《现代数字通信与编码理论》讲义

Principles of Advanced Digital Communications and Coding

白宝明

西安电子科技大学

综合业务网国家重点实验室

2010 年 7 月

Notice

This is a draft. The notes are work in progress.

Comments will be much appreciated; please send them to me at bmbai@mail.xidian.edu.cn

Course Information for 0122229

现代数字通信与编码理论

（Principles of advanced digital communications and coding）

The goal of this class is to introduce the information transmission techniques used in modern communication systems, with emphasis on information-theoretic and advanced coding aspects. This is done by understanding the following course contents:

各种信道模型（包括功率受限、带宽受限、ISI、衰落、多天线等）及其 Shannon 容量的计算；最新的可达容量限的信道码的编译码原理；现代编码通信系统的性能分析技术。

Prerequisite: Principles of communications

Error control coding (preferable but not necessary)

Instructor: Prof. Baoming BAI

Assistant:

Time and place: Monday 8:30 – 10:05 a.m. and Wednesday 3:35 – 5:10 p.m. in Classroom J2-04

Grading: 50% Homework

50% Project (The project will involve in writing a report as well as an oral presentation)

Class WWW page:

Outline

Preliminaries

- Phase splitter and analytic signal
- Complex baseband representation of passband signals
- Signal space representations
- Circularly symmetric Gaussian processes
- Some facts from information theory

Digital Transmission of Information over Ideal AWGN Channels (10 hours)

- Discrete-time AWGN channel model
- Signal constellation
- PAM and QAM transmission systems
- Capacity for *M*-PAM and *M*-QAM signaling
- The gap between uncoded performance and the Shannon limit
- Performance analysis of small signal constellations
- Design of signal constellations

Performance Analysis of Coded Communication Systems

- Approaching capacity with coding

- Techniques for performance analysis of coded communication systems
- Bhattacharyya bound and Gallager bound

Introduction to Modern Coding Theory (8 hours)

- Trellis representation of codes and decoding on a trellis (linear block codes, VA, BCJR)
- Turbo codes and the iterative decoding principles
- Performance analysis
- Codes defined on graphs and the sum-product algorithm
- LDPC codes

Bandwidth-Efficient Coded-Modulation Techniques (for Ideal Band-limited Channels) (8 hours)

- Lattice constellations
- Shaping gain
- TCM principles and performance analysis techniques
- Multi-dimentional TCM and multiple TCM
- Turbo-TCM codes
- Multilevel coding and multistage decoding
- Bit-interleaved coded modulation using Turbo codes and LDPC codes (Gallager mapping)
- Constellation shaping techniques

Transmission over Linear Gaussian Channels (6 hours)

- Linear Gaussian channels
- Equivalent discrete-time model
- Principles of “water pouring” and evaluation of the channel capacity
- Optimal receiver in the presence of both ISI and AWGN
- Optimal detection: MAP, ML sequence detection
- Symbol-by-symbol equalization methods: MMSE-LE, ZF-LE and MMSE-DFE
- Tomlinson-Harishima precoding
- Coding for ISI channels
- Principles of Turbo equalizations
- Approaching capacity with parallel transmission: COFDM

Communications over Fading Channels (5 hours)

- Wireless channel models
- Capacity of wireless channels
- Diversity techniques
- Coding for fading channels (including adaptive coding & modulation)
- Bound on the probability of decoding error
- Information-theoretic aspects of spread-spectrum communications

MIMO Wireless Communications (5 hours)

- Multi-antenna (MIMO) channel models
- Capacity of MIMO wireless channels
- Diversity and spatial multiplexing
- Approaching capacity with space-time coding
- Performance analysis and design criteria for space-time codes on fading channels
- Various space-time coding schemes

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, 623-656, July-Oct. 1948; Reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. Illinois Press, 1949.

[2] R. G. Gallager, "Claude E. Shannon: A Retrospective on His Life,Work, and Impact," *IEEE Trans. Inform. Theory*, vol.47, no.7, pp. 2681-2695, Nov. 2001.

[3] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol.44, no.6, pp. 2384-2415, Oct. 1998.

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

[5] D. J. Costello, J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol.44, no.6, pp.2531-2560, Oct. 1998.

[6] A. R. Calderbank, "The art of signaling: Fifty years of coding theory," *IEEE Trans. Inform. Theory*, vol.44, no.6, pp.2561-2595, Oct. 1998.

[7] J. G. Proakis, *Digital Communications*. 4rd ed. New York: McGraw-Hill, 2000.

[8] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Kluwer Academic Publishers, Boston, 1994.

[9] G. D. Forney and R. Gallager, *Principles of Digital Communications*. Course notes. MIT.

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

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[12] J. L. Massey, *Applied Digital Information Theory*. Course notes. ETH.

[13] S. G. Wilson, *Digital Modulation and Coding*. Prentice-Hall, 1996.

[14] E. Biglierli, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: MacMillan, 1991.

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

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

[17] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Course notes. EPFL.

[18] C. Schlegel and L. Perez, *Trellis and Turbo Coding*. IEEE Press, 2004.

[19] *Proceedings of The IEEE*, Special issue on wireless commun., vol.92, no.2, Feb. 2004.

[20] *IEEE Signal Processing Magazine*, Jan. 2004.

[21] *IEEE Communication Magazine*, Aug. 2003.

[22] M. Medard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inform. Theory*, vol.48, no.4, pp.840-852, April 2002.

[23] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol.47, no.4, pp.1290-1301, May 2001.

[24] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol.47, no.4, pp.1273-1289, May 2001.

[25] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Communications, vol.6, no.3, pp.311-335, Mar. 1998.

[26] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol.10, no.6, pp.585-596, Nov.-Dec. 1996.

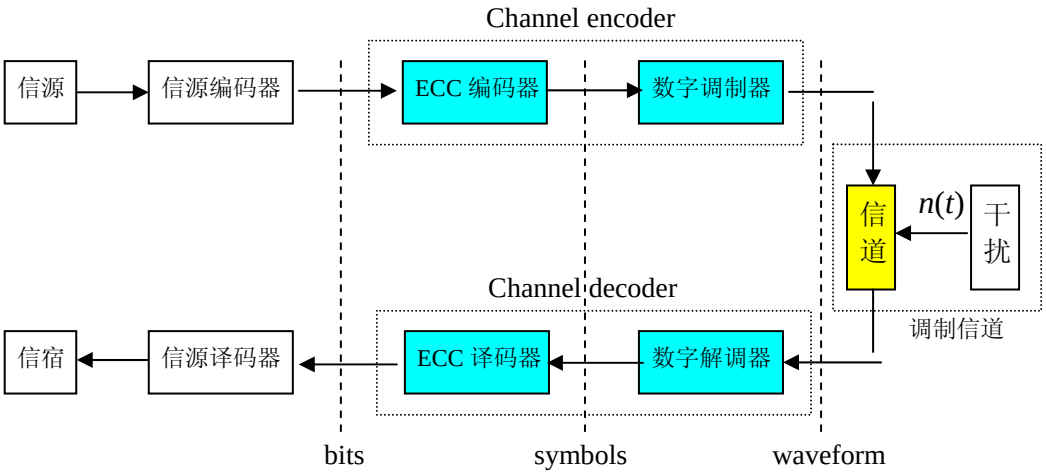
[27] C. Berrou and A.Glavieux, “Near optimum error correcting coding and decoding: Turbo-codes,” *IEEE Trans. Commun.*, vol. 44, no.10, pp.1261-1271, Oct. 1996.
[28] C. Heegard and S. B. Wicker, *Turbo Coding*, Norwell, MA: Kluwer, 1998.
[29] B. Vucetic and Jinhong Yuan, *Space-Time Coding*, Wiley, 2003.
[30] Shu Lin and D. J. Costello, Jr. *Error Control Coding: Fundamentals and Applications*. 2nd ed. Prentice-Hall, 2004.
[31] William E. Ryan and Shu Lin, *Channel Codes: Classical and Modern*. Cambridge University Press, 2009.
[32] *Proceedings of The IEEE*, Special issue on Turbo-information processing: Algorithms, implementations and applications, vol.95, no.6, June 2007.

Introduction

Digital communication is a field in which theoretical ideas have had an unusually powerful impact on actual system design. The basis of the theory was developed 58 years ago by Claude Shannon, and is called *information theory*. The goal of this course is to get acquainted with some of these ideas and to gain deep understanding on how to efficiently and reliably communicate through a channel, especially to better understand the advanced techniques for signal transmission and coding used in modern digital communication systems. We will focus on point-to-point systems consisting of a single transmitter, a channel and a receiver.

A. Block diagram of a digital communication system

In 1948, Claude E. Shannon of the Bell Telephone Laboratories published one of the most remarkable papers in the history of engineering. This paper ("A Mathematical Theory of Communication", *Bell System Tech. Journal*, Vol. 27, July and October 1948, pp. 379 - 423 and pp. 623 - 656) laid the groundwork of an entirely new scientific discipline, "information theory", in which Shannon first introduced the following figure to model a digital communication system.



The *source encoder* involves the efficient representation of source signals. It has the function of converting the input from its original form, e.g., speech waveforms, image waveforms, and text, into a sequence of bits. The objective of doing this is as efficiently as possible. i.e., transmitting as few bits as possible, subject to the need to reconstruct the input adequately at the output. In this case source encoding is often called *data compression*. Shannon showed that the ultimate data compression is the *entropy* of the source.

The *channel encoder* box in the figure above has the function of mapping the binary sequence at the source/channel interface into channel inputs. The channel inputs might be waveforms, or might be discrete sequences. The general objective here is to map binary inputs at the maximum bit rate possible into waveforms or sequences such that the channel decoder can recreate the original bits with low probability of error. One simple approach to performing this is called *modulation and demodulation*. From the geometric signal-space viewpoint, the modulation process may be thought of a two-step process: first mapping

binary digits into signals (e.g. signal levels) and then signals into waveforms.

Since high error probability is frequently incurred with simple modulation and demodulation in the presence of noise, the error-correcting codes was introduced and the channel coder is separated into two layers, first an error-correcting encoder, and then a simple modulator. Shannon showed that, with appropriate coding schemes, arbitrarily low error probabilities can be achieved at any data rate below a certain data rate called the *channel capacity*.

By the 1980's, channel coding usually involved a two layer system similar to that above, where an error-correcting code is followed by a modulator. At the receiver, the waveform is first demodulated, and then the error correction code is decoded. Since the Ungerboeck's work in 1982, it has been recognized that coding and modulation should be considered as a unit, resulting in the schemes called *coded modulation*. In such schemes, the lower error probability can be achieved without sacrificing bandwidth efficiency.

"The purpose of the modulation system is to create a good discrete channel from the modulator input to the demodulator output, and the purpose of the coding system is to transmit the information bits reliably through this discrete channel at the highest practicable rate." -- Massey

In this course, we will study the concepts and fundamental principles involved in advanced digital communication systems. We will focus on the channel coding component in the above figure. As we will see later, many advanced techniques used in modern digital communication systems (including mobile communication systems) are developed using information-theoretic ideas. This course will attempt to reflect these new evolutions. Some of exposition has benefited from the excellent notes written by Gallager and Forney for the MIT courses 6.450 and 6.451.

We will present the material in such a unified way that the channel model and the corresponding channel capacity are introduced first, and then the coding and signal process techniques for approaching these optimal performance limits are presented, and followed by the discussion on the performance of the actual systems with these channel coding schemes.

B. Relevant results from information theory

- The communications problem can be broken down without loss of reliability or efficiency into the separate components shown in the above diagram.Reliable communication can be achieved at any rate below the capacity of the communications channel.
- We add controlled redundancy to data transmitted over the channel. This redundancy lowers the raw data rate, but reduces the error rate after using the redundancy to correct errors. (*distance gain*)
- The net effect is to increase the rate at which clean data is delivered

C. Historical notes

- Hamming codes: 1950
- Convolutional codes: 1955 (by Elias)
- BCH, Reed-Solomon codes: 1960
- LDPC codes: 1962 (by Gallager) (rediscoved in late 1990's)
- Concatenated codes: 1966 (by Forney)
- Viterbi algorithm: 1967
- TCM: 1982 (by Ungerboeck)

- Turbo codes: 1993 (by Claude Berrou)
- Space-time codes: 1998 (by V. Tarokh)
- Dirty-paper coding, Cooperation via distributed coding and network coding: 2000-

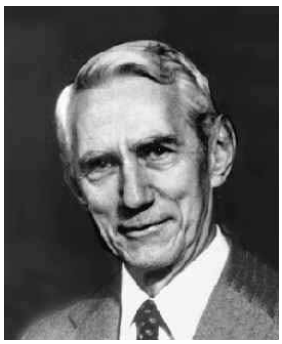
■ **Most of important achievements in digital communications are based on the results of information theory and coding**.

D. Giants in the field of digital communications

■ **Harry Nyquist** (1928)

- Analog signals of bandwidth W can be represented by $2W$ samples/s
- Channels of bandwidth W support transmission of $2W$ symbols/s

■ **Claude Shannon** (1948)



4/30/1916 – 2/24/2001

- His information theory addressed all the big questions in a single stroke.
- He thought of both information sources and channels as random and used probability models for them.
- Most modern communication systems are designed according to the principles laid down by Shannon.

We conclude this section, which should have provided some motivation for the use of coding, with an adage from R. E. Blahut: "To build a communication channel as good as we can is a waste of money – use coding instead!"

Chapter 1 Preliminaries

In this chapter we will briefly review some basic concepts and principles, which will be used as the basis of discussions later.

A. *Phase splitter and analytic signal*

- If $x(t)$ is a real-valued signal, then its Fourier transform $X(f)$ satisfies the symmetry property

$$X(-f) = X^*(f)$$

where $X^*(f)$ is the complex conjugate of $X(f)$.

The symmetry property says that knowing $X(f)$ for $f \geq 0$ is sufficient to entirely describe $X(f)$ and thus to describe $x(t)$.

- A *phase splitter* (also known as Hilbert filter) is a complex filter with impulse response $h_+(t)$ and transfer function $H_+(f)$, where

$$H_+(f) = \begin{cases} 1, & f \geq 0 \\ 0, & f < 0 \end{cases}$$

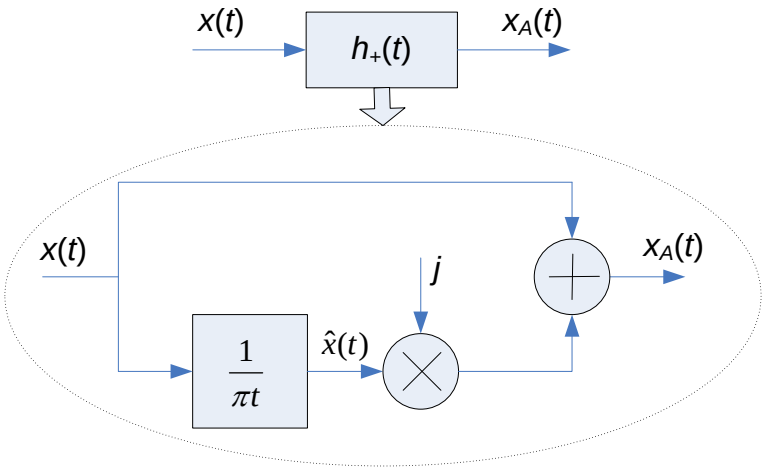


Figure 1.0 A Hilbert filter

- If the real-valued input to a phase splitter is $x(t)$, then the output is

$$x_A(t) = \frac{1}{\sqrt{2}}[x(t) + j\hat{x}(t)], \text{ or}$$

$$X_A(f) = \sqrt{2}X(f)H_+(f)$$

where $\hat{x}(t)$ is the *Hilbert transform* of $x(t)$. We introduce the factor $\sqrt{2}$ so that $x(t)$ and $x_A(t)$ have the same energy (or power). Notice that $x_A(t)$ is a complex-valued signal.

- A signal with only nonnegative frequency components is called an *analytic signal*.
- $x(t)$ can be recovered from $x_A(t)$ by

$$x(t) = \sqrt{2}\Re[x_A(t)]$$

B. *Complex baseband representation of passband signals*

- Suppose that $x(t)$ is a real-valued passband signal with a spectrum centered at $f = f_c$. The *complex baseband equivalent signal* (sometimes also called complex envelope) of $x(t)$ can be represented as

$$x_b(t) = \underbrace{\frac{1}{\sqrt{2}}[x(t) + j\hat{x}(t)]}_{\text{analytic passband signal } x_A(t)} e^{-j2\pi f_c t} \tag{1.1}$$

  In terms of Fourier transforms

$$X_b(f) = X_A(f + f_c) = \begin{cases} \sqrt{2}X(f + f_c), & f + f_c \geq 0 \\ 0, & f + f_c < 0 \end{cases}$$

- The original passband signal can be recovered from $x_b(t)$ by

$$x(t) = \sqrt{2}\Re\left[x_b(t)e^{j2\pi f_c t}\right] \tag{1.2}$$

The relationship between $x(t), x_A(t)$ and $x_b(t)$ is shown in Fig. 1.1 in terms of their spectrum.

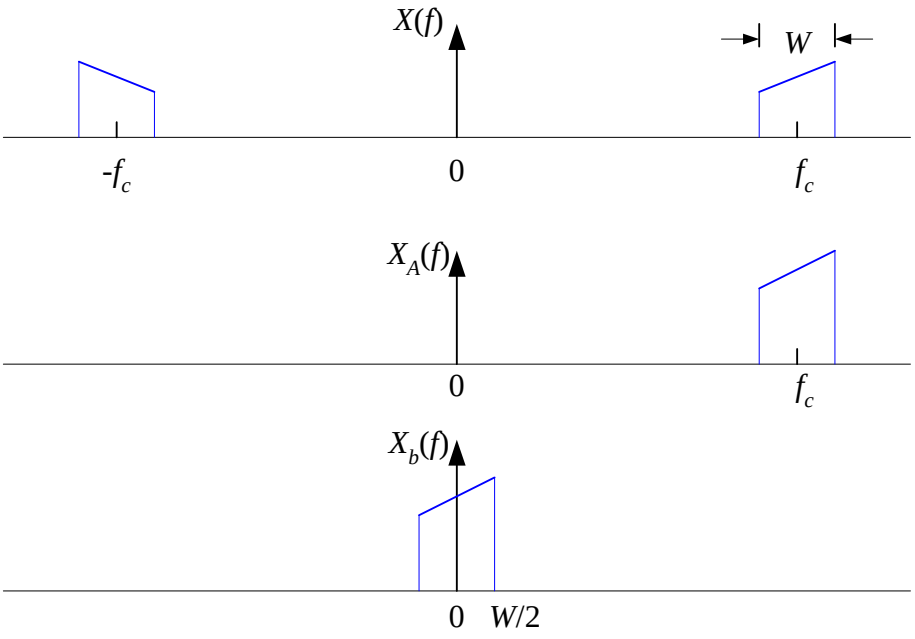


Figure 1.1 Fourier transform of a passband signal $x(t)$ and the transform of the corresponding complex baseband signal.

- Baseband to passband and go back

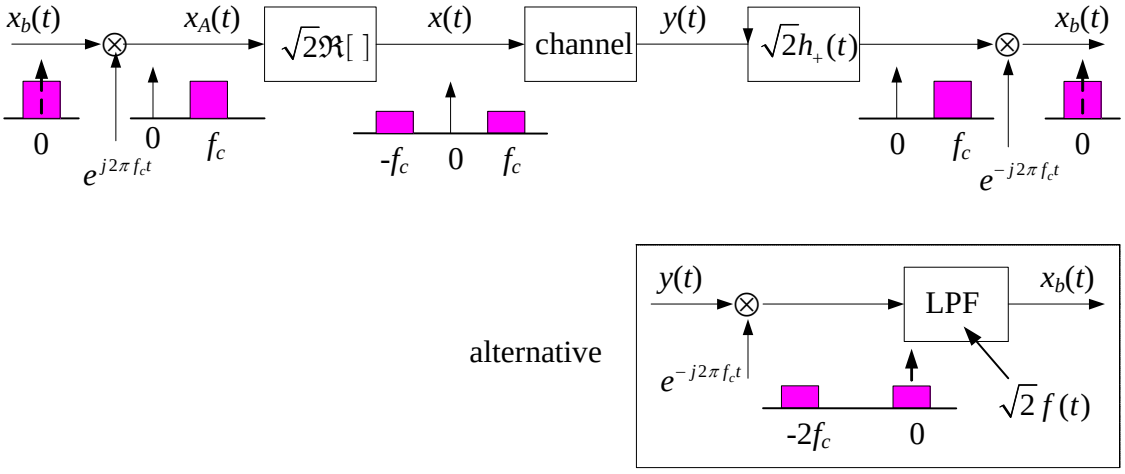


Figure 1.2

- Baseband equivalent channel (at carrier frequency f_c)

$$h_A(t) = \frac{1}{\sqrt{2}}[h(t) + j\hat{h}(t)]$$

$$h_b(t) = h_A(t)e^{-j2\pi f_c t}$$

- An alternative representation of a real signal is derivative of the complex envelope representation. The real and imaginary parts of the complex envelope $x_b(t)$ are referred to as the *in-phase* and *quadrature* components of $x(t)$, respectively, and are denoted by $x_I(t) = \Re\{x_b(t)\}$ and $x_Q(t) = \Im\{x_b(t)\}$. From (1.2), we have the in-phase and quadrature representation of a real signal $x(t)$ given by

$$\begin{aligned} x(t) &= \sqrt{2}\Re\left[x_b(t)e^{j2\pi f_c t}\right] \\ &= \sqrt{2}\Re\left[x_b(t)\right]\cos(2\pi f_c t) - \sqrt{2}\Im\left[x_b(t)\right]\sin(2\pi f_c t) \\ &= \sqrt{2}x_I(t)\cos(2\pi f_c t) - \sqrt{2}x_Q(t)\sin(2\pi f_c t) \end{aligned} \tag{1.3}$$

A quadrature modulator performing upconversion and a quadrature demodulator performing downconversion are shown in Fig. 1.3, respectively.

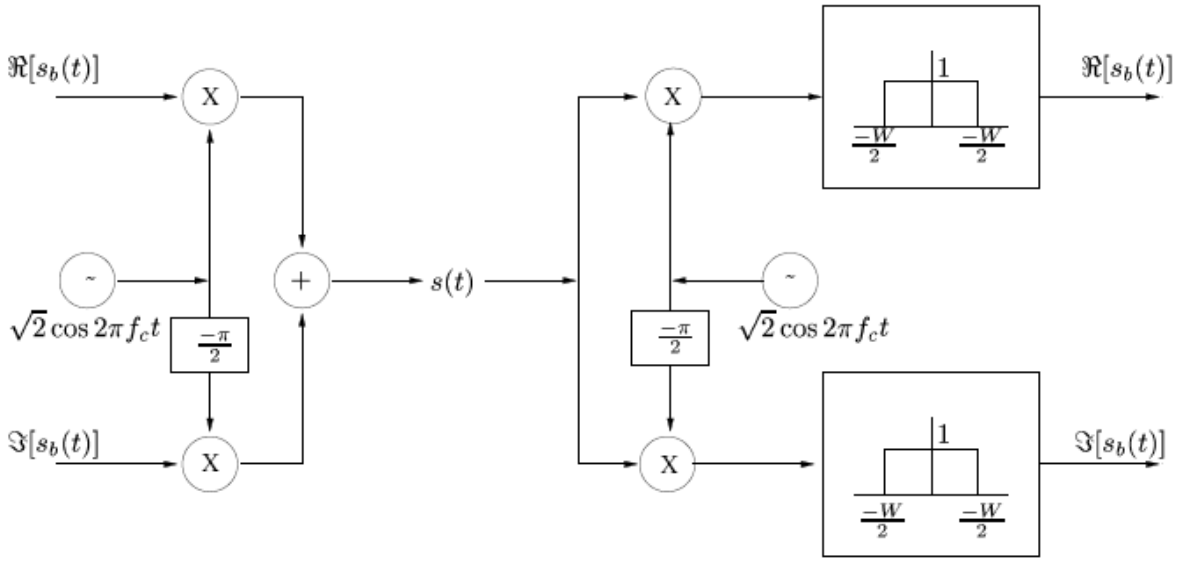


Figure 1.3 Quadrature modulation and demodulation

C. A complete system diagram

The next step in implementing a digital communication system is to convert the discrete-time signal sequence into a baseband waveform (such as a PAM or QAM modulated waveform), and vice versa. This is performed via baseband modulation and demodulation. For example, with QAM transmission, the baseband complex waveform can be expressed as

$$x_b(t) = \sum_n x_n p(t - nT) = \sum_{n\in\mathbb{Z}} (x'_n + jx''_n)\, p(t - nT)$$

where $\{x_n\}$ is a discrete-time sequence of complex symbols to be transmitted, T is the symbol interval/duration, and $x'_n = \Re\{x_n\}$ and $x''_n = \Im\{x_n\}$. The real waveform $p(t)$ is a basic modulation pulse. At the receiver, the sequence $\{x_n\}$ can be retrieved from the sampled outputs $y_n = y_b(nT)$. Figure 1.4 shows a complete system diagram with $p(t) = \text{sinc}(t)$ which is defined as

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi \text{t}}$$

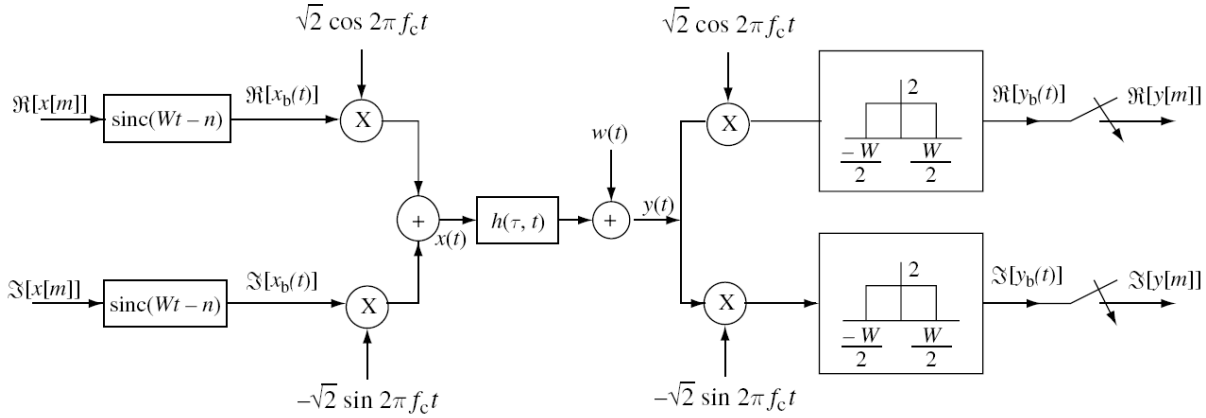


Figure 1.4 A complete system diagram from the baseband transmitted symbol to the baseband received symbol.

D. Signal space representations

- A signal space is a linear space (or vector space) in which vectors represent signals.
- In an *n*-dimensional complex vector space $\mathbb{C}^n$, the *inner product* of two vectors $\mathbf{u} = (u_1, ..., u_n)$ and $\mathbf{v} = (v_1, ..., v_n)$ is defined as

  $$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i^*$$

  A vector space equipped with an inner product is called an *inner product space*.
- A special notation is used for <**u**, **v**>,

  $$\langle \mathbf{u}, \mathbf{u} \rangle = \| \mathbf{u} \|^2 = \sum_{i=1}^{n} | u_i |^2$$

  where ||**u**|| is called the *norm* of vector **u** and geometrically is the length of the vector.
- Two vectors **u**, **v** are said to be *orthogonal* if <**u**, **v**> = 0.
- *Schwarz inequality:* Let **u** and **v** be vectors in an inner product space (either on $\mathbb{R}$ or $\mathbb{C}$). Then

$$\langle \mathbf{u}, \mathbf{v} \rangle \le \| \mathbf{u} \| \cdot \| \mathbf{v} \|$$

- Orthonormal bases: In an inner product space, a set of vectors $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots$ is *orthonormal* if

$$\langle \boldsymbol{\phi}_j, \boldsymbol{\phi}_k \rangle = \delta_{jk} = \begin{cases} 1, & \text{for } j = k \\ 0, & \text{for } j \neq k \end{cases}$$

- One dimensional projections: The vector **u** can be viewed as the sum of two vectors

$$\mathbf{u} = \mathbf{u}_{|\mathbf{v}} + \mathbf{u}_{\perp \mathbf{v}}$$

where $\mathbf{u}_{|\mathbf{v}}$ is collinear with **v**, and $\mathbf{u}_{\perp \mathbf{v}}$ is orthogonal to **v**. The vector $\mathbf{u}_{|\mathbf{v}}$ is called the *projection* of **u** onto **v**.

- Finite dimensional projections: If *S* is a subspace of an inner product space *V*, and $\mathbf{u} \in V$, the projection of **u** on *S* is defined to be a vector $\mathbf{u}_{|S} \in S$ such that

$$\langle \mathbf{u} - \mathbf{u}_{|S}, \mathbf{v} \rangle = 0 \quad \text{for every vector } \mathbf{v} \in S.$$

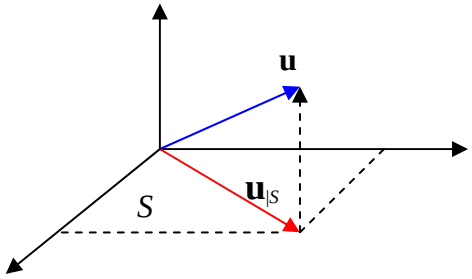


Figure 1.5

- *Projection theorem*: Let *S* be an *n*-dimensional subspace of an inner product space *V* and assume that $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n$ is an orthonormal basis for *S*. Then any $\mathbf{u} \in V$ may be decomposed as $\mathbf{u} = \mathbf{u}_{|S} + \mathbf{u}_{\perp S}$, where $\mathbf{u}_{|S} \in S$ and $\langle \mathbf{u}_{\perp S}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in S$. Futhermore, $\mathbf{u}_{|S}$ is uniquely determined by

$$\mathbf{u}_{|S} = \sum_{j=1}^{n} \langle \mathbf{u}, \boldsymbol{\phi}_j \rangle \boldsymbol{\phi}_j$$

- A consequence of projection theorem is that the projection $\mathbf{u}_{|S}$ is the unique closest vector in *S* to **u**; that is, for all $\mathbf{v} \in S$,

$$\| \mathbf{u} - \mathbf{u}_{|S} \| \le \| \mathbf{u} - \mathbf{v} \|$$

with equality iff $\mathbf{v} = \mathbf{u}_{|S}$. See figure 1.5.

- *Gram-Schmidt orthogonalization procedure*: It produces an orthonormal basis $\{\boldsymbol{\phi}_j\}$ for an arbitrary *n*-dimensional subspace *S* with the original basis $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. See, e.g., [Proakis 2000, ch4] for details.

E. Circularly symmetric Gaussian processes

- A vector **X** with *M* jointly Gaussian real-valued random variables has the p.d.f.

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{M/2} \sqrt{\det(K_{\mathbf{X}})}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_{\mathbf{X}})^T K_{\mathbf{X}}^{-1} (\mathbf{x} - \mathbf{m}_{\mathbf{X}}) \right)$$

where $K_{\mathbf{X}} = E[(\mathbf{x} - \mathbf{m}_{\mathbf{X}})(\mathbf{x} - \mathbf{m}_{\mathbf{X}})^T]$ is the covariance matrix, and $\mathbf{m}_{\mathbf{X}} = E[\mathbf{x}]$ is

the vector mean.

- A *complex-valued Gaussian random process* consists of two jointly Gaussian real-valued processes, a real part and an imaginary part. By jointly Gaussian, we mean that any arbitrary set of samples of real and imaginary parts is a jointly Gaussian set of random variables.
- Let $Z(t)$ be a zero mean complex-valued Gaussian process. Let $R(t) = \Re[Z(t)]$ and $I(t) = \Im[Z(t)]$. By definition, both $R(t)$ and $I(t)$ are zero mean real Gaussian processes.
- Thus, $R(t)$ and $I(t)$ are fully characterized by their 2nd order statistics,

$$R_R(\tau) = \mathrm{E}[R(t+\tau)R(t)], \quad R_I(\tau) = \mathrm{E}[I(t+\tau)I(t)],$$
$$R_{RI}(\tau) = \mathrm{E}[R(t+\tau)I(t)]$$

- The complex-valued process $Z(t)$ is *strictly stationary* if $R(t)$ and $I(t)$ are *jointly wise-sense stationary*, and hence jointly strictly stationary.
- By definition, the complex-valued process $Z(t)$ is wise-sense stationary if the *autocorrelation function*

$$R_Z(\tau) = \mathrm{E}[Z(t+\tau)Z^*(t)]$$

is independent of t.

Notice that this is not the same as saying that the real and imaginary parts are jointly wide-sense stationary, since $R_Z(\tau)$ could not by itself contain information equivalent to $R_R(\tau)$, $R_I(\tau)$ and $R_{RI}(\tau)$.

- Thus, we require more than $R_Z(\tau)$ to fully specify the statistics of $Z(t)$. In addition to $R_Z(\tau)$, it suffices to know the *complementary autocorrelation function* defined as

$$\tilde{R}_Z(\tau) = \mathrm{E}[Z(t+\tau)Z(t)]$$

- Using the relations $2R(t) = Z(t) + Z^*(t)$ and $2jI(t) = Z(t) - Z^*(t)$, it is easy to show that

$$2R_R(\tau) = \mathrm{Re}\{R_Z(\tau)\} + \mathrm{Re}\{\tilde{R}_Z(\tau)\}$$
$$2R_I(\tau) = \mathrm{Re}\{R_Z(\tau)\} - \mathrm{Re}\{\tilde{R}_Z(\tau)\}$$
$$2R_{RI}(\tau) = \mathrm{Im}\{\tilde{R}_Z(\tau)\} - \mathrm{Im}\{R_Z(\tau)\}$$

- With these equations, we can see that if $Z(t)$ is wise-sense stationary, and in addition $\tilde{R}_Z(\tau)$ is not a function of t, then $R(t)$ and $I(t)$ are jointly wise-sense stationary, and $Z(t)$ strictly stationary.

- *Circularly symmetric Gaussian random variables*: Let $Z = R + jI$ be a zero mean Gaussian variable. Z will be called circularly symmetric if

$$\mathsf{E}[Z^2] = \mathsf{E}[R^2] - \mathsf{E}[I^2] + 2j\mathsf{E}[RI] = 0$$

Note that R and I are i.i.d. iff $\mathrm{E}[Z^2] = 0$.

- The source of the terminology: $e^{j\phi}Z$ has the same distribution as Z. It is

$$p_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\|z\|^2}{2\sigma^2}\right\}$$

- A complex-valued zero mean Gaussian process is circularly symmetric if

$$\mathrm{E}[Z(t+\tau)Z(t)] = 0 \quad \text{for all } t \text{ and } \tau$$

- *Property*:
  - A circularly symmetric Gaussian process is strictly stationary iff it is wide-sense stationary.
  - For a wide-sense stationary circularly symmetric Gaussian process,

$$R_R(\tau) = R_I(\tau) = \frac{1}{2}\mathrm{Re}\{R_Z(\tau)\}, \quad R_{RI}(\tau) = -\frac{1}{2}\mathrm{Im}\{R_Z(\tau)\}$$

- Circularly symmetric processes with a real-valued $R_Z(\tau)$ have a real and imaginary part that are independent at all time, since $R_{RI}(\tau) = 0$.
- Circularly symmetry is preserved by linear (time-invariant or time-varying) systems.

■ A white complex-valued Gaussian process has an autocorrelation function

$$R_Z(\tau) = N_0\delta(\tau), \quad R_Z(k) = 2\sigma^2\delta_k$$

for continuous and discrete time, respectively.

■ For a circularly symmetric white Gaussian process, the real and imaginary parts are identically distributed, and are independent of each other.

D. Basics of information theory

Entropy and mutual information

■ For a discrete random variable X with sample space Ω_X, its entropy is defined as

$$H(X) = \mathbb{E}[-\log P_X(x)] = -\sum_{x\in\Omega_X} P_X(x)\log P_X(x)$$

■ The mutual information between two random variables X and Y are given by

$$I(X,Y) = H(X) - H(X\,|\,Y) = H(Y) - H(Y\,|\,X)$$

$$I(X;Y) = \sum_x p(x)\sum_y p(y|x)\log\frac{p(y|x)}{p(y)}$$

$$I(X;Y) = \sum_x p(x)\int_{-\infty}^{\infty} p(y|x)\log\frac{p(y|x)}{p(y)}dy$$

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y|x)\log\frac{p(y|x)}{p(y)}dxdy$$

Channel capacity and the coding theorem

- **(Operational) Channel capacity:**
  Maximum rate R for which reliable communication can be achieved.
- **Information channel capacity:**
  Maximum of mutual information over all possible input statistics $P(X)$

$$C \equiv \max_{P_X} I(X;Y) = \max_{P_X}\ [H(Y) - H(Y\,|\,X)]$$

Suggested Reading

[1] R. G. Gallager, *Principles of Digital Communication*. Cambridge University Press, 2009. (影印版，人民邮电出版社，2010)