

MODELING COMPENSATION FOR OPTICAL FIBER COMMUNICATION SYSTEMS

JOHN ZWECK AND SUSAN E. MINKOFF*

Abstract. Today the vast majority of telecommunication and Internet messages are sent along fiber optic cables buried underground. Binary data (encoded as a sequence of pulses of light) may travel thousands of kilometers to reach its final destination. The fibers that are used for this data transfer necessarily contain manufacturing impurities that lead to fast and slow polarization states for the propagating signal. This imperfection in the fiber results in a random distortion effect known as polarization mode dispersion (PMD). As binary data travels along these fibers, the pulses spread causing the ones to decrease in value and the zeros to increase. Thus, the received message may contain errors. To decrease the likelihood of errors in the received signal, a device known as a compensator can be placed at the receiver. Determining an optimal setting for the compensator involves rotating the fiber in the compensator to best align its slow axis with the fast axis of the transmission fiber. Such a rotation should cancel out some of the effects of PMD. Modeling this system numerically requires that one generate fiber realizations with large amounts of PMD. To measure rotation angle goodness of fit between compensation and transmission fiber requires that one choose a feedback signal for the compensator. We compare the eye opening, spectral line, and degree of polarization ellipsoid feedback signals. While the eye opening feedback mechanism is the most accurate measure, it is difficult to optimize numerically. The degree of polarization and spectral line feedback signals act as smooth surrogates for the eye.

Key words. optical fiber communication, photonics, pde-constrained optimization, polarization mode dispersion, Monte Carlo methods

AMS subject classifications. 65C05, 78A48, 78A50, 90B18, 90C90, 94A99

1. Introduction. The vast majority of long-distance telecommunications and Internet traffic is carried by optical fiber communication systems [31]. In an optical communication system binary data is encoded onto a sequence of pulses of light which are then transmitted over long distances through optical fiber. A material property of optical fiber called birefringence causes the pulses to spread and distort as they propagate. This distortion of the optical signal is called polarization-mode dispersion (PMD), and is governed by the linear-PMD equation, which is a special case of the Manakov-PMD equation. The Manakov-PMD equation models the propagation of light through dispersive, nonlinear, birefringent optical fiber and is derived from Maxwell's equations [25], [43]. In the 1980's scientists realized that polarization mode dispersion in optical fibers would have a significant impact on the performance of high-data rate systems [18], [30]. By the time the signal reaches its destination it may not be possible to correctly decode the transmitted binary message, *i.e.*, bit errors may occur. The engineering goal is to ensure that the bit error ratio, which is the probability that a bit error occurs, is as small as possible (typically 10^{-9} – 10^{-15}). A major difficulty in achieving this goal is that the birefringence of optical fiber, and hence the bit error ratio, varies randomly over time. To reduce the probability of a large bit error ratio, engineers have proposed using devices called optical PMD compensators.

In birefringent fiber, the speed of the light depends on its polarization state. To first approximation, a given optical fiber has two special principal states of polarization: fast and slow. The delay in the arrival times between light traveling in these two states is called the differential group delay (DGD). If the DGD is too large, bit errors

*Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland, 21250 (zweck@math.umbc.edu, sminkoff@math.umbc.edu).

will occur. A PMD compensator is a device that is placed after the transmission fiber just prior to the receiver. It is designed to reduce the deleterious effect that PMD has on the performance of a communication system. A *simple PMD compensator* consists of a device called a polarization controller that is used to change the polarization state of the light, followed by a piece of compensating fiber with a fixed DGD. The polarization controller rotates the fast polarization state of the transmission fiber onto the slow polarization state of the compensation fiber, thereby reducing the total DGD. In this paper, we study the problem of how best to optimize the performance of a simple PMD compensator. For each random realization of the birefringence of the transmission fiber, the goal is to find the rotation of the polarization controller that will minimize the bit-error ratio.

The physical and statistical properties of PMD, the equations that govern the propagation of light through birefringent optical fiber, and Monte Carlo-based models for the random variation in the birefringence of optical fiber have been widely studied over the last 25 years, and several excellent review articles have recently appeared [24], [31], [43].

Because the random fiber realizations that produce unacceptably large PMD are extremely rare, it is very difficult to observe them experimentally. It is also not practical to simulate them using a standard Monte Carlo search of the state space of all possible fiber realizations. However it is precisely these rare, large-PMD fiber realizations that are most important to consider when assessing the effectiveness of a PMD compensator. Recently, Biondini and Kath [2]–[4] developed a multiple importance sampling algorithm that uses biased Monte Carlo simulations to efficiently generate realizations of the fiber which have large amounts of PMD. This advance made it possible to perform simulation studies that more accurately assess the performance of PMD compensators [35], [36], [38], [39], [64].

In recent years several different approaches have been proposed for reducing bit errors due to PMD. Comprehensive reviews of these ideas can be found in [7], [26], [52]. One approach is to install newly designed low-PMD fiber [45]. However replacing fiber is prohibitively expensive for existing systems. Another approach is to design the shape of the transmitted light pulses to make the signal more resilient to PMD [52]. In addition to these passive methods, active PMD compensation methods have also been proposed. Active compensation techniques can either be applied to the optical signal just before it enters the receiver (optical compensation) or after the optical signal has been converted back into an electrical signal in the receiver (electrical compensation) [39].

We focus in this paper on optical PMD compensation only. Optical PMD compensation is complicated by the fact that the DGD and principal states depend on the frequency of the light. If the PMD in the transmission fiber were actually frequency independent then a simple optical PMD compensator like that described above could completely eliminate the effects of PMD. In the realistic case of frequency-dependent PMD it is still theoretically possible (but not experimentally viable) to completely eliminate the effects of PMD via solution of an inverse problem [52]. Compensators that account for at least some frequency-dependent PMD have a relatively large number of degrees of freedom, *i.e.*, their objective functions are defined on a high dimensional space. However, such devices are costly to build and operate. Since low cost is desirable, we chose to study the simple optical PMD compensator described above, which has only two degrees of freedom.

An obvious choice for the feedback mechanism to use for compensation is the

bit-error ratio. However, it is not actually possible to compute the bit-error ratio in a real system. Instead, for a simple compensator, the most common feedback signals are the power in a spectral line [27] and the degree of polarization (DOP) [33]. The goal is to maximize these feedback signals since small values of the DGD usually result in large values of the monitored signal. In addition, it has been shown that the performance of a compensator can be improved by scrambling the polarization state of the input signal [47]. Therefore, the monitor signals we chose to study for this paper are the spectral line and polarization-scrambled DOP. We compare these two feedback mechanisms to a third — the eye opening [8] — which is highly correlated to the bit-error ratio. Unfortunately, the eye opening monitor is not very practical since it is difficult to build and operate (requiring complex fast electronic circuitry).

Several simulation studies have compared the performance of different PMD compensators. Sunnerud *et al.* [52], [53] and Buchali and Bülow [7] compared compensators with a few (2-5) degrees of freedom. Their results show that compensators with three or more degrees of freedom are somewhat more effective than the simple compensator. Although they used the spectral line and DOP monitors (without polarization scrambling) they did not study the properties of these objective functions, or compare them to the eye opening monitor. Moreover, because they used standard Monte Carlo simulations they were not able to access the rare large-PMD fibers that are of real interest when quantifying the performance of a compensator. I. Lima *et al.* [39] used multiple importance sampling to study the performance of a simple compensator with a fixed DGD. They showed that the optimal value for the fixed DGD in the compensator is about 2–3 times larger than the mean DGD of the transmission line, averaged over fiber realizations. However, they only used the eye opening objective function in their work. In addition, none of the papers just cited carefully studies how the performance of a compensator depends on the choice of algorithm used to optimize the objective function.

We compare the spectral line, polarization-scrambled DOP, and eye opening objective functions, both for a particular fiber realization and statistically over many fiber realizations. In the special case that the PMD is frequency-independent, we derive analytical formulae for the spectral line and polarization-scrambled DOP objective functions. Each objective function can be regarded as a doubly periodic function on a two-dimensional plane that parametrizes a certain set of rotations of three-dimensional space. Our formulae show that the polarization-scrambled DOP has a single maximum, while the spectral line can have at least two maxima. However, there may be more local maxima in the general case of frequency-dependent PMD.

We also systematically study the combined effect that the choice of feedback signal and optimization algorithm have on the performance of a simple PMD compensator. By using importance sampling with a large number of fiber realizations, we are able to assess performance for the very rare large-PMD realizations of the fiber that are most important to consider when studying PMD compensators. Because the eye opening is so highly correlated to the bit-error ratio, it is very reasonable to assume that the best performance is obtained when the rotation of the polarization controller is given by the global maximum of the eye opening objective function. However, it is difficult to locate the global maximum of the eye opening. On the other hand, we will show that it is much easier to locate the global maxima for the spectral line and polarization-scrambled DOP, which indicates that these two objective functions are smoother than the eye opening.

The main conclusion of our study is that the spectral line and polarization-

scrambled DOP act as smooth surrogates for the eye opening (or bit-error ratio) objective function. This result is obtained by comparing the performance of the compensators with different local and global optimization algorithms, and by statistical studies that compare the location of local and global maxima of the three different objective functions. We show that the best trade off between computational cost and performance is obtained using the polarization-scrambled DOP objective function with a multilevel optimization algorithm. This algorithm uses a global genetic algorithm followed by a local conjugate gradient algorithm. Preliminary work [64] compared the performance of the simple compensator with the three objective functions, but used only local optimization algorithms and did not compare the structure of the objective functions.

In §2, we review the basic mathematical models for optical fiber communications systems with PMD. In §3, we describe the PMD compensator we study in this work, and in §4 we derive formulae for the polarization-scrambled DOP and spectral line objective functions in the special case of only first-order PMD. In §5 we review the importance sampling algorithm and in §6 we describe the optimization algorithms we used. Finally, in §7 we present the results of our optimization case study.

2. Mathematical Models for Optical Communication Systems. In this section, we review the linear-PMD equation and the coarse-step algorithm that is used to generate random realizations of the fiber. We also explain how to calculate and measure the performance of an optical communication system.

2.1. The governing equations. In this subsection, we review the Manakov-PMD equation that describes the propagation of an optical signal through birefringent optical fiber. The birefringence of optical fiber, which is the physical cause of PMD in the signal, varies randomly along the fiber and over the course of time due to temperature variations and mechanical vibrations.

Light in an optical fiber propagates in two eigenmodes which are distinguished from each other by their polarization states. To model the propagation of light in an optical fiber, we choose coordinates, (x, y, z) , so that the positive z -axis is the propagation direction along the fiber and the (x, y) -plane is orthogonal to the fiber. The electric field of light propagating at a carrier frequency, ω_0 , is the real part of the complex-valued vector field, \mathbf{E} , which we express as [43]

$$\mathbf{E}(x, y, z, \tau) = \kappa [U_1(z, \tau)\mathbf{R}_1 + U_2(z, \tau)\mathbf{R}_2] \exp[i\beta(\omega_0, z)z - i\omega_0\tau]. \quad (2.1)$$

Here τ denotes physical time, and the dispersion relation of the fiber is determined by the frequency dependence of the wavenumber, $\beta(\omega_0, z)$. The vector fields \mathbf{R}_1 and \mathbf{R}_2 are two eigenfunctions that describe the (x, y) -dependence of the electric field, \mathbf{E} . These vector fields are orthogonal to the propagation direction z , and to each other. The functions U_1 and U_2 describe the slow variation of the envelope of the electric field about the rapidly varying carrier wave given by the exponential factor in equation (2.1). The column vector $\mathbf{U}(z, t) = (U_1(z, t), U_2(z, t))^T$, which is called the *Jones vector* of the light, models the *polarization state* of the light at (z, t) . Here we have transformed from physical time, τ , to retarded time $t = \tau - \frac{\partial\beta}{\partial\omega}(\omega_0, z)z$, which defines a coordinate system that is moving with the group velocity $[\frac{\partial\beta}{\partial\omega}(\omega_0, z)]^{-1}$. The normalization constant κ is chosen so that $|\mathbf{U}|^2 = |U_1|^2 + |U_2|^2$ corresponds to the *optical power* of the signal. The data is encoded onto the signal by allocating a time slot to each bit and varying the power of the signal so that the power is large in the time slots allocated to the ones and small in the time slots allocated to the zeros.

If the refractive index is perfectly axially symmetric, then the two eigenmodes are equal and the signal is not affected by PMD. However, in real fiber this degeneracy is broken due to imperfections in the fiber. Consequently, real optical fiber has a small *birefringence*: Light propagating in the two different eigenmodes travels at slightly different group velocities. Therefore, if the power of an optical pulse is split between the two polarization eigenmodes, \mathbf{R}_1 and \mathbf{R}_2 , then as it propagates through the fiber, the power will spread in the time domain and can become severely distorted. This phenomenon is called *polarization-mode dispersion* (PMD). As a result, optical power will be transferred between the time slots allocated to the different bits, potentially resulting in bit errors at the receiver. Polarization-mode dispersion is particularly difficult to combat because it is inherently stochastic in nature. As the distance, z , along the fiber increases, the eigenfunctions \mathbf{R}_1 and \mathbf{R}_2 rotate rapidly and randomly in the xy -plane, but are otherwise unchanged. At each fixed distance, they also rotate randomly over time on a scale of minutes to hours due to temperature variations and mechanical vibrations. Consequently, a PMD compensator must be continually optimized to correct for PMD-induced distortions in the received optical signal.

The equation governing the z -evolution of \mathbf{U} is the *coupled nonlinear Schrödinger equation* (CNLS). The CNLS is derived from Maxwell's equations by averaging over the rapid variations of the carrier wave, $\exp[i\beta(\omega_0, z)z - i\omega_0\tau]$, and over the eigenfunctions, \mathbf{R}_1 and \mathbf{R}_2 [43]. The CNLS states that

$$\frac{\partial \mathbf{U}}{\partial z} = g\mathbf{U} + i\Delta\mathbf{B}\mathbf{U} - \Delta\mathbf{B}'\frac{\partial \mathbf{U}}{\partial t} - \frac{i}{2}\beta''\frac{\partial^2 \mathbf{U}}{\partial t^2} + i\gamma[|\mathbf{U}|^2\mathbf{U} - \frac{1}{3}(\mathbf{U}^\dagger\sigma_2\mathbf{U})\sigma_2\mathbf{U}]. \quad (2.2)$$

Here, the scalar coefficient g is the loss coefficient of the fiber. The factor $\Delta\mathbf{B} = \Delta\mathbf{B}(\omega_0, z)$ is the *birefringence matrix*, which is a 2×2 Hermitian matrix that models the anisotropy and asymmetry of the linear dielectric response tensor of the fiber, averaged with respect to the eigenfunctions \mathbf{R}_1 and \mathbf{R}_2 . The matrix $\Delta\mathbf{B}'$ is defined by $\Delta\mathbf{B}' = \frac{\partial \Delta\mathbf{B}}{\partial \omega}(\omega_0, z)$, and the scalar $\beta'' = \frac{\partial^2 \beta}{\partial \omega^2}(\omega_0, z)$ is the chromatic (frequency-dependent) dispersion. The scalar coefficient γ measures the strength of the Kerr nonlinearity in the fiber. The Kerr nonlinearity arises from the fact that the refractive index of optical fiber has a small dependence on the optical power $|\mathbf{U}|^2$ of the light. Finally, \dagger denotes conjugate transpose, and σ_2 is the second of the three Pauli spin matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.3)$$

Because of the polarization properties of glass [48], optical fiber is *linearly birefringent*, which means that

$$\Delta\mathbf{B} = \Delta\beta(\cos\theta\sigma_3 + \sin\theta\sigma_1), \quad (2.4)$$

where $\Delta\beta$ is the magnitude of birefringence and θ is an orientation angle whose significance will be explained later. Since θ is a very weak function of frequency, we assume that $\Delta\mathbf{B}' = \Delta\beta'(\cos\theta\sigma_3 + \sin\theta\sigma_1)$. Wai and Menyuk [58] proposed a model for the random variation of the birefringence along the fiber in which $\Delta\beta'\cos\theta$ and $\Delta\beta'\sin\theta$ are independent Gaussian random processes with mean zero and the same standard deviation [58]. Recently, Galtarossa *et al.* [19], [20] experimentally validated this model.

The random variations in the fiber birefringence occur on a length scale of 1–100 m and result in very rapid changes in the polarization state of the light as it propagates.

However, since optical communication systems are at least several hundred kilometers long, it is not computationally feasible to simulate propagation through birefringent fiber by using a numerical method that takes small enough steps along the fiber to track these rapid changes in the polarization state of the light. This problem can be overcome by transforming the rapid changes in the polarization state of the light at the carrier frequency out of the CNLS to obtain the *Manakov-PMD equation* [42], [43]:

$$\frac{\partial \mathbf{W}}{\partial z} = g\mathbf{W} - \Delta\beta' \bar{\sigma}_3 \frac{\partial \mathbf{W}}{\partial t} - \frac{i}{2} \beta'' \frac{\partial^2 \mathbf{W}}{\partial t^2} + i\gamma |\mathbf{W}|^2 \mathbf{W}. \quad (2.5)$$

Here $\mathbf{W}(z, t) = \mathbf{Q}(z)\mathbf{U}(z, t)$, where $\mathbf{Q}(z)$ is a unitary transformation and $\bar{\sigma}_3 = \mathbf{T}(z)^{-1} \sigma_3 \mathbf{T}(z)$ for a matrix $\mathbf{T}(z)$ that is determined by the birefringence parameters $\Delta\beta$ and θ . To explain the rationale for making a transformation of the form $\mathbf{Q}(z)$, we first note that the Fourier conjugate of the retarded time, t , is frequency, ω , measured relative to the carrier frequency, ω_0 , of the optical signal. Even though the Fourier transform, $\hat{\mathbf{U}}(z, \omega)$, of $\mathbf{U}(z, t)$ varies rapidly with the propagation distance z , it only has a very weak dependence on frequency, ω . Therefore, it makes sense to transform \mathbf{U} so that the new coordinates exactly follow the rapid changes of the polarization state of the signal at the center frequency, ω_0 , *i.e.*, so that $\hat{\mathbf{W}}(z, 0)$ is constant in z . In this new coordinate system, the Fourier transform, $\hat{\mathbf{W}}(z, \omega)$, of the solution of the Manakov-PMD equation measures the slow variation of the polarization state of the light at each frequency, ω , with respect to the polarization state of the light at the carrier frequency, ω_0 .

2.2. The linear PMD equation. In this subsection we review the linear PMD equation which is a special case of the Manakov-PMD equation. The linear PMD equation is appropriate for studying the statistical behavior of PMD and PMD compensators. We will use this equation to explain how PMD results in the spreading and distortion of optical pulses.

The linear PMD equation is obtained by omitting all but the second term on the right-hand side of equation (2.5). Another widely studied special case of the Manakov-PMD equation is the scalar nonlinear Schrödinger equation which is the equation obtained when there is no birefringence ($\Delta\beta' = 0$) and the optical signal at the transmitter is polarized [25], [43]. In many systems, the primary source of bit errors is not the nonlinear effects that are modeled by the scalar nonlinear Schrödinger equation, but rather the effects of PMD that are modeled by the linear PMD equation. Moreover, because PMD is a stochastic effect, and because bit errors are so rare, PMD must be studied statistically using a large number of randomly chosen realizations of the birefringence of the optical fiber. Simulations that do not include the Kerr nonlinearity are computationally several orders of magnitude faster than those that do.

We now explain how PMD gives rise to the spreading and distortion of optical pulses, and introduce the concept of differential group delay. The linear PMD equation is most readily analyzed in the frequency domain, where

$$\frac{\partial \hat{\mathbf{W}}}{\partial z} = i\Delta\beta' \bar{\sigma}_3(z) \omega \hat{\mathbf{W}}. \quad (2.6)$$

The solution to this equation can be expressed in the form $\hat{\mathbf{W}}(z, \omega) = \hat{f}(\omega) \hat{\mathbf{A}}(z, \omega)$, where \hat{f} is a real scalar-valued function, and $|\hat{\mathbf{A}}(z, \omega)|^2 = 1$. As a function of time, the

power of the optical signal is given by $|f|^2$, where f is the inverse Fourier transform of \hat{f} . The vector $\hat{\mathbf{A}}$ is the polarization state of the signal. Suppose that $\mathbf{F} = \mathbf{F}(z, \omega)$ is a matrix such that

$$\frac{\partial \hat{\mathbf{A}}}{\partial \omega} = i\mathbf{F}\hat{\mathbf{A}}. \quad (2.7)$$

It can be shown that \mathbf{F} is a Hermitian matrix that is determined by the quantities $\Delta\beta'$ and $\bar{\sigma}_3$ that characterize the birefringence of the optical fiber [43]. The absolute difference of the two real eigenvalues of \mathbf{F} is called the *differential group delay* (DGD), and the eigenvectors of \mathbf{F} are called the *principal states of polarization*. The DGD and the principal states of polarization depend on both the propagation distance, z , and the frequency, ω .

To see how the DGD is related to pulse spreading, suppose for simplicity that the matrix \mathbf{F} is ω -independent, at least over the frequency bandwidth of the signal, $\hat{\mathbf{W}}$. In this case, we say that the fiber birefringence generates only *first-order PMD*. Diagonalizing \mathbf{F} , we see that

$$\hat{\mathbf{A}}(z, \omega) = \mathbf{P}e^{i\omega\mathbf{D}(z)}\mathbf{P}^\dagger\hat{\mathbf{A}}(z, 0), \quad (2.8)$$

where $\mathbf{P} = (\mathbf{v}_1, \mathbf{v}_2)$ is unitary, and $\mathbf{D}(z) = \text{diag}\left(-\frac{\tau(z)}{2}, \frac{\tau(z)}{2}\right)$. Here $\tau(z)$ is the DGD. (By ignoring a common phase, we can assume that $\text{trace}(\mathbf{F}) = 0$.) Therefore, the optical signal is given by

$$\mathbf{W}(z, t) = c_1 f\left(t + \frac{\tau(z)}{2}\right)\mathbf{v}_1 + c_2 f\left(t - \frac{\tau(z)}{2}\right)\mathbf{v}_2, \quad (2.9)$$

where c_1 and c_2 are complex constants with $|c_1|^2 + |c_2|^2 = 1$. The DGD is therefore the time delay between light that is launched in the two different principal states of polarization. If the power of an optical pulse is split between the two principal states (*i.e.*, $c_k \neq 0$ for $k = 1, 2$), and the DGD is large, then the power of the pulse is spread out and distorted as a function of time and bit errors are more likely to occur at the receiver.

2.3. The Stokes representation. In our discussion so far, we have modeled the polarization state of light using the Jones representation. In this subsection, we review an alternate approach based on the Stokes representation [5], [24], [43], [44]. Using the Stokes representation, we can regard the propagation of the polarization state of light through birefringent optical fiber as a random walk on a two-dimensional sphere.

With the Jones representation, polarization states are represented as unit vectors $\mathbf{U} \in \mathbf{C}^2$, *i.e.*, as points on the sphere $S^3 \subset \mathbf{C}^2$, while with the Stokes representation they are represented using unit Stokes vectors, $\mathbf{S} = (S_1, S_2, S_3) \in S^2 \subset \mathbf{R}^3$. The sphere of unit Stokes vectors is called the *Poincaré sphere*. The mapping $\psi : S^3 \rightarrow S^2$ from Jones space to Stokes space is defined by $\mathbf{S} = \psi(\mathbf{U}) = \mathbf{U}^\dagger \vec{\sigma} \mathbf{U}$ where $\vec{\sigma} = (\sigma_3, \sigma_1, -\sigma_2)$ is the Pauli spin vector. For each \mathbf{S} , the inverse image $\psi^{-1}(\mathbf{S})$ is the circle in S^3 that consists of all Jones vectors of the form $\mathbf{U}_\varphi = e^{i\varphi}\mathbf{U}$, where $\varphi \in S^1$ and $\psi(\mathbf{U}) = \mathbf{S}$. The angle φ is an overall phase that plays no role in the theory of PMD. Since the matrix acting on the right-hand side of equation (2.6) is an element of the Lie algebra $\mathfrak{su}(2)$ of trace-free skew-Hermitian matrices, the solution of the linear-PMD equation is of the form $\hat{\mathbf{W}}(z, \omega) = \mathbf{T}(z, \omega)\hat{\mathbf{W}}(0, \omega)$ where the PMD-transmission matrix, \mathbf{T} , is an element of the Lie Group $\text{SU}(2)$ of unitary matrices

with determinant 1. The action of an element \mathbf{T} of $SU(2)$ on $S^3 \subset \mathbf{C}^2$ is equivalent to the action of an element \mathbf{R} of the special orthogonal group, $SO(3)$, on S^3 , where \mathbf{T} is mapped to \mathbf{R} by the 2-1 and onto map $\Psi : SU(2) \rightarrow SO(3)$ that is determined by $\mathbf{R}\vec{\sigma} = \mathbf{T}^\dagger \vec{\sigma} \mathbf{T}$. In particular, if \mathbf{R} is a rotation by an angle θ about an axis $\hat{r} \in S^2 \subset \mathbf{R}^3$, then $\mathbf{T} = \pm[\cos(\theta/2)\mathbf{I} - i\sin(\theta/2)\hat{r} \cdot \vec{\sigma}]$. The Lie algebra $\mathfrak{so}(3)$ of $SO(3)$ consists of all anti-symmetric matrices. The induced map between Lie algebras, $\Psi_* : \mathfrak{su}(2) \rightarrow \mathfrak{so}(3)$, is defined by $\Psi_*(i\vec{\beta} \cdot \vec{\sigma}) = \vec{\beta} \times$. Here $\vec{\beta} = (\beta_1, \beta_2, \beta_3) \in \mathbf{R}^3$ and

$$\vec{\beta} \times = \begin{pmatrix} 0 & -\beta_3 & \beta_2 \\ \beta_3 & 0 & -\beta_1 \\ -\beta_2 & -\beta_1 & 0 \end{pmatrix} \quad (2.10)$$

determines an isomorphism between $\mathfrak{so}(3)$ and \mathbf{R}^3 . Therefore, if $\hat{\mathbf{S}}$ is the Stokes vector that is equivalent to the Jones vector $\hat{\mathbf{A}}$ defined below (2.6), then in Stokes space, the linear PMD equation (2.6) is given by

$$\frac{\partial \hat{\mathbf{S}}}{\partial z} = \vec{\beta} \times \hat{\mathbf{S}}, \quad (2.11)$$

where $\vec{\beta} = \vec{\beta}(\omega, z)$ is the *local birefringence vector* of the fiber. If the local birefringence vector is constant, then the polarization state $\hat{\mathbf{S}}$ at a single frequency traces a circle on the Poincaré sphere centered at $\vec{\beta}$. However, in real linearly birefringent fibers, the local birefringence vector, $\vec{\beta}$, moves randomly on the equator of the sphere.¹ Therefore, the polarization state of the light, $\hat{\mathbf{S}}$, moves randomly over the entire sphere. More specifically, given a length of birefringent fiber and an optical signal with a given input polarization state, consider the probability distribution of output polarization states, $\hat{\mathbf{S}}$, on the Poincaré sphere, where the samples are generated from different realizations of the fiber birefringence. The length scale required for the probability distribution of $\hat{\mathbf{S}}$ to become uniform on the sphere is on the order of a kilometer, which is short compared to the total length of a communication system. These observations provide the primary motivation for the coarse-step method for modeling PMD, which we will discuss in §2.4 below.

To explain how the DGD evolves as a function of distance along the fiber, we introduce the *polarization dispersion vector*, $\vec{\Omega} = \vec{\Omega}(\omega, z) = (\Omega_1, \Omega_2, \Omega_3)$, which is defined in terms of the Hermitian matrix \mathbf{F} in equation (2.7) by $\mathbf{F} = \text{trace}(\mathbf{F})\mathbf{I} + \frac{1}{2}\vec{\Omega} \cdot \vec{\sigma}$. Then the Stokes formulation of equation (2.7) is

$$\frac{\partial \hat{\mathbf{S}}}{\partial \omega} = \vec{\Omega} \times \hat{\mathbf{S}}. \quad (2.12)$$

Consequently, the unit vectors $\pm \vec{\Omega}/|\vec{\Omega}|$ represent the two principal states of polarization of the fiber on the Poincaré sphere, and the magnitude, $|\vec{\Omega}|$, is the DGD. Finally, the *dynamical PMD equation*, which describes the evolution of the polarization dispersion vector is given by [46]

$$\frac{\partial \vec{\Omega}}{\partial z} = \frac{\partial \vec{\beta}}{\partial \omega} + \vec{\beta} \times \vec{\Omega}. \quad (2.13)$$

This equation can be derived by differentiating (2.11) with respect to ω and (2.12) with respect to z .

¹The angle θ in (2.4) is half the angle between $\vec{\beta}$ and the positive X -axis.

2.4. The probability space of fiber realizations. Each choice of a set of random local birefringence vectors along the fiber is called a *fiber realization*. In this subsection, we describe the coarse-step method that is used to generate different fiber realizations.

When a PMD compensator is used, each time the fiber realization changes, the compensator needs to be reset via optimization. In reality, the fiber realization can both drift gradually over time due to temperature fluctuations and change abruptly to an unrelated realization in response to large disturbances such as a truck passing overhead [6], [32], [52], [57]. We assume that when the fiber realization changes, there is no correlation between the old and new realizations. Statistically we model PMD by defining an appropriate space of fiber realizations, imposing a probability distribution on this space, and devising a method for randomly sampling the space of fiber realizations. To do so, it is commonly assumed that optical fiber is homogeneous. In other words, the statistical properties of the local birefringence vector $\vec{\beta}$ do not depend on distance, z , along the fiber. We will consider a space of fiber realizations consisting of fibers of a given length with a prescribed average DGD. In reality, such a space of fiber realizations could correspond either to fibers that are fabricated using the same manufacturing process, or to realizations of a single fiber whose birefringence is randomly varying over time.

The most commonly used method for generating fiber realizations with a given average DGD is the *coarse-step method* [13], [59] which we now describe using the Stokes' representation [44]. The coarse-step method can be regarded as a computationally efficient, statistically correct, numerical method for solving the linear-PMD (or Manakov-PMD) equation. In the coarse-step method, the action of a birefringent fiber on the Stokes vector, $\hat{\mathbf{S}}$, of the light is modeled as the concatenation of the action of N segments of fixed-birefringence fiber, each of which is preceded by a random rotation of $\hat{\mathbf{S}}$ on the Poincaré sphere. If $\hat{\mathbf{S}}^{(n)}(\omega)$ denotes the Stokes vector of light at frequency ω after the n -th fiber segment, then

$$\hat{\mathbf{S}}^{(n)}(\omega) = \mathbf{R}(\omega) \mathbf{Q}_n \hat{\mathbf{S}}^{(n-1)}(\omega). \quad (2.14)$$

Here, the matrix $\mathbf{R}(\omega)$ is the element of $\text{SO}(3)$ that is the rotation about the X -axis through an angle $\Delta\beta'\omega\Delta z$. This rotation models the propagation of light through each of the fixed-birefringence fiber segments. The quantity $\Delta\beta'$ is the magnitude of the frequency derivative of the local birefringence vector of each of the N fiber segments, and is related to the average DGD by $\Delta\beta' = (3\pi N/8)^{1/2} \overline{\text{DGD}}/L$, where L is the length of the fiber [42]. The matrix \mathbf{Q}_n is a frequency-independent, random rotation that is chosen using the canonical uniform probability distribution on $\text{SO}(3)$. Specifically, \mathbf{Q}_n can be expressed as an Euler-angle rotation matrix [23] of the form $\mathbf{Q}_n = \mathbf{R}_X(\psi_n) \mathbf{R}_Y(\theta_n) \mathbf{R}_X(\phi_n)$. Here, $\mathbf{R}_X(\psi)$ is a rotation about the X -axis through an angle ψ . The angles ψ_n and ϕ_n are uniformly distributed in $[0, 2\pi]$, and $\cos\theta_n$ is uniformly distributed in $[-1, 1]$. In the special case that $N = 1$, the coarse-step method generates only first-order PMD. For a large enough number of fiber segments, the coarse-step method produces the same statistical properties as are obtained using Wai and Menyuk's model for the randomly varying birefringence [42]. Moreover, the coarse-step method is much more computationally efficient, since it takes steps on the order of a kilometer, rather than the meter-long steps required to track the rapid variations in the birefringence.

The dynamical PMD equation (2.13) can also be solved using a coarse-step approach. If $\vec{\Omega}^{(n)} = \vec{\Omega}^{(n)}(z, \omega)$ is the polarization dispersion vector after the n -th fiber

segment, then

$$\vec{\Omega}^{(n)} = R(\omega) \left[\mathbf{Q}_n \vec{\Omega}^{(n-1)} + \Delta \vec{\Omega}_n \right], \quad (2.15)$$

where $\Delta \vec{\Omega}_n = \Delta \beta' \Delta z \vec{e}_X$ is the polarization dispersion vector of the n -th segment. Here $\vec{e}_X = (1, 0, 0)^T$, and we regard $\vec{\Omega}$ as a column vector. In §3 below, we will use Eq. (2.15) to explain the basic idea behind PMD compensation.

There is a large literature on the statistical properties of PMD. The most important result is that in the limit as $N \rightarrow \infty$, the DGD is Maxwellian distributed with distribution $f_{\text{DGD}}(x) = \frac{2\pi x^2}{\alpha^3} \exp(-x^2/2\alpha^2)$, where $\alpha = (\pi/8)^{1/2} \overline{\text{DGD}}$ [46]. In particular, there is an extremely small probability that the DGD of a fiber realization is significantly larger than the average DGD.

2.5. The receiver and performance evaluation metrics. To evaluate the degree to which a PMD compensator reduces the probability of errors due to PMD, we also need a model of the receiver. The purpose of the receiver is to convert the transmitted optical signal into an electrical current, to determine a *clock time* that is used to set the beginning and ending points of the time intervals for each of the bits being transmitted [55], and finally to make a decision as to whether the voltage of the received electrical current in each of these bit slots is to be received as a one or a zero. This decision is based on a choice of *decision voltage*. If the received voltage is larger than the decision voltage, the bit is declared to be a one. Otherwise a zero is received.

A receiver model should include an algorithm to evaluate the performance of the communication system. There are several ways to measure performance — the most fundamental means is via the *bit error ratio*, which is given by $\text{BER} = \frac{1}{2}(p_{1|0} + p_{0|1})$. Here $p_{1|0}$ is the probability of receiving a one given that a zero was transmitted, and $p_{0|1}$ is the probability of receiving a zero given that a one was transmitted. In a real system bit errors occur due to the combined effect of PMD and noise from optical amplifiers. Since we did not include noise in our simulations, we measured the performance using a quantity called the *eye opening* [62] rather than the BER. The eye opening of a noise-free signal at the receiver is defined to be the difference between the smallest electrical voltage of a one and the largest electrical voltage of a zero at the clock time.² There is a strong correlation between the BER and the eye opening [41], [49], [63]. To study the degree to which PMD degrades the performance of the system, we define the *eye opening penalty* for a particular fiber realization to be the ratio between the eye opening for a version of the system that just has a transmitter and receiver (*i.e.*, no fiber and hence no PMD), to the eye opening for the system with a transmitter, PMD due to the given fiber realization, and a receiver. The more the PMD reduces the eye opening, the larger the eye opening penalty for that fiber realization. System designers typically specify that a *system outage* occurs if the eye opening penalty exceeds a specified threshold, such as 2 dB [52].³ They require that the *outage probability*, which is the probability that such an outage occurs, be on the order of 10^{-3} to 10^{-6} , corresponding to a few minutes to hours of outage per year. A major goal of this paper is to present a numerical model that can be used

²The term “eye opening” is used here because in Fig. 3.2 the image of each signal under the mapping $t \mapsto t \bmod T$ looks like an eye, where T is the bit period.

³A linear factor of x corresponds to $10 \log_{10} x$ decibels (dB). So, if the eye opening penalty for a particular fiber realization is 2 dB, then the eye opening is 63% of the value one would see in a system without PMD.

to assess the degree to which a PMD compensator can reduce the outage probability due to PMD.

3. Polarization Mode Dispersion Compensation. In this section, we discuss the optical PMD compensator whose performance we study in this paper. As discussed in §2, imperfections in optical fibers result in two principal states of polarization for the light. Light traveling in the fast principal state arrives at the receiver ahead of the light traveling in the slow state. In practice the power in the optical signal is split between these two principal states, so that the optical pulses used to encode the binary data become spread out and distorted. Consequently, over long distances, the message being transmitted will be corrupted by errors. To compensate for this spread of information, physical devices called optical PMD compensators can be used.

A variety of designs have been proposed for optical PMD compensators [52], [53]. We chose to study a simple compensator that is easy to build and operate. To motivate the design of this compensator, we consider the important special case that the transmission fiber has only first-order PMD, *i.e.*, that the polarization dispersion vector, $\vec{\Omega}_T$, of the transmission line is frequency independent. Recall from (2.9) that if the polarization state of the signal is not closely aligned with either of the principal states of the transmission fiber, then the larger the DGD, $|\vec{\Omega}_T|$, the more the signal will be distorted due to PMD, and the greater the probability of a bit error. The simple PMD compensator we study is a device that can at least partially cancel out the DGD of the transmission fiber. The idea is to insert a segment of *compensation fiber* between the transmission fiber and the receiver, and to rotate the polarization state of the signal between the transmission and compensation fibers so as to align the fast principal state $\vec{\Omega}_T$ of the transmission fiber with the slow principal state $-\vec{\Omega}_C$ of the compensation fiber. Then by equation (2.15), the total polarization dispersion vector, $\vec{\Omega}_R$, at the receiver is given by

$$\vec{\Omega}_R = \mathbf{Q}\vec{\Omega}_T + \vec{\Omega}_C, \quad (3.1)$$

where \mathbf{Q} is the rotation between the transmission and compensation fibers. Consequently, if the DGD of the compensation fiber were equal to that of the transmission fiber, then the total DGD at the receiver would be zero, $|\vec{\Omega}_R| = 0$.

Simple compensators based on this principle have been built, and a diagram of one is shown in Fig. 3.1. The optical signal generated in the transmitter propagates through the transmission fiber. The compensator itself is located immediately prior to the receiver and consists of a polarization controller that can be adjusted to transform the polarization state of the fiber by any desired rotation, followed by a short segment of compensation fiber. The compensation fiber is designed to be polarization maintaining in that its principal states of polarization and its DGD are fixed. Notice that since the transmission DGD, $|\vec{\Omega}_T|$, of a fiber realization may not equal the fixed DGD, $|\vec{\Omega}_C|$, of the compensator, the total DGD, $|\vec{\Omega}_R|$, in Eq. (3.1) may not be zero. In addition, since this compensator consists only of one polarization controller and one segment of polarization maintaining fiber, it can only compensate for the DGD at the carrier frequency, and not for PMD distortions at all frequencies in the signal.

After passing through the compensation fiber, the optical signal is monitored and a feedback loop is used to adjust the rotation performed by the polarization controller. Given a realization of the birefringence in the transmission fiber, the feedback loop

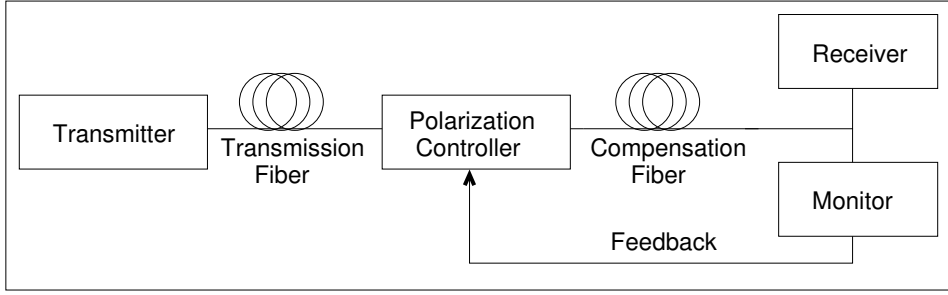


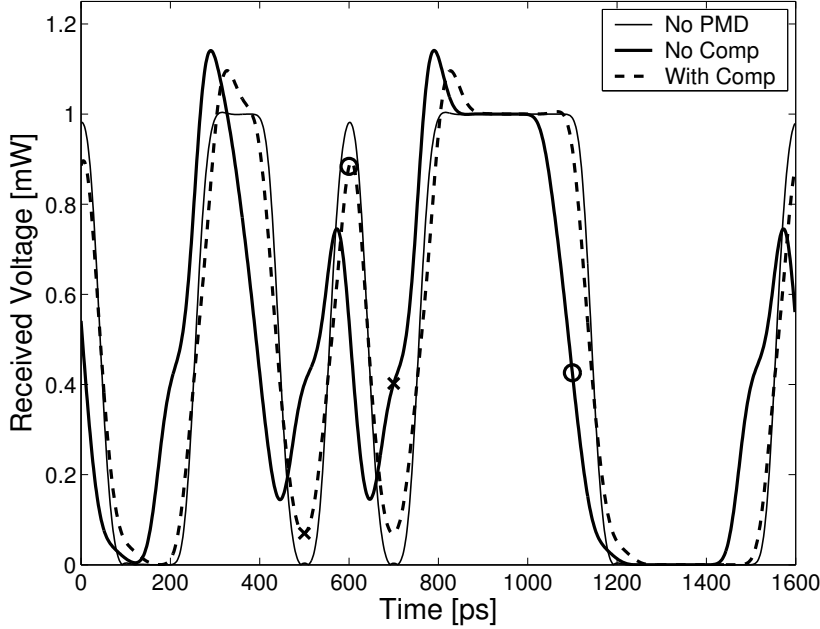
FIG. 3.1. An optical communication system with a simple PMD compensator

can be modeled by optimizing the function from the state space $SO(3)$ of all possible rotations of the polarization controller to \mathbf{R} , that is given by the monitor.

In Figure 3.2, we show the effect that PMD can have on a signal and how a PMD compensator can decrease the signal distortion due to PMD. The results we show are for a particular fiber realization. Different fiber realizations could have quite different effects on the signal. We plot the received electrical current as a function of time in three cases. The thin solid curve shows the case where there is no PMD in the transmission line and hence no pulse distortion. (This signal is called the *back-to-back* signal as the transmitter and receiver abut each other.) The data pattern 1001101011110000 can be easily recognized in the signal. Notice that the voltage in the signal does not return to zero between two consecutive ones. Signals like this are called *non return-to-zero* and are commonly used in communication systems. The thick solid curve shows the same signal after it has traveled through the transmission fiber with PMD. The DGD was 75 ps at the carrier frequency and there was a strong frequency-dependence to the polarization dispersion vector. The signal has clearly been distorted due to PMD. The circle at 1100 ps shows the minimum-voltage one, and the cross at 700 ps shows the maximum-voltage zero. The eye opening, which is the difference between the height of the circle and the cross, is 0.02 mV. Therefore the eye opening after the transmission fiber is very small compared to the back-to-back eye opening which is 0.98 mV. Finally the thick dashed curve shows this same signal after compensation. The circle at 600 ps shows the minimum-voltage one, and the cross at 500 ps shows the maximum-voltage zero. The compensator has increased the eye opening to 0.81 mV, which is a marked improvement over the uncompensated case.

In this paper, we study the performance of this PMD compensator for three choices of monitor. Different monitors and different random realizations of the birefringence in the transmission fiber correspond to different choices of the objective function to be optimized. Since our ultimate goal is to minimize the bit error ratio, the monitor should be chosen so that the monitored value is strongly correlated (or anti-correlated) to the bit error ratio, *i.e.*, to the eye opening. The most obvious choice of monitor is the bit error ratio itself, or the eye opening. However, it is not possible to measure the bit error ratio in a real system, and it is often not feasible to measure the eye opening in real time.

We will now briefly describe each of the monitors used in our numerical experiments — the eye opening, spectral line, and the degree of polarization (DOP) ellipsoid. The *eye-opening monitor* measures the eye opening of the signal after propagation through the system, relative to the eye opening of the back-to-back signal. Optical

FIG. 3.2. *Compensated vs Uncompensated Signals*

signals typically have carrier frequencies of about 200 THz (or about 2×10^{12} Hz) and a bandwidth of about 20 GHz (or 2×10^{10} Hz). The optical signal is sent through a photodetector to convert it from an optical to an electrical signal. The electric signal has a shifted frequency spectrum (shifted relative to the optical signal) in the range $[0, 10]$ GHz. The *spectral line* monitoring technique requires that an electric filter be used to monitor the power in a particular frequency (or tone). In our work we have used a filter to monitor the power in the 5 GHz tone using a window of width 0.5 GHz. The spectral line feedback mechanism attempts to maximize the power in this tone relative to the reference back-to-back signal which has gone straight from transmitter to receiver without encountering the optical fiber (hence without being affected by PMD).

In the case of only first-order PMD, we will show in Eq. (4.17) below that the power in the 5 GHz spectral line decreases monotonically with increasing DGD between DGD values of 0 and 100 ps (the width of each bit slot; see also Figure 1 in [51]). Thus it would be hoped that for moderate amounts of DGD, an optimization algorithm which maximizes the power in this spectral line (or frequency) will compensate for the DGD present in the fiber. As has been discussed earlier, the eye-opening is correlated to the DGD, and we see now that for first-order PMD, the amount of DGD is correlated to the power in the spectral line. Thus this feedback measure achieves our aim of providing a real-valued function correlated to the eye-opening. In the case of higher-order PMD, this correlation becomes more complicated.

We now explain why the functions to be optimized can be regarded as being defined on the 2-sphere, S^2 , rather than on the three-dimensional manifold, $SO(3)$. As in our discussion of the coarse-step method in §2.4, the rotation \mathbf{Q} in (3.1) can be expressed in the form $\mathbf{Q} = \mathbf{R}_X(\psi)\mathbf{R}_Y(\theta)\mathbf{R}_X(\phi)$, where $\psi, \theta, \phi \in [0, 2\pi]$, and $\mathbf{R}_X(\phi)$ is a rotation by an angle ϕ about the X -axis. We assume that the polarization dispersion

vector, $\vec{\Omega}_C$, of the compensator is parallel to the X -axis. Consequently, $\vec{\Omega}_R = \mathbf{R}_X(\psi) [\mathbf{R}_Y(\theta) \mathbf{R}_X(\phi) \vec{\Omega}_T + \vec{\Omega}_C]$. Since the final rotation, $\mathbf{R}_X(\psi)$, does not affect the values of any of the objective functions we can ignore it and regard $\mathbf{Q} = \mathbf{Q}(\phi, \theta)$ as a function of (ϕ, θ) alone, *i.e.*,

$$\vec{\Omega}_R(\phi, \theta) = \mathbf{R}_Y(\theta) \mathbf{R}_X(\phi) \vec{\Omega}_T + \vec{\Omega}_C. \quad (3.2)$$

Since $\mathbf{R}_X(\pi) = \text{diag}(1, -1, -1)$, $\mathbf{R}_Y(-\theta) \mathbf{R}_X(\pi) = \mathbf{R}_X(\pi) \mathbf{R}_Y(\theta)$. Therefore, the rotation $\mathbf{Q}(\phi + \pi, -\theta)$ has the same effect as the rotation $\mathbf{Q}(\phi, \theta)$. Consequently, we can regard the objective functions as being defined on the rectangle $R = \{(\phi, \theta) \in [-\pi, \pi] \times [0, \pi]\}$. Moreover, ignoring a final rotation about the X -axis, $\mathbf{Q}(\phi, 0) = I$ and $\mathbf{Q}(\phi, \pi) = \mathbf{R}_Y(\pi)$ are constants, independent of ϕ . Therefore, the space of rotations performed by the polarization controller is actually diffeomorphic to the sphere, S^2 : The rotation $\mathbf{Q}(\phi, \theta)$ corresponds to the point on the sphere with spherical coordinates (ϕ, θ) , where $\phi = \phi_0$ is a circle of longitude and $\theta = \theta_0$ is a circle of latitude. Consequently, the objective functions are actually defined on S^2 .

Next, we examine the structure of the objective function for the eye-opening and spectral-line monitors. In Fig. 3.3, we show a contour plot of the eye opening objective function for a particular fiber realization. In this figure, we have parametrized the sphere in the space of rotations using spherical coordinates (ϕ, θ) so that the horizontal lines $\theta = -\pi$ and $\theta = \pi$ map to the south and north poles respectively, and the vertical lines $\phi = \pm\pi$ both map to the same great semicircle of longitude. The eye-opening objective function in Fig. 3.3 has two local maxima located close to $(\phi, \theta) = (-\pi, \pi/2)$ and $(\phi, \theta) = (0, \pi/2)$. The spectral-line objective function in Fig. 3.4 has a similar structure, although it is smoother than the eye-opening objective function. In particular we note that the steep ridge located between $\phi = 0$ and $\phi = -\pi/2$ for the eye objective function has been considerably smoothed out in the spectral line objective function diagram. Thus local optimization techniques would be more effective at finding optima for the spectral line than for the eye in this case.

To motivate the degree of polarization ellipsoid monitor, we observe that when a signal propagates through fiber with PMD there are two basic reasons why the eye-opening at the receiver can be large: either the total DGD is small or the input state of polarization of the signal is aligned with one of the principal states of the fiber [47]. In Fig. 3.3, the eye-opening is large near $(-\pi, \pi/2)$ because the total DGD is small there, and it is large near $(0, \pi/2)$ because the rotation of the polarization controller is such that the principal state of the entire length of fiber, $\vec{\Omega}_R$, is parallel to the input state of polarization of the signal. In a real system, the input polarization state can drift over time. Consequently, it is better to operate a PMD compensator near where the DGD is minimized rather than near where the input polarization state of the signal is aligned with a principal state of the fiber [52]. The degree of polarization (DOP) ellipsoid is one such monitor.

The *DOP ellipsoid* monitor is obtained by polarization-scrambling the degree of polarization (DOP) of the signal. The DOP is used to monitor PMD because polarized signals become depolarized when they propagate through fiber with PMD. The degree to which the signal is depolarized depends on the DGD and on the input state of polarization. Any optical signal can be decomposed as the sum of polarized and unpolarized components [5]. The degree of polarization (DOP) is the ratio of the power in the polarized component to the total power. If $\mathbf{U}(t) = f(t) \mathbf{U}_0$ denotes the Jones vector of an input polarized signal, then the total power is $S_0 = \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 d\omega$,

where \hat{f} is the Fourier transform of f . The power in the polarized component of the signal is given by the length of the average Stokes vector,

$$\mathbf{S} = \int_{-\infty}^{\infty} \hat{\mathbf{S}}(\omega) |\hat{f}(\omega)|^2 d\omega, \quad (3.3)$$

where $\hat{\mathbf{S}}(\omega)$ is the Stokes vector at frequency ω defined in §2.3. The DOP is given by $\text{DOP} = |\mathbf{S}|/S_0$.

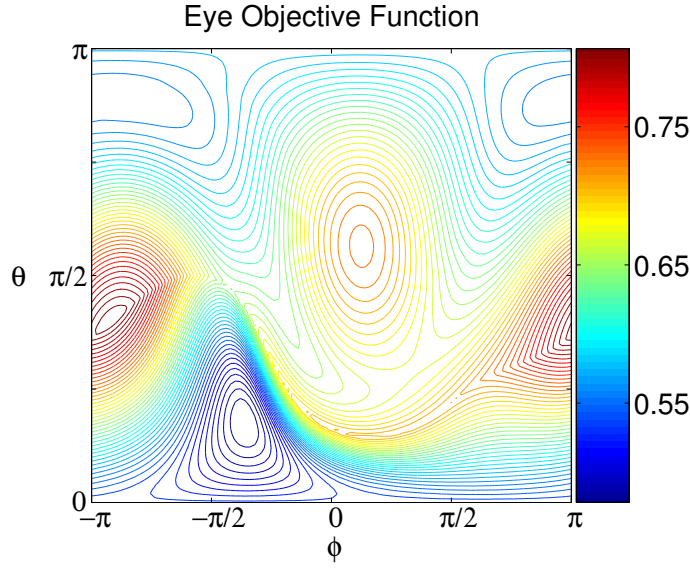
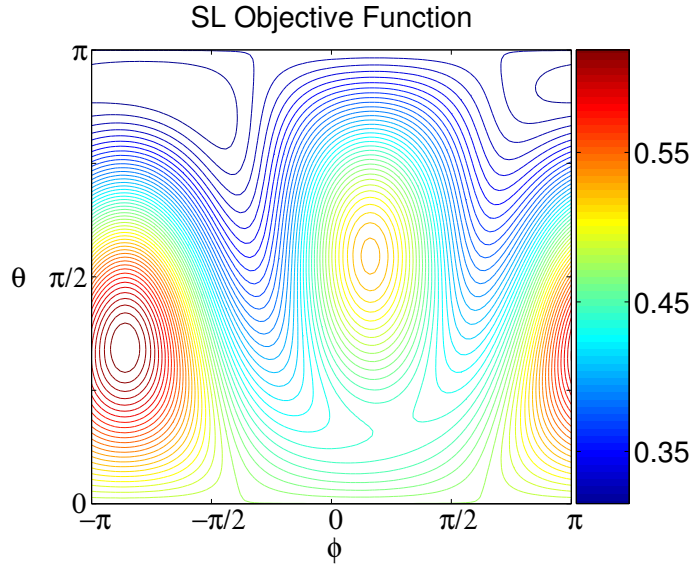
To explain why a large DGD results in a low DOP, consider a fiber with only first-order PMD for which the polarization dispersion vector is fixed at the north pole. Then, as a function of frequency, ω , the Stokes vector $\hat{\mathbf{S}}(\omega)$ traces out an arc of a circle of latitude on the sphere. By equation (2.12), the larger the DGD, $|\vec{\Omega}|$, the longer the arc. Suppose, for example, that the input polarization state is such that the circle of latitude is the equator and the DGD is large enough so that $\hat{\mathbf{S}}(\omega)$ traces out the entire equatorial circle as ω varies over the bandwidth of the signal. Then, by symmetry, the integral in Eq. (3.3) is close to zero and so the DOP is small. More generally, since equation (3.3) is a weighted average of vectors, the larger the DGD, the smaller the DOP.

The DOP ellipsoid is defined [9], [14], [15], [50], [54] so that for each unit vector \mathbf{S}_{in} , polarized light with Stokes vector \mathbf{S}_{in} is sent in to a fiber with PMD, and the average output Stokes vector, \mathbf{S}_{out} , is measured. The set of all such vectors \mathbf{S}_{out} forms the DOP ellipsoid. The length of the shortest principal axis of the DOP ellipsoid is approximately given by [14]

$$\lambda_{\min} \approx 1 - \frac{1}{2} \overline{\Delta\omega^2} |\vec{\Omega}|^2. \quad (3.4)$$

Here $\overline{\Delta\omega^2} = (1/S_0) \int_{-\infty}^{\infty} \omega^2 |\hat{f}(\omega)|^2 d\omega$ is a measure of the square of the bandwidth of the signal and $|\vec{\Omega}|$ is the DGD at the carrier frequency. Therefore, with the DOP ellipsoid feedback mechanism, we aim to maximize λ_{\min} and therefore to minimize the DGD. To fit the ellipsoid we used a Euclidean invariant linear least-squares algorithm that minimized the algebraic distance to 36 output Stokes vectors \mathbf{S}_{out} [21]. In Fig. 3.5, we show a plot of the DOP ellipsoid objective function for the same fiber realization as in Figs. 3.3 and 3.4. This objective function has a single maximum that is located near the maximum of the eye-opening objective function (and which corresponds to minimizing the DGD). However, unlike the case of the eye-opening and spectral-line objective functions, it does not have a second local maximum corresponding to the case that the input state of polarization is aligned with one of the principal states of the fiber. The shortest axis of the ellipsoid corresponds to the worst possible choice of input polarization state. The worst state is the one for which the power in the signal is evenly split between the two principal states of polarization of the fiber, rather than being aligned with one of them. Therefore the DOP ellipsoid objective function depends on the DGD but not on the input polarization state of the signal.

4. Analysis of the Objective Functions for First-Order PMD. In §3 we saw that for a particular fiber realization the eye opening and spectral line objective functions had two local maxima, whereas the DOP ellipsoid objective function had only one local maximum which corresponded to minimizing the DGD. In this section we show that this behavior is typical by deriving formulae for the DOP ellipsoid and spectral line objective functions in the special case of first-order PMD, *i.e.* in the case that the PMD vector $\vec{\Omega}_T$ of the transmission fiber is frequency independent. In §7

FIG. 3.3. *Eye opening objective function for a typical fiber realization.*FIG. 3.4. *Spectral line objective function for a typical fiber realization. (Same fiber as is shown in Figure 3.3.)*

we will use numerical simulation to quantify the performance of a PMD compensator, and we will use the analysis in this section to help explain those results. In §7 we also provide statistical evidence that for an arbitrary fiber realization, we can regard the objective function as being a perturbation of an objective function for a fiber with only first-order PMD. This statistical result is to be expected since for many fiber realizations $\vec{\Omega}_T$ has only a weak dependence on frequency across the bandwidth of

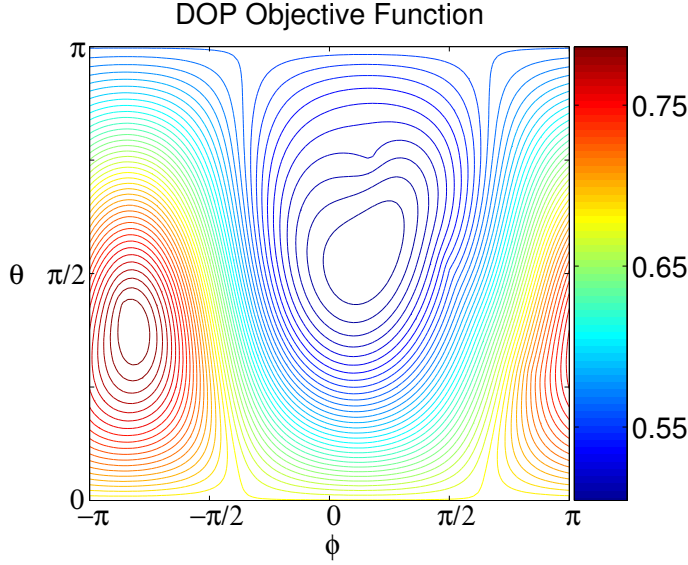


FIG. 3.5. *DOP ellipsoid objective function for a typical fiber realization. (Same fiber as is shown in Figure 3.3.)*

the signal.

As we explained in §3, the domain for the two-dimensional optimization problem we wish to solve is the unit sphere, $S^2 \subset \mathbf{R}^3$. In other words, we want to solve the problem

$$\max_{p \in S^2} J(p), \quad (4.1)$$

where $J : S^2 \rightarrow \mathbf{R}$ is an objective function defined by one of the three feedback mechanisms discussed earlier. In §7, we will in fact solve an unconstrained optimization problem of the form,

$$\max_{(\phi, \theta) \in \mathbf{R}^2} J(\phi, \theta), \quad (4.2)$$

where (ϕ, θ) are spherical coordinates, and where J is now regarded as a doubly periodic function on \mathbf{R}^2 . This unconstrained form of the problem is easier to solve computationally.

For our analysis of the DOP ellipsoid, we assume that the polarization dispersion vector $\vec{\Omega}_R$ of the entire system from transmitter to receiver, including the compensation fiber, is frequency independent. Let τ_T and τ_C be the DGD of the transmission and compensation fibers respectively. We can assume that $\vec{\Omega}_C = \tau_C \vec{e}_X$. If we let Ψ be the angle between $\vec{\Omega}_T$ and $\vec{\Omega}_C$, then in spherical coordinates $\vec{\Omega}_T = \tau_T(\cos \Psi, \cos \beta \sin \Psi, \sin \beta \sin \Psi)^T$, for some angle β . Substituting Eq. (3.2) into Eq. (3.4), we find that the DOP ellipsoid objective function is given by

$$J(\phi, \theta) = 1 - \frac{1}{2} \overline{\Delta \omega^2} [\tau_C^2 + \tau_T^2 + 2\tau_C \tau_T H(\phi, \theta)], \quad (4.3)$$

where

$$H(\phi, \theta) = \cos \Psi \cos \theta + \sin \Psi \sin \theta \sin(\phi - \beta). \quad (4.4)$$

If $\Psi \neq 0, \pi$, then the optimization problem (4.2) has six critical points (ϕ, θ) in $[-\pi, \pi] \times [0, \pi]$. The global maximum of H has value 1 and is located at $(\phi, \theta) = (\beta + \pi/2, \Psi)$, and the global minimum is -1 at $(\phi, \theta) = (\beta - \pi/2, \pi - \Psi)$. There are saddle points at $(\beta, 0)$, $(\beta + \pi, 0)$, (β, π) , and $(\beta + \pi, \pi)$. These saddle points lie on the singularity set $\{\theta = 0\} \cup \{\theta = \pi\}$ of the spherical coordinate mapping from $[-\pi, \pi] \times [0, \pi]$ to S^2 , and are mapped to $(0, 0, \pm 1) \in S^2$. However, they do not correspond to critical points for the optimization problem (4.1) on S^2 , since $(0, 0, \pm 1)$ are not critical points of

$$\max_{(x,y,z) \in \mathbf{R}^3} H(x,y,z) = \cos(\Psi)z + \sin(\Psi)y \quad \text{subject to } x^2 + y^2 + z^2 = 1. \quad (4.5)$$

If $\Psi = 0$ or π , then $H(\phi, \theta) = \pm \cos \theta$ has global optima at $\theta = 0, \pi$ and these critical points are also global optima for the problem (4.1) on the sphere. To summarize, in the case of only first-order PMD, the DOP ellipsoid objective function on S^2 has one maximum and one minimum which is antipodal to the maximum.

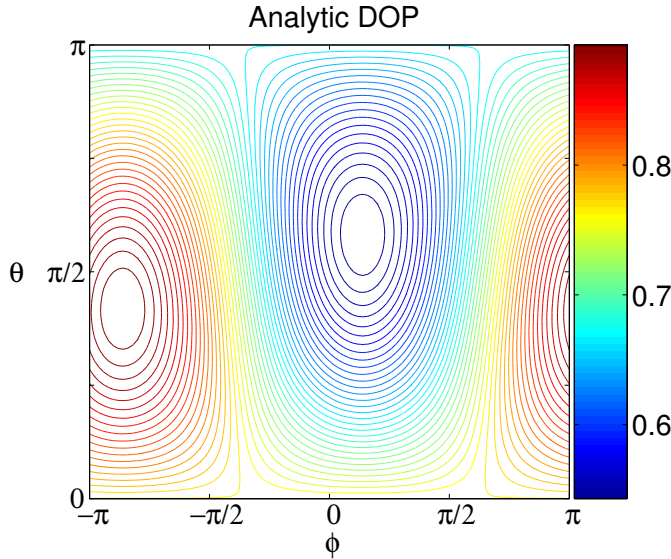


FIG. 4.1. *DOP ellipsoid objective function obtained using Eq. (4.3) for the first-order PMD approximation of the fiber in Fig. 3.5*

Our analysis agrees well with the numerically computed objective function in Fig. 3.5. Given a fiber realization, we can obtain an associated fiber realization with only first-order PMD by computing the polarization dispersion vector, $\vec{\Omega}_T$, at the carrier frequency of the signal. For the fiber realization in Fig. 3.5, the parameters in the formula for $\vec{\Omega}_T$ are $\tau_C = 30$ ps, $\tau_T = 2.68 \tau_C$, $\Psi = 105^\circ$, and $\beta = -65^\circ$. We chose $\sqrt{\Delta\omega^2} = 8.75 \times 10^9$ Hz, so as to fit the result in Fig. 3.5, as was done in [14]. In Fig. 4.1, we show the DOP ellipsoid objective function given by Eq. (4.3) with these parameters. The close agreement between the analytical and numerical results is noteworthy since the PMD of the fiber realization used for Fig. 3.5 depends strongly on frequency. (In fact, the second-order PMD, $|\vec{\Omega}_\omega|$, which is the length of the frequency derivative of the polarization dispersion vector, is 3.1 times its mean value.)

For the analysis of the spectral line objective function, we work in Jones space with matrices in $SU(2)$ acting on \mathbf{C}^2 , rather than in Stokes space. Suppose that the input optical signal at the transmitter is of the form $f(t)\mathbf{u}_0$, where f is a real-valued scalar function, and $\mathbf{u}_0 \in \mathbf{C}^2$ is constant. Let \mathbf{R} be the element of $SU(2)$ that corresponds to the product $\prod_n \mathbf{Q}_n$ of the random rotations used in the coarse-step method to model a realization of the transmission fiber. If we approximate the transmission fiber by a fiber with only first-order PMD, then by Eq. (2.8) the output optical signal v after the PMD compensator is the \mathbf{C}^2 -valued function whose Fourier transform is given by

$$\hat{v}(\omega; \phi, \theta) = \hat{f}(\omega) \mathbf{B}_C(\omega) \mathbf{Q}(\phi, \theta) \mathbf{P} e^{i\omega \mathbf{D}} \mathbf{P}^\dagger \mathbf{R} \mathbf{u}_0, \quad (4.6)$$

where $\mathbf{P} = (\mathbf{v}_1, \mathbf{v}_2) \in SU(2)$ is the matrix of principal states of the transmission fiber and $\mathbf{D} = \text{diag}(-\tau_T/2, \tau_T/2)$. In addition,

$$\mathbf{Q}(\phi, \theta) = \begin{pmatrix} e^{-i\phi/2} \cos \theta/2 & -ie^{i\phi/2} \sin \theta/2 \\ -ie^{-i\phi/2} \sin \theta/2 & e^{i\phi/2} \cos \theta/2 \end{pmatrix} \quad (4.7)$$

is the element of $SU(2)$ that models the action of the polarization controller, and

$$\mathbf{B}_C(\omega) = \begin{pmatrix} e^{-i\omega\tau_C/2} & 0 \\ 0 & e^{i\omega\tau_C/2} \end{pmatrix} \quad (4.8)$$

models the DGD in the compensation fiber. Let $\sigma = \frac{\tau_T + \tau_C}{2}$ and $\eta = \frac{\tau_T - \tau_C}{2}$. Then

$$v(t; \phi, \theta) = \begin{pmatrix} \mathbf{A}_{11}f(t + \sigma) + \mathbf{A}_{12}f(t - \eta) \\ \mathbf{A}_{21}f(t + \eta) + \mathbf{A}_{22}f(t - \sigma) \end{pmatrix}, \quad (4.9)$$

where $\mathbf{A} = \mathbf{Q}(\phi, \theta)(c_1\mathbf{v}_1, c_2\mathbf{v}_2)$. Here $c_k = \mathbf{v}_k^* \mathbf{R} \mathbf{u}_0 \in \mathbf{C}$ is the projection onto the principal state \mathbf{v}_k of the Jones vector of the signal at the central frequency after the transmission fiber.

If we ignore the optical filter and model the electrical filter as a Dirac function centered at frequency ω_{SL} , then the spectral line objective function is given by

$$J(\phi, \theta) = \left| \hat{P}(\omega_{\text{SL}}; \phi, \theta) \right|^2, \quad \text{where } P(t; \phi, \theta) = |v(t; \phi, \omega)|^2 \quad (4.10)$$

is the received optical power.

To calculate an explicit formula for J we first express quadratic functions of \mathbf{A} in terms of the generalized Stokes vectors

$$\mathbf{S}_{jk} = \mathbf{v}_j^\dagger \vec{\sigma} \mathbf{v}_k, \quad j, k \in \{1, 2\}, \quad (4.11)$$

where $\vec{\sigma}$ is the Pauli spin 3-vector defined in §2.3 whose components are elements of $SU(2)$. Notice that \mathbf{S}_{11} and $\mathbf{S}_{22} = -\mathbf{S}_{11} \in \mathbf{R}^3$ are the standard Stokes vectors of \mathbf{v}_1 and \mathbf{v}_2 , and that $\bar{\mathbf{S}}_{21} = -\mathbf{S}_{12} \in \mathbf{C}^3$. If \mathbf{Q}_k denotes the k -th row of $\mathbf{Q}(\phi, \theta)$, then $\mathbf{Q}_k^\dagger \mathbf{Q}_k = \frac{1}{2}(\mathbf{I} + (-1)^{k-1} \mathbf{r}(\phi, \theta) \cdot \vec{\sigma})$ in $SU(2)$, where

$$\mathbf{r}(\phi, \theta) = (\cos \theta, \sin \phi \sin \theta, -\cos \phi \sin \theta) \quad (4.12)$$

parametrizes a sphere. Then,

$$|\mathbf{A}_{22}|^2 = \frac{1}{2}|c_{22}|^2(1 - \mathbf{r} \cdot \mathbf{S}_{22}) \quad \text{and} \quad \Re(\mathbf{A}_{22}\bar{\mathbf{A}}_{21}) = -\frac{1}{2}\mathbf{r} \cdot \Re(c_1\bar{c}_2\mathbf{S}_{21}), \quad (4.13)$$

where \Re denotes the real part, with similar formulae for other quadratic functions of \mathbf{A} . Finally, let

$$g_{a,b}(t) := f(t+a)f(t+b) \quad \text{and} \quad G_{a,b} := \widehat{g}_{a,b}(\omega_{\text{SL}}) \in \mathbf{C}. \quad (4.14)$$

Combining Eqs. (4.6)–(4.14), the objective function for the optimization problem (4.2) on the plane is of the form

$$J(\phi, \theta) = \frac{1}{4} |\alpha + \mathbf{r}(\phi, \theta) \cdot \mathbf{s}|^2, \quad (4.15)$$

where $\alpha \in \mathbf{C}$ and $\mathbf{s} \in \mathbf{C}^3$ are defined by

$$\begin{aligned} \alpha &= |c_1|^2 (G_{\sigma, \sigma} + G_{\eta, \eta}) + |c_2|^2 (G_{-\sigma, -\sigma} + G_{-\eta, -\eta}) \\ \mathbf{s} &= \{ |c_1|^2 (G_{\sigma, \sigma} - G_{\eta, \eta}) + |c_2|^2 (G_{-\sigma, -\sigma} - G_{-\eta, -\eta}) \} \mathbf{S}_{11} \\ &\quad + 2(G_{\sigma, -\eta} - G_{-\sigma, \eta}) \Re(c_1 \bar{c}_2 \mathbf{S}_{21}). \end{aligned} \quad (4.16)$$

In the special case that there is no compensation ($\tau_C = 0, (\phi, \theta) = (0, 0)$) we obtain the well known formula for the electrical power P_{SL} in frequency ω_{SL} as a function of the DGD τ [28]:

$$P_{\text{SL}}(\tau) = [1 - 4\gamma(1 - \gamma) \sin^2(\frac{\omega_{\text{SL}} \tau}{2})] P_{\text{SL}}(0), \quad (4.17)$$

where $\gamma = |c_1|^2$ is the power splitting factor and

$$P_{\text{SL}}(0) = \left| \int |f(t)|^2 e^{i\omega_{\text{SL}} t} dt \right|^2. \quad (4.18)$$

Notice the correlation between the power P_{SL} and the DGD τ : For the 5 GHz spectral line with $\gamma = \frac{1}{2}$, $P_{\text{SL}}(\tau)/P_{\text{SL}}(0)$ decreases from 1 to 0 as τ increases from 0 to 100 ps.

If we reformulate the optimization problem to be of the form (4.1) and make an orthogonal change of coordinates we obtain

$$\max_{\mathbf{x} \in \mathbf{R}^3} J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_0)^T \Lambda (\mathbf{x} - \mathbf{x}_0) \quad \text{subject to} \quad \|\mathbf{x}\| = 1, \quad (4.19)$$

where $\mathbf{x}_0 \in \mathbf{R}^3$ and Λ is a real diagonal matrix. Using the method of Lagrange multipliers, we find that in the non-degenerate case there are no more than six critical points on S^2 , including a global maximum and global minimum. However, it is not possible to obtain explicit formulae for the critical points since they are the zeros of a degree six polynomial.

In Fig. 4.2 we plot the analytical objective function given by formula (4.15). This plot corresponds to an approximation of the fiber when only first-order PMD is present, and it should be compared to the objective function shown in Fig. 3.4. The difference between the two objective functions is due to the large amount of second-order PMD present in the fiber realization used for Fig. 3.4. Note however, that the global maximum is in approximately the same location in both figures. There are four critical points of the objective function on S^2 : Two maxima, one minimum, and a saddle point. As in the case of the DOP ellipsoid, the saddle points on $\{\theta = 0\} \cup \{\theta = \pi\}$ are not critical points of Eq. (4.19). Notice that the eye opening objective function in Fig. 3.3 has five critical points on S^2 .

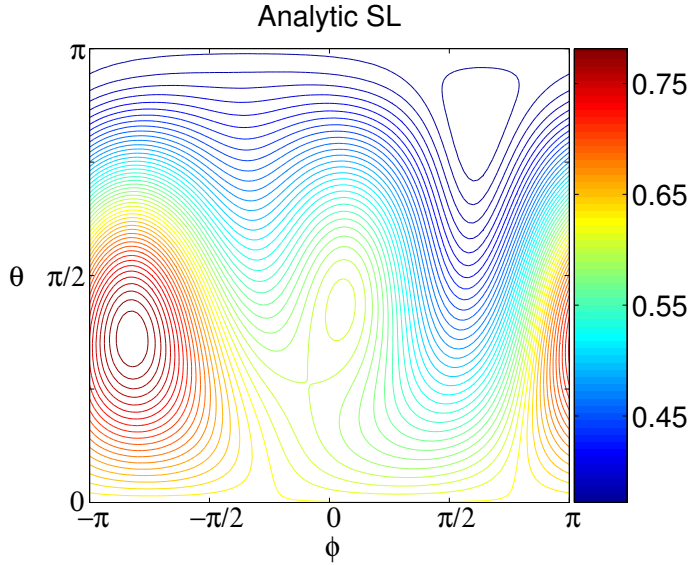


FIG. 4.2. Spectral line objective function obtained using Eq. (4.15) for the first-order PMD approximation of the fiber in Fig. 3.4

5. Importance sampling for PMD. In this subsection, we review the importance sampling algorithm we used to accurately compute outage probabilities due to PMD. Importance sampling increases the computational efficiency of Monte Carlo sampling from the space of fiber realizations.

The problem of evaluating the performance of a PMD compensator is quite challenging because system designers require compensators to maintain a high degree of integrity: the system should lose accuracy for no more than a few minutes per year. It is too time consuming to gather enough samples to accurately measure such low probabilities in a laboratory experiment. It is also not feasible to accurately evaluate the performance of a PMD compensator using numerical simulations based on standard Monte-Carlo sampling.

PMD-induced outages occur when the eye opening is small. In a system without a PMD compensator, small eye openings are correlated to large DGD values, which occur very rarely since large DGD values correspond to sampling from the tail of a Maxwellian distribution. As we discussed in §3, in systems with PMD compensators, the DGD in the transmission line can be at least partially canceled out by the DGD of the compensator at the carrier frequency. Therefore, after the compensator, large DGD values *at that frequency* are exceedingly rare. In general, however, the polarization dispersion vector, and hence the DGD, depends on frequency. It is useful to quantify the strength of this dependence using *second-order PMD* (SOPMD), which is defined to be the length, $|\vec{\Omega}_\omega|$, of the partial derivative of the polarization dispersion vector with respect to frequency. After the signal has traversed the compensation fiber, the DGD may be small at the carrier frequency, but the SOPMD may still be large enough that the eye opening will be small. To summarize, outages only tend to occur in the very rare case that either the DGD or the SOPMD are large relative to their average values. Recently, variance reduction techniques have been developed to greatly increase the computational efficiency of Monte Carlo simulations of

PMD. Variance reduction techniques, which have been successfully applied in many contexts [16], [29], [34], simulate low probability events of interest by concentrating Monte Carlo simulations in those regions of the probability state space that are most likely to give rise to these events. More specifically, in the context of PMD [4], let $\boldsymbol{\theta}$ denote a particular fiber realization in the space Θ of all possible fiber realizations, and let p_θ be the probability density function on Θ . Let $X : \Theta \rightarrow \mathbf{R}^K$ be a random variable on Θ , such as the eye opening or the two-dimensional quantity, $(|\vec{\Omega}|, |\vec{\Omega}_\omega|)$. Let $I : \mathbf{R}^K \rightarrow \{0, 1\}$ be the indicator function for a prescribed range of values $R \subset \mathbf{R}^K$, *i.e.*, $I(x) = 1$ if $x \in R$ and $I(x) = 0$ otherwise. In practice, R could be a bin in the histogram of X . Our goal is compute the probability, P , that $X(\boldsymbol{\theta})$ lies in R ,

$$P = \int_{\Theta} I(X(\boldsymbol{\theta})) p_\theta(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (5.1)$$

Using a standard Monte Carlo simulation, an estimator, \hat{P} , for P is given by

$$\hat{P} = \frac{1}{M} \sum_{m=1}^M I(X(\boldsymbol{\theta}_m)), \quad (5.2)$$

where we have drawn M samples $\boldsymbol{\theta}_m$ according to p_θ . If the events that lie in the region R are very rare, *i.e.*, $P \ll 1$, then the relative variance of the Monte Carlo estimator \hat{P} is $\sigma_{\hat{P}}/\hat{P} \sim (MP)^{-1/2}$. So for example, if $P = 10^{-6}$, as is typically the case for an outage probability, then about $M = 10^8$ samples are required to ensure that the relative variance of \hat{P} is on the order of 10%.

If a variance reduction technique is used, instead of drawing samples according to p_θ , we draw them according to a *biasing distribution*, p_θ^* , chosen so that $p_\theta^*(\boldsymbol{\theta})$ is relatively large when $X(\boldsymbol{\theta}) \in R$. Then the probability P can be expressed in the form

$$P = \int_{\Theta} I(X(\boldsymbol{\theta})) L(\boldsymbol{\theta}) p_\theta^*(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (5.3)$$

where $L = p_\theta/p_\theta^*$ is called the *likelihood ratio*. If we now use a Monte Carlo simulation to draw samples $\boldsymbol{\theta}_m^*$ from Θ according to the biasing distribution p_θ^* , then an estimator \hat{P}^* for P is given by

$$\hat{P}^* = \frac{1}{M} \sum_{m=1}^M I(X(\boldsymbol{\theta}_m^*)) L(\boldsymbol{\theta}_m^*). \quad (5.4)$$

If the biasing distribution p_θ^* is chosen appropriately, many more of the samples $\boldsymbol{\theta}_m$ will fall into the region R and contribute to the sum in (5.4). To ensure that \hat{P}^* is computed correctly, each sample is weighted by its likelihood ratio, which is small where p_θ^* is large relative to p_θ . If the biasing distribution is chosen appropriately, then the relative variance of P can be much smaller than for an unbiased Monte Carlo simulation.

One important question to address is how the biasing distribution should be determined. When using importance sampling, the researcher must use a combination of physical intuition and mathematical analysis to determine appropriate biasing distributions. Recently, two different variance reduction techniques—an importance sampling algorithm and a multicanonical Monte Carlo method—have been developed for simulations of PMD. These two methods take different approaches to solve the

problem of finding an appropriate biasing distribution. The multicanonical Monte Carlo method of Berg and Neuhaus [1], is an iterative method that was adapted for simulations of PMD by Yevick [60], [61], and later by A. Lima [37]. At the n -th iteration of the method, samples are drawn from a biasing distribution $p_{\theta}^{*,n}$, and at the end of each iteration, $p_{\theta}^{*,n}$ is updated in such a way that as n increases there are approximately an equal number of hits in each bin of the histogram of the eye opening for that iteration. Consequently, after sufficiently many iterations, the relative variance between the bins of the eye opening histogram will be small (even for the low-probability, small eye-opening bins).

The second technique (importance sampling) was developed by Biondini and Kath [2]–[4], [17] to generate large values of first- and second-order PMD. Building on the work of I. Lima [38], [39] and A. Lima [35], we used this algorithm to generate the results in this paper. To accurately compute outage probabilities for a PMD compensator, it is necessary to sample a large region in the $(|\vec{\Omega}|, |\vec{\Omega}_{\omega}|)$ -plane. Since this cannot be done efficiently using a single choice of biasing distribution [39], several biasing distributions, p_j^* , are used, each of which targets a different region of the $(|\vec{\Omega}|, |\vec{\Omega}_{\omega}|)$ -plane. To compute the probability P in (5.3), we associate a weight function $w_j : \Theta \rightarrow \mathbf{R}$ to the distribution p_j^* and define a multiple importance sampling Monte Carlo estimator, \hat{P} , for P by

$$\hat{P} = \sum_{j=1}^J \frac{1}{M_j} \sum_{m=1}^{M_j} w_j(\theta_{j,m}) I(X(\theta_{j,m})) L(\theta_{j,m}), \quad (5.5)$$

where M_j samples are drawn using the j -th distribution, and $\theta_{j,m}$ is the m -th such sample. A formula for the relative variance of \hat{P} which was used for the results in this paper is given in [4].

The choices of the biasing distributions and the weights can have a large effect on the relative variance of the estimator \hat{P} in (5.5). For this paper, we used the simple and efficient choice of weights that is given by the *balance heuristic* [56]. With this heuristic,

$$w_j(\theta) = \frac{M_j p_j^*(\theta)}{\sum_{j'=1}^J M_{j'} p_{j'}^*(\theta)}, \quad (5.6)$$

i.e., the weight $w_j(\theta)$ is the probability of realizing the sample θ using p_j^* , relative to the probability of realizing this sample using all J biasing distributions. Therefore, the distribution p_j^* is weighted most heavily in those regions of the sample space Θ where it is largest.

To define a biasing distribution that targets a region of the $(|\vec{\Omega}|, |\vec{\Omega}_{\omega}|)$ -plane, Fogal, Biondini and Kath first determined fiber realizations that maximize a specified linear combination of $|\vec{\Omega}|$ and $|\vec{\Omega}_{\omega}|$. Consider for example the simplest case of maximizing the DGD. From Eq. (2.15) we see that to maximize the DGD, the rotation matrices \mathbf{Q}_n should be chosen so that the vector $\mathbf{Q}_n \vec{\Omega}^{(n-1)}$ is aligned with $\Delta \vec{\Omega}_n$. This set of rotations, $\{\mathbf{Q}_n\}_{n=1}^N$, defines a particular fiber realization. Once this fiber realization has been determined, a biasing distribution is chosen that preferentially selects nearby fiber realizations. In this way, a family of biasing distributions can be chosen, each of which targets a different region of the $(|\vec{\Omega}|, |\vec{\Omega}_{\omega}|)$ -plane. For the results in

this paper, we used the ten biasing distributions described in [39]. This multiple importance sampling algorithm has been successfully used to simulate low-probability regions in the $(|\vec{\Omega}|, |\vec{\Omega}_\omega|)$ -plane. For example, the joint pdf of $(|\vec{\Omega}|, |\vec{\Omega}_\omega|)$ has been calculated down to probability levels of 10^{-8} or less with a relative variance of less than 10% using a total of only 6×10^5 samples [39].

To evaluate the performance of a PMD compensator, we actually need to generate fiber realizations that produce low-probability, small values of the eye opening. Although the importance sampling algorithm is not explicitly designed to generate small eye opening values, A. Lima [36] demonstrated that there is a strong correlation between small eye openings and large values of $(|\vec{\Omega}|, |\vec{\Omega}_\omega|)$. She reached this conclusion in a study of PMD compensators by comparing results obtained using importance sampling with those using the multicanonical Monte Carlo algorithm. The multiple importance sampling algorithm has several advantages over the multicanonical Monte Carlo algorithm, at least for simulations of PMD: the relative variance is easier to calculate; the algorithm is more computationally efficient; and it can be easily parallelized by using different processors to draw samples from the different biasing distributions.

6. Optimization. The original optimization studies we performed were carried out using an object-oriented package that performed local optimization only. The Hilbert Class Library (or HCL) code was developed at Rice University [22], and preliminary results obtained using this optimization software are discussed in [64]. As described in that paper, we used HCL’s limited-memory BFGS (LMBFGS) [40] algorithm with line search [10] to determine an appropriate rotation for the polarization controller. While these early results were intriguing and allowed us to compare the performance of the optimization algorithm with the spectral line and DOP ellipsoid feedback mechanisms, the engineering problem is to find the “best” rotation for the polarization controller, *i.e.*, to find the global optimum. While LMBFGS is a fast (Newton-based) technique that has been “globalized” to accommodate arbitrary initial guesses via the line search feature of HCL, the technique is not guaranteed to find the global optimum. Therefore, after completing the previous study, we decided to incorporate an object-oriented optimization package that contains a variety of optimization tools (including some global techniques) into our optics simulator. The Design Analysis Kit for Optimization and Terascale Applications (DAKOTA) is the optimization package we used to obtain the results presented in this paper. The package was developed at Sandia National Labs to allow users to optimize their (generally pde-based) simulation models for purposes of engineering design, etc.

Our optimization problem is an unconstrained problem of the form (4.2) (although simple bound constraints may be placed on the rotation angles from periodicity), and our goal is to find the global optimum given reasonable computer time limitations. The design variables are continuous, and no analytic gradient information is available.

The experiments described in this paper fall into three categories of optimization jobs – local optimization, multistart jobs which use local optimization repeatedly on the same problem, and global hybrid optimization runs. The local optimization was performed using DAKOTA’s *OPT++* library [11]. The *OPT++* library contains mostly gradient-based nonlinear programming algorithms for constrained or unconstrained optimization. In this paper we chose *OPT++*’s conjugate gradient (CG) method which is appropriate for unconstrained optimization. The optimization in all these experiments is performed over two rotation angles (ϕ and θ). Numerical (finite difference) gradients are used in the CG algorithm with a relative finite difference step

size of 0.0001. The gradient stopping tolerance is also set to the default value of 10^{-4} .

The second set of numerical experiments invokes DAKOTA's multistart capabilities. In each multistart job, a series of local optimization runs are completed, each using a different starting point. The multistart jobs also use *OPT++*'s conjugate gradient routine. Numerical finite difference gradients are used with the same tolerances as in the single starting point local optimization experiments. These multistart jobs were run using either two, four, or nine equally-spaced starting points. Since the ϕ values range over the interval $[-\pi, \pi]$, and θ takes values over the interval from $[0, \pi]$, the nine starting points were chosen to be the equally-spaced points $(\phi, \theta) = \{(-\pi, 0), (-\pi/3, 0), (\pi/3, 0), (-\pi, \pi/3), (-\pi/3, \pi/3), (\pi/3, \pi/3), (-\pi, 2\pi/3), (-\pi/3, 2\pi/3), (\pi/3, 2\pi/3)\}$. Note that periodicity implies these equally-spaced points cover the (ϕ, θ) -domain uniformly. (The edges of the domain wrap.) For the four point runs, the initial guesses are $(\phi, \theta) = \{(-\pi, 0), (-\pi, \pi/2), (0, 0), (0, \pi/2)\}$. Finally, the two-point runs use $\{(-\pi, 0), (0, \pi/2)\}$ as starting guesses.

The third type of optimization jobs are hybrid multilevel optimization schemes which use a global optimization method initially and then switch to a local Newton-based scheme once the optimization is close enough to the solution to ensure fast convergence to the optimum. In our case we chose a genetic algorithm (GA) for the global method. GA's are based on Darwin's theory of survival of the fittest [12]. The GA starts with a random selection of points called a "population". The values of the parameters being optimized over form a string of mathematical "DNA" (a conglomeration of values for the parameters being optimized) which then is adapted to a best fit (or optimal configuration) by a process of natural selection, breeding, and mutation [12]. GA's are convenient when there are multiple local optima or gradients cannot be computed easily. In those cases, GA's can be used to determine regions of the solution space where the global optimum may be located [12]. Global methods such as GA's, however, are slow to complete convergence to a minimizer and are best used in conjunction with a fast local method. Typical GA behavior shows a rapid decrease in the objective function initially, but then a steady slowing of progress towards the minimum. Often only a few initial GA iterations suffice to move the focus of the optimization to the region of interest.

In our GA runs, we chose a population size of five points (representing five (ϕ, θ) pairs) and ran the GA for a maximum of 25 function evaluations. The stopping tolerance for convergence was chosen to be (a loose) 5×10^{-2} . Once the GA run is finished, control is passed to a local method (in our case we chose the Fletcher-Reeves conjugate gradient algorithm from the *CONMIN* package in DAKOTA). The *CONMIN* package contains both constrained and unconstrained minimization algorithms similar to *OPT++*. The same default stopping tolerances were used for the local optimization part of the multilevel jobs as for the purely local *OPT++* jobs.

7. Numerical Results. We now describe a suite of numerical experiments that we ran to investigate the relative performance of the three compensation feedback mechanisms (DOP ellipsoid, spectral line, and eye opening). The experiments used the local and global optimization routines from DAKOTA discussed in §6. Specifically, for each of the three feedback mechanisms we ran 5 optimization jobs involving 200,000 fiber realizations (or Monte Carlo simulations) each. For each feedback mechanism we optimized the compensator using local optimization (conjugate gradients), multilevel global optimization, and a multistart global routine involving varying numbers of starting guesses. The aim in all cases was to compensate for the transmission fiber DGD. As discussed in §6, the multilevel strategy starts with a small number of GA

iterations. After narrowing the optimization search area via the GA, the algorithm passes control to a fast local gradient-based optimization routine (CG in our case) to find the (hopefully) global optima. For the multistart runs, local optimization is “globalized” by running the local optimization to completion from a few different starting points. We ran three sets of multistart jobs for each feedback mechanism. In the first multistart job we used two initial points. In the second we used four starting guesses, and in the last simulation we used nine points. The multistart algorithm is an ad hoc globalization of local methods but suffers from the fact that the starting points are chosen at random. No prior (or current) knowledge of the objective function surface is used.

The goal is to drive the outage probability down to a small acceptable level. System designers typically require a probability of no more than 10^{-3} – 10^{-6} that the eye is 40% closed relative to its original back-to-back value (*i.e.*, 2 dB down from the ideal case of no PMD). Note that an eye corresponding to a 1 dB decrease from the back-to-back signal is 79% open. An eye that is 2 dB down is 63% open, and 3 dB down corresponds to an eye that is only 50% open. Values of 1–2 dB reduction are of greatest interest. A value of 3 dB down corresponds to an eye that is so closed as to indicate system failure.

Parameters in the optical communications simulator which were fixed in these simulations include specifications about the transmitted signal, transmission fiber and receiver. We modeled the transmitted signal by allocating a time interval of duration 100 ps to each binary digit (corresponding to a data rate of 10 Gb/s). We used the 16-bit data pattern 1001101011110000, and we encoded the binary data onto the amplitude of the signal by setting $\mathbf{U}(0, t) = (\chi_1(t), 0)^T$. Here $\chi_1(t) = 1$ when t is in the time interval of a one, and $\chi_1(t) = 0$ when t is in the time interval of a zero. The signal was polarized since $U_2(0, t) = 0$. We then applied a Gaussian filter to \mathbf{U} to produce a smooth signal. The width of the filter was chosen so that the time required for the signal power to increase from 10% to 90% of its maximum value of 1 milliWatt was 30 ps. The transmission fiber was modeled using the coarse-step method (described in §2.4) with 80 fiber segments. The average DGD of the transmission fiber was set at 30 ps, and the DGD of the compensation fiber was fixed at 30 ps. We modeled the receiver using a 60 GHz Gaussian-shaped optical filter, a photodetector that converts the optical power to electrical voltage, and a fifth-order electrical Bessel filter with half-width of 8 GHz.

Our first conclusion from these numerical experiments is that the shape of the objective function depends strongly on the feedback mechanism. The eye opening feedback signal is highly correlated to the bit error ratio which ultimately is the quantity to optimize. Unfortunately the eye feedback mechanism results in a fairly rough objective function. On the other hand, both the spectral line and DOP feedback mechanisms are smoother than the eye and so are easier to optimize. Figures 3.3, 3.4, and 3.5 are plots of the objective functions for the eye, spectral line, and DOP feedback mechanisms for a typical fiber realization. We see that for this example, the eye objective function has a steep ridge. Features such as this can make the progress of local methods towards the maximum difficult. The DOP and spectral line objective functions are considerably smoother.

In Figure 7.1 we show results for the eye opening feedback mechanism. We plot outage probability as a function of the eye opening penalty (in dB down from the back-to-back signal) for six cases: (1) uncompensated signal, (2) compensated case using local optimization, (3) compensated using the global multilevel strategy, and

(4–6) compensated using the global multistart strategy with a varying numbers of initial guesses. We note that for the majority of fiber realizations, the eye diagram is mostly open and so the eye opening penalty is small. In other words, there is a large probability of a small eye opening penalty. In Figure 7.1, the outage probability is the probability that the eye opening penalty exceeds the value on the horizontal axis. The eye opening penalty will exceed 0 dB, whenever the eye is more closed with PMD than without it, which occurs almost all the time. Consequently, in Figure 7.1 the outage probability is approximately 1.0 when the eye opening penalty is 0 dB. It is only for the very rare fiber realizations with large DGD that the eye opening penalty is large, *i.e.*, that the eye is mostly closed. Therefore, there is a very small outage probability that the eye opening penalty exceeds 2 dB. Clearly, if the typical case resulted in a partially or fully closed eye, the state of optical communications would be considerably more dire. The need for importance sampling to bias the Monte Carlo simulations towards the “bad cases” reflects that in most cases, the DGD of the fiber is small. As expected, the local CG method does compensate for the DGD by reducing the outage probability from 10^{-2} to about 5×10^{-4} at an eye opening penalty of 2 dB. However, the best global method in this case (multistart with four points) further decreases the outage probability to about 10^{-5} . We believe that multistart with four points is finding the global optimum for the eye feedback signal since the multistart algorithm with nine points is unable to reduce the outage probability further. Consequently, this result represents the best possible performance for this fixed DGD compensator.

As Table 7.1 indicates, the cost of the multistart jobs is high. Multistart four and nine point schemes require $3\text{--}7 \times$ as many function evaluations as the multilevel technique (200 or 450 iterations respectively versus about 65). In the case of the eye feedback mechanism, multilevel optimization reduces the outage probability to about 4×10^{-5} at 2 dB. In general the number of function evaluations required (for all feedback mechanisms and all levels of DGD) increases from least expensive to most costly in the following order: local conjugate gradient method, the multilevel technique, and finally multistart. The multistart job cost increases approximately linearly with the number of starting points used. Table 7.1 shows the number of function evaluations for all the algorithms and feedback mechanisms at a fixed eye opening penalty (2 dB).

In Figs. 7.2 and 7.3 we show the outage probability versus eye opening penalty for the local and global optimization methods and the DOP and spectral line feedback mechanisms, respectively. These figures include curves for two different conjugate gradient simulations. The first run (CG IC1) used a starting point of $(\phi, \theta) = (0, \frac{\pi}{2})$, and the second (CG IC2) started from $(0, 0)$. For the DOP ellipsoid, the local CG algorithm (CG IC1) took an average of 50 function evaluations and resulted in an outage probability of 7×10^{-5} at 2 dB. The multilevel scheme took 65 function evaluations to arrive at an outage probability of 3×10^{-5} . Multistart takes considerably more function evaluations to arrive at the same outage probability value (see Table 7.1). The lack of model-based knowledge built into the multistart scheme is clearly to blame for this dramatic increase in the number of function evaluations without a corresponding outage probability reduction.

In §3 and §4 we saw that the DOP ellipsoid is very smooth and usually has only one maximum whereas the spectral line can have at least two maxima. Starting from the south pole (IC2), there is a small probability that the CG algorithm will stall near a saddle point since the objective function can be very flat due to distortions inherent in the spherical coordinate mapping. For the DOP ellipsoid, starting from

the equator (IC1) the CG algorithm will usually head straight to the top of the only hill. Since this maximum is between the equator and the saddle points at the south pole, there is very little chance that the algorithm will go via these saddle points and hence very little chance it will get stuck there. Therefore, the outage probability is lower with IC1 than with IC2. Different initial conditions lead to outage probability curves which lie between the curves for IC1 and IC2.

Figure 7.4 compares the best local and best global schemes for the three feedback mechanisms. For each of the pdfs used to determine the curves in this figure, importance sampling produces a relative variation in each bin of the histogram of less than 10%. One can surmise the relative smoothness of the objective functions for the three feedback mechanisms by noting the distance between the local and global curves in the figure. For the eye feedback signal, the local method is least effective at reducing outage probability. Yet the global scheme is more effective for this objective function than it is for either DOP or spectral line. In other words, for the eye feedback mechanism, the outage probability for the best global optimization algorithm is over an order of magnitude smaller than for local optimization. A much smaller decrease occurs for the spectral line feedback signal. (All these comparisons were made at the 2 dB point in the plots.)

The DOP and spectral line feedback signals act as smooth surrogate approximations to the more realistic but bumpier eye signal. The DAKOTA Reference Manual [11] describes surrogate-based optimization as an iterative process that periodically recalibrates an approximate model via data from a true model. The DOP and spectral line functions are not true surrogates as in this work we fix the objective functions used throughout a particular optimization run. No updates to the shape of objective function occur during the course of the optimization. Finally, we note (Table 7.1) that for the spectral line, the multilevel scheme does a better job of reducing the objective function than the two-point multistart scheme but at 50% less cost. Our conclusions are that the simpler smooth DOP and spectral line feedback mechanisms do a good job of approximating the realistic (but bumpy) eye. Moreover, multilevel appears to do the best job of minimizing the number of function evaluations while still reducing the outage probability.

Average Number of Function Evaluations					
	CG	Multilevel	Multistart 2 pts	Multistart 4 pts	Multistart 9 pts
SL	53	64	95	191	454
DOP	52	65	99	199	458
Eye	52	68	99	194	459

TABLE 7.1

Average number of function evaluations per optimization routine and specified feedback mechanism. The average is taken over 200,000 MC simulations.

In Fig. 7.5 we assess the statistical effect that higher-order PMD has on the performance of the spectral line and DOP ellipsoid feedback mechanisms. For each fiber realization we chose the setting of the polarization controller by optimizing the analytical objective functions given by Eqs. (4.3) and (4.15) that we obtained using the first-order PMD approximation of the fiber. Given the solution to this surrogate optimization problem, we then computed the performance of the compensator using the original all-order PMD fiber realization. For the DOP ellipsoid we calculated the global maximum analytically using the formula given above Eq. (4.5), while for

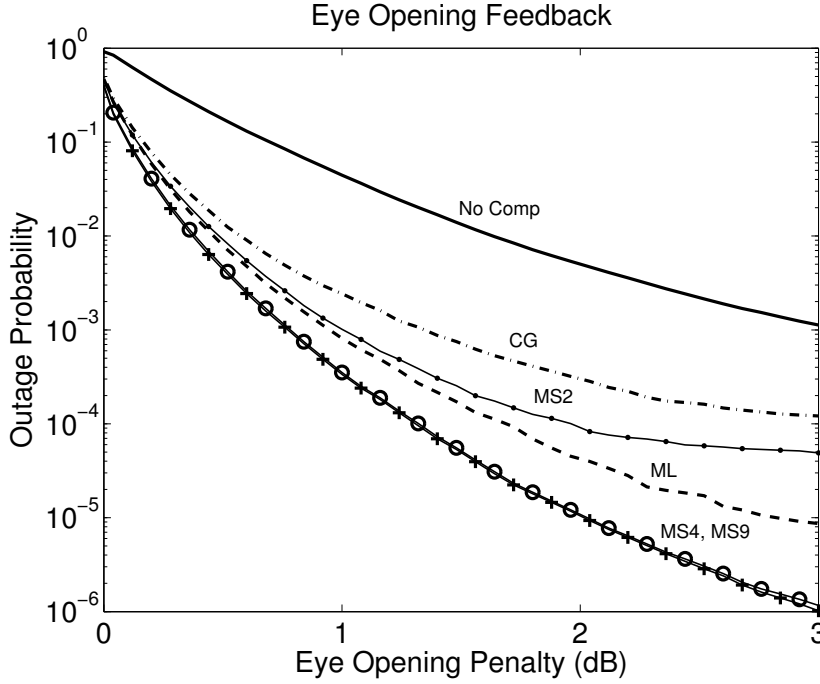


FIG. 7.1. Outage probability as a function of eye opening penalty for different optimization algorithms applied to the eye opening feedback mechanism. The curves shown include the no compensation case (thick solid line), local (conjugate gradient) optimization (dot-dash line), the multilevel hybrid optimization scheme (thick dashed line), and three curves for the multistart method (thin dotted line for multistart with two starting guesses; thin line with circles for multistart with four initial guesses; and thin line with crosses for multistart with nine starting points for the local optimization runs). Note that there is no difference in optimization results for the multistart scheme using four and nine starting points in the case of the eye opening feedback mechanism.

the spectral line we optimized the analytical objective function numerically using the multistart strategy with four starting points. In Fig. 7.5, for the DOP ellipsoid, we compare the “analytical” outage probability curve (thin line with pluses) to the one we obtained numerically using the all-order PMD objective functions (thick dashed line). We also compare the analytic and numerical outage probability curves for the spectral line, which we show with a thin line and a thin line with circles, respectively. For both the DOP ellipsoid and the spectral line, the fairly close agreement between the analytic and numerical curves confirms that we can regard the all-order PMD objective functions as being perturbations of objective functions for fibers with only first-order PMD.

Next, we explain the relative performance of the different methods in Fig. 7.4 by comparing the relative location of the local and global maxima of the different objective functions. This discussion will also quantify the degree to which the spectral line and DOP ellipsoid objective functions act as smooth surrogates of the eye opening. We begin by discussing the three feedback mechanisms optimized via global methods. First, we observe that across the entire range of eye opening penalty values in Fig. 7.4, the outage probability is always larger for the DOP ellipsoid than for the spectral line, and that the eye opening feedback mechanism has the lowest outage probability. The primary reason for the poorer performance with the DOP ellipsoid is that the distance

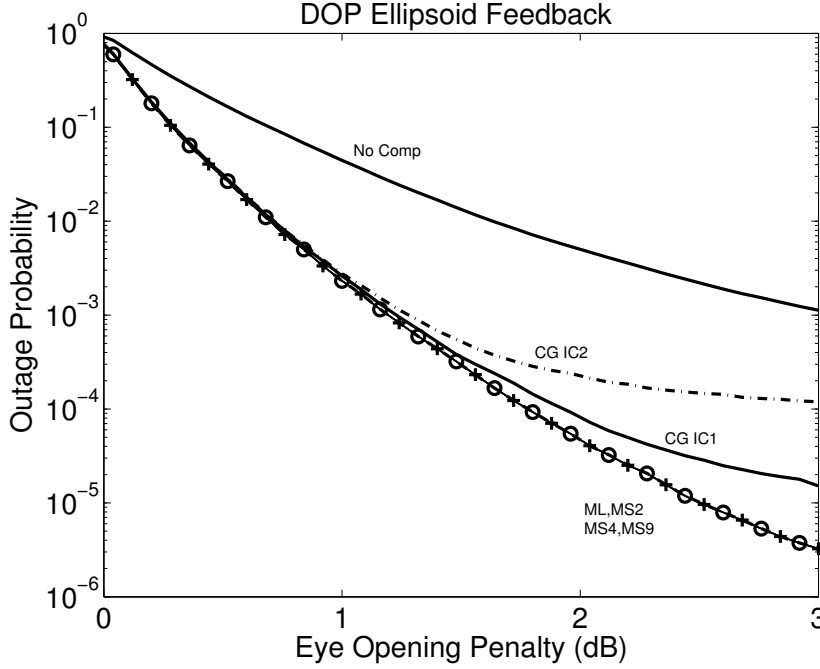


FIG. 7.2. Outage probability as a function of eye opening penalty for different optimization algorithms applied to the DOP feedback mechanism. The curves shown include the no compensation case (thick solid line), local (conjugate gradient) optimization (thick solid line again) starting from the initial guess $(\phi, \theta) = (0, \pi/2)$, and local optimization (CG) (dot-dashed line) starting from the initial guess $(\phi, \theta) = (0, 0)$. The curves for the multilevel hybrid optimization scheme (thick dashed line), multistart method with two starting guesses (thin dotted line), multistart with four initial guesses (thin line with circles), and multistart with nine initial guesses (thin line with crosses) lie on top of each other.

between the global maxima of the DOP ellipsoid and eye objective functions tends to be larger than between the global maxima of the spectral line and eye objective functions.

To verify this observation, we gathered statistics of the distances between the global maxima for the different feedback mechanisms. To define a physically meaningful notion of distance between two rotations of the polarization controller in a compensator, we begin by recalling that any rotation of S^2 can be expressed as a rotation $R_{\vec{r}}(\Psi)$ by an angle $\Psi \in [0, \pi]$ about an axis $\vec{r} \in S^2$. Given two rotations R_1 and R_2 , consider the rotation E_0 such that $R_2 = E_0 R_1$, and let Ψ_0 be the angle such that $E_0 = R_{\vec{r}_0}(\Psi_0)$. In our compensator model the rotations R_1 and $R_X(\pi)R_1$ have the same effect on the signal. Therefore, we also consider the rotation E_π defined by $R_2 = E_\pi R_X(\pi)R_1$ and let Ψ_π be the angle such that $E_\pi = E_0 R_X(\pi) = R_{\vec{r}_\pi}(\Psi_\pi)$. Finally, we define an angle $\Psi \in [0, \pi]$ by $\Psi = \min\{\Psi_0, \Psi_\pi\}$. The angle Ψ is our measure of distance between two rotations R_1 and R_2 performed by the polarization controller. (Note though that the distance function given by Ψ does not satisfy the triangle inequality.) In the case of only first-order PMD, the angle between global maximum and global minimum of the DOP ellipsoid objective function given by Eq. (4.3) is 180° .

In Fig. 7.6 we plot the probability density function (pdf) of the angle between

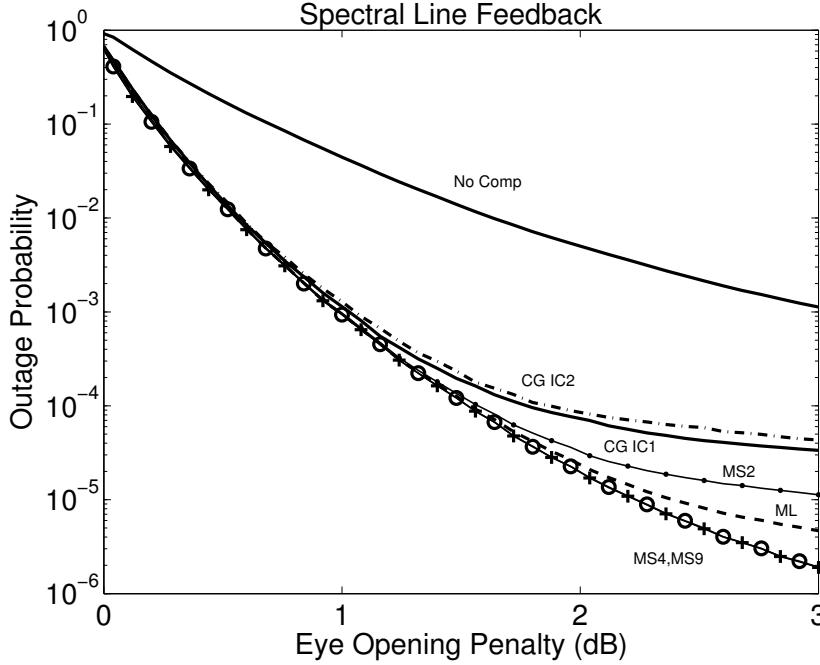


FIG. 7.3. Outage probability as a function of eye opening penalty for different optimization algorithms applied to the spectral line feedback mechanism. The curves shown include the no compensation case (thick solid line), local (conjugate gradient) optimization (thick solid line again) starting from the initial guess $(\phi, \theta) = (0, \pi/2)$, and local optimization (CG) (dot-dashed line) starting from the initial guess $(\phi, \theta) = (0, 0)$. The curve for the multilevel hybrid optimization scheme is denoted by a thick dashed line, and the multistart method with two starting guesses is indicated by a thin dotted line. The multistart method with four initial guesses (thin line with circles), and multistart with nine initial guesses (thin line with crosses) coincide for this feedback mechanism.

the numerically computed global maxima of the spectral line and the eye opening (thick dashed curve) and between the global maxima of the DOP ellipsoid and eye opening (thick solid curve). The most likely angle is 7° for the spectral line and 14° for the DOP ellipsoid. We observe that the probability that the angle is between 30° and 90° degrees is significantly greater for the DOP ellipsoid than for the spectral line. Therefore the global maximum of the spectral line objective function is usually closer to that of the eye opening than is the global maximum of the DOP ellipsoid. Consequently, the eye opening tends to be somewhat larger for the spectral line than for the DOP ellipsoid. This observation explains why the outage probability is smaller for the spectral line than for the DOP ellipsoid when global optimization is used. One of the physical reasons for this performance difference is that the DOP ellipsoid objective function is defined by minimizing the output DOP over all possible input polarization states. However, when we computed the eye opening penalties for the outage probability curves, we chose the input polarization state to be $(S_1, S_2, S_3) = (1, 0, 0)$ rather than choosing it to be the state which resulted in the smallest DOP at the receiver. If we maximize the DOP for the input polarization state with the smallest output DOP, we do not obtain the same global maximum as we would if we maximized the DOP (or spectral line) when the input polarization state is $(1, 0, 0)$.

We also observe in Fig. 7.6 that the probability the angle is between 130° and

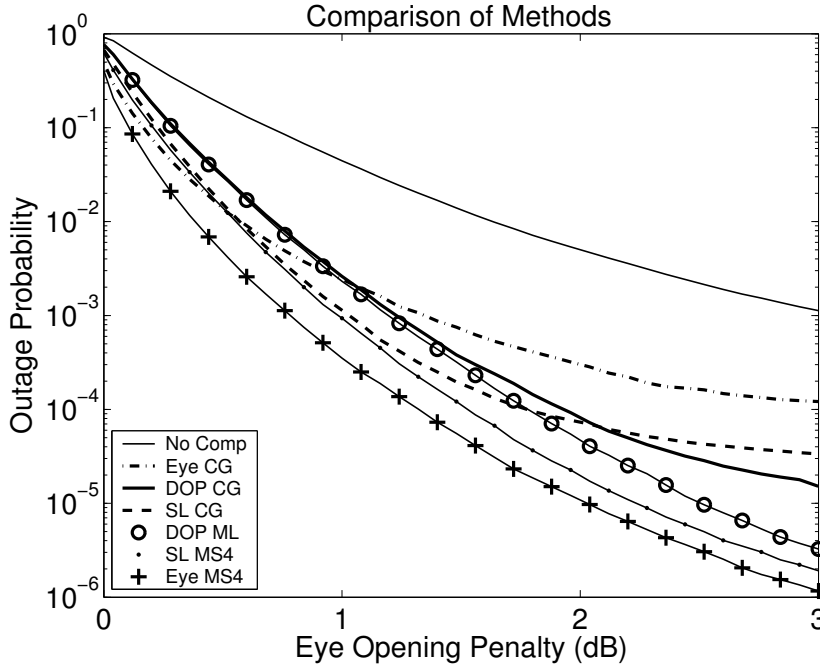


FIG. 7.4. A comparison of the local and best global optimization results for the three feedback mechanisms. For local optimization of the spectral line and DOP ellipsoid we show the results for the starting point IC1 $(\phi, \theta) = (0, \pi/2)$.

170° is much larger for the spectral line than for the DOP ellipsoid. The physical reason for this feature can be explained using the formulae for the objective functions we derived in §4: In the case of only first-order PMD the DOP ellipsoid has only one maximum, whereas the spectral line objective function can have at least two maxima. Therefore, in a small proportion of cases the global maximum of the spectral line can be far from that of the eye opening. For example, suppose as in Figs. 3.3 and 3.4 that the eye opening has local maxima at (ϕ_1, θ_1) and (ϕ_2, θ_2) , and that the spectral line also has local maxima located near these two points. It could happen that the global maximum for the eye opening is at (ϕ_1, θ_1) whereas the global maximum for the spectral line is located near (ϕ_2, θ_2) .

We also found that if we only gather statistics over those fiber realizations for which the second-order PMD is large relative to the DGD, then the most likely angle between the global maxima of the DOP ellipsoid and the eye opening feedback increases markedly. In contrast, the most likely angle between the global maxima of the spectral line and the eye opening is unchanged. (We do not show these results.) These results suggest that when higher-order PMD is introduced, the global maxima of the DOP ellipsoid and the eye opening tend to move apart from each other, whereas the global maximum of the spectral line remains closer to that of the eye opening. This difference in behavior with higher-order PMD provides a second explanation for why the outage probability is smaller for the spectral line than for the DOP ellipsoid when global optimization is used.

To summarize, when global optimization is used, it is important to choose an objective function with the property that the global maximum is close to that of

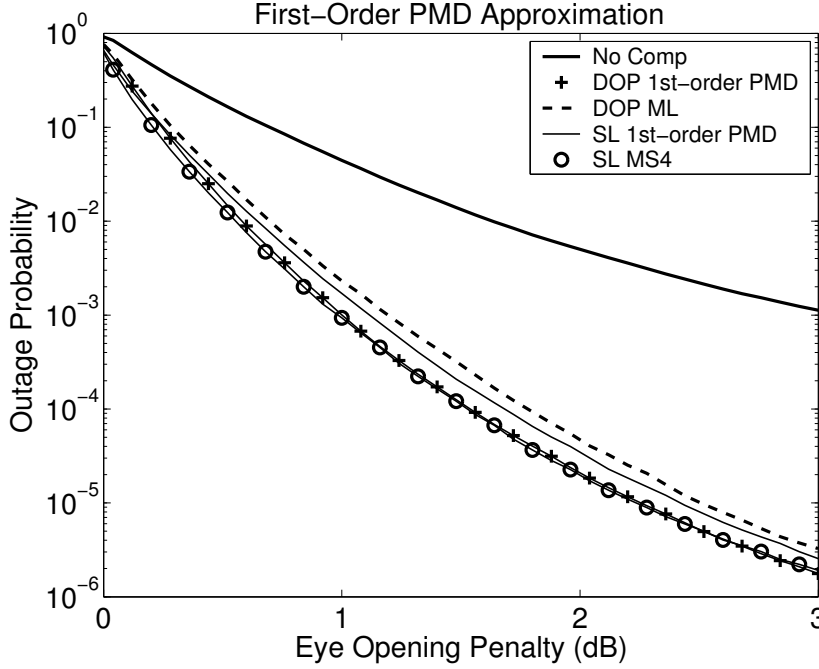


FIG. 7.5. A comparison of outage probability for the first-order PMD approximation of the objective functions coming from the DOP ellipsoid (thin line with pluses) and the spectral line (thin line) versus the all-order PMD objective functions for the DOP ellipsoid (thick dashed line) and the spectral line (thin line with circles). These results were obtained using global optimization. The result without compensation is shown as a thick line.

the eye opening. Therefore, with global optimization, the spectral line is a better surrogate for the eye opening than is the DOP ellipsoid.

Finally, we explain the relative performance with the three feedback signals when local optimization is used. Looking again at Figure 7.4, when the eye opening penalty is larger than 2 dB, the eye opening feedback does not perform as well as the spectral line with local optimization. The performance with the DOP ellipsoid is comparable to that with the spectral line, but as we saw in Fig. 7.2, it can depend significantly on the choice of starting point for the local optimization. To explain the poorer performance with the eye opening feedback, for each feedback mechanism we plot the pdf of the angle, Ψ , between the local maxima reached with conjugate gradients and the global maximum (see Figure 7.7). First, we observe that the DOP ellipsoid and the spectral line have slightly higher probabilities of a very small angle ($\Psi \approx 0$) between the local and global maxima than is the case for the eye opening feedback mechanism. However, there is a significantly higher probability that the angle is between 10° and 50° for the eye opening. This high probability is not seen for the DOP or spectral line objective functions. These observations provide further evidence of the roughness of the eye opening objective function and help to explain why the eye objective function is the worst feedback mechanism for local optimization when the eye opening penalties are large. The other significant feature in Fig. 7.7 is that the probability that the angle exceeds 100° is extremely small for the DOP ellipsoid, but it is relatively large for the other two objective functions. This feature is also present

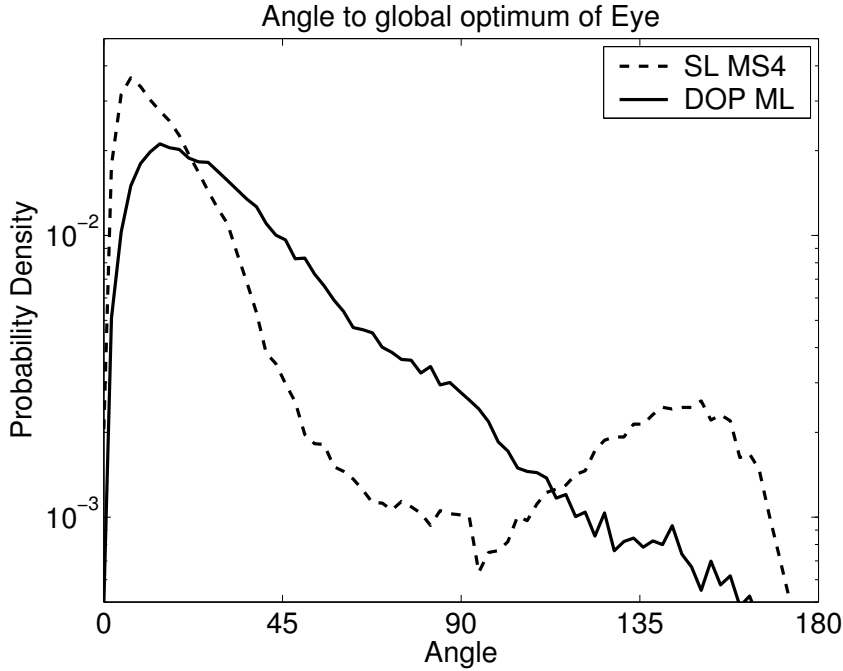


FIG. 7.6. The pdf of the angle between the global maximum of the eye objective function and the maximum obtained for the spectral line using multistart with four starting points (thick dashed curve) and for the DOP ellipsoid using the multilevel algorithm (thick solid curve).

in the pdfs if we only gather statistics over those rare fiber realizations for which the eye opening penalty for the DOP ellipsoid (or spectral line) is larger than 1 dB.

We found that the curves in Fig. 7.7 have approximately the same shape when we gather statistics over fibers with large second-order PMD. Consequently, even with higher-order PMD, the DOP ellipsoid objective function tends to be very smooth and to have only one maximum, just as the analysis in §4 showed for fibers with only first-order PMD. Why then for the DOP ellipsoid is the outage probability at 3 dB smaller with global than with local optimization in Fig. 7.4? One possible reason is that for some fiber realizations we observed that the DOP ellipsoid objective function has suboptimal, wide flat regions. Also, when second-order PMD is large enough, there can be small bumps in the bottom of the valleys which may present a problem for a local algorithm.

8. Conclusions. In an optical fiber communication system, binary data is transmitted through optical fiber using a sequence of pulses of light. The birefringence of the fiber causes the pulses to spread and distort as they propagate and increases the probability that bit errors will occur. A simple optical PMD compensator can be used to reduce these distortions and errors. Since the birefringence varies randomly over time, the compensator must be continually optimized with the aid of a feedback signal. To evaluate the performance of a compensator an optimization problem must be solved for a large number of random realizations of the birefringence. In each case, the goal is to locate the operating point at which the bit-error ratio is smallest. Since it is not possible to measure the bit-error ratio in a real system, we studied three com-

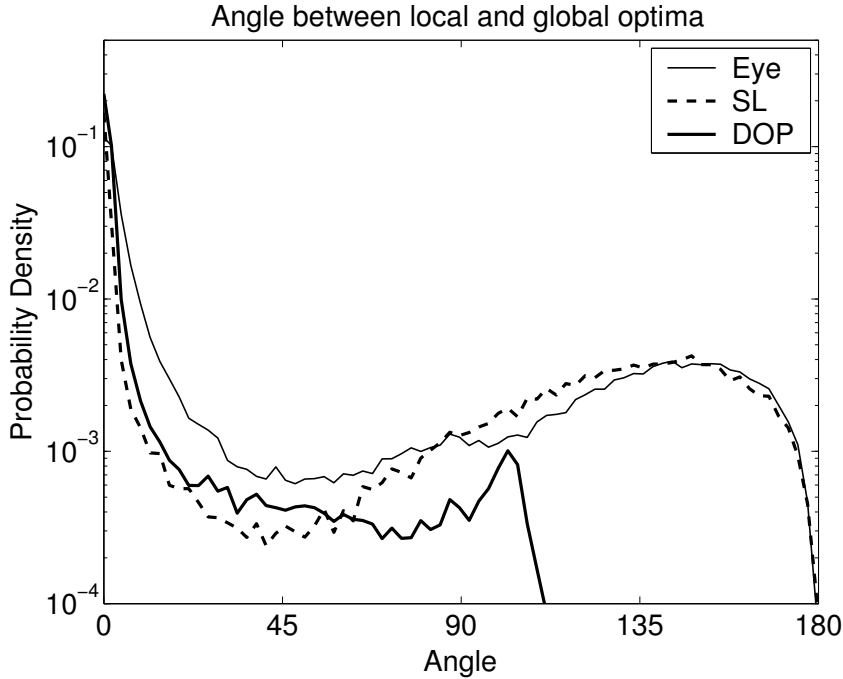


FIG. 7.7. The pdf of the angle between the local and global maxima for the eye opening (thin solid curve), spectral line (dashed curve) and DOP ellipsoid (thick solid curve). The local and global optimization algorithms are the ones shown in Fig. 7.4.

monly used feedback signals: the eye opening, spectral line, and degree of polarization (DOP) ellipsoid. To adequately sample the very rare fiber realizations that result in a large uncompensated bit-error ratio, we performed Monte Carlo simulations with multiple importance sampling. We quantified the degree to which the performance of a compensator depends on the choice of feedback signal and optimization algorithm by computing the probability that the eye opening penalty exceeds a given threshold, *i.e.*, that the bit-error ratio is large.

Although the eye opening is highly correlated to the bit-error ratio, its objective function is quite rough and is therefore hard to optimize. Our results show that the spectral line and DOP objective functions act as smooth surrogate approximations to the rougher eye opening. In the special case of first-order PMD, we proved that the spectral line objective function can have as many as six critical points on the sphere, whereas the DOP ellipsoid has only one maximum and one minimum. We verified that these conclusions also hold statistically over a wide range of fiber realizations with higher order PMD. Since the spectral line objective function is similar to the eye opening, the performance is somewhat better with the spectral line than with the DOP ellipsoid when global optimization is used. However, the DOP ellipsoid objective function is smoother and easier to optimize than the spectral line. In conclusion, since it is most desirable to have a low outage probability for large eye opening penalties, we suggest that multilevel optimization with the DOP ellipsoid feedback gives a good trade-off between the requirements of high performance, computational cost, and complexity of the feedback mechanism.

Acknowledgements. This work was supported by a grant from the NASA Goddard/UMBC Center for Advanced Study of Photonics Research. We thank Paul Leo (YAFO Networks), Curtis Menyuk, Aurenice Lima, Ivan Lima, Brian Marks, (all at UMBC) and Michael Eldred and Tony Giunta (both at Sandia National Labs) for generously sharing their knowledge with us. We also thank the reviewers for helpful comments.

REFERENCES

- [1] B. A. BERG AND T. NEUHAUS, *Multicanonical algorithms for first order phase transitions*, Phys. Lett. B, 267 (1991), pp. 249–253.
- [2] G. BIONDINI AND W. L. KATH, *PMD emulation with Maxwellian length sections and importance sampling*, Photonics Technol. Lett., 16 (2004), pp. 789–791.
- [3] G. BIONDINI, W. L. KATH, AND C. R. MENYUK, *Importance sampling for polarization-mode dispersion*, Photonics Technol. Lett., 14 (2002), pp. 310–312.
- [4] G. BIONDINI, W. L. KATH, AND C. R. MENYUK, *Importance sampling for polarization-mode dispersion: Techniques and applications*, J. Lightwave Technol., 22 (2004), pp. 1201–1215.
- [5] M. BORN AND E. WOLF, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Cambridge University Press, Cambridge, 7th ed., 1999.
- [6] M. BRODSKY, P. MAGILL, AND N. FRIGO, *Polarization-mode dispersion of installed recent vintage fiber as a parametric function of temperature*, Photonics Technol. Lett., 16 (2004), pp. 209–211.
- [7] F. BUCHALI AND H. BÜLOW, *Adaptive PMD compensation by electrical and optical techniques*, J. Lightwave Technol., 22 (2004), pp. 1116–1126.
- [8] F. BUCHALI, S. LANGE, J.-P. THIERY, W. BAUMERT, AND H. BÜLOW, *Fast eye monitor for 10 Gbit/s and its application for optical PMD compensation*, in Proc. OFC'01, Anaheim, CA, 2001, paper TuP5.
- [9] P. C. CHOU, J. M. FINI, AND H. A. HAUS, *Real-time principal state characterization for use in PMD compensators*, Photonics Technol. Lett., 13 (2001), pp. 568–570.
- [10] J. DENNIS AND R. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Non-linear Equations*, Prentice-Hall, Inc., New Jersey, 1983.
- [11] M. ELDRED, A. GIUNTA, B. VAN BLOEMEN WAANDERS, J. WOJTKIEWICZ, S.F., W. HART, AND M. ALLEVA, *DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis. Version 3.0 reference manual*, Technical Report SAND2001-3515, Sandia National Labs, 2002.
- [12] ———, *DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis. Version 3.0 users manual*, Technical Report SAND2001-3796, Sandia National Labs, 2002.
- [13] S. G. EVANGELIDES JR., L. F. MOLLENAUER, J. P. GORDON, AND N. S. BERGANO, *Polarization multiplexing with solitons*, J. Lightwave Technol., 10 (1992), pp. 28–25.
- [14] J. M. FINI, *Coherent Multi-Photon Interference and Compensation of Polarization Dispersion*, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2001.
- [15] J. M. FINI, P. C. CHOU, AND H. A. HAUS, *Estimation of polarization dispersion parameters for compensation with reduced feedback*, in Proc. OFC'01, paper WAA6, Anaheim, CA, 2001.
- [16] G. S. FISHMAN, *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag, New York, NY, 1996.
- [17] S. L. FOGAL, G. BIONDINI, AND W. L. KATH, *Multiple importance sampling for first- and second-order polarization-mode dispersion*, Photonics Technol. Lett., 14 (2002), pp. 1273–1275.
- [18] G. J. FOSCHINI AND C. D. POOLE, *Statistical theory of polarization dispersion in single mode optical fibers*, J. Lightwave Technol., 9 (1991), pp. 1439–1456.
- [19] A. GALTAROSSA, L. PALMIERI, M. SCHIANO, , AND T. TAMBOSSO, *Measurement of birefringence correlation length in long, single-mode fibers*, Opt. Lett., 26 (2001), pp. 962–964.
- [20] A. GALTAROSSA, L. PALMIERI, M. SCHIANO, AND T. TAMBOSSO, *Statistical characterization of fiber random birefringence*, Opt. Lett., 25 (2000), pp. 1322–1324.
- [21] W. GANDER, G. H. GOLUB, AND R. STREBEL, *Least squares fitting of circles and ellipses*, BIT, 34 (1994), pp. 558–578.
- [22] M. S. GOCKENBACH, M. J. PETRO, AND W. W. SYMES, *C++ classes for linking optimiza-*

- tion with complex simulations, *ACM Transactions on Mathematical Software*, 25 (1999), pp. 191–212.
- [23] H. GOLDSTEIN, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1980.
 - [24] J. P. GORDON AND H. KOEGLNIK, *PMD fundamentals: polarization-mode dispersion in optical fibers*, *Proc. Nat. Acad. Sci.*, 97 (2000), p. 4541.
 - [25] A. HASEGAWA AND F. TAPPERT, *Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion*, *Appl. Phys. Lett.*, 23 (1973), pp. 142–144.
 - [26] H. F. HAUNSTEIN, W. SAUER-GREFF, A. DITTRICH, K. STICHT, AND R. URBANSKY, *Principles for electronic equalization of polarization-mode dispersion*, *J. Lightwave Technol.*, 22 (2004), pp. 1169–1182.
 - [27] F. HEISLMANN, D. A. FISHMAN, AND D. L. WILSON, *Automatic compensation of first-order polarization mode dispersion in a 10 Gb/s transmission system*, in *Proc. European Conf. Opt. Comm.*, vol. 1, Madrid, Spain, 1998, pp. 529–530.
 - [28] G. ISHIKAWA AND H. OOI, *Polarization-mode dispersion sensitivity and monitoring in 40 Gb/s OTDM and 10 Gb/s NRZ transmission experiments*, in *Proc. OFC'98*, paper WC5, San Jose, CA, 1998.
 - [29] M. C. JERUCHIM, P. BALABAN, AND K. S. SHAMUGAN, *Simulation of Communication Systems*, Plenum, New York, NY, 1992.
 - [30] I. P. KAMINOW, *Polarization in optical fibers*, *IEEE Journal of Quantum Electronics*, QE-17 (1981), pp. 15–22.
 - [31] I. P. KAMINOW AND T. LI, *Optical Fiber Telecommunications IV-B: Systems and Impairments*, Academic Press, San Diego, CA, 2002.
 - [32] M. KARLSSON, J. BRENTTEL, AND P. A. ANDREKSON, *Long-term measurement of PMD and polarization drift in installed fiber*, *J. Lightwave Technol.*, 18 (2000), pp. 941–951.
 - [33] N. KIKUCHI, *Analysis of signal degree of polarization degradation used as control signal for optical polarization mode dispersion compensation*, *J. Lightwave Technol.*, 19 (2001), pp. 480–486.
 - [34] D. P. LANDAU AND K. BINDER, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, New York, NY, 2000.
 - [35] A. O. LIMA, I. T. LIMA JR., C. R. MENYUK, G. BIONDINI, B. S. MARKS, AND W. L. KATH, *Statistical analysis of the performance of PMD compensators using multiple importance sampling*, *Photonics Technol. Lett.*, 15 (2003), pp. 1716–1718.
 - [36] A. O. LIMA, I. T. LIMA JR., C. R. MENYUK, AND J. ZWECK, *Performance evaluation of single-section and three-section PMD compensators using extended Monte Carlo methods*, in *Optical Fiber Communications Conference*, paper OME27, Anaheim, CA, 2005.
 - [37] A. O. LIMA, I. T. LIMA JR., J. ZWECK, AND C. R. MENYUK, *Efficient computation of PMD-induced penalties using Multicanonical Monte Carlo simulations*, in *European Conference on Optical Communications*, paper We364, Rimini, Italy, 2003, Institute of Electrical Engineers, pp. 538–539.
 - [38] I. T. LIMA, G. BIONDINI, B. S. MARKS, W. L. KATH, AND C. R. MENYUK, *Analysis of PMD compensators with fixed DGD using importance sampling*, *Photonics Technol. Lett.*, 14 (2002), pp. 627–629.
 - [39] 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. Lightwave Technol.*, 22 (2004), pp. 1023–1032.
 - [40] D. LIU AND J. NOCEDAL, *On the limited memory BFGS method for large scale optimization*, *Mathematical Programming*, 45 (1989), pp. 503–528.
 - [41] D. MARCUSE, *Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers*, *J. Lightwave Technol.*, 8 (1990), pp. 1816–1823.
 - [42] D. MARCUSE, C. R. MENYUK, AND P. K. A. WAI, *Application of the Manakov-PMD equation to studies of signal propagation in optica; fibers with randomly varying birefringence*, *J. Lightwave Technol.*, 15 (1997), pp. 1735–1745.
 - [43] C. R. MENYUK, *Application of multiple-length-scale methods to the study of optical fiber transmission*, *J. Eng. Math.*, 36 (1999), pp. 113–136.
 - [44] C. R. MENYUK, B. S. MARKS, I. T. LIMA, J. ZWECK, Y. SUN, AND G. M. CARTER, *Polarization effects in long-haul undersea systems*, in *Undersea Fibre Communication Systems*, J. Chesnoy, ed., Elsevier Press, San Diego, CA, 2002, ch. 7, pp. 270–306.
 - [45] D. A. NOLAN, X. CHEN, AND M. J. LI, *Fibers with low polarization-mode dispersion*, *J. Lightwave Technol.*, 22 (2004), pp. 1066–1077.
 - [46] C. D. POOLE, J. H. WINTERS, AND J. A. NAGEL, *Dynamical equation for polarization dispersion*, *Opt. Lett.*, 16 (1991), pp. 372–4.
 - [47] H. Y. PUA, K. PEDDANARAPPAGARI, B. ZHU, C. ALLEN, K. DEMAREST, AND R. HUI, *An adap-*

- tive first-order polarization-mode dispersion compensation system aided by polarization scrambling: Theory and demonstration*, J. Lightwave Technol., 18 (2000), pp. 832–841.
- [48] S. C. RASHLEIGH, *Origins and control of polarization effects in single-mode fibers*, J. Lightwave Technol., 1 (1983), pp. 312–331.
 - [49] J. L. REBOLA AND A. V. T. CARTAXO, *Q-factor estimation and impact of spontaneous-spontaneous beat noise on the performance of optically preamplified systems with arbitrary optical filtering*, J. Lightwave Technol., 21 (2003), pp. 87–95.
 - [50] H. ROSENFELDT, C. KNOTHE, R. ULRICH, E. BRINKMEYER, U. FEISTE, C. SCHUBERT, J. BERGER, R. LUDWIG, H. G. WEBER, AND A. EHRHARDT, *Automatic PMD compensation at 40 Gb/s and 80 Gb/s using a 3-dimensional DOP evaluation for feedback*, in Proc. OFC'01, paper PD27, Anaheim, CA, 2001.
 - [51] D. SANDEL, M. YOSHIDA-DIEROLF, R. NOÉ, A. SCHÖPFLIN, E. GOTTWALD, AND G. FISCHER, *Automatic polarization mode dispersion compensation in 40 Gb/s optical transmission system*, Electron. Lett., 34 (1998), pp. 2258–2259.
 - [52] H. SUNNERUD, M. KARLSSON, C. XIE, AND P. A. ANDREKSON, *Polarization-mode dispersion in high-speed fiber-optic transmission systems*, J. Lightwave Technol., 20 (2002), pp. 2204–2219.
 - [53] H. SUNNERUD, C. XIE, M. KARLSSON, R. SAMUELSSON, AND P. A. ANDREKSON, *A comparison between different PMD compensation techniques*, J. Lightwave Technol., 20 (2002), pp. 368–378.
 - [54] P. M. SYLLA, C. J. K. RICHARDSON, M. VANLEEUEWEN, M. SAYLORS, AND J. GOLDHAR, *DOP ellipsoids for systems with frequency-dependent principal states*, Photonics Technol. Lett., 13 (2001), pp. 1310–1312.
 - [55] P. R. TRISCHITTA AND E. L. VARMA, *Jitter in Digital Transmission Systems*, Artech House, Boston, MA, 1989.
 - [56] E. VEACH, *Robust Monte Carlo Methods For Light Transport Simulation*, PhD thesis, Stanford University, 1997.
 - [57] D. WADDY, L. CHEN, AND X. BAO, *Theoretical and experimental study of the dynamics of polarization-mode dispersion*, Photonics Technol. Lett., 14 (2002), pp. 468–470.
 - [58] P. K. A. WAI AND C. R. MENYUK, *Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence*, J. Lightwave Technol., 14 (1996), pp. 148–157.
 - [59] P. K. A. WAI, C. R. MENYUK, AND H. H. CHEN, *Stability of solitons in randomly varying birefringent fibers*, Opt. Lett., 16 (1991), pp. 1231–1233.
 - [60] D. YEVICK, *Multicanonical communication system modeling—application to PMD statistics*, Photonics Technol. Lett., 14 (2002), pp. 1512–1514.
 - [61] ———, *The accuracy of multicanonical system models*, Photonics Technol. Lett., 15 (2003), pp. 224–226.
 - [62] J. ZHOU AND M. J. O'MAHONY, *Optical transmission system penalties due to fiber polarization mode dispersion*, Photonics Technol. Lett., 6 (1994), pp. 1265–1267.
 - [63] J. ZWECK, I. T. LIMA, JR., Y. SUN, A. O. LIMA, C. R. MENYUK, AND G. M. CARTER, *Modeling receivers in optical communication systems with polarization effects*, Optics and Photonics News, 14 (2003), pp. 30–35.
 - [64] J. ZWECK, S. E. MINKOFF, A. O. LIMA, I. T. LIMA JR., AND C. R. MENYUK, *A comparative study of feedback controller sensitivity to all orders of PMD for a fixed DGD compensator*, in Proc. OFC'03, paper ThY, Atlanta, GA, 2003.