Importance Sampling for Polarization-Mode Dispersion: Techniques and Applications

Gino Biondini, William L. Kath, *Member, OSA*, and Curtis R. Menyuk, *Fellow, IEEE, Fellow, OSA*

Invited Paper

***Abstract*—The basic theory of importance sampling (IS) as relevant to polarization-mode dispersion (PMD) in optical fibers is discussed, and its application to Monte Carlo (MC) simulations of PMD-induced transmission impairments is demonstrated. The use of IS allows rare PMD events to be simulated much more efficiently than with standard MC methods. As a consequence, methods employing IS provide natural and effective tools to assess PMD-induced impairments and outages in optical transmission systems at realistic probability levels.**

***Index Terms*—Importance sampling (IS), Monte Carlo (MC) simulations, optical fiber communications, polarization-mode dispersion (PMD).**

I. Introduction

POLARIZATION-MODE dispersion (PMD) is a potentially major impairment in high-bit-rate terrestrial and undersea wavelength-division-multiplexed (WDM) systems. In its simplest manifestation, PMD splits a pulse between the fast and the slow axes in an optical fiber; at the same time, higher orders of PMD induce depolarization and polarization-dependent chromatic dispersion (PCD). At high bit rates, all of these effects can lead to unacceptable transmission penalties. One of the main difficulties with PMD is that it is a stochastic phenomenon, and, as a consequence, the penalties it produces change randomly over distance and time as the ambient temperature and other environmental parameters vary. To account for random fluctuations, a maximum power penalty (typically 1 dB) is usually assigned to PMD, and one demands that the outage probability (i.e., the probability of the PMD-induced penalty exceeding this allowed value) is very small—typically 1 min per year, that is, 10^{-6} or less. Because of this stringent requirement, it has been difficult to use either Monte Carlo (MC) simulations or laboratory measurements to fully assess system outage probabilities due to the extremely large number of PMD configurations that are necessary to obtain reliable estimates.

One measure of PMD is the differential group delay (DGD) [1]–[3], which quantifies the splitting between the fast and slow axes. In optical fibers, the DGD has a Maxwellian probability distribution. The rare events where the DGD is significantly larger than its mean are particularly important, since they are the ones most likely to result in system outages. In the absence of effective tools for calculating outage probabilities, system designers have resorted to stopgap techniques. One such technique is to artificially produce specific values of DGD, determine the penalties at these values, and then weight the results using the Maxwellian distribution. A fundamental problem with this method, however, is that there is not necessarily a direct relationship between the DGD and the power penalty. In particular, the penalty generally depends also on higher order PMD. Thus, different fiber configurations can give the same DGD but not contribute equally to the penalty.

The technique of *importance sampling* (IS), which is one member of a family of methods known as variance reduction techniques, addresses these difficulties. The technique allows low-probability events to be efficiently simulated by concentrating MC simulations in the most significant regions of interest in sample space. The method is well known and has been applied with success in other contexts [4]–[10]. Recently, we have developed IS techniques for the study of PMD-induced effects generated by a concatenation of birefringent sections [11]–[13]. In this context, IS provides a tool that can be used in numerical simulations—and, in principle, in experiments—to accurately estimate PMD-induced system penalties. In the simplest case, the regions of interest in the sample space are the configurations that lead to large values of DGD and/or second-order PMD (SOPMD). This method has been recently applied for numerically calculating PMD-induced transmission penalties [14], [15]. Here, we give a detailed description of the technique and its applications. It should be noted that other variance reduction techniques have also recently been proposed, such as the multi-canonical MC method [16], [17] and the Brownian-bridge method [18]. We also note that a limited form of IS using variance scaling has recently been implemented in a hardware PMD emulator [19].

The structure of this paper is as follows. Section II provides a basic introduction to IS in a general setting, including the technique of multiple IS. Section III discusses the application of

Manuscript received July 5, 2003; revised November 14, 2003. This work was supported in part by the National Science Foundation under Grant Number DMS-0 101 476.

G. Biondini is with the Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208 USA, and also with the Department of Mathematics, Ohio State University, Columbus, OH 43210 USA.

W. L. Kath is with the Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208 USA (e-mail: kath@northwestern.edu).

C. R. Menyuk is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21228 USA.

Digital Object Identifier 10.1109/JLT.2004.825321

IS to numerical simulations of transmission effects where large DGD values are the primary cause of penalties. Because PMD is frequency dependent, however, the DGD at any single frequency is not sufficient to uniquely describe PMD effects, and in many practical cases, higher orders of PMD need to be considered. Thus, in Section IV, an IS technique that employs multiple IS to generate arbitrary combinations of first- and second-order PMD is discussed. Finally, in Section V, the methods are demonstrated by calculating the distribution functions of various quantities of interest for a concatenation of a finite number of birefringent sections, and the application of the same techniques to the direct numerical estimation of PMD-induced outage probabilities are presented.

II. Importance Sampling: Overview

IS has been extensively discussed in the literature [4]–[10], and we refer the reader to the references for further details. Here, we only review the basic aspects of the theory that are necessary for PMD emulation.

A. Rare Events and MC Simulations

Consider a quantity X which depends upon a random vector $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots \theta_N)$. In our case, the components θ_n will be the angles describing the relative orientations of the sections of a PMD emulator, but obviously this framework can be used more generally. The random variable X could be any generic quantity such as the DGD, a specific combination of first- and second-order PMD, the amount of pulse broadening, or any combination thereof. Suppose we are interested in calculating the probability P that $X(\boldsymbol{\theta})$ falls in a given range R. (For example, P could be the probability that the power penalty is larger than 1 dB, the outage probability, or any other probability of interest.) The probability P can be represented as the expectation value of an indicator function $I_R(X(\boldsymbol{\theta}))$, where $I_R(X) = 1$ if X falls in the prescribed range R and $I_R(X) = 0$ otherwise. That is, P is represented by the N-dimensional integral

$$P = \int_{\Theta} I_R(X(\boldsymbol{\theta})) p_\theta(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \mathsf{E}[I(X)] \tag{1}$$

where $p_\theta(\boldsymbol{\theta})$ is the joint probability density function (pdf) for the RVs, $\mathsf{E}[\cdot]$ denotes the expectation value with respect to $p(\boldsymbol{\theta})$, and Θ represents the entire configuration space. (In the following, we will omit the subscripts θ and R unless they are needed to avoid ambiguities.)

In many cases of interest, the variable X depends in a highly nontrivial way upon the RVs $\boldsymbol{\theta}$, thus making a direct calculation of the integral in (1) impossible. Thus, one usually resorts to MC simulations and writes an estimator $\hat{P}$ for the probability P by replacing the integral in (1) by

$$\hat{P} = \frac{1}{M} \sum_{m=1}^{M} I(X(\boldsymbol{\theta}_m)) \tag{2}$$

where M is the total number of samples and where the samples $\boldsymbol{\theta}_m$ are drawn according to the distribution $p(\boldsymbol{\theta})$ [4]–[10]. Equation (2) simply expresses the relative number of samples falling in the range of interest. By design, $\mathsf{E}[\hat{P}] = P$. In practice, however, many MC samples may be necessary to obtain a good estimate. In particular, if one is interested in low-probability events (that is, if $P \ll 1$), an impractical number of samples is necessary in order to see even a single event, and an even larger number is required to obtain an accurate estimate.

B. Importance-Sampled MC Simulations

When the main contribution to the desired probability comes from regions of the sample space Θ where $p_\theta(\boldsymbol{\theta})$ is small, IS can be used to speed up simulations. Here, the idea is to first rewrite the probability P in (1) as

$$\begin{aligned} P &= \int_{\Theta} I(X(\boldsymbol{\theta})) \frac{p_\theta(\boldsymbol{\theta})}{p^*(\boldsymbol{\theta})} p^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \mathsf{E}^*[I(X(\boldsymbol{\theta}))L(\boldsymbol{\theta})] \end{aligned} \tag{3}$$

where $\mathsf{E}^*[\cdot]$ denotes the expectation value with respect to $p^*(\boldsymbol{\theta})$, and where

$$L(\boldsymbol{\theta}) = p_\theta(\boldsymbol{\theta}) / p^*(\boldsymbol{\theta}) \tag{4}$$

is called the *likelihood ratio*. The density $p^*(\boldsymbol{\theta})$ is usually called the *biasing distribution*. As before, we then estimate the corresponding integral through MC simulations; that is, we write an importance-sampled MC estimate for P as

$$\hat{P}^* = \frac{1}{M} \sum_{m=1}^{M} I(X(\boldsymbol{\theta}_m^*)) L(\boldsymbol{\theta}_m^*) \tag{5}$$

where now the samples $\boldsymbol{\theta}_m^*$ are drawn according to $p^*(\boldsymbol{\theta})$. Using a biased probability distribution allows the desired regions of sample space to be visited much more frequently. At the same time, the likelihood ratio $L(\boldsymbol{\theta})$ automatically adjusts the results so that all of the different realizations are correctly weighted, thus contributing properly to the final estimate.

Not all biasing schemes are equivalent, of course. First of all, in order to obtain an unbiased estimator for P, the biased probability distribution $p^*(\boldsymbol{\theta})$ should generate with nonzero probability all of the configurations that contribute to the result P (i.e., all the samples $\boldsymbol{\theta}$ such that $X(\boldsymbol{\theta}) \in R$). One way to assess the efficiency of an MC estimator is to compute its variance, which determines the number of samples that are necessary on average to obtain a desired level of accuracy. If the samples $\boldsymbol{\theta}_m$ are statistically independent, the variance $\sigma_{\hat{P}}^2$ of $\hat{P}$ in (2) is $\sigma_{\hat{P}}^2 = \sigma_I^2/M$, where the variance σ_I^2 is

$$\sigma_I^2 = \mathsf{E}[I^2] - \mathsf{E}[I]^2 = P - P^2 \tag{6}$$

since $I(X)^2 = I(X)$. Thus, if P represents a rare event (that is, if $P \ll 1$), then $\sigma_I \sim \sqrt{P}$, and the coefficient of variation (or relative variance) of the MC estimator will be proportional to $\sigma_{\hat{P}}/P \sim 1/\sqrt{MP}$. In particular, if P is small (as in the case of the bit-error ratio or the outage probability), an exceedingly large number of unbiased MC samples are necessary to obtain a reliable estimate, i.e., one whose relative variance is a fraction smaller than 1.

If (5) is used instead, we have $\sigma_{\hat{P}^*}^2 = \sigma_{I^*}^2/M$, where the (biased) variance of $I^* = IL$, defined as in (3), is

$$\begin{aligned} \sigma_{I^*}^2 &= \mathsf{E}^*[I^2L^2] - \mathsf{E}^*[I\,L]^2 \\ &= \int_{\Theta} I(X(\boldsymbol{\theta}))L(\boldsymbol{\theta}) p_\theta(\boldsymbol{\theta})\, d\boldsymbol{\theta} - P^2. \end{aligned} \tag{7}$$

With appropriate choices of biasing distribution, this variance can be made much smaller than its unbiased counterpart. This can be seen by observing that (7) is that same as (6), except for the addition of the factor $L(\boldsymbol{\theta})$ in the first term; if $p^*(\boldsymbol{\theta})$ is large where $p_\theta(\boldsymbol{\theta})$ is small, then $L(\boldsymbol{\theta})$ will be small, and the first term on the right-hand side (RHS) of (7) will be smaller than the equivalent term in (6); thus, the variance will be reduced. Indeed, the use of IS results in substantial speed-ups of numerical simulations precisely because it is highly effective at reducing the variance of estimators.

An importance-sampled MC estimate of $\sigma^2_{I^*}$ in (7) is

$$\hat{\sigma}^2_{I^*} = \frac{1}{M-1}\sum_{m=1}^{M}(I(X(\boldsymbol{\theta}^*_m))L(\boldsymbol{\theta}^*_m) - \hat{P}^*)^2. \tag{8}$$

It is straightforward to verify that $\mathsf{E}^*[\hat{\sigma}^2_{I^*}] = \sigma^2_{I^*}$ so that this sample variance is unbiased (in the statistical sense). From this, we then immediately have the result for the sample variance in the estimator $\hat{P}^*$, namely

$$\hat{\sigma}^2_{\hat{P}^*} = \frac{1}{M}\hat{\sigma}^2_{I^*}. \tag{9}$$

It should be noted that direct use of (9) in numerical simulations requires that the user store all the individual samples until the estimator $\hat{P}^*$ is calculated. If this is not possible, a recursion relation for $\hat{P}^*$ and $\hat{\sigma}^2_{\hat{P}^*}$ can be employed, as follows:

$$\hat{P}_m = \frac{m-1}{m}\hat{P}_{m-1} + \frac{1}{m}I(X(\boldsymbol{\theta}^*_m))L(\boldsymbol{\theta}^*_m) \tag{10a}$$

$$\hat{S}_m = \hat{S}_{m-1} + \frac{m-1}{m}(I(X(\boldsymbol{\theta}^*_m))L(\boldsymbol{\theta}^*_m) - \hat{P}_{m-1})^2 \tag{10b}$$

with $\hat{P}^* \equiv \hat{P}_M$, and where

$$\hat{\sigma}^2_{\hat{P}^*} = \frac{1}{M(M-1)}\hat{S}_M. \tag{11}$$

C. Asymptotic Efficiency

The main difficulty with IS is the choice of biasing distribution, which requires some knowledge of the system. From (7), it is easy to see that the best choice

$$p^*_{\text{opt}}(\boldsymbol{\theta}) = \frac{1}{P}I(X(\boldsymbol{\theta}))p(\boldsymbol{\theta}) \tag{12}$$

yields exactly zero variance. Of course, this choice is not practical, since it requires the knowledge of the parameter being estimated, P. It does, however, offer some insight into what a good biasing distribution should look like: (12) generates only configurations that do contribute to the desired event (that is, only those for which $I(X) = 1$), and it selects them with a relative weight that is proportional to the original density. The rationale is that, of all the configurations that do lead to the desired event, the most important ones are those for which the original density is largest. This illustrates a general principle: in order for IS to be effective, *the biasing distribution should preferentially select the most likely configurations that lead to the desired event*. In other words, if the events sought are the transmission errors, the biasing distribution should "encourage typical errors" [10].

Techniques have been proposed for choosing biasing distributions [4]–[10], but finding good distributions in any given situation still remains a nontrivial task, and it represents the most difficult step when applying IS. The problem is twofold: 1) to understand which regions of sample space are most important, namely, find the most likely configurations that result in the desired events; and 2) to devise an efficient way to bias the simulations toward these regions. This cannot be done blindly; a bad approach can result in a less efficient method and, in extreme cases, an importance-sampled estimator with a variance that is even larger than the unbiased case [20]. One biasing choice that is at first glance promising but often fails is variance scaling, i.e., the idea of producing rare events merely by increasing the variance of the sample distribution. This may push random samples toward events of interest, but if the number of dimensions of the configuration space is large, the number of ways of getting unimportant events generally increases even more rapidly.[1] Variance scaling has been used to guide sampling in a hardware PMD emulator [19]; this type of device is therefore limited to a small number of degrees of freedom (i.e., a small number of adjustable birefringent elements).

Often, for simplicity, the choice of biasing distribution is restricted to a specific family of distributions, usually dependent on one or more parameters. In this case, we refer to the choice of parameters that yields the lowest variance as the optimal one. A more useful concept, however, is that of *asymptotic efficiency* [20], [21], which is formulated in the framework of large deviations theory [22]. Consider a set of probabilities P_n dependent on a parameter n; e.g., the probability that the random variable $X(\boldsymbol{\theta})$ takes values that are larger than n times its mean, or, in systems affected by PMD, the probability that the power penalty exceeds n dB. Furthermore, suppose that the probabilities P_n decay exponentially as n increases. In this case, it is reasonable to expect that the computational cost required for an accurate estimation of P_n will grow with n. Next, consider a sequence of biasing distributions $p^*_n(\boldsymbol{\theta})$. Roughly speaking, the sequence is said to be asymptotically efficient if the computational burden grows less than exponentially fast.

In general, it is quite difficult to show rigorously that a method is asymptotically efficient [20]. It is important to realize, however, that optimal biasing is limited to a predetermined class of biasing distributions, and therefore it only provides a relative form of comparison. Asymptotic efficiency, on the other hand, is concerned with the actual computational cost associated to the importance-sampled estimator. It is possible for a biasing distribution to be optimal within a subclass but not asymptotically efficient (see Section IV-D for examples).

D. Multiple IS

In many practical cases, including all the PMD examples to be discussed in the following sections, no single choice of biasing distribution can efficiently capture all the regions of sample space that give rise to the events of interest. In these

[1]This is often referred to as the dimensionality problem of variance scaling [9]. Another situation where IS can fail arises when there are two significant but disjoint regions in sample space, and the biasing directs the samples only toward one of them. In this case, it is possible in principle to obtain incorrect results even when the sample variance is low.

cases, it is necessary to use IS with more than one biasing distribution. The simultaneous use of different biasing methods is called *multiple IS* [23], [24]. When using several biasing distributions $p_j^*(\boldsymbol{\theta})$, a difficulty arises about how to correctly weight the results coming from different distributions. One solution to this problem can be found by assigning a weight $w_j(\boldsymbol{\theta})$ to each distribution and by rewriting the probability P as

$$P = \sum_{j=1}^{J} P_j = \sum_{j=1}^{J} \int_{\Theta} w_j(\boldsymbol{\theta}) I(X(\boldsymbol{\theta})) L_j(\boldsymbol{\theta}) p_j^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (13)$$

where J is the number of different biasing distributions used and $L_j(\boldsymbol{\theta}) = p(\boldsymbol{\theta})/p_j^*(\boldsymbol{\theta})$ is the likelihood ratio for the jth distribution. Note that the weights $w_j(\boldsymbol{\theta})$ depend on the value of the random variables for each individual sample. From (13), a multiply-importance-sampled MC estimator for P can now be written as

$$\hat{P} = \sum_{j=1}^{J} \hat{P}_j = \sum_{j=1}^{J} \frac{1}{M_j} \sum_{m=1}^{M_j} w_j(\boldsymbol{\theta}_{j,m}) I(X(\boldsymbol{\theta}_{j,m})) L_j(\boldsymbol{\theta}_{j,m}) \quad (14)$$

where M_j is the number of samples drawn from the jth distribution $p_j^*(\boldsymbol{\theta})$, and $\boldsymbol{\theta}_{j,m}$ is the mth such sample.

Several ways exist to choose the weights $w_j(\boldsymbol{\theta})$. Generally, $\hat{P}$ is an unbiased estimator for P (i.e., the expectation value of $\hat{P}$ is equal to P) for any choice of weights such that $\sum_{j=1}^{J} w_j(\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$ [23], [24]. Thus, each choice of weights corresponds to a different way of partitioning of the total probability. The simplest possibility is just to set $w_j(\boldsymbol{\theta}) = 1/J$ for all $\boldsymbol{\theta}$, meaning that each distribution is assigned an equal weight in all regions of sample space. This choice is not advantageous, however, as will be seen shortly.

If $\hat{P}$ is a multiply-importance-sampled MC estimator defined according to (14), then, similarly to (9), one can show that an unbiased estimator of its variance is

$$\hat{\sigma}_{\hat{P}}^2 = \sum_{j=1}^{J} \frac{1}{M_j(M_j-1)} \times \sum_{m=1}^{M_j} \left(w_j(\boldsymbol{\theta}_{j,m}) L_j(\boldsymbol{\theta}_{j,m}) I(X(\boldsymbol{\theta}_{j,m})) - \hat{P}_j \right)^2. \quad (15)$$

Similarly to (10), recursion relations can also be written so that $\hat{\sigma}^2$ can be obtained without the need of storing all the individual samples until the end of the simulation, as follows:

$$\hat{\sigma}_{\hat{P}}^2 = \sum_{j=1}^{J} \frac{1}{M_j(M_j-1)} \hat{S}_{j,M_j} \quad (16)$$

with $\hat{P} = \sum_{j=1}^{J} \hat{P}_{j,M_j}$ and

$$\hat{P}_{j,m} = \frac{m-1}{m} \hat{P}_{j,m-1} + \frac{1}{m} w_j(\boldsymbol{\theta}_{j,m}) L_j(\boldsymbol{\theta}_{j,m}) I(X(\boldsymbol{\theta}_{j,m})) \quad (17a)$$

$$\hat{S}_{j,m} = \hat{S}_{j,m-1} + \frac{m-1}{m} \left(w_j(\boldsymbol{\theta}_{j,m}) \times L_j(\boldsymbol{\theta}_{j,m}) I(X(\boldsymbol{\theta}_{j,m})) - \hat{P}_{j,m-1} \right)^2. \quad (17b)$$

E. The Balance Heuristic

When using multiple IS, the choice of weights $w_j(\boldsymbol{\theta})$ is almost as important as the choice of biasing distributions $p_j(\boldsymbol{\theta})$. Different weighting functions result in different values for the variance of the combined estimator. A poor choice of weights can result in a large variance, thus partially negating the gains obtained by IS. The best weighting strategies are obviously the ones that yield the smallest value.

For example, consider the case where the weighting functions are constant over the whole domain Θ. In this case

$$P = \sum_{j=1}^{J} w_j \int_{\Theta} I(X(\boldsymbol{\theta})) L_j(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sum_{j=1}^{J} w_j \mathsf{E}_j^*[I(X(\boldsymbol{\theta})) L_j(\boldsymbol{\theta})]. \quad (18)$$

That is, the estimator is simply a weighted combination of the estimators obtained by using each of the biasing techniques. Unfortunately, the variance of P is also a weighted sum of the individual variances $\sigma_P^2 = \sum_{j=1}^{J} w_j \sigma_j^2$, and if any of the sampling techniques is bad in a given region, then P will also have a high variance.

A relatively simple and particularly useful choice of weights is the balance heuristic [23]. In this case, the weights $w_j(\boldsymbol{\theta})$ are assigned according to

$$w_j(\boldsymbol{\theta}) = \frac{M_j p_j^*(\boldsymbol{\theta})}{\sum_{j'=1}^{J} M_{j'} p_{j'}^*(\boldsymbol{\theta}).} \quad (19)$$

The quantity $q_j(\boldsymbol{\theta}) = M_j p_j^*(\boldsymbol{\theta})$ is proportional to the expected number of hits from the jth distribution. Thus, the weight associated with a sample $\boldsymbol{\theta}$ with the balance heuristic is given by the relative likelihood of realizing that sample with the jth distribution relative to the total likelihood of realizing that same sample with all distributions. Thus, (19) weights each distribution $p_j^*(\boldsymbol{\theta})$ most heavily in those regions of sample space where $p_j^*(\boldsymbol{\theta})$ is largest. ((19) can also be written in terms of likelihood ratios, a form that is particularly convenient for use in (14).)

The balance heuristic has been mathematically proven to be asymptotically close to optimal as the number of realizations becomes large [23]. Because of its effectiveness and simplicity, we used the balance heuristic in this paper. The same choice was also used in [12]–[15]. Of course, other strategies are possible (see [23], [24], and Appendix I). In specific cases, some of these alternatives might offer better results—namely, lower variances. In any given situation, however, it is difficult to tell *a priori* which choice will be best.

III. IS FOR THE DGD

We now discuss the application of IS to the numerical simulation of those rare events in which the DGD at a particular frequency assumes large values. As explained in Section II, the key step to IS is determining an efficient biasing distribution or an efficient set of biasing distributions. We shall see that information can be extracted from the equations governing PMD that leads to very efficient DGD biasing distributions.

A. Transmission Matrices and the PMD Vector

Let us first establish some basic notation. The action of any lossless transmission element on an optical pulse can be described, up to a polarization-independent factor, by a unitary 2×2 frequency-dependent transmission matrix $\mathsf{U}(z,\omega)$, which can be expressed as [3]

$$\begin{aligned}\mathsf{U} &= \exp[-i(\varphi/2)\hat{\mathbf{r}}\cdot\boldsymbol{\sigma}] \\ &= \mathsf{I}\cos(\varphi/2) - i\,\hat{\mathbf{r}}\cdot\boldsymbol{\sigma}\,\sin(\varphi/2)\end{aligned} \tag{20}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is a vector of Pauli matrices [3], and $\hat{\mathbf{r}}$ is a unit vector representing a rotation axis (hereafter, the circumflex accent will be used to denote unit vectors).[2] The orthogonal 3×3 Müller matrix $\mathsf{R}(z,\omega)$ corresponding to the Jones matrix $\mathsf{U}(z,\omega)$ is given by the expression

$$\begin{aligned}\mathsf{R} &= \exp\left[\varphi(\hat{\mathbf{r}}\times)\right] \\ &= (\cos\varphi)\mathsf{I} + (1-\cos\varphi)\hat{\mathbf{r}}\hat{\mathbf{r}}^T + (\sin\varphi)\hat{\mathbf{r}}\times\end{aligned} \tag{21}$$

which represents a frequency-dependent rotation through an angle φ about the axis $\hat{\mathbf{r}}$. Note that, in general, both φ and $\hat{\mathbf{r}}$ are also frequency dependent.

As described in [3], the polarization effects can be uniquely characterized (up to an arbitrary constant rotation) by the PMD vector $\boldsymbol{\tau}(z,\omega)$ (also called the polarization dispersion vector), which is the real three-dimensional (3-D) vector defined by $\boldsymbol{\tau}\times = (\partial\mathsf{R}/\partial\omega)\mathsf{R}^T$, or equivalently by $\boldsymbol{\tau}\cdot\boldsymbol{\sigma} = 2i(\partial\mathsf{U}/\partial\omega)\mathsf{U}^{-1}$. The DGD is given by the length of this PMD vector. For linearly birefringent elements, $\hat{\mathbf{r}}$ is independent of frequency, and φ has a linear frequency dependence $\varphi(\omega) = 2b'\ell\omega = \varphi(\omega_0) + 2b'\ell(\omega-\omega_0)$, where b' is the frequency derivative of the birefringence strength of the element, ℓ is its length, and ω_0 is a reference frequency [3]. In this case, the PMD vector is simply given by $\Delta\boldsymbol{\tau} = 2b'\ell\hat{\mathbf{r}}$. In addition, if no circular birefringence is present (e.g., for linearly birefringent elements), $\Delta\boldsymbol{\tau}$ lies on the equatorial plane of the Poincaré sphere [29].

B. The First-Order PMD Concatenation Equation

A standard technique for simulating PMD effects is the coarse-step method [26], which approximates the continuous birefringence variations present in real fibers using a concatenation of fixed-length birefringent sections. Many experimental PMD generation techniques also employ a concatenation of birefringent elements, such as high-birefringence fibers [27] or birefringent waveplates [28], [29]. These can be connected by either polarization scramblers (e.g., polarization controllers [27]), or rotatable connectors [28].

In all cases, the total PMD vector after the $(n+1)$st section, $\boldsymbol{\tau}^{(n+1)}$, is obtained from the PMD concatenation equation [3]

$$\boldsymbol{\tau}^{(n+1)} = \mathsf{R}_{n+1}\boldsymbol{\tau}^{(n)} + \Delta\boldsymbol{\tau}_{n+1} \tag{22}$$

[2]Throughout this work, we use the conventions of [3] regarding the representations of an optical pulse (in particular, the choice of sign for the carrier frequency), of the Pauli matrices $\{\sigma_1, \sigma_2, \sigma_3\}$ and, as a consequence, of the Stokes' parameters. In these conventions, the third component of a Stokes' vector is $+1$ for right-handed circular polarization. We refer the reader to in [3, App. B] for further details and for the connection with other conventions.

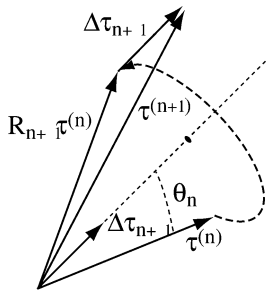

Fig. 1. Graphical representation of the first-order PMD concatentation.

as illustrated in Fig. 1. Here, $\Delta\boldsymbol{\tau}_{n+1}$ is the differential contribution to the PMD vector coming from the $(n+1)$st section. For fixed-length sections, the magnitude of $\Delta\boldsymbol{\tau}_{n+1}$ is constant, and only its direction varies. For linearly birefringent elements, $\mathsf{R}_{n+1}\Delta\boldsymbol{\tau}_{n+1} = \Delta\boldsymbol{\tau}_{n+1}$ and (22) can alternatively be written as[3]

$$\boldsymbol{\tau}^{(n+1)} = \mathsf{R}_{n+1}\left(\boldsymbol{\tau}^{(n)} + \Delta\boldsymbol{\tau}_{n+1}\right). \tag{23}$$

When polarization controllers are present, an additional rotation matrix Q_{n+1} precedes $\boldsymbol{\tau}^{(n)}$ in (22) and (23) to represent the action of the controller. It is possible to factor out altogether this rotation Q_{n+1} from the concatenation equation by redefining $\boldsymbol{\tau}^{(n+1)}$; the resulting equation is identical to (23) with a new rotation matrix $\mathsf{R}'_{n+1} = \mathsf{R}_{n+1}\mathsf{Q}_{n+1}$ except that the new contributions $\Delta\boldsymbol{\tau}'_{n+1} = \mathsf{Q}_{n+1}^{-1}\Delta\boldsymbol{\tau}_{n+1}$ are now uniformly distributed on the Poincaré sphere. Thus, except for the rotation R'_{n+1}, the case of polarization scramblers is formally equivalent to a 3-D random walk. Since a theoretical result for a 3-D random walk is available, this allows an independent check of the results for IS with polarization scramblers. To distinguish the different cases, in the following we refer to the situation where there are no polarization controllers as the case of rotatable waveplates.

C. Deterministic Biasing Directions

When applying IS to the PMD concatenation (22), we need to target the rare events in which the DGD assumes values much larger than its mean. The appropriate variables to control to do this are the polar angles θ_n between the PMD vector after the first n sections $\boldsymbol{\tau}^{(n)}$ and the differential contribution in the $(n+1)$st section $\Delta\boldsymbol{\tau}_{n+1}$, as shown in Fig. 1. The critical step of biasing is to determine at each fiber section the deterministic direction around which the random samples should preferentially be selected.

It should be intuitively clear that the configurations that lead to large DGDs are those in which the individual contributions to the PMD vector from each section tend to be aligned with each other. It is not useful to attempt this alignment with respect to some fixed direction, however, because random sampling is still to be performed, and after a sufficiently large number of random variations, it is possible for the fluctuations to build up

[3]Equations (20) and (21) express the Jones matrix U and the Müller matrix R in terms of frequency-dependent SU(2) and O(3) rotations, respectively. These formulas are valid for birefringent elements with any degree of ellipticity. In general, however, PMD vector $\Delta\boldsymbol{\tau}$ is parallel to the rotation axis $\hat{\mathbf{r}}$ only for linearly birefringent elements, corresponding to the case when $\hat{\mathbf{r}}$ is independent of frequency and ϕ varies linearly with frequency (e.g., see [3, eq. (5.13)]).

to the point where the direction of the PMD vector is arbitrary [1]. The correct strategy, therefore, is to choose the alignment of each successive section with respect to the direction defined by the total PMD vector up to that point.

When polarization scramblers are present, the length and orientation of the successive differential PMD vectors $\Delta\boldsymbol{\tau}_{n+1}$ can be regarded as fixed, while the output of the polarization scramblers varies. We therefore bias the simulations by making the scramblers preferentially rotate $\boldsymbol{\tau}^{(n)}$ toward the direction of $\Delta\boldsymbol{\tau}_{n+1}$. More specifically, we bias θ_n (the angle between the PMD vector at the output of the nth scrambler and the $(n+1)$st differential PMD vector) toward zero. This choice does not uniquely determine the orientation of the PMD vector at the scrambler output, however, because $\boldsymbol{\tau}^{(n)}$ can still be rotated by an arbitrary amount about $\Delta\boldsymbol{\tau}_{n+1}$ while keeping θ_n constant. We take this additional rotational angle to be uniformly distributed.

When using rotatable waveplates, the relative orientations between sections are the primary variables determining the total DGD, since different configurations are then generated solely by rotating sections relative to one another. In the case of linearly birefringent elements, each section's PMD vector $\Delta\boldsymbol{\tau}_{n+1}$ lies in the equatorial plane. The largest increase in the DGD is obtained when $\Delta\boldsymbol{\tau}_{n+1}$ is aligned with the projection of the total PMD vector up to that point $\boldsymbol{\tau}^{(n)}$ onto the equatorial plane. Thus, the biasing toward large DGDs is done by choosing $\Delta\boldsymbol{\tau}_{n+1}$ to be preferentially aligned with the projection of $\boldsymbol{\tau}^{(n)}$ onto the equatorial plane.

If frequency dependence is desired (as, for example, when simulating pulse transmission), or in the case of rotatable waveplates, the differential phase retardations φ_{n+1} in the Müller matrix R_{n+1} must also be specified. Recall that each beat length of a birefringent section generates a 2π retardation. In practice, sections with significant DGDs will be many beat lengths long, and unless the section lengths are precise to within a small fraction of a beat length, these retardations will vary from section to section. One way to proceed is to choose the retardation angles $\varphi_{n+1}(\omega_0)$ to be uniformly distributed between 0 and 2π. Alternatively, one could assume a random distribution of lengths ℓ_{n+1}; if the variance of the lengths is large compared with the beat length, however, an approximately uniform distribution of the angles $\varphi_{n+1}(\omega_0)$ results. Of course, if the lengths are very precise, specific phase retardations should be used. In either case, the results do not depend significantly upon the particular retardation angles used, except for certain clearly pathological cases such as identical angles with φ_{n+1} equal to 0 or π.

D. Biasing Distributions

Once the deterministic biasing directions have been identified, a biasing distribution must be chosen that preferentially selects the random samples to be close to these directions. In the unbiased case, the angles θ_n are independent random variables, with $\cos\theta_n$ uniformly distributed in $[-1,1]$. When applying IS, we choose $\cos\theta_n = 2x^{1/\alpha} - 1$, where x is a uniform random variable in $[0,1]$ and $\alpha \geq 1$ is a biasing parameter. Other choices are possible for the biasing distribution of the θ_n; the effectiveness of the method is not very sensitive to the particular distribution used. This choice yields the likelihood ratio as

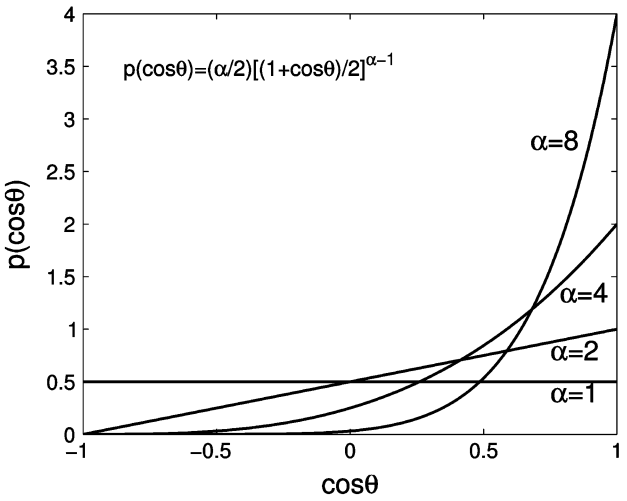

Fig. 2. Angular biasing distributions $p(\cos\theta) = p_\alpha(\theta)/\sin\theta$ in the case of polarization scramblers for different values of the biasing strength α.

$$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} \frac{p_1(\theta_n)}{p_\alpha(\theta_n)} \tag{24}$$

where

$$p_\alpha(\theta) = (\alpha/2)\sin\theta\,[(1+\cos\theta)/2]^{\alpha-1}. \tag{25}$$

The value $\alpha = 1$ reproduces the unbiased case, while increasing values of α biases the distribution toward $\theta = 0$, thus preferentially producing configurations with increasingly large values of DGD. If the individual sections are independent, the overall likelihood ratio is the product of all of the individual likelihood ratios. Here, the orientation of each section depends upon the PMD vectors of all previous sections. Nevertheless, the overall likelihood ratio can still be written as a product of individual likelihood ratios (see Appendix II).

For the case of birefringent waveplates, the individual contributions lie on the equatorial plane of the Poincaré sphere. In this case, we choose $\Delta\boldsymbol{\tau}_{n+1}$ so that the angle θ_n between it and the projection of $\boldsymbol{\tau}^{(n)}$ onto the equatorial plane is preferentially close to zero. This can be done, for example, by taking $\theta = \pi\mathrm{sgn}(2x-1)|2x-1|^\alpha$, where x is again uniformly distributed between 0 and 1.[4] As before, other choices are possible. If $\alpha = 1$ (unbiased case), θ is uniformly distributed in $[-\pi,\pi]$, while for $\alpha > 1, \theta$ is concentrated near 0, meaning that the waveplates' PMD vectors $\Delta\boldsymbol{\tau}_{n+1}$ will be preferentially aligned with the projection of the total PMD vector to that point onto the equatorial plane. As explained previously, this preferential alignment tends to increase the length of the total PMD vector. The likelihood ratio is still given by (24), but now

$$p_\alpha(\theta) = (1/\alpha\pi)|\theta/\pi|^{1-\alpha}. \tag{26}$$

As illustrated in Figs. 2 and 3 for the case of polarization scramblers, each biasing distribution allows us to sample a different

[4]In [11], it was incorrectly stated that the same distribution as that used for scramblers was utilized. The correct expression was used in the simulations, however.

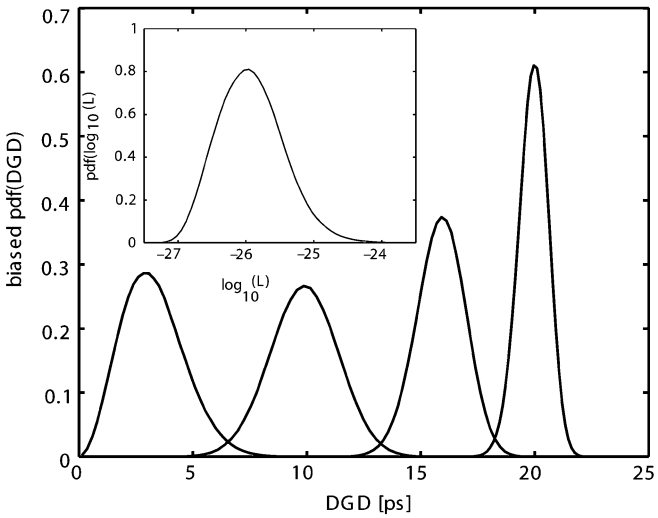

Fig. 3. Biased DGD distributions (numerically reconstructed from MC simulations) resulting from the distributions in Fig. 2 for a concatenation of 50 birefringent elements with 0.5-ps DGD each (corresponding to a mean DGD of 3.26 ps) with polarization scramblers. Inset: Likelihood ratios and their distribution for samples with DGD values within a 0.2-ps interval around 20 ps and simultaneously with second-order PMD lying within a 6 ps^2 interval around 20 ps^2 for the same emulator as in Fig. 3. The distribution was numerically reconstructed through MC simulations with $\alpha = 8$.

range of DGDs, and simulations with different values of biasing strength α can be combined using the techniques described in Section II-D in order to cover the whole range of DGD values.

We emphasize that different configurations with the same DGD (and/or SOPMD) can have very different likelihood ratios, and thus their relative contribution can vary substantially. It is therefore not sufficient merely to generate *some* realizations with a given DGD and/or SOPMD. The inset in Fig. 3 is an illustration of this point; here, we plot the distribution of likelihood ratios for samples generated by biasing the DGD but that produce a DGD and SOPMD in a narrow range. Note that the horizontal axis has a logarithmic scale and shows a variation of almost three orders of magnitude of the likelihood ratios. Similar considerations apply to the pulse broadening, the power penalty, or other quantities of interest: different fiber realizations with the same values of first- and second-order PMD can result in different values for any of these quantities.

IV. IS for Both First- and Second-Order PMD

Because the PMD vector is frequency dependent, in general the DGD at any given frequency is not the sole determiner of system outages. In particular, second-order PMD, which includes both depolarization and PCD, is known to produce additional system penalties [30], [31]. This is especially important when PMD compensation is applied, since first-order PMD is typically reduced to moderate values (or perfectly cancelled at a particular frequency); in this case, it is possible for SOPMD to become the primary cause of system penalties [15], [30]–[33]. Thus, when using IS to calculate outage probability, it is important to generate sufficient statistics of the frequency derivative of the PMD vector $\boldsymbol{\tau}_\omega$, which quantifies SOPMD. We now describe an IS technique that employs multiple biasing schemes to generate arbitrary combinations of first- and second-order PMD,

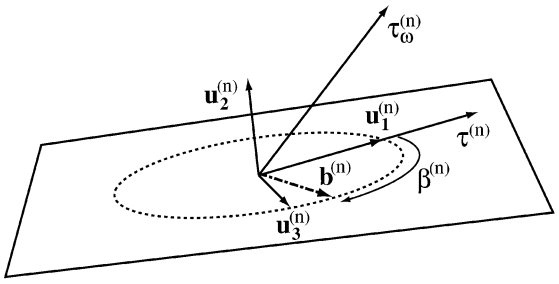

Fig. 4. Diagram showing the preferential direction $\mathbf{b}$ for biasing the simulations in the plane identified by $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\tau}^{(n)} \times \boldsymbol{\tau}_\omega^{(n)}$.

effectively targeting all regions of the $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane. As a result, the method generates much more complete PMD statistics than first-order biasing alone.

A. The Second-Order PMD Concatenation Equation

A concatenation equation similar to (22) holds for the second-order PMD vector $\boldsymbol{\tau}_\omega$ [3], as follows:

$$\boldsymbol{\tau}_\omega^{(n+1)} = \mathsf{R}_{n+1}\boldsymbol{\tau}_\omega^{(n)} + \Delta\boldsymbol{\tau}_{n+1} \times \boldsymbol{\tau}^{(n+1)} + \Delta\boldsymbol{\tau}_{\omega,n+1}. \quad (27)$$

For linearly birefringent elements, $\Delta\boldsymbol{\tau}_{\omega,n+1} = 0$. In addition, $\mathsf{R}_{n+1}\Delta\boldsymbol{\tau}_{n+1} = \Delta\boldsymbol{\tau}_{n+1}$, and (22) can be rewritten in a form similar to (23)

$$\boldsymbol{\tau}_\omega^{(n+1)} = \mathsf{R}_{n+1}\left(\boldsymbol{\tau}_\omega^{(n)} + \Delta\boldsymbol{\tau}_{n+1} \times \boldsymbol{\tau}^{(n)}\right). \quad (28)$$

In a manner similar to the first-order concatenation equation, when polarization scramblers are present, an additional rotation matrix Q_{n+1} precedes $\boldsymbol{\tau}_\omega^{(n)}$ in (27) and (28). As before, the concatenation equation can then be written in a form identical to (28), with a new rotation matrix $\mathsf{R}'_{n+1} = \mathsf{R}_{n+1}\mathsf{Q}_{n+1}$ and with the new contributions $\Delta\boldsymbol{\tau}'_{n+1} = \mathsf{Q}_{n+1}^{-1}\Delta\boldsymbol{\tau}_{n+1}$ uniformly distributed on the Poincaré sphere.

As demonstrated previously, the appropriate variables to control when applying IS are the orientations of the individual PMD vectors of each section $\Delta\boldsymbol{\tau}_{n+1}$. IS works by biasing these vectors toward specific directions $\mathbf{b}^{(n)}$ which maximally increase the particular quantity of interest (in the previous section, the specific directions were those of the preceeding total PMD vectors, and the quantity of interest was the total DGD). In what follows, we will formalize this idea by characterizing the vector $\mathbf{b}^{(n)}$ relative to a orthonormal frame of reference $\mathcal{U}^{(n)}$ formed by the unit vectors $\{\hat{\mathbf{u}}_1^{(n)}, \hat{\mathbf{u}}_2^{(n)}, \hat{\mathbf{u}}_3^{(n)}\}$, where

$$\hat{\mathbf{u}}_1^{(n)} = \boldsymbol{\tau}^{(n)} \Big/ \left|\boldsymbol{\tau}^{(n)}\right| \quad (29a)$$

$$\hat{\mathbf{u}}_2^{(n)} = \boldsymbol{\tau}_{\omega,\perp}^{(n)} \Big/ \left|\boldsymbol{\tau}_{\omega,\perp}^{(n)}\right| \quad (29b)$$

$$\hat{\mathbf{u}}_3^{(n)} = \hat{\mathbf{u}}_1^{(n)} \times \hat{\mathbf{u}}_2^{(n)}. \quad (29c)$$

Here, $\boldsymbol{\tau}_{\omega,\perp}^{(n)}$ is the component of $\boldsymbol{\tau}_\omega^{(n)}$ perpendicular to $\boldsymbol{\tau}^{(n)}$, as illustrated in Fig. 4. The magnitudes of $\boldsymbol{\tau}_{\omega,\|}$ (the component of $\boldsymbol{\tau}_\omega^{(n)}$ parallel to $\boldsymbol{\tau}^{(n)}$) and $\boldsymbol{\tau}_{\omega,\perp}$ quantify the PCD and the depolarization, respectively.

As before, in order to apply IS, we first need to find the deterministic configurations that maximize the desired quantity.

(Once these directions are found, random samples around these directions are generated.) For example, the total DGD is maximized by choosing $\Delta\boldsymbol{\tau}_{n+1}$ to be preferentially aligned with the previous PMD vector $\boldsymbol{\tau}^{(n)}$, i.e., using $\mathbf{b}^{(n)} = \hat{\mathbf{u}}_1^{(n)}$. We refer to this case as first-order biasing. While first-order biasing yields the largest DGDs, it does not produce particularly large values of SOPMD, because when $\Delta\boldsymbol{\tau}_{n+1}$ is parallel to $\boldsymbol{\tau}^{(n)}$, the contribution to $\boldsymbol{\tau}_\omega^{(n+1)}$ is zero. In fact, if all the $\Delta\boldsymbol{\tau}_{n+1}$ were parallel to $\boldsymbol{\tau}^{(n)}$, no SOPMD would be produced at all. Random fluctuations add significant SOPMD, but large values are not specifically targeted.

When just a single section is added, one can easily show that, in the case of polarization scramblers, the orientation that maximizes the contribution to $\boldsymbol{\tau}_\omega^{(n+1)}$ is to align $\Delta\boldsymbol{\tau}_{n+1}$ with the direction of $\hat{\mathbf{u}}_3^{(n)}$. In the case of rotatable waveplates, a constrained (Lagrange multiplier) maximization shows that the optimum is to align $\Delta\boldsymbol{\tau}_{n+1}$ with the projection of $\hat{\mathbf{u}}_3^{(n)}$ onto the equatorial plane. Unfortunately, looking at a single section is not sufficient, because (27) shows that the rate at which $\boldsymbol{\tau}_\omega^{(n+1)}$ increases depends both upon the orientation of $\Delta\boldsymbol{\tau}_{n+1}$ and upon the size of $\boldsymbol{\tau}^{(n)}$. Thus, when many sections are used, maximizing the SOPMD also depends upon the growth of $\boldsymbol{\tau}^{(n)}$.

B. Continuum Limit and Deterministic Biasing Directions

When the number of sections in the emulator is large, it is possible to employ a continuum approximation to find the deterministic configurations that generate the maximum second-order PMD. Specifically, we let $\lim_{\Delta z\to 0}\Delta\boldsymbol{\tau}_{n+1}/\Delta z = \mathbf{b}(z)$. The magnitude of $\mathbf{b}(z)$ describes the rate at which PMD is added by the birefringent sections. In this continuum limit, for the case of polarization scramblers, one obtains (after factoring out a common rotation R'_{n+1})

$$\frac{d\boldsymbol{\tau}}{dz} = \mathbf{b} \tag{30a}$$

$$\frac{d\boldsymbol{\tau}_\omega}{dz} = \mathbf{b}\times\boldsymbol{\tau}. \tag{30b}$$

We want to write these equations in the frame of reference $\mathcal{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3\}$ defined by the analogs of (29). (This frame rotates with the first- and second-order PMD vectors.) To this end, we introduce the rotated quantities $\boldsymbol{\tau}' = \mathsf{T}\boldsymbol{\tau}, \boldsymbol{\tau}'_\omega = \mathsf{T}\boldsymbol{\tau}_\omega$ and $\mathbf{b}' = \mathsf{T}\,\mathbf{b}$. Written in the new frame of reference, both (30a) and (30b) have an additional term in the RHS: $\boldsymbol{\Omega}\times\boldsymbol{\tau}'$ and $\boldsymbol{\Omega}\times\boldsymbol{\tau}'_\omega$, respectively, where $\boldsymbol{\Omega}\times = (d\mathsf{T}/dz)\mathsf{T}^{-1}$. Requiring that $\tau'_2 = \tau'_3 = d\tau'_2/dz = d\tau'_3/dz = 0$ implies $\Omega_2 = b'_3/\tau'$ and $\Omega_3 = -b'_2/\tau'$, and requiring $\tau'_{\omega,3} = d\tau'_{\omega,3}/dz = 0$ implies $\Omega_1 = (b'_2\tau'^2 + b'_3\tau'_{\omega,\|})/\tau'_{\omega,\perp}\tau'$. Omitting primes for simplicity, we then obtain the following system of equations:

$$\frac{d\tau}{dz} = b_1 \tag{31a}$$

$$\frac{d\tau_{\omega,\|}}{dz} = b_2\frac{\tau_{\omega,\perp}}{\tau} \tag{31b}$$

$$\frac{d\tau_{\omega,\perp}}{dz} = b_3\tau - b_2\frac{\tau_{\omega,\|}}{\tau} \tag{31c}$$

where (b_1, b_2, b_3) are now the components of $\mathbf{b}$ with respect to the reference frame $\mathcal{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3\}$.

The goal is now to find the function $\mathbf{b}(z)$ that maximizes SOPMD. Considering first for simplicity only the case when $\mathbf{b}(z)$ is independent of z (i.e., constant biasing direction), the solution of (31) is

$$\tau = b_1 z, \quad \tau_{\omega,\|} = \frac{b_1^2 b_2 b_3}{4b_1^2 + b_2^2}z^2, \quad \tau_{\omega,\perp} = \frac{2b_1^3 b_3}{4b_1^2 + b_2^2}z^2. \tag{32}$$

The choice that maximizes second-order PMD is now easily found to be $b_1 = b_3 = b/\sqrt{2}$ and $b_2 = 0$, with $b > 0$, which splits the difference between the growth of first-order PMD and SOPMD. We refer to directions with zero $\hat{\mathbf{u}}_2$ component as *in-plane* biasing. Similar calculations can be done to find the choices that maximize the PCD or the depolarization. For the depolarization, the solution is the same, which suggests that the total second-order PMD is maximized by the depolarization. For the PCD, the solution is $b_1 = (b/2)\sqrt{-1+r}, b_2 = \pm b/\sqrt{2}, b_3 = (b/2)\sqrt{3-r}$, where $r = \sqrt{11/3}$ and $b > 0$, which suggests that the maximum value of PCD is obtained with off-plane biasing.

We strongly emphasize, however, that solutions with constant biasing directions do not generate the maximum SOPMD. Fortunately, (31) can be solved exactly for any choice of function $\mathbf{b}(z)$: the exact solution is

$$\tau(z) = \int_0^z b_1(z')\,dz' \tag{33a}$$

$$\tau_{\omega,\|}(z) = \int_0^z b_3(z')\tau(z')\sin[B(z,z')]\,dz' \tag{33b}$$

$$\tau_{\omega,\perp}(z) = \int_0^z b_3(z')\tau(z')\cos[B(z,z')]\,dz' \tag{33c}$$

where

$$B(z,z') = \int_{z'}^z \frac{b_2(z'')}{\tau(z'')}\,dz''. \tag{33d}$$

Starting with this (33), the choice of $\mathbf{b}(z)$ that maximizes the magnitude of SOPMD can now be found using calculus of variations (seeAppendix III). The result is that the maximum SOPMD is obtained for in-plane contributions

$$b_1(z) = b\cos\beta(z), \quad b_2(z) = 0, \quad b_3(z) = b\sin\beta(z) \tag{34}$$

with, for the case of equal-length sections (namely, for $|\mathbf{b}(z)| = b$), a linearly varying angle profile

$$\beta(z) = \beta_{\max}z/z_{\max} \tag{35}$$

with $\beta_{\max} = \pi/2$. (The solution for the case of nonequal-length sections can be obtained easily from this one; see Appendix III.) With multiple biasing strengths, this biasing choice generates region 3 in Fig. 5. We refer to this choice, which yields the largest values of $|\boldsymbol{\tau}_\omega|$, as pure second-order biasing.

In many practical situations, a more complete coverage of the $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane is needed. In this case, intermediate biasing choices must also be used in addition to pure first- and second-order biasing. These intermediate choices can be obtained by using the calculus of variations to maximize a linear combination of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, as obtained from (33). The resulting form of $\mathbf{b}(z)$ is the same as previously, except that the value of the final angle $\beta_{\max}$ now varies between 0 and π, the particular value depending upon the specific linear combination of first- and second-order PMD being maximized. A selection of an-

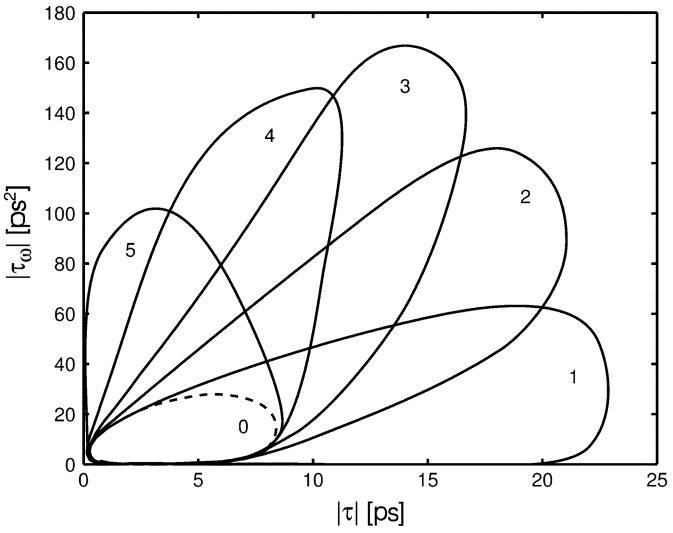

Fig. 5. Regions of the $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane targeted by the various biasing methods. Region 1 corresponds to first-order biasing ($\beta = 0$), region 2 to pure second-order biasing ($\beta_{\max} = \pi/2$), and regions 3, 4, and 5 to $\beta_{\max} = \pi/4$, $3\pi/4$, and π, respectively. The dashed line shows the much smaller region obtained with unbiased runs. As in Fig. 3, 50 sections with 0.5-ps DGD each and polarization scramblers were used.

gles, together with the resulting regions in the $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane, is shown in Fig. 5. (Note region 1 is the result for the case of biasing the DGD.) The advantage of using multiple biasing—as opposed to just pure first- or second-order biasing or no biasing at all—is evident. Each value of $\beta_{\max}$ generates samples lying in a region that emanates in a roughly radial fashion from the location where the joint pdf is maximum. Together, a set of angles $\beta_{\max}$ can be used to cover the entire $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane.

When birefringent waveplates are used, the differential contributions $\Delta\boldsymbol{\tau}_{n+1}$ are confined to the equatorial plane of the Poincaré sphere. In this case, there is no full theory. Nevertheless, in the limited cases for which the optimization can be done (single steps of optimizing the DGD or the second-order PMD), the result is that one should use the projection onto the equatorial plane, which is the strategy that we adopt here: the appropriate biasing directions are obtained by projecting the vectors $\mathbf{b}$ determined previously onto the equatorial plane. We have found this to work extremely well in practice, i.e., with roughly the same degree of efficiency as for polarization scramblers, where the full theory is available.

C. Biasing Distributions and Maximum SOPMD

For both PMD generation models, once the deterministic biasing direction $\mathbf{b}^{(n)} = \mathbf{b}(n\,\Delta z)$ has been selected, the biasing distribution for the random orientation of $\Delta\boldsymbol{\tau}_{n+1}$ is chosen as described in Section III. That is, for the case of polarization scramblers, the angle θ_n between $\mathbf{b}^{(n)}$ and $\Delta\boldsymbol{\tau}_{n+1}$ is biased toward zero (as before), but the remaining degrees of orientation degrees of freedom are uniformly distributed. Note that, as a result, when random angles are selected, $\Delta\boldsymbol{\tau}_{n+1}$ is likely to have a component off the $(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_3)$ plane, meaning that PCD will be generated. For the case of birefringent waveplates, the biasing is done by choosing the angle between $\Delta\boldsymbol{\tau}_{n+1}$ and the projection of $\mathbf{b}^{(n)}$ onto the equatorial plane to be biased toward zero.

It should be emphasized that, even though the actual biasing directions must be calculated dynamically for each MC sample (since they depend on the orientation of the first- and second-order PMD vectors up to that point), the determination of the biasing direction relative to the local coordinate system was done once and for all. Although the calculations are a bit complicated, the final result (namely, linear profiles for the angle $\beta(z)$ with different target values) is quite simple, can be easily implemented, and provides an effective recipe for simulating transmission effects characterized by large values of PMD.

One way to evaluate the effectiveness of a biasing scheme (or of a PMD generation method) is to evaluate the maximum PMD that it can produce. For a device with equal-length birefringent sections, the exact solution of the continuum model allows us to obtain an approximate value for the maximum SOPMD. More precisely, by evaluating the exact solution of the continuum model with pure first-order or pure second-order biasing [i.e., substituting (34) with $\beta_{\max} = 0$ or $\pi/2$ into (33)], we obtain $|\boldsymbol{\tau}|_{\max} = bz_{\max}$ and $|\boldsymbol{\tau}_\omega|_{\max} = b^2 z_{\max}^2/\pi$, respectively. Obviously, however, the maximum DGD for a device with a finite number of equal-length birefringent sections is $|\boldsymbol{\tau}|_{\max} = N\,\Delta\tau$, where N is the number of sections and $\Delta\tau$ is the DGD of each section. Thus, we can eliminate the product $bz_{\max}$ and obtain an approximation for the maximum SOPMD as $|\boldsymbol{\tau}_\omega|_{\max} = N^2(\Delta\tau)^2/\pi$. With 50 birefringent sections and polarization scramblers, the approximate formula for the maximum SOPMD is within 0.7% of the actual value (determined by numerical optimization). If constant biasing directions are used instead, the maximum attainable value of second-order PMD, in the continuum limit, is $|\boldsymbol{\tau}_\omega|_{\max} = N^2(\Delta\tau)^2/4$, i.e., 20% less.

D. Asymptotic Efficiency and PCD

The algorithms that we have constructed to obtain targeted values of first- and/or SOPMD appear to be asymptotically efficient in almost all cases. This is evidenced by the numerical results, which show that the number of MC samples that are necessary to achieve a small variance does not grow rapidly as the simulations are pushed further into the tails of the pdfs. We believe this to be the case because the continuum approximation to the PMD concatenation equations appears to characterize their behavior with relatively high accuracy for most cases of interest. We conjecture that it might be possible to justify this claim in a more precise way, but we have not done so. On the other hand, while the solutions with $\mathbf{b}(z) = \text{const}$ provide an optimal biasing choice within the framework of constant directions, they do not result in asymptotically efficient methods. This is easily verified, for example, by trying to evaluate the pdf of SOPMD further and further out into the tails.

The situation appears to be different for the PCD: although we have found an optimal biasing in the context of constant biasing directions, we have not so far been able to do the same in the more general case. Numerical evidence shows that the solution for constant biasing directions is not optimal. Numerical optimization also suggests, however, that the configurations that produce the largest PCD are ones that are near the solution given by (34) and (35) with $\beta_{\max} = \pi$. In this particular situation, $|\boldsymbol{\tau}|$ first increases as z (or the number of sections) increases and then decreases until it becomes close to zero at the end. At

the same time, $|\boldsymbol{\tau}_\omega|$ increases monotonically. Thus, near the last section, a situation arises where $|\boldsymbol{\tau}_\omega|$ is relatively large and $|\boldsymbol{\tau}|$ is very small. In this situation, because $|\boldsymbol{\tau}|$ is small, it is relatively easy for small variations (which arise because the sections' orientations are still being chosen randomly), to perturb $\boldsymbol{\tau}$ so that it points in the direction of $|\boldsymbol{\tau}_\omega|$. In this case, large PCD results. Thus, the largest PCD values seem to arise when $|\boldsymbol{\tau}|$ is small and $|\boldsymbol{\tau}_\omega|$ is large. This method of generating large PCD appears to be reasonably efficient.

V. Examples and Applications

The techniques described in the Sections III and IV have two main applications: the calculation of PMD statistics—most notably for PMD emulators, where few analytical results exist—and the calculation of PMD-induced transmission effects, such as pulse broadening, power penalties, and outage probabilities. Here, we demonstrate how the techniques can be applied by discussing a few specific examples.

A. Numerical Calculation of Probability Densities

We first discuss the basic strategy that can be used to set up the numerical simulations. Suppose it is desired to construct the pdf of some quantity. This is usually done with MC simulations and sorting the results into bins. The desired result is more than a single quantity; more precisely, we are trying to simultaneously estimate all the integrals

$$p_i = \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} p_x(x)\,dx = \frac{1}{\Delta x}\int_\Theta I_{R_i}(X(\boldsymbol{\theta}))p_\theta(\boldsymbol{\theta})\,d\boldsymbol{\theta} \tag{36}$$

where x is the quantity of interest, $x_i = i\Delta x$, with Δx being the bin size, and $R_i = [x_i, x_{i+1}]$. The integrals in (36) are of the same type as (1); thus, we can apply the IS techniques presented in Section II. It is clear, however, that in general no single biasing distribution can efficiently generate all the possible values of X (e.g., see Fig. 3 for the DGD). Thus, we need to resort to multiple IS, using the distributions described in Section III-D with different values of biasing angles and biasing strengths. Therefore, the procedure is as follows:

1) choose a set of biasing directions (values of $\beta_{\max}$) and biasing strengths (different values of α);
2) perform a predetermined number of MC simulations for each direction and biasing strength;
3) sort the results of all the MC samples into bins, keeping track of the likelihood ratio for each sample, and combine the individual samples using one of the weighting strategies presented previously.

In all of the results presented here, we have used the balance heuristic.

For concreteness, in the following, we consider a specific scenario, namely, PMD generation resulting from the concatenation of 50 birefringent sections with 0.5 ps of DGD per section (corresponding to a mean DGD of 3.26 ps) and employing either polarization scramblers or rotatable waveplates.

B. Probability Densities of First- and Second-Order PMD

As explained previously, in the case of polarization scramblers, the evolution of the PMD vector is equivalent to a 3-D random walk, and an exact solution is available for the pdf (either in terms of an integral or as a Fourier series) that can be used to verify the accuracy of the importance-sampled MC simulations [34], [35]. For moderate values of the DGD and for numbers of sections not too small, the pdf $p_N(r)$ is well approximated by a Maxwellian distribution

$$p_{\mathrm{DGD}}(r) = \frac{\sqrt{2}r^2}{\sqrt{\pi}\,\sigma^3}e^{-r^2/2\sigma^2} \tag{37}$$

where $\sigma^2 = \langle\tau^2\rangle/3$. The degree of agreement, of course, improves as the number of sections is increased [35].

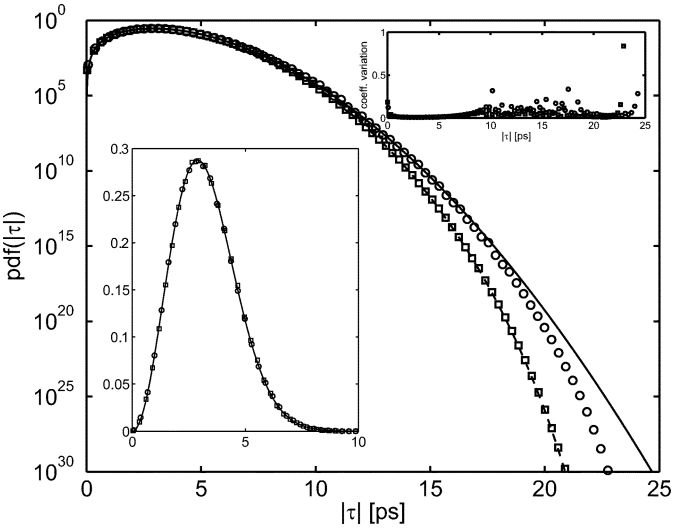

Fig. 6. Importance-sampled pdf of the DGD for 50 sections with 0.5-ps DGD each and with polarization scramblers (squares) or birefringent waveplates (circles). Solid curve: Maxwellian distribution with mean DGD of 3.26 ps; dashed line: exact (linear scale) and asymptotic (logarithmic scale) solutions from [34]. Bottom inset: the pdfs on a linear scale. Top inset: the coefficient of variation. A total of 2×10^6 MC realizations were used.

To our knowledge, no closed-form expression exists for the pdf of the DGD in the case of birefringent waveplates or for the pdf of the SOPMD of a concatenation of finite-length sections, with or without polarization scramblers. An expression exists, however, for the pdf of the magnitude of SoPMD for an optical fiber in the long-length regime [1], i.e., as the number of sections goes to infinity and the DGD per section goes to zero, namely

$$p_{\mathrm{SOPMD}}(x) = \frac{32\,x}{\pi\langle\tau\rangle^4}\mathrm{sech}\frac{4x}{\langle\tau\rangle^2}\tanh\frac{4x}{\langle\tau\rangle^2} \tag{38}$$

where $\langle\tau\rangle^2 = (8/3\pi)\langle\tau^2\rangle$ is the square of the average DGD.

Fig. 6 shows the pdf of the DGD, obtained with pure first-order biasing, while Fig. 7 shows the pdf of second-order PMD, obtained with the pure second-order biasing technique. In both cases, the biasing strengths $\alpha = 1, 2, \ldots, 10$ were used, with 200 000 realizations each. The individual samples are combined by sorting the values of DGD and/or second-order PMD obtained from all of the simulations into 80 bins and adjusting the contribution *of each individual sample* for the bias using the weights and likelihood ratios, as explained previously. The solid lines show the analytical pdfs in the long-length regime [36], while the dashed lines in Fig. 6 show exact and asymptotic solutions from [34] for the case of scramblers. For illustration purposes, we also show the coefficient of variation, that is, the

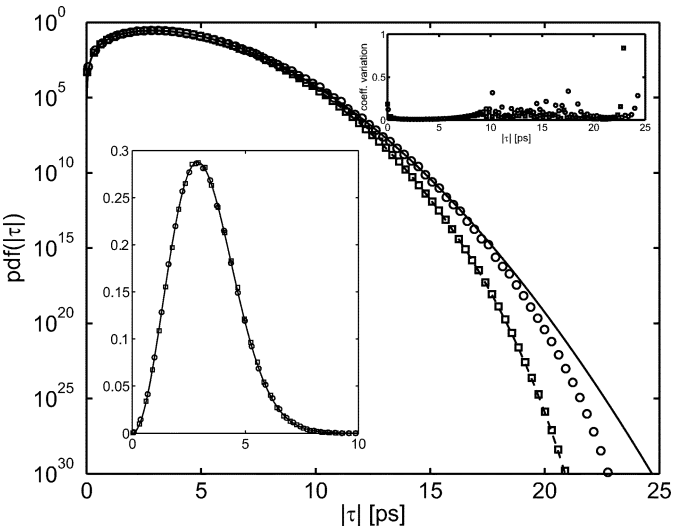

Fig. 7. Pdf of second-order PMD for a concatenation of 50 sections with 0.5-ps DGD each, using scramblers (squares) or waveplates (circles). The solid lines show the pdf for real fiber, from [36]. Bottom inset: the pdf on a linear scale. Top inset: the coefficient of variation. As in Fig. 6, a total of 2×10^6 MC realizations were used.

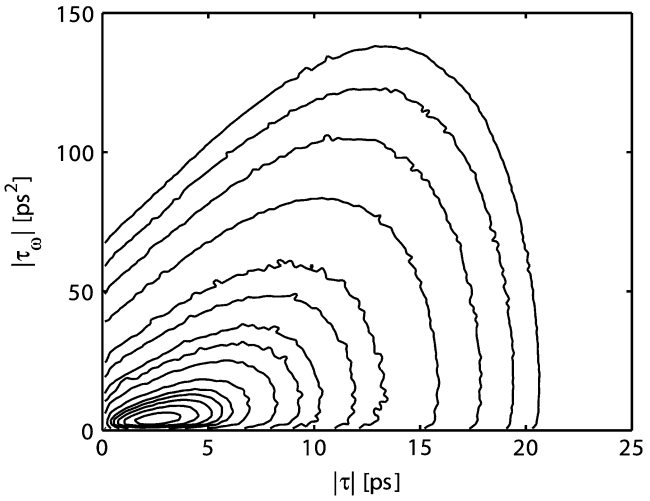

Fig. 8. Contour plots of the joint pdf for a concatenation of 50 sections with 0.5-ps DGD each and polarization scramblers. The contours are at 10^{-n} with $n = 30, 25, 20, 15, 10, 8, 6, 5, 4, 3, 2.5, 2.25, 2, 1.75$, and 1.5. A total of 10^6 MC realizations were used.

ratio of the standard deviation to the estimated value. The estimated standard deviation for the probability in each bin was calculated using the method described in Section II.

The numerically calculated pdf of the DGD for the case of scramblers agrees extremely well with the exact solution (dashed line under the squares). In all cases, the accuracy of the numerical results improves as the number of samples (and the number of bins) is increased. It should be noted, however, that extremely good results are obtained with the 2 000 000 MC samples employed here, as evident in Fig. 6. In particular, a good approximation is achieved even for probabilities well below 10^{-12}; to obtain comparable accuracy with straightforward MC simulations at this probability level, at least 10^{15} samples would be required. Thus, in this situation, IS provides a speed-up of more than ten orders of magnitude.

For both scramblers and waveplates, the pdfs deviate significantly from the real fiber distributions in the tails, since here the PMD is generated by an emulator with finite number of birefringent sections, which by necessity has finite maximums for both $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$. Similarly, such a device has unaccessible regions in the $(|\boldsymbol{\tau}|, |\boldsymbol{\tau}_\omega|)$ plane. The emulator should be chosen so that these regions where the pdfs are not close to those of real fiber are unimportant for determining the outage probability of the systems to be tested. We also note that the finite value of the birefringence correlation length suggests that the DGD distribution in fiber might also deviate from Maxwellian in the tails. In this case, however, the equivalent number of waveplates—which can be estimated from measurements of the fiber autocorrelation length [37]—is very large for typical communications distances, which indicates that the deviations from the continuum limit, if they happen, would occur at extremely low probability levels.

The emulator with rotatable sections yields better agreement with real fiber than the emulator with polarization scramblers, with regard to the distributions of both first- and second-order PMD. Note, however, that a concatenation of a small number of equal-length, rotatable birefringent sections is not a good model for real fiber due to artificial periodicities of the PMD vector's autocorrelation function (ACF) in the frequency domain [27], [38] (not shown here). Using nonequal section lengths can resolve this problem and produces an emulator with both good properties in the frequency domain [27]–[29], [34]–[39]. On the other hand, in [13], it was shown that PMD emulators with nonequal section lengths yield a worse agreement with the tails of the real fiber distributions of first- and second-order PMD than do emulators with equal section lengths. Therefore, when using nonequal section lengths, a larger number of sections must be used in order to obtain a comparable degree of tail approximation. One way to obviate this problem is to use PMD emulators with randomly varying sections lenghts which follow a Maxwellian distribution [40]. In this case, it is possible to obtain a very good degree of tail approximation with a very small number of sections and at the same time avoid the artificial periodicities of the ACF even when the mean lengths are the same [13]. The application of IS techniques to PMD generation devices employing Maxwellian-distributed length sections is described elsewhere [13].

C. Joint Distribution of First- and Second-Order PMD

It is also possible to use these methods to determine the joint pdf of first- and second-order PMD. When combined with measurements of receiver performance, the joint pdf can be used to estimate the total outage probability due to PMD [41].

In the long-length regime, the characteristic function of first- and second-order PMD was derived in [1] based on a physical model of fiber birefringence. To our knowledge, however, no exact analytical expression exists for the joint pdf. (An approximate pdf is available [1], but only for the case where second-order PMD is a small perturbation compared with first-order.) Similarly, the joint pdf for PMD emulators with a finite number of sections is not known analytically.

The joint pdf can be calculated with MC simulations, however. Fig. 8 shows the joint pdf of the magnitude of first- and second-order PMD (a two-dimensional reduction of the full 3-D joint pdf of first- and second-order PMD [1]) for an emulator

with polarization scramblers, as calculated with the multiple biasing technique. The joint pdf confirms why the optimal first- and second-order biasing methods are the correct ones to use if one is only interested in first- or SOPMD statistics, respectively. As seen from Fig. 8, for any fixed value of DGD (a vertical slice in Fig. 8), the maximum of the pdf occurs in a zone that falls within region 1 in Fig. 5. Similarly, for any given second-order PMD (a horizontal slice in Fig. 8), the maximum of the pdf occurs in a zone falling within region 2 in Fig. 5. The multiple biasing method, of course, is not limited to just these particular combinations of first- and second-order PMD but is able to produce any combination of the two.

D. Outage Probabilities

The end goal of the methods that have been discussed here is not merely to produce PMD statistics, of course, but rather to accurately estimate the probability of PMD-induced transmission penalties and system outages. The methods outlined here can accomplish this goal as long as PMD-induced power penalties are correlated with either the DGD or both the DGD and the second-order PMD. Several studies indicate that this is indeed the case [12], [14], [43].

To perform such calculations, it is necessary to transmit a sequence of pulses. Thus, the full frequency-dependent Jones matrix of the fiber must be evaluated. However, it is straightforward to use the multiple IS techniques described in this work with the concatenation equations in Jones space, rather than Stokes' space [3]. The application of the algorithms described here requires prescribing the orientation of the individual PMD vectors of all sections and, in the case of birefringent waveplates, the rotation angles. Once this information is known, the Jones matrix of each birefringent section can be immediately reconstructed using (20). One can then determine the full input–output transfer matrix in Jones space—which is needed for pulse propagation—by simply multiplying the individual transmission matrices of the individual sections. Alternatively, it is possible to use the algorithms to guide the construction of Jones matrices for each birefringent section and then, rather than combining them together, use them separately as the coarse-step part of a more complete fiber propagation model [30], e.g., including chromatic dispersion and nonlinear effects.

The simplest implementation of the method in this context is to bias at one particular frequency, such as the central frequency of the channel of interest, but to construct the Jones matrices at all frequencies. The PMD will only be large within the fiber's autocorrelation bandwidth [42]. If large PMD is desired at several frequencies, one can bias at each of these frequencies, one at a time. This will not typically make the PMD simultaneously large at two widely separate frequencies, however, such an event has a probability that is roughly the square of the probability of the PMD being large at any single frequency. If one specifically wants the PMD to be large at several frequencies simultaneously, it is possible to use birefringent sections with identical lengths and a specified free-spectral range to artificially make the DGD spectrum periodic.

Regardless of the specific implementation used, once large PMD has been generated and its effects upon pulse propagation has been simulated, the last step is to use the PMD-distorted pulses as the input to a receiver model [14], [43]. As long as the receiver's performance is correlated with the PMD, large PMD will tend to generate large penalties [41]. Throughout the process, the likelihood ratios will keep track of the biasing used to generate these large penalties, and thus the result will be rare PMD-induced penalties with, automatically, their associated probabilities. This procedure has been used in [14] and [43]–[45]. Because the details of its implementation have been better described elsewhere, we will not repeat them here.

E. Experimental Implementation

Experimental PMD emulators employing a concatenation of birefringent elements have been built [31]–[33], and it is possible that an emulator could be constructed to use the importance-sampled algorithms described in Sections III and IV. Essentially, in order for it to be possible to use the algorithm, it is necessary that the emulator be stable, repeatable, and predictable. The biggest difficulty associated with an experimental implementation of the algorithms is the requirement that the orientation of the PMD vector and its frequency derivative be known after each birefringent section. One possible way of doing this would be to monitor the PMD vector after each birefringent section; such an implementation would obviously be prohibitively cumbersome and expensive.

If the PMD vector is not monitored, then the next alternative is for each element of the emulator to be sufficiently well characterized that the PMD vector can be calculated if the orientations of the different birefringent elements is specified. Here, the problem is that the differential phase retardations φ_{n+1} in the Müller matrix R_{n+1} must then be known, otherwise the Müller matrix will rotate the PMD vector to an unknown location on the Poincaré sphere, and it will then be impossible to properly determine the preferential orientation of the next differential PMD vector. As mentioned previously, each beat length of a birefringent section generates a 2π retardation. Because sections with significant DGDs will be many beat lengths long, this means that the section lengths must be precise to within a small fraction of a beat length and that these sections lengths must be stable with respect to temperature and other environmental fluctuations. Fortunately, it has been demonstrated that it is possible to satisfy all of these constraints experimentally [46]. Finally, we should note that a hardware PMD emulation device that makes use of multiple IS through variance scaling (a limited version of IS that works for a small number of birefringent sections) has been recently demonstrated and used in experiments to characterize system behavior [19].

Appendix I
Stratified Importance Sampling

As mentioned in Section II, several strategies are possible for choosing the weights in multiple IS. In Section II-E, we presented the balance heuristic, which is the choice we used in this work. Other weighting choices (such as the cutoff heuristic or the power heuristic, for example) are discussed in [23] and [24]. Here, we present another strategy, which is not discussed in those references and which was used in [14]. First, we recall the maximum heuristic from [23] and [24]

$$w_j(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } q_j(\boldsymbol{\theta}) = q_{\max}(\boldsymbol{\theta}), \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

where, as before, $q_j(\boldsymbol{\theta}) = M_j p_j^*(\boldsymbol{\theta})$ is proportional to the expected number of hits from $p_j^*(\boldsymbol{\theta})$. Here, samples from the jth distribution are used only in the region where $q_j(\boldsymbol{\theta}) = q_{\max}(\boldsymbol{\theta})$. This strategy is conceptually simple; it does not tend to work as well as the balance heuristic [23], however. Intuitively, this is because too many samples are thrown away.

A similar strategy is stratified IS, which is a variant of stratified MC [6]. In this case, one partitions the event space Θ into a number of disjoint subsets $\{\Theta_1, \ldots, \Theta_J\}$. For example, the disjoint sets Θ_j could be chosen by prescribing specific ranges of a control variable $g(\boldsymbol{\theta})$; that is, $\Theta_j = \{\boldsymbol{\theta} \in \Theta | g_j \le g(\boldsymbol{\theta}) < g_{j+1}\}$, with $g_0 = \min[g(\boldsymbol{\theta})]$ and $g_J = \max[g(\boldsymbol{\theta})]$. The weights $w_j(\boldsymbol{\theta})$ are then chosen according to

$$w_j(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \Theta_j, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

Thus, each distribution $p_j^*(\boldsymbol{\theta})$ only contributes to the estimator when the samples it generates fall in a predetermined region of sample space. This strategy was used in studies of PMD compensation [14]; in this case, the control variable was the total DGD. Note, however, that unlike the maximum heuristic, this choice might not use the distribution with $q_j(\boldsymbol{\theta}) = q_{\max}$ in any given region of sample space.

Appendix II
Likelihood Ratios for Correlated Sections

The algorithms described in Sections III and IV work by selecting the angles θ_{n+1} for the $(n+1)$st section with respect to a biasing direction $\mathbf{b}^{(n)}$, which depends upon the value of the PMD vector and its frequency derivative after the nth section $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\tau}_\omega^{(n)}$, respectively. These values depend on the orientations of all the previous sections and in particular upon the angles $\{\theta_1, \ldots, \theta_n\}$. Thus, although the different angles are independent of one another, the choice of each successive PMD vector $\Delta\boldsymbol{\tau}_{n+1}$ depends upon the previous choices. Because they are not independent of one another, it is not immediately clear that the overall likelihood ratio should be simply the product of the individual likelihood ratios for each section. Nevertheless, we show here that this is the case.

Suppose we have a set of random samples $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$. In our case, the $\mathbf{z}_n$'s are the individual PMD vectors of each section, i.e., $\mathbf{z}_n = \Delta\boldsymbol{\tau}_n$. Using recursively the law of conditional probabilities, namely $p(AB) = p(A|B)p(B)$, we can express the joint probability distribution as

$$\begin{aligned} p(\mathbf{z}_1, \ldots, \mathbf{z}_N) &= p(\mathbf{z}_N \,|\, \mathbf{z}_{N-1}, \mathbf{z}_{N-2}, \ldots, \mathbf{z}_2, \mathbf{z}_1) \\ &\quad \times p(\mathbf{z}_{n-1} \,|\, \mathbf{z}_{N-2}, \ldots, \mathbf{z}_2, \mathbf{z}_1) \cdots p(\mathbf{z}_2 \,|\, \mathbf{z}_1) p(\mathbf{z}_1). \end{aligned} \tag{41}$$

Thus, even if the random variables are not independent and one cannot factor the joint pdf as $p(\mathbf{z}_1, \ldots, \mathbf{z}_N) = p(\mathbf{z}_1) \ldots p(\mathbf{z}_N)$ (i.e., into the product of individual distributions), it is still possible to write the joint pdf as a product if each random variable is selected using a conditional probability distribution.

This is precisely the situation for the algorithms described in Sections III and IV, since the choice of the individual PMD vector for each new section depends on the previous PMD vectors *only through the value of the total PMD vector and its frequency derivative up to that point*. That is, the process is a Markov chain: the transitional probabilities only depend upon the current state of the system, not on the details of its history; in other words, the system is memoryless. Once the orientation of the total PMD vector and its derivative have been determined (and with them the preferential biasing direction $\mathbf{b}^{(n)}$), $\Delta\boldsymbol{\tau}_n$ is chosen relative to $\mathbf{b}^{(n)}$ by selecting the distribution for the angle θ_n to be either (25) (for scramblers) or (26) (for waveplates). Since both the unbiased and biased cases can be constructed in this way, both the numerator and denominator of the overall likelihood ratio can be factored, and therefore the overall likelihood ratio can be expressed as product of individual likelihood ratios, as shown in (24).

Appendix III
Deterministic Maximum of Second-Order PMD for the Continuum Equations

As mentioned in Section IV-B, the choice of $\mathbf{b}(z)$ that maximizes the magnitude of second-order PMD can be found using calculus of variations on the exact solution of the continuum model (i.e., (33)). First, note that $|\mathbf{b}(z)|$ can be scaled out of the equations by making the change of variable $dz'/dz = |\mathbf{b}(z)|$: the result is the same equations, where now $|\mathbf{b}(z)| = 1$. Thus, the case of equal-length sections can be considered without loss of generality, and all other cases can be obtained from it using a simple distance scaling. Second, note that, from (31), $d(|\boldsymbol{\tau}_\omega|^2)/dz = d(\tau_{\omega,||}^2 + \tau_{\omega,\perp}^2)/dz = 2b_3\,\tau_{\omega,\perp}\tau$. Thus, b_2 does not directly contribute to increase the magnitude of second-order PMD; rather, as seen from (31b) and (31b), (31c), it induces a rotation of $\boldsymbol{\tau}_\omega$ about $\hat{\mathbf{u}}_3$. Since the maximum $\tau_{\omega,\perp}$ under such a rotation occurs when $\tau_{\omega,||} = 0$, we expect the maximum second-order PMD to be generated for $b_2 = 0$, i.e., when $\mathbf{b}(z)$ lies in the $\hat{\mathbf{u}}_1$-$\hat{\mathbf{u}}_3$ plane. Thus, we look for the choice of $\mathbf{b}(z)$ such that, to first order, $\delta\tau_{\omega,\perp}(z_{\max}) = 0$ when $\mathbf{b}(z) \to \mathbf{b}(z) + \delta\mathbf{b}(z)$. For simplicity, we only present the calculations in the special case $|\mathbf{b}(z)| = b = \text{const}$ and for in-plane contributions, in this case

$$b_1(z) = b\cos\beta(z), \quad b_2(z) = 0, \quad b_3(z) = b\sin\beta(z). \tag{42}$$

In this framework, it is easy to show that, when $\beta(z) \to \beta(z) + \delta\beta(z)$, it is $\tau_{\omega,\perp}(z_{\max}) \to \tau_{\omega,\perp}(z_{\max}) + \delta\tau_{\omega,\perp}(z_{\max})$, where

$$\begin{aligned} \delta\tau_{\omega,\perp}(z_{\max}) = b \int_0^{z_{\max}} \Bigg(&\tau(z)\,\cos\beta(z) \\ &- b\,\sin\beta(z) \int_z^{z_{\max}} \sin\beta(z')\,dz' \Bigg)\,\delta\beta(z)\,dz \end{aligned} \tag{43}$$

where now $\tau(z) = b\int_0^z \cos\beta(z')\,dz'$. In order for the RHS of (43) to be zero for all $\delta\beta(z)$, it is necessary that the integrand be zero, i.e.,

$$\tau(z) = b\,\tan\beta(z) \int_z^{z_{\max}} \sin\beta(z')\,dz'. \tag{44}$$

Substituting $\tau(z)$ and differentiating (44) twice, we then get $d^2\beta/dz^2 = 0$, and by substituting back into (44), we can fix the integration constants

$$\beta(z) = \beta_{\max}\, z/z_{\max} \tag{45}$$

with $\beta_{\max} = \pi/2$. Thus, the maximum growth of second-order PMD is obtained for in-plane contributions with a linearly varying angle profile. The same solution, namely the in-plane linearly varying angle profile with $\beta_{\max} = \pi/2$, is obtained in the more general scenario in which all three components of $\mathbf{b}(z)$ are nonzero [13]. The calculations in this case, however, are considerably more involved, and thus for brevity they will not be reported here.

A similar calculation is possible in the case where one wants to maximize a linear combination of first- and second-order PMD. In this case, the solution is similar, the only difference being that the linearly varying angle profile has a different final target value, namely

$$\beta(z) = \beta_{\max} z / z_{\max} \tag{46}$$

with $\beta_{\max}$ varying from 0 to π. Furthermore, by integrating (33) using the above choice for $\mathbf{b}(z)$, it is possible to explicitly determine the final values of first- and second-order PMD that are obtained. Finally, we reiterate that these calculations have been generalized to the case where all three components of $\mathbf{b}(z)$ are nonzero, and to PMD emulators where the section lengths are also random [40], in which case they allow us to determine the most likely configurations that result in *any* value of first- and/or second-order PMD [13].

ACKNOWLEDGMENT

The authors would like to thank J. N. Damask, S. L. Fogal, A. Galtarossa, and I. T. Lima for many valuable discussions.

REFERENCES

[1] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightwave Technol.*, vol. 9, p. 1439, Nov. 1991.

[2] N. Gisin, R. Passy, J. C. Bishoff, and B. Perny, "Experimental investigations of the statistical properties of polarization mode dispersion in single-mode fibers," *IEEE Photon. Technol. Lett.*, vol. 5, p. 819, July 1993.

[3] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization-mode dispersion in optical fibers," in *Proc. Nat. Acad. Sci.*, vol. 97, 2000, p. 4541.

[4] P. Bratley, B. L. Fox, and L. E. Schrage, *A Guide to Simulation*. New York, NY: Springer-Verlag, 1987.

[5] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*. New York: Plenum, 1992.

[6] G. S. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. New York, NY: Springer-Verlag, 1996.

[7] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* Cambridge, MA/New York, 2000.

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

[9] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, "On importance sampling in digital communications—Part I: Fundamentals," *IEEE J. Select. Areas Commun.*, vol. 11, p. 289, 1993.

[10] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE J. Select. Areas Commun.*, vol. 15, p. 597, May 1997.

[11] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization-mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, p. 310, Mar. 2002.

[12] 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, p. 1273, 1487, Sept. 2002.

[13] G. Biondini and W. L. Kath, "PMD emulation with Maxwellian length sections and importance sampling," *IEEE Photon. Technol. Lett.*, vol. 16, p. 789, Mar. 2004.

[14] I. T. Lima, G. Biondini, B. S. Marks, W. L. Kath, and C. R. Menyuk, "Analysis of polarization-mode dispersion compensators using importance sampling," *IEEE Photon. Technol. Lett.*, vol. 14, p. 627, Mar. 2002.

[15] S. L. Fogal, G. Biondini, and W. L. Kath, "Importance-sampled pulse broadening before and after PMD compensation," in *Proc. OFC'2002*, pp. 372–374. Paper ThA2.

[16] D. Yevick, "Multicanonical communication system modeling—Application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 1512–1514, Nov. 2002.

[17] ——, "The accuracy of multicanonical system models," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 224–226, Feb. 2003.

[18] M. Shtaif, "The Brownian-bridge method for simulating polarization mode dispersion in optical communications systems," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 51–53, Jan. 2003.

[19] L. S. Yan, M. C. Hauer, P. Ebrahimi, Y. Wang, Y. Q. Shi, X. S. Yao, A. E. Willner, and W. L. Kath, "Measurement of Q degradation due to polarization mode dispersion using multiple importance sampling," *Electron. Lett.*, vol. 39, pp. 974–975, 2003.

[20] J. S. Sadowsky and J. A. Bucklew, "On large deviations theory and asymptotically efficient Monte-Carlo estimation," *IEEE Trans. Inform. Theory*, vol. 36, p. 579, May 1990.

[21] P. W. Glynn and W. Whitt, "The asymptotic efficiency of simulation estimators," *Oper. Res.*, vol. 40, p. 505, 1992.

[22] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. New York: Springer, 1993.

[23] E. Veach, "Robust Monte Carlo Methods for Light Transport Simulation," Ph.D. Thesis, Stanford University, 1997.

[24] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. Amer. Stat. Assoc.*, vol. 95, p. 135, 2000.

[25] A. Galtarossa, L. Palmieri, M. Schiano, and T. Tambosso, "Statistical characterization of fiber random birefringence," *Opt. Lett.*, vol. 25, p. 1322, 2000.

[26] 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. Lightwave Technol.*, vol. 15, p. 1735, Sept. 1997.

[27] R. Khosravani, I. T. Lima, P. Ebrahimi, E. Ibragimov, A. E. Willner, and C. R. Menyuk, "Time and frequency domain characteristics of polarization-mode dispersion emulators," *IEEE Photon. Technol. Lett.*, vol. 13, p. 127, Feb. 2001.

[28] J. N. Damask, G. J. Simer, K. B. Rochford, and P. R. Myers, "Demonstration of a programmable PMD Source," *IEEE Photon. Technol. Lett.*, vol. 15, p. 296, 2004.

[29] J. N. Damask, G. J. Simer, K. B. Rochford, and P. R. Myers, "Demonstration of a programmable PMD source," *IEEE Photon. Technol. Lett.*, vol. 15, p. 296, Feb. 2003.

[30] P. Ciprut, B. Gisin, N. Gisin, R. Passy, J. P. Von der Weid, F. Prieto, and C. W. Zimmer, "Second-order polarization-mode dispersion: Impact on analog and digital transmissions," *J. Lightwave Technol.*, vol. 16, p. 757, May 1998.

[31] C. Francia, F. Bruyére, D. Penninckx, and M. Chbat, "PMD second-order effects on pulse propagation in single-mode optical fibers," *IEEE Photon. Technol. Lett.*, vol. 10, p. 1739, Dec. 1998.

[32] W. Shieh, "On the second-order approximation of PMD," *IEEE Photon. Tech. Lett.*, vol. 12, p. 290, 2000.

[33] Q. Yu and A. E. Willner, "Performance limits of first-order PMD compensation using fixed and variable DGD elements," *IEEE Photon. Technol. Lett.*, vol. 14, p. 304, Mar. 2002.

[34] B. D. Hughes, *Random Walks and Random Environments*. Oxford, U.K.: Clarendon, 1995.

[35] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *J. Lightwave Technol.*, vol. 19, p. 324, Mar. 2001.

[36] G. J. Foschini, L. E. Nelson, R. M. Jopson, and H. Kogelnik, "Probability densities of second-order polarization-mode dispersion including polarization-dependent chromatic dispersion," *IEEE Photon. Technol. Lett.*, vol. 12, p. 293, Mar. 2000.

[37] A. Galtarossa, L. Palmieri, M. Schiano, and T. Tambosso, "Measurements of birefringence correlation length in long single-mode fibers," *Opt. Lett.*, vol. 26, p. 962, 2001.

[38] I. T. Lima, R. Khoshravani, P. Ebrahimi, E. Ibragimov, A. E. Willner, and C. R. Menyuk, "Comparison of polarization mode dispersion emulators," *J. Lightwave Technol.*, vol. 19, p. 1872, Dec. 2001.

[39] B. S. Marks, I. T. Lima, and C. R. Menyuk, "Autocorrelation function for polarization mode dispersion emulators with rotators," *Opt. Lett.*, vol. 27, p. 1150, 2002.

[40] J. H. Lee, M. S. Kim, and Y. C. Chung, "Statistical PMD emulator using variable DGD elements," *IEEE Photon. Technol. Lett.*, vol. 15, p. 54, Jan. 2003.

[41] J. N. Damask, G. Gray, P. Leo, G. Simer, K. Rochford, and D. Veasey, "Method to measure and estimate total outage probability for PMD-impaired systems," *IEEE Photon. Technol. Lett.*, vol. 15, p. 48, Jan. 2003.

[42] M. Karlsson and J. Brentel, "Autocorrelation function of the polarization-mode dispersion vector," *Opt. Lett.*, vol. 24, p. 465, 1999.

[43] A. O. Lima, I. T. Lima Jr., T. Adali, and C. R. Menyuk, "A novel polarization diversity receiver for PMD mitigation," *IEEE Photon. Technol. Lett.*, vol. 14, p. 465, Apr. 2002.

[44] A. O. Lima, I. T. Lima, C. R. Menyuk, G. Biondini, B. S. Marks, and W. L. Kath, "Statistical analysis of PMD compensators with multiple importance sampling," *Photon. Technol. Lett.*, vol. 15, no. 12, Dec. 2003. to appear.

[45] I. T. Lima, A. O. Lima, G. Biondini, C. R. Menyuk, and W. L. Kath, "A comparative study of single-section polarization-mode dispersion compensators," *J. Lightwave Technol.*, vol. 22, pp. 1023–1032, Apr. 2004.

[46] J. N. Damask, P. R. Myers, A. Boschci, and G. J. Simer, "Demonstration of a coherent PMD source," *IEEE Photon. Technol. Lett.*, vol. 15, p. 1612, 2004.

Gino Biondini was born November 18, 1966 in Perugia, Italy. He received the Laurea in physics and the Doctorate in theoretical physics from the University of Perugia, Perugia, Italy, in 1991 and 1997, respectively.

He was a Postdoctoral Research Associate at the University of Colorado from 1997 to 1999, and from 1999 to 2001, he was a Research Assistant Professor at Northwestern University, Evanston, IL, where he started working on the problem of rare events in optical fiber communications. Since 2001, he has been a Zassenhaus Assistant Professor in the Mathematics Department at Ohio State University, Columbus. Starting in fall 2004, he will join the Mathematics Department of the State University of New York (SUNY), Buffalo. He is the author of more than 30 refereed publications and one patent application, and he has given numerous invited presentations at national and international conferences. His research interests include nonlinear wave equations, solitons and integrable systems, nonlinear optics and optical fiber communications, applied probability and stochastic processes, and variance reduction techniques.

William L. Kath received the B.S. degree in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1978 and the Ph.D. degree in applied mathematics from the California Institute of Technology (Caltech), Pasadena, in 1981.

He was with Caltech until 1984, where he was an NSF Postdoctoral Fellow (1981–1982), and a Von Kármán Instructor of Applied Mathematics (1982–1984). He joined the Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, in 1984, and from 1985 to 1990, he was an NSF Presidential Young Investigator. His research interests include optical fibers and waveguides, solitons, polarization-mode dispersion, parametric amplification, and computational neuroscience. He is the author or coauthor of more than 130 publications.

He is a Member of the Optical Society of America (OSA) and the Society for Industrial and Applied Mathematics.

Curtis R. Menyuk (F'98) was born March 26, 1954. He received the B.S. and M.S. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1976 and the Ph.D. from the University of California at Los Angeles (UCLA) in 1981.

He has worked as a Research Associate at the University of Maryland, College Park, and at Science Applications International Corporation, McLean, VA. In 1986, he became an Associate Professor in the Department of Electrical Engineering at the University of Maryland Baltimore County (UMBC), and he was the founding member of this department. In 1993, he was promoted to Professor. He was on partial leave from UMBC from fall 1996 until fall 2002. From 1996 to 2001, he worked part-time for the Department of Defense, codirecting the optical networking program at the Department of Defense (DoD) Laboratory for Telecommunications Sciences, Adelphi, MD, from 1999 to 2001. From 2001 to 2002, he was Chief Scientist at PhotonEx Corporation. He has authored or coauthored more than 180 archival journal publications as well as numerous other publications and presentations. He has also edited two books. For the last 17 years, his primary research area has been theoretical and computational studies of fiber-optic communications. The equations and algorithms that he and his research group at UMBC have developed to model optical fiber transmission systems are used extensively in the telecommunications industry.

Dr. Menyuk is a Fellow of the Optical Society of America (OSA). He is a Member of the Society for Industrial and Applied Mathematics and the American Physical Society and a former UMBC Presidential Research Professor.