Signal Processing Techniques for

Software Radios

2nd Edition

Behrouz Farhang-Boroujeny

Department of Electrical and Computer Engineering
University of Utah

This book comes with a downloadable CD that can be found at the authors website: http://www.ece.utah.edu/∼farhang/

This book is self -published by the author at the Lulu publishing house; http://www.lulu.com/

2nd Edition, Version 2.0

To Diana
for her continuous support
understanding and love
throughout my career

Contents

Preface

Advancement in digital signal processing algorithms as well as programmable digital signal processing platforms, being FPGA, a digital signal processor, or both, have led to the notion of Software Defined Radio. The term software radio refers to a piece of equipment that consists of a programmable digital signal processor core, and the necessary air interfaces (consisting of DACs, ADCs, analog circuitries, and antennae) to build a complete communication system. The ultimate goal is to reduce the analog parts to a bare minimum and put as much load as possible on the programmable part of the system. Ideally, ADC and DAC blocks should be moved as close as possible to the antennae. Hence, operations such as modlation/demodulation and various filtering operations are all performed in digital domain using software programs. Thus, through replacement of software, the same equipment (software radio platform) can be used to implement a range of radios for various applications.

This book puts together a collection of signal processing algorithms, filter design methods, and signal processing techniques (tricks) to provide the practicing engineers with the tools necessary for efficient implementation of software radios. The book is divided in two parts. Chapters 2 to 5 and Chapter 7 develop the theoretical foundation based on which various functions in a physical layer of software radio are implemented. The rest of the book discusses and presents implementation details of data modems.

A short review of Fourier analysis and linear time-invariant systems, as a brain refresher, is presented in Chapter 2. Chapter 3 introduces digital transmission systems and establishes most of the notations that are used throughout the book. The related background theory of digital signal processing is presented in Chapter 4. Design of various pulse-shaping filters has been given a special treatment in this chapter. Chapter 5 contains a detailed presentation of multirate signal processing and a number of application examples related to software radios. A brief theory of adaptive filters along with commonly used adaptive filtering algorithms is presented in Chapter 7.

To demonstrate the implementation of various algorithms on a software radio platform and also to demonstrate their performance, MATLAB scripts (programs) are presented throughout the book. The choice of MAT-

LAB, as against any other programming language, has a number of reasons. First, MATLAB is widely used in both industry and academia. It is thus assumed that most of the readers of this book have some level of familiarity with MATLAB. Second, the vast graphic functions available in MATLAB facilitate visual presentation of the results. Third, MATLAB is simple enough to allow even those who are not familiar with it still follow most of the presented scripts and functions.

The book come with an accompanying (downloadable) CD that contains all the MATLAB codes that are developed throughout the book as well as a number of test files that the reader needs for doing some of the end of the chapter problems and particularly to complete the design project that is introduced in Chapter 12 (see below). To download the CD, please visit the author's website at http://www.ece.utah.edu/∼farhang/.

The second part of the book begins with Chapter 6. This is an introductory chapter to the rest of the book. The main goal here is to set-up the basic MATLAB scripts that one needs in realizing any transceiver (modem). Eye-patterns as visualization tools in study of quality of the demodulated and (possibly) equalized signals are introduced in this chapter. In addition, the chapter provides a top view of digital communication systems and touches on the channel impairments, such as, noise, frequency offset and multipath effects.

Chapter 8 introduces phase-locked loop (PLL) and provides an in depth theoretical study as well as design of PLLs. The developed theory will be put in use in Chapter 10, where a number of carrier recovery and carrier tracking schemes are presented.

Chapter 10 is devoted to symbol timing recovery. The chapter begins with a study of statistical properties of digital data signals and finds that they are cyclostationary processes. In particular, we find that the ensemble average of the squared value of any digital data signal is a periodic function of time whose frequency matches the baud rate of the received data symbols. We will take advantage of this property in development of a number of symbol synchronization/timing recovery methods.

Channel equalization is discussed in Chapter 11. We distinguish between equalizers in which the spacing between adjacent taps is equal to data symbols and those that have smaller tap-spacing. The former is called a symbol-spaced equalizer and the latter is a fractionally-spaced equalizer, because of obvious reasons. Symbol-spaced and fractionally-spaced equalizers are contrasted against each other by deriving cost functions that can be evaluated for numerical comparisons. This chapter also covers the subject of cyclic equalization in detail. Cyclic equalization is a clever, yet simple method for synchronization of a receiver with its transmitter counterpart.

Chapter 12, titled "Putting it together", aims at bringing together what has been learnt in the previous chapters, in the context of a hands on project. A transmitter is introduced at the beginning of the chapter and the reader is guided to develop various receivers to extract the transmitted

information. The reader does not know what information has been transmitted. However, it turns out to be very rewarding when he/she observes a meaningful text which shows correct detection of the transmitted information. The author's experiment with this project in classrooms has been a very fun and rewarding exercise. According to students, this is where they really learn what has been covered in the previous chapters.

The topics covered in Chapter 2 through 12 mostly relate to single carrier communication systems. However, the modern broadband communication systems use multicarrier techniques for effective transmission at very high bit rates. Chapters 13 and 14, that are the new additions to this edition of this book present two classes of multicarrier methods. The widely adopted multicarrier method in the current standards, known as orthogonal frequency division multiplexing (OFDM), is presented in Chapter 13. An alternative class of multicarrier methods that use filter banks is the subject of Chapter 14. Although, multicarrier filter banks are not yet fully developed and have received less attention in the industry, they have a number of useful properties that, compared with OFDM, make them a better choice in the applications of multiuser multicarrier networks. The authors intention of presenting the multicarrier filter banks in this book is to promote this important class of multicarrier systems and to motivate further research in this important area.

As a class text, this book can be taught in a number of different ways. In author's opinion, the goal should be to reach and cover most of the last chapter of the book (Chapter 12 / the project) before the class ends. It might be a big challenge if one attempts to fully cover the whole book in one semester. A realistic approach will be to teach fundamental parts plus some selected topics and aim for reaching Chapter 12 three or four weeks before the end of the semester, so that students will have enough time to complete the class project. What I consider as fundamental parts are:

- Chapter 2 (Fourier Analysis and Linear Time-Invariant Systems): Most of this chapter should have been covered in the previous classes. It can be left as self-study for students.

- Chapter 3 (Digital Transmission Systems): Depending on the program that students have gone through, they may or may not have seen the content of this chapter.

- Chapter 4 (Sampling and Discrete Time Systems): Assuming students have gone through a DSP class, they should have already seen the first three sections of this chapter. It may be thus necessary to only teach Section 4.4 (Digital Filters). I consider Sections 4.4.1 through 4.4.3 as fundamental. One can leave out Section 4.4.4.

- Chapter 5 (Multirate Signal Processing): This is a relatively long chapter on multirate signal processing. Although the beauty of the

signal processing materials, here, are very tempting to teach everything, it may work better if one takes certain material out of the classroom and let them remain as self-study material for the students. The essential sections are 5.1 through 5.5.

- Chapter 6 (An Overview of Transceiver Systems): With the exception of Section 6.7 that may be left as self-study for interested students, this chapter should be fully covered.
- Chapter 7 (Adaptive Systems): It is not necessary to cover all details of this chapter. What I consider as essential are Sections 7.1 and 7.2.
- Chapter 8 (Phase-Locked Loop): It is not necessary to go through all the details of this chapter. I usually spend one hour in highlighting the pertinent points of phase-locked loops.
- Chapter 9 (Carrier Acquisition and Tracking): All sections in this chapter are interesting for most of the students and fun to teach. However, if time does not permit, take out Section 9.2.2.
- Chapter 10 (Timing Recovery): All sections in this chapter are also interesting for most of the students and fun to teach. However, if time does not permit, take out Sections 10.2.2 and 10.2.3.
- Chapter 11 (Channel Equalization): Most of this chapter should be covered. One may only take out some details of the derivations.
- Chapter 12 (Putting It Together): Not much to teach here. Leave it to students to go through it as a class project.

An alternative method of teaching this book is to present the whole book in a two semester/quarter class.

Once again, the author wishes to remind the readers of this book, particularly, the course instructors that this is an on going book project. Please send your suggestion and corrections for improving this text to the author at the below e-mail.

Behrouz Farhang-Boroujeny
farhang@ece.utah.edu

Acknowledgments

I wish to thank Professor Jonathon Chambers of the Advanced Signal Processing Group, Department of Electronic and Electrical Engineering at Loughborough University, UK, for diligent and careful reading of the first print of this book and making numerous invaluable suggestions.

I also wish to thank the members of the Software Defined Radio (SDR) Forum Educational Working Group for their encouragement. Special thanks go to Dr. Robert W. McGwier of the Institute for Defense Analysis for using the initial draft of this book in couple of classes that he taught at the College of New Jersey and University of Maryland. His inspiring comments on the usefulness of the material have been a big motive for a faster development of this book.

I am grateful to my students at the University of Utah who have enthusiastically given me their invaluable feedback in improving the content of this book.

Behrouz Farhang-Boroujeny

Chapter 1

Introduction

1.1 Software Radio

The term software-defined radio (SDR), or simply software radio, refers to a radio communication system that can be configured to receive a wide range of modulated digital signals across a large frequency spectrum by means of a programmable hardware platform. An SDR platform may be based on a general-purpose processor, a special-purpose digital signal processor, a field-programmable-gate-array (FPGA), or any combination of these. In addition, an SDR platform includes an air interface consisting of (reconfigurable) radio frequency (RF) front-end circuitries and antennae.

Figure 1.1 presents a block diagram of an SDR receiver platform. The transmitter counterpart of this platform is somewhat similar, with the order of the blocks reversed. The RF front-end circuitries consist of a low-noise amplifier (LNA) with an automatically controllable gain, and often a (programmable) RF to IF (intermediate frequency) demodulator. The antenna may be a single antenna or an array of antennae that may be configured (through software) for a desirable directivity. In addition, an SDR platform may be equipped with an array of antennae that can be configured for reception of RF signals from a very wide spectrum, e.g., 2 MHz to 5 GHz. It is also desirable to have control over the bandwidth of the RF/IF signals in order to remove irrelevant spurious signals and noise before conversion from analog to digital. A compromise choice on the bandwidth has to be made to trade the system flexibility versus performance and/or cost.

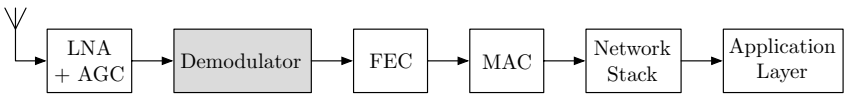


Figure 1.1: An SDR receiver platform.

Clearly, maximum flexibility/reconfigurability is the ultimate goal in any SDR system. This is achieved by inserting an analog-to-digital converter (ADC) in the chain of the blocks in Figure 1.1 as close as possible to the antenna. In practice, one has a limited control on reconfiguring the blocks to the left of ADC, i.e., the analog circuitries/elements. On the other hand, since in an SDR all the blocks after ADC are implemented in software or on reconfigurable devices, such as FPGA, the user has almost full freedom to tailor these blocks for different applications. Hence, ideally, the ADC should be located right after the cascade of the antenna and the LNA. In cases where carrier frequency is relatively high and thus ADC operation at RF is not feasible or is too expensive to implement, the ADC is put at an IF stage in the receiver chain.

In Figure 1.1, demodulation may be performed in a single step, through direct conversion of RF signal to baseband, or through a *superheterodyne* structure where down-conversion occurs through one or more IF stages. In traditional communication systems, where all blocks are implemented using analog circuitries, the use of superheterodyne structure has proven useful because it simplifies the filtering operations required to extract the desired signal. In software radios, also, the use of superheterodyne structure is found beneficial. Here, significant reduction in computational complexity is achieved through use of *multirate* signal processing techniques. Other elements that are not explicitly shown, but have to be included as sub-blocks in the demodulator block in Figure 1.1, are carrier frequency and phase recovery, symbol timing recovery, and channel equalizer. A brief review of these blocks is presented in Section 1.3, below.

Other blocks in the receiver chain are:

- *Forward error correcting (FEC):* Redundant bits are often added to the raw data bits prior to transmission. This is called *channel coding*. At the receiver, the FEC decoder takes into account the presence of redundant bits and corrects the bit errors that may have occurred as a result of channel impairments. The FEC decoder capability depends on the amount of the redundant bits that have been added to the raw data and the way such bits have been included (i.e., the type of channel code). To name, convolutional codes, turbo codes and low-density parity check (LDPC) codes are the common codes in use today. The combination of modulator/demodulator and FEC encoder/decoder in communication link is often referred to as physical layer, or, in short, PHY.

- *Medium access control (MAC):* MAC is the next processing layer that sits at the top of PHY. The tasks under MAC layer include data framing and the associated synchronization structures and control, resource allocation and MAC addressing, and payload encapsulation with possible fragmentation/defragmentation structures.

- *Network stack:* In addition to PHY and MAC, an operational network involves a number of other layers. These may include, but not necessary all, *network layer, transport layer, session layer,* and *presentation layer.* The network layer is responsible for determining the route(s) that data from each node/user in the network should take to reach its destination safely. The basic function of the transport layer is to accept data from the session layer, split it into smaller units if needed, pass these units to the network layer, and ensure that the units arrive correctly at the destination. The session layer allows different nodes/users to establish sessions between them. The major services offered by the session layer include keeping track of whose turn is to transmit and preventing two parties from attempting the same critical operation at the same time. In order to ensure that various nodes with different data representations can communicate, standard data structures are applied at the presentation layer before passing a user data to the session layer.

- *User application layer:* The application layer uses a variety of protocols depending on the requested services. For instance the familiar HTTP (HyperText Transfer Protocol) is the protocol used when one attempts to access the internet. Other popular protocols are FTP (File Transfer Protocol) and SMTP (Simple Mail Transfer Protocol). The latter is what we use for our e-mail exchanges.

The content of this book is limited to the details of the demodulation block in Figure 1.1 and its counterpart (the modulator) at the transmitter. The term *modem* is often used to refer to the combination of modulator and demodulator blocks in a communication system.

1.2 A Brief History of Modems

The history of modems may be dated back to the era of teletype and telegraph data transmission as early as 1919. Transmission in this era was mostly through wired lines and performed at baseband, i.e., no modulation was performed. Data rates were in the range of 100 bits/s or lower.

The invention of digital computers in the early 1950s and the resulting applications led to significant interest in higher speed data transmission over telephone lines. The earliest telephone-line modems used frequency-shift keying (FSK) to achieve speeds of 300 bits/s (Bell 103) and later 1200 bits/s (Bell 202). The concepts of phase-shift keying (PSK) and quadrature amplitude modulation (QAM) were also developed in 1950s and 1960s and by the end of 1960s modems that were capable of transmitting information at a rate of 2400 bits/s (Bell 201) or higher had been commercialized. The Bell 201 achieved a transmission rate of 2400 bits/s using 4-PSK signaling over a bandwidth of 1200 Hz. The Milgo 4400/48 was the first commer-

cialized 4800 bits/s modem. It was based on 8-PSK signaling and had a bandwidth span of 1600 Hz. To be able to cover such bandwidth (which was considered wide for voice grade telephone lines, at the time), the Milgo 4400/48 modems were equipped with manually adjustable equalizers. The development of automatic equalizers in 1960s soon led to the introduction of modems with bandwidth of 2400 Hz, essentially the full telephone line bandwidth. In 1971, the Codex 9600C which could transmit at a rate of 9600 bits/s, using 16-QAM signaling, was introduced to the market. The rest of 1970s was spent by researchers to fine-tune the latter developments. Sophisticated carrier recovery and timing recovery loops, pulse shaping filters, equalization methods, and channel coding techniques were developed during these years.

Further developments on telephone line modems, achieving transmission rates greater than 9600 bits/s, occurred in 1980s and 1990s. Telephone line modems that could transmit at 56000 bits/s were marketed by early 2000s.

Some developments on radio (wireless) modems were also made in parallel with telephone line modems, though, initially, at a much lower pace. Compared to telephone line modems literature, the reports on radio modems prior to 1980 are rather sporadic. It appears that most of the theoretical developments on modems have been done in the context of telephone line channels. However, the developments are also applicable to radio channels and, thus, are widely deployed in radio modems as well.

Widespread development of radio modems was started with the introduction of cellular and cordless telephone networks in early and mid 1980s. The first-generation of the wireless devices was based on analog technology with FM modulation. The new era of digital communications started in late 1980s and early 1990s as second-generation (2G) wireless cellular phone began to gain popularity. Digital transmission was the enabling technology for integration of speech and data services. Data services include facsimile, paging, local area networks and internet access. The local area networks and internet access, in particular, have been the driving force for further development of the digital wireless systems in 1990s and more recent years. The third generation (3G) of wireless communication systems that emerged by the turn of the millennium includes a variety of standards, such as, IEEE 802.15/WPAN (Wireless Personal Area Networks), IEEE 802.11/WiFi/WLAN (Wireless Local Area Networks), and IEEE 802.16/WiMAX/WMAN (Wireless Metropolitan Area Networks). They cover variety of network sizes and support data rates in the orders of Mega bits per second and higher. Variety of signaling/modulation techniques have been considered in these standards, ranging from binary phase-shift keying (BPSK) to large QAM constellations, as well as single carrier and multicarrier modulations. Moreover, since wireless channels, in general, should be shared among different users, variety of accessing methods such as time-division multiple access (TDMA), frequency-division multiple access (FDMA), and code-division multiple access (CDMA) have been consid-

ered. More recent advancements are MIMO communications where space diversity adds to time, frequency and code diversity to improve on the link reliability as well as transmission rate. Some of the recent standards (e.g. IEEE 802.11n and 802.16e) suggest data rates in the range of 100 Mega bits per second or higher. Another interesting emerging technology is cognitive radio. Cognitive radios, yet to be developed, will be intelligent systems that will be aware of spectral activities in their environment and pick their transmission band and rate according to available resources.

1.3 Signal Processing in Modems

This text presents a collection of signal processing techniques that are commonly used in implementation of data modems and, thus, are not tied to any specific standard. Figures 1.2 and 1.3, respectively, present detailed block diagrams of the transmitter and receiver of a modem.

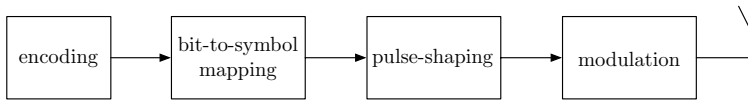


Figure 1.2: The transmitter part of a data modem.

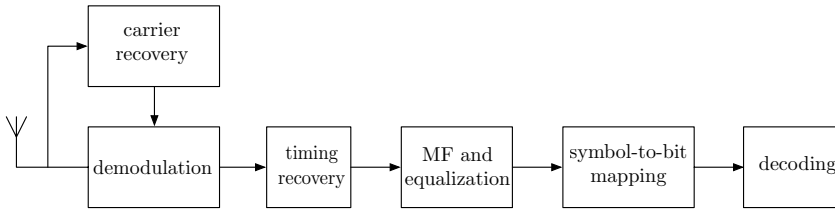


Figure 1.3: The receiver part of a data modem.

The transmitter blocks and their functional role are as follows:

- *Encoding:* In order to implement a communication system robust against channel impairments, redundant bits are added to raw data bits. The redundancies are considered at the receiver to correct errors that may have occurred as a result of the channel distortion and/or noise. As noted above, the process of adding redundant bits to the raw data bits is called channel coding. In this text, to keep our attention on signal processing aspects of modems only, the topic of coding is completely left out. It is believed that this treatment is appropriate as coding and signal processing in modems are mostly orthogonal. In

addition, many excellent texts on the topic of coding are available, e.g., (Lin and Costello, 2004).

- *Bit-to-symbol mapping:* Information bits (usually, after channel coding) are mapped to a sequence of symbols from a pre-selected alphabet, for transmission. This allows one to bundle a few bits in each symbol and thus increase the bandwidth efficiency of transmission channel, i.e., transmit more bits per unit of bandwidth. It is also possible to combine coding and bit-to-symbol mapping to improve on the system performance. This is called *trellis coding*. Again, to limit our discussion to signal processing aspects of modems, this text only considers simple mapping of (possibly, coded) bits to symbols.

  An alphabet, in general, is a set of complex numbers. The common alphabets are (i) a set of equally distanced points on a circle as in Figure 1.4(a); and (ii) a set of points in a square-grid as in Figure 1.4(b). These are called constellations. Also shown in Figures 1.4(a) and (b) are one possible choice of bit-to-symbol mapping for each alphabet. The alphabet shown in Figure 1.4(a) is called phase-shift keying (PSK), because information bits are mapped into a set of phase angles; here, there are eight phase angles $2\pi k/8$, for $k = 0, 1, \cdots, 7$. The alphabet shown in Figure 1.4(b) is called quadrature amplitude modulation (QAM), for obvious reasons that will become clear as we go through later chapters of this text.

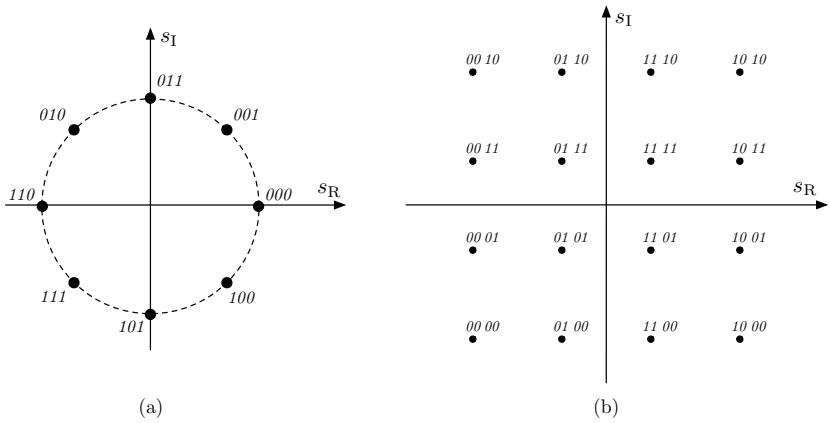


Figure 1.4: Examples of symbol alphabets; (a) phase-shift keying (PSK), and (b) quadrature amplitude modulation (QAM). s_R and s_I are real and imaginary parts of a complex-valued symbol s.

- *Pulse-shaping:* In any communication system, it is always desirable (and necessary) to minimize the transmission bandwidth. The pulse-

shaping is used to limit the bandwidth of symbol sequences. This, effectively, is a filtering operation. Pulse-shaping filters are covered in Chapters 3 and 4.

- *Modulation:* The output of the pulse-shaping filter is a baseband signal which usually is not appropriate for transmission. In wireless channels, one often needs to shift the baseband signal to an assigned RF channel (a passband) before directing it to an antenna for transmission. The process of shifting a baseband signal to a radio channel is called modulation. As was noted above, modulation may be done in a number of steps through a superheterodyne structure, i.e., by first moving the spectrum from baseband to one or more IF bands and then to the desired RF band. When the latter operations are performed in digital domain, on an SDR platform, as the signal is moved to higher frequency bands, its sampling rate should be increased proportionately. Hence, modulation in an SDR is effectively a multirate signal processing. Chapter 5 is devoted to the theory of multirate signal processing. Examples that show the effective application of multirate techniques in SDRs, including modulation and demodulation, are also presented in this chapter.

On the other hand, the receiver blocks and their functional role are as follows:

- *Demodulation:* This is the dual of the modulation block at the transmitter. The goal is to shift the desired signal from the RF band to the baseband. Here, also, following the superheterodyne principle, the demodulation may be performed by shifting the spectrum of the desired signal first to an IF band, and then to the baseband. At various stages of this process, filters are used to remove undesired signals and noise. Particularly, a filter that is matched to the transmitter pulse-shaping filter is used at the last stage of the demodulation process to clean up the acquired baseband signal and to maximize the signal-to-noise ratio (SNR) at this final stage. This, which is called matched filter (MF), may also be combined with a channel equalizer whose role is to combat channel distortion (see below).

- *Carrier Recovery:* A demodulation operation is always more complex than its modulation counterpart, because the receiver has to tune the carrier that it uses for demodulation to match that of the transmitter. The receiver begins with a local free running oscillator. A mechanism, called phase-locked loop (PLL), is then used to fine-tune the local oscillator to match the frequency and phase of the carrier of the received signal. The subject of PLL is covered in Chapter 8 and a number of practical carrier recovery and tracking algorithms are presented in Chapter 9.

- *Timing Recovery:* The receiver must also synchronize itself with the rate of data symbols embedded in the incoming signal. Timing information is often extracted by noting that digital data signals are cyclostationary, meaning that their statistical properties exhibit some cyclic properties. The period of these cycles matches the symbol rate. Hence, a PLL that locks to this cyclic behavior can be used for symbol timing synchronization and thus symbol timing recovery. The topic of timing recovery is covered in Chapter 10.

- *Equalization:* Often, the presence of channel introduces distortion in data signals. In wireless channels, the main source of distortion is multipath. When data rate is high, the presence of multipath results in interference among neighboring data symbols. This is called intersymbol interference (ISI). To combat ISI, a filter that compensates for the distortion introduced by channel is used. This is called equalization. An equalizer should be trained to learn about the channel and to adjust its parameters to realize a filter equivalent to the inverse of the channel. The theory of adaptive equalizers is developed in Chapter 11. This chapter also provides some tips on additional mechanisms that are needed for other synchronization tasks for correct operation of modems. This chapter as well as Chapters 9 and 10 benefit significantly from the theory of adaptive filters that is presented in Chapter 7.

Chapter 2

Fourier Analysis and Linear Time-Invariant Systems

In this chapter, we briefly review the basic tools from the theory of signals and linear time-invariant (LTI) systems that are used in the analysis and design of communication systems. An LTI system is characterized in the time domain by its *impulse response* and in the frequency domain by its *transfer function*. Moreover, the impulse response and transfer function of an LTI system are related through *Fourier transform*. At the same time, the Fourier transform is an analysis tool that is used to expand a signal as a sum of spectral components; a finite or infinite set of sinusoidal waveforms. This leads to the notion of *signal spectra*. Since signal spectra and transfer functions are fundamental to the design and analysis of communication systems, most of the discussions in this chapter are related to these topics.

2.1 Fourier Series

According to the theory of Fourier, a periodic signal $x_{T_0}(t)$ with period of T_0 and frequency of $f_0 = 1/T_0$ can be expressed as

$$x_{T_0}(t) = \sum_{n=-\infty}^{\infty} x_n e^{j2\pi n f_0 t} \tag{2.1}$$

where $j = \sqrt{-1}$ and the x_n's are the *Fourier series coefficients* of the signal $x_{T_0}(t)$. The Fourier coefficients x_n's can be evaluated by performing the integral

$$x_n = \frac{1}{T_0} \int_{\alpha}^{\alpha+T_0} x_{T_0}(t) e^{-j2\pi n f_0 t} dt \tag{2.2}$$

where α is an arbitrary constant that determines the start of the period and is chosen such that to simplify the evaluation of the integral.

According to (2.1), the periodic signal $x_{T_0}(t)$ can be synthesized by adding (complex) sine-waves of frequencies 0, $\pm f_0$, $\pm 2f_0$, $\cdots$. The coefficient x_0 which associates with the frequency of 0 is called DC (directly coupled) component of $x_{T_0}(t)$. The frequency f_0 is called the fundamental frequency of $x_{T_0}(t)$, and accordingly x_1 and x_{-1} are referred to as the *fundamental components* of $x_{T_0}(t)$. Accordingly, f_0 is called the fundamental frequency. The remaining terms are referred to as *harmonics*.

When $x_{T_0}(t)$ is real-valued, we have

$$\begin{aligned} x_{-n} &= \frac{1}{T_0}\int_{\alpha}^{\alpha+T_0} x_{T_0}(t)e^{j2\pi n f_0 t}dt \\ &= \frac{1}{T_0}\left[\int_{\alpha}^{\alpha+T_0} x_{T_0}(t)e^{-j2\pi n f_0 t}dt\right]^* \\ &= x_n^* \end{aligned} \tag{2.3}$$

which implies

$$|x_n| = |x_{-n}| \quad \text{and} \quad \angle x_n = -\angle x_{-n}. \tag{2.4}$$

That is, the magnitude of x_n is an even function of n and its phase is an odd function of n. In other words, the Fourier series coefficients of a real-valued signal $x_{T_0}(t)$ have *Hermitian symmetry*.

The Fourier series expansion of (2.1) is known as the *exponential Fourier series* and is applicable to both real- and complex-valued periodic signals. When $x_{T_0}(t)$ is real-valued, another form of Fourier series, known as *trigonometric Fourier series*, may be applied. In the trigonometric Fourier series, $x_{T_0}(t)$ is expanded as

$$x_{T_0}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n \cos(2\pi n f_0 t) + b_n \sin(2\pi n f_0 t)]. \tag{2.5}$$

This can be obtained by using the Euler's formula and substituting

$$e^{j2\pi n f_0 t} = \cos(2\pi n f_0 t) + j\sin(2\pi n f_0 t) \tag{2.6}$$

in (2.1), and defining

$$\begin{aligned} a_n &= x_n + x_{-n} \\ &= 2\Re\{x_n\} \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} b_n &= j(x_n - x_{-n}) \\ &= -2\Im\{x_n\}. \end{aligned} \tag{2.8}$$

Also, applying the Euler's formula to expand $e^{-j2\pi nf_0t}$ in (2.2) and using (2.7) and (2.8), we obtain

$$a_n = \frac{2}{T_0}\int_{\alpha}^{\alpha+T_0} x_{T_0}(t)\cos(2\pi nf_0t)dt \tag{2.9}$$

and

$$b_n = \frac{2}{T_0}\int_{\alpha}^{\alpha+T_0} x_{T_0}(t)\sin(2\pi nf_0t)dt. \tag{2.10}$$

Example 2.1:

Find the Fourier series coefficients of the signal $x_{T_0}(t)$ shown in Figure 2.1.

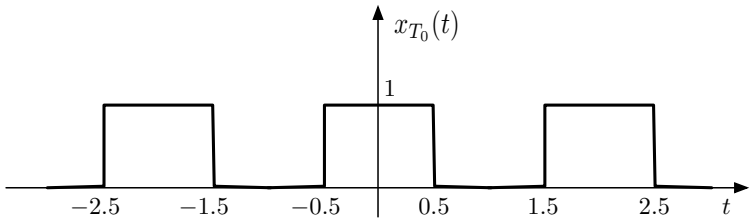


Figure 2.1: An example of a periodic signal.

Solution:

Here, $T_0 = 2$ and, thus, $f_0 = 0.5$. Choosing $\alpha = -T_0/2 = -1$, we have

$$\begin{aligned} x_n &= \frac{1}{2}\int_{-1}^{1} x_{T_0}(t)e^{-j2\pi nf_0t}dt \\ &= \frac{1}{2}\int_{-0.5}^{0.5} e^{-j\pi nt}dt \\ &= \frac{1}{-j2\pi n}\left(e^{-j\pi n/2} - e^{j\pi n/2}\right) \\ &= \frac{1}{2}\frac{\sin(\pi n/2)}{\pi n/2} \\ &= \frac{1}{2}\text{sinc}(n/2) \end{aligned} \tag{2.11}$$

where $\text{sinc}(x)$ is defined as

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{2.12}$$

From (2.11) and recalling (2.7) and (2.8), we obtain

$$a_n = \text{sinc}(n/2) \quad \text{and} \quad b_n = 0.$$

Hence,

$$x_{T_0}(t) = \frac{1}{2} + \sum_{n=1}^{\infty}\text{sinc}(n/2)\cos(\pi nt). \tag{2.13}$$

Figure 2.2 presents a set of approximations of $x_{T_0}(t)$ over one period and for various number of the terms in (2.13).

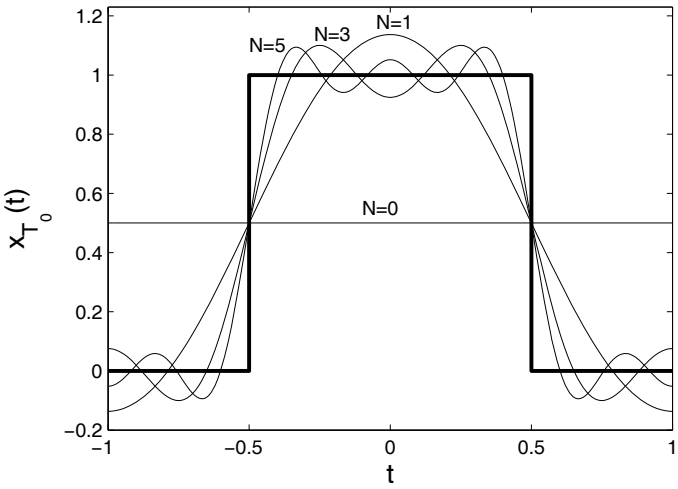


Figure 2.2: Fourier series approximations to rectangular pulse of Figure 2.1.

The magnitude spectra of $x_{T_0}(t)$ are presented in Figure 2.3.

Note that, in the above example, $x_{T_0}(t)$ has a rather wide spectrum, i.e., the signal energy remains significant over a large number of harmonics. The presence of many harmonics here is a direct consequence of the sharp edges in the rectangular waveform. Pulse shapes/signals with smoother edges have spectra that are confined within a more limited band. Examples are *raised-cosine* pulse shapes that are discussed in the next chapter for data transmission over band-limited channels.

2.2 Fourier Transform

The Fourier transform can be thought as an extension to the Fourier series when the period $T_0 \to \infty$. In this case, the Fourier series coefficients are defined over continuous points on the frequency axis. This is because in the Fourier series, the spacing between spectral lines (i.e., x_n's) is $f_0 = 1/T_0$ and when $T_0 \to \infty$, $f_0 \to 0$. Also, because of the divide by T_0 in (2.2), x_n approaches zero, assuming that the integral on the right-hand side of (2.2) is finite. To resolve this problem, the Fourier transform of a non-periodic signal is defined as

$$\mathcal{F}[x(t)] = X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{2.14}$$

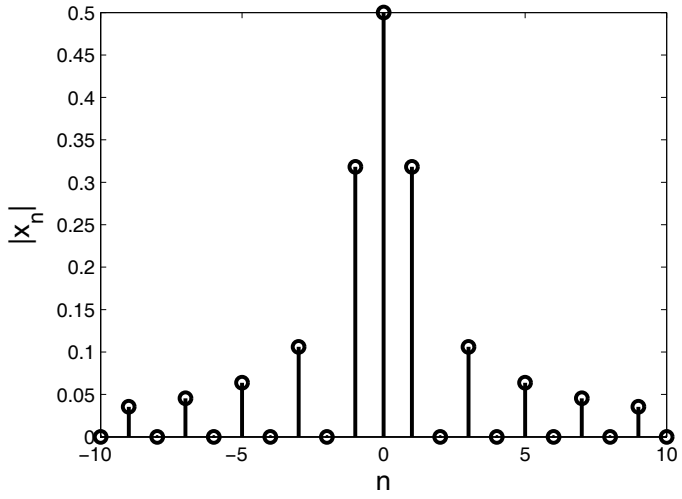


Figure 2.3: Magnitude spectra of the rectangular pulse of Figure 2.1.

and, accordingly, the inverse Fourier transform of $X(f)$ is $x(t)$ and is given by

$$\mathcal{F}^{-1}[X(f)] = x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df. \tag{2.15}$$

Similar to (2.1), we may have the following interpretation of (2.15). A signal $x(t)$ can be constructed by adding an infinite set of sine-waves. Here, the sine-wave components are the continuous in frequency set $[X(f)df]e^{j2\pi ft}$, for $-\infty < f < \infty$. When $x(t)$ is real-valued, the positive and negative frequency components $[X(f)df]e^{j2\pi ft}$ and $[X(-f)df]e^{-j2\pi ft}$ are complex conjugate of each other. This implies that

$$X(-f) = X^*(f). \tag{2.16}$$

The equality (2.16) is the dual of (2.3). Moreover, when $x(t)$ is real-valued

$$\begin{aligned} \mathcal{F}[x(-t)] &= \int_{-\infty}^{\infty} x(-t) e^{-j2\pi ft} dt \\ &= \int_{-\infty}^{\infty} x(t) e^{j2\pi ft} dt \\ &= \left[\int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \right]^* \\ &= X^*(f) \end{aligned} \tag{2.17}$$

where the second line is obtained by replacing t with $-t$.

When $x(t)$ is complex-valued, (2.16) no longer holds, and

$$\mathcal{F}[x^*(-t)] = X^*(f). \tag{2.18}$$

Clearly, when $x(t)$ is real-valued, (2.18) reduces to (2.17).

The Fourier transform has a number of properties which prove very useful in the study of signals and LTI systems. The most important properties of the Fourier transform are summarized as follows.

1. **Linearity:** The Fourier transform of a linear combination of two or more signals is the linear combination of the corresponding Fourier transforms, e.g., for constants α and β,

$$\mathcal{F}[\alpha x(t) + \beta y(t)] = \alpha\mathcal{F}[x(t)] + \beta\mathcal{F}[y(t)]. \tag{2.19}$$

2. **Duality:** If $X(f) = \mathcal{F}[x(t)]$, then

$$\mathcal{F}[X(t)] = x(-f). \tag{2.20}$$

3. **Conjugate symmetry:** When $x(t)$ is real-valued,

$$X(f) = X^*(-f). \tag{2.21}$$

4. **Modulation:** If $X(f) = \mathcal{F}[x(t)]$, then

$$\mathcal{F}[e^{j2\pi f_0 t}x(t)] = X(f - f_0). \tag{2.22}$$

   In other words, multiplication of a signal $x(t)$ by a complex sinusoidal $e^{j2\pi f_0 t}$ in the time domain corresponds to a frequency shift f_0 in the frequency domain.

   Also, using (2.22), and recalling that $\cos(2\pi f_0 t) = \frac{1}{2}(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t})$, we obtain

$$\mathcal{F}[x(t)\cos(2\pi f_0 t)] = \frac{1}{2}[X(f - f_0) + X(f + f_0)]. \tag{2.23}$$

5. **Time shifting:** If $\mathcal{F}[x(t)] = X(f)$, then

$$\mathcal{F}[x(t - t_0)] = e^{-j2\pi f t_0}X(f). \tag{2.24}$$

   This which may be thought as a dual of the modulation property, implies that a shift in the time domain results in a linear phase shift in the frequency domain.

6. **Time scaling:** If $\mathcal{F}[x(t)] = X(f)$ and $a \neq 0$ is a constant, then

$$\mathcal{F}[x(at)] = \frac{1}{|a|}X\left(\frac{f}{a}\right). \tag{2.25}$$

   This property has the following interpretation. An expansion in the time domain results in a contraction in the frequency domain, and vice versa.

7. **Differentiation:** If $\mathcal{F}[x(t)] = X(f)$, then

$$\mathcal{F}\left[\frac{d^n x(t)}{dt^n}\right] = (j2\pi f)^n X(f). \tag{2.26}$$

8. **Convolution:** Convolution in the time domain is equivalent to multiplication in the frequency domain:

$$\mathcal{F}[x(t) \star y(t)] = X(f)Y(f). \tag{2.27}$$

Similarly, multiplication in the time domain is equivalent to convolution in the frequency domain:

$$\mathcal{F}[x(t)y(t)] = X(f) \star Y(f). \tag{2.28}$$

Note: If you need to refresh your memory on the definition and the significance of convolution, this appears in the Section 2.3.

9. **Parseval's relation:** If $\mathcal{F}[x(t)] = X(f)$ and $\mathcal{F}[y(t)] = Y(f)$, then

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt = \int_{-\infty}^{\infty} X(f)Y^*(f)df. \tag{2.29}$$

For the particular case when $y(t) = x(t)$, this simplifies to

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{2.30}$$

which is known as *Rayleigh's relation.*

Table 2.1 presents a list of the frequently used Fourier transform pairs. In this table, $\delta(t)$ is the *delta* function which by definition is non-zero only for $t = 0$ and satisfies

$$\int_{-\infty}^{\infty} \delta(t)dt = 1. \tag{2.31}$$

Also, $u(t)$ and $\text{sgn}(t)$ are the *step* and *signum* functions which, respectively, are defined as

$$u(t) = \begin{cases} 0, & t < 0 \\ 0.5, & t = 0 \\ 1, & t > 0 \end{cases} \tag{2.32}$$

and

$$\text{sgn}(t) = \begin{cases} -1, & t < 0 \\ 0, & t = 0 \\ 1, & t > 0. \end{cases} \tag{2.33}$$

Other functions that appear in Table 2.1 are the *rectangular* and *triangular* pulse shapes $\Pi(t)$ and $\Lambda(t)$, respectively. These are defined as

$$\Pi(t) = \begin{cases} 1, & |t| < 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{2.34}$$

and

$$\Lambda(t) = \begin{cases} 1+t, & -1 < t \leq 0 \\ 1-t, & 0 < t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.35}$$

Table 2.1: Table of Fourier transform pairs.

	$x(t)$	$X(f)$
1.	$\delta(t)$	1
2.	1	$\delta(f)$
3.	$\delta(t-t_0)$	$e^{-j2\pi f t_0}$
4.	$e^{j2\pi f_0 t}$	$\delta(f-f_0)$
5.	$\cos(2\pi f_0 t)$	$\frac{1}{2}[\delta(f-f_0)+\delta(f+f_0)]$
6.	$\sin(2\pi f_0 t)$	$\frac{1}{2j}[\delta(f-f_0)-\delta(f+f_0)]$
7.	$e^{-\alpha t}u(t), \quad \alpha > 0$	$\frac{1}{\alpha+j2\pi f}$
8.	$t^n e^{-\alpha t}u(t), \quad \alpha > 0$	$\frac{n!}{(\alpha+j2\pi f)^{n+1}}$
9.	$e^{-\alpha\lvert t\rvert}, \quad \alpha > 0$	$\frac{2\alpha}{\alpha^2+(2\pi f)^2}$
10.	$e^{-\pi t^2}$	$e^{-\pi f^2}$
11.	$u(t)$	$\frac{1}{2}\delta(f)+\frac{1}{j2\pi f}$
12.	$\text{sgn}(t)$	$\frac{1}{j\pi f}$
13.	$\Pi(t)$	$\text{sinc}(f)$
14.	$\text{sinc}(t)$	$\Pi(f)$
15.	$\Lambda(t)$	$\text{sinc}^2(f)$
16.	$\text{sinc}^2(t)$	$\Lambda(f)$
17.	$\sum_{n=-\infty}^{\infty}\delta(t-nT_0)$	$f_0\sum_{n=-\infty}^{\infty}\delta(f-nf_0), \quad f_0=\frac{1}{T_0}$

Most of the pairs in Table 2.1 can be derived straightforwardly by direct application of the definition of Fourier transform and its properties. For instance,

$$\begin{aligned} \mathcal{F}[\delta(t)] &= \int_{-\infty}^{\infty}\delta(t)e^{-j2\pi ft}dt \\ &= \int_{-\infty}^{\infty}\delta(t)dt = 1 \end{aligned} \tag{2.36}$$

where the second line follows since $\delta(t)$ is non-zero only when $t=0$, and for $t=0$, $e^{-j2\pi ft}=1$. Once we have this result, line 2 in the table follows from the duality property. Moreover, applying the time shifting property to $\mathcal{F}[\delta(t)]=1$, we obtain $\mathcal{F}[\delta(t-t_0)]=e^{-j2\pi ft_0}$ (line 3 in the table).

And using the duality property, from this result, we obtain $\mathcal{F}\left[e^{j2\pi f_0 t}\right] = \delta(f - f_0)$ (line 4 in the table).

Another important Fourier pair in Table 2.1 is the one in line 17. This can be shown as follows. We recall that a periodic signal $x_{T_0}(t)$ can be expanded as

$$x_{T_0}(t) = \sum_{n=-\infty}^{\infty} x_n e^{j2\pi n f_0 t}. \tag{2.37}$$

Using the linear property of the Fourier transform and line 4 of Table 2.1, from (2.37), we obtain

$$X_{T_0}(f) = \sum_{n=-\infty}^{\infty} x_n \delta\left(f - n f_0\right). \tag{2.38}$$

This shows that the Fourier transform of a periodic signal with period of $T_0 = 1/f_0$ consists of impulses at the multiples of the fundamental frequency f_0. A special case of this is when

$$x_{T_0}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0). \tag{2.39}$$

In this case, we have

$$x_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \delta(t) e^{-j2\pi f_0 t} dt = \frac{1}{T_0} = f_0. \tag{2.40}$$

Substituting this result in (2.38), leads to the Fourier pair in line 17 of Table 2.1. This result is used in Chapter 4 in the derivation of the sampling theorem.

Other important Fourier pairs in Table 2.1 are those in lines 13 to 16. Line 13 can be shown simply by direct evaluation of the Fourier transform of $\Pi(t)$. Line 14, then, follows by invoking the duality property of the Fourier transform. Derivations of lines 15 and 16 are presented after reviewing the convolution integral in the next section.

2.3 Linear Time-Invariant Systems

A system, in general, is characterized by a rule (or a set of rules) that determine how the system input and output are related. Let $x(t)$ and $y(t)$ denote the system input and output, respectively, and $T[\cdot]$ be the rule relating them, viz. $y(t) = T[x(t)]$. A system $T[\cdot]$ is *linear* if for any pair of inputs $x_1(t)$ and $x_2(t)$ and arbitrary constants α and β,

$$T[\alpha x_1(t) + \beta x_2(t)] = \alpha T[x_1(t)] + \beta T[x_2(t)]. \tag{2.41}$$

Moreover, if the system $y(t) = T[x(t)]$ is *time-invariant,* then

$$y(t - t_0) = T[x(t - t_0)]. \tag{2.42}$$

A system that satisfies both (2.41) and (2.42) is called *linear time-invariant* (LTI). Many systems in practice, including most of the communication systems, are LTI. The most important property of the LTI systems is that any LTI system is completely characterized by its impulse response. This is formalized through the convolution integral.

2.3.1 Convolution integral

Consider an LTI system with the impulse response $h(t)$. Also, consider the arbitrary signal $x(t)$ shown in Figure 2.4. The signal $x(t)$ may be partitioned into a set of pulses each of width Δt as in Figure 2.4. For a small Δt, each of these pulses may be approximated by an impulse with the size of $x(t_n)\Delta t$. Accordingly, we have

$$x(t) \approx \sum_{n=-\infty}^{\infty} [x(t_n)\Delta t]\delta(t - t_n). \tag{2.43}$$

Passing both sides of (2.43) through an LTI system $h(t)$ and recalling the linearity and time shifting properties, we obtain

$$y(t) \approx \sum_{n=-\infty}^{\infty} [x(t_n)\Delta t]h(t - t_n). \tag{2.44}$$

Replacing t_n by τ and Δt by $d\tau \to 0$, and, accordingly, the summation by an integral, we obtain

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau. \tag{2.45}$$

This which is called *convolution integral* can also be rearranged as

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau. \tag{2.46}$$

The notation $y(t) = x(t) \star h(t) = h(t) \star x(t)$ is often used instead of the integral (2.45) or (2.46).

Example 2.2:

Evaluate the convolution $\Pi(t) \star \Pi(t)$ and show that it is equal to $\Lambda(t)$.

Solution:

$$\Pi(t) \star \Pi(t) \quad = \quad \int_{-\infty}^{\infty} \Pi(\tau)\Pi(t - \tau)d\tau$$

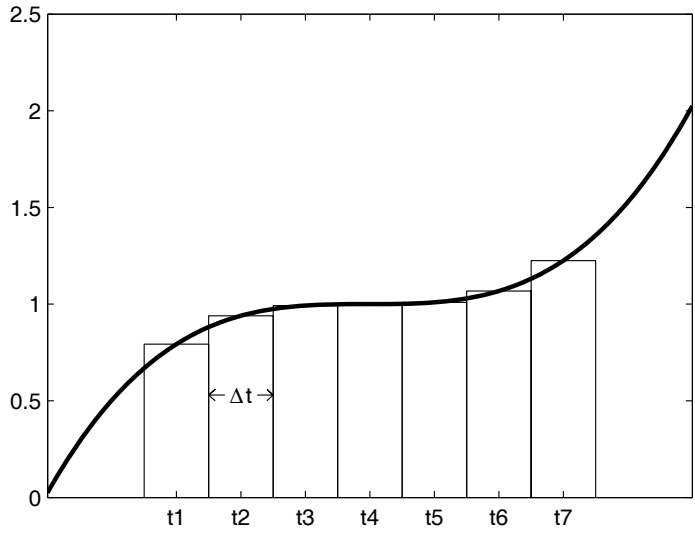


Figure 2.4: An arbitrary signal and its approximation by a sequence of rectangular pulses.

$$= \begin{cases} \int_{-0.5}^{t+0.5} dt, & -1 < t \leq 0 \\ \int_{t-0.5}^{0.5} dt, & 0 < t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.47}$$

$$= \begin{cases} 1+t, & -1 < t \leq 0 \\ 1-t, & 0 < t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.48}$$

2.3.2 Transfer function

From (2.45) and (2.46), we note that the input-output relationship of an LTI system is fully specified by its impulse response. In the particular case where $x(t) = e^{j2\pi ft}$, using (2.46), we obtain

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) e^{j2\pi f(t-\tau)} d\tau \\ &= \left[\int_{-\infty}^{\infty} h(\tau) e^{-j2\pi f\tau} d\tau \right] e^{j2\pi ft} \\ &= H(f)x(t) \end{aligned} \tag{2.49}$$

where $H(f) = \int_{-\infty}^{\infty} h(\tau) e^{-j2\pi f\tau} d\tau$ is the Fourier transform of $h(t)$. The significance of this result is that it shows when a complex sinusoid $e^{j2\pi ft}$ is applied to an LTI system, the LTI system acts like an amplifier or attenuator

with the gain of $H(f)$. Accordingly, $H(f)$ is referred to as the system *transfer function*.

On the other hand, according to the inverse Fourier transform relation (2.15), $x(t)$ can be constructed by adding the sinusoidal components $[X(f)df]e^{j2\pi ft}$, for $-\infty < f < \infty$. Applying the above result, the system response to each of these components is $H(f)[X(f)df]e^{j2\pi ft}$. Moreover, adding these components gives the system output, viz.,

$$y(t) = \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft}df. \tag{2.50}$$

From (2.50), we obtain

$$Y(f) = H(f)X(f) \tag{2.51}$$

which implies that $\mathcal{F}[h(t) \star x(t)] = H(f)X(f)$. This also proves the convolution property of the Fourier transform.

Another interesting observation can be made by studying the response of an LTI system with the transfer function $H(f)$ and real impulse response $h(t)$ to the real-valued sine-wave $A\cos(2\pi ft+\phi)$. Using the Euler's formula, we have

$$A\cos(2\pi ft + \phi) = \frac{A}{2}e^{j(2\pi ft+\phi)} + \frac{A}{2}e^{-j(2\pi ft+\phi)} \tag{2.52}$$

Since the system is linear, the output $y(t)$ is the sum of the responses to the two input components:

$$\begin{aligned} y(t) &= \frac{A}{2}H(f)e^{j(2\pi ft+\phi)} + \frac{A}{2}H(-f)e^{-j(2\pi ft+\phi)} \\ &= \frac{A}{2}H(f)e^{j(2\pi ft+\phi)} + \frac{A}{2}H^*(f)e^{-j(2\pi ft+\phi)} \\ &= \frac{A}{2}|H(f)|e^{j\angle H(f)}e^{j(2\pi ft+\phi)} + \frac{A}{2}|H(f)|e^{-j\angle H(f)}e^{-j(2\pi ft+\phi)} \\ &= A|H(f)|\cos(2\pi ft + \phi + \angle H(f)) \end{aligned} \tag{2.53}$$

where the second identity follows since for a system with real impulse response $h(t)$, $H(-f) = H^*(f)$. As seen, the system output is a sine-wave with the same frequency, but with modified amplitude and phase according to the transfer function $H(f)$.

Example 2.3:

Using the result of Example 2.2, show the correctness of lines 15 and 16 of Table 2.1.

Solution:

Since $\mathcal{F}[\Pi(t)] = \text{sinc}(f)$ and $\Lambda(t) = \Pi(t) \star \Pi(t)$, we get

$$\mathcal{F}[\Lambda(t)] = \mathcal{F}[\Pi(t)]\mathcal{F}[\Pi(t)] = \text{sinc}(f)\text{sinc}(f) = \text{sinc}^2(f).$$

This is line 15 of Table 2.1.

Line 16 of Table 2.1 follows from this by invoking the duality property of the Fourier transform.

2.4 Energy and Power Spectral Density

2.4.1 Energy-type signals

A signal $x(t)$ is called an *energy-type signal* if

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt \tag{2.54}$$

is finite. The *energy spectral density* of an energy-type signal is defined as

$$\Phi_{xx}(f) = |X(f)|^2. \tag{2.55}$$

In light of the Rayleigh's relation (2.30), $\Phi_{xx}(f)$ gives the distribution of energy of the energy signal $x(t)$ around the frequency f. On the other hand, the *autocorrelation function* of the energy signal $x(t)$ is defined as

$$\begin{aligned} \phi_{xx}(\tau) &= \int_{-\infty}^{\infty} x(t+\tau)x^*(t)dt \\ &= x(\tau) \star x^*(-\tau). \end{aligned} \tag{2.56}$$

Taking Fourier transform on both sides of (2.56), and using (2.18) and convolution property of the Fourier transform, one finds that

$$\Phi_{xx}(f) = \mathcal{F}[\phi_{xx}(\tau)]. \tag{2.57}$$

This suggests an alternative method of obtaining the energy spectral density of an energy signal: evaluate the autocorrelation function of $x(t)$, using (2.56) and, then, apply Fourier transform to the result.

2.4.2 Power-type signals

When E_x is not finite, but

$$P_x = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 dt \tag{2.58}$$

is finite, $x(t)$ is called *power-type signal*. For a power-type signal $x(t)$, we define the *time-average autocorrelation function*

$$\phi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)x^*(t)dt \tag{2.59}$$

and the *power-spectral density*

$$\Phi_{xx}(f) = \mathcal{F}[\phi_{xx}(\tau)]. \tag{2.60}$$

Note that we use similar notations to refer to the autocorrelation and spectral density of energy-type and power-type signals.

2.4.3 Random signals

When a signal $x(t)$ is a sample realization of a random process $X(t)$, the autocorrelation function, for a pair of time variables t and τ, is defined as

$$\phi_{xx}(t+\tau, t) = E[x(t+\tau)x^*(t)] \tag{2.61}$$

where $E[\cdot]$ denotes the ensemble average over infinite number of realizations of $X(t)$.

When a random process $X(t)$ is stationary, the ensemble average (2.61) is independent of t. In such a case, $\phi_{xx}(t+\tau, t)$ is replaced by $\phi_{xx}(\tau)$ and accordingly (2.61) is replaced by

$$\phi_{xx}(\tau) = E[x(t+\tau)x^*(t)]. \tag{2.62}$$

Moreover, when the random process $X(t)$ satisfies certain conditions, the ensemble average (2.62) may be replaced by a time average such as (2.59) over one realization of $X(t)$. In such a case, we say $X(t)$ is an ergodic process.

When a signal is non-stationary, Wiener-Khinchin theorem suggests the following formula for computation of the power spectral density of $X(t)$:

$$\Phi_{xx}(f) = \mathcal{F}[\bar{\phi}_{xx}(\tau)] \tag{2.63}$$

where

$$\bar{\phi}_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \phi_{xx}(t+\tau, t)dt \tag{2.64}$$

i.e., the time-average of $\phi_{xx}(t+\tau, t)$.

It follows from (2.63) that

$$\bar{\phi}_{xx}(\tau) = \mathcal{F}^{-1}[\Phi_{xx}(f)] = \int_{-\infty}^{\infty} \Phi_{xx}(f)e^{j2\pi f\tau}df. \tag{2.65}$$

Letting $\tau = 0$, this reduces to

$$\bar{\phi}_{xx}(0) = \int_{-\infty}^{\infty} \Phi_{xx}(f)df. \tag{2.66}$$

Also, using (2.61) and (2.64), we get

$$\bar{\phi}_{xx}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} E[|x(t)|^2]dt. \tag{2.67}$$

Combining (2.66) and (2.67), we obtain

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} E[|x(t)|^2]dt = \int_{-\infty}^{\infty} \Phi_{xx}(f)df. \tag{2.68}$$

This which has similarity with the Rayleigh's relation (2.30) shows that the average power of $x(t)$ over time and its various realizations, i.e., its ensemble average, can be calculated by integrating its spectrum. Because of this similarity, we refer to (2.68) as *Rayleigh's relation for random processes.*

An example of non-stationary signals

Most of the discussion throughout this text deals with signals that originate from digital data sequences. A realization of such signals finds the following form

$$x(t) = \sum_{n=-\infty}^{\infty} s[n]p(t - nT_b) \tag{2.69}$$

where $s[n]$'s are information symbols, $p(t)$ is a pulse-shape, and T_b is the symbol/baud interval. Here, one finds that $\phi_{xx}(t+\tau, t)$ is not independent of t, thus, $X(t)$ is non-stationary. However, $\phi_{xx}(t+\tau, t)$ is periodic in t with a period of T_b. Accordingly, $X(t)$ is referred to as a *cyclostationary* process.

Here, we should use (2.63) to obtain the power spectral density of $x(t)$. Using the periodicity property of $\phi_{xx}(t+\tau, t)$, we evaluate its mean over one period, viz.,

$$\bar{\phi}_{xx}(\tau) = \frac{1}{T_b} \int_{-T_b/2}^{T_b/2} \phi_{xx}(t+\tau, t)dt. \tag{2.70}$$

Substituting (2.69) in (2.61) and assuming that the pulse-shape $p(t)$ is real-valued, we obtain

$$\phi_{xx}(t+\tau, t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{ss}[n+m, n]p(t+\tau-(n+m)T_b)p(t-nT_b) \tag{2.71}$$

where

$$\phi_{ss}[n+m, n] = E[s[n+m]s^*[n]] \tag{2.72}$$

is the autocorrelation function of the sequence $s[n]$. Assuming that $s[n]$ is stationary, thus $\phi_{ss}[n+m, n]$ is independent of n and hence may be written as $\phi_{ss}[m]$, (2.71) reduces to

$$\phi_{xx}(t+\tau, t) = \sum_{m=-\infty}^{\infty} \phi_{ss}[m] \sum_{n=-\infty}^{\infty} p(t+\tau-(n+m)T_b)p(t-nT_b). \tag{2.73}$$

Next, substituting (2.73) in (2.70), we get

$$\begin{aligned}
\bar{\phi}_{xx}(\tau) &= \sum_{m=-\infty}^{\infty} \phi_{ss}[m] \sum_{n=-\infty}^{\infty} \frac{1}{T_b} \int_{-T_b/2}^{T_b/2} p(t+\tau-(n+m)T_b)p(t-nT_b)dt \\
&= \sum_{m=-\infty}^{\infty} \phi_{ss}[m] \sum_{n=-\infty}^{\infty} \frac{1}{T_b} \int_{nT_b-T_b/2}^{nT_b+T_b/2} p(t+\tau-mT_b)p(t)dt \\
&= \frac{1}{T_b} \sum_{m=-\infty}^{\infty} \phi_{ss}[m] \int_{-\infty}^{\infty} p(t+\tau-mT_b)p(t)dt \\
&= \frac{1}{T_b} \sum_{m=-\infty}^{\infty} \phi_{ss}[m]\phi_{pp}(\tau - mT_b).
\end{aligned} \tag{2.74}$$

Taking Fourier transform on both sides of (2.74), we obtain

$$\begin{aligned}
\Phi_{xx}(f) &= \int_{-\infty}^{\infty} \bar{\phi}_{xx}(\tau) e^{-j2\pi f\tau} d\tau \\
&= \frac{1}{T_b} \sum_{m=-\infty}^{\infty} \phi_{ss}[m] \int_{-\infty}^{\infty} \phi_{pp}(\tau - mT_b) e^{-j2\pi f\tau} d\tau \\
&= \frac{1}{T_b} \sum_{m=-\infty}^{\infty} \phi_{ss}[m] e^{-j2\pi fmT_b} \int_{-\infty}^{\infty} \phi_{pp}(\tau - mT_b) e^{-j2\pi f(\tau - mT_b)} d\tau \\
&= \frac{1}{T_b} \sum_{m=-\infty}^{\infty} \phi_{ss}[m] e^{-j2\pi fmT_b} \int_{-\infty}^{\infty} \phi_{pp}(\tau) e^{-j2\pi f\tau} d\tau \\
&= \frac{1}{T_b} \Phi_{ss}(e^{j2\pi f}) |P(f)|^2 \qquad (2.75)
\end{aligned}$$

where $P(f) = \mathcal{F}[p(t)]$ and

$$\Phi_{ss}(e^{j2\pi f}) = \sum_{m=-\infty}^{\infty} \phi_{ss}[m] e^{-j2\pi fmT_b}. \qquad (2.76)$$

When the data symbols $s[n]$ are independent of one another, have zero mean and variance of $\sigma_s^2 = E[|s[n]|^2]$,

$$\phi_{ss}[m] = \begin{cases} \sigma_s^2, & m = 0 \\ 0, & m \neq 0. \end{cases} \qquad (2.77)$$

Using (2.77), (2.76) reduces to

$$\Phi_{ss}(e^{j2\pi f}) = \sigma_s^2. \qquad (2.78)$$

Substituting (2.78) in (2.75), we get

$$\Phi_{xx}(f) = \frac{\sigma_s^2}{T_b} |P(f)|^2. \qquad (2.79)$$

Example 2.4:

In (2.69), let $s[n]$ be a sequence of independent random binary numbers taking values of $+1$ and -1 with the same probabilities, and $p(t) = \Pi(t/T_b)$. Find the power spectral density $\Phi_{xx}(f)$.

Solution:

Since $P(f) = \mathcal{F}[\Pi(t/T_b)] = T_b \text{sinc}(fT_b)$ and $\sigma_s^2 = 1$, we have

$$\Phi_{xx}(f) = \frac{1}{T_b} \left(T_b \text{sinc}(fT_b)\right)^2 = T_b \text{sinc}^2(fT_b).$$

2.4.4 Passing a random signal through an LTI system

When a signal passes through an LTI system with the transfer function $H(f)$, each frequency component in the signal is affected by an amplifying/attenuating gain which is determined by the value of $H(f)$ at the corresponding frequency. Accordingly, the energy/power spectral density of the system output, $y(t)$, is obtained as

$$\Phi_{yy}(f) = \Phi_{xx}(f)|H(f)|^2. \tag{2.80}$$

2.5 Problems

1. Find the exponential Fourier series and the trigonometric Fourier series of the following periodic functions.

   (a) $x_1(t) = \sum_{n=-\infty}^{\infty} \Pi\left(\frac{t-5n}{2}\right)$.

   (b) $x_2(t) = \sum_{n=-\infty}^{\infty} \Lambda\left(\frac{t-5n}{2}\right)$.

2. Using the trigonometric Fourier series derived in Problem 1, develop a computer program (MATLAB is our preferred choice) to confirm the correctness of the series.

3. Using the tabulated Fourier transform pairs in Table 2.1 and the Fourier transform properties find the Fourier transforms of:

   (a) $x_1(t) = \Pi\left(\frac{t}{2}\right)$.

   (b) $x_2(t) = \Pi\left(\frac{t-1}{2}\right)$.

   (c) $x_3(t) = \text{sinc}(2(t+3)$.

   (d) $x_4(t) = t^2 e^{-3t} u(t)$.

   (e) $x_5(t) = \cos(2\pi t)$.

4. (a) Using the results in line 3 of Table 2.1, find the Fourier transform of $x(t) = \delta(t-t_0) + \delta(t+t_0)$.

   (b) Using the result of Part (a) and the duality property, prove line 5 of Table 2.1.

5. Using the result of line 4 of Table 2.1, prove lines 5 and 6 of the same table.

6. Prove the identity (2.18).

7. Prove line 7 of Table 2.1 and explain why the condition $\alpha > 0$ is required. Also, show that when $\alpha > 0$, $\frac{1}{\alpha - j2\pi f} = \mathcal{F}\left[e^{\alpha t}u(-t)\right]$.

8. Using the result of Problem 7, prove line 9 of Table 2.1.

9. Apply the duality property of the Fourier transform to obtain the following results of Table 2.1.

   (a) Line 4 from line 3.

   (b) Line 14 from line 13.

   (c) Line 16 from line 15.

10. Using line 12 of Table 2.1 and the duality property, find the Fourier transform of $x(t) = \dfrac{1}{t}$.

11. Consider the energy signal $x(t) = \Pi(t)$. Find the energy spectral density of $x(t)$ using (2.55) and (2.57) and show that both lead to the same result.

12. Consider the baseband communication channel shown in Figure 2.5, where $s[n]$ is a sequence of transmitted symbols with

$$\phi_{ss}[m] = \begin{cases} \sigma_s^2, & m = 0 \\ 0, & m \neq 0, \end{cases}$$

   $p(t)$ is the impulse response of the transmit filter, and $c(t)$ is the channel impulse response. Find the power spectral density of the received signal $y(t)$, when

   (a) $c(t) = \delta(t)$ (ideal channel).

   (b) $c(t) = a_0\delta(t - \tau_0) + a_1\delta(t - \tau_1)$ (a 2-path channel).

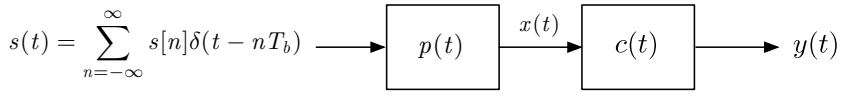


Figure 2.5: A baseband communication channel.

Chapter 3

Digital Transmission Systems

This chapter is devoted to a review of digital transmission systems. Several aspects of digital transmission through band-limited channels are discussed. We begin with introducing the concept of digital pulse amplitude modulation (PAM) and the role of pulse-shaping in limiting the bandwidth of PAM signals. We also discuss the problem of joint design of transmitter and receiver filters for optimum detection. In later sections of the chapter, various carrier modulations are reviewed. The chapter ends with a discussion of baseband equivalent of passband channels.

3.1 Pulse Amplitude Modulation

In baseband digital PAM, a sequence of (coded) information bits is first partitioned into blocks of M bits. For example, when $M = 2$, the sequence 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0 is partitioned as {1,0}, {1,1}, {0,1}, {0,0}, {1,1}, {1,0}. Then, each group is mapped to a symbol $s[n]$ whose amplitude is determined by the bits in the group. In the latter example, since two bits can take four different combinations, the data symbols will have four different amplitudes. Table 3.1 presents two possible mappings of a pair of bits to four levels of PAM symbols. In the direct mapping, the ordered pairs {0,0}, {0,1}, {1,0} and {1,1} are mapped to the amplitude levels -3, -1, $+1$ and $+3$, respectively. In the Gray mapping, on the other hand, the mapping is done so that there is only one bit change in going from each PAM symbol to one of its adjacent neighbors. Since most of the errors caused by noise are likely due to detection of an adjacent symbol in place of the actual transmitted symbol, Gray mapping minimizes the bit-error-rate (BER) of the system.

Mathematically, a PAM signal at the output of the transmitter has the

Table 3.1: Two possible mapping of a pair of bits to PAM symbols.

Direct mapping	Gray mapping	PAM symbols
0,0	0,0	-3
0,1	0,1	-1
1,0	1,1	$+1$
1,1	1,0	$+3$

form

$$x(t) = \sum_{n=-\infty}^{\infty} s[n]p(t - nT_b) \tag{3.1}$$

where $s[n]$ are PAM symbols, T_b is the symbol/baud duration, and $p(t)$ is the symbol pulse-shape. Figure 3.1 presents a block schematic that provides an instructive view of (3.1). Here, the data source is modeled as a sequence of impulses whose amplitudes are determined by the PAM symbols, $s[n]$. The symbol pulse-shape $p(t)$ is the impulse response of an LTI system that takes $s(t)$ as its input and generates $x(t)$ at its output.

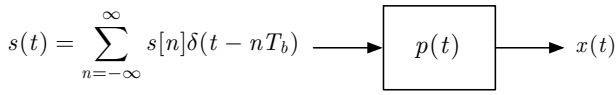


Figure 3.1: Generation of a PAM signal.

A naive choice for $p(t)$ is a rectangular pulse with a width of T_b and height of one, i.e., $p(t) = \Pi(t/T_b)$. Figure 3.2 presents an example of the transmit signal $x(t)$ when $p(t) = \Pi(t/T_b)$. In practice, the pulse-shape $p(t) = \Pi(t/T_b)$ is non-attractive because the presence of the sharp pulse edges in the time domain translate to an excessively wide bandwidth in the frequency domain. This point can be seen easily by evaluating the Fourier transform of $x(t)$. When data symbols $s[n]$ are statistically independent, the power spectrum of $s(t)$ is given by

$$\Phi_{ss}(f) = \frac{\sigma_s^2}{T_b}. \tag{3.2}$$

Hence, the power spectrum of $x(t)$ is obtained as (see (2.79))

$$\Phi_{xx}(f) = \frac{\sigma_s^2}{T_b}|P(f)|^2 \tag{3.3}$$

where $P(f)$ is the Fourier transform of $p(t)$. Since $p(t) = \Pi(t/T_b)$ has the Fourier transform $P(f) = T_b\text{sinc}(fT_b)$ (this can be easily shown by

recalling line 13 of Table 2.1 and using the time-scaling property of the Fourier transform), from (3.3), we obtain

$$\Phi_{xx}(f) = \sigma_s^2 T_b \text{sinc}^2\left(fT_b\right) \tag{3.4}$$

Figure 3.3 presents a plot of $10\log_{10}\left(\Phi_{xx}(f)/(\sigma_s^2 T_b)\right)$ as function of f. As

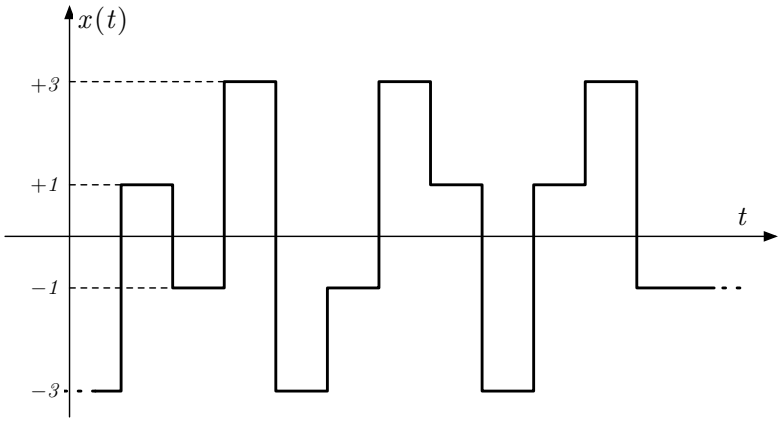


Figure 3.2: An example of a four level PAM signal with $p(t) = \Pi(t/T_b)$.

seen, this spectrum has very poor side lobes that remain significant over a wide range of frequencies. The first side lobe is only 13 dB below the main lobe. The rest of the side lobes decrease in magnitude, but very gradually.

3.2 Pulse-Shape Designs for Band-Limited Communications

As noted above, the choice of $p(t) = \Pi(t/T_b)$ as a pulse-shape is inappropriate as it leads to a transmitted signal with a very wide bandwidth. In practice, the available frequency bands for transmission of digital data streams are limited and should be kept to a minimum. Therefore, pulse-shapes whose Fourier transforms have a narrow width are of interest and used in practice. The pulse-shape $p(t)$ should also have an additional property that is known as *Nyquist criterion for intersymbol interference (ISI) free communication* or, simply, *Nyquist criterion.* A pulse-shape $p(t)$ is called Nyquist, i.e., it satisfies the Nyquist criterion, if

$$p(nT_b) = \begin{cases} 1, & n = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

When this holds,

$$x(nT_b) = s[n]. \tag{3.6}$$

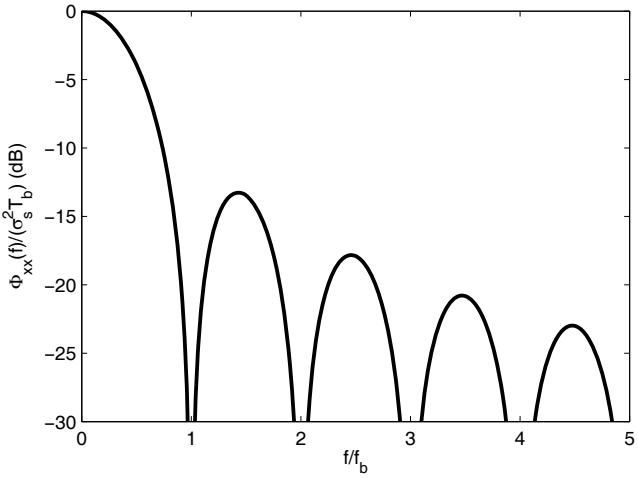


Figure 3.3: Power spectrum of a PAM signal when $p(t) = \Pi(t/T_b)$.

That is, the sampled values of $x(t)$ at the time instants $t = nT_b$ are the data symbols $s[n]$. In other words, a Nyquist pulse results in a transmit signal $x(t)$ that if sampled at proper time instants gives sample values that are equal to the transmitted symbols.

The Fourier transform of a Nyquist pulse $p(t)$ satisfies the following identity

$$\sum_{k=-\infty}^{\infty} P\left(f - \frac{k}{T_b}\right) = T_b. \tag{3.7}$$

This identity which may also be referred to as *Nyquist criterion* will be proved in the next chapter after stating the Nyquist sampling theorem (see Section 4.1.4). Figure 3.4 visualizes this property of the Nyquist pulse. Important to note here is that the mid-frequency point of the transition band of $P(f)$ is at $f = f_b/2$ and the response of $P(f)$ has odd symmetry with respect to the point $(f_b/2, T_b/2)$. Accordingly, $P(f)$ finds its minimum bandwidth when the width of its transition band approaches zero. Figure 3.5 presents the response of such a filter in both the frequency domain and the time domain. This is known as the *brick-wall filter*.

We note that for the brick-wall filter, $P(f) = T_b\Pi\left(\frac{f}{f_b}\right)$. Taking the inverse Fourier transform of $P(f)$, we obtain $p(t) = \text{sinc}(t/T_b)$. A plot of this pulse-shape is presented in Figure 3.5(b). As one would expect, $p(t)$ satisfies the Nyquist criterion (3.5), i.e., it is equal to 1 at $t = 0$ and finds value of zero at non-zero multiples of T_b.

Unfortunately, the zero width of the transition band of the brick-wall filter makes it unrealizable. Therefore, Nyquist filters with non-zero, but reasonable, transition bandwidth are designed and used in practice.

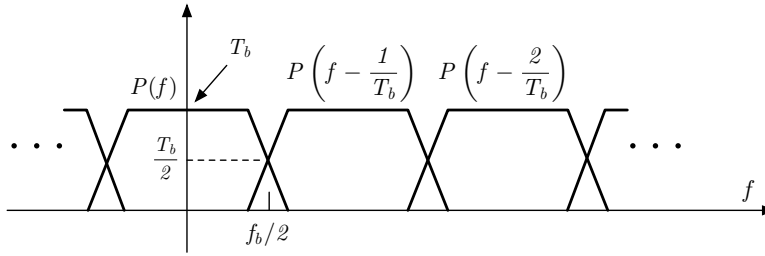


Figure 3.4: Visualization of the Nyquist pulse in the frequency domain.

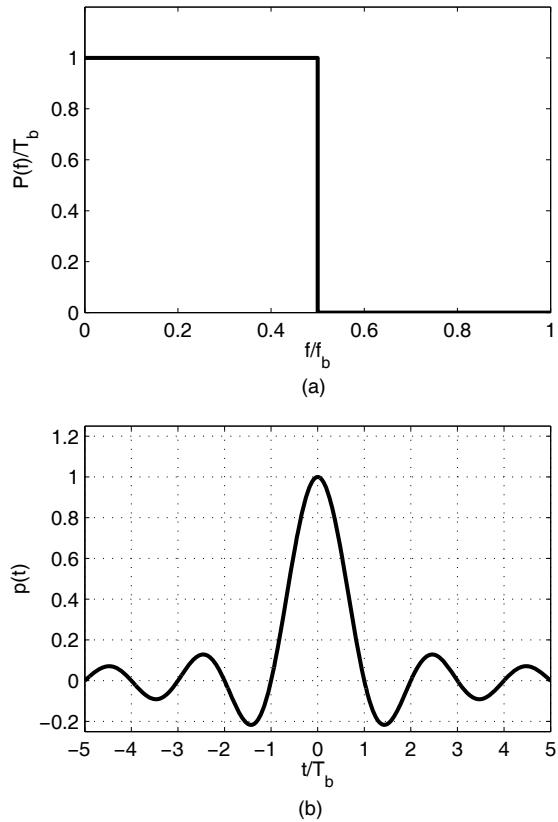


Figure 3.5: The Nyquist pulse with minimum bandwidth. (a) The frequency domain response. (b) The time domain response.

3.2.1 Raised-cosine filter

A widely used Nyquist filter is the so called '*raised-cosine filter*'. In its non-causal form, the raised-cosine filter is characterized by the frequency response

$$P_{\rm rc}(f) = \begin{cases} T_b, & |f| \le \frac{1-\alpha}{2T_b} \\ \frac{T_b}{2}\left\{1+\cos\left[\frac{\pi T_b}{\alpha}\left(|f|-\frac{1-\alpha}{2T_b}\right)\right]\right\}, & \frac{1-\alpha}{2T_b} \le |f| \le \frac{1+\alpha}{2T_b} \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

where $0 \le \alpha \le 1$, called *roll-off factor*, is a filter parameter that is related to its response as in Figure 3.6. As seen, the larger α, the wider is the bandwidth of the filter. For $\alpha = 0$, the raised-cosine filter will become the brick-wall filter. The band edge $1/2T_b = f_b/2$ is called *Nyquist frequency*.[1] For $\alpha > 0$, the bandwidth occupied by $P_{\rm rc}(f)$ beyond the Nyquist frequency is called *excess bandwidth*. The excess bandwidth is usually expressed as a percentage of the Nyquist frequency. For example, when $\alpha = 0.25$, the excess bandwidth is 25%, and when $\alpha = 1$, the excess bandwidth is 100%.

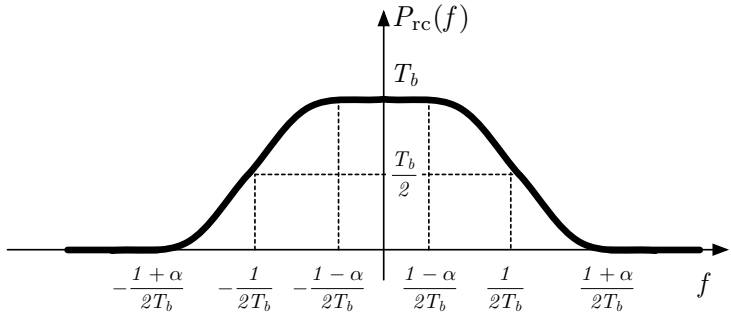


Figure 3.6: Magnitude response of the raised-cosine filter.

Taking the inverse Fourier transform of $P_{\rm rc}(f)$, the impulse response of the raised-cosine filter is obtained as

$$\begin{aligned} p_{\rm rc}(t) &= \frac{\sin(\pi t/T_b)}{\pi t/T_b}\frac{\cos(\pi\alpha t/T_b)}{1-4\alpha^2t^2/T_b^2} \\ &= {\rm sinc}(t/T_b)\frac{\cos(\pi\alpha t/T_b)}{1-4\alpha^2t^2/T_b^2}. \end{aligned} \tag{3.9}$$

A method of obtaining (3.9) is outlined in Problem 6.

Figure 3.7 presents plots of $p_{\rm rc}(t)$ for values of $\alpha = 0.25$, 0.5, 0.75, and 1. We note that as α increases, the size of the side lobes of $p_{\rm rc}(t)$ decreases.

[1] Note that the 'Nyquist frequency' that is defined here is different from the 'Nyquist rate' that is defined in the next chapter.

Clearly, having smaller side lobes is advantageous as it results in smaller level of ISI when the sampling time at the receiver is not perfect. However, this is at the cost of additional (excess) bandwidth.

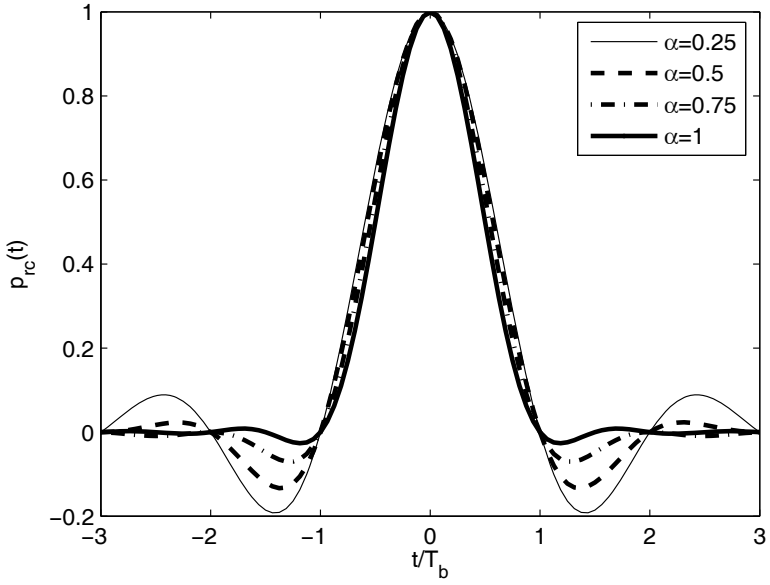


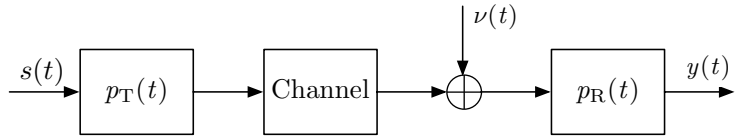
Figure 3.7: Plots of impulse response $p_{\rm rc}(t)$ of the raised-cosine filter for various choices of α.

Strictly speaking the raised-cosine pulse-shape has a duration that extends from $-\infty$ to $+\infty$. As will be discussed in the next chapter, in software radio, a preferred choice for implementation of the pulse-shaping filters is to use a finite impulse response (FIR) filter. In the next chapter, we discuss the impact of truncating the impulse response of the raised-cosine filter to a finite duration. We will also discuss direct methods of designing Nyquist FIR filters with band limited frequency response.

3.2.2 Matched filtering and square-root raised-cosine filter

In the above discussion, two unrealistic assumptions were made implicitly. Firstly, we ignored the fact that any signal that passes through a communication channel is subject to a distortion caused by the channel response. Secondly, additive noise exists in any communication channel and that further distorts the received signal. Figure 3.8 presents a communication system where the channel as well as the additive noise are included. Moreover, the pulse-shaping filter $p(t)$ is divided in to a transmit filter, $p_{\rm T}(t)$, and a receive filter, $p_{\rm R}(t)$. In the following, we ignore the channel and explore a

possible choice of the pair $p_{\rm T}(t)$ and $p_{\rm R}(t)$ that maximizes the SNR at the output of the receive filter $p_{\rm R}(t)$.



$$s(t) = \sum_{n=-\infty}^{\infty} s[n]\delta(t - nT_b)$$

Figure 3.8: Block diagram of a baseband communication system.

In the absence of the channel, from Figure 3.8, we obtain

$$y(t) = s(t) \star (p_{\rm T}(t) \star p_{\rm R}(t)) + \nu(t) \star p_{\rm R}(t). \tag{3.10}$$

Assuming that $p_{\rm T}(t) \star p_{\rm R}(t)$ is a Nyquist pulse (like $p(t)$, above) and $y(t)$ is sampled at the correct time instants, each sample of $y(t)$ will have contribution from one of the data symbols. We thus concentrate on transmission of one isolated symbol; say, we choose $s(t) = s_n(t) = s[n]\delta(t - nT_b)$. The received signal corresponding to this input is

$$y_n(t) = s[n]\delta(t - nT_b) \star (p_{\rm T}(t) \star p_{\rm R}(t)) + \nu(t) \star p_{\rm R}(t). \tag{3.11}$$

Next, we note that

$$\mathcal{F}\left[s[n]\delta(t - nT_b) \star (p_{\rm T}(t) \star p_{\rm R}(t))\right] = s[n]P_{\rm T}(f)P_{\rm R}(f)e^{-j2\pi nT_b f}. \tag{3.12}$$

Evaluating the inverse Fourier transform of the right-hand side of (3.12) for the time instant $t = nT_b$, we obtain

$$\begin{aligned} y_n^s(nT_b) &= \int_{-\infty}^{\infty} \left(s[n]P_{\rm T}(f)P_{\rm R}(f)e^{-j2\pi nT_b f}\right) e^{j2\pi nT_b f} df \\ &= s[n] \int_{-\infty}^{\infty} P_{\rm T}(f)P_{\rm R}(f) df. \end{aligned} \tag{3.13}$$

We also assume that $t = nT_b$ is the time instant when the Nyquist pulse-shape $\delta(t - nT_b) \star (p_{\rm T}(t) \star p_{\rm R}(t))$ is at its peak and thus $t = nT_b$ is the correct sampling time for $s[n]$. We thus argue that $y_n^s(nT_b)$ is the sampled signal value for $s[n]$, and accordingly at $t = nT_b$ the signal energy part of $y_n(t)$ is

$$|y_n^s(nT_b)|^2 = |s[n]|^2 \left| \int_{-\infty}^{\infty} P_{\rm T}(f)P_{\rm R}(f) df \right|^2. \tag{3.14}$$

On the other hand, assuming that $\nu(t)$ is a white noise with the power spectral density $N_0/2$, the power of the noise part of $y_n(t)$, for any t, is obtained as

$$E[|y_n^{\nu}(t)|^2] = \frac{N_0}{2}\int_{-\infty}^{\infty}|P_{\rm R}(f)|^2 df. \tag{3.15}$$

Using (3.14) and (3.15), the signal-to-noise ratio (SNR) of the received signal at the output of the receive filter $p_{\rm R}(t)$, at $t = nT_b$, is obtained as

$$\text{SNR}_o = \frac{|y_n^s(nT_b)|^2}{E[|y_n^{\nu}(nT_b)|^2]} = \frac{|s[n]|^2\left|\int_{-\infty}^{\infty}P_{\rm T}(f)P_{\rm R}(f)df\right|^2}{\frac{N_0}{2}\int_{-\infty}^{\infty}|P_{\rm R}(f)|^2 df}. \tag{3.16}$$

We wish to find the pair of filters $P_{\rm T}(f)$ and $P_{\rm R}(f)$ such that SNR_o is maximized. For that we recall the Cauchy-Schwarz inequality, which states for any pair of energy functions $G_1(f)$ and $G_2(f)$

$$\left|\int_{-\infty}^{\infty}G_1(f)G_2(f)df\right|^2 \le \int_{-\infty}^{\infty}|G_1(f)|^2 df\int_{-\infty}^{\infty}|G_2(f)|^2 df \tag{3.17}$$

and the equality holds when $G_2(f) = CG_1^*(f)$ for any arbitrary constant C.

Using the Cauchy-Schwarz inequality in (3.16), one finds that to maximize the SNR at the output of the receiver filter, $P_{\rm T}(f)$ and $P_{\rm R}(f)$ should be chosen such that

$$P_{\rm R}(f) = CP_{\rm T}^*(f). \tag{3.18}$$

The pair of filters that are related according to (3.18) are referred to as *matched filters*.

Without any loss of generality, we let $C = 1$ and note that the identity $P_{\rm R}(f) = P_{\rm T}^*(f)$ implies that

$$\mathcal{F}[p_{\rm T}(t)\star p_{\rm R}(t)] = |P_{\rm T}(f)|^2 = |P_{\rm R}(f)|^2. \tag{3.19}$$

Since $p_{\rm T}(t)\star p_{\rm R}(t)$ has to be a Nyquist pulse, if we choose it to be a raised-cosine pulse-shape, (3.19) implies that

$$|P_{\rm T}(f)| = |P_{\rm R}(f)| = \sqrt{P_{\rm rc}(f)}. \tag{3.20}$$

For obvious reasons, when $P_{\rm T}(f)$ and $P_{\rm R}(f)$ are chosen according to (3.20), they are called square-root raised-cosine filters. We use the subscript 'sr-rc' to refer to such filters, e.g., $|P_{\rm sr-rc}(f)| = \sqrt{P_{\rm rc}(f)}$.

Another interesting observation is made by noting that $P_{\rm R}(f) = P_{\rm T}^*(f)$ implies that $p_{\rm R}(t) = p_{\rm T}(-t)$ and

$$\begin{aligned} p(t) &= p_{\rm T}(t)\star p_{\rm R}(t) \\ &= \int_{-\infty}^{\infty}p_{\rm T}(\tau)p_{\rm R}(t-\tau)d\tau \end{aligned}$$

$$= \int_{-\infty}^{\infty} p_{\rm T}(\tau) p_{\rm R}(-(\tau - t)) d\tau$$
$$= \int_{-\infty}^{\infty} p_{\rm T}(\tau) p_{\rm T}(\tau - t) d\tau. \qquad (3.21)$$

The last line of (3.21) shows that the pulse-shape $p(t) = p_{\rm T}(t) \star p_{\rm R}(t)$ is the autocorrelation function of the transmit pulse-shape $p_{\rm T}(t)$. Moreover, the time where the matched filtered signal is sampled corresponds to the peak of $p(t)$ and this happens at $t = 0$. In addition, at $t = 0$, we have

$$p(0) = \int_{-\infty}^{\infty} p_{\rm T}(\tau) p_{\rm R}(0 - \tau) d\tau = \int_{-\infty}^{\infty} |p_{\rm T}(\tau)|^2 d\tau. \qquad (3.22)$$

This shows that at $t = 0$ the time reverse of $p_{\rm R}(t)$ matches $p_{\rm T}(t)$ perfectly and thus they correlate to a peak; another reason for the name matched filter.

Taking the inverse Fourier transform of $P_{\text{sr-rc}}(f)$, the impulse response of the square-root raised-cosine filter is obtained as

$$p_{\text{sr-rc}}(t) = \frac{1}{\sqrt{T_b}} \frac{\sin\left((1-\alpha)\frac{\pi t}{T_b}\right) + \frac{4\alpha t}{T_b} \cos\left((1+\alpha)\frac{\pi t}{T_b}\right)}{\frac{\pi t}{T_b}\left(1 - \left(\frac{4\alpha t}{T_b}\right)^2\right)}. \qquad (3.23)$$

Figure 3.9 presents a set of plots of square-root raised-cosine pulse-shapes for various choices of the parameter α. We note that unlike the raised-cosine pulse-shapes which had regular zero-crossings at the multiples of T_b, the square-root raised-cosine pulse-shapes have no such regular zero-crossings.

3.2.3 Causality

In the above, for the simplicity of discussion, the raised-cosine filter $P_{\rm rc}(f)$ as well as the square-root raised-cosine filter $P_{\text{sr-rc}}(f)$ were defined to be zero-phase. As a consequence, the impulse responses $p_{\rm rc}(t)$ and $p_{\text{sr-rc}}(t)$ are found to be non-causal, i.e., they are non-zero prior to $t = 0$. Such filters are clearly not realizable. To obtain realizable filters, $p_{\rm rc}(t)$ and $p_{\text{sr-rc}}(t)$ must be truncated to a finite duration and shifted to right so that they begin from or after $t = 0$. The time shift of the impulse responses do not affect the Nyquist property of the filters, though, it delays the filters outputs. However, the truncation of the impulse responses affects the responses of the filters. In particular, $P_{\rm T}(f)$ and $P_{\rm R}(f)$ will no longer be zero for $f > \frac{1+\alpha}{2T_b}$. Moreover, $p(t) = p_{\rm T}(t) \star p_{\rm R}(t)$ will satisfy the Nyquist criterion only approximately. In the next chapter, filter design techniques that aim at maximizing the stopband attenuation while the Nyquist condition is perfectly satisfied will be introduced.

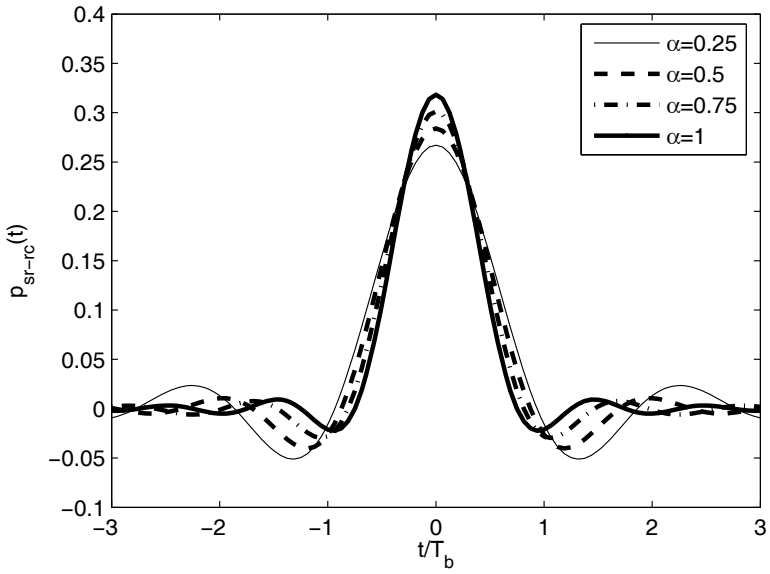


Figure 3.9: Plots of impulse response $p_{\text{sr-rc}}(t)$ of the square-root raised-cosine filter for various choices of the parameter α.

3.3 Modulation Techniques

In the baseband digital PAM that was introduced in Section 3.1, digital information is transmitted through a channel that is centered around $f = 0$. However, most communication channels only allow transmission of signals over some passbands. For example radio propagation through wireless channels can only be established over frequency bands that are in the Mega or Giga Hertz range. Hence, a baseband signal has to be shifted to the desired passband before transmission. At the receiver, on the other hand, the received signal should be shifted back to the baseband before extracting the transmitted information. The processes of shifting a baseband signal to passband and vice versa are called *modulation* and *demodulation*, respectively.

In this section, we review three modulation methods: (i) carrier-amplitude modulation; (ii) carrier-phase modulation; and (iii) quadrature amplitude modulation. We also show that the first two methods can be thought of as special cases of the third method and thus put more emphasis on the latter method.

3.3.1 Carrier-amplitude modulation

Consider the baseband signal

$$x(t) = \sum_{n=-\infty}^{\infty} s[n]p_{\mathrm{T}}(t - nT_b) \tag{3.24}$$

and note that it can be written as

$$x(t) = \sum_{n=-\infty}^{\infty} x_n(t) \tag{3.25}$$

where

$$x_n(t) = s[n]p_{\mathrm{T}}(t - nT_b) \tag{3.26}$$

is the nth component of $x(t)$ corresponding to the data symbol $s[n]$. The spectrum of $x_n(t)$ is thus obtained as

$$|X_n(f)| = |s[n]| \cdot |P_{\mathrm{T}}(f)|. \tag{3.27}$$

Figure 3.10(a) presents a plot of the spectrum of $x_n(t)$ which is assumed to be contained within the frequency range $-B \leq f \leq B$.

The carrier-amplitude modulated (AM) of $x(t)$ is obtained as in Figure 3.11 by multiplying/modulating the carrier wave $\cos(2\pi f_c t)$ with the baseband signal $x(t)$, viz.,

$$x_{\mathrm{AM}} = x(t)cos(2\pi f_c t). \tag{3.28}$$

Accordingly, the nth component of the modulated signal is

$$x_{\mathrm{AM},n}(t) = x_n(t)\cos(2\pi f_c t). \tag{3.29}$$

Recalling (2.23), the spectrum of $x_{\mathrm{AM},n}(t)$ is obtained as

$$|X_{\mathrm{AM},n}(f)| = \frac{|s[n]|}{2}[|P_{\mathrm{T}}(f - f_c)| + |P_{\mathrm{T}}(f + f_c)|]. \tag{3.30}$$

This which clearly shows a shift of the spectrum to around $\pm f_c$ is presented in Figure 3.10(b).

Assuming that the channel has a flat gain, the received signal may be written as

$$y_{\mathrm{AM}}(t) = Ax(t)\cos(2\pi f_c t + \phi) + \nu(t) \tag{3.31}$$

where A is the amplitude of the channel gain, ϕ is the carrier phase shift induced by the channel, and $\nu(t)$ is the channel noise.

To recover $x(t)$, $y_{\mathrm{AM}}(t)$ is passed through a modulator circuit similar to the one in Figure 3.11. Ignoring the channel noise, the modulator output is

$$\begin{aligned} y_{\mathrm{AM}}(t)\cos(2\pi f_c t + \phi) &= Ax(t)\cos^2(2\pi f_c t + \phi) \\ &= \frac{A}{2}x(t) + \frac{A}{2}x(t)\cos(4\pi f_c t + 2\phi). \end{aligned} \tag{3.32}$$

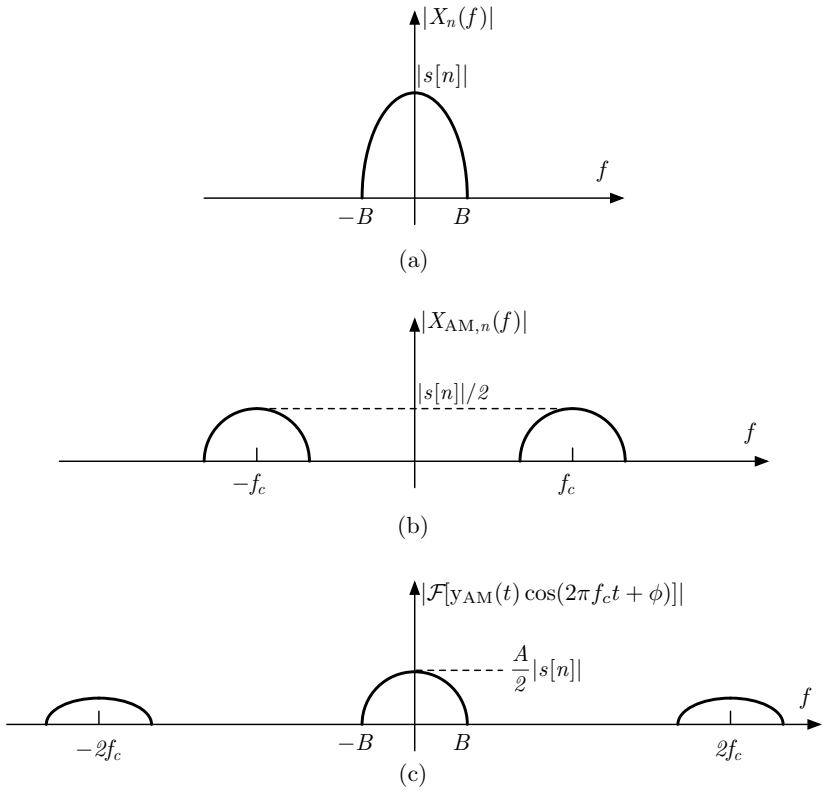


Figure 3.10: Plots of spectra of (a) $x_n(t)$, (b) $x_{\mathrm{AM},n}(t)$, and (c) $x_{\mathrm{AM},n}(t)\cos(2\pi f_c t)$.

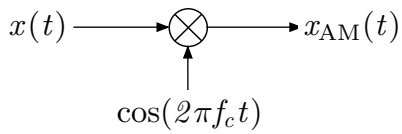


Figure 3.11: Amplitude modulation of a sinusoidal carrier by a baseband PAM signal.

Figure 3.10(c) depicts the result of this operation in the frequency domain. Clearly the second term on the right-hand side of (3.32) is $x(t)$ modulated to the carrier frequency $2f_c$. This can be removed by passing $y_{\rm AM}(t)\cos(2\pi f_c t+\phi)$ through a lowpass filter. By choosing the lowpass filter to be the matched filter $p_{\rm R}(t) = \frac{2}{A}p_{\rm T}(-t)$, the filter output will satisfy the Nyquist condition and also maximizes the SNR.

Note that successful recovery of $x(t)$ at the receiver requires the knowledge of the carrier frequency, f_c, and the carrier phase, ϕ. The estimates of f_c and ϕ are obtained from $x_{\rm AM}(t)$ using a phase-locked loop (PLL). Hence, the receiver for an AM signal has the structure shown in Figure 3.12. The subject of PLL is covered in Chapter 8.

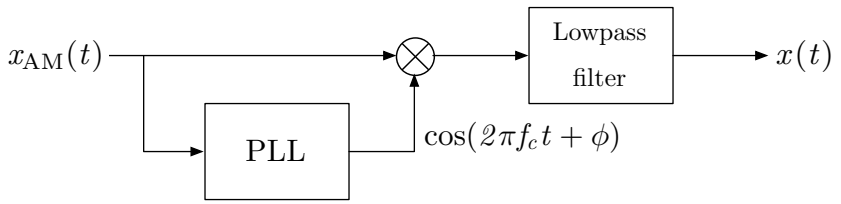


Figure 3.12: Demodulation of a carrier-amplitude modulated signal.

3.3.2 Quadrature amplitude modulation

In quadrature amplitude modulation (QAM) two PAM signals are transmitted simultaneously over the same carrier/frequency band, by modulating $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$. The terminology 'quadrature' refers to the fact that $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are in the quadrature phase with respect to each other, i.e., have a phase difference of $\frac{2\pi}{4} = \frac{\pi}{2}$. Separation of the two PAM signals at the receiver, as explained below, is possible because of the orthogonality of $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$, i.e., since $\int \sin(2\pi f_c t)\cos(2\pi f_c t)dt = 0$.

A quadrature-amplitude-modulated signal has the form

$$x_{\rm QAM}(t) = x_{\rm R}(t)\cos(2\pi f_c t) - x_{\rm I}(t)\sin(2\pi f_c t) \tag{3.33}$$

where $x_{\rm R}(t)$ and $x_{\rm I}(t)$ are the baseband PAM signals similar to $x(t)$ in (3.24). Assuming the channel has a flat gain, as in the case of AM, the received signal is obtained as

$$y_{\rm QAM}(t) = Ax_{\rm R}(t)\cos(2\pi f_c t+\phi) - Ax_{\rm I}(t)\sin(2\pi f_c t+\phi) + \nu(t). \tag{3.34}$$

Ignoring the channel noise, $\nu(t)$, and observing the following equalities, one can easily see that $x_{\rm R}(t)$ and $x_{\rm I}(t)$ may be extracted from $x_{\rm QAM}(t)$, by

modulating $\cos(2\pi f_c t+\phi)$ and $-\sin(2\pi f_c t+\phi)$, respectively, with $x_{\rm QAM}(t)$ and lowpass filtering the results:

$$y_{\rm QAM}(t)\cos(2\pi f_c t+\phi)=\frac{A}{2}x_{\rm R}(t)+\frac{A}{2}x_{\rm R}(t)\cos(4\pi f_c t+2\phi) \\ -\frac{A}{2}x_{\rm I}(t)\sin(4\pi f_c t+2\phi) \qquad (3.35)$$

$$y_{\rm QAM}(t)\times(-\sin(2\pi f_c t+\phi))=\frac{A}{2}x_{\rm I}(t)-\frac{A}{2}x_{\rm I}(t)\cos(4\pi f_c t+2\phi) \\ -\frac{A}{2}x_{\rm R}(t)\sin(4\pi f_c t+2\phi). \qquad (3.36)$$

To summarize, the block diagrams of a QAM modulator and a QAM demodulator are shown in Figures 3.13 and 3.14, respectively.

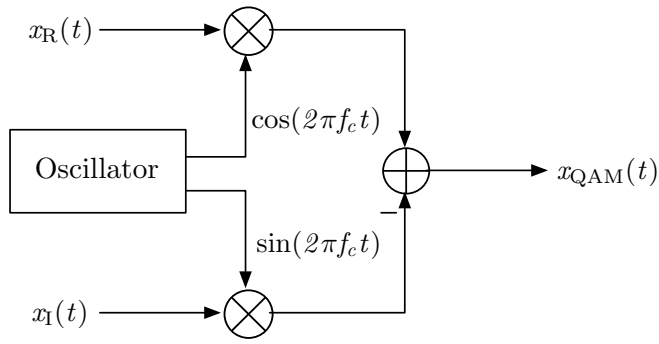


Figure 3.13: QAM modulator.

Bandwidth efficiency

A key point that needs special attention is that both $x_{\rm AM}(t)$ and $x_{\rm QAM}(t)$ take up the same amount of bandwidth, assuming that the symbol rate in $x_{\rm R}(t)$ and $x_{\rm I}(t)$ of Figure 3.13 is the same as that in $x(t)$ of Figure 3.11. However, the rate of information transmission in QAM is twice as much as that of AM, obviously because each of $x_{\rm R}(t)$ and $x_{\rm I}(t)$ in (3.33) carry the same amount of information as $x(t)$ in (3.28). It is thus seen that from bandwidth point of view, QAM is twice more efficient than AM. On the other hand, since $x(t)$ is a real-valued signal, $X(f)$ possesses the conjugate symmetry property that is given in (2.21). One can take advantage of this symmetry of $X(f)$ and establish a transmission of $x(t)$ within half

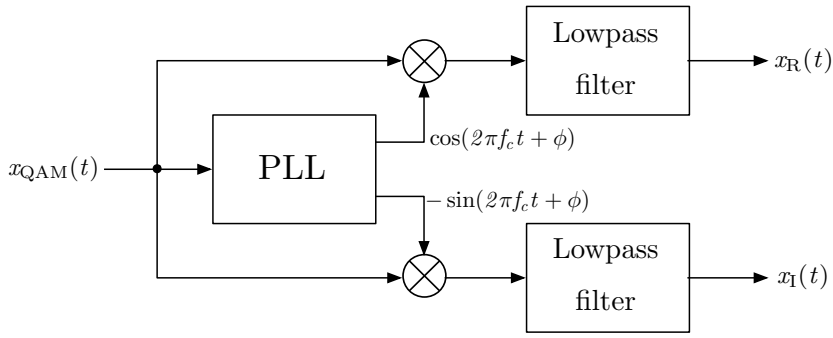


Figure 3.14: QAM demodulator.

of the bandwidth used by $x_{\rm AM}(t)$, and thus achieve the same bandwidth efficiency as QAM. This concept is demonstrated in Figures 3.15(a), (b) and (c), where the positive and negative frequency portions of $X(f)$ are identified as lower and upper bands, they are separated and shifted to the carrier band for modulation, and later shifted back to the baseband for reconstruction of $x(t)$. This method that effectively transmits only one side of the spectrum $X(f)$ is called single-side band (SSB) modulation. On the same basis, the AM modulation that was introduced in Section 3.3.1 is called double-side band (DSB) modulation.

Although, SSB may have the same bandwidth efficiency as QAM, it is rarely used for digital data transmission. This is because the process of filtering one of the side-bands of the spectrum is rather expensive and thus makes the SSB modulator costly. In addition, perfect filtering is not possible. This, in practice, results in SSB signals whose bandwidth is more than one-half of that of QAM signals. Such signals are called vestigial-side band (VSB) modulated signals.

To summarize, because of the bandwidth inefficiency of DSB-AM and the complexity of implementation of SSB-AM, digital AM is rarely used. Hence, most of the digital transmission systems are QAM-based and in the rest of this book we limit our discussion to this type of systems.

QAM formulation based on complex signals

A convenient formulation of quadrature amplitude modulated signals is obtained through complex representation of the underlying signals. To begin with, the pair of PAM symbols $s_{\rm R}[n]$ and $s_{\rm I}[n]$ are combined in a complex-valued symbol

$$s[n] = s_{\rm R}[n] + j s_{\rm I}[n] \tag{3.37}$$

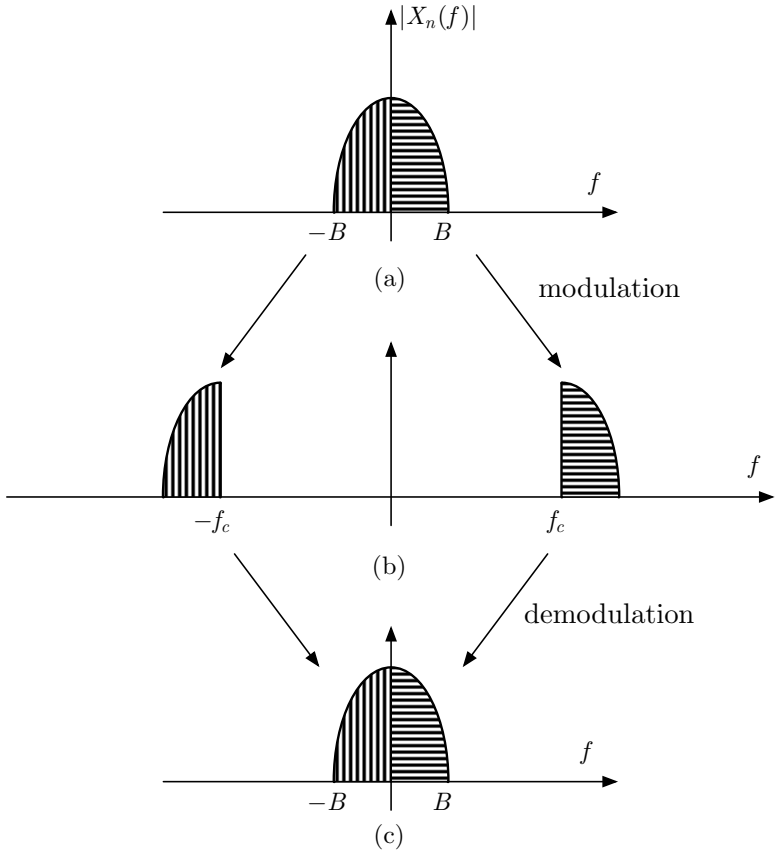


Figure 3.15: Single side band modulation. (a) Baseband spectrum. (b) Modulated spectrum. (c) Demodulated spectrum.

where $s_{\rm R}[n]$ and $s_{\rm I}[n]$ are real and imaginary parts of $s[n]$, respectively. The corresponding band-limited baseband signal is thus obtained as

$$\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} s[n] p_{\rm T}(t - nT_b) \\ &= x_{\rm R}(t) + j x_{\rm I}(t). \end{aligned} \tag{3.38}$$

The modulated signal $x_{\rm QAM}(t)$ is then constructed by modulating the complex carrier $e^{j2\pi f_c t}$ and taking the real part of the result, viz.,

$$\begin{aligned} x_{\rm QAM}(t) &= \Re\left[x(t) e^{j2\pi f_c t}\right] \\ &= x_{\rm R}(t)\cos(2\pi f_c t) - x_{\rm I}(t)\sin(2\pi f_c t). \end{aligned} \tag{3.39}$$

Demodulation of $x(t)$ to baseband can be established by modulating $e^{-j2\pi f_c t}$ with $x_{\rm QAM}(t)$ and lowpass-filtering of the result. This is easily confirmed by noting that

$$x_{\rm QAM}(t) = \frac{1}{2}\left[x(t) e^{j2\pi f_c t} + x^*(t) e^{-j2\pi f_c t}\right]. \tag{3.40}$$

Thus,

$$x_{\rm QAM}(t) e^{-j2\pi f_c t} = \frac{1}{2}\left[x(t) + x^*(t) e^{-j4\pi f_c t}\right]. \tag{3.41}$$

and the second term on the right-hand side of (3.41) is removed by a lowpass filter.

In the light of the above formulation, it is common to represent the complex-valued QAM symbols in a complex plane. Two examples of QAM symbols (also, known as constellations) are presented in Figures 3.16(a) and (b). Figure 3.16(a) is a 4-QAM constellation where each symbol carries 2 bits of information. Figure 3.16(b), on the other hand, shows a 16-QAM constellation. In this case, each symbol carries 4 bits of information.

3.3.3 Carrier-phase modulation

In carrier-phase modulation (PM) the information symbols determine the phase of the carrier. A carrier-phase modulated signal thus finds the following form

$$x_{\rm PM}(t) = \sum_{n=-\infty}^{\infty} p_{\rm T}(t - nT_b)\cos(2\pi f_c t + \theta_n) \tag{3.42}$$

where θ_n are the information carrying phase angles. Expanding $\cos(2\pi f_c t + \theta_n)$, (3.42) can be rearranged as

$$x_{\rm PM}(t) = x_{\rm R}(t)\cos(2\pi f_c t) - x_{\rm I}(t)\sin(2\pi f_c t) \tag{3.43}$$

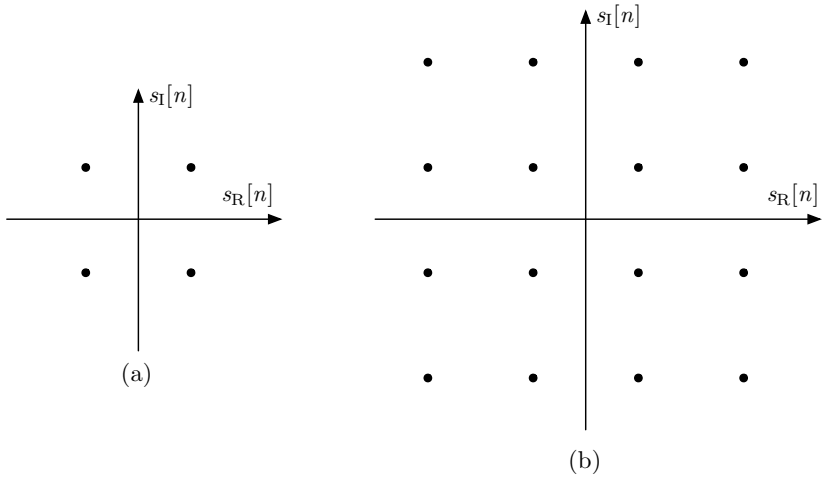


Figure 3.16: Examples of QAM constellations. (a) 4-QAM: each symbol carries 2 bits of information. (b) 16-QAM: each symbol carries 4 bits of information.

where

$$x_R(t) = \sum_{n=-\infty}^{\infty} p_T(t - nT_b) \cos\theta_n \tag{3.44}$$

and

$$x_I(t) = \sum_{n=-\infty}^{\infty} p_T(t - nT_b) \sin\theta_n. \tag{3.45}$$

If we let $\cos\theta_n = s_R[n]$ and $\sin\theta_n = s_I[n]$, this, clearly, shows that PM may be thought as a special case of QAM where constellation points are on a circle, at the angles θ_1, θ_2, $\cdots$. Hence, the modulator and demodulator block diagrams that were introduced in Figures 3.13 and 3.14 are applicable to PM as well. Figures 3.17(a), (b), (c) and (d) show examples of PM constellations for 2-ary, 4-ary, 8-ary and 16-ary signalling. Comparing Figures 3.16 and 3.17, it may be noted that 4-ary PM and 4-QAM have the same constellation. They are often referred to as quadrature phase-shift keying (QPSK). This is the name that we also use in the rest of this book.

In the case of the 2-ary PM, the angle θ_n only takes values of 0 and π. In this case, $\sin\theta_n = 0$, for any n, and $\cos\theta_n = \pm 1$. Using these in (3.44) and (3.45), we get

$$x_R(t) = \sum_{n=-\infty}^{\infty} s[n] p_T(t - nT_b) \tag{3.46}$$

and

$$x_{\mathrm{I}}(t) = 0, \tag{3.47}$$

respectively. Using these, (3.43) reduces to the same form as (3.28). This shows that 2-ary PM is simply a 2-level/binary AM signal. The common term for 2-ary PM/AM signal is binary phase-shift keying (BPSK).

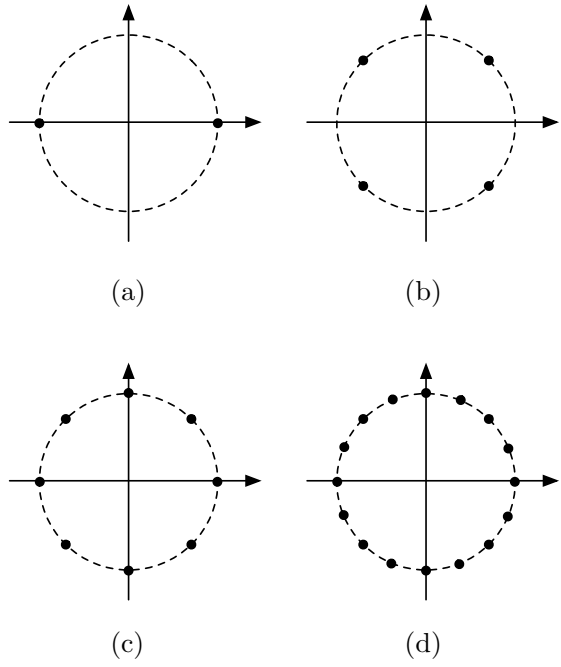


Figure 3.17: Examples of constellations of phase modulators. (a) 2-ary PM or BPSK. (b) 4-ary PM or QPSK. (c) 8-ary PM. (d) 16-ary PM.

It is also important to note that the number of bits carried by each symbol in 2-ary PM/BPSK is 1 bit, in 4-ary PM/QPSK is 2 bits, in 8-ary PM is 3 bits, and in 16-ary PM is 4 bits.

3.4 Binary to Symbol Mapping

One of the first tasks in any transmitter is to map the (coded) information bits to data symbols for transmission. Mapping is straightforward and is done as follows. Without any loss of generality, let us consider the case where data symbols are to be selected from a 16-QAM set. Since each 16-QAM symbol takes 16 different combinations, it can carry 4 bits of information. Accordingly, the sequence of information bits is first partitioned

to groups of 4 bits each, and each group is mapped to a 16-QAM symbol. Although any mapping that uniquely assigns each combination of 4 bits to one of the 16 possible choices of the QAM symbols may be acceptable, it is more common to use Gray mapping, where mapping is done such that adjacent symbols differ by only one bit; also, see the discussion in Section 3.1.

Figure 3.18 shows a possible Gray mapping of 4 binary bits to a 16-QAM constellation. One simple method of thinking about this mapping is to imagine that the first two bits determine $s_{\mathrm{R}}[n]$ and the other two bits determine $s_{\mathrm{I}}[n]$, and for mappings along $s_{\mathrm{R}}[n]$ and $s_{\mathrm{I}}[n]$ axes we have used Table 3.1.

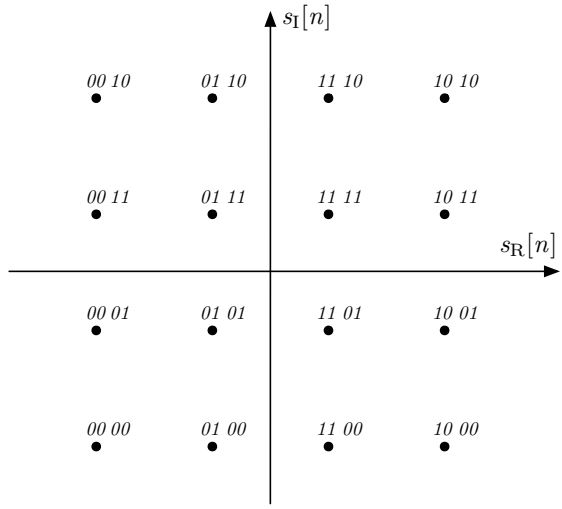


Figure 3.18: A Gray mapping of 4 binary bits to a 16-QAM constellation.

3.5 Differential Encoding and Decoding

In practice, a receiver is usually trained to match its carrier phase and frequency with those of the received signal for a (near) perfect demodulation. Furthermore, pilot symbols that are known to the receiver are often transmitted at the beginning of each communication session for the receiver training. However, there are also situations where the receiver has to rely only on the statistics of the incoming signal and use such information to extract the carrier phase and frequency. As we will see in later chapters of this book, carrier frequency can always be found without any problem. However, there will always be some ambiguity in phase. For instance, if the transmit symbols belong to QPSK constellation, for an integer k, a change

of $k\pi/2$ in the carrier phase has no noticeable effect on the observed constellation points. This can be easily seen if we refer to Figure 3.16(b) and note that any rotation of $k\pi/2$ has no impact on the constellation. Hence, unless the receiver knows exactly what has been transmitted, it will remain ambiguous about the phase by some integer factor of $\pi/2$. Clearly, this conclusion remains the same for other square QAM constellations as well, e.g., 16-QAM. For PM symbols, the phase ambiguity will depend on the constellation size. If data symbols belong to an M-ary PM, the receiver will remain ambiguous to any phase change/rotation of $2\pi/M$.

The problem of the phase ambiguity can be solved using the method of differential encoding and decoding. Since this method works best with the phase modulated signals and can be best explained in this context, we continue our discussion while limiting ourselves to this type of modulation. Extension of the results to a special class of QAM symbols is left as problems at the end of the chapter; see Problem 8.

In the conventional PM, information bits are coded as the phase angle θ_n of a carrier as in (3.42). The receiver, using a demodulator similar to the one in Figure 3.14, extracts an estimate of $s[n] = \cos\theta_n + j\sin\theta_n = e^{j\theta_n}$, from which it obtains the associated information bits. In the differential PM, the information symbols are encoded as follows. The carrier phase in the time interval corresponding to the $(n-1)$st symbol is set equal to θ_{n-1} and in the next time interval is shifted to θ_n, and θ_n is selected such that $\theta_n - \theta_{n-1}$ be the phase angle representing the transmitted information. With this arrangement, the receiver will remain insensitive to any carrier phase ambiguity. This can be explained as follows. Let the carrier phase error be ϕ. In that case, the receiver, at the time instant $n-1$, detects the phase angle $\phi + \theta_{n-1}$, and at the time instant n, detects the phase angle $\phi+\theta_n$, and hence recovers the correct phase difference $(\phi+\theta_n)-(\phi+\theta_{n-1}) = \theta_n - \theta_{n-1}$.

3.6 Baseband Equivalent of a Passband Channel

Following (3.40) and (3.41), Figure 3.19 presents the block diagram of a digital QAM communication system. In the study of communication systems, we are often interested in the compound effect of the transmit filter $p_{\mathrm{T}}(t)$, the channel $c(t)$, and the receive filter $p_{\mathrm{R}}(t)$, i.e., the channel model between the input $s(t)$ and the output $y(t)$. Since both $s(t)$ and $y(t)$ are baseband signals, but the channel is passband, this channel model is called the *baseband equivalent of the passband channel*. Moreover, since $s(t)$ and $y(t)$ are complex signals, the impulse response of the baseband equivalent of the passband channel, in general, is complex-valued.

The transfer function between $s(t)$ and $x_{\mathrm{QAM}}(t)$ is obtained as follows. Without any loss of generality, we let $s(t) = \delta(t)$. The Fourier transform of

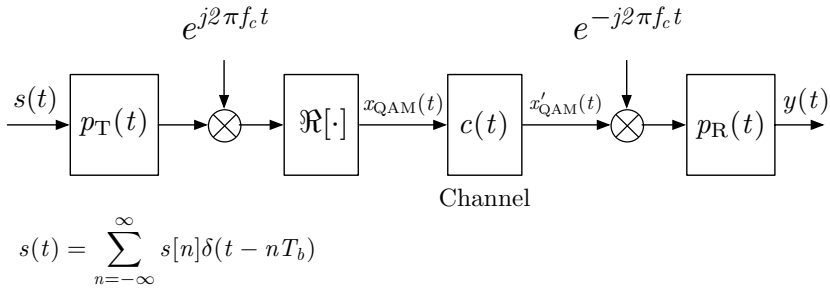


Figure 3.19: Block diagram of a digital QAM communication system.

$x_{\mathrm{QAM}}(t)$ is thus obtained as

$$X_{\mathrm{QAM}}(f) = \frac{1}{2}[P_{\mathrm{T}}(f - f_c) + P_{\mathrm{T}}(f + f_c)]. \tag{3.48}$$

The Fourier transform of the received signal (prior to demodulation) is $X_{\mathrm{QAM}}(f)$ times the channel transfer function, $C(f) = \mathcal{F}[c(t)]$, viz.,

$$X'_{\mathrm{QAM}}(f) = \frac{1}{2}[P_{\mathrm{T}}(f - f_c)C(f) + P_{\mathrm{T}}(f + f_c)C(f)]. \tag{3.49}$$

Next, using (2.22), we note that

$$\begin{aligned}\mathcal{F}\left[x'_{\mathrm{QAM}}(t)e^{-j2\pi f_c t}\right] &= X'_{\mathrm{QAM}}(f + f_c) \\ &= \frac{1}{2}[P_{\mathrm{T}}(f)C(f + f_c) + P_{\mathrm{T}}(f + 2f_c)C(f + f_c)].\end{aligned} \tag{3.50}$$

Since the second term on the right-hand side of (3.50) is centered around $-2f_c$, it will be removed by the (lowpass) receive filter $p_{\mathrm{R}}(t)$. The Fourier transform of $y(t)$ is thus obtained as $Y(f) = \frac{1}{2}P_{\mathrm{T}}(f)C(f + f_c)P_{\mathrm{R}}(f)$. On the other hand, since the input was an impulse, $Y(f)$ is also the desired transfer function and thus the baseband equivalent of the passband channel has the transfer function

$$C_{\mathrm{BB}}(f) = \frac{1}{2}P(f)C(f + f_c) \tag{3.51}$$

where $P(f) = P_{\mathrm{T}}(f)P_{\mathrm{R}}(f)$. In the rest of this book, for mathematical simplicity, we ignore the coefficient $\frac{1}{2}$ on the right-hand side of (3.51) and define the baseband equivalent of the passband channel as

$$C_{\mathrm{BB}}(f) = P(f)C(f + f_c). \tag{3.52}$$

Finally, we obtain the impulse response of the channel between $s(t)$ and $y(t)$ as

$$c_{\mathrm{BB}}(t) = \mathcal{F}^{-1}[C_{\mathrm{BB}}(f)]. \tag{3.53}$$

Example 3.1:

A multipath passband communication channel that operates at the carrier frequency 100 MHz has two paths with the gains of 1 and 0.5 and the delays of 0 and 7.5 μs, respectively. The matched filters at the transmitter and receiver sides are square-root raised-cosine pulse-shapes designed for $T_b = 10$ μs and $\alpha = 0.5$.

(a) Derive an equation for the equivalent baseband channel transfer function $C_{\rm BB}(f)$.

(b) Find $C_{\rm BB}(f)$ if the delay of the second path changes to 7.5025 μs.

(c) In both cases draw the plots of the magnitude and phase response of $C_{\rm BB}(f)$.

(d) In both cases obtain and plot the impulse response of the baseband equivalent of the passband channel.

Solution:

(a) The impulse response of the channel is $c(t) = \delta(t) + 0.5\delta(t - 7.5)$. Hence, $C(f) = 1+0.5e^{-j2\pi f\times 7.5} = 1+0.5e^{-j15\pi f}$. Substituting this in (3.52), we obtain

$$C_{BB}(f) = \left(1 + 0.5e^{-j15\pi(f+f_c)}\right) P(f) = \left(1 + 0.5e^{-j15\pi f}\right) P(f)$$

where $P(f)$ is the raised-cosine filter (3.8) with $T_b = 10$ μs and $\alpha = 0.5$.

(b) Following the same line of derivation as in Part (a), in this case, we obtain

$$\begin{aligned} C_{BB}(f) &= \left(1 + 0.5e^{-j0.5\pi}e^{-j15.005\pi f}\right) P(f) \\ &= \left(1 - 0.5je^{-j15.005\pi f}\right) P(f). \end{aligned}$$

(c) The plots are shown in Figure 3.20. As seen, a small change in one of the path delays can significantly affect the baseband equivalent of the channel response.

(d) Taking the inverse Fourier transforms of the results in Parts (a) and (b), we obtain:

$$\text{for case (a)}, \quad c_{BB}(t) = p(t) + 0.5p(t - 7.5)$$

and

$$\text{for case (b)}, \quad c_{BB}(t) = p(t) - 0.5jp(t - 7.5025).$$

These are plotted in Figure 3.21. Interesting to note here is that while the case (a) leads to a real-valued impulse response, the case (b) has a complex-valued impulse response. As noted above $c_{BB}(t)$, in general, is expected to be a complex-valued time response.

Generalizing the above example, if the channel is a multipath channel that is characterized by the impulse response

$$c(t) = \sum_{i=0}^{L-1} a_i\delta(t - \tau_i) \tag{3.54}$$

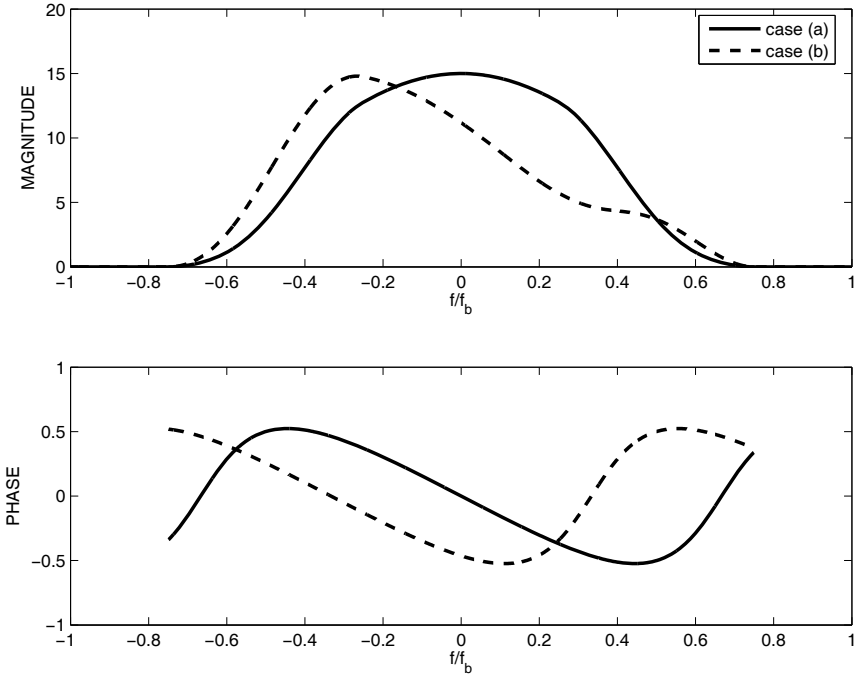


Figure 3.20: Magnitude and phase responses of the baseband equivalent channels of Example 3.1.

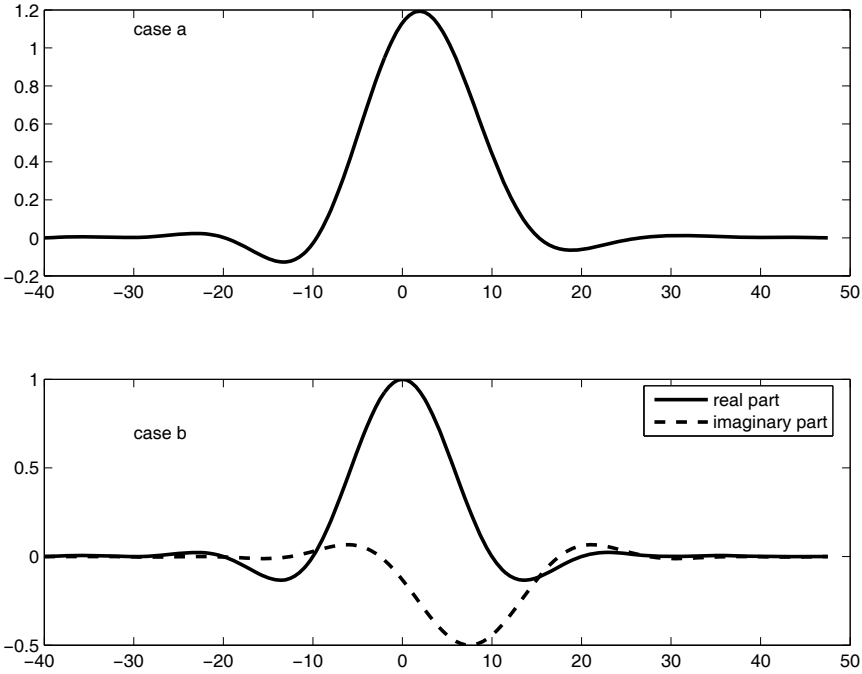


Figure 3.21: Impulse responses of the baseband equivalent channels of Example 3.1.

where L is the number of paths, and a_i and τ_i are the ith path gain and delay, the baseband equivalent of the channel will be

$$c_{\mathrm{BB}}(t) = \sum_{i=0}^{L-1} a_i e^{-j2\pi f_c \tau_i} p(t - \tau_i). \tag{3.55}$$

3.7 Problems

1. A radio channel with a bandwidth of 2 MHz is available for transmission. Using 16-QAM modulation, we wish to transmit 6 Mb/s of coded information through this channel. Find the maximum roll-off factor that may be used for pulse-shaping in this system.

2. A communication system with a bandwidth of 6 MHz uses raised-cosine pulse-shapes with a roll-off factor $\alpha = 0.5$. For each of the following modulation methods find the system bit rate: (a) BPSK. (b) QPSK. (c) 16-QAM. (d) 64-QAM.

3. (a) For the 4-QAM symbols presented in Figure 3.16(a) suggest a method of Gray mapping, i.e., assigning 2 bits to each point of the constellation such that each pair of neighbor points differ in at most one bit.
   (b) Repeat (a) for the 16-QAM symbols presented in Figure 3.16(b).
   (c) Repeat (a) for the PSK symbols presented in Figure 3.17(b) and (c).

4. Consider the AM modulator shown in Figure 3.22(a). Here, the block $\Re[\cdot]$ means taking the real part of the input. Assuming that $s(t)$ is a real-valued signal with the amplitude and phase spectra shown in Figure 3.22(b), find and plot:
   (a) The amplitude and phase spectra of $x(t)$.
   (b) The amplitude and phase spectra of $y(t)$.
   (c) By comparing the results of (a) and (b) show that
   $$\mathcal{F}\left[\Re[x(t)]\right] = \frac{1}{2}X(f) + \frac{1}{2}X^*(-f).$$

5. This problem guides you to develop a SSB modulator using a special signal processing block called *Hilbert transformer*. Hilbert transformer is an LTI system which is characterized by the transfer function
   $$H_{\mathrm{Hilbert}}(f) = -j\mathrm{sgn}(f)$$
   where
   $$\mathrm{sgn}(f) = \begin{cases} +1 & f \geq 0 \\ -1 & f < 0 \end{cases}$$

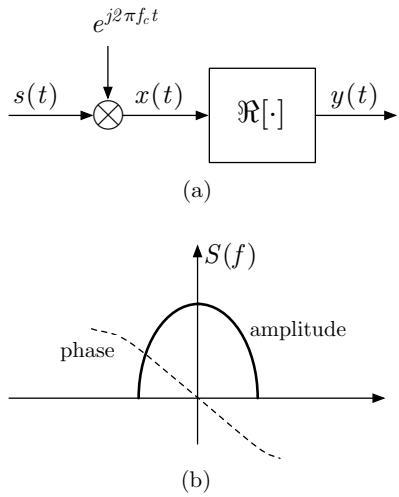


Figure 3.22: (a) An AM modulator. (b) The amplitude and phase spectra of the input $s(t)$

In the time domain, Hilbert transformer is characterized by the impulse response $h_{\text{Hilbert}}(t) = \mathcal{F}^{-1}[H_{\text{Hilbert}}(f)] = \frac{1}{\pi t}$ which is obtained by using line 12 of Table 2.1 and invoking the duality property of the Fourier transform.

Now consider the system shown in Figure 3.23, where $\hat{p}_{\text{T}}(t)$ is the Hilbert transform of $p_{\text{T}}(t)$.

(a) By taking $s(t) = s[0]\delta(t)$ (i.e., a single symbol) and careful examination of the Fourier transform of $y(t)$, show that the system of Figure 3.23 is an SSB modulator.

(b) By using the results of (a), argue that the system of Figure 3.23 is also an SSB modulator for the more general case where

$$s(t) = \sum_{n=-\infty}^{\infty} s[n]\delta(t - nT_b).$$

(c) Assuming that $p_{\text{T}}(t)$ is a square-root raised-cosine pulse with the roll-off factor $\alpha = 0.5$, present a plot of $p_{\text{T}}(t)$ and a plot of $\hat{p}_{\text{T}}(t) = p_{\text{T}}(t) \star \frac{1}{\pi t}$. You may use MATLAB to perform the convolution $p_{\text{T}}(t) \star \frac{1}{\pi t}$. Discuss any potential problem that you may see with regard to the implementation of the SSB modulator of Figure 3.23.

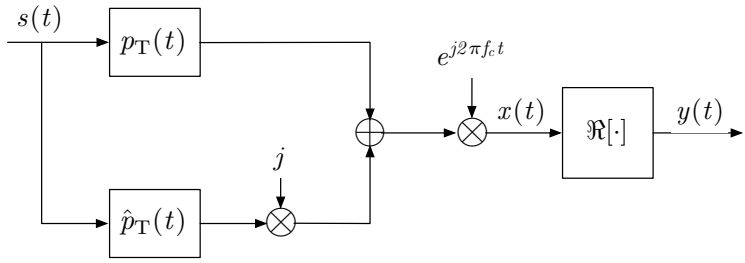


Figure 3.23: An SSB modulator.

6. (a) Present the plots of the frequency domain functions $\Pi(fT_b)$ and $G(f) = \frac{\pi T_b}{2\alpha}\cos(\pi f T_b/\alpha)\Pi(fT_b/\alpha)$, and by convolving them show that $P_{\mathrm{rc}}(f) = T_b\Pi(fT_b) \star G(f)$.

   (b) Using the result of Part (a), confirm the correctness of (3.9).

7. (a) Starting with (3.8) and recalling the identity $P_{\mathrm{sr-rc}}(f) = \sqrt{P_{\mathrm{rc}}(f)}$, show that

$$P_{\mathrm{rc}}(f) = \begin{cases} \sqrt{T_b}, & |f| \le \frac{1-\alpha}{2T_b} \\ \sqrt{T_b}\cos\left[\frac{\pi T_b}{2\alpha}\left(|f| - \frac{1-\alpha}{2T_b}\right)\right], & \frac{1-\alpha}{2T_b} \le |f| \le \frac{1+\alpha}{2T_b} \\ 0, & \text{otherwise} \end{cases}$$

   (b) By taking the inverse Fourier transform of the result of Part (a), confirm the correctness of (3.23).

8. The discussion on differential encoding/decoding in Section 3.5 was limited to phase modulation. In this problem, we extend the idea to PAM as well as QAM systems.

   (a) Consider the case where the transmitted information is mapped in the following manner. The transmitted baseband signal is a PAM signal that can take two possible levels a and b, both positive numbers. If the current information bit is 0, the current level of PAM signal remains the same for the next symbol period, i.e., there will be no level transition. On the other hand, if the current information bit is 1 the PAM signal level changes to b, if the current level is a, or vice versa. For this transmitter, suggest a receiver system.
   Hint: Think of a rectifier followed by a lowpass filter and add the rest.

   (b) Star QAM (SQAM) is a special QAM that allows differential encoding and decoding. Figure 3.24 presents the constellation of a 16-SQAM. Following the discussion in Section 3.5 and the

results of Part (a), development a method of differential encoding and decoding for the 16-SQAM.

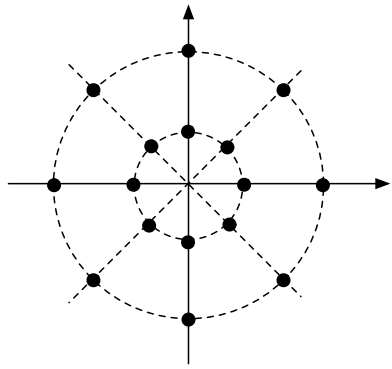


Figure 3.24: A SQAM constellation.

9. A multipath passband communication channel that operates at the carrier frequency f_c MHz has two paths with the gains of a and b and the delays of 0 and τ μs, respectively. The matched filters at the transmitter and receiver sides are square-root raised-cosine pulse-shapes designed for symbol duration of T_b μs and the roll-off factor α.

   (a) Derive a general equation for the equivalent baseband channel transfer function $C_{\mathrm{BB}}(f)$ and its time domain equivalent, $c_{BB}(t)$.

   (b) Evaluate and plot $c_{BB}(t)$ for the cases where:

      i. $T_b = 1$ μs, $\alpha = 0.25$, $f_c = 200$ MHz, $a = 1$, $b = 0.7$, and $\tau = 0.3$ μs.

      ii. $T_b = 1$ μs, $\alpha = 0.25$, $f_c = 200$ MHz, $a = 1$, $b = 0.7$, and $\tau = 0.302$ μs.

      iii. $T_b = 1$ μs, $\alpha = 0.25$, $f_c = 201$ MHz, $a = 1$, $b = 0.7$, and $\tau = 0.3$ μs.

      iv. $T_b = 1$ μs, $\alpha = 0.25$, $f_c = 200$ MHz, $a = 1$, $b = 0.7$, and $\tau = 0.012$ μs.

   (c) Try some other examples of your own. Can you find a combination of a, b, f_c and τ that results in a very small $c_{BB}(t)$, for all values of t, i.e., a case where the channel is in a deep fade?

Index