Marquette University
e-Publications@Marquette

MSCS Faculty Research and Publications

Mathematics, Statistics and Computer Science,
Department of

4-1-2010

Importance Sampling for Dispersion-managed Solitons

Elaine T. Spiller
Marquette University, elaine.spiller@marquette.edu

Gino Biondini
State University of New York, Buffalo

Importance Sampling for Dispersion-Managed Solitons[*]

Elaine T. Spiller[†] and Gino Biondini[‡]

Abstract. The dispersion-managed nonlinear Schrödinger (DMNLS) equation governs the long-term dynamics of systems which are subject to large and rapid dispersion variations. We present a method to study large, noise-induced amplitude and phase perturbations of dispersion-managed solitons. The method is based on the use of importance sampling to bias Monte Carlo simulations toward regions of state space where rare events of interest—large phase or amplitude variations—are most likely to occur. Implementing the method thus involves solving two separate problems: finding the most likely noise realizations that produce a small change in the soliton parameters, and finding the most likely way that these small changes should be distributed in order to create a large, sought-after amplitude or phase change. Both steps are formulated and solved in terms of a variational problem. In addition, the first step makes use of the results of perturbation theory for dispersion-managed systems recently developed by the authors. We demonstrate this method by reconstructing the probability density function of amplitude and phase deviations of noise-perturbed dispersion-managed solitons and comparing the results to those of the original, unaveraged system.

1. Introduction. Noise is one of the major factors limiting the performance of coherent optical fiber communication systems [5, 17, 18, 21] as well as certain mode-locked lasers [22, 31, 41]. In both kinds of systems the noise is typically small compared to the signal, and so are the noise-induced perturbations to the pulse parameters on average. Those rare events in which many small noise-induced parameter changes build up coherently, however, are precisely the ones that are most likely to lead to significant pulse distortion and thus to system failures. Because optical fiber communication systems and Ti:sapphire femtosecond lasers are both designed to operate with very high accuracies, failures in these systems result from the occurrence of unusually large deviations, which makes calculating their error rates a challenging task. On one hand, direct Monte Carlo (MC) computations of failure rates are impractical due to the number of samples necessary to obtain a reliable estimate. On the other hand, analytical estimates are extremely difficult both because the noise couples nonlinearly to the signal during propagation and because solutions of (unforced) dispersion managed systems are often only available numerically.

[†]Department of Mathematics, Statistics, and Computer Science, Marquette University, Milwaukee, WI 53201 (elaine.spiller@marquette.edu).
[‡]Department of Mathematics, State University of New York, Buffalo, NY 14260 (biondini@buffalo.edu).

The effect of noise on systems modeled by the nonlinear Schrödinger (NLS) equation can be studied using importance sampling (IS) [29, 30, 37, 38, 39]. The idea behind IS is to sample the noise from a biased distribution that makes the rare events occur more frequently than would naturally be the case, while simultaneously correcting for the biasing. In order to successfully implement IS, however, one must bias toward the most likely noise realizations that lead to system failures. In the above works, this was achieved by taking advantage of well-known soliton perturbation theory results. No inverse scattering transform exists, however, for dispersion-managed systems. The response of dispersion-managed solitons subject to noise was studied in [19, 24, 25, 26, 36]. Here, as in [23], we address this problem by employing the dispersion-managed NLS (DMNLS) equation, which governs the long-term dynamics of dispersion-managed optical systems [1, 4, 12]. In [23] we developed a perturbation theory for dispersion-managed solitons, and we presented some importance-sampled MC (ISMC) simulations for which the choice of biasing was straightforward. Here we first discuss in detail the methods that are needed for the actual implementation of IS, namely, the precise formulation of the optimal biasing problem which utilizes the perturbation theory for dispersion-managed solitons. We then extend the previous results by actually solving the optimal biasing problem for all of the four dispersion-managed soliton parameters, and we use the results to perform ISMC numerical simulations to reconstruct the probability density function (PDF) of the final amplitude *as well as* the phase of the dispersion-managed solitons subject to noise, in which case the choice of biasing is nontrivial.

The outline of this work is the following. In section 2 we define more precisely the problem of interest, we briefly recall how the DMNLS equation arises as a model that describes the long-term dynamics of dispersion-managed systems, we review the results of perturbation theory for the DMNLS equation (using this occasion to correct a few typos in the original work), and we use these results to solve the stochastic differential equations (SDEs) for the evolution of the soliton parameters under the effect of noise in the limit of small deviations, obtaining the variance of the noise-induced parameter fluctuations. In section 3 we briefly review the main idea behind IS and we define the optimal biasing problem, namely, the problem of finding the most likely noise realizations that produce a given parameter change at the output. In section 4 we then solve the optimal biasing problem for each of the four soliton parameters. In section 5 we describe the ISMC simulations of the DMNLS equation perturbed by noise, and we compare the results to those arising from standard MC simulations of the original, unaveraged system. Finally, section 6 offers some concluding remarks.

2. Noise-perturbed dispersion-managed systems. The propagation of light pulses in dispersion-managed optical fiber communication systems [28] and Ti:sapphire femtosecond lasers [41] is effectively described by an NLS equation with distance-dependent coefficients plus a source term:

$$(2.1) \qquad i\frac{\partial q}{\partial z} + \frac{1}{2}\, d(z/z_a)\, \frac{\partial^2 q}{\partial t^2} + g(z/z_a)|q|^2 q = iS(t,z).$$

Here z is the dimensionless propagation distance, t is the dimensionless retarded time, $q(t,z)$ is the dimensionless slowly varying electric field envelope (rescaled to account for loss and amplification in communication systems), $d(z/z_a)$ is the local value of dispersion, and $g(z/z_a)$

describes the periodic power variations, which are due to loss and amplification in communication systems and to the fact that nonlinearity in Ti:sapphire lasers is present only inside the crystal and not the whole cavity. The choice of $d(z/z_a)$ is called a dispersion map, and z_a is the dispersion map period, which in lasers is the total length of the cavity. The source term $S(t, z)$ can represent various kinds of perturbations. Here, we will focus on the physically interesting case where the pulses are subject to spontaneous emission noise, which originates from the optical amplifiers in communication systems and from the gain medium in Ti:sapphire lasers. That is, we take (as in [2, 15])

$$(2.2) \qquad S(t, z) = \sum_{n=1}^{N_a} v_n(t)\delta(z - nz_a),$$

where N_a is the number of dispersion maps traversed (and hence the number of amplifiers in communication systems), z_a is the dispersion map period, $\delta(z)$ is the Dirac delta distribution, and $\nu_n(t)$ is white Gaussian noise, satisfying

$$(2.3) \qquad \mathbb{E}[v_n(t)] = 0, \quad \mathbb{E}[v_n(t)v_{n'}^*(t')] = \sigma^2\delta(t - t')\delta_{nn'},$$

where $\mathbb{E}[\cdot]$ denotes the ensemble average and σ^2 is the dimensionless noise variance. That is, at each amplifier we have the jump condition

$$(2.4) \qquad q(t, nz_a^+) = q(t, nz_a^-) + \sigma v_n(t)$$

(obtained by integrating (2.1) in a neighborhood of $z = nz_a$). The amount of noise added at each amplifier is typically very small (that is, $\sigma \ll 1$), and therefore the noise-induced parameter fluctuations are also typically small. On rare occasions, however, many small contributions accumulate to produce large deviations at the output. Since these events are by definition rare and atypical, it is very difficult to estimate precisely how frequently they occur. The problem at hand is to obtain the PDF of the noise-induced parameter fluctuations at the output.

2.1. The DMNLS equation and its soliton solution. Generically, (2.1) contains large and rapidly varying terms. As a consequence, the long-term behavior of its solutions is difficult to study. Fortunately, a different approach is possible. It has been shown that, once the compression/expansion cycle of the pulse in each dispersion map is properly factored out, the core pulse shape obeys the DMNLS equation [1, 12]. Namely, in the Fourier domain it is

$$(2.5) \qquad \hat{q}(\omega, z) = \hat{u}(\omega, z)\,e^{-iC(z/z_a)\omega^2/2},$$

with

$$(2.6) \qquad C(z/z_a) = \int_0^z \left(d(z'/z_a) - \bar{d}\right)\mathrm{d}z',$$

where $\bar{d}$ is the average dispersion. Hereafter, $\hat{f}(\omega, z)$ denotes the Fourier transform of a generic function $f(t, z)$, defined as

$$(2.7) \qquad \hat{f}(\omega, z) = \int f(t, z)\,e^{-i\omega t}\,\mathrm{d}t.$$

Throughout this work, integrals are complete unless limits are explicitly given. The exponential factor in (2.5) accounts for the rapid breathing (compression and expansion cycle) of the pulse, while the slowly varying core $\hat{u}(\omega, z)$ satisfies a perturbed DMNLS equation:

$$(2.8a) \qquad i\frac{\partial u}{\partial z} + \frac{1}{2}\bar{d}\frac{\partial^2 u}{\partial t^2} + \iint u_{(t+t')}u_{(t+t'')}u^*_{(t+t'+t'')}R_{(t't'')}\,\mathrm{d}t'\mathrm{d}t'' = iS(t, z),$$

or, in the Fourier domain,

$$(2.8b) \qquad i\frac{\partial \hat{u}}{\partial z} - \frac{1}{2}\bar{d}\omega^2\hat{u} + \iint \hat{u}_{(\omega+\omega')}\hat{u}_{(\omega+\omega'')}\hat{u}^*_{(\omega+\omega'+\omega'')}r_{(\omega'\omega'')}\,\mathrm{d}\omega'\mathrm{d}\omega'' = i\hat{S}(\omega, z),$$

where the asterisk denotes complex conjugate, and where for brevity we denote $u_{(t)} = u(t, z)$, $\hat{u}_{(\omega)} = \hat{u}(\omega, z)$, etc. The kernels $R(t', t'')$ and $r(\xi)$ quantify the average nonlinearity over a dispersion map, mitigated by dispersion management:

$$(2.9) \qquad R(t't'') = \iint e^{i\omega't'+i\omega''t''}r(\omega'\omega'')\,\mathrm{d}\omega'\mathrm{d}\omega'', \qquad r(\xi) = \frac{1}{(2\pi)^2}\int_0^1 e^{iC(\zeta)\xi}g(\zeta)\,\mathrm{d}\zeta.$$

(A straightforward calculation shows that, like $r(\xi)$, $R(\cdot)$ depends only on the product $t't''$ and not on t' and t'' separately.) In what follows we will neglect the periodic power variations and take $g(z) = 1$. This is a good approximation in systems with distributed amplification [6].

The DMNLS equation and its solutions depend parametrically on a quantity s, called the *reduced map strength*, which quantifies the size of the dispersion variations around the mean. That is,

$$(2.10) \qquad s = \frac{1}{4z_a}\int_0^{z_a}\left|d(z/z_a) - \bar{d}\right|\mathrm{d}z.$$

As $s \to 0$, it is $r(\xi) \to 1/(2\pi)^2$, $R(t't'') \to \delta(t)\delta(t')$, and the DMNLS equation (2.8a) reduces to the NLS equation with constant coefficients. A special but physically interesting case is that of a piecewise constant, two-step dispersion map, namely,

$$(2.11) \qquad d(z/z_a) = \begin{cases} \bar{d} + 2s/(\theta z_a), & 0 \le z < \theta z_a, \\ \bar{d} - 2s/[(1-\theta)z_a], & \theta z_a \le z < 1, \end{cases}$$

for some $0 < \theta < 1$. In this case the kernels assume a particularly simple form [1]:

$$(2.12) \qquad r(\xi) = \frac{1}{(2\pi)^2}\frac{\sin s\xi}{s\xi}, \qquad R(\xi) = \frac{1}{2\pi|s|}\,\mathrm{ci}(\xi/s).$$

Note that the above definition of map strength s, which is consistent with [1, 3], differs from another common definition (e.g., cf. [3, 35]). In particular, the map strength as defined here is only a property of the fiber and does not change with a pulse's width.

Dispersion-managed solitons are traveling-wave solutions of the DMNLS equation. If

$$(2.13) \qquad u_o(t, z; s) = f(t; s)\,e^{i\lambda^2 z/2},$$

then $\hat{f}(\omega; s)$ satisfies the following nonlinear integral equation:

$$(2.14) \qquad \left(\lambda^2 + \bar{d}\omega^2\right) \hat{f}_{(\omega)} = 2 \iint \hat{f}_{(\omega+\omega')} \hat{f}_{(\omega+\omega'')} \hat{f}^*_{(\omega+\omega'+\omega'')} r_{(\omega'\omega'')} \, d\omega' d\omega''.$$

Equation (2.14) can be efficiently solved numerically, as discussed in [23]. The invariances of the DMNLS equation then yield from $u_o(t, z; s)$ a four-parameter family of dispersion-managed solitons:

$$(2.15) \qquad u_{\mathrm{dms}}(t, z; s) = A\, f(A(t-T); A^2 s)\, e^{i\Theta(t,z)}$$

(where $\lambda = 1$ was used without loss of generality), where A and Ω are the dispersion-managed soliton amplitude and frequency, $\Theta(t, z) = \Omega(t - T) + \Phi$ is the local phase, and

$$(2.16a) \qquad T(z) = \bar{d}\Omega z + t_o,$$

$$(2.16b) \qquad \Phi(z) = (A^2 + \bar{d}\Omega^2)z/2 + \phi_o$$

are, respectively, the mean time and the mean phase. (The present definition of Φ differs from that in [23, 30] and is chosen so that Φ is the phase at the center of the pulse.) It should be noted that $f(t; s)$ need not necessarily be real and even: one can insert an initial guess with a nonzero value of phase and mean position and obtain a perfectly valid solution with nonzero values of T and Φ thanks to the invariances of the integral equation and the well-known translation properties of Fourier transforms. This property is expedient in the numerical simulations.

2.2. Perturbation theory for the DMNLS equation. If $u(t, z) = u_{\mathrm{dms}} + w$ solves the perturbed DMNLS equation (2.8a) and $w = o(u_{\mathrm{dms}})$, then $w(t, z)$ satisfies the perturbed linearized DMNLS equation

$$(2.17a) \qquad L[w, u] = S(t, z),$$

where $L[w, u]$ is the linearized DMNLS operator [33]:

$$L[w, u] = \frac{\partial w}{\partial z} - \frac{i}{2}\bar{d}\frac{\partial^2 w}{\partial t^2} - 2i \iint u_{(t+t'')} u^*_{(t+t'+t'')} w_{(t+t')} R_{(t',t'')} \, dt' dt''$$

$$(2.17b) \qquad\qquad - i \iint u_{(t+t')} u_{(t+t'')} w^*_{(t+t'+t'')} R_{(t',t'')} \, dt' dt''.$$

The operator $L[u, v]$ admits a nontrivial nullspace (see below), which is related to the invariances of the DMNLS equation, as shown in [23]. Because of the existence of this nontrivial nullspace, secularities will arise in the solution of the perturbed linearized DMNLS equation (2.17a). The secular terms correspond to the portions of the noise components that alter the soliton parameters, as shown in [23]. The parameter changes are found as usual by removing secular terms, i.e., by requiring that $w(t, z)$ remain bounded in time. This yields [23]

$$(2.18a) \qquad \frac{dA}{dz} = S_A(z), \qquad \frac{d\Omega}{dz} = S_\Omega(z), \qquad \frac{dT}{dz} = \bar{d}\Omega + S_T(z),$$

$$(2.18b) \qquad \frac{d\Phi}{dz} = \frac{1}{2}(A^2 + \bar{d}\Omega^2) + \Omega\, S_T(z) + S_\Phi(z),$$

where the source terms are, for $Q = A, \Omega, T, \Phi$,

$$(2.19) \qquad S_Q(z) = \langle \mathrm{e}^{i\Theta} \underline{y}_Q, S \rangle / \langle \underline{y}_Q, y_Q \rangle,$$

and y_Q and $\underline{y}_Q$ are the neutral modes and their adjoints, as defined below.

In the special case of noise-induced solution perturbations, or of any other changes in the initial condition of the type

$$(2.20) \qquad u(t, nz_a^+) = u(t, nz_a^-) + \Delta u_n(t),$$

the induced dispersion-managed soliton parameter changes at each map period are [23]

$$(2.21) \qquad Q(nz_a^+) = Q(nz_a^-) + \Delta Q_n,$$

where for $Q = A, \Omega, T$ it is

$$(2.22a) \qquad \Delta Q_n = \frac{\langle \mathrm{e}^{i\Theta} \underline{y}_Q, \, \Delta u_n(t) \rangle}{\langle \underline{y}_Q, y_Q \rangle}$$

while

$$(2.22b) \qquad \Delta \Phi_n = \frac{\langle \mathrm{e}^{i\Theta} \underline{y}_\Phi, \, \Delta u_n(t) \rangle}{\langle \underline{y}_\Phi, y_\Phi \rangle} + \Omega \frac{\langle \mathrm{e}^{i\Theta} \underline{y}_T, \, \Delta u_n(t) \rangle}{\langle \underline{y}_\Omega, y_\Omega \rangle},$$

and where $\langle f, g \rangle$ is the inner product of two functions, defined as

$$(2.23) \qquad \langle f, g \rangle = \mathrm{Re} \int f^*(t) g(t) \, \mathrm{d}t.$$

Above and throughout this work, $y_A(t)$ and $y_\Omega(t)$ are the neutral eigenmodes, and $y_T(t)$ and $y_\Phi(t)$ are the generalized eigenmodes of the linearized DMNLS operator around the soliton solution (2.15) once the exponential phase has been removed (cf. [23] for further details), while $\underline{y}_A(t)$, $\underline{y}_\Omega(t)$, $\underline{y}_T(t)$, and $\underline{y}_\Phi(t)$ are the corresponding adjoint modes, that is, the modes of the adjoint linearized DMNLS operator. Explicitly [23],

$$(2.24a) \qquad y_\Phi = iU, \qquad y_T = -\frac{\partial U}{\partial \xi}, \qquad y_\Omega = i\xi U, \qquad y_A = \frac{1}{A}\left(U + \xi \frac{\partial U}{\partial \xi} + 2s \frac{\partial U}{\partial s}\right),$$

$$(2.24b) \qquad \underline{y}_\Phi = \frac{i}{A}\left(U + \xi \frac{\partial U}{\partial \xi} + 2s \frac{\partial U}{\partial s}\right), \qquad \underline{y}_T = \frac{\xi}{A} U, \qquad \underline{y}_\Omega = -\frac{i}{A}\frac{\partial U}{\partial \xi}, \qquad \underline{y}_A = U,$$

where $\xi = t - T(z)$, and $U(t, z) = u(t, z)\,\mathrm{e}^{-i\Theta} = A\,f(A\xi(t, z); A^2 s)$ is the envelope of the dispersion-managed soliton solution (2.15). These modes and adjoint modes are biorthogonal, that is,

$$(2.25) \qquad \langle \underline{y}_Q, y_{Q'} \rangle = \langle \underline{y}_Q, y_Q \rangle \, \delta_{QQ'},$$

where $\delta_{QQ'}$ is the Kronecker delta, and they form a basis of the nullspace of the linearized DMNLS operator. (Note that, unlike the NLS equation, the DMNLS equation also admits
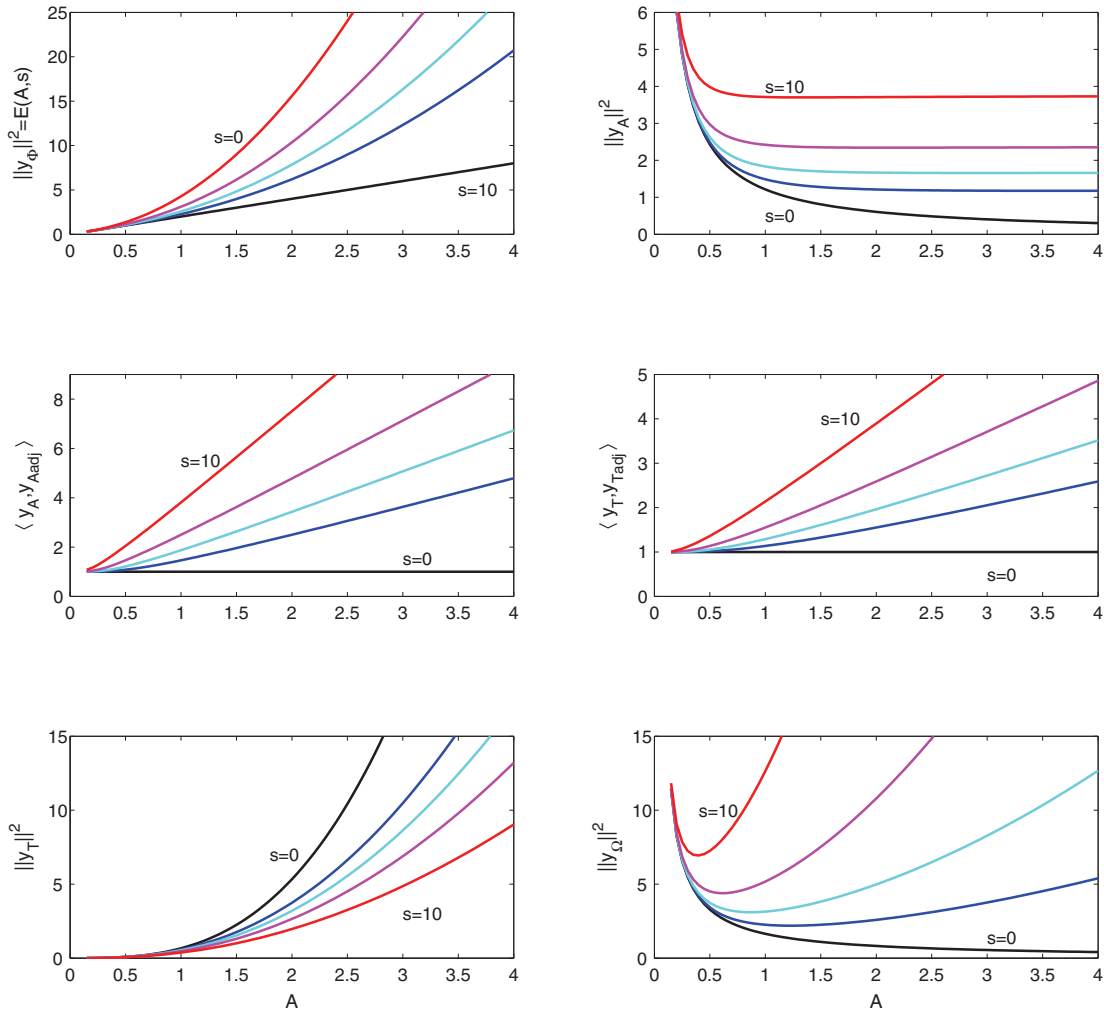
Figure 1. *Norms and inner products of the linear modes and their adjoints as a function of the soliton amplitude for different values of map strength: $s = 0$ (black), $s = 1$ (blue), $s = 2$ (cyan), $s = 4$ (magenta), and $s = 10$ (red). For the case $s = 0$, corresponding to the constant dispersion case (NLS equation), these functions are known explicitly; cf. (2.28). Note also that $\|y_\Phi\|^2 = \|\underline{y}_A\|^2 = E(A; s)$ provides a map that will be useful in the numerical simulations.*

internal modes [8, 20, 40]. These, however, do not play a role in our discussion.) All of these modes are generated using the symmetries of DMNLS, and all of these results reduce to the standard perturbation theory for NLS when $s = 0$. When $s \neq 0$, however, the inner products depend on both the soliton amplitude and the map strength, as shown in Figure 1. In particular, it is [23]

$$(2.26) \qquad \langle \underline{y}_A, y_A \rangle = \langle \underline{y}_\Phi, y_\Phi \rangle = \left(\frac{1}{2} + s \frac{\partial}{\partial s} \right) E/A, \qquad \langle \underline{y}_T, y_T \rangle = \langle \underline{y}_\Omega, y_\Omega \rangle = \frac{1}{2} E/A,$$

where $E(A; s) = \|u\|^2 = \int |u(t, z)|^2 \, dt$ is the pulse energy. Figure 1 shows the above norms and inner products for $s \neq 0$ as a function of A for various values of map strength. For the NLS equation (that is, when $s = 0$), it is $E = 2A$, so the above modes and adjoint modes become orthonormal:

$$(2.27) \qquad \langle \underline{y}_A, y_A \rangle = \langle \underline{y}_\Omega, y_\Omega \rangle = \langle \underline{y}_T, y_T \rangle = \langle \underline{y}_\Phi, y_\Phi \rangle = 1.$$

The norms of the linear modes and their adjoints also take on a simple form for the NLS equation:

$$(2.28) \quad \|y_A\|^2 = (12+\pi^2)/(18A), \qquad \|\underline{y}_A\|^2 = 2A, \qquad \|y_\Omega\|^2 = \pi^2/(6A), \qquad \|\underline{y}_\Omega\|^2 = 2A/3.$$

The remaining norms are obtained from these via (2.24).

As discussed in [23], the linear modes are related to changes in the solution parameters, and their norms and inner products will be important in solving the optimal biasing problem for IS, which we will turn to in sections 3 and 4.

2.3. Noise-induced changes and soliton parameter variances. When the perturbing term $S(t, z)$ in (2.8a) represents amplifier noise, (2.18) becomes a system of nonlinear SDEs for the evolution of the dispersion-managed soliton parameters in the presence of noise. Recall that the source terms in those SDEs are given by (2.19). For the present discussion (as well as in section 4), we employ a continuum approximation of the noise. That is, we consider $S(t, z)$ to be a Gaussian white-noise process with zero mean and autocorrelation function

$$(2.29) \qquad \mathbb{E}[S(t, z)S^*(t, z)] = \sigma^2 \, \delta(t - t')\delta(z - z').$$

(An alternative treatment that preserves the discreteness of the amplifier locations is described in [30] for the NLS equation. But note also that for systems with distributed amplification noise is indeed added continuously throughout the transmission line.) Since the adjoint modes are orthogonal, the sources $S_Q(z)$ are independent white-noise processes, with autocorrelation function

$$(2.30) \qquad \mathbb{E}[S_Q(z)S_{Q'}^*(z')] = \sigma_Q^2 \delta(z - z'),$$

where

$$(2.31) \qquad \sigma_Q^2 = \mathrm{var}[S_Q(z)] = \mathbb{E}\big[\langle e^{i\Theta} \underline{y}_Q, S \rangle^2 / \langle \underline{y}_Q, y_Q \rangle^2\big] = \sigma^2 \|\underline{y}_Q\|^2 / \langle \underline{y}_Q, y_Q \rangle^2.$$

In the case of the NLS equation, (2.27) and (2.28) yield these variances explicitly as

$$(2.32) \quad \sigma_A^2/\sigma^2 = 2A, \quad \sigma_\Omega^2/\sigma^2 = 2A/3, \quad \sigma_T^2/\sigma^2 = \pi^2/(6A^3), \quad \sigma_\Phi^2/\sigma^2 = (12 + \pi^2)/(18A).$$

For all values of s, these variances depend on the soliton amplitude A (as well as on the map strength s) and therefore on the propagation distance z. As a result, it is not possible to integrate the SDEs (2.18b) in closed form, even in the case of constant dispersion.

In those situations where the amplitude deviations are not large, one can approximate the variances $\sigma_A^2, \ldots, \sigma_\Phi^2$ of the source terms as constant. Equations (2.18b) can then be integrated exactly, to obtain

$$(2.33a) \qquad A(z) = A_o + W_A(z), \qquad \Omega(z) = \Omega_o + W_\Omega(z),$$

$$(2.33b) \qquad T(z) = T_o + \int_0^z \bar{d}\Omega(z') \, \mathrm{d}z' + W_T(z),$$

$$(2.33c) \qquad \Phi(z) = \frac{1}{2} \int_0^z \left(A^2(z') + \bar{d}\Omega^2(z') \right) \mathrm{d}z' + \int_0^z \Omega(z') S_T(z') \, \mathrm{d}z' + W_\Phi(z),$$

where, for $Q = A, \Omega, T, \Phi$,

$$(2.34a) \qquad W_Q(z) = \int_0^z S_Q(z') \, \mathrm{d}z'$$

is a Wiener process with zero mean and autocorrelation function

$$(2.34b) \qquad \mathbb{E}[W_Q(z)W_{Q'}(z')] = \sigma_Q^2 \delta_{QQ'} \min(z, z').$$

The mean values of the soliton parameters at the output are then

$$(2.35a) \qquad \mathbb{E}[A(L)] = A_o, \qquad \mathbb{E}[\Omega(L)] = \Omega_o, \qquad \mathbb{E}[T(L)] = T_o + \bar{d}\Omega_o L,$$

$$(2.35b) \qquad \mathbb{E}[\Phi(L)] = \frac{1}{2}(A_o^2 + \bar{d}\Omega_o^2) L + \frac{1}{4}(\sigma_A^2 + \bar{d}\sigma_\Omega^2) L^2.$$

Tedious but straightforward stochastic calculus [32] also yields the variances of the noise-perturbed output soliton parameters as

$$(2.36a) \qquad \mathrm{var}[A(L)] = \sigma_A^2 L, \qquad \mathrm{var}[\Omega(L)] = \sigma_\Omega^2 L, \qquad \mathrm{var}[T(L)] = \sigma_T^2 L + \frac{1}{3}\bar{d}\sigma_\Omega^2 L^3,$$

(2.36b)

$$\mathrm{var}[\Phi(L)] = (\sigma_\Phi^2 + \bar{d}\Omega_o^2 \sigma_\Omega^2) L + \bar{d}\Omega_o \sigma_T^2 \sigma_\Omega^2 L^2 + \frac{1}{3}(A_o^2 \sigma_A^2 + \bar{d}\Omega_o^2 \sigma_\Omega^2) L^3 + \frac{1}{12}(\sigma_A^4 + \bar{d}^2 \sigma_\Omega^4) L^4.$$

Note that, unlike the other soliton parameters, the mean phase is affected by the noise. The cubic dependence of timing jitter on distance (which arises as a result of the coupling between carrier frequency and group velocity) in the third equation of (2.36a) is of course the well-known Gordon–Haus jitter [13]. Similarly, the cubic dependence of the phase jitter on distance (due to the Kerr effect and Galilean invariance) is called the Gordon–Mollenauer jitter [14]. Note that additional contributions are present for the phase variance, which also happens for NLS [14], but the term is quadratic in the noise variance.

Remarkably, these results are formally identical to those for the NLS equation. In particular, (2.36b) reduces to the results of [16] when $\Omega_o = 0$. Here, however, we show that they also hold for $s \neq 0$. Moreover, even though the form of (2.36) is the same for $s = 0$ and $s \neq 0$, the variances σ_Q^2 of the source terms vary with s, as can be seen from Figure 1 and (2.31). Also, the functional dependence of the variances on the soliton amplitude is dramatically different because of the different dependence of the norms and inner products on the amplitude,

as discussed earlier. As is the case for the NLS equation, the knowledge of variances as a function of distance is not sufficient to accurately predict the frequency of the rare events, as discussed in detail in [30]. Finally, we emphasize that the SDEs (2.18b) were derived by applying perturbation theory and using the Stratonovich definition of calculus. This is correct only if one is dealing with the Stratonovich interpretation of the noise-driven NLS equation (2.8a). If one wants to study a nonanticipatory noise process, however, the Itô formulation is required [9].

3. IS for the DMNLS equation. We now discuss how the results of section 2 can be used to guide the implementation of IS for the DMNLS equation. For simplicity we set $\bar{d} = 1$ throughout the rest of this work.

3.1. IS and the optimal biasing problem. The idea behind IS is straightforward. In order to calculate the probability of a desired rare event, one samples the noise from a biased distribution that makes the rare events occur more frequently than would naturally be the case, and simultaneously corrects for the biasing.

Consider a set of random variables $\mathbf{X} = (x_1, \ldots, x_N)$ distributed according to a joint probability distribution $p(\mathbf{X})$. The probability P that a function $y(\mathbf{X})$ falls into a desired range Y_d can be expressed via the multidimensional integral

$$(3.1a) \qquad P = \mathbb{P}[y \in Y_d] = \mathbb{E}[I(y(\mathbf{X}))] = \int I(y(\mathbf{x}))p(\mathbf{x})(\mathrm{d}\mathbf{x}),$$

where the *indicator function* $I(y)$ equals 1 when $y \in Y_d$ and 0 otherwise, and $(\mathrm{d}\mathbf{x})$ is the volume element in $\mathbb{R}^N$. When evaluation of the integral is impossible (as it often is due to the large dimensionality of sample space and the complicated form of the map from $\mathbf{X}$ to y), one needs to resort to numerical methods. An unbiased estimator for P can be constructed via MC quadrature as

$$(3.1b) \qquad \hat{P}_{\mathrm{mc}} = \frac{1}{M} \sum_{m=1}^{M} I(y(\mathbf{X}_m)),$$

where the M samples $\mathbf{X}_m$ are drawn from the distribution $p(\mathbf{X})$. If P is very small, however, an unreasonable number of samples are necessary to produce any events for which $y \in Y_d$, let alone enough to accurately approximate the integral. However, one can rewrite (3.1) as follows:

$$(3.2a) \qquad \mathbb{P}[y \in Y_d] = \int I(y(\mathbf{x}))L(\mathbf{x})p_*(\mathbf{x})d\mathbf{x},$$

$$(3.2b) \qquad \hat{P}_{\mathrm{ismc}} = \frac{1}{M} \sum_{m=1}^{M} I(y(\mathbf{X}_{*,m}))L(\mathbf{X}_{*,m}),$$

where the samples $\mathbf{X}_{*,m}$ are now drawn from the *biasing distribution* $p_*(\mathbf{X})$, and where the quantity $L(\mathbf{X}) = p(\mathbf{X})/p_*(\mathbf{X})$ is the *likelihood ratio*. Note that if one sampled from a biased distribution but did not correct with the likelihood ratio, one would be "counting the observed events too much." When a large range of values of the quantity of interest is desired, it might

be necessary to use several biasing distributions. The individual samples can then be efficiently combined using a practice called *multiple importance sampling* [7].

When appropriate biasing distributions are selected, and the contributions of the individual samples are corrected with the likelihood ratios, ISMC simulations can accurately estimate the probability of the rare events in exam with a small fraction of the number of samples that standard MC methods would require. In order for IS to work, however, the biasing distribution $p_*(\mathbf{X})$ must preferentially concentrate the MC samples around the *most likely* system realizations that lead to the rare events of interest. In our case the random variables are the noise components added after each dispersion map period. Thus, to successfully apply IS we must find the most likely noise realizations that lead to a desired value of the soliton parameters at the output. We approach this problem by decomposing it into two logically distinct steps: (i) finding the most likely noise realizations that produce a given parameter change at each amplifier; and (ii) finding the most likely way in which individual parameter changes at each amplifier combine to produce a total change at the output. The two-step approach is justified by the fact that the noise at different amplifiers is statically independent.

3.2. Biasing at a single amplifier. Recall that at the nth amplifier a zero-mean, delta-correlated white-noise process, $\nu_n(t)$, is added to the signal as a consequence of spontaneous emission. On average, this noise does not introduce parameter changes at a single amplifier. However, we can induce an average parameter change by biasing the mean of this process. To so do, we begin by considering a generic perturbation to the solution at the nth amplifier, $b_n(t)$. Recall that the noise-induced change to a soliton parameter Q (with $Q = A, \Omega, T, \Phi$) is found by taking the projection of the perturbation onto the adjoint mode of the linear DMNLS operator associated with Q (and normalized by the norm of that adjoint mode). That is, if $u(t, nz_a^+) = u(t, nz_a^-) + b_n(t)$, the change to parameter Q due to the perturbation $b_n(t)$ is given by

$$(3.3) \qquad \Delta Q_n = \mathrm{Re} \int \underline{y}_Q^* b_n(t)\, \mathrm{d}t \Big/ \int |\underline{y}_Q|^2\, \mathrm{d}t.$$

(For the purposes of the present discussion it is convenient to redefine $\underline{y}'_\Phi = \underline{y}_\Phi + \Omega\, \underline{y}_T$.) If $b_n(t)$ is an unbiased, zero-mean process, it is obviously $\mathbb{E}[\Delta Q_n] = 0$; but different noise realizations lead to different noise-induced parameter changes. The problem of finding the optimal biasing at each amplifier is to find the most likely noise realization subject to the constraint of achieving, on average, a desired parameter change at that amplifier. In other words, given a specific parameter change ΔQ_n at the nth amplifier (with $Q = A, \Omega, T, \Phi$), what is the form of $b_n(t)$ that is *most likely* to produce this prescribed change? (Once these $b_n(t)$ have been found, one can then use them to bias the MC simulations by concentrating the random samples $\nu_n(t)$ around these deterministic values.) For white Gaussian noise, maximizing its probability amounts to minimizing the negative of the log-likelihood, i.e., the negative of the argument of the exponential in the noise PDF. That is, we need to minimize the L_2-norm

$$(3.4) \qquad \|b_n(t)\|^2 = \int |b_n(t)|^2\, \mathrm{d}t$$

subject to achieving the desired parameter change ΔQ_n, which is given by (3.3). (Keep in mind that the question of how the ΔQ_n's themselves should be prescribed is a separate issue that we will address at length in section 4.) One can formulate this as a variational problem as in [30]. That is, we consider the functional

$$(3.5) \qquad J[b_n, b_n^*] = \int |b_n(t)|^2 \, dt + \mu \left(\Delta Q - \mathrm{Re} \int \underline{y}_Q^* b_n(t) \, dt \right),$$

which depends on the biasing function $b_n(t)$ and on the Lagrange multiplier μ. The minimum of this functional yields the optimal biasing, that is, the most likely noise realization that satisfies the parameter constraint (3.3). Taking the functional derivative of $J[\ ]$ with respect to b_n and setting it to zero yields $b_n(t) = \mu \underline{y}_Q(t)$. The value of μ can be determined by taking the inner product of b_n with the adjoint eigenmode $\underline{y}_Q$ and then simplifying with the constraint, which yields $\mu = \Delta Q_n / \|\underline{y}_Q\|^2$. Combining the two results, we then obtain the optimal biasing function as

$$(3.6) \qquad b_n(t) = \Delta Q_n \frac{\mathrm{Re} \int \underline{y}_Q^* y_Q \, dt}{\int |\underline{y}_Q|^2 \, dt} \, \underline{y}_Q.$$

We emphasize that the above calculation and, therefore, the optimal biasing function are completely deterministic. Note also that, even though adding a noise component along the eigenmode associated with parameter Q is the cleanest way to change that parameter (that is, the way to obtain a prescribed ΔQ_n without also adding radiation to the signal), the optimal way to bias the noise in order to obtain a prescribed parameter change is to add noise along the *adjoint* eigenmode. In other words, the most likely noise realization that produces that change is proportional to the adjoint mode [23]. For the constant dispersion case, this was first reported in [29]. Notice, however, that the coefficient of proportionality is scaled by the norm of the adjoint mode. This norm is unity in the constant dispersion case but varies with amplitude in the dispersion-managed case, as is demonstrated in Figure 1.

4. Biasing across all amplifiers. We now address the following question: how should one distribute the bias for the soliton parameters among all amplifiers in order to achieve a specified parameter change at the output? In other words, what is the most likely set of individual parameter changes $\{\Delta A_n, \Delta \Omega_n, \Delta T_n, \Delta \Phi_n\}_{n=1}^{N_a}$ that realizes a given value of ΔQ_{target} (with Q equal to A, Ω, T, or Φ, as before) at the output?

4.1. Biasing across all amplifiers for amplitude changes. Here we seek to minimize how much we bias the noise over all amplifiers to achieve a given output amplitude change. We begin by examining the amplitude evolution from one amplifier to the next, namely,

$$(4.1) \qquad A_{n+1} = A_n + \Delta A_{n+1}.$$

The results of section 3.2 yield the most likely noise realization that achieves a given amplitude change at a single amplifier as

$$(4.2) \qquad b_n = \frac{\langle y_A, \underline{y}_A \rangle}{\|\underline{y}_A\|^2} \Delta A_n \, \underline{y}_{A_n}.$$

Recall that the norms and inner products of the linear modes depend on the soliton amplitude (cf. Figure 1) and therefore also indirectly on distance. At each amplifier, we will be biasing the mean of the white Gaussian noise with b_n according to (4.2), and our goal is to maximize the probability of achieving a given amplitude at the output. It should be clear that maximizing this probability is equivalent to minimizing the sum of the L_2-norm of the biasing function *over all amplifiers.* That is, we need to minimize the sum

$$(4.3) \qquad \sum_{n=1}^{N_a} \|b_n\|^2 = \sum_{n=1}^{N_a} \frac{1}{\sigma_A^2} |\Delta A_n|^2$$

subject to the constraint

$$(4.4) \qquad \sum_{n=1}^{N_a} \Delta A_n = A_{\text{target}} - A_o,$$

where σ_A^2 is given by (2.31). To solve this problem we consider a continuum approximation. That is, we replace (4.1) by the first equation of (2.18) with $S(t,z) = b(t,z)$, and $b(t,z)$ given by the continuum analogue of (4.2); that is,

$$(4.5) \qquad b(t,z) = \frac{\langle y_A, \underline{y}_A \rangle}{\|\underline{y}_A\|^2} \, \underline{y}_A(t) \, \dot{A}.$$

We thus need to find a function $A(z)$ that minimizes the continuum limit of (4.3). That is, we seek to minimize the integral from $z = 0$ to $z = L$ of the L_2-norm of $b(t,z)$, namely, the functional

$$(4.6) \qquad J[A] = \int_0^L \frac{1}{\sigma_A^2} \, \dot{A}^2 \, dz,$$

subject to the fixed boundary conditions $A(0) = A_o$ and $A(L) = A_{\text{target}}$. Hereafter the dot will denote total differentiation with respect to z, and L the total transmission distance. After some straightforward algebra, the Euler–Lagrange equation associated with J becomes

$$2\ddot{A} \frac{1}{\sigma_A^2} + \dot{A}^2 \frac{\partial}{\partial A} \left(\frac{1}{\sigma_A^2} \right) = 0,$$

which is readily integrated to give

$$(4.7) \qquad \dot{A} = c \, \sigma_A,$$

where c is an integration constant which determines the total amount of biasing being applied and thereby the value of the amplitude at the output.

One can now integrate (4.7) to find the *optimal biasing path* $A(z)$ that realizes a desired amplitude at the output. Once the optimal amplitude path $A(z)$ has been obtained as a function of distance, one can then calculate ΔA_n, which was the only unknown in the optimal biasing perturbation b_n as given in (4.2). Note, however, that it is not actually necessary to
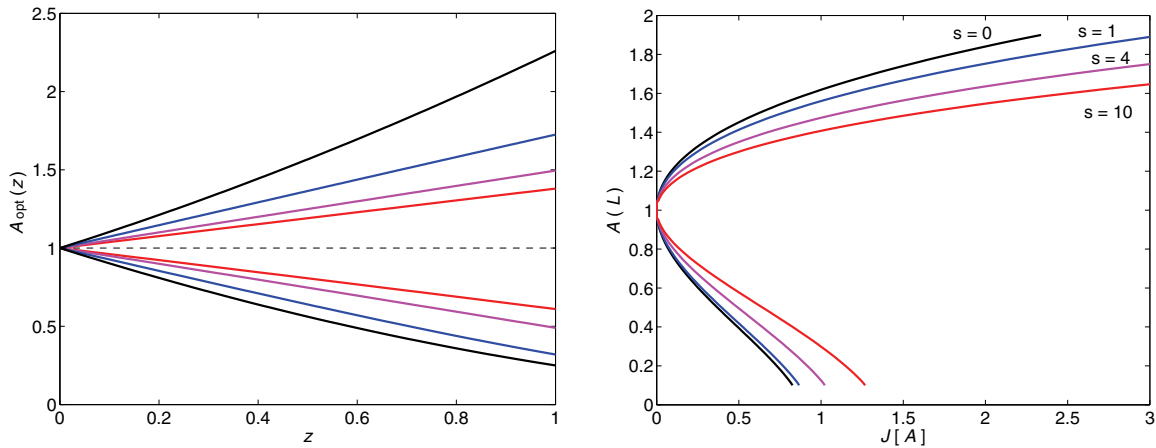
Figure 2. *Left: Optimal paths for the amplitude for dispersion-managed solitons to obtain large (positive and negative) deviations at $L = 1$, computed for the same amount of total noise norm and for different values of map strength: $s = 0$ (black), $s = 1$ (blue), $s = 4$ (magenta), and $s = 10$ (red). The dashed line shows the unperturbed amplitude. Right: The output value of the amplitude as a function of the total noise norm for the same values of map strength as before.*

obtain $A(z)$ to find the optimal biasing function, since once $\dot{A}$ is determined via (4.7), the optimal biasing $b(t, z)$ in the continuum approximation is obtained via (4.5), and the optimal biasing $b_n(t)$ in the discrete case is then approximated by taking $\Delta A_n = \dot{A} z_a$. On the other hand, integrating (4.7) is necessary to explicitly connect the biasing strength to the final value of the amplitude, for which it is necessary to explicitly determine the integration constant c in terms of the given boundary conditions.

Equation (4.7) can be integrated explicitly in the special case of the NLS equation (that is, in the case of constant dispersion, $s = 0$). In this case, use of (2.32) reduces (4.7) to $\dot{A} = c\sqrt{A}$, which is trivially integrated to

$$(4.8) \qquad A_{\mathrm{nls}}(z) = \left[\left(\sqrt{A_{\mathrm{target}}} - \sqrt{A_o}\,\right) z/L + \sqrt{A_o}\,\right]^2.$$

When $s \neq 0$, the functional dependence of σ_A on A is not known, and therefore it is not possible to integrate (4.7) analytically. Since we have numerical expressions for the norms and inner products, however, we can proceed by numerically integrating $\dot{A}$, obtaining an expression for $z = z(A)$, and then inverting this expression to find the optimal biasing path. Such paths are shown in Figure 2 for several values of reduced map strength.

Another case in which (4.7) can be integrated exactly is the large-s limit. As can be seen from Figure 1, $\langle y_A, \underline{y}_A \rangle$ is (to a good approximation) linear in A, while for large values of s $\|\underline{y}_A\|^2$ is approximately quadratic in A. Since σ_A^2 is proportional to $\|\underline{y}_A\|^2 / \langle \underline{y}_A, y_A \rangle^2$, for large values of s (4.7) is approximated by $\dot{A} \propto \sigma_A^2 = \mathrm{const}$. This means that the optimal path to reach a given target value of amplitude is for the amplitude to increase/decrease linearly. This justifies the use of a constant biasing strength in [23]. Indeed, it should be clear from Figure 2 that *the optimal amplitude paths are very nearly linear for almost all nonzero values of map strength.* This further justifies the use of an equal biasing distribution among amplifiers.

4.2. Biasing across all amplifiers for frequency changes. The search for larger-than-normal frequency changes is perhaps the simplest of the four cases. Recall that the frequency offset changes according to

$$(4.9) \qquad \Omega_{n+1} = \Omega_n + \Delta\Omega_{n+1},$$

and thus the setup for the optimal frequency biasing problem quite closely follows that of the optimal amplitude biasing problem. As before, (3.6) implies that the optimal biasing function for the frequency at a single amplifier is

$$(4.10) \qquad b_n = \frac{\langle y_\Omega, \underline{y}_\Omega \rangle}{\|\underline{y}_\Omega\|^2} \, \Delta\Omega_n \, \underline{y}_{\Omega_n}.$$

Again we consider the continuum limit, and we seek a function $\Omega(z)$ that minimizes the functional

$$(4.11) \qquad J[\Omega] = \int_0^L \frac{1}{\sigma_\Omega^2} \dot{\Omega}^2 \, dz,$$

with σ_Ω^2 as before, subject to the fixed boundary conditions $\Omega(0) = 0$ and $\Omega(L) = \Omega_{\text{target}}$. The Euler–Lagrange equation associated to this functional is trivially integrated to give $\dot{\Omega} = c \, \sigma_\Omega^2$, where c is an integration constant. Even though σ_Ω depends on the A, it is independent of Ω, and therefore it is constant to first order in the perturbation expansion. We can therefore integrate trivially and use the boundary conditions $\Omega(0) = 0$ and $\Omega(L) = \Omega_{\text{target}}$ to obtain $c = \Omega_{\text{target}}/\sigma_\Omega^2$, implying

$$(4.12) \qquad \Omega(z) = \Omega_{\text{target}} \, z/L.$$

That is, the optimal way to achieve a given target value of frequency is to divide the total change equally among all amplifiers. In the discrete case, we have $\Delta\Omega_n = \Omega_{\text{target}}/N$. Since the result is identical to the NLS case [29, 30], we will forgo any further discussion of frequency biasing in the remainder of this work.

4.3. Biasing across all amplifiers for timing changes. The setup for the optimal biasing problem in the case of large timing changes differs from the cases of amplitude and frequency changes because changes in the pulse timing are induced both by *direct* timing jitter and by *frequency-induced* timing jitter via the coupling between carrier frequency and group velocity (cf. the third equation of (2.18)). Of course, this is the well-known Gordon–Haus effect [13].

From an amplifier to the next amplifier the pulse timing evolves as

$$(4.13) \qquad \Omega_{n+1} = \Omega_n + \Delta\Omega_{n+1}, \qquad T_{n+1} = T_n + \Omega_n z_a + \Delta T_{n+1}.$$

Thus we must bias to induce both timing and frequency changes at each amplifier. Owing to (3.6) and the orthogonality of the modes, the most likely noise realization to obtain given simultaneous changes in both the frequency and the timing is then

$$(4.14) \qquad b_{n+1} = \frac{\langle y_\Omega, \underline{y}_\Omega \rangle}{\|\underline{y}_\Omega\|^2} \, \Delta\Omega_{n+1} \, \underline{y}_{\Omega_n} + \frac{\langle y_T, \underline{y}_T \rangle}{\|\underline{y}_T\|^2} \, \Delta T_{n+1} \, \underline{y}_{T_n}.$$

After again taking the continuum limit and imposing the evolution equations for the soliton parameters (the second and third equations of (2.18)) as constraints, we now obtain

$$(4.15) \qquad b(t,z) = \frac{\langle y_\Omega, \underline{y}_\Omega \rangle}{\|\underline{y}_\Omega\|^2} \dot{\Omega}\, \underline{y}_\Omega + \frac{\langle y_T, \underline{y}_T \rangle}{\|\underline{y}_T\|^2} (\dot{T} - \Omega)\, \underline{y}_{T_n}.$$

Correspondingly, our optimization problem is

$$(4.16) \qquad J[\Omega, T] = \int_0^L \left[\frac{1}{\sigma_\Omega^2} \dot{\Omega}^2 + \frac{1}{\sigma_T^2} (\dot{T} - \Omega)^2 \right] \mathrm{d}z.$$

Importantly, unlike the previous cases, the direct timing biasing ΔT_{n+1} is *not* given by $\dot{T}\, z_a$, but rather by $(\dot{T} - \Omega)\, z_a$. This is because the ODE for dT/dz has a deterministic part (cf. the third equation of (2.18)).

    The Euler–Lagrange equations associated with T and Ω are, respectively,

$$(4.17a) \qquad \frac{d}{dz}\left[\frac{1}{\sigma_T^2}(\dot{T} - \Omega) \right] = 0,$$

$$(4.17b) \qquad \frac{d}{dz}\left[\frac{1}{\sigma_\Omega^2}\dot{\Omega} \right] + \frac{1}{\sigma_T^2}(\dot{T} - \Omega) = 0.$$

As before, neglecting amplitude changes in the optimal trajectory, the norms and inner products in σ_Ω and σ_T are constants of the motion. As a result, we can easily integrate the first equation of (4.17) to give

$$(4.18a) \qquad \dot{T} - \Omega = c,$$

where c is an integration constant. Substituting (4.18a) into the second equation of (4.17), integrating twice, and using the boundary conditions $\Omega(0) = 0$ and $\dot{\Omega}(L) = 0$ then yields, respectively,

$$(4.18b) \qquad \dot{\Omega} = c\frac{\sigma_\Omega^2}{\sigma_T^2}(L - z)$$

and

$$(4.19a) \qquad \Omega(z) = c\frac{\sigma_\Omega^2}{\sigma_T^2}(Lz - z^2/2).$$

(Note that the boundary condition $\dot{\Omega}(L) = 0$ is natural because a frequency change at $z = L$ would not result in an additional timing change, since frequency-induced timing changes can only accumulate with distance.) One can now substitute (4.19a) back into (4.18a), integrate the latter, and impose the boundary conditions $T(0) = 0$ and $T(L) = T_{\text{target}}$, obtaining

$$(4.19b) \qquad T(z) = c\left(z + \frac{\sigma_\Omega^2}{2\sigma_T^2}(Lz^2 - z^3/3) \right),$$

where the integration constant is

$$(4.20) \qquad c = T_{\text{target}} / \left(L + \frac{\sigma_\Omega^2}{\sigma_T^2} L^3/3 \right).$$

As noted before, the right-hand sides of (4.18) (with c given by (4.20)) give the biasing functions for the direct timing and frequency contributions, respectively.

In the case of the NLS equation (constant dispersion), (2.32) yields the ratio of the variances explicitly as $\sigma_\Omega^2/\sigma_T^2 = A^4/(9\pi^2)$. The functional dependence of the optimal biasing function in the general case, however, is exactly the same as in the constant dispersion case, the only difference between the two cases being the precise proportion of direct timing and indirect frequency biasing (as shown by (4.18)). Therefore, since the case of constant dispersion was discussed in [30], we will also omit further discussion of time biasing.

4.4. Biasing across all amplifiers for phase changes. The optimal biasing problem for the phase is not conceptually different from those for the other soliton parameters. The calculations, however, turn out to be the most complicated among the four cases.

The mean phase Φ evolves according to (2.18b). Note, however, that, among the terms in the right-hand side, changes in Ω^2 and $\langle e^{-i\Theta}\underline{y}_T, S\rangle\Omega$ are second-order in the noise, while changes in A^2 are first-order in the noise, because $\Omega(0) = 0$ while $A(0) \neq 0$. Motivated by these considerations, we neglect the terms $\Omega^2/2$ and $\langle e^{-i\Theta}\underline{y}_T, S\rangle\Omega$ in (2.18b). That is, we introduce the auxiliary quantity

$$(4.21) \qquad \phi(z) = \Phi(z) - \int_0^z \left(\Omega^2/2 - \langle e^{-i\Theta}\underline{y}_T, S\rangle\,\Omega \right) \mathrm{d}z,$$

and we then consider the optimal biasing problem for $\phi(z)$. Using (2.18b) in (4.21) then, we simply have

$$(4.22) \qquad \frac{d\phi}{dz} = \frac{1}{2}A^2 + \frac{\langle e^{i\Theta}\underline{y}_\Phi, S\rangle}{\langle \underline{y}_\Phi, y_\Phi\rangle}.$$

Equation (4.22) shows that, similarly to the case of timing changes, changes in the soliton phase result both from direct phase perturbations and by amplitude-induced phase perturbations via the coupling between amplitude and phase due to Kerr nonlinearity. This is the well-known Gordon–Mollenauer effect [14]. Thus, if we seek to induce larger-than-normal phase changes, we need to bias both the soliton phase directly and the soliton amplitude.

It should be noted that, while substituting $\phi(z)$ for $\Phi(z)$ greatly simplifies the optimal biasing problem, its use can be justified a posteriori. Indeed, the ISMC simulations discussed in section 5 show that the uncertainties in the reconstructed tails of the noise-induced PDF of the actual output phase are compatible to those for the other soliton parameters, thereby suggesting that our choice of biasing is optimal for the phase as well.

In the case of noise, (4.22) and the first equation of (2.18) imply the following changes of amplitude and phase parameters from one amplifier to the next:

$$(4.23) \qquad A_{n+1} = A_n + \Delta A_{n+1}, \qquad \phi_{n+1} = \phi_n + \frac{1}{2}z_a\, A_n^2 + \Delta\phi_{n+1},$$

where ΔA_{n+1} is the amplitude change and $\Delta\phi_{n+1}$ is the *direct* phase change. Recall that the optimal biasing at a single amplifier is known and given by (3.6). Thus the most likely noise realizations that produce the given values of ΔA_{n+1} and $\Delta\phi_{n+1}$ have a mean that equals the following optimal biasing function:

$$(4.24) \qquad b_{n+1} = \Delta A_{n+1} \frac{\langle y_A, \underline{y}_A \rangle}{\|\underline{y}_A\|^2} \underline{y}_{A_n} + \Delta\phi_{n+1} \frac{\langle y_\Phi, \underline{y}_\Phi \rangle}{\|\underline{y}_\Phi\|^2} \underline{y}_{\Phi_n}.$$

Taking the continuum limit $z_a \to 0$ as before, and using (4.22) and the first equation of (2.18), the optimal biasing function becomes

$$(4.25) \qquad b(t, z) = \frac{\langle y_A, \underline{y}_A \rangle}{\|\underline{y}_A\|^2} \dot{A} \, \underline{y}_A + \frac{\langle y_\Phi, \underline{y}_\Phi \rangle}{\|\underline{y}_\Phi\|^2} (\dot\phi - A^2/2) \, \underline{y}_\Phi.$$

(Note that, similarly to section 4.3, the direct phase biasing $\Delta\phi_{n+1}$ is given not by $\dot\phi \, z_a$ but by $(\dot\phi - A^2/2) \, z_a$.) Again, minimizing the sum of the L_2-norm of the biasing function over all amplifiers is equivalent in the continuum limit to finding functions $A(z)$ and $\phi(z)$ that minimize the functional

$$(4.26) \qquad J[A, \phi] = \int_0^L \left[\frac{1}{\sigma_A^2} \dot{A}^2 + \frac{1}{\sigma_\Phi^2} (\dot\phi - A^2/2)^2 \right] dz.$$

The Euler–Lagrange equations associated with ϕ and A are, respectively,

$$(4.27a) \qquad \frac{d}{dz} \left[\frac{1}{\sigma_\Phi^2} (\dot\phi - A^2/2) \right] = 0,$$

$$(4.27b) \qquad 2 \frac{d}{dz} \left[\frac{1}{\sigma_A^2} \dot{A} \right] - \frac{\partial}{\partial A} \left[\frac{1}{\sigma_A^2} \dot{A}^2 + \frac{1}{\sigma_\Phi^2} (\dot\phi - A^2/2)^2 \right] = 0.$$

Equation (4.27a) yields immediately

$$(4.28) \qquad \dot\phi - A^2/2 = c \, \sigma_\Phi^2,$$

which, in turn, when substituted into (4.27b), yields

$$(4.29) \qquad 2\ddot{A} \frac{1}{\sigma_A^2} + \dot{A}^2 \frac{\partial}{\partial A} \left[\frac{1}{\sigma_A^2} \right] + c^2 \frac{\partial}{\partial A} \left[\sigma_\Phi^2 \right] + 2c \, A = 0.$$

The system composed of (4.29) and (4.28), together with the boundary conditions $A(0) = A_o$, $\dot{A}(L) = 0$, $\phi(0) = 0$, and $\phi(L) = \phi_{\text{target}}$, describes the optimal amplitude and phase paths around which to bias the ISMC simulations. Note that, as in section 4.3, the boundary condition $\dot{A}(L) = 0$ is natural because an amplitude change at $z = L$ would not result in an additional phase change since amplitude-induced phase changes accumulate with distance.

When $c = 0$, (4.29) can be integrated to obtain (after use of the boundary conditions) the constant solution $A(z) = A_o$. Thus, $c = 0$ corresponds to the case of no biasing being applied. When $c \neq 0$, however, one must solve (4.29) numerically. This can be done using a relaxation method, as discussed in Appendix C, or, alternatively, a numerical continuation software such
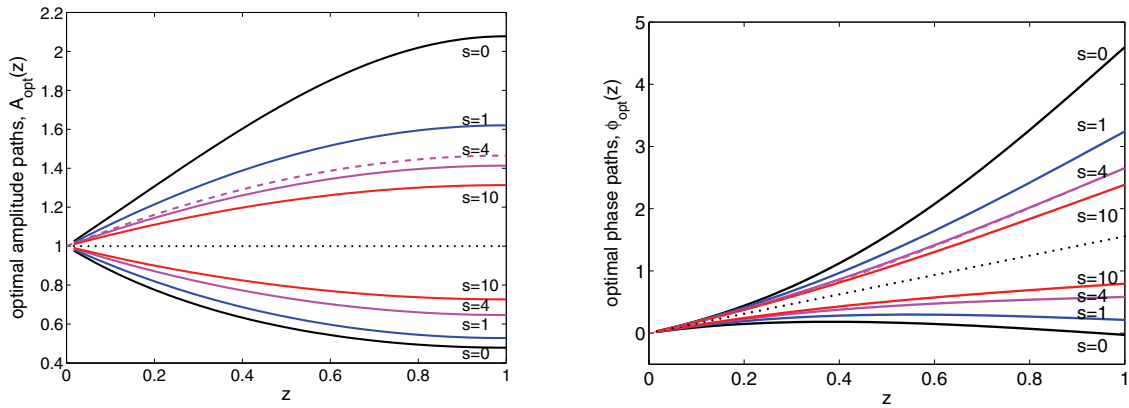
Figure 3. *Optimal paths that the amplitude (left) and phase (right) take to reach larger- or smaller-than-normal noise-induced output phase values of ϕ at $z = 1$ for the same total noise norm with different values of map strength. The dotted lines show the unperturbed amplitude and phase paths. The dashed lines show solutions to (4.28)–(4.29) where σ_A^2 and σ_ϕ^2 are replaced with large-s approximations as described in Appendix D. Note that the dashed line in the right figure lies on top of the phase path for $s = 4$. Colors are the same as in Figure 2.*

as XPP/AUTO [10]. In either case, different output phases can be targeted by solving (4.29) exactly in the case $c = 0$ and then extending the solution set by varying c. For the large map strength case, σ_A^2 and σ_Φ^2 are well approximated by monomials in A. Appendix D describes this approximation, the resulting equation, and its solution(s) in detail. We will refer to solutions of this approximate equation as "large-s" amplitude and phase paths.

In the constant dispersion case, use of (2.27) and (2.28) simplifies (4.29) to

$$(4.30) \qquad \ddot{A} - \frac{1}{2A}\dot{A}^2 - c^2\frac{\pi^2 + 12}{18A} + 2cA^2 = 0,$$

which agrees with our previous results [37]. Note, however, that even (4.30) must be integrated numerically when $c \neq 0$.

Figure 3 shows the optimal paths A and ϕ for $s = 0, 1, 4, 10$ with the same total noise norm. Also plotted in Figure 4 is the output phase as a function of the total noise norm where that norm is computed by evaluating the functional $J[A, \phi]$ (4.26) with the optimal paths of A and ϕ (which are plotted in Figure 3). Note that as the map strength increases, one must bias the noise harder (i.e., the minimum $J[A, \phi]$ is larger) in order to achieve the same output phase. This demonstrates the robustness of dispersion-managed solitons to direct and amplitude-induced phase distortion.

When the above calculation is reduced to the constant dispersion case, we once again recover the result of [30, 37]. It is worth mentioning, however, that the direct phase biasing is *not constant* in z, unlike the case of constant dispersion and unlike the case of direct time biasing in section 4.3.

5. Numerical simulations and results. We now describe how the results of section 4 can be used to implement IS for the DMNLS equation and in turn to compute the PDF of
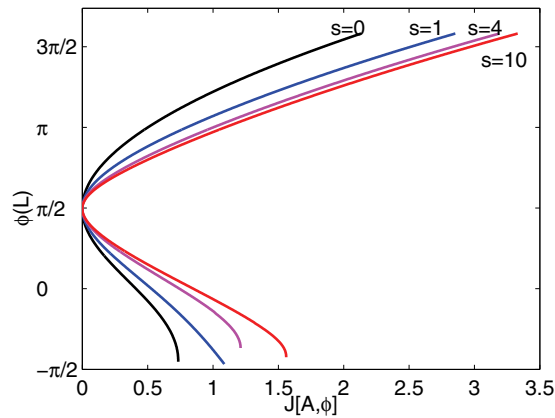
Figure 4. *The output value of phase, $\Phi(L)$, as a function of the total noise norm. Colors and parameters are the same as in Figure* 3.

the dispersion-managed soliton parameters at the output as a result of noise. When seeking larger-than-normal deviations of a parameter Q, one must perform the following steps at each map period:

1. Recover the underlying dispersion-managed soliton from the noisy signal.
2. Obtain the linear modes and adjoint modes of the linearized DMNLS around the given dispersion-managed soliton.
3. Generate an unbiased noise realization, shift its mean with the appropriately scaled adjoint modes, and update the likelihood ratio.
4. Add the noise to the pulse, and propagate the new noisy signal to the next map period.

One then repeats this process until the signal reaches the output. Note that, to induce a larger than normal parameter change, the noise is biased by concentrating the MC samples around $b_n(t)$. That is, we take $\nu_{n,\text{biased}}(t) = b_n(t) + \nu_n(t)$, where $b_n(t)$ is the optimal biasing function determined in sections 3 and 4 and $\nu_n(t)$ is an unbiased white-noise process satisfying (2.3).

For the simulations described next, we combined several thousand ISMC samples generated with a few different biasing targets, using multiple IS [7] in order to reconstruct the PDF of the quantity of interest.

We performed studies on two physical systems. In system (a) we sought large output amplitude changes, while in system (b) we sought large output phase changes. (As mentioned earlier, we omit performing simulations for frequency and timing since the biasing choices for frequency and for timing are the same as for the NLS equation, which has already been demonstrated in [29, 30].) Amplitude deviations are of course important for systems using amplitude-shift-keyed formats. Similarly, phase deviations are important for phase-shift-keying formats. In both cases we chose system parameters based on realistic values for optical fiber communication systems. (Typical values of system parameters for the DMNLS in femtosecond lasers can be obtained from [34].) We took a piecewise constant dispersion map with an average dispersion of $0.15 \,\text{ps}^2/\text{km}$ and a map strength $s = 4$. We set the unit time to $17 \,\text{ps}$, and we used the resulting dispersion length of $1,923 \,\text{km}$ to normalize the distance z, resulting in $\bar{d} = 1$. We considered amplifiers spaced every $100 \,\text{km}$ (resulting in $z_a = 0.052$).

Table 1

Physical system parameters for the numerical simulations.

System	Distance	N_a	Loss coeff	Unit/peak power	Spontaneous emission factor	Optical SNR	Noise variance
(a)	4000 km	40	0.21 dB/km	2.96 mW	1.5	13.8	$1.873 \cdot 10^{-3}$
(b)	6000 km	60	0.25 dB/km	3.51 mW	1.65	9.3	$9.486 \cdot 10^{-4}$
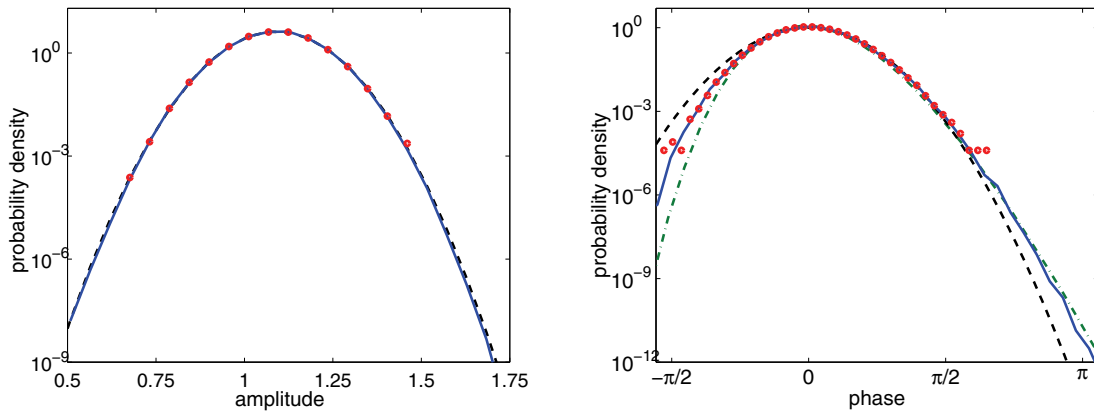
Figure 5. *(Left) PDF of dispersion-managed soliton amplitude. Solid curve: Results from ISMC simulations of the DMNLS with 70,000 samples. Dots: Results from standard MC simulations of the NLS equation with dispersion management with 300,000 samples. Dashed curve: Gaussian fit to that simulation. (Right) PDF of output phase. Solid curve: Results from ISMC simulation of DMNLS with 50,000 samples. Dots: Results from standard MC simulations of the NLS equation with dispersion management with 250,000 samples. Dashed curve: A Gaussian PDF fit to mean and variance of ISMC simulation. Dashed-dot curve: ISMC simulations of the noise-driven SDEs (2.18b). Note how unbiased MC simulations of the NLS equation with dispersion management clearly deviate from the Gaussian but agree well with ISMC simulations of the DMNLS as far down in probability as the unbiased simulations can accurately reach.*

We also set the dispersion map period to be aligned with them. The remaining system parameters for both systems are given in Table 1. We took a nonlinear coefficient of $1.7 \, (\text{W} \cdot \text{km})^{-1}$, and we used the power needed to have $\bar{g} = 1$ (which depends on the fiber loss; see Table 1) to normalize pulse powers. Note that the local dispersion is large in magnitude, but the average dispersion is small.

Even though at each map period the noise added induces only a small change in the dispersion-managed soliton parameters, these small changes can accumulate into large changes at the output resulting in a significantly distorted signal. Note that the linearized DMNLS is used only to guide the biasing. That is, for each individual sample in the ISMC simulations, the full DMNLS is solved to propagate the signal. In other words, use of IS enables full nonlinear simulation of large, noise-induced parameter changes. For amplitude biasing, the optimal paths become closer to linear as s increases. It was demonstrated in [30] that linear biasing works well for NLS, that is, for $s = 0$.

Figure 5(left) shows the PDF of the dispersion-managed soliton output energy in system (a), as computed from ISMC simulations of DMNLS (solid line), from standard MC simulations of the original NLS with dispersion management (dots), and a Gaussian fit to

them (dashed line). The PDFs generated from ISMC simulations of DMNLS and standard MC simulations of NLS both agree very well with each other as far down in probability as the unbiased MC simulations can reach. This comparison provides a further strong validation of DMNLS as a model of dispersion-managed lightwave systems and suggests that amplitude-jitter in dispersion-managed systems with large map strength is nearly Gaussian distributed.

Figure 5(right) shows the PDF of the dispersion-managed soliton output phase in system (b) as computed from ISMC simulations of DMNLS (solid blue line). Also plotted are MC simulations of the NLS with a varying dispersion coefficient (red dots), a Gaussian fit to those simulations (black dashed line), and an ISMC simulation of the parameter SDEs (2.18b) (green dashed-dot line). The value of the PDF corresponding to π deviations from the mean output phase is a quantity of interest in femtosecond lasers since any phase deviation beyond that value would represent a phase slip. Figure 5(right) shows that the Gaussian approximation to the phase PDF near π can be off by several orders of magnitude. Note that the phase of the dispersion-managed soliton in Figure 5(right) is obtained from the filtered output pulse. Note also that numerical estimates of the PDF at negative phase deviations did not seem to converge as quickly.

Both figures in Figure 5 show excellent agreement between the noise-perturbed DMNLS equation (2.8a) and the noise-perturbed NLS equation with varying dispersion (2.1). That is, the distributions for both amplitude and phase agree well as far down as the tails can be computed using standard MC. An even stronger result, however, is that the simulations show *pathwise* agreement between noise-perturbed DMNLS and noise-perturbed NLS with varying dispersion for both the amplitude (compared at amplifiers) and phase. That is, *the phases and amplitudes agree for each individual MC sample as a function of noise across the whole transmission line.* These results, which may appear rather surprising given the "softness" of the phase, the variations in amplitude within the dispersion map for NLS with varying dispersion, and the complexity of the system (nonlinearity, dispersion, noise, large deviations, etc.), provide additional confirmation of the validity and robustness of the DMNLS equation in capturing the dynamics of dispersion-managed systems. Its usefulness is also increased by the availability of tools such as the perturbation theory that we used in this work.

The phase PDF presented in Figure 5 is centered around the computed mean. There is a discrepancy between the theoretical and simulated values of the mean phase. This discrepancy also appears for the NLS equation (i.e., for $s = 0$) [30] and is not well understood. As with the NLS equation, the discrepancy does not affect the variance or the shape of the PDF. It is also worth noting that this effect is quadratic in the noise variance, and therefore it is not picked up by the equations of perturbation theory in section 2.2, which is carried out only to first order. As a result, the discrepancy does not affect the method or the results presented here. It also does not affect the biasing calculation. We are still biasing toward the most likely path to some total phase change. The only difference is that the total change we get at the output is slightly more than what we biased for. This does not present a problem for our simulations since we sought to compute the phase PDF over a wide range of phase values, and thus we targeted several different output phases within the ISMC simulations.

6. Conclusions. In summary, we have outlined perturbation theory for the DMNLS equation and used that theory to guide ISMC simulations which we employed to directly calculate PDFs of phase and amplitude in dispersion-managed systems. We then compared these PDFs

to results from the original, unaveraged system. This comparison between the PDFs arising from simulations of the DMNLS and of the original NLS with varying dispersion provides a further significant validation of DMNLS as a model for dispersion-managed systems.

We should also mention that the PDF of noise-induced phase jitter was also addressed in [39]. That work, however, was limited to the NLS equation. Also, the results of the MC simulation in that work go down to only 10^{-6} in probability, which does not provide a real validation of the method, since one can often reach intermediate results with suboptimal biasing, as demonstrated in [7]. A more accurate test of the method is whether the coefficient of variation (as reflected visually in the smoothness of the curves for the PDF) remains small across the whole range of values of the PDF, i.e., arbitrarily far down as one wants to go into the tails. Visually, that is precisely what happens in our case, and this is therefore a good indication that our choice of biasing is indeed near optimal.

We believe that the methods described in this work will provide one of the steps that will enable the accurate computation of failure rates in dispersion-managed systems affected by noise, and we plan to work on such computations in the near future.

Appendix A. One-parameter family of shape functions for dispersion-managed solitons.
The nonlinear integral equation (2.14), whose solution yields the shape of the dispersion-managed solitons, contains three free parameters, $\bar{d}$, s, and A. The scaling invariance of the DMNLS equation, however, implies that, for each fixed value of $\bar{d}$, the dispersion-managed solitons can be described in terms of just a one-parameter family of shape functions, as we show next.

Let $u(t, z; A, s) = f(t; A, s)\, e^{iA^2 t/2}$ be a stationary dispersion-managed soliton, and let

$$(A.1) \qquad \hat{f}(\omega; A, s) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t; A, s)\, dt.$$

Then $\hat{f}$ solves the nonlinear integral equation (2.14), which with (2.12) becomes

$$(A.2) \qquad 2\pi^2\left(A^2 + \bar{d}\omega^2\right)\hat{f}_{(\omega)} = \iint \hat{f}_{(\omega+\omega')}\hat{f}_{(\omega+\omega'')}\hat{f}^*_{(\omega+\omega'+\omega'')} \frac{\sin(s\omega'\omega'')}{s\omega'\omega''}\, d\omega' d\omega''.$$

Now, for all $s \neq 0$, let $y = \sqrt{s}\,\omega$ and $\hat{f}(\omega; A, s) = \hat{\Psi}(\sqrt{s}\,\omega; A^2 s)$. Then $\hat{\Psi}(y; \mu)$ solves

$$(A.3) \qquad 2\pi^2\left(\mu + \bar{d}y^2\right)\hat{\Psi}_{(y)} = \iint \hat{\Psi}_{(y+y')}\hat{\Psi}_{(y+y'')}\hat{\Psi}^*_{(y+y'+y'')} \frac{\sin(y'y'')}{y'y''}\, dy' dy''.$$

The one-parameter family of functions $\hat{\Psi}(y; \mu)$ or its inverse Fourier transform

$$(A.4) \qquad \Psi(x; \mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy} \hat{\Psi}(y; \mu)\, dy$$

provides the shape of the dispersion-managed solitons for all values of A and s. It is easy to show that, for all values of A and s, it is

$$(A.5) \qquad f(t; A, s) = \frac{1}{\sqrt{s}} \Psi(t/\sqrt{s}; A^2 s).$$
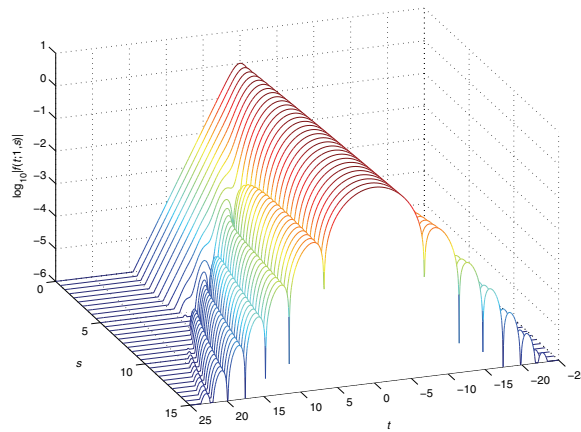
Figure 6. *One-parameter family of shape functions of dispersion-managed solitons.*

The transformation from ω to y becomes singular at $s = 0$. The apparent discontinuity of (A.5) at $s = 0$ is removable, however. To see this, we can compute the limit of the right-hand side of (A.5) as $s \to 0$ using the inverse transformation (namely, from y to ω) and expressing the shape functions as $\hat{\Psi}(y; \mu) = \hat{f}(\sqrt{s}\, y; \sqrt{\mu/s}, s)$, where $\hat{f}$ solves (A.2). In the limit $s \to 0$ we then obtain $\hat{f}(\omega; A, 0) = \pi \operatorname{sech}[\pi\omega/(2A)]$, which yields simply $f(t; A, 0) = A \operatorname{sech} At$. Note also that no problems exist at $A = 0$, since when $A = 0$ (A.2) yields the trivial solution for all s.

The above results allow one to precompute the one-parameter family of solutions only once with high accuracy and then rescale it as needed to obtain the desired shape in all situations. For example, one can compute $f(t; 1, s)$ for all values of s, obtaining the family of shapes shown in Figure 6. Then, for each value of A and s, one can use (A.5) to obtain

$$(A.6) \qquad\qquad f(t; A, s) = A\, f(At; 1, A^2 s).$$

The norms and inner products of the linear modes can be obtained in a similar way. Indeed,

$$(A.7a) \qquad \|\underline{y}_A\|^2/A = E/A = \frac{1}{\mu}\|\Psi\|^2, \qquad \|\underline{y}_\Omega\|^2/A = \frac{1}{\mu^3}\left\|\frac{\partial\Psi}{\partial x}\right\|^2,$$

where $\mu = A^2 s$ and $\Psi = \Psi(x; \mu)$. Similarly,

$$(A.7b) \qquad \|y_A\|^2 A = \frac{1}{\mu}\left\|\Psi + x\frac{\partial\Psi}{\partial x} + 2\mu\frac{\partial\Psi}{\partial\mu}\right\|^2, \qquad \|y_\Omega\|^2 A = \mu\|x\Psi\|^2.$$

Indeed, this is how the curves in Figure 1 were produced.

Appendix B. Numerical integration of the DMNLS equation. Numerical calculations of the pulse phase are more sensitive compared to those of amplitude, frequency, or position. Hence, careful choice of the numerical method is crucial in order to obtain reliable results. As shown in [1, 23], the DMNLS equation (2.8a) can be written as

$$(B.1) \qquad\qquad \frac{\partial u}{\partial z} = \frac{i}{2}\bar{d}\frac{\partial^2 u}{\partial t^2} + iN[u],$$

where the nonlinear term is $N[u] = \mathcal{F}^{-1}\big[\,\hat{N}[\hat{u}]\,\big]$, with

(B.2) $$\hat{N}[\hat{u}] = \big\langle \mathrm{e}^{iC(\zeta)\omega^2/2}\mathcal{F}\big[|q|^2q\big]\big\rangle,$$

where $\mathcal{F}$ is the Fourier transform operator (defined by (2.7)), the angled brackets denote average with respect to ζ over one period of the dispersion map, and $q(t, z, \zeta)$ is obtained from $u(t, z)$ and $C(\zeta)$ via (2.5). (Of course, for constant dispersion (that is, for the NLS equation) it is $C(\zeta) = 0$, and the average becomes an identity.) Thus, the double integral in the DMNLS equation can be efficiently evaluated using fast Fourier transforms (see [23] for further details). Once the integral is computed, the time stepping can be performed using any of the numerical methods for partial differential equations (PDEs). (The role of "time" here is of course played by the propagation variable z.) Several methods are accurate enough to reproduce the numerical results of [23].

Equation (B.1) is a semilinear PDE, and its numerical integration is a stiff problem because of the second derivative in the right-hand side. To overcome this difficulty, we use an integrating factor (IF) approach [11]. Introducing

(B.3a) $$\hat{v}(\omega, z) = \mathrm{e}^{i\bar{d}\omega^2 z/2}\hat{u}(\omega, z),$$

we have

(B.3b) $$\frac{\partial\hat{v}}{\partial z} = i\mathrm{e}^{i\bar{d}\omega^2 z/2}\hat{N}\big[\mathrm{e}^{-i\bar{d}\omega^2 z/2}\hat{v}\big].$$

Instead of solving the DMNLS equation directly, we then numerically integrate (B.3b), and, at each value of z, we recover $u(t, z)$ by trivially inverting (B.3a). As shown in [27] for various other PDEs, this approach removes the stiffness from the problem by treating the linear term exactly. The actual time stepping for (B.3b) is performed using a fourth-order Runge–Kutta method. As documented in [11], the IF-RK4 approach yields comparable efficiency to other numerical methods with the same order of accuracy.

An alternative approach with comparable efficiency to IF is a split-step Fourier method. Note, however, that, for $s \neq 0$, such an approach is not as convenient as in the case of constant dispersion ($s = 0$). This is because, unlike what happens for the NLS equation, when $s \neq 0$, it is not possible to integrate the nonlinear part exactly. Therefore, one would have to integrate the nonlinear part approximately (e.g., using a Runge–Kutta method, as was done in [11] for the Korteweg–de Vries equation).

Appendix C. Relaxation method for the optimal biasing problem for the phase. For simplicity, we just present the method in the case $s = 0$, that is, for the NLS equation. The method, however, works with trivial modifications for the DMNLS equation.

Recall that the optimal biasing for the phase is obtained by solving the boundary value problem (BVP) given by the second-order ordinary differential equation (ODE)

(C.1) $$\ddot{A} - \frac{\dot{A}^2}{2A} - \frac{bc^2}{A} + 2cA^2 = 0$$

subject to the boundary conditions

(C.2) $$A(0) = 1, \qquad \dot{A}(L) = 0,$$

where $b/A = (12 + \pi^2)/18A$ is the variance of the source term in the phase ODE and c is an arbitrary constant. Once $A(z)$ is known, $\phi(z)$ is recovered by solving the auxiliary ODE

$$\text{(C.3)} \qquad \dot{\phi} = \frac{1}{2}A^2 + \frac{bc}{A},$$

plus the initial condition $\phi(0) = 0$. The value of c is related to that of the output phase, $\phi(L)$. In particular, the case $c = 0$ has the unique solution $A(z) = 1$. (This is true even when $s \neq 0$.)

To solve the above BVP, we use a relaxation method. First, it is convenient to convert the problem into a first-order system. Let $\mathbf{x} = (x, y)^t$, where $x = A$ and $y = A'$. The ODE (C.1) then yields

$$\text{(C.4)} \qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \qquad \mathbf{f}(\mathbf{x}) = \begin{pmatrix} y \\ y^2/2x + bc^2/x - 2cx^2 \end{pmatrix}$$

together with the boundary conditions $x(0) = 1$ and $y(L) = 0$. Now we discretize the system of ODEs. We partition the interval $[0, L]$ into $\{z_o, \ldots, z_M\}$, with $z_m = m\Delta z$ for $m = 0, \ldots, M$ and $\Delta z = L/M$. Let $\mathbf{x}_m = \mathbf{x}(z_m)$, and introduce the $2 \times (M + 1)$ matrix $X = (\mathbf{x}_0, \ldots, \mathbf{x}_M)$ which collects the values of the dependent variables at all grid points. We then approximate (C.4) with

$$\text{(C.5)} \qquad \mathbf{E}_m(X) = 0, \qquad m = 1, \ldots, M,$$

where $\mathbf{E}_m(X)$ is the discrete version of $\mathbf{x}' = \mathbf{f}(\mathbf{x})$ evaluated at $z = z_m$:

$$\text{(C.6)} \qquad \mathbf{E}_m(X) = \mathbf{x}_m - \mathbf{x}_{m-1} - (\Delta z/2)\big(\mathbf{f}(\mathbf{x}_m) + \mathbf{f}(\mathbf{x}_{m-1})\big).$$

The above are a system of $2M$ nonlinear equations for the $2M$ unknowns $x_1, \ldots, x_M$ and $y_0, \ldots, y_{M-1}$. (Note that the boundary conditions for $\mathbf{x}(z)$ imply simply $x_0 = 1$ and $y_M = 0$.) To solve this nonlinear system of equations, we use Newton's method. Introduce the $2 \times M$ matrix $E(X) = (\mathbf{E}_1, \ldots, \mathbf{E}_M)$, and expand $E(X)$ in Taylor series as

$$\text{(C.7)} \qquad E(X + \Delta X) = E(X) + J_E(X)\Delta X + O(\Delta X^2),$$

where $J_E(X)$ is the Jacobian matrix of E as a function of the $2M$ independent variables $x_1, \ldots, x_M$ and $y_0, \ldots, y_{M-1}$. (Note that there is a slight abuse of notation here, since X also contains the quantities x_0 and y_M, which are fixed.) The requirement $E(X + \Delta X) = 0$ then yields, to leading order,

$$\text{(C.8)} \qquad \Delta X = -(J_E(X))^{-1}E(X)$$

(again with a slight abuse of notation, since x_0 and y_M are unaffected by the iteration). Componentwise, writing $E = (E_{jm})$ with $j = 1, 2$ and $m = 1, \ldots, M$ and $X = (X_{jm})$ with $j = 1, 2$ and $0 = 1, \ldots, M$, we have

(C.9)
$$\big(J_E(X)\big)_{im}^{\ jk} = \frac{\partial E_{im}}{\partial X_{jk}} = \left(\delta_i^{\ j} - (\Delta z/2)\big(J_f(\mathbf{x}_m)\big)_i^{\ j}\right)\delta_m^{\ k} - \left(\delta_i^{\ j} + (\Delta z/2)\big(J_f(\mathbf{x}_{m-1})\big)_i^{\ j}\right)\delta_{m-1}^{\ k},$$

where

(C.10) $$J_f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\[2mm] \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -y^2/(2x^2) - bc^2/x^2 - 4cx & y/x \end{pmatrix}.$$

The implementation of this algorithm requires one to properly deal with the tensorial nature of $J_E(X)$, but this can be done effectively using computer algebra packages such as MATLAB.

The case $s \neq 0$ is slightly more complicated because we have only numerical representations of the source variances. In this case, one can fit a fourth-order polynomial in A to each curve and then proceed as above. The same approach can be used when employing numerical continuation software such as XPP/AUTO.

Appendix D. Large map strength approximation of the optimal phase paths. The optimal biasing problem for the phase requires that amplitude follows the second-order differential equation (4.30), which is rewritten below for clarity, derived in section 4.4,

(D.1) $$2\ddot{A}\frac{1}{\sigma_A^2} + \dot{A}^2\frac{\partial}{\partial A}\left[\frac{1}{\sigma_A^2}\right] + c^2\frac{\partial}{\partial A}\left[\sigma_\Phi^2\right] + 2c\,A = 0,$$

where c is an arbitrary integration constant. The amplitude is subject to the boundary conditions

(D.2) $$A(0) = A_o, \qquad A'(L) = 0.$$

Again, $\Phi(z)$ is obtained from $A(z)$ by solving the auxiliary ODE $\phi' = A^2/2 + c\sigma_\Phi^2$ plus the initial condition $\phi(0) = 0$. The value of c is related to that of the output phase, $\phi(L)$, and the case $c = 0$ has the unique solution $A(z) = 1$.

For large values of map strength, the variances of the source terms can be approximated by monomial functions of A, and thus the optimal biasing equation can be written in terms of the amplitude. Specifically, recall that $\sigma_Q^2 = \sigma^2 \|\underline{y}_Q\|^2/\langle \underline{y}_Q, y_Q \rangle^2$. For large map strengths we approximate

(D.3) $$\|\underline{y}_A\|^2 = E(A, s) = k_1 A^2, \qquad \|\underline{y}_\Phi\|^2 = k_2, \qquad \langle \underline{y}_A, y_A \rangle = \langle \underline{y}_\Phi, y_\Phi \rangle = k_3 A.$$

This implies $\sigma_A^2 = \sigma^2 k_1/k_3^2$ and $\sigma_\Phi^2 = \sigma^2 k_2/(k_3 A)^2$. Note that the constants, k_1, k_2, k_3, depend on the map strength. For the comparison plotted in Figure 3, we used the value $s = 4$.

The optimal biasing for the phase then simplifies into the BVP given by the ODE

(D.4) $$A'' - \frac{bk^2}{A^3} + kA = 0,$$

where $k = c\sigma^2 k_1/k_3^2$ and $b = k_1/k_2$. This ODE can be integrated once to give $(A')^2 + bk^2/A^2 + kA^2 = 2k\,C$, where the right-hand side is another arbitrary integration constant. (Note that the above equation implies $C = A^2(L)/2 + bk/2A^2(L)$.) This is a separable first-order ODE, which can be conveniently written as

(D.5) $$\frac{(AA')^2}{2CA^2 - A^4 - bk} = k.$$

The quantity

$$\Delta = 2CA^2 - A^4 - bk, \tag{D.6}$$

which is the denominator in the left-hand side of the ODE (D.5), must have constant sign equal to that of k. This implies that $C > \max(bk, 0)$. This also means that the range of A is limited to the following:

$$C - \sqrt{C^2 - bk} < A^2 < C + \sqrt{C^2 - bk} \qquad \text{when } k > 0, \tag{D.7a}$$

$$A^2 > C + \sqrt{C^2 - bk} \qquad \text{when } k < 0, \tag{D.7b}$$

while for $k = 0$ it is simply $A = A_o$. We will consider the case $k \geq 0$. (The case $k \leq 0$ can be handled in a similar way.) Letting $e = A^2 - C$, (D.5) becomes simply

$$e'/\sqrt{\Delta} = \sqrt{k}, \tag{D.8}$$

where now $\Delta = C^2 - bk - e^2$. Rescaling yet again by letting $y = e/\sqrt{C^2 - bk}$, we obtain $y'/\sqrt{1 - y^2} = \sqrt{k}$, which is trivially integrated to give

$$y(z) = \sin(\sqrt{k}z + \alpha), \tag{D.9}$$

where $\alpha = \arcsin(y_o)$ and $y_o = (A_o^2 - C)/\sqrt{C^2 - bk}$. Without loss of generality, we can choose the fundamental branch of the arcsine. So we have $|\alpha| \leq \pi/2$.

We now use the boundary condition at $z = L$ obtaining

$$\cos(\sqrt{k}L + \alpha) = 0. \tag{D.10}$$

This condition fixes y_o via α and therefore provides an equation that determines the integration constant C. Explicitly, if $[\sqrt{k}L/\pi] = n$, we get $\alpha = (n + \frac{1}{2})\pi - \sqrt{k}L$, implying

$$\frac{A_o^2 - C}{\sqrt{C^2 - bk}} = (-1)^n \cos(\sqrt{k}L). \tag{D.11}$$

Note that when $k = 0$ we have $n = 0$, and the equation yields simply $C = A_o^2/2$. The above condition yields

$$C^2 \sin^2 \sqrt{k}L - 2A_o^2 C + (A_o^4 + bk \cos^2 \sqrt{k}L) = 0. \tag{D.12}$$

This is solved by $C = C_\pm$, with

$$C_\pm = \left[A_o^2 \pm \sqrt{A_o^4 \cos^2 \sqrt{k}L - \tfrac{1}{4}bk \sin^2 2\sqrt{k}L} \right] \csc^2 \sqrt{k}L. \tag{D.13}$$

It is necessary to choose the branch $C = C_+$ to achieve the output phase which the optimal phase path reaches at $z = L$ for $s = 4$ case in Figure 3.

The optimal amplitude path thus obtained is

$$A(z) = \sqrt{C + \sqrt{C^2 - bk}\, \sin(\sqrt{k}z + \alpha)}, \tag{D.14}$$

with $\alpha = \alpha(C)$ and $C = C_+$. We can now substitute $A(z)$ into the ODE for ϕ to find the optimal phase path and the output phase as a function of c:

$$(D.15) \qquad \phi' = A^2/2 + bc/A^2.$$

The problem is thus reduced to quadratures. Explicitly, its solution is

$$(D.16) \quad \phi(z) = Cz/2 - \frac{\delta}{2\sqrt{k}}\big[\cos(\sqrt{c}z + \alpha) - \cos\alpha\big]$$
$$+ \frac{2\sqrt{b}c}{k}\bigg\{\arctan\bigg[\frac{\delta + C\tan\big(\frac{1}{2}(\sqrt{k}z + \alpha)\big)}{\sqrt{bk}}\bigg] - \arctan\bigg[\frac{\delta + C\tan\big(\frac{1}{2}\alpha\big)}{\sqrt{bk}}\bigg]\bigg\},$$

where $\delta = \sqrt{C^2 - bk}$ and C is as before. This yields $\phi(L)$ explicitly as a function of c. The phase and amplitude paths which this approximation yields are plotted in Figure 3.

REFERENCES

[1] M. J. ABLOWITZ AND G. BIONDINI, *Multiple scale dynamics in communication systems with strong dispersion management*, Optics Letters, 23 (1998), pp. 1668–1670.

[2] M. J. ABLOWITZ, G. BIONDINI, S. CHAKRAVARTY, R. B. JENKINS, AND J. R. SAUER, *Four-wave mixing in wavelength-division-multiplexed soliton systems: Damping and amplification*, Optics Letters, 21 (1996), pp. 1646–1648.

[3] M. J. ABLOWITZ, G. BIONDINI, AND E. S. OLSEN, *Incomplete collisions of wavelength-division multiplexed dispersion-managed solitons*, J. Opt. Soc. Amer. B Opt. Phys., 18 (2001), pp. 577–583.

[4] M. J. ABLOWITZ, B. ILAN, AND S. T. CUNDIFF, *Carrier-envelope phase slip of ultrashort dispersion managed solitons*, Optics Letters, 29 (2004), pp. 1808–1810.

[5] G. P. AGRAWAL, *Fiber Optics Communication Systems*, 3rd ed., Wiley, New York, 2002.

[6] J. D. ANIA-CASTANON, T. J. ELLINGHAM, R. IBBOTSON, X. CHEN, L. ZHANG, AND S. K. TURITSYN, *Ultralong Raman fiber lasers as virtually lossless optical media*, Phys. Rev. Lett., 96 (2006), 023902.

[7] G. BIONDINI, W. L. KATH, AND C. R. MENYUK, *Importance sampling for polarization mode dispersion: Techniques and applications*, IEEE J. Lightwave Technol., 22 (2004), pp. 1201–1215.

[8] A. D. CAPOBIANCO, G. NALESSO, A. TONELLO, F. CONSOLANDI, AND C. DE ANGELIS, *Noise evolution and soliton internal modes in dispersion-managed fiber systems*, Optics Letters, 28 (2003), pp. 1754–1756.

[9] S. A. DEREVYANKO, S. K. TURITSYN, AND D. A. YAKUSHEV, *Fokker-Planck equation approach to the description of soliton statistics in optical fiber transmission systems*, J. Opt. Soc. Amer. B Opt. Phys., 22 (2005), pp. 743–752.

[10] E. J. DOEDEL, A. R. CHAMPNEYS, T. F. FAIRGRIEVE, Y. A. KUZNETSOV, B. SANDSTEDE, AND X. J. WANG, *AUTO97: Continuation and Bifurcation Software for Ordinary Differential Equations*, available from http://cmvl.cs.concordia.ca, 1997.

[11] B. FORNBERG AND T. A. DRISCOLL, *A fast spectral algorithm for nonlinear wave equations with linear dispersion*, J. Comput. Phys., 155 (1999), pp. 456–467.

[12] I. R. GABITOV AND S. K. TURITSYN, *Averaged pulse dynamics in a cascaded transmission system with passive dispersion compensation*, Optics Letters, 21 (1996), pp. 327–329.

[13] J. P. GORDON AND H. A. HAUS, *Random walk of coherently amplified solitons in optical fiber transmission*, Optics Letters, 11 (1986), pp. 665–667.

[14] J. P. GORDON AND L. F. MOLLENAUER, *Phase noise in photonic communications systems using linear amplifiers*, Optics Letters, 15 (1990), pp. 1351–1353.

[15] A. HASEGAWA AND Y. KODAMA, *Guiding-center soliton in optical fibers*, Optics Letters, 15 (1990), pp. 1443–1445.

[16] K.-P. HO, *Phase statistics of the soliton*, J. Opt. Soc. Amer. B Opt. Phys., 21 (2004), pp. 266–272.

[17] E. IANNONE, F. MATERA, A. MECOZZI, AND M. SETTEMBRE, *Nonlinear Optical Communication Networks*, Wiley, New York, 1998.

[18] S. KUMAR, *Analysis of nonlinear phase noise in coherent fiber-optic systems based on phase shift keying*, IEEE J. Lightwave Technol., 27 (2009), pp. 4722–4733.

[19] S. KUMAR AND F. LEDERER, *Gordon-Haus effect in dispersion-managed soliton systems*, Optics Letters, 22 (1998), pp. 1870–1872.

[20] T. I. LAKOBA AND D. E. PELINOVSKY, *Persistent oscillations of scalar and vector dispersion-managed solitons*, Chaos, 10 (2000), pp. 539–550.

[21] A. P. T. LAU AND J. M. KAHN, *Signal design and detection in presence of nonlinear phase noise*, IEEE J. Lightwave Technol., 25 (2007), pp. 3008–3016.

[22] J. J. MCFERRAN, E. N. IVANOV, A. BARTELS, G. WILPERS, C. W. OATES, S. A. DIDDAMS, AND L. HOLLBERG, *Low-noise synthesis of microwave signals from an optical source*, IEEE Electron. Lett., 41 (2005), pp. 650–651.

[23] J. LI, E. T. SPILLER, AND G. BIONDINI, *Noise-induced perturbations of dispersion-managed solitons*, Phys. Rev. A, 75 (2007), 053818.

[24] C. J. MCKINSTRIE, *Gordon-Haus timing jitter in dispersion-managed systems with distributed amplification*, Opt. Comm., 200 (2001), pp. 165–177.

[25] C. J. MCKINSTRIE AND T. I. LAKOBA, *Probability density function for energy perturbations of isolated optical pulses*, Opt. Expr., 11 (2003), pp. 3628–3648.

[26] A. MECOZZI, *Probability density functions of the nonlinear phase noise*, Optics Letters, 29 (2004), pp. 673–675.

[27] P. A. MILEWSKI AND E. G. TABAK, *A pseudospectral procedure for the solution of nonlinear wave equations with examples from free-surface flows*, SIAM J. Sci. Comput., 21 (1999), pp. 1102–1114.

[28] L. F. MOLLENAUER AND J. P. GORDON, *Solitons in Optical Fibers: Fundamentals and Applications*, Academic Press, New York, 2006.

[29] R. O. MOORE, G. BIONDINI, AND W. L. KATH, *Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems*, Optics Letters, 28 (2003), pp. 105–107.

[30] R. O. MOORE, G. BIONDINI, AND W. L. KATH, *A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons*, SIAM J. Appl. Math., 67 (2007), pp. 1418–1439.

[31] N. R. NEWBURY AND W. C. SWANN, *Low-noise fiber-laser frequency combs*, J. Opt. Soc. Amer. B Opt. Phys., 24 (2007), pp. 1756–1770.

[32] A. PAPOULIS, *Probability, Random Variables and Stochastic Processes*, McGraw–Hill, New York, 1991.

[33] D. E. PELINOVSKY, *Instabilities of dispersion-managed solitons in the normal dispersion regime*, Phys. Rev. E, 62 (2000), pp. 4283–4293.

[34] Q. QURAISHI, S. T. CUNDIFF, B. ILAN, AND M. J. ABLOWITZ, *Dynamics of nonlinear and dispersion managed solitons*, Phys. Rev. Lett., 94 (2005), 243904.

[35] L. J. RICHARDSON, W. FORYSIAK, AND N. J. DORAN, *Dispersion-managed soliton propagation in short-period dispersion maps*, Optics Letters, 25 (2000), pp. 1010–1012.

[36] N. J. SMITH, W. FORYSIAK, AND N. J. DORAN, *Reduced Gordon-Haus jitter due to enhanced power solitons in strongly dispersion-managed systems*, Electron. Lett., 32 (1996), pp. 2085–2086.

[37] E. T. SPILLER AND W. L. KATH, *A method for determining most probable errors in nonlinear lightwave systems*, SIAM J. Appl. Dyn. Syst., 7 (2008), pp. 868–894.

[38] E. T. SPILLER, W. KATH, R. MOORE, AND C. MCKINSTRIE, *Computing large signal distortions and bit-error ratios in DPSK transmission systems*, IEEE Photon. Technol. Lett., 17 (2005), pp. 1022–1024.

[39] A. TONELLO, S. WABNITZ, I. GABITOV, AND R. INDIK, *Importance sampling of Gordon-Mollenauer soliton phase noise in optical fibers*, IEEE Photon. Technol. Lett., 18 (2006), pp. 886–888.

[40] A. TONELLO, A. D. CAPOBIANCO, G. NALESSO, F. GRINGOLI, AND C. DE ANGELIS, *Perturbations, internal modes and noise in dispersion-managed soliton transmission*, Opt. Comm., 246 (2005), pp. 393–403.

[41] J. YE AND S. T. CUNDIFF, EDS., *Femtosecond Optical Frequency Comb: Principle, Operation and Applications*, Springer, New York, 2004.