Chapter 3

COMPARISON OF OPTICAL MODULATION FORMATS

It is seen from Chapter 2 that impairments caused by nonlinearity, such as XPM and FWM, can be minimized provided that the local dispersion of the transmission fiber or channel spacing is sufficiently large. Due to the fact that the effect of dispersion is a linear process, the dispersion accumulated along the transmission fiber can be equalized by performing dispersion compensation either periodically along the link or at the receiver end. This, however, increases system complexity compared with systems employing optical fibers whose dispersion is so small that the dispersion compensation is not necessary. On the other hand, large channel spacing means than the available bandwidth cannot be utilized effectively. The main purpose of spacing channels sufficiently far apart or employing fibers with adequate dispersion is to suppress the interchannel impairments. Still, the intrachannel interaction between dispersion and nonlinearity causes performance degradation on a given channel. Although the dispersion can be compensated, the fiber nonlinearity cannot. The interaction between dispersion and nonlinearity along the main transmission fiber causes waveform degradation even when dispersion compensation is employed. However, intrachannel performance degradation can be minimized if the transmitted signal possesses some special characteristic. This leads to the choice of novel of optical modulation formats.

As the trend in the channel bit rate is to move from 10 Gb/s to 40 Gb/s, optical modulation formats become more important since the performance degradation caused by the interaction between the dispersion and nonlinearity increases with the channel bit rate. It should be noted, however, that the receiver is still the square law detector, which is insensitive to the signal phase. The photodetector converts the received optical signal to electrical signal. The received electrical signal is then amplified, and compared with the threshold to decide whether bit 1 or bit 0 was transmitted. To facilitate the synchronization between the received signal and the local clock at the receiver, the impairments caused by dispersion and nonlinearity, such as timing jitter and amplitude

jitter, have to be minimized. The purpose of this chapter is to compare different optical modulation formats in terms of signal characteristics, system implementation, and power spectrum. Different modulation formats have different characteristics, which lead to unique performance improvements. This chapter begins with a discussion of the benefits of the RZ signal format over the NRZ format. It is then followed by the intrachannel impairments in systems with/without dispersion compensation in Section 3.2. Section 3.3 is a theoretical analysis of pulse chirping when fiber dispersion, nonlinearity and loss are considered. Chirping the pulse is actually the first kind of optical modulation format considered, and is discussed in Section 3.4. Two variants of chirp are discussed in Section 3.4: sinusoidal same phase modulation (SaPM) and sinusoidal alternate phase modulation (APM). The remaining sections are dedicated to other modulation formats, including square-wave phase modulation (SWM), optical duobinary, and alternate mark inversion (AMI). The SWM signal format is discussed in Section 3.5 whereas the optical duobinary and the AMI signal format are discussed together in Section 3.6. All of the above techniques have been considered in the literature. A purpose of this dissertation is not only to compare these alternative techniques, but to propose and evaluate a new signal format, continuous-wave square-wave (CWSW) which is discussed in a separate chapter.

3.1 ADVANTAGES OF RETURN-TO-ZERO (RZ) SIGNAL FORMAT

In the past, the optical modulation format employed in lightwave communication systems was non-return-to-zero (NRZ). At that time, the most deleterious effect of an optical fiber was the dispersion, which increases with the bit rate. However, at high bit rate, such as 40 Gb/s, the effect of nonlinearity becomes more important, and the RZ signal format is proving to be superior to NRZ at such high bit rates. The advantages of the RZ signal format over the NRZ are discussed below.

Due to the fact that the RZ signal format has larger bandwidth than the NRZ, the RZ pulses are broadened more rapidly by dispersion. However, this turns out to be beneficial because as a pulse is broadened, its peak decreases. The severe pulse broadening in the case of RZ signals makes it more robust to the effect of nonlinearity [74]-[76] due to the fact that the nonlinear effect is proportional to the signal intensity.

This is very important in high-bit-rate systems in which high launched power is required to provide adequate SNR at the receiver. The effect of accumulated dispersion can be removed by employing dispersion compensation, which in effect recovers the waveform back to its original form. Since the pulse width of the RZ signal is narrower than that of the NRZ signal, the RZ pulse has higher peak power than the NRZ for a given average power. Thus, the eye opening of the RZ signal format is wider than that of the NRZ, resulting in better receiver sensitivity than the NRZ for a given average power [77]-[79]. This implies that for a required receiver sensitivity, the transmitted power can be lowered by employing the RZ signal format rather than the NRZ. The better receiver sensitivity in the case of the RZ signal also suggests that the transmission distance can be increased compared with the NRZ signal for the same transmitted power [80]. When the SPM is considered, its interaction with dispersion depends strongly on the pulse width. In the case of the NRZ signal format, the transmitted signal consists of pulses having different pulse widths depending on the data pattern. Thus, the effect of SPM depends on the data pattern. On the other hand, the transmitted RZ signal consists of a sequence of identical pulses corresponding to the data pattern; hence, the pattern-dependence of SPM-induced waveform distortion can be avoided [81]. In addition, the interaction between the SPM and dispersion in the anomalous dispersion regime can be exploited effectively in the case of the RZ signal, resulting in uniform pulse compression.

The advantages of the RZ signal over the NRZ are not limited to reducing the intrachannel nonlinear effect. In fact, the RZ signal format is significantly superior to NRZ when the interchannel nonlinear effects are considered. As discussed in Chapter 2, the XPM and FWM are strongest when pulses at different wavelengths completely overlap one another in the time domain. This condition can be easily satisfied in the case of the NRZ signal format. Generally, the pulse width of the RZ signal is shorter than that of the NRZ signal. Thus, the NRZ pulses at different wavelengths take longer time than the RZ pulses to walk off from one another. This in effect favors the generation of FWM [82], [83], and XPM [84]-[86]. Therefore, NRZ systems are more susceptible to interchannel nonlinear effects than RZ systems. It should be noted that NRZ systems are more robust to linear cross talk among channels than the RZ systems due to the fact that the NRZ signal has narrower bandwidth than RZ signal. However, with proper spectral

filtering or pulse shaping at the transmitter to remove unwanted high-frequency components, the linear cross talk among channels in RZ systems can be minimized, causing RZ systems to further outperform NRZ systems [76], [87], [88].

Shown in Fig. 3.1 is the schematic diagram of the transmitter for RZ systems. The continuous wave (CW) output signal from a laser is externally modulated by NRZ data. Note that in a semiconductor laser diode, direct amplitude modulation causes a change in the refractive index of the laser [1]; thus, amplitude modulation is always associated with phase modulation. In addition, the bandwidth of the laser does not permit modulation at high bit rate. Therefore, external modulation is generally used instead of direct modulation on the laser when the bit rate is high. The role of the RZ modulator is to perform NRZ-to-RZ conversion, and the output signal from the RZ modulator is a conventional RZ signal. It should be noted that the operation of the data modulator and RZ modulator is mathematically equivalent to multiplying the output CW signal from the laser by NRZ data and by a RZ pulse train; thus, the order of the operations does not matter. Shown in Fig. 3.1 (b) is the power spectral density (PSD) of the RZ signal relative to the carrier frequency. Note that the frequency is normalized by the bit rate. The PSD of the signal is calculated by taking the Fourier transform of the autocorrelation of the signal. The spectrum of the RZ signal contains the carrier and two sidebands spaced away from the carrier on each side by the bit rate. The existence of the carrier is due to the fact that all pulses contained in the signal have identical sign, which contributes to a DC signal when the baseband-equivalent model is considered

It is clearly seen in this section that the RZ signal format provides significant advantages over the NRZ format in many system performance aspects. Still, the RZ signal format suffers the impairments caused by intrachannel interaction between dispersion and nonlinearity. To further enhance system performance, those impairments have to be minimized. However, it is important to understand the mechanisms that cause the performance degradation. The sources of intrachannel performance degradation are discussed in the next section.

3.2 INTRACHANNEL IMPAIRMENTS

It is widely known that dispersion alone causes a pulse to spread. In RZ systems in which an isolated pulse is used for carrying bit 1, a pulse can spread beyond its own bit slot due to the dispersion. The overlap between pulses and pulse spreading cause intersymbol interference (ISI). An illustration of ISI is shown in Fig. 3.2. The dashed curves corresponds to the intensity of individual pulses broadened by dispersion, and the solid curve is the resultant intensity of those two broadened pulses when they interfere with each other. As seen from Fig. 3.2 (a) dispersion causes pulses to overlap each other. Because those two pulses are identical, the constructive interference occurs midway between the two original pulses, resulting in the growth of a spurious pulse. In addition, the overlap between the tail of one pulse and the center portion of another pulse shifts the peak away from the center of the original pulse as shown in Fig. 3.2 (a).

If those two pulses represent the bit 1's surrounding a bit 0 on each side (bit pattern 101), the growth of the spurious pulse caused by the overlap between the surrounding pulses increases the signal level for bit 0, hence decreasing the eye opening at the receiver. On the other hand, when the two pulses shown in Fig. 3.2 (a) represents two adjacent bit 1's (bit pattern 11), the overlap between two pulses results in the growth of a spurious pulse, which in effect can distort the eye profile so that it is no longer RZ. The situation is opposite when two pulses have opposite signs (positive and negative) as shown in Fig. 3.2 (b). When two pulses have opposite signs, the growth of a spurious pulse is suppressed due to destructive interference. If those two pulses represents the bit 1's surrounding a bit 0 on each side (bit pattern 101), the ISI in the bit slot 0 is minimized. When bit pattern 11 is considered, the RZ eye profile is preserved when those two pulses representing bit 1's have opposite signs. In addition, constructive interference between the tail of one pulse and the center portion of another pulse helps maintain the signal level of the resultant signal at the centers of the original pulses although the peak is shifted from its original location. This implies that if adjacent pulses carrying bit 1's have opposite signs (180° phase shift at the boundary between the two pulses), the growth of spurious pulses can be prevented, hence minimizing the ISI and improving the system performance. In fact, this is the mechanism by which the SWM, optical duobinary and

AMI signal formats improve system performance. The practical implementation of providing 180° phase shift between two adjacent pulses is discussed later in this chapter.

It is clearly seen that for systems with no dispersion compensation, performance degradation is mainly due to the growth of a spurious pulse caused by overlap between pulses. Ideally, if the optical fiber is linear, the effect of dispersion, hence the growth of a spurious pulse as shown in Fig. 3.2 (a), can be completely removed by performing dispersion compensation. At the fiber output, the signal waveform would be identical to that at the fiber input except for its intensity due to fiber loss, and the system performance would be limited only by fiber loss. This suggests that the transmission distance can be increased infinitely as long as the loss and dispersion along the transmission link are compensated. Unfortunately, the fiber nonlinearity prevents perfect dispersion compensation although the loss along the transmission link can be compensated by either in-line amplifiers or transmitting as high power as possible. Note that in-line amplifiers introduce noise into the system; hence, there exists a noise-limited maximum transmission distance. In systems employing dispersion compensation either at the receiver or periodically along the transmission link, the nonlinearity gives rise to two effects called intrachannel four-wave mixing (IFWM), and intrachannel cross-phase modulation (IXPM). The terms for these phenomena are borrowed from the interchannel nonlinear interactions among channels with the frequency-related parameters being replaced by time-related parameters. Both phenomena are discussed in the following sections.

3.2.1 Intrachannel Four-Wave Mixing (IFWM)

In systems in which the accumulated dispersion is compensated, the intrachannel four-wave mixing (IFWM) among pulses, which overlap one another during the propagation, results in the growth of spurious (ghost) pulses [89]-[91]. The energy is transferred from the original pulses involved in the IFWM process to the spurious (ghost) pulses, and the energy transfer increases linearly with the transmission distance [92], [93]. Fig. 3.3 is an illustration of IFWM among pulses. The dashed curves in the upper half of Fig. 3.3 correspond to the intensities of original pulses at the fiber input whereas the solid curves are the intensities of dispersed pulses. The lower half of Fig. 3.3 is the

resultant intensity of the pulses after dispersion compensation is performed. It is clearly seen that the IFWM among pulses surrounding an empty bit slot 0 results in the generation of a spurious (ghost) pulse in the bit slot 0. The locations of spurious pulses caused by IFWM among overlapping pulses can also coincide with the original pulses themselves, which leads to amplitude jitter as shown in the lower half of Fig. 3.3.

Unlike the FWM among WDM channels, large local dispersion increases the IFWM effect [94], [95]. This can be intuitively understood from the fact that large local dispersion implies strong overlap among pulses, which in effect increases the number of interacting pulses. However, it should be noted that the strength of the IFWM also depends on the phase matching among overlapping pulses. It has been theoretically shown in [96] that constant-phase inversion between consecutive bit slots minimizes the effect of IFWM due to the reduction of phase matching among interacting pulses, and the optimum phase inversion is found to be equal to 180° (consecutive bit slots have opposite signs) [97], [98]. In addition, the growth of a spurious pulse in an empty bit slot 0 can be suppressed if the signs of a pair of interacting pulses that are equidistance from the bit slot 0 are opposite [99]. It should be noted that although the growth of a spurious pulse in bit slot 0 is suppressed, timing jitter still occurs.

3.2.2 Intrachannel Cross-Phase Modulation (IXPM)

The overlap between pulses during propagation not only causes the IFWM, discussed in the previous section, but also gives rise to the intrachannel cross-phase modulation (IXPM). The physical origin of the IXPM can be understood as follows. Recall that the instantaneous-frequency shift across the pulse depends on the intensity derivative of the pulse during propagation due to the intensity-dependence of refractive index. When overlapping pulses are considered, the intensities of the overlapping tails of interfering pulses also affect the instantaneous-frequency shifts across the considered pulse, which results in the shift in the mean frequency of the considered pulse. Due to the dispersion, the shift in the mean frequency causes the interfered pulse to travel at a speed different from that of an isolated pulse, hence leading to timing jitter [100], [101]. It should be noted that this effect is similar to interchannel XPM in WDM systems with the additional frequency shifts being caused by the neighboring pulses instead of neighboring

channels. Moreover, the IXPM does not depend on the relative phase of the interacting pulses, and does not result in energy fluctuation among interacting pulses. (IFWM is responsible for the intrachannel nonlinear interaction, which depends on the relative phase of interacting pulses, and causes energy transfer among interacting pulses.) No energy transfer among interacting pulses in the case of IXPM can be understood from the fact that the interfering pulses only disturb the phase, not the amplitude, of the considered pulse, which is similar to interchannel XPM.

When two pulses are considered, the mean frequency shift of an interacting pulse caused by IXPM is proportional to [93]

$$\frac{d\Delta f_{XPM}}{dz} \sim F(x) = \frac{\exp(-2/x^2)}{x^3} \tag{3.2}$$

where x is the ratio between the pulse width and the pulse separation. The magnitude of $F(x)$ as a function of x is shown in Fig. 3.4. Generally, the effect of IXPM depends on two factors: the level of pulse overlap and the intensity derivative of the interacting pulses. The intensity derivative in this context means the first derivative of pulse intensity. As interacting pulses are dispersed during propagation, their intensity derivatives are decreased while the pulse overlap region is increased. Therefore, when pulses strongly overlap each other ($x >> 1$), the intensity derivative of the pulses involved in the IXPM process is small due to the pulses being severely dispersed. On the other hand, the intensity derivative of the pulses is strong when the pulse overlap is small ($x << 1$). Hence, the IXPM is strongest when pulses partially overlap ($x \approx 1$) due to the compromise between the intensity derivatives of the interacting pulses and the level of overlap. This suggests that in order to minimize the effect of IXPM, the local dispersion of the transmission fiber should be sufficiently large so that the interacting pulses are dispersed rapidly, thus reducing their interactions. However, large local dispersion implies strong IFWM due to strong pulse overlap. Thus, there is the trade-off between the suppressions of IXPM and IFWM.

It is clearly seen that the performance degradations are mainly due to the nonlinear effects that occur when there is overlap between dispersed pulses regardless of whether dispersion compensation is deployed. Another technique that can minimize the overlap between pulses is to make use of the pulse compression effect. In a nonlinear

optical fiber, the pulse compression effect can be achieved when the operating wavelength is in the anomalous regime in which the effect of SPM counteracts dispersion. In this regime, if the effect of the SPM is sufficiently strong, the pulse width is reduced during the initial propagation and then increased again. Thus, the overlap between pulses can be reduced if the transmission distance is still within the range in which pulse compression is effective. However, the pulse compression effect caused by the interaction between the dispersion and SPM does not exist in the normal dispersion regime. Moreover, a pulse is broadened more severely than under the presence of dispersion alone in the normal dispersion regime. Generally, the effect of the dispersion and SPM is to alter the phase of the pulse. Thus, if the phase variation across the pulse is intentionally manipulated at the transmitter to counteract dispersion (and SPM, if the operating wavelength is in the normal dispersion regime), the pulse compression effect can be achieved. This technique is generally referred to as prechirping, and it is the mechanism used by the SaPM and APM signal formats to improve system performance. It should be noted that the initial phase variation imposed onto a pulse at the transmitter also interacts with the SPM, which depends on the operating wavelength regime and the fiber loss. This makes the prechirping technique complicated. In the next section, theoretical analysis of prechirping under the presence of fiber loss and nonlinearity is performed. The main purpose is to find the maximum transmission distance and corresponding optimum prechirp at which the prechirp technique can restore the output pulse width to its initial value when an optical fiber is not lossless.

3.3 EFFECT OF FIBER LOSS ON OPTIMUM CHIRP FOR DISPERSIVE NONLINEAR FIBER

Generally, pulse propagation in an optical fiber is governed by the nonlinear Schrödinger equation (NLSE) [57]. This equation is a nonlinear partial differential equation, whose exact solution cannot be found in closed form except in some special cases. Thus, numerical methods are generally used to approximate the solution. However, the numerical methods do not yield insight in understanding how the optical fiber affects the propagation of the optical signal. A breakthrough came in 1983 when D. Anderson [102] showed that the NLSE could be approximated by using the variational method

[103] to get a closed-form solution. In this section, we utilize the variational method to find a closed-form expression of the pulse width for a chirped Gaussian pulse. From that expression, we can further analytically evaluate the optimum prechirp that yields the maximum transmission distance at which the output pulse width is equal to the input pulse width. M. Brandt-Pearce *et al* have derived the closed form expression of the pulse width by using the variational method [104]. However, that work is under the assumption that an optical fiber is lossless. This would be true in a system with distributed optical amplifier repeaters where the optical amplifier gain compensates for the fiber attenuation. This is not true in practice where lumped gain amplifiers are used, and the signal power varies between amplifiers. Previous work [104] considers constant power equal to the average power between amplifiers. The purpose here is to take into account the fact that the signal power is not constant along the transmission path. That is, dispersion, SPM, and attenuation are all taken into account. The results are verified with split-step Fourier (SSF) simulation [57].

3.3.1 Nonlinear Schrödinger Equation (NLSE)

Since the optical field that travels along an optical fiber is in fact an electromagnetic field, its behavior is governed by the well-known Maxwell's equations. The expression of the pulse envelope of the optical field can be derived from the Maxwell's equations, and is governed by [57]

$$i\frac{\partial A}{\partial z} = -\frac{i}{2}\alpha A + \frac{1}{2}\beta_2 \frac{\partial^2 A}{\partial T^2} - \gamma |A|^2 A, \tag{3.2}$$

where $A(z,T)$ is the slowly-varying amplitude of the pulse envelope, α is the attenuation constant, β_2 is the group-velocity dispersion (GVD) parameter, γ is the nonlinear coefficient, and T is a retarded time which is given by

$$T = t - \frac{z}{v_g}, \tag{3.3}$$

where t is time, z is the distance of interest along a fiber link, and v_g is the group velocity. It should be noted that (3.2) is derived under the assumption that the bandwidth

of the pulse $\Delta\omega$ is small compared to the carrier frequency ω_0, and that β_2 is far enough from $\beta_2 = 0$ that β_3 is negligible in comparison.

When $\Delta\omega << \omega_0$ is satisfied, β_2 can be obtained by using a Taylor series to expand the mode-propagation constant $\beta(\omega)$ about the carrier frequency ω_0. That is, [57]

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \ldots, \tag{3.4}$$

where

$$\beta_m = \left(\frac{d^m \beta}{d\omega^m}\right)_{\omega=\omega_0}, \tag{3.5}$$

where m =0, 1, 2, … . Note that $v_g = 1/\beta_1$. Moreover, β_2 can be related to the commonly-used dispersion coefficient D of an optical fiber by (2.4). The nonlinear coefficient γ, can be expressed as a function of physical parameters of a optical fiber as

$$\gamma = \frac{n_2 \omega_0}{c A_{eff}}, \tag{3.6}$$

where n_2 is the nonlinear-index coefficient, A_{eff} is the effective core area of a fiber.

It is more useful to use the normalized form of the pulse envelope $A(z,T)$ in (3.2). First, we define $A(z,T)$ as

$$A(z,T) = \sqrt{P_0}\exp\left(-\frac{\alpha z}{2}\right)U(z,T), \tag{3.7}$$

where P_0 is the peak power of the pulse at $z = 0$, and $U(z,T)$ is the normalized pulse envelope. By substituting $A(z,T)$ given by (3.7) into (3.2), we obtain

$$i\frac{\partial U}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 U}{\partial T^2} - \gamma P_0 e^{-\alpha z}|U|^2 U. \tag{3.8}$$

This is the NLSE for a normalized pulse envelope $U(z,T)$.

3.3.2 Methodology

In order to understand the effect of loss, fiber nonlinearity, and dispersion on a pulse traveling along an optical fiber, (3.8) has to be solved. Unfortunately, (3.8) is a

nonlinear partial differential equation that cannot be solved analytically without some simplification; thus, a numerical approach is usually used. However, a numerical approach gives only limited insight to understanding the impact of initial pulse design on the normalized pulse envelope $U(z,T)$.

There is another approach that can be used to approximate the solution of (3.8). It is called the variational method. The classical variational method works by replacing certain equations of mathematical physics by Lagrangians, integrals of which often represent quantities to be maximized or minimized. The integrals have the form

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[L(x,y,w,w_x,w_y)\right]dxdy\,, \tag{3.9}$$

where

$$w = w(x,y)\,, \tag{3.10}$$

$$w_x = \frac{\partial w(x,y)}{\partial x}\,, \tag{3.11}$$

$$w_y = \frac{\partial w(x,y)}{\partial y}\,. \tag{3.12}$$

It can be shown that (3.9) is extremized with respect to w when $f = L$ satisfies Euler's equation

$$\frac{\partial f}{\partial w} - \frac{\partial}{\partial x}\frac{\partial f}{\partial w_x} - \frac{\partial}{\partial y}\frac{\partial f}{\partial w_y} = 0\,, \tag{3.13}$$

where

$$f = f(x,y,w,w_x,w_y)\,. \tag{3.14}$$

The problem then becomes finding a Lagrangian L such that the NLSE is Euler's equation corresponding to that Lagrangian. D. Anderson [102] has found a Lagrangian version of the NLSE in the form

$$L\left(z,T,U^*(z,T),U^*{}_z,U^*{}_T\right) = \frac{i}{2}\left(UU_z^* - U^*U_z\right) - \frac{\beta_2}{2}\left|U_T\right|^2 - \frac{\gamma P_0 e^{-\alpha z}}{2}\left|U\right|^4, \tag{3.15}$$

where $U^*(z,T)$ is the complex conjugate of $U(z,T)$. Substituting (3.15) into (3.9), and using the variables corresponding to (3.8), (3.9) becomes

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[L(z,T,U^*(z,T),U^*{}_z,U^*{}_T)\right]\,dzdT\,. \tag{3.16}$$

Note that the Euler's equation is a necessary condition for L to extremize J, given by (3.16). If Euler's equation were also a sufficient condition for an extremum of J, then such extrema would always be solutions of the NLSE. In practice, however, the extrema are only approximations. Moreover, one cannot generally calculate extrema of J unless special forms for $U(z,T)$ are assumed. The process of extremalizing over special classes of $U(z,T)$ is known as the 'Ritz' method. In the 'Ritz' method, $U^*(z,T)$ that extremizes (3.16) is the approximate solution to the Euler's equation, which is the NLSE for the Lagrangian given by (3.15). In this analysis, we will postulate that the pulse envelope $U(z,T)$ has a Gaussian shape, the characteristic of which is usually governed by three parameters: pulse amplitude $G(z)$, pulse width $a(z)$, and chirp $b(z)$. The problem is then reduced to finding the values of these parameters that extremalize J in (3.16).

Following D. Anderson [102], we will then cite governing equations for $G(z)$, $a(z)$, and $b(z)$ based on extremalizing J. By performing the integration with respect to T in (3.16), the expression for J then contains only a single integral with respect to z. The values of $G(z)$, $G^*(z)$, $a(z)$, and $b(z)$ that extremize J can be found by differentiating J with respect to these parameters and setting the results equal to 0. The resultant expressions are coupled differential equations, which are then used as a basis to find the analytical expressions of parameters that govern the pulse envelope. While this variational method is only partially rigorous, it does provide relatively simple approximations, which will be provided later and compared with numerical solutions of the NLSE. Note that variational approximation, as used here, assumes a pulse remains Gaussian, and only $G(z)$, $a(z)$, and $b(z)$ change with distance z.

3.3.3 Gaussian Pulse Envelope

The normalized pulse envelope $U(z,T)$ used in this analysis is assumed to be Gaussian. In fact, it is not the only possible pulse shape that may be used in the

variational method [102]. A justification for using a Gaussian pulse is that an initial Gaussian pulse remains Gaussian as it propagates in a linear fiber [57]. For this reason, the approximate solution of the NLSE obtained from the variational method is actually exact in a linear fiber with a Gaussian input pulse. By continuity, the approximation is still valid when the nonlinearity is not too large. In practice, the approximation holds because effective nonlinearity decreases with the increase in distance z. When $U(z,T)$ is a Gaussian pulse with chirp, it is governed by

$$U(z,T) = G(z)\exp\left[-T^2\left(\frac{1}{2a^2(z)} - ib(z)\right)\right], \tag{3.17}$$

where $G(z)$ is the complex pulse amplitude with $G(0)=1$, $a(z)$ is a pulse width, and $b(z)$ is the chirp parameter at a distance z. It is appropriate at this point to introduce the definition of the RMS half-width of a pulse envelope, which is defined as [57]

$$\sigma(z) = \left[< T^2 > - < T >^2\right]^{1/2}, \tag{3.18}$$

where

$$< T^n >= \frac{\int_{-\infty}^{\infty} T^n \left|U(z,T)\right|^2 dT}{\int_{-\infty}^{\infty} \left|U(z,T)\right|^2 dT}. \tag{3.19}$$

By using (3.18) together with (3.17) and (3.19), the RMS half-width $\sigma(z)$ is related to $a(z)$ by

$$\sigma(z) = a(z)/\sqrt{2}. \tag{3.20}$$

By substituting $U(z,T)$ given by (3.17) into (3.16), and following the procedure outlined in the previous section to determine the values of $G(z)$, $G^*(z)$, $a(z)$, and $b(z)$ that extremize (3.16), the following expressions are obtained.

$$\frac{da(z)}{dz} = -2\beta_2 a(z)b(z) \tag{3.21}$$

$$\frac{db(z)}{dz} = 2\beta_2 b^2(z) - \frac{\beta_2}{2a^4(z)} - \frac{\gamma P_0 e^{-\alpha z} E_0}{2\sqrt{2}a^3(z)} \tag{2.22}$$

$$E_0 = a(z)\left|A(z)\right|^2 = \text{constant} \tag{3.23}$$

By further differentiating (3.21) with the help of (3.22), the second derivative of $a(z)$ with respect to z is given by

$$\frac{d^2a(z)}{dz^2} = \frac{{\beta_2}^2}{a^3(z)} + \frac{\beta_2\gamma P_0 e^{-\alpha z}E_0}{\sqrt{2}\ a^2(z)}. \tag{3.24}$$

Note that three key equations that are used to find the expression of $a(z)$ are (3.21), (3.23), and (3.24). To solve these equations, firstly $a^2(z)$ is expanded into a power series, which is given by

$$a^2(z) = c_0 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4 + \ldots . \tag{3.25}$$

The coefficients of the power series of $a^2(z)$ in (3.25) can then be found by using (3.21), (3.23), and (3.24). As seen from (3.25), there are an infinite number of terms in the power series of $a^2(z)$. However, in this analysis the power series will be expanded up to the fourth order of z (up to the fifth term in (3.25)). It can easily be shown that if the order of the power series is less than three, the effect of fiber loss, which is represented by α, does not appear in the expression of $a^2(z)$, which will be seen later in this analysis. The reasons for including terms up to fourth order will be discussed later.

The first coefficient of the power series given by (3.25) is c_0. When $z = 0$, (3.25) becomes

$$c_0 = a_0^2, \tag{3.26}$$

where $a_0 = a(0)$. In order to find subsequent coefficients in (3.25), (3.25) is recursively differentiated with respect to z, and then z is set to zero after each differentiation. With the help of (3.21), (3.23), and (3.24), the expression of $a^2(z)$ normalized by a_0^2 up to the fourth order in z is finally obtained, and is given by

$$\begin{aligned}\frac{a^2(z)}{a_0^2} = 1 &+ \left[2\,\mathrm{sgn}(\beta_2)C\right]\left(\frac{z}{L_D}\right) + \left[1 + \frac{\mathrm{sgn}(\beta_2)N^2}{\sqrt{2}} + C^2\right]\left(\frac{z}{L_D}\right)^2 \\ &+ \left[-\frac{\mathrm{sgn}(\beta_2)\alpha L_D}{3\sqrt{2}} + \frac{C}{3\sqrt{2}}\right]N^2\left(\frac{z}{L_D}\right)^3 \\ &+ \left[\frac{\mathrm{sgn}(\beta_2)\alpha^2 L_D^2}{12\sqrt{2}} - \frac{\mathrm{sgn}(\beta_2)C^2}{6\sqrt{2}} + \frac{N^2}{24} + \frac{\mathrm{sgn}(\beta_2)}{12\sqrt{2}}\right]N^2\left(\frac{z}{L_D}\right)^4,\end{aligned} \tag{3.27}$$

where

$$\mathrm{sgn}(\beta_2) = \begin{cases} +1, & \beta_2 > 0 \\ -1, & \beta_2 < 0 \end{cases}, \tag{3.28}$$

$$L_D = \frac{a_0^2}{|\beta_2|}, \tag{3.29}$$

$$L_{NL} = \frac{1}{\gamma P_0}, \tag{3.30}$$

$$N^2 = \frac{L_D}{L_{NL}}, \tag{3.31}$$

$$C = -2a_0^2 b_0, \tag{3.32}$$

where L_D is the dispersion length, L_{NL} is the nonlinear length, N^2 is a nonlinearity parameter, C is the initial normalized chirp, and $b_0 = b(0)$. Generally, N^2 indicates the dominant effect in an optical fiber. That is, if $N^2 << 1$, the dispersion is the dominant effect. On the other hand, the nonlinearity is dominant over the dispersion when $N^2 >> 1$. Note that both N and C are dimensionless parameters.

It is clearly seen from (3.27) that the expression for $a^2(z)/a_0^2$ as a function of system parameters is complicated. The optimum initial chirp C_{opt} that gives $a(z) = a_0$ has to be evaluated numerically. Thus, (3.27) does not seem to be useful for evaluating C_{opt}. If the third and the fourth order terms in (3.27) are neglected in order to simplify the expression of $a^2(z)/a_0^2$, (3.27) becomes

$$\frac{a^2(z)}{a_0^2} = 1 + \left[2\,\mathrm{sgn}(\beta_2)C\right]\left(\frac{z}{L_D}\right) + \left[1 + \frac{\mathrm{sgn}(\beta_2)N^2}{\sqrt{2}} + C^2\right]\left(\frac{z}{L_D}\right)^2. \tag{3.33}$$

Furthermore, it may be argued [104] that asymptotically $a^2(z)$ must increase quadratically with z so that (3.33) has the correct asymptotic dependence. This can be understood from the fact that for $\beta_2 > 0$, $\frac{d^2a(z)}{dz^2}$ given by (3.24) is larger than zero, and

so $\frac{da(z)}{dz}$ is increasing. When $b(0)=0$ (no chirp), $\frac{da(0)}{dz}$ in (3.21) is equal to zero. Thus, $\frac{da(z)}{dz}>0$ for all z when $\beta_2>0$ and $b(0)=0$. As $z\to\infty$, $a(z)\to\infty$. By (3.24), it follows that $\frac{d^2a(z)}{dz^2}\to 0$ as $z\to\infty$ so that $a(z)$ is asymptotically linear. Therefore, $a^2(z)$ is asymptotically quadratic with z for large z. If z is relatively small, it may be possible to more accurately approximate $a^2(z)$ with a power series expanded up to a term higher than the quadratic term.

As we can see from (3.33), the attenuation constant α does not appear in the expression for the normalized pulse width squared $a^2(z)/a_0^2$. However, if the peak input power P_0 in (3.30) is replaced by the average power P_{avg} defined as

$$P_{avg}=\frac{1}{z}\int_0^z P_0 e^{-\alpha\xi}d\xi$$
$$=\frac{P_0(1-e^{-\alpha z})}{\alpha z}, \tag{3.34}$$

then the effect of loss can be included in (3.33). This differs from previous approaches [104] where z in (3.34) is set equal to the distance between amplifiers, and P_{avg} is then independent of distance z. By replacing P_0 in (3.30) by P_{avg} given by (3.34), (3.33) becomes

$$\frac{a^2(z)}{a_0^2}=1+\left[2\,\mathrm{sgn}(\beta_2)C\right]\left(\frac{z}{L_D}\right)+\left[1+\frac{\mathrm{sgn}(\beta_2)}{\sqrt{2}}\cdot\frac{N^2(1-e^{-\alpha z})}{\alpha z}+C^2\right]\left(\frac{z}{L_D}\right)^2. \tag{3.35}$$

This approach is equivalent to scaling the nonlinear length L_{NL} by the average power. Thus, (3.35) can be rewritten as

$$\frac{a^2(z)}{a_0^2}=1+\left[2\,\mathrm{sgn}(\beta_2)C\right]\left(\frac{z}{L_D}\right)+\left[1+\frac{\mathrm{sgn}(\beta_2)\tilde{N}^2(z)}{\sqrt{2}}+C^2\right]\left(\frac{z}{L_D}\right)^2, \tag{3.36}$$

where

$$\tilde{L}_{NL}(z)=\frac{1}{\gamma P_{avg}}, \tag{3.37}$$

$$\tilde{N}^2(z) = \frac{L_D}{\tilde{L}_{NL}}, \tag{3.38}$$

where $\tilde{L}_{NL}(z)$, and $\tilde{N}^2(z)$ are the effective nonlinear length, and the effective nonlinearity parameter, respectively. Note that replacing N^2 in (3.35) by $\tilde{N}^2$ to get (3.36) is strictly based on physical principles. (3.35) is derived from the variational method, which is itself not completely rigorous, and (3.36) is actually a postulate that we will test by comparing the results of this analysis with numerical solutions of the NLSE . For comparison, $a^2(z)/a_0^2$ as a function of z are evaluated by using (3.27), (3.33), and (3.36), and compared with $a^2(z)/a_0^2$ obtained from SSF simulation for given system parameters. They are plotted in Fig. 3.5. It is clearly seen that in fact the approximation of $a^2(z)/a_0^2$ given by (3.36) is more accurate than (3.27) or (3.33). Note that (except for very small value of z) (3.33) is more accurate than (3.27). Thus, it is not beneficial to include higher order terms in the power series of $a^2(z)/a_0^2$.

As seen from Fig. 3.5, with the proper amount of initial C, a pulse is compressed at its initial stage of propagation, and then broadened again. Thus, the transmission distance z, at which the output pulse width is equal to the input pulse width, can be maximized with the optimum choice of C.

3.3.4 Optimum Chirp

In a practical fiber-optic communication system, it is desirable that the pulse be broadened as little as possible. The most desired scenario is that the output pulse width is equal to the input pulse width. Since (3.36) is quadratic in C, it can be solved to find C_{opt} that gives the output width $a(z)$ equal to the input width a_0 in closed-form. Setting $a^2(z)/a_0 = 1$ in (3.36) and solving for C yields an expression for the optimum initial chirp,

$$C_{opt} = -\mathrm{sgn}(\beta_2)\left(\frac{z}{L_D}\right)^{-1} \pm \sqrt{\left(\frac{z}{L_D}\right)^{-2} - \left[1 + \frac{\mathrm{sgn}(\beta_2)\tilde{N}^2(z)}{\sqrt{2}}\right]}, \tag{3.39}$$

where the plus and minus sign corresponds to the normal and anomalous dispersion regimes, respectively. This is because the sign of C has to be opposite to $\mathrm{sgn}(\beta_2)$ in order to obtain the pulse-compression effect [57].

Since the chirp is a real quantity, the expression in the square root in (3.39) has to be greater than or equal to zero. Therefore, there exists a maximum distance $z_{\max}$ at which the initial chirp can be utilized to make the output pulse width $a(z_{\max})$ equal to the input pulse width a_0. By using (3.39), it is found that $z_{\max}$ has to satisfy

$$z_{\max}^2 = \frac{L_D^2}{\left[1+\dfrac{\mathrm{sgn}(\beta_2)\tilde{N}^2(z_{\max})}{\sqrt{2}}\right]}. \tag{3.40}$$

In the normal dispersion case ($\mathrm{sgn}(\beta_2) > 0$), the denominator is always larger than unity. On the other hand, the denominator is less than unity in the anomalous case. Thus, $z_{\max}$ is less/greater than L_D in the normal/anomalous dispersion cases.

Of most interest is the C_{opt} value that maximizes the transmission distance z at which the output pulse width can be made equal to the input pulse width. By using the assumption that the bit period $T_b = 4\sigma_0$, 40Gb/s corresponds to $a_0 = 9$ ps. We take $|D| = 0.7$ ps/(km·nm), which corresponds to $|\beta_2| = 0.895$ ps^2/km at $\lambda_0 = 1.552$ μm, giving a dispersion of length $L_D = 87$ km. Furthermore, the single-mode optical fiber used in the numerical evaluation has $n_2 = 3\cdot 10^{-20}$ m^2/W, $A_{eff} = 70$ μm^2, and $\alpha = 0.2$ dB/km. Substituting z in (3.39) by the $z_{\max}$ found from (3.40) yields C_{opt} that maximizes the transmission distance. C_{opt} that gives $a(z_{\max}) = a_0$ is plotted in Fig. 3.6. As seen from the plot, in the normal dispersion regime ($D < 0$, curves — and --- in Fig. 3.6), C_{opt} is negative, and its magnitude increases with increasing N^2. This is because in this regime the nonlinearity enhances the effect of dispersion; thus, when N^2 increases, a larger amount of chirp is required. As stated earlier, loss reduces the effect of nonlinearity; thus, fiber loss causes C_{opt} (curve ---) to be less dependent on N^2. That is, the fiber loss decreases the variation of C_{opt}, which is beneficial.

On the other hand, C_{opt} is positive, and decreases with increasing N^2 in the anomalous regime ($D > 0$, curves -*- and -o- in Fig. 3.6) as seen in Fig. 3.6. This can be understood from the fact that the nonlinearity tends to cancel the dispersion in this regime. Therefore, a smaller chirp C_{opt} is needed as N^2 increases. Since loss reduces the effect of nonlinearity, the required C_{opt} in the lossy case (curve -o-) is larger than C_{opt} in the lossless case (curve -*-) for a given N^2. Again, loss reduces the variation of C_{opt} (curve -o-) with N^2. In the lossless case when $N^2 = \sqrt{2}$, (3.40) yields $z_{\max} = \infty$, which results in $C_{opt} = 0$. Under this condition, the soliton formation is achieved; i.e., a pulse travels in a lossless optical fiber without being distorted when $N^2 = \sqrt{2}$. This is why the curve of C_{opt} in the lossless case (curve -*-) stops at $N^2 = \sqrt{2}$. Due to the step size in the numerical calculation of C_{opt}, C_{opt} in the lossless case does not go to zero at $N^2 = \sqrt{2}$ in Fig. 3.6.

3.3.5 Maximum Transmission Distance

By using (3.40), $z_{\max}$ as a function of N^2 can be easily evaluated. The results are plotted in Fig. 3.7. The system parameters used in this figure are the same as those used in Fig. 3.6. It is clearly seen that without nonlinearity ($N^2 = 0$), $z_{\max}$ is equal to the dispersion length L_D. In the normal dispersion regime ($D < 0$, curves — and --- in Fig. 3.3), nonlinearity ($N^2 > 0$) decreases $z_{\max}$ because the nonlinearity multiplies the dispersion effect. The fiber loss decreases the nonlinear effect; thus, $z_{\max}$ in the lossy case (curve ---) is larger than $z_{\max}$ in the lossless case (curve —) for a given N^2.

It is clearly seen from Fig. 3.7 that fiber loss destroys the soliton formation, which occurs in the anomalous dispersion regime in the lossless case ($D > 0$, curves -*- and -o- in Fig. 3.7). Since the nonlinearity decreases the effect of dispersion in this regime, $z_{\max}$ increases with N^2 (curves -*- and -o-). Loss also decreases $z_{\max}$ for a given N^2 because the nonlinearity is beneficial in this regime, and loss reduces the effective nonlinearity. Over the range of N^2, the maximum transmission distance is approximately 100 km,

which corresponds to the loss of 20 dB. Thus, in practice, the transmission distance z is usually less than the maximum transmission distance $z_{\max}$ owing to noise considerations. Nevertheless, for a given z if the peak input power P_0 is replaced by P_{avg}, we get good results.

Because the analytical expressions in this analysis are obtained from approximations, it is important to discuss the accuracy of the results. The accuracy is measured in terms of the error of $a(z_{\max})/a_0$ as a function of N^2. That is, the error is determined by

$$\text{Error}(\%) = \frac{1 - a_{SSF}(z_{\max})}{a_{SSF}(z_{\max})} \cdot 100, \tag{3.41}$$

where $a_{SSF}(z_{\max})$ is the pulse width at $z_{\max}$ evaluated from the SSF simulation with C_{opt}, and $z_{\max}$ corresponds to those plotted in Fig. 3.6 and Fig. 3.7, respectively.

The error given by (3.41) as a function of N^2 is plotted as shown in Fig. 3.8. It is clearly seen that error is less than 5% up to $N^2 = 3$ for the normal dispersion case (curve ---). On the other hand, the error is kept below 15% up to $N^2 = 2$ in the anomalous dispersion case. In both the normal and anomalous dispersion cases, (3.36) reduces the error caused by the neglect of loss by more than a half for a given N^2. This implies that the effect of loss cannot be disregarded. The result shown in Fig. 3.8 also suggests that (3.36) provides more accurate results in the normal dispersion regime than in the anomalous dispersion regime. In addition, the results indicate that the effect of loss is not negligible. The prechirp technique can be utilized to make the output pulse width return to its initial width. However, there exists a maximum transmission distance for which this can be achieved.

3.4 SINUSOIDAL PHASE MODULATION

The results in Section 3.3 indicates that prechirping a pulse used for carrying bit 1 is a potential technique to avoid pulse overlap by means of the pulse compression effect. However, quadratically prechirping a pulse is not practical in terms of system implementation. In general, phase modulating the RZ signal is accomplished by driving a

phase modulator with a periodic electrical signal. In order to practically prechirp the RZ signal, the quadratic phase variation has to be approximated. Since most of the pulse energy is concentrated around the center of the pulse, the characteristic of the chirp around the pulse center is much more important that that around the tails. When a quadratic chirp and a sinusoidal chirp are compared as sketched in Fig. 3.9 (a), the trends of both chirps across an RZ pulse are similar. Hence, the sinusoidal chirp should be a good approximation of the quadratic chirp. In practice, the RZ signal is synchronously phase modulated so that its phase variation is sinusoidal in the time domain, and the amplitude of the sinusoidal phase variation is properly adjusted to yield the optimum system performance [105]-[107]. Note that the frequency of the sinusoidal phase variation has to be equal to the bit rate R_b in order that all pulses contained in the RZ signal are identically chirped. In this dissertation, we use the notation sinusoidal same phase modulation (SaPM) to represent this signal format. Note that the notation SaPM is used instead of SPM because SPM usually stands for self-phase modulation.

The transmitter configuration of the SaPM signal format is shown in Fig. 3.9 (b). The transmitter of the SaPM signal is similar to that of the conventional RZ signal except that there is an additional device, a phase modulator, to perform phase modulation. It should be noted that the phase modulation is done in the optical domain, which can be accomplished by a lithium niobate ($LiNbO_3$) phase modulator [108]. In this device, the phase modulation is obtained by using the electro-optic effect by which the refractive index of the device is proportional to the applied electric field. As a signal propagates in this device, the change in the refractive index with time causes a corresponding change in the phase of the signal. The phase modulator shown in Fig. 3.9 (b) is driven by a synchronized sinusoidal signal having frequency f_p equal to the bit rate R_b. The amount of prechirp is proportional to the amplitude of the driving sinusoidal signal, and the sign of the prechirp can be chosen by synchronizing either positive peak or negative peak of the driving signal to a pulse peak. The peak (not peak-to-peak) amplitude of the sinusoidal phase variation of the signal is called the modulation index or modulation depth B.

The pulse compression effect generally delays the pulse-to-pulse interactions and the peak reduction, which makes the SaPM signal format perform better than the RZ

signal format. However, at large transmission distance where two pulses overlap each other, a spurious pulse is generated due to the fact that adjacent pulses have identical phase variation, similar to the example shown in Fig. 3.2 (a). Since the frequency of the sinusoidal signal driving the phase modulator is equal to the bit rate R_b, the effect of the phase modulation on the PSD of the signal is to expand the signal bandwidth. The PSD of the SaPM signal is shown in Fig. 3.9 (c) when the modulation index B is equal to 1.5 radians. When the PSD of the SaPM signal is compared with that of the RZ signal shown in Fig. 3.1 (b), it is apparent that the SaPM signal has a broader spectral width than the RZ signal. Larger occupied bandwidth is in fact traded for the improvement obtained from the pulse compression effect for the SaPM signal format. The larger the modulation index B, the broader is the spectral width of the signal. Hence, in WDM systems employing SaPM signal format, the linear cross talk among channels increases with the modulation index B. In practice, the modulation index B has to be optimally adjusted to improve system performance in all channels while minimizing the linear cross talk among channels [109]. Doing this, the dispersion-managed transmission of 64 WDM channels, each at 12.3 Gb/s, over the distance of 9000 km was successfully demonstrated by researchers at TyCom laboratories [109].

It should be noted that when the frequency of the sinusoidal signal driving the phase modulator is changed from the bit rate to half the bit rate, adjacent pulses in the RZ signal are chirped oppositely [110]. This modulation format is called sinusoidal alternating phase modulation (APM) in this dissertation. For the APM signal, some pulses are compressed during propagation while others spread more severely than an unchirped pulse. That is, one pulse is broadened whereas its neighboring pulses are compressed. This suppresses the growth of a spurious pulse between adjacent pulses due to both pulses having opposite phase variation; hence, the RZ eye profile is preserved at the receiver. However, the flaw of the APM signal format is that not all pulses are chirped identically; hence, an isolated pulse having the wrong chirp is broadened more severely. This causes severe ISI in an empty bit slot 0 when its neighboring pulses have the wrong chirp, and severe reduction of signal level for bit 1.

Since the frequency of the sinusoidal phase variation of the APM signal is equal to half the bit rate, the sidebands of the APM signal are located at frequencies equal to

multiples of half the bit rate away from the carrier frequency, as shown in Fig. 3.9 (d), which is the PSD of the APM signal when the modulation index is set to 1.5. Note that the carrier and the sidebands at the frequencies of multiples of the bit rate away from the carrier frequency are barely seen in Fig. 3.9 (d) because their magnitudes depend on the modulation index, which is intentionally set to 1.5 for comparison with the SaPM. Comparing the PSD of the APM signal with that of the SaPM signal, the PSD of the APM signal is narrower. This is due to the fact that the frequency of the phase variation in the case of APM is half that of the SaPM. Still, the PSD of the APM signal is broader than that of the RZ signal due to the phase modulation.

3.5 SQUARE-WAVE PHASE MODULATION (SWM)

The difference between the square-wave phase modulation (SWM) signal format and the RZ format is that there is a 180° phase shift at the border between two adjacent bit slots in the case of the SWM [111]-[114]. In other words, adjacent pulses have opposite signs in the case of an SWM signal. The transmitter configuration of the SWM signal is shown in Fig. 3.10 (a). The NRZ signal is fed to the SWM modulator, which performs the NRZ-to-RZ conversion, and at the same time imposes periodic square-wave phase variation onto the signal. Note that the SWM modulator is a combined amplitude and phase modulator. In practice, it is implemented as a push-pull type Mach-Zehnder (MZ) modulator biased at its transmission null so that the output electric field is minimum when there is no driving signal. The signal that drives the SWM modulator is a sinusoidal clock signal having a frequency of half the bit rate. Mathematically, this is equivalent to synchronously multiplying the NRZ signal by a periodic RZ pulse train having alternate sign between adjacent pulses, as shown in Table 3.1 where D_m, S_m, and $SWM_m = D_m \times S_m$ are the NRZ data sequence, an alternate-sign RZ pulse train, and the sign of the resultant SWM signal, respectively. The alternate signs between adjacent pulses prevents the growth of spurious pulses [see Fig. 3.2 (b)], and minimizes the effect of IFWM as discussed early in this chapter, which leads to performance improvement compared to the RZ signal. Using this format,1.6-Tb/s transmission consisting of 40 SWM signals, each at the bit rate of 42.7 Gb/s, over 2000 km of dispersion-managed fibers was demonstrated in 2001 [115].

When an alternate-sign RZ pulse train is considered, due to its periodicity it can be represented by the complex exponential Fourier series as

$$S(\tau) = \sum_{k=-\infty}^{\infty} c_k e^{j\pi k r_b \tau}, \tag{3.42}$$

where c_k are the complex Fourier coefficients. The Fourier transform of $S(\tau)$ is given by

$$F\{S(\tau)\} = \sum_{k=-\infty}^{\infty} c_k \delta(f - kR_b/2), \tag{3.43}$$

where F represents the Fourier transform operation, and $\delta(f)$ is the Dirac delta function. From (3.43), $S(\tau)$ has spectral lines at multiples of half the bit rate ($f = kR_b/2$), weighted by c_k. Since multiplication in the time domain is equivalent to convolution in the frequency domain, the Fourier transform of the SWM signal is, therefore, the convolution between the Fourier transform of the NRZ data sequence and that of the alternate-sign RZ pulse train, which is given by (3.43). By using the shifting property of the Dirac delta function, the Fourier transform of the SWM signal consists of replicas of the Fourier transform of the NRZ signal weighted by c_k and shifted in frequency by $kR_b/2$. Due to the fact that the RZ pulse train $S(\tau)$ has alternate signs between adjacent pulses, $c_k = 0$ for k even ($k = 0, \pm 2, \ldots$). Thus, the Fourier transform of the SWM signal contains the weighted replicas of the Fourier transform of the NRZ signal at the frequencies of odd multiples of half the bit rate ($f = \pm R_b/2, \pm 3R_b/2, \ldots$) away from the carrier frequency. Shown in Fig. 3.10 (b) is the PSD of the SWM signal. The sidebands are located at the frequencies of odd multiples of half the bit rate away from the carrier frequency. Those sidebands are in fact the shifted versions of the carrier of the NRZ signal. Note that the carrier does not exist in the PSD of the SWM signal because adjacent pulses have opposite signs.

One may notice that if the SWM data sequence SWM_m shown in Table 3.1 is multiplied again by the alternate-sign RZ pulse train S_m, then the sign of the resultant pulse is reverted back to the original data sequence D_m. Similar to the case of the SWM signal discussed earlier, this implies that the Fourier transform of an RZ signal itself contains the weighted replicas of the Fourier transform of a SWM signal at the

frequencies of odd multiples of half the bit rate ($f = \pm R_b/2, \pm 3R_b/2, \ldots$) away from the carrier frequency. Therefore, by single-sideband filtering an RZ signal, the SWM signal can be partially obtained. The signal format generated from this technique is generally called vestigial sideband-RZ (VSB-RZ) [116]-[118]. The transmitter configuration of the VSB-RZ signal format is similar to that of the RZ signal shown in Fig. 3.1 (a) except that there is a sharp roll-off optical bandpass filter placed after the RZ modulator. The center frequency of the bandpass filter is detuned from the carrier frequency of the RZ signal so that only a single sideband is allowed to pass. It should be noted that some spectral components of the unwanted adjacent sidebands are within the passband of the optical filter, hence interfering with the desired sidebands. Moreover, approximately half of the original signal power is lost due to the filtering. Although the bandwidth of the VSB-RZ signal is generally narrower than the SWM signal due to strong filtering, narrower bandwidth does not offer any advantages compared with the SWM signal in terms of dispersion tolerance and interchannel linear cross talk [119], [120]. This is due to the fact that the VSB-RZ signal format does not fully possess the benefits of the SWM signal. Moreover, alternate signs between adjacent pulses makes the SWM signal robust to the effect of spectral filtering. Thus, the bandwidth of the SWM signal can be further reduced to minimize interchannel linear cross talk by symmetrically filtering the SWM signal while the capability of spurious pulse suppression is still preserved.

3.6 OPTICAL DUOBINARY AND ALTERNATE MARK INVERSION (AMI)

In terms of classical digital signal formats, the duobinary signal format is classified as a partial response or correlative signal format in which ISI is intentionally and deterministically introduced into the transmitted signal in order to achieve the Nyquist rate for a given transmission bandwidth [121], [122]. At the receiver the inserted ISI can be removed by way of a decoding process. However, as discussed below, the optical duobinary signal format does not work exactly in the same fashion as the conventional electrical duobinary signal format.

3.6.1 Optical Duobinary

The schematic diagram of the transmitter for the optical duobinary signal format is shown in Fig. 3.11 (a). It is seen that the transmitter configuration of the optical duobinary signal format is identical to that of the RZ signal format. However, the signal that drives the data modulator is not the original NRZ data, but the duobinary-encoded data sequence [123]-[125]. Each step in the encoding process is illustrated as shown in Table 3.2. The first step in the process is to feed the original NRZ data sequence D_m to the differential precoder in order to avoid error propagation at the receiver caused by the preceding received data not being recovered correctly [122], [126]. The precoded data sequence P_m is then passed through the delay-and-add circuit (encoder). The data sequence output from the encoder is denoted by B_m in Table 3.2. It is clearly seen that an original data bit 0 results in the B_m being either 0 or 2 whereas an original data bit 1 corresponds to $B_m = 1$. If the encoded data sequence B_m is used to directly drive the data modulator, the suppression of a spurious pulse caused by the overlap between two pulses cannot be achieved since the sign of B_m is always positive, hence the sign of the transmitted pulses will always be the same. In practice, the encoded data B_m is fed to an electrical amplifier before driving the data modulator. The role of the electrical amplifier is to amplify the signal to the desired level, and to remove the DC component of the signal (the amplifier is AC coupled). Thus, the encoded sequence that drives the data modulator is $B_m - 1$. As seen from Table 3.2, an original data bit 0 ($D_m = 0$) corresponds to $B_m - 1 = \pm 1$ while an original data bit 1 ($D_m = 1$) causes $B_m - 1$ to take the value of 0 regardless of their positions. The NRZ optical duobinary signal is then converted to RZ optical duobinary by the RZ modulator. At the receiver, the square-law detector neglects the phase of a received pulse; therefore, the received data sequence is the absolute value of the transmitted data sequence $B_m - 1$. The received data sequence $|B_m - 1|$ has to be inverted to recover the original data sequence D_m. It should be noted that if the original data sequence D_m is inverted before being fed to the precoder at the transmitter, the received data sequence $|B_m - 1|$ is D_m.

Although the precoding and encoding processes in the optical duobinary signal are identical to those in the electrical duobinary signal, the decoder is not required in the case of optical duobinary signal. This is because the receiver is the square-law detector. In addition, the purpose of having a data bit transmitted in two levels is to combat the impairments caused by the optical fiber. The advantage of transmitting an original data bit 0 ($D_m = 0$) as either $+1$ or -1 can be explained as follows. As seen from Table 3.2, the unique property of the RZ optical duobinary signal format is that the transmitted blocks of pulses surrounding an empty bit slot always have opposite signs. To be more precise, for the optical duobinary signal format, the signs of two blocks of pulses are opposite to each other when they are separated by an odd number of empty bit slots ($B_m - 1 = 0$), and all pulses in each block have the same sign [127]. Theoretically, this is an advantage of the RZ optical duobinary signal over the RZ signal because in the case of a transmitted empty bit slot ($B_m - 1 = 0$) surrounded by pulses, blocks of RZ pulses on each side have opposite signs. This in effect prevents the growth of a spurious (ghost) pulse in a transmitted empty bit slot caused by the overlap between surrounding pulses as explained in Section 3.2.

When the PSD of the RZ optical duobinary signal format is considered, it is completely distinct from the other modulation formats discussed so far. Mathematically, the baseband complex envelope of the transmitted optical duobinary signal can be written as [127]

$$E_D(\tau) = \sum_m o_m g(\tau - mT_b), \tag{3.44}$$

where o_m is either $+1$ or -1 with equal probability, and $g(\tau)$ is given by

$$g(\tau) = \frac{1}{2}\left[h(\tau) + h(\tau - T_b)\right], \tag{3.45}$$

where $h(\tau)$ is the transmitted pulse shape. By using (3.44) and (3.45), it can be shown that the PSD of the optical duobinary signal $E_D(\tau)$ is given by

$$P_D(f) = \left|H(f)\right|^2 \cos^2(\pi f T_b), \tag{3.46}$$

where $H(f)$ is the Fourier transform of $h(\tau)$. Since $h(\tau)$ is generally a time-limited pulse, $H(f)$ does not contain a line spectrum. Therefore, the PSD of the duobinary

signal given by (3.46) does not have a line spectrum, which differs from all modulation formats previously discussed, as shown in Fig. 3.11 (b). When $h(\tau)$ is an RZ pulse having the pulse width of less than the bit period T_b, the spectral width of $|H(f)|^2$ is larger than $1/T_b$. Due to the fact that the cosine function in (3.46) is periodic in frequency, the envelope of $P_D(f)$ is mainly governed by $|H(f)|^2$ [127]. This suggests that the bandwidth of the RZ optical duobinary signal is not smaller than that of the RZ signal provided that both signals employ the same elementary pulse shape. It was shown in [127] and [128] that the optical duobinary signal does not provide any significant advantage over the RZ signal unless bandlimited spectral filtering is performed at the transmitter. This is a result of the fact that both signal formats contain the same level of high-frequency spectral components, and those high-frequency components are severely affected by dispersion.

The sign inversion between two blocks of pulses surrounding an empty bit slot on each side makes the optical duobinary signal more robust to ISI caused by the bandlimiting filter than the RZ signal. In practice, the bandlimited spectral filtering can be done either electrically or optically. In the optical domain, the signal bandwidth can be reduced by placing a narrowband optical bandpass filter at the fiber input [128], [129]. In the electrical domain, the signal bandwidth can be limited as follows. When the delay-and-add circuit shown in Fig. 3.11 (a) is considered, its transfer function is given by

$$H_{duo}(f) = 2e^{-j\pi fT_b}\cos(\pi fT_b), \tag{3.47}$$

which is periodic in frequency. The high frequency components can be eliminated electrically by replacing the delay-and-add circuit by an electrical lowpass filter whose transfer function duplicates the center lobe of $H_{duo}(f)$. That bandlimiting filter is called the duobinary filter, and the 5th order Bessel filter having a bandwidth of approximately one-fourth the bit rate is generally used [130]-[132]. It should be noted that bandlimited spectral filtering reduces the signal bandwidth while the capability of spurious-pulse suppression in an empty bit slot is still preserved.

3.6.2 Alternate Mark Inversion

The only difference between the optical duobinary and the alternate mark inversion (AMI) signal formats is the encoding process. As shown in Fig. 3.12, the-delay and-subtract circuit (AMI encoder) is used instead of the delay-and-add circuit in the case of the AMI signal format. This causes the output encoded sequence B_m of the AMI signal to differ significantly from that of the optical duobinary signal as illustrated in Table 3.3. By employing the AMI encoder, the original data bit 1 corresponds the $B_m = \pm 1$ while the original data bit 0 is transformed to $B_m = 0$. In fact, the AMI encoding process causes two consecutive transmitted pulses to have opposite signs regardless of the number of empty bit slots ($B_m = 0$) in between. Unlike the optical duobinary signal, an inverter is not needed either at the receiver or at the transmitter because the received data sequence $|B_m| = D_m$ as demonstrated in Table 3.3. This is one advantage of the AMI signal format over the optical duobinary. When the signs of the transmitted pulses B_m are considered, the AMI signal format possesses the benefits offered by both the optical duobinary and the SWM signal formats. Similar to the SWM signal format, adjacent pulses in a block of pulses have opposite signs. Hence, the growth of a spurious pulse caused by the overlap between adjacent pulses due to the dispersion is suppressed and the IFWM among pulses is minimized. Moreover, the signs of the nearest neighboring pulses surrounding a block of empty bit slots (bit slot 0's) are opposite, which is identical to the optical duobinary signal format when the number of adjacent empty bit slots is odd. This in effect minimizes the ISI in a bit slot 0 caused by the surrounding bit 1's. The potential of the AMI signal format was experimentally demonstrated by researchers at NTT Network Innovation Laboratory [133]. In their experiment, eight 43-Gb/s WDM channels spaced at 100 GHz apart were successfully transmitted over 163 km of transmission fibers consisting of the SSMF and RDF.

Similar to the optical duobinary signal, the baseband complex envelope of the transmitted AMI signal is given by (3.44). However, $g(\tau)$ in the case of the AMI signal is given by

$$g(\tau) = \frac{1}{2}\left[h(\tau) - h(\tau - T_b)\right] \quad . \tag{3.48}$$

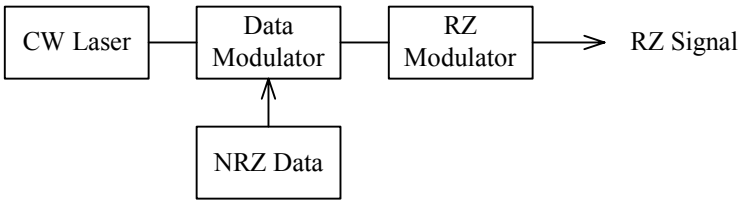
Note that the only difference between (3.45) and (3.46) is the operation between $h(\tau)$ and $h(\tau - T_b)$. By substituting (3.48) into (3.44), the PSD of the AMI signal is given by

$$P_D(f) = |H(f)|^2 \sin^2(\pi f T_b). \tag{3.49}$$

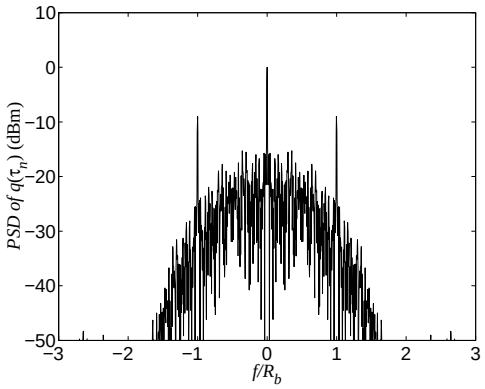
From (3.49), it is seen that the baseband-equivalent PSD of the AMI signal always has a null at DC, which corresponds to the null at the carrier frequency in the optical signal. The example of the PSD of the AMI signal relative to the carrier frequency is shown in Fig. 3.12 (b). Similar to the optical duobinary signal, the PSD of the AMI signal does not contain a line spectrum. This makes the optical duobinary and AMI signal formats unique in terms of the PSD profile.

3.7 SUMMARY

It is clearly seen that the RZ signal format provides many advantages over the NRZ signal in terms of tolerance to interchannel and intrachannel impairments. However, the system performance can be further improved by imposing some special characteristics onto the RZ signal. Each RZ optical modulation format has its own mechanism to combat the impairments caused by dispersion and nonlinearity in an optical fiber. Those mechanisms are spurious pulse suppression and pulse compression, which are obtained by manipulating the phase of the signal. However, the RZ pulse shape itself can also lead to performance improvement. This is discussed in the next chapter.

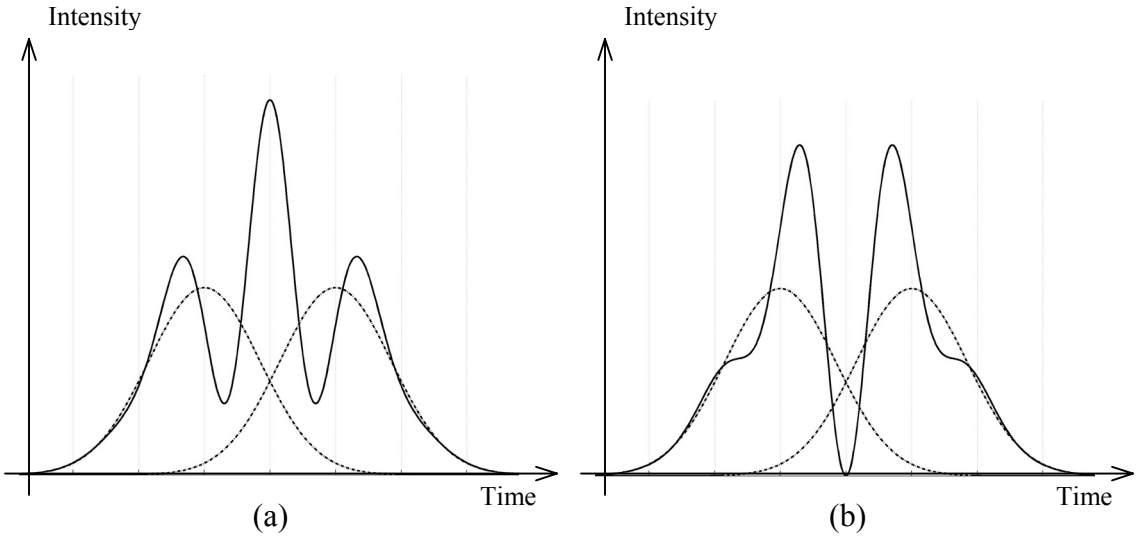


(a)



(b)

Fig. 3.1: RZ signal generation: (a) transmitter configuration. (b) Power spectral density of RZ signal.



(a) (b)

Fig. 3.2: Illustration of ISI caused by dispersion. The dashed curves correspond to the intensity of individual pulses broadened by dispersion. The solid curve is the resultant intensity of two pulses when they interfere each other. (a) Two pulses have identical sign. (b) Two pulses have opposite signs (positive and negative).

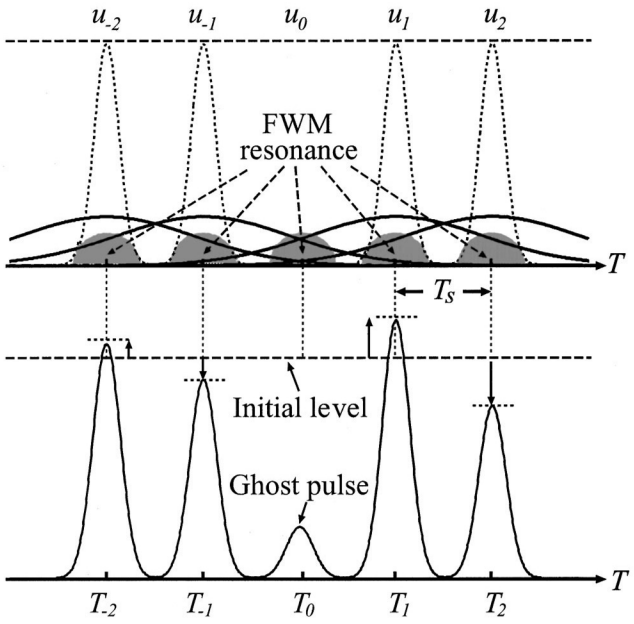


Fig. 3.3: Interactions among pulses due to IFWM [93]. Horizontal axis is time, and vertical axis is intensity. The corresponding bit pattern is 11011.

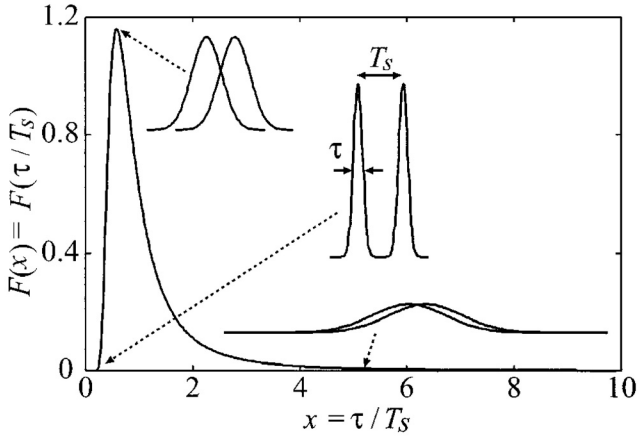


Fig. 3.4: Dimensionless function $F(x)$ representing the magnitude of IXPM between two interacting pulses where x is the ratio between the pulse width and the pulse separation [93].

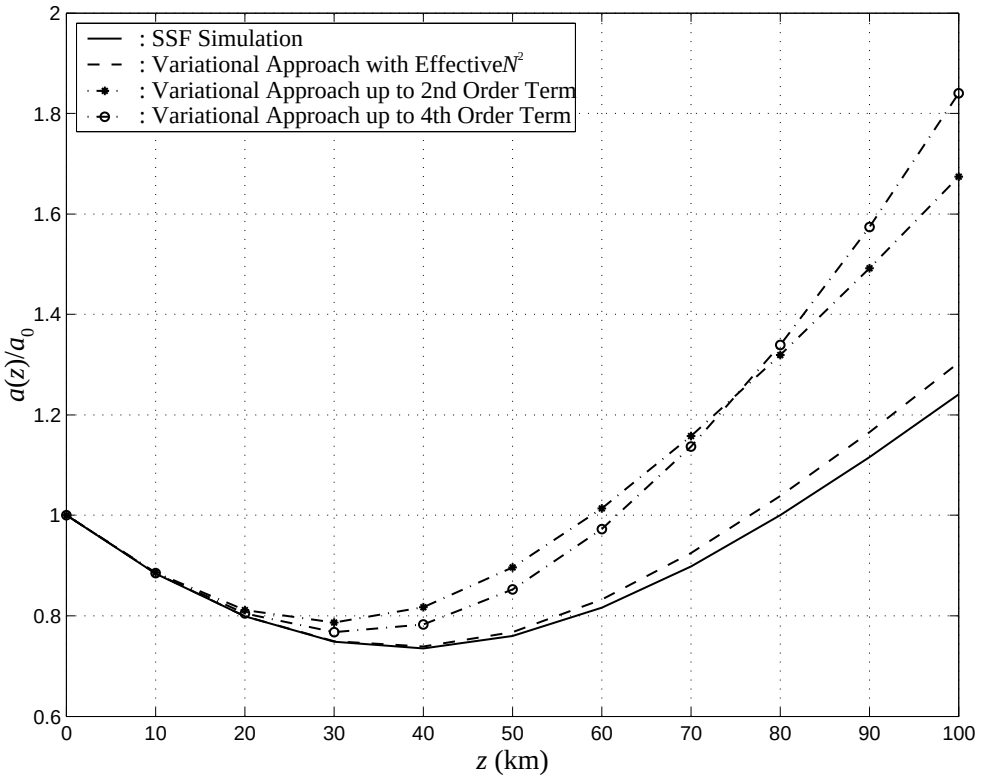


Fig. 3.5: Normalized pulse width $a(z)/a_0$ as a function of distance z when $\lambda_0 = 1.552$ μm, $n_2 = 3 \cdot 10^{-20}$ m^2/W, $A_{eff} = 70$ μm^2, $a_0 = 9$ ps (40 Gb/s), $P_0 = 10$ mW ($L_D = 87$ km, $L_{NL} = 58$ km and $N^2 = 1.52$), $D =$ -0.7 ps/(km·nm), $\alpha =$ 0.2 dB/km, and $C =$ -1.13.

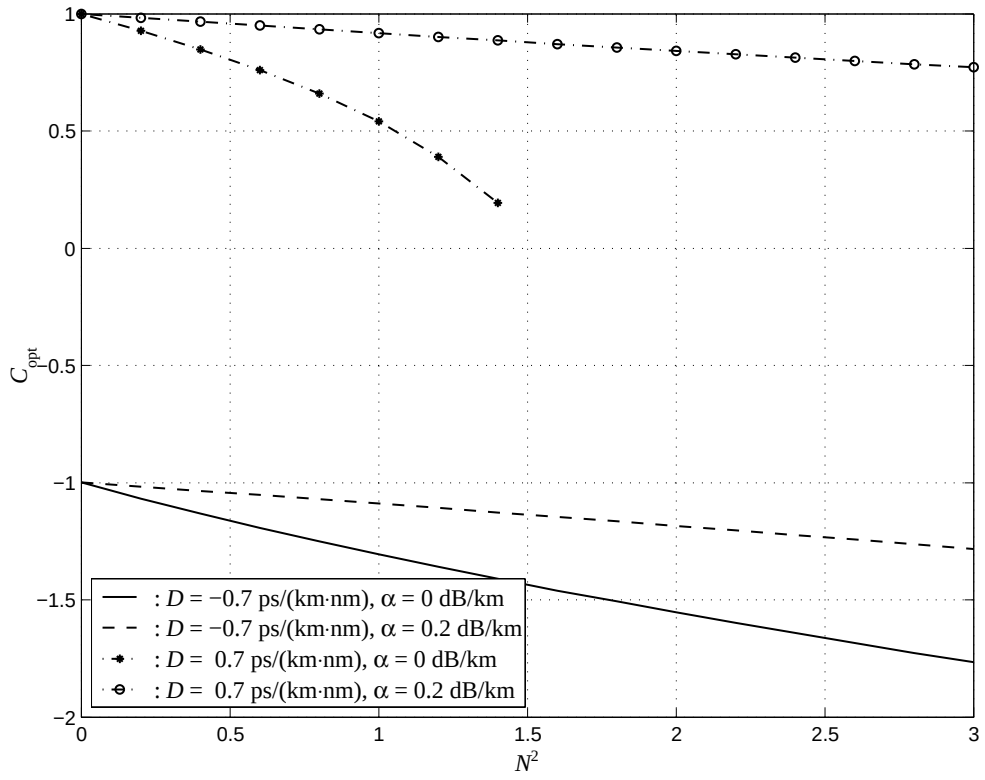


Fig. 3.6: Optimum initial chirp C_{opt} which gives $a(z_{\max})/a_0 = 1$ obtained from variational approach as a function of nonlinearity parameter N^2. Common parameters in all plots are: $\lambda_0 = 1.552$ μm, $n_2 = 3 \cdot 10^{-20}$ m^2/W, $A_{eff} = 70$ μm^2, and $a_0 = 9$ ps (40 Gb/s).

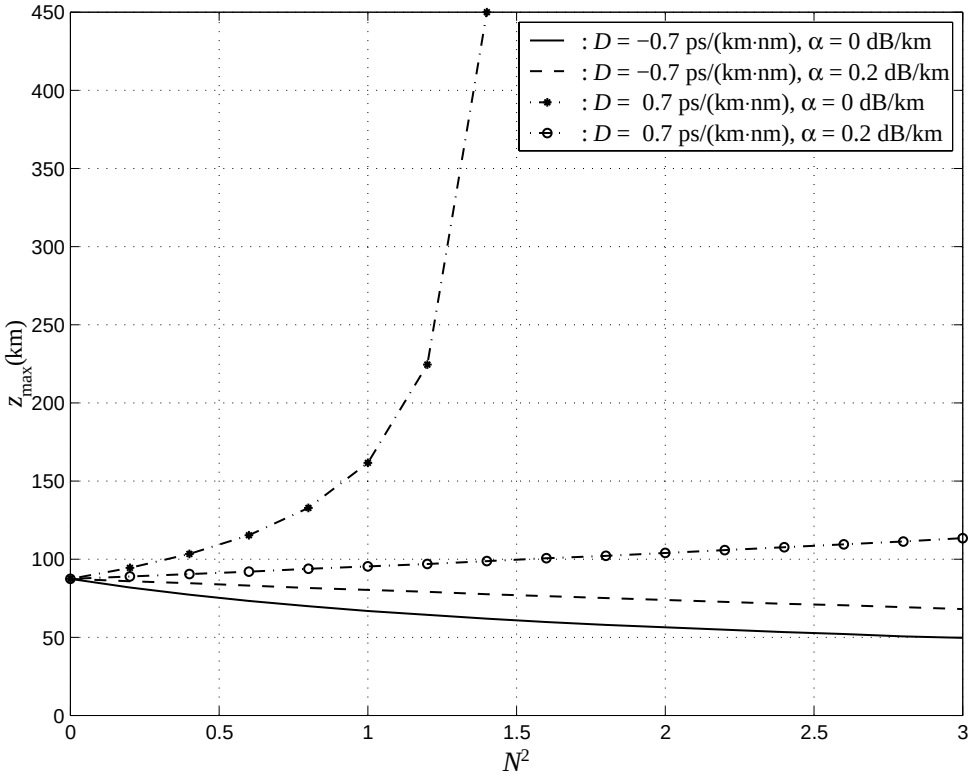


Fig. 3.7: Maximum distance $z_{\max}$ as a function of nonlinearity parameter N^2. Common parameters in all plots are: $\lambda_0 = 1.552$ μm, $n_2 = 3 \cdot 10^{-20}$ m^2/W, $A_{eff} = 70$ μm^2, and $a_0 = 9$ ps (40 Gb/s).

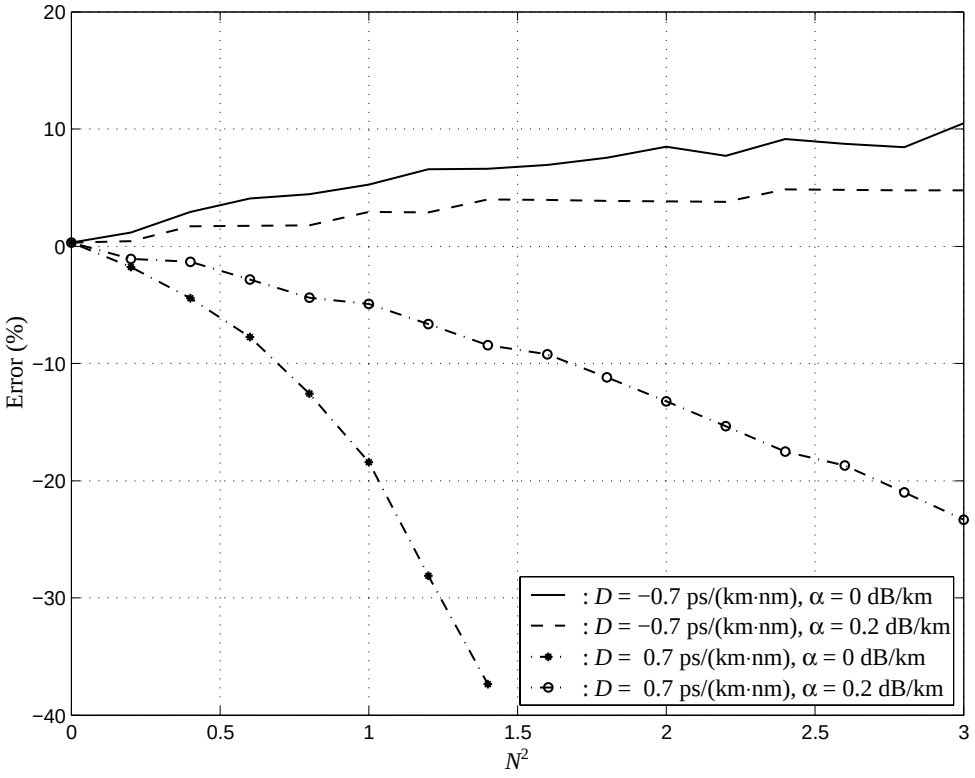


Fig. 3.8: Error in percent of the normalized pulse width at $z_{\max}$ obtained from variational approach as a function of nonlinearity parameter N^2. Common parameters in all plots are: $\lambda_0 = 1.552$ μm, $n_2 = 3 \cdot 10^{-20}$ m^2/W, $A_{eff} = 70$ μm^2, and $a_0 = 9$ ps (40 Gb/s).

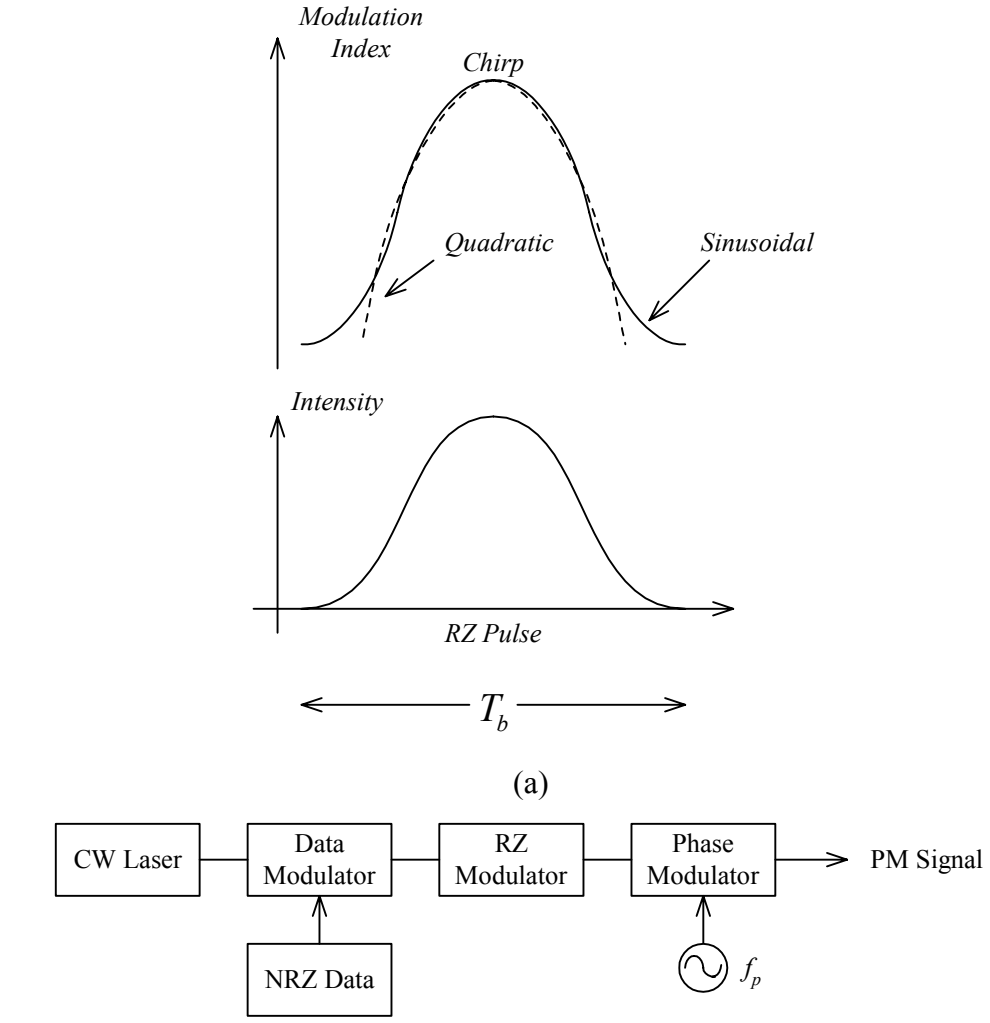
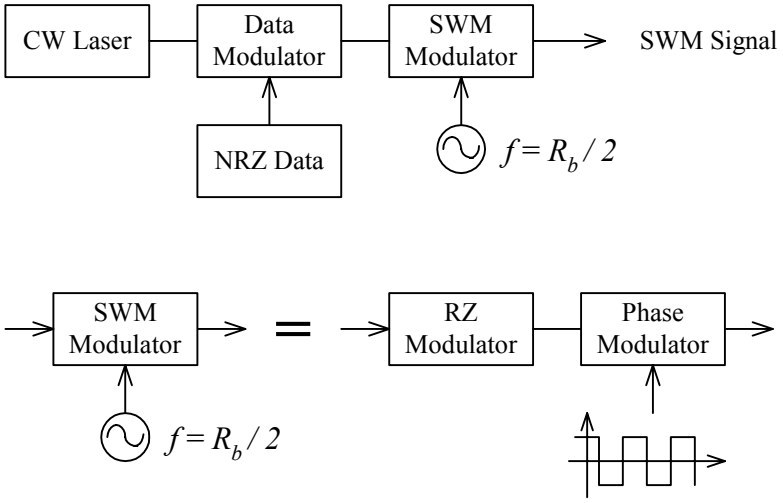
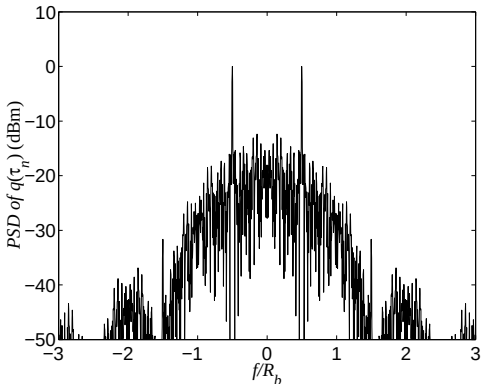


Fig. 3.9: SaPM signal generation: (a) Comparison between quadratic chirp and sinusoidal chirp. (b) Transmitter configuration. (c) Power spectral density of SaPM signal when the modulation index is 1.5. (d) Power spectral density of APM signal when the modulation index is 1.5.



(a)

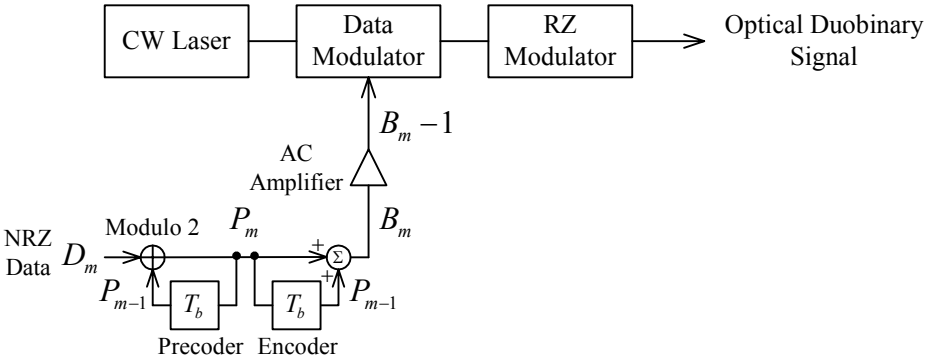


(b)

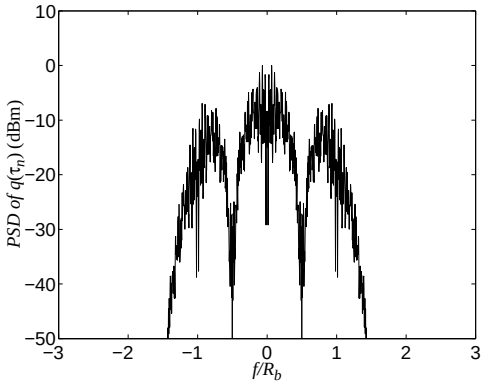
Fig. 3.10: SWM signal generation. (a) Transmitter configuration. (b) Power spectral density of SWM signal.

Table 3.1: Illustration of SWM signal generation.

D_m	1	0	0	1	1	1	1	1	0	1	1	0	1	1	0	1
S_m	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1
SWM_m	+1	0	0	−1	+1	−1	+1	−1	0	−1	+1	0	+1	−1	0	−1



(a)

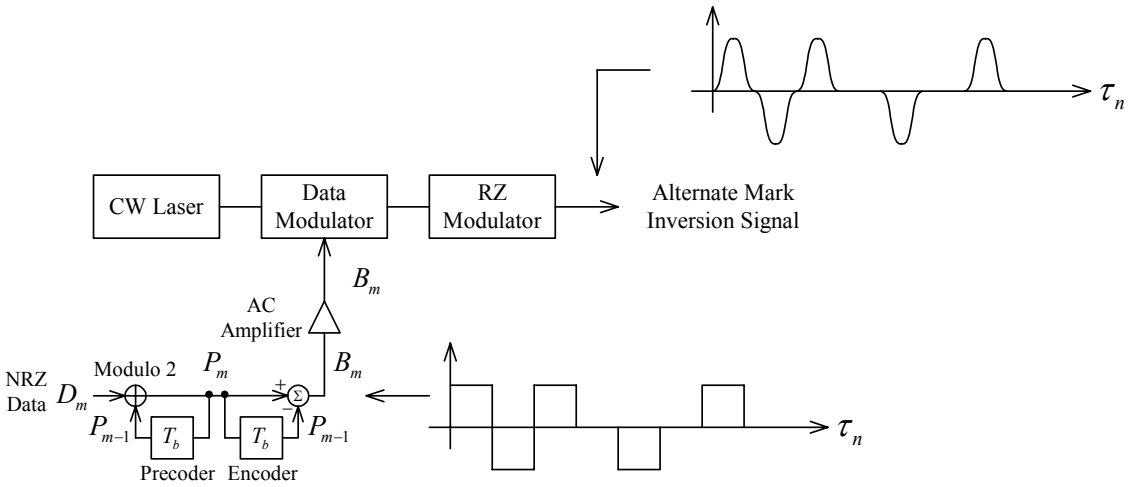


(b)

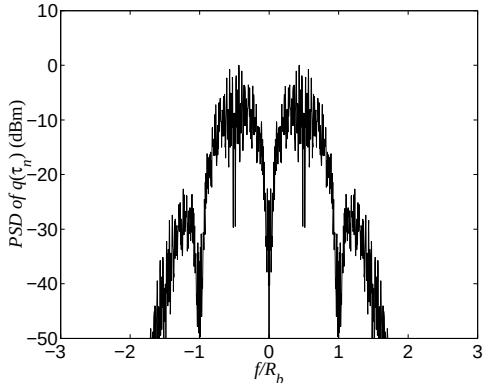
Fig. 3.11: Optical duobinary signal generation. (a) Transmitter configuration. (b) Power spectral density of optical duobinary signal.

Table 3.2: Illustration of duobinary precoding and encoding processes.

D_m		0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	0
P_m	0	0	0	1	0	1	0	1	1	0	1	1	0	1	1	0	0
B_m		0	0	1	1	1	1	1	2	1	1	2	1	1	2	1	0
$B_m - 1$		−1	−1	0	0	0	0	0	+1	0	0	+1	0	0	+1	0	−1
$\lvert B_m - 1 \rvert$		1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1
$\overline{\lvert B_m - 1 \rvert}$		0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	0



(a)



(b)

Fig. 3.12: Alternate mark inversion signal generation. (a) Transmitter configuration. (b) Power spectral density of alternate mark inversion signal.

Table 3.3: Illustration of alternate mark inversion precoding and encoding processes.

D_m		0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	0
P_m	0	0	0	1	0	1	0	1	1	0	1	1	0	1	1	0	0
B_m		0	0	1	−1	1	−1	1	0	−1	1	0	−1	1	0	−1	0
$\lvert B_m \rvert$		0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	0