

Outage Statistics in a Waveplate Hinge Model of Polarization-Mode Dispersion

Jinglai Li, Gino Biondini, William L. Kath, and Herwig Kogelnik

Abstract—The properties of the waveplate hinge model of polarization-mode dispersion (PMD) are studied in detail, and its statistics are compared to those of the traditional hinge model based on the assumption of an isotropic output after each hinge. In particular, the probability density function of the differential group delay for each individual frequency band is computed using a combination of importance sampling and the cross-entropy method. The outage probability is then obtained combining these results with the outage map method, allowing the fraction of bands with unacceptable outage probabilities to be quantified by the noncompliant capacity ratio (NCR). The results show that the traditional hinge model significantly overestimates the NCR compared to the waveplate hinge model.

Index Terms—Cross-entropy (CE) method, hinge model, importance sampling (IS), noncompliant capacity ratio (NCR), optical fiber communications, outage probability, polarization-mode dispersion (PMD).

I. INTRODUCTION

POLARIZATION-MODE DISPERSION (PMD) is one of the obstacles facing the next generation of optical fiber communication systems [1]. Since the birefringence of installed optical fiber varies randomly with respect to both time and wavelength, the effect of PMD on a system is stochastic in nature. This problem is dealt with by allotting a power penalty to PMD. The outage probability (i.e., the probability of the PMD-induced penalty exceeding this allowed value), then becomes a key measure of the reliability of a system.

PMD is quantified by the PMD vector [2], which is frequency- and distance-dependent. The magnitude of the PMD vector, called the differential group delay (DGD), is the difference in group delays between the two principal states of polarization. It has traditionally been assumed that: 1) each individual fiber section behaves identically and independently of one another, so that the probability density function (PDF) of the total DGD follows a Maxwellian distribution [3], and that 2) the process is ergodic, in the sense that time averages

coincide with averages with respect to frequency. Recent measurements of installed fiber links, however, have reported non-Maxwellian variations in the temporal statistics of the DGD of different frequency channels in wavelength-division-multiplexed (WDM) systems [4], [5]. This behavior has been shown to be consistent with the assumption that fiber links are composed of a concatenation of a small number of long, stable sections joined by short, unprotected sections, or “hinges,” which are subject to environmental effects [5]. The hinges themselves contribute little to the total system DGD, but their variations are responsible for the temporal dynamics of PMD within each channel, whereas the longer sections act as if they are frozen in time. As a result, the statistics of the polarization-induced transmission outages of these systems differ from those predicted by traditional models; in particular, the measurements contradict the traditional assumption that all channels have identical outage probabilities.

In the “hinge model” of PMD [6]–[9], the DGD of each stable long section is assumed to follow a Maxwellian distribution as wavelength varies but is frozen in time. It is also assumed that the time-dependent hinges act as polarization controllers that scatter the PMD vector uniformly over the Poincaré sphere. The system outages under these assumptions have been well studied, and analytical expressions for the outage probability have been reported [10], [11]. In the following, we refer to this as the *isotropic hinge model*, because of the assumption of uniform output of the hinges. Recent experimental studies [12] suggest, however, that this model should be modified to encompass more realistic hinge behavior. Based on these experiments and on previous theory [13], we [14] proposed a modified, *anisotropic hinge model*, which assumes that the long sections are joined by “waveplate-type” hinges, which produce a random rotation about a static axis. We refer to this model as the *waveplate hinge model*. The goal of this work is to perform a detailed comparison between the two models. Our results show that the effects of the two types of hinges are quite different, and, as a consequence, the outage statistics predicted by them differ significantly from one another.

The most demanding part of comparing the two hinge models is to compute the outage probability associated with any given wavelength band. Outage probabilities are required to be very small—typically a minute per year, that is, 10^{-6} or less. Because of this constraint, it is difficult to use standard Monte Carlo (MC) simulations or laboratory measurements to assess system outage probabilities, due to the extremely large number of PMD configurations that must be visited to obtain reliable estimates. Importance sampling (IS) has been used to address such problems and allow efficient computation of PMD-induced transmission penalties [15]–[19]. The idea is to do MC simulations with

Manuscript received December 09, 2009; revised March 18, 2010; accepted April 12, 2010. Date of publication April 22, 2010; date of current version June 28, 2010. This work was supported in part by the National Science Foundation under Grant DMS-0709070 and Grant DMS-0908399.

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

G. Biondini is with the Department of Mathematics, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: biondini@buffalo.edu).

H. Kogelnik is with the Bell Labs, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: herwig@alcatel-lucent.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2010.2048890

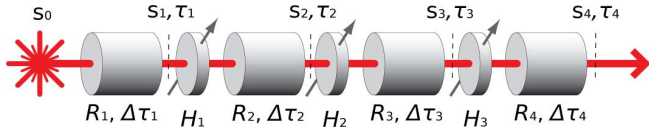


Fig. 1. Schematic representation of the waveplate hinge model: long sections (characterized by their Müller matrices R_n and sectional PMD vectors $\Delta\vec{\tau}_n$) joined by waveplate-like hinges (characterized by their Müller matrices H_n). Depending on the system, an additional hinge may be present at the input (before the first long section).

a biased distribution so that the events of interest occur more frequently than they would normally, and correct for the bias using the likelihood ratios. Generally speaking, choosing an appropriate biasing distribution is the most difficult aspect of IS. In contrast to previously studied models of PMD generation, to the best of our knowledge it is not possible to identify analytically an effective biasing distribution for the waveplate hinge model, since the equations that govern the growth of PMD are substantially more complicated. Therefore, in this work we employ the cross-entropy (CE) method, which finds a good biasing distribution iteratively. We then use the resulting distributions in importance-sampled MC simulations to compute the statistics of the PMD-induced outages in the waveplate hinge model.

The layout of this paper is as follows. Section II describes the physical model, and Section III discusses outage probabilities and the noncompliant capacity ratio (NCR) in the waveplate hinge model. Section IV introduces the CE method and applies it to reconstructing the DGD distributions. Section V discusses the results of extensive numerical simulations, and Section VI provides some additional remarks.

II. WAVEPLATE HINGE MODEL OF PMD

As shown in Fig. 1, the hinge model represents a transmission link as a finite sequence of fiber sections with fixed properties, connected by variable “hinges,” each with negligible DGD. Consider a system with N birefringent sections and $N - 1$ hinges, with one hinge between any two sections. At each frequency, the total PMD for a finite number of sections is determined by the PMD concatenation equations. The PMD vector $\vec{\tau}_{n+1}$ after the $(n + 1)$ st section is [2]

$$\vec{\tau}_{n+1} = R_{n+1}H_n\vec{\tau}_n + \Delta\vec{\tau}_{n+1} \quad (1)$$

and the total PMD vector is $\vec{\tau}_N$. Here, $\Delta\vec{\tau}_n$ is the individual PMD vector of the n th section, R_{n+1} is the rotation matrix of the $(n+1)$ st section and H_n is the rotation matrix of the n th hinge. In particular [2]

$$R_n = \cos\theta_n \mathbf{I} + (1 - \cos\theta_n)\hat{r}_n\hat{r}_n^T + \sin\theta_n \hat{r}_n \times \quad (2)$$

where \mathbf{I} is the 3×3 identity matrix, $\hat{r}_n = (r_1^{(n)}, r_2^{(n)}, r_3^{(n)})^T$, the superscript T denotes matrix transpose, and

$$\hat{r}_n \times = \begin{pmatrix} 0 & -r_3^{(n)} & r_2^{(n)} \\ r_3^{(n)} & 0 & -r_1^{(n)} \\ -r_2^{(n)} & r_1^{(n)} & 0 \end{pmatrix}.$$

For birefringent sections whose frequency dependence is linear, the rotation axis \hat{r}_n points in the direction of $\Delta\vec{\tau}_n$ [2], but this

is not true for more general polarization elements. Each of the sectional PMD vectors $\Delta\vec{\tau}_n$ is assumed to be uniformly oriented on the Poincaré sphere but is frozen in time, at least over the time scales of interest. In addition, the DGD $\Delta\tau_n = |\Delta\vec{\tau}_n|$ of each section is assumed to obey a Maxwellian distribution with respect to frequency.

The only temporal variation in (1) arises from the hinge rotation matrix H_n . Previous studies of the statistical properties of the isotropic hinge model have assumed that each hinge acts as a polarization controller that scatters the previous PMD vector evenly across the Poincaré sphere. Experimental studies do not seem to support this assumption of the isotropic hinge model, however, in that the hinge output appears to be far from uniform [12]. In fact, in Appendix I, we show that, even if each hinge were a rotator whose rotation axis varied uniformly across the Poincaré sphere and whose rotation angle was uniform, the output would be far from uniformly distributed. Thus, it appears difficult to see how a perfectly uniform scatterer could arise physically.

It is known that any polarization element acts as a waveplate with a fixed axis and variable rotation for small changes of the parameters that control its output, such as those induced by temperature [13]. Indeed, the experimental data in [12] seems to be consistent this prediction. Motivated by these considerations, in what follows we will assume that each hinge is characterized by a variable rotation about a fixed rotation axis. The rotation matrix H_n then takes the same form as R_n , namely,

$$H_n = \cos\phi_n(t)\mathbf{I} + (1 - \cos\phi_n(t))\hat{h}_n\hat{h}_n^T + \sin\phi_n(t)\hat{h}_n \times. \quad (3)$$

We will also assume that the orientation of each hinge axis is arbitrarily distributed on the Poincaré sphere, that hinge axes in different wavelengths bands are statistically independent of one another, and that the rotation angles about each hinge axis are uniform $[0, 2\pi]$. By analogy with the behavior of a wave plate, in [14] we referred to this model as the *waveplate hinge model*. Note, however, that, unlike ordinary waveplates, the rotation axes \hat{h}_n are not confined to the equator of the Poincaré sphere. Thus, one should think of the hinges as *generalized wave plates*.

PMD-induced power penalties depend not only on the total DGD τ , but also on the power splitting ratio γ , a quantity in $[0, 1]$ defined as the fraction of the pulse energy aligned with the fast principal state of polarization (SOP), namely [2],

$$\gamma = \frac{1 - \hat{s} \cdot \hat{\tau}}{2}. \quad (4)$$

Here, $\hat{\tau} = \vec{\tau}/\tau$ is the unit-length vector in the direction of the PMD vector, and \hat{s} is the unit-length Stokes vector identifying the SOP of the optical signal. The vectors \hat{s} and $\hat{\tau}$ can be evaluated either at the input of the system or at the output. In many cases, the SOP of the signal launched at the input is fixed, and we assume that this is the case throughout this paper. Since the result of (1) gives the PMD vector at the output, however, the Stokes vector at the output is also needed in order to evaluate (4). Fortunately, it is easy to show that \hat{s} satisfies its own concatenation equation

$$\hat{s}_{n+1} = R_n H_n \hat{s}_n \quad (5)$$

where \hat{s}_n indicates the SOP of the signal at the input of the n th section.

III. OUTAGES IN THE WAVEPLATE HINGE MODEL

We compute outage probabilities for both the isotropic and the anisotropic hinge model with the outage map method [20]. In this method, for each particular frequency band, the first-order PMD-induced outage probability is taken to be

$$P_{\text{out}}(\mathbf{z}) = \int \int_{(\tau, \gamma) \in B} p(\tau, \gamma; \mathbf{z}) d\tau d\gamma \quad (6)$$

where $p(\tau, \gamma; \mathbf{z})$ is the joint PDF of the total DGD τ and the power splitting ratio γ , and B is the region of the $\tau\gamma$ -plane in which an outage occurs. The regions B , which depend on the outage specifications, transmission format and receiver design, have been determined from extensive numerical simulations [21]. Here and below, \mathbf{z} is a multi-component random variable that collects the ‘‘frozen’’ degrees of freedom of the transmission link, i.e., all sectional PMD vectors, the rotation axes and angles of all sectional Müller matrices, and all the hinge rotation axes.

In previous calculations of the outage probability [11], [20], the power splitting ratio γ and the DGD τ were assumed to be statistically independent, and γ was taken to be uniformly distributed. In such a case, one of the integrals becomes trivial, and the outage probability is

$$P_{\text{out}}(\mathbf{z}) = \int p_{\text{DGD}}(\tau; \mathbf{z}) \Delta\gamma(\tau) d\tau \quad (7)$$

where $\Delta\gamma(\tau)$ is the width of the outage region B at a fixed value of DGD τ (i.e., the range of values of γ for which the PMD-induced penalty exceeds a prescribed value ϵ when the DGD has the value τ). A good approximation for $\Delta\gamma(\tau)$ can be written as [20]

$$\Delta\gamma(\tau) = \sqrt{1 + 4b\epsilon/A} \sqrt{1 - (\tau_o/\tau)^2} \quad (8)$$

where $\tau_o = 2T/\sqrt{A/\epsilon + 4b}$, while T is the bit interval, and where the values of the parameters A and b were determined from extensive numerical simulations.

The use of (7) in previous works had been justified by invoking the presence of a polarization scatterer with characteristics similar to those of an isotropic hinge at the input of the system. The validity of this approximation may be questioned, however. First of all, some systems may not have a module at the input that acts as a polarization scatterer. Moreover, any polarization-dependent module at the input is more likely to act like a waveplate hinge, rather than a scatterer (again, see Appendix I). In general, then, the probability distribution of γ cannot be expected to be uniform, especially for the waveplate hinge model. We therefore calculate outage probabilities for the waveplate hinge model using the more correct (6). The joint PDF is evaluated using ISMC methods, as described in Sections IV–VI. For the isotropic hinge model, the PDF of the DGD is obtained, as described in [11].

Recall that both isotropic and anisotropic hinge models invalidate the assumption that all frequency channels have identical outage probabilities, i.e., in a WDM system described by a

hinge model, each band will have its own value of outage probability. In other words, P_{out} becomes itself a random quantity, dependent on the particular values of the sectional PMD vectors and hinge rotation axes of that wavelength band. (Note that here, as in [20], different wavelength bands are defined to be sufficiently separated in frequency that their PMD is statistically uncorrelated. Hence, a wavelength band may contain one or more channels depending on the channel separation.) To characterize this effect, an additional measure, termed the NCR, was introduced in [20]. The NCR is defined as the probability that a given wavelength band is not compliant with a given outage specification:

$$\text{NCR} = \mathbb{P}[P_{\text{out}}(\mathbf{z}) > P_{\text{spec}}] = \mathbb{E}_{\mathbf{z}}[I_{\text{NCR}}(P_{\text{out}}(\mathbf{z}))] \quad (9)$$

where the indicator function $I_{\text{NCR}}(P_{\text{out}})$ is 1 if $P_{\text{out}} > P_{\text{spec}}$, and 0 otherwise, and the expectation value is now taken over all possible realizations of the sectional PMD vectors, Müller matrices, and hinge rotation axes, each weighted with the appropriate distribution. In the traditional model of PMD, all channels are statistically identical, and therefore, all are either compliant or noncompliant, and the NCR is either 0 or 1. This is not the case, however, if PMD is described by a hinge model [20]. In Section V, we will show that the NCR of the isotropic hinge model and that of the waveplate hinge model differ significantly.

IV. IS-CE FOR THE DGD

To efficiently calculate the PDF of the DGD and the corresponding outage probabilities in the waveplate hinge model, we use a combination of IS and the CE method. General references on the CE method are [22]–[24], while the use of IS for studying PMD is described in [15]–[19]. In summary, our method comprises a two-step approach: first, one searches for good biasing distributions $p_*(\mathbf{x})$ by iteratively minimizing their CE distance from the optimal biasing distribution; then one implements IS using the distributions obtained. Appendix VI gives a concise introduction to IS and CE, as adapted and tailored for the problem at hand. Shortly, we discuss the detailed implementation of the method for the anisotropic hinge model.

Recall that the total DGD τ depends on the hinge rotation angles $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{N-1})$ via the matrices H_n . Recall also that, the hinge rotation angles are the only variables that change with time over the time scales of interest. One way to implement IS to compute the PMD-induced outage probability for a given wavelength band (i.e., for a given set of ‘‘frozen’’ parameters \mathbf{z}) is then to bias the set of hinge rotation angles toward the choice $\boldsymbol{\Phi} = (\Phi_1, \Phi_2, \dots, \Phi_{N-1})$ that maximizes the total DGD [15]. Thus, the problem is to find these biasing angles $\boldsymbol{\Phi}$. Once these angles are known, one can use some biasing distribution to generate angles preferentially biased toward them. Here, we will use

$$p_{\alpha, \boldsymbol{\Phi}}(\boldsymbol{\phi}) = \prod_{n=1}^{N-1} p_{\alpha, \Phi_n}(\phi_n) \quad (10a)$$

with

$$p_{\alpha, \Phi_n}(\phi_n) = (\alpha/2\pi^\alpha) |\phi_n - \Phi_n - \pi|^{\alpha-1}. \quad (10b)$$

The value $\alpha = 1$ reproduces the unbiased distribution $p(\boldsymbol{\phi}) = 1/(2\pi)^{N-1}$, whereas values $\alpha > 1$ concentrate ϕ_n around Φ_n . (Other choices of the biasing distributions are also possible, as the effectiveness of IS does not seem to be very sensitive to the particular choice [15].) We refer to α as the *biasing strength* parameter.

In previous studies of PMD [15], we were able to identify the optimal biasing directions analytically, but this does not seem possible for the waveplate hinge model. As a consequence, here we find Φ by solving an optimization problem

$$\Phi = \arg \max_{\boldsymbol{\phi} \in [0, 2\pi]^{N-1}} \tau(\boldsymbol{\phi}). \quad (11)$$

To solve (11), we conduct an iterative stochastic optimization, namely, we progressively narrow the MC samples around the choice of angles that yield the largest DGD. The algorithm proceeds as follows:

- 1) Set $j = 0$, $\Phi_0 = (0, \dots, 0)$, and $T_0 = 0$.
- 2) Generate M samples $\boldsymbol{\phi}^{(1)}, \dots, \boldsymbol{\phi}^{(M)}$ distributed according to $p_{\alpha_j, \Phi_j}(\boldsymbol{\phi})$, with $\alpha_j = j + 1$. Let $\tau^{(m)} = \tau(\boldsymbol{\phi}^{(m)})$ and

$$T_{j+1} = \max_{m=1, \dots, M} \tau^{(m)}. \quad (12)$$

- 3) If $T_{j+1} < T_j$ (i.e., if no further increase has been obtained), or if some other stopping criterion is satisfied (e.g., $j > j_{\max}$), stop. Otherwise, set $\Phi_{j+1} = \boldsymbol{\phi}^{(k)}$, where k is the index such that $\tau^{(k)} = T_{j+1}$; increase j to $j + 1$ and reiterate from step 2).

Note that the above search algorithm does not use CE since in this case one cannot easily solve (27).

In general, one is interested in the *most likely* configurations that produce the event of interest. Here, however, the angles are evenly distributed in $[0, 2\pi]$, and therefore, every configuration is equally likely. Hence, we can simply look for the collection of rotation angles that maximizes the total DGD. Note that the algorithm assumes that there is only one choice of angles that produces the largest DGD. This condition is verified in all but an extremely small fraction of cases.

As usual, to correctly account for the contribution of each sample when performing ISMC simulations one needs to use the importance function, or likelihood ratio, $L_{\alpha, \Phi}(\boldsymbol{\phi}) = p(\boldsymbol{\phi})/p_{\alpha, \Phi}(\boldsymbol{\phi})$, where as before $p(\boldsymbol{\phi})$ is the unbiased distribution. In particular, for the biasing distribution [see (10a) and (10b)], it is

$$L_{\alpha, \Phi}(\boldsymbol{\phi}) = \left[\frac{\pi^{\alpha-1}}{\alpha} \right]^{N-1} \prod_{n=1}^{N-1} |\phi_n - \Phi_n - \pi|^{1-\alpha}.$$

Another result needed in the simulations is the set of values of the biasing strength parameter α that are necessary to obtain given target values of total DGD. (In general, these values do

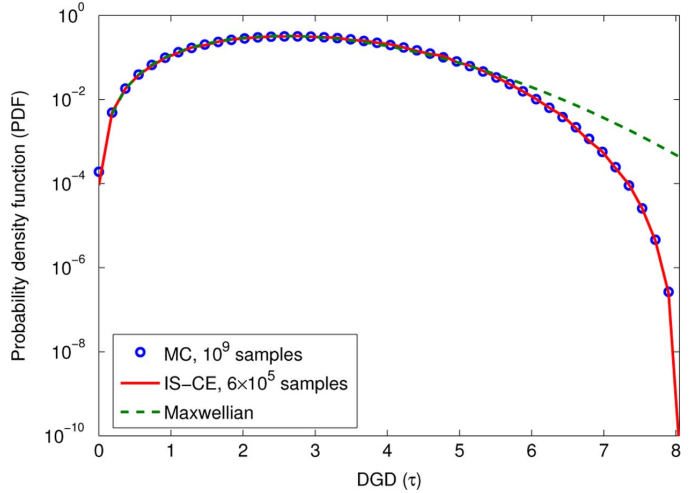


Fig. 2. PDF of the total DGD as computed by the IS-CE method and by standard MC simulations with 10^9 samples. In the IS-CE simulations, 350 000 samples were used to search for the optimal biasing angles, 250 000 samples to find the optimal biasing strengths (50 000 samples for each iteration) and another 250 000 samples in the multiple ISMC simulations to reconstruct the PDF (50 000 samples for each biasing distribution). The corresponding values of α are [1.00 1.83 2.70 4.12 6.61]. A Maxwellian distribution with the same mean DGD is also shown for comparison.

not coincide with the α_j used in the stochastic optimization algorithm described above.) To determine these values we use the CE method. Specifically, we use the distributions [see (10a) and (10b)] as the family of biasing distributions, and we search for the value of α by following the steps described in Section VI. Remarkably, in this case, (27) can be solved exactly, as we show in Appendix II, yielding (13), shown at the bottom of the page, where α_j in the right-hand side of (13) is the value used to draw the samples $\boldsymbol{\phi}^{(1)}, \dots, \boldsymbol{\phi}^{(M)}$. Note that the computation of the α 's using (13) is affected by random sampling. Moreover, since each value of α is used in the next stage of the iteration, successive values will be more affected by error. Hence, a sufficiently large number of samples should be used to ensure the accuracy of the results.

Summarizing, the specifics of the overall IS-CE algorithm for computing the PDF of the total DGD are as follows.

- 1) Find the set of rotation angles Φ that maximizes the total DGD τ using the optimization method described above.
- 2) Choose a quantile parameter q and a limiting value p_{final} of the PDF of the DGD. (For example, in the simulations described in Fig. 2 below, $q = 0.01$ and $p_{\text{final}} = 10^{-15}$ were used.) Set the iteration counter j to 0, let $\alpha_0 = 1$ (i.e., start with an unbiased distribution) and $T_0 = 0$.
- 3) Generate M samples $\boldsymbol{\phi}^{(1)}, \dots, \boldsymbol{\phi}^{(M)}$ distributed according to $p_{\alpha_j, \Phi}(\boldsymbol{\phi})$. Compute $\tau^{(m)}$ for $m = 1, \dots, M$ and let T_{j+1} be the $(1 - q)$ th sample quantile of the DGD. (That

$$\alpha_{j+1} = - \frac{(N-1) \sum_{m=1}^M I_{R_j}(\tau^{(m)}) L_{\alpha_j, \Phi}(\boldsymbol{\phi}^{(m)})}{\sum_{m=1}^M I_R(\tau^{(m)}) L_{\alpha_j, \Phi}(\boldsymbol{\phi}^{(m)}) \sum_{n=1}^{N-1} \ln \left| \frac{(\phi_n^{(m)} - \Phi_n - \pi)}{\pi} \right|} \quad (13)$$

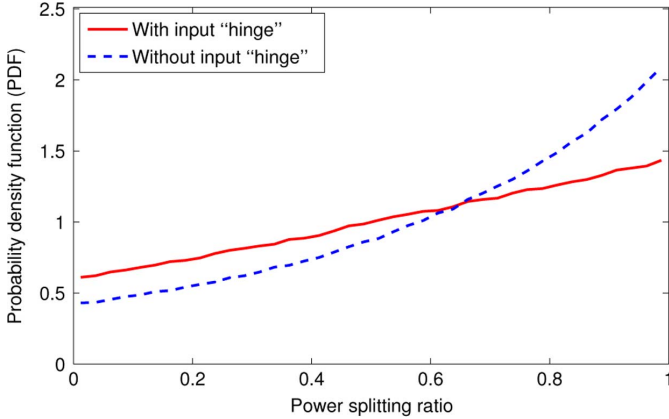


Fig. 3. PDF of the power splitting ratio for the sample transmission link that results in Fig. 2. The solid line shows the case in which a waveplate hinge is present at the input, with rotation axis $\hat{h}_0 = (0.7630, -0.1600, -0.6263)^T$, while the dashed line shows the case without a hinge at the input. The input SOP was $\hat{s}_0 = (0, 1, 0)^T$.

is, T_{j+1} is the value such that a fraction q of samples have DGD greater or equal than T_{j+1} .)

- 4) If $T_{j+1} < T_j$ or $\mathbb{P}[\tau > T_{j+1}] < p_{\text{final}}$, proceed to step 5). Otherwise, compute α_{j+1} from (13), with $I_{R_j} = 1$ for those samples in the $(1 - q)$ th quantile and $I_{R_j} = 0$, otherwise; increase j to $j+1$ and reiterate from step 3.
- 5) Perform a multiple IS simulation [15], [25] with biasing parameters $\alpha_1, \dots, \alpha_j$ to obtain the full PDF of the DGD.

To validate the method, we considered as an example a randomly selected wavelength band in a transmission link of ten sections with mean DGD of 3 ps. The corresponding results are shown in Figs. 2 and 3. The sample sectional PMD vectors and rotation axes and angles for each long section are given in Table I and the sample hinge rotation axes in Table II. Note that these values were only used to produce Figs. 2 and 3, and not for the system simulations discussed in Section V. We reconstructed the PDF of the DGD, $p_{\text{DGD}}(\tau)$, both by using the IS-CE method and by using a standard MC simulation. (Here and below, the rotation axis and angle of each frozen section were taken to be uncorrelated with the sectional PMD vector, to model the case of long fiber sections.) The results are shown in Fig. 2. As is seen from the figure, excellent agreement exists between the results of both simulations. For the same number of samples, however, use of the IS-CE method results in an efficiency improvement of several orders of magnitude over standard MC simulations. Note that the values of the biasing strengths are affected by random sampling error; a sufficient number of samples must therefore be used in the CE iteration to ensure the accuracy of the results.

As explained in Section III, in the waveplate hinge model the DGD does not completely determine the outage probability, as the power splitting ratio γ cannot be taken to be uniformly distributed in the waveplate hinge model. To illustrate this, we also reconstructed the distributions of γ for the transmission link in Fig. 2. The resulting distributions are shown in Fig. 3: one is for a system with a waveplate hinge at the input, and the other for a system with no hinge at the input. In neither case is the distribution uniform, however. As a consequence, one must use (6) rather than (7) to compute P_{out} . Fortunately,

the IS-CE algorithm above can be applied equally well to (6), as we describe below.

A possible variant is to use a different biasing strength parameter for each hinge angle. This may in principle result in a more efficient method for some specific configurations. To implement this variant, one would take the parameter vector to be $\Phi = (\Phi_1, \dots, \Phi_{N-1}, \alpha_1, \dots, \alpha_{N-1})$ and then use CE to find the optimal values of the angles and biasing strengths simultaneously. Unfortunately, doing so results in a system of nonlinear equations that is difficult to solve. Moreover, the use of this strategy did not noticeably improve the efficiency of the calculations of the NCR in the tests we performed.

V. SYSTEM SIMULATIONS AND DISCUSSION

We now apply the methods described above to numerical simulations aimed at computing outage probabilities and NCR for the waveplate hinge model, and we compare these to the corresponding results for the isotropic hinge model. We simulate a system with a transmission rate of 40 Gb/s and an allocated power margin of 1 dB. We use the outage map given in Fig. 2(a) of [21] for a nonreturn-zero (NRZ) format with a 10%–90% rise time of $0.6T$, and a target bit-error-ratio (BER) of 10^{-9} . For the isotropic hinge model, these specifications result in $b = 1.38 \times 10^{-8}$, $A = 65.4$, and $\tau_o = 6.25$ ps in (8). For the waveplate hinge model, we employ ISMC simulations to evaluate (6) using the algorithm given in Section IV, the only difference being in step 5), where one computes a multiply importance-sampled estimator for the outage probability as

$$\hat{P}_{\text{out}}(\mathbf{z}) = \frac{1}{M} \sum_{j=1}^J \sum_{m=1}^M w_j \left(\phi^{(j,m)} \right) \times I_{\text{out}} \left(\tau^{(j,m)}, \gamma^{(j,m)} \right) L_{\alpha_j, \Phi} \left(\phi^{(j,m)} \right). \quad (14)$$

Here, $\phi^{(j,m)}$ is the m th sample obtained from the j th biasing distribution, $\tau^{(j,m)}$ and $\gamma^{(j,m)}$ are the corresponding DGD and splitting ratio, the $w_j(\phi)$ are the IS weights [15], [25], and the indicator function $I_{\text{out}}(\tau, \gamma)$ is 1, if (τ, γ) lies in the outage region [21], and it is 0, otherwise. In practice, $I_{\text{out}}(\tau, \gamma)$ was taken to be 1 if $\tau/T > f(\gamma)$, and 0, otherwise, where $f(x)$ was a polynomial fit to the boundary of the outage region in Fig. 2(a) of [21]: $f(x) = 697.08x^9 - 2837.95x^8 + 4782.98x^7 - 4261.49x^6 + 2089.59x^5 - 488.66x^4 - 6.01x^3 + 32.21x^2 - 7.64x + 0.95$.

As an alternative to using (14), the methods of Section IV could be used to compute the joint PDF of τ and γ . Doing so would clearly be more computationally laborious, but would allow one to rescale the joint PDF to different values of mean DGD, in a similar way as described in [11]. Since here we compute NCR values for only a few mean DGDs, however, it was not advantageous for us to use this approach.

Since each frequency band has its own statistically independent set of “frozen” parameters (sectional PMD vectors, rotation matrices for the long sections, and rotation axes for the hinges), frequency averages are by definition ensemble averages. For both the isotropic and the anisotropic models, the NCR is then estimated via MC simulations of (9) as

$$\widehat{\text{NCR}} = \frac{1}{B} \sum_{b=1}^B I_{\text{NC}}(P_{\text{out}}(\mathbf{z}^{(b)})) \quad (15)$$

TABLE I
 SECTIONAL PMD VECTORS AND MÜLLER MATRIX PARAMETERS FOR FIGS. 2 AND 3

n	1	2	3	4	5	6	7	8	9	10
$\Delta \vec{\tau}_n$	0.0765	-0.2738	-0.7844	-0.3835	-0.5884	0.8792	-0.7681	-0.5015	-0.3249	0.3942
	0.3903	-0.1560	0.5536	0.4790	0.7964	0.6766	-0.0434	0.2959	-0.5034	-0.5078
	-0.6943	-0.7212	0.0067	0.1377	0.1721	-0.4067	-0.1965	0.8849	-0.1464	-0.7142
\hat{r}_n	0.1583	-0.2944	0.2353	0.2588	-0.5198	-0.2982	0.4653	-0.0702	0.0385	0.5821
	-0.3531	-0.9553	0.4631	-0.8885	-0.6729	-0.2006	-0.8785	0.9083	-0.2496	0.1903
	0.9221	-0.0268	0.8545	-0.3790	0.5264	-0.9332	-0.1086	-0.4124	-0.9676	-0.7905
θ_n	5.0253	0.0159	0.0397	0.1140	6.0709	6.0346	0.0133	0.1937	0.1761	0.1130

 TABLE II
 HINGE ROTATION AXES FOR FIGS. 2 AND 3

n	1	2	3	4	5	6	7	8	9
	-0.6730	-0.9383	0.4622	0.6254	-0.6891	0.0683	0.9053	0.9219	-0.5233
\hat{h}_n	0.1778	0.1125	0.6008	0.6129	-0.6819	0.6361	0.1231	0.0542	-0.7700
	0.7180	0.3272	-0.6522	0.4829	-0.2454	0.7686	-0.4065	0.3836	0.3649

where B is the number of sample frequency bands and $\mathbf{z}^{(b)}$ is the b th MC sample and P_{out} for the anisotropic model is estimated using (14). With regard to the ergodic property and the relation between time averages and frequency averages, we should remark that nothing here deviates from the usual assumptions in the hinge model, and therefore the same arguments apply regarding the temporal scales that limit the applicability of the model (e.g., see [1] and [11]).

Fig. 4 compares the value of the NCR predicted by the isotropic hinge model (dotted-dashed lines) with that predicted by the waveplate hinge model (solid lines), for links composed of six sections with a mean DGD 2.5 ps, and for links of ten sections with mean DGD 3 ps, respectively. In both cases, $B = 10000$ sample bands were used in the calculation of the NCR, with $M = 50000$ MC samples for each iteration step and for each biasing distribution, and with $q = 0.01$ and $p_{\text{final}} = 10^{-10}$. Note how the NCR for the waveplate hinge model is almost always significantly smaller than the NCR for the isotropic hinge model. Note also that assuming that the power splitting ratio is uniformly distributed yields incorrect results (dashed lines), at least for the numbers of sections considered. For comparison purposes, we also show (dotted lines) the upper bound that was used in [14], obtained by letting $\Delta\gamma(\tau) = 1$ for all $\tau > \tau_0$. Here and in all numerical simulations, the input SOP was taken to be $\hat{s}_0 = (0, 1, 0)^T$. Since the distributions of sectional PMD vectors and hinge axes are uniform over the Poincaré sphere, however, the choice of input SOP has no effect on the NCR.

Fig. 5 shows the value of the NCR obtained when a waveplate hinge is added before the input of the line. Comparison of Figs. 4 and 5 seems to indicate that the results are not very sensitive to the presence of this element. We also note that the NCR values for the isotropic and anisotropic hinge model often cross each other in the bottom left part of Figs. 4 and 5. These portions of the curves are not very interesting, however, because they have such large values of P_{spec} as to be impractical. (For example,

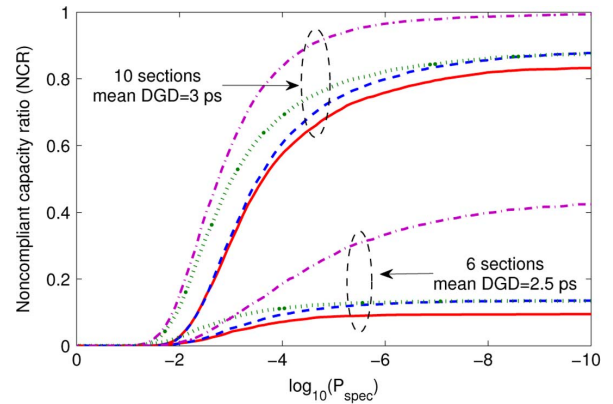


Fig. 4. NCR as a function of $\log_{10}(P_{\text{spec}})$, for fiber links of six sections with mean DGD of 2.5 ps and for those of ten sections with mean DGD of 3 ps, respectively. The solid lines are the NCR of the waveplate hinge model, with P_{out} computed by using (6), the dashed lines are the approximation assuming γ is uniformly distributed, and the dotted lines are the upper boundary [14]. The dotted-dashed lines are the NCR in the isotropic hinge model.

$P_{\text{spec}} = 0.1$ means that noncompliant bands suffer outages for more than a month per year.) Moreover, by definition a small NCR means that only a small fraction of wavelength bands is noncompliant. Therefore, estimating very small values of NCR incurs its own kind of rare event simulation problem.

To learn more about the difference between the outage statistics predicted by the isotropic and the waveplate hinge models, we also look at the ensemble average of the PDF of the DGD, $p_{\text{ens}}(\tau) = \mathbb{E}_{\mathbf{z}}[p_{\text{DGD}}(\tau; \mathbf{z})]$, which is the DGD distribution averaged over all frequency bands. As usual, we estimate this quantity using the sample mean:

$$\hat{p}_{\text{ens}}(\tau) = \frac{1}{B} \sum_{b=1}^B p_{\text{DGD}}(\tau; \mathbf{z}^{(b)}) \quad (16)$$

where B is the number of frequency bands, and for each band $p_{\text{DGD}}(\tau; \mathbf{z}^{(b)})$ is computed using the methods described in

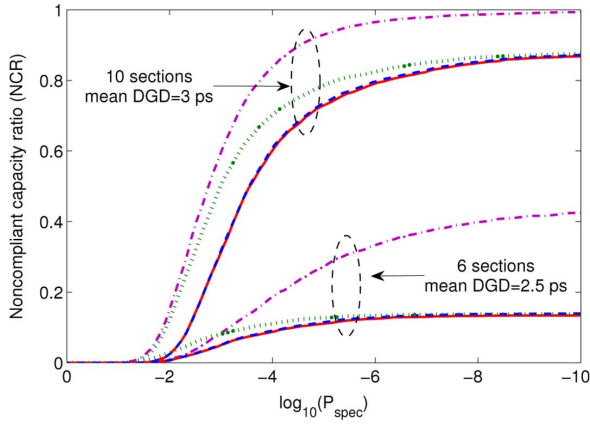


Fig. 5. Same as Fig. 4, except for the additional presence of a waveplate “hinge” at the system input.

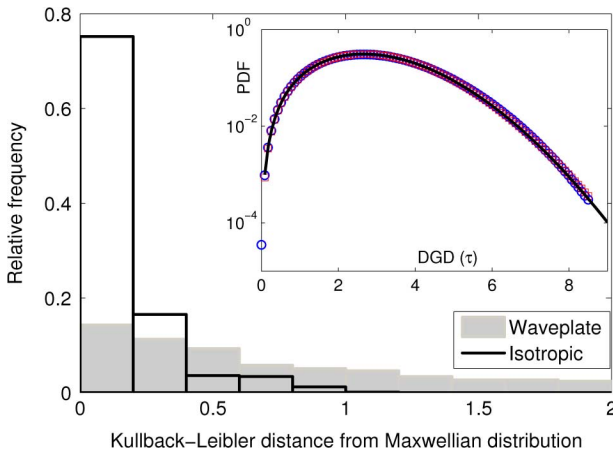


Fig. 6. (Inset) Ensemble average of the PDF of the DGD in the isotropic model (squares) and in the waveplate model (circles), as well as the Maxwellian distribution (solid line). (Main plot) Histogram showing the distribution of the KL distance between the Maxwellian PDF and the PDF of the DGD in both hinge models.

Section IV. The simulation results (see the inset of Fig. 6) show that the $p_{\text{ens}}(\tau)$ predicted by both the isotropic and the anisotropic hinge models agree very well with a Maxwellian distribution having the same mean. This result, which could have been expected, might appear at first to be at variance with the result that the NCR for the anisotropic hinge model is so much smaller than that of the isotropic hinge model.

To reconcile these two results one must look at the deviations from the ensemble average. Recall that a measure of how close two distributions are to one another is given by the Kullback-Leibler (KL) distance. We therefore compared the KL distance from the Maxwellian distribution to the DGD distributions associated with each hinge model. [Specifically, we compute (23), where f and g are, respectively, the PDF of the DGD and the Maxwellian distribution.] The histogram of the resulting KL distances is shown in Fig. 6. It is immediately seen that there is more variability associated with the DGD distributions for the waveplate hinge model than for the isotropic hinge model, i.e., at any given frequency band, the PDF of the DGD for the isotropic hinge model is expected to be closer to the Maxwellian distribution than the corresponding PDF for the waveplate hinge model.

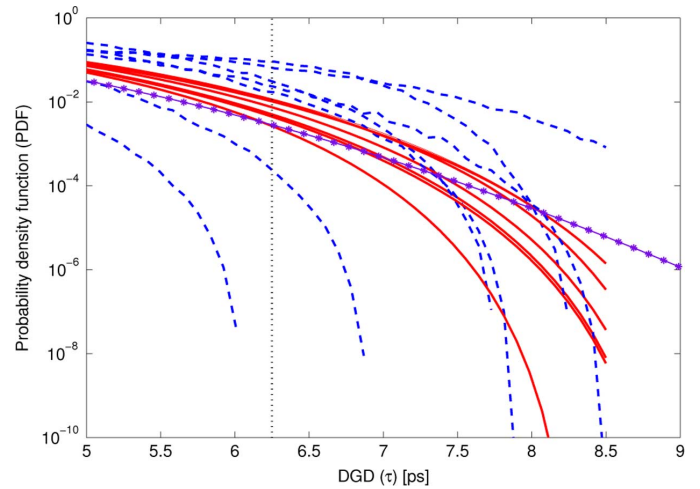


Fig. 7. PDF tails of the DGD for seven randomly selected frequency bands of the waveplate (dashed) and isotropic (solid) hinge models. The vertical line indicates the cutoff value $\tau_o = 6.25$ ps, below which no signal launch polarization can produce penalties exceeding 1 dB. Also shown (solid line with asterisks) is the average Maxwellian tail.

As a further example, Fig. 7 shows the tails of the DGD distribution for seven randomly selected bands for both the isotropic and anisotropic model.

To understand how the difference in the PDFs affects the NCR, note that, for a given outage specification P_{spec} , if the outage probability associated with the average PDF $p_{\text{ens}}(\tau)$ is not compliant, then a larger fraction of DGD distributions generated by the waveplate hinge model is likely to remain compliant than those generated by the isotropic hinge model, thus resulting in a smaller value of NCR. This is consistent with the NCR curves in Fig. 4: the waveplate hinge model produces a smaller NCR at the outage probabilities of practical interest. Conversely, when the outage probability associated with $p_{\text{ens}}(\tau)$ complies with P_{spec} , the family of distributions with larger ensemble variance then has more members *noncompliant* with it, and, therefore, results a larger NCR. This situation, however, only happens at very large values of P_{spec} , and correspondingly small values of NCR (not distinguishable in Figs. 4 and 5).

Another way to interpret this result is to observe that the more randomization is present, the closer the DGD distribution can be expected to be to Maxwellian. A comparison of different types of hinges, or rotators (see Appendix I) shows that waveplate hinges are far from random. Surprisingly, however, even a rotation about a random axis through a random angle does not produce an output distribution that is uniform on the Poincaré sphere. This shows that a perfect scatterer, which is the type of hinge assumed for the isotropic hinge model, is a highly idealized object.

VI. CONCLUSION

We have studied the outage statistics of the waveplate hinge model in detail, and reported on an IS implementation that uses the CE method to determine the biasing strategy. This method allows one to compute the outage probability without assuming a uniform distribution of the power splitting ratio. Furthermore, we have demonstrated that the outage statistics in the waveplate hinge model differ significantly from those in the isotropic hinge

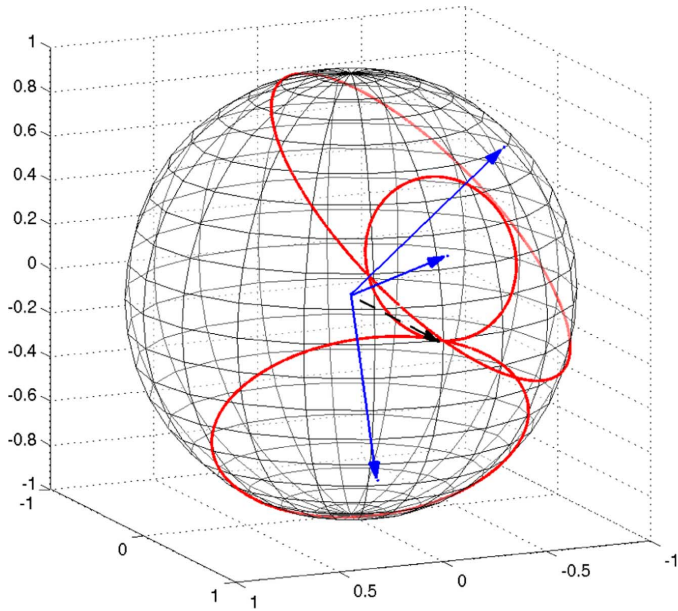


Fig. 8. Distribution of the output SOP (solid curves) from a generalized waveplate, for three different hinge axes. As expected, in each case, the output SOP is uniformly distributed on a circle. The solid arrow shows the input SOP $\hat{s} = (0, 1, 0)^T$, and the dashed arrows show the three rotation axes: $\hat{h} = (-0.0512, 0.930, 0.363)^T$, $\hat{h} = (-0.536, 0.455, 0.711)^T$, and $\hat{h} = (0.188, 0.713, -0.676)^T$.

model and, in particular, have shown that the waveplate hinge model predicts a smaller NCR when the specified outage probability is in typical ranges of interest.

To obtain these results we developed and implemented a hybrid variance reduction technique (VRT) that combines IS with CE. We should note that other VRTs have also recently been applied to simulate rare PMD events, such as the multi-canonical MC (MMC) method [26], [27] and the Markov chain MC (MCMC) method [28]. The implementation of these other methods may be somewhat simpler. Nevertheless, the use of IS has certain advantages. For instance, when a pure IS can be implemented, it usually needs much less tailoring than iterative methods, and it is usually more computationally efficient [18], [27]. Moreover, IS lends itself naturally to the computation of standard deviations for its estimators [15], which is usually a more difficult task with MMC or MCMC [27].

While the waveplate hinge model was formulated in an attempt to more properly model the behavior of PMD for installed fiber links, it is of course possible that such links are not exactly described by either the isotropic or waveplate hinge model. The results we have presented here, however, indicate that the statistics of PMD-induced outages strongly depend on the specific details by which PMD is physically generated in the system, and that as a result, the issue of determining the correct model with which to describe PMD in actual installed systems deserves further study. We hope that additional experimental data will become available in the near future to allow the development of a more refined hinge model that can produce accurate predictions of the outage probabilities of installed systems.

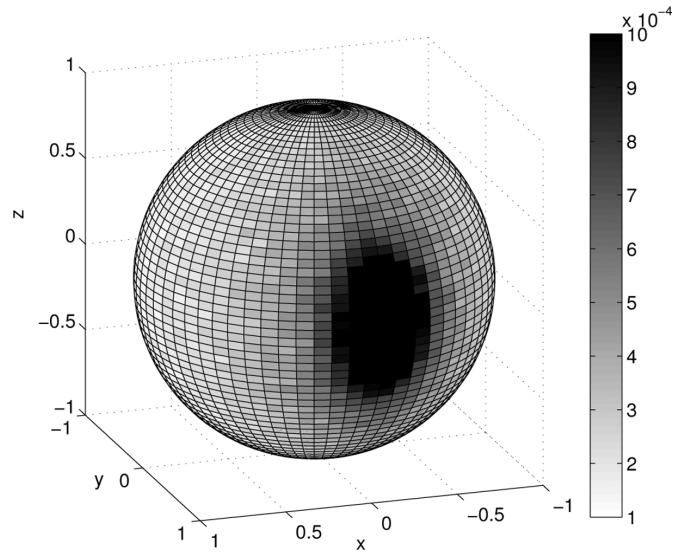


Fig. 9. Distribution of the output polarization states for a random rotator. The input vector is the same as in for Fig. 8.

APPENDIX A

OUTPUT DISTRIBUTION OF RANDOMLY ROTATED VECTORS

Here, we compare the behavior of different models of polarization “hinges.” Specifically, we consider the following three models.

- 1) A generalized waveplate rotator that rotates the input vector about a *fixed* axis through a random angle. This is the physical model of hinges considered in this work.
- 2) A *purely random* rotator that rotates the input vector about a random axis through a random angle, with the axis uniformly distributed on the sphere and the angle uniform in $[0, 2\pi]$.
- 3) A *perfect* rotator that distributes the output vector uniformly on the Poincaré sphere.

We reconstruct the distributions of the output polarization states on the Poincaré sphere using MC simulation for each type of hinge, with a fixed input state $\hat{s} = (0, 1, 0)$. A million samples were used in each simulation.

As is expected, the waveplate hinge (case 1) uniformly distributes the output vectors on a circle of the Poincaré sphere (see Fig. 8). It might be surprising, however, that, even if the rotation axis and angle are randomized (case 2), the output vectors are heavily concentrated around the direction of the input vector, and are hence far from being uniformly distributed. This phenomenon, illustrated in Fig. 9, can be understood by realizing that the rotated output will coincide with the input for any orientation of the rotation axis when the rotation angle is 0 or 2π , whereas an output in the opposite direction can only be realized when the rotation axis is orthogonal to the input and the rotation angle is exactly π .

In fact, it can be shown that, to produce a uniform output distribution (case 3), in addition to uniformly randomizing the rotation axis on the Poincaré sphere, one must draw the rotation angle ϕ for each sample according to the nonuniform distribution [29]

$$p(\phi) = (2/\pi) \sin^2(\phi/2) \quad (17)$$

where $\phi \in [0, \pi]$. These results provide additional evidence that isotropic output is difficult to achieve in practice, and that, consequently, the isotropic assumption appears to be unrealistic.

APPENDIX B IMPORTANCE SAMPLING AND CROSS-ENTROPY

Suppose \mathbf{x} is a d -dimensional random variable with PDF $p(\mathbf{x})$, and we are interested in the probability Q that a measurable function of \mathbf{x} , say $y(\mathbf{x})$, falls in some region R . In our case, \mathbf{x} collects the hinge rotation angles, and $y(\mathbf{x})$ is the DGD at the output. The quantity Q can be expressed as

$$\begin{aligned} Q &= \mathbb{P}[y(\mathbf{x}) \in R] = \int_{y \in R} p(\mathbf{x}) dV(\mathbf{x}) \\ &= \mathbb{E}[I_R(y(\mathbf{x}))] = \int I_R(y(\mathbf{x})) p(\mathbf{x}) dV(\mathbf{x}) \end{aligned} \quad (18)$$

where $dV(\mathbf{x}) = dx_1 \cdots dx_d$ is the volume element in \mathbb{R}^d and $I_R(y)$ is the indicator function, defined as $I_R(y) = 1$ for $y \in R$, and $I_R(y) = 0$, otherwise. One can estimate this probability using a standard MC estimator as

$$\hat{Q} = \frac{1}{M} \sum_{m=1}^M I_R(y(\mathbf{x}^{(m)})) \quad (19)$$

where M is the total number of samples and $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$ are independent identically distributed (i.i.d.) samples drawn according to the distribution $p(\mathbf{x})$. It is easy to see that $\mathbb{E}[\hat{Q}] = Q$. If $Q \ll 1$, however, the estimate becomes impractical due to the large number of samples needed. IS can often be used to resolve this problem [15]. IS works by rewriting (19) as

$$Q = \int I_R(y(\mathbf{x})) L(\mathbf{x}) p_*(\mathbf{x}) dV(\mathbf{x}) = \mathbb{E}_* [I_R(y(\mathbf{x})) L(\mathbf{x})] \quad (20)$$

where $p_*(\mathbf{x})$ is the biasing distribution, $L(\mathbf{x}) = p(\mathbf{x})/p_*(\mathbf{x})$ is the importance function, or likelihood ratio, and $\mathbb{E}_*[\cdot]$ denotes expectation values with respect to $p_*(\mathbf{x})$. One can then estimate Q by

$$\hat{Q}^* = \frac{1}{M} \sum_{m=1}^M I_R(y(\mathbf{x}^{(m)})) L(\mathbf{x}^{(m)}) \quad (21)$$

where samples are drawn according to $p_*(\mathbf{x})$.

The main issue with IS, of course, is to design a good biasing strategy. To do so, one must find a biasing distribution $p_*(\mathbf{x})$ that causes samples to fall in the region of interest more frequently than would happen normally. An analytical formula for a good biasing distribution for the anisotropic hinge model has not been found. It is well known that an optimal biasing distribution exists in principle

$$p_{\text{opt}} = \frac{I_R(y(\mathbf{x})) p(\mathbf{x})}{Q} \quad (22)$$

This choice is not practical, however, as it requires knowledge of the sought-after probability Q . One can, however, find a good

biasing distribution by requiring it be “close” to the optimal biasing distribution, in terms of some measure of distance. An especially convenient way to do so is to minimize the KL distance [30], which for two functions $f(\mathbf{x})$ and $g(\mathbf{x})$, is defined as

$$\begin{aligned} \mathcal{D}(g, f) &= \mathbb{E}_g \left[\ln \frac{g(\mathbf{x})}{f(\mathbf{x})} \right] \\ &= \int \ln(g(\mathbf{x})) g(\mathbf{x}) dV(\mathbf{x}) \\ &\quad - \int \ln(f(\mathbf{x})) g(\mathbf{x}) dV(\mathbf{x}) \end{aligned} \quad (23)$$

which is also known as the cross-entropy between two probability distributions. (The KL distance is not a true metric; for example, it is not symmetric.) Here, we take $g(\mathbf{x}) = p_{\text{opt}}(\mathbf{x})$ and $f(\mathbf{x}) = p_*(\mathbf{x})$. Since the first integral on the right-hand side of (23) is fixed, minimizing the cross-entropy between $p_{\text{opt}}(\mathbf{x})$ and $p_*(\mathbf{x})$ is equivalent to maximizing $\int \ln(p_*(\mathbf{x})) p_{\text{opt}}(\mathbf{x}) dV(\mathbf{x})$. In turn, recalling (22), this problem is equivalent to maximizing $\mathbb{E}[I_R(y(\mathbf{x})) \ln p_*(\mathbf{x})]$.

Suppose that, as is the case in practice, the potential biasing distributions for \mathbf{x} are selected from a parameterized family $\{p_*(\mathbf{x}; \mathbf{v})\}_{\mathbf{v} \in V}$, where \mathbf{v} is a vector of parameters and V is the corresponding parameter space, and let the unbiased distribution be denoted by $p_*(\mathbf{x}; \mathbf{u}) = p(\mathbf{x})$. In this case, one must look for the member of this family that is closest to the optimal distribution. Based on the above discussion, one must then maximize the integral

$$D(\mathbf{v}) = \int I_R(y(\mathbf{x})) \ln(p_*(\mathbf{x}; \mathbf{v})) p(\mathbf{x}) dV(\mathbf{x}). \quad (24)$$

This is usually done numerically. Because the optimal biasing distribution is typically far from the unbiased distribution, however, the member of the family that is closest to optimal is also likely to be far from $p(\mathbf{x})$. Thus, determining the best choice for \mathbf{v} also becomes a rare event simulation. The problem, of course, is that the region R of interest is generally far from the region in sample space where the unbiased distribution $p(\mathbf{x})$ is large.

The solution to this problem is to avoid attempting to jump to the region R directly, but rather to use a sequence of intermediate regions R_j that reach the desired region in a series of steps. Let $D_j(\mathbf{v})$ be the integral in (24) with R replaced by R_j . Starting with a biasing parameter $\mathbf{w}_1 = \mathbf{u}$ (i.e., with the unbiased distribution), one uses MC sampling to minimize the CE distance between the parametrized distribution and the optimal distribution that reaches R_1 . This step, which is done by finding the maximum of $D_1(\mathbf{v})$ over this first set of samples, will give a biasing parameter \mathbf{w}_2 . One then uses this value to define a biasing distribution, and performs an MC simulation with this distribution to minimize the CE distance between the parametrized distribution and the optimal distribution that reaches R_2 . Of course, since a biasing distribution is being used, each step of the procedure is an ISMC simulation of a stochastic optimization program, i.e., at step j , one must compute

$$\mathbf{w}_{j+1} = \arg \max_{\mathbf{v} \in V} \hat{D}_j(\mathbf{v}) \quad (25)$$

where

$$\begin{aligned} \hat{D}_j(\mathbf{v}) &= \frac{1}{M} \sum_{m=1}^M I_{R_j} \left(y \left(\mathbf{x}^{(m)} \right) \right) L \left(\mathbf{x}^{(m)} \right) \\ &\quad \times \ln \left(p_* \left(\mathbf{x}^{(m)}; \mathbf{v} \right) \right) \end{aligned} \quad (26)$$

and where $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$ are i.i.d. samples generated according to $p_*(\mathbf{x}; \mathbf{w}_j)$. The optimal biasing distribution then can be adaptively determined by performing the following steps [22].

- 1) Set $j = 1$ and the initial parameter $\mathbf{w}_1 = \mathbf{u}$.
- 2) Generate MC samples according to $p_*(\mathbf{x}; \mathbf{w}_j)$.
- 3) Solve (25) to find \mathbf{w}_{j+1} .
- 4) If the iteration has converged (see [22] and below), stop; otherwise, increase j to $j + 1$ and reiterate from step 2).

Once the iteration has converged, one can then perform ISMC simulations using the biasing distribution $p_*(\mathbf{x}; \mathbf{w}_{\text{final}})$.

The regions R_j above are defined implicitly in terms of decreasing levels of probability. A practical alternative is to define the regions R_j in terms of *sample quantiles* of some quantity of interest $y(\mathbf{x})$ [22]. For example, if the goal is to produce large values of DGD, at each iteration, we can define the region as those parts of sample space that produce the largest fraction, e.g., 1%, of all of the randomly generated DGD samples. In this way, the iteration systematically moves to larger and larger values of DGD. The iteration is then considered to have converged when the sample quantiles cross some predefined threshold, or when the probability associated with that DGD value is sufficiently small.

One can think of CE as an adaptive IS method [31], [32] in which the KL distance is the measure used to do the optimization. A major issue associated with the algorithm is of course how to accomplish step 3). Solving (25) can be complicated in general. It may therefore seem that one has simply traded the original problem for an equally difficult one. If the function $D(\mathbf{v})$ is convex and differentiable, however, the solutions of (25) can be obtained by solving a system of equations [22]

$$\frac{1}{M} \sum_{m=1}^M I_R \left(y \left(\mathbf{x}^{(m)} \right) \right) L \left(\mathbf{x}^{(m)}; \mathbf{u}, \mathbf{v} \right) \times \nabla_{\mathbf{v}} \ln p \left(\mathbf{x}^{(m)}; \mathbf{v} \right) = 0 \quad (27)$$

which, in many important applications, can be solved analytically. In other words, a key advantage of the CE method is that, for a wide class of distributions, the set of parameters that minimizes the KL distance can be found straightforwardly—i.e., the parameter update can be done explicitly. As we show in next section, this is indeed the case when applying this method to calculate the PDF of the DGD. Therefore, in our case, the use of CE *does* result in a significant step forward toward the solution of the problem.

APPENDIX C

CROSS-ENTROPY MAXIMIZATION STEP FOR THE DGD

Here, we outline the solution of the optimization problem (27) for the waveplate hinge model, which determines the optimal value of the biasing strength in order to obtain a desired value of total DGD.

When (10a) and (10b) are substituted into (27), one obtains

$$\frac{1}{M} \sum_{m=1}^M I_{R_j} \left(\tau^{(m)} \right) L_{\alpha_j, \Phi} \left(\phi^{(m)} \right) \times \frac{\partial}{\partial \alpha} \ln \left[\left(\frac{\alpha}{2\pi} \right)^{N-1} \prod_{n=1}^{N-1} \left| \frac{\phi_n^{(m)} - \Phi_n - \pi}{\pi} \right|^{\alpha-1} \right] = 0 \quad (28)$$

where as before, $L_{\alpha, \Phi}(\phi)$ is the likelihood ratio, and α_j is the actual value of α used to draw the samples. Simple manipulation of (28) yields

$$\sum_{m=1}^M I_{R_j} \left(\tau^{(m)} \right) L_{\alpha_j, \Phi} \left(\phi^{(m)} \right) \times \left(\frac{N-1}{\alpha} + \sum_{n=1}^{N-1} \left| \frac{\phi_n^{(m)} - \Phi_n - \pi}{\pi} \right| \right) = 0 \quad (29)$$

which can be solved immediately for α , yielding (13).

ACKNOWLEDGMENT

The authors would like to thank P. J. Winzer for many insightful discussions.

REFERENCES

- [1] H. Kogelnik, L. E. Nelson, and R. M. Jopson, "Polarization mode dispersion," in *Optical Fiber Telecommunications IVB*, I. P. Kaminow and T. Li, Eds. San Diego, CA: Academic, 2002, pp. 725–861.
- [2] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization-mode dispersion in optical fibers," *Proc. Nat. Acad. Sci. USA*, vol. 97, pp. 4541–4550, 2000.
- [3] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1439–1456, Nov. 1991.
- [4] M. Karlsson, J. Brentel, and P. A. Andrekson, "Long-term measurement of PMD and polarization drift in installed fibers," *J. Lightw. Technol.*, vol. 18, no. 7, pp. 941–951, Jul. 2000.
- [5] M. Brodsky, M. Boroditsky, P. Magill, N. J. Frigo, and M. Tur, "Persistence of spectral variations in DGD statistics," *Opt. Exp.*, vol. 13, pp. 4090–4095, 2005.
- [6] Boroditsky, P. Magill, N. J. Frigo, and M. Tur, "Channel-to-channel variation of non-Maxwellian statistics of DGD in a field installed system," in *Proc. ECOC 2004*, vol. 3, pp. 306–309, Paper We1.4.1..
- [7] M. Boroditsky, M. Brodsky, P. Magill, N. J. Frigo, C. Antonelli, and A. Mecozzi, "Outage probabilities for fiber routes with finite number of degrees of freedom," *IEEE Photon. Technol. Lett.*, vol. 17, no. 2, pp. 345–347, Feb. 2005.
- [8] C. Antonelli and A. Mecozzi, "Theoretical characterization and system impact of the hinge model of PMD," *J. Lightw. Technol.*, vol. 24, no. 11, pp. 4064–4074, Nov. 2006.
- [9] M. Brodsky, N. J. Frigo, M. Boroditsky, and M. Tur, "Polarization mode dispersion of installed fibers," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4584–4599, Dec. 2006.
- [10] C. Antonelli and A. Mecozzi, "Statistics of the DGD in PMD emulators," *IEEE Photon. Technol. Lett.*, vol. 16, no. 8, pp. 1840–1842, Aug. 2004.
- [11] J. Li, G. Biondini, H. Kogelnik, and P. Winzer, "Noncompliant capacity ratio for systems with an arbitrary number of polarization hinges," *J. Lightw. Technol.*, vol. 26, no. 14, pp. 2110–2117, Jul. 2008.
- [12] M. Brodsky, J. C. Martinez, N. J. Frigo, and A. Sirenko, "Dispersion compensation module as a polarization hinge," in *Proc. ECOC 2005*, vol. 3, pp. 335–336.
- [13] W. Shieh and H. Kogelnik, "Dynamic eigenstates of polarization," *IEEE Photon. Technol. Lett.*, vol. 13, no. 1, pp. 40–42, Jan. 2001.
- [14] J. Li, G. Biondini, W. L. Kath, and H. Kogelnik, "Anisotropic hinge model for polarization-mode dispersion in installed fibers," *Opt. Lett.*, vol. 33, pp. 1924–1926, 2008.
- [15] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization mode dispersion: Techniques and applications," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1201–1215, Apr. 2004, vol. 24, pp. 1065–1066, 2006.
- [16] I. T. Lima, Jr., A. O. Lima, G. Biondini, C. R. Menyuk, and W. L. Kath, "A comparative study of single-section polarization-mode dispersion compensators," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1023–1032, Apr. 2004.
- [17] G. Biondini and W. L. Kath, "PMD emulation with Maxwellian length sections and importance sampling," *IEEE Photon. Technol. Lett.*, vol. 16, no. 3, pp. 789–791, Mar. 2004.
- [18] G. Biondini and W. L. Kath, "Polarization-dependent chromatic dispersion and its impact on return-to-zero transmission formats," *IEEE Photon. Technol. Lett.*, vol. 17, no. 9, pp. 1866–1868, Sep. 2005.

- [19] B. R. Stone, G. Biondini, and W. L. Kath, "Statistics of polarization-mode dispersion emulators with unequal sections," *SIAM J. Appl. Math.*, vol. 69, pp. 552–564, 2008.
- [20] H. Kogelnik, P. J. Winzer, L. E. Nelson, R. M. Jopson, M. Boroditsky, and M. Brodsky, "First-order PMD outage for the hinge model," *IEEE Photon. Technol. Lett.*, vol. 17, no. 6, pp. 1208–1210, Jun. 2005.
- [21] P. J. Winzer, H. Kogelnik, and K. Ramanan, "Precise outage specifications for first-order PMD," *IEEE Photon. Technol. Lett.*, vol. 16, no. 6, pp. 449–451, Mar. 2004.
- [22] P. T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, pp. 19–67, 2005.
- [23] T. Homem-de-Mello, "A study on the cross-entropy method for rare event probability estimation," *INFORMS J. Comput.*, vol. 19, pp. 381–394, 2007.
- [24] R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning*. New York: Springer-Verlag, 2004.
- [25] E. Veach, "Robust Monte Carlo Methods For Light Transport Simulation," Ph.D. dissertation, Stanford Univ., Stanford, CA, , 1997.
- [26] D. Yevick, "Multicanonical communication system modeling—Application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, no. 11, pp. 1512–1514, Nov. 2002.
- [27] A. O. Lima, I. T. Lima, Jr., and C. R. Menyuk, "Error estimation in multicanonical Monte Carlo simulations with applications to polarization-mode-dispersion emulators," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3781–3789, Nov. 2005.
- [28] M. Secondini and E. Forestieri, "All-order PMD outage probability evaluation by Markov chain Monte Carlo simulations," *IEEE Photon. Technol. Lett.*, vol. 17, no. 7, pp. 1417–1419, Jul. 2005.
- [29] R. E. Miles, "On random rotations in R^3 ," *Biometrika*, vol. 52, pp. 636–639, 1965.
- [30] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, pp. 79–86, 1951.
- [31] J. S. Stadler and S. Roy, "Adaptive importance sampling," *J. Sel. Areas Commun.*, vol. 11, pp. 309–316, 1993.
- [32] D. Remondo, R. Srinivasan, V. F. Nicola, W. C. van Etten, and H. E. P. Tattje, "Adaptive importance sampling for performance evaluation and parameter optimization of communication systems," *IEEE Trans. Commun.*, vol. 48, pp. 557–565, 2000.

Jinglai Li, photograph and biography not available at the time of publication.

Gino Biondini, photograph and biography not available at the time of publication.

William L. Kath, photograph and biography not available at the time of publication.

Herwig Kogelnik, photograph and biography not available at the time of publication.