SIAM J. APPL. MATH.
Vol. 67, No. 5, pp. 1418–1439

A METHOD TO COMPUTE STATISTICS OF LARGE, NOISE-INDUCED PERTURBATIONS OF NONLINEAR SCHRÖDINGER SOLITONS*

R. O. MOORE†, G. BIONDINI‡, AND W. L. KATH§

Abstract. We demonstrate in detail the application of importance sampling to the numerical simulation of large noise-induced perturbations in soliton-based optical transmission systems governed by the nonlinear Schrödinger equation. The method allows one to concentrate numerical Monte Carlo simulations around the noise realizations that are most likely to produce the large pulse deformations connected with errors, and yields computational speedups of several orders of magnitude over standard Monte Carlo simulations. We demonstrate the method by using it to calculate the probability density functions associated with pulse amplitude, frequency, and timing fluctuations in a prototypical soliton-based communication system.

Key words. nonlinear Schrödinger equation, optical fibers, solitons, noise, Monte Carlo simulations, importance sampling

AMS subject classifications. 35Q51m, 35Q55, 65C20, 65C05, 78A40

DOI. 10.1137/060650775

1. Introduction.

The development of high-bit-rate data transmission over optical fibers is one of the major technological achievements of the late 20th century. The information-carrying capacity of such systems has increased by several orders of magnitude over the past quarter-century. There are limits imposed on capacities, however, by various transmission impairments that distort and degrade the signal in a number of ways [1, 2]. One common source of impairments in lightwave communication systems is the amplified spontaneous emission (ASE) noise generated by the erbium-doped fiber amplifiers (EDFAs) used to compensate loss in the fiber [1, 2]. This additive noise perturbs the propagating pulses, producing amplitude, frequency, timing, and phase jitter, which can then lead to bit errors [3, 4, 5].

Since ASE noise is a stochastic phenomenon, Monte Carlo simulations can be used to determine its effects on a system. The direct calculation of bit error rates with standard Monte Carlo simulations is impossible, however. Because data transmission rates are so high (currently, 10 Gb/s or more per channel, with tens of channels usually present per fiber) and errors must be handled by much slower electronic equipment, error rates are required to be very small, typically one error per trillion bits or lower. As a result, an exceedingly large number of Monte Carlo realizations would be needed to observe even a single transmission error, and even more would be required to obtain reliable error estimates.

*Received by the editors January 24, 2006; accepted for publication (in revised form) March 30, 2007; published electronically July 20, 2007.
http://www.siam.org/journals/siap/67-5/65077.html

†Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 (rmoore@njit.edu). The work of this author was supported in part by the National Science Foundation under grant DMS-0511091.

‡Department of Mathematics, State University of New York, Buffalo, NY 14260 (biondini@buffalo.edu). The work of this author was supported by National Science Foundation grant DMS-0506101.

§Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208-3125 (kath@northwestern.edu). The work of this author was supported by the National Science Foundation (grants DMS-0101476 and DMS-0406513) and by the Air Force Office of Scientific Research (FA9550-04-1-0289).

To overcome this limitation, one approximation is to calculate numerically the variances of pulse amplitude and/or timing and then to extrapolate the results into the tails of the probability density function (pdf) by assuming it to be Gaussian. It is clear, however, that this procedure is inadequate, since nonlinearity and pulse interactions can both contribute to make the resulting distributions non-Gaussian [6, 7, 8]. Nonlinearity arises from several sources, such as self-phase modulation due to the fiber's nonlinear refractive index [9, 10] as well as the nonlinear conversion of optical energy into electrical energy by the photodetector [8].

Various techniques have recently been proposed to address the difficulty in calculating accurate statistics for rare events [8, 11, 12, 13, 14, 15, 16]. One promising approach is a technique known as importance sampling (IS) [17, 18]. In general, IS works by concentrating Monte Carlo samples on those configurations that are most likely to lead to transmission errors, thus significantly speeding up the simulations. Previously, we have successfully applied this technique to the direct simulation of transmission impairments caused by polarization-mode dispersion [14, 15]. More recently, we have presented numerical results demonstrating that IS can also be applied to simulations of ASE-induced transmission impairments [16]. The purpose of this paper is to describe in detail the methods used to produce these numerical simulations. For simplicity, we consider the case of a soliton-based transmission system (where, in the absence of noise, the pulse shape remains fixed), but it is anticipated that the technique can be extended to more realistic systems and more general transmission formats. The advantages of the method are substantial, allowing an increase in efficiency of several orders of magnitude over standard Monte Carlo simulations.

2. Solitons and amplifier noise. The propagation of pulses in an optical fiber with periodically spaced amplification is governed by the nonlinear Schrödinger (NLS) equation [9, 10, 16], which in dimensionless units is

$$i\frac{\partial u}{\partial z} + \frac{1}{2}\frac{\partial^2 u}{\partial t^2} + |u|^2 u = i\sum_{n=1}^{N_a} \delta(z - nz_a) f_n(t). \tag{2.1}$$

Here z and t are distance and retarded time, u is the pulse's electromagnetic field envelope, N_a is the number of amplifiers, and z_a is the amplifier spacing. (The nondimensionalization and our choice of units are described in detail in Appendix A.) In this equation, the periodic power variations due to fiber loss and amplification have been averaged out [9].

The term $f_n(t)$ is the noise added at each amplifier; when the pulse reaches an amplifier at $z = nz_a$ (where z_a is the dimensionless amplifier spacing and $n = 1, 2, \ldots, N_a$, with N_a being the total number of amplifiers in the transmission line), a small amount of noise $f_n(t)$ is added to u: $u(nz_a^+, t) = u(nz_a^-, t) + f_n(t)$, as seen by integrating (2.1) across $z = nz_a$. (Recall that we have averaged out loss and gain; this means that the noise is the only effect remaining at the amplifiers.) The amplifier noise $f_n(t)$ can be modeled as classical zero-mean white noise:

$$\langle f_n(t) \rangle = 0, \tag{2.2a}$$

$$\langle f_m(t) f_n(t') \rangle = 0, \tag{2.2b}$$

$$\langle f_m(t) f_n^\dagger(t') \rangle = \sigma^2 \delta_{mn} \delta(t - t'), \tag{2.2c}$$

where $\langle \cdot \rangle$ denotes ensemble average, the superscript † denotes the complex conjugate, δ_{mn} and $\delta(t-t')$ are the Kronecker and Dirac deltas, respectively, and σ^2 is a parameter describing the noise power. Technically speaking, (2.2c) is not mathematically

correct, since as written it implies an infinite noise bandwidth and thus produces infinite noise power. Any physical system (or any numerical computation) necessarily has a finite noise bandwidth [19]. When calculating amplitude, frequency, and timing fluctuations, however, the specific value of the noise bandwidth is unimportant if it is larger than the soliton bandwidth (this is the case in practice), because only those components of the noise within the same spectral range as the soliton will affect these soliton parameters.

Without the noise term, (2.1) admits the well-known soliton solution

$$u_s(z,t) = u_0(z,t)\, e^{i\Theta(z,t)}, \tag{2.3a}$$

with

$$u_0(z,t) = A\,\mathrm{sech}[A(t-T(z))], \qquad \Theta(z,t) = \Omega\, t + \Phi(z), \tag{2.3b}$$

where $T(z) = T_0 + \Omega z$ and $\Phi(z) = \Phi_0 + \frac{1}{2}(A^2 - \Omega^2)z$ and where the four soliton parameters A, Ω, T_0, and Φ_0 are constant. When noise is added at each amplifier, part of the noise is incorporated into the soliton, where it produces small changes of the soliton parameters [9, 10]. The rest of the noise propagates along with the perturbed soliton. This process is repeated at each amplifier, resulting in a random walk of the four quantities A, Ω, T, and Φ [4, 5]. For typical system configurations, the noise amplitude at each amplifier is small, and thus the noise-induced changes of the soliton parameters at each individual amplifier are also usually small. In rare cases, however, these individual contributions combine to produce large deviations, resulting in potential transmission errors. Because these large pulse deformations are rare, estimating their probability is difficult.

2.1. Soliton perturbation theory. Soliton perturbation theory (SPT) is a standard method that has been used to approximate the effect of noise upon propagating pulses (e.g., see [9, 20, 21, 22, 23]). Rather than using it directly to obtain an analytical approximation to the perturbed pulse, however, here we will use it only as a tool to guide numerical simulations. The key information that is needed to do this comes from the dependence of the soliton solution, equation (2.3), upon the parameters A, Ω, T_0, and Φ_0. Since any value of these parameters is permitted, no resistance is encountered if the noise at an amplifier changes one of them. This lack of resistance allows large variations to build up after many amplifiers. Furthermore, frequency fluctuations change the group velocity of the pulse, and subsequent propagation integrates this velocity shift to produce a large timing shift (as reflected in the dependence of T on Ω).

The small noise-induced changes of the soliton parameters at a single amplifier can be estimated by decomposing $u(z,t)$ into its soliton and radiative (nonsoliton) components,

$$u(z,t) = [u_0(z,t) + v(z,t)]\, e^{i\Theta}, \tag{2.4}$$

and by linearizing the NLS equation (2.1) around the soliton solution (2.3):

$$\frac{\partial v}{\partial z} = L\, v, \qquad L\, v := \frac{i}{2}\frac{\partial^2 v}{\partial t^2} - \frac{i}{2}A^2 v + 2i|u_0|^2 v + i u_0^2\, v^\dagger. \tag{2.5}$$

Importantly, the linearized NLS operator L is non–self-adjoint and nonnormal, and its generalized nullspace admits four modes (localized eigenfunctions) $v_k(z,t)$ ($k = A, \Omega, T, \Phi$) satisfying [9, 10, 24]

$$L\, v_A = A\, v_\Phi, \qquad L\, v_\Omega = v_T, \qquad L\, v_T = 0, \qquad L\, v_\Phi = 0. \tag{2.6}$$

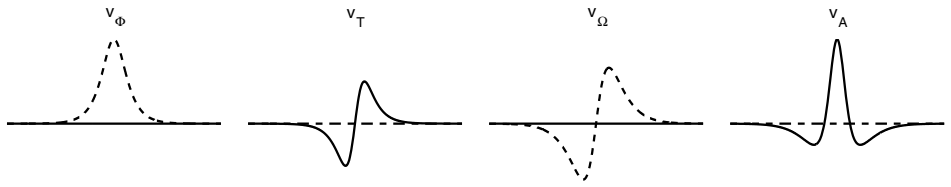

FIG. 1. *The real (solid) and imaginary (dashed) parts of the four neutral modes of the linearized NLS equation associated with the soliton solution* (2.3).

Explicitly, these four modes are

$$v_A(z,t) = \frac{1}{A}\frac{\partial}{\partial t}[(t-T)u_0], \tag{2.7a}$$

$$v_\Omega(z,t) = i(t-T)\,u_0, \tag{2.7b}$$

$$v_T(z,t) = -\frac{\partial u_0}{\partial t}, \tag{2.7c}$$

$$v_\Phi(z,t) = iu_0. \tag{2.7d}$$

Each of these linear modes is associated with a continuous symmetry of the NLS equation [23]: invariance with respect to scaling, Galilean boosts, translations, and phase rotations, respectively. Note also from (2.6) that the timing and phase modes are actual eigenfunctions, whereas the amplitude and frequency modes are generalized eigenfunctions. This is related to two of these symmetries giving rise to modified conservation laws, which are directly related to a change of the pulse frequency producing a subsequent change in its velocity or a change of the pulse amplitude producing a change in its phase upon propagation [23]. These four modes, shown in Figure 1, are also associated with changes of the soliton parameters:[1]

$$\frac{\partial u_s}{\partial A} = v_A\, e^{i\Theta}, \tag{2.8a}$$

$$\frac{\partial u_s}{\partial \Omega} = v_\Omega\, e^{i\Theta} + T v_\Phi\, e^{i\Theta}, \tag{2.8b}$$

$$\frac{\partial u_s}{\partial T} = v_T\, e^{i\Theta}, \tag{2.8c}$$

$$\frac{\partial u_s}{\partial \Phi} = v_\Phi\, e^{i\Theta}. \tag{2.8d}$$

In fact, removing secular terms from the forced linearized NLS equation obtained by adding the right-hand side of (2.1) to (2.5), one finds the local changes of the soliton parameters at the nth amplifier produced by the noise $f_n(t)$ [9, 10]:

$$\Delta A_n = \mathrm{Re}\int \underline{v}_A^\dagger(z,t)e^{-i\Theta} f_n(t)\,dt, \tag{2.9a}$$

$$\Delta \Omega_n = \mathrm{Re}\int \underline{v}_\Omega^\dagger(z,t)e^{-i\Theta} f_n(t)\,dt, \tag{2.9b}$$

$$\Delta T_n = \mathrm{Re}\int \underline{v}_T^\dagger(z,t)e^{-i\Theta} f_n(t)\,dt, \tag{2.9c}$$

$$\Delta \Phi_n = \mathrm{Re}\int (\underline{v}_\Phi(z,t) - T\,\underline{v}_\Omega(z,t))^\dagger e^{-i\Theta} f_n(t)\,dt, \tag{2.9d}$$

[1] Note that the derivatives on the left-hand side of (2.8) are taken with respect to the variables A, Ω, T, and Φ, with the other three variables kept constant. A different choice of parametrization for the soliton solution in (2.3) would lead to different results.

where the integrals are from $-\infty$ to ∞. The functions $\underline{v}_k(z,t)$ are the modes of the *adjoint* linearized NLS operator, defined by $L^{\mathrm{adj}}\underline{v} = -\frac{i}{2}\partial_t^2\underline{v} + \frac{i}{2}A^2\underline{v} - 2i|u_0|^2\underline{v} + iu_0^2\,\underline{v}^\dagger$ and the inner product $\langle \underline{v}, v\rangle = \mathrm{Re}\int \underline{v}^\dagger v\,dt$ [10]. They are

$$\underline{v}_A = -iv_\Phi, \qquad \underline{v}_\Omega = iv_T/A, \qquad \underline{v}_T = iv_\Omega/A, \qquad \underline{v}_\Phi = iv_A. \tag{2.10}$$

(The adjoint modes are easily obtained from those in (2.7), noting that $L^{\mathrm{adj}}(\underline{v}) = iL(i\underline{v})$.) Together, the modes of L and L^{adj} form a biorthonormal basis of the tangent space corresponding to infinitesimal changes in the soliton parameters at a given amplifier,[2] and the source terms in (2.9) represent the projection of the noise onto this basis.

2.2. Noise-induced amplitude, frequency, timing, and phase jitter. Equations (2.9) establish a direct projection from the infinite-dimensional noise which is added at each amplifier to a discrete random walk for the four soliton parameters. In particular, the equations can be easily integrated, including the unperturbed evolution in between amplifiers, to obtain the final values of amplitude A, frequency Ω, timing T, and phase Φ:

$$A_{\mathrm{out}} = A_0 + \sum_{n=1}^{N_a} \Delta A_n, \tag{2.11a}$$

$$\Omega_{\mathrm{out}} = \Omega_0 + \sum_{n=1}^{N_a} \Delta\Omega_n, \tag{2.11b}$$

$$T_{\mathrm{out}} = T_0 + N_a z_a \Omega_0 + \sum_{n=1}^{N_a} \Delta T_n + \sum_{n=1}^{N_a} (N_a + 1 - n)\, z_a \Delta\Omega_n, \tag{2.11c}$$

$$\begin{aligned}\Phi_{\mathrm{out}} = \Phi_0 + \frac{1}{2}N_a z_a (A_0^2 - \Omega_0^2) + \sum_{n=1}^{N_a}\Delta\Phi_n + \sum_{n=1}^{N_a}(N_a+1-n) z_a (A_0\Delta A_n - \Omega_0\Delta\Omega_n) \\ + \frac{1}{2}\sum_{n=1}^{N_a}\sum_{m=1}^{N_a}[N_a - \max(n,m)] z_a (\Delta A_n \Delta A_m - \Delta\Omega_n\Delta\Omega_m).\end{aligned} \tag{2.11d}$$

The fourth term in (2.11c) and the fourth and fifth terms in (2.11d) arise from the above-mentioned Galilean invariance of the NLS equation. Whereas amplitude and timing jitter are the most important failure mechanisms for communication lines using standard on-off keying receivers, phase fluctuations are of critical importance when the receivers are phase-sensitive, as is the case for phase-shift or differential phase-shift keying [25].

Owing to (2.2) and (2.9), ΔA_n, ΔT_n, $\Delta\Omega_n$, and $\Delta\Phi_n$ are Gaussian-distributed random variables at each amplifier, with variances

$$\langle \Delta A_{n+1}^2\rangle = A_n\sigma^2, \tag{2.12a}$$

$$\langle \Delta\Omega_{n+1}^2\rangle = \sigma^2 A_n/3, \tag{2.12b}$$

$$\langle \Delta T_{n+1}^2\rangle = \pi^2\sigma^2/(12A_n^3), \tag{2.12c}$$

$$\langle \Delta\Phi_{n+1}^2\rangle = (1 + \pi^2/12 + T_n^2)\sigma^2/3A_n, \tag{2.12d}$$

[2]Note that, as usual, a true linear mode of the linearized NLS operator corresponds to a generalized mode of the adjoint operator, and vice-versa.

respectively. Note that all of these variances depend on the value of the soliton amplitude immediately prior to arrival at the amplifier, and that the phase variance depends on the soliton position; this is because the time-dependent term in the phase in (2.3b) is not defined to be zero at $t = T(z)$. For small deviations of amplitude and position, one can approximate these variances with their initial values (assuming without loss of generality that the initial position is zero):

(2.13)
$$\sigma_A^2 := A_0\sigma^2, \quad \sigma_\Omega^2 := \sigma^2 A_0/3, \quad \sigma_T^2 := \pi^2\sigma^2/(12A_0^3), \quad \sigma_\Phi^2 := (1+\pi^2/12)\sigma^2/3A_0.$$

The variances of the final soliton amplitude, frequency, and position timing are then easily computed to be

$$\langle A_{\text{out}}^2\rangle \simeq N_a\sigma_A^2, \tag{2.14a}$$

$$\langle \Omega_{\text{out}}^2\rangle \simeq N_a\sigma_\Omega^2, \tag{2.14b}$$

$$\langle T_{\text{out}}^2\rangle \simeq N_a\sigma_T^2 + N_a(N_a+1)(2N_a+1)\,\sigma_\Omega^2\, z_a^2/6, \tag{2.14c}$$

$$\langle \Phi_{\text{out}}^2\rangle \simeq N_a\sigma_\Phi^2 + N_a(N_a+1)(2N_a+1)\,\sigma_A^2\, z_a^2/6, \tag{2.14d}$$

respectively. The cubic dependence on N_a of the growth of $\langle T_{\text{out}}^2\rangle$ and $\langle \Phi_{\text{out}}^2\rangle$ is a discrete analogue to the cubic dependence on distance in a distributed noise approximation, used by Gordon and Haus and by Gordon and Mollenauer, respectively, to derive upper limits for the error-free propagation distance of a soliton transmission system [4, 5].

These calculations, however, are not sufficient to give an accurate estimate of noise-induced transmission penalties, for several reasons. First, the variances in (2.14) are correct only for small deviations of the pulse amplitude. Second, even though the noise is Gaussian-distributed, the full noise-induced changes of the soliton parameters are not necessarily Gaussian. In particular, the variance of each amplitude shift depends on the previous value of the amplitude, which causes the distribution of A to deviate significantly from Gaussian. A Gaussian approximation will therefore be valid only in the limit of exceedingly small amplitude shifts, and quite possibly only in the core region of the pdf and not in the tails. The timing T and frequency Ω also deviate very slightly from Gaussian due to the local dependence of their variances on A (cf. (2.12); see also [13]). Since T, Ω, and Φ have no influence on the random walk of A (or on T, in the case of the phase), however, their statistical behavior is expected to be dominated by the mean value of A. Finally, even if the noise-induced changes of the soliton parameters were approximately Gaussian-distributed, calculating the probability densities in the tails from the (analytically or numerically obtained) variances would require an exponential extrapolation, and any errors or uncertainties would be magnified correspondingly.

3. Importance sampling. The main idea behind importance sampling is to bias the probability distribution functions used to generate the random Monte Carlo samples so that errors occur more frequently than would be the case otherwise [17, 18]. Before we delve into the implementation of importance sampling for amplifier noise, let us briefly present the basic ideas in a general setting.

Let X denote a collection of random variables (RVs) identifying a particular system realization. (In our case, X will be a vector or matrix which determines the noise at all the amplifiers.) Consider a measurable quantity $y(X)$ associated with each system configuration and, therefore, with each value of X. (In our case, $y(X)$ will be the final pulse amplitude or timing.) Suppose that we are interested in calculating

the probability P that $y(X)$ falls in some prescribed range. This probability can be represented as the expectation value of an *indicator function* $I(y(X))$ such that $I(y) = 1$ if the random variable y falls in the prescribed range and $I(y) = 0$ otherwise. That is, the probability P is represented by the multidimensional integral

$$P = \int I(y(x))p_X(x)\,dx = \mathbb{E}[I(y(X))], \tag{3.1}$$

where $p_X(x)$ is the joint pdf of the RVs X, $\mathbb{E}[\,\cdot\,]$ denotes the expectation value with respect to the distribution $p_X(x)$, and the integral is over the entire configuration space. In many cases of interest, a direct evaluation of the integral in (3.1) is impossible. One can then resort to Monte Carlo simulations and write an estimator $\hat{P}$ for P, replacing the integral in (3.1) by

$$\hat{P}_{\rm mc} = \frac{1}{M}\sum_{m=1}^{M} I(y(X_m)), \tag{3.2}$$

where M is the total number of Monte Carlo samples, and where each X_m is a random sample drawn from $p_X(x)$. Equation (3.2) simply expresses the relative number of samples falling in the range of interest. If one is interested in low probability events, however (that is, if $P \ll 1$), an impractically large number of samples is usually necessary in order to see even a single event, and an even larger number is required in order to obtain an accurate estimate.

When the main contribution to P comes from regions of sample space where $p_X(x)$ is small, IS can be used to speed up the simulations. The idea is first to rewrite the the probability P in (3.1) as

$$P = \int I(y(X))\,r(x)p^*(x)\,dx = \mathbb{E}^*[I(y(X))r(X)], \tag{3.3}$$

where $\mathbb{E}^*[\,\cdot\,]$ denotes the expectation value with respect to the *biasing distribution* $p^*(x)$, and where $r(x) = p_X(x)/p^*(x)$ is called the *likelihood ratio* [17]. As before, we then estimate the corresponding integral via Monte Carlo simulations; that is, we write an importance-sampled Monte Carlo estimate for P as

$$\hat{P}_{\rm is} = \frac{1}{M}\sum_{m=1}^{M} I(y(X_m^*))r(X_m^*), \tag{3.4}$$

where now the samples X_m^* are drawn according to the distribution $p^*(x)$. By design, the estimator $\hat{P}_{\rm is}$ is unbiased; i.e., $\mathbb{E}^*[\hat{P}_{\rm is}] = P$. If $p^*(x) \equiv p_X(x)$ (unbiased simulations), $r(x) = 1$ and (3.4) agrees with (3.2) (i.e., one recovers the standard Monte Carlo method). The use of a biasing pdf, however, allows the desired regions of sample space to be visited much more frequently. At the same time, the likelihood ratio automatically adjusts each contribution so that all of the different realizations add properly to give a correct final estimate.

The crucial step when trying to apply IS is to determine a proper choice of the biasing distribution $p^*(x)$ in order to reduce the variance of the estimator $\hat{P}_{\rm is}$. Naturally, not all biasing distributions are appropriate for accomplishing this. One might think that the simplest choice is to increase the overall noise variance, in an attempt to increase the probability of generating errors. It is well-known, however, that this biasing method (which is usually referred to as *variance scaling*) is effective

only in low-dimensional systems [17]. In general, in order for IS to be effective, $p^*(x)$ should concentrate the Monte Carlo samples near the regions that are *most likely* to generate rare events of interest, which in our case means determining the most likely noise instantiations at each amplifier which produce large pulse amplitude or timing variations at the fiber output. The proper choice of biasing distributions when one is interested in amplitude and timing jitter will be discussed in the next section.

If one seeks to reconstruct a broad region of the pdf for the quantity of interest, no single choice of biasing distribution can be expected to capture efficiently all the regions of sample space that give rise to the events of interest. In this case, several different biasing distributions $p_q^*(x)$ can be used and their results combined using a method known as *multiple importance sampling* [26, 27, 28]. With this technique, a weight $w_q(x)$ is associated with each biasing distribution. An importance-sampled estimator for P is then written as

$$\hat{P}_{\text{mis}} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{M_q}\sum_{m=1}^{M_q} w_q(X_{mq}^*)I(y(X_{mq}^*))r_q(X_{mq}^*), \tag{3.5a}$$

where Q is the total number of biasing distributions, M_q is the number of samples drawn from $p_q^*(x)$, X_{mq}^* is the mth such sample, and $r_q(x) = p_X(x)/p_q^*(x)$. Several strategies are possible for selecting the weights; the estimator $\hat{P}$ will be unbiased as long as $\sum_{q=1}^{Q} w_q(x) = 1$ for all x. A particularly simple and effective choice is the *balance heuristic* [26]:

$$w_q(x) = \frac{M_q p_q^*(x)}{\sum_{q'=1}^{Q} M_{q'} p_{q'}^*(x)}. \tag{3.5b}$$

Note that $M_q p_q^*(x)$ is proportional to the expected number of hits from the qth distribution. Thus, the weight of a sample with the balance heuristic is given by the likelihood of realizing that sample with the qth distribution relative to the total likelihood of realizing the same sample with all distributions.

4. IS for amplitude, frequency, timing, and phase jitter. We now turn our attention to the application of IS to Monte Carlo simulations of noise-induced amplitude, frequency, timing, and phase jitter. As explained earlier, in order for IS to be effective we need to bias the simulations towards the events that are most likely to produce the rare events of interest. Therefore, the strategy to bias the simulations toward predetermined target values of each soliton parameter consists of two logically distinct steps: First, we must determine how to bias the noise at each amplifier in order to obtain a specified change of amplitude, frequency, timing, and/or phase. Second, we must determine how to select the individual changes at each amplifier to obtain the desired total change at the end of the transmission line. To accomplish these goals, we need to (i) find the most likely noise configurations that result in a specified soliton parameter change at each amplifier, and (ii) find the most likely combination of individual amplitude, frequency, timing, and phase changes at all of the amplifiers that result in the desired value at the end of the transmission line. The fact that so much is known about NLS solitons greatly simplifies these tasks. The key information comes from the dependence of the soliton solution upon the parameters A, Ω, T_0, and Φ_0. The noise-induced changes in these soliton parameters can be calculated using SPT, as explained in section 2.1. In turn this knowledge can be used to bias the noise at each amplifier to induce larger-than-normal changes of the soliton parameters at the fiber output.

To make these ideas more definite, suppose that we are interested in large deviations of a quantity Y. Later on, this quantity will be identified with the amplitude A, frequency Ω, timing T, or phase Φ of the soliton. For now, let $Y_{\rm in}$ denote the value of Y at the fiber input, and suppose that we want to direct the simulations towards a target value $Y_{\rm out}$. As explained above, we need to (i) find the most likely noise realization at each amplifier to produce a given shift $C_n = Y_n - Y_{n-1}$, and (ii) find the most likely combination of individual contributions $\{C_n\}_{n=1}^{N_a}$ such that the final value of Y is $Y_{\rm out}$. We address these two issues separately.

To solve problem (i), at each amplifier we need to find the noise instantiation that maximizes the probability of realizing a prescribed shift in one of the soliton parameters. In any numerical implementation of (2.1), noise is added to the propagating signal by adding one independent Gaussian random variable to the real part of the optical field, and one to the imaginary part, at each discretized time point (when split-step spectral methods are used to solve (2.1), one can alternatively add an independent Gaussian random variable to the real part and to the imaginary part of every Fourier mode; this is equivalent to the above procedure in the time domain). Maximizing the probability of this Gaussian perturbation amounts to minimizing the sum of the squares of all of the random variables (one for the real part and one for the imaginary part at each time point), and in a continuous-time limit this corresponds to seeking a noise-produced perturbation $f(t)$ that minimizes the L^2-norm

$$\|f\|^2 = \int |f(t)|^2\, dt. \tag{4.1}$$

There is no weighting of the noise because every perturbation of comparable size is equally probable. Of course, we are interested in the noise perturbation that produces not just the most probable change, but rather the most probable change in one of the soliton parameters. This means that the minimization should be performed subject to the constraint

$$\Delta Y_n = \mathrm{Re} \int \underline{v}_Y^\dagger(z_n, t) f(t)\, dt = C_n, \tag{4.2}$$

where $\underline{v}_Y(z_n, t)$ is one of the adjoint linear modes evaluated at $z_n = nz_a$, consistent with (2.9), and C_n is for now an arbitrary constant. This constrained minimization problem can be expressed in Lagrange multiplier form by defining the functional

$$M_n = \int |f(t)|^2\, dt + \lambda \left[\int \underline{v}_Y^\dagger(z_n, t) f(t)\, dt + \int \underline{v}_Y(z_n, t) f^\dagger(t)\, dt - 2C_n \right]. \tag{4.3}$$

The solution to this problem, which is easily obtained using the calculus of variations, is

$$f(t) = C_n \, \frac{\underline{v}_Y(z_n, t)}{\|\underline{v}_Y(z_n, \cdot)\|^2}. \tag{4.4}$$

Here it should be noted that, even though a noise perturbation proportional to one of the linear modes produces a "clean" change in the soliton parameters (that is, a change without additional radiative components), (4.4) implies that the most likely way to realize the same parameter change occurs when the noise is proportional to the corresponding *adjoint* mode, a result which is not evident a priori.

Once the maximum likelihood noise configurations at each amplifier are known, it remains to solve problem (ii), namely to find the coefficients $\{C_1, \ldots, C_{N_a}\}$ that lead

with maximum probability to a prescribed change in the soliton parameter Y over N_a amplifiers. From (2.3), it is immediately apparent that (if one neglects the dependence of the variances in (2.12) on the amplitude) the soliton amplitude A and frequency Ω are not affected by changes to the other parameters, while the soliton timing T and phase Φ are affected both by direct perturbations to T and Φ, respectively, and by integrated changes to the frequency Ω and amplitude A, respectively. We therefore consider these problems separately.

4.1. Amplitude and frequency shifts. In the case of amplitude shifts, the problem is to find the most likely noise realization that produces a prescribed total change in the soliton amplitude, $\Delta A_{\rm tot} = A_{\rm out} - A_{\rm in}$. This amounts to another constrained minimization problem, where we now need to choose the set of individual amplitude shifts at each amplifier, $\{\Delta A_n\}_{n=1}^{N_a}$, in order to minimize the cumulative L^2-norm

$$\sum_{n=1}^{N_a} \|f_n\|^2 = \sum_{n=1}^{N_a} \frac{\Delta A_n^2}{\|\underline{v}_A(z_n,\cdot)\|^2} \tag{4.5}$$

(where (4.4) was used) under the constraint

$$\sum_{n=1}^{N_a} \Delta A_n = \Delta A_{\rm tot}. \tag{4.6}$$

Evaluating the norm of $\underline{v}_A$ using (2.9a), we can also rewrite this optimization problem in Lagrange multiplier form as

$$M = \sum_{n=1}^{N_a} \frac{\Delta A_n^2}{A_n} + \lambda \left[\Delta A_{\rm tot} - \sum_{n=1}^{N_a} \Delta A_n \right], \tag{4.7}$$

where, obviously, at each amplifier $A_n = A_{\rm in} + \sum_{n'=1}^{n} \Delta A_{n'}$. To find the minimum of M we then look for zeros of the gradient of M with respect to all the individual amplitude changes ΔA_n. If the total amplitude change over the N_a amplifiers is not too large, we can write $\Delta A_{\rm tot} = \epsilon\,\Delta$ and employ a perturbation expansion of ΔA_n and λ in powers of ϵ, namely, $\Delta A_n = \epsilon a_n^{(1)} + \epsilon^2 a_n^{(2)} + \cdots$ and of $\lambda = \epsilon\lambda_0 + \epsilon^2\lambda_2 + \cdots$. Minimizing and then matching orders of ϵ gives

$$a_n^{(1)} = \frac{\Delta}{N_a}, \tag{4.8a}$$

$$a_n^{(2)} = -\frac{1}{4}\frac{\Delta^2}{N_a^2}\,(N_a - 2n + 1), \tag{4.8b}$$

and so on. These constants determine the appropriate biasing for amplitude jitter at leading order and up to second order. Note that, at leading order, the desired average change in amplitude over the entire span of amplifiers is simply divided evenly among the individual amplifiers. In practice, this leading order approximation has been found to be adequate for all cases considered, even when the actual amplitude shifts computed were reasonably large.

The optimal biasing problem for the frequency Ω is similar but simpler, in that the L^2-norm of the associated linear mode is independent of Ω. In particular, we seek to minimize

$$\sum_{n=1}^{N_a} \|f_n\|^2 = \sum_{n=1}^{N_a} \frac{\Delta\Omega_n^2}{\|\underline{v}_\Omega(z_n,\cdot)\|^2} \tag{4.9}$$

under the constraint

$$\sum_{n=1}^{N_a} \Delta\Omega_n = \Delta\Omega_{\text{tot}}. \tag{4.10}$$

This leads to

$$M = \sum_{n=1}^{N_a} \frac{\Delta\Omega_n^2}{A_n} + \lambda \left[\Delta\Omega_{\text{tot}} - \sum_{n=1}^{N_a} \Delta\Omega_n \right], \tag{4.11}$$

where, due to the orthogonality of $\underline{v}_\Omega$ and $\underline{v}_A$, the amplitude remains unaffected by the biasing applied to Ω. For this reason, we assume $A_n = 1 \ \forall \ n = 1, \ldots, N_a$, which simply gives

$$\Delta\Omega_n = \frac{\Delta\Omega_{\text{tot}}}{N_a}. \tag{4.12}$$

Note that this assumption would need to be modified if one wished to compute the joint distribution of amplitude and frequency (or amplitude and timing), however.

4.2. Timing and phase shifts. We next look at the most likely noise realization resulting in a prescribed timing shift of the soliton at the fiber output. Because of the Galilean invariance of the NLS equation, in this case we need to consider frequency shifts as well as timing shifts. In other words, we seek to find the most likely set of frequency and timing shifts at each amplifier, $\{\Delta\Omega_n, \Delta T_n\}_{n=1}^{N_a}$, that produce a final value $T_{\text{out}} = T_{\text{in}} + \Delta T_{\text{tot}}$ of timing. Because of the orthogonality of $\underline{v}_T$ and $\underline{v}_\Omega$, this amounts to choosing ΔT_n and $\Delta\Omega_n$ in order to minimize the cumulative L^2-norm

$$\sum_{n=1}^{N_a} \frac{\Delta T_n^2}{\|\underline{v}_T(z_n, \cdot)\|^2} + \sum_{n=1}^{N_a} \frac{\Delta\Omega_n^2}{\|\underline{v}_\Omega(z_n, \cdot)\|^2} \tag{4.13}$$

under the constraint

$$\Delta T_{\text{tot}} = N_a z_a \Omega_{\text{in}} + \sum_{n=1}^{N_a} \Delta T_n + \sum_{n=1}^{N_a} (N_a + 1 - n) \Delta\Omega_n z_a \tag{4.14}$$

for a prescribed value of ΔT_{tot}. Again, we can evaluate the norms in (4.13) and rewrite the above problem in Lagrange multiplier form:

$$M = \frac{6}{\pi^2} \sum_{n=1}^{N_a} A_n^3 \Delta T_n^2 + \frac{3}{2} \sum_{n=1}^{N_a} \frac{\Delta\Omega_n^2}{A_n} + \lambda \left[\sum_{n=1}^{N_a} \Delta T_n + \sum_{n=1}^{N_a} (N + 1 - n) \, \Delta\Omega_n z_a - \Delta T_{\text{tot}} \right]. \tag{4.15}$$

If the noise has components only along $\underline{v}_T$ and $\underline{v}_\Omega$, the soliton amplitude again remains unaffected, so that minimizing the action M gives

$$\Delta T_n = \frac{\pi^2}{12\sigma_{T,\text{tot}}^2} \Delta T_{\text{tot}}, \tag{4.16a}$$

$$\Delta\Omega_n = \frac{(N_a + 1 - n) z_a}{3\sigma_{T,\text{tot}}^2} \Delta T_{\text{tot}}, \tag{4.16b}$$

where

$$\sigma_{T,\mathrm{tot}}^2 = N_a \left[\frac{\pi^2}{12} + \frac{z_a^2}{18} (N_a + 1)(2N_a + 1) \right]. \tag{4.16c}$$

Note that the rule for biasing the frequency given by (4.16b) is rather different from the rule given by (4.12). Whereas the former is designed to produce a given total change in *frequency* with highest likelihood, the latter is designed to produce a given total change in *timing* with highest likelihood, so that frequency shifts occurring earlier in the propagation are weighted much more heavily. In fact, comparing (2.10), (4.14), and (4.16), it can easily be seen that the relative weight of each term is proportional to the variance of its term in the final result. In other words, the most probable way of obtaining a given timing shift at the end of the fiber is to perform relatively larger frequency shifts at the beginning of the fiber, since these are the ones that can accumulate over the longest distances and therefore produce larger deviations for the same "effort" (i.e., for the same contribution to the cumulative L^2-norm of (4.1)).

Just as the optimal noise instantiation to obtain a given total timing shift depends on both frequency and timing shifts at each amplifier, the most probable way of obtaining a prescribed total phase shift requires shifting *three* parameters at each amplifier: phase, amplitude, and frequency. Under the conditions that $\Omega_0 = 0$ and that the individual amplitude shifts are kept small, however, the terms in (2.11d) involving Ω_0 and those involving products of shifts can be neglected, leaving as the action

$$\begin{aligned} M = &\sum_{n=1}^{N_a} \frac{A_n \Delta\Phi_n^2}{\pi^2/18 + 2/3} + \sum_{n=1}^{N_a} \frac{\Delta A_n^2}{2A_n} \\ &+ \lambda \left[\sum_{n=1}^{N_a} \Delta\Phi_n + \sum_{n=1}^{N_a} (N + 1 - n) A_0 \Delta A_n z_a - \left(\Phi_{\mathrm{out}} - \frac{1}{2} N_a z_a A_0^2 \right) \right]. \end{aligned} \tag{4.17}$$

This action has markedly similar form to (4.15) and demonstrates again that the effect of amplitude shifts on phase through self-phase modulation is completely analogous to the effect of frequency shifts on position through Galilean invariance. Here, however, the fact that amplitude appears in the summations is problematic, as amplitude is one of the parameters being shifted in our biasing scheme. To resolve this, we take an approach similar to that for direct amplitude shifting; i.e., we use a perturbation expansion in ΔA_{tot}. At leading order, the optimal phase and amplitude shifts take the same form as the above optimal timing and frequency shifts:

$$\Delta\Phi_n = \frac{\pi^2/36 + 1/3}{\sigma_{\Phi,\mathrm{tot}}^2} \left(\Delta\Phi_{\mathrm{tot}} - \frac{1}{2} N_a z_a \right), \tag{4.18a}$$

$$\Delta A_n = \frac{(N_a + 1 - n) z_a}{\sigma_{\Phi,\mathrm{tot}}^2} \left(\Delta\Phi_{\mathrm{tot}} - \frac{1}{2} N_a z_a \right), \tag{4.18b}$$

where

$$\sigma_{\Phi,\mathrm{tot}}^2 = N_a \left[\frac{\pi^2}{36} + \frac{1}{3} + \frac{z_a^2}{6} (N_a + 1)(2N_a + 1) \right]. \tag{4.18c}$$

4.3. Implementation issues. Having found the most probable configurations that produce given values of amplitude, frequency, timing, and phase shifts, we now

discuss how to use them to guide the biasing of the importance-sampled Monte Carlo simulations.

For concreteness, suppose that we are numerically solving a discretized version of the NLS equation (2.1). As explained earlier, at each amplifier we add random noise f_n. If J is the total number of discrete time points in the computational domain (or, equivalently, the total number of complex Fourier modes), the noise is represented by a vector $\mathbf{x}_n = (x_1, \ldots, x_{2J})^T$ giving the real and imaginary noise components at each discretized time location. In the unbiased case, the x_j are independent identically distributed (i.i.d.) normal RVs with mean zero and variance $\sigma_a^2 = \sigma^2/(2\Delta t)$; explicitly, the probability distribution is $p_{\mathbf{x}}(\mathbf{x}) = \exp[-\mathbf{x}^T\mathbf{x}/2\sigma_a^2]/(2\pi\sigma_a^2)^J$. Let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be the $2J \times N$ matrix that denotes all of the noise components at all of the amplifiers, and suppose that we are interested in reconstructing the pdf of a variable $y(X)$. (Here, y will be the amplitude A, the frequency Ω, the timing T, or the phase Φ.)

At each amplifier, we will perform the biasing by selecting a *deterministic* biasing vector $\mathbf{b}_n$ that will be added to the noise vector $\mathbf{x}_n$ drawn from the unbiased distribution. That is, we will form a biased noise realization as $\mathbf{x}_n^* = \mathbf{x}_n + \mathbf{b}_n$. This corresponds to choosing, at each amplifier, the biasing pdf $p_{\mathbf{x}}^*(\mathbf{x}_n^*) = p_{\mathbf{x}}(\mathbf{x}_n^* - \mathbf{b}_n) = p_{\mathbf{x}}(\mathbf{x}_n)$, which therefore gives a likelihood ratio of $r_{\mathbf{x}}(\mathbf{x}_n^*) = p_{\mathbf{x}}(\mathbf{x}_n + \mathbf{b}_n)/p_{\mathbf{x}}(\mathbf{x}_n)$. One can then obtain the overall likelihood ratio of the noise over N_a amplifiers[3] as

$$r(X^*) = \prod_{n=1}^{N_a} \frac{p_{\mathbf{x}}(\mathbf{x}_n^*)}{p_{\mathbf{x}}(\mathbf{x}_n)},$$

where $X^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_a}^*)$ and $\mathbf{x}_n^* = \mathbf{x}_n + \mathbf{b}_n$, as before. Of course, in order for this strategy to be effective, the choice of the biasing vectors $\mathbf{b}_n$ is crucial. The means by which we choose these biasing vectors in the case of amplitude, frequency, timing, and phase jitter is discussed next.

In order to perform the biasing, at each amplifier we first need to find the soliton parameters associated with the solution immediately before amplification (i.e., the addition of noise). We do this either by solving the Zakharov–Shabat eigenvalue problem [9, 29] or by using the moment integrals for the soliton parameters [10]. (A more detailed discussion of this soliton reconstruction process is contained in Appendices B and C.) The soliton parameters uniquely determine the soliton solution, which in turn determines the linear modes. Since the deterministic biasing term is expressed in the form of a linear combination of modes (as determined in the previous subsection), knowing the soliton parameters allows us to select the proper biasing of the Monte Carlo simulations as given by (4.4) and the determination of C_n in subsections 4.1 and 4.2.

In particular, if one wants to bias the amplitude, one chooses a shift in the mean of the noise $f(t)$ equal to

$$\Delta A_n \frac{\underline{v}_A(z_n, t)}{\|\underline{v}_A(z_n, \cdot)\|^2} = \Delta A_n \frac{\underline{v}_A(z_n, t)}{2A_n}, \tag{4.19}$$

[3]Note that the biased noise realizations at each amplifier are not statistically independent, since at each amplifier the choice of the biasing vector $\mathbf{b}_n$ depends on the current state of the soliton and therefore on the accumulated effect of the noise from the previous amplifiers. Nevertheless, it is easy to show that the overall likelihood ratio for such a Markov process can still be written as a product of the individual likelihood ratios (e.g., see [28]).

where ΔA_n is given by (4.8). Similarly, if one wants to bias the pulse frequency, one chooses a shift in the mean of the noise equal to

$$\Delta\Omega_n \frac{\underline{v}_\Omega(z_n,t)}{\|\underline{v}_\Omega(z_n,\cdot)\|^2} = \Delta\Omega_n \frac{3\,\underline{v}_\Omega(z_n,t)}{2A_n}, \tag{4.20}$$

where $\Delta\Omega_n$ is given by (4.12). To bias the soliton position, however, one must also bias the frequency, and in this case one chooses a shift in the mean of the noise equal to

$$\Delta T_n \frac{\underline{v}_T(z_n,t)}{\|\underline{v}_T(z_n,\cdot)\|^2} + \Delta\Omega_n \frac{\underline{v}_\Omega(z_n,t)}{\|\underline{v}_\Omega(z_n,\cdot)\|^2} = \Delta T_n \frac{6A_n^3\,\underline{v}_T(z_n,t)}{\pi^2} + \Delta\Omega_n \frac{3\,\underline{v}_\Omega(z_n,t)}{2A_n}, \tag{4.21}$$

where ΔT_n and $\Delta\Omega_n$ are now given by (4.16). Finally, to bias the phase one must also bias the amplitude, giving a mean noise shift of

$$\Delta\Phi_n \frac{\underline{v}_\Phi(z_n,t)}{\|\underline{v}_\Phi(z_n,\cdot)\|^2} + \Delta A_n \frac{\underline{v}_A(z_n,t)}{\|\underline{v}_A(z_n,\cdot)\|^2} = \Delta\Phi_n \frac{A_n\,\underline{v}_\Phi(z_n,t)}{\pi^2/18+2/3} + \Delta A_n \frac{\underline{v}_A(z_n,t)}{2A_n}. \tag{4.22}$$

In the discretized version of the problem, this biasing term, i.e., the shift of the mean of the noise $f(t)$, can also be represented as a vector, $\mathbf{b}_n$. Once the biasing direction and strength have been chosen, the actual biasing is straightforward: an unbiased noise realization $\mathbf{x}_n$ is generated, and the biased noise realization $\mathbf{x}_n^*$ is obtained by simply adding $\mathbf{b}_n$ to $\mathbf{x}_n$; that is, $\mathbf{x}_n^* = \mathbf{x}_n + \mathbf{b}_n$, as explained above.

5. Numerical results. To demonstrate the effectiveness of applying IS to Monte Carlo simulations of amplitude, frequency, timing, and phase jitter, we have performed simulations using the procedure described above. In dimensionless units, we took an amplifier spacing of $z_a = 0.1$, a total propagation distance of $N_a z_a = 20$, and a dimensionless noise strength of $\sigma^2 = 6.3 \times 10^{-5}$. The physical parameters generating these values are given in Appendix A. In the simulations, we extracted the soliton parameters at the intermediate amplifiers using moments (see Appendix C), but computed the values at the final distance using the more accurate Zakharov–Shabat eigenvalue problem (see Appendix B).

Figure 2 shows the results of 50,000 importance-sampled Monte Carlo simulations, selectively targeting amplitude fluctuations. Five biasing targets with 10,000 samples per target were used. Different choices of biasing generate the different regions of the pdfs shown in Figure 3, and the results from all Monte Carlo realizations are combined using the balance heuristic described in section 3. Even with a relatively small number of Monte Carlo samples, the method produces values of amplitude and timing jitter far down into the tails of the probability distributions. As shown in Figure 2, these results agree with unbiased Monte Carlo simulations in the main portion of the pdf (the only region that can be reconstructed with unbiased simulations). For smaller probability values, however, the deviations from Gaussian are evident.

A simple model of amplitude fluctuations can be obtained via soliton perturbation theory [10]: $A_{n+1} = A_n + \sqrt{A_n}\, s_{n+1}$, where the s_n are i.i.d. normal RVs with mean zero and variance σ^2. (Of course, this model cannot be correct when the noise is not a small perturbation of the soliton; an obvious erroneous result of this is that there is a very slight probability for negative A_n's to occur. Fortunately, it will be seen that such unphysical cases occur with extremely small probability, and therefore can be ignored.) Note that this model reflects a prepoint approximation of the jump conditions in (2.1). While this approach is closer in spirit to the Markov process created

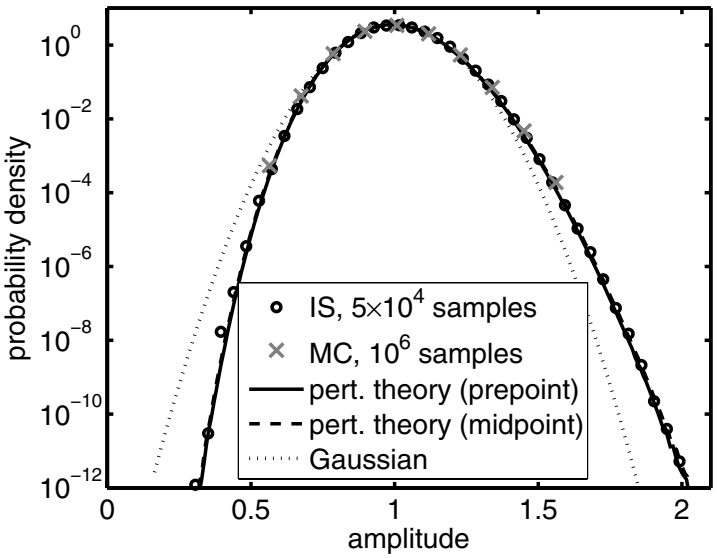

FIG. 2. *The pdf of amplitude jitter in a soliton-based transmission system, obtained from* 50,000 *importance-sampled Monte Carlo simulations. Results from a simple model from perturbation theory and an approximate Gaussian curve are also shown.*

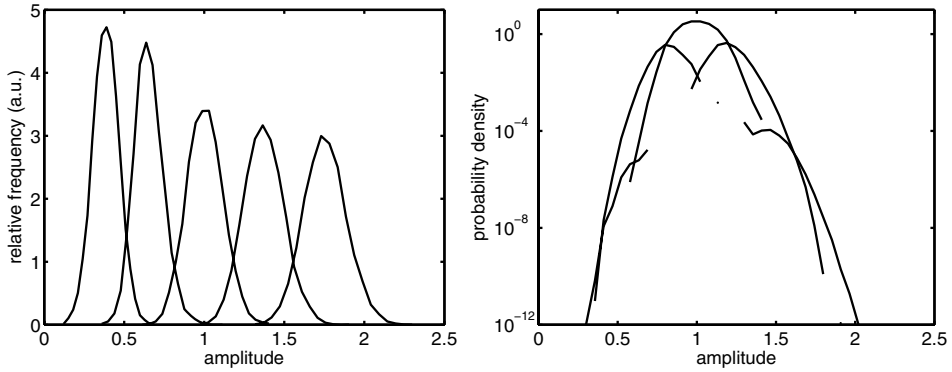

FIG. 3. *(left) Relative frequency plots showing the different ranges of amplitude generated by biasing distributions with five different targets. From left to right, the targets are* $\Delta A_{\rm tot} = \{-0.8, -0.4, 0, 0.4, 0.8\}$. *(right) The relative contribution of each biasing distribution to the final result of Figure* 2 *when weighted by the balance heuristic.*

by the biased Monte Carlo simulations, it is unclear whether the physical process is more accurately represented by this approximation or by a midpoint approximation, given by $A_{n+1} = A_n + (\sqrt{A_{n+1}} + \sqrt{A_n})\, s_{n+1}/2$. (For example, in one interpretation of a quantum-mechanical analysis of noise induced by loss and gain in a periodically amplified system, half of the noise is contributed in a distributed manner by the loss [30].) For comparison, in Figure 2 we show the pdfs obtained from both models, using importance-sampled numerical simulations for the prepoint approximation and simple analysis (see Appendix D) for the midpoint approximation. Although agreement with the importance-sampled simulations of the full NLS equation (2.1) is very

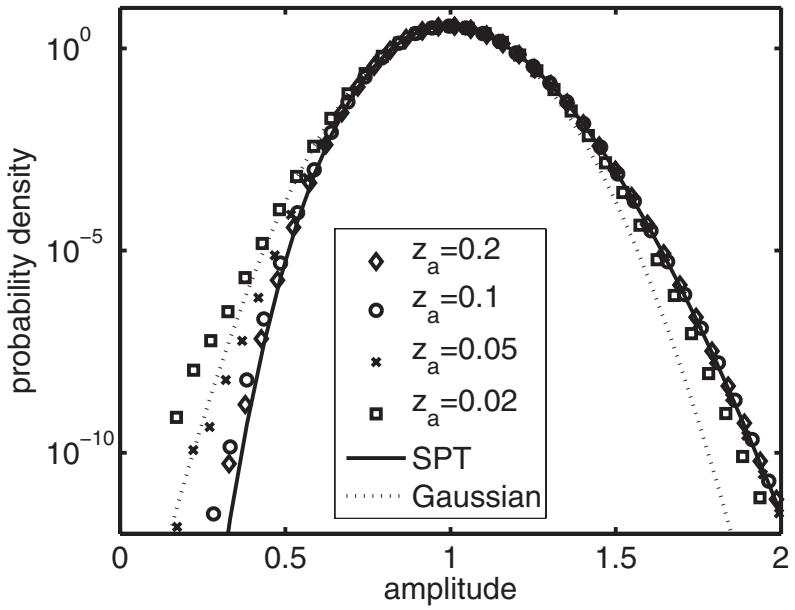

FIG. 4. *The pdf of amplitude jitter when the amplifier separation z_a is varied while the number of amplifiers N_a and the noise strength at each amplifier are held fixed. Even though SPT predicts the solid curve for each of these runs, the agreement with PDE numerics appears to be good only for sufficiently large z_a. Each of the numerical runs was obtained using 5×10^5 simulations.*

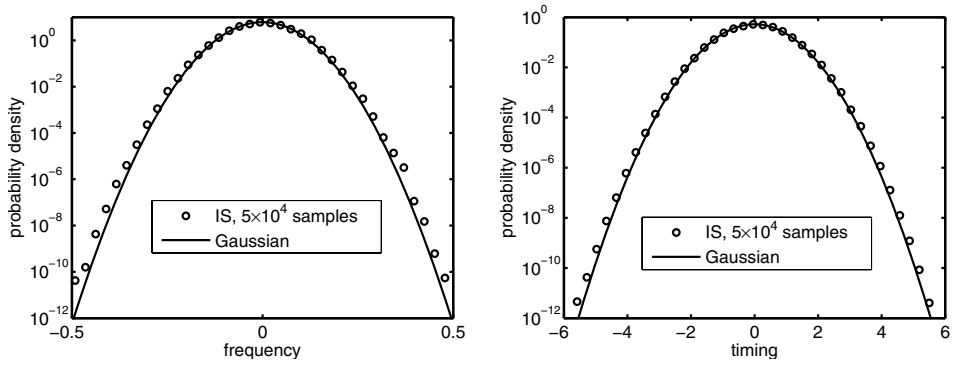

FIG. 5. *The pdfs of frequency (left) and timing jitter (right) in a soliton-based transmission system, each obtained from* 50,000 *importance-sampled Monte Carlo simulations.*

good throughout the range of amplitude values considered, with a slight deviation at small amplitudes, this agreement deteriorates significantly at both small and large amplitudes when the amplifier spacing z_a is decreased, as shown in Figure 4. As z_a is increased, the agreement appears to improve. It is not clear why the numerical results disagree with SPT here; nevertheless, the biasing obtained using SPT is sufficiently close to the correct biasing that the IS simulations accurately capture the pdf.

Figure 5 shows results similar to those of Figure 2, but for importance-sampled Monte Carlo simulations targeting frequency and timing fluctuations. The distributions of frequency and timing jitter agree well over a larger range of probability values with Gaussian curves whose variances are calculated from the theoretical results,

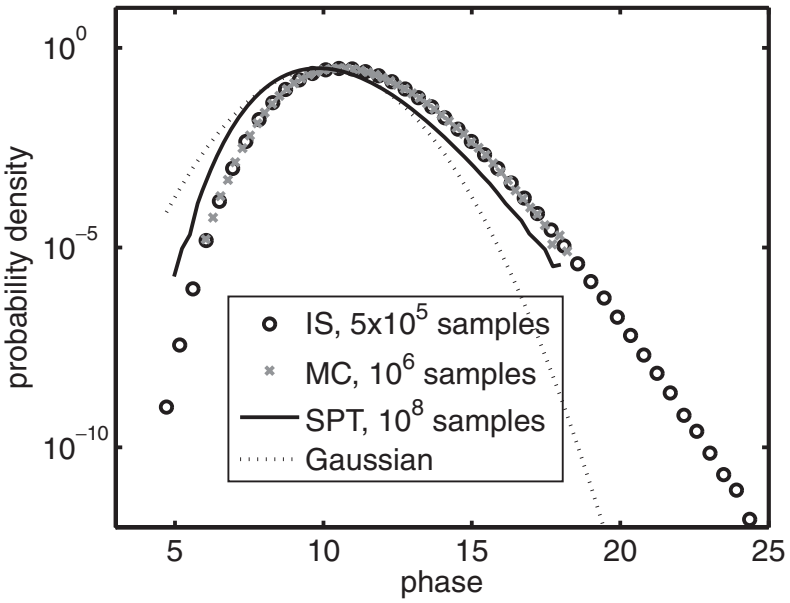

FIG. 6. *Comparison of the pdf of phase jitter with the Gaussian obtained from the linearized equations from SPT and with Monte Carlo simulations of the full PDE and of the nonlinear SPT equations. In this case, the last two results disagree significantly.*

(2.14b) and (2.14c). Note, however, that both for frequency and for timing jitter small deviations from Gaussian behavior are observed in the tails, where the numerically reconstructed pdfs are slightly but systematically larger than the predicted Gaussians. This discrepancy is due to the amplitude dependence of the variances of frequency and timing fluctuations [13]. Such dependence was neglected when the deterministic biasing choices were determined. Nevertheless, because random sampling is used, the numerical simulations access not only the deterministic biasing directions, but also nearby points in sample space around them. If errors in the deterministic biasing directions are not too large, a reasonably large number of random samples will find the correct regions in sample space that contribute most significantly to the pdf.

Finally, the pdf of phase jitter obtained using 5×10^5 IS trials is shown in Figure 6, along with 10^6 Monte Carlo runs to demonstrate the accuracy of the biased runs. We have unwrapped the phase to better illustrate deviations in the pdf tails. As expected, the pdf disagrees with the Gaussian obtained by linearizing the soliton perturbation equations. Somewhat surprising, however, is the fact that it also disagrees with 10^8 Monte Carlo simulations of the full nonlinear SPT equations. This suggests that dispersive radiation plays an important role in the case of phase jitter, rendering the (first-order) SPT equations ineffective in reproducing the correct jitter statistics. These equations are nevertheless still sufficiently accurate to provide effective biasing for the IS runs.

6. Conclusion. In summary, we have presented the application of importance sampling to numerical simulations of large noise-induced perturbations of nonlinear Schrödinger solitons, and we have demonstrated the method by calculating the pdfs of amplitude, frequency, timing, and phase jitter in a soliton-based transmission system. These results show that IS can be an effective tool for assessing the impact of noise in such systems.

Appendix A. NLS nondimensionalization and units. Here we describe the nondimensionalization procedure and the choice of units. The NLS equation is written in dimensional units as

$$i\frac{\partial E}{\partial Z} + \frac{|\beta''|}{2}\frac{\partial^2 E}{\partial T^2} + \gamma|E|^2E = i\sum_{n=1}^{N_a}\delta(Z - nZ_a)F_n(T), \tag{A.1}$$

where $|E|^2$ is optical power in watts, Z and T are dimensional distance in km and retarded time in ps, Z_a is the amplifier spacing, and β'' is the group velocity dispersion parameter in ps^2/km. The nonlinear coefficient is $\gamma = \omega_0 n_2/cA_{\text{eff}}$, where ω_0 is the carrier frequency, n_2 is the Kerr nonlinear-index coefficient, c is the vacuum speed of light, and A_{eff} is the effective area of the fiber core. The periodic cycle of loss and gain introduced by the chain of amplifiers has already been averaged out of (A.1); for details, see [30]. The delta-correlated white noise added at each amplifier then has noise strength

$$\langle F_m(T)F_n^\dagger(T')\rangle = \frac{\hbar\omega_0\eta_{\text{sp}}(G-1)^2}{G\ln G}\delta_{mn}\delta(T-T'), \tag{A.2}$$

where G is the power gain at each amplifier and η_{sp} is the spontaneous emission factor.

We then let $z = Z/L$, $t = T/T_0$, and $u = E/E_0$, where $L = T_0^2/|\beta''|$ is the dispersion length, $T_0 = T_{\text{FWHM}}/1.76$ is the soliton (sech) width, and $E_0 = 1/\sqrt{L\gamma}$ is the characteristic optical power for critical balance between nonlinearity and group velocity dispersion. This reduces (A.1) and (A.2) to (2.1) and (2.2), with

$$\sigma^2 = \frac{\hbar\omega_0\eta_{\text{sp}}\gamma T_0}{|\beta''|}\frac{(G-1)^2}{G\ln G}. \tag{A.3}$$

In the simulations we used a pulse full width at half maximum (FWHM) of 17.6 ps (i.e., a sech width of $T_0 = 10$ ps), an amplifier spacing of $Z_a = 50$ km and a fiber loss of 0.2 dB/km (yielding a power gain of $G = 10$), a spontaneous emission factor of η_{sp} of 1.4, a fiber dispersion $\beta'' = -0.2$ ps^2/km, and a total transmission distance of 10,000 km. The nonlinear coefficient of the fiber was taken to be 2.0 km^{-1}W^{-1}. The dimensionless parameters corresponding to these values are given in section 5.

Appendix B. Soliton extraction via Zakharov–Shabat eigenvalue problem. As discussed in the main text, the first step in implementing importance sampling is to find the soliton part of the solution at each amplifier. One way to do this is to solve the Zakharov–Shabat (Z-S) eigenvalue problem [31, 32]. Given a solution u of the NLS equation at a particular value of z, one can discretize the Z-S eigenvalue problem and solve it numerically [29, 33]. In the case of noisy solutions, which may not be smooth, it may be more robust to use a completely integrable discrete version, such as the Ablowitz–Ladik eigenvalue problem [29].

Unfortunately, an eigenvalue of the Z-S problem (or its discrete equivalent, the Ablowitz–Ladik problem) gives only two of the soliton parameters, the amplitude and the frequency, and there is apparently no way to determine the *exact* values of the soliton's position and phase. One can, however, obtain values that are relatively unaffected by noise, even when this perturbation is large.

To do this, one makes use of the trace formula for the NLS equation [34],

$$u = 2i\sum_{k=1}^{N}\frac{b_k}{a_k'}\psi_1^2 - 2i\sum_{k=1}^{N}\frac{b_k^\dagger}{a_k'^\dagger}\psi_2^{\dagger 2} - \frac{1}{\pi}\int_{-\infty}^{\infty}\left\{\frac{b}{a}(\xi)\psi_1^2(t,\xi) + \frac{b^\dagger}{a^\dagger}(\xi)\psi_2^{\dagger 2}(t,\xi)\right\}d\xi.$$

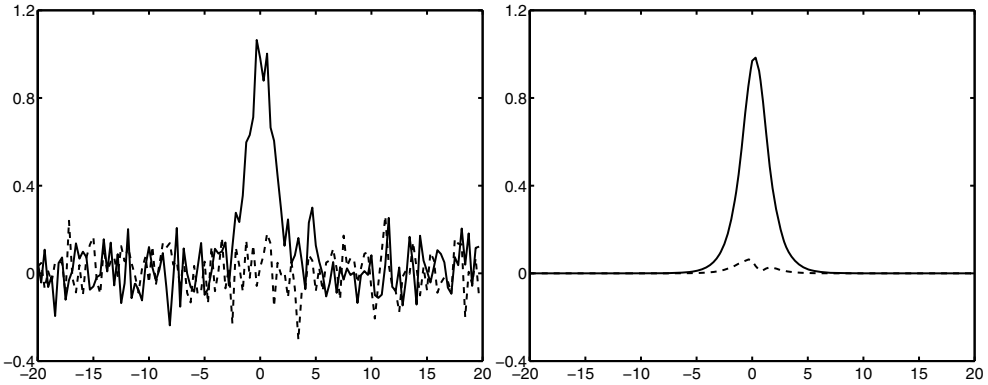

FIG. 7. *A noisy soliton (left) and the "clean" soliton (right) recovered from the Jost functions of the Z-S eigenproblem. Apart from a small ripple in the imaginary component, the Jost functions are seen to produce a reconstructed soliton that is largely free of radiation.*

Essentially, this shows that one can break the solution up into two contributions, one from the eigenfunctions of the Z-S scattering problem and the other from the continuous spectrum. Here, $\psi_1(x,\zeta)$ and $\psi_2(t,\zeta)$ are the components of one set of *Jost functions*, i.e., solutions of the Z-S scattering problem satisfying special boundary conditions, namely,

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta t} \quad \text{as } t \to +\infty \quad \text{or} \quad \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta t} \quad \text{as } t \to -\infty.$$

The coefficients b_k, a'_k, $b(\xi)$, and $a(\xi)$ are determined by the connection between these two sets of functions,

$$\begin{pmatrix} \phi_1(t,\zeta) \\ \phi_2(t,\zeta) \end{pmatrix} = a(\zeta) \begin{pmatrix} \psi_2^\dagger(t,\zeta^\dagger) \\ -\psi_1^\dagger(t,\zeta^\dagger) \end{pmatrix} + b(\zeta) \begin{pmatrix} \psi_1(t,\zeta) \\ \psi_2(t,\zeta) \end{pmatrix}.$$

At an eigenvalue ζ_k of the Z-S scattering problem (with $\operatorname{Im}\zeta_k > 0$), one has $a(\zeta_k) = 0$. At such an eigenvalue, $a'_k = a'(\zeta_k)$ and $b_k = b(\zeta_k)$.

In order to extract the position and phase from a noisy soliton in a robust way, the idea is to discard the contribution from the continuous spectrum and use only the discrete part of the trace formula. Because of the exponential decay of the eigenfunctions, this "reconstructed" or "nonlinearly filtered" solution will be smooth, and thus definitions of position and phase using moments will not have any difficulties caused by long noisy tails present in the solution. Of course, one must recognize that the trace formula does not precisely partition the solution into "soliton" and "dispersive radiation" components, as the discrete part of the trace formula does not produce solutions which are exactly hyperbolic-secant shaped. Nevertheless, the solutions appear very much like solitons, as shown in Figure 7.

As written, the trace formula is a little difficult to use, as one still needs the coefficients b_k and a'_k. Fortunately, the ratio of these coefficients can be computed more conveniently. From the orthogonality relation [32, equation A6.6e], one has

$$\int_{-\infty}^{\infty} (\psi_{2k}\phi_{1k} + \psi_{1k}\phi_{2k})\, dt = ia'_k \quad \Rightarrow \quad 2b_k \int_{-\infty}^{\infty} \psi_{1k}\psi_{2k}\, dt = ia'_k,$$

since $\phi_{1k} = b_k\psi_{1k}$ and $\phi_{2k} = b_k\psi_{2k}$. Thus, the discrete part of the trace formula becomes

$$u = -\sum_{k=1}^{N}\left(\frac{\psi_{1k}^2}{\int_{-\infty}^{\infty}\psi_{1k}\psi_{2k}\,dt} + \frac{{\psi_{2k}^{\dagger}}^2}{\int_{-\infty}^{\infty}\psi_{1k}^{\dagger}\psi_{2k}^{\dagger}\,dt}\right).$$

Because the numerator and denominator in this expression are both quadratic in ψ_k, this means that one does not need to normalize the Jost eigenfunctions when computing their contribution to the solution.

Appendix C. Soliton extraction via moments. The method of obtaining the soliton parameters described above is very effective but computationally intensive, requiring the numerical determination of selected eigenvalues and eigenfunctions of a large matrix. The resolution afforded by this method is critical for the final calculation of the soliton parameters in order to determine correct statistics; the formulation of the biasing vectors, however, does not require the same degree of precision. In particular, if the applied biasing vector differs from the optimal biasing vector by a small random amount (caused, for example, by sensitivity of the parameter measurement technique to the presence of radiation), this can be expected to produce only a small reduction in the efficiency of generating an accurate and unbiased estimate through IS. As long as the biased Monte Carlo simulations sample a large enough region around the deterministic biasing direction, the method will remain efficient. Another way to interpret this is that the Monte Carlo sampling corrects for slight inaccuracies in the determination of the optimal biasing direction.

We have therefore used the following filtered moments to generate approximate values for the soliton parameters at each amplifier. We first obtain an estimate for the soliton frequency

$$\Omega_{\text{est}} = \frac{\int \omega\,|\tilde{u}|^2\,d\omega}{\int |\tilde{u}|^2\,d\omega}, \tag{C.1}$$

where $\tilde{u}$ is the Fourier transform of u. We then use this to band-pass filter the soliton to reduce the noise,

$$\tilde{u}_{\text{filt}} = \tilde{u}\,e^{-(\omega-\Omega_{\text{est}})^2/2W_{\text{filt}}^2}. \tag{C.2}$$

This filtered image of the noisy soliton is used to obtain the final parameter estimates,

$$A = \frac{1}{2}\int |u_{\text{filt}}|^2\,dt, \qquad \Omega = \frac{\int \omega|\tilde{u}_{\text{filt}}|^2\,d\omega}{\int |\tilde{u}_{\text{filt}}|^2\,d\omega}, \tag{C.3}$$

$$T = \frac{\int t|u_{\text{filt}}|^2\,dt}{\int |u_{\text{filt}}|^2\,dt}, \qquad \Phi = \frac{\int \arctan(\operatorname{Im} u_{\text{filt}}/\operatorname{Re} u_{\text{filt}})|u_{\text{filt}}|^2\,dt}{\int |u_{\text{filt}}|^2\,dt}. \tag{C.4}$$

Appendix D. Pdf for the midpoint model. The model of the soliton amplitude's random walk obtained by applying first-order SPT and a particular midpoint approximation to (2.1) was given in section 5 as

$$A_{n+1} = A_n + \frac{1}{2}\left(\sqrt{A_{n+1}} + \sqrt{A_n}\right)s_{n+1}, \tag{D.1}$$

where the s_n are i.i.d. normal RVs with mean zero and variance σ^2. To obtain the pdf for this process, it is convenient to introduce $a_n = \sqrt{A_n}$ and to collect terms, noting

that one can then complete the square by adding $s_n^2/16$ to both sides of the resulting equation:

$$\left(a_{n+1} - \frac{1}{4}s_{n+1}\right)^2 = \left(a_n + \frac{1}{4}s_{n+1}\right)^2. \tag{D.2}$$

Taking the positive branch of this square root then gives the much simpler process $a_{n+1} = a_n + s_{n+1}/2$, which is seen immediately to result in a Gaussian distribution for a_n with mean a_0 and variance $\sigma_a^2 = n\sigma^2/4$. Finally, a simple transformation yields the pdf for A_n:

$$p(A_n) = \frac{1}{2}\frac{1}{\sqrt{2\pi A_n \sigma_a^2}} \exp\left(-\frac{(\sqrt{A_n} - \sqrt{A_0})^2}{2\sigma_a^2}\right). \tag{D.3}$$

REFERENCES

[1] E. Iannone, F. Matera, A. Mecozzi, and M. Settembre, *Nonlinear Optical Communication Networks*, Wiley, New York, 1998.

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

[3] D. Marcuse, *Calculation of bit-error probability for lightwave system with optical amplifiers and post-detection Gaussian noise*, J. Lightwave Tech., 9 (1991), pp. 505–513.

[4] J. P. Gordon and H. A. Haus, *Random walk of coherently amplified solitons in optical fiber transmission*, Opt. Lett., 11 (1986), pp. 665–667.

[5] J. P. Gordon and L. F. Mollenauer, *Phase noise in photonic communication systems using linear amplifiers*, Opt. Lett., 15 (1990), pp. 1351–1353.

[6] C. R. Menyuk, *Non-Gaussian corrections to the Gordon-Haus distribution resulting from soliton interactions*, Opt. Lett., 20 (1995), pp. 270–272.

[7] T. Georges, *Study of the non-Gaussian timing jitter statistics induced by soliton interaction and filtering*, Opt. Commun., 123 (1996), pp. 617–623.

[8] R. Holzlöhner, V. S. Grigoryan, C. R. Menyuk, and W. L. Kath, *Accurate calculation of eye diagrams and bit error rates in optical transmission systems using linearization*, J. Lightwave Tech., 20 (2002), pp. 389–400.

[9] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Oxford University Press, Oxford, UK, 1995.

[10] E. Iannone, F. Matera, A. Mecozzi, and M. Settembre, *Nonlinear Optical Communication Networks*, Wiley, New York, 1998.

[11] R. Holzlöhner and C. R. Menyuk, *The use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communication systems*, Opt. Lett., 23 (2003), pp. 1894–1896.

[12] G. E. Falkovich, I. Kolokolov, V. Lebedev, and S. K. Turitsyn, *Statistics of soliton-bearing systems with additive noise*, Phys. Rev. E (3), 63 (2001), 025601.

[13] K.-P. Ho, *Non-Gaussian statistics of the soliton timing jitter due to amplifier noise*, Opt. Lett., 28 (2003), pp. 2165–2167.

[14] G. Biondini, W. L. Kath, and C. R. Menyuk, *Importance sampling for polarization-mode dispersion*, Photon. Technol. Lett., 14 (2002), pp. 310–312.

[15] S. L. Fogal, G. Biondini, and W. L. Kath, *Multiple importance sampling for first- and second-order polarization-mode dispersion*, Photon. Technol. Lett., 14 (2002), pp. 1273–1275; *Correction*, 14 (2002), p. 1487.

[16] R. O. Moore, G. Biondini, and W. L. Kath, *Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems*, Opt. Lett., 28 (2003), pp. 105–107.

[17] P. J. Smith, M. Shafi, and H. Gao, *Quick simulation: A review of importance sampling techniques in communications systems*, IEEE J. Select. Areas Commun., 15 (1997), pp. 597–613.

[18] R. Srinivasan, *Importance Sampling: Applications in Communications and Detection*, Springer-Verlag, New York, 2002.

[19] C. J. McKinstrie and T. I. Lakoba, *Probability-density function for energy perturbations of isolated optical pulses*, Optics Express, 11 (2003), pp. 3628–3648.

[20] D. J. Kaup, *Closure of the squared Zakharov-Shabat eigenstates*, J. Math. Anal. Appl., 54 (1976), pp. 849–864.

[21] D. J. Kaup, *A perturbation expansion for the Zakharov–Shabat inverse scattering transform*, SIAM J. Appl. Math., 31 (1976), pp. 121–133.
[22] Y. S. Kivshar and B. A. Malomed, *Dynamics of solitons in nearly integrable systems*, Rev. Modern Phys., 61 (1989), pp. 763–915.
[23] W. L. Kath, *A modified conservation law for the phase of the nonlinear Schrödinger equation*, Methods Appl. Anal., 4 (1997), pp. 141–155.
[24] J. Yang, *Complete eigenfunctions of linearized integrable equations expanded around a soliton solution*, J. Math. Phys., 41 (2000), pp. 6614–6638.
[25] E. T. Spiller, W. L. Kath, R. O. Moore, and C. J. McKinstrie, *Computing large signal distortions and bit-error ratios in DPSK transmission systems*, Photon. Technol. Lett., 17 (2005), pp. 1022–1024.
[26] E. Veach, *Robust Monte Carlo Methods for Light Transport Simulation*, Ph.D. thesis, Department of Computer Science, Stanford University, Palo Alto, CA, 1997.
[27] A. Owen and Y. Zhou, *Safe and effective importance sampling*, J. Amer. Statist. Assoc., 95 (2000), pp. 135–143.
[28] G. Biondini, W. L. Kath, and C. R. Menyuk, *Importance sampling for polarization mode dispersion: Techniques and applications*, J. Lightwave Technol., 22 (2004), pp. 1210–1215.
[29] J. A. C. Weideman and B. M. Herbst, *Finite difference methods for an AKNS eigenproblem*, Math. Comput. Simulation, 43 (1997), pp. 77–88.
[30] W. L. Kath, A. Mecozzi, P. Kumar, and C. G. Goedde, *Long-term storage of a soliton bit stream using phase-sensitive amplification: Effects of soliton-soliton interactions and quantum noise*, Opt. Commun., 157 (1998), pp. 310–326.
[31] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Soviet Physics JETP, 34 (1972), pp. 62–69.
[32] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *The inverse scattering transform-Fourier analysis for nonlinear problems*, Studies in Appl. Math., 53 (1974), pp. 249–315.
[33] S. Burtsev, R. Camassa, and I. Timofeyev, *Numerical algorithms for the direct spectral transform with applications to nonlinear Schrödinger type systems*, J. Comput. Phys., 147 (1998), pp. 166–186.
[34] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, SIAM Stud. Appl. Math. 4, SIAM, Philadelphia, 1981.