referred to as a *multiple-input, multiple-output (MIMO)* wireless communication system, which includes receive diversity and transmit diversity as special cases of space diversity. The novel feature of the MIMO system is that, in a rich Rayleigh scattering environment, it can provide a high spectral efficiency, which may be explained as follows: The signals radiated simultaneously by the transmit antennas arrive at the input of each receive antenna in an uncorrelated manner due to the rich scattering mechanism of the channel. The net result is the potential for a spectacular increase in the spectral efficiency of the wireless link. Most importantly, the spectral efficiency increases roughly linearly with the number of transmit or receive antennas, whichever is less. This result assumes that the receiver has knowledge of channel state information. The spectral efficiency of the MIMO system can be further enhanced by including a feedback channel from the transmitter to the receiver, whereby the channel state is also made available to the transmitter, and with it, the transmitter is enabled to exercise control over the transmitted signal.

Increasing spectral efficiency in the face of multipath fading is one important motivation for using MIMO transmission schemes. Another important motivation is the development of *space–time codes*, whose aim is the joint coding of multiple transmit antennas so as to provide protection against channel fading, noise, and interference. In this context, of particular interest is a class of block codes referred to as *orthogonal* and *generalized complex orthogonal space–time block codes*. In this class of codes, the *Alamouti code*, characterized by a two-by-two transmission matrix, is the only full-rate complex orthogonal space–time block code. The Alamouti code satisfies the condition for complex orthogonality or unitarity in both the spatial and temporal sense. In contrast, the generalized complex orthogonal space–time codes can accommodate more than two transmit antennas; they are therefore capable of providing a larger coding gain than the Alamouti code for a prescribed bit error rate and total transmission rate at the expense of a reduced code rate and increased computational complexity. However, unlike the Alamouti code, the generalized complex orthogonal space–time codes satisfy the condition for complex orthogonality only in the temporal sense. Accordingly, the complex orthogonal space–time codes, including the Alamouti code and generalized forms, permit the use of linear receivers.

The complex orthogonal property of the Alamouti code is exploited in the development of a *differential space–time block coding scheme*, which eliminates the need for channel estimation and thereby simplifies the receiver design. This simplification is, however, attained at the expense of degradation in receiver performance, compared with the coherent version of the Alamouti code, which assumes knowledge of the channel state information at the receiver.

Space was also discussed in the context of space-division multiple-access (SDMA), the mechanization of which relies on the use of highly directional antennas. SDMA improves system capacity by allowing a greater reuse of the available spectrum through a combination of two approaches: minimization of the effects of interference and increased signal strength for both the user terminal and the base station. Advanced techniques such as phased-array antennas and adaptive antennas, which have been researched extensively under the umbrellas of signal processing and radar for more than three decades, are well suited for implementing the practical requirements of both approaches.

Under the three theme examples, we discussed three different BLAST architectures issues relating to antenna diversity, spectral efficiency, as well as keyhole channels. Each of the BLAST architectures, namely, diagonal-BLAST, vertical-BLAST, and Turbo-BLAST, offers distinct features of its own. Diagonal-BLAST (D-BLAST) makes it possible to closely approximate the ergodic channel capacity in a rich scattering environment and may therefore be viewed as the benchmark BLAST architecture. But it is impractical, as it suffers from a serious space–time edge wastage. Vertical-BLAST (V-BLAST) mitigates the computational difficulty problem of D-BLAST at the expense of a reduced channel capacity. Turbo-BLAST uses a random layered space–time code at the transmitter and incorporates the turbo coding principle in designing an iterative receiver. In so doing, Turbo-BLAST offers a significant improvement in spectral efficiency over V-BLAST, yet the computational complexity is maintained at a manageable level. In terms of performance, Turbo-BLAST outperforms V-BLAST for a prescribed (N_t, N_r) antenna configuration, but does not perform as well as D-BLAST.

The different BLAST architectures were discussed in Theme Example 1. The material presented in Theme Example 2 taught us the following:

- The two basic forms of diversity, namely, transmit diversity and receive diversity, play complementary roles, with both of them being located at the base station.
- For low SNR and fixed spectral efficiency, V-BLAST outperforms space–time block codes (STBCs) on $\{N_t, N_r\}$ antenna configurations with $N_r > N_t$.
- Assuming the use of forward error-correction channel codes, a two-by-two STBC system could provide an adequate performance for wireless communications at low SNR.
- Diversity order is determined experimentally by measuring the asymptotic slope of the average frame error rate (or average symbol error rate) plotted versus the signal-to-noise ratio on a log-log scale.
- MIMO systems provide a trade-off between outage capacity and diversity order, depending on how the system is configured.

The degenerate occurrence of keyhole channels, discussed in Theme Example 3 arises when the rank of the channel matrix is reduced to unity, in which case the capacity of the MIMO link is equivalent to that of a single-input, single-output link operating at the same signal-to-noise ratio. Fortunately, the physical occurrence of keyhole channels is a rare phenomenon.

One last comment is in order: the discussion of channel capacity presented in the chapter focused on *single-user* MIMO links. Although, indeed, wireless systems in current use cater to the needs of multiple users, the focus on single users may be justified on the following grounds:

- The derivation of MIMO channel capacity is much easier to undertake for single users than multiple users.

- Capacity formulas are known for many single-user MIMO cases, whereas the corresponding multiuser ones are unsolved.

Simply put, very little is known about the channel capacity of *multiuser* MIMO links, unless the channel state is known at both the transmitter and receiver.[28]

NOTES AND REFERENCES

[1] For detailed discussions of the receive diversity techniques of selection combining, maximal-ratio combining, and square-law combining, see Schwartz et al. (1966), Chapter 10.

[2] The term "maximal-ratio combiner" was coined in a classic paper on linear diversity combining techniques by Brennan (1959).

[3] The three-point exposition presented in Section 6.2.3 on maximal-ratio combining follows S. Stein in Schwartz (1966), pp. 653–654.

[4] For expository discussions of the many facets of MIMO wireless communications, see the papers by Gesbert et al. (2003), Diggavi et al. (2003), and Goldsmith et al. (2003). The paper by Diggavi et al. includes an exhaustive list of references to MIMO wireless communications and related issues. For books on wireless communications using multiple antennas, see Hottinen et al. (2003) and Vucetic and Yuan (2003).

[5] Impulsive noise due to human-made electromagnetic interference is discussed in Blackard and Rappaport (1993), and Wang and Poor (1999); see also Chapter 2.

[6] The formula of Eq. (6.56), defining the ergodic capacity of a flat-fading channel, is derived in Ericson (1970).

[7] The log-det capacity formula of Eq. (6.59) for MIMO wireless links operating in rich scattering environments was derived independently by Teletar (1995) and Foschini (1996); Teletar's report was published subsequently as a journal article (1999). For a detailed derivation of the log-det capacity formula, see Appendix G.

[8] The Gaussian approximation of the probability density function of the instantaneous channel capacity of a MIMO wireless link, which is governed by the log-det formula, is discussed in detail in Hochwald et al. (2003).

[9] The result that at high signal-to-noise ratios the outage probability and frame (burst) error probability are the same is derived in Zheng and Tse (2002).

[10] MIMO wireless communications systems incorporating the use of feedback channels are discussed in Vishwanath et al. (2001), Simon and Moustakas (2003), and Hochwald et al. (2003). The latter paper introduces the notion of rate feedback by quantizing the instantaneous channel capacity of the MIMO link.

[11] The effect of correlation fading on the channel capacity of MIMO wireless communications is discussed in Shiu et al. (2000) and Smith et al. (2003).

[12] Space–time trellis codes are discussed in Tarokh et al. (1998).

[13] The Alamouti code was pioneered by Siavash Alamouti (1998); the code has been adopted in third-generation (3G) wireless systems, in which it is known as space–time transmit diversity (STTD).

[14] The generalized space–time orthogonal codes were originated by Tarokh et al. (1999a,b).

[15] The decoding algorithms (written in MATLAB) for the Alamouti code **S**, and the orthogonal space–time codes $\mathbf{G}_3$, $\mathbf{G}_4$, $\mathbf{H}_3$, and $\mathbf{H}_4$ due to Tarokh et al. are presented in the Solutions Manual to this book. It should, however, be noted that there are minor errors in the original decoding algorithms for $\mathbf{H}_3$, and $\mathbf{H}_4$ listed in the Appendix to the paper by Tarokh et al. (1998). These errors have been corrected in the pertinent MATLAB codes.

[16] Differential space–time block coding, based on the Alamouti code, was first described by Tarokh and Jafarkhani (2000). See also the article by Diggavi et al. (2002), which combines this form of differential coding with orthogonal frequency-division multiplexing (OFDM) for signal transmission over fading frequency-selective channels; OFDM was discussed in Chapter 3.

[17] Chapter 3 of Liberti and Rappaport (1999) describes more general models for phased arrays other than linear and where gain in elevation angle as well as azimuth is of interest. Chapter 8 of the same book describes various algorithms for adapting the weighting vector, depending upon the direction of arrival of the signal.

[18] The circular model for effective scatterers was proposed in Lee (1982).

[19] In Chapter 7 of Liberti and Rappaport (1999), the single-bounce elliptical model is described in greater detail. Note that the model does not take into account the effects of diffraction.

[20] The D-BLAST architecture was pioneered by Foschini (1996) and discussed further in the papers by Foschini and Gans (1999) and Foschini et al. (2003).

[21] The first experimental results in the V-BLAST architecture were originally reported in the article by Golden et al. (1999); see also the paper by Foschini et al. (2003), in which this particular form of BLAST is referred to as horizontal-BLAST, or H-BLAST.

[22] The Turbo-BLAST architecture was first described by Sellathurai and Haykin (1998), with additional results reported subsequently in the papers by the same authors (2000, 2002, 2003).

[23] The experimental results presented in Figs. 6.40 through 6.42 are reproduced from the paper by Sellathurai and Haykin (2002) with permission of the IEEE.

[24] According to deHaas (1927, 1928), the possibility of using antenna diversity for mitigating short-term fading effects in radio communications was apparently first discovered in experiments with spaced receiving antennas operating in the high-frequency (HF) band. For additional historical notes, see Chapter 10 by Seymour Stein in Schwartz et al. (1966).

[25] For definitions of the diversity order and multiplexing gain of MIMO wireless communication systems and the implications of these definitions in terms of system behavior, see Digavi (2003).

[26] Keyhole channels, also dubbed pinhole channels, were described independently by Gesbert et al. (2002) and Chizhik et al. (2002).

[27] The GBGP model for MIMO wireless links is described in Gesbert et al. (2002).

[28] Multiuser MIMO wireless systems are discussed in Diggavi et al. (2003) and Goldsmith et al. (2003).

ADDITIONAL PROBLEMS

Diversity-on-receive techniques

Problem 6.21 A receive-diversity system uses a selection combiner with two diversity paths. An outage occurs when the instantaneous signal-to-noise ratio γ drops below $0.25\gamma_{\text{av}}$, where γ_{av} is the average signal-to-noise ratio. Determine the probability of outage experienced by the receiver.

Problem 6.22 The average signal-to-noise ratio in a selection combiner is 20 dB. Compute the probability that the instantaneous signal-to-noise ratio of the device drops below $\gamma = 10$ dB for the following number of receive antennas:

(a) $N_r = 1$
(b) $N_r = 2$
(c) $N_r = 3$
(d) $N_r = 4$

Comment on your results.

Problem 6.23 Repeat Problem 6.22 for $\gamma = 15$ dB.

Problem 6.24 In Section 6.2.2, we derived the optimum values of Eq. (6.18) for complex weighting factors of the maximal-ratio combiner using the Cauchy–Schwartz inequality. This problem addresses the same issue, using the standard maximization procedure. To simplify matters, the number N_r of diversity paths is restricted to two, with the complex weighting parameters denoted by a_1 and a_2.

Let

$$a_k = x_k + jy_k \qquad k = 1, 2$$

Then the complex derivative with respect to a_k is defined by

$$\frac{\partial}{\partial a_k^*} = \frac{1}{2}\left(\frac{\partial}{\partial x_k} + j\frac{\partial}{\partial y_k}\right) \qquad k = 1, 2$$

Applying this formula to the combiner's output signal-to-noise ratio γ_c of Eq. (6.14), derive Eq. (6.18).

Problem 6.25 In this problem, we develop an approximate formula for the probability of error, P_e, produced by a maximal-ratio combiner for coherent FSK. We start with Eq. (6.25), and for small γ_{mrc}, we may use the following approximation for the probability density function:

$$f_\Gamma(\gamma_{\text{mrc}}) = \frac{1}{\gamma_{\text{av}}^{N_r}(N_r - 1)!}\gamma_{\text{mrc}}^{N_r - 1}$$

(a) Using the conditional probability of error for coherent BFSK, that is,

$$\text{Prob}(\text{error}|\gamma_{\text{mrc}}) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{1}{2}\gamma_{\text{mrc}}}\right)$$

derive the approximation

$$P_e \approx \frac{1}{2\left(\frac{1}{2}\gamma_{\text{av}}\right)^{N_r}(N_r-1)!}\int_0^\infty \text{erfc}(\sqrt{y})y^{N_r-1}dy$$

where $y = \frac{1}{2}\gamma_{\text{mrc}}$.

(b) Integrating the definite integral by parts and using the definition of the complementary error function, show that

$$P_e \approx \frac{1}{2\sqrt{\pi}\left(\frac{1}{2}\gamma_{\text{av}}\right)^{N_r}N_r!}\int_0^\infty e^{-y}y^{N_r-\frac{1}{2}}dy$$

(c) Finally, using the definite integral

$$\int_0^\infty e^{-y}y^{N_r-\frac{1}{2}}dy = \left(N_r-\frac{1}{2}\right)!$$

obtain the desired approximation

$$P_e \approx \frac{1}{2\sqrt{\pi}\left(\frac{1}{2}\gamma_{\text{av}}\right)^{N_r}}\frac{\left(N_r-\frac{1}{2}\right)!}{N_r!}$$

Problem 6.26

(a) Using the approximation for $f_\Gamma(\gamma_{\text{mrc}})$ given in Problem 6.25, determine the probability of symbol error for a maximal-ratio combiner that uses noncoherent BFSK.

(b) Compare your result of part(a) with that of Problem 6.25 for coherent BFSK.

Problem 6.27

(a) Continuing the approximation to $f_\Gamma(\gamma_{\text{mrc}})$, determine the probability of symbol error for a maximal-ratio combiner that uses coherent BPSK.

(b) Compare your result of part(a) with that of Problem 6.25 for coherent BFSK.

Problem 6.28 As discussed in Section 6.2.3, an *equal-gain combiner* is a special form of the maximal-ratio combiner for which the weighting factors are all equal. For convenience of presentation, the weighting parameters are set to unity. Assuming that the instantaneous signal-to-noise ratio γ is small compared with the average signal-to-noise ratio γ_{av}, derive an approximate formula for the probability density function of γ.

Problem 6.29 Compare the performances of the following linear diversity-on-receive techniques:

(a) Selection combiner
(b) Maximal-ratio combiner
(c) Equal-gain combiner

Base the comparison on signal-to-noise improvement, expressed in dB, for $N_r = 2, 3, 4, 5$, and 6 diversity branches.

Problem 6.30 Show that the maximum-likelihood decision rule for the maximal-ratio combiner may be formulated in the following equivalent forms:

(a) Choose symbol s_i over s_k if and only if

$$(\alpha_1^2 + \alpha_2^2)|s_i|^2 - y_1 s_i^* - y_1^* s_i < (\alpha_1^2 + \alpha_2^2)|s_k|^2 - y_1 s_k^* - y_1^* s_k \qquad k \neq i$$

(b) Choose symbol s_i over s_k if and only if

$$(\alpha_1^2 + \alpha_2^2 - 1)|s_i|^2 + d^2(y_1, s_i) < (\alpha_1^2 + \alpha_2^2 - 1)|s_k|^2 + d^2(y_1, s_k) \qquad k \neq i$$

Here, $d^2(y_1, s_i)$ denotes the squared Euclidean distance between the received signal y_1 and constellation points s_i.

Problem 6.31 It may be argued that, in a rather loose sense, transmit-diversity and receive-diversity antenna configurations are the *dual* of each other, as illustrated in Fig. 6.46.

(a) Taking a general viewpoint, justify the mathematical basis for this duality.

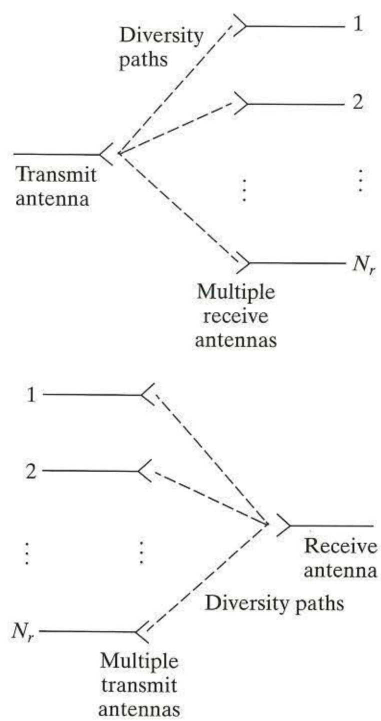


FIGURE 6.46 Diagram for Problem 6.31.

(b) However, we may cite the example of frequency-division diplexing (FDD), in which, in a strict sense, the duality depicted in Fig. 6.44 is violated. How is it possible for the violation to arise in this example?

MIMO Channel Capacity

Problem 6.32 In this problem, we continue with the solution to Problem 6.9, namely,

$$C \to \left(\frac{\lambda_{\text{av}}}{\log_e 2}\right)\rho \qquad \text{as } N \to \infty$$

where $N_t = N_r = N$ and λ_{av} is the average eigenvalue of $\mathbf{H}\mathbf{H}^\dagger = \mathbf{H}^\dagger\mathbf{H}$,

(a) Justify the asymptotic result given in Eq. (6.61)—that is,

$$\frac{C}{N} \geq \text{constant}$$

What is the value of the constant?

(b) What conclusion can you draw from the asymptotic result?

Problem 6.33 By and large, the treatment of the ergodic capacity of a MIMO channel, as presented in Sections 6.3 and 6.5, focused on the assumption that the channel is Rayleigh distributed. In this problem, we expand on that assumption by considering the channel to be Rician distributed. In such an environment, we may express the channel matrix as

$$\mathbf{H} = a\mathbf{H}_{\text{sp}} + \mathbf{H}_{\text{sc}}$$

where $\mathbf{H}_{\text{sp}}$ and $\mathbf{H}_{\text{sc}}$ denote the specular and scattered components, respectively. To be consistent with the MIMO model described in Section 6.3, the entries of both $\mathbf{H}_{\text{sp}}$ and $\mathbf{H}_{\text{sc}}$ have unit amplitude variance, with $\mathbf{H}_{\text{sp}}$ being deterministic and $\mathbf{H}_{\text{sc}}$ consisting of iid complex Gaussian-distributed variables with zero mean. The scaling parameter a is related to the Rice K-factor by the formula

$$K = 10\log_{10} a^2 \text{dB}$$

(a) Considering the case of a pure line of sight (LOS), show that the MIMO channel has the deterministic capacity

$$C = \log_2(1 + N_r a^2 \rho) \quad \text{bits/s/Hz}$$

where N_r is the number of receive antennas and ρ is the total signal-to-noise ratio at each receiver input.

(b) Compare the result obtained in part (a) with that pertaining to the pure Rayleigh distributed MIMO channel.

(c) Explore the more general situation, involving the combined presence of both the specular and scattered components in the channel matrix $\mathbf{H}$.

Problem 6.34 Suppose that an additive, temporally stationary Gaussian interference $\mathbf{v}(t)$ corrupts the basic channel model of Eq. (6.48). The interference $\mathbf{v}(t)$ has zero mean and correlation matrix $\mathbf{R}_v$. Evaluate the effect of $\mathbf{v}(t)$ on the ergodic capacity of the MIMO link.

Problem 6.35 Consider a MIMO link for which the channel may be considered to be essentially constant for τ uses of the channel.

(a) Starting with the basic channel model of Eq. (6.48), formulate the input–output relationship of this link, with the input described by the N_r-by-τ matrix

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_\tau]$$

(b) How is the log-det capacity formula of the link correspondingly modified?

Orthogonal Space-Time Block Codes

Problem 6.36 The objective of this problem is to fill in the mathematical details that lie behind the formulas of Eqs. (6.104) and (6.105) for the maximum-likelihood estimates $\hat{s}_1$ and $\hat{s}_2$.

(a) Starting with Eq. (6.102) for the combiner output $\tilde{y}_k$ and using Eq. (6.103) for the probability density function of the additive complex Gaussian noise $\tilde{v}_k$, formulate the expression for the likelihood function of transmitted symbol $\tilde{s}_k$, $k = 1,2$.

(b) Hence, using the result of part (a), derive the formulas of Eqs. (6.103) and (6.104).

Problem 6.37 Figure 6.47 shows the extension of orthogonal space–time codes to the Alamouti code, using two antennas on transmit and receive. The sequence of signal encoding and transmissions is identical to that of the single-receiver case of Fig. 6.18. Table 6.5(a) defines the channels between the transmit and receive antennas. Table 6.5(b) defines the outputs of the receive antennas at times t' and $t' + T$, where T is the symbol duration.

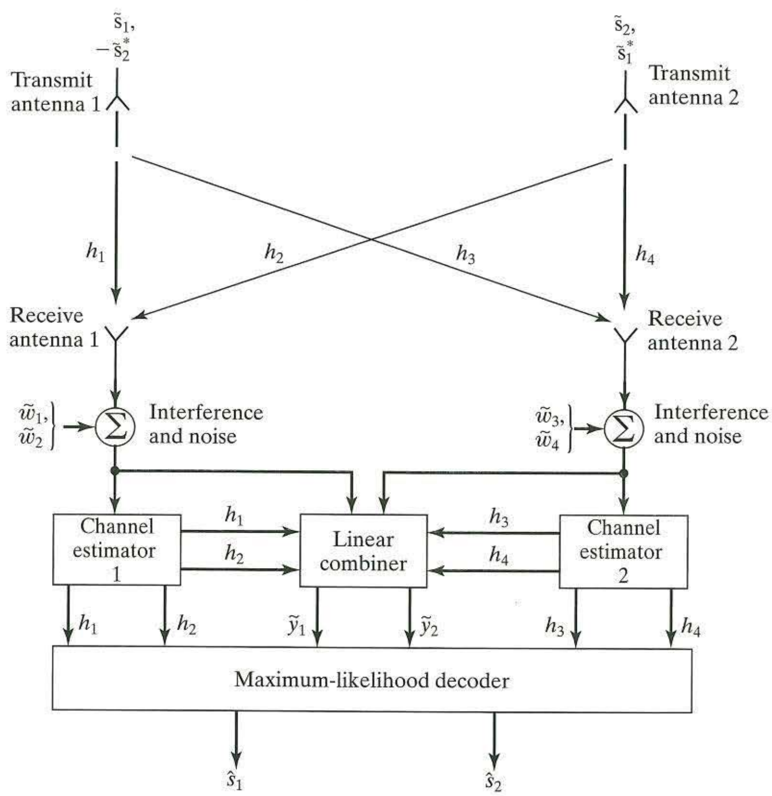


FIGURE 6.47 Diagram for Problem 6.37.

TABLE 6.5 Table for Problem 6.36.

(a)

	Receive antenna 1	Receive antenna 2
Transmit antenna 1	h_1	h_3
Transmit antenna 2	h_2	h_4

(b)

	Receive antenna 1	Receive antenna 2
Time t'	$\tilde{x}_1$	$\tilde{x}_3$
Time $t' + T$	$\tilde{x}_2$	$\tilde{x}_4$

(a) Derive expressions for the received signals $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$, and $\tilde{x}_4$, including the respective additive noise components, in terms of the transmitted symbols.

(b) Derive expressions for the line of combined outputs in terms of the received signals.

(c) Derive the maximum-likelihood decision rule for the estimates $\tilde{s}_1$ and $\tilde{s}_2$.

Problem 6.38 This problem explores a new interpretation of the Alamouti code. Let

$$\tilde{s}_i = s_i^{(1)} + js_i^{(2)} \qquad i = 1, 2$$

where $s_i^{(1)}$ and $s_i^{(2)}$ are both real numbers. The complex entry $\tilde{s}_i$ in the two-by-two Alamouti code is represented by the two-by-two real orthogonal matrix

$$\begin{bmatrix} s_i^{(1)} & s_i^{(2)} \\ -s_i^{(2)} & s_i^{(1)} \end{bmatrix} \qquad i = 1, 2$$

Likewise, the complex-conjugated entry $\tilde{s}_i^*$ is represented by the two-by-two real orthogonal matrix

$$\begin{bmatrix} s_i^{(1)} & -s_i^{(1)} \\ s_i^{(2)} & s_i^{(2)} \end{bmatrix} \qquad i = 1, 2$$

(a) Show that the two-by-two complex Alamouti code $\mathbf{S}$ is equivalent to the four-by-four *real* transmission matrix

$$\mathbf{S}_4 = \left[\begin{array}{cc:cc} s_1^{(1)} & s_1^{(2)} & s_2^{(1)} & s_2^{(2)} \\ -s_1^{(2)} & s_1^{(1)} & -s_2^{(2)} & s_2^{(1)} \\ \hdashline -s_2^{(1)} & s_2^{(2)} & s_1^{(1)} & -s_1^{(2)} \\ -s_2^{(2)} & -s_2^{(1)} & s_1^{(2)} & s_1^{(1)} \end{array}\right]$$

(b) Show that $\mathbf{S}_4$ is an orthogonal matrix.

(c) What is the advantage of the complex code $\mathbf{S}$ over the real code $\mathbf{S}_4$?

Problem 6.39

(a) Show that the generalized complex orthogonal space-time codes of Eqs. (6.107) and (6.108) satisfy the temporal orthogonality condition

$$\mathbf{G}^{\dagger}\mathbf{G} = \mathbf{I}$$

where the superscript † denotes Hermitian transposition and $\mathbf{I}$ denotes the identity matrix.

(b) Likewise, show that the sporadic complex orthogonal space-time codes of Eqs. (6.109) and (6.110) satisfy the temporal orthogonality condition

$$\mathbf{H}^{\dagger}\mathbf{H} = \mathbf{I}$$

Problem 6.40 Applying the maximum-likelihood decoding rule, derive the optimum receivers for the generalized complex orthogonal space–time codes of Eqs. (6.107) and (6.108).

Problem 6.41 Repeat Problem 6.40 for the sporadic complex orthogonal space–time codes of Eqs. (6.109) and (6.110).

Problem 6.42 Show that the channel capacity of the Alamouti code is equal to the sum of the channel capacities of two single-input, single-output systems.

Differential Space–Time Block Coding

Problem 6.43 Equation (6.116) defines the input–output matrix relationship of the differential space–time block coding system described in Section 6.7. Starting with Eqs. (6.98) and (6.99), derive Eq. (6.116).

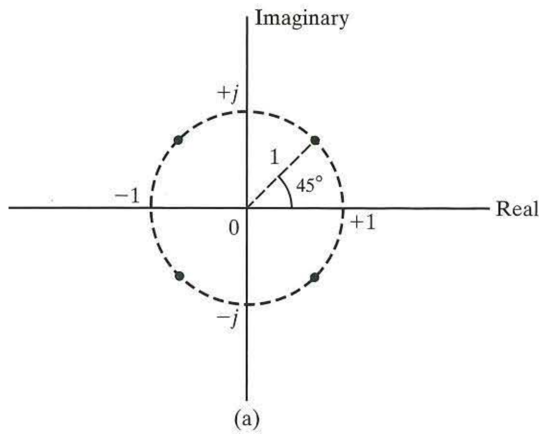
Problem 6.44 The constellation expansion illustrated in Fig. 6.44 is based on the polar baseband representation {-1, +1} for BPSK transmissions of the Alamouti code on antennas 1 and 2. Explore the constellation expansion property of differential space–time coding for the following two situations:

(a) Frame of reference: dibit 00

(b) Frame of reference: dibit 11

Comment on your results.

Problem 6.45 In this problem, we investigate the use of QPSK for transmission of the Alamouti code on antennas 1 and 2. The corresponding input block of data will be in the form of quadbits (i.e., 4-bit blocks). Perform the investigation for each of the two QPSK constellations depicted in Fig. 6.48. Use 0000 as the frame of reference.



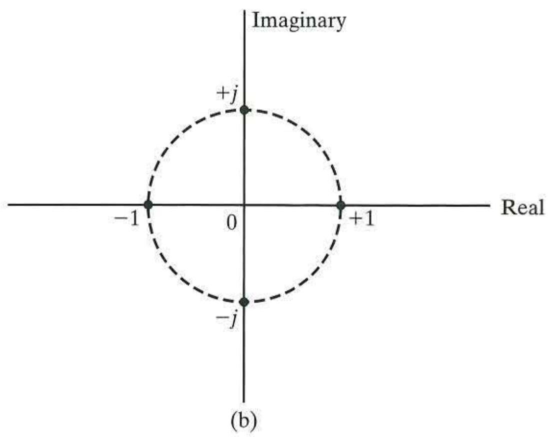


FIGURE 6.48 Diagram for Problem 6.44.

Problem 6.46 Repeat Problem 6.45 for the frame of reference 1111.

Problem 6.47 In the analytic study of differential space–time block coding presented in Section 6.7, we ignored the presence of channel noise. This problem addresses the extension of Eq. (6.116) by including the effect of channel noise.

(a) Starting with Eq. (6.101), expand the formulas of Eqs. (6.116) and (6.117) by including the effect of channel noise modeled as additive white Gaussian noise.

(b) Using the result derived in part (a), expand the formula of Eq. (6.121) by including the effect of channel noise, which consists of the following components:

- **(i)** Two signal-dependent noise terms
- **(ii)** A multiplicative noise term consisting of the product of two additive white Gaussian noise terms

(c) Show that, when the signal-to-noise ratio is high, the noise term (ii) of part (b) may be ignored, with the result that the remaining two signal-dependent noise terms (i) double the average power of noise compared with that experienced in the coherent detection of the Alamouti code.

Theme Examples

Problem 6.48 In this problem, we repeat Experiment 1 of Section 6.10, but this time we investigate the effect of increasing signal-to-noise ratio (SNR) on the symbol error rate (SER) for a prescribed modulation scheme, still operating in a Rayleigh fading environment.

(a) Using 4-PSK for both STBC and V-BLAST, plot the SER versus SNR for the following antenna configurations:

(i) $N_t = 2, N_r = 2$
(ii) $N_t = 2, N_r = 4$

(b) What conclusions do you draw from the experimental results of part (a)?

Problem 6.49 Continuing with Problem 6.48, suppose the STBC and V-BLAST systems use 4-PSK. This time, however, we wish to display the spectral efficiency in bits/s/Hz versus the SNR. How would you expect the performance curve of STBC to compare against that of V-BLAST? Explain.

Problem 6.50 Compare the relative merits of STBC systems versus BLAST systems in terms of the following issues:

- Capacity
- Diversity order
- Multiplexing gain
- Computational complexity

Problem 6.51 In Chapter 2, we discussed the reciprocity theorem in the context of a single-input, single-output wireless communication link. Show that the theorem also applies to Eq. (6.146); that is, show that the channel matrix **H** of the MIMO link satisfies the Hermitian property.

APPENDIX A

Fourier Theory

A.1 THE FOURIER TRANSFORM[1]

Let $g(t)$ denote a *nonperiodic deterministic signal*, expressed as some function of time t. By definition, the *Fourier transform* of the signal $g(t)$ is given by the integral

$$G(f) = \int_{-\infty}^{\infty} g(t)\exp(-j2\pi ft)dt \tag{A.1}$$

where $j = \sqrt{-1}$, and the variable f denotes *frequency*. Given the Fourier transform $G(f)$, the original signal $g(t)$ is recovered exactly using the formula for the *inverse Fourier transform*:

$$g(t) = \int_{-\infty}^{\infty} G(f)\exp(j2\pi ft)df \tag{A.2}$$

Note that in Eqs. (A.1) and (A.2) we have used a lowercase letter to denote the time function and an uppercase letter to denote the corresponding frequency function. The functions $g(t)$ and $G(f)$ are said to constitute a *Fourier-transform pair.*

For the Fourier transform of a signal $g(t)$ to exist, it is sufficient, but not necessary, that $g(t)$ satisfy three conditions, known collectively as *Dirichlet's conditions*:

1. The function $g(t)$ is single valued, with a finite number of maxima and minima in any finite time interval.
2. The function $g(t)$ has a finite number of discontinuities in any finite time interval.
3. The function $g(t)$ is absolutely integrable; that is,

$$\int_{-\infty}^{\infty} |g(t)|\,dt < \infty$$

We may safely ignore the question of the existence of the Fourier transform of a time function $g(t)$ when $g(t)$ is an accurately specified description of a physically realizable signal. In other words, physical realizability is a sufficient condition for the existence of a Fourier transform. Indeed, we may go one step further and state that all finite-energy signals are Fourier transformable.

The absolute value of the Fourier transform $G(f)$, plotted as a function of frequency f, is referred to as the *amplitude spectrum* or *magnitude spectrum* of the signal $g(t)$. By the same token, the argument of the Fourier transform, plotted as a function of frequency f, is referred to as the *phase spectrum* of the signal $g(t)$. The amplitude spectrum is denoted by $|G(f)|$ and the phase spectrum is denoted by $\theta(f)$. When $g(t)$ is a real-valued function of time t, the amplitude spectrum $|G(f)|$ is symmetrical about the origin $f = 0$, whereas the phase spectrum $\theta(f)$ is antisymmetrical about $f = 0$.

Strictly speaking, the theory of the Fourier transform is applicable only to time functions that satisfy the Dirichlet conditions. (Among such functions are energy signals.) However, it would be highly desirable to extend this theory in two ways to include power signals (i.e., signals whose average power is finite). It turns out that this objective can be met through the "proper use" of the *Dirac delta function,* or *unit impulse.*

The Dirac delta function, denoted by $\delta(t)$, is defined as having zero amplitude everywhere except at $t = 0$, where it is infinitely large in such a way that it contains unit area under its curve; that is,

$$\delta(t) = 0 \qquad t \neq 0 \tag{A.3}$$

and

$$\int_{-\infty}^{\infty} \delta(t)dt = 1 \tag{A.4}$$

An implication of this pair of relations is that the delta function is an *even function of time*; that is, $\delta(-t) = \delta(t)$. Another important property of the Delta function is the *replication property* described by

$$\int_{-\infty}^{\infty} g(\tau)\delta(t-\tau)d\tau = g(t) \tag{A.5}$$

which states that the convolution of any function with the delta function leaves that function unchanged.

Tables A.1 and A.2 build on the formulas of Eqs. (A.1) through (A.5). In particular, Table A.1 summarizes the properties of the Fourier transform, while Table A.2 lists a set of Fourier-transform pairs.

In the time domain, a linear system (e.g., filter) is described in terms of its *impulse response*, defined as *the response of the system (with zero initial conditions) to a unit impulse or delta function* $\delta(t)$ *applied to the input of the system* at time $t = 0$. If the system is *time invariant*, then the shape of the impulse response is the same, no matter when the unit impulse is applied to the system. Thus, assuming that the unit impulse or delta function is applied at time $t = 0$, we may denote the impulse response of a linear time-invariant system by $h(t)$. Let this system be subjected to an arbitrary excitation $x(t)$, as in Fig. A.1(a). Then the response $y(t)$ of the system is determined by the formula

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \end{aligned} \tag{A.6}$$

TABLE A.1 Summary of Properties of the Fourier Transform.

Property	Mathematical Description
1. Linearity	$ag_1(t) + bg_2(t) \rightleftharpoons aG_1(f) + bG_2(f)$ where a and b are constants
2. Time scaling	$g(at) \rightleftharpoons \frac{1}{\lvert a\rvert}g\left(\frac{f}{a}\right)$ where a is a constant
3. Duality	If $g(t) \rightleftharpoons G(f)$, then $G(t) \rightleftharpoons g(-f)$
4. Time shifting	$g(t-t_0) \rightleftharpoons G(f)\exp(-j2\pi f t_0)$
5. Frequency shifting	$\exp(j2\pi f_c t)g(t) \rightleftharpoons G(f-f_c)$
6. Area under $g(t)$	$\int_{-\infty}^{\infty} g(t)dt = G(0)$
7. Area under $G(f)$	$g(0) = \int_{-\infty}^{\infty} G(f)df$
8. Differentiation in the time domain	$\frac{d}{dt}g(t) \rightleftharpoons j2\pi f G(f)$
9. Integration in the time domain	$\int_{-\infty}^{t} g(\tau)d\tau \rightleftharpoons \frac{1}{j2\pi f}G(f) + \frac{G(0)}{2}\delta(f)$
10. Conjugate functions	If $g(t) \rightleftharpoons G(f)$, then $g^*(t) \rightleftharpoons G^*(-f)$
11. Multiplication in the time domain	$g_1(t)g_2(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\lambda)G_2(f-\lambda)d\lambda$
12. Convolution in the time domain	$\int_{-\infty}^{\infty} g_1(\tau)g_2(t-\tau)d\tau \rightleftharpoons G_1(f)G_2(f)$
13. Correlation theorem	$\int_{-\infty}^{\infty} g_1(t)g_2^*(t-\tau)\,dt = G_1(f)G_2^*(f)$
14. Rayleigh's energy theorem	$\int_{-\infty}^{\infty} \lvert g(t)\rvert^2 dt = \int_{-\infty}^{\infty} \lvert G(f)\rvert^2 df$

The formula of Eq. (A.6) is called the *convolution integral.* Three different time scales are involved in it: the *excitation time* τ, *response time* t, *and system-memory time* $t-\tau$. Equation (A.6) is the basis of the time-domain analysis of linear time-invariant systems. It states that the present value of the response of a linear time-invariant system is the integral over the past history of the input signal, weighted according to the impulse response of the system. Thus, the impulse response acts as a *memory function* for the system.

TABLE A.2 Fourier-Transform Pairs.

Time Function	Fourier Transform
$\mathrm{rect}\left(\frac{t}{T}\right)$	$T\,\mathrm{sinc}(fT)$
$\mathrm{sinc}(2Wt)$	$\frac{1}{2W}\mathrm{rect}\left(\frac{f}{2W}\right)$
$\exp(-\pi t^2)$	$\exp(-\pi f^2)$
$\begin{cases} 1-\frac{\lvert t\rvert}{T}, & \lvert t\rvert < T \\ 0, & \lvert t\rvert \geq T \end{cases}$	$T\,\mathrm{sinc}^2(fT)$
$\delta(t)$	1
1	$\delta(f)$
$g(t-t_0)$	$\exp(j2\pi f t_0)$
$\exp(j2\pi f_c t)$	$\delta(f-f_c)$
$\cos(2\pi f_c t)$	$\frac{1}{2}[\delta(f-f_c)+\delta(f+f_c)]$
$\sin(2\pi f_c t)$	$\frac{1}{2j}[\delta(f-f_c)-\delta(f+f_c)]$
$\sum_{i=-\infty}^{\infty}\delta(t-iT_0)$	$\frac{1}{T_0}\sum_{n=-\infty}^{\infty}\delta\left(f-\frac{n}{T_0}\right)$
$\sum_{n=-\infty}^{\infty}x(t-nT_0)$	$\frac{1}{T_0}\sum_{m=-\infty}^{\infty}X\left(\frac{m}{T_0}\right)\delta\left(f-\frac{m}{T_0}\right)$

Notes: $\delta(t)$ = delta function, or unit impulse
$\mathrm{rect}(t)$ = rectangular function of unit amplitude and unit duration centered on the origin
$\mathrm{sinc}(t)$ = sinc function

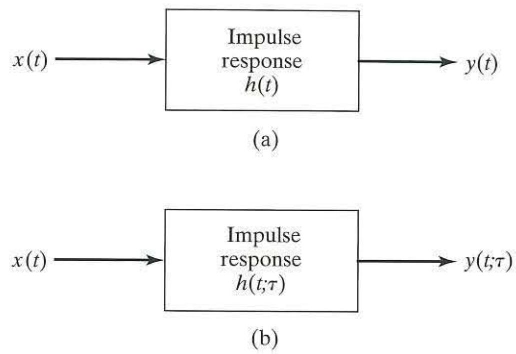


FIGURE A.1 (a) Linear system. (b) Linear time-varying system.

From Table A.1, we note that when two functions of time are convolved with each other, the operation of convolution is transformed into the multiplication of the Fourier transforms of the functions in the frequency domain. Hence, applying this property to Eq. (A.6), we may express the Fourier transform of the output signal $y(t)$ as

$$Y(f) = H(f)X(f) \tag{A.7}$$

where $X(f)$ is the Fourier transform of the input signal $x(t)$. The other quantity in Eq. (A.7), namely, $H(f)$, is called the *transfer function* of the system. It is formally defined as the Fourier transform of the impulse response $h(t)$ and is given by

$$H(f) = \int_{-\infty}^{\infty} h(t)\exp(-j2\pi ft)dt \tag{A.8}$$

Thus, the impulse response $h(t)$ provides a time-domain description of a linear time-invariant system, whereas the transfer function $H(f)$ provides an equivalent description of the system in the frequency domain.

A.2 LINEAR TIME-VARYING SYSTEMS

Consider next the case of a linear time-varying system, exemplified by a wireless communication channel. As the name implies, the impulse response of a *linear time-varying system* depends on the time at which the unit impulse is applied to the input of the system. We thus denote the impulse response of such a system by $h(t;\tau)$, where $(t - \tau)$ is the time at which the unit impulse is applied to the system and t is the time at which the resulting response is measured. (See Fig. A.1(b).) Suppose, then, that an input signal $x(t)$ is applied to a linear time-varying system with impulse response $h(t;\tau)$. Then the resulting response of the system is defined by

$$y(t) = \int_{-\infty}^{\infty} h(t;\tau)x(t-\tau)d\tau \tag{A.9}$$

where the integration is performed with respect to τ. Correspondingly, the transfer function of the system is written as $H(f;\tau)$, which is related to the impulse response $h(t;\tau)$ via the Fourier transform through the relationship

$$H(f;\tau) = \int_{-\infty}^{\infty} h(t;\tau)\exp(-j2\pi ft)\,dt \tag{A.10}$$

Equation (A.8) is a special case of Eq. (A.10) in that, for a linear time-invariant system, we have $H(f;\tau) = H(f)$ for all τ.

A.3 SAMPLING THEOREM

In continuous-wave modulation, the carrier is typically a sinusoidal wave. In pulse modulation, by contrast, the carrier is a uniform train of pulses that are relatively short compared with the fundamental period of the carrier. The sampling theorem, described next, is basic to all the different forms of pulse modulation used in practice.

To set the stage for a statement of the sampling theorem, consider a strictly band-limited signal $x(t)$ whose frequency content is confined to a bandwidth W; that is,

$$X(f) = 0 \qquad \text{for } |f| \geq W \tag{A.11}$$

For such a signal, the *sampling theorem* may be stated in two parts:

1. The strictly band-limited signal $x(t)$ is uniquely represented by a set of samples $x(nT_0)$, $n = 0, \pm 1, \pm 2, \ldots$, provided that the sampling rate $f_0 = 1/T_0$ is greater than twice the highest frequency component of $x(t)$; in other words, $f_0 > 2W$.
2. The original signal $x(t)$ is reconstructed from the set of samples $x(nT_0)$ for $n = 0, \pm 1, \pm 2, \ldots$, and $T_0 > 1/(2W)$, without loss of information, by passing this uniformly sampled signal through an ideal low-pass construction filter of bandwidth W hertz.

For a proof of the sampling theorem, we may invoke the duality property of the Fourier transform. From Table A.1, that duality property states,

> If $g(t) \rightleftharpoons G(f)$, then $G(-t) \rightleftharpoons g(f)$, where the time function $G(-t)$ is obtained by substituting $-t$ for f in the Fourier transform $G(f)$ and the frequency function $g(f)$ is obtained by substituting f for t in the inverse Fourier transform $g(t)$.

From the last entry of Table A.2, we also have the Fourier-transform pair

$$\sum_{m=-\infty}^{\infty} g(t - mT_0) \rightleftharpoons f_0 \sum_{n=-\infty}^{\infty} G(nf_0)\delta(f - nf_0) \tag{A.12}$$

where $f_0 = 1/T_0$ is the *sampling rate* and $\delta(f)$ is the Dirac delta function defined in the frequency domain. Applying the duality property to Eq. (A.12), invoking the even-function property of the delta function, and using T_0 in place of f_0 to maintain proper dimensionality in the result, we obtain

$$T_0 \sum_{n=-\infty}^{\infty} G(-nT_0)\delta(t-nT_0) \rightleftharpoons \sum_{m=-\infty}^{\infty} g(f-mf_0) \tag{A.13}$$

To put this relation in the context of the strictly band-limited signal $x(t)$, we set $G(-t) = x(t)$ and $g(f) = X(f)$, in which case we may recast Eq. (A.13) in the desired form:

$$x_\delta(t) = T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\delta(t-nT_0) \rightleftharpoons \sum_{m=-\infty}^{\infty} X(f-mf_0) = X_\delta(f) \tag{A.14}$$

Figure A.2 presents a time–frequency description of Eq. (A.14), assuming that $X(f) = 0$ for $|f| > W$ and $f_0 > 2W$. Parts (a) and (b) of the figure depict the spectra $X(f)$ and $X_\delta(f)$, respectively, where $x(t) \rightleftharpoons X(f)$ and $x_\delta(t) \rightleftharpoons X_\delta(f)$.

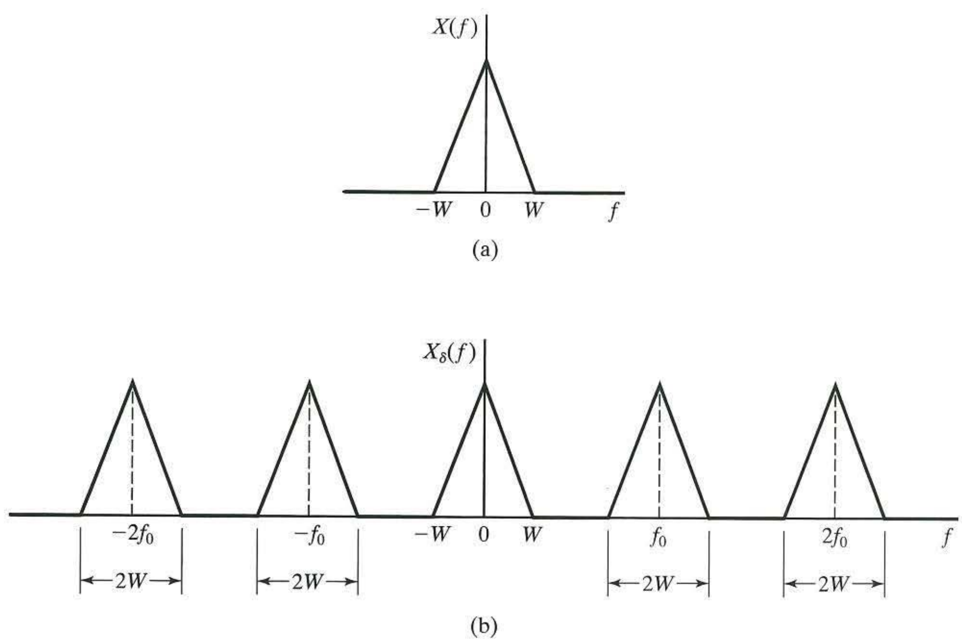


FIGURE A.2 (a) Spectrum of a signal $x(t)$ limited to the band $-W \le f < W$.
(b) Spectrum of the instantaneously sampled signal $x_\delta(t)$ for a sampling rate $f_0 > 2W$.

Given the *instantaneously sampled signal* $x_\delta(t)$ and assuming a sampling rate $f_0 > 2W$, how do we use $x_\delta(t)$ to reconstruct the original signal $x(t)$? We may do so by employing a *reconstruction system* formulated as an ideal low-pass filter of bandwidth W. To find the output of this filter in response to the sampled signal $x_\delta(t)$, we proceed in two stages:

1. Take the Fourier transform of sampled signal $x_\delta(t)$, and limit the spectrum to the frequency band $|f| \le W$ permitted by the low-pass reconstruction filter, thereby obtaining the spectrum

$$X(f) = \begin{cases} T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\exp(-j2\pi nT_0 f) & |f| \le W \quad |f| \le W \\ 0 & |f| > W \end{cases} \tag{A.15}$$

2. Take the inverse Fourier transform of the spectrum defined in Eq. (A.15), yielding the original signal

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} X(f)\exp(j2\pi ft)dt \\ &= \int_{-W}^{W}\left(T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\exp(-j2\pi nT_0 f)\right)\exp(j2\pi ft)dt \\ &= T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\left(\frac{\sin(2\pi(t-nT_0)W)}{2\pi(t-nT_0)W}\right) \\ &= T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\,\mathrm{sinc}(2(t-nT_0)W) \end{aligned} \tag{A.16}$$

where the function

$$\mathrm{sinc}(\lambda) = \frac{\sin(\pi\lambda)}{\pi\lambda} \tag{A.17}$$

is called the *sinc function*. Equation (A.16) states that, provided that the sampling rate f_0 satisfies the condition $f_0 > 2W$, the original signal $x(t)$ may be reconstructed as the weighted sum of the reconstruction kernel sinc($2Wt$), where the nth component of the sum consists of the time-shifted kernel sinc($2(t-nT_0)W$), weighted by the corresponding sample $x(nT_0)$.

Equation (A.16) verifies part (2) of the sampling theorem. Part (1) of the theorem is, in reality, merely a reformulation of part (2).

A.4 SAMPLED CONVOLUTION THEOREM

Suppose that we have two time functions, $x(t)$ and $h(t)$, with $x(t)$ limited to the frequency band $-W < f < W$. The function $x(t)$ is uniformly sampled at the rate $f_0 > 2W$ and then convolved with $h(t)$. The convolution product, denoted by $y(t)$, may be viewed as the output of a linear time-invariant system with impulse response $h(t)$, which is driven by the instantaneously sampled version of $x(t)$. The requirement is to evaluate $y(t)$.

The instantaneously sampled version of $x(t)$ is defined by (see the left-hand side of Eq. (A.14))

$$x_\delta(t) = T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\delta(t-nT_0) \tag{A.18}$$

where $T_0 = 1/f_0$ is the sampling period. The convolution of $x_\delta(t)$ with $h(t)$ is defined by the integral

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)x_\delta(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)\left(T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\delta(t-nT_0-\tau)\right)d\tau \\ &= T_0 \sum_{n=-\infty}^{\infty} x(nT_0)\int_{-\infty}^{\infty} h(\tau)\delta(t-nT_0-\tau)d\tau \end{aligned} \tag{A.19}$$

Invoking the replication property of the delta function described by Eq. (A.5), we may reduce the integral in the last line of Eq. (A.19) to

$$\int_{-\infty}^{\infty} h(\tau)\delta(t-nT_0-\tau)d\tau = h(t-nT_0) \tag{A.20}$$

Accordingly, Eq. (A.19) simplifies to

$$y(t) = T_0 \sum_{n=-\infty}^{\infty} x(nT_0)h(t-nT_0) \tag{A.21}$$

which is the desired result. Equation (A.21) is a statement of the *sampled convolution theorem*:

> *The convolution of a continuous-time function with the instantaneously sampled version of a band-limited signal is a scaled version of the convolution sum of two time series: the original instantaneously sampled signal and the instantaneously sampled version of the continuous-time function.*

Note that Eq. (A.21) is a generalization of the expression on the left-hand side of Eq. (A.14), with the impulse response $h(t)$ taking on the role of the delta function $\delta(t)$.

A.5 OUTPUT SAMPLING OF A LINEAR TIME-VARYING CHANNEL

In the case of a linear time-invariant communication channel, it is a straightforward matter to apply the sampled convolution theorem of Eq. (A.21) to the channel output in order to proceed with the use of digital signal processing in the receiver. When, however, the channel is linear, but time varying, as, for example, in wireless communications, we have to exercise care in the selection of a suitable sampling rate for the channel output, the reason being that the impulse response of the channel, $h(t;\tau)$, depends on a second time variable, namely, τ. The problem is now more complicated, in that we have a *two-dimensional temporal situation* to handle, with the two dimensions being defined by both t and τ. For frequency analysis of the channel output and, therefore, the determination of an appropriate sampling rate, we require the use of a two-dimensional Fourier transform. Such an analysis is beyond the scope of this book; the interested reader is referred to Note 2 for further pursual of the sampling rate required for linear time-varying channels. Suffice it to say, if W_I is the bandwidth (i.e., highest frequency component) of the input signal and W_C is the bandwidth of the channel time variations, then the sampling rate for the channel output must be larger than $(W_I + W_C)$. When the channel is time invariant, W_C is zero, and this result reduces to the standard form of the sampling theorem.

A.6 CORRELATION THEOREM

Thus far, we have discussed attention on the Fourier perspectives of two basic signal-processing operations: filtering (i.e., convolution) and sampling. Another signal-processing operation basic to the study of communication systems is correlation. To be specific, consider a pair of complex-valued signals $g_1(t)$ and $g_2(t)$, which may exhibit some degree of *similarity* in their time behaviors. The similarity is quantified by the integral

$$R_{12}(\tau) = \int_{-\infty}^{\infty} g_1(t) g_2^*(t-\tau)\,dt \qquad \text{(A.22)}$$

where the asterisk denotes complex conjugation. The function $R_{12}(\tau)$ is called the *cross-correlation function* between $g_1(t)$ and $g_2(t)$. The *time lag* τ is introduced into one of the two signals—$g_2(t)$ in the case under consideration here—in order to explore the similarity between them. To that end, τ is made variable. Intuitively, if, on the one hand, $g_1(t)$ and $g_2(t)$ are highly similar, then we expect $R_{12}(\tau)$ to peak around some value of τ. If, on the other hand, $g_1(t)$ and $g_2(t)$ are highly dissimilar, then $R_{12}(\tau)$ would be relatively flat over a broad range of values of τ.

With Fourier analysis as the subject of interest in this appendix, it is natural that we consider the Fourier transformation of $R_{12}(\tau)$. To pursue this transformation, we first use the inverse Fourier transform that defines $g(t)$ in terms of $G(f)$ and thus rewrite Eq. (A.22) in the form of a double integral (after rearranging terms):

$$R_{12}(\tau) = \int_{-\infty}^{\infty} G_1(f) \int_{-\infty}^{\infty} g_2^*(t-\tau) e^{j2\pi ft} dt df \tag{A.23}$$

Clearly, $R_{12}(\tau)$ is unchanged by introducing the product of the exponential $e^{j2\pi f\tau}$ and its complex conjugate $e^{-j2\pi f\tau}$ into the integral in Eq. (A.3), as is shown by

$$R_{12}(\tau) = \int_{-\infty}^{\infty} G_1(f) e^{j2\pi f\tau} \left[\int_{-\infty}^{\infty} g_2(t-\tau) e^{-j2\pi f(t-\tau)} d(t-\tau) \right]^* df \tag{A.24}$$

Now, the inner integral inside the square brackets in Eq. (A.24) is recognized as the Fourier transform of $g_2(t)$, which is denoted by $G_2(f)$. Accordingly, bearing in mind the complex conjugation around the square brackets, we may finally simplify Eq. (A.24) as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} G_1(f) G_2^*(f) e^{j2\pi f\tau} df \tag{A.25}$$

from which we immediately infer that $G_1(f)G_2^*(f)$ is the Fourier transform of $R_{12}(\tau)$.

In words, the *correlation theorem* may be stated as follows:

> *Given a pair of Fourier-transformable signals $g_1(t)$ and $g_2(t)$ whose cross-correlation function $R_{12}(\tau)$ is defined by Eq. (A.22), the Fourier transform of $R_{12}(\tau)$ is defined by the product $G_1(f)G_2^*(f)$, where $G_1(f)$ and $G_2(f)$ are the Fourier transforms of $g_1(t)$ and $g_2(t)$, respectively.*

In applying the correlation theorem, careful attention has to be paid to the order and manner in which the functions $g_1(t)$ and $g_2(t)$ appear in Eq. (A.22) and the corresponding order of subscripts in $R_{12}(\tau)$.

Moreover, there are some similarities and basic differences between the cross-correlation and convolution integrals that should be noted:

1. In the convolution integral of Eq. (A.6), the integration is with respect to the lag variable t. By contrast, the integration in the cross-correlation integral of Eq. (A.22) is with respect to the time variable τ.
2. When both integrals are transformed into the frequency domain, the result of each transformation is expressed as a product of two Fourier transforms—but with a difference. In the case of the convolution integral, the product is simply equal to the Fourier transforms of the two signals, with the result that *convolution* is commutative. In the case of the cross-correlation integral, the Fourier transform of the particular signal delayed in the correlation process is complex

conjugated. Consequently, unlike the convolution of two integrals, the cross-correlation is *not* commutative; that is,

$$R_{12}(\tau) \rightleftharpoons G_1(f)G_2^*(f) \tag{A.26}$$

which, in general, is different from

$$R_{21}(\tau) \rightleftharpoons G_2(f)G_1^*(f) \tag{A.27}$$

A.6.1 Autocorrelation Function

When $g_1(t) = g_2(t) = g(t)$, we have the *autocorrelation function* of the signal $g(t)$, defined by

$$R_g(\tau) = \int_{-\infty}^{\infty} g(t)g^*(t-\tau)dt \tag{A.28}$$

Correspondingly, Eq. (A.26) reduces to

$$R_g(\tau) \rightleftharpoons |G(f)|^2 \tag{A.29}$$

Note that the autocorrelation function $R_g(\tau)$ is an even function of the lag τ, as is shown by

$$R_g(-\tau) = R_g(\tau) \text{ for all } \tau \tag{A.30}$$

Expanding the pair of relations summarized under Eq. (A.29), we have

$$S_g(f) = \int_{-\infty}^{\infty} R_g(\tau)e^{-j2\pi f\tau}d\tau \tag{A.31}$$

and

$$R_g(\tau) = \int_{-\infty}^{\infty} S_g(f)e^{j2\pi f\tau}df \tag{A.32}$$

where we have introduced the definition

$$S_g(f) = |G(f)|^2 \text{ for all } f \tag{A.33}$$

The new function $S_g(f)$ is called the *energy density spectrum* of the signal $g(t)$. The pair of equations (A.31) and (A.32) constitute the *Wiener–Khintchine relations* for signals with finite energy.

A.7 PARSEVAL'S RELATIONSHIPS

The energy of a complex-valued signal g(t) is defined by

$$E_g = \int_{-\infty}^{\infty} |g(t)|^2 dt \tag{A.34}$$

Putting $\tau = 0$ in Eq. (A.28) and using the definition of Eq. (A.34), we readily see that

$$E_g = R_g(0) \tag{A.35}$$

which states that the value of the autocorrelation function $R_g(\tau)$ at the origin $\tau = 0$ is equal to the energy of the signal $g(t)$.

Putting $\tau = 0$ in Eq. (A.32) and using Eqs. (A.34) and (A.35), we obtain *Parseval's energy theorem*, which states that

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df \tag{A.36}$$

In words, Parseval's energy theorem asserts that the energy of a nonperiodic signal $g(t)$ is equal to the total area under the curve of the energy density spectrum $S_g(f)$.

To deal with a periodic signal $g(t)$ of fundamental period T, we use *Parseval's power theorem*, which states that

$$\frac{1}{T}\int_0^T |g(t)|^2 dt = \sum_{k=-\infty}^{\infty} |G_k|^2 \tag{A.37}$$

where G_k are the *complex Fourier coefficients*, in terms of which the periodic signal

$$g(t) = \sum_{k=-\infty}^{\infty} G_k e^{j2\pi t/T} \tag{A.38}$$

is defined.

To deal with a periodic signal, we may use *Parseval's power theorem* to calculate the average power of the signal. To formulate this theorem, recall that a complex-valued periodic signal $g(t)$ with fundamental period T may be expanded into the *Fourier series*

$$g(t) = \sum_{k=-\infty}^{\infty} G_k e^{j2\pi k f_0 t} \tag{A.39}$$

where

$$G_k = \frac{1}{T}\int_0^T g(t) e^{-j2\pi k f_0 t} \qquad k = 0, \pm 1, \pm 2, \ldots \tag{A.40}$$

are the *complex Fourier coefficients*. The fundamental frequency of the signal is itself defined by

$$f_0 = \frac{1}{T} \tag{A.41}$$

By definition, for the average power of the periodic signal $g(t)$, we have

$$P = \frac{1}{T}\int_0^T |g(t)|^2 dt \tag{A.42}$$

Accordingly, Parseval's power theorem states that we may also evaluate P by using the formula

$$P = \sum_{k=-\infty}^{\infty} |G_k|^2 \tag{A.43}$$

where the G_k are themselves defined by Eq. (A.40).

Notes and References

[1] For an authoritative treatment of the many facets of the Fourier transform and its applications, see Bracewell (1986).

[2] For a careful discussion of the sampling rate required for linear time-varying systems, see Kailath (1959) and Médard (1995).

APPENDIX B

Bessel Functions

B.1 BESSEL FUNCTIONS OF THE FIRST KIND

Bessel functions of the first kind of integer order ν are defined as the solution of the integral equation

$$\begin{aligned} J_\nu(z) &= \frac{1}{\pi}\int_0^\pi \cos(z\sin\theta - \nu\theta)d\theta \\ &= \frac{j^{-\nu}}{\pi}\int_0^\pi e^{jz\cos\theta}\cos(\nu\theta)d\theta \end{aligned} \tag{B.1}$$

where j is the square root of -1. The special case $\nu = 0$ reduces to

$$J_0(z) = \frac{1}{\pi}\int_0^\pi \cos(z\sin\theta)d\theta \tag{B.2}$$

For a real argument z, the Bessel functions are real valued, continuously differentiable, and bounded in magnitude by unity. The even-numbered Bessel functions are symmetric and the odd-numbered Bessel functions are antisymmetric.

The Bessel function $J_\nu(z)$ may also be expressed as the infinite series

$$J_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{4}z^2\right)^k}{k!\Gamma(\nu+k+1)} \tag{B.3}$$

where $\Gamma(k)$ is the *gamma function*; for integer values, $\Gamma(k+1) = k!$.

We plot $J_0(z)$ and $J_1(z)$ for real-valued z in Fig. B.1. The values of these functions for a subset of z are given in Table B.1.

Problem B.1 Using the first line of Eq. (B.1), derive the second line of the equation. ■

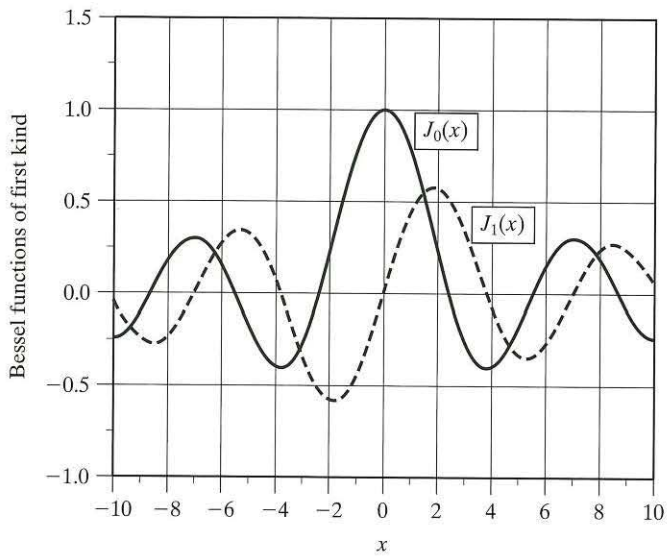


FIGURE B.1 Plots of Bessel functions of the first kind, $J_0(x)$ and $J_1(x)$.

TABLE B.1 Values of Bessel Functions and Modified Bessel Functions of the First Kind.

x	$J_0(x)$	$J_1(x)$	$I_0(x)$	$I_1(x)$
0.00	1.0000	0.0000	1.0000	0.0000
0.20	0.9900	0.0995	1.0100	0.1005
0.40	0.9604	0.1960	1.0404	0.2040
0.60	0.9120	0.2867	1.0920	0.3137
0.80	0.8463	0.3688	1.1665	0.4329
1.00	0.7652	0.4401	1.2661	0.5652
1.20	0.6711	0.4983	1.3937	0.7147
1.40	0.5669	0.5419	1.5534	0.8861
1.60	0.4554	0.5699	1.7500	1.0848
1.80	0.3400	0.5815	1.9896	1.3172
2.00	0.2239	0.5767	1.1796	1.5906
2.20	0.1104	0.5560	2.6291	1.9141
2.40	0.0025	0.5202	3.0493	2.2981
2.60	−0.0968	0.4708	3.5533	2.7554
2.80	−0.1850	0.4097	4.1573	3.3011
3.00	−0.2601	0.3391	4.8808	3.9534
3.20	−0.3202	0.2613	5.7472	4.7343
3.40	−0.3643	0.1792	6.7848	5.6701
3.60	−0.3918	0.0955	8.0277	6.7927
3.80	−0.4026	0.0128	9.5169	8.1404
4.00	−0.3971	−0.0660	11.3019	9.7595

B.2 MODIFIED BESSEL FUNCTIONS OF THE FIRST KIND

Modified Bessel functions of the first kind of integer order v are defined as the solution of the integral equation

$$I_v(z) = \frac{1}{\pi}\int_0^{\pi} e^{z\cos\theta}\cos(v\theta)d\theta \tag{B.4}$$

For the special case $v = 0$, Eq. (B.4) reduces to

$$I_0(z) = \frac{1}{\pi}\int_0^{\pi} e^{z\cos\theta}d\theta \tag{B.5}$$

For a real argument z, the modified Bessel functions are real valued, continuously differentiable, and grow exponentially as $|z|$ increases. The even-numbered modified Bessel functions are symmetric and the odd-numbered ones are antisymmetric.

The modified Bessel function may also be expressed as the infinite series

$$I_v(z) = \left(\frac{1}{2}z\right)^v \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(v+k+1)} \tag{B.6}$$

We plot $I_0(z)$ and $I_1(z)$ for real-valued z in Fig. B.2. The values of these functions for a subset of z are given in Table B.1.

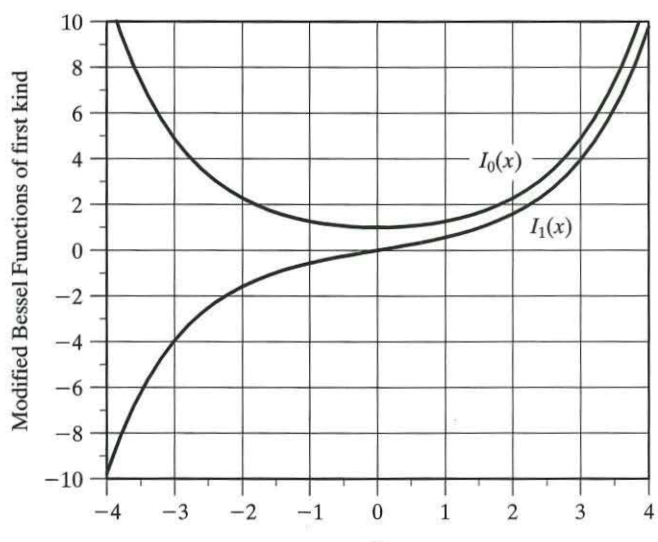


FIGURE B.2 Plots of modified Bessel functions of the first kind, $I_0(z)$ and $I_1(z)$.

APPENDIX C

Random Variables and Random Processes

C.1 SETS, EVENTS, AND PROBABILITY

Probability theory is centered on fundamental principles relating sets. In this appendix, we will consider an abstract space Ω that has elements ω. The space Ω may consist of a finite number of elements, may be countably infinite, such as the set of integers, or may be uncountable, such as the set of real numbers. We let Ξ represent all the possible subsets of Ω, including the empty set, $\varnothing$, and the complete set Ω.

Probability is a measure on any set Q in Ξ. Conceptually, if Q represents a set of elements, or *an event*, then $\text{Prob}(Q)$ is the probability of that event. Empirically, if we make N observations of this space and determine how many times n out of N trials that the observations belong to Q, then the empirical definition of the probability of the event Q is

$$\text{Prob}(Q) = \lim_{N \to \infty} (n/N) \tag{C.1}$$

This is intuitively what we think of as probability: what fraction of the time a certain event occurs.

A probability measure must satisfy three properties:

$$\begin{aligned} &\text{Prob}(\Omega) = 1; \\ &\text{Prob}(\varnothing) = 0; \\ &\text{Prob}(A \cup B) \le \text{Prob}(A) + \text{Prob}(B) \quad \text{for any } A, B \text{ in } \Xi. \end{aligned} \tag{C.2}$$

In calculations, we are often interested in the *conditional probability* that an event A occurs, given that an event B has occurred. This is defined as

$$\text{Prob}[A|B] = \frac{\text{Prob}[A \cap B]}{\text{Prob}[B]} \tag{C.3}$$

The conditional probability is a probability measure in its own right and satisfies all of the properties of Eq. (C.2).

Bayes' theorem allows us to convert between conditioning on one event to conditioning on a different event and is given by

$$\text{Prob}(A_i|B) = \frac{\text{Prob}(B|A_i)\text{Prob}(A_i)}{\sum_{j=1}^{N} \text{Prob}(B|A_j)\text{Prob}(A_j)} \tag{C.4}$$

Bayes' theorem is often used in inferential analysis, as the expressions for conditional probability based on some events are often much simpler than those based on other events.

C.2 RANDOM VARIABLES

A *random variable* is a mapping from the abstract space Ω to the real numbers, represented as $X:\Omega \rightarrow \Re$, where $\Re$ is the set of real values. Conceptually, we usually consider X as the physical realization of some unknowable process. For example, X could be the voltage measured across a resistor due to thermal noise. In that case, Ω could be the state of all the electrons in the resistor.

A *discrete random variable* can take on only a discrete set of values. Sometimes these are denoted as $\{x_i\}$, where i indexes the possible values of X. For example, the number of paths in a multipath signal is a discrete random variable; it may take only the values 1,2,3,.... A *continuous random variable* can take on a continuum of values. Often, this continuum is the set of real values or the set of nonnegative reals. For example, thermal noise voltage observed at a specific instant of time is a continuous random variable.

To characterize the probabilistic behavior of random variables, we simply extend the set concepts used in the previous section. In the thermal-noise example, we may be interested in determining the probability that $X \leq x$ for some value x. We would write this as $\text{Prob}(X \leq x)$, but its mathematical meaning is

$$\text{Prob}(X \leq x) = \text{Prob}(\{\omega \in \Omega : X(\omega) \leq x\}) \tag{C.5}$$

That is, it is a measure on the set of those ω's such that the random variable maps the $X(\omega)$ to a value less than x. Probability is a measure on the underlying abstract set Ω. The physical realization is usually more easily understood than the abstract, but understanding the underlying concepts is often useful for resolving some probability issues.

C.3 PROBABILITY DISTRIBUTIONS AND DENSITIES

The probability that a random variable X is less than a given x is written as

$$F_X(x) = \text{Prob}(X \leq x) \tag{C.6}$$

which is called the *cumulative distribution function* of the random variable X. This function is right continuous and increases monotonically, with $F(-\infty) = 0$ and $F(\infty) = 1$.

A discrete random variable will have a discrete distribution function that consists of steps at the finite or countable number of points where $\text{Prob}(X = x_i) > 0$. A continuous random variable will have a continuous distribution function. If the distribution function is a continuously differentiable function of x, then we define the *probability density function* as

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{C.7}$$

Probability density functions play an important role in defining the conditional probabilities of continuous random variables. Consider the two joint events $X \le x$ and $y \le Y \le y + \delta y$. We may use Bayes' rule to express the conditional probability of the first event, given the second, as

$$\begin{aligned} F_{X|y \le Y \le y + \delta y}(x) &= \frac{\text{Prob}(X \le x, y \le Y \le y + \delta y)}{\text{Prob}(y \le Y \le y + \delta y)} \\ &= \frac{F_{XY}(x, y + \delta y) - F_{XY}(x, y)}{F_Y(y + \delta y) - F_Y(y)} \\ &= \frac{\int_{-\infty}^{x} \int_{y}^{y + \delta y} f_{X,Y}(u, v)\,du\,dv}{\int_{y}^{y + \delta y} f_Y(v)\,dv} \end{aligned} \tag{C.8}$$

Differentiating Eq. (C.8) with respect to x by using Leibniz's rule, we may write

$$\begin{aligned} f_{X|y \le Y \le y + \delta y}(x) &= \frac{\int_{y}^{y + \delta y} f_{X,Y}(u, v)\,dv}{\int_{y}^{y + \delta y} f_Y(v)\,dv} \\ &\approx \frac{f_{X,Y}(x, y)\,\delta y}{f_Y(y)\,\delta y} \end{aligned}$$

Finally, in the limit, as δy approaches zero, assuming that $f_Y(y) \neq 0$, we have

$$f_{X|Y = y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{C.9}$$

It is important to note that Eq. (C.9) describes a probability density function of x for a fixed y.

C.4 EXPECTATION OF RANDOM VARIABLES

The *expected value,* or *mean value,* of a random variable X is written as $\mathbf{E}[X]$, where $\mathbf{E}$ is the *statistical expectation operator*. For a discrete random variable, the expected value of X is given by

$$\mathbf{E}[X] = \sum_{i=1}^{N} x_i \mathrm{Prob}(X = x_i) \tag{C.10}$$

For a continuous random variable that has a probability density function, the expected value of X is given by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx \tag{C.11}$$

If X_1 and X_2 are any two random variables, then

$$\mathbf{E}[X_1 + X_2] = \mathbf{E}[X_1] + \mathbf{E}[X_2] \tag{C.12}$$

and

$$\mathbf{E}[\alpha X_1] = \alpha \mathbf{E}[X_1] \tag{C.13}$$

where α is a constant. That is, expectation is a *linear* operator. In general, if $g(X)$ is any well-defined function of X, then the expected value of $g(X)$ is given by

$$\mathbf{E}[g(X)] = \sum_{i=1}^{N} g(x_i) P(X = x_i) \quad \text{or} \quad \mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x_i) f_X(x) dx \tag{C.14}$$

depending upon whether the random variable is discrete or continuous, respectively. Other common statistical parameters of interest are the *second moment* or *mean-square value*

$$\mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx \tag{C.15}$$

and the *variance*

$$\mathrm{Var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 f_X(x) dx \tag{C.16}$$

An analogous result holds for the discrete case.

C.5 COMMON PROBABILITY DISTRIBUTIONS AND THEIR PROPERTIES

Binomial distribution. Consider a discrete random variable X that can take the values 0 and 1 with probabilities $(1 - p)$ and p, respectively. Suppose N independent observations of this random variable are made and labeled X_i for $1 \le i \le N$. Define the new random variable

$$Y = \sum_{i=1}^{N} X_i \tag{C.17}$$

Then Y is said to have a *binomial distribution* with parameter p; that is,

$$F_Y(y) = \sum_{j=0}^{y} \binom{N}{j} p^{N-y}(1-p)^y \qquad \text{for } 0 \le y \le N \tag{C.18}$$

The expected value and variance of Y are Np and $Np(1-p)$, respectively.

Gaussian distribution. A common continuous random variable is the Gaussian random variable. The density function of a Gaussian random variable is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{C.19}$$

where the mean of the Gaussian random variable is μ and its variance is σ^2. The distribution function of a Gaussian random variable does not have a closed-form solution, but it is usually expressed in terms of the error function as

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(s)ds \\ &= \begin{cases} \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right) & x \ge 0 \\ 1 - F_X(-x) & x < 0 \end{cases} \end{aligned} \tag{C.20}$$

where the *error function* is given by (see Appendix E)

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}dz \tag{C.21}$$

A linear transformation of a Gaussian random variable is also a Gaussian random variable. That is, if $X_1, X_2, ..., X_N$ are Gaussian random variables, then the composite random variable

$$Y = \sum_{i=1}^{N} b_i X_i \tag{C.22}$$

is also a Gaussian random variable. The mean of Y is given by

$$\mathbf{E}[Y] = \sum_{i=1}^{N} b_i E[X_i] \tag{C.23}$$

If the $\{X_i\}$ are independent Gaussian random variables, then the variance of Y is given by

$$\operatorname{Var}(Y) = \sum_{i=1}^{N} b_i^2 \operatorname{Var}(X_i) \tag{C.24}$$

Rayleigh distribution. If X_1 and X_2 are zero-mean Gaussian random variables with a common variance σ^2, then the random variable

$$R = \sqrt{X_1^2 + X_2^2} \tag{C.25}$$

has a *Raleigh distribution* given by (see Section C.6)

$$\text{Prob}(R < r) = 1 - e^{-r^2/2\sigma^2} \qquad r \geq 0 \tag{C.26}$$

The corresponding probability density function is

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \qquad r \geq 0 \tag{C.27}$$

The first and second moments of Y are given by

$$\mathbf{E}[R] = \sqrt{\frac{\pi}{2}}\sigma \quad \text{and} \quad \mathbf{E}[R^2] = 2\sigma^2 \tag{C.28}$$

and the variance of Y is given by $(2 - \pi/2)\sigma^2$.

Rician distribution. If X_1 and X_2 are Gaussian random variables with means μ_1 and μ_2, respectively, and a common variance σ, then the new random variable

$$R = \sqrt{X_1^2 + X_2^2} \tag{C.29}$$

has a *Rician distribution*. The probability density function of R is given by

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2 + s^2)/2\sigma^2} I_0\left(\frac{rs}{\sigma^2}\right) \qquad r \geq 0 \tag{C.30}$$

where $I_0(\cdot)$ is the *modified Bessel function of order zero* (see Appendix B) and $s = \sqrt{\mu_1^2 + \mu_2^2}$. There is no known closed-form solution for the distribution function of a Rician random variable.

Chi-square distribution. If $\{X_i\}$, $i = 1, ..., N$ are zero-mean Gaussian random variables with a common variance σ^2, then the random variable

$$Y = \sum_{i=1}^{N} X_i^2 \tag{C.31}$$

is said to have a *Chi-square distribution with N degrees of freedom*. The first two moments of Y are

$$\mathbf{E}[Y] = N\sigma^2 \tag{C.32}$$

and

$$\mathbf{E}[Y^2] = 2N\sigma^4 + N^2\sigma^4 \tag{C.33}$$

When $m = N/2$ is an integer, the probability density function of Y is given by

$$f_Y(y) = \frac{1}{(2\sigma^2)^m (m-1)!} y^{m-1}$$

and the cumulative distribution function of Y is given by

$$F_Y(y) = 1 - e^{-y/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!}\left(\frac{y}{2\sigma^2}\right)^k$$

This is one of the most common distribution functions in communications systems applications. For the case of odd N, there is no closed-form solution for $F_Y(y)$.

C.6 TRANSFORMATIONS OF RANDOM VARIABLES

Let X_1 and X_2 be continuous random variables with a joint probability density function $f_{X_1, X_2}(x_1, x_2)$, and consider the transformation defined by

$$y_1 = h_1(x_1, x_2), \text{ and } y_2 = h_2(x_1, x_2), \tag{C.34}$$

which are assumed to be one-to-one and continuously differentiable.The *Jacobian of this transformation* is defined by the matrix determinant

$$\mathbf{J}\left(\frac{y_1, y_2}{x_1, x_2}\right) = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} \\ \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} \neq 0 \tag{C.35}$$

The joint probability density function of Y_1 and Y_2 is given by

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2)\left|\mathbf{J}\left(\frac{y_1, y_2}{x_1, x_2}\right)\right| \tag{C.36}$$

If the transformations are not one-to-one, then other approaches must be taken. For example, consider the transformation $Y = \sqrt{X_1^2 + X_2^2}$. The cumulative distribution function of Y is given by

$$F_Y(y) = \iint_A f_{X_1, X_2}(x_1, x_2)\,dx_1\,dx_2 \tag{C.37}$$

where A is the set of all (x_1, x_2) such that $\sqrt{x_1^2 + x_2^2} \le y$. If X_1 and X_2 are independent, zero-mean Gaussian random variables with a variance of unity, then

$$F_Y(y) = \iint_A \frac{1}{\sqrt{2\pi}} e^{-(x_1^2 + x_2^2)/2}\,dx_1\,dx_2 \tag{C.38}$$

If we make the transformations $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$, then Eq. (C.38) becomes

$$\begin{aligned} F_Y(y) &= \frac{1}{2\pi}\int_0^y\int_0^{2\pi} e^{-r^2/2}\, r\,dr\,d\theta \\ &= 1 - e^{-y^2/2} \end{aligned} \tag{C.39}$$

This is the Rayleigh distribution described in Section C.5.

C.7 CENTRAL-LIMIT THEOREM

Consider a sequence $\{X_n\}$ of independent and identically distributed (i.i.d.) random variables with means $\mathbf{E}[X_i] = m$ and variances $\mathbf{E}[(X_i - m)^2] = \sigma^2$. Let Y_n be a new sequence of random variables defined by the partial sums

$$Y_n = \sum_{i=1}^{n} X_i \tag{C.40}$$

Let μ_n be the mean of Y_n and S_n be the variance of Y_n. Define the normalized random variable

$$Z_n = \frac{Y_n - \mu_n}{S_n} = \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{n}\sigma} \tag{C.41}$$

Then *the random variable Z_n has a distribution that is asymptotically unit normal.* That is, as n becomes large, the distribution of Z_n approaches that of a zero-mean Gaussian random variable with unit variance. This result is referred to as the *central-limit theorem*. The theorem also holds if the variables X_i are not identically distributed, but there are some restrictions.

C.8 RANDOM PROCESSES

A random process is a mapping $X{:}[\Omega, T] \to \Re$, where T represents a time interval such that $X(.,t)$ is a random variable for each fixed time t. To distinguish between a random variable X and a random process X, we usually write the latter as $X(t)$. If X is a discrete random variable for all t, then we say that $X(t)$ is a *discrete random process*. If X is a continuous random variable for all t, then we say that $X(t)$ is a *continuous random process*.

For each fixed value of t, we speak of the distribution function

$$F_{X(t)}(x) = \text{Prob}(X(t) < x) \tag{C.42}$$

We also speak of the joint distribution function

$$F_{X(t_1)X(t_2)}(x_1, x_2) = \text{Prob}(X(t_1) < x_1, X(t_2) < x_2) \tag{C.43}$$

For a fixed ω in Ω, the time function $X(\omega, t)$ is known as a *sample function* or *realization* of the random process.

C.9 PROPERTIES OF RANDOM PROCESSES

Let $X(t)$ be a complex random process, and define the *autocorrelation function* of that process as

$$R_X(t, s) = \mathbf{E}[X(t)X^*(s)] \tag{C.44}$$

where the asterisk denotes complex conjugation. That is, the autocorrelation function is the expectation of the product of two random variables that are parameterized by t and s. Recognizing that $R_X(0) = \mathbf{E}[|X(t)|^2]$, we see that the autocorrelation is a generalization of the second moment of a random variable. The autocorrelation is a deterministic function.

A process whose joint distribution is invariant with time translation, that is,

$$\begin{aligned}&\text{Prob}(X(t_1) < x_1, X(T_2) < x_2, \ldots, X(t_n) < x(n)) = \\ &\text{Prob}(X(t_1 + h) < x_1, X(t_2 + h) < x_2, \ldots, X((t_n + h) < x_n))\end{aligned} \tag{C.45}$$

is said to be *stationary to order n* if Eq. (C.46) holds for all h and for a particular n. Many of the random processes dealt with in wireless communications are assumed to be stationary. A process is said to be *wide-sense stationary* if

$$E[X(t)] = \text{constant for all } t \text{ and } R_X(t, s) = R_X(t - s) \tag{C.46}$$

Random processes whose joint distribution functions are multivariate Gaussian are referred to as *Gaussian random processes*. If a Gaussian random process is wide-sense stationary, then it is also stationary.

C.10 SPECTRA OF RANDOM PROCESSES

In Appendix A, we defined the spectrum of a finite-energy signal $x(t)$ as the Fourier transform of that signal. However, in considering a random process $X(t)$, we have an ensemble of sample functions. To get around this difficulty, we note that the autocorrelation function of a stationary random process, namely, $R_X(\tau)$, satisfies the conditions of Fourier transformability. Accordingly, we may define the *power spectrum* or *power spectral density* of a random process $X(t)$ as the Fourier transform of its autocorrelation function $R_X(\tau)$. Denoting this new parameter by $S_X(f)$, we may thus formally write

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}d\tau \text{ and } R_X(z) = \int_{-\infty}^{\infty} S_X(f)e^{j2\pi f\tau}df \tag{C.47}$$

Note that $S_X(f)$ is measured in watt/Hz. Stated another way, the total area under the curve of $S_X(f)$, plotted as a function of frequency f, defines the average power of the process.

The autocorrelation function $R_X(\tau)$ and power spectral density $S_X(f)$ form a Fourier-transform pair, which means that the autocorrelation function $R_X(\tau)$ is the inverse Fourier transform of the power spectral density $S_X(f)$. The Fourier transform pair, linking $S_X(f)$ to $R_X(\tau)$ and vice versa, is called the *Wiener–Khintchine relation* for random processes.

An idealized example of a random process is a special form of noise commonly referred to as *white noise* $W(t)$. White noise has the property that it is uncorrelated for all nonzero time offsets. Consequently, the autocorrelation function of such a process is defined by a delta function, or

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau) \tag{C.48}$$

where $N_0/2$ is the two-sided noise density in watt/Hz. The noise is referred to as "white" because the corresponding power spectrum is flat; that is,

$$S_W(f) = \frac{N_0}{2} \quad \text{for all } f \tag{C.49}$$

In other words, white noise contains all frequency components at equal strength, analogous to white light in the visible part of the spectrum. This relationship does not make any assumptions about the distribution of the random variable $n(t)$ at time t; it could be Gaussian or otherwise.

Another example of a random process is a *random binary wave*, defined by

$$x(t) = b_n \qquad t_0 + nT < t < t_0 + (n+1)T \tag{C.50}$$

where t_0 is a random starting time between $[0,T]$ and $\{b_n\}$ is a sequence of independent, zero-mean random variables with values ± 1. The autocorrelation function of the process described in Eq. (C.51) is

$$\begin{aligned} R_X(t, t+\tau) &= \mathbf{E}[x(t)x(t+\tau)] \\ &= \mathbf{E}_{t_0}\mathbf{E}_b[x(t)x(t+\tau)] \\ &= \mathbf{E}_{t_0}\begin{cases} 1 & t_0 + nT < t < t+\tau < t_0 + (n+1)T \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{C.51}$$

where the expectation $\mathbf{E}$ has been split over the two random independent variables t_0 and $\{b_n\}$ and we have used the fact that $\mathbf{E}[b_n b_m] = 0$ if $n \neq m$. If we evaluate the expectation of Eq. (C.51) over t_0, we obtain

$$R_X(t, t+\tau) = \begin{cases} T-\tau & T > \tau > 0 \\ T+\tau & -T < \tau \le 0 \\ 0 & \text{otherwise} \end{cases} \tag{C.52}$$

$$= \begin{cases} T-|\tau| & |\tau| < T \\ 0 & \text{otherwise} \end{cases}$$

Thus, $R_X(\tau)$ is stationary with a triangular autocorrelation function. The spectrum of the binary random wave is the Fourier transform of this autocorrelation function:

$$\begin{aligned} S_X(f) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \\ &= T^2 \frac{\sin^2(\pi f T)}{(\pi f T)^2} \\ &= T^2 \operatorname{sinc}^2(fT) \end{aligned} \tag{C.53}$$

C.11 LINEAR FILTERING OF RANDOM PROCESSES

In communications, we filter signals for various reasons. If the signal is a random process, how do we characterize the output? Let $y(t)$ be the output resulting from applying a linear time-invariant causal filter $h(t)$ to a realization of an input random process, namely, $x(t)$, as represented by the convolution integral

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \tag{C.54}$$

This means that the filter is applied to the particular realization $x(t) = X(\omega,t)$ of the random process, and that realization is referred as a sample path integral. A sufficient condition for $Y(t)$, the random variable composed of $\{y(\omega,t)\}$, to be a well-defined random variable for all t is that

$$\int_{-\infty}^{\infty} |h(\tau)| \mathbf{E}[|X(t-\tau)|] d\tau < \infty \tag{C.55}$$

If $Y(t)$ is well defined, then we may determine the expectation of $Y(t)$ using

$$\mathbf{E}[Y(t)] = \mathbf{E}\left[\int_{-\infty}^{\infty} h(t-\tau)X(\tau)d\tau\right] = \int_{-\infty}^{\infty} h(t-\tau)\mathbf{E}[X(\tau)]d\tau \tag{C.56}$$

and similarly for other moments of $Y(\tau)$. The interchange of the order of integration and expectation is allowed because both of these operations are linear.

If $X(t)$ is a stationary process with autocorrelation $R_X(\tau)$ and corresponding spectral density $S_X(f)$, then the spectral density of the output $Y(t)$ is given by

$$S_Y(f) = |H(f)|^2 S_X(f) \tag{C.57}$$

That is, for stationary random processes, the output power spectral density of a linear continuous-time filter is equivalent to the product of two quantities: the squared magnitude response of the filter and the input power spectral density. In addition, if the input is wide-sense stationary and the linear system is time invariant, then the output will be stationary as well. The spectral relationship of Eq. (C.57) for a stationary random process $X(t)$ may be viewed as the counterpart of Eq. (A.7) for a signal $x(t)$ with finite energy.

Analogous to the result for linear transformations of random variables, we have the result that if a linear filter has an input which is a Gaussian random process, then the output will also be a Gaussian random process.

C.12 COMPLEX RANDOM VARIABLES AND PROCESSES

In certain situations, we have to deal with the statistical characterization of complex random variables and complex random processes. (A case in point is that of the complex baseband representation of a narrowband process considered in the next section.) When we refer to a *complex random variable* $Z = X + jY$, we mean that X and Y are (real) random variables and they are described by their joint distribution function.

Similarly, if $Z(t) = X(t) + jY(t)$ is a *complex random process*, then $X(t)$ and $Y(t)$ are random processes that are characterized by their joint distributions at each time t. For example, the autocorrelation of a stationary $Z(t)$ is given by

$$\begin{aligned} R_Z(\tau) &= \mathbf{E}[Z(t)Z^*(t-\tau)] \\ &= \mathbf{E}[(X(t)+jY(t))(X(t-\tau)-jY(t-\tau))] \\ &= \mathbf{E}[X(t)X(t-\tau)] + \mathbf{E}[Y(t)Y(t-\tau)] + j(\mathbf{E}[Y(t)X(t-\tau)] - \mathbf{E}[X(t)Y(t-\tau)]) \\ &= R_X(\tau) + R_Y(\tau) + j(R_{YX}(\tau) - R_{XY}(\tau)) \end{aligned} \tag{C.58}$$

where

$$R_{XY}(\tau) = \mathbf{E}[X(t)Y(t-\tau)] \tag{C.59}$$

C.13 COMPLEX REPRESENTATION OF NARROWBAND RANDOM PROCESSES

Let $X(t)$ be a narrowband random process centered on some frequency f_c. In a manner similar to that described in Chapter 3, we may introduce a complex baseband process $\tilde{X}(t)$ by writing

$$X(t) = \mathrm{Re}[\tilde{X}(t)\exp(j2\pi f_c t)] \tag{C.60}$$

where Re[·] denotes the real part of the quantity enclosed inside the square brackets. The *complex baseband process* $\tilde{X}(t)$ is itself defined by

$$\tilde{X}(t) = X_I(t) + jX_Q(t) \tag{C.61}$$

where $X_I(t)$ is the *in-phase component* and $X_Q(t)$ is the *quadrature component*. Equivalently, we may express the original process $X(t)$ in terms of these two components as follows:

$$X(t) = X_I(t)\cos(2\pi f_c t) - X_Q(t)\sin(2\pi f_c t) \tag{C.62}$$

Correspondingly, for sample functions of $X(t,\omega)$, $X_I(t,\omega)$, and $X_Q(t,\omega)$, we may write

$$X(t, \omega) = X_I(t, \omega)\cos(2\pi f_c t) - X_Q(t, \omega)\sin(2\pi f_c t) \tag{C.63}$$

C.14 STATIONARY AND ERGODICITY

A random process is said to be *ergodic* if time averages of a sample function are equal to the corresponding ensemble average (or expectation) at a particular point in time. Mathematically, for a random process $X(t,\omega)$, this relationship can be expressed as

$$\mathbf{E}[X(t_0, \omega)] = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t, \omega)\,dt \tag{C.64}$$

where the left-hand side is the *ensemble average* (i.e., the expectation over all realizations ω at a particular point in time) and the right-hand side is the *time average* of the random process for a particular realization ω_0. In many physical applications, it is assumed that stationary processes are ergodic and that time averages and expectations can be used interchangeably.

NOTES AND REFERENCES

[1] For a detailed description of random variables and processes, see Leon-Garcia (1994).

APPENDIX D

Matched Filters

D.1 MATCHED-FILTER RECEIVER

Consider a known signal $s(t)$ corrupted by additive white Gaussian noise $w(t)$, resulting in the received signal

$$x(t) = s(t) + w(t) \qquad 0 \le t \le T \tag{D.1}$$

What is the optimum receiver for *detecting* the known signal $s(t)$ in the received signal $x(t)$? To answer this fundamental question, we first note the following two important points:

1. The power spectral density of white noise, with sample function $w(t)$, is defined by

$$S_w(f) = \frac{N_0}{2} \quad \text{for all } f \text{ in the entire interval} \quad -\infty < f < \infty \tag{D.2}$$

The power spectral density of white noise is illustrated in Fig. D.1(a). For a stationary random process, the autocorrelation function is the inverse Fourier transform of the power spectral density. (See Appendix C.) It follows, therefore, that the autocorrelation function of white noise consists of a Dirac delta function $\delta(\tau)$, weighted by $N_0/2$, as shown in Fig. D.1(b). That is,

$$\begin{aligned} R_w(\tau) &= \mathbf{E}[w(\tau)w(t-\tau)] \\ &= \frac{N_0}{2}\delta(\tau) \end{aligned} \tag{D.3}$$

where $\mathbf{E}$ is the statistical expectation operator. Accordingly, any two different samples of white noise are uncorrelated, no matter how closely together in time they are taken. If the white noise $w(t)$ is also Gaussian, then the two samples are statistically independent. In a sense, white Gaussian noise represents the *ultimate in randomness.*

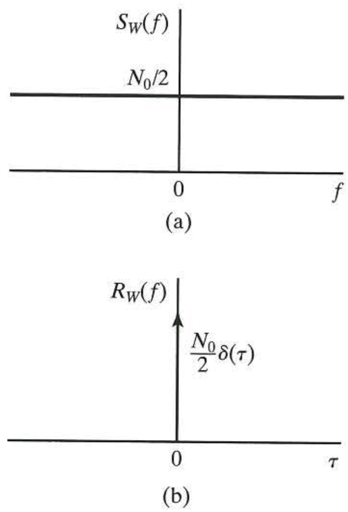


FIGURE D.1 (a) Power spectrum of the additive white noise $W(t)$.
(b) Autocorrelation function of $W(t)$.

2. Since the signal $s(t)$ is known and therefore deterministic, it follows that $s(t)$ and $w(t)$ are as uncorrelated (i.e., dissimilar) as they could ever be.

In light of point 2, we may intuitively state that, for the problem described herein, the optimum receiver consists of a *correlator* with two inputs, one being the noisy received signal $x(t)$ and the other being a locally generated replica of the known signal $s(t)$, as shown in Fig. D.2. For obvious reasons, this optimum receiver is known as the *correlation receiver*.

Another way of constructing the optimum receiver is to use a *matched filter*, defined as a linear filter whose impulse response $h(t)$ is a time-reversed, delayed version of the known signal $s(t)$; that is,

$$h(t) = \begin{cases} s(T-t) & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{D.4}$$

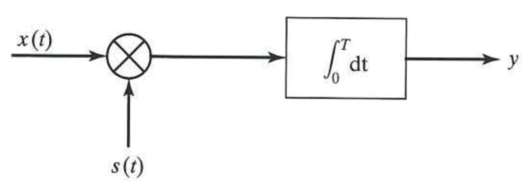


FIGURE D.2 Correlation receiver.

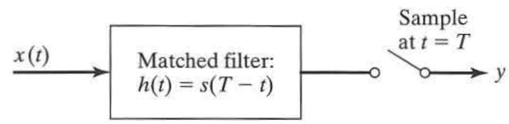


FIGURE D.3 Matched-filter receiver.

Figure D.3 shows a *matched filter receiver*, which consists of a matched filter followed by a sampler that is activated at the end of the signaling interval $t = T$. The important point to note here is that the correlation receiver of Fig. D.2 and the matched filter receiver of Fig. D.3 are equivalent insofar as their overall output samples are concerned. Specifically, for the same input signal and at the end of a signaling interval, the resulting output samples produced by these two receivers are identical.

D.2 PROBABILITY OF DETECTION

To detect a signal with a correlation receiver or a matched-filter receiver, the output sample is compared against a threshold and then a decision is made by the receiver, depending on whether the threshold is exceeded or not. In so doing, the receiver makes a decision in favor of one of two hypotheses:

Hypothesis H_1: The known signal $s(t)$ is present in the received signal $x(t)$, a decision that is made when the threshold is exceeded.

Hypothesis H_0: The received signal $x(t)$ consists solely of noise $w(t)$, a decision that is made when the threshold is not exceeded.

Clearly, the receiver is subject to *errors* due to the random behavior of the additive noise $w(t)$ in the received signal $x(t)$.

To calculate the *average probability of error* incurred by the receiver, we proceed by using Eq. (D.1) as the input signal applied to the correlation receiver of Fig. D.2. The resulting output sample is

$$\begin{aligned} y &= \int_0^T x(t)s(t)dt \\ &= \int_0^T (s(t) + w(t))s(t)dt \\ &= \int_0^T s^2(t)dt + \int_0^T w(t)s(t)dt \\ &= E + \int_0^T w(t)s(t)dt \end{aligned} \tag{D.5}$$

where

$$E = \int_0^T s^2(t)dt \tag{D.6}$$

is the *energy* of the known signal $s(t)$.

Since, by assumption, the white noise $w(t)$ is the sample function of a Gaussian process $W(t)$, it follows that the receiver output y is the sample of a Gaussian-distributed random variable Y. To complete the characterization of the receiver output, we need to determine its mean and variance.

The mean of the random variable Y is

$$\begin{aligned}\mu_Y &= \mathbf{E}[Y] \\ &= E + \mathbf{E}\left[\int_0^T W(t)s(t)dt\right] \\ &= E + \mathbf{E}\int_0^T [W(t)]s(t)dt \\ &= E\end{aligned} \tag{D.7}$$

where we have used two facts: First, the known signal $s(t)$ is deterministic and therefore unaffected by the expectation operator $\mathbf{E}$. Second, by assumption, the mean of the white noise process $W(t)$ is zero.

The variance of the random variable Y is

$$\begin{aligned}\sigma_Y^2 &= \mathbf{E}[(Y-\mu_Y)^2] \\ &= \mathbf{E}\left[\int_0^T\int_0^T W(t_1)W(t_2)s(t_1)s(t_2)dt_1dt_2\right] \\ &= \int_0^T\int_0^T \mathbf{E}[W(t_1)W(t_2)]s(t_1)s(t_2)dt_1dt_2\end{aligned} \tag{D.8}$$

Invoking the use of Eq. (D.2), we may write

$$\mathbf{E}[W(t_1)W(t_2)] = \frac{N_0}{2}\delta(t_1-t_2) \tag{D.9}$$

Substituting Eq. (D.9) into (D.8) yields

$$\begin{aligned}\sigma_Y^2 &= \frac{N_0}{2}\int_0^T\int_0^T \delta(t_1-t_2)s(t_1)s(t_2)dt_1dt_2 \\ &= \frac{N_0}{2}\int_0^T s^2(t_1)dt_1 \\ &= \frac{N_0E}{2}\end{aligned} \tag{D.10}$$

where E is the *signal energy.*

Putting all the pieces together, we can now say that the correlation receiver output y is the sample value of a Gaussian-distributed random variable Y with mean

$\mu_Y = E$ and variance $\sigma_Y^2 = N_0E/2$. Accordingly, we may express the probability density function of the random variable Y as

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right) \\ &= \frac{1}{\sqrt{\pi N_0 E}}\exp\left(-\frac{(y-E)^2}{N_0E}\right) \end{aligned} \tag{D.11}$$

which is plotted in Fig. D.4.

Let λ denote the *threshold* against which the correlator output y is compared. As stated previously, when $y > \lambda$, the receiver decides in favor of hypothesis H_1; otherwise it decides in favor of hypothesis H_0. Accordingly, the conditional probability of error, given that the known signal $s(t)$ is present in the receiver input, is defined by

$$\text{Prob(say } H_0|H_1 \text{ is true)} = \int_{-\infty}^{\lambda} f_Y(y)dy \tag{D.12}$$

which is illustrated graphically in Fig. D.4. Substituting Eq. (D.11) into (D.12) yields

$$\text{Prob(say } H_0|H_1 \text{ is true)} = \frac{1}{\sqrt{\pi N_0 E}}\int_{-\infty}^{\lambda} \exp\left(-\frac{(y-E)^2}{N_0E}\right)dy \tag{D.13}$$

To simplify matters, let

$$z = \frac{y-E}{\sqrt{N_0E}} \tag{D.14}$$

which means that

$$dz = \frac{dy}{\sqrt{N_0E}}$$

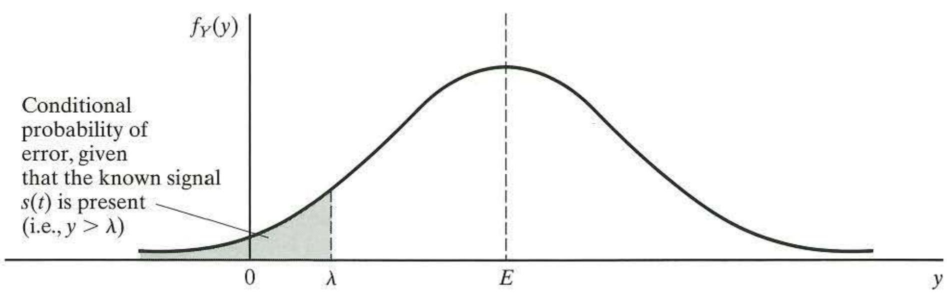


FIGURE D.4 Probability distribution of the correlation receiver output.

Hence, we may rewrite Eq. (D.13) as

$$\begin{aligned}\text{Prob(say } H_0|H_1 \text{ is true)} &= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{(\lambda-E)/\sqrt{N_0E}} \exp(-z^2)dz \\ &= \frac{1}{\sqrt{\pi}}\int_{(E-\lambda)/\sqrt{N_0E}}^{\infty} \exp(-z^2)dz\end{aligned} \tag{D.15}$$

At this point in the discussion, we digress briefly to introduce a function that is closely related to the Gaussian distribution: the *error function*, defined by

$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}}\int_0^u \exp(-z^2)dz \tag{D.16}$$

Table E.1 of Appendix E gives values of the error function erf(u) for the argument u in the interval $0 \le u \le 3.3$. The error function has two useful properties:

1. *Symmetry property*, described by

$$\operatorname{erf}(-u) = -\operatorname{erf}(u) \tag{D.17}$$

2. *Asymptote property*, which, for the argument u approaching infinity, is described by

$$\begin{aligned}\operatorname{erf}(\infty) &= \frac{2}{\sqrt{\pi}}\int_0^\infty \exp(-z^2)dz \\ &= 1\end{aligned} \tag{D.18}$$

Another function, the *complementary error function*, is defined by

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}}\int_u^\infty \exp(-z^2)dz \tag{D.19}$$

which is related to the error function by the formula

$$\operatorname{erfc}(u) = 1 - \operatorname{erf}(u) \tag{D.20}$$

We may now reformulate the conditional probability of error of Eq. (D.15) in terms of the complementary error function by writing

$$\text{Prob(say } H_0|H_1 \text{ is true)} = \frac{1}{2}\operatorname{erfc}\left(\frac{E-\lambda}{\sqrt{N_0E}}\right) \tag{D.21}$$

From Eq. (D.21), the following points are noteworthy:

- The signal energy E and noise spectral density N_0 have different physical interpretations, in that E is measured in joules whereas N_0 is measured in watts/hertz; yet these two units are in fact equal.

- Insofar as the signal component is concerned, the probability of error is independent of the waveform of the known signal $s(t)$, and the only parameter that matters is the signal energy E.
- The threshold λ is measured in joules.

D.3 ANOTHER PROPERTY OF THE MATCHED FILTER

Equation (D.21) sums up one important property of the matched-filter receiver in the combined presence of signal and noise at the filter input. For another important property of the matched filter, consider the case of a noiseless input. Then, with the input $x(t) = s(t)$ and the impulse response $h(t) = s(T - t)$, the resulting filter output is defined by the convolution integral

$$\begin{aligned} y(t) &= R_s(T-t) \\ &= R_s(t-T) \qquad \textit{if } w(t) = 0 \end{aligned} \tag{D.22}$$

The integral of Eq. (D.22) is recognized as the deterministic autocorrelation function of the signal component $s(t)$ for a lag of $T - t$, namely, $R_s(T - t)$. Accordingly, we may write

$$\begin{aligned} y(t) &= R_s(T-t) \\ &= R_s(T-t) \qquad wt = 0 \end{aligned} \tag{D.23}$$

where, in the second line, we have used the fact that the autocorrelation function of a signal of finite energy is an even function of the lag (see Appendix A). In words, Eq. (D.23) states that *the output of a filter matched to an input signal is equal to the autocorrelation function of that signal, delayed by an amount equal to the duration of the signal.*

D.4 MATCHED FILTERING FOR COMPLEX SIGNALS

The material presented thus far on matched filtering applies to real-valued signals. When dealing with complex-valued signals, we make a simple modification to Eq. (D.4). Specifically, the impulse response of a filter matched to a complex-valued signal $s(t)$ is defined by

$$h(t) = \begin{cases} s^*(T-t) & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{D.24}$$

where the asterisk denotes complex conjuction. Except for this minor modification, everything else presented in the Appendix remains intact.

A P P E N D I X E

Error Function

E.1 DEFINITIONS

The *error function*, denoted by erf(u), is defined in a number of different ways in the literature. We shall use the following definition:

$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-z^2)\,dz \tag{E.1}$$

The error function has two useful properties:

1.

$$\operatorname{erf}(-u) = -\operatorname{erf}(u) \tag{E.2}$$

   This is known as the *symmetry property.*

2. As u approaches infinity, erf(u) approaches unity; that is,

$$\frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-z^2)\,dz = 1 \tag{E.3}$$

   This is known as the *asymptote property.*

The *complementary error function* is defined by

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-z^2)\,dz \tag{E.4}$$

The complementary error function is related to the error function as follows:

$$\operatorname{erfc}(u) = 1 - \operatorname{erf}(u) \tag{E.5}$$

Table E.1 gives values of the error function erf(u) for u in the range from 0 to 3.3.

TABLE E.1 The Error Function[a].

u	erf(u)	u	erf(u)
0.00	0.00000	1.10	0.88021
0.05	0.05637	1.15	0.89612
0.10	0.11246	1.20	0.91031
0.15	0.16800	1.25	0.92290
0.20	0.22270	1.30	0.93401
0.25	0.27633	1.35	0.94376
0.30	0.32863	1.40	0.95229
0.35	0.37938	1.45	0.95970
0.40	0.42839	1.5	0.96611
0.45	0.47548	1.55	0.97162
0.50	0.52050	1.6	0.97635
0.55	0.56332	1.65	0.98038
0.60	0.69386	1.70	0.98379
0.65	0.64203	1.75	0.98667
0.70	0.67780	1.80	0.98909
0.75	0.71116	1.85	0.99111
0.80	0.74210	1.90	0.99279
0.85	0.77067	1.95	0.99418
0.90	0.79691	2.00	0.99532
0.95	0.82089	2.50	0.99959
1.00	0.84270	3.00	0.99998
1.05	0.86244	3.30	0.999998

[a] The error function is tabulated extensively in several references; see for example, Abramowitz and Stegun (1965, pp. 297–316).

E.2 BOUNDS ON THE COMPLEMENTARY ERROR FUNCTION

Substituting $u - x$ for z in Eq. (E.4), we get

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}} \exp(-u^2) \int_{-\infty}^{0} \exp(2ux) \exp(-x^2) dx$$

For any real x, the value of $\exp(-x^2)$ lies between the successive partial sums of the power series

$$1 - \frac{x^2}{1!} + \frac{(x^2)^2}{2!} - \frac{(x^2)^3}{3!} + \dots$$

Therefore, for $u > 0$, we find, on using $(n+1)$ terms of this series, that erfc(u) lies between the values taken by

$$\frac{2}{\sqrt{\pi}} \exp(-u^2) \int_{\infty}^{0} \left(1 - x^2 + \frac{x^4}{2} - \dots \pm \frac{x^{2n}}{n!}\right) \exp(2ux) dx$$

for even n and for odd n. Putting $2ux = -\upsilon$ and using the integral

$$\int_{0}^{\infty} \upsilon^n \exp(-\upsilon) d\upsilon = n!$$

we obtain the following *asymptotic expansion* for erfc(u), assuming that $u > 0$:

$$\operatorname{erfc}(u) \approx \frac{\exp(-u^2)}{\sqrt{\pi}u}\left[1 - \frac{1}{2u^2} + \frac{1 \cdot 3}{2^2 u^4} - \dots \pm \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n u^{2n}}\right] \tag{E.6}$$

For large positive values of u, the successive terms of the series on the right-hand side of Eq. (E.6) decrease very rapidly. We thus deduce two simple bounds on erfc(u), one lower and the other upper, given by the inequality[1]

$$\frac{\exp(-u^2)}{\sqrt{\pi}u}\left(1 - \frac{1}{2u^2}\right) < \operatorname{erfc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi}u} \tag{E.7}$$

For large positive u, a second bound on the complementary error function erfc(u) is obtained by omitting the multiplying factor $1/u$ in the upper bound of Eq. (E.7):

$$\operatorname{erfc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi}} \tag{E.8}$$

In Fig. E.1, we have plotted erfc(u), the two bounds defined by Eq. (E.7), and the upper bound of Eq. (E.8). We see that, for $u \geq 1.5$, the bounds on erfc(u), defined by Eq. (E.7), become increasingly tight.

E.3 THE *Q*-FUNCTION

Consider a *standardized* Gaussian random variable X of zero mean and unit variance. The probability that an observed value of the random variable X will be greater than v is given by the *Q-function*:

$$Q(v) = \frac{1}{\sqrt{2\pi}} \int_v^{\infty} \exp\left(-\frac{x^2}{2}\right) \tag{E.9}$$

The *Q-function defines the area under the standardized Gaussian tail.* Inspection of Eqs. (E.4) and (E.9) reveals that the Q-function is related to the complementary error function as follows:

$$Q(v) = \frac{1}{2}\operatorname{erfc}\left(\frac{v}{\sqrt{2}}\right) \tag{E.10}$$

Conversely, putting $u = v$, we have

$$\operatorname{erfc}(u) = 2Q(\sqrt{2}u) \tag{E.11}$$

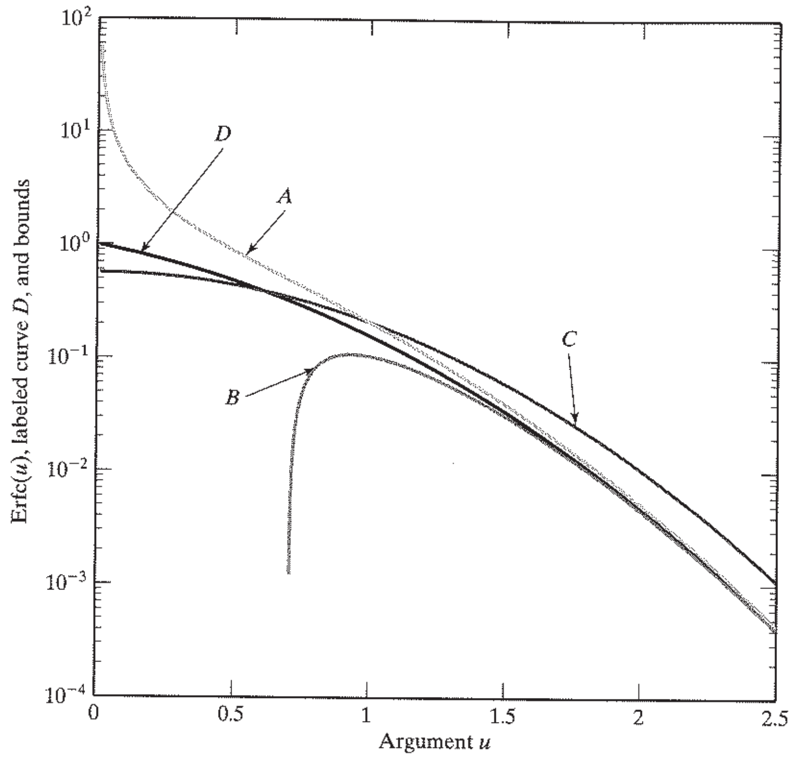


FIGURE E.1 The complementary error function erfc(u) and its bounds:

Curve A: $\dfrac{e^{-u^2}}{\sqrt{\pi}u}$

Curve B: $\dfrac{e^{-u^2}}{\sqrt{\pi}u}\left(1 - \dfrac{1}{2u^2}\right)$

Curve C: $\dfrac{e^{-u^2}}{\sqrt{\pi}}$

Curve D: erfc(u)

NOTES AND REFERENCES

[1] The derivation of Eq. (E.7) follows Blachman (1966).

APPENDIX F

MAP Algorithm

F.1 SEPARABILITY THEOREM

Following the terminology introduced in Section 4.12.4, let the vectors $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$ denote estimates of the state probabilities in a turbo decoder at time t that are based on past and future data, respectively. According to the *separability theorem*, the state probabilities at time t are related to $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$ by

$$\lambda(t) = \frac{\boldsymbol{\alpha}(t) \cdot \boldsymbol{\beta}(t)}{\|\boldsymbol{\alpha}(t) \cdot \boldsymbol{\beta}(t)\|_1} \tag{F.1}$$

where the numerator is the vector product of $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$ and the denominator is the L_1 norm of this product, as defined in Section 4.12.

Proof: For any m, for which $\text{Prob}(s(t) = m) \neq 0$,

$$\begin{aligned}
\lambda_m(t) &= \text{Prob}(s(t) = m|\mathbf{y}) \\
&= \frac{\text{Prob}(s(t) = m, \mathbf{y})}{\text{Prob}(\mathbf{y})} \\
&= \text{Prob}(\mathbf{y}|s(t) = m) \cdot \frac{\text{Prob}(s(t) = m)}{\text{Prob}(\mathbf{y})} \\
&= \text{Prob}(\mathbf{y}_{[1,\,t]}|s(t) = m) \cdot \text{Prob}(\mathbf{y}_{[t+1,\,T]}, s(t) = m) \cdot \frac{\text{Prob}(s(t) = m)}{\text{Prob}(\mathbf{y})} \\
&= \frac{\text{Prob}(\mathbf{y}_{[1,\,t]}|s(t) = m)}{\text{Prob}(s(t) = m)} \cdot \frac{\text{Prob}(\mathbf{y}_{[t+1,\,T]}, s(t) = m)}{\text{Prob}(s(t) = m)} \cdot \frac{\text{Prob}(s(t) = m)}{\text{Prob}(\mathbf{y})} \\
&= \text{Prob}(s(t) = m|\mathbf{y}_{[1,\,t]}) \cdot \text{Prob}(s(t) = m|\mathbf{y}_{[t+1,\,T]}) \cdot \left(\frac{\text{Prob}(\mathbf{y}_{[1,\,t]}|\mathbf{y}_{[t+1,\,T]})}{\text{Prob}(s(t) = m)\text{Prob}(\mathbf{y})}\right)
\end{aligned} \tag{F.2}$$

where the second and third lines in the development follow from Bayes' rule, the fourth line follows from the fact that the decoding process is a Markov process, and the fifth and sixth lines are further manipulations using Bayes' rule.

Often, the a priori probabilities, $\text{Prob}(s(t) = m)$, are independent of m. (In practice, some of the a priori probabilities in a turbo decoding process may be zero due to certain features of the trellis—for example, at start-up. In that case, this requirement applies only to those states whose a priori probabilities are nonzero.) This is usually the case for a time-invariant trellis, for which the inputs are equiprobable. In that case, the bracketed term in the last line on the right-hand side is independent of m. Since the summation of the left-hand side of Eq. (F.2) over m must equal unity, it follows that this bracketed term must normalize the right-hand side to sum to unity. Thus, when we identify the first and second terms in the product of the last line as $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$, we have

$$\lambda(t) = \frac{\boldsymbol{\alpha}(t) \cdot \boldsymbol{\beta}(t)}{\|\boldsymbol{\alpha}(t) \cdot \boldsymbol{\beta}(t)\|_1} \tag{F.3}$$

which proves the theorem.

A P P E N D I X G

Capacity of MIMO Links

G.1 PRELIMINARIES[1]

The purpose of this appendix is to present a derivation of the log-det capacity formula of Eq. (6.59). To prepare the way for the derivation, we briefly review some basic concepts in information theory.

Consider a continuous random variable X with probability density function $f_X(x)$. The *differential entropy* of the random variable X, measured in *bits*, is defined by

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)dx \\ &= -\mathbf{E}[\log_2 f_X(x)] \text{ bits} \end{aligned} \tag{G.1}$$

where $\mathbf{E}$ is the statistical expectation operator. It is important to note that the symbol X in the entropy $h(X)$ is *not* the argument of a function; rather, it merely serves the purpose of a label for the source of information.

When we have a continuous random vector $\mathbf{X}$ consisting of N random variables $X_1, X_2, ..., X_N$, we may generalize Eq. (G.1) and define the differential entropy of $\mathbf{X}$ as the N-fold integral

$$\begin{aligned} h(\mathbf{X}) &= -\int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\log_2 f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\ &= -\mathbf{E}[\log_2 f_{\mathbf{X}}(\mathbf{x})] \text{ bits} \end{aligned} \tag{G.2}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function of the random vector $\mathbf{X}$.

The logarithmic description of entropy is evident from both Eqs. (G.1) and (G.2). This particular form of description is in perfect accord with the notion of entropy in thermodynamics.

Equations (G.1) and (G.2) apply to random data, real or complex. The difference between these two forms of data manifests itself in the way in which the pertinent probability density functions are defined, as illustrated in the next example.

EXAMPLE G.1 Complex Multidimensional Gaussian Distribution

Consider an N-dimensional complex Gaussian-distributed vector $\mathbf{X}$. Each element of $\mathbf{X}$ consists of an in-phase component $X_{k,I}$ and a quadrature component $X_{k,Q}$, so

$$X_k = X_{k,I} + jX_{k,Q} \qquad k = 1,2,\dots,N \tag{G.3}$$

or, vectorially,

$$\mathbf{X} = \mathbf{X}_I + j\mathbf{X}_Q \tag{G.4}$$

It is assumed that $\mathbf{X}$ has zero mean. The requirement is to determine the differential entropy of $\mathbf{X}$.

If the components $\mathbf{X}_I$ and $\mathbf{X}_Q$ are orthogonal—that is, if we have

$$E[\mathbf{X}_I\mathbf{X}_Q^T] = \mathbf{O} \tag{G.5}$$

and if they are both Gaussian distributed, then they are statistically independent, or

$$f_{\mathbf{X}_I,\mathbf{X}_Q}(\mathbf{x}_I,\mathbf{x}_Q) = f_{\mathbf{X}_I}(\mathbf{x}_I)f_{\mathbf{X}_Q}(\mathbf{x}_Q) \tag{G.6}$$

The in-phase component $\mathbf{X}_I$ and quadrature component $\mathbf{X}_Q$ share the same formula for their joint probability density functions. We therefore make two important observations:

1. The components $\mathbf{X}_I$ and $\mathbf{X}_Q$ have exactly the same entropy.
2. Since the differential entropy is logarithmic in nature, it follows that the differential entropies of $\mathbf{X}_I$ and $\mathbf{X}_Q$ are additive in terms of calculating the differential entropy of $\mathbf{X}$.

Hence, we may write

$$h(\mathbf{X}_I) = h(\mathbf{X}_Q) \tag{G.7}$$

and

$$\begin{aligned} h(\mathbf{X}) &= h(\mathbf{X}_I) + h(\mathbf{X}_Q) \\ &= 2h(\mathbf{X}_I) \end{aligned} \tag{G.8}$$

The joint probability density function of the complex Gaussian vector $\mathbf{X}$ with zero mean and correlation matrix $\mathbf{R_x}$ is defined by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^N \det(\mathbf{R_x})}\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{R_x}^{-1}\mathbf{x}\right) \tag{G.9}$$

where $\mathbf{R_x}^{-1}$ is the inverse of $\mathbf{R_x}$ and $\det(\mathbf{R_x})$ is the determinant of $\mathbf{R_x}$. Substituting Eq. (G.9) into (G.2), using the fact that the volume under $f_{\mathbf{X}}(\mathbf{x})$ is unity, and then simplifying terms, we get

$$h(\mathbf{X}) = N + N\log_2(2\pi) + \log_2\{\det(\mathbf{R_x})\} \text{ bits} \tag{G.10}$$

which is uniquely defined by the correlation matrix $\mathbf{R_x}$.

For the special case of a scalar complex Gaussian random variable X, $N = 1$ and Eq. (G.10) reduces to

$$h(X) = 1 + \log_2(2\pi\sigma_X^2) \text{ bits} \qquad (X: \text{complex}) \tag{G.11}$$

where σ_X^2 is the variance of X. If X is real, we have

$$h(X) = \frac{1}{2}[1 + \log_2(2\pi\sigma_X^2)] \text{ bits} \qquad (X: \text{real}) \tag{G.12}$$

For a given variance σ_X^2, the Gaussian random variable X has the largest differential entropy attainable by any random variable in its class (i.e., real or complex). A similar remark applies to a multivariate Gaussian distribution. ■

For the discussion at hand, we need one other notion: mutual information, which applies to a pair of related random variables or random vectors. To be specific, consider a pair of random variables X and Y with joint probability density function $f_{X,Y}(x,y)$. The *mutual information* between X and Y is defined by

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left(\frac{f_{X|Y}(x|y)}{f_X(x)}\right)dx\,dy \tag{G.13}$$

where $f_{X|Y}(x|y)$ is the conditional probability density function of X, given that $Y = y$. In words, the mutual information $I(X;Y)$ is a *measure of the uncertainty about the random variable X that is resolved by observing the second random variable Y.*

On the basis of Eq. (G.13), we may derive the following properties of mutual information that hold in general:

$$I(X;Y) \geq 0 \tag{G.14}$$

$$I(X;Y) = I(Y;X) \tag{G.15}$$

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) \\ &= h(Y) - h(Y|X) \end{aligned} \tag{G.16}$$

Here, $h(X)$ and $h(Y)$ are the differential entropies of X and Y, respectively, and

$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2 f_{X|Y}(x|y)\,dx\,dy \tag{G.17}$$

is the conditional differential entropy of X, given Y.

Formulas similar to Eqs. (G.13) through (G.17) apply to a related pair of random vectors $\mathbf{X}$ and $\mathbf{Y}$.

With the definitions of differential entropy, conditional differential entropy, and mutual information at hand, we are ready to proceed with the derivation of the log-det capacity formula.

G.2 LOG-DET CAPACITY FORMULA OF MIMO LINK[2]

Consider a communication link with multiple antennas. Let the N_t-by-1 vector $\mathbf{s}$ denote the transmitted signal vector and the N_r-by-1 vector $\mathbf{x}$ denote the received signal vector. These two vectors are related by the *input–output relation of the channel*, namely,

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{G.18}$$

where $\mathbf{H}$ is the *channel matrix* of the link and $\mathbf{w}$ is the additive channel noise vector. The vectors $\mathbf{s}$, $\mathbf{w}$, and $\mathbf{x}$ are realizations of the random vectors $\mathbf{S}$, $\mathbf{W}$, and $\mathbf{X}$, respectively.

In the rest of this section, the following assumptions are made:

1. The channel is stationary and ergodic.
2. The channel matrix $\mathbf{H}$ is made up of i.i.d. Gaussian elements.
3. The channel state $\mathbf{H}$ is known to the receiver, but not the transmitter.
4. The transmitted signal vector $\mathbf{s}$ has zero mean and correlation matrix $\mathbf{R_s}$.
5. The additive channel noise vector $\mathbf{w}$ has zero mean and correlation matrix $\mathbf{R_w}$.
6. Both $\mathbf{s}$ and $\mathbf{w}$ are governed by Gaussian distributions.

With both $\mathbf{H}$ and $\mathbf{x}$ unknown to the transmitter, the primary issue of interest is to determine $I(\mathbf{s};\mathbf{x},\mathbf{H})$, which denotes the mutual information between the transmitted signal vector $\mathbf{s}$ and both the received signal vector $\mathbf{x}$ and the channel matrix $\mathbf{H}$. Extending the definition of mutual information given in Eq. (G.13) to the problem at hand, we write

$$I(\mathbf{S};\mathbf{X},\mathbf{H}) = \iiint_{\mathcal{H}\mathcal{X}\mathcal{S}} f_{\mathbf{S},\mathbf{X},\mathbf{H}}(\mathbf{s},\mathbf{x},\mathbf{H})\log_2\left(\frac{f_{\mathbf{S}|\mathbf{X},\mathbf{H}}(\mathbf{s}|\mathbf{x},\mathbf{H})}{f_{\mathbf{X},\mathbf{H}}(\mathbf{x},\mathbf{H})}\right)d\mathbf{s}\,d\mathbf{x}\,d\mathbf{H} \tag{G.19}$$

where $\mathcal{S}$, $\mathcal{X}$, and $\mathcal{H}$ are the respective spaces pertaining to the random vectors $\mathbf{S}$ and $\mathbf{X}$ and the matrix $\mathbf{H}$. According to Bayes' rule, we have

$$f_{\mathbf{S},\mathbf{X},\mathbf{H}}(\mathbf{s},\mathbf{x},\mathbf{H}) = f_{\mathbf{S},\mathbf{X}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H})f_{\mathbf{H}}(\mathbf{H})$$

We may therefore rewrite Eq. (G.19) in the equivalent form

$$\begin{aligned}
I(\mathbf{S};\mathbf{X},\mathbf{H}) &= \int_{\mathcal{H}} f_{\mathbf{H}}(\mathbf{H})\left[\iint_{\mathcal{X}\mathcal{S}} f_{\mathbf{S},\mathbf{X}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H})\log_2\left(\frac{f_{\mathbf{S}|\mathbf{X},\mathbf{H}}(\mathbf{s}|\mathbf{x},\mathbf{H})}{f_{\mathbf{X},\mathbf{H}}(\mathbf{x},\mathbf{H})}\right)d\mathbf{s}\,d\mathbf{x}\right]d\mathbf{H} \\
&= \mathbf{E_H}\left[\iint_{\mathcal{X}\mathcal{S}} f_{\mathbf{S},\mathbf{X}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H})\log_2\left(\frac{f_{\mathbf{S}|\mathbf{X},\mathbf{H}}(\mathbf{s}|\mathbf{x},\mathbf{H})}{f_{\mathbf{X},\mathbf{H}}(\mathbf{x},\mathbf{H})}\right)d\mathbf{s}\,d\mathbf{x}\right] \\
&= \mathbf{E_H}[I(\mathbf{s};\mathbf{x}|\mathbf{H})]
\end{aligned} \tag{G.20}$$

where the expectation is with respect to the channel matrix $\mathbf{H}$, and

$$I(\mathbf{s};\mathbf{x}|\mathbf{H}) = \iint_{\mathcal{X}\mathcal{S}} f_{\mathbf{S},\mathbf{X}|\mathbf{H}}(\mathbf{s},\mathbf{x}|\mathbf{H})\log_2\left(\frac{f_{\mathbf{S}|\mathbf{X},\mathbf{H}}(\mathbf{s}|\mathbf{x},\mathbf{H})}{f_{\mathbf{X},\mathbf{H}}(\mathbf{x},\mathbf{H})}\right)d\mathbf{s}\,d\mathbf{x}$$

is the conditional mutual information between the transmitted signal vector $\mathbf{s}$ and the received signal vector $\mathbf{x}$, given the channel matrix $\mathbf{H}$. However, by assumption, the state of the channel is unknown to the transmitter. It follows, therefore, that insofar as the receiver is concerned, $I(\mathbf{s};\mathbf{x}|\mathbf{H})$ is a random variable—hence the expectation with respect to $\mathbf{H}$ in Eq. (G.20). The quantity resulting from this expectation is deterministic, defining the mutual information jointly between the transmitted signal vector **s** and both the received signal vector **x** and the channel matrix **H**. The result so obtained is indeed consistent with what we know about the notion of joint mutual information.

Next, applying the vector form of the first line in Eq. (G.16) to the mutual information $I(\mathbf{s};\mathbf{x}|\mathbf{H})$, we may write

$$I(\mathbf{s};\mathbf{x}|\mathbf{H}) = h(\mathbf{x}|\mathbf{H}) - h(\mathbf{x}|\mathbf{s},\mathbf{H}) \tag{G.21}$$

where $h(\mathbf{x}|\mathbf{H})$ is the conditional differential entropy of the input **x**, given **H**, and $h(\mathbf{x}|\mathbf{s},\mathbf{H})$ is the conditional differential entropy of the input **x**, given both **s** and **H**. Both of these entropies are random quantities, as they depend on **H**.

To proceed further, we now invoke the assumed Gaussian nature of both **s** and **H**, in which case **x** also assumes a Gaussian description. Under these assumptions, we may use Eq. (G.10) to express the entropy of the received signal **x** of dimension N_r, given **H**, as

$$h(\mathbf{x}|\mathbf{H}) = N_r + N_r \log_2(2\pi) + \log_2\{\det(\mathbf{R_x})\} \text{ bits} \tag{G.22}$$

where $\mathbf{R_x}$ is the correlation matrix of **x**. Recognizing that the transmitted signal vector **s** and the channel noise vector **w** are independent of each other, we find, from Eq. (G.18), that the correlation matrix of the received signal vector **x** is given by

$$\begin{aligned}
\mathbf{R_x} &= \mathbf{E}[\mathbf{x}\mathbf{x}^\dagger] \\
&= \mathbf{E}[(\mathbf{Hs}+\mathbf{w})(\mathbf{Hs}+\mathbf{w})^\dagger] \\
&= \mathbf{E}[(\mathbf{Hs}+\mathbf{w})(\mathbf{s}^\dagger\mathbf{H}^\dagger+\mathbf{w}^\dagger)] \\
&= \mathbf{E}[\mathbf{H}\mathbf{s}\mathbf{s}^\dagger\mathbf{H}^\dagger] + E[\mathbf{w}\mathbf{w}^\dagger] \quad \text{because} \quad \mathbf{E}[\mathbf{s}\mathbf{w}^\dagger] = \mathbf{0} \\
&= \mathbf{H}\mathbf{E}[\mathbf{s}\mathbf{s}^\dagger]\mathbf{H}^\dagger + \mathbf{R_w} \\
&= \mathbf{H}\mathbf{R_s}\mathbf{H}^\dagger + \mathbf{R_w}
\end{aligned} \tag{G.23}$$

where

$$\mathbf{R_s} = \mathbf{E}[\mathbf{s}\mathbf{s}^\dagger] \tag{G.24}$$

and

$$\mathbf{R_w} = \mathbf{E}[\mathbf{w}\mathbf{w}^\dagger] \tag{G.25}$$

Hence, using Eq. (G.23) in (G.22), we get

$$h(\mathbf{x}|\mathbf{H}) = N_r + N_r \log_2(2\pi) + \log_2\{\det(\mathbf{R_w} + \mathbf{H}\mathbf{R_s}\mathbf{H}^\dagger)\} \text{ bits} \tag{G.26}$$

Next, we note that, since the vectors **s** and **w** are independent, and since the sum of **w** plus **Hs** equals **x**, as indicated in Eq. (G.18), then the conditional differential entropy of **x**, given both **s** and **H**, is simply equal to the differential entropy of the additive channel noise vector **w**:

$$h(\mathbf{x}|\mathbf{s},\mathbf{H}) = h(\mathbf{w}) \tag{G.27}$$

Again invoking the formula of Eq. (G.10), we have

$$h(\mathbf{w}) = N_r + N_r \log_2(2\pi) + \log_2\{\det(\mathbf{R_w})\} \text{ bits} \tag{G.28}$$

Using Eqs. (G.26), (G.27), and (G.28) in Eq. (G.21), we get

$$\begin{aligned} I(\mathbf{s};\mathbf{x}|\mathbf{H}) &= \log_2\left\{\det(\mathbf{R_w} + \mathbf{H}\mathbf{R_s}\mathbf{H}^H)\right\} - \log_2\{\det(\mathbf{R_w})\} \\ &= \log_2\left\{\frac{\det(\mathbf{R_w} + \mathbf{H}\mathbf{R_s}\mathbf{H}^H)}{\det(\mathbf{R_w})}\right\} \end{aligned} \tag{G.29}$$

As was remarked previously, the conditional mutual information $I(\mathbf{s};\mathbf{x}|\mathbf{H})$ is a random variable. Hence, using Eq. (G.29) in (G.20), we finally formulate the ergodic capacity of the MIMO link as the expectation

$$C = \mathbf{E_H}\left[\log_2\left\{\frac{\det(\mathbf{R_w} + \mathbf{H}\mathbf{R_s}\mathbf{H}^H)}{\det(\mathbf{R_w})}\right\}\right] \text{ bits/s/Hz} \tag{G.30}$$

which is subject to the constraint

$$\max_{\mathbf{R_s}} \operatorname{tr}[\mathbf{R_s}] \le P \qquad P = \text{constant transmit power}$$

where tr[.] denotes the *trace* operator, which extracts the sum of the diagonal elements of the enclosed matrix.

Equation (G.30) is the desired log-det formula for the ergodic capacity of the MIMO link. This formula is of general applicability in that correlations among the elements of the transmitted signal vector **s** and among those of the channel noise vector **w** are permitted. However, the assumptions made in its derivation involve the Gaussian aspects of **s**, **H**, and **w**.

One last comment is in order: the white Gaussian input spectrum

$$\mathbf{R_s} = \sigma_s^2 \mathbf{I}_{N_t}$$

is not necessarily optimal; nevertheless, its application does yield a lower bound to the ergodic capacity C.

G.3 MIMO CAPACITY FOR CHANNEL KNOWN AT THE TRANSMITTER[3]

The log-det formula of Eq. (G.30) for the ergodic capacity of a MIMO flat-fading channel assumes that the state of the channel is known only at the receiver. What if the state is also known perfectly at the transmitter? Then the state of the channel becomes known to the entire system, which means that we may treat the channel matrix $\mathbf{H}$ as a constant. Hence, unlike the partially known case discussed in Section G.2, there is no longer the need for invoking the expectation operator in formulating the log-det capacity. Rather, the problem becomes one of constructing the optimal $\mathbf{R_s}$ (i.e., the correlation matrix of the transmitted signal vector $\mathbf{s}$) that maximizes the ergodic capacity. To simplify the construction procedure, we consider a MIMO link with $N_r = N_t = N$. Accordingly, using the assumption of additive white Gaussian noise with variance $\sigma_{\mathbf{w}}^2$ in the log-det capacity formula of Eq. (G.2), we get

$$C = \log_2\left\{\det\left(I_N + \frac{1}{\sigma_{\mathbf{w}}^2}\mathbf{H}\mathbf{R_s}\mathbf{H}^\dagger\right)\right\} \text{ bits/s/Hz} \tag{G.31}$$

We can now formally postulate the optimization problem at hand as follows:

Maximize the ergodic capacity C of Eq. (G.31) with respect to the correlation matrix $\mathbf{R_s}$, subject to two requirements:

1. *Nonnegative definite $\mathbf{R_s}$, which is a necessary requirement for a correlation matrix.*
2. *Global power constraint*

$$\text{tr}[\mathbf{R_s}] = P \tag{G.32}$$

where P is the total transmit power.

To proceed with construction of the optimal $\mathbf{R_s}$, we first use the *determinant identity*:

$$\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA}) \tag{G.33}$$

Applying this identity to Eq. (G.31) yields

$$C = \log_2\left\{\det\left(\mathbf{I}_N + \frac{1}{\sigma_{\mathbf{w}}^2}\mathbf{R_s}\mathbf{H}^\dagger\mathbf{H}\right)\right\} \text{ bits/s/Hz} \tag{G.34}$$

Diagonalizing the matrix product $\mathbf{H}^\dagger\mathbf{H}$ by invoking the *eigendecomposition* of a Hermitian matrix, we may write

$$\mathbf{U}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{U} = \Lambda \tag{G.35}$$

where Λ is a diagonal matrix made up of the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ and $\mathbf{U}$ is a unitary matrix whose columns are the associated eigenvectors. (The eigendecomposition of a Hermitian matrix is discussed in Appendix H.)We may rewrite Eq. (G.35) in the form

$$\mathbf{H}^\dagger\mathbf{H} = \mathbf{U}\Lambda\mathbf{U}^\dagger \tag{G.36}$$

Substituting Eq. (G.36) into Eq. (G.34), we get

$$C = \log_2\left\{\det\left(\mathbf{I}_N + \frac{1}{\sigma_{\mathbf{w}}^2}\mathbf{R}_{\mathbf{s}}\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\dagger}\right)\right\} \text{ bits/s/Hz} \tag{G.37}$$

Applying the determinant identity of Eq. (G.33) to Eq. (G.37) yields

$$\begin{aligned} C &= \log_2\left\{\det\left(\mathbf{I}_N + \frac{1}{\sigma_{\mathbf{w}}^2}\Lambda\mathbf{U}^{\dagger}\mathbf{R}_{\mathbf{s}}\mathbf{U}\right)\right\} \\ &= \log_2\left\{\det\left(\mathbf{I}_N + \frac{1}{\sigma_{\mathbf{w}}^2}\Lambda\bar{\mathbf{R}}_{\mathbf{s}}\right)\right\} \text{ bits/s/Hz} \end{aligned} \tag{G.38}$$

where

$$\bar{\mathbf{R}}_{\mathbf{s}} = \mathbf{U}^{\dagger}\mathbf{R}_{\mathbf{s}}\mathbf{U} \tag{G.39}$$

Note that the transformed matrix $\bar{\mathbf{R}}_{\mathbf{s}}$ is nonnegative definite. Note also that

$$\begin{aligned} \mathrm{tr}[\bar{\mathbf{R}}_{\mathbf{s}}] &= \mathrm{tr}[\mathbf{U}^{\dagger}\mathbf{R}_{\mathbf{s}}\mathbf{U}] \\ &= \mathrm{tr}[\mathbf{U}\mathbf{U}^{\dagger}\mathbf{R}_{\mathbf{s}}] \\ &= \mathrm{tr}[\mathbf{R}_{\mathbf{s}}] \end{aligned} \tag{G.40}$$

It follows, therefore, that maximization of the capacity of Eq. (G.38) can be carried equally well over the transformed correlation matrix $\bar{\mathbf{R}}_{\mathbf{s}}$.

One other important point to note is that any nonnegative definite matrix $\mathbf{A}$ satisfies the *Hadamard inequality*

$$\det(\mathbf{A}) \le \prod_k a_{kk} \tag{G.41}$$

where the a_{kk} are the diagonal elements of the matrix $\mathbf{A}$. Hence, applying this inequality to the determinant in Eq. (G.38), we may write

$$\det\left(\mathbf{I}_N + \frac{1}{\sigma_{\mathbf{w}}^2}\Lambda\bar{\mathbf{R}}_{\mathbf{s}}\right) \le \prod_{k=1}^{N}\left(1 + \frac{1}{\sigma_{\mathbf{w}}^2}\lambda_k\bar{r}_{s,\,kk}\right) \tag{G.42}$$

where λ_k is the kth eigenvalue of the matrix product $\mathbf{H}\mathbf{H}^{\dagger}$ and $\bar{r}_{s,\,kk}$ is the kth diagonal element of the transformed matrix $\bar{\mathbf{R}}_{\mathbf{s}}$. The equality in Eq. (G.42) holds only when $\bar{\mathbf{R}}_{\mathbf{s}}$ is a diagonal matrix, which is the very condition that maximizes the ergodic capacity C.

To proceed further, we now use Eq. (G.38) and Eq. (G.42) with the equality sign to express the capacity as

$$\begin{aligned} C &= \log_2\left\{\prod_{k=1}^{N}\left(1+\frac{1}{\sigma_{\mathbf{w}}^2}\lambda_k\bar{r}_{s,\,kk}\right)\right\} \\ &= \sum_{k=1}^{N}\log_2\left(1+\frac{1}{\sigma_{\mathbf{w}}^2}\lambda_k\bar{r}_{s,\,kk}\right) \\ &= \sum_{k=1}^{N}\log_2\left\{\lambda_k\left(\lambda_k^{-1}+\frac{1}{\sigma_{\mathbf{w}}^2}\bar{r}_{s,\,kk}\right)\right\} \\ &= \sum_{k=1}^{N}\log_2\lambda_k+\sum_{k=1}^{N}\log_2\left(\lambda_k^{-1}+\frac{1}{\sigma_{\mathbf{w}}^2}\bar{r}_{s,\,kk}\right) \end{aligned} \tag{G.43}$$

where only the second summation is clearly adjustable. We may therefore reformulate the optimization problem at hand as follows:

Given the set of eigenvalues $\{\lambda_k\}_{k=1}^{N}$ *pertaining to the matrix product* $\mathbf{HH}^\dagger$, *determine the optimal set of autocorrelations* $\{\bar{r}_{s,\,kk}\}_{k=1}^{N}$ *that maximizes the summation*

$$\sum_{k=1}^{N}\left(\frac{1}{\lambda_k}+\frac{1}{\sigma_{\mathbf{w}}^2}\bar{r}_{s,\,kk}\right)$$

subject to the constraint

$$\sum_{k=1}^{N}\bar{r}_{s,\,kk} = P \tag{G.44}$$

The global power constraint of Eq. (G.44) follows from Eq. (G.40) and the trace definition

$$\mathrm{tr}[\bar{\mathbf{R}}_{\mathbf{s}}] = \sum_{k=1}^{N}\bar{r}_{s,\,kk} \tag{G.45}$$

The solution to this optimization problem may be determined through the *water-filling procedure*, which is well known in information theory.[3] Effectively, the solution to the water-filling problem says that, in a multiple-channel scenario, we transmit more signal power in the better channels and less signal power in the poorer channels. To be specific, imagine a vessel whose bottom is defined by the set of N dimensionless discrete levels

$$\left\{\frac{\mu-(\sigma_{\mathbf{w}}^2/\lambda)}{\lambda_k}\right\}_{k=1}^{N}$$

and pour "water" into the vessel in an amount corresponding to the total transmit power P. That power is optimally divided among the N eigenmodes of the MIMO link in accordance with their corresponding "water levels" in the vessel, as illustrated in Fig. G.1 for a MIMO link with $N = 6$. The "water-fill level", denoted by the dimensionless parameter μ and indicated by the dashed line in the figure, is chosen to satisfy the constraint of Eq. (G.44). On the basis of the spatially discrete water-filling picture portrayed in Fig. G.1, we may now finally postulate the optimal $\bar{r}_{s,\,kk}$ to be

$$\bar{r}_{s,\,kk} = \left(\mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_k}\right)^+ \qquad k = 1,2,\ldots,N \tag{G.46}$$

where the superscript "+" signifies retaining only those terms on the right-hand side of the equation that are positive (i.e., the terms that pertain to those eigenmodes of the MIMO link for which the water levels lie below the constant μ). Correspondingly, the maximum value of the capacity of the MIMO link, in accordance with the first line of Eq. (G.43) and Eq. (G.46), is defined by

$$\begin{aligned} C &= \sum_{k=1}^{N} \log_2\left(1 + \frac{1}{\sigma_{\mathbf{w}}^2}\lambda_k \bar{r}_{s,\,kk}\right) \\ &= \sum_{k=1}^{N} \log_2\left\{1 + \frac{1}{\sigma_{\mathbf{w}}^2}\lambda_k\left(\mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_k}\right)^+\right\} \\ &= \sum_{k=1}^{N} \log_2\left(\frac{\mu\lambda_k}{\sigma_{\mathbf{w}}^2}\right)^+ \end{aligned} \tag{G.47}$$

where, as stated previously, the constant μ is chosen to satisfy the global power constraint of Eq. (G.44).

The optimal results of Eqs. (G.46) and (G.47), assuming that the channel state is known to both the transmitter and receiver, were derived by considering a MIMO link with $N_r = N_t = N$.

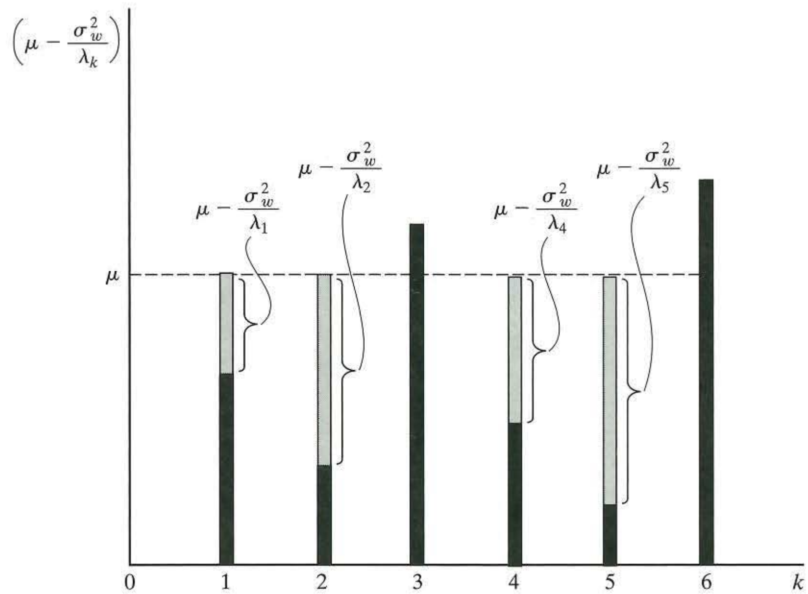


FIGURE G.1 Water-filling interpretation of the optimization procedure. For the example portrayed in the figure, we have the following source autocorrelation values

$$\bar{r}_{s,11} = \mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_1}$$

$$\bar{r}_{s,22} = \mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_2}$$

$$\bar{r}_{s,33} = 0$$

$$\bar{r}_{s,44} = \mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_4}$$

$$\bar{r}_{s,55} = \mu - \frac{\sigma_{\mathbf{w}}^2}{\lambda_5}$$

$$\bar{r}_{s,66} = 0$$

All the other nondiagonal elements of the source correlation matrix $\overline{\mathbf{R}}_{\mathbf{s}}$ are zero.

NOTES AND REFERENCES

[1] For a detailed exposition of the many facets of information theory, see Cover and Thomas (1991).

[2] The first detailed derivation of the log-det capacity formula for a stationary MIMO link was presented by Telatar in an AT&T technical memorandum published in 1995 and republished as a journal paper in 1999.

[3] The waterfilling procedure is described in Cover and Thomas (1991).

A P P E N D I X H

Eigendecomposition

H.1 UNITARY TRANSFORMATION OF A HERMITIAN MATRIX[1]

Consider a square complex matrix $\mathbf{R}$ of dimensions M by M. The matrix $\mathbf{R}$ is assumed to be Hermitian; that is,

$$\mathbf{R}^{\dagger} = \mathbf{R} \tag{H.1}$$

where the superscript $\dagger$ denotes Hermitian transposition. With $\mathbf{R}$ as the matrix of interest, the *eigenvalue problem* is defined by

$$\mathbf{R}\mathbf{q} = \lambda\mathbf{q} \tag{H.2}$$

where $\mathbf{q}$ is an M-by-1 vector and λ is a scalar.

In general, there are M distinct values of the scalar λ that satisfy Eq. (H.2); these values are roots of the *characteristic equation*

$$\det(\mathbf{R} - \lambda\mathbf{I}) = 0 \tag{H.3}$$

where $\mathbf{I}$ is the M-by-M identity matrix.

Typically, the off-diagonal elements of matrix $\mathbf{R}$ are nonzero. The *diagonalization* of $\mathbf{R}$ is achieved by expanding on the transformation described in Eq. (H.2). Specifically, we may write

$$\mathbf{Q}^{\dagger}\mathbf{R}\mathbf{Q} = \mathbf{\Lambda} \tag{H.4}$$

where

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M) \tag{H.5}$$

is a *diagonal matrix* and

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M] \tag{H.6}$$

is a *unitary matrix*. The scalars $\lambda_1, \lambda_2, \ldots, \lambda_M$ constituting the matrix $\mathbf{\Lambda}$ are called the *eigenvalues* of matrix $\mathbf{R}$, and the M-by-1 vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M$ constituting the matrix $\mathbf{Q}$ are the associated *eigenvectors* of $\mathbf{R}$.

By definition, the unitary matrix $\mathbf{Q}$ satisfies the relation

$$\mathbf{Q}\mathbf{Q}^{\dagger} = \mathbf{Q}^{\dagger}\mathbf{Q} = \mathbf{I} \tag{H.7}$$

In expanded form, we may rewrite Eq. (H.7) as

$$\mathbf{q}_i^{\dagger}\mathbf{q}_k = \begin{cases} 1 & \text{for } k = i \\ 0 & \text{for } k \neq i \end{cases} \tag{H.8}$$

According to Eq. (H.7), the inverse of the matrix $\mathbf{Q}$, namely, $\mathbf{Q}^{-1}$, is equal to the Hermitian transpose of $\mathbf{Q}$, or

$$\mathbf{Q}^{-1} = \mathbf{Q}^{\dagger} \tag{H.9}$$

In light of Eqs. (H.5) through (H.9), we may rewrite Eq. (H.4) in the equivalent form

$$\begin{aligned} \mathbf{R} &= \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\dagger} \\ &= \sum_{k=1}^{M} \lambda_k \mathbf{q}_k \mathbf{q}_k^{\dagger} \end{aligned} \tag{H.10}$$

Equation (H.10) is called the *spectral decomposition theorem*, which states that the *Hermitian matrix* $\mathbf{R}$ *can be expanded as the linear combination of the rank-one matrix products* $\left\{\mathbf{q}_k\mathbf{q}_k^{\dagger}\right\}_{k=1}^{M}$*, and the corresponding eigenvalues* $\{\lambda_k\}_{k=1}^{M}$ *are the scaling factors of the linear combination.*

H.2 RELATIONSHIP BETWEEN EIGENDECOMPOSITION AND SINGULAR-VALUE DECOMPOSITION[2]

Consider next a rectangular complex matrix $\mathbf{A}$ with dimensions L by M. Let the M-by-M matrix $\mathbf{R}$ be related to the matrix $\mathbf{A}$ as follows:

$$\mathbf{R} = \begin{cases} \mathbf{A}\mathbf{A}^{\dagger} & \text{for } M \geq L \\ \mathbf{A}^{\dagger}\mathbf{A} & \text{for } M < L \end{cases} \tag{H.11}$$

Then, according to the *singular-value decomposition (SVD) theorem*, the matrix $\mathbf{A}$ may be diagonalized as

$$\mathbf{U}^{\dagger}\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{H.12}$$

where **D** is a diagonal matrix, the **0**'s are null matrices, and **U** and **V** are respectively L-by-L and M-by-M unitary matrices; that is,

$$\mathbf{U}^{\dagger} = \mathbf{U}^{-1} \tag{H.13}$$

and

$$\mathbf{V}^{\dagger} = \mathbf{V}^{-1} \tag{H.14}$$

Specifically, we may make the following statements:

- The diagonal matrix

$$\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_W), \quad W = \min(L,M) \tag{H.15}$$

  defines the *singular values* of matrix **A**.
- The unitary matrix

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L] \tag{H.16}$$

  defines the L *left-singular vectors* of matrix **A**.
- The second unitary matrix

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M] \tag{H.17}$$

  defines the M *right-singular vectors* of matrix **A**.

Moreover, depending on whether the dimension L is greater than M or the other way around, we have two different cases in describing the relationships between singular-value decomposition and eigendecomposition:

Case 1. $L > M$
In this case, the dimension $W = M$ and the singular values $d_1, d_2, \ldots, d_M$ are equal to the square roots of the eigenvalues of the matrix product $\mathbf{R} = \mathbf{A}^{\dagger}\mathbf{A}$. Correspondingly, the right-singular vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M$ are the associated eigenvectors.

Case 2. $L < M$
In this second case, the dimension $W = L$ and the singular values $d_1, d_1, \ldots, d_L$ are equal to the square roots of the eigenvalues of the alternative matrix product $\mathbf{R} = \mathbf{A}\mathbf{A}^{\dagger}$. Correspondingly, the left-singular vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$ are the associated eigenvectors.

NOTES AND REFERENCES

[1] The eigendecomposition of a square matrix is discussed in Chapter 5 of Strang (1980). The discussion presented therein focuses on square matrices that are real.

[2] The singular-value decomposition of a rectangular matrix is discussed in Chapter 7 of Strang (1980). Here again, the discussion focuses on real matrices. The chapter also discusses issues relating to the computation of eigenvalues.

APPENDIX I

Adaptive Array Antennas

I.1 NEED FOR ADAPTIVITY

The goal of wireless communications is to allow as many users as possible to communicate reliably without regard to location and mobility. From the discussion presented in Chapter 2, we find that this goal is seriously impeded by three major channel impairments:

1. *Multipath* can cause severe fading due to phase cancellation between different propagation paths. Fading leads to a reduction in available signal power and therefore a degraded noise performance at the receiver.
2. *Delay spread* results from differences in propagation delays among the multiple propagation paths. When the delay spread exceeds about 10% of the symbol duration, the intersymbol interference experienced by the received signal reaches a significant level, thereby causing a reduction in the attainable data rate.
3. *Cochannel interference* arises in cellular systems in which the available frequency channels are divided into different sets, each of which is assigned to a specific cell and with several cells in the system using the same set of frequencies. Cochannel interference limits the *system capacity* (i.e., the largest possible number of users that can be reliably served by the system).

Typically, cellular systems use 120° sectorization at each base station, and only one user accesses a sector of a base station at a given frequency. We may combat the effects of multipath fading and cochannel interference at the base station by using three identical, but separate, *antenna arrays*, one for each sector of the base station. (The compensation of delay spread is considered later in the section.) Figure I.1 shows the block diagram of an *array signal processor*; it is assumed that there are N users whose signals are received at a particular sector of the base station and that the array for that sector consists of K identical antenna elements. A particular user is treated as the one of interest, and the remaining $N-1$ users give rise to cochannel interference. In addition to the cochannel interference, each component of the array signal processor's input is corrupted by additive white Gaussian noise (AWGN). The analysis presented herein is

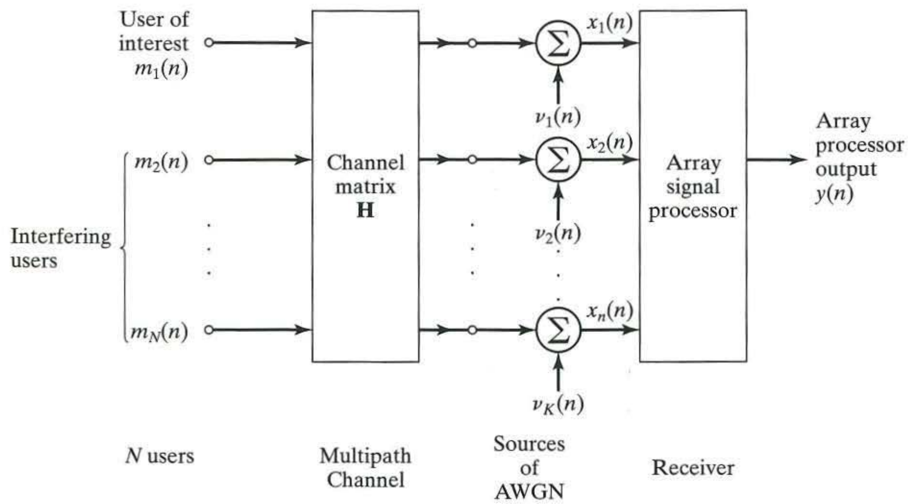


FIGURE I.1 Block diagram of array signal processor that involves K antenna elements and that is being driven by a multipath channel.

for baseband signals, which, in general, are complex valued. This means that both the channel and the array signal processor require complex characterizations of their own. The structure depicted in Fig. I.1 is drawn for one output pertaining to the user of interest. The array signal processor is duplicated for users at other frequencies at the base station.(Figure I.1 refers to a situation different from that considered in Chapter 6 and Appendix G, hence the difference in notation.)

The multipath channel is characterized by the channel matrix, which is denoted by $\mathbf{H}$. The matrix $\mathbf{H}$ has dimensions K-by-N and may therefore be expanded into N column vectors, as shown by:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N] \tag{I.1}$$

Column vector $\mathbf{h}_i$, $i = 1,2,\ldots,N$ is of dimension K, and represents the multipath component pertaining to user i.

Given the configuration described in Fig. I.1, the goal is to design a *linear array signal processor* for the receiver that satisfies two requirements:

1. The cochannel interference produced by the $N-1$ interfering users is cancelled.
2. The output signal-to-noise ratio (SNR) for the user of interest is maximized.

Hereafter, these two requirements are referred to as design requirements 1 and 2.

To proceed with this design task, it is assumed that the multipath channel is described by flat Rayleigh fading. Then, in light of the material presented in Example 6.2, we find that the use of diversity permits the treatment of the column vectors $\mathbf{h}_2, \mathbf{h}_3, \ldots, \mathbf{h}_N$ as *linearly independent*, which is justified, provided that the spacing between antenna elements of the array is large enough (10 to 20 times the wavelength) for independent fading. To simplify the presentation, we suppose that user 1 is the user of interest and the remaining $N-1$ users are responsible for co-channel interference, as indicated in Fig. I.1. The key design issue is how to find the *weight vector*, denoted by $\mathbf{w}$, that characterizes the array signal processor. Toward that end, we may proceed as follows:

1. We choose the K-dimensional weight vector $\mathbf{w}$ to be orthogonal to the vectors $\mathbf{h}_2, \mathbf{h}_3, \ldots, \mathbf{h}_N$, which are associated with the interfering users. This choice fulfills design requirement 1 (i.e., the cancellation of cochannel interference).
2. To satisfy design requirement 2 (i.e., maximization of the SNR), we will briefly digress from the issue at hand to introduce the notion of a subspace. Given a *vector space*, or just a *space*, formed by a set of linearly independent vectors, a *subspace* of the space is a subset that satisfies two conditions:
   (i) If we add any two vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ in the subspace, their sum $\mathbf{z}_1$ and $\mathbf{z}_2$ is still in the subspace.
   (ii) If we multiply any vector $\mathbf{z}$ in the subspace by any scalar a, the multiple $a\mathbf{z}$ is still in the subspace.

   Define the subspace $\mathcal{W}\perp\{\mathbf{h}_2, \mathbf{h}_3, \ldots, \mathbf{h}_N\}$. Then, returning to the issue of how to maximize the output SNR for user 1, we first construct a subspace denoted by $\mathcal{W}$ whose dimension is equal to the difference between the number of antenna elements and the number of interfering users—that is, $K-(N-1)=K-N+1$. Next, we project the complex conjugate of the channel vector $\mathbf{h}_1$ (pertaining to user 1) onto the subspace $\mathcal{W}$. The projection so computed defines the weight vector $\mathbf{w}$.

EXAMPLE I.1 Subspace Method for Determining the Weight Vector

To illustrate the two-step subspace method for determining the weight vector $\mathbf{w}$, consider the simple example of a system involving two users characterized by the channel vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, and an antenna array consisting of three elements; that is, $N=2$ and $K=3$. Then, for this example, the subspace $\mathcal{W}$ is two-dimensional, since

$$K-N+1=3-2+1=2$$

With user 1 viewed as the user of interest and user 2 viewed as the interferer, we may construct the signal-space diagram shown in Fig. I.2. The subspace $\mathcal{W}$, shown shaded in this figure, is orthogonal to channel vector $\mathbf{h}_2$. The weight vector $\mathbf{w}$ of the array signal processor is determined by the projection of the complex-conjugated channel vector of user 1 (i.e., $\mathbf{h}_1^*$) onto the subspace $\mathcal{W}$, as depicted in the figure.

The important conclusion drawn from this discussion is that a linear receiver using optimum combining with K antenna elements and involving $N-1$ interfering users has the same performance as a linear receiver with $K-N+1$ antenna elements, without interference, independent of the multipath environment. For this equivalence to be realized, we of course require that $K > N-1$. Provided that this condition is satisfied, the receiver cancels the cochannel interference with a diversity improvement equal to $K-N+1$, which represents an N- fold increase in system capacity.

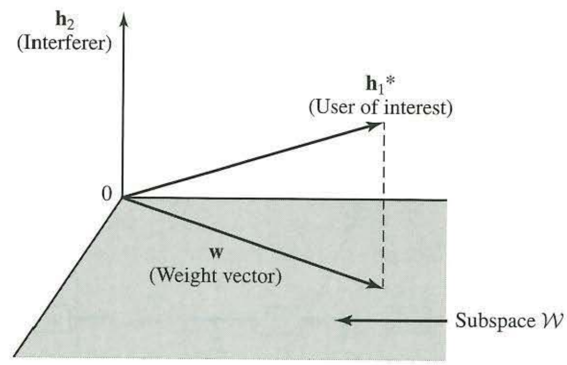


FIGURE I.2 Signal-space diagram for Example I.1, involving a user of interest, a single interferer, and an antenna array of three elements. The subspace $\mathcal{W}$, shown shaded, is two dimensional in this example.

The design of an array signal processor in accordance with the two-step subspace procedure described herein is of the *zero-forcing* kind. We say this because, given K antenna elements, the array has enough degrees of freedom to *force* the output due to the $N-1$ interfering users, represented by the linearly independent channel vectors $\mathbf{h}_2, \ldots, \mathbf{h}_K$, to zero so long as K is greater than $N-1$. Note also that this procedure includes $N=1$ (i.e., a single user with no interfering users) as a special case. In this case, the channel matrix consists of vector $\mathbf{h}_1$ that lies in the subspace $\mathcal{W}$, and the zero-forcing solution $\mathbf{w}$ equals $\mathbf{h}_1$*. ■

The analysis presented thus far has been entirely of a *spatial* kind that ignores the effect of delay spread. What if the delay spread is significant compared with the symbol duration, and cannot, therefore, be ignored? Recognizing that delay-spread is responsible for intersymbol interference, we may incorporate a *linear* equalizer in each antenna branch of the array to compensate for delay spread. The resulting array signal processor takes the form shown in Fig. I.3, which combines temporal and spatial processing. Spatial processing is provided by the antenna array, and the temporal processing is provided by a bank of finite-duration impulse response (FIR) filters. For obvious reasons, this structure is called a *space–time processor*.

I.1.1 Adaptive Antenna Arrays[1]

The subspace design procedure for the array signal processor in Fig. I.1 assumes that the channel impairments are stationary and that we have knowledge of the channel matrix **H**. In reality, however, multipath fading, delay spread, and cochannel interference are all *nonstationary* in their own individual ways. Also, the channel characterization may be unknown. To deal with these practical issues, we need to make the receiving array signal processor in Fig. I.1 *adaptive*. Bearing in mind the scope of this book, we confine the discussion to adaptive spatial processing, assuming that the delay spread is negligible. We further assume that the multipath fading phenomenon is slow enough to justify the *least-mean-square (LMS) algorithm* to perform the adaptation.

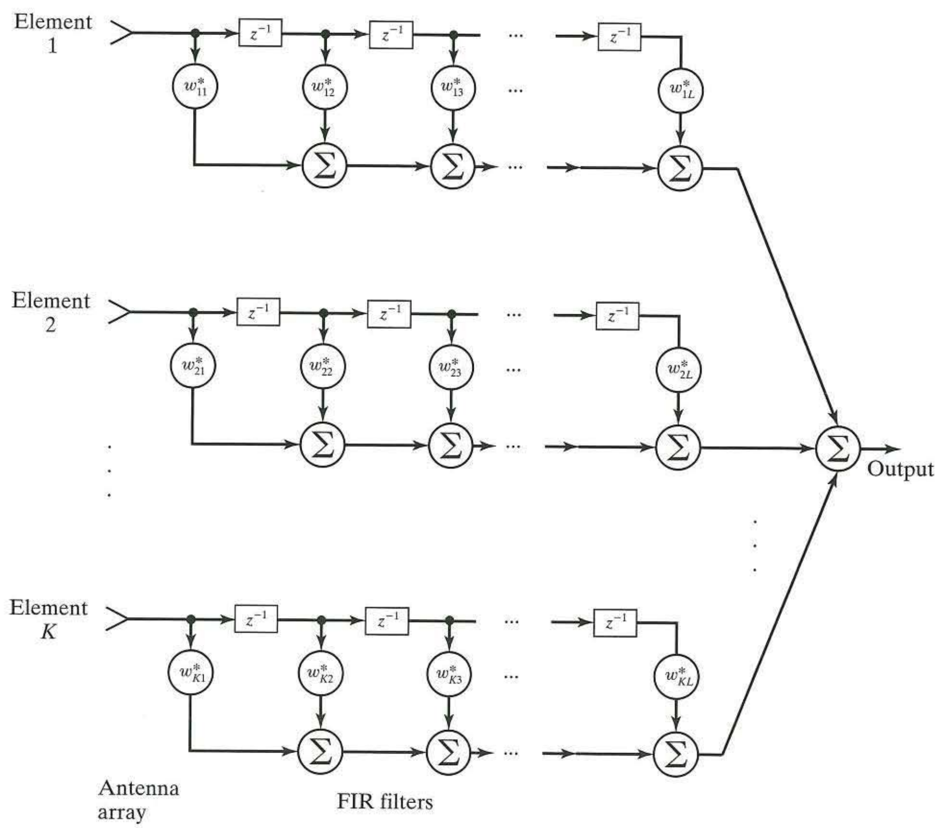


FIGURE I.3 Baseband space–time processor. The blocks labeled z^{-1} are unit-delay elements, with each delay being equal to the symbol period. The filter coefficients are complex valued. The FIR filters are all assumed to be of length L.

I.1.2 Least-Mean-Square (LMS) Algorithm

Figure I.4 shows the structure of an *adaptive antenna array*, in which the output of each antenna element is multiplied by an adjustable (controllable) weight w_k, $k = 1,2,...,K$, and then the weighted elemental outputs of the array are summed to produce the array output signal, denoted by y. The adaptive antenna array does not require knowledge of the direction of arrival of the desired signal originating from a user of interest, so long as the system is supplied with a *reference signal*, which is *correlated* with the desired signal. For example, the reference signal could correspond to a training sequence that is transmitted on a periodic basis. The output signal of the array is subtracted from the reference signal, denoted by d, to generate an *error signal* e, which is used to apply the appropriate adjustments to the elemental weights of the array. In this way, a feedback system to control the elemental weights is built into the operation of the antenna array, thereby making it adaptive to changes in the environment. Note that the block diagram is drawn for baseband processing. In a practical system, a quadrature hybrid is

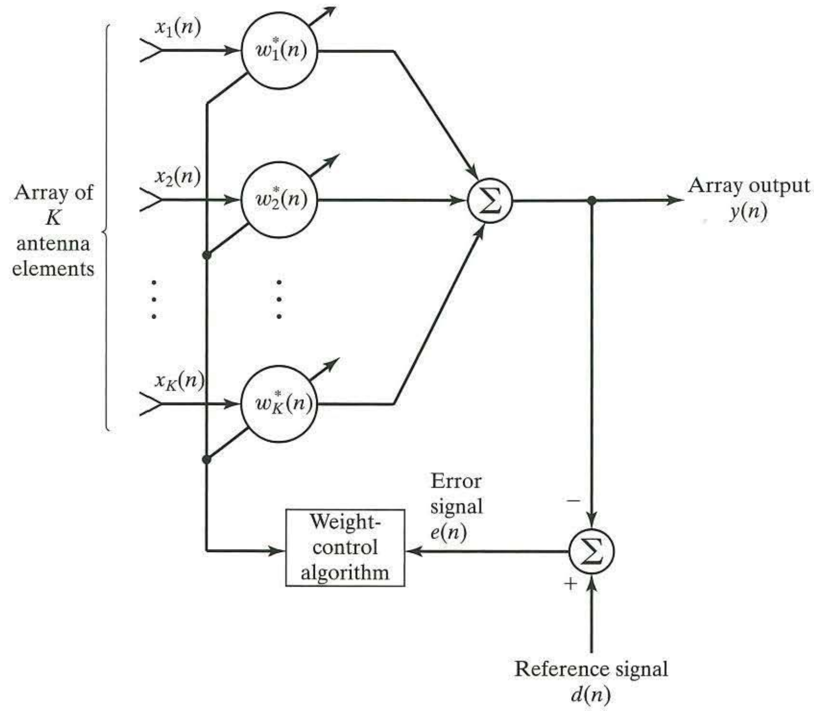


FIGURE I.4 Block diagram of adaptive antenna array.

used for each antenna element of the array to split the complex-valued received signal at each element into two components: one real and the other imaginary. The use of a hybrid has been omitted from the figure, to simplify the diagram.

Let $x_k(n)$ denote the output of the kth element in the array at discrete time n, and let $w_k(n)$ denote the corresponding value of the weight connected to this element. Then the output signal of the array (consisting of K antenna elements) is

$$y(n) = \sum_{k=1}^{K} w_k^*(n)x_k(n) \tag{I.2}$$

where $w_k^*(n)x_k(n)$ is the inner product of the complex-valued quantities $w_k(n)$ and $x_k(n)$. Denoting the reference signal as $d(n)$, we may evaluate the *error signal* as

$$e(n) = d(n) - y(n) \tag{I.3}$$

To optimize the performance of the adaptive antenna array, it is customary to use the *mean-square error*

$$J = E[|e(n)|^2] \tag{I.4}$$

as the cost function to be *minimized*. Minimization of the cost function J tends to suppress the interfering signals and thereby enhance the desired signal in the array output. The LMS algorithm minimizes the instantaneous value of the cost function J and, through successive iterations, approaches the minimum mean-square error (MMSE) (i.e., the optimum solution for the elemental weights) ever more closely. An adaptive antenna array based on the minimum mean-square error criterion is highly likely to provide a better solution than one based on the zero-forcing criterion embodied in the two-step subspace method.

The adjustment applied to the kth elemental weight is

$$\Delta w_k(n) = \mu e^*(n) x_k(n) \qquad k = 1, 2, \cdots, K \tag{I.5}$$

where μ is the *step-size parameter*. The updated value of this weight is

$$w_k(n+1) = w_k(n) + \Delta w_k(n) \qquad k = 1, 2, \cdots, K \tag{I.6}$$

Equations (I.2), (I.3), (I.5) and (I.6), in that order, constitute the *complex LMS algorithm*.[2] The algorithm is initiated by setting $w_k(0) = 0$ for all k.

The advantages of an adaptive antenna array using the complex LMS algorithm are threefold:

1. Simplicity of implementation
2. Only a linear growth in complexity with the number of antenna elements
3. Robust performance with respect to disturbances.

However, the system suffers from the following drawbacks:

- A slow rate of convergence, which is typically 10 times the number of adjustable weights. This limits the use of the complex LMS algorithm to a slow-fading environment, for which the Doppler spread is small compared with the reciprocal of the duration of the observation interval.
- Sensitivity of the convergence behavior to variations in the reference signal and cochannel interference powers.

These limitations of the complex LMS algorithm can be overcome by using an algorithm known as *direct matrix inversion* (DMI).[3] Unlike the LMS algorithm, the DMI algorithm operates in *batch* mode, in that the computation of the elemental weights is based on a batch of L snapshots. The batch size L is chosen as a compromise between two conflicting requirements:

- The size L should be small enough for the batch of snapshots used in the computation to be justifiably treated as pseudostationary.
- The size L should be large enough for the computed values of the elemental weights to approach the MMSE solution.

The DMI algorithm is the optimum combining technique for array antennas deployed in many base stations today. The algorithm may be reformulated for recursive computation if desired.[4]

When the teletraffic is high, the base stations are ordinarily configured as microcells, which are small cells such as an office floor or a station deployed along a highway with directional antennas. In such a configuration, there are many inexpensive base stations in close proximity to each other. The use of adaptive antenna arrays provides the means for an alternative configuration in which there are fewer (but more expensive) base stations, further apart from each other than in the corresponding microcellular system.

NOTES AND REFERENCES

[1] For a discussion of adaptive antenna arrays and their theory, design, and applications, see Compton (1988).

[2] The least-mean-square (LMS) algorithm is discussed Haykin (2002) and Widrow and Stearns (1985).

[3] The direct matrix inversion (DMI) algorithm, also referred to as the sample matrix inversion method, is discussed in Compton (1988); see pp. 331–332.

[4] The recursive least-squares (RLS) algorithm provides an iterative method for implementing the method of least squares, which lies behind the DMI; for details, see Haykin (2002).

Bibliography

Abramowitz, M. and I. Stegun, eds. *Handbook of Mathematical Functions.* New York: Dover, 1964.

Abramson, N. "The ALOHA system—Another alternative for computer communications," in 1970 Fall Joint Compt. Conf., *AFIPS Conf. Proc.*, vol. 37. Montvale, NJ: AFIPS Press, 1970, pp. 281–285.

Abramson, N. (ed.). *Multiple Access Communications*, New York: IEEE Press, 1993.

Alamouti, S. "A simple transmitter diversity scheme for wireless communications." *IEEE J. Selected Areas in Communications*, Vol. 16, pp. 1451–1458, 1998.

Anderson, J.B., T. Aulin, and C.E. Sundberg. *Digital Phase Modulation*. New York: Plenum Press, 1986.

Anderson, J.B., *Digital Transmission Engineering*. IEEE Press, 1999.

ANSI/IEEE Std 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1999 edition.

Bahai, A.R.S. and Saltzberg, B.R. *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, Kluwer Academic/Plenum Publishers, 1999.

Bahl, L., J. Cocke, F. Jelinek, and J. Raviv. "Optimal decoding of linear codes for minimizing symbol error rate." *IEEE Trans. Inform. Theory*, Vol. IT-20, 284–287, 1974.

Barker, R.H. "Group synchronization of binary digital systems." pp. 273–2876, in W. Jackson (ed.), *Communication Theory.* New York: Academic Press, 1953.

Benedetto, J.S. and G. Montrosi. "Unveiling turbo codes: Some results on parallel concatenated coding schemes." *IEEE Trans. Inform. Theory*, Vol. 42, pp. 409–428, March 1996.

Benedetto, S. and E. Biglieri. *Principles of Digital Transmission with Wireless Applications.* New York: Kluwer Academic/Plenum Publishers, 1999.

Berrou, C. "The ten-year old turbo codes are entering into service." *IEEE Communications Magazine*, pp. 110–116, August 2003.

Berrou, C. and A. Glavieux. "Near optimum error correcting coding and decoding: Turbo-codes." *IEEE Trans. Commun.*, Vol. 44, pp. 1261–1271, 1996.

Berrou, C., A. Glavieux, and P. Thitimajshima. "Near Shannon-limit error correction coding and decoding: Turbo codes," in *Proc. 1993 International Conference on Communications*, pp. 1064–1070, Geneva, Switzerland, May 1993.

Bertsekas, D. and R. Gallager. *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1992.

Blachman, N.M. *Noise and its Effect on Communication*, McGraw-Hill, 1966.

Blackard, K.L. and T.S. Rappaport. "Measurements and models of radio frequency impulse noise for indoor wireless communications." *IEEE J. Selected Areas in Communications*, Vol. 11, pp. 991–1001, Sept. 1993.

Boutros, J., N. Gresset, and L. Brunel. "Turbo coding and decoding for multiple antenna channels." *International Symposium on Turbo Codes, Proceedings*, pp. 185–86, Brest, France, Sept. 2003.

Bracewell, R.N. *The Fourier Transform and Its Applications*, 2nd ed., rev. New York: McGraw-Hill, 1986.

Brennan, D.G. "Linear diversity combining techniques." *Proceedings of the IRE*, Vol. 47, pp. 1075–1102, June 1959.

Caire, G., G. Taricco, and E. Biglieri. "Bit-interleaved coded modulation." *IEEE Trans. Inform. Theory*, Vol. 44, pp. 927–946, 1998.

Carlson, A.B. *Communication Systems*, 2nd ed. New York: McGraw-Hill, 1975.

Cassioli, D., M.Z. Win, and A.F. Molisch. "The Ultra-Wide Bandwidth Indoor Channel: From Statistical Model to Simulations," *IEEE J. Selected Areas in Commun.*, Vol. 20, pp. 1247–1257, 2002.

Chennakeshu, S. and G.J. Saulnier. "Differential detection of $\pi/4$-shifted-DQPSK for digital cellular radio." *IEEE Trans. Vehicular Technology*, Vol. 42, pp. 46–57, 1993.

Cheung, K.W., J.H.M. Sau, and R.D. Murch. "A new empirical model for indoor propagation prediction." *IEEE Trans. Vehicular Technology*, 47(3), pp. 996–1001, August 1998.

Chizhik, D., G. Foschini, and R.A. Valenzuali. "Capacities of multi-element transmit and receive antennas: Correlations and keyholes." *Electronics Letters*, Vol. 36, pp. 1099–1100, 2000.

Chugg, K., A. Anastasopoulos, and X. Chen. *Iterative Detection: Adaptivity, complexity, and applications*. Boston: Kluwer, 2001.

Clarke, R.H. "A statistical theory of mobile radio reception." *Bell System Tech. J.*, Vol. 47, pp. 957–1000, 1968.

Clarke, G.C., and Cain, J.B., *Error-Correction Coding for Digital Communications*. New York: Plenum, 1981.

Compton, R.T. *Adaptive Antennas: Concepts and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

Cover, T.M. and J.A. Thomas. *Elements of Information Theory*. New York: Wiley, 1991.

deBuda, R. "Coherent demodulation of frequency-shift keying with low deviation ratio." *IEEE Trans. Communications*, Vol. COM-20, pp. 429–435, 1972.

deHaas, A. *Radio-Nieuws*, Vol. 10, pp. 357-364, December 1927, and Vol. 11, pp. 80-88, February 1928; portions of these two papers were translated by B.B. Barrow, under the title "Translation of a historic paper on diversity reception." *Proc. IRE*, Vol. 49, pp. 367–369, January 1961.

deJager, F. and C.B. Dekker. "Tamed frequency modulation—A novel method to achieve spectrum economy in digital transmission." *IEEE Trans. Communications*, Vol. 26, pp. 534–542, May 1978.

Diggavi, S.N., N. Al-Dhahir, A. Stanoulis, and A.R. Calderbank. "Differential space–time coding for frequency on selective channels." *IEEE Communications Letters*, Vol. 6, pp. 253–255, 2002.

Diggavi, S.N., N. Al-Dhahir, A. Stanoulis, and A.R. Calderbank. "Great expectations: The value of spatial diversity in wireless networks." *Proceedings of the IEEE*, Vol. 91, 2003.

Diggavi, S.N. "Role of spatial diversity in wireless networks," Workshop on New Directions for Statistical Signal Processing in the 21st Century," Lake Louise, Alberta, Canada, 5–10 October, 2003.

Divsalar, D., and F. Pollara. "Turbo codes for PCS applications." *Proceedings of International Communications Conference, Seattle, Washington*, pp. 54-59, June 1995.

Dixon, R.C. *Spread Spectrum Systems with Commercial Applications*, 3rd ed. New York: Wiley, 1994.

Doelz, M.I. and E.H. Heald. Minimum shift data communication system, U.S. Patent 2977417, March 1961.

Elias, P. "Coding for noisy channels." *IRE Convention Record, Part 4*, pp. 37–46, March 1955.

Ericson, T. "A Gaussian channel with slow fading." *IEEE Trans. Inform. Theory*, Vol. 47, pp. 2321–2334, 2001.

Forney, G.D. Jr. "The Viterbi algorithm." *Proceedings of the IEEE*, Vol. 61, pp. 268–278, 1973.

Foschini, G.J. "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas." *Bell Labs Technical J.*, Vol. 1, No. 2, pp. 41–59, 1996.

Foschini, G.J., and M.J. Gans. "On limits of wireless communications in a fading environment when using multiple antennas." *Wireless Personal Communications*, Vol. 6, pp. 311–335, 1998.

Foschini, G.J., D. Chizkik, M.J. Gans, C. Papadies, and R.A. Valenzueli. "Analysis and performance of some basic space–time architectures." *IEEE J. Select. Areas Communications*, Vol. 21, pp. 303–320, 2003.

Foschini, G.J., G.D. Golden, R.A. Valenzuela, and P.W. Wolniansky. "Simplified processing for wireless communication at high spectral efficiency." *IEEE J. Selected Areas in Communications*, Vol. 17, pp. 1841–1852, 1999.

Gallager, R.G. *Information Theory and Reliable Communications*. New York: John Wiley and Sons, 1968.

Garg, V.K. *IS-95 CDMA and cdma 2000*. Upper Saddle River, NJ: Prentice Hall, 2000.

Gesbert, D., H. Bolseskei, D. Gore, and A. Paulraj. "Outdoor MIMO wireless channels: Models and performance prediction." *IEEE Trans. Communications*, Vol. 50, pp. 225–234, 2002.

Gesbert, D., M. Shafi, D.S. Shiu, P. Smith, and A. Naguib. "From theory to practice: An overview of MIMO space–time coded wireless systems." *IEEE J. Selected Areas in Communications*, Vol. 21, pp. 281–302, 2003.

Gold, R. "Optimal binary sequences for spread spectrum multiplexing." *IEEE Trans. Inform. Theory*, Vol. IT-14, pp. 154–156, 1968.

Golden, G.D., J.G. Forschini, R.A. Valenzuela, and P.W. Woliansky. "Detection algorithm and initial laboratory results using V-BLAST space–time communication architecture." *Electronics Letters*, Vol. 35, pp. 14–15, 1999.

Goldsmith, A., S.A. Jafar, N. Jindal, and S. Vishwaneth. "Fundamental capacity of MIMO channels." *IEEE J. Selected Areas in Communications*, Vol. 21, pp. 684–702, 2003.

Hagenauer, J. "The turbo principle: Tutorial introduction and state of the art," International Symposium Turbo Codes. Brest, France, Sept. 1997.

Hanzo, L., T. Liew, and B. Yeap. *Turbo Coding, Turbo Equalization and Space–Time Coding*. Hoboken, NJ: Wiley, 2002.

Hata M., "Empirical formal for propagation loss in land mobile radio services," *IEEE Trans. Vehic. Technol.*, Vol. 29, August 1980.

Haykin, S. *Adaptive Filter Theory*, 4th ed., Upper Saddle River, NJ: Prentice Hall, 2002.

Haykin, S. *Communication Systems*. 4th ed. New York: John Wiley, 2001.

Heegard, C. and S.B. Wicker. *Turbo Coding*. Boston: Kluwer Academic Publishers, 1999.

Holma, H. and A.Toskala (eds.). *WCDMA for UMTS*. New York: Wiley, 2000.

Holmes, J.K. *Coherent Spread Spectrum Systems.* Melbourne, FL: 1990.

Hochwald, B.M., T.L. Marzetta, and V. Tarokh. "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," submitted to *IEEE Trans. Inform. Theory.*

Hottinen, A., O. Tirkkonen, and R. Wichman. *Multi-antenna Transceiver Techniques for 3G and Beyond*. Hoboken, NJ: Wiley, 2003.

ITU Recommendations. The ITU publishes a variety of recommendations regarding propagation models, interference models, and other topics of interest to radio engineers.

Jakes, W.C. (ed.). *Microwave Mobile Communications.* New York: Wiley, 1974. Reprinted by IEEE Press, 1994.

Kailath, T. *Sampling models for linear time-variant filters*, Technical Report 352, MIT RLE, Cambridge, MA, May 25, 1959.

Kaplan, E.D. (ed.). *Understanding GPS: Principles and Applications.* Boston: Artech House, 1996.

Lee, J.S. and L.E. Miller. *CDMA Systems Engineering Handbook*. Boston: Artech House, 1998.

Lee, W.C.Y. *Mobile Communications Engineering*. New York: McGraw-Hill, 1982.

Leon-Garcia, A. *Probability and Random Processes*, 2nd ed., Reading, MA: Addison-Wesley, 1994.

Liberti, J.C., Jr. and T.S Rappaport. *Smart Antennas for Wireless Communications.* Upper Saddle River, NJ: Prentice Hall, 1999.

Liew, T.H. and L. Hanzo. "Space–time codes and concatenated channel codes for wireless communications." *Proceedings of the IEEE*, Vol. 90, pp. 187–219, 2002.

Lin, S., and D.J. Costello, Jr. *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.

Lodge, J.H. and M.L. Moher. "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels." *IEEE Trans. Communications*, Vol. 38, pp. 787–794, 1990.

McEliece, R. *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.

Médard, M. *The Capacity of Time Varying Multiple User Channels in Wireless Communications.* Doctor-of-Science Thesis in Electrical Engineering, MIT, Cambridge, MA, Sept. 1995.

Mehrotra, A. *GSM System Engineering*. Boston: Artech House, 1997.

Michelson, A.M. and A.H. Levesque. *Error-control Techniques for Digital Communication*. New York: Wiley, 1985.

Moher, M. "An iterative multiuser decoder for near capacity communications." *IEEE Trans. on Communications*, Vol. 46, pp. 870–880, July 1998.

Moher, M.L. and J.H. Lodge. "TCMP–a modulation and coding strategy for Rician fading channels." *IEEE J. Selected Areas Communications*, Vol.7, pp. 1347–1355, 1989.

Monzingo, R.A. and T.W. Miller. *Introduction to Adaptive Arrays.* New York: Wiley, 1980.

Nakahima, N., et al. "A system design for TDMA mobile radios." *Proc. 40th IEEE Vehicular Technology Conference*, pp. 295–298, May 1990.

Noble, D.E. "The history of land-mobile radio communications." *Proceedings of the IRE*, Section 28 Vehicular Communications, pp. 1405–1414, May 1962.

Nyquist, H. "Thermal agitation of electric charge in conductors." *Physical Review*, second series, Vol. 32, pp. 110–113, 1928.

O'Hara, B. and A. Petrick. *IEEE 802.11 Handbook: A Designer's Companion*. New York: IEEE Press, 1999.

Okumura, Y., Ohmori, E., and Fukuda, K., "Field strength and its variability in VHF and UHF land mobile radio services," *Review of the Electrical Communications Laboratory,* Vol. 16, Sept–Oct 1968.

Oppenheim, A.V., R.W. Schafer, and J.R. Buck. *Discrete-time Signal Processing*, 2nd. ed. Upper Saddle River, NJ: Prentice Hall, 1999.

Pahlavan, K., and Levesque, A., *Wireless Information Networks,* New York: Wiley-Interscience, 1995.

Papadias, C. and G. Foschini. "On the capacity of certain space–time coding schemes." *EURASIP Journal on Applied Signal Processing* 2002:5, pp. 447–458, Hindawi Publishing Corporation, 2002.

Parsons, D. *The Mobile Propagation Channel*. New York: Wiley, 1992.

Paulraj, A. "Diversity Techniques." chapter in *CRC Handbook on Communications*, ed. J. Gibson. Boca Raton, FL: CRC Press, 11, pp. 213–223, Dec. 1996.

Peterson, R.L., R.E. Ziemer, and D.E. Borth. *Introduction to Spread Spectrum Communications*. Upper Saddle River, NJ: Prentice Hall, 1995.

Proakis, J.G. *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

Ramo, S., J.R. Whinnery, and T. Van Duzer. *Fields and Waves in Communications Electronics.* New York: Wiley, 1965.

Rappaport, T.S. *Wireless Communications: Principles and Practice.* Upper Saddle River, N.J.: Prentice Hall, 1999.

Rappaport, T.S. *Wireless Communications: Principles and Practice*, 2nd. ed. Upper Saddle River, N.J.: Prentice Hall, 2002.

Rice, S.O. "Mathematical analysis of random noise," *Bell System Technical J.*, Vol. 23, pp. 282–332, 1944; Vol. 24, pp. 46–156, 1945. Reprinted in N. Wax (ed.), *Selected papers on Noise and Stochastic Processes*, New York: Dover, 1954.

Roddy, W. *Satellite Communications.* 2nd ed. New York: McGraw-Hill, 1996.

Saleh, A.A.M. and R.A. Valenzuela. "A statistical model for indoor multipath propagation." *IEEE J. Selected Areas Commun.*, 5, pp. 128–137, February 1987.

Scholtz, R.A. "The origins of spread spectrum communications." *IEEE Trans. Communications*, Vol. 30, pp. 822–854, 1982.

Schwartz, M., W.R. Bennett, and S. Stein. *Communications Systems and Techniques.* New York: McGraw-Hill, 1966. Reprinted by IEEE Press, 1995.

Sellathurai, M. and S. Haykin. "Joint beamformer estimation and co-antenna interference cancelation for turbo-BLAST," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Salt Lake City, UT, May 2001.

Sellathurai, M. and S. Haykin. "Turbo-BLAST: A novel technique for multi-transmit multi-receive wireless communications." *Multiaccess, Mobility, and Teletraffic for Wireless Communications*, Vol. 5, pp. 13–24, Kluwer Academic Publishers.

Sellathurai, M. and S. Haykin. "Turbo-BLAST for high speed wireless communications." *Proceedings of wireless communications and network conf.*, Chicago, Sept. 2000.

Sellathurai, M. and S. Haykin. "Turbo-BLAST for Wireless Communications: Theory and Experiments." *IEEE Trans. Signal Processing*, Special Issue on MIMO Wireless Communications, Vol. 50, pp. 2538–2546, 2002.

Sellathurai, M. and S. Haykin. "Turbo-BLAST for Wireless Communications: first experimental results," *IEEE Trans. on Vehicular Technology*, Vol. 52 pp. 530–535, 2003.

Sellathurai, M. and S. Haykin. "Turbo-BLAST: Performance Evaluation in Correlated Rayleigh Fading Environment." *IEEE J. Selected Areas of Communications*, Special issue, Vol. 21 pp. 340–349, 2003.

Shankar, P.M. *Introduction to Wireless Systems*, Hoboken, NJ: Wiley, 2002.

Shannon, C. "A mathematical theory of communication." *Bell System Technical J.*, Vol. 27, pp. 379–423, 623–656, 1948.

Shiu D.-S., G.J. Foschini, M.J. Gans, and J.M. Kahn. "Fading correlation and the effect on the capacity of multielement antenna systems." *IEEE Trans. Communications*, Vol. 48, pp. 502–513, 2000.

Simon, M.K., J.K. Omura, R.A. Scholtz, and B.K. Levitt. *Spread Spectrum Communications*, 3 Vols., Rockville, MD: Computer Science Press, 1985.

Simon, M.K., J.K. Omura, R.A. Scholtz, and B.K. Levitt. *Spread Spectrum Communications Handbook*, New York: McGraw-Hill, 1994.

Simon, S. and A. Moustakas. "Optimizing MIMO antenna systems with channel covariance feedback." *IEEE J. Selected Areas Communications*, Special Issue on MIMO Systems, Vol. 21, Iss. 3, pp. 406–417, 2003.

Sklar, B. *Digital Communications, Fundamentals and Applications*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2001.

Smith, P.J., M. Shafi, and G. Lebrun. "MIMO capacity in Rician fading channels: An exact characterization," under preparation, 2003.

Spilker, J.J., Jr. *Digital Communications by Satellite.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

Starr, T., J.M. Cioffi, and P.J. Silverman. *Understanding Digital Subscriber Line Technology.* Upper Saddle River, NJ: Prentice Hall, 1999.

Steele, R. and L. Hanzo. *Mobile Radio Communications*, 2nd ed. New York: Wiley/IEEE Press, 1999.

Stein, S. "Linear diversity combining techniques," Chapter 10 in M. Schwartz, W.R. Bennett, and S. Stein, Communication Systems and Techniques, McGraw-Hill, 1966.

Strang, G. *Linear Algebra and Its Applications*, 2nd ed. New York: Academic Press, 1980.

Stüber, G.L. *Principles of Mobile Communication*, 2nd ed. Boston: Kluwer Academic, 2001.

Stutzman, W.L. and G.A. Thiele. *Antenna Theory and Design*, 2nd ed. New York: Wiley, 1998.

Tabagi, F.A. and L. Kleinroke. "Packet switching in radio channels." *IEEE Trans. Communications*, Part I, Vol. 23, pp. 1400–1416, Part II, Vol. 23, pp. 1417–1433, 1975.

Tabagi, F.A. and L. Kleinroke. "Packet switching in radio channels." *IEEE Trans. Communications*, Part III, Vol. 24, pp. 832–844, 1976.

Tarokh, V. and H. Jafarkhani. "A differential detection scheme for transmit diversity." *IEEE J. Selected Areas in Communications*, Vol. 18, pp. 1169–1174, 2000.

Tarokh, V., H. Jafarkhani, and A. Calderbank. "Space–time block codes from orthogonal designs." *IEEE Trans. Inform. Theory*, Vol. 45, pp. 1456–1467, 1999.

Tarokh, V., H. Jafarkhani, and A.R. Calderbank. "Space–time block coding for wireless communications: Performance results." *IEEE J. Select. Areas Communications*, Vol. 17, pp. 451–460, Mar. 1999.

Tarokh, V., A. Naguib, N. Seshadri, and A. Calderbank. "Combined array processing and space–time coding." *IEEE Trans. Inform. Theory*, Vol. 45, pp. 1121–1128, 1999.

Tarokh, V., A. Naguib, N. Seshadri, and A. Calderbank. "Space–time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths." *IEEE Trans. Communications*, Vol. 47, pp. 199–207, 1999.

Tarokh, V., N. Seshadri, and A.R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction." *IEEE Trans. Information Theory*, Vol. 44, No. 2, pp. 744–765, 1998.

Technology Industry Association (TIA), Project No. PN-2759, "Cellular system dual-mode mobile station - Base station compatibility standard," (IS-54, Revision B) Jan. 9, 1992.

Telatar, I.E. "Capacity of multi-antenna Gaussian channels." *European Trans. Telecommunications*, Vol. 10, pp. 585–595, 1999. (Originally published as AT&T Technical Memorandum, 1995).

ten Brink, S. "Convergence of iterative decoding." *Electronic Letters*, Vol. 53, No. 10, pp. 806–808, May 1999.

Tobagi, F., and Kleinrock, L., "Packet Switching in Radio Channels: Part II - The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution", *IEEE Trans. on Communications,* Vol. 23, pp. 1417–1433, December 1975.

Ungerboeck, G. "Channel coding with multilevel/phase signals." *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 55–67, 1982.

Ungerboeck, G. "Trellis-coded modulation with redundant signal sets." *Parts 1 and 2, IEEE Communications Magazine*, Vol. 25, No. 2, pp. 5–21, 1987.

Verdú, S. "Minimum probability of error for asynchronous Gaussian multiple-access channels." *IEEE Trans. Inform. Theory*, Vol. 32, pp. 85–96, 1986.

Verdú, S. *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

Vishwanath, S., S. Jafar, and A. Goldsmith. "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback." *International Communications Conference*, 2001, Helsinki, Finland.

Viterbi, A.J. *CDMA: Principles of Spread Spectrum Communications*. Reading, MA: Addison-Wesley, 1995.

Viterbi, A.J. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm." *IEEE Trans. Inform. Theory*, Vol. IT-13, pp. 260–269, 1967.

Viterbi, A.J. and J.K. Omura. *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

Wang, X., and H.V. Poor. "Robust multiuser detection in non-Gaussian channels." *IEEE Trans. Signal Processing*, Vol. 47, pp. 289–305, 1999.

Widrow, B. and S.D. Stearns. *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

Win, M.Z. and R.A. Scholtz. "Impulse radio: how it works," *IEEE Commun. Letters*, Vol. 2, pp. 36–38, 1998.

Wolniansky, P.W., J.G. Foschini, G.D. Golden, and R.A. Valenzuela. "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel." *Proceedings of ISSSE*, Pisa, Italy, Sept. 1998.

Wozencraft, J.M. and I.M. Jacobs. *Principles of Communication Engineering*. New York: Wiley, 1965.

Zehavi, E. "8-PSK trellis codes for a Rayleigh fading channel." *IEEE Trans. Communications*, Vol. 40, pp. 873–883, May 1992.

Zheng, L. and D.N.C. Tse. "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel." *IEEE Trans. Inform. Theory*, Vol. 48, pp. 359–383, 2002.

Index

T

Modern Wireless Communications

SIMON HAYKIN • MICHAEL MOHER

The rapid growth of wireless communications and its pervasive use are changing the way we communicate in some fundamental ways. *Modern Wireless Communications* distinguishes itself with the treatment of these techniques in depth and yet in a principled manner that can be taught at the undergraduate level. The order in which these techniques are treated follows the evolution of spectral utilization of the radio channel. As a consequence, much of the material in the book is presented in the order in which it was actually developed, giving students insight into the logic and need for each development.

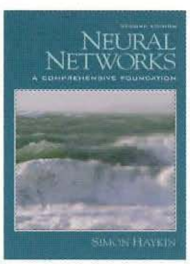
Features of the Book

Each chapter includes:

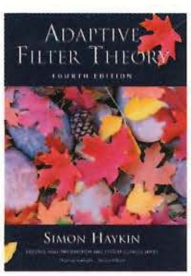
- Examples with detailed solutions
- Problems that help the reader develop an improved understanding of the issues, along with the answers
- End-of-chapter problems that provide an abundance of additional examples, whose solutions will further help the reader develop a deeper understanding of the material covered
- Notes and References that guide the reader to related sources for further reading.
- Theme Examples, which discuss pertinent issues of practical relevance.

Modern Wireless Communications is indispensable both in a course on wireless communications or as a useful reference for practicing electrical engineers.

Other Books by the Author



Neural Networks: A Comprehensive Foundation, Second Edition. ISBN: 0-13-273350-1.
Neural Networks represents the first comprehensive treatment of neural networks from an engineering perspective. Thorough, well-organized, and completely up-to-date, it examines all the important aspects of this emerging technology and provides broad coverage of the subject, including the learning process, back propogation, radial basis functions, recurrent networks, self-organizing systems, modular networks, temporal processing, neurodynamics, and VLSI implementations. The book includes chapter objectives, computer experiments, problems, worked examples, a bibliography, photographs, illustrations, and a thorough glossary to reinforce key concepts. The author's concise and fluid writing style makes the material more accessible.



Adaptive Filter Theory, Fourth Edition. ISBN: 0-13-090126-1.
The third edition of this highly successful book has been updated and refined to keep current with the field and develop concepts in as unified and accessible a manner as possible. Haykin examines both the mathematical theory behind various linear adaptive filters with finite-duration impulse response (FIR) and the elements of supervised neural networks.

ISBN 0-13-022472-3
90000
9 780130 224729

PEARSON
Prentice Hall

Upper Saddle River, NJ 07458
www.prenhall.com