



Channel model and the achievable information rates of the optical nonlinear frequency division-multiplexed systems employing continuous b-modulation

STANISLAV DEREVYANKO,^{1,*}  MUYIWA BALOGUN,¹  OFER ALUF,¹
DMITRY SHEPELSKY,^{2,3}  AND JAROSLAW E. PRILEPSKY⁴ 

¹*School of Electrical and Computer Engineering, Ben Gurion University of the Negev, Beer Sheva 84105, Israel*

²*B. Verkin Institute for Low Temperature Physics and Engineering, Kharkiv, 61103, Ukraine*

³*V. Karazin Kharkiv National University, Kharkiv, 61022, Ukraine*

⁴*Aston Institute of Photonic Technologies, Aston University, Birmingham B4 7ET, UK*

**stasd@bgu.ac.il*

Abstract: Following the rise in interest in transmission systems employing the nonlinear Fourier transform (NFT) for the nonlinearity mitigation, we present the theoretical analysis of the achievable information rates in these systems, addressing the case of continuous *b*-modulated systems. Using adiabatic perturbation theory and the asymptotic analysis by means of Riemann-Hilbert problem, we obtain a remarkably simple input-output relation for arbitrary *b*-modulated transmission. Based on this model, we estimated the spectral efficiency for various single polarization (scaled and unscaled) *b*-modulated systems and observed an excellent agreement between our theory and the numerical results in the regime when the inline amplifier noise is the dominant source of spectral distortion.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

The growing demand for the IP traffic volume, escalating due to emerging number of versatile novel on-line services, imposes progressively higher requirements on the throughput of contemporary core fiber communication systems. Due to the ever-increasing traffic demand, we are quickly approaching the limits of available fiber capacity provided by the current (mostly linear) transmission technologies [1–3]. However the nonlinear transmission effects arising largely due to the Kerr nonlinear media response in silica fibers are responsible for the capacity degradation at high signal-to-noise (SNR) levels [2–5], thus making it impossible to gain more line capacity by simply raising the signal power. To reduce the negative fiber nonlinearity impact, there have been proposed a plethora of nonlinearity mitigation techniques [6]. Among the existing methods, the nonlinear Fourier transform (NFT) [7,8] has recently attracted a lot of attention, since within this essentially nonlinear technique, the modulation and transmission effectively takes place inside the nonlinear Fourier (NF) domain, where the nonlinear cross-talk degrading the performance of conventional WDM systems is virtually absent. Such a transmission method employing the NF modes instead of linear counterparts is known as nonlinear frequency division multiplexing (NFDm). The idea of using the NF spectrum components for modulation and transmission was first proposed in [9], where only discrete nonlinear modes were used. Later on the use of continuous spectrum was suggested in [7,10] and a hybrid scheme using both discrete and continuous spectrum was developed in [11,12]. NFT has also been successfully generalized for two polarizations [13–18].

While having the potential to overcome the nonlinearity induced limits, the NFT-based methods still suffer from several limitations. For example, the coupling of NF modes still occurs in the

real systems due to the deviation of the true transmission channel from the idealised model (i.e. from the unperturbed nonlinear Schrödinger equation (NLSE) containing only the effects of the second order dispersion and Kerr nonlinearity). The effects emerging in the multi-user NFDm environment compared to the WDM are analyzed in [19,20]. Inline amplifier noise is also not accounted for by the idealized model and the interaction of the signal evolving inside the NF domain with noise is an area that is still poorly studied despite some recent progress reported in [21–24]. In particular, there is lack of analytical estimates for the achievable information rates for the NFT-based systems despite some results for discrete [25,26] and continuous spectrum [21,23,27]. The main difficulty here is that even in the large SNR regime when one can use perturbation theory, the resulting channel law is strongly input-dependent, which makes capacity estimates notoriously difficult. However, better and more accurate analytical derivations for the channel law are useful as they could be used for estimates for the achievable information rates (AIR) with the mismatched decoding [28–30].

Recently, a specific type of the NFT operating with the continuous spectrum has come to vogue: namely, the so-called b -modulation [31]. In this scheme the data is encoded on the so-called second Jost coefficient $b(\xi)$ [32,33]. The original suggestion [16,17,31,34] favoured b -modulation over the original r -modulation [10] as it allows, in principle, the generation of signals having a compact support, which helps improve spectral efficiency. At the same time, unlike the r -modulated system, the b -modulated system suffers from the so-called "energy barrier" [34] as the value of $|b|$ is upper-bounded by unity. The possible ways of circumventing this limit is using smartly tailored nonlinear signal carriers [34]. Another proposition is to apply the nonlinear spectrum scaling which squeezes the absolute value of b coefficient into the allowed interval at the expense of losing the time-limitedness of the pulse [16,17].

In this paper we report two new results: i) the analytical model (channel-law) for b -modulated nonlinear spectra and ii) the capacity lower bound for these systems. We study both scaled and unscaled b -modulation systems. Some preliminary results of this paper regarding the channel model were reported in [35].

2. Main model and basics of NFT

2.1. Nonlinear Schrödinger equation (NLSE)

The principal master model for the slow-varying envelope of electrical field, $U(z, t)$, propagating down a single-mode optical fiber under the assumption of ideal distributed amplification (i.e. assuming the flat zero-level gain-loss profile) is written as a noise-perturbed NLSE (see [5,36]):

$$\frac{\partial U}{\partial Z} + i \frac{\beta_2}{2} \frac{\partial^2 U}{\partial T^2} - i\gamma |U|^2 U = N(Z, T), \quad (1)$$

with Z being the distance along the fiber, T is the time in the frame co-moving with the velocity of the envelope U . We restrict ourselves to the case of a single polarization only and the parameter β_2 is the characteristic of chromatic dispersion ($\beta_2 < 0$ for the anomalous dispersion case considered in our study), γ is the nonlinear Kerr coefficient. Model (1) represents lossless NLSE perturbed by the zero-mean additive white Gaussian noise (AWGN) term $N(z, t)$ characterized by the power spectral density (PSD) per unit of propagation distance, N_{ASE} , and the autocorrelation function [5,36]:

$$\mathbb{E} [N(Z, T) \bar{N}(Z', T')] = (N_{\text{ASE}}/L) \delta(T - T') \delta(Z - Z'), \quad (2)$$

where the overbar stands for the complex conjugate, $\mathbb{E}[\dots]$ is the expectation value, $\delta(\dots)$ is the Dirac delta-function, and L is the signal propagation distance.

The model above is exact in the case of ideal distributed Raman amplification (IDRA) and neglecting other effects such as the third order dispersion, Raman frequency shift and so on. For the IDRA, we have: $N_{\text{ASE}}/L = h \nu_s K_T \alpha$, where $\alpha \approx 0.2$ dB/km is the fiber loss coefficient at the

carrier wavelength $\lambda_s = 1.55 \mu\text{m}$, $K_T \approx 1$ is the temperature-dependent factor that characterizes the Raman pump providing the gain, $\nu_s = 193.55 \text{ THz}$ is the carrier frequency of the signal.

However, to be more realistic and to describe most transmission experiments, we shall assume the model (1), and (2) to be a path-averaged limit of the lumped erbium doped fiber amplification (EDFA) transmission [37] (see also Ref. [38] for the study of applicable limits of path-averaged model in continuous NFDM systems). In this case the optical pulse is propagated through lossy spans each of length L_a where at the end of each span the signal is regenerated via EDFA gain $G = \exp(\alpha L_a)$. The resulting path-averaged model is the same as in the distributed Raman scheme but with the renormalized nonlinear coefficient $\gamma_{eff} = \gamma(G - 1)/(G \ln G)$ [37,39], and the total accumulated noise ASE is then given by [5]: $N_{ASE} = (G - 1) N_a h \nu_s n_{sp}$, where N_a is the total number of spans ($L = N_a L_a$), and $n_{sp} < 1$ is the spontaneous emission factor commonly evaluated via amplifier noise figure: $F = G^{-1} + 2 n_{sp}(1 - G^{-1})$ [40]. In this case the resulting averaged model (1) represents an approximation valid when the amplifier span is much less than the typical scale of variation of the pulse. Note that although the analytical NFT analysis below assumes the distributed model, the numerical verification of the results was obtained by simulating the lumped amplification link, taking into account periodic power variations and noise injection from span to span. Now let us introduce the normalized units,

$$T/T_s \rightarrow t, \quad Z/Z_s \rightarrow z, \quad U/\sqrt{P_s} \rightarrow q, \quad (3)$$

with $Z_s = T_s^2/|\beta_2|$, $P_s = (\gamma Z_s)^{-1}$. The resulting normalized NLSE becomes:

$$\frac{\partial q}{\partial z} - \frac{i}{2} \frac{\partial^2 q}{\partial t^2} - i|q|^2 q = n(z, t). \quad (4)$$

The noise term $n(z, t)$ has been normalized in accordance to (3), so that it is delta-correlated with the intensity: $2D = N_{ASE}Z_s/(P_s T_s L)$ [21].

2.2. NFT operations

The NFT decomposes the signal $q(z, t)$ at a fixed location $z = z_0$ and returns the corresponding NF spectrum. The reciprocal operation, the inverse NFT (INFT), returns the signal $q(z_0, t)$ taking the full NFT spectrum as an input. In what follows, we shall consider the case of the signal vanishing at infinity and anomalous dispersion fiber.

Direct NFT operation: The direct NFT is computed from specific (auxiliary) solutions $v_{1,2}(t, \zeta) = v_{1,2}(t, \zeta; z_0)$ (in the future we shall omit the z_0 notation for brevity) to the linear Zakharov-Shabat problem (ZSP), which we write down as [33]:

$$\frac{dv_1}{dt} = q(t)v_2 - i\zeta v_1, \quad \frac{dv_2}{dt} = -\bar{q}(t)v_1 + i\zeta v_2, \quad (5)$$

for different values of the complex spectral parameter $\zeta = \xi + i\eta$. The overbar in (5) and below stands for the complex conjugate. Under the assumption that $\int_{-\infty}^{\infty} |q(t)| dt < \infty$, specific solutions (the so-called Jost functions) $\phi_{1,2}(t, \xi)$ and $\psi_{1,2}(t, \xi)$, for real ξ , to the ZSP can be obtained from the boundary conditions:

$$\begin{aligned} \phi_1(t, \xi) &\rightarrow e^{-i\xi t}, & \phi_2(t, \xi) &\rightarrow 0 & \text{for } t \rightarrow -\infty, \\ \psi_1(t, \xi) &\rightarrow 0, & \psi_2(t, \xi) &\rightarrow e^{i\xi t} & \text{for } t \rightarrow +\infty. \end{aligned} \quad (6)$$

Furthermore, we will also need the expression for the conjugated pairs: $\tilde{\phi}_1 = \bar{\phi}_2$ and $\tilde{\phi}_2 = -\bar{\phi}_1$ as well as $\tilde{\psi}_1 = \bar{\psi}_2$ and $\tilde{\psi}_2 = -\bar{\psi}_1$. These also solve ZSP as expressed in (5). In practical realization of the transmission schemes, the pulse $q(t)$ is truncated to have a finite duration, i.e. we operate in the so-called burst mode [41], and thus one sets the initial conditions at the trailing

and leading ends of the finite-extent pulse. The functions $a(\xi)$ and $b(\xi)$ forming the core of the NFT signal decomposition, also known as the *Jost scattering coefficients* are defined through the Jost solution as:

$$a(\xi) = \lim_{t \rightarrow \infty} \phi_1(t, \xi) e^{i\xi t}, \quad b(\xi) = \lim_{t \rightarrow \infty} \phi_2(t, \xi) e^{-i\xi t}. \quad (7)$$

These scattering coefficients satisfy: $|a(\xi)|^2 + |b(\xi)|^2 = 1$. The NF spectrum of the signal $q(t)$ consists of two parts. The first (continuous) part is given by the right reflection coefficient (RC),

$$r(\xi) = b(\xi)/a(\xi), \quad \xi \in \mathbb{R}. \quad (8)$$

The second part of the NF spectrum consists of the discrete NF spectrum responsible for the soliton component of the pulse. However, we do not consider the solitonic degrees of freedom in our study. The z -dependence of the RC and scattering coefficients is as follows:

$$r(\xi, z) = r(\xi, z_0) e^{2i\xi^2(z-z_0)}, \quad b(\xi, z) = b(\xi, z_0) e^{2i\xi^2(z-z_0)}, \quad a(\xi, z) = a(\xi, z_0). \quad (9)$$

INFT operation. The INFT maps the scattering data onto the field $q(t)$. This is achieved via the solution of the Gelfand-Levitan-Marchenko equations (GLME) [33,42], or, alternatively, by means of the Riemann-Hilbert problem [43]. However, we do not present here the explicit expressions as these are not necessary for our further exposition (safe for long distance asymptotes in the [Appendix](#)).

2.3. r and b -modulation

In all general NFT-based schemes similarly to the conventional modulated systems, one starts with a sequence of symbols c_k chosen from a modulation constellation of cardinality M . These symbols, however, are used to modulate the nonlinear spectrum (at the transmitter, i.e. at $z = 0$):

$$u_{in}(\xi) = \sqrt{S} \sum_{k=1}^{N_{sc}} c_k \psi(\xi - k \Delta\xi), \quad (10)$$

where $\psi(\xi)$ is the chosen nonlinear carrier spectral shape, $\Delta\xi$ is the nonlinear carrier spacing, and S is the effective parameter that defines an average power per carrier (the constellation points are supposed to have standard dimensionless spacing). Often one defines a time scale $T_0 = \pi T_s / \Delta\xi$ in the real world units, which can be loosely defined as NFDm symbol duration [16,17]. This definition must, however, be handled with caution as it is known [10] that the nonlinear spectral spacing defines the signal support in the time domain only in the low power limit when the nonlinear spectrum becomes (up to a factor of 2 frequency scaling and complex conjugation) identical to its linear counterpart [42]. As for the carrier shapes, the most popular is the sinc-shape chosen in analogy to linear orthogonal frequency division multiplexing (OFDM) and providing a good localization of resulting profiles in time domain, but other carrier forms can also be employed, see below. Therefore, we shall carry out most of our theoretical analysis for the general carrier shape and will only fix it in the numerical simulations when the specific choice is required. Note that in all implementations of the NFT, the carrier is usually oversampled both in the nonlinear and time domain.

The idea of the NFT-based methods is to take the spectral shape (10) (amplitude-scaled, if desired, by a local transformation) and use it either as the reflection coefficient $r(\xi)$ or as a b -coefficient. The former technique is commonly known as the r -modulation (or the nonlinear inverse synthesis [10]), and the latter technique as the b -modulation. Note that r , a and b coefficients are not independent. They are related to each other by the following relations

(assuming continuous spectrum only) [7,32]:

$$\begin{aligned} |a(\xi)|^2 + |b(\xi)|^2 &= 1, & |a(\xi)|^2 &= (1 + |r(\xi)|^2)^{-1}, \\ \angle a(\xi) &= -\frac{1}{2} \mathcal{H} \{ \log(|a(\xi)|^2) \}, & b(\xi) &= r(\xi) a(\xi), \end{aligned} \quad (11)$$

where $\mathcal{H}\{\dots\}$ denotes the Hilbert transform.

It follows from the first expression in (11) that the absolute value of the b -coefficient is bounded by unity: $|b(\xi)| < 1$. Therefore, the original unscaled b -modulation suggested in [44], while providing an attractive property of localized signal in the time domain (and hence high spectral efficiency at the transmitter), suffered from the “energy barrier” issue – a limitation on the input power S in (10) imposed by the above constraint [44,45]. This limitation was partially circumvented in [34] by using cleverly tailored flat-top carrier shapes and optimized constellations. The authors in [16,17] have chosen another path by imposing no limit on the initial power factor S in (10) but, instead, providing a nonlinear squeezing transformation: $|b| = f(|u|)$ and in the latter case the exact localization property in the time domain was lost, but instead one earns the freedom in signal power up-tuning. It was argued in [16,17] that most of the spectral efficiency in the NFT communication is lost due to the need to insert large guard bands between the bursts [8] and not due to the initial pulse shape within the burst so such a loss of compactness can be tolerated. The specific choice of the exponential function to the best of our knowledge is not dictated by any particular performance gains, but was chosen in [16,17] to provide convenient expression for the total energy of the NFT burst, E_b , via the nonlinear version of the Parseval identity [7,8]: the exponential scaling yields the expression for E_b , which is formally identical to the energy of a pulse sequence with the *linear* spectrum given by Eq. (10). In Table 1 below we summarise some popular choices of scaling used in recent literature. Note that there is still no agreed standard notation in the NFT area and some authors, e.g., [7], use q instead of r to denote the continuous spectrum.

Table 1. Some of the scaling functions used in NFT-based transmission systems, with $u(\xi)$ being the information-bearing modulated function of type (10).

Modulation	Scaling	Ref
r	None: $r(\xi) = u(\xi)$	[10,21,22,24,46]
	$r(\xi) = (e^{ u(\xi) ^2/2} - 1)^{1/2} e^{i\angle u(\xi)}$	[19,20]
b	None: $b(\xi) = u(\xi)$	[31,34]
	$b(\xi) = (1 - e^{- u(\xi) ^2})^{1/2} e^{i\angle u(\xi)}$	[16,17]

The setup of a typical NFDM system using either r or b modulation is shown in Fig. 1. Our data are encoded into the spectral waveform according to Eq. (10), after which an amplitude scaling transform f is applied and the received spectral shape is associated either with r or b component of the nonlinear spectrum. After that, we apply precompensation to the initial encoded spectral shape inside the NF domain, i.e. we unroll the half of accumulated phase in the nonlinear domain (9). This is done to reduce the required guard band of each burst and, thus, increase the spectral efficiency [47]. After the precompensation, the INFT is performed (we use the fast NFT implementation from [48]) and the received time domain signal after inserting the corresponding guard band, is launched into a single mode optical fiber modelled by Eq. (4). For the simulations, we used a standard split-step propagation algorithm and the fiber parameters $\beta_2 = -21 \text{ ps}^2/\text{km}$, $\gamma = 1.27 \text{ W}^{-1} \text{ km}^{-1}$, $\alpha = 0.2 \text{ dB/km}$. The noise was added in a lumped fashion as described in Section 2.1.

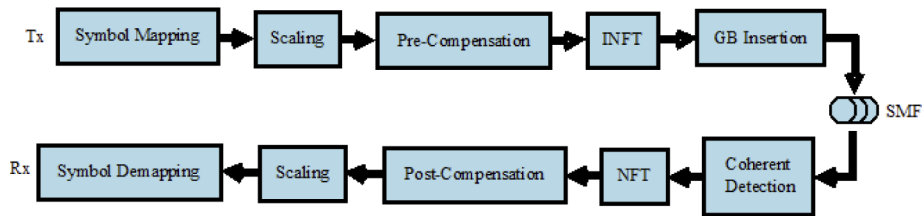


Fig. 1. Basic elements of the NFD optical communication system.

At the receiver the above operations are performed in the reversed order, with the post-compensation equalizing the remaining half of the propagation effects, and the received spectral sequence $u_r(\xi)$ is demodulated according to the chosen carrier shape.

The presence of the noise breaks the exact integrability of the NLSE and makes the full equalization of the nonlinear spectrum impossible. In addition, the numerical errors of both forward and inverse NFT as well as suboptimal processing window and inter-burst interference, all make an independent contribution to the NF spectrum distortion [24]. There are currently plenty of numerical studies of the performance of NFT systems for both r and b modulation case including the achievable information rates (AIR) for these channels [16], but the *analytical* models describing the properties of the amplifier noise in the NS domain are quite few and far between, see Refs. [21,23,27] for the examples related to r -modulation. It was, however, recently reported [16] that the b -modulated systems demonstrate higher tolerance to noise as well as higher achievable spectral efficiency compared to r -modulation. It is, therefore, highly desirable to develop an analytical channel model for the b -modulated NFD systems that could serve two goals: i) provide a possible explanation for the bit error rate (BER) improvement, and ii) give a channel law that could be used to provide capacity estimates via AIR in such systems. In the following section, we build such a model using the perturbation theory for b -coefficient and the asymptotic expression for the Jost functions $\phi_{1,2}$ and $\psi_{1,2}$ taken at large enough propagation distances.

3. Analytical model for b -modulated NFT channel

In this section, we examine noise characteristics for the b -modulated transmission system. We build on our previous approach for r -modulation [21], where the perturbation theory of Kaup and Newell [49,50] was applied to the amplifier noise in Eq. (4) in conjunction with the long-distance asymptotes for the Jost functions [51]. Here, our goal will be similar but we reformulate the approach for the b -modulation. Define by $b_{in}(\xi)$ the modulated sequence (10) after possible prescaling of the amplitude but prior-to precompensation. The pre-compensated b -modulated input is given by $b(\xi, 0) = b_{in}(\xi) \exp(-2i s \xi^2 L/Z_s)$ where $s \in [0, 1]$ specifies the amount of precompensation: when $s = 0$, no precompensation is applied and all compensation is performed at the receiver, while $s = 1$ implies full precompensation. In our simulation we shall use the most common choice, and split the precompensation half-way by setting $s = 1/2$, in order to optimize spectral efficiency. After the postcompensation the output b -coefficient is given by $b_{out}(\xi) = b(\xi, L) \exp[-2i(1-s)\xi^2 L/Z_s]$, from which u_{out} can be restored via the inverse amplitude scaling, if needed.

First, we will demonstrate that in the small-noise limit, the statistics of the output b -coefficient b_{out} conditioned on the input b_{in} are Gaussian but with input dependent covariance. Next, we will show that at long enough distances, one can get relatively simple analytical expressions for these correlation functions.

According to the adiabatic perturbation theory [49,50], the perturbed z -evolution of the a and b coefficient is described by the following set of equations:

$$\frac{\partial a}{\partial z} = I[\phi, \psi], \quad \frac{\partial b}{\partial z} - 2i\xi^2 b = -I[\phi, \tilde{\psi}], \quad (12)$$

where $I[u, v]$ is the projection of the perturbation (in our case of the AWGN noise $n(z, t)$) on the corresponding unperturbed squared Jost functions of the unperturbed system:

$$I[u, v; \xi] = \int_{-\infty}^{\infty} dt \left[n(t, z) u_2(\xi, t, z) v_2(\xi, t, z) + \bar{n}(t, z) u_1(\xi, t, z) v_1(\xi, t, z) \right].$$

The system (12) is incomplete unless one specifies the z -dependence of Jost functions. Since our treatment is a perturbative one, it is sufficient to evaluate $\psi(\zeta, t; z)$ and $\phi(\zeta, t; z)$ in the zeroth order. The important consequence is that since the perturbation terms in the r.h.s. of Eqs. (12) are *linear* with respect to noise term, the effective noise in NF domain is Gaussian albeit its correlation properties depend on the input data and their unperturbed evolution.

Since in this work we concentrate on b -modulation, the second equation from Eqs. (12) allows us to define the noise term in the NFT domain $\mathcal{N}(\xi, b_{in}(\xi))$ according to the following expression:

$$b_{out}(\xi) = b_{in}(\xi) + \mathcal{N}(\xi, L). \quad (13)$$

In the first order of the perturbation theory, given the input $b(\xi, 0)$, the noise term \mathcal{N} is zero mean complex Gaussian and is completely characterised by its (conditional) covariance and pseudocovariance:

$$\mathbb{E} \left[\mathcal{N}(\xi) \bar{\mathcal{N}}(\xi') \right] = 2D e^{2is l(\xi^2 - \xi'^2)} \int_0^l e^{-2i(\xi^2 - \xi'^2)z} A(z; \xi, \xi') dz, \quad (14)$$

$$\mathbb{E} \left[\mathcal{N}(\xi) \mathcal{N}(\xi') \right] = 2D e^{2is l(\xi^2 + \xi'^2)} \int_0^l e^{-2i(\xi^2 + \xi'^2)z} B(z; \xi, \xi') dz, \quad (15)$$

depending on the normalized distance $l = L/Z_s$ and noise intensity D .

The functions A and B are built on the time integral of the products of the Jost functions of the unperturbed system:

$$A = \int_{-\infty}^{+\infty} dt \left[\phi_2(\xi, t, z) \tilde{\psi}_2(\xi, t, z) \bar{\phi}_2(\xi', t, z) \bar{\tilde{\psi}}_2(\xi', t, z) + \phi_1(\xi, t, z) \tilde{\psi}_1(\xi, t, z) \bar{\phi}_1(\xi', t, z) \bar{\tilde{\psi}}_1(\xi', t, z) \right], \quad (16a)$$

$$B = \int_{-\infty}^{+\infty} dt \left[\phi_2(\xi, t, z) \tilde{\psi}_2(\xi, t, z) \phi_1(\xi', t, z) \tilde{\psi}_1(\xi', t, z) + \phi_1(\xi, t, z) \tilde{\psi}_1(\xi, t, z) \phi_2(\xi', t, z) \tilde{\psi}_2(\xi', t, z) \right]. \quad (16b)$$

The Jost functions entering the expressions above generally are not available in the closed form. However, when the propagation distances are large, one can obtain them in an asymptotic form that was first given in [51] and later formalized rigorously by using Riemann-Hilbert technique [52,53]. We note here that the usage of asymptotic expressions for Jost functions ϕ_i, ψ_i , and so on,

in Eqs. (16) and, subsequently, in (14) and (15), is based i) on the assumption that they are reached with the sufficient accuracy at the distances much smaller than the propagation distance L , and this is proven via numerical simulations, and ii) on the cancellations of fast integrand oscillations in integrals entering Eqs. (14) and (15), such that we can keep only the leading term in the resulting approximate expressions. The details of the Jost asymptotics derivation are somewhat technical and can be found in the [Appendix](#). Briefly, finding the Jost functions asymptotics relies on the fact that the off-diagonal elements of jump matrix for the Riemann-Hilbert problem are fast oscillating in the large distance limit, allowing one to evaluate the Jost functions using the nonlinear steepest descent method. The technique can be reckoned as the nonlinear analogue to the evaluation of contour integrals containing oscillatory integrands by the conventional (linear) steepest descent method. A more physically intuitive reasoning given in [51] (see also [54]) is that any solution of the NLSE model evolves asymptotically in a quasilinear way and is centered at the resonant line $t = -2z\xi$. This observation, in turn, can be used to obtain a simplified asymptotic solution to the direct ZSP.

Regardless of the method chosen for finding the Jost function asymptotics, the end result of the calculation is remarkably simple. It turns out that in the large L limit the local covariance is diagonal and symmetric:

$$\begin{aligned} \mathbb{E} \left[\mathcal{N}(\xi) \tilde{\mathcal{N}}(\xi') \right] &= 2\pi D l B_1(\xi) \delta(\xi - \xi'), \quad \text{where} \\ B_1(\xi) &= |a(\xi, L)|^2 = |a(\xi, 0)|^2 = 1 - |b(\xi, 0)|^2 = 1 - |b_{in}(\xi)|^2, \\ \mathbb{E} \left[\mathcal{N}(\xi) \mathcal{N}(\xi') \right] &= 2\pi D l B_2(\xi) \delta(\xi - \xi'), \quad \text{where } B_2(\xi) = 0. \end{aligned} \quad (17)$$

In the above, the functions $B_{1,2}(\xi)$ represent the power spectral density (PSD) and pseudodensity normalized to power spectra density of the linear problem (expressed in normalized units) $n_{ASE} = 2\pi D l$ [5], while recalling that in the linear limit, the linear angular frequency is equal to twice the nonlinear case, so that $\omega = 2\xi$ and $|a(\xi)| \rightarrow 1$ [42]. Note that the n_{ASE} scales linearly with the normalized propagation distance, l , i.e. the noise accumulates along the link.

The expressions in Eqs. (17) are the main analytical result of the paper and according to it, the ASE-induced contribution to the signal distortion *diminishes* as the input approaches the energy barrier: $|b(\xi)| \rightarrow 1$. This is in stark contrast with the r -modulated systems, where it was shown [21,22] that as the energy of the burst grows so does the PSD of the noise in the NFT domain. However, as we shall presently see, simply pushing the b -modulated signal to the energy barrier does not necessarily improve performance insofar as, apart from the ASE, there are other factors contributing to the nonlinear spectrum distortion. This is studied in the numerical simulations described in the next section.

4. Numerical verification of the properties of the nonlinear noise

Throughout the paper we will use the (approximate) burst duration at the transmitter as the time normalization parameter: $T_S = T_0$. Note that T_0 differs from the actual time duration of the pulse because of i) precompensation and ii) the nonlinear nature of the INFT, which makes the duration of any time domain pulse generated by inverse synthesis from a given nonlinear spectrum difficult to control [10]. Only in the limit of the linear system without precompensation does T_0 actually represent the input burst duration.

We start with the validation of the theoretical predictions (17). To this end, we consider an NFDM input sequence (10) with $N_{sc} = 128$, 16-QAM modulated subcarriers. We have considered both scaled and un-scaled b -modulation setups and no precompensation was employed for the noise study, i.e. $s = 0$.

In the case of the scaled b -modulation, the carrier function is the un-normalized sinc: $\psi_0(\xi) = \sin(\xi)/\xi$ with the carrier spacing $\Delta\xi = \pi$ and the exponential scaling $|b(\xi)| = (1 - \exp(-|u(\xi)|^2))^{1/2}$ was used. For the unscaled simulations, we have used the transformed root-raised cosine carriers (28) with the roll-off factor $\beta = 0.05$ described in Section 6.2. In both cases the maximum propagation distance was $L = 960$ km: $N_a = 12$ spans of $L_a = 80$ km.

First, we used Eq. (13) to extract the current realization of the NFT noise. Then we computed the average PSD \bar{B}_1 and the absolute value of its average pseudo-covariance counterpart, $|\bar{B}_2|$, by averaging: i) over 600 realization of the input symbols and inline noise and ii) over the nonlinear spectrum. The procedure is similar to that used in [24] for the r -modulation. The results obtained by such a procedure are presented in Figs. 2, 3, panes (a) and (b). The input power was $P_{in} = -6.8$ dBm, corresponding to $S = 0.1$, in the scaled case, and $P_{in} = -5.55$ dBm (corresponding to $S = 0.0544$) in the unscaled case. According to the theoretical predictions (17), the covariance matrix is diagonal while the pseudocovariance should vanish. And, indeed, one can see that the covariance matrix averaged over input symbols exhibits strong diagonal correlations, while the pseudo-covariance is at least an order of magnitude smaller as predicted by our theory. Therefore, in the following, we will ignore the pseudocovariance and assume the noise in the nonlinear domain to be symmetric.

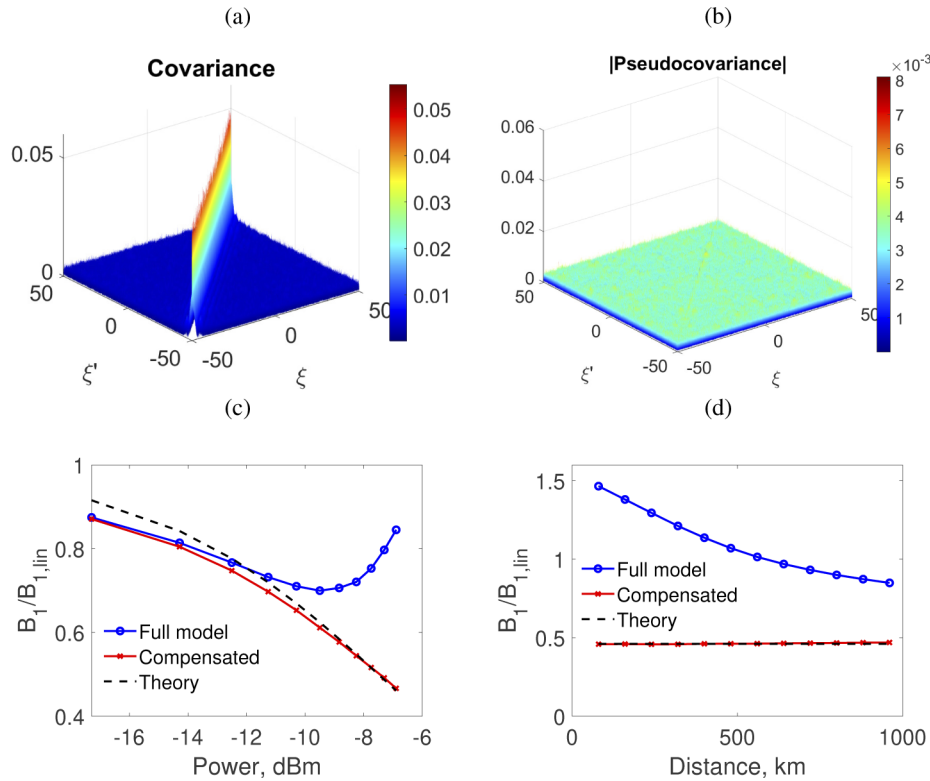


Fig. 2. Scaled b -modulation. 2D covariance (a) and pseudocovariance (b). The normalized averaged PSD as a function of the input power for $L = 960$ km (c), and propagation distance for $P_{in} = -6.8$ dBm (d). The other parameters are summarized in the beginning of Section 4. The “compensated” lines identify the case where the contribution of the processing noise was subtracted from the overall result.

The nonlinear PSD was calculated numerically, and the results are given in Figs. 2, 3 in panes (c), (d), where to obtain the representative quantity, the diagonal part of the covariance matrix

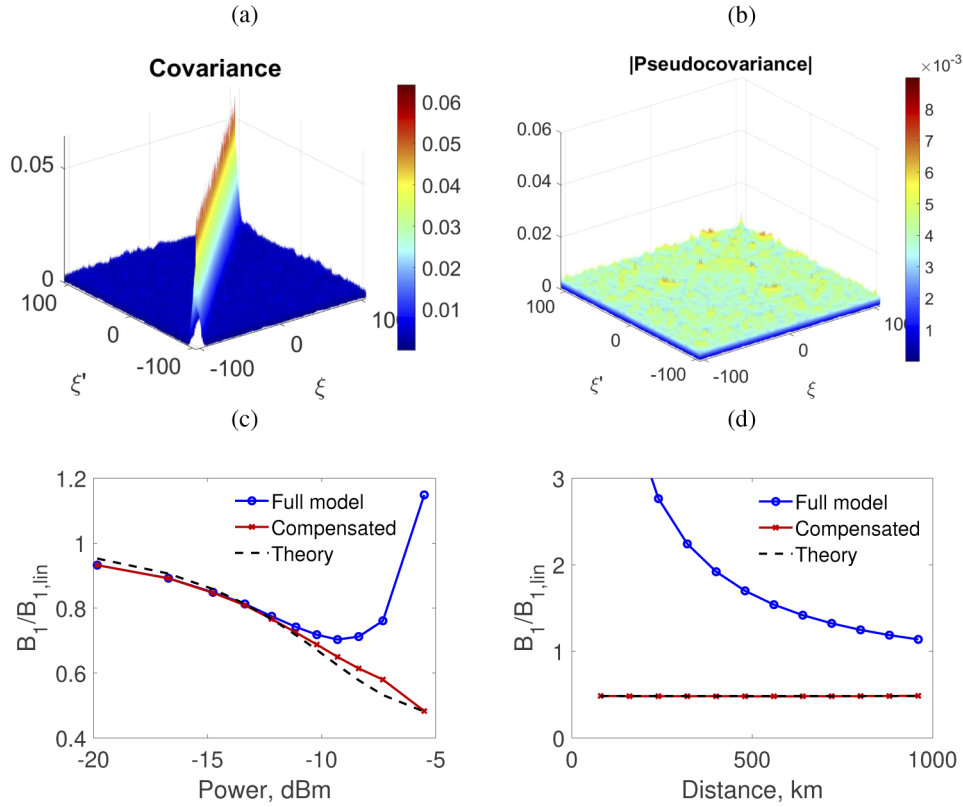


Fig. 3. Unscaled b-modulation. 2D covariance (a) and pseudocovariance (b). The normalized averaged PSD as a function of the input power for $L = 960$ km (c), and propagation distance for $P_{in} = -5.55$ dBm (d). The other parameters are summarized in the beginning of Section 4. The “compensated” lines identify the case where the contribution of the processing noise was subtracted from the overall result.

was averaged over the input symbols and the nonlinear bandwidth. The covariance values in these figures are normalized to the linear value $B_{1,lin} = n_{ASE} = 2\pi D l$. The results were compared to the theoretical prediction (18), where the averaging was performed analytically for the scaled and the unscaled form of b -modulation, in which we used different randomly generated input sequences (10). For the unscaled RRC-shaped carriers, the averaging can be carried out explicitly using the fact that different subcarriers do not overlap in the nonlinear domain. For the scaled case, the averaging amounts to averaging the exponential $\exp(-|u_{in}(\xi)|^2)$, which cannot be carried out explicitly. However, for a large number of subcarriers, the central limit theorem applied to i.i.d. input symbols c_k allows us to substitute $u_{in}(\xi)$ with an effective zero-mean complex symmetric Gaussian process, where the variance can be easily obtained. Combining both results, we get the theoretical prediction:

$$\frac{B_1}{B_{1,lin}} = \mathbb{E}\left[1 - |b_{in}(\xi)|^2\right] \rightarrow \begin{cases} \approx (1 + S\sigma_k^2)^{-1} & \text{scaled,} \\ = 1 - S\sigma_k^2(1 + \beta)^{-1} & \text{unscaled,} \end{cases} \quad (18)$$

where $\sigma_k^2 = \mathbb{E}[|c_k|^2] = 2(M-1)/3$ for M-QAM. Note that for the unscaled b-modulation, we have a natural upper bound for the power, S_{max} , corresponding to the requirement $|b_{in}(\xi)| < 1$. We also notice here that we have a small mismatch between the theory and numerical results in Figs. 2(c)

and 3(c), especially visible for the former. This is because the exact theoretical averaging over the input symbols in Eq. (18) can be obtained only in the limit of infinite number of subcarriers when the central limit theorem applies. Note, however, that even for the exponentially scaled case of Figs. 2(c), where the central limit theorem is less applicable, the maximum discrepancy with the compensated PSD is about 5%, while for the unscaled case of Fig. 3(c), they are almost indistinguishable.

The dimensionless parameter S is related to the real-world unit power via the nonlinear Parseval identity for the burst energy [7,55]:

$$P = \frac{E_b}{T_b} = -\frac{T_s P_s}{\pi T_b} \int_{-\infty}^{\infty} \ln [1 - |b_{in}(\xi)|^2] d\xi, \quad (19)$$

where the burst duration is estimated by the quasilinear dispersive spread of the pulse: $T_b = \varepsilon N_{sc} B |\beta_2| L$, where B is the nonlinear bandwidth of a single subcarrier and $\varepsilon \sim 1$ is carrier-dependent form factor usually determined empirically by considering a noise-free transmission.

As some of us have shown in the previous publication [24], only part of the signal distortion in (13) is due to the ASE noise. Another contribution comes from the processing noise, which arises due to the finite accuracy of the forward/inverse NFT routines and is expected to become more and more prominent as the energy of the burst grows (i.e. $|b_{in}| \rightarrow 1$). And, indeed, as seen in Figs. 2, 3(c), (d), when the input power increases, the PSD of the noise initially goes down as predicted by the model (17) but then begins to rise. To separate the ASE contribution from the processing noise we followed the procedure of Ref. [24] and subtracted the results of separate runs obtained with the ASE noise switched off from the full PSD. The original, uncompensated results are labelled as “Full model” in the graphs while the PSD graphs obtained after the subtraction of the processing noise are labeled as “Compensated”. As one can readily notice, this compensated model corresponding to the isolated contribution of the ASE noise only is fitted extremely well by our theory. The remaining processing noise that becomes relevant at high input powers near the energy barrier, currently lacks an analytical description. While it will be beneficial to include its contribution (if only phenomenologically) into the model (17), we leave this to further studies. Moreover, the properties of the processing noise depend on the particular numerical (I)NFT method used, and so it required a specific discussion in the context of the features and accuracy of particular NFT routines.

5. Input-output channel model

Let us study the implications on the derived approximate analytical model (17). The input-output relation (13) (after pre- and post-compensation) can be written in the following form:

$$b_{out}(\xi) = b_{in}(\xi) + (1 - |b_{in}(\xi)|^2)^{1/2} N(\xi), \quad \mathbb{E} \left[N(\xi) \bar{N}(\xi') \right] = n_{ASE} \delta(\xi - \xi'), \quad (20)$$

where we have introduced the signal-independent linear noise $N(\xi)$, and $\mathbb{E} \left[N(\xi) N(\xi') \right] = 0$. Also, in the quasilinear limit $|b_{in}| \rightarrow 0$, one recovers an AWGN channel in the (linear) spectral domain.

In the case of unscaled transmission, the mapping $b(\xi) = u(\xi)$ and Eq. (20) already defines a nonlinear spectrum waveform channel. For the scaled transmission, we shall assume an arbitrary real, differentiable, invertible scaling function $f(x)$ such as $0 \leq f(x) \leq 1$. In Refs. [16,17], the choice was $f(x) = (1 - \exp(-x^2))^{1/2}$, but we shall keep the scaling function general for now. To get an analytical expressions for the symbol statistics, we will need to assume that $|b_{in}| \gg |N(\xi)|$, so that the second term in (20) can be treated perturbatively. For the first order in the perturbation

term, the amplitude and phase can be expressed as:

$$\begin{aligned} |u_{out}(\xi)| &= |u_{in}(\xi)| + \frac{\sqrt{1-f^2(|u_{in}(\xi)|)}}{f'(|u_{in}(\xi)|)} \operatorname{Re} \left[\bar{N} e^{i\angle u_{in}(\xi)} \right], \\ e^{i\angle u_{out}(\xi)} &= e^{i\angle u_{in}(\xi)} \left[1 + \frac{i\sqrt{1-f^2(|u_{in}(\xi)|)}}{f(|u_{in}(\xi)|)} \operatorname{Im} \left[N e^{-i\angle u_{in}(\xi)} \right] \right]. \end{aligned} \quad (21)$$

Combining the two results together to the same first order accuracy, one can obtain a *linearized* input-output relation:

$$\begin{aligned} u_{out}(\xi) &= u_{in}(\xi) + g_+(|u_{in}(\xi)|) N + g_- (|u_{in}(\xi)|) e^{2i\angle u_{in}(\xi)} \bar{N}, \\ g_{\pm}(x) &= \frac{\sqrt{1-f^2(x)}}{2} \left(\frac{1}{f'(x)} \pm \frac{x}{f(x)} \right). \end{aligned} \quad (22)$$

The linearization procedure used to obtain the input-output relation (22) imposes more severe restrictions than the original perturbative model. Indeed, one can see that if we choose larger input power (controlled by the scaling parameter S in expression (10)), this will make $|u_{in}|$ tend to $+\infty$ and, thus, $1 - f^2(|u_{in}|)$ will tend to zero. So the phase fluctuations in the relations (21) will be squeezed to zero. The same cannot be said, however, of the amplitude fluctuations because of the derivative term in the denominator. Case in point is the exponential squeezing factor of Refs. [16,17], for which the pre-factor in the amplitude relation for $|u|$ is given by $(\exp(x^2) - 1)^{1/2}/x$, $x = |u_{in}|$, and is exponentially diverging. This is no surprise since any monotonic upper bounded function converges to an asymptote (which is 1 in the b -modulated case). Therefore, even small changes in the values of function $|b| = f(|u|)$, as prescribed by our theory, Eq. (20), will induce large changes in the argument $|u|$, which makes the simplified model (22) inapplicable. It is, however, very difficult (if possible at all) to obtain the auxiliary channel law analytically without resorting to linearization. We will, therefore, use model (22) bearing in mind that it is strictly applicable if the input b -coefficient is not too close to unity (which renders linearisation inapplicable) or zero (which violates the validity of the perturbation theory result (25)). The unscaled modulation scheme, which modulates directly the b -coefficient, is free from the former limitation: it follows from Eq. (25) that it makes sense to operate as close to the “energy barrier” $|b_{in}| = 1$ as possible inasmuch as the noise term gets effectively reduced. This fact was noticed in our preliminary report [35].

Now let us apply the channel model obtained in this section to the capacity estimates for the b -modulated NFD systems.

6. Achievable information rates

Finding Shannon capacity of the NFD-modulated optical communication channel (or of the general nonlinear fiber channel for that matter) remains an unsolved problem. The capacity is defined as a maximum of mutual information (MI) defined below, over the probability distribution of the vector of symbols forming the NFD burst subject to (average) input power constraint, see [5]. Since the exact solution for capacity is only available for very simple channels (e.g. for the AWGN channel), then one can only work with various lower bounds obtained by evaluating the MI either analytically or numerically for a given (suboptimal) input distribution $P_x(x)$. For the r -modulated system, the first capacity estimates were obtained in Ref. [21] for continuous input alphabet representing complex values of the initial reflection coefficient: $X_i = r(\xi_i, 0)$.

However, the genuine channel law for the NFD systems is still unknown, and the models like (25), (28) developed here for b -modulated systems, or those derived in Ref. [21,23] for r -modulated systems, still render at best just some approximations of the genuine channel.

Therefore, an alternative approach has recently gained popularity: the technique relying on the AIR and the concept of mismatched decoding [28–30].

Let us begin with some definitions. At the transmitter, the bit stream of k bit forming a message is mapped onto a complex vector of N_{sc} symbols: $\vec{X} = \{c_1, \dots, c_{N_{sc}}\}$, where each symbol c_k is drawn from a modulation constellation χ_M of cardinality M . These symbols are then scaled by the power-tuning factor $S^{1/2}$ and mapped onto the nonlinear spectrum via relation (10) and, possibly, further scaling transformations. Our channel is defined as discrete input - complex output (see e.g. [5]), where the output to the input sequence \vec{X} is the continuous complex vector $\vec{Y} \in \mathbb{C}^{N_{sc}}$. For a general communication channel with memory, the MI is defined as [5,29]:

$$I(\vec{X}, \vec{Y}) = N_{sc}^{-1} \mathbb{E} \left[\log_2 P(\vec{Y}|\vec{X})/P(\vec{Y}) \right], \quad (23)$$

where $P(\vec{Y}|\vec{X})$ is the channel law with memory, $P(\vec{Y})$ is the distribution of the output complex vector induced by the discrete input distribution, and the averaging is performed over the joint input-output distribution $P(\vec{X}, \vec{Y})$. The rigorous definition of MI implies taking the limit $N_{sc} \rightarrow \infty$, but in practical settings the number of subcarriers is always limited. The capacity of the channel, C , is defined as supremum of the MI over all input distributions $P(\vec{X})$ subject to the fixed average input burst power, P_{in} . The AIR is providing a lower bound for the channel capacity by lower-bounding the MI of a given channel for a given input. Specifically, we shall assume that our symbols, c_k , are i.i.d. uniformly drawn from the constellation χ_M . As for the channel law, we substitute the (unknown) genuine channel law $P(\vec{Y}|\vec{X})$ and the induced output PDF $P(\vec{Y})$ with the *auxiliary* ones, $\tilde{P}(\vec{Y}|\vec{X})$ and $\tilde{P}(\vec{Y})$ based on our approximate analytical model, Eqs. (25) and (28). The averaging in (30), nonetheless, is still performed over the *genuine* channel by averaging over the large number of the Monte-Carlo runs of different input bursts for a given input power and collecting the resulting vector of complex outputs, \vec{Y} . The procedure described defines the AIR, and it can be proven that it provides a lower bound for the MI of a genuine channel and, hence, for the capacity [28].

6.1. Scaled b -modulated transmission

The auxiliary channel law, $\tilde{P}(\vec{Y}|\vec{X})$, depends on the particular type of the b -modulated scheme. We begin with the scaled version and the (approximate) input-output relation (28) for the complex spectral wave-form. We shall assume as in the previous section, the nonlinear version of the OFDM scheme where the nonlinear spectral carriers in Eq. (10) are sincs. In the linear OFDM, the demodulation of the sampled time signal at the receiver is performed by means of a discrete Fourier transform. Although through the numerical NFFT algorithm at the receiver one already obtains the nonlinear spectral form u_{out} the removal of guard band and oversampling necessitates the additional linear DFT transform pair (see [16]). Mathematically, this is equivalent to the “matched filtering” albeit in the nonlinear spectral domain, so that the output complex amplitudes forming the output vector $\vec{Y} = \{Y_1, \dots, Y_{N_{sc}}\}$ are obtained via the orthogonal projection:

$$Y_k = \frac{t_0}{\pi \sqrt{S}} \int_{-\infty}^{\infty} u_{out}(\xi) \frac{\sin(\xi t_0 - k\pi)}{\xi t_0 - k\pi} d\xi, \quad (24)$$

with $t_0 = T_0/T_s = 1$ being the useful symbol duration in the normalized units. Then, from (10) and (28) it follows that our communication channel conditioned on the input, is *input-dependent Gaussian with memory*. Moreover, the noise is no longer circularly symmetric and is characterised by both covariance and pseudo-covariance matrices, $\Sigma_{kk'}$ and $\tilde{\Sigma}_{kk'}$, correspondingly, conditioned

on the input \vec{X} (below $CN\{\dots\}$ stands for the multivariate complex normal distribution):

$$\begin{aligned} \tilde{P}(\vec{Y}|\vec{X}) &= CN\left\{\vec{Y}|\vec{X}, \hat{\Sigma}, \hat{\tilde{\Sigma}}\right\}, \\ \Sigma_{kk'}(\vec{X}) &= \mathbb{E}\left[(Y_k - X_k)(\bar{Y}_{k'} - \bar{X}_{k'})\right] = \frac{n_{ASE} t_0^2}{\pi^2 S} \int_{-\infty}^{\infty} \text{sinc}\left[\frac{\xi t_0}{\pi} - k\right] \text{sinc}\left[\frac{\xi t_0}{\pi} - k'\right] \\ &\quad \times \left[g_-^2(|u_{in}(\xi)|) + g_+^2(|u_{in}(\xi)|)\right] d\xi, \\ \tilde{\Sigma}_{kk'}(\vec{X}) &= \mathbb{E}\left[(Y_k - X_k)(Y_{k'} - X_{k'})\right] = \frac{n_{ASE} t_0^2}{\pi^2 S} \int_{-\infty}^{\infty} \text{sinc}\left[\frac{\xi t_0}{\pi} - k\right] \text{sinc}\left[\frac{\xi t_0}{\pi} - k'\right] \\ &\quad \times 2 g_+(|u_{in}(\xi)|) g_- (|u_{in}(\xi)|) e^{2i \angle u_{in}(\xi)} d\xi. \end{aligned} \tag{25}$$

From the above, one can see that as the distance between the symbols k and k' grows, the memory fades due to the diminishing overlap between the sincs. To make the results more tractable, we shall neglect this overlap altogether here and assume the contribution of the diagonal $k = k'$ terms only, so that $\Sigma_{kk'} \approx \sigma_k^2(\vec{X}) \delta_{kk'}$, and $\tilde{\Sigma}_{kk'} \approx \tilde{\sigma}_k^2(\vec{X}) \delta_{kk'}$. Furthermore, in the same approximation we can substitute $|u_{in}(\xi)| \approx \sqrt{S} |X_k| |\text{sinc}(\xi t_0/\pi - k)|$ in all the integrals for the covariance and pseudo-covariance and obtain the memoryless channel with the variance and pseudovariance of each transmitted symbol X_k depending on this symbol alone:

$$\begin{aligned} \sigma^2(X_k) &= \frac{n_{ASE} t_0}{\pi S} \int \text{sinc}^2[x] \left(g_-^2\left[\sqrt{S}|X_k| |\text{sinc}x|\right] + g_+^2\left[\sqrt{S}|X_k| |\text{sinc}x|\right]\right) dx, \\ \tilde{\sigma}^2(X_k) &= \frac{n_{ASE} t_0}{\pi S} \int \text{sinc}^2[x] 2 g_- \left[\sqrt{S}|X_k| |\text{sinc}x|\right] g_+ \left[\sqrt{S}|X_k| |\text{sinc}x|\right] e^{2i \sqrt{S} \angle(X_k \text{sinc}x)} dx. \end{aligned}$$

Note that for a given power value S these values can be precomputed for all constellation points $X_k \in \chi_M$ and then used in a look-up fashion in the numerical simulations.

In the same approximation for the i.i.d. input \vec{X} , the AIR (30) simplifies to

$$\begin{aligned} AIR(S) &= \mathbb{E}\left[\log_2 CN\left\{Y_k|X_k, \sigma^2(X_k), \tilde{\sigma}^2(X_k)\right\} - \log_2 P(Y_k)\right], \\ P(Y_k) &= \frac{1}{M} \sum_{x \in \chi_M} CN\left\{Y_k|x, \sigma^2(x), \tilde{\sigma}^2(x)\right\}, \end{aligned} \tag{26}$$

where the averaging is performed over multiple generated input-output pairs $\{X_k, Y_k\}$ obtained by simulating the real channel.

The above results are approximate inasmuch as they neglect the channel memory due to the small overlap between the distant sinc carriers. A better channel matching (and hence tighter capacity bounds) can be obtained by considering the full model (32), but this makes the estimation task more challenging numerically and so we leave it to further studies.

6.2. Unscaled b -modulated channel

In this subsection we provide the results for the unscaled b -modulation transmission considered in [31,55]. In this scheme $b_{in}(\xi) = u_{in}(\xi)$ given by (10) with the spectral carrier shape ψ_ξ taken to be a spectrally flat with well defined bandwidth $B = \Delta\xi$ and possibly a roll-off factor. We are assuming that the carriers are real functions normalized to have unit peak, and the matched

filtering sampling corresponds to the output sequence:

$$Y_k = \frac{1}{A_\psi \sqrt{S}} \int_{(k-1/2)B}^{(k+1/2)B} u_{out}(\xi) \psi(\xi - kB) d\xi, \quad (27)$$

where $A_\psi \sim B$ is the effective area under the carrier pulse. Note that unlike the case of OFDM-related sinc-shaped carriers considered in the previous subsection, the individual subcarriers here are not orthogonal as such and, therefore, the effective demodulation is achieved by demanding the negligible overlap between them. The orthogonal multiplexing with overlapping subcarriers in the unscaled b-modulation is problematic because the power scaling factor S must be adjusted for each random burst according to the requirement that $\sup_{\xi} |b_{in}(\xi)| < 1$, which is difficult to enforce [34,55]. But for the non-overlapping spectrally flat subcarriers of unit height, one must simply ensure that the power-control parameter satisfies: $S < S_{max} = 1/\max_k |X_k|^2$, i.e. it is upper-bounded by the reciprocal maximum energy of a given constellation.

As for the spectral shape of each non-overlapping carrier, one of the simplest choices that we adopt here is to use that of the root raised cosine (RRC) filter:

$$\psi_{RRC}(\xi) = \begin{cases} 1, & |\xi| < \frac{\pi}{2t_0} (1 - \beta), \\ \frac{1}{\sqrt{2}} \left(1 + \sin \left[\frac{\pi - 2|\xi|t_0}{2\beta} \right] \right)^{1/2}, & \frac{\pi}{2t_0} (1 - \beta) \leq |\xi| \leq \frac{\pi}{2t_0} (1 + \beta), \\ 0, & |\xi| > \frac{\pi}{2t_0} (1 + \beta), \end{cases} \quad (28)$$

with the roll-off factor $0 < \beta < 1$ and the effective ξ -bandwidth is given by: $B_\xi = (\pi/t_0)(1 + \beta)$. The area under the squared subcarrier is: $A_\psi = \pi/t_0$. The RRC subcarriers, however, have a non-localized linear inverse FT and, therefore, relatively large tails in the time domain. We note that most of the results in this subsection are quite general as long as the subcarriers are non-overlapping.

Regardless of the carrier shape used it follows that our auxiliary communication channel conditioned on the input is again input-dependent Gaussian. Nonetheless, unlike the case of nonlinear OFDM subcarriers considered in the previous subsection, the channel now is memoryless because different subcarriers in (10) do not overlap. The input-output relation for this channel is quite simple:

$$Y_k = X_k + \frac{1}{\sqrt{SA_\psi}} \int_{-B_\xi/2}^{B_\xi/2} \psi(\xi) \sqrt{1 - S|X_k|^2 \psi^2(\xi)} N(\xi) d\xi.$$

This corresponds to the circularly symmetric complex Gaussian process with the mean X_k and the following variance:

$$\sigma^2(X_k) = \mathbb{E}[|Y_k - X_k|^2] = \frac{n_{ASE}}{SA_\psi^2} \int_{-B_\xi/2}^{B_\xi/2} \psi^2(\xi) [1 - S|X_k|^2 \psi^2(\xi)] d\xi. \quad (29)$$

In the case of RRC carriers, the integration in (29) can be carried out analytically yielding

$$\sigma^2(X_k) = \frac{n_{ASE} t_0}{\pi S} [1 - (1 - \beta/4) S |X_k|^2].$$

The achievable information rate can now be estimated from (26) setting the pseudo-covariance to zero.

7. Simulation results

In this section we analyze the performance of both scaled and unscaled b-modulated systems and compare the results of the simulations with the predictions of the theory given above. In all the simulations the 12 spans of 80 km were considered and our main quantity of interest was the spectral efficiency (SE) defined as the AIR per unit bandwidth per unit time. In our simulations, as it is common for almost all current NFT-based systems, the transmission is performed in a burst mode with the width of the burst, T_b , usually estimated via the *linear* spread of the pulse. The total linear bandwidth $W = N_{sc}/T_0$ was fixed to be 56 GHz for the scaled b-modulation and to the half of this amount for the unscaled one. The number of subcarriers was a free parameter that was optimized to achieve optimal performance as was done, e.g., in [16]. The fiber channel parameters, amplification scheme, the noise figure are the same as detailed in Section 2.1. The same two types of subcarriers as in the previous section were used (sincs for scaled, RRCs for unscaled b-modulation). To achieve maximal spectral efficiency half-way pre-compensation, $s = 1/2$, was used [47]. We used 32-QAM constellations throughout and the signal was demodulated according to the prescribed rules (24) and (27) respectively. Also a constant phase offset of M-QAM constellation was compensated using the first burst in the simulation as a pilot.

Following Ref. [16], we introduce an SE loss parameter η , which determines how “wasteful” the guard band is with respect to the information content of the pulse. Assuming that a single burst carries N_{sc} modulated symbols each containing $\log_2 M$ bits over the total linear bandwidth W , the connection between the SE and the AIR (26) is:

$$SE = \frac{N_{sc} AIR}{W T_b} = \frac{AIR}{\eta}, \quad (30)$$

where $\eta = (T_b/T_0)(1 + \beta)$, and the roll-off factor β should be put to zero for the scaled sinc-carrier modulation. For the unscaled RRC transmission, we used a range of values between $\beta = 0.1$ and $\beta = 0.15$.

In the last equality in (30), we have made one simplifying assumption: namely, we assume that the linear bandwidth W does not change significantly during the propagation and can be approximated by its nonlinear counterpart (i.e. $W = N_{sc}/T_0$ for the scaled b-modulation based on sincs or $W = N_{sc}(1 + \beta)/T_0$ for the unscaled RRC). This is strictly true only for moderate input powers. However, we have checked that in all cases for the optimal launch powers, the average energy contained outside the linear bandwidth did not exceed 0.17% for the scaled case and 0.02% for the un-scaled one.

The expression in (30) has a few interesting consequences. Firstly, the AIR is upper bounded by the entropy of the discrete input, i.e. $\log_2 M$ and hence the SE cannot exceed $\log_2 M/\eta$ bits/s/Hz. Since the numerator is bounded, increasing the guard band η eventually degrades the SE. On the other hand, as η decreases one starts to clip the tails of the burst, which introduces the additional numerical error and diminishes AIR. It is theoretically possible to have a non-zero limit of SE as $\eta \rightarrow 0$ – the informational content diminishes to zero but it is spread over a vanishing window – so the “efficiency” remains finite. This result is clearly meaningless since no reliable communication is possible in this limit. We therefore, need to define the area of the parameters where the expression in (30) still makes sense. For this expression, we chose the value of η , where the corresponding BER can be made lower than the 7% hard decision FEC threshold: $BER_{FEC} = 4.5 \times 10^{-3}$ [56]. Note that this criterion is somewhat arbitrary and we did not implement the actual FEC in our setup, but used this value as the benchmark of the tolerable quality for the transmission. In our simulations the BER was estimated using standard techniques of error vector magnitude evaluation [57]. It was subsequently translated into the Q-factor using the standard relation: $Q = 20 \log \left[\sqrt{2} \operatorname{erfc}^{-1}(2 BER) \right]$. We also estimated BER via direct error

counting and found a very good agreement below FEC threshold (low Q-factor, high BER) while above the threshold the BER was too low and the direct counting data were too noisy to be used.

The Q-factors for both scaled and unscaled b-modulated systems are shown in Figs. 4(a) and 5(a) respectively. Quite unsurprisingly, the larger values of η parameter translate into the larger guard bands T_b and, as a result, in the larger Q-factor values. The HD-FEC threshold is shown as solid horizontal line there. The minimum value of η above HD-FEC was found to be 1.3 in the scaled case and 1.61 in the un-scaled b-modulation. The latter value is higher, which is non-surprising as the RRC spectra after the INFT render a poorly localized time-domain shapes and suffer from numerical error resulting from the burst truncation. The number of subcarriers N_{sc} for both schemes varied between $N_{sc} = 210$ and $N_{sc} = 1024$, and the reported curves represent the results optimised over N_{sc} . The maximum achieved Baud rates N_{sc}/T_b were 43 Gbaud ($\eta = 1.3$, $N_{sc} = 700$) for the scaled b-modulation and 20 Gbaud ($\eta = 1.61$, $N_{sc} = 322$) for the un-scaled case, corresponding to the bit rates of 215 Gb/s and 100 Gb/s, respectively. To demonstrate that the performance deterioration at high input powers is due to the processing noise discussed in Section 4 and not ASE, we have also included the results of the noise-free runs with the same parameters. One can see from Figs. 4(a), 5(a) that this is indeed the case as both noise-free and noisy curves virtually coalesce in this regime (especially for the scaled modulation).

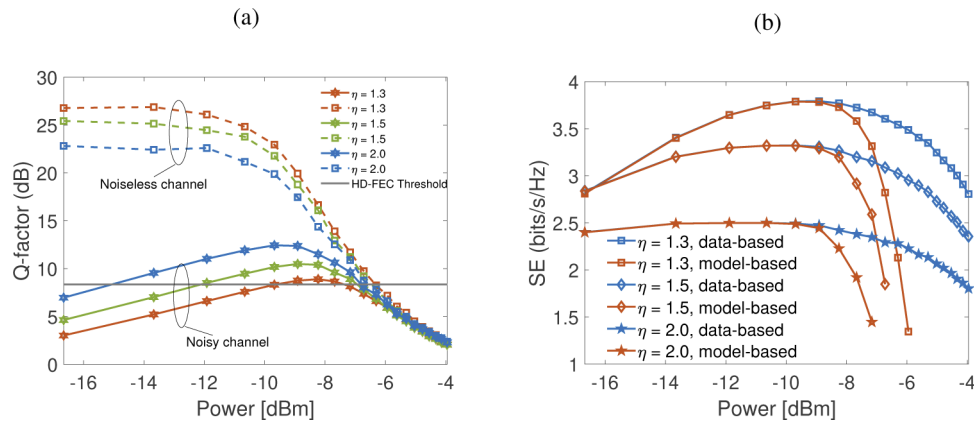


Fig. 4. Scaled b-modulation. Q-factor (a) and the spectral efficiency (b) as functions of input power.

Next, we picked three particular values of parameter η and determined the SE values for these three specific cases. The results are given in Figs. 4(b), 5(b). Two sets of curves are given: the first one is model-based using (26) with the symbol-variances and pseudo-variances given by the theoretical analysis from the previous section. Superimposed on these are the data-based results that fit these pseudo-covariances directly from the data. As one can see, in both cases the data-based SE provides a tighter bound, which is not surprising given that our theory does not take into account the processing noise for which no theoretical model exists to date. For this reason we are not providing the noise-free estimates for the SE (as we did for the Q-factor), as the processing noise takes over at the higher values of the input power, and the performance there is starting to degrade noticeably.

Most importantly, our theory works very well in the ASE-dominated regime up to the optimal value: the relative SE difference between the theory and the data-driven estimates does not exceed 2%. The maximum found value of SE was 3.8 bits/s/Hz for the scaled transmission and 3.04 bits/s/Hz for the un-scaled transmission. The former value is very close to the reported record theoretical SE per polarization from Ref. [16], In fact, our result is slightly higher as the maximal

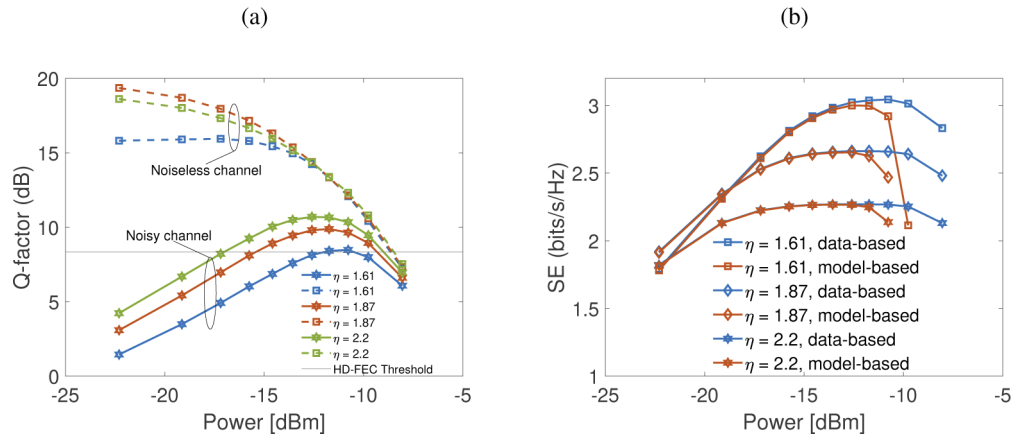


Fig. 5. Unscaled b-modulation. Q-factor (a) and the spectral efficiency (b) as functions of input power.

dual-polarization b-modulation SE was estimated as 7.2 bits/s/Hz in that reference. Note that here, we obtain our results under strict HD-FEC constraint and if these were relaxed, as in Ref. [16], then even higher values of SE (up to 4.5 bits/s/Hz observed in our simulations) can be achieved.

8. Conclusion

In this paper we have developed a rigorous theory for the ASE-induced spectral noise in the NFDN transmission systems using b-modulation. Based on perturbation theory for the path-averaged NLSE and far-distance asymptotes of the Jost functions derived from Riemann-Hilbert formalism, we obtained a remarkably simple input-output relation (25) for the effective b-modulation channel. This results predicts that the ASE part of the spectral distortion is squeezed as the input power grows, and the b -coefficient is pushed to the limit $|b| = 1$. The eventual ASE degradation observed at high power values is attributed solely to the processing noise stemming from the inadequacy of the path-averaged model, burst truncation, and insufficient sampling rate resulting in the (I)NFT numerical routines performance drop [24].

We have used this model to provide general analytical results for the symbol-wise covariance and pseudo-covariance addressing the arbitrary form of b-modulation coding: both the scaled case (where the constraint $|b(\xi)| < 1$ is enforced by applying a squeezing transformation to the information sequence (10)) and the unscaled, that uses flat spectral carriers (we chose root raised cosines), within which the maximum value of the b -coefficient is easily controlled. For both schemes we have obtained an excellent agreement between the theory and the data-based results for the achievable information rate and spectral efficiency in the region where the spectrum is degraded by the ASE. The theory predicts the record spectral efficiency of 3.8 bits/s/Hz for the scaled b-modulation system with the BER remaining below 7% HD-FEC.

As the future avenue of research, we suggest exploring new nonlinear carrier shapes (rather than sines and RRCs) that will can help reducing the SE penalty η while our theoretical result (25) can be used to estimate the AIR, i.e. capacity per symbol. One interesting candidate for such shapes is the recently suggested flat-top carriers [34]. Another promising way to increase SE can be the adaptation of faster than Nyquist signalling for the NFDN as recently reported in [58]. Yet another perspective direction can be combining the discrete NF modes (solitons) with the b-modulation, which can result in the gain in the transmission rate [59] where, however,

the correct analytically-tractable channel model is yet to be built using the results of general nonlinear perturbation theory [27].

Note added in proof. When this manuscript was under review, a paper of Civelli et al. [60] was published, where the authors analyzed modulation and detection strategies for the NFDM systems, including the b-modulation case with two scaling functions (linear and exponential) considered in our work. The authors of Ref. [60] also proposed a bevy of advanced de/modulation strategies accounting for the specific NFDM channel features, which can be employed to increase noise tolerance and boost the transmission quality, and we expect that by using such strategies we can improve the estimates obtained in the current paper even further.

Appendix: Long-distance approximation of the Jost functions

The approximations for A and B functions defined in Eqs. (16) follow from the large z approximations for the Jost solutions in the absence of the discrete spectrum. The latter can be obtained using the Riemann–Hilbert formalism for the inverse problem in the NFT, and we give the respective derivations below. The resulting approximations for the Jost solution are as follows (we use a short-hand notation $a(\xi, 0) = a(\xi)$, $b(\xi, 0) = b(\xi)$):

$$\begin{aligned} \phi_1(t, z, \xi) &= \begin{pmatrix} e^{-i\xi t + i p_+(\mathcal{N}, \xi)} \\ 0 \end{pmatrix} H\left[-\frac{\mathcal{N}}{2} - \xi\right] + \begin{pmatrix} a(\xi) e^{-i\xi t - i p_-(\mathcal{N}, \xi)} \\ b(\xi) e^{2i\xi^2 z + i\xi t + i p_-(\mathcal{N}, \xi)} \end{pmatrix} H\left[\xi + \frac{\mathcal{N}}{2}\right], \\ \phi_2(t, z, \xi) &= \begin{pmatrix} \bar{b}(\xi) e^{-2i\xi^2 z - i\xi t + i p_+(\mathcal{N}, \xi)} \\ a(\xi) e^{i\xi t - i p_+(\mathcal{N}, \xi)} \end{pmatrix} H\left[-\frac{\mathcal{N}}{2} - \xi\right] + \begin{pmatrix} 0 \\ e^{i\xi t + i p_-(\mathcal{N}, \xi)} \end{pmatrix} H\left[\xi + \frac{\mathcal{N}}{2}\right], \end{aligned} \tag{31}$$

where $H(\xi)$ is the Heaviside step function, $\mathcal{N} = t/z$, and

$$p_{\pm}(\mathcal{N}, \xi) = \pm \frac{1}{2\pi} \int_{-\mathcal{N}/2}^{\pm\infty} \frac{\log(1 + |r(s)|^2)}{s - \xi} ds.$$

The expressions for $\phi_{1,2}$ from Eqs. (31) coincide with those derived in Ref. [21] using a different procedure.

Indeed, substituting (31) into (16), using the fact that $(\tilde{\psi}_1, \tilde{\psi}_2)^T = (\bar{\psi}_2, -\bar{\psi}_1)^T$, and assuming that ξ is close to ξ' , we obtain:

$$\begin{aligned} A(z; \xi, \xi') &= \int_{-\infty}^{-2\xi z} dt |a(\xi)|^2 \exp\left\{-2i(\xi - \xi')t + 2i[p_+(\mathcal{N}, \xi) - p_+(\mathcal{N}, \xi')]\right\} \\ &+ \int_{-2\xi z}^{\infty} dt |a(\xi)|^2 \exp\left\{-2i(\xi - \xi')t - 2i[p_-(\mathcal{N}, \xi) - p_-(\mathcal{N}, \xi')]\right\}. \end{aligned}$$

Here, we have put $\xi = \xi'$ in all the prefactors keeping the dependence only in the exponential terms. Note that the main contribution to the integral comes from the terms linearly oscillating in t , whereas the corrections $p_{\pm}(\mathcal{N}, \xi)$ depending on t via the integration limit decay as $t \rightarrow +\infty$ and $t \rightarrow -\infty$, respectively, such that their contribution can be neglected when $\xi' \rightarrow \xi$. Therefore, the only time dependence of the integrand comes from the oscillating factors $\exp(\pm 2i(\xi - \xi')t)$, and, using the identity $\int_{-\infty}^{\infty} e^{i\xi t} dt = 2\pi\delta(\xi)$, we arrive at

$$A(z; \xi, \xi') = \pi |a(\xi)|^2 \delta(\xi - \xi').$$

On the other hand, substituting (31) into the expression for B in (16), we notice that each summand contains a component ϕ_j or ψ_j that, according to (31), vanishes. This implies that

$B(z; \xi, \xi') = 0$ in the same approximation. Both results upon substitution into (14)–(15), yield Eqs. (17).

In turn, the approximations (31) follow from the representation of the Jost solutions in terms of the solution of a Riemann–Hilbert factorization problem (RHP) [43] and the subsequent large- z analysis of this problem [52,61]. Namely,

$$\begin{pmatrix} \frac{\phi(t, z, \xi)}{a(\xi)}, \psi(t, z, \xi) \end{pmatrix} = M_+(t, z, \xi) \begin{pmatrix} e^{-i\xi t} & 0 \\ 0 & e^{i\xi t} \end{pmatrix}, \tag{32}$$

where $M_+(t, z, \xi)$ is the limiting values of $M(t, z, \zeta)$ as $\zeta \in \mathbb{C}$ approaches $\xi \in \mathbb{R}$ from above, and $M(t, z, \zeta)$, $\zeta \in \mathbb{C} \setminus \mathbb{R}$ is the solution of the following RHP: given $r(\xi)$, $\xi \in \mathbb{R}$, find a 2×2 -valued function $M(t, z, \zeta)$ analytic in $\zeta \in \mathbb{C} \setminus \mathbb{R}$, such that

(i) its limiting values $M_{\pm}(\cdot, \cdot, \xi)$ as $\zeta \rightarrow \xi$ from $\mathbb{C}_{\pm} = \{\zeta \mid \pm \Im \zeta > 0\}$ are related by

$$M_+(t, z, \xi) = M_-(t, z, \xi)J(t, z, \xi), \quad \xi \in \mathbb{R},$$

where

$$J(t, z, \xi) = \begin{pmatrix} 1 + |r(\xi)|^2 & \bar{r}(\xi)e^{-2i\xi t - 2i\xi^2 z} \\ r(\xi)e^{2i\xi t + 2i\xi^2 z} & 1 \end{pmatrix};$$

(ii) $M(t, z, \zeta) \rightarrow I$ as $\zeta \rightarrow \infty$.

The large- z analysis of the solution of the RHP above follows the ideas of the nonlinear steepest descent method for oscillatory RHP [52,61] and consists of a series of transformations reducing the original RHP with oscillatory (with z) jump conditions to a RHP with the jump matrix rapidly decaying to I as $z \rightarrow \infty$. This series involves: (i) factorizing the jump matrix on \mathbb{R} into triangular factors in accordance with the “signature table”, i.e., the distribution of signs of $\Im \theta(\varkappa, \zeta)$ with $\theta(\varkappa, \zeta) = \varkappa \zeta + \zeta^2$ having $\zeta = -\varkappa/2$ as the intersection point; (ii) adopting the triangular factors into the solution of the RHP problem, replacing the (original) jump on \mathbb{R} by the jumps across $\Sigma = \cup_{j=1}^4 \Sigma_j$ with $\Sigma_j = \{\zeta \mid \zeta = -\varkappa/2 + \tau e^{i(-\frac{\pi}{4} + \frac{\pi j}{2})}, \tau \in (0, \infty)\}$. The resulting transformation producing a RHP for $M^{(appr)}$ with the jump conditions $M_+^{(appr)}(\varkappa, z, \zeta) = M_-^{(appr)}(\varkappa, z, \zeta)J^{(appr)}(\varkappa, z, \zeta)$, $\zeta \in \Sigma$, has the form:

$$M^{(appr)} = M\Delta P,$$

where

$$\Delta = \begin{pmatrix} \delta^{-1}(\varkappa, \zeta) & 0 \\ 0 & \delta(\varkappa, \zeta) \end{pmatrix},$$

with

$$\delta(\varkappa, \zeta) = \exp \left\{ \frac{1}{2\pi i} \int_{-\infty}^{-\varkappa/2} \frac{\log(1 + |r(s)|^2)}{s - \zeta} ds \right\}, \quad \zeta \in \mathbb{C} \setminus \left(-\infty, -\frac{\varkappa}{2}\right],$$

and

$$P(\varkappa, z, \zeta) := \begin{cases} \begin{pmatrix} 1 & 0 \\ -r\delta^{-2}e^{2iz\theta} & 1 \end{pmatrix}, & \zeta \in G_1, \\ \begin{pmatrix} 1 & \bar{r}\delta^2e^{-2iz\theta} \\ 0 & 1 \end{pmatrix}, & \zeta \in G_4, \\ \begin{pmatrix} 1 & -\frac{\bar{r}\delta^2}{1+|r|^2}e^{-2iz\theta} \\ 0 & 1 \end{pmatrix}, & \zeta \in G_2, \\ \begin{pmatrix} 1 & 0 \\ \frac{r\delta^{-2}}{1+|r|^2}e^{2iz\theta} & 1 \end{pmatrix}, & \zeta \in G_3, \\ I, & \text{otherwise.} \end{cases},$$

In the expression above, G_j , $j = 1, \dots, 4$, are the sectors between \mathbb{R} and Σ_j . According to the signature table, the jump matrices $J^{(appr)}(\varkappa, z, \zeta)$ decay exponentially to I as $z \rightarrow \infty$ for any $\zeta \in \Sigma \setminus \{-\varkappa/2\}$. Thus, the first approximation for $M^{(appr)}$ is: $M^{(appr)}(\varkappa, z, \zeta) \approx I$. This leads to the approximation:

$$M \approx P^{-1}\Delta^{-1},$$

which, in view of (32), $1 + |r(\xi)|^2 = |a(\xi)|^{-2}$ for $\xi \in \mathbb{R}$, and

$$a(\xi) = \exp \left\{ \frac{i}{2\pi} \int_{-\infty}^{\infty} \frac{\log(1 + |r(s)|^2)}{s - \xi - i0} ds \right\},$$

gives (31).

We note that the nonlinear steepest descent method allows improving the approximation for $M^{(appr)}$ to $M^{(appr)}(\varkappa, z, \zeta) = I + O(z^{-1/2})$, where the term of order $z^{-1/2}$ can be written explicitly [52,53].

Funding. Leverhulme Trust (RPG-2018-063); Israel Science Foundation (466/18).

Acknowledgments. DS and JEP acknowledge the Erasmus+ mobility exchange programme between Aston University and V. Karazin Kharkiv National University. SD and JEP acknowledge the Erasmus+ mobility exchange programme between Aston University and Ben Gurion University of the Negev.

Disclosures. The authors declare no conflicts of interest.

References

1. E. Agrell, M. Karlsson, A. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, and B. J. Eggleton, "Roadmap of optical communications," *J. Opt.* **18**(6), 063002 (2016).
2. P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. Killely, "Maximizing the optical network capacity," *Phil. Trans. R. Soc. A* **374**(2062), 20140440 (2016).
3. A. Ellis, M. McCarthy, M. Al Khateeb, M. Sorokina, and N. Doran, "Performance limits in optical communications due to fiber nonlinearity," *Adv. Opt. Photonics* **9**(3), 429–503 (2017).
4. P. Mitra and J. Stark, "Nonlinear limits to the information capacity of optical fiber communications," *Nature* **411**(6841), 1027–1030 (2001).
5. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightwave Technol.* **28**(4), 662–701 (2010).
6. J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities," *Opt. Express* **25**(3), 1916–1936 (2017).
7. M. Yousefi and F. Kschischang, "Information transmission using the nonlinear fourier transform, part i–iii," *IEEE Trans. Inf. Theory* **60**(7), 4312–4369 (2014).
8. S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, "Nonlinear fourier transform for optical data processing and transmission: advances and perspectives," *Optica* **4**(3), 307–322 (2017).
9. A. Hasegawa and T. Nyu, "Eigenvalue communication," *J. Lightwave Technol.* **11**(3), 395–399 (1993).
10. J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn, "Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels," *Phys. Rev. Lett.* **113**(1), 013901 (2014).

11. V. Aref, S. Le, and H. Buelow, "Modulation over nonlinear fourier spectrum: Continuous and discrete spectrum," *J. Lightwave Technol.* **36**(6), 1289–1295 (2018).
12. S. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the kerr nonlinearity limit," *Nat. Photonics* **11**(9), 570–576 (2017).
13. J. W. Goossens, M. I. Yousefi, Y. Jaouan, and H. Hafermann, "Polarization-division multiplexing based on the nonlinear fourier transform," *Opt. Express* **25**(22), 26437–26452 (2017).
14. S. Gaiarin, A. M. Perego, E. P. da Silva, F. Da Ros, and D. Zibar, "Dual-polarization nonlinear fourier transform-based optical communication system," *Optica* **5**(3), 263–270 (2018).
15. S. Civelli, S. Turitsyn, M. Secondini, and J. Prilepsky, "Polarization-multiplexed nonlinear inverse synthesis with standard and reduced-complexity nft processing," *Opt. Express* **26**(13), 17360–17377 (2018).
16. 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," *J. Lightwave Technol.* **37**(6), 1570–1578 (2019).
17. X. Yangzhang, V. Aref, S. T. Le, H. Buelow, D. Lavery, and P. Bayvel, "Experimental demonstration of dual-polarisation nfdm transmission with *b*-modulation," *IEEE Photonics Technol. Lett.* **31**(11), 885–888 (2019).
18. F. Da Ros, S. Civelli, S. Gaiarin, E. P. da Silva, N. De Renzis, M. Secondini, and D. Zibar, "Dual-polarization nfdm transmission with continuous and discrete spectral modulation," *J. Lightwave Technol.* **37**(10), 2335–2343 (2019).
19. X. Yangzhang, M. I. Yousefi, A. Alvarado, D. Lavery, and P. Bayvel, "Nonlinear frequency-division multiplexing in the focusing regime," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, (2017), pp. 1–3.
20. M. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing," *IEEE Trans. Inf. Theory* **66**(1), 478–495 (2020).
21. S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, "Capacity estimates for optical transmission based on the nonlinear fourier transform," *Nat. Commun.* **7**(1), 12710 (2016).
22. S. Civelli, E. Forestieri, and M. Secondini, "Why noise and dispersion may seriously hamper nonlinear frequency-division multiplexing," *IEEE Photonics Technol. Lett.* **29**(16), 1332–1335 (2017).
23. I. Tavakkolnia and M. Safari, "Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers," *J. Lightwave Technol.* **35**(11), 2086–2097 (2017).
24. M. Pankratova, A. Vasylichenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, "Signal-noise interaction in optical-fiber communication systems employing nonlinear frequency-division multiplexing," *Phys. Rev. Appl.* **13**(5), 054021 (2020).
25. N. A. Shevchenko, J. E. Prilepsky, S. A. Derevyanko, A. Alvarado, P. Bayvel, and S. K. Turitsyn, "A lower bound on the per soliton capacity of the nonlinear optical fibre channel," in *2015 IEEE Information Theory Workshop-Fall (ITW)*, (IEEE, 2015), pp. 104–108.
26. N. A. Shevchenko, S. A. Derevyanko, J. E. Prilepsky, A. Alvarado, P. Bayvel, and S. K. Turitsyn, "Capacity lower bounds of the noncentral chi-channel with applications to soliton amplitude modulation," *IEEE Trans. on Commun.* **66**(7), 2978–2993 (2018).
27. J. E. Prilepsky, S. A. Derevyanko, and S. K. Turitsyn, "Nonlinear fourier transform-based optical transmission: Methods for capacity estimation," in *Optical Communication Systems: Limits and Possibilities*, A. Ellis and M. Sorokina, eds. (Jenny Stanford Publishing, 2019), chap. 7, pp. 243–272.
28. D. Arnold, H. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory* **52**(8), 3498–3508 (2006).
29. M. Secondini, E. Forestieri, and G. Prati, "Achievable information rates in nonlinear wdm fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightwave Technol.* **31**(23), 3839–3852 (2013).
30. A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable information rates for fiber optics: Applications and computations," *J. Lightwave Technol.* **36**(2), 424–439 (2018).
31. S. Wahls, "Generation of time-limited signals in the nonlinear fourier domain via *b*-modulation," in *2017 European Conference on Optical Communication (ECOC)*, (IEEE, 2017), pp. 1–3.
32. M. Ablowitz, B. Prinari, and A. Trubatch, *Discrete and Continuous Nonlinear Schrödinger systems* (Cambridge University Press, 2003).
33. A. Shabat and V. Zakharov, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Sov. Phys. JETP* **34**, 62–69 (1972).
34. T. Gui, G. Zhou, C. Lu, A. P. T. Lau, and S. Wahls, "Nonlinear frequency division multiplexing with *b*-modulation: Shifting the energy barrier," *Opt. Express* **26**(21), 27978–27990 (2018).
35. S. Derevyanko, D. Shepelsky, M. Pankratova, A. Vasylichenkova, N. Chichkov, and J. Prilepsky, "Analytical model of nonlinear noise in the *b*-modulated optical transmission systems," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, (2020), p. SF2L.5.
36. R. J. Essiambre, R. Tkach, and R. Ryf, "Fiber nonlinearity and capacity: Single mode and multimode fibers," in *Optical Fiber Telecommunications, Vol. VIB: Systems and Networks*, I. Kaminow, T. Li, and A. Willner, eds. (Academic Press, 2013).
37. S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis technique for optical links with lumped amplification," *Opt. Express* **23**(7), 8317–8328 (2015).
38. M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, "On the design of nft-based communication systems with lumped amplification," *J. Lightwave Technol.* **35**(24), 5464–5472 (2017).
39. A. Hasegawa and Y. Kodama, *Solitons in optical communications*, 7 (Oxford University Press, USA, 1995).

40. L. Mollenauer and J. Gordon, *Solitons in Optical Fibers: Fundamentals and Applications* (Academic Press, 2006).
41. S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers," *Opt. Express* **22**(22), 26720–26741 (2014).
42. M. Ablowitz, D. Kaup, A. Newell, and H. Segur, "The inverse scattering transform-fourier analysis for nonlinear problems," *Stud. Appl. Math.* **53**(4), 249–315 (1974).
43. L. D. Faddeev and L. A. Takhtajan, *Hamiltonian methods in the theory of solitons*, vol. 36 (Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1987).
44. S. Wahls, "Generation of time-limited signals in the nonlinear fourier domain via b-modulation," in *2017 European Conference on Optical Communication (ECOC)*, (2017), pp. 1–3.
45. S. Chimmalgi and S. Wahls, "Bounds on the transmit power of b-modulated nfdm systems in anomalous dispersion fiber," *Entropy* **22**(6), 639 (2020).
46. I. Tavakkolnia and M. Safari, "Signaling on the continuous spectrum of nonlinear optical fiber," *Opt. Express* **25**(16), 18685–18702 (2017).
47. I. Tavakkolnia and M. Safari, "Dispersion pre-compensation for nft-based optical fiber communication systems," in *2016 Conference on Lasers and Electro-Optics (CLEO)*, (IEEE, 2016), pp. 1–2.
48. S. Wahls, S. Chimmalgi, and P. J. Prins, "Fnft: A software library for computing nonlinear fourier transforms," *J. Open Source Software* **3**(23), 597–598 (2018).
49. D. Kaup, "A perturbation expansion for the zakharov–shabat inverse scattering transform," *SIAM J. Appl. Math.* **31**(1), 121–133 (1976).
50. D. J. Kaup and A. C. Newell, "Solitons as particles, oscillators, and in slowly changing media: a singular perturbation theory," *Proc. R. Soc. Lond. A* **361**(1707), 413–446 (1978).
51. V. E. Zakharov and S. V. Manakov, "Asymptotic behavior of non-linear wave systems integrated by the inverse scattering method," *Sov. Phys. JETP* **44**, 358–364 (1976).
52. P. A. Deift, A. R. Its, and X. Zhou, *In Important developments in Soliton Theory* (Springer, New York, 1993).
53. M. Borghese, R. Jenkins, and K. D.-R. McLaughlin, "Long time asymptotic behavior of the focusing nonlinear schrödinger equation," *Ann. I. H. Poincaré* **35**(4), 887–920 (2018).
54. M. J. Ablowitz and H. Segur, *Solitons and the inverse scattering transform*, vol. 4 (Siam, 1981).
55. S. Wahls, "Second order statistics of the scattering vector defining the d-t nonlinear fourier transform," in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, (VDE, 2017), pp. 1–6.
56. E. Agrell and M. Secondini, "Information-theoretic tools for optical communications engineers," in *2018 IEEE Photonics Conference (IPC)*, (2018), pp. 1–5.
57. R. A. Shafik, M. S. Rahman, and A. R. Islam, "On the extended relationships among evm, ber and snr as performance metrics," in *2006 International Conference on Electrical and Computer Engineering*, (2006), pp. 408–411.
58. X. Zhang, P. Sun, L. Xi, Z. Zheng, S. Du, J. Wei, Y. Wu, and X. Zhang, "Nonlinear-frequency-packing nonlinear frequency division multiplexing transmission," *Opt. Express* **28**(10), 15360–15375 (2020).
59. A. Vasylichenkova, J. E. Prilepsky, N. B. Chichkov, and S. K. Turitsyn, "Multieigenvalue communication paired with b-modulation," in *45th European Conference on Optical Communication (ECOC 2019)*, (IET, 2019), pp. 1–4.
60. S. Civelli, E. Forestieri, and M. Secondini, "Mitigating the impact of noise on nonlinear frequency division multiplexing," *Appl. Sci.* **10**(24), 9099 (2020).
61. P. Deift and X. Zhou, "A steepest descent method for oscillatory riemann-hilbert problems. asymptotics for the mkdv equation," *Ann. of Math.* **137**(2), 295–368 (1993).