Chapter 11

CAPACITY OF OPTICAL CHANNELS

The overall data rate of a wavelength-division-multiplexed (WDM) system can be increased by many methods. When a wider optical bandwidth is used, more channels can be transmitted to increase the overall system throughput. There are many research activities to open up further optical amplifier bands to increase the system throughput (Bigo, 2004, Bromage, 2004, Islam, 2002, Ono et al., 2003), mostly using Raman amplifier to provide optical gain outside the Erbium-doped fiber amplifier (EDFA) gain bandwidth.

The overall data rate also increases linearly with the spectral efficiency. The usage of a wider optical bandwidth typical requires new optical amplifier technologies and further optical components, so raising spectral efficiency is often the more practical and economical alternatives. From Chapter 9, phase-modulated optical communications enable efficient increase of spectral efficiency without a large degradation on receiver sensitivity. When the optical signal is contaminated by noise, the important issue is the ultimate limits of the spectral efficiency that are determined by the information-theoretic capacity per unit bandwidth (Cover and Thomas, 1991, Shannon, 1948, Yeung, 2002). With the allowance of high complexity and long delay, those limits can closely be approached using Turbo or low-density parity check codes (Berrou, 2003, Berrou et al., 1993, Chung et al., 2001). Recently, those advance error-correction codes are implemented for high-speed optical communications (Mizuochi et al., 2004) after the usage was proposed for sometime (Ait Sab and Lemaire, 2001, Bosco et al., 2003, Cai et al., 2003a, Vasic and Djordjevic, 2002).

In this chapter, we calculate the spectral efficiency limits, considering various system design issues, like unconstrained and constant-intensity
modulation with coherent or direct detection, and in either linear or nonlinear propagation regime. In most of the cases, optical amplifier noises are assumed to be the dominant noise source. Coherent detection allows information to be encoded in two degrees of freedom per polarization, and its spectral efficiency limits are several b/s/Hz in typical terrestrial systems, even considering nonlinear effects. Using constant-intensity modulation or direct detection, only one degree of freedom per polarization can be exploited, reducing spectral efficiency. Using binary modulation, regardless of detection technique, spectral efficiency cannot exceed 1 b/s/Hz per polarization.

When the number of signal and/or noise photons is small, the channel capacity of optical communication systems is also limited by the particle nature of photons. Coherent communication is equivalent to detecting the real and imaginary parts of the coherent states. Direct detection is equivalent to counting the number of photons in the number states. In the coherent states, if both signal and noise are expressed in terms of photon number, quantum effects add one photon to the noise variance, usually providing a channel capacity slightly smaller than the classical limit. The quantum limit of direct detection is determined by photon statistics and also yields a slightly smaller channel capacity than the classical limit.

While the signal-to-noise ratio (SNR) of a fiber link is proportional to the launched power, fiber nonlinearities induce spurious tones via four-wave-mixing and multiplicative noises via both self- and cross-phase modulation. Fiber nonlinearities certainly also limits the spectral efficiencies. This chapter also reviews the studies on the impact of fiber nonlinearities on the spectral efficiency of lightwave communications.

1. Optical Channel with Coherent Detection

The channel capacity, or the maximum spectral efficiency limit, of a discrete-time channel with X and Y as input and output, respectively, is equal to the maximum mutual information between input and output of

\[ C = \max_{p(x)} E \left\{ \log \frac{p(y|x)}{p(y)} \right\}, \] (11.1)

where \( p(x) \) and \( p(y) \) are the probability density function (p.d.f.) of the input of X and output Y, respectively, and \( p(y|x) \) is the conditional p.d.f. of the output given the input. The channel capacity can be rewritten as

\[ C = \max_{p(x)} \{ H(Y) - H(Y|X) \}, \] (11.2)
where the entropy of the output of $H(Y)$ and the conditional entropy of $H(Y|X)$ are

$$H(Y) = - \int p(y) \log p(y) dy,$$

$$H(Y|X) = - \int \int p(x)p(y|x) \log p(y|x) dx dy.$$ (11.3, 11.4)

Intuitively, in special case when $H(Y|X)$ is a constant, the channel capacity can be found by using an output density of $p(y)$ to maximize $H(Y)$. However, in general, when the output p.d.f. of $p(y)$ was given to maximize the output entropy of $H(Y)$, an input p.d.f. of $p(x)$ cannot be found with the condition of $p(y) = \int p(y|x)p(x) dx$. Later in this chapter, the channel capacities of some channels are derived by this special method. Using log(.) instead of log2(.) to calculate entropy, the capacity of Eq. (11.2) has a unit of nat/s/Hz that is 1.44 times less than b/s/Hz.

In additional to Eq. (11.2), the input signal has the constraint of

$$\int g(x)p(x) dx = A,$$ (11.5)

where $g(x) = x$ and $g(x) = x^2$ [or $g(x) = |x|^2$ for multi-dimensional input] are the most common mean and power constraint. As a p.d.f., we also have the probability constraint of

$$\int p(x) dx = 1 \text{ and } \int p(y) dy = 1.$$ (11.6)

### 1.1 Kuhn-Tucker Condition

With constraints of Eqs. (11.5) and (11.6), the optimal problem to find the channel capacity of Eq. (11.2) does not have a simple analytical solution in most cases. Variational principle may be used to derive the Kuhn-Tucker condition for optimality.

Using Lagrange multipliers of $\lambda$ and $\mu$, the cost function of Eq. (11.2) becomes

$$C_{\text{Lag}} = \int p(x) h_{Y|X}(x) dx - \int p(y) \log p(y) dy$$

$$+ \lambda \int g(x)p(x) dx + \mu \int p(x) dx,$$ (11.7)

where

$$h_{Y|X}(x) = -\int p(y|x) \log p(y|x) dy.$$ (11.8)
is a given function determined by the conditional p.d.f. of $p(y|x)$. Using variational method, with $p(x) \rightarrow p(x) + \delta_x(x)$ and $p(y) \rightarrow p(y) + \delta_y(y)$, where both $\delta_x(x)$ and $\delta_y(y)$ are both very small perturbation with the relationship of $\delta_y(y) = \int p(y|x)\delta_x(x)dx$, we obtain

$$\delta C_{\text{log}} = \int h_{Y|X}(x)\delta_x(x)dx - \int [\log p(y) + 1]\delta_y(y)dy$$

$$+ \lambda \int g(x)\delta_x(x)dx + \mu \int \delta_x(x)dx.$$ 

$$= \int \left\{ h_{Y|X}(x) + \lambda g(x) + \mu - \int [\log p(y) + 1]p(y|x)dy \right\}\delta_x(x)dx. \quad (11.9)$$

If $\delta_x(x)$ is a continuous function, we need to solve the integral equation of $\int p(y|x)\log p(y)dy = h_{Y|X}(x) + \lambda g(x) + \mu - 1$. However, $\delta_x(x)$ is not necessary a continuous function but may be a discrete function, we obtain the Kuhn-Tucker condition of

$$\int p(y|x) \log p(y)dy + h_{Y|X}(x) + \lambda [g(x) - A] + C \geq 0, \quad (11.10)$$

with $\lambda > 0$, where the equal sign is satisfied at the locations when $p(x) \neq 0$ [or $\delta_x(x) \neq 0$]. When the optimal $p(x)$ is multiplied to Eq. (11.10) and integrates over the whole region, we obtain $C = H(Y) - H(Y|X)$ for the optimal $p(x)$, conform to the definition of Eq. (11.2).

### 1.2 Unconstrained Channel

For an optical channel with only amplifier noise and power constrained on the input signal of $X$, the output of the channel is given by

$$Y = X + N, \quad (11.11)$$

where $N$ is two-dimensional Gaussian distributed noise. For zero mean $X$ and $Y$, the variance of $Y$ is the summation of $\sigma_y^2 = \sigma_x^2 + 2\sigma_n^2$, where $\sigma_n^2$ is the noise variance per dimension and $\sigma_x^2$ and $\sigma_y^2$ are the variance or power of the input and output, respectively, where the input constraint of Eq. (11.5) is $\int ||x||^2p(x)dx = \sigma_x^2$. The unconstrained channel of Eq. (11.11) requires the usage of coherent detection to recover the two-dimensional component of $Y$. The conditional p.d.f. of the channel is

$$p(y|x) = \frac{1}{2\pi\sigma_n^2} \exp \left[ -\frac{||x-y||^2}{2\sigma_n^2} \right], \quad (11.12)$$

with $x$ and $y$ as two-dimensional vectors. The conditional entropy of $H(Y|X) = h_{Y|X}(x) = \log (2\pi\sigma_n^2) + 1$ is independent of the channel input.
The output density that maximizes $H(Y)$ of Eq. (11.3) is found to be zero-mean two-dimensional Gaussian distribution with overall variance of $\sigma_y^2$ and $H(Y) = \log(\pi \sigma_y^2) + 1$. The Gaussian distribution of the input signal is shown in Fig. 11.1(a).

With only power constraint, the channel capacity is

$$C = \log (1 + \rho_s), \quad (11.13)$$

where $\rho_s = \sigma_x^2/2\sigma_n^2$ is the SNR of the channel.

The unconstrained spectral efficiency limit of Eq. (11.13) was derived by Shannon (1948) and can be found in most textbooks on information theory (Cover and Thomas, 1991, Yeung, 2002).

For continuous $p(x)$, because $\int ||y||^2 p(y|x)dy = ||x||^2 + 2\sigma_n^2$, for $p(y) = e^{-||x||^2/2\sigma_y^2}/2\pi \sigma_y^2$, the Kuhn-Tucker condition becomes

$$\int p(y|x) \log p(y)dy = -\frac{||x||^2}{2\sigma_y^2} - \frac{\sigma_n^2}{\sigma_y^2} - \log(2\pi \sigma_y^2). \quad (11.14)$$

The Kuhn-Tucker condition is conformed by the capacity of Eq. (11.13) with a Lagrange multiplier of $\lambda = 1/2\sigma_y^2$.

1.3 Constant-Intensity Modulation

The input of $X$ in the optical channel of Eq. (11.11) may be a constant-intensity signal similar to a phase- or frequency-modulated signal. In wireless communications, constant-intensity signal is used such that nonlinear amplifiers can be used in the transmitter. When constant-intensity signal is used in optical fiber, both self- and cross-phase modulation gives a constant phase shift to the channel itself or all other WDM channels. Constant-intensity signal may increase the spectral efficiency of an optical signal if self- or cross-phase modulation is the dominant impairment. However, constant-intensity modulation cannot solve the problem
of nonlinear phase noise of Chapter 5. If self-phase modulation induced nonlinear phase noise is the dominant impairment, constant-intensity modulation should not be used.

For constant intensity signal, the input of $X$ should uniformly distributed as a circle with a radius of $A$ as shown in Fig. 11.1(b). The SNR of the channel is $\rho_s = A^2 / 2\sigma_n^2$. While $H(Y|X)$ is the same as that of the unconstrained channel of Eq. (11.11), the output entropy of $H(Y)$ must calculate differently. The two-dimensional output density is equal to

$$p(y) = \frac{1}{4\pi^2\sigma_n^2} \int_{-\pi}^{+\pi} \exp \left[ -\frac{(y_1 - A \cos \theta)^2 + (y_2 - A \sin \theta)^2}{2\sigma_n^2} \right] d\theta,$$

or

$$p(y) = \frac{1}{2\pi\sigma_n^2} \exp \left[ -\frac{y_1^2 + y_2^2 + A^2}{2\sigma_n^2} \right] I_0 \left( \frac{\sqrt{y_1^2 + y_2^2}A}{\sigma_n^2} \right).$$

(11.15)

(11.16)

The channel capacity is equal to

$$C = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(y) \log p(y) dy_1 dy_2 - \log (2\pi\sigma_n^2)$$

$$= -2\pi \int_0^{+\infty} r f(r) \log f(r) dr - \log (2\pi\sigma_n^2)$$

(11.17)

where

$$f(r) = \frac{1}{2\pi\sigma_n^2} \exp \left[ -\frac{r^2 + A^2}{2\sigma_n^2} \right] I_0 \left( \frac{rA}{\sigma_n^2} \right).$$

(11.18)

is the substitute of $r = \sqrt{y_1^2 + y_2^2}$ to $p(y)$ of Eq. (11.16). The integration of Eq. (11.17) has changed from rectangular to polar coordinate and finds that the p.d.f. of $p(y)$ is independent of angle. With large SNR of $\rho_s$, the spectral limit of Eq. (11.17) can be simplified using the asymptotic expression of

$$I_0(\alpha) \sim \frac{e^\alpha}{\sqrt{2\pi\alpha}}$$

(11.19)

and

$$f(r) \sim \frac{1}{(2\pi)^{3/2} A\sigma} \exp \left[ -\frac{(r - A)^2}{2\sigma_n^2} \right].$$

(11.20)

In Eq. (11.20), outside the exponential, we also approximate $r$ with $A$. The asymptotic result for $-2\pi \int_0^{+\infty} r f(r) \log f(r) dr$ is

$$\log \left[ (2\pi)^{3/2} A\sigma \right] + \frac{1}{2}.$$
The asymptotic spectral efficiency under a constant-intensity constraint is

$$C \sim \log \frac{A}{\sigma_n} + \frac{1}{2} \log \frac{2\pi}{e} = \frac{1}{2} \log \rho_s + 1.10 \log 2. \tag{11.22}$$

The asymptotic limit of Eq. (11.22) is half of the unconstrained limit of Eq. (11.13) plus 1.10 b/s/Hz. While the unconstrained signal can use two dimensions, the constant-intensity modulation can only use a single dimension that gives the factor of 1/2 in Eq. (11.22). The factor of 1.10 may come from the usage of a circle instead of just only the $x$-axis of the signal.

In general, constant-intensity signal requires coherent detection or interferometric detection. The constant-intensity spectral efficiency of Eq. (11.22) was first derived by Geist (1990) and re-derived independently by Aldis and Burr (1993) and Ho and Kahn (2002). The derivation here is mostly based on Ho and Kahn (2002) and Kahn and Ho (2004).

2. **Intensity-Modulation/Direct-Detection Channel**

While this book focuses on phase-modulated optical communications, the IMDD channel is very popular as a low-cost solution. The previous section derives the capacity for system with coherent detection, for comparison propose, we derive the capacity of IMDD system in this section.

In IMDD channel limited by amplifier noise, the discrete-time model of the channel is

$$Y = |X + N|^2, \tag{11.23}$$

where the additive noise is the same as that for unconstrained channel of Eq. (11.11). We assume that the random variables of $X$ and $N$ are all complex numbers but $Y$ is positive real random variable. The mean of the output is $m_y = \sigma_x^2 + 2\sigma_n^2$. The conditional p.d.f. of $p(y|x)$ is a noncentral chi-square ($\chi^2$) distribution with two degrees of freedom of

$$p(y|x) = \frac{1}{2\sigma_n^2} \exp \left[ -\frac{y + x_1^2 + x_2^2}{2\sigma_n^2} \right] I_0 \left( \frac{\sqrt{y(x_1^2 + x_2^2)}}{\sigma_n^2} \right) \tag{11.24}$$

with noncentrality parameter of $x_1^2 + x_2^2$ from the two components of the two-dimensional input of $X$. 


2.1 Some Approximated Results

The channel capacity of the IMDD channel of Eq. (11.23) is difficult to derive analytically. Some approximated capacities for IMDD channel are given here.

One-Dimensional Gaussian Channel Approximation

If the channel of Eq. (11.23) is rewritten as \( Y = |X|^2 + X \cdot N^* + X^* \cdot N + |N|^2 \), in high SNR, the signal is of \(|X|^2\) is a \(\chi^2\) random variable with two degrees of freedom with a variance of \(2\sigma_x^4\). Ignored the small noise of \(|N|^2\), the noise of \(X^* \cdot N + X \cdot N^*\) has a variance of \(4\sigma_x^2\sigma_n^2\). With a SNR of \(\sigma_x^4/2\sigma_n^2\) as a one-dimensional Gaussian channel, the channel capacity is approximately equal to

\[
C \sim \frac{1}{2} \log \rho_s,
\]

where \(\rho_s\) is the same as that for Eq. (11.13). Compared with the Shannon limit of Eq. (11.13), in this approximation, the capacity for IMDD channel is approximately half of the Shannon limit.

The channel capacity of Eq. (11.25) is first derived by Desurvire (2000) and also used in Wegener et al. (2004). The approximated capacity of Eq. (11.25) is very simple. Here, we assume that \(X\) is Gaussian distribution but Desurvire (2000) does not assume an input distribution of \(X\). Note that \(|X|^2 > 0\) as a positive random variable. Gaussian distribution with a variance of \(2\sigma_x^2\) may have negative value, giving larger entropy than \(|X|^2\). This approximation gives larger channel capacity than the exact results.

Maximum Output Entropy

For an output of \(Y > 0\) with a mean constraint of \(m_y\), the output entropy of \(H(Y)\) is maximized by the exponential distribution of

\[
p(y) = \frac{1}{m_y} \exp \left(-\frac{y}{m_y}\right), \quad y > 0,
\]

with \(H(Y) = \log m_y + 1\). In order to obtain the exponential distribution of \(p(y)\) in Eq. (11.26), \(X + N\) may have a two-dimensional Gaussian distribution with overall variance of \(m_y\). Because the noise of \(N\) is a two-dimensional Gaussian distribution with per dimension variance of \(\sigma_n^2\), the input \(X\) may also have a two-dimensional Gaussian distribution with an overall variance of \(\sigma_x^2 = m_y - 2\sigma_n^2\).

If the input of \(X\) is two-dimensional Gaussian distributed with overall variance of \(\sigma_x^2\), its intensity of \(x_1^2 + x_2^2\) has the exponential distribution
of \( p(y_i) = \exp(-y_i/\sigma_n^2)/\sigma_x^2 \). The channel capacity is

\[
C = \log m_y + 1 + \int_0^{+\infty} p(y_i)dy_i \int_0^{+\infty} g(y, y_i) \log g(y, y_i)dy,
\]

where

\[
g(y, y_i) = \frac{1}{2\sigma_n^2} \exp \left[ \frac{y + y_i}{2\sigma_n^2} \right] I_0 \left( \frac{\sqrt{y} - \sqrt{y_i}}{\sigma_n^2} \right).
\]

Using the asymptotic expression of Eq. (11.19) for large SNR, we obtain

\[
g(y, y_i) \sim \frac{1}{2\sqrt{2\pi}\sigma_n^2 y_i} \exp \left[ -\frac{(\sqrt{y} - \sqrt{y_i})^2}{2\sigma_n^2} \right],
\]

and

\[
-\int_0^{+\infty} g(y, y_i) \log g(y, y_i)dy \sim \frac{1}{2} \log \left( 8\pi e \sigma_n^2 y_i \right)
\]

for a one-dimensional Gaussian p.d.f. with variance of \( 4\sigma_n^2 y_i \), that is approximately equal to the variance of the noncentral \( \chi^2 \) distribution of Eq. (11.24) or (11.28) with \( y_i = x_1^2 + x_2^2 \). In the integration of Eq. (11.30), we use the approximation of \( y \sim y_i \) for large SNR. We obtain

\[
H(Y|X) \sim \frac{1}{2} \log \left( 8\pi e \sigma_n^2 \right) + \frac{1}{2} \log \sigma_x^2 - \frac{1}{2} \gamma_e,
\]

where \( \gamma_e = 0.577 \) is the Euler gamma constant. We obtain

\[
C \sim \frac{1}{2} \log \frac{\sigma_x^2}{2\sigma_n^2} + \frac{1}{2}(1 + \gamma_e) - \frac{1}{2} \log 4\pi
\]

\[
= \frac{1}{2} \log \rho_s - 0.688 \log 2.
\]

The channel capacity of Eq. (11.32) was derived in Hall (1994), Hall and O'Rourke (1993), and Kahn and Ho (2004) as an approximation. Because \( h_{Y|X}(x) \) is not a constant for IMDD channel, the p.d.f. that maximizes the output entropy does not necessary also give the channel capacity.

The channel model of Eq. (11.23) assumes that a polarizer precede the receiver to filter out the noise from the polarization orthogonal to the signal. Without the polarizer, the output signal is equal to

\[
Y = |X + N_1|^2 + |N_2|^2,
\]

where \( N_1 \) and \( N_2 \) are two independent two-dimensional Gaussian distributed random variables from both polarizations. Unlike the model of Eq. (11.23), we are not able to find an input density of \( X \) for the
model of Eq. (11.33) to give the output density of Eq. (11.26) for $Y$. We assume a Gaussian input density for $X$ to give a lower bound for the ultimate spectral limit. The characteristic function of the output $Y$ for the model of Eq. (11.33) is

$$
\Psi_Y(\nu) = \frac{1}{[1 - j\nu(\sigma_x^2 + 2\sigma_n^2)][1 - j2\nu\sigma_n^2]},
$$

(11.34)

where the first factor is the characteristic function of $|X + N_1|^2$ and the second factor is that for $|N_2|^2$. Taking an inverse Fourier transform, the p.d.f. of the output $Y$ is

$$
p(y) = \frac{1}{\sigma_x^2} \left[ \exp \left( -\frac{y}{\sigma_x^2 + 2\sigma_n^2} \right) - \exp \left( -\frac{y}{2\sigma_n^2} \right) \right].
$$

(11.35)

The conditional entropy of $H(Y|X)$ can be calculated using the conditional p.d.f. of noncentral $\chi^2$ distribution with four degrees of freedom. In the channel model of Eq. (11.33), both $H(Y)$ and $H(Y|X)$ must be evaluated numerically. Without going into detail, with Gaussian input, the asymptotic limit of Eq. (11.32) is valid for both the one or two polarization noise model of Eqs. (11.23) and (11.33), respectively.

**Half-Gaussian Input Distribution**

The Gaussian input distribution that maximizes the output entropy cannot give the maximum spectral efficiency. As a counter example, if the input electric field is one-dimensionally Gaussian distributed with variance of $\sigma_x^2$, the input intensity of $y_i = x_i^2$ has a p.d.f. of $p(y_i) = e^{-y_i/2\sigma_x^2}/\sqrt{2\pi\sigma_x^2}$. Using the asymptotic results of Eq. (11.30) and similar to Eq. (11.31) but using different $p(y_i)$, we obtain

$$
H(Y|X) \sim \frac{1}{2} \log \left( 4\pi \sigma_n^2 \sigma_x^2 \right) + \frac{1}{2} - \frac{1}{2} \gamma_e.
$$

(11.37)

The output distribution can be found analytically as

$$
p(y) = \frac{1}{2\sigma_n \sqrt{\sigma_n^2 + \sigma_x^2}} \exp \left[ -\frac{y(2\sigma_n^2 + \sigma_x^2)}{4\sigma_n^2(\sigma_n^2 + \sigma_x^2)} \right] I_0 \left[ \frac{y\sigma_x^2}{4\sigma_n^2(\sigma_n^2 + \sigma_x^2)} \right]
$$

(11.38)

with asymptotic entropy of $H(Y) \sim \log \sigma_x^2 + \frac{1}{2} (\log \pi - 1 - \gamma_e)$. The asymptotic channel capacity is

$$
C \sim \frac{1}{2} \log \rho_s - \frac{1}{2} \log 2,
$$

(11.39)
Capacity of Optical Channels

Figure 11.2. The maximum spectral efficiency of optical channel in linear regime. Those for IMDD channels are approximation using various input distributions. The dashed lines are asymptotic limits for constant-intensity and IMDD signal. The two curves with Gaussian input for IMDD signal include noise from one or two polarizations.

that is 0.5 b/s/Hz worse than half of the Shannon limit.

Mecozzi and Shtaif (2001) uses the above approximation to find the maximum spectral efficiency of IMDD channel. This is an example to shown that to maximize the output entropy does not necessary give the channel capacity.

Figure 11.2 shows the ultimate spectral efficiency of optical channel in linear regime as a function of SNR $\rho_p$. The unconstrained capacity is directly calculated from Eq. (11.13). The capacity of constant-intensity modulation is calculated by Eq. (11.17) by numerical integration. For Gaussian input, the capacity of IMDD channel is calculated using Eq. (11.27) using numerical integration including and excluding the noise from the polarization orthogonal to the signal. For half-Gaussian input, the channel capacity is calculated directly using Eqs. (11.3) and (11.4).

The asymptotic limits of constant-intensity modulation of Eq. (11.22) and IMDD channel of Eqs. (11.32) and (11.39) are both plotted as dashed lines. From Fig. 11.2, the asymptotic limit of Eq. (11.22) is very accurate for constant-intensity modulation in a wide range of SNR. The asym-
totic limit for IMDD is valid for a SNR larger than 10 dB. Figure 11.2 also shows the limit of $\frac{1}{2} \log_2 \rho_s$ that is larger than other limits.

2.2 Exact Capacity by Numerical Calculation

For the IMDD channel of Eq. (11.23) with the conditional p.d.f. of Eq. (11.24), the channel output is the intensity $Y \geq 0$ but monotonic one-to-one transfer to the amplitude of $R = \sqrt{Y} \geq 0$ does not change the channel capacity. Using the input and output amplitude random variables of $S$ and $R$, respectively, the conditional p.d.f. becomes a Rice distribution of

$$p(r|s) = \frac{r}{\sigma^2_n} \exp \left[ -\frac{r^2 + s^2}{2\sigma^2_n} \right] I_0 \left( \frac{rs}{\sigma^2_n} \right), \quad r, s \geq 0. \quad (11.40)$$

The channel capacity, or the maximum spectral efficiency limit, is also equal to the maximum mutual information of

$$C = \max_{p(s)} E \left\{ \log_2 \frac{p(r|s)}{p(r)} \right\}, \quad (11.41)$$

where $E\{\cdot\}$ denotes expectation, $p(s)$ and $p(r) = \int_0^\infty p(r|s)p(s)ds$ are the p.d.f. of the input and output amplitudes, respectively. The channel capacity is

$$C = \max_{p(s)} \{H(R) - H(R|S)\}, \quad (11.42)$$

where the entropy of the output of $H(R)$ and the conditional entropy of $H(R|S)$ are

$$H(R) = -\int p(r) \log_2 p(r) dr, \quad (11.43)$$

$$H(R|S) = -\int \int p(s)p(r|s) \log p(r|s) dr ds, \quad (11.44)$$

where all integrations are from 0 to $+\infty$. The capacity of Eq. (11.42) should be evaluated together with the average power and probability constraints of

$$\int s^2 p(s) ds = \sigma^2_x, \quad \int p(s) ds = 1. \quad (11.45)$$

Based on different assumptions, three algorithms are used to find the optimal input distribution to maximize the channel capacity given by Eq. (11.42).

At large amplitude of $r, s \gg \sigma_n$, the conditional p.d.f. of $p(r|s)$ is approximately a one-dimensional Gaussian distribution with a variance of $\sigma^2_n$. In the Kuhn-Tucker condition of Eq. (11.10), the corresponding
function of $h_{R|S}(s) = \frac{1}{2} \log(2\pi e \sigma_n^2)$ is a constant at large amplitude. At large amplitude of $r$ and $s$, similar to Gaussian channel, the output amplitude may have a tail distribution of $p(r) \sim e^{-\lambda r^2}$ where $\lambda > 0$ is the Lagrange multiplier. The input amplitude also has a tail distribution of $p(s) \sim e^{-\kappa s^2}$ where $\lambda$ and $\kappa$ has the relationship of $\lambda^{-1} = 2\sigma_n^2 + \kappa_s^{-1}$. With a constant $h_{R|S}(s)$, the tail distribution of both $p(s)$ and $p(r)$ approaches a continuous distribution. However, the above argument is not sufficient to prove that both the input and output amplitude is continuously distributed with a Gaussian tail. Alternatively, the input and output may have many points very close to each other at large amplitude. In practice, the small probability at large amplitude does not affect the capacity of a practical channel.

At small amplitude of $s$ approaches zero, the function of $H_{R|S}(s)$ approaches $H_{R|S}(0) = \frac{1}{2} \log(2\pi e \sigma_n^2)$. At low intensity, the input p.d.f. has discrete points as shown in Fig. 11.1(c).

**Arimoto Algorithm**

The Arimoto algorithm can calculate the channel capacity iteratively (Arimoto, 1972, Blahut, 1972, Cover and Thomas, 1991). The single optimization of Eq. (11.41) can change to double iterative optimization of

$$C = \max_{p(s)} \max_{q(s|r)} \int \int p(s)p(r|s) \log \frac{q(s|r)}{p(s)} dr ds.$$  \hspace{1cm} (11.46)

Given an input distribution of $p(s)$, the optimal conditional p.d.f. for $q(s|r)$ is

$$q(s|r) = \frac{p(s)p(r|s)}{\int p(s)p(r|s) ds}.$$  \hspace{1cm} (11.47)

Given $q(s|r)$, with the condition of Eq. (11.45), the optimal input distribution of $p(s)$ is

$$p(s) = \frac{\exp \left[ \int p(r|s) \log q(s|r) dr - \lambda s^2 \right]}{\int \exp \left[ \int p(r|s) \log q(s|r) dr - \lambda s^2 \right] ds},$$  \hspace{1cm} (11.48)

in which the multiplier of $\lambda$ can be found by the power constraint of $\int s^2 p(s) ds = \sigma_n^2$. In the Arimoto algorithm, the two procedures of Eqs. (11.47) and (11.48) should be operated iteratively to find the optimal distribution. In practice, with an original input amplitude distribution of $p_{old}(s)$, the two steps of Eqs. (11.47) and (11.48) can be combined into a single step to give a new input amplitude distribution of $p_{new}(s)$
In the expression of Eq. (11.49), inside the exponential is a factor the same as Kuhn-Tucker condition of Eq. (11.10) with \( \int p(r|s) \log p(r) dr + h_{R|S}(s) \). Comparing Eq. (11.49) with the condition Eq. (11.10), the ratio of \( r(s) \) decreases the probability where the condition of Eq. (11.10) is greater than zero and increases the probability where the condition of Eq. (11.10) is small than zero. At locations in which the Kuhn-Tucker condition is conformed, the input probability is converged to a fixed value. The Arimoto algorithm can operate for an initial p.d.f. of \( p_{\text{ini}}(s) \) either discrete or continuous. For example, for the case of Fig. 11.1(c) with a discrete point at \( s = 0 \), the initial p.d.f. can be \( p_{\text{ini}}(s) = po\delta(s) + (1 - p_0)p_{\text{ini,1}}(s) \). However, in the region that \( p_{\text{ini}}(s) = 0 \), the algorithm cannot obtain a nonzero probability of \( p(s) > 0 \) afterward. With initial nonzero probability of \( p_{\text{ini}}(s) > 0 \), the algorithm can converge to a very small probability of \( p(s) \) approaching zero.

To implement the Arimoto algorithm of Eqs. (11.49) and (11.50) for continuous amplitude p.d.f., the channel needs to be first discretized to \( p(s_i) \Delta s \) for the input and \( p(r_j) \Delta r \) for the output. The Gaussian tail is also given by the multiplier of \( \lambda \) and \( \kappa_s \).

**Numerical Optimization**

If an artificial peak-power (or equivalently peak-amplitude) constraint is imposed to the IMDD channel, the optimal input distribution is discrete (Abou-Faycal et al., 2001, Smith, 1971). Practical system should have a peak-amplitude constraint of \( s \leq s_{\text{max}} \) limited by the maximum rating of the transmitter or fiber nonlinearities. Optical amplifiers also cannot provide an infinitely large amplitude, even with very small probability. Given certain number of discrete points, numerical nonlinear programming algorithm can be used to find the optimal distribution. The input distribution of \( p(s) = \sum_{k=1}^{K} p_k \delta(s - s_k) \) is fully determined by \( 2K \) parameters of \( p_k \) and \( s_k, k = 1, \ldots, K \). The channel capacity of Eq. (11.42) can be maximized for those \( 2K \) parameters with the constraints of

\[
0 \leq s_1 < s_2 < \cdots < s_{K-1} < s_K \leq s_{\text{max}}, \tag{11.51}
\]
and
\[ \sum_k p_k = 1, \quad \text{and} \quad \sum_k p_k s_k^2 = \sigma_x^2. \] (11.52)

To determine whether those 2\( K \) parameters are the global optimum of the channel capacity of Eq. (11.42), the Kuhn-Tucker condition of Eq. (11.10) can be used to verify the optimality of the parameters. If the Kuhn-Tucker condition cannot be conformed, an additional discrete points can be added to the optimization procedure until the conformance of Eq. (11.10). In the condition of Eq. (11.10), equality must be satisfied at \( s_k \) and the inequality in \( s \neq s_k \). In all cases, \( s_1 = 0 \) is one of the solution. Depending on the ratio of \( s_{\text{max}}^2/\sigma_2^2 \) and SNR of \( \rho_s \), the maximum point of \( s_K \) is usually but not always equal to \( s_{\text{max}} \).

The nonlinear programming algorithm can be initiated with two discrete points of \( K = 2 \) with the number of discrete points increasing until the conformance of the Kuhn-Tucker condition. The channel capacity is also increased with the number of discrete points and the nonlinear programming algorithm can stop with a stable capacity or one of the discrete point has zero probability. At high SNR, the Kuhn-Tucker condition of Eq. (11.10) is difficult to ideally verify. A convergent capacity can be used instead at high SNR.

With fixed positions of \( s_k, k = 1, \ldots, K \), the Arimoto algorithm can iteratively find the optimal probability of \( p_k, k = 1, \ldots, K \). While possible, the Arimoto algorithm requires lengthy computation to find the optimal positions of \( s_k \).

**Combined Numerical Optimization and Arimoto Algorithm**

As discussed earlier, the Arimoto algorithm can be modified to include some discrete points, especially for a single discrete point at zero intensity. Instead of using continuous distribution as initial assumption, the algorithm is modified with discrete probability at zero intensity. With more than one discrete point, other than the single point at zero intensity, the positions of other discrete points must be optimized accordingly. With the prior assumption that there are several discrete points at small input amplitude, numerical optimization can be used to find the locations of those optimal discrete points and Arimoto algorithm is used to find the optimal probability (both continuous and discrete parts). The two procedures are used alternatively with increase channel capacity in each iteration.

Based on the above three algorithms, the optimal input distribution is evaluated to maximize the channel capacity of Eq. (11.42). Figure 11.3 shows the channel capacity as a function of SNR for IMDD channel of Eq. (11.23). Different algorithms give different input distributions as
shown in Fig. 11.4 for $\rho_s = 10$ dB but the same channel capacity in Fig. 11.3. The Shannon limit of Eq. (11.13) is also shown in Fig. 11.3 for comparison.

Calculated using numerical optimization, Figure 11.4 shows the optimal 9 discrete points with the corresponding probability for a peak-power constraint 10 times the average power. Theoretically, the larger is the peak-power constraint, the larger is the channel capacity. Numerically, the usage of 10 times the average power as peak-power constraint gives a channel capacity virtually the same as other algorithms.

Without the artificial peak-power constraint, the optimal input distribution has continuous component or infinite number of points very close to each other in its tail. The Arimoto algorithm gives only a single discrete point at zero intensity. Instead of using continuous distribution as initial assumption, the algorithm is modified with discrete probability at zero intensity. Figure 11.4 shows the optimal input distribution with discrete probability at zero intensity (overlapped with the square there) and continuous-distribution as dash-dotted line. Note that unlike Rayleigh channel in Abou-Faycal et al. (2001), the Arimoto algorithm converges very fast for IMDD channel. Figure 11.4 also shows the optimal input distribution with two discrete points at low intensity (empty
Figure 11.4. The input and output probability density as a function of normalized input and output amplitude of $r/\sigma_n$, $s/\sigma_n$ for SNR of $\rho_s = 10$ dB. [Adapted from Ho (2005b)]

All three algorithms converge to the same channel capacity without observable difference. However, Figure 11.4 shows that the input distribution of $p(s)$ has significant difference from one algorithm to others. At $\rho_s = 10$ dB, the three algorithms give a channel capacity within ±0.05% of each other. The output distributions of $p(r)$ in Fig. 11.4 from the three input distributions also have no significant difference at small amplitude. Only the tail distribution of $p(r)$ has major difference when the input distribution is totally or partially discrete.

Figure 11.3 also shows the channel capacity for binary signal (two discrete points in the input distribution) calculated by numerical optimization. Binary signal achieves the channel capacity for SNR less than about 5 dB. The channel capacity of binary signal was also calculated in Mecozzi and Shtaif (2001) with the assumption that the two levels are equally probable. Except for large SNR ($\rho_s > 12$ dB), the optimal binary signal has larger probability at zero-intensity and smaller probability at nonzero-intensity. For example, if only 10% probability
at nonzero-intensity, the nonzero-intensity is 10 times the averaged intensity as compared with twice the averaged intensity for equal-probable case. Compared to similar curve in Mecozzi and Shtaif (2001), the binary signal can achieve better channel capacity at low SNR. By sending occasional pulses with large intensity above the noise, the system is similar to the essence of return-to-zero (RZ) signaling. Systems with powerful forward error correction (Mizuochi et al., 2004) can operate around $\rho_s = 5$ to 7 dB. Binary instead of multilevel signals may be sufficient for those systems.

Based on the same channel model of Eq. (11.23), the half-Gaussian distribution of Mecozzi and Shtaif (2001) and Fig. 11.2 provides a lower bound and is 0.07 to 0.21 b/s/Hz worse than the optimal distribution calculated by numerical algorithms. The optimal distribution has a tail profile of $e^{-\kappa r^2}$, similar to that of half-Gaussian distribution. For very large SNR ($\rho_s > 30$ dB), the discrete region of Fig. 11.4 at low intensity becomes insignificant and the half-Gaussian distribution should be very close to the optimal distribution.

### 2.3 Thermal Noise Dominated IMDD Channel

For an IMDD system without optical amplifiers, the system is limited by thermal noise of $N_{th}$ from the receiver circuitry. The photodetector gives the intensity of the signal as $|X|^2$, with additive thermal noise, the receiver output is

$$Y = |X|^2 + N_{th},$$

where $|X|^2 \geq 0$ because optical intensity is always positive. Naturally, there is a peak constraint of the instantaneous optical power of $|X|^2 \leq P_{\text{max}}$. From Smith (1971), the optimal input distribution to give the ultimate spectral efficiency is a discrete distribution.

Using numerical optimization, Figure 11.5 shows the channel capacity of thermal noise limited IMDD channel with peak intensity constraint. The SNR of Fig. 11.5 is defined as $m_{|X|^2}/\sigma_{th}$, where $m_{|X|^2} = E\{|X|^2\}$ is the average of the optical intensity and $\sigma_{th}^2 = E\{N_{th}^2\}$ is the variance of the Gaussian distributed zero-mean thermal noise. This definition of SNR is consistent with the definition of $Q$ factor for binary equal probable signal in Eq. (3.139). Figure 11.5 shows the channel capacity when the peak intensity of $P_{\text{max}}$ is either 3 or 10 times larger than the average optical intensity of $m_{|X|^2}$.

At high SNR, the output of $Y$ may be considered as confined to $Y \geq -\sigma_{th}$, the optimal distribution to maximize the output entropy is exponential distribution with p.d.f. of $p(y) = \exp[-(y + \sigma_{th})/m_y]/m_y, y \geq -\sigma_{th}$, where $m_y = m_{|X|^2} + \sigma_{th}$, where $\sigma_{th}$ is added to take into ac-
count the small negative value of the output $Y$. The output entropy is $H(Y) \sim \log (m_y) + 1$ and the channel capacity has an asymptotic limit of

$$C \sim \log (m_y) + 1 - \frac{1}{2} \left[ \log \left( 2\pi \sigma_{th}^2 \right) + 1 \right]$$

$$= \log \left( 1 + \frac{m_y x^2}{\sigma_{th}^2} \right) - 0.604 \log 2. \quad (11.54)$$

This asymptotic limit is also shown in Fig. 11.5. Comparing the channel capacity of Fig. 11.5 limited by thermal noise with similar channel capacity of Fig. 11.3 limited by amplifier noise, the channel capacity limited by thermal noise is significantly larger at low SNR and slightly smaller at high SNR. At low SNR, IMDD channel limited by thermal noise can have negative output but that limited by amplifier noise always has positive output. The channel capacity improves with the possibility of negative output.

The asymptotic limit for the channel capacity limited by amplifier noise is 0.1 b/s/Hz larger than that limited by thermal noise at high SNR. The conditional entropy of thermal-noise-limited channel is a constant independent of input signal. The conditional entropy $H(R|S)$ of IMDD channel limited by amplifier noise at small input is half of that at large input. The channel capacity increases slightly for channel limited by amplifier noise due to the reduction of conditional entropy.
The discrete nature of the optimal input distribution was first shown in Smith (1971) for Gaussian channel with peak-amplitude constraint and later used by Abou-Faycal et al. (2001) for Rayleigh channel. Many channels have discrete optimal input distribution (Huang and Meyn, 2003, Shamai and Bar-David, 1995).

Most short-distance optical communication systems without optical amplifiers are primarily limited by thermal noise. Those systems usually use single-channel on-off keying and the channel capacity is usually not a major issue.

3. Quantum-Limited Capacity

When the number of signal and/or noise photons is small, quantum effects must be considered to compute the capacity of an optical channel. While a classical continuous-time channel can be converted to a discrete-time channel using the sampling theorem (Shannon, 1948), the particle-based quantum channel does not have the corresponding sampling theorem. However, one may assume that the measurement is made within a time interval limited by the channel bandwidth. Hence, most studies of quantum-limited capacity assume a discrete-time channel. Corresponding to a narrow-band WDM channel, this type of channel is generally referred to as narrow-band channel.

In coherent detection of an optical signal, the input signal can be assumed as a coherent state (Caves and Drummond, 1994, Gardiner, 1985, Yamamoto and Haus, 1986). If there is an average of \( \bar{n}_S \) signal photon and \( \bar{n}_N \) noise photon, the channel capacity is the same as that of Eq. (11.13) with one additional noise photon of \( C = \log \left( 1 + \frac{\bar{n}_S}{1 + \bar{n}_N} \right). \) (11.55)

The classic SNR of \( \rho_s \) is the ratio of \( \bar{n}_S \) to \( \bar{n}_N \). If a quantum SNR is defined by the ratio of \( \bar{n}_S/(1 + \bar{n}_N) \), the quantum capacity of Eq. (11.55) is the same as that of the classic limit of Eq. (11.13). Intuitively, there is a minimum of one noise photon in coherent state. Here, the SNR is expressed as the ratio of photons instead of power like Eq. (3.36). The relationship of SNR to number of photons is very obvious from Table 3.1. For typical optical communication systems with amplifier noise, the quantum-limited capacity is slightly less than the classic limit of Eq. (11.13). However, for large signal and noise photons of both \( \bar{n}_S \) and \( \bar{n}_N \), the difference between quantum and classic limit is small. However, for “quantum” thermal noise limited system, the number of noise photons of \( \bar{n}_N \) is very small (Hall, 1994), usually in the other of
10^{-3} or less. Of course, the quantum thermal noise is not the same as receiver thermal noise.

The IMDD channel considered in previous section is equivalent to the classical limit of a photon-counting channel. The continuous-time photon-counting channel is modeled by information theorists as a Poisson channel with unlimited bandwidth but peak and average power constraints (Davis, 1980, Kabanov, 1978, Massey, 1981, Wyner, 1988). A discrete-time Poisson channels can be used to model a quantum-limited, band-limited photon-counting channel (Caves and Drummond, 1994, Gordon, 1962, Hall, 1994, Stern, 1960, Yamamoto and Haus, 1986).

Optical amplification alters the photon statistics of amplified light. While an amplified signal has Poisson statistics, amplified spontaneous emission (ASE) noise obeys Bose-Einstein statistics. For a signal having $n_S$ signal photons and an average of $\bar{n}_{\text{ASE}}$ ASE photons, the ASE noise is equivalent to Poisson-distributed light where the mean number of photons has an exponential distribution of

$$p_{n N}(\bar{n}) = \frac{1}{\bar{n}_N} \exp\left(-\frac{n}{\bar{n}_N}\right), \quad \bar{n} \geq 0. \quad (11.56)$$

With $n_S$ of signal photon, the average number of photons has an distribution of $p_{n N}(\bar{n} - n_S)$ with $\bar{n} \geq n_S$. The output photon-number distribution is equal to

$$p_{n S}(n) = \frac{1}{n!} \int_{n_S}^{+\infty} p_{n N}(\bar{n} - n_S)e^{-\bar{n}}\bar{n}^n d\bar{n}$$

$$= \frac{e^{-n_S}}{n!} \int_{0}^{+\infty} p_{n N}(\bar{n})e^{-\bar{n}}(\bar{n} + n_S)^n d\bar{n}, \quad (11.57)$$

or

$$p_{n S}(n) = \frac{\bar{n}^n}{(1 + \bar{n}_N)^{n+1}} \exp\left(-\frac{n_S}{1 + \bar{n}_N}\right) L_n\left(-\frac{n_S/\bar{n}_N}{1 + \bar{n}_N}\right), \quad (11.58)$$

where $L_n(\cdot)$ is the Laguerre polynomial of

$$L_n(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{(-x)^k}{k!}. \quad (11.59)$$

The distribution of Eq. (11.58) includes only the ASE photons from the same polarization of the signal.

If the input distribution is $p_S(n_S)$, the output photon-number distribution is $p_{N}(n) = \int_{0}^{+\infty} p_S(n_S)p_{n S}(n)dn_S$. The maximum spectral efficiency is

$$C = \max_{p_S(n_S)} \left\{ H(N) - H(N|S) \right\}, \quad (11.60)$$
where
\[ H(N) = -\sum_{n=0}^{\infty} p_N(n) \log p_N(n), \]  
(11.61)
and
\[ H(N|S) = -\int_{0}^{\infty} p_S(n_S) \sum_{n=0}^{\infty} p_{n_S}(n) \log p_{n_S}(n) dn_S. \]  
(11.62)

As an approximation, the maximum spectral efficiency can be derived by maximizing the output entropy of \( H(N) \). Previous section already shows that the maximizing of the output entropy of \( H(N) \) does not necessarily give the channel capacity. However, the calculation here gives a lower bound of the channel capacity. If the input photon is exponential distributed with \( p_S(n_S) = \exp(-n_S/\bar{n}_S)/\bar{n}_S \), the output photon distribution is \( p_N(n) = n^n/(1+n)^{n+1} \) and \( H(N) \) is
\[ H(N) = \log(1 + \bar{n}) + \bar{n} \log(1 + 1/\bar{n}), \]  
(11.63)
where \( \bar{n} = \bar{n}_S + \bar{n}_N \). With an input exponential distribution of \( p_S(n_S) \) and the conditional p.d.f. of Eq. (11.58), the conditional entropy of Eq. (11.62) can be calculated numerically.

For a large number of photons and high SNR, the summation of \( \sum_{n=0}^{\infty} p_{n_S}(n) \log p_{n_S}(n) \) can be approximated by integration. If the conditional probability of \( p_{n_S}(n) \) is assumed to be Gaussian distributed with variance of \( \sigma^2_{n_S}(n_S) = n_S + \bar{n}_N + 2n_S\bar{n}_N + \bar{n}_N^2 \) for the signal shot noise, ASE shot noise, signal-spontaneous beat noise, and spontaneous-spontaneous beat noise, respectively (Desurvire, 1994). Based on the Gaussian approximation, we obtain
\[ \sum_{n=0}^{\infty} p_{n_S}(n) \log p_{n_S}(n) \approx -\frac{1}{2} \log \left[ 2\pi e \sigma^2_{n_S}(n_S) \right]. \]  
(11.64)

Using Eq. (11.64) to calculate \( H(N|S) \), the asymptotic limit is
\[ C \approx \frac{1}{2} \log \left( \frac{n_S e^{1+\gamma_o}}{n_N \pi} \right) - 1 \]  
\[ = \frac{1}{2} \log \frac{n_S}{n_N} - 0.688 \log 2, \]  
(11.65)
the same as that of Eq. (11.32).

Figures 11.6 present spectral efficiency limits given by Eq. (11.60) for a photon-counting channel with ASE noise. Figure 11.6(a) shows the spectral efficiency as a function of the classical SNR \( \rho_s \) for various values
Figure 11.6. The quantum-limited maximum spectral efficiency of photon-counting optical channel as a function of (a) the SNR of $\bar{n}_S/\bar{n}_N$ (b) the average number of signal photons $\bar{n}_S$. 
of the mean number of signal photons $\bar{n}_S$. Figure 11.6(a) also shows the asymptotic limit of Eq. (11.65) for large number of signal photons and high SNR. Figures 11.6(a) and (b) show the spectral efficiency of Eq. (11.60) as a function of the mean signal photon number $\bar{n}_S$ for various values of the SNR $\rho_s$. The case with infinite SNR corresponds to the spectral efficiency for Poisson-distributed photons (Gordon, 1962, Stern, 1960).

Unlike the classical case in Fig. 11.2 in which the spectral efficiency depends only on the SNR, the quantum-limited spectral efficiency depends on both the SNR and the number of signal photons. Even at high SNR, high spectral efficiency cannot be achieved with a small number of signal photons. As shown in Desurvire (2002, 2003), the effect of fiber nonlinearity upon quantum-limited spectral efficiency is equivalent to an increase in the number of ASE photons.

4. Channel Capacity in Nonlinear Regime

Fiber nonlinearities limit the transmission distance and the overall capacity of a WDM system. The major fiber nonlinearities are the Kerr effect, stimulated Raman scattering, and stimulated Brillouin scattering (Agrawal, 2001). The Kerr effect leads to self- and cross-phase modulation, and four-wave mixing. In the Kerr effect, from Eq. (5.1), the intensity of the aggregated optical signal perturbs the fiber refractive index, thereby modulating the signal phase. In WDM systems, self- and cross-phase modulations arise when the phase of a channel is modulated by its own intensity and by the intensity of other channels, respectively. Four-wave-mixing arises when two channels beat with each other, causing intensity modulation at the different frequency, thereby phase-modulating all the channels and generating new frequency components. Previous chapter considers both self- and cross-phase modulation but not four-wave-mixing. The new frequency components from four-wave mixing can be considered as additive noise in additional to amplifier noise.

Fiber propagation with the Kerr effect is modeled using the nonlinear Schrödinger equation of Eq. (7.9) for single channel systems and coupled nonlinear Schrödinger equations for WDM systems. Among the various nonlinearities, the Kerr effect has the greatest impact on a WDM system and thus its channel capacity. Early studies focused on the effect of fiber nonlinearity on specific modulation and detection techniques, including on-off keying with direct detection (Chraplyvy, 1990, Chraplyvy and Tkach, 1993, Forghieri et al., 1997, Wu and Way, 2004) or simple modulations with coherent detection (Shibata et al., 1990, Waarts
Recently, the combined effect of amplifier noise and Kerr nonlinearity on the Shannon capacity has been studied. Mitra and Stark (2001) argued that the capacity of WDM systems is fundamentally limited mostly by cross-phase modulation. As a signal propagates, chromatic dispersion converts cross-phase-modulation-induced phase modulation to intensity noise. Capacity limitations caused by cross-phase modulation are further studied in Green et al. (2002), Stark et al. (2001), and Wegener et al. (2004). In fibers with nonzero dispersion, cross-phase modulation has a much greater impact than four-wave-mixing. WDM systems with many channels are likely to be limited by cross-phase modulation, perhaps allowing self-phase modulation to be ignored to first order for systems with many channels. However, the methods of Mitra and Stark (2001), Stark et al. (2001), and Green et al. (2002) do not quantify the importance of self- relative to cross-phase modulation, and cannot be applied to single-channel systems limited primarily by self-phase modulation.

With constant-intensity modulation, such as phase or frequency modulation (Ho and Kahn, 2002), ideally, both self- and cross-phase modulation cause only time-invariant phase shifts, eliminating both phase and intensity distortion. If one could further neglect four-wave-mixing, propagation would be linear; the capacity would be given by the expressions of previous section, and increasing the launched power would lead to a monotonic increase in spectral efficiency. In reality, chromatic dispersion converts phase or frequency modulation to intensity modulation (Norimatsu and Iwashita, 1993, Wang and Petermann, 1992), and laser intensity noise and imperfect modulation cause additional intensity fluctuations. Hence, it is difficult to maintain constant intensity along an optical fiber. Furthermore, in reality, constant-intensity modulation is subject to four-wave-mixing. As shown earlier, constant-intensity modulation is also fundamentally limited by nonlinear phase noise.

Tang (2001a,b, 2002) solved the nonlinear Schrödinger equation using a series expansion, similar to the Volterra series in Peddanarappagari and Brandt-Pearce (1997, 1998). The linear term is considered to be signal and all higher-order terms are considered to be noise. If sufficient number of terms is included, methods based on series expansion are very accurate. Tang (2001a,b, 2002) has included many terms, yielding a quasi-exact closed-form treatment.

In a single-channel system, the channel capacity is limited by self-phase modulation. In a WDM system, when all channels are detected together, the impact of cross-phase modulation can be reduced using a multi-user detection or interference-cancellation scheme. Using perturbation methods, Narimanov and Mitra (2002) found the channel capacity
of single-channel systems. For a single-channel system with zero average dispersion, Turitsyn et al. (2003) solves the nonlinear Schrödinger equation analytically to find the channel capacity. Basically, Turitsyn et al. (2003) finds the p.d.f. of signal with nonlinear phase noise of Eq. (6.66) and uses it to find the asymptotic channel capacity of the channel. With large nonlinear phase noise, IMDD signal without the usage of phase information can be used instead.

To quantify the SNR in the presence of cross-phase modulation, Mitra and Stark (2001) introduced a nonlinear intensity scale $I_0$. For transmitted power per channel well below $I_0$, increasing the power raises the SNR, increasing capacity. As the transmitted power approaches $I_0$, cross-phase modulation noise increases rapidly, causing capacity to decrease precipitously. This nonlinear intensity scale $I_0$ is also applicable to the models of Tang (2001a,b, 2002), Narimanov and Mitra (2002), constant-intensity modulation of Ho and Kahn (2002), and even the nonlinear quantum limit of Dosurvire (2002, 2003). In each of those models, the launched power that maximizes the channel capacity increases with $I_0$. In WDM systems, the nonlinear intensity scale $I_0$, and thus the capacity, increases with fiber dispersion, channel spacing and signal bandwidth, and decreases with the total number of spans and the total number of channels. In Mitra and Stark (2001), the nonlinear intensity scale for cross-phase modulation was found to be

$$I_0 = \left[ \frac{BD\Delta\lambda}{2\gamma^2 \log(M)L_{\text{eff}}} \right]^{1/2}$$

(11.66)

and the maximum spectral efficiency is lowered bound by

$$C = \log\left[ 1 + \frac{e^{-(I_t/I_0)^2}I_t}{I_n + (1 - e^{-(I_t/I_0)^2})I_t} \right]$$

(11.67)

where $B$ is the number of symbol per second, $D$ is the dispersion coefficient and $\Delta\lambda$ is the channel spacing of the WDM system, $2M + 1$ is the overall number of channels, $\gamma$ is the fiber nonlinear coefficient, $I_t$ and $I_n$ are the signal and noise power per channel, respectively. For a system with $N_A$ spans, the overall effective nonlinear length is approximately equal to $L_{\text{eff}} = N_A/\alpha$ where $\alpha$ is the fiber attenuation coefficient.

Using the spectral efficiency lower bound of Eq. (11.67), the power per channel that maximizes spectral efficiency is approximately equal to $(I_0^2I_n/2)^{1/3}$, and the maximum spectral efficiency is approximately equal to

$$\frac{2}{3} \log \left( \frac{2I_0}{I_n} \right)$$

(11.68)
Figure 11.7. The maximum spectral efficiency of optical channel in nonlinear regime for both unconstrained or constant-intensity signal. The unconstrained signal is limited by cross-phase modulation (XPM) but constant-intensity signal is limited by four-wave-mixing (FWM).

In a WDM system limited by four-wave-mixing instead of cross-phase modulation, the spectral efficiency bound is also given by Eq. (11.67), and the nonlinear intensity scale is given by:

\[
I_0^{-1} = \frac{N_A \gamma^2}{9} \sum_{p,q,p \neq 0,q \neq 0 \mid |p| + |q| \leq M} \frac{D_{pq}^2}{\alpha^2 + \Delta k_{pq}^2},
\]

(11.69)

where \(-M \leq p, q \leq M\), \(D_{pq} = 3\) if \(p = q\) and \(D_{pq} = 6\) if \(p \neq q\), and \(\Delta k_{pq} = 2\pi \lambda^2 D \Delta f^2 q/c\), \(\lambda\) is the optical wavelength, and \(c\) is the speed of light. The expression Eq. (11.69) has been derived for the center (worst-case) channel. The additive noise from four-wave-mixing was studied by Eiselt (1999), Tkach et al. (1995), and Forghieri et al. (1997).

Figure 11.7 shows the spectral efficiency as a function of input power density, including the Shannon limit of Eq. (11.13), the numerical expression of Eq. (11.17) for constant-intensity signal, and the results of Mitra and Stark (2001) limited by cross-phase modulation. In the absence of four-wave-mixing, as input power is increased, spectral efficiency increases monotonically for the Shannon limit and for constant-intensity modulation, but the spectral efficiency computed following Mitra and Stark (2001) reaches a maximum value limited by cross-
phase modulation. When four-wave-mixing is modeled as extra additive Gaussian noise, spectral efficiencies for the Shannon limit and for constant-intensity modulation reach maximum values limited by four-wave-mixing, while the spectral efficiency computed following Mitra and Stark (2001) remains unchanged. The maximum spectral efficiency of constant-intensity signal is about 2.8 bit/s/Hz, compared with 2.3 bit/s/Hz computed following Mitra and Stark (2001).

The system of Fig. 11.7 has \(2M + 1 = 101\) WDM channels and \(N_A = 10\) fiber spans; uses optical fiber having attenuation coefficient \(\alpha = 0.2\) dB/km, nonlinear coefficient of \(\gamma = 1.24\) rad/W/km, and dispersion coefficient \(D = 17\) ps/km/nm; operates around the wavelength of \(\lambda = 1.55\) \(\mu\)m with channel bandwidth \(B = 40\) GHz, and channel separation \(\Delta f = 1.5B\); uses optical amplifiers with noise figure of 4 dB and gain of 30 dB. Using an overall effective length of \(L_{\text{eff}} = N_A/\alpha\), the nonlinear intensity scale of Eq. (11.66) is \(I_0 = 11.2\) mW. In Fig. 11.7, we assume that all four-wave-mixing components from the same span and from each fiber span combine incoherently by ignoring the phase dependence between four-wave-mixing components.

Limited by four-wave-mixing, constant-intensity modulation may provide better spectral efficiencies than those of Mitra and Stark (2001) in the regime in which cross-phase modulation dominates over four-wave-mixing. Because four-wave-mixing decreases more rapidly than cross-phase modulation as channel spacing is increased, the improvement obtained using constant-intensity modulation is more significant for systems having large channel spacing. Both four-wave-mixing and cross-phase modulation decrease with an increase of fiber dispersion. Four-wave-mixing dominates over cross-phase modulation for zero-dispersion optical fiber. Of course, although the system for Fig. 11.7 shows that constant-intensity modulation has better maximum spectral efficiency than unconstrained modulation. Depending on system parameters, unconstrained modulation may have maximum spectral efficiency better than constant-intensity modulation (Kahn and Ho, 2004).

While the nonlinear Schrödinger equation with noise provides a very accurate model for nonlinear propagation in fiber, the equation does not have an analytical solution except in some special cases (Turitsyn et al., 2003). While all of the works are based on this accurate formulation, they make different assumptions and approximations, leading to different estimation of the channel capacity.

In order to illustrate the major qualitative differences between the various models, we consider a simplified memoryless monotonic transfer characteristic of \(y = f(x) = x + \epsilon x^3\), where \(x\) and \(y\) are the input and output, respectively, and \(\epsilon\) is a small number. While there is no nonlinear
fiber channel, or other channel type, having transfer characteristic of $f(x)$, this simple function with linear term $x$ and nonlinear term of $cx^3$ yields insight into fiber systems. For a monotonic, one-to-one function such as $f(x)$, if we interpret both terms $x$ and $cx^3$ as signal, then in the absence of any noise, the entropy of the output given the input, $H(Y|X)$, is equal to zero. The mutual information between the input and the output, and thus the channel capacity, equals the entropy of the input $x$.

We draw an equivalence to the most models by considering the linear term $x$ to be signal and the nonlinear term $cx^3$ to be noise. In WDM systems with many channels, the nonlinear term $cx^3$ includes “intermodulation products” corresponding to the cross-phase modulation and four-wave-mixing caused by other channels. As all channels are typically independent from one another, the models concerning cross-phase-modulation-induced distortion can indeed model cross-phase modulation as noise independent from the signal (Green et al., 2002, Mitra and Stark, 2001, Stark et al., 2001, Wegener et al., 2004). Likewise, the model concerning four-wave-mixing components from other channels for constant-intensity modulation can model four-wave-mixing to be noise independent from the signal (Ho and Kahn, 2002). In a single-channel system (Turitsyn et al., 2003), the nonlinear distortion caused by self-phase modulation depends on the signal and cannot be modeled as signal-independent noise. While only contributions from cross-phase modulation are modeled as signal-independent noise in Green et al. (2002), Mitra and Stark (2001), Stark et al. (2001), and Wegener et al. (2004), the series expansion model of Tang (2001a,b, 2002) regards all higher-order terms as noise independent of the signal. In fact, the two main groups of models are not complete because Tang (2001a,b, 2002) cannot account for the dependence of higher-order terms on the signal and Green et al. (2002), Mitra and Stark (2001), Stark et al. (2001), and Wegener et al. (2004) cannot include the higher-order terms caused by self-phase modulation. In WDM systems with many channels, the methods of Green et al. (2002), Mitra and Stark (2001), Stark et al. (2001), Tang (2001a,b), and Wegener et al. (2004) can be considered to be equivalent if the effect of self-phase modulation is negligible compared to cross-phase modulation. In single-channel systems, where self-phase modulation must be considered, only the method of Turitsyn et al. (2003) yields the probability density of the channel output including nonlinearity and uses it to calculate the channel capacity. In all cases, if the nonlinear term of $cx^3$ is considered as noise, the channel capacity decreases at high launched power and a nonlinear intensity scale similar to Eqs. (11.66) and (11.69) can be evaluated. In the single-channel system
of Narimanov and Mitra (2002), the channel capacity curve behaves like the curves in Fig. 4 with four-wave mixing.

The single-channel capacity of Turitsyn et al. (2003) has been evaluated for fiber links with zero average dispersion. Only nonlinear phase noise of Chapter 5 modulates the signal phase. The channel capacity is calculated using the probability density of the signal with nonlinear phase noise of Eq. (6.66). In the limit of very high nonlinear phase noise, the capacity degenerates to that of direct detection in the last section, which increases logarithmically with launched power. The nonlinear phase noise causes no amplitude noise.

The impact of Kerr nonlinearity can be reduced or canceled using phase conjugation (Brener et al., 2000, Pepper and Yariv, 1980). In WDM systems with many channels, Kerr effect compensation reduces or cancels the nonlinear terms originating from other channels. In such a case, Kerr effect compensation yields the obvious benefit of reducing the “noise”. In single-channel systems, Kerr effect compensation changes the statistics of the signal with noise. While mid-span or distributed Kerr effect compensation can improve the capacity, Kerr effect compensation just before the receiver does not improve capacity, and may actually reduce capacity by adding more noise.

5. Summary

Increasing spectral efficiency is often the most economical means to increase WDM system capacity. In this chapter, we find the information-theoretical spectral efficiency limits for various modulation and detection techniques in both classical and quantum regimes, considering both linear and nonlinear fiber propagation regimes. Spectral efficiency limits for unconstrained modulation with coherent detection are several b/s/Hz in terrestrial WDM systems, even considering nonlinear effects. Spectral efficiency limits are reduced significantly using either constant-intensity modulation or direct detection. Using binary modulation, regardless of detection technique, spectral efficiency cannot exceed 1 b/s/Hz per polarization.

Optical signals propagating in fibers offer several degrees of freedom, including time, frequency and polarization. The combined coding over these degrees of freedom has been seldom explored as a means to increase transmission capacity in fibers, especially as a way to combat or benefit from fiber nonlinearity and polarization-mode dispersion.

Both fiber chromatic dispersion and polarization-mode dispersion does not limit the maximum spectral efficiency of the optical channel. In the ultimate limit, the bandwidth per channel can be very small to reduce the impacts of both chromatic and polarization-mode dispersion. The
spectral efficiency can also be doubled using polarization-division multiplexing (PDM). For fiber with polarization-mode dispersion, the two orthogonal polarized signals can propagate along the two principle states of polarization. Of course, the transmitter requires active tracking such that the signal can follow the two principle states of polarization for a time varying fiber channel.