Information Transmission using the Nonlinear Fourier Transform, Part III: Spectrum Modulation

Mansoor I. Yousefi and Frank R. Kschischang Fellow, IEEE

Abstract—Motivated by the looming “capacity crunch” in fiber-optic networks, information transmission over such systems is revisited. Among numerous distortions, inter-channel interference in multiuser wavelength-division multiplexing (WDM) is identified as the seemingly intractable factor limiting the achievable rate at high launch power. However, this distortion and similar ones arising from nonlinearity are primarily due to the use of methods suited for linear systems, namely WDM and linear pulse-train transmission, for the nonlinear optical channel. Exploiting the integrability of the nonlinear Schrödinger (NLS) equation, a nonlinear frequency-division multiplexing (NFDM) scheme is presented, which directly modulates non-interacting signal degrees-of-freedom under NLS propagation. The main distinction between this and previous methods is that NFDM is able to cope with the nonlinearity, and thus, as the the signal power or transmission distance is increased, the new method does not suffer from the deterministic cross-talk between signal components which has degraded the performance of previous approaches. In this paper, emphasis is placed on modulation of the discrete component of the nonlinear Fourier transform of the signal and some simple examples of achievable spectral efficiencies are provided.

Index Terms—Fiber-optic communications, nonlinear Fourier transform, Darboux transform, multisoliton transmission.

I. INTRODUCTION

This paper is a continuation of Part I [1] and Part II [2] on data transmission using the nonlinear Fourier transform (NFT). [Part I] describes the mathematical tools underlying this approach to communications. Numerical methods for implementing the NFT at the receiver are discussed in [Part II]. The aims of this paper are to provide methods for implementing the inverse NFT at the transmitter, to discuss the influence of noise on the received spectra, and to provide some example transmission schemes, which illustrate some of the spectral efficiencies achievable by this method.

The proposed nonlinear frequency-division multiplexing (NFDM) scheme can be considered as a generalization of orthogonal frequency-division multiplexing (OFDM) to integrable nonlinear dispersive communication channels [1]. The advantages of NFDM stem from the following:

1) NFDM removes deterministic inter-channel interference (cross-talk) between users of a network sharing the same fiber channel;

2) NFDM removes deterministic inter-symbol interference (ISI) (intra-channel interactions) for each user;

3) spectral invariants as carriers of data are remarkably stable and noise-robust features of the nonlinear Schrödinger (NLS) flow;

4) with NFDM, information in each channel of interest can be conveniently read anywhere in a network independently of the optical path length(s).

As described in [Part I], the nonlinear Fourier transform of a signal with respect to a Lax operator consists of discrete and continuous spectral functions, in one-to-one correspondence with the signal. In this paper we focus mainly on discrete spectrum modulation, which captures a large class of input signals of interest (see Section VI). For this class of signals, the inverse NFT is a map from $2N$ complex parameters (discrete spectral degrees-of-freedom) to an $N$-soliton pulse in the time domain. This special case corresponds to an optical communication system employing multisoliton transmission and detection in the focusing regime.

A physically important integrable channel is the optical fiber channel. Despite substantial effort, fiber-optic communications using fundamental solitons (i.e., 1-solitons) has faced numerous challenges in the past decades. This is partly because the spectral efficiency of conventional soliton systems is typically quite low ($\rho \sim 0.2$ bits/s/Hz), but also because on-off keyed solitons interact with each other, and in the presence of noise the system reach is limited by the Gordon-Haus effect [3]. Although solutions have been suggested to alleviate these limitations [5], most current research is focused on the use of spectrally-efficient pulse shapes, such as sinc and raised-cosine pulses, with digital backpropagation at the receiver, e.g., [4]. Although these approaches provide a substantial spectral efficiency at low to moderate signal-to-noise ratios (SNRs), their efficacy saturates after a finite $\text{SNR} \sim 25 \text{ dB}$ where $\rho \sim 5$ bits/s/Hz. This, as we shall see in Section VI, is due to the incompatibility of the wavelength-division multiplexing (WDM) with the flow of the NLS equation, causing severe inter-channel interference.

There is a vast body of literature on solitons in mathematics, physics, and engineering; see, e.g., [3], [5–8] and references therein. Classical, path-averaged and dispersion-managed fundamental 1-solitons are well-studied in fiber optics [3]. The existence of optical $N$-soliton pulses in optical fibers is also well known [3]; this previous work is mostly confined to the pulse-propagation properties of $N$-solitons, is usually
limited to small $N$ (e.g., $N = 2, 3$), or focuses on specific isolated $N$-solitons (e.g., pulses of the form $A \text{sech}(t)$). Signal processing problems (e.g., detection and estimation) involving soliton signals in the Toda lattice and other models have been considered by Singer [3]. There is also a related work by Hasegawa and Nu, “eigenvalue communication” [16], which is reviewed in Section VII after NFDM is explained.

While a fundamental soliton can be modulated, detected and analyzed in the time domain, $N$-solitons are best understood via their spectrum in the complex plane. In this paper, these pulses are obtained by implementing a simplified inverse NFT at the transmitter using the Darboux transform and are demodulated at the receiver by recovering their spectral content using the forward NFT. Since the spectral parameters of a multisoliton naturally do not interact with one another (at least in the absence of noise), there is potentially a great advantage in directly modulating these non-interacting degrees-of-freedom. Sending an $N$-soliton train for large $N$ and detecting it at the receiver—a daunting task in the time domain due to the interaction of the individual components—can be efficiently accomplished, with the help of the NFT, in the nonlinear frequency domain.

The paper is organized as follows. In Section II we revisit the wavelength-division multiplexing method commonly used in optical fiber networks and identify inter-channel interference as the capacity bottleneck in this method. This section provides further motivation for the NFT approach taken here. In Section III we study algorithms for implementing the inverse nonlinear Fourier transform at the transmitter for signals having only a discrete spectrum. Among several methods, the Darboux transform is found to provide a suitable approach. The first-order statistics of the discrete spectrum in the presence of noise are calculated in Section IV. In Section V we calculate some spectral efficiencies achievable using very simple NFT examples. Finally, we provide some remarks on the use of the NFT method in Section VI and conclude the paper in Section VII.

II. ORIGIN OF CAPACITY LIMITS IN WDM OPTICAL NETWORKS

Recent studies on the capacity of WDM optical fiber networks suggest that the information rates of such networks is ultimately limited by the impacts of the nonlinearity, namely inter-channel and intra-channel nonlinear interactions [4]. These distortions have deterministic and signal-dependent stochastic components that grow with the input signal power, diminishing the achievable rate at high signal-to-noise ratios. In these studies, for a class of ring constellations, the achievable rates of the WDM method increases with input power power, $\mathcal{P}$, reaching a peak at a certain critical input power, and then asymptotically vanishes as $\mathcal{P} \to \infty$ (see e.g., [4] and references therein).

In this section we briefly review WDM, the method commonly used to multiplex many channels in practical optical fiber systems. We identify the origin of capacity limitations in this model and explain that this method and similar ones, which are borrowed from linear systems theory, are poorly suited for efficient communication over nonlinear optical fiber networks. In particular, certain capacity limitations in the prior work are an artifact of these methods (notably WDM) and may not be fundamental. In subsequent sections, we continue the development of the NFDM approach that is able to overcome some of these limitations, in a manner that is fundamentally compatible with the structure of the nonlinear fiber-optic channel.

A. System Model

For convenience, we reproduce the system model given in [Part I]. We consider a standard single-mode fiber with dispersion coefficient $\beta_2$, nonlinearity parameter $\gamma$ and length $L$. After appropriate normalization (see, e.g., [Part I]), the evolution of the slowly-varying part $q(t,z)$ of a narrowband signal as a function of retarded time $t$ and distance $z$ is well modeled by the stochastic nonlinear Schrödinger equation

$$jq_z(t,z) = q_t + 2|q(t,z)|^2 q(t,z) + n(t,z),$$  \hspace{1cm} (1)

where subscripts denote differentiation and $n(t,z)$ is a band-limited white Gaussian noise process, i.e., with

$$\mathbb{E} \{ n(t,z)n^*(t',z') \} = \frac{\sigma_0^2}{PT_0} \delta_W(t-t') \delta(z-z'),$$

where $\delta_W(x) = 2W \text{sinc}(2Wx)$, $P = \frac{\sqrt{\sigma_0}}{\mathcal{L}}$ and $T_0 = \sqrt{\frac{|\beta_2|L}{2}}$. The noise spectral density is given by $\sigma_0^2 = n_0 \alpha h \nu$, with parameters given in Table I. It is assumed that the transmitter is bandlimited to $W$ for all $\ell$, $0 \leq \ell \leq \mathcal{L}$, and power limited to $\mathcal{P}$, i.e.,

$$\mathbb{E} \left[ \int_0^\mathcal{T} |q(t,0)|^2 dt \right] = \mathcal{P}.$$

Note that the fiber loss is assumed to be perfectly compensated by distributed Raman amplification, and hence (1) contains no loss term. A derivation of (1) and a discussion about sources of noise in fiber-optic channels can be found in [3, 9].

Stochastic partial differential equations (PDEs) such as (1) are usually interpreted via their equivalent integral representations. Integrating a function with unbounded variation can be problematic, since, e.g., a Riemann approximation

$$\int_z^{z+dz} g(z) dB(z) \approx g(l) (B(z + dz) - B(z)),$$

for some $l \in [z, z + dz]$, where $B$ is the Wiener process, would depend on the choice of $l$. The choice of $l$ leads to various interpretations for a stochastic PDE, notably Itô and Stratonovich representations, in which, respectively, $l = z$ and $l = z + dz/2$. Since noise is bandlimited in its temporal component, the stochastic PDE (1) is essentially a finite-dimensional system and there is no difficulty in the rigorous interpretation of (1).

| Parameter | Value |
|-----------|-------|
| $n_0$     | 1     |
| $\hbar$  | $6.626 \times 10^{-34}$ J·s |
| $\nu$    | 193.55 THz |
| $\alpha$ | 0.046 km$^{-1}$ |
| $\gamma$ | 1.27 W$^{-1}$km$^{-1}$ |
| $W$      | 125 GHz |
B. Capacity of WDM Optical Fiber Networks

Fiber-optic communication systems use wavelength-division multiplexing to transmit information. Similar to frequency-division multiplexing, information is multiplexed in distinct wavelengths. This helps to separate the signals of different users in a network, when they have to share the same links between different nodes.

Fig. 1 shows the system model of a link in an optical fiber network between a source and a destination. There are $N$ fiber spans between multiple users at the transmitter (TX) and multiple users at the receiver (RX). The signal of some of these users is destined to a receiver other than the RX shown in Fig 1. As a result, at the end of each span there is a ROADM that may drop the signal of some of the users or, if there are unused frequency bands, add the signal of potential external users. We are interested in evaluating the per-degree-of-freedom capacity (bits/s/Hz) of the fiber link from the transmitter (TX) to the receiver (RX).

In WDM, the following signal is transmitted over the channel

$$q(t, 0) = \sum_{k=0}^{N-1} \left( \sum_{\ell=1}^{WT} s_k^{\ell} \phi_{k}(t) \right) e^{j2k\pi W t}, \quad 0 \leq t \leq T,$$

where $s_k^{\ell}$ are the symbols transmitted by user $k$, $\phi_{k}(t)$ are pulse shapes, $W \geq 1/T$ is the per-user bandwidth, and $N$ is the number of WDM channels. For the purposes of illustration, we can assume that each user sends an isolated pulse in the time interval $[0, T]$. Thus each user operates at a single frequency in a bandwidth $W = 1/T$ and $q(t, 0) = \sum_{k=0}^{N-1} q_k(0) \exp(jk2\pi t/T)$, where $\{q_k(0)\}$ are the Fourier series coefficients at $z = 0$. As this signal evolves along $z$, we have a Fourier series with variable coefficients

$$q(t, z) = \sum_{k=0}^{N-1} q_k(z)e^{j2k\pi z}.$$  \(3\)

Substituting the periodic solution (3) into (1), we get the NLS equation in the discrete frequency domain

$$j \partial q_k(z) \partial z = -4\pi^2 W^2 k^2 q_k(z) + 2|q_k(z)|^2 q_k(z) $$
$$+ 2q_k(z) \sum_{\ell \neq k} |q_{\ell}(z)|^2 \quad \text{dispersion (SPM)}$$
$$+ 2 \sum_{m \neq k} q_m(z)q_{\ell}(z)q_{m+\ell-k}(z) + v_k(z),$$  \(4\)

in which $v_k$ are the noise coordinates in frequency and where we have identified the dispersion, self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM) terms in the frequency domain.

It is important to note that the optical WDM channel is a (nonlinear) multiuser interference channel [10]. The (inter-channel) interference terms are the XPM and FWM. There is no ISI in the assumed isolated pulse transmission model [11] with one degree of freedom per user. However in a pulse-train transmission model where $WT > 1$, replacing $\{q_k\}$ via the inverse transform shows that the other two effects, the dispersion and SPM, cause inter-symbol interference (intra-channel interaction). Performance of a WDM transmission system depends on how interference and ISI are treated, and in particular the availability of the user signals at the receiver. Several cases can be considered.

Assuming a given (perhaps suboptimal) receiver, (3) is discretized as a map from degrees-of-freedom $s_k^{\ell}$ at the input to their estimate $s_k^{\ell}$ at the output:

$$(s_1^{\ell}, s_1^{\ell}, \ldots, s_2^{\ell}, s_2^{\ell}, \ldots) \rightarrow (s_1^{\ell}, s_1^{\ell}, \ldots, s_2^{\ell}, s_2^{\ell}, \ldots).$$

If joint transmission and detection of all degrees of freedom $s_k^{\ell}$ is practical, the channel is essentially a single user channel whose capacity is non-decreasing with average input power

$$\mathcal{P} = \frac{1}{NW T} \sum_{k=0}^{N-1} \sum_{\ell=1}^{WT} E[s_k^{\ell}]^2.$$  \(5\)

If joint transmission and detection is not possible (e.g., in frequency $k$ or in time $\ell$), then one must adopt the strategy of treating interference as noise. In this case, one can examine various strategies, e.g., interference alignment [11] or signal-space orthogonalization. If one of these strategies can be successfully applied, then, again, the capacity is non-decreasing with $\mathcal{P}$. If none of these strategies is applicable, or if additional constraints are present, the capacity of a WDM system may in some cases decrease with $\mathcal{P}$.

a) Single-user Memoryless Channels: The capacity-cost function of a single user discrete memoryless channel (DMC) with unbounded input alphabet (real- or complex-valued symbols with no peak power), is a non-decreasing function of a (equality or inequality constrained) linear cost [12]. This follows from the concavity of the mutual information and the linearity of the cost as a function of the input distribution (see [12] for a detailed discussion about more general scenarios). This, of course, holds true for any set of transition probabilities (a DMC), including that obtained from nonlinear channels.

Care should be taken when one has more than one degree-of-freedom, if additional constraints are present, or if the capacity-cost program is non-convex. For instance, when a signal propagates according to a nonlinear evolution equation, its spectrum can spread continuously. The amount of spectral broadening depends on the pulse shape and, in particular, on the signal intensity. In the absence of optical filtering, the per degree-of-freedom capacity (bits/s/Hz) of the single-user nonlinear channel is expected to be non-decreasing with input power (even under the suboptimal backpropagation receiver); however, it may asymptotically vanish in the presence of optical filtering. Here, by per degree-of-freedom capacity, we mean the ratio of the sum rate to the maximum number of

1Note that this may not hold for finite alphabet channels (with equality power constraint), or when constraints limit the support of the input distribution. This can occur, e.g., in the presence of a peak power constraint, or a maximum bandwidth constraint, in a channel—such as [4]—in which each input symbol has a power or bandwidth cost. Therefore, in computer simulations, one should be wary of peaks in $C(\mathcal{P})$ which are close to the peak power.
signal space dimensions occupied by the transmitted signals during the entire evolution (capacity per maximum input cost—see Section V-A and 29). Spectral broadening is an important issue in zero-dispersion fibers where the channel is strongly nonlinear [13].

b) Single-user Channels with Memory: The SPM term of each user is available at the receiver for that user. Its deterministic part, if needed, can be removed, e.g., by backpropagation and its (signal-dependent) stochastic part can be handled by coding and optimal detection over a long block of data (e.g., maximum likelihood sequence detection). Deterministic or stochastic intra-channel effects do not limit the capacity if such joint detection is performed.

This may not be true for a receiver which performs isolated symbol detection. Whereas ISI is a deterministic distortion in traditional linear channels, for a nonlinear waveform channel ISI can have a stochastic component as well. As a result, unlike a waveform additive white Gaussian noise channel where ISI is removed completely by channel inversion, in nonlinear systems such channel inversion (e.g., by backpropagation) leaves a residual stochastic ISI for each symbol. In this case, the receiver based on backpropagation and isolated symbol detection gives rise to an ISI-limited system with a suboptimal performance.

c) Multi-user Channels with Interference: Inter-channel interference in the frequency domain is the dual of intra-channel ISI in the time domain; the difference is that cooperation and joint detection is generally not possible among users. In an optical fiber network, many users have to share the same optical fiber link. The signals of some of these users join and leave the optical link at intermediate points along the fiber, leaving behind a residual nonlinear impact. Thus we should assume that each user has access only to the signal in its own frequency band, and the signal of users \( k' \), where \( k' \neq k \), is unknown to the \( k \)th user.

In a nonlinear channel where, by definition, additivity is not preserved under the action of the channel, multiplexing signals in a linear fashion, e.g., by adding them in time or in frequency— in reconfigurable optical add-drop multiplexers (ROADMs)—leads to inter-channel interference. The interference resulting from these signals is unavoidably treated as noise and ultimately limits the achievable rates of such optical fiber systems.

Assuming that the coefficients \( q_k(z) \) remain independent in (4) during the evolution, the XPM term is statistically regulated, so that as \( N \to \infty \),

\[
\sum_{l \neq k} |q_l(z)|^2 \to \mathcal{P}.
\]

Thus XPM does not influence the capacity of optical fibers significantly.

FWM, on the other hand, is stronger (cubic in the signal amplitude) and is not averaged out like XPM is [14]. It is obvious that this interference will blow up when increasing the common average power \( \mathcal{P} (N \gg 1) \), ultimately overwhelming the signal and limiting the achievable rate. The per-degree-of-freedom achievable rates of the WDM method versus power, \( I(\mathcal{P}) \), is noise-limited in the low SNR regime, following \( \log(1 + \text{SNR}) \), and interference-limited in the high SNR regime, decreasing to zero [4], [13], [15]; see Fig. 4.

To summarize the preceding discussion, the transmission rates achievable over the nonlinear Schrödinger channel depend on the method of transmission and detection, as well as the assumptions on the model. One can assume a single-user or a multiple-user channel, with or without memory, and with or without optical filtering. It is thus important, when comparing different results, to clarify which modelling assumptions have been made.

C. Inter-channel Interference as the Capacity Bottleneck

It follows that while intra-channel interference can be handled by signal processing, inter-channel interference ultimately limits the achievable rate of optical fiber networks. The current practice in fiber-optic communications is to send a linear sum of signals in time (i.e., a pulse train) and in frequency (i.e., WDM) in the form of (2), both of which are not particularly well suited for the nonlinear fiber channel. Thus we identify ROADM as the main culprit for capacity limitation in currently deployed optical fiber networks. A modification of these devices is needed so that the incoming signals are multiplexed in a nonlinear fashion, exciting non-interacting degrees-of-freedom of the NLS equation.

To illustrate the effect of inter-channel interference in the application of the WDM method to the nonlinear fiber channel, we have simulated transmission of 5 WDM channels over 2000 km of standard single mode fiber with parameters from Table I. At the end of each span, a ROADM filters the central channel of interest (COI) at 40 GHz bandwidth, and adds four independent side-band signals randomly from a common constellation. At the channel output, the COI is filtered and backpropagated according to the inverse NLS equation. Fig. 2
Fig. 2. (a) 5 WDM channels in the frequency, with the channel of interest (COI) at the center. Neighbor channels are dropped and added at the end of each span, creating a leftover interference for the COI. (b) Channel of interest at the input (dotted rectangle) and at the output after backpropagation (solid curve). The mismatch is due to the fact that the backpropagation is performed only on the channel of interest and the interference signals cannot be backpropagated. (c) Inter-channel interference is increased with signal intensity.

Fig. 3. Achievable rates of the WDM method under weak and strong inter-channel interference.

The estimated information rate of this system is shown in Fig. 3. It is clear from both Fig. 2 and Fig. 3 that as the average transmitted power is increased, the signal-to-noise ratio in the COI, and as a result the information rate, vanishes to zero. Note that this effect can also be predicted by a simple SNR analysis at the receiver.

Using the mathematical and numerical tools described in Parts I and II, this paper aims to show that it is possible to exploit the integrability of the nonlinear Schrödinger equation and induce a $k$-user interference channel on the NLS equation so that both the deterministic inter-channel and inter-symbol interferences are simultaneously zero for all users of a multiuser network. Such interference cancellation follows directly from the integrability of the cubic nonlinear Schrödinger equation in $1 + 1$ dimensions, and is generally not feasible for other types of nonlinearity (even if the nonlinearity is weaker than cubic!). The achievable rates of the suggested scheme are expected to be higher than those of the WDM method and the scheme has numerous other potential advantages, mentioned in Section II.

Remark 1. Note that for an information-theoretic study, it is not necessary to perform deterministic signal processing such as backpropagation. One is only concerned with transition probabilities, which include effects such as rotations or other deterministic transformations. Backpropagation just aids the system engineer to simplify the task of the signal recovery (being an invertible operation, it does not change the information content of the received signal).

III. THE DISCRETE SPECTRAL FUNCTION

A. Background

Here we briefly recall the definition of the discrete spectral function in the context of the nonlinear Schrödinger equation. We first consider the deterministic version of (1), where the noise is zero. Later, we will treat noise as a perturbation of the noise-free equation.

The nonlinear Fourier transform of a signal in (1) arises via spectral analysis of the operator

$$L = j\left(\frac{\partial}{\partial t} - q(t) - \frac{\partial}{\partial x} - q^*(t) - \frac{\partial}{\partial x}\right) = j(D\Sigma_3 + Q),$$

where $D = \frac{\partial}{\partial t}$,

$$Q = \begin{pmatrix} 0 & -q \\ -q^* & 0 \end{pmatrix} \quad \text{and} \quad \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
propagating according to (1) is obtained by solving the the Zakharov-Shabat eigenproblem \( Lu = \lambda v \), or equivalently

\[
v_t = \begin{pmatrix} -j\lambda & q(t) \\ -q^*(t) & j\lambda \end{pmatrix} v, \quad v(-\infty, \lambda) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-j\lambda t},
\]

where the initial condition was chosen based on the assumption that the signal \( q(t) \) vanishes as \( |t| \to \infty \). The system of ordinary differential equations (6) is solved from \( t = -\infty \) to \( t = +\infty \) to obtain \( v(\pm \infty, \lambda) \). The nonlinear Fourier coefficients \( a(\lambda) \) and \( b(\lambda) \) are then defined as

\[
a(\lambda) = \lim_{t \to -\infty} v_1(t, \lambda) e^{j\lambda t}, \\
b(\lambda) = \lim_{t \to -\infty} v_2(t, \lambda) e^{-j\lambda t}.
\]

Finally, the discrete spectral function is defined on the upper half complex plane \( \mathbb{C}^+ = \{ \lambda : \Im(\lambda) > 0 \} \):

\[
\hat{q}(\lambda_j) = \frac{b(\lambda_j)}{a(\lambda_j)}, \quad j = 1, \ldots, N,
\]

where prime \(^'\) denotes differentiation and \( \lambda_j \) are the isolated zeros of \( a(\lambda) \) in \( \mathbb{C}^+ \), i.e., solutions of \( a(\lambda_j) = 0 \). The continuous spectral function is defined on the real axis \( \lambda \in \mathbb{R} \) as \( \hat{q}(\lambda) = b(\lambda)/a(\lambda) \).

### B. Modulating the Discrete Spectrum

Let the nonlinear Fourier transform of the signal \( q(t) \) be represented by \( q(t) \leftrightarrow (\hat{q}(\lambda), \tilde{q}(\lambda_j)) \). When the continuous spectrum \( \hat{q}(\lambda) \) is set to zero, the nonlinear Fourier transform consists only of discrete spectral functions \( \tilde{q}(\lambda_j) \), i.e., \( N \) complex numbers \( \lambda_1, \ldots, \lambda_N \) in \( \mathbb{C}^+ \) together with the corresponding \( N \) complex spectral amplitudes \( \tilde{q}(\lambda_1), \ldots, \tilde{q}(\lambda_N) \). In this case, the inverse nonlinear Fourier transform can be worked out in closed-form, giving rise to \( N \)-soliton pulses [17]. The simplified expressions, however, quickly get complicated when \( N > 2 \), and tend to be limited to low-order solitons.

One can, however, create and modulate these multisolitons numerically. In this section we study various schemes for the implementation of the inverse NFT at the transmitter when \( \hat{q} = 0 \).

1) Discrete Spectrum Modulation by Solving the Riemann-Hilbert System: The inverse nonlinear Fourier transform can be obtained by solving a Riemann-Hilbert system of integro-algebraic equations or, alternatively, by solving the Gelfand-Levitan-Marchenko integral equations. Great simplifications occur when \( \hat{q}(\lambda) \) is zero. For instance, in this case the integral terms in the Riemann-Hilbert system vanish and the integro-algebraic system of equations is reduced to an algebraic linear system, whose solution gives rise to \( N \)-soliton pulses.

Let \( V(t, \lambda_j) \) and \( \tilde{V}(t, \lambda_j^* \) denote the scaled eigenvectors associated with \( \lambda_j \) and \( \lambda_j^* \) defined by their boundary conditions at \( +\infty \) (they are denoted by \( V^1 \) and \( V^1 \) in [Part I]). Setting the continuous spectral function \( \hat{q}(\lambda) \) to zero in the Riemann-Hilbert system of [Part I], we obtain an algebraic system of equations

\[
\tilde{V}(t, \lambda_j^*) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{i=1}^{N} \tilde{q}(\lambda_i) e^{2j\lambda_i t} V(t, \lambda_i) / \lambda_i - \lambda_j^*,
\]

\[
V(t, \lambda_j) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sum_{i=1}^{N} \tilde{q}(\lambda_i) e^{-2j\lambda_i t} \tilde{V}(t, \lambda_i^*) / \lambda_i^*- \lambda_j.
\]

Define

\[
\mathbf{V}_{2 \times N} = \begin{pmatrix} V(t, \lambda_1) & V(t, \lambda_2) & \cdots & V(t, \lambda_N) \\ \tilde{V}(t, \lambda_1^*) & \tilde{V}(t, \lambda_2^*) & \cdots & \tilde{V}(t, \lambda_N^*) \end{pmatrix},
\]

\[
\mathbf{K}_{N \times N} = \begin{pmatrix} \tilde{q}_{12} e^{2j\lambda_1 t} & \tilde{q}_{22} e^{2j\lambda_2 t} & \cdots & \tilde{q}_{2N} e^{2j\lambda_N t} \\ \tilde{q}_{11} e^{2j\lambda_1 t} & \tilde{q}_{21} e^{2j\lambda_2 t} & \cdots & \tilde{q}_{2N} e^{2j\lambda_N t} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_{1N} e^{2j\lambda_1 t} & \tilde{q}_{2N} e^{2j\lambda_2 t} & \cdots & \tilde{q}_{NN} e^{2j\lambda_N t} \end{pmatrix},
\]

\[
\mathbf{J}_{1 \times N} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix},
\]

\[
\mathbf{J}_{2 \times N} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{pmatrix},
\]

\[
\mathbf{F}_{N \times 1} = \begin{pmatrix} \tilde{q}_1 e^{2j\lambda_1 t} & \tilde{q}_2 e^{2j\lambda_2 t} & \cdots & \tilde{q}_N e^{2j\lambda_N t} \end{pmatrix}^T.
\]

In this notation, the algebraic equations (7) are simplified to

\[
\tilde{V} = \mathbf{J}_1 + \mathbf{VK} \quad \mathbf{V} = \mathbf{J}_2 - \tilde{\mathbf{V}} \mathbf{K}^*.
\]

Note that \( \mathbf{K}^* \) is the complex conjugate of \( \mathbf{K} \) (not Hermitian). Therefore

\[
\mathbf{V} = (\mathbf{J}_2 - \mathbf{J}_1 \mathbf{K}^*) (\mathbf{I}_N + \mathbf{KK}^*)^{-1} = \mathbf{J} (\mathbf{I}_N + \mathbf{KK}^*)^{-1},
\]

and \( \mathbf{V}_2 = \mathbf{e}^T (\mathbf{I}_N + \mathbf{KK}^*)^{-1} \). The \( N \)-soliton formula is given by

\[
q(t) = -2je^T (\mathbf{I}_N + \mathbf{K}^* \mathbf{K})^{-1} \mathbf{F}^*.
\]

The right hand side is a complex scalar and has to be evaluated for every \( t \) to determine samples of \( q(t) \) everywhere.

**Example 1.** It is useful to see that the (scaled) eigenvector for a single soliton with spectrum \( \hat{q}(1 + j\omega / 2, z) = \hat{q}_0 e^{2\alpha z} e^{-j(\alpha^2 - \omega^2)z} \) is

\[
v(t, \lambda; z) = \frac{1}{2} \operatorname{sech}[\omega(t-t_0)] \left( e^{-j\Phi} \right),
\]

where \( \Phi = \alpha t + (\alpha^2 - \omega^2)z - \angle \hat{q}_0 - \frac{\pi}{2} \) and \( t_0 = \frac{1}{2} \log \left| \frac{\hat{q}_0}{\hat{q}_0} \right| - 2\alpha z \). The celebrated equation for the single soliton obtained from (8) is

\[
q(t) = -j\omega e^{-j\alpha t} e^{-j\omega \phi} \operatorname{sech} \left( \omega(t \pm t_0) \right).
\]

From the phase-symmetry of the NLS equation, the factor \(-j\) in (9) can be dropped. The real and imaginary part of the eigenvalue are the frequency and amplitude of the soliton. Note that the discrete spectral amplitude \( \hat{q}(\lambda) \) is responsible for the phase and time-center of the soliton.
Unfortunately the Riemann–Hilbert system is found to be occasionally ill-conditioned for large $N$ and, in the current form, is not the best method for numerical generation of $N$-solitons.

2) Discrete Spectrum Modulation via the Hirota Bilinearization Scheme: It is also possible to generate multisolitons without solving a Riemann–Hilbert system or directly using the NFT. A method which is particularly analytically insightful is the Hirota direct method [17]. It prescribes, in some sense, a nonlinear superposition for integrable equations.

The Hirota method for an integrable equation works by introducing a transformation of the dependent variable $q$ to convert the original nonlinear equation to one or more homogeneous bilinear PDEs. For integrable equations, the nonlinearity usually is canceled or separated out. The resulting bilinear equations have solutions that can be expressed as a sum of exponentials. Computationally, bilinear equations are solved perturbatively by expanding the unknowns in terms of the powers of a small parameter $\epsilon$. For integrable equations, this series truncates, rendering approximate solutions of various orders to be indeed exact. The bilinearization transformation has been found for many integrable equations [17], taking on similar forms that usually involve the derivatives of the logarithm of the transformed variable.

Let us substitute $q(t, z) = G(t, z)/F(t, z)$, where, without loss of generality, we may assume that $F(t, z)$ is real-valued. To keep track of the effect of nonlinearity, let us restore the nonlinearity parameter $\gamma$ in the NLS equation. Plugging $q = G/F$ into the NLS equation

$$j q_z = q_{tt} + 2|q|^2 q,$$

we get

$$j (G_z F - FG_z) = F G_{tt} - 2 F^2 G_t - G F_{tt}$$

$$+ 2 |F|^2 + |G|^2 G$$

$$= 2 F G_{tt} - 2 F^2 G_t + G F_{tt}$$

$$+ 2 |F|^2 + |G|^2 F - F F_{tt} G,$$

where we have added and subtracted $2 G F_{tt}$. Equation (10) is trilinear in $F$ and $G$. It can be made bilinear by setting

$$j (G_z F - FG_z) = F G_{tt} - 2 F^2 G_t + G F_{tt},$$

$$F_{tt}^2 + |G|^2 F - F F_{tt} = 0.$$  

(11)

It is very convenient (though not necessary) to organize using the Hirota $D$ operator

$$D^n (a(t), b(t)) = \left( \frac{\partial}{\partial t} - \frac{\partial}{\partial t'} \right)^n a(t)b(t)|_{t=t'},$$

resulting in

$$(j D_z + D^2_{tt}) FG = 0,$$

$$D^2_{tt} FF = 2|G|^2,$$  

(12)

(13)

Note that the $D$ operator acts on a pair of functions to produce another function. Note further that (12) does not depend on the nonlinearity parameter $\gamma$. That is to say, the nonlinearity has been separated from equation (12). For some other integrable equations (e.g., the Korteweg–de Vries equation) for which one gets only one bilinear PDE, the nonlinearity parameter is in fact canceled completely.

Bilinear Hirota equations (12)–(13) have solutions in the form of sum of exponentials. As shown in Appendix A $F$ and $G$ are obtained as [17], [18]:

$$F(t, z) = \sum_{b=0,1} \delta_1 (b) \exp \left( b^T x + b^T R b \right),$$

$$G(t, z) = \sum_{b=0,1} \delta_2 (b) \exp \left( b^T x + b^T R b \right),$$

where $b = [b_i]_{i=1}^{2N}$ is a binary vector, $b_i = \{0, 1\}$, $x = [X_i]_{i=1}^{2N}$,

$$[X_i]_{i=1}^{N} = \omega_i t - k_i z + X_i(0),$$

$k_i = j \omega_i$, $[X_i]_{i=N+1}^{2N} = [X^*]_{i=1}^{N}$, $R = [r_{ij}]_{i,j=1}^{2N}$ is the Riemann spectrum,

$$r_{ij} = \begin{cases} 
0 & i \geq j \\
-2 \log (\omega_i + \omega_j) & i = 1, 2, \ldots, N; j = N + 1, \ldots, 2N \\
-2 \log (\omega_i - \omega_j) & i = N + 1, \ldots, 2N; j = N + 1, \ldots, 2N
\end{cases}$$

and

$$\delta_1 (b) = \begin{cases} 
1 \sum_{i=1}^{N} b_i = N \sum_{i=1}^{N} b_{i+N}, \\
0 & \text{otherwise}
\end{cases}$$

$$\delta_2 (b) = \begin{cases} 
1 \sum_{i=1}^{N} b_i = 1 + \sum_{i=1}^{N} b_{i+N}, \\
0 & \text{otherwise}
\end{cases}$$

Note that, using the identity $\partial_t \log F = \frac{F_t F - F^2}{F^2}$, (13) is reduced to $|q(t, z)|^2 = \gamma^{-1} \partial_t \log F$. Therefore the amplitude of $q$ is captured by the real-valued function $F$, while $G$ contains the phase of the signal.

Two important observations follow from the Hirota method. Firstly, multisoliton solutions of the NLS equation in the $F, G$ domain ($q = G/F$) are the summation of exponentially decaying functions $e^{-\alpha t} e^{j \omega t}$, each located at a frequency $\omega$. That is to say, while plane waves $e^{j (\omega t - kz)}$ are the natural basis functions that solve linear PDEs, for integrable systems, exponentially decaying functions are suitable. The addition of the decaying factor $e^{-\alpha t}$ (when $t > 0$) is the point at which the nonlinear Fourier transform diverges from the linear Fourier transform [18]. Secondly, for each individual soliton term, the Hirota method adds two-way interaction terms, three-way interaction terms, etc., until all the interactions are accounted for. In this way, the interference between individual components is removed, as shown schematically in Fig. [1]. Tables [1] and [II] show these interaction terms for $N = 1, 2, 3$.

While the Hirota method reveals important facts about signal degrees-of-freedom in the NLS equation, it may not be the best method to compute multisolitons numerically. There are $\binom{2N}{N} \sim N^{2N}$ and $\binom{2N}{N+1} \sim N^{2N}$ terms in $F$ and $G$ respectively, and unless one truncates the interaction terms at some step, the complexity quickly grows, making it hard to compute $N$-solitons for $N > 10$.

3) Recursive Discrete Spectrum Modulation Using Darboux Transformation: Multisoliton solutions of the NLS equation can be constructed recursively using the Darboux transformation (DT). The Darboux transformation, originally introduced...
in the context of the Sturm-Liouville differential equations and later used in nonlinear integrable systems, provides the possibility to construct a solution of an integrable equation from another solution [19]. For instance, one can start from the trivial solution \( q = 0 \) of the NLS equation, and recursively obtain all higher order \( N \)-soliton solutions. This approach is particularly well suited for numerical implementation.

Let \( x(t; \lambda; q) \) denote a solution of the system

\[
\begin{align*}
    x_t &= P(\zeta, q)x, \\
    x_z &= M(\zeta, q)x,
\end{align*}
\]

for the signal \( q \) and complex number \( \zeta = \lambda \) (not necessarily an eigenvalue of \( q \)), where the \( P \) and \( M \) are \( 2 \times 2 \) matrix operators defined in [Part I]. It is clear that \( \dot{x} = [x^2, -x]^{\top} \) satisfies (14) for \( \zeta \to \zeta^* \), and furthermore, by cross-elimination, \( q \) is a solution of the integrable equation underlying (14).

The Darboux theorem is stated as follows.

**Theorem 1** (Darboux Transformation). Let \( \phi(t, \lambda; q) \) be a known solution of (14) and set \( \Sigma = STS^{-1} \), where \( S = [\phi(t, \lambda), \phi(t, \lambda)] \) and \( \Gamma = \text{diag}(\lambda, \lambda^*) \). If \( v(t, \mu; q) \) satisfies
Fig. 6. Darboux iterations for the construction of an $N$-soliton.

then $u(t, \mu; \tilde{q})$ obtained from the Darboux transform

$$u(t, \mu; \tilde{q}) = (\mu I - \Sigma) v(t, \mu; q),$$

satisfies (14) as well, for

$\tilde{q} = q - 2j(\lambda^* - \lambda) \frac{\phi_2^* \phi_1}{|\phi_1|^2 + |\phi_2|^2}$.

Furthermore, both $q$ and $\tilde{q}$ satisfy the integrable equation underlying the system (14).

Proof: See Appendix B

Theorem 1 immediately provides the following observations.

1) From $\phi(t, \lambda; q)$ and $v(t, \mu; q)$, we can obtain $u(t, \mu; \tilde{q})$ according to (15). If $\mu$ is an eigenvalue of $q$, then $\mu$ is an eigenvalue of $\tilde{q}$ as well. Furthermore, since $u(t, \mu = \lambda; \tilde{q}) = 0$, $\lambda$ is also an eigenvalue of $\tilde{q}$. It follows that the eigenvalues of $\tilde{q}$ are the eigenvalues of $q$ together with $\lambda$.

2) $\tilde{q}$ is a new solution of the equation underlying (14), obtained from $q$ according to (16), and $u(t, \mu; \tilde{q})$ is one of its eigenvectors.

These observations suggest a two-step iterative algorithm to generate $N$-solitons, as illustrated in the Figs. 5-6. Denote a $k$-soliton pulse with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ by

$q(t; \lambda_1, \lambda_2, \ldots, \lambda_k) := q^{(k)}$. The update equations for the recursive Darboux method are given in Table III. Note that $v(t, \lambda_j, q^{(k+1)})$ can also be obtained directly by solving the Zakharov-Shabat system (6) for $q^{(k+1)}$. It is however more efficient to update the required eigenvector according to Table III. The algorithm is initialized from the trivial solution $q = 0$. The initial eigenvectors in Fig 6 are chosen to be the (non-canonical) eigenvectors $v(t, \lambda_j) = [A_j e^{-j \lambda_j t}, B_j e^{j \lambda_j t}]^T$. The coefficients $A_j$ and $B_j$ control the spectral amplitudes and the shape of the pulses. For a single soliton, $A_j = \exp(j \mathcal{L} \tilde{q})$ and $B_j = |\tilde{q}|$.

Remark 2. It can be observed numerically that the discrete spectrum of a long random bandlimited wavetrain, e.g., that formed from a sinc basis, appears to contain most of the wavetrain energy. This implies that a realization of a stationary bandlimited stochastic process $x(t), t \in [-T/2, T/2]$, is likely to be a multisoliton as $T \to \infty$. In other words, a large class of input signals of interest are indeed multisolitons (though they may not look so in appearance).

C. Evolution of the Discrete Spectrum

Recall that the imaginary and real parts of the eigenvalues correspond, respectively, to soliton amplitude (energy) and frequency. If the discrete spectrum of the signal lies completely on the imaginary axis, the $N$-soliton does not travel while propagating (with respect to a traveling observer). The individual components of an $N$-soliton pulse with frequencies $\lambda_i$ off the $j\omega$ axis travel in retarded time with speeds proportional to $\Re \lambda_i$ (frequency).

The manner of $N$-soliton propagation thus depends on the choice of the eigenvalues. An $N$-soliton signal is essentially composed of $N$ single solitons coupled together, similar to a molecule which groups a number of atoms. If the eigenvalues have non-zero distinct real parts, various components travel at different speeds and eventually when $z \to \infty$ the $N$-soliton decomposes into $N$ separate solitons

$$q(t, z) \to \sum_{i=1}^{N} \omega_i e^{-j(\alpha_i t + k_i z + \phi_i)} \sech(\omega_i (t - t_i - \alpha_i z)),$$

where $\lambda_i = (\alpha_i + j\omega_i)/2$ are eigenvalues, $k_i = (\alpha_i^2 - \omega_i^2)/2$, and $t_i$ is the time center. This breakdown of a signal to its individual components, while best observed in the case of multisolitons, is simply a result of group velocity dispersion and exists for all pulses similarly (including sinc functions). The extent of breakdown and shift depends on a variety of factors, such the length of the fiber, number of mass points, fiber dispersion and dispersion-management schemes. If dispersion is not managed, the effects of pulse broadening must be carefully considered.

D. Demodulating the Discrete Spectrum

To demodulate a multisoliton pulse, the eigenproblem (6) needs to be solved. There is limited work in the mathematical literature concerning the numerical solution of the Zakharov-Shabat spectral problem (6). In Part II, we have studied methods by which the nonlinear Fourier transform of a signal may be computed numerically. In particular, in this paper we use the layer-peeling and Ablowitz-Ladik methods described.
in Part [II] to estimate the discrete spectrum. The reader is referred to [Part II] for the details.

**IV. STATISTICS OF THE SPECTRAL DATA**

In this section we generalize the deterministic model considered so far to include the effects of amplified spontaneous emission (ASE) noise during signal propagation. We present a method to approximate the statistics of the spectral data at the receiver.

**Remark 3.** In Section [I-B] we identified inter-channel interference in multiuser WDM networks as the intractable factor limiting the achievable rates of the current methods at high launch powers. In comparison with the effect of inter-channel interference on achievable rate, noise is of secondary importance, and hence we do not intend to provide here a comprehensive analysis.

The generalized NLS equation that includes an additive space-time noise term is generally not integrable\(^2\) Furthermore, the addition of noise disturbs the vanishing or periodic boundary conditions usually assumed in the development of the nonlinear Fourier transform. One may therefore question whether the NFT is in fact well defined in this case. Fortunately, since the ASE noise power in optical fibers is quite small compared to the signal power for SNR > 10 dB, one can treat the noise as a small perturbation, and still safely use the NFT.

Calculation of the exact statistics of the spectral data at the receiver can be quite cumbersome. This is essentially because the NLS equation with additive noise, unlike the noise-free equation, has little or no structure, giving rise to complicated variational representations for the noise statistics. Even if exact expressions could be obtained, it is unlikely that they would be suitably tractable for data communications studies. One can, however, approximate these statistics using a perturbation theory, or simulate them on a computer. In this paper we follow a perturbation theory approach.

**Remark 4.** Note that in this paper, we have not included the effects of fiber loss in our model. This assumption is justified in systems using distributed (Raman) amplification (DRA), which essentially trades loss for noise, which is treated in this section.

Noise can be introduced in a lumped or distributed fashion. The former case arises in systems using erbium-doped fiber amplifiers located at the end of each fiber span. In this case, the randomness is on initial conditions at the input of each fiber span, and the NFT can be used without approximation. We refer to this type of noise (where the randomness comes from the initial conditions and not the medium) as lumped noise.

If noise is injected continuously throughout the fiber as a result of DRA, one has distributed noise. Here we can discretize the fiber into a large number of small fiber segments and add lumped noise at the end of each segment. Each such injection of noise acts as a random perturbation of the initial data at the input of the next segment. The DRA can thus be approximately treated similar to the lumped noise case. Indeed, when the stochastic NLS equation is simulated numerically, the step size in the split-step Fourier method can be as large as the distance between erbium-doped fiber amplifiers (EDFAs), i.e., the dynamics of the pulse propagation under DRA and a large number of EDFAs is about the same.

In this section, we study the effect of the two types of perturbations on the NFDM channel model. We assume that the noise vanishes, or is negligible, as \(|t| \to \infty\) and has a finite energy such that the signal remains absolutely integrable almost surely.

**A. Perturbation of Eigenvalues**

1) **Lumped Noise:** The NFT arises in the spectral analysis of the \(L\) operator. We can easily analyze the perturbations of the eigenvalues \(\lambda\) as a result of the changes in the signal \(q(t)\).

Let us denote the nonlinear Fourier transform of \(q(t)\) in the absence of noise by \((\hat{q}(\lambda), \hat{q}(\lambda_j))\). As the signal \(q(t)\) at the input of each small segment is perturbed to \(q(t) + \epsilon u(t)\) for some small parameter \(\epsilon\) and (normalized) noise process \(n(t)\), (the discrete) eigenvalues and spectral amplitudes deviate slightly from their nominal values. Separating the signal and noise terms, the perturbed \(v\) and \(\lambda\) satisfy

\[
(L + \epsilon R)v = \lambda v, \quad R = \begin{pmatrix} 0 & n^* \\ -n & 0 \end{pmatrix},
\]

(17)
where $R$ is the matrix containing the noise. The study of the nonlinear Fourier transform in the presence of (small) input noise is thus a perturbation theory of the non-self-adjoint operator $L + \epsilon R$.

Perturbation theory of Hermitian operators is well-studied (e.g., in quantum mechanics). The Zakharov-Shabat operator in (17) is however non-self-adjoint. Unfortunately most useful properties of self-adjoint operators (in particular, the existence of a complete orthonormal basis from eigenvectors) do not carry over to non-self-adjoint operators. For either type of operator, deterministic perturbation analysis already exists in the literature [21]–[24]. These results, however, are non-stochastic and the scattering data is still lacking. A very interesting work is [25] in which authors calculate the distribution of the spectral data for the special case in which the channel is noise-free and the input is a white Gaussian stochastic process. There is also numerous work pertaining to the statistics of the parameters of a single soliton; see e.g., [26] and references therein.

For the non-self-adjoint operators $L$, the orthogonality that we require is between the space of left and right eigenvectors of $L$ associated with distinct eigenvalues; that is to say, between eigenvectors of $L$ associated with $\mu$ and eigenvalues of the adjoint operator $L^*$ associated with $\lambda \neq \lambda^*$. Let us equip the space of eigenvectors with the usual $L^2$ inner product

$$\langle u, v \rangle = \int_{-\infty}^{\infty} (u_1 v_1^* + u_2 v_2^*) \, dt.$$  

It can be verified that the operator $\Sigma_3 L$ is self-adjoint, i.e., $\langle u, \Sigma_3 L v \rangle = \langle \Sigma_3 L u, v \rangle$, where $\Sigma_3 = \text{diag}(1, -1)$ is the Pauli matrix.

We use a small noise approximation, expanding unknown variables in noise level $\epsilon$ as

$$v(t) = v^{(0)}(t) + \epsilon v^{(1)}(t) + \epsilon^2 v^{(2)}(t) + \cdots,$$
$$\lambda = \lambda^{(0)} + \epsilon \lambda^{(1)} + \epsilon^2 \lambda^{(2)} + \cdots.$$  

We assume these variables are analytic functions of $\epsilon$ so that the above series are convergent. Plugging (18) and equating like powers of $\epsilon$, we obtain

$$L v^{(0)} = \lambda^{(0)} v^{(0)},$$

$$(L - \lambda^{(0)}) v^{(1)} = -(R - \lambda^{(1)}) v^{(0)},$$

$$(L - \lambda^{(0)}) v^{(2)} = -(R - \lambda^{(1)}) v^{(1)} + \lambda^{(2)} v^{(0)},$$

and so on.

The first term implies that $v^{(0)}$ and $\lambda^{(0)}$ are eigenvalue and eigenvector of the (nominal) operator $L$. To eliminate $v_1$ from the second equation, we take the inner product on both sides of (20) with some vector $u$; the left hand side of the resulting expression is

$$\langle u, (L - \lambda^{(0)}) v^{(1)} \rangle = \langle (L - \lambda^{(0)})^* u, v^{(1)} \rangle = \langle (L^* - \lambda^{(0)*}) u, v^{(1)} \rangle.$$  

To have the right-hand side of (21) vanish, we can choose $u$ to be an eigenvector of the adjoint operator $L^*$ associated with an eigenvalue $\mu = \lambda^{(0)*}$, i.e., $(L^* - \lambda^{(0)*}) u = 0$. Since $L^*(q) = L(-q)$, if $L v = \lambda^{(0)} v$, it can be verified that $L^* u = \lambda^{(0)} u$ for $u = [v_1, -v_2] = \Sigma_3 v$. Setting $u = u^{(0)} = \Sigma_3 v^{(0)}(t, \lambda^*)$

$$\lambda^{(1)} = \frac{\langle u^{(0)}, R v^{(0)} \rangle}{\langle u^{(0)}, v^{(0)} \rangle}.$$  

Using similar calculations we obtain $\lambda^{(2)}$

$$\lambda^{(2)} = \frac{\langle u, R v^{(1)} \rangle}{\langle u, v^{(0)} \rangle} - \lambda^{(1)} \frac{\langle u, v^{(1)} \rangle}{\langle u, v^{(0)} \rangle},$$

and so on.

To summarize, the fluctuations of a discrete eigenvalue $\lambda_n$ is given by

$$\hat{\lambda}_n = \lambda_n + \epsilon \frac{\langle u_n, R v_n \rangle}{\langle u_n, v_n \rangle} + O(\epsilon^2), \quad n = 1, 2, \ldots, N,$$

where $\hat{\lambda}_n$ denotes the eigenvalue after noise addition. It follows that the perturbation of the eigenvalues is distributed, to the first order, according to a zero-mean complex Gaussian distribution.

Continuing this approach to find higher-order fluctuations of eigenvectors, $u^{(k)}$, $k \geq 1$, is not straightforward because the underlying operator is not self-adjoint.

2) Distributed Noise: Consider now the perturbed NLS equation

$$j q_z = q_{tt} + 2|q|^2 q + \epsilon n(t, z),$$  

where $\epsilon$ is a small parameter (noise level), and $n(t, z)$ represents the combined effects of the signal loss and the distributed noise.

Let us represent (22) with the same $L$ and $M$ of the noise-free equation and now let $\lambda$ vary with $z$. The equality of mixed derivatives $v_{tz} = v_{zt}$ gives

$$-j \lambda z_q - j q_{zt} + 2j |q|^2 q = 0.$$  

This, upon re-arranging and using (22), simplifies to

$$\lambda z v = \epsilon R v, \quad R = -R.$$  

Note that, as before, we do not have $v(t, z)$ a priori because, according to (5), it depends on the noisy signal $q(t, z)$ and $\lambda(z)$, both of which are unknown. However, if the noise level $\epsilon$ is small, we can expand $v(t, z)$ and $\lambda$ in powers of $\epsilon$ as

$$v(t, z) = \langle \Sigma_3 v(t, \lambda^*) v(t, \lambda^*) \rangle,$$

and equating like powers of $\epsilon$, we obtain

$$\lambda(t, z) = \frac{\langle u^{(0)}, R v^{(0)} \rangle}{\langle u^{(0)}, v^{(0)} \rangle}.$$  

It follows from (23) that the distribution of the deviation of the eigenvalues is approximately a zero-mean conditionally Gaussian random variable. The variance of this random variable is signal dependent, and although eigenvectors of an $N$-soliton can be represented as a series from Darboux transform, it is best calculated numerically if $N \geq 2$.

Example 2. Consider the single soliton of Example 1. It can be verified that $\langle \Sigma_3 v(t, \lambda^*) v(t, \lambda^*) \rangle = \frac{\pi}{2} (1 + \frac{2}{q_0^2}) = \frac{1}{2}$, where
we assumed $|\tilde{q}| = \omega$ so that the soliton is centered at $t_0 = 0$. Furthermore,

$$
\langle v(t, \lambda^*), \overline{Rv(t, \lambda)} \rangle = -\int (v_1(t, \lambda^*)v_1^*(t, \lambda)n^* + v_2(t, \lambda^*)v_2^*(t, \lambda)n) \, dt
$$

$$
= -\int \frac{1}{2} \text{sech}(\omega t)(v_2(t, \lambda^*) + v_2^*(t, \lambda)) \, \overline{\Re dt}
$$

$$
+j \int \frac{1}{2} \text{sech}(\omega t)(v_2(t, \lambda^*) - v_2^*(t, \lambda)) \, \overline{\Im dt}
$$

$$
= -\frac{1}{2} \int \text{sech}(\omega t) \overline{\Re dt} - \frac{j}{2} \int \text{sech}(\omega t) \text{tanh}(\omega t) \overline{\Im dt},
$$

where $\overline{n} = n \exp(j\Phi)$ has the same statistics as $n$. It follows that

$$
\alpha_z = -\int \text{sech}(\tau) \overline{\Re dt},
$$

$$
\omega_z = -\int \text{sech}(\tau) \text{tanh}(\tau) \overline{\Im dt}. \tag{25}
$$

The Gordon-Haus effect easily follows from the $\alpha_z$ equation. Note that in fiber optics noise is added to the signal, i.e., $n$ in this subsection should be replaced with $jn$.

### B. Perturbation of Spectral Amplitudes

Using a similar perturbation approach, we can study the influence of noise on spectral amplitudes as well.

As reported in [Part II], discrete spectral amplitudes are chaotic even when noise is zero. We thus here consider first-order fluctuation of continuous spectral amplitudes, under the lumped noise model.

Continuous spectral amplitudes are obtained by solving the following Riccati equation [Part I]

$$
\frac{dy(t, \lambda)}{dt} + (\bar{q}(t) + \epsilon n(t)) y^2(t, \lambda) + (\bar{q}^*(t) + \epsilon n^*(t)) = 0,
$$

$$
y(-\infty, \lambda) = 0, \tag{26}
$$

where

$$
\bar{q}(t) = q(t) \exp(2j\lambda t) \text{ and } \bar{q}^*(t) = \lim_{t \to \infty} e^{2j\lambda t}y(t, \lambda).
$$

One can write a Fokker-Planck equation for $y(t, \lambda)$ in $[26, 13]$. If the signal $q(t)$ is known, e.g., $q(t) = 0$ as in $[25]$, the resulting equation can be solved. In general, however, we can expand

$$
y(t, \lambda) = y_0(t, \lambda) + \epsilon y_1(t, \lambda) + \epsilon^2 y_2(t, \lambda) + \cdots,
$$

$$
\rho(\lambda) = \rho_0(\lambda) + \epsilon \rho_1(\lambda) + \epsilon^2 \rho_2(\lambda) + \cdots,
$$

and equate like powers of $\epsilon$. We obtain that $\rho_0(\lambda)$ is the spectral amplitude when noise is zero. Define

$$
G(t, \lambda) = \int_{-\infty}^{t} \bar{q}(\tau, \lambda)y_0(\tau, \lambda) \, d\tau,
$$

and $\hat{n}(t, \lambda) = n(t)y_0^2(t, \lambda) + n^*(t)$. Then

$$
\rho_1(\lambda) = -e^{-jG(\infty, \lambda)} \int_{-\infty}^{\infty} \hat{n}(\tau, \lambda)e^{jG(\tau, \lambda)} \, d\tau.
$$

This is a conditionally Gaussian random variable (for each $\lambda$) with variance

$$
E|\rho_1(\lambda)|^2 = -e^{-2jG(\infty, \lambda)} \int_{-\infty}^{\infty} |y_0(\tau, \lambda)|^4 e^{-2jG(\tau, \lambda)} \, d\tau,
$$

where we assumed that noise is delta-correlated. If for some signals, $E|\rho_1(\lambda)|^2$ is unbounded, the above perturbation expansion fails and a slow-scale variable $T = ct$ needs to be introduced.

### V. SOME ACHIEVABLE SPECTRAL EFFICIENCIES USING NFT

We turn now to a numerical study of $N$-soliton data transmission schemes, providing simulation results to quantify the gains that are achievable.

We start off with classical on-off keying soliton transmission, in which, in any symbol period $T_s$, either zero or a fundamental soliton is sent. To improve upon the classical binary soliton transmission system, we consider an $(N \geq 2)$-soliton system, occupying the same time interval as the on-off keyed soliton system, and maintaining the same bandwidth requirements. The location of the eigenvalues and the values of the discrete spectral amplitudes can be jointly modulated for this purpose. We shall see that the effective useful region in the upper-half plane to exploit the potential of discrete eigenvalue modulation is limited by a variety of factors. In this paper, we will provide only a brief discussion of the possibility of modulating the continuous spectrum.

Modulating the nonlinear spectrum generates pulses with variable width, power, and bandwidth. We take the average time, average power, and the maximum bandwidth to properly convert bits/symbol to bits/s and bits/s/Hz. To improve upon the on-off keying solitons, first we continuously modulate one eigenvalue in a given region (i.e., a classical soliton with varying amplitude, width and phase). We next consider multisoliton systems with a number of constellations on eigenvalues and discrete spectral amplitudes.

Throughout this section, we consider a 2000 km single-mode single-channel optical link in which fiber loss is perfectly compensated in a distributed manner using Raman amplification. Fiber parameters are given in Table[I]. Dispersion compensation is not applied, as it is an advantage of the NFT approach that no dispersion management or nonlinearity compensation is required. We let pulses interact naturally, as atoms in a molecule, and perform signal processing at the receiver on these groups. The method however works for dispersion managed fibers as well, and in general with any operation that does not change integrability.

### A. Spectral Efficiency of 1-Soliton Systems

Soliton transmission systems typically do not have high spectral efficiencies. This is because the amplitude and the width of a single soliton are inversely related, and hence they require a lot of time or bandwidth per degree of freedom provided. Errors in a soliton transmission system occur either because of the Gordon-Haus timing jitter effect (which is the primary source of the errors, if not managed) [3] or amplitude
(energy) fluctuations. It follows from Galilean invariance \(5\) that the Gordon-Haus effect exists for all kinds of pulses to the same extent and is not specific to solitons. This classical effect can be reduced with the help of suitably designed filters. We do not treat Gordon-Haus jitter analytically; however, system simulation do naturally include this effect.

Let us first consider a classical soliton system with only one eigenvalue \(\lambda = (\alpha + j\omega)/2\). The joint density \(f_{\lambda,\Omega}(\alpha, \omega)\) at any fixed distance \(z\) can be obtained from \((24)–(25)\) (or by extracting the dynamics of \(\alpha\) and \(\omega\) from the stochastic NLS equation, resulting in a pair of coupled stochastic ordinary differential equations).

Multiplying the stochastic NLS equation \((22)\) by \(q^*\), subtracting from its conjugate, integrating over time, and using integration by parts in the dispersion term, we get

\[
\frac{\partial E}{\partial z} = 2T \int_{-\infty}^{\infty} q(\tau, z)Z(\tau, z)d\tau,
\]

where \(E(z) = \int |q(\tau, z)|^2d\tau\) is the energy, and \(Z = -n^*\) is a noise process similar to \(r\). Thus energy fluctuation is a signal-dependent Gaussian random variable \(E(z) \sim N_{\mathbb{R}}(E(0), \sigma^2 \int E(z)d\tau) \approx N_{\mathbb{R}}(E(0), \sigma^2 zE(0))\) \((E(0) \gg \sigma^2)\). Ignoring the energy of the continuous spectrum, we have \(E \approx 2\omega\) and therefore

\[
\omega(z) = \omega(0) + \sigma \sqrt{\frac{\omega(0)}{2}} N_{\mathbb{R}}(0, 1).
\]

The conditional probability density function (PDF) \(f_{\Omega|\Omega_0}(\omega|\omega_0)\) is

\[
\tilde{f}(\omega|\omega_0) = \frac{1}{\sqrt{\pi} \sigma^2 z\omega_0} e^{-\frac{(\omega-\omega_0)^2}{\sigma^2 z\omega_0}},
\]

and the PDF of \(r = \sqrt{\omega(z)}\) is approximately a Rician distribution

\[
f(r|r_0) = \frac{r}{\sigma^2} e^{-\frac{r^2 + r_0^2}{2\sigma^2}} I_0\left(\frac{rr_0}{\sigma^2}\right) \\
\approx \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(r-r_0)^2}{2\sigma^2}}, \quad r, r_0 \geq \sigma,
\]

which is signal-independent in the high SNR regime.

Note that we are interpreting the stochastic NLS equation as a soliton-bearing system in the small-noise limit. A soliton of the deterministic NLS equation launched into a system described by the stochastic NLS equation would, of course, have a growing continuous spectrum too; in addition, there would be a very small chance of creating additional solitons out of noise at some distance, or the soliton spectrum might collapse into the real axis in the \(\lambda\) plane. All these effects are negligible if the energy of the launched soliton is large enough and the propagation distance is not exceedingly long.

In \([13]\) we have shown that a half-Gaussian density

\[
f(\omega_0) = \frac{1}{\sqrt{2\pi} P} e^{-\frac{\omega_0^2}{2P}}, \quad \omega_0 \geq 0,
\]

gives the asymptotic capacity for \((28)\)

\[
C \sim \frac{1}{2} \log(1 + \text{SNR}) - \frac{1}{2},
\]

where \(\text{SNR} = \frac{P}{\pi^2}\).

Translating capacity in bits/symbol to spectral efficiency in bits/Hz depends on the receiver architecture. Assuming that the receiver is able to decode pulses with variable widths, the spectral efficiency \(\rho(P)\) is obtained by

\[
\rho(P) = 1 \frac{1}{\omega_0 \in S} \frac{1}{\omega_0} I(\omega; \omega_0),
\]

where \(T(\omega_0)\) and \(P(\omega_0)\) are the width and the power of a single soliton with amplitude \(\omega_0\) and \(\text{BW}(S) = \max_{\omega_0 \in S} \text{BW}(\omega_0)\) is the maximum passband bandwidth that the signal set \(S\) requires for transmission (set by the min \(S\)). For a one soliton signal, we have approximately

\[
T(\omega_0) = \frac{7}{\omega_0}, \quad P(\omega_0) = \frac{\omega_0^2}{6.2}, \quad \text{BW}(\omega_0) = 0.95\omega_0,
\]

where the width \(T(\omega_0)\) includes a guard time—four times the full width at half maximum power (FWHM)—so as to minimize the intra-channel interactions.

Using \((30)\), the maximum spectral efficiency of a baseline on-off keying system is obtained to be about \(\rho \approx 0.15\) bits/Hz at the average power \(P_0 = 0.16\) mW. Note that the per unit cost capacity problem \((29)\) is non-convex and hence finding the global optimum may prove to be challenging. Here we simply optimize mutual information and scale it by \(\text{BW}(S) \times ET(\omega_0)\) evaluated at the mutual-information-maximizing input distribution.

Fig.\(\text{7}\) shows the achievable rate and the spectral efficiency of a 1-soliton system with amplitude modulation using various detection methods. Note that since we do not solve the optimization problem \((29)\), the spectral efficiencies shown in the Fig.\(\text{7b}\) are only lower bounds on the actual achievable values. Fig.\(\text{7c}\)–(d) shows the constellation at the transmitter and the “noise balls” (of radius equal to one standard deviation of the distance to the transmitted point) accumulated over 30,000 simulation trials at the receiver. The actual number of signal levels is 64 in the simulations. Calculation of the approximate rate is performed using the Arimoto-Blahut algorithm and is confirmed by numerical interior point optimization.

Note that, as is clear from Fig.\(\text{7a}\)–(b), sampling signals at \(t = 0\) is clearly a bad idea; it is shown here just to see the effects of the timing jitter on 1-soliton systems.

B. Spectral Efficiency of 2-Soliton Systems

To illustrate how the NFT method works, we start off with two simple examples. These two examples are intended to explain the details of transmission and detection using the NFT, but they have not been optimized for performance.

1) Modulating Eigenvalues: Consider the following signal set with 4 elements:

\[
\begin{align*}
S_1 & : 0, \\
S_2 & : \tilde{q}(0.5j) = 1, \\
S_3 & : \tilde{q}(0.25j) = 0.5, \\
S_4 & : \tilde{q}(0.25j, 0.5j) = (1, 1).
\end{align*}
\]

where \(\tilde{q}(0.25j, 0.5j) = (1, 1)\).
Table IV shows the energy, duration, power, and the bandwidth of these signals.

We compare this with a standard on-off keying (OOK) soliton transmission system, consisting of $S_1$ and $S_2$. From the signal parameters given in Table IV it follows that the OOK system provides about $\rho_0 = 0.33$ bits/s/Hz spectral efficiency at $P_0 = 0.1876$ mW and $R_0 = 7.42$ Gbits/s data rate. Note that the noise level is so small compared to the imaginary part of the eigenvalues that this scheme essentially achieves a transmission rate of 2 bits/symbol.

The full constellation defined in (31) has average power $0.46P_0$ and average time duration $1.65T_1$, where $P_0$ and $T_1$ are the power and the time duration of the fundamental soliton. The new signal set therefore provides a spectral efficiency of about $\frac{\log 4}{1.65T_0} = 0.2121 \times \rho_0$ bits/s/Hz and operates at $R = 1.2121 \times R_0$ for about the same average power ($0.5P_0$). Note that without $S_1$ the average power would be higher and in addition the improvement in the spectral efficiency would be slightly smaller compared to the on-off keying system. Signal $S_4$ is the new signal (a 2-soliton) that goes beyond conventional methods. Such signals do not cost much in terms of time × maximum bandwidth product, while they add additional elements to the signal set. These additional signals can generally be best decoded with the help of the nonlinear Fourier transform.

In this example, the receiver needs to estimate the pulse-duration. This can be done in many ways, e.g., using the NFT computations already performed: zeros of the signal in time can be detected when $|v_2^2 e^{j\lambda t}, v_3^2 e^{-j\lambda t}|$ reaches a constant value in steady state. This can be checked at times $t = T_1$, $t = 2T_1$ and $t = 2.58T_1$. If one of the signals is zero at the end of another signal, one can monitor the energy of the continuous spectrum to make sure that it is small. If the symbol duration is fixed to be the maximum 2.58$T_1$, the addition of $S_3$ and $S_4$ increases both time interval and cardinality of signal set such that the spectral efficiency and data rate remain constant $\log(6)/2.58$, while operating at 77% of the on-off keying signal power.

Since solitons with purely imaginary eigenvalues do not suffer from major temporal or spectral broadening, spectral efficiencies at the fiber output are essentially the same as those at the input to the fiber.

2) Modulating Both Eigenvalues and Spectral Amplitudes: We can improve upon the previous example by also modulating the spectral amplitudes. Consider the following signal

![Diagram](image-url)

Table IV shows the energy, duration, power, and the bandwidth of these signals.

| Signal | Energy | Duration FWHM | 99% Duration | Power | Bandwidth |
|--------|--------|---------------|--------------|-------|-----------|
| $S_1$  | 0      | $T_0$         | $T_1$        | 0     | $W_0$     |
| $S_2$  | $E_0$  | $T_0$         | $T_1$        | $P_0$ | $W_0$     |
| $S_3$  | 0.5 $E_0$ | 2$T_0$       | 2.58$T_1$    | 0.25$P_0$ | 0.5$W_0$ |
| $S_4$  | 1.5 $E_0$ | 4.25$T_0$    | 2.58$T_1$    | 0.58$P_0$ | 0.5$W_0$ |

**TABLE IV**

Parameters of the signal set in Section V-B1. Here $E_0 = 4 \times 0.5 = 2, T_0 = 1.763$ at FWHM power, $T_1 = 5.2637$ (99% energy), $P_0 = 0.38$ and $W_0 = 0.5714$. The scale parameters are $T_0' = 25.246$ ps and $P_0' = 0.5$ mW at dispersion 0.5 ps/(nm·km).
provides about log 16 benchmark on-off keying system. Therefore the new signal set $P$ values for spectral amplitudes provides up to and having a large peak to average power ratio, large (99%) simple and is able to decode NFT signals rather efficiently.

We make a 3-PAM constellation on $\tilde{q}_i \in \{0.5, 1, 1.5\}$. This creates a signal set with 16 elements. Here pulses are extended uniformly chosen in the interval $[0, T_1]$. The pulse width $T_1$ is chosen to be the maximum $3T_1$, where $P_0$ and $T_1$ are the power and the symbol-duration of the benchmark on-off keying system. Therefore the new signal set provides about $N \log_2 16 = 4$ bits/s/Hz and operates at $R = 1.79 \times P_0$ bits/s/Hz and operates at $R = 1.79 \times R_0$ for about the same average power. If we fix symbol durations to be the maximum $3T_1$, then the improvement is $\rho = 2.2P_0 = 0.73$ bits/s/Hz, at 80% of the average power.

Again, since the real part of the eigenvalues is not modulated, signals do not suffer from major temporal or spectral broadening.

Remark 5. Note that modulating the eigenvalues includes only the amplitude information (similarly to $M$-ary frequency-shift-keying). To excite the other half of the degrees-of-freedom representing the phase, discrete spectral amplitudes should also be considered. While $|\tilde{q}(\lambda_j)|$ may be noisy, the phase $\angle \tilde{q}(\lambda_j)$ or a function of $\{\tilde{q}(\lambda_j)\}_{j=1}^N$ can be investigated for this purpose. Note however that the asymptotic behavior of the spectral efficiency can be obtained, within a constant factor, by considering the eigenvalues alone.

C. Spectral Efficiency of $N$-Soliton Systems, $N \geq 3$

To achieve greater spectral efficiencies, a dense constellation in the upper-half complex plane needs to be considered. A spectral constellation with $n$ possible eigenvalues in $\mathbb{C}^+$ (from which $k$ eigenvalues are chosen, $0 \leq k \leq n$) and $m$ possible values for spectral amplitudes provides up to

$$\log \left( \sum_{k=0}^{n} \binom{n}{k} m^k \right) = n \log (m + 1),$$

bits per symbol (fewer if a subset is chosen). One can continue the approach presented in the previous examples by increasing $n$ and $m$. The receiver architecture presented in [Part I] is fairly simple and is able to decode NFT signals rather efficiently.

Some choices of spectral parameters may translate to pulses having a large peak to average power ratio, large (99%) bandwidth, or large (99%) time duration at $z = 0$ or during their propagation; hence, the signal set should be expurgated to avoid such undesirable signals. We have not yet found rules for modulating the spectrum so that such undesirable signals are not generated. For the small examples given here, we can check pulse properties directly; however, appropriate design criteria for the spectral data (particularly the discrete spectral amplitudes) should be developed.

In this simulation, we assume a constellation with 30 points uniformly chosen in the interval $0 \leq \lambda \leq 2$ on the imaginary axis and create all $N$-solitons, $1 \leq N \leq 6$. We then prune signals with undesirable bandwidth or duration from this large signal set. The remaining multi-solitons are used as carriers of data in the typical fiber system considered earlier. Here a spectral efficiency of 1.5 bits/s/Hz is achieved. For this calculation, we take the maximum pulse width (containing 99% of the signal energy) and the maximum bandwidth of the signal set. By increasing $n$ and $m$, pulse widths get large and the shift of the signal energy from the symbol period, due to the Gordon-Haus effect, becomes less significant. Gordon-Haus effect for solitons is as important as it is for sinc function transmission and backpropagation.

We would note that the spectral efficiency reported here was achieved using only a rather simplistic design approach. We believe that a more sophisticated search over the design space, in particular exploiting the possibility of more cleverly modulating spectral amplitudes and phase and choosing eigenvalues in a region not limited to the $j\omega$ axis, is likely to yield significantly higher spectral efficiencies.

The recent work of [28] also describes optical transmission schemes based on N-soliton transmission and the inverse scattering transform, with reports on some achievable spectral efficiencies.

D. Spectral Efficiencies Achievable by Modulating the Continuous Spectrum

In addition to the discrete spectrum, the continuous spectrum can, in some cases, also be modulated.

Here we consider the modulation of classical raised-cosine pulses, with 50% excess bandwidth. The continuous spectrum of an isolated raised-cosine pulse is purely continuous at low amplitudes, resembling its ordinary Fourier transform. We modulated the amplitude of an isolated raised-cosine pulse, propagated the pulse over a fiber channel, and estimated the continuous spectrum of the received signal. The received spectrum was then compared with the spectrum of all possible transmitted waveforms at the transmitter using the log-Euclidean distance

$$d(\hat{q}_2(\lambda), \hat{q}_1(\lambda)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \log \left( 1 + |\hat{q}_2(\lambda) - \hat{q}_1(\lambda)|^2 \right) d\lambda.$$

Fig. 8(a) shows the estimated rates achievable by a typical single-channel fiber-optic system, comparing detection after filtering, backpropagation, and matched-filtering, with detection using the nonlinear Fourier transform. A multi-ring QAM modulation was used, with rings at 16 distinct amplitude values and 32 different phase values per ring. The complex plane was quantized into the Voronoi regions corresponding to the ring constellation, to induce a discrete channel, whose capacity was computed via the Blahut-Arimoto algorithm. The NFT is calculated at either 64 or 1024 uniformly spaced discrete points on the real axis over a range containing most of the pulse energy. The 1024-point NFT is compared with a 1024-point FFT implementation of the Split-Step Fourier Method in backpropagation. As can be seen, the NFT and backpropagation methods achieve approximately the same rates. The slight improvement in the NFT method can be attributed to the stability of the continuous spectral data.
A potential gain is achievable in fiber-optic systems by employing methods, such as NFDM, which are less prone to inter-channel interference.

Recall that in NFDM the real and imaginary axes of the complex plane correspond, approximately, to the signal frequency and energy. To use NFDM as a multiuser communication method, we can partition the complex plane into disjoint regions, e.g., vertical bins. Each bin can be assigned to a user and can contain one or more degrees-of-freedom. To multiplex user signals, traditional ROADMs must be replaced with nonlinear add-drop multiplexers (NADMs) which function according to the NFT. In principle, each NADM would compute the spectrum of its input signals and filter the signals to be dropped in C^+ or C^-. It would then place the spectrum of the signals to be added in empty bands and produce the output signal by taking the inverse NFT. In this way, each user is assigned a region in the complex plane and does not interfere with other users (at least in the absence of noise—see Section VI-B).

B. Noise in the Spectral Coordinates

In a nonlinear interference channel, the interference can have two components. The first, termed “deterministic interference,” arises from the (deterministic) interaction of the signals of other users with the user of interest, and in general is present even in the absence of noise. The second, termed “stochastic interference,” arises from the (stochastic) coupling of noise with the signals of other users, interfering with the channel of interest. Thus, noise can affect the channel of interest directly (in-band noise) and indirectly (by introducing interference). Typically deterministic interference is stronger than stochastic interference.

The NLS equation with additive noise has no known integrability structure, in the sense of possessing a set of non-interacting degrees-of-freedom. As a result, while the NFT method does not suffer from strong deterministic interference, a weak stochastic interference is expected to be present. In other words, even when users are multiplexed so that they do not interact in the absence of noise, the addition of noise can potentially re-introduce (stochastic) interference. In comparison, WDM with backpropagation is subject to both strong and weak interference.

Finally, note that an additive (in-band) white noise in the time domain has coordinates in the spectral domain which may not be independent. Such correlations should be accounted for when designing signal detectors.

C. NFDM versus OFDM

NFDM and OFDM are essentially identical in the absence of noise. However, as noted above, in cases where the presence of noise breaks the integrability structure, then, unlike OFDM, the signal degrees-of-freedom are not independent in NFDM.

Note also that, while the ordinary Fourier transform of a signal is a function of a real variable, the NFT of a signal is generally defined on the whole complex plane, i.e., nonlinear frequencies are complex-valued. Since complex frequencies in the upper-half complex plane C^+ are isolated points, this component in the support of the NFT can be modulated too. OFDM is conceptually in analogy with the continuous spectrum modulation, where only the spectral amplitudes are modulated. The discrete spectrum differs further from the continuous spectrum in that it is the frequencies in C^+ themselves that contribute to the signal energy, and not their spectral amplitudes.

The analogy between NFDM and OFDM can be better understood in fibers with negative dispersion parameter. In this case, q^* is replaced with q in (3) and the (noise-free) channel is still integrable. Here, similar to OFDM, the nonlinear frequencies are real and information is encoded only in spectral amplitudes. This case is appealing analytically and numerically, since the underlying L operator is Hermitian.

D. Advantages and Disadvantages of NFDM

a) Advantages: Some of the advantages of using NFDM were outlined in Section I. In short, all deterministic distortions (SPM, XPM, FWM, ISI and interference) are zero for all users of a multiple-user network.

b) Disadvantages:

1) NFDM critically relies on the integrability of the channel. Loss, higher-order dispersion, and other perturbations caused by filters and communication equipment not taken into account in this study contribute to deviations from integrability. There are, however, several reasons to believe that the overall channel from the transmitter to the receiver can still be close to an integrable channel:
Fig. 8. Achievable rates in (a) a single channel, and (b) a WDM optical fiber system using the nonlinear Fourier transform and backpropagation. The SNR is calculated at the system bandwidth and can be adjusted to represent the optical signal-to-noise ratio.

![Graph showing achievable rates in a single channel and a WDM system](image)

Fig. 9. Partitioning $C^+$ for multiuser communication using the NFT.

![Diagram illustrating partitioning](image)

a) Using Raman amplification, the effects of loss are minimal (indeed traded with noise perturbation).

b) The NFT is applicable as long as the system can support soliton transmission. Solitons have been implemented in practice in the presence of communication equipment (filters, multiplexers, analog-to-digital (A/D) converters, etc). This is an indicator that the overall channel is still nearly integrable.

c) Mathematically one has stability results for solitons [29]. A soliton passing through a filter might be slightly distorted, but it re-organizes its shape so as to revert back to its original shape (or to form a soliton living nearby in the complex plane).

d) Considering that the performance of the WDM method degrades asymptotically with SNR, it may be worthwhile to investigate NFDM, and identify and minimize the perturbations of integrability.

2) The nonlinear Fourier coefficients at the receiver are calculated using $O(n)$ operations per nonlinear frequency, where $n$ is the number of signal samples in time. In contrast with with the $O(n \log n)$ complexity of the FFT, the complexity of an $n$-point NFT is, at present, $O(n^2)$ (assuming a fixed number of Newton steps are needed in the case of discrete frequencies). The complexity of the transmitter can be even higher. As a result, the NFT is currently computationally difficult to implement or sometimes to simulate. It is therefore of interest to develop faster algorithms.

3) Optical or electrical signal processing of $N$-solitons and their required hardware (e.g., A/Ds) may not be as simple as those in linear systems. The NFT decoder typically requires signal samples at increments smaller than the Nyquist period. An interpolation step may be needed to find all the necessary data.

E. Achievable Rates Using NFDM

As mentioned earlier, a weak stochastic interference can be present in NFDM. As a result, although the achievable rate of NFDM is expected to be higher than that of WDM with backpropagation, it too may ultimately peak at some finite SNR, and decline at higher SNRs. We have not yet simulated the capacity at high SNRs to see when this may potentially happen. However, note that, regardless of the method of transmission, the noisy channel will fundamentally be interference-limited due to the lack of integrability.

Note that the decline of the achievable rates in WDM simulations in prior work is mostly due to the (strong) deterministic interference. For instance, capacity simulations in [4], [30] which show that the rates of the WDM method vanish at high powers even if noise is set to zero. This is not the case for multiuser NFDM whose achievable rate is unbounded when noise is absent. We expect improvements in the information rates using NFDM, due to the immunity to the (strong) deterministic interference.

F. Eigenvalue Communication

As noted in [Part I], the work of Hasegawa and Nyu on “eigenvalue communication” [16] is strongly related to the NFT-based approach taken in this paper. In [16], the authors since, for instance, the NLS equation with a generic additive potential does not have conserved quantities.
considered single-user channels, encoding information in conserved quantities, and use the inverse scattering transform (IST) as a means to decode these conserved quantities. The idea is illustrated for pulses of the form $A \text{sech}(t)$.

The use of conserved quantities in a communication channel is desirable since it facilitates deterministic signal processing at the receiver and may simplify communication design. However, fundamentally it does not offer any capacity improvement if one uses an equivalent set of non-conserved quantities (e.g., amplitude and phase). It is thus not necessary to aim at extracting and modulating quantities conserved by the channel.

In contrast, our primary motivation for introducing NFT-based methods stems from recent capacity studies and our observation, made in Section II, that inter-channel interference is the dominant rate-limiting impairment in multiuser fiber optic networks. The crucial property of the NFT is its capability to eliminate deterministic interference in multiuser networks. This observation was not made in [16]; indeed, the use of the IST as a solver, or a bridge to conserved quantities, does not provide any advantage compared to alternative signal processing methods such as digital backpropagation.

### G. The Discrete Nonlinear Fourier Transform

To implement NFDM in practice, the discrete nonlinear Fourier transform, in which time domain signal is discrete and periodic, should be implemented. The development of the discrete nonlinear Fourier transform exists in the mathematics literature, taking a similar form to the continuous one [18], [31].

### VII. Conclusions

Motivated by recent studies showing that the achievable rates of current methods in optical fiber networks vanish at high launch power due to the impact of nonlinearity, in [Part I], [Part II], and this paper we have revisited information transmission in such nonlinear systems. In these papers, we suggested using the nonlinear Fourier transform to transmit information over integrable communication channels such as the optical fiber channel, which is governed by the nonlinear Schrödinger equation. In this transmission scheme, information is encoded in the nonlinear Fourier transform of the signal, consisting of two components: a discrete and a continuous spectral function. With this new method, deterministic distortions arising from the dispersion and nonlinearity, such as inter-symbol and inter-channel interference are zero for a single user channel or all users of a multiuser network.

We took the first steps towards the design of a communication system implementing the nonlinear Fourier transform. We proposed a Darboux-transform-based algorithm for modulating the discrete spectrum at the transmitter, and we provided a first-order perturbation analysis of the influence of noise on the received spectrum. Furthermore, we provided examples illustrating how the NFT can be used for data transmission. Although these small examples clearly demonstrate improvements over their benchmark systems, more sophisticated large-scale simulations are required to demonstrate the potential to achieve high spectral-efficiencies.

Because nonlinearity is a key feature of fiber-optic networks, the development of nonlinearity-compatible transmission schemes, like those based on the nonlinear Fourier transform, is likely to continue to be a fruitful research direction.

### Appendix A

**Solution of Hirota Equations**

Because (12) and (13) are homogeneous in the order of derivatives that occur in each term, exponential functions are candidate solutions. This suggests that $F$ and $G$ can be expanded (linearized) as

$$F(t,z) = f^0(t,z) + \epsilon f^{(1)}(t,z) + \epsilon^2 f^{(2)}(t,z) + \cdots, \quad G(t,z) = g^0(t,z) + \epsilon g^{(1)}(t,z) + \epsilon^2 g^{(2)}(t,z) + \cdots,$$

for some small parameter $\epsilon$. We then obtain

$$FG = f^0(0)g^0(0) + \epsilon \left( f^{(1)}(0)g^{(1)}(0) + f^{(2)}(0)g^{(1)}(0) + \cdots \right) + \epsilon^2 \left( f^{(1)}(0)g^{(2)}(0) + f^{(2)}(0)g^{(1)}(0) + \cdots \right) + \cdots,$$

which can then be substituted into (12)–(13). Equating like powers of $\epsilon$, one obtains bilinear equations for sub-components $f^{(i)}g^{(j)}$ using the $D$ operator. As shown below (for the NLS), this series truncates for integrable systems and exact solutions of various finite order are obtained. To begin finding unknowns recursively, we can set initially $f^{(0)} = 1$.

The zero-order term in (12) gives $g_0^2 + q_0^2 = 0$, i.e.,

$$g_0 = \omega e^{\lambda_0}, \quad \lambda_0 = (\alpha + i\omega)/2.$$

Here $\lambda_0 = \lambda + i\phi_0$. The corresponding term in (13) gives $c = 0$, and thus $g_0^2 = 0$. At this stage, we can set the higher order terms to zero and get the trivial solution $q = 0$.

The $\epsilon$ term in (12) gives $g^{(1)} = \alpha e^{X_1}$. The corresponding term in (13) gives $D^2 f^{(1)} = 0$, i.e., $f^{(1)} = 0$.

The $\epsilon^2$ term in (13) is $D^2 f^{(2)} = 2|g^{(1)}|^2$, or $f^{(2)} = 2|g^{(1)}|^2$. Trying $f^{(2)} = \alpha e^{X_1+X_1^*}$, we get

$$\alpha = (\lambda - \lambda^* - 2)/2.$$
APPENDIX B

PROOF OF THE DARBOUX THEOREM

See [19] for a more general theorem and its proof; here we give a simple proof for Theorem 1.

Let φ(t, λ; q) be a known eigenvector associated with λ and q, i.e., satisfying φt = P(λ, λ)φ. Its adjoint φ∗(t, λ) = [φ∗(t, λ)] satisfies φt(λ, λ) = P(λ, λ∗)φ(t, λ). Denote this known solution as S = [φ, φ], Γ = diag(λ, λ∗), and Σ = STS−1.

We can verify that St = JΣ + QΣ, where J = diag(j, −j) and Q = offdiag(q, −q). In addition we have Σt = [JΣ + Q, Σ].

Given that φ(t, λ; q) is known, the Darboux transformation maps \{v(t, μ; q), \tilde{v}(t, μ; q)\} to \{u(t, μ; \tilde{q}), \tilde{u}(t, μ; \tilde{q})\} according to

\[V = U Λ - Σ V,\]

where \[V = [v, \tilde{v}], U = [u, \tilde{u}], Λ = \text{diag}(μ, μ^*).\]

We have \[V_t = JV Λ + QV\]

\[U_t = V_t Λ - (Σ_t V + Σ_t V_t)\]

\[= (JVA + QV)Λ - (\{JΣ + Q, Σ\}V + Σ(JVA + QV))\]

\[= (JVA + QV)Λ - JΣJVΛ - \{JΣ + Q, Σ\}V\]

\[= J(VΛ - ΣVΛ) + ΣJVA - ΣJVA\]

\[= QΛ + \{J, Σ\}VΛ - \{JΣ + Q, Σ\}V + QVΛ\]

\[= JΛ + [J, Σ]VΛ - \{JΣ + Q, Σ\}V + QVΛ\]

\[= JΛ + [J, Σ]VΛ - ΣJΣVΛ - QΣV + QVΛ\]

\[= JΛ + [J, Σ]VΛ - QΣV + QVΛ\]

\[= JΛ + [J, Σ]U + Q(VΛ - ΣV)\]

\[= JΛ + Q + [J, Σ]U\]

\[= JΛ + QΛ + \tilde{Q}U,\]

where \[\tilde{Q} = Q + [J, Σ].\]

In the same manner we can show that u and \(\tilde{u}\) satisfy the M-equation \[v_z = M(λ, \tilde{q})\] and \[\tilde{v}_t = M(λ^*, \tilde{q}).\]

REFERENCES

[1] M. I. Yousefi and F. R. Kschischang, “Information transmission using the nonlinear Fourier transform,” Part I: Mathematical tools,” submitted to IEEE Trans. Inf. Theory, Feb. 2012.

[2] ——, “Information transmission using the nonlinear Fourier transform, Part II: Numerical methods,” submitted to IEEE Trans. Inf. Theory, Feb. 2012.

[3] L. F. Mollenauer and J. P. Gordon, Solitons in Optical Fibers: Fundamentals and Applications. Academic Press, 2006.

[4] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks,” IEEE J. Lightwave Techn., vol. 28, no. 4, pp. 662–701, Feb. 2010.

[5] M. J. Ablowitz and H. Segur, Solitons and the Inverse Scattering Transform, 1st ed. Society for Industrial and Applied Mathematics, 2000.

[6] L. D. Faddeev and L. A. Takhtajan, Hamiltonian Methods in the Theory of Solitons. Springer Verlag, 2007.

[7] A. Hasegawa and Y. Kodama, Solitons in Optical Communications. Clarendon Press Oxford, UK, 1995, vol. 320.

[8] A. C. Singer, “Signal processing and communication with solitons,” Ph.D. dissertation, Massachusetts Institute of Technology, 1996.

[9] G. P. Agrawal, Nonlinear Fiber Optics. Springer, 2006.

[10] A. El Gamal and Y. H. Kim, Network Information Theory. Cambridge University Press, 2011.

[11] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the k-user interference channel,” IEEE Trans. Inf. Theory, vol. 54, no. 8, p. 3425, 2008.

[12] E. Agrell, “The channel capacity increases with power,” CoRR, vol. abs/1108.0391, 2011.

[13] M. I. Yousefi and F. R. Kschischang, “On the per-sample capacity of nondispersive optical fibers,” IEEE Trans. Inf. Theory, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[14] B. W. Gobel, “Information-theoretic aspects of fiber-optic communication channels,” Ph.D. dissertation, Technische Universität München, 2010.

[15] A. Mecozi and R. J. Essiambre, “Nonlinear Shannon limit in pseudo-linear coherent systems,” IEEE J. Lightwave Techn., vol. 30, no. 12, pp. 2011–2024, 2012.

[16] A. Hasegawa and T. Nuy, “Eigenvalue communication,” IEEE J. Lightwave Techn., vol. 11, no. 3, pp. 395–399, 1993.

[17] R. Hirota, The Direct Method in Soliton Theory. Cambridge University Press, 2004.

[18] A. R. Osborne, Nonlinear Ocean Waves and the Inverse Scattering Transform. Academic Press, 2010.

[19] V. Matveev and M. A. Salle, Darboux Transformations and Solitons. Springer-Verlag, 1991.

[20] M. J. Ablowitz, B. Prinari, and A. D. Trubatch, Discrete and Continuous Nonlinear Schrödinger Systems. Cambridge University Press, 2003.

[21] V. I. Karpman and V. V. Solov’ev, “A perturbational approach to the two-soliton systems,” Physica D: Nonlinear Phenomena, vol. 3, no. 3, pp. 487–502, 1981.

[22] V. I. Karpman and E. M. Maslov, “A perturbation theory for the Korteweg-de Vries equation,” Phys. Lett. A, vol. 60, no. 4, pp. 307–308, 1977.

[23] D. J. Kaup, “A perturbation expansion for the Zakharov-Shabat inverse scattering transform,” SIAM J. Appl. Math., vol. 31, no. 1, pp. 121–133, 1976.

[24] V. Konotop and L. Vázquez, Nonlinear Random Waves. World Scientific Publishing Company Inc., 1994.

[25] P. Kazarinou et al., A. M. Moustakas, “Nonlinear Schrödinger equation with random Gaussian input: Distribution of inverse scattering data and eigenvalues,” Phys. Rev. E, vol. 78, no. 1, p. 016603, 2008.

[26] G. Falkovich, I. Kolokolov, V. Lebedev, M. Mezentsiev, and S. Turitsyn, “Non-Gaussian error probability in optical soliton transmission,” Physica D: Nonl. Phenom., vol. 195, no. 1, pp. 1–28, 2004.

[27] M. I. Yousefi and F. R. Kschischang, “A Fokker-Planck differential equation approach for the zero-dispersion optical fiber channel,” in 2010 IEEE Int. Symp. Info. Theory (ISIT), 2010, pp. 206–210.

[28] E. Meron, M. Feder, and M. Shaiﬁ, “On the achievable communication rates of generalized soliton transmission systems,” CoRR, vol.abs/1207.0297, 2012.

[29] T. Tao, “Why are solitons stable?” Bull. Amer. Math. Soc, vol. 46, no. 1, pp. 1–33, 2009.

[30] M. I. Yousefi, “Information transmission using the nonlinear Fourier transform,” Ph.D. dissertation, University of Toronto, Nov. 2012.

[31] Y. C. Ma and M. J. Ablowitz, “The periodic cubic Schrödinger equation,” Studies in Appl. Math., vol. 65, pp. 113–158, 1981.