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, multi-soliton 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. 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 multi-soliton 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 [3], 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 [4]. Although these approaches provide a substantial spectral efficiency at low to moderate signal-to-noise ratios (SNRs), their efficacy saturates after a finite SNR $\sim 20-30$ dB where $\rho \sim 5-9$ bits/s/Hz. This, as we shall see in Section II, 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 $A \operatorname{sech}(t)$). Signal processing problems (*e.g.*, detection and estimation) involving soliton signals in the Toda lattice and other models have been considered by Singer [8]. There is also a related work by Hasegawa and Nyu, "eigenvalue communication" [9], which is reviewed and compared with our approach in Section VI-F, 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 multi-soliton 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 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) eigenvalues and the continuous spectral amplitudes 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], [10]. The distortions arising from these interactions have deterministic and (signal-dependent) stochastic components that grow with the input signal power, diminishing the achievable rate[1] at high powers. In these studies, for a class of ring constellations, the achievable rates of the WDM method increases with average input power power, $\mathcal{P}$, reaching a peak at a certain critical

---

[1]In this paper, the term "*achievable rate*" refers to a lower bound to the capacity. It is obtained by optimizing mutual information under some assumptions, *e.g.*, considering a sub-optimal transmitter and receiver or a method of communication, a subset of all possible input distributions, etc.

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 factors limiting the achievable rates 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 β_2, nonlinearity parameter γ and length $\mathcal{L}$. After appropriate normalization (see, *e.g.*, [Part I, Section I, particularly Eq. (3)]), 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

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

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

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

where $\delta_B(x) = 2B\operatorname{sinc}(2Bx)$, where B is the normalized noise bandwidth and where E denotes the expected value. It is assumed that the transmitter is bandlimited to B and power limited to $\mathcal{P}$, *i.e.*,

$$\mathsf{E}\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} |q(t, 0)|^2 dt = \mathcal{P},$$

where $\mathcal{T} \to \infty$ is the communication time. Note that the power constraint in this paper is an equality constraint.

In fiber-optic communication systems, noise can be introduced in a lumped or distributed fashion. The former case arises in systems using erbium-doped fiber amplifiers (EDFAs) located at the end of each fiber span [3]. We refer to this type of noise as *lumped noise*. If noise is injected continuously throughout the fiber as a result of distributed Raman amplification (DRA), as in (1), one has *distributed noise* [4]. Here the fiber loss is assumed to be perfectly compensated by the amplifier. In this paper, we consider

TABLE I
FIBER PARAMETERS

n_{sp}	1.1	excess spontaneous emission factor
h	$6.626 \times 10^{-34}\,\mathrm{J} \cdot \mathrm{s}$	Planck's constant
ν	193.55 THz	center frequency
α	0.046 km^{-1}	fiber loss (0.2 dB/km)
γ	1.27 W^{-1}km^{-1}	nonlinearity parameter

the DRA model, hence (1) explicitly contains no loss term — see also Remark 5. In this model, the noise spectral density is given by $\sigma^2 = \sigma_0^2 \mathcal{L}/(P_n T_n)$, where $\sigma_0^2 = n_{\text{sp}} \alpha h\nu$, with parameters given in Table I, and where $P_n = 2/(\gamma \mathcal{L})$ and $T_n = \sqrt{|\beta_2|\mathcal{L}/2}$ are normalization scale factors. The unnormalized signal and noise bandwidth can be obtained by dividing the corresponding normalized bandwidth by T_n. A derivation of (1) and a discussion about sources of noise in fiber-optic channels can be found in [3], [11].

Mathematically, stochastic partial differential equations (PDEs) such as (1) are usually interpreted via their equivalent integral representations. Integrating a stochastic process with unbounded variation, such as white noise, can be problematic. Consider, for instance, the Riemann integral $\int_z^{z+\mathrm{d}z} g(z)\mathrm{d}B(z) \approx g(l)(B(z + \mathrm{d}z) - B(z))$, where $B(z)$ is the Wiener process and $l \in [z, z + \mathrm{d}z]$. Since the integrand is not approximately constant in $[z, z + \mathrm{d}z]$, the value of the integral depends on l. The choice of l leads to various interpretations for a stochastic PDE, notably Itō and Stratonovich representations, in which, respectively, $l = z$ and $l = z + \mathrm{d}z/2$. Fortunately, since in our application 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).

B. Achievable Rates of WDM Optical Fiber Networks

Fiber-optic communication systems often 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, where 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 reconfigurable optical add-drop multiplexer (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 optical fiber link from the transmitter to the receiver.

In WDM, the following (baseband) signal is transmitted over the channel

$$q(t, 0) = \sum_{k=0}^{N-1} \left(\sum_{\ell=1}^{M} s_k^\ell \phi_\ell(t) \right) e^{j2\pi kWt}, \qquad (2)$$

where k and l are user and time indices, $\{s_k^\ell\}_{l=1}^M$ are symbols transmitted by user k, $W = B/N$ is the per-user bandwidth, $\{\phi_\ell(t)\}_{l=1}^\infty$ is an orthonormal basis for the space of finite energy signals with Fourier spectrum in $[-W/2, W/2]$, N is the number of WDM users and M is number of symbols per user. To illustrate the essential aspects, we temporarily simplify (2) by assuming that $M = 1$ and $\phi_1(t) = 1$, $\forall t$, *i.e.*, each user sends a pure sinusoid, so as to work with Fourier series instead of Fourier integrals. Thus each user operates at a single frequency centered in a band of width W and

$$q(t, 0) = \sum_{k=0}^{N-1} q_k(0)e^{j2\pi kWt}, \qquad (3)$$

where $\{q_k(0)\}$ are the Fourier series coefficients at $z = 0$.

As the periodic signal (3) evolves in the nonlinear optical fiber, new frequency components are created and the signal may not remain periodic as in (3). However, assuming a small W, there are a large number of frequencies (users) at $z = 0$ and we can assume a Fourier series with variable coefficients for $q(t, z)$ at $z > 0$

$$q(t, z) = \sum_{k=0}^{N-1} q_k(z)e^{j2\pi kWt}. \qquad (4)$$

Substituting (4) into (1), we get the NLS equation in the discrete frequency domain

$$j\frac{\partial q_k(z)}{\partial z} = \underbrace{-4\pi^2 W^2 k^2 q_k(z)}_{\text{dispersion}} + \underbrace{2|q_k(z)|^2 q_k(z)}_{\text{SPM}}$$
$$+ \underbrace{4q_k(z) \sum_{\ell \neq k} |q_\ell(z)|^2}_{\text{XPM}}$$
$$+ \underbrace{2 \sum_{\substack{\ell \neq m \\ \ell \neq k}} q_\ell(z)q_m^*(z)q_{k+m-\ell}(z)}_{\text{FWM}} + n_k(z), \quad (5)$$

in which n_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[2].

It is important to note that the optical WDM channel is a nonlinear multiuser *interference channel* with memory [12]. The inter-channel interference terms are the XPM and FWM. There is no ISI in the assumed isolated pulse transmission model (5) with one degree-of-freedom per user. However in a pulse-train transmission model where $M > 1$, replacing $\{q_k\}$ by $\{s_k\}$ 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.

The received signal $q(t, \mathcal{L})$ associated with (2) can be projected into the space spanned by $\phi_l(t) \exp(j2\pi kWt)$, $l = 1, 2, \ldots, M'$, $k = 0, 1, \ldots, N' - 1$, for some N' and M', similar to (2). In this manner, the channel is discretized as a map from a finite number of degrees-of-freedom $s = \{s_k^\ell\}_{k,l=0,1}^{N-1,M}$ at the channel input, to their corresponding values $\hat{s} = \{\hat{s}_k^\ell\}_{k,l=0,1}^{N'-1,M'}$ at the channel output. In general $N' \neq N$ and $M' \neq M$, since signal bandwidth and duration at the transmitter and receiver might be different.

If one has a finite number of degrees-of-freedom s and $\hat{s}$, has access to all of them and joint transmission and detection

---
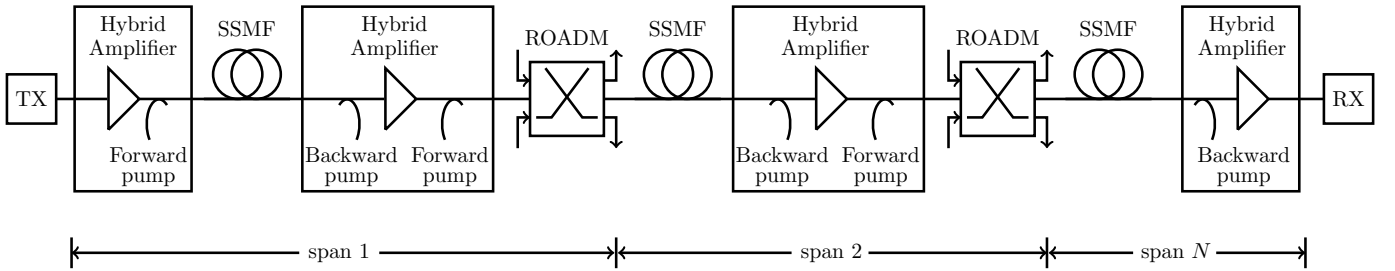
[2]Some authors define XPM differently.

Fig. 1. Fiber-optic communication system (after [4]).

of s and $\hat{s}$ is practical, the channel is essentially a single-user vector channel $s \mapsto \hat{s}$, whose capacity is non-decreasing with average input power [13]

$$\mathcal{P} = \frac{1}{NM} \sum_{k=0}^{N-1} \sum_{l=1}^{M} \mathsf{E}|s_k^\ell|^2.$$

If joint transmission and detection is not possible, *e.g.*, in frequency k or in time ℓ, then one can either treat interference as noise (as currently assumed in WDM networks), or examine various strategies to manage interference. By analogy with the linear N-user interference channel [14], [15], these strategies include signal-space orthogonalization (*e.g.*, as achieved by NFDM in the deterministic model) and, if enough information about the channel is known, interference cancellation (particularly in the strong regime) and interference alignment. If one of these interference management strategies can be successfully applied, then, again, the capacity of each user can be non-decreasing with average input power. If none of these strategies is applicable so that interference is treated as noise, or if additional constraints are present, the achievable rate of the channel of interest in WDM can saturate or decrease with average input power. Below, we clarify these cases in more detail.

It is obvious that capacity is a non-decreasing function of cost under an inequality constraint. Below, we assume an average cost defined by an equality constraint. This may not be a suitable definition from a practical point of view, but it is certainly of theoretical interest and it is also the convention in optical fiber communication.

1) Single-user Memoryless Channels: The capacity-cost function of a single-user vector discrete-time memoryless channel with input alphabet having a symbol with unbounded cost is a non-decreasing function of an equality-constrained average cost [13]. The argument of [13] goes as follows. If a rate R is achievable at cost $\mathcal{P}$ by some input distribution $p(x)$, then, for a small positive ϵ, a rate of at least $(1-\epsilon)R$ is achievable at cost $\mathcal{P}' > \mathcal{P}$ using the distribution $(1-\epsilon)p(x) + \epsilon\delta(x - x_1)$ where x_1 is a symbol of large cost. Intuitively, sending a symbol x_1 of large cost with small probability allows average cost to grow with negligible impact on the achieved rate. Since the channel is memoryless, transmission of x_1 does not affect the other symbols transmitted. Essentially the transmitter remains in a low-power state most of the time, which effectively turns the equality constraint to an inequality constraint. See [13] for details, as well as for a discussion about more general scenarios.

The monotonicity of $C(\mathcal{P})$, of course, holds true for any set of transition probabilities in a discrete-time memoryless channel, including those obtained from nonlinear channels. Consider, for instance, a nonlinear memoryless channel, *e.g.*, the continuous-time zero-dispersion optical fiber channel in the presence of a filter at the receiver. When a signal propagates in this channel, its spectrum can spread continuously. The amount of spectral broadening depends on the pulse shape and, in particular, on the signal intensity. Thus a signal with large cost may also require a large transmission bandwidth and may be filtered out by the receiver filter. However, a large-cost dummy signal does not need to be decoded. Thus, as for any discrete-time memoryless channel, the capacity (bits/symbol) is non-decreasing with the average input power $\mathcal{P}$.

Although $C(\mathcal{P})$ is monotonic, it may saturate, *i.e.*, approach a finite constant for large values of $\mathcal{P}$. In the zero-dispersion optical fiber example, nonlinearity will cause a signal-dependent spectral broadening, and large-energy signals may broaden beyond the bandwidth of the receiver filter. Thus the nonlinearity could potentially cause the capacity $C(\mathcal{P})$ to saturate; a precise analysis would depend on the definitions of bandwidth and time duration.

Of course a saturating $C(\mathcal{P})$ is a serious limitation to data communications. Firstly, from a practical standpoint, a capacity that saturates is equivalent to one that that peaks. Secondly, in many channels one may not be able to increase the average cost in the particular manner described above. For instance, it is not possible to send a symbol with arbitrarily large cost in a channel in which each symbol has a finite cost or in the presence of a peak-power constraint. Thirdly, in some cases increasing the average cost will limit the admissible input distributions and decrease the capacity[3].

2) Single-user Channels with Memory: The argument of [13] can be repeated for channels with finite memory, *i.e.*, when the influence of a large-cost symbol vanishes in a finite time interval. Whenever such a large-cost symbol is transmitted, the receiver can simply wait for the channel to settle before resuming normal operation. As before, in the limit of small ϵ, the loss in data rate is negligible. Of course, as before, saturation can occur; for example, [16] gives an example of a channel with memory where $C(\mathcal{P})$ can saturate

---

[3]As a simple example, a binary-input channel with input costs $c_0 < c_1$ can achieve a capacity $C(c_a)$ with average cost c_a in the range $c_0 \leq c_a \leq c_1$. However, since $C(c_0) = C(c_1) = 0$, $C(c_a)$ is non-monotonic. This situation can be observed in computer simulations at average powers close to the peak power.

even if optimal detection is performed.

In fiber-optic channels, 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 nonlinear intra-channel effects do not cause the capacity to vanish if the channel has finite memory (as in the previous paragraph) and such joint detection is performed.

Note that using a sub-optimal receiver in the fiber-optic channel may cause the achievable rate to saturate with input power. Consider, by way of analogy, a usual linear channel $y^m = T(s^m) + n^m$, where $s^m, y^m, n^m \in \mathbb{C}^m$ are, respectively, the input, output and noise blocks and $T : \mathbb{C}^m \mapsto \mathbb{C}^m$ is an invertible linear transformation. If T is not a multiplication (diagonal) operator, the channel is subject to ISI. Inverting the channel at the receiver completely removes this ISI: $\hat{s}_i = s_i + \hat{n}_i$, where $\hat{s} = T^{-1}y$ and $\hat{n} = T^{-1}n$, giving an additive noise channel with colored, but signal-independent noise.[4] A (suboptimal) receiver which ignores noise correlation and performs isolated symbol detection achieves a rate (lower bound to the capacity) going to infinity with average input power. In contrast, now consider a nonlinear channel $y^m = F(s^m, n^m)$, where $F : \mathbb{C}^m \times \mathbb{C}^m \mapsto \mathbb{C}^m$ is a nonlinear transformation, *i.e.*, each output component y_i is a nonlinear function of signal $s^m \in \mathbb{C}^m$ and noise $n^m \in \mathbb{C}^m$. If $F(s^m, 0)$ is invertible, channel inversion at the receiver, in general, gives $\hat{s}_i = s_i + h_i(s^m, n^m)$, for some function h_i. As a result, in nonlinear systems, channel inversion (*e.g.*, by backpropagation) leaves a residual "stochastic ISI" $h_i(s^m, n^m)$ for each symbol. This form of ISI is absent when the noise is zero. In this case, a receiver based on backpropagation and isolated symbol detection gives rise to an ISI-limited communication system with suboptimal performance. This occurs when assuming a memoryless model for the fiber-optic channel. Treating stochastic ISI as noise can lead to a bounded achievable rate. This is one reason that $I(\mathcal{P})$ has a peak in some of the prior work.

The case of channels with infinite memory can be more involved. Here sending a symbol with large cost may render the rest of transmission useless. Thus care must be taken in nonlinear channels in which the memory grows with signal.

3) Multiuser Channels with Interference: Inter-channel interference in the frequency domain is mathematically the dual of intra-channel ISI in the time domain. The difference is that 1) cooperation and joint detection is generally not possible among users, and 2) while the bandwidth is usually limited, transmission time can be practically unlimited. In an optical fiber network, many users have to share the same optical fiber link. Some user signals 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 $k' \neq k$ is unknown to the user k.

---

[4]Of course, naive channel inversion may result in noise enhancement, which we ignore for the purposes of this discussion.

In optical fiber networks, not much information is known about interference signals. The location, or the number, of ROADMs, and how many signals and with what properties have joined or left the link may be unknown. These signals may not co-originate, so cooperation and precoding among users may not be possible. In the absence of such information, it is difficult to perform techniques such as interference cancellation or alignment. In this paper, we refer to such system assumptions as the *network scenario*. The WDM simulations in the literature, and our analysis in this section, assumes a network scenario, in which the nonlinear interference is unavoidably treated as noise.

In a nonlinear channel where, by definition, additivity is not preserved under the action of the channel, multiplexing user signals in a linear fashion, *e.g.*, by adding them in time (time-division multiplexing), in frequency (WDM or traditional OFDM) or in space (multi-mode communications) leads to inter-channel interference. In the case of WDM where the available bandwidth is very limited, the interference is significant and, if treated as noise in a network scenario, ultimately limits the achievable rates of such optical fiber systems. In Section VI-G a brief discussion of other transmission techniques is given.

Among the two interference terms FWM and XPM, FWM is cubic in signal amplitude and has a larger variance. It is obvious that this interference grows rapidly when increasing the common average power $\mathcal{P}$ (or the number of users N), ultimately overwhelming the signal and limiting the achievable rate. The per-degree-of-freedom achievable rates of the channel of interest in WDM method versus average power in a network scenario, $I(\mathcal{P})$, 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], [10], [17], [18]; see also Fig. 3 and Remark 2.

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 multiuser channel, with or without memory, and with or without optical filtering. It is thus important, when comparing different results, to clarify which modeling assumptions have been made.

C. Inter-channel Interference as the Capacity Bottleneck in WDM Optical Fiber Networks

In the previous section we argued that, while intra-channel interference can be handled by signal processing and coding, inter-channel interference ultimately limits the achievable rate of optical fiber networks. The current practice in fiber-optic communication is to send a linear sum of signals in time (*e.g.*, a pulse train) and in frequency (*e.g.*, WDM) in the form of (2), the linear orthogonality of which is corrupted by the nonlinear fiber channel. This corresponds to modulating *linear-algebraic modes* in the nonlinear channel (*e.g.*, sending sinc functions). Thus we identify conventional linear multiplexing as a major culprit limiting the achievable rate of current approaches in optical fiber networks. A modification of add-drop multiplexers is needed so that the incoming signals are multiplexed in
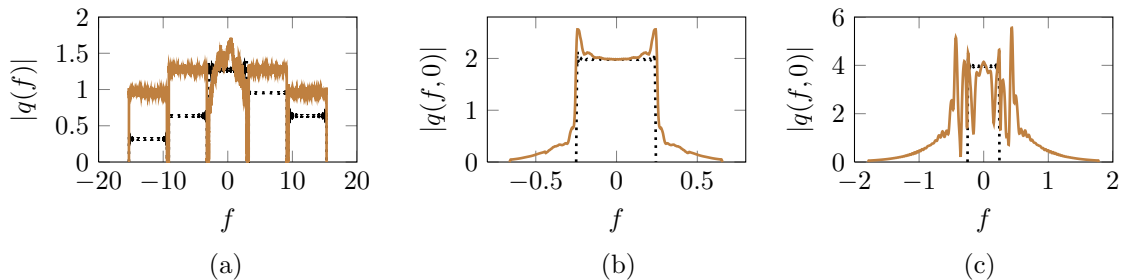
Fig. 2. (a) 5 WDM channels, with the channel of interest at the center. The dotted and solid graphs represent, respectively, the input and (noisy) output after backpropagation. Neighbor channels are dropped and added at the end of the each span in the link, creating a residual interference for the channel of interest. (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.

a nonlinear fashion, exciting non-interacting signal degrees-of-freedom under the NLS propagation. This corresponds to modulating appropriate *nonlinear modes* supported by the channel (*e.g.*, in the case of focusing regime, sending N-soliton functions).

To illustrate the effect of the inter-channel interference in the application of the WDM method to the nonlinear fiber channel, we have simulated the 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 signals in neighboring bands, with symbols chosen uniformly from a common constellation. At the channel output, the COI is filtered and backpropagated according to the inverse NLS equation. Fig. 2 compares the input and output frequency-domain waveforms, after backpropagation of the COI. Fig. 2(a) shows five random instances of the multiplexed signals; note that, because out-of-band signals are filtered and replaced at various points along the fiber, the out-of-band signals at the receiver are not related to the transmitted ones. Only the COI is backpropagated. Comparing Fig. 2(c) with Fig. 2(b), it can be seen that the nonlinear inter-channel interference is stronger at higher powers.

The simulated achievable rates of WDM are shown in Fig. 3. Here the distribution of the user of interest is optimized and interference signals correspond to independent symbols chosen uniformly from a common multi-ring constellation. The two cases of large and small inter-channel interference shown in the figure correspond to large and small user peak powers, by scaling user constellations. 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; see also [19].

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. Here by "deterministic interference" we

mean interference terms that are present even in the absence of noise. This lack of interference is a consequence of 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!). This results in a deterministic "orthogonalization" for the nonlinear optical fiber channel for any value of dispersion, nonlinearity, signal power or transmission distance.

Remark 1. Note that, as a consequence of the data-processing inequality, 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.

Remark 2. An appropriate information-theoretic framework for WDM is to describe the achievable rate region of the N-user nonlinear interference channel with memory by a joint rate $(R_1, \ldots, R_n)$. This is, however, difficult to achieve. If one isolates a single channel, the corresponding rate, R_i, would depend on the distribution of the signals of the other users (and not just their average powers). In WDM, users can operate at low powers most of the time, as prescribed in Section II-B (a), and each get a rate potentially saturating with power (even by regarding interference as noise). However, in a network scenario, interfering users may transmit data according to any distribution — including, in the extreme case, sending a symbol with power $\mathcal{P}$ all the time. The rates shown in Fig. 3, vanishing at high powers, are obtained when interfering users send data based on uniform distributions, while the distribution of the user of interest is optimized. The average power for each user was increased in the manner explained in the description of the simulation, and not as in [13] (prescribed in Section II-B (a)). As noted earlier, this need not to be elaborated since the vanishing and saturating scenarios are essentially equivalent. Some of the non-monotonic achievable rate graphs in the literature, similar to Fig. 3, given appropriate assumptions, can be interpreted as rates saturating with power at the location of the peak, by staying in a low power regime most of the time, if needed.
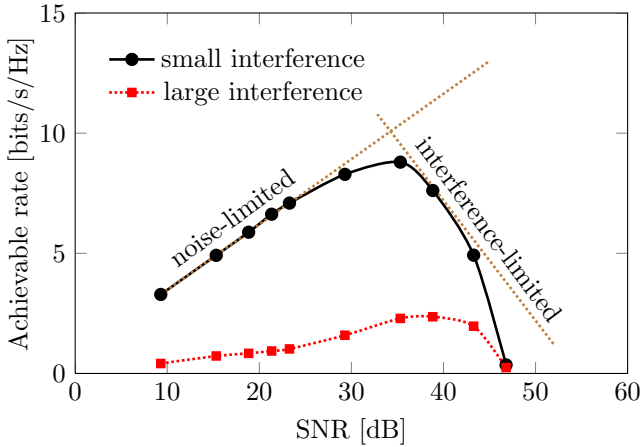
Fig. 3. Achievable rates of the WDM method in a network scenario.

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 the spectral analysis of the operator

$$L = j \begin{pmatrix} \frac{\partial}{\partial t} & -q(t) \\ -q^*(t) & -\frac{\partial}{\partial t} \end{pmatrix} = j \left(D\Sigma_3 + Q \right), \qquad (6)$$

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

$$Q = \begin{pmatrix} 0 & -q \\ -q^* & 0 \end{pmatrix} \text{ and } \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let $v(t, \lambda)$ be an eigenvector of L with eigenvalue λ. Following [Part I, Section IV], the discrete spectral function of the signal propagating according to (1) is obtained by solving the the Zakharov-Shabat eigenproblem $Lv = \lambda v$, or equivalently

$$v_t = \begin{pmatrix} -j\lambda & q(t) \\ -q^*(t) & j\lambda \end{pmatrix} v, \quad v(t \to -\infty, \lambda) \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-j\lambda t}, \quad (7)$$

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 (7) is solved from $t = -\infty$ to $t = +\infty$ to obtain $v(+\infty, \lambda)$. 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 $\mathbb{C}^+ = \{\lambda : \Im(\lambda) > 0\}$:

$$\tilde{q}(\lambda_j) = \frac{b(\lambda_j)}{a_\lambda(\lambda_j)}, \quad j = 1, \ldots, N,$$

where subscript λ denotes differentiation and λ_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 $q(t) \longleftrightarrow (\hat{q}(\lambda), \tilde{q}(\lambda_j))$. 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 [20]. 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 multi-solitons 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 — see *e.g.*, [Part I, Section VII. A-B, particularly Eqs. (30a)–(30d)]. 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 solutions are N-soliton signals.

Let $V(t, \lambda_j)$ and $\tilde{V}(t, \lambda_j^*)$ denote the scaled eigenvectors associated with λ_j and λ_j^* defined by their boundary conditions at $+\infty$ (they are denoted by V^1 and $\tilde{V}^1$ in [Part I]). Setting the continuous spectral function $\hat{q}(\lambda)$ to zero in the Riemann-Hilbert system of [Part I, Eqs. (30a)–(30d)], we obtain an algebraic system of equations

$$\tilde{V}(t, \lambda_m^*) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{i=1}^N \frac{\tilde{q}(\lambda_i) e^{2j\lambda_i t} V(t, \lambda_i)}{\lambda_m^* - \lambda_i},$$

$$V(t, \lambda_m) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sum_{i=1}^N \frac{\tilde{q}^*(\lambda_i) e^{-2j\lambda_i^* t} \tilde{V}(t, \lambda_i^*)}{\lambda_m - \lambda_i^*}. \qquad (8)$$

Let K be an $N \times N$ matrix with entries

$$[K]_{ij} = \frac{\tilde{q}_i e^{2j\lambda_i t}}{\lambda_j^* - \lambda_i}, \quad 1 \le i, j \le N.$$

Let $e_{N \times 1}$ be the all one column vector, $e_i = 1$, $i = 1, \cdots, N$, and define variables

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

$$(J_1)_{2 \times N} = \begin{pmatrix} e^T \\ 0 \end{pmatrix}, \quad (J_2)_{2 \times N} = \begin{pmatrix} 0 \\ e^T \end{pmatrix},$$

$$J_{2 \times N} = J_2 - J_1 K^* = \begin{pmatrix} -e^T K^* \\ e^T \end{pmatrix},$$

$$\tilde{J}_{2 \times N} = J_1 + J_2 K = \begin{pmatrix} e^T \\ e^T K \end{pmatrix},$$

$$F_{N \times 1} = \begin{pmatrix} \tilde{q}_1 e^{2j\lambda_1 t} & \tilde{q}_2 e^{2j\lambda_2 t} & \cdots & \tilde{q}_N e^{2j\lambda_N t} \end{pmatrix}^T.$$

Using these variables, the algebraic equations (8) are simplified to

$$\tilde{U} = J_1 + UK, \qquad U = J_2 - \tilde{U}K^*.$$

Note that K^* is the complex conjugate of K (not the conjugate transpose). The solution of the above system is

$$U = (J_2 - J_1 K^*)(I_N + KK^*)^{-1} = J(I_N + KK^*)^{-1},$$
$$\tilde{U} = (J_1 + J_2 K)(I_N + KK^*)^{-1} = \tilde{J}(I_N + KK^*)^{-1}.$$

From [Part I, Section. VII. B, Eq. 32], the N-soliton formula is given by

$$q(t) = -2je^T (I_N + K^* K)^{-1} F^*. \tag{9}$$

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

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

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

where $\Phi = \alpha t + (\alpha^2 - \omega^2)z - \angle\tilde{q}_0 - \frac{\pi}{2}$ and $t_0 = \frac{1}{\omega}\log\left|\frac{\tilde{q}_0}{\omega}\right| - 2\alpha z$. The celebrated equation for the single-soliton obtained from (9) is

$$q(t) = -j\omega e^{-j\alpha t} e^{-j\phi_0} \operatorname{sech}\left(\omega(t-t_0)\right). \tag{10}$$

From the phase-symmetry of the NLS equation, the factor $-j$ in (10) 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 $\tilde{q}(\lambda)$ 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. The Nth row of the K matrix is proportional to $\exp(2j\lambda_N t)$. Thus this row gets a large scale factor as $\Im(\lambda)$ is increased ($t < 0$) which then makes $I_N + K^* K$ ill-conditioned at large negative times. As a result, the Riemann-Hilbert system, at least 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 multi-solitons without solving a Riemann-Hilbert system or directly using the NFT. A method which is particularly analytically insightful is the Hirota direct method [20]. 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 sums of exponentials. Computationally, bilinear equations are solved perturbatively by expanding the unknowns in terms of the powers of a small parameter ϵ. For integrable equations, this series truncates, rendering approximate solutions of various orders to be indeed exact. The bilinear transformation has been

found for many integrable equations [20], 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 γ in the NLS equation. Plugging $q = G/F$ into the NLS equation

$$jq_z = q_{tt} + 2\gamma|q|^2 q,$$

we get

$$j(G_z F - FG_z) = FG_{tt} - 2F_t G_t - GF_{tt}$$
$$+ 2\frac{F_t^2 + \gamma|G|^2}{F}G$$
$$= FG_{tt} - 2F_t G_t + GF_{tt}$$
$$+ 2\frac{F_t^2 + \gamma|G|^2 - FF_{tt}}{F}G, \tag{11}$$

where we have added and subtracted $2GF_{tt}$. Equation (11) is trilinear in F and G. It can be made bilinear by setting

$$j(G_z F - GF_z) = FG_{tt} - 2F_t G_t + GF_{tt}, \tag{12a}$$
$$F_t^2 + \gamma|G|^2 - FF_{tt} = 0. \tag{12b}$$

This is a special solution for (11) which, as we shall see, corresponds to N-soliton solutions. It is very convenient (though not necessary) to organize (12a)–(12b) using the Hirota D operator

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

resulting in

$$(jD_z + D_t^2)FG = 0, \tag{13a}$$
$$D_t^2 FF = 2\gamma|G|^2. \tag{13b}$$

Note that the D-operator acts on a pair of functions to produce another function. Note further that (13a) does not depend on the nonlinearity parameter γ. That is to say, the nonlinearity has been separated from equation (13a). 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 (13a)–(13b) have solutions in the form of a sum of exponentials. As shown in Appendix A, F and G are obtained as [20], [21]:

$$F(t, z) = \sum_{b=\{0,1\}^{2N}} \delta_1(b) \exp\left(b^T X + b^T Rb\right),$$

$$G(t, z) = \sum_{b=\{0,1\}^{2N}} \delta_2(b) \exp\left(b^T X + b^T Rb\right),$$

where $b = [b_i]_{i=1}^{2N}$ is a binary column vector, $b_i = \{0, 1\}$, $X = [X_i]_{i=1}^{2N}$, $X_i = \zeta_i t - k_i z + \phi_i$, $\zeta_{i+N} = \zeta_i^*$, $k_{i+N} = k_i^*$, $\phi_{i+N} = \phi_i^*$, $X_{i+N} = X_i^*$, $k_i = j\zeta_i^2$ is the dispersion relation, $R_{2N \times 2N}$ is the Riemann matrix,

$$R_{ij} = \begin{cases} 0, & i \geq j, \\ 2\log(\zeta_i - \zeta_j) - \log\gamma, & (N + \frac{1}{2} - i)(N + \frac{1}{2} - j) > 0, \\ -2\log(\zeta_i + \zeta_j) + \log\gamma, & i \leq N \text{ and } j \geq N + 1, \end{cases}$$
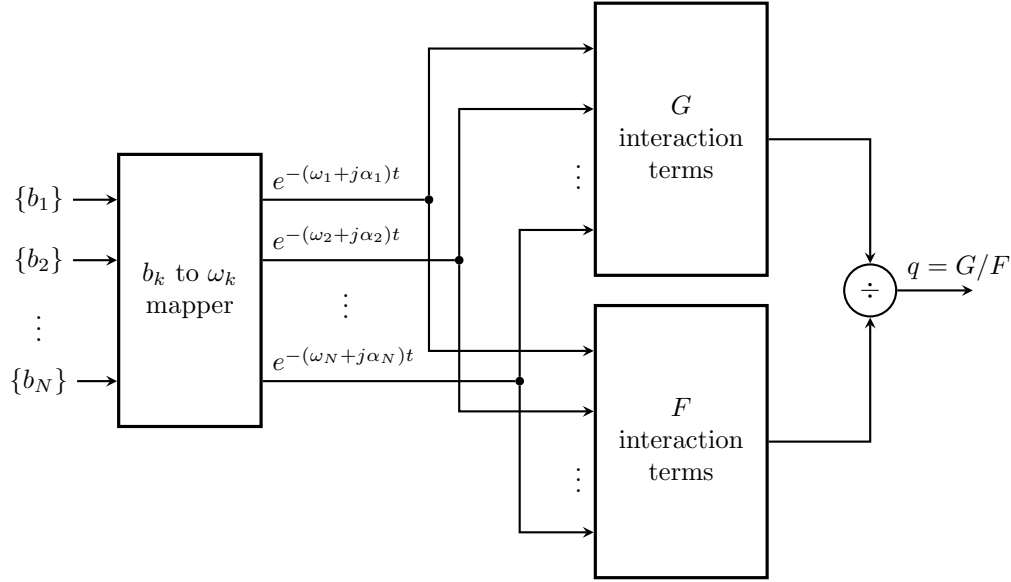
Fig. 4. Hirota modulator for creating N-solitons.

and

$$\delta_1(b) = \begin{cases} 1, & \sum_{i=1}^{N} b_i = \sum_{i=N+1}^{2N} b_i, \\ 0, & \text{otherwise}, \end{cases}$$

$$\delta_2(b) = \begin{cases} 1, & \sum_{i=1}^{N} b_i = 1 + \sum_{i=N+1}^{2N} b_i, \\ 0, & \text{otherwise}. \end{cases}$$

Matrix R is upper triangular (with zeros on diagonal) and such that when it is partitioned into four $N \times N$ blocks, the 11, 12 and 22 blocks capture, respectively, the interaction between X_i and X_j, X_i and X_j^*, and X_i^* and X_j^* variables. The entries in the 11, 12 and 22 blocks are, respectively, given by $2\log(\zeta_i - \zeta_j) - \log\gamma$, $-2\log(\zeta_i + \zeta_{N-j}^*) + \log\gamma$ and $2\log(\zeta_{N-i}^* - \zeta_{N-j}^*) - \log\gamma$. Eigenvalues λ_i are related to ζ_i via $\zeta_i = -2j\lambda_i$, whereas the Hirota spectral amplitudes ϕ_i are generally different from those of other methods.

Functions F and G are in the form of the sum of all possible exponentials such that in F the number of non-conjugate and conjugate variables X_i and X_j^* is the same while in G the former is one more than the latter. For each exponential term, terms R_{ij} corresponding to the interaction between all possible pairs X_i and X_j, $i \neq j$, in the exponent are added; see Table II.

Note that functions F and G both contribute to the signal amplitude, whereas F, being real-valued, does not contribute to the signal phase. Using the identity $\partial_{tt}\log F = \left(F_{tt}F - F_t^2\right)/F^2$, (13b) is reduced to $|q(t,z)|^2 = \gamma^{-1}\partial_{tt}\log F$. Therefore it is also possible to derive the amplitude of q solely in terms of a function of F. This is because F and G are not independent.

Two important observations follow from the Hirota method. Firstly, multi-soliton solutions of the NLS equation in the F and G domain ($q = G/F$) are the summation of exponentially decaying/growing functions $e^{\pm\omega t}e^{j\alpha t - kz}$, each located at a frequency α. That is to say, while plane waves $e^{j\alpha t - kz}$ are

the natural Fourier basis functions that solve linear PDEs, for integrable systems, exponentially decaying/growing functions are suitable. The addition of the decaying/growing factor $e^{\pm\omega t}$ is the point at which the nonlinear Fourier transform diverges from the linear Fourier transform [21]. 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. 4. Table II shows 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 multi-solitons numerically. There are $\binom{2N}{N} \sim 2^{2N}$ and $\binom{2N}{N+1} \sim 2^{2N}$ 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: Multi-soliton solutions of the NLS equation can be constructed recursively using the Darboux transformation. 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 [22]. 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

$$\begin{aligned} x_t &= P(\zeta, q)x, \\ x_z &= M(\zeta, q)x, \end{aligned} \tag{14}$$

for the signal q and complex number $\zeta = \lambda$ (not necessarily an eigenvalue of q), where the P and M are 2×2 matrix operators defined in [Part I]. It is clear that $\tilde{x} = [x_2^*, -x_1^*]^T$ satisfies

TABLE II
THE STRUCTURE OF THE INTERACTION TERMS IN F AND G^1.

	F
$N=1$	$1 + e^{X_1 + X_1^* + R_{13}}$
2	$1 + e^{X_1 + X_1^* + R_{13}} + e^{X_2 + X_2^* + R_{24}} + e^{X_1 + X_2^* + R_{14}} + e^{X_2 + X_1^* + R_{23}} + e^{X_1 + X_2 + X_1^* + X_2^* + R_{12} + R_{13} + R_{14} + R_{23} + R_{24} + R_{34}}$
3	$1 + e^{X_1 + X_1^*} + e^{X_2 + X_2^*} + e^{X_3 + X_3^*} + \left(e^{X_1 + X_2^*} + e^{X_2 + X_1^*} + e^{X_1 + X_3^*} + e^{X_3 + X_1^*} + e^{X_2 + X_3^*} + e^{X_3 + X_2^*}\right) +$ $\left(e^{X_1 + X_2 + X_1^* + X_2^*} + e^{X_1 + X_2 + X_1^* + X_3^*} + e^{X_1 + X_2 + X_2^* + X_3^*} + e^{X_1 + X_3 + X_1^* + X_2^*} + e^{X_1 + X_3 + X_1^* + X_3^*} + e^{X_1 + X_3 + X_2^* + X_3^*}\right.$ $\left. + e^{X_2 + X_3 + X_1^* + X_3^*} + e^{X_2 + X_3 + X_1^* + X_2^*} + e^{X_2 + X_3 + X_2^* + X_3^*}\right) + e^{X_1 + X_2 + X_3 + X_1^* + X_2^* + X_3^*}$

	G
$N=1$	e^{X_1}
2	$e^{X_1} + e^{X_2} + e^{X_1 + X_2 + X_1^* + R_{12} + R_{13} + R_{23}} + e^{X_1 + X_2 + X_2^* + R_{12} + R_{14} + R_{24}}$
3	$e^{X_1} + e^{X_2} + e^{X_3} + \left(e^{X_1 + X_2 + X_1^*} + e^{X_1 + X_2 + X_2^*} + e^{X_1 + X_2 + X_3^*} + e^{X_1 + X_3 + X_1^*} + e^{X_1 + X_3 + X_2^*} + e^{X_1 + X_3 + X_3^*}\right.$ $\left. e^{X_2 + X_3 + X_1^*} + e^{X_2 + X_3 + X_2^*} + e^{X_2 + X_3 + X_3^*}\right) + \left(e^{X_1 + X_2 + X_3 + X_1^* + X_2^*} + e^{X_1 + X_2 + X_3 + X_1^* + X_3^*} + e^{X_1 + X_2 + X_3 + X_2^* + X_3^*}\right)$

^{1} For $N = 3$, terms R_{ij} in the exponent are not shown, due to space limitation.

(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 $\Sigma = S\Gamma S^{-1}$, where $S = [\phi(t, \lambda; q), \tilde{\phi}(t, \lambda; q)]$ and $\Gamma = \text{diag}(\lambda, \lambda^*)$. If $v(t, \mu; q)$ satisfies (14), then $u(t, \mu; \tilde{q})$ obtained from the Darboux transform*

$$u(t, \mu; \tilde{q}) = (\mu I - \Sigma)\, v(t, \mu; q), \qquad (15)$$

satisfies (14) as well, for

$$\tilde{q} = q + 2j(\lambda^* - \lambda)\frac{\phi_1 \phi_2^*}{|\phi_1|^2 + |\phi_2|^2}. \qquad (16)$$

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 μ is an eigenvalue of q, then μ is an eigenvalue of $\tilde{q}$ as well. Furthermore, since $u(t, \mu = \lambda; \tilde{q}) \neq 0$, λ is also an eigenvalue of $\tilde{q}$. It follows that the eigenvalues of $\tilde{q}$ are the eigenvalues of q together with λ.

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 (7) 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 $q^{(0)} = 0$. The initial eigenvectors in Fig. 6 are chosen to be the (non-canonical) eigenvectors $v(t, \lambda_j; 0) = [A_j e^{-j\lambda_j t}, B_j e^{j\lambda_j t}]^T$. The coefficients A_j and B_j control the spectral amplitudes and the shape of the pulses. For a single-soliton, $A_j = \exp(j\angle\tilde{q})$ and $B_j = |\tilde{q}|$.

Remark 3. In this paper we mostly use Darboux method for numerical generation of N-solitons and discrete spectrum simulations. Hirota method, on the other hand, is preferred for the analytical examination of N-soliton and insight into the NFT. The Riemann-Hilbert approach is more general and captures the continuous spectrum too (though it is sometimes ill-behaved).

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 λ_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

$$q(t, z) \to \sum_{i=1}^{N} \omega_i e^{-j\alpha_i t + j(\alpha_i^2 - \omega_i^2)z + j\phi_i} \text{sech}(\omega_i(t - 2\alpha_i z - t_i)),$$

where $\lambda_i = (\alpha_i + j\omega_i)/2$ are eigenvalues and t_i is the time center. This breakdown of a signal to its individual components, while best observed in the case of multi-solitons, is simply a result of group velocity dispersion and exists for
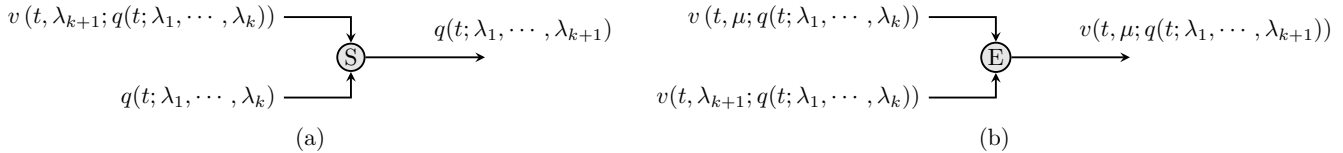
Fig. 5.   Updates in the Darboux transformation: (a) signal update; (b) eigenvector update.
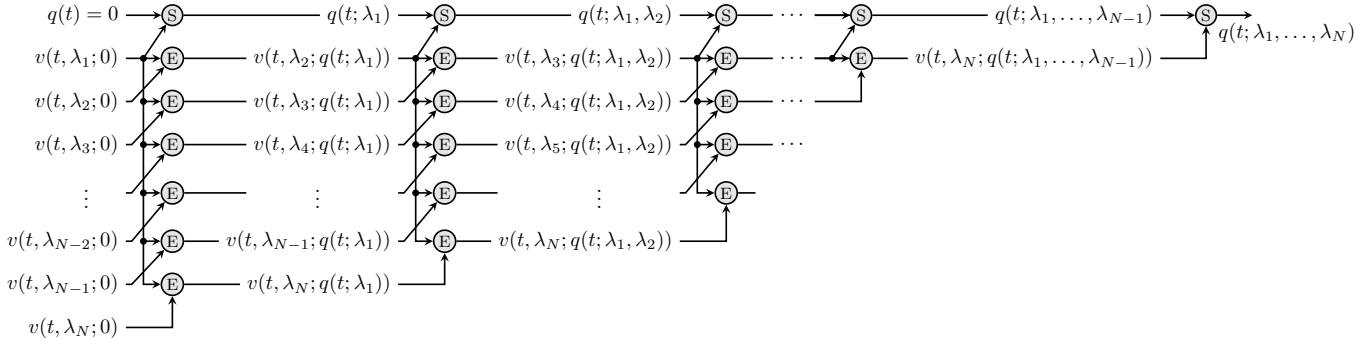

Fig. 6.   Darboux iterations for the construction of an N-soliton.

all pulses similarly (including sinc functions). The extent of breakdown and shift depends on a variety of factors, such as the length of the fiber, number of mass points, fiber dispersion and dispersion-management schemes. The effects of pulse broadening must be carefully considered, especially if dispersion is not managed.

D. Demodulating the Discrete Spectrum

To demodulate a multi-soliton pulse, the eigenproblem (7) needs to be solved. There is limited work in the mathematical literature concerning the numerical solution of the Zakharov-Shabat spectral problem (7). 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 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 4. In Section II-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, noise is a weaker form of distortion, and hence we do not intend to provide here a comprehensive analysis.

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 $\gg 0$ dB used in long-haul fiber-optic communications, one can treat 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 5. 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 ideal Raman amplification, which compensates loss but adds an equal amount of noise. Therefore loss is essentially traded with noise, which is treated in this section.

If noise is added in a lumped fashion, we have the deterministic NLS equation with random initial data at the input of each fiber span. In this case, the NFT can be used without approximation.

If noise is injected continuously throughout the fiber as a result of DRA, we have the stochastic NLS equation (1) that includes an additive space-time noise term. This equation is generally not integrable[5]. However, 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

---

[5]In a special case, the NLS equation with a certain real-valued multiplicative potential can still be integrable [23, Appendix D].

TABLE III
UPDATE EQUATIONS FOR THE RECURSIVE DARBOUX METHOD

Eigenvector update:

$$v_1(t, \lambda_j; q^{(k+1)}) = \frac{1}{\left\| v(t, \lambda_{k+1}; q^{(k)}) \right\|^2} \left\{ \left\{ (\lambda_j - \lambda_{k+1}) \left| v_1(t, \lambda_{k+1}; q^{(k)}) \right|^2 + (\lambda_j - \lambda_{k+1}^*) \left| v_2(t, \lambda_{k+1}; q^{(k)}) \right|^2 \right\} v_1(t, \lambda_j; q^{(k)}) \right.$$

$$\left. + (\lambda_{k+1}^* - \lambda_{k+1}) v_1(t, \lambda_{k+1}; q^{(k)}) v_2^*(t, \lambda_{k+1}; q^{(k)}) v_2(t, \lambda_j; q^{(k)}) \right\},$$

$$v_2(t, \lambda_j; q^{(k+1)}) = \frac{1}{\left\| v(t, \lambda_{k+1}; q^{(k)}) \right\|^2} \left\{ \left\{ (\lambda_j - \lambda_{k+1}^*) \left| v_1(t, \lambda_{k+1}; q^{(k)}) \right|^2 + (\lambda_j - \lambda_{k+1}) \left| v_2(t, \lambda_{k+1}; q^{(k)}) \right|^2 \right\} v_2(t, \lambda_j; q^{(k)}) \right.$$

$$\left. + (\lambda_{k+1}^* - \lambda_{k+1}) v_1^*(t, \lambda_{k+1}; q^{(k)}) v_2(t, \lambda_{k+1}; q^{(k)}) v_1(t, \lambda_j; q^{(k)}) \right\},$$

for $k = 0, \ldots, N-2$ and $j = k+2, \ldots, N$.

Signal update:

$$q^{(k+1)} = q^{(k)} + 2j(\lambda_{k+1}^* - \lambda_{k+1}) \frac{v_1(t, \lambda_{k+1}; q^{(k)}) v_2^*(t, \lambda_{k+1}; q^{(k)})}{\left\| v(t, \lambda_{k+1}; q^{(k)}) \right\|^2}.$$

of the next segment. The DRA can thus be approximately treated similar to the lumped noise case. In this case, NFT is used under such approximation.

In this section, we study the effect of the lumped and distributed noise 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 (NFT assumptions).

A. Perturbation of Eigenvalues

1) Lumped Noise: The NFT arises in the spectral analysis of the L operator (6). We can easily analyze the perturbations of the eigenvalues of the L operator to the first order in the noise level ϵ.

Let us denote the nonlinear Fourier transform of $q(t)$ in the absence of noise by $(\hat{q}(\lambda), \tilde{q}(\lambda_j))$. As the signal $q(t)$ at the input of each small segment is perturbed to $q(t) + \epsilon n(t)$ for some small parameter ϵ 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 λ satisfy

$$(L + \epsilon R)v = \lambda v, \qquad R = \begin{pmatrix} 0 & n \\ -n^* & 0 \end{pmatrix}, \qquad (17)$$

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 [24]–[27]. These results, however, are non-stochastic and the distribution of the scattering data is still lacking. A very interesting work is [28] 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 much work pertaining to the statistics of the parameters of a single-soliton; see *e.g.*, [29] 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 λ and eigenvectors of the adjoint operator L^* associated with $\mu \neq \lambda^*$. 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 ϵ as

$$v(t) = v^{(0)}(t) + \epsilon v^{(1)}(t) + \epsilon^2 v^{(2)}(t) + \cdots, \quad (18a)$$
$$\lambda = \lambda^{(0)} + \epsilon \lambda^{(1)} + \epsilon^2 \lambda^{(2)} + \cdots. \quad (18b)$$

We assume these variables are analytic functions of ϵ so that the above series are convergent. Plugging (18a)–(18b) into (17) and equating like powers of ϵ, we obtain

$$L v^{(0)} = \lambda^{(0)} v^{(0)}, \quad (19)$$
$$(L - \lambda^{(0)}) v^{(1)} = -(R - \lambda^{(1)}) v^{(0)},$$
$$(L - \lambda^{(0)}) v^{(2)} = -(R - \lambda^{(1)}) v^{(1)} + \lambda^{(2)} v^{(0)},$$

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 (19) with some vector u; the left hand side of the resulting expression is

$$\langle u, (L - \lambda^{(0)}) v^{(1)} \rangle = \langle (L - \lambda^{(0)})^* u, v^{(1)} \rangle$$
$$= \langle (L^* - \lambda^{(0)*}) u, v^{(1)} \rangle. \quad (20)$$

To have the right-hand side of (20) 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 $Lv = \lambda^{(0)}v$, it can be verified that $L^*u = \lambda^{(0)}u$ for $u = [v_1, -v_2] = \Sigma_3 v$. Setting $u = u^{(0)} = \Sigma_3 v^{(0)}(t, \lambda^*)$,

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

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

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

and so on.

To summarize, the fluctuations of the nth eigenvalue λ_n with nominal eigenvector v_n is given by

$$\hat{\lambda}_n = \lambda_n + \epsilon\frac{\langle u_n, Rv_n\rangle}{\langle u_n, v_n\rangle} + O(\epsilon^2), \quad n = 1, 2, \ldots, N, \quad (21)$$

where $u_n = \Sigma_3 v_n$ and '$^\wedge$' 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, $v^{(k)}$, $k \geq 1$, is not straightforward because the underlying operator is not self-adjoint.

2) Distributed Noise: Consider now the perturbed NLS equation

$$jq_z = q_{tt} + 2|q|^2q + \epsilon n(t, z), \quad (22)$$

where ϵ is a small parameter (noise power) and the normalized noise term $n(t, z)$ 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 λ vary with z. The equality of mixed derivatives $v_{tz} = v_{zt}$ gives

$$\begin{pmatrix} -j\lambda_z & q_z + jq_{tt} + 2j|q|^2q \\ -q_z^* + jq_{tt}^* + 2j|q|^2q^* & j\lambda_z \end{pmatrix} v = 0.$$

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

$$\lambda_z v = \epsilon\bar{R}v, \quad \bar{R} = -R.$$

Note that, as before, we do not have $v(t, z)$ a priori because, according to (6), it depends on the noisy signal $q(t, z)$ and $\lambda(z)$, both of which are unknown. However, if the noise level ϵ is small, we can expand $v(t, z)$ and λ in powers of ϵ as in (18a) and (18b) to obtain $(\lambda^{(0)})_z = 0$ and $\bar{R}v^{(0)} = (\lambda^{(1)})_z v^{(0)}$ (*i.e.*, $(\lambda^{(1)})_z$ appears as a time-independent eigenvalue of $\bar{R}(t)$). Taking the inner-product with $u^{(0)} = \Sigma_3 v^{(0)}(t, \lambda^*)$ on both sides of $\bar{R}v^{(0)} = (\lambda^{(1)})_z v^{(0)}$, we obtain the first-order variation of eigenvalues

$$(\lambda_1)_z = \frac{\langle u^{(0)}, \bar{R}v^{(0)}\rangle}{\langle u^{(0)}, v^{(0)}\rangle}. \quad (23)$$

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 $\langle \Sigma_3 v(t, \lambda^*), v(t, \lambda)\rangle = -\frac{\omega}{2}(1 + \frac{\omega^2}{|\bar{q}|^2}) = -\frac{1}{\omega}$, where we assumed $|\bar{q}| = \omega$ so that the soliton is centered at $t_0 = 0$. Furthermore,

$$\langle \Sigma_3 v(t, \lambda^*), \bar{R}v(t, \lambda)\rangle$$
$$= -\int (v_1(t, \lambda^*)v_2^*(t, \lambda)n^* + v_2(t, \lambda^*)v_1^*(t, \lambda)n)\,dt$$
$$= -\int \frac{1}{4}\text{sech}^2(\omega t)(v_2(t, \lambda^*) - v_2^*(t, \lambda))\Re\bar{n}dt$$
$$\quad -j\int \frac{1}{4}\text{sech}^2(\omega t)(v_2(t, \lambda^*) + v_2^*(t, \lambda))\Im\bar{n}dt$$
$$= \frac{1}{2}\int \text{sech}(\omega t)\tanh(\omega t)\Re\bar{n}dt - \frac{j}{2}\int \text{sech}(\omega t)\Im\bar{n}dt,$$

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

$$\alpha_z = -\int \text{sech}(\tau)\tanh(\tau)z_1(\tau, z)d\tau, \quad (24a)$$
$$\omega_z = \int \text{sech}(\tau)z_2(\tau, z)d\tau, \quad (24b)$$

where $z_1 = \Re z$, $z_2 = \Im z$ and $z(t, z) = \bar{n}(t/\omega, z)$, *i.e.*, z_1 and z_2 are independent Gaussian processes each with total power $\omega W z\sigma_0^2$. The Gordon-Haus effect can be observed from the α_z equation. Note that in fiber optics noise is added to the signal, *i.e.*, n in this subsection should be replaced with jn.

Remark 6. Note that the higher-order terms $\lambda^{(i)}$ in expansion (21) are signal-dependent (even though they are normalized by $\langle u_n, v_n\rangle$). For instance, Example 2, and as well as Fig. 8 (b), show that noise variance grows with signal. In this case higher-order terms need to be included as well for precision. The perturbation approach given in this section takes into account only the first term in the expansion, and as a result mostly describes the bulk of the distribution in the small noise limit. Note, however, that the first term is also signal-dependent. Thus the effect of the noise growth with signal is accounted for in the above analysis.

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], in the Riemann-Hilbert approach the 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]

$$\frac{dy(t, \lambda)}{dt} = -(\bar{q}(t, \lambda) + \epsilon n(t))y^2(t, \lambda) - (\bar{q}^*(t, \lambda) + \epsilon n^*(t)), \quad (25)$$

where $y(-\infty, \lambda) = 0$ and

$$\bar{q}(t, \lambda) = q(t)\exp(2j\lambda t), \quad \hat{q}(\lambda) = \lim_{t\to\infty} y(t, \lambda).$$

One can write a Fokker-Planck equation for the probability distribution of $y(t, \lambda)$ in (25) [30], [31]. If the signal $q(t)$ is

known, *e.g.*, $q(t) = 0$, the resulting equation can be solved. In general, however, we can expand

$$y(t, \lambda) = y_0(t, \lambda) + \epsilon y_1(t, \lambda) + \epsilon^2 y_2(t, \lambda) + \cdots,$$
$$\hat{q}(\lambda) = \hat{q}_0(\lambda) + \epsilon \hat{q}_1(\lambda) + \epsilon^2 \hat{q}(\lambda) + \cdots,$$

and equate like powers of ϵ. We obtain that $\hat{q}_0(\lambda)$ (and its corresponding $y_0(t, \lambda)$) is the spectral amplitude when noise is zero (assuming $\bar{q}(t, \lambda) \approx \bar{q}(t, \lambda_0)$). Define

$$G(t, \lambda) = 2 \int\limits_{-\infty}^{t} \bar{q}(\tau, \lambda) y_0(\tau, \lambda) \mathrm{d}\tau,$$

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

$$\hat{q}_1(\lambda) = e^{-G(\infty, \lambda)} \int_{-\infty}^{\infty} \hat{n}(\tau, \lambda) e^{G(\tau, \lambda)} \mathrm{d}\tau.$$

This is a conditionally Gaussian random variable (for each λ) with variance

$$\mathsf{E} |\hat{q}_1(\lambda)|^2 = \sigma_0^2 e^{-2\Re G(\infty, \lambda)} \int_{-\infty}^{\infty} \big(|y_0(\tau, \lambda)|^4 + 1\big) e^{2\Re G(\tau, \lambda)} \mathrm{d}\tau,$$

where we assumed that noise is delta-correlated. If for some signals, $\mathsf{E}|\hat{q}_1(\lambda)|^2$ is unbounded, the above perturbation expansion fails and a slow-scale variable $T = \epsilon t$ needs to be introduced.

In summary, in this section we showed that a simple first-order perturbation analysis, though inaccurate, gives insights into the nature of the statistics in the nonlinear spectral domain. In the next section, we discuss the impact of the noise on the achievable spectral efficiencies.

V. Some Achievable Spectral Efficiencies Using the NFT

We now turn to a numerical study of data modulation in the nonlinear Fourier domain, providing some simulation results and examples of achievable spectral efficiencies. First, in Sections V-A to V-C, we set the continuous spectrum to zero and modulate the discrete spectrum. This special case corresponds to an N-soliton data transmission system. Then, in Section V-D, we set the discrete spectrum to zero and modulate the continuous spectrum.

For the discrete spectrum modulation, we begin with a classical on-off keying soliton transmission ($N = 1$), in which, in any symbol period T_s, either zero or a fundamental soliton is sent. We then increase N and the spectral efficiency by considering 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.

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/Hz. As a first step to

improve upon the on-off keying solitons, first we continuously modulate one eigenvalue in a given region (*i.e.*, a classical soliton but with varying amplitude, width and phase). We next consider multi-soliton 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 optical 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 operations that do not change integrability.

A. Spectral Efficiency of 1-Soliton Systems

Traditional 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 simulations naturally include this effect.

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

Note that 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 small chance of creating additional solitons out of noise at some distance, or the soliton spectrum might collapse into the real axis in the λ plane. All these effects are negligible if noise is small enough, $2Wz\sigma^2 \ll E(0)$, and the propagation distance is not exceedingly long. Thus, at a length scale determined by $2Wz\sigma^2 \ll E(0)$, we can still think of the noisy signal as a soliton with re-modulated parameters.

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

$$\frac{\partial E}{\partial z} = 2\Im \int\limits_{-\infty}^{\infty} q(\tau, z) Z(\tau, z) \mathrm{d}\tau,$$

where $E(z) = \int |q(\tau, z)|^2 \mathrm{d}\tau$ is the energy, and $Z = -\epsilon n^*$ is a noise process similar to ϵn. Replacing $q(t, z) \to q(t, 0)$ in the
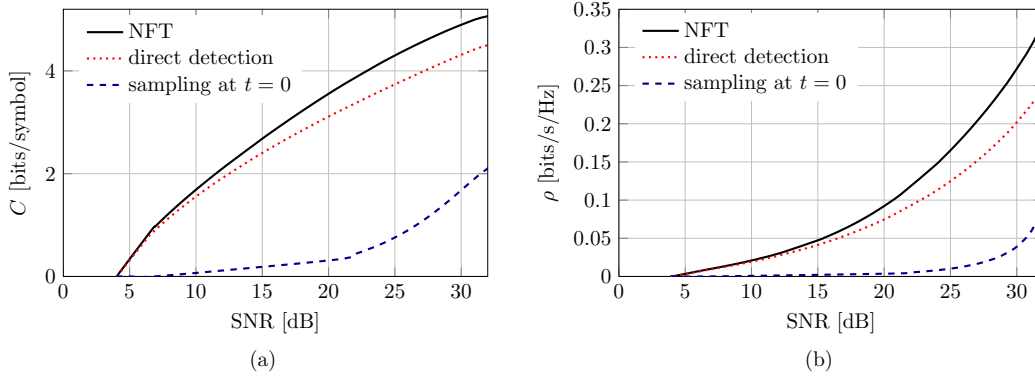
Fig. 7. (a) Capacity (bits/symbol) and, (b) spectral efficiency (bits/s/Hz), of soliton systems using direct detection, sampling at $t = 0$, and the NFT.

small noise limit, we obtain that energy fluctuation is a signal-dependent conditionally Gaussian random variable $E(z) \approx \mathcal{N}_{\mathbb{R}}\big(E(0), \sigma^2 E(0)\big)$ ($\sigma^2 = 2Wz\sigma_0^2/P$). Ignoring the energy of the continuous spectrum in the small noise limit $\sigma^2 \ll 1$, we have $E \approx 2\omega$ and therefore

$$\omega(z) = \omega(0) + \sigma\sqrt{\frac{\omega(0)}{2}}\tilde{n}, \quad \omega(0) \gg \sigma^2, \qquad (26)$$

where $\tilde{n} \sim \mathcal{N}_{\mathbb{R}}(0,1)$. The conditional probability density function (PDF) is

$$f_{\Omega|\Omega_0}(\omega|\omega_0) = \frac{1}{\sqrt{\pi\sigma^2\omega_0}}e^{-\frac{(\omega-\omega_0)^2}{\sigma^2\omega_0}}, \quad \omega_0 = \omega(0), \quad (27)$$

and the PDF of $r = \sqrt{\omega}$ given $r_0 = \sqrt{\omega_0}$, $\omega, \omega_0 \geq 0$, is approximately a Rician distribution

$$\begin{aligned} f_{R|R_0}(r|r_0) &= \frac{1}{\sqrt{\pi\sigma^2}}e^{-\frac{(r-r_0)^2}{\sigma^2}} \\ &\approx \frac{2r}{\sigma^2}e^{-\frac{r^2+r_0^2}{\sigma^2}}I_0\big(\frac{2rr_0}{\sigma^2}\big), \quad r, r_0 \gg \sigma, \end{aligned}$$

which is signal-independent in the high SNR regime.

In [31] we have shown that a half-Gaussian density

$$f_{\Omega_0}(\omega_0) = \frac{2}{\sqrt{2\pi\mathcal{P}}}e^{-\frac{\omega_0^2}{2\mathcal{P}}}, \quad \omega_0 \geq 0,$$

gives the asymptotic capacity for (27)

$$C \sim \frac{1}{2}\log(\text{SNR}),$$

where $\text{SNR} = \frac{\mathcal{P}}{\sigma^2}$.

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

$$\rho(\mathcal{P}) = \max_{\substack{f(\omega_0) \\ \omega_0 \in S}} \frac{1}{\text{BW}(S)\mathsf{E}T(\omega_0)}I(\omega_0;\omega), \qquad (28)$$

$$\mathsf{E}P(\omega_0) \leq \mathcal{P},$$

where $T(\omega_0)$ and $P(\omega_0)$ are the width and the power of a single-soliton with amplitude ω_0 and $\text{BW}(S) = \max_{\omega_0 \in S}\text{BW}(\omega_0)$ is the maximum passband bandwidth that the signal set

requires for transmission. For a one soliton signal, we have approximately

$$T(\omega_0) = \frac{7}{\omega_0}, \quad P(\omega_0) = \frac{\omega_0^2}{6.2}, \quad \text{BW}(\omega_0) = 0.95\omega_0, \quad (29)$$

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 (29), the maximum spectral efficiency of a baseline on-off keying system is obtained to be about $\rho \approx 0.15$ bits/s/Hz at the average power $\mathcal{P}_0 = 0.16$ mW. Note that the per unit cost capacity problem (28) is non-convex and hence finding the global optimum may prove to be challenging. Here we simply optimize mutual information and scale it by $\text{BW}(S) \times \mathsf{E}T(\omega_0)$ evaluated at the mutual-information-maximizing input distribution.

Fig. 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 (28), the spectral efficiencies shown in the Fig. 7(b) are only lower bounds on the actual achievable values. Figs. 8(a)–(b) show 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 Figs. 7(a)–(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.
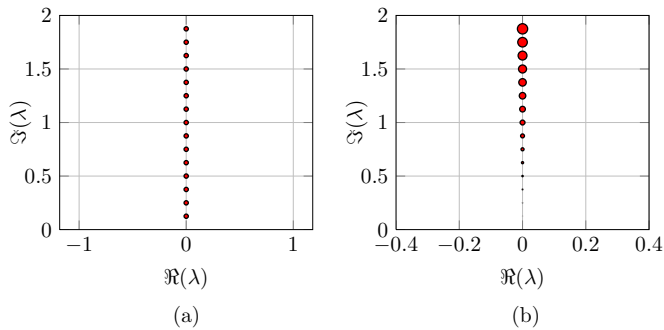
Fig. 8. (a) Eigenvalue constellation at the transmitter. (b) Noise balls at the receiver in the NFT approach. The signal-dependency of the noise balls can be analytically seen *e.g.*, through (26). The radius of each circle is the standard deviation of the received eigenvalue.

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

$$
\begin{aligned}
S_1 &: \quad 0, \\
S_2 &: \quad \tilde{q}(0.5j) = 1, \\
S_3 &: \quad \tilde{q}(0.25j) = 0.5, \\
S_4 &: \quad \tilde{q}(0.25j, 0.5j) = (1, 1).
\end{aligned} \tag{30}
$$

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 $\mathcal{P}_0 = 0.1876$ 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 (30) has average power $\mathcal{P} = 0.46P_0$ and average time duration $T = 1.65T_1$, 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 $\rho = \log 4/(1.65T_1W_0) = 1.2121 \times \rho_0$ bits/s/Hz and operates at $R = 1.2121 \times R_0$ for about the same average power ($\mathcal{P} = 0.1748$ mW). Note that without S_4 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 pulse shapes. 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 $[v_1^2 e^{j\lambda t}, v_2^2 e^{-j\lambda t}]$ 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 $2.58T_1$, 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 Eigenvalues and Spectral Amplitudes: We can improve upon the previous example by modulating the spectral amplitudes as well. Consider the following signal set

$$
\begin{aligned}
S_1 &: \quad 0, \\
S_2 - S_5 &: \quad \tilde{q}(0.5j) = \tilde{q}_1, \\
S_6 - S_9 &: \quad \tilde{q}(0.25j) = \tilde{q}_2, \\
S_{10} - S_{16} &: \quad \tilde{q}(0.25j, 0.5j) = (\tilde{q}_3, \tilde{q}_4).
\end{aligned}
$$

We make a 3-ary constellation on $\tilde{q}_i \in \{0.5, 1, 1.5\}$. This creates a signal set with 16 elements. Here pulses are extended to $3T_1$ time duration. The resulting constellation has average power $\mathcal{P} = 1.06P_0$ and average time duration $T = 2.236T_1$, 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 $\rho = \log 16/(2.236T_1W_0) = 1.79 \times \rho_0$ 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 $\rho = 2.2\rho_0 = 0.73$ 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 7. 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 $\{\tilde{q}(\lambda_j)\}_{j=1}^{j=N}$ can be investigated for this purpose.

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

TABLE IV
PARAMETERS OF THE SIGNAL SET IN (30)[1].

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	$0.5\,E_0$	$2T_0$	$2T_1$	$0.25P_0$	$0.5W_0$
S_4	$1.5\,E_0$	$4.25T_0$	$2.58T_1$	$0.58P_0$	$0.5W_0$

[1] Here $E_0 = 2$, $T_0 = 1.763$ at FWHM power, $T_1 = 5.2637$ (99% energy), $P_0 = 0.38$ and $W_0 = 0.5714$. The normalization factors in the NLS equation are $T_n = 25.246$ ps and $P_n = 0.5$ mW at dispersion 0.5 ps/(nm − km) and $\gamma = 1.27$ W^{-1}km^{-1}.
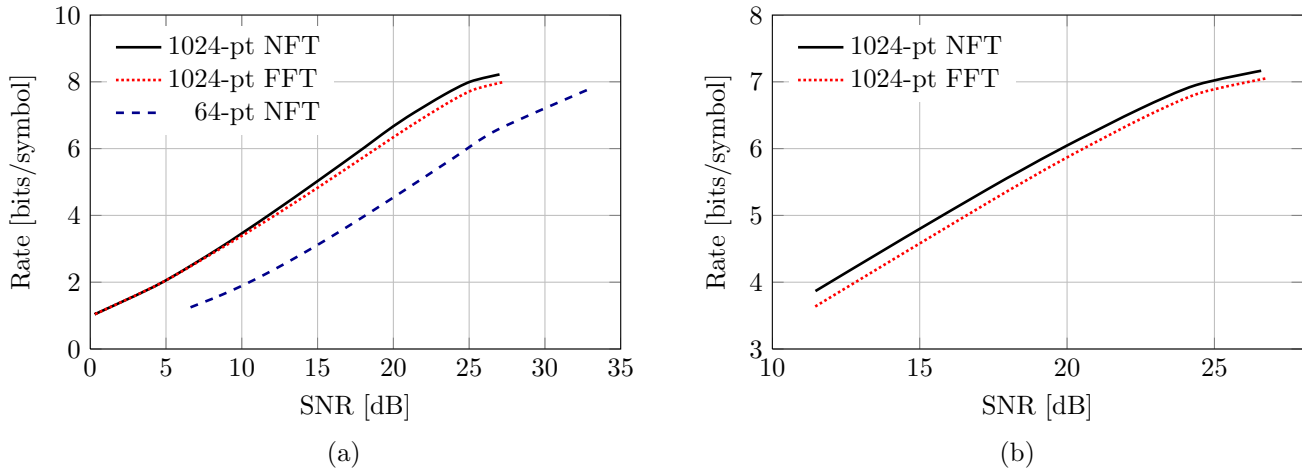
Fig. 9. Achievable rates in (a) a single-channel, and (b) a WDM optical fiber system using the nonlinear Fourier transform and backpropagation. The SNR is calculated at the system bandwidth and can be adjusted to represent the *optical* signal-to-noise ratio. Note that since we have used raised-cosine functions with 50% excess bandwidth, $\rho = \frac{2}{3}C$.

values for spectral amplitudes provides up to

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

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. The signal set should thus 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 $0 \le \lambda \le 2$ on the imaginary axis and create all N-solitons, $1 \le N \le 6$. 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 fiber system described at the beginning of Section V. 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 eigen-

values in a region not limited to the $j\omega$ axis, is likely to yield significantly higher spectral efficiencies.

The recent work of [32] 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 an optical 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

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

Fig. 9(a) shows the estimated achievable rates in 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 phase-shift keying 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 discretize the channel input. Capacity was computed via the Arimoto-Blahut 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 fast Fourier transform (FFT) implementation of the split-step Fourier method in the backpropagation scheme. The simulation is performed for a standard single-mode fiber with dispersion parameter $D = 17\,\text{ps}/(\text{nm} - \text{km})$. 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 compared to the traditional time data (amplitude and phase at the output of a matched filter).

This simulation was repeated for 5 WDM channels with the system architecture of Fig. 1. Here, low-pass filters and ROADMs are placed at the end of each fiber segment. The spectral efficiency in this case is obviously lower due to the inter-channel interference. Here too, NFT and FFT-based backpropagation produce approximately the same results.

From Fig. 9 it follows that at low SNRs NFT and backprop-agation achieve about the same rates. As the SNR is increased, the spectral efficiency of backpropagation degrades due to ISI (under the commonly-assumed isolated symbol detection, in which the memory is not accounted for) or inter-channel interference (in a network scenario). However, the spectral efficiency of the NFT scheme may be higher due to its inherent immunity to cross-talk, provided that users are multiplexed appropriately in the nonlinear frequency domain in a multiuser setup; see Section VI-E.

We have not yet simulated the spectral efficiency at SNRs beyond those shown in Figs. 9 (a)–(b) due to high numerical complexity. Up to the SNRs $25 - 30$ dB shown in Figs. 9 (a)–(b) signals only have a continuous spectrum. This made simulations doable. Beyond these SNRs, discrete mass points start to appear. This is in fact the SNR level where the nonlinearity starts to become significant (because, in the focusing regime, the discrete spectrum is the component of the NFT which is primarily responsible for the nonlinearity).

In the N-soliton transmission simulations of the previous subsections, we began with a desired spectral constellation at the transmitter. As a result, the initial conditions needed in the Newton algorithm at the receiver were known and the detection was computationally feasible. The difficulty there, however, was that the properties of the resulting signals in the time-domain (bandwidth and time duration) were not properly understood. In contrast, in this subsection we started with a signal set in the time domain *i.e.*, raised-cosine signals. When a signal is scaled according to a time-domain constellation, discrete mass points appear at unknown places in the complex plane. The difficulty here is that the properties of the resulting signals in the spectral domain are not properly understood. As a result, it is difficult to search for these discrete mass points without a priori information on their locations. The potential advantage of NFDM thus remains to be illustrated.

VI. DISCUSSION

In this section, we make a few remarks about the NFT method, clarify some of its properties and potential practical issues, and suggest some possible directions for future investigation.
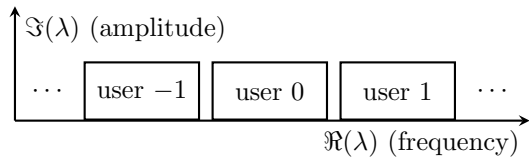

Fig. 10. Partitioning $\mathbb{C}^+$ for multiuser communication using the NFT.

A. Multiuser Communications Using the NFT

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 $\mathbb{C}^+$. It would then places the spectrum of the signals to be added in empty bands and produces the output signal by taking the inverse NFT. In this way, each user is assigned a region in the complex plane and, in the absence of noise, does not interfere with other users; see Fig. 10 as well as Section VI-B. In such a manner, NFDM results in an orthogonalization for the deterministic nonlinear Schrödinger channel.

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 introduce stochastic interference. In comparison, conventional WDM with backpropagation is subject to both deterministic and stochastic 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 similar 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 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 $\mathbb{C}^+$ 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 (6) and (7) 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

1) Advantages: Some of the advantages of using NFDM were outlined in Section I. In short, deterministic distortions, *i.e.*, distortions that arise even in a noise-free system (SPM, XPM, FWM, ISI and interference) are zero for all users of a multiuser network.

2) Disadvantages:

a) Aplicable only to Perfectly Integrable Models: 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:

- Using Raman amplification, the effects of loss are minimal (and indeed are traded with noise perturbation).
- 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.
- Mathematically one has stability results for solitons [33]. 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 with a nearby discrete spectrum in the complex plane).
- Considering that the performance of the WDM method degrades asymptotically with SNR, it may be worthwhile to identify sources of perturbations to integrability in NFDM and minimize them. This way, rather than engineering

the channel to be linear, the near-integrable channel is engineered to be integrable, so as to support its natural nonlinear modes, which characteristically tend to form in the medium.

b) High Complexity: The nonlinear Fourier coefficients at the receiver are calculated using $\mathcal{O}(n)$ operations per nonlinear frequency, where n is the number of signal samples in time. In comparison with with the $\mathcal{O}(n \log n)$ complexity of the FFT, the complexity of an n-point NFT is, at present, $\mathcal{O}(n^2)$ (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 even to simulate. It is therefore of interest to develop faster algorithms.

c) Hardware Limitations: 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 stochastic interference can be present in NFDM. As a result, although the achievable rate of NFDM in a multiuser channel can be higher than that of WDM with backpropagation (at least in the limit that noise goes to zero), it too may ultimately peak at some finite SNR. 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 subject to interference due to the lack of integrability[6].

Note that the decline of the achievable rates in WDM simulations in prior work is mostly due to the deterministic distortions. Our capacity simulations, as well as those in [4, Fig. 36], show that in the absence of noise the rates of the WDM method still vanish (or saturate, depending on assumptions on interfering users) at high powers in the network scenario. This is not the case for multiuser NFDM whose achievable rate is unbounded when noise is absent. We thus expect improvements in the information rates using NFDM, due to the immunity to the deterministic distortions.

F. Eigenvalue Communication of Hasegawa and Nyu

As noted in introduction, as well as in [Part I], the work of Hasegawa and Nyu on "eigenvalue communication" [9] is related to the NFT-based approach taken in this paper. Hasegawa and Nyu make the observation that the time-domain data is distorted while the eigenvalues remain constant and can be used for communication. The authors then considered single-user channels, encoded information in conserved quantities, and used the inverse scattering transform (IST) as a means to decode these conserved quantities. The idea is illustrated for pulses of the form $A \operatorname{sech}(t)$.

---

[6]Since, for instance, the NLS equation with a generic additive potential does not have conserved quantities.

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). In fact, it is clear that the *entire signal itself* is a conserved quantity under backpropagation and can be perfectly recovered. Thus the approach of [9] does not achieve anything more than what BP does. It is therefore not necessary to aim at extracting and modulating quantities conserved by a channel. Furthermore, the application of conserved quantities in a multiuser setup is in and of itself not useful, since conservation does not imply separability which underlies the NFDM approach.

In contrast, our primary motivation for introducing NFT-based methods stems from recent capacity studies [4] and our observation, made in Section II, that a major reason that the capacity rolls off in prior work, after abstracting away non-essential aspects, is predominantly because these methods, in essence, modulate linear-algebraic modes. When used in a nonlinear channel, this leads to a significant inter-channel interference and ISI. We noted that the NLS equation, however, supports nonlinear modes which have a crucial *independence property*, that can be used to build a interference-free multiuser system. The mathematical framework necessary to reveal signal degrees-of-freedom is the IST. As a result, in [Part I] we began with IST, shifted the focus from the scattering theory, considered IST as a Fourier transform, filtered its literature accordingly and presented a signal-and-systems perspective. This then paved the way for [Part II] and this paper whose overall goal was to construct NFDM, which can be viewed as a generalization of OFDM to nonlinear systems. The existence of an OFDM-like scheme for a nonlinear system is rather surprising. In [Part II] and this paper, we developed details of NFDM transmitter and receiver. Note that NFDM may not have eigenvalues as in [9]. Indeed the analogy between NFDM and OFDM can be best understood in the defocusing regime where, similarly to the ordinary Fourier transform, one has a transformation $\hat{q}(\lambda)$. This signal transformation underlies NFDM.

G. Linear Multiplexing Methods other than WDM

In this paper, we mostly focused on WDM as an example a linear multiplexing method. There are however other methods as well which, in essence, modulate linear-algebraic modes for transmission and behave similarly. Time-division multiplexing (TDM), OFDM and multi-mode communication are other examples.

Although mathematically these methods are not exact orthogonalizations, within a certain range of system and user parameters they can potentially be approximate orthogonalizations. For instance, at low powers where the nonlinearity is weak, each of n WDM users achieves $1/n$ of the degrees-of-freedom. Among these schemes, TDM is different in that, the dimension in which the multiplexing occurs, *i.e.*, time, can be practically unlimited. TDM can be capacity-achieving if a sufficient guard time, depending on the users' powers and

transmission distance, is introduced between users' signals. However, a system designed for one transmission distance may fail to operate at a larger distance: As the signal propagates further in distance, at some point orthogonalization is lost. Also, in principle, as the signal power tends to infinity, the transmission time of the user of interest tends to infinity as well — in effect, TDM turns the multiuser channel into a single-user one in order to address the interaction problem associated with the linear multiplexing. Ignoring practical issues in a network scenario, a TDM system designed for the worst-case system and user parameters can be nearly capacity-achieving in that regime. In fact, this is also true for WDM: if one has infinite bandwidth, users' signals can be widely separated such that in the range of system and user parameters their interaction is negligible.

The distinguishing feature of NFDM is that, deterministically, it is an exact orthogonalization for any transmission distance, signal power, dispersion or nonlinearity parameter. Users' signals may overlap in time and frequency, but they are separated in the nonlinear Fourier domain.

H. 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 [21], [34]–[36] — although it is not as fully developed as the continuous one. There are also important differences in the way that the spectrum is defined.

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 (13a) and (13b) are homogeneous equations in the order of derivatives that occur in each term, exponential functions are candidate solutions. Let us expand F and G as

$$F(t,z) = f_0(t,z) + \epsilon f_1(t,z) + \epsilon^2 f_2(t,z) + \cdots,$$
$$G(t,z) = g_0(t,z) + \epsilon g_1(t,z) + \epsilon^2 g_2(t,z) + \cdots,$$

for some small parameter ϵ. The bilinear terms in (13a)–(13b) are

$$FG = \sum_{n=0}^{\infty} \epsilon^n \left(\sum_{k=0}^{n} f_k g_{n-k} \right), \; FF = \sum_{n=0}^{\infty} \epsilon^n \left(\sum_{k=0}^{n} f_k f_{n-k} \right),$$
$$|G|^2 = \sum_{n=0}^{\infty} \epsilon^n \left(\sum_{k=0}^{n} g_k g_{n-k}^* \right).$$

Substituting these expressions into (13a)–(13b) and equating like powers of ϵ gives rise to a recursive procedure to obtain $\{f_{n+1}, g_{n+1}\}$ from $\{f_n, g_n\}$. To begin finding unknowns recursively, we can set initially $f_0 = 1$. The zero-order term ϵ^0 then gives $g_0 = 0$. Starting with $\{f_0 = 1, g_0 = 0\}$, the recursive equations are:

$$\begin{cases} H1.g_n = -\sum_{k=1}^{n-1} H(f_k g_{n-k}), \\ \partial_{tt} f_n = -\frac{1}{2} \sum_{k=1}^{n-1} D_t^2 f_k f_{n-k} + \gamma \sum_{k=1}^{n-1} g_k g_{n-k}^*. \end{cases}$$

where $H = jD_z + D_t^2$. It can be seen that $f_{2n+1} = g_{2n} = 0$. At each iteration n one of f_n or g_n is non-zero, which is used in the next iteration. For a general nonlinear system, iterations continue indefinitely. However, for integrable equations, as shown below for the NLS equation, this series truncates and exact solutions of various finite order are obtained.

Before proceeding, it is useful to know properties of the Hirota operators. Let $X_i = \zeta_i t - k_i z + \phi_i$, with dispersion relation $k_i = j\zeta_i^2$. Then:

1) $D_t^2 e^{X_i} e^{X_j} = D_t^2 e^{X_j} e^{X_i} = (\zeta_j - \zeta_i)^2 e^{X_i + X_j}$;
2) $D_z e^{X_i} e^{X_j} = -D_z e^{X_j} e^{X_i} = (k_j - k_i)e^{X_i + X_j}$;
3) $He^{X_i} e^{X_j} = \left(j(k_j - k_i) + (\zeta_j - \zeta_i)^2 \right) e^{X_i + X_j}$;
4) $H(1.e^X) = (jk + \zeta^2)e^X, \; X = \zeta t - kz + \phi$;
5) $H(e^{X_i} e^{X_j}) = \alpha_{ij} H(1.e^{X_i + X_j})$, where $\alpha_{ij} = \left(j(k_j - k_i) + (\zeta_j - \zeta_i)^2 \right) / \left(j(k_j + k_i) + (\zeta_j + \zeta_i)^2 \right)$;
6) $H(e^{X_i^*} e^{X_i^* + X_j}) = H(e^{X_i + X_j^*} e^{X_i}) = 0$;
7) $H(1.g) = He^{X_i} e^{X_j}$ has a solution $g = \alpha_{ij} e^{X_i + X_j}$.

The first iteration $n = 1$ is

$$\begin{cases} H1.g_1 = 0, \\ (f_1)_{tt} = 0, \end{cases}$$

which gives

$$g_1 = \sum_{i=1}^{N} e^{X_i}, \quad f_1 = 0,$$

where $N \in \mathbb{N}$ is the order of the multi-soliton solution.

We first consider $N = 1$ and then show how an additional mass point can be added to obtain the $N = 2$ solution.

A. Case N=1

We have $f_0 = 1$, $g_0 = f_1 = 0$ and $g_1 = e^{X_1}$. For $n = 2$

$$\begin{cases} H(1.g_2) = -H(f_1 g_1) = 0, \\ (f_2)_{tt} = -\frac{1}{2} D_t^2 f_1 f_1 + \gamma g_1 g_1^* = \gamma e^{X_1 + X_1^*}. \end{cases}$$

At this point, since one variable X_1 already exists and $N = 1$, we aim at truncating the process by choosing zero solutions and avoiding the inclusion of terms $\exp(X_i)$, $i \geq 2$. We thus consider the solution

$$g_2 = 0, \quad f_2 = e^{X_1 + X_1^* + R_{13}},$$

where $R_{13} = -2\log(\zeta_1 + \zeta_1^*) + \log\gamma$. It can be verified that the above solution then gives all other terms $f_k = g_k = 0$, $k \geq 3$, and the iteration truncates. The 1-soliton solution is obtained as

$$q = \frac{G}{F} = \frac{e^{X_1}}{1 + e^{X_1 + X_1^* + R_{13}}}$$
$$= \frac{1}{2} e^{-\frac{1}{2}R_{13}} e^{j\Im X_1} \text{sech}\left(\Re X_1 + \frac{1}{2}R_{13} \right).$$

By setting $\zeta_1 = -2j\lambda_1 = \omega - j\alpha$ where $\lambda_1 = (\alpha + j\omega)/2$ is the eigenvalue, $k_1 = j\zeta_1^2 = 2\alpha\omega - j(\alpha^2 - \omega^2)$, $\phi_1 = \log\frac{2\omega^2}{\tilde{q}}$, we get

$$q = \omega e^{-j\alpha t + j(\alpha^2 - \omega^2)z - j\angle\tilde{q}} \text{sech}\left(\omega(t - 2\alpha z - t_0) \right),$$

where $t_0 = \frac{1}{\omega}\log\frac{\gamma|\tilde{q}|}{\omega}$ and $\angle$ denotes phase.

B. Case N=2

For $n = 1$ we get $f_0 = 1$, $g_0 = f_1 = 0$ and $g_1 = e^{X_1} + e^{X_2}$. The next iteration $n = 2$ is

$$\begin{cases} H(1.g_2) = -Hf_1 g_1 = 0, \\ (f_2)_{tt} = -\frac{1}{2} D_t^2 f_1 f_1 + \gamma g_1 g_1^* \\ \quad = \gamma \left(e^{X_1 + X_1^*} + e^{X_1 + X_2^*} + e^{X_2 + X_1^*} + e^{X_2 + X_2^*} \right). \end{cases}$$

We choose the solution $g_2 = 0$ and

$$f_2 = e^{X_1 + X_1^* + R_{13}} + e^{X_1 + X_2^* + R_{14}} + e^{X_2 + X_1^* + R_{23}}$$
$$+ e^{X_2 + X_2^* + R_{24}},$$

where

$$R_{ij} = \begin{cases} 2\log(\zeta_i - \zeta_j) - \log\gamma, & i,j \leq N \text{ or } i,j \geq N+1, \\ -2\log(\zeta_i + \zeta_j) + \log\gamma, & i \leq N \text{ and } j \geq N+1, \end{cases}$$

and $\zeta_{i+N} = \zeta_i^*$.

The $n = 3$ iteration is

$$\begin{cases} H1.g_3 = -H(f_1g_2 + f_2g_1) = -Hf_2(e^{X_1} + e^{X_2}), \\ \partial_{tt}f_3 = 0. \end{cases}$$

The second equation gives $f_3 = 0$. The first one, using property 6), simplifies to

$$H(1.g_3) = -H(e^{X_1^* + X_2 + R_{23}}.e^{X_1}) - H(e^{X_2 + X_2^* + R_{24}}.e^{X_1}) \\ - H(e^{X_1^* + X_1 + R_{13}}.e^{X_2}) - H(e^{X_1 + X_2^* + R_{14}}.e^{X_2}).$$

Using property 7), a solution is

$$g_3 = -\Big(\alpha_{23,1}e^{X_1 + X_1^* + X_2 + R_{23}} + \alpha_{24,1}e^{X_1 + X_2 + X_2^* + R_{24}} \\ + \alpha_{13,2}e^{X_1 + X_1^* + X_2 + R_{13}} + \alpha_{14,2}e^{X_1 + X_2^* + X_2 + R_{14}} \Big) \\ = -\big(\alpha_{13,2}e^{R_{13}} + \alpha_{23,1}e^{R_{23}} \big) e^{X_1 + X_1^* + X_2} \\ - \big(\alpha_{14,2}e^{R_{14}} + \alpha_{24,1}e^{R_{24}} \big) e^{X_1 + X_2 + X_2^*}. \tag{31}$$

Here

$$\alpha_{(i+N)j,i} = \frac{j(k_i - (k_i^* + k_j))) + (\zeta_i - (\zeta_i^* + \zeta_j)^2}{j(k_i + (k_i^* + k_j))) + (\zeta_i + (\zeta_i^* + \zeta_j)^2} \\ = \frac{2\zeta_j^2 - 2|\zeta_i|^2 - 2\zeta_i\zeta_j + 2\zeta_i^*\zeta_j}{2\zeta_i^{*2} + 2|\zeta_i|^2 + 2\zeta_i\zeta_j + 2\zeta_i^*\zeta_j} \\ = \frac{\zeta_j - \zeta_i}{\zeta_i + \zeta_i^*},$$

and similarly,

$$\alpha_{(i+N)i,j} = \frac{j(k_j - (k_i^* + k_i))) + (\zeta_j - (\zeta_i^* + \zeta_i))^2}{j(k_j + (k_i^* + k_i))) + (\zeta_j + (\zeta_i^* + \zeta_i))^2} \\ = \frac{\zeta_i - \zeta_j}{\zeta_i^* + \zeta_j}.$$

The first coefficient $a = \alpha_{13,2}e^{R_{13}} + \alpha_{23,1}e^{R_{23}}$ in (31) is

$$a = \frac{\zeta_1 - \zeta_2}{\zeta_1^* + \zeta_2}e^{R_{13}} + \frac{\zeta_2 - \zeta_1}{\zeta_1 + \zeta_1^*}e^{R_{23}} \\ = \frac{\zeta_1 - \zeta_2}{\zeta_1^* + \zeta_2} \times \frac{\gamma}{(\zeta_1^* + \zeta_1)^2} + \frac{\zeta_2 - \zeta_1}{\zeta_1 + \zeta_1^*} \times \frac{\gamma}{(\zeta_2 + \zeta_1^*)^2} \\ = -\frac{\gamma(\zeta_1 - \zeta_2)^2}{(\zeta_1^* + \zeta_1)^2(\zeta_1^* + \zeta_2)^2} \\ = -e^{R_{12} + R_{13} + R_{23}}.$$

In the same manner, the second coefficient $b = \alpha_{14,2}e^{R_{14}} + \alpha_{24,1}e^{R_{24}}$ in (31) is obtained as

$$b = \frac{-\gamma(\zeta_1 - \zeta_2)^2}{(\zeta_1 + \zeta_2^*)^2(\zeta_2 + \zeta_2^*)^2} = -e^{R_{12} + R_{14} + R_{24}}.$$

It follows that

$$g_3 = e^{X_1 + X_1^* + X_2 + R_{12} + R_{13} + R_{23}} + e^{X_1 + X_2 + X_2^* + R_{12} + R_{14} + R_{24}}.$$

The $n = 4$ iteration is:

$$\begin{cases} H1.g_4 = 0, \\ (f_4)_{tt} = -\frac{1}{2}D_t^2(f_2f_2) + \gamma(g_3g_1^* + g_1g_3^*). \end{cases}$$

The first equation gives $g_4 = 0$. For the second one, using property 1), we have

$$\frac{1}{2}D_t^2 f_2.f_2 = \frac{1}{2}D_t^2\Big(e^{X_1 + X_1^* + R_{13}} + e^{X_1 + X_2^* + R_{14}} \\ + e^{X_2 + X_1^* + R_{23}} + e^{X_2 + X_2^* + R_{24}} \Big)^2 \\ = (\zeta_2^* - \zeta_1^*)^2 e^{2X_1 + X_1^* + X_2^* + R_{13} + R_{14}} \\ + (\zeta_2 - \zeta_1)^2 e^{X_1 + 2X_1^* + X_2 + R_{13} + R_{23}} \\ + (\zeta_2 + \zeta_2^* - \zeta_1 - \zeta_1^*)^2 e^{X_1 + X_1^* + X_2 + X_2^* + R_{13} + R_{24}} \\ + (\zeta_1^* + \zeta_2 - \zeta_1 - \zeta_2^*)^2 e^{X_1 + X_1^* + X_2 + X_2^* + R_{14} + R_{23}} \\ + (\zeta_2 - \zeta_1)^2 e^{X_1 + X_2 + 2X_2^* + R_{14} + R_{24}} \\ + (\zeta_2^* - \zeta_1^*)^2 e^{X_1^* + 2X_2 + X_2^* + R_{23} + R_{24}} \\ = (\zeta_2 + \zeta_2^* - \zeta_1 - \zeta_1^*)^2 e^{X_1 + X_1^* + X_2 + X_2^* + R_{13} + R_{24}} \\ + (\zeta_1^* + \zeta_2 - \zeta_1 - \zeta_2^*)^2 e^{X_1 + X_1^* + X_2 + X_2^* + R_{14} + R_{23}} \\ + \gamma\Big\{ e^{2X_1 + X_1^* + X_2^* + R_{13} + R_{14} + R_{34}} \\ + e^{X_1 + 2X_1^* + X_2 + R_{13} + R_{23} + R_{12}} \\ + e^{X_1 + X_2 + 2X_2^* + R_{14} + R_{24} + R_{12}} \\ + e^{X_1^* + 2X_2 + X_2^* + R_{23} + R_{24} + R_{34}} \Big\}. \tag{32}$$

Also

$$\gamma(g_3g_1^* + \text{c.c.}) = \gamma\Big[(e^{R_{13} + R_{23}} + e^{R_{14} + R_{24}})e^{R_{12}} + \text{c.c.} \Big] \\ \times e^{X_1 + X_1^* + X_2 + X_2^*} \\ + \gamma\Big\{ e^{X_1 + 2X_1^* + X_2 + R_{12} + R_{13} + R_{23}} \\ + e^{X_1 + X_2 + 2X_2^* + R_{12} + R_{14} + R_{24}} \\ + e^{2X_1 + X_1^* + X_2^* + R_{12}^* + R_{13}^* + R_{23}^*} \\ + e^{X_1^* + 2X_2 + X_2^* + R_{12}^* + R_{14}^* + R_{24}^*} \Big\}. \tag{33}$$

The terms inside braces in (32) and (33), involving $2X_1$, $2X_2$, $2X_1^*$ and $2X_2^*$, cancel out using identities $R_{12}^* = R_{34}$, $R_{13}^* = R_{13}$, $R_{14}^* = R_{23}$, $R_{24}^* = R_{24}$. The terms containing $\exp(X_1 + X_1^* + X_2 + X_2^*)$, after some algebra, simplify and we obtain

$$f_4 = e^{X_1 + X_1^* + X_2 + X_2^* + R_{12} + R_{13} + R_{14} + R_{23} + R_{24} + R_{34}}.$$

One can check that $f_k = g_k = 0$, $k \geq 5$. It follows that

$$G = e^{X_1} + e^{X_2} + e^{X_1 + X_1^* + X_2 + R_{12} + R_{13} + R_{23}} \\ + e^{X_1 + X_2 + X_2^* + R_{12} + R_{14} + R_{24}}, \\ F = 1 + e^{X_1 + X_1^* + R_{13}} + e^{X_1 + X_2^* + R_{14}} + e^{X_2 + X_1^* + R_{23}} \\ + e^{X_2 + X_2^* + R_{24}} \\ + e^{X_1 + X_1^* + X_2 + X_2^* + R_{12} + R_{13} + R_{14} + R_{23} + R_{24} + R_{34}}.$$

C. General Formula

The calculations in the previous subsections illustrate how one can add an additional term $\exp(X_N)$ to g_1 and update F and G. Using similar calculations, N-soliton solutions are obtained inductively. Furthermore, at this point the structure of F and G, noted earlier, becomes clear.

In the Hirota method, eigenvalues $\lambda_i = (\alpha_i + j\omega_i)/2$ are related to ζ_i by $\zeta_i = -2j\lambda_i = \omega_i - j\alpha_i$. For $N = 1$, the spectral amplitude is determined via $\phi_1 = \log(2\omega_1^2/\tilde{q}_1)$. While eigenvalues in the Riemann-Hilbert, Hirota and Darboux methods are the same, other spectral parameters are different.

APPENDIX B
PROOF OF THE DARBOUX THEOREM

See [22] for the proof of a more general theorem. Here we give a simple proof for Theorem 1.

Let $\phi(t, \lambda; q)$ be a known eigenvector associated with λ and q, *i.e.*, satisfying $\phi_t = P(\lambda, q)\phi$. Its adjoint $\tilde{\phi}(t, \lambda; q) = [\phi_2^*, -\phi_1^*]$ satisfies $\tilde{\phi}_t = P(\lambda^*, q)\tilde{\phi}$. Denote this known solution as $S = [\phi, \tilde{\phi}]$, $\Gamma = \mathrm{diag}(\lambda, \lambda^*)$, and $\Sigma = S\Gamma S^{-1}$.

We can verify that $S_t = JS\Gamma + QS$, where $J = \mathrm{diag}(-j, j)$ and $Q = \mathrm{offdiag}(q, -q^*)$. In addition we have $\Sigma_t = [J\Sigma + Q, \Sigma]$.

Given that $\phi(t, \lambda; q)$ is known, the Darboux transformation maps $\{v(t, \mu; q), \tilde{v}(t, \mu; q)\}$ to $\{u(t, \mu; \tilde{q}), \tilde{u}(t, \mu; \tilde{q})\}$ according to

$$U = V\Lambda - \Sigma V,$$

where $V = [v, \tilde{v}]$, $U = [u, \tilde{u}]$, $\Lambda = \mathrm{diag}(\mu, \mu^*)$.

We have $V_t = JV\Lambda + QV$ and

$$
\begin{aligned}
U_t &= V_t\Lambda - (\Sigma_t V + \Sigma V_t)\\
&= (JV\Lambda + QV)\Lambda - ([J\Sigma + Q, \Sigma]V + \Sigma(JV\Lambda + QV))\\
&= (JV\Lambda + QV)\Lambda - \Sigma JV\Lambda - ([J\Sigma + Q, \Sigma] + \Sigma Q)V\\
&= J(V\Lambda - \Sigma V)\Lambda + J\Sigma V\Lambda - \Sigma JV\Lambda\\
&\quad + QV\Lambda - ([J\Sigma + Q, \Sigma] + \Sigma Q)V\\
&= JU\Lambda + [J, \Sigma]V\Lambda - ([J\Sigma + Q, \Sigma] + \Sigma Q)V + QV\Lambda\\
&= JU\Lambda + [J, \Sigma]V\Lambda - (J\Sigma^2 + Q\Sigma - \Sigma J\Sigma)V + QV\Lambda\\
&= JU\Lambda + [J, \Sigma]V\Lambda - [J, \Sigma]\Sigma V - Q\Sigma V + QV\Lambda\\
&= JU\Lambda + [J, \Sigma](V\Lambda - \Sigma V) + Q(V\Lambda - \Sigma V)\\
&= JU\Lambda + (Q + [J, \Sigma])U\\
&= JU\Lambda + \tilde{Q}U,
\end{aligned}
$$

where $\tilde{Q} = Q + [J, \Sigma]$.

It follows that u and $\tilde{u}$ satisfy the P-equations $u_t = P(\mu, \tilde{q})u$ and $\tilde{u}_t = P(\mu^*, \tilde{q})\tilde{u}$. In the same manner we can show that u and $\tilde{u}$ satisfy the M-equations $u_z = M(\mu, \tilde{q})u$ and $\tilde{u}_z = M(\mu^*, \tilde{q})\tilde{u}$.

REFERENCES

[1] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools," *Arxiv e-prints,* arXiv:1202.3653, Feb. 2012. [Online]. Available: http://arxiv.org/abs/1202.3653

[2] ——, "Information transmission using the nonlinear Fourier transform, Part II: Numerical methods," *Arxiv e-prints,* arXiv:1204.0830, Apr. 2012. [Online]. Available: http://arxiv.org/abs/1204.0830

[3] L. F. Mollenauer and J. P. Gordon, *Solitons in Optical Fibers: Fundamentals and Applications*, 1st ed. Amsterdam, The Netherlands: Elsevier 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. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[5] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, ser. SIAM Stud. Appl. Numer. Math. Philadelphia, PA, USA: SIAM, 1981, vol. 4.

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

[7] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, ser. Oxford Series in Optical and Imaging Sciences. Oxford, U.K.: Clarendon Press, 1995, vol. 7.

[8] A. C. Singer, "Signal processing and communication with solitons," Ph.D. dissertation, Dept. Electr. Eng., Massachusetts. Inst. of Technol., Cambridge, MA, USA, 1996.

[9] A. Hasegawa and T. Nyu, "Eigenvalue communication," *IEEE J. Lightw. Technol.*, vol. 11, no. 3, pp. 395–399, Mar. 1993.

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

[11] G. P. Agrawal, *Nonlinear Fiber Optics*, 5th ed. San Francisco, CA, USA: Academic Press, 2012.

[12] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.

[13] E. Agrell, "Conditions for a monotonic channel capacity," *ArXiv e-prints,* arXiv:1209.2820v2, Feb. 2014. [Online]. Available: http://arxiv.org/abs/1209.2820v2

[14] G. Kramer, "Review of rate regions for interference channels," in *2006 Int. Zurich Seminar on Commun. (IZS)*. IEEE, Feb. 2006, pp. 162–165.

[15] V. 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, pp. 3425–3441, Aug. 2008.

[16] T. Koch, A. Lapidoth, and P.-P. Sotiriadis, "Channels that heat up," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3594–3612, Aug. 2009.

[17] B. W. Göbel, "Information-theoretic aspects of fiber-optic communication channels," Ph.D. dissertation, Techn. U. München, 2010.

[18] A. Mecozzi and R.-J. Essiambre, "Nonlinear Shannon limit in pseudo-linear coherent systems," *IEEE J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.

[19] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *IEEE J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.

[20] R. Hirota, *The Direct Method in Soliton Theory*, ser. Cambridge Tracts in Mathematics. Cambridge, U.K.: Cambridge University Press, 2004, vol. 155.

[21] A. R. Osborne, *Nonlinear Ocean Waves and the Inverse Scattering Transform*, 1st ed., ser. Int. Geophys. New York, NY, USA: Academic Press, 2010, vol. 97.

[22] V. B. Matveev and M. A. Salle, *Darboux Transformations and Solitons*, ser. Springer Series in Nonlinear Dynamics. New York, NY, USA: Springer-Verlag, 1991.

[23] M. J. Ablowitz, B. Prinari, and A. D. Trubatch, *Discrete and Continuous Nonlinear Schrödinger Systems*, 1st ed., ser. Lond. Math. Soc. Lec. Note Series. Cambridge, U.K.: Cambridge University Press, 2003, vol. 302.

[24] V. I. Karpman and V. V. Solov'ev, "A perturbational approach to the two-soliton systems," *Physica D: Nonl. Phenom.*, vol. 3, no. 3, pp. 487–502, 1981.

[25] 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, Mar. 1977.

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

[27] V. V. Konotop and L. Vázquez, *Nonlinear Random Waves*. Singapore: World Scientific, 1994.

[28] P. Kazakopoulos and A. L. Moustakas, "Nonlinear Schrödinger equation with random Gaussian input: Distribution of inverse scattering data and eigenvalues," *Phys. Rev. E*, vol. 78, no. 1, pp. 016603 (1–7), Jul. 2008.

[29] G. Falkovich, I. Kolokolov, V. Lebedev, V. Mezentsev, and S. Turitsyn, "Non-Gaussian error probability in optical soliton transmission," *Physica D: Nonl. Phenom.*, vol. 195, nos. 1–2, pp. 1–28, Aug. 2004.

[30] 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. on Inf. Theory (ISIT 2010)*, Jun. 2010, pp. 206–210.

[31] ——, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[32] E. Meron, M. Feder, and M. Shtaif, "On the achievable communication rates of generalized soliton transmission systems," *Arxiv e-prints,* arXiv:1207.0297, Jul. 2012. [Online]. Available: http://arxiv.org/abs/1207.0297

[33] T. Tao, "Why are solitons stable?" *Bull. Amer. Math. Soc.*, vol. 46, no. 1, pp. 1–33, Jan. 2009.

[34] Y. C. Ma and M. J. Ablowitz, "The periodic cubic Schrödinger equation," *Stud. Appl. Math.*, vol. 65, no. 2, pp. 113–158, Oct. 1981.

[35] E. R. Tracy, "Topics in nonlinear wave theory with applications," Ph.D. dissertation, U. of Maryland, College Park, 1984.

[36] E. R. Tracy and H. H. Chen, "Nonlinear self-modulation: An exactly solvable model," *Phys. Rev. A*, vol. 37, no. 3, pp. 815–839, Feb. 1988.