1 Lecture 1

Channel Model for Point-to-Point Communications

Point-to-point communications systems are well modeled using a bandpass additive noise channel model of the form shown in Figure 1.1.

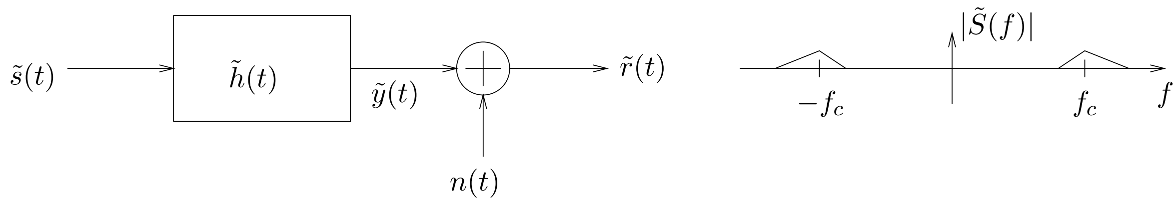


Figure 1.1: Real bandpass channel model for point-to-point communications

- The message bearing signal $\tilde{s}(t)$ is a real-valued bandpass signal whose spectrum is concentrated in the vicinity of some carrier frequency f_c.

- Distortions introduced by the channel are characterized by a linear *time invariant* system with impulse response $\tilde{h}(t)$, and frequency response $\tilde{H}(f)$ concentrated around f_c. The channel response $\tilde{h}(t)$ may or may not be known at the receiver. In the simplest case, the response $\tilde{h}(t)$ corresponds to a an ideal bandpass filter with bandwidth corresponding to that of the signal $\tilde{s}(t)$.

- The additive noise process $n(t)$ is a WSS bandpass random process. It may be idealized by *White Gaussian Noise* (WGN) for the purposes of analysis.

- The received signal $\tilde{r}(t)$ is a real-valued bandpass process as well.

In the following we convert the bandpass channel model into a more convenient and equivalent complex baseband channel model. (Also see chapter 4 of [1].)

Complex baseband representation for signal

Since the signal $\tilde{s}(t)$ is real, its spectrum $\tilde{S}(f)$ is symmetric about $f = 0$. Hence all of the information about the signal $\tilde{s}(t)$ is contained in the positive half of the spectrum $\tilde{S}(f)$, which we define to be

$$\tilde{S}_+(f) = \sqrt{2}\tilde{S}(f)u(f) \,. \tag{1.1}$$

where $u(\cdot)$ is the unit step function. The factor of $\sqrt{2}$ in the above equation makes the signal $\tilde{s}_+(t)$ have the same energy as the signal $\tilde{s}(t)$. The inverse Fourier transform of the spectrum $\tilde{S}_+(f)$ is easily shown to be the complex signal

$$\tilde{s}_+(t) = \frac{1}{\sqrt{2}}[\tilde{s}(t) + j\hat{\tilde{s}}(t)] \,, \tag{1.2}$$

where the signal $\hat{\tilde{s}}(t)$ is the Hilbert transform of $\tilde{s}(t)$ (i.e., the Fourier transform of $\hat{\tilde{s}}(t)$ is $-j\text{sgn}(f)\tilde{S}(f)$). The signal $\tilde{s}_+(t)$ is called the *pre-envelope* of $\tilde{s}(t)$. If we shift the spectrum of $\tilde{S}_+(f)$ down to the origin, we get the baseband signal $s(t)$ with

$$S(f) = \tilde{S}_+(f + f_c), \quad \text{and} \;\; s(t) = \tilde{s}_+(t)\mathrm{e}^{-j2\pi f_c t} \, . \tag{1.3}$$

Note that since $S(f)$ is not necessarily symmetric around the origin, the signal $s(t)$ is in general complex-valued. The signal $s(t)$ is called the *complex envelope* or the *complex baseband representation* of the real signal $\tilde{s}(t)$. From (1.2) and (1.3), we get

$$\tilde{s}(t) = \text{Re}[\sqrt{2}\,\tilde{s}_+(t)] = \text{Re}[\sqrt{2}\,s(t)\mathrm{e}^{j2\pi f_c t}] \, . \tag{1.4}$$

The complex envelope $s(t)$ can be written in terms of its real and imaginary parts as

$$s(t) = s_I(t) + j s_Q(t) \, . \tag{1.5}$$

From this and (1.4) we get

$$\tilde{s}(t) = \sqrt{2}[s_I(t)\cos 2\pi f_c t - s_Q(t)\sin 2\pi f_c t] = \sqrt{2}\,a(t)\cos[2\pi f_c t + \psi(t)] \, , \tag{1.6}$$

where

$$a(t) = \sqrt{s_I^2(t) + s_Q^2(t)}, \quad \text{and} \;\; \psi(t) = \tan^{-1}\frac{s_Q(t)}{s_I(t)} \, . \tag{1.7}$$

The signal $a(t)$ is called the *envelope* of $s(t)$, and $\psi(t)$ is called the *phase* of $s(t)$. It is to be noted that every bandpass signal can be written in the forms given in (1.6).

Equation (1.6) also suggests a practical way to generate the (components of) complex envelope from the passband signal. It is easy to see that if we multiply $\tilde{s}(t)$ by $\sqrt{2}\cos(2\pi f_c t)$ and low-pass filter (LPF) the output, we produce $s_I(t)$. Similarly, if we multiply by $-\sqrt{2}\sin(2\pi f_c t)$ and LPF the output, we get $s_Q(t)$.

The conversion from passband to baseband and vice-versa is illustrated below in Figure 1.2

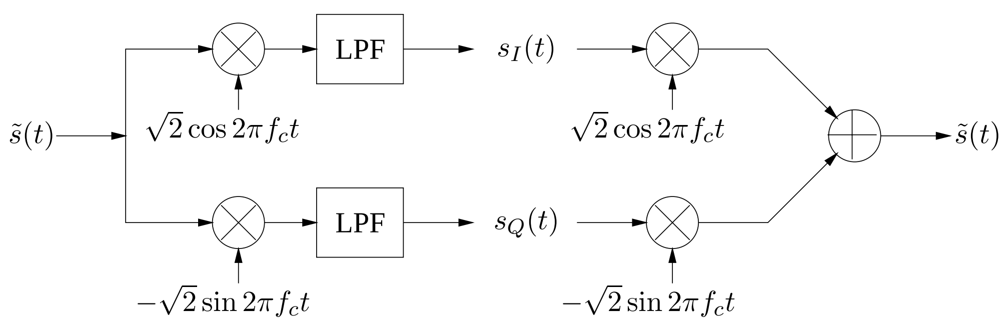


Figure 1.2: Conversion from passband to baseband and vice-versa.

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

2 Lecture 2

Complex baseband representation of channel response

Referring to Figure 1.1, since the output of the channel $\tilde{y}(t)$ is a bandpass signal, it has the complex baseband representation $y(t) = \tilde{y}_+(t)\,\mathrm{e}^{-j2\pi f_c t}$. The signal $\tilde{y}(t)$ is related to $\tilde{s}(t)$ through the convolution integral, i.e., $\tilde{y}(t) = \tilde{h} \star \tilde{s}(t)$. The question that we ask now is whether the complex envelopes $s(t)$ and $y(t)$ are related in a similar fashion, and if so, what is the corresponding complex impulse response? You will show in HW#1 that this in indeed the case and that the corresponding complex baseband channel response is given by:

$$H(f) = \frac{1}{\sqrt{2}}\tilde{H}_+(f+f_c) \quad \Longrightarrow \quad h(t) = \frac{1}{\sqrt{2}}\tilde{h}_+(t)\mathrm{e}^{-j2\pi f_c t} \;\text{ and }\; \tilde{h}(t) = 2\mathrm{Re}[h(t)\mathrm{e}^{j2\pi f_c t}]\,. \tag{2.1}$$

Note the additional factor of $\sqrt{2}$ in the equation relating $h(t)$ and $\tilde{h}(t)$.

Note that

$$y(t) = h \star s(t) = (h_I + jh_Q) \star (s_I + js_Q)(t) \tag{2.2}$$

implies that the I and Q components of $y(t)$ can be computed separately as

$$y_I(t) = h_I \star s_I(t) - h_Q \star s_Q(t)\,, \;\text{ and } y_Q(t) = h_I \star s_Q(t) + h_Q \star s_I(t)\,. \tag{2.3}$$

This suggests a way to implement the passband filter $\tilde{h}$ using real baseband operations.

Complex baseband representation of noise

We now consider the last component of the channel in Figure 1.1, namely, the additive noise term $n(t)$. Clearly, since $n(t)$ is a sample path of a WSS bandpass process, it can be represented by a complex envelope which we denote by $w(t)$.

$$w(t) = n_+(t)\mathrm{e}^{-j2\pi f_c t} = \frac{1}{\sqrt{2}}\,[n(t) + j\hat{n}(t)]\mathrm{e}^{-j2\pi f_c t}, \text{ and } \; n(t) = \mathrm{Re}[\sqrt{2}\,w(t)\mathrm{e}^{j2\pi f_c t}]\,. \tag{2.4}$$

The complex process $w(t)$ has some very interesting properties.

- ∘ If the process $n(t)$ is zero mean, then the process $w(t)$ is obviously zero mean as well.

- ∘ Let $w(t) = w_I(t) + jw_Q(t)$, where $w_I(t)$ and $w_Q(t)$ are the real in-phase and quadrature processes, respectively. Assume that $n(t)$ is zero mean and WSS. From (2.4) and the WSS property of $n(t)$, it is easily established that $w_I(t)$ and $w_Q(t)$ are jointly WSS processes (see HW#1.) Furthermore, we can show that the auto- and cross-correlation functions satisfy[1]

$$R_{w_I}(\tau) = R_{w_Q}(\tau)\,, \quad \text{and } \; R_{w_I w_Q}(\tau) = -R_{w_Q w_I}(\tau)\,. \tag{2.5}$$

A complex process with the above property is said to be a *proper complex* process. We will study such processes in greater detail later in Lecture 3. The autocorrelation function of $w(t)$ is defined as

$$R_w(\tau) = \mathrm{E}[w(t+\tau)w^\star(t)]\,. \tag{2.6}$$

[1]For jointly WSS $X(t)$ and $Y(t)$, we define $R_{XY}(\tau) = \mathrm{E}[X(t+\tau)Y(t)]$

Note the complex conjugate in the above definition. From (2.5), it is easy to show that

$$R_w(\tau) = 2R_{w_I}(\tau) + j2R_{w_Q w_I}(\tau)\ . \tag{2.7}$$

Furthermore, we can show that

$$R_n(\tau) = \text{Re}[R_w(\tau)\mathrm{e}^{j2\pi f_c \tau}]\ . \tag{2.8}$$

Let $S_w(f)$ of denote the PSD of $w(t)$. From (2.8), it is easy to show that

$$S_n(f) = \frac{1}{2}\left[S_w(f - f_c) + S_w(-f - f_c)\right] \tag{2.9}$$

and that

$$S_w(f) = 2S_n(f + f_c)\, u(f + f_c)\ . \tag{2.10}$$

○ If $n(t)$ is a stationary *Gaussian* process, then $w(t)$ is a stationary *proper complex Gaussian* (PCG) process.(See Lecture 3.)

Based on the complex baseband representations of the signal, channel response and noise, we have the following complex baseband system shown in Figure 2.1 which is equivalent to the bandpass system of Figure 1.1. Note that the signal $r(t)$ is the complex envelope of $\tilde{r}(t)$, i.e., $r(t) = \tilde{r}_+(t)\mathrm{e}^{-j2\pi f_c t}, \forall t$.

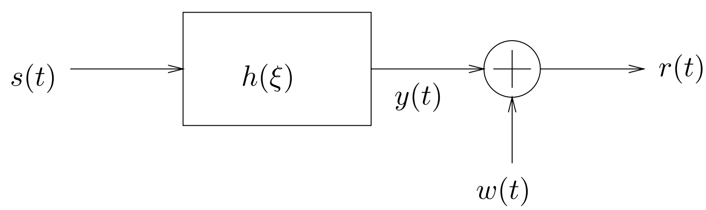


Figure 2.1: Complex baseband channel model for point-to-point communications

Idealization by White Gaussian Noise

If the process $n(t)$ is idealized by a white Gaussian noise (WGN) process with PSD $N_0/2$ in Figure 1.1, then the above does not apply since WGN is not a bandpass process. In this case we replace the process $n(t)$ by a bandpass noise process $n_{\text{BP}}(t)$ with a flat spectrum over the bandwidth of the channel, with the understanding that $n_{\text{BP}}(t)$ is the noise process seen after front-end processing at the receiver. The PSD of the process $n_{\text{BP}}(t)$ is:

$$S_{n_{\text{BP}}}(f) = \begin{cases} \frac{N_0}{2} & \text{if } |f - f_c| < \frac{B}{2} \\ 0 & \text{otherwise} \end{cases}\ , \tag{2.11}$$

where B is the channel bandwidth.

Now the process $n_{\text{BP}}(t)$ has complex envelope $w_{\text{LP}}(t)$ which is a lowpass complex Gaussian process. By the symmetry of the spectrum $S_{n_{\text{BP}}}(f)$ around f_c, we have

$$S_{w_{\text{LP}}}(f) = \begin{cases} N_0 & \text{if } |f| < \frac{B}{2} \\ 0 & \text{otherwise} \end{cases}\ . \tag{2.12}$$

Since $S_{w_{\rm LP}}(f)$ is symmetric about $f = 0$, $R_{w_{\rm LP}}(\tau)$ is a real function. From (2.7), we then conclude that the real and imaginary parts of the process $w_{\rm LP}(t)$ are uncorrelated (and hence independent). Note that

$$R_{w_{\rm LP}}(\tau) = N_0 B \text{sinc}(B\tau) \, . \tag{2.13}$$

We now idealize $w_{\rm LP}(t)$ by a white noise process for the same reason we idealize the bandpass noise by WGN. If we consider the limiting form of (2.13) as $B \to \infty$, we get

$$R_w(\tau) = N_0 \delta(\tau) \tag{2.14}$$

Again since $R_w(\tau)$ is real, from (2.7) we see that $R_{w_I w_Q}(\tau)$ must be 0 for all τ. Also,

$$R_{w_I}(\tau) = R_{w_Q}(\tau) = \frac{1}{2} R_w(\tau) = \frac{N_0}{2} \delta(\tau) \, . \tag{2.15}$$

and

$$S_{w_I}(f) = S_{w_Q}(f) = \frac{1}{2} S_w(f) = \frac{N_0}{2} \text{ for all } f \, . \tag{2.16}$$

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

3 Lecture 3

Proper Complex and Circularly Complex Gaussian Processes

The complex baseband representation of bandpass communication systems described in the previous section will be used throughout this course. We will hence need to be comfortable in dealing with complex random processes. Fortunately, all the complex random processes we have to deal with have an elegant structure (which is why they are called "proper") that makes analysis in the complex baseband domain considerably more convenient than in the real passband domain. In this lecture we study this structure in more detail.

Proper Complex Random Vectors

Let $\mathbf{Y} = \mathbf{Y}_I + j\mathbf{Y}_Q$ be a complex random vector. Define the real covariance matrices

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{Y}_I} &= \text{cov}[\mathbf{Y}_I, \mathbf{Y}_I] = \text{E}[(\mathbf{Y}_I - \mathbf{m}_{\mathbf{Y}_I})(\mathbf{Y}_I - \mathbf{m}_{\mathbf{Y}_I})^\top] \\
\mathbf{\Sigma}_{\mathbf{Y}_Q} &= \text{cov}[\mathbf{Y}_Q, \mathbf{Y}_Q] = \text{E}[(\mathbf{Y}_Q - \mathbf{m}_{\mathbf{Y}_Q})(\mathbf{Y}_Q - \mathbf{m}_{\mathbf{Y}_Q})^\top] \\
\mathbf{\Sigma}_{\mathbf{Y}_I\mathbf{Y}_Q} &= \text{cov}[\mathbf{Y}_I, \mathbf{Y}_Q] = \text{E}[(\mathbf{Y}_I - \mathbf{m}_{\mathbf{Y}_I})(\mathbf{Y}_Q - \mathbf{m}_{\mathbf{Y}_Q})^\top] \\
\mathbf{\Sigma}_{\mathbf{Y}_Q\mathbf{Y}_I} &= \text{cov}[\mathbf{Y}_Q, \mathbf{Y}_I] = \text{E}[(\mathbf{Y}_Q - \mathbf{m}_{\mathbf{Y}_Q})(\mathbf{Y}_I - \mathbf{m}_{\mathbf{Y}_I})^\top]
\end{aligned} \tag{3.1}$$

Also define the complex covariance matrices

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{Y}} &= \text{E}\left[(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})^\dagger\right] \\
\tilde{\mathbf{\Sigma}}_{\mathbf{Y}} &= \text{E}\left[(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})^\top\right]
\end{aligned} \tag{3.2}$$

where $\mathbf{\Sigma}_{\mathbf{Y}}$ and $\tilde{\mathbf{\Sigma}}_{\mathbf{Y}}$ are, respectively, the *covariance* and *pseudocovariance* matrices of $\mathbf{Y}$. Note that

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{Y}} &= \left(\mathbf{\Sigma}_{\mathbf{Y}_I} + \mathbf{\Sigma}_{\mathbf{Y}_Q}\right) + j\left(\mathbf{\Sigma}_{\mathbf{Y}_Q\mathbf{Y}_I} - \mathbf{\Sigma}_{\mathbf{Y}_I\mathbf{Y}_Q}\right) \\
\tilde{\mathbf{\Sigma}}_{\mathbf{Y}} &= \left(\mathbf{\Sigma}_{\mathbf{Y}_I} - \mathbf{\Sigma}_{\mathbf{Y}_Q}\right) + j\left(\mathbf{\Sigma}_{\mathbf{Y}_Q\mathbf{Y}_I} + \mathbf{\Sigma}_{\mathbf{Y}_I\mathbf{Y}_Q}\right)
\end{aligned} \tag{3.3}$$

Definition 3.1. $\mathbf{Y}$ is said to be a *proper* complex random vector if pseudocovariance matrix $\tilde{\mathbf{\Sigma}}_Y = 0$, i.e. if

$$\mathbf{\Sigma}_{\mathbf{Y}_I} = \mathbf{\Sigma}_{\mathbf{Y}_Q} \quad \text{and} \quad \mathbf{\Sigma}_{\mathbf{Y}_Q\mathbf{Y}_I} = -\mathbf{\Sigma}_{\mathbf{Y}_I\mathbf{Y}_Q}\,. \tag{3.4}$$

Note that for proper complex $\mathbf{Y}$, it follows from (3.3) that

$$\mathbf{\Sigma}_{\mathbf{Y}} = 2\mathbf{\Sigma}_{\mathbf{Y}_I} + j2\mathbf{\Sigma}_{\mathbf{Y}_Q\mathbf{Y}_I}\,. \tag{3.5}$$

The scalar case

In the special case where $\mathbf{Y}$ is a scalar, denoted by Y, it is clear that

$$\Sigma_{Y_Q Y_I} = \text{E}[(Y_I - m_I)(Y_Q - m_Q)] = \Sigma_{Y_I Y_Q} \tag{3.6}$$

In this case, if Y is proper, which means $\mathrm{E}[(Y - m_Y)^2] = 0$, then $\Sigma_{Y_Q Y_I} = -\Sigma_{Y_I Y_Q}$ together with (3.6) implies that $\Sigma_{Y_Q Y_I} = 0$, i.e. that Y_I and Y_Q are uncorrelated. Furthermore, from (3.5), we get

$$\Sigma_Y = \sigma_Y^2 = \mathrm{E}[|Y - m_Y|^2] = 2\Sigma_{Y_I} = 2\sigma_{Y_I}^2 = 2\sigma_{Y_Q}^2 \,. \tag{3.7}$$

Thus for a complex random scalar Y to be proper, the in-phase and quadrature components must have the same variance and be uncorrelated. Also we see that variance of Y is twice the variance of each of the components.

In the following, we give some general results for proper complex random vectors that also justify the use of the term "proper" in decribing such complex random vectors.

Theorem 3.1. *Let $\mathbf{Y}$ be a proper complex random n-vector, and suppose the random m-vector $\mathbf{Z}$ is defined by*

$$\mathbf{Z} = \mathbf{A}\mathbf{Y} + \mathbf{b} \,. \tag{3.8}$$

Then $\mathbf{Z}$ is also proper complex.

That is, "properness" is preserved under affine transformations.

Theorem 3.2. *Let $\mathbf{Y} = \mathbf{Y}_I + j\mathbf{Y}_Q$ be a proper complex Gaussian vector, i.e. $\mathbf{Y}_I, \mathbf{Y}_Q$ are jointly Gaussian. Then*

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &:= p_{\mathbf{Y}_I \mathbf{Y}_Q}(\mathbf{y}_I, \mathbf{y}_Q) \\ &= \frac{1}{\pi^n \det(\mathbf{\Sigma_Y})} \exp\left\{ -(\mathbf{y} - \mathbf{m_Y})^\dagger \mathbf{\Sigma}_{\mathbf{Y}}^{-1} (\mathbf{y} - \mathbf{m_Y}) \right\} \end{aligned} \tag{3.9}$$

(*Check:* From the fact that $\mathbf{\Sigma_Y}$ is Hermitian and positive definite [1], it follows that $\det(\mathbf{\Sigma_Y})$ is real and positive. For the same reason the quantity inside the exponential is also real and positive.)

Note that (3.9) does not hold if $\mathbf{Y}$ is not proper. Proper complex Gaussian vectors are also called *circularly* complex Gaussian vectors. This is because the pdf of $\mathbf{Y}$ is unchanged if we rotate each component about its mean by some angle θ. That is, $\mathbf{Z} = (\mathbf{Y} - \mathbf{m_Y})e^{j\theta} + \mathbf{m_Y}$ has the same pdf as $\mathbf{Y}$ (prove this!).

In the special case of a proper complex Gaussian scalar Y, the components Y_I and Y_Q are independent Gaussian random variables with variance equal to $\sigma_Y^2/2$, i.e.,

$$p_Y(y) = p_{Y_I Y_Q}(y_I, y_Q) = \frac{1}{\pi \sigma_Y^2} \exp\left\{ -\frac{(y_I - m_I)^2 + (y_Q - m_Q)^2}{\sigma_Y^2} \right\} \,. \tag{3.10}$$

Note that this joint pdf is circularly symmetric about the mean (m_I, m_Q).

Corollary 3.1. *If $\mathbf{Y}$ is a proper complex Gaussian (PCG) vector, then $\mathbf{Z} = \mathbf{A}\mathbf{Y} + \mathbf{b}$ is also a PCG vector.*

Proper Complex Processes

Let $Y(t) = Y_I(t) + jY_Q(t)$ be a complex random process. Parallel to the vector case, we define covariance and pseudocovariance functions:

$$\begin{aligned} C_Y(t+\tau, t) &= \mathrm{E}\left[(Y(t+\tau) - m_Y(t+\tau))(Y(t) - m_Y(t))^\star \right] \\ \tilde{C}_Y(t+\tau, t) &= \mathrm{E}\left[(Y(t+\tau) - m_Y(t+\tau))(Y(t) - m_Y(t)) \right] \end{aligned} \tag{3.11}$$

It is easy to show that

$$\begin{aligned} C_Y(t+\tau,t) &= [C_{Y_I}(t+\tau,t) + C_{Y_Q}(t+\tau,t)] + j[C_{Y_Q Y_I}(t+\tau,t) - C_{Y_I Y_Q}(t+\tau,t)] \\ \tilde{C}_Y(t+\tau,t) &= [C_{Y_I}(t+\tau,t) - C_{Y_Q}(t+\tau,t)] + j[C_{Y_Q Y_I}(t+\tau,t) + C_{Y_I Y_Q}(t+\tau,t)] \end{aligned} \tag{3.12}$$

Definition 3.2. $Y(t)$ is proper complex if the pseudocovariance $\tilde{C}_{YY}(t+\tau,t) = 0$, i.e., if

$$C_{Y_I}(t+\tau,t) = C_{Y_Q}(t+\tau,t) \quad \text{and} \quad C_{Y_Q Y_I}(t+\tau,t) = -C_{Y_I Y_Q}(t+\tau,t) \tag{3.13}$$

Note that for a zero mean process $Y(t)$, the covariance functions C above may be replaced by correlation functions R. Also, if $Y_I(t)$ and $Y_Q(t)$ are jointly WSS processes, then $Y(t)$ is WSS, and $(t+\tau,t)$ in the above equations may be replaced by τ.

For a proper complex process $Y(t)$,

$$C_Y(t+\tau,t) = 2C_{Y_I}(t+\tau,t) + j2C_{Y_Q Y_I}(t+\tau,t)\ . \tag{3.14}$$

Theorem 3.3. *For any n, and any $t_1, t_2, \ldots, t_n$, the samples $Y(t_1), Y(t_2), \ldots, Y(t_n)$ of a proper complex process $Y(t)$ form a proper complex random vector*

Definition 3.3. A proper complex process $Y(t)$ is said to be proper complex Gaussian if, for all n, and all $t_1, t_2, \ldots, t_n$, the samples $Y(t_1), Y(t_2), \ldots, Y(t_n)$ are jointly PCG.

We now give the continuous-time version of Theorem 3.1.

Theorem 3.4. *If a proper complex process $Y(t)$ is passed through a linear (possibly time-varying) system to form*

$$Z(t) = \int_{s=-\infty}^{\infty} h(t,s)Y(s)ds\ . \tag{3.15}$$

Then $Z(t)$ is a proper complex process as well. In addition, if $Y(t)$ is proper complex Gaussian process, then $Z(t)$ a is PCG process as well.

References

[1] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge, New York, 1985.

[2] F. D. Neeser and J. L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. Inform. Th.*, 39(4), July 1993.

4 Lecture 4

Signal Space Concepts

In order to proceed with the design and analysis of digital communication systems (in complex baseband) it is important for us to understand some properties of the space in which the message bearing signal $s(t)$ lies.

Inner Product and Norm

- Let $x(t)$ and $y(t)$ be complex valued signals with $t \in [a, b]$. If a and b are not specified, it is assumed that $t \in (-\infty, \infty)$.

  Definition 4.1. *(Inner Product)*
$$\langle x(t), y(t) \rangle \triangleq \int_a^b x(u) y^*(u) du \, . \tag{4.1}$$

  The inner product satisfies the necessary axioms (see, e.g., [1, Chap 3]):

  ① $\langle x(t), y(t) \rangle = \langle y(t), x(t) \rangle^*$
  ② $\langle x(t) + y(t), z(t) \rangle = \langle x(t), z(t) \rangle + \langle y(t), z(t) \rangle$
  ③ $\langle \alpha x(t), y(t) \rangle = \alpha \langle x(t), y(t) \rangle$
  ④ $\langle x(t), x(t) \rangle \geq 0$, and $\langle x(t), x(t) \rangle = 0$ iff $x(t) = 0$ for all t.

- Signals $x(t)$ and $y(t)$ are said to be orthogonal if $\langle x(t), y(t) \rangle = 0$. The orthogonality of $x(t)$ and $y(t)$ is sometimes denoted by $x(t) \perp y(t)$.

  Definition 4.2. *(Norm)* The inner product defined above induces the following norm:
$$\|x(t)\| = \sqrt{\langle x(t), x(t) \rangle} \, . \tag{4.2}$$

  It is easy to show that the above quantity is a valid norm in that it satisfies the required axioms. (Based on ④ above, all that one needs to verify is the triangle inequality.)

Properties of Inner Product and Norm

① Cauchy-Schwarz Inequality:
$$|\langle x(t), y(t) \rangle| \leq \|x(t)\| \|y(t)\| \tag{4.3}$$
with equality iff $x(t) = \alpha y(t)$ for some complex α.

② Parallelogram Law:
$$\|x(t) + y(t)\|^2 + \|x(t) - y(t)\|^2 = 2\|x(t)\|^2 + 2\|y(t)\|^2 \, . \tag{4.4}$$

③ Pythagorean Theorm: If $x(t) \perp y(t)$ then
$$\|x(t) + y(t)\|^2 = \|x(t)\|^2 + \|y(t)\|^2 \, . \tag{4.5}$$

Signal Space and Basis Functions

- If all we know about the signal $s(t)$ is that it has finite energy, i.e., $\|s(t)\| \leq \infty$, then we can consider $s(t)$ to belong to the (infinite-dimensional) Hilbert space of complex signals with finite energy and with inner product as defined above. This Hilbert space is denoted by $\mathcal{L}_2[a,b]$.

  One can find a (countably infinite) set of functions $\{\psi_i(t)\}_{i=1}^{\infty}$ in $\mathcal{L}_2[a,b]$ that are orthornormal, i.e., $\langle \psi_i(t), \psi_\ell(t) \rangle = \delta_{i\ell}$, such that for any $s(t) \in \mathcal{L}_2[a,b]$, we have

$$s(t) = \sum_{i=1}^{\infty} x_i \psi_i(t) \ . \tag{4.6}$$

  The set $\{\psi_i(t)\}_{i=1}^{\infty}$ is called a *complete basis* for $\mathcal{L}_2[a,b]$

  Example 4.1. On $\mathcal{L}_2[0,T]$, we have the Fourier basis, defined by:

$$\psi_i(t) = \frac{1}{\sqrt{T}} e^{j2\pi it/T} \ , \ i = 0, \pm 1, \pm 2, \ldots \tag{4.7}$$

- Suppose we further impose constraint that the complex baseband signal $s(t)$ is approximately bandlimited to $W/2$ Hz (and time-limited to $[-T/2, T/2]$, say), and impose no other constraints on the signal space. Then the appropriate basis functions for the signal space are the Prolate Spheroidal Wave Functions (PSWF's). See the papers by Slepian, Landau and Pollack [2, 3, 4] for a description of PSWF's. This basis is optimum in the sense that, although there are a countably infinite number of functions in the set, at most WT of these are enough to capture most of the energy for any signal in this signal space. So the signal space of complex signals that are approximately bandlimited to $W/2$ Hz and time limited to $[-T/2, T/2]$ is approximately finite dimensional.

- More typically in communication systems, $s(t)$ is one of M possible signals $s_0(t)$, $s_1(t)$, …, $s_{M-1}(t)$. If we let $\mathcal{S} = \text{span}\{s_0(t), \ldots, s_{M-1}(t)\}$, then $\dim(\mathcal{S}) = n \leq M$. The signal $s(t)$ can then be considered to belong to the n-dim space $\mathcal{S}$. One can find an orthonormal basis for $\mathcal{S}$ by the standard *Gram-Schmidt procedure*:

$$\phi_0(t) = s_0(t), \quad \psi_0(u) = \begin{cases} \frac{\phi_0(u)}{\|\phi_0(t)\|} & \text{if } \|\phi_0(t)\| \neq 0 \\ \text{stop} & \text{otherwise} \end{cases} \tag{4.8}$$

$$\phi_1(u) = s_1(u) - \langle s_1(t), \psi_0(t) \rangle \psi_0(u), \quad \psi_1(u) = \begin{cases} \frac{\phi_1(u)}{\|\phi_1(t)\|} & \text{if } \|\phi_1(t)\| \neq 0 \\ \text{stop} & \text{otherwise} \end{cases} \tag{4.9}$$

$$\phi_\ell(u) = s_\ell(u) - \sum_{i=0}^{\ell-1} \langle s_\ell(t), \psi_i(t) \rangle \psi_i(u), \quad \psi_\ell(u) = \begin{cases} \frac{\phi_\ell(u)}{\|\phi_\ell(t)\|} & \text{if } \|\phi_\ell(t)\| \neq 0 \\ \text{stop} & \text{otherwise} \end{cases} \tag{4.10}$$

- For signal $s(t) \in \mathcal{S}$, we can write

$$s(t) = \sum_{\ell=0}^{n} s_\ell \psi_\ell(t) \ , \ \text{ with } s_\ell = \langle s(t), \psi_\ell(t) \rangle \ . \tag{4.11}$$

  The signal $s(t) \in \mathcal{S}$ is equivalent to the vector $\mathbf{s} = [s_1 \ s_2 \cdots s_n]$ in the sense that

$$\|s(t)\| = \sqrt{\mathbf{s}^\dagger \mathbf{s}} = \|\mathbf{s}\| \ \text{ (show this!)} \tag{4.12}$$

and for $s_k(t), s_m(t) \in \mathcal{S}$

$$\langle s_k(t), s_m(t)\rangle = \mathbf{s}_m^\dagger \mathbf{s}_k = \langle \mathbf{s}_k, \mathbf{s}_m\rangle \ \text{(show this!)}\ . \tag{4.13}$$

Signal Energy, Correlation and Distance

- The *energy* of a signal $s(t)$ is denoted by $\mathcal{E}$ and is given by

$$\mathcal{E} = \|s(t)\|^2\ . \tag{4.14}$$

- The *correlation* between two signals $s_k(t)$ and $s_m(t)$, which is a measure of the similarity between these two signals, is given by

$$\rho_{km} = \frac{\langle s_k(t), s_m(t)\rangle}{\|s_k(t)\|\|s_m(t)\|} = \frac{\langle s_k(t), s_m(t)\rangle}{\sqrt{\mathcal{E}_k\mathcal{E}_m}}\ . \tag{4.15}$$

- The *distance* between two signals $s_k(t)$ and $s_m(t)$, which is also a measure of the similarity between these two signals, is given by

$$d_{km} = \|s_k(t) - s_m(t)\| = \left(\mathcal{E}_k + \mathcal{E}_m - 2\sqrt{\mathcal{E}_k\mathcal{E}_m}\mathrm{Re}[\rho_{k,m}]\right)^{\frac{1}{2}}\ . \tag{4.16}$$

  If $\mathcal{E}_k = \mathcal{E}_m = \mathcal{E}$, then

$$d_{k,m} = [2\mathcal{E}(1 - \mathrm{Re}[\rho_{k,m}])]^{\frac{1}{2}}\ . \tag{4.17}$$

References

[1] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969.

[2] D. Slepian and H. O. Pollack. Prolate Spheroidal Wave Functions, Fourier analysis and uncertainty-I. *Bell Syst. Tech. J.*, pages 43–63, January 1961.

[3] H. J. Landau and H. O. Pollack. Prolate Spheroidal Wave Functions, Fourier analysis and uncertainty-II. *Bell Syst. Tech. J.*, pages 64–85, January 1961.

[4] H. J. Landau and H. O. Pollack. PSWFs-III: The dimension of the space of essentially time- and band-limted signals. *Bell Syst. Tech. J.*, pages 1295–1320, July 1962.

[5] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

5 Lecture 5

Digital Modulation

- After possible source and error control encoding, we have a sequence $\{m_n\}$ of message symbols to be transmitted on the channel. The message symbols are assumed to come from a finite alphabet, say $\{0, 1, \ldots, M-1\}$. In the simplest case of binary signaling, $M = 2$. Each symbol in the sequence is assigned to one of M waveforms $\{s_0(t), \ldots, s_{M-1}(t)\}$.

- *Memoryless modulation versus modulation with memory.* If the symbol to waveform mapping is fixed from one interval to the next, i.e., $m \mapsto s_m(t)$, then the modulation is memoryless. If the mapping from symbol to waveform in the n-th symbol interval depends on previously transmitted symbols (or waveforms) then the modulation is said to have memory.

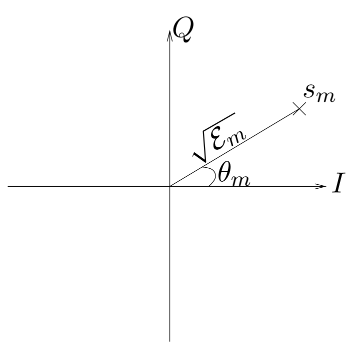
- For memoryless modulation, to send the sequence $\{m_n\}$ of symbols at the rate of $1/T_\mathrm{s}$ symbols per second, we transmit the signal
$$s(t) = \sum_n s_{m_n}(t - nT_\mathrm{s}) \ . \tag{5.1}$$

- *Linear versus nonlinear modulation.* A digital modulation scheme is said to be linear if we can write the mapping from the sequence of symbols $\{m_n\}$ to the transmitted signal $s(t)$ as concatenation of a mapping from the sequence $\{m_n\}$ to a complex sequence $\{c_n\}$, followed by a *linear* mapping from $\{c_n\}$ to $s(t)$. Otherwise the modulation is nonlinear.

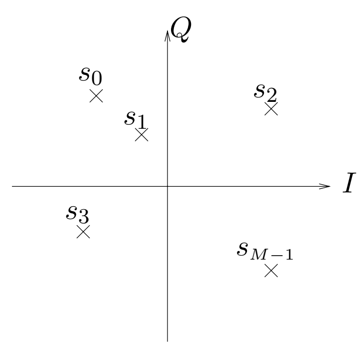
Linear Memoryless Modulation

- In this case, the mapping from symbols to waveforms can be written in complex baseband as:
$$s_m(t) = \sqrt{\mathcal{E}_m}\, e^{j\theta_m} g(t) \ , \ \ m = 0, 1, \ldots, M-1 \ , \tag{5.2}$$
where $g(t)$ is a real-valued, unit energy, pulse shaping waveform.

- The signal $s_m(t)$ can be represented by a point in the complex plane, i.e., the signal space corresponding to a symbol interval is a 1-d (complex) space with basis function $g(t)$.



Representation of $s_m(t)$



Signal constellation

- In real passband,
$$\tilde{s}_m(t) = \mathrm{Re}[\sqrt{2} s_m(t)\, e^{j2\pi f_c t}] = \sqrt{2\mathcal{E}_m}\ \cos(2\pi f_c t + \theta_m)\ . \tag{5.3}$$

- It is easy to see that the signal energy is the same in both the real passband and complex baseband domains and equals $\mathcal{E}_m$.

- The average symbol energy for the constellation is given by
$$\mathcal{E}_s = \frac{1}{M} \sum_{m=0}^{M-1} \mathcal{E}_m\ . \tag{5.4}$$

- The average bit energy for the constellation (assuming that $M = 2^\nu$, for some integer ν) is given by
$$\mathcal{E}_\text{b} = \frac{\mathcal{E}_\text{s}}{\log_2 M} = \frac{\mathcal{E}_\text{s}}{\nu}\ . \tag{5.5}$$

- The distance between signals s_k and s_m is $d_{k,m} = \|s_k - s_m\|$, and the minimum distance is given by
$$d_\text{min} = \min_{k,m} d_{k,m}\ . \tag{5.6}$$

- A measure of goodness of the constellation is the ratio
$$\zeta = \frac{d^2_\text{min}}{\mathcal{E}_\text{b}}\ . \tag{5.7}$$
Note that ζ is independent of scaling of the constellation.

- Some commonly used signal constellations are:
  - *Pulse Amplitude Modulation (PAM).* Information only in amplitude:
  $$\theta_m = 0 \text{ and } \sqrt{\mathcal{E}_m} = (2m + 1 - M)\frac{d}{2},\ \ m = 0, 1, \ldots, M-1\ . \tag{5.8}$$
  We can compute ζ as a function of M. For example, $\zeta = 4$ for $M = 2$.
  - *Phase Modulation or Phase Shift Keying (PSK).* Information only in phase:
  $$\theta_m = \frac{2\pi m}{M} \text{ and } \sqrt{\mathcal{E}_m} = \mathcal{E}\ ,\ \ m = 0, 1, \ldots, M-1\ . \tag{5.9}$$
  For QPSK, $\zeta = 4$ (as in BPSK).
  - *Quadrature Amplitude Modulation (QAM).* Information in phase and amplitude. We can design constellations to maximize ζ for a given M. Rectangular constellations are convenient for demodulation. For rectangular 16-QAM, $\zeta = 1.6$.

Orthogonal Memoryless Modulation

- Here the signal set is given by

$$s_m(t) = \sqrt{\mathcal{E}}\, g_m(t), \quad m = 0, 1, \ldots, M-1 \tag{5.10}$$

where $\{g_m(t)\}$ are (possibly complex) unit energy signals, i.e., $\|g_m(t)\| = 1$.

- The correlation between signals $s_k(t)$ and $s_m(t)$ is given by:

$$\rho_{km} = \frac{\langle s_k(t), s_m(t)\rangle}{\mathcal{E}} = \langle g_k(t), g_m(t)\rangle \tag{5.11}$$

- There are two kinds of orthogonality:
  - ○ Orthogonality only in the real component of the correlation, i.e. Re$\{\rho_{km}\} = 0$, for $k \neq m$. This form of orthogonality is enough for coherent demodulation.
  - ○ Complete orthogonality, i.e., $\rho_{km} = 0$, for $k \neq m$. This is required for noncoherent demodulation.

- *Examples of orthogonal signal sets*
  - ○ *Separation in time:*

$$g_m(t) = g\left(t - mT_\text{s}/M\right) \tag{5.12}$$

where $g(t)$ is such that $\langle g(t - kT_\text{s}/M), g(t - mT_\text{s}/M)\rangle = \delta_{km}$. For example, $g(t) = p_{T_\text{s}/M}(t)$, a rectangular pulse of width T_s/M.

This signal set is completely orthogonal. We can also create a signal set of twice the size which satisfies orthogonality only in the real component of the correlation by adding $\{jg_m(t)\}$ to the above signal set.

  - ○ *Separation in frequency:*

$$g_m(t) = e^{j2\pi m\Delta_f t}\, p_{T_\text{s}}(t) \tag{5.13}$$

It is easy to show that

$$\rho_{km} = \text{sinc}[T_\text{s}(k-m)\Delta_f]\; e^{j\pi T_\text{s}(m-k)\Delta_f} \tag{5.14}$$

and that

$$\text{Re}\{\rho_{km}\} = \text{sinc}[2T_\text{s}(k-m)\Delta_f]\ . \tag{5.15}$$

Thus the smallest value of Δ_f such that $\rho_{km} = 0$, for $k \neq m$, is $1/T$, and such that Re$\{\rho_{km}\} = 0$, for $k \neq m$, is $1/2T$.

  - ○ *Separation in time and frequency:* One way to do this is to pick $\{g_m(t)\}$ to be the Walsh functions on $[0, T_s]$ (see, e.g., [1, page 424]).

References

[1] G. Stuber. *Principles of Mobile Communication*. Kluwer Academic, Norwell, MA, 1996.

[2] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

6 Lecture 6

Continuous Phase Modulation (CPM)

- CPM is a form of nonlinear modulation with memory. A natural way to introduce CPM is via memoryless frequency shift keying (FSK) orthogonal modulation.

- Memoryless FSK is a formed by linear scaling of the carrier frequency based on the symbol. If we map the symbol $m \in \{0, 1, \ldots, M-1\}$ in a one-one manner to $B \in \{\pm 1, \pm 2, \ldots, \pm(M-1)\}$, then

$$s_m(t) = \sqrt{\mathcal{E}} e^{j\pi\Delta_f Bt} \,. \tag{6.1}$$

  To send the sequence $\{m_n\}$, we transmit the signal:

$$s(t) = \sum_n \sqrt{\mathcal{E}} e^{j\pi\Delta_f B_n(t-nT_s)} \,. \tag{6.2}$$

  The problem with memorlyless FSK is that abrupt frequency switching from one symbol interval to the next can result in large spectral side lobes outside the main spectral lobe in the power spectrum of $s(t)$. The solution to this problem is continuous phase FSK (CPFSK)

CPFSK

- Start with a real baseband PAM signal

$$d(t) = \sum_n B_n v(t - nT_s) \tag{6.3}$$

  where $B_n \in \{\pm 1, \pm 2, \ldots, \pm(M-1)\}$, and

$$v(t) = \frac{1}{2T_s} p_{T_s}(t) \tag{6.4}$$

- Use $d(t)$ to frequency modulate the carrier to form:

$$s(t) = \sqrt{\frac{\mathcal{E}}{T_s}} \exp\left\{ j \left[4\pi T_s f_d \int_{-\infty}^{t} d(\tau) d\tau + \phi_0 \right] \right\} \tag{6.5}$$

  where f_d is the frequency deviation factor and ϕ_0 is the initial phase.

- The phase of $s(t)$ is continuous and is given by

$$\phi(t; \mathbf{B}) = 4\pi T_s f_d \int_{-\infty}^{t} d(\tau) \mathrm{d}\tau = 4\pi T_s f_d \int_{-\infty}^{t} \left[\sum_n B_n v(\tau - nT_s) \right] d\tau \,. \tag{6.6}$$

  For $t \in [nT_s, (n+1)T_s]$, we can simplify the above expression as

$$\phi(t; \mathbf{B}) = \theta_n + 2\pi h B_n \frac{(t - nT_s)}{2T_s} \tag{6.7}$$

where the phase at time nT_s, θ_n, is given by

$$\theta_n = \phi(nT_\mathrm{s}; \mathbf{B}) = \pi h \sum_{k=-\infty}^{n} B_k \tag{6.8}$$

and the modulation index, h, is given by $h = 2f_d T_\mathrm{s}$.

- Equation (6.7) can be rewritten as:

$$\phi(t; \mathbf{B}) = 2\pi h \sum_{k=-\infty}^{\infty} B_k q(t - kT_\mathrm{s}) \tag{6.9}$$

where

$$q(t) = \frac{t}{2T_\mathrm{s}} \, \mathbb{1}_{\{t \in [0,T_\mathrm{s}]\}} + \frac{1}{2} \, \mathbb{1}_{\{t > T_\mathrm{s}\}} \, . \tag{6.10}$$

CPM as a Generalization of CPFSK

- Based on (6.9), we can generalize CPFSK to a continuous phase modulation scheme in which the carrier phase is varied as:

$$\phi(t; \mathbf{B}) = 2\pi \sum_{k=-\infty}^{\infty} B_k h_k q(t - kT_\mathrm{s}) \tag{6.11}$$

where h_k could be constant with k or varied cyclically between a finite set of values, and

$$q(t) = \int_{-\infty}^{t} v(\tau) d\tau = \int_{0}^{t} v(\tau) d\tau \tag{6.12}$$

with $v(t)$ being a causal signal that is normalized so that $\int_0^\infty v(\tau) d\tau = 1/2$.

- *Full response versus partial response*

  ◦ If $v(t)$ has support $[0, T_\mathrm{s}]$, then it is said to be a full response pulse. For example, the rectangular full response pulse is given by

$$v(t) = \frac{1}{2T_\mathrm{s}} p_{T_\mathrm{s}}(t) \ , \ \text{ with } q(t) = \frac{t}{2T_\mathrm{s}} \, \mathbb{1}_{\{t \in [0,T_\mathrm{s}]\}} + \frac{1}{2} \, \mathbb{1}_{\{t > T_\mathrm{s}\}} \, . \tag{6.13}$$

  The time domain raised cosine (TDRC) full response pulse is given by:

$$v(t) = \frac{1}{2T_\mathrm{s}} \left[1 - \cos\left(\frac{2\pi t}{T_\mathrm{s}} \right) \right] p_{T_\mathrm{s}}(t) \tag{6.14}$$

  with

$$q(t) = \left[\frac{t}{2T_s} - \frac{1}{4\pi} \sin\left(\frac{2\pi t}{T_\mathrm{s}} \right) \right] \mathbb{1}_{\{t \in [0,T_\mathrm{s}]\}} + \frac{1}{2} \, \mathbb{1}_{\{t > T_\mathrm{s}\}} \, . \tag{6.15}$$

  ◦ If $v(t)$ is of support $[0, LT_\mathrm{s}]$, where $L > 1$, the modulation is said to be partial response. For example, the rectangular partial response pulse with $L = 2$ is given by

$$v(t) = \frac{1}{4T_\mathrm{s}} p_{2T_\mathrm{s}}(t) \ , \ \text{ with } q(t) = \frac{t}{4T_\mathrm{s}} \, \mathbb{1}_{\{t \in [0,2T_\mathrm{s}]\}} + \frac{1}{2} \, \mathbb{1}_{\{t > 2T_\mathrm{s}\}} \, . \tag{6.16}$$

Phase Trajectories for CPM

- The phase trajectory for CPM is a plot of the phase ϕ as a function of time for all possible sequences $\{B_n\}$.
- Starting with $\phi(0, \mathbf{B}) = 0$, we can draw a tree with edges corresponding to the values taken by the B_n's. This is called a *phase tree*. If $M = 2$ we have a binary tree.
- If we draw the phase tree with the phase modulo $[-\pi, \pi]$, then the tree collapses to a trellis.
- For a full response rectangular pulse with $h_k = h$, the edges are straight lines (see [1, Pg. 194]) and:

$$\phi(t; \mathbf{B}) = \pi h \sum_{k=-\infty}^{n-1} B_k + \pi h B_n \left(\frac{t - nT_\mathrm{s}}{T_\mathrm{s}} \right) , \quad \text{for } t \in [nT_\mathrm{s}, (n+1)T_\mathrm{s}] . \tag{6.17}$$

  For a full response TDRC pulse, the edges are sinusoidal curves.

- For a $L = 2$ partial response rectangular pulse (see (6.16)) with $h = h_k$,

$$\phi(t; \mathbf{B}) = \pi h \sum_{k=-\infty}^{n-2} B_k + \frac{\pi h (t - (n-1)T_\mathrm{s}) B_{n-1}}{2T_\mathrm{s}} + \frac{\pi h (t - nT_\mathrm{s}) B_n}{2T_\mathrm{s}} , \quad \text{for } t \in [nT_\mathrm{s}, (n+1)T_\mathrm{s}] . \tag{6.18}$$

- *State Trellis.* A simpler description of the phase trajectories is given in terms of a state trellis, in which only the transitions between phase states at the symbol boundaries is drawn. For the example of (6.17)

$$\theta_n = \pi h \sum_{k=-\infty}^{n-1} B_k . \tag{6.19}$$

  For rational $h = m/p$, it is of interest to list the set of possible states in the state trellis. It is easy to show that if m is even

$$\Theta_\mathrm{s} = \left\{ 0, \pm\frac{\pi m}{p}, \pm\frac{2\pi m}{p}, \ldots, \pm\frac{(p-1)\pi m}{2p} \right\} \tag{6.20}$$

  and we have a total of p states. If m is odd, we have a total of $2p$ states and

$$\Theta_\mathrm{s} = \left\{ 0, \pm\frac{\pi m}{p}, \pm\frac{2\pi m}{p}, \ldots, \pm\frac{(2p-1)\pi m}{2p}, \pi \right\} . \tag{6.21}$$

  For the example of (6.18), with $h = m/p$,

$$\theta_n = \pi h \sum_{k=-\infty}^{n-2} B_k + \frac{\pi h B_{n-1}}{2} . \tag{6.22}$$

  Here we have a maximum of pM states if m is even, and a maximum of $2pM$ states if m is odd. In general, for partial response covering L symbol intervals,

$$\text{max \# states} = \begin{cases} pM^{L-1} & \text{if } m \text{ is even} \\ 2pM^{L-1} & \text{if } m \text{ is odd} \end{cases} \tag{6.23}$$

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

7 Lecture 7

Minimum Shift Keying (MSK)

- MSK is a special case of CPFSK with $M = 2$ and $h = 1/2$. Note that, as in any CPFSK modulation,

$$q(t) = \frac{t}{2T_s} \mathbb{1}_{\{t \in [0,T_s]\}} + \frac{1}{2} \mathbb{1}_{\{t > T_s\}} . \tag{7.1}$$

- The phase of the MSK signal is given by:

$$\phi(t; \mathbf{B}) = \frac{\pi}{2} \sum_{k=-\infty}^{n-1} B_k + \pi B_n \left(\frac{t - nT_s}{2T_s} \right) , \quad \text{for } t \in [nT_s, (n+1)T_s] . \tag{7.2}$$

- The frequency in interval $[nT_s, (n+1)T_s]$ is given by

$$f_n = \frac{1}{2\pi} \frac{\pi B_n}{2T_s} = \frac{B_n}{4T_s} = \pm \frac{1}{T_s} . \tag{7.3}$$

  The frequency difference $\Delta_f = \frac{1}{2T_s}$ is the smallest frequency separation for orthogonality in a symbol period. Hence the name *minimum* shift keying.

- MSK can also be considered to be a special case of *offset* QPSK (or OQPSK)[2]. It can be shown that:

$$\begin{aligned} s(t) &= \sqrt{\frac{\mathcal{E}}{T_s}} \exp\left(j\pi \sum_{k=-\infty}^{\infty} B_k q(t - kT_s) \right) \\ &= \sqrt{\frac{\mathcal{E}}{2}} \sum_{n=-\infty}^{\infty} [B_{2n} g(t - 2nT_s) + jB_{2n+1} g(t - 2nT_s - T_s)] \\ &= \sqrt{\frac{\mathcal{E}}{2}} \sum_{n=-\infty}^{\infty} [B_{2n} g(t - 2nT_b) + jB_{2n+1} g(t - 2nT_b - T_b)] \end{aligned} \tag{7.4}$$

  where

$$g(t) = \frac{1}{\sqrt{T_b}} \sin\left(\frac{\pi t}{2T_b} \right) p_{2T_b}(t) . \tag{7.5}$$

Power Spectra of Digitally Modulated Signals

- Consider the linearly modulated signal of the form

$$s(t) = \sum_{n=-\infty}^{\infty} B_n g(t - nT_s) , \tag{7.6}$$

  where $\{B_n\}$ is a complex sequence produced by mapping the symbol sequence $\{m_n\}$ to the complex plane.

[2]Note that $T_s = T_b$ for MSK since it is a form of binary modulation. So the comparison with OQPSK made by fixing T_b, with the understanding that $T_s = 2T_b$ for OQPSK.

- If we model $\{m_n\}$ as a random sequence, then $\{B_n\}$ is a random complex sequence. We assume that $\{B_n\}$ is WSS (discrete-time) process with mean $\mu_B = \mathrm{E}[B_n]$, ACF $R_B[k] = \mathrm{E}\left[B_{n+k}B_n^\star\right]$, and PSD

$$S_B(f) = \sum_{k=-\infty}^{\infty} R_B[k]\, e^{-j2\pi fk} \,. \tag{7.7}$$

- The modulated signal $s(t)$ is a random process with mean

$$\mu_s(t) = \mathrm{E}[s(t)] = \mu_B \sum_{n=-\infty}^{\infty} g(t - nT_\mathrm{s}) \tag{7.8}$$

and ACF, which can be shown to equal

$$R_s(t+\tau, t) = \mathrm{E}[s(t+\tau)s^\star(t)] = \sum_{\ell=-\infty}^{\infty} R_B[\ell] \sum_{n=-\infty}^{\infty} g(t+\tau-\ell T_\mathrm{s} - nT_\mathrm{s})g(t - nT_\mathrm{s}) \,. \tag{7.9}$$

Since $\mu_s(t)$ and $R_s(t+\tau,t)$ are periodic in t with period T_s, $s(t)$ is a *cyclostationary* process.

- The PSD of the cyclostationary process $s(t)$ is given by the Fourier transform of the average ACF

$$\bar{R}_s(\tau) = \frac{1}{T_\mathrm{s}} \int_{-T_\mathrm{s}/2}^{T_\mathrm{s}/2} R_s(t+\tau, t)dt \,. \tag{7.10}$$

- It is easy to show based on (7.9) that

$$\bar{R}_s(\tau) = \sum_{\ell=-\infty}^{\infty} R_B[\ell] \frac{1}{T_\mathrm{s}} \int_{\infty}^{\infty} g(t+\tau-\ell T_\mathrm{s})g(t)dt \,. \tag{7.11}$$

- If we define the time ACF of the deterministic function $g(t)$ by

$$R_g(\tau) = \int_{-\infty}^{\infty} g(t+\tau)g(t)dt \tag{7.12}$$

then it is easy to show that

$$S_g(f) = \mathcal{F}\left[R_g(\tau)\right] = |G(f)|^2 \,. \tag{7.13}$$

- Based on (7.11) and (7.12), we obtain

$$\bar{R}_s(\tau) = \frac{1}{T_\mathrm{s}} \sum_{\ell=-\infty}^{\infty} R_B[\ell] R_g(\tau - \ell T_\mathrm{s}) \tag{7.14}$$

and hence

$$\boxed{S_s(f) = \mathcal{F}\left[\bar{R}_s(\tau)\right] = \frac{1}{T_\mathrm{s}}\, S_B(fT_\mathrm{s})\, |G(f)|^2} \,. \tag{7.15}$$

Special Case: Uncorrelated Symbols

◦ The sequence $\{B_n\}$ is uncorrelated with each element having mean μ_B and variance σ_B^2. Then

$$R_B[k] = \begin{cases} \sigma_B^2 + \mu_B^2 & \text{if } k = 0 \\ \mu_B^2 & \text{if } k \neq 0 \end{cases} = \mu_B^2 + \sigma_B^2 \delta[k] \, . \tag{7.16}$$

◦ Thus

$$S_B(f) = \sigma_B^2 + \mu_B^2 \sum_{k=-\infty}^{\infty} e^{-j2\pi f k} \, . \tag{7.17}$$

◦ *Poisson Sum Formula*

$$\sum_{k=-\infty}^{\infty} e^{-j2\pi f T_s k} = \frac{1}{T_\text{s}} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_\text{s}}\right) \, . \tag{7.18}$$

◦ From (7.15), (7.17) and (7.18), we obtain

$$S_s(f) = \frac{1}{T_\text{s}} \sigma_B^2 |G(f)|^2 + \frac{\mu_B^2}{T_\text{s}^2} \sum_{k=-\infty}^{\infty} \left| G\left(\frac{k}{T_\text{s}}\right)\right|^2 \delta\left(f - \frac{k}{T_\text{s}}\right) \tag{7.19}$$

The second term in PSD corresponds to lines in the spectrum at the fundamental frequency $1/T_\text{s}$. We can suppress these spectral lines by forcing B_n to be zero mean. This is done in practice by ensuring that the symbols are equally likely and symmetrical positioned around 0 in the complex plane.

◦ For the special case of MPSK, it is easy see that, for all M,

$$S_s(f) = \frac{\mathcal{E}}{T_\text{s}} |G(f)|^2 \tag{7.20}$$

Generalization of PSD Analysis

• Consider linear modulation with memory of the form where $\{B_n\}$ is first passed through a linear filter, with transfer function $H(z)$, to produce a new complex sequence $\{A_n\}$. The sequence $\{A_n\}$ is modulated to form:

$$s(t) = \sum_{n=-\infty}^{\infty} A_n g(t - nT_\text{s}) \, . \tag{7.21}$$

The PSD of $s(t)$ can be controlled by $H(z)$ (see problem 6 of HW#2).

• For offset QPSK and MSK,

$$s(t) = \sqrt{\frac{\mathcal{E}}{2}} \sum_{n=-\infty}^{\infty} [B_{2n} g(t - 2nT_\text{b}) + jB_{2n+1} g(t - 2nT_\text{b} - T_\text{b})] \tag{7.22}$$

with $B_n \in \{+1, -1\}$. If $\{B_n\}$ is an uncorrelated sequence with $\text{P}\{B_n = +1\} = \text{P}\{B_n = -1\} = 1/2$, then it easy to show that

$$\bar{R}_s(\tau) = \frac{\mathcal{E}}{2} \sum_n g(t - nT_\text{b}) g(t + \tau - nT_\text{b}) \tag{7.23}$$

Thus $s(t)$ is cyclostationary with period T_b in this case, and following steps similar to those used above for memoryless linear modulation we get

$$S_s(f) = \frac{\mathcal{E}}{2T_\mathrm{b}} |G(f)|^2 = \frac{\mathcal{E}}{T_\mathrm{s}} |G(f)|^2 \text{ (for OQPSK)} . \tag{7.24}$$

Thus OQPSK has no spectral advantage over QPSK. However, as we discussed in class OQPSK is used in practice to avoid abrupt zero-crossings in the passband signal.

- The PSD analysis for CPM is considerably more complicated (see [1, Section 4.4.2] for details).

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

8 Lecture 8

Likelihood Functions and Optimum Detection/Estimation

- *Estimation Problem:* Information about unknown parameter $\lambda \in \mathcal{S}_\lambda$ is available through a random observation Y whose probability distribution depends on λ. The goal is to estimate λ from Y. If card($\mathcal{S}_\lambda$) is finite, then the estimation problem is called a *detection* problem.

- *Likelihood function:* The pdf of Y when the parameter value is λ is denoted by $p_\lambda(y)$ and is called the likelihood function for λ. There are two ways to interpret $p_\lambda(y)$. If we consider λ to be a deterministic but unknown parameter, then $p_\lambda(y)$ is simply a member of the family of densities $\{p_\lambda(y), \lambda \in \mathcal{S}_\lambda\}$. If λ is considered to be a realization of a random variable Λ, then $p_\lambda(y)$ equals the conditional pdf $p_{Y|\Lambda}(y|\lambda)$.

- The *maximum likelihood (ML) estimate* of λ given $Y = y$ is given by

$$\hat{\lambda}_{\mathrm{ML}}(y) = \arg\max_{\lambda} p_\lambda(y) \,. \tag{8.1}$$

  For example, if $Y \sim \mathcal{N}(\lambda, \sigma^2)$, then it is easy to see that $\hat{\lambda}_{\mathrm{ML}}(y) = y$.

- If the observation is a random vector $\mathbf{Y}$ and the parameter is a vector $\boldsymbol{\lambda}$, then the likelihood function is $p_{\boldsymbol{\lambda}}(\mathbf{y})$, and the joint ML estimate of $\boldsymbol{\lambda}$ given $\mathbf{y}$ is:

$$\hat{\boldsymbol{\lambda}}_{\mathrm{ML}}(\mathbf{y}) = \arg\max_{\boldsymbol{\lambda}} p_{\boldsymbol{\lambda}}(\mathbf{y}) \,. \tag{8.2}$$

  For example, if $\mathbf{Y} = [Y_1 \; Y_2 \; \cdots \; Y_n]$ has components that are i.i.d. Gaussian with unknown mean μ and unknown variance σ^2, then it is easy to show that joint ML estimates of μ and σ^2 are simply the sample mean and samlple variance, respectively.

Bayesian Estimation

- If the parameter is assumed to be random with known prior distribution $p_\Lambda(\lambda)$, then the estimation procedure is said to be *Bayesian*. There are many forms of Bayesian estimators. Two important ones are given below.

- *Maximum A Posteriori (MAP):*

$$\hat{\lambda}_{\mathrm{MAP}}(y) = \arg\max_{\lambda} p_{\Lambda|Y}(\lambda|y) = \arg\max_{\lambda} p_{Y|\Lambda}(y|\lambda) p_\Lambda(\lambda) = \arg\max_{\lambda} p_\lambda(y) p_\Lambda(\lambda) \,. \tag{8.3}$$

  If $p_\Lambda(\lambda)$ is uniform on $\mathcal{S}_\lambda$ (this may not be possible for some $\mathcal{S}_\lambda$), then $\hat{\lambda}_{\mathrm{MAP}}(y) = \hat{\lambda}_{\mathrm{ML}}(y)$.

- *Minimum Mean Squared Error (MMSE):* Assuming that λ belongs to a Hilbert space, $\hat{\lambda}_{\mathrm{MMSE}}(y)$ is the estimator that minimizes $\mathrm{E}[\|\hat{\lambda}(Y) - \Lambda\|^2]$. It can be shown that [1, Pg. 143]

$$\hat{\lambda}_{\mathrm{MMSE}}(y) = \mathrm{E}\left[\Lambda \mid Y = y\right] \,. \tag{8.4}$$

  (More about MMSE estimators later in the course.)

- *Minimum Probability of Error (MPE).* If $\mathrm{card}(\mathcal{S}_\lambda) = M < \infty$, then without loss of generality we can consider $\mathcal{S}_\lambda = \{0, 1, \ldots, M-1\}$. Let $\pi_m = \mathrm{P}\{\Lambda = m\}$. We can define the probability of error as:

$$\mathrm{P_e} = \mathrm{P}\{\hat{\lambda}(Y) \neq \Lambda\} = \sum_{m=0}^{M-1} \pi_m \mathrm{P}(\{\hat{\lambda}(Y) \neq m\} | \{\Lambda = m\}) \; . \tag{8.5}$$

  Then $\hat{\lambda}_{\mathrm{MPE}}(y)$ is the estimator that minimizes $\mathrm{P_e}$

- *Relationship between MAP and MPE Estimators.* The probability of correct decisions is given by:

$$\mathrm{P_c} = 1 - \mathrm{P_e} = \sum_{m=0}^{M-1} \pi_m \mathrm{P}(\{\hat{\lambda}(Y) = m\} | \{\Lambda = m\}) = \sum_{m=0}^{M-1} \int_{\Gamma_m} \pi_m p_m(y) dy \tag{8.6}$$

  where Γ_m is the decision region for parameter value m. It is clear that $\mathrm{P_c}$ is maximized by placing y in decision region Γ_m if $\pi_m p_m(y)$ is larger than $\pi_j p_j(y)$ for all $j \neq m$. Thus

$$\hat{\lambda}_{\mathrm{MPE}}(y) = \arg\max_\lambda \pi_\lambda p_\lambda(y) = \hat{\lambda}_{\mathrm{MAP}}(y) \; . \tag{8.7}$$

Examples

○ Suppose the likelihood function is given by:

$$p_\lambda(y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{|y - \mu_\lambda|^2}{2\sigma^2} \right] \; . \tag{8.8}$$

Then

$$\hat{\lambda}_{\mathrm{ML}}(y) = \arg\max_\lambda p_\lambda(y) = \arg\min_\lambda |y - \mu_\lambda|^2 \tag{8.9}$$

$$\hat{\lambda}_{\mathrm{MPE}}(y) = \hat{\lambda}_{\mathrm{MAP}}(y) = \arg\max_\lambda \log \pi_\lambda - \frac{|y - \mu_\lambda|^2}{2\sigma^2} \; . \tag{8.10}$$

○ In the special case of binary signaling, i.e., $\mathcal{S}_\lambda = \{0, 1\}$,

$$\hat{\lambda}_{\mathrm{ML}}(y) = \arg\max_\lambda p_\lambda(y) = \begin{cases} 1 & \text{if } p_1(y) \geq p_0(y) \\ 0 & \text{otherwise} \end{cases} \tag{8.11}$$

This detector can be rewritten in the form:

$$\hat{\lambda}_{\mathrm{ML}}(y) = \begin{cases} 1 & \text{if } L(y) \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{8.12}$$

where $L(y) = p_1(y)/p_0(y)$ is called the *likelihood ratio* of the observations. The optimum detector is a special case of a *likelihood ratio test* (LRT).

References

[1] H. V. Poor. *An Introduction to Signal Detection and Estimation, 2nd Edition.* Springer-Verlag, New York, 1994.

9 Lecture 9

Log-likelihood Ratio for Detection/Estimation in Discrete-Time WGN

- Suppose the observation is a vector $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]$, with

$$Y_k = s_{\lambda,k} + w_k \ , \quad k = 1, \dots, n \tag{9.1}$$

  where $\{w_k\}$ are i.i.d. $\mathcal{CN}(0, 2\sigma^2)$

- The likelihood function is given by:

$$p_\lambda(\mathbf{y}) = \prod_{k=1}^{n} \frac{1}{2\pi\sigma^2} \, \exp\left[-\frac{|y_k - s_{\lambda,k}|^2}{2\sigma^2}\right] \ . \tag{9.2}$$

- For the purposes of optimum detection/estimation based on $\mathbf{Y}$, it is okay to divide $p_\lambda(\mathbf{y})$ by the "noise only" pdf

$$\tilde{p}(\mathbf{y}) = \prod_{k=1}^{n} \frac{1}{2\pi\sigma^2} \, \exp\left[-\frac{|y_k|^2}{2\sigma^2}\right] \ . \tag{9.3}$$

  This division yields the *likelihood ratio*

$$L_\lambda(\mathbf{y}) = \prod_{k=1}^{n} \exp\left[-\frac{|y_k - s_{\lambda,k}|^2 - |y_k|^2}{2\sigma^2}\right] \ . \tag{9.4}$$

  (Note that $L(\mathbf{y})$ is the Radon-Nikodym derivative of measure P_λ w.r.t. the measure $\tilde{P}$ [1, Page 443].)

- Taking the logarithm of $L(\mathbf{y})$, we get an equivalent likelihood function, which simplifies to

$$\log L_\lambda(\mathbf{y}) = \frac{1}{2\sigma^2} \sum_{k=1}^{n} \left[2\mathrm{Re}[y_k s^\star_{\lambda,k}] - |s_{\lambda,k}|^2\right] \ . \tag{9.5}$$

- The MAP estimate for λ based on $\mathbf{y}$ is given by:

$$\hat{\lambda}_{\mathrm{MAP}}(\mathbf{y}) = \arg\max_\lambda \log L_\lambda(\mathbf{y}) + \log p_\Lambda(\lambda) \ . \tag{9.6}$$

  The other estimators have similar expressions in terms of $\log L_\lambda$.

Log-likelihood Ratio for Detection/Estimation in Continuous-Time WGN

- Suppose the observation is a random process on interval $[a, b]$ of the form:

$$Y(t) = s_\lambda(t) + w(t) \ , \quad t \in [a, b] \tag{9.7}$$

  where $w(t)$ is a proper complex WGN process with PSD N_0.

- Using Grenander's Theorem [2, Page 272], we showed in class that the log-likelihood ratio for the estimation of λ based on $\{Y(t), t \in [a,b]\}$ is given by:

$$\log L_\lambda(Y) = \frac{1}{2\sigma^2}\left[2\mathrm{Re}\langle Y(t), s_\lambda(t)\rangle - \|s_\lambda(t)\|^2\right] \tag{9.8}$$

where $\sigma^2 = N_0/2$.

- *Sufficient Statistics.* Suppose that for all $\lambda \in \mathcal{S}_\lambda$, $s_\lambda(t)$ has a representation in terms of a finite set of orthonormal basis functions $\{\psi_k(t)\}_{k=1}^n$, i.e.,

$$s_\lambda(t) = \sum_{k=1}^n s_{\lambda,k}\psi_k(t)\ , \quad \text{with } s_{\lambda,k} = \langle s_\lambda(t), \psi_k(t)\rangle\ . \tag{9.9}$$

(This is certainly true if card$(\mathcal{S}_\lambda) \leq \infty$.) Then it is easy to see that

$$\langle Y(t), s_\lambda(t)\rangle = \sum_{k=1}^n s_{\lambda,k}^\star Y_k \tag{9.10}$$

with $Y_k = \langle Y(t), \psi_k(t)\rangle$. Thus

$$\log L_\lambda(Y) = \frac{1}{2\sigma^2}\sum_{k=1}^n \left[2\mathrm{Re}[Y_k s_{\lambda,k}^\star] - |s_{\lambda,k}|^2\right]\ . \tag{9.11}$$

This means that even though $Y(t)$ has components outside the span of $(\psi_1(t), \ldots, \psi_n(t))$, these components are irrelevant for the computation of the computation of the log-likelihood ratio, i.e., they are irrelevant for optimum detection/estimation of λ based on $\{Y(t), t \in [a,b]\}$.

The correlations $\{Y_k\}_{k=1}^n$ are said to form *sufficient statistics* for optimum detection/estimation.

- It is easy to check that

$$Y_{\text{res}}(t) = Y(t) - \sum_{k=1}^n Y_k\psi_k(t) = w(t) - \sum_{k=1}^n w_k\psi_k(t) \tag{9.12}$$

is independent of $\{Y_k\}_{k=1}^n$ and is a function of the noise only. This is the justification that is given in many texts for ignoring $Y_{\text{res}}(t)$ and using only the sufficient statistics, and is sometimes referred to as the *Principle of Irrelevance* [3, Page 220].

References

[1] P. Billingsley. *Probability and Measure*. Wiley, New York, 1986.

[2] H. V. Poor. *An Introduction to Signal Detection and Estimation, 2nd Edition*. Springer-Verlag, New York, 1994.

[3] J. Wozencraft and I. Jacobs. *Principles of Communication Engineering*. John Wiley and Sons, New York, 1965.

10 Lecture 10

Examples of Optimum Detection/Estimation in Continuous-Time WGN

- From last class, for the continuous observation

$$Y(t) = s_\lambda(t) + w(t)\ , \ \ t \in [a, b] \tag{10.1}$$

if $s_\lambda(t) \in \text{span}(\psi_1(t), \ldots, \psi_n(t))$, then $Y_k = \langle Y(t), \psi_k(t) \rangle$ form sufficient statistics. We may pose the problem of optimum detection of λ based on $\{Y(t), t \in [a, b]$ in terms of $\{Y_k\}_{k=1}^n$ without loss of optimality.

Example 1

◦ Suppose

$$Y(t) = \mu_\lambda g(t) + w(t)\ , \ \ t \in [a, b] \tag{10.2}$$

where $\|g(t)\|^2 = 1$.

◦ Clearly $s_\lambda(t) = \mu_\lambda g(t)$ is spanned by the single basis function $g(t)$. Thus $Y = \langle Y(t), g(t) \rangle$ is sufficient, and is given by:

$$Y = \mu_\lambda + w \tag{10.3}$$

where $w = \langle w(t), g(t) \rangle \sim \mathcal{CN}(0, N_0)$.

◦ If $\lambda \in \{0, 1, \ldots, M-1\}$, then, for equal priors,

$$\hat{\lambda}_{\text{MPE}} = \arg\max_m p_m(y) = \arg\min |y - \mu_m|^2\ . \tag{10.4}$$

Example 2

◦ Suppose

$$Y(t) = s_\lambda(t) + w(t)\ , \ \ t \in (-\infty, \infty) \tag{10.5}$$

with

$$s_\lambda(t) = \sum_{n=0}^{N-1} \mu_{m_n} g(t - nT_\text{s}) \tag{10.6}$$

where $m_n \in \{0, 1, \ldots, M-1\}$ and $g(t)$ satisfies the zero-ISI condition:

$$\langle g(t - nT_\text{s}), g(t - \ell T_\text{s}) \rangle = \delta_{n\ell}\ . \tag{10.7}$$

The parameter to be estimated is the sequence $\lambda = (m_0, \ldots, m_{N-1})$.

◦ The basis functions for this example are $\psi_k(t) = g(t - kT_\text{s})$, $k = 0, 1, \ldots, N-1$.

◦ Projecting $Y(t)$ on to these basis functions we get:

$$Y_k = \langle y(t), \psi_k(t) \rangle = \mu_{m_k} + w_k \tag{10.8}$$

where $\{w_k\}$ are easily seen to be i.i.d. $\mathcal{CN}(0, N_0)$ random variables.

- ◦ Note that $\{Y_k\}$ can be formed by passing $Y(t)$ through a LTI sytem with impulse response $h(t) = g(T_s - t)$, and sampling the output every T_s seconds.

- ◦ The likelihood function (based on the sufficient statistics) is given by:

$$p_\lambda(\mathbf{y}) = \prod_{k=0}^{N-1} \frac{1}{\pi N_0} \exp\left[-\frac{|y_k - \mu_{m_k}|^2}{N_0}\right] . \tag{10.9}$$

- ◦ The ML solution for λ is given by:

$$\hat{\lambda}_{\text{ML}} = \arg \min_{m_0,\ldots,m_{N-1}} \sum_{k=0}^{N-1} |y_k - \mu_{m_k}|^2 . \tag{10.10}$$

It is clear from the above that the components of the ML solution satisfy:

$$\hat{\lambda}_{\text{ML},k} = \arg\min_{m_k} |y_k - \mu_{m_k}|^2 , \quad k = 0, 1, \ldots, N-1 . \tag{10.11}$$

Thus symbol-by-symbol detection is optimum for the ML scheme.

- ◦ For the MPE solution, we need priors on all possible sequences $(m_0, m_1, \ldots, m_{N-1})$. If the symbols in the sequence are i.i.d. with each symbol being equally likely to take on the M possible values, then we have uniform priors on the sequences and $\hat{\lambda}_{\text{MPE}} = \hat{\lambda}_{\text{ML}}$, with symbol-by-symbol detection being optimum. If there is dependence between the symbols due to coding, then $\hat{\lambda}_{\text{MPE}}$ is different from $\hat{\lambda}_{\text{ML}}$ and symbol-by-symbol detection is not optimum anymore, i.e., we need to do sequence detection.

- ◦ If the zero-ISI condition on $g(t)$ is not met, again symbol-by-symbol detection is not optimum.

Digital Communication on an Ideal AWGN Channel

- If we assume an ideal channel filter, then passing the signal through the channel leaves it unchanged except for the introduction of a delay τ. In this case

$$y(t) = s_\lambda(t-\tau)e^{j\phi} + w(t) \tag{10.12}$$

where $\phi = -2\pi f_c \tau + \phi_0$.

- The delay τ can usually be accurately estimated at the receiver. But even with a fairly accurate estimate of τ, we may be left with an unknown phase offset at the receiver[3]. Thus after delay estimation, we may move the time axis to the right by τ to get the equivalent model

$$y(t) = s_\lambda(t)e^{j\phi} + w(t) . \tag{10.13}$$

- If ϕ is accurately estimated at the receiver and is used in the demodulation, then we have *coherent* demodulation. In this case, we may multiply $Y(t)$ of (10.13) by $e^{-j\phi}$ to get the equivalent model

$$y(t) = s_\lambda(t) + w(t) \tag{10.14}$$

since multiplying $w(t)$ by $e^{-j\phi}$ does not change its statistical properties.

If ϕ cannot be accurately estimated at the receiver, we have resort to *noncoherent* demodulation.

[3]Errors in τ are magnified by f_c in the phase ϕ.

Coherent demodulation for linear memoryless modulation

∘ The received signal is given by:

$$y(t) = \sum_{n=0}^{N-1} \sqrt{\mathcal{E}_{m_n}} e^{j\theta_{m_n}} g(t - nT_\mathrm{s}) + w(t) \ . \tag{10.15}$$

This is a special case of Example 2 above.

∘ Assuming that the sequences $\{m_n\}$ are equally likely, symbol-by-symbol MPE detection is optimum. Without loss of generality, we can consider demodulation of the symbol corresponding to symbol interval $[0, T_\mathrm{s}]$. The sufficient statistic for optimum detection of this symbol is given by:

$$y = \langle y(t), g(t) \rangle = \sqrt{\mathcal{E}_m} e^{j\theta_m} + w \tag{10.16}$$

where we have dropped the subscript 0 for convenience. This detection problem is a special case of Example 1 above, and we see that

$$\hat{m}_\text{MPE}(y) = \hat{m}_\text{ML}(y) = \arg\min_m |y - \sqrt{\mathcal{E}_m} e^{j\theta_m}|^2 \ . \tag{10.17}$$

∘ Let Γ_m denote the region in the complex plane where a decision in favor of symbol m is made. These decision regions are obtained using the mimumum distance criterion of (10.17).

∘ *Probability of (symbol) error.* The probability of error, conditioned on symbol m being sent is given by:

$$\mathrm{P}_{\mathrm{e},m} = 1 - \mathrm{P}_{\mathrm{c},m} \ , \ \ \text{with } \mathrm{P}_{\mathrm{c},m} = \int_{\Gamma_m} p_m(y) dy \ . \tag{10.18}$$

The average probability of error (assuming equally likely symbols) is given by:

$$\mathrm{P}_\mathrm{e} = \frac{1}{M} \sum_{m=0}^{M-1} \mathrm{P}_{\mathrm{e},m} \ . \tag{10.19}$$

For symmetric constellations, $\mathrm{P}_\mathrm{e} = \mathrm{P}_{\mathrm{e},m}$ for all m.

∘ For MPE detection, P_e can easily be calculated exactly in some special cases such as BPSK and QPSK. For BPSK

$$\mathrm{P}_\mathrm{e} = Q\left(\sqrt{\frac{2\mathcal{E}_\mathrm{s}}{N_0}}\right) \tag{10.20}$$

and for QPSK

$$\mathrm{P}_\mathrm{e} = 2Q\left(\sqrt{\frac{\mathcal{E}_\mathrm{s}}{N_0}}\right) - Q^2\left(\sqrt{\frac{\mathcal{E}_\mathrm{s}}{N_0}}\right) \ . \tag{10.21}$$

∘ *Union Bound on P_e*

$$\mathrm{P}_{\mathrm{e},m} = \mathrm{P}\left(\bigcup_{\ell \neq m} \{\text{decide } \ell\} \middle| \{m \text{ sent}\} \right) \leq \sum_{\ell \neq m} \mathrm{P}\left(\{\text{decide } \ell\} | \{m \text{ sent}\}\right) = Q\left(\sqrt{\frac{d_{m,\ell}^2}{2N_0}}\right) \tag{10.22}$$

where $d_{m,\ell}$ is distance between the points m and ℓ in the constellation.

○ *Intelligent Union Bound (IUB).* The Union Bound is generally too conservative. A better bound is obtained by keeping only the terms in the Union Bounds that are required to cover the error region.

○ *Nearest Neighbor Approximation (NNA).* Let

$$d_{\text{min},m} = \min_{\ell \neq m} d_{m,\ell} \tag{10.23}$$

and let the number of neighbors that are at this minimum distance be $N_{d_{\text{min}}}(m)$. Then

$$\text{P}_{\text{e},m} \approx N_{d_{\text{min}}}(m) Q\left(\sqrt{\frac{d^2_{\text{min},m}}{2N_0}}\right) . \tag{10.24}$$

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

11 Lecture 11

Coherent Detection of Orthogonally Modulated Signals

- The received signal is given by:

$$y(t) = \sum_{n=0}^{N-1} \sqrt{\mathcal{E}} g_{m_n}(t - nT_\text{s}) + w(t) . \tag{11.1}$$

  where $\{g_m(t)\}_{m=0}^{M-1}$ are orthonormal signals.

- Assuming memoryless modulation and equally likely symbol sequences, symbol-by-symbol detection is optimum. For symbol corresponding to $[0, T_\text{s}]$

$$y(t) = \sqrt{\mathcal{E}} g_m(t) + w(t) . \tag{11.2}$$

- The signal $s_m(t) = \sqrt{\mathcal{E}} g_m(t)$ belongs to span$(g_0(t), \ldots, g_{M-1}(t))$, and hence $y_k =< y(t), g_k(t) >$, $k = 0, \ldots, M-1$, form sufficient statistics.

- If symbol m is sent, then the vector of sufficient statistics is given by

$$\mathbf{y} = [w_0 \ \cdots \ \sqrt{\mathcal{E}} + w_m \ \cdots \ w_{M\text{–}1}]^\top \tag{11.3}$$

  where $\{w_k\}$ are easily seen to be i.i.d. $\mathcal{CN}(0, N_0)$ random variables.

- The likelihood function is given by:

$$p_m(\mathbf{y}) = \frac{1}{(\pi N_0)^{M/2}} \exp\left[-\frac{(\mathbf{y}_I - \boldsymbol{\mu}_m)^\top (\mathbf{y}_I - \boldsymbol{\mu}_m)}{N_0} \right] \frac{1}{(\pi N_0)^{M/2}} \exp\left[-\frac{\mathbf{y}_Q^\top \mathbf{y}_Q}{N_0} \right] \tag{11.4}$$

  where $\boldsymbol{\mu}_m = [0 \ldots 0 \ \sqrt{\mathcal{E}} \ 0 \ldots 0]^\top$, with the m-th element being $\sqrt{\mathcal{E}}$.

- For equal priors,

$$\hat{m}_\text{MPE}(y) = \hat{m}_\text{ML}(y) = \arg\max_m p_m(\mathbf{y}) = \arg\min_m (\mathbf{y}_I - \boldsymbol{\mu}_m)^\top (\mathbf{y}_I - \boldsymbol{\mu}_m) = \arg\max_m \ y_{m,I} \tag{11.5}$$

- The probability of error for the MPE decision rule is calculated as follows. First note that by symmetry, $\text{P}_{\text{e},m} = \text{P}_\text{e} = \text{P}_{\text{e},0}$. Also $\text{P}_{\text{e},0} = 1 - \text{P}_{\text{c},0}$, where $\text{P}_{\text{c},0}$ is the probability of correct decision when symbol 0 is sent. Now

$$\text{P}_{\text{c},0} = \text{P}\left(\{y_{0,I} > \max_{k \neq 0} y_{k,I}\} \mid \{\text{‘0’ sent}\} \right) = \text{P}\left\{ \sqrt{\mathcal{E}} + w_{0,I} > \max_{k \neq 0} w_{k,I} \right\} . \tag{11.6}$$

  Let $X = \max_{k \neq 0} w_{k,I}$. Then it is easy to show that

$$\text{P}(X \leq x) = \left[1 - Q\left(\frac{x}{\sqrt{N_0/2}} \right) \right]^{M-1} . \tag{11.7}$$

Now, X is independent of $w_{0,I}$. Thus

$$\begin{aligned}\mathrm{P}_{\mathrm{c},0} &= \int_{-\infty}^{\infty} \mathrm{P}\left(X < \sqrt{\mathcal{E}} + w\right) p_{w_{0,I}}(w) dw = \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{\infty} \left[1 - Q\left(\frac{\sqrt{\mathcal{E}} + w}{\sqrt{N_0/2}}\right)\right]^{M-1} e^{-\frac{w^2}{N_0}}\ dw \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [1 - Q(t)]^{M-1}\ e^{-\frac{1}{2}\left(t - \sqrt{2\gamma_{\mathrm{s}}}\right)^2}\ dt\end{aligned} \tag{11.8}$$

where $\gamma_{\mathrm{s}} = \frac{\mathcal{E}_{\mathrm{s}}}{N_0} = \frac{\mathcal{E}}{N_0}$.

- In the special case of $M = 2$ (e.g. binary FSK), the calculation of P_{e} is much easier. In particular,

$$\mathrm{P}_{\mathrm{e}} = \mathrm{P}_{\mathrm{e},0} = 1 - \mathrm{P}\left[\sqrt{\mathcal{E}} + w_{0,I} > w_{1,I}\right] = Q\left(\sqrt{\frac{\mathcal{E}}{N_0}}\right) = Q\left(\sqrt{\gamma_{\mathrm{s}}}\right)\ . \tag{11.9}$$

since $w_{1,I} - w_{0,I}$ is $\mathcal{N}(0, 2N_0)$.

- Finally, we note that we do not require complete orthogonality among the signals $\{g_m(t)\}$ to get this performance. We can show that it is enough for only $\mathrm{Re}[\rho_{k,m}] = 0$ for $k \neq m$.

Estimation in the Presence of Nuisance Parameters

- Suppose the distribution of the observation Y depends on $\lambda \in \mathcal{S}_\lambda$ as well as $\theta \in \mathcal{S}_\theta(\lambda)$, i.e., the likelihood function given θ and λ is $p_{\lambda,\theta}(y)$. But we are not interested in estimating θ, i.e., it is a *nuisance parameter* for the estimation of λ based on Y. There are two approaches to estimating λ in this situation.

- *Joint ML Approach.* We assume that λ and θ are deterministic but unknown (non-Bayesian model), and do a joint ML estimation of θ and λ, but keep only the estimate of λ. Thus

$$\hat{\lambda}_{\mathrm{ML}}^{(J)}(y) = \arg\max_{\lambda} p_{\lambda}^{\max}(y) \tag{11.10}$$

where

$$p_{\lambda}^{\max}(y) = \max_{\theta \in \mathcal{S}_\theta(\lambda)} p_{\lambda,\theta}(y)\ . \tag{11.11}$$

- *MAP Approach.* Here we assume λ and θ are realizations of random variables Λ and Θ (Bayesian model). Then it is clear that

$$p_{\lambda}^{\mathrm{avg}}(y) = p_{Y|\Lambda}(y|\lambda) = \int_\theta p_{\lambda,\theta}(y) p_{\Theta|\Lambda}(\theta|\lambda) d\theta \tag{11.12}$$

and hence

$$\hat{\lambda}_{\mathrm{MAP}}(y) = \arg\max_{\lambda} p_{\lambda}^{\mathrm{avg}}(y)\ p_\Lambda(\lambda)\ . \tag{11.13}$$

If $\mathrm{card}(\mathcal{S}_\lambda) = M < \infty$, then using steps similar to those used in Lecture 8, one can show that

$$\mathrm{P}_{\mathrm{e}} = \mathrm{P}\{\hat{\Lambda} \neq \Lambda\} = \int_\theta \mathrm{P}(\{\hat{\Lambda} \neq \Lambda\}|\{\Theta = \theta\}) p_\Theta(\theta) d\theta \tag{11.14}$$

is minimized by $\hat{\lambda}_{\mathrm{MAP}}$. Thus $\hat{\lambda}_{\mathrm{MPE}} = \hat{\lambda}_{\mathrm{MAP}}$.

- In the analysis of digital communication systems, we are typically interested in $\hat{\lambda}_{\rm MPE}$ (or $\hat{\lambda}_{\rm MAP}$). In the absence of nuisance parameters, the justification we gave for considering $\hat{\lambda}_{\rm ML}$ was that it equalled $\hat{\lambda}_{\rm MAP}$ for uniform priors. This justification does not extend to case where we have nuisance parameters, since $\hat{\lambda}_{\rm MAP}$ and $\hat{\lambda}_{\rm ML}^{(J)}$ are not necessarily equal even if we have uniform priors on Λ. We may justify the use of $\hat{\lambda}_{\rm ML}^{(J)}$ based on asymptotic properties of ML estimation (see, e.g., [1, Section IV.D]). Also, in some cases of interest such as the one considered in the next section, $\hat{\lambda}_{\rm MAP}$ and $\hat{\lambda}_{\rm ML}^{(J)}$ are indeed equal.

Noncoherent Detection of Orthogonally Modulated Signals

- In this case, the received signal for symbol corresponding to $[0, T_{\rm s}]$ is

$$y(t) = \sqrt{\mathcal{E}} g_m(t) e^{j\phi} + w(t) \tag{11.15}$$

  where ϕ is assumed to be unknown at the receiver.

- The correlations $y_k =< y(t), g_k(t) >$, $k = 0, \ldots, M-1$, still form sufficient statistics, and conditioned on symbol m being sent:

$$\mathbf{y} = [w_0 \ \cdots \ \sqrt{\mathcal{E}} e^{j\phi} + w_m \ \cdots \ w_{M-1}]^\top \tag{11.16}$$

- The likelihood function is given by:

$$p_{m,\phi}(\mathbf{y}) = \frac{1}{(\pi N_0)^M} \exp\left[-\frac{(\mathbf{y} - \boldsymbol{\mu}_m)^\dagger (\mathbf{y} - \boldsymbol{\mu}_m)}{N_0} \right] , \tag{11.17}$$

  where $\boldsymbol{\mu}_m = [0 \ldots 0 \ \sqrt{\mathcal{E}} \mathrm{e}^{j\phi} \ 0 \ldots 0]^\top$, with the m-th element being $\sqrt{\mathcal{E}} \ e^{j\phi}$. This is a detection problem with nuisance parameter ϕ.

- For the joint ML approach,

$$\hat{m}_{\rm ML}^{(J)}(\mathbf{y}) = \arg\max_m p_m^{\rm max}(y) \tag{11.18}$$

  where

$$\begin{aligned} p_m^{\rm max}(y) &= \max_{\phi \in [0,2\pi]} p_{m,\phi}(\mathbf{y}) = \max_{\phi \in [0,2\pi]} \frac{1}{(\pi N_0)^M} \exp\left[-\frac{(\sum_{k=0}^{M-1} |y_k|^2) - 2\mathrm{Re}[y_m \sqrt{\mathcal{E}} e^{-j\phi}] + \mathcal{E}}{N_0} \right] \\ &= \frac{1}{(\pi N_0)^M} \exp\left[-\frac{(\sum_{k=0}^{M-1} |y_k|^2) - 2|y_m|\sqrt{\mathcal{E}} + \mathcal{E}}{N_0} \right] . \end{aligned} \tag{11.19}$$

  Thus

$$\hat{m}_{\rm ML}^{(J)}(\mathbf{y}) = \arg\max_m |y_m| \, . \tag{11.20}$$

- For the MAP approach, assuming that ϕ is uniformly distributed on $[0, 2\pi]$ (and that m is random with uniform priors)

$$\hat{m}_{\rm MAP}(\mathbf{y}) = \hat{m}_{\rm MPE}(\mathbf{y}) = \arg\max_m p_m^{\rm avg}(\mathbf{y}) \tag{11.21}$$

where

$$
\begin{aligned}
p_m^{\text{avg}}(\mathbf{y}) &= \int_0^{2\pi} p_{m,\phi}(\mathbf{y}) \frac{1}{2\pi} d\phi \\
&= \frac{1}{(\pi N_0)^M} \exp\left[-\frac{(\sum_{k=0}^{M-1} |y_k|^2) + \mathcal{E}}{N_0} \right] \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\frac{2\text{Re}[y_m \sqrt{\mathcal{E}} e^{-j\phi}]}{N_0} \right) d\phi \qquad (11.22) \\
&= \frac{1}{(\pi N_0)^M} \exp\left[-\frac{(\sum_{k=0}^{M-1} |y_k|^2) + \mathcal{E}}{N_0} \right] I_0\left(\frac{2|y_m|\sqrt{\mathcal{E}}}{N_0} \right) .
\end{aligned}
$$

Since $I_0(x) \uparrow$ as $x \uparrow$,

$$
\hat{m}_{\text{MAP}}(\mathbf{y}) = \hat{m}_{\text{MPE}}(\mathbf{y}) = \arg\max_m I_0\left(\frac{2|y_m|\sqrt{\mathcal{E}}}{N_0} \right) = \arg\max_m |y_m| . \qquad (11.23)
$$

- Thus $\hat{m}_{\text{ML}}^{(J)} = \hat{m}_{\text{MAP}} = \hat{m}_{\text{MPE}}$ in this case.
- *Probability of error for MPE decision detection.* As before, $\text{P}_\text{e} = \text{P}_{\text{e},0} = 1 - \text{P}_{\text{c},0}$. And

$$
\text{P}_{\text{c},0} = \text{P}\left(\{|y_0| > \max_{k\neq 0} |y_k|\} \mid \{\text{‘0’ sent}\} \right) . \qquad (11.24)
$$

Now, when ‘0’ is sent, $|y_0|$ is Ricean$(\sqrt{\mathcal{E}}, N_0)$ and $|y_k|$, $k \neq 0$, is Rayleigh with second moment N_0, i.e.,

$$
p_{|y_0|}(r) = \frac{r}{\sigma^2} \exp\left[-\frac{r^2 + \mathcal{E}}{2\sigma^2} \right] I_0\left(\frac{r\sqrt{\mathcal{E}}}{\sigma^2} \right) \mathbb{1}_{\{r \geq 0\}} \qquad (11.25)
$$

and for $k \neq 0$,

$$
p_{|y_k|}(r) = \frac{r}{\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2} \right] \mathbb{1}_{\{r \geq 0\}} \qquad (11.26)
$$

where $\sigma^2 = N_0/2$. You will show in Problem 6 of HW#3 that

$$
\text{P}_{\text{c},0} = \int_0^\infty x I_0\left(x\sqrt{\frac{2\mathcal{E}_\text{s}}{N_0}} \right) \exp\left[-\left(\frac{x^2}{2} + \frac{\mathcal{E}_\text{s}}{N_0} \right) \right] \left[1 - \exp\left(-\frac{x^2}{2} \right) \right]^{M-1} dx . \qquad (11.27)
$$

- In the special case of $M = 2$ (e.g., binary FSK) we get the considerably simpler expression:

$$
\text{P}_\text{e} = \text{P}_{\text{e},0} = \frac{1}{2} \exp\left[-\frac{\mathcal{E}_\text{s}}{2N_0} \right] = \frac{1}{2} \exp\left[-\frac{\gamma_\text{s}}{2} \right] . \qquad (11.28)
$$

References

[1] H. V. Poor. *An Introduction to Signal Detection and Estimation, 2nd Edition*. Springer-Verlag, New York, 1994.

12 Lecture 12

Probability of Bit Error for M-ary Modulation

- Assuming that $M = 2^\nu$ for some positive integer ν, we can map the symbols of any M-ary signaling scheme to ν-bit vectors. To compare modulation schemes with different constellation sizes, it is useful to plot the average *bit* error probability for M-ary modulation versus the bit SNR $\gamma_\text{b} = \frac{\mathcal{E}_\text{b}}{N_0} = \frac{\mathcal{E}_\text{s}}{\nu N_0}$.

- For M-ary orthogonal signaling, it is easy to show that irrespective of how bits are assigned to symbols, we have
$$\text{P}_\text{b} = \frac{2^{\nu-1}}{2^\nu - 1}\text{P}_\text{e} \; . \tag{12.1}$$
We can hence get P_b as a function of γ_b, based on expressions for P_e in terms of γ_s.

- For linear modulation, finding an exact expression for P_b as a function of γ_b is difficult except in special cases such as BPSK and QPSK.

- *Nearest Neighbor Approximation (NNA) for* P_b. Let the symbol m be represented by the bit vector $\mathbf{b}_m = [b_{1,m} \; \cdots \; b_{\nu,m}]^\top$, and define:
$$N_{d_\text{min}}(\mathbf{b}_m, i) = \text{\# NN's of } \mathbf{b}_m \text{ that differ from } \mathbf{b}_m \text{ in the } i\text{-th bit position} \; . \tag{12.2}$$
Then
$$\text{P}\left(\{\hat{b}_{m,i} \neq b_{m,i}\} \mid \{\mathbf{b}_m \text{ sent}\}\right) \approx N_{d_\text{min}}(\mathbf{b}_m, i)\, Q\left(\sqrt{\frac{d^2_\text{min}(\mathbf{b}_m)}{2N_0}}\right) \tag{12.3}$$
and
$$\text{P}_{\text{b},i} = \text{P}\left\{i\text{-th bit position in error}\right\} \approx \frac{1}{M}\sum_{m=0}^{M-1} N_{d_\text{min}}(\mathbf{b}_m, i)\, Q\left(\sqrt{\frac{d^2_\text{min}(\mathbf{b}_m)}{2N_0}}\right) \; . \tag{12.4}$$
Finally
$$\text{P}_\text{b} = \frac{1}{\nu}\sum_{i=1}^{\nu} \text{P}_{\text{b},i} \; . \tag{12.5}$$
For Gray coded constellations, the NNA approximation for P_b is at most equal to P_e/ν.

Differential Phase Modulation and Detection

- Consider MPSK signalling on an ideal AWGN channel with phase offset $\phi(t)$ that may change with time, i.e.,
$$y(t) = \sum_{n=0}^{N-1} \sqrt{\mathcal{E}} e^{j\theta_n} g(t - nT_\text{s}) e^{j\phi(t)} + w(t) \; . \tag{12.6}$$
Suppose $\phi(t)$ changes slowly with time so that we can assume that it is constant over two consecutive symbol intervals.

- For standard MPSK
$$\theta_n = \theta_{m_n} = \frac{2\pi m_n}{M}\,, \;\; m_n \in \{0, 1, \ldots, M-1\}\,. \tag{12.7}$$
We know that this scheme performs poorly if we cannot estimate ϕ at the receiver. But if $\phi(t)$ changes slowly, a differential modulation approach can be taken where the sequence $\{\theta_n\}$ is generated from $\{m_n\}$ as
$$\theta_n - \theta_{n-1} = \Delta_{m_n} = \frac{2\pi m_n}{M} \;\; (\text{with } \theta_0 = 0)\,. \tag{12.8}$$

- Sufficient statistics for demodulation are still given by $y_n = \langle y(t), g(t - nT_\mathrm{s})\rangle$, $n = 0, 1, \ldots, N-1$. Note that
$$y_n = \sqrt{\mathcal{E}}\, e^{j\theta_n} e^{j\phi_n} + w_n \tag{12.9}$$
where $\phi_n \approx \phi_{n-1}$ for all n.

- Since the information about m_n is contained in the phase difference between y_n and y_{n-1}, it is convenient to form the statistics:
$$r_n = \frac{y_n y_{n-1}^\star}{\sqrt{\mathcal{E}}} \approx \sqrt{\mathcal{E}} e^{j\Delta_{m_n}} + X_n \;\; (\text{with } r_0 = y_0) \tag{12.10}$$
where
$$X_n = w_n e^{-j\phi_n} e^{-j\theta_{n-1}} + w_{n-1}^\star e^{-j\phi_n} e^{-j\theta_n} + \frac{w_n w_{n-1}^\star}{\sqrt{\mathcal{E}}} \tag{12.11}$$
Since $\{y_n\}$ can be recovered from $\{r_n\}$, there is no loss of optimality if we use $\{r_n\}$ in place of $\{y_n\}$ for detection.

- The statistics $\{r_n\}$ are related to the symbols $\{m_n\}$ in the same way as in standard PSK with coherent detection, except that X_n is not a sequence of i.i.d. $\mathcal{CN}(0, N_0)$ random variables. The fact that $\{X_n\}$ are correlated implies that symbol-by-symbol detection is not optimum, even if $g(t)$ satisfies the zero-ISI condition and the symbols are independent. The MAP (or MPE) detector for this problem is quite complicated and impractical, and hence the following suboptimum detector is used.

- *Differential Detector.* This detector makes a decision on m_n based on r_n alone. In particular, $\hat{m}_n$ is chosen via a minimum distance criterion as
$$\hat{m}_n = \arg\min_m |r_n - \sqrt{\mathcal{E}} e^{j\Delta_m}|^2 \tag{12.12}$$
"pretending" that X_n is a zero-mean PCG random variable.

- *Performance of Differential Detector.* At high SNR, we may ignore the cross-term in the equation for X_n given in (12.11), and conclude that X_n is approximately $\mathcal{CN}(0, 2N_0)$. Then it is clear that the performance of DPSK is worse than PSK with coherent demodulation by approximately 3 dB. An exact analysis of the performance of the differential detector can be done in the special cases of DBPSK and DQPSK (see [1, Page 275]). For DBPSK, we get the surprisingly simple expression:
$$\mathrm{P_b} = \frac{1}{2} \exp(-\gamma_\mathrm{b})\,. \tag{12.13}$$

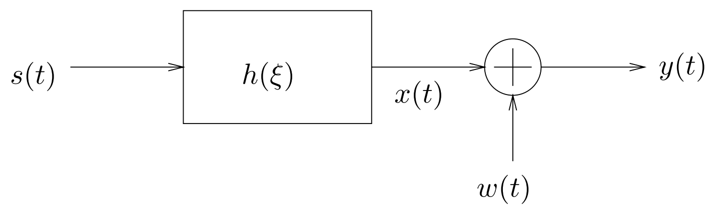


Figure 12.1: Complex baseband point-to-point communications channel

Channel Model for Mobile Communications

In Lecture 2, we developed the complex baseband model for point-to-point communications shown in Figure 12.1. Our goal now is to modify this channel model to incorporate the effects of the mobility. We will focus on terrestrial mobile communications channels – satellite channels are more "well-behaved". The following are points worth noting in making the transition to the mobile communication channel model.

- The additive noise term $w(t)$ is always present whether the channel is point-to-point or mobile, and usually $w(t)$ is modelled as proper complex WGN.

- For point-to-point communications the channel response is generally well modelled by a linear time invariant (LTI) system ($h(\xi)$ may or may not be known at the receiver). For mobile communications, the channel response is time-varying, and we will see that it is well modelled as a linear time-varying (LTV) system.

To study the mobile communications channel, consider the situation where the mobile station (MS) is at location (x, y) or (d, φ) in a coordinate system with the base station (BS) at the origin as shown in Figure 12.2. A 3-d model may be more appropriate in some situations, but for simplicity we will consider a 2-d model. Also, we restrict our attention now to the channel connecting one pair of transmit (Tx) and receive (Rx) antennas.

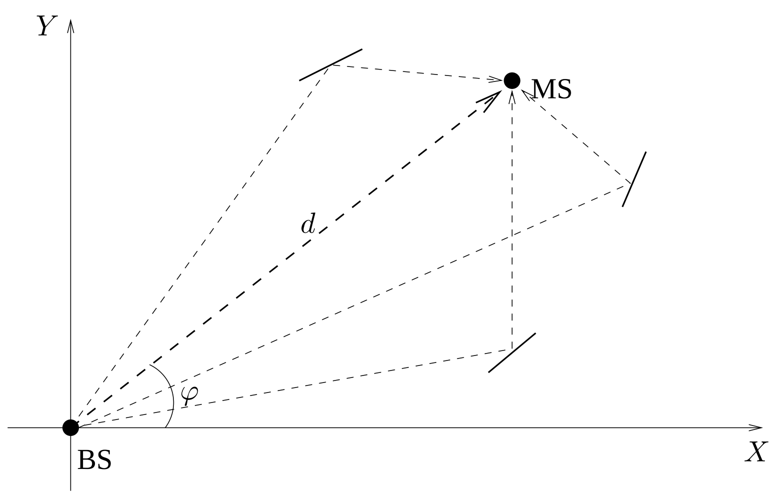


Figure 12.2: Multipath channel seen at location (d, φ) for one Tx&Rx antenna pair

If the mobile is fixed at location (d, φ), the channel that it sees is time-invariant. The response of this time-invariant channel is a function of the location, and is determined by all paths connecting the BS and the MS. Thus we have the system shown in Figure 12.3, where $h_{d,\varphi}(\xi)$ is the impulse response of a causal LTI system, which is a function of the multipath profile between the BS and MS.

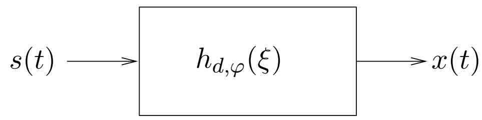


Figure 12.3: Causal LTI system representing multipath profile at location (d, φ)

Referring to Figure 12.2, suppose the n-th path connecting the BS and MS has amplitude gain $\beta_n(d, \varphi)$ and delay $\tau_n(d, \varphi)$. The delay of $\tau_n(d, \varphi)$ introduces a carrier phase shift of

$$\phi_n(d, \varphi) = -2\pi f_c \tau_n(d, \varphi) + \text{constant} \tag{12.14}$$

where the constant depends on the reflectivity of the surface(s) that reflect the path. Then we can write the output $x(t)$ in terms of the input $s(t)$ as

$$x(t) = \sum_n \beta_n(d, \varphi)\, \mathrm{e}^{j\phi_n(d,\varphi)}\, s(t - \tau_n(d, \varphi)) \tag{12.15}$$

which implies that the impulse response is

$$h_{d,\varphi}(\xi) = \sum_n \beta_n(d, \varphi)\, \mathrm{e}^{j\phi_n(d,\varphi)}\, \delta(t - \tau_n(d, \varphi)) \tag{12.16}$$

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

13 Lecture 13

Channel Model for Mobile Communications (continued)

As the MS moves, (d, φ) change with time and the linear system associated with the channel becomes LTV. There are two scales of variation:

- The first is a small-scale variation due to rapid changes in the phase ϕ_n as the mobile moves over distances of the order of a wavelength of the carrier $\lambda_c = c/f_c$, where c is the velocity of light. This is because movements in space of the order of a wavelength cause changes in τ_n of the order of $1/f_c$, which in turn cause changes in ϕ_n of the order of 2π. (Note that for a 900 MHz carrier, $\lambda_c \approx 1/3$ m.)

  Modeling the phases ϕ_n as independent Uniform$[0, 2\pi]$ random variables, we can see that the average power gain in the vicinity of (d, φ) is given by $\sum_n \beta_n^2(d, \varphi)$. We denote this average power gain by $G(d, \varphi)$.

- The second is a large-scale variation due to changes in $\{\beta_n(d, \varphi)\}$ – both in the number of paths and their strengths. These changes happen on the scale of the distance between objects in the environment.

To study these two scales of variation separately, we redraw Figure 12.3 in terms of two components as shown in Figure 13.1.

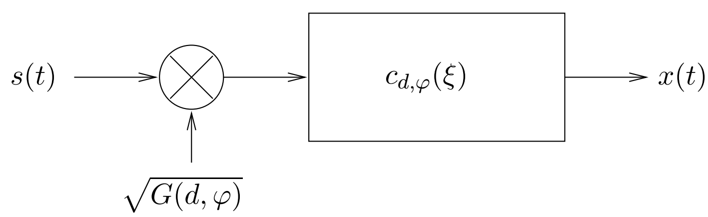


Figure 13.1: Small-scale and large-scale variation components of channel

Here $c_{d,\varphi}$ is normalized so that the average power gain introduced by $c_{d,\varphi}$ is 1, i.e.

$$c_{d,\varphi}(\xi) = \sum_n \beta_n(d, \varphi) e^{j\phi_n(d,\varphi)} \delta(t - \tau_n(d, \varphi)) \tag{13.1}$$

where $\{\beta_n(d, \varphi)\}$ is normalized so that $\sum_n \beta_n^2(d, \varphi) = 1$. The large-scale variations in (average) amplitude gain are then lumped into the multiplicative term $\sqrt{G(d, \varphi)}$.

The goal of wireless channel modeling is to find useful analytical models for the variations in the channel. Models for the large scale variations are useful in cellular capacity-coverage optimization and analysis, and in radio resource management (handoff, admission control, and power control) [1, Chapter]. Models for the small scale variations are more useful in the design of digital modulation and demodulation schemes (that are robust to these variations). We hence focus on the small scale variations in this class.

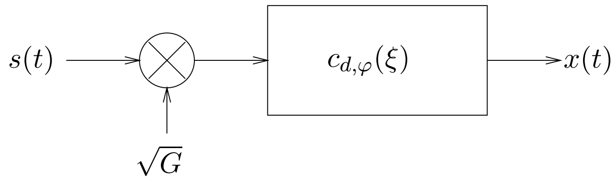


Figure 13.2: Small-scale variations in the channel (with large-scale variations treated as constant).

Small-scale Variations in Gain

Recall from Section that the small scale variations in the channel are captured in a linear system with response

$$c_{d,\varphi}(\xi) = \sum_n \beta_n(d,\varphi) e^{j\phi_n(d,\varphi)} \delta(t - \tau_n(d,\varphi)) \,, \tag{13.2}$$

where the $\{\beta_n(d,\varphi)\}$ are normalized so that $\sum_n \beta_n^2(d,\varphi) = 1$. As (d,φ) changes with t, the channel corresponding to the small-scale variations becomes time-varying and we get:

$$c(t;\xi) := c_{d(t),\varphi(t)}(\xi) = \sum_n \beta_n(t)\, e^{j\phi_n(t)}\, \delta(\xi - \tau_n(t)) \,. \tag{13.3}$$

Treating the large scale variations $\sqrt{G(d,\varphi)}$ as roughly constant (see Figure 13.2), we obtain:

$$x(t) = \sqrt{G} \int_0^\infty c(t;\xi) s(t-\xi) d\xi \,. \tag{13.4}$$

Finally, we may absorb the scaling factor $\sqrt{G}$ into the signal $s(t)$, with the understanding that the power of $s(t)$ is the received signal power after passage through the channel. Then

$$x(t) = \int_0^\infty c(t;\xi) s(t-\xi) d\xi \,. \tag{13.5}$$

Doppler shifts in phase

For movements of the order of a few wavelengths, $\{\beta_n(t)\}$ and $\{\tau_n(t)\}$ are roughly constant, and the time variations in $c(t;\xi)$ are mainly due to changes in $\{\phi_n(t)\}$, i.e.,

$$c(t;\xi) \approx \sum_n \beta_n\, e^{j\phi_n(t)}\, \delta(\xi - \tau_n) \,. \tag{13.6}$$

From this equation it is clear that the magnitude of the impulse response $|c(t;\xi)|$ is roughly independent of t. A typical plot of $|c(t;\xi)|$ is shown in Figure 13.3. The width of the delay profile (delay spread) is of the order of tens of microseconds for outdoor channels, and of the order of hundreds of nanoseconds for indoor channels. Note that the paths typically appear in clusters in the delay profile (why?).

To study the phase variations $\phi_n(t)$ in more detail, consider a mobile that is traveling with velocity v and suppose that the n-th path has an angle of arrival $\theta_n(t)$ with respect to the velocity vector as shown in Figure 13.4. (Note that we may assume that θ_n is roughly constant over the time horizon corresponding to a few

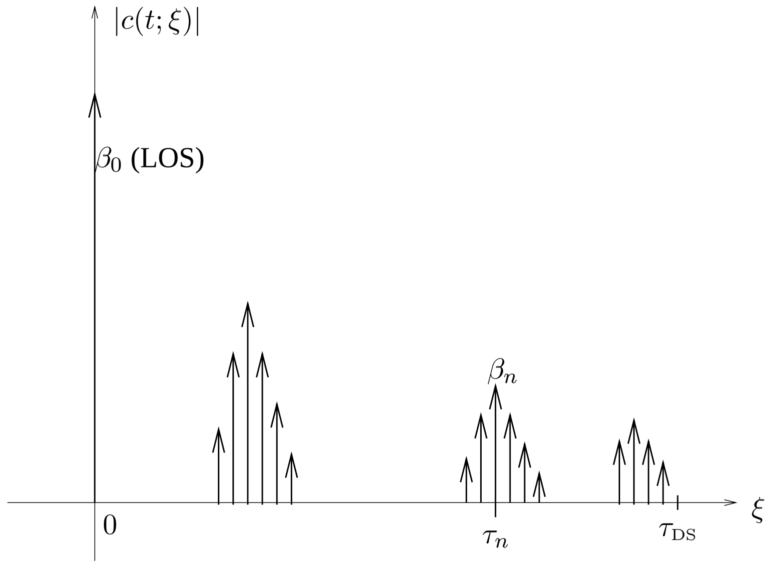


Figure 13.3: Typical delay profile of channel with LOS path having delay 0.

wavelengths.) Then for small Δ_t,

$$\phi_n(t + \Delta_t) - \phi_n(t) \approx \frac{2\pi f_c v \Delta_t \cos\theta_n}{c} = \frac{2\pi v \Delta_t}{\lambda_c} \cos\theta_n \ , \tag{13.7}$$

where λ_c is the carrier wavelength and c is the velocity of light. The frequency shift introduced by the movement of the mobile is hence given by

$$\lim_{\Delta t \to 0} \frac{\phi_n(t + \Delta_t) - \phi_n(t)}{2\pi\Delta_t} = \frac{v}{\lambda_c} \cos\theta_n = f_m \cos\theta_n \ , \tag{13.8}$$

where $f_m = v/\lambda_c$ is called the *maximum Doppler frequency*. We will use this model for the variations in $\phi_n(t)$ to characterize small-scale variations statistically in the following section.

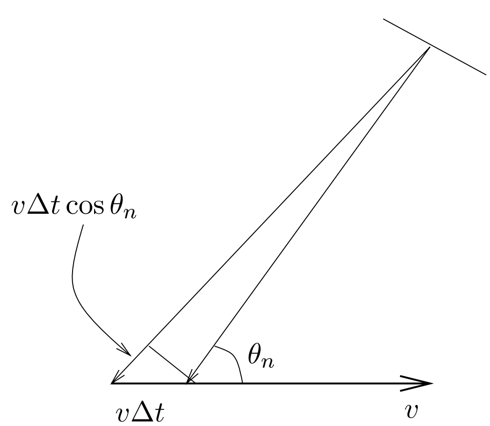


Figure 13.4: Doppler Shifts

Definition 13.1. The quantity $\tau_{\mathrm{DS}} = \max \tau_n(d, \varphi) - \min \tau_n(d, \varphi)$ is called the *delay spread* of the channel.

Without loss of generality, we may assume that the delay corresponding to the first path arriving at the receiver is 0. Then $\min \tau_n(d, \varphi) = 0$, $\tau_{\mathrm{DS}} = \max \tau_n(d, \varphi)$, and (13.5) can be rewritten as:

$$x(t) = \int_0^{\tau_{\mathrm{DS}}} c(t; \xi) s(t - \xi) d\xi \ . \tag{13.9}$$

Frequency Nonselective (Flat) Fading

If the bandwidth of transmitted signal $s(t)$ is much smaller than $1/\tau_{\mathrm{DS}}$, then $s(t)$ does not change much over time intervals of the order of τ_{DS}. Thus (13.9) can be approximated as

$$x(t) \approx s(t) \int_0^{\tau_{\mathrm{DS}}} c(t, \xi) = s(t) \sum_n \beta_n e^{j\phi_n(t)} \ . \tag{13.10}$$

This implies that the multipath channel simply scales the transmitted signal without introducing significant frequency distortion. The variations with time of this scale factor are referred to as frequency nonselective, or *flat*, fading.

Note that the distortions introduced by the channel depend on the relationship between the delay spread of the channel and the bandwidth of the signal. The same channel may be frequency selective or flat, depending on the bandwidth of the input signal. With a delay spread of 10 μs corresponding to a typical outdoor urban environment, an AMPS signal (30 kHz) undergoes flat fading, whereas an IS-95 CDMA signal (1.25 MHz) undergoes frequency selective fading.

For flat fading, the channel model simplifies to

$$x(t) = E(t) s(t) \tag{13.11}$$

where

$$E(t) = \int_0^{\tau_{\mathrm{DS}}} c(t, \xi) d\xi = \sum_n \beta_n e^{j\phi_n(t)} \ . \tag{13.12}$$

References

[1] G. Stuber. *Principles of Mobile Communication*. Kluwer Academic, Norwell, MA, 1996.

14 Lecture 14

Purely Diffuse Scattering - Rayleigh Fading

Our goal is to model $\{E(t)\}$ statistically, but before we do that we distinguish between the cases where the multipath does or does not have a line-of-sight (LOS) component. In the latter case, the multipath is produced only from reflections from objects in the environment. This form of scattering is purely diffuse and can be assumed to form a continuum of paths, with no one path dominating the others in strength. When there is a LOS component, it usually dominates all the diffuse components in signal strength.

To model $\{E(t)\}$ statistically, we first fit a stochastic model to the phases $\{\phi_n(t)\}_{n=1,2,\ldots}$.

Assumption 14.1. *The phases* $\{\phi_n(t)\}_{n=1,2,\ldots}$ *are well modeled as independent stochastic processes, with* $\phi_n(t)$ *being uniformly distributed on* $[0, 2\pi]$ *for each* t *and* n*.*

Using this assumption, we get the following results:

① $\{E(t)\}$ is a zero-mean process. This is because

$$\mathrm{E}\left[E(t)\right] = \sum_n \beta_n \mathrm{E}\left[e^{j\phi_n(t)}\right] = 0$$

② The process $\{e_n(t)\}$ defined by $e_n(t) = \beta_n \mathrm{e}^{j\phi_n(t)}$ is a *proper complex* random process.

Proof. We need to show that the pseudocovariance function of $\{e_n(t)\}$ equals zero.

$$\begin{aligned}\tilde{C}(t+\tau, t) &= \mathrm{E}[e_n(t)e_n(t+\tau)] \\ &= \mathrm{E}[(\beta_n \mathrm{e}^{j\phi_n(t)})(\beta_n \mathrm{e}^{j\phi_n(t+\tau)})] \\ &= \beta_n^2 \mathrm{E}[\mathrm{e}^{j(\phi_n(t)+\phi_n(t+\tau))}] \\ &\approx \beta_n^2 \mathrm{E}[\mathrm{e}^{j(2\phi_n(t)+2\pi f_m \tau \cos\theta_n)}] = 0\end{aligned}$$

where the approximation on the last line follows from (13.7). ■

③ If the number of paths is large, we may apply the Central Limit Theorem to conclude that $\{E(t)\}$ is a proper complex Gaussian (PCG) random process.

First order statistics of $\{E(t)\}$ for purely diffuse scattering

For fixed t, $E(t) = E_I(t) + jE_Q(t)$ is PCG random variable with

$$\mathrm{E}\left[|E(t)|^2\right] = \sum_n \beta_n^2 = 1\ . \tag{14.1}$$

Since $E(t)$ is proper, $E_I(t)$ and $E_Q(t)$ are uncorrelated and have the same variance, which equals half the variance of $E(t)$. Since $E(t)$ is also Gaussian, $E_I(t)$ and $E_Q(t)$ are independent as well. Thus $E_I(t)$ and $E_Q(t)$ are independent $\mathcal{N}(0, 1/2)$ random variables.

Envelope and Phase Processes

The envelope process $\{\alpha(t)\}$ and the phase process $\{\phi(t)\}$ are defined by

$$\alpha(t) = |E(t)| = \sqrt{E_I^2(t) + E_Q^2(t)}\,, \text{ and } \phi(t) = \tan^{-1}\left(\frac{E_Q(t)}{E_I(t)}\right)\,. \tag{14.2}$$

We can write $x(t) = E(t)s(t)$ in terms of $\alpha(t)$ and $\phi(t)$ as:

$$x(t) = \alpha(t)\mathrm{e}^{j\phi(t)}s(t)\,. \tag{14.3}$$

This means that for flat fading, the channel is "seen" as a single path with gain $\alpha(t)$ and phase shift $\phi(t)$. Note that α and ϕ vary much more rapidly than the gain and phase of the individual paths β_n and ϕ_n (why?).

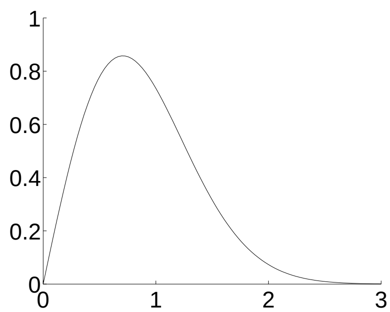


Figure 14.1: Rayleigh pdf

For fixed t, using the fact that $E_I(t)$ and $E_Q(t)$ are independent $\mathcal{N}(0, 1/2)$ random variables, it is easy to show that $\alpha(t)$ and $\phi(t)$ are independent random variables with $\alpha(t)$ having a Rayleigh pdf and $\phi(t)$ being uniform on $[0, 2\pi]$. The pdf of $\alpha(t)$ is given by

$$p_\alpha(x) = 2x\mathrm{e}^{-x^2}u(x)\,. \tag{14.4}$$

It is easy to show that $\mathrm{E}[\alpha] = \sqrt{\pi/4}$ and $\mathrm{E}[\alpha^2] = \mathrm{E}\left[|E(t)|^2\right] = 1$.

Since the envelope has a Rayleigh pdf, purely diffuse fading is referred to as *Rayleigh fading*.

Scattering with a LOS component – Ricean Fading

◇ If there is a LOS (specular) path with parameters θ_0, β_0 and $\phi_0(t)$ in addition to the diffuse components, then

$$E(t) = \beta_0\, e^{j\phi_0(t)} + \sqrt{1-\beta_0^2}\,\check{E}(t) \tag{14.5}$$

where $\{\check{E}(t)\}$ is a zero mean PCG, Rayleigh fading process with variance $\mathrm{E}\left[|\check{E}(t)|^2\right] = 1$.

Note: $\{E(t)\}$ is also a zero-mean process, but it is not Gaussian since the LOS component $\{\beta_0\, e^{j\phi_0(t)}\}$ dominates the diffuse components in power. However, conditioned on $\{\phi_0(t)\}$, $\{E(t)\}$ is a PCG process with mean $\{\beta_0\, e^{j\phi_0(t)}\}$.

◇ *Rice Factor:* The Rice factor κ is defined by

$$\kappa = \frac{\text{power in the specular component}}{\text{total power in diffuse components}} = \frac{\beta_0^2}{1-\beta_0^2}\,. \tag{14.6}$$

From the definition of κ it follows that

$$\beta_0 = \sqrt{\frac{\kappa}{\kappa+1}}, \quad \text{and} \quad 1-\beta_0^2 = \frac{1}{(\kappa+1)}\,. \tag{14.7}$$

⋄ For fixed t, the pdf of the envelope $\alpha(t)$ can be found by first computing the joint pdf of $\alpha(t)$ and $\phi(t)$, conditioned on $\phi_0(t)$. This is straightforward since, conditioned on $\phi_0(t)$, $E(t)$ is a CCG random variable with mean $\beta_0 e^{j\phi_0(t)}$.

⋄ We can then show that the pdf of $\alpha(t)$ conditioned on $\phi_0(t)$ is not a function of $\phi_0(t)$, and we get:

$$p_{\alpha|\phi_0}(x) = \frac{2x}{1-\beta_0^2}\; I_0\left(\frac{2x\beta_0}{1-\beta_0^2}\right)\; \exp\left[-\frac{x^2+\beta_0^2}{1-\beta_0^2}\right]\; u(x) \;\; = \;\; p_\alpha(x) \tag{14.8}$$

This pdf is called a Ricean pdf [1] and it can be rewritten in terms of κ as:

$$p_\alpha(x) = 2x(\kappa+1)\; I_0\left(2x\sqrt{\kappa(\kappa+1)}\right)\; \exp\left[-x^2(\kappa+1)-\kappa\right]\; u(x) \tag{14.9}$$

where $I_0(\cdot)$ is the *zeroth order modified Bessel function of 1st kind* [2], i.e.,

$$I_0(y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(y\cos\phi)d\phi\;. \tag{14.10}$$

⋄ It is easy to see that when $\kappa = 0$, $p_\alpha(x)$ of (14.9) reduces to a Rayleigh pdf.

⋄ The pdf of $\alpha^2(t)$ is easily computed as:

$$p_{\alpha^2}(x) = \frac{p_\alpha(\sqrt{x})}{2\sqrt{x}} = (\kappa+1)\; I_0\left(2\sqrt{x\kappa(\kappa+1)}\right)\; \exp\left[-x(\kappa+1)-\kappa\right]\; u(x)\;. \tag{14.11}$$

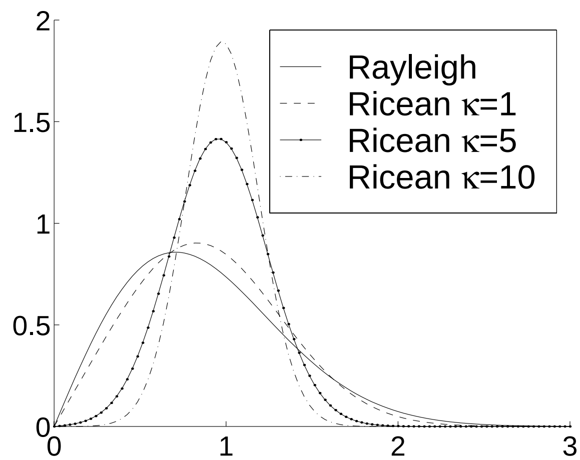


Figure 14.2: Ricean pdf for various Rice factors.

References

[1] S. Rice. Statistical properties of a sine wave plus noise. *Bell Syst. Tech. J.*, 27(1):109–157, January 1948.

[2] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1964.

15 Lecture 15 (2.5 hrs)

Signaling Through Slow Flat Fading Channels

◇ We assume that the long-term variations in the channel are absorbed into $\mathcal{E}_m$. Then $\mathcal{E}_m$ represents the average received symbol energy (for symbol m) over the time frame for which the multipath profile may be assumed to be constant. Then the received signal is given by:

$$y(t) = E(t)s(t) + w(t) = \alpha(t)\mathrm{e}^{j\phi(t)}s(t) + w(t) \tag{15.1}$$

where $\mathrm{E}[\alpha^2(t)] = 1$.

◇ For slow fading, $\alpha(t)$ and $\phi(t)$ may be assumed to be constant over each symbol period. Thus, for memorlyless modulation and symbol-by-symbol demodulation, $y(t)$ for demodulation over symbol period $[0, T_s]$ may be written as

$$y(t) = \alpha e^{j\phi} s_m(t) + w(t) \ \text{ (conditioned on symbol } m \text{ being transmitted)} \tag{15.2}$$

Average probability of error for slow, flat fading

◇ The error probability is a function of the received signal-to-noise ratio (SNR), i.e., the received symbol energy divided by the noise power spectral density. We denote the symbol SNR by γ_{s}, and the corresponding bit SNR by γ_{b}, where $\gamma_{\mathrm{b}} = \gamma/\nu$ and $\nu = \log_2 M$.

◇ For slow, flat fading, the received SNR is

$$\gamma_{\mathrm{s}} = \frac{\alpha^2 \mathcal{E}_{\mathrm{s}}}{N_0} \,. \tag{15.3}$$

The average SNR (averaging over α^2) is given by

$$\overline{\gamma}_{\mathrm{s}} = \mathrm{E}[\alpha^2]\frac{\mathcal{E}_{\mathrm{s}}}{N_0} = \frac{\mathcal{E}_{\mathrm{s}}}{N_0} \,. \tag{15.4}$$

The corresponding bit SNR's are given by

$$\gamma_{\mathrm{b}} = \frac{\gamma}{\nu}, \ \text{ and } \overline{\gamma}_{\mathrm{b}} = \frac{\mathcal{E}_{\mathrm{s}}}{N_0 \nu} = \frac{\mathcal{E}_{\mathrm{b}}}{N_0} \,. \tag{15.5}$$

◇ Suppose the symbol error probability with SNR γ_{s} is denoted by $\mathrm{P_e}(\gamma_{\mathrm{s}})$. Then the average error probability (averaged over the fading) is

$$\overline{\mathrm{P}}_{\mathrm{e}} = \int_0^\infty \mathrm{P_e}(x) p_{\gamma_{\mathrm{s}}}(x) dx \tag{15.6}$$

where $p_{\gamma_{\mathrm{s}}}(x)$ is the pdf of γ_{s}.

◇ For Rayleigh fading, α^2 is exponential with mean 1; hence γ_{s} is exponential with mean $\overline{\gamma}_{\mathrm{s}}$, i.e.,

$$p_{\gamma_{\mathrm{s}}}(x) = \frac{1}{\overline{\gamma}_{\mathrm{s}}} \exp\left[-\frac{x}{\overline{\gamma}_{\mathrm{s}}}\right] u(x) \,. \tag{15.7}$$

⋄ For Ricean fading, α^2 has the pdf given in (14.11), and hence $\gamma_{\rm s}$ has pdf

$$p_{\gamma_{\rm s}}(x) = \frac{\kappa+1}{\overline{\gamma}_{\rm s}}\, I_0\left(2\sqrt{\frac{x\kappa(\kappa+1)}{\overline{\gamma}_{\rm s}}}\right)\ \exp\left[-\frac{x(\kappa+1)}{\overline{\gamma}_{\rm s}} - \kappa\right]\ u(x)\,. \tag{15.8}$$

⋄ *$\overline{P}_e$ for Rayleigh Fading*

∘ Useful result (see problem 3 of HW#4):

$$\int_0^\infty Q(\sqrt{x}\,)\,\frac{{\rm e}^{-x/\gamma}}{\gamma}dx = \frac{1}{2}\left[1-\sqrt{\frac{\gamma}{2+\gamma}}\,\right]\,. \tag{15.9}$$

∘ BPSK

$${\rm P_b}(\gamma_{\rm b}) = Q(\sqrt{2\gamma_{\rm b}}\,)\,. \tag{15.10}$$

Using (15.7) and (15.9), we get

$$\overline{\rm P}_{\rm b} = \int_0^\infty Q(\sqrt{2x}\,)p_{\gamma_{\rm b}}(x)dx = \frac{1}{2}\left[1-\sqrt{\frac{\overline{\gamma}_{\rm b}}{1+\overline{\gamma}_{\rm b}}}\,\right] \approx \frac{1}{4\overline{\gamma}_{\rm b}}\ \text{(for large}\,\overline{\gamma}_{\rm b}). \tag{15.11}$$

∘ Binary coherent orthogonal modulation (e.g. FSK)

$${\rm P_b}(\gamma_{\rm b}) = Q(\sqrt{\gamma_{\rm b}}\,)\,. \tag{15.12}$$

Here $\overline{\rm P}_{\rm b}$ is the same as that for BPSK with $\overline{\gamma}_{\rm b}$ replaced by $\overline{\gamma}_{\rm b}/2$, i.e.,

$$\overline{\rm P}_{\rm b} = \frac{1}{2}\left[1-\sqrt{\frac{\overline{\gamma}_{\rm b}}{2+\overline{\gamma}_{\rm b}}}\,\right] \approx \frac{1}{2\overline{\gamma}_{\rm b}}\ \text{(for large}\ \overline{\gamma}_{\rm b}). \tag{15.13}$$

∘ Binary DPSK

$${\rm P_b}(\gamma_{\rm b}) = \frac{1}{2}{\rm e}^{-\gamma_{\rm b}}\,. \tag{15.14}$$

In this we case we may integrate directly to get

$$\overline{\rm P}_{\rm b} = \int_0^\infty \frac{1}{2}{\rm e}^{-x}\ \frac{{\rm e}^{-x/\overline{\gamma}_{\rm b}}}{\overline{\gamma}_{\rm b}}dx = \frac{1}{2(1+\overline{\gamma}_{\rm b})} \approx \frac{1}{2\overline{\gamma}_{\rm b}}\ \text{(for large}\ \overline{\gamma}_{\rm b}). \tag{15.15}$$

∘ Binary noncoherent orthogonal modulation (FSK)

$${\rm P_b}(\gamma_{\rm b}) = \frac{1}{2}{\rm e}^{-\gamma_{\rm b}/2}\,. \tag{15.16}$$

Here $\overline{\rm P}_{\rm b}$ is the same as that for DPSK with $\overline{\gamma}_{\rm b}$ replaced by $\overline{\gamma}_{\rm b}/2$, i.e.,

$$\overline{\rm P}_{\rm b} = \frac{1}{2+\overline{\gamma}_{\rm b}} \approx \frac{1}{\overline{\gamma}_{\rm b}}\ \text{(for large}\ \overline{\gamma}_{\rm b}). \tag{15.17}$$

∘ Similar expressions may be derived for other M-ary modulation schemes. Note that without fading the error probabilities decrease exponentially with SNR, whereas with fading the error probabilities decrease much more slowly with SNR (inverse linear in case of Rayleigh fading).

◇ $\overline{\mathrm{P}}_e$ *for Ricean Fading*

- ○ *Direct approach:* Compute $\overline{\mathrm{P}}_\mathrm{e}$ using (15.6) and (15.8). This is cumbersome except in some special cases.
- ○ *Nakagami-m approach:* Approximate $p_{\gamma_\mathrm{s}}(x)$ by a Nakagami-m distribution for which integration of P_e to produce $\overline{\mathrm{P}}_\mathrm{e}$ is relatively easy. (See Problem 6 of HW#4).
- ○ *Complex Gaussian approach:* We begin by rewriting γ_s of (15.3) as

$$\gamma_\mathrm{s} = \alpha^2 \overline{\gamma}_\mathrm{s} = (E_I^2 + E_Q^2)\overline{\gamma}_\mathrm{s} = Y_I^2 + Y_Q^2 = |Y|^2 \,, \tag{15.18}$$

where $Y_I = \sqrt{\overline{\gamma}_\mathrm{s}}\, E_I$, $Y_Q = \sqrt{\overline{\gamma}_\mathrm{s}}\, E_Q$, and $Y = Y_I + jY_Q$ is PCG, with mean m_Y and variance σ_Y^2, conditioned on the LOS phase ϕ_0, given by (see (14.7))

$$m_Y = \sqrt{\overline{\gamma}_\mathrm{s}}\, \beta_0 \mathrm{e}^{j\phi_0} = \sqrt{\frac{\overline{\gamma}_\mathrm{s}\kappa}{\kappa+1}}\, \mathrm{e}^{j\phi_0}, \text{ and } \sigma_Y^2 = \mathrm{E}[|Y|^2] = \frac{\overline{\gamma}_\mathrm{s}}{\kappa+1} \,. \tag{15.19}$$

- ○ Without loss of generality, we may assume that $\phi_0 = 0$, since the pdf of γ_s is independent of ϕ_0.
- ○ *General expression for* $\overline{\mathrm{P}}_e$

$$\begin{aligned}\overline{\mathrm{P}}_\mathrm{e} &= \int_0^\infty \mathrm{P}_\mathrm{e}(x) p_{\gamma_\mathrm{s}}(x) dx = \int_{y\in\mathbb{C}} \mathrm{P}_\mathrm{e}\left(|y|^2\right) p_Y(y)\, dy \\ &= \frac{1}{\pi\sigma_Y^2} \int_{y\in\mathbb{C}} \mathrm{P}_\mathrm{e}\left(|y|^2\right)\, \exp\left(-\frac{|y-m_Y|^2}{\sigma_Y^2}\right)\, dy\end{aligned} \tag{15.20}$$

- ○ *Useful result 1*

$$Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta \;\; \text{(problem 3 of HW\#4)} \,. \tag{15.21}$$

This alternative representation was introduced recently by Simon and Divsalar [1] as a way to compute general expressions for the error rates for digital modulation on fading channels. For more recent results, see the book by Simon and Alouini [2].

- ○ *Useful result 2* The following result is also very useful in computing closed-form expressions for the error probability in some special cases.

$$I_n(c) = \frac{1}{\pi}\int_0^{\pi/2} \left(\frac{\sin^2\theta}{\sin^2\theta + c}\right)^n d\theta = [A(c)]^n \sum_{i=0}^{n-1} \binom{n-1+i}{i} [1-A(c)]^i \tag{15.22}$$

with $A(c) = \frac{1}{2}\left[1 - \sqrt{c/(1+c)}\right]$. This result is derived in [3]. Note that $I_n(c)$ also has the following alternative expression whose form is similar to that obtained in Problem 6 of HW#4.

$$I_n(c) = \frac{1}{\pi}\int_0^{\pi/2} \left(\frac{\sin^2\theta}{\sin^2\theta + c}\right)^n d\theta = \frac{1}{2} - \left[\frac{1}{2} - A(c)\right] \sum_{i=0}^{n-1} \binom{2i}{i} [A(c)]^i \, [1-A(c)]^i \,. \tag{15.23}$$

- ○ $\overline{P}_b$ *for Binary Signaling with Ricean Fading*

• BPSK

$$\begin{aligned}\overline{\mathrm{P}}_{\mathrm{b}} &= \frac{1}{\pi\sigma_Y^2}\int_{y\in\mathbb{C}} Q\left(\sqrt{2|y|^2}\right)\exp\left(-\frac{1}{\sigma_Y^2}|y-m_Y|^2\right)\,dy \\ &= \frac{1}{\pi}\int_0^{\pi/2}\left[\frac{1}{\pi\sigma_Y^2}\int_{y\in\mathbb{C}}\exp\left(-\frac{|y|^2}{\sin^2\theta}\right)\,\exp\left(-\frac{|y-m_Y|^2}{\sigma_Y^2}\right)\,dy\right]\,d\theta \qquad (15.24) \\ &= \frac{1}{\pi}\int_0^{\pi/2}\frac{(\kappa+1)\sin^2\theta}{\overline{\gamma}_{\mathrm{b}}+(\kappa+1)\sin^2\theta}\,\exp\left(-\frac{\overline{\gamma}_{\mathrm{b}}\kappa}{(\kappa+1)\sin^2\theta+\overline{\gamma}_{\mathrm{b}}}\right)\,d\theta\,.\end{aligned}$$

where the last line follows after completion of squares inside the exponential to compute the complex Gaussian integral.
Note that for $\kappa = 0$ (i.e. Rayleigh fading), we have

$$\overline{\mathrm{P}}_{\mathrm{b}} = \frac{1}{\pi}\int_0^{\pi/2}\frac{\sin^2\theta}{\overline{\gamma}_{\mathrm{b}}+\sin^2\theta}\,d\theta\,. \qquad (15.25)$$

Using (15.22) with $n = 1$, we can immediately see that the above expression is the same as the one obtained in (15.11). Also, for $\kappa\to\infty$, we see that we get back AWGN performance.

• Binary coherent FSK. Same as BPSK with $\overline{\gamma}_{\mathrm{b}}$ replaced by $\overline{\gamma}_{\mathrm{b}}/2$.

• Binary DPSK.

$$\begin{aligned}\overline{\mathrm{P}}_{\mathrm{b}} &= \frac{1}{\pi\sigma_Y^2}\int_{y\in\mathbb{C}}\frac{1}{2}\exp\left(-|y|^2\right)\,\exp\left(-\frac{1}{\sigma_Y^2}|y-m_Y|^2\right)\,dy \\ &= \frac{\kappa+1}{2(\kappa+1+\overline{\gamma}_{\mathrm{b}})}\,\exp\left[-\frac{\kappa\overline{\gamma}_{\mathrm{b}}}{\kappa+1+\overline{\gamma}_{\mathrm{b}}}\right]\,.\end{aligned} \qquad (15.26)$$

where the second line follows easily by completion of squares as done in class. Again, it is easy to check that we get the Rayleigh result when $\kappa = 0$ and the AWGN result as $\kappa\to\infty$.

• Binary noncoherent FSK. Same as DPSK with $\overline{\gamma}_{\mathrm{b}}$ replaced by $\overline{\gamma}_{\mathrm{b}}/2$.

References

[1] M. K. Simon and D. Divsalar. Some new twists to problems involving the gaussian integral. *IEEE Trans. Commun.*, 46(2):200–210, February 1998.

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

[3] M.-S. Alouini and M.K. Simon. Multichannel reception of digital signals over correlated nakagami fading channels. In *Proc. 36th Annual Allerton Conf.*, Monticello, IL, September 1998.

16 Lecture 16

Diversity Techniques for Flat Fading Channels

- Performance with fading is considerably worse than without fading, especially when the fading is Rayleigh.
- Performance may be improved by sending the same information on many (independently) fading channels
- For signaling on L channels, the received signal on the ℓ-th channel is:
$$y_\ell(t) = \alpha_\ell e^{j\phi_\ell} s_{m,\ell}(t) + w_\ell(t), \quad \ell = 1, 2, \ldots, L, \;\; m = 0, 1, \ldots, M-1 \, . \tag{16.1}$$
where the noise $w_\ell(t)$ is assumed to be independent across channels.
- If $\{\alpha_\ell e^{j\phi_\ell}\}_{\ell=1}^{L}$ are independent, we get maximum diversity against fading.
- How do we guarantee independence of channels? By separating them either in time, frequency or space.
  - ○ frequency separation must be $\gg \frac{1}{\tau_{\rm DS}}$, where $\tau_{\rm DS}$ is the delay spread
  - ○ time separation must be $\gg \frac{1}{f_m}$, where f_m is the maximum Doppler frequency
  - ○ spatial separation must be $\gg \frac{\lambda_c}{2}$, where λ_c is the carrier wavelength.

Memoryless linear modulation with diversity

- When symbol m is sent on the channels
$$y_\ell(t) = \alpha_\ell e^{j\phi_\ell} \sqrt{\mathcal{E}_{s,\ell}} \, a_m \, e^{j\theta_m} \, g_\ell(t) + w_\ell(t), \;\; \ell = 1, 2, \ldots, L, \; m = 0, 1, \ldots, M-1 \, , \tag{16.2}$$
where $g_\ell(t)$ is a (possibly complex) unit energy shaping function on channel ℓ, $\mathcal{E}_{s,\ell}$ is the average symbol energy on channel ℓ, and the a_m's are normalized so that $\sum_m a_m^2 = 1$. We assume that the fading and noise are independent across channels. Note that $\{w_\ell(t)\}$ are independent PCG processes with PSD N_0.
- *Optimum receiver:* If we assume that the phases $\{\phi_\ell\}$ and the amplitudes $\{\alpha_\ell\}$ are estimated perfectly at the receiver, the optimum test statistic is formed by Maximal Ratio Combining (MRC) as
$$y = \sum_{\ell=1}^{L} \alpha_\ell \sqrt{\mathcal{E}_\ell} \, e^{-j\phi_\ell} \int y_\ell(t) g_\ell(t) dt \, . \tag{16.3}$$
We proved that his was optimum in class; also see [1, 2] and Problem 1 of HW#5.
- The sufficient statistic y may be rewritten as
$$y = \sum_{\ell=1}^{L} \alpha_\ell^2 \sqrt{\mathcal{E}_{s,\ell}} \, a_m \, e^{j\theta_m} + \sum_{\ell=1}^{L} \alpha_\ell \sqrt{\mathcal{E}_{s,\ell}} \, w_\ell \, , \tag{16.4}$$
where $\{w_\ell\}$ are independent $\mathcal{CN}(0, N_0)$.

- The MPE (ML) decision rule is the same as without diversity except that the constellation is scaled in amplitude based on the fading on the channels.

Special Case: BPSK with diversity

- The sufficient statistic in this case takes the form

$$y = \pm \sum_{\ell=1}^{L} \alpha_\ell^2 \mathcal{E}_{b,\ell} + w \ , \tag{16.5}$$

where $w = \sum_{\ell=1}^{L} \alpha_\ell \sqrt{\mathcal{E}_{b,\ell}}\, w_\ell$ is PCG with

$$\mathrm{E}[|w|^2] = N_0 \sum_{\ell=1}^{L} \mathcal{E}_{b,\ell}\, \alpha_\ell^2 \ . \tag{16.6}$$

- The MPE decision rule for equal priors (or the ML decision rule)is to decide $+1$ (bit '1') if $y_I > 0$, and -1 (bit '0') if $y_I < 0$.

- For fixed $\{\alpha_\ell\}$,

$$\begin{aligned} \mathrm{P_b} &= \mathrm{P}(\{y_I > 0\} \mid \{\text{bit '0' sent}\}) = \mathrm{P}\left\{ w_I > \sum_{\ell=1}^{L} \alpha_\ell^2 \mathcal{E}_{b,\ell} \right\} \\ &= Q\left(\sqrt{2 \sum_{\ell=1}^{L} \frac{\alpha_\ell^2 \mathcal{E}_{b,\ell}}{N_0}} \right) = Q\left(\sqrt{2 \sum_{\ell=1}^{L} \gamma_{b,\ell}} \right) = Q\left(\sqrt{2\gamma_b}\right) \end{aligned} \tag{16.7}$$

where $\gamma_{b,\ell}$ is the received bit SNR on the ℓ-th channel, and $\gamma_b = \sum_{\ell=1}^{L} \gamma_{b,\ell}$ is the total received bit SNR.

- The average BER is given by

$$\overline{\mathrm{P}}_\mathrm{b} = \int_0^\infty Q\left(\sqrt{2x}\right) p_{\gamma_b}(x) dx \ . \tag{16.8}$$

Thus, we may evaluate $\overline{\mathrm{P}}_\mathrm{b}$ by first finding the pdf $p_{\gamma_b}(x)$. This works well for Rayleigh fading. However, as shown below, $\overline{\mathrm{P}}_\mathrm{b}$ is more easily evaluated in the general case of Ricean fading using the complex Gaussian approach of (15.20), and we get the Rayleigh fading result as a special case.

- *General Ricean analysis using the complex Gaussian approach:*

Assume that α_ℓ is Ricean with Rice factor κ_ℓ. Write $\gamma_{b,\ell} = |Y_\ell|^2$ where $\{Y_\ell\}$ are PCG with means and variances:

$$m_\ell = \sqrt{\overline{\gamma}_{b,\ell}}\, \beta_{0,\ell}\, \mathrm{e}^{j\phi_{0,\ell}} = \sqrt{\frac{\overline{\gamma}_{b,\ell}\, \kappa_\ell}{\kappa_\ell + 1}}\, \mathrm{e}^{j\phi_{0,\ell}}, \ \text{ and } \sigma_\ell^2 = \mathrm{E}[|Y_\ell|^2] = \frac{\overline{\gamma}_{b,\ell}}{\kappa_\ell + 1} \ . \tag{16.9}$$

Then

$$\begin{aligned}
\overline{\mathrm{P}}_{\mathrm{b}} &= \int_0^\infty Q\left(\sqrt{2x}\right) p_{\gamma_b}(x)dx \\
&= \int_{\mathbf{y}} Q\left(\sqrt{2\sum_{k=1}^{L}|y_\ell|^2}\right) \prod_{\ell=1}^{L} \frac{1}{\pi\sigma_\ell^2} \exp\left(-\frac{|y_\ell - m_\ell|^2}{\sigma_\ell^2}\right) dy_1 \dots dy_L \\
&= \frac{1}{\pi}\int_0^{\pi/2} \prod_{\ell=1}^{L} \left[\int_{y_\ell} \exp\left(-\frac{|y_\ell|^2}{\sin^2\theta}\right) \exp\left(-\frac{|y_\ell - m_\ell|^2}{\sigma_\ell^2}\right) dy_\ell\right] d\theta \\
&= \frac{1}{\pi}\int_0^{\pi/2} \prod_{\ell=1}^{L} \frac{(\kappa_\ell+1)\sin^2\theta}{\overline{\gamma}_{b,\ell} + (\kappa_\ell+1)\sin^2\theta} \exp\left(-\frac{\overline{\gamma}_{b,\ell}\,\kappa_\ell}{(\kappa_\ell+1)\sin^2\theta + \overline{\gamma}_{b,\ell}}\right) d\theta\,.
\end{aligned} \tag{16.10}$$

This is best we can do for general Ricean fading. Further simplification is possible for Rayleigh fading.

- *Special case: Rayleigh fading*
If the fading is Rayleigh on all channels, i.e., $\kappa_\ell = 0$, for $\ell = 1, 2, \dots, L$, then

$$\overline{\mathrm{P}}_{\mathrm{b}} = \frac{1}{\pi}\int_0^{\pi/2} \prod_{\ell=1}^{L} \frac{\sin^2\theta}{\overline{\gamma}_{b,\ell} + \sin^2\theta} \; d\theta\,. \tag{16.11}$$

◦ Case 1: $\overline{\gamma}_{b,\ell}$'s are distinct for $\ell = 1, 2, \dots, L$. Here

$$\prod_{\ell=1}^{L} \frac{\sin^2\theta}{\overline{\gamma}_{b,\ell} + \sin^2\theta} = \sum_{\ell=1}^{L} C_\ell \frac{\sin^2\theta}{\overline{\gamma}_{b,\ell} + \sin^2\theta}\,, \tag{16.12}$$

where

$$C_\ell = \prod_{i\neq\ell} \frac{\overline{\gamma}_{b,\ell}}{\overline{\gamma}_{b,\ell} - \overline{\gamma}_{b,i}}\,. \tag{16.13}$$

Thus

$$\overline{\mathrm{P}}_{\mathrm{b}} = \sum_{\ell=1}^{L} C_\ell \frac{1}{\pi}\int_0^{\pi/2} \frac{\sin^2\theta}{\overline{\gamma}_{b,\ell} + \sin^2\theta} d\theta = \sum_{\ell=1}^{L} \frac{C_\ell}{2}\left[1 - \sqrt{\frac{\overline{\gamma}_{b,\ell}}{1+\overline{\gamma}_{b,\ell}}}\right] \tag{16.14}$$

where the second equality follows from (15.22)

◦ Case 2: $\overline{\gamma}_{b,\ell}$'s are identical, i.e. $\overline{\gamma}_{b,\ell} = \overline{\gamma}_{\mathrm{b}}/L$ for all ℓ. Here

$$\overline{\mathrm{P}}_{\mathrm{b}} = \frac{1}{\pi}\int_0^{\pi/2} \left(\frac{\sin^2\theta}{\frac{\overline{\gamma}_{\mathrm{b}}}{L} + \sin^2\theta}\right)^L d\theta = \left[A\left(\frac{\overline{\gamma}_{\mathrm{b}}}{L}\right)\right]^L \sum_{\ell=0}^{L-1} \binom{L-1+\ell}{\ell} \left[1 - A\left(\frac{\overline{\gamma}_{\mathrm{b}}}{L}\right)\right]^\ell \tag{16.15}$$

with

$$A\left(\frac{\overline{\gamma}_{\mathrm{b}}}{L}\right) = \frac{1}{2}\left[1 - \sqrt{\frac{\overline{\gamma}_{\mathrm{b}}}{L+\overline{\gamma}_{\mathrm{b}}}}\right]\,. \tag{16.16}$$

Note that the equation for $\overline{\mathrm{P}}_{\mathrm{b}}$ given in (16.15) is identical to that for BPSK in Nakagami-m fading with $m = L$ (see Problem 6 of HW#4).

For large $\overline{\gamma}_{\mathrm{b}}$,

$$A\left(\frac{\overline{\gamma}_{\mathrm{b}}}{L}\right) \approx \frac{L}{4\overline{\gamma}_{\mathrm{b}}} \quad \text{and} \quad 1 - A\left(\frac{\overline{\gamma}_{\mathrm{b}}}{L}\right) \approx 1\,. \tag{16.17}$$

Thus

$$\overline{\mathrm{P}}_{\mathrm{b}} \approx \left(\frac{L}{4\overline{\gamma}_{\mathrm{b}}}\right)^{L} \sum_{\ell=1}^{L} \binom{L-1+\ell}{\ell} = \left(\frac{L}{4\overline{\gamma}_{\mathrm{b}}}\right)^{L} \binom{2L-1}{L}\,. \tag{16.18}$$

Note that with diversity $\overline{\mathrm{P}}_{\mathrm{b}}$ decreases at $(\overline{\gamma}_{\mathrm{b}})^{-L}$ which is a significant improvement over the inverse linear performance obtained without diversity. (See Figure 16.1.)

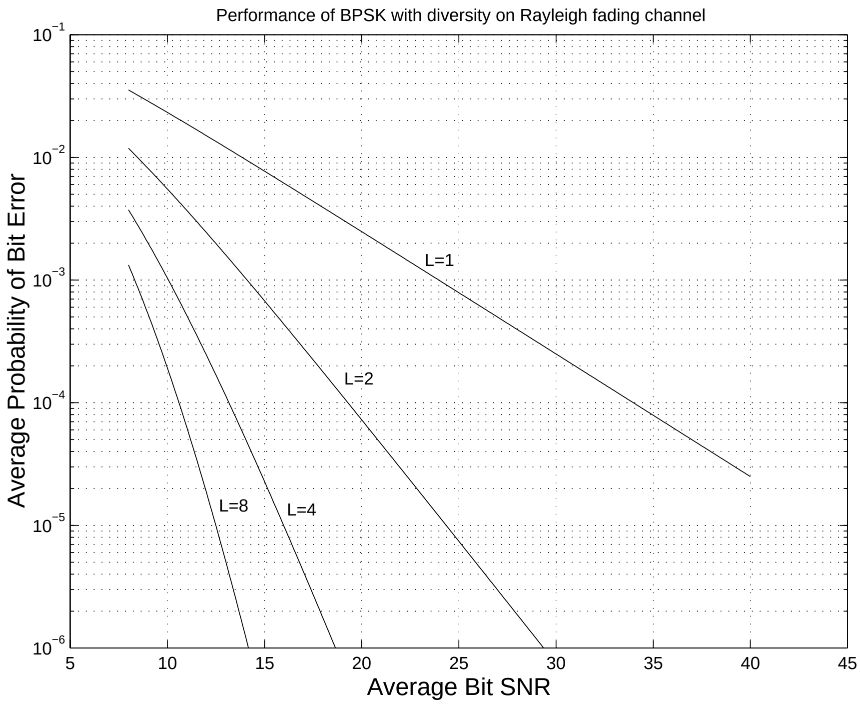


Figure 16.1: BPSK with diversity on Rayleigh fading channels.

References

[1] R. Price. Optimum detection of random signals in noise. *IRE Trans. Inform. Th.*, pages 125–135, December 1956.

[2] T. Kailath. Correlation detection of signals perturbed by a random channel. *IRE Trans. Inform. Th.*, pages 361–368, June 1960.

17 Lecture 17

Error Control Coding for Fading Channels

- The diversity approach to mitigating fading involves sending the same information on multiple independently fading channels. If diversity is obtained by the use of one transmit antenna and multiple receive antennas, we must have the same information on all channels. In other situations it may be possible to send different pieces of information on the various channels, i.e., *code* across the channels.

- A *code* a mapping that takes a sequence of information symbols and produces a (larger) sequence of code symbols so as to be able to detect/correct errors in the transmission of the symbols.

- The simplest class of codes is the class of *binary linear* block codes. Here each vector of k information bits $\mathbf{x}_i = [x_{i,1} \ldots x_{i,k}]$ is mapped to vector of n code bits $\mathbf{c}_i = [c_{i,1} \ldots c_{i,n}]$, with $n > k$. The rate R of the code is defined to be the ratio k/n.

- A binary linear block code can be defined in terms of a $k \times n$ generator matrix $\mathbf{G}$ with binary entries such that the code vector $\mathbf{c}_i$ corresponding to an information vector $\mathbf{x}_i$ is given by:

$$\mathbf{c}_i = \mathbf{x}_i \, \mathbf{G} \tag{17.1}$$

(The multiplication and addition are the standard binary or GF(2) operations.)

- *Example:* $(7, 4)$ Hamming Code

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & | & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & | & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & | & 0 & 1 & 1 \end{bmatrix} \qquad \mathbf{x}_i \, \mathbf{G} = \mathbf{c}_i \tag{17.2}$$

Note that the codewords of this code are in *systematic* form with 4 information bits followed by 3 *parity* bits, i.e.,

$$\mathbf{c}_i = [x_{i,1} \; x_{i,2} \; x_{i,3} \; x_{i,4} \; c_{i,5} \; c_{i,6} \; c_{i,7}] \tag{17.3}$$

with $c_{i,5} = x_{i,1} + x_{i,2} + x_{i,3}$, $c_{i,6} = x_{i,2} + x_{i,3} + x_{i,4}$, and $c_{i,7} = x_{i,1} + x_{i,2} + x_{i,4}$. It is easy to write down the 16 codewords of the (7,4) Hamming code. It is also easy to see that the minimum (Hamming) distance between the codewords, d_{min}, equals 3.

- *General Result.* If $d_{\text{min}} = 2t + 1$, then the code can correct t errors.

- *Example:* Repitition Codes. A rate-$\frac{1}{n}$ repitition code is defined by codebook:

$$0 \mapsto [0\,0 \ldots 0] \text{ , and} 1 \mapsto [1\,1 \ldots 1] \tag{17.4}$$

The minimum distance of this code is n and hence it can correct $\lfloor \frac{n-1}{2} \rfloor$ errors. The optimum decoder for this code is simply a majority logic decoder. A rate-$\frac{1}{2}$ repitition code can detect one error, but cannot correct any errors. A rate-$\frac{1}{3}$ repitition code can correct one error.

Coding Gain

- The *coding gain* of a code is the gain in SNR, at a given error probability, that is achieved by using a code before modulation.

- The coding gain of a code is a function of: (i) the error probability considered, (ii) the modulation scheme used, and (iii) the channel. We now compute the coding gain for BPSK signaling in AWGN for some simple codes. Before we proceed, we introduce the following notation:

$$\begin{aligned}\gamma_{\mathrm{c}} &= \text{SNR per code bit} \\ \gamma_{\mathrm{b}} &= \text{SNR per information bit} = \frac{\gamma_{\mathrm{c}}}{R}\end{aligned}$$

- *Example:* Rate-$\frac{1}{2}$ repitition code, BPSK in AWGN

$$\mathrm{P}\{\text{code bit in error}\} = Q(\sqrt{2\gamma_{\mathrm{c}}}\,) = Q(\sqrt{\gamma_{\mathrm{b}}}\,) \tag{17.5}$$

For an AWGN channel, bit errors are independent across the codeword. It is easy to see that with majority logic decoding

$$\mathrm{P}_{\mathrm{ce}} = \mathrm{P}\{\text{decoding error}\} = [Q(\sqrt{\gamma_{\mathrm{b}}}\,)]^2 + \frac{1}{2}\cdot 2\cdot Q(\sqrt{\gamma_{\mathrm{b}}}\,)\,[1-Q(\sqrt{\gamma_{\mathrm{b}}}\,)] = Q(\sqrt{\gamma_{\mathrm{b}}}\,)\,. \tag{17.6}$$

Thus

$$\mathrm{P}_{\mathrm{b}}(\text{with coding}) = Q(\sqrt{\gamma_{\mathrm{b}}}\,) > Q(\sqrt{2\gamma_{\mathrm{b}}}\,) = \mathrm{P}_{\mathrm{b}}(\text{without coding})\,. \tag{17.7}$$

The rate-$\frac{1}{2}$ repitition code results in a 3 dB coding loss for BPSK in AWGN at all error probabilities!

- *Example:* Rate-$\frac{1}{3}$ repitition code, BPSK in AWGN

$$\mathrm{P}\{\text{code bit in error}\} = Q(\sqrt{2\gamma_{\mathrm{c}}}\,) = Q(\sqrt{2\gamma_{\mathrm{b}}/3}\,) = p \;\; (\text{say})\,. \tag{17.8}$$

Then it is again easy to show that with majority logic decoding

$$\mathrm{P}_{\mathrm{b}}(\text{with coding}) = \mathrm{P}_{\mathrm{ce}} = p^3 + 3p^2(1-p) = 3p^2 - 2p^3 = 3\left[Q\left(\sqrt{\frac{2\gamma_{\mathrm{b}}}{3}}\right)\right]^2 - 2\left[Q\left(\sqrt{\frac{2\gamma_{\mathrm{b}}}{3}}\right)\right]^3. \tag{17.9}$$

Furthermore one can show that the above expression for P_{ce} is always larger than P_{b} without coding (which equals $Q(\sqrt{2\gamma_{\mathrm{b}}}\,)$). Hence this code also has a coding loss (negative coding gain) for BPSK in AWGN.

- *Example:* Rate-$\frac{1}{n}$ repitition code, BPSK in AWGN

$$\mathrm{P}\{\text{code bit in error}\} = Q(\sqrt{2\gamma_{\mathrm{c}}}\,) = Q(\sqrt{2\gamma_{\mathrm{b}}/n}\,) = p \;\; (\text{say})\,. \tag{17.10}$$

Then we can generalize the previous two examples to get:

$$\mathrm{P}_{\mathrm{b}}(\text{with coding}) = \mathrm{P}_{\mathrm{ce}} = \begin{cases} \sum_{q=\frac{n+1}{2}}^{n} \binom{n}{q}\, p^q\, (1-p)^{n-q} & \text{if } n \text{ is odd} \\ \frac{1}{2}\binom{n}{\frac{n}{2}}\, p^{\frac{n}{2}}\, (1-p)^{\frac{n}{2}} + \sum_{q=\frac{n}{2}+1}^{n} \binom{n}{q}\, p^q\, (1-p)^{n-q} & \text{if } n \text{ is even} \end{cases} \tag{17.11}$$

It is possible to show that the Rate-$\frac{1}{n}$ repitition code (with hard decision decoding) results in a coding loss for all n. (One way to show this is to see that even with optimum (soft-decision) decoding, the coding gain for a Rate-$\frac{1}{n}$ repitition code is 0.)

∘ *Example:* $(7,4)$-Hamming code, BPSK in AWGN

$$\text{P}\{\text{code bit in error}\} = Q(\sqrt{2\gamma_\text{c}}\,) = Q(\sqrt{8\gamma_\text{b}/7}\,) = p \ \ (\text{say})\,. \tag{17.12}$$

Since the code can correct one code bit error (and will always have decoding error with more than one code bit error), we have

$$\text{P}_\text{ce} = \text{P}\{2 \text{ or more code bits in error}\} = \sum_{q=2}^{7} \binom{7}{q} \, p^q \ (1-p)^{7-q}\,. \tag{17.13}$$

In general, it is difficult to find an exact relationship between the probability of information bit error P_b and the probability of codeword error P_ce. However, it is easy to see that $\text{P}_\text{b} \leq \text{P}_\text{ce}$ always (see problem 4 of HW#5). Thus the above expression for P_ce serves as an upper bound for the P_b. Based on this bound, we can show (see Fig. 17.1) that for small enough P_b we obtain a positive coding gain from this code. Of course, this coding gain comes with a reduction in information rate (or bandwidth expansion).

∘ *Example:* Rate R, t-error correcting code, BPSK in AWGN

$$\text{P}\{\text{code bit in error}\} = Q(\sqrt{2\gamma_\text{c}}\,) = Q(\sqrt{2R\gamma_\text{b}}\,) = p \ \ (\text{say})\,. \tag{17.14}$$

Now we can only bound P_ce since the code may not be "perfect" like the $(7,4)$-Hamming code. Thus

$$\text{P}_\text{b} \leq \text{P}_\text{ce} \leq \text{P}\{t+1 \text{ or more code bits in error}\} = \sum_{q=t+1}^{n} \binom{n}{q} \, p^q \ (1-p)^{n-q}\,. \tag{17.15}$$

Soft decision decoding (for BPSK in AWGN)

$$\mathbf{x}_i = [x_{i,1} \ldots x_{i,k}] \xrightarrow{\mathcal{C}} [c_{i,1} \ldots c_{i,n}] = \mathbf{c}_i \tag{17.16}$$

∘ The code bits are sent using BPSK. If $c_{i,\ell} = 0$, -1 is sent; if $c_{i,\ell} = 1$, $+1$ is sent $\Rightarrow (2c_{i,\ell} - 1)$ is sent. Thus, the received signal corresponding to the codeword $\mathbf{c}_i$ in AWGN is given by

$$y(t) = \sum_{\ell=1}^{n} (2c_{i,\ell} - 1) \ \sqrt{\mathcal{E}_\text{c}} \ g(t - \ell T_c) + w(t) \tag{17.17}$$

where T_c is the code bit period, and $g(\cdot)$ is a unit energy pulse shaping function that satisfies the zero-ISI condition w.r.t. T_c.

∘ The task of the decoder is to classify the received signal $y(t)$ into one of 2^k classes corresponding to the 2^k possible codewords. This is a 2^k-ary detection problem, for which the sufficient statistics are given by projecting $y(t)$ onto $g(t - \ell T_c)$, $\ell = 1, 2, \ldots, n$. Alternatively, we could filter $y(t)$ with $g(T_c - t)$ and sample the output at rate $1/T_c$. The output of the matched filter for the ℓ-th code bit interval is given by:

$$y_\ell = (2c_{i,\ell} - 1)\sqrt{\mathcal{E}_\text{c}} + w_\ell \tag{17.18}$$

where $\{w_\ell\}$ are i.i.d. $\mathcal{CN}(0, N_0)$.

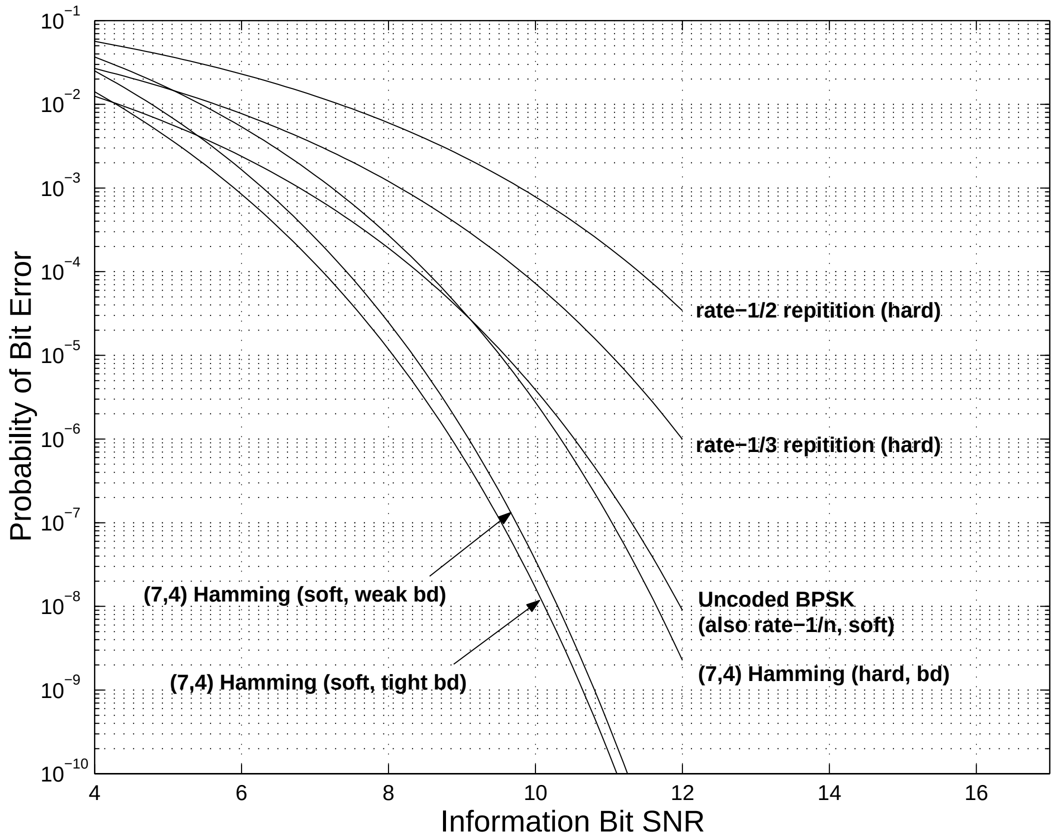


Figure 17.1: Performance of block codes for BPSK in AWGN

- For hard decision decoding, $\operatorname{sgn}(y_{\ell,I})$, $\ell = 1, 2, \ldots, n$, are sent to the decoder. This is suboptimum but is used in practice for block codes since optimum (soft decision) decoding is very complex for large n, and efficient hard-decision decoders can be designed for many good block codes. For convolutional codes, there is no reason to resort to hard decision decoding, since soft decision decoding can be done without significantly increased complexity.

- For optimum (soft decision) decoding, $\{y_\ell\}_{\ell=1}^n$ are sent directly to the decoder for optimum decoding of the codewords.

- *MPE (ML) Decoding:* Let $p_j(\mathbf{y})$ denote the conditional pdf of $\mathbf{y}$, given $\mathbf{c}_j$ that is transmitted. Then, assuming all codewords are equally likely to be transmitted, the MPE estimate of the transmitted codeword index is given by

$$\hat{i}_{\text{MPE}} = \hat{i}_{\text{ML}} = \arg\max_j \; p_j(\mathbf{y}) \, . \tag{17.19}$$

References

[1] J. G. Proakis. *Digital Communications*. Mc-Graw Hill, New York, 3rd edition, 1995.

18 Lecture 18

Soft decision decoding for BPSK in AWGN (continued)

○ The MPE estimate of the codeword index is given by

$$\begin{aligned}
\hat{i}_{\text{MPE}} &= \arg\max_j \; p_j(\mathbf{y}) \\
&= \arg\max_j \prod_{\ell=1}^{n} \frac{1}{\pi N_0} \; \exp\left[-\frac{|y_\ell - (2c_{j,\ell}-1)\sqrt{\mathcal{E}_\text{c}}|^2}{N_0}\right] \\
&= \arg\max_j \sum_{\ell=1}^{n} y_{\ell,I}\, c_{j,\ell} \qquad (18.1)
\end{aligned}$$

○ If we restrict attention to linear block codes, we can assume that the "all-zeros" codeword is part of the code book. Without loss of generality, we can set $\mathbf{c}_1 = [0\,0\cdots 0]$. Furthermore linear block codes have the symmetry property that the probability of codeword error, conditioned on $\mathbf{c}_i$ being transmitted, is the same for all i.

○ Thus

$$\begin{aligned}
\text{P}_{\text{ce}} &= \text{P}\left(\{\hat{i} \neq 1\} | \{i=1\}\right) = \text{P}\left(\bigcup_{j=2}^{2^k} \{\hat{i} \neq j\} \middle| \{i=1\}\right) \\
&= \text{P}\left(\bigcup_{j=2}^{2^k} \left\{\sum_{\ell=1}^{n} c_{j,\ell}\, y_{\ell,I} > 0\right\} \middle| \{i=1\}\right) \\
&\leq \sum_{j=2}^{2^k} \text{P}\left(\left\{\sum_{\ell=1}^{n} c_{j,\ell}\, y_{\ell,I} > 0\right\} \middle| \{i=1\}\right) \qquad (18.2)
\end{aligned}$$

where the last line follows from the Union Bound.

○ Now when $\{i=1\}$, i.e., if $\mathbf{c}_1 = \mathbf{0}$ is sent, then

$$y_{\ell,I} = -\sqrt{\mathcal{E}_\text{c}} + w_{\ell,I}\ . \qquad (18.3)$$

Thus

$$\begin{aligned}
\text{P}\left(\left\{\sum_{\ell=1}^{n} c_{j,\ell}\, y_{\ell,I} > 0\right\} \middle| \{i=1\}\right) &= \text{P}\left\{\sum_{\ell=1}^{n} c_{j,\ell}\,(-\sqrt{\mathcal{E}_\text{c}} + w_{\ell,I}) > 0\right\} \\
&= \text{P}\left\{\sum_{\ell=1}^{n} c_{j,\ell} w_{\ell,I} \;>\; \sqrt{\mathcal{E}_\text{c}} \sum_{\ell=1}^{n} c_{j,\ell}\right\} \\
&= Q\left(\sqrt{\frac{2\mathcal{E}_\text{c}}{N_0} \sum_{\ell=1}^{n} c_{j,\ell}}\right) \qquad (18.4)
\end{aligned}$$

where the last line follows from the fact that $c_{j,\ell}^2 = c_{j,\ell}$.

∘ *Definition:* The Hamming weight ω_i of a codeword $\mathbf{c}_i$ is the number of 1's in the codeword, i.e.,

$$\omega_i = \sum_\ell c_{j,\ell} \,. \tag{18.5}$$

∘ Thus

$$\mathrm{P_b} \le \mathrm{P_{ce}} \le \sum_{j=2}^{2^k} Q(\sqrt{2\ \gamma_\mathrm{c}\ \omega_j}\ ) = \sum_{j=2}^{2^k} Q(\sqrt{2\ R\ \gamma_\mathrm{b}\ \omega_j}\ ) \,. \tag{18.6}$$

∘ To compute the bound on $\mathrm{P_b}$ we need the weight distribution of the code. For example, for the (7,4) Hamming code, it can be shown that there are 7 codewords of weight 3, 7 of weight 4, and 1 of weight 7.

∘ We can obtain a weaker bound on $\mathrm{P_b}$ using only the minimum distance d_min of the code as

$$\mathrm{P_b} \le \mathrm{P_{ce}} \le \sum_{j=2}^{2^k} Q(\sqrt{2\ R\ \gamma_\mathrm{b}\ d_\mathrm{min}}\ ) = (2^k - 1)\ Q(\sqrt{2\ R\ \gamma_\mathrm{b}\ d_\mathrm{min}}\ ) \tag{18.7}$$

∘ *Example:* Rate-$\frac{1}{n}$ repitition code

This code has only one non-zero codeword with weight equal to n. With only one term in the Union Bound, the bound (18.6) becomes an equality. Furthermore, $\mathrm{P_b} = \mathrm{P_{ce}}$ in this special case. Thus

$$\mathrm{P_b} = \mathrm{P_{ce}} = Q\left(\sqrt{2\ \frac{1}{n}\ \gamma_\mathrm{b}\ n}\right) = Q(\sqrt{2\ \gamma_\mathrm{b}}\ ) \,. \tag{18.8}$$

This means that repitition codes with soft decision decoding have zero coding gain for AWGN channels. (We will see in the next section that repitition codes can indeed provide gains for *fading* channels.)

∘ *Example:* $(7,4)$-Hamming code

Using the weight distribution given above, we immediately obtain:

$$\mathrm{P_b} \le 7Q\left(\sqrt{\frac{32}{7}\ \gamma_\mathrm{b}}\right) + 7Q\left(\sqrt{\frac{24}{7}\ \gamma_\mathrm{b}}\right) + Q\left(\sqrt{8\ \gamma_\mathrm{b}}\right) \,. \tag{18.9}$$

We can also obtain the following weaker bound based on (18.7):

$$\mathrm{P_b} \le 15\ Q\left(\sqrt{\frac{24}{7}\ \gamma_\mathrm{b}}\right) \,. \tag{18.10}$$

See Figure 17.1 for the performance curves for soft decision making for the repitition and (7,4) Hamming codes. We can see that soft decision decoding improves performance by about 2 dB over hard decision decoding for the (7,4) Hamming code.

Coding and Interleaving for Slow, Flat Fading Channels

◇ If we send the n bits of the codeword directly on the fading channel, they will fade together if the fading is slow compared to the code bit rate. This results in bursts of errors over the block length of the code. If the burst is longer than the error correcting capability of the code, we have a decoding error.

◇ To avoid bursts of errors, we need to guarantee that the n bits of the codeword fade independently. One way to do this is via *interleaving* and *de-interleaving*.

◇ The interleaver follows the encoder and rearranges the output bits of the encoder so that the code bits of a codeword are separated by N bits, where NT_c is chosen to be much greater than the coherence time T_{coh} of the channel.

◇ The de-interleaver follows the demodulator (and precedes the decoder) and simply inverts the operation of the interleaver so that the code bits of the codeword are back together for decoding. For hard decision decoding the de-interleaver receives a sequence of bit estimates from the demodulator, and for soft decision decoding, the decoder receives a sequence of (quantized) matched filter outputs.

◇ In the following, we assume perfect interleaving and de-interleaving, so that the bits of the codeword fade independently.

Coded BPSK on Rayleigh fading channel – Hard decision decoding

◇ The received signal corresponding to the codeword $\mathbf{c}_i$ received in AWGN with independent fading on the bits is:

$$y(t) = \sum_{\ell}^{n} \alpha_\ell \, e^{\phi_\ell} \sqrt{\mathcal{E}_{\text{c}}} \, (2c_{i,\ell} - 1) \, g_\ell(t) + w(t) \tag{18.11}$$

where $\{g_\ell(t)\}_{\ell=1}^{n}$ are shifted versions of the pulse shaping function $g(t)$ corresponding to the appropriate locations in time after interleaving.

◇ Assuming that $\{\phi_\ell\}$ are known at the receiver, the output of the matched filter for the ℓ-th code bit interval is given by:

$$y_\ell = (2c_{i,\ell} - 1)\alpha_\ell \sqrt{\mathcal{E}_{\text{c}}} + w_\ell \tag{18.12}$$

where $\{w_\ell\}$ are i.i.d. $\mathcal{CN}(0, N_0)$.

◇ For hard decision decoding, we send sgn$(y_{\ell,I})$ to the decoder. Note that we do *not* need to know the fade levels $\{\alpha_\ell\}$ to make hard decisions. However, the error probability corresponding to the hard decisions is a function of the fade levels.

◇ The conditional code bit error probability is given by:

$$\text{P}(\{\ell\text{-th code bit is in error}\} \mid \alpha_\ell) = Q(\sqrt{2\gamma_{\text{c},\ell}}\,) \tag{18.13}$$

where

$$\gamma_{\text{c},\ell} = \frac{\alpha_\ell^2 \mathcal{E}_{\text{c}}}{N_0} = \alpha_\ell^2 \overline{\gamma}_{\text{c}} \,. \tag{18.14}$$

◇ For Rayleigh fading, $\{\gamma_{\text{c},\ell}\}_{\ell=1}^{n}$ are i.i.d. exponential with mean $\overline{\gamma}_{\text{c}} = R\overline{\gamma}_{\text{b}}$.

◇ Using this fact, we can show that for a t-error correcting code the average probability of codeword error $\overline{\text{P}}_{\text{c e}}$ is given by:

$$\overline{\text{P}}_{\text{c e}} \leq \sum_{q=t+1}^{n} \binom{n}{q} \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{c}}}{1+\overline{\gamma}_{\text{c}}}} \right)^{q} \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{c}}}{1+\overline{\gamma}_{\text{c}}}} \right)^{n-q} \tag{18.15}$$

and $\overline{\text{P}}_{\text{b}} \leq \overline{\text{P}}_{\text{c e}}$ in general. (See Problem 4 of HW#5.)

◇ For a rate-$\frac{1}{n}$ repitition code, it is easy to show that

$$\overline{\text{P}}_{\text{b}} = \overline{\text{P}}_{\text{c e}} = \begin{cases} \sum_{q=\frac{n+1}{2}}^{n} \binom{n}{q} \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{q} \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{n-q} & \text{if } n \text{ is odd} \\ \frac{1}{2}\binom{n}{\frac{n}{2}} \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{\frac{n}{2}} \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{\frac{n}{2}} & \\ \qquad + \sum_{q=\frac{n}{2}+1}^{n} \binom{n}{q} \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{q} \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{n+\overline{\gamma}_{\text{b}}}} \right)^{n-q} & \text{if } n \text{ is even} \end{cases} \tag{18.16}$$

In the special case of a rate-$\frac{1}{2}$ repitition code we obtain:

$$\overline{\text{P}}_{\text{b}} = \overline{\text{P}}_{\text{c e}} = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\overline{\gamma}_{\text{b}}}{2+\overline{\gamma}_{\text{b}}}} \tag{18.17}$$

which is 3 dB worse than the error probability without coding.

19 Lecture 19 (2.5 hrs)

Coded BPSK on Rayleigh fading channel – Soft decision decoding

$\diamond$ Assuming that $\{\phi_\ell\}$ are known at the receiver, the output of the matched filter for the ℓ-th code bit interval is given by:

$$y_\ell = (2c_{i,\ell} - 1)\alpha_\ell \sqrt{\mathcal{E}_\text{c}} + w_\ell \tag{19.1}$$

where $\{w_\ell\}$ are i.i.d. $\mathcal{CN}(0, N_0)$.

$\diamond$ MPE decoding (for fixed and known $\{\alpha_\ell\}$): We follow the same steps as in the pure AWGN case to get

$$\begin{aligned} \hat{i}_\text{MPE} &= \arg\max_j \prod_{\ell=1}^{n} \frac{1}{\pi N_0} \exp\left[-\frac{|y_{\ell,I} - (2c_{j,\ell} - 1)\alpha_\ell \sqrt{\mathcal{E}_\text{c}}\,|^2}{N_0}\right] \\ &= \arg\max_j \sum_{\ell=1}^{n} \alpha_\ell \, r_{\ell I} \, c_{j,\ell} \end{aligned} \tag{19.2}$$

Note that, unlike in hard decision decoding, we need to know the fade levels $\{\alpha_\ell\}$ for soft decision decoding.

$\diamond$ Again, as in pure AWGN case, we can set $\mathbf{c}_1 = \mathbf{0}$ and compute a bound on P_ce for fixed $\{\alpha_\ell\}$ as:

$$\text{P}_\text{ce}(\alpha_1, \ldots, \alpha_n) \le \sum_{j=2}^{2^k} \text{P}\left(\left\{\sum_{\ell=1}^{n} \alpha_\ell \, c_{j,\ell} \, y_{\ell I} > 0\right\} \Big| \{i = 1\}\right) = Q\left(\sqrt{2\sum_{\ell=1}^{n} c_{j,\ell} \, \gamma_{\text{c},\ell}}\right) . \tag{19.3}$$

$\diamond$ Let

$$\beta_j = \sum_{\ell=1}^{n} c_{j,\ell} \, \gamma_{\text{c},\ell} \, . \tag{19.4}$$

Then β_j is the sum of ω_j i.i.d. exponential random variables, where ω_j is the weight of $\mathbf{c}_j$. Thus

$$p_{\beta_j}(x) = \left(\frac{1}{\overline{\gamma}_\text{c}}\right)^{w_j} \frac{x^{\omega_j - 1}}{(\omega_j - 1)!} \exp\left(-\frac{x}{\overline{\gamma}_\text{c}}\right) \mathbb{1}_{\{x \ge 0\}} \tag{19.5}$$

$\diamond$ Thus the average codeword error probability (averaged over the distribution of the $\{\alpha_\ell\}$) is given by:

$$\overline{\text{P}}_\text{c e} \le \sum_{j=2}^{2^k} \int_0^\infty Q(\sqrt{2x}\,) p_{\beta_j}(x) \, dx \tag{19.6}$$

and clearly $\overline{\text{P}}_\text{b} \le \overline{\text{P}}_\text{c e}$.

It is easy to show that (see Problem 5 of HW#5):

$$\overline{\text{P}}_\text{b} \le \overline{\text{P}}_\text{c e} \le \sum_{j=2}^{2^k} \left(\frac{1 - \tilde{\sigma}}{2}\right)^{\omega_j} \sum_{q=0}^{\omega_j - 1} \binom{\omega_j - 1 + q}{q} \left(\frac{1 + \tilde{\sigma}}{2}\right)^q \tag{19.7}$$

where

$$\tilde{\sigma} = \sqrt{\frac{\overline{\gamma}_{\text{c}}}{\overline{\gamma}_{\text{c}} + 1}} = \sqrt{\frac{R\,\overline{\gamma}_{\text{b}}}{R\,\overline{\gamma}_{\text{b}} + 1}} \,. \tag{19.8}$$

◇ Also, since $\int_0^\infty Q(\sqrt{2x}\,)p_{\beta_j}(x)dx$ decreases as ω_j increases, we have the following weaker bound on $\overline{\text{P}}_{\text{c e}}$ in terms of d_{min}.

$$\overline{\text{P}}_{\text{b}} \le \overline{\text{P}}_{\text{c e}} \le (2^k - 1)\left(\frac{1-\tilde{\sigma}}{2}\right)^{d_{\text{min}}} \sum_{q=0}^{d_{\text{min}}-1} \binom{d_{\text{min}} - 1 + q}{q} \left(\frac{1+\tilde{\sigma}}{2}\right)^q . \tag{19.9}$$

◇ For large SNR, i.e., $\overline{\gamma}_{\text{c}} \gg 1$, we have

$$\frac{1+\tilde{\sigma}}{2} \approx 1\,, \quad \text{and } \frac{1-\tilde{\sigma}}{2} \approx \frac{1}{4\,\overline{\gamma}_{\text{c}}} = \frac{1}{4\,R\,\overline{\gamma}_{\text{b}}} \tag{19.10}$$

and hence the bound in (19.9) can be approximated by

$$\text{Bound} \approx (2^k - 1)\left(\frac{1}{4\,R\,\overline{\gamma}_{\text{b}}}\right)^{d_{\text{min}}} \binom{2d_{\text{min}} - 1}{d_{\text{min}}} . \tag{19.11}$$

This means that $\overline{\text{P}}_{\text{b}}$ decreases as $(\overline{\gamma}_{\text{b}})^{-d_{\text{min}}}$ for large SNR.

◇ *Example:* Rate-$\frac{1}{n}$ repitition code

$$\overline{\text{P}}_{\text{b}} = \overline{\text{P}}_{\text{c e}} = \left(\frac{1-\tilde{\sigma}}{2}\right)^{n} \sum_{q=0}^{n-1} \binom{n - 1 + q}{q} \left(\frac{1+\tilde{\sigma}}{2}\right)^q \tag{19.12}$$

with $\tilde{\sigma} = \sqrt{\overline{\gamma}_{\text{b}}/(\overline{\gamma}_{\text{b}} + n)}$

The average bit error probability is the same as that obtained with n-th order diversity and maximum ratio combining (as expected).

◇ *Example:* $(7,4)$-Hamming code

$$\overline{\text{P}}_{\text{b}} \le \overline{\text{P}}_{\text{c e}} \le 7f(3) + 7f(4) + f(7) \tag{19.13}$$

where

$$f(\omega) = \left(\frac{1-\tilde{\sigma}}{2}\right)^{\omega} \sum_{q=0}^{\omega-1} \binom{\omega - 1 + q}{q} \left(\frac{1+\tilde{\sigma}}{2}\right)^q \tag{19.14}$$

and $\tilde{\sigma} = \sqrt{4\overline{\gamma}_{\text{b}}/(4\overline{\gamma}_{\text{b}} + 7)}$

◇ Performance plots for these codes for both hard and soft decision making are shown in Figure 19.1.

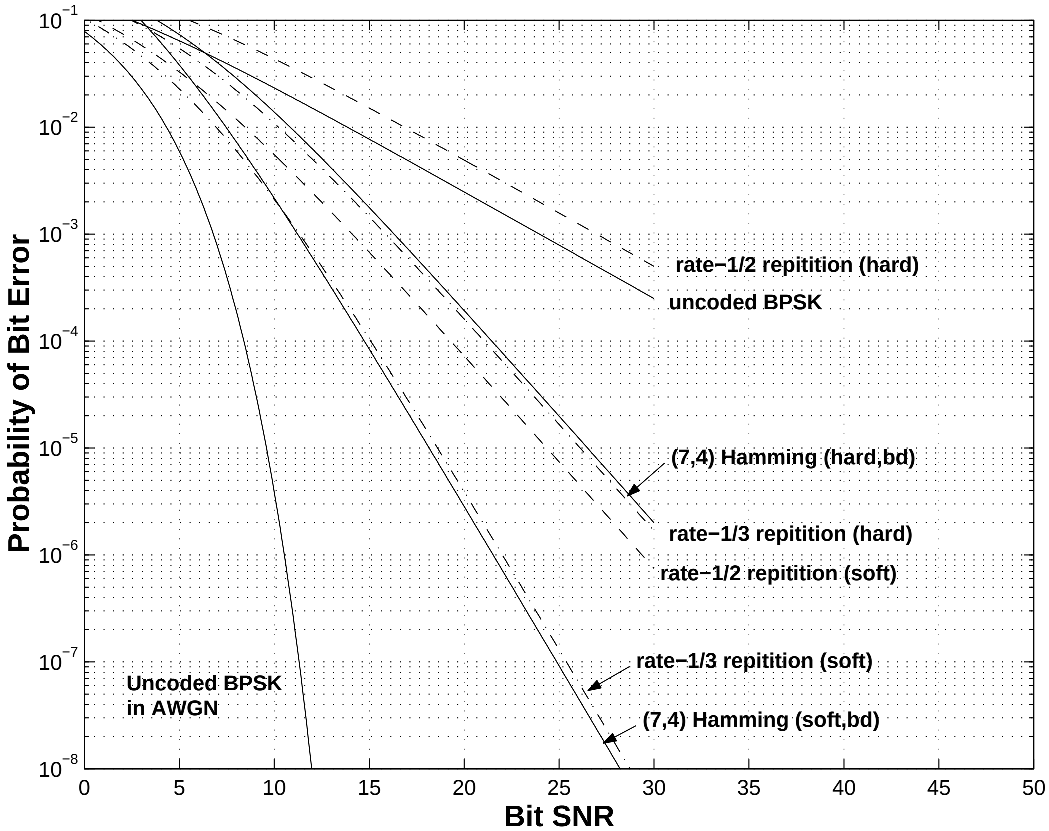


Figure 19.1: Performance of block codes for BPSK in Rayleigh fading with perfect interleaving

Code Division Multiple Access (CDMA)

Spread Spectrum Modulation

- *Informal definition of spread spectrum signal* (Viterbi [1]): A spread spectrum information bearing signal is one whose bandwidth is much larger than what is needed to transmit data reliably.

- *Precise definition* (Massey [2]): A spread spectrum signal is one for which the (essential) Fourier bandwidth is much larger than the Shannon bandwidth, where Shannon bandwidth refers to half the number of dimensions in signal space occupied by the signal per second.

- *Spreading versus Coding* (VVV [3]): Spreading is a linear mapping in signal space that is energy and distance preserving. Spreading provides no coding gain against AWGN; it is hence akin to repitition coding. Coding is necessarily a nonlinear mapping in signal space. Every bandwidth expansion scheme can be written as coding followed by spreading.

- Why spread spectrum?

  ◇ Military applications
    - ○ immunity to narrowband jammers
    - ○ low probability of intercept (LPI)

  ◇ Commercial applications
    - ○ multiaccess capability
    - ○ randomization of interference
    - ○ diversity gain against fading

Direct sequence spread spectrum (DS/SS)

Consider the complex baseband signal

$$s(t) = \sum_n s_{m_n}(t - nT_s) \tag{19.15}$$

where $m_k \in \{0, 1, \ldots, M-1\}$. The signal $s(t)$ occupies a bandwidth W that depends on the modulation scheme used. To spread spectrum, we simply multiply $s(t)$ by a high frequency chip waveform $c(t)$ that has bandwidth NW, where N is said to be the *processing gain*.

DS/SS Linear Modulation

- Without spreading

$$s(t) = \sqrt{\mathcal{E}} \sum_n z^{(n)} g_{T_s}(t - nT_s) \tag{19.16}$$

  where $z^{(n)} \in \{\sqrt{\mathcal{E}_0}\, e^{j\theta_0}, \ldots, \sqrt{\mathcal{E}_{M-1}}\, e^{j\theta_{M-1}}\}$ is the complex symbol that is transmitted during the n-th symbol interval, and $g_{T_s}(\cdot)$ is a unit energy pulse shaping function that satisfies the zero ISI (Nyquist) condition

$$\int g_{T_s}(t - iT_s)\, g_{T_s}(t - jT_s)\, dt \approx \delta[i-j] \tag{19.17}$$

- Examples of $g_{T_s}(\cdot)$ (two extreme cases)
  - ◦ *Sinc pulse:*

$$g_{T_s}(t) = \frac{1}{\sqrt{T_s}} \, \text{sinc}\left(\frac{t}{T_s} - 0.5\right) . \tag{19.18}$$

This is the pulse with smallest bandwidth satisfying the Nyquist condition of (19.17). The Fourier transform of this pulse is given by

$$G_{T_s}(f) = \sqrt{T_s}\, \text{rect}(fT_s)\, e^{-j\pi T_s f} \tag{19.19}$$

where

$$\text{rect}(x) = \begin{cases} 1 & \text{if } |x| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{19.20}$$

  - ◦ *Rectangular pulse:*

$$g_{T_s}(t) = \frac{1}{\sqrt{T_s}} \, \text{rect}\left(\frac{t}{T_s} - 0.5\right) . \tag{19.21}$$

and $G_{T_s}(f) = \sqrt{T_s}\, \text{sinc}(fT_s)\, e^{-j\pi T_s f}$.

This pulse is convenient for analysis since the waveforms do not overlap from symbol to symbol, and the pulse autocorrelation function has a convenient triangular form. However, it has poor bandwidth properties and has been shown to result in poor performance in CDMA systems [4]

  - ◦ In practice, since sinc pulses cannot be used due to their infinite time extent, pulses that are approximately bandlimited and are time limited to a few symbol periods are used.

References

[1] A. J. Viterbi. Spread spectrum communications-myths and realities. *IEEE Commun. Mag.*, 17(3):11–18, May 1979.

[2] J. L. Massey. Information theory aspects of spread-spectrum communications. In *Proc. IEEE ISSSTA'94*, pages 16–20, Oulu, Finland, July 1994.

[3] V. V. Veeravalli. The coding-spreading tradeoff in CDMA systems. In *Proc. 27th Annual Allerton Conference*, pages 831–40, Monticello, IL, September 1999.

[4] A. Mantravadi and V. V. Veeravalli. On chip-matched filtering and discrete sufficient statistics for asynchronous band-limited CDMA systems. *IEEE Trans. Commun.*, 2000. Submitted. See http://www.comm.csl.uiuc.edu/vvv/cv/pubs/ for a copy.

20 Lecture 20

DS/SS Linear Modulation

- *Spreading Spectrum*

  The transmitted signal for DS/SS linear modulation is given by:

$$s(t) = \sqrt{\mathcal{E}_{\rm s}} \sum_n z^{(n)} c^{(n)}(t - nT_s) \tag{20.1}$$

  where $c^{(n)}(\cdot)$ is a unit energy waveform that replaces $g_{T_s}(\cdot)$ in (19.16), and is given by:

$$c^{(n)}(t) = \sum_{j=0}^{N-1} c_j^{(n)} g_{T_c}(t - jT_c)\ . \tag{20.2}$$

  The sequence $\{c_j^{(n)}\}_{j=0}^{N-1}$ is the *chip sequence* for the n-th symbol interval, and can be written compactly using the vector notation

$$\mathbf{c}^{(n)} = [c_0^{(n)}\ c_1^{(n)}\ \cdots\ c_{N-1}^{(n)}]^\top \tag{20.3}$$

  with $\mathbf{c}^{(n)}$ normalized such that $\mathbf{c}^{(n)\dagger}\ \mathbf{c}^{(n)} = 1$. There are two special cases that we can consider:

  - *Short sequences:* $\mathbf{c}^{(n)} = \mathbf{c}$ for all n.
  - *Long sequences:* $\mathbf{c}^{(n)}$ is different for each n, and the sequence may repeat after a long period that spans several symbols. Such sequences are generated using pseudorandom number generators and are also called *random* sequences.

  The chip sequences are typically binary valued, i.e., $c_j^{(n)} \in \{+\frac{1}{\sqrt{N}}, -\frac{1}{\sqrt{N}}\}$, but in general they can be complex valued and satisfy $\mathbf{c}^{(n)\dagger}\ \mathbf{c}^{(n)} = 1$

  The chip pulse $g_{T_c}(\cdot)$ is a unit energy function that satisfies the zero ICI (Nyquist) condition

$$\langle g_{T_c}(t - iT_c), g_{T_c}(t - jT_c)\rangle \approx \delta[i - j] \tag{20.4}$$

  and just as with $g_{T_s}(\cdot)$, there is a range of choices for $g_{T_c}(\cdot)$, with the sinc pulse and the rectangular pulse being extreme cases. Since $g_{T_c}(\cdot)$ has unit energy and $\mathbf{c}^{(n)\dagger}\ \mathbf{c}^{(n)} = 1$, it follows that $c^{(n)}(\cdot)$ is a unit energy waveform. It is also clear from (20.4) that $c^{(n)}(\cdot)$ satisfies the zero ISI condition given in (19.17).

Single User Communications with DS/SS Linear Modulation

- Consider single user communications over an AWGN channel with DS/SS linear modulation. The received signal is given by:

$$y(t) = e^{j\phi} \sqrt{\mathcal{E}_{\rm s}} \sum_n z^{(n)} c^{(n)}(t - nT_s) + w(t) \tag{20.5}$$

  where ϕ is the phase offset introduced by the channel.

- Assuming zero ISI, symbol-by-symbol detection is optimum, and the sufficient statistic for detecting the symbol corresponding to interval $[0, T_{\rm s}]$ (say) is given by

$$y = \langle y(t), e^{j\phi} c(t) \rangle = \sqrt{\mathcal{E}_{\rm s}}\, z + w \tag{20.6}$$

where we have dropped the superscript "(0)" for convenience, and where $w \sim \mathcal{CN}(0, N_0)$. Note that we need to know the sequence $\mathbf{c} = [c_0\ c_1\ \cdots\ c_{N-1}]^\top$ in addition to ϕ to compute the above correlation.

- For soft decision decoding, we send y to the decoder. The performance metric for soft decisions is the signal-to-noise ratio in the statistic y, which is given by:

$$\text{SNR} = \frac{\text{E}[|\text{E}[y|z]|^2]}{\text{var}(y|z)} = \frac{\mathcal{E}_{\rm s}}{N_0}\ . \tag{20.7}$$

- For hard decision decoding, the MPE decision rule is given by:

$$\hat{z}_{\text{MPE}} = \arg\max_{z \in \mathcal{S}} p(y|z) = \arg\min_{z \in \mathcal{S}} |y - \sqrt{\mathcal{E}_{\rm s}}\, z|^2 \tag{20.8}$$

where $\mathcal{S} = \{a_0 e^{j\theta_0}, \ldots, a_{M-1} e^{j\theta_{M-1}}\}$.

For binary signaling

$$\hat{z}_{\text{MPE}} = \text{sgn}(y_I)\ , \tag{20.9}$$

and the bit-error rate (BER) is given by

$$\text{P}_{\rm b} = Q\left(\sqrt{\frac{2\mathcal{E}_{\rm b}}{N_0}}\right)\ . \tag{20.10}$$

Note that the performance is the same as without spreading – spreading results in zero coding gain in AWGN, i.e., it is akin to repitition coding with soft decision decoding.

Mulitiuser Communications

- Users are indexed by $k = 1, 2, \ldots, K$, with K being the total number of users.

- The signal of user k (for linear modulation) is given by

$$s_k(t) = \sqrt{\mathcal{E}_{{\rm s},k}} \sum_n z_k^{(n)}\, c_k^{(n)}(t - nT_s)\ . \tag{20.11}$$

where $\mathcal{E}_{{\rm s},k}$ is the average symbol energy of user k, $z_k^{(n)}$ is the n-th symbol of user k, T_s is the symbol period, and $c_k^{(n)}(\cdot)$ is the signaling waveform for the n-th symbol of user k. Note that $c_k^{(n)}(\cdot)$ is not necessarily a spreading waveform.

- *Signal separation*
  - For FDMA, $\{s_k(t)\}_{k=1}^K$ occupy orthogonal frequency slots (possibly separated by guard bands).
  - For TDMA, $\{s_k(t)\}_{k=1}^K$ occupy orthogonal time slots (possibly separated by guard times).
  - For CDMA, $\{s_k(t)\}_{k=1}^K$ have their energy spread out roughly uniformly over time and frequency. The signals are not necessarily orthogonal, and they may not even be linearly independent.

DS/SS CDMA

- For DS/CDMA, the signaling waveform for n-th symbol of user k is the spreading waveform given by

$$c_k^{(n)}(t) = \sum_{j=0}^{N-1} c_{k,j}^{(n)} \, g_{T_c}(t - jTc) \tag{20.12}$$

  with $T_s = NT_c$. The corresponding chip sequence can be written compactly as

$$\mathbf{c}_k^{(n)} = [c_{k,0}^{(n)} \; c_{k,1}^{(n)} \; \cdots \; c_{k,N-1}^{(n)}]^\top \tag{20.13}$$

  and ${\mathbf{c}_k^{(n)}}^\dagger \mathbf{c}_k^{(n)} = 1$.

- *Synchronous versus asynchronous users*

  ◦ In general, the received signal in AWGN is given by

$$y(t) = \sum_{k=1}^{K} A_k \; \sum_n z_k^{(n)} \, c_k^{(n)}(t - nT_s - \tau_k) \, e^{j\phi_k} \; + \; w(t) \tag{20.14}$$

  where $A_k = \sqrt{\mathcal{E}_{\mathrm{s},k}}$.

  ◦ Without loss of generality, we may assume that $\tau_k \in [0, T_s]$.

  ◦ For synchronous users, $\tau_k = 0$ and $\phi_k = 0$ for all k.

  ◦ For asynchronous users, one-shot (symbol-by-symbol) detection is not optimum. We hence need to consider a frame of length μT_s, $\mu > 1$, for detection. The asynchronous user problem over μT_s can be converted to an equivalent "synchronous-user" problem with $\mu + 2(K-1)(\mu+1)$ users.

  ◦ For long (random) sequence CDMA, the performance for asynchronous users with multishot detection can be approximated by the performance for synchronous users with one-shot detection.

References

[1] S. Verdu. *Multiuser Detection*. Cambridge University Press, United Kingdom, 1998.

21 Lecture 21

Synchronous user model

$$y(t) = \sum_{k=1}^{K} \sqrt{\mathcal{E}_k} \, \sum_n z_k^{(n)} \, c_k^{(n)}(t - nT_s) \; + \; w(t) \, . \tag{21.1}$$

- Sufficient statistics for detection are given by:

$$y_k^{(n)} = \langle y(t), c_k^{(n)}(t - nT_\mathrm{s}) \rangle \tag{21.2}$$

- Using the zero-ICI condition satisfied by $g_{T_c}(\cdot)$ it is easy to show that

$$y_k^{(n)} = \sum_{\ell=1}^{K} A_\ell \, z_\ell^{(n)} \, \langle c_\ell^{(n)}(t - nT_\mathrm{s}), c_k^{(n)}(t - nT_\mathrm{s}) \rangle \; + \; w_k^{(n)} \, . \tag{21.3}$$

- Note that $y_k^{(n)}$ is a function of only $\{z_\ell^{(n)}\}_{\ell=1}^K$, but not a function of $\{z_\ell^{(n')}\}_{\ell=1}^K$ for any $n' \neq n$. Also, $\{w_\ell^{(n')}\}_{\ell=1}^K$ and $\{w_\ell^{(n)}\}_{\ell=1}^K$ are independent for $n' \neq n$. Thus, one-shot multiuser processing is optimum, i.e., the symbol decisions for the users can be made one symbol interval at time without loss of optimality.

- Without loss of generality, consider symbol interval $[0, T_s]$, i.e., $n = 0$, and drop the superscript "0" for convenience. Then

$$y_k = \sum_{\ell=1}^{K} A_\ell \, z_\ell \, \langle c_\ell(t - nT_\mathrm{s}), c_k(t - nT_\mathrm{s}) \rangle \; + \; w_k \, . \tag{21.4}$$

  with $w_k = \langle w(t), c_k(t) \rangle$.

$$y_k = A_k z_k + \sum_{\ell \neq k} A_\ell z_\ell \rho_{\ell,k} + w_k \tag{21.5}$$

  where

$$\rho_{\ell,k} = \langle c_\ell(t), c_k(t) \rangle == \sum_{i=0}^{N-1} \; c_{\ell,i} \, c_{k,i}^\star = \mathbf{c}_k^\dagger \, \mathbf{c}_\ell \tag{21.6}$$

  and $w_k \sim \mathcal{CN}(0, N_0)$. The noise components are not independent; in particular, $\mathrm{E}[w_k w_\ell^\star] = N_0 \rho_{\ell,k}$.

- If $K \leq N$, the chip sequences can be made orthogonal, i.e., $\rho_{\ell,k} = \delta[k - \ell]$, and hence

$$y_k = A_k z_k + w_k \tag{21.7}$$

  which is the same as the expression for the MF output for a single user in AWGN. Thus single-user detection is optimum in this case, and the performance obtained is the same as that without the multiple-access interference (MAI) from other users. An example of an orthogonal sequence set is the set of Walsh-Hadamard sequences, which are used in the forward link of IS-95 based CDMA systems.

Single user detection

- In general when the sequences are not necessarily orthogonal

$$y_k = A_k z_k + I_k + w_k \tag{21.8}$$

where

$$I_k = \sum_{\ell \neq k} A_\ell \, z_\ell \, \rho_{\ell,k} \, . \tag{21.9}$$

If we approximate I_k by a zero mean, PCG random variable, then the conditional pdf of Y_k given z_k is PCG with mean $A_k z_k$.

- A single user (SU) detector treats I_k as a CCG random variable and makes a decision on z_k based purely on y_k, i.e., ignoring $\{y_\ell\}_{\ell \neq k}$.

- For SU hard decision making, the ML decision for z_k based on y_k is given by:

$$\hat{z}_{k,\text{SU-MF}} \approx \arg \min_{z_k \in \mathcal{S}} |y_k - A_k z_k|^2 \, . \tag{21.10}$$

For binary signaling, $z_k \in \{+1, -1\}$, and we obtain:

$$\hat{z}_{k,\text{SU-MF}} \approx \text{sgn}\left[\text{Re}(Y_k)\right] \, . \tag{21.11}$$

- For SU soft decision making, we send y_k to the decoder and the decoder may use knowledge of A_k in decoding.

- For hard decisions, the performance metric of interest is of course the probability of error $\text{P}_{\text{e},k} = \text{P}\{\hat{z}_k \neq z_k\}$.

- For soft decisions, a useful performance metric is the signal-to-intereference ratio (SIR) in the soft decision statistic, defined by:

$$\text{SIR}_k = \frac{\text{E}\left[|\text{E}[y_k | z_k]|^2 \right]}{\text{Var}(y_k | z_k)} \, . \tag{21.12}$$

Single User Detection – Performance Analysis

- *Binary signaling assumption*: For the analysis in this section we make the simplifying assumption that symbols and spreading sequences are binary, i.e.,

$$z_k \in \{+1, -1\} \, , \text{ and } \; c_{k,i} \in \{+\frac{1}{\sqrt{N}}, -\frac{1}{\sqrt{N}}\} \, . \tag{21.13}$$

- *Case 1: Orthogonal users*

  - Since $\rho_{\ell,k} = 0$ for $\ell \neq k$,

$$y_k = A_k \, z_k + w_k \, . \tag{21.14}$$

◦ The SIR for user k is given by:

$$\text{SIR}_k = \frac{\text{E}\left[|\text{E}[y_k|z_k]|^2\right]}{\text{Var}(y_k|z_k)} = \frac{\text{E}\left[(A_k z_k)^2\right]}{N_0} = \frac{\mathcal{E}_{\text{b},k}}{N_0} \,. \tag{21.15}$$

◦ The BER for user k for the MPE SU decision rule of (21.11) is given by:

$$\text{P}_{\text{b},k} = \text{P}\left(\{\hat{z}_k = 1\}|\{z_k = -1\}\right) = \text{P}\left(\{y_k > 0\}|\{z_k = -1\}\right) = Q\left(\sqrt{\frac{2\mathcal{E}_{\text{b},k}}{N_0}}\right) = Q(\sqrt{2\text{SIR}_k}\,) \,. \tag{21.16}$$

As expected since the interference is completely cancelled out.

- *Case 2: Synchronous users with non-orthogonal short spreading sequences*

◦ Assuming that the bits of the users are i.i.d. Bernouilli$(\pm 1, 0.5)$

$$\text{E}[y_k|z_k] = A_k z_k \quad \Longrightarrow \quad \text{E}\left[|\text{E}[y_k|z_k]|^2\right] = A_k^2 \tag{21.17}$$

and

$$\text{Var}(y_k|z_k) = \text{Var}\left(\sum_{\ell \neq k} A_\ell z_\ell \rho_{\ell,k} + w_k\right) = \sum_{\ell \neq k} A_\ell^2 \rho_{\ell,k}^2 + N_0 \,. \tag{21.18}$$

Thus

$$\text{SIR}_k = \frac{A_k^2}{\sum_{\ell \neq k} A_\ell^2 \rho_{\ell,k}^2 + N_0} = \frac{\frac{\mathcal{E}_{\text{b},k}}{N_0}}{1 + \sum_{\ell \neq k} \frac{\mathcal{E}_{\text{b},\ell}}{N_0} \rho_{\ell,k}^2} \tag{21.19}$$

◦ The BER for user k for the MPE decision rule of (21.11) is to be computed by averaging over the distribution of the bits of the other users. We do this by first computing $\text{P}_{\text{b},k}$ conditioned on the bits of the others users, and then average over the distribution of these bits.

$$\begin{aligned}
\text{P}_{\text{b},k} &= \text{P}\left(\{\hat{z}_k = -1\}|\{z_k = +1\}\right) = \text{E}\left[\text{P}\left(\{\hat{z}_k = -1\} \,\middle|\, \{z_k = +1\}, \{z_\ell\}_{\ell \neq k}\right)\right] \\
&= \text{E}\left[\text{P}\left(\{y_{k,I} < 0\} \,\middle|\, \{z_k = +1\}, \{z_\ell\}_{\ell \neq k}\right)\right] = \text{E}\left[Q\left(\frac{A_k + \sum_{\ell \neq k} A_\ell z_\ell \rho_{\ell,k}}{\sqrt{N_0/2}}\right)\right] \\
&= \text{E}\left[Q\left(\sqrt{\frac{2\mathcal{E}_{\text{b},k}}{N_0}} + \sum_{\ell \neq k} \sqrt{\frac{2\mathcal{E}_{\text{b},\ell}}{N_0}}\, z_\ell\, \rho_{\ell,k}\right)\right] \\
&= \sum_{\tilde{\mathbf{z}} \in \{+1,-1\}^{K-1}} Q\left(\sqrt{\frac{2\mathcal{E}_{\text{b},k}}{N_0}} + \sum_{\ell \neq k} \sqrt{\frac{2\mathcal{E}_{\text{b},\ell}}{N_0}}\, z_\ell\, \rho_{\ell,k}\right)
\end{aligned} \tag{21.20}$$

where $\tilde{\mathbf{z}} = [z_0 \;\cdots\; z_{k-1} \; z_{k+1} \;\cdots\; z_K]$. Note that from the above equation we can immediately conclude that

$$\text{P}_{\text{b},k} \geq Q\left(\sqrt{\frac{2\mathcal{E}_{\text{b},k}}{N_0}}\right) = \text{P}_{\text{b},k}(\text{orthogonal signaling}) \quad \text{[Why?]} \tag{21.21}$$

Also note that the number of terms in the sum grows exponentially with K, and hence it is difficult to compute $\mathrm{P}_{\mathrm{b},k}$ exactly when the K is large.

∘ Gaussian approximation for $\mathrm{P}_{\mathrm{b},k}$: For large K

$$\sum_{\ell \neq k} A_\ell z_\ell \rho_{\ell,k} + w_{k,I} \;\approx\; \sim \mathcal{N}(0, \sigma_1^2) \tag{21.22}$$

where

$$\sigma_1^2 = \mathrm{Var}(y_{k,I}|z_k) = \sum_{\ell \neq k} A_\ell^2 \rho_{\ell,k}^2 + \frac{N_0}{2} \; . \tag{21.23}$$

Thus we can approximate $\mathrm{P}_{\mathrm{b},k}$ as

$$\mathrm{P}_{\mathrm{b},k} = \mathrm{P}\left(\{\hat{z}_k = -1\} | \{z_k = +1\}\right) \approx Q\left(\frac{A_k}{\sigma}\right) \; . \tag{21.24}$$

• *Case 3: Synchronous users with long (random) spreading sequences*

∘ Assuming that the bits of the users are i.i.d. Bernouilli($\pm 1, 0.5$), and the chips of the users are i.i.d. Bernouilli($\pm\frac{1}{\sqrt{N}}, 0.5$), we can show that (see HW#6)

$$\mathrm{SIR}_k = \frac{A_k^2}{\frac{1}{N}\sum_{\ell \neq k} A_\ell^2 + N_0} = \frac{\frac{\mathcal{E}_{\mathrm{b},k}}{N_0}}{1 + \frac{1}{N}\sum_{\ell \neq k} \frac{\mathcal{E}_{\mathrm{b},\ell}}{N_0}} \tag{21.25}$$

For equal power users, $\mathcal{E}_{\mathrm{b},\ell} = \mathcal{E}_{\mathrm{b}}$ for all ℓ, and we obtain

$$\mathrm{SIR}_k = \frac{\frac{\mathcal{E}_{\mathrm{b}}}{N_0}}{1 + \frac{K-1}{N}\frac{\mathcal{E}_{\mathrm{b}}}{N_0}} \approx \frac{N}{K-1} \;\text{ for large } K \text{ or large } \mathcal{E}_{\mathrm{b}}/N_0 \tag{21.26}$$

∘ The BER for user k for the MPE decision rule of (21.11) is to be computed by averaging over the distribution of the bits as well as the chips. The procedure is similar to that used for Case 2.

$$\mathrm{P}_{\mathrm{b},k} = \mathrm{E}\left[Q\left(\sqrt{\frac{2\mathcal{E}_{\mathrm{b},k}}{N_0}} + \sum_{\ell \neq k}\sqrt{\frac{2\mathcal{E}_{\mathrm{b},\ell}}{N_0}}\, z_\ell\, \rho_{\ell,k}\right)\right] \tag{21.27}$$

where the expectation is taken over the distribution of the bits and the chips. It is clear that computing this expectation is even more cumbersome than in Case 2.

∘ Gaussian approximation for $\mathrm{P}_{\mathrm{b},k}$: For large K, and equal powers, using steps similar to those used in Case 2, we can approximate $\mathrm{P}_{\mathrm{b},k}$ as

$$\mathrm{P}_{\mathrm{b},k} \approx Q\left(\sqrt{\frac{K-1}{N}}\right) \approx Q\left(\sqrt{\mathrm{SIR}_k}\right) \; . \tag{21.28}$$

∘ See Figures 21.1 and 21.2 for typical numerical results.

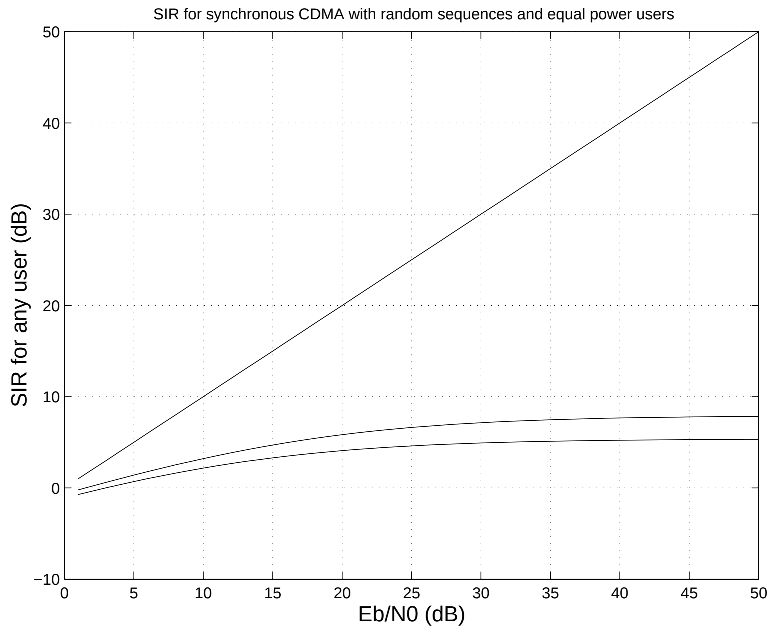


Figure 21.1: SIR for synchronous CDMA with $N = 31$

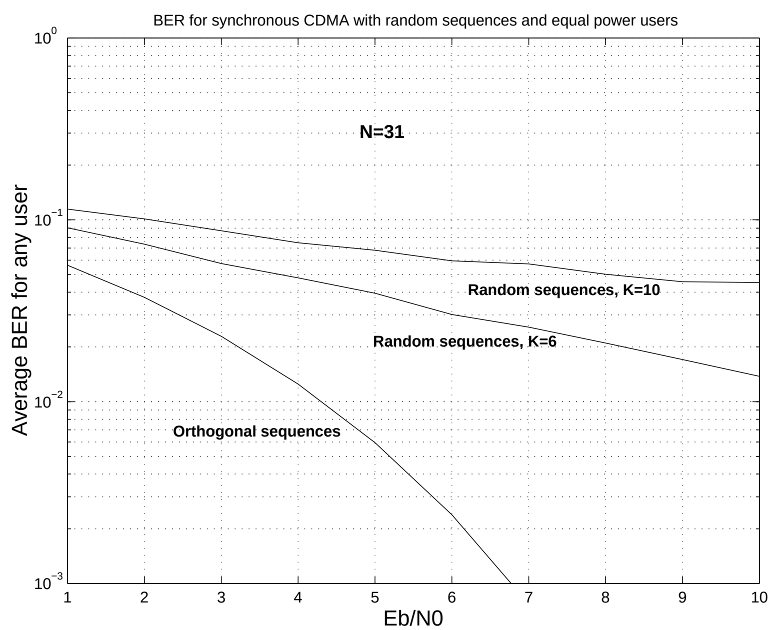


Figure 21.2: BER for synchronous CDMA with $N = 31$