Journal of Engineering Mathematics **36:** 113–136, 1999.

Application of multiple-length-scale methods to the study of optical fiber transmission

C. R. MENYUK
Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, U.S.A.
Laboratory for Telecommunications Sciences, c/o: USARL, Bldg. 601, Rm. 131, 2800 Powder Mill Road, Adelphi, MD 20783-1197, U.S.A.

Received 20 February 1998; accepted in revised form 5 January 1999

Abstract. It is natural to apply multiple-length-scale methods to the study of optical-fiber transmission because the key length scales span 13 orders of magnitude and cluster in three main groups. At the lowest scale, corresponding to micrometers, the full set of Maxwell's equations should be used. At the intermediate scale, corresponding to the range from one centimeter to tens of meters, the coupled nonlinear Schrödinger equation should be used. Finally, at the longest length scale, corresponding to the range from tens to thousands of kilometers, the Manakov-PMD equation should be used, and, when polarization mode dispersion can be neglected and the fiber gain and loss can be averaged out, one arrives at the scalar nonlinear Schrödinger equation. As an illustrative example of multiple-scale-length techniques, the nonlinear Schrödinger equation will be derived, carefully taking into account the actual length scales that are important in optical-fiber transmission.

Key words: optical fibers, multiple length scales, nonlinearity, birefringence, communications.

1. Introduction

The history of nonlinear optics dates back nearly forty years to shortly after the invention of the laser [1, 2], while nonlinear fiber optics dates back more than twenty years to the invention of low-loss fibers [3, Chapter 1]. Given this long history, which has included a Nobel prize in Physics, it always surprises mathematicians who are newcomers to this field when they discover the relatively small impact that twentieth century nonlinear dynamics has had on nonlinear optics. Nearly all theoretical work until recently has been carried out by experimentalists 'on the fly', while conducting their experimental research and has been relatively simple in character. This situation existed because most optical nonlinearities have been weak and could be adequately studied by use of simple mode-coupling theories in which one first treats systems linearly using Fourier mode decompositions and then treats the nonlinearity as a weak coupling between the Fourier modes. Traditional texts on nonlinear optics such as Bloembergen's [1] and Shen's [2] begin by treating two-mode coupling, then three-mode coupling and four-mode coupling, at which point one has reached the end of the book! This theory can be algebraically complex, mostly because of complexities in the dielectric properties of the nonlinear materials, but there is little of conceptual interest to challenge the applied mathematician.

There are notable exceptions. It was recognized some time ago that self-phase modulation could not be treated by a simple mode-coupling approach [3, Chapter 4]. However, the theory is once again very simple from a conceptual standpoint. Solitons are a more important

exception [4, Chapter 1]. Hasegawa and Tappert [5] recognized over twenty years ago that with low-loss fibers and sufficiently high input powers, the nonlinearity would be strong and solitons would propagate in optical fibers. Here, there has certainly been sufficient scope for the applied mathematician, and a large amount of work has been done, not only on optical fiber solitons but also on solitons in other optical contexts. While much of this work has been done for experimentally unrealizable configurations [6] and solitons have yet to be used in commercial high-data-rate communications, this work has led to a much better understanding of how strong nonlinearity affects signal transmission in commercial optical fiber communication systems. Moreover, dispersion-managed solitons, which in contrast to standard solitons are only periodically stationary, are currently being considered for use in optical communication systems.

It has not been widely recognized in the applied-mathematics community that the invention of the erbium-doped fiber amplifier nearly a decade ago has greatly increased the potential scope for applications of nonlinear dynamics to optical-fiber communication problems [7, pp. 196–203]. Prior to the invention of the erbium-doped fiber amplifier, optical-fiber communication systems used repeaters spaced every twenty kilometers or so to completely regenerate the communication signal electronically. Thus, there was no opportunity for the nonlinearity to accumulate and become strong. Maximum data rates were under 1 Gbit/sec and only one channel at a time could be transmitted. By contrast, systems based on erbium-doped fiber amplifiers can transmit more than 100 Gbits/sec in a single wavelength channel. Moreover, the amplifiers have a wide bandwidth of 20 nm or more, so that it is possible to use wavelength division multiplexing (WDM) in which multiple wavelength channels propagate [8, pp. XV–XVI]. Systems are currently being studied that can transmit terabits/sec of information! However, the amplifiers only compensate for fiber attenuation, allowing the effects of chromatic dispersion, randomly varying birefringence, and the Kerr nonlinearity to all accumulate. Additionally, the amplifiers emit spontaneous emission noise, setting a lower limit on the signal intensity that is required for an acceptable signal-to-noise ratio. As a consequence, strong nonlinearity is inevitable in modern-day, optical-fiber communication systems!

It cannot be sufficiently stressed that strong nonlinearity is important – indeed, nearly inevitable – in modern-day optical-fiber communication systems regardless of the format used! Understanding the effects of nonlinearity is every bit as important in an NRZ (non-return to zero) system as it is in a soliton system. Thus, the use of the nonlinear Schrödinger equation and its extensions is becoming increasingly widespread in the engineering community, and simulation packages are commercially available for solving these equations and are being bought by major telephone service providers [9]. Understanding and coping with the effects of strong nonlinearity is crucially important in modern-day optical communication systems, and the techniques of nonlinear dynamics that have been developed over the last 200 years have a great role to play.

Multiple-scale methods are among the most important techniques that are available in the applied mathematician's arsenal. When Poincaré wrote *Les méthodes nouvelles de la mécanique céleste* [10] a century ago, they had already been in use for nearly a century to study planetary orbits. As a simple case, we can imagine a system that depends on two variables θ_1 and θ_2, so that

$$\frac{\mathrm{d}\theta_1}{\mathrm{d}t} = f(\theta_1, \theta_2), \qquad \frac{\mathrm{d}\theta_2}{\mathrm{d}t} = \varepsilon g(\theta_1, \theta_2), \tag{1}$$

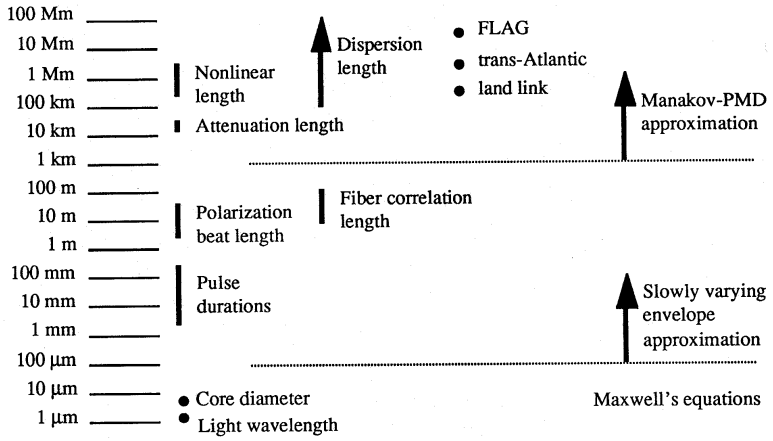

Figure 1. Illustration of the key length scales in optical fiber communication systems.

where ε indicates a small parameter so that θ_2 changes slowly compared to θ_1. For example, one could imagine the three-body system that consists of the sun, Jupiter, and one of Jupiter's moons. In this case, θ_1 designates the angle between the moon and Jupiter, while θ_2 designates the angle between Jupiter and the sun. To solve (1), one first holds θ_2 fixed while solving for θ_1; one then averages over θ_1 to solve for θ_2. We have thus decomposed a three-body problem, which is in general intractable, to two easily solved two-body problems. Multiple-scale techniques have been used throughout the twentieth century to solve a wide array of engineering problems [11].

A large range of length scales exists in optical communication systems, as shown in Figure 1. There are over 13 orders of magnitude that separate the smallest length scale, the wavelength of light (1·55 μm), from the largest scale which corresponds to FLAG (fiber loop around the globe), the most extensive communication system currently being built (23,000 km). These scales naturally cluster into three groups. The short scale, on the order of micrometers, corresponds to the wavelength of light and the core diameter. At this length scale, one should use the full set of Maxwell's equations to solve for the evolution. It is at this length scale that the dispersion relation $\beta(\omega)$ is determined. The intermediate length scale, on the order of meters, corresponds to the fiber beat length and to the fiber correlation length – the length scale on which the orientation of the fiber's axes of birefringence changes randomly. At this length scale, one should use the coupled nonlinear Schrödinger equation which is obtained by appropriately averaging over Maxwell's equations. The last length scale, on the order of tens of kilometers and more, corresponds to the length scales for fiber attenuation, chromatic dispersion, and the Kerr nonlinearity. At this length scale, one obtains the Manakov-PMD equation after averaging over the coupled nonlinear Schrödinger equation. When polarization-mode dispersion is negligible, higher-order dispersion is negligible, and the initial signal is in a single polarization state, one can average over the rapidly varying gain and loss to obtain the nonlinear Schrödinger equation.

It is remarkable that no derivation of the nonlinear Schrödinger equation that is valid for physically realistic optical fibers exists within the scientific literature. First, all the derivations that have been published in textbooks assume that optical fibers are perfectly circular. Not only is this assumption false, but, in fact, just the opposite is true. The magnitude of an effect is inversely proportional to its corresponding scale length, and thus relative to the Kerr effect

and chromatic dispersion, the birefringence must be considered large but rapidly varying. A conceptually correct derivation must take this essential fact into account. Second, even if a perfectly round fiber is assumed, most of the published derivations – with notable exceptions such as Kodama's elegant derivation [12] – contain contradictory assumptions and errors.

In this article, I will be studying the application of multiple length scale techniques to optical fiber transmission. Most of the issues treated in this paper can be found scattered throughout the literature; however, they have never been brought together in a unified context before and the logical connection between them exposed. In Section 2 of this paper, I will show how one goes from the short length scale in Figure 1 to the intermediate length scale by averaging over the rapid transverse variations and carrier oscillations. At the intermediate length scale, one obtains the coupled nonlinear Schrödinger equation. I previously derived this equation, neglecting the transverse fiber variations [6] and Kodama [12], for example, has derived the nonlinear Schrödinger equation keeping the transverse variations, but assuming a perfectly circular fiber. To my knowledge, the derivation presented here is the first really complete derivation of the coupled nonlinear Schrödinger equation. In Section 3, I average over the rapidly and randomly varying birefringence to obtain the Manakov-PMD equation. From there, I will show how one can average over the rapidly varying gain and loss and rapidly varying dispersion to obtain the nonlinear Schrödinger equation. This path from Maxwell's equations to the nonlinear Schrödinger equation is not only conceptually complete, but makes clear the limitations on the nonlinear Schrödinger equation that exist in real optical fibers.

Rederiving the nonlinear Schrödinger equation might seem pointless from a practical standpoint. After all, it is a well-established experimental fact that it works in at least certain contexts! Here, I appeal for support to Poincaré's statement in the Introduction to *Les méthodes nouvelles de la mécanique céleste* [10, pp. 1–5]:

> 'Le but final de la Mécanique céleste est de résoudre cette grande question de savoir si la loi de Newton explique à elle seule tous les phénomènes astronomiques; le seul moyen d'y parvenir est de faire des observations aussi précises que possibles et de les comparer ensuite aux résultats du calcul. Ce calcul ne peut être qu'approximatif et il ne servirait à rien, d'ailleurs, de calculer plus de décimales que les observations n'en peuvent faire connaître. Il est donc inutile de demander plus de précision qu'aux observations; mais on ne doit pas en demander moins.'[1]

Celestial mechanics had its origin in the very practical problem of determining the motion of the celestial bodies for the benefit of navigators, but Poincaré emphasized its value in determining the validity of Newton's law of gravitation. In this as in much else he proved prescient since the evidence for Einstein's general theory of relativity stemmed in part from the small deviations of Mercury's motions from the predictions of Newton's law. Of course, Poincaré had great respect for practical applications of mathematical theory. He was trained as a mine engineer and contributed to the understanding of practical problems throughout his distinguished career. At one point he even participated in debunking the claims of another mathematician who had 'proved' mathematically that Dreyfus was a liar [13, pp. 100–101]!

[1] The final goal of celestial mechanics is to resolve the great question as to whether Newton's law can explain by itself all the astronomical phenomena; the only way to arrive at the answer is to make observations that are as accurate as possible and to compare them to calculations. These calculations can only be approximate and, besides which, it serves no purpose to calculate more decimals than the observations allow one to know. It is thus pointless to ask for more accuracy than is available from the observations, but one should not ask for less. (author's translation)

Nonetheless, he held firm in his belief that the greatest value of his work was in the light that it shed on the fundamental foundations of the science of his day. It is my view that optical fibers are not only of evident practical value, but are also an unparalleled laboratory for nonlinear phenomena. Thus, their study has the potential to yield fundamental insights whose impact can only be imagined.

2. The coupled nonlinear Schrödinger equation

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

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r}, t)] + \frac{1}{\varepsilon_0 c^2} \frac{\partial^2 \mathbf{D}(\mathbf{r}, t)}{\partial t^2} = 0, \tag{2}$$

where the electric displacement is defined as

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{3}$$

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

$$\mathbf{P} = \mathbf{P}_{\mathrm{L}} + \mathbf{P}_{\mathrm{NL}}, \tag{4}$$

where

$$\mathbf{P}_{\mathrm{L}}(\mathbf{r}, t) = \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t_1 X_{\mathrm{L}}(\mathbf{r}, t - t_1) \cdot \mathbf{E}(\mathbf{r}, t_1), \tag{5a}$$

$$\begin{aligned} \mathbf{P}_{\mathrm{NL}}(\mathbf{r}, t) \;=\; & \delta_3 \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t_1 \int_{-\infty}^{t} \mathrm{d}t_2 \int_{-\infty}^{t} \mathrm{d}t_3 \\ & \times X_{\mathrm{NL}}(\mathbf{r}, t - t_1, t - t_2, t - t_3) \vdots \mathbf{E}(\mathbf{r}, t_1)\mathbf{E}(\mathbf{r}, t_2)\mathbf{E}(\mathbf{r}, t_3). \end{aligned} \tag{5b}$$

The quantities X_{L} and X_{NL} are the material dielectric response tensors. We are assuming that the response is local in space, consistent with the small size of the glass molecules relative to the wavelength of light, although we are also assuming that it is nonlocal in time, consistent with the existence of dispersion in glass. The time integrals end at t, not ∞, to be consistent with causality. Alternatively, we can demand that X_{L} and X_{NL} equal zero when any of their temporal arguments is negative. The parameter δ_3 is an ordering parameter that is used to indicate that nonlinearity is small so that its corresponding length scale is long, being part of the longest or third group of length scales shown in Figure 1.

We now specialize to the optical fiber geometry by letting z correspond to distance along the fiber and $\mathbf{r}_\perp = (x, y)$ correspond to the transverse coordinates. To begin the analysis, we write

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{F}(\delta_1 z, \mathbf{r}_\perp, \delta_1 t, \omega_0) \exp\{\mathrm{i}[\beta(\omega_0) z - \omega_0 t]\} + \text{complex conjugate}, \tag{6}$$

where ω_0 is the carrier frequency of the signal, also referred to as the central frequency. We will shortly choose the central wavenumber $\beta(\omega_0)$ so that $\mathbf{F}$ is slowly varying relative to the wavelength of light. The small parameter δ_1 is added to indicate that the variation of the

modal structure **F** in z and t occurs slowly relative to the light's wavelength and period. The corresponding length scale, labeled pulse durations in Figure 1, is on the order of a centimeter or more, corresponding to the roughly 50 ps or more rise and fall time in an NRZ signal or the roughly 20–50 ps duration of a soliton signal. The dielectric tensor $X_{\rm L}(\mathbf{r}, \tau) = X_{\rm L}(\delta_2 z, \mathbf{r}_\perp, \tau)$ changes on a slow length scale, on the order of tens of meters and more, shown in Figure 1 as the length scale for the beat length and the fiber autocorrelation length. We designate this length scale with the small parameter δ_2. In our analysis, we will assume $\delta_1 \gg \delta_2 \gg \delta_3$, corresponding inversely to the distinct length scales just noted. The change of the dielectric tensor on the length scale of meters corresponding to the small parameter δ_2 can induce changes in $\beta(\omega_0)$, so that more generally

$$\beta(\omega_0)z \to \int_0^z \beta(\delta_2 z_1, \omega_0)\,\mathrm{d}z_1, \tag{7}$$

however, the magnitude of these variations is typically one part in $10^6 - 10^7$ and leads to no observable effects. Note that δ_1 and δ_2 both correspond to length scales within the middle or second group of length scales, but we are taking advantage of the separation of scales within this middle group to specify two parameters. Later in this paper, we will also subdivide the third length scale, taking advantage of the discrepancy between the length scale corresponding to the gain and loss and the length scales corresponding to nonlinearity and dispersion.

To make further progress, we will specialize the dielectric responses so that they correspond to the optical fiber physics. To begin with, the nonlinear response can be considered isotropic so that $X_{\rm NL} \to \chi_{\rm NL}$, a scalar, and (5b) becomes

$$\begin{aligned}\mathbf{P}_{\rm NL}(\mathbf{r}, t) \;=\; & \delta_3 \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t_1 \int_{-\infty}^{t} \mathrm{d}t_2 \int_{-\infty}^{t} \mathrm{d}t_3 \\ & \times \chi_{\rm NL}(\delta_2 z, \mathbf{r}_\perp, t - t_1, t - t_2, t - t_3)[\mathbf{E}(\mathbf{r}, t_1) \cdot \mathbf{E}(\mathbf{r}, t_2)]\mathbf{E}(\mathbf{r}, t_3).\end{aligned} \tag{8}$$

While there may be an anisotropic contribution to the nonlinear response, it would presumably be small, and there is no experimental evidence that it plays any role. By contrast, anisotropies in the linear response along with axial asymmetry play a crucial role since they lead to fiber birefringence. [The distinction between anisotropy and asymmetry is that asymmetry implies that the dielectric tensor is not rotationally symmetric, varying as a function of angle around the axis of the fiber. Anisotropy implies that the dielectric tensor itself has a preferred orientation. Both are present in real fibers, and it is a matter of some debate which contributes more to the fiber birefringence.] We therefore need to keep the complete linear response, and we write

$$X_{\rm L}(\delta_2 z, \mathbf{r}_\perp, \tau) = \chi_{\rm L}(\delta_2 z, \rho, \tau) + \delta_2 \Delta X_{\rm L}(\delta_2 z, \mathbf{r}, \tau), \tag{9}$$

where $\Delta X_{\rm L}$ contains both the dielectric anisotropies and asymmetries; *i.e.*, if one expresses $\mathbf{r} = (\rho, \theta)$, then one finds that

$$\text{Trace}\left(\int_0^{2\pi} \Delta X_{\rm L}\,\mathrm{d}\theta\right) = 0 \quad \text{at all } \rho.$$

In the first stage of the multiple-length-scale expansion, we ignore contributions of order δ_2 and δ_3, keeping only the contributions of order 1 and powers of δ_1. The smaller contributions

will be restored later. From Maxwell's equation, Equation (2), it follows that

$$-\frac{\partial^2 \mathbf{E}_\perp(\mathbf{r},t)}{\partial z^2}+\hat{\mathbf{e}}_z\frac{\partial}{\partial z}\nabla_\perp\cdot\mathbf{E}_\perp(\mathbf{r},t)+\frac{\partial}{\partial z}\nabla_\perp E_z(\mathbf{r},t)+\nabla_\perp\times[\nabla_\perp\times\mathbf{E}(\mathbf{r},t)]$$
$$+\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left[\mathbf{E}(\mathbf{r},t)+\int_{-\infty}^{t}\chi_{\mathrm{L}}(\mathbf{r},t-t_1)\mathbf{E}(\mathbf{r},t_1)\,\mathrm{d}t_1\right]\equiv\hat{\mathcal{L}}(\mathbf{E})=0, \tag{10}$$

where $\mathbf{E}=(\mathbf{E}_\perp, E_z)$ with $\mathbf{E}_\perp=(E_x, E_y)$. Inserting (6) into (10) and focusing on the positive frequency component of the electric field, we obtain to lowest order in δ_1,

$$\beta^2\mathbf{F}_\perp+\mathrm{i}\beta\hat{\mathbf{e}}_z\nabla_\perp\cdot\mathbf{F}_\perp+\mathrm{i}\beta\nabla_\perp F_z+\nabla_\perp\times(\nabla_\perp\times\mathbf{F})$$
$$-\frac{\omega_0^2}{c^2}[1+\tilde{\chi}_{\mathrm{L}}(\rho,\omega_0)]\mathbf{F}\equiv\hat{\mathcal{L}}_0(\mathbf{F})=0, \tag{11}$$

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

$$\tilde{\chi}_{\mathrm{L}}(\rho,\omega)=\int_0^\infty\chi_{\mathrm{L}}(\rho,\tau)\exp(\mathrm{i}\,\omega\tau)\,\mathrm{d}\tau \tag{12}$$

is the Fourier transform of $\chi_{\mathrm{L}}(\rho,\tau)$, and the arguments of $\beta(\omega_0)$ and $\mathbf{F}(\mathbf{r}_\perp,\omega_0)$ have been suppressed. Equation (11) determines both $\beta(\omega_0)$ and $\mathbf{F}(\mathbf{r}_\perp,\omega_0)$ subject to the boundary condition that $\mathbf{F}(\rho,\omega_0)\to 0$ as $\rho\to\infty$. The solutions to (11) can all be written as a superposition of solutions of the form

$$\mathbf{R}_{\mathrm{o}}^{(m)}(\mathbf{r}_\perp,\omega_0) = \hat{\mathbf{e}}_\rho R_\rho^{(m)}(\rho)\sin(m\theta)+\hat{\mathbf{e}}_\theta R_\theta^{(m)}(\rho)\cos(m\theta)$$
$$+\mathrm{i}\hat{\mathbf{e}}_z R_z^{(m)}(\rho)\sin(m\theta), \tag{13a}$$

and

$$\mathbf{R}_{\mathrm{e}}^{(m)}(\mathbf{r}_\perp,\omega_0) = \hat{\mathbf{e}}_\rho R_\rho^{(m)}(\rho)\cos(m\theta)-\hat{\mathbf{e}}_\theta R_\theta^{(m)}(\rho)\sin(m\theta)$$
$$+\mathrm{i}\hat{\mathbf{e}}_z R_z^{(m)}(\rho)\cos(m\theta), \tag{13b}$$

where m is a nonnegative integer and the $R_j^{(m)}$ are all real. If the set $(\mathbf{F}_\perp, F_z;\beta)$ is a solution to (11), then so is the set $(-\mathbf{F}_\perp, F_z;-\beta)$, corresponding physically to the existence of both forward and backward propagating waves. In elementary derivations of the nonlinear Schrödinger equation, it is common to find that $\nabla\times(\nabla\times\mathbf{E})$ is replaced by $-\nabla^2\mathbf{E}+\nabla(\nabla\cdot\mathbf{E})$ and the second term simply dropped. However, as Kodama [12] has pointed out, the second term is larger in realistic cases than the terms due to birefringence, chromatic dispersion, and nonlinearity that we are keeping. Thus, its neglect must be carefully justified.

In single mode fibers, which are now used almost universally in high-data-rate optical-fiber communication systems, it is a physical fact that $\mathbf{F}$ is doubly-degenerate and there are no other propagating modes in the wavelength range that is used. In this case, $\mathbf{F}$ is a superposition of the two modes with the form shown in (13) and $m=1$. By contrast, in multimode fibers there are typically many modes at the same frequency, corresponding to a range of allowed m-values, including circularly symmetric modes that are nondegenerate, corresponding to $m=0$. The variation of χ_{L} is less than 1% between the core and the cladding; so, the modes in a single mode fiber are nearly plane waves with a radially modulated amplitude, and $R_z^{(m)}$ is small

compared to $R_\rho^{(m)}$ and $R_\theta^{(m)}$ by a factor that approximately equals the square root of the index difference. In special cases, as for example when both the core and cladding have a uniform index, one can show explicitly that, as the index difference goes to zero, the fiber has a doubly-degenerate single mode [14, Chapters 8 and 10]. I know of no proof that this result holds more generally with the complex index profiles, like W-profiles, that have become commonplace in modern-day optical fibers, although it is highly plausible physically. In practice, researchers determine both the parameter regime in which fibers have a single mode and the numerical variation of $\beta(\omega_0)$ by solving (11) numerically [14, Chapters 8 and 10].

We next write

$$\mathbf{F} = \left(\frac{\omega_0}{2\varepsilon_0 c^2 \beta(\omega_0)}\right)^{1/2} [u_1 \mathbf{R}_1(\mathbf{r}_\perp, \omega_0) + u_2 \mathbf{R}_2(\mathbf{r}_\perp, \omega_0)], \tag{14}$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are the two orthogonal eigenmodes of a single mode fiber corresponding to $\mathbf{R}_{\rm o}^{(1)}$ and $\mathbf{R}_{\rm e}^{(1)}$, shown in (13), normalized so that

$$\int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho \, \mathrm{d}\rho |\mathbf{R}_{1\perp}|^2 = \int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho \, \mathrm{d}\rho |\mathbf{R}_{2\perp}|^2 = 1. \tag{15}$$

The quantities u_1 and u_2 are constant coefficients, and the factor $[\omega_0/2\varepsilon_0 c^2 \beta(\omega_0)]^{1/2}$ 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 which $R_z^{(m)}$ is neglected.

At this point, we include the variation of $\mathbf{F}$ with z and t that is due to the higher-order contributions in δ_1 as shown in (6). In this case, u_1 and u_2 become slowly varying functions of z and t, *i.e.*, $u_1 \to u_1(\delta_1 z, \delta_1 t)$ and $u_2 \to u_2(\delta_1 z, \delta_1 t)$, while $\mathbf{F}$ becomes a sum over the narrow bandwidth of modes that make up the signal. Writing the Fourier transform of $u_1(\delta_1 z, \delta_1 t)$ as $\tilde{u}_1(\delta_1 z, \Omega) = \int_{-\infty}^{\infty} \mathrm{d}t u_1(\delta_1 z, \delta_1 t) \exp(\mathrm{i}\Omega t)$ and writing the Fourier transform of $u_2(\delta_1 z, \delta_1 t)$ similarly, we find that

$$\begin{aligned}\mathbf{F} = \left(\frac{\omega_0}{2\varepsilon_0 c^2 \beta(\omega_0)}\right)^{1/2} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} [&\tilde{u}_1(\delta_1 z, \omega - \omega_0) \mathbf{R}_1(\mathbf{r}_\perp, \omega) + \tilde{u}_2(\delta_1 z, \omega - \omega_0) \\ &\times \mathbf{R}_2(\mathbf{r}_\perp, \omega)] \exp[-\mathrm{i}(\omega - \omega_0)t].\end{aligned} \tag{16}$$

Expanding in a narrow bandwidth around $\omega = \omega_0$, we find

$$\begin{aligned}\mathbf{F} = &\left(\frac{\omega_0}{2\varepsilon_0 c^2 \beta(\omega_0)}\right)^{1/2} \\ &\Bigg[u_1(\delta_1 z, \delta_1 t) \mathbf{R}_1(\mathbf{r}_\perp, \omega_0) + u_2(\delta_1 z, \delta_1 t) \mathbf{R}_2(\mathbf{r}_\perp, \omega_0) \\ &+ \mathrm{i}\mathbf{R}_1'(\mathbf{r}_\perp, \omega_0) \frac{\partial u_1(\delta_1 z, \delta_1 t)}{\partial t} + \mathrm{i}\mathbf{R}_2'(\mathbf{r}_\perp, \omega_0) \frac{\partial u_2(\delta_1 z, \delta_1 t)}{\partial t} \\ &- \tfrac{1}{2}\mathbf{R}_1''(\mathbf{r}_\perp, \omega_0) \frac{\partial^2 u_1(\delta_1 z, \delta_1 t)}{\partial t^2} - \tfrac{1}{2}\mathbf{R}_2''(\mathbf{r}_\perp, \omega_0) \frac{\partial^2 u_2(\delta_1 z, \delta_1 t)}{\partial t^2} + \cdots \Bigg],\end{aligned} \tag{17}$$

where $x'(\omega_0) \equiv \partial x/\partial\omega|_{\omega=\omega_0}$ and $x''(\omega_0) \equiv \partial^2 x/\partial\omega^2|_{\omega=\omega_0}$.

The goal is to determine the evolution of u_1 and u_2. We may focus on the evolution of u_1 since u_1 and u_2 are uncoupled and satisfy identical evolution equations, neglecting the contributions of order δ_2 at this point. Taking the dot product of (11) with $\mathbf{R}_1^*$ and integrating over the transverse profile, we find that $\beta^2(\omega_0)+\tilde{\psi}(\omega_0)\beta(\omega_0)-(\omega_0/c)^2[\tilde{\varepsilon}(\omega_0)/\varepsilon_0]=0$, where

$$\tilde{\varepsilon}(\omega_0) = \varepsilon_0 \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \rho \,\mathrm{d}\rho\, \mathbf{R}_1^*(\mathbf{r}_\perp, \omega_0) \cdot \left[1 + \tilde{\chi}_{\mathrm{L}}(\rho, \omega_0) - \frac{c^2}{\omega_0^2}\nabla_\perp \times \nabla_\perp \times\right] \mathbf{R}_1(\mathbf{r}_\perp, \omega_0), \tag{18}$$

$$\tilde{\psi}(\omega_0) = \mathrm{i}\int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \rho \,\mathrm{d}\rho \times [R_{1z}^*(\mathbf{r}_\perp, \omega_0)\nabla_\perp \cdot \mathbf{R}_{1\perp}(\mathbf{r}_\perp, \omega_0) + \mathbf{R}_{1\perp}^*(\mathbf{r}_\perp, \omega_0) \cdot \nabla_\perp R_{1z}(\mathbf{r}_\perp, \omega_0)].$$

Letting $\xi = \delta_1 z$ and $s = \delta_1 t$, we now insert (6) and (17) into (10), take the dot product of the result with $\mathbf{R}_1^*$, and integrate over ρ and θ to obtain

$$\int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \rho \,\mathrm{d}\rho \mathbf{R}_1^* \cdot \left[\hat{\mathcal{L}}_0(\mathbf{R}_1)u_1 - \mathrm{i}\delta_1 \frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \beta}\frac{\partial u_1}{\partial \xi} + \mathrm{i}\delta_1 \frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega}\bigg|_\beta \frac{\partial u_1}{\partial s} - \tfrac{1}{2}\delta_1^2 \frac{\partial^2 \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \beta^2}\frac{\partial^2 u_1}{\partial \xi^2} + \delta_1^2 \frac{\partial}{\partial \beta}\frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega}\bigg|_\beta \frac{\partial^2 u_1}{\partial \xi \,\partial s} - \tfrac{1}{2}\delta_1^2 \frac{\partial^2 \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega^2}\bigg|_\beta \frac{\partial^2 u_1}{\partial s^2} + \cdots\right] \equiv \hat{\ell}(u_1) = 0, \tag{19}$$

where $\hat{\mathcal{L}}_0$ is defined in (11), and $\partial^n \hat{\mathcal{L}}_0(\mathbf{R}_1)/\partial\omega^n|_\beta$ indicates the nth derivative of $\hat{\mathcal{L}}_0(\mathbf{R}_1)$, evaluated at $\omega = \omega_0$ and holding β fixed. Noting that $\hat{\mathcal{L}}_0(\mathbf{R}_1) = 0$ and that

$$0 = \frac{\mathrm{d}\hat{\mathcal{L}}_0(\mathbf{R}_1)}{\mathrm{d}\omega} = \beta' \frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \beta} + \frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega}\bigg|_\beta, \tag{20}$$

we obtain at first order in δ_1,

$$\mathrm{i}\frac{\partial u_1}{\partial \xi} + \mathrm{i}\beta'(\omega_0)\frac{\partial u_1}{\partial s} = 0. \tag{21}$$

Continuing to second order, one must first note that through this order

$$\frac{\partial^2 u_1}{\partial \xi \,\partial s} = -\beta'(\omega_0)\frac{\partial^2 u_1}{\partial s^2}, \qquad \frac{\partial^2 u_1}{\partial \xi^2} = [\beta'(\omega_0)]^2 \frac{\partial^2 u_1}{\partial s^2}. \tag{22}$$

We also note that

$$0 = \frac{\mathrm{d}^2 \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\mathrm{d}\omega^2} = \tfrac{1}{2}\beta'' \frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \beta} + \tfrac{1}{2}(\beta')^2 \frac{\partial^2 \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\beta^2} + \beta' \frac{\partial}{\partial \beta}\frac{\partial \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega}\bigg|_\beta + \frac{1}{2}\frac{\partial^2 \hat{\mathcal{L}}_0(\mathbf{R}_1)}{\partial \omega^2}\bigg|_\beta. \tag{23}$$

After substitution, we then find that

$$\mathrm{i}\frac{\partial u_1}{\partial \xi} + \mathrm{i}\beta'(\omega_0)\frac{\partial u_1}{\partial s} - \delta_1 \tfrac{1}{2}\beta''(\omega_0)\frac{\partial^2 u_1}{\partial s^2} = 0. \tag{24}$$

Continuing through arbitrarily high order in δ_1 yields

$$\mathrm{i}\frac{\partial u_1(z,t)}{\partial z} + \mathrm{IFT}\{[\beta(\omega_0+\Omega) - \beta(\omega_0)]\tilde{u}_1(z,\Omega)\} = 0, \tag{25}$$

where IFT$(\cdot)$ indicates the inverse Fourier-transform with respect to the Fourier-transform variable Ω, and we have returned to z and t coordinates. This result implies that

$$\begin{aligned}\hat{\ell}(u_1) \;=\; & -\mathrm{i}[2\beta(\omega_0) + \tilde{\psi}(\omega_0)] \\ & \times\left(\frac{\partial u_1(z,t)}{\partial z} - \mathrm{i}\ \mathrm{IFT}\{[\beta(\omega_0+\Omega) - \beta(\omega_0)]\tilde{u}_1(z,\Omega)\}\right),\end{aligned} \tag{26}$$

where $\hat{\ell}(u_1)$ was defined in (19).

At some point as we expand in powers of δ_1, the additional terms in (25) become less important than the terms of order δ_2 and δ_3 that have been neglected thus far. However, that point is problem-dependent. To model systems with a narrow bandwidth, such as single-channel systems, it is usually sufficient to keep only second derivative terms in time which leads to the usual group-velocity dispersion. When the channel is near the zero-dispersion point or there are many channels in the system, then it is necessary to keep the third derivative terms that lead to higher-order dispersion. Typically, that is all that is necessary when modeling optical-fiber transmission alone.

I will stress that *it is not correct to set* $\partial^2 u_1/\partial\xi^2$ *to zero* as is done in some elementary derivations of the nonlinear Schrödinger equation that can be found in standard textbooks. If one does that and carries out the expansion to second order, one does not obtain $\frac{1}{2}\beta''$ as the coefficient in front of the second derivative in time. This mistake has led to untold confusion.

A couple of salient points emerge from Equations (25) and (26). First, the slowly-varying envelope approximation, the physical assumption that there are only forward-going waves in the optical fiber, and the reduction of the propagation equation from a second-order equation in z to a first-order equation are all tightly coupled. In fact, when laser light is injected into an optical fiber, there is a small region about the entry point in the fiber where both forward-going (transmitted) and backward-going (reflected) waves exist, but the light that is transmitted quickly sorts itself out into the correct modal pattern. Second, when the additional contributions due to the birefrigence and the nonlinearity are added, (25) takes on the form

$$\mathrm{i}\frac{\partial u_1(z,t)}{\partial z} + \mathrm{IFT}\{[\beta(\omega_0+\Omega) - \beta(\omega_0)]\tilde{u}_1(z,\Omega)\} - \frac{1}{2\beta(\omega_0) + \tilde{\psi}(\omega_0)}S(z,t) = 0, \tag{27}$$

where $S(z,t)$ contains the additional contributions. However, the magnitude of $\tilde{\psi}(\omega_0)$ is typically about 10^{-3} of the magnitude of $\beta(\omega)$. Thus, it is safe to ignore it as is always done in practice.

There is a final issue that we must address before discussing birefringence and nonlinearity. In the derivation of (25), we implicitly assumed that $\beta(\omega)$ is real. In fact, $\beta(\omega)$ has a small imaginary part that is responsible for fiber attenuation. This attenuation has a well-documented

frequency dependence [14, Chapter 5]; however, it is usually sufficient in practice to ignore this variation. In this case, rewriting (6) as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{F}(\delta_1 z, \mathbf{r}_\perp, \delta_1 t, \omega_0) \exp(\mathrm{i}\{\mathfrak{Re}[\beta(\omega_0)]z - \omega_0 t\}) + \text{complex conjugate}, \tag{28}$$

(26) becomes

$$\mathrm{i}\frac{\partial u_1(z,t)}{\partial z} + \mathrm{IFT}[\{\beta(\omega_0 + \Omega) - \mathfrak{Re}[\beta(\omega_0)]\}\tilde{u}_1(z, \Omega)] = 0. \tag{29}$$

After expansion through the third derivative in time, we obtain

$$\mathrm{i}\frac{\partial u_1}{\partial z} - \mathrm{i}g(\omega_0)u_1 + \mathrm{i}\beta'(\omega_0)\frac{\partial u_1}{\partial t} - \tfrac{1}{2}\beta''(\omega_0)\frac{\partial^2 u_1}{\partial t^2} - \tfrac{1}{6}\mathrm{i}\,\beta'''(\omega_0)\frac{\partial^3 u_1}{\partial t^3} = 0, \tag{30}$$

where $g(\omega_0) \equiv -\mathfrak{Im}[\beta(\omega_0)]$, $\beta'(\omega_0) \equiv \mathfrak{Re}[\beta'(\omega)]|_{\omega=\omega_0}$, $\beta''(\omega_0) \equiv \mathfrak{Re}[\beta''(\omega)]|_{\omega=\omega_0}$, and $\beta'''(\omega_0) \equiv \mathfrak{Re}[\beta'''(\omega)]|_{\omega=\omega_0}$. Note that we are redefining β', β'', ... to assure that they are real. The imaginary parts are in effect assumed to be small. There are two important practical cases in which the expansion given in (30) may not be legitimate. The first is when erbum-doped fiber amplifiers are used in conjunction with WDM and the second is when filters are used. When the bandwidth of the signal is large enough so that the frequency response cannot be approximated by a simple Taylor expansion – as often occurs in practice – then the full response function given in (29) must be used.

We next turn to the effects of fiber asymmetry and anisotropy that lead to birefringence. Including these effects, (10) becomes

$$\hat{\mathcal{L}}[\mathbf{E}(\mathbf{r}, t)] + \delta_2 \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\int_{-\infty}^{t} \mathrm{d}t_1 \Delta X_\mathrm{L}(\delta_2 z, \mathbf{r}_\perp, t - t_1) \cdot \mathbf{E}(\mathbf{r}, t_1) = 0, \tag{31}$$

where ΔX_L is defined in (9) and $\hat{\mathcal{L}}(\mathbf{E})$ is defined in (10). Inserting (6), taking the dot product of (31) with $\mathbf{R}_1^*$, and integrating over the transverse profile of the fiber, just as we did when neglecting asymmetry and anisotropy, we obtain an evolution equation for u_1. We can similarly obtain an evolution equation for u_2. These equations may be written

$$\begin{aligned}
&\hat{\ell}(u_j) - \delta_2 \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \rho\, \mathrm{d}\rho \mathbf{R}_j^*(\mathbf{r}_\perp, \omega_0) \\
&\quad \times \left\{ \frac{\omega_0^2}{c^2}\Delta\tilde{X}_\mathrm{L}(\mathbf{r}_\perp, \omega_0) \cdot [u_1\mathbf{R}_1(\mathbf{r}_\perp, \omega_0) + u_2\mathbf{R}_2(\mathbf{r}_\perp, \omega_0)] \right. \\
&\qquad + \mathrm{i}\delta_1 \left[\frac{\omega_0^2}{c^2}\Delta\tilde{X}_\mathrm{L}(\mathbf{r}_\perp, \omega_0) \cdot \mathbf{R}_1(\mathbf{r}_\perp, \omega_0)\right]' \frac{\partial u_1}{\partial s} \\
&\qquad \left. + \mathrm{i}\delta_1 \left[\frac{\omega_0^2}{c^2}\Delta\tilde{X}_\mathrm{L}(\mathbf{r}_\perp, \omega_0) \cdot \mathbf{R}_2(\mathbf{r}_\perp, \omega_0)\right]' \frac{\partial u_2}{\partial s} + \cdots \right\} = 0,
\end{aligned} \tag{32}$$

where $\hat{\ell}(u_j)$ is defined in (19) and (26), $j = 1$ or 2, $\Delta\tilde{X}_\mathrm{L}$ is the Fourier transform of ΔX_L, and the prime indicates as usual the derivative with respect to ω, evaluated at $\omega = \omega_0$. The

z-dependence of $\Delta\tilde{X}_{\mathrm{L}}$ and the z and t dependences of the u_j and their derivatives have been suppressed. We now proceed by analogy with our previous derivation of (29) to find through second order in δ_1

$$\mathrm{i}\frac{\partial \mathbf{U}}{\partial \xi} - \frac{\mathrm{i}}{\delta_1} g(\omega_0)\mathbf{U} + \frac{1}{\delta_1}\Delta\mathsf{B}(\omega_0)\mathbf{U} + \mathrm{i}\mathsf{B}'(\omega_0)\frac{\partial \mathbf{U}}{\partial s} - \delta_1 \tfrac{1}{2}\mathsf{B}''(\omega_0)\frac{\partial^2 \mathbf{U}}{\partial s^2} = 0, \tag{33}$$

where $\mathbf{U}$ indicates the two-element column vector $\binom{u_1}{u_2}$. The quantities $\mathsf{B}(\omega_0)$ and $\Delta\mathsf{B}(\omega_0)$ are 2×2 matrices defined as $\mathsf{B}(\omega_0) \equiv \beta(\omega_0)\mathsf{I} + \delta_2[\omega_0^2/2\beta(\omega_0)c^2][\mathsf{E}(\omega_0)/\varepsilon_0]$ and $\Delta\mathsf{B}(\omega_0) \equiv \mathsf{B}(\omega_0) - \beta(\omega_0)\mathsf{I}$, where I is the unit matrix and

$$\mathsf{E}_{ij}(\omega_0) = \varepsilon_0 \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \rho \, \mathrm{d}\rho \mathbf{R}_i^* \cdot \Delta\tilde{X}_{\mathrm{L}} \cdot \mathbf{R}_j, \tag{34}$$

with $i = 1, 2$ and $j = 1, 2$. Note that inverse powers of δ_1 appear at the lowest order in (33). Hence, it is important that the length scales associated with attenuation and birefringence are long compared to the pulse durations as is indeed physically the case. Consequently, the ratios g/δ_1 and δ_2/δ_1 are small. The contribution of $\Delta\mathsf{B}''$ to the evolution of $\mathbf{U}$ is too small to be observable in practice, and it will be neglected from hereon. By contrast, the contribution of $\Delta\mathsf{B}'$ to the evolution of $\mathbf{U}$ is very important since it leads to polarization mode dispersion. We now make three physical assumptions that are appropriate for optical fibers. The first is that $\Delta\mathsf{B}$ is Hermitian so that there is no polarization dependent loss. While polarization dependent loss can be present in some discrete components and can play a role over very long distances [15], there is no evidence that it is present in the optical fiber itself. The second is that there is no helicity in glass so that we may write

$$\Delta\mathsf{B} = \Delta\beta(\cos\,\theta\sigma_3 + \sin\,\theta\sigma_1), \tag{35}$$

where

$$\sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \sigma_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are standard Pauli matrices, and θ is an orientation angle. Finally, we assume that $\Delta\mathsf{B}'$ is oriented in the same direction as $\Delta\mathsf{B}$ so that

$$\Delta\mathsf{B}' = \Delta\beta'(\cos\,\theta\sigma_3 + \sin\,\theta\sigma_1). \tag{36}$$

This assumption is valid because physically the birefringence is due to anisotropies and asymmetries that are very weak functions of frequency. We note that the orientation angle $\theta = \theta(\delta_2 z)$ varies on the length scale corresponding to δ_2 since by a physical coincidence the beat lengths and fiber correlation lengths are comparable in magnitude as shown in Figure 1.

We now turn to consideration of the nonlinear contributions. Since the contributions of the birefringence and the nonlinearity to the evolution of $\mathbf{U}$ simply add through the first order in δ_2 and δ_3, which is the only order beyond zero order that it is necessary in practice to keep, we may consider the nonlinear contributions separately from the contributions due to anisotropy and asymmetry and then simply add both sets of contributions at the end. Thus, (10) becomes

$$\hat{\mathcal{L}}(\mathbf{E}) + \frac{1}{\varepsilon_0 c^2}\frac{\partial^2}{\partial t^2}\mathbf{P}_{\mathrm{NL}} = 0, \tag{37}$$

where $\mathbf{P}_{\mathrm{NL}}$ is given in (8). We now insert (6) into (37) to determine the evolution of $\mathbf{U}$. One immediately finds that components at $\omega = \pm 3\omega_0$ are produced in addition to the components at $\pm\omega_0$. Strictly speaking, we should write

$$\mathbf{E}(\mathbf{r}, t) = \sum_{n=0}^{\infty} \mathbf{F}^{(n)}(\delta_1 z, \mathbf{r}_\perp, \delta_1 t, \omega_0) \exp\{(2n+1)\mathrm{i}[\beta(\omega_0)z - \omega_0 t]\} + \text{complex conjugate}, \tag{38}$$

where $\mathbf{F}^{(0)} \equiv \mathbf{F}$, and solve self-consistently for all the harmonics. This issue has been thoroughly discussed by Kodama [12]. Because all the harmonics at $n = 1$ and higher are nonresonant, they are vanishingly small and negligible in practice. Keeping only the first harmonic contributions we may write

$$\mathbf{P}_{\mathrm{NL}}(\mathbf{r}, t) = \mathbf{Q}_{\mathrm{NL}}(\delta_1 z, \mathbf{r}_\perp, \delta_1 t, \omega_0) \exp\{\mathrm{i}[\beta(\omega_0)z - \omega_0 t]\} + \text{complex conjugate} \tag{39}$$

and we then find

$$\begin{aligned}\mathbf{Q}_{\mathrm{NL}}(\delta_1 t) = {} & \delta_3 \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t_1 \int_{-\infty}^{t} \mathrm{d}t_2 \int_{-\infty}^{t} \mathrm{d}t_3 \chi_{\mathrm{NL}}(t - t_1, t - t_2, t - t_3) \\ & \{2\mathbf{F}(\delta_1 t_1) \cdot \mathbf{F}^*(\delta_1 t_2)\mathbf{F}(\delta_1 t_3) \exp[\mathrm{i}\omega_0(t - t_1 + t_2 - t_3)] \\ & + \mathbf{F}(\delta_1 t_1) \cdot \mathbf{F}(\delta_1 t_2)\mathbf{F}^*(\delta_1 t_3) \exp[\mathrm{i}\omega_0(t - t_1 - t_2 + t_3)]\},\end{aligned} \tag{40}$$

where the dependences on $\mathbf{r}_\perp$ and z have been suppressed for clarity. We have used the symmetry of χ_{NL} with respect to the exchange $t_1 \leftrightarrow t_2$ to simplify (40) from three terms to two.

Physically, there are three major contributions to χ_{NL}. The first is the contribution from the electronic resonances which have a characteristic times scale that is a fraction of a femtosecond; these contributions can be considered instantaneous in practice. The second is the contribution from the Raman effect due to vibrational resonances of the individual glass molecules; it has a characteristic time on the order of 100 femtoseconds. This contribution can usually be considered instantaneous for single-channel communication systems, but that is not the case for some high-power undersea systems in which the input is hundreds of milliwatts. It is also not necessarily the case for WDM systems with a large bandwidth [16]. The third major contribution to the nonlinearity is from the Brillouin effect which is due to vibrational modes of the entire fiber (sound waves); these contributions have a characteristic time of ten nanoseconds or more. This contribution can become important at powers of 10 mW or more, which is commonly obtained in practice; however, the Brillouin effect requires a large degree of coherence in the communication signal to manifest itself, and it is easily avoided in practice by dithering the signal or otherwise spreading its bandwidth [16]. Optical fibers have low attenuation because the frequency of propagation ω_0 is far from any of the quantum level differences (resonances) in the glass. As a consequence, the noninstantaneous contributions are limited in their allowed behavior. One can show from either first quantum-mechanical principles [17] or the uncertainty principle [18] that χ_{NL} differs significantly from zero only when $t = t_3$ and $t_1 = t_2$ to within ω_0^{-1} in the first term in (40) and when $t = t_1$ and $t_2 = t_3$ or vice versa to within ω_0^{-1} in the second term. Consequently, we

may write

$$\mathbf{Q}_{\mathrm{NL}}(\delta_1 t) = \delta_3\varepsilon_0\Bigg\{\chi_{\mathrm{I}}[2\mathbf{F}(\delta_1 t)\cdot\mathbf{F}^*(\delta_1 t)\mathbf{F}(\delta_1 t)+\mathbf{F}(\delta_1 t)\cdot\mathbf{F}(\delta_1 t)\mathbf{F}^*(\delta_1 t)]$$
$$+\int_{-\infty}^{t}\mathrm{d}t_1 a(t-t_1)\mathbf{F}(\delta_1 t_1)\cdot\mathbf{F}^*(\delta_1 t_1)\mathbf{F}(\delta_1 t)$$
$$+\int_{-\infty}^{t}\mathrm{d}t_1 b(t-t_1)\mathbf{F}(\delta_1 t)$$
$$\cdot[\mathbf{F}^*(\delta_1 t_1)\mathbf{F}(\delta_1 t_1)+\mathbf{F}(\delta_1 t_1)\mathbf{F}^*(\delta_1 t_1)]\Bigg\}. \tag{41}$$

We have taken advantage of the slow variation of $\mathbf{F}(\delta_1 t)$ relative to ω_0^{-1} to neglect its changes on this time scale.

We now proceed by analogy with our previous calculations to determine the evolution equation for $\mathbf{U}$. We insert (39) and (41) into (37), take the dot product of the result with $\mathbf{R}_1^*$, and integrate over the transverse profile of the fiber to determine an evolution equation for u_1. We then proceed similarly to find an evolution equation for u_2. Following convention, we next define an effective area

$$\frac{1}{A_{\mathrm{eff}}}\equiv\int_0^{2\pi}\mathrm{d}\theta\int_0^{\infty}\rho\,\mathrm{d}\rho(\mathbf{R}_j^*\cdot\mathbf{R}_j)^2, \tag{42}$$

where $j = 1$ or 2. This area is approximately but not exactly equal to the fiber core area. We keep only the lowest-order contribution in δ_1 for the nonlinear term since there is no experimental evidence that the higher-order terms contribute measurably to the evolution. We now find, adding together the contributions from fiber anisotropy and asymmetry and the contribution from the nonlinearity, that the evolution of $\mathbf{U}$ is given by

$$\mathrm{i}\frac{\partial\mathbf{U}(z,t)}{\partial z}+\mathrm{IFT}(\{\beta(\omega_0+\Omega)-\mathfrak{Re}[\beta(\omega_0)]\}\tilde{\mathbf{U}}(z,\Omega))$$
$$+(\cos\theta\sigma_3+\sin\theta\sigma_1)\left(\Delta\beta(\omega_0)\mathbf{U}(z,t)+\mathrm{i}\Delta\beta'(\omega_0)\frac{\partial\mathbf{U}(z,t)}{\partial t}\right)$$
$$+\frac{\omega_0}{c}\frac{n_2}{A_{\mathrm{eff}}}f_{\mathrm{I}}\{|\mathbf{U}(z,t)|^2\mathbf{U}(z,t)-\tfrac{1}{3}[\mathbf{U}^\dagger(z,t)\sigma_2\mathbf{U}(z,t)]\sigma_2\mathbf{U}(z,t)\}$$
$$+\frac{\omega_0^3}{4\varepsilon_0\beta^2(\omega_0)c^4A_{\mathrm{eff}}}\int_{-\infty}^{t}\mathrm{d}t_1\{a(t-t_1)|\mathbf{U}(z,t_1)|^2\mathbf{U}(z,t)$$
$$+b(t-t_1)\mathbf{U}(z,t)\cdot[\mathbf{U}(z,t_1)\mathbf{U}^*(z,t_1)+\mathbf{U}^*(z,t_1)\mathbf{U}(z,t_1)]\}=0. \tag{43}$$

The dependence on the ordering parameters has been suppressed. The dependences of the parameters β, $\Delta\beta$, $\Delta\beta'$, θ, n_2, and A_{eff} on z have also been suppressed. The parameter

$$n_2 f_{\mathrm{I}}=\frac{3\omega_0^2\chi_{\mathrm{I}}}{4\varepsilon_0\beta^2(\omega_0)c^3} \tag{44}$$

has a factor $f_{\mathrm{I}} \simeq 0{\cdot}82$ because when the Kerr coefficient n_2 is measured experimentally, it includes contributions of the Raman effect which have been explicitly separated here [19]. Interpreting $\mathbf{U}$ as a column matrix $\binom{u_1}{u_2}$, the quantity $\mathbf{U}^\dagger$ is a row matrix (u_1^*, u_2^*) so that $\mathbf{U}^\dagger\sigma_2\mathbf{U} = -\mathrm{i}(u_1^*u_2 - u_2^*u_1)$, where

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

is one of the standard Pauli matrices. In obtaining (43), we ignore small corrections that result from the deviation of the factors $\int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho\,\mathrm{d}\rho\ (\mathbf{R}_1 \cdot \mathbf{R}_1^*)(\mathbf{R}_2 \cdot \mathbf{R}_2^*)$ and $\int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho\,\mathrm{d}\rho(\mathbf{R}_i \cdot \mathbf{R}_i)(\mathbf{R}_j^* \cdot \mathbf{R}_j^*)$ $(i = 1, 2, j = 1, 2)$ from $1/A_{\mathrm{eff}}$. These deviations are less than 1% in all practical cases and will have no measurable effect. Similarly, we may ignore the deviation of the factors $\int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho\,\mathrm{d}\rho(\mathbf{R}_1 \cdot \mathbf{R}_2)(\mathbf{R}_1^* \cdot \mathbf{R}_2^*)$ and $\int_0^{2\pi} \mathrm{d}\theta \int_0^\infty \rho\,\mathrm{d}\rho(\mathbf{R}_i \cdot \mathbf{R}_{3-i}^*)(\mathbf{R}_j \cdot \mathbf{R}_{3-j}^*)$ $(i = 1, 2, j = 1, 2)$ from zero.

The Brillouin contribution to $a(\tau)$ and $b(\tau)$ is on such a long time scale that it must be studied in conjunction with reduced models that only take into acount some average features of the signal channel rather than its detailed temporal behavior [16]. By contrast, the Raman contributions are on a short time scale so that it is often legitimate to make the replacement

$$\begin{aligned}
&\int_{-\infty}^{t} \mathrm{d}t_1\{a_{\mathrm{R}}(t - t_1)|\mathbf{U}(t_1)|^2\mathbf{U}(t) + b_{\mathrm{R}}(t - t_1)\mathbf{U}(t) \cdot [\mathbf{U}(t_1)\mathbf{U}^*(t_1) + \mathbf{U}^*(t_1)\mathbf{U}(t_1)]\} \\
&\quad \simeq [\tilde{a}_{\mathrm{R}}(0) + \tilde{b}_{\mathrm{R}}(0)]|\mathbf{U}(t)|^2\mathbf{U}(t) + \tilde{b}_{\mathrm{R}}(0)\mathbf{U}(t) \cdot \mathbf{U}(t)\mathbf{U}^*(t) \\
&\qquad + \mathrm{i}\tilde{a}'_{\mathrm{R}}(0)\mathbf{U}(t)\frac{\partial}{\partial t}|\mathbf{U}(t)|^2 + \mathrm{i}\tilde{b}'_{\mathrm{R}}(0)\mathbf{U}(t) \cdot \frac{\partial}{\partial t}[\mathbf{U}(t)\mathbf{U}^*(t) + \mathbf{U}^*(t)\mathbf{U}(t)],
\end{aligned} \tag{45}$$

where $a_{\mathrm{R}}(\tau)$ and $b_{\mathrm{R}}(\tau)$ are the Raman contributions to $a(\tau)$ and $b(\tau)$, $\tilde{a}_{\mathrm{R}}(\Omega)$ and $\tilde{b}_{\mathrm{R}}(\Omega)$ are their Fourier transforms, while $\tilde{a}'_{\mathrm{R}}(0) = \mathrm{d}\tilde{a}_{\mathrm{R}}/\mathrm{d}\Omega|_{\Omega=0}$ and $\tilde{b}'_{\mathrm{R}}(0) = \mathrm{d}\tilde{b}_{\mathrm{R}}/\mathrm{d}\Omega|_{\Omega=0}$. The first two terms on the right-hand side of (45) can be absorbed directly into the Kerr coefficient, and consistency demands that $\tilde{a}_{\mathrm{R}}(0) = \tilde{b}_{\mathrm{R}}(0) = \chi_{\mathrm{I}}(1 - f_{\mathrm{I}})$. Kodama has shown that if instead of stopping at the zeroth order in δ_1 in the expansion of the nonlinear terms, we continue to first order, then we obtain derivatives of the electric field analogous to the terms on the right-hand side of (45) [12]. However, these contributions have real coefficients and appear to be negligible in practice.

It might seem a bit surprising and even contradictory to readers with a background in soliton physics that one often needs in practice to keep the third and sometimes even higher derivatives in the dispersive term (second term on the left-hand side) of (43), while it is possible to neglect the higher-order corrections to the nonlinearity. Textbook derivations of the nonlinear Schrödinger equation and its modifications typically assume that $\delta_3 = \delta_1^2$. With this assumption the third derivative in the dispersive term appears at the same order as the first derivative contributions to the nonlinearity. This assumption makes physical sense for standard solitons because there is a direct tie between the soliton's bandwidth and its peak intensity; as the pulse duration decreases, increasing the bandwidth, the peak intensity must also increase. It is important to recognize that this assumption is not true in general, and, in fact, it is not true for any current communication system. Indeed, the trend has been toward lowering the signal intensity in current NRZ systems to reduce the nonlinear impairments while raising the bandwidth through the use of WDM. However, even in short pulse experiments in which one

might expect to observe the higher-order nonlinear corrections, other effects like the Raman effect, gain saturation, and geometric effects due to lumped elements appear to overshadow these terms in practice. Thus, they have never to my knowledge been unambiguously observed.

Equation (43) is the most general form of the coupled nonlinear Schrödinger equation that is of any use in practice and is the central result of this section. It is usually simplified further in applications. In most optical-fiber transmission systems the noninstantaneous contributions to the nonlinearity can be neglected. While it is easy to retain the full expression $\mathrm{IFT}(\{\beta(\omega_0+\Omega)-\mathfrak{Re}[\beta(\omega_0)]\}\tilde{\mathbf{U}}(z,\Omega))$ in numerical simulations of transmission systems, which are typically based on split-step, semi-spectral methods, it is sufficient when simulating the optical fiber itself to keep terms up to the third derivative in time. While we must often keep the full expression when dealing with the amplifiers and filters that are inevitably in any real system, these are normally lumped elements in the transmission line and can be treated separately from the optical fiber. Finally, it is conventional to use retarded time, $t_{\mathrm{retarded}} = t - \beta'(\omega_0)z$, rather than standard time. This transformation is of critical importance in simulations because it allows one to view the signal evolution in a time window of limited duration. [More generally, this retarded time should be written $t_{\mathrm{retarded}} = t - \int_0^z \beta'(\omega_0, z_1)\,\mathrm{d}z_1$, to take into account the slow variations of the fiber on the length scale corresponding to δ_2; however, just as in the case of the variations of $\beta(\omega_0, z)$, the variations of $\beta'(\omega_0, z)$ are negligible in practice.] With these simplifications, (43) becomes

$$\mathrm{i}\frac{\partial \mathbf{U}}{\partial z} - \mathrm{i}g\mathbf{U} + (\cos\theta\sigma_3 + \sin\theta\sigma_1)\left(\Delta\beta\mathbf{U} + \mathrm{i}\Delta\beta'\frac{\partial \mathbf{U}}{\partial t}\right)$$

$$-\tfrac{1}{2}\beta''\frac{\partial^2\mathbf{U}}{\partial t^2} - \tfrac{1}{6}\mathrm{i}\beta'''\frac{\partial^3\mathbf{U}}{\partial t^3} + \gamma[|\mathbf{U}|^2\mathbf{U} - \tfrac{1}{3}(\mathbf{U}^\dagger\sigma_2\mathbf{U})\sigma_2\mathbf{U}] = 0, \tag{46}$$

where t now refers to retarded time and $\gamma = \omega_0 n_2/cA_{\mathrm{eff}}$. Equation (46) is the version of the coupled nonlinear Schrödinger equation that will be the basis of our analysis in the next section, although it must be borne in mind that there are practical cases in which one or more terms of the fuller expression in (43) that have been left out of (46) become important. It should also be borne in mind that the birefringent coefficients θ, $\Delta\beta$, and $\Delta\beta'$ all vary on the length scale corresponding to δ_2, while g and β'' vary on the length scale corresponding to δ_3. The variations of β''' and n_2 are typically negligible. In addition to its variations on the length scale corresponding to δ_3, the dispersion β'' also has variations on the length scale corresponding to δ_2, but the signal averages over these variations, as will be discussed in the next section, and they are negligible.

3. The Manakov-PMD equation and its reductions

A. The Manakov-PMD equation

We now turn to the next stage of the analysis – averaging over the rapidly and randomly varying birefringence on the length scale corresponding to δ_2 to determine the appropriate evolution equations on the length scale corresponding to δ_3. We begin by first considering the evolution on the length scale corresponding to δ_2, neglecting the evolution on the longer length scale corresponding to δ_3. Thus, only keeping terms of order δ_2 and taking the Fourier transform, (46) becomes

$$\mathrm{i}\frac{\partial\tilde{\mathbf{U}}(z,\Omega)}{\partial z} + [\cos\theta(z)\sigma_3 + \sin\theta(z)\sigma_1][\Delta\beta(z) + \Delta\beta'(z)\Omega]\tilde{\mathbf{U}}(z,\Omega) = 0. \tag{47}$$

The random variation of θ, $\Delta\beta$, and $\Delta\beta'$ implies that a complete characterization of the statistical properties of $\tilde{\mathbf{U}}$ can only be accomplished once the statistical properties of θ, $\Delta\beta$, and $\Delta\beta'$ are known. Unfortunately, very little is known about the actual statistical properties of these quantities. On the other hand, numerical studies [20, 21] supported by analytical work [22, 23] show that, as long as θ is uniformly distributed, the distribution of $\Delta\beta$ does not matter much in determining the statistical properties of the field $\tilde{\mathbf{U}}$. It has generally been assumed in theoretical work to date that $\Delta\beta' = \Delta\beta/\omega_0$, consistent with the observation that the phase and group velocities are nearly identical in optical fibers [24]. Specifically two very different models of the fiber statistics have been compared both numerically [20, 21] and using Ito's method [23]. In the first model, $\Delta\beta$ is held fixed, and in the second $\Delta\beta\ \cos\theta$ and $\Delta\beta\ \sin\theta$ are both assumed to have independent Gaussian distributions. These models yield nearly identical results for the statistical quantities characterizing the field evolution. Additionally, it has been shown that a wide range of models will yield Poole's [25] classic expression for the PMD length. However, from a mathematical standpoint, there is no proof that the field statistics are nearly model-independent, and this issue remains open.

We begin the analysis by transforming $\tilde{\mathbf{U}}$ in a way that would diagonalize its evolution, were it not for the z-variation of θ, letting $\tilde{\mathbf{V}} = \mathsf{R}^{-1}\tilde{\mathbf{U}}$, where $\mathsf{R}^{-1} = \cos(\theta/2)\mathsf{I} + \mathrm{i}\,\sin(\theta/2)\sigma_2$. Equation (47) now becomes

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

where $\theta_z \equiv \mathrm{d}\theta/\mathrm{d}z$. We next write

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

where T satisfies (48) at $\Omega = 0$, *i.e.*,

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

and $\mathsf{T}\,(z = 0) = \mathsf{I}$. It follows that $\tilde{\mathbf{W}}(z, \Omega = 0) = \tilde{\mathbf{W}}(z = 0, \Omega = 0)$ is constant. Thus, the transformation $\tilde{\mathbf{W}}(z,\Omega) = \mathsf{T}^{-1}\mathsf{R}^{-1}\tilde{\mathbf{U}}(z,\Omega)$ freezes the motion of the frequency $\Omega = 0$ in SU(2) so that $\tilde{\mathbf{W}}(z,\Omega)$ measures the relatively slow motion of the other frequencies with respect to $\Omega = 0$. Explicitly, one now finds that

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

where $\bar{\sigma}_3(z) = \mathsf{T}^{-1}(z)\sigma_3\mathsf{T}(z)$. One of the important effects of the randomly varying birefringence is to spread an initial input pulse in the optical fiber. To calculate this spreading, we first write $\tilde{\mathbf{W}}(z,\Omega) = \alpha(\Omega)\tilde{\mathbf{A}}(z,\Omega)$, where $|\tilde{\mathbf{A}}(z,\Omega)|^2 = 1$. It follows from (51) that $\alpha(\Omega)$ is independent of z. The conventional Stokes parameters on the unit Poincaré sphere that are used in optics are defined in terms of $\tilde{\mathbf{A}}$ as

$$S_1 \equiv \tilde{\mathbf{A}}^\dagger\sigma_3\tilde{\mathbf{A}}, \qquad S_2 \equiv \tilde{\mathbf{A}}^\dagger\sigma_1\tilde{\mathbf{A}}, \qquad S_3 \equiv -\tilde{\mathbf{A}}^\dagger\sigma_2\tilde{\mathbf{A}}. \tag{52}$$

These Stokes parameters together define a Stokes vector $\mathbf{S} = (S_1, S_2, S_3)$.

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

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

Since the transformation relating $\tilde{\mathbf{A}}(z, \Omega)$ at different values of Ω is evidently unitary, the matrix F must be Hermitian. From the compatibility conditions for (51) and (53), we find

$$\frac{\partial \mathsf{F}}{\partial z} = \mathrm{i}\Delta\beta'\Omega(\bar{\sigma}_3\mathsf{F} - \mathsf{F}\bar{\sigma}_3) + \Delta\beta'\bar{\sigma}_3, \tag{54}$$

so that the trace is constant as a function of z. The eigenvalues of F will be designated as $T_{\mathrm{off}} \pm T_{\mathrm{PMD}}$. The eigenvectors of F are conventionally referred to as the principal states while the difference between the eigenvalues, $2T_{\mathrm{PMD}}$ is conventionally referred to as the differential delay. The matrix F is directly related to the spreading. Defining the mean signal time

$$\begin{aligned}\overline{T(z)} &= \frac{\int_{-\infty}^{\infty} t|\mathbf{W}(z, t)|^2\,\mathrm{d}t}{\int_{-\infty}^{\infty}|\mathbf{W}(z, t)|^2\,\mathrm{d}t} \\ &= \frac{-\mathrm{i}\int_{-\infty}^{\infty}[\alpha'(\Omega)\alpha(\Omega) + \alpha^2(\Omega)\tilde{\mathbf{A}}^{\dagger}(z, \Omega)\tilde{\mathbf{A}}'(z, \Omega)]\,\mathrm{d}\Omega}{\int_{-\infty}^{\infty}\alpha^2(\Omega)\,\mathrm{d}\Omega},\end{aligned} \tag{55}$$

where the primes indicate derivatives with respect to Ω, and the mean square signal time

$$\overline{T^2(z)} = \frac{\int_{-\infty}^{\infty} t^2|\mathbf{W}(z, t)|^2\,\mathrm{d}t}{\int_{-\infty}^{\infty}|\mathbf{W}(z, t)|^2\,\mathrm{d}t} = \frac{\int_{-\infty}^{\infty}|\alpha'(\Omega)\tilde{\mathbf{A}}(z, \Omega) + \alpha(\Omega)\tilde{\mathbf{A}}'(z, \Omega)|^2\,\mathrm{d}\Omega}{\int_{-\infty}^{\infty}\alpha^2(\Omega)\,\mathrm{d}\Omega}, \tag{56}$$

we may define the signal spread as $\Sigma(z) \equiv \left[\overline{T^2(z)} - \overline{T(z)}^2\right]^{1/2}$. By analogy with the definition of the Stokes vector, it is conventional to write

$$\mathsf{F}(z, \Omega) \equiv T_{\mathrm{off}}(\Omega)\mathsf{I} + \tfrac{1}{2}[\Theta_1(z, \Omega)\sigma_3 + \Theta_2(z, \Omega)\sigma_1 - \Theta_3(z, \Omega)\sigma_2], \tag{57}$$

which expresses $F(z, \Omega)$ in terms of the Pauli matrices. Writing $\Theta^2 = \Theta_1^2 + \Theta_2^2 + \Theta_3^2$, it follows that $T_{\mathrm{PMD}}^2 = \frac{1}{4}\Theta^2$. The quantity Θ just equals the differential delay, and the vector $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \Theta_3)$ is referred as the dispersion vector. Noting from (53) that $\tilde{\mathbf{A}}'(z, \Omega) = i\mathsf{F}(z, \Omega)\tilde{\mathbf{A}}(z, \Omega)$ and using the definitions of the Stokes vector in (52) and the dispersion vector in (57), we obtain

$$\begin{aligned}\Sigma^2(z) = \frac{1}{\int_{-\infty}^{\infty}\alpha^2(\Omega)\,\mathrm{d}\Omega}\Bigg\{&\int_{-\infty}^{\infty}\mathrm{d}\Omega[\alpha'(\Omega)]^2 \\ &+ \int_{-\infty}^{\infty}\mathrm{d}\Omega\alpha^2(\Omega)\Bigg[T_{\mathrm{off}}(\Omega)\Bigg(T_{\mathrm{off}}(\Omega) + \boldsymbol{\Theta}(z, \Omega)\cdot\mathbf{S}(z, \Omega) \\ &- \int_{-\infty}^{\infty}\mathrm{d}\Omega_1\alpha^2(\Omega_1)[T_{\mathrm{off}}(\Omega_1) + \boldsymbol{\Theta}(z, \Omega_1)\cdot\mathbf{S}(z, \Omega_1)]\Bigg) + \tfrac{1}{4}\Theta^2(\Omega, z) \\ &- \tfrac{1}{4}\boldsymbol{\Theta}(z, \Omega)\cdot\mathbf{S}(z, \Omega)\int_{-\infty}^{\infty}\mathrm{d}\Omega_1\alpha^2(\Omega_1)\boldsymbol{\Theta}(z, \Omega_1)\cdot\mathbf{S}(z, \Omega_1)\Bigg]\Bigg\}.\end{aligned} \tag{58}$$

We now specialize to the case in which the initial signal is in a fixed polarization state as a function of wavelength. This case is common for present-day communication systems, but there are important exceptions – notably long-distance, undersea systems in which it is common to scramble the signal's polarization. The reason for focusing on the case of a single-input polarization state is that it allows us to unambiguously focus on the spreading due to the effects of the optical fiber as opposed to the initial signal profile itself. In this case, $T_{\text{off}}(\Omega)$ and $\mathbf{\Theta}(z = 0, \Omega)$ both equal zero, and (58) may be rewritten

$$\Sigma^2(z) = \Sigma^2(z=0) + \frac{1}{\int_{-\infty}^{\infty} \alpha^2(\Omega)\,\mathrm{d}\Omega} \int_{-\infty}^{\infty} \mathrm{d}\Omega \alpha^2(\Omega) \times \left[\tfrac{1}{4}\Theta^2(\Omega, z) - \tfrac{1}{4}\mathbf{\Theta}(z, \Omega) \cdot \mathbf{S}(z, \Omega) \int_{-\infty}^{\infty} \mathrm{d}\Omega_1 \alpha^2(\Omega_1) \mathbf{\Theta}(z, \Omega_1) \cdot \mathbf{S}(z, \Omega_1) \right]. \quad (59)$$

Over sufficiently small lengths, before the spreading has become important, $\mathbf{\Theta}$ and $\mathbf{S}$ are nearly constant over the bandwidth of the signal, and (59) reduces to

$$\Sigma^2(z) - \Sigma^2(z=0) \simeq \tfrac{1}{4}|\mathbf{\Theta}(z, \Omega = 0) \times \mathbf{S}(z, \Omega = 0)|^2. \quad (60)$$

From a physical standpoint, what happens is that a signal launched in either one of the two principal states is advanced or retarded by T_{PMD}, while a signal launched in any other polarization state decomposes into the two principal states, so that part of the signal is advanced and part is retarded. When this effect is concatenated over a long length of fiber, or alternatively (but equivalently) the length has become long enough so that the approximation in (60) is no longer valid, then significant spreading occurs. This length scale is given by $T_{\text{PMD}}(z)\Delta\nu = 1$, where $\Delta\nu$ is the bandwidth of the signal (in cycles per unit time). The definition of the bandwidth is a bit arbitrary. In practice, most workers use the full width at half maximum, although the root mean square is also occasionally used in theoretical work. The two are the same within a factor of order one.

To determine the length scale on which the spreading occurs, we may first write

$$\bar{\sigma}_3 \equiv \begin{pmatrix} \cos\theta_S & \sin\theta_S \exp(\mathrm{i}\phi_S) \\ \sin\theta_S \exp(-\mathrm{i}\phi_S) & -\cos\theta_S \end{pmatrix}, \quad (61)$$

which defines the angles θ_S and ϕ_S. Calculating $\Theta_1^2(z)$ at $\Omega = 0$, we find

$$\Theta_1^2(z) = 8 \int_0^z \mathrm{d}z_1 \{\Delta\beta'(z_1) \cos[\theta_S(z_1)]\} \int_0^{z_1} \mathrm{d}z_2 \{\Delta\beta'(z_2) \cos[\theta_S(z_2)]\}. \quad (62)$$

To make further progress, one must know the autocorrelation function $C(z_1, z_2) \equiv \langle \Delta\beta'(z_1) \cos[\theta_S(z_1)] \Delta\beta'(z_2) \cos[\theta_S(z_2)] \rangle$ over an ensemble of fibers. A wide variety of physical models of the optical fiber all lead to the conclusion that [20–25]

$$C(z_1, z_2) = \tfrac{1}{3}\langle [\Delta\beta'(z)]^2 \rangle \exp[-(z_1 - z_2)/Z_{\text{c}}], \quad (63)$$

where Z_{c} is the characteristic correlation length of the optical fiber shown in Figure 1. So, we obtain

$$\langle \Theta_1^2(z) \rangle = \tfrac{8}{3}\langle [\Delta\beta'(z)]^2 \rangle \{Z_{\text{c}} z + Z_{\text{c}}^2 [\exp(-z/Z_{\text{c}}) - 1]\}. \quad (64)$$

Noting that $\langle\Theta_3^2(z)\rangle = \langle\Theta_2^2(z)\rangle = \langle\Theta_1^2(z)\rangle$, we find that $\langle T_{\rm PMD}^2\rangle = \frac{3}{4}\langle\Theta_1^2(z)\rangle$ which is Poole's classic result [25]. When $z \gg Z_{\rm c}$, then it has been shown (at the same level of rigor as this paper) that an ergodic theorem holds [22] so that

$$T_{\rm PMD}(z) = \frac{1}{2\nu_0 Z_{\rm b}}(2Z_{\rm c}z)^{1/2}, \tag{65}$$

where ν_0 is the signal's central frequency (in cycles per unit time) and $Z_{\rm b}$ is the beat length. We have made the replacement $\langle(\Delta\beta')^2\rangle = 1/2\nu_0 Z_{\rm b}$ which is only true to within about 10%. However, the beat length is notoriously difficult to characterize accurately over a long length of fiber; so, this relationship is sufficiently accurate for practical purposes. Using the condition $T_{\rm PMD}(z)\Delta\nu = 1$, we conclude that the characteristic length scale $Z_{\rm PMD}$ over which PMD becomes important is given by

$$Z_{\rm PMD} \simeq 2\frac{Z_{\rm b}^2\nu_0}{Z_{\rm c}\Delta\nu}. \tag{66}$$

In practice, accurate techniques exist for measuring $T_{\rm PMD}$ over an ensemble of frequencies [24, Chapter 6]. By contrast, $Z_{\rm c}$ is unknown and must be inferred from the measured $T_{\rm PMD}$, the measured beat length, and (66). In practice $Z_{\rm PMD}$ can vary from tens of kilometers to thousands of kilometers, depending on the bandwidth of the signal and the quality of the optical fiber. Thus, this effect acts on the length scale corresponding to δ_3. By contrast, as shown in Figure 1, the correlation length $Z_{\rm c}$ varies from meters, in the best case when the fiber is rapidly twisted as it is manufactured, to around a hundred meters. When we make the transformation $\tilde{\mathbf{U}}(z,\Omega) \to \tilde{\mathbf{W}}(z,\Omega)$ defined in (49) and just prior, we thus obtain an equation that only varies on the length scale corresponding to δ_3. In effect, we have averaged over the large but rapidly varying birefringence.

Thus, we turn to (46) in the previous section and make the substitution $\mathbf{U}(z,t) = \mathsf{R}(z)\mathsf{T}(z)\mathbf{W}(z,t)$ which leads to the equation

$$\begin{aligned}&\mathrm{i}\frac{\partial\mathbf{W}}{\partial z} - \mathrm{i}g\mathbf{W} - \tfrac{1}{2}\beta''\frac{\partial^2\mathbf{W}}{\partial t^2} - \tfrac{1}{6}\mathrm{i}\beta'''\frac{\partial^3\mathbf{W}}{\partial t^3} + \gamma|\mathbf{W}|^2\mathbf{W}\\ &\quad = -\mathrm{i}\Delta\beta'\bar{\sigma}_3\frac{\partial\mathbf{W}}{\partial t} + \tfrac{1}{3}\gamma[(\mathbf{W}^\dagger\bar{\sigma}_2\mathbf{W})\bar{\sigma}_2\mathbf{W} - \tfrac{1}{3}|\mathbf{W}|^2\mathbf{W}],\end{aligned} \tag{67}$$

where $\bar{\sigma}_2(z) \equiv \mathsf{T}^{-1}(z)\sigma_2\mathsf{T}(z)$. A nice feature of this result is that the transformation of (46) into (67) that separates out explicitly the rapidly from the slowly varying contributions is exact!

The second term on the right-hand side of (67) is referred to as nonlinear PMD and contains the effect of incomplete mixing on the Poincaré sphere. A careful analysis shows that this term is always negligible in practice for communication systems [26]. When in addition $Z_{\rm PMD}$ is longer than the system length, then the first term on the right-hand side of (67) due to the usual linear PMD can be neglected. In this case, the coupling between the two components of $\mathbf{W}$ disappears, and (67) can be replaced by a scalar equation,

$$\mathrm{i}\frac{\partial\psi}{\partial z} - \mathrm{i}g\psi - \tfrac{1}{2}\beta''\frac{\partial^2\psi}{\partial t^2} - \tfrac{1}{6}\mathrm{i}\beta'''\frac{\partial^3\psi}{\partial t^3} + \tfrac{8}{9}\gamma|\psi|^2\psi = 0. \tag{68}$$

Moreover, if $Z_{\rm PMD}$ is long compared to the nonlinear and dispersive scale lengths, then it is possible to treat it perturbatively, and in this case its effect is usually small. It should, however,

be clearly understood that the usual linear PMD plays an important role in many, although not all modern-day communication systems. When it is important, the two polarization modes in a single-mode fiber are coupled to each other through the off-diagonal, randomly varying elements in $\bar{\sigma}_3$ which is what leads to the spreading discussed earlier. This coupling cannot be transformed away [22]. Thus, (68) is not valid in this case and, consequently, neither is the nonlinear Schrödinger equation or any other scalar approximation to (68). One must instead use the full vector equation, Equation (67). There is no practical barrier to doing so since the vector equation evolves on the length scale corresponding to δ_3 and highly efficient algorithms exist to solve it.

B. The nonlinear Schrödinger equation

We now turn to the final and conceptually simplest stage of the discussion. As shown in Figure 1, the length scale on which the gain and loss act is typically 30–120 km, while the length scale on which the nonlinearity and the dispersion act is typically many hundreds or even thousands of kilometers. We can take advantage of this separation between the length scales to average over the gain-and-loss variations and dispersion variations that occur on a length scale of about 100 km or less to obtain a new averaged equation. When in addition the third-order dispersion is negligible, one obtains the nonlinear Schrödinger equation. The average over the rapidly varying dispersion is conceptually important because, when optical fibers are manufactured, there are always significant, uncontrolled fluctuations in the dispersion over length scales of hundreds of meters. This variation should be contrasted with the intentional changes in the dispersion that are used for dispersion management on a length scale of 100 km or more and which cannot be averaged over. I will return to this point shortly.

The approach presented here has been described by Mollenauer *et al.* [27] and by Doran and Blow [28]. It is based on the observation that gain and loss change the amplitude of the signal, but do not change its shape. Thus, writing

$$\psi(z,t) = u(z,t)\exp\left[\int_0^z g(z_1)\,\mathrm{d}z_1\right], \tag{69}$$

we find

$$\mathrm{i}\frac{\partial u(z,t)}{\partial z} - \tfrac{1}{2}\beta''(z)\frac{\partial^2 u(z,t)}{\partial t^2} - \tfrac{1}{6}\mathrm{i}\beta'''(z)\frac{\partial^3 u(z,t)}{\partial t^3} + \tfrac{8}{9}\gamma G(z)|u(z,t)|^2 u(z,t) = 0, \tag{70}$$

where $G(z) = \exp[2\int_0^z g(z_1)\,\mathrm{d}z_1]$. In almost all communication systems, the amplifiers are regularly spaced. The gain must integrate to zero over one amplification period Z_{amp} because otherwise the signal does not propagate stably through the system, but suffers exponential growth or loss. Amplifiers are usually operated in a regime of gain saturation to avoid this difficulty. One may now divide the transmission line into N discrete segments where $N = Z_{\mathrm{system}}/Z_{\mathrm{amp}}$ and Z_{system} is the total length of the system. Then, in the nth segment, we can write

$$G(z) = \bar{G}_n(z) + R(z), \tag{71}$$

where, for $(n-1)Z_{\mathrm{amp}} < z < nZ_{\mathrm{amp}}$,

$$\bar{G}_n(z) = \frac{1}{Z_{\mathrm{amp}}}\int_{(n-1)Z_{\mathrm{amp}}}^{nZ_{\mathrm{amp}}} \mathrm{d}z\,\exp\left[2\int_{(n-1)Z_{\mathrm{amp}}}^{z} g(z_1)\,\mathrm{d}z_1\right]. \tag{72}$$

In fiber-optic communication sytems, $g(z) = -\alpha$, where α is the attenuation coefficient, except right at $z = nZ_{\text{amp}}$, the amplifier locations, where $g(z)$ becomes large and positive. In this case, one obtains

$$\bar{G}_n(z) = \frac{1}{2\alpha Z_{\text{amp}}}[1 - \exp(-2\alpha Z_{\text{amp}})] \equiv \bar{G}, \tag{73}$$

where $\bar{G}$ is a constant. Similarly, we may write

$$\beta''(z) = \bar{\beta}_n''(z) + r(z), \tag{74}$$

where $\bar{\beta}_n''(z)$ is the average value in the nth segment and $r(z)$ is the remainder. It is often the case, as for example in dispersion-managed systems, that $\bar{\beta}_n''(z)$ varies. Then, (70) becomes

$$\mathrm{i}\frac{\partial u}{\partial z} - \tfrac{1}{2}\bar{\beta}_n''(z)\frac{\partial^2 u}{\partial t^2} + \tfrac{8}{9}\gamma\bar{G}|u|^2u = \tfrac{1}{6}\mathrm{i}\beta'''(z)\frac{\partial^3 u}{\partial t^3} - \tfrac{1}{2}r(z)\frac{\partial^2 u}{\partial t^2} - \tfrac{8}{9}R(z)|u|^2u. \tag{75}$$

When all the terms on right-hand side are negligible and, in addition, $\bar{\beta}_n''(z)$ is constant, then (75) becomes the nonlinear Schrödinger equation! Note that the nonlinearity has been effectively lowered by the factor $\bar{G}$.

Equation (75) can be used as the starting point for a multiple-length-scale analysis to obtain higher-order corrections [4, Chapter 7]. Assuming that third-order dispersion is negligible and $\bar{\beta}_n''(z)$ is constant, in order to simplify the arguments, we obtain at lowest order the nonlinear Schrödinger equation

$$\mathrm{i}\frac{\partial u}{\partial z} - \tfrac{1}{2}\bar{\beta}''\frac{\partial u}{\partial t} + \tfrac{8}{9}|u|^2u = 0. \tag{76}$$

Writing $u(z,t) = u^{(0)}(z,t) + u^{(1)}(z,t) + \cdots$, where $u^{(0)}(z,t)$ is the solution of (76), we then find that

$$\begin{aligned}\mathrm{i}\frac{\partial u^{(1)}}{\partial z} - \tfrac{1}{2}\beta''\frac{\partial^2 u^{(1)}}{\partial t^2} + \tfrac{16}{9}\gamma\bar{G}|u^{(0)}|^2u^{(1)} + \tfrac{8}{9}\gamma\bar{G}[u^{(0)}]^2u^{(1)*}\\ = -\tfrac{1}{2}r(z)\frac{\partial^2 u^{(0)}}{\partial t^2} - \tfrac{8}{9}\gamma R(z)|u^{(0)}|^2u^{(0)}.\end{aligned} \tag{77}$$

The solution to (77) yields a rapidly oscillating correction to $u^{(0)}$. Additionally, it can also yield a slow secular growth because $u^{(0)}$ changes slightly over one amplifier period. At the next order, secularities are inevitable. When considering specific solutions to (76), like soliton solutions, it is possible to modify the expansion of $u(z,t)$ in a way that removes the secularities, and this issue is thoroughly discussed by Hasegawa and Kodama [4, Chapter 7]. For our purposes here it is sufficient to note that when higher-order corrections including the secular contributions are negligible, then the nonlinear Schrödinger equation is a valid model of the signal evolution in the optical fiber. It bears emphasis that, while the averaging approach just described was first applied to soliton systems, I have made no mention of solitons in the derivation of the nonlinear Schrödinger equation, and the result is equally valid for any communication format.

The derivation that has just been given for the nonlinear Schrödinger equation begs the question of determining when the higher-order corrections that were neglected become important. If the separation between the length scale on which the gain and the loss occurs and on which the nonlinearity and dispersion act had been many orders of magnitude, then we would expect the corrections to be negligible, but in practice these scales are only separated by about a factor of ten, so that the spatially varying gain and loss often lead to important changes in the evolution from what the nonlinear Schrödinger equation would predict. It is standard, therefore, when using simulations to design communication systems, to use (68) or possibly (67), often with the effects of spontaneous emission noise added. We should not conclude from this situation, however, that averaging over the gain and loss is useless. The nonlinear Schrödinger equation has many special properties, and a large amount of theoretical work has been carried out based on it – particularly for solitons – that has yielded important insights for communication systems.

As a final point, I turn to consideration of the case in which $\bar{\beta}_n''(z)$ varies along the optical-fiber transmission line. It has been found in practice that it is very beneficial to use large values of dispersion with alternating signs in an optical fiber transmission system since the alternating expansion and contraction of the signal as well as its spatially varying chirp serves to reduce the deleterious effects of the nonlinearity. In practical systems, the greatest advantage is obtained when the dispersion in each leg of the dispersion map is large enough, so that its scale length is roughly comparable to the map length, but not larger. However, we must analyze these systems using numerical techniques, and it is of conceptual interest at least to consider what happens when the dispersive scale length is short compared to the map length so that the local dispersion dominates the evolution. In this limit, one can apply multiple-scale-length techniques directly to (77), letting β'' vary as a function of z. This approach has recently been used by Bronski and Kutz [29] to analyze the onset of the modulational instability in nonreturn-to-zero systems and independently by Turitsyn [30] and by Kodama *et al.* [31] to investigate the existence of very-long-term radiation in dispersion-managed soliton systems. Clearly, the story is far from finished!

Acknowledgment

Work at the University of Maryland Baltimore County was supported by the Department of Energy and the National Science Foundation. The author thanks D. Marcuse for useful comments on an earlier version of this paper.

References

1. N. Bloembergen, *Nonlinear Optics*. Reading (MA): Benjamin (1965) 229 pp.
2. Y. R. Shen, *The Principles of Nonlinear Optics*. New York: Wiley (1984) 563 pp.
3. G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego: Academic (1995) 592 pp.
4. A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*. Oxford: Clarendon (1995) 320 pp.
5. A. Hasegawa and F. D. Tappert, Transmission of stationary nonlinear pulses optical pulses in dispersive dielectric fibers. I. Anomalous dispersion. *Appl. Phys. Lett.* 23 (1973) 142–144.
6. C. R. Menyuk, Pulse propagation in an elliptically birefringent Kerr medium. *IEEE J. Quantum Electron.* 25 (1989) 2674–2682.
7. C. R. Menyuk, Impairments due to nonlinearity and birefringence in optical fiber transmission systems. In: A. E. Willner and C. R. Menyuk (eds), *System Technologies*. Washington: Optical Society of America TOPS Vol. 12 (1997) 523 pp.

8. S. B. Alexander, The WDM Revolution. In: A. E. Willner and C. R. Menyuk (eds), *System Technologies*. Washington: Optical Society of America TOPS Vol. 12 (1997) 523 pp.
9. Examples of currently available commercial software packages are BroadNeD and OPALS by Virtual Photonics F.C.
10. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste, tome I*. Paris: Gauthier-Villars (1892) 385 pp.
11. A large number of such problems can be found cited by H. Minorsky, *Nonlinear Oscillations*. Princeton (NJ): Van Nostrand (1962) 714 pp.; A. Nayfeh, *Perturbation Methods*. New York: Wiley (1973) 425 pp.
12. Y. Kodama, Optical solitons in a monomode fiber. *J. Stat. Phys.* 39 (1985) 597–614.
13. P. Appell, *Henri Poincaré*. Paris: Plon-Nourrit (1925) 121 pp.
14. J. Gowar, *Optical Communication Systems*. New York: Prentice Hall (1993) 696 pp.
15. C. R. Menyuk, D. Wang and A. N. Pilipetskii, Repolarization of polarization-scrambled signals due to polarization dependent loss. *IEEE Photon. Technol. Lett.* 9 (1997) 1247–1249.
16. A. R. Chraplyvy, Limitations on lightwave communications imposed by optical-fiber nonlinearities. *J. Lightwave Technol.* 8 (1990) 1548–1557.
17. R. W. Hellwarth, Third-order optical susceptibilities of liquids and solids. *Prog. Quant. Electron.* 5 (1977) 1–68.
18. C. R. Menyuk, M. N. Islam and J. P. Gordon, Raman effect in birefringent optical fibers. *Opt. Lett.* 16 (1991) 566–568.
19. R. H. Stolen, J. P. Gordon, W. J. Tomlinson and H. A. Haus, Raman response function of silica core fibers. *J. Opt. Soc. Am. B* 6 (1989) 1159–1166.
20. P. K. A. Wai and C. R. Menyuk, Polarization decorrelation in optical fibers with randomly varying birefringence. *Opt. Lett.* 19 (1994) 1517–1519.
21. P. K. A. Wai and C. R. Menyuk, Anisotropic diffusion of the state of polarization in optical fibers with randomly varying birefringence. *Opt. Lett.* 20 (1995) 2490–2492.
22. C. R. Menyuk and P. K. A. Wai, Polarization evolution and dispersion in fibers with spatially varying birefringence. *J. Opt. Soc. Am. B* 11 (1994) 1288–1296.
23. P. K. A. Wai and C. R. Menyuk, Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence. *J. Lightwave Technol.* 14 (1996) 148–157.
24. C. D. Poole and J. Nagel, Polarization effects in lightwave systems. In: I. P. Kaminow and T. L. Koch (eds), *Optical Fiber Telecommunications, Vol. IIIA*. San Diego: Academic (1997) 608 pp.
25. C. D. Poole, Statistical treatment of polarization dispersion in single-mode fiber. *Opt. Lett.* 13 (1988) 687–689.
26. D. Marcuse, C. R. Menyuk and P. K. A. Wai, Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence. *J. Lightwave Technol.* 15 (1997) 1735–1746.
27. L. F. Mollenauer, S. G. Evangelides, Jr. and H. A. Haus, Long-distance soliton propagation using lumped amplifiers and dispersion shifted fiber. *J. Lightwave Technol.* 9 (1991) 194–197.
28. K. J. Blow and N. J. Doran, Average soliton dynamics and the operation of soliton systems with lumped amplifiers. *IEEE Photon. Technol. Lett.* 3 (1991) 369–371.
29. J. C. Bronski and J. N. Kutz, Modulational stability of plane waves in nonreturn-to-zero communications systems with dispersion management. *Opt. Lett.* 21 (1996) 937–939.
30. S. K. Turitsyn, Theory of average pulse propagation in high-bit-rate optical transmission systems with strong dispersion management. *JETP Lett.* 65 (1997) 845–851, (*Pisma Zh. Eksp. Teor. Fiz.* 65 (1997) 812–817).
31. Y. Kodama, S. Kumar and A. Maruta, Chirped nonlinear pulse propagation in a dispersion-compensated system. *Opt. Lett.* 22 (1997) 1689–1691.