Chapter 3

Statistical Multipath Channel Models

In this chapter we examine fading models for the constructive and destructive addition of different multipath components introduced by the channel. While these multipath effects are captured in the ray-tracing models from Chapter 2 for deterministic channels, in practice deterministic channel models are rarely available, and thus we must characterize multipath channels statistically. In this chapter we model the multipath channel by a random time-varying impulse response. We will develop a statistical characterization of this channel model and describe its important properties.

If a single pulse is transmitted over a multipath channel the received signal will appear as a pulse train, with each pulse in the train corresponding to the LOS component or a distinct multipath component associated with a distinct scatterer or cluster of scatterers. An important characteristic of a multipath channel is the time delay spread it causes to the received signal. This delay spread equals the time delay between the arrival of the first received signal component (LOS or multipath) and the last received signal component associated with a single transmitted pulse. If the delay spread is small compared to the inverse of the signal bandwidth, then there is little time spreading in the received signal. However, when the delay spread is relatively large, there is significant time spreading of the received signal which can lead to substantial signal distortion.

Another characteristic of the multipath channel is its time-varying nature. This time variation arises because either the transmitter or the receiver is moving, and therefore the location of reflectors in the transmission path, which give rise to multipath, will change over time. Thus, if we repeatedly transmit pulses from a moving transmitter, we will observe changes in the amplitudes, delays, and the number of multipath components corresponding to each pulse. However, these changes occur over a much larger time scale than the fading due to constructive and destructive addition of multipath components associated with a fixed set of scatterers. We will first use a generic time-varying channel impulse response to capture both fast and slow channel variations. We will then restrict this model to narrowband fading, where the channel bandwidth is small compared to the inverse delay spread. For this narrowband model we will assume a quasi-static environment with a fixed number of multipath components each with fixed path loss and shadowing. For this quasi-static environment we then characterize the variations over short distances (small-scale variations) due to the constructive and destructive addition of multipath components. We also characterize the statistics of wideband multipath channels using two-dimensional transforms based on the underlying time-varying impulse response. Discrete-time and space-time channel models are also discussed.

3.1 Time-Varying Channel Impulse Response

Let the transmitted signal be as in Chapter 2:

$$s(t) = \Re\left\{u(t)e^{j2\pi f_c t}\right\} = \Re\left\{u(t)\right\}\cos(2\pi f_c t) - \Im\left\{u(t)\right\}\sin(2\pi f_c t), \tag{3.1}$$

where $u(t)$ is the complex envelope of $s(t)$ with bandwidth B_u and f_c is its carrier frequency. The corresponding received signal is the sum of the line-of-sight (LOS) path and all resolvable multipath components:

$$r(t) = \Re\left\{\sum_{n=0}^{N(t)} \alpha_n(t)u(t-\tau_n(t))e^{j(2\pi f_c(t-\tau_n(t))+\phi_{D_n})}\right\}, \tag{3.2}$$

where $n = 0$ corresponds to the LOS path. The unknowns in this expression are the number of resolvable multipath components $N(t)$, discussed in more detail below, and for the LOS path and each multipath component, its path length $r_n(t)$ and corresponding delay $\tau_n(t) = r_n(t)/c$, Doppler phase shift $\phi_{D_n}(t)$ and amplitude $\alpha_n(t)$.

The nth resolvable multipath component may correspond to the multipath associated with a single reflector or with multiple reflectors clustered together that generate multipath components with similar delays, as shown in Figure 3.1. If each multipath component corresponds to just a single reflector then its corresponding amplitude $\alpha_n(t)$ is based on the path loss and shadowing associated with that multipath component, its phase change associated with delay $\tau_n(t)$ is $e^{-j2\pi f_c\tau_n(t)}$, and its Doppler shift $f_{D_n}(t) = v\cos\theta_n(t)/lambda$ for $\theta_n(t)$ its angle of arrival. This Doppler frequency shift leads to a Doppler phase shift of $\phi_{D_n} = \int_t 2\pi f_{D_n}(t)dt$. Suppose, however, that the nth multipath component results from a reflector cluster[1]. We say that two multipath components with delay τ_1 and τ_2 are **resolvable** if their delay difference significantly exceeds the inverse signal bandwidth: $|\tau_1 - \tau_2| >> B_u^{-1}$. Multipath components that do not satisfy this resolvability criteria cannot be separated out at the receiver, since $u(t-\tau_1) \approx u(t-\tau_2)$, and thus these components are **nonresolvable**. These nonresolvable components are combined into a single multipath component with delay $\tau \approx \tau_1 \approx \tau_2$ and an amplitude and phase corresponding to the sum of the different components. The amplitude of this summed signal will typically undergo fast variations due to the constructive and destructive combining of the nonresolvable multipath components. In general wideband channels have resolvable multipath components so that each term in the summation of (3.2) corresponds to a single reflection or multiple nonresolvable components combined together, whereas narrowband channels tend to have nonresolvable multipath components contributing to each term in (3.2).

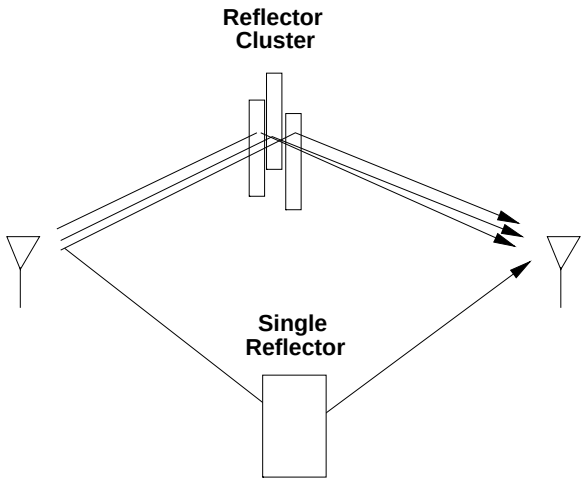


Figure 3.1: A Single Reflector and A Reflector Cluster.

Since the parameters $\alpha_n(t)$, $\tau_n(t)$, and $\phi_{D_n}(t)$ associated with each resolvable multipath component change over time, they are characterized as random processes which we assume to be both stationary and ergodic. Thus, the received signal is also a stationary and ergodic random process. For wideband channels, where each term in

[1]Equivalently, a single "rough" reflector can create different multipath components with slightly different delays.

(3.2) corresponds to a single reflector, these parameters change slowly as the propagation environment changes. For narrowband channels, where each term in (3.2) results from the sum of nonresolvable multipath components, the parameters can change quickly, on the order of a signal wavelength, due to constructive and destructive addition of the different components.

We can simplify $r(t)$ by letting

$$\phi_n(t) = 2\pi f_c \tau_n(t) - \phi_{D_n}. \tag{3.3}$$

Then the received signal can be rewritten as

$$r(t) = \Re\left\{\left[\sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} u(t-\tau_n(t))\right] e^{j2\pi f_c t}\right\}. \tag{3.4}$$

Since $\alpha_n(t)$ is a function of path loss and shadowing while $\phi_n(t)$ depends on delay and Doppler, we typically assume that these two random processes are independent.

The received signal $r(t)$ is obtained by convolving the baseband input signal $u(t)$ with the equivalent lowpass time-varying channel impulse response $c(\tau, t)$ of the channel and then upconverting to the carrier frequency[2]:

$$r(t) = \Re\left\{\left(\int_{-\infty}^{\infty} c(\tau, t) u(t-\tau) d\tau\right) e^{j2\pi f_c t}\right\}. \tag{3.5}$$

Note that $c(\tau, t)$ has two time parameters: the time t when the impulse response is observed at the receiver, and the time $t - \tau$ when the impulse is launched into the channel relative to the observation time t. If at time t there is no physical reflector in the channel with multipath delay $\tau_n(t) = \tau$ then $c(\tau, t) = 0$. While the definition of the time-varying channel impulse response might seem counterintuitive at first, $c(\tau, t)$ must be defined in this way to be consistent with the special case of time-invariant channels. Specifically, for time-invariant channels we have $c(\tau, t) = c(\tau, t + T)$, i.e. the response at time t to an impulse at time $t - \tau$ equals the response at time $t + T$ to an impulse at time $t + T - \tau$. Setting $T = -t$, we get that $c(\tau, t) = c(\tau, t - t) = c(\tau)$, where $c(\tau)$ is the standard time-invariant channel impulse response: the response at time τ to an impulse at zero or, equivalently, the response at time zero to an impulse at time $-\tau$.

We see from (3.4) and (3.5) that $c(\tau, t)$ must be given by

$$c(\tau, t) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} \delta(\tau - \tau_n(t)), \tag{3.6}$$

where $c(\tau, t)$ represents the equivalent lowpass response of the channel at time t to an impulse at time $t - \tau$. Substituting (3.6) back into (3.5) yields (3.4), thereby confirming that (3.6) is the channel's equivalent lowpass

[2] See Appendix A for discussion of the lowpass equivalent representation for bandpass signals and systems.

time-varying impulse response:

$$\begin{aligned} r(t) &= \Re\left\{\left[\int_{-\infty}^{\infty} c(\tau,t)u(t-\tau)d\tau\right]e^{j2\pi f_c t}\right\} \\ &= \Re\left\{\left[\int_{-\infty}^{\infty}\sum_{n=0}^{N(t)}\alpha_n(t)e^{-j\phi_n(t)}\delta(\tau-\tau_n(t))u(t-\tau)d\tau\right]e^{j2\pi f_c t}\right\} \\ &= \Re\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)e^{-j\phi_n(t)}\left(\int_{-\infty}^{\infty}\delta(\tau-\tau_n(t))u(t-\tau)d\tau\right)\right]e^{j2\pi f_c t}\right\} \\ &= \Re\left\{\left[\sum_{n=0}^{N(t)}\alpha_n(t)e^{-j\phi_n(t)}u(t-\tau_n(t))\right]e^{j2\pi f_c t}\right\}, \end{aligned}$$

where the last equality follows from the sifting property of delta functions: $\int \delta(\tau-\tau_n(t))u(t-\tau)d\tau = \delta(t-\tau_n(t))*u(t) = u(t-\tau_n(t))$. Some channel models assume a continuum of multipath delays, in which case the sum in (3.6) becomes an integral which simplifies to a time-varying complex amplitude associated with each multipath delay τ:

$$c(\tau,t) = \int \alpha(\xi,t)e^{-j\phi(\xi,t)}\delta(\tau-\xi)d\xi = \alpha(\tau,t)e^{-j\phi(\tau,t)}. \tag{3.7}$$

To give a concrete example of a time-varying impulse response, consider the system shown in Figure 3.2, where each multipath component corresponds to a single reflector. At time t_1 there are three multipath components associated with the received signal with amplitude, phase, and delay triple (α_i,ϕ_i,τ_i), $i=1,2,3$. Thus, impulses that were launched into the channel at time $t_1-\tau_i, i=1,2,3$ will all be received at time t_1, and impulses launched into the channel at any other time will not be received at t_1 (because there is no multipath component with the corresponding delay). The time-varying impulse response corresponding to t_1 equals

$$c(\tau,t_1) = \sum_{n=0}^{2}\alpha_n e^{-j\phi_n}\delta(\tau-\tau_n) \tag{3.8}$$

and the channel impulse response for $t=t_1$ is shown in Figure 3.3. Figure 3.2 also shows the system at time t_2, where there are two multipath components associated with the received signal with amplitude, phase, and delay triple $(\alpha_i',\phi_i',\tau_i')$, $i=1,2$. Thus, impulses that were launched into the channel at time $t_2-\tau_i', i=1,2$ will all be received at time t_2, and impulses launched into the channel at any other time will not be received at t_2. The time-varying impulse response at t_2 equals

$$c(\tau,t_2) = \sum_{n=0}^{1}\alpha_n' e^{-j\phi_n'}\delta(\tau-\tau_n') \tag{3.9}$$

and is also shown in Figure 3.3.

If the channel is time-invariant then the time-varying parameters in $c(\tau,t)$ become constant, and $c(\tau,t)=c(\tau)$ is just a function of τ:

$$c(\tau) = \sum_{n=0}^{N}\alpha_n e^{-j\phi_n}\delta(\tau-\tau_n), \tag{3.10}$$

for channels with discrete multipath components, and $c(\tau)=\alpha(\tau)e^{-j\phi(\tau)}$ for channels with a continuum of multipath components. For stationary channels the response to an impulse at time t_1 is just a shifted version of its response to an impulse at time $t_2, t_1 \neq t_2$.

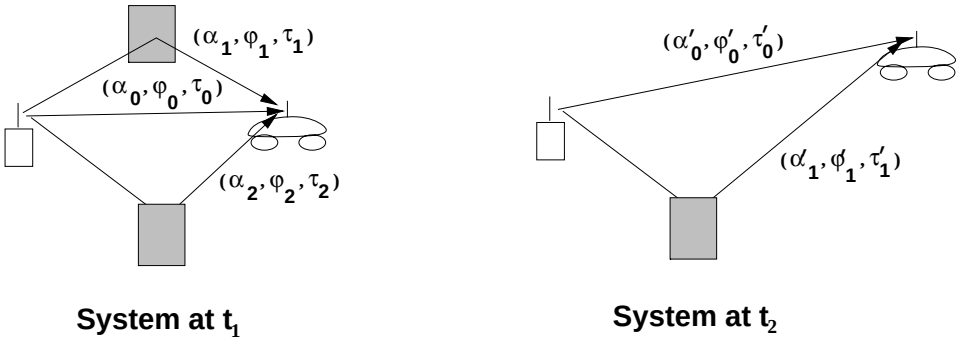


Figure 3.2: System Multipath at Two Different Measurement Times.

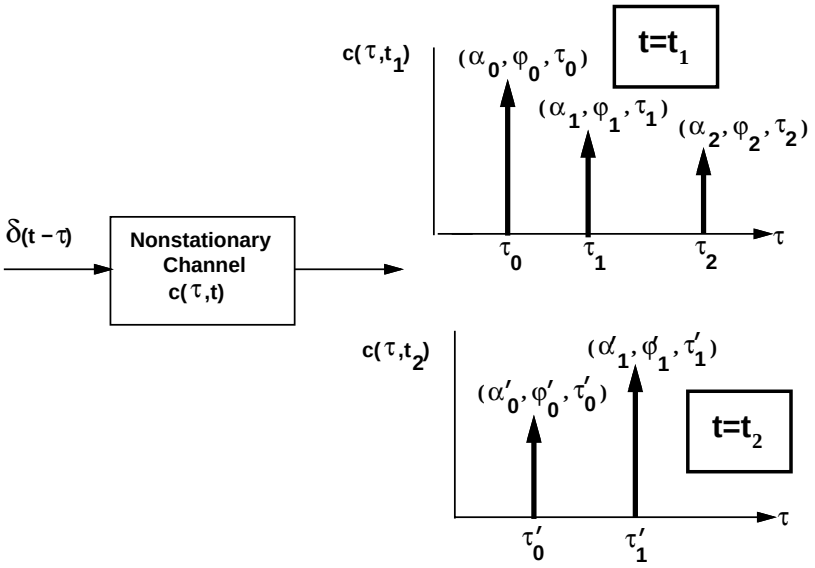


Figure 3.3: Response of Nonstationary Channel.

Example 3.1: Consider a wireless LAN operating in a factory near a conveyor belt. The transmitter and receiver have a LOS path between them with gain α_0, phase ϕ_0 and delay τ_0. Every T_0 seconds a metal item comes down the conveyor belt, creating an additional reflected signal path in addition to the LOS path with gain α_1, phase ϕ_1 and delay τ_1. Find the time-varying impulse response $c(\tau, t)$ of this channel.

Solution: For $t \neq nT_0$, $n = 1, 2, \ldots$ the channel impulse response corresponds to just the LOS path. For $t = nT_0$ the channel impulse response has both the LOS and reflected paths. Thus, $c(\tau, t)$ is given by

$$c(\tau,t) = \begin{cases} \alpha_0 e^{j\phi_0}\delta(\tau-\tau_0) & t \neq nT_0 \\ \alpha_0 e^{j\phi_0}\delta(\tau-\tau_0) + \alpha_1 e^{j\phi_1}\delta(\tau-\tau_1) & t = nT_0 \end{cases}$$

Note that for typical carrier frequencies, the nth multipath component will have $f_c\tau_n(t) >> 1$. For example, with $f_c = 1$ GHz and $\tau_n = 50$ ns (a typical value for an indoor system), $f_c\tau_n = 50 >> 1$. Outdoor wireless

systems have multipath delays much greater than 50 ns, so this property also holds for these systems. If $f_c\tau_n(t) >> 1$ then a small change in the path delay $\tau_n(t)$ can lead to a very large phase change in the nth multipath component with phase $\phi_n(t) = 2\pi f_c\tau_n(t) - \phi_{D_n} - \phi_0$. Rapid phase changes in each multipath component gives rise to constructive and destructive addition of the multipath components comprising the received signal, which in turn causes rapid variation in the received signal strength. This phenomenon, called *fading*, will be discussed in more detail in subsequent sections.

The impact of multipath on the received signal depends on whether the spread of time delays associated with the LOS and different multipath components is large or small relative to the inverse signal bandwidth. If this channel delay spread is small then the LOS and all multipath components are typically nonresolvable, leading to the narrowband fading model described in the next section. If the delay spread is large then the LOS and all multipath components are typically resolvable into some number of discrete components, leading to the wideband fading model of Section 3.3. Note that some of the discrete components in the wideband model are comprised of nonresolvable components. The delay spread is typically measured relative to the received signal component to which the demodulator is synchronized. Thus, for the time-invariant channel model of (3.10), if the demodulator synchronizes to the LOS signal component, which has the smallest delay τ_0, then the delay spread is a constant given by $T_m = \max_n \tau_n - \tau_0$. However, if the demodulator synchronizes to a multipath component with delay equal to the mean delay $\overline{\tau}$ then the delay spread is given by $T_m = \max_n |\tau_n - \overline{\tau}|$. In time-varying channels the multipath delays vary with time, so the delay spread T_m becomes a random variable. Moreover, some received multipath components have significantly lower power than others, so it's not clear how the delay associated with such components should be used in the characterization of delay spread. In particular, if the power of a multipath component is below the noise floor then it should not significantly contribute to the delay spread. These issues are typically dealt with by characterizing the delay spread relative to the channel power delay profile, defined in Section 3.3.1. Specifically, two common characterizations of channel delay spread, average delay spread and rms delay spread, are determined from the power delay profile. Other characterizations of delay spread, such as excees delay spread, the delay window, and the delay interval, are sometimes used as well [6, Chapter 5.4.1],[28, Chapter 6.7.1]. The exact characterization of delay spread is not that important for understanding the general impact of delay spread on multipath channels, as long as the characterization roughly measures the delay associated with significant multipath components. In our development below any reasonable characterization of delay spread T_m can be used, although we will typically use the rms delay spread. This is the most common characterization since, assuming the demodulator synchronizes to a signal component at the average delay spread, the rms delay spread is a good measure of the variation about this average. Channel delay spread is highly dependent on the propagation environment. In indoor channels delay spread typically ranges from 10 to 1000 nanoseconds, in suburbs it ranges from 200-2000 nanoseconds, and in urban areas it ranges from 1-30 microseconds [6].

3.2 Narrowband Fading Models

Suppose the delay spread T_m of a channel is small relative to the inverse signal bandwidth B of the transmitted signal, i.e. $T_m << B^{-1}$. As discussed above, the delay spread T_m for time-varying channels is usually characterized by the rms delay spread, but can also be characterized in other ways. Under most delay spread characterizations $T_m << B^{-1}$ implies that the delay associated with the ith multipath component $\tau_i \leq T_m \forall i$, so $u(t-\tau_i) \approx u(t) \forall i$ and we can rewrite (3.4) as

$$r(t) = \Re\left\{u(t)e^{j2\pi f_c t}\left(\sum_n \alpha_n(t)e^{-j\phi_n(t)}\right)\right\}. \tag{3.11}$$

Equation (3.11) differs from the original transmitted signal by the complex scale factor in parentheses. This scale factor is independent of the transmitted signal $s(t)$ or, equivalently, the baseband signal $u(t)$, as long as the

narrowband assumption $T_m << 1/B$ is satisfied. In order to characterize the random scale factor caused by the multipath we choose $s(t)$ to be an unmodulated carrier with random phase offset ϕ_0:

$$s(t) = \Re\{e^{j(2\pi f_c t + \phi_0)}\} = \cos(2\pi f_c t - \phi_0), \tag{3.12}$$

which is narrowband for any T_m.

With this assumption the received signal becomes

$$r(t) = \Re\left\{\left[\sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)}\right] e^{j2\pi f_c t}\right\} = r_I(t)\cos 2\pi f_c t + r_Q(t)\sin 2\pi f_c t, \tag{3.13}$$

where the in-phase and quadrature components are given by

$$r_I(t) = \sum_{n=1}^{N(t)} \alpha_n(t)\cos\phi_n(t), \tag{3.14}$$

and

$$r_Q(t) = \sum_{n=1}^{N(t)} \alpha_n(t)\sin\phi_n(t), \tag{3.15}$$

where the phase term

$$\phi_n(t) = 2\pi f_c \tau_n(t) - \phi_{D_n} - \phi_0 \tag{3.16}$$

now incorporates the phase offset ϕ_0 as well as the effects of delay and Doppler.

If $N(t)$ is large we can invoke the Central Limit Theorem and the fact that $\alpha_n(t)$ and $\phi_n(t)$ are stationary and ergodic to approximate $r_I(t)$ and $r_Q(t)$ as jointly Gaussian random processes. The Gaussian property is also true for small N if the $\alpha_n(t)$ are Rayleigh distributed and the $\phi_n(t)$ are uniformly distributed on $[-\pi, \pi]$. This happens when the nth multipath component results from a reflection cluster with a large number of nonresolvable multipath components [1].

3.2.1 Autocorrelation, Cross Correlation, and Power Spectral Density

We now derive the autocorrelation and cross correlation of the in-phase and quadrature received signal components $r_I(t)$ and $r_Q(t)$. Our derivations are based on some key assumptions which generally apply to propagation models without a dominant LOS component. Thus, these formulas are not typically valid when a dominant LOS component exists. We assume throughout this section that the amplitude $\alpha_n(t)$, multipath delay $\tau_n(t)$ and Doppler frequency $f_{D_n}(t)$ are changing slowly enough such that they are constant over the time intervals of interest: $\alpha_n(t) \approx \alpha_n$, $\tau_n(t) \approx \tau_n$, and $f_{D_n}(t) \approx f_{D_n}$. This will be true when each of the resolvable multipath components is associated with a single reflector. With this assumption the Doppler phase shift is[3] $\phi_{D_n}(t) = \int_t 2\pi f_{D_n} dt = 2\pi f_{D_n} t$, and the phase of the nth multipath component becomes $\phi_n(t) = 2\pi f_c \tau_n - 2\pi f_{D_n} t - \phi_0$.

We now make a *key* assumption: we assume that for the nth multipath component the term $2\pi f_c \tau_n$ in $\phi_n(t)$ changes rapidly relative to all other phase terms in this expression. This is a reasonable assumption since f_c is large and hence the term $2\pi f_c \tau_n$ can go through a 360 degree rotation for a small change in multipath delay τ_n. Under this assumption $\phi_n(t)$ is uniformly distributed on $[-\pi, \pi]$. Thus

$$\mathrm{E}[r_I(t)] = \mathrm{E}[\sum_n \alpha_n \cos\phi_n(t)] = \sum_n \mathrm{E}[\alpha_n]\mathrm{E}[\cos\phi_n(t)] = 0, \tag{3.17}$$

[3] We assume a Doppler phase shift at $t = 0$ of zero for simplicity, since this phase offset will not affect the analysis.

where the second equality follows from the independence of α_n and ϕ_n and the last equality follows from the uniform distribution on ϕ_n. Similarly we can show that $\mathrm{E}[r_Q(t)] = 0$. Thus, the received signal also has $\mathrm{E}[r(t)] = 0$, i.e. it is a zero-mean Gaussian process. When there is a dominant LOS component in the channel the phase of the received signal is dominated by the phase of the LOS component, which can be determined at the receiver, so the assumption of a random uniform phase no longer holds.

Consider now the autocorrelation of the in-phase and quadrature components. Using the independence of α_n and ϕ_n, the independence of ϕ_n and ϕ_m, $n \neq m$, and the uniform distribution of ϕ_n we get that

$$\begin{aligned}
\mathrm{E}[r_I(t)r_Q(t)] &= \mathrm{E}\left[\sum_n \alpha_n \cos\phi_n(t) \sum_m \alpha_m \sin\phi_m(t)\right] \\
&= \sum_n \sum_m \mathrm{E}[\alpha_n\alpha_m]\mathrm{E}[\cos\phi_n(t)\sin\phi_m(t)] \\
&= \sum_n \mathrm{E}[\alpha_n^2]\mathrm{E}[\cos\phi_n(t)\sin\phi_n(t)] \\
&= 0. \qquad (3.18)
\end{aligned}$$

Thus, $r_I(t)$ and $r_Q(t)$ are uncorrelated and, since they are jointly Gaussian processes, this means they are independent.

Following a similar derivation as in (3.18) we obtain the autocorrelation of $r_I(t)$ as

$$A_{r_I}(t,\tau) = \mathrm{E}[r_I(t)r_I(t+\tau)] = \sum_n \mathrm{E}[\alpha_n^2]\mathrm{E}[\cos\phi_n(t)\cos\phi_n(t+\tau)]. \qquad (3.19)$$

Now making the substitution $\phi_n(t) = 2\pi f_c\tau_n - 2\pi f_{D_n}t - \phi_0$ and $\phi_n(t+\tau) = 2\pi f_c\tau_n - 2\pi f_{D_n}(t+\tau) - \phi_0$ we get

$$\mathrm{E}[\cos\phi_n(t)\cos\phi_n(t+\tau)] = .5\mathrm{E}[\cos 2\pi f_{D_n}\tau] + .5\mathrm{E}[\cos(4\pi f_c\tau_n + -4\pi f_{D_n}t - 2\pi f_{D_n}\tau - 2\phi_0)]. \qquad (3.20)$$

Since $2\pi f_c\tau_n$ changes rapidly relative to all other phase terms and is uniformly distributed, the second expectation term in (3.20) goes to zero, and thus

$$A_{r_I}(t,\tau) = .5\sum_n \mathrm{E}[\alpha_n^2]\mathrm{E}[\cos(2\pi f_{D_n}\tau)] = .5\sum_n \mathrm{E}[\alpha_n^2]\cos(2\pi v\tau\cos\theta_n/\lambda), \qquad (3.21)$$

since $f_{D_n} = v\cos\theta_n/\lambda$ is assumed fixed. Note that $A_{r_I}(t,\tau)$ depends only on τ, $A_{r_I}(t,\tau) = A_{r_I}(\tau)$, and thus $r_I(t)$ is a wide-sense stationary (WSS) random process.

Using a similar derivation we can show that the quadrature component is also WSS with autocorrelation $A_{r_Q}(\tau) = A_{r_I}(\tau)$. In addition, the cross correlation between the in-phase and quadrature components depends only on the time difference τ and is given by

$$A_{r_I,r_Q}(t,\tau) = A_{r_I,r_Q}(\tau) = \mathrm{E}[r_I(t)r_Q(t+\tau)] = -.5\sum_n \mathrm{E}[\alpha_n^2]\sin(2\pi v\tau\cos\theta_n/\lambda) = -\mathrm{E}[r_Q(t)r_I(t+\tau)]. \qquad (3.22)$$

Using these results we can show that the received signal $r(t) = r_I(t)\cos(2\pi f_c t) + r_Q(t)\sin(2\pi f_c t)$ is also WSS with autocorrelation

$$A_r(\tau) = \mathrm{E}[r(t)r(t+\tau)] = A_{r_I}(\tau)\cos(2\pi f_c\tau) + A_{r_I,r_Q}(\tau)\sin(2\pi f_c\tau). \qquad (3.23)$$

In order to further simplify (3.21) and (3.22), we must make additional assumptions about the propagation environment. We will focus on the **uniform scattering environment** introduced by Clarke [4] and further developed by Jakes [Chapter 1][5]. In this model, the channel consists of many scatterers densely packed with respect to angle, as shown in Fig. 3.4. Thus, we assume N multipath components with angle of arrival $\theta_n = n\Delta\theta$, where $\Delta\theta = 2\pi/N$. We also assume that each multipath component has the same received power, so $\mathrm{E}[\alpha_n^2] = 2P_r/N$, where P_r is the total received power. Then (3.21) becomes

$$A_{r_I}(\tau) = \frac{P_r}{N}\sum_{n=1}^{N}\cos(2\pi v\tau\cos n\Delta\theta/\lambda). \tag{3.24}$$

Now making the substitution $N = 2\pi/\Delta\theta$ yields

$$A_{r_I}(\tau) = \frac{P_r}{2\pi}\sum_{n=1}^{N}\cos(2\pi v\tau\cos n\Delta\theta/\lambda)\Delta\theta. \tag{3.25}$$

We now take the limit as the number of scatterers grows to infinity, which corresponds to uniform scattering from all directions. Then $N \to \infty$, $\Delta\theta \to 0$, and the summation in (3.25) becomes an integral:

$$A_{r_I}(\tau) = \frac{P_r}{2\pi}\int \cos(2\pi v\tau\cos\theta/\lambda)d\theta = P_r J_0(2\pi f_D\tau), \tag{3.26}$$

where

$$J_0(x) = \frac{1}{\pi}\int_0^{\pi} e^{-jx\cos\theta}d\theta$$

is a Bessel function of the 0th order[4]. Similarly, for this uniform scattering environment,

$$A_{r_I,r_Q}(\tau) = \frac{P_r}{2\pi}\int \sin(2\pi v\tau\cos\theta/\lambda)d\theta = 0. \tag{3.27}$$

A plot of $J_0(2\pi f_D\tau)$ is shown in Figure 3.5. There are several interesting observations from this plot. First we see that the autocorrelation is zero for $f_D\tau \approx .4$ or, equivalently, for $v\tau \approx .4\lambda$. Thus, the signal decorrelates over a distance of approximately one half wavelength, under the uniform θ_n assumption. This approximation is commonly used as a rule of thumb to determine many system parameters of interest. For example, we will see in Chapter 7 that obtaining independent fading paths can be exploited by antenna diversity to remove some of the negative effects of fading. The antenna spacing must be such that each antenna receives an independent fading path and therefore, based on our analysis here, an antenna spacing of $.4\lambda$ should be used. Another interesting characteristic of this plot is that the signal recorrelates after it becomes uncorrelated. Thus, we cannot assume that the signal remains independent from its initial value at $d = 0$ for separation distances greater than $.4\lambda$. As a result, a Markov model is not completely accurate for Rayleigh fading, because of this recorrelation property. However, in many system analyses a correlation below .5 does not significantly degrade performance relative to uncorrelated fading [8, Chapter 9.6.5]. For such studies the fading process can be modeled as Markov by assuming that once the correlation is close to zero, i.e. the separation distance is greater than a half wavelength, the signal remains decorrelated at all larger distances.

[4]Note that (3.26) can also be derived by assuming $2\pi v\tau\cos\theta_n/\lambda$ in (3.21) and (3.22) is random with θ_n uniformly distributed, and then taking expectation with respect to θ_n. However, based on the underlying physical model, θ_n can only be uniformly distributed in a dense scattering environment. So the derivations are equivalent.

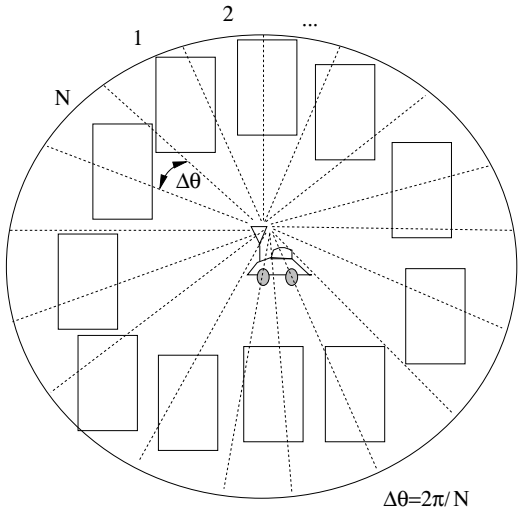


Figure 3.4: Dense Scattering Environment

The power spectral densities (PSDs) of $r_I(t)$ and $r_Q(t)$, denoted by $S_{r_I}(f)$ and $S_{r_Q}(f)$, respectively, are obtained by taking the Fourier transform of their respective autocorrelation functions relative to the delay parameter τ. Since these autocorrelation functions are equal, so are the PSDs. Thus

$$S_{r_I}(f) = S_{r_Q}(f) = \mathcal{F}[A_{r_I}(\tau)] = \begin{cases} \frac{P_r}{2\pi f_D} \frac{1}{\sqrt{1-(f/f_D)^2}} & |f| \leq f_D \\ 0 & \text{else} \end{cases} \tag{3.28}$$

This PSD is shown in Figure 3.6.

To obtain the PSD of the received signal $r(t)$ under uniform scattering we use (3.23) with $A_{r_I,r_Q}(\tau) = 0$, (3.28), and simple properties of the Fourier transform to obtain

$$S_r(f) = \mathcal{F}[A_r(\tau)] = .25[S_{r_I}(f-f_c) + S_{r_I}(f+f_c)] = \begin{cases} \frac{P_r}{4\pi f_D} \frac{1}{\sqrt{1-\left(\frac{|f-f_c|}{f_D}\right)^2}} & |f-f_c| \leq f_D \\ 0 & \text{else} \end{cases}, \tag{3.29}$$

Note that this PSD integrates to P_r, the total received power.

Since the PSD models the power density associated with multipath components as a function of their Doppler frequency, it can be viewed as the distribution (pdf) of the random frequency due to Doppler associated with multipath. We see from Figure 3.6 that the PSD $S_{r_i}(f)$ goes to infinity at $f = \pm f_D$ and, consequently, the PSD $S_r(f)$ goes to infinity at $f = \pm f_c \pm f_D$. This will not be true in practice, since the uniform scattering model is just an approximation, but for environments with dense scatterers the PSD will generally be maximized at frequencies close to the maximum Doppler frequency. The intuition for this behavior comes from the nature of the cosine function and the fact that under our assumptions the PSD corresponds to the pdf of the random Doppler frequency $f_D(\theta)$. To see this, note that the uniform scattering assumption is based on many scattered paths arriving uniformly from all angles with the same average power. Thus, θ for a randomly selected path can be regarded as a uniform random variable on $[0, 2\pi]$. The distribution $p_{f_\theta}(f)$ of the random Doppler frequency $f(\theta)$ can then be obtained from the distribution of θ. By definition, $p_{f_\theta}(f)$ is proportional to the density of scatterers at Doppler frequency f. Hence, $S_{r_I}(f)$ is also proportional to this density, and we can characterize the PSD from the pdf $p_{f_\theta}(f)$. For this characterization, in Figure 3.7 we plot $f_D(\theta) = f_D \cos(\theta) = v/\lambda \cos(\theta)$ along with a dotted line straight-line segment approximation $\underline{f}_D(\theta)$ to $f_D(\theta)$. On the right in this figure we plot the PSD $S_{r_i}(f)$ along with a dotted

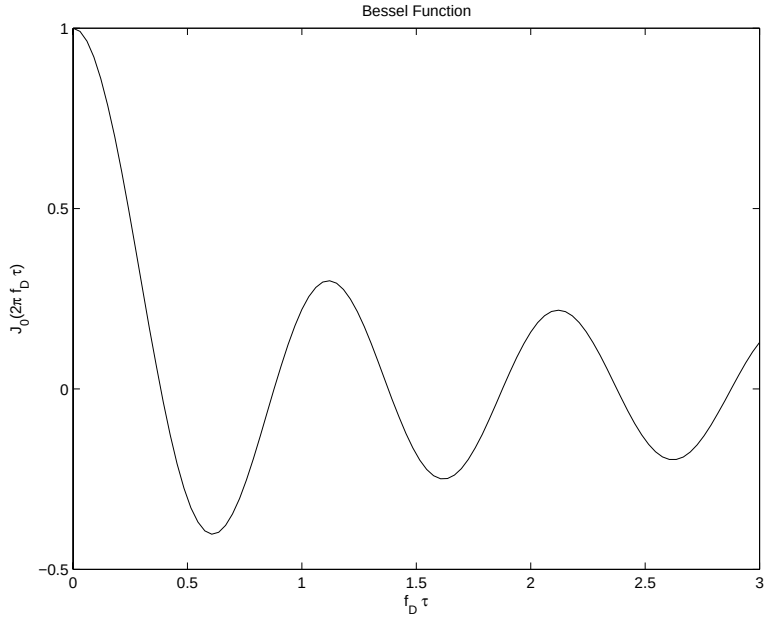


Figure 3.5: Bessel Function versus $f_d\tau$

line straight line segment approximation to it $\underline{S}_{r_i}(f)$, which corresponds to the Doppler approximation $\underline{f}_D(\theta)$. We see that $\cos(\theta) \approx \pm 1$ for a relatively large range of θ values. Thus, multipath components with angles of arrival in this range of values have Doppler frequency $f_D(\theta) \approx \pm f_D$, so the power associated with all of these multipath components will add together in the PSD at $f \approx f_D$. This is shown in our approximation by the fact that the segments where $\underline{f}_D(\theta) = \pm f_D$ on the left lead to delta functions at $\pm f_D$ in the pdf approximation $\underline{S}_{r_i}(f)$ on the right. The segments where $\underline{f}_D(\theta)$ has uniform slope on the left lead to the flat part of $\underline{S}_{r_i}(f)$ on the right, since there is one multipath component contributing power at each angular increment. Formulas for the autocorrelation and PSD in nonuniform scattering, corresponding to more typical microcell and indoor environments, can be found in [5, Chapter 1], [11, Chapter 2].

The PSD is useful in constructing simulations for the fading process. A common method for simulating the envelope of a narrowband fading process is to pass two independent white Gaussian noise sources with PSD $N_0/2$ through lowpass filters with frequency response $H(f)$ that satisfies

$$S_{r_I}(f) = S_{r_Q}(f) = \frac{N_0}{2}|H(f)|^2. \tag{3.30}$$

The filter outputs then correspond to the in-phase and quadrature components of the narrowband fading process with PSDs $S_{r_I}(f)$ and $S_{r_Q}(f)$. A similar procedure using discrete filters can be used to generate discrete fading processes. Most communication simulation packages (e.g. Matlab, COSSAP) have standard modules that simulate narrowband fading based on this method. More details on this simulation method, as well as alternative methods, can be found in [11, 6, 7].

We have now completed our model for the three characteristics of power versus distance exhibited in narrowband wireless channels. These characteristics are illustrated in Figure 3.8, adding narrowband fading to the path loss and shadowing models developed in Chapter 2. In this figure we see the decrease in signal power due to path loss decreasing as d^γ with γ the path loss exponent, the more rapid variations due to shadowing which change on the order of the decorrelation distance X_c, and the very rapid variations due to multipath fading which change on the order of half the signal wavelength. If we blow up a small segment of this figure over distances where path loss

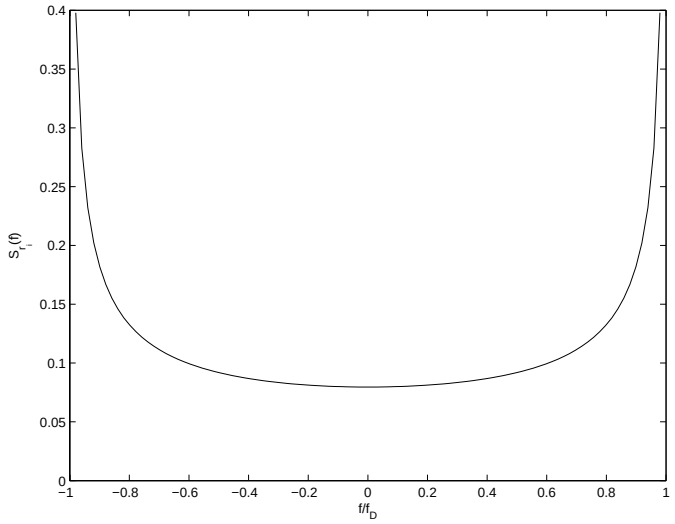


Figure 3.6: In-Phase and Quadrature PSD: $S_{r_I}(f) = S_{r_Q}(f)$

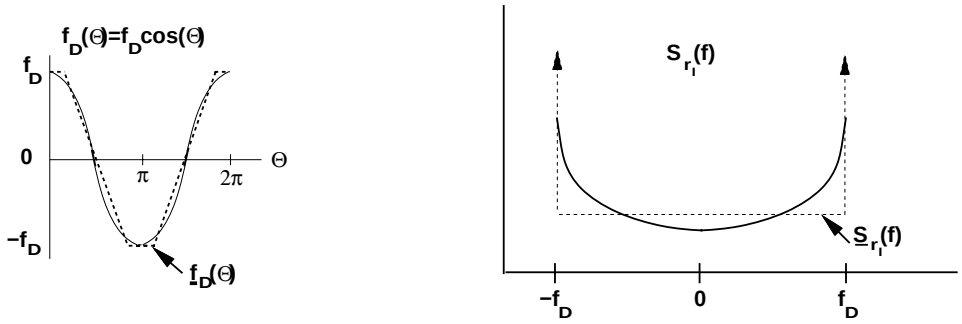


Figure 3.7: Cosine and PSD Approximation by Straight Line Segments

and shadowing are constant we obtain Figure 3.9, where we show dB fluctuation in received power versus linear distance $d = vt$ (not log distance). In this figure the average received power P_r is normalized to 0 dBm. A mobile receiver traveling at fixed velocity v would experience the received power variations over time illustrated in this figure.

3.2.2 Envelope and Power Distributions

For any two Gaussian random variables X and Y, both with mean zero and equal variance σ^2, it can be shown that $Z = \sqrt{X^2 + Y^2}$ is Rayleigh-distributed and Z^2 is exponentially distributed. We saw above that for $\phi_n(t)$ uniformly distributed, r_I and r_Q are both zero-mean Gaussian random variables. If we assume a variance of σ^2 for both in-phase and quadrature components then the signal envelope

$$z(t) = |r(t)| = \sqrt{r_I^2(t) + r_Q^2(t)} \tag{3.31}$$

is Rayleigh-distributed with distribution

$$p_Z(z) = \frac{2z}{P_r}\exp[-z^2/P_r] = \frac{z}{\sigma^2}\exp[-z^2/(2\sigma^2)], \quad x \geq 0, \tag{3.32}$$

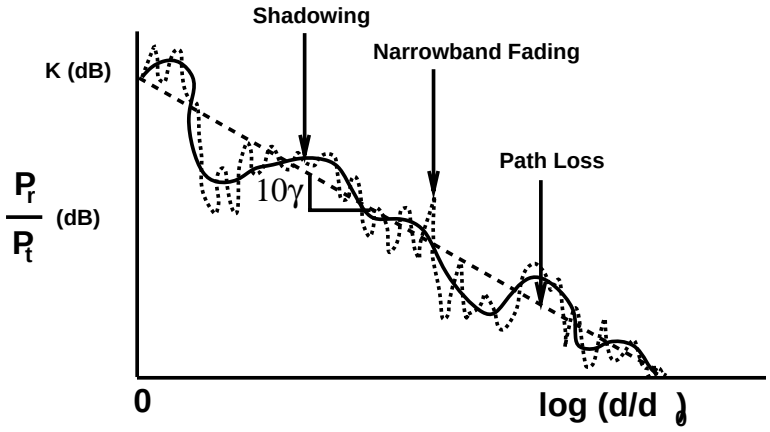


Figure 3.8: Combined Path Loss, Shadowing, and Narrowband Fading.

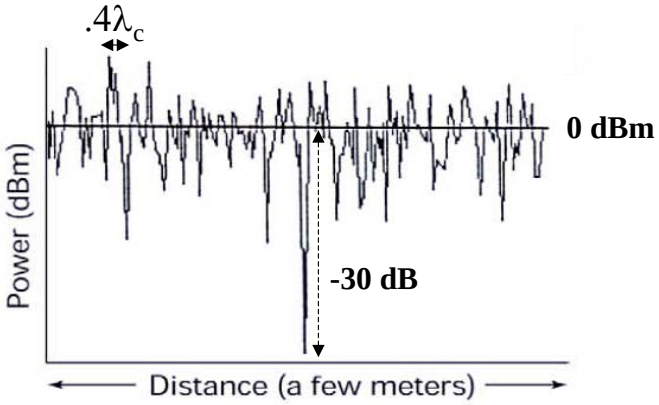


Figure 3.9: Narrowband Fading.

where $P_r = \sum_n E[\alpha_n^2] = 2\sigma^2$ is the average received signal power of the signal, i.e. the received power based on path loss and shadowing alone.

We obtain the power distribution by making the change of variables $z^2(t) = |r(t)|^2$ in (3.32) to obtain

$$p_{Z^2}(x) = \frac{1}{P_r} e^{-x/P_r} = \frac{1}{2\sigma^2} e^{-x/(2\sigma^2)}, \quad x \geq 0. \tag{3.33}$$

Thus, the received signal power is exponentially distributed with mean $2\sigma^2$. The complex lowpass equivalent signal for $r(t)$ is given by $r_{LP}(t) = r_I(t) + jr_Q(t)$ which has phase $\theta = \arctan(r_Q(t)/r_I(t))$. For $r_I(t)$ and $r_Q(t)$ uncorrelated Gaussian random variables we can show that θ is uniformly distributed and independent of $|r_{LP}|$. So $r(t)$ has a Rayleigh-distributed amplitude and uniform phase, and the two are mutually independent.

Example 3.2: Consider a channel with Rayleigh fading and average received power $P_r = 20$ dBm. Find the probability that the received power is below 10 dBm.
Solution. We have $P_r = 20$ dBm =100 mW. We want to find the probability that $Z^2 < 10$ dBm =10 mW. Thus

$$p(Z^2 < 10) = \int_0^{10} \frac{1}{100} e^{-x/100} dx = .095.$$

If the channel has a fixed LOS component then $r_I(t)$ and $r_Q(t)$ are not zero-mean. In this case the received signal equals the superposition of a complex Gaussian component and a LOS component. The signal envelope in this case can be shown to have a Rician distribution [9], given by

$$p_Z(z) = \frac{z}{\sigma^2} \exp\left[\frac{-(z^2+s^2)}{2\sigma^2}\right] I_0\left(\frac{zs}{\sigma^2}\right), \quad z \geq 0, \tag{3.34}$$

where $2\sigma^2 = \sum_{n,n\neq 0} \mathrm{E}[\alpha_n^2]$ is the average power in the non-LOS multipath components and $s^2 = \alpha_0^2$ is the power in the LOS component. The function I_0 is the modified Bessel function of 0th order. The average received power in the Rician fading is given by

$$P_r = \int_0^\infty z^2 p_Z(z) dx = s^2 + 2\sigma^2. \tag{3.35}$$

The Rician distribution is often described in terms of a fading parameter K, defined by

$$K = \frac{s^2}{2\sigma^2}. \tag{3.36}$$

Thus, K is the ratio of the power in the LOS component to the power in the other (non-LOS) multipath components. For $K = 0$ we have Rayleigh fading, and for $K = \infty$ we have no fading, i.e. a channel with no multipath and only a LOS component. The fading parameter K is therefore a measure of the severity of the fading: a small K implies severe fading, a large K implies more mild fading. Making the substitution $s^2 = KP/(K+1)$ and $2\sigma^2 = P/(K+1)$ we can write the Rician distribution in terms of K and P_r as

$$p_Z(z) = \frac{2z(K+1)}{P_r} \exp\left[-K - \frac{(K+1)z^2}{P_r}\right] I_0\left(2z\sqrt{\frac{K(K+1)}{P_r}}\right), \quad z \geq 0. \tag{3.37}$$

Both the Rayleigh and Rician distributions can be obtained by using mathematics to capture the underlying physical properties of the channel models [1, 9]. However, some experimental data does not fit well into either of

these distributions. Thus, a more general fading distribution was developed whose parameters can be adjusted to fit a variety of empirical measurements. This distribution is called the Nakagami fading distribution, and is given by

$$p_Z(z) = \frac{2m^m z^{2m-1}}{\Gamma(m)P_r^m}\exp\left[\frac{-mz^2}{P_r}\right], \quad m \geq .5, \tag{3.38}$$

where P_r is the average received power and $\Gamma(\cdot)$ is the Gamma function. The Nakagami distribution is parameterized by P_r and the fading parameter m. For $m = 1$ the distribution in (3.38) reduces to Rayleigh fading. For $m = (K+1)^2/(2K+1)$ the distribution in (3.38) is approximately Rician fading with parameter K. For $m = \infty$ there is no fading: P_r is a constant. Thus, the Nakagami distribution can model Rayleigh and Rician distributions, as well as more general ones. Note that some empirical measurements support values of the m parameter less than one, in which case the Nakagami fading causes more severe performance degradation than Rayleigh fading. The power distribution for Nakagami fading, obtained by a change of variables, is given by

$$p_{Z^2}(x) = \left(\frac{m}{P_r}\right)^m \frac{x^{m-1}}{\Gamma(m)}\exp\left(\frac{-mx}{P_r}\right). \tag{3.39}$$

3.2.3 Level Crossing Rate and Average Fade Duration

The envelope level crossing rate L_Z is defined as the expected rate (in crossings per second) at which the signal envelope crosses the level Z in the downward direction. Obtaining L_Z requires the joint distribution of the signal envelope $z = |r|$ and its derivative with respect to time $\dot{z}$, $p(z, \dot{z})$. We now derive L_Z based on this joint distribution.

Consider the fading process shown in Figure 3.10. The expected amount of time the signal envelope spends in the interval $(Z, Z + dz)$ with envelope slope in the range $[\dot{z}, \dot{z} + d\dot{z}]$ over time duration dt is $A = p(Z, \dot{z})dzd\dot{z}dt$. The time required to cross from Z to $Z + dz$ once for a given envelope slope $\dot{z}$ is $B = dz/\dot{z}$. The ratio $A/B = \dot{z}p(Z, \dot{z})d\dot{z}dt$ is the expected number of crossings of the envelope z within the interval $(Z, Z + dz)$ for a given envelope slope $\dot{z}$ over time duration dt. The expected number of crossings of the envelope level Z for slopes between $\dot{z}$ and $\dot{z} + d\dot{z}$ in a time interval $[0, T]$ in the downward direction is thus

$$\int_0^T \dot{z}p(Z, \dot{z})d\dot{z}dt = \dot{z}p(Z, \dot{z})d\dot{z}T. \tag{3.40}$$

So the expected number of crossings of the envelope level Z with negative slope over the interval $[0, T]$ is

$$N_Z = T\int_{-\infty}^0 \dot{z}p(Z, \dot{z})d\dot{z}. \tag{3.41}$$

Finally, the expected number of crossings of the envelope level Z per second, i.e. the level crossing rate, is

$$L_Z = \frac{N_Z}{T} = \int -\infty^0 \dot{z}p(Z, \dot{z})d\dot{z}. \tag{3.42}$$

Note that this is a general result that applies for any random process.

The joint pdf of z and $\dot{z}$ for Rician fading was derived in [9] and can also be found in [11]. The level crossing rate for Rician fading is then obtained by using this pdf in (3.42), and is given by

$$L_Z = \sqrt{2\pi(K+1)}f_D\rho e^{-K-(K+1)\rho^2}I_0(2\rho\sqrt{K(K+1)}), \tag{3.43}$$

where $\rho = Z/\sqrt{P_r}$. It is easily shown that the rate at which the received signal power crosses a threshold value γ_0 obeys the same formula (3.43) with $\rho = \sqrt{\gamma_0/P_r}$. For Rayleigh fading ($K = 0$) the level crossing rate simplifies to

$$L_Z = \sqrt{2\pi}f_D\rho e^{-\rho^2}, \tag{3.44}$$

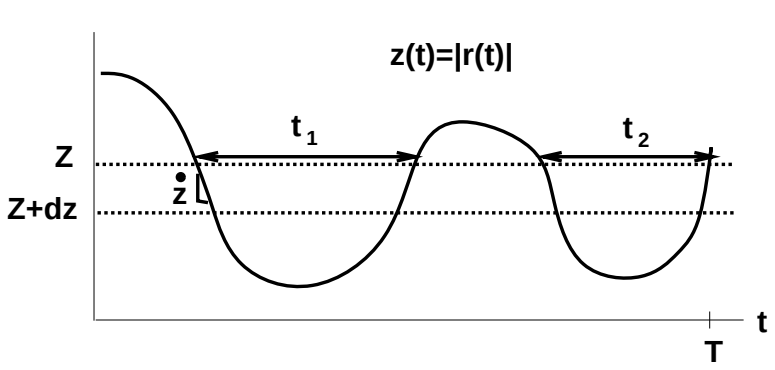


Figure 3.10: Level Crossing Rate and Fade Duration for Fading Process.

where $\rho = Z/\sqrt{P_r}$.

We define the average signal fade duration as the average time that the signal envelope stays below a given target level Z. This target level is often obtained from the signal amplitude or power level required for a given performance metric like bit error rate. Let t_i denote the duration of the ith fade below level Z over a time interval $[0, T]$, as illustrated in Figure 3.10. Thus t_i equals the length of time that the signal envelope stays below Z on its ith crossing. Since $z(t)$ is stationary and ergodic, for T sufficiently large we have

$$p(z(t) < Z) = \frac{1}{T} \sum_i t_i. \tag{3.45}$$

Thus, for T sufficiently large the average fade duration is

$$\overline{t}_Z = \frac{1}{TL_Z} \sum_{i=1}^{L_Z T} t_i \approx \frac{p(z(t) < Z)}{L_Z}. \tag{3.46}$$

Using the Rayleigh distribution for $p(z(t) < Z)$ yields

$$\overline{t}_Z = \frac{e^{\rho^2} - 1}{\rho f_D \sqrt{2\pi}} \tag{3.47}$$

with $\rho = Z/\sqrt{P_r}$. Note that (3.47) is the average fade duration for the signal envelope (amplitude) level with Z the target amplitude and $\sqrt{P_r}$ the average envelope level. By a change of variables it is easily shown that (3.47) also yields the average fade duration for the signal power level with $\rho = \sqrt{P_0/P_r}$, where P_0 is the target power level and P_r is the average power level. Note that average fade duration decreases with Doppler, since as a channel changes more quickly it remains below a given fade level for a shorter period of time. The average fade duration also generally increases with ρ for $\rho >> 1$. That is because as the target level increases relative to the average, the signal is more likely to be below the target. The average fade duration for Rician fading is more difficult to compute, it can be found in [11, Chapter 1.4].

The average fade duration indicates the number of bits or symbols affected by a deep fade. Specifically, consider an uncoded system with bit time T_b. Suppose the probability of bit error is high when $z < Z$. Then if $T_b \approx \overline{t}_Z$, the system will likely experience single error events, where bits that are received in error have the previous and subsequent bits received correctly (since $z > Z$ for these bits). On the other hand, if $T_b << \overline{t}_Z$ then many subsequent bits are received with $z < Z$, so large bursts of errors are likely. Finally, if $T_b >> \overline{t}_Z$ the fading is averaged out over a bit time in the demodulator, so the fading can be neglected. These issues will be explored in more detail in Chapter 8, when we consider coding and interleaving.

Example 3.3:

Consider a voice system with acceptable BER when the received signal power is at or above half its average value. If the BER is below its acceptable level for more than 120 ms, users will turn off their phone. Find the range of Doppler values in a Rayleigh fading channel such that the average time duration when users have unacceptable voice quality is less than $t = 60$ ms.

Solution: The target received signal value is half the average, so $P_0 = .5P_r$ and thus $\rho = \sqrt{.5}$. We require

$$\bar{t}_Z = \frac{e^{.5} - 1}{f_D\sqrt{\pi}} \leq t = .060$$

and thus $f_D \geq (e-1)/(.060\sqrt{2\pi}) = 6.1$ Hz.

3.2.4 Finite State Markov Channels

The complex mathematical characterization of flat fading described in the previous subsections can be difficult to incorporate into wireless performance analysis such as the packet error probability. Therefore, simpler models that capture the main features of flat fading channels are needed for these analytical calculations. One such model is a finite state Markov channel (FSMC). In this model fading is approximated as a discrete-time Markov process with time discretized to a given interval T (typically the symbol period). Specifically, the set of all possible fading gains is modeled as a set of finite channel states. The channel varies over these states at each interval T according to a set of Markov transition probabilities. FSMCs have been used to approximate both mathematical and experimental fading models, including satellite channels [13], indoor channels [14], Rayleigh fading channels [15, 19], Ricean fading channels [20], and Nakagami-m fading channels [17]. They have also been used for system design and system performance analysis in [18, 19]. First-order FSMC models have been shown to be deficient in computing performance analysis, so higher order models are generally used. The FSMC models for fading typically model amplitude variations only, although there has been some work on FSMC models for phase in fading [21] or phase-noisy channels [22].

A detailed FSMC model for Rayleigh fading was developed in [15]. In this model the time-varying SNR associated with the Rayleigh fading, γ, lies in the range $0 \leq \gamma \leq \infty$. The FSMC model discretizes this fading range into regions so that the jth region R_j is defined as $R_j = \gamma : A_j \leq \gamma < A_{j+1}$, where the region boundaries $\{A_j\}$ and the total number of fade regions are parameters of the model. This model assumes that γ stays within the same region over time interval T and can only transition to the same region or adjacent regions at time $T + 1$. Thus, given that the channel is in state R_j at time T, at the next time interval the channel can only transition to R_{j-1}, R_j, or R_{j+1}, a reasonable assumption when $f_D T$ is small. Under this assumption the transition probabilities between regions are derived in [15] as

$$p_{j,j+1} = \frac{N_{j+1}T_s}{\pi_j}, \quad p_{j,j-1} = \frac{N_j T_s}{\pi_j}, \quad p_{j,j} = 1 - p_{j,j+1} - p_{j,j-1}, \tag{3.48}$$

where N_j is the level-crossing rate at A_j and π_j is the steady-state distribution corresponding to the jth region: $\pi_j = p(\gamma \in R_j) = p(A_j \leq \gamma < A_{j+1})$.

3.3 Wideband Fading Models

When the signal is not narrowband we get another form of distortion due to the multipath delay spread. In this case a short transmitted pulse of duration T will result in a received signal that is of duration $T + T_m$, where T_m is the multipath delay spread. Thus, the duration of the received signal may be significantly increased. This is illustrated in Figure 3.11. In this figure, a pulse of width T is transmitted over a multipath channel. As discussed in Chapter 5, linear modulation consists of a train of pulses where each pulse carries information in its amplitude and/or phase corresponding to a data bit or symbol[5]. If the multipath delay spread $T_m << T$ then the multipath components are received roughly on top of one another, as shown on the upper right of the figure. The resulting constructive and destructive interference causes narrowband fading of the pulse, but there is little time-spreading of the pulse and therefore little interference with a subsequently transmitted pulse. On the other hand, if the multipath delay spread $T_m >> T$, then each of the different multipath components can be resolved, as shown in the lower right of the figure. However, these multipath components interfere with subsequently transmitted pulses. This effect is called intersymbol interference (ISI).

There are several techniques to mitigate the distortion due to multipath delay spread, including equalization, multicarrier modulation, and spread spectrum, which are discussed in Chapters 11-13. ISI migitation is not necessary if $T >> T_m$, but this can place significant constraints on data rate. Multicarrier modulation and spread spectrum actually change the characteristics of the transmitted signal to mostly avoid intersymbol interference, however they still experience multipath distortion due to frequency-selective fading, which is described in Section 3.3.2.

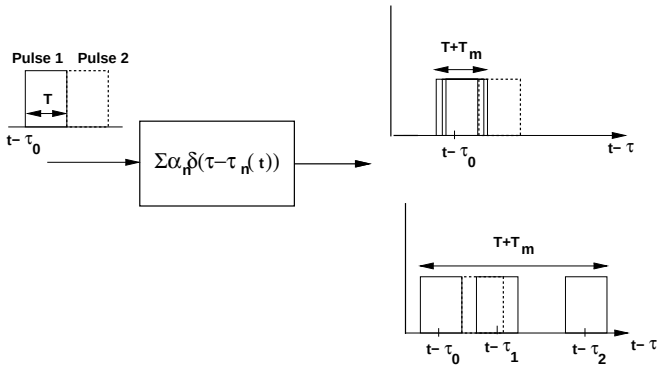


Figure 3.11: Multipath Resolution.

The difference between wideband and narrowband fading models is that as the transmit signal bandwidth B increases so that $T_m \approx B^{-1}$, the approximation $u(t - \tau_n(t)) \approx u(t)$ is no longer valid. Thus, the received signal is a sum of copies of the original signal, where each copy is delayed in time by τ_n and shifted in phase by $\phi_n(t)$. The signal copies will combine destructively when their phase terms differ significantly, and will distort the direct path signal when $u(t - \tau_n)$ differs from $u(t)$.

Although the approximation in (3.11) no longer applies when the signal bandwidth is large relative to the inverse of the multipath delay spread, if the number of multipath components is large and the phase of each component is uniformly distributed then the received signal will still be a zero-mean complex Gaussian process with a Rayleigh-distributed envelope. However, wideband fading differs from narrowband fading in terms of the resolution of the different multipath components. Specifically, for narrowband signals, the multipath components have a time resolution that is less than the inverse of the signal bandwidth, so the multipath components characterized

[5]Linear modulation typically uses nonsquare pulse shapes for bandwidth efficiency, as discussed in Chapter 5.4

in Equation (3.6) combine at the receiver to yield the original transmitted signal with amplitude and phase characterized by random processes. These random processes are characterized by their autocorrelation or PSD, and their instantaneous distributions, as discussed in Section 3.2. However, with wideband signals, the received signal experiences distortion due to the delay spread of the different multipath components, so the received signal can no longer be characterized by just the amplitude and phase random processes. The effect of multipath on wideband signals must therefore take into account both the multipath delay spread and the time-variations associated with the channel.

The starting point for characterizing wideband channels is the equivalent lowpass time-varying channel impulse response $c(\tau, t)$. Let us first assume that $c(\tau, t)$ is a continuous[6] deterministic function of τ and t. Recall that τ represents the impulse response associated with a given multipath delay, while t represents time variations. We can take the Fourier transform of $c(\tau, t)$ with respect to t as

$$S_c(\tau, \rho) = \int_{-\infty}^{\infty} c(\tau, t) e^{-j2\pi\rho t} dt. \tag{3.49}$$

We call $S_c(\tau, \rho)$ the **deterministic scattering function** of the lowpass equivalent channel impulse response $c(\tau, t)$. Since it is the Fourier transform of $c(\tau, t)$ with respect to the time variation parameter t, the deterministic scattering function $S_c(\tau, \rho)$ captures the Doppler characteristics of the channel via the frequency parameter ρ.

In general the time-varying channel impulse response $c(\tau, t)$ given by (3.6) is random instead of deterministic due to the random amplitudes, phases, and delays of the random number of multipath components. In this case we must characterize it statistically or via measurements. As long as the number of multipath components is large, we can invoke the Central Limit Theorem to assume that $c(\tau, t)$ is a complex Gaussian process, so its statistical characterization is fully known from the mean, autocorrelation, and cross-correlation of its in-phase and quadrature components. As in the narrowband case, we assume that the phase of each multipath component is uniformly distributed. Thus, the in-phase and quadrature components of $c(\tau, t)$ are independent Gaussian processes with the same autocorrelation, a mean of zero, and a cross-correlation of zero. The same statistics hold for the in-phase and quadrature components if the channel contains only a small number of multipath rays as long as each ray has a Rayleigh-distributed amplitude and uniform phase. Note that this model does not hold when the channel has a dominant LOS component.

The statistical characterization of $c(\tau, t)$ is thus determined by its **autocorrelation function**, defined as

$$A_c(\tau_1, \tau_2; t, \Delta t) = E[c^*(\tau_1; t) c(\tau_2; t + \Delta t)]. \tag{3.50}$$

Most channels in practice are wide-sense stationary (WSS), such that the joint statistics of a channel measured at two different times t and $t + \Delta t$ depends only on the time difference Δt. For wide-sense stationary channels, the autocorrelation of the corresponding bandpass channel $h(\tau, t) = \Re\{c(\tau, t) e^{j2\pi f_c t}\}$ can be obtained [16] from $A_c(\tau_1, \tau_2; t, \Delta t)$ as[7] $A_h(\tau_1, \tau_2; t, \Delta t) = .5\Re\{A_c(\tau_1, \tau_2; t, \Delta t) e^{j2\pi f_c \Delta t}\}$. We will assume that our channel model is WSS, in which case the autocorrelation becomes indepedent of t:

$$A_c(\tau_1, \tau_2; \Delta t) = E[c^*(\tau_1; t) c(\tau_2; t + \Delta t)]. \tag{3.51}$$

Moreover, in practice the channel response associated with a given multipath component of delay τ_1 is uncorrelated with the response associated with a multipath component at a different delay $\tau_2 \neq \tau_1$, since the two components are caused by different scatterers. We say that such a channel has uncorrelated scattering (US). We abbreviate

[6] The wideband channel characterizations in this section can also be done for discrete-time channels that are discrete with respect to τ by changing integrals to sums and Fourier transforms to discrete Fourier transforms.

[7] It is easily shown that the autocorrelation of the passband channel response $h(\tau, t)$ is given by $E[h(\tau_1, t) h(\tau_2, t + \Delta t)] = .5\Re\{A_c(\tau_1, \tau_2; t, \Delta t) e^{j2\pi f_c \Delta t}\} + .5\Re\{\hat{A}_c(\tau_1, \tau_2; t, \Delta t) e^{j2\pi f_c (2t + \Delta t)}\}$, where $\hat{A}_c(\tau_1, \tau_2; t, \Delta t) = E[c(\tau_1; t) c(\tau_2; t + \Delta t)]$. However, if $c(\tau, t)$ is WSS then $\hat{A}_c(\tau_1, \tau_2; t, \Delta t) = 0$, so $E[h(\tau_1, t) h(\tau_2, t + \Delta t)] = .5\Re\{A_c(\tau_1, \tau_2; t, \Delta t) e^{j2\pi f_c \Delta}\}$.

channels that are WSS with US as WSSUS channels. The WSSUS channel model was first introduced by Bello in his landmark paper [16], where he also developed two-dimensional transform relationships associated with this autocorrelation. These relationships will be discussed in Section 3.3.4. Incorporating the US property into (3.51) yields

$$E[c^*(\tau_1;t)c(\tau_2;t+\Delta t)] = A_c(\tau_1;\Delta t)\delta[\tau_1-\tau_2] \stackrel{\triangle}{=} A_c(\tau;\Delta t), \tag{3.52}$$

where $A_c(\tau;\Delta t)$ gives the average output power associated with the channel as a function of the multipath delay $\tau = \tau_1 = \tau_2$ and the difference Δt in observation time. This function assumes that τ_1 and τ_2 satisfy $|\tau_1 - \tau_2| > B^{-1}$, since otherwise the receiver can't resolve the two components. In this case the two components are modeled as a single combined multipath component with delay $\tau \approx \tau_1 \approx \tau_2$.

The **scattering function** for random channels is defined as the Fourier transform of $A_c(\tau;\Delta t)$ with respect to the Δt parameter:

$$S_c(\tau,\rho) = \int_{-\infty}^{\infty} A_c(\tau,\Delta t)e^{-j2\pi\rho\Delta t}d\Delta t. \tag{3.53}$$

The scattering function characterizes the average output power associated with the channel as a function of the multipath delay τ and Doppler ρ. Note that we use the same notation for the deterministic scattering and random scattering functions since the function is uniquely defined depending on whether the channel impulse response is deterministic or random. A typical scattering function is shown in Figure 3.12.

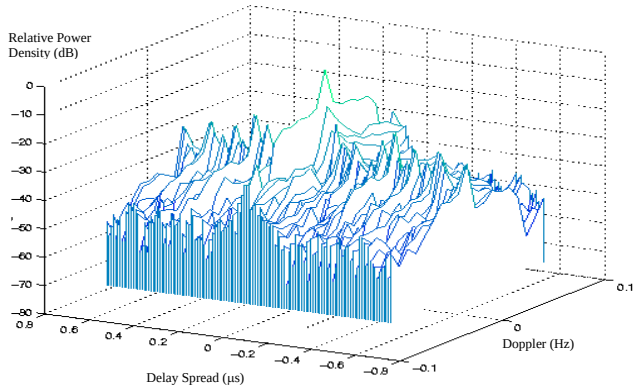


Figure 3.12: Scattering Function.

The most important characteristics of the wideband channel, including the power delay profile, coherence bandwidth, Doppler power spectrum, and coherence time, are derived from the channel autocorrelation $A_c(\tau,\Delta t)$ or scattering function $S(\tau,\rho)$. These characteristics are described in the subsequent sections.

3.3.1 Power Delay Profile

The **power delay profile** $A_c(\tau)$, also called the **multipath intensity profile**, is defined as the autocorrelation (3.52) with $\Delta t = 0$: $A_c(\tau) \stackrel{\triangle}{=} A_c(\tau,0)$. The power delay profile represents the average power associated with a given multipath delay, and is easily measured empirically. The average and rms delay spread are typically defined in terms of the power delay profile $A_c(\tau)$ as

$$\mu_{T_m} = \frac{\int_0^\infty \tau A_c(\tau)d\tau}{\int_0^\infty A_c(\tau)d\tau}, \tag{3.54}$$

and

$$\sigma_{T_m} = \sqrt{\frac{\int_0^\infty (\tau - \mu_{T_m})^2 A_c(\tau) d\tau}{\int_0^\infty A_c(\tau) d\tau}}. \tag{3.55}$$

Note that if we define the pdf p_{T_m} of the random delay spread T_m in terms of $A_c(\tau)$ as

$$p_{T_m}(\tau) = \frac{A_c(\tau)}{\int_0^\infty A_c(\tau) d\tau} \tag{3.56}$$

then μ_{T_m} and σ_{T_m} are the mean and rms values of T_m, respectively, relative to this pdf. Defining the pdf of T_m by (3.56) or, equivalently, defining the mean and rms delay spread by (3.54) and (3.55), respectively, weights the delay associated with a given multipath component by its relative power, so that weak multipath components contribute less to delay spread than strong ones. In particular, multipath components below the noise floor will not significantly impact these delay spread characterizations.

The time delay T where $A_c(\tau) \approx 0$ for $\tau \geq T$ can be used to roughly characterize the delay spread of the channel, and this value is often taken to be a small integer multiple of the rms delay spread, i.e. $A_c(\tau) \approx 0$ for $\tau > 3\sigma_{T_m}$. With this approximation a linearly modulated signal with symbol period T_s experiences significant ISI if $T_s << \sigma_{T_m}$. Conversely, when $T_s >> \sigma_{T_m}$ the system experiences negligible ISI. For calculations one can assume that $T_s << \sigma_{T_m}$ implies $T_s < \sigma_{T_m}/10$ and $T_s >> \sigma_{T_m}$ implies $T_s > 10\sigma_{T_m}$. When T_s is within an order of magnitude of σ_{T_m} then there will be some ISI which may or may not significantly degrade performance, depending on the specifics of the system and channel. We will study the performance degradation due to ISI in linearly modulated systems as well as ISI mitigation methods in later chapters.

While $\mu_{T_m} \approx \sigma_{T_m}$ in many channels with a large number of scatterers, the exact relationship between μ_{T_m} and σ_{T_m} depends on the shape of $A_c(\tau)$. A channel with no LOS component and a small number of multipath components with approximately the same large delay will have $\mu_{T_m} >> \sigma_{T_m}$. In this case the large value of μ_{T_m} is a misleading metric of delay spread, since in fact all copies of the transmitted signal arrive at rougly the same time and the demodulator would synchronize to this common delay. It is typically assumed that the synchronizer locks to the multipath component at approximately the mean delay, in which case rms delay spread characterizes the time-spreading of the channel.

Example 3.4:

The power delay spectrum is often modeled as having a one-sided exponential distribution:

$$A_c(\tau) = \frac{1}{\overline{T}_m} e^{-\tau/\overline{T}_m}, \ \tau \geq 0.$$

Show that the average delay spread (3.54) is $\mu_{T_m} = \overline{T}_m$ and find the rms delay spread (3.55).

Solution: It is easily shown that $A_c(\tau)$ integrates to one. The average delay spread is thus given by

$$\mu_{T_m} = \frac{1}{\overline{T}_m} \int_0^\infty \tau e^{-\tau/\overline{T}_m} d\tau = \overline{T}_m.$$

$$\sigma_{T_m} = \sqrt{\frac{1}{\overline{T}_m} \int_0^\infty \tau^2 e^{-\tau/\overline{T}_m} d\tau - \mu_{T_m}^2} = 2\overline{T}_m - \overline{T}_m = \overline{T}_m.$$

Thus, the average and rms delay spread are the same for exponentially distributed power delay profiles.

Example 3.5:

Consider a wideband channel with multipath intensity profile

$$A_c(\tau) = \begin{cases} e^{-\tau/.00001} & 0 \le \tau \le 20 \ \mu\text{sec.} \\ 0 & \text{else} \end{cases} .$$

Find the mean and rms delay spreads of the channel and find the maximum symbol rate such that a linearly-modulated signal transmitted through this channel does not experience ISI.

Solution: The average delay spread is

$$\mu_{T_m} = \frac{\int_0^{20*10^{-6}} \tau e^{-\tau/.00001} d\tau}{\int_0^{20*10^{-6}} e^{-\tau/.00001} d\tau} = 6.87 \ \mu\text{sec.}$$

The rms delay spread is

$$\sigma_{T_m} = \sqrt{\frac{\int_0^{20*10^{-6}} (\tau - \mu_{T_m})^2 e^{-\tau} d\tau}{\int_0^{20*10^{-6}} e^{-\tau} d\tau}} = 5.25 \ \mu\text{sec.}$$

We see in this example that the mean delay spread is roughly equal to its rms value. To avoid ISI we require linear modulation to have a symbol period T_s that is large relative to σ_{T_m}. Taking this to mean that $T_s > 10\sigma_{T_m}$ yields a symbol period of $T_s = 52.5$ μsec or a symbol rate of $R_s = 1/T_s = 19.04$ Kilosymbols per second. This is a highly constrained symbol rate for many wireless systems. Specifically, for binary modulations where the symbol rate equals the data rate (bits per second, or bps), high-quality voice requires on the order of 32 Kbps and high-speed data requires on the order of 10-100 Mbps.

3.3.2 Coherence Bandwidth

We can also characterize the time-varying multipath channel in the frequency domain by taking the Fourier transform of $c(\tau, t)$ with respect to τ. Specifically, define the random process

$$C(f; t) = \int_{-\infty}^{\infty} c(\tau; t) e^{-j2\pi f \tau} d\tau. \tag{3.57}$$

Since $c(\tau; t)$ is a complex zero-mean Gaussian random variable in t, the Fourier transform above just represents the sum[8] of complex zero-mean Gaussian random processes, and therefore $C(f; t)$ is also a zero-mean Gaussian random process completely characterized by its autocorrelation. Since $c(\tau; t)$ is WSS, its integral $C(f; t)$ is as well. Thus, the autocorrelation of (3.57) is given by

$$A_C(f_1, f_2; \Delta t) = \text{E}[C^*(f_1; t) C(f_2; t + \Delta t)]. \tag{3.58}$$

[8] We can express the integral as a limit of a discrete sum.

We can simplify $A_C(f_1, f_2; \Delta t)$ as

$$\begin{aligned} A_C(f_1, f_2; \Delta t) &= E\left[\int_{-\infty}^{\infty} c^*(\tau_1; t)e^{j2\pi f_1\tau_1}d\tau_1 \int_{-\infty}^{\infty} c(\tau_2; t+\Delta t)e^{-j2\pi f_2\tau_2}d\tau_2\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[c^*(\tau_1; t)c(\tau_2; t+\Delta t)]e^{j2\pi f_1\tau_1}e^{-j2\pi f_2\tau_2}d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} A_c(\tau, \Delta t)e^{-j2\pi(f_2-f_1)\tau}d\tau. \\ &= A_C(\Delta f; \Delta t) \end{aligned} \tag{3.59}$$

where $\Delta f = f_2 - f_1$ and the third equality follows from the WSS and US properties of $c(\tau; t)$. Thus, the autocorrelation of $C(f; t)$ in frequency depends only on the frequency difference Δf. The function $A_C(\Delta f; \Delta t)$ can be measured in practice by transmitting a pair of sinusoids through the channel that are separated in frequency by Δf and calculating their cross correlation at the receiver for the time separation Δt.

If we define $A_C(\Delta f) \stackrel{\triangle}{=} A_C(\Delta f; 0)$ then from (3.59),

$$A_C(\Delta f) = \int_{-\infty}^{\infty} A_c(\tau)e^{-j2\pi\Delta f\tau}d\tau. \tag{3.60}$$

So $A_C(\Delta f)$ is the Fourier transform of the power delay profile. Since $A_C(\Delta f) = \mathrm{E}[C^*(f; t)C(f + \Delta f; t]$ is an autocorrelation, the channel response is approximately independent at frequency separations Δf where $A_C(\Delta f) \approx 0$. The frequency B_c where $A_C(\Delta f) \approx 0$ for all $\Delta f > B_c$ is called the **coherence bandwidth** of the channel. By the Fourier transform relationship between $A_c(\tau)$ and $A_C(\Delta f)$, if $A_c(\tau) \approx 0$ for $\tau > T$ then $A_C(\Delta f) \approx 0$ for $\Delta f > 1/T$. Thus, the minimum frequency separation B_c for which the channel response is roughly independent is $B_c \approx 1/T$, where T is typically taken to be the rms delay spread σ_{T_m} of $A_c(\tau)$. A more general approximation is $B_c \approx k/\sigma_{T_m}$ where k depends on the shape of $A_c(\tau)$ and the precise specification of coherence bandwidth. For example, Lee has shown that $B_c \approx .02/\sigma_{T_m}$ approximates the range of frequencies over which channel correlation exceeds 0.9, while $B_c \approx .2/\sigma_{T_m}$ approximates the range of frequencies over which this correlation exceeds 0.5. [12].

In general, if we are transmitting a narrowband signal with bandwidth $B << B_c$, then fading across the entire signal bandwidth is highly correlated, i.e. the fading is roughly equal across the entire signal bandwidth. This is usually referred to as **flat fading**. On the other hand, if the signal bandwidth $B >> B_c$, then the channel amplitude values at frequencies separated by more than the coherence bandwidth are roughly independent. Thus, the channel amplitude varies widely across the signal bandwidth. In this case the channel is called **frequency-selective**. When $B \approx B_c$ then channel behavior is somewhere between flat and frequency-selective fading. Note that in linear modulation the signal bandwidth B is inversely proportional to the symbol time T_s, so flat fading corresponds to $T_s \approx 1/B >> 1/B_c \approx \sigma_{T_m}$, i.e. the case where the channel experiences negligible ISI. Frequency-selective fading corresponds to $T_s \approx 1/B << 1/B_c = \sigma_{T_m}$, i.e. the case where the linearly modulated signal experiences significant ISI. Wideband signaling formats that reduce ISI, such as multicarrier modulation and spread spectrum, still experience frequency-selective fading across their entire signal bandwidth which causes performance degradation, as will be discussed in Chapters 12 and 13, respectively.

We illustrate the power delay profile $A_c(\tau)$ and its Fourier transform $A_C(\Delta f)$ in Figure 3.13. This figure also shows two signals superimposed on $A_C(\Delta f)$, a narrowband signal with bandwidth much less than B_c and a wideband signal with bandwidth much greater than B_c. We see that the autocorrelation $A_C(\Delta f)$ is flat across the bandwidth of the narrowband signal, so this signal will experience flat fading or, equivalently, negligible ISI. The autocorrelation $A_C(\Delta f)$ goes to zero within the bandwidth of the wideband signal, which means that fading will be independent across different parts of the signal bandwidth, so fading is frequency selective and a linearly-modulated signal transmitted through this channel will experience significant ISI.

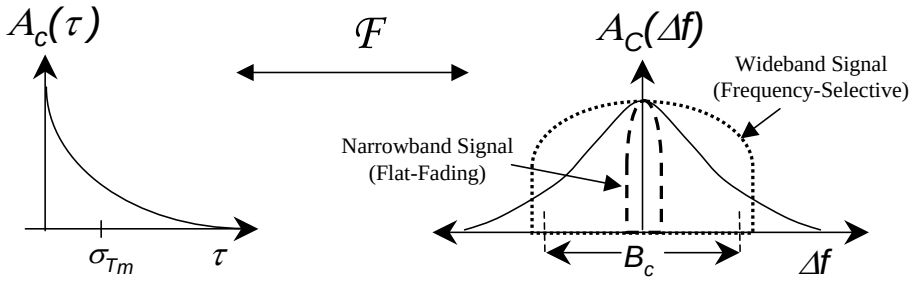


Figure 3.13: Power Delay Profile, RMS Delay Spread, and Coherence Bandwidth.

Example 3.6: In indoor channels $\sigma_{T_m} \approx 50$ ns whereas in outdoor microcells $\sigma_{T_m} \approx 30\mu$sec. Find the maximum symbol rate $R_s = 1/T_s$ for these environments such that a linearly-modulated signal transmitted through these environments experiences negligible ISI.

Solution. We assume that negligible ISI requires $T_s >> \sigma_{T_m}$, i.e. $T_s \geq 10\sigma_{T_m}$. This translates to a symbol rate $R_s = 1/T_s \leq .1/\sigma_{T_m}$. For $\sigma_{T_m} \approx 50$ ns this yields $R_s \leq 2$ Mbps and for $\sigma_{T_m} \approx 30\mu$sec this yields $R_s \leq 3.33$ Kbps. Note that indoor systems currently support up to 50 Mbps and outdoor systems up to 200 Kbps. To maintain these data rates for a linearly-modulated signal without severe performance degradation due to ISI, some form of ISI mitigation is needed. Moreover, ISI is less severe in indoor systems than in outdoor systems due to their lower delay spread values, which is why indoor systems tend to have higher data rates than outdoor systems.

3.3.3 Doppler Power Spectrum and Channel Coherence Time

The time variations of the channel which arise from transmitter or receiver motion cause a Doppler shift in the received signal. This Doppler effect can be characterized by taking the Fourier transform of $A_C(\Delta f; \Delta t)$ relative to Δt:

$$S_C(\Delta f; \rho) = \int_{-\infty}^{\infty} A_C(\Delta f; \Delta t) e^{-j2\pi\rho\Delta t} d\Delta t. \tag{3.61}$$

In order to characterize Doppler at a single frequency, we set Δf to zero and define $S_C(\rho) \stackrel{\triangle}{=} S_C(0; \rho)$. It is easily seen that

$$S_C(\rho) = \int_{-\infty}^{\infty} A_C(\Delta t) e^{-j2\pi\rho\Delta t} d\Delta t \tag{3.62}$$

where $A_C(\Delta t) \stackrel{\triangle}{=} A_C(\Delta f = 0; \Delta t)$. Note that $A_C(\Delta t)$ is an autocorrelation function defining how the channel impulse response decorrelates over time. In particular $A_C(\Delta t = T) = 0$ indicates that observations of the channel impulse response at times separated by T are uncorrelated and therefore independent, since the channel is a Gaussian random process. We define the **channel coherence time** T_c to be the range of values over which $A_C(\Delta t)$ is approximately nonzero. Thus, the time-varying channel decorrelates after approximately T_c seconds. The function $S_C(\rho)$ is called the **Doppler power spectrum** of the channel: as the Fourier transform of an autocorrelation

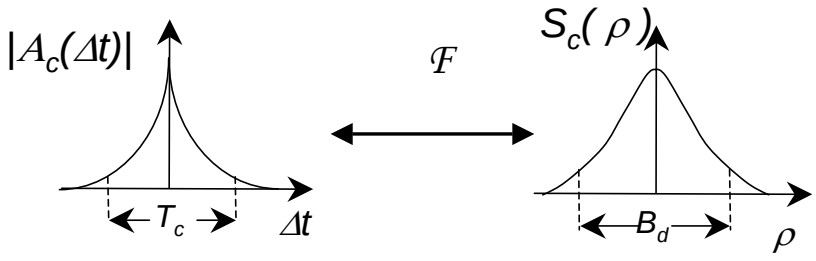


Figure 3.14: Doppler Power Spectrum, Doppler Spread, and Coherence Time.

it gives the PSD of the received signal as a function of Doppler ρ. The maximum ρ value for which $|S_C(\rho)|$ is greater than zero is called the **Doppler spread** of the channel, and is denoted by B_D. By the Fourier transform relationship between $A_C(\Delta t)$ and $S_C(\rho)$, $B_D \approx 1/T_c$. If the transmitter and reflectors are all stationary and the receiver is moving with velocity v, then $B_D \leq v/\lambda = f_D$. Recall that in the narrowband fading model samples became independent at time $\Delta t = .4/f_D$, so in general $B_D \approx k/T_c$ where k depends on the shape of $S_c(\rho)$. We illustrate the Doppler power spectrum $S_C(\rho)$ and its inverse Fourier transform $A_C(\Delta_t)$ in Figure 3.14.

Example 3.7:

For a channel with Doppler spread $B_d = 80$ Hz, what time separation is required in samples of the received signal such that the samples are approximately independent.

Solution: The coherence time of the channel is $T_c \approx 1/B_d = 1/80$, so samples spaced 12.5 ms apart are approximately uncorrelated and thus, given the Gaussian properties of the underlying random process, these samples are approximately independent.

3.3.4 Transforms for Autocorrelation and Scattering Functions

From (3.61) we see that the scattering function $S_c(\tau;\rho)$ defined in (3.53) is the inverse Fourier transform of $S_C(\Delta f;\rho)$ in the Δf variable. Furthermore $S_c(\tau;\rho)$ and $A_C(\Delta f;\Delta t)$ are related by the double Fourier transform

$$S_c(\tau;\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_C(\Delta f;\Delta t)e^{-j2\pi\rho\Delta t}e^{j2\pi\tau\Delta f}d\Delta t d\Delta f. \tag{3.63}$$

The relationships among the four functions $A_C(\Delta f;\Delta t)$, $A_c(\tau;\Delta t)$, $S_C(\Delta f;\rho)$, and $S_c(\tau;\rho)$ are shown in Figure 3.15

Empirical measurements of the scattering function for a given channel are often used to approximate empirically the channel's delay spread, coherence bandwidth, Doppler spread, and coherence time. The delay spread for a channel with empirical scattering function $S_c(\tau;\rho)$ is obtained by computing the empirical power delay profile $A_c(\tau)$ from $A_c(\tau,\Delta t) = \mathcal{F}_\rho^{-1}[S_c(\tau;\rho)]$ with $\Delta t = 0$ and then computing the mean and rms delay spread from this power delay profile. The coherence bandwidth can then be approximated as $B_c \approx 1/\sigma_{T_m}$. Similarly, the Doppler

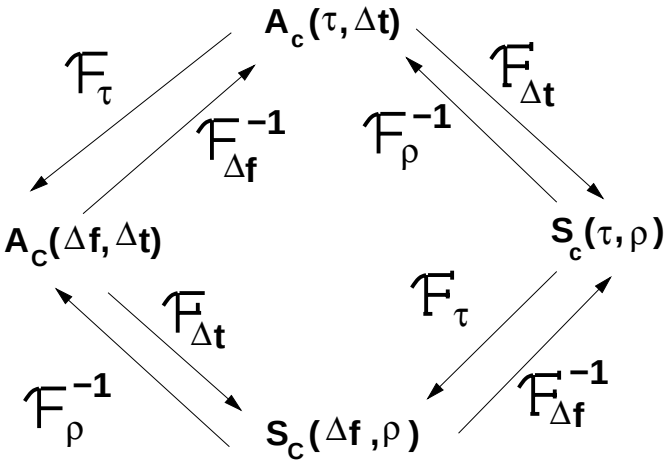


Figure 3.15: Fourier Transform Relationships

spread B_D is approximated as the range of ρ values over which $S(0;\rho)$ is roughly nonzero, with the coherence time $T_c \approx 1/B_D$.

3.4 Discrete-Time Model

Often the time-varying impulse response channel model is too complex for simple analysis. In this case a discrete-time approximation for the wideband multipath model can be used. This discrete-time model, developed by Turin in [3], is especially useful in the study of spread spectrum systems and RAKE receivers, which is covered in Chapter 13. This discrete-time model is based on a physical propagation environment consisting of a composition of isolated point scatterers, as shown in Figure 3.16. In this model, the multipath components are assumed to form subpath clusters: incoming paths on a given subpath with approximate delay τ_n are combined, and incoming paths on different subpath clusters with delays r_n and r_m where $|r_n - r_m| > 1/B$ can be resolved, where B denotes the signal bandwidth.

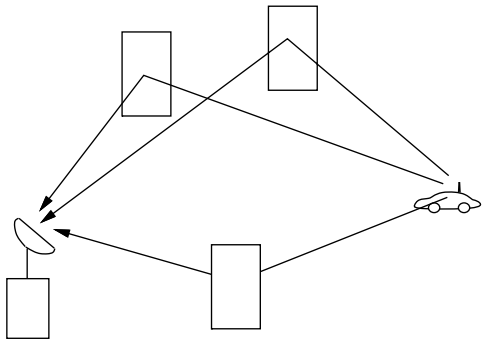


Figure 3.16: Point Scatterer Channel Model

The channel model of (3.6) is modified to include a fixed number $N+1$ of these subpath clusters as

$$c(\tau;t) = \sum_{n=0}^{N} \alpha_n(t) e^{-j\phi_n(t)} \delta(\tau - \tau_n(t)). \tag{3.64}$$

The statistics of the received signal for a given t are thus given by the statistics of $\{\tau_n\}_0^N$, $\{\alpha_n\}_0^N$, and $\{\phi_n\}_0^N$. The model can be further simplified using a discrete time approximation as follows: For a fixed t, the time axis is divided into M equal intervals of duration T such that $MT \geq \sigma_{T_m}$, where σ_{T_m} is the rms delay spread of the channel, which is derived empirically. The subpaths are restricted to lie in one of the M time interval bins, as shown in Figure 3.17. The multipath spread of this discrete model is MT, and the resolution between paths is T. This resolution is based on the transmitted signal bandwidth: $T \approx 1/B$. The statistics for the nth bin are that r_n, $1 \leq n \leq M$, is a binary indicator of the existence of a multipath component in the nth bin: so r_n is one if there is a multipath component in the nth bin and zero otherwise. If $r_n = 1$ then (a_n, θ_n), the amplitude and phase corresponding to this multipath component, follow an empirically determined distribution. This distribution is obtained by sample averages of (a_n, θ_n) for each n at different locations in the propagation environment. The empirical distribution of (a_n, θ_n) and (a_m, θ_m), $n \neq m$, is generally different, it may correspond to the same family of fading but with different parameters (e.g. Ricean fading with different K factors), or it may correspond to different fading distributions altogether (e.g. Rayleigh fading for the nth bin, Nakagami fading for the mth bin).

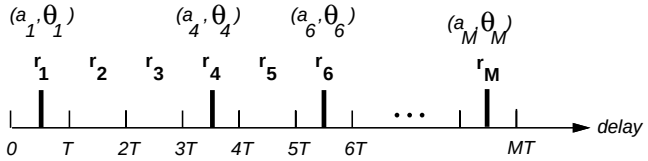


Figure 3.17: Discrete Time Approximation

This completes the statistical model for the discrete time approximation for a single snapshot. A sequence of profiles will model the signal over time as the channel impulse response changes, e.g. the impulse response seen by a receiver moving at some nonzero velocity through a city. Thus, the model must include both the first order statistics of $(\tau_n, \alpha_n, \phi_n)$ for each profile (equivalently, each t), but also the temporal and spatial correlations (assumed Markov) between them. More details on the model and the empirically derived distributions for N and for $(\tau_n, \alpha_n, \phi n)$ can be found in [3].

3.5 Space-Time Channel Models

Multiple antennas at the transmitter and/or receiver are becoming very common in wireless systems, due to their diversity and capacity benefits. Systems with multiple antennas require channel models that characterize both spatial (angle of arrival) and temporal characteristics of the channel. A typical model assumes the channel is composed of several scattering centers which generate the multipath [23, 24]. The location of the scattering centers relative to the receiver dictate the angle of arrival (AOA) of the corresponding multipath components. Models can be either two dimensional or three dimensional.

Consider a two-dimensional multipath environment where the receiver or transmitter has an antenna array with M elements. The time-varying impulse response model (3.6) can be extended to incorporate AOA for the array as follows.

$$c(\tau, t) = \sum_{n=0}^{N(t)} \alpha_n(t) e^{-j\phi_n(t)} \overline{a}(\theta_n(t)) \delta(\tau - \tau_n(t)), \tag{3.65}$$

where $\phi_n(t)$ corresponds to the phase shift at the origin of the array and $\overline{a}(\theta_n(t))$ is the array response vector given by

$$\overline{a}(\theta_n(t)) = [e^{-j\psi_{n,1}}, \ldots, e^{-j\psi_{n,M}}]^T, \tag{3.66}$$

where $\psi_{n,i} = [x_i \cos\theta_n(t) + y_i \sin\theta_n(t)]2\pi/\lambda$ for (x_i, y_i) the antenna location relative to the origin and $\theta_n(t)$ the AOA of the multipath relative to the origin of the antenna array. Assume the AOA is stationary and identically distributed for all multipath components and denote this random AOA by θ. Let $A(\theta)$ denote the average received signal power as a function of θ. Then we define the mean and rms angular spread in terms of this power profile as

$$\mu_\theta = \frac{\int_{-\pi}^{\pi} \theta A(\theta) d\theta}{\int_{-\pi}^{\pi} A(\theta) d\theta}, \tag{3.67}$$

and

$$\sigma_\theta = \sqrt{\frac{\int_{-\pi}^{\pi} (\theta - \mu_\theta)^2 A(\theta) d\theta}{\int_{-\pi}^{\pi} A(\theta) d\theta}}, \tag{3.68}$$

We say that two signals received at AOAs separated by $1/\sigma_\theta$ are roughly uncorrelated. More details on the power distribution relative to the AOA for different propagation environments along with the corresponding correlations across antenna elements can be found in [24]

Extending the two dimensional models to three dimensions requires characterizing the elevation AOAs for multipath as well as the azimuth angles. Different models for such 3-D channels have been proposed in [25, 26, 27]. In [23] the Jakes model is extended to produce spatio-temporal characteristics using the ideas of [25, 26, 27]. Several other papers on spatio-temporal modeling can be found in [29].

Bibliography

[1] R.S. Kennedy. *Fading Dispersive Communication Channels*. New York: Wiley, 1969.

[2] D.C. Cox. "910 MHz urban mobile radio propagation: Multipath characteristics in New York City," *IEEE Trans. Commun.*, Vol. COM-21, No. 11, pp. 1188–1194, Nov. 1973.

[3] G.L. Turin. "Introduction to spread spectrum antimultipath techniques and their application to urban digital radio," *IEEE Proceedings*, Vol. 68, No. 3, pp. 328–353, March 1980.

[4] R.H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, pp. 957-1000, July-Aug. 1968.

[5] W.C. Jakes, Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.

[6] T.S. Rappaport, *Wireless Communications - Principles and Practice*, 2nd Edition, Prentice Hall, 2001.

[7] M. Pätzold, *Mobile fading channels: Modeling, analysis, and simulation*, Wiley, 2002.

[8] M.K. Simon and M.-Sl. Alouini, *Digital Communication over Fading Channels*, New York: Wiley, 2000.

[9] S.O. Rice, "Mathematical analysis of random noise," *Bell System Tech. J.*, Vol. 23, No. 7, pp. 282–333, July 1944, and Vol. 24, No. 1, pp. 46–156, Jan. 1945.

[10] J.G. Proakis, *Digital Communications*, 3rd Ed., New York: McGraw-Hill, 1995.

[11] G.L. Stuber, *Principles of Mobile Communications*, Kluwer Academic Publishers, 2nd Ed., 2001.

[12] W.C.Y. Lee, *Mobile Cellular Telecommunications Systems*, New York: Mcgraw Hill, 1989.

[13] F. Babich, G. Lombardi, and E. Valentinuzzi, "Variable order Markov modeling for LEO mobile satellite channels," *Electronic Letters*, pp. 621–623, April 1999.

[14] A.M. Chen and R.R. Rao, "On tractable wireless channel models," *Proc. International Symp. on Pers., Indoor, and Mobile Radio Comm.*, pp. 825–830, Sept. 1998.

[15] H.S. Wang and N. Moayeri, "Finite-state Markov channel - A useful model for radio communication channels," *IEEE Trans. Vehic. Technol.*, pp. 163–171, Feb. 1995.

[16] P.A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Comm. Syst.*, pp. 360–393, Dec. 1963.

[17] Y. L. Guan and L. F. Turner, "Generalised FSMC model for radio channels with correlated fading," *IEE Proc. Commun.*, pp. 133–137, April 1999.

[18] M. Chu and W. Stark,“Effect of mobile velocity on communications in fading channels,” *IEEE Trans. Vehic. Technol.*, Vol 49, No. 1, pp. 202–210, Jan. 2000.

[19] C.C. Tan and N.C. Beaulieu, “On first-order Markov modeling for the Rayleigh fading channel,” *IEEE Trans. Commun.*, Vol. 48, No. 12, pp. 2032–2040, Dec. 2000.

[20] C. Pimentel and I.F. Blake, “”Modeling burst channels using partitioned Fritchman’s Markov models, *IEEE Trans. Vehic. Technol.*, pp. 885–899, Aug. 1998.

[21] C. Komninakis and R. D. Wesel, ”Pilot-aided joint data and channel estimation in flat correlated fading,” *Proc. of IEEE Globecom Conf. (Comm. Theory Symp.)*, pp. 2534–2539, Nov. 1999.

[22] M. Peleg, S. Shamai (Shitz), and S. Galan, “Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channels,” *IEE Proceedings - Communications*, Vol. 147, pp. 87–95, April 2000.

[23] Y. Mohasseb and M.P. Fitz, “A 3-D spatio-temporal simulation model for wireless channels,” *IEEE J. Select. Areas Commun.* pp. 1193–1203, Aug. 2002.

[24] R. Ertel, P. Cardieri, K.W. Sowerby, T. Rappaport, and J. H. Reed, “Overview of spatial channel models for antenna array communication systems,” *IEEE Pers. Commun. Magazine*, pp. 10–22, Feb. 1998.

[25] T. Aulin, “A modified model for fading signal at the mobile radio channel,” *IEEE Trans. Vehic. Technol.*, pp. 182–202, Aug. 1979.

[26] J.D. Parsons and M.D.Turkmani, “Characterization of mobile radio signals: model description.” *Proc. Inst. Elect. Eng.* pt. 1, pp. 459–556, Dec. 1991.

[27] J.D. Parsons and M.D.Turkmani, “Characterization of mobile radio signals: base station crosscorrelation.” *Proc. Inst. Elect. Eng.* pt. 2, pp. 459–556, Dec. 1991.

[28] D. Parsons, *The Mobile Radio Propagation Channel*. New York: Wiley, 1994.

[29] L.G. Greenstein, J.B. Andersen, H.L. Bertoni, S. Kozono, and D.G. Michelson, (Eds.), *IEEE Journal Select. Areas Commun.* Special Issue on Channel and Propagation Modeling for Wireless Systems Design, Aug. 2002.

Chapter 3 Problems

1. Consider a two-path channel consisting of a direct ray plus a ground-reflected ray where the transmitter is a fixed base station at height h and the receiver is mounted on a truck also at height h. The truck starts next to the base station and moves away at velocity v. Assume signal attenuation on each path follows a free-space path loss model. Find the time-varying channel impulse at the receiver for transmitter-receiver separation $d = vt$ sufficiently large such that the length of the reflected path can be approximated by $r+r' \approx d+2h^2/d$.

2. Find a formula for the multipath delay spread T_m for a two-path channel model. Find a simplified formula when the transmitter-receiver separation is relatively large. Compute T_m for $h_t = 10$m, $h_r = 4$m, and $d = 100$m.

3. Consider a time-invariant indoor wireless channel with LOS component at delay 23 nsec, a multipath component at delay 48 nsec, and another multipath component at delay 67 nsec. Find the delay spread assuming the demodulator synchronizes to the LOS component. Repeat assuming that the demodulator synchronizes to the first multipath component.

4. Show that the minimum value of $f_c\tau_n$ for a system at $f_c = 1$ GHz with a fixed transmitter and a receiver separated by more than 10 m from the transmitter is much greater than 1.

5. Prove that for X and Y independent zero-mean Gaussian random variables with variance σ^2, the distribution of $Z = \sqrt{X^2 + Y^2}$ is Rayleigh-distributed and the distribution of Z^2 is exponentially-distributed.

6. Assume a Rayleigh fading channel with the average signal power $2\sigma^2 = -80$ dBm. What is the power outage probability of this channel relative to the threshold $P_o = -95$ dBm? How about $P_o = -90$ dBm?

7. Assume an application that requires a power outage probability of .01 for the threshold $P_o = -80$ dBm, For Rayleigh fading, what value of the average signal power is required?

8. Assume a Rician fading channel with $2\sigma^2 = -80$ dBm and a target power of $P_o = -80$ dBm. Find the outage probability assuming that the LOS component has average power $s^2 = -80$ dBm.

9. This problem illustrates that the tails of the Ricean distribution can be quite different than its Nakagami approximation. Plot the CDF of the Ricean distribution for $K = 1, 5, 10$ and the corresponding Nakagami distribution with $m = (K+1)^2/(2K+1)$. In general, does the Ricean distribution or its Nakagami approximation have a larger outage probability $p(\gamma < x)$ for x large?

10. In order to improve the performance of cellular systems, multiple base stations can receive the signal transmitted from a given mobile unit and combine these multiple signals either by selecting the strongest one or summing the signals together, perhaps with some optimized weights. This typically increases SNR and reduces the effects of shadowing. Combining of signals received from multiple base stations is called *macrodiversity*, and in this problem we explore the benefits of this technique. Diversity will be covered in more detail in Chapter 7.

    Consider a mobile at the midpoint between two base stations in a cellular network. The received signals (in dBW) from the base stations are given by

    $$P_{r,1} = W + Z_1,$$

    $$P_{r,2} = W + Z_2,$$

    where $Z_{1,2}$ are $\mathcal{N}(0, \sigma^2)$ random variables. We define outage with macrodiversity to be the event that both $P_{r,1}$ and $P_{r,2}$ fall below a threshould T.

(a) Interpret the terms W, Z_1, Z_2 in $P_{r,1}$ and $P_{r,2}$.

(b) If Z_1 and Z_2 are independent, show that the outage probability is given by

$$P_{out} = [Q(\Delta/\sigma)]^2,$$

where $\Delta = W - T$ is the fade margin at the mobile's location.

(c) Now suppose Z_1 and Z_2 are correlated in the following way:

$$Z_1 = a\,Y_1 + b\,Y,$$

$$Z_2 = a\,Y_2 + b\,Y,$$

where Y, Y_1, Y_2 are independent $\mathcal{N}(0, \sigma^2)$ random variables, and a, b are such that $a^2 + b^2 = 1$. Show that

$$P_{out} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left[Q\left(\frac{\Delta + by\sigma}{|a|\sigma} \right) \right]^2 e^{-y^2/2} dy.$$

(d) Compare the outage probabilities of (b) and (c) for the special case of $a = b = 1/\sqrt{2}$, $\sigma = 8$ and $\Delta = 5$ (this will require a numerical integration).

11. The goal of this problem is to develop a Rayleigh fading simulator for a mobile communications channel using the method based on filtering Gaussian processes based on the in-phase and quadrature PSDs described in 3.2.1. In this problem you must do the following:

(a) Develop simulation code to generate a signal with Rayleigh fading amplitude over time. Your sample rate should be at least 1000 samples/sec, the average received envelope should be 1, and your simulation should be parameterized by the Doppler frequency f_D. Matlab is the easiest way to generate this simulation, but any code is fine.

(b) Write a description of your simulation that clearly explains how your code generates the fading envelope using a block diagram and any necessary equations.

(c) Turn in your well-commented code.

(d) Provide plots of received amplitude (dB) vs. time for $f_D = 1, 10, 100$ Hz. over 2 seconds.

12. For a Rayleigh fading channel with average power $P_r = 30$dB, compute the average fade duration for target fade values $P_0 = 0$ dB, $P_0 = 15$ dB, and $P_0 = 30$dB.

13. Derive a formula for the average length of time a Rayleigh fading process with average power P_r stays **above** a given target fade value P_0. Evaluate this average length of time for $P_r = 20$ dB, $P_0 = 25$ dB, and $f_D = 50$ Hz.

14. Assume a Rayleigh fading channel with average power $P_r = 10$ dB and Doppler $f_D = 80$ Hz. We would like to approximate the channel using a finite state Markov model with eight states. The regions R_j corresponds to $R_1 = \gamma : -\infty \leq \gamma \leq -10$dB, $R_2 = \gamma : -10\text{dB} \leq \gamma \leq 0$dB, $R_3 = \gamma : 0\text{dB} \leq \gamma \leq 5$dB, $R_4 = \gamma : 5\text{dB} \leq \gamma \leq 10$dB, $R_5 = \gamma : 10\text{dB} \leq \gamma \leq 15$dB, $R_6 = \gamma : 15\text{dB} \leq \gamma \leq 20$dB, $R_7 = \gamma : 20\text{dB} \leq \gamma \leq 30$dB, $R_8 = \gamma : 30\text{dB} \leq \gamma \leq \infty$. Find the transition probabilties between each region for this model.

15. Consider the following channel scattering function obtained by sending a 900 MHz sinusoidal input into the channel:

$$S(\tau,\rho) = \begin{cases} \alpha_1\delta(\tau) & \rho = 70\text{Hz.} \\ \alpha_2\delta(\tau - .022\mu\text{sec}) & \rho = 49.5\text{Hz.} \\ 0 & \text{else} \end{cases}$$

where α_1 and α_2 are determined by path loss, shadowing, and multipath fading. Clearly this scattering function corresponds to a 2-ray model. Assume the transmitter and receiver used to send and receive the sinusoid are located 8 meters above the ground.

(a) Find the distance and velocity between the transmitter and receiver.

(b) For the distance computed in part (a), is the path loss as a function of distance proportional to d^{-2} or d^{-4}? *Hint: use the fact that the channel is based on a 2-ray model.*

(c) Does a 30 KHz voice signal transmitted over this channel experience flat or frequency-selective fading?

16. Consider a wideband channel characterized by the autocorrelation function

$$A_c(\tau,\Delta t) = \begin{cases} \text{sinc}(W\Delta t) & 0 \le \tau \le 10\mu\text{sec.} \\ 0 & \text{else} \end{cases},$$

where $W = 100$Hz and $sinc(x) = \sin(\pi x)/(\pi x)$.

(a) Does this channel correspond to an indoor channel or an outdoor channel, and why?

(b) Sketch the scattering function of this channel.

(c) Compute the channel's average delay spread, rms delay spread, and Doppler spread.

(d) Over approximately what range of data rates will a signal transmitted over this channel exhibit frequency-selective fading?

(e) Would you expect this channel to exhibit Rayleigh or Ricean fading statistics, and why?

(f) Assuming that the channel exhibits Rayleigh fading, what is the average length of time that the signal power is continuously below its average value.

(g) Assume a system with narrowband binary modulation sent over this channel. Your system has error correction coding that can correct two simultaneous bit errors. Assume also that you always make an error if the received signal power is below its average value, and never make an error if this power is at or above its average value. If the channel is Rayleigh fading then what is the maximum data rate that can be sent over this channel with error-free transmission, making the approximation that the fade duration never exceeds twice its average value.

17. Let a scattering function $S(\tau,\rho)$ be nonzero over $0 \le \tau \le .1$ ms and $-.1 \le \rho \le .1$ Hz. Assume that the power of the scattering function is approximately uniform over the range where it is nonzero.

(a) What are the multipath spread and the doppler spread of the channel?

(b) Suppose you input to this channel two identical sinusoids separated in time by Δt. What is the minimum value of Δf for which the channel response to the first sinusoid is approximately independent of the channel response to the second sinusoid.

(c) For two sinusoidal inputs to the channel $u_1(t) = \sin 2\pi ft$ and $u_2(t) = \sin 2\pi f(t + \Delta t)$, what is the minimum value of Δt for which the channel response to $u_1(t)$ is approximately independent of the channel response to $u_2(t)$.

(d) Will this channel exhibit flat fading or frequency-selective fading for a typical voice channel with a 3 KHz bandwidth? How about for a cellular channel with a 30 KHz bandwidth?

Chapter 4

Capacity of Wireless Channels

The growing demand for wireless communication makes it important to determine the capacity limits of these channels. These capacity limits dictate the maximum data rates that can be transmitted over wireless channels with asymptotically small error probability, assuming no constraints on delay or complexity of the encoder and decoder. Channel capacity was pioneered by Claude Shannon in the late 1940s, using a mathematical theory of communication based on the notion of mutual information between the input and output of a channel [1, 2, 3]. Shannon defined capacity as the mutual information maximized over all possible input distributions. The significance of this mathematical construct was Shannon's coding theorem and converse, which proved that a code did exist that could achieve a data rate close to capacity with negligible probability of error, and that any data rate higher than capacity could not be achieved without an error probability bounded away from zero. Shannon's ideas were quite revolutionary at the time, given the high data rates he predicted were possible on telephone channels and the notion that coding could reduce error probability without reducing data rate or causing bandwidth expansion. In time sophisticated modulation and coding technology validated Shannon's theory such that on telephone lines today, we achieve data rates very close to Shannon capacity with very low probability of error. These sophisticated modulation and coding strategies are treated in Chapters 5 and 8, respectively.

In this chapter we examine the capacity of a single-user wireless channel where the transmitter and/or receiver have a single antenna. Capacity of single-user systems where the transmitter and receiver have multiple antennas is treated in Chapter 10 and capacity of multiuser systems is treated in Chapter 14. We will discuss capacity for channels that are both time-invariant and time-varying. We first look at the well-known formula for capacity of a time-invariant AWGN channel. We next consider capacity of time-varying flat-fading channels. Unlike in the AWGN case, capacity of a flat-fading channel is not given by a single formula, since capacity depends on what is known about the time-varying channel at the transmitter and/or receiver. Moreover, for different channel information assumptions, there are different definitions of channel capacity, depending on whether capacity characterizes the maximum rate averaged over all fading states or the maximum constant rate that can be maintained in all fading states (with or without some probability of outage).

We will consider flat-fading channel capacity where only the fading distribution is known at the transmitter and receiver. Capacity under this assumption is typically very difficult to determine, and is only known in a few special cases. Next we consider capacity when the channel fade level is known at the receiver only (via receiver estimation) or that the channel fade level is known at both the transmitter and the receiver (via receiver estimation and transmitter feedback). We will see that the fading channel capacity with channel side information at both the transmitter and receiver is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation. The optimal power allocation in this case is a "water-filling" in time, where power and data rate are increased when channel conditions are favorable and decreased when channel conditions are not favorable.

We will also treat capacity of frequency-selective fading channels. For time-invariant frequency-selective

channels the capacity is known and is achieved with an optimal power allocation that water-fills over frequency instead of time. The capacity of a time-varying frequency-selective fading channel is unknown in general. However, this channel can be approximated as a set of independent parallel flat-fading channels, whose capacity is the sum of capacities on each channel with power optimally allocated among the channels. The capacity of this channel is known and is obtained with an optimal power allocation that water-fills over both time and frequency.

We will consider only discrete-time systems in this chapter. Most continuous-time systems can be converted to discrete-time systems via sampling, and then the same capacity results hold. However, care must be taken in choosing the appropriate sampling rate for this conversion, since time variations in the channel may increase the sampling rate required to preserve channel capacity [4].

4.1 Capacity in AWGN

Consider a discrete-time additive white Gaussian noise (AWGN) channel with channel input/output relationship $y[i] = x[i] + n[i]$, where $x[i]$ is the channel input at time i, $y[i]$ is the corresponding channel output, and $n[i]$ is a white Gaussian noise random process. Assume a channel bandwidth B and transmit power P. The channel SNR, the power in $x[i]$ divided by the power in $n[i]$, is constant and given by $\gamma = P/(N_0 B)$, where N_0 is the power spectral density of the noise. The capacity of this channel is given by Shannon's well-known formula [1]:

$$C = B \log_2(1 + \gamma), \tag{4.1}$$

where the capacity units are bits/second (bps). Shannon's coding theorem proves that a code exists that achieves data rates arbitrarily close to capacity with arbitrarily small probability of bit error. The converse theorem shows that any code with rate $R > C$ has a probability of error bounded away from zero. The theorems are proved using the concept of mutual information between the input and output of a channel. For a memoryless time-invariant channel with random input x and random output y, the channel's **mutual information** is defined as

$$I(X;Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right), \tag{4.2}$$

where the sum is taken over all possible input and output pairs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ for $\mathcal{X}$ and $\mathcal{Y}$ the input and output alphabets. The log function is typically with respect to base 2, in which case the units of mutual information are bits per second. Mutual information can also be written in terms of the **entropy** in the channel output y and conditional output $y|x$ as $I(X;Y) = H(Y) - H(Y|X)$, where $H(Y) = -\sum_{y \in \mathcal{Y}} p(y) \log p(y)$ and $H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log p(y|x)$. Shannon proved that channel capacity equals the mutual information of the channel maximized over all possible input distributions:

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right). \tag{4.3}$$

For the AWGN channel, the maximizing input distribution is Gaussian, which results in the channel capacity given by (4.1). For channels with memory, mutual information and channel capacity are defined relative to input and output sequences x^n and y^n. More details on channel capacity, mutual information, and the coding theorem and converse can be found in [2, 5, 6].

The proofs of the coding theorem and converse place no constraints on the complexity or delay of the communication system. Therefore, Shannon capacity is generally used as an upper bound on the data rates that can be achieved under real system constraints. At the time that Shannon developed his theory of information, data rates over standard telephone lines were on the order of 100 bps. Thus, it was believed that Shannon capacity, which

predicted speeds of roughly 30 Kbps over the same telephone lines, was not a very useful bound for real systems. However, breakthroughs in hardware, modulation, and coding techniques have brought commercial modems of today very close to the speeds predicted by Shannon in the 1950s. In fact, modems can exceed this 30 Kbps Shannon limit on some telephone channels, but that is because transmission lines today are of better quality than in Shannon's day and thus have a higher received power than that used in Shannon's initial calculation. On AWGN radio channels, turbo codes have come within a fraction of a dB of the Shannon capacity limit [7].

Wireless channels typically exhibit flat or frequency-selective fading. In the next two sections we consider capacity of flat-fading and frequency-selective fading channels under different assumptions regarding what is known about the channel.

Example 4.1: Consider a wireless channel where power falloff with distance follows the formula $P_r(d) = P_t(d_0/d)^3$ for $d_0 = 10$ m. Assume the channel has bandwidth $B = 30$ KHz and AWGN with noise power spectral density of $N_0 = 10^{-9}$ W/Hz. For a transmit power of 1 W, find the capacity of this channel for a transmit-receive distance of 100 m and 1 Km.

Solution: The received SNR is $\gamma = P_r(d)/(N_0B) = .1^3/(10^{-9} \times 30 \times 10^3) = 33 = 15$ dB for $d = 100$ m and $\gamma = .01^3/(10^{-9} \times 30 \times 10^3) = .033 = -15$ dB for $d = 1000$ m. The corresponding capacities are $C = B\log_2(1+\gamma) = 30000\log_2(1+33) = 152.6$ Kbps for $d = 100$ m and $C = 30000\log_2(1+.033) = 1.4$ Kbps for $d = 1000$ m. Note the significant decrease in capacity at farther distances, due to the path loss exponent of 3, which greatly reduces received power as distance increases.

4.2 Capacity of Flat-Fading Channels

4.2.1 Channel and System Model

We assume a discrete-time channel with stationary and ergodic time-varying gain $\sqrt{g[i]}, 0 \leq g[i]$, and AWGN $n[i]$, as shown in Figure 4.1. The channel power gain $g[i]$ follows a given distribution $p(g)$, e.g. for Rayleigh fading $p(g)$ is exponential. We assume that $g[i]$ is independent of the channel input. The channel gain $g[i]$ can change at each time i, either as an i.i.d. process or with some correlation over time. In a **block fading channel** $g[i]$ is constant over some blocklength T after which time $g[i]$ changes to a new independent value based on the distribution $p(g)$. Let $\overline{P}$ denote the average transmit signal power, $N_0/2$ denote the noise power spectral density of $n[i]$, and B denote the received signal bandwidth. The instantaneous received signal-to-noise ratio (SNR) is then $\gamma[i] = \overline{P}g[i]/(N_0B)$, $0 \leq \gamma[i] < \infty$, and its expected value over all time is $\overline{\gamma} = \overline{P}\overline{g}/(N_0B)$. Since $\overline{P}/(N_0B)$ is a constant, the distribution of $g[i]$ determines the distribution of $\gamma[i]$ and vice versa.

The system model is also shown in Figure 4.1, where an input message $\mathbf{w}$ is sent from the transmitter to the receiver. The message is encoded into the codeword $\mathbf{x}$, which is transmitted over the time-varying channel as $x[i]$ at time i. The channel gain $g[i]$, also called the **channel side information** (CSI), changes during the transmission of the codeword.

The capacity of this channel depends on what is known about $g[i]$ at the transmitter and receiver. We will consider three different scenarios regarding this knowledge:

1. **Channel Distribution Information (CDI):** The distribution of $g[i]$ is known to the transmitter and receiver.

2. **Receiver CSI:** The value of $g[i]$ is known at the receiver at time i, and both the transmitter and receiver know the distribution of $g[i]$.

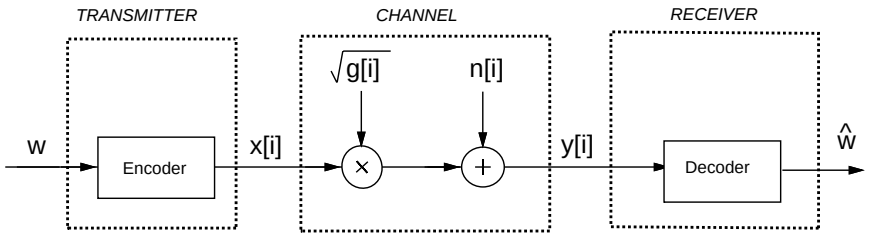


Figure 4.1: Flat-Fading Channel and System Model.

3. **Transmitter and Receiver CSI:** The value of $g[i]$ is known at the transmitter and receiver at time i, and both the transmitter and receiver know the distribution of $g[i]$.

Transmitter and receiver CSI allow the transmitter to adapt both its power and rate to the channel gain at time i, and leads to the highest capacity of the three scenarios. Note that since the instantaneous SNR $\gamma[i]$ is just $g[i]$ multipled by the constant $\overline{P}/(N_0B)$, known CSI or CDI about $g[i]$ yields the same information about $\gamma[i]$. Capacity for time-varying channels under assumptions other than these three are discussed in [8, 9].

4.2.2 Channel Distribution Information (CDI) Known

We first consider the case where the channel gain distribution $p(g)$ or, equivalently, the distribution of SNR $p(\gamma)$ is known to the transmitter and receiver. For i.i.d. fading the capacity is given by (4.3), but solving for the capacity-achieving input distribution, i.e. the distribution achieving the maximum in (4.3), can be quite complicated depending on the fading distribution. Moreover, fading correlation introduces channel memory, in which case the capacity-achieving input distribution is found by optimizing over input blocks, which makes finding the solution even more difficult. For these reasons, finding the capacity-achieving input distribution and corresponding capacity of fading channels under CDI remains an open problem for almost all channel distributions.

The capacity-achieving input distribution and corresponding fading channel capacity under CDI is known for two specific models of interest: i.i.d. Rayleigh fading channels and FSMCs. In i.i.d. Rayleigh fading the channel power gain is exponential and changes independently with each channel use. The optimal input distribution for this channel was shown in [10] to be discrete with a finite number of mass points, one of which is located at zero. This optimal distribution and its corresponding capacity must be found numerically. The lack of closed-form solutions for capacity or the optimal input distribution is somewhat surprising given the fact that the fading follows the most common fading distribution and has no correlation structure. For flat-fading channels that are not necessarily Rayleigh or i.i.d. upper and lower bounds on capacity have been determined in [11], and these bounds are tight at high SNRs.

FSMCs to approximate Rayleigh fading channels was discussed in Chapter 3.2.4. This model approximates the fading correlation as a Markov process. While the Markov nature of the fading dictates that the fading at a given time depends only on fading at the previous time sample, it turns out that the receiver must decode all past channel outputs jointly with the current output for optimal (i.e. capacity-achieving) decoding. This significantly complicates capacity analysis. The capacity of FSMCs has been derived for i.i.d. inputs in [13, 14] and for general inputs in [15]. Capacity of the FSMC depends on the limiting distribution of the channel conditioned on all past inputs and outputs, which can be computed recursively. As with the i.i.d. Rayleigh fading channel, the complexity of the capacity analysis along with the final result for this relatively simple fading model is very high, indicating the difficulty of obtaining the capacity and related design insights into channels when only CDI is available.

4.2.3 Channel Side Information at Receiver

We now consider the case where the CSI $g[i]$ is known at the receiver at time i. Equivalently, $\gamma[i]$ is known at the receiver at time i. We also assume that both the transmitter and receiver know the distribution of $g[i]$. In this case there are two channel capacity definitions that are relevant to system design: Shannon capacity, also called **ergodic capacity**, and **capacity with outage**. As for the AWGN channel, Shannon capacity defines the maximum data rate that can be sent over the channel with asymptotically small error probability. Note that for Shannon capacity the rate transmitted over the channel is constant: the transmitter cannot adapt its transmission strategy relative to the CSI. Thus, poor channel states typically reduce Shannon capacity since the transmission strategy must incorporate the effect of these poor states. An alternate capacity definition for fading channels with receiver CSI is capacity with outage. Capacity with outage is defined as the maximum rate that can be transmitted over a channel with some outage probability corresponding to the probability that the transmission cannot be decoded with negligible error probability. The basic premise of capacity with outage is that a high data rate can be sent over the channel and decoded correctly except when the channel is in deep fading. By allowing the system to lose some data in the event of deep fades, a higher data rate can be maintained than if all data must be received correctly regardless of the fading state, as is the case for Shannon capacity. The probability of outage characterizes the probability of data loss or, equivalently, of deep fading.

Shannon (Ergodic) Capacity

Shannon capacity of a fading channel with receiver CSI for an average power constraint $\overline{P}$ can be obtained from results in [16] as

$$C = \int_0^\infty B \log_2(1+\gamma) p(\gamma) d\gamma. \tag{4.4}$$

Note that this formula is a probabilistic average, i.e. Shannon capacity is equal to Shannon capacity for an AWGN channel with SNR γ, given by $B\log_2(1+\gamma)$, averaged over the distribution of γ. That is why Shannon capacity is also called ergodic capacity. However, care must be taken in interpreting (4.4) as an average. In particular, it is incorrect to interpret (4.4) to mean that this average capacity is achieved by maintaining a capacity $B\log_2(1+\gamma)$ when the instantaneous SNR is γ, since only the receiver knows the instantaneous SNR $\gamma[i]$, and therefore the data rate transmitted over the channel is constant, regardless of γ. Note, also, the capacity-achieving code must be sufficiently long so that a received codeword is affected by all possible fading states. This can result in significant delay.

By Jensen's inequality,

$$\mathbf{E}[B\log_2(1+\gamma)] = \int B\log_2(1+\gamma)p(\gamma)d\gamma \le B\log_2(1+\mathbf{E}[\gamma]) = B\log_2(1+\overline{\gamma}), \tag{4.5}$$

where $\overline{\gamma}$ is the average SNR on the channel. Thus we see that the Shannon capacity of a fading channel with receiver CSI only is less than the Shannon capacity of an AWGN channel with the same average SNR. In other words, fading reduces Shannon capacity when only the receiver has CSI. Moreover, without transmitter CSI, the code design must incorporate the channel correlation statistics, and the complexity of the maximum likelihood decoder will be proportional to the channel decorrelation time. In addition, if the receiver CSI is not perfect, capacity can be significantly decreased [20].

Example 4.2: Consider a flat-fading channel with i.i.d. channel gain $g[i]$ which can take on three possible values: $g_1 = .05$ with probability $p_1 = .1$, $g_2 = .5$ with probability $p_2 = .5$, and $g_3 = 1$ with probability $p_3 = .4$. The transmit power is 10 mW, the noise spectral density is $N_0 = 10^{-9}$ W/Hz, and the channel bandwidth is 30 KHz. Assume the receiver has knowledge of the instantaneous value of $g[i]$ but the transmitter does not. Find the

Shannon capacity of this channel and compare with the capacity of an AWGN channel with the same average SNR.

Solution: The channel has 3 possible received SNRs, $\gamma_1 = P_t g_1/(N_0 B) = .01*(.05^2)/(30000*10^{-9}) = .8333 = -.79$ dB, $\gamma_2 = P_t g_2/(N_0 B) = .01 \times (.5^2)/(30000 * 10^{-9}) = 83.333 = 19.2$ dB, and $\gamma_3 = P_t g_3/(N_0 B) = .01/(30000 * 10^{-9}) = 333.33 = 25$ dB. The probabilities associated with each of these SNR values is $p(\gamma_1) = .1$, $p(\gamma_2) = .5$, and $p(\gamma_3) = .4$. Thus, the Shannon capacity is given by

$$C = \sum_i B \log_2(1+\gamma_i) p(\gamma_i) = 30000(.1 \log_2(1.8333) + .5 \log_2(84.333) + .4 \log_2(334.33)) = 199.26 \text{ Kbps}.$$

The average SNR for this channel is $\overline{\gamma} = .1(.8333) + .5(83.33) + .4(333.33) = 175.08 = 22.43$ dB. The capacity of an AWGN channel with this SNR is $C = B \log_2(1 + 175.08) = 223.8$ Kbps. Note that this rate is about 25 Kbps larger than that of the flat-fading channel with receiver CSI and the same average SNR.

Capacity with Outage

Capacity with outage applies to slowly-varying channels, where the instantaneous SNR γ is constant over a large number of transmissions (a transmission burst) and then changes to a new value based on the fading distribution. With this model, if the channel has received SNR γ during a burst then data can be sent over the channel at rate $B \log_2(1+\gamma)$ with negligible probability of error[1]. Since the transmitter does not know the SNR value γ, it must fix a transmission rate independent of the instantaneous received SNR.

Capacity with outage allows bits sent over a given transmission burst to be decoded at the end of the burst with some probability that these bits will be decoded incorrectly. Specifically, the transmitter fixes a minimum received SNR γ_{min} and encodes for a data rate $C = B \log_2(1+\gamma_{min})$. The data is correctly received if the instantaneous received SNR is greater than or equal to γ_{min} [17, 18]. If the received SNR is below γ_{min} then the bits received over that transmission burst cannot be decoded correctly with probability approaching one, and the receiver declares an outage. The probability of outage is thus $p_{out} = p(\gamma < \gamma_{min})$. The average rate correctly received over many transmission bursts is $C_o = (1 - p_{out}) B \log_2(1+\gamma_{min})$ since data is only correctly received on $1 - p_{out}$ transmissions. The value of γ_{min} is a design parameter based on the acceptable outage probability. Capacity with outage is typically characterized by a plot of capacity versus outage, as shown in Figure 4.2. In this figure we plot the normalized capacity $C/B = \log_2(1+\gamma_{min})$ as a function of outage probability $p_{out} = p(\gamma < \gamma_{min})$ for a Rayleigh fading channel (γ exponential) with $\overline{\gamma} = 20$ dB. We see that capacity approaches zero for small outage probability, due to the requirement to correctly decode bits transmitted under severe fading, and increases dramatically as outage probability increases. Note, however, that these high capacity values for large outage probabilities have higher probability of incorrect data reception. The average rate correctly received can be maximized by finding the γ_{min} or, equivalently, the p_{out}, that maximizes C_o.

Example 4.3: Assume the same channel as in the previous example, with a bandwidth of 30 KHz and three possible received SNRs: $\gamma_1 = .8333$ with $p(\gamma_1) = .1$, $\gamma_2 = 83.33$ with $p(\gamma_2) = .5$, and $\gamma_3 = 333.33$ with $p(\gamma_3) = .4$. Find the capacity versus outage for this channel, and find the average rate correctly received for outage probabilities $p_{out} < .1$, $p_{out} = .1$ and $p_{out} = .6$.

[1]The assumption of constant fading over a large number of transmissions is needed since codes that achieve capacity require very large blocklengths.

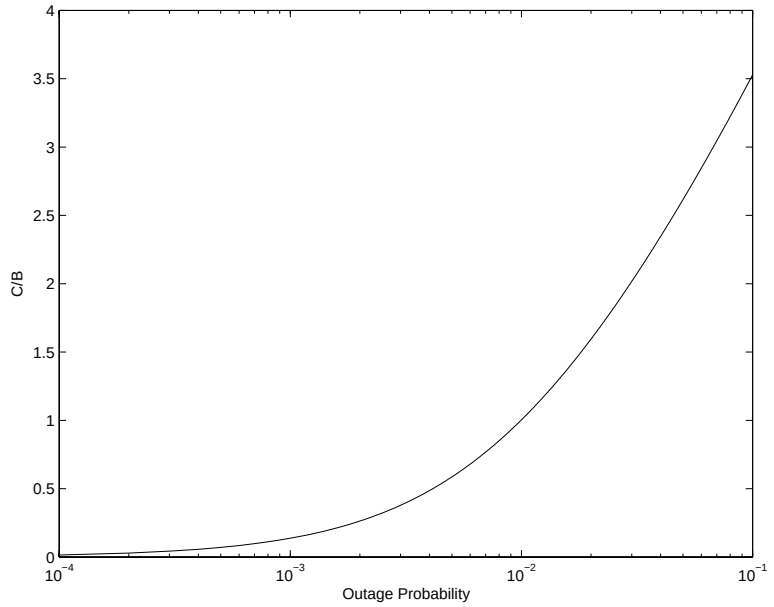


Figure 4.2: Normalized Capacity (C/B) versus Outage Probability.

Solution: For time-varying channels with discrete SNR values the capacity versus outage is a staircase function. Specifically, for $p_{out} < .1$ we must decode correctly in all channel states. The minimum received SNR for p_{out} in this range of values is that of the weakest channel: $\gamma_{min} = \gamma_1$, and the corresponding capacity is $C = B\log_2(1+\gamma_{min}) = 30000\log_2(1.833) = 26.23$ Kbps. For $.1 \leq p_{out} < .6$ we can decode incorrectly when the channel is in the weakest state only. Then $\gamma_{min} = \gamma_2$ and the corresponding capacity is $C = B\log_2(1+\gamma_{min}) = 30000\log_2(84.33) = 191.94$ Kbps. For $.6 \leq p_{out} < 1$ we can decode incorrectly if the channel has received SNR γ_1 or γ_2. Then $\gamma_{min} = \gamma_3$ and the corresponding capacity is $C = B\log_2(1+\gamma_{min}) = 30000\log_2(334.33) = 251.55$ Kbps. Thus, capacity versus outage has $C = 26.23$ Kbps for $p_{out} < .1$, $C = 191.94$ Kbps for $.1 \leq p_{out} < .6$, and $C = 251.55$ Kbps for $.6 \leq p_{out} < 1$.

For $p_{out} < .1$ data transmitted at rates close to capacity $C = 26.23$ Kbps are always correctly received since the channel can always support this data rate. For $p_{out} = .1$ we transmit at rates close to $C = 191.94$ Kbps, but we can only correctly decode these data when the channel SNR is γ_2 or γ_3, so the rate correctly received is $(1-.1)191940 = 172.75$ Kbps. For $p_{out} = .6$ we transmit at rates close to $C = 251.55$ Kbps but we can only correctly decode these data when the channel SNR is γ_3, so the rate correctly received is $(1-.6)251550 = 125.78$ Kbps. It is likely that a good engineering design for this channel would send data at a rate close to 191.94 Kbps, since it would only be received incorrectly at most 10% of this time and the data rate would be almost an order of magnitude higher than sending at a rate commensurate with the worst-case channel capacity. However, 10% retransmission probability is too high for some applications, in which case the system would be designed for the 26.23 Kbps data rate with no retransmissions. Design issues regarding acceptable retransmission probability will be discussed in Chapter 14.

4.2.4 Channel Side Information at Transmitter and Receiver

When both the transmitter and receiver have CSI, the transmitter can adapt its transmission strategy relative to this CSI, as shown in Figure 4.3. In this case there is no notion of capacity versus outage where the transmitter sends bits that cannot be decoded, since the transmitter knows the channel and thus will not send bits unless they can be decoded correctly. In this section we will derive Shannon capacity assuming optimal power and rate adaptation relative to the CSI, as well as introduce alternate capacity definitions and their power and rate adaptation strategies.

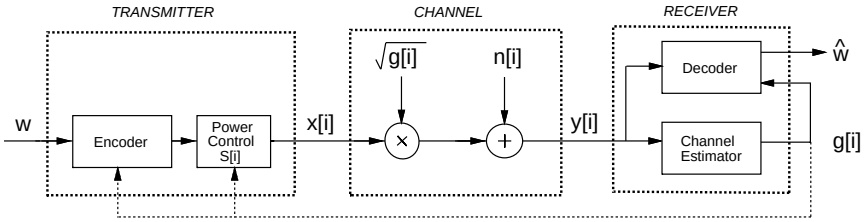


Figure 4.3: System Model with Transmitter and Receiver CSI.

Shannon Capacity

We now consider the Shannon capacity when the channel power gain $g[i]$ is known to both the transmitter and receiver at time i. The Shannon capacity of a time-varying channel with side information about the channel state at both the transmitter and receiver was originally considered by Wolfowitz for the following model. Let $s[i]$ be a stationary and ergodic stochastic process representing the channel state, which takes values on a finite set $\mathcal{S}$ of discrete memoryless channels. Let C_s denote the capacity of a particular channel $s \in \mathcal{S}$, and $p(s)$ denote the probability, or fraction of time, that the channel is in state s. The capacity of this time-varying channel is then given by Theorem 4.6.1 of [19]:

$$C = \sum_{s \in \mathcal{S}} C_s p(s). \tag{4.6}$$

We now apply this formula to the system model in Figure 4.1. We know the capacity of an AWGN channel with average received SNR γ is $C_\gamma = B \log_2(1+\gamma)$. Let $p(\gamma) = p(\gamma[i] = \gamma)$ denote the probability distribution of the received SNR. From (4.6) the capacity of the fading channel with transmitter and receiver side information is thus[2]

$$C = \int_0^\infty C_\gamma p(\gamma) d\gamma = \int_0^\infty B \log_2(1+\gamma) p(\gamma) d\gamma. \tag{4.7}$$

We see that without power adaptation, (4.4) and (4.7) are the same, so transmitter side information does not increase capacity unless power is also adapted.

Let us now allow the transmit power $P(\gamma)$ to vary with γ, subject to an average power constraint $\overline{P}$:

$$\int_0^\infty P(\gamma) p(\gamma) d\gamma \leq \overline{P}. \tag{4.8}$$

With this additional constraint, we cannot apply (4.7) directly to obtain the capacity. However, we expect that the capacity with this average power constraint will be the average capacity given by (4.7) with the power optimally

[2]Wolfowitz's result was for γ ranging over a finite set, but it can be extended to infinite sets [21].

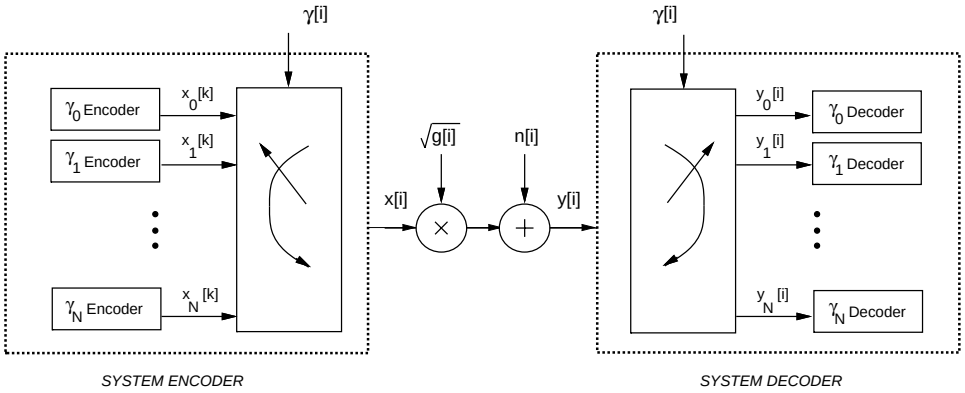


Figure 4.4: Multiplexed Coding and Decoding.

distributed over time. This motivates defining the fading channel capacity with average power constraint (4.8) as

$$C = \max_{P(\gamma):\int P(\gamma)p(\gamma)d\gamma=\overline{P}} \int_0^\infty B\log_2\left(1+\frac{P(\gamma)\gamma}{\overline{P}}\right)p(\gamma)d\gamma. \tag{4.9}$$

It is proved in [21] that the capacity given in (4.9) can be achieved, and any rate larger than this capacity has probability of error bounded away from zero. The main idea behind the proof is a "time diversity" system with multiplexed input and demultiplexed output, as shown in Figure 4.4. Specifically, we first quantize the range of fading values to a finite set $\{\gamma_j : 1 \le j \le N\}$. For each γ_j, we design an encoder/decoder pair for an AWGN channel with SNR γ_j. The input x_j for encoder γ_j has average power $P(\gamma_j)$ and data rate $R_j = C_j$, where C_j is the capacity of a time-invariant AWGN channel with received SNR $P(\gamma_j)\gamma_j/\overline{P}$. These encoder/decoder pairs correspond to a set of input and output ports associated with each γ_j. When $\gamma[i] \approx \gamma_j$, the corresponding pair of ports are connected through the channel. The codewords associated with each γ_j are thus multiplexed together for transmission, and demultiplexed at the channel output. This effectively reduces the time-varying channel to a set of time-invariant channels in parallel, where the jth channel only operates when $\gamma[i] \approx \gamma_j$. The average rate on the channel is just the sum of rates associated with each of the γ_j channels weighted by $p(\gamma_j)$, the percentage of time that the channel SNR equals γ_j. This yields the average capacity formula (4.9).

To find the optimal power allocation $P(\gamma)$, we form the Lagrangian

$$J(P(\gamma)) = \int_0^\infty B\log_2\left(1+\frac{\gamma P(\gamma)}{\overline{P}}\right)p(\gamma)d\gamma - \lambda\int_0^\infty P(\gamma)p(\gamma)d\gamma. \tag{4.10}$$

Next we differentiate the Lagrangian and set the derivative equal to zero:

$$\frac{\partial J(P(\gamma))}{\partial P(\gamma)} = \left[\left(\frac{B/\ln(2)}{1+\gamma P(\gamma)/\overline{P}}\right)\frac{\gamma}{\overline{P}} - \lambda\right]p(\gamma) = 0. \tag{4.11}$$

Solving for $P(\gamma)$ with the constraint that $P(\gamma) > 0$ yields the optimal power adaptation that maximizes (4.9) as

$$\frac{P(\gamma)}{\overline{P}} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma} & \gamma \ge \gamma_0 \\ 0 & \gamma < \gamma_0 \end{cases} \tag{4.12}$$

for some "cutoff" value γ_0. If $\gamma[i]$ is below this cutoff then no data is transmitted over the ith time interval, so the channel is only used at time i if $\gamma_0 \le \gamma[i] < \infty$. Substituting (4.12) into (4.9) then yields the capacity formula:

$$C = \int_{\gamma_0}^\infty B\log_2\left(\frac{\gamma}{\gamma_0}\right)p(\gamma)d\gamma. \tag{4.13}$$

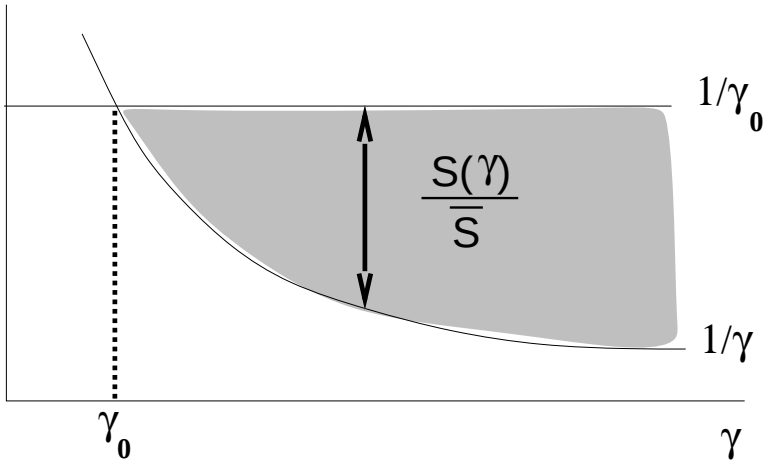


Figure 4.5: Optimal Power Allocation: Water-Filling.

The multiplexing nature of the capacity-achieving coding strategy indicates that (4.13) is achieved with a time-varying data rate, where the rate corresponding to instantaneous SNR γ is $B\log_2(\gamma/\gamma_0)$. Since γ_0 is constant, this means that as the instantaneous SNR increases, the data rate sent over the channel for that instantaneous SNR also increases. Note that this multiplexing strategy is not the only way to achieve capacity (4.13): it can also be achieved by adapting the transmit power and sending at a fixed rate [22]. We will see in Section 4.2.6 that for Rayleigh fading this capacity can exceed that of an AWGN channel with the same average power, in contrast to the case of receiver CSI only, where fading always decreases capacity.

Note that the optimal power allocation policy (4.12) only depends on the fading distribution $p(\gamma)$ through the cutoff value γ_0. This cutoff value is found from the power constraint. Specifically, by rearranging the power constraint (4.8) and replacing the inequality with equality (since using the maximum available power will always be optimal) yields the power constraint

$$\int_0^\infty \frac{P(\gamma)}{\overline{P}} p(\gamma) d\gamma = 1. \tag{4.14}$$

Now substituting the optimal power adaptation (4.12) into this expression yields that the cutoff value γ_0 must satisfy

$$\int_{\gamma_0}^\infty \left(\frac{1}{\gamma_0} - \frac{1}{\gamma}\right) p(\gamma) d\gamma = 1. \tag{4.15}$$

Note that this expression only depends on the distribution $p(\gamma)$. The value for γ_0 cannot be solved for in closed form for typical continuous pdfs $p(\gamma)$ and thus must be found numerically [23].

Since γ is time-varying, the maximizing power adaptation policy of (4.12) is a "water-filling" formula in time, as illustrated in Figure 4.5. This curve shows how much power is allocated to the channel for instantaneous SNR $\gamma(t) = \gamma$. The water-filling terminology refers to the fact that the line $1/\gamma$ sketches out the bottom of a bowl, and power is poured into the bowl to a constant water level of $1/\gamma_0$. The amount of power allocated for a given γ equals $1/\gamma_0 - 1/\gamma$, the amount of water between the bottom of the bowl ($1/\gamma$) and the constant water line ($1/\gamma_0$). The intuition behind water-filling is to take advantage of good channel conditions: when channel conditions are good (γ large) more power and a higher data rate is sent over the channel. As channel quality degrades (γ small) less power and rate are sent over the channel. If the instantaneous channel SNR falls below the cutoff value, the channel is not used. Adaptive modulation and coding techniques that follow this same principle were developed in [24, 25] and are discussed in Chapter 9.

Note that the multiplexing argument sketching how capacity (4.9) is achieved applies to any power adaptation policy, i.e. for any power adaptation policy $P(\gamma)$ with average power $\overline{P}$ the capacity

$$C = \int_0^\infty B \log_2 \left(1 + \frac{P(\gamma)\gamma}{\overline{P}}\right) p(\gamma) d\gamma. \tag{4.16}$$

can be achieved with arbitrarily small error probability. Of course this capacity cannot exceed (4.9), where power adaptation is optimized to maximize capacity. However, there are scenarios where a suboptimal power adaptation policy might have desirable properties that outweigh capacity maximization. In the next two sections we discuss two such suboptimal policies, which result in constant data rate systems, in contrast to the variable-rate transmission policy that achieves the capacity in (4.9).

Example 4.4: Assume the same channel as in the previous example, with a bandwidth of 30 KHz and three possible received SNRs: $\gamma_1 = .8333$ with $p(\gamma_1) = .1$, $\gamma_2 = 83.33$ with $p(\gamma_2) = .5$, and $\gamma_3 = 333.33$ with $p(\gamma_3) = .4$. Find the ergodic capacity of this channel assuming both transmitter and receiver have instantaneous CSI.

Solution: We know the optimal power allocation is water-filling, and we need to find the cutoff value γ_0 that satisfies the discrete version of (4.15) given by

$$\sum_{\gamma_i \geq \gamma_0} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_i}\right) p(\gamma_i) = 1. \tag{4.17}$$

We first assume that all channel states are used to obtain γ_0, i.e. assume $\gamma_0 \leq \min_i \gamma_i$, and see if the resulting cutoff value is below that of the weakest channel. If not then we have an inconsistency, and must redo the calculation assuming at least one of the channel states is not used. Applying (4.17) to our channel model yields

$$\sum_{i=1}^{3} \frac{p(\gamma_i)}{\gamma_0} - \sum_{i=1}^{3} \frac{p(\gamma_i)}{\gamma_i} = 1 \Rightarrow \frac{1}{\gamma_0} = 1 + \sum_{i=1}^{3} \frac{p(\gamma_i)}{\gamma_i} = 1 + \left(\frac{.1}{.8333} + \frac{.5}{83.33} + \frac{.4}{333.33}\right) = 1.13$$

Solving for γ_0 yields $\gamma_0 = 1/1.13 = .89 > .8333 = \gamma_1$. Since this value of γ_0 is greater than the SNR in the weakest channel, it is inconsistent as the channel should only be used for SNRs above the cutoff value. Therefore, we now redo the calculation assuming that the weakest state is not used. Then (4.17) becomes

$$\sum_{i=2}^{3} \frac{p(\gamma_i)}{\gamma_0} - \sum_{i=2}^{3} \frac{p(\gamma_i)}{\gamma_i} = 1 \Rightarrow \frac{.9}{\gamma_0} = 1 + \sum_{i=2}^{3} \frac{p(\gamma_i)}{\gamma_i} = 1 + \left(\frac{.5}{83.33} + \frac{.4}{333.33}\right) = 1.0072$$

Solving for γ_0 yields $\gamma_0 = .89$. So by assuming the weakest channel with SNR γ_1 is not used, we obtain a consistent value for γ_0 with $\gamma_1 < \gamma_0 \leq \gamma_2$. The capacity of the channel then becomes

$$C = \sum_{i=2}^{3} B \log_2(\gamma_i/\gamma_0) p(\gamma_i) = 30000(.5 \log_2(83.33/.89) + .4 \log_2(333.33/.89)) = 200.82 \text{ Kbps}.$$

Comparing with the results of the previous example we see that this rate is only slightly higher than for the case of receiver CSI only, and is still significantly below that of an AWGN channel with the same average SNR. That is because the average SNR for this channel is relatively high: for low SNR channels capacity in flat-fading can exceed that of the AWGN channel with the same SNR by taking advantage of the rare times when the channel is in a very good state.

Zero-Outage Capacity and Channel Inversion

We now consider a suboptimal transmitter adaptation scheme where the transmitter uses the CSI to maintain a constant received power, i.e., it inverts the channel fading. The channel then appears to the encoder and decoder as a time-invariant AWGN channel. This power adaptation, called **channel inversion**, is given by $P(\gamma)/\overline{P} = \sigma/\gamma$, where σ equals the constant received SNR that can be maintained with the transmit power constraint (4.8). The constant σ thus satisfies $\int \frac{\sigma}{\gamma} p(\gamma) d\gamma = 1$, so $\sigma = 1/\mathbf{E}[1/\gamma]$.

Fading channel capacity with channel inversion is just the capacity of an AWGN channel with SNR σ:

$$C = B \log_2 [1 + \sigma] = B \log_2 \left[1 + \frac{1}{\mathbf{E}[1/\gamma]} \right]. \tag{4.18}$$

The capacity-achieving transmission strategy for this capacity uses a fixed-rate encoder and decoder designed for an AWGN channel with SNR σ. This has the advantage of maintaining a fixed data rate over the channel regardless of channel conditions. For this reason the channel capacity given in (4.18) is called **zero-outage capacity**, since the data rate is fixed under all channel conditions and there is no channel outage. Note that there exist practical coding techniques that achieve near-capacity data rates on AWGN channels, so the zero-outage capacity can be approximately achieved in practice.

Zero-outage capacity can exhibit a large data rate reduction relative to Shannon capacity in extreme fading environments. For example, in Rayleigh fading $\mathbf{E}[1/\gamma]$ is infinite, and thus the zero-outage capacity given by (4.18) is zero. Channel inversion is common in spread spectrum systems with near-far interference imbalances [26]. It is also the simplest scheme to implement, since the encoder and decoder are designed for an AWGN channel, independent of the fading statistics.

Example 4.5: Assume the same channel as in the previous example, with a bandwidth of 30 KHz and three possible received SNRs: $\gamma_1 = .8333$ with $p(\gamma_1) = .1$, $\gamma_2 = 83.33$ with $p(\gamma_2) = .5$, and $\gamma_3 = 333.33$ with $p(\gamma_3) = .4$. Assuming transmitter and receiver CSI, find the zero-outage capacity of this channel.

Solution: The zero-outage capacity is $C = B \log_2[1 + \sigma]$, where $\sigma = 1/\mathbf{E}[1/\gamma]$. Since

$$\mathbf{E}[1/\gamma] = \frac{.1}{.8333} + \frac{.5}{83.33} + \frac{.4}{333.33} = .1272,$$

we have $C = 30000 \log_2(1 + 1/.1272) = 9443$ Kbps. Note that this is less than half of the Shannon capacity with optimal water-filling adaptation.

Outage Capacity and Truncated Channel Inversion

The reason zero-outage capacity may be significantly smaller than Shannon capacity on a fading channel is the requirement to maintain a constant data rate in all fading states. By suspending transmission in particularly bad fading states (outage channel states), we can maintain a higher constant data rate in the other states and thereby significantly increase capacity. The **outage capacity** is defined as the maximum data rate that can be maintained in all nonoutage channel states times the probability of nonoutage. Outage capacity is achieved with a **truncated channel inversion** policy for power adaptation that only compensates for fading above a certain cutoff fade depth γ_0:

$$\frac{P(\gamma)}{\overline{P}} = \begin{cases} \frac{\sigma}{\gamma} & \gamma \geq \gamma_0 \\ 0 & \gamma < \gamma_0 \end{cases}, \tag{4.19}$$

where γ_0 is based on the outage probability: $p_{out} = p(\gamma < \gamma_0)$. Since the channel is only used when $\gamma \geq \gamma_0$, the power constraint (4.8) yields $\sigma = 1/\mathbf{E}_{\gamma_0}[1/\gamma]$, where

$$\mathbf{E}_{\gamma_0}[1/\gamma] \triangleq \int_{\gamma_0}^{\infty} \frac{1}{\gamma} p(\gamma) d\gamma. \tag{4.20}$$

The outage capacity associated with a given outage probability p_{out} and corresponding cutoff γ_0 is given by

$$C(p_{out}) = B \log_2 \left(1 + \frac{1}{\mathbf{E}_{\gamma_0}[1/\gamma]}\right) p(\gamma \geq \gamma_0). \tag{4.21}$$

We can also obtain the **maximum outage capacity** by maximizing outage capacity over all possible γ_0:

$$C = \max_{\gamma_0} B \log_2 \left(1 + \frac{1}{\mathbf{E}_{\gamma_0}[1/\gamma]}\right) p(\gamma \geq \gamma_0). \tag{4.22}$$

This maximum outage capacity will still be less than Shannon capacity (4.13) since truncated channel inversion is a suboptimal transmission strategy. However, the transmit and receive strategies associated with inversion or truncated inversion may be easier to implement or have lower complexity than the water-filling schemes associated with Shannon capacity.

Example 4.6: Assume the same channel as in the previous example, with a bandwidth of 30 KHz and three possible received SNRs: $\gamma_1 = .8333$ with $p(\gamma_1) = .1$, $\gamma_2 = 83.33$ with $p(\gamma_2) = .5$, and $\gamma_3 = 333.33$ with $p(\gamma_3) = .4$. Find the outage capacity of this channel and associated outage probabilities for cutoff values $\gamma_0 = .84$ and $\gamma_0 = 83.4$. Which of these cutoff values yields a larger outage capacity?

Solution: For $\gamma_0 = .84$ we use the channel when the SNR is γ_2 or γ_3, so $\mathbf{E}_{\gamma_0}[1/\gamma] = \sum_{i=2}^{3} p(\gamma_i)/\gamma_i = .5/83.33 + .4/333.33 = .0072$. The outage capacity is $C = B \log_2(1 + 1/\mathbf{E}_{\gamma_0}[1/\gamma]) p(\gamma \geq \gamma_0) = 30000 \log_2(1 + 138.88) * .9 = 192.457$. For $\gamma_0 = 83.34$ we use the channel when the SNR is γ_3 only, so $\mathbf{E}_{\gamma_0}[1/\gamma] = p(\gamma_3)/\gamma_3 = .4/333.33 = .0012$. The capacity is $C = B \log_2(1 + 1/\mathbf{E}_{\gamma_0}[1/\gamma]) p(\gamma \geq \gamma_0) = 30000 \log_2(1 + 833.33) * .4 = 116.45$ Kbps. The outage capacity is larger when the channel is used for SNRs γ_2 and γ_3. Even though the SNR γ_3 is significantly larger than γ_2, the fact that this SNR only occurs 40% of the time makes it inefficient to only use the channel in this best state.

4.2.5 Capacity with Receiver Diversity

Receiver diversity is a well-known technique to improve the performance of wireless communications in fading channels. The main advantage of receiver diversity is that it mitigates the fluctuations due to fading so that the channel appears more like an AWGN channel. More details on receiver diversity and its performance will be given in Chapter 7. Since receiver diversity mitigates the impact of fading, an interesting question is whether it also increases the capacity of a fading channel. The capacity calculation under diversity combining first requires that the distribution of the received SNR $p(\gamma)$ under the given diversity combining technique be obtained. Once this distribution is known it can be substituted into any of the capacity formulas above to obtain the capacity under diversity combining. The specific capacity formula used depends on the assumptions about channel side information, e.g. for the case of perfect transmitter and receiver CSI the formula (4.13) would be used. Capacity under both maximal ratio and selection combining diversity for these different capacity formulas was computed

in [27]. It was found that, as expected, the capacity with perfect transmitter and receiver CSI is bigger than with receiver CSI only, which in turn is bigger than with channel inversion. The performance gap of these different formulas decreases as the number of antenna branches increases. This trend is expected, since a large number of antenna branches makes the channel look like AWGN, for which all of the different capacity formulas have roughly the same performance.

Recently there has been much research activity on systems with multiple antennas at both the transmitter and the receiver. The excitement in this area stems from the breakthrough results in [29, 28, 30] indicating that the capacity of a fading channel with multiple inputs and outputs (a MIMO channel) is M times larger then the channel capacity without multiple antennas, where $M = \min(M_t, M_r)$ for M_t the number of transmit antennas and M_r the number of receive antennas. We will discuss capacity of multiple antenna systems in Chapter 10.

4.2.6 Capacity Comparisons

In this section we compare capacity with transmitter and receiver CSI for different power allocation policies along with the capacity under receiver CSI only. Figures 4.6, 4.7, and 4.8 show plots of the different capacities (4.4), 4.9), (4.18), and (4.22) as a function of average received SNR for log-normal fading (σ=8 dB standard deviation), Rayleigh fading, and Nakagami fading (with Nakagami parameter $m = 2$). Nakagami fading with $m = 2$ is roughly equivalent to Rayleigh fading with two-antenna receiver diversity. The capacity in AWGN for the same average power is also shown for comparison. Note that the capacity in log-normal fading is plotted relative to average dB SNR (μ_{dB}), not average SNR in dB ($10\log_{10}\mu$): the relation between these values, as given by (2.45) in Chapter 2, is $10\log_{10}\mu = \mu_{dB} + \sigma_{dB}^2\ln(10)/20$.

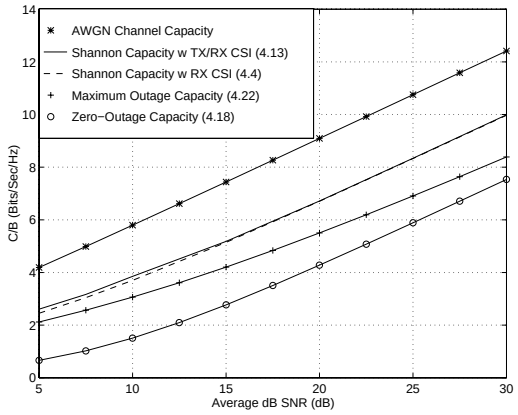


Figure 4.6: Capacity in Log-Normal Shadowing.

Several observations in this comparison are worth noting. First, we see in the figure that the capacity of the AWGN channel is larger than that of the fading channel for all cases. However, at low SNRs the AWGN and fading channel with transmitter and receiver CSI have almost the same capacity. In fact, at low SNRs (below 0 dB), capacity of the fading channel with transmitter and receiver CSI is larger than the corresponding AWGN channel capacity. That is because the AWGN channel always has the same low SNR, thereby limiting it capacity. A fading channel with this same low average SNR will occasionally have a high SNR, since the distribution has infinite range. Thus, if all power and rate is transmitted over the channel during these very infrequent high SNR values, the capacity will be larger than on the AWGN channel with the same low average SNR.

The severity of the fading is indicated by the Nakagami parameter m, where $m = 1$ for Rayleigh fading and $m = \infty$ for an AWGN channel without fading. Thus, comparing Figures 4.7 and 4.8 we see that, as the severity

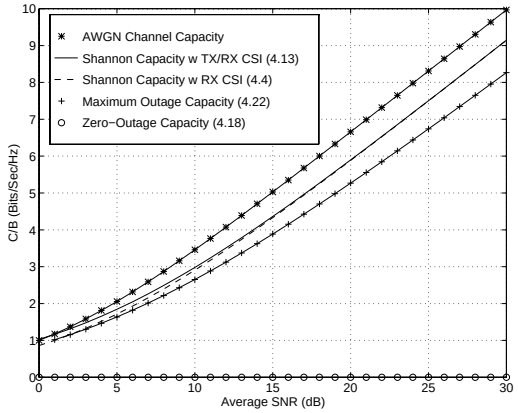


Figure 4.7: Capacity in Rayleigh Fading.

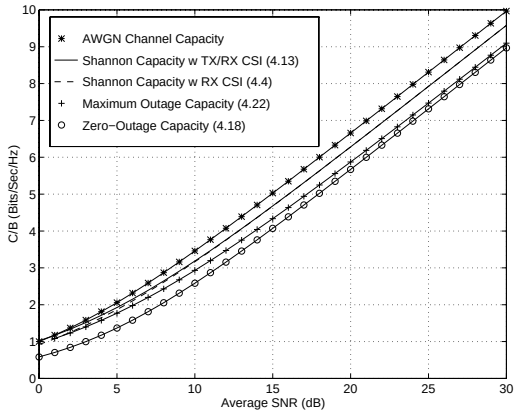


Figure 4.8: Capacity in Nakagami Fading ($m = 2$).

of the fading decreases (Rayleigh to Nakagami with $m = 2$), the capacity difference between the various adaptive policies also decreases, and their respective capacities approach that of the AWGN channel.

The difference between the capacity curves under transmitter and receiver CSI (4.9) and receiver CSI only (4.4) are negligible in all cases. Recalling that capacity under receiver CSI only (4.4) and under transmitter and receiver CSI without power adaptation (4.7) are the same, this implies that when the transmission rate is adapted relative to the channel, adapting the power as well yields a negligible capacity gain. It also indicates that transmitter adaptation yields a negligible capacity gain relative to using only receiver side information. In severe fading conditions (Rayleigh and log-normal fading), maximum outage capacity exhibits a 1-5 dB rate penalty and zero-outage capacity yields a very large capacity loss relative to Shannon capacity. However, under mild fading conditions (Nakagami with $m = 2$) the Shannon, maximum outage, and zero-outage capacities are within 3 dB of each other and within 4 dB of the AWGN channel capacity. These differences will further decrease as the fading diminishes ($m \to \infty$ for Nakagami fading).

We can view these results as a tradeoff between capacity and complexity. The adaptive policy with transmitter and receiver side information requires more complexity in the transmitter (and it typically also requires a feedback path between the receiver and transmitter to obtain the side information). However, the decoder in the receiver is relatively simple. The nonadaptive policy has a relatively simple transmission scheme, but its code design must use the channel correlation statistics (often unknown), and the decoder complexity is proportional to the channel

decorrelation time. The channel inversion and truncated inversion policies use codes designed for AWGN channels, and are therefore the least complex to implement, but in severe fading conditions they exhibit large capacity losses relative to the other techniques.

In general, Shannon capacity analysis does not show how to design adaptive or nonadaptive techniques for real systems. Achievable rates for adaptive trellis-coded MQAM have been investigated in [25], where a simple 4-state trellis code combined with adaptive six-constellation MQAM modulation was shown to achieve rates within 7 dB of the Shannon capacity (4.9) in Figures 4.6 and 4.7. More complex codes further close the gap to the Shannon limit of fading channels with transmitter adaptation.

4.3 Capacity of Frequency-Selective Fading Channels

In this section we consider the Shannon capacity of frequency-selective fading channels. We first consider the capacity of a time-invariant frequency-selective fading channel. This capacity analysis is similar to that of a flat-fading channel with the time axis replaced by the frequency axis. Next we discuss the capacity of time-varying frequency-selective fading channels.

4.3.1 Time-Invariant Channels

Consider a time-invariant channel with frequency response $H(f)$, as shown in Figure 4.9. Assume a total transmit power constraint P. When the channel is time-invariant it is typically assumed that $H(f)$ is known at both the transmitter and receiver: capacity of time-invariant channels under different assumptions of this channel knowledge are discussed in [18].

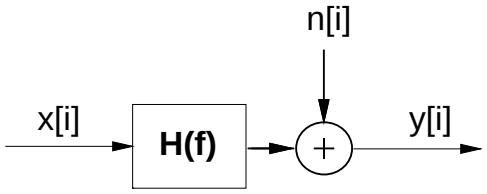


Figure 4.9: Time-Invariant Frequency-Selective Fading Channel.

Let us first assume that $H(f)$ is block-fading, so that frequency is divided into subchannels of bandwidth B, where $H(f) = H_j$ is constant over each block, as shown in Figure 4.10. The frequency-selective fading channel thus consists of a set of AWGN channels in parallel with SNR $|H_j|^2 P_j/(N_0 B)$ on the jth channel, where P_j is the power allocated to the jth channel in this parallel set, subject to the power constraint $\sum_j P_j \leq P$.

The capacity of this parallel set of channels is the sum of rates associated with each channel with power optimally allocated over all channels [5, 6]

$$C = \sum_{\max P_j : \sum_j P_j \leq P} B \log_2 \left(1 + \frac{|H_j|^2 P_j}{N_0 B}\right). \tag{4.23}$$

Note that this is similar to the capacity and optimal power allocation for a flat-fading channel, with power and rate changing over frequency in a deterministic way rather than over time in a probabilistic way. The optimal power allocation is found via the same Lagrangian technique used in the flat-fading case, which leads to the water-filling power allocation

$$\frac{P_j}{P} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma_j} & \gamma_j \geq \gamma_0 \\ 0 & \gamma_j < \gamma_0 \end{cases} \tag{4.24}$$

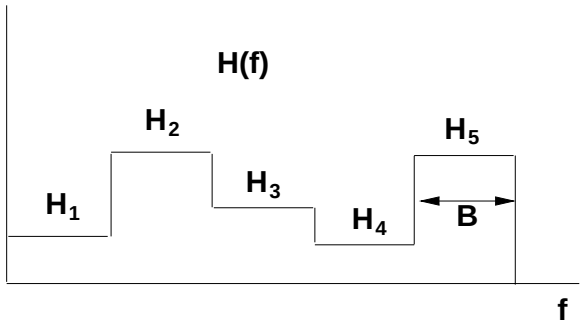


Figure 4.10: Block Frequency-Selective Fading

for some cutoff value γ_0, where $\gamma_j = |H_j|^2 P/(N_0 B)$ is the SNR associated with the jth channel assuming it is allocated the entire power budget. This optimal power allocation is illustrated in Figure 4.11. The cutoff value is obtained by substituting the power adaptation formula into the power constraint, so γ_0 must satisfy

$$\sum_j \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_j}\right) = 1. \tag{4.25}$$

The capacity then becomes

$$C = \sum_{j:\gamma_j \geq \gamma_0} B \log_2(\gamma_j/\gamma_0). \tag{4.26}$$

This capacity is achieved by sending at different rates and powers over each subchannel. Multicarrier modulation uses the same technique in adaptive loading, as discussed in more detail in Chapter 12.

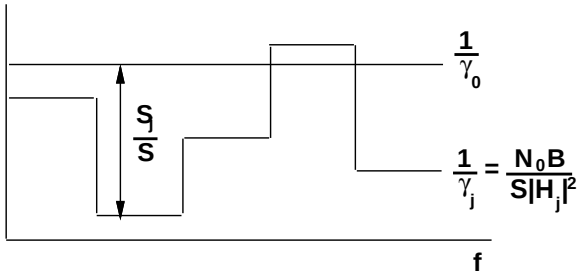


Figure 4.11: Water-Filling in Block Frequency-Selective Fading

When $H(f)$ is continuous the capacity under power constraint P is similar to the case of the block-fading channel, with some mathematical intricacies needed to show that the channel capacity is given by

$$C = \max_{P(f):\int P(f)df \leq P} \int \log_2\left(1 + \frac{|H(f)|^2 P(f)}{N_0}\right) df. \tag{4.27}$$

The equation inside the integral can be thought of as the incremental capacity associated with a given frequency f over the bandwidth df with power allocation $P(f)$ and channel gain $|H(f)|^2$. This result is formally proven using a Karhunen-Loeve expansion of the channel $h(t)$ to create an equivalent set of parallel independent channels [5, Chapter 8.5]. An alternate proof decomposes the channel into a parallel set using the discrete Fourier transform (DFT) [12]: the same premise is used in the discrete implementation of multicarrier modulation described in Chapter 12.4.

The power allocation over frequency, $P(f)$, that maximizes (4.27) is found via the Lagrangian technique. The resulting optimal power allocation is water-filling over frequency:

$$\frac{P(f)}{P} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma(f)} & \gamma(f) \geq \gamma_0 \\ 0 & \gamma(f) < \gamma_0 \end{cases} \tag{4.28}$$

This results in channel capacity

$$C = \int_{f:\gamma(f)\geq\gamma_0} \log_2(\gamma(f)/\gamma_0) df. \tag{4.29}$$

Example 4.7: Consider a time-invariant frequency-selective block fading channel consisting of three subchannels of bandwidth $B = 1$ MHz. The frequency response associated with each channel is $H_1 = 1$, $H_2 = 2$ and $H_3 = 3$. The transmit power constraint is $P = 10$ mW and the noise PSD is $N_0 = 10^{-9}$ W/Hz. Find the Shannon capacity of this channel and the optimal power allocation that achieves this capacity.

Solution: We first first find $\gamma_j = |H_j|^2 P/(N_b)$ for each subchannel, yielding $\gamma_1 = 10$, $\gamma_2 = 40$ and $\gamma_3 = 90$. The cutoff γ_0 must satisfy (4.25). Assuming all subchannels are allocated power, this yields

$$\frac{3}{\gamma_0} = 1 + \sum_j \frac{1}{\gamma_j} = 1.14 \Rightarrow \gamma_0 = 2.64 < \gamma_j \; \forall j.$$

Since the cutoff γ_0 is less than γ_j for all j, our assumption that all subchannels are allocated power is consistent, so this is the correct cutoff value. The corresponding capacity is $C = \sum_{j=1}^{3} B \log_2(\gamma_j/\gamma_0) = 1000000(\log_2(10/2.64) + \log_2(40/2.64) + \log_2(90/2.64)) = 10.93$ Mbps.

4.3.2 Time-Varying Channels

The time-varying frequency-selective fading channel is similar to the model shown in Figure 4.9, except that $H(f) = H(f, i)$, i.e. the channel varies over both frequency and time. It is difficult to determine the capacity of time-varying frequency-selective fading channels, even when the instantaneous channel $H(f, i)$ is known perfectly at the transmitter and receiver, due to the random effects of self-interference (ISI). In the case of transmitter and receiver side information, the optimal adaptation scheme must consider the effect of the channel on the past sequence of transmitted bits, and how the ISI resulting from these bits will affect future transmissions [31]. The capacity of time-varying frequency-selective fading channels is in general unknown, however upper and lower bounds and limiting formulas exist [31, 32].

We can approximate channel capacity in time-varying frequency-selective fading by taking the channel bandwidth B of interest and divide it up into subchannels the size of the channel coherence bandwidth B_c, as shown in Figure 4.12. We then assume that each of the resulting subchannels is independent, time-varying, and flat-fading with $H(f, i) = H_j[i]$ on the jth subchannel.

Under this assumption, we obtain the capacity for each of these flat-fading subchannels based on the average power $\overline{P}_j$ that we allocate to each subchannel, subject to a total power constraint $\overline{P}$. Since the channels are independent, the total channel capacity is just equal to the sum of capacities on the individual narrowband flat-fading channels subject to the total average power constraint, averaged over both time and frequency:

$$C = \max_{\{\overline{P}_j\}:\sum_j \overline{P}_j \leq \overline{P}} \sum_j C_j(\overline{P}_j), \tag{4.30}$$

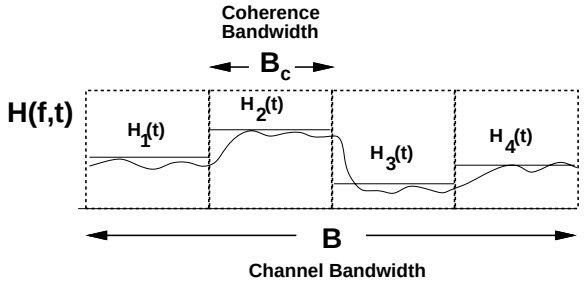


Figure 4.12: Channel Division in Frequency-Selective Fading

where $C_j(\overline{P}_j)$ is the capacity of the flat-fading subchannel with average power $\overline{P}_j$ and bandwidth B_c given by (4.13), (4.4), (4.18), or (4.22) for Shannon capacity under different side information and power allocation policies. We can also define $C_j(\overline{S}_j)$ as a capacity versus outage if only the receiver has side information.

We will focus on Shannon capacity assuming perfect transmitter and receiver channel CSI, since this upper-bounds capacity under any other side information assumptions or suboptimal power allocation strategies. We know that if we fix the average power per subchannel, the optimal power adaptation follows a water-filling formula. We also expect that the optimal average power to be allocated to each subchannel should also follow a water-filling, where more average power is allocated to better subchannels. Thus we expect that the optimal power allocation is a two-dimensional water-filling in both time and frequency. We now obtain this optimal two-dimensional water-filling and the corresponding Shannon capacity.

Define $\gamma_j[i] = |H_j[i]|^2\overline{P}/(N_0B)$ to be the instantaneous SNR on the jth subchannel at time i assuming the total power $\overline{P}$ is allocated to that time and frequency. We allow the power $P_j(\gamma_j)$ to vary with $\gamma_j[i]$. The Shannon capacity with perfect transmitter and receiver CSI is given by optimizing power adaptation relative to both time (represented by $\gamma_j[i] = \gamma_j$) and frequency (represented by the subchannel index j):

$$C = \max_{P_j(\gamma_j):\sum_j \int_0^\infty P_j(\gamma_j)p(\gamma_j)d\gamma_j \le \overline{P}} \sum_j \int_0^\infty B_c \log_2\left(1 + \frac{P_j(\gamma_j)\gamma_j}{\overline{P}}\right) p(\gamma_j)d\gamma_j. \tag{4.31}$$

To find the optimal power allocation $P_j(\gamma_j)$, we form the Lagrangian

$$J(P_j(\gamma_j)) = \sum_j \int_0^\infty B_c \log_2\left(1 + \frac{P_j(\gamma_j)\gamma_j}{\overline{P}}\right) p(\gamma_j)d\gamma_j - \lambda \sum_j \int_0^\infty P_j(\gamma_j)p(\gamma_j)d\gamma_j. \tag{4.32}$$

Note that (4.32) is similar to the Lagrangian for the flat-fading channel (4.10) except that the dimension of frequency has been added by summing over the subchannels. Differentiating the Lagrangian and setting this derivative equal to zero eliminates all terms except the given subchannel and associated SNR:

$$\frac{\partial J(P_j(\gamma_j))}{\partial P_j(\gamma_j)} = \left[\left(\frac{B/\ln(2)}{1 + \gamma_j P(\gamma_j)/\overline{P}}\right)\frac{\gamma_j}{\overline{P}} - \lambda\right] p(\gamma_j) = 0. \tag{4.33}$$

Solving for $P_j(\gamma_j)$ yields the same water-filling as the flat fading case:

$$\frac{P_j(\gamma_j)}{\overline{P}} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma_j} & \gamma_j \ge \gamma_0 \\ 0 & \gamma_j < \gamma_0 \end{cases}, \tag{4.34}$$

where the cutoff value γ_0 is obtained from the total power constraint over both time and frequency:

$$\sum_j \int_0^\infty P_j(\gamma)p_j(\gamma)d\gamma_j = \overline{P}. \tag{4.35}$$

Thus, the optimal power allocation (4.34) is a two-dimensional waterfilling with a common cutoff value γ_0. Dividing the constraint (4.35) by $\overline{P}$ and substituting in the optimal power allocation (4.34), we get that γ_0 must satisfy

$$\sum_j \int_{\gamma_0}^{\infty} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_j} \right) p(\gamma_j) d\gamma_j = 1. \tag{4.36}$$

It is interesting to note that in the two-dimensional water-filling the cutoff value for all subchannels is the same. This implies that even if the fading distribution or average fade power on the subchannels is different, all subchannels suspend transmission when the instantaneous SNR falls below the common cutoff value γ_0. Substituting the optimal power allocation (4.35) into the capacity expression (4.31) yields

$$C = \sum_j \int_{\gamma_0}^{\infty} B_c \log_2 \left(\frac{\gamma_j}{\gamma_0} \right) p(\gamma_j) d\gamma_j. \tag{4.37}$$

Bibliography

[1] C. E. Shannon *A Mathematical Theory of Communication. Bell Sys. Tech. Journal*, pp. 379–423, 623–656, 1948.

[2] C. E. Shannon *Communications in the presence of noise*. *Proc. IRE*, pp. 10-21, 1949.

[3] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. Illinois Press, 1949.

[4] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, pp. 933-946, May 2000.

[5] R.G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[6] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[7] C. Heegard and S.B. Wicker, *Turbo Coding*. Kluwer Academic Publishers, 1999.

[8] I. Csiszár and J. Kórner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. New York: Academic Press, 1981.

[9] I. Csiszár and P. Narayan, "The capacity of the Arbitrarily Varying Channel," *IEEE Trans. Inform. Theory*, Vol. 37, No. 1, pp. 18–26, Jan. 1991.

[10] I.C. Abou-Faycal, M.D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inform. Theory*, pp. 1290–1301, May 2001.

[11] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inform. Theory*, pp. 2426-2467, Oct. 2003.

[12] W. Hirt and J.L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, Vol. 34, No. 3, pp. 380-388, May 1988

[13] A.J. Goldsmith and P.P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*. pp. 868–886, May 1996.

[14] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliot channel," *IEEE Trans. Inform. Theory*, Vol. IT-35, No. 6, pp. 1277–1290, Nov. 1989.

[15] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of Finite State Markov Channels with general inputs," *Proc. IEEE Intl. Symp. Inform. Theory*, pg. 289, July 2003. Also submitted to *IEEE Trans. Inform. Theory*.

[16] R.J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, Vol IT-30, No. 1, pp. 44-53, Jan. 1984.

[17] G.J. Foschini, D. Chizhik, M. Gans, C. Papadias, and R.A. Valenzuela, "Analysis and performance of some basic space-time architectures," newblock *IEEE J. Select. Areas Commun.*, pp. 303–320, April 2003.

[18] W.L. Root and P.P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math*, pp. 1350-1393, Nov. 1968.

[19] J. Wolfowitz, *Coding Theorems of Information Theory*. 2nd Ed. New York: Springer-Verlag, 1964.

[20] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inform. Theory*, pp. 1118-1134, Nov. 1997.

[21] A.J. Goldsmith and P.P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, pp. 1986-1992, Nov. 1997.

[22] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, pp. 2007–2019, Sept. 1999.

[23] M.S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity combining techniques," *IEEE Transactions on Vehicular Technology,* pp. 1165–1181, July 1999.

[24] S.-G. Chua and A.J. Goldsmith, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. on Communications*, pp. 1218-1230, Oct. 1997.

[25] S.-G. Chua and A.J. Goldsmith, "Adaptive coded modulation," *IEEE Trans. on Communications*, pp. 595-602, May 1998.

[26] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Vehic. Technol.*, Vol. VT-40, No. 2, pp. 303–312, May 1991.

[27] M.-S. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," IEEE Transactions on Vehicular Technology, pp. 1165 -1181, July 1999.

[28] E. Teletar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Internal Tech. Memo, June 1995.

[29] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, pp. 41-59, Autumn 1996.

[30] G. Foschini and M. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, pp. 311-335, March 1998.

[31] A. Goldsmith and M Medard, "Capacity of time-varying channels with channel side information," *IEEE Intl. Symp. Inform. Theory*, pg. 372, Oct. 1996. Also to appear: *IEEE Trans. Inform. Theory*.

[32] S. Diggavi, "Analysis of multicarrier transmission in time-varying channels," *Proc. IEEE Intl. Conf. Commun.* pp. 1191–1195, June 1997.

Chapter 4 Problems

1. Capacity in AWGN is given by $C = B\log_2(1 + S/(N_0B))$. Find capacity in the limit of infinite bandwidth $B \to \infty$ as a function of S.

2. Consider an AWGN channel with bandwidth 50 MHz, received power 10 mW, and noise PSD $N_0 = 2 \times 10^{-9}$W/Hz. How much does capacity increase by doubling the received power? How much does capacity increase by doubling the channel bandwidth?

3. Consider two users simultaneously transmitting to a single receiver in an AWGN channel. This is a typical scenario in a cellular system with multiple users sending signals to a base station. Assume the users have equal received power of 10 mW and total noise at the receiver in the bandwidth of interest of 0.1 mW. The channel bandwidth for each user is 20 MHz.

   (a) Suppose that the receiver decodes user 1's signal first. In this decoding, user 2's signal acts as noise (assume it has the same statistics as AWGN). What is the capacity of user 1's channel with this additional interference noise?

   (b) Suppose that after decoding user 1's signal, the decoder re-encodes it and subtracts it out of the received signal. Then in the decoding of user 2's signal, there is no interference from user 1's signal. What then is the Shannon capacity of user 2's channel?

   Note: We will see in Chapter 14 that the decoding strategy of successively subtracting out decoded signals is optimal for achieving Shannon capacity of a multiuser channel with independent transmitters sending to one receiver.

4. Consider a flat-fading channel of bandwidth 20 MHz where for a fixed transmit power $\overline{S}$, the received SNR is one of six values: $\gamma_1 = 20$ dB, $\gamma_2 = 15$ dB, $\gamma_3 = 10$ dB, $\gamma_4 = 5$ dB, and $\gamma_5 = 0$ dB and $\gamma_6 = -5$ dB. The probability associated with each state is $p_1 = p_6 = .1$, $p_2 = p_4 = .15$, $p_3 = p_5 = .25$. Assume only the receiver has CSI.

   (a) Find the Shannon capacity of this channel.

   (b) Plot the capacity versus outage for $0 \leq p_{out} < 1$ and find the maximum average rate that can be correctly received (maximum C_o).

5. Consider a flat-fading channel where for a fixed transmit power $\overline{S}$, the received SNR is one of four values: $\gamma_1 = 30$ dB, $\gamma_2 = 20$ dB, $\gamma_3 = 10$ dB, and $\gamma_4 = 0$ dB. The probability associated with each state is $p_1 = .2$, $p_2 = .3$, $p_3 = .3$, and $p_4 = .2$. Assume both transmitter and receiver have CSI.

   (a) Find the optimal power control policy $S(i)/\overline{S}$ for this channel and its corresponding Shannon capacity per unit Hertz (C/B).

   (b) Find the channel inversion power control policy for this channel and associated zero-outage capacity per unit bandwidth.

   (c) Find the truncated channel inversion power control policy for this channel and associated outage capacity per unit bandwidth for 3 different outage probabilities: $p_{out} = .1$, $p_{out} = .01$, and p_{out} (and the associated cutoff γ_0) equal to the value that achieves maximum outage capacity.

6. Consider a cellular system where the power falloff with distance follows the formula $P_r(d) = P_t(d_0/d)^\alpha$, where $d_0 = 100$m and α is a random variable. The distribution for α is $p(\alpha = 2) = .4$, $p(\alpha = 2.5) = .3$, $p(\alpha = 3) = .2$, and $p(\alpha = 4) = .1$ Assume a receiver at a distance $d = 1000$ m from the transmitter, with

an average transmit power constraint of $P_t = 100$ mW and a receiver noise power of .1 mW. Assume both transmitter and receiver have CSI.

(a) Compute the distribution of the received SNR.

(b) Derive the optimal power control policy for this channel and its corresponding Shannon capacity per unit Hertz (C/B).

(c) Determine the zero-outage capacity per unit bandwidth of this channel.

(d) Determine the maximum outage capacity per unit bandwidth of this channel.

7. Assume a Rayleigh fading channel, where the transmitter and receiver have CSI and the distribution of the fading SNR $p(\gamma)$ is exponential with mean $\overline{\gamma} = 10$dB. Assume a channel bandwidth of 10 MHz.

(a) Find the cutoff value γ_0 and the corresponding power adaptation that achieves Shannon capacity on this channel.

(b) Compute the Shannon capacity of this channel.

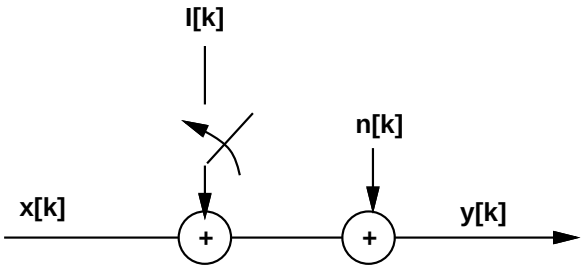
(c) Compare your answer in part (b) with the channel capacity in AWGN with the same average SNR.

(d) Compare your answer in part (b) with the Shannon capacity where only the receiver knows $\gamma[i]$.

(e) Compare your answer in part (b) with the zero-outage capacity and outage capacity with outage probability .05.

(f) Repeat parts b, c, and d (i.e. obtain the Shannon capacity with perfect transmitter and receiver side information, in AWGN for the same average power, and with just receiver side information) for the same fading distribution but with mean $\overline{\gamma} = -5$dB. Describe the circumstances under which a fading channel has higher capacity than an AWGN channel with the same average SNR and explain why this behaivor occurs.

8. Time-Varying Interference: This problem illustrates the capacity gains that can be obtained from interference estimation, and how a malicious jammer can wreak havoc on link performance. Consider the following interference channel.



The channel has a combination of AWGN $n[k]$ and interference $I[k]$. We model $I[k]$ as AWGN. The interferer is on (i.e. the switch is down) with probability .25 and off (i.e. the switch is up) with probability .75. The average transmit power is 10 mW, the noise spectral density is 10^{-8} W/Hz, the channel bandwidth B is 10 KHz (receiver noise power is N_oB), and the interference power (when on) is 9 mW.

(a) What is the Shannon capacity of the channel if neither transmitter nor receiver know when the interferer is on?

(b) What is the capacity of the channel if both transmitter and receiver know when the interferer is on?

(c) Suppose now that the interferer is a malicious jammer with perfect knowledge of $x[k]$ (so the interferer is no longer modeled as AWGN). Assume that neither transmitter nor receiver have knowledge of the jammer behavior. Assume also that the jammer is always on and has an average transmit power of 10 mW. What strategy should the jammer use to minimize the SNR of the received signal?

9. Consider the malicious interferer from the previous problem. Suppose that the transmitter knows the interference signal perfectly. Consider two possible transmit strategies under this scenario: the transmitter can ignore the interference and use all its power for sending its signal, or it can use some of its power to cancel out the interferer (i.e. transmit the negative of the interference signal). In the first approach the interferer will degrade capacity by increasing the noise, and in the second strategy the interferer also degrades capacity since the transmitter sacrifices some power to cancel out the interference. Which strategy results in higher capacity? *Note: there is a third strategy, where the encoder actually exploits the structure of the interference in its encoding. This strategy is called dirty paper coding, and is used to achieve Shannon capacity on broadcast channels with multiple antennas.*

10. Show using Lagrangian techniques that the optimal power allocation to maximize the capacity of a time-invariant block fading channel is given by the water filling formula in (4.24).

11. Consider a time-invariant block fading channel with frequency response

$$H(f) = \begin{cases} 1 & f_c - 20\text{MHz} \le f < f_c - 10\text{MHz} \\ .5 & f_c - 10\text{MHz} \le f < f_c \\ 2 & f_c \le f < f_c + 10\text{MHz} \\ .25 & f_c + 10\text{MHz} \le f < f_c + 20\text{MHz} \\ 0 & \text{else} \end{cases}$$

For a transmit power of 10mW and a noise power spectral density of $.001\mu$W per Hertz, find the optimal power allocation and corresponding Shannon capacity of this channel.

12. Show that the optimal power allocation to maximize the capacity of a time-invariant frequency selective fading channel is given by the water filling formula in (4.28).

13. Consider a frequency-selective fading channel with total bandwidth 12 MHz and coherence bandwidth $B_c = 4$ MHz. Divide the total bandwidth into 3 subchannels of bandwidth B_c , and assume that each subchannel is a Rayleigh flat-fading channel with independent fading on each subchannel. Assume the subchannels have average gains $\mathbf{E}[|H_1(t)|^2] = 1$, $\mathbf{E}[|H_2(t)|^2] = .5$, and $\mathbf{E}[|H_3(t)|^2] = .125$. Assume a total transmit power of 30 mW, and a receiver noise spectral density of $.001\mu$W per Hertz.

    (a) Find the optimal two-dimensional water-filling power adaptation for this channel and the corresponding Shannon capacity, assuming both transmitter and receiver know the instantaneous value of $H_j(t), j = 1, 2, 3$.

    (b) Compare the capacity of part (a) with that obtained by allocating an equal average power of 10 mW to each subchannel and then water-filling on each subchannel relative to this power allocation.