J. Opt. Fiber Commun. Rep. 1, 14–31 (2004)
© 2004 Springer Science + Business Media Inc.
DOI: 10.1007/s10297-004-0007-z
Editors: Curtis R. Menyuk, Andrea Galtarossa

JOURNAL OF
Optical and Fiber
Communications Reports

# Applications of importance sampling to polarization mode dispersion

Gino Biondini[1] and William L. Kath[2]

[1] Department of Mathematics, State University of New York
Buffalo, NY 14260 USA
Email: biondini@buffalo.edu

[2] Department of Engineering Sciences and Applied Mathematics
Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208 USA
Email: kath@northwestern.edu

**Abstract.** We describe the application of importance sampling to Monte-Carlo simulations of polarization-mode dispersion (PMD) in optical fibers. The method allows rare PMD events to be simulated much more efficiently than with standard Monte-Carlo methods, thus making it possible to assess the effect of PMD on system outage probabilities at realistic bit error ratios.

## 1. Introduction

Polarization-mode dispersion (PMD) has emerged as a major impairment in both terrestrial and undersea wavelength-division-multiplexed (WDM) systems. To first order in frequency, PMD splits a pulse between the fast and the slow axis of the fiber; at the same time, higher orders of PMD induce depolarization and polarization-dependent chromatic dispersion. At high bit rates, all of these effects lead to unacceptable transmission penalties. A key difficulty with PMD is its stochastic nature. PMD is caused by random variations of an optical fiber's birefringence with distance. In addition, the penalties it produces change randomly over time as the ambient temperature and other environmental parameters vary. In system design, a maximum power penalty (typically 1 dB) is usually assigned to PMD, and one demands that the outage probability (that is, the probability of the PMD-induced penalty exceeding this allowed value) is very small (typically a minute per year, $10^{-6}$ or less). Because of this small probability,

it is very difficult to use either Monte-Carlo simulations or laboratory experiments to determine a system's outage probability, due to the extremely large number of system configurations that must be explored in order to obtain a reliable estimate.

A measure of PMD is provided by the PMD vector $\boldsymbol{\tau}$ [1, 2, 3], the magnitude of which is the differential group delay (DGD). 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. Of course, the PMD vector is frequency-dependent, and therefore 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 polarization-dependent chromatic dispersion (PCD), is known to produce additional system penalties [4, 5]. 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 second-order PMD to become the primary cause of system penalties [6, 7, 8, 9].

In the absence of effective tools for calculating outage probabilities, system designers have resorted to stopgap techniques. One such technique is to produce artificially large DGD values, determine the penalties at these large DGDs, and then weight the results using the DGD's well-known Maxwellian probability distribution [1]. A fundamental problem with this method, however, is that there is no direct relationship between the DGD (and/or second-order PMD) and the power penalty. In addition, it is possible for different configurations to give both the same DGD and the same second-order PMD, but not contribute equally to the penalty (because of different higher-order PMD), and therefore they should be weighted differently in calculations of outage probability.

The technique of *importance sampling* (IS) [10, 11, 12], which is one of a family of methods known as variance reduction techniques [13, 14], addresses these difficulties and provides a tool that can be used in numerical simulations — and, in principle, in experiments — to accurately estimate PMD-induced system penalties. The technique allows low probability events to be efficiently simulated by enabling one to concentrate Monte-Carlo simulations in the most significant regions of interest in sample space. We have applied IS to the numerical calculation of PMD-induced effects generated by a concatenation of birefringent sections [15, 16, 17]. In this case, the regions of interest in the sample space are the configurations that lead to large values of DGD and/or second-order PMD. This method has also been applied to numerically calculate PMD-induced transmission penalties [7, 9, 18]. Here we give a complete description of the technique and its application.

In Sections 2–4 we introduce the technique by first discussing some basic principles and by considering the simpler case in which large DGD values are the events of interest. Of course, in many practical situations it is necessary to also consider the frequency dependence of the PMD vector. Thus, in Section 5 we present a multiple biasing technique that allows one to target arbitrary amounts of first- and second-order PMD. Use of this composite biasing method in importance-sampled Monte-Carlo simulations yields a more comprehensive determination of PMD-induced system penalties than first-order biasing alone. Finally, in Sections 6–7 we discuss some technical issues that are needed to implement the method. To demonstrate the method, in Section 9 we calculate the probability density functions of the DGD, the magnitude of the second-order PMD vector and their joint pdf for a device composed by a concatenation of birefringent sections.

## 2. The PMD Concatenation Equations

A standard technique for simulating PMD effects is the coarse-step method [19], which approximates the continuous birefringence variations present in real fibers by a concatenation of fixed length birefringent sections. Many experimental PMD generation techniques also employ a concatenation of birefringent elements, such as high-birefringence fibers [20] or birefringent waveplates [21], connected by either polarization scramblers (e.g., polarization controllers [20]) or rotatable connectors [21].

In all such cases, the total PMD vector after the $(n+1)$-st fiber section can be obtained from the PMD concatenation equation [3],

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

illustrated in Fig. 1. Here $\boldsymbol{\tau}_n$ is the PMD vector after $n$ sections and $\Delta\boldsymbol{\tau}_{n+1}$ is the differential contribution to the PMD vector of the $(n+1)$-st section. For fixed length sections, the magnitude of $\Delta\boldsymbol{\tau}_{n+1}$ is fixed, and only its direction varies. For linearly birefringent elements, $\Delta\boldsymbol{\tau}_{n+1}$ lies on the equatorial plane of the Poincaré sphere [22]. The Müller matrix $\mathsf{R}_{n+1} = \exp\left[\varphi_{n+1}(\hat{\mathbf{r}}_{n+1}\times)\right]$ represents a rotation through an angle $\varphi_{n+1}$ about the axis $\hat{\mathbf{r}}_{n+1}$. This rotation is frequency dependent; for linearly birefringent elements, $\hat{\mathbf{r}}_{n+1} = \Delta\boldsymbol{\tau}_{n+1}/|\Delta\boldsymbol{\tau}_{n+1}|$ and $\varphi_{n+1}(\omega) = 2b'\ell_{n+1}\omega = \varphi_{n+1}(\omega_0) + 2b'\ell_{n+1}(\omega-\omega_0)$, where $b'$ is the birefringence strength of each element and $\ell_{n+1}$ is its length [3]. A similar concatenation equation holds for the frequency derivative of the PMD vector, $\boldsymbol{\tau}_\omega$ [3]:

$$\boldsymbol{\tau}_\omega^{(n+1)} = \mathsf{R}_{n+1}\boldsymbol{\tau}_\omega^{(n)} + \Delta\boldsymbol{\tau}_{n+1}\times\boldsymbol{\tau}^{(n+1)}\,. \tag{2}$$

If linearly birefringent elements are used, $\mathsf{R}_{n+1}\Delta\boldsymbol{\tau}_{n+1} = \Delta\boldsymbol{\tau}_{n+1}$ and Eqs. (1)–(2) can be written as

$$\boldsymbol{\tau}^{(n+1)} = \mathsf{R}_{n+1}\big(\boldsymbol{\tau}^{(n)} + \Delta\boldsymbol{\tau}_{n+1}\big), \tag{3}$$

$$\boldsymbol{\tau}_\omega^{(n+1)} = \mathsf{R}_{n+1}\big(\boldsymbol{\tau}_\omega^{(n)} + \Delta\boldsymbol{\tau}_{n+1}\times\boldsymbol{\tau}^{(n)}\big). \tag{4}$$

When polarization controllers are present, an additional rotation matrix $\mathsf{Q}_{n+1}$ preceeds $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\tau}_\omega^{(n)}$ in Eqs. (3)–(4). It is possible to factor out this rotation $\mathsf{Q}_{n+1}$ from the concatenation equations: the resulting equations are formally equivalent to Eqs. (3)–(4) 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. This property facilitates the implementation of IS with polarization scramblers. In the following, we refer to the situation where there are no polarization controllers as the case of rotatable waveplates.

## 3. Importance Sampling

In order to apply IS to the PMD concatenation Eqs. (1)–(2) we need to target the rare events in which the DGD and/or second-order PMD assume values much larger than their mean. Here, for brevity, we only present the implementation of importance sampling for PMD emulators which employ fixed-length birefringent sections [15, 16]. Importance sampling methods for PMD emulators with variable-length sections are

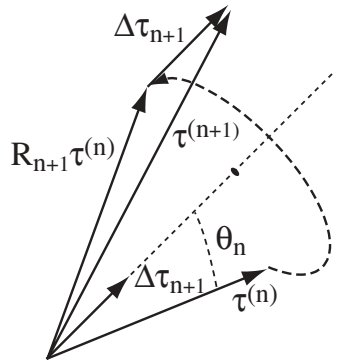


**Fig. 1.** Graphical representation of the first-order PMD concatentation, Eq. (1).

discussed in [17]. When fixed-length birefringent sections are employed, we have found that the appropriate variables to control are the polar angles $\theta_n$ between the PMD vector after the first $n$ sections, $\boldsymbol{\tau}_n$, and the differential contribution of the $(n+1)$-st section, $\Delta\boldsymbol{\tau}_{n+1}$, as shown in Fig. 1. (Of course, a second set of angles, giving the orientation of $\Delta\boldsymbol{\tau}_{n+1}$ about $\boldsymbol{\tau}_{n+1}$ at fixed $\theta_n$, is needed to uniquely specify each birefringent section. These angles do not need to be controlled and can be left uniformly distributed.) For example, the configurations that lead to these large DGD values are the ones in which the individual contributions to the PMD vector from each section tend to be aligned with each other. These configurations are the ones where the angles $\theta_n$ tend to be closer to zero than on average.

Suppose we are interested in determining the probability $P$ that a random variable which depends upon the angles $(\theta_1, \theta_2, \ldots \theta_N) \equiv \boldsymbol{\theta}$ falls in a given range, where $N$ is the total number of sections. In our first example we will target the total DGD, $|\boldsymbol{\tau}_N|$, but later on we will consider an arbitrary combination of DGD and second-order PMD. The method can be applied to any random variable, such as the amount of pulse broadening, the power penalty, or any combination thereof. The probability $P$ can be represented as the expectation value of an indicator function $I(\boldsymbol{\theta})$, where $I = 1$ if the random variable of interest falls in the prescribed range and $I = 0$ otherwise. That is, $P$ is represented by the $N$-dimensional integral

$$P = \int_{\Theta} I(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}\,, \tag{5}$$

where the $p(\boldsymbol{\theta})$ is the unbiased joint probability distribution function for the angles and $\Theta$ represents the configuration space. When using importance sampling, we rewrite the integrand in Eq. (5) as $I(\boldsymbol{\theta})p(\boldsymbol{\theta}) = I(\boldsymbol{\theta})L(\boldsymbol{\theta})p^*(\boldsymbol{\theta})$, where $p^*(\boldsymbol{\theta})$ is the biasing distribution, and where $L(\boldsymbol{\theta}) = p(\boldsymbol{\theta})/p^*(\boldsymbol{\theta})$ is the importance sampling likelihood ratio [10, 11, 12]. We then estimate the corresponding integral through Monte Carlo simulations; that is, we write an importance-sampled Monte-Carlo estimate of the above probability $P$ as [10, 11, 12]

$$\hat{P} = \frac{1}{M} \sum_{m=1}^{M} I(\boldsymbol{\theta}_m) L(\boldsymbol{\theta}_m)\,, \tag{6}$$

where $M$ is the total number of trials, and where the samples $\boldsymbol{\theta}_m$ are drawn according to the biased distribution $p^*(\boldsymbol{\theta})$. If $p^*(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta})$ (that is, for unbiased Monte Carlo simulations), Eq. (6) simply is the relative frequency of trials falling in the range of interest. The problem with this choice is that, when simulating low probability events, an exceedingly large number of samples is necessary in order for the desired events to occur, and an even larger number is required in order to obtain an accurate estimate. 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 probability.

Not all biasing schemes are equivalent, of course. In order to obtain a fair estimator for the desired quantity $P$, the biased probability distribution $p^*(\boldsymbol{\theta})$ should generate all of the configurations that contribute to the final result. The best methods are the ones that require the smallest numbers of realizations to be performed for all of the significant regions of sample space to be visited; such methods are called asymptotically efficient [23]. The algorithms that we have chosen to obtain targeted values of first- and/or second-order PMD work by orienting the sections so that the successive differential PMD vectors are preferentially aligned in certain directions with respect to the first- and second-order PMD vectors up to that point. These algorithms appear to be asymptotically efficient, as evidenced by the numerical results to be presented later, which show that the tails of the probability distribution can be reached with approximately the same amount of effort as the main part of the distribution.

## 4. IS for the DGD

When polarization scramblers are employed, the successive differential PMD vectors $\Delta\boldsymbol{\tau}_{n+1}$ can be regarded as fixed, while the orientations of the previous PMD vectors $\boldsymbol{\tau}_n$ at the output of the polarization scrambler vary. We therefore bias the simulations by making the scramblers preferentially orient $\boldsymbol{\tau}_n$ near the direction of $\Delta\boldsymbol{\tau}_{n+1}$. More specifically, we set the polarization scramblers so that $\theta_n$ (the angle between the PMD vector at the output of the $n$-th scrambler and the $(n+1)$-st differential PMD vector) is biased towards zero. This choice does not uniquely determine the orientation of the PMD vector at the scrambler output, of course, because $\boldsymbol{\tau}_n$ can still be rotated by an arbitrary amount about $\Delta\boldsymbol{\tau}_{n+1}$ while keeping $\theta_n$ constant. We assume that this additional rotational angle is uniformly distributed.

In the unbiased case, the angles $\theta_n$ are independent random variables, with $\cos\theta_n$ uniformly distributed in $[-1, 1]$. When applying importance sampling, 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 biased distribution used for the $\theta_n$; the effectiveness of the method does not seem to be very sensitive to the particular distribution used. The above choice yields the likelihood ratio as

$$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} \frac{p_1(\cos\theta_n)}{p_\alpha(\cos\theta_n)}\,, \tag{7}$$

where $p_\alpha(\cos\theta) = (\alpha/2)[(1+\cos\theta)/2]^{\alpha-1}$ [24]. The value $\alpha = 1$ reproduces the unbiased case (i.e., $\cos\theta$ is uniformly distributed), while increasing values of $\alpha$ bias

the configurations towards increasingly larger values of DGD, as illustrated in Figs. 2–3. Thus, each biasing distribution allows us to sample a different range of DGDs, and simulations with different values of $\alpha$ can be combined in order to cover the whole range of DGD values. An efficient strategy of combining results from different biasing distributions is described in Section 6.

We emphasize that different configurations with the same value of DGD (and/or second order PMD) can have quite different likelihood ratios, and thus their relative contribution to the final result can vary substantially. As a consequence, different configurations producing the same DGD and/or second-order PMD are expected to give very different contributions to the power penalty. This point is illustrated in Fig. 4, where we plot the distribution of likelihood ratios for some specific samples. Note that the horizontal axis uses a logarithmic scale, which implies a variation of almost three orders of magnitude of the likelihood ratios.

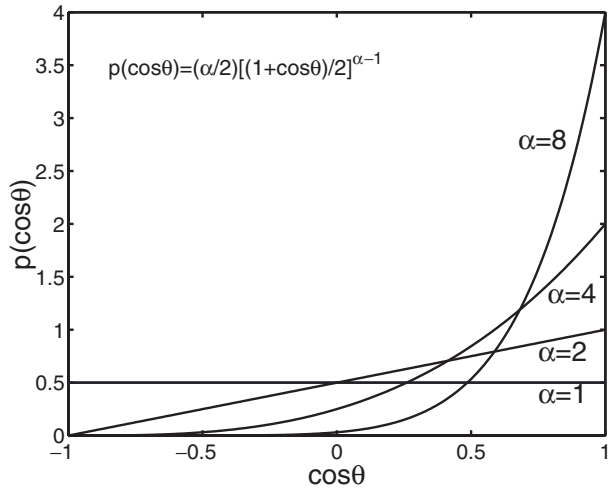


**Fig. 2.** The angular biasing distributions $p_\alpha(\cos\theta)$ for different values of $\alpha$. Note that the distribution for the angle $\theta$ is $p_\alpha(\theta) = p_\alpha(\cos\theta)\sin\theta$.

When using rotatable waveplates, the relative orientations between sections are the primary variables determining the total DGD. The biasing toward large DGD values is done by choosing the next differential PMD vector, $\Delta\boldsymbol{\tau}_{n+1}$, to be preferentially aligned with the projection of the total PMD vector obtained thus far, $\boldsymbol{\tau}_n$, onto the equatorial plane. Specifically, we choose $\Delta\boldsymbol{\tau}_{n+1}$ so that the angle $\theta_n$ between it and the projection of $\boldsymbol{\tau}_n$ is biased toward zero (mod $2\pi$). This is done, for example, by taking $\theta = \pi\left[1 + \mathrm{sgn}(2x-1)|2x-1|^{1/\alpha}\right]$, where $x$ is again uniformly distributed between 0 and 1. Other choices are possible, of course. If $\alpha = 1$ (unbiased case), $\theta$ is uniformly distributed between 0 and $2\pi$, while for $\alpha > 1$, $\theta$ is concentrated near 0 and $2\pi$.

If frequency dependence is desired, or in the case of rotatable waveplates, the differential phase retardations $\varphi_{n+1}$ in the Müller matrix $\mathsf{R}_{n+1}$ must also be specified. Recall that each beat length of a birefringent section generates a $2\pi$ retardation. In practice, sections with significant DGDs will be many beat lengths long, and unless

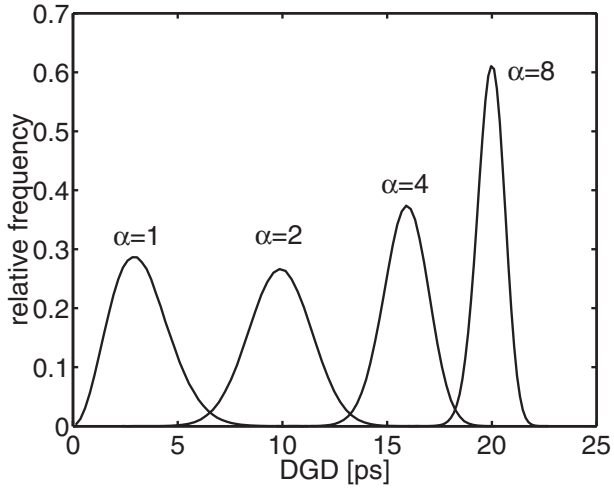


**Fig. 3.** The DGD frequency histograms for the biasing distributions in Fig. 2 for a concatenation of 50 birefringent elements with 0.5 ps DGD each (corresponding to a mean DGD of 3.26 ps) with polarization scramblers.

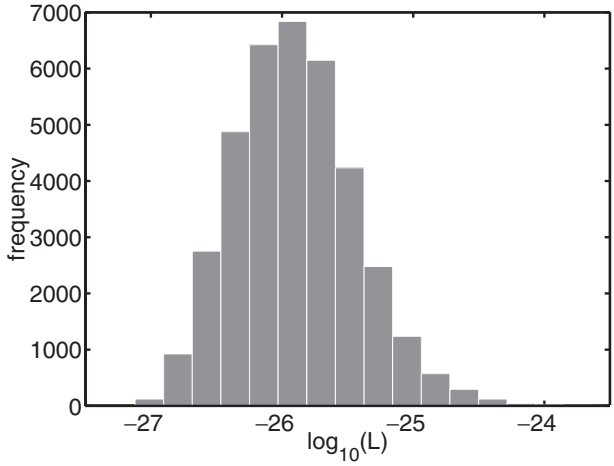


**Fig. 4.** The likelihood ratios and their relative frequency for samples with DGD values within a 0.2 ps interval around 20 ps and simultaneously with second-order PMD lying within a 6 $ps^2$ interval around 20 $ps^2$. A concatenation of 50 sections with 0.5 ps DGD each and polarization scramblers was used, as in Fig. 3.

the section lengths are precise to within a small fraction of a beat length, these retardations will vary from section to section. We therefore choose the retardation angles $\varphi_{n+1}(\omega_0)$ to be uniformly distributed between 0 and $2\pi$. (Alternatively, one could assume a random distribution of lengths $\ell_{n+1}$; 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. If, of course, the lengths are very precise, then specific phase retardations should be used.) In the case of rotatable waveplates, different configurations are then generated solely by rotating sections relative to one another. The results do not depend significantly upon the particular retardation angles used, except for certain clearly pathological cases such as identical angles with $\varphi_{n+1}$ equal to 0 or $\pi$.

## 5. IS for Both First- and Second-order PMD

When calculating the PMD-induced outage probability of a system using Monte-Carlo simulations, it is also important to generate sufficient statistics of the frequency derivative of the PMD vector, $\boldsymbol{\tau}_\omega$, which quantifies the second-order PMD. It is possible to use importance sampling with multiple biasing distributions to generate arbitrary combinations of first- and second-order PMD, 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.

As shown in the previous section, the appropriate variables to control when applying importance sampling 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 (e.g., the total DGD). In the following, we will characterize the vector $\mathbf{b}^{(n)}$ relative to the orthonormal frame of reference $\mathcal{U}$ formed by the unit vectors $\{\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \mathbf{u}_3^{(n)}\}$, where

$$\mathbf{u}_1^{(n)} = \boldsymbol{\tau}^{(n)}/|\boldsymbol{\tau}^{(n)}|\,, \qquad \mathbf{u}_2^{(n)} = \boldsymbol{\tau}_{\omega,\perp}^{(n)}/|\boldsymbol{\tau}_{\omega,\perp}^{(n)}|\,, \qquad \mathbf{u}_3^{(n)} = \mathbf{u}_1^{(n)} \times \mathbf{u}_2^{(n)}\,. \quad (8)$$

Here $\boldsymbol{\tau}_{\omega,\perp}^{(n)}$ is the component of $\boldsymbol{\tau}_\omega^{(n)}$ perpendicular to $\boldsymbol{\tau}^{(n)}$, as illustrated in Fig. 5. The magnitudes of $\boldsymbol{\tau}_{\omega,\parallel}$ (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 described earlier, first-order biasing is achieved by choosing $\Delta\boldsymbol{\tau}_{n+1}$ to be preferentially aligned with the previous PMD vector $\boldsymbol{\tau}^{(n)}$, i.e., using $\mathbf{b}^{(n)} = \mathbf{u}_1^{(n)}$. With multiple biasing strengths (a combination of different $\alpha$'s), this choice produces region 1 in Fig. 6. First-order biasing yields the largest values of DGD. It does not, however, produce very large values of $\boldsymbol{\tau}_\omega$, because when $\Delta\boldsymbol{\tau}_{n+1}$ is parallel to $\boldsymbol{\tau}_n$ the contribution from that section to $\boldsymbol{\tau}_\omega^{(n+1)}$ is zero. Thus, if all the $\Delta\boldsymbol{\tau}_{n+1}$ were parallel to $\boldsymbol{\tau}^{(n)}$, no second-order PMD would be produced. Random fluctuations add second-order PMD, but large values are not specifically targeted. Over a single section, one can easily show that the choice that maximizes the contribution to $\boldsymbol{\tau}_\omega^{(n+1)}$ is to align $\Delta\boldsymbol{\tau}_{n+1}$ with the direction of $\mathbf{u}_3^{(n)}$. The rate at which $\boldsymbol{\tau}_\omega^{(n+1)}$ increases, however, also depends on $\boldsymbol{\tau}^{(n)}$, and thus, when optimizing over many sections, the growth of $\boldsymbol{\tau}^{(n)}$ must also be considered.

When the number of sections in the emulator is large, we have found it useful to employ a continuum approximation to find the deterministic configuration that generates

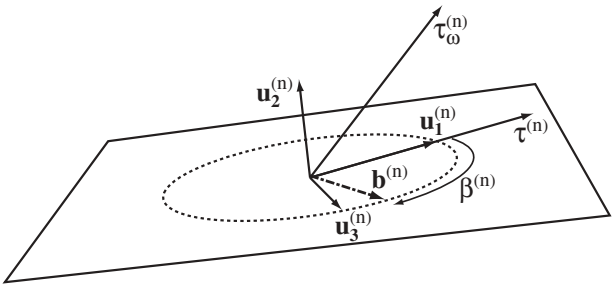


**Fig. 5.** Diagram showing the preferential direction **b** for biasing the simulations in the plane identified by $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\tau}^{(n)} \times \boldsymbol{\tau}_\omega^{(n)}$.

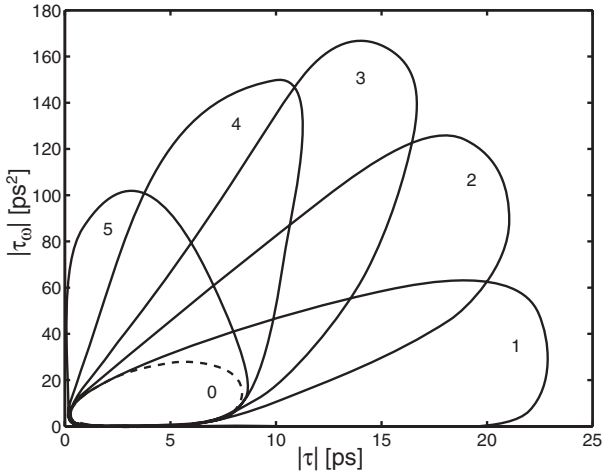


**Fig. 6.** The regions of the $|\boldsymbol{\tau}|$-$|\boldsymbol{\tau}_\omega|$ plane targeted by the various biasing methods. Region 1 corresponds to first-order biasing ($\beta = 0$), region 3 to pure second-order biasing ($\beta_{\max} = \pi/2$), and regions 2, 4 and 5 to $\beta_{\max} = \pi/4$, $3\pi/4$ and $\pi$, respectively. The dashed line shows the much smaller region obtained with unbiased runs. As in Figs. 3–4, 50 sections with 0.5 ps DGD each and polarization scramblers were used.

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 relative rate at which PMD is added by the birefringent sections. For simplicity, we concentrate upon the case $|\mathbf{b}(z)| = b =$const. In the continuum limit, for the case of polarization scramblers one obtains

$$\frac{d\tau}{dz} = b_1 \,, \tag{9}$$

$$\frac{d\tau_{\omega,\parallel}}{dz} = b_2 \, \frac{\tau_{\omega,\perp}}{\tau} \,, \tag{10}$$

$$\frac{d\tau_{\omega,\perp}}{dz} = b_3 \tau - b_2 \, \frac{\tau_{\omega,\parallel}}{\tau} \,, \tag{11}$$

where $(b_1, b_2, b_3)$ are the components of $\mathbf{b}$ with respect to $\mathcal{U}$. This system of equations can be solved exactly for any $\mathbf{b}(z)$:

$$\tau(z) = \int_0^z b_1(z')\, dz' \,, \tag{12}$$

$$\tau_{\omega,\parallel} = \int_0^z b_3(z')\tau(z') \sin[B(z,z')]\, dz' \,, \tag{13}$$

$$\tau_{\omega,\perp} = \int_0^z b_3(z')\tau(z') \cos[B(z,z')]\, dz' \,, \tag{14}$$

where

$$B(z,z') = \int_{z'}^z \frac{b_2(z'')}{\tau(z'')}\, dz'' \,. \tag{15}$$

Note that $d[|\boldsymbol{\tau}_\omega|^2]/dz = 2b_3\, \tau_{\omega,\perp}\tau$. Thus, $b_2$ does not directly contribute toward increasing the magnitude of second-order PMD; rather, it induces a rotation of $\boldsymbol{\tau}_\omega$ about $\mathbf{u}_3$, as seen from Eqs. (10)–(11). Since the maximum $\tau_{\omega,\perp}$ under such a rotation occurs when $\tau_{\omega,\parallel} = 0$, we expect the maximum second-order PMD to be generated for $b_2 = 0$, i.e., when $\mathbf{b}$ lies in the $\mathbf{u}_1$-$\mathbf{u}_3$ plane. Calculus of variations can then be used on Eq. (14) to show that the maximum growth of second-order PMD is obtained when

$$\mathbf{b}(z) = b\left[\mathbf{u}_1 \cos\beta(z) + \mathbf{u}_3 \sin\beta(z)\right] \,, \tag{16}$$

where $\beta(z) = \beta_{\max}\, z/z_{\max}$ and $\beta_{\max} = \pi/2$. With multiple biasing strengths, this generates region 3 in Fig. 6. We refer to this choice, which yields the largest values of $|\boldsymbol{\tau}_\omega|$, as optimal second-order biasing.

Of course, a more complete coverage of the $|\boldsymbol{\tau}|$-$|\boldsymbol{\tau}_\omega|$ plane is often needed. In this case, intermediate biasing choices must also be used in addition to optimal first- and second-order biasing. These intermediate choices can be obtained by using calculus of variations to maximize a linear combination of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, as obtained from Eqs. (12)–(14). The resulting form of $\mathbf{b}(z)$ is the same as above, except that the value of the final angle $\beta_{\max}$ now varies between 0 and $\pi$, the particular value depending upon the specific linear combination of first- and second-order PMD being maximized. A selection of angles, together with the resulting regions in the $|\boldsymbol{\tau}|$-$|\boldsymbol{\tau}_\omega|$ plane, is shown in Fig. 6. The advantage of using multiple biasing — as opposed to just pure first- or second-order biasing or no biasing at all — is evident. When a range of biasing strengths are used, 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. (A more careful analysis shows the samples to follow a family of parabolas.) Together, a selection of angles $\beta_{\max}$ and biasing strengths $\alpha$ covers the entire $|\boldsymbol{\tau}|$-$|\boldsymbol{\tau}_\omega|$ plane.

When birefringent waveplates are used, the contributions $\Delta\boldsymbol{\tau}_{n+1}$ are confined to the equatorial plane of the Poincaré sphere. In this case, the appropriate biasing directions are obtained by projecting the vectors $\mathbf{b}$ determined above onto the equatorial plane. For both models, once the biasing direction $\mathbf{b}$ has been selected, the biased distribution for the orientation of $\Delta\boldsymbol{\tau}_{n+1}$ is chosen as described previously. Furthermore, if desired, the frequency dependence of the PMD vector can be added as described as in Section 4.

## 6. Multiple Importance Sampling

The simultaneous use of different biasing methods is called *multiple importance sampling* [25]. When using several biasing distributions $p_j^*(\boldsymbol{\theta})$, a weight $w_j(\boldsymbol{\theta})$ is assigned to the samples from each distribution, and a multiply-importance-sampled Monte Carlo estimator for $P$ is written as

$$\hat{P} = \sum_{j=1}^{J} \hat{P}_j = \sum_{j=1}^{J} \frac{1}{M_j} \sum_{m=1}^{M_j} w_j(\boldsymbol{\theta}_{j,m}) L_j(\boldsymbol{\theta}_{j,m}) I(\boldsymbol{\theta}_{j,m}) \,, \tag{17}$$

where $M_j$ is the number of samples drawn from the $j$th distribution $p_j^*(\boldsymbol{\theta})$, and $\boldsymbol{\theta}_{j,m}$ is the $m$th such sample. Here, $J$ is the number of different biasing distributions, $L_j(\boldsymbol{\theta}) = p(\boldsymbol{\theta})/p_j^*(\boldsymbol{\theta})$ is the likelihood ratio of the $j$th distribution and the $w_j(\boldsymbol{\theta})$ are the weights used to combined the results.

Several ways exist to choose the weights $w_j(\boldsymbol{\theta})$. The quantity $\hat{P}$ will be an unbiased estimator for $P$ (i.e., the expectation value of $\hat{P}$ is $P$) for all choices such that $\sum_{j=1}^{J} w_j(\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$. Each choice of weights corresponds to a different way of partitioning of the total probability. The simplest possibility, of course, 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 region of sample space. This choice, however, is not the most advantageous; a better strategy is to favor those biasing distributions that are expected to produce the best results in each given region.

A particularly useful choice of weights is the so-called *balance heuristics* [26]. In this case, the weights $w_j(\boldsymbol{\theta})$ are assigned according to

$$w_j(\boldsymbol{\theta}) = \frac{p_j^*(\boldsymbol{\theta})}{\sum_{j'=1}^{J} p_{j'}^*(\boldsymbol{\theta})} \,. \tag{18}$$

[Obviously, Eq. (18) can also be written in terms of likelihood ratios.] Thus, the weight of a sample obtained with the $j$th distribution is given by the probability of realizing that sample with the $j$th distribution relative to the total probability of realizing the same sample with all distributions. Thus, Eq. (18) weights more heavily those samples coming from a biasing distribution that it more likely to generate samples in that region, while samples that are unlikely to have occurred [because $p_j^*(\boldsymbol{\theta})$ is small] are weighted less heavily. Other strategies for assigning the weights are possible. However, the use of balance heuristics has been shown to become asymptotically close to optimal as the number of realizations becomes large [26]. Of course, the outputs of each Monte-Carlo realization is always adjusted for the bias through the IS likelihood ratio.

## 7. Variance Estimation

The variance of a Monte Carlo estimator is obviously a critical parameter, since its value determines the number of samples that will be necessary, on average, to obtain a required level of accuracy, and because its numerical estimate provides a measure of the accuracy of the results.

From probability theory we know that the variance $\sigma^2_{\hat{P}}$ of the importance-sampled Monte Carlo estimator $\hat{P}$ in Eq. (6) is $\sigma^2_{\hat{P}} = \sigma^2_P/M$. We can then write a Monte Carlo estimate of $\sigma^2_{\hat{P}}$ as

$$\hat{\sigma}_{\hat{P}}^2 = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left( I(\boldsymbol{\theta}_m) L(\boldsymbol{\theta}_m) - \hat{P} \right)^2 . \tag{19}$$

When no biasing is used (that is, for straightforward Monte Carlo simulations), we have $L(\boldsymbol{\theta}) = 1$, which yields $\sigma_{\hat{P}}^2 = P - P^2$. If $P$ is a rare event we also have $P \ll 1$, which means that the relative accuracy of the Monte Carlo estimator is $\hat{\sigma}_{\hat{P}}/\hat{P} \sim 1/(M\,\hat{P}^{1/2})$. Thus if $\hat{P}$ is exponentially small (as in the case of the bit error rate or the outage probability), an exceedingly large number of unbiased Monte Carlo simulations are necessary to obtain a reliable estimate. Indeed, it is because importance sampling is highly effective in reducing the variance of the Monte Carlo estimator that its use allows substantial speed ups of numerical simulations.

Of course, if $\hat{P}$ is obtained through multiply-importance-sampled Monte Carlo simulations according to Eq. (17), the variance in Eq. (19) should also be modified accordingly. The proper estimator for $\hat{\sigma}_{\hat{P}}^2$ is, in this case,

$$\hat{\sigma}_{\hat{P}}^2 = \sum_{j=1}^{J} \frac{1}{M_j(M_j-1)} \sum_{m=1}^{M_j} \left( I(\boldsymbol{\theta}_{j,m}) w_j(\boldsymbol{\theta}_{j,m}) L_j(\boldsymbol{\theta}_{j,m}) - \hat{P}_j \right)^2 . \tag{20}$$

Note also that different weighting strategies result in different values of $\hat{\sigma}_{\hat{P}}^2$, and the most effective methods are the ones that yield the smallest variances [26]. It is outside the scope of this paper to discuss the various choices; we will limit ourselves to note that the balance heuristics has been shown to give excellent results in a wide variety of situations.

## 8. Experimental Implementation

Experimental PMD emulators employing a concatenation of birefringent elements have been built [20, 21], and a PMD emulator that incorporates a simplified form of importance sampling which works with a small number of sections has also been built [27]. It is possible that an emulator could be constructed to use the importance-sampled algorithm described above. 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 a possible experimental implementation of the algorithm 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; obviously, such an implementation would 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 $\varphi_{n+1}$ in the Müller matrix $\mathsf{R}_{n+1}$ must then be well characterized. If they are not known relatively precisely, then 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 earlier, each beat length of a birefringent section generates a $2\pi$ retardation. Because sections with significant DGDs will be

many beat lengths long, this means that the section lengths must 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 [28].

## 9. Examples

To demonstrate the ideas explained in the previous sections, we have considered a concatenation of 50 birefringent sections with 0.5 ps of DGD per section (corresponding to a mean DGD of 3.26 ps), employing either polarization scramblers or rotatable waveplates. The average DGD is given by $\langle\tau\rangle = \sqrt{8\langle\tau^2\rangle/(3\pi)}$, where $\langle\tau^2\rangle = \sum_{n=1}^{N} a_n^2$, the $a_n$ are the individual DGDs of each section and $N$ is the number of sections. Figure 7 shows the pdf of the DGD, obtained with optimal first-order biasing, while Fig. 8 shows the pdf of second order PMD, obtained with the optimal second-order biasing technique. In both cases, the biasing parameters $\alpha = 1, 2, \ldots, 10$ 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 bins and adjusting the contribution *of each individual sample* for the bias using its likelihood ratio, as explained earlier. The solid lines show the analytical pdfs for real fiber [29].

When polarization scramblers are present, the evolution of the PMD vector is equivalent to a three-dimensional random walk, and an exact solution is available for the pdf of the DGD [30, 31]:

$$p_N(r) = \frac{2r}{\pi} \int_0^\infty x \, \sin(rx) \prod_{n=1}^{N} \frac{\sin(a_n x)}{a_n x} \, dx \,. \tag{21}$$

This integral also has Fourier and asymptotic expansions [30] which can be compared with numerical results. For moderate values of the DGD $r$, the pdf $p_N(r)$ is well approximated by a Maxwellian distribution, which has the pdf $p_{\mathrm{M}}(r) = 3\sqrt{6/\pi}\,\left(r^2/\langle\tau^2\rangle^{3/2}\right) e^{-3r^2/2\langle\tau^2\rangle}$. The degree of agreement, of course, improves as the number of sections is increased [31].

The numerically calculated pdf for the case of scramblers agrees extremely well with the exact solution (dashed line under the squares). For both scramblers and waveplates, the accuracy of the numerical results improves as the number of simulations (and the number of bins) is increased. It should be noted, however, that very good results are obtained with the 2,000,000 Monte-Carlo simulations employed here, as evident in Fig. 7, and reasonable results can be obtained with far fewer numbers of trials. In particular, a good approximation is achieved for probabilities below $10^{-12}$; to obtain comparable accuracy at this probability with straightforward Monte-Carlo simulations, at least $10^{14}$ or $10^{15}$ trials would be required. Thus, for these cases importance sampling speeds up simulations by several orders of magnitude.

The emulator with rotatable sections yields better agreement with real fiber than the emulator with polarization scramblers. 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 in the frequency domain [20, 32]. For both scramblers and waveplates, the pdfs deviate

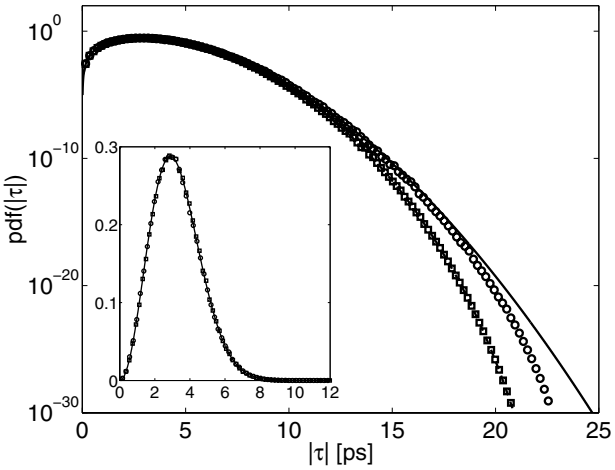


**Fig. 7.** Importance-sampled pdf of the DGD for 50 sections with 0.5 ps DGD each and with polarization scramblers (squares) or birefringent waveplates (circles). Solid curve: Maxwellian distribution with mean DGD of 3.26 ps; dashed line (coincident with squares): exact and asymptotic solutions from [30]. Inset: the pdfs on a linear scale. A total of $2 \times 10^6$ Monte-Carlo realizations were used.

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

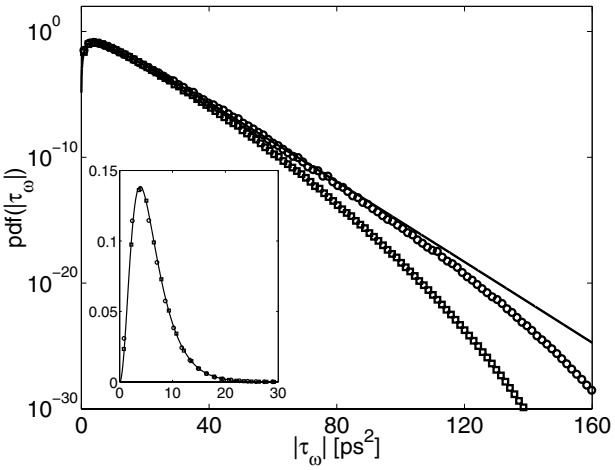


**Fig. 8.** The pdf of second-order PMD for a concatenation of 50 sections with 0.5 ps DGD each, using scramblers (squares) or waveplates (circles). The solid lines show the pdf for real fiber, from Ref. [29]. Inset: the pdf on a linear scale. As in Fig. 7, $2 \times 10^6$ Monte-Carlo realizations were used.

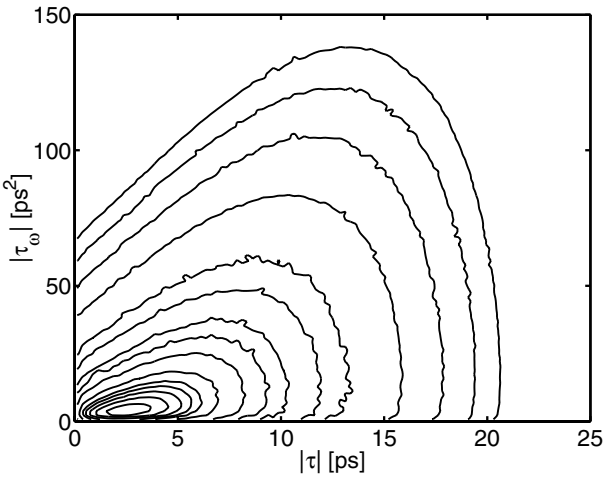


**Fig. 9.** Contour plots of the joint pdf for a concatenation of 50 sections with 0.5 ps DGD each and polarization scramblers. The contours are at $10^{-n}$ with $n = 30$, 25, 20, 15, 10, 8, 6, 5, 4, 3, 2.5, 2.25, 2, 1.75, and 1.5. A total of $10^6$ Monte-Carlo realizations were used.

has finite maximums for both $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$. As a result, large values may be inaccessible in the $|\boldsymbol{\tau}|$-$|\boldsymbol{\tau}_\omega|$ plane. (The emulator should be chosen so that these regions are unimportant in determining the outage probability of the system being tested.) For a device with equal length birefringent sections, an approximate value for the maximum second-order PMD is obtained from the exact solution of the continuum model, evaluated in the case of optimal second-order biasing: $|\boldsymbol{\tau}_\omega|_{\max} = N^2(\Delta\tau)^2/\pi$, where $\Delta\tau$ is the DGD of each section. Of course, $|\boldsymbol{\tau}|_{\max} = N\,\Delta\tau$. 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.

Figure 9 shows the joint pdf of the magnitudes of first- and second-order PMD (a two-dimensional reduction of the full three-dimensional joint pdf of first- and second-order PMD [1]) for an emulator with polarization scramblers, as calculated with the multiple biasing technique. The characteristic function associated with the joint pdf of first- and second-order PMD in an optical fiber (the limit of an infinite number of infinitesimal birefringent sections) was given in [1], but, to our knowledge, no exact analytical expression exists for the joint pdf. (An approximate pdf was given in [1], however, for the case where second-order PMD is a small perturbation compared to first-order.) Similarly, the joint pdf for PMD emulators with a finite number of sections is not known analytically.

The results for the joint pdf confirm why the optimal first- and second-order biasing methods are the correct ones to use if one is only interested in first- or second-order PMD statistics, respectively. As seen from Fig. 9, for any fixed value of DGD (a vertical slice in Fig. 9), the maximum of the pdf occurs in a zone that falls within region 1 in Fig. 6. Similarly, for any given second-order PMD (a horizontal slice in Fig. 9), the maximum of the pdf occurs in a zone falling within region 3 in Fig. 6. 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.

## 10. Conclusions

We have discussed the application of importance sampling to Monte-Carlo simulations of PMD generated by birefringent sections connected by either polarization scramblers or rotatable connectors. Importance sampling biases the Monte-Carlo simulations so that configurations with large values of first- and/or second-order PMD occur more frequently than they would normally. As a result, the method allows rare PMD events to be simulated much more efficiently than with standard methods. Importance-sampled Monte-Carlo techniques therefore provide a natural and effective means to assess PMD-induced impairments in optical transmission systems at realistic bit-error rates.

## Acknowledgments

We thank J. N. Damask, A. Galtarossa, I. T. Lima and C. R. Menyuk for many interesting and valuable discussions. This work was supported in part by the National Science Foundation under grant numbers DMS-0101476 and DGE-9987577.

## References

1. G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers", IEEE J. Lightwave Technol. **9**, 1439 (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. **5**, 819 (1993)
3. J. P. Gordon and H. Kogelnik, "PMD fundamentals: polarization-mode dispersion in optical fibers", Proc. Nat. Acad. Sci. **97**, 4541 (2000)
4. P. Ciprut, B. Gisin, N. Gisin, R. Passy, J. P. Von der Weid, F. Prieto, C. W. Zimmer, "Second-order polarization-mode dispersion: impact on analog and digital transmissions", IEEE J. Lightwave Technol. **16**, 757 (1998)
5. C. Francia, F. Bruyére, D. Penninckx, M. Chbat, "PMD second-order effects on pulse propagation in single-mode optical fibers", IEEE Photon. Technol. Lett. **10**, 1739 (1998)
6. W. Shieh, "On the Second-Order Approximation of PMD", IEEE Photon. Technol. Lett. **12**, 290 (2000)
7. S. L. Fogal, G. Biondini and W. L. Kath, "Importance-sampled pulse broadening before and after PMD compensation", OFC2002, paper ThA2
8. Q. Yu and A. E. Willner, "Performance limits of first-order PMD compensation using fixed and variable DGD elements", IEEE Photon. Technol. Lett. **14**, 304 (2002)
9. A. O. Lima, I. T. Lima, C. R. Menyuk, G. Biondini, B. S. Marks and W. L. Kath, "Statistical analysis of the performance of polarization-mode dispersion compensators with multiple importance sampling", IEEE Photon. Technol. Lett. **15**, 1716 (2003)

10. M. C. Jeruchim, "Techniques for estimating the bit error rate in the simulation of digital communication systems", IEEE J. Select. Areas Commun. **2**, 153 (1984)

11. 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. **11**, 289 (1993)

12. P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems", IEEE J. Select. Areas Commun. **15**, 597 (1997)

13. P. Bratley, B. L. Fox and L. E. Schrage, *A guide to simulation* (Springer-Verlag, New York, 1987)

14. G. S. Fishman, *Monte Carlo: concepts, algorithms and applications* (Springer-Verlag, New York, 1996)

15. G. Biondini, W. L. Kath and C. R. Menyuk, "Importance sampling for polarization-mode dispersion", IEEE Photon. Technol. Lett. **14**, 310 (2002).

16. S. L. Fogal, G. Biondini and W. L. Kath, "Multiple importance sampling for first- and second-order polarization mode dispersion", IEEE Photon. Technol. Lett. **14**, 1273 (2002); 1487 (2002)

17. G. Biondini and W. L. Kath, "Polarization mode dispersion emulation with Maxwellian length sections and importance sampling", OFC2003, paper ThA2; IEEE Photon. Technol. Lett. **16** #3, to appear (2004)

18. I. T. Lima, G. Biondini, B. S. Marks, W. L. Kath and C. R. Menyuk, "Analysis of polarization-mode dispersion compensators using importance sampling", Photon. Technol. Lett. **14**, 627 (2002).

19. 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. **15**, 1735 (1997)

20. 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", Phot. Technol. Lett. **13**, 127 (2001)

21. J. N. Damask, "A programmable polarization-mode dispersion emulator for systematic testing of 10 Gb/s PMD compensators", OFC2000, paper ThB3

22. A. Galtarossa, L. Palmieri, M. Schiano, and T. Tambosso, "Statistical characterization of fiber random birefringence", Opt. Lett. **25**, 1322 (2000)

23. J. S. Sadowsky and J. A. Bucklew, "On large deviations theory and asymptotically efficient Monte-Carlo estimation", IEEE Trans. Inf. Theory **36**, 579 (1990)

24. A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1965)

25. A. Owen and Y. Zhou, "Safe and effective importance sampling", J. Amer. Stat. Assoc. **95**, 135 (2000).

26. E. Veach, *Robust Monte Carlo Methods For Light Transport Simulation*, Ph.D. thesis (Stanford University, 1997).

27. 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 polarisation mode dispersion using multiple importance sampling", Elec. Lett. **39**, 974–975 (2003)

28. J. N. Damask, P. Myers and T. Boschci, "Coherent PMD and programmable PMD generation", OFC2003, paper ThA6

29. G. J. Foschini, L. E. Nelson, R. M. Jopson, H. Kogelnik, "Probability densities of second-order polarization-mode dispersion including polarization-dependent chromatic dispersion", IEEE Photon. Technol. Lett. **12**, 293 (2000).

30. B. D. Hughes, *Random Walks and Random Environments* (Clarendon Press, Oxford, 1995).

31. M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems", J. Lightwave Technol. **19**, 324 (2001)

32. I. T. Lima, R. Khoshravani, P. Ebrahimi, E. Ibragimov, A. E. Willner, and C. R. Menyuk, "Polarization mode dispersion emulator", OFC2000, paper ThB4