Probability-density function for energy perturbations of isolated optical pulses

C. J. McKinstrie

Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733

mckinstrie@lucent.com

T. I. Lakoba

Department of Mathematics, University of Vermont, Burlington, VT 05405

lakobati@emba.uvm.edu

Abstract: The mathematical methods required to model simple stochastic processes are reviewed briefly. These methods are used to determine the probability-density function (PDF) for noise-induced energy perturbations of isolated (solitary) optical pulses in fiber communication systems. The analytical formula is consistent with the numerical solution of the energy-moment equation. System failures are caused by large energy perturbations. For such perturbations the actual PDF differs significantly from the (idealized) Gauss PDF that is often used to predict system performance.

OCIS codes: (060.2230) fiber optics communications; (190.5530) pulse propagation and solitons; (030.6600) statistical optics

---

References and links

1. L. F. Mollenauer, J. P. Gordon and P. V. Mamyshev, "Solitons in high bit-rate, long-distance transmission," in *Optical Fiber Telecommunications IIIA*, edited by I. P. Kaminow and T. L. Koch (Academic, San Diego, 1997), pp. 373–460.
2. E. Iannone, F. Matera, A. Mecozzi and M. Settembre, *Nonlinear Optical Communications Networks* (Wiley, New York, 1998).
3. K. Ishida, T. Kobayashi, J. Abe, K. Kinjo, S. Kuroda and T. Misuochi, "A comparative study of 10 Gb/s RZ-DPSK and RZ-ASK WDM transmission over transoceanic distances," OFC 2003, paper ThE2.
4. J. X. Cai, D. G. Foursa, C. R. Davidson, Y. Cai, G. Domagala, H. Li, L. Liu, W. W. Patterson, A. N. Pilipetskii, M. Nissov and N. S. Bergano, "A DWDM demonstration of 3.73 Tb/s over 11,000 km using 373 RZ-DPSK channels at 10 Gb/s," OFC 2003, paper PD22.
5. J. P. Gordon and H. A. Haus, "Random walk of coherently amplified solitons in optical fiber transmission," Opt. Lett. **11**, 665–667 (1986).
6. C. J. McKinstrie, "Gordon–Haus timing jitter in dispersion-managed systems with distributed amplification," Opt. Commun. **200**, 165–177 (2001) and references therein.
7. J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communication systems using linear amplifiers," Opt. Lett. **23**, 1351–1353 (1990).
8. C. J. McKinstrie, C. Xie and T. I. Lakoba, "Efficient modeling of phase jitter in dispersion-managed soliton systems," Opt. Lett. **27**, 1887–1889 (2002) and references therein.
9. S. N. Vlasov, V. A. Petrishchev and V. I. Talanov, "Averaged dersciption of wave beams in linear and nonlinear media," Radiophys. Quantum Electron. **14**, 1062–1070 (1971).
10. W. J. Firth, "Propagation of laser beams through inhomogeneous media," Opt. Commun. **22**, 226–230 (1977).
11. D. Anderson, "Variational approach to pulse propagation in optical fibers," Phys. Rev. A **27**, 3135–3145 (1983).
12. D. J. Kaup, "Perturbation theory for solitons in optical fibers," Phys. Rev. A **42**, 5689–5694 (1990).
13. H. A. Haus and Y. Lai, "Quantum theory of soliton squeezing: a linearized approach," J. Opt. Soc. Am. B **7**, 386–392 (1990).
14. D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," J. Lightwave Technol. **8**, 1816–1823 (1990).

15. J. P. Gordon and L. F. Mollenauer, "Effects of fiber nonlinearities and amplifier spacing on ultra-long distance transmission," J. Lightwave Technol. **9**, 170–173 (1991).
16. P. A. Humblet and M. Azizoğlu, "On the bit error rate of lightwave systems with optical amplifiers," J. Lightwave Technol. **9**, 1576–1582 (1991).
17. J. S. Lee and C. S. Shim, "Bit-error-rate analysis of optically preamplified receivers using an eigenfunction expansion method in optical frequency domain," J. Lightwave Technol. **12**, 1224–1229 (1994).
18. T. Yoshino and G. P. Agrawal, "Photoelectron statistics of solitons corrupted by amplified spontaneous emission," Phys. Rev. A **51**, 1662–1668 (1995).
19. B. A. Malomed and N. Flytzanis, "Fluctuational distribution function of solitons in the nonlinear Schrödinger system," Phys. Rev. E **48**, R5–R8 (1993).
20. G. E. Falkovich, I. Kolokolov, V. Lebedev and S. K. Turitsyn, "Statistics of soliton-bearing systems with additive noise," Phys. Rev. E **63**, 25601R (2001).
21. 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 Technol. **20**, 389–400 (2002).
22. 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).
23. C. J. McKinstrie and P. J. Winzer, "How to apply importance-sampling techniques to simulations of optical systems," http://arxiv.org/physics/0309002.
24. H. Kim and A. H. Gnauck, "Experimental investigation of the performance limitation of DPSK systems due to nonlinear phase noise," IEEE Photon. Technol. Lett. **15**, 320–322 (2003).
25. C. W. Gardiner, *Handbook of Stochastic Methods, 2nd Ed.* (Springer, Berlin, 2002).
26. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
27. J. G. Proakis, *Digital Communications, 3rd Ed.* (McGraw-Hill, New York, 1995).
28. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), result 29.3.81.
29. W. Horsthemke and R. Lefever, *Noise-Induced Transitions* (Springer, Berlin, 1983), Sections 5.5 and 6.7.
30. R. Graham, "Hopf bifurcation with fluctuating control parameter," Phys. Rev. A **25**, 3234–3258 (1982).
31. R. Graham and A. Schenzle, "Carleman imbedding of multiplicative stochastic processes," Phys. Rev. A **25**, 1731–1754 (1982).
32. I. S. Gradsteyn and I. M. Rhyzhik, *Table of Integrals, Series and Products, 5th Ed.* (Academic, San Diego, 1994), result 6.633.2.
33. P. J. Winzer, S. Chandrasekhar and H. Kim, "Impact of filtering on RZ-DPSK reception," IEEE Photon. Technol. Lett. **15**, 840–842 (2003) and references therein.
34. J. D. Hoffman, *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1992).

---

1. Introduction

In the last two decades there has been considerable interest in optical communication systems based on return-to-zero (RZ) pulses. In amplitude-shift-keyed (ASK) systems [1, 2] information is transmitted by the presence (or absence) of a pulse. In differential-phase-shift-keyed (DPSK) systems [3, 4] information is transmitted by the phase difference between neighboring pulses. Long-haul systems need amplifiers to compensate for fiber loss. Unfortunately, amplifier noise causes the amplitude (energy), phase, frequency and arrival-time of the pulses to fluctuate (jitter). These pulse-parameter fluctuations (perturbations) impair system performance. For ASK systems the dominant impairment is arrival-time jitter [5, 6], in which noise-induced frequency shifts are converted into arrival-time shifts by dispersion. For DPSK systems the dominant impairment is phase jitter [7, 8], in which energy (power) shifts are converted into phase shifts by nonlinearity. Although energy jitter is not the dominant impairment for either type of system, it should be considered first because it affects both types of system and is the simplest jitter phenomenon to model. Furthermore, the mathematical methods and physical insights developed in the study of energy jitter can be applied to studies of the other jitter phenomena.

Light-wave propagation in a fiber is governed by the nonlinear Schrödinger equation (NSE). The standard way to model pulse-parameter perturbations is to use the moment [9], variation [10, 11] or soliton-perturbation [12, 13] method to convert the NSE into a set of stochastic differential equations (SDEs) for the pulse parameters. If one assumes that the parameter perturbations are small, one can model their growth using linearized SDEs, which produce Gauss probability-density functions (PDFs). This approach produces (approximate) formulas that de-

scribe how the means and variances of the PDFs depend on the system parameters (such as length): One can use these formulas to distinguish reasonable system designs from bad ones. However, system failures are caused by large parameter perturbations that occur infrequently. For such parameter perturbations the linearized SDEs (and the associated Gauss PDFs) are not valid: One cannot rely on the aforementioned formulas to make quantitative predictions about system performance.

Non-Gauss PDFs are produced by two independent mechanisms: nonlinear propagation in fibers and nonlinear (quadratic) detection at receivers. The effects of the latter mechanism on the bit-error-ratios (BERs) of ASK systems with (idealized) integrate-and-dump receivers were studied in detail [14, 15, 16]. Propagation was not analyzed, but characteristic functions were used to calculate the BERs associated with specified signals and specified amounts of amplitude noise. Subsequently, Karhunen–Loève expansions were used to extend these analyses to include more-realistic receivers [17, 18]. Recent studies included the effects of propagation. Malomed [19] used a probability-diffusion, or Fokker-Planck, equation (FPE) to study the statistics of damped soliton systems driven by noise. (These systems are not like communication systems, in which fiber loss is compensated by amplification.) The statistics of undamped soliton systems perturbed by noise (which are like communication systems) were studied recently by Falkovich [20], who used a path-integral formalism to predict a non-Gauss energy PDF. Non-Gauss energy PDFs were observed in recent numerical simulations [21, 22, 23], and a non-Gauss phase PDF (which is a manifestation of a non-Gauss energy PDF) was observed in a recent experiment [24].

In this paper energy jitter is studied in detail. The paper is organized as follows: In Section 2 the mathematical methods required to model nonlinear stochastic processes are reviewed briefly. These methods include stochastic calculus, which facilitates the solution of SDEs, and the conversion of SDEs to FPEs, for which many solution methods exist. In Section 3 the NSE is used to derive two (equivalent) SDEs for the energy in a bit period. Because the rate at which amplifier noise changes the bit energy depends on the current energy, these SDEs are nonlinear. Energy jitter is analyzed in Section 4, for (idealized) systems with uniformly-distributed amplification (UDA). Formulas are obtained for the PDF of the output energies. In Section 5 the analytical results are validated by numerical solutions of the energy SDEs and their (common) associated FPE. The effects of nonuniformly-distributed amplification (NDA) are studied in Section 6. Finally, in Section 7 the main results of this paper are summarized.

2. Simple stochastic process

The mathematical methods required to model stochastic processes were described in detail by Gardiner [25]. Readers who are familiar with these methods should proceed to Section 3 directly. Others should consider a simple stochastic process that involves a single measurable quantity X. Suppose that this process is modeled by the SDE

$$\dot{X} = a(X, z) + b(X, z) r(z), \tag{1}$$

where $\dot{X} = dX/dz$, a and b are deterministic functions of X and z, and r is a random function of z.

A common (idealized) model for r is white noise, which is characterized by the equations

$$\langle r(z) \rangle = 0, \tag{2}$$
$$\langle r(z) r(z') \rangle = \boldsymbol{\delta}(z - z'), \tag{3}$$

where $\langle \, \rangle$ denotes an ensemble average and $\boldsymbol{\delta}$ denotes a Dirac delta function. (Boldface was used to distinguish the symbol for the delta function from the symbol for a small change in value.)

According to Eq. (3), the correlation distance of white noise is zero. The related function

$$W(z) = \int_0^z r(z')dz' \tag{4}$$

defines a Wiener process [25]. It follows from Eqs. (2)–(4) that

$$\langle W(z) \rangle = 0, \tag{5}$$
$$\langle W(z)W(z') \rangle = \min(z, z'). \tag{6}$$

Let δz be a short, but finite, distance interval and let $\delta W = W(\delta z)$. Then it follows from Eqs. (5) and (6) that a typical noise increment $\delta W \sim \delta z^{1/2}$.

Although the white-noise model is common (and convenient), white noise cannot exist in a real physical system. To understand why not, recall that the space-autocorrelation function

$$G_z(\zeta) = \lim_{L \to \infty} \int_{-L/2}^{L/2} r(z + \zeta)r^*(z)dz/L \tag{7}$$

and the ensemble-autocorrelation function

$$G_e(z, z') = \langle r(z)r^*(z') \rangle. \tag{8}$$

If $r(z)$ represents a stationary stochastic process, the ensemble-autocorrelation is a function of $\zeta = z - z'$ (rather than z and z' separately) and the space average (7) equals the ensemble average (8): $G_e = G_z = G$ [26]. Define the Fourier transform

$$\bar{r}(k) = \int_{-L/2}^{L/2} r(z) \exp(-i2\pi kz)dz \tag{9}$$

and the spectral density

$$S(k) = \lim_{L \to \infty} |\bar{r}(k)|^2/L. \tag{10}$$

[For cases in which $|r|^2$ is (noise) power, S is the power per unit inverse-wavelength.] Then the Wiener–Khinchin (statistical-autocorrelation) theorem [26] states that

$$S(k) = \int_{-\infty}^{\infty} G(\zeta) \exp(-i2\pi k\zeta)d\zeta, \tag{11}$$
$$G(\zeta) = \int_{-\infty}^{\infty} S(k) \exp(i2\pi k\zeta)dk. \tag{12}$$

For white noise the autocorrelation function $G(\zeta) = \delta(\zeta)$, so the associated spectral density $S(k) = 1$: The spectrum has infinite bandwidth and contains infinite power. Such a spectrum cannot exist. [In transmission systems the frequency spectra are limited by the amplifier (or filter) bandwidths, which are finite. They are converted to wavenumber spectra by propagation.]

A more-realistic model for r is weakly-colored noise, the simplest example of which is the rectangular spectral density

$$S(k) = \text{rect}(k/K). \tag{13}$$

This spectrum has a bandwidth of K (which we assume is much broader than any other relevant bandwidth) and contains finite power. The associated autocorrelation function

$$G(\zeta) = K\text{sinc}(\pi K\zeta) \tag{14}$$

has a correlation distance of order $1/K$ (which is much shorter than any other relevant distance scale). One can still use Eqs. (2) and (3) to characterize weakly-colored noise, provided that

one interprets δ as the sinc function in Eq. (14). For weakly-colored noise Eq. (4) does not define a Wiener process. However, Eq. (5) is still satisfied exactly and Eq. (6) is satisfied approximately. In particular, the increment $\int_0^{\delta z} r(z')dz' \sim \delta z^{1/2}$, provided that δz is much longer than the correlation distance.

Because Eq. (1) has a random driving term, each $X(z)$ is one member of an ensemble of solutions. This description of the stochastic process is based on an SDE for the dependent variable X. An alternative description is based on the PDF for the independent variable x. Let $P(x,z)dx$ be the probability that $X(z)$ is in the range $(x, x+dx)$. Then $P(x,z)$ satisfies a probability-diffusion equation, which is called the Fokker-Planck equation (FPE). (In the mathematics literature the probability-diffusion equation is called the Kolmogorov equation.)

The rules of stochastic calculus and the relation between an SDE and its associated FPE both depend on an approximate formula for the small, but finite, increment

$$\delta X = \int_0^{\delta z} \{a[X(z)] + b[X(z)]r(z)\}dz, \tag{15}$$

where, for simplicity of notation, the explicit z-dependence of a and b was suppressed and the initial position was denoted by 0 (rather than z). The first term on the right side of Eq. (15) is of order δz, whereas the second term is of order $\delta z^{1/2}$. By omitting terms of order $\delta z^{3/2}$ and higher, one can make the simplification

$$\int_0^{\delta z} a[X(z)]dz \approx a_0 \delta z, \tag{16}$$

where $a_0 = a(X_0) = a[X(0)]$: The first contribution only depends on the value of X at the beginning of the interval. In contrast, because the second term is larger than the first, one must account for the cumulative change in X. By doing so, one finds that

$$\int_0^{\delta z} b[X(z)]r(z)dz \approx \int_0^{\delta z} \left[b_0 + b_0'b_0 \int_0^z r(z')dz' \right] r(z)dz, \tag{17}$$

where $b' = \partial b/\partial X$. It follows from Eqs. (16) and (17) that

$$\delta X \approx a_0 \delta z + b_0 \int_0^{\delta z} r(z)dz + b_0'b_0 \int_0^{\delta z} \int_0^z r(z)r(z')dz'dz. \tag{18}$$

The first term on the right side of Eq. (18) is deterministic and of order δz. The second is random (with zero mean) and of order $\delta z^{1/2}$. The third is also random, but of order δz. Consequently, its contribution to the random part of δX is insignificant. However, its mean value, if nonzero, is of order δz. It follows from Eq. (18) that

$$\langle \delta X \rangle \approx a_0 \delta z + b_0'b_0 \int_0^{\delta z} \int_0^z \langle r(z)r(z') \rangle dz'dz. \tag{19}$$

According to Eq. (3), the double integral in Eq. (19) contains a delta function, which implies that the noise correlation length $l_c = 0$. As we explained above, no real physical process can have $l_c = 0$. The requirement that l_c be much shorter than the next-shortest length δz introduces an ambiguity into the right side of Eq. (15), as viewed from the perspective of deterministic calculus. Indeed, if r were a deterministic function of z Eq. (15) could be rewritten as

$$\delta X = \int_0^{\delta z} \{a[X(z) + b[X(z)]r(z+l)\}dz + O(l). \tag{20}$$

Equation (20) would be consistent with Eq. (15), provided only that $l \ll \delta z$. The ambiguity, which exists for stochastic processes, is whether one treats $l \gg l_c$ or $l \ll l_c$. The choice one makes determines whether one obtains Ito or Stratonovich calculus, respectively.

In the Ito formalism the correlation in Eq. (19) vanishes: $\langle r(z+l)r(z') \rangle = 0$ because $z' \leq z$ and $l \gg l_c$. One obtains the same result by making the intuitive argument that white noise emitted at position z is not correlated with noise emitted at any previous position, no matter how close, and interpreting the upper limit of the z'-integral as z_-. It follows from either approach that

$$\langle \delta X \rangle \approx a_0 \delta z. \tag{21}$$

It also follows (unambiguously) from Eqs. (3) and (18) that

$$\langle \delta X^2 \rangle \approx b_0^2 \int_0^{\delta z} \int_0^{\delta z} \langle r(z)r(z') \rangle dz' dz, \tag{22}$$

$$= b_0^2 \delta z. \tag{23}$$

In most physical models the functions r and b are meant to be evaluated at the same position z, so the parameter l, defined in Eq. (20), should satisfy $l \ll l_c$. In the Stratonovich formalism $l = 0$. Let I and J denote the double integrals that appear in Eqs. (19) and (22), respectively. Then, by using the sinc function in Eq. (14) to evaluate these integrals, one finds that $J = \delta z + O(1/K)$ and $I = J/2$. Because the correlation distance is much shorter than the interval under consideration ($K\delta z \gg 1$)

$$\langle \delta X \rangle \approx (a_0 + b_0' b_0/2)\delta z, \tag{24}$$

$$\langle \delta X^2 \rangle \approx b_0^2 \delta z. \tag{25}$$

In general, the short, but finite correlation distance associated with weakly-colored noise modifies the mean value of δX (which is called the drift), but does not affect its variance. These conclusions do not depend sensitively on the choice of autocorrelation function. For the special case in which the noise is additive ($b' = 0$), the drift formulas (21) and (24) are identical.

The simplest SDEs can be solved directly. (For example, linear SDEs with additive noise.) However, most SDEs cannot. Just as changes of variables enable the solution of some deterministic differential equations, so also do they enable the solution of some SDEs. In stochastic calculus the change-of-variable rule depends on the relation between the SDE and the associated finite-difference equation (FDE).

In the Ito formulation Eq. (1) is equivalent to the FDE

$$\delta X = a\delta z + b\delta W. \tag{26}$$

Consistent with Eqs. (5) and (6), the noise increment has mean $\langle \delta W \rangle = 0$ and variance $\langle \delta W^2 \rangle = \delta z$. Furthermore, noise increments associated with different positions are uncorrelated. Let $Y = f(X)$, where f is an arbitrary differentiable (and invertible) function. Then

$$\delta Y \approx f_X \delta X + f_{XX} \delta X^2/2, \tag{27}$$

where $f_X = df/dX$. It follows from the discussion of Eq. (18) that the larger deterministic contribution to δX^2 must be retained [Eq. (23)], whereas the other (smaller deterministic and random) contributions need not. Consequently,

$$\dot{Y} = f_X \dot{X} + f_{XX} b^2/2. \tag{28}$$

The change-of-variables rule in Ito calculus differs from the rule in deterministic calculus.

In the Stratonovich formulation Eq. (1) is equivalent to the FDE

$$\delta X = (a + bb_X/2)\delta z + b\delta W, \tag{29}$$

where $b_X = \partial b/\partial X$. It follows from Eqs. (27) and (29) that

$$\delta Y = f_X(a + bb_X/2)\delta z + f_X b\delta W + f_{XX}b^2\delta z/2. \tag{30}$$

To complete the change of variables, one must replace the X-derivatives on the right side of Eq. (30) by Y-derivatives. Let g be the inverse of f, so that $f_X = 1/g_Y$ and $f_{XX} = -g_{YY}/g_Y^3$. Then, by using these facts, one can rewrite Eq. (30) in the canonical form

$$\delta Y = [(a/g_Y) + (b/g_Y)(b/g_Y)_Y/2]\delta z + (b/g_Y)\delta W. \tag{31}$$

This FDE is equivalent to the SDE

$$\dot{Y} = (a + br)/g_Y, \tag{32}$$
$$= f_X \dot{X}. \tag{33}$$

The change-of-variables rule in Stratonovich calculus is the same as the deterministic rule.

Many SDEs resist solution by the change-of-variable method. However, it is sometimes possible to solve the FPEs associated with such SDEs. The derivation of these FPEs is based on the Chapman–Kolmogorov equation [25]

$$P(x,\delta z) = \int_{-\infty}^{\infty} T(\delta x, \delta z|x - \delta x, 0)P(x - \delta x, 0)d(\delta x), \tag{34}$$

where T is the transition-probability function and the initial position was denoted by 0 (rather than z). Suppose that $X(0)$ has the value x. Then $T(\delta x, \delta z|x, 0)d(\delta x)$ is the probability that $X(\delta z)$ has a value between $x + \delta x$ and $x + \delta x + d(\delta x)$. Because δx is a small increment (the transition probability is only significant for small values of δx), one can expand the integrand TP in a Taylor series about x. The result is

$$\begin{aligned}
P + \delta z \partial_z P \approx\ & P\int_{-\infty}^{\infty} T(\delta x, \delta z)d(\delta x) \\
& - \partial_x\left[P\int_{-\infty}^{\infty} \delta x T(\delta x, \delta z)d(\delta x)\right] \\
& + \partial_{xx}\left[P\int_{-\infty}^{\infty} \delta x^2 T(\delta x, \delta z)d(\delta x)/2\right],
\end{aligned} \tag{35}$$

where the explicit dependence of P and T on x and $z = 0$ was suppressed for simplicity of notation. For white noise the random contribution to δX is $b\delta W$ [Eqs. (4) and (18)]. The increment of a Wiener process (δW) has Gaussian statistics [25]. It follows from this fact and Eq. (18) that the Ito transition probability

$$T(\delta x, \delta z) = \frac{\exp[-(\delta x - a\delta z)^2/2b^2\delta z]}{(2\pi b^2\delta z)^{1/2}}. \tag{36}$$

By substituting this formula in Eq. (35) one finds that the first integral is 1, the second is $a\delta z$ and the third is $b^2\delta z$. By cancelling like terms one obtains the Ito FPE

$$\partial_z P = -\partial_x(aP) + \partial_{xx}(b^2 P)/2. \tag{37}$$

For reference, notice that this derivation does not require the shape of T to be known exactly: The cumulative transition probability is always 1, and the mean and variance of δX follow directly from Eqs. (19) and (22), respectively.

If one repeats the preceding derivation, using Eqs. (24) and (25) instead of Eqs. (21) and (23), one obtains the Stratonovich FPE

$$\partial_z P = -\partial_x[(a+b\partial_x b/2)P] + \partial_{xx}(b^2 P)/2, \tag{38}$$

which one can rewrite in the canonical form

$$\partial_z P = -\partial_x(aP) + \partial_x[b\partial_x(bP)]/2. \tag{39}$$

It is clear from Eqs. (37) and (38) that the Ito and Stratonovich FPEs associated with the SDE (1) are formally distinct. However, for systems with additive noise the associated FPEs are identical. [See the discussion after Eq. (25).] For such systems, let Y be an arbitrary function of X. Then the FPEs for $P(y)$ are also identical. The proof of this statement is similar to the derivations of the Ito and Stratonovich change-of-variables rules.

We end this section by noting another difference between the Ito and Stratonovich formulations. Consider the ensemble average $\langle f(X)r \rangle$, where f is an arbitrary function. At any position z the quantity X depends on the noise emitted at all previous positions. The Ito formulation is based on the assumption that noise emitted at position z is not correlated with noise emitted at any previous position, no matter how close. Consequently, in the Ito formulation

$$\langle f(X)r \rangle = 0 \tag{40}$$

by assumption. In contrast, the Stratonovich formulation is based on the assumption that the correlation distance of the noise is short, but finite. At any reference position $z = 0$

$$X(0) \approx X(-\delta z) + b[X(-\delta z)] \int_{-\delta z}^{0} r(z')dz', \tag{41}$$

from which it follows that

$$\langle f[X(0)]r(0) \rangle \approx \langle f[X(-\delta z)]r(0) \rangle + f'[X(-\delta z)]b[X(-\delta z)] \int_{-\delta z}^{0} \langle r(z')r(0) \rangle dz'. \tag{42}$$

Terms proportional to a, b' and f'' were omitted from Eqs. (41) and (42) because they are of higher order in δz. By assumption, the correlation distance of the noise is shorter than δz, so $\langle f[X(-\delta z)]r(0) \rangle = 0$. It follows from Eq. (14) that $\int_{-\delta z}^{0} \langle r(z')r(0) \rangle dz' = 1/2$. Consequently, in the Stratonovich formulation

$$\langle f(X)r \rangle = f'(X)b(X)/2. \tag{43}$$

Equations (40) and (43) are closely related to the finite-increment formulas (21) and (24). Suppose that the SDE (1) is written in the form $\dot{X} = r_x(X,z)$, where r_x is the total rate of change. It follows from the preceding results that in the Ito formulation

$$\langle r_x(z) \rangle = a(X), \tag{44}$$

whereas in the Stratonovich formulation

$$\langle r_x(z) \rangle = a(X) + b'(X)b(X)/2. \tag{45}$$

Let δr_x denote the deviatoric rate of change $r_x - \langle r_x \rangle$. Then

$$\langle \delta r_x(z)\delta r_x(z') \rangle = b^2(X)\delta(z-z'). \tag{46}$$

In the Ito formulation Eq. (46) is exact, whereas in the Stratonovich formulation it is approximate (non-delta-like terms of order 1 were omitted).

3. Energy equation

Light-wave propagation in a fiber is governed by the NSE [1, 2]

$$\partial_z A = [g(z) - \alpha]A/2 - i\beta \partial_{tt}^2 A/2 + i\gamma |A|^2 A + R(z,t), \tag{47}$$

where $\partial_z = \partial/\partial z$, A is the (complex) electric-field amplitude, g is the amplifier gain rate, α, β and γ are the fiber loss, dispersion and nonlinearity coefficients, respectively, and R is a random driving (source) term that models the effects of amplifier noise. Because R is independent of A, the noise is said to be additive. At each position the rate at which noise changes the amplitude is a random function of time. Furthermore the rates of change at different positions are independent. These properties are quantified by the equations

$$\langle R(z,t) \rangle = 0, \tag{48}$$
$$\langle R(z,t)R(z',t') \rangle = 0, \tag{49}$$
$$\langle R^*(z,t)R(z',t') \rangle = S(z)\delta(z-z')\delta(t-t'), \tag{50}$$

where $\langle\,\rangle$ denotes an ensemble average and δ denotes a delta function. (Boldface was used to distinguish the symbol for a delta function from the symbol for a small change in value.) The source strength $S = n_{sp}\hbar\omega g$, where n_{sp} is the spontaneous noise factor (which has a typical value in the range 1.1–1.3) and $\hbar\omega$ is the photon energy.

It is instructive to consider the noise energy that accumulates between the initial point $z = 0$ and the neighboring point $z = \delta z$, in the absence of a signal pulse. It follows from Eq. (47) that

$$\delta A = \int_0^{\delta z} R(z,t)dz + O(\delta z^{3/2}). \tag{51}$$

Let T be the bit period. Then the noise energy in a bit period is

$$\delta E = \int_{-T/2}^{T/2} |\delta A(t)|^2 dt. \tag{52}$$

By combining formulas (51) and (52) one finds that the expected noise energy

$$\langle \delta E \rangle = \int_{-T/2}^{T/2} \int_0^{\delta z} \int_0^{\delta z} \langle R^*(z,t)R(z',t) \rangle dz' dz dt. \tag{53}$$

By using Eq. (50) to evaluate the integrals in Eq. (53), one finds that $\langle \delta E \rangle = S\delta z T \delta(0)$. If one tries to interpret $\delta(t)$ as the Dirac delta function $\delta_D(t)$, one obtains the unphysical result that the frequency bandwidth and energy of the noise are infinite. In real systems the noise bandwidth and energy are limited by the (common) bandwidth of the amplifiers (or filters). Consequently, one should interpret $\delta(t)$ as the delta-like function $F\text{sinc}(\pi F t)$, where F is the amplifier (or filter) bandwidth. By doing so, one obtains the physical result

$$\langle \delta E \rangle = S\delta z FT, \tag{54}$$

where $S\delta z$ is the energy per noise mode and FT is the number of modes [14, 15, 16].

In real systems the wavenumber bandwidth of the noise is also finite. However, when one models the spatial evolution of the bit energy one can interpret $\delta(z)$ as the Dirac delta function $\delta_D(z)$, and use Ito calculus, or one can interpret $\delta(z)$ as the delta-like function $K\text{sinc}(\pi K z)$, where K is the noise bandwidth, and use Stratonovich calculus. The former assumption implies that the correlation distance of the noise is zero, whereas the latter assumption implies that the

correlation distance is short, but finite. We will illustrate both approaches in this section. The required elements of stochastic calulus were reviewed in Section 2.

The total (signal and noise) energy in a bit period is

$$E = \int_{-T/2}^{T/2} |A(t)|^2 dt. \tag{55}$$

The wave power depends quadratically on A. By applying the Ito change-of-variable rule [Eq. (28)] to the real and imaginary parts of Eq. (47), combining the results and integrating the transformed equation over a bit period, one can show that

$$d_z E = [g(z) - \alpha]E + S(z)FT + \int_{-T/2}^{T/2} [A^*(z,t)R(z,t) + A(z,t)R^*(z,t)] dt. \tag{56}$$

The term SFT is present in Eq. (56) because the Ito rule differs from the deterministic rule. No energy-flux terms are present. On average, the noise-energy flux at the time boundaries $(-T/2$ and $T/2)$ is zero. The signal-energy flux can be neglected if the signal pulse is isolated (does not interact with other pulses in the same channel or other channels).

Our goal is to rewrite Eq. (56) in the canonical form

$$d_z E = a(E,z) + b(E,z)r(z), \tag{57}$$

where a and b are deterministic functions and r is a random driving term with unit strength $[\langle r(z) \rangle = 0$ and $\langle r(z)r(z') \rangle = \delta(z - z')]$. Henceforth, the explicit dependence of a and b on z will be suppressed. Let r_e denote the total rate of change of the bit energy [right side of Eq. (56) or (57)] and let δr_e denote the deviatoric rate of change $r_e - \langle r_e \rangle$. Then, for an Ito SDE, r_e has the canonical properties [Eqs. (44) and (46)]

$$\langle r_e(z) \rangle = a(E), \tag{58}$$
$$\langle \delta r_e(z) \delta r_e(z') \rangle = b^2(E)\delta(z - z'). \tag{59}$$

One can deduce formulas for a and b by calculating the requisite moments of the right side of Eq. (56). In the Ito formulation the correlation distance of the noise is zero and $A(z,t)$ depends only on the noise emitted at previous positions $(z' < z)$. It follows from these facts that

$$\langle A^*(z,t)R(z,t) \rangle = 0. \tag{60}$$

By applying this result to the right side of Eq. (56), one finds that

$$\langle r_e(z) \rangle = (g - \alpha)E + SFT, \tag{61}$$
$$\langle \delta r_e(z) \delta r_e(z') \rangle = 2SE\delta(z - z'). \tag{62}$$

Thus, Eq. (56) is equivalent to the Ito SDE

$$d_z E = (g - \alpha)E + SFT + (2SE)^{1/2} r(z), \tag{63}$$

where g and S depend on z. Notice that the rate at which noise changes the bit energy depends on the current energy. Equation (63) is valid for any combination of distributed and lumped amplification.

The preceding description of the energy evolution is based on the Ito SDE (63) for the independent variable E. There is an alternative description, which is based on the PDF for the

independent variable e. Let $P(e,z)de$ be the probability that $E(z)$ is in the range $(e, e+de)$. Then [according to Eq. (37)] the energy PDF satisfies the Ito FPE

$$\partial_z P = (\alpha - g)\partial_e(eP) - SFT\partial_e P + S\partial_{ee}^2(eP). \tag{64}$$

Now consider the Stratonovich analysis. It follows from Eqs. (47) and (55), and the Stratonovich change-of-variable rule [Eq. (33)], that

$$d_z E = [g(z) - \alpha]E + \int_{-T/2}^{T/2}[A^*(z,t)R(z,t) + A(z,t)R^*(z,t)]dt. \tag{65}$$

In contrast to Eq. (56), the term SFT is absent because the Stratonovich rule is the same as the deterministic rule. Our goal is to rewrite Eq. (65) in the canonical form (57). For a Stratonovich SDE r_e has the canonical properties [Eqs. (45) and (46)]

$$\langle r_e(z)\rangle = a(E) + b'(E)b(E)/2, \tag{66}$$
$$\langle \delta r_e(z)\delta r_e(z')\rangle = b^2(E)\delta(z - z'), \tag{67}$$

where $b' = \partial b/\partial E$ and non-delta-like terms of order 1 were omitted. In the Stratonovich formulation the correlation distance of the noise is short, but finite, and $A(z,t)$ depends on the noise emitted at all previous positions and the current position ($z' \le z$). It follows from these facts [and a short calculation similar to that which produced Eq. (43)] that

$$\langle A^*(z,t)R(z,t)\rangle = SF/2. \tag{68}$$

By applying this result to the right side of Eq. (65), one finds that

$$\langle r_e(z)\rangle = (g - \alpha)E + SFT, \tag{69}$$
$$\langle \delta r_e(z)\delta r_e(z')\rangle = 2SE\delta(z - z'). \tag{70}$$

Thus, Eq. (65) is equivalent to the Stratonovich SDE

$$d_z E = (g - \alpha)E + S(FT - 1/2) + (2SE)^{1/2}r(z), \tag{71}$$

where g and S depend on z. It follows from Eq. (71) [and Eq. (38)] that the energy PDF satisfies the Stratonovich FPE

$$\partial_z P = (\alpha - g)\partial_e(eP) - SFT\partial_e P + S\partial_{ee}^2(eP). \tag{72}$$

Although the Ito (63) and Stratonovich (71) SDEs differ slightly, the Ito (64) and Stratonovich (72) FPEs are identical. This result was obtained because the noise term in the NSE (47) is additive [see the discussion after Eq. (39)].

In this section the NSE was used to derive an (Ito or Stratonovich) SDE for the bit (pulse) energy. Solitons in constant-dispersion systems are characterized by four shape parameters: the energy, phase, frequency shift and time delay. Return-to-zero pulses in dispersion-managed systems are characterized by the aforementioned shape parameters, together with the chirp and width. By following a procedure similar to that described above (moment method), one can derive a set of six coupled (Ito or Stratonovich) SDEs, one for each shape parameter. Not only is energy jitter an important phenomenon in its own right, it is also a paradigm for a class of jitter phenomena: Whatever mathematical tools and physical insights are developed in the study of energy jitter can be applied to these other jitter phenomena.

4. Analysis of energy jitter

In the absence of noise the energy is constant. In the presence of noise it undergoes a random walk. For the purposes of illustration, it is enough to consider systems with UDA ($g = \alpha$). The effects of NDA ($g \neq \alpha$) will be discussed in Section 6. Let E_0 denote the initial pulse energy, and let $X = E/E_0$, $\mu = FT$ and $\zeta = Sz/E_0$ denote the normalized bit energy, mode number and normalized distance, respectively. Then, for systems with UDA, the normalized energy is governed by the Ito SDE

$$d_\zeta X = \mu + (2X)^{1/2} r(\zeta), \tag{73}$$

where r is a random driving term with unit strength. (In the Stratonovich SDE the parameter μ is replaced by $\mu - 1/2$.) By definition, the initial condition is $X(0) = 1$. Equation (73) is difficult to solve exactly.

For systems with high signal-to-noise ratios, the probability that X differs significantly from 1 is small. By making the (standard) approximation $X^{1/2} \approx 1$ in Eq. (73), one obtains the linearized equation

$$d_\zeta X \approx \mu + 2^{1/2} r(\zeta). \tag{74}$$

Equation (74) defines X (approximately) as a Gauss (normal) random variable with mean $m_n = 1 + \mu\zeta$ and variance $v_n = 2\zeta$. The associated PDF is

$$P(x, \zeta) \approx \exp[-(x - m_n)^2/2v_n]/(2\pi v_n)^{1/2}. \tag{75}$$

(From a logical standpoint the PDF of the non-negative quantity X cannot be exactly Gaussian, because, if it were, the probability of $X < 0$ would be finite for all $\zeta > 0$. From a practical standpoint this inconsistency is tolerable if the probability of $X < 0$ is exponentially small for system lengths of interest.)

For example, consider a 10-Gb/s system with UDA, $\alpha = 0.21$ dB/Km, $\beta = -0.30$ ps^2/Km ($D = 0.38$ ps/Km-nm) and $\gamma = 1.7$/Km-W. Then a soliton (sech pulse) with a full-width at half-maximum of 30 ps has an input energy of 21 fJ (time-averaged power of 0.21 mW). If the system length $l = 10$ Mm, the output noise power in both polarizations, in a frequency bandwidth of 12 GHz (wavelength bandwidth of 0.1 nm), is 1.7 μW: The (optical) signal-to-noise ratio (SNR) is 21 dB. (Systems with NDA produce the same noise power in shorter distances.) For this system the normalized output-energy variance $2Sl/E_0$ is 6.6×10^{-3} and the output-energy deviation is 8.1×10^{-2}.

Although the standard approximation facilitates the solution of Eq. (73), it constrains the associated PDF (75) to be a symmetric function of x (relative to the mean value). Equation (73) defines X as the sum of independent random increments. Because the size of each increment depends on the current value of X, the actual PDF cannot be a symmetric function of x. Thus, the standard approximation is inadequate. One can remove the coefficient $X^{1/2}$ from the random term in Eq. (73) by making the change of variables $Y = X^{1/2}$, in which case

$$d_\zeta Y = \overline{\mu}/Y + r(\zeta)/2^{1/2}, \tag{76}$$

where $\overline{\mu} = (2\mu - 1)/4$. By making the approximation $Y \approx 1$ in the drift term, one obtains the linearized equation

$$d_\zeta Y \approx \overline{\mu} + r(\zeta)/2^{1/2}. \tag{77}$$

Equation (77) defines Y (approximately) as a normal random variable with mean $m_n = 1 + \overline{\mu}\zeta$ and variance $v_n = \zeta/2$. It follows from this result that X is a non-central chi-squared random variable with mean $m_x = m_n^2 + v_n$ and variance $v_x = 2v_n^2 + 4m_n^2 v_n$ [27]. For the process under consideration $m_x \approx 1 + \mu\zeta$ and $v_x \approx 2\zeta$. The associated PDF can be written approximately as

$$P(x, \zeta) \approx \exp[-(x^{1/2} - m_n)^2/2v_n]/(8\pi x v_n)^{1/2}. \tag{78}$$

Formula (78) exhibits clearly the asymmetric and non-Gaussian nature of the energy PDF. However, its accuracy remains to be determined.

Equation (73) and the initial condition $X(0) = 1$ are equivalent to the FPE

$$\partial_\zeta P = -\mu \partial_x P + \partial_{xx}^2(xP) \tag{79}$$

and the initial condition $P(x,0) = \delta(x-1)$. The moments of x are defined by the formula

$$\langle x^n(\zeta)\rangle = \int_0^\infty x^n P(x,\zeta) d\zeta. \tag{80}$$

By multiplying Eq. (75) by x and x^2, and integrating by parts, one can show that

$$d_\zeta \langle x \rangle = \mu, \tag{81}$$
$$d_\zeta \langle x^2 \rangle = 2(1+\mu)\langle x \rangle, \tag{82}$$

from which it follows that the mean is $1 + \mu\zeta$ and the variance is $2\zeta + \mu\zeta^2 \approx 2\zeta$. For typical systems $\zeta^2 \sim 10^{-5}$, so the approximate PDF (78) predicts the energy mean and variance accurately.

Because the diffusion term in Eq. (79) involves the first power of x, one can solve Eq. (79) by Laplace transforming in x, making the *a priori* assumption that $P(0,\zeta) = 0$, and using the method of characteristics to solve the transformed equation. The result is

$$\overline{P}(s,\zeta) = \frac{\exp[-s/(1+s\zeta)]}{(1+s\zeta)^\mu}. \tag{83}$$

Solution (83) is also called the moment generating function [27], because

$$\langle x^n(\zeta)\rangle = (-1)^n \lim_{s\to 0} d^n \overline{P}(s,\zeta)/ds^n. \tag{84}$$

It follows from Eqs. (83) and (84) that the mean and variance of the PDF are $1 + \mu\zeta$ and $2\zeta + \mu\zeta^2$, respectively, in agreement with the moment results. One can invert the transformed solution by rewriting the numerator as $\exp(-1/\zeta)\exp[1/(\zeta^2 s + \zeta)]$. By using the shift theorem and the relation [28]

$$\exp(k/s)/s^\mu \to (x/k)^{(\mu-1)/2} I_{\mu-1}[2(kx)^{1/2}], \tag{85}$$

where $I_n(z)$ is the modified Bessel function of order n, one finds that

$$P(x,\zeta) = [x^{(\mu-1)/2}/\zeta]\exp[-(1+x)/\zeta]I_{\mu-1}(2x^{1/2}/\zeta). \tag{86}$$

Notice that $P(0,\zeta) = 0$, as assumed. For short distances, one can use the relation $I_n(z) \sim \exp(z)/(2\pi z)^{1/2}$ as $z \to \infty$ to rewrite solution (86) as

$$P(x,t) \approx \frac{x^{\mu/2}\exp[-(x^{1/2}-1)^2/\zeta]}{(4\pi x^{3/2}\zeta)^{1/2}}. \tag{87}$$

The similarities between formulas (78) and (87) are evident. [The main effect of the term $x^{\mu/2}$ in formula (87) is to shift the mean from 1 to $1 + \mu\zeta$.]

One can also use a Sturm–Liouville expansion to solve the FPE (79). Consider the evolution of the conditional probability $P(x,\zeta|x_0,0)$ of observing the value x at the position ζ, provided that x_0 was observed at $\zeta = 0$. (In this context one should consider the normalization energy E_0 as a reference energy rather than the initial energy.) The conditional PDF satisfies the forward

FPE (79) with respect to the variable x. Moreover, it can be shown [25], similarly to Section 2, that the backward FPE (or backward Kolmogorov equation), which $P(x,\zeta|x_0,0)$ satisfies with respect to x_0, has the form of the equation adjoint to (79). Specifically,

$$\partial_\zeta P = (\mu - 2)\partial_{x_0}P + \partial^2_{x_0 x_0}(x_0 P). \tag{88}$$

The initial condition for Eqs. (79) and (88) is

$$P(x,0|x_0,0) = \delta(x - x_0). \tag{89}$$

The boundary conditions belong to the natural type in the Gikhman–Skorokhod classification [29], which means that one should not and cannot specify such conditions. However, one can impose the normalization conditions

$$\int_0^\infty P(x,\zeta|x_0,0)dx = 1 = \int_0^\infty P(x,\zeta|x_0,0)dx_0. \tag{90}$$

Further discussion of the natural boundary conditions can be found in [30]. We will point out issues relevant to our analysis as we proceed.

Following [29], we seek the solution to Eqs. (79) and (88) in the form

$$P(x,\zeta|x_0,0) = \int_0^\infty \exp(-\lambda\zeta)p(x,\lambda)\tilde{p}(x_0,\lambda)d\lambda, \tag{91}$$

where $p(x,\lambda)$ and $\tilde{p}(x_0,\lambda)$ are the original and adjoint eigenfunctions, which satisfy the related eigenvalue equations

$$xd^2_{xx}p - (\mu - 2)d_x p = \lambda p, \tag{92}$$
$$d^2_{xx}(x\tilde{p}) + (\mu - 2)d_x\tilde{p} = \lambda\tilde{p}. \tag{93}$$

According to the standard Sturm–Liouville theory of second-order differential operators, these eigenfunctions form a bi-orthogonal set satisfying the orthogonality relation

$$\int_0^\infty p(x,\lambda)\tilde{p}(x,\lambda')dx = \delta(\lambda - \lambda'). \tag{94}$$

It is the existence of the integral in Eq. (94) that plays the role of the boundary conditions for Eqs. (92) and (93) [30]. One can also show that p and $\tilde{p}$ are related by the simple equation

$$p(x,\lambda) = q(x)\tilde{p}(x,\lambda), \tag{95}$$

where $q(x)$ is the particular (but not necessarily normalizable) solution of the FPE (79) satisfying

$$d_x(xq) - \mu q = 0. \tag{96}$$

In order that a function satisfying condition (90) be representable in the form of expansion (91), the sets of eigenfunctions $p(x,\lambda)$ and $\tilde{p}(x,\lambda)$ must be complete. The completeness relation has the form

$$\int_0^\infty p(x,\lambda)\tilde{p}(x',\lambda)d\lambda = \delta(x - x'). \tag{97}$$

Indeed, if Eq. (97) is satisfied and $f(x)$ is any function for which the following integrals exist, then

$$f(x) = \int_0^\infty \delta(x - x')f(x')dx', \tag{98}$$

$$= \int_0^\infty c(\lambda)p(x,\lambda)d\lambda, \text{ where } c(\lambda) = \int_0^\infty f(x')\tilde{p}(x',\lambda)dx', \tag{99}$$

$$= \int_0^\infty \tilde{c}(\lambda)\tilde{p}(x,\lambda)d\lambda, \text{ where } \tilde{c}(\lambda) = \int_0^\infty f(x')p(x',\lambda)dx'. \tag{100}$$

Equations (98)–(100) are consistent with Eqs. (91) and (94). For a standard Sturm–Liouville problem, in which the operator has real coefficients, the completeness relation (97) is known to hold in a certain functional space [29, 31]. In more general problems, which might occur for multi-variable FPEs, the completeness relations of the eigenfunctions must be established on a case-by-case basis [30]. As we will show shortly, in the case of Eqs. (92) and (93), together with the boundary conditions stated after Eq. (94), the completeness relation (97) is equivalent to the orthogonality relation (94). We reiterate that, in general, the completeness relation has to be established independently in order to justify the eigenfunction expansion (91).

We now return to the derivation of the conditional PDF. The eigenvalue problem (92)–(94) has the well-known solution

$$p(x, \lambda) = x^{(\mu-1)/2} J_{|\mu-1|}[2(\lambda x)^{1/2}], \qquad (101)$$

$$\tilde{p}(x, \lambda) = x^{(1-\mu)/2} J_{|\mu-1|}[2(\lambda x)^{1/2}], \qquad (102)$$

where $J_n(z)$ is the Bessel function of order n. The completeness relation for the eigen-functions (101) and (102) follows from the orthogonality relation (94), because the arguments of the Bessel functions depend symmetrically on λ and x. Notice that the functions $x^{(\mu-1)/2} N_{|\mu-1|}[2(\lambda x)^{1/2}]$, where $N_n(z)$ is the Neumann function of order n, also solve (92) and are bounded at $x = 0$. However, they are not included in the set of eigenfunctions $p(x, \lambda)$ because they are not required for completeness. Notice also that for $\mu > 3/2$ the individual eigenfunctions $p(x, \lambda)$ are not normalizable due to their divergence at $x = \infty$. This behavior does not invalidate expansion (91), but does make its convergence non-uniform in ζ: One cannot interchange the limits $x \to \infty$ and $\zeta \to \infty$ in the resulting solution [31]. By substituting Eqs. (101) and (102) into Eq. (91), and using the relation [32]

$$\int_0^\infty \exp(-\gamma^2 u^2) J_n(\alpha u) J_n(\beta u) u \, du = \exp[-(\alpha^2 + \beta^2)/4\gamma^2] I_n(\alpha\beta/2\gamma^2)/2\gamma^2, \qquad (103)$$

one finds that

$$P(x, \zeta | x_0, 0) = (x/x_0)^{(\mu-1)/2} \exp[-(x_0 + x)/\zeta] I_{\mu-1}[2(x_0 x)^{1/2}/\zeta]/\zeta. \qquad (104)$$

Even though we did not impose boundary conditions on Eqs. (79) and (88), the eigenfunctions (101) and the solution (104) satisfy the condition of vanishing probability flux,

$$-\mu P + \partial_x(xP) = 0, \qquad (105)$$

at the boundaries $x = 0$ and $x = \infty$. Not only does the Sturm–Liouville analysis validate the Laplace-transform analysis ($x_0 = 1$), it also obviates the need for *a priori* boundary conditions.

Solution (104) is identical to the solution obtained in previous studies [14, 15, 16] by a different approach. We expect that the approach presented herein, which is based on a FPE, can be generalized in a straightforward way to obtain the PDFs of other pulse parameters.

Energy PDFs are plotted in Fig. 1 for $\mu = 5$ and $\zeta = 3.3 \times 10^{-3}$, which correspond to the physical parameters described in Section 3. The solid, dot-dashed and dashed curves represent the exact PDF (86), the approximate PDF (78) and the Gauss PDF (75), respectively. All three PDFs have (approximately) the same mean and variance. When plotted on a linear scale (Fig. 1a), the differences between the PDFs are barely perceptible. When plotted on a logarithmic scale (Fig. 1b), it is clear that the exact and approximate PDFs have enhanced high-energy tails and diminished low-energy tails relative to the Gauss PDF. In particular, at high energies these (logarithmic) PDFs decrease linearly with energy, rather than quadratically [20]. To compensate for this distortion, their peaks of are shifted to slightly lower energies (Fig. 1a). For the stated
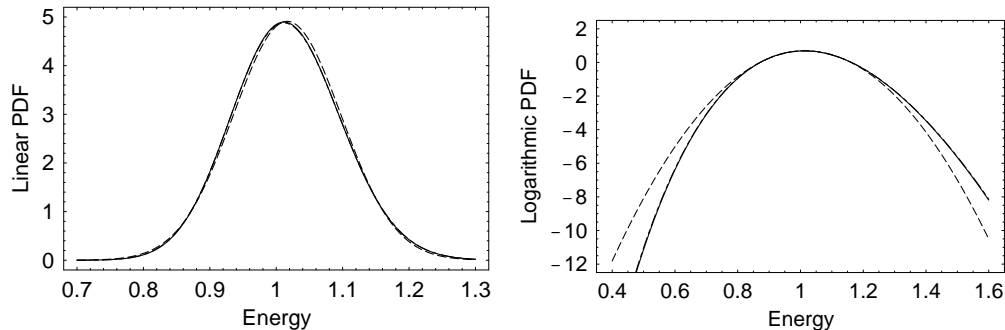
Fig. 1. PDFs for the normalized energy E/E_0. The solid, dot-dashed and dashed curves represent the exact PDF (86), the approximate PDF (78) and Gauss PDF (75), respectively. The differences between the exact and approximate PDFs are barely perceptible. Because the noise-induced energy perturbations depend on the pulse energy, the exact and approximate PDFs (which account for this dependence) have enhanced high-energy tails and diminished low-energy tails relative to the Gauss PDF (which does not).

parameters, the differences between the exact and approximate PDFs are insignificant. The probability of observing the low (normalized) energy $X = 0.5$ is about three orders of magnitude smaller than that predicted by the Gauss PDF and the probability of observing the high energy $X = 1.5$ is about two orders of magnitude larger. As the energy deviation $|E - E_0|$ increases, so also do the differences between the PDFs. These results show the danger of using Gaussian statistics to predict system performance.

In the past, the Gauss approximation was made often to facilitate predictions of bit-error-ratios (BERs). For ASK systems, or DPSK systems with unbalanced detectors, the Gauss approximation predicts BERs that are correct to within an order of magnitude [14, 15, 16, 33]. However, this agreement is fortuitous: The Gauss approximation overestimates the probability of a 1 by several orders of magnitude (Fig. 1*b*), but underestimates the probability of a 0 by a similar amount, and the two errors cancel. For DPSK systems with balanced detectors the Gauss approximation overestimates the BER by several orders of magnitude [33].

5. Simulation of energy jitter

One can verify the results of Section 4 by solving the FPE (79) numerically. We did so by using the Crank–Nicholson (CN) scheme [34], which is the standard finite-difference scheme for diffusion equations.

One can also simulate energy jitter by solving the SDE (73) numerically, and using a random-number generator to specify the noise increments. We will discuss finite-difference schemes for the general SDE (1). Schemes for the energy SDE are special cases of the general schemes.

If one applies the Euler scheme to Eq. (1), one obtains the FDE

$$X_{n+1} = X_n + a(X_n)\delta z + b(X_n)\delta W_{n+1}, \tag{106}$$

where the subscript n denotes the distance $n\delta z$. The noise increments have zero mean and satisfy the correlation equation

$$\langle \delta W_m \delta W_n \rangle = \delta z \delta_{mn}, \tag{107}$$

where δ_{mn} is the Kronecker delta function. It follows from Eqs. (106) and (107) that δW_{n+1} is not correlated with X_n. Equation (106) is the Ito FDE (26).

In contrast, if one applies the Heun (predictor-corrector) scheme to Eq. (1), one obtains the coupled FDEs

$$X^p_{n+1} \ = \ X_n + a(X_n)\delta z + b(X_n)\delta W_{n+1}, \tag{108}$$

$$X^c_{n+1} \ = \ X_n + [a(X_n) + a(X^p_{n+1})]\delta z/2$$
$$+ [b(X_n) + b(X^p_{n+1})]\delta W_{n+1}/2. \tag{109}$$

By using Eq. (108) to approximate the predicted terms in Eq. (109), one finds that

$$X^c_{n+1} \approx X_n + a_n\delta z + (b_n + b'_n b_n \delta W_{n+1}/2)\delta W_{n+1}. \tag{110}$$

Let $\delta X_{n+1} = X^c_{n+1} - X_n$. Then it follows from Eqs. (107) and (110) that

$$\langle \delta X_{n+1} \rangle \ \approx \ (a_n + b'_n b_n/2)\delta z, \tag{111}$$

$$\langle \delta X^2_{n+1} \rangle \ \approx \ b^2_n \delta z. \tag{112}$$

Thus, the Heun scheme is equivalent to the FDE

$$X_{n+1} = X_n + (a_n + b'_n b_n/2)\delta z + b_n \delta W_{n+1}. \tag{113}$$

Equation (113) is the Stratonovich FDE (29).

Although the Heun scheme provides the means to simulate energy jitter in a manner that is consistent with the Stratonovich formulation, it was motivated by mathematical considerations (numerical accuracy), rather than physical considerations. One can also derive a Stratonovich scheme from Eq. (106), by modifying the properties of the noise increments in accordance with the discussion of Section 3 [Eq. (14)]. However, to do so accurately requires that each increment be correlated with many of its neighboring increments: The scheme that results is physically motivated, but cumbersome.

For some stochastic processes the Stratonovich SDE occurs naturally, whereas for others the Ito SDE occurs naturally. The Heun scheme for the former is more accurate than the Euler scheme for the latter [34]. If higher accuracy is required for an Ito SDE, one can convert it to the equivalent Stratonovich SDE (which has a different drift term) and apply the Heun scheme to the latter.
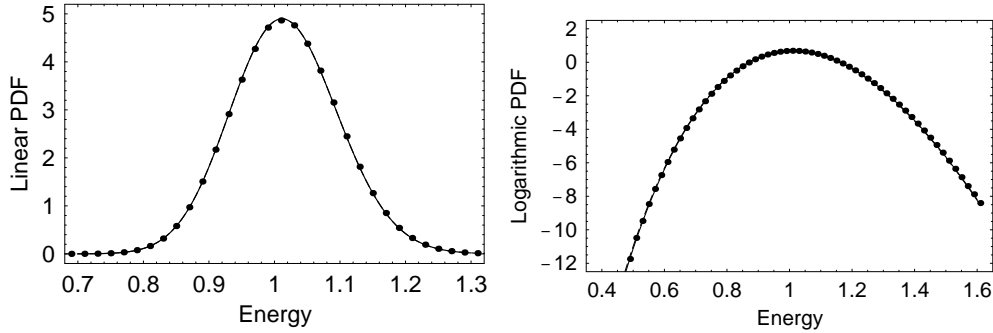

Fig. 2. PDFs for the normalized energy E/E_0. The solid curve represents the analytical solution (86), whereas the dashed curve represents the numerical solution of the FPE (79) and the dots represent the results of importance-sampled simulations based on the FDE (113).

Energy PDFs are plotted in Fig. 2 for $\mu = 5$ and $\zeta = 3.3 \times 10^{-3}$, which correspond to the physical parameters described in Section 3. The solid curve denotes the analytical solution (86),

whereas the dashed curve denotes the numerical solution of the FPE (79) and the dots represent the results of numerical simulations based on the Stratonovich FDE (113), with $a = \mu - 1/2$ and $b = (2X)^{1/2}$, and the correlation equation (107). The simulations were made with an ensemble of 3×10^6 pulses. Importance sampling [22, 23] was used to improve the simulation statistics (measure the probabilities of rare energy perturbations accurately). There is excellent agreement between the analytical predictions and the numerical results. Simulations of energy jitter in systems with lumped amplification were made recently by Moore [22]. There was excellent agreement between the numerical solutions of the energy SDE and the NSE, on which the energy SDE is based. Thus, there is no reason to doubt the accuracy of formula (86) for (model) systems with UDA.

6. Effects of nonuniformly-distributed amplification

For systems with NDA, the normalized energy $X = E/E(0)$ is governed by the Ito SDE

$$d_z X = \mu \sigma_x(z) + \nu(z)X + [2\sigma_x(z)X]^{1/2} r(z), \tag{114}$$

where r is a random driving term with unit strength. The excess gain rate $\nu(z) = g(z) - \alpha$ and the source strength

$$\sigma_x(z) = n_{\mathrm{sp}} \hbar \omega g(z) / E(0) \tag{115}$$

both depend on z.

In the absence of noise the energy varies periodically with distance (oscillates). The spatial period l is the distance between the Raman pumps (or the Erbium amplifiers). In the presence of noise the energy undergoes an oscillatory random walk. Energy decision thresholds (for the detection of 1s) are defined as fractions of the mean energy. In the context of such thresholds, what is important is not the absolute energy variance, but the relative energy variance (which is measured relative to the mean energy). Thus, one can simplify the mathematical analysis of energy jitter and clarify the physical significance of the results by writing the energy variable as a product of oscillatory and non-oscillatory factors. Let $X = Y \exp\left[\int_0^z \nu(z')dz'\right]$. Then the modified energy Y is governed by the equation

$$d_z Y = \mu \sigma_y(z) + [2\sigma_y(z)Y]^{1/2} r(z), \tag{116}$$

where the modified source strength

$$\sigma_y(z) = [n_{\mathrm{sp}} \hbar \omega / E(0)]g(z)e^{-\int_0^z \nu(z')dz'}. \tag{117}$$

Equation (116) is similar to Eq. (73). However, for systems with NDA, the source strength (117) oscillates. One can account for these oscillations by defining the dimensionless distance variable

$$\zeta(z) = \int_0^z \sigma_y(z')dz'. \tag{118}$$

By using the fact that $\delta(z - z')/d_z\zeta = \delta(\zeta - \zeta')$, one can rewrite Eq. (116) in the canonical form

$$d_\zeta Y = \mu + (2Y)^{1/2} r(\zeta). \tag{119}$$

Equation (119) is identical to Eq. (73): The modified energy in systems with NDA evolves in the same way as the absolute energy in systems with UDA. Because the modified and absolute energies only differ by a scale factor, their PDFs are similar: One can apply the results of Section 4 to systems with NDA, provided that one modifies the definition of ζ and rescales the energy variable.

Comparisons of different nonlinear systems are usually based on the equality of path-averaged energies. For systems with UDA, the path-averaged energy E_a equals the input energy $E(0)$ and the path-averaged source strength $\sigma_a = (n_{sp}\hbar\omega\alpha/E_a)$. For systems with NDA, the path-averaged energy

$$E_a = E(0) \int_0^l e^{\int_0^z v(z')dz'}dz \Big/ l. \tag{120}$$

It follows from Eq. (118) that the distance increment

$$\zeta(l) = \sigma_a \int_0^l e^{\int_0^z v(z')dz'}dz \int_0^l e^{-\int_0^z v(z')dz'}dz \Big/ l. \tag{121}$$

For typical system parameters, energy jitter grows on a (long) distance scale of order 1 Mm, whereas the (short) distance between pumps (amplifier spacing) is of order 0.1 Mm. In the context of energy jitter, the slow (but steady) accumulation of the distance increments is more important than the (small) rapid oscillations in ζ, so one can make the approximation

$$\zeta(z) \approx \rho\sigma_a z, \tag{122}$$

where the (dimensionless) ratio ρ is the product of the integrals in Eq. (121) divided by l^2. It follows from this definition that $\rho \geq 1$, which reflects the fact that NDA produces more noise than UDA.

For example, consider a 10-Gb/s system with $\alpha = 0.21$ dB/Km, $\beta = -0.30$ ps^2/Km ($D = 0.38$ ps/Km-nm), $\gamma = 1.7$/Km-W and backward Raman amplification. The gain rate

$$g(z) \approx \frac{\alpha_p\alpha l \exp[\alpha_p\text{mod}(z,l)]}{\exp(\alpha_p l) - 1}, \tag{123}$$

where $\alpha_p = 0.25$ dB/Km is the fiber loss rate at the pump wavelength. A soliton with a full-width at half-maximum of 30 ps has a path-averaged energy of 21 fJ and an input energy of 75 fJ (time-averaged input power of 0.75 mW). If the system length $l = 5$ Mm the output noise power in both polarizations, in a frequency bandwidth of 12 GHz (wavelength bandwidth of 0.1 nm), is 5.4 μW: The (optical) SNR is 21 dB. (The SNR of this 5-Mm system with NDA is comparable to the SNR of the 10-Mm system with UDA that was discussed in Section 4.)
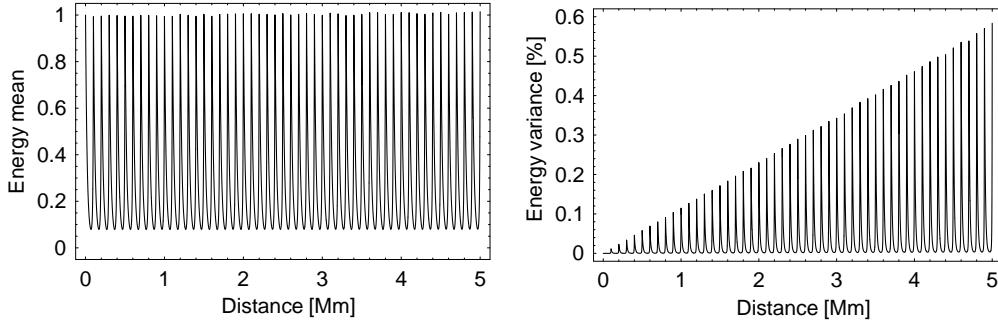

Fig. 3. Mean and variance of the normalized energy $E/E(0)$ plotted as functions of distance. The curves were obtained by solving Eqs. (124) and (125) numerically.

The mean and variance of the normalized energy X are plotted as functions of distance in Fig. 3 for the aforementioned physical parameters and $\mu = 5$. One can determine the mean and variance by using the formulas of Section 4, and evaluating the scale factor $\exp\left[\int_0^z v(z')dz'\right]$

and the distance variable ζ numerically. One can also determine them by solving numerically the moment equations

$$d_z m_x = \nu(z)m_x + \mu\sigma_x(z), \tag{124}$$

$$d_z v_x = 2\nu(z)v_x + 2\sigma_x(z)m_x, \tag{125}$$

which follow from Eq. (114) and the associated FPE. We used the latter approach. It is clear from the figure that the oscillations in excess-gain cause the mean energy to oscillate and the energy variance to oscillate as it grows. These large oscillations obscure the underlying simplicity of the energy-jitter phenomenon.
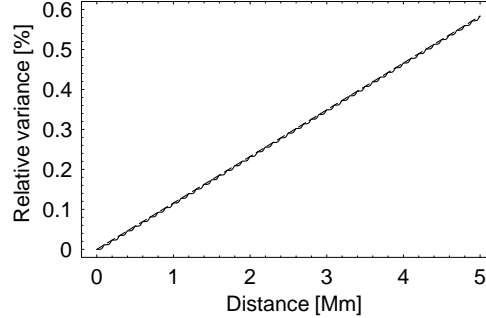

Fig. 4. Relative energy variance plotted as a function of distance. The solid curve was obtained by solving Eqs. (124) and (125) numerically, whereas the dashed line was obtained from an approximate analytical formula.

The variance of the modified energy Y is plotted as a function of distance in Fig. 4. The solid curve was obtained by solving Eqs. (124) and (125) numerically, whereas the dashed line was obtained from the formula $v_y = 2\zeta + \mu\zeta^2$. The distance variable ζ was evaluated by using Eq. (122), with $\rho = 1.8$. As stated above, the oscillations in ζ and (hence) v_y are small: From a practical standpoint v_y grows monotonically.

7. Summary

There are two ways to model stochastic (noisy) processes. The first involves a stochastic differential equation (SDE) for the physical quantity X, in which X is the dependent variable. The second involves a probability-diffusion, or Fokker-Planck, equation (FPE) for the probability-density function (PDF) $P(x)$, in which x is an independent variable. This paper contained brief descriptions of stochastic calculus, which facilitates the solution of SDEs, and the relation between SDEs and their associated FPEs, for which many solution methods exist. The Ito formulation is based on the assumption that the correlation-time of the noise is zero, whereas the Stratonovich formulation is based on the complementary assumption that the correlation-time is nonzero, but short. Both formulations were described.

A detailed study was made of energy jitter in optical communication systems. The nonlinear Schrödinger equation (NSE), which governs light-wave propagation in a fiber, was used to derive exact (Ito and Stratonovich) SDEs for the energy of an isolated bit (pulse and noise). Because the rate at which amplifier noise changes the bit energy depends on the current energy, these SDEs are nonlinear (multiplicative noise) and their (common) associated FPE has variable coefficients. The linearized SDE (additive noise) predicts that the energy PDF is Gaussian. Accurate formulas for the energy PDF were obtained from the nonlinear SDEs and their

associated FPE. For typical system parameters (which were listed in Secs. 4 and 6) the actual PDF differs significantly from the Gauss PDF. Let E_0 be the (unperturbed) output energy in a noiseless system and let E be the (perturbed) output energy in a noisy system. Then the probability of observing the low energy $E = 0.5E_0$ is about three orders of magnitude smaller than that predicted by the Gauss PDF and the probability of observing the high energy $E = 1.5E_0$ is about two orders of magnitude larger: One cannot predict system performance accurately by linearizing the energy SDE (assuming Gaussian statistics). Reduced low-energy probabilities bode well for ASK systems (because the receiver eyes are closed by low energies, which are less likely), whereas increased high-energy probabilities bode ill for DPSK systems [because energy (power) jitter drives phase jitter and phase shifts of either sign close the receiver eyes.]

In (idealized) systems with uniformly-distributed amplification (UDA) the energy variance increases monotonically with distance. In systems with nonuniformly-distributed amplification (NDA) the absolute energy variance oscillates as it grows. However, the relative energy variance (which is measured relative to the mean energy) grows monotonically. The (relative) growth rate in systems with NDA is higher than the (absolute) growth rate in systems with UDA because more noise is emitted in the former.

Finally, the mathematical methods and physical insights developed in this study of energy jitter can be applied to studies of other jitter phenomena (such as arrival-time and phase jitter).