ARTICLE

Capacity estimates for optical transmission based on the nonlinear Fourier transform

Stanislav A. Derevyanko[1], Jaroslaw E. Prilepsky[2] & Sergei K. Turitsyn[2]

What is the maximum rate at which information can be transmitted error-free in fibre–optic communication systems? For linear channels, this was established in classic works of Nyquist and Shannon. However, despite the immense practical importance of fibre–optic communications providing for >99% of global data traffic, the channel capacity of optical links remains unknown due to the complexity introduced by fibre nonlinearity. Recently, there has been a flurry of studies examining an expected cap that nonlinearity puts on the information-carrying capacity of fibre–optic systems. Mastering the nonlinear channels requires paradigm shift from current modulation, coding and transmission techniques originally developed for linear communication systems. Here we demonstrate that using the integrability of the master model and the nonlinear Fourier transform, the lower bound on the capacity per symbol can be estimated as 10.7 bits per symbol with 500 GHz bandwidth over 2,000 km.

[1] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel. [2] Aston Institute of Photonic Technologies, Aston University, Birmingham B4 7ET, UK. Correspondence and requests for materials should be addressed to S.A.D. (email: stasd@bgu.ac.il) or to J.E.P. (email: y.prylepskiy1@aston.ac.uk) or to S.K.T. (email: s.k.turitsyn@aston.ac.uk).

I t is hard to overestimate the impact that optical fibre transmission systems have had on everyday life in the 'information society' era. Although these systems have undergone a long process of increasing engineering complexity and sophistication[1], the key physical effects that affect system performance remain much the same as before[1–8]. These are: chromatic dispersion, fibre Kerr nonlinearity and optical noise. Most of the current optical networks exploit methodologies that were originally developed for linear channels. Thus, it is not surprising that nonlinearity has a detrimental impact on such systems[3–8], since the only role that it can play within the 'linear communications' is to serve as a source of signal distortion; examples of the beneficial impact of nonlinearity are relatively scarce[9–11]. It has been predicted that, within the next decade, the existing optical fibre technology will approach the 'nonlinear transmission limit' (an infamous capacity crunch problem[8]), which caps the achievable rate of error-free data transmission[3–6,8] (with the first capacity limit estimates taking into account both noise and nonlinearity attributed to the work of Splett et al.[12]). Thus, to 'unlock' the capacity of nonlinear channels, it is necessary to shift the relevant information and communications technology paradigm by introducing truly nonlinear transmission and signal processing techniques. In this work, we adapt techniques developed in nonlinear science to optical communications and use these principally new tools to determine an estimate for the lower bound on nonlinear channel capacity.

The ubiquitous master model governing signal propagation in fibre–optic links is the nonlinear Schrödinger equation (NLSE)[1,2,5]. The NLSE belongs to the unique class of integrable equations that can be solved via the inverse scattering transform[13]. The latter is an extension of the Fourier transform onto nonlinear systems and is often called the nonlinear Fourier transform (NFT)[14,15]. This term indicates that the basic principle of how NFT works is the same as in the linear case: similar to reducing the effect of chromatic dispersion in a linear propagation to a phase rotation in frequency space through the Fourier transform, the NFT transforms the effects of both nonlinearity and dispersion into a trivial linear evolution of the nonlinear spectral data. Therefore, it stands to reason that truly nonlinear techniques of the chromatic dispersion and fibre nonlinearity compensation should rely on NFT-based algorithms in place of linear counterparts. In 1993, Hasegawa and Nyu proposed using discrete eigenvalues (corresponding to solitons) emerging from the NFT to encode and transmit information, as these are not affected by dispersion and nonlinearity[10,16]. They termed this approach 'eigenvalue communications'. Later, Yousefi and Kschischang[17] used NFT for nonlinear signal multiplexing in multi-user channels. The objective of their approach was to solve the problem of nonlinear crosstalk that occurs in wavelength–division–multiplexed systems. Both ideas have received various generalizations and extensions, and some first experimental implementations have already been reported (see below). In this paper, we refer to both approaches by the umbrella term of NFT.

The existing optical transmission methods employing NFT can be categorized into two general groups. The first one[18,19] employs NFT as an efficient tool for solving NLSE backwards, in a manner similar to digital back propagation[20]. The second approach implies the use of nonlinear modes themselves for the data encoding and transmission[17,21–26]. The first consideration of the multiplexing in the nonlinear Fourier domain was presented in ref. 17. We note the recent experiments of Osaka group[27,28], Bülow et al.[29,30] and Dong et al.[26], demonstrating the feasibility of the NFT-based optical transmission. Furthermore, the current NFT-based approaches can be classified according to what part of

the nonlinear spectrum is used for modulation. The authors of (refs 26–28,31,32) exploited discrete spectra. The novel concept of using the continuous nonlinear spectrum as information carrier was put forward in refs 17,21–23,33. In particular, a method of nonlinear inverse synthesis (NIS) was proposed in refs 21–23: its purpose is to generate the time domain waveforms starting from a continuous nonlinear spectrum that exactly matches the linear spectrum of the data to be transmitted.

In the following, we address the fundamental question as to whether the achievable information capacity of fibre channels can be enhanced using NFT. In this work, we show that the use of NFT/NIS methods makes it possible to favourably estimate the lower bound of the capacity per symbol for the long-haul fibre networks in the multichannel/multicarrier environment, compared with the conventional modulation techniques. We demonstrate that in a wide range of input power levels, the well-established results from the NLSE perturbation theory[34,35] can be used to formulate an asymptotic channel model in the NFT domain. Using very conservative estimates for the lower bound of capacity[3,36], we derive the estimates for the lower bound for the capacity per symbol of NIS-based transmission (within an approximate model), predicting the lower-bound values of ~ 11 (bits/symbol) for 5×100 GHz wavelength-division-multiplexing (WDM) Nyquist and orthogonal frequency division multiplexing (OFDM) transmission at 2,000 km. This bound improves logarithmically with the channel bandwidth or subcarrier spacing, see equation (12). Our results also reveal an improvement over the achievable information rates reported recently[37,38], although our goal here is rather to show to the wider community the potential benefits of using the NFT. We also demonstrate that even in the presence of the small inline noise the channel remains free from the nonlinear crosstalk that is thought to be one of the main sources of the spectral efficiency degradation[3–5,8]. Since the capacity estimates used to derive these bounds are known to be loose for nonlinear and non-Gaussian information channels[39], the actual value of the achievable capacity is anticipated to be higher.

Results

Model description and basics of NFT and NIS method. The common channel model for optical communications inside a single-mode fibre is the NLSE written for the electrical field envelope $q(z,t)$, perturbed by additive white Gaussian noise (AWGN)[1,2,4,5]. We will mostly work in standard dimensionless units (Supplementary Note 1), and consider the most practically useful case of anomalous dispersion:

$$\frac{\partial q}{\partial z} - \frac{i}{2}\frac{\partial^2 q}{\partial t^2} - i|q|^2 q = \eta(z,t), \qquad (1)$$

with z being a normalized distance along the fibre, t is time in frame co-moving with the envelope and the circularly-symmetric AWGN term η (having zero mean) is completely characterized by the spectral power density of noise D defined via the auto-correlation function: $\mathbb{E}[\eta(z,t)\bar{\eta}(z',t')] = 2D\delta(t-t')\delta(z-z')$, where the overbar means complex conjugate and $\mathbb{E}(\ldots)$ is the expectation value. Such a form of the optical channel corresponds to the amplification scheme, in which the distributed Raman gain exactly compensates for the intrinsic fibre loss[4,5]. Traditional (linear) modulation techniques work in time or linear frequency domain, where the evaluation of the maximum achievable error-free transmission rate of channel (1) in symbols per second—that is, the Shannon capacity[40,41]—is quite a nontrivial and challenging task[42]. We address the same problem in our work, but specifically for the NFT-based transmission.

The details of the NFT for the NLSE can be found in a great number of works on the subject[10,13–15,17]. Performing the

direct NFT on a pulse $q(t)$ amounts to solving the so-called Zakharov–Shabat problem, written for two auxiliary functions $v_{1,2}(t)$:

$$\frac{dv_1}{dt} = q(t)v_2 - i\zeta v_1, \quad \frac{dv_2}{dt} = -\bar{q}(t)v_1 + i\zeta v_2, \quad (2)$$

where the input pulse shape $q(t)$ acts as a potential. Here ζ is a (generally complex) eigenvalue, $\zeta = \xi + i\rho$, and $q(t)$ decays as $t \to \pm\infty$. To define scattering data (the analogue of Fourier spectrum), for real $\zeta = \xi$, one selects a specific solution of equation (2), $\Phi(t,\xi) = [\phi_1, \phi_2]^T$, by the 'initial condition' at the trailing end of the pulse: $\Phi|_{t\to-\infty} = [e^{-i\xi t}, 0]^T$. Then, the solution at the leading end must necessarily take the form $\Phi_{t\to+\infty} = [a(\xi)e^{-i\xi t}, b(\xi)e^{i\xi t}]^T$, where the functions $a(\xi)$ and $b(\xi)$ are called scattering coefficients. The continuous part of the nonlinear spectrum is defined by the ratio named a reflection coefficient: $r(\xi) = b(\xi)/a(\xi)$, and the discrete complex eigenvalues, ζ_n, are the zeros of the coefficient $a(\xi)$ analytically extended into the upper half plane of ζ. The forward NFT operation corresponds to mapping of the initial field, $q(0,t)$, onto a set of scattering data: $\Lambda = [\{r(\xi), \xi \in \mathbb{R}\}; \{\varsigma_n, \gamma_n = b(\varsigma_n)/a'_\varsigma(\varsigma_n)\}]$, where the index n runs over all discrete eigenvalues of Zakharov–Shabat problem.

Figure 1 depicts the simplified flowchart of operations for the NIS NFT-based transmission scheme, see also[21–24], and ref. 25 for the experimental set-up scheme. Within the NIS, the parameters of nonlinear modes serve as elementary information carriers, and at the detector one retrieves the data encoded directly from the nonlinear spectrum using the NFT operation. The main advantages of the NIS again the other NFT-based counterparts are as follows. First, insofar as the continuous nonlinear spectrum of our signal matches the linear spectrum of data to be transmitted, the 'learning curve' for system designers is not very steep, as one can avoid dealing with 'non-traditional' encoding schemes. Second, the transmission looks very similar to that through a linear dispersive channel. Third, for the continuous spectrum, one can immediately take advantage of the existing efficient modulation formats and adapt those directly for nonlinear spectral communications. In addition, this scheme has been shown to provide higher noise tolerance and the potential for lower numerical complexity than in the case of digital back propagation, outperforming linear compensation in terms of transmission quality[21–23]. Thereby, we will use the NIS as our scheme of choice when providing the capacity estimates of the nonlinear fibre channel, though our approach can be generalized to various other NFT systems.

In our study, we will employ only the continuous part of the nonlinear spectrum, that is, our data are encoded on and retrieved from the quantity $r(\xi)$. The evolution of $r(\xi)$ in the noise-free NLSE channel is trivial: $r(\xi,z) = r(\xi,0)e^{2i\xi^2 z}$, so that
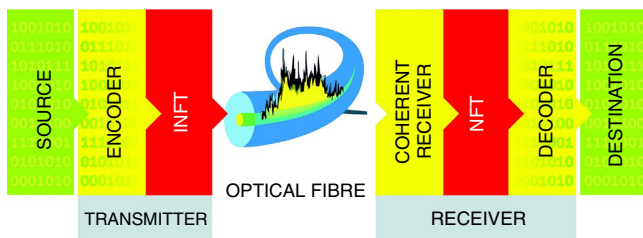

Figure 1 | Simplified flowchart of the tranceiver scheme. Shown are: the NFT-based synthesis operation at the transmitter and NFT-based demodulation at the receiver (both transmitter and receiver can also include an FT operation depending on the explicit processing algorithm). For a more detailed block-diagram and a thorough description of the NIS-based optical communication system, see (refs 22,23).

the orthogonality of nonlinear normal modes is preserved during the evolution. The inverse NFT (INFT) maps the encoded scattering data Λ at the transmitter onto the field $q(t)$; see Fig. 1. This is achieved via the solution of Gelfand–Levitan–Marchenko equations[13,14,21,22]. Then, after the propagation over a fibre, at the receiver one reads the waveform $q(t,L)$ and retrieves the nonlinear spectrum $r(\xi;L)$ by solving the Zakharov–Shabat problem (2), that is, by the forward NFT. Unwinding the accumulated phase rotation inside the nonlinear domain, we finally recover the initial data, and this completes the NIS scheme (Fig. 1). Further details about the basics of the NIS can be found in refs 21–23 as well as in Supplementary Note 4.

NFT data evolution in the presence of AWGN. The first goal of our study is to formulate the stochastic model for the data evolution inside the NFT domain. When the noise is small compared with signal power (the exact conditions are given in the Methods section), one can apply the inverse scattering transform perturbation theory[34,35], which yields a self-consistent stochastic channel description inside the NFT domain. Namely, the dynamics of the continuous nonlinear spectrum is given by a linear equation with additive noise:

$$\frac{\partial r(\xi,z)}{\partial z} = 2i\xi^2 r(\xi,z) + \Gamma(z,\xi). \quad (3)$$

The nonlinear spectral noise $\Gamma(z,\xi)$ is still a zero mean complex Gaussian process, but it possesses several properties distinguishing it from its space-time domain progenitor $\eta(z,t)$. It is fully characterized by two complex autocorrelation functions: $\mathbb{E}[\Gamma(z,\xi)\bar{\Gamma}(z',\xi')] = 2D A(z;\xi,\xi')\delta(z-z')$ and $\mathbb{E}[\Gamma(z,\xi)\Gamma(z',\xi')] = 2D B(z;\xi,\xi')\delta(z-z')$. The explicit form of the functions A and B is determined by the projection of η on the nonlinear normal modes and is given in the Supplementary Note 2.

Thus, the signal evolution inside the nonlinear spectral domain amounts to the dispersive phase rotation affected by noise. One can note the similarities between the linear Fourier channel and its nonlinear counterpart (equation (3)): to see this, we can drop nonlinearity in equation (1) and rewrite it in the linear frequency domain as: $\partial q(\omega,z)/\partial z = -(i\omega^2/2) q(\omega,z) + \eta_\omega(z)$, where the noise, as in equation (1), has zero mean and the only nonzero autocorrelation function $\mathbb{E}[\eta_\omega(z)\bar{\eta}_{\omega'}(z')] = 4\pi D \delta(z-z')\delta(\omega-\omega')$. This similarity becomes even more striking if we recall that in the limit of low power the following relation between the FT and NFT spectra holds[14,33]: $r(\xi,z)|_{|q|\to 0} \to -\bar{q}(\omega,z)|_{\omega=-2\xi}$. Applying this transformation to equation (3), we indeed recover circular AWGN with the linear power spectral density (PSD) $2D$. However, the seeming simplicity of the evolution inside the NFT channel is deceptive. First, the new noise Γ is no longer circular, in contrast to its linear counterpart $\eta_\omega(z)$. Next, this noise is neither homogeneous nor uncorrelated, as A and B are generally functions of both 'frequencies' ξ and ξ'. The most important distinctive property of Γ, however, is that it depends on the initial spectrum, $r(\xi,0)$. From the information theory perspective, the latter means that equation (3) defines an input-dependent Gaussian channel with memory[41].

The continuous NFT channel model. For the information-theoretic analysis, the channel model given by a stochastic equation (3) must be reformulated as an input–output probabilistic model, that is, the conditional probability density function (PDF) of the channel output given the channel input. We define the continuous channel output Y_ξ as the solution of equation (3) at the receiver located at distance $z = L$ with the

compensated phase rotation and filtering:

$$Y_\xi \equiv H(\xi)e^{-2i\xi^2 L}r(\xi, L) = X_\xi + N(\xi, X_\xi), \qquad (4)$$

where $X_\xi \equiv r(\xi, 0)$, $H(\xi)$ is the rectangular bandpass filtering function in the nonlinear frequency domain (applied at the receiver), that selects only a given channel of interest (COI). The effective filtered noise $N(\xi, X_\xi)$ (with zero mean) has the following correlation properties:

$$\mathbb{E}[N(\xi)\bar{N}(\xi')] = 2D \int_0^L e^{-2i(\xi^2 - \xi'^2)z} A(z; \xi, \xi') dz,$$
$$\mathbb{E}[N(\xi)N(\xi')] = 2D \int_0^L e^{-2i(\xi^2 + \xi'^2)z} B(z; \xi, \xi') dz. \qquad (5)$$

Naturally, due to the filtering the above relations hold within the COI only. We do not include any add/drop elements and optical–electrical conversion in our considerations here. But the possibility of including such elements and the lack of side information regarding them from the point of view of COI (which is a commonplace situation) makes the interference from other channels being effectively random, and in all our further calculations, we only consider a single (central) COI, and reckon the encoded information in other (than COI) channels as an additional contribution to the noise PSD (Supplementary Note 6). To evaluate autocorrelation functions (5), one needs to know the full z-evolution of the unperturbed Jost functions $\Phi(z; t, \xi)$, and this problem does not have a closed form solution in the general case. However, in the regime of a long fibre system one can either use the large z asymptotic solutions of Zakharov–Shabat problem (2)[43,44], or the assumption of a finite temporal extent of the pulse, which is always the case for the NIS in a burst mode[19,23] (Supplementary Note 3). Then, assuming large L, one obtains a remarkably simple result that explicitly depends only on the initial spectral data:

$$\mathbb{E}[N(\xi)\bar{N}(\xi')] \approx \mathcal{N}\pi\delta(\xi - \xi')E_1(\xi),$$
$$\mathbb{E}[N(\xi)N(\xi')] \approx \mathcal{N}\pi\delta(\xi - \xi')E_2(\xi), \qquad (6)$$

where $E_1(\xi) \equiv 1 + |X_\xi|^2 + |X_\xi|^4$ is an effective PSD (normalized to its linear value), $E_2(\xi) \equiv X_\xi^2$, $\mathcal{N} \equiv 2DL$. The latter quantity is the accumulated noise variance per sample in the time domain, and we have omitted the non-diagonal terms of order unity as small compared with those $\sim L$ (Supplementary Note 3).

Two important observations regarding the properties of the noise $N(\xi)$ can be made. First, within the nonlinear bandwidth of the COI, the noise PSD in the NFT domain, $E_1(\xi)$, grows nonlinearly with the spectral power of the input. Second, the channel model (4), (6) is local in the nonlinear frequency ξ. So, for example, in the case of dense WDM, one can simply match the nonlinear bandwidth of the filter with that of the COI and prevent both direct and noise-induced channel crosstalks without losing any of the informational content of our message, since the signal-dependent nonlinear spectral broadening is virtually absent. It is this remarkable property of the nonlinear spectrum (which holds as long as the effective signal-to-noise ratio (SNR$_{eff}$) defined by equation (11) below is large and the propagation distance is not too small) that makes the NIS-based transmission potentially free from the crosstalk and bandwidth-related sources of the capacity degradation that plague most of the conventional transmission systems[3–5]. In the Methods section, we elaborate this statement further considering practical time-sampled multi-channels and in the Supplementary Note 6, we verify the PSD results above by a direct numerical simulation.

Sampling and the discrete input–output model. So far we have defined our channel (4)–(6) using continuous field representation. The advantage of such an approach is that it allowed us to consider the multitude of the conventional schemes within the same theoretical framework. However, in digital communications, the signal is modulated and sampled in the time domain, and for each time sample, the information is encoded via complex amplitude level sets corresponding to discrete or continuous constellations[5]. Therefore, to make our results pertinent to the recently proposed NFT communication systems[21–23], we shall consider two closely related standard frequency multiplexed schemes, namely, dense Nyquist WDM[5] and OFDM[45,46] both adapted to the NIS scheme.

We start from a general encoded sequence in time domain:

$$q_{in}(t) = \sum_{\alpha=0}^{N_b-1} \sum_{k=0}^{N_{ch}-1} c_{\alpha k} s(t - \alpha T_s)e^{i\Omega_k t}, \qquad (7)$$

where N_b is the length of the symbol sequence (that is, burst), N_{ch} is the number of WDM channels or alternatively OFDM subcarriers, $s(t)$ is the base wave-shape defining the particular format, T_s, is the symbol width, Ω_k is an individual channel/subcarrier frequency. Here unless otherwise specified, we use normalized units. It is the discrete set of coefficients $c_{\alpha k}$ that now bears our informational content, and real and imaginary parts of $c_{\alpha k}$ form the components of the $2M$-dimensional input X, with $M = N_b \times N_{ch}$. Within the NIS scheme, Fig. 1, we do not actually synthesize the waveform (7). Instead, we use its linear spectrum and use it as the nonlinear spectrum of a new optical signal $q(z = 0, t)$ to be launched into the fibre, utilising the mapping rule between the initial Fourier spectrum and the NFT reflection coefficient: $r_{in}(\xi, 0) = X_\xi = \bar{q}_{in}(\omega)|_{\omega = -2\xi}$. Note that the correlation properties of the nonlinear noise, (4)–(6), now explicitly depend on the amplitudes of the input sequence (7). The actual optical signal in the time domain is generated by applying the INFT. The resulting waveform is then fed into the optical fibre model (1). At large values of SNR$_{eff}$, as defined by equation (11), the nonlinear spectrum evolves according to equation (3), and the input–output interrelation is then given by equations (4) and (5), assuming that at the receiver one uses NFT (2) to obtain the nonlinear spectrum, compensates for the propagation-accumulated phase and then retrieves the modulation coefficients of each symbol, $c_{\alpha k}$, using standard linear demodulation schemes[22,23] (Supplementary Note 4). For the WDM case, the received nonlinear spectrum, $r(\xi, L)$, is bandpass filtered for a given COI, see equation (4), while for the OFDM, the filtering is assumed over the total signal bandwidth. Note that the quantity $W_0 = T_s^{-1}$ serves as the single channel bandwidth for the Nyquist WDM and carrier spacing for the OFDM case.

Since the channel in the nonlinear frequency domain (4)–(5) is characterized by additive Gaussian input-dependent noise (that is, the channel law inside the NFT domain is Gaussian with the input-dependent covariance), the discrete channel in the NIS scheme has the same property:

$$c_{\alpha k}(L) = c_{\alpha k}(0) + N_{\alpha k}, \qquad (8)$$

where $N_{\alpha k}$ is the projection of the spectral noise $N(\xi, X_\xi)$ onto the corresponding subcarrier in the OFDM case and the Nyquist-sampled noise vectors for the COI in the WDM case. Introducing $2M$ real and imaginary parts of $c_{\alpha k}$ as discrete real-valued input and output, **X** and **Y** correspondingly, one gets for the input–output conditional PDF the multivariate Gaussian distribution with the $2M \times 2M$ quadrature correlation matrix $\hat{\Sigma}$ whose elements are obtained from the correlation functions (5).

Since the intensities A and B from (5) depend on $\mathbf{X}$ (that is, on $r(0,\xi)$), so does the correlation matrix: $\hat{\Sigma} = \hat{\Sigma}(\mathbf{X})$. Using input (7) and asymptotic expressions (6), one obtains for the components of $\hat{\Sigma}(N_{ak} = N_{ak}^{R} + iN_{ak}^{I})$:

$$
\mathbb{E}\left[N_{\alpha k}^{R/I} N_{\alpha' k'}^{R/I}\right] = \frac{\mathcal{N}}{4\pi} \mathrm{Re}\left[\int d\omega \left\{ e^{i\omega(\alpha-\alpha')T_s} f_k(\omega)\bar{f}_{k'}(\omega)E_1\left(-\frac{\omega}{2}\right) \right.\right.
$$
$$
\left.\left. \pm e^{i\omega(\alpha+\alpha')T_s} f_k(\omega)f_{k'}(\omega)\bar{E}_2\left(-\frac{\omega}{2}\right) \right\}\right],
$$
$$
\mathbb{E}\left[N_{\alpha k}^{R} N_{\alpha' k'}^{I}\right] = \frac{\mathcal{N}}{4\pi} \mathrm{Im}\left[\int d\omega \left\{ - e^{i\omega(\alpha-\alpha')T_s} f_k(\omega)\bar{f}_{k'}(\omega)E_1\left(-\frac{\omega}{2}\right) \right.\right.
$$
$$
\left.\left. + e^{i\omega(\alpha+\alpha')T_s} f_k(\omega)f_{k'}(\omega)\bar{E}_2\left(-\frac{\omega}{2}\right) \right\}\right],
$$

(9)

where $E_{1,2}(\xi)$ is defined below equation (6). Coefficients $f_k(\omega)$ are the format-dependent form factors closely related to the linear Fourier transform of the pulse form $s(t)$ from (7), see Supplementary Note 6. In the WDM case, this form factor is cut off by filtering and is only nonzero when the frequency belongs to the COI of width $2\pi/T_s$. Since the channels do not overlap, the noise components from different channels are uncorrelated. This is the consequence of the already mentioned property of the asymptotic absence of the channel crosstalk in the continuous model (4), (6). For the OFDM, the integration is restricted to the total nonlinear bandwidth of $2\pi N_{ch}/T_s$.

Capacity per symbol estimates for WDM/OFDM NIS transmission. For an arbitrary vector, information channel the input–output mutual information $I(\mathbf{X},\mathbf{Y})$ is defined as[5,40,41]:

$$
I(\mathbf{X};\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}),
$$
$$
H(\mathbf{Y}) = -\int d^{2M}\mathbf{Y} P_Y(\mathbf{Y}) \log_2(P_Y(\mathbf{Y})),
$$

(10)

where H designates the entropy. The Shannon capacity per symbol, C, is the maximum of $I(\mathbf{X};\mathbf{Y})/M$ over the input distribution $P_X(\mathbf{X})$ subject to the average power per sample constraint $\mathbb{E}\left[\|\mathbf{X}\|^2\right]/M \leq S$. For any additive Gaussian channel, the expression for the channel entropy $H(\mathbf{Y}|\mathbf{X})$ is obtained by averaging the determinant of the conditional correlation matrix $\Sigma(\mathbf{X})$ over the input distribution. Our channel (8) possesses the non-diagonal input-dependent correlation matrix (9) that makes the direct optimization of the mutual information functional extremely difficult and only some lower bound for the channel capacity can be obtained. This is a common situation in case when the physical signal propagation is a nonlinear dynamical process. A standard approach for the lower-bound estimate is to use Gaussian input $\mathbf{X}_G$ with independently distributed real quadrature samples each having the variance $S/2$, which in the continuous limit corresponds to a Gaussian process with constant spectral density proportional to S (ref. 5). Another popular choice is the so-called ring constellation input,[4,5] where for each complex sample the amplitude is fixed while the phase is uniformly distributed. Here we shall use the Gaussian input unless otherwise specified.

Analytical expressions for the mutual information for the channel given by equations (8)–(9) are generally intractable even with the Gaussian independent identically distributed (i.i.d.) input. This is because of the forbiddingly complex dependence of the noise correlation matrix (9) on the input. Further standard step to achieve a tractable analytical result at the expense of the accuracy of the estimate is to use the effective Gaussian input–output model and the Pinsker's formula, keeping in mind that this bound may not be tight at all[39]. Despite that, this procedure is rather standard and its further details are outlined in the Methods section. There it is shown that in the limit of large effective SNR defined by the equation below, the Pinsker lower bound for the capacity, C_G, in bits per symbol is found to be (real-

world units are assumed)

$$
C_G = \log_2\left[\frac{ST_s/N_{ASE}}{2(E_{in}(S)/E_{NL})^2 + (E_{in}(S)/E_{NL}) + 1}\right],
$$
$$
= \log_2(\mathrm{SNR}_{eff}),
$$

(11)

where the second line is the definition of the effective SNR, $N_{ASE} = h\nu_0 K_T \chi L$ is the PSD of the accumulated ASE noise (see Supplementary Note 1 and refs 4,5 for the physical meaning of each parameter), $E_{in}(S) = ST_s N_b N_{ch}$ is the average energy of the effective initial optical burst (before the NIS module); if one wants to express equation (11) in terms of S instead of E_{in}, this can be done by means of this linear dependence given above. The quantity $E_{NL} = |\beta_2| N_{ch}/(\gamma T_s)$ represents a typical energy scale where the nonlinear effects become pronounced. This formula holds for both OFDM and Nyquist-based NIS transmission, and it is the main result of our paper. It is accurate up to terms of order $O[1/\mathrm{SNR}_{eff}]$; the general applicability criteria are discussed in the Methods section. Note that, for a fixed propagation distance L and symbol rate, it displays a characteristic peaky behaviour (that is, reaches a local maximum) in both average input power, S, and burst energy, E_{in}, which is common to many Gaussian-based lower-bound estimates for conventional transmission formats[3–5,37].

Implications for long-haul optical systems. We can now put the obtained results into perspective by considering model of the fibre–optical communication systems operating on long-haul distances. For a fixed distance, the number of channels/sub-carriers, burst duration and symbol rate the argument of log in (11) is a monotonically growing function of E_{in} (or S) up to $E_{in}^* = E_{NL}/\sqrt{2}$, which corresponds to the maximum of the estimated bound:

$$
C_G^{max} = \log_2\left(\frac{2\sqrt{2}-1}{7}\frac{E_{NL}}{N_{ASE}N_b N_{ch}}\right)
$$
$$
= \log_2\left(\frac{|\beta_2|}{K_T h\nu_0 \chi \gamma T_b L}\right) - 1.94 \text{ bits/symbol}
$$

(12)

where $T_b = T_s N_b$ is the duration of the burst before the NIS module, and the signal bandwidth is $W_0 = T_s^{-1}$. From equation (12), it is seen that the estimate deteriorates slowly (logarithmically) with the product $L \times T_b$. On the other hand, it does not depend on the number of channels, N_{ch}, which is a direct consequence of the absence of the channel crosstalk.

Let us now address the physical meaning of equation (12), in particular, explain why a shorter burst duration brings about a higher capacity than a longer one. One can note that the denominator of equation (11) (which is actually the effective noise power in the NFT domain) grows with the ratio of the burst energy and the nonlinear energy. In other words, unlike the linear situation where the size of the burst does not affect the spectral properties of the noise, here this density grows both with the burst size and the energy of the pulse. The first circumstance is due to the fact that the noise in the nonlinear spectrum depends on the signal in time domain in the nonlocal manner: it depends on the integral characteristics of the time domain pulse rather than local ones (pretty much as the usual linear Fourier transform has the spectrum that depends on the whole time domain evolution of the pulse). The longer the burst duration T_b, the more nonlinear noise it accumulates as it is dragged along—hence the capacity decreases. Now let us turn to the dependence of C_G^{max} on the fibre parameters. By decreasing β_2 and increasing the nonlinearity parameter γ, we effectively increase the dispersion length L_D, while decreasing the nonlinear length $L_{NL} = 1/(\gamma P)$. In other words, one makes the system more nonlinear and lowers

the energy threshold E_{NL} where the nonlinear effects are important. This, in turn, makes the aforementioned noise accumulation in the NFT domain more pronounced, such that its nonlinear spectral density grows, which has the adverse effect on SNR. This accumulation effect is clearly seen in the direct numerical simulation given in Supplementary Fig. 1 and Supplementary Note 6.

Let us provide two typical examples of Nyquist and OFDM system parameters close to those reported both in conventional-[4,5] and NIS-based[21–23] systems. The goal here is not to compete with the record state-of-the-art experiments, but rather to give an idea of the required power levels and achievable rates. For the Nyquist transmission, we pick five channels with individual bandwidth $W_0 = T_s^{-1} = 100$ GHz, corresponding to the overall bandwidth of 500 GHz and for the OFDM we pick 100 subcarriers with the spacing of $W_0 = 5$ GHz and the same effective total bandwidth. The optimal initial energy is then $E_{in}^* = 6$ pJ for both Nyquist and OFDM cases. The required power levels in the optical domain can be estimated by specifying the burst size, N_b (Supplementary Note 7). For the considered parameters, the lower bound on the capacity per symbol can be estimated as ∼10.7 bits per symbol over 500 GHz bandwidth at 2,000 km. Figure 2 plots the estimate (12) as a function of distance for different burst sizes. The result deteriorates with the burst size for both Nyquist and OFDM transmission as predicted by (12). Note however that for a fixed symbol rate the penalty of going from, say, 1,000 to 2,000 km is only 1 bit per symbol, so ∼1,000 km mark at least 12 bit per symbol can be achieved and so on. For the Nyquist case, varying the burst size is equivalent to changing the number of symbols in the burst while keeping the bandwidth fixed, whereas for the OFDM, one needs to change the number of subcarriers to keep the bandwidth fixed.

Discussion

We have developed a theoretical approach based on the perturbation theory for the NFT data for the estimate of the lower bound of the capacity per symbol for the NFT-based optical transmission, which becomes asymptotically exact in the limit of large effective SNR. Considering transmission over 500 GHz bandwidth, the lower bound on the capacity per symbol is estimated as ∼10.7 bit per symbol at 2,000 km. The accurate estimates of the spectral efficiency corresponding to the capacity per symbol require massive system optimization in terms of achievable symbol rates. The NFT technique is still in the emerging early stages and the accurate estimates of the spectral efficiency corresponding to the capacity per symbol found in this paper would requir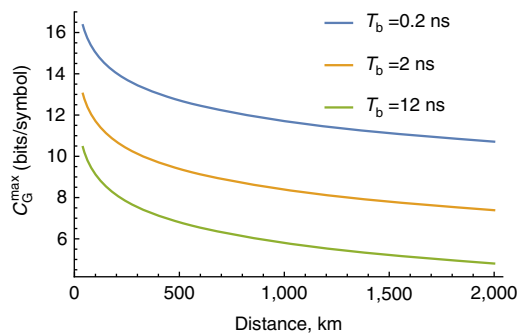e the massive system optimization in terms of achievable symbol rates. In particular, one will need to obtain an explicit dependence of the linear frequency and time domain dependence of the pulse width and bandwidth on the system parameters. This seems to us to be a difficult task to achieve analytically and will require the full numerical optimization of various NFT systems that is well beyond the scope of a single paper. But some preliminary results and considerations can already be found in the Methods section and Supplementary Note 7. However, we would like to stress that the estimates of capacity per symbol made in our work show great promise of the NFT technique and give an important guidance for the development of future systems. Moreover, some of the results presented here have high self-sufficient value. For example, equation (3) describes a continuous channel model for a generic NFT-based system in the presence of inline noise, while equations (4), (5) and (9) develop it further, introducing a discrete time channel model for the NFT-based transmission dealing with the continuous spectrum. The model predicts the absence of the nonlinear spectral broadening and channel crosstalk that makes it applicable to multi-user routed optical transmission systems. The developed channel models can be applied for the transmission system design, optimization and digital signal processing. Note also that in the Methods section, we also present the pioneering results for the simulations of the nonlinear spectral domain WDM transmission, addressing the issue of the crosstalk between the channels and revealing the absence of the latter in the considered NFT systems. Finally, we believe that the capacity estimate (11) is too conservative and can be improved. Indeed, our channel model with memory (8) is very close to a recently studied simpler model[47], where it was shown that by a proper coding one can achieve a non-decreasing lower bound for capacity. In fact this result can be proven rigorously for any static, memoryless, power constrained communication channel[48].

When the paper was under the review, a very recent ArXiv publication[49] came to our attention. It reports simulation results for the normal dispersion NFT channel. Interestingly, the capacity rates shown there are close to those predicted in our paper for a slightly more relevant to the long-haul transmission case of the anomalous dispersion. This further supports our belief that the NFT-based methods are important tool for overcoming the capacity crunch.

Methods

A lower bound of the capacity of the nonlinear channel. To derive a lower bound using the mutual information (10) with the $2M$-dimensional Gaussian input $\mathbf{X}_G$, here we largely follow a standard information theory approach, see, for example, ref. 3. Namely, we replace the channel output $\mathbf{Y}$ with another Gaussian, $\mathbf{Y}_G$ such that the joint Gaussian input–output PDF $P_G(\mathbf{X}_G, \mathbf{Y}_G)$ has the same binary correlation function as the original distribution $P(\mathbf{X}_G, \mathbf{Y})$. This effective Gaussian channel provides yet another lower bound for the capacity and has one important advantage that its capacity, C_G, can be calculated directly via the so-called Pinsker formula[36,39]:

$$C \geq C_G = \frac{1}{2M} \log_2 \frac{\det \hat{\Sigma}_X \det \hat{\Sigma}_Y}{\det \hat{\Sigma}_{XY}}, \tag{13}$$

where $\hat{\Sigma}_{XY}$ is the full input–output correlation matrix, while $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ are input–input and output–output covariance matrices. The result is verified by the direct substitution of the multivariate Gaussian PDF $P(\mathbf{X}_G, \mathbf{Y}_G)$ into the mutual information functional (10).

The fact that the conditional probability of the output is Gaussian simplifies the calculation of the determinant. Moreover, in Supplementary Note 5, it is shown that when SNR_{eff}, defined in equation (11), is large, equation (13) simplifies to

$$C_G \approx \frac{1}{2M} \log \det [S/(2\hat{N}(S))], \tag{14}$$

where the effective noise matrix $\hat{N}(S)$ is the $2M \times 2M$ correlation matrix (9) averaged over the i.i.d. Gaussian input with the variance $S/2$. It is the characteristic value of the noise matrix that defines the effective SNR in the problem and controls the validity of the model. To calculate it explicitly, one has to specify a particular

Figure 2 | Lower bound for the maximum achievable capacity. Data obtained via equation (12) through the analysis of the approximate model (8), (9), versus transmission distance over 500 GHz bandwidth at different burst sizes.

transmission format. This is done in some detail in Supplementary Note 6, for the OFDM and Nyquist modulation of the input waveform (7). The result reads

$$\hat{N}(S) = \frac{\mathcal{N}}{T_s}[(SN_b T_s^2)^2 + SN_b T_s^2/2 + 1/2]\hat{\mathbf{I}}.$$

Plugging this result for the average noise matrix $\hat{N}(S)$ into the asymptotic estimate (14) and going back to the dimensional variables as discussed in the Supplementary Note 1, one obtains equation (11).

The applicability of the obtained results. In the limit of small power and short burst, when $E_{in} \ll E_{NL}$, the definition of the SNR_{eff} coincides with the linear one. However, in the nonlinear regime, the effective SNR deteriorates if one considers either a high-power regime or a very long burst. The overall consistency criteria for combining perturbation theory and asymptotic analysis can be written as $SNR_{eff} \gg 1$, so that the validity condition for equation (3) is met, and the propagation distance L must be much greater than the dispersion length (defined via the total transmission bandwidth $W = N_{ch}W_0$), to assume the diagonal form of the correlation functions (6) that was further used to get equation (9). For the fixed fibre and burst parameters considered in the text, and the given symbol length T_s, assuming the optimal energy $E_{in}^* = E_{NL}/\sqrt{2}$ corresponding to the optimal variance level $S_* = |\beta_2|/(\sqrt{2}\gamma T_s^2 N_b)$, all the above conditions turn into a restricted window of distances in the real-world units:

$$\frac{1}{W^2 |\beta_2|} \ll L \ll \frac{|\beta_2|}{K_T h\nu_s \chi\gamma T_b}. \qquad (15)$$

For $W_0 = 100$ GHz and 5 WDM channels with the total bandwidth $W = 5 W_0 = 0.5$ THz, and the burst size $T_b = 12$ ns in a standard telecom fibre, the above reads as $0.2 \ll L(\text{km}) \ll 2.1 \times 10^5$, and this condition is easily met in all realistic implementations. Now let us study what are the theoretical restrictions on the input parameter S. In the nonlinear regime, when $E_{in} \gtrsim E_{NL}$, the quadratic term in the denominator of SNR_{eff} in equation (11) dominates and the condition $SNR_{eff} \gg 1$ is equivalent to the following restriction:

$$S \ll \frac{|\beta_2|^2}{\gamma^2 T_s T_b^2 N_{ASE}}.$$

Thus, for the example considered above, that is, for the Nyquist modulation with five channels each having the bandwidth $W_0 = T_s^{-1} = 100$ GHz, $T_b = 12$ ns and $L = 2,000$ km considered in the text, the perturbative approach is valid up to the optical power levels $S_{opt} \sim 11$ dBm.

At this point, we would like to explicitly clarify and explain some details of our results obtained. First, one has to keep in mind that in spite of our referring to the estimates as 'lower bounds', our results are not an exact bound (in a rigorous

mathematical sense) for the NLSE channel model: we presented the lower bound to an approximation of an initial model given by equations (8) and (9). Then, we have not considered full multiple-input–multiple-output system capacity and do not actually include add-drop multiplexers in the link believing, following ref. 5, that the case of COI with no side information corresponds to the worst case scenario capacity wise. We note that different assumptions for the input statistics can affect the capacity estimates[50]. We are assuming here that all channels are transmitting symbols with the same statistics and input power that is known at the receiver. According to the recently proposed classification of work[50], this corresponds to the so-called adaptive interferer distribution. To avoid a possible confusion, one should notice that the result given by equation (11) does not conflict with the lower-bound estimation for the zero-dispersion channel obtained in ref 39: when the dispersion β_2 goes to 0, the window of applicability of the result (11), given by two formulas above, closes, such that one cannot perform a correct comparison.

Another theory aspect is that, strictly speaking, the integrability of the NLSE (1) is lost due to the noise action (again, in a mathematical sense) even when one resides within the applicability limits protocolled above. However, when the conditions above are met, the NFT-type analysis still describes the system behaviour correctly, as it is guaranteed by the perturbation theory[34,35]. On the other hand, the perturbation theory used here cannot describe non-adiabatic phenomena like, for example, the creation of new solitonic eigenstates. Supplementary Fig. 4 from the Supplementary Note 7 demonstrates that there is no significant noise influence on the NFT domain bandwidth within the theory applicability range. Hoverer, there is still lack of the detailed study of how the signal bandwidth within the NFT domain behaves in response to the noise action when one is far beyond the perturbative regime.

Evidence of the absence of the nonlinear channel crosstalk. In this section, we provide the results of the numerical simulations corroborating the predictions of the theory elaborated in the Results section above. One of the main challenges of the NFT method is to demonstrate that it is free from the nonlinear inter-channel interference that is thought to be the capacity bottleneck in the conventional systems.

We aim at studying how both linear and nonlinear spectra evolve with the propagation distance. Since in the NIS scheme the nonlinear spectrum at the encoder and decoder coincides with the linear spectrum of the generated and detected sequences correspondingly, we shall present the results for the latter spectra in the real-world units. To be specific, we consider two cases: for the first, one we take 100 subcarriers OFDM over 500 GHz, and for the second example, we use the five-channel Nyquist-based WDM modulation with the same rate, as it was considered in the main text. For each format, in all our simulations, we used the same fixed realization of the symbol coefficients c_{zk} in the input form (7). As an example, we utilized the quadrature phase shift keying modulation of coefficients
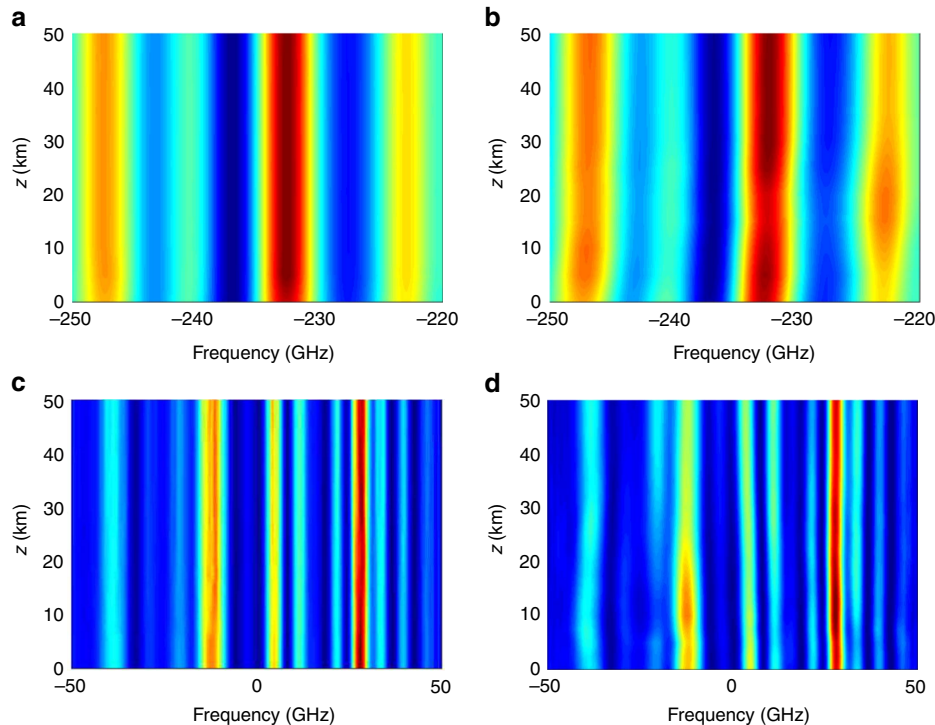
Figure 3 | Linear versus Nonlinear Spectrum. (**a**) The evolution of the nonlinear spectrum of a 100 OFDM subcarriers with the 5 GHZ carrier spacing. (**b**) The evolution of the same sequence without the NIS module. (**c**) The evolution of the nonlinear spectrum for $N_{ch} = 5$ WDM channels in the NIS-based transmission. (**d**) The same but without the NIS block.

throughout. In all the cases, we compare the propagation in the NIS scheme with the same one in the absence of the NIS blocks, that is, when the signal (7) is actually synthesized and launched directly into the fibre (without digital NFT pre- and post-processing). To achieve a fair comparison, we made sure that the average optical power of the pulse launched into the fibre was the same both with and without NIS, so that $S_{opt} \sim 12$ dBm for OFDM and ~ 18 dBm for the Nyquist case. For the OFDM case, the input power levels were chosen to correspond to the optimal launch energy $E_{in}^* = E_{NL}/\sqrt{2}$. For the Nyquist case, the power levels were chosen to be higher, to illustrate the stability of the NIS-based transmission. Note that in both cases the average signal level drops very quickly with the propagation distance due to the dispersion broadening, which in the absence of the soliton component decays almost as fast as it does in the linear case. Therefore, the nonlinear interaction between the different frequencies (which is expected to plague the conventional transmission, see the right column in Fig. 3) only takes place during the initial stages of evolution, and so there is no need to consider long spans. Also, due to the dispersion-induced power degradation the PSD of the signal very quickly becomes comparable with that of the noise making the spectral evolution curves uninformative. Therefore, in this section, we only present the spectral traces for the noise-free case, as the main goal here is to illustrate that the nonlinear spectra are immune to the nonlinear cross- and self-phase modulation. The results for the noise PSD under the same initial conditions are presented in Supplementary Note 6. The pulse evolution was studied by means of standard split-step scheme[1] with a spatial step of 200 m. The forward and inverse NFT operations required for simulating the spectra in the left column were obtained using transfer matrix and Toeplitz matrix inversion respectively (see, for example, ref. 22).

In both cases, we are only showing the magnified part of the spectrum. For the OFDM, the higher four subcarriers are shown while for the Nyqusit case only the central COI is plotted. From Fig. 3, one can clearly see that the linear Fourier spectrum gets distorted during the conventional transmission (the right column), while its nonlinear counterpart remains robust (the left column).

Additionally in the Supplementary Note 6, we show how the nonlinear PSD evolves during the propagation in the multichannel environment similar to that of Fig. 3.

Together, with the simulations described above, we conducted a set of numerical experiments aimed at studying whether the soliton modes can emerge in the NIS scheme due to the noise action. Decreasing the effective correlation lengths of the numerical noise (the z interval of the noise injection during the NLSE simulations and the elementary time sample duration), we found that the amount of the total energy contained into the soliton degrees of freedom became $<2\%$ at 1,500 km when the time correlation duration was 5.25 ps and z correlation length 500 m, for a single channel with $W_0 = 100$ GHz, $S = 22.3$ dBm (32 Nyquist pulses in the burst). As we observed a steady tendency for the decrease of solitonic signal part with the contraction of correlation lengths in both time and space, we believe that, within the applicability limits of the perturbation theory, for the 'very white' noise, the effect of solitonic constituents emerging from noise is of higher smallness order and can be neglected (at least, for the ideal Raman amplification case) within the leading order of perturbation approach developed in our work.

References

1. Agrawal, G. P. *Fiber-Optic Communication Systems* 4th edn (Wiley-Blackwell, 2010).
2. Iannoe, E., Matera, F., Mecozzi, A. & Settembre, M. *Nonlinear Optical Communication Networks* (John Wiley & Sons, 1998).
3. Mitra, P. P. & Stark, J. B. Nonlinear limits to the information capacity of optical fibre communications. *Nature* **411**, 1027–1030 (2001).
4. Essiambre, R.-J., Foschini, G. J., Kramer, G. & Winzer, P. J. Capacity limits of information transport in fiber-optic networks. *Phys. Rev. Lett.* **101**, 163901 (2008).
5. Essiambre, R. *et al.* Capacity limits of optical fiber networks. *J. Lightwave Technol.* **28**, 662–701 (2010).
6. Ellis, A. D., Zhao, J. & Cotter, D. Approaching the non-linear Shannon limit. *J. Lightwave Technol.* **28**, 423–433 (2010).
7. Killey, R. I. & Behrens, C. Shannon's theory in nonlinear systems. *J. Mod. Optics* **58**, 1–10 (2011).
8. Richardson, D. J. Filling the light pipe. *Science* **330**, 327–328 (2010).
9. Meron, E., Shtaif, M. & Feder, M. Beneficial use of spectral broadening resulting from the nonlinearity of the fiber-optic channel. *Opt. Lett.* **37**, 4458–4460 (2012).
10. Hasegawa, A. & Kodama, Y. *Solitons in Optical Communications* (Oxford Univ. Press, 1995).
11. Mollenauer, L. F. & Gordon, J. P. *Solitons in Optical Fibers: Fundamentals and Applications* (Academic Press, 2006).
12. Splett, A., Kurzke, C. & Petermann, K. in *Proceedings of European Conference on Optical Communications (ECOC)* 41–44 (Montreux, Switzerland, 1993).
13. Zakharov, V. E. & Shabat, A. B. Exact theory of 2-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. JETP* **34**, 62–69 (1972).
14. Ablowitz, N. J., Kaup, D. J., Newell, A. C. & Segur, H. The inverse scattering transform-Fourier analysis for nonlinear problems. *Stud. Appl. Math.* **53**, 249–315 (1974).
15. Ablowitz, M. J. & Segur, H. *Solitons and the Inverse Scattering Transform* (SIAM, 1981).
16. Hasegawa, A. & Nyu, T. Eigenvalue communication. *J. Lightwave Technol.* **11**, 395–399 (1993).
17. Yousefi, M. I. & Kschischang, F. R. Information transmission using the nonlinear Fourier transform, Parts I–III. *IEEE Trans. Inform. Theory* **60**, 4312–4369 (2014).
18. Turitsyna, E. G. & Turitsyn, S. K. Digital signal processing based on inverse scattering transform. *Opt. Lett.* **38**, 4186–4188 (2013).
19. Wahls, S., Le, S. T., Prilepsky, J. E., Poor, H. V. & Turitsyn, S. K. in *Proceedings of IEEE 16th International Workshop in Signal Processing Advances in Wireless Communications (SPAWC)*, 445–449 (Stockholm, Sweden, 2015).
20. Ip, E. & Kahn, J. Compensation of dispersion and nonlinear impairments using digital backpropagation. *J. Lightwave Technol.* **26**, 3416–3425 (2008).
21. Prilepsky, J. E. *et al.* Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels. *Phys. Rev. Lett.* **113**, 013901 (2014).
22. Le, S. T., Prilepsky, J. E. & Turitsyn, S. K. Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers. *Opt. Express* **22**, 26720–26741 (2014).
23. Le, S. T., Prilepsky, J. E. & Turitsyn, S. K. Nonlinear inverse synthesis technique for optical links with lumped amplification. *Opt. Express* **23**, 8317–8328 (2015).
24. Le, S. T., Prilepsky, J. E., Rosa, P., Ania-Castañón, J. D. & Turitsyn, S. K. Nonlinear inverse synthesis for optical links with distributed Raman amplification. *J. Lightwave Technol.* **34**, 1778–1785 (2016).
25. Le, S. T. *et al.* Demonstration of nonlinear inverse synthesis transmission over transoceanic distances. *J. Lightwave Technol.* **34**, 2459–2466 (2016).
26. Dong, Z. *et al.* Nonlinear frequency division multiplexed transmissions based on NFT. *IEEE Photon. Technol. Lett.* **27**, 1621–1623 (2015).
27. Terauchi, H. & Maruta, A. in *18th OptoElectronics and Communications Conference held jointly with 2013 International Conference on Photonics in Switching (OECC/PS)* (Kyoto, Japan, 2013).
28. Maruta, A. in *Proceedings of the 20th OptoElectronics and Communications Conference (OECC)* (Shanghai, China, 2015).
29. Bülow, H. Experimental demonstration of optical signal detection using nonlinear fourier transform. *J. Lightwave Technol.* **33**, 1433–1439 (2015).
30. Aref, V., Bülow, H., Schuh, K. & Idler, W. in *Proceedings of European Conference on Optical Communications (ECOC)* (Valencia, Spain, 2015).
31. Meron, E., Feder, M. & Shtaif, M. On the achievable communication rates of generalized soliton transmission systems. Preprint at http://arXiv:1207.0297 (2012).
32. Oda, S., Maruta, A. & Kitayama, K. All-optical quantization scheme based on fiber nonlinearity. *IEEE Photonics Technol Lett.* **16**, 587–589 (2004).
33. Prilepsky, J. E., Derevyanko, S. A. & Turitsyn, S. K. Nonlinear spectral management: linearization of the lossless fiber channel. *Opt. Express* **21**, 24344–24367 (2013).
34. Kaup, D. J. Perturbation expansion for Zakharov-Shabat inverse scattering transform. *SIAM J. Appl. Math.* **31**, 121–133 (1976).
35. Kaup, D. J. & Newell, A. C. Solitons as particles, oscillators, and in slowly changing media—singular perturbation-theory. *Proc. R. Soc. London A Math.* **361**, 413–446 (1978).
36. Pinsker, M. S. *Information and Informational Stability of Random Variables and Processes* 160–201 (Holden Day, 1964).
37. Dar, R., Shtaif, M. & Feder, M. New bounds on the capacity of the nonlinear fiber-optic channel. *Opt. Lett.* **39**, 398–401 (2014).
38. Secondini, M. & Forestieri, E. On XPM mitigation in WDM fiber-optic systems. *IEEE Photonics Technol. Lett.* **26**, 2252–2255 (2014).
39. Turitsyn, K. S., Derevyanko, S. A., Yurkevich, I. V. & Turitsyn, S. K. Information capacity of optical fiber channels with zero average dispersion. *Phys. Rev. Lett.* **91**, 203901 (2003).
40. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Technol. J.* **27**, 379–423 (1948).
41. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 2nd edn (Wiley, 2006).
42. Terekhov, I. S., Vergeles, S. S. & Turitsyn, S. K. Conditional probability calculations for the nonlinear Schrödinger equation with additive noise. *Phys. Rev. Lett* **113**, 230602 (2014).
43. Zakharov, V. E. & Manakov, S. V. Asymptotic behavior of non-linear wave systems integrated by the inverse scattering method. *Sov. Phys. JETP* **44**, 106–112 (1976).

44. Novokshenov, V. I. Asymptotics as $t \to \infty$ of the solution of the Cauchy-problem for the non-linear Schrödinger-equation. *Sov. Phys. Doklady* **251,** 799–802 (1980).

45. Schmogrow, R. *et al.* Real-time Nyquist pulse generation beyond 100 Gbit/s and its relation to OFDM. *Opt. Express* **20,** 317–337 (2012).

46. Shieh, W., Bao, H. & Tang, Y. Coherent optical OFDM: theory and design. *Opt. Express* **16,** 841–850 (2008).

47. Agrell, E., Alvarado, A., Durisi, G. & Karlsson, M. Capacity of a nonlinear optical channel with finite memory. *J. Lightwave Technol.* **16,** 2862–2876 (2014).

48. Agrell, E. Conditions for a monotonic channel capacity. *IEEE Trans. Commun.* **63,** 738–748 (2015).

49. Yousefi, M. I. & Yangzhang, X. Nonlinear frequency-division multiplexing. Preprint at http://arXiv:1603.04389 (2016).

50. Agrell, E. & Karlsson, M. Influence of behavioral models on multiuser channel capacity. *J. Lightwave Technol.* **33,** 3507–3515 (2015).

Acknowledgements

Author contributions

S.K.T. initiated the study. S.A.D. and J.E.P. derived the results, performed the simulations and prepared the figures together with the accompanying Supplementary Notes and Figures. S.K.T., S.A.D. and J.E.P. contributed equally to the preparation of the manuscript.

Additional information