

# Statistical Errors in Biasing Monte Carlo Simulations With Applications to Polarization-Mode Dispersion Compensators

Curtis R. Menyuk, *Fellow, IEEE, Fellow, OSA*

*Invited Paper*

**Abstract**—The possible sources of error in Monte Carlo (MC) simulations are errors in physical modeling, coding errors (bugs), statistical errors, and algorithmic errors. While most algorithmic errors lead to large statistical errors, subtle algorithmic errors that do not lead to statistical errors are at least theoretically possible. These sources of error are reviewed, with the emphasis on statistical errors. Methods for calculating the statistical error in two types of biasing MC simulation, i.e., 1) standard importance sampling and 2) multicanonical, are described. The former requires *a priori* knowledge of how to bias the simulation, while the latter does not. Examples are drawn from the work of the author and his colleagues on calculating the effects of polarization-mode dispersion in optical fiber communication systems. Potential pitfalls when MC simulation codes are not carefully validated and statistical errors are not carefully monitored are described. A proposal for “best practice” in which statistical errors are always presented in conjunction with MC simulations is made.

**Index Terms**—Coding errors, importance sampling, Monte Carlo (MC) simulations, multicanonical simulations, PMD compensators, polarization-mode dispersion (PMD), statistical errors.

## I. INTRODUCTION

IT WOULD BE hard to exaggerate the importance of Monte Carlo (MC) methods. In a recent list of the ten most important computer algorithms, MC methods appear first [1], [2]. While this list is in chronological order—not order of importance—the appearance of MC simulations at the head of the list is still highly significant. The development of electronic computers in the 1950s was motivated at least in part by the desire to carry out MC simulations [3], [4], and conversely, advances in computer technology have enabled the development of increasingly sophisticated MC techniques and their application to increasingly difficult problems [5].

Manuscript received May 14, 2006; revised July 17, 2006. This work was supported by Grants from the Department of Energy (DoE) and the National Science Foundation (NSF).

The author is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: menyuk@umbc.edu).

Color versions of Figs. 2–6, 8, and 9 are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2006.883131

While MC methods make it easy to model complex nonlinear systems that cannot be accurately modeled in any other way and to validate other more approximate techniques that may be far more rapid computationally, they suffer from poor convergence since the statistical errors are expected to decrease only proportional to  $N^{1/2}$ , where  $N$  is the number of samples. This difficulty is exacerbated when the events of interest are expected to occur rarely, as is often the case in optical fiber communications systems. If, for example, one wishes to consider the probability of obtaining 80 or more heads in 100 fair coin tosses, then one would have to simulate on the order of  $10^{15}$  sequences of 100 coin tosses because the probability of obtaining a sequence with 80 or more heads is about  $2.4 \times 10^{-13}$  [6]. As a more germane example, if one wanted to accurately determine the outage probability due to polarization-mode dispersion (PMD) in a system in which this probability is  $10^{-5}$  or less, then one would need to simulate many millions of fiber realizations.

It was recognized in the 1960s that it would be possible to circumvent or at least reduce this difficulty in some cases by biasing MC simulations to make the rare events of interest occur more frequently in the simulations than they do in the systems being modeled [7], [8]. This biasing must be done in a controlled way so that one can weight the result of the simulation to take into account that the rare events of interest have occurred more frequently than in the system being modeled. The goal of any biasing MC simulation is what statisticians refer to as “variance reduction.” If we consider a statistic of interest like the outage probability  $P_{\text{out}}$  of a system with PMD, then the expected relative error in our estimate will be given approximately by  $(1/P_{\text{out}}N)^{1/2}$  in a conventional unbiased MC simulation, where  $N$  is the number of samples. To obtain anything useful from a conventional MC simulation, it must be the case that  $P_{\text{out}}N \gg 1$ , which corresponds to our intuition that we must simulate enough samples so that an outage occurs many times. By contrast, in a biased MC simulation, the expected relative error becomes approximately  $(1/P_{\text{bias}}N)^{1/2}$ , where  $P_{\text{bias}}$  is the probability of observing an outage in the biased simulation, regardless of the true outage probability. When  $P_{\text{bias}}/P_{\text{out}} \gg 1$ , then far fewer samples are needed to calculate the outage probability with high accuracy. Conversely, we have reduced the variance (the square of the expected statistical error) of our estimate of  $P_{\text{out}}$  for a fixed number of samples  $N$ .

A difficulty is that one does not know *a priori* either  $P_{\text{out}}$ —which is after all what one wants to calculate—or  $P_{\text{bias}}$ . One must calculate the statistical error as part of the simulation procedure and verify that it is acceptably small in any MC simulation, biased or not. It can happen in practice that biasing an MC simulation does not significantly reduce the variance relative to a conventional MC simulation and may even increase it and, thus, offers no advantage [7], [8]. As the dimensionality of the configuration space—the space from which the samples are randomly selected—grows, it becomes increasingly difficult to find biases that significantly reduce the variance.

Despite these difficulties, biasing MC methods have been increasingly used to study optical fiber communications systems and components since their introduction to study PMD compensators [9], [10]. Other applications have included determining the performance of soliton [11] and quasilinear [12], [13] amplitude-shift-keyed systems that are dominated by noise impairments or that are dominated by impairments due to cross-phase modulation [14], [15], determining the performance of soliton [16] or quasilinear [17] differential-phase-shift-keyed systems that are impaired by noise, studying induced crosstalk [18], [19], and calculating the noise in semiconductor amplifiers [20], [21]. While the focus here is on PMD applications, the techniques that will be presented here can be used in all these contexts, and indeed, more generally, in any context where biased MC methods are used. A cautionary note is, however, in order. The dimensionality of the configuration space in PMD-related problems is typically small—several hundreds at most—while the configuration spaces in other applications can be much larger. As the dimensionality of the configuration space grows, it becomes increasingly difficult to effectively bias MC simulations.

At the same time, acceptance of biasing MC methods by the optical fiber communications community has been slower than the author believes their power warrants. One reason is that biasing MC methods are subject in principle to errors that conventional MC methods are not. While these difficulties are real—in principle—and should be watched for, other sources of error that conventional MC methods share with biasing MC methods are far more important in practice, and the use of biasing MC methods can help alleviate these shared difficulties. In particular, we note that coding errors (bugs) are a potential source of error in any computation. One of the best ways of eliminating coding errors is to compare codes based on two completely different algorithms that calculate the same quantity. Comparing biasing and conventional MC simulations or two different types of biasing simulation can be very helpful in this regard. This point is worthy of emphasis, and we will return to it several times throughout this paper. It might seem odd to the reader that a paper whose focus is statistical errors pays so much attention to coding errors. In fact, coding errors are the “soft underbelly” of computational work—often encountered and rarely discussed. The usefulness of biasing MC methods in eliminating coding errors is an important motivation for their development and, thus, for the development of the methods for calculating statistical errors that are discussed in this paper.

The remainder of this paper is organized as follows: In Section II, we describe the major sources of error in MC

simulations and discuss their relative importance. In Section III, we describe the model system that we use for the optical fiber transmission line, the PMD compensator, and the receiver. In this paper, we will focus on two different biasing methods, i.e., 1) standard importance sampling, which relies on *a priori* knowledge of where in the configuration space the important regions are located and 2) multicanonical MC (MMC) method, which is an iterative learning method and does not require this *a priori* knowledge. In Section IV, we discuss standard importance sampling, while in Section V, we discuss the multicanonical method. In both cases, our focus is on how to calculate the statistical errors. It is fairly straightforward with standard importance sampling, but it is not at all straightforward with the multicanonical method. Section VI contains the conclusions.

## II. SOURCES OF ERROR

In any computation, one must make physical approximations or assumptions to render the computation tractable. Prior to the advent of importance sampling methods to calculate the outage due to PMD, it was not possible to directly compute outage probabilities of  $10^{-5}$  or lower in systems with PMD compensators, and it was common to assume when modeling compensators with a fixed differential group delay (DGD) element that the element should be chosen to minimize the mean DGD after compensation. When importance sampling became available, it was no longer necessary to make this assumption, which proved to be incorrect [9], [10], [22]. The reason is that outages are caused by rare events with large DGDs. By picking a fixed group delay element to be several times larger than the mean DGD, it is possible to reduce the DGD of those few events that would otherwise lead to an outage and reduce the overall probability of an outage, at the cost of slightly “increasing” the DGD in the vast majority of cases and thus slightly increasing the mean DGD and hence the mean penalty after compensation. At the same time, physical assumptions that are common to both standard and biasing importance sampling simulations—like the underlying fiber transmission, compensator, and receiver models—cannot be checked by comparing two different simulations. Careful comparison to experiments is essential.

Another source of error—perhaps the most important in practice—is coding errors (bugs). The effects of coding errors can be very subtle in practice and difficult to eliminate. It is the author’s experience that newcomers to simulation often attribute strange behavior in their simulations to errors in the physics—or subtle errors in the computational algorithms—when in the vast majority of cases they are observing coding errors. Thus, it is very important to validate a simulation code by checking it against another independently written computer code whenever possible. When it is not possible, the simulation should be checked against known analytical limits. However, this second option is never as effective as the first. By their very nature, analytical limits are special, and coding errors often do not make themselves apparent in these limits. One of the primary benefits that biasing MC methods offers—and for that matter the use of more than one biasing MC algorithm—is

that they allow the user to validate MC simulations. Several examples will be presented in the remainder of this paper.

A third source of errors is statistical errors. Because a finite number of samples are selected in any MC simulation, the value that one calculates for the statistic of interest, such as the expected outage probability of a compensator, will change if the simulation is repeated with a new seed for the random number generator. The expected standard deviation of a statistic is what we refer to as the statistical error. The statistical error can be estimated during the simulation. This estimate will also be subject to a statistical error. However, one does not typically need to know the value of the statistical error with high accuracy; one simply wants to verify that it is small compared to the statistic of interest. In a standard MC simulation with independently drawn samples of equal weight, the statistical error is not difficult to estimate. Moreover, when making a histogram of a quantity of interest, like the eye-opening penalty, the smoothness of the histogram gives some indication that the statistical error is small. In biasing MC simulations that use standard importance sampling, it is still not too difficult to estimate the statistical error, although the variation of the weights is an added complication. By contrast, it is quite difficult to estimate the statistical errors in learning methods like the MMC method, and a way to do so has only recently been found [23]. Moreover, it is important to make these estimates because the samples from the bin-to-bin histograms of the statistic of interest—in our case the DGD or the penalty—are correlated, and hence, the smoothness of the histogram is not a reliable indicator of convergence. The principal purpose of this paper is to show how to calculate the statistical error for both standard importance-sampled and MMC simulations and to illustrate the importance of making these estimates with examples.

The last source of error—and the least important in practice—are subtle algorithmic errors due to contributions from disjoint regions of the configuration space that are not detectable by calculating the statistical errors. One might consider the following scenario: As was previously discussed, optical PMD compensators are typically tuned to optimally compensate a fixed range of DGD values, e.g., three to four times the mean DGD [10], [22]. After electrical detection, a receiver might have a signal processing unit that compensates for most but not all of the PMD errors caused by DGD values from two to three times the mean DGD, but does not succeed at all with errors caused by DGD values of four to five times the mean DGD. There are no errors caused by DGD values in the range of three to four times the mean DGD because the optical compensator eliminates them. To assign numbers to this scenario, we might imagine that one DGD value in ten in our simulation produces a DGD value in the range of two to three times the mean DGD, of which the signal processing algorithm is known *a priori* to correct all but one error in  $10^5$ ; we might imagine that the simulation produces one sample in  $10^6$  that is in the DGD range of four to five; finally, we might imagine that we collect  $10^5$  samples in our simulation. Unless by chance with a probability of one in ten we collect a sample in the DGD range of four to five, we will estimate a probability of error of  $10^{-6}$  with a low statistical error, and we will be off by a factor of 2!

It is somewhat easier to construct a scenario in which a biasing MC simulation will similarly fail. We might suppose that a PMD compensator is constructed so that a large DGD leads to an error and a large second-order PMD leads to an error, but a combination of first- and second-order PMDs with a similar probability of occurring does not lead to an error. We might also suppose that we bias the simulation so that we obtain many samples with high DGD, but no samples with a large second-order PMD. In this scenario, we will once again obtain the wrong answer with a low statistical error.

The reader would be forgiven for thinking that both scenarios just outlined are somewhat artificial. It is not difficult to construct purely mathematical examples of this difficulty [6]; however, the author knows of no concrete examples from realistic settings in optical fiber communications in which this difficulty appears. One reason may be that the configuration spaces in the mathematical settings that illustrate this difficulty are low dimensional (just one dimensional in the example in [6]), while the dimensions of typical configuration spaces in realistic problems are quite large; in large-dimensional spaces, it is difficult to obtain regions that are truly isolated from one another.

However, the author knows of several instances in which differences between biasing MC simulations and standard MC simulations or two types of biasing MC simulation were due to coding errors and disappeared once the errors were corrected—allowing the mutual validation of both simulation codes, in addition to the solution of problems that could not be solved using standard MC simulations. Since the mathematical possibility exists of errors due to disjoint regions in the configuration space that also produce low statistical errors, the user of any MC simulation—biased or unbiased—should keep this possibility in mind. At the same time, given that coding errors or statistical errors are a far more likely source of discrepancy between standard and biasing MC simulations, it is the author's view that researchers should not invoke this possibility as a reason not to validate their MC simulations whenever possible by comparison to an MC simulation of a different type.

### III. SIMULATION MODEL

Our simulation model consists of three principle components, i.e., 1) a fiber transmission model, 2) a compensator model, and 3) a receiver model. These models are described in detail in [22], [24], and [25], and only the essentials are repeated here.

We simulate a system that is nominally a 10-Gb/s nonreturn-to-zero system. The pulses are generated by using perfect rectangular pulses that are filtered by a Gaussian filter to produce a 30-ps rise time. We do not take into account noise, polarization-dependent loss, chromatic dispersion, or fiber nonlinearity in the transmission. Consequently, the data rate can be rescaled to any desired amount by rescaling all the times and filter frequencies. Since the sources of intersymbol interference in our system only couple neighboring bits, it is sufficient to simulate 8-bit de Bruijn sequences that include all possible 3-bit sequences. In some of our work, we have used longer strings; the results are consistent [10], [23]–[27].

We represent the optical fiber transmission using the coarse-step model [28], [29]. The output field at each radial frequency  $\omega$  is represented by a Jones vector  $\mathbf{E}_{\text{out}}(\omega)$ , which is related to the input Jones vector  $\mathbf{E}_{\text{in}}(\omega)$  via a transmission matrix, i.e.,

$$\mathbf{E}_{\text{out}} = \bar{\mathbf{T}}(\omega)\mathbf{E}_{\text{in}}(\omega) \quad (1)$$

where  $\bar{\mathbf{T}}(\omega)$  consists of a concatenation of  $N$  linearly birefringent sections, so that

$$\bar{\mathbf{T}}(\omega) = \prod_{n=1}^N \bar{\mathbf{T}}_n(\omega) \quad (2)$$

where  $\bar{\mathbf{T}}_n(\omega) = \bar{\mathbf{P}}(\omega)\bar{\mathbf{S}}_n$  is the transmission matrix of the  $n$ th fiber section. The matrix  $\bar{\mathbf{S}}_n$  shown in (3) at the bottom of the page, is a Jones matrix that corresponds to a uniformly distributed rotation on the Poincaré sphere, while

$$\bar{\mathbf{P}}(\omega) = \begin{pmatrix} \exp(-i\omega\tau_s/2) & 0 \\ 0 & \exp(i\omega\tau_s/2) \end{pmatrix} \quad (4)$$

models the frequency-dependent phase rotation of the light through a birefringent section. The parameter  $\tau_s$  is the DGD in a single section. The angles  $\xi_n$ ,  $\psi_n$ , and  $\phi_n$  are independent from each other and at each  $n$ . The angles  $\psi_n$  and  $\phi_n$  are uniformly distributed between 0 and  $2\pi$ , while the random variables  $\cos \xi_n$  are uniformly distributed between  $-1$  and  $1$ . These angles are the same as the Euler angles in classical mechanics [30], and the Müller matrix  $\mathbf{S}_n$ , which is equivalent to the Jones matrix  $\bar{\mathbf{S}}_n$  in (3), is composed of elementary rotations about the  $x$ -axis and  $y$ -axis of the Poincaré sphere, i.e.,  $\mathbf{S}_n = \mathbf{R}_x(\psi_n)\mathbf{R}_y(\xi_n)\mathbf{R}_x(\phi_n)$ . Similarly, the Müller matrix corresponding to the Jones matrix  $\bar{\mathbf{P}}(\omega)$  in (4) is composed of an elementary rotation around the  $x$ -axis, i.e.,  $\mathbf{P}(\omega) = \mathbf{R}_x(-\omega\tau_s)$ . We should choose  $\tau_s = (3\pi/8N)^{1/2}\langle\tau\rangle$ , where  $\langle\tau\rangle$  is the expected DGD at the fiber output. In the simulations reported here, we chose  $N = 80$  unless otherwise stated, which is sufficient to obtain a Maxwellian distribution for the DGD in the output probability range down to  $10^{-6}$  [22], [31], [32].

The coarse-step model presented here more closely reproduces the actual transmission behavior than does the still commonly used rotator plate model. As long as the length of the birefringent sections is long compared to the field correlation length—which is required for either the rotating plate model or the coarse-step model to make sense—then the electric field will be randomly scattered on the Poincaré sphere, as assumed by the coarse-step model, and not rotated, as assumed by the rotating plate model. From a more practical standpoint, if  $\omega\tau_s = 2\pi$ , then a Stokes' vector that is initially on the equator of the Poincaré sphere will not be able to leave, and simulation will yield incorrect statistics for the DGD [33]. In

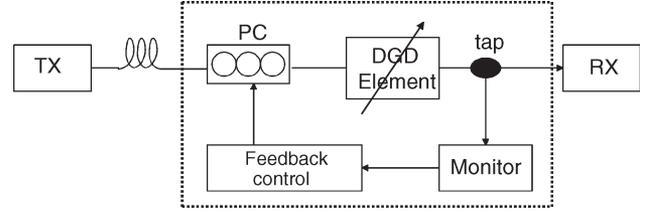


Fig. 1. Schematic illustration of a single-section compensator with a monitor and a feedback element. In practical systems, the compensator will usually be part of the receiver, so that the monitor and feedback control are integrated with the detection circuit. This figure was created by A. O. Lima.

single-channel studies of DGD, this difficulty can be avoided by choosing enough sections so that  $|\omega\tau_s| \ll 2\pi$  for the largest frequency in the signal's bandwidth. However, when systems that include nonlinearity and chromatic dispersion as well as PMD are being modeled, so that the full Manakov-PMD equation should be used [28], [29], it is often not practical to fulfill the condition  $|\omega\tau_s| \ll 2\pi$ . In that case, one must use the coarse-step method.

The optical fiber transmission is followed by an optical PMD compensator. In the examples presented in this paper, we will discuss the results from single-section compensators with fixed DGD, single-section compensators with variable DGD, and three-section compensators. Fig. 1 shows a schematic illustration of a fixed-DGD compensator. The compensator, shown inside the dashed lines, has polarization controller, a DGD element, and a monitor-and-feedback control element. In actual systems, optical PMD compensators are built into the receiver and are not independent as shown here. Practical optical compensators typically use either the degree of polarization of the optical signal or radio frequency tones at one-fourth or one-half the data rate as the feedback signal [34], [35]. The system performance can vary by orders of magnitude, depending on implementation details.

In the studies presented in this paper, we consider two different feedback models. In the first feedback model, we calculate the optical compensation that leads to the smallest eye-opening penalty. Details on how we find the global optimum may be found in [22], [24], and [25]. The value of this feedback model is that it indicates the best that one can possibly do given the transmission and receiver systems. In the second feedback model, we exactly compensate for the DGD at the central frequency or, with three-section compensators, both the first- and second-order PMDs at the central frequency. The value of this feedback model is that it allows one to compare the computational models to analytical studies [22], [24], [34], [35].

The expansion for the polarization dispersion vector—equivalent to the one in [36]—is given by [22], [24], [25]

$$\tau_{\text{tot}}(\omega) = \tau_c + \mathbf{T}_c(\omega)\mathbf{R}_{\text{pc}}\tau_f(\omega) \quad (5)$$

$$\bar{\mathbf{S}}_n = \begin{pmatrix} \cos(\xi_n/2) \exp[i(\psi_n + \phi_n)/2] & i \sin(\xi_n/2) \exp[i(\psi_n - \phi_n)/2] \\ i \sin(\xi_n/2) \exp[-i(\psi_n - \phi_n)/2] & \cos(\xi_n/2) \exp[-i(\psi_n + \phi_n)/2] \end{pmatrix} \quad (3)$$

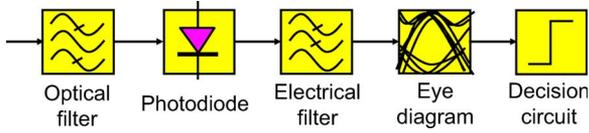


Fig. 2. Schematic illustration of the receiver model. This figure was created by R. Holzlöhner.

where  $\tau_c$  is the polarization dispersion vector of the compensator,  $R_{pc}$  is the polarization rotation due to the polarization controller expressed in the Stokes' space, and  $T_c(\omega)$  is the polarization transformation expressed in the Stokes' space due to the DGD of the compensator. We may write  $R_{pc} = R_x(\phi_{pc})R_y(\psi_{pc})R_x(-\phi_{pc})$ , where  $\phi_{pc}$  and  $\psi_{pc}$  will transform an arbitrary input Stokes' vector into a given output Stokes' vector. We have  $T_c(\omega) = R_{\tau_c}(-\omega\tau_c)$ , where  $R_{\tau_c}$  indicates a rotation about the Stokes' vector pointing in the  $\tau_c$  direction.

A schematic illustration of the receiver model is shown in Fig. 2. This basic model has been the subject of careful studies by us and by others [37], dating back to early work by Marcuse [38] and Humblet and Azizoğlu [39], which followed the invention of the erbium-doped amplifier. Prior to that, most work on optical receivers focused on reducing electrical noise, which was the major source of impairments in optical communications systems based on repeaters [40]. The first element in our receiver model is an optical filter, which in the studies reported here is a Gaussian filter with a full-width at half-maximum bandwidth of 60 GHz. The second element is a square-law photodetector. The third element is an electrical filter, which in the studies reported here is a fifth-order Bessel filter with a 3-dB bandwidth of 8.6 GHz. Since there is no noise in our model system, the electrical filter is the primary source of intersymbol interference, and we optimized its bandwidth empirically. The final element is the detection algorithm, which we based on an algorithm described by Trischitta and Varma [41]. We delay the signal by half a bit period and then subtract it from the original signal, which is then squared. As a result, a strong tone is produced at 10 GHz. The decision time is set equal to the time at which the phase of this tone is  $\pi/2$ .

Our penalty measure is the eye-opening penalty. We define the eye opening as the difference between the lowest mark and the highest space at the decision time. The eye-opening penalty is the ratio between the back-to-back and the PMD-distorted eye opening—after a compensator when one is used. When we study optimal compensators, we use this penalty as the feedback element. Thus, just like in real systems, the compensator and the receiver are not isolated from one another.

#### IV. STANDARD IMPORTANCE SAMPLING

We recall that the configuration space of the system is the high-dimensional space that defines all the possible choices of the random variables. In our system, the configuration space is defined by the  $3N$  variables,  $\xi_n$ ,  $\psi_n$ , and  $\phi_n$ , where  $1 \leq n \leq N$ . When  $N = 80$ , our most common choice, the configuration space has  $3 \times 80 = 240$  dimensions. The eye-opening penalty  $\Delta Q$  is a complicated highly nonlinear function of the  $3N$  variables that define the configuration space—particularly in

a system with a PMD compensator. We do not have enough *a priori* knowledge of the configuration to bias the penalty directly. Instead, we may take advantage of the strong correlation that exists between  $\Delta Q$  and the DGD and bias the DGD. As we will see, it is not sufficient to bias just the DGD in systems with compensators, and one must bias the first-order PMD as well. We also note that it is necessary to select several different biases to obtain sufficient coverage of the important regions of the configuration space. This technique is referred to as multiple importance sampling [22], [24]–[27], [32].

We define an indicator function  $I(\mathbf{x})$  on the configuration space  $\{\mathbf{x}\}$  as a function that equals 1 when  $\mathbf{x}$  is in a region of interest and equals 0 otherwise. Thus, for example,  $I(\mathbf{x})$  might equal 1 when  $\Delta Q > 1$  dB and equal 0 otherwise. The probability  $P_I$  that the indicator function is 1 may be estimated as [22], [24]–[27], [32]

$$\hat{P}_I = \sum_{j=1}^J \frac{1}{M_j} \sum_{i=1}^{M_j} I(\mathbf{x}_{ij}) w_j(\mathbf{x}_{ij}) L_j(\mathbf{x}_{ij}) \quad (6)$$

where  $L_j(\mathbf{x}_{ij}) = p(\mathbf{x}_{ij})/p_j^*(\mathbf{x}_{ij})$  is the likelihood ratio of the  $i$ th sample drawn from the  $j$ th biasing distribution, and where  $M_j$  is the number of samples drawn from the  $j$ th biasing distribution  $p_j^*$ . Hats indicate estimators. The quantity  $p(\mathbf{x})$  is the probability density function (pdf) of the unbiased distribution, and  $J$  is the number of different biasing distributions. The weights  $w_j(\mathbf{x}_{ij})$  allow one to combine different biasing distributions, and we choose them using the balanced heuristic method invented by Veach [42] and described in [22], [24]–[27], and [32]. A confidence interval for the estimator  $\hat{P}_I$  of the indicator function  $I(\mathbf{x})$  can be defined from the estimator of the variance of  $\hat{P}_I$ , which is given by

$$\hat{\sigma}_{\hat{P}_I}^2 = \sum_{j=1}^J \frac{1}{M_j(M_j-1)} \sum_{i=1}^{M_j} \left[ I(\mathbf{x}_{ij}) w_j(\mathbf{x}_{ij}) L_j(\mathbf{x}_{ij}) - \hat{P}_I \right]^2 \quad (7)$$

where

$$\hat{P}_{I_j} = \frac{1}{M_j} \sum_{i=1}^{M_j} I(\mathbf{x}_{ij}) w_j(\mathbf{x}_{ij}) L_j(\mathbf{x}_{ij}) \quad (8)$$

is the contribution of the samples drawn from the  $j$ th biasing distribution to the estimator  $\hat{P}_I$ . The confidence interval of the estimator  $\hat{P}_I$  equals the range  $[\hat{P}_I - \hat{\sigma}_{\hat{P}_I}, \hat{P}_I + \hat{\sigma}_{\hat{P}_I}]$ . The relative variation equals  $\hat{\sigma}_{\hat{P}_I}/\hat{P}_I$ .

To underline the importance of carefully monitoring statistical errors, we may compare studies in which we bias only the DGD  $\tau = |\tau|$  and in which we bias both the DGD and the second-order PMD  $|\tau_\omega|$ . The subscript  $\omega$  indicates the derivative with respect to  $\omega$ .

When biasing just the DGD, the appropriate parameters to bias are the angles  $\theta_n$  between the polarization dispersion vector in the first  $n$  sections and the polarization dispersion vector in the  $(n+1)$ th section at the center frequency such that  $\cos \theta_n$  is biased toward 1. In earlier work, we selected the  $\cos \theta_n$  from the pdf  $p_\alpha^*(\cos \theta) = (\alpha/2)[(1 + \cos \theta)/2]^{\alpha-1}$  [10], and

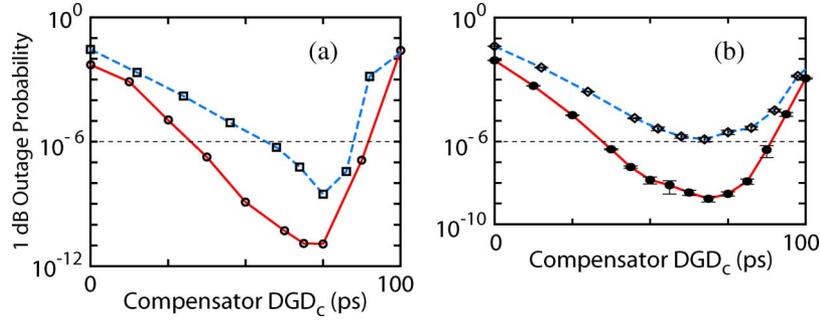


Fig. 3. Calculation of the 1-dB outage probability in a single-section fixed-DGD compensator, as a function of the DGD  $\tau_c$  of the fixed element. (a) Only  $|\tau|$  is biased. (b) Both  $|\tau|$  and  $|\tau_\omega|$  are biased. The value  $\tau_c = 0$  corresponds to the uncompensated case. The solid lines show the results when  $\langle\tau\rangle = 25$  ps, while the dashed lines show the results when  $\langle\tau\rangle = 30$  ps. There are no error bars in (a); the error bars in (b) show the confidence interval. The dashed line shows the  $10^{-6}$  outage probability level. Part (a) of this figure first appeared in [10], and part (b) first appeared in [22].

in later work, we selected the  $\cos \theta_n$  from the pdf  $p_\alpha^*(\cos \theta) = \alpha \{\exp[-\alpha(1 - \cos \theta)]\} / [1 - \exp(-2\alpha)]$  [22]. In earlier work, we used three different values of  $\alpha$ :  $\alpha = 1$  (unbiased),  $\alpha = 1.4$ , and  $\alpha = 1.9$ , with  $10^4$  samples each. We combined the results using a rejection algorithm described in [22]. In later work, we used  $\alpha = 0$  (unbiased),  $\alpha = 0.5$ , and  $\alpha = 1$ , with  $10^4$  samples each, and we used the balanced heuristic method to combine the results. The results are insensitive to these choices. In uncompensated systems, it has been shown that biasing the DGD alone produces highly accurate results with a limited number of samples [22], [24]–[27], [32].

Biasing the DGD alone is not sufficient in a system with compensators. It is not obvious how to bias both  $|\tau|$  and  $|\tau_\omega|$  together. Fortunately, Fogal *et al.* [43], [44] (see also [32]) have invented an effective algorithm for doing that, based on the calculus of variations. One biases the polarization dispersion vector in the  $n$ th section in a direction  $\hat{\mathbf{b}}$  that is located in a plane that contains  $\tau^{(n-1)}$  and  $\tau^{(n-1)} \times \tau_\omega^{(n-1)}$ , where  $\tau^{(n-1)}$  and  $\tau^{(n-1)} \times \tau_\omega^{(n-1)}$  are the polarization dispersion vector and its frequency derivative in the after  $n - 1$  fiber sections. One chooses  $\hat{\mathbf{b}}$  so that the expected angle  $\beta_n$  between  $\hat{\mathbf{b}}$  and  $\tau^{(n-1)}$  increases linearly up to some maximum  $\beta_N = \beta$ , so that  $\beta_n = (n/N)\beta$ . When  $\beta = 0$ , we are only biasing the DGD, and as  $\beta$  increases, we increasingly bias the second-order PMD. Some experimentation and on the order of ten combinations of  $\alpha$  and  $\beta$  are required to obtain good coverage of the  $|\tau| - |\tau_\omega|$  plane.

In Fig. 3, we show a comparison of the 1-dB outage probability in a system with a fixed-DGD compensator with optimal biasing that was simulated just biasing  $|\tau|$  and biasing both  $|\tau|$  and  $|\tau_\omega|$ . The key technical point that these results make is that the optimal choice of the compensator DGD is several times the mean DGD of the optical fiber transmission line. The results of Fig. 3(a) are sufficient to demonstrate the incorrectness of previous work, based on standard MC simulations, that purported to show that the optimal choice of the compensator DGD is close to the mean DGD. Nonetheless, the discrepancy by several orders of magnitude in the 1-dB outage probability at the minimum is clearly unacceptable. The reader should note that Fig. 3(a) does not have statistical error bars, while Fig. 3(b) does. The large errors in Fig. 3(a) are entirely statistical, and, indeed, it was the comparison of these two figures that convinced the author that statistical errors should *always* be presented whenever MC simulations are presented.

To understand what happened in this case, we now turn to Fig. 4. To generate this figure, we used the following procedure [25], [26]: We divided the  $|\tau| - |\tau_\omega|$  plane into  $25 \times 25 = 625$  evenly spaced bins and then estimated the average eye-opening penalty and its expected variance using the following estimators:

$$\hat{\mu} = \frac{\hat{D}}{\hat{C}}, \quad \hat{\sigma}_\mu^2 = \frac{\hat{\sigma}_D^2}{\hat{C}^2} + \frac{\hat{\mu}^2 \sigma_C^2}{\hat{C}^2} \quad (9)$$

where  $\hat{C} = \sum_{j=1}^J \hat{C}_j$  and  $\hat{D} = \sum_{j=1}^J \hat{D}_j$ , with

$$\hat{C}_j = \frac{1}{M_j} \sum_{I=1}^{M_j} I(\mathbf{x}_{ij}) w_j(\mathbf{x}_{ij}) L(\mathbf{x}_{ij})$$

$$\hat{D}_j = \frac{1}{M_j} \sum_{i=1}^{M_j} f(\mathbf{x}_{ij}) I(\mathbf{x}_{ij}) w_j(\mathbf{x}_{ij}) L(\mathbf{x}_{ij}). \quad (10)$$

The meaning of all quantities is the same as in (8), except that  $f(\mathbf{x}_{ij})$ , which was not defined previously, is the eye-opening penalty corresponding to the point in the configuration space  $\mathbf{x}_{ij}$ . In the work presented here, we have  $J = 10$ . Finally, we have

$$\hat{\sigma}_C^2 = \sum_{j=1}^J \frac{1}{M_j(M_j - 1)} \sum_{i=1}^{M_j} \left[ I(\mathbf{x}_{ij}) w(\mathbf{x}_{ij}) L_j(\mathbf{x}_{ij}) - \hat{C}_j \right]^2$$

$$\hat{\sigma}_D^2 = \sum_{j=1}^J \frac{1}{M_j(M_j - 1)} \sum_{i=1}^{M_j} \left[ f(\mathbf{x}_{ij}) I(\mathbf{x}_{ij}) w(\mathbf{x}_{ij}) L_j(\mathbf{x}_{ij}) - \hat{D}_j \right]^2. \quad (11)$$

The estimators for the mean eye-opening penalty  $\hat{\mu}$  and its variance  $\hat{\sigma}_\mu^2$  in (9) are biased. The bias in the estimators can be reduced by computing  $\hat{C}$  with a much larger number of samples than are used to compute  $\hat{D}$ . This approach is practical because the computational cost of generating the fiber samples that are required to calculate  $\hat{C}$  is significantly smaller than the cost to compute penalties after compensation, i.e.,  $\hat{D}$ , in cases where it is necessary to optimize the compensator. Here, we used  $10^7$  samples per biased simulation to calculate  $\hat{C}$  and  $10^5$  samples per biased simulation to calculate  $\hat{D}$ . The maximum value of

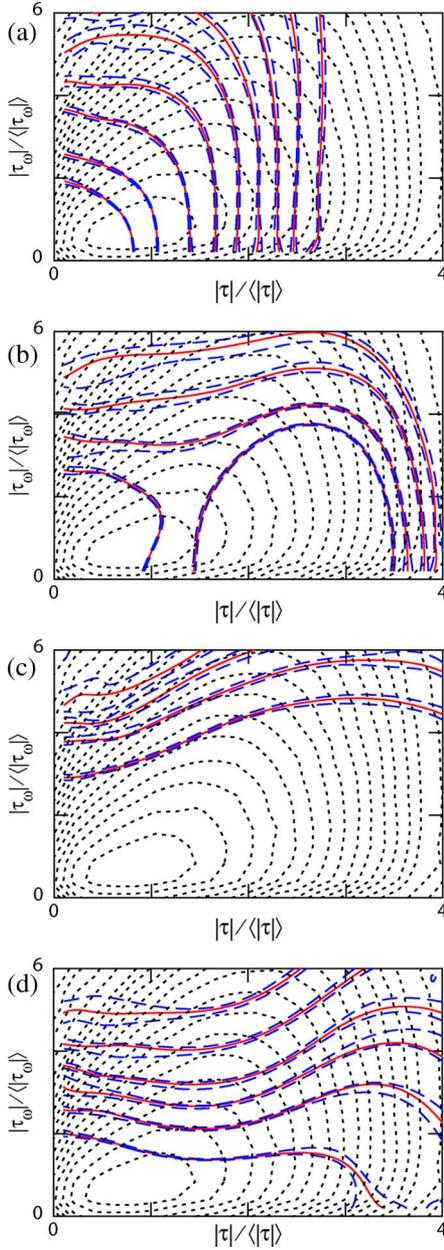


Fig. 4. Contour plots of the eye-opening plots and the associated confidence intervals, indicated by the solid lines and the dashed lines. The dotted lines show the contours of the joint pdf of the normalized DGD and second-order PMD. The contours are at  $3 \times 10^{-N}$ ,  $N = 1, \dots, 7$ , and at  $10^{-N}$ ,  $N = 1, \dots, 11$ . (a) Uncompensated system. The penalty curves are at 0.1, 0.2, 0.4, 0.6, 0.9, 1.2, 1.6, 2.2, and 3.2 dB. (b) Fixed-DGD compensator with  $\tau_c = 2.5\langle|\tau|\rangle$ . The penalty curves are at 0.1, 0.2, 0.3, and 0.4 dB. (c) Variable-DGD compensator with optimization of the eye opening. The penalty curves are at 0.1, 0.2, 0.3, and 0.4 dB. (d) Variable-DGD compensator with the DGD minimized at the central frequency of the signal: The penalty curves are at 0.1, 0.2, 0.3, 0.4, 0.6, and 0.9 dB. This figure first appeared as four separate figures in [26].

$\hat{\sigma}_C^2/\hat{C}^2$  is  $1.21 \times 10^{-4}$ . The maximum value for  $\hat{\sigma}_D^2/\hat{D}^2$  is 0.14, but it is much smaller in almost all bins, typically between 0.001 and 0.002. All penalty curves in Fig. 4 are smoothed using  $N$ -order Bezier smoothing, where  $N$  is the number of points in the curve.

The region in the  $|\tau| - |\tau_\omega|$  plane that dominates the penalty is where the corresponding penalty contour intersects the contour of the joint pdf of  $|\tau|$  and  $|\tau_\omega|$  with the highest prob-

ability. In the uncompensated system, the penalty curves are vertical where this intersection occurs, indicating that large-DGD regions of the configuration space dominate the penalty. By contrast, in all the compensated systems, except the system with a fixed-DGD compensator when the penalty is below 0.1 dB, this intersection occurs when the penalty curves are nearly horizontal, indicating that large  $|\tau_\omega|$  regions of the configuration space dominate the penalty. Thus, we must bias  $|\tau_\omega|$  as well as  $|\tau|$  to obtain accurate results. We note that we monitored the statistical error during the entire process of generating and analyzing Fig. 4.

While it is apparent that biasing  $|\tau|$  is insufficient, it is not at all apparent that biasing just  $|\tau|$  and  $|\tau_\omega|$  will be sufficient—particularly when compensators with several sections are used. Biasing higher orders or separately biasing the components of  $|\tau|$  and  $|\tau_\omega|$  would require us to adequately sample a high-dimensional projection of the original configuration space—higher than the two dimensions that we show in Fig. 4. Moreover, the procedure for jointly biasing  $|\tau|$  and  $|\tau_\omega|$  is nontrivial, and it is not at all clear how to jointly bias higher  $\omega$  derivatives as well. Using the MMC method, which allows us to directly bias the eye-opening penalty  $\Delta Q$ , sidesteps these difficulties, as we will show in the next section. At the same time, it makes monitoring the statistical errors considerably more difficult.

## V. MMC SIMULATIONS

The MMC is an iterative learning method that—in contrast to standard importance sampling—requires little *a priori* knowledge of where in the configuration space the samples of importance are located. For this reason, since its invention by Berg and Neuhaus [45]–[47] and its introduction into studies of optical communications systems by Yevick [48], it has been successfully applied in a variety of contexts in optical fiber communications [12], [13], [17]–[21], [27]. In the present context, it has two advantages. The first is that it allows us to directly bias the penalty, rather than the DGD, its first derivative with respect to frequency, and possibly even higher derivatives. The second is that it is based on different principles from standard importance sampling. As a consequence, comparison of simulations based on standard importance sampling and the multicanonical method allows us to mutually validate both approaches, giving us confidence that we are avoiding subtle algorithmic errors and—far more important in practice—do not have coding errors.

We begin by briefly reviewing our implementation of the MMC. More details may be found in [23], [25], and [27]. The basic approach that we use was first used by Berg and Neuhaus [45]–[47] for statistical physics applications and later by Yevick [48] to calculate DGD pdfs. In this method, one carries out  $J$  iterations of a Metropolis MC simulation, in which one uses the results of the previous iteration to rebias the next iteration. In a Metropolis MC calculation, one makes a random walk through the configuration space, rather than drawing the samples completely at random. Thus, in the present context, given a point in the configuration space, i.e.,  $\mathbf{x}_m = [\xi_i^{(m)}, \phi_i^{(m)}, \psi_i^{(m)}, i = 1, \dots, N]$ ,

we make a provisional transition to a new point  $\mathbf{x}_{m+1,p} = [\xi_{i,p}^{(m+1)}, \phi_{i,p}^{(m+1)}, \psi_{i,p}^{(m+1)}, i = 1, \dots, N]$  by adding a small random perturbation  $\Delta \mathbf{x} = [\Delta \xi_i, \Delta \phi_i, \Delta \psi_i, i = 1, \dots, N]$  to  $\mathbf{x}_m$ . In the first iteration,  $j = 1$ , we carry out a standard Metropolis random walk. To bias the simulation toward higher values of  $\Delta Q$ , in subsequent iterations, we accept the transition from  $\mathbf{x}_m$  to  $\mathbf{x}_{m+1,p}$  with a probability that is given by  $\min\{1, P^j[\Delta Q(\mathbf{x}_m)]/P^j[\Delta Q(\mathbf{x}_{m+1,p})]\}$ , where  $P^j[\Delta Q]$  is the estimate of the pdf of  $\Delta Q$  that we obtained in the previous iteration. If the step is accepted, then we set  $\mathbf{x}_{m+1} = \mathbf{x}_{m+1,p}$  and  $\Delta Q_{m+1} = \Delta Q(\mathbf{x}_{m+1,p})$ . If the step is rejected, then we set  $\mathbf{x}_{m+1} = \mathbf{x}_m$  and  $\Delta Q_{m+1} = \Delta Q_m$ . In either case, we update the appropriate bin of the histogram in which we are accumulating the samples of  $\Delta Q$  that we will use to estimate  $P^{j+1}[\Delta Q]$ . In each iteration, the random walk consists of  $M_j$  steps, where  $j$  does not have to be the same in every iteration. This random walk is a Markov process in the configuration space.

Writing the number of elements in the  $k$ th histogram bin  $k = 1, \dots, K$  of  $\Delta Q$  on the  $j$ th iteration as  $H_k^j$ , a reasonable estimate for  $P_k^{j+1}$ , which is the estimate of the pdf for the  $(j+1)$ th iteration corresponding to the  $k$ th histogram bin, would be

$$P_k = \frac{P_k^j H_k^j}{\sum_{k'=1}^K P_{k'}^j H_{k'}^j}. \quad (12)$$

However, this approach does not take advantage of any of the information that has been accumulated prior to the  $j$ th iteration and does not take advantage of the expected correlations from bin to bin. Berg and Neuhaus [45]–[47] have developed a better algorithm, in which we set

$$P_{k+1}^{j+1} = P_k^{j+1} \frac{P_{k+1}^j}{P_k^j} \left( \frac{H_{k+1}^j}{H_k^j} \right)^{\hat{g}_k^j} \quad (13)$$

where  $\hat{g}_k^j$ , which is the relative statistical significance of the  $k$ th bin in the  $j$ th iteration, is defined as

$$\hat{g}_k^j = \frac{g_k^j}{\sum_{j'=1}^j g_k^{j'}}, \quad \text{with} \quad g_k^j = \frac{H_{k+1}^j H_k^j}{H_{k+1}^j + H_k^j}. \quad (14)$$

They derive this algorithm, assuming that the pdf is exponentially distributed with an exponent that is a slowly varying function of the control quantity— $\Delta Q$  in our case. Yevick [49] has demonstrated the greater effectiveness of (13) relative to (12) in obtaining the pdf of the DGD.

Fig. 5 shows the 1-dB outage probability as a function of the eye-opening penalty for uncompensated systems with both 15 and 30 ps of mean DGD and for a system with a one-stage variable-DGD compensator that minimizes the DGD at the central frequency, in a system with a mean DGD of 30 ps. Both standard importance-sampled and MMC simulations are compared, and the agreement is excellent—serving to mutually validate the results for both methods. The agreement is an indication that biasing just the first- and second-order PMDs is sufficient for the standard importance-sampled simulations

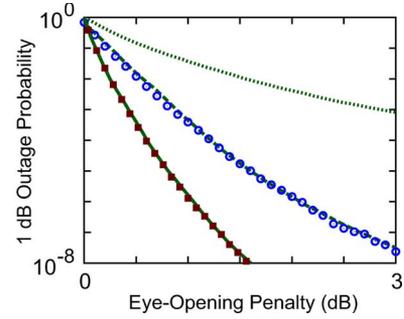


Fig. 5. Calculation of the 1-dB outage probability. Dotted line: Results for an uncompensated system with a mean DGD of 30 ps. Dashed line and open circles: Results for a system with a variable-DGD compensator, obtained using MMC and standard importance sampling, respectively, for a system with a mean DGD of 30 ps. Solid line and squares: Results for an uncompensated system with a mean DGD of 15 ps, obtained using MMC and standard importance sampling, respectively. This figure was obtained from [25].

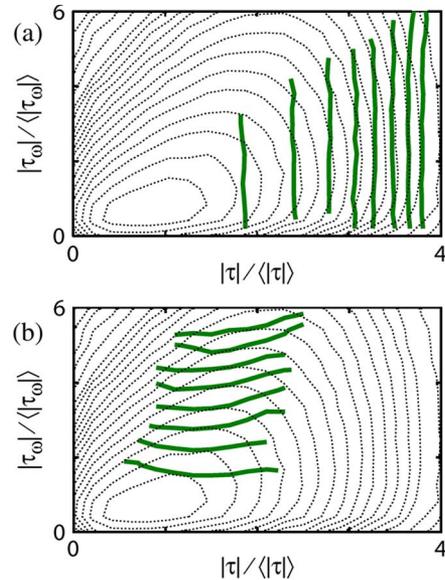


Fig. 6. Contour lines produced by the MMC method for (a) uncompensated system and (b) system with a variable-DGD compensator. The dotted lines show the contours of the joint pdf of normalized  $|\tau|$  and  $|\tau_\omega|$ . The contours are the same as in Fig. 4. The solid lines show the penalty contours. (a) Corresponding from left to right to 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, and 1.6 dB. (b) Corresponding from bottom to top to the same values. This figure was obtained from [25].

in this case and that neither algorithm has subtle algorithmic errors or coding errors. The MMC simulations reported here used 50 iterations with 2000 steps each. Increased insight into the working of the MMC method can be obtained from Fig. 6. In Fig. 6(a), we show the contours for the eye-opening penalty of an uncompensated system with a mean DGD of 15 ps. The penalty contours are only plotted where the MMC method produced samples. The method automatically placed its samples in the regions of the  $|\tau| - |\tau_\omega|$  plane that have the highest probability of occurrence and contribute the most to the penalty. In Fig. 6(b), we show similar results for the compensated system with a mean DGD of 30 ps. Again, the MMC method automatically placed its samples in the regions of the  $|\tau| - |\tau_\omega|$  plane that have the highest probability of occurrence.

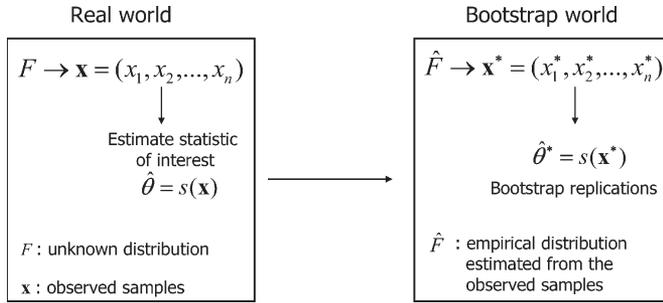


Fig. 7. Schematic illustration of the bootstrap procedure. Left: Operation of drawing a true realization from the actual unknown distribution  $F$ . Right: Same operation, applied to drawing bootstrap realizations. This figure was created by A. O. Lima.

The reader should note that no statistical errors were calculated for Figs. 5 and 6. At the time that these figures were produced, no effective procedures existed for calculating the statistical errors for MMC simulations. The highly nonlinear iterative procedure that is used in these simulations induces large correlations from iteration to iteration and bin to bin, so that analytical formulas analogous to (7) and (11) do not exist. An effective procedure for calculating the statistical error may be found in [23] and [25]. The method is a bootstrap resampling method [50], [51] that uses a computed estimate of the probability of a transition from bin  $k$  to bin  $l$  of the histogram of the control quantity; we refer to this method as the transition matrix method. In a bootstrap method, one estimates a complex statistical quantity by extracting a sequence of samples from an unknown distribution and computing the quantity. One then obtains an ensemble of sequences from the original sequence by repeatedly and independently drawing samples from the original sequence *with replacement* in exactly the same way that the original samples were drawn from the unknown distribution. Since each sequence is independent, one can then estimate the statistic of interest—in our case, the variance—using standard techniques. The bootstrap method is used when it is computationally far more rapid to resample the original sequence than it is to generate new sequences, allowing for an efficient estimate of the statistic of interest.

Efron’s bootstrap [50], [51] is a general-purpose technique for obtaining statistical estimates without making *a priori* assumptions about the distribution of the data. A schematic illustration of this procedure is shown in Fig. 7. Suppose one draws a random sequence that is represented as a random vector  $\mathbf{x} = (x_1, \dots, x_n)$  with  $n$  samples from an unknown pdf  $F$  and one wishes to estimate the standard deviation of a parameter of interest  $\hat{\theta} = s(\mathbf{x})$ . Since there is only one realization of the random vector, there is only realization of  $\hat{\theta}$ , and one cannot use the traditional standard deviation formula to compute the error. However, one can use the random vector  $\mathbf{x}$  to determine an empirical distribution  $\hat{F}$  from the unknown distribution  $F$ . One can then generate bootstrap samples from  $\hat{F}$ ,  $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ , to obtain  $\hat{\theta}^* = s(\mathbf{x}^*)$  by drawing  $n$  samples with replacement from  $\mathbf{x}$ . The star notation indicates that  $\mathbf{x}^*$  is not the actual random vector  $\mathbf{x}$ , but rather a resampled version of  $\mathbf{x}$  obtained from the estimated distribution  $\hat{F}$ . One then generates as many bootstrap realizations as one needs and

then generates independent bootstrap-sampled estimates of  $\hat{\theta}$ ,  $\hat{\theta}_1^* = s(\mathbf{x}_1^*), \dots, \hat{\theta}_B^* = s(\mathbf{x}_B^*)$ , where  $B$  is the total number of bootstrap realizations. Then, one can estimate the error in  $\hat{\theta}$  using the usual standard deviation formula on the bootstrap realizations.

The transition matrix method that we describe in this paper is related to the bootstrap resampling method as follows.

- 1) The pdf  $\hat{F}$  is an estimate of the transition matrix obtained from a single standard-MMC simulation.
- 2) The  $\mathbf{x}_1^*, \dots, \mathbf{x}_B^*$  are the collection of realizations obtained from the ensemble of pseudo-MMC simulations. We note that  $x_b^*$  should be computed using exactly the same number of iterations and exactly the same number of samples per iteration as in the original standard-MMC simulation.
- 3) Each  $\hat{\theta}_b^*$ , where  $b = 1, \dots, B$ , is a value for the probability of the  $k$ th bin of the histogram of the control quantity  $p_k^*$  obtained from each of the pseudo-MMC simulations.
- 4) Given that one has  $B$  independent realizations of  $p_k^*$ , one can obtain an estimate for the statistical error in each bin in the histogram of the pdf of the control quantity using the traditional standard deviation formula [50], [51], i.e.,

$$\sigma_{\hat{\theta}^*} = \left[ \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2 \right]^{1/2}, \text{ where } \bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*. \tag{15}$$

The transition matrix method has two parts. In the first part, we obtain an estimate of the pdf of the control quantity—in our case  $\Delta Q$  or the DGD—and an estimate of the one-step transition probability matrix  $\mathbf{\Pi}$ . To do so, we run a standard MMC simulation, as outlined earlier and described in more detail in [23] and [25]. At the same time, we compute an estimate of the transition probability  $\pi_{kl}$ , which is the probability that a sample in bin  $k$  will move to bin  $l$  after a single step in the MMC algorithm. We stress that a transition attempt must be recorded, whether or not it is accepted by the Metropolis algorithm, after the fiber undergoes a random perturbation. The transition matrix is a matrix that contains the probability that a transition will take place from one bin to any other bin when applying a random perturbation. It is independent of the procedure for rejecting or accepting samples, which is how the biasing is implemented in the MMC method. An estimate of the transition matrix that is statistically as accurate as the estimate of the pdf using MMC can be obtained by considering all the transitions that were attempted in the MMC ensemble. We use this information to build a  $K \times K$  one-step transition probability matrix, where  $K$  is the number of bins in the histogram of the pdf. The transition matrix  $\mathbf{\Pi}$  consists of elements  $\pi_{kl}$ , where the sum of the row elements of  $\mathbf{\Pi}$  equals 1. The elements  $\pi_{kl}$  are computed as

$$\pi_{kl} = \frac{\sum_{m=1}^{M_{\text{total}}-1} I_k(E_m) I_l(E_{m+1})}{\sum_{m=1}^{M_{\text{total}}-1} I_k(E_m)}, \text{ if } \sum_{m=1}^{M_{\text{total}}-1} I_k(E_m) \neq 0 \tag{16}$$

and  $\pi_{kl} = 0$ , otherwise. In (16),  $M_{\text{total}}$  is the total number of samples in the MMC simulation, and  $E_m$  is the value of the control quantity after  $m$  steps. The indicator function  $I_k(E)$  is chosen to compute the probability of having a sample of the control quantity inside bin  $k$  of the histogram. Thus, we have  $I_k(E) = 1$  inside the range of bin  $k$ , and otherwise,  $I_k(E) = 0$ .

In the second part of the procedure, we carry out a series of MMC simulations, which we refer to as pseudo-MMC simulations. In each step, if we start, for example, in bin  $k$  of the histogram, we pick a provisional new bin  $l$  using the probability  $\pi_{kl}$  in the transition matrix  $\Pi$ . We then accept or reject this provisional transition using the same criteria as in the full standard-MMC simulation, and the number of samples in the bins of the histogram is updated accordingly. It is possible to carry out hundreds of these pseudo-MMC simulations in a fraction of the computer time that it takes to carry out a single standard-MMC simulation in the problems that we studied.

To validate this procedure, we applied it to calculations of the DGD distributions of 15- and 80-section PMD emulators. It is computationally far more rapid to calculate the pdf of the DGD than it is to calculate the pdf of  $\Delta Q$  in a compensated system. Thus, it is feasible to check our approach by running many independent standard-MMC simulations and comparing the estimated statistical error from the transition matrix method to the estimated statistical error from a large number of independent standard-MMC simulations. We used 14 MMC iterations with 4000 samples each to compute the pdf of the normalized DGD when we used a 15-section emulator and 30 MMC iterations with 8000 samples each when we used an 80-section PMD emulator.

We monitor the accuracy of our computation by calculating the relative variation of the pdf of the normalized DGD,  $\hat{\sigma}_{\hat{P}_{\text{DGD}}}/\hat{P}_{\text{DGD}}$ . In Fig. 8, we show the relative variation for both 15- and 80-section PMD emulators. The symbols show the relative variation when we apply the transition matrix method with 1000 pseudo-MMC simulations based on a single standard-MMC simulation, while the solid and the dot-dashed lines show the results when we use 1000 standard-MMC simulations. As expected, the result from an ensemble of pseudo-MMC simulations shows a systematic deviation from the result of an ensemble of standard-MMC simulations for both emulators. The systematic deviation changes, depending on which standard-MMC simulation is used to generate the pseudo-ensemble.

In Fig. 8, the two dashed lines show the confidence interval of the relative variation with the 15-section PMD emulator computed using the transition matrix method, i.e., the confidence interval for the results that are shown with circles. While the relative variation that is computed using the transition matrix method from a single MMC simulation will vary from one standard-MMC simulation to another, the results obtained from different standard-MMC simulations are likely to be within this confidence interval with a well-defined probability. In the case shown here, where we had an ensemble of 1000 standard-MMC simulations available, we calculated this confidence interval using this ensemble. More generally, the number of standard-MMC simulations available will be small (typically only one), and one must estimate the confidence interval from the pseudo-

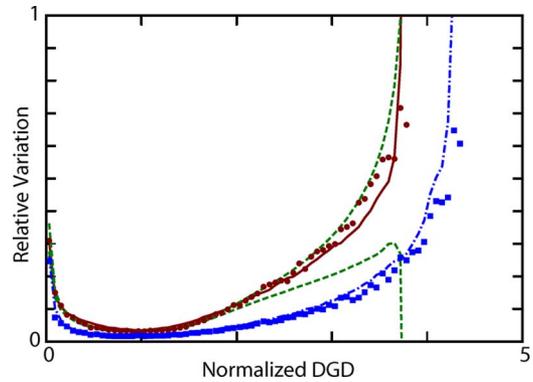


Fig. 8. Relative variation  $\hat{\sigma}_{\hat{P}_{\text{DGD}}}/\hat{P}_{\text{DGD}}$  of the pdf of the normalized DGD. Circles: Transition matrix method for the 15-section PMD emulator. Solid line: 1000 standard-MMC simulations for the 15-section PMD emulator. Dashed lines: Confidence interval for the 15-section PMD emulator. Squares: Transition matrix method for the 80-section PMD emulator. Dot-dashed line: 1000 standard-MMC simulations for the 80-section PMD emulator. This figure first appeared in [23].

MMC simulations through a process referred to as “bootstrapping the bootstrap” or the double-bootstrap procedure [23], [25], [50], [51]. We have verified that the confidence interval calculated using the double-bootstrap procedure on a single standard-MMC simulation agrees well with the confidence interval calculated using an ensemble of standard-MMC simulations in the cases considered here.

We observed excellent agreement between the results obtained with the transition matrix method based on a single standard-MMC simulation and the results obtained with 1000 standard standard-MMC simulations for both the 15- and 80-section PMD emulators when the relative error is smaller than 15%. For a larger relative error, the actual error is within the confidence interval predicted by the double-bootstrap procedure. The curves for the 80-section PMD emulator have a larger DGD range because a fiber with 80 birefringent sections is able to produce larger DGD values than is possible with a fiber with 15 birefringent sections [31].

At this point, we will mention parenthetically a potential difficulty that was brought to the author’s attention by Neal Radford [52]. The random walk that occurs on each iteration of the MMC algorithm is a Markov process in the configuration space. However, it does not follow that it is a Markov process in the space of the control quantity. The reason is that the map from the configuration space to the control quantity space is a many-to-one map, so that a manifold of points in the configuration space correspond to a single point in the control quantity space. As a consequence, multistep correlations can be induced by the random walk in the control quantity space. In our pseudo-MMC simulations, we treat the random walk in the control space as a Markov process, which implies that it may not have the same statistics as the original random walk in the configuration space. Whether this potential difficulty will lead in practice to errors in the estimates of the statistical errors or is merely one of those statistical bugaboos, which, like contributions from disjoint regions of the configuration space, seem to rarely be important in practice, is uncertain at this point. Certainly, this issue appeared to have no effect on

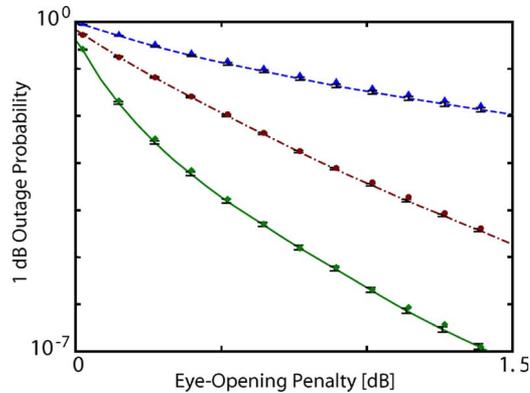


Fig. 9. One-decibel outage probability as a function of the eye-opening penalty for a system with a mean DGD of 30 ps. Dashed line (MMC) and triangles (standard importance sampling): Uncompensated system. Dot-dashed lines (MMC) and circles (standard importance sampling): System with a single-section compensator. System with a three-section compensator: Solid line (MMC) and diamonds (standard importance sampling). The error bars show the confidence interval for the MMC simulation. This figure first appeared in [27].

our estimate of the statistical error in emulators, which we directly verified. Moreover, in contrast to our calculation of the statistical quantities of interest, it is not really necessary to calculate the statistical error to high accuracy—just well enough to ensure that it is small relative to the quantity of interest. Nonetheless, it would have been useful to monitor the multistep correlations of the control quantity to ensure that they are small, and we intend to do so in future applications of this approach.

We now turn to applications of the MMC method to multisection compensators [25], [27]. In this case, it is possible to eliminate both first- and second-order PMDs at the central frequency of the signal. In Fig. 9, we show a comparison of the 1-dB outage probability as a function of the eye-opening penalty for a system with a mean DGD of 30 ps for an uncompensated system, a single-section compensator, and a three-section compensator. We used both standard importance sampling and the MMC method. The importance-sampled results were obtained by biasing both the DGD and the second-order PMD, as described in Section IV. We show with error bars the statistical error for the MMC method, calculated using the transition matrix method. The statistical error for the standard importance sampling was smaller in all cases, and the corresponding error bars are not shown. The agreement between the two biasing methods is excellent even for the three-section compensator, where one might expect standard importance sampling to fail since only the first two orders of PMD are biased. This agreement indicates that the three-section compensator produces third-order and higher orders of PMD after compensation that are strongly correlated to the first two orders of PMD before compensation. We have verified this inference with detailed studies [25], [27].

In conclusion, the comparison of standard importance sampling to the MMC method not only allowed us to mutually validate both calculations but also yielded insights that were not obtained from either method alone.

## VI. CONCLUSION

It is important to carefully monitor statistical errors when carrying out MC simulations. Failure to do so has led to errors in the past. Effective procedures for calculating the statistical errors in standard MC simulations have long been known and are easily implemented. Moreover, when each sample in the simulation is independently drawn, the statistical errors in each bin of a histogram will be independent, and the smoothness of the histogram is often a good indication of low statistical errors. While calculating the statistical errors with standard importance sampling is more complicated, analytical formulas exist and have been successfully implemented. By contrast, calculating the statistical errors using multicanonical methods is highly nontrivial. Recently, however, effective methods have been developed that are based on the calculation of a transition matrix with a standard multicanonical simulation and use of this transition matrix to draw a large number of independent samples. This method is closely related to Efron's bootstrap—a general-purpose statistical method.

At the same time, there are even more important sources of error that must be continually kept in mind. First is incorrect physical assumptions about the system being modeled. As the parameter regime being modeled changes, one can often make a transition from a regime where an assumption that was valid—or at least unimportant—becomes incorrect and an important source of errors. No amount of code validation will guard against them. Only careful and continuous checking against well-formulated experiments will eliminate these errors.

Second are coding errors. It is the author's experience that newcomers to simulation—particularly those with strong experimental backgrounds—are typically far too complacent about the possibility of coding errors in their computer codes. Oftentimes, they will validate their codes by comparison to analytical results in special limits or—worse yet—by comparison to experiments. Special analytical limits are by their very nature special, and codes that work in these limits often fail elsewhere. Using experiments to validate computer codes conflates the issue of determining whether a code's physical foundations are valid with the issue of removing coding errors. It is the author's view that whenever possible—and in the case of MC simulations it is always possible—a code should be validated by comparison to an independent simulation, preferably based on an independent algorithmic approach. Biasing MC simulations was first developed to solve problems that simply could not be solved by using standard MC methods. However, once developed, they also have the major advantage of allowing their users to validate their standard MC codes.

Finally, there are subtle algorithmic errors. While most algorithmic errors lead to unacceptably large statistical errors—indeed all practical cases that lie within the author's experience—it is at least theoretically possible to have subtle algorithmic errors without large statistical errors. Thus, it is important to not only monitor the statistical errors but also compare MC simulations that are based on different algorithmic approaches. In the author's view, there has been too much concern in the optical communications community with the possibility of subtle algorithmic errors in biasing MC simulations

and too little concern with the possibility of errors—particularly coding errors—in standard MC simulations. Thus, there has been a tendency when standard and biasing MC simulations are compared and do not agree to simply attribute the difference to subtle algorithmic errors in the biasing simulations and not carry out the full process of code validation. In these cases, an exaggerated fear of subtle algorithmic errors in biasing MC simulations has not only become a barrier to further progress but also become a barrier to the efficient development of reliable codes.

Good scientific practice should always be a matter of debate and will change over time. It is the author's view that the development of biasing MC methods, along with reliable methods for testing their statistical accuracy, has changed what should be viewed as "best practice" in carrying out MC simulations and reporting their results. First, any MC simulation should be carried out using two different codes, preferably based on different algorithms, with enough cases spread through the parameter regime of interest that one has thoroughly validated both codes. Second, when attributing the failure of an MC method to algorithmic errors, the author should be required to carefully document the source of the errors. Vague assertions that halt the process of code validation should not be acceptable. If it is understood that these proposals for best practice are offered to hopefully spark a debate and that they represent an ideal toward which the author and his colleagues have strived, while occasionally falling short, then the author will have achieved his goals.

#### ACKNOWLEDGMENT

The author would like to thank the Editors of this special issue for the invitation to write this paper and, in particular, M. Brodsky, who urged him along. The intellectual debt that this article owes to Dr. I. T. Lima, Jr., and Dr. A. O. Lima is profound. Almost all the technical work that appears here first appeared in articles that they first-authored, meeting presentations they first-authored, or their dissertations. Special thanks are due as well to the author's long-time collaborators, Dr. G. Biondini, Dr. W. Kath, Dr. B. Marks, and Dr. J. Zweck. Special thanks are due to R. Holzlöhner, who introduced the author to multicanonical methods. The opinions expressed by the author are, however, entirely his own.

#### REFERENCES

- [1] B. A. Cipra, "The best of the 20th century: Editors name top 10 algorithms," *SIAM News*, vol. 33, no. 4, 2000.
- [2] I. Beichl and F. Sullivan, "The Metropolis algorithm," *Comput. Sci. Eng.*, vol. 2, no. 1, pp. 65–69, Jan. 2000.
- [3] A. S. Householder, Eds., "Monte Carlo method," in *National Bureau of Standards Applied Mathematics Series*, ser. Series No. 12. Washington, DC: U.S. Gov. Printing Office, 1951.
- [4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.
- [5] G. S. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer-Verlag, 1996.
- [6] G. Biondini and W. L. Kath, "Tutorial: Calculating PMD statistics and outage probabilities with importance sampling," in *Proc. LEOS Summer Topical Meetings*, Vancouver, BC, Canada, Jul. 14–16, 2003.
- [7] N. Madras, "Lectures on Monte Carlo methods," in *Fields Institute Monographs*. Providence, RI: Amer. Math. Soc., 2002.
- [8] R. Srinivasan, *Importance Sampling*. New York: Springer-Verlag, 2002.
- [9] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, no. 3, pp. 310–312, Mar. 2002.
- [10] I. T. Lima, Jr., G. Biondini, B. S. Marks, W. L. Kath, and C. R. Menyuk, "Analysis of PMD compensators with fixed DGD using importance sampling," *IEEE Photon. Technol. Lett.*, vol. 14, no. 5, pp. 627–629, May 2002.
- [11] R. O. Moore, B. Biondini, and W. L. Kath, "Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems," *Opt. Lett.*, vol. 28, no. 2, pp. 105–107, Jan. 2003.
- [12] R. Holzlöhner and C. R. Menyuk, *Opt. Lett.*, vol. 28, no. 20, pp. 1894–1896, Oct. 2003.
- [13] W. Pellegrini, J. Zweck, C. R. Menyuk, and R. Holzlöhner, "Computation of bit error ratios for a dense WDM system using the noise covariance matrix and multicanonical Monte Carlo methods," *IEEE Photon. Technol. Lett.*, vol. 17, no. 8, pp. 1644–1646, Aug. 2005.
- [14] O. V. Sinkin, V. S. Grigoryan, R. Holzlöhner, A. Kalra, J. Zweck, and C. R. Menyuk, "Calculation of error probability in WDM RZ systems in presence of bit-pattern-dependent nonlinearity and of noise," presented at the Optical Fiber Commun. Conf., Anaheim, CA, Feb. 23–27, 2004, Paper TuN4.
- [15] O. V. Sinkin, V. S. Grigoryan, J. Zweck, and C. R. Menyuk, "Calculation of the bit-error ratio in wavelength-division-multiplexed return-to-zero systems when the nonlinear penalty is dominated by collision-induced timing jitter," presented at the Optical Fiber Commun. Conf., Anaheim, CA, Mar. 5–10, 2006, Paper JThB3.
- [16] 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 Photon. Technol. Lett.*, vol. 17, no. 5, pp. 1022–1024, May 2005.
- [17] Y. Yadin, M. Shtatif, and M. Orenstein, "Bit-error rate of optical DPSK in fiber systems by multicanonical Monte Carlo simulations," *IEEE Photon. Technol. Lett.*, vol. 17, no. 6, pp. 1355–1357, Jun. 2005.
- [18] T. Kamalakis, D. Varoutas, and T. Spicopoulos, "Statistical study of in-band crosstalk noise using the multicanonical Monte Carlo method," *IEEE Photon. Technol. Lett.*, vol. 16, no. 10, pp. 2242–2244, Oct. 2004.
- [19] I. Neokosmidis, T. Kamalakis, A. Chipouras, and T. Spicopoulos, "Estimation of the four-wave mixing noise probability-density function by the multicanonical Monte Carlo method," *Opt. Lett.*, vol. 30, no. 1, pp. 11–13, Jan. 2005.
- [20] A. Bilenca and G. Eisenstein, "Statistical noise properties of an optical pulse propagating in a nonlinear semiconductor amplifier," *IEEE J. Quantum Electron.*, vol. 41, no. 1, pp. 36–44, Jan. 2005.
- [21] —, "Fokker–Planck and Langevin analysis of noise accompanying the amplification of optical pulses in semiconductor optical amplifiers," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 22, no. 8, pp. 1632–1639, Aug. 2005.
- [22] I. T. Lima, Jr., A. O. Lima, G. Biondini, C. R. Menyuk, and W. L. Kath, "A comparative study of single-section polarization-mode dispersion compensators," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1023–1032, Apr. 2004.
- [23] A. O. Lima, I. T. Lima, Jr., and C. R. Menyuk, "Error estimation in multicanonical Monte Carlo simulations with applications to polarization-mode-dispersion emulators," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3781–3789, Nov. 2005.
- [24] I. T. Lima, Jr., "Investigation of the performance degradation due to polarization effects in optical fiber communication systems," Ph.D. dissertation, Univ. Maryland Baltimore County, Cantonville, MD, Dec. 2003.
- [25] A. O. Lima, "Advanced Monte Carlo methods for computation of penalties induced by polarization mode dispersion in optical fiber transmission systems," Ph.D. dissertation, Univ. Maryland Baltimore County, Cantonville, MD, Aug. 2005.
- [26] A. O. Lima, I. T. Lima, Jr., C. R. Menyuk, G. Biondini, B. S. Marks, and W. L. Kath, "Statistical analysis of the performance of PMD compensators using multiple importance sampling," *IEEE Photon. Technol. Lett.*, vol. 15, no. 12, pp. 1716–1718, Dec. 2003.
- [27] A. O. Lima, C. R. Menyuk, and I. T. Lima, Jr., "Comparison of two biasing Monte Carlo methods for calculating outage probabilities in systems with multisection PMD compensators," *IEEE Photon. Technol. Lett.*, vol. 17, no. 12, pp. 2580–2582, Dec. 2005.
- [28] P. K. A. Wai and C. R. Menyuk, "Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 14, no. 2, pp. 148–157, Feb. 1996.

- [29] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1746, Sep. 1997.
- [30] H. Goldstein, *Classical Mechanics*. Reading, MA: Addison-Wesley, 1980, ch. 4.
- [31] I. T. Lima, Jr., R. Khosravani, P. Ebrahimi, E. Ibragimov, A. E. Willner, and C. R. Menyuk, "Comparison of polarization mode dispersion emulators," *J. Lightw. Technol.*, vol. 19, no. 12, pp. 1872–1881, Dec. 2001.
- [32] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization-mode dispersion: Techniques and applications," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1201–1215, Apr. 2004.
- [33] A. Galtarossa and L. Palmieri, "Relationship between pulse broadening due to polarisation mode dispersion and differential group delay in long singlemode fibres," *Electron. Lett.*, vol. 34, no. 5, pp. 492–493, Mar. 1998.
- [34] H. Kogelnik, R. M. Jopson, and L. E. Nelson, "Polarization-mode dispersion," in *Optical Fiber Telecommunications*, vol. IVB, I. Kaminow and T. Li, Eds. San Diego, CA: Academic, 2002, ch. 15, pp. 725–861. See in particular, Sec. 7.4.
- [35] H. Bülow and S. Lanne, "PMD compensation techniques," in *Polarization Mode Dispersion*, A. Galtarossa and C. R. Menyuk, Eds. New York: Springer-Verlag, 2005, pp. 225–245.
- [36] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proc. Nat. Acad. Sci.*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.
- [37] I. T. Lima, Jr., A. O. Lima, Y. Sun, H. Jiao, J. Zweck, C. R. Menyuk, and G. M. Carter, "A receiver model for optical fiber communication systems with arbitrarily polarized noise," *J. Lightw. Technol.*, vol. 23, no. 3, pp. 1478–1490, Mar. 2005.
- [38] D. Marcuse, "Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 8, no. 12, pp. 1816–1823, Dec. 1990.
- [39] P. A. Humblet and M. Azizoglu, "On the bit error rate of lightwave systems with optical amplifiers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1576–1582, Nov. 1991.
- [40] S. D. Personick, "Receiver design," in *Optical Fiber Telecommunications*, S. E. Miller and A. G. Chynoweth, Eds. New York: Academic, 1979, ch. 19, pp. 627–651.
- [41] P. R. Trischitta and E. L. Varma, *Jitter in Digital Transmission Systems*. Boston, MA: Artech House, 1989.
- [42] E. Veach, "Robust Monte Carlo methods for light transport simulation," Ph.D. dissertation, Stanford Univ., Stanford, CA, Dec. 1997.
- [43] S. L. Fogal, G. Biondini, and W. L. Kath, "Multiple importance sampling for first- and second-order polarization-mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, no. 9, pp. 1273–1275, Sep. 2002.
- [44] —, "Correction to 'Multiple importance sampling for first- and second-order polarization-mode dispersion'," *IEEE Photon. Technol. Lett.*, vol. 14, no. 10, p. 1487, Oct. 2002.
- [45] B. A. Berg and T. Neuhaus, "The multicanonical ensemble: A new approach to simulate first-order phase transition," *Phys. Rev. Lett.*, vol. 68, no. 1, pp. 9–12, Jan. 1992.
- [46] B. A. Berg, "Algorithmic aspects of multicanonical simulations," *Nuclear Phys. B, Proc. Suppl.*, vol. 63A-C, no. 1–3, pp. 982–984, Apr. 1998.
- [47] —, "Multicanonical simulations step by step," *Comput. Phys. Commun.*, vol. 153, no. 3, pp. 397–406, Jul. 2003.
- [48] D. Yevick, "Multicanonical communication system modeling—Application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, no. 11, pp. 1512–1514, Nov. 2002.
- [49] —, "The accuracy of multicanonical system models," *IEEE Photon. Technol. Lett.*, vol. 15, no. 2, pp. 224–226, Feb. 2003.
- [50] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: SIAM, 1982.
- [51] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman and Hall, 1993.
- [52] N. M. Radford, Univ. Toronto, Toronto, ON, Canada, private communication.



**Curtis R. Menyuk** (SM'88–F'98) was born on March 26, 1954. He received the B.S. and M.S. degrees from the Massachusetts Institute of Technology, Cambridge, in 1976 and the Ph.D. degree from the University of California, Los Angeles, in 1981, all in physics.

He was a Research Associate at the University of Maryland, College Park, and at Science Applications International Corporation, McLean, VA. In 1986, he joined the University of Maryland Baltimore County (UMBC) as an Associate Professor in the Department of Electrical Engineering, in which he was a Founding Member. In 1993, he was promoted to Presidential Research Professor and was on partial leave from Fall 1996 to Fall 2002. From 1996 to 2001, he worked part-time for the Department of Defense (DoD), where he codirected the Optical Networking program at the DoD Laboratory for Telecommunications Sciences, Adelphi, MD, from 1999 to 2001. From 2001 to 2002, he was a Chief Scientist at PhotonEx Corporation. He has authored or coauthored more than 200 archival journal publications as well as numerous other publications and presentations and has edited three books. The equations and algorithms that he and his research group at UMBC have developed to model optical fiber systems are used extensively in the telecommunications and photonics industry. For the last 18 years, his primary research interests have been theoretical and computational studies of lasers, nonlinear optics, and fiber optic communications.

Dr. Menyuk is a member of the Society for Industrial and Applied Mathematics and the American Physical Society. He is also a Fellow of the Optical Society of America.