In phase-modulated optical communication systems, like other means of communication, an essential function is to modulate the transmitted signal with information. Digital modulation usually carries information using the amplitude, phase, or frequency of the carrier (Haykin, 1988, Proakis, 2000). Digital modulation for coherent optical communications is fundamentally the same as that for wireless or wireline communications. Optical carrier is used in optical communications, electromagnetic wave is used in radio frequency (RF) communications, and electrical signal is used in copper-wire based wireline communications. In this chapter, we will first review the basic modulation formats for digital communications and then the methods to modulate them to the optical carrier.

Phase-modulated or coherent optical communication systems, like intensity-modulated/direct-detection (IMDD) systems, mostly use semiconductor diode laser as the light source. The basic structure, modulation response, and noise properties of semiconductor laser are briefly reviewed. With low quality signal, a semiconductor laser can be directly modulated to generate on-off keying signal.

Both amplitude- and phase-modulated signals are usually generated by amplitude and phase modulator, respectively. The principle and basic structures of those modulators are presented here. An amplitude modulator can be used to generate on-off keying and binary phase-modulated signals. Various types of return-to-zero (RZ) differential phase-shift keying (RZ-DPSK) transmitter are designed with RZ duty cycle of 1/3, 1/2, and 2/3.

The frequency modulation characteristic of semiconductor laser is also discussed later in this chapter.
1. Basic Modulation Formats

In addition to the amplitude or intensity, coherent optical communication systems utilize the phase or frequency of the optical carrier to carry information. Figure 2.1 shows basic modulation formats for coherent optical communications. The binary amplitude-shift keying (ASK) signal is the same as on-off keying. On and off states are represented by the presence or absence of light, the same as nonzero and zero amplitude, respectively. While the term of on-off keying is used for IMDD systems, ASK is the term commonly used for coherent optical communications. In binary ASK systems, the two transmitted signals to represent “0” and “1” states are

\[ s_1(t) = A \cos \omega_c t, \quad 0 < t \leq T \]  
\[ s_2(t) = 0, \quad 0 < t \leq T \]

where \( A \) is a constant amplitude, \( \omega_c \) is the angular frequency of the optical carrier, and \( T \) is the signal duration that is also equal to the bit interval for the binary signal.

Phase-shift keying (PSK) in Fig. 2.1 uses the phase of the carrier to carry information. In binary PSK systems, the two binary signals are

\[ s_1(t) = A \cos \omega_c t, \quad 0 < t \leq T \]  
\[ s_2(t) = -A \cos \omega_c t, \quad 0 < t \leq T \]

Ideally, the phase difference between the two antipodal signals is 180° for optimal performance. For the same signal-to-noise ratio (SNR) and
assumed ideal signal demodulation, binary PSK signal can achieve the lowest error probability.

Frequency-shift keying (FSK) in Fig. 2.1 uses the frequency of the carrier to encode information. In binary FSK systems, the two FSK signals are

\[ s_1(t) = A \cos \omega_1 t, \quad 0 < t \leq T \]  
\[ s_2(t) = A \cos \omega_2 t, \quad 0 < t \leq T \]

where \( \omega_1 \) and \( \omega_2 \) are two angular frequencies to represent "0" and "1" state, respectively. Ideally, the two signals of \( s_1(t) \) and \( s_2(t) \) should be orthogonal with each other.

In addition to the basic modulation formats of binary ASK, PSK, and FSK in Fig. 2.1, more complicated formats can be constructed by adding more levels into the basic signals or combining two components of, for example, amplitude and phase. More general discussion of digital modulation formats can be found in Proakis (2000) and Haykin (1988). The baseband signal can also be return-to-zero (RZ) instead of non-return-to-zero (NRZ) linecode of Fig. 2.1.

A single-mode fiber can support two orthogonal polarizations, another alternative to the formats in Fig. 2.1 is polarization-shift keying (PolSK) (Benedetto and Poggiolini, 1992, Calvani et al., 1988, Imai et al., 1990b). Polarization-division multiplexing (PDM) utilizes the two orthogonal polarizations to transmit two independent data streams (Bigo et al., 2001, Frignac et al., 2002, Hill et al., 1992, Noé et al., 2001). In general, PDM gives better spectral efficiency than PolSK. PDM is similar to polarization multiplexing in microwave systems to utilize the two polarizations of electromagnetic wave as two, ideally, independent channels.

There are many methods to generate the basic modulation formats of Fig. 2.1 in optical domain. This chapter will study some popular methods to generate the digital modulated signal.

### 2. Semiconductor Diode Lasers

In all optical communication systems, a light source must be used to originate an optical signal. Because of its small size, low power consumption, reliability, and compatible with electronic circuits, semiconductor diode lasers are the most widely used light source for communication applications. Virtually all optical communication systems use semiconductor laser as light source. Erbium-doped fiber amplifiers (EDFA) are also commonly pumped by high-power semiconductor lasers (Becker et al., 1999, Desurvire, 1994, Nakazawa et al., 1989). Other than lightwave communications, semiconductor laser also find its applications for opti-
Figure 2.2. Structure of a semiconductor laser with a Fabry-Perot cavity.

2.1 Basic Structures

Figure 2.2 shows the basic structure of a semiconductor laser with a Fabry-Perot cavity. The active region is the media to provide optical gain sandwiched between the $p$- and $n$-type semiconductor materials. The $p-n$ junction is preference to be a double-heterostructure junction in which the active region using a material having a band-gap smaller than both $p$- and $n$-type materials. The usage of heterostructure structure for semiconductor laser has two major advantages. Because of the band-gap difference, the active region can effectively confine electrons and holes into the active layer. Electron and hole recombination generates light or amplifies the passing through light through spontaneous and stimulated emission, respectively. The active region based on heterostructure also has larger refractive index. Similar to the principle of light guiding in an optical fiber, the active region acts as a dielectric slab waveguide to confine the light.

The laser cavity of the semiconductor laser of Fig. 2.2 is a Fabry-Perot resonator with two reflective mirrors at both sides. When light reaches one of the facets, part of the light reflects back to the active region as an optical feedback signal. In the most basic structure, the mirrors of semiconductor laser are formed simply by cleaving with a reflectivity of

$$r_m = \left( \frac{n_r - 1}{n_r + 1} \right)^2,$$

where $n_r$ is the refractive index of the gain medium. Typical for most semiconductor lasers, $n_r = 3.5$ and $r_m = 30\%$. For a semiconductor laser to generate light, the round-trip optical gain must be equal to the
overall cavity loss. When current is injected into the $p$-$n$ junction, the semiconductor laser reaches its lasing threshold when the optical gain is just equal to the cavity loss. In semiconductor laser, the optical gain is very large such that high facet loss can be tolerated.

For a cavity length of $L_l$, the modes of the Fabry-Perot cavity have a spacing of $c/(2n_r L_l)$, where $c$ is the speed of light in free space. Note that the refractive index of $n_r$ may change to $n_{eff}$ to include the effect of the slab waveguide. For a typical length of $L_l = 200 - 400 \, \mu m$, the mode spacing of the laser is about 100 to 200 GHz, or about 1 to 2 nm at the wavelength around 1.55 $\mu m$.

The gain medium of semiconductor laser has a gain bandwidth on the order of 100 nm (about 12 THz). At least tens of modes can be supported within the gain bandwidth. In the semiconductor gain medium, due to heterogeneous broadening, the gain is reduced for the lasing wavelength with strong optical power. When the gain at the strong mode is reduced, the other modes can reach the same gain as the strong mode and start lasing. With tens of modes within the gain bandwidth, a semiconductor laser may have tens of lasing wavelengths. To limit the mode number in semiconductor laser, frequency selective loss can be introduced into the laser cavity. In addition to the Fabry-Perot resonator to limit the lasing wavelength to tones that are $c/2n_r L_l$ apart, frequency selective structure can be introduced close to the active region of Fig. 2.2.

Figure 2.3 shows a distributed-feedback (DFB) laser with a grating close to the active region. The grating gives periodic variation of the waveguide refractive index. Instead of two facets, optical feedback is provided in the whole cavity in a DFB laser. The waves propagating in the forward and backward directions are coupled with each other. With the grating, coupling occurs only for wavelength $\lambda_B$ with the Bragg condition of

$$\Lambda_g = m \frac{\lambda_B}{2n_{eff}},$$

where $\Lambda_g$ is the grating period, $n_{eff}$ is the effective refractive index of the active region of the DFB laser, and $m$ is the order of Bragg diffraction. The coupling between the forward and backward waves is the strongest.
for the first-order Bragg diffraction of \( m = 1 \). In order to operate a laser at \( \lambda_B = 1.55 \mu m \), \( \Lambda_g \) is about 235 nm for \( m = 1 \) for \( n_{\text{eff}} = 3.3 \). Such grating can be fabricated by the interferometric pattern of two short wavelength laser beams or electron beam writing.

First operated in the liquid nitrogen temperature of 77 K° (Hall et al., 1962) and then the room temperature (Hayashi et al., 1970), various types of semiconductor laser are described in more details in the books of Agrawal and Dutta (1986), Casey and Panish (1978), and Coldren and Corzine (1995). The DFB structure was first proposed by Kogelnik and Shank (1971). In additional to the DFB structure, there is also a distributed Bragg reflection (DBR) structure with the grating at one end of the waveguide of Fig. 2.3 (Suematsu et al., 1983). As a variation of DFB laser, \( \lambda/4 \)-shifted DFB laser has a \( \lambda/4 \) shift in the middle of the grating to provide a \( \pi/2 \) phase shift (Akiba et al., 1987). Other advanced semiconductor laser structures were reviewed in Suematsu et al. (1992), including the low-cost surface-emitting semiconductor lasers (Iga et al., 1988). DFB laser for WDM applications was reviewed in Funabashi et al. (2004).

In additional to single wavelength DFB laser, there are also monolithically integrated widely tunable multiple wavelength semiconductor lasers (Coldren, 2000, Coldren et al., 2004), external cavity diode lasers tuned by micro-electro-mechanical systems (MEMS) (Anthon et al., 2002), or an array of DFB lasers selected by MEMS (Pezeshki et al., 2002). Those widely tunable lasers can cover the whole C- or L-band for EDFA.

The output intensity of a semiconductor laser follows its injected current. On-off keying signal may be generated by the presence or absence of injection current to turn-on and -off the semiconductor laser. Later in this section presents the rate equations to govern the laser dynamic.

Even when a semiconductor laser has a constant injection current, the output intensity of the semiconductor laser is noisy due to the spontaneous emission in the lasing material. External modulated system that uses a diode laser with constant injection current is still affected by the semiconductor laser noise. The laser phase noise directly affects a phase-modulated signal and laser intensity noise directly affects an on-off keying or amplitude-modulated signal. As a second-order effect, laser phase noise may convert to amplitude noise due to fiber chromatic dispersion.

### 2.2 Rate Equations and Laser Dynamic

The dynamic and noise characteristics of a semiconductor laser can be described by the laser rate equations. In its simplest way, a semiconductor laser converts electrons to photons. The laser electric field
Digital Modulation of Optical Carrier

$E_L$ and the carrier density $n_c$ as a function of time are governed by the following rate equations

$$\frac{dE_L}{dt} = -j\Delta\omega(n_c)E_L + \frac{1}{2} \left[ G(n_c) - \frac{1}{\tau_p} \right] E_L + F_n(t), \quad (2.9)$$

$$\frac{dn_c}{dt} = I_c - \frac{n_c}{\tau_c} - G(n_c)|E_L|^2, \quad (2.10)$$

where $\Delta\omega(n_c)$ is the deviation of oscillation angular frequency from the natural frequency of the laser cavity as a function of carrier density of $n_c$, $G(n_c)$ is the power gain as a function of carrier density $n_c$, $\tau_p$ is the lifetime of the photon particle, $F_n(t)$ is the Langevin random force caused by spontaneous emission, and $I_c$ is the rate of carrier injection, and $\tau_c$ is the carrier lifetime. In the equation of Eq. (2.9), the electric field of $E_L$ should have a unit such that the photon density in the laser cavity is $p_c = |E_L|^2$.

The physical meaning of the above rate equations is obvious. The first term of the right hand side of Eq. (2.9) is the frequency detuning due to carrier injection. In the stationary condition, $\Delta\omega(n_c) = 0$. The second term of the right hand side of Eq. (2.9) is the gain of electric field provided by the lasing medium. Because the laser light losses due to material absorption, emitted out from the laser cavity, the electric field (or photon) has a photon lifetime of $\tau_p$, depending on the laser cavity. If the overall loss of the cavity is $r_e$ due to mirror and internal material loss, the photon lifetime is equal to $\tau_p^{-1} = v_g r_e$ where $v_g$ is the group velocity of light. The last term of the right hand side of Eq. (2.9) is due to spontaneous emission. The Langevin force of $F_n(t)$ is usually modeled as a complex value white Gaussian noise. The carrier rate equation of Eq. (2.10) also has simple physical meaning. $I_c$ is the rate of carrier increase due to current injection, $\tau_c$ is the carrier lifetime, and the last term of Eq. (2.10) is the carrier converting to photons. For a current density of $I_d$, the rate of carrier injection is $I_c = I_d/e$, where $e$ is the electron charge. The detuning of $\Delta\omega(n_c)$ just affects the phase of the electric field $E_L$ and thus the term of $-j\Delta\omega(n_c)E_L$ with the multiplication of “$j$”.

The refractive index of semiconductor material is equal to $n_r(n_c) = n_0 + \Delta n_r(n_c) + j\Delta n_i(n_c)$, where $n_0$ is the refractive index without carrier injection, $n_r(n_c)$ and $n_i(n_c)$ are the real and imaginary parts of the refractive index as a function of carrier density $n_c$. The gain of $G(n_c)$ is proportional to the imaginary refractive index of $-\Delta n_i(n_c)$. The lasing frequency should be aligned with the mode spacing of $n_r(n_c)/L_{lc}$. The term of $\Delta n_r(n_c)$ induces fluctuation in resonance frequency of the cavity and gives $\Delta\omega(n_c)$, generating frequency modulation by carrier injection.
When the semiconductor laser is biased well above the lasing threshold with a small signal injection current of

\[ I_c = I_{c0} + R\{v_m e^{j\omega_m t}\}, \quad (2.11) \]

where \( v_m \) and \( \omega_m \) are the modulation amplitude and frequency, respectively. The electric field and the carrier density have stationary values of \( E_0 = \sqrt{P_{c0}} e^{j\phi_0} \) and \( N_0 \). The rate equations of Eqs. (2.9) and (2.10) can be linearized around \( E_0 \) and \( N_0 \) for \( E_L(t) = \sqrt{P_{c0} + \Delta p(t)} e^{j\phi_0 + j\varphi_n(t)} \) and \( n_c(t) = N_0 + \Delta n(t) \). First of all, both frequency detuning \( \Delta \omega(n_c) \) and medium gain \( G(n_c) \) are linearized by

\[ \Delta \omega(n_c) = -\frac{\partial \omega}{\partial n_c} \Delta n, \quad (2.12) \]

\[ G(n_c) = G(N_0) + \frac{\partial G}{\partial n_c} \Delta n. \quad (2.13) \]

Ignores the Langevin noise of \( F_n \), the rate equations become

\[ \frac{d\phi_n}{dt} = \frac{\partial \omega}{\partial n_c} \Delta n, \quad (2.14) \]

\[ \frac{d\Delta p}{dt} = P_{c0} \frac{\partial G}{\partial n_c} \Delta n, \quad (2.15) \]

\[ \frac{d\Delta n}{dt} = \Delta i_c - \left( \frac{1}{\tau_c} + P_{c0} \frac{\partial G}{\partial n_c} \right) \Delta n - G(N_0) \Delta p, \quad (2.16) \]

where \( \Delta i_c = R\{v_m e^{j\omega_m t}\} \). The photon density is

\[ \Delta p(\omega_m) = \frac{P_{c0} \frac{\partial G}{\partial n_c} v_m}{\omega_R^2 - \omega_m^2 + j\gamma_e \omega_m}, \quad (2.17) \]

where

\[ \omega_R^2 = P_{c0} G(N_0) \frac{\partial G}{\partial n_c}, \quad (2.18) \]

\[ \gamma_e = \frac{1}{\tau_c} + P_{c0} \frac{\partial G}{\partial n_c} \quad (2.19) \]

as the resonance frequency and damping rate of relaxation oscillation. In semiconductor laser, the differential gain of \( \frac{\partial G}{\partial n_c} \) is almost a constant above lasing threshold. The relaxation frequency of \( \omega_R \) of Eq. (2.18) increases with both the photon density of \( P_{c0} \) and the gain of the laser \( G(N_0) \). In general, \( \omega_R \) increases almost linearly with \( P_{c0} \).

The normalized modulation response of a semiconductor laser is the transfer function of

\[ H_p(\omega_m) = \frac{\Delta p(\omega_m)}{\Delta p(0)} = \frac{\omega_R^2}{\omega_R^2 - \omega_m^2 + j\gamma_e \omega_m}. \quad (2.20) \]
The instantaneous frequency of a laser is \( \frac{1}{2\pi} \frac{d\Phi_n(t)}{dt} \). The response of the laser frequency can be written as

\[
\dot{\phi}_n(\omega_m) = \frac{j\omega_m}{2P_{e0}} \frac{\partial \omega_m}{\partial n_c} \Delta p(\omega_m).
\]  
(2.21)

The factor of

\[
\alpha = \frac{\frac{\partial \omega_m}{\partial n_c}}{\frac{1}{2} \frac{\partial G}{\partial n_c}}
\]  
(2.22)

is the linewidth enhancement factor. The linewidth enhancement factor of \( \alpha \) can also be rewritten as

\[
\alpha = \frac{\frac{\partial n_r}{\partial n_c}}{\frac{\partial n_i}{\partial n_c}},
\]  
(2.23)

where \( n_r \) and \( n_i \) represent the real and imaginary parts of the effective refractive index. The linewidth enhancement factor of commonly used DFB laser is about 3 to 4.

The response of the laser phase is

\[
\dot{\Phi}_n(\omega_m) = \frac{\alpha}{2P_{e0}} j\omega_m \Delta p(\omega_m).
\]  
(2.24)

The normalized transfer function of the laser frequency is

\[
H_{\Phi}(\omega_m) = \frac{j \omega_m \omega_R^2}{\omega_R^2 - \omega_m^2 + j \gamma_c \omega_m}.
\]  
(2.25)

The transfer function of Eq. (2.25) will be revisited later in this chapter about direct frequency modulation of a semiconductor laser.

Figure 2.4 shows the photon response of direct-modulation of a high-speed semiconductor laser used for 40-Gb/s applications (Sato, 2002, Sato et al., 2002). The laser response has a second-order response of Eq. (2.20). Similar to Eq. (2.18), the relaxation oscillation frequency increases with the bias current or the output power.

The method to analyze laser dynamic here follows the books of Petermann (1991) and Okoshi and Kikuchi (1988). Laser dynamic can also be studied based on the rate equations of the carrier density and photon density (Lau and Yariv, 1985, Tucker, 1985, Tucker and Kaminow, 1984). Further increase of current injection generates more photons, but the optical gain is reduced to \( G(n_c)(1 - \epsilon_{NL} P_c) \), where \( \epsilon_{NL} \) is a
nonlinear-gain parameter due to saturation effect. In semiconductor material, due to heterogeneous broadening, the factor of $\epsilon_{NL}$ depends on wavelength. Nonlinear gain saturation is a major factor that affects the laser dynamic (Lau and Yariv, 1985, Olshansky et al., 1987, Ralston et al., 1993, Tucker, 1985). The relaxation peaks of the amplitude response of Fig. 2.4 is reduced due to the contribution of $\epsilon_{NL}$ (Ralston et al., 1993) or laser parasitic (Tucker and Kaminow, 1984). High-speed semiconductor laser with a speed of 20 to 40 GHz has been developed (Kjebon et al., 1997, Matsui et al., 1997, Morton et al., 1992, Ralston et al., 1993, Weisser et al., 1996). Semiconductor laser can be directly modulated for 40-Gb/s data rate (Sato, 2002, Sato et al., 2002). The parameters for the rate equations of Eqs. (2.9) and (2.10) can be found using parameter extraction methods (Cartledge and Srinivasan, 1997, Lee et al., 2002b, Salgado et al., 1997).

2.3 Laser Noises

When the semiconductor laser is biased well above the lasing threshold with a fixed current with constant carrier injection of $I_{c0}$ without any modulation, with the Langevin noise, the rate equations of Eqs. (2.14)
Digital Modulation of Optical Carrier

(2.16) can be linearized to

\[
\frac{d\phi_n}{dt} = \frac{\partial \omega}{\partial n_c} \Delta n + \frac{F_i}{\sqrt{P_{c0}}} ,
\]

(2.26)

\[
\frac{d\Delta p}{dt} = P_{c0} \frac{\partial G}{\partial n_c} \Delta n + 2 \sqrt{P_{c0}} F_r ,
\]

(2.27)

\[
\frac{d\Delta n}{dt} = - \left( \frac{1}{\tau_c} + P_{c0} \frac{\partial G}{\partial n_c} \right) \Delta n - G(N_0) \Delta p ,
\]

(2.28)

where \( F_r \) and \( F_i \) are the real and imaginary parts of the Langevin force of \( F_n(t) \).

In very low frequency, the time derivative of Eqs. (2.26) to (2.28) is equal to zero, this is similar to find the direct-current (d.c.) operating point of the linearized equations except that both the phase and amplitude equations of Eqs. (2.26) and (2.27) have noisy driving force of \( F_i \) and \( F_r \), respectively.

From the equation of Eq. (2.27) with \( d\Delta p/dt = 0 \), we obtain

\[
\Delta n = - \frac{F_r}{\sqrt{P_{c0}} \frac{\partial G}{\partial n_c}} ,
\]

(2.29)

that the number of carriers in the laser is fluctuated due to the spontaneous emission of \( F_r \). As spontaneous emission depletes the injected carriers, the number of carriers becomes random. From the rate equation of Eq. (2.28), together with Eq. (2.29), we obtain

\[
\Delta p = \frac{1}{\tau_c} + P_{c0} \frac{\partial G}{\partial n_c} \frac{F_r}{G(N_0) \sqrt{P_{c0}} \frac{\partial G}{\partial n_c}} ,
\]

(2.30)

for the effect of spontaneous emission to the number of photons or the amplitude of the emitted electric field. From Eq. (2.30), the amplitude of the emitted electric field increases with spontaneous emission. The carrier fluctuation of Eq. (2.29) and the photon density fluctuation of Eq. (2.30) have opposite sign. As photon number increase, the carrier has to be consumed for photon creation.

The fluctuation of the angular frequency of Eq. (2.14) is

\[
\frac{d\phi_n}{dt} = - \frac{\alpha F_r}{\sqrt{P_{c0}}} + \frac{F_i}{\sqrt{P_{c0}}} .
\]

(2.31)

From Eq. (2.31), the linewidth enhancement factor of \( \alpha \) gives more phase noise to the laser. The reason of linewidth enhancement of a factor of \( \alpha \)
is due the fluctuation of amplitude of $|E_L|$ changing the carrier intensity because of the rate equations of Eq. (2.9) and the linearized version of Eq. (2.27). However, the refractive index of the lasing material also changes with carrier density through $\Delta \omega(n_c)$. The frequency or phase of the semiconductor laser changes due to the detuning effect from the variation of the refractive index of the lasing material.

From the phase equation of Eq. (2.31), the instantaneous frequency is a Gaussian process and the phase is

$$\phi_n(t) = \int_0^t n_\phi(t)dt,$$  \hspace{1cm} (2.32)

where

$$\sigma_{n_\phi}^2 = \frac{(1 + \alpha^2)\sigma_{Fc}^2}{P_{c0}}$$, \hspace{1cm} (2.33)

where $\sigma_{Fc}^2$ is the spontaneous diffusion coefficient of $\sigma_{Fc}^2 = E\{F^2_c\} = E\{F_{e2}^2\}$.

The phase noise of Eq. (2.32) is a Wiener process with autocorrelation of

$$E\{\phi_n(t_1)\phi_n(t_2)\} = \sigma_{n_\phi}^2 \min(t_1, t_2).$$ \hspace{1cm} (2.34)

For a laser with only phase fluctuation, in normalized form, the electric field is $v(t) = e^{j\phi_n(t)}$ with autocorrelation of

$$R(\tau) = E\{v(t)v^*(t-\tau)\} = E\left\{e^{j\phi_n(t)-j\phi_n(t-\tau)}\right\}.$$ \hspace{1cm} (2.35)

The phase difference of $\phi_n(t) - \phi_n(t-\tau)$ is zero-mean Gaussian distributed with a variance of $\sigma_{n_\phi}^2|\tau|$. The autocorrelation function of the electric field is

$$R(\tau) = \frac{1}{\sqrt{2\pi|\tau|\sigma_{n_\phi}}} \int_{-\infty}^{+\infty} \exp\left(j\phi + \frac{-\phi^2}{2\sigma_{n_\phi}^2|\tau|}\right) d\tau = \exp\left(-\frac{1}{2}\sigma_{n_\phi}^2|\tau|\right).$$ \hspace{1cm} (2.36)

The laser linewidth is

$$S_{\phi_n}(f) = \int_{-\infty}^{+\infty} R(\tau)e^{j2\pi f\tau} d\tau = \frac{\sigma_{n_\phi}^2}{\frac{1}{4}\sigma_{n_\phi}^4 + (2\pi f)^2}$$ \hspace{1cm} (2.37)

as the Lorentzian line-shape.
The full-width-half-maximum (FWHM) linewidth of the semiconductor laser is equal to

$$\Delta f_L = \frac{\sigma_{n_\phi}^2}{2\pi}. \quad (2.38)$$

Because the noise diffusion of $\sigma_{n_\phi}^2$ increases by a factor of $1 + \alpha^2$, the Lorentzian line-shape of semiconductor laser increases by the same factor of $1 + \alpha^2$. When the laser phase noise is white Gaussian noise with spectral density of $\sigma_{n_\phi}^2$, the frequency noise has a spectral density of

$$S_\phi(f) = \frac{\Delta f_L}{2\pi f^2}. \quad (2.39)$$

Without the coupling from carrier density to phase change due to the real part of the refractive of $n_r$, i.e., $\alpha = 0$, we obtain $d\phi_{in}/dt = F_\phi/A_0$. The laser linewidth is given by the well-known Schawlow-Townes formula (Sargent et al., 1974, Yariv, 1997),

$$\Delta f_{L_0} = \frac{2\pi\hbar\omega_L(\Delta \nu_{1/2})^2n_{SP}}{P_0}, \quad (2.40)$$

and

$$\Delta f_L = (1 + \alpha^2)\Delta f_{L_0}, \quad (2.41)$$

where $\omega_L$ is the angular frequency of the laser, $P_0$ is the laser output power, $\Delta \nu_{1/2}$ is the linewidth of the passive laser resonator, and $n_{SP}$ is the spontaneous emission factor representing the population inversion. For a Fabry-Perot resonator, the linewidth of the laser cavity is

$$\Delta \nu_{1/2} = \frac{c}{2n_rL_l} \frac{1 - r_c}{\sqrt{r_c}}, \quad (2.42)$$

where $c/2n_rL_l$ is the free-spectral range with $c$ as the speed of light and $L_l$ as the cavity length, $r_c$ is the round-trip loss of the Fabry-Perot resonator, and $\pi\sqrt{r_c}/(1 - r_c)$ is the finesse of the Fabry-Perot resonator.

The Schawlow-Townes formula is only for the case that the angular frequency of the laser of $\omega_L$ matches to that of the laser cavity. The laser linewidth is broadened due to frequency detuning (Lax, 1967a,b). In semiconductor laser, the refractive index of semiconductor material changes due to carrier injection from the factor of $\Delta \omega(n_c)$. The fluctuation of the laser cavity causes further linewidth broadening. In early 1980s, Mooradian and his coworkers measured the spectral characteristic of semiconductor laser in detail (Fleming and Mooradian, 1981a,b, Welford and Mooradian, 1982). The linewidth broadening by the factor of $1 + \alpha^2$ is exactly due to carrier injection effect (Henry, 1982, 1986,
Henry et al., 1981). The linewidth increase by the factor of $1 + \alpha^2$ is confirmed by measurement.

The linearized equations of Eqs. (2.14) to (2.16) can be solved exactly without the low-frequency approximation. When $\frac{d}{dt}$ is replaced by $j\omega_m$, with constant injection current, the noise spectrum becomes

$$S_{\phi_n}(\omega_m) = \Delta f_{L_0} \left[1 + \frac{\alpha^2 \omega_R^4}{(\omega_R^2 - \omega_m^2)^2 + \gamma_e^2 \omega_m^2}\right],$$  \hspace{1cm} (2.43)

$$S_p(\omega_m) = 8\pi P_{c_0}^2 \Delta f_{L_0} \frac{\omega_m^2 + \gamma_e^2}{(\omega_R^2 - \omega_m^2)^2 + \gamma_e^2 \omega_m^2}.$$ \hspace{1cm} (2.44)

The laser linewidth is equal to $\Delta f_L = (1 + \alpha^2)\Delta f_{L_0} = S_{\phi_n}(0)$. The amplitude noise of $S_p(\omega_m)$ is called the relative-intensity noise (RIN) of the laser.

Figure 2.5 shows the change of RIN and frequency noise with laser bias current or output power. The relaxation peaks of both amplitude and frequency noise are located at about the same frequency. Amplitude noise decreases rapidly with output optical power, frequency noises also decreases slightly with output power (or bias current). While amplitude noise decreases by a ratio of $-40$ dB per decade at very high frequency, the frequency noise approaches a constant of $\Delta f_{L_0}$ at high frequency. At low frequency, the frequency noise is $(1 + \alpha^2)\Delta f_{L_0}$, a factor of $1 + \alpha^2$ larger.
than the high frequency limit. In Fig. 2.5, frequency noise decreases with bias current because $\Delta f_{L0}$ is inversely proportional to the laser output power.

Laser linewidth broadening in semiconductor laser was first observed by the group of Mooradian (Fleming and Mooradian, 1981a,b, Welford and Mooradian, 1982) and was first analyzed by Henry (1982). The impact of relaxation oscillation to the spectrum of semiconductor laser was first observed by Vahala et al. (1983) and analyzed using semiclassical methods by Henry (1983), Kikuchi and Okoshi (1985), Spano et al. (1983), and Vahala and Yariv (1983a,b). Yamamoto (1984a,b) analyzed both amplitude and frequency noise of semiconductor laser based on the quantum mechanical Langevin equations. The theory of laser linewidth was reviewed by the paper of Henry (1986) and the book of Petermann (1991). This section follows the books of Petermann (1991) and Okoshi and Kikuchi (1988).

Figure 2.6 shows a comparison of the measured and theoretical RIN from a high-speed semiconductor laser. The measurement results match well to the theoretical curves. For continuous-wave operation, DFB lasers for WDM applications typically have an output power more than 100 mW as laser chip, a RIN in the range lower than -150 to -160 dB/Hz, and a laser linewidth of few MHz (Funabashi et al., 2004). As RIN is inversely proportional to output power, high-speed laser of Ralston et al. (1993) has larger RIN than WDM laser due to an output power of only
10 to 20 mW. After coupling to an optical fiber, typical WDM laser has a power more than 20 mW.

Laser RIN is usually not an issue for common IMDD and phase-modulated high-speed optical communication systems. Laser phase noise adds directly to the signal phase and degrades a phase- or frequency-modulated signal. Laser phase noise is not an issue for high-speed 10- and 40-Gb/s phase-modulated systems, but for low-speed systems.

Laser RIN increases with optical feedback (Ho et al., 1993, Shikada et al., 1988, Tkach and Chraplyvy, 1986), an optical isolator must be used to eliminated optical feedback into the laser cavity for low-noise laser.

3. External Modulators

External modulators provide the best signal quality for both phase and amplitude modulated signals. An electro-optical crystal with proper orientation provides phase modulation with a voltage applied in the right direction. Lithium Niobate (LiNbO₃) is the most commonly used electro-optical crystal to fabricate external modulator. In this section, external modulator using LiNbO₃ is assumed by default.

3.1 Phase Modulator

The generation of a phase-modulated signal requires an external modulator capable of changing the optical phase in response to an applied voltage. In LiNbO₃, if an electric field is applied along the z-axis of the crystal, the refractive index of the material is changed by

$$\Delta n = \frac{1}{2} n_r^3 r_{33} E_z,$$

(2.45)

where $r_{33}$ is the electro-optic coefficient for a change of refractive index $n_r$ for light propagating along the $z$-direction, and $E_z$ is the electric field along the $z$ direction. In the structure of Fig. 2.7, when a voltage is applied to the electro-optical material, the electric field of $E_z$ is approximately equal to $V/d$, where $d$ is the distance between the two electrodes. The total phase shift over an interaction length of $L_i$ is

$$\Delta \phi_0 = \frac{2\pi \Delta n L_i}{\lambda_c} = \pi n_r^3 r_{33} \frac{V L_i}{d \lambda_c},$$

(2.46)

where $\lambda_c$ is the wavelength of the optical signal in vacuum.

The voltage necessary to provide a phase shift of $180^o$ is equal to

$$V_\pi = \frac{d \lambda_c}{n_r^3 r_{33} L_i}.$$

(2.47)
In the design of a phase modulator, one of the main objectives is to reduce the voltage of \( V_\pi \). The electro-optic coefficient of \( r_{33} \) in LiNbO\(_3\) is the largest among all coefficients for various orientations (Turner, 1966). The parameter of \( n_r^3 r_{33} \) of LiNbO\(_3\) is equal to \( 328 \times 10^{-6} \text{um/V} \) (Alferness, 1982).

In the bulk modulator structure of Fig. 2.7, the modulator \( V_\pi \) can be reduced by increasing the interaction length of \( L_4 \) or reducing the electrode distance of \( d \). However, the capacitance of the capacitor formed between the two electrodes is proportional to \( L_4/d \). The reduction of \( V_\pi \) increases the capacitance and slows down the operation of the modulator.

Figure 2.8 shows a traveling-wave phase modulator. In principle, if the light in the optical waveguide and modulated electrical signal in the microwave electrodes are traveling in the same speed, the traveling-wave phase modulator of Fig. 2.8 has infinite bandwidth. LiNbO\(_3\) has a refractive index of about \( n_r = 2.1 \) to 2.2. The microwave signal must slow down to close to the speed of \( c/n_r \) of the optical signal. The design of the microwave waveguide/electrode must match to the traveling speed of optical signal in LiNbO\(_3\) for high-speed operation.

In the waveguide phase modulator of Fig. 2.8, there are two methods to couple electric field into the optical waveguide. Figure 2.9 shows two waveguide structures using either \( x \)- or \( z \)-cut LiNbO\(_3\) crystal. The electric field lines are along the \( z \)-axis in both cases. Using the \( x \)-cut
material, the microwave electrodes for the transmission line are located in either side of the optical waveguide. Using the z-cut material, the electrode for applied drive signal is located exactly on the top of the optical waveguide. Usually, the z-cut phase modulator has better coupling efficiency between the electric field and the optical waveguide. The $V_T$ of the modulator is still the same as that of Eq. (2.47) with the distance of $d$ taking into account the overlapping between electric and optical fields, i.e., an effective separation of $d$ instead of the physical separation.

**Effects of Velocity Mismatch**

If the microwave electrode impedance is matched to the connecting cable and signal source, the microwave drive signal along the transmission line is

$$V(z, t) = V_0 \Re \left\{ e^{j\beta_m z - j\omega_m t} \right\},$$

(2.48)

where $\omega_m$ is the angular frequency of the microwave signal, and $\beta_m = \omega_m n_m/c + j \alpha_m$ is the propagation constant of the microwave signal with $n_m$ as the effective refractive index of the microwave waveguide, $z$ is the distance from the beginning of the microwave and optical waveguide, and $\alpha_m$ is the loss coefficient of the microwave signal. Traveling with a speed of $c/n_r$, the photon entering the optical waveguide at any time $t_0$ experiences a voltage of

$$V(z, t_0) = V_0 \Re \left\{ \exp \left[ j\beta_m z - j\omega_m \left( t_0 + \frac{zn_r}{c} \right) \right] \right\}$$

$$= V_0 e^{-\alpha_m z} \Re \left\{ e^{j\omega_m d_{12} z - j\omega_m t_0} \right\}$$

(2.49)

at a distance of $z$ from the input of the waveguide, where $d_{12} = (n_m - n_r)/c$ is the walk-off parameter between microwave and optical fields.
For a waveguide length of $L$, the overall phase shift is

$$\Delta \phi(\omega_m) = \frac{\Delta \phi_0 \left(1 - e^{-\alpha_m L} + j d_{12} \omega_m L \right)}{L_i} \frac{1}{\alpha_m - j d_{12} \omega_m}$$  \hspace{1cm} (2.50)$$

as a function of the modulation frequency of $\omega_m$, where $\Delta \phi_0$ is the phase shift from Eq. (2.46) and $L_i = (1 - e^{-\alpha_m L})/\alpha_m$ is the effective interaction length. For a d.c. voltage with $\omega_m = 0$, the phase shift is $\Delta \phi(0) = \Delta \phi_0$. Without waveguide loss of $\alpha_m = 0$, the frequency response is proportional to the sinc function of $\sin(x)/x$ with $x = d_{12} \omega_m L/2$.

Figure 2.10 shows the frequency response of a phase modulator with velocity mismatch. The microwave loss is assumed to be either without loss of $\alpha_m = 0$ or $\alpha_m = \alpha_{m0} \sqrt{f_m}$ due to skin effect, where $f_m = \omega_m / 2\pi$ is the modulation frequency. The loss coefficient is $\alpha_{m0} = 0.2$ dB/cm/$\sqrt{\text{GHz}}$ (Kondo et al., 2002, Sugiyama et al., 2002). The length of the modulator is assumed to be $L = 5$ cm (Sugiyama et al., 2002). With microwave loss, Figure 2.10 shows that the velocity matching of only 95% reduces the bandwidth from 50 GHz to less than 22 GHz.

In the design of external modulator, in addition to impedance match and loss reduction, the design of the microwave electrode for velocity matching is an important issue (Gopalakrishnan et al., 1994, Kawano et al., 1991, Koshiba et al., 1999).
3.2 Amplitude Modulator

An amplitude-shift keying (ASK) signal of Fig. 2.1 can be generated using an amplitude or intensity modulator to turn-on and -off the light. Amplitude modulator functions as a very fast switch. While semiconductor laser can be directly modulated, direct modulation of a semiconductor laser comes with frequency chirp and limited the transmission distance (Corvini and Koch, 1987). An external amplitude modulator usually provides very a high quality signal.

Figure 2.11 shows an amplitude modulator based on the simple equal path-length Mach-Zehnder interferometer. The input optical signal is split into two paths via a Y junction. For illustration purpose, one of the optical paths is phase-modulated and another path remains unmodulated. If the Y junction splits the input signal of $E_i$ into two equal electric fields of $E_i/\sqrt{2}$ each, ignored the path delay, some constant phase shifts, and waveguide loss, the combined signal at another end of the Y junction is

$$
E_o = \frac{E_i}{2} \left[ 1 + \exp \left( j \pi \frac{V(t)}{V_\pi} \right) \right] = E_i \cos \left[ \frac{\pi}{2} \frac{V(t)}{V_\pi} \right] \exp \left[ j \frac{\pi}{2} \frac{V(t)}{V_\pi} \right], \quad (2.51)
$$

where $V_\pi$ is the parameter of the phase modulator at the lower path of Fig. 2.11. From Eq. (2.51), in term of optical intensity, the input and output relationship of a Mach-Zehnder intensity modulator is

$$
\frac{|E_o|^2}{|E_i|^2} = \cos^2 \left[ \frac{\pi}{2} \frac{V(t)}{V_\pi} \right], \quad (2.52)
$$

as a nonlinear sinusoidal transfer function shown in Fig. 2.12. The voltage to turn the modulator from minimum to maximum transmission points is $V_\pi$. Because of some amount of constant phase shift, $V(t) = 0$ in practical modulator is not necessary corresponding to the maximum transmission point as shown in Figure 2.12.
Digital Modulation of Optical Carrier

Figure 2.12. Input-output transfer characteristic of a Mach-Zehnder modulator.

However, in additional to the change of optical intensity, the input and output relationship of Eq. (2.51) also accompanies with phase modulation of \( \exp[j\phi(t)] \) with \( \phi(t) = \pi V(t)/2V_\pi \). The ratio of phase to amplitude modulation is called chirp coefficient (Koyama and Oga, 1988). The Mach-Zehnder modulator in the schematic of Fig. 2.11 has a chirp coefficient of \( \pm 1 \) with the plus or minus sign depending on whether the amplitude modulator is operated in the positive or negative inflection points of Fig. 2.12, respectively. In additional to chirp, the unmodulated path of Fig. 2.11 reduces the modulation efficiency of the modulator. When the microwave electrodes are properly designed, both paths of the Mach-Zehnder modulator can be modulated to improve the modulator efficiency.

Figure 2.13 shows three different waveguide structures for a Mach-Zehnder modulator using x- and z-cut LiNbO\(_3\) crystal. Using the x-cut crystal, the two paths of the Mach-Zehnder modulator are phase-modulated with opposite phase shifts in a push-pull structure. For the input-output relationship of Fig. 2.12, the \( V_\pi \) is reduced by half with respect to the single electrode schematic of Fig. 2.11. Because the two paths are modulated with opposite phases of \( \pm \pi V(t)/2V_\pi \) with a parameter of \( V_\pi \) representing the combined \( V_\pi \) of the two phase modulators, the output field corresponding to Eq. (2.51) is

\[
E_o = \frac{E_i}{2} \left[ \exp \left( -j\pi \frac{V(t)}{2V_\pi} \right) + \exp \left( j\pi \frac{V(t)}{2V_\pi} \right) \right] = E_i \cos \left[ \frac{\pi V(t)}{2V_\pi} \right].
\]

(2.53)

In the single-drive x-cut push-pull modulator of Fig. 2.13, no phase modulation accompanies the amplitude modulation from Eq. (2.53) and
the modulator has zero-chirp. In term of optical intensity, the transfer characteristic follows Eq. (2.52) or Fig. 2.12. The input and output transfer characteristic of Eq. (2.53) may also provide 0 and π phase modulation to an optical single. For \( V(t) \) from 0 to \( V_\pi \), \( E_o \) and \( E_i \) have the same phase. For \( V(t) \) from \( V_\pi \) to \( 2V_\pi \), \( E_o \) and \( E_i \) has opposite phases. Phase modulation provided by zero-chirp modulator has the advantage that amplitude jitter does not give phase jitter, i.e., variation of drive voltage does not transfer to variation of output phase. However, the phase modulation is limited to 0 and π, equivalent to ±1. For phase-modulation operation to minimize loss and obtain optimal amplitude jitter compression, the modulator should operate between two maximum transmission points for a peak-to-peak voltage swing of \( 2V_\pi \).

Figure 2.13 shows two waveguide configurations using z-cut LiNbO₃ crystal. Assume that the two paths have identical structure, the chirp coefficient is equal to (Koyama and Oga, 1988)

\[
\alpha = \frac{V_1(t) + V_2(t)}{V_1(t) - V_2(t)},
\]  

(2.54)

where \( V_1(t) \) and \( V_2(t) \) is the drive voltage of the two paths, respectively. When the two paths are driven by complementary signal with \( V_1(t) = -V_2(t) \), the modulated signal has zero-chirp. In the dual-drive structure, the modulator chirp is adjustable (Gnauck et al., 1991).

The single-drive configuration of Fig. 2.13 for z-cut modulator usually has nonzero chirp coefficient because the electric field lines pass through the two optical waveguides are usually not the same. In normal operation, the chirp coefficient for single-drive z-cut modulator is about ±0.75 (Schiess and Carlden, 1994). With proper waveguide arrangement, the
single-drive $z$-cut modulator can be designed as zero-chirp modulator with reduced efficiency. If the dual-drive $z$-cut modulator is operated as a single-drive modulator with one hot electrode connected to ground, the chirp coefficient is also about the same as a single-drive $z$-cut modulator of about $\pm 0.75$.

In on-off keying for IMDD system, the modulator chirp can be used to compensate for fiber dispersion (Gnauck et al., 1991). In ASK for coherent optical communications, additional phase modulation is not desirable and should be reduced. Currently, most commercially available amplitude modulator is LiNbO$_3$ Mach-Zehnder modulator of Fig. 2.13 (Wooten et al., 2000). Earlier amplitude modulator has various structures besides the Mach-Zehnder interferometer (Alferness, 1981, 1982, 1990, Korotky and Alferness, 1988, Kubota et al., 1980). Instead of just a phase modulator like Fig. 2.8 (Kawano et al., 1989, Tench et al., 1987), after 90's, high-speed modulators are designed mostly for amplitude modulation (Burns et al., 1999, Dolfi et al., 1988, Fujiwara et al., 1990, Howerton et al., 2000, Noguchi et al., 1998, 1995, Wooten et al., 2000). Currently, dual-drive 40 Gb/s modulator with a $V_\pi$ of less than 2 V has been demonstrated experimentally (Sugiyama et al., 2002). Even experimental $x$-cut modulator has a $V_\pi$ less than 3 V (Aoki et al., 2004, Kondo et al., 2002). Commercially available LiNbO$_3$ amplitude modulator has a $V_\pi$ from 4 to 6 V. In additional to LiNbO$_3$, semiconductor materials also have electro-optical effect and can be used to fabricate external modulator (Alferness, 1981, Citlles and Ashley, 1994, Leonberger and Donnelly, 1990, Tsuzuki et al., 2004, Walker, 1987, 1991, Yu et al., 1996). Even silicon has electro-optical effect that can be used for modulator (Dainczi et al., 2000, Jackson et al., 1998, Liu et al., 2004e). External modulator can also use the electro-optical effect in polymer (Chen et al., 1997b, Dalton et al., 1999, Lee et al., 2000, Oh et al., 2001, Shi et al., 1996, 2000). With a refractive index smaller than that of LiNbO$_3$, the speed of electrical and optical signals may be matched in polymer modulator without major engineering efforts. The electro-optic coefficient of other materials are listed in Liu (1996), Saleh and Teich (1991), and Yariv (1997).

The $Y$ junction of the Mach-Zehnder modulator of Fig. 2.11 should split the optical signal into two equal parts. If the power splitting ratio is not equal to unity but a factor of, for example, $\gamma_s$, the ratio of the power at the maximum to the minimum transmission point of Fig. 2.12 is equal to

$$r_{ex} = \left(\frac{1 + \gamma_s}{1 - \gamma_s}\right)^2.$$  (2.55)
The ratio of $r_{\text{ex}}$ is the extinction ratio. Most commercially available LiNbO$_3$ Mach-Zehnder modulator has an extinction ratio larger than 20 dB at d.c. However, the extinction ratio for the eye-diagram is usually limited to 10 to 12 dB because of waveform ripples of the drive signal. Finite d.c. extinction ratio also induces chirp to the optical signal (Kim and Gnauck, 2002, Walklin and Conradi, 1997).

Semiconductor materials can be used to make electroabsorption modulator (EAM) based on Frank-Keldysh effect for bulk semiconductor and quantum-confined Stark effect for quantum well (Bennett and Soref, 1987, Miller et al., 1986, Wood, 1988). Together with amplitude modulation, electroabsorption modulator always gives a chirped output because of the accompany frequency modulation. Electroabsorption modulator can be integrated with many other components (Akulova et al., 2002, Frateschi et al., 2004, Johansson et al., 2004, Mason et al., 2002) or operated in very high-speed (Akage et al., 2001, Choi et al., 2002, Kawanishi et al., 2001, Miyazaki et al., 2003).

3.3 Operation of Amplitude Modulator

When a Mach-Zehnder modulator is biased at the middle positive inflection point and the drive signal has a peak-to-peak voltage of $V_\pi$, the baseband representation of the electric field at the output of the modulator is

$$E_o(t) = \frac{E_i}{2} \left\{ \exp \left[ \frac{j(1 + \alpha)\pi}{2} \frac{V(t)}{V_\pi} \right] \right.$$

$$+ j \exp \left[ -\frac{j(1 - \alpha)\pi}{2} \frac{V(t)}{V_\pi} \right] \right\},$$

where $\alpha$ is the chirp coefficient, and $V(t)$ is the binary drive signal of

$$V(t) = \sum_{k=-\infty}^{+\infty} \frac{b_k V_\pi}{2} p(t - kT),$$

where $b_k = \pm 1$ is the transmitted random data stream, $p(t)$ is the pulse shape of the drive signal, and $T$ is the bit interval of the data. The two terms in Eq. (2.56) correspond to the two phase-modulated paths of the Mach-Zehnder modulator. The differential phase shift between the two phase modulators is $\pi V(t)/V_\pi$. The intensity of the modulator output of

$$|E_o(t)|^2 = |E_i(t)|^2 \sin \left[ \frac{\pi}{4} + \frac{\pi}{2} \frac{V(t)}{V_\pi} \right]^2$$

is independent of the chirp coefficient. When $V(t) = -V_\pi/2$, the output of $|E_o(t)| = 0$ locates at the minimum transmission point. When
Digital Modulation of Optical Carrier

Figure 2.14. (a) Bessel-filtered pulse shape and (b) the corresponding eye-diagram of the optical intensity.

\[ V(t) = +V_\pi/2, \] the output of \[ |E_o(t)| = |E_i(t)|^2 \] is at the maximum transmission point. An external modulator is normally biased at the negative inflection point to give negative chirp for dispersion compensation (Gnauck et al., 1991). The bias at positive inflection point is used here for illustration propose.

The relation of Eq. (2.56) can be used to model all types of Mach-Zehnder modulators having different values of chirp coefficient \( \alpha \). For example, a dual-drive modulator has adjustable chirp, a single-drive \( z \)-cut modulator has a chirp coefficient of \( \alpha = \pm 0.75 \), and an \( x \)-cut push-pull Mach-Zehnder modulator has zero chirp. Because most external modulators have very large extinction ratio, the expression Eq. (2.56) assumes an infinite extinction ratio.

Ideally, the Mach-Zehnder modulator should be driven by a rectangular pulse that has been filtered by a Bessel low-pass filter. Figures 2.14 show both the pulse shape and the corresponding eye diagram when a fifth-order Bessel filter having a bandwidth of either 0.75/\( T \) or 0.5/\( T \) is used\(^1\), where \( T \) is the bit interval for the binary signal. Owing to the nonlinear transfer characteristic of the Mach-Zehnder modulator, a bandwidth of 0.5/\( T \) is sufficient to provide a good eye-diagram in optical intensity, but then the receiver must have a very wide bandwidth to preserve the eye opening to the decision circuits.

While the eye-diagram of Fig. 2.14(b) does not change with the chirp coefficient, the phase of the output electric field of \( E_o(t) \) of Eq. (2.56) varies with chirp coefficient. Figure 2.15 shows the electric field locus changing from the off state of \((0, 0)\) to the on state at the unit circle.

---
\(^1\)The fifth-order Bessel filter has a response of

\[ H(s) = \frac{945}{945 + 945s + 420s^2 + 105s^3 + 15s^4 + s^5} \]

where \( s = jf/f_0 \) with \( f_0 \) adjust for different bandwidth.
The phase change from turn-on to turn-off of an amplitude modulator biased at the positive middle inflection point. The dashed circle is the on state and the origin is the off state. The middle intersection corresponds the bias of $V(t) = 0$ shown as dash-line. The intersection point corresponds to the bias point of $V(t) = 0$ located at $(1/2, 1/2)$. Using a zero-chirp modulator with $\alpha = 0$, the electric field has a constant angle of $\pi/4$. However, with a chirp modulator, the phase changes with the instantaneous drive voltage. With a chirp coefficient of $\alpha = \pm 1$, the output electric field changes along a circle ending at $(0, 1)$ and $(1, 0)$, respectively. The locus of Fig. 2.15 depends only on the chirp coefficient but independent to the waveform of the drive signal.

Figures 2.16 show the single-sided optical spectrum of the optical signal modulated by a random pulse stream with a Bessel-filtered pulse given by Fig. 2.14(a). Fifth-order Bessel filters have bandwidths of $0.75/T$ and $0.5/T$ for Figs. 2.16(a) and (b), respectively. Figs. 2.16 also show the optical spectra for different values of the chirp coefficient of $\alpha = 0, \pm 0.5, \pm 1$. For comparison, Figs. 2.16 also plot the electrical spectrum of the drive signal $V(t)$ for comparison.

Modulator chirp broadens the optical spectrum of the signal. Figures 2.16 show that the approximation of the optical spectrum using the electrical spectrum of the drive signal underestimates the spectral spreading. The differences between the power spectra of $E_o(t)$ and $V(t)$ are more significant when the low-pass filter having the small bandwidth of $0.5/T$ is used. Figures 2.16 show that chirp broadens the optical spectrum, causing more broadening when a Bessel filter having smaller bandwidth, corresponding to longer rise and fall times, is used.
Figure 2.16. The power spectral density of an on-off keying optical signal when a Bessel-filtered pulse is used as the drive signal. The Bessel filter has a bandwidth of (a) 0.75/T and (b) 0.5/T. The dashed lines are the electrical spectrum of the drive signal. [Adapted from Ho (2004e)]

In Figs. 2.16, there are major differences at the integer normalized frequencies of $fT = \pm 1, \pm 2, \ldots$. While the electrical spectrum of the drive signal of $V(t)$ has notches at those normalized frequencies, the optical spectrum has discrete tones at those same frequencies, and also has a tone at $f = 0$ due to the d.c. value of the electric field. The differences between the power spectra of $E_o(t)$ and that of $V(t)$ are more significant when the Bessel filtered pulse has smaller bandwidth. When the pulse bandwidth is 0.75/T, the difference is about 3 dB at the second lobe of $fT = 1.5$ for $\alpha = 0$. When the pulse bandwidth is 0.5/T, the second lobe is at about $fT = 1.25$, the difference increases to about 5 to 12 dB even for a chirp coefficient of $\alpha = 0$.

The effect of the chirp coefficient also depends on the pulse bandwidth. With a bandwidth of 0.75/T, the chirp coefficient does not change the optical power spectral density as much as when the bandwidth is 0.5/T. By comparing Figs. 2.16(a) and (b), we may also conclude that the modulator chirp has a greater effect for a drive signal having longer rise or fall times.

The operation of amplitude modulator is well-known in most optical communication textbooks (Agrawal, 2002, Keiser, 1999, Liu, 1996). The power spectrum of signal with modulator chirp was first analyzed by Ho and Kahn (2004c) using the method of Greenstein (1977), Ho (1999b), and Ho et al. (2000c). Of course, the optical spectrum is routinely measured or simulated for phase-modulated or IMDD optical communication systems.
For the long-term operation of an amplitude modulator based on LiNbO₃, the bias voltage must be actively controlled to compensate for the d.c. drift (Korotky and Veselka, 1996, Nagata, 2000, Nagata et al., 2004). Low-speed controller is usually used for bias control to maintain the optimal biasing point for the modulator.

### 3.4 Generation of RZ-DPSK Signals

Among all digital modulation formats for phase-modulated optical communications, from Table 1.2, DPSK signal may be the most popular scheme using the phase difference of optical carrier to carry information. DPSK signals typically use return-to-zero (RZ) linecode.

Figure 2.17 is a transmitter to generate RZ-DPSK signal using the cascade of an amplitude and phase modulator. The amplitude or intensity modulator is preference to have zero-chirp and generate an RZ pulse train. The phase modulator modulates the phase of the RZ pulse train to generate the RZ-DPSK signal. Similar to Fig. 1.4(a), the drive signal of the phase modulator should be preceded by a precoder to calculate the cumulative parity of the input sequence. The precoder is required such that the phase difference can be either demodulated using the delay-and-multiplier circuits of Fig. 1.4(b) or the direct-detection interferometer of Fig. 1.4(c). In Fig. 2.17, the phases of all pulses given by the amplitude modulator are assumed to be identical or alternating.

The pulse generator of Fig. 2.17 is an amplitude modulator driven by a sinusoidal signal synchronized with the data source. The drive signal of the phase modulator must be arrived to the modulator at the same time as the pulse for optimal performance.
Figure 2.18. Pulse generation for RZ-DPSK modulation: (a) is the driving signal, (b), (c), and (d) are the output intensity. The output pulse trains of (b) and (c) use a driving swing of 2V\textsubscript{T} and bias at maximum and minimum transmission points, respectively. The output pulse trains of (d) use a driving swing of V\textsubscript{T} and biases at either the positive or negative inflection point.

Figure 2.18 shows three different methods to generate the pulse train for RZ-DPSK transmitter of Fig. 2.17. Figure 2.18(a) is the sinusoidal drive signal. Figures 2.18(b) and (c) are the output pulse train in optical intensity when the amplitude modulator is biased at the maximum and minimum transmission point of Fig. 2.12, respectively. From Figs. 2.18(b) and (c), when biased at either minimum or maximum transmission points, the pulse rate is twice the frequency of the sinusoidal drive signal of Fig. 2.18(a). In addition to pulse train generation, the frequency doubling property had also been used for electro-optical mixer (Ho et al., 1997, Sun et al., 1996).

The sinusoidal drive signal of Fig. 2.18(a) has a peak-to-peak voltage of 2V\textsubscript{T} for the output pulse trains of Figs. 2.18(b) and (c). The duty cycles of the pulse train of Figs. 2.18(b) and (c) are 1/3 and 2/3, respectively. The pulse train of Fig. 2.18(b) biased at the maximum transmission point has the same optical phase at all pulses. However, the pulse train of Fig. 2.18(c) biased at the minimum transmission point has the opposite optical phase of 0 and π at adjacent pulses. The pulse train of Fig. 2.18(c) is commonly called carrier-suppressed RZ (CSRZ) pulse. Even with opposite phases at adjacent pulses, the precoder of Fig. 2.17 is still the same but operated with an inverted transmitted data, or the received signal can be inverted after the decision circuits.

The pulse train of Fig. 2.18(d) is generated by the sinusoidal drive signal of Fig. 2.18(a) with V\textsubscript{T} of peak-to-peak voltage swing. The amplitude modulator is biased at either the positive or negative inflection point. The rate of the pulse train of Fig. 2.18(d) is the same as the frequency of the sinusoidal drive signal of Fig. 2.18(a). The phase of all
pulses in the pulse train of Fig. 2.18(d) is identical. The duty cycle of the pulse train is 1/2.

To obtain DPSK signal, the phase modulator in Fig. 2.17 may be a simple phase modulator or a zero-chirp amplitude modulator. From the transfer function of Eq. (2.53) for zero-chirp modulator, amplitude modulator can provide phase modulation. In steady state, if \( V(t) = \{0, 2V_\pi\} \), the output electric field is \( E_0 = \pm E_i \) with a phase difference of 180°. Note that amplitude ripples deviated from 0 and 2\( V_\pi \) do not give a phase other than 0 and 180°. The nonlinear transfer characteristic of Eq. (2.52) also gives intensity compression to the output signal. The amplitude ripples are compressed in the output intensity.

In additional to the dual-modulator configuration of Fig. 2.18 to generate RZ-DPSK signal, a single-drive zero-chirp modulator with push-pull drive signal can be used to generate NRZ-DPSK signal. The drive signal is a NRZ signal with a peak-to-peak voltage swing of 2\( V_\pi \). The modulator must be biased at the minimum transmission point. The drive voltage swings between two maximum transmission points of the transfer characteristic of Fig. 2.12 with phase difference of 180°. The operation of the modulator with and without the phase difference of 180° generates NRZ-DPSK signal. Normally, the modulator is operated with a bias at \( V_\pi \) with a signal swing between \( \pm V_\pi \).

The transmitter of RZ-DPSK signal similar to Fig. 2.17 was first used by Gnauck et al. (2002) and explained in more detail in both Xu et al. (2004) and Gnauck and Winzer (2005). The locations of RZ and DPSK modulators of Fig. 2.17 can be interchanged without changing the function of the transmitter. Wen et al. (2004) proposed method to generate RZ-DPSK signal using only a single dual-drive modulator.

4. Direct Frequency Modulation of a Semiconductor Laser

The laser frequency is determined by the mode spacing of \( c/2L_i n_r \), where the refractive index of \( n_r = n_0 + \Delta n_r(n_c) \) is a function of carrier density of \( n_c \). If the lasing mode remains the same, the frequency changes by a factor of \( \Delta n_r(n_c)/n_0 \) due the change of the laser refractive index. If the phase modulator discussed in Sec. 2.3.1 is placed inside the laser cavity, frequency modulated laser output can also be generated.

From the analysis of Sec. 2.2.2, a semiconductor laser can be directly frequency modulated with a frequency response given by Eq. (2.25). The frequency response of Eq. (2.25) was derived based on the rate equations of Eqs. (2.9) and (2.10) for the dynamic characteristic of a semiconductor laser. However, the thermal effect of carrier heating and cooling due to current injection is ignored in the rate equations of Eqs. (2.9) and
(2.10). In Petermann (1991), taking into account the thermal effect, the response of Eq. (2.25) becomes

\[ H_\phi(\omega_m) = \frac{j\omega_m^2 \omega_R^2}{\omega_R^2 - \omega_m^2 + j\gamma_c \omega_m} + k_{th} \frac{1}{j\omega_m + \omega_{th}}, \]  

(2.59)

where \( k_{th} \) is the relative strength of the thermal effects and \( \omega_{th} \) is the first-order cut-off frequency due to thermal effect.

Figure 2.19 shows the measured frequency response of a semiconductor laser including both the thermal and carrier response, conformed closely to the transfer characteristic of Eq. (2.59). The thermal effect is similar to a first-order response from the second term of Eq. (2.59). The carrier effect is mainly the same as the first term of Eq. (2.59) or (2.25) but the low frequency part is constant instead of falling down to zero due to some large signal effect (Kobayashi et al., 1982).

In order to use the frequency response like that of Fig. 2.19 for frequency modulation, equalization circuits can be used to flatten the response (Alexander et al., 1989, Gimlett et al., 1987, Iwashita et al., 1986). Another method is to use coding to match the spectral density of the data with the response of frequency modulation (Emura et al., 1984, Hooijmans et al., 1990, Noć et al., 1989, Vodhanel and Enning, 1988, Vodhanel et al., 1988).

Some laser structures provide a flat frequency-modulation response. When the reflector section of a DBR laser is modulated by injection current, the frequency response is very flat. A multi-electrode DFB laser can inject complementary currents to different electrodes to eliminate the thermal effect and provide flat frequency-modulation (Kobayashi et al., 1991).
Direct frequency modulation of a semiconductor laser gives continuous-phase FSK (CPFSK) signal. Due to the availability of low-cost semiconductor laser as the signal source, CPFSK was one of the most popular signal formats in early studies of coherent optical communications (Emura et al., 1990a, Imai et al., 1991, Iwashita and Matsumoto, 1987, Iwashita and Takachio, 1988, Ryu et al., 1991b, Saito et al., 1992, Takachio et al., 1989). Recently, there are suggestions to use CPFSK signal for signal transmission (Idler et al., 2004). In the system of Idler et al. (2004), the CPFSK receiver and the optical fiber are both linear optical system. When the optical filter for the CPFSK signal followed the transmitter, a high quality on-off keying signal can be generated.

In addition to direct frequency modulation, a semiconductor laser can also be directly phase-modulated by injection locking (Kobayashi and Kimura, 1982, Kobayashi et al., 1991). However, the requirement of injection locking does not give a simple transmitter implementation.

5. Summary

This chapter briefly introduces methods to generate an optical signal with modulation in amplitude, phase, or frequency to carry digital information. Semiconductor lasers are widely used as the light source for typical lightwave communication systems that also enable directly either amplitude or frequency modulation.

To provide better signal quality, external modulator is used to apply amplitude or phase modulation to an optical signal. External modulator based on electro-optical effect is also discussed in this chapter. Amplitude modulator is mostly used for on-off keying signal to block or unblock an optical signal. Amplitude modulator is also able to generate either PSK or DPSK signals for a phase difference of either 0 or $\pi$.

The popular RZ-DPSK signals are generated by the cascade of a RZ and equivalently a phase modulator. The RZ duty cycle is typically $1/3$, $1/2$, and $2/3$. 