Interaction of Polarization Mode Dispersion and Nonlinearity in Optical Fiber Transmission Systems

Curtis R. Menyuk, *Fellow, IEEE*, *Fellow, OSA*, and Brian S. Marks, *Member, IEEE*

Invited Tutorial

***Abstract*—Aspects of the interaction of the Kerr nonlinearity and polarization mode dispersion (PMD) are reviewed. The basic equation that governs this interaction on the length scale of interest in optical fiber communications systems is the Manakov-PMD equation. This equation is derived using multiple-length-scale techniques. The focus of the derivation is the elucidation of common misunderstandings and pitfalls rather than mathematical rigor. It is shown that the scalar nonlinear Schrödinger equation is valid when PMD is absent and the signal is initially in a single polarization state. Two examples are then presented that illustrate the complexity of the interaction between nonlinearity and PMD. The first example considers the interaction of a nonlinearly induced chirp with PMD. As the power increases, one can obtain an improved eye opening relative to the case when PMD is absent. The second example considers the effect of nonlinear polarization rotation in a wavelength-division-multiplexed system. When nonlinear polarization rotation is important, the principal states of polarization become time dependent and PMD compensation becomes ineffective. This problem can be mitigated through the use of line codes.**

***Index Terms*—Kerr nonlinearity, Manakov-polarization mode dispersion (PMD) equation, optical fiber transmission, PMD.**

I. Introduction

TRANSMISSION in optical fiber communications systems is impaired and ultimately limited by the four "horsemen" of optical fiber communications systems—chromatic dispersion, amplified spontaneous emission noise from amplifiers, polarization effects, and fiber nonlinearity [1].

Polarization effects are difficult to analyze and study because they are due to random variations in the spatially varying birefringence of the optical fibers, as well as polarization-dependent loss and gain in the amplifiers [2], [3]. The random variation in the fibers changes on a time scale of milliseconds to days, so that it is important to characterize the behavior statistically [2], [3]. Modeling these effects properly is a challenge.

By contrast, the Kerr nonlinearity—the principal source of nonlinearity—is not difficult to model. It leads to a phase rotation that is proportional to the intensity at every point in time. However, it couples to the other transmission impairments in a complex way that is often difficult to analyze. It has been known since the late nineteenth century that even simple nonlinearities can lead to stunningly complex dynamics, including chaos [4], and the Kerr nonlinearity is no exception.

Polarization effects and nonlinearity are usually studied separately for two reasons. First, each is hard to understand on its own. Studying them together can seem almost hopeless. Second, in practice, they usually appear separately in communications systems, with one or the other interacting with the amplified spontaneous emission noise and the chromatic dispersion to limit the transmission distance and/or the per-channel data rate.

However, there are good reasons for studying them together. They can interact to produce effects that are potentially harmful—like nonlinear polarization rotation [5], which can limit the effectiveness of polarization division multiplexing [6]. Conversely, there are a number of cases where they interact to produce effects that are potentially useful, like all-optical switching [7].

Polarization mode dispersion (PMD) has become increasingly important as the per-channel data rates have increased and is now arguably the most important of the polarization effects. It is due to the differential rotation of the polarization states in neighboring frequencies [2], [3], as will be discussed later in more detail. The interaction between PMD and nonlinearity can be particularly complex. Sometimes, in combination with chirp, a system with both PMD and nonlinearity is less impaired than a system with the same amount of nonlinearity and no PMD [8]–[10]. More often, the combination of PMD and nonlinearity can be harmful. Nonlinear polarization rotation can alter the polarization states of the bits, so that they vary from one bit to the next in a way that is difficult to predict. In this case, conventional PMD compensation becomes impossible [11].

Another motivation for studying the interaction of nonlinearity and polarization effects is that this interaction can provide

Manuscript received September 12, 2005; revised March 7, 2006. This work was supported by the Air Force Office of Scientific Research, the U.S. Department of Energy, and the National Science Foundation. This work was also supported by the Laboratory for Physical Sciences, College Park, MD, by the Laboratory for Telecommunications Sciences, Adelphi, MD, and by the Naval Research Laboratory, Washington, DC.

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

B. S. Marks was with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA. He is now with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405 USA (e-mail: bsmarks@indiana.edu).

Digital Object Identifier 10.1109/JLT.2006.875953

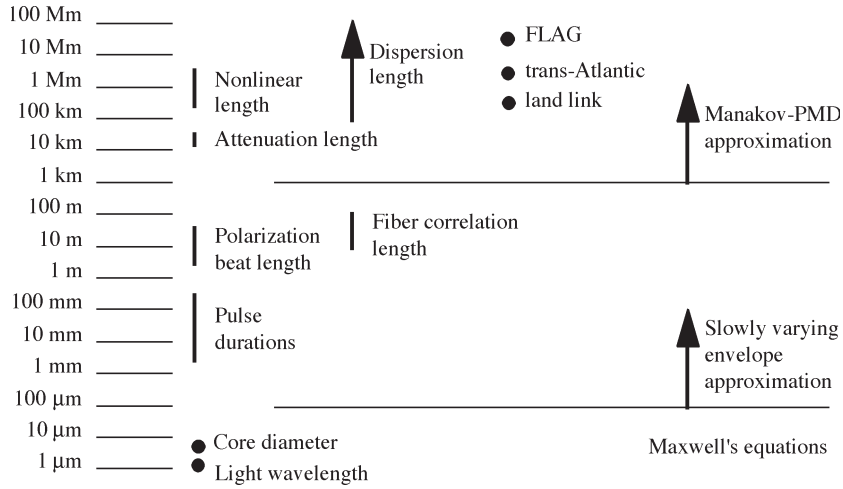

Fig. 1. Key length scales in optical fiber communication systems. (Originally from [26]).

insight into the linear properties of optical fibers. For example, if fibers are linearly birefringent, then the ratio of the cross-phase-modulation coefficient to the self-phase-modulation coefficient is 2/3. By contrast, in a circularly birefringent fiber, this ratio is 2. With arbitrary ellipticity, it varies between the two extremes of 2/3 and 2 [12]. This ratio has been measured to be nearly equal to 2/3 [13], [14], providing powerful evidence—although by no means the only evidence—that standard optical fibers are to good approximation linearly birefringent. Another example is that signals are expected to disperse linearly in exactly the same way, regardless of whether any orientation of the randomly varying axes of birefringence is equally likely, as is the case in standard communications fiber, or there is some fixed axis that is most likely, as is the case in polarization-maintaining fiber (compare [15] and [16]). By contrast, solitons will be stable in the former case, but they will split in two in the latter case (compare [17] and [18]). The stability of solitons in standard fiber thus provides strong evidence that the orientation of the axes of the birefringence is uniformly distributed to a good approximation.

In addition to the practical reasons for studying the interaction between polarization effects and nonlinearity, there are some fundamental reasons as well. First, one of the outstanding questions in mathematical physics is understanding when the complex, chaotic behavior that is generally present in nonlinear systems can be avoided. Mathematical systems in which this behavior is never present are referred to as integrable [4]. An example of an integrable system is the nonlinear Schrödinger equation [19], which can accurately model transmission in optical fibers in certain cases [1], [20]. Integrable partial differential equations like the nonlinear Schrödinger equation typically have soliton solutions. However, solitons—or at least solitonlike solutions—are often also present in nonintegrable systems. Indeed, for deep theoretical reasons, it has been hypothesized that solitons are robust when integrable systems are subjected to Hamiltonian deformations [21]. Without entering into what Hamiltonian deformations are here, we will simply note that the addition of the birefringence to the nonlinear Schrödinger equation produces a new equation—the coupled nonlinear Schrödinger equation—that is a Hamiltonian deformation of the nonlinear Schrödinger equation [22]. The stability of solitons in this system [17], [23] provides evidence that the robustness hypothesis just cited is correct.

Another fundamental reason for studying the interaction of polarization effects and nonlinearity that is of more direct relevance to optical fiber technology is to determine when the nonlinear Schrödinger equation is expected to hold. To understand this issue, the reader should peruse Fig. 1, in which the different length scales that are present in typical optical fiber communications systems are shown. The length scales vary over 13 orders of magnitude from the smallest length, the wavelength of light (1.55 μm), to the largest length, the length of a transglobal communications system (23 000 km) like fiber loop around the globe (FLAG). There is a similar large variation in time scales. The length scales shown in Fig. 1 cluster into three groups. The shortest group, on the order of micrometers, corresponds to the wavelength of light and the fiber core diameter. The intermediate group, on the order of meters, corresponds to the fiber beat length and to the fiber correlation length—the length scale on which the fiber's axes of the birefringence change randomly. The longest group, on the order of tens of kilometers and more, corresponds to length scales for fiber attenuation and amplifier spacing, chromatic dispersion, and the Kerr nonlinearity. A point to note is that the strength of an effect is inversely proportional to its length scale. Therefore, the birefringence is a large effect compared to the chromatic dispersion.

Why then should the nonlinear Schrödinger equation, which completely ignores polarization effects, ever hold? The answer to this question lies in a multiple-scale analysis [24], [25]. While the birefringence is large, it is also rapidly and randomly varying. On a long length scale, its effects can average out. The operative word here is "can." These effects do not always average out, and one must then take into account the vector effects that are associated with the birefringence. A careful analysis shows that it is possible to use scalar equations like the nonlinear Schrödinger equation when the length scale associated with PMD is long compared to the system length and when the initial signal starts out in a single polarization state [26]. Moreover, when the PMD length is long compared to the

nonlinear length, the vector effects can be treated perturbatively and are often small [26].

Thus, any attempt to determine the basic equations that govern light propagation in optical fibers at the lengths of interest in communications systems leads one naturally to contemplate the relationship between nonlinearity and the birefringence. This point bears emphasis because almost all derivations of the nonlinear Schrödinger equation that appear in the literature—including several textbooks—simply ignore the birefringence. An unfortunate consequence is that the unwary reader may attempt to use these equations where they do not apply. Indeed, there are numerous instances in the literature in which insufficient attention to the length scales of interest have led to confusion and error.

Another common source of misunderstanding in the literature is the tendency not to explicitly write the independent variables upon which the dependent variables depend. On occasion, this lack of specificity may be followed by a Fourier transformation or some other transformation in which part of the equation is transformed and part is not. Sometimes, this sort of transformation can be patched up mathematically by using a two-time-scale or two-length-scale analysis in which one of the time or length variables is transformed and the other is not. Sometimes, it is just plain wrong. In any case, the failure to be explicit is a source of confusion.

The remainder of this paper consists of two parts. In Section II, we show how by averaging over the field variations on the scale of a single wavelength, it is possible to obtain the coupled nonlinear Schrödinger equation from Maxwell's equations and how by averaging over the randomly varying birefringence, it is possible to obtain the Manakov-PMD equation and from that the scalar nonlinear Schrödinger equation. Since a detailed mathematical derivation has been published in [26], we will refer the reader there for details and focus here on physical understanding, with an emphasis on common misunderstandings and pitfalls. At every point, we have been explicit about which length scales are being treated as fast and slow and about which independent variables we are using. In Section III, we will review two examples of how nonlinearity interacts with PMD. The first will be the PMD improvement that appears in some nonlinear systems [8], [10]. The second will be the impact of nonlinear polarization rotation on the principal states of polarization [11], [27], [28]. In both cases, the emphasis will be on elucidating the physics and the key length scales. Section IV contains the conclusions.

II. Derivation of the Basic Equations

A. Coupled Nonlinear Schrödinger Equation

Our starting point is Maxwell's equation in a dielectric medium, which may be written

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r},t)] + \frac{1}{\epsilon_0 c^2}\frac{\partial^2 \mathbf{D}(\mathbf{r},t)}{\partial t^2} = 0 \tag{1}$$

where $\mathbf{r}$ and t correspond to space and time. The electric displacement is defined as

$$\mathbf{D}(\mathbf{r},t) = \epsilon_0 \mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t) \tag{2}$$

where the polarizability $\mathbf{P}$ has both a linear and a nonlinear response. One thus writes

$$\mathbf{P}(\mathbf{r},t) = \mathbf{P}_{\mathrm{L}}(\mathbf{r},t) + \mathbf{P}_{\mathrm{NL}}(\mathbf{r},t) \tag{3}$$

where

$$\mathbf{P}_{\mathrm{L}}(\mathbf{r},t) = \epsilon_0 \int_{-\infty}^{t} dt_1 \mathsf{X}_{\mathrm{L}}(\mathbf{r}, t-t_1)\cdot \mathbf{E}(\mathbf{r},t_1) \tag{4a}$$

$$\mathbf{P}_{\mathrm{NL}}(\mathbf{r},t) = \epsilon_0 \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \int_{-\infty}^{t} dt_3 \mathsf{X}_{\mathrm{NL}}(\mathbf{r}, t-t_1, t-t_2, t-t_3) \vdots \mathbf{E}(\mathbf{r},t_1)\mathbf{E}(\mathbf{r},t_2)\mathbf{E}(\mathbf{r},t_3). \tag{4b}$$

The quantities X_{L} and X_{NL} are the material dielectric response tensors. The time integrals end at t, not ∞, to be consistent with causality. Alternatively, one can demand that X_{L} and X_{NL} equal zero when any of their temporal arguments is negative.

We now specialize to the optical fiber geometry by allowing z to correspond to distance along the fiber and $\mathbf{r}_\perp = (x,y)$ to correspond to the transverse dimensions. We write

$$\mathbf{E}(\mathbf{r},t) = \mathbf{F}(z,\mathbf{r}_\perp,t,\omega_0)\exp\left[i\beta(\omega_0)z - i\omega_0 t\right] + \text{complex conjugate} \tag{5}$$

where ω_0 is the radial carrier frequency of the signal, also referred to as the central frequency. Our basic assumption is that the dependence of $\mathbf{F}$ on z and t is slow compared to the dependence of $\mathbf{F}$ on $\mathbf{r}_\perp$ [1], [26]. Referring to Fig. 1, the length scale for the variation of $\mathbf{F}$ due to $\mathbf{r}_\perp$ is the shortest scale or micrometers, while the length scale for the variation with z is centimeters or meters. We also assume that the chromatic dispersion, nonlinearity, and the birefringence do not change on the shortest length scale. Indeed, the assumption that the change in the dispersion is slower than the shortest length scale is implicit in the form of (5). In general we would have to replace $\beta(\omega_0)z$ with $\int_0^z \beta(z_1,\omega_0)dz_1$.

Another issue that we will mention briefly here is our choice of the convention for the carrier frequency in (1). It is common in the engineering literature to replace $\exp[i\beta(\omega_0)z - i\omega_0 t]$ with $\exp[j\omega_0 t - j\beta(\omega_0)z]$, where $j=\sqrt{-1}$ instead of i and $\omega_0 t$ appears with a positive sign, rather than a negative sign. There is no strong reason to prefer one convention over the other, and both are used about equally in the literature on polarization. However, this choice affects the meaning of the Stokes parameters. Therefore, the reader must pay careful attention to which convention is being used in order to avoid confusion. The literature is often not specific in its choice, and one finds on occasion that a single publication switches in mid-stream without informing its readers. The implication of the different conventions is thoroughly discussed in [29] and [30].

The first step in applying multiple length scale methods to the shortest length scale is to neglect the variation that occurs in $\mathbf{F}$ on a longer length scale. Hence, we write

$$\mathbf{E}(\mathbf{r},t) = \mathbf{F}(\mathbf{r}_\perp,\omega_0)\exp\left[i\beta(\omega_0)z - i\omega_0 t\right] + \text{complex conjugate.} \quad (6)$$

In keeping with our assumption that nonlinearity can be neglected on the shortest length scale, we write $\mathbf{P} = \mathbf{P}_{\rm L}$. We write the dielectric tensor as

$$\mathsf{X}_{\rm L}(\mathbf{r},t) = \chi_{\rm L}(\mathbf{r},t)\mathsf{I} + \Delta\mathsf{X}_{\rm L}(\mathbf{r},t) \quad (7)$$

where $\chi_{\rm L}$ is a scalar, I is the identity tensor, and the tensor $\Delta\mathsf{X}_{\rm L}$ includes the effects of the birefringence. At this length scale, we also neglect the variation of $\chi_{\rm L}$ with z and assume that $\chi_{\rm L}$ is cylindrically symmetric, so that writing $\mathbf{r}_\perp = (\rho,\theta)$, we find $\chi_{\rm L}(\mathbf{r},t) = \chi_{\rm L}(\mathbf{r}_\perp,t) = \chi_{\rm L}(\rho,t)$. The tensor $\Delta\mathsf{X}_{\rm L}$ contains both anisotropy due to a preferred orientation in the dielectric tensor at a fixed point and asymmetry due to lack of rotational symmetry about the axis of the fiber. The division of $\mathsf{X}_{\rm L}$ into $\chi_{\rm L}$ and $\Delta\mathsf{X}_{\rm L}$ is ambiguous. To resolve this ambiguity, we choose $\Delta\mathsf{X}_{\rm L}$ so that Trace $(\int_0^{2\pi}\Delta\mathsf{X}_{\rm L}d\theta) = 0$ at all ρ. With this choice, the variable $\chi_{\rm L}$ represents the average or nonbirefringent contribution to $\mathsf{X}_{\rm L}$. Substituting (6) into (1), we obtain the standard eigenvalue problem for optical fibers [31]

$$\begin{aligned}&\beta^2(\omega_0)\mathbf{F}_\perp(\mathbf{r}_\perp,\omega_0) + i\beta(\omega_0)\hat{\mathbf{e}}_z\nabla_\perp\cdot\mathbf{F}_\perp(\mathbf{r}_\perp,\omega_0)\\ &\quad + i\beta(\omega_0)\nabla_\perp F_z(\mathbf{r}_\perp,\omega_0) + \nabla_\perp\times[\nabla_\perp\times\mathbf{F}(\mathbf{r}_\perp,\omega_0)]\\ &\quad - \frac{\omega_0^2}{c^2}\left[1+\tilde{\chi}_{\rm L}(\rho,\omega_0)\right]\mathbf{F}(\mathbf{r}_\perp,\omega_0) = 0\end{aligned} \quad (8)$$

where $\mathbf{F}_\perp = (F_x, F_y)$, and

$$\tilde{\chi}_{\rm L}(\rho,\omega_0) = \int_0^\infty \chi_{\rm L}(\rho,t)\exp(i\omega_0 t)dt. \quad (9)$$

We neglect $\Delta\mathsf{X}_{\rm L}$ in keeping with the assumption that the birefringence can be neglected at this length scale.

The solutions to (8) are thoroughly described in [31] and are reviewed in [26] from the standpoint of multiple scale expansions. In single-mode fibers (SMFs), which are now almost universally used in high-data-rate communications systems, it is a physical fact that $\mathbf{F}$ is doubly degenerate. We can thus write

$$\mathbf{F}(\mathbf{r}_\perp,\omega_0) = \left(\frac{\omega_0}{2\epsilon_0 c^2\beta(\omega_0)}\right)^{1/2} \times\left[u_1\mathbf{R}_1(\mathbf{r}_\perp,\omega_0) + u_2\mathbf{R}_2(\mathbf{r}_\perp,\omega_0)\right] \quad (10)$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are two orthogonal eigenmodes of the SMF. These modes satisfy the relation $\hat{\mathbf{e}}_z\times\mathbf{R}_1 = \mathbf{R}_2$ and are normalized so that

$$\int_0^{2\pi}d\theta\int_0^\infty \rho d\rho|\mathbf{R}_{1\perp}(\mathbf{r}_\perp,\omega_0)|^2 = \int_0^{2\pi}d\theta\int_0^\infty\rho d\rho|\mathbf{R}_{2\perp}(\mathbf{r}_\perp,\omega_0)|^2 = 1 \quad (11)$$

where $\mathbf{R}_{1\perp}$ and $\mathbf{R}_{2\perp}$ are the transverse components of $\mathbf{R}_1$ and $\mathbf{R}_2$. The quantities u_1 and u_2 are constant coefficients, and the factor $[\omega_0/2\epsilon_0 c^2\beta(\omega_0)]^{1/2}$ was added in keeping with the usual convention in which $|u_1|^2 + |u_2|^2$ equals the signal power in the weak guiding approximation.

In elementary derivations of the nonlinear Schrödinger equation, the term $i\beta\nabla_\perp\cdot\mathbf{F}_\perp$ and the term $\nabla_\perp(\nabla_\perp\cdot\mathbf{F})$ that comes from the expansion $\nabla_\perp\times[\nabla_\perp\times\mathbf{F}] = \nabla_\perp(\nabla_\perp\cdot\mathbf{F}) - \nabla_\perp^2\mathbf{F}$ are often neglected. However, as Kodama originally pointed out [32], it is not correct to just drop them. While they are small compared to other terms on the shortest length scale, they are large compared to terms on the longer length scales that contribute to polarization effects, chromatic dispersion, and nonlinearity. The reason that one obtains the right equation after dropping these terms is that their effect is simply to change $\mathbf{R}_1$, $\mathbf{R}_2$, and $\beta(\omega_0)$ slightly, without affecting the structure of the equations that apply on the longer length scales.

A subtle point is that in writing (10), we are focusing on forward-propagating solutions. Maxwell's equations have solutions at $\omega = \omega_0$ that propagate both forward and backward. However, when light is injected into an optical fiber, it rapidly sorts itself out into its forward- and backward-propagating solutions, and only the forward-propagating components propagate through the fiber. Mathematically, the existence of both forward- and backward-propagating solutions is due to the second derivative of z in (1). By contrast, the equations that describe the propagation on the slow length and time scales are only first order in z since only solutions with forward propagation exist on the slow time scale. Equation (10) is where we have made the physical assumption that will lead mathematically to a first-order equation in z.

After determining the fast variation, which in our case means determining $\mathbf{R}_1$, $\mathbf{R}_2$, and $\beta(\omega_0)$, the next step in applying multiple-scale methods is to average over the rapid variation to determine the slow variation. In our particular case, we let the coefficients u_1 and u_2 become functions of z and t so that

$$\mathbf{F}(\mathbf{r},t) = \left(\frac{\omega_0}{2\epsilon_0 c^2\beta(\omega_0)}\right)^{1/2} \times\left[u_1(z,t)\mathbf{R}_1(\mathbf{r}_\perp,\omega_0) + u_2(z,t)\mathbf{R}_2(\mathbf{r}_\perp,\omega_0)\right]. \quad (12)$$

The goal is to determine the equations that govern the evolution of $u_1(z,t)$ and $u_2(z,t)$.

If we did not have additional terms to account for the presence of the birefringence and nonlinearity, then the field $\mathbf{F}_\perp(\mathbf{r},t)$ would consist of a linear superposition of solutions to (8), in which we solve (8) for a range of radial frequency values surrounding ω_0, rather than just ω_0, and we then add these solutions together with amplitudes that are determined by the initial conditions. Writing the Fourier transform of $u_1(z,t)$ as

$$\tilde{u}_1(z,\omega) = \int_{-\infty}^\infty dt\, u_1(z,t)\exp(i\omega t) \quad (13)$$

we find that ω in (13) corresponds physically to the frequency offset from ω_0. The z-evolution of each frequency component must physically be governed by the dispersion relation $\beta(\omega)$.

Consistency with the dispersion relation then demands that $\tilde{u}_1$ obey the equation

$$i\frac{\partial \tilde{u}_1(z,\omega)}{\partial z} + [\beta(\omega_0+\omega) - \beta(\omega_0)]\,\tilde{u}_1(z,\omega) = 0 \qquad (14)$$

and that $\tilde{u}_2$ obey an identical equation. In the time domain, one then finds

$$i\frac{\partial u_1(z,t)}{\partial z} + \text{IFT}\left\{[\beta(\omega_0+\omega) - \beta(\omega_0)]\,\tilde{u}_1(z,\omega)\right\} = 0 \qquad (15)$$

where IFT$(\cdot)$ indicates the inverse Fourier transform. Alternatively, we may write

$$i\frac{\partial u_1(z,t)}{\partial z} + i\beta'(\omega_0)\frac{\partial u_1(z,t)}{\partial t} - \frac{1}{2}\beta''(\omega_0)\frac{\partial^2 u_1(z,t)}{\partial t^2} + \cdots = 0 \qquad (16)$$

where $\beta'(\omega_0)$, $\beta''(\omega_0), \ldots$ indicate the first and higher derivatives of β evaluated at the radial frequency ω_0.

Equation (16) can be obtained directly in the time domain through an iterative procedure described in [26] and [32]. This iterative procedure is based on the assumption that $\beta(\omega_0+\omega)$ can be expanded in a Taylor series around about $\omega = 0$ and that the terms of this series diminish rapidly when integrated over the bandwidth of the signal. It allows the user to work directly in the time domain to include the effects of nonlinearity and the birefringence, but its application to optical fibers keeping the full transverse variation is complex [26]. Here, we will work directly in the plane wave approximation, which was described in [12], and we will summarize the difference when the full transverse variation is taken into account, as in [26]. This approach is far simpler, contains all the essential physics at the intermediate and longest length scales, and gives the right answer, although its coefficients have to be replaced by the true dispersion coefficients and nonlinear coefficient. It might seem surprising at first that a method that approximates the fiber with a medium of infinite transverse extent can yield the correct answer after merely "renormalizing" its coefficients. The validity of this approach follows from two observations. The first is that single-mode optical fibers are weakly guiding, so that there will be two polarizations present, just like in a plane wave. The second is that the multiple-scale technique averages over the transverse variations to yield the evolution along the fiber. The dispersive and nonlinear coefficients are determined by the averaging, but the structure of the equations is not.

In the plane wave approximation, (12) becomes simply

$$\begin{aligned}\mathbf{F}(z,t) &= \left(\frac{\omega_0}{2\epsilon_0 c^2 \beta(\omega_0)}\right)^{1/2} [u_1(z,t)\hat{\mathbf{e}}_1 + u_2(z,t)\hat{\mathbf{e}}_2] \\ &= \left(\frac{\omega_0}{2\epsilon_0 c^2 \beta(\omega_0)}\right)^{1/2} \mathbf{U}(z,t) \end{aligned} \qquad (17)$$

where $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ are two orthogonal unit vectors transverse to the direction of propagation that satisfy $\hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_2$. These unit vectors replace $\mathbf{R}_1$ and $\mathbf{R}_2$. We have also defined $\mathbf{U}(z,t) = [u_1(z,t), u_2(z,t)]^t$. We note that in making this replacement, we have effectively changed the units and physical meaning of u_1 and u_2, because $\mathbf{R}_1$ and $\mathbf{R}_2$ both have units of $1/(\text{area})^{1/2}$, while $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ are unitless. In the plane wave approximation, $|u_1|^2 + |u_2|^2$ equals the signal power per unit area, rather than the signal power. This change is necessary because the signal power, along with the signal's transverse extent, is now infinite. Maxwell's equation (1) becomes

$$\frac{\partial^2 \mathbf{E}(z,t)}{\partial z^2} - \frac{1}{\epsilon_0 c^2}\frac{\partial^2}{\partial t^2}\left[\epsilon_0 \mathbf{E}(z,t) + \mathbf{P}(z,t)\right] = 0. \qquad (18)$$

Combining (4a) and (7), the equation for the linear polarizability becomes

$$\begin{aligned}\mathbf{P}_{\text{L}}(z,t) &= \epsilon_0 \int_{-\infty}^{t} dt_1\,[\chi_{\text{L}}(t-t_1)\mathsf{I} + \Delta\mathsf{X}_{\text{L}}(z,t-t_1)]\cdot\mathbf{E}(z,t_1) \\ &= \epsilon_0 \exp\left[i\beta(\omega_0)z - i\omega_0 t\right] \\ &\quad\times\left\{\int_{-\infty}^{t} dt_1 \chi_{\text{L}}(t-t_1)\mathbf{F}(z,t_1)\exp\left[i\omega_0(t-t_1)\right]\right. \\ &\quad\left. + \int_{-\infty}^{t} dt_1 \Delta\mathsf{X}_{\text{L}}(z,t-t_1)\cdot\mathbf{F}(z,t_1)\exp[i\omega_0(t-t_1)]\right\} \\ &\quad + \text{c.c.}\end{aligned} \qquad (19)$$

where c.c. stands for complex conjugate. Making a Taylor expansion of $\mathbf{F}$ in the neighborhood of $t_1 = t$ inside the integrals of (19), we obtain

$$\begin{aligned}\mathbf{P}_{\text{L}}(z,t) &= \epsilon_0 \exp\left[i\beta(\omega_0)z - i\omega_0 t\right] \\ &\quad\times\left\{\tilde{\chi}_{\text{L}}(\omega_0)\mathbf{F}(z,t) + i\tilde{\chi}'_{\text{L}}(\omega_0)\frac{\partial \mathbf{F}(z,t)}{\partial t}\right. \\ &\quad - \frac{1}{2}\tilde{\chi}''_{\text{L}}(\omega_0)\frac{\partial^2 \mathbf{F}(z,t)}{\partial t^2} + \cdots + \Delta\tilde{\mathsf{X}}_{\text{L}}(\omega_0) \\ &\quad\left.\cdot\,\mathbf{F}(z,t) + i\Delta\tilde{\mathsf{X}}'_{\text{L}}(\omega_0)\cdot\frac{\partial \mathbf{F}(z,t)}{\partial t} + \cdots\right\} \\ &\quad + \text{c.c.}\end{aligned} \qquad (20)$$

As in the case of (13)–(16), the tilde indicates a Fourier transform, and the primes indicate derivatives with respect to frequency.

We will limit the discussion in this article to the terms that we have shown in (20). In practice, it is sometimes necessary to keep nonbirefringent terms to higher order. It has never been necessary to date to keep higher order terms in the birefringence.

We now determine the nonlinear contribution to the polarizability $\mathbf{P}_{\text{NL}}$. We first make the assumption that the nonlinear response is isotropic. Within the plane wave approximation (4b) now becomes

$$\begin{aligned}\mathbf{P}_{\text{NL}}(z,t) &= \epsilon_0 \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \int_{-\infty}^{t} dt_3 \\ &\quad\times \chi_{\text{NL}}(t-t_1, t-t_2, t-t_3) \\ &\quad\times [\mathbf{E}(z,t_1)\cdot\mathbf{E}(z,t_2)]\,\mathbf{E}(z,t_3).\end{aligned} \qquad (21)$$

There is no experimental evidence of a nonisotropic contribution to the cubic nonlinearity, although nonisotropic effects can induce a quadratic nonlinearity in phosphorus-doped fibers [33]. We now make the assumption that the nonlinearity is rapid compared to ω_0^{-1} or the period of the light. In this case, allowing

$$\bar{\chi}_{\rm NL} = \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \int_{-\infty}^{t} dt_3\, \chi_{\rm NL}(t-t_1, t-t_2, t-t_3) \quad (22)$$

we obtain

$$\mathbf{P}_{\rm NL}(z,t) = \epsilon_0 \bar{\chi}_{\rm NL} \left[\mathbf{E}(z,t)\cdot\mathbf{E}(z,t)\right]\mathbf{E}(z,t). \quad (23)$$

It may seem surprising at first that only one material parameter appears, here denoted $\bar{\chi}_{\rm NL}$, rather than the two material parameters introduced by Maker and Terhune [5] and denoted by them as a and b. The reason is that Maker and Terhune do not assume that the nonlinear response is fast compared to the light period; they only assume that it is fast compared to the time variation of the wave envelope [5]. To understand this point, we substitute $\mathbf{E}(z,t) = \mathbf{F}(z,t)\exp[i\beta(\omega_0)z - i\omega_0 t] +$ c.c. into (21) and obtain

$$\begin{aligned}\mathbf{P}_{\rm NL}&(z,t)\\ &= \epsilon_0 \exp[i\beta(\omega_0)z - i\omega_0 t]\\ &\quad\times\{2\tilde{\chi}_{\rm NL}(\omega_0,-\omega_0,\omega_0)\left[\mathbf{F}(z,t)\cdot\mathbf{F}^*(z,t)\right]\mathbf{F}(z,t)\\ &\qquad+\tilde{\chi}_{\rm NL}(\omega_0,\omega_0,-\omega_0)\left[\mathbf{F}(z,t)\cdot\mathbf{F}(z,t)\right]\mathbf{F}^*(z,t)\}\\ &\quad+\text{c.c.}\end{aligned} \quad (24)$$

where

$$\tilde{\chi}_{\rm NL}(\omega_1,\omega_2,\omega_3) = \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \int_{-\infty}^{t} dt_3 \chi_{\rm NL}(t_1,t_2,t_3) \times \exp(i\omega_1 t_1 + i\omega_2 t_2 + i\omega_3 t_3)$$

and we note that $\tilde{\chi}_{\rm NL}(\omega_1,\omega_2,\omega_3) = \tilde{\chi}_{\rm NL}(\omega_2,\omega_1,\omega_3)$. We have kept only the lowest order term in a Taylor expansion of $\mathbf{F}$ and its derivatives. The coefficients $2\tilde{\chi}_{\rm NL}(\omega_0,-\omega_0,\omega_0)$ and $\tilde{\chi}_{\rm NL}(\omega_0,\omega_0,-\omega_0)$ are proportional to the Maker and Terhune a and b coefficients, respectively. When the response is faster than the period of the light, then

$$\begin{aligned}\tilde{\chi}_{\rm NL}(\omega_0,-\omega_0,\omega_0) &= \tilde{\chi}_{\rm NL}(\omega_0,\omega_0,-\omega_0)\\ &= \tilde{\chi}_{\rm NL}(0,0,0)\\ &= \bar{\chi}_{\rm NL}\end{aligned} \quad (25)$$

and (24) reduces to

$$\begin{aligned}\mathbf{P}_{\rm NL}(z,t) =& \epsilon_0\bar{\chi}_{\rm NL}\exp[i\beta(\omega_0)z - i\omega_0 t]\\ &\times\{2\left[\mathbf{F}(z,t)\cdot\mathbf{F}^*(z,t)\right]\mathbf{F}(z,t)\\ &+\left[\mathbf{F}(z,t)\cdot\mathbf{F}(z,t)\right]\mathbf{F}^*(z,t)\}+\text{c.c.}\end{aligned} \quad (26)$$

which is equivalent to (23) in the limit that we are considering.

One may question the physical assumption upon which (26) is based. Indeed, both the Brillouin and Raman interactions have times scales that are far longer than the period of the light waves and so must be considered separately [1]. Moreover, even in the case of the Kerr effect, while most of the contribution (82%) comes from the electronic response, which is fast compared to the light period, the remainder comes from the Raman response, which is not [34]. Additionally, acoustic interactions can make an important contribution [35]. These considerations limit the validity of (26) and are discussed in detail in [26]. From a practical standpoint, they do not appear to affect the applicability of (26) to communications systems.

We may now substitute (20) and (26) into Maxwell's equation (18), and use the definition $\mathbf{F} = [\omega_0/2\epsilon_0 c^2\beta(\omega_0)]^{1/2}\mathbf{U}$, to obtain

$$\begin{aligned}&-\beta^2(\omega_0)\mathbf{U}(z,t) + 2i\beta(\omega_0)\frac{\partial\mathbf{U}(z,t)}{\partial z} + \frac{\partial^2\mathbf{U}(z,t)}{\partial z^2}\\ &\quad+\frac{\omega_0^2}{c^2}\tilde{\epsilon}(\omega_0)\mathbf{U}(z,t) + i\left[\frac{\omega_0^2}{c^2}\tilde{\epsilon}'(\omega_0) + 2\frac{\omega_0}{c^2}\tilde{\epsilon}(\omega_0)\right]\frac{\partial\mathbf{U}(z,t)}{\partial t}\\ &\quad-\left[\frac{\omega_0^2}{2c^2}\tilde{\epsilon}''(\omega_0) + 2\frac{\omega_0}{c^2}\tilde{\epsilon}'(\omega_0) + \frac{1}{c^2}\tilde{\epsilon}(\omega_0)\right]\frac{\partial^2\mathbf{U}(z,t)}{\partial t^2}\\ &\quad+\frac{\omega_0^2}{c^2}\epsilon_0\Delta\tilde{\chi}_{\rm L}(\omega_0)\cdot\mathbf{U}(z,t)\\ &\quad+i\left[\frac{\omega_0^2}{c^2}\epsilon_0\Delta\tilde{\chi}_{\rm L}'(\omega_0) + \frac{2\omega_0}{c^2}\epsilon_0\Delta\tilde{\chi}_{\rm L}(\omega_0)\right]\cdot\frac{\partial\mathbf{U}(z,t)}{\partial t}\\ &\quad+\frac{\omega_0^3}{2c^4\beta(\omega_0)}\bar{\chi}_{\rm NL}\{[2\mathbf{U}(z,t)\cdot\mathbf{U}^*(z,t)]\,\mathbf{U}(z,t)\\ &\qquad+[\mathbf{U}(z,t)\cdot\mathbf{U}(z,t)]\,\mathbf{U}^*(z,t)\} = 0\end{aligned} \quad (27)$$

where

$$\tilde{\epsilon}(\omega_0) = \epsilon_0\left[1+\tilde{\chi}_{\rm L}(\omega_0)\right]. \quad (28)$$

We now proceed to determine the evolution equation, making use of the inequalities

$$\begin{aligned}\left|\beta^2(\omega_0)\mathbf{U}(z,t)\right| &\gg \left|\beta(\omega_0)\frac{\partial\mathbf{U}(z,t)}{\partial z}\right| \gg \left|\frac{\partial^2\mathbf{U}(z,t)}{\partial z^2}\right|\\ \left|\omega_0^2\mathbf{U}(z,t)\right| &\gg \left|\omega_0\frac{\partial\mathbf{U}(z,t)}{\partial t}\right| \gg \left|\frac{\partial^2\mathbf{U}(z,t)}{\partial t^2}\right|.\end{aligned} \quad (29)$$

We note that it is not necessary for the inequalities in (29) to hold at every value of z and t. It is only necessary for them to hold at "most" values, so that one may apply perturbation theory and obtain reasonably accurate results. What is meant by "most" in this context can be made rigorous by applying the concepts of perturbation theory [24], [25]. In keeping with the physics of Fig. 1, we assume that the nonlinearity appears at the same order of the perturbation expansion as the dispersion. In modern-day dispersion-managed communications systems, the nonlinear scale length is hundreds to thousands of kilometers and is often far larger than the scale length of the local dispersion, indicating that the local dispersion is a far larger effect than the local nonlinearity. At the same time, the local

dispersion can be far larger than the averaged dispersion, which is typically comparable in magnitude to the nonlinearity. Small or large, it is important to keep the nonlinear contribution in the equations. Physically, the nonlinearity leads to qualitatively different behavior, which is reflected mathematically by a change in the structure of the equations. Unlike some quantitatively larger linear contributions, the nonlinearity cannot be absorbed into a redefinition of the coefficients of the linear equation.

At lowest order in the perturbation expansion, we obtain the dispersion relation

$$\beta^2(\omega_0) - \frac{\omega_0^2}{c^2}\tilde{\epsilon}(\omega_0) = 0. \tag{30}$$

At the next order, we obtain the group velocity equation

$$2i\beta(\omega_0)\frac{\partial \mathbf{U}(z,t)}{\partial z} + i\left[\frac{\omega_0^2}{c^2}\tilde{\epsilon}(\omega_0)\right]'\frac{\partial \mathbf{U}(z,t)}{\partial t} + \frac{\omega_0^2}{c^2}\epsilon_0 \Delta\tilde{\mathsf{X}}_{\mathrm{L}}(\omega_0)\cdot\mathbf{U}(z,t) = 0 \tag{31}$$

where we recall that the prime indicates the derivative with respect to angular frequency, evaluated at ω_0. Equation (31) can be written in the form

$$i\frac{\partial \mathbf{U}(z,t)}{\partial z} + i\beta'(\omega_0)\frac{\partial \mathbf{U}(z,t)}{\partial t} + \Delta\mathsf{B}(\omega_0)\cdot\mathbf{U}(z,t) = 0 \tag{32}$$

where $\Delta\mathsf{B}(\omega_0) = \epsilon_0[\omega_0^2/2c^2\beta(\omega_0)]\Delta\tilde{\mathsf{X}}_{\mathrm{L}}(\omega_0)$. By including the term $\Delta\mathsf{B}(\omega_0)\cdot\mathbf{U}(z,t)$ at this order, instead of with the dispersion relation (30), we are effectively assuming that the components of $\Delta\mathsf{B}(\omega_0)$ are all much less than $\beta(\omega_0)$. This assumption is physically reasonable since the birefringence in communications fibers is approximately 10^{-6} or less. To proceed to the next and final order, we must eliminate $\partial^2\mathbf{U}/\partial z^2$ consistently with (32). To do so, we take the z-derivative of (32) and then use (32) again to eliminate the first-order derivatives in z. We thus obtain

$$\frac{\partial^2 \mathbf{U}(z,t)}{\partial z^2} = [\beta'(\omega_0)]^2\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} - 2i\beta'(\omega_0)\Delta\mathsf{B}(\omega_0)\cdot\frac{\partial \mathbf{U}(z,t)}{\partial t} - [\Delta\mathsf{B}(\omega_0)]^2\cdot\mathbf{U}(z,t). \tag{33}$$

We then find that

$$i\frac{\partial \mathbf{U}(z,t)}{\partial z} + \Delta\mathsf{B}_1(\omega_0)\cdot\mathbf{U}(z,t) + i\left[\beta'(\omega_0)\mathsf{I} + \Delta\mathsf{B}'(\omega_0)\right]\cdot\frac{\partial \mathbf{U}(z,t)}{\partial t} - \frac{1}{2}\beta''(\omega_0)\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} + \gamma\left\{|\mathbf{U}(z,t)|^2\mathbf{U}(z,t) - \frac{1}{3}\left[\mathbf{U}^\dagger(z,t)\sigma_2\mathbf{U}(z,t)\right]\sigma_2\mathbf{U}(z,t)\right\} = 0 \tag{34}$$

where $\Delta\mathsf{B}_1(\omega_0) = \Delta\mathsf{B} - (\Delta\mathsf{B})^2/2\beta$, $\mathbf{U}^\dagger$ is the row vector (u_1^*, u_2^*), and $\gamma = [3\omega_0^3/4c^4\beta^2(\omega_0)]\bar{\chi}_{\mathrm{NL}}$ is the Kerr coefficient. The matrix

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

is the second Pauli matrix [26]. The term $\Delta\mathsf{B}'\cdot\partial\mathbf{U}/\partial t$ may be written as $\Delta\mathsf{B}_1'\cdot\partial\mathbf{U}/\partial t + (\Delta\mathsf{B}/\beta)\cdot\Delta\mathsf{B}'\cdot\partial\mathbf{U}/\partial t$. The second term must be dropped in order to be consistent with the ordering assumptions that we made when writing (30) and (31). Thus, we have $\Delta\mathsf{B}(\omega_0)\cdot\partial\mathbf{U}/\partial t = \Delta\mathsf{B}_1(\omega_0)\cdot\partial\mathbf{U}/\partial t$. At this point, we rewrite $\Delta\mathsf{B}_1(\omega_0)$ as $\Delta\mathsf{B}(\omega_0)$. Effectively, the small correction to $\Delta\mathsf{B}(\omega_0)$ in $\Delta\mathsf{B}_1(\omega_0)$ has been absorbed into a redefinition of $\Delta\mathsf{B}(\omega_0)$ [12], [26]. There is a subtlety here that bears brief mention. After this redefinition, $\Delta\mathsf{B}(\omega_0)$ is strictly speaking no longer traceless. However, the small trace in $\Delta\mathsf{B}(\omega_0)$ can be absorbed into a redefinition of $\beta'(\omega_0)$, so that $\Delta\mathsf{B}(\omega_0)$ once again becomes traceless. While the algebra involved in these redefinitions is somewhat intricate, the physical meaning is simple. The higher order corrections to $\mathsf{B}(\omega_0)$ do not change the qualitative behavior of the system. Hence, mathematically, they can be absorbed into a redefinition of the coefficients of the linear equation.

When deriving (34), it is important to take (33) into account. Otherwise, one does not obtain $\beta''(\omega_0)$ self-consistently as the coefficient of $\partial^2\mathbf{U}/\partial t^2$. To see this point explicitly, we note that in the absence of the birefringence and nonlinearity, the second-order equation becomes explicitly

$$2i\beta(\omega_0)\frac{\partial \mathbf{U}(z,t)}{\partial z} + i\left[\frac{\omega_0^2}{c^2}\tilde{\epsilon}(\omega_0)\right]'\frac{\partial \mathbf{U}(z,t)}{\partial t} - \left[\frac{\omega_0^2}{2c^2}\tilde{\epsilon}''(\omega_0) + \frac{2\omega_0}{c^2}\tilde{\epsilon}'(\omega_0) + \frac{1}{c^2}\tilde{\epsilon}(\omega_0)\right]\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} + [\beta'(\omega_0)]^2\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} = 0. \tag{35}$$

From the expression $\beta(\omega_0) = (\omega_0/c)[\tilde{\epsilon}(\omega_0)]^{1/2}$, one can verify that

$$\beta(\omega_0)\beta''(\omega_0) = \frac{\omega_0^2}{2c^2}\tilde{\epsilon}''(\omega_0) + \frac{2\omega_0}{c^2}\tilde{\epsilon}'(\omega_0) + \frac{1}{c^2}\tilde{\epsilon}(\omega_0) - [\beta'(\omega_0)]^2 \tag{36}$$

which is the consistency requirement. It is stated in some derivations of the nonlinear Schrödinger equation that may be found in articles and even well-known textbooks that one simply "drops" the second derivative in z. In fact, one must self-consistently replace it as shown here in order to obtain the correct answer. This point bears emphasis because it has been a significant point of confusion for many years.

If we explicitly take into account the transverse variations in optical fibers, then (34) still has the same form [26]. There are two changes. The first is that the expressions for $\beta(\omega_0)$ and $\Delta\mathsf{B}(\omega_0)$ will include geometric contributions as well as material contributions. Since, in practice, one just uses the measured values for $\beta(\omega_0)$ and $\Delta\mathsf{B}(\omega_0)$, this difference has no

practical consequences. The second is that γ becomes equal to $[3\omega_0^3/4c^4\beta^2(\omega_0)A_{\rm eff}]\bar{\chi}_{\rm NL}$, where

$$\frac{1}{A_{\rm eff}} = \int_0^{2\pi} d\theta \int_0^{\infty} \rho d\rho \left[\mathbf{R}_j(\rho,\theta)\cdot\mathbf{R}_j^*(\rho,\theta)\right]^2 \qquad (37)$$

with $j = 1, 2$. The additional factor of $A_{\rm eff}^{-1}$ compensates for the difference in the units of u_1 and u_2 in the plane wave approximation. In practice, one either measures γ directly from the signal's self-phase modulation or estimates $A_{\rm eff}$ from the core area. There are other small changes in the nonlinear terms that are due to geometric factors and are negligible in practice [26]. An important point to note is that in contrast to $\mathbf{F}(\mathbf{r},t)$ and the other field vectors, the vector $\mathbf{U}(z,t)$ remains only a function of z and t, as can be seen from the definition in (12).

We now consider the properties of $\Delta\mathsf{B}(\omega_0)$ and the impact on the coefficients for self-phase and cross-phase modulation. For the moment, we will neglect variations that occur on the intermediate length scale in Fig. 1 and treat $\Delta\mathsf{B}(\omega_0)$ as constant in z. Since $\Delta\mathsf{B}(\omega_0)$ is traceless, we can write it as

$$\Delta\mathsf{B}(\omega_0) = \Delta\beta[\cos\phi\cos\theta\sigma_3 + \cos\phi\sin\theta\sigma_1 + \sin\phi\sigma_2] \qquad (38)$$

where $\Delta\beta(\omega_0)$ gives the magnitude of the birefringence and $\phi(\omega_0)$ and $\theta(\omega_0)$ are angles that determine the ellipticity of the fiber's polarization eigenstate and the orientation of the axes of the birefringence, respectively. All three quantities are functions of ω_0 in principle, although we have not written this explicitly in (38). The matrices σ_j are the standard Pauli matrices. The coefficients of σ_3, σ_1, and $-\sigma_2$ correspond, respectively, to the first, second, and third components of the Stokes vector. The mismatch between the standard definition of the Pauli matrices and the components of the Stokes vector can be inconvenient and led Gordon and Kogelnik [29] to redefine the Pauli matrices. Since ϕ and θ are independent of z, we can pick the axis of the birefringence so that $\theta = 0$. We will also assume that θ and ϕ are independent of ω_0. This physically reasonable assumption effectively states that that the orientation of a fiber's birefringence and the ellipticity of its eigenstates are independent of frequency. There is no experimental evidence to contradict this assumption, and we will use it throughout this paper. Equation (34) now becomes

$$\begin{aligned} i\frac{\partial\mathbf{U}(z,t)}{\partial z} &+ \Delta\beta(\omega_0)(\cos\phi\sigma_3 + \sin\phi\sigma_2)\mathbf{U}(z,t) \\ &+ i\Delta\beta'(\omega_0)(\cos\phi\sigma_3 + \sin\phi\sigma_2)\frac{\partial\mathbf{U}(z,t)}{\partial t} \\ &- \frac{1}{2}\beta''(\omega_0)\frac{\partial^2\mathbf{U}(z,t)}{\partial t^2} \\ &+ \gamma\Big\{ |\mathbf{U}(z,t)|^2\,\mathbf{U}(z,t) \\ &\quad - \frac{1}{3}\left[\mathbf{U}^\dagger(z,t)\sigma_2\mathbf{U}(z,t)\right]\sigma_2\mathbf{U}(z,t)\Big\} = 0. \end{aligned} \qquad (39)$$

Writing $\mathbf{U} = u_+\hat{\mathbf{e}}_+ + u_-\hat{\mathbf{e}}_-$, where $\hat{\mathbf{e}}_+ = \cos(\phi/2)\hat{\mathbf{e}}_1 + i\sin(\phi/2)\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_- = i\sin(\phi/2)\hat{\mathbf{e}}_1 + \cos(\phi/2)\hat{\mathbf{e}}_2$, and noting that $\hat{\mathbf{e}}_+^*\cdot\hat{\mathbf{e}}_+ = \hat{\mathbf{e}}_-^*\cdot\hat{\mathbf{e}}_- = 1$ and $\hat{\mathbf{e}}_+^*\cdot\hat{\mathbf{e}}_- = \hat{\mathbf{e}}_+\cdot\hat{\mathbf{e}}_-^* = 0$, we find that the linear portions of the equations for $u_+(z,t)$ and $u_-(z,t)$ decouple. Thus, the unit vectors $\hat{\mathbf{e}}_+$ and $\hat{\mathbf{e}}_-$ correspond to the orthogonal polarization eigenstates—the states that do not evolve as a function of distance along the fiber for a fixed value of ϕ. The equations become

$$\begin{aligned} i\frac{\partial u_\pm}{\partial z} &\pm \Delta\beta u_\pm \pm i\Delta\beta'\frac{\partial u_\pm}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 u_\pm}{\partial t^2} \\ &+ \gamma\left(|u_+|^2 + |u_-|^2\right)u_\pm \\ &- \frac{\gamma}{3}\left[\sin\phi\left(|u_+|^2 - |u_-|^2\right) - i\cos\phi\left(u_+^*u_- - u_-^*u_+\right)\right] \\ &\times\left[\pm\sin\phi u_\pm \mp i\cos\phi u_\mp\right] = 0. \end{aligned} \qquad (40)$$

While we have not explicitly written the dependence of $u_\pm$ on z and t and the dependence of β and its frequency derivatives on ω_0 for the sake of compactness, these dependences should be kept in mind. In the case that $\cos\phi = 1$, corresponding to linear birefringence, (40) becomes simply

$$\begin{aligned} i\frac{\partial u_\pm}{\partial z} &\pm \Delta\beta u_\pm \pm i\Delta\beta'\frac{\partial u_\pm}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 u_\pm}{\partial t^2} \\ &+ \gamma\left(|u_\pm|^2u_\pm + \frac{2}{3}|u_\mp|^2u_\pm + \frac{1}{3}u_\mp^2u_\pm^*\right) = 0 \end{aligned} \qquad (41)$$

which is a well-known result.

The second term in (41) corresponds physically to phase beating due to the birefringence, and its scale length is typically meters, which corresponds to the intermediate group in Fig. 1. All other terms in (41) are typically associated with the third and longest group of scale lengths. In keeping with the usual multiple scale length procedure, we may average over the rapid variations due to the birefringent beating. To so do, it is useful to first transform (41) into the form

$$\begin{aligned} i\frac{\partial u_\pm}{\partial z} &\pm i\Delta\beta'\frac{\partial u_\pm}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 u_\pm}{\partial t^2} \\ &+ \gamma\left[|u_\pm|^2u_\pm + \frac{2}{3}|u_\mp|^2u_\pm + \frac{1}{3}u_\mp^2u_\pm^*\exp(\mp 4i\Delta\beta z)\right] = 0 \end{aligned} \qquad (42)$$

where we have set $u_\pm(z,t) = w_\pm(z,t)\exp[\pm i\Delta\beta(\omega_0)z]$ and then replaced $w_\pm$ with $u_\pm$. This transformation shifts the central wavenumbers of u_+ and u_- so that they no longer coincide at $\omega = 0$. We emphasize that u_+ and u_- have been transformed and are no longer exactly the same entities in (42) as they are in (41), since this point has led to some confusion. We also note that $|u_+|^2$ and $|u_-|^2$ are unchanged by this transformation, so that experimental observations in ON–OFF-keyed systems with square-law photodetectors are unaffected. Averaging over the rapidly varying final term, (42) becomes

$$i\frac{\partial u_\pm}{\partial z} \pm i\Delta\beta'\frac{\partial u_\pm}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 u_\pm}{\partial t^2} + \gamma\left(|u_\pm|^2 + \frac{2}{3}|u_\mp|^2\right)u_\pm = 0. \qquad (43)$$

A mistake that one finds in the literature is to drop the term proportional to $\Delta\beta$ in (41) or, equivalently, to set $\exp(\pm i\Delta\beta z) = 1$. In this case, the equation corresponds to the practically unrealizable limit in which the beat length is very long compared to the physical length scales corresponding to the other terms

in (41), which include the chromatic dispersion, nonlinearity, and the polarization-induced differential group delay.

In the general case, in which the ellipticity is arbitrary, (40) is somewhat complex. However, in the limit in which the beat length is small compared to the other scale lengths represented in (40), (40) simply becomes [12]

$$i\frac{\partial u_\pm}{\partial z} \pm i\Delta\beta'\frac{\partial u_\pm}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 u_\pm}{\partial t^2} + \gamma\left(1-\frac{1}{3}\sin^2\phi\right)|u_\pm|^2 u_\pm + \gamma\left(1+\frac{1}{3}\sin^2\phi-\frac{1}{3}\cos^2\phi\right)|u_\mp|^2 u_\pm = 0 \quad (44)$$

where, just as in going from (41) to (42), we first set $u_\pm(z,t) = w_\pm(z,t)\exp[\pm i\Delta\beta(\omega_0)z]$, we next average over the rapidly varying terms, and we finally replace $w_\pm$ with $u_\pm$. The limit in which the beat length is small compared to the other physical length scales in (41) is the usual physical limit in optical fiber communications systems.

From (44), we infer that the strength of the cross-phase modulation relative to the self-phase modulation coefficient depends on the ellipticity. It is 2/3 for linear birefringence, two for circular birefringence, and one when $\phi = \tan^{-1}(1/\sqrt{2}) \simeq 35^\circ$. Thus, measuring the strength of the cross-phase modulation relative to the self-phase modulation allows us to infer the ellipticity of the fiber's birefringence—a striking example of how a nonlinear measurement allows one to infer a linear property of the fiber. Measurements by Botineau and Stolen [14] of this ratio indicate that optical fibers are linearly birefringent to a good approximation. Other linear measurements (see [36] and references cited therein), as well as more recent polarization optical time domain reflectometry measurements [37] support this conclusion, although there is some evidence of a small residual ellipticity due to twisting. A difficulty in making this assessment is that twisting will induce an apparent ellipticity even when the birefringence is locally linear, and this geometric ellipticity is expected to be 10–20 times larger than the helicity induced in the material. This geometric ellipticity will also affect the ratio between the cross-phase and self-phase modulation as long as the nonlinear scale length is long compared to the twist length [38]. Thus, in this limit there is no way to distinguish between geometric and intrinsic ellipticity. However, when the nonlinear scale length becomes comparable to the twist length, it is possible to distinguish these two sources of ellipticity. This experiment could be done in principle with high-power lasers.

Recent theoretical work shows that a small ellipticity makes no difference in the results that are calculated when it is assumed that the fibers are linearly birefringent [39]. Thus, the small ellipticity can be neglected for all practical purposes, and we do so throughout the remainder of this paper.

B. Randomly Varying Birefringence and the Nonlinear Schrödinger Equation

We now focus on the second group of length scales in Fig. 1. On this length scale, we must take into account the variation of the birefringence as a function of z. Since we are assuming that the birefringence is linear, we may write $\Delta\mathsf{B}(z,\omega_0) = \Delta\beta(z,\omega_0)\{\cos[\theta(z)]\sigma_3 + \sin[\theta(z)]\sigma_1\}$ and similarly $\Delta\mathsf{B}'(z,\omega_0) = \Delta\beta'(z,\omega_0)\{\cos[\theta(z)]\sigma_3 + \sin[\theta(z)]\sigma_1\}$. As noted earlier, we are assuming that $\theta(z)$ is independent of frequency. From here on, we will write $\Delta\beta(z)$ rather than $\Delta\beta(z,\omega_0)$, $\Delta\beta'(z)$ rather than $\Delta\beta'(z,\omega_0)$, and β'' rather than $\beta''(\omega_0)$, but the dependence on ω_0 is not being neglected. We do not make a note of it explicitly since it appears as a constant parameter in the remainder of the discussion. Equation (34) now becomes

$$\begin{aligned} &i\frac{\partial \mathbf{U}(z,t)}{\partial z} + \Delta\beta(z)\left\{\cos\left[\theta(z)\right]\sigma_3 + \sin\left[\theta(z)\right]\sigma_1\right\}\mathbf{U}(z,t) \\ &\quad + i\Delta\beta'(z)\left\{\cos\left[\theta(z)\right]\sigma_3 + \sin\left[\theta(z)\right]\sigma_1\right\}\frac{\partial \mathbf{U}(z,t)}{\partial t} \\ &\quad - \frac{1}{2}\beta''\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} \\ &\quad + \gamma\Big\{|\mathbf{U}(z,t)|^2\,\mathbf{U}(z,t) \\ &\qquad - \frac{1}{3}\left[\mathbf{U}^\dagger(z,t)\sigma_2\mathbf{U}(z,t)\right]\sigma_2\mathbf{U}(z,t)\Big\} = 0. \end{aligned} \quad (45)$$

The third term in (45) is responsible for the usual linear PMD, as we will show.

The quantities $\Delta\beta(z)$ and $\theta(z)$ vary on a length scale that is referred to as the fiber correlation length and is typically on the order of 10–100 m. It is included in the second group of length scales in Fig. 1. We will use the multiple length scale technique to average over these variations, which are rapid relative to the length scales in the third group of length scales, and we will obtain an equation that is valid for this third group of length scales.

Focusing on the linear terms in (45) and transforming to the Fourier domain, we have

$$i\frac{\partial \tilde{\mathbf{U}}(z,\omega)}{\partial z} + \left\{\cos\left[\theta(z)\right]\sigma_3 + \sin\left[\theta(z)\right]\sigma_1\right\} \times \left[\Delta\beta(z) + \Delta\beta'(z)\omega\right]\tilde{\mathbf{U}}(z,\omega) = 0. \quad (46)$$

Transforming $\tilde{\mathbf{U}}(z,\omega)$ in a way that would diagonalize its evolution were it not for the z-variation of $\theta(z)$, we let

$$\tilde{\mathbf{V}}(z,\omega) = \mathsf{R}^{-1}(z)\tilde{\mathbf{U}}(z,\omega) \quad (47)$$

where $\mathsf{R}^{-1} = \cos(\theta/2)\mathsf{I} + \sin(\theta/2)\sigma_2$. Equation (46) now becomes

$$i\frac{\partial \tilde{\mathbf{V}}(z,\omega)}{\partial z} + \left\{\left[\Delta\beta(z) + \Delta\beta'(z)\omega\right]\sigma_3 + \left[\theta_z(z)/2\right]\sigma_2\right\}\tilde{\mathbf{V}}(z,\omega) = 0 \quad (48)$$

where $\theta_z(z) \equiv d\theta(z)/dz$. We next write

$$\tilde{\mathbf{W}}(z,\omega) = \mathsf{T}^{-1}(z)\tilde{\mathbf{V}}(z,\omega) \quad (49)$$

where T satisfies (48) at $\omega = 0$ so that

$$i\frac{\partial \mathsf{T}(z)}{\partial z} + \left\{\Delta\beta(z)\sigma_3 + \left[\theta_z(z)/2\right]\sigma_2\right\}\mathsf{T}(z) = 0 \quad (50)$$

with $\mathsf{T}(z=0)=\mathsf{I}$. It follows that $\tilde{\mathbf{W}}(z,\omega=0)$ is constant. In effect, we have mathematically "frozen" the rapid motion on the Poincaré sphere due to rapidly varying birefringence at $\omega=0$. This transformation will enable us to observe the differential motion as ω varies, which is the physical origin of PMD. Explicitly, we now find

$$i\frac{\partial\tilde{\mathbf{W}}(z,\omega)}{\partial z}+\Delta\beta'(z)\omega\bar{\sigma}_3(z)\tilde{\mathbf{W}}(z,\omega)=0 \tag{51}$$

where $\bar{\sigma}_3(z)=\mathsf{T}^{-1}(z)\sigma_3\mathsf{T}(z)$.

To verify that (51) includes the effects of PMD, we investigate two important properties of PMD. The first is the spread of pulses in the time domain. It is useful to first define the Stokes vector. We begin by allowing $\tilde{\mathbf{W}}(z,\omega)=\alpha(\omega)\tilde{\mathbf{A}}(z,\omega)$, where $|\tilde{\mathbf{A}}(z,\omega)|^2=1$. One can show by direct substitution of $\tilde{\mathbf{W}}(z,\omega)=\alpha(\omega)\tilde{\mathbf{A}}(z,\omega)$ in (51) that $\partial\alpha(\omega)/\partial z=0$ and hence $\alpha(\omega)$ is independent of z. As a consequence, the effect of the differential polarization rotation in (51) is to trade energy among the components of the Jones vector $\tilde{\mathbf{W}}(z,\omega)$ and to rotate the overall phase, but there is no effect on the total energy. We define the components of the Stokes vector as

$$\begin{aligned}S_1(z,\omega)&=\tilde{\mathbf{A}}^\dagger(z,\omega)\sigma_3\tilde{\mathbf{A}}(z,\omega)\\ S_2(z,\omega)&=\tilde{\mathbf{A}}^\dagger(z,\omega)\sigma_1\tilde{\mathbf{A}}(z,\omega)\\ S_3(z,\omega)&=-\tilde{\mathbf{A}}^\dagger(z,\omega)\sigma_2\tilde{\mathbf{A}}(z,\omega).\end{aligned} \tag{52}$$

We now define a matrix $\tilde{\mathsf{F}}(z,\omega)$ such that

$$i\frac{\partial\tilde{\mathbf{A}}(z,\omega)}{\partial\omega}+\tilde{\mathsf{F}}(z,\omega)\tilde{\mathbf{A}}(z,\omega)=0. \tag{53}$$

Since the transformation relating $\tilde{\mathbf{A}}(z,\omega)$ at different values of ω is unitary, the matrix $\tilde{\mathsf{F}}(z,\omega)$ must be Hermitian. Hence, its eigenvalues are real and its eigenvectors are orthogonal. Taking the ω-derivative of (51) and the z-derivative of (53), we find that the compatibility condition for these two derivatives is the evolution equation

$$\begin{aligned}\frac{\partial\tilde{\mathsf{F}}(z,\omega)}{\partial z}&=i\Delta\beta'(z)\omega\left[\bar{\sigma}_3(z)\tilde{\mathsf{F}}(z,\omega)-\tilde{\mathsf{F}}(z,\omega)\bar{\sigma}_3(z)\right]\\ &\quad+\Delta\beta'(z)\bar{\sigma}_3(z)=0.\end{aligned} \tag{54}$$

The trace of $\tilde{\mathsf{F}}(z,\omega)$ is independent of z, and we will designate the eigenvalues of $\tilde{\mathsf{F}}(z,\omega)$ as $T_{\rm off}(\omega)\pm T_{\rm PMD}(z,\omega)$. The eigenvectors of $\tilde{\mathsf{F}}(z,\omega)$ are conventionally referred to as the principal states, while the difference between the eigenvalues $2T_{\rm PMD}(z,\omega)$ is conventionally referred to as the differential group delay. The matrix $\tilde{\mathsf{F}}(z,\omega)$ is directly related to the pulse spreading. Defining the mean signal time and the mean-square signal time as

$$\begin{aligned}\overline{T(z)}&=\frac{\int_{-\infty}^{\infty}t\,|\mathbf{W}(z,t)|^2\,dt}{\int_{-\infty}^{\infty}|\mathbf{W}(z,t)|^2\,dt}\\ \overline{T^2(z)}&=\frac{\int_{-\infty}^{\infty}t^2\,|\mathbf{W}(z,t)|^2\,dt}{\int_{-\infty}^{\infty}|\mathbf{W}(z,t)|^2\,dt}\end{aligned} \tag{55}$$

we may define the mean-square signal spread $\Sigma^2(z)=\overline{T^2(z)}-\overline{T(z)}^2$. By analogy with the Stokes vector, it is conventional to write

$$\begin{aligned}\tilde{\mathsf{F}}(z,\omega)&=T_{\rm off}(\omega)\mathsf{I}\\ &\quad+\frac{1}{2}\left[\Omega_1(z,\omega)\sigma_3+\Omega_2(z,\omega)\sigma_1-\Omega_3(z,\omega)\sigma_2\right].\end{aligned} \tag{56}$$

The vector $\mathbf{\Omega}=(\Omega_1,\Omega_2,\Omega_3)$ is referred to as the polarization dispersion vector. One then finds the well-known result [2], [3]

$$\Sigma^2(z)-\Sigma^2(z=0)\simeq\frac{1}{4}\left|\mathbf{\Omega}(z,\omega=0)\times\mathbf{S}(z,\omega=0)\right|^2 \tag{57}$$

which holds when the length is short enough so that the spreading has not become large. More details, along with more general expressions, may be found in [26], [40]. From a physical standpoint, what happens is that a signal launched in either one of the principal states is either advanced or retarded by $T_{\rm PMD}$ relative to $T_{\rm off}$, while a signal launched in any other state decomposes into the two principal states, so that part is advanced and part is retarded.

The second result that we will verify is the length scale on which spreading occurs. To determine this length scale, we first write

$$\bar{\sigma}_3(z)\equiv\begin{pmatrix}\cos\theta_s(z) & \sin\theta_s(z)\exp\left[i\phi_s(z)\right]\\ \sin\theta_s(z)\exp\left[-i\phi_s(z)\right] & -\cos\theta_s(z)\end{pmatrix} \tag{58}$$

which effectively defines θ_s and ϕ_s. When the fiber is untwisted so that the birefringence variation is purely random, then a wide variety of physical models all lead to the conclusion that the fiber is exponentially correlated so that

$$\begin{aligned}C(z_1,z_2)&\equiv\left\langle\Delta\beta'(z_1)\cos\left[\theta_s(z_1)\right]\Delta\beta'(z_2)\cos\left[\theta_s(z_2)\right]\right\rangle\\ &=\frac{1}{3}\left\langle\left[\Delta\beta'(z)\right]^2\right\rangle\exp\left(-|z_1-z_2|/z_{\rm corr}\right)\end{aligned} \tag{59}$$

where $z_{\rm corr}$ is the fiber correlation length, and the brackets indicate an ensemble average over fiber realizations. One then finds that

$$\begin{aligned}\left\langle\Omega_1^2(z)\right\rangle&=8\int_0^z dz_1\int_0^z dz_2\\ &\quad\times\left\langle\Delta\beta'(z_1)\cos\left[\theta_s(z_1)\right]\Delta\beta'(z_2)\cos\left[\theta_s(z_2)\right]\right\rangle\end{aligned} \tag{60}$$

from which we conclude

$$\begin{aligned}\left\langle\Omega_1^2(z)\right\rangle&=\frac{8}{3}\left\langle\left[\Delta\beta'(z)\right]^2\right\rangle\\ &\quad\times\left\{z_{\rm corr}z+z_{\rm corr}^2\left[\exp(-z/z_{\rm corr})-1\right]\right\}.\end{aligned} \tag{61}$$

We note that this result is independent of frequency, except for a slow variation due to the dependence of $\Delta\beta'$ on ω_0. Similar

results hold for $\langle\Omega_2^2(z)\rangle$ and $\langle\Omega_3^2(z)\rangle$. Therefore, we conclude that

$$\begin{aligned}\left\langle T_{\mathrm{PMD}}^2(z)\right\rangle &= \frac{1}{4}\left\langle|\mathbf{\Omega}(z)|^2\right\rangle \\ &= 2\left\langle\left[\Delta\beta'(z)\right]^2\right\rangle \\ &\quad\times\left\{z_{\mathrm{corr}}z + z_{\mathrm{corr}}^2\left[\exp(-z/z_{\mathrm{corr}})-1\right]\right\} \quad (62)\end{aligned}$$

which is another well-known result [2], [3].

We now make the substitution $\mathbf{U}(z,t)=\mathsf{R}(z)\mathsf{T}(z)\mathbf{W}(z,t)$ in (45), where R and T are defined after (47). We obtain

$$\begin{aligned}&i\frac{\partial\mathbf{W}(z,t)}{\partial z}-\frac{1}{2}\beta''\frac{\partial^2\mathbf{W}(z,t)}{\partial t^2}+\frac{8}{9}\gamma|\mathbf{W}(z,t)|^2\mathbf{W}(z,t)\\ &\quad=-i\Delta\beta'(z)\bar{\sigma}_3(z)\frac{\partial\mathbf{W}(z,t)}{\partial t}\\ &\qquad+\frac{1}{3}\gamma\Big\{\left[\mathbf{W}^\dagger(z,t)\bar{\sigma}_2(z)\mathbf{W}(z,t)\right]\bar{\sigma}_2(z)\mathbf{W}(z,t)\\ &\qquad\quad-\frac{1}{3}|\mathbf{W}(z,t)|^2\mathbf{W}(z,t)\Big\} \quad (63)\end{aligned}$$

where $\bar{\sigma}_2(z)=\mathsf{T}^{-1}(z)\sigma_2\mathsf{T}(z)$. This equation explicitly separates the rapidly oscillating portions on the right side of the equation from the slowly varying portions on the left. The factor of 8/9 that multiplies the nonlinear factor when the slowly varying portion is properly separated from the rapidly varying portion has been verified experimentally [35], [41]. As we have already shown, the first term on the right side of (63) leads to the conventional linear PMD. Equation (63) is the Manakov-PMD equation. The second term on the right, which is due to incomplete mixing on the Poincaré sphere, is referred to as nonlinear PMD. It is negligible as long as the correlation length is short compared to the nonlinear scale length [42], which is always true in communications systems. However, it can play an important role in short pulse experiments [43].

When the PMD can be neglected, the Manakov-PMD equation, (63) becomes

$$i\frac{\partial\mathbf{U}(z,t)}{\partial z}-\frac{1}{2}\beta''\frac{\partial^2\mathbf{U}(z,t)}{\partial t^2}+\frac{8}{9}\gamma\left|\mathbf{U}(z,t)\right|^2\mathbf{U}(z,t)=0 \quad (64)$$

where we have replaced $\mathbf{W}$ with $\mathbf{U}$. As in the case where we previously transformed $u_\pm$ into $w_\pm$ and then replaced $w_\pm$ with $u_\pm$ in deriving (42) from (41), we have changed the meaning of $\mathbf{U}$. This Jones vector no longer corresponds to a polarization state in a fixed frame. The evolution of the polarization state at $\omega=0$ has been removed. As before, this transformation does not matter in ON–OFF-keyed systems with square law detectors because $|\mathbf{W}(z,t)|^2=|\mathbf{U}(z,t)|^2$. Equation (64) is the Manakov equation. If we assume that at $z=0$, $\mathbf{U}(z=0,t)$ is in a single polarization state so that $\mathbf{U}(z=0,t)=u(z=0,t)\hat{\mathbf{e}}_+$, then there is no coupling to the orthogonal polarization state, and one may replace (64) with the scalar equation

$$i\frac{\partial u(z,t)}{\partial z}-\frac{1}{2}\beta''\frac{\partial^2u(z,t)}{\partial t^2}+\frac{8}{9}\gamma\left|u(z,t)\right|^2u(z,t)=0. \quad (65)$$

Thus, we find that the conditions for the nonlinear Schrödinger equation to hold are that PMD is negligible and the signal is initialized in a single polarization state. Of course, the polarization state of the signal is varying rapidly as a function of z as the signal propagates along the fiber. This rapid variation was hidden by the transformation from $\mathbf{U}$ to $\mathbf{W}$. However, when the nonlinear Schrödinger equation applies, this variation is the same at every point in time and does not affect the evolution of the intensity envelope. The physical meaning of the mathematical transformations that we have carried out is to isolate these rapid variations and determine when they do not affect the intensity envelope.

An issue that we have not touched up to this point is gain and loss. In practice, gain and loss always play an important role, so that $\beta(\omega_0)$ becomes complex. Returning to (6), we would replace the equation for the electric field with

$$\mathbf{E}(\mathbf{r},t)=\mathbf{F}(\mathbf{r}_\perp,\omega_0)\exp\left\{i\mathrm{Re}\left[\beta(\omega_0)\right]z-i\omega_0t\right\}+\text{c.c.} \quad (66)$$

where $\mathrm{Re}[\beta(\omega_0)]$ indicates the real part of $\beta(\omega_0)$, and we write $g(z)=\mathrm{Im}[\beta(\omega_0,z)]$, where we account for the z-variation of the imaginary part of the propagation constant explicitly. We note that the length scale associated with the attenuation is the third or the longest length scale. With this change, the derivation of the nonlinear Schrödinger equation proceeds as before, ultimately terminating at the equation

$$\begin{aligned}i\frac{\partial u(z,t)}{\partial z}+ig(z)u(z,t)-\frac{1}{2}\beta''\frac{\partial^2u(z,t)}{\partial t^2}\\ +\frac{8}{9}\gamma\left|u(z,t)\right|^2u(z,t)=0 \quad (67)\end{aligned}$$

where $g(z)$ is the attenuation coefficient, which is negative at the amplifier locations and positive otherwise. Referring to Fig. 1, we see that the typical length scale for $g(z)$ is about 20 km. This length scale is often significantly shorter than the length scales for nonlinearity and chromatic dispersion, which may be several hundred kilometers. In this case, it is possible to treat the gain and loss as rapidly varying and obtain new averaged equations [20], [44], [45]. In practice, the difference between length scales is usually not wide enough to make this procedure worthwhile and one solves (67) with the gain and the loss included. However, this procedure has been carried out for soliton systems to demonstrate that solitons still exist in the presence of gain and loss and to determine the changes in the soliton shape that gain and loss cause [44], [45].

III. Interaction of Nonlinearity and PMD

We now consider two examples of the interaction between nonlinearity and PMD. In the first example, the nonlinearity induces a chirp in the signal, which in combination with higher order PMD can lead to pulse compression and a reduction in the eye-opening penalty that would be present in the absence of

PMD. In the second example, the nonlinear interaction between two channels induces a nonlinear rotation in the principal states of the bits in each channel that varies bit to bit. While this rotation does not directly induce an eye-closing penalty, it makes it impossible to use most current PMD compensation techniques, since these techniques typically rely on each bit having the same principal states, and, more generally, the same polarization variation as a function of time within a bit slot.

These two examples serve to underline two points. The first is that the impact of fiber nonlinearity can be quite complex. The second, which is closely related to the first, is that one cannot simply add the penalties due to PMD and nonlinearity without careful justification.

The basic equation that governs our investigation is the Manakov-PMD equation that we write in the form

$$i\frac{\partial \mathbf{U}(z,t)}{\partial z} + i\Delta\beta'(z)\bar{\sigma}_3(z)\frac{\partial \mathbf{U}(z,t)}{\partial t} - \frac{1}{2}\beta''\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} + \frac{8}{9}\gamma\left|\mathbf{U}(z,t)\right|^2\mathbf{U}(z,t) = 0. \quad (68)$$

We do not include the nonlinear PMD term in (63) since we will be considering pulses in the picosecond range, as is typical in optical fiber communications systems, and the nonlinear PMD term is negligible in this range [42]. We have returned to using $\mathbf{U}$ as the dependent variable for the wave envelope, using it to replace $\mathbf{W}$. We remind the reader that the rapid polarization evolution at $\omega = 0$ has been removed.

A. Improvement of a Nonlinear Chirped Signal Due to Higher Order PMD

It has been known since the early work of Poole and Giles [46] that a chirped signal can be compressed as well as spread by its interaction with fiber PMD. While Poole and Giles [46] focused on chirp that is due to dispersion, other mechanisms such as an initial chirp due to the laser transmitter can also produce the same effect [47], [48]. We show schematically the physical origin of this effect in Fig. 2. Due to chirp, the leading edge of the pulse (earlier times) has a different frequency from the trailing edge (later times). In the case shown in Fig. 2(a), the leading edge of the pulse has a frequency that differs from the mean ω_0 by $+\Delta\omega$ and the trailing edge has a frequency that differs by $-\Delta\omega$. When second-order PMD is large relative to the first-order PMD, the principal state of polarization can change significantly over the bandwidth of the signal. We show this effect in Fig. 2(b). Here, we show the variation of the principal state of polarization as a curve on the Poincaré sphere. Thus, the state of polarization of the signal, which is shown as a dot, is primarily aligned to the fast axis when the frequency is $\omega_0 - \Delta\omega$, corresponding to the trailing edge, and is primarily aligned to the slow axis when the frequency is $\omega_0 + \Delta\omega$, corresponding to the leading edge, as shown in Fig. 2(c). As a consequence, the pulse compresses during propagation, as shown in Fig. 2(d).

We are interested here in the case when the chirp is due to nonlinearity [8], [10]. In contrast to other mechanisms, the pulse compression and hence the signal improvement depends on the pulse power. Systems that show no signal improvement

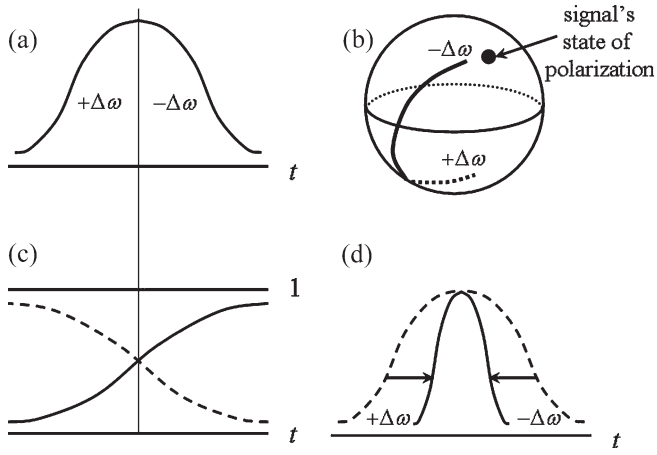

Fig. 2. (a) Chirp leads to a time-dependent frequency variation. In the presence of large second-order PMD, this chirp also leads to (b) variation of the principal states over the bandwidth of the signal and (c) time-dependent variation in the fraction of energy in a signal pulse that is aligned with the fast (—) and slow (- - -) axes, even though the signal is in a single polarization state. (d) This variation leads to pulse compression.

at low powers may show signal improvement at higher powers. The theoretical development is based on (53)–(57) with two important modifications. The first is that we set $T_{\rm off} = 0$, which is physically equivalent to assuming that the pulse is originally in a single polarization state. Second, (53) must be modified to take account of the presence of chirp so that it becomes

$$i\frac{\partial \tilde{\mathbf{A}}(z,\omega)}{\partial \omega} + \left[\tilde{\mathsf{F}}(z,\omega) + \phi'(z,\omega)\mathsf{I}\right]\tilde{\mathbf{A}}(z,\omega) = 0 \quad (69)$$

where $\phi(z,\omega)$ is an overall phase that is defined by $\tilde{\mathbf{A}}(z,\omega) = |\tilde{\mathbf{A}}(z,\omega)|\exp[i\phi(z,\omega)]$, and $\phi'(z,\omega) = \partial\phi(z,\omega)/\partial\omega$. With these modifications, we find that

$$\overline{T(z)} = \frac{\int_{-\infty}^{\infty}\alpha^2(\omega)\left[\frac{1}{2}\mathbf{\Omega}(z,\omega)\cdot\mathbf{S}(z,\omega) + \phi'(z,\omega)\right]d\omega}{\int_{-\infty}^{\infty}\alpha^2(\omega)d\omega}$$

$$\overline{T^2(z)} = \frac{\int_{-\infty}^{\infty}\alpha^2(\omega)\left|\frac{1}{2}\mathbf{\Omega}(z,\omega) + \phi'(z,\omega)\mathbf{S}(z,\omega)\right|^2 d\omega}{\int_{-\infty}^{\infty}\alpha^2(\omega)d\omega} \quad (70)$$

which is equivalent to Karlsson's result [40] and from which $\Sigma^2(z) = \overline{T^2(z)} - \overline{T(z)}^2$ can be readily determined. Focusing on the case in which PMD is weak, so that it makes sense to just consider first and second-order PMD, one then finds that $\alpha^2(\omega)$ varies rapidly compared to $\mathbf{\Omega}(z,\omega)$ and $\mathbf{S}(z,\omega)$. We will also assume that while the chirp given by $\phi''(z,\omega) = \partial^2\phi(z,\omega)/\partial\omega^2$ may be large, its variation with frequency is small. We may then make the Taylor expansion

$$\overline{T(z)} = \left[\frac{1}{2}\mathbf{\Omega}\cdot\mathbf{S} + \phi'\right]_0 + \frac{1}{2}\overline{\omega^2}\frac{\partial^2}{\partial\omega^2}\left[\frac{1}{2}\mathbf{\Omega}\cdot\mathbf{S} + \phi'\right]_0$$

$$\overline{T^2(z)} = \left|\frac{1}{2}\mathbf{\Omega} + \phi'\mathbf{S}\right|_0^2 + \frac{1}{2}\overline{\omega^2}\frac{\partial^2}{\partial\omega^2}\left|\frac{1}{2}\mathbf{\Omega} + \phi'\mathbf{S}\right|_0^2 \quad (71)$$

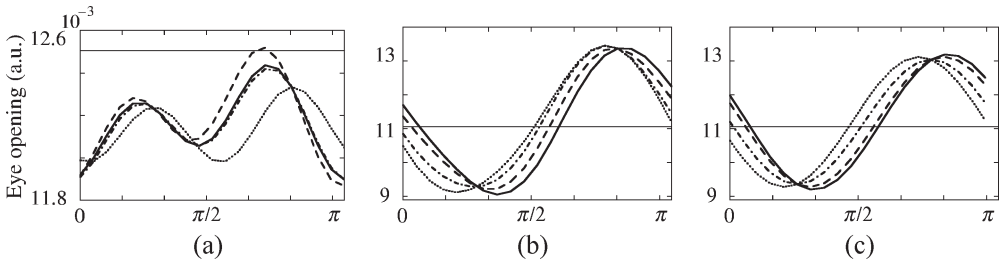

Fig. 3. Dependence of the eye-opening penalty on the angle ϕ for four values of θ $(0, \pi/4, \pi/2, 3\pi/4)$. (a) Simulation with an input power of 0.01 mW. (b) Simulation with an input power of 10 mW. (c) Analysis including only second-order PMD. In this case, second-order PMD dominates the behavior at high power. (Originally from [8]).

where all terms with a subscript 0 are evaluated at $\omega = 0$. We have not written explicitly the dependence of $\mathbf{\Omega}$ and $\mathbf{S}$ on z for the sake of compactness, but it is still present. We have defined the mean-square bandwidth $\overline{\omega^2} = \int_{-\infty}^{\infty} \omega^2 \alpha^2(\omega) d\omega / \int_{-\infty}^{\infty} \alpha^2(\omega) d\omega$, and we have assumed that the signal is centered at $\omega = 0$, so that $\overline{\omega} = \int_{-\infty}^{\infty} \omega \alpha^2(\omega) d\omega / \int_{-\infty}^{\infty} \alpha^2(\omega) d\omega = 0$. If we calculate $\Sigma^2(z)$ directly from (71), we find that a plethora of terms appear as a consequence, only some of which are included in discussions prior to Karlsson [40], as Karlsson has pointed out. Conversely, however, if we assume that $\phi' = 0$ at $\omega = 0$ and that ϕ'' is large compared to the frequency derivatives of $\mathbf{\Omega}$ and $\mathbf{S}$ so that the only terms of interest are the ones in which ϕ'' appears, and we assume that $\overline{\omega^2}$ is small, which is consistent with our initial Taylor expansion, then we obtain

$$\begin{aligned}\Sigma^2(z) - \Sigma^2(z=0) &= \frac{1}{4} |\mathbf{\Omega}(z,\omega) \times \mathbf{S}(z,\omega)|_0^2 - \frac{1}{4} |\mathbf{\Omega}(z=0,\omega) \times \mathbf{S}(z=0,\omega)|_0^2 \\ &+ \overline{\omega^2} \Bigg\{ [\phi''(z,\omega)]^2 - [\phi''(z=0,\omega)]^2 + \frac{1}{2}\phi''(z,\omega)\mathbf{\Omega}'(z,\omega) \\ &\cdot \mathbf{S}(z,\omega) - \frac{1}{2}\phi''(z=0,\omega)\mathbf{\Omega}'(z=0,\omega) \cdot \mathbf{S}(z=0,\omega) \Bigg\}_0 \end{aligned} \tag{72}$$

where $\mathbf{\Omega}' = \partial \mathbf{\Omega} / \partial \omega$, and we have used the result $\mathbf{\Omega} \cdot \mathbf{S}' = \mathbf{\Omega} \cdot (\mathbf{\Omega} \times \mathbf{S}) = 0$. This result is consistent with Bruyère [47].

From (72), we directly obtain the result first obtained by Poole and Giles [46] that when $\mathbf{\Omega}' \cdot \mathbf{S}$ has the opposite sign from ϕ'', pulse contraction can occur relative to the result with no PMD. If there is an initial chirp whose magnitude exceeds the final chirp, then there can be a pulse contraction relative to the initial pulse duration. The physical source of the chirp is irrelevant in (72), and a chirp due to the initial pulse generation, dispersion, and Kerr nonlinearity all lead to the same conclusion. However, nonlinearity has the special property that the chirp is power dependent, so that one can go from one regime to another by simply changing the input power.

In initial studies to demonstrate the reality of a PMD improvement when a nonlinearly induced chirp is present, and to determine its power dependence, Ibragimov *et al.* [8] simulated a 1.55-μm system with a 10-GHz nonreturn-to-zero pulse train in a dispersion-managed transmission line with a 100-km amplifier spacing. They used a standard fiber with a dispersion of 17 ps/nm-km and a compensating fiber with a dispersion of -120 ps/nm-km. The effective area of the standard fiber was 80 μm^2 and that of the compensating fiber was 30 μm^2. The length of the compensating fiber between the amplifiers was chosen to provide close-to-zero average dispersion. The power extinction ratio was set at -15 dB. The losses were 0.29 dB/km in the standard fiber and 0.5 dB/km in the compensating fiber. The overall propagation distance was 578 km. The average differential group delay due to the PMD was 23.2 ps, which is approximately a quarter of a bit period. The simulations used a 64-bit pseudorandom bit string and solved the Manakov-PMD equation using the coarse step method [42].

To calculate the penalty due to PMD, the input optical signal had two polarization components, $u_1 = U_0 \cos\theta$ and $u_2 = U_0 \sin\theta \exp(i\phi)$, where U_0 is determined by the input power. The reader should note that θ and ϕ defined here refer to the polarization state of the field envelope and are unrelated to θ and ϕ defined after (39), which refer to the polarization state of the fiber. Fig. 3(a) and (b) shows the eye opening as a function of θ, measured in arbitrary units for four different values of ϕ $(0, \pi/4, \pi/2, 3\pi/4)$ at two different powers and with the same fiber realization. The fiber PMD is 1.0 ps/km$^{1/2}$. Fig. 3 also shows the eye opening when the PMD is zero, which appears as a horizontal line. For this fiber realization, we find that there is some residual second-order PMD even when the power is very low. When the power equals 10 mW, the behavior is completely dominated by the second-order PMD. Fig. 3(c) shows the behavior that would be expected at 10 mW if only second order was present, and one finds that it nearly matches the variation in Fig. 3(b). When the second-order PMD dominates, about half the input polarization states yield a penalty that is lower than the penalty in the absence of PMD.

Fig. 4 shows another fiber realization with all other parameters the same. In this case, the first-order PMD dominates the behavior at low power and even when the power reaches 10 mW, a large residual contribution from the first-order PMD is visible. Nonetheless, even at 5 mW and certainly at 10 mW, there are values of θ and ϕ that lead to a reduction of the penalty that would be present with no PMD. The difference between the cases shown in Figs. 3 and 4 can be traced back to the behavior of the differential group delay as a function of frequency. In the case of Fig. 3, there is a sharp dip in the differential group delay in the center of the signal spectrum, while in the case of Fig. 4, the differential group delay is nearly flat across the spectrum.

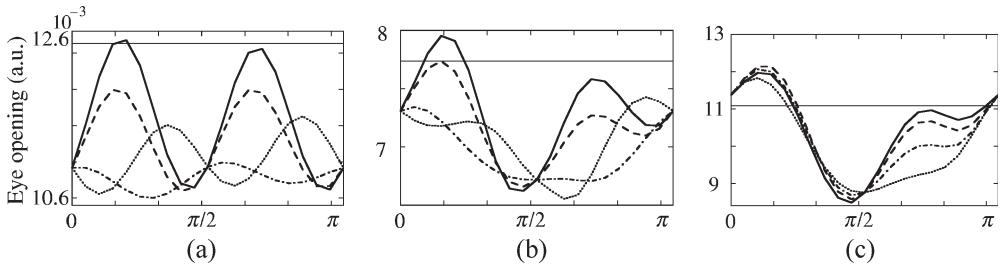

Fig. 4. Dependence of the eye-opening penalty on the angle ϕ for four values of θ $(0, \pi/4, \pi/2, 3\pi/4)$. Simulations with input powers of (a) 0.01 mW, (b) 5 mW, and (c) 10 mW. In this case, the first-order PMD dominates the behavior at low power, and its presence is visible at high power, although second-order PMD is becoming increasingly important. (Originally from [8]).

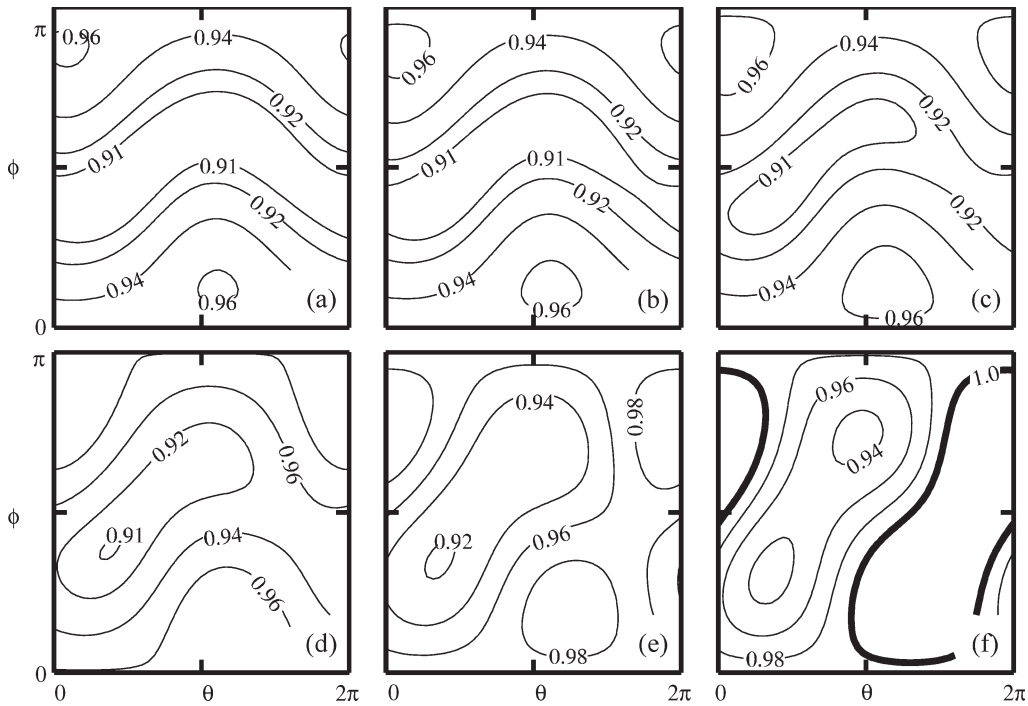

Fig. 5. Contours of constant PMD enhancement in the θ-ϕ plane of input polarization states. In each plot, contours are shown for PMD enhancements of 0.91, 0.92, 0.94, 0.96, 0.98, and 1.0. The thick contour corresponds to a PMD enhancement of 1.0. The input peak power for the return-to zero (RZ) signal is (a) 0, (b) 3, (c) 6, (d) 9, (e) 12, and (f) 15 dBm. The dispersion compensation is 100% of the residual dispersion of the line. (Originally from [10]).

More recently, Marks and Menyuk [10] simulated a 10-Gb/s return-to-zero system with five periods of a dispersion map, each of which consists of 100 km of standard fiber with a dispersion of 17 ps/nm-km and 15 km of dispersion-compensating fiber (DCF) with a dispersion of -120 ps/nm-km. They used pre- and postcompensating fiber at the beginning and at the end of the transmission to control the net dispersion of the entire system so that it is 100% dispersion compensated. The 100% dispersion compensation implies that when PMD is not present, the system performs optimally when nonlinearity is also not present and degrades as the signal power, and hence, the nonlinearity increases. The 1.55-μm signal is an initially unchirped Gaussian pulse with a 30% duty cycle. The fiber's nonlinear coefficient equals 1.3 $\mathrm{W}^{-1}\mathrm{km}^{-1}$, and the average differential group delay of the link is 33 ps. The receiver is modeled as an ideal square law detector followed by an 8-GHz electrical fifth-order Bessel filter. The authors investigated the effect of increasing the system nonlinearity by increasing the peak power of the input pulses from 0 up to 15 dBm. At the highest power, the propagation length is several times larger than the nonlinear scale length.

The simulation solves the Manakov-PMD equation using the coarse step method [42]. The eye closure is defined as the difference between the average power of the marks and the average power of the spaces at the clock recovery time. The PMD enhancement is defined as the ratio of the eye closure with the PMD to the eye closure when the fiber's PMD is set to zero. A PMD enhancement greater than 1 implies that including PMD increases the eye opening.

For one sample fiber realization, chosen so that the enhancement is particularly visible, Fig. 5 shows contours of constant PMD enhancement as a function of (θ, ϕ). For this fiber realization, the link's differential group delay is 21.9 ps, and the polarization-dependent chromatic dispersion is 80.2 ps^2. Fig. 5(a)–(f) shows the variation of the contours as the power increases in 3-dB increments from 0 to 15 dBm. The thick

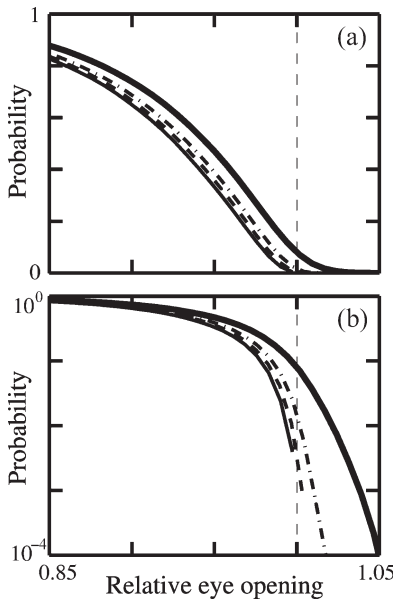

Fig. 6. Curves showing the probability of exceeding a given relative eye opening (complement of the cumulative density function) generated from a simulation consisting of 10^5 fiber realizations. The thin solid curve is for the system in which the fiber nonlinearity is set to zero. For the other curves, the same fiber realizations were used, and the nonlinearity was nonzero and held fixed with the peak power of the launched RZ pulse set to 9 dBm for the dashed curve, 12 dBm for the dot-dashed curve, and 15 dBm for the thick solid curve. (a) Linear scale and (b) log scale. (Originally from [10]).

contour in Fig. 5(f) corresponds to a PMD enhancement of 1; every input polarization state within this contour yields an improved performance relative to zero PMD.

Marks and Menyuk [10] simulated 100 000 randomly chosen fiber realizations in the same system configuration to explore the statistical significance of the PMD enhancement. For each fiber realization, they evaluated the eye opening both with and without PMD and both with and without nonlinearity. Fig. 6 shows the complement of the cumulative distribution functions for launched return-to-zero peak power of 9 dBm (dashed curve), 12 dBm (dot-dashed curve), and 15 dBm (thick solid curve). For comparison, this figure also shows the complement of the cumulative distribution function when the nonlinearity is set to zero (thin solid curve). The eye-opening values presented in Fig. 6 are normalized by the eye opening with nonlinearity and no PMD. The input polarization state was fixed for all fiber realizations and generally does not correspond to a principal state axis.

For low values of the peak power, the relative eye opening with nonlinearity is nearly the same as the relative eye opening without nonlinearity. As the peak power increases, however, the eye opening with both PMD and nonlinearity is larger than in the same case with the fiber nonlinearity set to zero for a large majority of the fiber realizations. Also, as the peak power increases, the eye opening with both PMD and nonlinearity is larger than in the case with no PMD for many fiber realizations. This result indicates that PMD enhancement occurs commonly as the power increases. As shown in Fig. 6, it is not difficult to find fiber realizations in which the PMD can increase the eye opening when nonlinearity is significant.

B. Nonlinear Polarization Rotation and Its Impact

The index of refraction can be changed by the optical power in a specific polarization state, resulting in a nonlinear birefringence [1], [5], [11]. In wavelength-division-multiplexed (WDM) systems, the nonlinear interaction between channels induces a nonlinear, bit-pattern-dependent phase change in an optical signal whenever there is optical power present at the other wavelengths, causing a bit-pattern-dependent change in the state of polarization. Fig. 7 illustrates this concept for a simple two-channel system. The bits at wavelength λ_1 that propagate alongside a long sequence of marks in the channel at λ_2 experience a small change in the fiber that changes their output state of polarization relative to the other bits in the channel. This effect becomes significant when the relative states of polarization of the channels are preserved over a distance that is long enough for the nonlinear interaction to accumulate, implying that the nonlinear change in the state of polarization is more prevalent in fibers with relatively low PMD, in which the polarization states of the channels remained correlated over a long distance. We thus find that while large PMD can lead to a pulse distortion in a single channel, it is actually helpful in reducing the nonlinear crosstalk between channels. The situation is analogous to the chromatic dispersion. While the chromatic dispersion may distort the signal inside a single channel, it actually reduces the nonlinear interaction between channels.

It is simplest to develop the theory of nonlinear polarization rotation in the limit where PMD is negligible. In this limit, the evolution is described by the Manakov equation with gain and loss

$$i\frac{\partial \mathbf{U}(z,t)}{\partial z} + ig(z)\mathbf{U}(z,t) - \frac{1}{2}\beta''(\omega_0)\frac{\partial^2 \mathbf{U}(z,t)}{\partial t^2} + \frac{8}{9}\gamma\left|\mathbf{U}(z,t)\right|^2\mathbf{U}(z,t) = 0 \quad (73)$$

which is (64), including the effects of gain and loss. We consider a WDM system, and we write

$$\mathbf{U}(z,t) = \sum_{m=1}^{N} \mathbf{U}_m(z,t)\exp(i\beta_m z - i\omega_m t) \quad (74)$$

where β_m is the wavenumber of the mth WDM channel relative to $\beta(\omega_0)$, and ω_m is the radial frequency of that channel relative to ω_0. When four-wave mixing and intrachannel chromatic dispersion can be neglected, we obtain for the lth channel

$$i\frac{\partial \mathbf{U}_l(z,t)}{\partial z} + ig(z)\mathbf{U}_l(z,t) + i\beta''(\omega_0)\omega_l\frac{\partial \mathbf{U}_l(z,t)}{\partial t} + \frac{8}{9}\gamma\sum_{m=1,\neq l}^{N}\left|\mathbf{U}_m(z,t)\right|^2\mathbf{U}_l(z,t) + \frac{8}{9}\gamma\sum_{m=1}^{N}\left[\mathbf{U}_l(z,t)\cdot\mathbf{U}_m^*(z,t)\right]\mathbf{U}_m(z,t) = 0 \quad (75)$$

where N is the total number of channels. Focusing on channel l, we may define a new time variable $t_l = t - z/[\beta''(\omega_0)\omega_l]$,

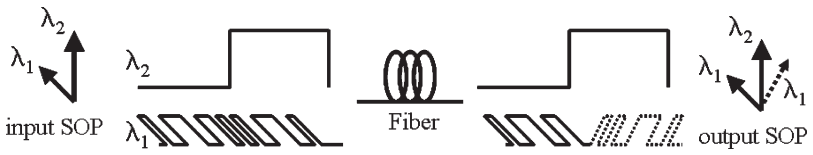

Fig. 7. Optical power induces a small nonlinear birefringence that randomizes the state of polarization, limiting the effectiveness of the first-order PMD compensation. (Originally from [11]).

in which the group velocity motion of the lth channel is removed. Equation (75) now becomes

$$i\frac{\partial \mathbf{U}_l(z,t_l)}{\partial z}+ig(z)\mathbf{U}_l(z,t_l)+\frac{8}{9}\gamma\sum_{m=1,\neq l}^{N}|\mathbf{U}_m(z,t_l)|^2\mathbf{U}_l(z,t_l)$$
$$+\frac{8}{9}\gamma\sum_{m=1}^{N}\left[\mathbf{U}_l(z,t_l)\cdot\mathbf{U}_m^*(z,t_l)\right]\mathbf{U}_m(z,t_l)=0. \quad (76)$$

It follows from (76) that

$$\frac{\partial\left|\mathbf{U}_l(z,t_l)\right|^2}{\partial z}+2g(z)\left|\mathbf{U}_l(z,t_l)\right|^2=0 \quad (77)$$

which in turn implies that the nonlinearity induces no waveform distortion and that there is no waveform distortion at all—merely gain and loss. Since the same result holds for all the channels, we can write [28]

$$|\mathbf{U}_m(z,t_l)|^2=P_m(z)p_m\left[t_l-\beta''(\omega_0)(\omega_m-\omega_l)z\right] \quad (78)$$

where $P_m(z)$ is the peak power in channel m at any point in z taken over time, and $p_m(t)$ gives the data modulation as a function of time.

Our ability to write $\mathbf{U}_m$ in the form (78) greatly simplifies the analysis. It is a consequence of neglecting intrachannel chromatic dispersion. This assumption requires additional comment. Modern-day long-haul WDM systems use dispersion management in which intrachannel chromatic dispersion—far from being negligible—is typically quite large, rapidly spreading each bit over the equivalent of several bit periods and then using fiber with dispersion of the opposite sign to compensate for that. Dispersion management significantly reduces the impact of the nonlinearity on system performance. However, the general problem does not appear possible to analyze except by using full computer simulations, and thus, work on (78) provides an important baseline from which to study the more general case. As a consequence, several experimental and theoretical studies of systems in which (78) holds have been carried out [6], [30]. Here, we follow the theoretical development of Vannucci *et al.* [28].

From (76), one can derive the equation

$$\frac{\partial\mathbf{S}_l(z,t_l)}{\partial z}+2g(z)\mathbf{S}_l(z,t_l)$$
$$=\frac{8}{9}\gamma\sum_{m=1}^{N}\left|\mathbf{U}_l(z,t_l)\right|^2\mathbf{S}_m(z,t_l)\times\mathbf{S}_l(z,t_l) \quad (79)$$

where $\mathbf{S}_l$ is the Stokes vector for a single channel. While (79) does not hold in general when the chromatic dispersion cannot be neglected, there are important limits besides the one considered here where it does hold. The first is the case of solitons, in which case $|\mathbf{U}_l|^2S_{lj}$, $j=1,2,3$ refers to the final three Stokes parameters of the individual solitons averaged over time [6]. The second is when an entire channel is considered, in which case $\mathbf{S}(z)$ is averaged over time and refers to the Stokes parameters of the entire channel [49]. This case holds more generally than the case being considered here, since it is not necessary for the intrachannel dispersion to be negligible. However, the Stokes parameters in this case provide no time domain information.

We now return to using the Stokes vector, and we combine (78) and (79) to obtain

$$\frac{\partial\mathbf{S}_l(z,t_l)}{\partial z}=\frac{8}{9}\gamma\exp(-2gz)\times\sum_{m=1}^{N}P_m(z=0)$$
$$\times p_m(t_l-\beta''\omega_{ml}z)\mathbf{S}_m(z,t_l)\times\mathbf{S}_l(z,t_l) \quad (80)$$

where $\omega_{ml}=\omega_m-\omega_l$. We are assuming that the loss is the same in all wavelength channels and is a constant, g. We now focus on the special case in which there are two wavelength channels. The first of these wavelengths is the signal and the second is the probe. There is a pivot vector $\mathbf{S}_{\rm pivot}=\mathbf{S}_{\rm signal}+\mathbf{S}_{\rm probe}$, which is constant as a function of z. The signal and the probe rotate around the pivot when both contain a mark, and the rotation stops when either contains a space. We may write the rotation angle as

$$\Psi(z,t_l)=\frac{8}{9}\gamma P_{\rm pivot}\int_0^z\exp(-2gz)p(t_l-\beta''\omega_{ml}z)dz \quad (81)$$

where $P_{\rm pivot}=(P_{\rm pump}^2+P_{\rm probe}^2+2P_{\rm pump}P_{\rm probe}\cos\theta)^{1/2}$, and $\cos\theta$ is the angle between $\mathbf{S}_{\rm pump}$ and $\mathbf{S}_{\rm signal}$. We now define the time-averaged rotation angle $\langle\Psi(t_l)\rangle=(8/9)\gamma P_{\rm pivot}L_{\rm eff}/4g$, which holds for nonreturn-to-zero signals for which $\langle p(t_l-\beta''\omega_{ml}z)\rangle=1/2$. For other signal formats, this average would be different. We also define the difference rotation angle $\Delta\Psi(t_l)=\Psi(t_l)-\langle\Psi(t_l)\rangle$, the power ratio $\mathrm{PR}=P_{\rm pump}/P_{\rm signal}$, and the relative angle between the signal and the pump $\theta_s=\theta-\tan^{-1}[\sin\theta/(\mathrm{PR}+\cos\theta)]$. We then find that $\mathbf{S}_{\rm signal}(z,t_l)=(\sin\theta_s\cos\Delta\Psi,\sin\theta_s\sin\Delta\Psi,\cos\theta_s)$, from which we conclude that the degree of polarization (DOP) at $z=L$, averaged over time τ, is given by

$$\mathrm{DOP}=|\langle\mathbf{S}_{\rm signal}(L,\tau)\rangle|$$
$$=\left[1-\sin^2\theta_s\left(1-\langle\cos\Delta\Psi\rangle^2-\langle\sin\Delta\Psi\rangle^2\right)\right]^{1/2}. \quad (82)$$

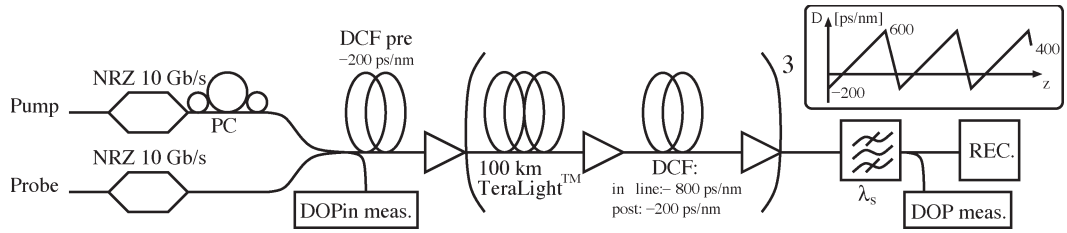

Fig. 8. Experimental setup for measuring the output probe DOP versus the input polarization angle θ. The inset shows the dispersion map. (Originally from [28]).

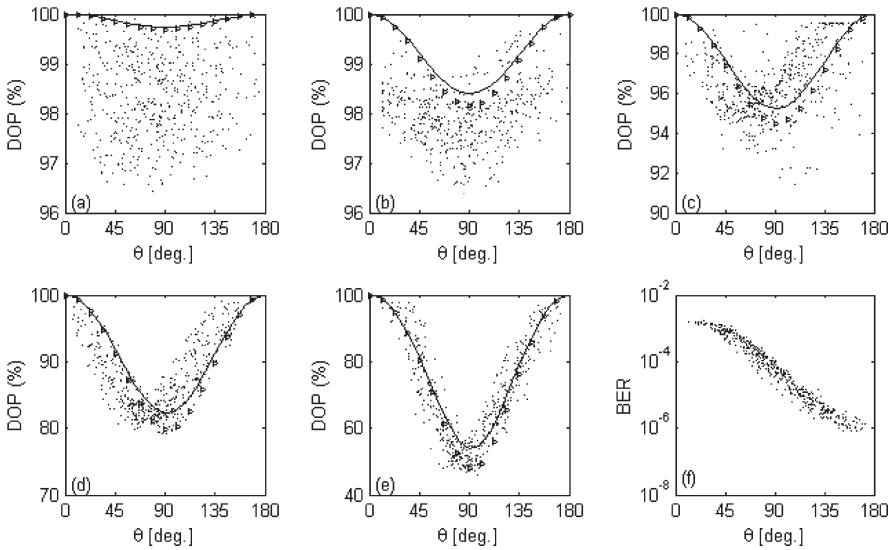

Fig. 9. Output probe DOP after three spans versus the input polarization angle θ for several pump-probe power ratios (PRs); measurements (dots), simulations (triangles), and theory (lines). (Originally from [28]). (a) PR $= 0$ dB, (b) PR $= 3.9$ dB, (c) PR $= 6.3$ dB, (d) PR $= 9.3$ dB, (e) PR $= 11.8$ dB, and (f) PR $= 11.8$ dB.

While it is not difficult in principle to solve (82) numerically, Vannucci *et al.* [28] obtain a closed-form expression for the DOP by making the following approximations: They first focus on periodic sequences of k marks followed by k spaces. Assuming that the contribution of this sequence is dominated by its lowest harmonic, they find that

$$\Delta\Psi(t) = \frac{16}{9\pi}\gamma P_{\text{pivot}} \frac{\sin(\pi t/kT)}{[4g^2 + (\beta'' \omega_{lm} L)^2]^{1/2}} \tag{83}$$

from which one concludes

$$\begin{aligned} \langle \cos \Delta\Psi \rangle &= J_0 \left(\frac{(16/9\pi)\gamma P_{\text{pivot}}}{[4g^2 + (\beta'' \omega_{lm} L)^2]^{1/2}} \right) \\ \langle \sin \Delta\Psi \rangle &= 0 \end{aligned} \tag{84}$$

where J_0 is the zeroth-order Bessel function. Noting that the sequence of k consecutive marks or spaces occurs with probability k^{-2}, they conclude that the DOP is given by

$$\begin{aligned} \text{DOP} = \Bigg(1 - \sin^2 \theta_s \Bigg\{ 1 - \Bigg[\sum_{k=1}^{\infty} \frac{1}{2^k} \\ \times J_0 \left(\frac{(16/9\pi)\gamma P_{\text{pivot}} N_{\text{span}}}{\left[4g^2 + (\beta'' \omega_{lm} L)^2\right]^{1/2}} \right) \Bigg]^2 \Bigg\} \Bigg)^{1/2} \end{aligned} \tag{85}$$

where N_{span} is the number of spans through which one repeats the process. Equation (85) can be evaluated numerically.

To observe the decrease in the DOP predicted in (85), Vannucci *et al.* [28] carried out measurements on the system shown in Fig. 8, with the dispersion map shown in the inset, which contained three spans of 100 km. The signal and the pump are spaced 0.8-nm apart, and the loss is 0.2 dB/km. They performed five sets of five hundred measurements of both the total input DOP and the DOP after filtering the signal channel, randomly changing the polarization controller each time. Fig. 9 shows a comparison of the experimental results, the results of a complete computer simulation, and the results from (85). The average signal power is 3 dBm, and the pump-probe power ratio (PR) is shown in each of the subfigures for each of the five cases. The experimental spread is due to amplifier noise. These results show clearly the reduction in the DOP that is induced by nonlinear polarization rotation.

Khosravani *et al.* [11] have carried out a set of simulations and experiments that demonstrate the deleterious effect of nonlinearity and PMD compensation when the PMD is large. Curiously, small PMD actually reduces the penalty. The reason is that when the PMD is large enough to randomize the polarization states between the channels, but not so large as to lead to penalties inside a single channel, then the PMD will reduce the nonlinear interactions and, hence, the nonlinear penalty. In their simulation studies, Khosravani *et al.* [11] focused on terrestrial

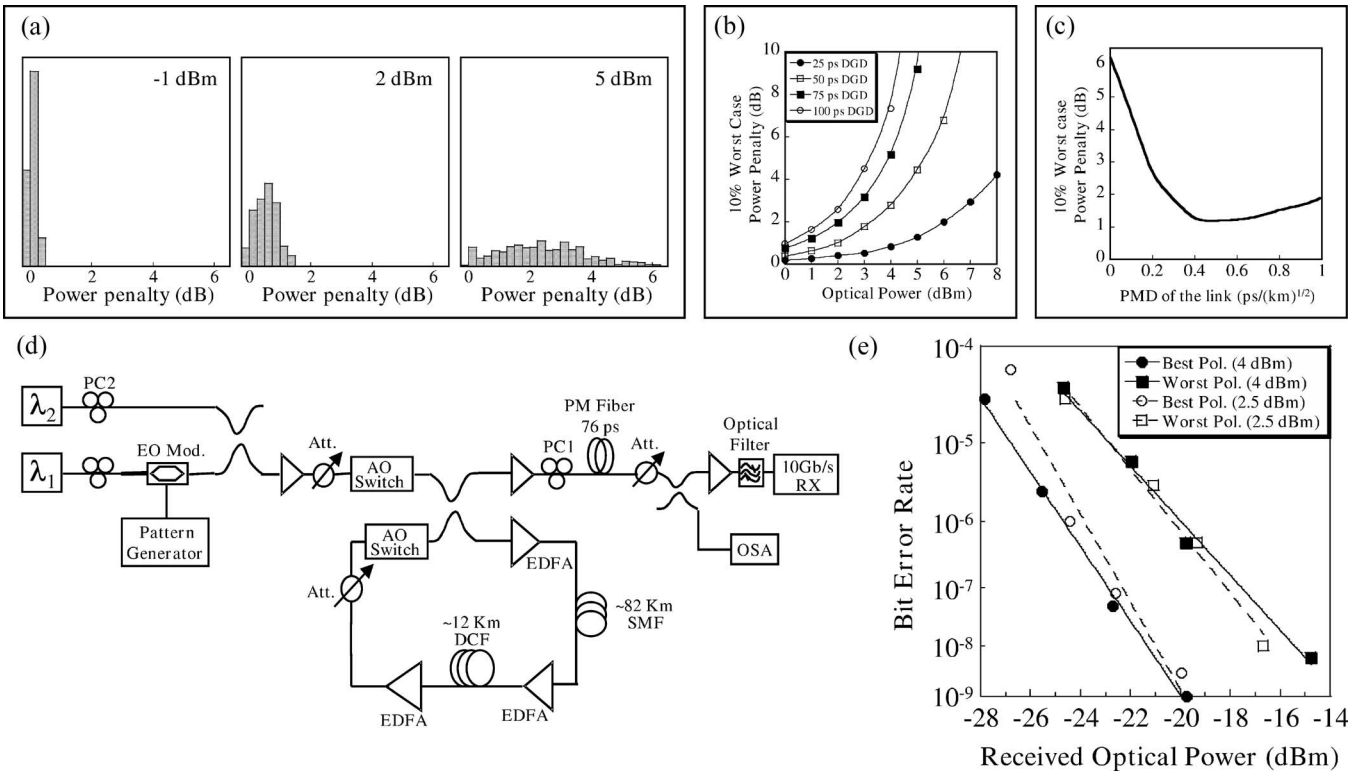

Fig. 10. SIMULATION–600-km transmission, two-channel 10-Gb/s system, 0.8-nm channel spacing. (a) Power penalty distributions due to different cross-phase-modulation-inducing optical powers (average power $= -1, 2,$ and 5 dBm) after the first-order PMD compensation, 50 ps initial differential group delay, (b) 10% worst case penalty after the first-order PMD compensation versus PMD of the link, 50-ps initial DGD, 5 dBm/channel; EXPERIMENT–six times recirculation in the loop, 0.8-nm channel spacing. (d) Experimental setup. (e) BER curves for the best and worst relative polarization between the two signals for 2.5- and 4-dBm input power on the cross-phase modulation-inducing channel to the SMF fiber. (Originally from [11]).

systems operating at 10 Gb/s. Each dispersion map period consisted of 85 km of SMF, 15 km of DCF, and two gain stages. The average input powers are set to 5 and -2 dBm for the SMF and DCF fibers, respectively. They considered six periods of the dispersion map, totaling 600 km. The channel spacing is 0.8 nm. The receiver model contains a low-pass filter with a 6-GHz cutoff frequency. The sampling time and decision threshold are optimized to account for PMD-induced bit pattern shifts. The amplified spontaneous emission noise is the dominant noise source; so, electrical receiver noise is ignored.

For the experiment, Khosravani *et al.* [11] used an optical fiber recirculating loop consisting of $\sim$ 82 km of SMF and $\sim$ 12 km of DCF. The signal passed through a single-section PMD compensator with $\sim$ 76 ps of differential group delay after six passes through the loop.

To evaluate the effects of nonlinear channel crosstalk on PMD compensation—due primarily to cross-phase modulation—they considered a simple two-channel system. The first channel contains a random 64-bit signal, with 50 ps of differential group delay applied to its two orthogonal components, and is transmitted over 600 km of low-PMD fiber (0.1 ps/km$^{1/2}$). The first-order PMD compensation is used at the end of the transmission link. The second channel is used to induce crosstalk, and it consists of a long series of marks followed by a long series of spaces, corresponding to the worst case pattern.

Fig. 10(a) shows the power penalty distributions for a two-channel system with different optical powers in the crosstalk-inducing channel. Fig. 10(b) shows the 10% worst-case penalty for different initial differential group delays and different crosstalk-inducing optical powers. These results show that although the initial differential group delay can be compensated after transmission in a system with cross-phase modulation, this compensation is no longer possible once crosstalk becomes important due to the uncertainty in principal states of polarization. Hence, initial values of the differential group delay are more susceptible to the uncertainty in the principal states of polarization, as the small deviations in the principal states of polarization lead to higher penalties after compensation. Average powers as low as 3 dBm can cause severe penalties after the first-order PMD compensation.

If the PMD of the link is not small, the different states of polarization of the wavelength channels become uncorrelated and change quickly over a short distance, which results in an averaging of the nonlinear effects and reduces the effect of crosstalk-induced variations of the principal states of polarization. Fig. 10(c) shows the power penalty caused by crosstalk-induced variations in a signal after the first-order PMD compensation. The penalty initially decreases as the PMD increases; however, as the PMD continues to increase, penalties that cannot be compensated due to the higher order PMD lead to an increase in the penalty.

In order to support the simulation results, Khosravani *et al.* [11] set up the experiment shown in Fig. 10(d). They transmitted two signals—one with modulated data and the other with a continuous wave. By adjusting the polarization controller

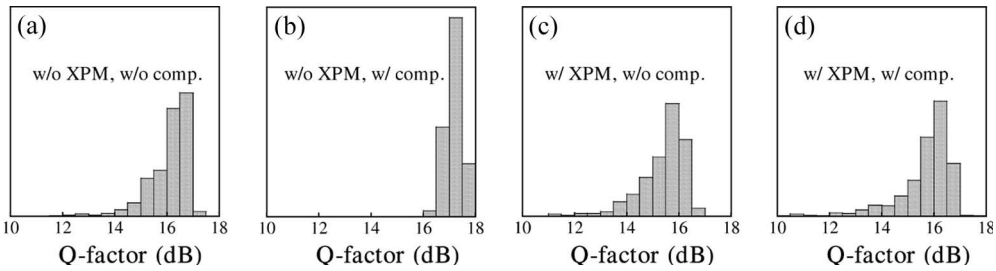

Fig. 11. Q-factor distribution for a 10-Gb/s signal (eight channels) after 600-km transmission. The first 100 km of the link has a high PMD of 3 ps/km$^{1/2}$, and the remaining 500 km has a low PMD of 0.1 ps/km$^{1/2}$. (a) No cross-phase modulation, before PMD compensation. (b) No cross-phase modulation, after PMD compensation. (c) With cross-phase modulation, after PMD compensation. (Originally from [11]).

(PC1) before the polarization-maintaining (PM) fiber, they first optimized the bit error rate (BER). Then, they changed the relative polarization states between the two optical signals by changing another polarization controller (PC2) in order to obtain the worst performance. Fig. 10(e) shows the BER curves for different optical powers in the crosstalk-inducing channel. A significant change in the system performance is apparent.

Fig. 11 shows the Q-factor distribution for an eight-channel system with 3 dBm per channel optical power, before and after the first-order PMD compensation. Again, it is assumed that all channels but one consist of a long series of marks followed by a long series of spaces, to simulate worst-case crosstalk effects. The first 100 km of the link is assumed to have a high PMD of 3 ps/km$^{1/2}$. The crosstalk had little effect on the Q-factor distribution without first-order compensation, but it caused significant additional penalties (> 4 dB) after compensation.

It is possible to greatly reduce or even eliminate this effect by avoiding long strings of marks in one polarization state. Pan *et al.* [27] demonstrated the effectiveness of this technique by replacing each mark in a single polarization state with a mark that is split into two polarizations. Many other variants can be imagined. Another possibility is to create a Manchester code in which a combination of $\hat{x}$-$\hat{y}$ polarizations indicates a mark and a combination of $\hat{y}$-$\hat{x}$ polarizations indicate a space, where $\hat{x}$ and $\hat{y}$ indicate two arbitrary orthogonal polarization states. This code has the additional advantage over the Pan *et al.* code of having constant power, which would also serve to reduce nonpolarization-dependent crosstalk-induced distortion. This Manchester code shares with the code of Pan *et al.* the defect that it requires doubling the transmission rate for a fixed data rate. Line codes exist that can substantially decrease the length of a string in a single polarization state while not substantially increasing the bandwidth. For example, one could create the polarization equivalent of a sliding window criterion code [50], in which the marks in that code are replaced by transmission in the $\hat{x}$-polarization and spaces in that code are replaced by transmission in the $\hat{y}$-polarization.

IV. Conclusion

In this paper, we have reviewed the interaction of the Kerr nonlinearity and PMD.

In Section II, we presented a derivation of the Mankov-PMD equation based on multiple-length-scale techniques. The length scales in optical fibers divide naturally into three different groups. The shortest group includes the wavelength of light and the fiber core diameter. At this length scale, one must use the complete set of Maxwell's equations to determine the light evolution. The intermediate group includes the birefringent beat length and the autocorrelation length for the orientation of the axes of the birefringence. At this length scale, we average over the rapid variations of the field amplitudes over one wavelength and the transverse variations to obtain the coupled nonlinear Schrödinger equation, which is the governing equation for the light evolution on this length scale. The longest group includes the nonlinear, dispersive, and PMD lengths. It is the length scale of interest for optical fiber communications systems. At this length scale, we average over the rapidly varying birefringence to obtain the Manakov-PMD equation, which is the governing equation for light evolution on this longest length scale. When the signal is initially in a single polarization state and the PMD may be neglected, then the light evolution is governed by the scalar nonlinear Schrödinger equation. In this derivation, we focused on elucidating common misconceptions and pitfalls, rather than on mathematical rigor.

Despite the simplicity of the Manakov-PMD equation and the transparent physical meaning of all its terms, the behavior that it predicts can be remarkably complex. In Section III, we present two examples that illustrate this complexity. The first example concerns the interaction of nonlinearity and PMD. It has been known since the early work of Poole and Giles that higher order PMD in combination with a frequency chirp can lead to a pulse compression during transmission, which in turn results in an increased eye opening after signal detection in the receiver. This PMD improvement can occur regardless of the physical origin of the chirp—dispersion, the transmitter, or nonlinearity. However, in the case of the Kerr nonlinearity, the chirp and hence the effect become power dependent. Thus, in the presence of significant PMD, it is possible to observe an improvement in the eye opening as the power increases. This effect has been studied for both nonreturn-to-zero and return-to-zero modulation formats and can be statistically significant.

The second example concerns the interaction of nonlinear polarization rotation in WDM systems with PMD compensators. When two pulses at different wavelengths interact, the Kerr nonlinearity induces a change in the polarization states of both unless both pulses are in the same or orthogonal

polarization states. In an ON–OFF-keyed system in which each pulse corresponds to a mark, the amount of change that one pulse undergoes depends on how many pulses with which it interacts in the other channels, which varies from pulse to pulse. Thus, pulses in a channel that are initially all in the same polarization state will end up in different polarization states, which in turn means that principal states of polarization change rapidly over time. Under these circumstances, PMD compensation schemes that rely on the slow variation of the principal states of polarization will prove ineffective. However, the use of a properly chosen line code can greatly mitigate this problem. The line code should be chosen so that every pulse in one WDM channel interacts with approximately the same number of pulses in the other channels.

The physics contained in the Manakov-PMD equation is not only of practical importance in optical fiber communications systems, but it is extraordinarily rich. We have little doubt that it will remain the focus of much future work in the years to come.

ACKNOWLEDGMENT

This work has been supported, over the years, by the Air Force Office of Scientific Research, the U.S. Department of Energy, and the National Science Foundation. Additional support was received from the Laboratory for Physical Sciences in College Park, MD, from the Laboratory for Telecommunications Sciences in Adelphi, MD, and from the Naval Research Laboratory in Washington, DC. The authors would like to thank their colleagues from these three DoD laboratories for many useful interactions, as well as for their help in obtaining financial support. They would also like to thank A. Bononi for permission to include Figs. 8 and 9 and A. Willner (AW) for permission to include Figs. 10 and 11. This tutorial article had its genesis in a special issue on the subject of polarization mode dispersion that appeared in the *Journal of Lightwave Technology* (April 2004). This special issue was edited by D. Chowdhury, M. Karlsson, and L. Nelson, who invited C. Menyuk (CM) to submit an invited paper on the subject of this tutorial. CM immediately asked B. Marks (BM) and AW to coauthor this article, and the latter two agreed. The authors are grateful for the original invitation. As CM proceeded to write the first draft of this article, it became apparent that it would not be finished in time for the special issue and that it was more appropriate as a tutorial than an invited article. After some hesitation, the authors concluded that the *Journal of Lightwave Technology* remained the most appropriate venue for this article. At that point, AW felt—despite an offer from M. Karlsson to edit this manuscript, for which the authors are grateful—that his responsibilities as an Editor of the *Journal of Lightwave Technology* were incompatible with his continued coauthorship of this article. CM and BM reluctantly acceded to his request to withdraw as a coauthor. They note, however, that AW made many important contributions to the final article. Whether these contributions are viewed as collegial or editorial, the authors are grateful that they had the opportunity to benefit from these contributions during this article's creation.

REFERENCES

[1] G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego, CA: Academic, 2001.

[2] C. D. Poole and J. Nagel, "Polarization effects in lightwave systems," in *Optical Fiber Telecommunications IIIA*, I. P. Kaminow and T. L. Koch, Eds. San Diego, CA: Academic, 1997, ch. 6, pp. 114–161.

[3] H. Kogelnik, R. M. Jopson, and L. E. Nelson, "Polarization-mode dispersion," in *Optical Fiber Telecommunications IVB*, I. P. Kaminow and T. Li, Eds. San Diego, CA: Academic, 2002, ch. 15, pp. 725–861.

[4] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*. New York: Springer-Verlag, 1992.

[5] P. D. Maker and R. W. Terhune, "Study of optical effects due to an induced polarization third order in the electric field strength," *Phys. Rev.*, vol. 137, no. 3A, pp. A801–A818, Feb. 1965.

[6] L. F. Mollenauer, J. P. Gordon, and F. Heisman, "Polarization scattering by soliton-soliton collisions," *Opt. Lett.*, vol. 20, no. 20, pp. 2060–2062, Oct. 1995.

[7] M. N. Islam, *Ultrafast Fiber Switching and Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[8] E. Ibragimov, C. R. Menyuk, and W. L. Kath, "PMD-induced reduction of nonlinear penalties in terrestrial optical fiber transmission," in *Proc. OFC*, 2000, pp. 195–197, Paper WL3.

[9] L. Möller, Y. Su, G. Raybon, S. Chandrasekhar, and L. L. Buhl, "Penalty interference of nonlinear intra-channel effects and PMD in ultra-high-speed TDM systems," *Electron. Lett.*, vol. 38, no. 6, pp. 281–283, Mar. 2002.

[10] B. S. Marks and C. R. Menyuk, "Polarization mode dispersion enhancement of nonlinear propagation," presented at the European Conference on Optical Communication (ECOC), Rimini, Italy, 2003, Paper Mo3.7.1.

[11] R. Khosravani, Y. Xie, L.-S. Yan, A. E. Willner, and C. R. Menyuk, "Limitations to first-order PMD compensation in WDM systems due to XPM-induced PSP changes," presented at the Optical Fiber Communication (OFC), Anaheim, CA, 2001, Paper WAA5.

[12] C. R. Menyuk, "Pulse propagation in an elliptically birefringent Kerr medium," *IEEE J. Quantum Electron.*, vol. 25, no. 12, pp. 2674–2682, Dec. 1989.

[13] R. H. Stolen, J. Botineau, and A. Ashkin, "Intensity discrimination of optical pulses with birefringent fibers," *Opt. Lett.*, vol. 7, no. 10, pp. 512–514, Oct. 1982.

[14] J. Botineau and R. H. Stolen, "Effect of polarization on spectral broadening in optical fibers," *J. Opt. Soc. Amer.*, vol. 72, no. 12, pp. 1592–1596, Dec. 1982.

[15] W. K. Burns, R. P. Moeller, and C. Chen, "Depolarization in a single-mode optical fiber," *J. Lightw. Technol.*, vol. LT-1, no. 1, pp. 44–49, 1983.

[16] C. D. Poole and R. E. Wagner, "Phenomenological approach to polarisation dispersion in long single-mode fibres," *Electron. Lett.*, vol. 22, no. 19, pp. 1029–1030, Sep. 1986.

[17] C. R. Menyuk, "Stability of solitons in birefringent optical fibers. I: Equal propagation amplitudes," *Opt. Lett.*, vol. 12, no. 8, pp. 614–616, Aug. 1987.

[18] P. K. A. Wai, C. R. Menyuk, and H. H. Chen, "Stability of solitons in randomly varying birefringent fibers," *Opt. Lett.*, vol. 16, no. 16, pp. 1231–1233, Aug. 1991.

[19] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*. Philadelphia, PA: SIAM, 1981.

[20] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*. Oxford, U.K.: Clarendon, 1995.

[21] C. R. Menyuk, "Soliton robustness in optical fibers," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 10, no. 9, pp. 1585–1591, Sep. 1993.

[22] ——, "Nonlinear pulse propagation in birefringent optical fibers," *IEEE J. Quantum Electron.*, vol. QE-23, no. 2, pp. 174–176, Feb. 1987.

[23] ——, "Stability of solitons in birefringent optical fibers. II: Arbitrary amplitudes," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 5, no. 2, pp. 392–402, Feb. 1988.

[24] J. Kevorkian and J. D. Cole, *Multiple Scale and Singular Perturbation Methods*. New York: Springer-Verlag, 1996.

[25] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I. Asymptotic Methods and Perturbation Theory*. New York: Springer-Verlag, 1999.

[26] C. R. Menyuk, "Application of multiple-length-scale methods to the study of optical fiber transmission," *J. Eng. Math.*, vol. 36, no. 1/2, pp. 113–136, Aug. 1999.

[27] Z. Pan, Q. Yu, and A. E. Willner, "Fast XPM-induced polarization-state fluctuations in WDM systems and their mitigation," in *Proc. OFC*, 2002, pp. 379–381, Paper ThA7.

[28] A. Vannucci, A. Bononi, A. Orlandini, E. Corbel, J. Thiéry, S. Lanne, and S. Bigo, "A simple formula for the degree of polarization degraded by XPM and its experimental validation," in *Proc. OFC*, 2003, pp. 498–499, Paper ThJ1.
[29] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proc. Natl. Acad. Sci. USA*, vol. 97, no. 9, pp. 4541–4550, Apr. 2000.
[30] C. R. Menyuk, B. S. Marks, I. T. Lima, Jr., J. Zweck, Y. Sun, G. M. Carter, and D. Wang, "Polarization effects in long-haul undersea systems," in *Undersea Fiber Communication Systems*, J. Chesnoy, Ed. San Diego, CA: Academic, 2002, ch. 7, pp. 270–305.
[31] A. W. Snyder and J. D. Love, *Optical Waveguide Theory*. New York: Chapman and Hall, 1983.
[32] Y. Kodama, "Optical solitons in a monomode fiber," *J. Stat. Phys.*, vol. 39, no. 5/6, pp. 597–614, 1985.
[33] U. Österberg and W. Margulis, "Experimental studies on efficient frequency doubling in glass optical fibers," *Opt. Lett.*, vol. 12, no. 1, pp. 57–59, Jan. 1987.
[34] R. H. Stolen, J. P. Gordon, W. J. Tomlinson, and H. A. Haus, "Raman response of silica core fibers," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 6, no. 6, pp. 1159–1166, Jun. 1989.
[35] E. L. Buckland and J. W. Boyd, "Electrostrictive contribution to the intensity-dependent refractive index of optical fibers," *Opt. Lett.*, vol. 21, no. 15, pp. 1117–1119, Aug. 1996.
[36] I. P. Kaminow, "Polarization in optical fibers," *IEEE J. Quantum Electron.*, vol. QE-17, no. 1, pp. 15–22, Jan. 1981.
[37] A. Galtarossa and L. Palmieri, "Spatially resolved PMD measurements," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1103–1115, Apr. 2004.
[38] C. R. Menyuk and P. K. A. Wai, "Elimination of nonlinear polarization rotation in twisted fibers," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 11, no. 7, pp. 1305–1309, Jul. 1994.
[39] T. Wanner, B. S. Marks, C. R. Menyuk, and J. Zweck, "Polarization decorrelation in optical fibers with randomly varying elliptical birefringence," *Opt. Lett.*, vol. 28, no. 19, pp. 1799–1801, Oct. 2003.
[40] M. Karlsson, "Polarization mode dispersion-induced pulse broadening in optical fibers," *Opt. Lett.*, vol. 23, no. 9, pp. 688–690, May 1998.
[41] S. V. Chernikov and J. R. Taylor, "Measurement of normalization factor of n_2 for random polarization in fiber," *Opt. Lett.*, vol. 21, no. 19, pp. 1559–1561, Oct. 1996.
[42] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1746, 1997.
[43] M. F. Arend, M. L. Dennis, I. N. Duling, III, E. A. Golovchenko, A. N. Pilipetskii, and C. R. Menyuk, "Nonlinear-optical loop mirror demutiplexer using a random birefringence fiber. Comparisons between theory and experiment," *Opt. Lett.*, vol. 22, no. 12, pp. 886–888, Jun. 1997.
[44] L. F. Mollenauer, S. G. Evangelides, Jr., and H. A. Haus, "Long-distance soliton propagation using lumped amplifiers and dispersion shifted fiber," *J. Lightw. Technol.*, vol. 9, no. 2, pp. 194–197, Feb. 1991.
[45] K. J. Blow and N. J. Doran, "Average soliton dynamics and the operation of soliton systems with lumped amplifiers," *IEEE Photon. Technol. Lett.*, vol. 3, no. 4, pp. 369–371, Apr. 1991.
[46] C. D. Poole and C. R. Giles, "Polarization-dependent pulse compression and broadening due to polarization dispersion in dispersion-shifted fiber," *Opt. Lett.*, vol. 13, no. 2, pp. 155–157, Feb. 1988.
[47] F. Bruyère, "Impact of first- and second-order PMD in optical digital transmission systems," *Opt. Fiber Technol.*, vol. 2, no. 3, pp. 269–280, Jul. 1996.
[48] H. Bülow, "System outage probability due to first- and second-order PMD," *IEEE Photon. Technol. Lett.*, vol. 10, no. 5, pp. 696–698, May 1998.
[49] D. Wang and C. R. Menyuk, "Calculation of penalties due to polarization effects in a long-haul WDM system using a Stokes parameter model," *J. Lightw. Technol.*, vol. 19, no. 4, pp. 487–494, Apr. 2001.
[50] Y. Cai, T. Adalı, C. R. Menyuk, and J. M. Morris, "Sliding window criterion codes and concatenation scheme for mitigating timing-jitter-induced errors in WDM fiber transmissions," *J. Lightw. Technol.*, vol. 20, no. 2, pp. 201–212, Feb. 2002.

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

He was a Research Associate at the University of Maryland, College Park, and at Science Applications International Corporation, McLean, VA. In 1986, he joined the University of Maryland Baltimore County (UMBC) as an Associate Professor in the Department of Electrical Engineering, in which he was a Founding Member. In 1993, he was promoted to Presidential Research Professor and was on partial leave from Fall 1996 to Fall 2002. From 1996 to 2001, he worked part-time for the Department of Defense (DoD), where he codirected the Optical Networking program at the DoD Laboratory for Telecommunications Sciences, Adelphi, MD, from 1999 to 2001. From 2001 to 2002, he was a Chief Scientist at PhotonEx Corporation. He has authored or coauthored more than 190 archival journal publications as well as numerous other publications and presentations and has edited three books. The equations and algorithms that he and his research group at UMBC have developed to model optical fiber systems are used extensively in the telecommunications and photonics industry. For the last 18 years, his primary research interests have been theoretical and computational studies of lasers, nonlinear optics, and fiber optic communications.

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

Brian S. Marks (M'01) was born in New Jersey in 1973. He grew up in Greenville, NC, attended the North Carolina (NC) School of Science and Mathematics, Durham, NC, and received the B.S. degrees in mathematics and physics from North Carolina State University, Raleigh, in 1995 and the Ph.D. degree in applied mathematics from Northwestern University, Evanston, IL, in 2000, where his dissertation focused on the optimization of dispersion maps for use with dispersion managed solitons.

After leaving Northwestern, he became a Research Associate and then a Research Assistant Professor at University of Maryland Baltimore County (UMBC), Baltimore. At UMBC, his research interests included polarization mode dispersion (PMD), analog fiber-optic communications systems, microstructure fiber modeling, and fiber amplifier modeling. In 2005, he joined the Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN. He is currently studying signal and image processing in neuroimaging.