

# **A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons**

**Richard O. Moore<sup>(1)</sup>, Gino Biondini<sup>(2)</sup>, William L.**

**Kath<sup>(3)</sup>**

**<sup>(1)</sup> Department of Mathematical Sciences**

**New Jersey Institute of Technology**

**323 Martin Luther King, Jr. Blvd., Newark, NJ 07102**

**<sup>(2)</sup> Department of Mathematics**

**State University of New York at Buffalo**

**Buffalo, NY 14260**

**<sup>(2)</sup> Department of Engineering Sciences and Applied Mathematics**

**Northwestern University**

**Evanston, IL 60208**

**CAMS Report 0506-35, Spring 2006**

**Center for Applied Mathematics and Statistics**

**NJIT**

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

R. O. MOORE<sup>1</sup>, G. BIONDINI<sup>2</sup> AND W. L. KATH<sup>3</sup>

**Abstract.** We show in detail how to apply 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.

**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 various 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 and/or timing jitter, which can then lead to bit errors [3, 4].

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, 10Gb/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.

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 [5-7]. Nonlinearity arises from several sources, such as self-phase modulation due to the fiber's nonlinear refractive index [8, 9] as well as the nonlinear conversion of optical energy into electrical energy by the photodetector [7].

Various techniques have recently been proposed to address this problem [7, 10-15]. One promising approach is a technique known as importance sampling (IS) [16, 17]. In general, importance sampling works by concentrating Monte Carlo samples on those configurations that are most likely to lead to transmission errors, thus significantly speeding up the simu-

---

<sup>1</sup>Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102 (rmoores@njit.edu)

<sup>2</sup>Department of Mathematics, State University of New York, Buffalo, NY 14260 (biondini@buffalo.edu)

<sup>3</sup>Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208-3125 (kath@northwestern.edu)

lations. Previously, we have successfully applied this technique to the direct simulation of transmission impairments caused by polarization-mode dispersion [13, 14]. More recently, we have presented numerical results demonstrating that importance sampling can also be applied to simulations of ASE-induced transmission impairments [15]. 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 simple 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 [8, 9, 15], which in dimensionless units is

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

Here  $z$  and  $t$  are distance and retarded time,  $u$  is the pulse's electromagnetic field envelope,  $N_a$  is the number of amplifiers,  $z_a$  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 [8].

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  $n = 1, 2, \dots, N_a$ , 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 Eq. (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:

$$\langle f_n(t) \rangle = 0, \quad (2.2a)$$

$$\langle f_m(t) f_n(t') \rangle = 0, \quad (2.2b)$$

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

where  $\langle \cdot \rangle$  denotes ensemble average,  $*$  denotes the complex conjugate,  $\delta_{mn}$  and  $\delta(t - t')$  are the Kronecker and Dirac deltas, respectively, and  $\sigma^2$  is a parameter describing the noise power. Technically speaking, Eq. (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 [18]. 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, Eq. (2.1) admits the well-known soliton solution

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

with

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

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  $A$ ,  $\Omega$ ,  $T_0$  and  $\Phi_0$  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 [8, 9]. 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  $A$ ,  $\Omega$ ,  $T$  and  $\Phi$  [4]. 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 is a standard method that has been used to approximate the effect of noise upon propagating pulses (e.g., see Refs. [8, 19-21]). Rather than using it directly to obtain an analytical approximation to the perturbed pulse, however, here it will only be used as a tool to guide numerical simulations. The key information that is needed to do this comes from the dependence of the soliton solution Eq. (2.3) upon the parameters  $A$ ,  $\Omega$ ,  $T_0$  and  $\Phi_0$ . 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  $\Omega$ ).

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

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

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

$$\frac{\partial v}{\partial z} = Lv, \quad Lv := \frac{i}{2} \frac{\partial^2 v}{\partial t^2} - \frac{i}{2} A^2 v + 2i|u_0|^2 v + iu_0^2 v^*. \quad (2.5)$$

Importantly, the linearized NLS operator  $L$  is non-self-adjoint and non-normal, and its generalized nullspace admits four modes (localized eigenfunctions)  $v_k(z, t)$  ( $k = A, \Omega, T, \Phi$ ) satisfying [8, 9, 22]

$$Lv_A = Av_\Phi, \quad Lv_\Omega = v_T, \quad Lv_T = 0, \quad Lv_\Phi = 0. \quad (2.6)$$

Explicitly, these four modes are:

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

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

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

$$v_\Phi(z, t) = iu_0. \quad (2.7d)$$

Each of these linear modes is associated with a continuous symmetry of the NLS equation [21]: invariance with respect to scaling, Galilean boosts, translations and phase rotations, respectively. Note also from Eq. (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 [21].

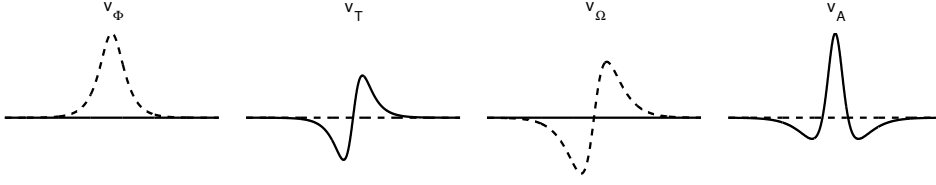


FIG. 1. The real (solid) and imaginary (dashed) parts of the four neutral modes of the linearized NLS equation associated with the soliton solution Eq. (2.3).

These four modes, shown in Figure 1, are also associated with changes of the soliton parameters<sup>1</sup>:

$$\partial u_s / \partial A = v_A e^{i\Theta}, \quad (2.8a)$$

$$\partial u_s / \partial \Omega = v_\Omega e^{i\Theta} + T v_\Phi e^{i\Theta} \quad (2.8b)$$

$$\partial u_s / \partial T = v_T e^{i\Theta}, \quad (2.8c)$$

$$\partial u_s / \partial \Phi = v_\Phi e^{i\Theta}. \quad (2.8d)$$

In fact, removing secular terms from the forced linearized NLS equation obtained by adding the right-hand side of Eq. (2.1) to Eq. (2.5), one finds the local changes of the soliton parameters at the  $n^{\text{th}}$  amplifier produced by the noise  $f_n(t)$  [8, 9]:

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

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

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

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

where the integrals are from  $-\infty$  to  $\infty$ . The functions  $\underline{v}_k(z, t)$  are the modes of the *adjoint* linearized NLS operator, defined by  $L^\dagger \underline{v} = -\frac{i}{2} \partial_t^2 \underline{v} + \frac{i}{2} A^2 \underline{v} - 2i|u_0|^2 \underline{v} + iu_0^2 \underline{v}^*$  and the inner product  $\langle \underline{v}, v \rangle = \text{Re} \int \underline{v}^* v dt$  [9]. They are

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

(The adjoint modes are easily obtained from those in Eqs. (2.7) noting that  $L^\dagger(\underline{v}) = iL(i\underline{v})$ .) Together, the modes of  $L$  and  $L^\dagger$  form a bi-orthonormal basis of the tangent space corresponding to infinitesimal changes in the soliton parameters at a given amplifier,<sup>2</sup> and the source terms in Eqs. (2.9) represent the projection of the noise onto this basis.

**2.2. Noise-induced amplitude, frequency and timing jitter.** Equations (2.9) establish a direct correspondence between the infinite-dimensional noise which is added at each amplifier and 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

<sup>1</sup>Note that the derivatives on the left-hand side of Eqs. (2.8) are taken with respect the variables  $A$ ,  $\Omega$ ,  $T$  and  $\Phi$  with the other three variables kept constant. A different choice of parametrization for the soliton solution in Eq. (2.3) would lead to different results.

<sup>2</sup>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.

the final values of amplitude  $A$ , frequency  $\Omega$  and timing  $T$ :

$$A_{\text{out}} = A_0 + \sum_{n=1}^{N_a} \Delta A_n, \quad (2.11a)$$

$$\Omega_{\text{out}} = \Omega_0 + \sum_{n=1}^{N_a} \Delta \Omega_n, \quad (2.11b)$$

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

(Since the phase of the pulse is discarded in standard on-off keying receivers, phase fluctuations will be neglected from here on. Of course these fluctuations would be important in phase-sensitive transmission formats such as phase-shift keying and differential phase-shift keying [23].) The fourth term in Eq. (2.11c) arises from the above-mentioned Galilean invariance of the NLS equation. Owing to Eqs. (2.2) and (2.9),  $\Delta A_n$ ,  $\Delta T_n$  and  $\Delta \Omega_n$  are Gaussian-distributed random variables at each amplifier, with variances

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

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

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

respectively. Note that all of these variances depend on the value of the soliton amplitude up to that point. For small deviations of this amplitude, one can approximate these variances with their initial values:

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

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

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

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

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

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

These calculations, however, are not sufficient to give an accurate estimate of noise-induced transmission penalties, for several reasons. First of all, the variances in Eq. (2.14) are only correct for small deviations of the pulse amplitude. Secondly, even though the noise is Gaussian-distributed, the 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 only be valid 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  $\Omega$  also deviate very slightly from Gaussian due to the local dependence of their variances on  $A$  (cf. Eqs. (2.12); see also Ref. [12]). Since  $T$  and  $\Omega$  have no influence on the random walk of  $A$ , 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 [16, 17]. 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 multi-dimensional integral

$$P = \int I(y(x))p_X(x)dx = E[I(y(X))], \quad (3.1)$$

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

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

where  $M$  is the total number of Monte Carlo samples, and 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, importance sampling can be used to speed up the simulations. The idea is first to rewrite the the probability  $P$  in Eq. (3.1) as

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

where  $E^*[\cdot]$  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* [16]. 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

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

where now the samples  $X_m^*$  are drawn according to  $p^*(x)$ . By design, the estimator  $\hat{P}_{\text{is}}$  is unbiased; i.e.,  $E^*[\hat{P}_{\text{is}}] = P$ . If  $p^*(x) \equiv p_X(x)$  (unbiased simulations),  $r(x) = 1$  and Eq. (3.4) agrees with Eq. (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 importance sampling is to determine a proper choice of the biasing distribution  $p^*(x)$  in order to reduce the variance of the estimator  $\hat{P}_{\text{is}}$ . 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 only effective in low-dimensional systems [16]. In general, in order for importance sampling 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* [24–26]. With this technique, a weight  $w_q(x)$  is associated with each biasing distribution. An importance-sampled estimator for  $P$  is then written as

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

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  $m$ -th 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* [24]:

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

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

**4. Importance sampling for amplitude, frequency and timing jitter.** We now turn our attention to the application of importance sampling to Monte Carlo simulations of noise-induced amplitude, frequency and timing jitter. As explained earlier, in order for importance sampling 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 and/or timing. 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 and timing 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  $A$ ,  $\Omega$ ,



$T_0$  and  $\Phi_0$ . The noise-induced changes in these soliton parameters can be calculated using soliton perturbation theory, 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  $\Omega$  or timing  $T$  of the soliton. For now, let  $Y_{\text{in}}$  denote the value of  $Y$  at the fiber input, and suppose that we want to direct the simulations towards a target value  $Y_{\text{out}}$ . 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  $Y_{\text{out}}$ . 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. Since the noise at each amplifier is Gaussian-distributed, in the continuous-time limit of discretized noise this corresponds to seeking a function  $f(t)$  that minimizes the  $L^2$ -norm

$$\|f\|^2 = \int |f(t)|^2 dt \quad (4.1)$$

subject to the constraint

$$\Delta Y_n = \text{Re} \int \underline{v}_Y^*(z_n, t) f(t) dt = C_n, \quad (4.2)$$

where  $\underline{v}_Y(z_n, t)$  is one of the adjoint linear modes evaluated at  $z_n = nz_a$ , consistent with Eq. (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

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

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

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

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), Eq. (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  $\{C_1, \dots, C_{N_a}\}$  that lead with maximum probability to a prescribed change in the soliton parameter  $Y$  over  $N_a$  amplifiers. From Eq. (2.3), it is immediately apparent that (if one neglects the dependence of the variances in Eq. (2.12) on the amplitude) the soliton amplitude  $A$  and frequency  $\Omega$  are not affected by changes to the other parameters, while the soliton timing  $T$  is affected by both direct perturbations to  $T$  and by integrated changes to its frequency  $\Omega$ . We therefore consider these problems separately.

**4.1. Amplitude and frequency shifts.** In this case the problem is to find the most likely noise realization that produces a prescribed total change in the soliton amplitude,  $\Delta A_{\text{tot}} = A_{\text{out}} - A_{\text{in}}$ . 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

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

(where Eq. (4.4) was used) under the constraint

$$\sum_{n=1}^{N_a} \Delta A_n = \Delta A_{\text{tot}}. \quad (4.6)$$

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

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

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

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

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

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  $\Omega$  is similar but simpler, in that the  $L^2$ -norm of the associated linear mode is independent of  $\Omega$ . In particular, we seek to minimize

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

under the constraint

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

This leads to

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

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

$$\Delta \Omega_n = \frac{\Delta \Omega_{\text{tot}}}{N_a}. \quad (4.12)$$

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

**4.2. Timing 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  $\Delta T_n$  and  $\Delta\Omega_n$  in order to minimize the cumulative  $L^2$ -norm

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

under the constraint

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

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

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

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

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

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

where

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

Note that the rule for biasing the frequency given by Eq. (4.16b) is rather different than the rule given by Eq. (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 Eqs. (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 Eq. (4.1)).

**4.3. Implementation issues.** Having found the most probable configurations that produce given values of amplitude, frequency and timing shifts, we now discuss how to use them to guide the biasing of the importance-sampled Monte Carlo simulations.

For concreteness, suppose 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  $\mathbf{x}_n = (x_1, \dots, x_{2J})^T$  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 random variables (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  $X = (\mathbf{x}_1, \dots, \mathbf{x}_N)$  be the  $2J \times N$  matrix that denotes all of the noise components at all of the amplifiers, and suppose we are interested in reconstructing the pdf of a variable  $y(X)$ . (Here,  $y$  will be the amplitude  $A$ , the frequency  $\Omega$  or the timing  $T$ .) Note that the noise bandwidth associated with the numerical simulation is  $J\pi/T$ .

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<sup>3</sup> as:

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

where  $X^* = (\mathbf{x}_1, \dots, \mathbf{x}_{N_a})$  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 and timing 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 [8, 27], or by using the moment integrals for the soliton parameters [9]. (A more detailed discussion of this soliton reconstruction process is contained in Appendix B and Appendix C.) The soliton parameters uniquely determine the soliton solution, which in turn determines the linear modes. Therefore, 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. In the discretized version of the problem, this biasing term 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 importance sampling to Monte Carlo simulations of amplitude, frequency and timing 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  $\sigma^2 = 6.9 \times 10^{-5}$ . The physical parameters generating these values are given in Appendix A.

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 Fig. 3, and the results from all Monte Carlo realizations are combined using the balance

<sup>3</sup>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 Ref. [26]).

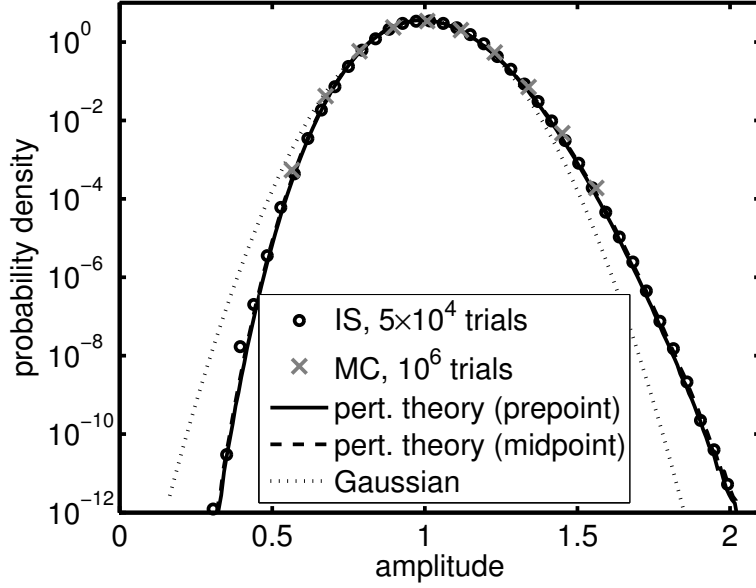


FIG. 2. The pdf of amplitude jitter in a soliton-based transmission system, obtained from 50,000 importance-sampled Monte Carlo simulations. Results from a simple model from perturbation theory and an approximate Gaussian curve are also shown.

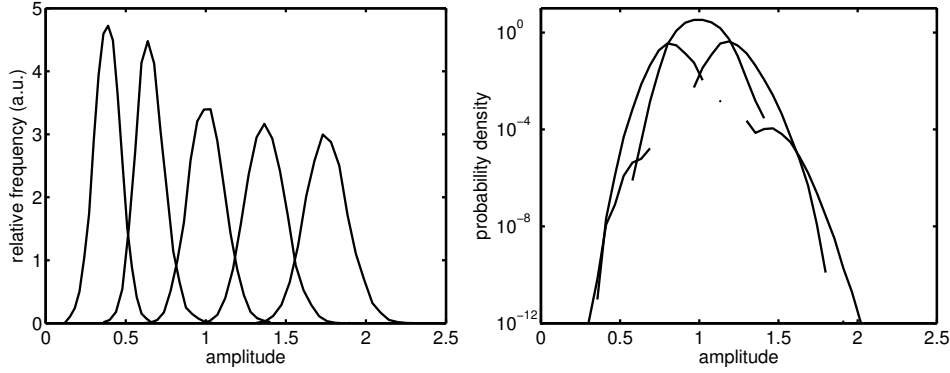


FIG. 3. (Left) Relative frequency plots showing the different ranges of amplitude generated by biasing distributions with five different targets. From left to right, the targets are  $\Delta A_{\text{tot}} = \{-0.8, -0.4, 0, 0.4, 0.8\}$ . (Right) The relative contribution of each biasing distribution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

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 Fig. 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 [9]:  $A_{n+1} = A_n + \sqrt{A_n} s_{n+1}$ , where the  $s_n$  are i.i.d. normal RVs with mean zero and

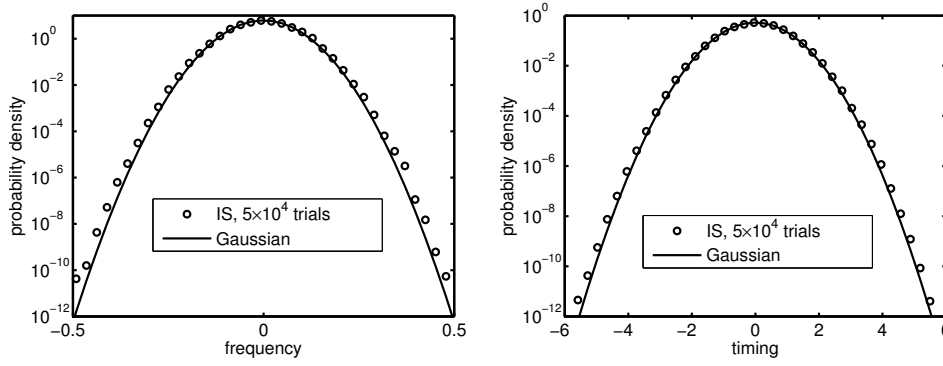


FIG. 4. The pdfs of frequency (left) and timing jitter (right) in a soliton-based transmission system, each obtained from 50,000 importance-sampled Monte Carlo simulations.

variance  $\sigma^2$ . (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 Eq. (2.1). While this approach is closer in spirit to the Markov process created 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 [28].) For comparison, in Fig. 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. The agreement with the importance-sampled simulations of the full NLS Eq. (2.1) is excellent throughout the range of amplitude values considered.

Figure 4 shows similar results as Fig. 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, Eqs. (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 [12]. 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.

**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 and timing 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.

**Acknowledgements.** This work was supported in part by the National Science Foundation (DMS-0101476 and DMS-0406513), and by the Air Force Office of Scientific Research (FA9550-04-1-0289).

### Appendix A. NLS nondimensionalization and units.

Here we describe the nondimensionalization procedure and the choice of units. The nonlinear Schrödinger (NLS) equation is written in dimensional units as

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

where  $|E|^2$  is optical power in watts,  $Z$  and  $T$  are dimensional distance and retarded time,  $Z_a$  is the amplifier spacing, and  $\beta''$  is the group velocity dispersion parameter in  $\text{ps}^2/\text{km}$ . The nonlinear coefficient is  $\gamma = \omega_0 n_2 / c A_{\text{eff}}$ , where  $\omega_0$  is the carrier frequency,  $n_2$  is the Kerr nonlinear-index coefficient,  $c$  is the vacuum speed of light and  $A_{\text{eff}}$  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 Eq. A.1; for details, see Ref. [28]. The delta-correlated white noise added at each amplifier then has noise strength

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

where  $G$  is the power gain at each amplifier and  $\eta_{\text{sp}}$  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 Eqs. A.1 and A.2 to Eqs. 2.1 and 2.2, with

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

In the simulations we used a pulse FWHM of 17.6 ps (i.e., a sech width of  $T_0=10$  ps), an amplifier spacing of  $Z_a = 50$  km, a fiber loss of 0.2 dB/km, yielding a power gain of  $G = 10$ , a spontaneous emission factor of  $\eta_{\text{sp}}$  of 1.5, a fiber dispersion  $\beta'' = -0.2 \text{ ps}^2/\text{km}$ , and a total transmission distance of 10,000 km. The nonlinear coefficient of the fiber was taken to be  $2.0 \text{ km}^{-1}\text{W}^{-1}$ . 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 [29, 30]

$$\begin{aligned} \frac{dv_1}{dt} &= -i\zeta v_1 + uv_2 \\ \frac{dv_2}{dt} &= i\zeta v_2 - u^*v_1. \end{aligned}$$

Here, the solution of the NLS equation,  $u$ , is the potential in the  $t$ -dependent spectral problem (frozen at a specific distance  $z$ )<sup>4</sup>. Each eigenvalue is associated with a soliton that is present

---

<sup>4</sup>Note the space and time variables appear reversed here in comparison with standard discussions of the Z-S scattering problem; this version is appropriate for the optical fiber context, of course.

in the solution  $u$ , and the associated eigenvalue  $\zeta$  gives two of the soliton parameters: the imaginary part is the amplitude, while the real part is the frequency.

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 [27, 31]. 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 [27].

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 [32]

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

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 *Jost functions*, i.e., the two components of the solution of the Z-S scattering problem satisfying the boundary conditions

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta t} \quad \text{as } t \rightarrow +\infty.$$

To define the coefficients  $b_k$ ,  $a'_k$ ,  $b(\xi)$  and  $a(\xi)$ , one needs to consider the two components of an additional solution,  $\phi_1(t, \zeta)$  and  $\phi_2(t, \zeta)$ , which satisfy

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta t} \quad \text{as } t \rightarrow -\infty.$$

This solution is connected to the previous one through the relations

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

At an eigenvalue  $\zeta_k$  of the Z-S scattering problem (with  $\text{Im } \zeta_k > 0$ ), one has  $a(\zeta_k) = 0$  and the two solutions become proportional to one another with proportionality constant  $b_k = b(\zeta_k)$ , giving an eigenfunction which decays both as  $t \rightarrow +\infty$  and  $t \rightarrow -\infty$ . At such an eigenvalue,  $a'_k = a'(\zeta_k)$ .

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 4.

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 [30, Eq. A6.6e], one has

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



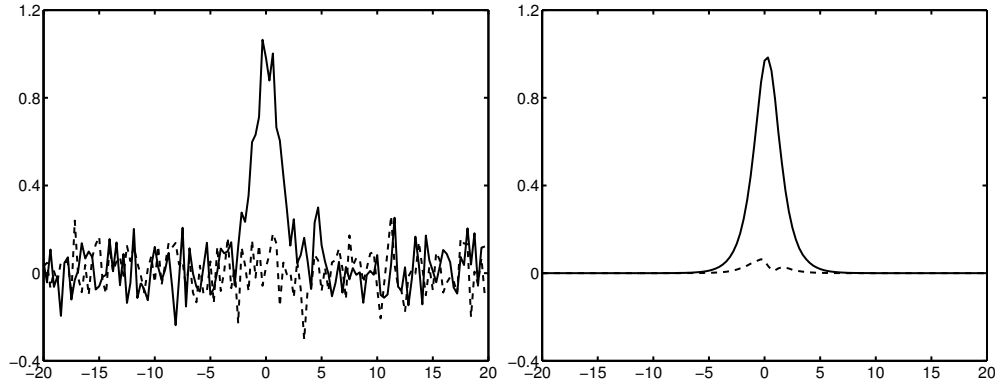


FIG. 4. A noisy soliton (left), and the “clean” soliton (right) recovered from the Jost functions of the Zakharov-Shabat eigenproblem. The Jost functions are mildly perturbed by radiation, as seen in the imaginary component (dashed line) of the clean soliton, but the radiation is otherwise stripped from the soliton.

Since  $\phi_{1k} = b_k \psi_{1k}$  and  $\phi_{2k} = b_k \psi_{2k}$ , however, this means that

$$\frac{b_k}{ia'_k} = \left[ 2 \int_{-\infty}^{\infty} \psi_{1k} \psi_{2k} dt \right]^{-1}.$$

Thus, the discrete part of the trace formula becomes

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

Because the numerator and denominator in this expression are both quadratic in  $\psi_k$ , 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 have minimal impact on the efficiency of generating an accurate and unbiased estimate through importance sampling. 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 method 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

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

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

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

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

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

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

#### Appendix D. Probability density function for midpoint model.

The model of the soliton amplitude's random walk obtained by applying first-order soliton perturbation theory and a particular midpoint approximation to Eq. 2.1 was given in Section 5 as

$$A_{n+1} = A_n + \frac{1}{2} (\sqrt{A_{n+1}} + \sqrt{A_n}) s_{n+1}, \quad (\text{D.1})$$

where the  $s_n$  are i.i.d. normal RVs with mean zero and variance  $\sigma^2$ . 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:

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

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  $A_n$ :

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

#### 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*, Third 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**, 505–513 (1991)
- [4] J. P. Gordon and H. A. Haus, *Random walk of coherently amplified solitons in optical fiber transmission*, *Opt. Lett.* **11**, 665–667 (1986)
- [5] C. R. Menyuk, *Non-Gaussian corrections to the Gordon-Haus distribution resulting from soliton interactions*, *Opt. Lett.* **20**, 270–272 (1995)
- [6] T. Georges, *Study of the non-Gaussian timing jitter statistics induced by soliton interaction and filtering*, *Opt. Commun.* **123**, 617–623 (1996)
- [7] 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**, 389–400 (2002)
- [8] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Oxford University Press, Oxford, 1995
- [9] E. Iannone, F. Matera, A. Mecozzi and M. Settembre, *Nonlinear Optical Communication Networks* (Wiley, New York, 1998)
- [10] 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**, 1894 (2003)

- [11] G. E. Falkovich, I. Kolokolov, V. Lebedev and S. K. Turitsyn, *Statistics of soliton-bearing systems with additive noise*, *Phys. Rev. E* **63**, 025601(R) (2001)
- [12] K.-P. Ho, *Non-Gaussian statistics of the soliton timing jitter due to amplifier noise*, *Opt. Lett.* **28**, 2165 (2003)
- [13] G. Biondini, W. L. Kath and C. R. Menyuk, *Importance sampling for polarization-mode dispersion*, *Photon. Technol. Lett.* **14**, 310–312 (2002)
- [14] S. L. Fogal, G. Biondini and W. L. Kath, *Multiple importance sampling for first- and second-order polarization-mode dispersion*, *Photon. Technol. Lett.* **14**, 1273–1275 (2002); 1487 (2002)
- [15] 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**, 105–107 (2003)
- [16] 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**, 597–613 (1997)
- [17] R. Srinivasan, *Importance Sampling: Applications in Communications and Detection* (Springer-Verlag, New York, 2002)
- [18] C. J. McKinstrie and T. I. Lakoba, *Probability-density function for energy perturbations of isolated optical pulses*, *Opt. Expr.* **11**, 3628–3648 (2003)
- [19] D. J. Kaup, *Closure of the squared Zakharov-Shabat eigenstates*, *J. Math. Anal. Appl.* **54**, 849–864 (1976); —, *A perturbation theory for inverse scattering transforms*, *SIAM J. Appl. Math.* **31**, 121–123 (1976)
- [20] Y. S. Kivshar and B. A. Malomed, *Dynamics of solitons in nearly integrable systems*, *Rev. Mod. Phys.* **61**, 763–915 (1989)
- [21] W. L. Kath, *A modified conservation law for the phase of the nonlinear Schrödinger equation*, *Meth. Appl. Analysis* **4**, 141–155 (1997)
- [22] J. Yang, *Complete eigenfunctions of linearized integrable equations expanded around a soliton solution*, *J. Math. Phys.* **41**, 6614–6638 (2000)
- [23] 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**, 1022 (2005)
- [24] E. Veach, *Robust Monte Carlo Methods for Light Transport Simulation*, Ph.D. thesis, Stanford University, 1997
- [25] A. Owen and Y. Zhou, *Safe and effective importance sampling*, *J. Amer. Stat. Assoc.* **95**, 135 (2000)
- [26] G. Biondini, W. L. Kath and C. R. Menyuk, *Importance sampling for polarization mode dispersion: Techniques and applications*, *J. Lightwave Technol.* **22**, 1210 (2004)
- [27] J. A. C. Weideman and B. M. Herbst, *Finite difference methods for an AKNS eigenproblem*, *Math. Comp. Simul.* **43**, 77–88 (1997)
- [28] 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.* **310**, 157 (1998)
- [29] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, *Sov. Phys. JETP* **34**, 62–69 (1972)
- [30] 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**, 249–315 (1974)
- [31] S. Burtsev, R. Camassa, and I. Timofeyev, *Numerical algorithms for the direct spectral transform with applications to nonlinear Schrödinger type systems*, *Journal of Computational Physics* **147**, 166 (1998)
- [32] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*. Society for Industrial and Applied Mathematics, Philadelphia, 1981.