A Path-Based Method for Simulating Large Deviations and Rare Events in Nonlinear Lightwave Systems

By J. Li and W. L. Kath

Errors in nonlinear lightwave systems are often associated with rare, noise-induced, large deviations of the signal. We present a method to determine the most probable manner in which such rare events occur by solving a sequence of constrained optimization problems. These results then guide importance-sampled Monte Carlo simulations to determine the events' probabilities. The method applies to a general class of intensity-based optical detectors and to arbitrarily shaped and multiple pulses.

1. Introduction

In lightwave systems, additive noise produces large signal deviations that lead to system errors [1–3], and determining the probabilities associated with such events is essential for evaluating system performance. Because such large deviations can be extremely rare (e.g., $\leq 10^{-9}$), standard Monte Carlo (MC) simulations are impractical due to the prohibitively large number of samples needed to produce reliable probability estimates.

A number of methods have been devised for improving the simulation efficiency of rare events [4]. Here, we will focus on importance sampling (IS) [5, 6]. IS generates samples using a biased distribution and corrects for the biasing using the likelihood ratio [4, 6]. A key issue with applying IS, however, is identifying the *most probable* regions of sample

This work was supported by National Science Foundation grants DMS-0709070 and DMS-1211912.
Address for correspondence: J. Li, Department of Mathematics, Institute of Natural Sciences, and MOE Key Laboratory of Scientific and Engineering Computing, Shanghai Jiaotong University, Shanghai 200240, China; e-mail: jinglaili@sjtu.edu.cn

DOI: 10.1111/sapm.12122

STUDIES IN APPLIED MATHEMATICS 00:1–15

space leading to the large deviations of interest, and then finding a biasing distribution that concentrates samples in such regions. Identifying these noise configurations not only improves error probability estimates but also can provide insight into mechanisms responsible for errors.

In soliton-based lightwave systems, the most probable noise configurations leading to large deviations have been identified using soliton perturbation theory, both for return-to-zero [7, 8] and differential phase-shift keyed formats [9, 10]. IS has also been used to determine the phase distributions of nonsoliton pulses using a root-mean-square approximation [11]. Studies further extended the method to dispersion-managed (DM) systems by taking advantage of a path-averaged governing equation and a singular perturbation technique [12–14]. More recently, a method applicable to arbitrarily shaped pulses was developed [15]; in this case, the most probable noise configurations were found using a combination of the singular value decomposition (SVD) and the cross-entropy (CE) method.

A related simulation technique is the multicanonical Monte Carlo (MMC) method [16]. MMC produces random samples with a Markov chain MC random walk, and uses an iterative procedure to locate the important regions of state space. When the biasing distribution can be found explicitly, the IS simulations tend to be more efficient.

Here, we pose a constrained optimization problem to identify the most probable noise configurations leading to errors in systems governed by the nonlinear Schrödinger equation (NLSE) with varying dispersion, optical noise, and a receiver subsystem that includes an optical filter and an integrate and dump receiver. Compared with the SVD/CE method [15], we avoid the expensive SVD computation by exploiting the mathematical structure of the governing equation, and replace the CE method with deterministic optimization, significantly improving efficiency. As a result, we are able to extend previous results to determine the large deviations in multiple bit patterns, accounting for pulse interactions.

The rest of this paper is organized as follows. Section 2 describes the lightwave system model, and Section 3 presents our method for determining large deviations. Section 4 applies the proposed method to determine transmission errors in lightwave systems, and Section 5 provides two numerical examples. Section 6 gives some final remarks.

2. Simulation model

Propagation of light in optical fibers is described by the NLSE, which in a dimensionless form is

$$\frac{\partial u}{\partial z} = \frac{i}{2} d(z) \frac{\partial^2 u}{\partial t^2} + i\,|u|^2 u + \sum_{n=1}^{N_a} s_n(t)\,\delta(z - z_n)\,, \tag{1}$$

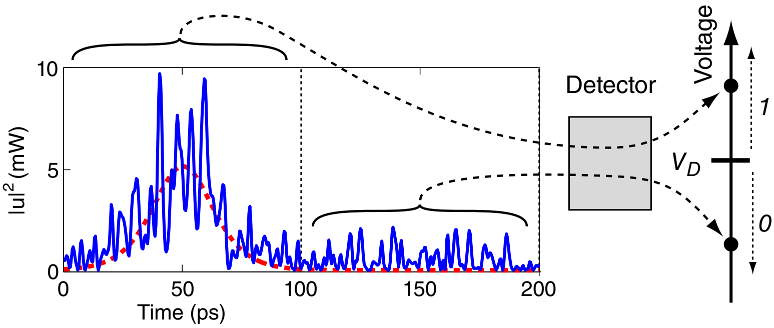

Figure 1. Optical detector schematic showing noisy (solid) and noise-free (dashed) optical signals; $|u|^2$ is the optical power.

where $u(t, z)$ is the optical field envelope, z and t are dimensionless distance and retarded time, respectively [1–3]. For DM systems, $d(z)$ is the dispersion map [1] (otherwise, $d(z) = 1$). The last term represents the random perturbations added at amplifiers due to amplified spontaneous emission (ASE) noise from in-line optical amplifiers [1–3]. N_a is the number of amplifiers, z_n are their locations, and $\delta(z)$ is the Dirac delta distribution. The terms $s_n(t)$ are approximated by i.i.d. Gaussian white noise with $\mathbb{E}[s_n(t)] = 0$ and $\mathbb{E}[s_n(t)s^*_{n'}(t')] = \sigma^2\delta(t - t')\,\delta_{nn'}$, where $\mathbb{E}[\,\cdot\,]$ denotes ensemble average, $\delta(t-t')$ is a Dirac delta function, $\delta_{nn'}$ is the Kronecker delta, and σ^2 is a combination of physical constants and system parameters that determines the noise power [3]. In real systems (and in any simulation), noise has a finite bandwidth, of course [17].

A model detection procedure is illustrated in Fig. 1. At the receiver, the signal is filtered, and optical power is converted into an electrical voltage V, which is then compared to a threshold level V_D to determine whether a "1" bit (if $V > V_D$) or a "0" bit (if $V < V_D$) was sent. Here, we assume an optical bandpass filter $\hat{f}(\omega)$ and an integrate-and-dump detector:

$$V = \int_{T_1}^{T_2} |F^{-1}[\hat{f}(\omega)F[u(t, z_{\text{end}})](\omega)](t)|^2 dt\,, \tag{2}$$

where $F[\cdot]$ and $F^{-1}[\cdot]$ are the Fourier and inverse transforms, respectively. The "integrate" part of this model approximates the electrical filters present in realistic receiver subsystems [17].

In what follows, we will omit writing the final distance $z = z_{\text{end}}$. For our purposes, it is more convenient to write Eq. (2) as:

$$V = \int W(t)|(f * u)(t)|^2 dt = \int W(t)\Big|\int f(t - \tau)u(\tau)d\tau\Big|^2 dt\,. \tag{3}$$

Here, $f * u$ is the time-domain convolution of f and u, and $W(t)$ is a window function: $W(t)$=1 for $t \in [T_1, T_2]$ and $W(t) = 0$ otherwise, Eq. (3) can also be written concisely as $V = \|W(f * u)\|^2$. Integrals are complete unless limits are explicit. The window function is used to specify a particular bit slot (time interval) of interest when the signal $u(\cdot, z)$ contains a series of pulses representing a bit sequence.

Noise can cause the received optical power to vary considerably, so that the detected voltage can be below the threshold value V_D. In this case, a transmitted "1" will be detected as a "0," i.e., an error will occur. A standard MC simulation of such events is straightforward. For each trial, one launches a pulse at z=0, and propagates it to the first amplifier by solving the deterministic part of equation (1) numerically. One then adds randomly generated ASE noise to the signal, propagates the signal to next amplifier, etc. At the end of the transmission line, the optical receiver is applied to obtain the output voltage. One repeats the entire process for different noise realizations and uses the voltage statistics to estimate the desired error probability. Because large deviations are infrequent, methods that can deal with rare events must be employed.

3. Identifying and simulating rare events with IS

Let $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ be an N-dimensional random variable and the event of interest be defined by a scalar function: $g(\mathbf{X}) \leq \rho$, where ρ is a constant (here, $g(\mathbf{X})$ will give the output voltage for a given noise realization). The probability of the event of interest, P, is expressed as:

$$P = \int I_\rho(g(\mathbf{x}))p(\mathbf{x})d\mathbf{x}\,, \tag{4a}$$

where $p(\mathbf{x})$ is the probability density function (PDF) of $\mathbf{X}$, and $I_\rho(g(\mathbf{X})) = 1$ for $g(\mathbf{X}) \leq \rho$ and 0 otherwise. Here, boldface indicates a column vector; a capital denotes a random variable. An unbiased estimator for P is

$$\hat{P} = \frac{1}{M}\sum_{m=1}^{M} I_\rho(g(\mathbf{X}_m)), \tag{4b}$$

where the samples $\{\mathbf{X}_m\}$ are drawn from the distribution $p(\mathbf{x})$ [4, 6]. If P is very small, however, an unreasonable number of samples are necessary to produce a reliable estimate of P. One can write Eqs. (4) as:

$$P = \int I_\rho(g(\mathbf{x}))L(\mathbf{x})p^*(\mathbf{x})d\mathbf{x}\,, \quad \hat{P}^* = \frac{1}{M}\sum_{m=1}^{M} I_\rho(g(\mathbf{X}^*_m))L(\mathbf{X}^*_m)\,, \tag{5}$$

where the samples $\mathbf{X}_m^*$ are now drawn from the *biasing distribution* $p^*(\mathbf{x})$, and where the quantity $L(\mathbf{x}) = p(\mathbf{x})/p^*(\mathbf{x})$ is the *likelihood ratio* [4, 6]. When an appropriate biasing distribution is used, the IS estimator (5) can accurately estimate the probability of the rare event of interest much more efficiently than with straightforward MC methods.

The challenge with implementing IS is to choose a good biasing distribution. For problems where $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ are i.i.d. zero-mean Gaussian random variables with $p(\mathbf{x}) \propto \exp(-\|\mathbf{x}\|^2)$, one choice is to translate the distribution's mean to $\mathbf{x}^*$ using $p^*(\mathbf{x}) = p(\mathbf{x}-\mathbf{x}^*)$, where $\mathbf{x}^*$ is the most probable location satisfying the constraint, i.e., solves [5, 18]

$$\min \|\mathbf{x}\|^2, \quad \text{subject to} \quad g(\mathbf{x}) \leq \rho. \tag{6}$$

Maximizing $\exp(-\|\mathbf{x}\|^2)$ clearly is equivalent to minimizing $\|\mathbf{x}\|^2$. The goal, of course, is to find the most probable deformation leading to a given output voltage change. Also, because larger deformations are typically expected to produce greater voltage changes, we expect the minimum to occur on the constraint boundary, thus we will replace $g(\mathbf{x}) \leq \rho$ with $g(\mathbf{x}) = \rho$ in the above.

When $g(\mathbf{x})$ is complicated and the dimensionality of $\mathbf{x}$ is high, this optimization problem can be difficult to solve. On the other hand, due to the random sampling nature of the IS method, a reasonable approximation to $\mathbf{x}^*$ can adequately guide the simulation. We thus propose the following adaptive method to obtain an approximate solution to Eq. (6):

1. Choose a sequence $g(0) = \rho_0 > \rho_1 > \cdots > \rho_{J-1} > \rho_J = \rho$ of constraint values, so that $\epsilon_j = \rho_{j-1} - \rho_j \ll 1$ for all $j = 0, \ldots, J$.
2. Find the solution to $\min \|\mathbf{x}\|^2$ subject to $g(\mathbf{x}) = \rho_0$ (it is trivial: $\mathbf{x}_0 = 0$); let $j = 1$.
3. Compute an (approximate) solution to Eq. (6) using the result of the previous step, $\mathbf{x}_{j-1}$, as the starting point. (This step is explained in more detail below.)
4. Stop if $j = J$; otherwise, let $j{=}j{+}1$ and go back to step 3.

The main part of the algorithm is to solve the intermediate optimization problem, Eq. (6), in step 3:

$$\min \|\mathbf{x}_j\|^2 = \min \|\mathbf{x}_{j-1} + \Delta\mathbf{x}_j\|^2, \quad \text{subject to} \quad g(\mathbf{x}_{j-1} + \Delta\mathbf{x}_j) = \rho_j. \tag{7}$$

Suppose that we have chosen ϵ_j sufficiently small, so that the constraint in Eq. (7) can be approximated by the linear equation $\mathbf{b}_j^T \Delta\mathbf{x}_j = -\epsilon_j$, where $\mathbf{b}_j$ is the gradient of $g(\mathbf{x})$ at $\mathbf{x}_{j-1}$: $\mathbf{b}_j = \partial g/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_{j-1}}$. This approximate problem can then be solved analytically,

$$\Delta\mathbf{x}_j = -\frac{\epsilon_j}{\|\mathbf{b}_j\|^2}\mathbf{b}_j + \frac{\mathbf{b}_j^T \mathbf{x}_{j-1}}{\|\mathbf{b}_j\|^2}\mathbf{b}_j - \mathbf{x}_{j-1}, \tag{8}$$

and therefore

$$\mathbf{x}_j = -\frac{\epsilon_j}{\|\mathbf{b}_j\|^2}\mathbf{b}_j + \frac{\mathbf{b}_j^T \mathbf{x}_{j-1}}{\|\mathbf{b}_j\|^2}\mathbf{b}_j. \tag{9}$$

As a result, this adaptive algorithm allows an explicit approximation to the solution $\mathbf{x}^*$ of Eq. (6). This is much more efficient than solving the full optimization problem numerically. In general, the gradient $\mathbf{b}_j$ could be computed using numerical differentiation [19], but as will be shown in Section 4, we can derive an explicit semianalytical expression for the gradient by exploiting the mathematical structure of Eq. (1).

4. Application to lightwave systems

We apply the method described in the previous section to lightwave communication systems. In a numerical simulation, the noise term $s_n(t)$ at an amplifier is discretized as a set of independent, identically distributed zero-mean normal random variables, one for each real and imaginary part of the signal at the N_t temporal grid points. Let $\mathbf{s}_n$ be the discrete counterpart of $s_n(t)$ in vector form and $\mathbf{X}^T = [\mathbf{s}_1^T, \mathbf{s}_2^T, \ldots, \mathbf{s}_{N_a}^T]$, where the superscript stands for matrix transpose. Note that $\mathbf{s}_n$ is of dimension $2N_t$ and $\mathbf{X}$ is of dimension $2N_t N_a$.

When a pulse is present, i.e., a binary "1" is being transmitted, the event of interest is that the detected voltage V is below a certain threshold value V_D. We need to solve the optimization problem [8,20],

$$\min_{u(t,z)} \iint \left| \frac{\partial u}{\partial z} - \frac{i}{2} d(z) \frac{\partial^2 u}{\partial t^2} - i|u|^2 u \right|^2 dt\, dz = \min_{\{s_n(t)\}} \sum_{n=1}^{N_a} \|s_n(t)\|^2, \tag{10}$$
$$\text{subject to} \quad V = \|W(f * u)\|^2 = V_D\,.$$

The optimization problem equation (10) can be solved either for $u(t,z)$, corresponding to the most probable pulse deformation, or $\{s_n(t)\}$, corresponding to the most probable noise configuration, and if either is known the other can be computed from Eq. (1). Here, we will use the method described in Section 3 to solve the problem in terms of $\{s_n\}$.

Specifically, we choose $V_0 = \rho_0 > \rho_1 > \cdots > \rho_{J-1} > \rho_J = V_D$ (where V_0 is the voltage detected in the absence of noise), and solve the problem following the steps described in Section 3. Assuming that the translated noise mean at step $j-1$ is $\{s_n^{(j-1)}(t)\}$, the optimization problem (7) becomes

$$\min \sum_{n=1}^{N_a} \|s_n^{(j)}(t)\|^2 = \sum_{n=1}^{N_a} \|s_n^{(j-1)}(t) + \Delta s_n^{(j)}(t)\|^2, \tag{11a}$$

subject to

$$\| W(f * u^{(j)})\|^2 = \rho_j \,, \tag{11b}$$

with

$$\frac{\partial u^{(j)}}{\partial z} = \frac{i}{2} d(z) \frac{\partial^2 u^{(j)}}{\partial t^2} + i|u^{(j)}|^2 u^{(j)} + \sum_{n=1}^{N_a} \left[s_n^{(j-1)}(t) + \Delta s_n^{(j)}(t) \right] \tag{11c}$$

$$\delta(z - z_n), u^{(j)}(0, t) = u_0(t),$$

where $u_0(t)$ is the initial condition.

Next, we derive a linear approximation of the constraint in Eqs. (11). The noise added at each amplifier is small, thus we linearize Eq. (11c) around the solution when $\Delta s_n^{(j)}(t) = 0$, i.e., $u^{(j-1)}(t, z)$. Because the result is now linear, each amplifier's contribution becomes independent,

$$\frac{\partial}{\partial z} \Delta u_n = \mathcal{L}[\Delta u_n], \qquad \Delta u_n(t, z_n) = \Delta s_n^{(j)}(t), \tag{12a}$$

for $n = 1, \ldots, N_a$, where

$$\mathcal{L}[\Delta u] = \frac{i}{2} d \frac{\partial^2 \Delta u}{\partial t^2} + i|u^{(j-1)}|^2 \Delta u + i(u^{(j-1)})^2 \Delta u^*. \tag{12b}$$

The approximate solution of Eq. (11c) thus gives

$$u^{(j)}(t, z_{\text{end}}) = u^{(j-1)}(t, z_{\text{end}}) + \sum_{n=1}^{N_a} \Delta u_n(t, z_{\text{end}}). \tag{13}$$

Substituting Eq. (13) into Eq. (11b) and linearizing the result yields

$$2 \operatorname{Re} \int \left[W(f * u^{(j-1)}(t, z_{\text{end}}) \right]^* \left[(f * \sum_{n=1}^{N_a} \Delta u_n(t, z_{\text{end}})) \right] dt = -\epsilon_j \tag{14}$$

with $\epsilon_j = \rho_j - \rho_{j-1}$. For conciseness, we write Eq. (14) as an inner product:

$$\left\langle 2W\mathcal{F}u^{(j-1)}, \mathcal{F} \sum_{n=1}^{N_a} \Delta u_n \right\rangle = \left\langle 2\mathcal{F}^\dagger W \mathcal{F} u^{(j-1)}, \sum_{n=1}^{N_a} \Delta u_n \right\rangle = -\epsilon_j \,, \tag{15}$$

where $\langle y(t), v(t) \rangle = \operatorname{Re} \int y^*(t)\, v(t)\, dt$, † denotes an adjoint operator (thus, $\langle v, \mathcal{F}y \rangle = \langle \mathcal{F}^\dagger v, y \rangle$), $\mathcal{F}u = f(t) * u(t)$, and $\mathcal{F}^\dagger u = f(-t)^* * u(t)$.

We define the operator $\Phi_{n,}$ so that $\Delta u_n(t, z_{\text{end}}) = \Phi_n \Delta s_n^{(j)}(t)$ gives the linear propagation of $\Delta u_n(t, z)$ from position z_n to the receiver z_{end} as in

Eq. (4). We then use this and $\langle v, \Phi_n \Delta u_n \rangle = \langle \Phi_n^\dagger v, \Delta u_n \rangle$ in Eq. (15) to obtain a linear mapping from $\Delta s_n^{(j)}$ to ϵ_j,

$$\sum_{n=1}^{N_a} \left\langle 2\Phi_n^\dagger \mathcal{F}^\dagger W \mathcal{F} u^{(j-1)},\ \Delta s_n^{(j)} \right\rangle = -\epsilon_j \,. \tag{16}$$

The adjoint $\Phi_n^\dagger v(t) = \Delta u'(t, z_n)$ is given by the solution of

$$\frac{\partial}{\partial z} \Delta u' = -\mathcal{L}^\dagger[\Delta u'] \,, \qquad \Delta u'(t, z_{\text{end}}) = v(t) \,, \tag{17a}$$

where

$$\mathcal{L}^\dagger[\Delta u'] = -\frac{i}{2} d \frac{\partial^2 \Delta u'}{\partial t^2} - 2i |u^{(j-1)}|^2 \Delta u' + i (u^{(j-1)})^2 \Delta u'^* ; \tag{17b}$$

more detail is given in the Appendix. Note $\Phi_n^\dagger v(t) = \Delta u'(t, z_n)$ gives the perturbation $\Delta u'(t, z_n)$ needed to produce a variation $v(t)$ at $z = z_{\text{end}}$.

An approximation to $\Phi_n^\dagger$ can be found numerically via the SVD [21], but this is not necessary. Equation (16) can be written more compactly as:

$$\sum_{n=1}^{N_a} \left\langle w_n^{(j)},\ \Delta s_n^{(j)} \right\rangle = -\epsilon_j , \tag{18}$$

where $w_n^{(j)} = 2\Phi_n^\dagger \mathcal{F}^\dagger W \mathcal{F} u^{(j-1)}$; we then use Eq. (9) to give the minimizer of $\sum_{n=1}^{N_a} \| s_n^{(j)}(t) \|^2$ subject to the linear constraint (18):

$$s_n^{(j)} = -\frac{\epsilon_j}{\sum\limits_{n=1}^{N_a} \| w_n^{(j)} \|^2} w_n^{(j)} + \frac{\sum\limits_{n=1}^{N_a} \left\langle w_n^{(j)}, s_n^{(j-1)} \right\rangle}{\sum\limits_{n=1}^{N_a} \| w_n^{(j)} \|^2} w_n^{(j)} \,. \tag{19}$$

Note the noise mean shift at iteration j depends upon the shift at iteration $j-1$, but because each is small the changes at each amplifier are independent of one another. As a result, the entire path or trajectory of the biased pulse is shifted by a small amount all at once. This continues until the most probable noise configuration reaching the voltage threshold V_D is found. The result is then used to guide IS simulations to find the associated probability [8, 9, 12]. The references give a full description for this portion of the procedure and thus we do not repeat the details here.

5. Simulation results

To demonstrate and validate our method, we first apply it to a previously studied system [12], where an averaged, nonlocal governing equation, the

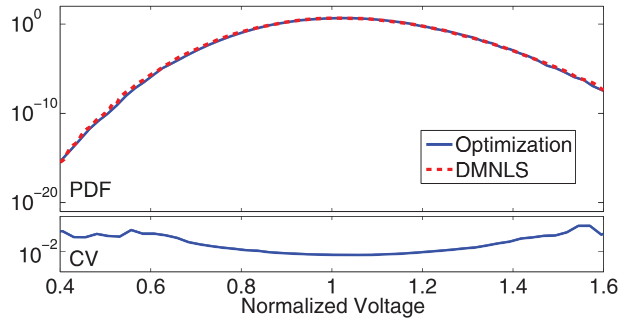

Figure 2. (Top) The voltage probability distribution of DM solitons (solid) reconstructed with our method, and result of [12] (dotted) for comparison. (Bottom) The coefficient of variation (CV) of the IS simulation.

dispersion-managed NLSE [22], was used to guide the IS simulations. This is only an approximation to Eq. (1); here, we simulate it directly. The total propagation distance is 4000 km, the average dispersion is 0.15 ps^2/km with a dispersion map strength $s = 4$ and dispersion map period $z_a = 100$ km. The amplifier spacing is also 100 km. The nonlinear coefficient is 1.7 $(\text{W km})^{-1}$, the fiber loss is 0.21 dB/km, and the transmitted pulses are DM solitons with 11.8 mW peak power. The noise spontaneous emission factor is 1.5, and a 10 GHz Gaussian optical filter is employed [15]. The system parameters can also be expressed in dimensionless units [12].

To employ our method to recover the detected voltage's full probability density, we draw samples using several different biasing distributions (i.e., with different values of V_D) and combine the results using multiple IS [8, 23]. We used seven biasing distributions targeting $V_D = 0.5$, 0.8, 0.9, 1, 1.1, 1.2, 1.5, each with 10,000 realizations. Figure 2 compares the normalized voltage distribution found with our method and that obtained previously [12]. There is a good agreement between the two results. We also show the coefficient of variation (standard deviation divided by the mean) of the IS simulation [24] in the bottom of the figure. These results show that our method is capable of accurately simulating large, rare pulse deformations without resorting to the averaged equation.

A more challenging example demonstrates the applicability of the method to multiple, arbitrarily shaped pulses. Specifically, we consider a chirped-return-to-zero (CRZ) system consisting of a dispersion compensating fiber (pre-DCF) followed by 75 dispersion map periods and with a final dispersion compensating fiber (post-DCF) [25]. Each dispersion map consists of an amplifier followed by 34 km of D_+ fiber and 17.44 km of D_- fiber, and the pre- and post-DCF consist of 51 and 44 km of D_+ fiber, respectively, resulting in a total propagation distance of approximately 4000 km. The

Table 1
Fiber Specifications

	Dispersion ps/(nm-km)	Slope ps/(nm^2-km)	Index Coeff. m^2/W	Eff. Area μm^2	Loss dB/km
$D+$	20.17	0.062	1.7×10^{-20}	106.7	0.19
$D-$	−40.8	−0.124	2.2×10^{-20}	36.1	0.25

optical fiber specifications are given in Table 1. The dispersion slope is modeled by adding a third derivative term to Eq. (1) [26]. The amplifiers have a gain of 10.82 dB and a spontaneous emission factor of 1.2. We use raised cosine pulses with a chirp parameter $A = -0.46$,

$$u(t,z) = \sqrt{\frac{P_{\text{peak}}}{2}\left(1 + \cos\left(\pi \sin\left(\frac{\pi t}{T}\right)\right)\right)} \exp\left(i A\pi \cos\left(\frac{2\pi t}{T}\right)\right),$$

where T=100 ps is the bit period, and P_{peak}=1 mW is the peak optical power. The receiver is a 50 GHz Gaussian filter and an integrate-and-dump detector within each bit period. The numerical simulations used the periodic bit pattern "01110100" containing all possible 3-bit combinations.

We first compute the most probable pulse deformations. In Fig. 3, we plot the optimally deformed pulses (toward a decreased voltage at the receiver after filtering). We then employ an IS simulation to reconstruct the PDF of the average power for the four pulses in the bit pattern. Initially, all the four are launched with an average power of 0.348 mW. In the simulation for each pulse, we bias the samples toward three different voltage variations: $\Delta V = 0$, -0.05, and -0.10 mW. Ten thousand samples are used with each biasing, thus a total of 30,000 samples are used to reconstruct the voltage PDF for each pulse. Figure 4 shows the results. The voltage PDFs for the four pulses are relatively close to one another, but the right-most pulse (i.e., the isolated one) has a moderately lower probability density in the region where errors usually occur. This is consistent with the simple hypothesis that this pulse experiences less distortion from interactions with the others during propagation because of the added separation between it and the rest of the bit pattern.

6. Discussion

We have presented an iterative, path-based method to predict large deviations and determine the associated probabilities of such rare events in nonlinear lightwave systems. We first formulated finding the most probable

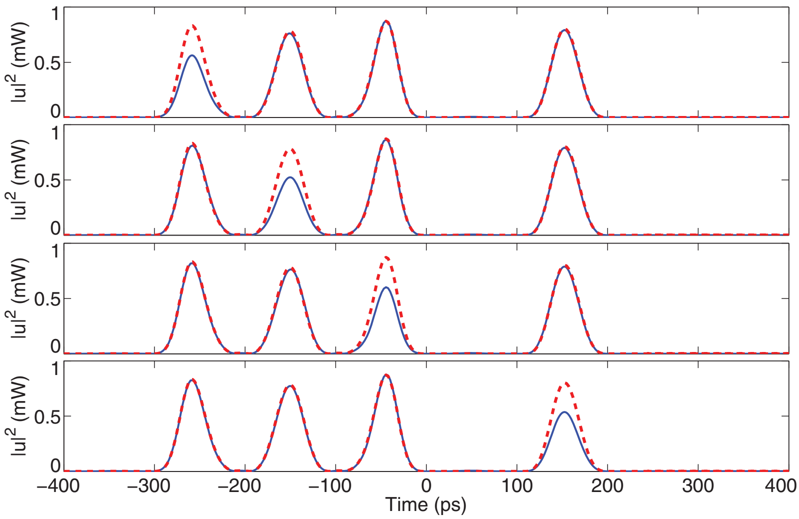

Figure 3. The profiles of the optimally deformed signals (solid lines) just before the receiver (after being filtered). Pulses 1–4 are numbered from top to bottom. Noise-free signals (dashed) are shown for comparison.

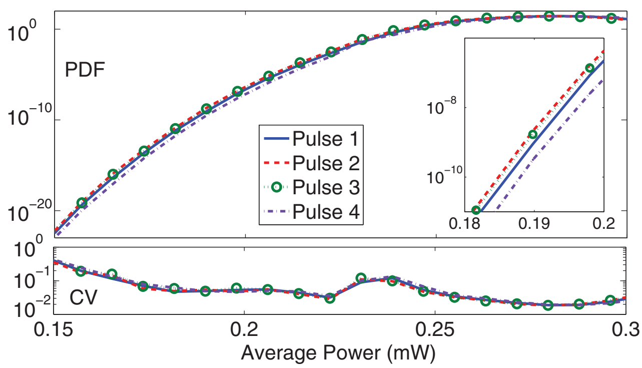

Figure 4. (Top) PDF for the average power for all the four pulses in the bit pattern "01110100." Pulses are numbered 1–4 from left to right in the bit pattern. (Bottom) The coefficient of variation (CV) or the IS simulations.

noise configurations leading to large deviations as a constrained optimization problem. Because a straightforward approach with standard numerical optimization techniques may have difficulties due to the nonlinearly and typical high dimensionality, we considered a sequence of problems with increasing pulse deformation. Exploiting the equation's structure, we can solve each subproblem semianalytically and construct an approximate solution to the overall optimization problem. Errors in the approximation can be controlled by taking sufficiently small increments; in addition, because IS randomly samples the region around the biasing point, the method appears relatively tolerant to such errors and provides good results even when relatively large steps are taken.

We considered two examples to demonstrate the effectiveness of the method: a DM soliton system and a CRZ system. In the first example, we demonstrated good agreement with results obtained previously with an alternate method using the perturbation theory of DM solitons. The second example demonstrated that our method applies to arbitrarily shaped and multiple pulses.

To assess fully the performance of an optical fiber communication (OFC) system, of course, one should also compute the probability that a "0" is misdetected as a "1." While the present method provides a framework for determining large deviations and rare events for ones, for zeros, it is not possible to linearize the behavior of the detector if no underlying pulse is present, and thus the full nonlinear behavior of the detector must be considered. An extension of the method to the case of zeros will be the subject of future work.

We also note that the method as currently implemented may have difficulty if more than one type of pulse deformation or large deviation occurs with roughly equal probability, leading to possible bifurcations of the most probable error mode [14]. In such situations, the iterative procedure described here would need to be modified to incorporate branch switching techniques from bifurcation theory to detect and follow such changes [27]. This should be possible because our iterative method is at its roots a continuation method of the type employed in bifurcation theory.

Acknowledgments

We would like to thank Gino Biondini, Graham Donovan, and Curtis Menyuk for many helpful discussions, and the reviewers for a number of constructive comments.

Appendix: Adjoint of the Linearized NLSE

Let Φ be the linear operator characterizing the propagation of a pulse from z_a to z_b via the linearized NLSE. Namely, when applied to a field $v(t)$, it maps $v(t)$ to $\Delta u(t, z_b)$, where $\Delta u(t, z)$ is governed by

$$\frac{\partial}{\partial z}\Delta u = \mathcal{L}[\Delta u]\,, \qquad \Delta u(t, z_a) = v(t), \tag{A.1}$$

where $\mathcal{L}$ is given by Eq. (12b). Also, let $\Phi^\dagger$ be the operator mapping a field $y(t)$ to $\Delta u'(t, z_a)$ via

$$\frac{\partial}{\partial z}\Delta u' = -\mathcal{L}^\dagger[\Delta u']\,, \qquad \Delta u'(t, z_b) = y(t)\,, \tag{A.2}$$

where $\mathcal{L}^\dagger$ is given by Eq. (17b). We will drop the superscript $(j-1)$ in Eqs. (12b) and (17b). We will show that $\langle y, \Phi v\rangle = \langle \Phi^\dagger y, v\rangle$. First, define

$$\vec{u} = \begin{pmatrix} \mathrm{Re}\,(u) \\ \mathrm{Im}\,(u) \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} \mathrm{Re}\,(v) \\ \mathrm{Im}\,(v) \end{pmatrix}, \qquad \Delta\vec{u} = \begin{pmatrix} \mathrm{Re}\,(\Delta u) \\ \mathrm{Im}\,(\Delta u) \end{pmatrix}, \ldots .$$

Thus, Eqs. (A.1) and (A.2) can be recast as:

$$\frac{\partial\, \Delta\vec{u}}{\partial z} = \mathsf{A}\Delta\vec{u}\,, \qquad \Delta\vec{u}(t, z_a) = \vec{v}(t)\,, \tag{A.3}$$

and

$$\frac{\partial\, \Delta\vec{u}'}{\partial z} = -\mathsf{A}^T\Delta\vec{u}'\,, \qquad \Delta\vec{u}'(t, z_b) = \vec{y}(t)\,, \tag{A.4}$$

respectively, where

$$\mathsf{A} = \begin{bmatrix} -\,\mathrm{Im}(u^2) & -(d/2)(\partial^2/\partial t^2) - 2|u|^2 + \mathrm{Re}(u^2) \\ (d/2)(\partial^2/t^2) + 2|u|^2 + \mathrm{Re}(u^2) & \mathrm{Im}(u^2) \end{bmatrix}.$$

The matrix Green's functions $\mathsf{G}(z, t; z_a, \tau)$ and $\mathsf{H}(z, t; z_b, \tau')$ associated with Eqs. (A.3) and (A.4) solve

$$\mathsf{G}_z = \mathsf{AG}\,, \quad z > z_a\,, \quad \mathsf{G}(z_a, t; z_a, \tau) = \delta(t-\tau)\mathsf{I}\,, \tag{A.5a}$$

$$-\,\mathsf{H}_z = \mathsf{A}^T\mathsf{H}\,, \quad z < z_b\,, \quad H(z_b, t; z_b, \tau') = \delta(t-\tau')\mathsf{I}\,, \tag{A.5b}$$

where I is the identity matrix. Then

$$\Delta\vec{u}(t, z) = \int \mathsf{G}(z, t; z_a, \tau)\,\vec{v}(\tau)\,d\tau\,, \tag{A.6}$$

$$\Delta\vec{u}'(t, z) = \int \mathsf{H}(z, t; z_b, \tau')\vec{y}(\tau')\,d\tau'\,. \tag{A.7}$$

We left multiply Eq. (A.5a) by H^T and subtract the transpose of Eq. (A.5b) right multiplied by G to obtain

$$\mathsf{H}^T\mathsf{G}_z + \mathsf{H}_z^T\mathsf{G} = \mathsf{H}^T\mathsf{AG} - (\mathsf{A}^T\mathsf{H})^T\mathsf{G}\,. \tag{A.8}$$

We will integrate this with respect to t from $-\infty$ to ∞ and with z from z_a to z_b (note the left-hand size is a perfect derivative). First, consider the t integral of the right-hand size and let $\mathsf{A} = \mathsf{A}_1 + \mathsf{A}_2$, where

$$\mathsf{A}_1 = \begin{bmatrix} -\,\mathrm{Im}(u^2) & -2|u|^2 + \mathrm{Re}(u^2) \\ 2|u|^2 + \mathrm{Re}(u^2) & \mathrm{Im}(u^2) \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} 0 & -\frac{d}{2}\frac{\partial^2}{\partial t^2} \\ \frac{d}{2}\frac{\partial^2}{\partial t^2} & 0 \end{bmatrix}. \tag{A.9}$$

Immediately, $(\mathsf{A}_1^T \mathsf{H})^T = \mathsf{H}^T \mathsf{A}_1$ because A_1 is just a matrix. Similarly,

$$\int \mathsf{H}^T \mathsf{A}_2 \mathsf{G}\, dt = \frac{d}{2} \int \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{11} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 g_{21}}{\partial t^2} & -\frac{\partial^2 g_{22}}{\partial t^2} \\ \frac{\partial^2 g_{11}}{\partial t^2} & \frac{\partial^2 g_{12}}{\partial t^2} \end{bmatrix} dt$$

$$= \frac{d}{2} \int \begin{bmatrix} \frac{\partial^2 h_{21}}{\partial t^2} & -\frac{\partial^2 h_{11}}{\partial t^2} \\ \frac{\partial^2 h_{22}}{\partial t^2} & -\frac{\partial^2 h_{12}}{\partial t^2} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} dt = \int (\mathsf{A}_2^T \mathsf{H})^T \mathsf{G}\, dt\,. \quad \text{(A.10)}$$

Therefore, integrating Eq. (A.8) with respect to both t and z and using Eqs. (17b), we find

$$\mathsf{H}^T(z_a, \tau; z_b, \tau') = \mathsf{G}(z_b, \tau'; z_a, \tau)\,. \quad \text{(A.11)}$$

Finally, we can connect the forward and adjoint propagators:

$$\begin{aligned} \langle y, \Phi v \rangle &= \int \vec{y}^{\,T}(t) \Delta \vec{u}(z_b, t) dt = \iint \vec{y}^{\,T}(t)\, \mathsf{G}\,(z_b, t; z_a, \tau)\, \vec{v}(\tau)\, d\tau\, dt \\ &= \iint \vec{v}^{\,T}(\tau) \mathsf{G}^T(z_b, t; z_a, \tau) \vec{y}(t)\, d\tau\, dt \\ &= \iint \vec{v}^{\,T}(\tau) \mathsf{H}(z_a, \tau; z_b, t) \vec{y}(t)\, dt\, d\tau \\ &= \int \vec{v}^{\,T}(\tau) \vec{w}(z_a, \tau) d\tau = \int \vec{w}^{\,T}(z_a, \tau) \vec{v}(\tau) d\tau = \langle \Phi^\dagger y, v \rangle\,. \end{aligned}$$

References

1. G. P. Agrawal, *Fiber-Optic Communication Systems*, (3rd ed.), John Wiley & Sons, Inc., New York, 2002.
2. A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Oxford University Press, Oxford, 1995.
3. E. Iannone, F. Matera, and M. Settembre, *Nonlinear Optical Communication Networks*, Wiley, New York, 1998.
4. G. S. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*, Springer, New York, 1995.
5. J. Bucklew, *Introduction to Rare Event Simulation*, Springer, New York, 2004.
6. R. Srinivasan, *Importance Sampling: Applications in Communications and Detection*, Springer, Berlin, 2002.
7. 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).
8. R. O. Moore, G. Biondini, and W. L. Kath, A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons, *SIAM Rev.* 50:523–549 (2008).
9. E. T. Spiller and W. L. Kath, A method for determining most probable errors in nonlinear lightwave systems, *SIAM J. Appl. Dyn. Sys.* 7:868–894 (2008).

10. 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, *IEEE Phot. Tech. Lett.* 17:1022–1024 (2005).
11. A. Tonello, S. Wabnitz, and R. Indik, Importance sampling of Gordon–Mollenauer soliton phase noise in optical fibers, *IEEE Phot. Tech. Lett.* 18:886–888 (2006).
12. J. Li, E. Spiller, and G. Biondini, Noise-induced perturbations of dispersion-managed solitons, *Phys. Rev. A* 75:053818 (2007).
13. E. Spiller and G. Biondini, Phase noise of dispersion-managed solitons, *Phys. Rev. A* 80:011805 (2009).
14. E. Spiller and G. Biondini, Importance sampling for dispersion-managed solitons, *SIAM J. Appl. Dyn. Sys.* 9:432 (2010).
15. G. Donovan and W. Kath, An iterative stochastic method for simulating large deviations and rare events, *SIAM J. Appl. Math.* 71:903–924 (2011).
16. R. Holzlohner and C. Menyuk, Use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems, *Opt. Lett.* 28:1894–1896 (2003).
17. D. Marcuse, Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers, *J. Lightwave Tech.* 8:1816 (1990).
18. J. Sadowsky and J. Bucklew, On large deviations theory and asymptotically efficient Monte-Carlo estimation, *IEEE Trans. Inf. Theory* 36:579–588 (1990).
19. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, (3rd ed.), Cambridge, 2007, Cambridge University press, Cambridge.
20. W. E, W. Ren, and E. Vanden-Eijnden, Minimum action method for the study of rare events, *Comm. Pure Appl. Math.* 57:637–656 (2004).
21. R. Lehoucq, D. Sorensen, and C. Yang, *ARPACK Users Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, PA, 1998.
22. M. Ablowitz and G. Biondini, Multiscale pulse dynamics in communication systems with strong dispersion management, *Opt. Lett.* 23:1668–1670 (1998).
23. A. Owen and Y. Zhou, Safe and effective importance sampling, *J. Am. Stat. Assoc.* 95:135–143 (2000).
24. G. Biondini, W. Kath, and C. Menyuk, Importance sampling for polarization-mode dispersion: techniques and applications, *J. Lightwave Tech.* 22:1201–1215 (2004).
25. J. Zweck and C. R. Menyuk, Validity of the additive white Gaussian noise model for quasi-linear long-haul return-to-zero optical fiber communications systems, *J. Lightwave Tech.* 27:3324–3335 (2009).
26. R.-M. Mu, V. Grigoryan, C. Menyuk, G. Carter, and J. Jacob, Comparison of theory and experiment for dispersion-managed solitons in a recirculating fiber loop, *IEEE J. Select. Top. Quantum Electron.* 6:248 (2000).
27. R. Seydel, *Practical Bifurcation and Stability Analysis*, (3rd ed.), Springer, New York, 2010.

Shanghai Jiaotong University
Northwestern University

(Received June 30, 2015)