), Paris, France, 29 June–2 July 2015; pp. 103–107.
9. Hari, S.; Yousefi, M.I.; Kschischang, F.R. Multieigenvalue communication. *J. Light. Technol.* **2016**, *34*, 3110–3117. [CrossRef]
10. Tavakkolnia, I.; Safari, M. Dispersion pre-compensation for NFT-based optical fiber communication systems. In Proceedings of the 2016 Conference on Lasers and Electro-Optics (CLEO), San Jose, CA, USA, 5–10 June 2016; pp. 1–2.
11. Chimmalgi, S.; Wahls, S. Bounds on the Transmit Power of b-Modulated NFDM Systems in Anomalous Dispersion Fiber. *Entropy* **2020**, *22*, 639. [CrossRef]
12. Span, A.; Aref, V.; Bülow, H.; ten Brink, S. Efficient precoding scheme for dual-polarization multi-soliton spectral amplitude modulation. *IEEE Trans. Commun.* **2019**, *67*, 7604–7615. [CrossRef]
13. Zhou, G.; Gui, T.; Lu, C.; Lau, A.P.T.; Wai, P.A. Improving Soliton Transmission Systems Through Soliton Interactions. *J. Light. Technol.* **2019**, *38*, 3563–3572. [CrossRef]
14. Yousefi, M.; Yangzhang, X. Linear and nonlinear frequency-division multiplexing. *IEEE Trans. Inf. Theory* **2019**, *66*, 478–495. [CrossRef]

15. Da Ros, F.; Civelli, S.; Gaiarin, S.; da Silva, E.P.; De Renzis, N.; Secondini, M.; Zibar, D. Dual-polarization NFDM transmission with continuous and discrete spectral modulation. *J. Light. Technol.* **2019**, *37*, 2335–2343. [CrossRef]

16. Derevyanko, S.A.; Turitsyn, S.; Yakushev, D. Non-Gaussian statistics of an optical soliton in the presence of amplified spontaneous emission. *Opt. Lett.* **2003**, *28*, 2097–2099. [CrossRef] [PubMed]

17. Derevyanko, S.A.; Prilepsky, J.E.; Yakushev, D.A. Statistics of a noise-driven Manakov soliton. *J. Phys. A Math. Gen.* **2006**, *39*, 1297. [CrossRef]

18. Zhang, Q.; Chan, T.H. Achievable rates of soliton communication systems. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 605–609.

19. Wahls, S.; Chimmalgi, S.; Prins, P. FNFT: A Software Library for Computing Nonlinear Fourier Transforms. *J. Open Source Softw.* **2018**, *3*, 597. [CrossRef]

20. Zhou, G.; Gui, T.; Chan, T.; Lu, C.; Lau, A.P.T.; Wai, P. Signal processing techniques for nonlinear Fourier transform systems. In *Optical Fiber Communication Conference*; Optical Society of America: Washington, DC, USA, 2019; p. M2H-5.

21. Derevyanko, S.A.; Prilepsky, J.E.; Turitsyn, S.K. Capacity estimates for optical transmission based on the nonlinear Fourier transform. *Nat. Commun.* **2016**, *7*, 1–9. [CrossRef]

22. Tavakkolnia, I.; Safari, M. Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers. *J. Light. Technol.* **2017**, *35*, 2086–2097. [CrossRef]

23. Shevchenko, N.A.; Derevyanko, S.A.; Prilepsky, J.E.; Alvarado, A.; Bayvel, P.; Turitsyn, S.K. Capacity lower bounds of the noncentral chi-channel with applications to soliton amplitude modulation. *IEEE Trans. Commun.* **2018**, *66*, 2978–2993. [CrossRef]

24. Meron, E.; Feder, M.; Shtaif, M. On the achievable communication rates of generalized soliton transmission systems. *arXiv* **2012**, arXiv:1207.0297.

25. Buchberger, A.; i Amat, A.G.; Aref, V.; Schmalen, L. Probabilistic eigenvalue shaping for nonlinear fourier transform transmission. *J. Light. Technol.* **2018**, *36*, 4799–4807. [CrossRef]

26. Hasegawa, A.; Kodama, Y. *Solitons in Optical Communications*; Number 7; Oxford University Press: Oxford, UK, 1995.

27. Safari, M. Efficient optical wireless communication in the presence of signal-dependent noise. In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 1387–1391.

28. Agrawal, G.P. Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 195–211.

29. Ablowitz, M.J.; ABLOWITZ, M.; Prinari, B.; Trubatch, A. *Discrete and Continuous Nonlinear Schrödinger Systems*; Cambridge University Press: Cambridge, UK, 2004; Volume 302.

30. Tavakkolnia, I.; Alvarado, A.; Safari, M. Capacity Estimates of Single Soliton Communication. In Proceedings of the 2018 European Conference on Optical Communication (ECOC), Rome, Italy, 23–27 September 2018; pp. 1–3.

31. Prucnal, P.R.; Saleh, B.E. Transformation of image-signal-dependent noise into image-signal-independent noise. *Opt. Lett.* **1981**, *6*, 316–318. [CrossRef] [PubMed]

32. Tsiatmas, A.; Willems, F.M.J.; Baggen, C.P.M.J. Square Root approximation to the Poisson Channel. In Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 1695–1699.

33. Bartlett, M. The square root transformation in analysis of variance. *Suppl. J. R. Stat. Soc.* **1936**, *3*, 68–78. [CrossRef]

34. Smith, J.G. The information capacity of amplitude-and variance-constrained sclar Gaussian channels. *Inf. Control* **1971**, *18*, 203–219. [CrossRef]

35. Fahs, J.; Tchamkerten, A.; Yousefi, M.I. On the Optimal Input of the Nondispersive Optical Fiber. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 131–135.

36. Dytso, A.; Yagli, S.; Poor, H.V.; Shitz, S.S. The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points. *IEEE Trans. Inf. Theory* **2019**, *66*, 2006–2022. [CrossRef]

37. Ganti, A.; Lapidoth, A.; Telatar, I.E. Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit. *IEEE Trans. Inf. Theory* **2000**, *46*, 2315–2328.

38. Gordon, J. Interaction forces among solitons in optical fibers. *Opt. Lett.* **1983**, *8*, 596–598. [CrossRef]

39. Nanthanasub, T.; Novaprateep, B.; Wichailukkana, N. The logarithmic concavity of modified Bessel functions of the first kind and its related functions. *Adv. Differ. Equ.* **2019**, *2019*, 379. [CrossRef]