0122229 现代数字通信与编码理论 December 16, 2008
XDU, Winter 2008 Lecture Notes

Chapter 7 Coding and Modulation for Fading Channels

The research on wireless communications has received many attentions. The underlying physical channels may be characterized as time-varying fading channels. In this chapter, we will study communications over fading channels from an information-theoretic point of view. Many state-of-the-art coding and modulation schemes, including space-time codes, will be introduced.

7.1 Wireless Channel Models

在许多无线通信中，由于电波的反射、散射和绕射等，使得发射机和接收机之间存在多条传播路径，并且每条路径的传播时延和衰耗因子都是时变的；在接收端，从不同路径到达的信号分量叠加在一起，导致接收信号的强度随时间变化，我们通常用“衰落”来描述这种现象。多径时变衰落是移动通信信道的主要特点。Many channel models arise according to different fading characteristics. Fig. 7.1 depicts an overview of fading channel manifestations.

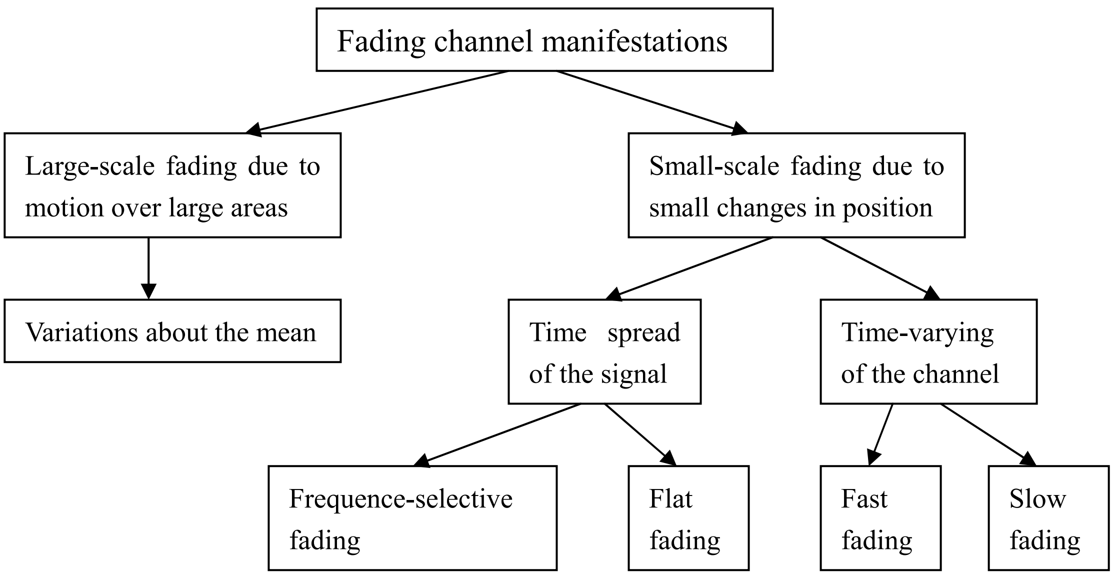


Figure 7.1

7.1.1 Large-Scale Fading: Path loss and Shadowing

Let $s(t) = \mathrm{Re}\left\{x(t)e^{j2\pi f_c t}\right\}$ be the transmitted signal, where $x(t) = |x(t)|e^{j\phi(t)} = R(t)e^{j\phi(t)}$ is the complex baseband equivalent signal. In a fading environment, the magnitude of the received baseband waveform $y(t)$ can be expressed as

$$|y(t)| = \alpha(t)R(t) = m(t)\cdot r_0(t)\cdot R(t),$$

where $m(t)$ is called the *large-scale-fading component* of the envelope (Since variations due to path loss and shadowing occur over relatively large distances), and $r_0(t)$ is called the *small-scale-fading component*. Fig. 7.2 illustrates the signal power received versus antenna displacement. Large-scale fading represents the average signal power attenuation due to motion over large areas. It consists of the path loss and the shadowing loss. The path loss is the signal attenuation due to the fact that the power received by an antenna at distance d from the transmitter decreases as d increases. Empirically, the power attenuation is proportional to $d^2 \sim d^4$. The shadowing loss is due to the absorption of the radiated signal by scattering structures. Furthermore, $r_0(t)$ is sometimes referred to as multipath (or Rayleigh) fading.

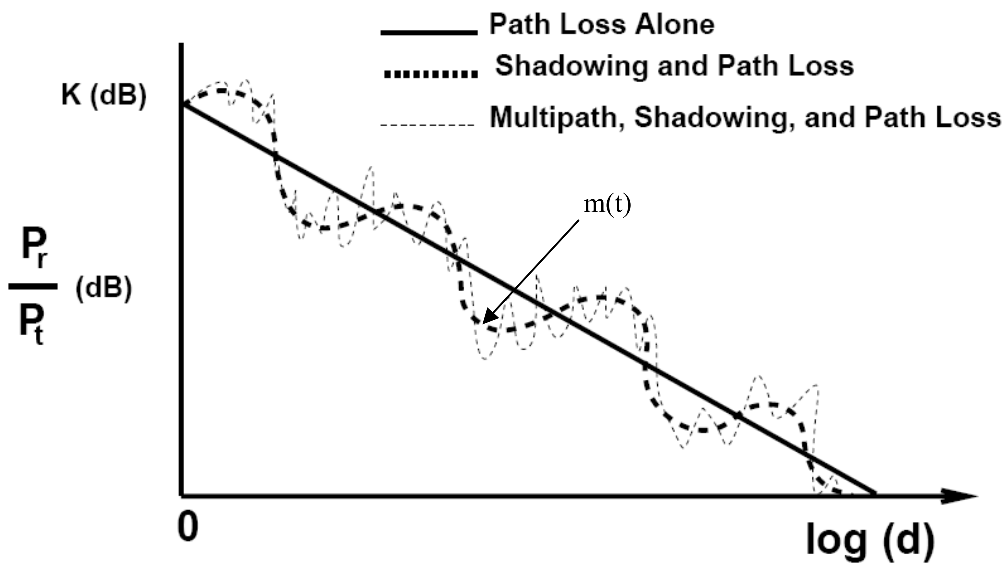


Figure 7.2 Path Loss, shadowing and multipath changes with distance.

■ Log-normal shadowing

In the log-normal atttenuation model, the ratio of transmit-to-receive signal power $\xi = P_t / P_r$ based on path loss and shadowing is assumed random with a log-normal distribution given by

$$p(\xi) = \frac{10/\ln 10}{\sqrt{2\pi}\sigma_{\xi_{dB}}} \exp\left\{-\frac{\left(10\log_{10}\xi - \mu_{\xi_{dB}}\right)^2}{2\sigma_{\xi_{dB}}^2}\right\}, \xi > 0 \qquad (7.1)$$

where $\mu_{\xi_{dB}}$ is the mean of $\xi_{dB} = 10\log_{10}\xi$ in dB and $\sigma_{\xi_{dB}}$ is the standard deviation of ξ_{dB} in dB. Note that if ξ is log-normal then the received power and received SNR will also be log-normal. The mean of ξ (the linear average path gain) can be obtained as

$$\mu_\xi = \mathrm{E}[\xi] = \exp\left\{\frac{\mu_{\xi_{dB}}}{\psi} + \frac{\sigma_{\xi_{dB}}^2}{2\psi^2}\right\}$$

where $\psi = 10/\ln 10$. The conversion from the above linear mean to the log mean (in dB) is given by

$$10\log_{10}\mu_{\xi} = \mu_{\xi_{\mathrm{dB}}} + \frac{\sigma^2_{\xi_{\mathrm{dB}}}}{2\psi}$$

With a change of variables we see that the distribution of the dB value of ξ is Gaussian with mean $\mu_{\xi_{dB}}$ and variance $\sigma^2_{\xi_{dB}}$:

$$p(\xi_{\mathrm{dB}}) = \frac{1}{\sqrt{2\pi}\sigma_{\xi_{dB}}}\exp\left\{-\frac{\left(\xi_{\mathrm{dB}} - \mu_{\xi_{\mathrm{dB}}}\right)^2}{2\sigma^2_{\xi_{\mathrm{dB}}}}\right\}$$

- Combined path loss and shadowing: In cellular systems, a generally accepted model is that the attenuation between the subscriber and the cell site is proportional to $10^{(\xi/10)}d^{-4}$, where d is distance from subscriber to cell site and ξ is a Gaussian random variable with zero mean and standard deviation of 8dB.

7.1.2 Small-Scale Fading: Multipath channel

- Time-Varying Channel Impulse Response for the Multipath fading channel
  Let the transmitted signal be

$$s(t) = \mathrm{Re}\left\{x(t)e^{j2\pi f_c t}\right\}$$

Assume that the antenna is traveling, and there are multiple scattered paths from the transmitter to the receiver, each associated with a time-varying propagation delay $\tau_n(t)$ and a time-varying multiplicative factor $\alpha_n(t)$. Neglecting noise, the received passband signal can be written as

$$r(t) = \sum_{n=0}^{N(t)} \alpha_n(t)s\left[t-\tau_n(t)\right] = \mathrm{Re}\left\{\sum_{n=0}^{N(t)} \alpha_n(t)x\left[t-\tau_n(t)\right]e^{j2\pi f_c[t-\tau_n(t)]}\right\} \tag{7.2}$$

where $\alpha_n(t)$ is the attenuation obtained from the path loss and shadowing model. Denote the Doppler shift at f_c on path n by $f_{D,n}(t) = f_D\cos\theta_n(t)$, where $f_D = v/\lambda_c = f_c\frac{v}{c}$ with v being the mobile velocity. In a simple case (under the assumption that each path delay is changing at a constant rate), we have $\tau_n(t) = \overline{\tau}_n - \frac{f_{D,n}(t)}{f_c}t$. Since all of these parameters change over time, they are characterized as random processes which we assume to be both stationary (over the short term) and ergodic. Thus, the received signal is also a stationary and ergodic random process.

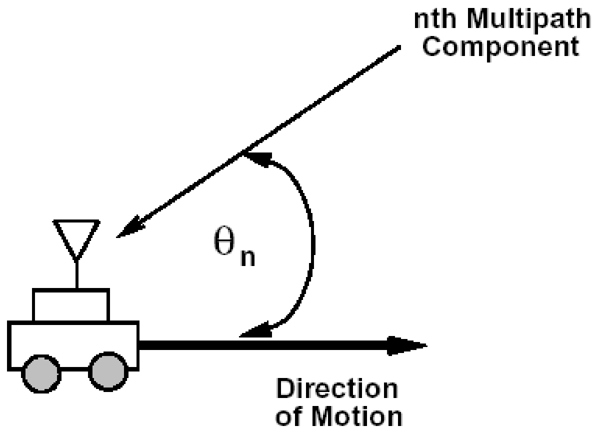


Fig. 7.4

Equation (7.2) can be rewritten as

$$r(t) = \mathrm{Re}\left\{ \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j2\pi f_c \tau_n(t)} x\left(t - \tau_n(t)\right) e^{j2\pi f_c t} \right\}$$

Since the channel may be subject to Doppler shifts, the recovered carrier, $\hat{f}_c$, at the receiver, might be different from the actual carrier f_c. Thus, the demodulated baseband signal is given by

$$y(t) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j2\pi f_c \tau_n(t)} x\left(t - \tau_n(t)\right) e^{j2\pi (f_c - \hat{f}_c) t}$$

By letting $\phi_n(t) = 2\pi f_c \tau_n(t) + 2\pi\left(\hat{f}_c - f_c\right)t$, we can then rewrite the complex baseband equivalent signal as

$$y(t) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} x\left(t - \tau_n(t)\right) \tag{7.3}$$

Since $\alpha_n(t)$ depends on attenuation while $\phi_n(t)$ depends on delay and Doppler shift, we assume that these two random processes are independent. Equ. (7.3) 描述了一个线性时变系统的输入-输出模型。

From (7.3), the equivalent lowpass time-varying impulse response (at time t to an impulse transmitted at time t-τ) is given by

$$h(\tau, t) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} \delta\left(\tau - \tau_n(t)\right) = \sum_{n=0}^{N(t)} \beta_n(t) \delta\left(\tau - \tau_n(t)\right) \tag{7.4}$$

with $\beta_n(t) = \alpha_n(t) e^{-j\phi_n(t)}$. This expression is quite nice. It says that the effect of mobile users, arbitrarily moving reflectors and absorbers, and all of the complexities of solving Maxwell's equations, finally reduce to an input/output relation between transmit and receive antennas which is simply represented as the impulse response of a linear time-varying channel filter.

For the time-varying impulse response $h(\tau, t)$, we can define a time-varying frequency response

$$H(f;t)=\int_{-\infty}^{\infty} h(\tau,t)e^{-j2\pi f\tau}d\tau=\sum_{n=0}^{N(t)}\beta_n(t)e^{-j2\pi f\tau_n(t)}$$

One way of interpreting $H(f;\,t)$ is to think of the system as a slowly varying function of t with a frequency response $H(f;\,t)$ at each fixed time t. Corresponding, $h(\tau,\,t)$ can be thought of as the impulse response of the system at a fixed time t. This is a legitimate and useful way of thinking about multipath fading channels, as the time-scale at which the channel varies is typically much longer than the delay spread of the impulse response at a fixed time.

7.1.3 Time and Frequency Coherence

Wireless channels change both in time and frequency. The time coherence shows us how quickly the channel changes in time, and similarly, the frequency coherence shows how quickly it changes in frequency.

7.1.3.1 Delay spread and Coherence bandwidth

- Multipath delay spread

The *multipath delay spread* of a channel $h(\tau,\,t)$ is given by

$$T_m=\max_n \tau_n(t)-\min_n \tau_n(t) \tag{7.5}$$

For a single transmitted impulse, it represents the *maximum excess delay*, after which the multipath signal power falls below some threshold level relative to the strongest component. The minimum delay $\min_n \tau_n(t)$ is often set to zero and the other multipath delays normalized with respect to this minimum delay.

Another often used parameter to characterize the delay spread of the channel is the root-mean-squared (rms) delay spread, defined as

$$\sigma_\tau=\sqrt{\overline{\tau^2}-(\overline{\tau})^2}$$

where $\overline{\tau}$ is the mean excess delay.

- Multipath Intensity Profile

The multipath intensity profile, also called the *power delay profile*, is defined as the autocorrelation

$$E\left[h(\tau_1,t)h^*(\tau_2,t)\right]=R_h(\tau_1)\delta(\tau_1-\tau_2)\triangleq R_h(\tau)$$

where $\tau=\tau_1=\tau_2$. It represents the average power associated with a given multipath delay.

We can also characterize the time-varying multipath channel in the frequency domain. The autocorrelation of () is given by

$$E\left[H(f_1,t)H^*(f_2,t+\Delta t)\right]=R_H(\Delta f,\Delta t)$$

where $\Delta f = f_2 - f_1$. If we define $R_H(\Delta f)\triangleq R_H(\Delta f,0)$, then it can be shown that

$$R_H(\Delta f) = F\left[R_h(\tau)\right] = \int_{-\infty}^{\infty} R_h(\tau) e^{-j2\pi\Delta f \tau} d\tau$$

The frequency bandwidth B_c where $R_H(\Delta f) \approx 0$ for all $\Delta f > B_c$ is called the *coherence bandwidth* of the channel. This bandwidth measures the frequency range over which the fading process is correlated (and is defined as the frequency bandwidth over which the correlation function of two samples of the channel response taken at the same time but different frequencies falls below a suitable value). Typically, the coherence bandwidth is related to the multipath delay spread T_m by

$$B_c = \frac{1}{T_m}$$

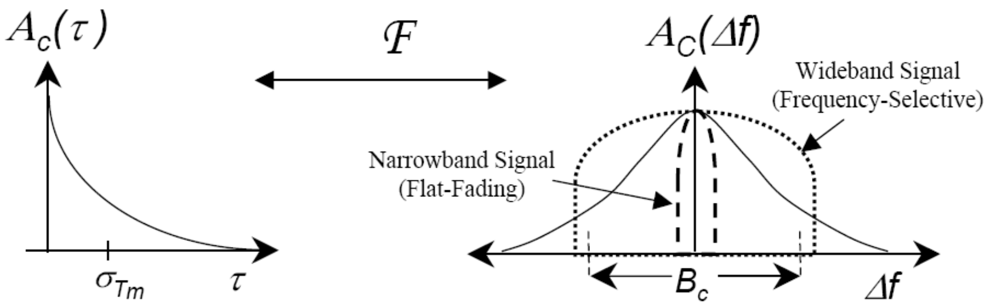


Figure 7.5. Multipath intensity profile, delay spread, and coherence bandwidth.

A more popular approximation of B_c, corresponding to a bandwidth interval having a correlation of at least 0.5, is

$$B_c \approx \frac{1}{5\sigma_\tau}$$

■ Frequency-flat and frequency-selective fading

In general, if we are transmitting a narrowband signal with bandwidth $W \ll B_c$, then fading across the entire signal bandwidth is highly correlated, and all the spectral components of the transmitted signal are affected in a similar manner. In this case, the fading is usually said to be *frequency-nonselective* or equivalently *frequency-flat*. On the other hand, if the spectral components of the transmitted signal are affected by different amplitude gains and phase shifts, the fading is said to be *frequency-selective*. This applies to *wideband* systems in which the transmitted bandwidth is bigger than the channel's coherence bandwidth.

■ A Discrete-Time Input/Output Model for Wideband Channels

由采样定理，带宽为 $W/2$ 的基带信号 $x(t)$可表示为

$$x(t) = \sum_k x_k \,\mathrm{sinc}(Wt - k)\,, \tag{7.6}$$

where $x_k = x\left(\frac{k}{W}\right)$ and $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$.

Using (7.6), the baseband output is given by

$$y(t) = \sum_k x_k \sum_n \beta_n(t) \operatorname{sinc}\left(Wt - W\tau_n(t) - k\right)$$

By sampling outputs at multiples of $1/W$, we have

$$y_m = \sum_k x_k \sum_n \beta_n(m/W) \operatorname{sinc}\left(m - k - W\tau_n(m/W)\right) \tag{7.7}$$

Let $l = m - k$. Then

$$y_m = \sum_l x_{m-l} \sum_n \beta_n(m/W) \operatorname{sinc}\left(l - W\tau_n(m/W)\right) \tag{7.8}$$

By defining

$$h_{l,m} = \sum_n \beta_n(m/W) \operatorname{sinc}\left(l - W\tau_n(m/W)\right) \tag{7.9}$$

(7.8) can be written as

$$y_m = \sum_l h_{l,m} x_{m-l} \tag{7.10}$$

We refer to $h_{l,m}$ as the lth (complex) channel filter tap at time m. This discrete-time model is illustrated in Fig. 7.6. If the lth tap is unchanging with m for each l, then the channel is linear time-invariant. If each tap changes slowly with m, then the channel is called *slowly time-varying*.

In practice, other transmit pulses, such as the raised cosing pulse, are often used in place of the sinc pulse. This necessitate sampling above the Nyquist sampling rate, but does not alter the essential nature of the model.

直观上，式(7.9)可解释如下：

Discrete the multipath delay axis τ of the impulse response into equal time delay segments called *excess delay bins*, where each bin has a time delay width equal to $\tau_{i+1} - \tau_i = 1/W$, where $\tau_0 = 0$ represents the first arriving signal at the receiver. Any number of multipath signals received within the ith bin are represented by (combined into) a single resolvable multipath component with delay τ_i. This single multipath component has an amplitude and phase corresponding to the sum of those non-resolvable components.

In general, wideband channels have resolvable multipath components which are represented by taps h_l's, whereas narrow-band channels tend to have non-resolvable multipath components, corresponding to a single tap.

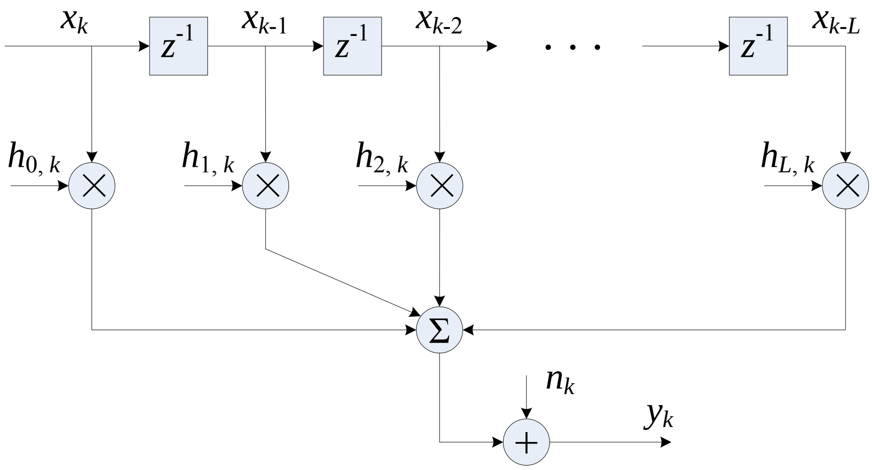


Fig. 7.6 The discrete tapped delay line channel model

7.1.3.2 *Doppler Spread And Coherence Time (Slow and Fast fading)*

Another important channel parameter is the time-scale of the variation of the channel. How fast do the taps $h_{l,m}$ vary as a function of time m? The distinction between slow and fast fading is important for the mathematical modeling of fading channels. This notion is related to the *coherence time* T_c of the channel, which measures the period of time over which the fading process is correlated (or equivalently, the period of time after which the correlation function of two samples of the channel response taken at the same frequency but different time instants drops below a certain predetermined threshold). The coherence time is also related to the channel *Doppler spread* f_d by

$$T_c = \frac{1}{B_d}$$

(Doppler spread is the largest difference between the Doppler shifts,

$B_d = f_{D,2} - f_{D,1} = 2 f_c v / c$.)

The fading is said to be slow if the symbol time duration T_s is smaller than the channel's coherence time T_c; otherwise it is considered to be fast. In slow fading a particular fade level will affect many successive symbols, which leads to burst errors, whereas in fast fading the fading decorrelates from symbol to symbol.

Note: A often used parameter in simulation is Fading rate (normalized Doppler spread) defined by $B_d T_s = B_d / W$.

■ Tap gain autocorrelation function and power spectral density

Modeling each $h_{l,m}$ as a complex random variable provides part of the statistical description that we need, but this is not the most important part. The more important issue is how these quantities vary with time. A statistical quantity that models this relationship is known as the tap gain autocorrelation function, $R_l[n]$. It is defined as

$$R_l[n] = \mathrm{E}\left\{h_{l,m}^* h_{l,m+n}\right\}$$

When using the dense-scatter channel model, the autocorrelation function has the form as

$$R_l[n] = J_0(n\pi B_d / W)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind. The Doppler power spectral density is given by

对于陆地移动通信信道，一种使用垂直天线的多普勒谱为：

$$S_s(f) = \frac{3}{2\pi f_D \sqrt{1 - \left(\frac{f - f_c}{f_D} \right)^2}}$$

相应的归一化相关函数为 $R[n] = J_0(2\pi n f_m T_s)$，$f_D$ 是最大多普勒频移，J_0 是第一类零阶贝塞尔(Bessel)函数。

7.1.3.3 Summary: 多径衰落信道的分类

一般地，人们根据发送信号的周期 T_s、信号带宽 W=1/T_s，与多径时延扩展 T_m 和多普勒扩展(Doppler spread) B_d 的相互关系，将多径衰落分为如下四类：(信道相干时间 T_c=1/B_d，相干带宽 B_c=1/T_m)

表 7.1 Types of small-scaling fading

	$T_c >> T_s$ (即 $B_d << W$)	$T_c < T_s$ (即 $B_d > W$)
$W << B_c$ (即 $T_s >> T_m$)	平慢衰落 (Loss in SNR)	平快衰落 (PLL failure, irreducible BER)
$W > B_c$ (即 $T_s < T_m$)	频率选择性慢衰落 (ISI distortion、SNR 损失)	频率选择性快衰落 (ISI distortion、PLL 失效)

多普勒扩展（即信道相干时间）反映了信道冲激响应在时域上的相关性，它控制着信道的衰落速率。对于慢衰落信道，通常认为信道增益在整个发送符号周期内是恒定的。多径时延扩展（即相干带宽）描述了信道冲激响应在频域上的相关性，对于平衰落(flat fading)信道，发送信号的各频率分量具有相同的增益，信号波形不失真，无码间串扰。

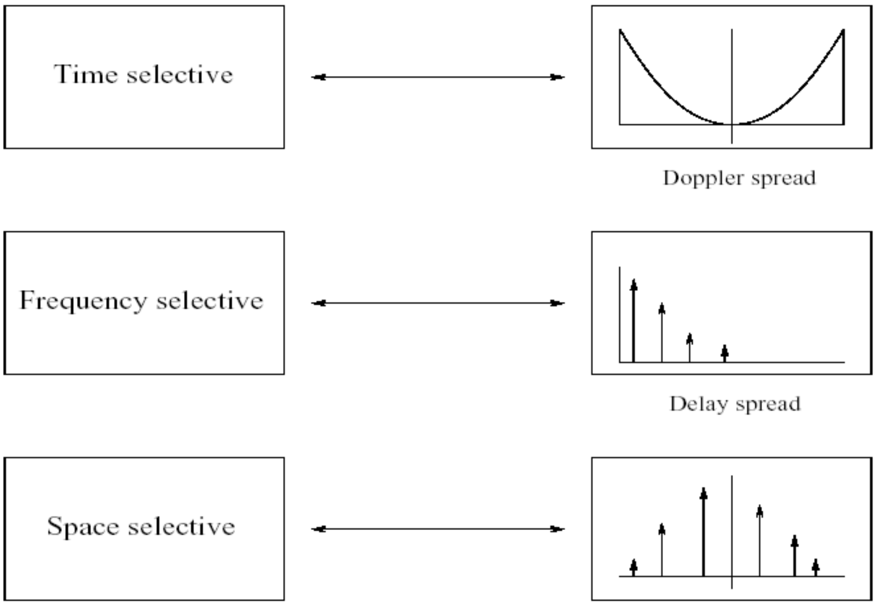


Figure 7.7 Microscopic fading

- Mitigating the Degradation Effects of Fading

To combat distortion

FREQ-SELECTIVE DISTORTION
Adaptive equalization
Spread spectrum (DS or FH)
OFDM

FAST-FADING DISTORTION
Robust modulation (e.g., noncoherent)
Signal redundancy to increase rate
Coding & interleaving

To combat loss in SNR

FLAT-FADING and SLOW FADING
Diversity
Error control codes
Interleaving

7.2 Statistical Models For Fading Channels

From (7.9), it is seen that each channel tap $h_{l,k}$ contains an aggregate of paths, with the delays smoothed out by the baseband signal bandwidth. Fortunately, the filter taps often contain a sufficient path aggregation so that a statistical model might have a chance of success.

- *Rayleigh fading*

The simplest probabilistic model for the channel filter taps is based on the assumption that there are a large number of statistically independent reflected and scattered paths with random amplitudes in the delay window corresponding to a single tap. Since the reflectors are far away relative to the carrier wavelength, it is also reasonable to assume that the phase for each path is uniformly distributed between 0 and 2π and that these phases are independent.

Thus each tap $h_{l,k}$ can be modeled as a sample value of a random variable $H_{l,k}$ which is the sum of a large number of small complex random variables each of uniformly distributed phase. By 根据中心极限定理，it can be easily shown that $H_{l,k}$ can be modeled as circularly symmetric complex Gaussian r.v.'s $\sim \mathcal{CN}(0,\sigma_l^2)$. 这样，$H_{l,k}$的幅度 $R = |H_{l,k}|$ 是服从Rayleigh分布的随机变量：

$$p_R(r) = \frac{r}{\sigma_l^2}\exp\left(-\frac{r^2}{2\sigma_l^2}\right), \quad r \ge 0 \tag{7.11}$$

这个模型描述的信道即为Rayleigh衰落信道。The squared magnitude $|h_{l,k}|^2$ is exponentially distributed with density

$$\frac{1}{\sigma_l^2}\exp\left(-\frac{r}{\sigma_l^2}\right), \quad r \ge 0$$

(this is the power distribution.)

Note that the word Rayleigh is almost universally used for this model, but the assumption is that the tap gains are circularly symmetric complex Gaussian random variables.

- *Rician fading*

When the line of sight (LOS) path (often called a specular path) is large and has a known magnitude, and there are also a large number of independent paths, the $h_{l,k}$, at least for one value of *l*, can be modeled as

$$h_l(m) = \sqrt{\frac{\kappa}{\kappa+1}}\sigma_l e^{j\theta} + \sqrt{\frac{1}{\kappa+1}}\mathcal{CN}\left(0,\sigma_l^2\right) \tag{7.12}$$

with the first term corresponding to the specular path arriving with uniform phase θ and the second term corresponding to the aggregation of the large number of reflected and scattered paths, independent of θ. The parameter k (so-called K-factor) is the ratio of the energy in the specular path to the energy in the scattered paths; the larger k is, the more deterministic is the channel. The magnitude of such a random variable is said to have a Rician distribution. Its density has the form as

$$p_R(r) = \frac{r}{\sigma_l^2}\exp\left(-\frac{r^2+s^2}{2\sigma_l^2}\right)I_0\left(\frac{rs}{\sigma_l^2}\right), \; r \ge 0 \tag{7.13}$$

where $2\sigma^2 = \sum_{n\ne 0} E[\alpha_n^2]$ is the average power in the non-LOS multipath component and $s^2 = \alpha_0^2$ is the power in the LOS component. $I_0(.)$ is the zero order modified Bessel function of the first kind.

The parameter k in (7.12) is then given by

$$K = \frac{s^2}{2\sigma^2}$$

For K=0, we have Rayleigh fading and for K=∞, we have no fading (AWGN channel). This model is often a better one than the Rayleigh model.

7.3 Channel Capacity for Fading Channels

Here we consider the flat fading channel. That is, we assume the bandwidth of the signal is much narrower than the coherence bandwidth of the channel.

- **<u>Channel state information:</u>**

As we will see, a crucial factor in determining the performance of transmission over a fading channel is the availability, at the transmitter side or the receiver side, of channel-state information (CSI), that is, the value taken on by the fading gains in a transmission path. In a fixed wireless environment, the fading gains can be expected to vary slowly, so their estimate can be obtained by the receiver with a reasonable accuracy, even in a system with a large number of antennas, and possibly fed back to the transmitter. In some cases, we may assume that a partial knowledge of the CSI is available. One way of obtaining this estimate is by periodically sending pilot signals (i.e., known training symbols to the receiver) on the same channel used for data (these pilot signals are used in wireless systems also for acquisition, synchronization, etc). Hereafter, we will use CSIT and CSIR to denote the availability of perfect CSI at the transmitter and receiver, respectively.

7.3.1 Information-Theoretic Notation

The Shannon capacity of a single-user time-invariant channel is defined as the maximum mutual information between the channel input and output. This maximum mutual information is shown by Shannon's capacity theorem to be the maximum data rate that can be transmitted over the channel with arbitrarily small error probability. When the channel is time-varying, channel capacity has multiple definitions, depending on what is known about the channel state or its distribution at the transmitter and/or receiver and whether capacity is measured based on averaging the rate over all channel states/distributions or maintaining a constant fixed or minimum rate. We will first introduce two important concepts: Ergodic capacity and outage capacity. Their definitions are connected to the ergodicity of the channel states.

Consider a single-user flat-fading channel. The discrete-time complex baseband equivalent channel, with k standing for the discrete-time index, can be modeled by

$$y_k = h_k x_k + n_k \tag{7.14}$$

where $\{x_k\}$ is the complex transmitted sequence subject to an average-power constraint of E_s joules/symbol, $\{y_k\}$ is the received sequence of complex signal samples. The circularly symmetric i.i.d. Gaussian noise samples are designated by $\{n_k\}$, where $\mathsf{E}[|n_k|^2] = 2\sigma^2 = N_0$.

The samples $\{h_k\}$ is the fading process and $\mathsf{E}\left[|h_k|^2\right] = 1$ is assumed for normalization.

Hence, $\mathsf{SNR} = E_s / N_0$ is the average received SNR.

Assume that the channel gain is fixed, i.e., $h_k = h$ for all k; and h is perfectly known at the receiver. Then conditional on a realization of the channel h, this is an AWGN channel with received SNR $|h|^2 \mathsf{SNR}$. The maximum rate of reliable communication supported by this channel is

$$C(h) = \max_{P_X} I(X;Y) = \log\left(1+|h|^2 \mathsf{SNR}\right) \tag{7.15}$$

When the channel is time-varying, this quantity is a function of the random channel gain h and is therefore random.

Assume $h_k = h^{(l)}$ remains constant over the lth coherence period of T_c symbols and is iid across different coherence periods. This is the so-called block fading model, see Fig.7.8.

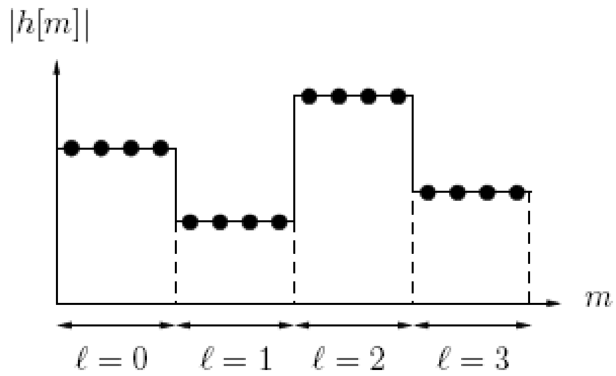


Figure 7.8

Suppose coding is done over L such coherence periods. If T_c >>1, we can effectively model this as L parallel sub-channels that fade independently. For finite L, the quantity

$$\frac{1}{L}\sum_{l=1}^{L}\log\left(1+|h^{(l)}|^2 \mathsf{SNR}\right)$$

is random. As $L \to \infty$, the law of large numbers says that

$$\frac{1}{L}\sum_{l=1}^{L}\log\left(1+|h^{(l)}|^2 \mathsf{SNR}\right) \to \mathsf{E}\left[\log\left(1+|h|^2 \mathsf{SNR}\right)\right]$$

Now we can average over many independent fades of the channel by coding over a large number of coherence time intervals and a reliable rate of communication of $\mathsf{E}\left[\log\left(1+|h|^2 \mathsf{SNR}\right)\right]$ can be achieved.

- *Ergodic capacity:* $C = \mathsf{E}_H[C(h)] = \mathsf{E}_H\left[I(h)\right]$, which is related to the mean behavior of the mutual information. The basic assumption here is that $T_s \gg T_c = 1/B_d$, meaning that the transmission time is so long as to reveal the long-term ergodic properties of the fading process $h(t, \tau)$ which is assumed to be an ergodic process in t. (i.e., the transmitted codewords span many coherence periods, such that the law of large numbers can be applied. A received codeword is affected by all possible fading states.)

In this case, standard capacity results in Shannon's sense are valid and coding theorems are proved by rather standard methods for time-varying and/or finite- (or infinite-) state channels. (The ergodic behavior is the key to generalize the use of the law of large numbers in the proof of the direct part of the memoryless channel coding theorem.) Hence, ergodic capacity is also called *Shannon capacity*.

For the above example,

$$C = \mathsf{E}_h\left[\log\left(1+|h|^2\,\mathsf{SNR}\right)\right] \quad \text{bits/2D} \tag{7.16}$$

When the instantaneous channel gains, called the *channel state information* (CSI), are known perfectly at both transmitter and receiver, the transmitter can adapt its transmission strategy relative to the instantaneous channel state. In this case, the Shannon (ergodic) capacity is the maximum mutual information averaged over all channel states. This ergodic capacity is typically achieved using an adaptive transmission policy where the power and data rate vary relative to the channel state variations. When only the receiver has perfect CSI the transmitter must maintain a fixed-rate transmission strategy optimized with respect to its CDI. In this case, ergodic capacity defines the rate that can be achieved based on averaging over all channel states.

- *Outage Capacity (capacity CDF):* The ergodic assumption is not necessarily satisfied in practical communication systems operating on fading channels. In fact, if stringent delay constraints are demanded, as is the case in speech transmission over wireless channels, the ergodicity requirement $T_s \gg T_c = 1/B_d$ cannot be satisfied. In this case, where no significant channel variability occurs during the whole transmission, there may not be a classical Shannon meaning attached to capacity in typical situations. For example, consider a situation of a channel whose fading is so slow that it remains constant for the whole duration of a codeword. This channel is referred to as *quasi-static fading channel*. It is nonergodic, as no codeword will be able to experience all the states of the channel, and hence (7.16) is not valid anymore. In fact, there may be a nonnegligible probability that the value of the actual transmitted rate, no matter how small, exceeds the instantaneous mutual information. This situation gives rise to error probabilities that *do not* decay with the increase of the blocklength. In these circumstances, the channel capacity (which we call *instantaneous* mutual information) is viewed as a random variable, as it depends on the instantaneous channel states. The capacity-versus-outage performance is then determined by the probability that the channel cannot support a given rate (capacity CDF): that is, we associate an outage probability

$$P_{\text{out}}(R) \equiv \Pr\left(C(h) < R\right)$$

to any given rate *R*. We see that outage probability expresses a tradeoff between rate and error probability. The maximum rate that can be supported by the channel with a given outage probability is referred to as *outage capacity*.

Example: Consider a binary channel where the output codeword is equal to the transmitted codeword with probability 1/2 and independent of the transmitted codeword with 1/2. The capacity of this channel is zero because there is *non nonzero rate* at which long codewords can be transmitted with an arbitrarily small error probability is unattainable. However, the capacity formula

$$C = \lim_{n\to\infty} \sup_{P_{X^n}} \frac{1}{n} I(X^n;Y^n)$$

yields C=1/2 bit per channel use. Dobrushin has proved that the above formula is valid for class of *information stable* channels. This example illustrates the difficulty of defining a capacity for nonergodic channels.

Consider the simple case of a flat Rayleigh fading with no dynamics (B_d=0), with channel-state information available to the receiver only. The channel capacity, viewed as a random variable, is given by

$$C(v) = \log_2(1 + v\mathsf{SNR}) \tag{7.17}$$

where $\mathsf{SNR} = E_s / N_0$ is the signal-to-noise ratio and $v = |h|^2$ is exponentially distributed. The capacity R (nats per unit bandwidth) per outage probability is given by

$$P_{\text{out}} \equiv \Pr\left(C(v) \le R\right) = \Pr\left(\ln(1 + v\mathsf{SNR}) \le R\right)$$

$$= 1 - \exp\left(-\frac{e^R - 1}{SNR}\right)$$

In this case only the zero rate R=0 is compatible with P_{out} = 0, thus eliminating any reliable communication in Shannon's sense. It is instructive to note that the ergodic Shannon capacity is no more than the expectation of (7.17).

ε-outage capacity: For a nonergodic channel, we may define an ε-outage capacity as the maximum rate R that can be transmitted with an outage probability $P_{out} = \varepsilon$. Note that the outage probability provides an estimate of word (frame) error probability when the transmitted codewords are long enough.

7.3.2 Ergodic Capacity for a Flat Rayleigh Fading Channel

We now consider the capacity for the simplest model of a single-user, flat fading case. The complex baseband representation of a flat fading channel is given by (7.14). For convenience, we repeat it as follows.

$$y_k = h_k x_k + n_k \tag{7.18}$$

Here, the samples $\{h_k\}$ is assumed to be of the complex circularly symmetric fading process with a one-dimensional probability distribution $p_v()$ of the power $v_k = |h_k|^2$, and a uniform distribution of phase in [-π, π]. As before, We assume that $E[v]=1$, and $\{n_k\}$ is i.i.d. $\mathcal{CN}(0, N_0)$ noise.

■ *Perfect CSI known to receiver only:*

This case is rather standard. Here we assume that $\{h_k\}$ is a stationary ergodic process, which gives rise to a capacity formula

$$\begin{aligned} C_{\text{CSIR}} &= \mathsf{E}_h\left[\log\left(1+|h|^2\,\mathsf{SNR}\right)\right] \\ &= \mathsf{E}_v\left[\log\left(1+v\cdot\mathsf{SNR}\right)\right] \\ &= \int_0^\infty p(v)\log\left(1+v\cdot\mathsf{SNR}\right)dv \\ &= -\frac{1}{\ln 2}\exp\left(\frac{1}{\mathsf{SNR}}\right)Ei\left(-\frac{1}{\mathsf{SNR}}\right) \qquad \text{bits/2D} \end{aligned} \tag{7.19}$$

where $Ei(x)=\int_{-\infty}^{x}\frac{e^t}{t}dt$.

Notice that a simple standard (Gaussian) long codebook will be efficient in this case. However, we should emphasize that contrary to the standard additive Gaussian noise channel, the length of the codebook dramatically depends on the dynamics of the fading process: in fact, it must be long enough for the fading to reflect its ergodic nature.

Let $\gamma=|h|^2\,\mathsf{SNR}$. By Jensen's inequality,

$$\mathsf{E}\left[\log\left(1+|h|^2\,\mathsf{SNR}\right)\right]=\mathsf{E}\left[\log\left(1+\gamma\right)\right]\le\log\left(1+\mathsf{E}[\gamma]\right)=\log\left(1+\bar{\gamma}\right)$$

where $\bar{\gamma}=\mathsf{SNR}\cdot\mathsf{E}[v]$ is the average SNR on the channel. Thus we see that the Shannon capacity of a fading channel with receiver CSI only is always less than the capacity of the AWGN channel with the same average SNR.

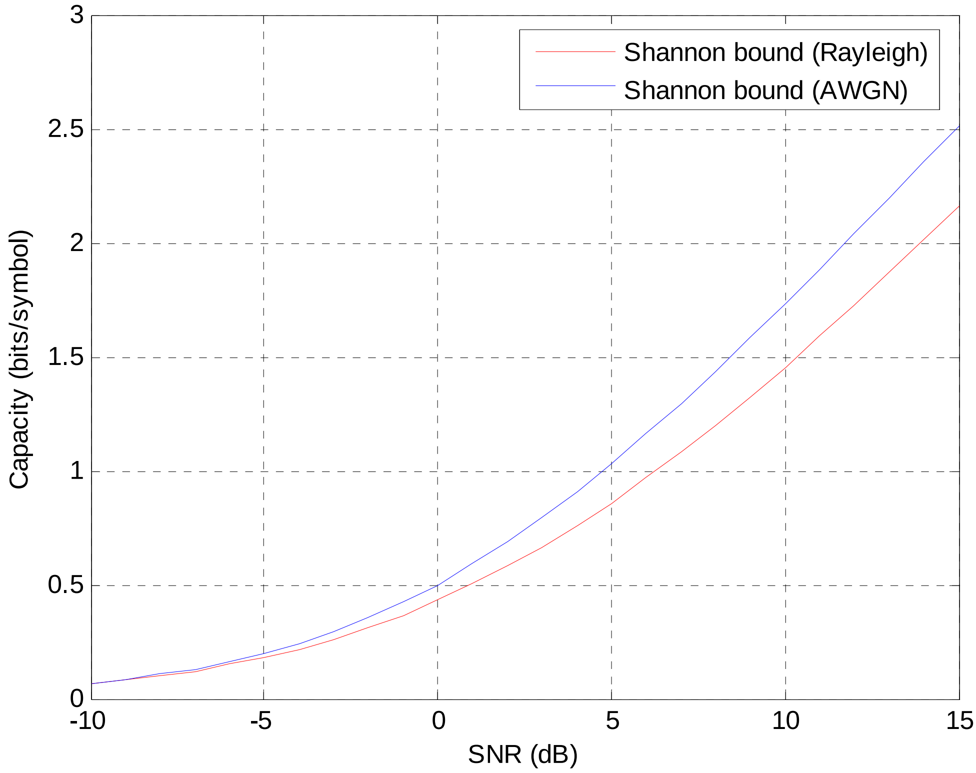


Figure 7.9. Capacity of the independent Rayleigh fading channel with CSIR. Assuming 1D signaling.

■ *Perfect channel-state information available to transmitter and receiver:*

Assume that the channel state information is available to both receiver and transmitter in a causal manner. Under the input-power constraint $E[|x_k|^2] \le P_{av}$, the capacity of the fading channel is given by

$$C_{\mathrm{CSITR}} = \mathsf{E}_v\left[\sup \log\left(1 + \frac{P_s(v)v}{N_0}\right)\right] \qquad \text{bits/2D} \tag{7.20}$$

where the supremum is over all nonnegative power assignments satisfying

$$\mathsf{E}_v\left[P_s(v)\right] \le P_{\mathrm{av}} \tag{7.21}$$

The optimal power assignment $P_s^*(v)$ satisfies [Goldsmith97]

$$P_s^*(v) = \left(\frac{1}{\lambda} - \frac{N_0}{v}\right)^+ \tag{7.22}$$

or

$$\frac{P_s^*(v)}{P_{\mathrm{av}}} = \begin{cases} \dfrac{1}{\gamma_0} - \dfrac{1}{\gamma}, & \gamma \ge \gamma_0 \\ 0, & \gamma < \gamma_0 \end{cases}$$

where the constant λ is determined by the average power constraint and the specific distribution of the fading power $p_v(v)$. Substituting (7.22) into (7.21), we have

$$\mathsf{E}\left[\left(\frac{1}{\lambda}-\frac{N_0}{v}\right)^+\right]=\int_{\lambda N_0}^{\infty} p_v(v)\left(\frac{1}{\lambda}-\frac{N_0}{v}\right)dv = P_{av}$$

In general, the value of λ cannot be solved in closed form and thus must be found numerically. Define $SNR_c(v)=v/N_0$ as the current "channel SNR". Substituing the optimal power control (7.22) into (7.20), the capacity is given by

$$C_{\mathrm{CSITR}} = \mathsf{E}_v\left[\log\left(1+\left(\frac{1}{\lambda}-\frac{N_0}{v}\right)^+\frac{v}{N_0}\right)\right]$$

$$=\int_{\lambda N_0}^{\infty} p_v(v)\log\left[\frac{v}{\lambda N_0}\right]dv$$

The optimal power control policy as in (7.22) gives rise to the *time-water-pouring* interpretation; that is, above a threshold v_0 the lower the deleterious fading (v is large), the larger the instantaneous transmitted power. If v_k is below this threshold, no data is transmitted over the kth signaling interval.

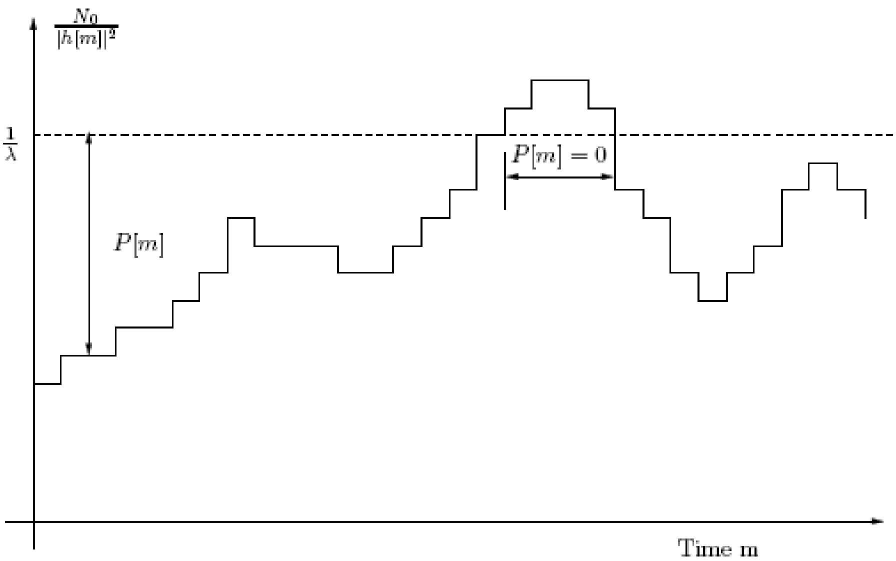
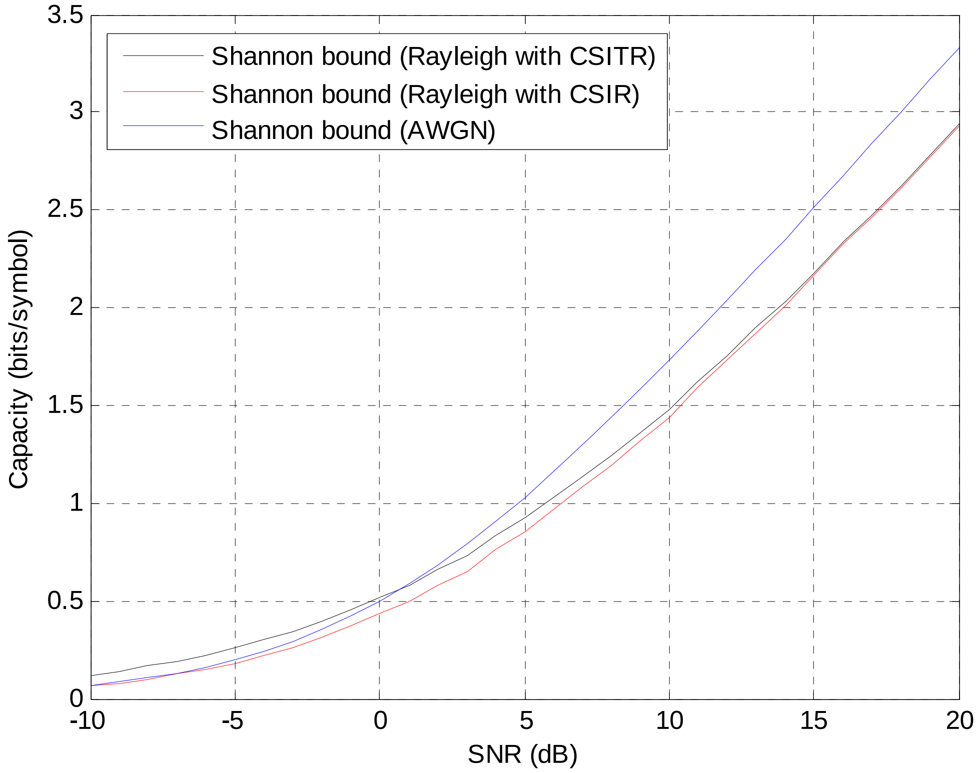


Fig.7.10 Optimization of transmit power assignment $P(v)$ by "water pouring" in time.

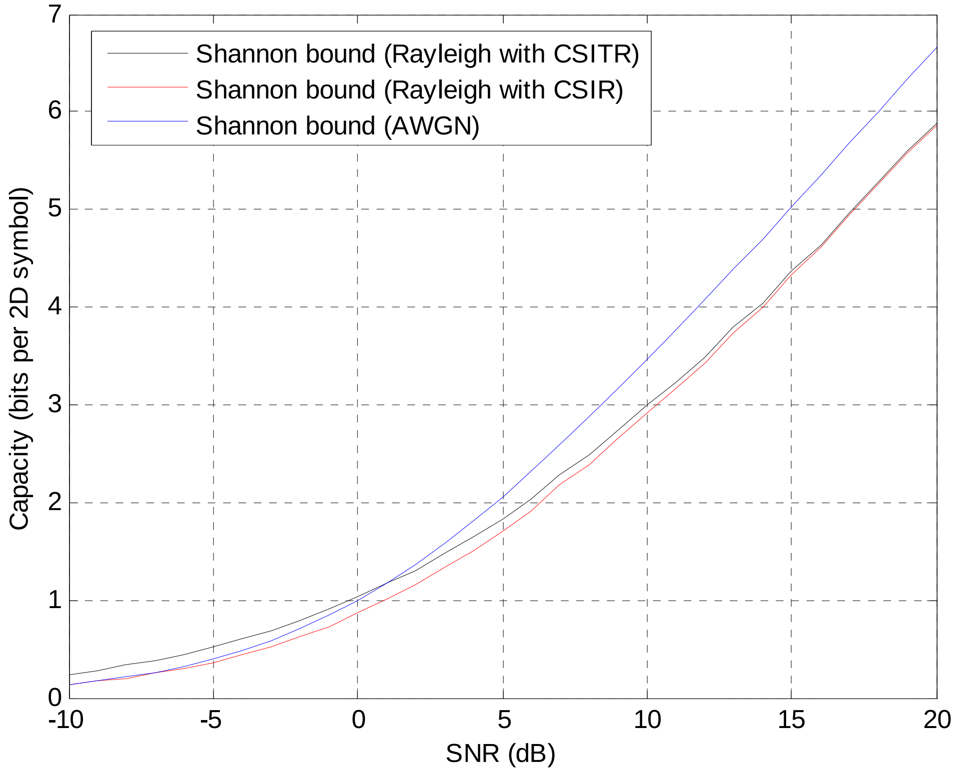
Clearly, the solution here advocates a variable-rate, variable power communication technique, where different codebooks with rate $\log\left(1+P_s^*(v)\mathrm{SNR}_c(v)\right)$ are used when the fading realization is v and the associated assigned power is $P_s^*(v)$.

By Jensen's inequality, (7.20) is always less than the capacity of an AWGN channel with the same average power. This reflects the fact that with fixed received (rather than transmitted) power the fading effect is always deleterious.

Fig.7.11 shows the effect on capacity of CSI at transmitter. Notice that, while the capacity with CSI at the receiver only never exceeds that of the AWGN channel, no such inequality holds for the capacity with CSI at transmitter and receiver. We observe that, unless the SNR is very low, CSI at transmitter increases capacity very little.



(a)



(b)

Figure 7.11. Capacity of the flat Rayleigh fading channel with CSIT and CSIR.
(a) 1D signaling. (b) 2D signaling

- *Unavailable channel-state information:*

In the absence of channel-state information at both transmitter and receiver, for i.i.d. states $\{h_k\}$ the full solution is available for circularly complex distribution of $\{h_k\}$. In fact, it has been shown that the capacity-achieving distribution has a discrete i.i.d. power $|X_k|$ and irrelevant phase. No general closed-form is known for this distribution; however, asymptotic results are available. Specifically, for relatively low values of the average signal-to-noise ratio (SNR)<8dB, only two signaling levels x=0 and $x=\sqrt{\alpha}$ and with respective probabilities $(1-P_\alpha, P_\alpha)$ suffice, where $\alpha P_\alpha = P_{\text{av}}$); and hence the optimum modulation scheme is on-off. Clearly, the capacity-achieving codes in this case deviate markedly from Gaussian codes, which achieve capacity when CSI is available either to the receiver or to both receiver and transmitter.

7.3.3 Capacity of the Independent Fading Channels with Constrained-Constellation

- Capacity with BPSK signaling

With BPSK modulation and coherent detection, the discrete-time model of a independent flat-fading channel is given by

$$y = a \cdot x + n$$

where $x \in \{\sqrt{E_s}, -\sqrt{E_s}\}$ is an input symbol with the average energy E_s, y is the channel output symbol, a is the fading coefficient (or called channel gain), and n is an AWGN sample. Suppose that $a=|h|$ has a Rayleigh distribution. If the perfect channel state information is known at the receiver, then the channel capacity can be expressed as

$$\begin{aligned} C_{\mathrm{CSIR}}^{\mathrm{BPSK}} &= \max_{P(x)}\{I(X;YA)\} = \max_{P(x)}\{I(X;A) + I(X;Y \mid A)\} \\ &= \max_{P(x)}\{I(X;Y \mid A)\} \\ &= \max_{P(x)}\{E_A[I(X;Y \mid A=a)]\} \\ &= \max_{P(x)}\left\{E_{p(x,y,a)}\left[\log\frac{p(y \mid x,a)}{p(y \mid a)}\right]\right\} \end{aligned} \tag{7.23}$$

For symmetric channels with discrete inputs, the maximization in the capacity definition is achieved by an equiprobable input distribution $P(X=\sqrt{E_s}) = P(X=-\sqrt{E_s}) = 1/2$. Noting that $p(x,y,a) = p(y \mid x,a)P_X(x)p_A(a)$, we have

$$\begin{aligned} C_{\mathrm{CSIR}}^{\mathrm{BPSK}} &= \int_a\int_y \sum_x \frac{1}{2} p_A(a)p(y \mid x,a)\log\frac{p(y \mid x,a)}{\sum_{x'} 1/2 \cdot p(y \mid x',a)}dyda \\ &= \int_a\int_y p_A(a)p(y \mid x=\sqrt{E_s},a)\log\frac{p(y \mid x=\sqrt{E_s},a)}{\sum_{x'} 1/2 \cdot p(y \mid x',a)}dyda \\ &= -\int_a\int_y p_A(a)p(y \mid x=\sqrt{E_s},a)\log\left[\frac{1}{2}\left(1+\frac{p(y \mid x=-\sqrt{E_s},a)}{p(y \mid x=\sqrt{E_s},a)}\right)\right]dyda \end{aligned} \tag{7.24}$$

For the case of no CSI is available at both transmitter and receiver, the channel capacity is

$$\begin{aligned} C_{\mathrm{NCSI}}^{\mathrm{BPSK}} &= \max_{P_X(x)}\{I(X;Y)\} \\ &= \max_{P_X(x)}\left\{E_{p(x,y)}\left[\log\frac{p(y \mid x)}{p(y)}\right]\right\} \end{aligned} \tag{7.25}$$

where

$$p(x,y) = \int_a p(x,y,a)da = \int_a p_A(a)P_X(x)p(y \mid x,a)da \tag{7.26}$$

Combining (7.25) and (7.26), and using some simplifications we obtain

$$C_{\mathrm{NCSI}}^{\mathrm{BPSK}} = -\int_a\int_y p_A(a)p(y \mid x=\sqrt{E_s},a)\log\left[\frac{1}{2}\left(1+\Phi(y)\right)\right]dyda \tag{7.27}$$

where

$$\Phi(y)=\frac{\int_a p_A(a)p(y\mid x=-\sqrt{E_s},a)da}{\int_a p_A(a)p(y\mid x=\sqrt{E_s},a)da}$$

Equations (7.24) and (7.27) can be computed using Monte Carlo or numerical integration, and the results are plotted in Fig. 7.12.

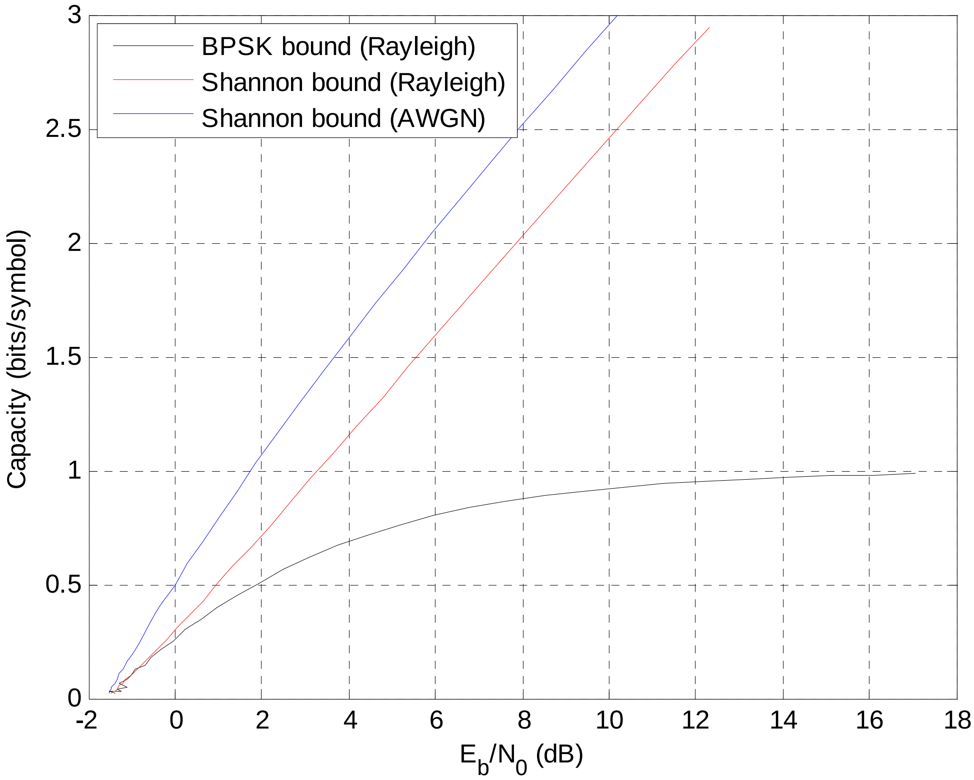


Figure 7.12 Capacity limits of BPSK signaling on the i.i.d. Rayleigh channel with perfect CSI at receiver

■ Capacity with *M*-PSK / *M*-QAM signaling

假定信号星座$\mathcal{X}$中的信号等概地使用，and CSI is perfectly known to the receiver。定义信道转移概率密度函数

$$p(y\mid x)=c\exp\left(-\frac{\|y-hx\|^2}{N_0}\right)$$

where *c* is a normalized constant. Using this expression for capacity (see (7.23)), we have

$$C_{\text{CSIR}}^{\text{MQAM}}=\mathsf{E}_{x,y,h}\left[\log_2\frac{p(y\mid x,h)}{p(y\mid h)}\right]$$

Since

$$\log_2\frac{p(y\mid x,h)}{p(y\mid h)}=\log_2\frac{p(y\mid x,h)}{\sum_{x'\in\mathcal{X}}p(y\mid x',h)P(x')}$$

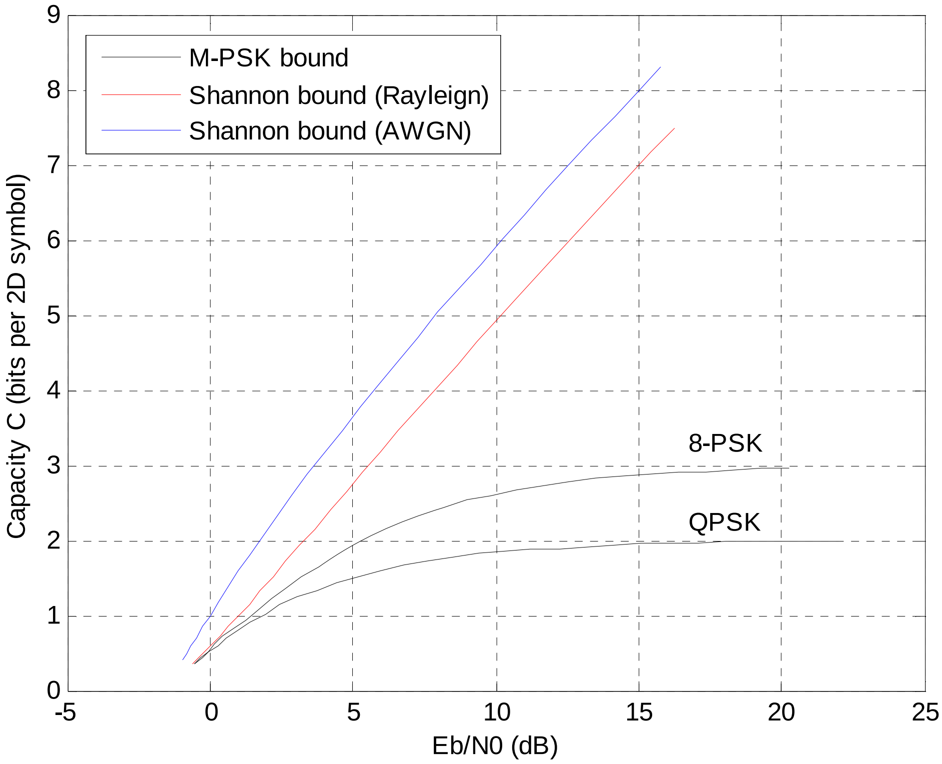
$$= \log_2 \frac{p(y \mid x, h)}{\frac{1}{M} \sum_{x' \in \mathcal{X}} p(y \mid x', h)}$$

$$= \log_2 M + \log_2 \frac{p(y \mid x, h)}{\sum_{x' \in \mathcal{X}} p(y \mid x', h)}$$

where $M = |\mathcal{X}|$, the ergodic capacity of the fading channel with M-QAM signaling is given by

$$C_{\mathrm{CSIR}}^{\mathrm{MQAM}} = \log_2 |\mathcal{X}| - \mathsf{E}_{x,y,h}\left[\log_2 \frac{\sum_{x' \in \mathcal{X}} p(y \mid x', h)}{p(y \mid x, h)} \right] \tag{7.28}$$

See Fig. 7.13 for numerical results.



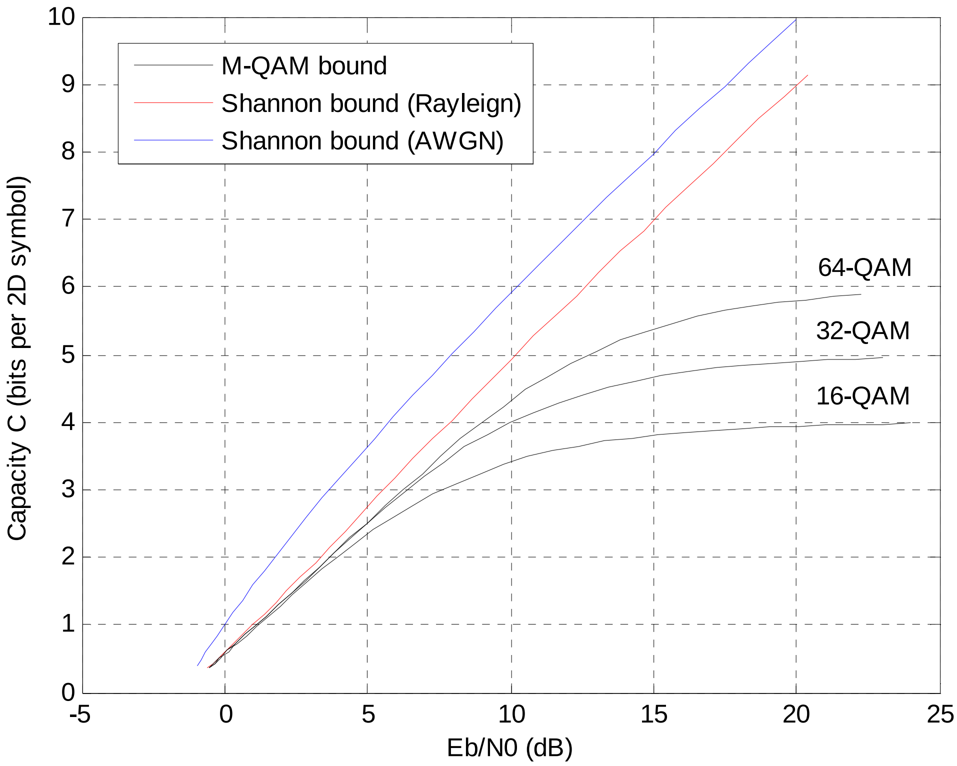


Fig. 7.13 Capacity limits of M-QAM signaling on the i.i.d. Rayleigh channel with perfect CSI at receiver

7.4 Coding for Fading Channels

通过前面的容量分析，我们看到采用编码方法可以实现逼近衰落信道容量限的信息传输，下面我们来讨论编码系统的设计准则与系统错误概率。作为对比，我们首先看一下衰落信道中未编码系统的性能。

7.4.1 Performance of Uncoded System in Independent Fading Channels

Consider the independent fading channel:

$$y = hx + n, \quad x \in \mathcal{X}$$

(未编码信号是标量)

The error probability with perfect CSI at receiver can be evaluated as follows.

(a) We first compute the error probability $P(e \mid h)$ by assuming h constant.

From Chapter 2, the pairwise error probability (PEP) conditioned on h is given by

$$P_2\left(x \to x' \mid h\right) = Q\left(\frac{|h(x - x')|}{\sqrt{2N_0}}\right)$$

Applying union bound, the conditional error probability $P(e \mid x, h)$ is upper-bounded by

$$P(e \mid x, h) \le \sum_{x' \ne x} P_2\left(x \to x' \mid h\right)$$

(b) Next we take the expectation of $P(e \mid h)$ with respect to the r.v. h

$$P_2\left(x \to x'\right) = \mathsf{E}_h\left[P_2\left(x \to x' \mid h\right)\right]$$

Under the assumption of Rayleigh fading, we have

$$P_2\left(x \to x'\right) = \frac{1}{2}\left(1 - \sqrt{\frac{|x-x'|^2/4N_0}{1+|x-x'|^2/4N_0}}\right)$$

Using the approximation

$$1 - \sqrt{\frac{z}{1+z}} \sim \frac{1}{2z}, \quad (z \to \infty)$$

we obtain, as $N_0 \to 0$,

$$P_2\left(x \to x'\right) \sim \frac{1}{|x-x'|^2/N_0} \sim \frac{1}{\mathsf{SNR}}$$

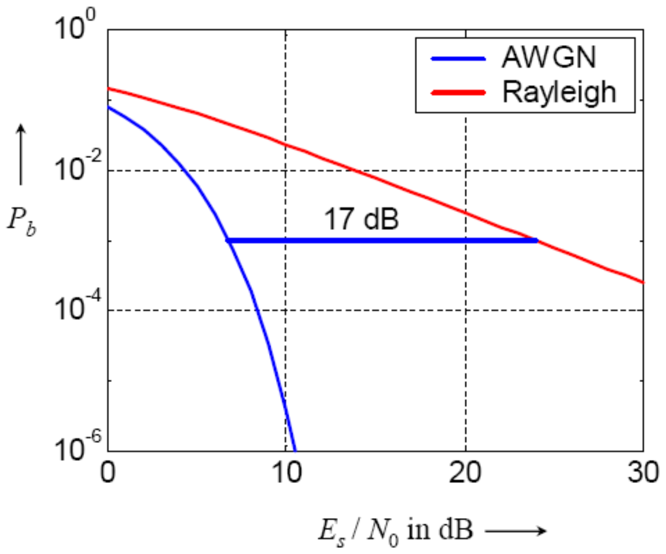
It is seen that the error probability is inversely proportional to SNR. In the following, we will show that coding can be used to improve the performance.

Example: Consider BPSK transmission over the flat Rayleigh fading channel. With $\mathcal{X} = \left\{\sqrt{E_s}, -\sqrt{E_s}\right\}$, we have $|x-x'|^2 = 4E_s$. Hence

$$P(e) = \frac{1}{2}\left(1 - \sqrt{\frac{E_s/N_0}{1+E_s/N_0}}\right)$$

Fig.7.14 shows the error probability of this signaling over the Rayleigh fading channel and the AWGN channel.

从 AWGN 信道与 independent Rayleigh 信道的容量曲线对比、未编码系统的错误概率曲线对比可以看出，它们在容量上的差别（gap）要比它们的错误概率曲线之间的差别小得多，this suggests that coding can be very beneficial to compensate for fading. (编码对衰落信道非常重要、有用)



- AWGN channel:

$$P_b = \frac{1}{2}\cdot \mathrm{erfc}\left(\sqrt{E_s/N_0}\right) = Q\left(\sqrt{2E_s/N_0}\right)$$

- Flat Rayleigh fading channel:

$$P_b = \frac{1}{2}\cdot\left[1-\sqrt{\frac{E_s/N_0}{1+E_s/N_0}}\right]$$

- Channel coding for fading channels essential (time diversity)

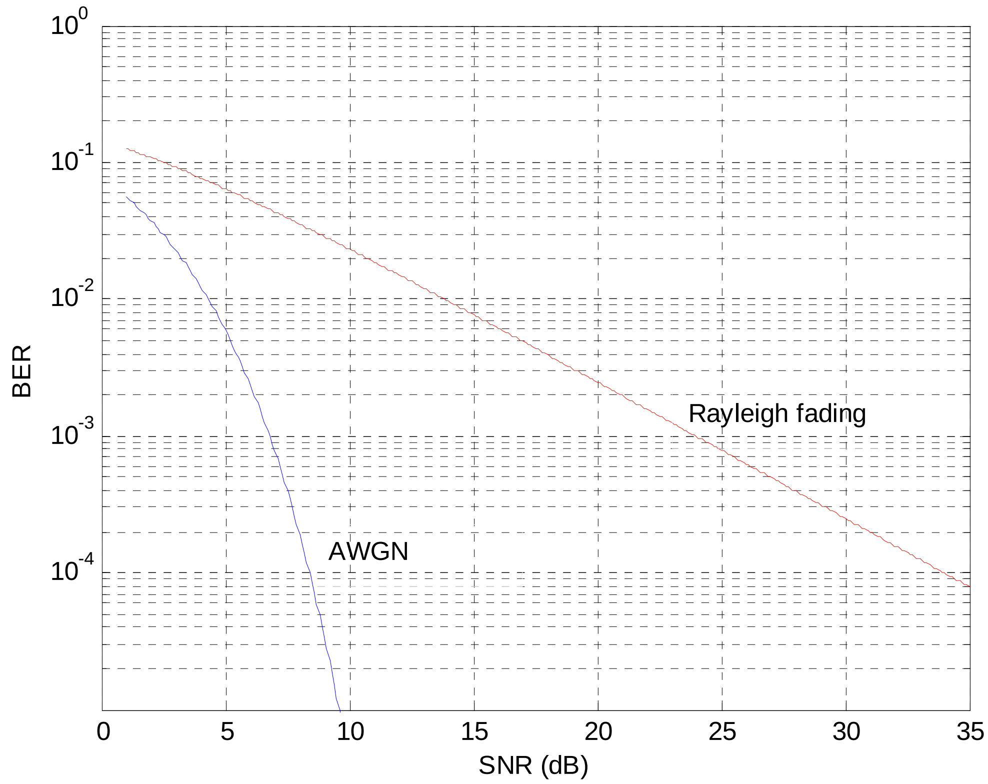


Figure 7.14 BER for AWGN and flat Rayleigh channel (BPSK modulation)

7.4.2 Performance of Coded Systems

Consider now coded transmission. The system model with coded signals can be written in vector form as

$$\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{n}$$

Using the inequality $Q(x) \le e^{-x^2/2}$, we have

$$P_2\left(\mathbf{x}\to\mathbf{x}'\,|\,\mathbf{h}\right)=Q\left(\frac{\left\|\mathbf{h}(\mathbf{x}-\mathbf{x}')\right\|}{\sqrt{2N_0}}\right)\le\exp\left(-\frac{\left\|\mathbf{h}(\mathbf{x}-\mathbf{x}')\right\|^2}{4N_0}\right)$$

Now, the assumption of independent Rayleigh fading yields

$$\begin{aligned}P_2\left(\mathbf{x}\to\mathbf{x}'\,|\,\mathbf{h}\right)&\le \mathsf{E}_{h_1,\ldots,h_n}\left[\exp\left(-\sum_{i=1}^{n}|\,h_i(x_i-x_i')\,|^2/4N_0\right)\right]\\&=\prod_{i=1}^{n}\mathsf{E}_{h_i}\left[\exp\left(-\,|\,h_i(x_i-x_i')\,|^2/4N_0\right)\right]\\&=\prod_{i=1}^{n}\frac{1}{1+|\,x_i-x_i'\,|^2/4N_0}\end{aligned}$$

Let $d_{\mathrm{H}}(\mathbf{x},\mathbf{x}')$ denote the symbol Hamming distance between **x** and **x**', and $\mathcal{I}=\{i\,|\,x_i\ne x_i',1\le i\le n\}$ the (index-) set of the places in which **x**' will differ **x**. Clearly, $|\,\mathcal{I}\,|=d_{\mathrm{H}}(\mathbf{x},\mathbf{x}')$. Then as $N_0\to 0$, we have

$$\begin{aligned}P\left(\mathbf{x}\to\mathbf{x}'\right)&\le\prod_{i\in\mathcal{I}}\frac{1}{1+|\,x_i-x_i'\,|^2/4N_0}\\&=\left[\prod_{i\in\mathcal{I}}\left|x_i-x_i'\right|^2\right]^{-1}\left(\frac{1}{4N_0}\right)^{-d_{\mathrm{H}}(\mathbf{x},\mathbf{x}')}\end{aligned}\tag{7.29}$$

By using the union bound, we see that the error probability is dominant by the pairwise errors with the smallest $d_{\mathrm{H}}(\mathbf{x},\mathbf{x}')$, denoted by $d_{\mathrm{H,min}}$. The minimum Hamming distance of the code is sometimes referred to as *code diversity*.

Notice also the effect of the *product distance*

$$d_p\left(\mathbf{x},\mathbf{x}'\right)\triangleq\prod_{i\in\mathcal{I}}\left|x_i-x_i'\right|^2$$

Its effect is to shift horizontally the curve of PEP vs. SNR. The smallest among the product distances is called *coding gain*.

Result (7.29) shows that a sensible criterion for the selection of a code for the independent Rayleigh fading channel with high SNR (low N_0) is ***the maximization of the minimum Hamming symbol-distance of the code***. The 2nd most important is to have a large product distance.

Example:

7.4.3 Coding for Block Fading Channels

To describe the block fading model conveniently, we define the codeword X as the $F\times v$ matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_F \end{bmatrix}$$

whose *m*th row contains the mth block $\mathbf{x}_m, m = 1,\ldots,F$, with length v. The mth block is sent over a constant-fading channel with gain $\mathbf{h}_m$. The channel output matrix is given by

$$\mathbf{Y} = \mathbf{HX} + \mathbf{N}$$

where $\mathbf{H} \triangleq diag(\mathbf{h}_1,\ldots,\mathbf{h}_F)$, and N is an $F \times v$ matrix of independent Gaussian noise R.Vs.

Let $\mathcal{I} = \{i \mid \mathbf{x}_i \neq \mathbf{x}'_i, 1 \leq i \leq F\}$ be the set of indices i such that $\|\mathbf{x}_i - \mathbf{x}'_i\| \neq 0$. Denote by $D_F(\mathbf{X}, \mathbf{X}')$ the number of rows in which **X** and **X'** differ (this is called the Hamming block distance between **X** and **X'**). Using similar analysis to …, it can be shown [Big04] that

$$P(\mathbf{X} \to \mathbf{X}') \leq \left[\prod_{i\in\mathcal{I}} \|x_i - x'_i\|^2\right]^{-1} \left(\frac{1}{4N_0}\right)^{-D_F(\mathbf{X},\mathbf{X}')}$$

We see that the error probability is (asymptotically for high SNR) inversely proportional to the product of the squared Euclidean distances between the signals transmitted in a block.

7.4.4 Diversity

从上述讨论得知，采用编码，我们能够逼近衰落信道的容量；为了能够进一步逼近 AWGN 信道的容量，我们需要引入 diversity 技术. As expected, capacity of the AWGN channel is approached with the increase of the number of diversity branches.

Diversity techniques provide the receiver with multiple independent looks at the signal to improve reception. Each of those independent looks is considered a diversity branch. The probability that all diversity branches will fade at the same time goes down as the number of branches increases. A tractable definition of the diversity, or diversity gain, is

$$G_d = -\lim_{\text{SNR}\to\infty} \frac{\log(P_e)}{\log(\text{SNR})}$$

where P_e is the probability of error at an received SNR equal to SNR. In other words, diversity is the slope of the error probability curve in terms of the received SNR in a log-log scale.

There are two important issues related to the concept of diversity. One is how to provide the replicas of the transmitted signal at the receiver with the lowest possible consumption of the power, bandwidth, decoding complexity and other resources. The second issue is how to use these replicas of the transmitted signal at the receiver in order to have the highest reduction in the probability of error.

There are many ways to obtain diversity. For example, replica of the transmitted signal can be sent in a different time slot, a different frequency, a different polarization, or a different antenna. The existing diversity techniques are listed below.

Diversity Types:
Time (e.g. coding & interleaving)
Frequency (e.g., DS or FH, OFDM)
Spatial (e.g., multiple antennas)
Polarization
Angular
Interference diversity
Coded diversity
Combining diversity

■ **Time diversity:** Time diversity exploits the time-varying nature of wireless channels. It can be obtained by using channel coding and interleaving. The information symbols are first encoded. These coded symbols are dispersed over time in different coherence periods so that different symbols in the codewords experience independent fades. This is done usually by using an appropriately designed interleaver, such that any two consecutive symbols at the output of the channel code are separated by more than the coherence time of the channel at the output of the interleaver.

The simplest way to obtain time diversity via coding and interleaving is by using repetition code. However, a repetition code does not effectively exploit all the degrees of freedom in the fading channel [Tse05], By using a more sophisticated channel code (such as linear block code, CC, and the codes stated in Chapter 4), a coding gain in addition to the diversity gain may be obtained.

■ **Frequency diversity:** Frequency diversity techniques exploit the variation of channel frequency response over the transmission bandwidth to extract the diversity gain. In direct-sequence spread spectrum systems, the multipath structure in the channel is exploited by modulating the information symbols with a pseudonoise sequence and by using the *rake receiver*. In COFDM systems, the frequency-selective channel is converted into a set of flat-fading channels in parallel. Then by using a channel code across the sub-carriers (flat-fading channels), a frequency diversity can be exploited.

■ **Spatial diversity:** Receive diversity relies on the availability of M>=2 receive antennas at the receiver that are spaced far enough apart such that the channels from the transmitter to each of the receive antennas can be assumed uncorrelated. Decorrelation between receive antennas can also be obtained with polarized antennas for M=2. Angular diversity uses directional antennas to achieve diversity. Different copies of the transmitted signal are collected from different angular directions.

There are several methods to combine signals from different receive antennas to obtain diversity. These methods, in general, can be classified into two categories of combining, namely, (1) selection combining, and (2) gain combining (including equal gain combing, maximal ratio combining, and MMSE combining).

Recently, transmit diversity techniques with the use of multiple transmit antennas have attracted much attention. It has been shown that, over fading channels, transmit diversity will achieve a higher capacity. A detailed discussion on this topic will be provided in Chapter 8.

7.5 Design of Error Control Systems for Fading Channels

（逼近容量限的编码调制系统设计）

- Information-Theoretic Inspired Signaling

In general, the capacity as well as the capacity-achieving distribution imply some underlying structure of optimal coding/signaling. Here we will highlight some recent results of primary practical importance.

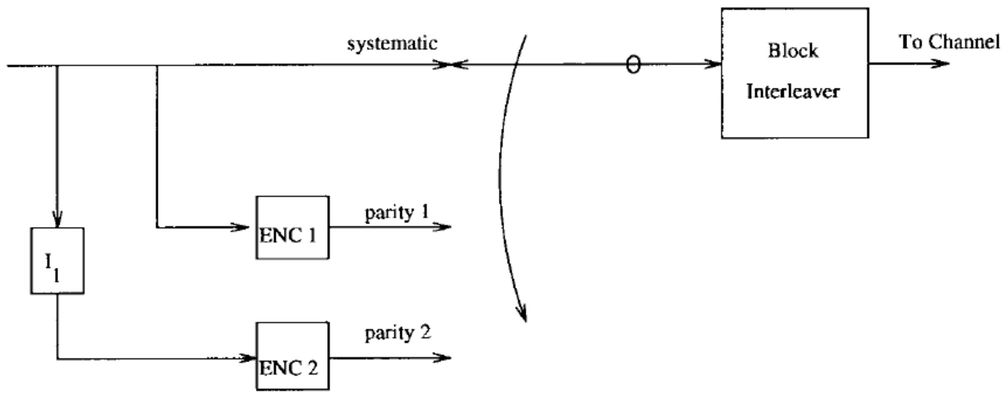
1) For the fast flat-fading model with no CSI available, the *discreteness and peaky nature* of the capacity-achieving inputs envelope gives rise to orthogonal coded pulse-position-like modulations (with efficient iterative detection). It has also been demonstrated in [] that the capacity achieving distribution remains discrete in its input norm for receiver space diversity as well.
2) For frequency-selective fading channel, the coded multicarrier modulation is a good choice. This signaling method is motivated by Shannon’s classical approach.
3) *Spectrally efficient modulation:*

As for the AWGN channel, information theory provides fundamental guidelines for the design of such systems in the realm of a faded time-varying channel. One of the most typical recent examples is *multilevel signaling*, which is an appealing scheme not only in the AWGN case, but also in the presence of fading. In fact, uses the chain rule of mutual information to demonstrate that capacity for the flat-fading channel can be achieved with a multilevel modulation scheme using multistage decoding. Interleavers are introduced on all stages, and rate selection is done, via an information-theoretic criterion (average mutual information for the stage conditioned on previously decoded stages), which if endowed with powerful binary codes, may achieve rates close to capacity, as inherent diversity is provided by the per-stage interleaver.

In [], the bit-interleaved coded-modulation scheme originally advocated by Zehavi, is investigated via information-theoretic tools in the AWGN and flat-fading channel, with known and unknown channel-state information at the receiver. It is concluded that Gray labeling (or pseudo-Gray labeling if the former cannot be achieved) yields overall rates similar to the rates achieved by the signal set itself, while Ungerboeck’s set partitioning inflicts significant degradation.

7.5.1 Turbo Codes for Fading Channels

正如上一节所述，分集在衰落信道的编码系统设计中起着重要作用。为了更好地将 Turbo 码应用于衰落信道，我们在 Turbo 编码之后引入一个信道交织器，如图 x 所示。With sufficiently interleaving we may obtain the (approximately) independent fading coefficients in a correlated fading channel.



7.5.2 LDPC Codes for Fading Channels

7.5.3 Coded-Modulation Systems for Fading Channels

In this subsection we will discuss some design principles of Turbo and LDPC coded modulation systems on Rayleigh fading channels.

From Section 7.4, it is clear that coded modulation (CM) scheme on Rayleigh fading channels must maximize the minimum Hamming distance rather than the minimum squared Euclidean distance. Many of CM schemes (e.g., TCM) for AWGN channels are not at all good on fading channels. For example, the simple TCM scheme shown in Fig. 5.x has minimum Hamming distance one due to the existence of parallel transition. It therefore has no diversity gain on a Rayleigh fading channel. MTCM introduced in Chapter 5 is designed to have larger minimum Hamming distance and is suitable for fading channels. Besides, BICM and MLC schemes perform well on fading channels as well as AWGN channels. We now discuss the design of Turbo and LDPC codes optimized for these systems.

7.6 Adaptive Coded Modulation

Rate compatible (RC) codes are codes where the coded bits of a higher rate code are also used by a lower rate code. This means that the higher rate code is embedded into the lower rate code. Such codes allow transmission of incremental redundancy in ARQ/FEC schemes and continuous rate variation to match the data rate to the channel rate. Rate compatible punctured convolutional (RCPC) codes are a class of such codes, in which high rate codes are obtained by puncturing a rate-1/n convolutional code (referred to as the mother code). Puncturing means that some of the coded bits are removed and never transmitted over the channel. On the other hand, rate compatible nested convolutional (RCNC) coding is a simpler way to obtain low rate code by extending a rate 1/n code to a rate 1/(n+1) code by finding a proper addition generator polynomial. This code is rate compatible with all higher rate codes in the same family.

7.6.2 Hybrid ARQ

ARQ is a simple error control approach/protocol designed for channels with feedback. Hybrid ARQ (H-ARQ) combines the simple ARQ and the FEC code to improve the throughput, in which both error correction and error detection are used. One common implementation is a concatenated coding system with an outer CRC code and an inner block or convolutional code designed for error correction. H-ARQ is roughly classified into 3 categories: namely, type-I, type-II and type-III ARQ. We will focus on type-II H-ARQ.

Throughout this subsection, it is assumed that the feedback channel is error free.

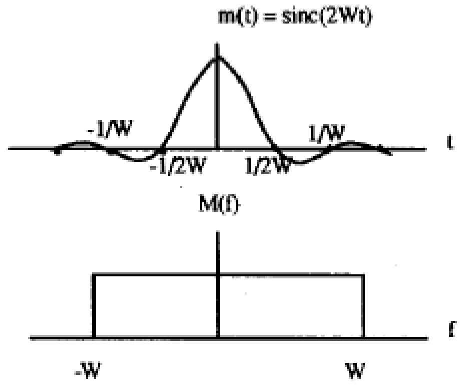
- Hybrid type-II ARQ

- Combined Hybrid type-II ARQ with adaptive modulation

7.7 Information-Theoretic Aspects of Spread-Spectrum Communications

The material presented in this section is mainly based on the paper by J. L. Messay in 1994. We often describe a spread-spectrum system as *one that uses much more bandwidth than it needs*. There seems to be a certain coarse truth in this description. In the following discussions, we will use two types of definitions of bandwidth:

- *Fourier bandwidth*: the bandwidth defined by the Fourier transform of a signal. For example, the signal sinc($2Wt$) has a Fourier bandwidth of W Hz.



- *Shannon bandwidth*: Using signal space representation, a stochastic process $s(t)$ can be represented as

$$s(t) = \sum_{i=1}^{N} s_i \phi_i(t) \qquad 0 \le t \le T \tag{7.30}$$

where $\{\phi_i(t)\}$ are orthonormal functions. When we do this in such a way as to minimize the dimensionality N of the representation, we have arrived at the Shannon bandwidth B, which is defined as

$$B = \frac{1}{2}\frac{N}{T} \qquad \text{(dim/s)} \tag{7.31}$$

That is, *the Shannon bandwidth is one-half the minimum number of dimensions per second required to represent the modulated signal in a signal space.*

Theorem [Fundamental theorem of bandwidth]: The Shannon bandwidth B of a modulated signal is at most equal to its Fourier bandwidth W; i.e., $B \le W$, with (rough) equality holds when the orthonormal functions are $\phi_i(t) = \sqrt{2W}\,\mathrm{sinc}(2Wt - i)$.

With the above notation, we now can define a spread-spectrum system as a communication system in which the modulated signal has a Fourier bandwidth substantially greater than its Shannon bandwidth. Correspondingly, the spreading factor is defined as

$$\gamma = \frac{W}{B}$$

For every communication system, $\gamma \ge 1$.

7.7.1 Some examples

Example 1*:* Consider a TDMA system with K users. Suppose that each user can send L data symbols during each TDMA frame of duration T seconds. By choosing sinc pulses, we can write the modulated signal of a user as

$$s(t) = \sum_{i=1}^{L} b_i \mathrm{sinc}\left(\frac{KL}{T}t - i\right), \quad 0 \le t \le T \tag{7.32}$$

where b_i are the data symbols. It is seen that

$$W = \frac{KL}{2T}$$

The dimension of signal space is $N=L$, so that the Shannon bandwidth $B = \frac{L}{2T}$. The speading factor is then $\gamma=K$. This TDMA system is indeed a spread-spectrum system when the number K of users is large.

Example 2*:* Consider a CDMA system in which each user modulates a user-specific spreading sequence of length L with one data symbol in each symbol period of duration T. We can write the modulated signal of a user as

$$s(t) = b_n \sum_{i=1}^{L} c_i \mathrm{sinc}\left(\frac{L}{T}t - i\right), \quad 0 \le t \le T \tag{7.33}$$

where b_n is the data symbol, and $\mathbf{c} = (c_1, c_2, ..., c_L)$ is the binary spreading sequence. The Fourier bandwidth is

$$W = \frac{L}{2T}$$

By inspection of (7.33) and since $\mathbf{c}$ is fixed, we see that $N=1$. Thus the Shannon bandwidth $B = \frac{1}{2T}$, and the speading factor is then $\gamma=L$.

Example 3*:* Consider the scenario that a user transmits an M-ary pulse-position modulation (PPM) signal in each T second interval; i.e.,

$$s(t) = A\mathrm{sinc}\left(\frac{M}{T}t - b_n\right), \quad 0\le t\le T \tag{7.34}$$

where $b_n \in \{1, 2, ..., M\}$ is the transmitted data symbol and A is a fixed amplitude. The Fourier bandwidth is

$$W = \frac{M}{2T}$$

From (7.34), N=M. Thus the Shannon bandwidth $B = \frac{M}{2T}$, and the speading factor is then γ=1. It follows that such an M-ary PPM communication system is never a spread-spectrum system, even though it utilizes a very large Fourier bandwidth.

<u>*Example 4*</u>*:* Consider the scenario that a user employs antipodal signaling to transmit the output coded symbols of a rate-1/*n* trellis encoder fed by random formation bits. Suppose that n such coded symbols $\{b_i\}$ are transmitted in every *T* second interval. Then

$$s(t) = \sum_{i=1}^{n} b_i \mathrm{sinc}\left(\frac{n}{T}t - i\right), \quad 0\le t\le T \tag{7.35}$$

The Fourier bandwidth is given by

$$W = \frac{n}{2T}$$

Since the coded symbols b_i in (7.35) will take on such a variety of different possible binary patterns that one cannot imbed the set of possible $s(t)$ in a smaller dimensional space than is required when all *n* binary symbols can be independently chosen. Thus we have $B = \frac{n}{2T}$ and γ=1. A non-trivially coded BPSK system is never a spread-spectrum system.

<u>*Example 5*</u>*:* Same as example 4 except that the code is a trivial R=1/n (repetition) code with two binary (±1) codewords, $(c_1, c_2, ..., c_n)$ and its negative. Then

$$s(t) = b_1 \sum_{i=1}^{n} c_i \mathrm{sinc}\left(\frac{n}{T}t - i\right), \quad 0\le t\le T$$

where b_1 is the information bit encoded. The Fourier bandwidth is

$$W = \frac{n}{2T}$$

Since this trivial code consists only of two codewords, the signal space has collapsed to a one-dimensional space; i.e., N=1. The Shannon bandwidth is thus $B = \frac{1}{2T}$, and hence γ=n. In fact, such trivial coding gives us a CDMA system.

From aforementioned examples, we can see that

- a large ratio of Fourier bandwidth to data rate does not imply that a system is

spread-spectrum (see example 3);
- in a coded CDMA system, the Fourier bandwidth expansion due to nontrivial coding is different from the kind of the Fourier bandwidth expansion due to direct-sequence multiplication.

7.7.2 Advantages of Spread-Spectrum Syetems

Now, a nature question is why we need spread-spectrum siganls. The original motivation is that a spread-spectrum signal has *low probability of interception* (LPI). If the signal is confined to a small number of N of dimensions within the global signal space of dimension $2WT=N\gamma$ in which all signals of bandwidth W and time-limited to $0\le t\le T$ must lie, and if there are parameters of the signal that can be varied to create a very large number of possible choices for the N-dimensional signal space occupied by the signal, then one can achieve LPI by selecting the value of these parameters at random. For the CDMA signal of example, there are 2^L possible choices of the binary parameters $c_1, c_2, ..., c_L$, but changing the sign of all parameters leaves the signal in the same one-dimensional signal space.

For the TDMA signal, the only parameter that can be varied is the choice of the L consecutive symbol periods (out of the total of KL such periods) in which data symbols will be transmitted. ➔time-hoping

Interpretation of LPI in frequency domain: *low power spectral density*

The twin brother of LPI is electromagnetic compatibility (EMC). If it is hard to determine whether a signal is present, then that signal cannot be interfering substantially with other commonly present signals.

7.7.3 Coding and Spreading

Consider a coded communication system. The main parameters that we are interested in are listed below:

1) the information rate R, measured in bits per second at the modulator input;
2) the capacity C of the channel created by the modulator and the band-limited AWGN waveform channel;
3) the average power, P, of the modulated signal;
4) the one-sided noise power spectral density N_0 of the AWGN;
5) the Fourier bandwidth W of the band-limited AWGN waveform channel;
6) the Shannon bandwidth B of the modulated signal.

With these quantities, the Shannon's famous capacity formula for band-limited AWGN channels is expressed as

$$C = B\log_2\left(1+\frac{P}{N_0 B}\right) \qquad \text{bits/s} \tag{7.36}$$

Equation (7.36) is precisely the equation that Shannon derives in []. Since $B\le W$, it follows that

$$C \le W\log_2\left(1+\frac{P}{N_0 W}\right) \qquad \text{bits/s} \tag{7.37}$$

with equality if and only if $B=W$. "The reason that Shannon wrote (7.37) with an equality sign rather than (7.36) in his final expression for the capacity of the AWGN channel is that he assumed that the choice of the modulator was up to the sender and that thus the sender would choose a modulator with $B=W$ to obtain (maximum) capacity."
So the use of spreading-spectrum is not for increasing capacity but for other reasons such as those considered in the previous subsection.

The Shannon bandwidth B is always proportional to the symbol rate R_s, which we define as the number of modulation symbols transmitted per second. Using a code with code rate R_c, we have

$$R_s = \frac{R}{R_c} \qquad \text{(symbols/s)}$$

Because the Shannon bandwidth is proportional to R_s, we see that *coding increase the Shannon bandwidth by a factor proportional to 1/R_s*. This is the true nature of the "bandwidth expansion" due to coding.

附录

1. 衰落信道的仿真方法

Rayleigh 平衰落信道的仿真方法

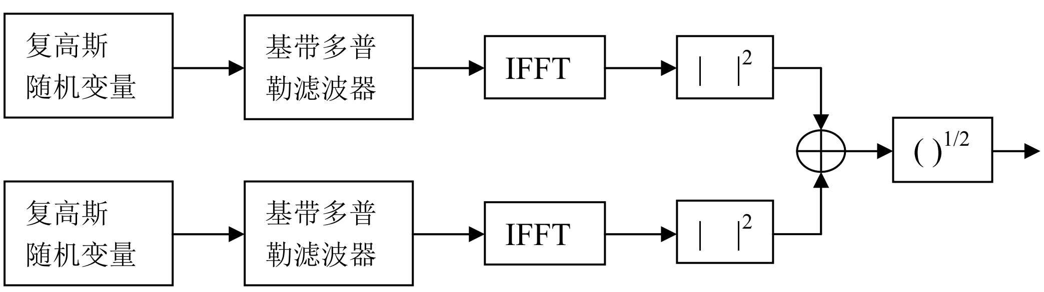
- 对于充分交织(fully interleaved)的 Rayleigh 慢衰落信道，衰落幅度 a_k=|h|之间互不相关，可以用 Rayleigh 随机变量来仿真，于是我们有如下构造方法：

$$a = \sqrt{x^2 + y^2}$$

式中 x、y 为独立同分布的高斯随机变量，其均值为 0，方差 $\sigma^2 = 0.5$。这样，接收信号的平均信噪比为

$$\bar{\gamma}_b = \frac{E_b}{N_0} E[a^2] = \frac{E_s}{2\sigma_n^2 R_c}$$

- 对于相关的 Rayleigh 慢衰落信道，衰落因子 a_k 是复高斯随机过程在 k 时刻的幅度，我们通过 Doppler 滤波器来产生所要求的响应。Doppler 滤波器可在时域实现，也可在频域实现，图 4.1 所示仿真方法是频域实现，图中的多普勒滤波器用于形成衰落的功率谱。



2. 相关衰落产生代码

1) Matlab

```
function Z = Rayfadsim(Fd, Ts, Ns)

%-----------------------------------------------------------
%     Z = Rayfadsim(Fd, Ts, Ns)
%-------------------------------------
%     This function can be used to generate multiple uncorrelated
%     Rayleigh channel fading waveforms with accurate second-order
%     statistics compared with those of Clarke's reference model.
%
%         Fd:    maximum Doppler frequency
%         Ts:    sampling period
%         Ns:    number of samples
%         Z:     (1, Ns) complex fading vector with unit variance,
%                i.e., R_{ZZ}(\tau) = J_0(2 \pi Fd \tau).
%-----------------------------------------------------------
%
%    Reference:
%         This simulator is built based on the following paper:
%
%         Yahong R. Zheng and Chengshan Xiao,``Improved models for
%         the generation of multiple uncorrelated Rayleigh fading
%         waveforms," IEEE Communicaitons Letters, vol.6, June 2002.
%
%-----------------------------------------------------------
% This matlab code was created by
%
%                 Dr. Chengshan Xiao
%                 Dept. of Electrical & Computer Eng.
%                 University of Missouri-Columbia
%                 Columbia, MO 65211, USA
%                 Phone: 573-884-5367
%                 Fax:      673-882-0397
%                 Email: xiaoc@missouri.edu
%                 http://www.ee.missouri.edu/faculty/xiao.htm
%
%    Date of Creation: November 12, 2001
%    Date of last modification: May 12, 2002
%
%    Remark: There are many ways to implement the 16 models provided
```

```
%               by Zheng and Xiao. This one is only an example and it
%               is good for generating long sequence of the fading,
%               with limited computer memory. However, this example
%               is not the most computationally efficient one due to
%               the ``for-loop" fact in Matlab. To remove the ``for-
%               loop" will need more computer memories.
%
%------------------------------------------------------------


%----[ set up fading channel parameters ]------------------------

        M=16; N=4*M;
        dop_gain = sqrt(1/M) * ones(1,M); % This gain differs sqrt(2)
                                         % from the paper to get Var(Z)=1.
        theta=(2*rand-1)*pi;
        doppler = Fd * cos(2*pi/N*[1:M]+theta/N-pi/N);
        phi=(2*rand(M,1)-1)*pi;
        varphi=(2*rand(M,1)-1)*pi;
        state = zeros(M,1);
        dop_update = (2*pi*doppler * Ts).';

%----[ generate fading channel samples ]-------------------------

      Z = zeros(1,Ns);
      for (k=1:Ns)
          Z(k) = dop_gain * [cos(state+phi)+i*sin(state+varphi)];
          state = dop_update + state;
      end

return
```

2) VC++

```
double Z[2*Ns] ;    // Fading Coeff
double *Rayfadsim(double Fd, double Ts, int Ns, int M)
{
      int i, j, N = 4 * M;
      double dop_gain, theta;
      double *doppler, *dop_update;
      double *phi, *varphi, *state;

      doppler = (double*)malloc(sizeof(double) * M);
```

```
    phi = (double*)malloc(sizeof(double) * M);
    varphi = (double*)malloc(sizeof(double) * M);
    state = (double*)malloc(sizeof(double) * M);
    dop_update = (double*)malloc(sizeof(double) * M);

    // set up fading channel parameters
    dop_gain = sqrt(1./M);          //  This gain differs sqrt(2), it is  from the paper to get Var(h)=1.
    theta = (2 * UniformRV() - 1) * PI;
    for (i = 0; i < M; i++) {
        doppler[i] = Fd * cos( ((2 * PI) / N) * (i + 1) + (theta / N) - (PI / N));
        phi[i] = (2 * UniformRV() - 1) * PI;
        varphi[i] = (2 * UniformRV() - 1) * PI;
        state[i] = 0;
        dop_update[i] = 2 * PI * doppler[i] * Ts;
    }
    for(i = 0; i < 2 * Ns; i++) {
        Z[i] = 0;
    }

    //generate fading channel samples
    for (i = 0; i < Ns; i++) {
        for (j = 0; j < M; j++) {
            Z[2 * i] += dop_gain * cos(state[j] + phi[j]);
            Z[2 * i + 1] += dop_gain * sin(state[j] + varphi[j]);
            state[j] += dop_update[j];
        }
    }

    free(doppler);
    free(phi);
    free(varphi);
    free(dop_update);
    free(state);
    free(Z);
}
```

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol.44, pp.2619-2692, Oct. 1998.

[2] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[3] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[4] E. Biglieri, *Coding for Wireless Channels*. Springer Science+Business Media, Inc., 2005.

[5] Massey, “Information theory aspects of spread spectrum communications,” conference, 1994.

[6] E. K. Hall and S. G. Wilson, “Design and analysis of turbo codes on Rayleigh fading channels,” *IEEE J. Select Areas Commun*, vol.16, no.2, pp.160-174, Feb. 1998.

[7] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), “The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels,” *IEEE Trans. Inform. Theory*, vol.47, no.4, pp.1290-1301, May 2001.