M. GHAVAMI · L. B. MICHAEL · R. KOHNO

ultra wideband

signals and systems in communication engineering

Companion Website

WILEY

Ultra Wideband Signals and Systems in Communication Engineering

M. Ghavami
King's College London, UK

L. B. Michael
Japan

R. Kohno
Yokohama National University, Japan

John Wiley & Sons, Ltd

Ultra Wideband Signals and Systems in Communication Engineering

Ultra Wideband Signals and Systems in Communication Engineering

M. Ghavami
King's College London, UK

L. B. Michael
Japan

R. Kohno
Yokohama National University, Japan

John Wiley & Sons, Ltd

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road, Etobicoke, Rexdale, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-86751-5

Typeset by the author using LaTex Software.
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

In this book we focus on the basic *signal processing* that underlies current and future ultra wideband systems. By looking at signal processing in this way we hope this text will be useful even as UWB applications mature and change or regulations regarding ultra wideband systems are modified. The current UWB field is extremely dynamic, with new techniques and ideas being presented at every communications and signal-processing conference. The basic signal-processing techniques presented in this text though will not change for some time to come. Thus, we have taken a somewhat theoretical approach, which we believe is longer lasting and more useful to the reader in the long term than an up-to-the-minute summary that is out of date as soon as it is published.

We restrict our discussion in general to *ultra wideband communication*, looking in particular at *consumer communication*. What we mean by this is that although there are many and varied specialized applications for UWB, particularly for the military, we assume that the majority of readers will either be in academia or in industry. In any case, as this is a basic text aimed primarily at the undergraduate or graduate student, these basics should stand readers in good stead in their efforts to understand more advanced papers and make a contribution in this field for themselves. We are painfully aware of the depth and breadth of this field and, regretfully, pass on interesting topics, such as UWB radar, including ground-penetrating radar, and most military applications. For the former there is already a great deal of information available, while for the latter most material is classified.

The introduction to this book presents a brief look at why UWB is considered to be such an exciting wireless technology for the near future. We examine Shannon's famous capacity equation and, as a consequence, look at the large-bandwidth possibilities for high data rate communication.

Chapter 1 presents the basic properties of UWB. We examine the power spectral density, basic pulse shape, and spectral shape of these pulses. The regulatory requirements laid down by the FCC are briefly described. Why UWB is considered to be a multipath-resistant form is also examined, and such basic features of merit such as capacity and speed of data transmission are considered. We finish the chapter with a look at the cost, size, and power consumption that is forecast for UWB devices and chipsets.

Chapter 2 examines in detail how to generate pulse waveforms for UWB systems, for both simple cases, such as the Gaussian pulse shape, and more complex orthogonal pulses. We examine the possibility of designing pulses to fit spectral masks, such as mandated by regulators, or to avoid interference with other frequency bands. We finish the chapter with a look at some practical constraints and the effects of imperfections on these designer pulse shapes.

Chapter 3 looks at different signal-processing techniques for UWB systems. The chapter begins with a review of basic signal-processing techniques, including both time and frequency domain techniques. The Laplace, Fourier, and z-transforms are reviewed and their application to UWB is discussed. Finally, some practical issues, such as pulse detection and amplification, are discussed.

The wireless indoor channel and how it should be modeled for UWB communications is considered in Chapter 4. Following our basic pattern we define and explore the basic concepts of wideband channel modeling and show a simplified UWB multipath channel model that is amenable to both theoretical analysis and simulation. Path loss effects and a two-ray model are presented. Finally, the frequency domain autoregressive model is discussed.

Chapter 5 takes a look at some of the fundamental communication concepts and how they should be applied to UWB. First, modulation methods applicable to UWB are presented. A basic communication system consisting of transmitter, receiver, and channel is discussed. Since most consumer communication systems do not consist of only one user, multiple access techniques are introduced. The simple capacity of a UWB system is derived. Since other wireless consumer communication systems have already become popular, a comparison between UWB and other wideband techniques is included. Finally, the chapter ends with a look at interference to and from UWB systems.

Chapter 6 is concerned with ultra wideband antennas and arrays of antennas. This is considered one of the most difficult problems that must be overcome before the widespread commercialization of UWB devices takes place. Antenna fundamentals are first introduced, including Maxwell's equations for free space, antenna field regions, directivity, and gain. The suitability of conventional antennas for UWB transmission and reception is discussed in detail. More suitable impulse

antennas are then introduced. Arrays of antennas and beamforming for UWB systems are given a brief treatment.

Positioning and location using both traditional techniques and UWB is discussed in Chapter 7. Traditional location systems are first introduced and their pros and cons discussed. The advantages of UWB, particularly the extremely precise positioning that is theoretically possible, is examined. Finally, several possible scenarios are discussed where the precise location capabilities and high data rate of UWB can be combined to produce some new and exciting applications.

Chapter 8 concludes the book with a brief look at some current applications that use UWB technology, as well as an overview of some current chipsets and possible future UWB products. Emphasis is on consumer communication; however, military applications are given a brief treatment.

For the reader who wants a fast-track understanding of UWB and some knowledge of the current situation, we recommend the introduction, Chapter 1 (Basic properties of UWB signals and systems), and Chapter 8 (Applications).

For students who want to look at UWB in more detail, they should then proceed to look at Chapters 2 (The generation of UWB waveforms), Chapter 3 (Signal processing techniques for UWB systems) and then Chapters 4 through to 7 as required; in other words, the entire book with the possible exception of Chapter 8. We have strived to make each chapter complete in itself as far as possible and provide as much basic theory as possible, including derivations where appropriate. We have made constant reference to the literature, a significant part of which is covered here.

As an extra resource we have set up a companion website for our book containing a solutions manual, matlab programs for the examples and problems, and a sample chapter. Also, for those wishing to use this material for lecturing purposes, electronic versions of most of the figures from our book are available. Please go to the following URL and take a look: `ftp://ftp/pub/books/ghavami:uwb`

We hope that you will find this book useful as both a reference, a learning tool, and a stepping stone to further your own efforts in this exciting field.

M. Ghavami
L. B. Michael
R. Kohno

London, May 2004

Acknowledgments

The authors would like to thank the following for their efforts and contributions to *Ultra Wideband Signals and Systems in Communication Engineering*:

Sarah Hinton, our editor, for her tireless and unending efforts to make this publication timely and well received, as well as for helping us with the ins and outs of writing a textbook.

Professor Hamid Aghvami, the director of the Centre for Telecommunications Research, London, for providing a rich research environment and for his encouragement during the preparation of the book.

Dr. Mario Tokoro of Sony Computer Science Laboratories, Tokyo, for providing a free research environment to explore new ideas. It was here that the authors first began to study ultra wideband signals and systems and their applications and the idea for this book was born.

The following made valuable contributions by reviewing and in some cases contributing material to the book:

Dr. B. Allen (King's College London)
S. Ciolino (King's College London)
R. S. Dilmaghani (King's College London)
D. Karveli (King's College London)
S. McGregor (King's College London)
C. Mitchell (Yokohama National University)
Dr. T. Otsuki (Tokyo University of Science)

In addition, S. Ciolino and R. S. Dilmaghani of King's College London helped to contribute material to this book.

M. Ghavami would like to thank:

my wife and my children who have suffered the long period of preparation of this book and who have been continually supportive.

L. B. Michael would like to thank:

my wife and children for their support and patience during the weekends and nights while I was preparing and editing material for this book.

List of Figures

List of Tables

Introduction

In this chapter we present a general background to UWB and try to explain without resorting to too many equations the reasons UWB is considered to be an exciting and breakthrough technology. We place UWB in its historical background and show that, while UWB is not necessarily entirely new in either the concept or the signal-processing techniques used, given the recent emphasis in wireless communication on sinusoidal system, UWB does present a paradigm shift for many engineers.

We believe the current (and for the foreseeable future) emphasis on low power, low interference and low regulation makes the use of UWB an attractive option for current and future wireless applications.

I.1 ULTRA WIDEBAND OVERVIEW

Historically, UWB radar systems were developed mainly as a military tool because they could "see through" trees and beneath ground surfaces. However, recently, UWB technology has been focused on consumer electronics and communications. Ideal targets for UWB systems are low power, low cost, high data rates, precise positioning capability and extremely low interference.

Although UWB systems are years away from being ubiquitous, the technology is changing the wireless industry today. UWB technology is different from conventional narrowband wireless transmission technology – instead of broadcasting on separate frequencies, UWB spreads signals across a very wide range of frequencies. The typical sinusoidal radio wave is replaced by trains of pulses at hundreds of millions of pulses per second. The wide bandwidth and very low power makes UWB transmissions appear as background noise.

I.2 A NOTE ON TERMINOLOGY

The name ultra wideband is an extremely general term to describe a particular technology. Many people feel other names, such as pulse communications, may be more descriptive and suitable. However, UWB has become the term by which most people refer to ultra wideband technology.

The question then arises as to how to spell UWB. Is it "ultrawideband", "ultra-wideband", "ultra wide band", "ultrawide band" or "ultra wideband"? In this text, quite arbitrarily, we decide to use the term *ultra wideband*. Our reasoning is that the term wideband communication has become very common in recent years and is one that most people are familiar with. To show that UWB uses an even larger bandwidth the extra large "ultra" is prefixed; however, both "ultra-wideband" and "ultra-wideband" seem unwieldy, so we use ultra wideband. Many people may disagree about our choice, even vehemently. We accept their arguments and suggest that time will show the most popular choice for UWB.

I.3 HISTORICAL DEVELOPMENT OF UWB

Most people would see UWB as a "new" technology, in the sense that it provides the means to do what has not been possible before, be that high data rates, smaller, lower powered devices or, indeed, some other new application. However, UWB is, rather, a *new engineering technology* in that no new physical properties have been discovered.

However, the dominant method of wireless communication today is based on sinusoidal waves. Sinusoidal electromagnetic waves have become so universal in radio communications that many people are not aware that the first communication systems were in fact pulse-based. It is this paradigm shift for today's engineers from sinusoids to pulses that requires the most shift in focus.

In 1893 Heinrich Hertz used a spark discharge to produce electromagnetic waves for his experiment. These waves would be called colored noise today. Spark gaps and arc discharges between carbon electrodes were the dominant wave generators for about 20 years after Hertz's first experiments.

However, the dominant form of wireless communications became sinusoidal, and it was not until the 1960s that work began again in earnest for time domain electromagnetics. The development of the sampling oscilloscope in the early 1960s and the corresponding techniques for generating sub-nanosecond baseband pulses sped up the development of UWB. Impulse measurement techniques were used to characterize the transient behavior of certain microwave networks.

From measurement techniques the main focus moved to develop radar and communications devices. In particular, radar was given a lot of attention because of the accurate results that could be obtained. The low-frequency components were useful in penetrating objects, and *ground-penetrating radar* was developed. See references [1] and [2] for more details about UWB radar systems.

In 1973 the first US patent was awarded for UWB communications [3]. The field of UWB had moved in a new direction. Other applications, such as automobile collision avoidance, positioning systems, liquid-level sensing and altimetry were developed. Most of the applications and development occurred in the military or work funded by the US Government under classified programs. For the military, accurate radar and low probability of intercept communications were the driving forces behind research and development.

It is interesting to note that in these early days, UWB was referred to as *baseband*, *carrier-free* and *impulse* technology. The US Department of Defense is believed to be the first to have started to use the term *ultra wideband*.

The late 1990s saw the move to commercialize UWB communication devices and systems. Companies such as Time Domain [4] and in particular startups like XtremeSpectrum [5] were formed around the idea of consumer communication using UWB.

For further historical reading, the interested reader is referred to [6] and [7].

I.4 KEY BENEFITS OF UWB

The key benefits of UWB can be summarized as

1. high data rates
2. low equipment cost
3. multipath immunity
4. ranging and communication at the same time

We will expand on these benefits in the coming chapters, but first we give a brief overview.

The high data rates are perhaps the most compelling aspect from a user's point of view and also from a commercial manufacturer's position. Higher data rates can enable new applications and devices that would not have been possible up

until now. Speeds of over 100 Mbps have been demonstrated, and the potential for higher speeds over short distances is there. The extremely large bandwidth occupied by UWB gives this potential, as we show in the next section.

The ability to directly modulate a pulse onto an antenna is perhaps as simple a transmitter as can be made, leading many manufacturers to get excited by the possibilities for extremely cheap transceivers. This is possible by eliminating many of the components required for conventional sinusoidal transmitters and receivers.

The narrow pulses used by UWB, which also give the extremely wide bandwidth, if separated out provide a fine resolution of reflected pulses at the receiver. This is important in any wireless communication, as pulses (or sinusoids) interfering with each other are the major obstacle to error-free communication.

Finally, the use of both precise ranging (object location) and high speed data communication in the same wireless device presents intriguing possibilities for new devices and applications. Simultaneous automotive collision avoidance radar and communication can give accident-free smooth traffic flow, or games where the players' position can be precisely known and a high speed wireless link seamlessly transfers a video signal to the players' goggles may seem the stuff of science fiction, but with UWB the possibilities for these and other applications are there, right now.

I.5 UWB AND SHANNON'S THEORY

Perhaps the benefits and possibilities of UWB can be best summarized by examining Shannon's famous capacity equation. This equation will be familiar to anyone who has studied communication or information theory. Capacity is important as more demanding audio-visual applications require higher and higher bit rates.

Shannon's equation is expressed as

$$C = B \log \left(1 + \frac{S}{N} \right) \tag{I.1}$$

where C is the maximum channel capacity, with units [bits/second]; B is the channel bandwidth [Hz]; S is the signal power in watts [W] and N is the noise power also in Watts.

This equation tells us that there are three things that we can do to improve the capacity of the channel. We can increase the bandwidth, increase the signal power or decrease the noise. The ratio S/N is more commonly known as the *signal-to-noise ratio* (SNR) of the channel. We also can see that the capacity of a channel grows linearly with increasing bandwidth B, but only logarithmically with signal power S.

The ultra wideband channel has an abundance of bandwidth and in fact can trade off some of the bandwidth for reduced signal power and interference from

other sources. Thus, from Shannon's equation we can see that UWB systems have a great potential for high-capacity wireless communications.

Another way of looking at wireless communication is the tradeoffs between

- the distance between transmitter and receiver
- simultaneous communication for many users
- sending the data very quickly
- sending and receiving a large amount of data

The first wireless communication systems, such as wireless communication at sea, were meant to communicate between ships separated by large distances. However, the amount of data that could be effectively transferred was extremely small and communication took a long time. Only one person can "talk" using Morse code at a time. More recently, cellular telephone systems have simultaneous communication for many users as their forté. The distance between the base station and the user is limited to at most a few kilometers. It can be classified as a system where a moderate amount of data can be sent reasonably quickly. An ultra wideband system is focused on the latter two attributes: a large amount of data that can be transmitted very quickly. This is at the expense of, in the main, distance. The precise tradeoffs are of course more complex and will depend upon the particular application.

I.6 CHALLENGES FOR ULTRA WIDEBAND

While UWB has many reasons to make it an exciting and useful technology for future wireless communications and many other applications, it also has some challenges which must be overcome for it to become a popular and ubiquitous technology.

Perhaps the most obvious one to date has been regulatory problems. Wireless communications have always been regulated to avoid interference between different users of the spectrum. Since UWB occupies such a wide bandwidth, there are many users whose spectrum will be affected and need to be convinced that UWB will not cause undue interference to their existing services. In many cases these users have paid to have exclusive use of the spectrum.

Other challenges include the industry coming to agreed standards for interoperability of UWB devices. At present no clear consensus has emerged, and the possibility of several competing UWB standards is extremely likely.

Many technical and implementation issues remain. The promise of low-cost devices is there, but the added complexity to combat interference and low-power operation may bring cost increases similar to current wireless devices.

I.7 SUMMARY

In this chapter we presented a general background to UWB and explained the reasons UWB is considered to be an exciting and breakthrough technology, particularly from the viewpoint of Shannon's famous capacity equation. We placed UWB in its historical background and showed the development of UWB from radar to communications applications. We showed the differences in the concept of the signal-processing techniques used for sinusoidal narrowband systems and those for pulse-based UWB systems.

Problems

Problem 1. Investigate the current regulations for UWB in your country. List other uses of the same wireless bandwidth.

Problem 2. Read and summarize a UWB journal or conference paper published before 1990. Discuss how the UWB technology described in that paper has changed. You may want to compare and contrast a more recent paper discussing the same topic.

Problem 3. Many wireless technologies, including UWB, were first used and developed by and for the military. Discuss your views on this progression of technology from the military to consumer markets. What are the possible pros and cons.

1

Basic properties of UWB signals and systems

1.1 INTRODUCTION

In this chapter the basic properties of UWB signals and systems are outlined, with details of each of the characteristics here explained in later chapters.

First, we examine the basic shape of an ultra wideband pulse in the time domain and see what the spectrum content is of these pulses. Generally speaking, the extremely short pulses with fast rise and fall times have a very broad spectrum and a very small energy content. We examine the regulatory aspects of power output and frequency by spectral masks.

Next, we see that because UWB pulses are extremely short they can be filtered or ignored. They can readily be distinguished from unwanted multipath reflections because of the fine time resolution. This leads to the characteristic of multipath immunity.

Furthermore, UWB pulses' low frequency components enable the signals to propagate effectively through materials such as bricks and cement.

The large bandwidth of UWB systems means extremely high data rates can be achieved, and we show that UWB systems have a potentially high spectral capacity.

UWB transmitters and receivers do not require expensive and large components such as modulators, demodulators and IF stages. This fact can reduce cost, size, weight, and power consumption of UWB systems compared with conventional narrow-band communication systems.

1.2 POWER SPECTRAL DENSITY

The power spectral density of UWB systems is generally considered to be extremely low, especially for communication applications. The power spectral density (PSD) is defined as

$$\text{PSD} = \frac{P}{B} \tag{1.1}$$

where P is the power transmitted in watts (W), B is the bandwidth of the signal in hertz (Hz), and the unit of PSD is watts/hertz (W/Hz).

Historically, wireless communications have only used a narrow bandwidth and can hence have a relatively high power spectral density. We can put this another way: since we know that frequency and time are inversely proportional, sinusoidal systems have narrow B and long time duration t. For a UWB system the pulses have a short t and very wide bandwidth B. It is helpful to review some traditional wireless broadcast and communication applications and calculate their PSDs as shown in Table 1.1.

Table 1.1 Power spectral density of some common wireless broadcast and communication systems.

System	Transmission power [W]	Bandwidth [Hz]	Power spectral density [W/MHz]	Classification
Radio	50 kW	75 kHz	666,600	narrowband
Television	100 kW	6 MHz	16,700	narrowband
2G Cellular	10 mW	8.33 kHz	1.2	narrowband
802.11a	1 W	20 MHz	0.05	wideband
UWB	1 mW	7.5 GHz	0.013	ultra wideband

The energy used to transmit a wireless signal is not infinite and, in general, should be as low as possible, especially for today's consumer electronic devices. If we have a fixed amount of energy we can either transmit a great deal of energy density over a small bandwidth or a very small amount of energy density over a large bandwidth. This comparison is shown for the PSD of two systems in Figure 1.1. The total amount of power can be calculated as the area under a frequency-power spectral density graph.

For UWB systems the energy is spread out over a very large bandwidth (hence the name ultra wideband) and, in general, is of a very low power spectral density. The major exception to this general rule of thumb is UWB radar systems which transmit at high power over a large bandwidth. However, here we will restrict ourselves to the communications area.

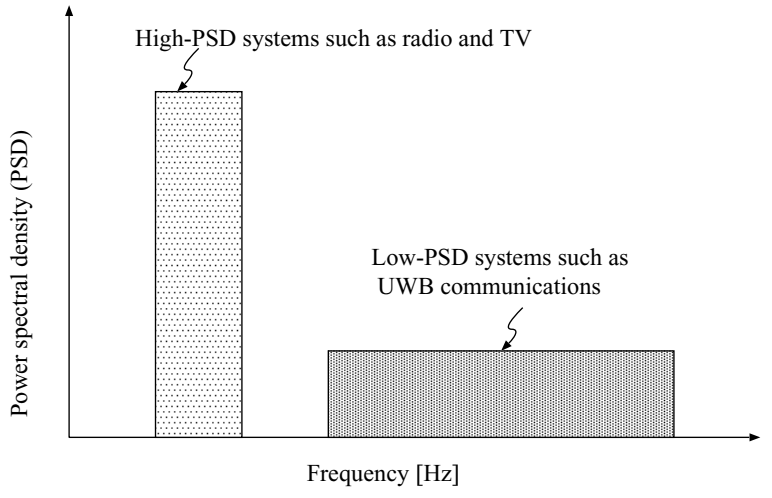


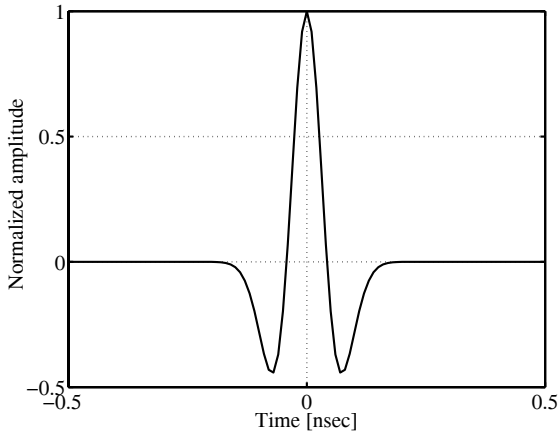
Fig. 1.1 Low-energy density and high-energy density systems.

One of the benefits of low-power spectral density is a *low probability of detection*, which is of particular interest for military applications: for example, covert communications and radar. This is also a concern for wireless consumer applications, where the security of data for corporations and individuals using current wireless systems is considered to be insufficient [8].

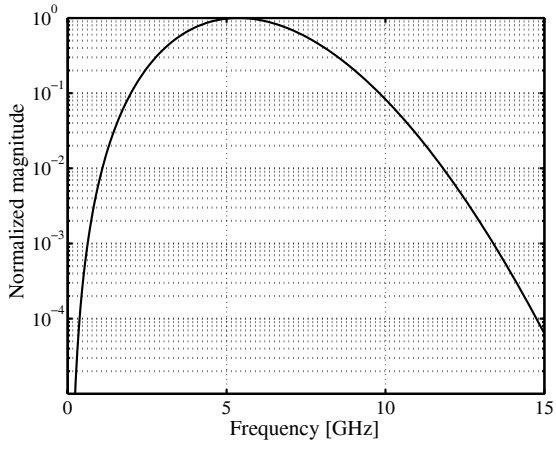
1.3 PULSE SHAPE

A typical received UWB pulse shape, sometimes known as a *Gaussian doublet*, is shown in Figure 1.2. More details regarding Gaussian and other waveforms are discussed in Chapter 2. This pulse is often used in UWB systems because its shape is easily generated. It is simply a square pulse which has been shaped by the limited rise and fall times of the pulse and the filtering effects of the transmit and receive antennas. A square pulse can be easily generated by switching a transistor on and off quickly.

We show a simple pulse generator model in Figures 1.3 and 1.4, which demonstrate the creation of Gaussian doublets at a transmitter, antenna effects and reception. We start with a rectangular pulse in Figure 1.4(a). Ultra wideband pulses are typically of nanosecond or picosecond order. The fast switching on and off leads to a pulse shape which is not rectangular, but has the edges smoothed off. The pulse shape approximates the Gaussian function curve. A Gaussian function



(a)



(b)

Fig. 1.2 (a) Idealized received UWB pulse shape w_{rx} and (b) idealized spectrum of a single received UWB pulse.

$G(x)$ is one which fits the well-known equation

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/\sqrt{2\sigma^2}} \tag{1.2}$$

where Eqn. (1.2) is assumed to be zero-mean. This is the origin of the name Gaussian pulse, monocycle or doublet. A simple circuit for creation of the Gaussian doublet is shown in Figure 1.3. Transmitting the pulses directly to the antennas

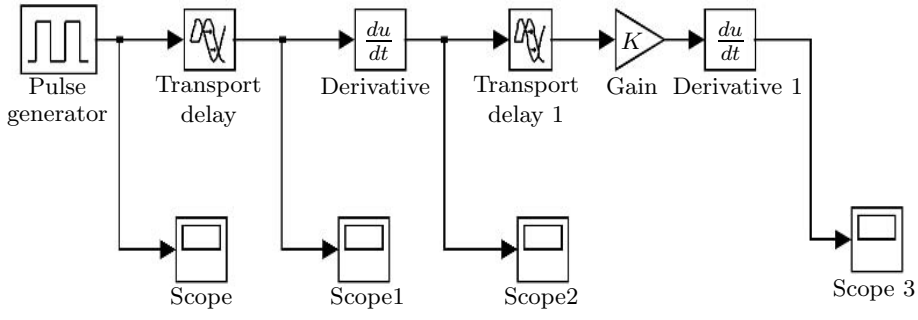


Fig. 1.3 A simple Matlab circuit model to create the Gaussian doublet.

results in the pulses being filtered due to the properties of the antennas. This filtering operation can be modeled as a derivative operation [9]. The same effect occurs at the receive antenna. Here, we model the channel as a delay and assume the pulse is amplified at the receiver.

In this chapter we will limit our discussion to this typical received pulse shape, which is assumed for the large majority of UWB research and is reasonably close to received pulses that have been measured. Details of pulse generation and a detailed discussion of pulse shaping can be found in Chapter 2.

This idealized received pulse shape p_{rx} can be written as [10]

$$p_{rx} = \left[1 - 4\pi \left(\frac{t}{\tau_m}\right)^2\right] e^{-2\pi(t/\tau_m)^2} \tag{1.3}$$

which is the equation used to generate the pulse shown in Figure 1.2(a). Here, τ_m is assumed to be 0.15. It should be mentioned that τ_m is the single parameter of Eqn. (1.3) and determines the time and frequency characteristics of the Gaussian doublet uniquely.

The spectrum of the Gaussian doublet is shown in Figure 1.2(b). The center frequency can be seen to be approximately 5 GHz, with the 3 dB bandwidth extending over several GHz. In comparison with narrowband or even wideband communication systems the large bandwidth is evident and, hence, the name ultra wideband communication can easily be inferred.

1.4 PULSE TRAINS

One pulse by itself does not communicate a lot of information. Information or data needs to be modulated onto a sequence of pulses called a *pulse train*.

When pulses are sent at regular intervals, which is sometimes called the *pulse repetition rate* or the *duty cycle*, the resulting spectrum will contain peaks of power

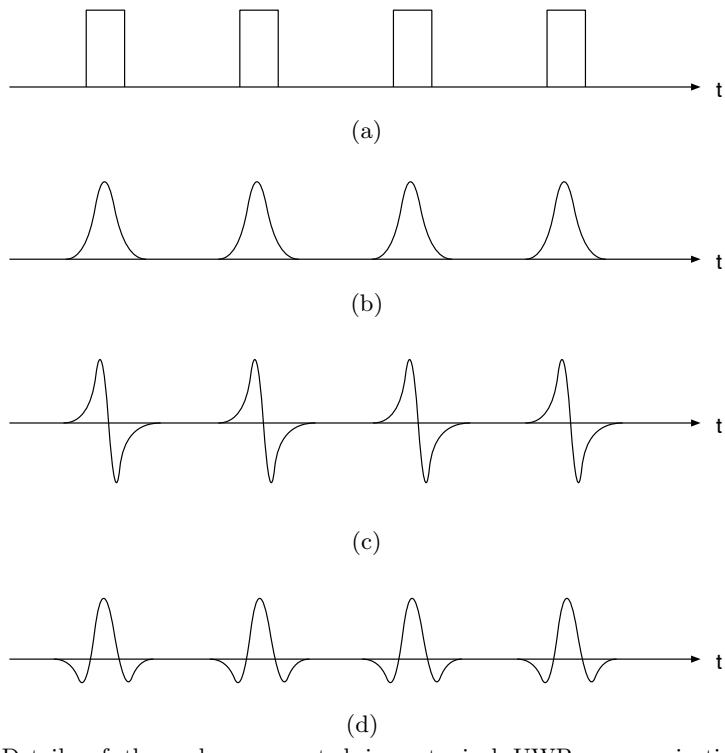


Fig. 1.4 Details of the pulses generated in a typical UWB communication system: (a) square pulse train; (b) Gaussian-like pulses; (c) first-derivative pulses; (d) received Gaussian doublets.

at certain frequencies. These frequencies are the inverse of the pulse repetition rate. These peak power lines are called *comb lines* because they look like a comb. See Figure 1.5(b) for an example.

These peaks limit the total transmit power undesirably. One method of making the spectrum more noise-like is to "dither" the signal by adding a small random offset to each pulse, either delaying the pulse or transmitting slightly before the regular pulse time. The resultant spectrum from such a random offset is shown in Figure 1.6 and should be compared with Figure 1.5(b).

As we will see in Chapter 5, by making this delay not completely random but cyclic according to a known pseudo-noise (PN) code, information can be modulated onto the pulse waveform. This is known as *pulse position modulation* (PPM) and has been investigated in different communication systems such as optical wireless communications.

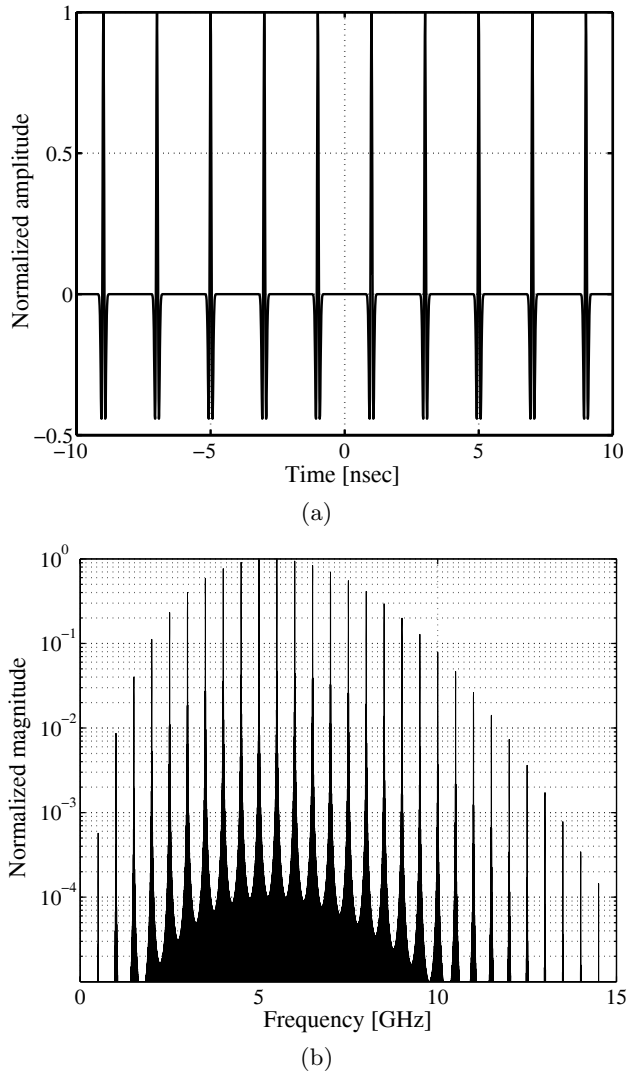


Fig. 1.5 (a) UWB pulse train and (b) spectrum of a UWB pulse train.

1.5 SPECTRAL MASKS

The spectrum of a UWB signal is one of the major issues confronting the industry and governments for the commercial use of UWB. In fact, the very name ultra wideband suggests that the issue of spectrum is at the very heart of the ultra wideband technology.

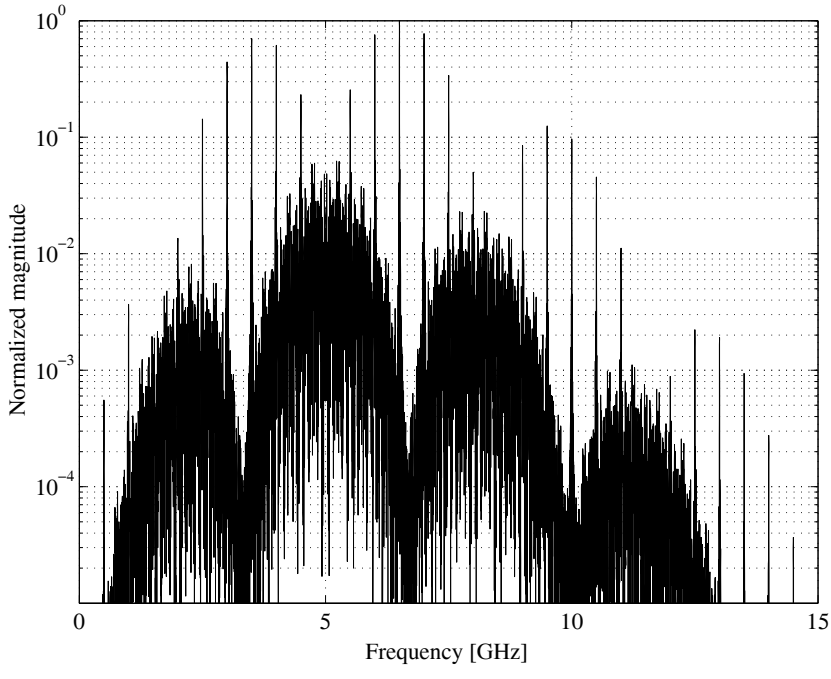


Fig. 1.6 Spectrum of a pulse train which has been "dithered" by shifting pulses forward and back of the original position.

All radio communication is subject to different laws and regulations about power output in certain frequency bands. This is to prevent interference to other users in nearby or the same frequency bands.

Ultra wideband systems cover a large spectrum and interfere with existing users. In order to keep this interference to the minimum the FCC and other regulatory groups specify *spectral masks* for different applications which show the allowed power output for specific frequencies.

In Figure 1.7 an example is shown of the FCC spectral mask for indoor UWB systems. A large contiguous bandwidth of 7.5 GHz is available between 3.1 GHz and 10.6 GHz at a maximum power output of −41.3 dBm/MHz.

The major reasons for the extremely low allowed power output in the frequency bands 0.96 GHz–1.61 GHz is due to pressure from groups representing existing services, such as mobile telephony, global positioning system (GPS), and military usage. The allowed level of −41.3 dBm/MHz itself is considered conservative, and many groups have lobbied for higher allowed power output.

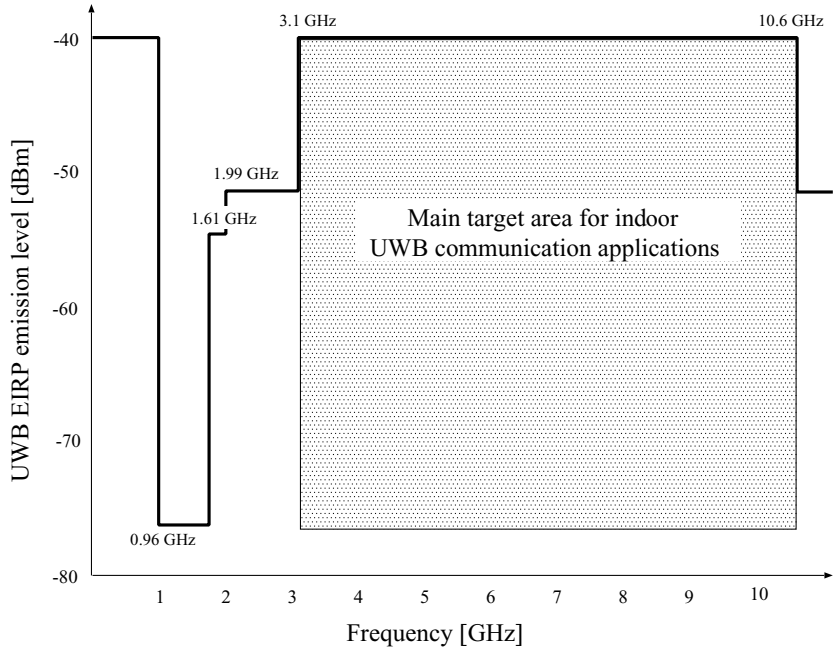


Fig. 1.7 Spectral mask mandated by FCC 15.517(b,c) for indoor UWB systems.

1.6 MULTIPATH

In this section we will look at the effects of multipath, particularly an indoor wireless channel, on the basic UWB pulse we have described. We will see that, because of the extremely short pulse widths, if these pulses can be resolved in the time domain then the effects of multipath, such as *inter-symbol interference* (ISI), can be mitigated.

Multipath is the name given to the phenomenon at the receiver whereby after transmission an electromagnetic signal travels by various paths to the receiver. See Figure 1.8 for an example of multipath propagation within a room. This effect is caused by reflection, absorption, diffraction, and scattering of the electromagnetic energy by objects in-between the transmitter and the receiver. If there were no objects to absorb or reflect the energy, this effect would not take place and the energy would propagate outward from the transmitter, dependent only on the transmit antenna characteristics. However, in the real world objects between the transmitter and the receiver cause the physical effects of reflection, absorption, diffraction, and scattering, and this gives rise to multiple paths. Due to the

different lengths of the paths, pulses will arrive at the receiver at different times, with the delay proportional to the path length.

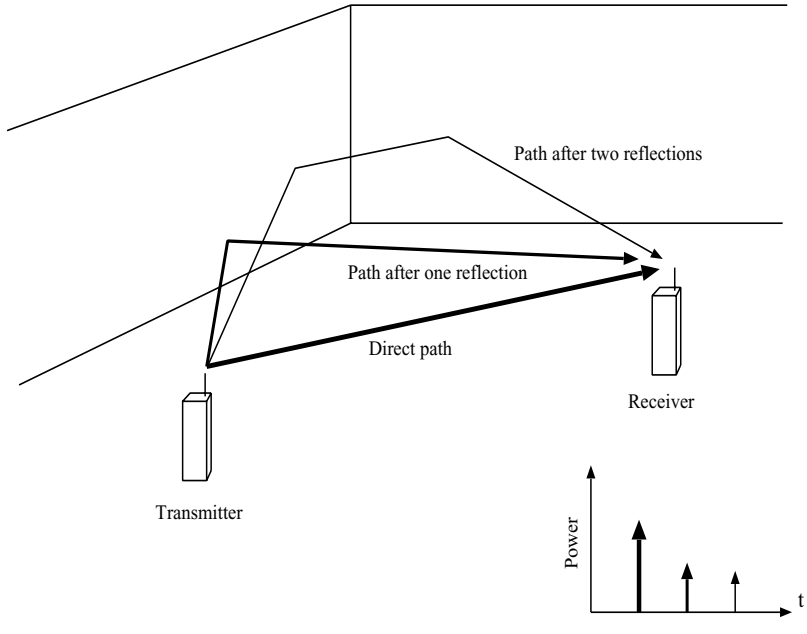


Fig. 1.8 A typical indoor scenario in which the transmitted pulse is reflected off objects within the room, thus creating multiple copies of the pulse at the receiver, with different delays.

UWB systems are often characterized as *multipath immune* or *multipath resistant*. Examining the pulses described previously, we can see that if pulses arrive within one pulse width they will interfere, while if they are separated by at least one pulse width they will not interfere. If pulses do not overlap, then they can be filtered out in the time domain or, in other words, ignored. Assuming one symbol per pulse, they will not produce interference with the same symbol. Alternatively the energy can be summed together by a *rake receiver*. Figures 1.9 and 1.10 demonstrate non-overlapping and overlapping pulses, respectively.

Example 1.1

Assuming a received pulse shape similar to Figure 1.2, how much extra distance must a second pulse travel to not interfere with the original pulse? If the pulse width was halved, what would be the separation between multiple paths needed?

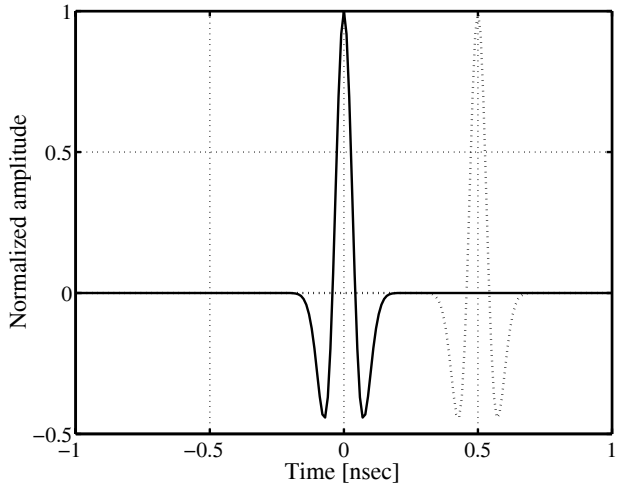


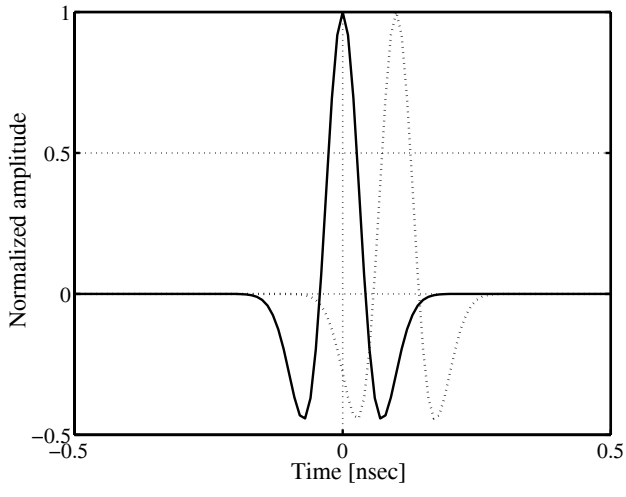
Fig. 1.9 Two pulses arriving with a separation greater than the pulse width will not overlap and will not cause interference.

Solution

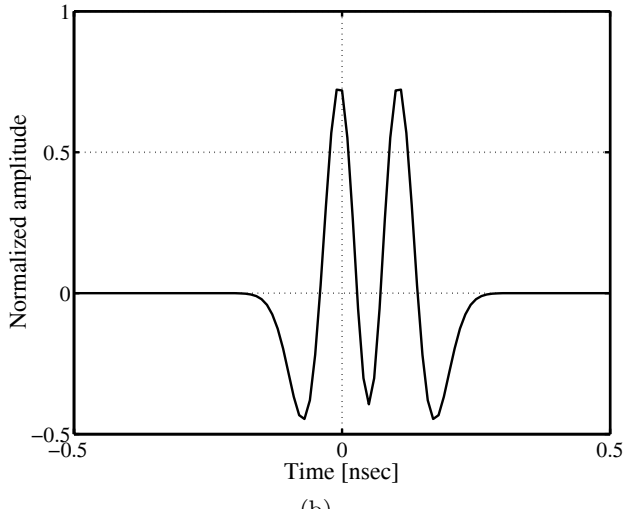
From Figure 1.2 the pulse width is approximately 0.4 ns. Using the relation that distance is the product of velocity by time travelled, $d = v \cdot t$, and since electromagnetic energy travels at a velocity of approximately 3×10^8 meters per second, the extra distance travelled via the second path to avoid interference at the receiver is 12 cm. If the pulse width was halved, the required distance between multipaths to avoid interference would be halved also, to 6 cm.

As we can see from Example 1.1 the separation distance required between multipaths decreases with decreasing pulse width. This is one reason for smaller pulse widths, particularly in indoor environments.

Another method to avoid multipath interference is to lower the duty cycle of the system. By transmitting pulses with time delays greater than the maximum expected multipath delay, unwanted reflections can be avoided at the receiver. This is inherently inefficient and places limits on the maximum speed of data transmission for a given modulation system. In the limit, if pulses were transmitted continuously, then the system would resemble a sinusoidal system. These issues are discussed in Chapter 5.



(a)



(b)

Fig. 1.10 (a) Two overlapping UWB pulses, and (b) the received waveform consisting of the overlapped pulses.

1.7 PENETRATION CHARACTERISTICS

One of the most important benefits of the UWB communication system that has been raised is the ability of pulses to easily penetrate walls, doors, partitions, and other objects in the home and office environment. In this section we will examine

the reported results for penetration of UWB pulses and comment on how that will affect communication in the home and office.

Frequency f and wavelength λ are related by the speed of light c as is shown in the well-known Eqn. (1.4).

$$\lambda[m] = \frac{c\ [\mathrm{m/s}]}{f\ [\mathrm{Hz}]} \tag{1.4}$$

In other words, as the frequency increases the wavelength becomes shorter, and for lower frequencies the wavelength is longer.

In conventional sinusoidal communication, lower frequency waves have the characteristic of being able to "pass through" walls, doors, and windows because the length of the wave is much longer than the material that it is passing through. On the other hand, higher frequency waves will have more of their energy reflected from walls and doors since their wavelength is much shorter.

Ultra wideband pulses are composed of a large range of frequencies, as is shown in Figure 1.2(b). One of the basic characteristics of early prototype UWB communication systems was its ability to "pass through walls", especially in comparison with IEEE 802.11 wireless local area network systems. The penetration capabilities of UWB come only from the lower frequency components which were for early systems mostly centered on 1 GHz. Since the 2002 ruling by the FCC (see Figure 1.7) the center frequency for most UWB systems has substantially increased. This means that the penetration characteristics of the signals have substantially decreased, especially in comparison with the IEEE 802.11b systems which are centered at 2.4 GHz.

1.8 SPATIAL AND SPECTRAL CAPACITIES

Another basic property of UWB systems is its high *spatial capacity*, measured in bits per second per square meter (bps/m^2) [11]. Spatial capacity is a relatively recent term, and its use stems from the interest in even higher data rates, even over extremely short distances.

Spatial capacity can be calculated as the maximum data rate of a system divided by the area over which that system can transmit. The transmission area can be calculated from the circular area assuming a transmitter in the center; however, in practice a rule of thumb is to use the square of the maximum transmission distance

$$\text{Spatial capacity } [\mathrm{bps/m^2}] = \frac{\text{Maximum data rate [bps]}}{\text{Transmission area } [\mathrm{m^2}]} \tag{1.5}$$

$$\text{Transmission area } [\mathrm{m^2}] = \pi \times (\text{Transmission distance})^2 \tag{1.6}$$

For narrowband systems the most popular measure of capacity has been *spectral capacity*, measured in bits per second per hertz (bps/Hz), because spectrum has

Table 1.2 Comparison of spatial capacity of various indoor wireless systems.

System name	Maximum data rate [Mbps]	Transmission distance [m]	Spatial capacity [kbps/m^2]	Spectral capacity [bps/Hz]
UWB	100	10	318.3	0.013
IEEE 802.11a	54	50	6.9	2.7
Bluetooth	1	10	3.2	0.012
IEEE 802.11b	11	100	0.350	0.1317

been the most limited resource. Power has generally only been limited by safety and commercial reasons, such as the battery life of mobile devices

$$\text{Spectral capacity [bps/Hz]} = \frac{\text{Maximum data rate [bps]}}{\text{Bandwidth [Hz]}} \tag{1.7}$$

For UWB systems, which operate in other licensed spectrum, the power has to be kept very low. This is compensated for by the use of extremely large bandwidths. Using the traditional measure of spectral capacity [bits/Hz] UWB has very low spectral capacity compared with existing systems. However, when comparing spatial capacity, UWB is extremely efficient. Table 1.2 shows the comparison of spatial and spectral capacity among various indoor wireless systems.

1.9 SPEED OF DATA TRANSMISSION

One of the advantages of UWB transmission for communications is its high data rate. While current chipsets are continually being improved, most UWB communication applications are targeting the range of 100 Mbps to 500 Mbps [12], which is roughly the equivalent of wired Ethernet to USB 2.0. It is significant that this data rate is 100 to 500 times the speed of Bluetooth, around 50 times the speed of the 802.11b, or 10 times the 802.11a wireless LAN (local area network) standards.

As can be seen in Table 1.3 the current target data rate for indoor wireless UWB transmission is between 110 Mbps and 480 Mbps. This is fast compared with current wireless and wired standards. In fact, the speed of transmission is currently being standardized into three different speeds: 110 Mbps with a minimum transmission distance of 10 m, 200 Mbps with a minimum transmission distance of 4 m and 480 Mbps with no fixed minimum distance.

The reasons for these particular distances lie mostly with different applications. For example, 10 m will cover an average room and may be suitable for wireless connectivity for home theater. A distance of less than 4 m will cover the distance

Table 1.3 Comparison of UWB bit rate with other wired and wireless standards.

Speed (Mbits/second)	Standard
480	UWB, USB 2.0
200	UWB (4 m minimum), 1394a (4.5 m)
110	UWB (10 m minimum)
90	Fast Ethernet
54	802.11a
20	802.11g
11	802.11b
10	Ethernet
1	Bluetooth

between appliances, such as a home server and a television. A distance of less than 1 m will cover the appliances around a personal computer.

1.10 COST

Among the most important advantages of UWB technology are those of low system complexity and low cost. Ultra wideband systems can be made nearly "all-digital", with minimal (radio frequency) RF or microwave electronics. The low component count leads to reduced cost, and smaller chip sizes invariably lead to low-cost systems. The simplest UWB transmitter could be assumed to be a pulse generator, a timing circuit, and an antenna.

However, as higher data rates are required more complex timing circuitry is needed. To provide a multiple access system, additional complexity is required. Rake receivers add further circuitry, and the cost increases. Furthermore, chipset costs depend heavily on the number of units manufactured.

To reduce costs, during later product cycles more functionality is implemented on fewer chips, reducing die area and, thus, manufactured cost.

Therefore, at this early stage it is extremely difficult to quantify the cost of UWB communication systems. To take one early example, it has been reported [13] that the XtremeSpectrum chipset is priced at $19.95 for 100,000 units.

1.11 SIZE

The small size of UWB transmitters is a requirement for inclusion in today's consumer electronics. In the 802.15 working groups, consumer electronics companies have targeted the size of the wireless circuit will be small enough to fit into a

Memory Stick or SD Card [12]. A chipset by XtremeSpectrum has a small size which enables compact flash implementation [13].

The main arguments for the small size of UWB transmitters and receivers are due to the reduction of passive components. However, antenna size and shape is another factor that needs to be considered. Ultra wideband antennas are considered in Chapter 6.

1.12 POWER CONSUMPTION

With proper engineering design the resultant power consumption of ultra wideband can be quite low. As with any technology, power consumption is expected to decrease as more efficient circuits are designed and more signal processing is done on smaller chips at lower operating voltages.

The current target for power consumption of UWB chipsets is less than 100 mW. Table 1.4 shows some figures for power consumption of current chipsets [12].

Table 1.4 Power consumption of UWB and other mobile communication chipsets.

Application chipset	Power consumption [mW]
802.11a	1500–2000
400 Mbps 1394 LSI	700
Mobile telephone RISC 32-bit MPU	200
Digital camera 12-bit A/D converter	150
UWB (target)	100
Mobile telephone TFT color display panel	75
MPEG-4 decoder LSI	50
Mobile telephone voice codec LSI	19

1.13 SUMMARY

In this chapter the basic properties of UWB signals were outlined, starting with the basic shape and spectrum of an ultra wideband pulse. We saw that the power output and spectrum of UWB pulses are limited by regulation.

We showed that because UWB pulses are extremely short, with the consideration of fading, they can be filtered or ignored. They can readily be distinguished from unwanted multipath reflections because of the fine time resolution. This leads to the characteristic of multipath immunity.

UWB pulses' low-frequency components enable the signals to propagate effectively through materials, such as bricks and cement.

The large bandwidth of UWB systems means extremely high data rates can be achieved, and we showed that UWB systems have a potentially high spectral capacity.

Finally, we stated that UWB transmitters and receivers do not require expensive and large components, such as modulators, demodulators, and IF stages. This fact can reduce cost, size, weight, and power consumption of UWB systems compared with conventional narrowband communication systems.

Problems

Problem 1. Investigate the transmission power and bandwidth of three systems *not* shown in Table 1.1. (*Hint:* radio and television systems may differ significantly from those shown in the table due to differing local conditions.) Calculate their power spectral densities.

Problem 2. Implement the Matlab circuit of Figure 1.3, and confirm the output of the four scopes. Apply some modifications to the circuit if necessary.

Problem 3. Based on the Matlab circuit of Figure 1.3, implement a three-path multipath channel model. Add Gaussian noise. Show the output of the four scopes for two different values of delay of the second and third paths. Comment on the effect of multipath to the receiver and receiver design.

Problem 4. Write a Matlab program to output Figures 1.5(a) and (b).

Problem 5. Investigate the current FCC (or the regulator in your own country) regulations for ultra wideband applications. What differences or similarities can you find with Figure 1.7.

Problem 6. Investigate and calculate the spatial and spectral capacities of three other wireless systems (outdoor wireless systems are acceptable) *not* shown in Table 1.2.

Problem 7. Investigate the reported speed of at least five current commercial UWB communication or laboratory prototypes. Plot a graph, with time as the horizontal axis and speed as the vertical axis.

Problem 8. Find the cost of five different wireless chipsets. Compare and contrast the complexity of the chipset, maturity of the system, and cost of the chipset. What, in your opinion, is the dominating factor for the cost of wireless chipsets?

Problem 9. For one standard or system (for example, UWB or IEEE standard 802.11a) find the cost of five commercial products. What fraction of the total product cost is contained in the wireless chipset?

2

Generation of ultra wideband waveforms

One of the essential functions in communications systems is the representation of a message symbol by an analog waveform for transmission through a channel. As was shown in Chapter 1, in UWB systems the conventional analog waveform is a simple pulse that in general is directly radiated to the air. These short pulses have typical widths of less than 1 ns and thus bandwidth of over 1 GHz.

In this chapter we will examine in detail how to generate pulse waveforms for UWB systems, for both simple cases of Gaussian wave shapes and more complex orthogonal pulses. We will discuss how to design pulses which meet requirements of spectral masks as mandated by government organizations. Finally, we will look at the practical constraints on pulse generation, the effects of imperfections on the pulse, and briefly discuss the effects of the channel on pulse shape.

2.1 INTRODUCTION

Sinusoidal electromagnetic waves have become so universal in radio communications that many people are not aware that the first communication systems were in fact pulse-based. Heinrich Hertz (1893) used a spark discharge to produce the electromagnetic waves for his experiment. These waves would be called colored noise today. Spark gaps and arc discharges between carbon electrodes were the dominant wave generators for about 20 years after Hertz's first experiments. Eventually, the development of rotating high-frequency generators and the electronic tube made the generation of sinusoidal currents and waves possible.

A strong incentive to use sinusoidal waves was provided by the need to operate several transmitters at the same time but to receive them selectively. This led to the development of transmitters and receivers on the basis of sinusoidal waves. Regulation followed common practice and led to the assignment of frequency bands for various radio services. However, a quotation from a textbook published in 1920 shows that nonsinusoidal waves were still used at that time.

Today's UWB systems employ nonsinusoidal wave shapes that should have certain properties when transmitted from the antenna. Emissions in UWB communication systems are constrained by the FCC regulation 47 CFR Section 15.5(d) [14], which states that

> Intentional radiators that produce class B emissions (damped wave) are prohibited.

Several non-damped waveforms have been proposed in the literature for UWB systems, such as Gaussian [15], Rayleigh, Laplacian, cubic [16], and modified Hermitian monocycles [17]. In all these waveforms the goal is to obtain a nearly flat frequency domain spectrum of the transmitted signal over the bandwidth of the pulse and to avoid a DC component. In order to understand the characteristics of different waveforms, we first discuss the theoretical definition of a damped wave.

2.1.1 Damped sine waves

What is a damped sine wave? According to [18] a damped sine wave is of the form

$$y_d(t) = Ae^{-\lambda t}\sin(2\pi f_0 t) \tag{2.1}$$

where A is an arbitrary amplitude, λ is the exponential decay coefficient, f_0 is the frequency of oscillation of the sine wave, and $t \geq 0$ is the time. Figure 2.1(a) demonstrates this waveform, and any wave of this general form can be considered a damped sine wave, or Class B, emission.

Example 2.1

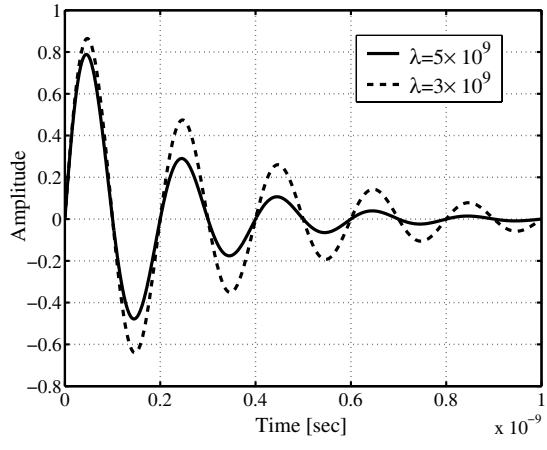
Calculate and sketch the time and frequency domain representations of Eqn. (2.1) for $A = 1$, $\lambda = 5 \times 10^9, 3 \times 10^9$, and $f = 5$ GHz.

Solution

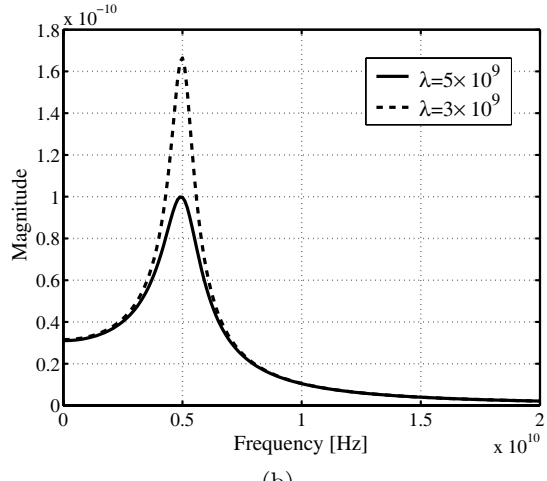
Using the Fourier transform we can derive the frequency domain formula from the time domain representation as follows

$$\begin{aligned} Y_d(f) &= \mathcal{F}\{y_d(t)\} \\ &= \int_0^\infty Ae^{-\lambda t}\sin(2\pi f_0 t)e^{-j2\pi f t}\,dt \\ &= \frac{2\pi f_0 A}{\lambda^2 - 4\pi^2(f^2 - f_0^2) + j4\pi\lambda f} \end{aligned} \tag{2.2}$$

where the final result of Eqn. (2.2) can be found by using a definite or indefinite table of integrals.
Substituting $A = 1$, $\lambda = 5 \times 10^9$, 3×10^9, and $f = 5$ the damped sine waves in the time domain are shown in Figure 2.1(a) and the frequency domain representation is given in Figure 2.1(b).



(a)



(b)

Fig. 2.1 Damped sine waves and their Fourier transforms.

The *effective bandwidth* W of a pulse is defined as

$$W = f_h - f_l \tag{2.3}$$

where f_l and f_h are the frequencies measured at nearest points with half of the maximum amplitude.

Example 2.2

For the damped sinusoids of Figure 2.1 calculate the effective bandwidth W in both cases. Comment on the relationship between W and the decay parameter λ.

Solution

The damped sinusoid of Figure 2.1 has a decay parameter $\lambda = 5 \times 10^9$. From Figure 2.1(b) we have $f_l = 3.25$ GHz and $f_h = 6.18$ GHz; hence, $W_d = 2.93$ GHz. When the decay parameter is $\lambda = 3 \times 10^9$ the effective bandwidth is $W_d = 1.69$ GHz.

The value of W depends on the frequency of oscillation f_0 and the decay parameter λ. If λ decreases, W also decreases.

Figure 2.1(b) clearly indicates that the damping oscillations of the waveform have led to a small effective bandwidth and a sharp peak in the spectrum. This is in contradiction with the required flatness of the spectrum of permitted waveforms by the FCC [14]. Waveforms with damped oscillatory tails cannot be used in UWB communication systems because they can seriously interfere with existing communication systems that are available in the area. For this and several other reasons that will be discussed in this chapter, the topic of wave shaping is very important and should be considered in the design of practical UWB transmitters.

2.2 GAUSSIAN WAVEFORMS

A class of waveforms are called *Gaussian waveforms* because their mathematical definition is similar to the Gauss function. The zero-mean Gauss function is described by Eqn. (2.4), where σ is standard deviation

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \tag{2.4}$$

The basis of these Gaussian waveforms is a *Gaussian pulse* represented by the following equation

$$y_{g_1}(t) = K_1 e^{-(t/\tau)^2} \tag{2.5}$$

where $-\infty < t < \infty$, τ is the time-scaling factor, and K_1 is a constant. More waveforms can be created by a sort of high-pass filtering of this Gaussian pulse. Filtering acts in a manner similar to taking the derivative of Eqn. (2.5). For example, a *Gaussian monocycle*, the first derivative of a Gaussian pulse, has the form

$$y_{g_2}(t) = K_2 \frac{-2t}{\tau^2} e^{-(t/\tau)^2} \tag{2.6}$$

where $-\infty < t < \infty$ and K_2 is a constant. A Gaussian monocycle has a single zero crossing. Further derivatives yield additional zero crossings, one additional zero crossing for each additional derivative. If the value of τ is fixed, by taking an additional derivative, the fractional bandwidth decreases, while the centre frequency increases.

A *Gaussian doublet* is the second derivative of Eqn. (2.5) and is defined by Eqn. (2.7)

$$y_{g_3}(t) = K_3 \frac{-2}{\tau^2} \left(1 - \frac{2t^2}{\tau^2}\right) e^{-(t/\tau)^2} \tag{2.7}$$

where $-\infty < t < \infty$ and K_3 is a constant. Figures 2.2(a) and (b) show a Gaussian pulse, a Gaussian monocycle, and a Gaussian doublet in time and frequency domains, respectively. In all three cases, $\tau = 50$ ps is assumed. The constants K_1, K_2, and K_3 are added to Eqns. (2.5)–(2.7) in order to determine the energy of the UWB Gaussian pulses. If E_1, E_2, and E_3 are the required energy of y_{g_1}, y_{g_2}, and y_{g_3}, respectively, it is possible to calculate the relation between them in a straightforward manner. The energy of $y_{g_1}(t)$ is defined and computed as

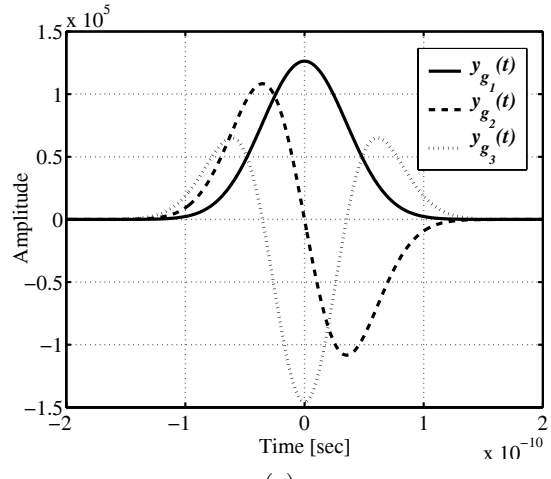
$$\begin{aligned} E_1 &= \int_{-\infty}^{\infty} y_{g_1}(t)^2 \, dt \\ &= \int_{-\infty}^{\infty} K_1^2 e^{-2(t/\tau)^2} \, dt \\ &= K_1^2 \tau \sqrt{\pi/2} \end{aligned} \tag{2.8}$$

Hence, the resulting formula for K_1 is as follows

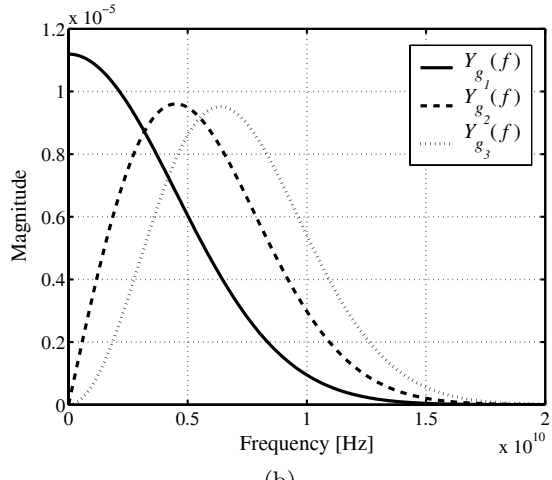
$$K_1 = \sqrt{\frac{E_1}{\tau\sqrt{\pi/2}}} \tag{2.9}$$

Employing an analogous procedure for K_2 and K_3 yields

$$K_2 = \sqrt{\frac{\tau E_2}{\sqrt{\pi/2}}} \tag{2.10}$$



(a)



(b)

Fig. 2.2 A Gaussian pulse, monocycle, and doublet in time and frequency domains. The Gaussian pulse has a large DC component.

and

$$K_3 = \tau \sqrt{\frac{\tau E_3}{3\sqrt{\pi/2}}} \tag{2.11}$$

Pulses of Figure 2.2(a) are all plotted for unit energy. The curves illustrated in Figure 2.2(b) are obtained by taking the Fourier transforms of Eqns. (2.5)–(2.7). For example, the frequency domain representation of $y_{g_1}(t)$ can be derived

as follows

$$\begin{aligned} Y_{g_1}(f) &= \mathcal{F}\{y_{g_1}(t)\} \\ &= \int_{-\infty}^{\infty} K_1 e^{-(t/\tau)^2} e^{-j2\pi ft}\, dt \\ &= K_1 \tau \sqrt{\pi}\, e^{-(\pi\tau f)^2} \end{aligned} \tag{2.12}$$

Since $y_{g_2}(t)$ and $y_{g_3}(t)$ are proportional to the first and second derivatives of $y_{g_1}(t)$, respectively, the Fourier transforms of $y_{g_2}(t)$ and $y_{g_3}(t)$ are the Fourier transforms of $y_{g_1}(t)$ multiplied by $(K_2/K_1)(j2\pi f)$ and $(K_3/K_1)(j2\pi f)^2$, respectively.

$$Y_{g_2}(f) = K_2 \tau \sqrt{\pi}(j2\pi f) e^{-(\pi\tau f)^2} \tag{2.13}$$

$$Y_{g_3}(f) = K_3 \tau \sqrt{\pi}(j2\pi f)^2 e^{-(\pi\tau f)^2} \tag{2.14}$$

The effective bandwidths of the Gaussian wave shapes of Figure 2.2 are now calculated. For the Gaussian pulse $y_{g_1}(t)$, which is actually a low-pass waveform, $f_l = 0$ and f_h is calculated by inserting $Y_{g_1}(f) = K_1 \tau \sqrt{\pi}/2$ into Eqn. (2.12). The result is $0.265/\tau$; hence

$$W_{g_1} = 5.3 \text{ GHz} \tag{2.15}$$

For the Gaussian monocycle defined in Eqn. (2.6) and shown in Figure 2.2 we can calculate $f_l = 1.44$ GHz and $f_h = 8.65$ GHz, which yields

$$W_{g_2} = 7.21 \text{ GHz} \tag{2.16}$$

For the Gaussian doublet defined in Eqn. (2.7) and shown in Figure 2.2 we can find $f_l = 3.07$ GHz and $f_h = 10.42$ GHz, which gives

$$W_{g_3} = 7.35 \text{ GHz} \tag{2.17}$$

The choice of which Gaussian waveform to use is usually driven by system design and application requirements. In any case the waveform should have a center frequency of about 3 to 10 GHz. All these three Gaussian waveforms have a dramatically different spectrum from a damped sine wave. One important characteristic of these waveforms is that they are almost uniformly distributed over their frequency spectrum and, therefore, are noise-like.

2.3 ORTHOGONAL WAVEFORMS AND HERMITE PULSES

The system of sine and cosine as orthogonal functions has been historically synonymous with communications. Whenever the term frequency is used, reference

is made implicitly to these functions, and, hence, the general theory of communication is based on the system of sine and cosine functions.

In recent years other complete systems of orthogonal functions have been used for theoretical investigations as well as for equipment design. Analogs to Fourier series, Fourier transform, frequency, power spectra, and amplitude, phase, and frequency modulation exist for many systems of orthogonal functions. This implies that theories of communication can be worked out on the basis of these systems. Most of these theories are of academic interest only. However, for the complete system of the orthogonal Hermite functions, the implementation of circuits by modern semiconductor techniques appears to be competitive in a number of applications with the implementation of circuits for the system of sine and cosine functions. In this section we present Hermite pulses as one example of orthogonal pulses that can be used for UWB communications.

Hermite functions and, indeed, Hermite pulses are not new [19]. The Hermite transform has already been used to shed light on spatiotemporal relationships in image processing [20]. The use of modified Hermite polynomials to create an orthogonal transform and even to create orthogonal wavelets for multicarrier data transmission over high-rate digital subscriber loops has been proposed [21]. The application of these pulses to communications seems limited to the latter reference, and, with the exception of the similarity to Hermite polynomials mentioned in [22], there seems to have been no attempt to systematically apply modified Hermite pulses to UWB communications before the proposal in [17]. Here, we present a systematic analysis of Hermite pulses and modify them for use in UWB communications systems.

2.3.1 Hermite polynomials

Charles Hermite was born in 1822 in France. He was a very brilliant mathematician. Some of his mathematical ideas are still widely used today, especially the Hermitian Forms that are used in physics and mathematics. Some of Hermite's other noted accomplishments that bear his name are Hermite polynomials, Hermite's differential equation, Hermite's formula of interpolation, and Hermitian matrices.

A polynomial is a finite sum of terms like $a_k x^k$, where k is a positive integer or zero. The functions defined by

$$h_{e_n}(t) = (-\tau)^n e^{t^2/2\tau^2} \frac{d^n}{dt^n}(e^{-t^2/2\tau^2}) \tag{2.18}$$

where $n = 0, 1, 2, \ldots$ and $-\infty < t < \infty$, are called *Hermite polynomials*. The parameter τ is the time-scaling factor. It should be mentioned that the definition of Eqn. (2.18) is one of various forms of Hermite polynomials used in the literature.

As examples of these polynomials we can write

$$
\begin{aligned}
h_{e_0}(t) &= 1 \\
h_{e_1}(t) &= \frac{t}{\tau} \\
h_{e_2}(t) &= \left(\frac{t}{\tau}\right)^2 - 1 \\
h_{e_3}(t) &= \left(\frac{t}{\tau}\right)^3 - 3\frac{t}{\tau} \\
h_{e_4}(t) &= \left(\frac{t}{\tau}\right)^4 - 6\left(\frac{t}{\tau}\right)^2 + 3 \\
h_{e_5}(t) &= \left(\frac{t}{\tau}\right)^5 - 10\left(\frac{t}{\tau}\right)^3 + 15\frac{t}{\tau} \\
h_{e_6}(t) &= \left(\frac{t}{\tau}\right)^6 - 15\left(\frac{t}{\tau}\right)^4 + 45\left(\frac{t}{\tau}\right)^2 - 15 \\
h_{e_7}(t) &= \left(\frac{t}{\tau}\right)^7 - 21\left(\frac{t}{\tau}\right)^5 + 105\left(\frac{t}{\tau}\right)^3 - 105\frac{t}{\tau} \\
h_{e_8}(t) &= \left(\frac{t}{\tau}\right)^8 - 28\left(\frac{t}{\tau}\right)^6 + 210\left(\frac{t}{\tau}\right)^4 - 420\left(\frac{t}{\tau}\right)^2 + 105
\end{aligned}
\tag{2.19}
$$

which are related by the following equations

$$h_{e_{n+1}}(t) = \frac{t}{\tau}h_{e_n}(t) - \tau\dot{h}_{e_n}(t) \tag{2.20}$$

$$\dot{h}_{e_n}(t) = \frac{n}{\tau}h_{e_{n-1}}(t) \tag{2.21}$$

where "˙" stands for derivative with respect to time.

Using Eqns. (2.20) and (2.21) the differential equation which is satisfied by Hermite polynomials is derived as Eqn. (2.22)

$$\tau^2\ddot{h}_{e_n}(t) - t\dot{h}_{e_n}(t) + nh_{e_n}(t) = 0 \tag{2.22}$$

2.3.2 Orthogonal modified Hermite pulses

By definition, two real-valued functions $g_m(t)$ and $g_n(t)$ that are defined on an interval $a \leq x \leq b$ are orthogonal if

$$(g_m \cdot g_n) \doteq \int_a^b g_m(t)g_n(t)\,dt = 0; \quad m \neq n \tag{2.23}$$

A set of real valued functions $g_1(t), g_2(t), g_3(t), \ldots$ is called an orthogonal set of functions in the set. The nonnegative square root of $(g_m \cdot g_m)$ is called the norm of $g_m(t)$ and is denoted by $\|g_m\|$; thus

$$\|g_m\| = \sqrt{(g_m \cdot g_m)} = \sqrt{\int_a^b g_m^2(t)\, dt} \tag{2.24}$$

Orthonormal sets of functions satisfy $\|g_m\| = 1$ for every value of m.

Hermite polynomials are not orthogonal; however, they can be modified to become orthogonal as follows [23]

$$\begin{aligned} h_n(t) &= k_n e^{-t^2/4\tau^2} h_{e_n}(t) \\ &= (-\tau)^n e^{t^2/4\tau^2} \tau^2 \frac{d^n}{dt^n}(e^{-t^2/2\tau^2}) \end{aligned} \tag{2.25}$$

where $n = 0, 1, 2, \ldots$ and $-\infty < t < \infty$. The result is a set of orthogonal functions $h_n(t)$ which can be easily derived for all values of n. As examples we can write

$$\begin{aligned} h_0(t) &= k_0 e^{-t^2/4\tau^2} \\ h_1(t) &= k_1 \frac{t}{\tau} e^{-t^2/4\tau^2} \\ h_2(t) &= k_2 \left[\left(\frac{t}{\tau}\right)^2 - 1 \right] e^{-t^2/4\tau^2} \\ h_3(t) &= k_3 \left[\left(\frac{t}{\tau}\right)^3 - 3\frac{t}{\tau} \right] e^{-t^2/4\tau^2} \\ h_4(t) &= k_4 \left[\left(\frac{t}{\tau}\right)^4 - 6\left(\frac{t}{\tau}\right)^2 + 3 \right] e^{-t^2/4\tau^2} \\ h_5(t) &= k_5 \left[\left(\frac{t}{\tau}\right)^5 - 10\left(\frac{t}{\tau}\right)^3 + 15\frac{t}{\tau} \right] e^{-t^2/4\tau^2} \\ h_6(t) &= k_6 \left[\left(\frac{t}{\tau}\right)^6 - 15\left(\frac{t}{\tau}\right)^4 + 45\left(\frac{t}{\tau}\right)^2 - 15 \right] e^{-t^2/4\tau^2} \\ h_7(t) &= k_7 \left[\left(\frac{t}{\tau}\right)^7 - 21\left(\frac{t}{\tau}\right)^5 + 105\left(\frac{t}{\tau}\right)^3 - 105\frac{t}{\tau} \right] e^{-t^2/4\tau^2} \\ h_8(t) &= k_8 \left[\left(\frac{t}{\tau}\right)^8 - 28\left(\frac{t}{\tau}\right)^6 + 210\left(\frac{t}{\tau}\right)^4 - 420\left(\frac{t}{\tau}\right)^2 + 105 \right] e^{-t^2/4\tau^2} \end{aligned} \tag{2.26}$$

The constants k_n, $n = 0, 1, \ldots$ determine the energy of the pulses. We call the functions derived in Eqn. (2.26) modified Hermite pulses (MHP), and it can be

shown that their general formula is the following

$$h_n(t) = k_n e^{-t^2/4\tau^2} n! \sum_{i=0}^{[n/2]} \left(-\frac{1}{2}\right)^i \frac{(t/\tau)^{n-2i}}{(n-2i)!\; i!} \tag{2.27}$$

where $[n/2]$ denotes the integer part of $n/2$. If it is desired that each $h_n(t)$ pulse has an energy of E_n, then we can show that

$$k_n = \sqrt{\frac{E_n}{\tau n!\sqrt{2\pi}}} \tag{2.28}$$

Figure 2.3(a) shows h_n for four values of n from 0 to 3. The scaling parameter is $\tau = 5 \times 10^{-11}$ sec, and the pulses have unit energy. The time duration of all five pulses is about 0.6 ns. Using Eqns. (2.20)–(2.22) and (2.25) it can be shown that, disregarding the constants k_n, all MHPs satisfy the following differential equations

$$\tau^2 \ddot{h}_n(t) + \left(n + \frac{1}{2} - \frac{1}{4}\frac{t^2}{\tau^2}\right) h_n(t) = 0 \tag{2.29}$$

$$\tau \dot{h}_n(t) + \frac{t}{2\tau} h_n(t) = n h_{n-1}(t) \tag{2.30}$$

$$h_{n+1}(t) = \frac{t}{2\tau} h_n(t) - \tau \dot{h}_n(t) \tag{2.31}$$

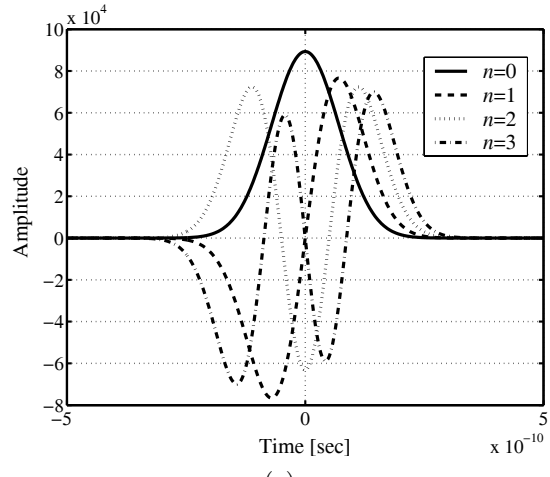
Denoting the Fourier transform of $h_n(t)$ as $H_n(f)$, Eqns. (2.29)–(2.31) can be written as

$$\ddot{H}_n(f) + 16\pi^2\tau^2 \left(n + \frac{1}{2} - 4\pi^2 f^2 \tau^2\right) H_n(f) = 0 \tag{2.32}$$

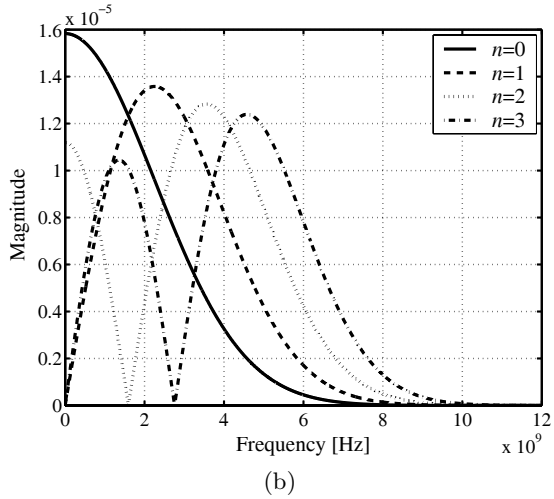
$$j8\pi^2 f \tau^2 H_n(f) + j\dot{H}_n(f) = 4\pi n \tau H_{n-1}(f) \tag{2.33}$$

$$H_{n+1}(f) = j\frac{1}{4\pi\tau}\dot{H}_n(f) - j2\pi f \tau H_n(f) \tag{2.34}$$

respectively. Note that the derivatives are with respect to the frequency.



(a)



(b)

Fig. 2.3 Time and frequency responses of the normalized MHP of orders $n = 0, 1, 2, 3$ normalized to unit energy.

Example 2.3

Using Eqn. (2.34) find the frequency domain formulas for $h_1(t)$, $h_2(t)$, and $h_3(t)$.

Solution

For the value of $n = 0$, we will have $h_0(t) = k_0 e^{-t^2/4\tau^2}$ and

$$H_0(f) = k_0 \tau 2\sqrt{\pi}\, e^{-4\pi^2 f^2 \tau^2}$$

and from Eqn. (2.34) the transform of some higher degrees of MHPs can be obtained as follows (now we include the constants k_n)

$$\begin{aligned} H_1(f) &= k_1\tau(-j4\pi f\tau)2\sqrt{\pi}\, e^{-4\pi^2\tau^2 f^2} \\ H_2(f) &= k_2\tau(1-16\pi^2 f^2\tau^2)2\sqrt{\pi}\, e^{-4\pi^2\tau^2 f^2} \\ H_3(f) &= k_3\tau(-j12\pi f\tau + j64\pi^3 f^3\tau^3)2\sqrt{\pi}\, e^{-4\pi^2\tau^2 f^2} \end{aligned} \tag{2.35}$$

Figure 2.3(b) shows the frequency domain representations of the MHPs for $n = 0, 1, 2, 3$. According to Figures 2.3(a) and (b) all MHP pulses have the following properties:

- The pulse duration is almost the same for all values of n.
- The pulse bandwidth is almost the same for every value of n.
- The pulses are mutually orthogonal.
- The pulses have nonzero DC components, and, in fact, the low-frequency components of the pulses are relatively significant.
- The number of zero crossings is equal to n.

The autocorrelation functions of the modified Hermite pulses for orders $n = 0, 1, 2, 3$ are shown in Figure 2.4. As can be seen the width of the main peak in the autocorrelation function becomes narrower as the order of the pulses increases. This suggests that higher order pulses will be more sensitive to pulse jitter.

2.3.3 Modulated and modified Hermite pulses

For gaining more flexibility in the frequency domain, the time functions can be multiplied and modified by an arbitrary phase-shifted sinusoid. Hence, the multiplied and modified normalized Hermite pulses are defined as follows

$$p_n(t) = \sqrt{2}\, h_n(t)\cos(2\pi f_c t + \phi_r) \tag{2.36}$$

where $n = 0, 1, 2, \ldots$, $-\infty < t < \infty$, f_c is the shifting frequency, and ϕ_r is an arbitrary phase that can be zero without loss of generality. The function $p_n(t)$ for $n = 0, 1, 2, 3$ in time and frequency domains is illustrated in Figures 2.5(a) and (b), respectively.

Comparing Figures 2.3 and 2.5 shows that $p_n(t)$ pulses have more oscillations than MHP pulses, but the pulse width has not been changed. The number of zero crossings in MHP is equal to n; however, in $p_n(t)$ it is a function of both n and f_c.

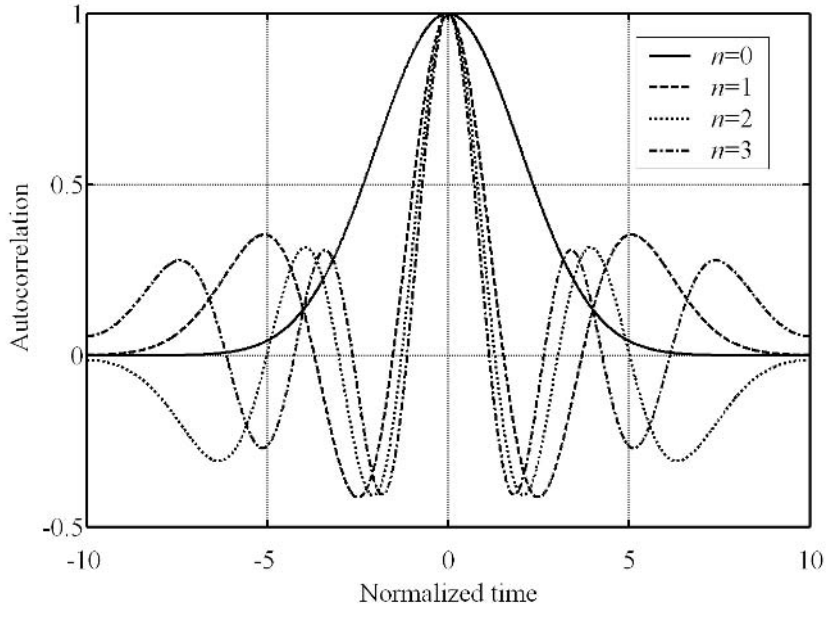


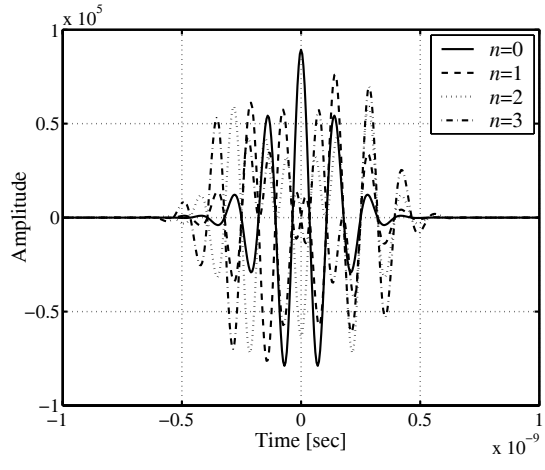
Fig. 2.4 Autocorrelation functions of modified normalized Hermite pulses of orders $n = 0, 1, 2, 3$. The width of the main peak in the autocorrelation function becomes narrower as the order of the pulse increases.

We also observe that the fractional bandwidth of MHP is higher, and increasing f_c decreases the fractional bandwidth accordingly.

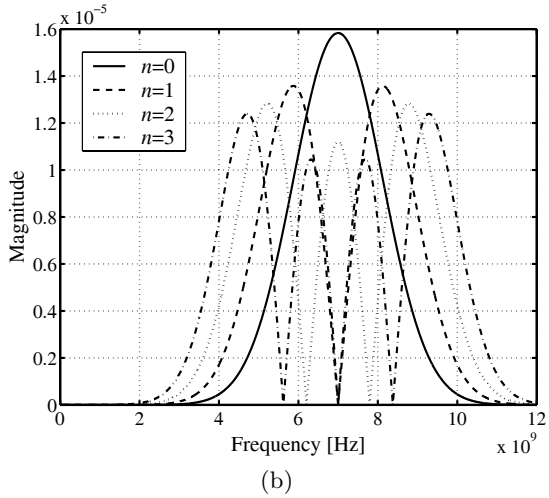
Equations (2.29) and (2.36) give the schematic diagram of a linear but time-variant system which creates pulses of different degrees. Figure 2.6 shows a simple schematic diagram for this system. The interesting feature of this structure is that, theoretically, by changing a simple gain of the circuit we can create different orders of the desired waveform.

According to Figures 2.5(a) and (b), all $p_n(t)$ functions have the following properties:

- The pulse duration is almost the same for all values of n.
- The pulse bandwidth is almost the same for every value of n.
- The fractional bandwidth can be easily controlled by f_c.
- The pulses are mutually orthogonal.
- The pulses have zero DC component.



(a)



(b)

Fig. 2.5 Time and frequency representations of $p_n(t)$ for orders $n = 0, 1, 2, 3$. All pulses have zero low-frequency components. Compared with Figure 2.3(a) the number of zero crossings has been increased. It can also be seen that the fractional bandwidth of the signals has reduced from 200% to about 100% and can be further reduced by increasing f_c.

2.4 ORTHOGONAL PROLATE SPHEROIDAL WAVE FUNCTIONS

In this section a multiple pulse generator which generates four different prolate spheroidal wave functions (PSWF) based on a source signal for use in UWB

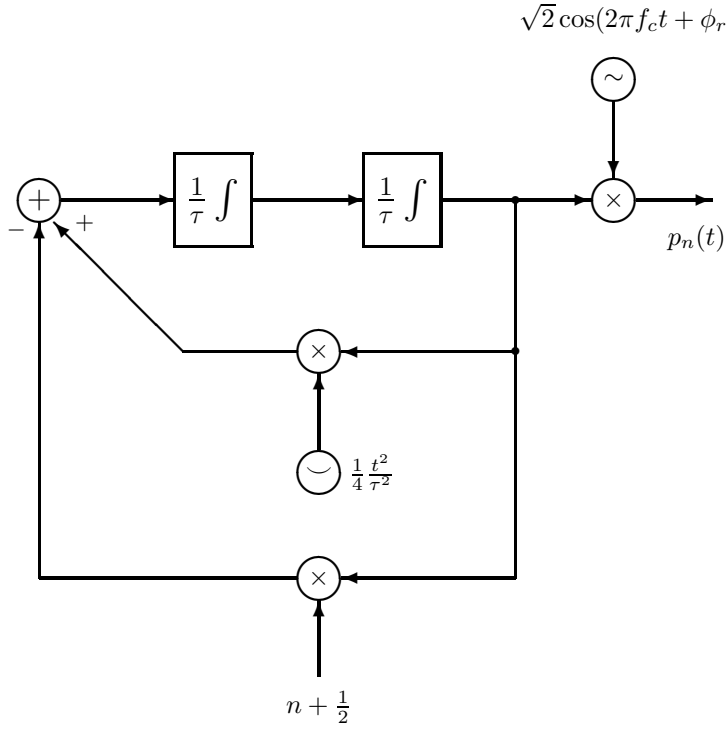


Fig. 2.6 The analog linear time-variant circuit producing different $p_n(t)$ functions.

communication systems is proposed. These sets of pulses can be applied to an M-ary communication system. This class of pulse shape yields orthogonal pulses that have a constant pulse width and bandwidth regardless of the pulse order, which is in contrast to the majority of other orthogonal pulse classes. The results encourage the implementation of a cheap and easily reproducible UWB pulse generator.

2.4.1 Introduction

All conventional methods for enabling distinction between different pulse trains require that the multi-user base station send data serially, one pulse after another. Otherwise, interference between pulses will make it impossible for proper reception at the remote receivers. Therefore, encoding schemes become more complicated as the number of remote receivers increases; hence, the data rate per user decreases.

It is conceivable for a wireless impulse transmitter to communicate with multiple receivers using mutually orthogonal pulses, wherein a different pair of orthogonal

pulses is assigned to each receiver to represent a digital signal [24]. Because all of the pulses are orthogonal, they can be transmitted simultaneously without interfering with each other. Moreover, since the transmitter and the receiver know in advance which pulses are for which receivers, each receiver will only receive signals including pulse shapes assigned to it. This would eliminate the need to modulate the pulse trains for identification purposes. However, with such a conceivable configuration, the transmitter would need to be provided with a separate pulse generator for each different pulse shape. The amount of hardware required for producing pulses would increase linearly with the number of receivers.

Considering the above mentioned drawbacks, several objectives can be considered for a UWB transmitter:

- To provide a multi-pulse generator capable of producing two or more orthogonal pulses without much more hardware than required to produce a single pulse.

- To provide a wireless impulse transmitter with a simple configuration capable of high communication rates without the limitations caused by circuitry in PPM and the distance attenuation problem of PAM.

- To provide a wireless impulse system with a simple configuration that enables increasing the number of users without reducing the communication rates.

In order to achieve the above described objectives a multi-pulse generator for generating four different PWSF pulses, based on a source signal, is introduced. Conventionally, four complete sets of hardware are needed in order to produce four different pulses of different shapes. However, it will be shown how different pulses of different shapes and a desirable number of the pulse orders can be produced with a single-source signal, simply by changing a constant value. A wireless impulse transmitter is presented accordingly for transmitting pulse trains to a plurality of receivers which include a multi-pulse generator, a pulse selector, and a transmission unit. The pulse selector selects pulses from different pulse shape generators according to input data. The transmission unit transmits the pulses selected by the pulse selector.

2.4.2 Fundamentals of PSWF

A function $h(t)$ is said to be time-limited if for $T > 0$, $h(t)$ vanishes for all $|t| > T/2$. In an analogous manner $H(f)$ is said to be band-limited with a bandwidth of Ω if it is zero outside of the band $(-\Omega, \Omega)$. Functions that are practically time and band-limited can be very useful in communication engineering. One of these kinds of functions is referred to as a prolate spheroidal wave function. This function is

the solution of [25]

$$\int_{-T/2}^{T/2} \psi_n(x) \frac{\sin \Omega(t-x)}{\pi(t-x)}\, dx = \lambda_n \psi_n(t) \tag{2.37}$$

or, alternatively, the solution of the differential equation

$$\frac{d}{dt}(1-t^2)\frac{d\psi_n(t)}{dt} + (\chi_n - c^2 t^2)\psi_n(t) = 0 \tag{2.38}$$

where $\psi_n(t)$ are prolate spheroidal wave functions of order n and χ_n is the eigenvalue of $\psi_n(t)$. The constant c is

$$c = \frac{\Omega T}{2} \tag{2.39}$$

where Ω is the bandwidth and T is the pulse duration.

In Eqn. (2.37), λ_n is the concentration of energy in the interval $[-T/2, T/2]$ given by

$$\lambda_n = \frac{\int_{-T/2}^{T/2} |\psi_n(t)|^2\, dt}{\int_{-\infty}^{\infty} |\psi_n(t)|^2\, dt} \tag{2.40}$$

whose values range from 0 to 1. Moreover, it can be shown that PSWFs have the following properties [25]

$$\int_{-\infty}^{\infty} \psi_m(t)\psi_n(t)\, dt = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \tag{2.41}$$

$$\int_{-T/2}^{T/2} \psi_m(t)\psi_n(t)\, dt = \begin{cases} \lambda_n, & m = n \\ 0, & m \neq n \end{cases} \tag{2.42}$$

From Eqns. (2.41) and (2.42) a double orthogonality of $\psi_n(t)$ can be observed. This is a useful property for the UWB communication system because it guarantees unique demodulation at the receiver. This is the reason orthogonal pulse sets have been proposed for use in UWB applications. These orthogonal pulse sets consist of coded monocycles or coded baseband waveforms.

If we solve the differential equation (2.38) for the highest derivative, we get

$$(1-t^2)\frac{d^2\psi_n(t)}{dt^2} - 2t\frac{d\psi_n(t)}{dt} + (\chi_n - c^2 t^2)\psi_n(t) = 0 \tag{2.43}$$

and consequently

$$\ddot{\psi}_n(t) = \frac{1}{1-t^2}[2t\dot{\psi}_n(t) - (\chi_n - c^2 t^2)\psi_n(t)] \tag{2.44}$$

As can be seen, different orders of the pulses can be simply obtained by changing the values of χ_n; hence, Eqn. (2.44) is the basis of the multi-pulse generator.

Different values of χ_n have been calculated and are shown in Table 2.1. (Details of the calculations can be found in [26].)

Table 2.1 Eigenvalue of different pulse orders for $c = 2$.

Pulse order, n	χ_n
0	1.12773
1	4.28712
2	8.22571
3	14.10020
4	22.05482
5	32.03526
6	44.02474
7	58.01837
8	74.01419
9	92.01130

Here are some properties of PSWF wave shapes:

- The pulse duration is exactly the same for all values of n.
- The pulse bandwidth is almost the same for all values of n.
- The pulses are double-orthogonal.
- The pulses have a nonzero DC component.
- Pulse duration and bandwidth can be controlled simultaneously.

If $\psi_n(t)$ is written in terms of the prolate angular function of the first kind

$$\begin{aligned}\psi_n(t) &= \psi_n(\Omega, T, t) \\ &= \frac{[2\lambda_n(c)/T]^{1/2} S_{0n}^1(c, 2t/T)}{\left\{\int_{-1}^{1} [S_{0n}^1(c, x)]^2 \, dx\right\}^{1/2}}\end{aligned} \tag{2.45}$$

and

$$\left\{\int_{-1}^{1} [S_{0n}^1(c, x)]^2 \, dx\right\}^{1/2} = \frac{2}{2n+1} \tag{2.46}$$

where S_{0n}^1 is the prolate angular function of the first kind and λ_n is the fraction of the energy of $\psi_n(t)$ that lies in the interval $[-1, 1]$.

The prolate angular function of the first kind is given by [27]

$$S_{0n}^{1}(c,t) = \begin{cases} \sum_{k=0,2,\ldots}^{\infty} d_k(c) P_k(c,t), & \text{for } n \text{ even} \\ \sum_{k=1,3,\ldots}^{\infty} d_k(c) P_k(c,t), & \text{for } n \text{ odd} \end{cases} \tag{2.47}$$

where $P_k(c,t)$ is the Legendre polynomials [28] and $d_k(c)$ are the series coefficients. Differentiation of $S_{0n}^{1}(c,t)$ is given by

$$\dot{S}_{0n}^{1}(c,t) = \sum_{k} d_k(c) \dot{P}_k(c,t) \tag{2.48}$$

If the associated Legendre functions are given in terms of unassociated Legendre functions, then [29]

$$P_k^m(t) = (-1)^m (1-t^2)^{1/2} \frac{d^m}{dt^m} P_k(t) \tag{2.49}$$

and for the first derivative

$$\frac{dP_k(t)}{dt} = \frac{-P_k^1(t)}{\sqrt{1-t^2}} \tag{2.50}$$

using Eqns. (2.48)–(2.50) yields

$$\dot{S}_{0n}^{1}(c,t) = \frac{-1}{\sqrt{1-t^2}} \sum_{k} d_k(c) P_k^1(t) \tag{2.51}$$

and

$$P_k^1(t) = \frac{-1}{2^k k!} (1-t^2)^{1/2} \frac{d^{k+1}}{dt^{k+1}} (t^2-1)^k \tag{2.52}$$

The second differentiation of $S_{0n}^{1}(c,t)$ is given by:

$$\ddot{S}_{0n}^{1}(c,t) = \sum_{k} d_k(c) \ddot{P}_k(c,t) \tag{2.53}$$

Using Eqn. (2.49) yields

$$\frac{d^2 P_k(t)}{dt^2} = \frac{P_k^2(t)}{1-t^2} \tag{2.54}$$

Using Eqns. (2.53) and (2.54) the following results:

$$\ddot{S}_{0n}^{1}(c,t) = \frac{1}{1-t^2} \sum_{k} d_k(c) P_k^2(t) \tag{2.55}$$

and

$$P_k^2(t) = \frac{1}{2^k k!} (1-t^2) \frac{d^{k+2}}{dt^{k+2}} (t^2-1)^k \tag{2.56}$$

2.4.3 PSWF pulse generator

A UWB communication system employing a PSWF pulse generators, a wireless UWB impulse transmitter and two remote receivers can be constructed as shown in Figure 2.7. Transmitter is a multi-user base station and the receivers 1 and 2 are mobile terminals.

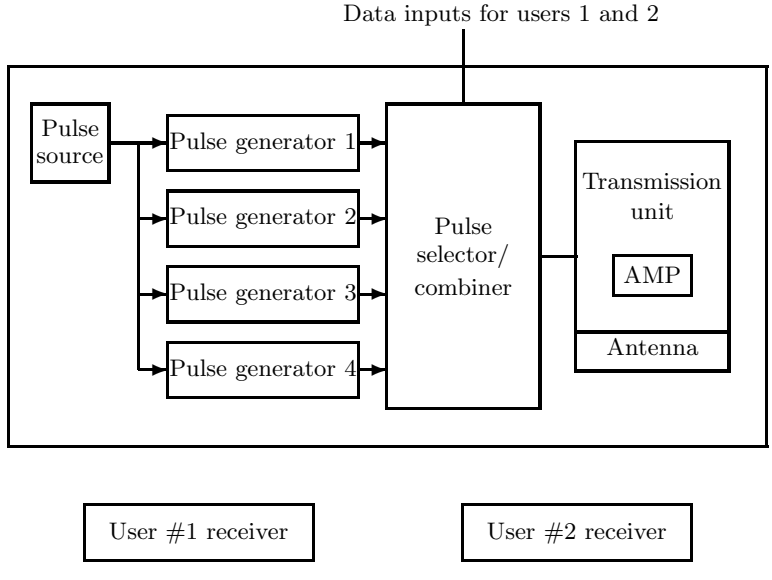


Fig. 2.7 Schematic diagram of a UWB communication system employing a PSWF pulse generator.

Transmitter includes a pulse source, four pulse generators, a pulse train selector/combiner, and a transmission unit. Pulse generators are individually connected to the pulse selector/combiner. The four pulse generators generate PSWF pulses of orders 2 to 5, respectively, and are in synchronization based on the inputs from pulse source 2. The pulse selector/combiner selects pulses from pulse generators based on a two-to-one correspondence between the different-order pulses and receivers 1 and 2: that is, the pulse selector/combiner selects pulses from pulse generators 1 and 2 based on input data to be sent to the receiver 1 and selects pulses from the pulse generators 3 and 4 based on input data to be sent to receiver 2. This is because the four pulse shapes generated by pulse generators are represented by symbol numbers 1 to 4 and each of receivers 1 and 2 is assigned two of the four pulse shapes (symbol numbers) to represent a binary channel: that is, 0 or 1. In this example, receiver 1 is assigned symbol numbers 1 and 2 and

receiver 2 is assigned symbol numbers 3 and 4. This correspondence relationship is summarized in Table 2.2.

Table 2.2 The relationship between receivers and assigned symbols.

Receiver	Binary	Symbol
1	0	1
1	1	2
2	0	3
2	1	4

The pulse selector/combiner also combines the pulses for receivers 1 and 2 and sends them to the transmission unit. The transmission unit amplifies the pulses using the amplifier and transmits them over the antenna. The transmission unit transmits pulses corresponding to different receivers simultaneously.

The configuration of pulse generators is illustrated in Figure 2.8. It should be noted that pulse generators are designed based on Eqn. (2.44). The results of PSWF generation are illustrated in Figure 2.9 for orders four and five.

Each of the first to fourth pulse generators has the same configuration for generating PSWF orthogonal pulses. In this way, four different order pulses can be generated with only slightly more hardware than required to produce one PSWF orthogonal pulse.

Each of receivers 1 and 2 has substantially the same configuration in order to demodulate incoming signals, with the exception of the two pulse shapes assigned to each. Here, an explanation will be provided for receiver 1 as a representative example. As shown in Figure 2.10, receiver 1 includes a reception unit W, a timing control circuit, a basic pulse supplier, a correlator, an orthogonal-to-digital data selector, and an output unit. The reception unit W includes an antenna, a filter, and an amplifier.

After receiving a pulse by the antenna the received signal is filtered and amplified. The basic pulse supplier supplies two different-order pulse waveforms and, in the present embodiment, the second and third-order pulse waveforms.

The correlator measures the similarity between each incoming pulse and the plurality of orthogonal pulse shapes from the basic pulse supplier to identify the corresponding symbol. Because orthogonal pulses are used the cross correlation between different order pulses is zero. Therefore, correlator can correctly distinguish among different-order pulses. The correlator performs the correlation process at a timing modified by the timing control circuit to allow for differences in time of flight between the transmitter and receiver 1; for example, when the transmitter and receiver 1, or both, are moved. It also allows inclusion of PPM and pseudo-noise (PN) code timing changes. In the example shown in Figure 2.10,

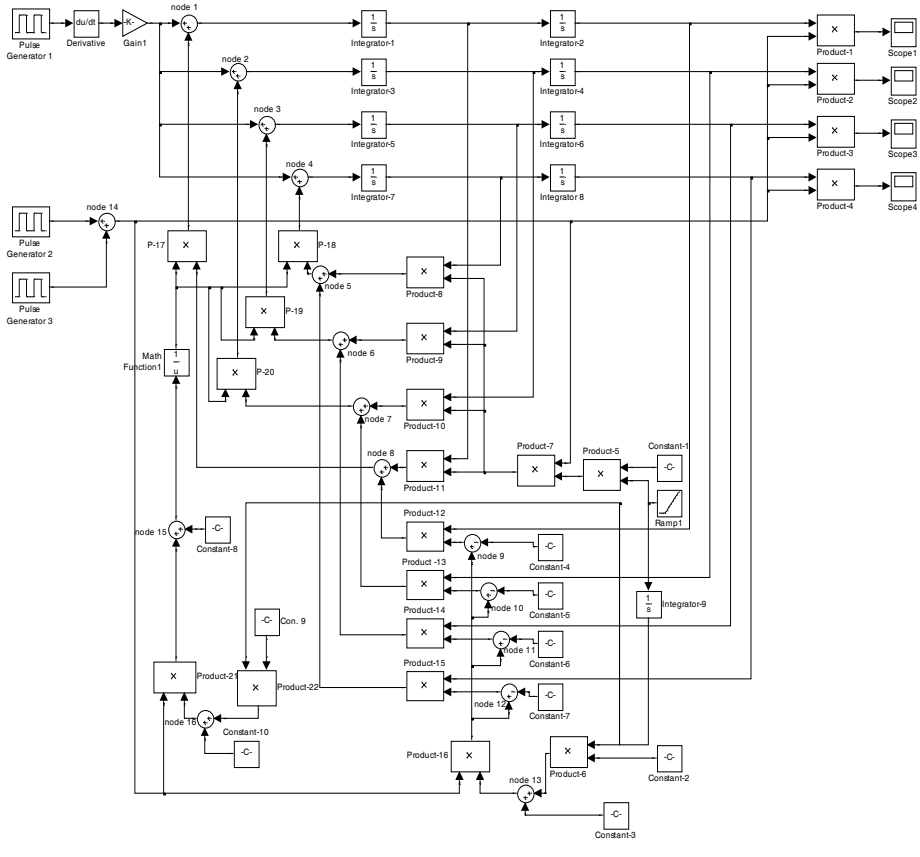
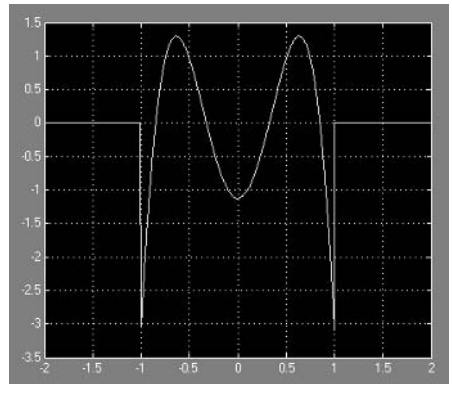


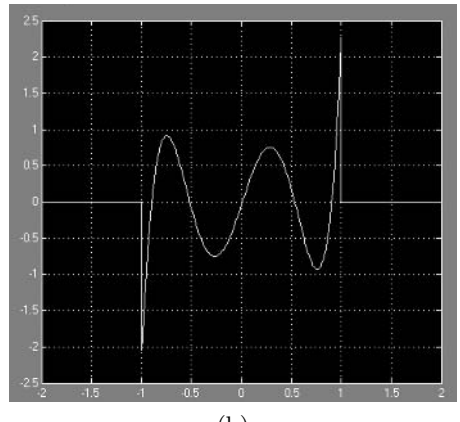
Fig. 2.8 Schematic diagram of four different PSWF pulse generators.

the correlator determines that four successive incoming pulses correspond to the symbol numbers 1, 2, 2, 1.

The orthogonal to digital data selector converts symbol numbers from the correlator into corresponding binary data. In Figure 2.10, orthogonal to digital data selector converts the symbol numbers 1, 2, 2, 1 from correlator into corresponding binary data of 0, 1, 1, 0. Receivers 1 and 2 have the ability to demodulate all the different orthogonal pulse shapes assigned to the system. This enables implementation of an M-ary modulation scheme when only one or other of receivers 1 and 2 are being used, to enhance the transmission rate of the system. For example, when only receiver 1 is used the transmitter assigns all of the different-order pulse waveforms to receiver 1 and assigns each pulse shape correspondence



(a)



(b)

Fig. 2.9 Output of the (a) fourth and (b) fifth-order PSWF generators.

with multi-bit symbols as indicated in Table 2.3 to create a 4-ary modulation scheme.

Although we have described only four pulse generators, six or more different-order pulse generators can be provided as well. In such a case, each pair of pulse generators could be assigned to a different receiver in order to increase the number of receivers that can be used within the system. Alternatively or in addition, each pair of generators could be assigned a different multi-bit symbol to increase the transmission rate of the system as described above. The transmission rate of the system could be enhanced by further use of conventional modulation schemes, such as PPM or PAM.

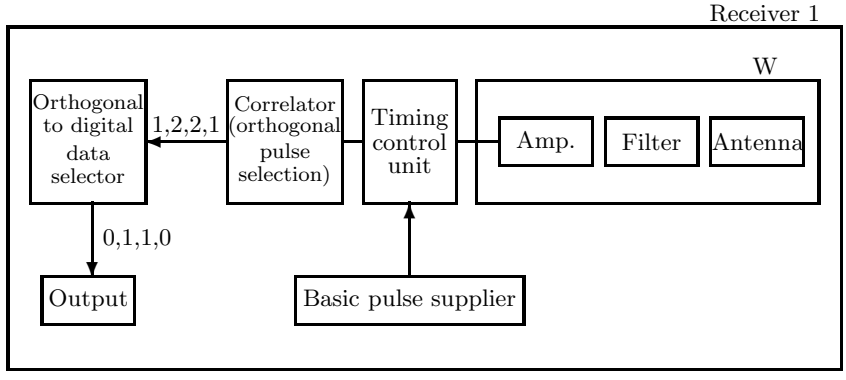


Fig. 2.10 Schematic diagram of the PSWF UWB receiver.

Table 2.3 Creation of a 4-ary modulation scheme.

Receiver	Binary	Symbol
1	00	1
1	01	2
1	10	3
1	11	4

2.5 DESIGNING WAVEFORMS FOR SPECIFIC SPECTRAL MASKS

It has already been explained in Chapter 1 that UWB signals are like noise. Transmitted energy is spread over a very wide bandwidth and the energy or power density of signals is very low. In this section we investigate another interesting tool for designing different pulse shapes. The pulses introduced in this section have deep spectral nulls which are at the same frequency as conventional narrowband communication systems. The result of these nulls is the extensive reduction of interference caused by our UWB system on nearby receivers.

2.5.1 Introduction

Single short pulse (or impulse) generation is the traditional and fundamental approach for generating UWB waveforms. By varying the pulse characteristics the characteristics of the energy in the frequency spectrum may be defined based on the desired design criteria. Generally, there are three parameters of interest when defining the properties of energy which is filling a specified frequency spectrum:

- The intended bandwidth of the transmitted energy should be carefully defined and considered.
- We have to limit the available energy to within the specified UWB frequency spectrum.
- Defining where generated energy should center within the spectrum of interest (that is, the *center frequency*).

Pulse duration in the time domain determines the bandwidth in the frequency domain. As a rule of thumb we may write

$$\frac{1}{\text{Duration}} \approx \text{Bandwidth} \tag{2.57}$$

Pulse repetition is a characteristic that may determine the center frequency of a band of transmitted energy if the repetition is regular.

Pulse shape determines the characteristics of how the energy occupies the frequency domain.

Since the highest spectral efficiency is one of the most important objectives for UWB communication, as much bandwidth as is practical should be used to take advantage of the capacity made available when bandwidth is very large.

As with traditional radio architectures, data can be modulated on the pulses in a number of ways including amplitude, phase, time position, or any combination of these. Single-pulse architectures offer relatively simple radio designs, but provide little flexibility where spectrum management is an objective. Examples of scenarios where managing the spectrum might be desirable are:

- Matching different regulatory requirements in different international regions.
- Dynamically sensing interfering technologies and suspending usage of contending frequencies.
- Choosing to use narrower bands of spectrum to either share spectrum in a local area or to enable lower cost devices that do not require large bandwidth for a specific application.

Another area where managing the spectrum might be desirable is in performance management. Increasing the performance of existing designs may require completely redesigning a radio for higher performance implementations, forgoing backward compatibility with earlier implementations [30].

2.5.2 Multi-band modulation

Multi-band modulation [31] is another approach to modulating information with wideband technology. Multi-band UWB modulation provides a method where, for

example, the 7.5 GHz of spectrum made available to UWB technology by the FCC (see Chapter 1) may be split into multiple smaller frequency bands. This could be accomplished by choosing to implement a few large bands or many small bands, all of which would be stacked across the legally available frequency spectrum.

As stated previously, ultra wideband technology uses pulses to modulate information over a wide band of frequencies. Pulse shape is the primary characteristic that determines the distribution of energy within the frequency domain, and properly shaping the pulse will concentrate more of the energy in the center lobe of the energy band, reducing side lobe energy and reducing chances for adjacent band interference. To effectively fill the specified spectrum, multiple frequency bands of energy must be generated with different center frequencies and be spaced across the spectrum. Center frequency selection is accomplished when using a pseudo-carrier oscillation in generating and shaping the required UWB pulse. The frequency of the pseudo-carrier oscillation determines the center frequency of the band, while the outline of the oscillation defines the pulse shape and, thus, defines the bandwidth.

The idea of multi-band modulation is to use multiple frequency bands to efficiently utilize the UWB spectrum by transmitting multiple UWB signals at the same time. The signals do not interfere with each other because they operate at different frequencies. Assume that the Gaussian pulse of Figure 2.2(a) is modulated by a sinusoid signal to form a transmitted pulse as follows

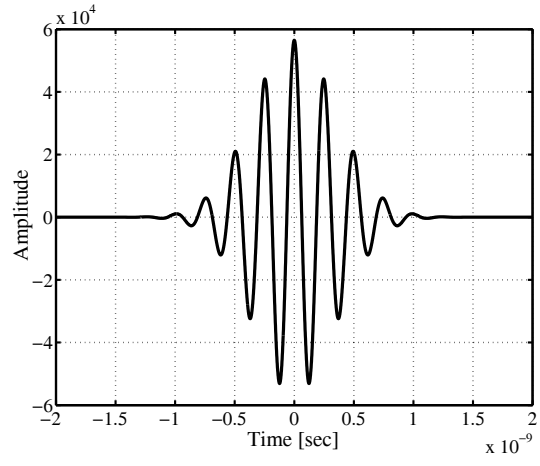
$$\begin{aligned} s_j(t) &= \sqrt{2}\, y_{g_1}(t) \cos(2\pi f_j t) \\ &= \sqrt{2}\, K_1 e^{-(t/\tau)^2} \cos(2\pi f_j t) \end{aligned} \tag{2.58}$$

where f_j is the modulating frequency, not much higher than the bandwidth of the Gaussian pulse. The waveform defined by Eqn. (2.58) is shown in Figure 2.11(a) for $f_j = 4$ GHz, $\tau = 0.5$ ns, and unit energy. The Fourier transform of Eqn. (2.58) can be calculated as follows

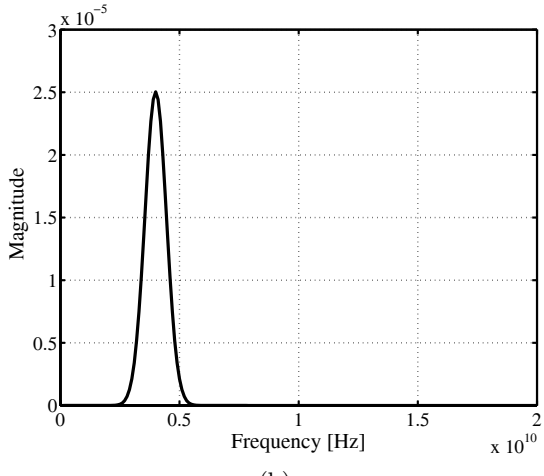
$$\begin{aligned} S_j(f) &= \frac{\sqrt{2}}{2}[Y_{g_1}(f - f_j) + Y_{g_1}(f + f_j)] \\ &= \frac{\sqrt{2\pi}}{2} K_1 \tau [e^{-(\pi\tau(f-f_j))^2} + e^{-(\pi\tau(f+f_j))^2}] \end{aligned} \tag{2.59}$$

The spectrum of the modulated Gaussian pulse for positive frequencies is illustrated in Figure 2.11(b) and extends around 4 GHz. Based on the discussion in Section 2.1 this kind of wave shape is of the damped sinusoid type and normally is not to be used for UWB communications. However, if a combination of several pulses like $s_j(t)$ contribute to the construction of the transmitted signal the result can be quite satisfactory.

As an example of multi-band modulation consider the case where five different frequencies, $f_1 = 4$, $f_2 = 5$, $f_3 = 6$, $f_4 = 7$, and $f_5 = 8$ GHz, are utilized for modulating Gaussian pulses and each of these signals is transmitted simultaneously to achieve a very high bit rate or used as a means of multiple access



(a)



(b)

Fig. 2.11 A modulated Gaussian pulse and its frequency domain presentation. The centre frequency is 4 GHz.

to allow multiple users to communicate at the same time. In addition, several standard digital modulation techniques can also be used on each individual UWB signal. The output of the modulated UWB signals can be added together before transmission. At the receiver the modulated UWB signals must be separated before demodulation. A combination of five modulated Gaussian pulses can be

simply done as follows

$$s_{mb}(t) = \sum_{j=1}^{5} s_j(t)$$
$$= \sqrt{2}\, K_1 e^{-(t/\tau)^2} \sum_{j=1}^{5} \cos(2\pi f_j t) \tag{2.60}$$

A plot of Eqn. (2.60) is shown in Figure 2.12(a) for $\tau = 0.5$ ns, $E_1 = 1$, and $K_1 = \sqrt{E_1/(\tau\sqrt{\pi/2})} = 39,947$. By using Eqns. (2.59) and (2.60) the spectrum of $s_{mb}(t)$ is easily calculated as follows

$$S_{mb}(f) = \sum_{j=1}^{5} S_j(f)$$
$$= \frac{\sqrt{2\pi}}{2} K_1 \tau \sum_{j=1}^{5} [e^{-(\pi\tau(f-f_j))^2} + e^{-(\pi\tau(f+f_j))^2}] \tag{2.61}$$

The result is shown in Figure 2.12(b). It is obvious from this figure that a wide frequency range from about 4 to about 8 GHz is covered by the spectrum, and therefore the damping sinusoid properties of Figure 2.12(b) are removed.

There are several advantages resulting from using multi-band UWB modulation. One advantage is the ability to efficiently use the entire 7.5 GHz of spectrum made available by the FCC. Choosing appropriate multi-band widths can ensure full usage of the entire spectrum, as compared with single-pulse systems. Figure 2.12(b) shows how multiple bands of wideband energy might efficiently fill the specified spectrum. Notice the rapid attenuation at the band edges.

Another compelling advantage is that separate bands stacked across the available spectrum may be treated independently, creating a new level of flexibility for UWB. Coexistence and interference are other interesting areas that could benefit from a multi-band approach. Any UWB system will be expected to share the approved spectrum with existing technologies. It is important first to protect existing technologies (UWB interfering with narrowband) using mitigation techniques. Overlay concepts must then be developed to ensure UWB is robust (narrowband interfering with UWB). It will be the responsibility of UWB developers to avoid interference with existing spectrum users if UWB is to be successful.

Some form of interference mitigation will also be possible by multi-band implementation. The most obvious technology that will require coexistence next to UWB is WLAN in the 5-GHz spectrum. A multi-band implementation would be able to identify potential interference and either reduce power in the contending band or turn it off completely.

Let us consider again the wave shape of Figure 2.12(a) that was given by the time domain equation (Eqn. 2.60) and remove the term corresponding to 5 GHz.

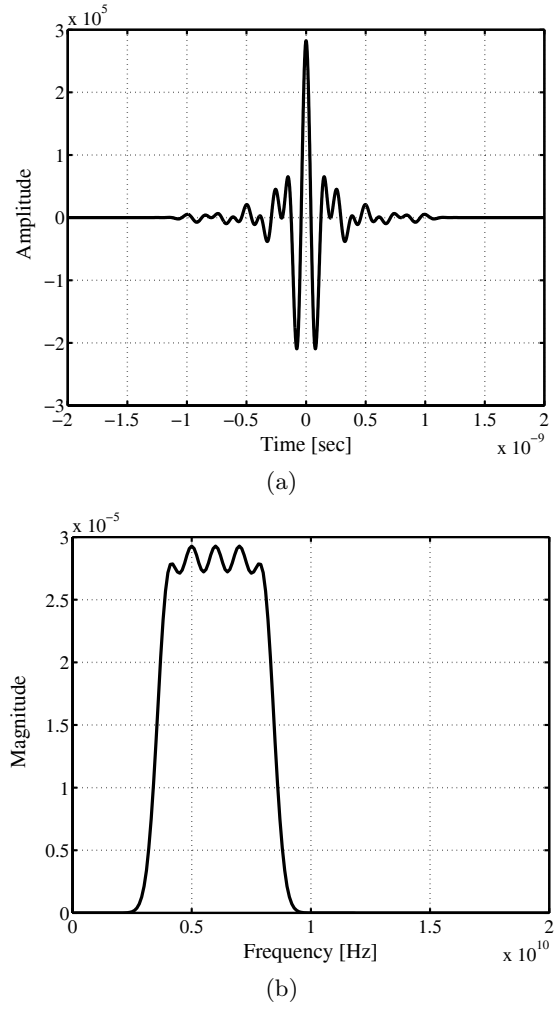
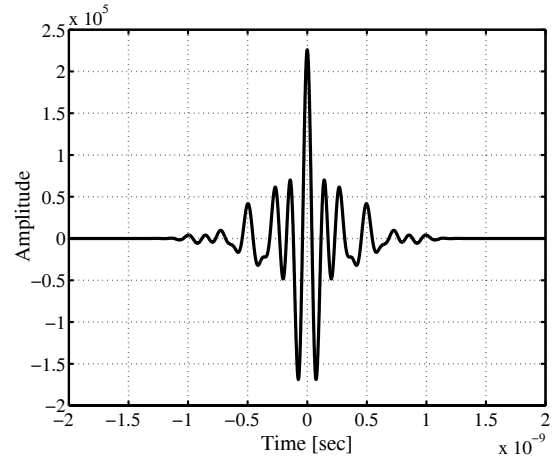
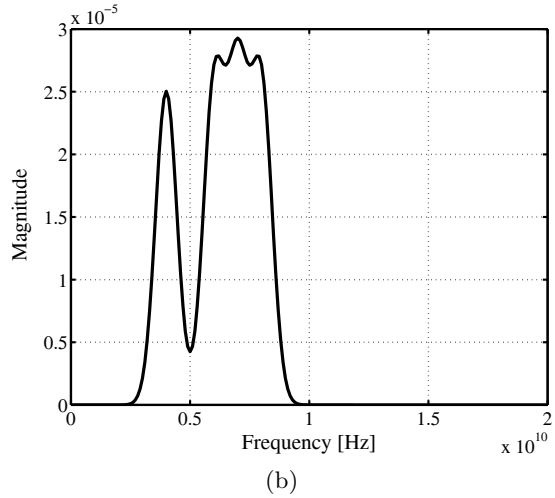


Fig. 2.12 A combination of five modulated Gaussian pulses and its frequency domain presentation.

This term belongs to f_2. The resulting pulse shape is illustrated in Figure 2.13(a). Although the removal of a frequency component might be hard to recognize when comparing Figures 2.12(a) and 2.13(a), the frequency response of Figure 2.13(b) is quite self-explanatory. As is clear from this figure the wave shape designed, based on the multi-band modulation technique, has no 5-GHz component to interfere with other communication systems operating in this frequency.



(a)



(b)

Fig. 2.13 A combination of four modulated Gaussian pulses and its frequency domain presentation after removing the 5-GHz band for interference mitigation.

The level of the null produced in Figure 2.13(b) can be adjusted adaptively by changing the number of bands and the parameter of the Gaussian pulse used for wave shaping. The following example will consider a modification of parameters in order to increase the depth of the desired null in the frequency spectrum.

Example 2.4

Modify the parameters of Figure 2.13 in such a way that the result presents a deeper null at the 5-GHz frequency band.

Solution

Let us modify the numerical values of the parameters of the example of Figure 2.13 as follows

$$f_1, f_2, \ldots, f_{13} = 3.6, 3.9, 4.2, 4.5, 5.5, 5.8, 6.1, \\ 6.4, 6.7, 7, 7.3, 7.6, 7.9 \text{ GHz} \tag{2.62}$$

and $\tau = 1.5$ ns. The individual pulses that form the combined UWB wave are narrower this time and they are arranged with a smaller frequency distance. Moreover, there is no frequency component at 5 GHz. In fact, the two nearest frequency components to 5 GHz are $f_4 = 4.5$ GHz and $f_5 = 5.5$ GHz. The resulting pulse shape and its spectrum are illustrated in Figures 2.14(a) and (b), respectively. We note that compared with Figure 2.13(b) the null produced in the frequency characteristic is much deeper, and consequently the potential interfering behavior of this UWB signal on 5-GHz communication systems will be considerably reduced.

Another way of interference mitigation with the assistance of wave shaping can be done by using a delay and sum combination of several multi-band modulated pulses. As an example consider the following signal generated at the transmitter side

$$\begin{aligned} s'_{mb}(t) &= \sum_{j=1,3,4,5} s_j(t - t_j) \\ &= \sqrt{2}\, K_1 e^{-(t-t_j/\tau)^2} \sum_{j=1,3,4,5} \cos(2\pi f_j (t - t_j)) \end{aligned} \tag{2.63}$$

where $f_1 = 4$, $f_3 = 6$, $f_4 = 7$, and $f_5 = 8$ GHz are utilized for modulating the Gaussian pulses, $\tau = 0.5$ ns, and $K_1 = \sqrt{1/(\tau\sqrt{\pi/2})} = 39,947$. The delay times are chosen as $t_1 = 0$, $t_3 = 4$, $t_4 = 6$, and $t_5 = 8$ ns. Figure 2.15(a) shows the time domain representation of the pulses of Eqn. (2.63), and Figure 2.15(b) demonstrates the frequency spectrum of these pulses. Again, it is clear that the wave shape designed this way has no 5-GHz component capable of interfering with other communication systems operating in this frequency.

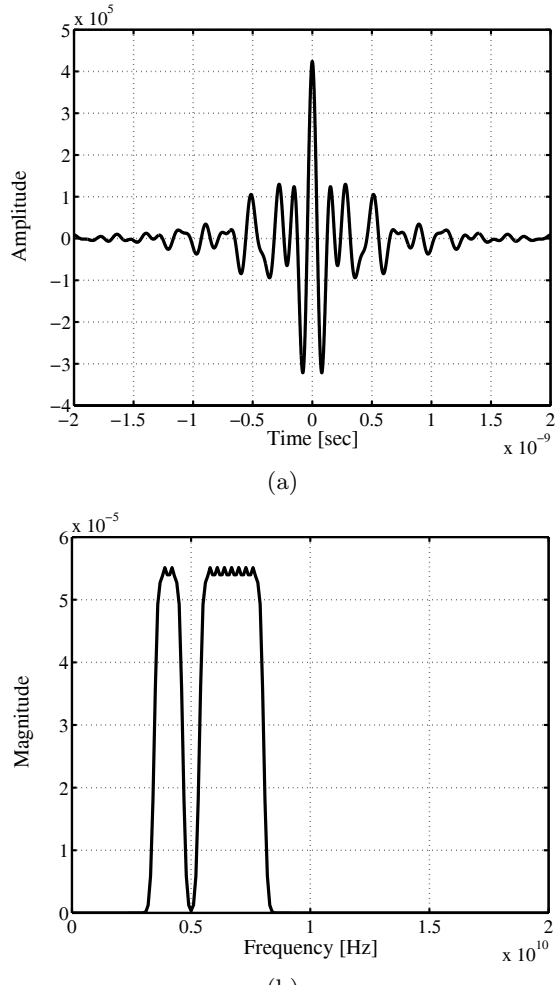


Fig. 2.14 Deeper null produced by changing the number of bands and the parameter of the Gaussian pulse used for wave shaping.

2.6 PRACTICAL CONSTRAINTS AND EFFECTS OF IMPERFECTIONS

The utilization of the radio frequency spectrum based on band sharing has long been discussed in different publications [32]. Instead of allocating the spectrum by frequency division we can construct radio waveforms from a set of orthogonal functions of inherent broad bandwidth and, in effect, allocate the spectrum by a scheme that is equivalent to code division. In this section we attempt to identify

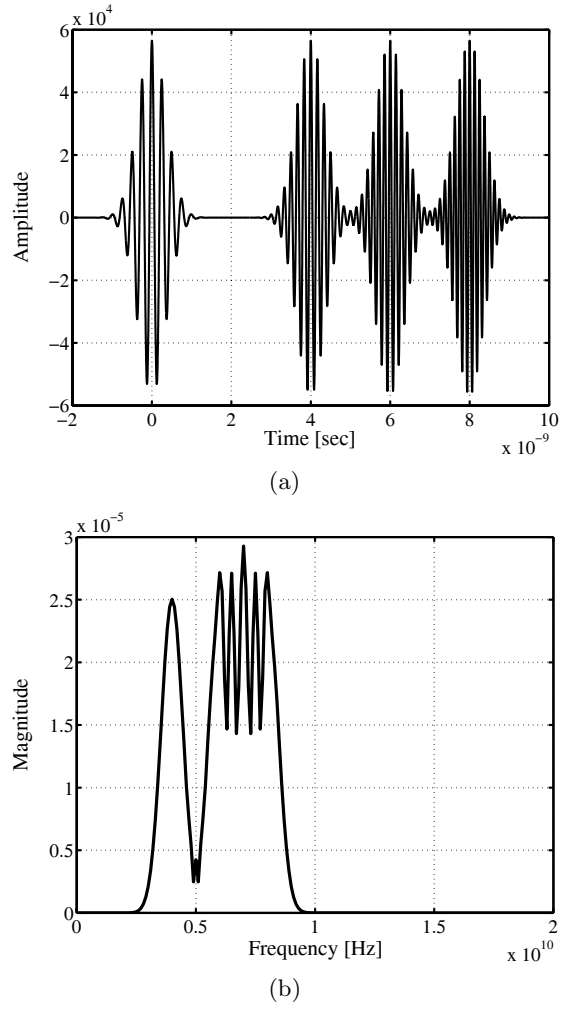


Fig. 2.15 A combination of four delayed modulated Gaussian pulses and its frequency domain presentation after removing the 5-GHz band for interference mitigation.

the key technical issues that govern the acceptability of the pulse-shaping scheme in radio frequency transmission. Some practical difficulties of implementing such a scheme for most applications are presented.

The main argument regarding the practical constraint of UWB pulse shapes is that the generation, transmission, and reception of UWB radio waves are less natural than sinusoids in most applications. Three reasons can be mentioned to support this idea [33]:

1) Radiation efficiency varies rapidly with frequency. Some UWB waveforms have considerable low-frequency content. It is a difficult practical problem to transfer useful low-frequency energy to a radiator and to accomplish its emission as radio waves, particularly if directivity is desired.

2) Relatively high-power switching circuits that might be essential in transmitters and receivers of UWB systems are usually not attractive in practice. In addition, the complexity and time-varying behavior of pulse-shaping circuits could possibly be a drawback regarding cost and reliability compared with conventional transmission systems.

3) The physical medium through which radio waves travel is time-variant on the scale of typical radio wave travel times and linear for normal field strength, but it is dispersive in general. Consequently, whereas sinusoids after having travelled through the channel are still almost sinusoids, complex functions of time no longer keep their exact shapes and, particularly, their orthogonality.

2.7 SUMMARY

This chapter discussed the various aspects and features of pulse generation in UWB communication systems. We started with the definition of damped sinusoids that are not appropriate for UWB systems due to spectral energy peaks and the possibility of interfering with other communication systems. A class of waveforms called Gaussian pulses, monocycles, and doublets was defined based on the Gauss function and differentiation. Several properties of these functions, such as energy, frequency spectrum, and bandwidth, were explained. The concept of a Gaussian pulse train for time modulation of UWB signals was also discussed. The effect of changing the pulse duration and repetition rate was examined.

A class of mathematical functions called Hermite polynomials were modified and made orthogonal. The fundamental equations representing the time and frequency domain characteristics of these functions were derived. We observed that a linear time-variant circuit can generate the pulses of different orders simultaneously.

As a second example of orthogonal sets of pulse shapes we defined the prolate spheroidal wave functions. A generator for producing four different orders of these functions in a UWB communication system was designed. Moreover, the pulses were applied to an M-ary transmission system and details for employment of the pulse generators were explained.

Designing waveforms for specific spectral masks using multi-band modulation was discussed. We noticed that by changing various parameters of the modulation (for example, the number of frequency components and the scaling parameter), we can create controlled deep nulls in the spectrum of the pulse shape and, hence, reduce the potential interfering behavior of the UWB system.

Finally, practical constraints and the effects of imperfections were briefly mentioned. We observed that radiation efficiency is a very rapidly varying characteristic with frequency. Moreover, the problem with the necessity of complicated electronic circuits was explained. We also note that the time variations and dispersive behavior of the medium through which the radio wave travels might change the shape of waveforms and, consequently, assumptions such as the orthogonality of the pulses might no longer be maintained.

Problems

Problem 1. In Section 2.2 we defined and investigated the properties of three basic Gaussian waveforms $y_{g_1}(t)$, $y_{g_2}(t)$, and $y_{g_3}(t)$ in Eqns. (2.5), (2.6), and (2.7), respectively. Proceed in the same way to derive $y_{g_4}(t)$ and:

(i) find its time domain equation;

(ii) calculate its frequency domain equation;

(iii) compare these characteristics with other Gaussian waveforms;

(iv) find the constant K_4; and

(v) calculate the bandwidth W_{g_4}.

Problem 2. Using the general formula in Eqn. (2.27) investigate the validity of MHP functions written in Eqn. (2.26).

Problem 3. By using the condition for orthogonality expressed in Eqn. (2.23) show that $h_0(t)$, $h_1(t)$, $h_2(t)$, and $h_4(t)$ are mutually orthogonal. Are they orthonormal as well?

Problem 4. Verify the validity of formulas for $H_1(f)$, $H_2(f)$, and $H_3(f)$ in Eqn. (2.35). Find a proper value for τ that gives a 3-dB bandwidth of 3 GHz for $H_2(f)$.

Problem 5. Assume that MHP pulse shapes have been applied to a differentiating process. Is their orthogonality maintained in spite of the differentiation?

Problem 6. Verify the eigenvalues χ_n given in Table 2.1. Use the details of the calculations in [26].

Problem 7. Show that

$$\sum_{n=0}^{\infty} \psi_n(0)\psi_n(t) = \mathrm{sinc}(t) \tag{2.64}$$

Problem 8. Show that

$$\sum_{n=0}^{\infty} \psi_n(k)\psi_n(m) = \delta_{km} \tag{2.65}$$

Problem 9. Show that

$$\sum_{k=-\infty}^{\infty} \psi_n(t-k) = \int_{-\infty}^{\infty} \psi_n(x)\,dx = \hat{\psi}_n(0) \tag{2.66}$$

Problem 10. Consider multi-band modulation using modulated Gaussian pulses, such as indicated in Eqn. (2.60). By using 10 different frequencies and a proper value for the parameter τ, design a pulse shape that produces a deep null at the frequency of 7 GHz.

3

Signal-processing techniques for UWB systems

The basic tools for a systematic way of solving problems of science and engineering usually include an appropriate mathematical formulation, a suitable mathematical solution, and, finally, a physical interpretation. In conventional narrowband signals and systems only the steady-state solutions to a given physical problem are considered. When the system is excited by a wideband signal the problem becomes very difficult.

It should be clearly stated that any analysis involving UWB signals and systems does not require new mathematical techniques. The treatment of systems excited by wideband waveforms can be adequately done by using, for example, two-sided Laplace transforms that have been available in the mathematical literature for over 200 years. Moreover, time domain analysis results contain equivalent information to frequency domain results, if an adequate interpretation of the results is performed. Relative or fractional bandwidth is totally irrelevant to the analysis, as Fourier or Laplace transforms are completely general in nature and require no such bandwidth restrictions.

3.1 THE EFFECTS OF LOSSY MEDIUM ON AN UWB TRANSMITTED SIGNAL

Analytical solution of the transient behavior of a class of microwave networks whose impulse response could be modeled with a train of impulses began in 1962

and was called time domain electromagnetics [7]. Early works demonstrated that time domain waveforms could be used to obtain wideband frequency domain information of two-port networks at 1 GHz. Later investigations presented a new method utilizing reflected and transmitted waves to obtain the complex permittivity and permeability of linear materials over a broad range of microwave frequencies.

In this section we do not intend to consider a detailed and sophisticated examination of the issues related to UWB electromagnetics. In fact, Chapter 4 is devoted to UWB propagation and channel modeling. The example mentioned here explains the problem of UWB pulse propagation in a lossy medium [2].

Maxwell's equations predict the propagation of electromagnetic energy away from time-varying sources (current and charge) in the form of waves. As shown in Figure 3.1, consider a linear, homogeneous, isotropic media characterized by (μ, ϵ, σ) in a source-free region (sources in region 1, source-free environment is region 2). Assume that a UWB signal propagates in this medium that has an effective lossy permittivity of ϵ and a conductivity σ. The magnetic permeability of the medium is μ. The target of the problem is to calculate the transient electric and magnetic fields if the excitation at the source is an impulsive signal, such as a UWB pulse.

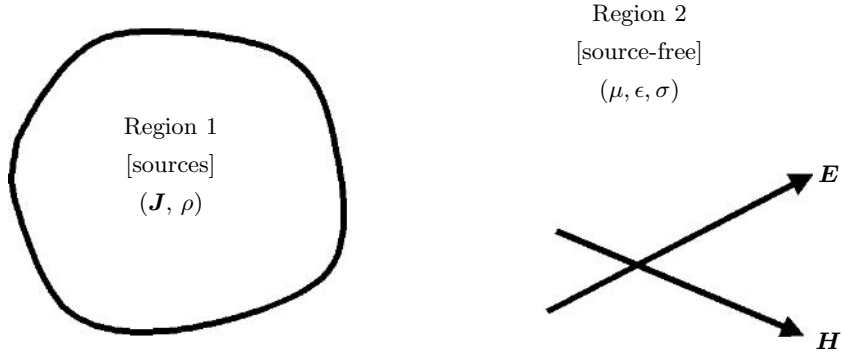


Fig. 3.1 Regions including the source and lossy medium for calculations of the electric and magnetic fields of a UWB signal.

If the electric and magnetic fields in region 2 are a function of time t and length z, then it can be shown that the electric and magnetic fields $E(z,t)$ and $H(z,t)$ satisfy the following equations

$$\mu\epsilon\frac{\partial^2 E(z,t)}{\partial t^2} + \mu\sigma\frac{\partial E(z,t)}{\partial t} - \frac{\partial^2 E(z,t)}{\partial z^2} = 0 \tag{3.1}$$

$$\mu\epsilon\frac{\partial^2 H(z,t)}{\partial t^2} + \mu\sigma\frac{\partial H(z,t)}{\partial t} - \frac{\partial^2 H(z,t)}{\partial z^2} = 0 \tag{3.2}$$

where z represents the direction of propagation and t is time. These equations are called instantaneous wave equations. The properties of an electromagnetic wave (direction of propagation, velocity of propagation, wavelength, frequency, attenuation, etc.) can be determined by examining the solutions to the wave equations that define the electric and magnetic fields of the wave.

The solution of this problem is similar to the solution of pulse propagation in a lossy transmission line. For example, if R, L, G, and C are the resistance, inductance, conductance and capacitance per unit of a lossy transmission line, respectively, then the voltage $v(z,t)$ on the transmission line is given by

$$LC\frac{\partial^2 v(z,t)}{\partial t^2} + (RC+LG)\frac{\partial v(z,t)}{\partial t} + RGv(z,t) - \frac{\partial^2 v(z,t)}{\partial z^2} = 0 \tag{3.3}$$

where the transmission line is assumed to be oriented in the z-direction. A solution to this problem has been derived using the Laplace transform. The solution for the current on the transmission line satisfies a similar differential equation to the one that was satisfied by the electric field. The original solution, first developed more than 35 years ago, is given as

$$v(z,t) = \left(e^{-\rho z/\nu}\delta\left[t-\frac{z}{\nu}\right] + \frac{\eta z}{\nu\chi}e^{-\rho t}I_1(\eta\chi)\right)\cdot u\left[t-\frac{z}{\nu}\right] \tag{3.4}$$

$$i(z,t) = \left(e^{-\rho z/\nu}\delta\left[t-\frac{z}{\nu}\right] + e^{-\rho t}\left(\frac{\eta t}{\chi}I_1(\eta\chi) - \eta I_0(\eta\chi)\right)\right)\cdot u\left[t-\frac{z}{\nu}\right] \tag{3.5}$$

where $u[\cdot]$ is the unit step function and $\delta[\cdot]$ is the delta function. The functions defined by I_0 and I_1 are the modified Bessel functions of the first kind, with order zero and one, respectively. Also

$$\rho = \frac{1}{2}\left(\frac{R}{L}+\frac{G}{C}\right) \tag{3.6}$$

$$\nu = \frac{1}{\sqrt{LC}} \tag{3.7}$$

$$\eta = \frac{1}{2}\left(\frac{R}{L}-\frac{G}{C}\right) \tag{3.8}$$

$$\chi = \sqrt{t^2 - \frac{z^2}{\nu^2}} \tag{3.9}$$

Because of the existence of the step function in the above equations the solution is guaranteed to be causal.

Now, propagation of a pulse in a lossy medium is exactly similar to pulse propagation in a lossy transmission line. Therefore, setting $R = 0$ in the above

expression leads to the following equivalent

$$\begin{aligned} v(z,t) &\Rightarrow E(z,t) \\ i(z,t) &\Rightarrow H(z,t) \\ L &\Rightarrow \mu \\ C &\Rightarrow \epsilon \\ G &\Rightarrow \sigma \end{aligned}$$

The electric and magnetic fields propagating in a lossy medium due to an arbitrarily shaped waveform excitation can easily be obtained by convolving the above expressions with the excitation waveform. It can be proven that electric and magnetic fields lie in a plane perpendicular to the direction of propagation and to each other. Figure 3.2 shows the relationship between $\boldsymbol{E}$ and $\boldsymbol{H}$ for the previously assumed uniform plane wave propagating in a slightly lossy medium.

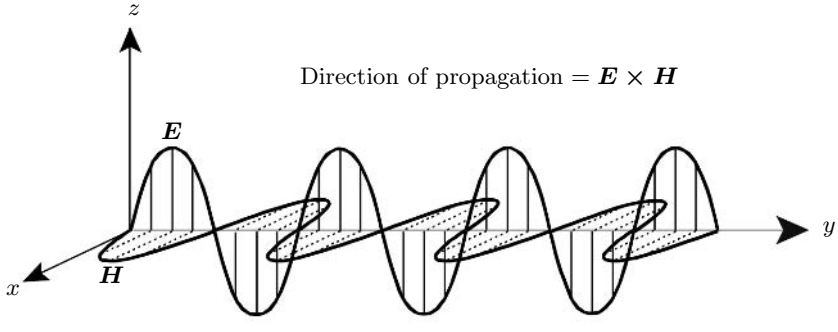


Fig. 3.2 Propagation of electric and magnetic fields.

Although this section may appear to bear little relevance to UWB signal processing it is important to consider the effects of transmission medium, such as antennas and physical propagation channel. The above analysis presents the fundamentals required for such effects to be analyzed and appreciated.

3.2 TIME DOMAIN ANALYSIS

The broadband characteristics of UWB communication systems can be obtained using either the time domain or the frequency domain. Each of these methods has its own advantages and disadvantages. In time domain analysis an important objective is to find the impulse response of a system. For example, the system can be a propagation channel, a transmitter antenna, or a receiver antenna. In this section various features and issues related to time domain processing of signals (in particular, UWB signals) are introduced.

3.2.1 Classification of signals

Electric signals can be classified in different ways [34]. Some examples of these classifications are now explained.

3.2.1.1 Continuous time and discrete time signals By the term *continuous time* signal we mean a real or complex function of time $s(t)$, where the independent variable t is continuous. If t is a discrete variable (i.e., $s(t)$ is defined at discrete times), then signal $s(t)$ is a *discrete time* signal. A discrete time signal is often identified as a sequence of numbers, denoted by $s(n)$, where n is an integer. For example, the functions

$$s(t) = (t^2 - 1)e^{-t^2/4} \tag{3.10}$$

and

$$s(n) = (n^2 - 1)e^{-n^2/4} \tag{3.11}$$

can represent two UWB pulses in the continuous and discrete time domains, respectively. A plot of these functions is illustrated in Figure 3.3.

3.2.1.2 Analogue and Digital Signals If a continuous time signal $s(t)$ can take on any values in a continuous time interval, then $s(t)$ is called an *analogue* signal. If a discrete time signal can take on only a finite number of distinct values $s(n)$, then the signal is called a *digital* signal.

Since UWB pulse are extremely short, it is difficult to use conventional analog-to-digital (A/D) conversion with a single converter. Further investigation into different techniques that could be used to acquire the analog subnanosecond pulsed signal that comes out from a UWB front end is required.

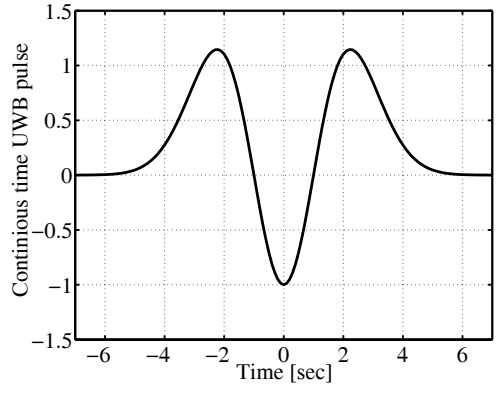
3.2.1.3 Deterministic and random signals *Deterministic signals* are those signals whose values are completely specified for any given time. For example, e^{-t^2} is well defined for any value of time.

Random signals are those signals that take random values at any given times. For instance, the noise at the receiver is a random signal and must be modeled probabilistically.

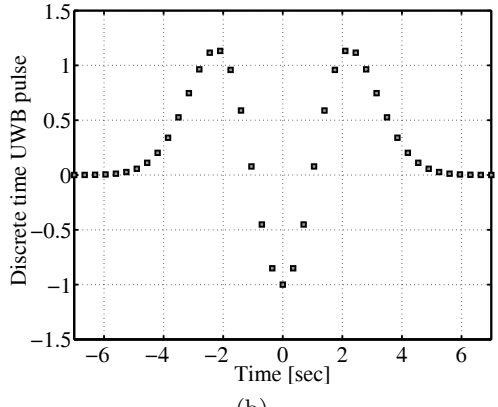
3.2.1.4 Periodic and nonperiodic signals A signal $s(t)$ is called a *periodic signal* if

$$s(t) = s(t + nT_0) \tag{3.12}$$

where T_0 is the *period* and integer $n > 0$. If $s(t) \neq s(t + T_0)$ for all t and any T_0, then $s(t)$ is a *non-periodic* or *aperiodic* signal.



(a)



(b)

Fig. 3.3 Examples of (a) continuous time and (b) discrete time UWB signals.

3.2.1.5 Power and energy signals A *complex* signal $s(t)$ is a *power* signal if the average normalized power P is finite, where

$$0 < P = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t) s^*(t)\, dt < \infty \tag{3.13}$$

and $s^*(t)$ is the complex conjugate of $s(t)$.

A *complex* signal $s(t)$ is an *energy* signal if the normalized energy E is finite, where

$$0 < E = \int_{-\infty}^{\infty} s(t)s^*(t)\,dt = \int_{-\infty}^{\infty} |s(t)|^2\,dt < \infty \tag{3.14}$$

Energy signals have finite energy. Power signals have finite and nonzero power. In fact, any signal with finite energy will have zero power and any signal with nonzero power will have infinite energy.

3.2.2 Some useful functions

For the time domain analysis of a system, special functions are defined. These functions, although not necessarily practical, can be approximated and employed for investigating the time domain characteristics of an unknown system.

3.2.2.1 Unit impulse function The *unit impulse function*, also known as the *Dirac delta function* $\delta(t)$, is shown in Figure 3.4 and is defined by

$$\int_{-\infty}^{\infty} s(\lambda)\delta(\lambda - t)\,d\lambda = s(t) \tag{3.15}$$

An alternative definition is

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1 \tag{3.16}$$

and

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{3.17}$$

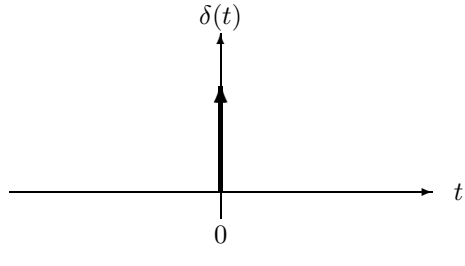


Fig. 3.4 Unit impulse function $\delta(t)$.

The response of a system to unit impulse function is called the *impulse response* of the system and is an important measure of its time domain characteristics.

3.2.2.2 Unit step function The *unit step function* $u(t)$ is shown in Figure 3.5 and is defined according to Eqn. 3.18.

$$u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases} \tag{3.18}$$

and the unit step function is related to the unit impulse function $\delta(t)$ by

$$u(t) = \int_{-\infty}^{t} \delta(\lambda)\, d\lambda \tag{3.19}$$

and

$$\frac{du(t)}{dt} = \delta(t) \tag{3.20}$$

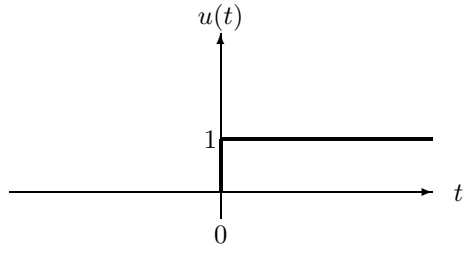


Fig. 3.5 Unit step function $u(t)$.

3.2.2.3 Sinc Function A *sinc function* is denoted by

$$\text{sinc}(t) = \frac{\sin \pi t}{\pi t} \tag{3.21}$$

and is illustrated in Figure 3.6.

3.2.2.4 Rectangular function A single *rectangular pulse* is denoted by

$$\Pi\left(\frac{t}{T}\right) = \begin{cases} 1, & |t| < \dfrac{T}{2} \\ 0, & |t| > \dfrac{T}{2} \end{cases} \tag{3.22}$$

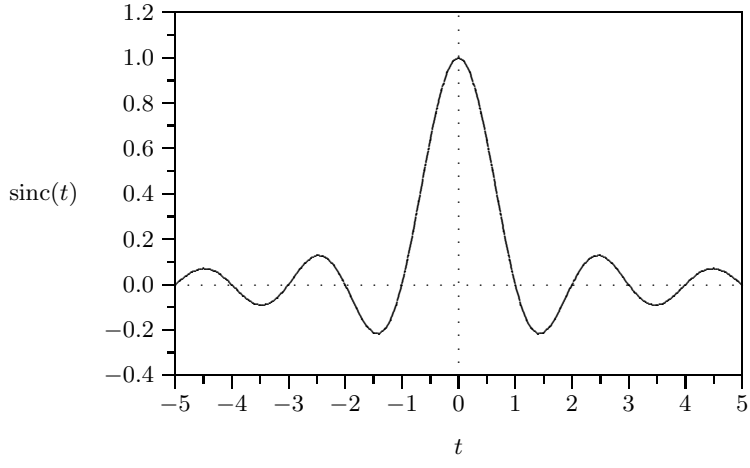


Fig. 3.6 Sinc function.

3.2.2.5 Triangular function A *triangular function* is denoted by

$$\Lambda\left(\frac{t}{T}\right) = \begin{cases} 1 - \dfrac{|t|}{T}, & |t| < T \\ 0, & |t| > T \end{cases} \tag{3.23}$$

3.2.3 Some useful operations

3.2.3.1 Time average The *time average operator* is given by

$$\langle[\cdot]\rangle = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [\cdot]\, dt \tag{3.24}$$

If a waveform is periodic, the time average operator can be reduced to the following:

$$\langle[\cdot]\rangle = \frac{1}{T_0} \int_{a-T_0/2}^{a+T_0/2} [\cdot]\, dt \tag{3.25}$$

where T_0 is the period of the waveform and a is an arbitrary real constant, which may be taken to be zero. Equation (3.25) readily follows from Eqn. (3.24) because integrals over successive time intervals that are T_0 seconds wide have identical area if the waveform is periodic. As these integrals are summed the total area and T are proportionally larger, resulting in a value for the time average that is the same as just integrating over one period and dividing by T_0. In summary, Eqn. (3.24) may be used to evaluate the time average of any type of waveform. Equation (3.25) is valid only for periodic waveforms.

3.2.3.2 Direct current value The *direct current* (dc or DC) value of a waveform is given by

$$\langle s(t)\rangle = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)\, dt \tag{3.26}$$

We can see that this is the time average of $s(t)$. Over a finite interval of interest, the dc value is

$$\langle s(t)\rangle = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} s(t)\, dt \tag{3.27}$$

3.2.3.3 Power and energy The *instantaneous power* (incremental work divided by incremental time) is given by

$$p(t) = v(t)i(t) \tag{3.28}$$

where $v(t)$ denotes voltage and $i(t)$ denotes current.

The *average power* is given by

$$\begin{aligned} P &= \langle p(t)\rangle \\ &= \langle v(t)i(t)\rangle \end{aligned} \tag{3.29}$$

The *root mean square* (rms) value of $s(t)$ is given by

$$S_{\text{rms}} = \sqrt{\langle s^2(t)\rangle} \tag{3.30}$$

If a load is resistive, the average power is given by

$$\begin{aligned} P &= \frac{\langle v^2(t)\rangle}{R} = \langle i^2(t)\rangle R \\ &= \frac{V_{\text{rms}}^2}{R} = I_{\text{rms}}^2 R = V_{\text{rms}} I_{\text{rms}} \end{aligned} \tag{3.31}$$

where R is the value of the resistive load. When $R = 1\ \Omega$, P becomes the *normalized power*.

The *average normalized power* of a *real valued* signal $s(t)$ is given by

$$\begin{aligned} P &= \langle s^2(t)\rangle \\ &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} s^2(t)\, dt \end{aligned} \tag{3.32}$$

The *total normalized energy* of a *real valued* signal $s(t)$ is given by

$$E = \lim_{T\to\infty} \int_{-T/2}^{T/2} s^2(t)\, dt \tag{3.33}$$

3.2.3.4 Decibel The decibel gain of a circuit is given by

$$\mathrm{dB} = 10 \log_{10} \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right) \tag{3.34}$$

If resistive loads are involved, it can be reduced to

$$\mathrm{dB} = 20 \log_{10} \left(\frac{V_{\text{rms out}}}{V_{\text{rms in}}} \right) + 10 \log_{10} \left(\frac{R_{\text{in}}}{R_{\text{load}}} \right) \tag{3.35}$$

or

$$\mathrm{dB} = 20 \log_{10} \left(\frac{I_{\text{rms out}}}{I_{\text{rms in}}} \right) + 10 \log_{10} \left(\frac{R_{\text{load}}}{R_{\text{in}}} \right) \tag{3.36}$$

If normalized powers are used

$$\mathrm{dB} = 20 \log_{10} \left(\frac{V_{\text{rms out}}}{V_{\text{rms in}}} \right) = 20 \log_{10} \left(\frac{I_{\text{rms out}}}{I_{\text{rms in}}} \right) \tag{3.37}$$

The *decibel power level* with respect to 1 mW is given by

$$\mathrm{dBm} = 10 \log_{10} \left(\frac{\text{Actual power level in watts}}{10^{-3}} \right) \tag{3.38}$$

or, alternatively the *decibel power level* with respect to 1 W is given by

$$\mathrm{dBW} = 10 \log_{10}(\text{actual power level in watts})$$

The *decibel voltage level* with respect to a 1-mV rms level is given by

$$\mathrm{dBmV} = 10 \log_{10} \left(\frac{V_{\text{rms}}}{10^{-3}} \right) \tag{3.39}$$

3.2.3.5 Cross-correlation The *cross-correlation* of two *real valued power* waveforms $s_1(t)$ and $s_2(t)$ is defined by

$$\begin{aligned} R_{12}(\tau) &= \langle s_1(t) s_2(t+\tau) \rangle \\ &= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_1(t) s_2(t+\tau)\, dt \end{aligned} \tag{3.40}$$

If $s_1(t)$ and $s_2(t)$ are periodic with the same period T_0, then

$$R_{12}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} s_1(t) s_2(t+\tau)\, dt \tag{3.41}$$

The cross-correlation of two *real valued energy* waveforms $s_1(t)$ and $s_2(t)$ is defined by

$$R_{12}(\tau) = \int_{-\infty}^{\infty} s_1(t)s_2(t+\tau)\,dt \tag{3.42}$$

Correlation is a useful operation to measure the similarity between two waveforms. To compute the correlation between waveforms it is necessary to specify which waveform is being shifted. In general, $R_{12}(t)$ is not equal to $R_{21}(t)$, where $R_{21}(t) = \langle s_2(t)s_1(t+\tau)\rangle$.

The cross-correlation of two complex waveforms is $R_{12}(\tau) = \langle s_1^*(t)s_2(t+\tau)\rangle$.

3.2.3.6 Autocorrelation The *autocorrelation* of a *real valued power* waveform $s_1(t)$ is defined by

$$\begin{aligned} R_{11}(\tau) &= \langle s_1(t)s_1(t+\tau)\rangle \\ &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} s_1(t)s_1(t+\tau)\,dt \end{aligned} \tag{3.43}$$

If $s_1(t)$ is *periodic* with fundamental period T_0, then

$$R_{11}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} s_1(t)s_1(t+\tau)\,dt \tag{3.44}$$

The autocorrelation of a *real valued energy* waveform $s_1(t)$ is defined by

$$R_{11}(\tau) = \int_{-\infty}^{\infty} s_1(t)s_1(t+\tau)\,dt \tag{3.45}$$

The autocorrelation of a *complex power* waveform is $R_{11}(\tau) = \langle s_1^*(t)s_1(t+\tau)\rangle$.

3.2.3.7 Convolution The *convolution* of a waveform $s_1(t)$ with a waveform $s_2(t)$ is given by

$$\begin{aligned} s_3(t) &= s_1(t) * s_2(t) \\ &= \int_{-\infty}^{\infty} s_1(\lambda)s_2(t-\lambda)\,d\lambda \\ &= \int_{-\infty}^{\infty} s_1(\lambda)s_2(-(\lambda-t))\,d\lambda \end{aligned} \tag{3.46}$$

where $*$ denotes the convolution operation. The third line of this equation is obtained by

1. Time reversal of $s_2(t)$ to obtain $s_2(-\lambda)$.

2. Time shifting of $s_2(-\lambda)$ to obtain $s_2(-(\lambda - t))$.

3. Multiplying $s_1(\lambda)$ and $s_2(-(\lambda - t))$ to form the integrand $s_1(\lambda)s_2(-(\lambda - t))$.

As an example let us convolve a rectangular waveform $s_1(t)$ with an exponential waveform $s_2(t)$ which are defined by the following equations

$$s_1(t) = \begin{cases} 1, & 0 < t < T \\ 0, & \text{elsewhere} \end{cases} \tag{3.47}$$

$$s_2(t) = e^{-t/T}u(t) \tag{3.48}$$

The steps involved in the convolution are illustrated in Figure 3.7.

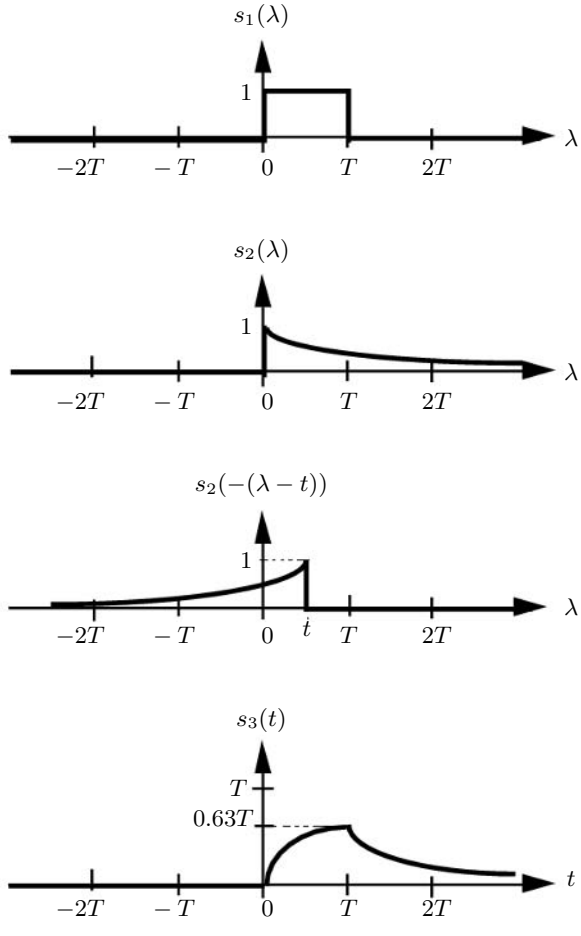


Fig. 3.7 Convolution of a rectangular waveform with an exponential waveform.

3.2.4 Classification of systems

A system is a mathematical model that relates the output signal to the input signal of a physical process. Representation of a system is shown in Figure 3.8. Classification of signals and systems will help us in finding a suitable mathematical model for a given physical process that is to be analyzed.

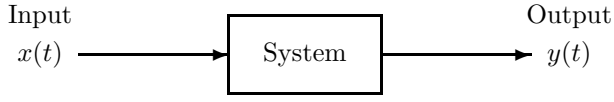


Fig. 3.8 Representation of a system with input $x(t)$ and output $y(t)$.

3.2.4.1 Linear and nonlinear systems Let $x_i(t)$ and $y_i(t)$, $i \geq 1$, be input and output signals of a system, respectively. As shown in Figure 3.9 a system is called a *linear* system if the input $x_1(t) + x_2(t) + \cdots + x_i(t) + \cdots$ produces a response $y_1(t) + y_2(t) + \cdots + y_i(t) + \cdots$ and $ax_i(t)$ produces $ay_i(t)$ for all input signals $x_i(t)$ and scalar a. This is known as the *superposition theorem*, and a linear system obeys this principle.

In practice, it may be found that a system is only linear over a limited range of input signals.

A *nonlinear* system does not obey the superposition theorem.

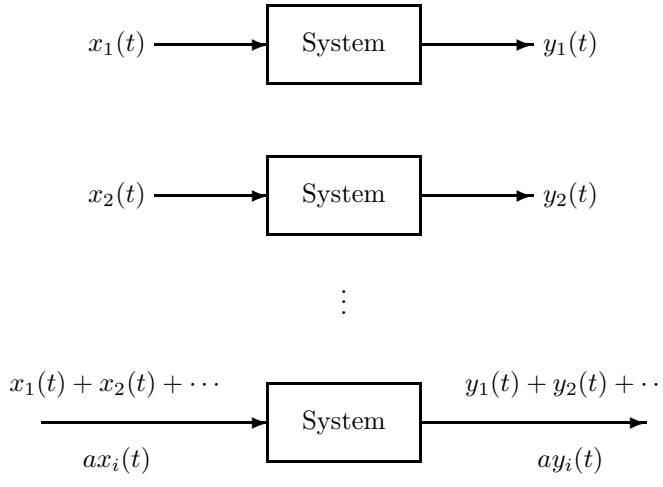


Fig. 3.9 Conditions for a system to be linear.

3.2.4.2 Causal and noncausal systems Let $x(t)$ and $y(t)$ be the input and output signals of a system (Figure 3.10). A *causal* (physically realizable) system produces an output response at time t_1 for an input at time t_0, where $t_0 \leq t_1$. In other words, a causal system is one whose response does not begin before the input signal is applied.

A *noncausal* system response will begin before the input signal is applied. It can be made realizable by introducing a positive time delay into the system.

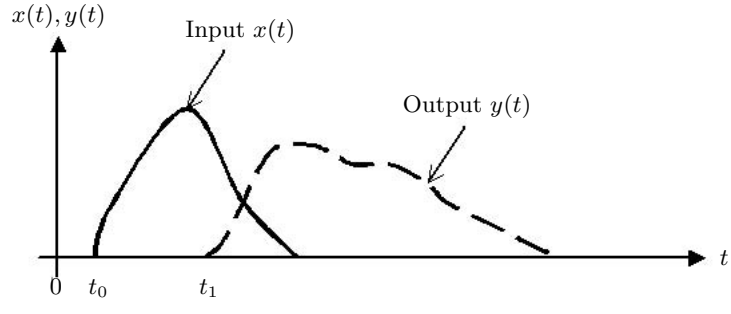


Fig. 3.10 Signals associated with a causal system.

3.2.4.3 Time-invariant and time-varying systems If the input $x(t - t_0)$ produces a response $y(t - t_0)$ where t_0 is any real constant, the system is called a *time-invariant* system (Figure 3.11). If the above condition is not satisfied, the system is called a *time-varying* system.

A system is called a *linear time-invariant* (LTI) system if the system is linear and time-invariant.

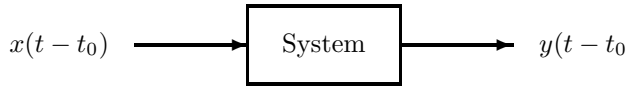


Fig. 3.11 Time-invariant system.

3.2.5 Impulse response

The *impulse response* $h(t)$ of an LTI system is defined as the response of the system when the input signal $x(t)$ is a delta function $\delta(t)$. The output $y(t)$ of an LTI system can be expressed as the convolution of the input signal $x(t)$ and the

impulse response $h(t)$ of the system, i.e.

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\lambda)h(t-\lambda)\,d\lambda \tag{3.49}$$

$$y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(\lambda)x(t-\lambda)\,d\lambda \tag{3.50}$$

For a causal LTI system

$$y(t) = \int_{0}^{\infty} x(\lambda)h(t-\lambda)\,d\lambda = \int_{0}^{\infty} h(\lambda)x(t-\lambda)\,d\lambda \tag{3.51}$$

3.2.6 Distortionless transmission

In communication systems, distortionless transmission is often desired. This implies that the output signal $y(t)$ is given by

$$y(t) = Kx(t-t_d) \tag{3.52}$$

where K is a constant and t_d is a time delay.

3.3 FREQUENCY DOMAIN TECHNIQUES

Frequency response is the gain and phase response of a circuit or other unit under test at all frequencies of interest. Although the formal definition of frequency response includes both the gain and phase, in common usage the frequency response often only implies the magnitude or gain.

The frequency response is directly related to the Fourier transform of a continuous time or discrete time signal.

3.3.1 Fourier transforms

The *Fourier transform* of a continuous time function $h(t)$ is denoted by $H(f)$ and is defined as follows

$$H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\,dt \tag{3.53}$$

The *discrete time Fourier transform* (DTFT) of a sequence $h(n)$ is given by

$$H(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} h(n)e^{-j\Omega n} \tag{3.54}$$

if the sum converges. Uniform convergence is assured for all sequences that are absolutely summable

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty \tag{3.55}$$

Note that Eqn. (3.54) corresponds to the definition of the two-sided z-transform which will be defined later. Moreover, from its definition it follows directly that the DTFT is periodic in Ω with period 2π

$$\begin{aligned} H(e^{j(\Omega+k2\pi)}) &= \sum_{n=-\infty}^{\infty} h(n)e^{-j(\Omega+k2\pi)n} \\ &= \sum_{n=-\infty}^{\infty} h(n)e^{-j\Omega n} = H(e^{j\Omega}) \end{aligned} \tag{3.56}$$

This implies that one period defines the DTFT completely.

Frequency response measurements require the excitation of the system under test at all relevant frequencies. The fastest way to perform the measurement is to use a broadband excitation signal that excites all frequencies simultaneously and use fast Fourier transform (FFT) techniques to measure at all of these frequencies at the same time. Noise and nonlinearity are best minimized by using random noise excitation, but short impulses or rapid sweeps (chirps) may also be used. When the desired resolution bandwidth of interest is low the fastest way to measure the frequency response functions is to use FFT-based techniques.

3.3.2 Frequency response approaches

There are various approaches for practical measurement of the frequency response of a system. Some of them are explained now.

3.3.2.1 Sine generator/voltmeter We apply a sine wave to the input of the system under test and measure the output voltage. Then, we repeat this process for each frequency. The gain of the system is the ratio of the output voltage to the input voltage.

Another variation of this method is that we apply a sine wave to the input and measure the phase of the output relative to the input at each frequency of interest. This method has the advantage of being low-cost and simple. It is also quite slow, and the following assumptions must be fulfilled in order for the measurement to be accurate:

1. The output voltage of the signal generator is stable during the measurement and also at all frequencies. If there is doubt about this the voltage must be measured at each frequency.

2. The system does not create significant distortion.

3. There is no significant noise on the output of the system. Otherwise, the measured output voltage will be too high. As a rule of thumb, if there is 1% distortion or noise in the system the error will be of the same order of magnitude.

4. The output must be statistically correlated to the input. This assumption is normally true in high-fidelity analog systems. However, in systems with complex transmission mechanisms and/or with digital encoding, echo cancelling, and other adaptive techniques, this assumption may not be fulfilled.

To account for all of the above, you can use digital signal-processing techniques including FFT and cross spectral methods.

Another variation that we might use is a swept sine wave generator and an associated voltmeter. The requirements of this method are as follows:

1. The sweep time for a given bandwidth must be greater than the reciprocal of the desired bandwidth. For example, if a resolution of 1 MHz is desired the sweep time for the 1 MHz must be at least 1 μs.

2. The integration time of the voltmeter must be short enough, otherwise it cannot respond fully.

This is a variant of the first method in that it uses continuous swept sine waves, instead of discretely stepped sine waves. This variation can be faster than the first technique, but it must fulfill the same assumptions.

Certain instruments may have "adaptive sweep", where the sweep rate adapts to the rate of change of the output signal. For example, when sweeping through a very sharp resonance, the sweep rate is reduced to fully resolve the resonance peak.

The third variation a swept sine with tracking filtering similar to that of the second variation is used, but this method has the advantage of being able to reject noise and distortion from the system by using a filter on the output that follows the frequency of the input. This method must meet all the requirements regarding averaging times.

A sophisticated variant of this method offsets (delays) the receiving filter with a fixed frequency offset (corresponding to a fixed time delay) and makes it possible to measure the frequency response of delayed signal paths.

When making this measurement we make sure that the output impedance of the sine generator is low compared with the input impedance of the unit under test. Otherwise, the actual input voltage applied may drop or be changed as a function of frequency.

3.3.2.2 Transient or noise excitation with cross spectral techniques It might be possible to use any signal that contains frequency components in the range of

interest. The signals are not required to have the same amplitude. However, all measurements using cross spectral techniques require simultaneous measurement of both input and output signals, using simultaneously sampling A/D converters.

The frequency response can be computed as

$$H_{xy}(f) = \frac{G_{xy}}{G_{xx}} \tag{3.57}$$

where G_{xy} is the cross spectrum and G_{xx} is the autospectrum of the input.

This technique computes the correlation between the input and output signal (as a function of frequency) and, hence, rejects noise and distortion. The more statistical samples that are included in the averaging the greater the noise and distortion rejection and, hence, the greater the accuracy of the measurement. The resulting statistical function, called the *cross spectrum*, is then normalized for the actual amplitude of the signal at each frequency on the input (called the *auto-spectrum*, or, more commonly, the averaged spectrum). This gives the frequency response function, which contains both magnitude and phase information. The magnitude is typically shown on a logarithmic Y axis (in dB), and the phase is often shown on a 0 to 360-degree scale.

This approach has the advantage of overcoming noise, distortion, and non-correlated effects. It also corrects for any loading effects on the input to the system. In addition, the technique can be extremely rapid, because it measures all frequencies of interest simultaneously. Its only weakness is that its signal-to-noise ratio (SNR) can be lower than the swept sine with the tracking filter technique.

3.3.2.3 Naturally occurring excitation Sometimes, it is not possible to insert an excitation signal into the system to be tested. However, if you want to measure the frequency response function you can use the naturally occurring "input signals" coming from the surrounding environment as excitation signals. Using cross spectral techniques you can measure the input signal and cross-correlate it with the output signal.

When making this measurement you should take extreme care to account for triggering and windowing conditions and also consider the potential time delays between the input and output. Thus, this technique is only recommended for experienced professionals with a thorough understanding of digital signal-processing techniques.

3.3.3 Transfer function

In the frequency domain the Fourier transform of

$$y(t) = h(t) * x(t) \tag{3.58}$$

is

$$Y(f) = H(f)X(f) \tag{3.59}$$

where $H(f)$ is the Fourier transform of $h(t)$. The function $H(f)$ is called the *transfer function* or *frequency response* of the LTI system.

Example 3.1

Figure 3.12(a) shows a two-stage RC circuit with two resistors R and two capacitors C. Find and sketch the impulse response and frequency response of the circuit.

Solution

Using well-known circuit theory techniques we can derive the differential equation relating $v_1(t)$ and $v_2(t)$ as follows

$$R^2C^2\ddot{v}_2(t) + 3RC\dot{v}_2(t) + v_2(t) = v_1(t) \tag{3.60}$$

The impulse response of the circuit is defined as $h(t) = v_2(t)$ if $v_1(t) = \delta(t)$. It can be shown that a solution in the form of

$$v_2(t) = (k_0 + k_1 e^{-(3+\sqrt{5})t/2RC} + k_2 e^{-(3-\sqrt{5})t/2RC})u(t) \tag{3.61}$$

can satisfy the differential equation. We also note that because the right-hand side of Eqn. (3.60) has an infinite discontinuity of type $\delta(t)$ at $t = 0$, the left-hand side should have the same level of discontinuity at the term involving the second derivative. Consequently, the first derivative of $v_2(t)$ will have a finite discontinuity at $t = 0$ and, hence, $v_2(t)$ should be continuous. Assuming zero initial charge at $t = 0$ on both capacitors C_1 and C_2 we should have $v_2(0) = 0$, hence

$$v_2(t) = (-(k_1 + k_2) + k_1 e^{-(3+\sqrt{5})t/2RC} + k_2 e^{-(3-\sqrt{5})t/2RC})u(t) \tag{3.62}$$

Substituting Eqn. (3.62) into Eqn. (3.60) and equating both sides of the resulting expression yields

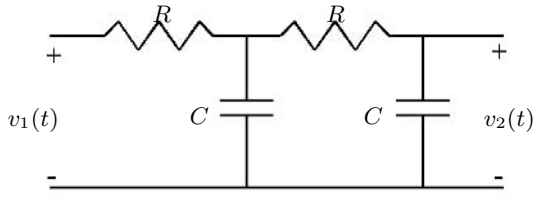
$$k_2 = -k_1 = \frac{1}{RC\sqrt{5}} \tag{3.63}$$

Therefore, the impulse response of the circuit becomes

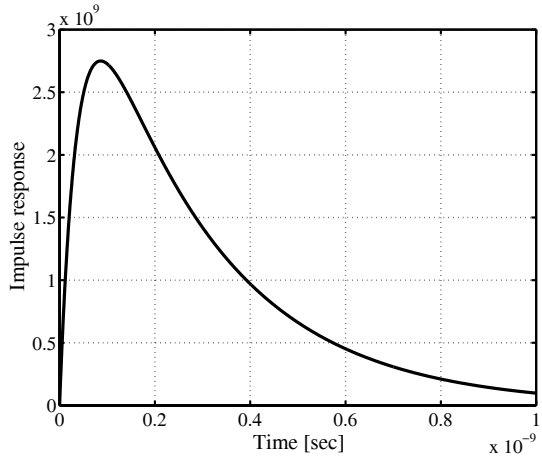
$$h(t) = v_2(t) = \frac{1}{RC\sqrt{5}}(-e^{-(3+\sqrt{5})t/2RC} + e^{-(3-\sqrt{5})t/2RC})u(t) \tag{3.64}$$

and is shown in Figure 3.12(b) for $R = 10\ \Omega$ and $C = 10$ pF.

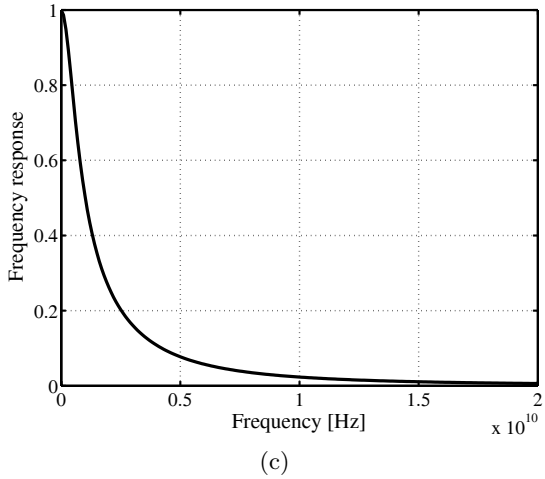
The frequency response or the transfer function of the circuit can be calculated by finding the Fourier transform of the impulse response. First, let us consider the Fourier transform of $h_1(t) = e^{\alpha t}u(t)$. Using Eqn. (3.53) we can



(a)



(b)



(c)

Fig. 3.12 A simple two-stage RC circuit and its time and frequency response for $R = 10\ \Omega$ and $C = 10$ pF.

write

$$\begin{aligned} H_1(f) &= \int_{-\infty}^{\infty} e^{\alpha t} u(t) e^{-j2\pi f t}\, dt \\ &= \int_{0}^{\infty} e^{(\alpha - j2\pi f)t}\, dt \\ &= \frac{1}{j2\pi f - \alpha} \end{aligned} \tag{3.65}$$

Now the Fourier transform of Eqn. (3.64) can be derived as follows:

$$\begin{aligned} H(f) &= \frac{1}{RC\sqrt{5}} \left[-\frac{1}{j2\pi f + (3+\sqrt{5})/2RC} + \frac{1}{j2\pi f + (3-\sqrt{5})/2RC} \right] \\ &= \frac{1}{1 + j6\pi RCf - 4\pi^2 R^2 C^2 f^2} \end{aligned} \tag{3.66}$$

An illustration of this equation is shown in Figure 3.12(c) and indicates the low-pass characteristic of the circuit.

3.3.4 Laplace transform

The one-sided Laplace transform of the function $h(t)$ is defined by

$$H(s) = \int_{0}^{\infty} h(t) e^{-st}\, dt \tag{3.67}$$

where t is real and $s = \sigma + j\omega$ is complex. There is also a two-sided Laplace transform obtained by setting the lower integration limit from 0 to $-\infty$. Since we will be analyzing only causal linear systems using the Laplace transform we can use either. However, it is customary in engineering treatments to use the one-sided definition. The one and two-sided Laplace transforms are also called the unilateral and bilateral Laplace transforms, respectively.

When evaluated along the $s = j\omega$ axis (i.e., $\sigma = 0$), the Laplace transform reduces to the Fourier transform. The Laplace transform can therefore be viewed as a generalization of the Fourier transform from the real line (a simple frequency axis) to the complex plane. One benefit of the more general Laplace transform is the ability to transform signals which have no Fourier transform. To see this we can write the Laplace transform as

$$\begin{aligned} H(s) &= \int_{0}^{\infty} h(t) e^{-(\sigma + j\omega)t}\, dt \\ &= \int_{0}^{\infty} [h(t) e^{-\sigma t}] e^{-j\omega t}\, dt \end{aligned} \tag{3.68}$$

We can thus interpret the Laplace transform as the Fourier transform of an exponentially enveloped input signal. For $\sigma > 0$ (the so-called "right-half plane"), this exponential weighting forces the Fourier-transformed signal toward zero as $t \to \infty$. As long as the signal $h(t)$ does not increase faster than e^{Bt} for some B, its Laplace transform will exist for all $\sigma > B$.

Example 3.2

Find the Laplace transform of the output signal $v_2(t)$ in Example 3.1.

Solution

Similar to the Fourier transform we might write

$$\begin{aligned} H_1(s) &= \int_0^\infty e^{\alpha t} e^{-st}\, dt \\ &= \frac{1}{s-\alpha} \end{aligned} \tag{3.69}$$

Now $V_2(s)$ will be easily calculated as follows:

$$\begin{aligned} V_2(s) &= \frac{1}{RC\sqrt{5}} \left[-\frac{1}{s + (3+\sqrt{5})/2RC} + \frac{1}{s + (3-\sqrt{5})/2RC} \right] \\ &= \frac{1}{1 + 3RCs + s^2} \end{aligned} \tag{3.70}$$

3.3.5 z-Transform

The z-transform, like the Laplace transform, is an indispensable mathematical tool for the design, analysis, and monitoring of systems. The z-transform is the discrete time counterpart of the Laplace transform and a generalization of the Fourier transform of a sampled signal. Like the Laplace transform the z-transform allows insight into the transient behavior, the steady-state behavior, and the stability of discrete time systems. A working knowledge of the z-transform is essential to the study of UWB signals and systems.

The *bilateral z-transform* of the discrete time signal $h(n)$ is defined to be

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) z^{-n} \tag{3.71}$$

where z is a complex variable. Since signals are typically defined to begin (become nonzero) at time $n = 0$ and since all filters are assumed to be causal, the lower summation limit given above may be written as 0 rather than $-\infty$ to yield the *unilateral z-transform*.

As mentioned earlier, the DTFT corresponds to the definition of the two-sided z-transform for the special case of $z = e^{j\Omega}$. The region of convergence for the z-transform is determined by the range of z, where the sequence $h(n)z^{-n}$ is absolutely summable.

The z-transform of a signal $h(n)$ can be regarded as a polynomial in z^{-1}, with coefficients given by the signal samples. As an example, the finite duration signal

$$h = [\ldots, 0, 0, 1, 2, 3, 0, 0, \ldots] \tag{3.72}$$

has the z-transform

$$H(z) = 1 + 2z^{-1} + 3z^{-2} \tag{3.73}$$

Mathematically, the entire signal is converted to a complex scalar indexed by z.

The z-transform of a signal h will always exist provided:

1. the signal starts at a finite time, and
2. it is *asymptotically exponentially bounded*, i.e., there exists a finite integer n_f and finite real numbers $A \geq 0$ and σ, such that $|h(n)| < Ae^{\sigma n}$ for all $n \geq n_f$. The bounding exponential may be growing with $n(\sigma > 0)$.

These are not the most general conditions for existence of the z-transform, but they suffice for our purposes here.

Example 3.3

Find the z-transform of the following functions

$$h_1(n) = a^n u(n) \tag{3.74}$$
$$h_2(n) = -a^n u(-n-1) \tag{3.75}$$

These discrete time functions are plotted in Figure 3.13 for $a = 0.9$. Discuss the convergence and stability of both exponential functions.

Solution

Using the definition of z-transform we can write

$$\begin{aligned} H_1(z) &= \sum_{n=-\infty}^{\infty} a^n u(n) z^{-n} \\ &= \sum_{n=0}^{\infty} a^n z^{-n} \\ &= \sum_{n=0}^{\infty} (az^{-1})^n \end{aligned} \tag{3.76}$$

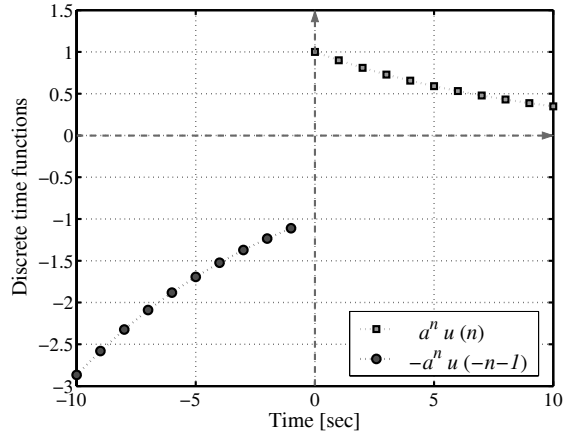


Fig. 3.13 Two discrete time exponential functions $h_1(n)$ and $h_2(n)$ for $a = 0.9$.

This series converges only if

$$\sum_{n=0}^{\infty} |az^{-1}|^n < \infty$$

equivalently, if $|az^{-1}| < 1$ or $|z| > a$. In this case we have

$$H_1(z) = \frac{1}{1 - az^{-1}} = \frac{z}{z - a} \tag{3.77}$$

For $h_2(n)$ the z-transform can be written as

$$\begin{aligned} H_2(z) &= \sum_{n=-\infty}^{\infty} [-a^n u(-n-1)] z^{-n} \\ &= -\sum_{n=-\infty}^{-1} a^n z^{-n} \\ &= -\sum_{n=1}^{\infty} a^{-n} z^n \\ &= 1 - \sum_{n=0}^{\infty} (a^{-1} z)^n \end{aligned} \tag{3.78}$$

This series converges only if $|z| < a$. In this case we have

$$H_2(z) = 1 - \frac{1}{1 - a^{-1}z} = \frac{1}{1 - az^{-1}} = \frac{z}{z - a} \tag{3.79}$$

It is interesting to note that both $h_1(n)$ and $h_2(n)$ have identical z-transforms although they are right-sided and left-sided in time, respectively. The convergence region of a z-transform is very important for its computation.

References [35] and [36] provide a detailed discussion of analog and digital signal processing, respectively.

3.3.6 The relationship between the Laplace transform, the Fourier transform, and the z-transform

The Laplace transform, the Fourier transform and the z-transform are closely related in that they all employ complex exponentials as their basis function. For right-sided signals (zero-valued for negative time index) the Laplace transform is a generalization of the Fourier transform of a continuous time signal and the z-transform is a generalization of the Fourier transform of a discrete time signal. In the previous section we have shown that the z-transform can be derived as the Laplace transform of a discrete time signal. In the following we explore the relation between the z-transform and the Fourier transform. Using the relationship

$$z = e^s = e^{\sigma} e^{j\omega} = re^{j2\pi f} \tag{3.80}$$

where $s = j\sigma$ and $\omega = 2\pi f$, we can rewrite the z-transform in the following form

$$H(z) = \sum_{n=-\infty}^{+\infty} h(n) r^{-n} e^{-j2\pi nf} \tag{3.81}$$

Note that when $r = e^{\sigma} = 1$ the z-transform becomes the Fourier transform of a sampled signal given by

$$H(z = e^{j2\pi f}) = \sum_{n=-\infty}^{+\infty} h(n) e^{-j2\pi nf} \tag{3.82}$$

Therefore, the z-transform is a simple generalization of the Fourier transform of a sampled signal. Like the Laplace transform the basis functions for the z-transform are damped or growing sinusoids of the form $z^{-n} = r^{-n} e^{-j2\pi nf}$. These signals are particularly suitable for transient signal analysis, such as UWB signals. Fourier basis functions are steady complex exponentials, $e^{-j2\pi nf}$, of time-invariant amplitudes and phases and are suitable for steady-state or time-invariant signal analysis.

A similar relationship exists between the Laplace transform and the Fourier transform of a continuous time signal. The Laplace transform is a one-sided transform with the lower limit of integration at $t = 0$, whereas the Fourier transform, Eqn. (3.53), is a two-sided transform with the lower limit of integration at $t = -\infty$. However, for a one-sided signal that is zero-valued for $t < 0$ the limits of integration for the Laplace and the Fourier transforms are identical. In that case if the variable s in the Laplace transform is replaced with the frequency variable $j2\pi$, then the Laplace integral becomes the Fourier integral. Hence, for a one-sided signal the Fourier transform is a special case of the Laplace transform corresponding to $s = j2\pi$ and $\sigma = 0$.

3.4 UWB SIGNAL-PROCESSING ISSUES AND ALGORITHMS

The distinctive properties of UWB signals require modification of analytical methods, signal-processing methods and new constructional solutions compared with conventional narrowband signals which merely require a description of envelope and phase for a full characterization. For UWB systems, representation of the signal with a single envelope and phase is neither convenient nor possible.

A UWB signal is typically composed of a train of subnanosecond pulses, resulting in a bandwidth over 1 GHz. Since the total power is spread over such a wide range of frequencies, its power spectral density is extremely low. This minimizes the interference caused to existing services that already use the same spectrum. On account of the large bandwidth used, UWB links are capable of transmitting data over tens of megabits per second. Other benefits include a low probability of interference, precise location capability and the possibility of transceiver implementation using simple architectures. It is mostly desired that UWB systems can be implemented using digital architectures because they bring low cost, ease of design, and flexibility. A single receiver would be able to support different bit rates, quality of service, and operating ranges. The vision of fully configurable software radios is an exciting one, and it appears UWB may be more amenable to such a realization than conventional, narrowband systems.

Figure 3.14 shows a generic block diagram for a digital UWB receiver [37]. A key component of such a system is the analog-to-digital converter (ADC or A/D). The ADC sampling rate for digitizing a UWB signal must be on the order of a few gigasamples/sec (GSPS). Even with the most modern process technologies this constitutes a serious challenge. Most reported data converters operating at this speed employ interleaving [38], with each channel typically based on a FLASH converter. The latter is the architecture of choice for high-speed designs, but is not suitable for high-resolution applications. An N-bit FLASH converter uses 2N comparators, so its power and area scale exponentially with resolution. Among recently reported high-speed ADCs (>1 GSPS) representing the state of the art, none has a resolution exceeding 8 bits. The minimum number of bits needed for

reliable detection of a UWB signal is, therefore, a critical parameter. If excessively large it can render an all-digital receiver infeasible.

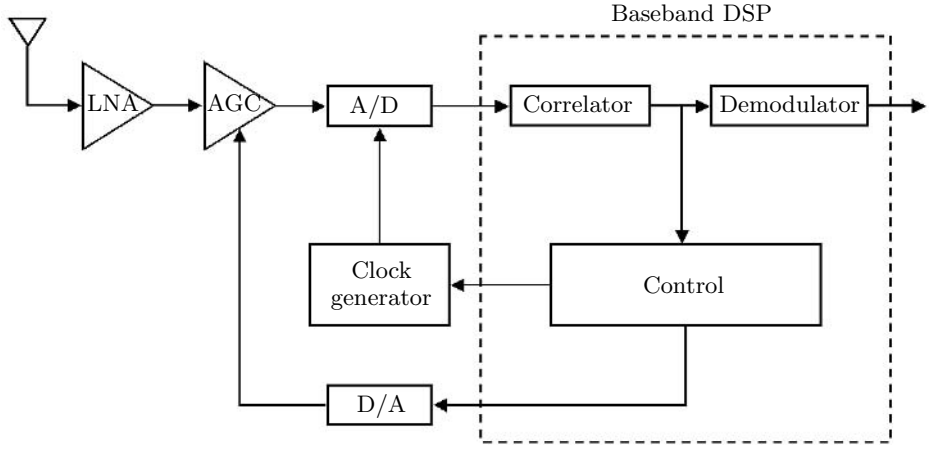


Fig. 3.14 Block diagram of a simple digital UWB receiver.

Information in such a system is typically transmitted using a collection of narrow pulses with a very low duty cycle of about 1%. Duty cycle is the ratio of pulse duration to pulse period. Each user is assigned a different pseudo-noise (PN) sequence that is used to encode the pulses in either position (pulse position modulation, PPM) or polarity (binary phase shift keying, BPSK, also known as bi-phase modulation, BPM). Channelization (the use of a single wideband and high-capacity facility to create many relatively narrowband and lower capacity channels by subdividing the wideband facility) is thus based on the assigned code, as in the case of CDMA systems.

Suppose the bitstream is denoted by a sequence of binary symbols b_j (with values $+1$ or -1) for $j = -\infty, \ldots, \infty$. A single bit is represented using N_c pulses, where N_c refers to the length of the PN code c_i. For BPSK the code modulates the polarity of a pulse within each frame. For PPM it modulates the pulse positions (incrementing or decrementing them by multiples of T_c). Data modulation is achieved by setting the sign of the block of N_c pulses for BPSK. For PPM we append an additional time-shift τ_{b_j} whose value depends on whether b_j is $+1$ or -1. Each frame has duration T_f; the duration of each bit is thus given by $N_c T_f$. Letting A denote the amplitude of each pulse $p(t)$, the transmitted signal $s(t)$ can be written as follows for the two different modulation schemes

$$s_{\text{BPSK}}(t) = A \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N_c-1} b_j c_i p(t - jN_c T_f - iT_f) \tag{3.83}$$

and

$$s_{\text{PPM}}(t) = A \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N_c-1} p(t - jN_cT_f - iT_f - c_iT_c - \tau b_j) \tag{3.84}$$

As mentioned above, each bit is represented by N_c pulses. This redundancy is one component of the signal's processing gain (PG). The other component is the duty cycle, a ratio of the short duration of each pulse to the large interval between successive ones. Processing gain refers to the boost in effective SNR as a UWB signal is processed by a correlating receiver

$$\text{PG/dB} = 10\log(N_c) + 10\log\left(\frac{1}{\text{duty-cycle}}\right) \tag{3.85}$$

Optimal detection of a noisy signal is based on matched filtering. This would entail correlating the received signal $r(t)$ against a template $s(t)$ that is an exact replica of the original transmitted signal and, then, feeding the correlator output to a slicer. However, generating exact replicas of subnanosecond pulses is a difficult problem and is highly susceptible to timing jitter.

A more tractable approach is to use a template signal comprising a train of rectangular pulses that are, in general, wider than the actual received pulses, but are coded with the same PN sequence.

Figure 3.15 shows the structure of the received and template signals, whereas Figure 3.16 depicts the operations necessary for demodulation. The signals $s_0(t)$ and $s_1(t)$ represent the template signals corresponding to a transmitted 0 and 1, respectively. They are related to one another by either a sign inversion or a time-shift, depending on whether BPSK or PPM was employed. Equations describing $s_0(t)$ and $s_1(t)$ can be obtained from Eqns. (3.83) and (3.84) by simply replacing the original pulse $p(t)$ with a rectangular pulse $\text{rect}(t)$ and setting b_j to $+1$ and -1, respectively. By using such template signals we are essentially performing a form of windowing. In other words, we are correlating the received signal against its underlying PN code over narrow windows.

The digital implementation of such a receiver entails sampling the received signal (once it has been sufficiently amplified), converting it to digital form, and, then, performing correlation. A figure analogous to Figure 3.16 describing this process can be obtained simply be replacing $r(t)$ and $s(t)$ with their discrete time counterparts $r(n)$ and $s(n)$.

3.5 DETECTION AND AMPLIFICATION

There are a number of approaches to the detection and amplification of the trains of UWB signals. In many cases there is an allocation of one-to-many in the assignment of bits to pulses to be transmitted. After 1945 the use of the correlation

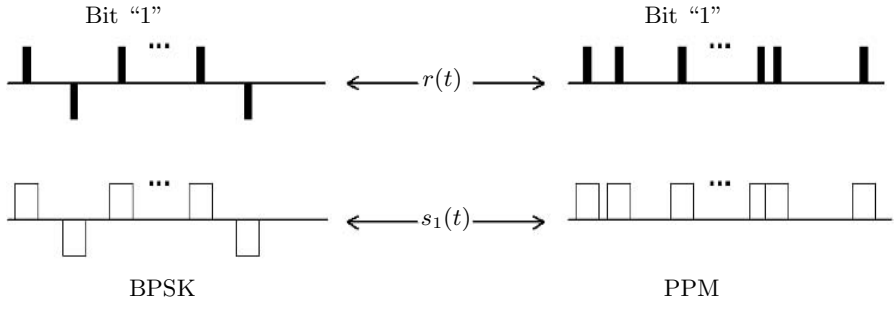


Fig. 3.15 The structure of the received and template signals.

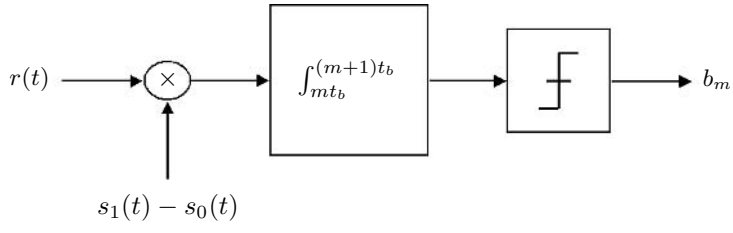


Fig. 3.16 Operations necessary for demodulation of a UWB signal.

detection receiver became commonplace. Skolnik in his introductory book [39] written in 1962 describes the use of correlation methods in the detection of weak signals and cites a number of earlier references. A synchronous detector is also shown by Fink and Christiansen in [40]. There are a variety of ways to trigger the receiver on pulse rise time, level detection, integration over time, etc.

In the case of the correlation receiver detector, UWB and gate pulses are multiplied to produce a short-output, unamplified pulse whenever there is coincidence. Next, the result is fed to a (short-term) integrator or averager to produce a reduced amplitude, stretched signal output. If the integrator time is sufficiently long (conventional correlator) or a second long-term integrator is employed, the output will then represent the average of the many high repetition rate pulses fed to the correlator. Unfortunately, the integrator not only acts as a detector but also reduces the input amplitude in the step from narrow-pulse, low-duty cycle to averaged output. The long-term integrating correlator thus effects a many-to-one detection of averaged inputs prior to any amplification.

Micro-power impulse radar (MIR), or *radar on a chip*, offered an alternative to correlation detection [41]. The MIR is an integrating peak detector as opposed to the multiply-and-average correlation receiver detector described above. In the

case of the MIR receiver detector, UWB and gate pulses are summed algebraically to form the input to a peak detector: i.e., the low-amplitude UWB pulse and the high-amplitude gate pulse, when summed, are above threshold for peak detection but individually are not. Moreover, it is not a single UWB pulse which, together with the gate pulse, provides the peak detected signal, but the (long-term) summing of a series of UWB inputs. The detector is thus triggered by the simultaneous occurrence of a summed series of low-amplitude UWB signals and a coincident large amplitude gate pulse which, together, are algebraically summed. The coincident summing method of a large gate input and summed low-amplitude signals effects a many-to-one, peak signal detection process.

3.6 SUMMARY

In this chapter the basic tools for processing UWB signals and systems were introduced. Initially, the fundamentals of impulse electromagnetics in a lossy medium was briefly explained. Basic tools for UWB waveform analysis in the time domain were presented. The concepts of phase and instantaneous frequency as a measure of the waveform informative value were introduced. Frequency response methods, such as Fourier, Laplace and z-transforms, were considered.

It should be emphasized that a specific UWB signal-processing technique does not exist. Both time and frequency domain approaches can be applicable when analyzing transmitter signaling, channel modeling, and antenna effect consideration. The most significant difference between the UWB and narrowband system is the variation of major parameters with frequency. Due to this fact, working with UWB signals and systems usually demands more complexity and higher degrees of calculations.

Problems

Problem 1. An LTI system has the following impulse response

$$h(t) = 2e^{-at}u(t) \tag{3.86}$$

Use convolution to find the response $y(t)$ to the following input

$$x(t) = u(t) - u(t-4) \tag{3.87}$$

Sketch $y(t)$ for the case when $a = 1$.

Problem 2. For each of the signals sketched in Figures 3.17(a) to (c) find the Fourier transform and plot the magnitude and phase of the resulting functions.

Problem 3. Compute the DTFT of the following signals and sketch $X(e^{j\Omega})$:

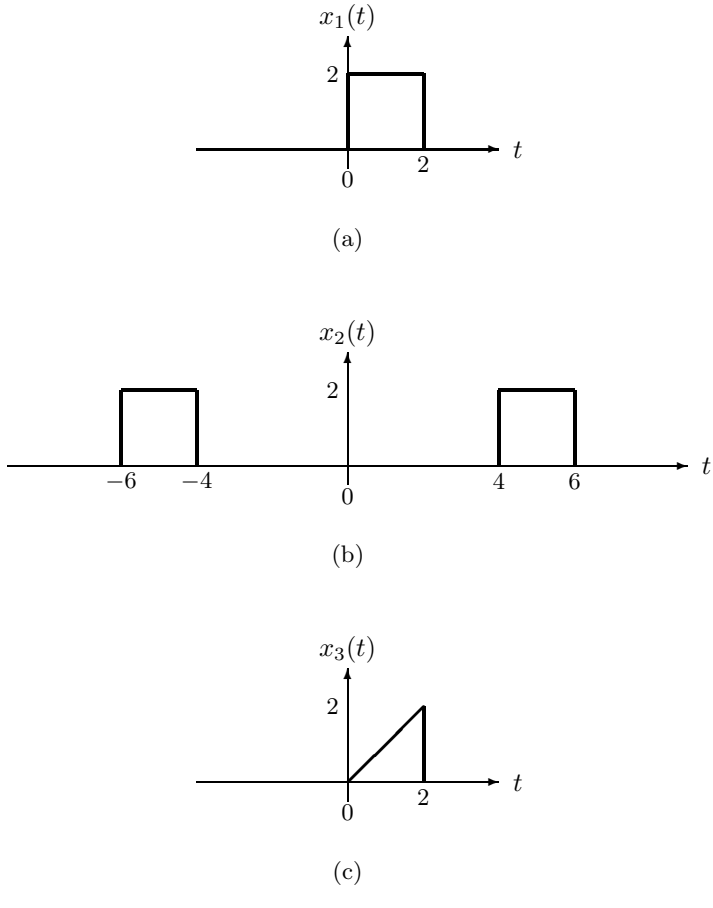


Fig. 3.17 Pulses for Problem 2.

(a) $x(n) = [1/4, 1/4, 1/4, 1/4]$

(b) $x(n) = [1, -2, 1]$

(c) $x(n) = 2(3/4)^n u(n)$

Problem 4. Find the response of the following systems to the inputs below. Sketch the magnitude of each frequency response for $-\pi < \Omega < \pi$ and determine the type of filter

$$x(n) = 2 + 2\cos\left(\frac{n\pi}{4}\right) + \cos\left(\frac{2n\pi}{3} + \frac{\pi}{2}\right) \tag{3.88}$$

(a) $H(e^{j\Omega}) = e^{-j\Omega}\cos(\Omega/2)$

(b) $H(e^{j\Omega}) = e^{-j\Omega/2}(1-\cos(\Omega/2))$

Problem 5. Compute the Laplace transforms of the following functions:

(a) $x(t) = 4\sin(100t-10)u(t-0.1)$

(b) $x(t) = tu(t) + 2(t-2)u(t-2) + (t-3)u(t-3)$

(c) $x(t) = u(t) - e^{-2t}\cos(10t)u(t)$

Problem 6. Sketch the response of each of the systems below to a step input:

(a) $H(s) = 10/s + 2$

(b) $H(s) = 0.2/(s+0.2)$

(c) $H(s) = 1/(s^2+4s+16)$

Problem 7. Find the transfer functions of the following discrete time systems:

(a) $y(n) + 0.5y(n-1) = 2x(n)$

(b) $y(n) + 2y(n-1) - y(n-2) = 2x(n) - x(n-1) + 2x(n-2)$

(c) $y(n) = x(n) - 2x(n-1) + x(n-2)$

Problem 8. Given the following system:

$$y(n) = x(n) - x(n-1) + x(n-2) \tag{3.89}$$

(a) Find the transfer function.

(b) Give the impulse response.

(c) Determine the stability.

(d) Sketch the frequency response and determine the type of filter.

Problem 9. Solve the following difference equation using the z-transform

$$\begin{aligned} y(n) + 3y(n-1) + 2y(n-2) &= 2x(n) - x(n-1) \\ y(-1) &= 0 \\ y(-2) &= 1 \\ x(n) &= u(n) \end{aligned} \tag{3.90}$$

4

Ultra wideband channel modeling

The propagation environment that a signal passes through from a transmitter to a receiver is referred to as the *channel*. Mobile cellular communication theory provides ample theory and measurement techniques for modeling signal propagation.

Thanks to its potential applications and unique capabilities the UWB wireless communication system has been the subject of extensive research in recent years. However, many important aspects of UWB-based communication systems have not yet been thoroughly investigated. In particular, indoor and outdoor channel modeling and propagation effects require careful examination before an actual implementation of UWB systems can be undertaken. Without this, system performance and interference to and from other spectrum users cannot be readily ascertained.

The propagation of UWB signals in indoor and indoor-outdoor environments is the single most important issue, with significant impacts on the future direction, scope, and, generally, the extent of the success of UWB technology. If the channel is well characterized, the effect of disturbances and other sources of perturbations can be reduced by proper design of the transmitter and receiver. Detailed characterization of UWB radio propagation is, therefore, a major requirement for successful design of UWB communication systems.

One important aspect of any radio channel-modeling activity is the investigation of the distribution functions of channel parameters. Typically, these distributions are obtained from measurements or simulations based on almost exact or simplified descriptions of the environment. However, such methods often only yield insights into the statistical behavior of the channel and are not able to give a physical

explanation of observed channel characteristics. Since these characteristics seem to yield complicated functions, both of properties of wave propagation and of the individual geometry of the surroundings for which the channel is modeled, analytical models are rare.

The typical UWB propagation channel is a function which depends only weakly on the geometry of the environment. Rough knowledge about the surroundings is supposed to be sufficient for its characterization. Otherwise, no measurement campaign conducted in one environment could be a valid approximation of the channel in another, similar situation. Compared with the deterministic one, the stochastic channel should hence be analytically more tractable. However, in a suitable approach the robustness of the channel should result from the modeling process and not from an assumption on which the derivation is based.

In UWB radio channel modeling a number of aspects have to be taken into account that amount to a changed overall behavior of the channel. The main difference between UWB and traditional channel-modeling techniques is due to the fact that in UWB propagation, frequency-dependent effects cannot be ignored. Parameters related to penetration, reflection, path loss, and many other effects should be considered frequency-variant and investigated more carefully.

4.1 A SIMPLIFIED UWB MULTIPATH CHANNEL MODEL

As a first example of channel-modeling techniques we examine a model which is not unique to UWB systems, but is considered a general model and is appropriate to explain the basic principles of channel modeling.

A convenient and simple model for characterization of UWB channels is the discrete time, multipath, impulse response model because the locations of ceilings, walls, doors, furniture, and people inside an indoor environment result in the transmit signal taking multiple paths from the transmitter to the receiver (Figure 4.1). Hence, signals arrive at the receiver with different amplitudes, phases, and delays. In this model the delay axis is divided into small time intervals called *bins*. Each bin is assumed to contain either one multipath component or no component. The possibility of more than one path in a bin is excluded. A reasonable bin width is the resolution of a specific measurement, since two paths arriving within a bin cannot be resolved as distinct paths. Using this model, each impulse response can be described by a sequence of zeros and nonzeros, where a nonzero indicates the presence of a path in a given bin and a zero represents the absence of a path in that bin.

The phenomena of multipath propagation can be represented conveniently and mathematically by the following discrete impulse response of the channel

$$h(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - lT_m) \tag{4.1}$$

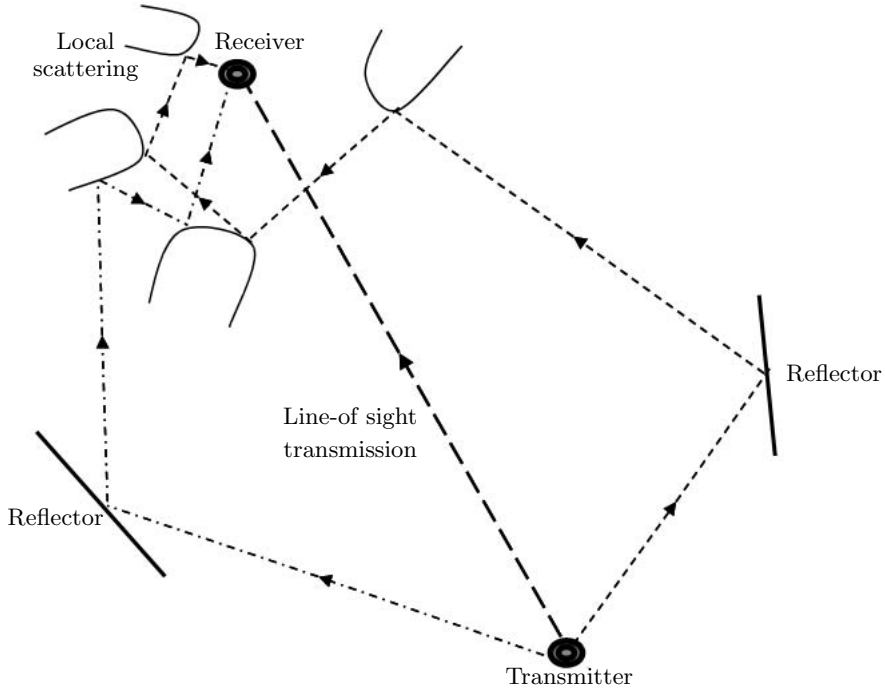


Fig. 4.1 A simple model of the indoor UWB radio multipath channel.

where α_l is the amplitude attenuation factor on path l and is a function of time and distance between transmitter and receiver. The parameter T_m is the minimum resolution time of the pulse, L is the number of resolvable multipath components, and $\delta(t)$ is the Dirac delta function. Sometimes, Eqn. (4.1) is referred to as the *multipath intensity profile*.

For simplicity, in order to avoid partial correlations of the waveform, let the minimum path resolution time be equal to the modulated symbol period for which the modulated waveform is nonzero (i.e., the UWB waveform is zero outside the interval $0 \leq t \leq T_m$).

According to Foerster [42] and Hashemi [43] the primary parameters that are important to characterize the indoor channel include the following:

- Number of resolvable multipath components.
- Multipath delay spread.
- Multipath intensity profile.

- Multipath amplitude-fading distribution.
- Multipath arrival times.

The following subsections describe each of these components in more detail.

This simple multipath channel model and its parameters have been adopted, primarily based on those described by Hashemi [43] for several reasons. First, it is based upon a large number of measurements of various indoor channels and uses a minimum resolution time of 5 nsec, which is representative of a broadband UWB channel. Second, the measurements are based on antenna separation distances of 5–30 meters, which is typical of UWB-based systems, due to the expected power-spectral density restrictions imposed by the FCC. Third, the analytical model described is relatively straightforward to realize for theoretical analysis or simulation.

4.1.1 Number of resolvable multipath components

The number of resolvable multipath components is important since it determines the design of a rake receiver. The distribution of the number of resolvable multipath components, L, for a large number of profiles collected in each building was investigated by Hashemi [43]. For each profile, L has been found by counting all multipath components within α dB of the strongest path, for example, $\alpha = 10, 20, 30$ dB. The mean and standard deviation of L for each building, each value of α, and each transmitter-receiver antenna separation was collected. Examination of the data showed that:

i. There is a clear dependence between the mean value of L and antenna separation.

ii. The mean value of L increases with increasing α. This is expected since, when α increases, more components are included in the calculation of L.

iii. Standard deviation of L increases with the increase in antenna separation. This is due to the fact that there are greater variations in the environment between the transmitter and receiver for large antenna separations. Also, standard deviations are very dependent upon the complexity and variations of the environment.

4.1.2 Multipath delay spread

Delayed signals through a channel are frequently described using one of the following three kinds of definitions of delay:

- *Average delay* which describes mean travel time of a signal from a transmitter to a receiver.

- *Delay spread* which is a metric of how much that signal is diluted in time.
- *Maximum* or *total delay spread* which indicates the largest delay due to the multipath.

The delay spread of a UWB multipath channel is usually described by its root mean square (rms) value. Figure 4.2 demonstrates the delay profile, which is the expected power per unit of time received with a certain excess delay and is obtained by averaging a large set of impulse responses. The maximum delay time spread is the total time interval during which reflections with significant energy arrive. The rms delay spread is the standard deviation value of the delay of reflections, weighted proportional to the energy in the reflected waves. For a digital signal with high bit rate this dispersion is experienced as frequency-selective fading and inter-symbol interference (ISI). No serious ISI is likely to occur if symbol duration is longer than, say, ten times the rms delay spread.

Typical values for the rms delay spread for indoor channels have been reported to be between 19 and 47 nsec in [44], and mean values between 20 and 30 nsec for 5 to 30 m antenna separation were reported in [43]. In addition, the multipath delay spread increases with increasing separation distance between receiving and transmitting antennas.

There also exists another relevant parameter, called *Doppler spread*, that tells us how signal energy smears out in frequency in situations where the environment or the transmit/receive antennas move. Doppler spread can be important when the bandwidth of the UWB signal is very large or when the mobile has considerable movement.

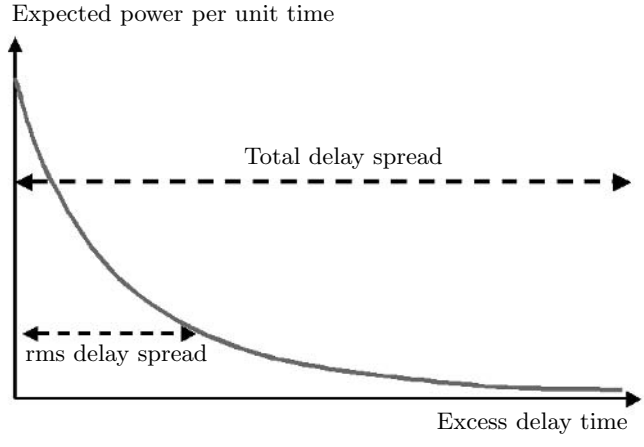


Fig. 4.2 Typical exponential delay profile with total and rms delay spread.

4.1.3 Multipath intensity profile

The parameter rms delay spread T_{rms} is related to the multipath intensity profile in the sense that it represents the standard deviation of it. Therefore, if the form of the multipath intensity profile is known, then there should be a one-to-one relationship between rms delay spread and the multipath intensity profile. Results derived in [43] and many other references suggest that an exponentially (linear in dB) decaying multipath intensity profile is a reasonable model. This means that the average received power for path l can be represented by

$$E[\alpha_l^2] = \Omega_0 e^{-\delta l} \tag{4.2}$$

where Ω_0 is used to normalize the total received power to unity and δ is the decay factor.

The rms delay spread of the channel is utilized to determine the proper values for the two parameters L, the total number of paths and the decay factor δ. Since T_{rms} is simply the standard deviation of the multipath intensity profile, it can be determined in closed form as a function of L and δ. On the other hand, for a given value of T_{rms} there are an infinite number of (L, δ) pairs. In order to come up with a reasonable pair to estimate a channel model for a given T_{rms}, the following method can be used.

Initially, it should be considered that we are primarily interested in multipath components that have a power within 30 dB (0.001) down from the direct path (sometimes called the line-of-sight (LOS) path) component, since it is expected that subsequent paths will have a negligible effect on performance. As a result, we can write

$$e^{-L\delta} < 0.001 \tag{4.3}$$

Hence, for a given value for L we have

$$\delta \approx -\ln(0.001)/L \tag{4.4}$$

Now, we can find the smallest L that results in an actual T_{rms} greater than or equal to the desired T_{rms}, to represent a channel with the greatest decay factor.

4.1.4 Multipath amplitude-fading distribution

Based on the fading statistics described in [43], amplitude fades are best modeled by a log–normal distribution with a standard deviation between 3 and 5 dB for local distributions. In the log–normal distribution the logarithm of the random variable has a normal distribution. The probability density and cumulative distribution functions for the log–normal distribution are

$$P(x) = \frac{1}{Sx\sqrt{2\pi}} e^{-(\ln x - M)^2/(2S^2)} \tag{4.5}$$

and

$$D(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\ln x - M}{S\sqrt{2}}\right)\right] \tag{4.6}$$

respectively, where M and S determine the statistics of the log–normal distribution. The error function erf is defined as

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt \tag{4.7}$$

Note that the channel impulse response measurements described by Hashemi [43] have a resolvability of 5 nsec. Based upon the intuitive understanding that UWB signals will probably experience less fading due to shorter pulse periods, the following results use a 3-dB standard deviation for the log–normal fading. Measurement results presented by Win and Scholtz [45] suggest that the fading is typically less than 5 dB for UWB impulse waveforms, which supports the smaller value for standard deviation. The mean value of log–normal fading is scaled such that it meets the exponentially decaying multipath intensity profile for the given delay spread of the channel.

4.1.5 Multipath arrival times

A simple statistical model for the arrival times of the paths is a Poisson process, since multipath propagation is caused by randomly located objects. Saleh and Valenzuela [46] have compared the path arrival distribution governed by a Poisson hypothesis with the empirical (measured) data to find the degree of closeness. The number of paths l in the first N bins of each measured profile was determined. To determine the empirical path index distribution the probability of receiving l paths in the first N bins $P_N(L = l)$ is plotted as a function of l. This procedure was repeated for different values of the number of bins N.

The probability $P_N(L = l)$ for the theoretical Poisson path index distribution is given by

$$P_N(L = l) = \frac{\mu^l}{l!} \tag{4.8}$$

where l is the path index and μ is the mean path arrival rate given by

$$\mu = \sum_{i=1}^{N} r_i \tag{4.9}$$

where r_i is the path occurrence probability for bin i computed from the empirical data.

Comparison of Poisson and empirical distributions in several papers, such as Suzuki [47], and Ganesh [48], have revealed that there exists a considerable discrepancy between the two. Another model called the modified Poisson model was proposed by Suzuki [47].

The modified Poisson model is a simple and reasonable model for calculating the multipath arrival time of UWB systems. This model is also called the Δ-K model and has two states: state 1 and state 2. Whenever there is an event (i.e., a path occurs) the mean arrival rate, which is the average number of paths arriving per unit time or distance interval, is increased (or decreased if necessary) by a factor K for the next Δ seconds, where K and Δ are parameters to be chosen. The concept of a continuous Poisson model is illustrated in Figure 4.3.

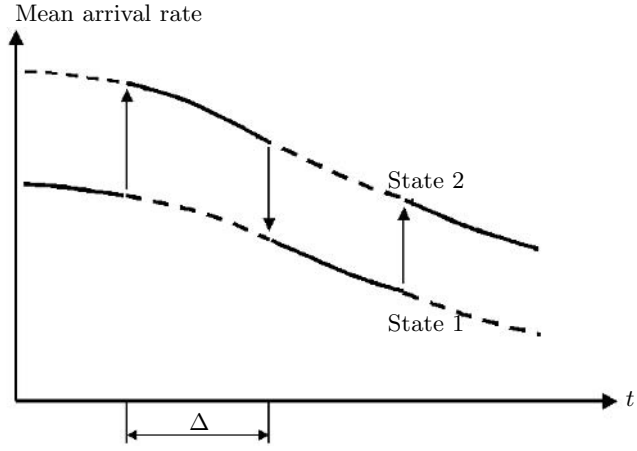


Fig. 4.3 Illustration of the modified Poisson process in the continuous case.

Note that when $K = 1$ or $\Delta = 0$ the process returns to a standard Poisson sequence. For $K > 1$ the probability that there will be another path within the next Δ seconds increases (i.e., the process has a clustering property). For $K < 1$ the incidence of a path decreases the probability of having another path within the next Δ seconds (i.e., events have a tendency to arrive rather more equally spaced than in a pure Poisson model).

The analysis of this statistical model on the discrete delay time axis is rather simple and straightforward. It can be stated as the branching process of Figure 4.4, where λ_i $(i = 1, 2, \dots)$ is the underlying path occupancy rate for the ith bin and, for simplicity, Δ is taken as one 30-m time interval (empirical path occupancy rates are obtained for each 100-ft time interval) and K as a constant. The solid lines are to be traced when there was a path in the previous bin. The process starts at state 1 since the line-of-sight path is not counted as a path. Therefore, the probability to have a path in the first bin is λ_1. If a path does not occur in bin 1,

then the probability of a path in bin 2 is λ_2, and if a path does occur in bin 1, then the probability of a path in bin 2 is $K\lambda_2$. The process proceeds similarly after this. For example, the probability that the process takes the arrowed route up to $N = 3$ is $\lambda_1(1 - K\lambda_2)\lambda_3$.

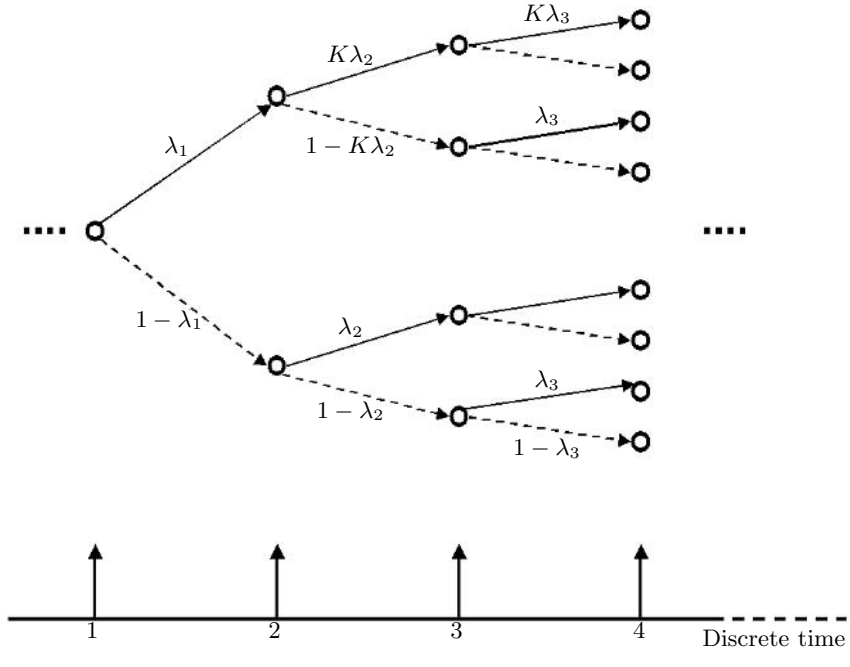


Fig. 4.4 Illustration of the modified Poisson process in the discrete case.

The path number distributions for this statistical model can be obtained, given the real (empirical) path occupancy rate r_i, by the following procedure. First, the underlying path occupancy rate λ_i is calculated from the values of r_i. As we easily see with the help of Figure 4.4, there exists the following relation between r_i and λ_i

$$\lambda_i = \frac{r_i}{(K-1)r_{i-1} + 1}; \quad \text{for } i \neq 1 \tag{4.10}$$

where $\lambda_1 = r_1$.

4.2 PATH LOSS MODEL

By definition, the attenuation undergone by an electromagnetic wave in transit between a transmitter and a receiver in a communication system is called *path loss* or *path attenuation*. Path loss may be due to many effects, such as:

- Free space loss.
- Refraction.
- Reflection.
- Diffraction.
- Clutter.
- Aperture-medium coupling loss.
- Absorption.

Note that path loss is usually expressed in decibels (dB). These effects are explained briefly in the following subsections.

4.2.1 Free space loss

Normally, the major loss of energy is due to the spreading out of the wavefront as it travels away from the transmitter. As the distance increases the area of the wavefront spreads out, much like the beam of a torch. This means the amount of energy contained within any unit of area on the wavefront will decrease as distance increases. By the time the energy arrives at the receiving antenna the wavefront is so spread out that the receiving antenna extends into only a very small fraction of the wavefront.

Free space loss is the signal attenuation that would result if all absorbing, diffracting, obstructing, refracting, scattering, and reflecting influences were sufficiently removed so as to have no effect on propagation.

4.2.2 Refraction

Refraction is defined, in the general case, as redirection of a wavefront passing through (a) a boundary between two dissimilar media or (b) a medium having a refractive index that is a continuous function of position (e.g., a graded index optical fiber). For two media of different refractive indices the angle of refraction is closely approximated by *Snell's law*.

4.2.2.1 Snell's law Snell's law is originally a law of geometric optics that defines the amount of bending that takes place when a light ray strikes a refractive boundary (e.g., an air-glass interface) at a non-normal angle. Snell's law states that

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{4.11}$$

where n_1 is the index of refraction of the medium in which the incident ray travels; θ_1 is the incident angle, with respect to the normal at the refractive boundary, at which the incident ray strikes the boundary; n_2 is the index of refraction of the medium in which the refracted ray travels; and θ_2 is the angle, with respect to the normal at the refractive boundary, at which the refracted ray travels. The incident ray and refracted ray travel in the same plane, on opposite sides of the normal at the point of incidence.

If a ray travels from a medium of lower refractive index into a medium of higher refractive index, it is bent toward the normal; if it travels from a medium of higher refractive index to a medium of lower index, it is bent away from the normal. If the incident ray travels in a medium of higher refractive index toward a medium of lower refractive index at such an angle that Snell's law would call for the sine of the refracted ray to be greater than unity (a mathematical impossibility); i.e.

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 > 1 \tag{4.12}$$

then the *refracted* ray in actuality becomes a *reflected* ray and is totally reflected back into the medium of higher refractive index, at an angle equal to the incident angle (and thus still obeys Snell's law). This reflection occurs even in the absence of a metallic reflective coating (e.g., aluminum or silver). This phenomenon is called *total internal reflection*. The smallest angle of incidence, with respect to the normal at the refractive boundary, which supports total internal reflection, is called the *critical angle*.

4.2.3 Reflection

The abrupt change in direction of a wave front at an interface between two dissimilar media so that the wave front returns into the medium from which it originated is called *reflection*. Reflection may be *specular* (mirror-like) or *diffuse* (i.e., not retaining the image, only the energy) according to the nature of the interface. Depending upon the nature of the interface (i.e., dielectric-conductor or dielectric-dielectric) the phase of the reflected wave may or may not be inverted.

4.2.4 Diffraction

Diffraction is the *spreading out* of waves. All waves tends to spread out at the edges when they pass through a narrow gap or past an object. Instead of saying

that the wave spreads out or bends round a corner we can say that it diffracts around the corner.

A narrow gap is one which is about the same size as the wavelength of the electromagnetic wave or less. The longer the wavelength of a wave the more it will diffract.

4.2.5 Wave clutter

Disorganized wave propagation due to a rough surface or interface is called wave clutter. The mechanisms leading to clutter are not well known so far, and they are a lot more complex compared with the narrowband case. Understanding the phenomenology of electromagnetic interactions between very short pulses and the complex dielectric ground surface provides an important input to the design of UWB systems, leading to improved clutter cancellation and improved detection performance. The roughness considerably influences the spectral content of the response.

4.2.6 Aperture-medium coupling loss

Coupling loss is the loss that occurs when energy is transferred from one medium to another. Aperture-medium coupling loss is the difference between the theoretical gain of the antenna and the gain that can be realized in operation. Aperture-to-medium coupling loss is related to the ratio of *scatter angle* to *antenna beamwidth*.

4.2.7 Absorption

In the transmission of electrical, electromagnetic, or acoustic signals the conversion of the transmitted energy into another form, usually thermal, is called *absorption*. Absorption is one cause of signal attenuation. The conversion takes place as a result of interaction between the incident energy and the material medium, at the molecular or atomic level.

4.2.8 Example of free space path loss model

There have been several proposed path loss models in the literature (e.g., see Cramer [22] and Ghassemzadeh [49]). Assuming perfect isotropically radiating antennas at the transmitter and receiver the received power as a function of frequency can be expressed as [50]

$$P_R(f) = \frac{P_T(f)G_T(f)G_R(f)c^2}{(4\pi d)^2 f^2} \tag{4.13}$$

where $P_T(f)$ is the average transmit power spectral density, c is the speed of light, $G_T(f)$ and $G_R(f)$ are the transmit and receive antenna frequency responses, respectively.

Clearly, Eqn. (4.13) depends on the frequency response of the antennas, which may be difficult to generalize, especially when we are dealing with wideband signals. However, the present regulations for UWB requires the transmitter to meet a certain electric field strength limit at a specified range, which is equivalent to a total, limited transmit power spectral density.

It is desirable to have the product $P_T(f)G_T(f)$ to be flat within the bandwidth of interest. Hence, as a first-order approximation a flat frequency response isotropic antenna is considered. Therefore, for a perfectly flat UWB waveform occupying the band $f_c - W/2$ to $f_c + W/2$ with power spectral density P_{av}/W and a flat frequency response of the receiving antenna with constant gain across the whole bandwidth (G_R), the total average received power at the output of the receiving antenna will be given by the following equation

$$\begin{aligned} P_{\text{Rav}} &= \int_{f_c - W/2}^{f_c + W/2} P_R(f)\, df \\ &= \frac{P_{\text{av}} G_R c^2}{W(4\pi d)^2} \left[\frac{1}{f_c - W/2} - \frac{1}{f_c + W/2} \right] \\ &= \frac{P_{\text{av}} G_R c^2}{W(4\pi d)^2 f_c^2} \left[\frac{1}{1 - (W/2f_c)^2} \right] \end{aligned} \tag{4.14}$$

which can be equally written as

$$P_{\text{Rav}} = P_{\text{av}}^{\text{NB}} \left[\frac{1}{1 - (W/2f_c)^2} \right] \tag{4.15}$$

where

$$P_{\text{av}}^{\text{NB}} = \frac{P_{\text{av}} G_R c^2}{W(4\pi d)^2 f_c^2} \tag{4.16}$$

corresponds to the well-known, narrowband, free space path loss model equation. Hence, the second term indicated in Eqn. (4.15) accounts for the difference between the narrowband and wideband models. For the largest fractional bandwidth allowed by current UWB regulations (occupying 3.1–10.6 GHz), P_{Rav} will differ from $P_{\text{av}}^{\text{NB}}$ by only 1.5 dB, and this difference becomes smaller for smaller fractional bandwidths. Also note that FCC rules result in $W < 2fc$, so the singularity in the above equation can be ignored at $W = 2fc$.

Alternatively, we can repeat the above analysis for a receiver antenna response of the following form

$$G_R(f) = \frac{4\pi A_R f^2}{c^2} \tag{4.17}$$

where A_R is the effective area of the antenna (e.g., the antenna has a fixed effective aperture). This type of response yields a greater gain for higher frequencies. In this case the above analysis results in the average received power given by

$$\begin{aligned} P_{\text{Rav}} &= \frac{P_{\text{av}} 4\pi A_R}{(4\pi d)^2} \\ &= \left[\frac{P_{\text{av}} c^2}{(4\pi d)^2 f_c^2}\right]\left[\frac{4\pi A_R f_c^2}{c^2}\right] \\ &= P_{\text{av}}^{\text{NB}} G_R(f_c) \end{aligned} \tag{4.18}$$

where $G_R(f_c)$ is the antenna gain at the center frequency of the transmitted waveform.

In conclusion, it appears that the narrowband model can be used to approximate the path loss for a UWB system, based on the assumptions discussed above.

4.3 TWO-RAY UWB PROPAGATION MODEL

Path loss is considered to be a fundamental parameter in channel modeling as it plays a key role in link budget analysis. Also, path loss serves as an input for the mean value of large-scale fading, which in turn determines the small-scale fading characteristics. In many channel models, path loss is modeled by adopting a power law dependence with distance from the transmitter $L_p = ad^{\gamma}$, which is empirically extracted from data collected by measurement campaigns. In some cases a simple free space law, where $\gamma = 2$, with additional losses is adopted [51]. Otherwise, the values of the exponent γ and the coefficient a of the mean path loss are either obtained through linear regression to fit empirical data [52] or statistically characterized by means of probability distributions [49]. A random variable with log–normal distribution is usually added to the mean path loss that takes into account variations associated with the shadowing phenomenon, in order to model the fading.

Path loss frequency dependence has sometimes been excluded in UWB propagation channel modeling, and, though its impact is negligible over the bandwidths of current wireless systems, the same assumption cannot be applied to UWB systems. This section describes a theoretical study of path loss over short ranges. A rigorous formulation of a two-ray link is considered and evaluated as a function of both frequency and distance, where all the assumptions adopted for deriving the traditional narrowband plane earth model are no longer valid. Considering the specific UWB application the analysis is carried out over distances of up to 10 meters and in the operational frequency band of 3 GHz to 10 GHz.

From this investigation an analytical path loss model for short-range, two-ray links with wide operational bandwidths is derived, that incorporates the frequency dependence, however, does not consider the effect of the transmitter and receiver

antennas. This simple model is applicable to line-of-sight UWB transmissions. The effect of path loss frequency selectivity and its impact on UWB pulse shape are also demonstrated. The analysis is performed by modeling the path loss as a low-pass filter and applying the associated impulse response to the UWB pulse signal. The observations confirm that path loss frequency dependence cannot be neglected in UWB systems, due to the extremely large bandwidth of the involved signals. Thus, the following issues [53] are emphasized in this section:

- Path loss frequency dependence.
- Observing a breakpoint at short distances of about 3 m.
- Filtering effect of path loss on UWB transmissions.

4.3.1 Two-ray path loss

A two-ray channel model is widely considered to be a good approximation for line-of-sight UWB links operating in a relatively clutter-free environment. According to the classical definition of two-ray propagation, the direct ray and the ray reflected by the ground are considered, as illustrated in Figure 4.5.

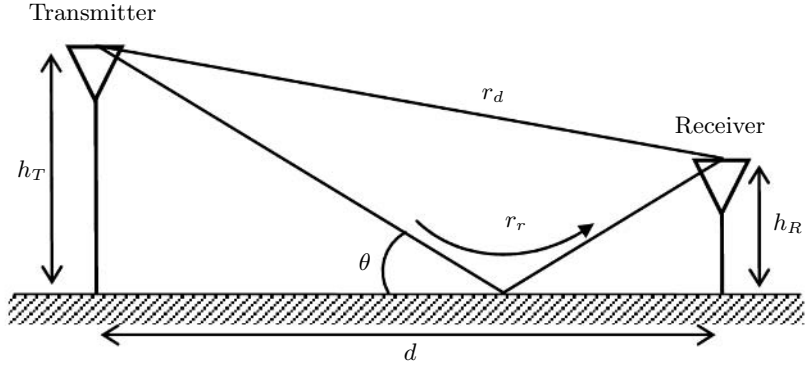


Fig. 4.5 Geometry of the two-ray model including a transmitter and a receiver.

The ground is regarded as a plane surface at this stage; thus, specular reflection is assumed. Indeed, the degree of roughness of a surface is frequency-dependent and results in angular spread of the reflected field. Its effect on UWB radio transmissions is a topic of research and needs to be incorporated into subsequent models.

Considering that the field at the receiver as the superposition of the contributions associated with the two rays the path loss can be expressed as $L_p = (G_p)^{-1}$,

where the path gain G_p is defined as

$$G_p = \left(\frac{\lambda}{4\pi}\right)^2 \left| \frac{e^{-jkr_d}}{r_d} + \frac{R_{H,V}e^{-jkr_r}}{r_r} \right|^2 \tag{4.19}$$

and $k = 2\pi/\lambda$ is the free space propagation constant, $\lambda = c/f$ is the free space wavelength, and c is the speed of light. Moreover, as shown in Figure 4.5

$$r_d = \sqrt{d^2 + (h_T - h_R)^2} = d\sqrt{1 + \left(\frac{h_T - h_R}{d}\right)^2} \tag{4.20}$$

and

$$r_r = \sqrt{d^2 + (h_T + h_R)^2} = d\sqrt{1 + \left(\frac{h_T + h_R}{d}\right)^2} \tag{4.21}$$

are the lengths of the direct path and of the reflected path, respectively. In Eqn. (4.19), $R_{H,V}$ indicates the Fresnel reflection coefficient for the horizontal or vertical polarization, defined as

$$R_H = \frac{\sin\theta - \sqrt{\varepsilon_r - \cos^2\theta}}{\sin\theta + \sqrt{\varepsilon_r - \cos^2\theta}} \tag{4.22}$$

and

$$R_V = \frac{\varepsilon_r \sin\theta - \sqrt{\varepsilon_r - \cos^2\theta}}{\varepsilon_r \sin\theta + \sqrt{\varepsilon_r - \cos^2\theta}} \tag{4.23}$$

where

$$\varepsilon_r(f) = \varepsilon_r'(f) - j\frac{\sigma(f)}{2\pi f \varepsilon_0} \tag{4.24}$$

is the dielectric constant of the reflecting surface. The frequency dependence of the relative permittivity $\varepsilon_r'(f)$ and the conductivity $\sigma(f)$ must be taken into account to appropriately characterize reflection and transmission mechanisms over the UWB frequency range.

In narrowband outdoor and indoor modeling, construction material properties have been defined for a fixed operating frequency. A database, obtained from power loss measurements in the frequency range 3 GHz–8 GHz, has been presented by Stone [54] where power attenuation through a set of construction materials with different thicknesses is described as a function of frequency. However, numerical estimates on dielectric constant and permittivity have not been provided as yet. Hence, at the present stage we consider constant values of $\varepsilon_r'(f) = \varepsilon_r'$ and $\sigma(f) = \sigma$.

The traditional plane earth model, as described by Rappaport [55], which is used in classical narrowband transmissions at long distances, and even in microcells

and picocells, is *not* applicable to short ranges, since the two hypotheses on which its derivation is based, namely that, the distance between transmitter and receiver is much greater than the antenna heights $d \gg h_T$ and $d \gg 4h_T h_R/\lambda$, are not satisfied. This implies that all the approximations held in the classical two-ray model [55] (for example, reflection angle $\theta \cong 0$ and reflection coefficient $|R_{H,V}| \cong 1$) are no longer valid and, obviously, the well-known fourth-order power dependence on distance does not hold for distances below the breakpoint $4h_T h_R/\lambda$. Thus, a rigorous analytical evaluation is required for short ranges, when distance d is comparable with h_T and h_R. By replacing Eqns. (4.20) and (4.21) in Eqn. (4.19), the path gain can be written as

$$G_p(d,f) = G_{FS}(d,f)\left|\frac{e^{-jkr_d}}{\sqrt{1+((h_T-h_R)/d)^2}} + \frac{R_{H,V}e^{-jkr_r}}{\sqrt{1+((h_T+h_R)/d)^2}}\right|^2 \quad (4.25)$$

where

$$G_{FS}(d,f) = \left(\frac{c}{4\pi df}\right)^2 \quad (4.26)$$

is the classical free space path gain depending on the transmitter-receiver distance d. Note that for short ranges the free space transmission distance r_d cannot in general be approximated with d, depending on the ratio $(h_T - h_R)/d$, as apparent in Eqn. (4.20).

Path gain in Eqn. (4.25) is clearly a function of both distance and frequency, as shown in Figure 4.6, where Eqn. (4.25) is evaluated in the case of vertical polarization of the field and total reflection from the ground ($\sigma \to \infty, R_V = 1$) and plotted in decibels, $L_{p,\mathrm{dB}} = -10\log(G_p)$, versus both distance and frequency.

The two-ray path loss clearly exhibits fading around a mean value due to the variation in the phase of the two-ray contributions. Therefore, mean path loss can be modeled as

$$\bar{L}_p(d,f) = \alpha d^{\gamma(f)} f^{\nu(d)} \quad (4.27)$$

where $d_{\min} \le d \le d_{\max}$ and $f_{\min} \le f \le f_{\max}$. The corresponding expression in decibels is written as

$$\bar{L}_{p_{\mathrm{dB}}}(d,f) = \alpha_{\mathrm{dB}} + \gamma(f)10\log(d) + \nu(d)10\log(f) \quad (4.28)$$

where $\gamma(f)$ and $\nu(d)$ turn out to be the slope coefficients and are functions of frequency and distance, respectively. In particular, at each frequency point f^*, the slope coefficient $\gamma(f^*)$ is determined by linear fitting of the decibel path loss curve, which results in a function of distance d only (i.e., $L_{p_{\mathrm{dB}}}(d)$ at $f = f^*$). Likewise, at each distance point d^* the value $\nu(d^*)$ is determined by linear fitting the curve $L_{p_{\mathrm{dB}}}(f)$ at $d = d^*$.

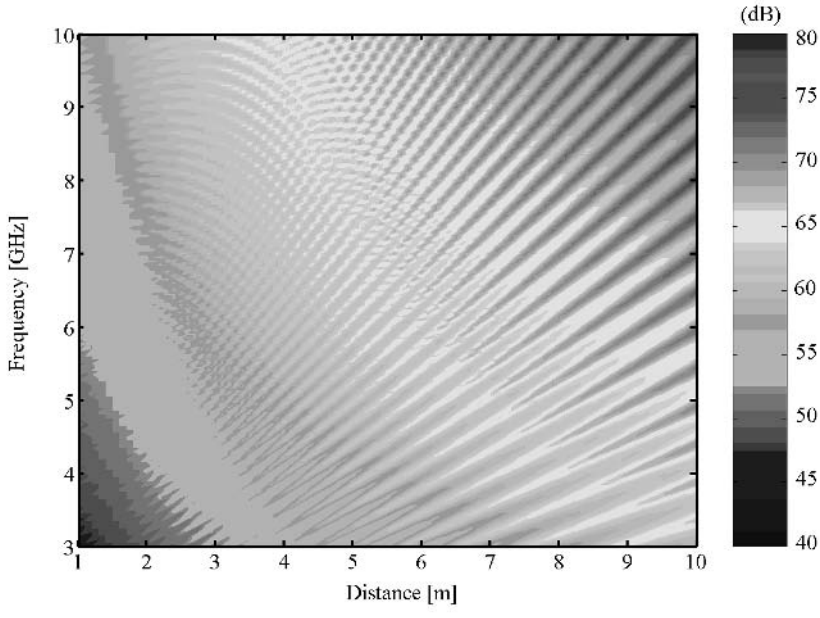


Fig. 4.6 Path loss versus distance and frequency: $h_T = 2.5$ m, $h_R = 1.2$ m, and $R_V = 1$.

4.3.2 Two-ray path loss model

An analysis of Eqn. (4.28) as a function of distance and frequency enables a power law path loss model to be derived:

$$\bar{L}_{p_{\mathrm{dB}}}(d,f)\Big|_{\substack{f_{\min}\le f\le f_{\max}\\ d_{\min}\le d\le d_{\max}}} = L_{p_0} + \gamma 10\log\left(\frac{d}{d_0}\right)\Big|_{f_{\min}\le f\le f_{\max}} + \nu 10\log\left(\frac{f}{f_0}\right)\Big|_{d_{\min}\le d\le d_{\max}} \tag{4.29}$$

where the slope coefficients γ and ν are constant values and

$$L_{p_0} = 10\log(L_{FS}(r_{d_0}, f_0)) \tag{4.30}$$

is the free space path loss at $r_{d_0} = r_d(d_0)$, where d_0 is a reference distance, and f_0 is a reference frequency. The values of γ and ν are obtained by averaging $\gamma(f)$ and $\nu(d)$ in the ranges $f_{\min} \le f \le f_{\max}$ and $d_{\min} \le d \le d_{\max}$.

In particular, in the distance range d from 1 m to 10 m and in the frequency range of f between 3 GHz and 10 GHz the path loss model exhibits the following features:

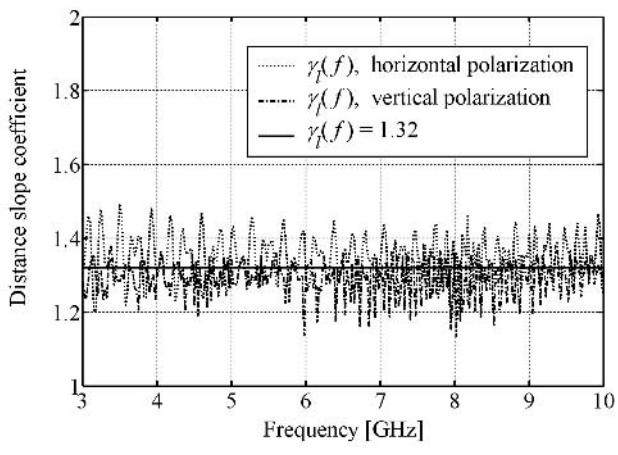
- **Distance dependence:** *Path loss has dual-slope behavior with respect to distance*, with a new breakpoint at a very short distance from the transmitter (a few meters). The exact position of breakpoint d_{BP} occurs at a closer distance from the transmitter as difference $\Delta h = h_T - h_R$ is reduced. The breakpoint defines two subranges 1 m$\leq d_1 \leq d_{BP}$ and $d_{BP} \leq d \leq 10$ m, over which $\gamma(f)$ is evaluated. The values $\gamma = \gamma_l$ and $\gamma = \gamma_u$ for use in Eqn. (4.29) are obtained as the mean value of $\gamma(f)$. Note that this breakpoint is valid at short distances and for operation over an extremely wide frequency range. Moreover, it should be emphasized that it is different from the breakpoint of the traditional plane earth model $4h_T h_R/\lambda$, which is clearly frequency-dependent and is located outside the maximum distance considered in this analysis and, therefore, outside of the expected operating range of UWB transmissions.

- **Frequency dependence:** *Path loss shows a $\nu = 2$ power law frequency dependence*, which is the same as the frequency dependence of free space transmissions (i.e., only a single ray). This result is due to the fact that, at this first stage of the study, the frequency dependence of permittivity, conductivity, and surface roughness have not been taken into account. The frequency dependence of the material properties and variations of ν are expected, confirming the frequency dependence of the clutter impact over the UWB frequency range.

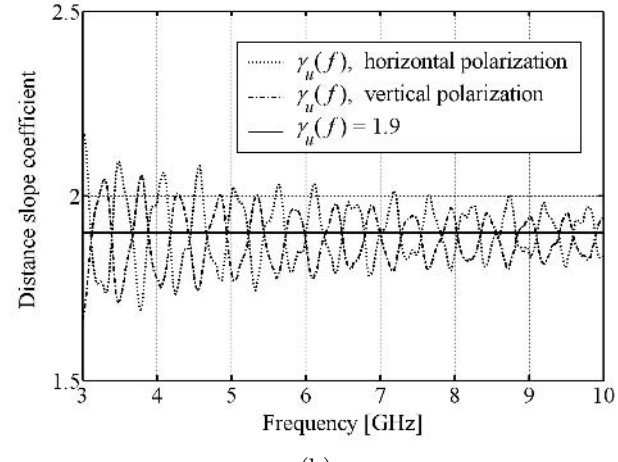
The model in Eqn. (4.29) is valid for both horizontal and vertical polarizations, as the two polarizations show the same mean path loss tendency. In the particular case of very conductive surfaces (i.e., reflection coefficient's absolute value is close to 1) the oscillations around the mean value have the same amplitude for both polarizations, but are opposite in phase; averaging the two path loss curves gives a fade-free path loss curve that is in close agreement with the power law path loss model.

Some results are shown in Figures 4.7 to 4.10, where the transmitter height is $h_T = 2.5$ m and the receiver height is $h_R = 1.2$ m. In this particular case the new short-distance breakpoint is located at $d_{BP} = 3.5$ m from the transmitter. Note that, with this choice of h_T and h_R, the location of the traditional breakpoint $4h_T h_R/\lambda$ from which a fourth power dependence is assumed in the classical plane earth model, would range between 120 m at 3 GHz and 400 m at 10 GHz. Relative permittivity $\varepsilon_r' = 6$ and conductivity $\sigma = 0.0166$ are assumed when computing the plane reflection coefficient Eqn. (4.24) and model the dielectric constant of a cement surface.

Figure 4.7 shows the plots of $\gamma(f)$ that result from evaluation over lower and upper distance subranges separated by the breakpoint: 1 m $\leq d_1 \leq$ 3.5 m in Figure 4.7(a) and 3.5 m $\leq d_u \leq$ 10 m in Figure 4.7(b). The values of γ are $\gamma_l = 1.32$ and $\gamma_u = 1.9$, respectively. Note that the values of the slope coefficient resemble those provided in previous literature [52], [49] for line-of-sight UWB transmissions,



(a)



(b)

Fig. 4.7 Path loss distance slope coefficient $\gamma(f)$ and mean value γ, 3 GHz $\leq f \leq$ 10 GHz, $h_T = 2.5$ m, $h_R = 1.2$ m: (a) 1 m $\leq d \leq$ 3.5 m, $\gamma_l = 1.32$; (b) 3.5 m $\leq d \leq$ 10 m, $\gamma_u = 1.9$.

but where the breakpoint is at very short distance from the transmitter values have not been included. Similarly, in Figure 4.8 the slope coefficient $\nu(d)$ is evaluated over the entire frequency range 3 GHz $\leq f \leq$ 10 GHz and plotted versus distance; its mean value is $\nu = 2$.

In Figure 4.9, path loss is plotted in decibels versus the distance d at the operational frequency $f = 5$ GHz. Note that this case does not show an ultra

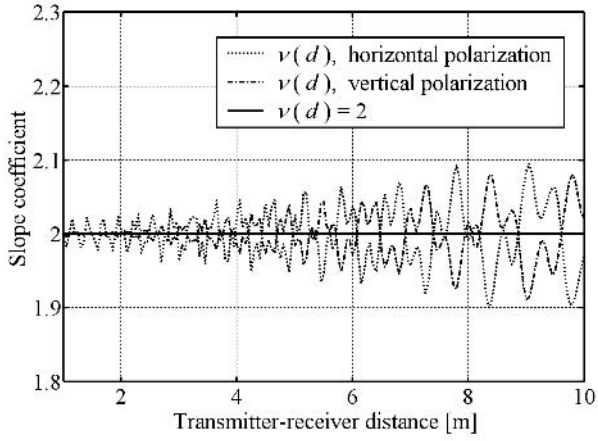


Fig. 4.8 Path loss frequency slope coefficient $\nu(d)$ and mean value $\nu = 2$ for 1 m $\leq d \leq$ 10 m, $h_T = 2.5$ m, and $h_R = 1.2$ m.

wideband situation. The two curves exhibiting fading are obtained by evaluating the rigorous analytical formulation of the two-ray path loss, and each one is related to a polarization. The solid line corresponds to the path loss model, with the breakpoint at $d_{BP} = 3.5$ m, and the dotted line is the free space path loss.

In Figure 4.10 the path loss evaluated at a distance $d = 2$ m is plotted versus frequency. Again, path loss curves for both horizontal and vertical polarizations are plotted: the solid line is the result obtained through our path loss model in Eqn. (4.29), with $\nu = 2$, and the dotted line, which is 3 dB below, is the free space path loss evaluated at the same distance as a variable of frequency.

4.3.3 Impact of path loss frequency selectivity on UWB transmission

UWB systems are clearly different from classical modulated or passband radio systems. Likewise, UWB channel modeling must consider the true nature of the UWB propagation channel. In particular, it is well known that the frequency selectivity of propagation mechanisms leads to distortion of signals arriving at the receiver. It is therefore necessary to investigate the impact on inter-symbol interference (ISI) and bit error rate (BER) performance of the system under such conditions. Provided the frequency response of the channel is known it can be combined with the frequency spectrum of the UWB pulse and the distortion level of the received signal can be evaluated. The frequency selectivity of the path loss has been neglected in link-level investigations of wireless systems until now.

In the following, a $\nu = 2$ frequency power law (i.e., free space) is assumed for the path loss and is in line with the observations made before, which neglect antenna effects. This produces a low-pass filtering effect on the transmitted signal

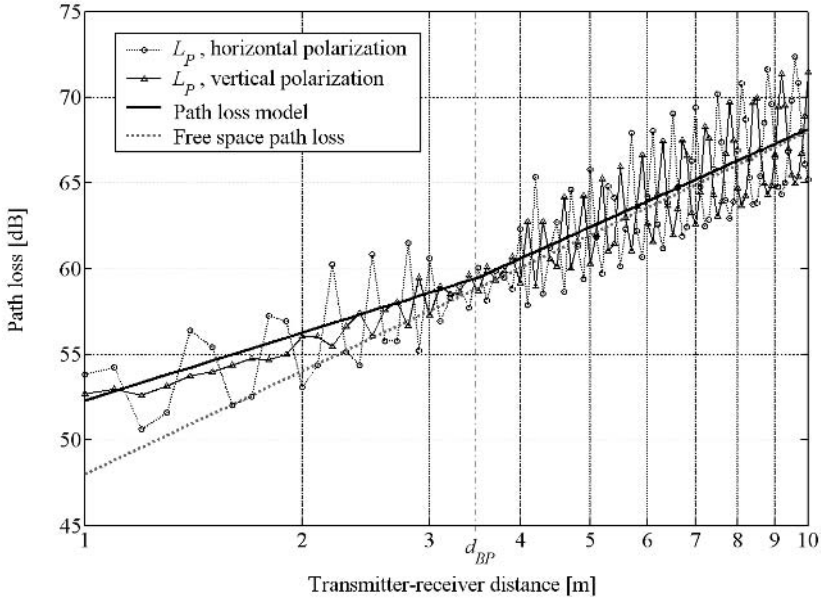


Fig. 4.9 Path loss versus distance for 1 m $\leq d \leq$ 10 m, $f = 5$ GHz , $h_T = 2.5$ m, and $h_R = 1.2$ m.

(see Figure 4.10). In fact, it can be modeled as a transfer function in the frequency domain by considering the path gain as:

$$G_p = |H_p(f)|^2 \tag{4.31}$$

Hence, from Eqn. (4.19), when assuming no reflected ray, the following transfer function can be defined:

$$H_p(f) = K_H \frac{c}{2d} \frac{e^{-j(2\pi/c)\,df}}{2\pi f} \tag{4.32}$$

where d is the distance between the transmitter and the receiver and K_H is the normalization constant necessary to meet the requirements on the energy associated with the filter.

By applying Fourier theory the corresponding path gain impulse response in the time domain is determined

$$h_p(t) = j\frac{H_0}{2}\,\mathrm{sign}(t - t_d) \tag{4.33}$$

where

$$H_0 = K_H \frac{c}{4d} \tag{4.34}$$

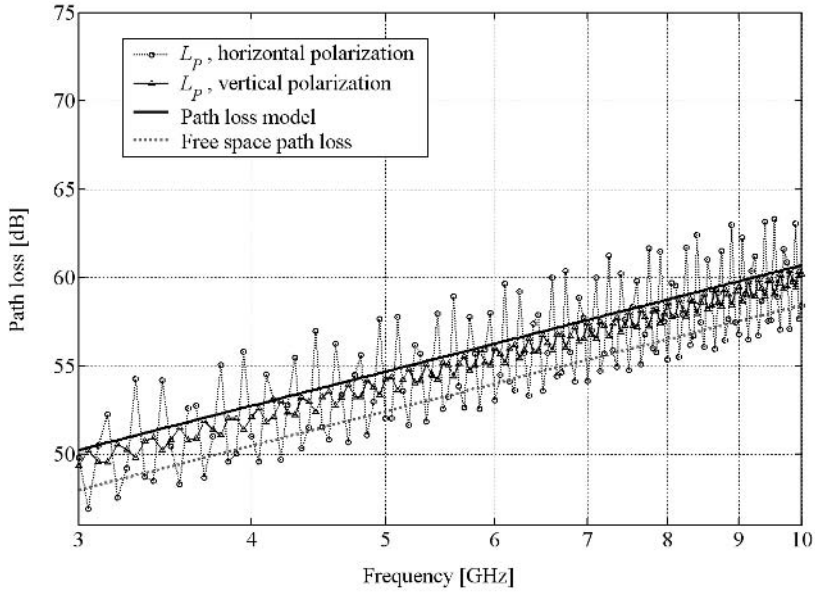


Fig. 4.10 Path loss versus frequency for 3 GHz $\leq f \leq$ 10 GHz, $d = 2$ m, $h_T = 2.5$ m and $h_R = 1.2$ m.

and $t_d = 2\pi/c$ is the propagation time delay between transmitter and receiver. The normalized impulse response $h_p(t)/(jH_0/2)$ is plotted in Figure 4.11, for a $d = 2$-m link that gives $t_d = 6.67$ ns.

By convoluting the path gain impulse response $h_p(t)$ with the transmitted signal $s(t)$, we obtain the signal at the receiver:

$$\begin{aligned} w(t) &= s(t) * h_p(t) \\ &= jH_0 \int s(t)\, dt \end{aligned} \tag{4.35}$$

thus, we can conclude that free space propagation produces an integration effect on a signal occupying a wide frequency spectrum. As an example, consider a typical UWB signal pulse $s(t)$ that is the second derivative of a Gaussian waveform

$$s(t) = K_s \frac{2}{\tau^2} \left(\frac{t^2}{\tau^2} - 1 \right) e^{-[t/\tau]^2} \tag{4.36}$$

and apply $h_p(t)$; as a result, the signal collected at a receiver located at $d = 2$ m from the transmitter is

$$w(t) = -j2\pi H_0 K_s \frac{t - t_d}{\tau} e^{-[(t-t_d)/\tau]^2} \tag{4.37}$$

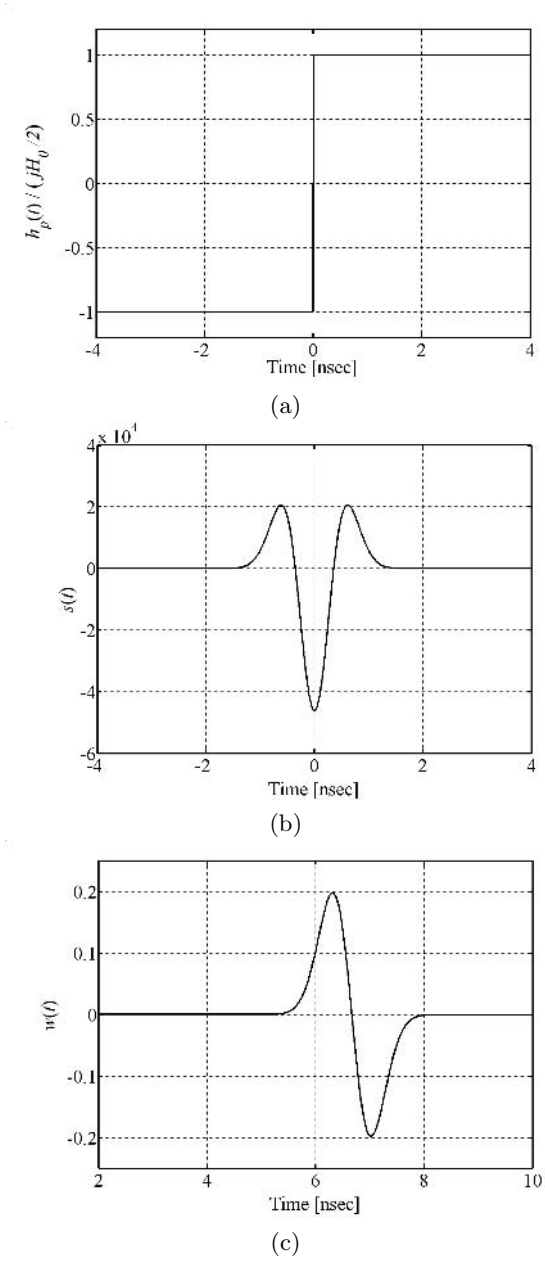


Fig. 4.11 Impact of path loss frequency selectivity on UWB signal waveforms: (a) normalized impulse response; (b) transmitted pulse waveform; (c) received pulse waveform.

Figures 4.11(b) and (c) show the waveforms defined in Eqns. (4.36) and (4.37), respectively, with $\tau = 0.5$ ns.

In conclusion, neglecting the frequency dependence of path loss in UWB propagation channel models seems to be an unsuitable choice, as the filtering effect on the UWB pulse shape is not considered, with possible consequences on the development of an effective UWB radio system. For example, the knowledge of the distortion to which the pulse is exposed would help in the choice of a suitable correlator when designing a UWB coherent receiver. Note, at this stage observation has been limited to the propagation channel (i.e., path loss) and antennas have not been included in the analysis. Nevertheless, effective modeling will also consider the pass-band behavior of transmitting and receiving antennas, which will contribute to the overall channel-filtering effect.

4.4 FREQUENCY DOMAIN AUTOREGRESSIVE MODEL

In this section an autoregressive (AR) frequency domain statistical model for UWB indoor radio propagation is described. Using AR modeling techniques the parameters of the channel model can be determined from the measured frequency responses. A frequency domain channel-sounding experiment is usually performed for this purpose.

By definition an AR model depends only on the previous outputs of the system. The transfer function representation of such a model can be written as follows

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} \\ &= \frac{b_0}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots} \end{aligned} \tag{4.38}$$

where $X(z)$ and $Y(z)$ denote the z-transforms of the input and output of the AR system, respectively. Equivalently, in discrete time domain we can write Eqn. (4.38) as follows

$$y(n) = b_0 x(n) + a_1 y(n-1) + a_2 y(n-2) + \cdots \tag{4.39}$$

The target of a frequency domain UWB channel model is to develop a statistical representation of the channel with a minimum number of parameters to regenerate the measured channel behavior accurately in computer simulations. The higher the complexity of the model the closer the statistical resemblance to that of measured data [56].

The main advantage of the frequency domain modeling of systems over time domain models is that it uses fewer parameters. On the other hand, it requires a complete characterization of the probability distributions of its parameters.

The frequency response of a system can be interpreted as the output of an AR model [57]. Autoregressive modeling of time domain data used for spectral

estimation and the techniques for determination of the coefficients or multipliers of the AR process are well known in the literature [58]. These results can be used for AR modeling of the frequency response observed from propagation measurements.

With the AR process assumption the frequency response at each location is a realization of an AR process of order p given by the following equation

$$H(f_n, x) - \sum_{i=1}^{p} a_i H(f_{n-1}, x) = V(f_n) \tag{4.40}$$

where $H(f_n, x)$ is the nth sample of the complex frequency domain measurement at location x and $\{V(f_n)\}$ is a complex white noise process. The parameters of the model are the complex constants a_i. Taking the z-transform of Eqn. (4.40), we can view the AR process $\{H(f_n, x)\}$ as the output of a linear filter with transfer function

$$\begin{aligned} G(z) &= \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \\ &= \prod_{i=1}^{p} \frac{1}{1 - p_i z^{-i}} \end{aligned} \tag{4.41}$$

driven by a zero-mean white Gaussian noise process $\{V(f_n)\}$. Using the AR model the channel frequency response can be identified with the p parameters of the AR model or the location of the p poles of the transfer function $G(z)$.

An AR model can be implemented using an infinite impulse response (IIR) filter as shown in Figure 4.12. In a simple AR model with a second-order IIR filter [56], the frequency response model has four complex parameters and one real parameter which is the noise standard deviation σ. Thus, the model is characterized by nine real parameters. For a model at a particular location with a transmitter-receiver separation of 0.6 m the estimated parameters are given in [56].

The agreement between the probability distribution of the model parameters using experimental data and the normal distribution is also shown for a particular location. Using this model the channel frequency response has been simulated as well. To verify model accuracy a comparison was made between the modeled probability density and that of the measured data. The model has been reported to reproduce the frequency selectivity and the multipath propagation characteristics observed in actual measurements.

4.4.1 Poles of the AR model

The AR model can be represented by the poles of its transfer function given by Eqn. (4.41)

$$1 - \sum_{i=1}^{p} a_i z^{-i} = \prod_{i=1}^{p} (1 - p_i z^{-i}) \tag{4.42}$$

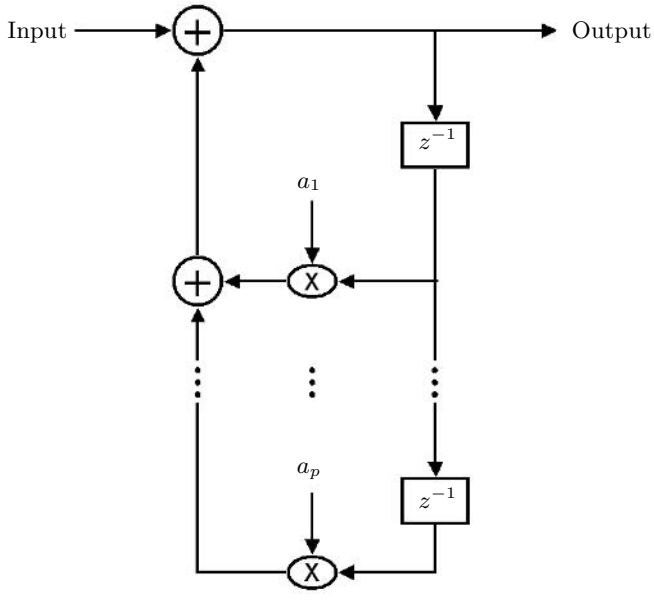


Fig. 4.12 Implementation of an AR model using an IIR formulation.

It has been demonstrated that the largest pole is very close to the unit circle. A pole close to the unit circle represents significant power at the delay related to the angle of the pole. The arrival times of significant paths in all measurements are from 3 to 74 ns. The delay is calculated as

$$\tau = \frac{\arg(p_i)}{2\pi f_s} \tag{4.43}$$

where f_s is sampling frequency. Experimenting with higher order models, it was observed that two significant poles exist in the range and this is adequate to represent the in-home UWB channel. The two significant poles can be interpreted as two significant clusters of multipath arrivals. The interpretation of a pole as defining a cluster of paths and the distance of the pole from the unit circle as defining the power in the cluster provides a useful physical interpretation of the AR model [57]. In contrast to other wideband systems the UWB channel represents itself as two pole clusters with a distinctive angular spread.

4.5 SUMMARY

In this chapter the propagation aspects of UWB signals in indoor and indoor-outdoor environments were discussed. We started with a simplified multipath

channel model that was not unique to UWB systems, but could be an appropriate approach. The primary parameters that are important to characterize the indoor channel, including the number of resolvable multipath components, multipath delay spread, multipath intensity profile, multipath amplitude-fading distribution, and multipath arrival times were briefly described.

As a second approach to UWB channel characterization we considered the path loss model. Various effects and phenomena involved, such as free space loss, refraction, reflection, diffraction, clutter, aperture-medium coupling loss, and absorption, were defined and explained. An example of a free space path loss consideration was also investigated and a simple two-ray UWB propagation mechanism was thoroughly studied.

Finally, the statistical frequency domain autoregressive model was explained. The target of a frequency domain model is to develop a statistical representation of the UWB channel with a minimum number of parameters to regenerate the measured channel behavior accurately in computer simulations. The higher the complexity of the model the closer the statistical resemblance to that of measured data.

Problems

Problem 1. What are the important parameters involved in multipath channel characterization of UWB systems? Briefly describe each one.

Problem 2. Discuss the importance of Doppler spread in UWB channel modeling.

Problem 3. Why can use of the Poisson process be a proper way for statistically modeling the arrival times of different paths?

Problem 4. Investigate the main differences between the path loss models proposed by Cramer [22] and Ghassemzadeh [49].

Problem 5. Derive Eqn. (4.25).

Problem 6. Describe the distance and frequency dependence behavior of the two-ray path loss model using Eqn. (4.29).

Problem 7. Explain how the transmitted and received signals of Figures 4.11(b) and (c) are related.

Problem 8. What is an AR model?

Problem 9. It has been shown that the largest pole of an autoregressive UWB model is very close to the unit circle. How does this property affect the performance of the system?

5

Ultra wideband communications

In this chapter we will look at the use of ultra wideband wireless communications. From the treatment of individual pulse shaping and generation which was introduced in Chapter 2, we now move on to examine various communications concepts. Particular attention will be paid to modulation methods including pulse position modulation, bi-phase modulation, orthogonal pulse modulation, and their combinations. Sequences of individual pulses onto pulse streams will be presented. Receiver design and pulse detection will be examined.

We also move from a single-user environment to examine multiple access techniques for UWB communications. The capacity of the wireless UWB channel will also be examined. The effect of UWB on existing wireless communication methods, such as the IEEE 802.11 wireless local area network standards and Bluetooth, is shown. Several methods to prevent interference from these narrowband systems to UWB will also be examined.

Finally, an important discussion on the relative merits and demerits of UWB as a communication method with other wideband communication techniques, such as CDMA and OFDM, is undertaken.

5.1 INTRODUCTION

Communication can generally be defined as the transmission of information from a source to a recipient. In this chapter we make our definition of communication much narrower, by restricting ourselves to wireless communication of digital data

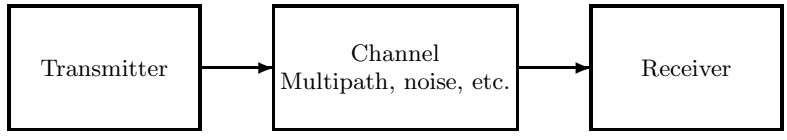


Fig. 5.1 Model of a general communications system.

streams of information using extremely short pulses. We deliberately ignore the kind of information contained in the digital data stream and do not use *media access control* (MAC) protocols, coding schemes, or retransmission schemes to reduce errors. We concentrate on what is commonly known as the *physical layer* in the international standards organization (ISO) protocol stack.

A general model of a communication system is shown in Figure 5.1. The three basic elements are as follows:

- The *transmitter*, whose primary task is to group the digital data stream into symbols, to map these symbols onto an analog waveform, and then to transmit them to the air through an antenna.
- The *channel*, which represents the effect of traveling through space, including reflections and distortions as the electromagnetic pulses impinge on other objects.
- The *receiver*, which collects the electromagnetic energy from the antenna, takes the extremely weak signal, reconstructs the pulse shape, and maps it to the appropriate symbols and then to the binary bitstream.

In this chapter we examine receiver and transmitter structures in more detail, focusing on the basic communication aspects, such as modulation. Channel effects have already be covered in Chapter 4.

5.2 UWB MODULATION METHODS

As we saw in Chapter 1, one single UWB pulse does not contain information of itself. We must add digital information to the analog pulse, by means of *modulation*. In UWB systems there are several basic methods of modulation, and we examine each in detail.

As a helpful categorization of modulation methods, we define two basic types of modulation method for UWB communication. These are shown in Figure 5.2 as *time-based* techniques and *shape-based* techniques.

By far the most common method of modulation in the literature is *pulse position modulation* (PPM) where each pulse is delayed or sent in advance of a regular time

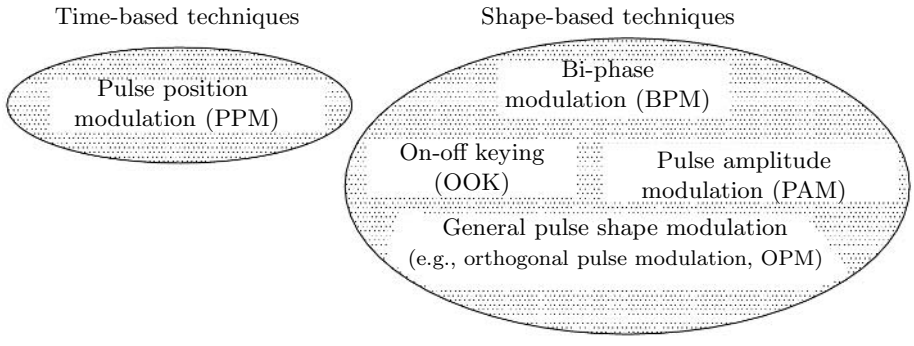


Fig. 5.2 Division of different modulation methods for UWB communications.

scale. Thus, a binary communication system can be established with a forward or backward shift in time. By specifying specific time delays for each pulse, an M-ary system can be created.

Another common method of modulation is to invert the pulse: that is, to create a pulse with opposite phase. This is known as *bi-phase modulation* (BPM).

An interesting modulation technique is *orthogonal pulse modulation*, which requires special pulse shapes to be generated which are orthogonal to each other.

Other well-known techniques for modulation are available. For example, *on-off keying* (OOK) is a modulation technique where the absence or presence of a pulse signifies the digital information of "0" or "1", respectively. *Pulse amplitude modulation* (PAM) is a technique where the amplitude of the pulse varies to contain digital information.

Furthermore, some traditional modulation techniques are not available to us. For example, the widely used *frequency modulation* (FM) is difficult to apply to UWB, since each pulse contains many frequency elements making it difficult to modulate. Note that this should not be confused with *frequency division multiplexing* (FDM) which is an entirely different technique to separate communication channels based on larger blocks of frequency (discussed later).

Let us examine each of these possible modulation techniques in turn. First, we examine the two most common techniques: PPM and BPM. A simple comparison of these two modulation methods is shown in Figure 5.3. In Figure 5.3(a) an unmodulated pulse train is shown for comparison. As an example for PPM, the pulse representing the information "1" is delayed in time (i.e., the pulse appears to be moved in position to the right). The pulse representing the information "0" is sent before the nonmodulated pulse (i.e., the pulse appears to be moved in position to the left) in Figure 5.3(b). For BPM the inverted pulse represents a "0" while the uninverted pulse represents a "1". This is clearly illustrated in Figure 5.3(c).

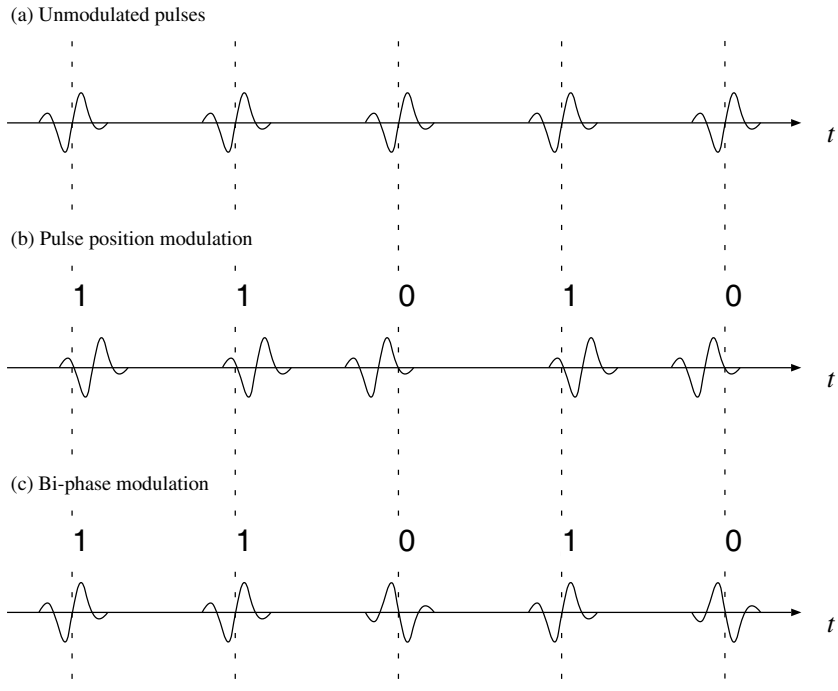


Fig. 5.3 Comparison of pulse position modulation and bi-phase modulation methods for UWB communication.

5.2.1 Pulse position modulation

As mentioned previously, the important parameter in pulse position modulation is the delay of the pulse. That is, by defining a basis pulse with arbitrary shape $p(t)$, we can modulate the data by the delay parameter τ_i to create pulses s_i as shown in Eqn. (5.1), where t represents time

$$s_i = p(t - \tau_i) \tag{5.1}$$

As an example we can let $\tau_1 = -0.75$, $\tau_2 = -0.25$, $\tau_3 = 0.25$, and $\tau_4 = 0.75$, to create a 4-ary PPM system. The four pulse shapes become

$$\begin{aligned} s_1 &= p(t + 0.75) \\ s_2 &= p(t + 0.25) \\ s_3 &= p(t - 0.25) \\ s_4 &= p(t - 0.75) \end{aligned} \tag{5.2}$$

The advantages of PPM mainly arise from its simplicity and the ease with which the delay may be controlled. On the other hand, for the UWB system extremely fine time control is necessary to modulate pulses to subnanosecond accuracy.

5.2.2 Bi-phase modulation

Bi-phase modulation can be defined as a kind of shape modulation. Since phase in a sinusoidal communication system is associated with the delay of a sine wave, the overuse of the term phase in UWB can be confusing. However, the use of BPM has become common in the UWB literature, so we continue to use it here. Bi-phase modulation is easily understood as the inversion of a particular pulse shape; therefore, we take the following equation

$$s_i = \sigma_i p(t), \quad \sigma_i = 1, -1 \tag{5.3}$$

to create a binary system based on inversion of the basis pulse $p(t)$. The parameter σ is often known as the *pulse weight*, but here we will refer to it as the shape parameter. For a binary system the two resultant pulse shapes s_1, s_2 are defined simply as $s_1 = p(t)$ and $s_2 = -p(t)$.

One of the reasons for the use of bi-phase modulation, especially in comparison with pulse position modulation (which is a mono-phase technique) is the 3-dB gain in power efficiency. This is simply a function of the type of modulation method. That is, bi-phase modulation is an *antipodal modulation method*,[1] whereas pulse position modulation, when separated by one pulse width delay for each pulse position, is an *orthogonal modulation method*. The interested reader should refer to any digital communications textbook for details (see [59]).

A simple example can illustrate this advantage of BPM. Since PPM must always delay pulses, in the limit when pulses are transmitted continuously PPM must always "waste" the time when pulses are not transmitted. If PPM delays by one pulse width, then BPM can send twice the number of pulses and, thus, twice the information, thus achieving a system which, given all other things being equal, has twice the data rate.

Another benefit of using BPM is that the mean of σ is zero. This has the important benefit of *removing* the comb lines or spectral peaks that were discussed in Chapter 1, without the need for "dithering". This of course assumes that transmitted bits are equally likely; however, this is a common and reasonable assumption in most digital communication systems. According to McCorkle [60], bi-phase modulation in UWB presents several other benefits:

> First, it exhibits a peak-to-average power ratio of less than 8 dB. Thus, an implementation using bi-phase does not require any external snap-recovery or

[1] Antipodal means opposite.

tunnel diodes or power-amplifier circuitry. Instead, it can be driven directly from a low-voltage high-speed CMOS IC.

Finally, for reasons of clocking, bi-phase modulation has reduced jitter requirements. In PPM, the clocking path must include elements to accurately control arbitrary time positions on a fast (pulse-to-pulse) basis. This control requires a series of wide-bandwidth circuits where jitter accumulates. But a bi-phase system needs only a stable, low-phase-noise clock as the pulses occur on a constant spacing. Synchronization circuits can be narrowband so that they do not add significant jitter. As a result, less power and real estate are needed to implement the required circuits.

5.3 OTHER MODULATION METHODS

Although the previously discussed PPM and BPM constitute the major approaches to modulation in UWB communication systems, other approaches have been proposed. A simple figure outlining these alternative modulation approaches is shown in Figure 5.4.

In Figure 5.4(a) an unmodulated pulse train is shown for comparison. In Figure 5.4(b) an example of pulse amplitude modulation is given where a pulse with large amplitude represents a "1" and smaller amplitude represents a "0" while in (c) on-off keying modulates data with the presence of a pulse representing "1" and absence representing "0". Figure 5.4(d) shows an example of orthogonal pulse modulation where a binary "1" is represented by a modified Hermitian pulse of order 3 and a binary "0" is represented by a modified Hermitian pulse of order 2 [17].

5.3.1 Orthogonal pulse modulation

Of the three unconventional modulation techniques, orthogonal pulse modulation (OPM) is simply a subset of general pulse shape modulation with the property that the pulse shapes are orthogonal to each other. The advantage of using orthogonal pulses is not strictly related to the modulation, but rather to the multiple access method, which is discussed later in this chapter. However, since arbitrary pulse shape modulation is not interesting in either a theoretical or practical sense,[2] here we consider OPM as an interesting subset of the general case of pulse shape modulation.

In narrowband sinusoidal communication, orthogonal sine and cosine functions form the basis for communication. In UWB we can design different pulse shapes that have the property of being orthogonal to each other. Unfortunately, a simple

[2]This is simply because additional circuitry, memory, and so on is needed to generate the arbitrary pulse shapes. If there is no reason for adding complexity to the system, we should not do it.

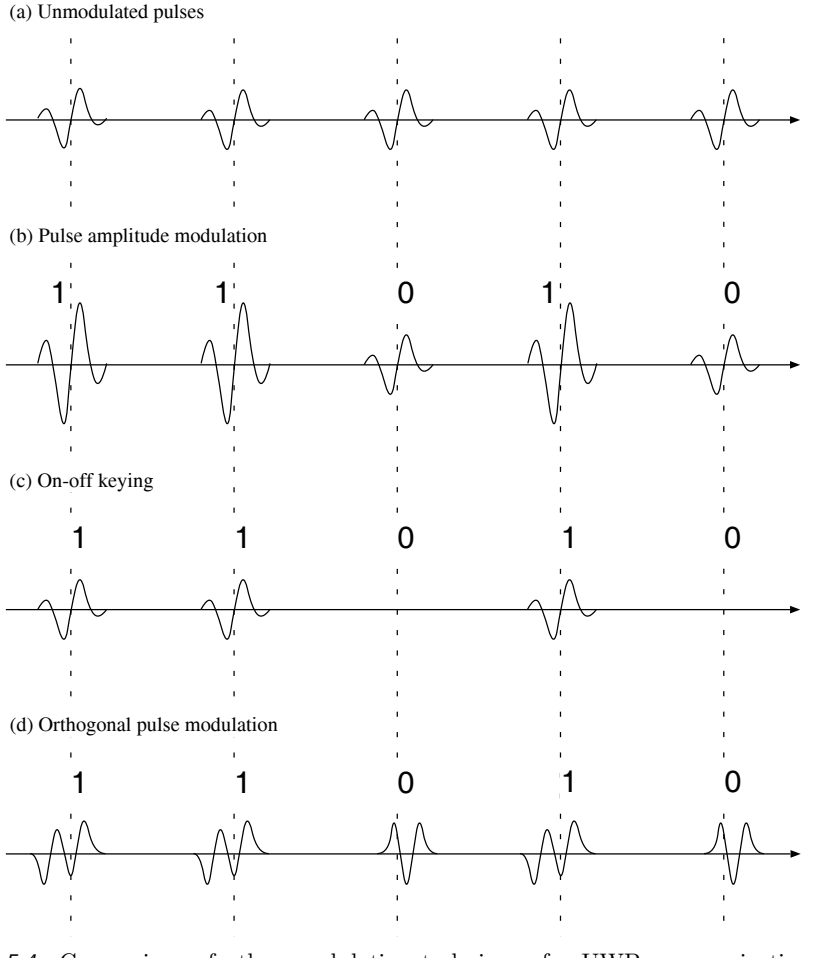


Fig. 5.4 Comparison of other modulation techniques for UWB communication: (a) an unmodulated pulse train, (b) pulse amplitude modulation, (c) on-off keying, and (d) orthogonal pulse modulation.

pulse shape parameter σ is inadequate to describe the set of pulses which we may encounter, and here we simply label each pulse as a general $p_1, p_2 \ldots p_i$ and assume that pulses are designed so as to be orthogonal. See [17] and Chapter 2 for examples on how to design orthogonal pulse shapes

$$s_i = \{p_1, p_2 \ldots p_i\} \tag{5.4}$$

The three modulation techniques presented previously – pulse position modulation, bi-phase modulation, and orthogonal pulse modulation – have been proposed

for use in ultra wideband communications. As far as the authors are aware, no serious attempt has been made to use either pulse amplitude modulation or on-off keying for UWB; however, for completeness here their possible use is discussed below.

5.3.2 Pulse amplitude modulation

Pulse amplitude modulation (PAM) for UWB can be represented as shown in Eqn. (5.5)

$$s_i = \sigma_i p(t) \quad \sigma_i > 0 \tag{5.5}$$

where the pulse shape parameter σ takes on positive values greater than zero. As an example we can set $\sigma_i = 1, 2$ and obtain the binary pulse set $s_1 = p(t)$, $s_2 = 2p(t)$.

In general, amplitude modulation is not the preferred way for most short-range communication. The major reasons include the fact that, in general, an amplitude-modulated signal which has a smaller amplitude is more susceptible to noise interference than its larger amplitude counterpart. Furthermore, more power is required to transmit the higher amplitude pulse.

In sinusoidal systems, amplitude-modulated systems are usually characterized by a relatively low bandwidth requirement and power inefficiency in comparison with angle modulation schemes. Thus, the major advantage (low bandwidth) can be seen to be anti-ethical to UWB, and in most UWB applications power efficiency is of high importance.

5.3.3 On-off keying

On-off keying (OOK) for UWB can be characterized as a type of pulse shape modulation where the shape parameter s is either 0 or 1, as shown in Eqn. (5.6)

$$s_i = \sigma_i p(t) \quad \sigma_i = 0, 1 \tag{5.6}$$

For example, the "on" pulse is created when $\sigma_i = 1$ and the "off" pulse when $\sigma_i = 0$; thus, $s_1 = p(t)$ and $s_2 = 0$.

The major difficulty of OOK is the presence of multipath, in which echoes of the original or other pulses make it difficult to determine the absence of a pulse. On-off keying is also a binary modulation method, similar to BPM, but it cannot be extended to an M-ary modulation method, as can PPM, PAM, and OPM.

5.3.4 Summary of UWB modulation methods

In this subsection we conclude the discussion of modulation methods for UWB communications with Table 5.1 which summarizes the advantages and disadvantages of each of the modulation methods.

Table 5.1 Advantages and disadvantages of various modulation methods.

Modulation methods	Advantages	Disadvantages
PPM	Simplicity	Needs fine time resolution
BPM	Simplicity, efficiency	Binary only
OPM	Orthogonal for multiple access	Complexity
PAM	Simplicity	Noise immunity
OOK	Simplicity	Binary only, noise immunity

5.4 PULSE TRAINS

We examined the creation of single pulses in Chapter 2. In this chapter we have looked at sets of pulses which are used for the modulation of digital information onto analog pulse shapes. We now turn our attention to sequences of pulses, called pulse trains, which will be able to transmit much larger volumes of information than a single set of pulses. In general, an unmodulated pulse train $s(t)$ with a regular pulse output can be written as

$$s(t) = \sum_{n=-\infty}^{\infty} p(t - nT) \tag{5.7}$$

where T is the period or the pulse-spacing interval and $p(t)$ is the basis pulse.

The effects of changing the pulse duration and repetition rate of each pulse have been examined and the results are as follows:

- Increasing the pulse rate in the time domain increases the magnitude in the frequency domain (i.e., the pulse rate influences the magnitude of the spectrum).
- The lower the pulse duration in the time domain the wider the spectral width (i.e., the pulse duration determines spectral width).
- A random pulse-to-pulse interval produces a much lower peak magnitude spectrum than a regular pulse-to-pulse interval since the frequency components are unevenly spread over the spectrum and the addition of magnitude is less effective. Therefore, the pulse-to-pulse interval controls the separation of spectral components.

5.4.1 Gaussian pulse train

As an example of a pulse train, let us consider the Gaussian doublets of Figure 2.2 with pulse position modulation.

As briefly mentioned in Chapter 1, assuming PPM is being used there is the problem of spectral peaks when a regular pulse train is used. These energy spikes can cause interference with other RF systems at short range and limit the amount of useful energy transmitted. One method to overcome these spectral peaks is to "dither" the transmitted signal by adding a random offset to each pulse, removing the common spectral components. However, when we attempt communication with this model the random offset is unknown at the receiver, making it extremely difficult to acquire and track the transmitted UWB signal. Another method with similar random properties, but using a known sequence, is to use *pseudo-random noise* (PN) codes to add an offset to the PPM signal. Since these codes are known and easily reproducible at the receiver, the problem for the receiver becomes mostly acquisition of the signal, but tracking is made much easier.

5.4.2 PN channel coding

The use of a PN time shift has other benefits besides just reducing the spectral peaks resulting from regular pulse emissions. Since the PN code is a channel code it can be used as a multiple access method to separate users in a similar manner to the code division multiple access (CDMA) scheme. By shifting each pulse at a pseudo-random time interval the pulses appear to be white background noise to users with a different PN code. Furthermore, the use of PN codes makes data transmission more secure in a hostile environment.

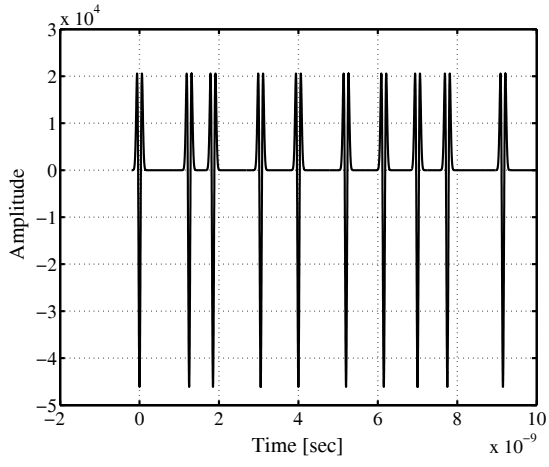
The impact of PN time offsets on energy distribution in the frequency domain is illustrated in Figure 5.5. Here, the basis pulse $p(t)$ used for transmission of the signals is assumed to be the Gaussian doublet of Figure 2.2: that is, $y_{g_3}(t)$. The sequence of pulses are formed as shown in Eqn. (5.8)

$$s(t) = \sum_{n=0}^{N_t-1} y_{g_3}(t - t_{d_n}) \tag{5.8}$$

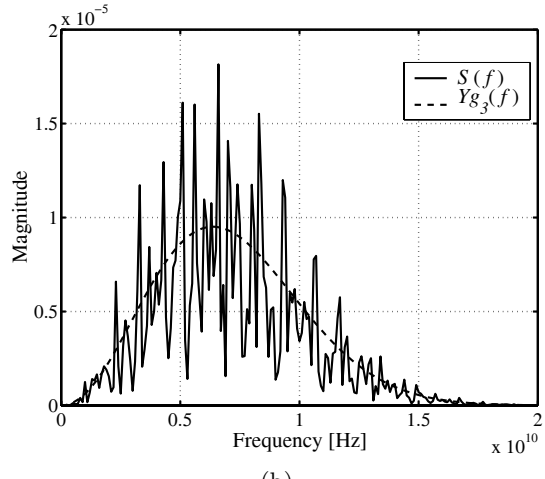
where $y_{g_3}(t)$ has been defined in Eqn. (2.7), N_t is the number of doublets, and t_{d_n} is the time delay associated with the pulse number n. This time delay is related to the input code and can be calculated in different ways: for example, as in Eqn. (5.9)

$$t_{d_n} = nT_f + T_{p_n} + T_{c_n} \tag{5.9}$$

where T_f is the pulse repetition time, T_{c_n} is the random time shift related to the PN code sequence, and T_{p_n} is the time shift from the pulse position modulation scheme. In practice T_f is much greater than T_{p_n} or T_{c_n}.



(a)



(b)

Fig. 5.5 A train of Gaussian doublets in time and frequency domains.

Using the Fourier transform of $y_{g_3}(t)$ indicated in Eqn. (2.14) we can conclude that the spectrum of $s(t)$ is as shown in Eqn. (5.10):

$$S(f) = \sum_{i=0}^{N_t-1} K_3 \tau \sqrt{\pi} (j2\pi f)^2 e^{-(\pi\tau f)^2} e^{-j2\pi f t_{d_n}} \tag{5.10}$$

where K_3 is calculated based on Eqn. (2.11) for $E_3 = 0.1$.

Example 5.1

Using Eqn. (5.8), plot the time domain function of a Gaussian doublet train. Use $N_t = 10$ and let the energy of each doublet be equal to 0.1, which gives a total energy of unity.

Using Eqn. (5.10), plot the frequency domain function of the Gaussian doublet train. Compare the result with the frequency domain representation of a single doublet $y_{g3}(t)$.

Solution

The curves are plotted in Figure 5.5(a) and (b).

5.4.3 Time-hopping PPM UWB system

We can now combine the techniques introduced to build a simple UWB transmitter. We will use a time-hopping code and binary pulse position modulation, with a single reference pulse shape $p(t)$. This system is perhaps the most common in the literature (e.g., see [10]). This system only requires a single template pulse for reception, and most of the complexity of this system resides in providing accurate timing for the generation of the transmitted sequence and subsequent reception.

In Figure 5.6 we show the output of this simple UWB transmitter. We describe it here for the single-user case, but we extend it easily and simply for the multi-user case by using different time-hopping codes, which in general will be pseudo-noise codes.

First, we note that there is one pulse transmitted in each frame of time T_f. The *pulse repetition frequency* (PRF), as described previously, is

$$\text{PRF} = \frac{1}{T_f} \tag{5.11}$$

The frame time should be at least long enough to overcome the delay spread of the channel, which, as described in Chapter 4, is generally of the order of hundreds of nanoseconds for an indoor environment, in order to avoid interference from reflected pulses. Thus, the frame time will be of the order of 1,000 times the actual pulse width. The unmodulated pulse stream is represented as

$$s(t) = \sum_{n=-\infty}^{\infty} p(t)(t - nT_f) \tag{5.12}$$

To modulate the data we add a small shift in the pulse position T_{p_n}, either forward T_{p_0} or back T_{p_1}, to represent the binary data stream. Often $T_{p_0} = -T_{p_1}$.

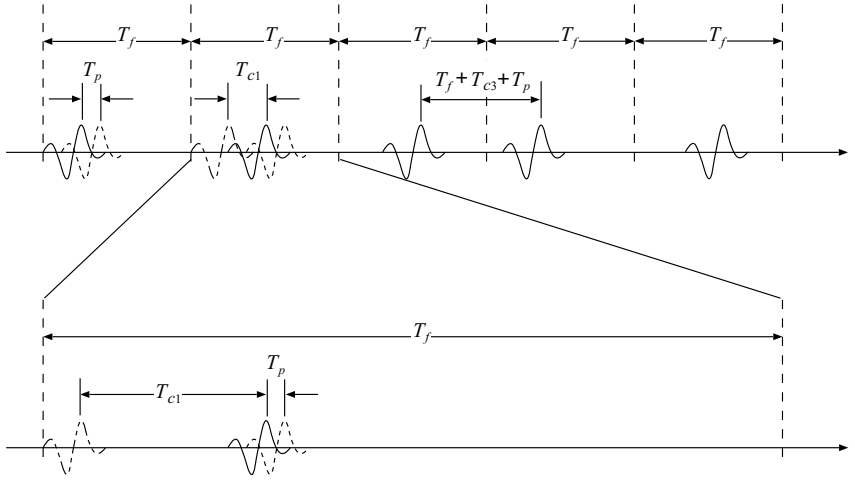


Fig. 5.6 A time-hopping, binary pulse position modulation system output.

The modulated data stream becomes

$$s(t) = \sum_{n=-\infty}^{\infty} p(t)(t - nT_f - T_{p_n}) \tag{5.13}$$

To avoid spectral lines and to provide a means of distinguishing users we finally add a time-shift based on a time-hopping code T_{c_n}, where the code repeats after a certain interval. The final output of the time-hopping, pulse position-modulated signal is then given by

$$s(t) = \sum_{n=-\infty}^{\infty} p(t)(t - nT_f - T_{p_n} - T_{c_n}) \tag{5.14}$$

5.5 UWB TRANSMITTER

A general UWB transmitter block diagram is shown in Figure 5.7. First, meaningful data are generated by applications that are quite separate from the physical layer transmitter. Applications might be an e-mail client or a web browser on a personal computer, a calendar application on a personal digital assistant (PDA) like the Sony Clié, or the digital stream of data from a DVD player. From the perspective of the physical layer the data may be anything at all. This part of the wireless device is often called the "back end". This terminology is not immediately apparent, but it is common to refer to it is from the receiver's point of view.

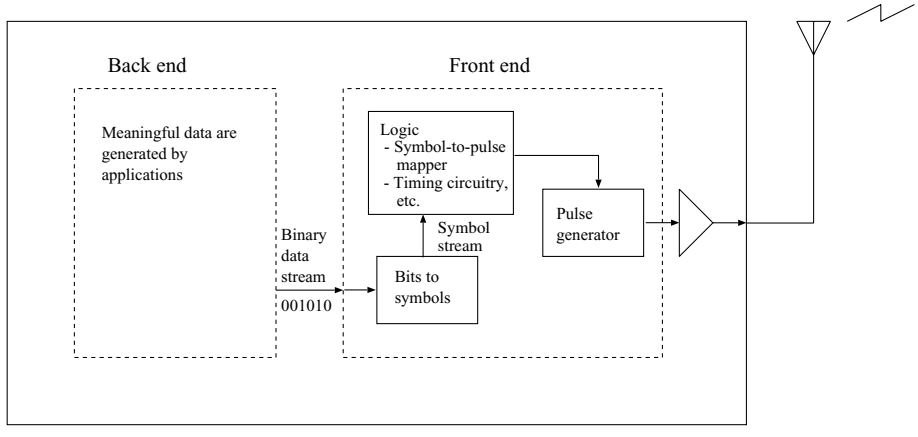


Fig. 5.7 A general UWB transmitter block diagram.

This binary information stream is then passed to the "front end", which is the part of the transmitter which we are concerned about. If higher modulation schemes are to be used the binary information should be mapped from bits to symbols, with each symbol representing multiple bits. These symbols are then mapped to an analog pulse shape. Pulse shapes are generated by the pulse generator. Precise timing circuitry is required to send the pulses out at intervals which are meaningful. If PPM is employed the timing must be even more precise, usually less than one pulse width.

Pulses can then be optionally amplified before being passed to the transmitter. In general though, to meet power spectral requirements large gain is typically not needed and may be omitted.

Although this is an extremely simplistic transmitter model, which omits any forward error-correcting scheme, it serves the purpose to show that UWB transmitters can be quite simple. This is to be compared with other wireless transmitters, such as OFDM. See Figure 5.17 for comparison.

5.6 UWB RECEIVER

A general UWB receiver block diagram is shown in Figure 5.8. The receiver performs the opposite operation of the transmitter to recover the data and pass the data to whatever "back end" application may require it.

There are two major differences between the transmitter and the receiver. One is that the receiver will almost certainly have an amplifier to boost the signal power

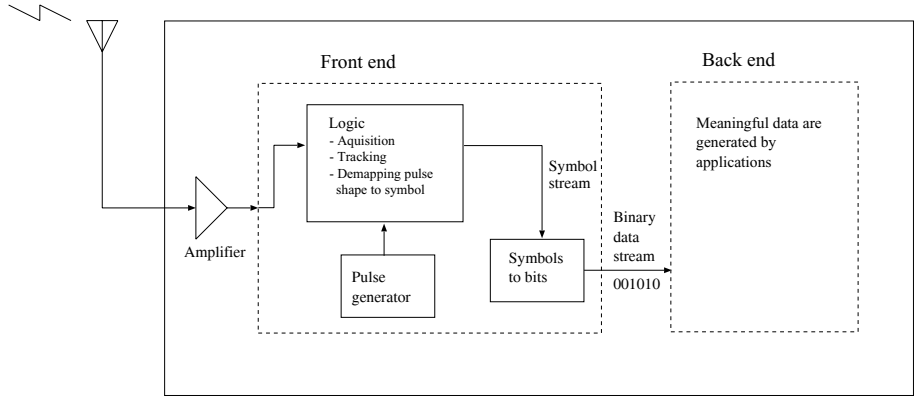


Fig. 5.8 A general UWB receiver block diagram.

of the extremely weak signals received. The other is that the receiver must perform the functions of *detection* or *acquisition* to locate the required pulses amongst the other signals and then to continue *tracking* these pulses to compensate for any mismatch between the clocks of the transmitter and the receiver.

Communication requires both the transmission and reception of signals. We have mostly concentrated on the wireless transmission side up to the moment. We will now focus on the detection of pulses; that is, the acquisition and tracking of the pulse trains.

5.6.1 Detection

Having generated a signal with the desired spectral features, it is also necessary to have an optimal receiving system. The optimal receive technique, the technique often used in UWB, is a correlation receiver, usually known as a *correlator*. A correlator multiplies the received RF signal by a *template waveform* and then integrates the output of that process to yield a single DC voltage. This multiply-and-integrate process occurs over the duration of the pulse and is performed in less than a nanosecond. With the proper template waveform the output of the correlator is a measure of the relative time positions of the received monocycle and the template.

If we assume PPM as the modulation method the correlator is an optimal early/late detector. As a very simple example, when the received pulse is $\frac{1}{4}$ of a pulse early the output of the correlator is $+1$, when it is $\frac{1}{4}$ of a pulse late the output is -1, and when the received pulse arrives centered in the correlation window the output is zero.

It is critical to note that the mean value of the correlator is zero. Thus, for in-band noise signals received by a UWB radio the correlator's output has an average value of zero. Moreover, the standard deviation or root mean square (rms) of the correlator output is related to the power of those in-band noise signals.

5.6.2 Pulse integration

When a monocycle is buried in the noise of other signals, it is extremely difficult to detect a single UWB pulse and the confidence with which we can receive the transmitted information is low. However, by adding together multiple correlator samples (i.e., multiple pulses), it becomes possible to receive transmitted signals with much higher confidence. This process is called *pulse integration*. Through pulse integration, UWB receivers can acquire, track, and demodulate UWB transmissions that are significantly below the noise floor. The measure of a UWB receiver's performance in the face of in-band noise signals is called the *processing gain*.

5.6.3 Tracking

Tracking is the process by which the receiver must continually check to see whether the pulses are arriving at the expected time and, if not, to adjust that time. A simple example will serve to show the process. Assume that the transmitter and receiver start with their clocks synchronized. As time passes the effects of heat and differences in manufacture cause one of the clocks or oscillators to become slightly faster. If this difference is not corrected, eventually the receiver will not be able to correctly demodulate the pulses. The time drift at subnanosecond orders must be vigilantly watched, in particular.

5.6.4 Rake receivers

As discussed in Chapter 4, the wireless channel suffers from *multipath*, where reflections and other effects of the channel cause multiple copies of the transmitted pulse to appear at the receiver. If a *rake receiver* is used, these extra pulses can be used to improve reception at the cost of increased receiver complexity. The increased complexity comes from the additional circuitry required to track multiple pulses and demodulate them. The name *rake* receiver comes from the common garden rake. The delay profile of the received pulse looks like an upturned garden rake.

5.7 MULTIPLE ACCESS TECHNIQUES IN UWB

Up to now we have implicitly assumed that there has only been one user using the UWB system at any one time. Although consumer UWB systems are in their

infancy, we must consider how best to design a UWB system where there is more than one user.

Naturally, all traditional multiple access methods should be considered; however, here we consider time, frequency, code, and space division multiple access. As an interesting special case for UWB, we look at orthogonal pulses as a novel multiple access technique.

5.7.1 Frequency division multiple access UWB

A common multiple access technique in narrowband communication is to divide users up based on frequency bandwidth. This is known as frequency division multiple access (FDMA). Each user uses a different carrier frequency to transmit and receive on.

In UWB, FDM is achieved by using pulses which have a narrower bandwidth than the total available bandwidth; however, they are still extremely broadband. Channelization can be achieved by multiplying by a sinusoidal carrier. The total effect can be thought of as an extremely broadband OFDM system.

5.7.2 Time division multiple access

In time division multiple access (TDMA), each user uses the same codes and the same bandwidth; however, a different time offset is needed to avoid interference. In general, this requires that all users be synchronized, which is not an easy task as the number of users increases. In general, this technique would only be applied to the downlink (from a central base station) to mobile users.

5.7.3 Code division multiple access

One multiple access technique possible in UWB is to assign a different spreading code to each user. This is known as code division multiple access (CDMA).

As an example, let us take Eqn. (5.14) and modify it so that we can separate k users out by different code. We refer to the kth user's output from the transmitter as

$$s^{(k)}(t) = \sum_{n=-\infty}^{\infty} p(t)(t - nT_f - T_{p_n}^{(k)} - T_{c_n}^{(k)}) \tag{5.15}$$

In Eqn. (5.15) the kth user has a different binary data stream, so we label this $T_{p_n}^{(k)}$; however, to distinguish the user we must have a distinct time-hopping code, which we distinguish as $T_{c_n}^{(k)}$.

5.7.4 Orthogonal pulse multiple access system

As an example of orthogonal pulse modulation let us look closely at the modified Hermite orthogonal pulse system proposed by [17].

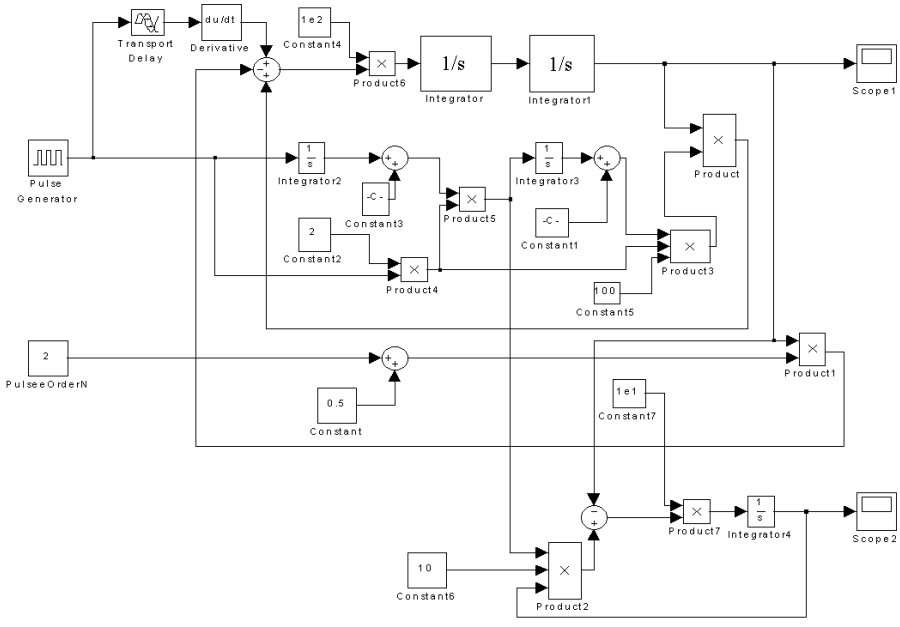
An M-ary communication system can be constructed from any set of orthogonal pulse shapes, such as $h_n(t)$ or $p_n(t)$. For simplicity let us consider only modified Hermitian pulse (MHP) waveforms.

We arbitrarily assume that 2-bit binary codes 00, 01, 10, and 11 are represented by MHP pulses of orders $n = 1, 2, 3, 4$. By assigning multiple-bit patterns to single pulse shapes, higher data rates can be achieved than simply by sending different pulse shapes. Furthermore, this can be extended to a coded scheme if desired.

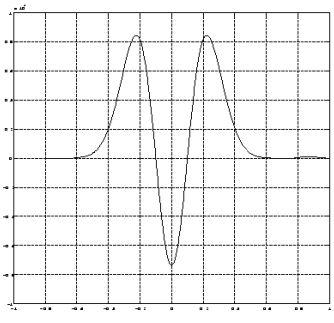
Since MHP pulses are orthogonal a multi-user system can be created using the same four pulse shapes, for example, by assigning MHP pulses of order $n = 1, 2$ to user 1 for the binary $0, 1$ and $n = 3, 4$ to user 2 for $0, 1$.

For a binary communication system using orthogonal pulse shape modulation, we wish to know whether a pulse representing either 0 or 1 is received. To achieve this, we need to generate a local copy of each pulse shape and integrate it with the received pulse. Conventionally, two complete sets of hardware are needed in order to produce two pulses of different shapes. However, because modified Hermite pulse shapes of lower order can be generated by integrating a pulse of higher order, a low-complexity multiple pulse generator can be constructed. We use the modified Hermite orthogonal pulse of the particular order from the first pulse generator to generate a different-order modified Hermite orthogonal pulse at the second pulse generator. The configuration of the second pulse generator is much less complex than if the different-order modified Hermite orthogonal pulses were produced from scratch based on a source signal. Also, only a single source signal is required to produce two different-order pulses.

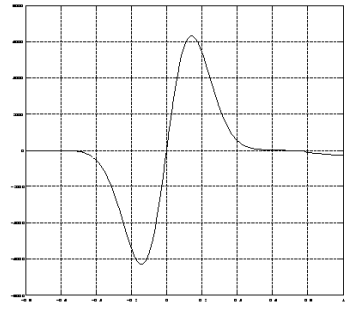
A Matlab model of the circuit to obtain double pulse generation is outlined in Figure 5.9. We can see that by utilizing one of the properties of the pulses (i.e., by differentiating or integrating them) another pulse can be created, with the order of the pulses being one more or one less than the original pulse, respectively. In Figure 5.9 the order $n = 2$ is input to the system, along with a pulse of specified width. The width of the input pulse is determined by the desired width of the output pulse and is approximately twice the length of the output pulse. While the input pulse is on, the pulse will be produced, but once the input pulse is zero any output will be suppressed. In an actual circuit the power would be removed from the input, so that no output would result in any case. In particular, the additional pulse is created by integrating the output from a pulse of order n. Thus, a pulse of order $n - 1$ is created.



(a) Matlab circuit for generating different pulses.



(b) Pulse output of order n=2 (scope 1).



(c) Pulse output of order n=1 (scope 2).

Fig. 5.9 (a) A circuit for generating multiple orthogonal pulses; (b) and (c) sample output pulses when input is $n = 2$.

5.8 CAPACITY OF UWB SYSTEMS

The capacity of a UWB multiple access system that is dependent on a specific pulse shape is computed by Zhao and Haimovich in [61], and, although this is a simplistic pulse shape and is only applicable for a pulse position modulation system, it serves to illustrate several points about UWB capacity. Thus, we follow the derivation here.

First, several assumptions are made. For simplicity and without any loss in generality, each UWB pulse is assumed to represent one symbol. Thus, the number of pulses per symbol $N_p = 1$. This means that the symbol interval time T_s and the frame time interval T_f are the same (i.e., $T_f = T_s$) and the energy per pulse E_p is the same as the energy per symbol E_p.

Next, we comment that the Shannon capacity formula $C = W \log_2(1 + \text{SNR})$ applies to a channel with continuous valued inputs and outputs, which is not the case for PPM UWB where the values are discrete.

For simplicity, the pulse shape is a rectangular waveform $p(t)$ where

$$p(t) = \sqrt{\frac{1}{T_p}}; \quad 0 \le t \le T_p \tag{5.16}$$

and T_p is the pulse time.

The correlation function $h(\tau)$ for $p(t)$ is

$$h(\tau) = \begin{cases} \dfrac{T_p + \tau}{T_p}, & -T_p \le \tau \le 0 \\ \dfrac{T_p - \tau}{T_p}, & 0 \le \tau \le T_p \end{cases} \tag{5.17}$$

In this derivation we model the time shift difference between users, with symbol Δ as a uniform distribution over the interval $[-T_f, T_f]$. It follows that probabilities

$$\begin{aligned} P(-T_p \le \Delta \le 0) &= P(0 \le \Delta \le T_p) \\ &= \frac{T_p}{2T_f} \\ &= \frac{1}{2\beta} \end{aligned} \tag{5.18}$$

where $\beta = T_f/T_p$ is the spreading ratio, which is defined as the time of the frame divided by the time of the pulse.

The mean value of $h(\Delta)$ can be calculated as

$$\begin{aligned} E[h(\Delta)] &= E[h(\Delta) \mid -T_p \le \Delta \le 0]P(-T_p \le \Delta \le 0) \\ &\quad + E[h(\Delta) \mid 0 \le \Delta \le T_p]P(0 \le \Delta \le T_p) \qquad (5.19) \\ &= \frac{1}{2\beta} \qquad (5.20) \end{aligned}$$

The variance of $h(\Delta)$ is denoted as σ_h^2. Since $E[h(\Delta)] = 1/(2\beta)$, we can write

$$\sigma_h^2 = E[h^2(\Delta)] - \left(\frac{1}{2\beta}\right)^2 \tag{5.21}$$

$$= E[h^2(\Delta) \mid -T_p \le \Delta \le 0]P(-T_p \le \Delta \le 0) + E[h(\Delta) \mid 0 \le \Delta \le T_p]P(0 \le \Delta \le T_p) - \frac{1}{4\beta^2} \tag{5.22}$$

$$= \frac{1}{3\beta} - \frac{1}{4\beta^2} \tag{5.23}$$

which can be approximated as

$$\sigma_h^2 \approx \frac{1}{3\beta} \tag{5.24}$$

for large values of the spreading ratio β.

The variance σ_I^2 of the multiple access interference term N_I can be calculated as

$$\sigma_I^2 = \sum_{\nu=2}^{N_u} A^{(\nu)^2} \sigma_h^2 \approx \sum_{\nu=2}^{N_u} \frac{A^{(\nu)^2}}{3\beta} \tag{5.25}$$

where ν is the user's index and N_u denotes the total number of users.

The SNR at the output of the receiver for each symbol can be calculated as

$$\rho_I = \frac{A^{(1)^2}}{\sigma_I^2 + N_0/2} = \frac{A^{(1)^2}}{\left(\sum_{\nu=2}^{N_u} A^{(\nu)^2}/3\beta\right) + N_0/2} \tag{5.26}$$

after considering additive white gaussian noise (AWGN).

With perfect power control, $A^{(1)} = A^{(\nu)} = \sqrt{E_p}$ and ρ_I can be expressed as

$$\rho_I = \frac{3\beta}{(N_u - 1) + 3\beta/\rho_0} \tag{5.27}$$

where $\rho_0 = 2E_p/N_0$. From this expression it can be seen that for a low number of users (i.e., $N_u < 3\beta/\rho_0$) the performance is noise-limited, while for a large number of users (i.e., $N_u > 3\beta/\rho_0$) the performance becomes interference-limited.

Using Eqn. (5.27) the single-user capacity $C_{M-\mathrm{PPM}}$ as a function of the channel symbol SNR ρ_I is given by

$$C_{M-\mathrm{PPM}}(\rho_I) = \log_2 M - E_{v|x_1} \log_2 \sum_{m=1}^{M} e^{\sqrt{\rho_I}(v_m - v_1)} \tag{5.28}$$

measured in bits per symbol. Here, v_m are random variables with m between 1 and M and have the following distribution conditional on the transmitted signal x_1

$$\begin{aligned} v_1 &: N(\sqrt{\rho_I}, 1) \\ v_m &: N(0,1), \quad m \neq 1 \end{aligned} \tag{5.29}$$

In Figure 5.10 the effects of thermal noise are ignored, and we concentrate on multiple access interference. The multiple access channel will achieve full user capacity when the number of users $N_u < 15$, and when the number of users is greater, user capacity will decrease.

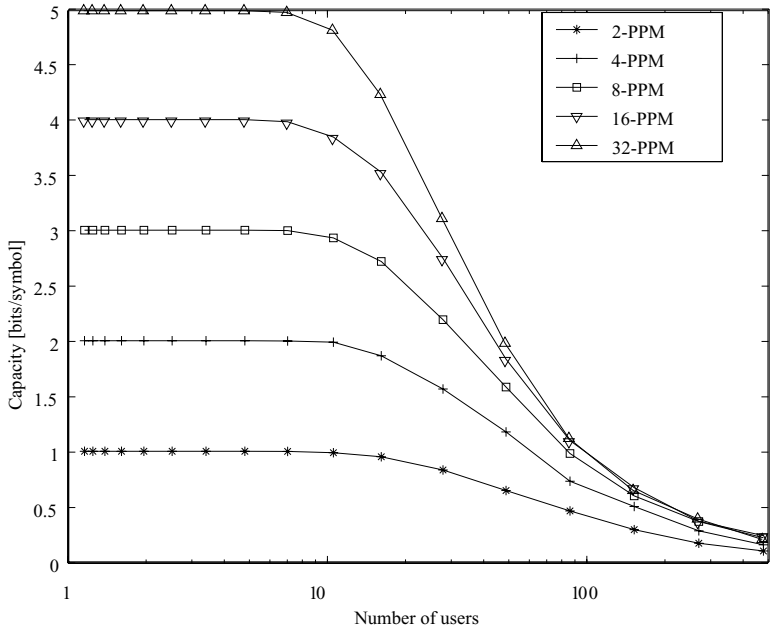


Fig. 5.10 User capacity for a multi-user UWB as a function of the number of users N_u for spreading ratio $\beta = 50$, ©IEEE 2002.

5.9 COMPARISON OF UWB WITH OTHER WIDEBAND COMMUNICATION SYSTEMS

In this section we discuss some of the important differences and similarities of UWB communication systems, spread spectrum (SS), and orthogonal frequency division multiplexing (OFDM) systems. Although this is a general discussion, we

have in mind indoor wireless local area network (WLAN) protocols (i.e., 802.11b and 802.11a). The reason is that in such a short space we cannot hope to cover in any depth the variations of spread spectrum communication or OFDM, and more particularly, many of the applications in which spread spectrum or OFDM are used or planned to be used are not within the current sphere of interest of UWB. To give some examples, spread spectrum is used within the so-called *third-generation* mobile telephone and data services. The communication from base station to mobile is of the order of hundreds of meters to kilometers. On the other hand, OFDM is being considered for so-called *fourth-generation* mobile systems. Moreover, OFDM is also used for digital television broadcasting, such as ISDB-T in Japan. Ultra wideband communication techniques are not currently being considered for these outdoor, long-range applications. However, indoor wireless LANs are within the possible application sphere of UWB and, thus, present a good opportunity to compare and contrast.

The IEEE standards by which these WLANs are commonly known to employ a direct sequence spread spectrum (DSSS) in 802.11b, which is centered at 2.4 GHz, and OFDM for the 802.11a standard at 5 GHz. It will be helpful to review each of these in turn. Then, we will proceed to a comparison and discuss the theoretical and practical difference between the three wideband communication systems.

5.9.1 CDMA

One of the most popular indoor wireless communication standards is the IEEE 802.11b standard for wireless local area networks. It operates in the 2.4-GHz unlicensed band. In 802.11b, spread spectrum techniques are used to take a narrowband data signal and spread it over the entire available frequency band, in order to combat interference from other users or noise sources. The 2.4-GHz band is known as the ISM band, which stands for industrial, scientific, and medical band. It accommodates many sources of electromagnetic radiation. One of the most common of these is the ordinary microwave oven.

There are two common techniques to spread the spectrum: the frequency-hopping spread spectrum (FHSS) and the direct sequence spread spectrum (DSSS). An overview of the frequency-time relationship is shown for these two methods in Figures 5.11 and 5.12.

In Figure 5.11 we can see that two users occupy a narrow frequency band for a short period of time. There are 79 frequency-hopping channels in the IEEE 802.11 standard and each is 1 MHz in width. Hops must take at most 224 μs.

In contrast, Figure 5.12 shows that each user occupies all of the available spectrum all the time and that different users are separated by their pseudo-noise (PN) codes. Thus, DSSS is also called code division multiple access (CDMA).

Although both DSSS and FHSS are specified as standards for IEEE 802.11 wireless LAN, in the more recent IEEE 802.11b standard, DSSS is the only physical

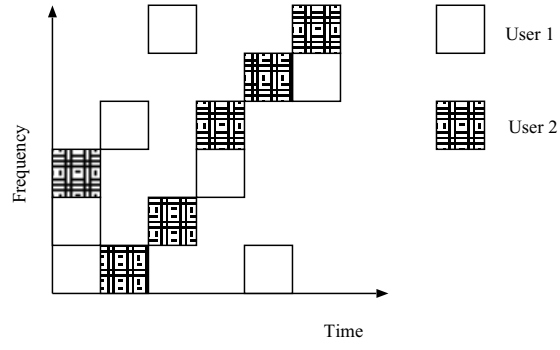


Fig. 5.11 Frequency-time relationship for two users using the frequency-hopping spread spectrum.

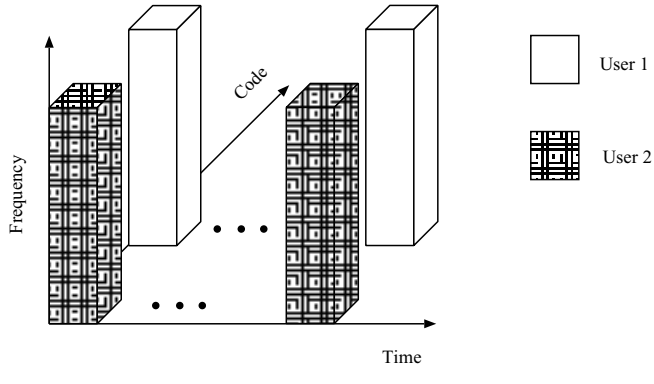


Fig. 5.12 Frequency-time relationship for two users using the direct sequence spread spectrum. The two users are separated by different codes.

layer defined. The 802.11a standard defines an OFDM physical layer (discussed later).

5.9.2 Comparison of UWB with DSSS and FHSS

In [62] a comparison of three different modulation techniques was considered: the direct sequence spread spectrum (DSSS), the frequency-hopping spread spectrum (FHSS), and ultra wideband. The setting was for each method to occupy 3.2 MHz, transmit at a rate of 3.125 Mbps, and support 30 simultaneous users.

For a DSSS system the signal-to-noise ratio (SNR) can be expressed as

$$\mathrm{SNR}_{\mathrm{DSSS}} = \frac{1}{(K-1)/3N - N_0/2E_b} \tag{5.30}$$

where K is the number of users and N is the number of chips per bit. The bit error rate (BER) can then be calculated from

$$\mathrm{BER}_{\mathrm{DSSS}} = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{\mathrm{SNR}_{\mathrm{DSSS}}}{2}}\right) \tag{5.31}$$

where "erfc" is the complementary error function, which is defined as:

$$\mathrm{erfc}(z) = 1 - \frac{1}{\sqrt{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k z^{2k+1}}{k!(2k+1)} \tag{5.32}$$

On the other hand the BER for FHSS can be calculated as

$$\begin{aligned}\mathrm{BER}_{\mathrm{FHSS}} = {} & \frac{1}{2}\left(1-\frac{1}{k}\right)^{M-1}\mathrm{erfc}\left(\sqrt{\frac{S}{2N}}\right) \\ & + \frac{1}{2}\sum_{i=1}^{M-1}\left(\frac{1}{k}\right)^{i}_{M-1} C_i\,\mathrm{erfc}\left(\sqrt{\frac{S}{2N+S_i}}\right)\end{aligned} \tag{5.33}$$

where k is the number of frequency-hopping slots, M is the number of users, S is the signal power, and N is the noise power. Thus, S_i represents the signal power from interfering users.

For the UWB result the average output SNR can be calculated by assuming a random time-hopping sequence. Let the number of active users be N_u. From [10] the SNR is

$$\mathrm{SNR} = \frac{(N_s A_1 m_p)^2}{\sigma^2_{\mathrm{rec}} + N_s\sigma_a^2\sum_{k=2}^{N_u} A_k^2} \tag{5.34}$$

where σ^2_{rec} is the variance of the receiver noise component at the pulse train integrator output. The monocycle waveform-dependent parameters m_p and σ_a^2 are given by

$$m_p = \int_{-\infty}^{\infty}\omega_{\mathrm{rec}}(x-\delta)v(x)\,dx \tag{5.35}$$

and

$$\sigma_a^2 = T_f^{-1}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\omega_{\mathrm{rec}}(x-s)v(x)\,dx\right]^2 ds \tag{5.36}$$

respectively [62], where A_1 is the monocycle amplitude, T_f is the frame time which is assumed to be 10 ns, and N_s is the number of impulses per symbol.

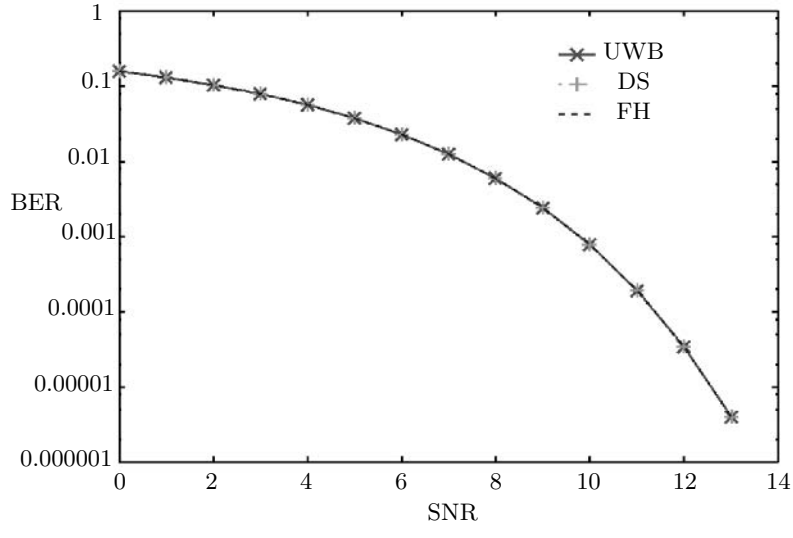


Fig. 5.13 Comparison of the BER of three wideband systems DSSS, FHSS, and UWB for a single user.

The first result is shown in Figure 5.13 for the single-user case where we can easily see that all three methods have the same bit error rate curve against SNR.

When 30 users are transmitting simultaneously, differences are seen. This is illustrated in Figure 5.14. An error floor is seen for frequency hopping, which is because the number of users is too large for the number of frequency slots available. Thus, collisions will always occur, generating interference even at high SNR.

An expanded view of Figure 5.14 is shown in Figure 5.15, which clearly shows that, theoretically under these assumed conditions, DSSS performs better than UWB. However, the chip bandwidth assumed for DSSS is 0.37 ns, which means that signal processing is much more difficult (and therefore more expensive) to produce than for UWB systems.

Thus, we can summarize by saying that similar performance is obtained by both DSSS and UWB wideband systems, given the same bandwidth constraints. However, from the practical perspective, UWB offers a possibly much cheaper implementation. As the bandwidth increases the signal-processing burden on DSSS and FHSS systems increase, making UWB more attractive. We further note that the bandwidth assumed in this example is only 3 MHz: UWB offers bandwidth possibilities that are at least three orders of magnitude greater.

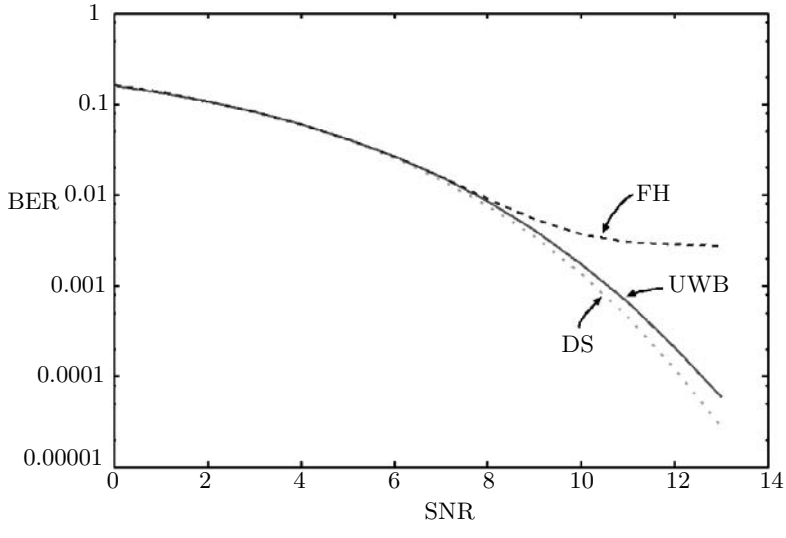


Fig. 5.14 Comparison of BER for the three systems when 30 users are simultaneously transmitting.

5.9.3 Orthogonal frequency division multiplexing

With orthogonal frequency division multiplexing (OFDM), multiple orthogonal carriers are transmitted simultaneously. By transmitting several symbols in parallel, symbol duration is increased proportionately, which reduces the effects of intersymbol interference (ISI) caused by a dispersive Rayleigh-fading environment.

The input sequences that determine which of the carriers is transmitted during the signaling interval is

$$s(t) = Ae^{2\pi f_i t} \cdot \Pi(t/T) \tag{5.37}$$

where

$$f_i = f_c + \frac{i}{T}, \quad i = 0, 1, \ldots, N-1 \tag{5.38}$$

and

$$\Pi(t/T) = \begin{cases} 1, & \dfrac{T}{2} \leq t \leq \dfrac{T}{2} \\ 0, & \text{otherwise} \end{cases} \tag{5.39}$$

The total number of subband carriers is N, and T is the symbol duration for the information sequence. In order that the carriers do not interfere with each other, the spectral peak of each subcarrier must coincide with the zero crossings of all

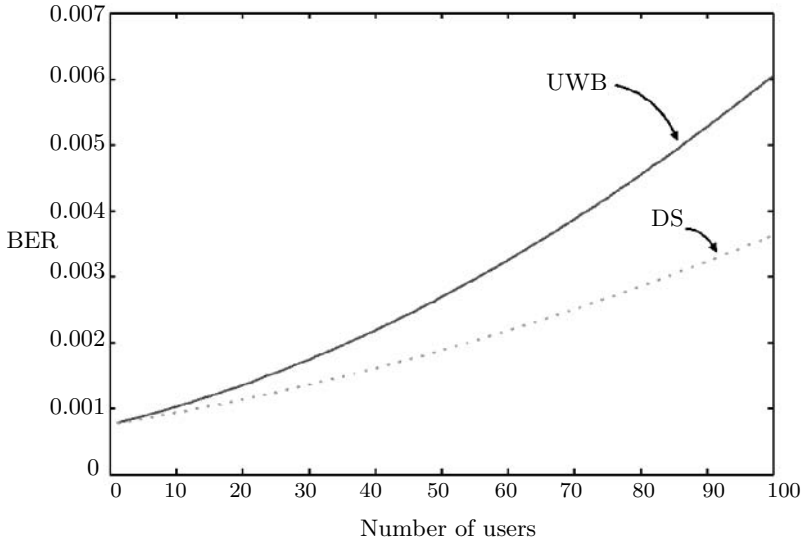


Fig. 5.15 Comparison of BER against the number of users for UWB and DSSS systems.

the other carriers (i.e., the subcarriers are orthogonal). This is shown graphically in Figure 5.16. The difference between the center lobe and the first zero crossing represents the minimum required spacing and is equal to $1/T$. An OFDM signal $s(n)$ can be constructed by assigning parallel bit streams to the subband carriers.

A block diagram of a typical 802.11a OFDM transmitter is shown in Figure 5.17. The modulation symbols are mapped to the subcarrier of the 64-point inverse discrete Fourier transform (IDFT) to create an OFDM symbol. However, only 48 subcarriers are used for data modulation, 4 subcarriers are used for pilot tones (which are used for channel estimation), and 12 subcarriers are not used.

The output of the inverse fast Fourier transform (IFFT) is converted to a serial sequence, and a guard interval or cyclic prefix is added.

A block diagram of a typical 802.11a OFDM receiver is shown in Figure 5.18. The receiver performs the inverse of the operations of the transmitter. Table 5.2 summarizes the key parameters of the IEEE 802.11a OFDM wireless local area network standard.

5.10 INTERFERENCE

Since UWB signals have such a broad bandwidth, they operate as an *overlay system* with other wireless communication methods. Figure 5.19 shows some of the

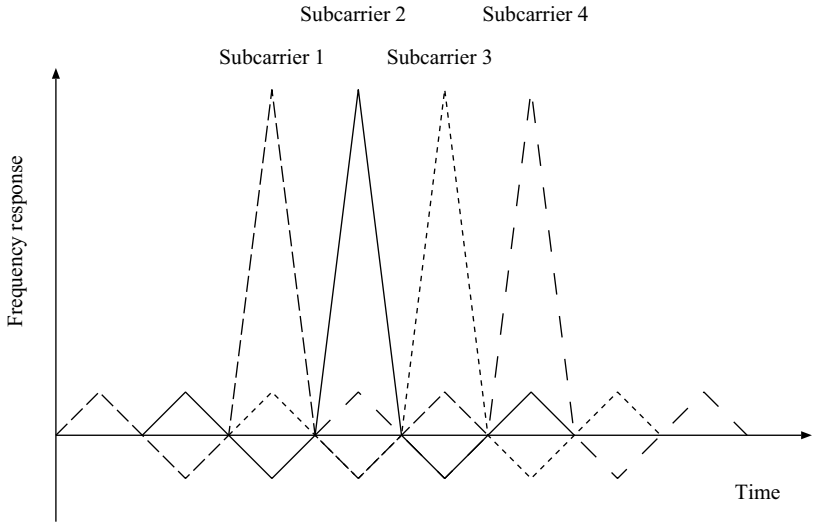


Fig. 5.16 Graphical representation of four orthogonal subcarriers to make up an OFDM symbol.

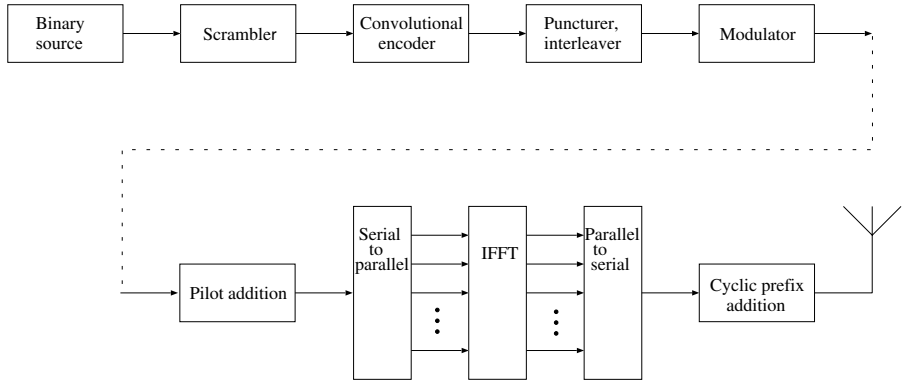


Fig. 5.17 Block diagram of a typical OFDM transmitter (IEEE 802.11a standard).

wireless systems with which UWB must contend. The problem is twofold. First, UWB must mitigate or be able to operate in the presence of these interferers. Second, UWB must not provide substantial interference to users of these other services. The definition of what is substantial interference is varied, but is generally taken to mean less than the legal maximum or less than the interference from unintentional radiators of electromagnetic energy.

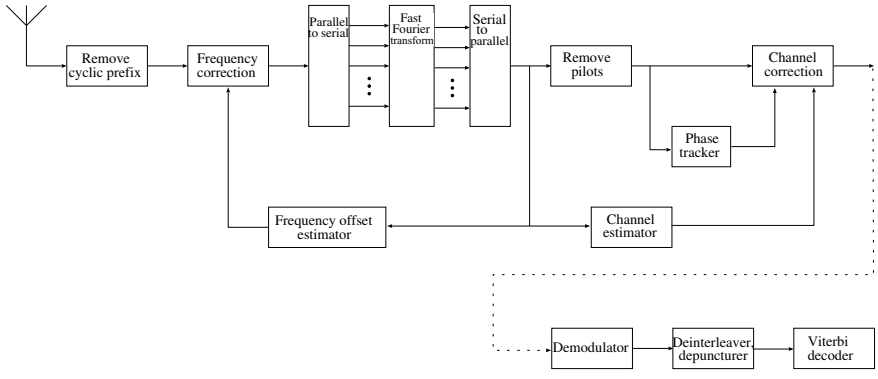


Fig. 5.18 Block diagram of a typical OFDM receiver (IEEE 802.11a standard).

Table 5.2 Key parameters of IEEE 802.11a OFDM wireless local area network standard.

Parameter	Values
Data Rate	6, 9, 12, 18, 24, 36, 48, 54 Mbps
Modulation	BPSK, QPSK, 16-QAM, 64-QAM
Subcarriers	52
Pilot tones	4
Symbol duration	4 μs
Subcarrier spacing	312.5 KHz
Signal bandwidth	16.66 MHz
Channel spacing	20 MHz

We can easily see that one of the reasons for avoiding the lower frequency bands is the plethora of wireless services with which UWB must contend. Between 3 GHz and 10 GHz the main source of interference for indoor wireless systems is assumed to be the 5-GHz wireless local area networks which are based on OFDM.

5.10.1 Wireless local area networks

In [63] the interference from IEEE 802.11a wireless LANs on UWB systems is discussed. It concluded that, with interfering system center frequency at 5.2 GHz and the UWB system center frequency at 4.2 GHz, the effect on BER performance is minimal when the signal power of the 802.11a signal is less than 10 dB of the UWB signal. The UWB system was simulated using a pulse waveform $s_0(t)$ that was modulated by a sine wave $\sin(2\pi f_0 t)$ at frequency f_0 Hz and given by the

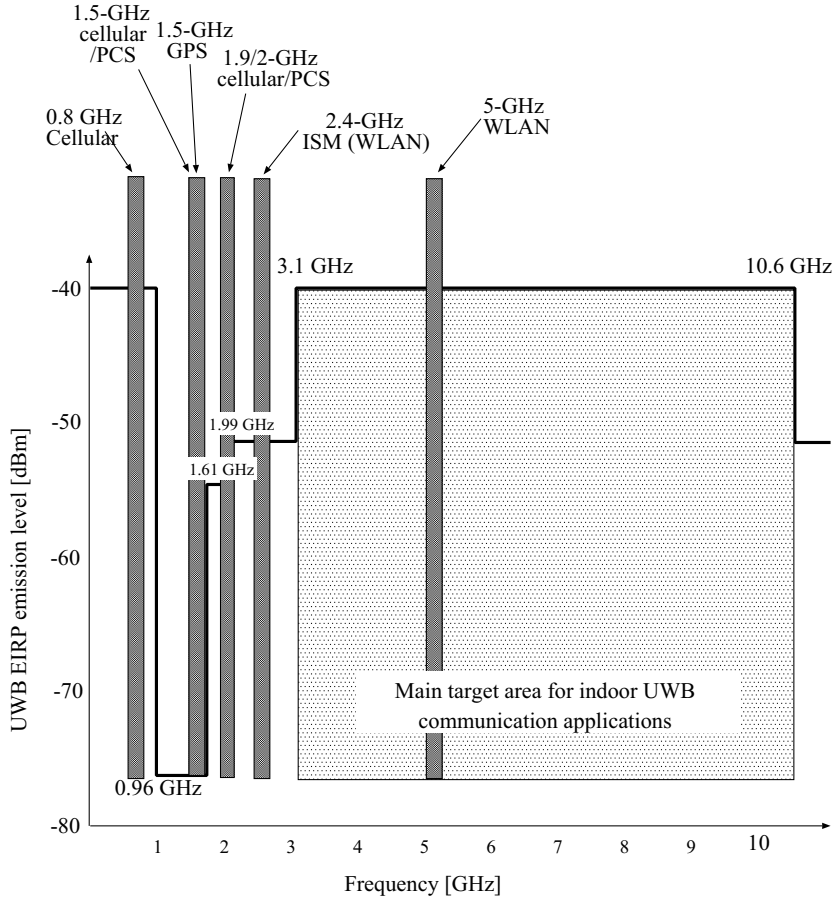


Fig. 5.19 Other wireless systems operating in the same bandwidth as UWB will both cause interference to and receive interference from each other.

equation

$$s_0(t) = e^{-at^2/\tau^2} \sin(2\pi f_0 t) \tag{5.40}$$

where, in the simulation, $a = \log_e 10$ and $\tau = 0.5$ ns. Pulses were sent every 5 ns using bi-phase modulation. The power of the UWB system for Eb/No (energy per bit divided by the noise power density) and the desired-to-undesired (DU) ratio was defined as the power of a single pulse in order to remove the effect of the pulse interval.

When the 802.11a signal is stronger, such as when the wireless LAN transmitter is close to the UWB receiver, significant interference occurs. To mitigate this interference, two techniques are suggested. One is to use a filter to remove unwanted interference. For example, a sixth-order Chebyshev filter with cutoff frequency at 4 GHz, ripple less than 0.2 dB, and −20 dB attenuation at 5.18 GHz can be used. This filter causes a 1-dB loss when there is no interferer present; however, significant performance improvement (no error floor for a DU signal ratio of 0 dB) is obtained when the interfering signal is present.

The second proposal is to use a multi-band UWB system. The highest frequency subcarriers are removed because they overlap with the 802.11a spectrum. Using a system with 11 subcarrier pulses at intervals of 200 MHz from 3.2 to 5.2 GHz gives

$$s_0(t) = e^{-at^2/\tau^2} \sin(2\pi f_0 t) \tag{5.41}$$

$$\approx \sum_{n=1}^{11} X_n e^{-at^2/\tau'^2} \sin(2\pi f_0 t) \tag{5.42}$$

where $\tau' = 0.5$ ns.

It was concluded that, at a DU ratio of 0 dB, removal of the two highest subcarriers gave the best BER performance, while at a DU ratio of −10 dB, removal of three subcarriers gave the best performance.

An experiment was performed to measure the performance of 802.11b wireless LAN networks under the influence of high-power UWB signals, and the results were reported in [64]. It was concluded that 802.11b can suffer from UWB interference given high-powered transmission and close range between the UWB transmitter and the wireless LAN receiver. The experimental setup is shown in Figure 5.20.

In this experiment, pulses with width 500 ps were generated from several UWB transmitters. The pulse repetition frequency in these prototypes was fixed at 87 MHz. The center frequency of the transmission was around 1.8 GHz. The peak-to-peak voltage for the pulse measured from the output port of the circuit board was approximately 300 mV. Omni-directional antennas were employed and had EIRP of −2 dBm to 3 dBm depending on the pulse repetition frequency. These prototypes are not compatible with FCC regulations, exceeding the mask for indoor transmission by 30 dB in the 2.4-GHz band.

Spectrum measurements on channel 1 ($f_c = 2.412$ GHz) showed peak UWB interference approximately 20 dBmV less than the peak 802.11b signal with 20 UWB transmitters at 100 cm. At 15 cm from the measuring antenna the UWB peak was approximately 5 dBmV less than the peak WLAN signal.

Signal-to-noise measurements were undertaken, and the conclusion was reached that, with the UWB transmitters used, if the distance from the WLAN receiver is greater than 50 cm, then no significant reduction in SNR occurred. For distances less than 50 cm, the SNR reduction was as much as 10–15 dB.

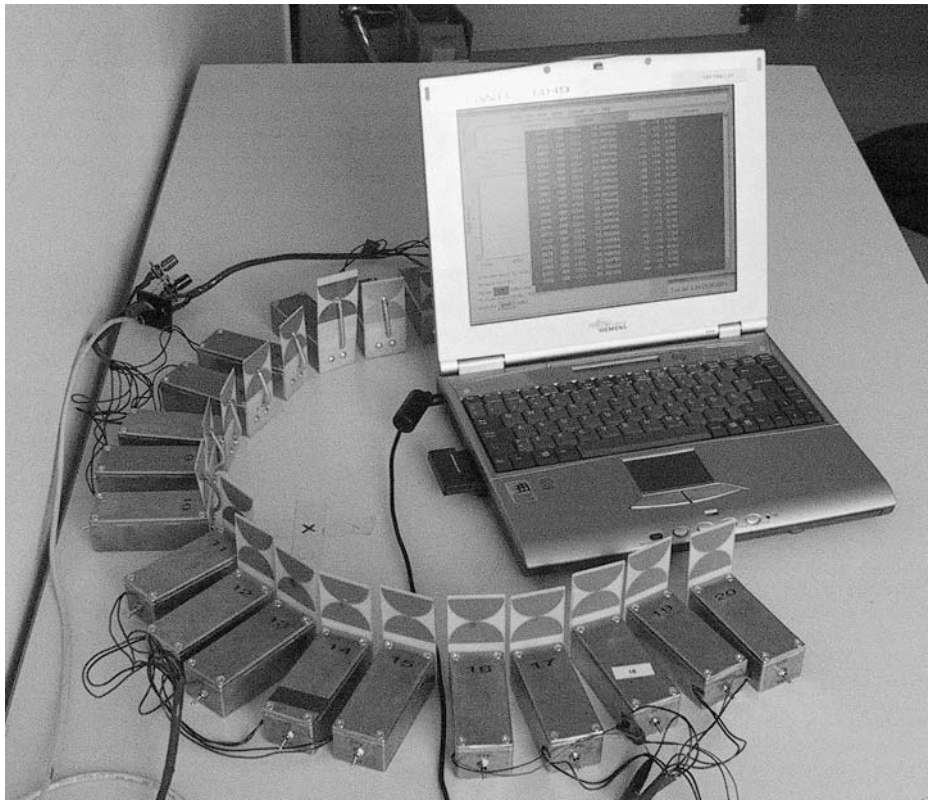


Fig. 5.20 Experimental setup used to find the interference to a wireless LAN card from high-powered UWB transmitters ©IEEE 2003.

Throughput measurements showed a similar trend: with distances less than 30 cm, WLAN throughput dropped dramatically when 15 or more high-powered UWB transmitters were used and if the distance was greater than 40 cm the deterioration was negligible.

5.10.2 Bluetooth

In [64] the performance of Bluetooth wireless LAN networks were considered under the influence of high-powered UWB signals.

It was concluded that the Bluetooth connection suffered less than the corresponding 802.11b channel. The reason for this was that the fixed pulse repetition frequency was used for the UWB transmitters, which gave a distinct line spectrum.

Since Bluetooth devices are able to monitor individual channel states these "bad" channels can be avoided. Only a degradation in throughput from 530 kbps to approximately 490 kbps was recorded for Bluetooth transceivers separated by 10 m, and no degradation in throughput was recorded for Bluetooth devices separated by 3 m. In these measurements the UWB devices were located in an arc 15 cm from the Bluetooth receiver.

5.10.3 GPS

In [65] a report into the interference of UWB transmitters on both high-grade aviation global positioning system (GPS) receivers and also low-cost OEM GPS receivers was made.

The report concluded that UWB transmitters could indeed affect GPS receivers and that care must be taken for transmission at GPS frequencies. Based on this report and others the FCC mandated a strict limit on emissions within the GPS frequency band.

In 1999 the U.S. Department of Transportation (DOT) approached Stanford University to research the compatibility of UWB and GPS and to conduct tests to help quantify any interference problems. GPS plays a major role in many commercial, military, and public systems. For example, aircraft rely on the information from GPS, as do most modern vehicles equipped with navigation systems. Mobile phones equipped with GPS receivers have been commercialized. It is anticipated that reliance on GPS and similar satellite systems will increase in the future.

The majority of the tests performed by Stanford University measured the UWB impact on the accuracy and loss-of-lock performance of a high-grade GPS aviation receiver. A smaller test set measured the UWB impact on the loss-of-lock performance for two different receivers: an aviation receiver and a low-cost commercial receiver. Tests were also undertaken to measure the UWB impact on the signal acquisition performance of a high-grade, general-purpose GPS receiver. This third receiver used the same hardware as the aviation receiver, but the firmware was changed so that the receiver did not utilize an acquisition strategy suited for aircraft dynamics.

It was found that spectral lines present in any UWB signal degraded the GPS signal significantly. A 17-dB degradation was measured without making any effort to place the UWB signals on the more sensitive GPS spectral lines. In practice, UWB lines will frequently find more sensitive lines than those found in these trials because (a) many GPS satellites will be in view and (b) the Doppler frequency for each satellite will change as the satellite moves across the sky, causing the frequency of the more sensitive lines to shift. Eventually, sensitive lines from one satellite or another will fall on the spectral lines from any nearby UWB transmitter that has such lines. The worst line for GPS satellite PRN 21 is 6.5 dB more sensitive than the victim line in these measurements.

Under the best circumstances, UWB signals with high PRFs appear as broadband noise. In other words, the equivalence factor is approximately 0 dB but only in the absence of in-band spectral lines. If UWB dithering codes or modulation indices are not chosen carefully and some spectral line content remains, then the UWB waveform is more damaging than white noise.

It was found that trends were observed in all three receivers: aviation receiver, general-purpose receiver and commercial receiver.

All tests showed the same sensitivity to UWB signal type. For example, the worst interference cases for all three receivers occurred when a discrete UWB spectral line fell into the GPS band. However, OEM tests must be more carefully interpreted because the OEM front end bandwidth is significantly narrower than the bandwidth for the aviation receiver and the standard filter used to measure noise power.

5.10.4 Cellular systems

In comparison with the GPS and other indoor communications techniques the effects of wireless UWB systems on cellular mobile telephone services has not been well covered in the literature. This is probably not surprising, recalling Figure 5.19. Most current mobile telephone systems fall below 1 GHz and, thus, are not in the frequency band in which it is anticipated that most UWB communication systems will use. The 1.5-GHz cellular band is heavily protected by current FCC regulations. It can be anticipated that 2-GHz services will be somewhat affected; however, this is at the edge of the main UWB bandwidth and is not as popular as other cellular services at the present time.

Some experiments have been performed. For example, in [66] the effect of a prototype UWB communication system meeting FCC requirements on a 1.9-GHz PCS band mobile phone was examined. Although the tests were subjective voice quality only, no discernible difference was found by the people making the telephone calls. In these tests the mobile phone was located approximately 1.5 m away from the UWB transmit antenna.

5.11 SUMMARY

In this chapter we have examined ultra wideband wireless communications.

Modulation methods including pulse position modulation, bi-phase modulation, orthogonal pulse modulation, and their combinations was presented. Receiver design and pulse detection was examined.

Multiple access techniques for UWB communications were examined.

A simple derivation of the capacity of the wireless UWB channel was examined. The effect of UWB on existing wireless communication methods, such as the IEEE 802.11 wireless local area network standards and Bluetooth, was examined. Several

methods to prevent interference from these narrowband systems on UWB was examined, in particular the use of filters and multi-band UWB.

Finally, a brief discussion on the relative merits and demerits of UWB as a communication method with other wideband communication techniques, such as CDMA and OFDM, was undertaken.

Problems

Problem 1. Investigate pseudo-noise (PN) codes and summarize three different codes. What is the processing gain of each PN code? Apply each of the codes to:

(a) an unmodulated pulse train;

(b) a 2-PPM pulse train; and

(c) a 4-PPM pulse train.

Compare the spectra obtained with 100 pulses and comment.

Problem 2. Implement the Matlab circuit-generating multiple orthogonal pulse shapes of Figure 5.9. Discuss the potential problems that might be present in practice.

Problem 3. In Figure 5.10 the capacity versus number of users is shown for a spreading ratio of $\beta = 50$. Calculate the curves for $\beta = 5$ and $\beta = 500$. At what number of users does the capacity begin to decrease? You may assume a pulse width of $T_p = 1$ ns and the transmission symbol rate $R = 1/(\beta T_p)$.

Problem 4. Calculate the aggregate capacity C of all traffic on the multiple access channel for all the PPM types shown in Figure 5.10. The aggregate capacity can be calculated from $C = N_u C_{M-PPM}$.

6

Ultra wideband antennas and arrays

Ultra short-pulse technologies are today generally considered practical only for low-power and short-range applications due to the limitations of pulse-forming electronics; however, this state of affairs is quickly changing. Building UWB communication systems requires a theoretical basis for computation and estimation of antenna design parameters and predicting performance. This is particularly important for designing antenna systems that determine the performance of precision range and direction measurements. In fact, antenna design is an important UWB radio frequency challenge. While wide bandwidth types of antennas are well understood in other applications, such as radar, UWB communication systems require an antenna with a flat group delay and a small size, so that the high and low-frequency signal components arrive at the receiver simultaneously and the antenna can fit into consumer electronics products like digital cameras and camcorders. It is expected that an appropriate antenna configuration should be part of a UWB chipset's reference design.

This chapter is about UWB antennas and arrays. These kinds of antennas are specifically designed to transmit and receive very short-time duration pulses of electromagnetic energy. An impulse antenna requires consideration of some concepts that may be new to many readers. The antenna's components act as distributed reactances and change the shape of the incoming waveforms that excite the antenna. Beamforming techniques for nonsinusoidal UWB signals are also considered in this chapter. In addition, radar UWB array systems and their properties are briefly discussed. It is believed that UWB antenna design remains one of the major issues in the progress of UWB technology.

6.1 ANTENNA FUNDAMENTALS

An antenna is a device that transmits and/or receives electromagnetic waves. Electromagnetic waves are often referred to as radio waves. Most conventional antennas are resonant devices that operate efficiently over a relatively narrow frequency band. An antenna must be tuned to the same frequency band as the radio system to which it is connected operates in, otherwise reception and/or transmission will be impaired. In this section we discuss some fundamentals of antenna systems, such as Maxwell's equations, far field and near-field regions, antenna patterns, and antenna gain.

6.1.1 Maxwell's equations for free space

Maxwell's equations represent one of the most elegant and concise ways of stating the fundamentals of electricity and magnetism. From them we can develop most of the working relationships in the field. Because of their conciseness they embody a high level of mathematical sophistication and are therefore not generally introduced in an introductory treatment of the subject, except perhaps as summary relationships.

Maxwell's equations tell us that a time-varying current can radiate as a free space electromagnetic field. The purpose of any antenna is to act as a launching means between guided and free space electromagnetic waves or vice versa. The "guiding" might be in the form of a wire, a coaxial cable, a stripline, or an actual waveguide.

All four Maxwell equations are available in two forms of integration and differentiation [67].

6.1.1.1 Gauss's law for electricity The electric flux out of any closed surface is proportional to the total charge enclosed within the surface

$$\text{Integral form:} \qquad \oint_A \mathbf{E} \cdot d\mathbf{A} = \frac{Q}{\epsilon_0} \tag{6.1}$$

where $\mathbf{E}$ is the electric field (in units of V/m), $d\mathbf{A}$ is the area of a differential square on the surface A with an outward-facing surface normal defining its direction, Q is the charge enclosed by the surface, and ϵ_0 (approximately 8.854 pF/m) is the permittivity of free space. Note that the integral form only works if the integral is over a closed surface.

The differential form (Eqn. 6.1) can be written as

$$\text{Differential form:} \qquad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{6.2}$$

where $\nabla \cdot \mathbf{E}$ is the divergence of $\mathbf{E}$ and ρ is the charge density.

The integral form of Gauss's law finds application in calculating electric fields around charged objects. By applying Gauss's law to the electric field of a point charge we can show that it is consistent with Coulomb's law. While the area integral of the electric field gives a measure of the net charge enclosed the divergence of the electric field gives a measure of the density of sources. It also has implications for the conservation of charge.

6.1.1.2 Gauss's law for magnetism The net magnetic flux out of any closed surface is zero

$$\text{Integral form:} \qquad \oint_A \mathbf{B} \cdot d\mathbf{A} = 0 \tag{6.3}$$

where $\mathbf{B}$ is the net magnetic flux (in units of tesla, T). Similar to Eqn. (6.2) we can write

$$\text{Differential form:} \qquad \nabla \cdot \mathbf{B} = 0 \tag{6.4}$$

These equations are related to the magnetic field's structure because it states that, given any volume element, the net magnitude of vector components that point outward from the surface must be equal to the net magnitude of vector components that point inward. Structurally, this means that magnetic field lines must be closed loops.

Another way of putting it is that field lines cannot originate from somewhere; attempting to follow the lines backward to their source or forward to their terminus ultimately leads back to the starting position. This implies that there are no *magnetic monopoles*.

6.1.1.3 Faraday's law of induction The line integral of the electric field around a closed loop is equal to the negative of the rate of change of the magnetic flux through the area enclosed by the loop

$$\text{Integral form:} \qquad \oint_L \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial \phi}{\partial t} \tag{6.5}$$

where $d\mathbf{l}$ is the differential length vector on the closed loop L and ϕ is the magnetic flux through the area A enclosed by the loop. The relation between ϕ and $\mathbf{B}$ can be written as:

$$\phi = \int_A \mathbf{B} \cdot d\mathbf{A} \tag{6.6}$$

This equation only works if surface A is not closed, because the net magnetic flux through a closed surface will always be zero. The line integral of Eqn. (6.5) is equal to the generated voltage or electromotive force (emf) in the loop, so Faraday's law

relates electric and magnetic fields and is the basis for electric generators. It also forms the basis for inductors and transformers.

Note that the negative sign is necessary to maintain conservation of energy. It is so important that it even has its own name, Lenz's law.

The differential form of Faraday's law of induction can be expressed in the following equation

$$\text{Differential form:} \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{6.7}$$

where $\nabla \times \mathbf{E}$ is the curl of $\mathbf{E}$.

6.1.1.4 Ampère's law The line integral of the magnetic field around a closed loop is comprised of two components. The first one is proportional to the electric current flowing through the loop and the second one is proportional to the rate of time variations of the integral of the electric flux out of the open surface A. This law is useful for calculation of the magnetic field for simple geometries

$$\text{Integral form:} \qquad \oint_L \mathbf{B} \cdot d\mathbf{l} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_A \mathbf{E} \cdot d\mathbf{A} \tag{6.8}$$

where I is total electric current and μ_0 is the permeability of the free space defined to be exactly $4\pi \times 10^{-7}$ W/Am. Usually the second term on the right-hand side is negligible and ignored, hence the integral form is known as *Ampère's law*.

The differential form of this law can be derived as follows

$$\text{Differential form:} \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{6.9}$$

where $\mathbf{j}$ is the current density.

6.1.2 Wavelength

Generally, the antenna size is referred relative to the signal wavelength if we are dealing with a narrowband system. For example, a half-wave dipole is approximately half a wavelength long. Wavelength is the distance a radio wave travels during one cycle. The formula for wavelength is

$$\lambda = \frac{c}{f} \tag{6.10}$$

where λ is the wavelength and is expressed in units of length, c is the speed of light in the medium, and f is the frequency.

For UWB signals we do not have a single operating frequency, hence wavelengths of these sorts of signals are usually expressed as the lower and upper available wavelengths in the system.

6.1.3 Antenna duality

Most antennas work equally well in both directions, being able to transmit or receive signals. The mathematical fundamentals of antennas work equally well either way. The only limits for duality of antenna are power and overload restrictions.

6.1.4 Impedance matching

For efficient transfer of energy, impedance of the radio, antenna, and transmission line connecting the radio to the antenna must be the same. For instance, radios typically are designed for 50-Ω impedance and the coaxial cables (transmission lines) used with them also have a 50-Ω impedance. Efficient antenna configurations often have an impedance other than 50 Ω, and Some sort of impedance-matching circuit is therefore required to transform the antenna impedance to 50 Ω.

6.1.5 VSWR and reflected power

The voltage standing wave ratio (VSWR) is a well-known indication of how good the impedance match is. The VSWR is often abbreviated as SWR. A high VSWR is an indication that the signal is reflected prior to being radiated by the antenna. VSWR and reflected power are different ways of measuring and expressing the same thing.

A VSWR of 2:1 or less is considered good. Most commercial antennas, however, are specified to be 1.5:1 or less over some bandwidth. Based on a 100-watt radio, a 1.5:1 VSWR equates to a forward power of 96 watts and a reflected power of 4 watts, or the reflected power is 4.2% of the forward power.

6.1.6 Antenna bandwidth

Bandwidth can be defined in terms of radiation patterns or VSWR/reflected power. Bandwidth is often expressed in terms of percent or fractional bandwidth (FB), because the percent bandwidth is constant relative to frequency. In this case, if the bandwidth is expressed in absolute units of frequency (e.g., GHz) the FB is then different depending upon whether the frequencies in question are near 3, 4, or 8 GHz.

By definition the FB of a signal is the ratio of the bandwidth to the center frequency as follows:

$$\mathrm{FB} = \frac{f_h - f_l}{(f_h + f_l)/2} \times 100\% \tag{6.11}$$

where f_h and f_l are the effective highest and lowest frequency components of the signal, respectively. Wideband arrays are designed with FB of up to 25% and UWB arrays are proposed with FB of 25 to 200%.

6.1.7 Directivity and gain

Directivity is the ability of an antenna to focus energy in a particular direction when transmitting or to receive energy in a better way from a particular direction when receiving. The relationship between gain and directivity is based on efficiency

$$\text{Gain} = \frac{\text{Efficiency}}{\text{Directivity}} \tag{6.12}$$

We can see the phenomena of increased directivity by comparing a light bulb with a spotlight. A 100-watt spotlight will provide more light in a particular direction than a 100-watt light bulb and less light in other directions. We could say the spotlight has more "directivity" than the light bulb. The spotlight is comparable with an antenna with increased directivity. An antenna with increased directivity that is hopefully implemented efficiently, is low loss, and, therefore, exhibits both increased directivity and gain.

Gain is given in reference to a standard antenna. The two most common reference antennas are the *isotropic antenna* and the *resonant half-wave dipole antenna*. The isotropic antenna radiates equally well in all directions. Real perfect isotropic antennas do not exist, but they provide useful and simple theoretical antenna patterns with which to compare real antennas. An antenna gain of 2 (3 dB) compared with an isotropic antenna would be written as 3 dBi. The resonant half-wave dipole can be a useful standard for comparison with other antennas at one frequency or over a very narrow band of frequencies. To compare the dipole with an antenna over a range of frequencies requires an adjustable dipole or a number of dipoles of different lengths. An antenna gain of 1 (0 dB) compared with a dipole antenna would be written as 0 dBd.

6.1.8 Antenna field regions

Most antennas have two operating regions which are called the *near field* and the *far-field* regions. These are sometimes called the Fraunhofer and Fresnel regions. In the antenna near field, behavior is highly complex and most energy drops off with the cube of distance. In the antenna far field, properties are more orderly and most energy falls off with the square of the distance. The crossover between near field and far field takes place at $2L^2/\lambda$ or around a wavelength for a normal antenna, where L is the antenna width. Figure 6.1 shows how the field strength drops with distance on most typical antennas.

6.1.9 Antenna directional pattern

Directional (radiation) pattern of an antenna or an array of antenna elements describes the relative strength of the radiated field in various directions from the antenna or array, at a fixed or constant distance. The radiation pattern is a

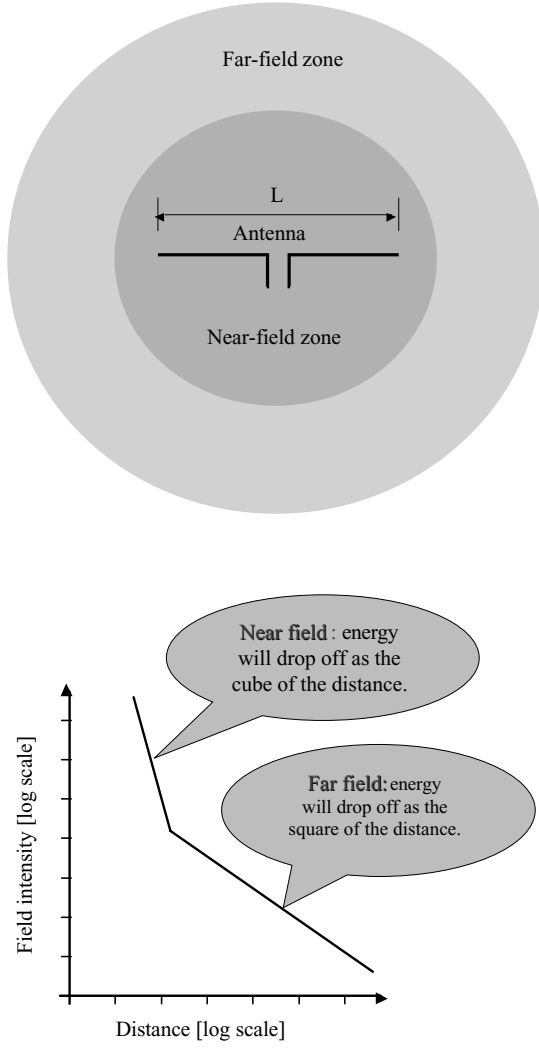


Fig. 6.1 Typical antennas have near field and far-field regions. The behavior of the two regions is radically different. Near-field mathematics is quite complex, whereas far-field mathematics is more orderly.

"reception pattern" as well, since it also describes the receiving properties of the antenna.

The radiation pattern is three-dimensional, but it is difficult to display the three-dimensional radiation pattern in a meaningful manner; it is also time-consuming

to measure a three-dimensional radiation pattern. Often, radiation patterns are a slice of the three-dimensional pattern, which is of course a two-dimensional radiation pattern that can be displayed easily on a screen or piece of paper. These pattern measurements are presented in either a rectangular or a polar format.

An antenna pattern defined by uniform radiation in all directions and produced by an isotropic radiator (or a point source: a nonphysical antenna which is the only nondirectional antenna) is called an *isotropic pattern*. A pattern which is uniform in a given plane is called an *omni-directional pattern*.

Generally, it is desired that an antenna send or gather in energy in a particular direction, thereby creating a special antenna radiation pattern. Patterns can be created by a single antenna element or by the arrangement and size of a set of antenna elements. For instance, a TV satellite antenna must have a very narrow beamwidth, because the desired satellite has a weak signal in any other direction than that directly toward the satellite. However the antenna of a global positioning system (GPS) navigation receiver usually has to follow six moving satellites at once. Hence, the antenna should have a half-hemispherical pattern that can receive equally well from any point in the sky.

Terrestrial TV transmitters create a "bagel" pattern since transmitting energy up toward the sky or into the earth is wasteful. Multi-tower AM broadcast antennas purposely design nulls toward neighboring stations to lower interference.

Directional patterns of wideband or UWB antennas and arrays are always functions of frequency. Sophisticated antenna design techniques and signal-processing algorithms have to be proposed in order to solve the problem of the frequency dependence of antennas and arrays.

6.1.10 Beamwidth

Depending on the radio system in which an antenna is being employed there can be many definitions of beamwidth. A common definition is the half-power beamwidth. The peak radiation intensity is found, and then the points on either side of the peak representing half the power of the peak intensity are located. The angular distance between the half-power points travelling through the peak is the beamwidth. Half the power is 3 dB, so the half-power beamwidth is sometimes referred to as the 3-dB beamwidth.

Inter-null beamwidth can also be used as the measure of the directivity of the antenna. By definition, inter-null beamwidth is the difference between the two angles corresponding to the first nulls on either side of the peak radiation.

Figure 6.2 shows the directional pattern parameters of an antenna (array). As indicated in this figure the *main lobe*, which is also called the *major lobe* or *main beam*, is the radiation lobe in the direction of maximum radiation. The *side lobe* is a radiation lobe in any direction other than the direction(s) of the intended radiation. The *back lobe* is defined as the radiation lobe opposite the main lobe.

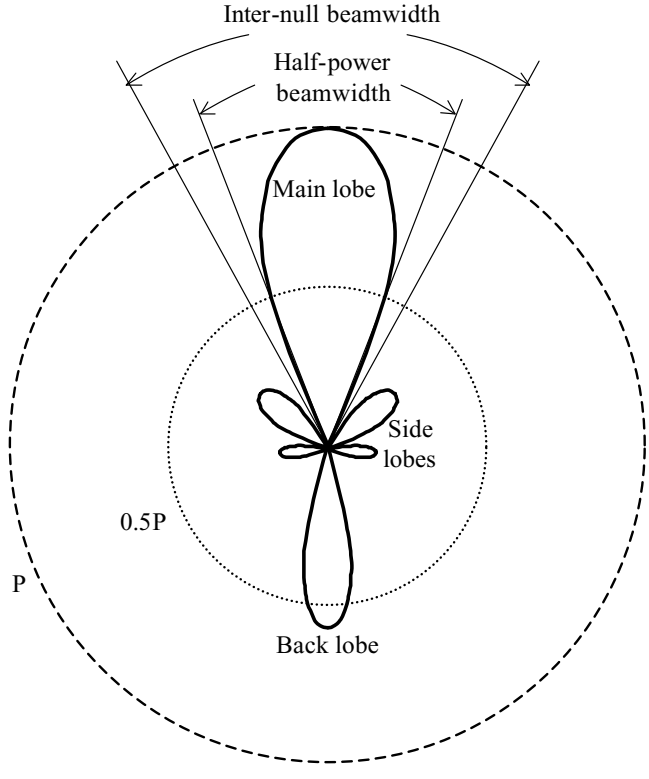


Fig. 6.2 Antenna directional pattern parameters. It is assumed that the power at the desired direction is P. Hence, the half-power circle is identified by $0.5P$ and determines the half-power beamwidth.

6.2 ANTENNA RADIATION FOR UWB SIGNALS

The radiation for short-duration UWB signals from an antenna is significantly different compared with the radiation produced by long-duration narrowband signals. These differences are mostly due to the following parameters [68]:

- The characteristics of the radiation of the signal from the antenna aperture.
- The time domain attributes of the radiation field.
- The amplitude of the radiation field.
- The spatial attributes of the radiation field.
- The properties of side radiation.

The development of new UWB sources and antennas has shown significant progress in recent years. In particular, research in impulse-radiating antennas has improved the performance of a number of other wideband systems. The development of a theoretical description of impulse-radiating antennas has been considered for some years, and theoretical models are now available for far and near-field radiations of various antenna types (e.g. reflector, horns, and antenna arrays).

Due to the strong impact of the waveform on the coupling of transient fields into a system the parameters of the waveform, such as rise time or pulse duration, are often more important than the pulse amplitude for susceptibility investigations. Based on the limited amplitude range of high-power pulse sources, field magnitudes are often scaled by changing the distance between the antenna and target or by changing the direction angle. The radiated impulse shape depends on the direction angle (e.g., rise time decreases for increasing direction angles).

We start with a short review of an analytical model for the radiation of aperture antennas. Following this a review of the frequency domain far-field definition is performed. Based on these theoretical reviews a time domain far-field definition will be developed with respect to waveform effects. Comparison of the developed definition and the classical frequency domain definitions shows that for fast transient radiation the well-known frequency domain estimation models fail because the dimensions of antenna systems and wave propagation lead to a special kind of dispersion. The importance of this near-field dispersion for practical applications has been demonstrated by both theoretical considerations and measurements with a half-impulse radiation antenna and an 8×8 antenna array [69].

The radiation behavior of most impulse-radiating antennas can be represented by the electrical field ($\boldsymbol{E_a}$) inside the aperture (A_a) using the equivalence principle [70]. In the frequency domain the electrical field strength at an arbitrary observation point r_p can be calculated by the integral

$$\boldsymbol{E_s}(\boldsymbol{r_p}, s) = \frac{1}{2\pi} \int_{A_a} \frac{sR + c}{cR^2} [(\boldsymbol{e_z} \times \boldsymbol{E_a}(\boldsymbol{r_q}, s)) \times \boldsymbol{e_R}] e^{-sR/c} \, dA_a \tag{6.13}$$

where $\boldsymbol{e_z}$ and $\boldsymbol{e_R}$ are the unit vectors of the z-axis and $\boldsymbol{R}$, respectively. The complex frequency is denoted by $s = p + j\omega$ and the speed of light is c. The source point is denoted by r_q and the radiation distance is calculated as follows

$$\begin{aligned} R &= |\boldsymbol{r_p} - \boldsymbol{r_q}| \\ &= r_p - \frac{2\boldsymbol{r_p} \cdot \boldsymbol{r_q}}{r_p + \sqrt{r_p^2 + r_q^2 - 2\boldsymbol{r_p} \cdot \boldsymbol{r_q}}} + \frac{r_q^2}{r_p + \sqrt{r_p^2 + r_q^2 - 2\boldsymbol{r_p} \cdot \boldsymbol{r_q}}} \end{aligned} \tag{6.14}$$

Equation (6.13) describes the radiated field under both near and far-field conditions.

For a large observation distance (far field) the approximations

$$R = \sqrt{r_p^2 + r_q^2 - 2\boldsymbol{r_p} \cdot \boldsymbol{r_q}} \approx r_p - 2\boldsymbol{r_p} \cdot \boldsymbol{r_q} \tag{6.15}$$

and

$$\frac{sR + c}{cR^2} \approx \frac{s}{cr_p} \tag{6.16}$$

can be used to simplify Eqn. (6.13)

$$\boldsymbol{E}_{\mathrm{Far}}(\boldsymbol{r_p}, s) = \frac{1}{2\pi c r_p} e^{-(s/c)r_p} \int_{A_a} [(\boldsymbol{e_z} \times \boldsymbol{E_a}(\boldsymbol{r_q}, s)) \times \boldsymbol{e_p}] e^{2(s/c)\boldsymbol{r_p}\cdot\boldsymbol{r_q}}\, dA_a \tag{6.17}$$

where $\boldsymbol{e_p}$ is the unit vector of $\boldsymbol{r_p}$. This far-field formulation for the radiated electrical field strength consists of two independent terms. The first one describes wave propagation (phase shift and amplitude attenuation) as a function of observation distance. The second term, the integral over the aperture area, is a function of the field distribution inside the aperture and the direction angle. In the far-field formulation just the phase (or time delay) and the attenuation of the radiated field are functions of the observation distance r_p. As a result the amplitude of the radiated field can be scaled by changing the distance if the far-field approximation is applicable. Because an error less than 5% is acceptable for normal applications the recent parameter of interest is the observation distance at which the errors caused by the approximations of Eqns. (6.15) and (6.16) are smaller than this limit.

Considering the far-field approximations, let us assess the approximation errors in the frequency domain first. The approximation (6.15) yields a relative phase shift of

$$\Delta\varphi_{\mathrm{rel}} \leq \frac{r_n^2}{1 + \sqrt{1 + r_n^2}} \tag{6.18}$$

with the normalized source distance $r_n = r_q/r_p$. It is simple to predict that, if $r_n \leq 0.3$ ($r_p \geq 3r_q$), the relative phase shift will be smaller than 5%.

The amplitude error resulting from the approximation (6.16) can be calculated by

$$\Delta_{\mathrm{rel}} = \frac{\lambda}{r_p 2\pi(1 + r_n^2)} + \frac{1 - \sqrt{1 + r_n^2}}{\sqrt{1 + r_n^2}} \tag{6.19}$$

Using the determined ratio between the observation distance and the source point, amplitude error will be smaller than 5% if $r_p \geq 3\lambda$. Note that for the case $r_n \leq 0.1$ and $r_p \geq 12\lambda$ the approximation error will be less than 1%.

6.2.1 Dispersion due to near-field effects

After the well-known frequency domain limits of far-field approximations have been expressed the behavior of fast transient radiation can be investigated. Assuming

$$\frac{\partial E_a}{\partial t} \gg \frac{E_a c}{r_p}$$

the time domain expression of Eqn. (6.13) is derived and

$$\begin{aligned}\boldsymbol{E_s}(\boldsymbol{r_p}, t) = \frac{1}{2\pi c r_p} \int_{A_a} \left[\left(\boldsymbol{e_z} \times \frac{\partial}{\partial t} \boldsymbol{E_a} \left(t + \frac{\boldsymbol{r_p} \cdot \boldsymbol{r_q}}{c} - \frac{r_p}{c} \right) \right) \times \boldsymbol{r_p} \right] \\ * \ \delta \Bigg(t + \frac{\boldsymbol{e_p} \cdot \boldsymbol{r_q}}{c} \frac{1 - \sqrt{1 + r_n^2 - 2\boldsymbol{e_p} \cdot \boldsymbol{e_q} r_n}}{1 + \sqrt{1 + r_n^2 - 2\boldsymbol{e_p} \cdot \boldsymbol{e_q} r_n}} \\ - \frac{r_q}{c} \frac{r_n}{1 + \sqrt{1 + r_n^2 - 2\boldsymbol{e_p} \cdot \boldsymbol{e_q} r_n}} \Bigg) \, dA_a \end{aligned} \tag{6.20}$$

where $\boldsymbol{e_q}$ is the unit vector of $\boldsymbol{r_q}$ and " $*$ " describes the convolution with respect to time, and $\delta(\cdot)$ is the Dirac delta function. The convolution with the delayed Dirac delta function yields a wider pulse in the near and intermediate field. This near-field dispersion can be ignored if the delay

$$\Delta t \leq \frac{r_q}{c} \frac{r_n}{1 + \sqrt{1 + r_n^2}} \tag{6.21}$$

between the shortest path and the length of a path from an arbitrary point on the aperture is short compared with the rise time of the radiated signal. Using the frequency domain far-field definition we get a delay of $\Delta t = 0.15 r_q / c$. Particularly for fast transient signals and large-aperture cross-sections, this time delay is not acceptable. For a distance-independent pulse shape of transient radiation it is necessary that Δt is small compared with the rise time of the radiated signal. For practical applications the rise time should be six times larger than the longest time delay ($\tau_r \geq 6\Delta t$). Based on this the transient far field is given by

$$r_p \geq \frac{3 r_q^2}{c \tau_r} = \frac{r_q}{c} \cdot \frac{3 r_q}{\tau_r} \tag{6.22}$$

Unlike the frequency domain equation in this time domain equation the time r_p/c which the wave needs to cross the aperture is weighted with the rise time as a measure of the shortest change in the shape of the signal. By comparing the phase shift with the phase of one oscillation of a continuous wave signal we can get a similar relation in the frequency domain.

6.3 SUITABILITY OF CONVENTIONAL ANTENNAS FOR THE UWB SYSTEM

Antennas can be classified as either *resonant* or *nonresonant*, depending on their design. In a resonant antenna, if the antenna works at its resonance frequency, almost all of the radio signal fed to the antenna is radiated. However, if the antenna is fed with a frequency other than a resonant one, a large portion of the signal will not be radiated. In case of a resonant antenna, if the frequency range is very wide, a separate antenna must be made for each frequency. On the other hand, a nonresonant antenna can cover a wide range of frequencies. However, special care must be taken when designing the antenna to make it efficient. Moreover, the physical size of currently available nonresonant antennas is not appropriate for portable UWB devices.

It is very difficult to compare a UWB antenna with a conventional antenna, because the traditional performance considerations are based upon continuous wave or narrowband theory. Basic concepts should be kept in mind when a conventional approach is used for UWB technology.

The goal of the UWB antenna designer is to design an antenna with a small size and a simple structure that can produce low distortions, but can provide large bandwidth and omni-directional patterns [71].

6.3.1 Resonant antennas

The most common and easiest antennas for communications are wire antennas. They are some of the cheapest, simplest, and most flexible antennas for many applications. They can be made of a very thin wire, thicker wire, or a cylinder.

One of the simplest practical antennas is the *dipole antenna* shown in Figure 6.3. A Hertzian electric dipole is shown in Figure 6.4 as a pair of electric charges that vary sinusoidally with time in such a way that at any instant the two charges have equal magnitude but opposite sign.

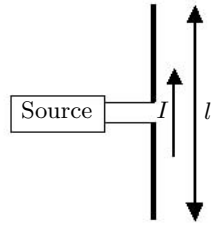


Fig. 6.3 A dipole antenna connected to a source.

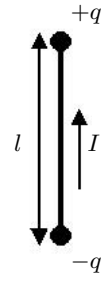


Fig. 6.4 A Hertzian electric dipole.

The Hertzian dipole is an inefficient radiator due to its need of a high voltage to produce a large current. However, this high voltage does not contribute to the radiated power. The resonant dipole (Figure 6.3) was the solution to the inefficiency of the Hertzian dipole (Figure 6.4).

Propagation time is an important parameter in radiating elements. Thus, if an alternating current is flowing in the radiating element as shown in Figure 6.5, the effect of the current is not immediately noticeable at point P, but only after the period of time necessary to propagate over the distance r.

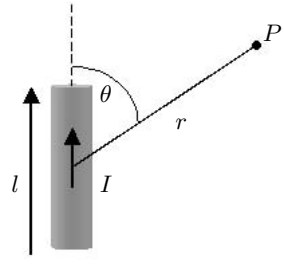


Fig. 6.5 Radiating element.

Current I is the time-varying current and at a distance of $r = 0$ is equal to

$$I = I_0 \sin(\omega t) \tag{6.23}$$

where ω is the radian frequency and t is time. From the Lorentz equations we can introduce the time of propagation. Therefore, Eqn. (6.23) can be rewritten as

$$I_p = I_0 \sin\left[\omega\left(t - \frac{r}{c}\right)\right] \tag{6.24}$$

where c is the speed of light. With respect to Figure 6.5, Eqn. (6.24) shows that current I_p at point P with distance r from the radiating element at time t occurs at the earlier time of $(t - r/c)$.

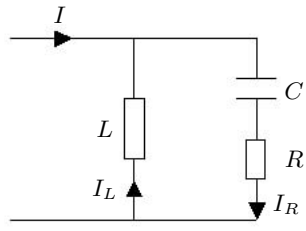


Fig. 6.6 RLC circuit of resonant antenna.

To get a clear understanding of resonant antennas, let us apply a sinusoidal current to the RLC circuit as shown in Figure 6.6. At resonance we will have $\omega^2 LC = 1$, which gives the resonant frequency ω_0 as follows

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{6.25}$$

where L is the inductance and C is the capacitance. At this frequency the current in the inductance I_L can be written as

$$I_L = I_0 \frac{Z}{R}\left(\cos\omega_0 t - \frac{R}{Z}\sin\omega_0 t\right) \tag{6.26}$$

where R is the resistance and Z is the impedance equal to

$$Z = \sqrt{\frac{L}{C}} = \frac{1}{C\omega_0} \tag{6.27}$$

Thus, Eqn. (6.26) can be rewritten as

$$I_L = I\left(\cos\omega_0 t - \frac{R}{Z}\sin\omega_0 t\right) \tag{6.28}$$

where

$$I = I_0 \frac{Z}{R} = \frac{I_0}{RC\omega_0} \tag{6.29}$$

If R is sufficiently small that it can be neglected the term $(R/Z)\sin\omega_0 t$ will be zero. Therefore, Eqn. (6.28) can be rewritten as

$$I_L = I\cos\omega_0 t \tag{6.30}$$

Current I_L shows the behavior of a resonant antenna. The principle of resonance in a resonant antenna has been used to increase the current. If an ultra wideband impulse is fed to this kind of antenna a *ringing* effect will occur. This severely distorts the pulse, spreading it out in time.

Another reason that dipole antennas are not suitable for the UWB system is due to the standing wave produced by the reflection from the end points of the antenna. This is shown graphically in Figure 6.7.

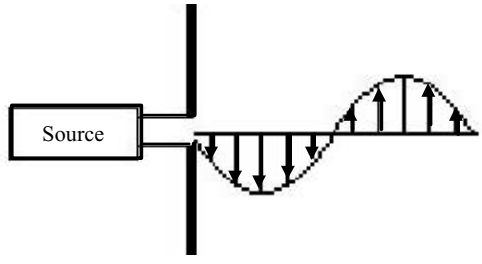


Fig. 6.7 Electromagnetic field and standing wave generated by a dipole antenna.

6.3.2 Nonresonant antennas

An antenna designed to have approximately constant input impedance and radiation characteristics over a wide range of frequencies is called a *nonresonant* antenna, or a *frequency-independent* antenna.

In a nonresonant antenna the maximum dimension (size) will be set by the limit of the lowest frequency to be radiated. The higher frequency limit will be decided by how precisely the input terminal region can be constructed. In fact, a true frequency-independent antenna is only a theoretical construct. In practice, frequency independence is only over a limited frequency bandwidth.

If the impedance and radiation characteristics of the antenna do not change significantly over about an octave or more, the antenna is called a *wideband antenna*.

From sinusoidal wave antenna theory we find that there are many types of antennas which can propagate nonsinusoidal waves. The *log-periodic* antenna and *spiral* antenna are examples of wideband antennas. However, these antennas are likely to be dispersive and inappropriate for very short pulses such as UWB signals. They radiate different frequency components from different parts of the antenna. Therefore, the radiated waveform will be both extended and distorted.

In essence, a trade-off is necessary to obtain an efficient, electrically small antenna which is suitable for the required application.

6.3.3 Difficulties with UWB antenna design

In UWB systems, antennas are significant pulse-shaping filters. Any distortion of the signal in the frequency domain (which is a filtering operation) causes distortion of the transmitted pulse shape, therefore increases the complexity of the detection mechanism at the receiver.

UWB antennas require the phase center and the VSWR to be constant across the whole bandwidth of operation. A change in phase center may cause distortion of the transmitted pulse and worse performance at the receiver. Therefore, the design of antennas for UWB signal radiation is one of the main challenges for

the UWB system, especially when low-cost, geometrically small and radio-efficient structures are required for typical consumer communication applications.

Conventional antennas are designed to radiate only over the relatively narrow range of frequencies used in conventional narrowband systems. If an impulse is fed to such an antenna, it tends to ring, which severely distorts the pulse and spreads it out in time. We have shown that resonant antennas are unsuitable for the UWB system, as they can only radiate sinusoidal waves at the resonance frequency.

On the other hand, to make an effective nonresonant antenna is not an easy task. One way to make a UWB antenna is to put the antenna resonating frequency above the UWB band. However, as the physical dimensions of antennas decrease, antennas will lose efficiency. Another way is to build an antenna with a lower quality factor, which results in a wider bandwidth but with lower efficiency.

The challenge of wide bandwidth antennas is well understood in other applications; however, work on wide bandwidth antennas for UWB applications is ongoing and far from complete.

6.4 IMPULSE ANTENNAS

A few high-quality, lab-grade, nondispersive UWB antennas are commercially available, though targeted at laboratory usage rather than commercial consumer products. One example is Farr Research, Inc. [72] which offers several antennas that operate across several decades of bandwidth. However, the high price range of these antennas make them less suitable for most commercial applications and infeasible for portable or handheld applications. There is a great need for a low-cost, easy-to-manufacture antenna that is omni-directional, radiation-efficient and has a stable UWB response.

6.4.1 Conical antenna

The conical antenna (Figure 6.8) suspended over a large metal ground plane is the preferred antenna for transmitting known transient electromagnetic waves. This type of antenna is used as a reference transient transmitting antenna. It radiates an electromagnetic field that is a perfect replica of the driving point voltage waveform.

The conical antenna can be analyzed as one-half of the bi-conical transmission line [73] with uniform characteristic impedance given by

$$Z_{\text{cone}} = \left(\frac{\eta}{2\pi}\right) \ln\left[\cot\left(\frac{\theta_0}{2}\right)\right] \tag{6.31}$$

where η is the free space impedance (377 Ω), θ_0 is the solid $\frac{1}{2}$ angle of the cone. A 4° cone has an impedance of 200 Ω, while a 47° cone is required to achieve an impedance of 50 Ω.

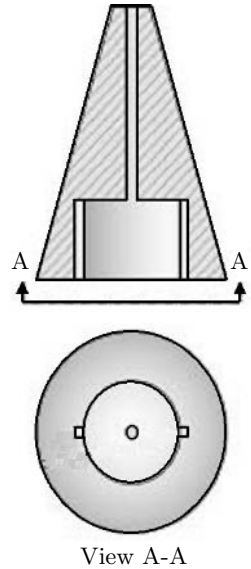


Fig. 6.8 Conical antenna.

The electric field generated at radius r and angle θ is given by

$$E_\theta(r,t) = \frac{V_{\text{base}}(t - r/c)}{r\sin(\theta)\ln[\cot(\theta_0/2)]} \tag{6.32}$$

for $t < l/c$, where l is the length of the antenna.

The driving point voltage at the base of the conical antenna is given by

$$V_{\text{base}}(t) = V_{\text{gen}}(t)\frac{R_{\text{cone}}}{R_{\text{cone}} + R_{\text{gen}}} \tag{6.33}$$

where $V_{\text{gen}}(t)$ is the source voltage, R_{cone} is the cone antenna resistance, and R_{gen} is the source resistance.

Equation (6.32) is only valid for the time window up to $T_w < l/c$. For $t > T_w$ the original incident wave launched on the antenna has reached the end of the antenna and is reflected, thus setting up a series of multiple reflections on the antenna's radiating element. The radiated E field is thus no longer a replica of the driving point voltage, but is corrupted by multiple reflections. If a conical antenna is used as a receiving antenna, its output is the *integral* of the incident E field.

6.4.2 Monopole antenna

The monopole antenna is sometimes used as a simpler version of the conical antenna for transmitting UWB signals that are similar in wave shape to the driving

point voltage. However, its radiated fields are not as uniform as those of the conical antenna. Its driving point impedance is not constant, but increases as a function of time. This leads to distortion of radiated electromagnetic (E-M) fields.

Time domain studies of a monopole antenna confirm this statement. When a monopole is used for receiving transient E-M fields, its output is the *integral* of the incident E field. The integration effect can be explained simply. Assume an impulsive E field is incident upon the monopole at a 90° angle to its axis. Thus, a current is induced simultaneously in each differential element dx of the antenna. These current elements di thus start to flow toward the output connector of the antenna. They do not arrive at the output simultaneously but in sequence. Thus, the output appears as a step function, which is the integral of the incident impulse. The integrating effect of this antenna only lasts for $t < l/c$.

6.4.3 D-dot probe antenna

The D-dot antenna is basically an extremely short, monopole antenna. The equivalent antenna circuit consists of a series capacitance and a voltage generator $V_a(t)$ as follows

$$V_a(t) = h_a E_i(t) \tag{6.34}$$

where h_a is the height of the antenna and $E_i(t)$ is the electrical field. For a very short monopole, antenna capacitance is very small and the capacitor thus acts like a differentiator to transient E-M fields. Therefore, the output from a D-dot probe antenna is the *derivative* of the incident E field.

To determine the actual wave shape of the incident E field we must integrate the output from the D-dot probe. When the frequencies become too high, the D-dot probe loses its derivative properties and it becomes a monopole antenna. This happens when the length of the D-dot probe approaches a quarter wavelength of the incident wave.

6.4.4 TEM horn antenna

Transient electromagnetic (TEM) horns are the most effective receiving antennas for making direct measurement of transient E-M fields (Figure 6.9). The TEM horn antenna is basically an open-ended, parallel plate transmission line. It is typically built using a taper from a large aperture at the receiving input down to a small aperture at the coax connector output. The height-to-width ratio of the parallel plate is maintained constant along the length of the antenna to maintain a uniform characteristic impedance. The output from a TEM antenna is

$$V_{\text{out}}(t) = \frac{h_{\text{eff}} E_i(t) R_{\text{load}}}{R_{\text{ant}} + R_{\text{load}}} \tag{6.35}$$

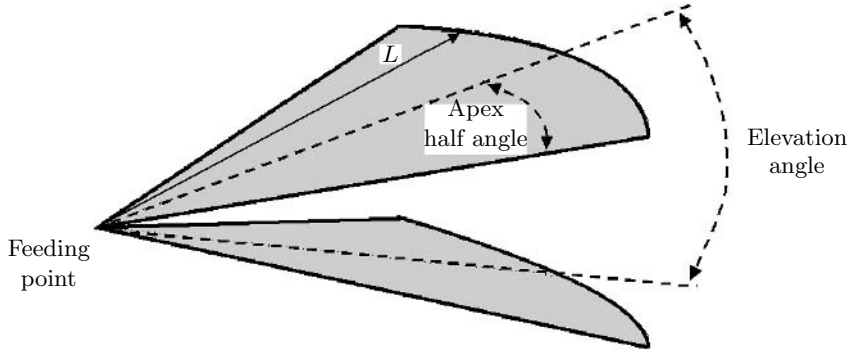


Fig. 6.9 Typical TEM horn antenna with length L.

where h_{eff} is the effective height at the aperture of the antenna, and R_{load} and R_{ant} are load and antenna resistances, respectively.

Ideally, to suppress multiple reflections, antenna impedance should match the 50-Ω output cable. However, to optimize sensitivity most TEM antennas are designed with an antenna impedance of 100 Ω. If a TEM antenna is used as a transmit antenna the radiated E field is the *derivative* of the input driving point voltage. We can see that the TEM horn antenna is capable of radiating and receiving very fast, but still clean pulses, which is of course important for many applications. The cleaner the pulse the cleaner the backscattered signal and the easier it will be to post-process and interpret the data.

6.4.5 Conclusion

Most practical UWB systems in fact will *not* use the antennas discussed above. To meet FCC spectrum requirements, considerable band-pass filtering will be needed to be done.

As a conclusion we can say that, although the classical antennas have proven their use in different applications, none of them can be used for UWB applications and there is a need to look for new types of antennas. To do so we can summarize the following technical and practical design goals for the UWB antenna:

- The antenna must be able to radiate or receive fast electromagnetic transients with frequencies between 3.1 and 10.6 GHz.
- We want an antenna which can be used off ground, not only for safety reasons but also to improve the mobility of the sensor.
- Another criterion to guarantee high mobility is the dimensions and weight of the antenna. Small antennas are also better for handheld applications.
- The antennas must be cheap to produce.

6.5 BEAMFORMING FOR UWB SIGNALS

Array signal processing or beamforming involves the manipulation of signals induced on the elements of an array system or transmitted from the elements and received at a distant point from the array. In a conventional narrowband beamformer the signals corresponding to each sensor element are multiplied by a complex weight to form the array output. As the signal bandwidth increases, the performance of the narrowband beamformer starts to deteriorate because the phase provided for each element and the desired angle is not a function of frequency and, hence, will change for the different frequency components of the communication wave.

For processing broadband signals a tapped delay line (TDL) can normally be used on each branch of the array. The TDL allows each element to have a phase response that varies with frequency, compensating for the fact that lower frequency signal components have less phase shift for a given propagation distance, whereas higher frequency signal components have greater phase shift as they travel the same length. This structure can be considered to be an equalizer that makes the response of the array the same across different frequencies. In addition to UWB signals, inherent baseband signals, such as audio and seismic signals, are examples of wideband signals. In sensor array processing applications, such as spread spectrum communications and passive sonar, there is also growing interest in the analysis of broadband sources and data.

6.5.1 Basic concepts

Most of the smart antennas proposed in the literature are narrowband beamformers. The antenna spacing of narrowband arrays is usually half of the wavelength of the incoming signal which is assumed to have a fractional bandwidth (FB) of less than 1%.

By definition the FB of a signal is the ratio of the bandwidth to the center frequency (Eqn. 6.11). Wideband arrays are designed for an FB of up to 25% and UWB arrays are proposed for an FB of 25 to 200%. Wideband and UWB arrays use the same antenna spacing for all frequency components of the arriving signals. Usually, the inter-element distance d is determined by the highest frequency of the input wave. In a uniform, one-dimensional linear array we can write

$$d = \frac{c}{2f_h} \tag{6.36}$$

Wideband and UWB antenna arrays use a combination of spatial filtering and temporal filtering. On each branch of the array a filter allows each element to have a phase response that varies with frequency. As a result, the phase shifts due to higher and lower frequencies are equalized by temporal signal processing.

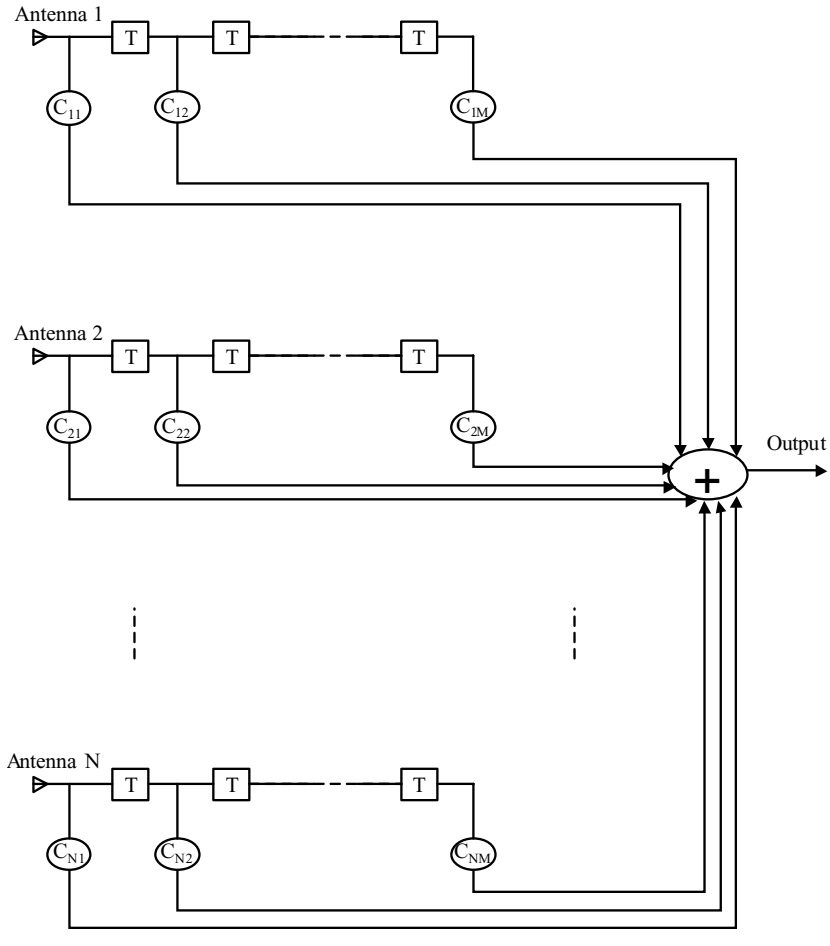


Fig. 6.10 General structure of a TDL wideband array antenna using N antenna elements and M taps.

Figure 6.10 shows the general structure of a wideband array antenna system in *receiving mode*. The TDL network permits adjustment of gain and phase as desired at a number of frequencies over the band of interest. The far-field wideband signal is received by N antenna elements. Each element is connected to $M-1$ delay lines, with the time delay of T seconds. The delayed input signal of each element is then multiplied by a real weight C_{nm}, where $1 \leq n \leq N$ and $1 \leq m \leq M$. If the input signals are denoted by $x_1(t), x_2(t), \ldots, x_N(t)$ the output signal which is

the sum of all intermediate signals can be written as

$$y(t) = \sum_{n=1}^{N} \sum_{m=1}^{M} C_{nm} x_n \left(t - (m-1)T\right) \tag{6.37}$$

In a linear array, such as Figure 6.10 [74], the signals $x_n(t)$ are related according to the angle of arrival and the distance between elements. Figure 6.11 shows that a time delay τ_n exists between the signal received at element n and at reference element $n = 1$. This amount of delay can be found as follows

$$\tau_n = (n-1)\frac{d}{c}\sin\theta \tag{6.38}$$

where d is inter-element spacing and c is propagation speed. It is assumed that the incoming signal is spatially concentrated around the angle θ. Using the time delays corresponding to the antenna elements we can now write $x_n(t)$ with respect to $x_1(t)$ as follows

$$\begin{aligned} x_n(t) &= x_1(t - \tau_n) \\ &= x_1\left(t - (n-1)\frac{d}{c}\sin\theta\right) \end{aligned} \tag{6.39}$$

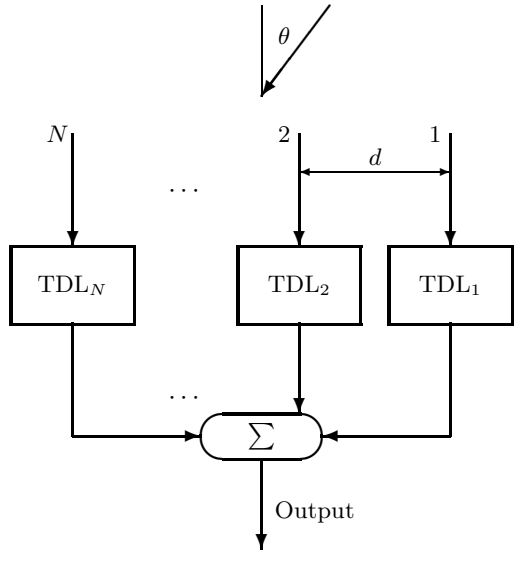


Fig. 6.11 The incoming signal arrives at the antenna array with angle θ.

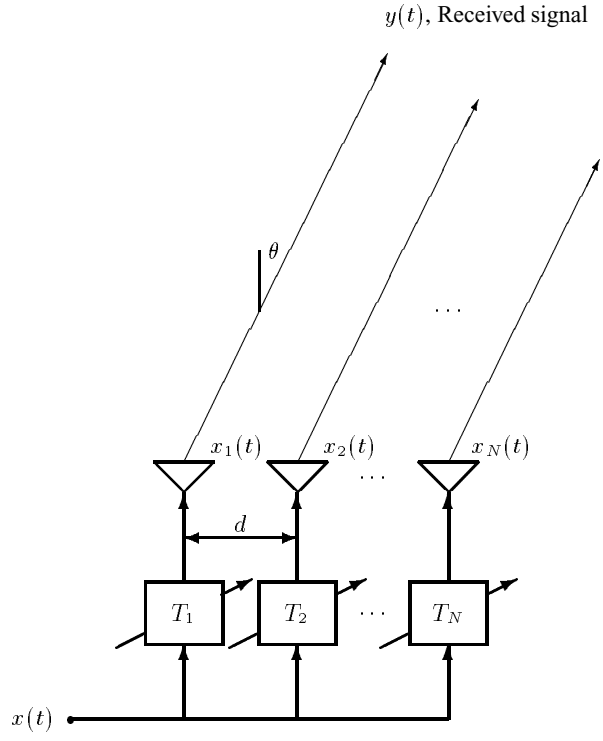


Fig. 6.12 Beam formation using adjustable delay lines.

Different methods and structures can be utilized for computation of the adjustable weights of a wideband beamforming network. In the following section we will explain some of them.

6.5.2 A simple delay-line transmitter wideband array

The problem of designing a uniformly spaced array of sensors for operation at a narrowband frequency domain is well understood. However, when a single frequency design is used over a wide bandwidth the array performance degrades significantly. At lower frequencies the beamwidth increases, resulting in reduced spatial resolution; at frequencies above the narrowband frequency the beamwidth decreases and grating lobes may be introduced into the array beam pattern.

Figure 6.12 shows the basic structure of a delay-line wideband transmitter array system. The adjustable delays T_n, $n = 1, 2, \ldots, N$ where N is the number of antenna, are controlled by the desired angle of the main lobe of the directional beam pattern θ_0 as follows

$$T_n = T_0 + (n-1)\frac{d}{c}\sin\theta_0 \tag{6.40}$$

where d is the inter-element spacing. The constant delay $T_0 \geq (N-1)d/c$ is required because without T_0 a negative delay will be obtained for negative values of θ_0 and cannot be implemented. The signal received at the far field in the direction of θ ($-90^\circ < \theta < +90^\circ$) is equal to

$$\begin{aligned} y(t) &= A(\theta)\sum_{n=1}^{N} x_n(t-\tau_n) \\ &= A(\theta)\sum_{n=1}^{N} x(t-T_n-\tau_n) \end{aligned} \tag{6.41}$$

where $x_n(t)$ indicates the transmitted signal from transducer n, τ_n is the delay due to the different distances between the elements and the receiver, and $A(\theta)$ is the overall gain of the elements and the path. The time delay τ_n regarding Figure 6.12 is equal to

$$\tau_n = \tau_0 - (n-1)\frac{d}{c}\sin\theta \tag{6.42}$$

where τ_0 is the constant transmission delay of the first element and is independent of θ. The gain $A(\theta)$ can be decomposed into two components as follows

$$A(\theta) = A_1(\theta)A_2 \tag{6.43}$$

where $A_1(\theta)$ is the angle-dependent gain of the elements and A_2 is attenuation due to the distance. Substituting Eqns. (6.42) and (6.43) into Eqn. (6.41) yields

$$y(t) = A_1(\theta)A_2\sum_{n=1}^{N} x\left(t-\alpha_0-(n-1)\frac{d}{c}(\sin\theta_0-\sin\theta)\right) \tag{6.44}$$

where $\alpha_0 = T_0 + \tau_0$. In the frequency domain we may write

$$\begin{aligned} H(f,\theta) &= \frac{Y(f,\theta)}{X(f)} \\ &= A_1(\theta)A_2 e^{-j2\pi f\alpha_0}\sum_{n=1}^{N} e^{-j2\pi f(n-1)(d/c)(\sin\theta_0-\sin\theta)} \\ &= A_1(\theta)A_2 e^{-j2\pi f\alpha_0} e^{-j\pi f(N-1)(d/c)(\sin\theta_0-\sin\theta)} \\ &\quad \times \frac{\sin[\pi f N(d/c)(\sin\theta_0-\sin\theta)]}{\sin[\pi f(d/c)(\sin\theta_0-\sin\theta)]} \end{aligned} \tag{6.45}$$

From this equation we can derive several properties of a wideband delay beamformer. An important characteristic of the beamformer is the directional patterns for different frequencies.

Example 6.1

Consider a UWB signal with a center frequency of 6 GHz and a bandwidth of 2 GHz. Calculate and sketch the normalized amplitude of Eqn. (6.45) for $\theta_0 = 10°$, $-90° < \theta < +90°$, $N = 10$, $d = 2.14$ cm, $c = 3 \times 10^8$ m/s, and perfect antennas (i.e., $A(\theta) = A_2$).

Solution

The result is plotted in Figure 6.13. We observe that at $\theta = \theta_0$ frequency independence is perfect, but as we move away from this angle the dependence increases. Nevertheless, the beamformer is considered wideband with a fractional bandwidth of $\frac{2}{6}$ or 33%.

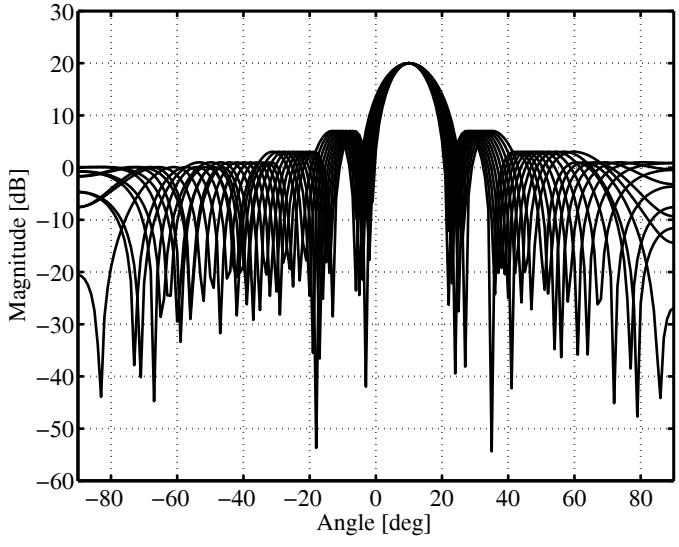


Fig. 6.13 Directional patterns of a delay beamformer for 11 frequencies uniformly distributed from 5 GHz to 7 GHz.

Increasing the inter-element spacing has positive and negative consequences. As we will shortly see, it will produce a sharper beam and it is clearly more practical. On the other hand, this increase will result in some extra main lobes in the same region of interest (i.e., $-90° < \theta < +90°$).

6.5.2.1 Angles of grating lobes We now derive the angles of grating lobes and conditions for their existence. Assuming perfect antennas (i.e., $A(\theta) = A_2$) we can

write from Eqn. (6.45) for $\theta = \theta_0$ as follows

$$|y(f, \theta_0)| = A_2 N \tag{6.46}$$

This situation can happen for some other angles, denoted by θ_g. To calculate θ_g, it follows from Eqn. (6.45) that

$$|y(f, \theta_g)| = A_2 N = A_2 \frac{\sin\left[\pi f N (d/c)(\sin\theta_0 - \sin\theta_g)\right]}{\sin\left[\pi f (d/c)(\sin\theta_0 - \sin\theta_g)\right]} \tag{6.47}$$

Now, Eqn. (6.47) should be solved for θ_g:

$$\sin\left[\pi f \frac{d}{c}(\sin\theta_0 - \sin\theta_g)\right] = 0$$

or

$$\pi f \frac{d}{c}(\sin\theta_0 - \sin\theta_g) = m\pi \tag{6.48}$$

where $m = \pm 1, \pm 2, \ldots$. The result is

$$\theta_g = \sin^{-1}\left(\sin\theta_0 - m\frac{c}{fd}\right) \tag{6.49}$$

The first grating lobes are given for $m = \pm 1$. The necessary condition for having no grating lobe for a beamformer is that θ_g does not exist for any values of $-90^\circ < \theta_0 < +90^\circ$. The worst case happens for $\theta_0 = \pm 90^\circ$ and the condition of no grating lobe can be inferred from Equation (6.49) as

$$\frac{c}{fd} \geq 2, \quad \text{or} \quad d \leq \frac{c}{2f} = \frac{\lambda}{2} \tag{6.50}$$

where λ indicates the wavelength. It is interesting to note that θ_g is not a function of N, but is very dependent on d. To show this more adequately, Figure 6.13 is replotted for $d = 8.57$ cm in Figure 6.14. We observe that the nearest grating lobes for $f = 6$ GHz are at 49.2 and -24.2 degrees and are inconsistent with Eqn. (6.49). The frequency dependence of the beam patterns increases as we move away from the desired angle.

6.5.2.2 Inter-null beamwidth Comparing Figures 6.13 and 6.14 reveals that the main beamwidth of Figure 6.14 is less than that of Figure 6.13. The corresponding equation can be derived easily. The inter-null beamwidth (INBW) is defined as the difference between the nearest two nulls around the desired angle. Starting from Eqn. (6.45) and equating it to zero gives the following

$$\pi f N \frac{d}{c}(\sin\theta_0 - \sin\theta) = m\pi \tag{6.51}$$

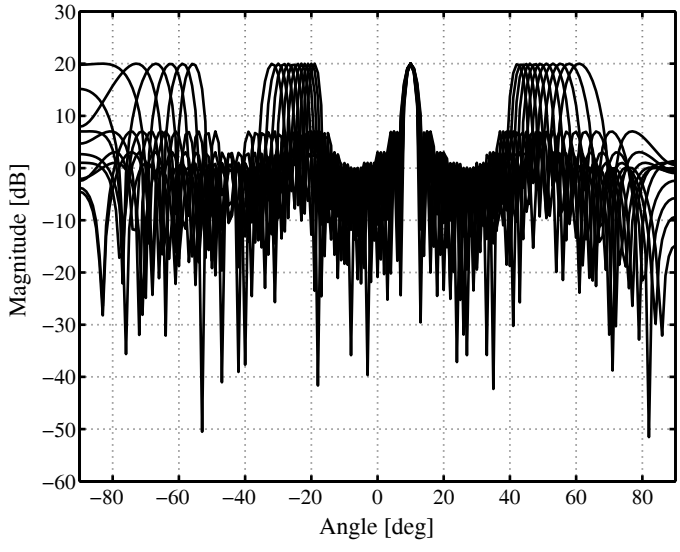


Fig. 6.14 Grating lobes appear as a result of the increase of spacing between antennas.

where $m = \pm 1, \pm 2, \ldots$. The first two angles around θ_0 are denoted by θ_1 and θ_2 and are computed from Eqn. (6.51) for $m = +1$ and $m = -1$, respectively

$$\theta_1 = \sin^{-1}\left(\sin\theta_0 - \frac{c}{fdN}\right) \tag{6.52}$$

$$\theta_2 = \sin^{-1}\left(\sin\theta_0 + \frac{c}{fdN}\right) \tag{6.53}$$

Hence, the INBW, $\Delta\theta = \theta_2 - \theta_1$, is written as

$$\text{INBW} = \sin^{-1}\left(\sin\theta_0 + \frac{c}{fdN}\right) - \sin^{-1}\left(\sin\theta_0 - \frac{c}{fdN}\right) \tag{6.54}$$

It is clear that for $|\sin\theta_0 \pm c/fdN| > 1$ there exists no null on the left or right side of the main angle θ_0. As a special case, for $\theta_0 = 0$ we have

$$\text{INBW}_0 = 2\sin^{-1}\left(\frac{c}{fdN}\right) \tag{6.55}$$

that is, increasing d lowers the INBW and produces sharper beams. It is easy to test Eqn. (6.54) for values of the first and second cases, which are illustrated in Figures 6.13 and 6.14.

As is obvious from Eqns. (6.54) or (6.55), INBW is a function of frequency f. To observe the effect of frequency variations on the beam pattern of the delay-line

beamforming network we repeat *Example 6.1* for a wide frequency range from 4 GHz to 8 GHz and inter-element spacing of 1.87 cm. The results are shown in Figure 6.15 for frequencies of 4, 5, 6, 7, and 8 GHz. The computed values of INBW for these frequencies are 47.3, 37.4, 31, 26.5, and 23.1 degrees, respectively.

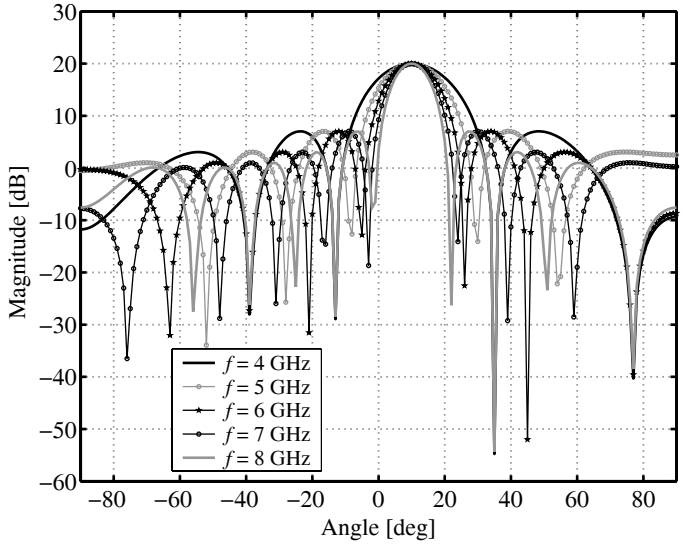


Fig. 6.15 Directional patterns of the delay beamformer for five different frequencies show that the beamwidth is very sensitive to frequency.

From the forgoing discussion we can conclude that pure delay-line wideband antenna arrays have the following properties:

- A relatively simple structure using only a variable delay element.
- No multiplier in the form of amplification or attenuation.
- A perfect frequency independence characteristic only for the desired angle of the array.
- Their INBW and side lobe characteristics vary considerably with frequency of operation.

Because of the existence of some distinctive differences between conventional and UWB antenna arrays the well-known conventional concepts of phased array antennas have to be modified appropriately to accommodate UWB signals. One significant difference from narrowband theory is that frequency domain analysis alone is insufficient to treat UWB arrays. In fact, the time domain may be a more natural setting for understanding and analyzing the radiation of UWB signals.

6.6 RADAR UWB ARRAY SYSTEMS

Advanced radar systems employ array antennas in their passive or active forms to achieve the required total power, high-resolution directive beam pattern, electronic and automatic beam steering, and interference suppression through side lobe nulling [75]. The conventional method of beamforming based on periodic sinusoidal waves results in a beam pattern, or array factor $F(\theta)$, of the form

$$F(\theta) = \frac{\sin(L\pi\theta/\lambda)}{L\pi\theta/\lambda} = \frac{\sin(Lf\pi\theta/c)}{Lf\pi\theta/c} \tag{6.56}$$

where L is the array size and θ is a function of the angle of incidence or radiation. The array factor in Eqn. (6.56) results in the well-known equation for resolution angle

$$\varepsilon = \frac{k\lambda}{L} = \frac{kc}{Lf} \tag{6.57}$$

where k is a constant usually set equal to one.

Array beamforming based on Gaussian pulses yields the array factor as follows [76]

$$A(\theta) = \frac{\text{erf}[\sqrt{\pi}L(\Delta f)\theta/2c]}{\sqrt{\pi}L(\Delta f)\theta/2c} \tag{6.58}$$

where erf[·] is the well-known error function and Δf is the approximate bandwidth of the pulses. If we define the parameter ρ as

$$\rho = \frac{L(\Delta f)}{c} \tag{6.59}$$

the variation of the array factor as a function of θ and ρ can be demonstrated as in Figure 6.16.

As an example, for $\Delta f = 4$ GHz, $c = 3\times10^8$ m/s, and $L = 0.3$ m we have $\rho = 5$, and from Figure 6.16 the array factor is equal to 1.128 and 0.141 for $\theta = 0^\circ$ and $\theta = 2^\circ$, respectively. An increase in frequency bandwidth and in the array length (e.g., the number of antenna elements) will increase the parameter ρ. An increase in ρ decreases the array factor at all angles other than zero degrees. In contrast to the beam patterns of Eqn. (6.56), which include distinguishable side lobes, $A(\theta)$ is a monotonically decreasing function of angle θ.

The resolution angle from the array factor $A(\theta)$ is given by

$$\varepsilon = \frac{Kc}{(\Delta f)L} = \frac{K(\Delta T)c}{L} \tag{6.60}$$

where K is a constant usually set equal to one. The resolution angle for nonsinusoidal signals in Eqn. (6.60) is a function of the array size L and the bandwidth Δf.

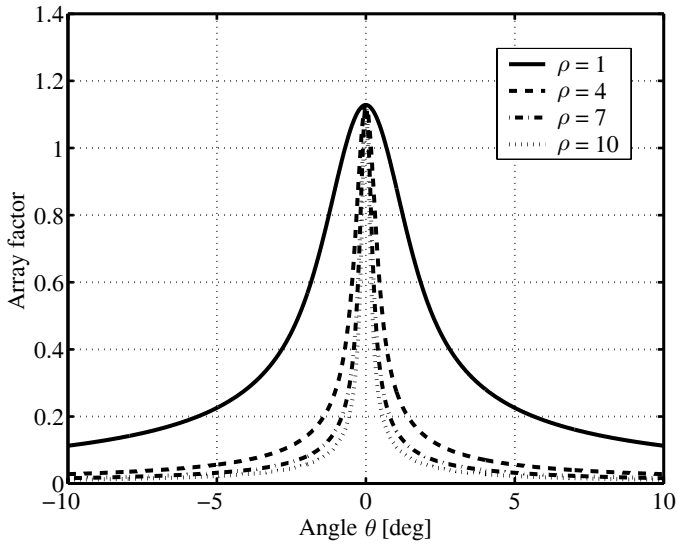


Fig. 6.16 Array factor for beamforming based on Gaussian pulses as a function of angle θ and ρ.

An increase in bandwidth results in simultaneous improvement in range resolution and angular resolution. This feature is of interest for UWB high-resolution imaging radar.

6.7 SUMMARY

In this chapter the basic knowledge regarding UWB antennas and arrays was presented. First, the fundamental equations governing antenna theory were mentioned and important parameters involved in antenna analysis were explained. We understood that the radiation of short-duration UWB signals from an antenna is significantly different from long-duration narrowband signals. By studying resonant and non-resonant antennas we showed that conventional antennas are not suitable for the UWB system.

Different antenna elements applicable to UWB systems, such as the conical antennas, monopole antennas, D-dot probe antennas, and TEM horn antennas were introduced. We concluded that, although the classical antennas have proven their use, none of them are exactly suitable for UWB applications, and, therefore, there is a need to look for new types of antennas.

As a conventional and straightforward method for UWB beamforming we studied the delay-line beamformer. Several parameters of this array, such as the beamwidth and angles of grating lobes, were derived and discussed. Finally, we introduced radar UWB array systems and several factors affecting their performance.

Problems

Problem 1. Derive the radiation distance as given in Eqn. (6.14).

Problem 2. Using the RLC model for resonance antenna given in Figure 6.6 verify the validity of Eqn. (6.26).

Problem 3. Discuss the difficulties of UWB antennas that are appropriate for communication systems with respect to the following:

(i) size;

(ii) cost;

(iii) VSWR;

(iv) frequency response; and

(v) on chip design.

Problem 4. You can find several algorithms for wideband smart antennas in the recent literature. Investigate one of them and discuss the possible applications of that technique for UWB systems.

Problem 5. In *Example 6.1* consider

$$d = \frac{4c}{f_h} \tag{6.61}$$

(i) Find d.

(ii) Calculate and sketch the inter-null beamwidth as a function of frequency f, $f_l < f < f_h$.

(iii) Compute the first two grating lobes at each sides of the main lobe for $f = 5, 6, 7$ GHz.

(iv) Sketch the beam patterns of the array for 21 frequencies equally distanced from 5 to 7 GHz.

Problem 6. Derive Eqn. (6.56).

Problem 7. Referring to [76], derive the formula for the array factor used in array beamforming based on the Gaussian pulses given in Eqn. (6.58).

7

Position and location with ultra wideband signals

Wireless UWB positioning techniques can provide real time indoor and outdoor precision tracking for many applications. Some potential uses include locator beacons for emergency services and mobile inventory, personnel and asset tracking for increased safety and security, and precision navigation capabilities for vehicles and industrial and agricultural equipment. The characteristics of UWB signals provide the potential of highly accurate position and location estimation.

In this chapter we explain the fundamentals of positioning and location using UWB signals and systems. Aspects of this topic, such as resolution and timing issues in a practical environment, are considered.

7.1 WIRELESS POSITIONING AND LOCATION

Positioning is defined as the "determination of the location of somebody or something". Use of electronic distance measurement techniques to position humans or objects derived from hyperbolic aircraft navigation systems was first developed during World War II. A variety of systems have been used since that time, most of which became quickly obsolete when the global positioning system (GPS) became fully operational. However, basic operating concepts have not changed significantly.

This section describes wireless distance measurement and positioning principles. Land-based or terrestrial positioning systems are distinguished from satellite or

extraterrestrial positioning systems. All these systems use time difference and trilateration techniques to estimate the position.

7.1.1 Types of wireless positioning systems

A good method of classifying wireless positioning systems is by their operating frequencies. The frequency generally determines the operating range and accuracy and, in turn, a system's suitability for a particular application. In general, the higher the frequency of the electronic positioning system the more accurate the resultant position becomes. Systems in the medium frequency range and below are typically hyperbolic phase/pulse differencing and can reach far beyond the visible or microwave horizons. These systems are more suited for long-range navigation purposes. Wireless indoor tracking systems that locate people and objects use high frequency and bandwidth radio signals.

7.1.1.1 Low-frequency positioning systems Low-frequency time-differencing positioning systems are suitable only for long-range navigation problems. Daily calibration is critical if absolute accuracy is to be maintained.

7.1.1.2 Medium-frequency positioning systems The first medium-frequency positioning systems were deployed in the mid-1950s and were used up to the early 1970s (they are no longer used today). Systems in this frequency range operated by time or phase-differentiating methods and required repeated calibration and continual monitoring.

7.1.1.3 Super high-frequency positioning systems Microwave and UWB systems in the super high frequency (SHF) range are most commonly used over relatively limited distances up to 100 meters and can provide the highest distance accuracy measurements.

7.1.2 Wireless distance measurement

Most wireless distance measurement systems operate either by resolving two-way phase delays of a modulated electromagnetic carrier mono-pulse signal between the object and the reference transmitter or by measuring the two-way propagation time of a coded electromagnetic pulse between these points. On the other hand, GPS operates in a similar manner to conventional systems, except that propagation distances from the satellites are one-way. Microwave pulsing systems measure the round trip propagation delay of a pulse.

For a pulsing system the round trip distance is computed by multiplying the measured elapsed delay, taking account of the internal system time delays, by the assumed velocity of propagation of electromagnetic energy. The distance, or *range*,

is computed by the equation

$$d = c\frac{t_m - t_d}{2} \tag{7.1}$$

where c is the assumed velocity of propagation (m/sec), t_m is the measured round trip time delay (sec), and t_d is the summation of internal system delays (sec). The resultant distance d will be in meters.

7.1.2.1 Distance determination Under ideal circumstances and with repeated measurements the time delay t_m can be measured fairly accurately and far more accurately when modulated phase comparison techniques are employed, such as with infrared and some microwave and UWB systems. However, at least two factors on the right side of Eqn. (7.1) are subject to both random and systematic errors. The only way to minimize these errors is by external and internal calibration of the equipment.

The whole internal system delay t_d can be controlled effectively on some modern pulsing systems. Such control is often termed *self-calibrating*. Other local anomalies, or inherent system measurement instabilities, cannot be controlled or corrected by the measurement system. Thus, an independent, on-site calibration must be performed if errors due to these sources become significant, which is normally the case. As a result a calibrated microwave positioning system can measure range to an accuracy between ±3 m and ±10 m (with 95% rms).

7.1.2.2 Velocity of propagation Variations in the velocity of propagation in air are caused by changes in air density due to temperature, humidity, and air pressure. The effect on land-based microwave positioning systems is more pronounced than on light waves. Assumed stability in the pulsing system time $(t_m - t_d)$ or phase measurement process cannot be guaranteed. Periodic, independent calibration is essential to check this stability.

7.1.2.3 Antenna considerations Electromagnetic wave propagation and refraction problems may exist in some areas. Weather, especially humidity and temperature, affects propagation through the air. Unwanted reflections of the signals can be received at the antenna. Directional antennas may be used to boost a signal into an antenna. This is possible by, for example, using a set of antenna elements and a beamforming algorithm. Circular polarization is a technique used to reduce multipath effects. Another technique used is antenna diversity, which switches from one antenna to another to reduce multipath fading effects.

7.1.2.4 Multipath propagation effects Multipath propagation is a major cause of systematic errors. Errors due to this effect are difficult to detect. Consideration of multipath during antenna placement, antenna design, and other internal electronic techniques and filters is required to identify and minimize multipath propagation

effects. Antenna spacing or systems with circular polarization are recommended to minimize the possibility of degrading effects.

7.1.3 Microwave positioning systems

These systems were first used in the early 1970s. They effectively replaced medium frequency positioning methods that had been used since the 1950s. Up until the mid-1990s, microwave positioning systems were the primary positioning systems nearly everywhere. After 1992, when full-coverage differential GPS became available, the use of microwave systems rapidly declined.

Positioning by microwave systems is accomplished by determining the coordinates of the intersection of two or more measured ranges from known control points. This process is termed *trilateration*. When two circular ranges are measured, two intersection points result, one on each side of the fixed baseline connecting the reference stations. The ambiguity is usually obvious and is controlled by either initializing the computing system with a coordinate on the desired side of the baseline or referencing the point relative to the baseline azimuth.

7.1.3.1 Automated tracking When automated positioning systems were used the range intersection coordinates were automatically computed and transformed relative to the project alignment coordinate system. These data were then applied to an analog or digital course indicator, allowing any particular cross-section or offset range to be tracked.

7.1.3.2 Positioning accuracy The positional accuracy of an intersection position (Figure 7.1) is a function of the range accuracy and angle of intersection of the ranges. The angle of intersection varies relative to the baseline. Assuming both ranges have equal value, positional accuracy can be estimated from

$$\text{Positional accuracy} = 2.447\sigma \csc(A) \tag{7.2}$$

where σ is the estimated standard error of measured range distance and A is the angle of intersection of two ranges at the destination.

Since A has a major effect on positional accuracy, quality control criteria will restrict surveys within intersection tolerances (e.g., A must be between 45° and 135°). The accuracy of microwave positioning systems is difficult to estimate since it is not constant with distance from a transceiver.

7.1.3.3 Multiple-range positioning technique This method is an expansion of the two-range method described above and was developed in 1979. In this case, three or more ranges are simultaneously observed and positional redundancy results. The position is determined from the computed coordinates of the intersections of three or more range circles. Since each range contains observational errors, all the circles will not intersect at the same point. In the case of three observed ranges,

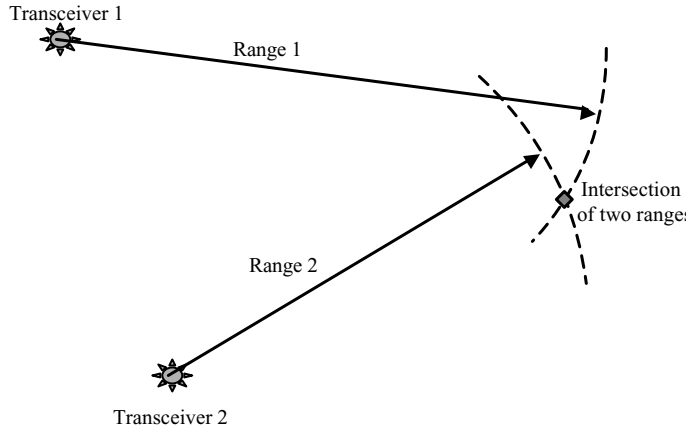


Fig. 7.1 Fundamentals of the range intersection method.

three different coordinates result. Four ranges result in six separate coordinates. The final position is derived by adjustment of these redundant coordinates, usually by a least squares minimization technique. Some automated microwave positioning systems simply use the strongest angle of intersection as the "adjusted" position, and others take the unweighted average of all the intersecting coordinates.

Using multiple ranging can minimize positional uncertainties. The coordinated position contains redundancy and can be adjusted. Such a process reduces geometrical constraints and provides an opportunity to evaluate the resultant positional accuracy. This is accomplished by evaluating the best estimate inside the so-called triangle of error which occurs when three or more position lines containing errors intersect. A plot of the simple case of three intersecting ranges is shown in Figure 7.2. The position of the target is obtained by adjusting the three ranges to a best fit.

Assessment of the range measurement accuracy may be obtained by computing the residual range errors v_1, v_2, and v_3 for each position (Figure 7.2). These are the corrections added to each range so that all ranges intersect at the same point. When a least squares type of adjustment is performed the sum of the squares of the residual errors v_i is minimal. The magnitudes of these residual range corrections provide the statistics for an accuracy estimate of the observed distances or, more practically, an approximate quality control indicator. When a least squares adjustment is performed, it is possible to obtain an accuracy estimate of the positional rms error. Automated software can provide such data at each position update. If known, different weights may be assigned to individual range observations. This proves useful when different types of positioning systems are mixed.

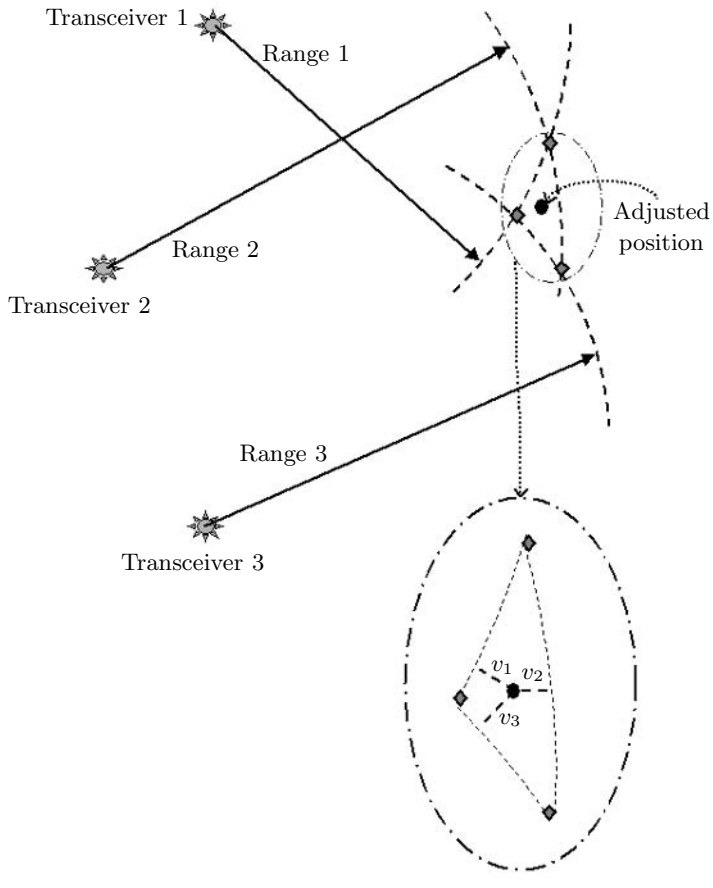


Fig. 7.2 Fundamentals of the multiple range intersection method.

The residual range errors v_i which result from comparing observed distances with the distances between the adjusted position and remote transmitters could be used to evaluate the accuracy of range measurements. A variety of methods have been used to compute these residual errors. An approximate estimate of range accuracy is obtained from the following:

$$\text{Estimated range error} = \sigma = \sqrt{\frac{\sum_{i=1}^{n} v_i^2}{n-1}} \tag{7.3}$$

where n is the number of observed ranges and $\sum_{i=1}^{n} v_i^2$ is the sum of the squared residuals.

Adding redundant ranges will not necessarily make a significant improvement in the positional accuracy because inherent random and systematic errors are still

present. It will, however, help detect the existence of large systematic errors that might have otherwise gone undetected using a nonredundant range-range system.

7.2 GLOBAL POSITIONING SYSTEM TECHNIQUES

The global positioning system or GPS (Figure 7.3) has rapidly become the standard surveying and navigation mode, replacing microwave and other types of ranging systems. Real time GPS positional accuracies now exceed those of any other positioning system. Most significantly, GPS does not require the time-consuming calibration that microwave equipment does. Numerous public and private differential GPS systems now exist which give a broad coverage with an accuracy in excess of standard nondifferential GPSs.

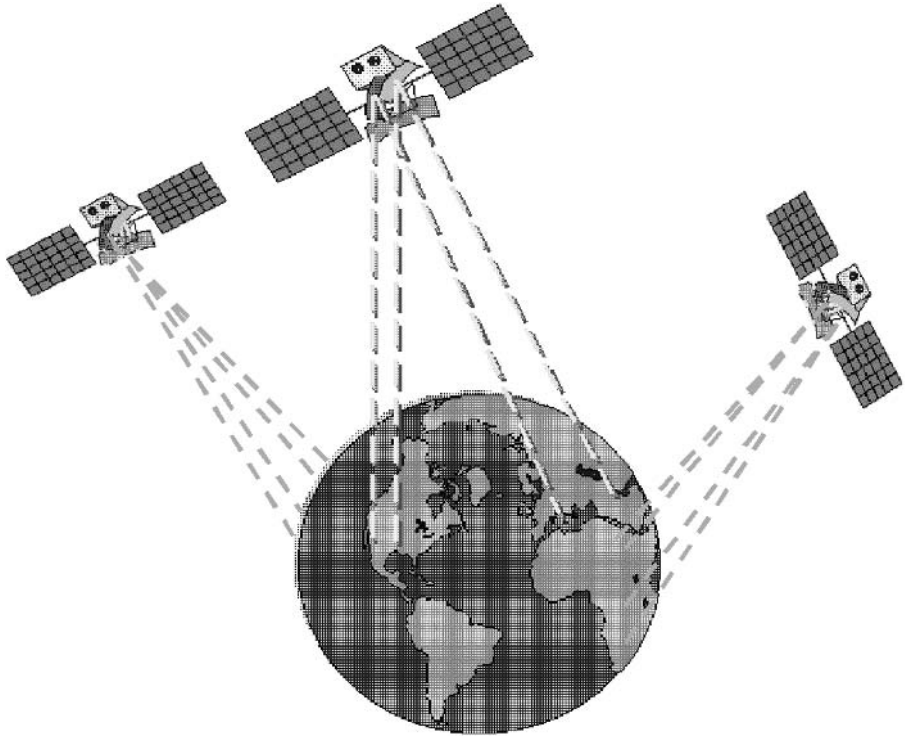


Fig. 7.3 Global positioning system (GPS) with satellite navigation.

Actually, GPS is a real time, all-weather, 24-hour, worldwide, 3-dimensional absolute satellite-based positioning system. This system consists of two positioning services:

- **The precise positioning service (PPS)** was developed for the military and other authorized users and provides an accuracy of 5-10 meters in absolute positioning mode.
- **The standard positioning service (SPS)** is available to civilian users and provides accuracy of 10-20 meters in absolute positioning mode.

7.2.1 Differential GPS (DGPS)

For many applications, absolute positioning does not provide sufficient accuracy. Differential GPS is a technique which can provide relative positioning with an accuracy from a few meters to a few millimeters, depending on the DGPS method used. Such a scheme utilizes an additional reference signal, transmitted from a known location, which is used to reduce the inherent error of standard GPS. Normally, DGPS utilizing code phase measurements can provide a relative accuracy of a few meters. However, DGPS utilizing carrier phase measurements can provide a relative accuracy of a few centimeters. Differential GPS requires two or more GPS receivers to be recording measurements simultaneously. With two stations recording observations at the same time, GPS processing software can reduce or eliminate common-mode errors. Both code and carrier-phase DGPS can be performed in real time, making it applicable for moving platforms.

7.2.2 GPS tracking modes

There are basically two general modes which are used to determine the distance between a GPS satellite and a ground-based receiver antenna. These measurements are made by signal-phase comparison techniques:

- Carrier-phase tracking.
- Code-phase tracking.

Either the satellite's carrier frequency phase or the phase of a digital code modulated on the carrier phase may be used, or tracked, to resolve the distance between the satellite and the receiver. The resulting positional accuracy is dependent upon the tracking method used.

GPS satellites actually broadcast on two carrier frequencies: L1 at 1575.42 MHz (19-cm wavelength) and L2 at 1227.60 MHz (24-cm wavelength). Modulated on these frequencies are the coarse acquisition (C/A) (300-m wavelength) and the precise (P) (30-m wavelength) codes. In addition, a 50-bps satellite navigation message containing the ephemeris and health status of each satellite is transmitted. The C/A and P codes are both present on the L1 frequency. Only the P code is present on the L2 frequency.

The higher frequencies of the carrier signal (L-band) have wavelengths of 19 and 24 cm, from which a distance can be resolved through post-processing software to

approximately 2 mm. The modulating code has a wavelength of 300 m and will only yield distances accurate to about 1 m.

7.2.3 GPS error sources

The accuracy of GPS is a function of the error and interference on the GPS signal and the processing technique used to reduce and remove these errors. The same types of phenomena as found in range-range microwave systems affect GPS signals. Both types of systems are highly affected by humidity and multipath.

7.2.3.1 Tropospheric error Humidity is included in this type of error. Humidity can delay a time signal up to approximately 3 m. Satellites low on the horizon will be sending signals across the face of the earth through the troposphere. Satellites directly overhead will transmit through much less troposphere. Masking the horizon angle to 15 degrees can minimize tropospheric error.

7.2.3.2 Ionospheric error Sunspots and other electromagnetic phenomena cause errors in GPS range measurements of up to 30 m during the day and as high as 6 m at night. The errors are not predictable, but can be estimated.

7.2.3.3 Multipath Multipath is the reception of reflected, refracted, or diffracted signals in lieu of a direct signal. Multipath signals can occur below or above the antenna. Multipath magnitude is less over water than over land, but it is still present and always changing. If possible the placement of the GPS receiver antenna should avoid areas where multipath is more likely to occur (e.g., rock outcrops, metal roofs, commercial roof-mounted heating and air conditioning, buildings, cars, ships, etc.). Increasing the height of the antenna is one method of reducing multipath at a reference station. Multipath occurrence on a satellite transmission can last several minutes while the satellite passes overhead. Masking out satellite signals from the horizon up to 15 degrees will also reduce multipath effects.

7.3 POSITIONING TECHNIQUES

In this section we will consider positioning techniques that are based on a different classification.

7.3.1 Introduction

According to the place where measurements and their evaluation take place, positioning systems can be classified as network-based, handset-based, or hybrid [77, 78, 79].

7.3.1.1 Network-based Systems In network-based systems, calculation of the position is computed in a control station which is located in a remote position from the object to be positioned. More specifically, a set of stationary receivers make appropriate measurements of a signal originating from or reflecting off the object to be positioned. These measurements are then used in the control station where the object's location is computed. Such systems involve lower mobile terminal costs than the self-positioning technique, but they are usually less accurate than handset-based systems.

7.3.1.2 Handset-based systems Handset-based systems involve the calculation of the position of the object to be positioned itself. More specifically, the positioning receiver uses appropriate measurements made from signals sent by location-known transmitters in order to determine its own position. The main advantage of handset-based systems is that this technology can better address privacy issues as the location is computed and stored in the receiver. However, mobile terminal costs will increase as additional signal processing is integrated in the receiver.

7.3.1.3 Hybrid systems Finally, hybrid systems incorporate a combination of handset and network-based technologies. Usually, the object to be positioned takes the measurements and transmits the results to the stationary network where the object's location is estimated. The main aim of this method is to produce a more robust estimate of location in a single process.

7.3.2 Network-based techniques

7.3.2.1 Received Signal Strength (RSS) With this technique the signal strength of the object to be positioned is measured at several stationary receivers. Ideally, each measurement will provide a circle of radius representing the distance between the object and the receiver that made this measurement, centered at the corresponding receiver. The object position is then given by the intersection of these circles. In two-dimensional positioning and assuming that no measurement error occurs, at least three circles are required in order to resolve the ambiguities arising from multiple crossings of the circles.

The accuracy of the position can be improved by increasing the number of measurements and then averaging the results. However, this approach requires an exact knowledge of the path loss in order to get an accurate estimation of the signal strength at the receivers. In a multipath fading environment, it is difficult to relate the distance with the received signal strength. Figure 7.4 shows an example of RSS positioning. Obviously, it is rare that all three distorted circles coincide exactly, and ambiguity in the coincidence point determines the systematic positioning error.

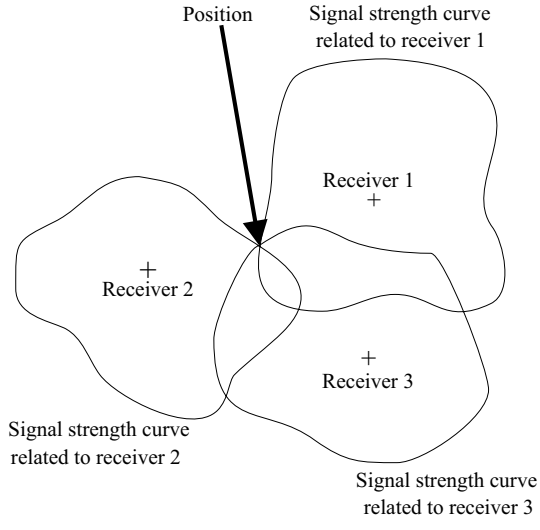


Fig. 7.4 In RSS positioning the intersections of the distorted circles determine the position of the object.

7.3.2.2 Angle of arrival (AOA)/Direction of arrival (DOA) In this method the angle of arrival of the signal sent by the object to be positioned is measured at several stationary receivers by steering the main lobe of a directional antenna or an adaptive antenna array. Each measurement forms a radial line from the receiver to the object to be positioned. In two-dimensional positioning the position of the object is defined at the intersection of two directional lines of bearing. In practice, more than two receivers may be employed to combat inaccuracies introduced by multipath propagation effects.

This method has the advantage of not requiring synchronization of the receivers nor an accurate timing reference. On the other hand, receivers require regular calibration in order to compensate for temperature variations and antenna mismatches. This can be done automatically.

Assuming that we know the coordinates of two receivers the derivation of the position is straightforward. Without loss of generality, we can assume that these coordinates are $(0, 0)$ for receiver 1 and $(0, y_2)$ for receiver 2. Given α and β, respectively, as the angles of arrival of the signal from the object at receiver 1 and receiver 2 we can define two straight lines by

$$y = \tan(\alpha)x \tag{7.4}$$

$$y = \tan(\beta)x + y_2 \tag{7.5}$$

Substituting Eqn. (7.4) into Eqn. (7.5) yields

$$x = x_0 = \frac{y_2}{\tan(\alpha) - \tan(\beta)} \tag{7.6}$$

Substituting x_0 into Eqn. (7.4) gives us a unique y_0, and thus the point defined by the coordinates (x_0, y_0) is the desired position. Figure 7.5 demonstrates an example of AOA positioning.

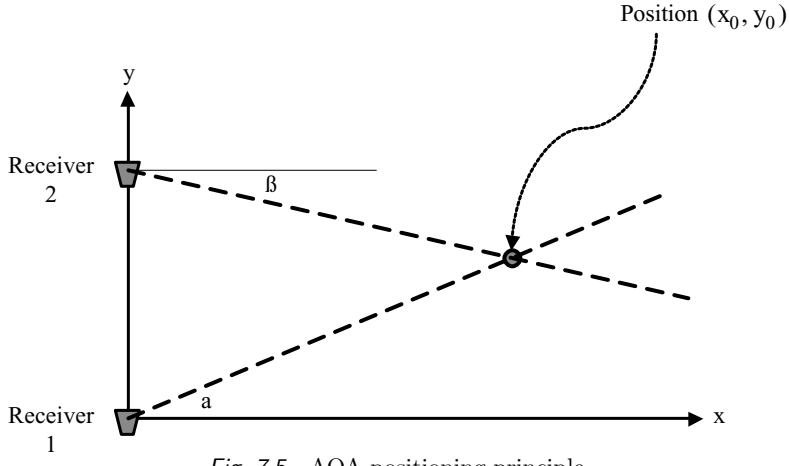


Fig. 7.5 AOA positioning principle.

7.3.2.3 Time of arrival (TOA) With this approach the time that the signal sent by the object to be positioned is measured at each receiver. Given that the propagation time of the signal is known and directly proportional to its traversed distance the measured time can provide a circle of radius representing the distance between the object and the receiver, centered at the latter. In two-dimensional positioning, at least three circles are required. This technique is easy to implement; however, it requires knowledge of the transmission time of the emitted signal as well as synchronization of the transmitter and receivers clocks. Otherwise, huge position errors can occur. For example, a clock inaccuracy of just 1 μs will lead to a position error of 300 m. Furthermore, this technique can suffer from multipath propagation effects.

Assuming that we know the coordinates of three receivers the derivation of the position is simple. Without loss of generality, we can assume that these coordinates are as follows

$$\text{Receiver } 1 : (0, 0)$$
$$\text{Receiver } 2 : (0, y_2)$$
$$\text{Receiver } 3 : (x_3, y_3)$$

Given that t_1, t_2, and t_3 denote the time it takes the signal to travel from the object to be positioned to the respective receivers and that c denotes the signal's speed of propagation the distance between the object and each of the receivers is given by

$$d_1 = c.t_1 = \sqrt{x^2 + y^2} \tag{7.7}$$

$$d_2 = c.t_2 = \sqrt{x^2 + (y - y_2)^2} \tag{7.8}$$

$$d_3 = c.t_3 = \sqrt{(x - x_3)^2 + (y - y_3)^2} \tag{7.9}$$

Each of these equations defines a circle whose x and y are unknowns. Squaring both sides of the above equations and some basic manipulation yields

$$y = \frac{y_2^2 + d_1^2 - d_2^2}{2y_2} \tag{7.10}$$

Substituting Eqn. (7.10) into Eqn. (7.7) gives two values for x. Only one of these values which is positive is correct, and this way we have calculated the desired position. Figure 7.6 demonstrates an example of TOA positioning.

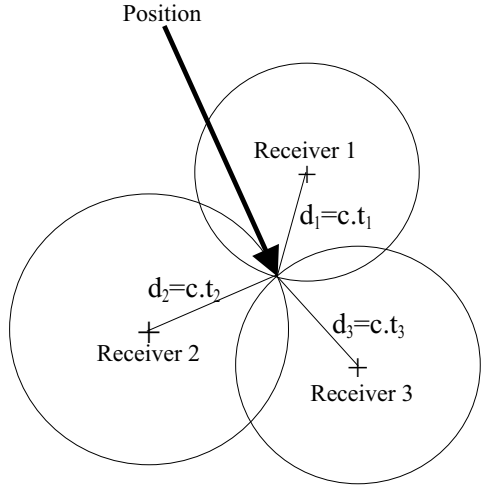


Fig. 7.6 TOA positioning principle.

7.3.2.4 Time difference of arrival (TDOA) In this technique [80] the difference in time at which the signal from the object to be positioned arrives at two different receivers is measured. Each time difference is then converted into a hyperboloid with a constant distance difference between the two receivers. In two-dimensional

positioning, at least two pairs of receivers are required; the position is the intersection of the two corresponding hyperboloids.

This technique requires synchronization of the receivers' clocks; however, knowledge of the absolute transmission time of the emitted signal is no longer required. Here again, multipath propagation effects can influence the accuracy of the position and the location of the receivers. It is important to note that the wider the bandwidth of the signal the lower the measurement error becomes.

The derivation of the position is straightforward if we assume that we know the coordinates of three receivers. Without loss of generality, it can be assumed again that these coordinates are

$$\begin{aligned}&\text{Receiver 1} : (0,0)\\ &\text{Receiver 2} : (0,y_2)\\ &\text{Receiver 3} : (x_3,y_3)\end{aligned}$$

Given that t_1, t_2, and t_3 denote the time it takes the signal to travel from the object to be positioned to the corresponding receivers the distance between the object and each of the receivers is given by

$$d_1 = c.t_1 \tag{7.11}$$

$$d_2 = c.t_2 \tag{7.12}$$

$$d_3 = c.t_3 \tag{7.13}$$

We can now define two hyperboloids using the TDOA algorithm, that is

$$\begin{aligned}d_{1,2} &= d_2 - d_1 = c.(t_2 - t_1)\\ &= \sqrt{x^2 + (y-y_2)^2} - \sqrt{x^2+y^2}\end{aligned} \tag{7.14}$$

and

$$\begin{aligned}d_{1,3} &= d_3 - d_1 = c.(t_3 - t_1)\\ &= \sqrt{(x-x_3)^2 + (y-y_3)^2} - \sqrt{x^2+y^2}\end{aligned} \tag{7.15}$$

where x and y are unknowns. Taking the square in Eqns. (7.14) and (7.15) yields

$$2d_{1,2}\sqrt{x^2+y^2} = y_2^2 - d_{1,2}^2 - (2y_2)y \tag{7.16}$$

$$2d_{1,3}\sqrt{x^2+y^2} = x_3^2 + y_3^2 - d_{1,3}^2 - (2x_3)x - (2y_3)y \tag{7.17}$$

Knowing that $x^2 + y^2$ is not equal to zero, Eqns. (7.16) and (7.17) lead to

$$x = by + a \tag{7.18}$$

where

$$b = \frac{2y_2 d_{1,3} - 2y_3 d_{1,2}}{2x_3 d_{1,2}} \tag{7.19}$$

and

$$a = \frac{x_3^2 d_{1,2} + y_3^2 d_{1,2} - y_2^2 d_{1,3} + d_{1,2}^2 d_{1,3} - d_{1,2} d_{1,3}^2}{2 x_3 d_{1,2}} \tag{7.20}$$

Substituting Eqn. (7.18) into Eqn. (7.16) gives

$$2 d_{1,2} \sqrt{(b^2+1) y^2 + (2ba) y + a^2} = y_2^2 - d_{1,2}^2 - (2 y_2) y \tag{7.21}$$

which results in

$$[4 d_{1,2}^2 (b^2+1) - 4 y_2^2] y^2 + [8 b a d_{1,2}^2 + 4 (y_2^2 - d_{1,2}^2) y_2] y + [4 a^2 d_{1,2}^2 - (y_2^2 - d_{1,2}^2)^2] = 0 \tag{7.22}$$

Eqn. (7.22) is a quadratic equation with two roots that are the y-coordinates of the intersection points of the hyperboloids. Using Eqn. (7.18) provides the corresponding x-coordinates. To remove the ambiguity we can define another hyperboloid by

$$\begin{aligned} d_{2,3} &= d_3 - d_2 = c.(t_3 - t_2) \\ &= \sqrt{(x - x_3)^2 + (y - y_3)^2} - \sqrt{x^2 + (y - y_2)^2} \end{aligned} \tag{7.23}$$

Substitution of $d_{1,3}$ by $d_{2,3}$ in Eqns. (7.16) to (7.22) yields two points. One of these points matches the previous ones. This point is the required position. Figure 7.7 demonstrates an example of TDOA positioning.

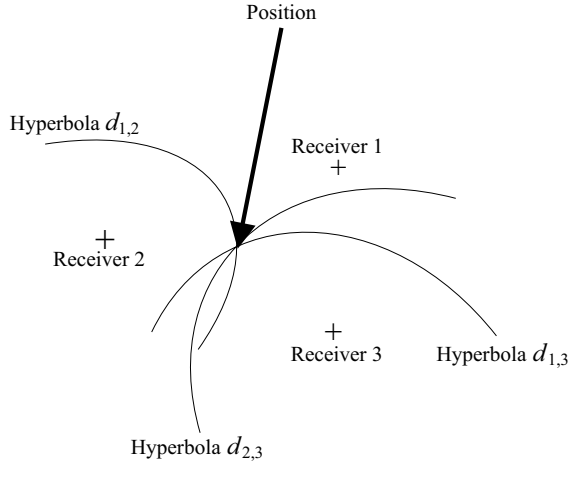


Fig. 7.7 TDOA positioning principle.

7.3.2.5 Multipath fingerprinting This positioning technique is based on matching the received signal fingerprint to a reference fingerprint previously measured for a known location in the network and stored in a central database. Each spot in a network would have a unique signature in terms of TOA, AOA, and RSS, observed from at least one receiver. Therefore, to build the database we have to divide the service area into nonoverlapping zones and record the received signal pattern corresponding to each zone in the database. This technique performs better than other techniques in multipath-rich environments, but the size of the database increases considerably when the service area becomes large.

7.3.3 Handset-based techniques

These are satellite-based positioning techniques, such as GPS, Galileo, or the global navigation satellite system (GLONASS). Currently, there are three potential global satellite navigation systems:

- The U.S. made GPS which is relatively accurate and reliable (described in Section 7.2).
- The European Galileo is still in development, but promises to be more accurate than the GPS system.
- The Russian GLONASS.

A receiver using such systems requires a clear view of the sky and signals from at least four satellites that are part of a constellation. For three-dimensional positioning, three time measurements (just like the TOA method) are used to define a sphere in space centered at the corresponding satellite and a fourth measurement is required to solve receiver clock bias caused by the unsynchronized clocks of the satellite and the receiver. Furthermore, each satellite sends its almanac to the GPS receiver to describe its position in the sky.

Such systems are most unreliable in buildings or urban environments, as the receiver might not be able to get four satellites in its line of sight or receive any satellite at all. Furthermore, without knowledge of the state of the satellite constellation a receiver may take several minutes to obtain a measured position.

As described before, differential GPS (DGPS) is used to improve the accuracy of a conventional GPS receiver. Along with the mobile receiver a stationary receiver is used to receive signals from all satellites that are visible from the station. A position measurement is made for each received signal, and correction data for each satellite are processed as the station's location is known. The data are then sent to the mobile GPS receiver which can improve the accuracy of the measured position by picking up information from those satellites it uses for the measurement. This method corrects errors due to atmosphere or clock bias, but cannot remove errors due to multipath-rich environments as the station cannot predict the mobile's nearby surroundings.

The usual way to calculate correction data is to measure the pseudo-range error of each satellite. This technique is based on the principle that the received signal from a satellite at two different receivers on earth located a few kilometers apart will suffer almost the same errors and delays due to the atmosphere. As the exact position of the stationary receiver is known as well as the true distance to the satellite from the stored almanac, it is possible to theoretically calculate how long it takes for the signal to reach the receiver. Compared with the measured time, the difference gives an error correction time factor which can be used to correct the pseudo-range given by

$$\text{Corrected pseudorange} = c \cdot (t_{\text{measured}} - t_{\text{correction}}) \tag{7.24}$$

where t_{measured} denotes the measured time for the signal to arrive and $t_{\text{correction}}$ represents the time correction factor. This corrected pseudo-range is then sent to the mobile GPS receiver which can then correct its own measurements.

The final technique discussed here uses the enhanced observed time difference (E-OTD) technique, which is similar to the TDOA method. The mobile object to be positioned listens to bursts sent from neighboring stationary transmitters with accurate timing sources and measures the time difference of arrival from these transmitters. The object to be positioned then uses these measurements to determine its own location through a trilateration scheme, as in the TDOA method. This technique requires at least two pairs of synchronized transmitters in order to remove the ambiguity in position.

7.3.4 Hybrid techniques

7.3.4.1 Assisted GPS (A-GPS) In this technique a reference GPS receiver, which sees the same satellites as the object to be positioned, is used. With this reference GPS receiver the network can predict the GPS signal that the object will receive and, thus, assist the object to be positioned by sending information about the position of the GPS satellites. The object can then make quick GPS measurements (the acquisition delays of a conventional GPS are eliminated) and send them back to the network where the exact position is estimated. Furthermore, the performance of conventional GPS receivers in low SNR conditions is improved with A-GPS technology.

7.3.4.2 Advanced forward link trilateration (AFLT) This technique is similar to the TDOA method. The object to be positioned measures the time of arrival of signals from the transmitters and reports them back to the network where the calculation of the object's position is processed. Here, at least three transmitters are required to get an optimal position fix. The accuracy of this method is usually better than using the TDOA technique as all the computations are made in the network.

7.3.5 Other techniques

As each technique has its own advantages and disadvantages, it is possible to combine several of these techniques in order to improve the accuracy or the cost of the positioning process. Some examples of combination are AOA+RSS, AOA+TDOA, E-OTD+A-GPS, and A-GPS+AFLT.

7.4 TIME RESOLUTION ISSUES

As time-based techniques are frequently used in geolocation, one of the main sources of error is due to the time resolution of the signal used to make the appropriate measurements. These signals can be classified in four main systems: narrowband, wideband, super-resolution, or ultra wideband ranging.

7.4.1 Narrowband systems

Usually, in the narrowband ranging technique the time-based approach is used and the distance between two points is determined by measuring the phase difference between transmitted and received signals. The relation between the phase of a received signal ϕ and the time of arrival of this signal t is given by

$$t = \frac{\phi}{w_0} \tag{7.25}$$

where w_0 denotes the frequency of the signal in radians per second. Narrowband ranging techniques can reach accuracies of the order of 1 m, like the DGPS system, but a direct line-of-sight path is required in order to obtain accurate results. In multipath-rich environments, such as urban or indoor propagation scenarios where a direct line-of-sight path is often not available, substantial measurement errors occur. In fact, the received signal is the sum of all the signals arriving from different paths with different amplitudes and delays. Thus, if a direct line-of-sight path is not present, then the received signal arriving from a non-line-of-sight scenario takes more time to reach the receiver than expected in a line-of-sight scenario. The delay thus observed is the source of error in narrowband ranging techniques.

7.4.2 Wideband systems

Another widely used technique is the wideband signal approach where the direct sequence spread spectrum (DSSS) method is the most commonly used form, as this technique performs better than competing systems at suppressing interference. In such a system a known pseudo-noise (PN) signal, which is modulated using a modulation technique (such as BPSK, QPSK, etc), is multiplied by the

carrier signal, which is thus replaced by a wide bandwidth signal with a spectrum equivalent to that of the noise signal.

Without loss of information, this technique allows the signal power to be less than or equal to the noise power. Usually, in order to measure the time of arrival of the signal a sliding correlator or a matched filter is used at the receiver which cross-correlates the received signal with a stored reference PN sequence. The arrival time of the first correlation peak is used as the time measurement.

In the time-based approach the resolution of the estimation is related to the bandwidth of the spread signal by

$$d = \frac{c}{\mathrm{BW}} \tag{7.26}$$

where d denotes the absolute resolution and BW is the bandwidth of the spread signal. For example, a signal of 300-MHz bandwidth allows an absolute resolution of 1 m to be obtained. This is true when a direct line-of-sight path is visible. However, when multipath propagation occurs, wideband ranging techniques cannot maintain this range of accuracy; however, it is still better than the narrowband technique.

7.4.3 Super-resolution techniques

This method has been introduced in order to get higher ranging accuracy for a given bandwidth. The time of arrival of a signal can be determined with high resolution using a frequency domain super-resolution technique [81]. This technique uses the frequency response of the channel.

The impulse response of the channel can be modeled as

$$h(t) = \sum_{k=0}^{L_p-1} \alpha_k \delta(t - \tau_k) \tag{7.27}$$

where L_p denotes the total number of multipath components and α_k and τ_k are the complex attenuation and propagation delay of path number k, respectively. Furthermore

$$\alpha_k = |\alpha_k| e^{j\theta_k} \tag{7.28}$$

where θ_k represents the phase of complex attenuation. In the model, as propagation delays τ_k, $0 \leq k \leq L_p - 1$, are in ascending order the parameter τ_0 represents the propagation delay of the shortest path, which is the path of the direct line of sight. This delay has to be estimated as it is required for computing the time of arrival. Then, taking the Fourier transform of Eqn. (7.27) the frequency domain channel response can be obtained as

$$H(f) = \sum_{k=0}^{L_p-1} \alpha_k e^{-j2\pi f \tau_k} \tag{7.29}$$

In the model, parameters α_k and τ_k are time-invariant random variables since the motion of the obstacles (people, objects, etc.) is very slow compared with the measurement time interval. The phase of complex attenuation θ_k is assumed to be a random variable uniformly distributed in $[0, 2\pi]$.

Practically, the frequency domain channel response can be obtained by deconvolution of the received signal of a DSSS system over the frequency band of high SNR, by sweeping the channel at different frequencies or by using a multi-carrier modulation technique. TOA estimation is then performed by first interchanging the role of time and frequency in Eqn. (7.29). The result is the harmonic signal model as follows

$$H(\tau) = \sum_{k=0}^{L_p-1} \alpha_k e^{-j2\pi f_k \tau} \tag{7.30}$$

Actually, a spectral estimation technique, such as the multiple signal classification (MUSIC) algorithm, can compute a time domain analysis from the frequency response of the channel. Hence, by sampling the channel frequency response $H(f)$ at L equally spaced frequencies we can get some measurement data. Taking into account the additive white Gaussian noise $w(l)$ with zero mean and variance σ_w^2 the sampled frequency response of the channel can be rewritten as

$$\begin{aligned} x(l) &= \hat{H}(f_l) \\ &= H(f_l) + w(l) \\ &= \sum_{k=0}^{L_p-1} \alpha_k e^{-j2\pi(f_0 + l\Delta f)\tau_k} + w(l) \end{aligned} \tag{7.31}$$

where $l = 0, 1, \ldots, L-1$. Let us now rewrite Eqn. (7.31) in vector form as

$$\begin{aligned} \mathbf{x} &= \mathbf{H} + \mathbf{w} \\ &= \mathbf{V}\boldsymbol{\alpha} + \mathbf{w} \end{aligned} \tag{7.32}$$

where

$$\mathbf{x} = [x(0)\ x(1)\ \ldots\ x(L-1)]^T \tag{7.33}$$

$$\mathbf{H} = [H(0)\ H(1)\ \ldots\ H(L-1)]^T \tag{7.34}$$

$$\mathbf{w} = [w(0)\ w(1)\ \ldots\ w(L-1)]^T \tag{7.35}$$

$$\mathbf{V} = [\mathbf{v}(\tau_0)\ \mathbf{v}(\tau_1)\ \ldots\ \mathbf{v}(\tau_{L_p-1})] \tag{7.36}$$

$$\boldsymbol{\alpha} = [\alpha'_0\ \alpha'_1\ \ldots\ \alpha'_{L_p-1}]^T \tag{7.37}$$

and

$$\mathbf{v}(\tau_k) = [1\ e^{-j2\pi\Delta f\tau_k}\ \ldots\ e^{-j2\pi(L-1)\Delta f\tau_k}]^T \tag{7.38}$$

$$\alpha'_k = \alpha_k e^{-j2\pi f_0 \tau_k} \tag{7.39}$$

where the superscript T denotes the matrix transpose operation. The MUSIC algorithm detects frequencies by performing an eigen-decomposition on the covariance matrix of the data vector described in Eqn. (7.32)

$$\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\} \tag{7.40}$$

$$= \mathbf{V}\mathbf{A}\mathbf{V}^H + \sigma_w^2\mathbf{I} \tag{7.41}$$

where $E\{\cdot\}$ is the expectation

$$\mathbf{A} = E\{\boldsymbol{\alpha}\boldsymbol{\alpha}^H\} \tag{7.42}$$

$\mathbf{I}$ is the identity matrix, and superscript H denotes the conjugate transpose operation. Since propagation delays τ_k in Eqn. (7.27) are all different the column vectors of $\mathbf{V}$ are linearly independent.

Given that the magnitude of α_k is constant and phase θ_k is a uniform random variable in $[0, 2\pi]$ the $L_p \times L_p$ covariance matrix $\mathbf{A}$ is nonsingular. Hence, assuming $L > L_p$ the rank of the matrix $\mathbf{V}\mathbf{A}\mathbf{V}^H$ is L_p. This means that the $L - L_p$ smallest eigenvalues of $\mathbf{R}_{xx}$ are all equal to σ_w^2 and are thus called noise eigenvectors, while the L_p largest eigenvalues are called signal eigenvectors. The signal vector $\mathbf{x}$ is contained in an L-dimensional subspace that can therefore be split into a signal subspace and a noise subspace. These subspaces are orthogonal and are defined by signal eigenvectors and noise eigenvectors, respectively.

Assuming that the eigenvectors are normalized we can write

$$\mathbf{Q}_w^H\mathbf{Q}_w = \mathbf{I} \tag{7.43}$$

where

$$\mathbf{Q}_w = [\mathbf{q}_{L_p}\ \mathbf{q}_{L_p+1}\ \cdots\ \mathbf{q}_{L-1}] \tag{7.44}$$

$\mathbf{q}_k$, $L_p \leq k \leq L-1$, are the noise eigenvectors. The projection matrix of the noise subspace is then given by

$$\mathbf{P}_w = \mathbf{Q}_w(\mathbf{Q}_w^H\mathbf{Q}_w)^{-1}\mathbf{Q}_w^H \tag{7.45}$$

$$= \mathbf{Q}_w\mathbf{Q}_w^H \tag{7.46}$$

Since the signal subspace is orthogonal to the noise subspace and the vector $\mathbf{v}_{\tau_k}$, $0 \leq k \leq L_p - 1$, belongs to the signal subspace we have

$$\mathbf{P}_w\mathbf{v}_{\tau_k} = 0 \tag{7.47}$$

The above equation means that the vector $\mathbf{v}_{\tau_k}$, $0 \leq k \leq L_p - 1$, is orthogonal to the noise subspace. Then, given that the projection matrix is idempotent, that is

$$\mathbf{P}_w^H\mathbf{P}_w = \mathbf{Q}_w\mathbf{Q}_w^H\mathbf{Q}_w\mathbf{Q}_w^H = \mathbf{Q}_w\mathbf{I}\mathbf{Q}_w^H \tag{7.48}$$

$$= \mathbf{Q}_w\mathbf{Q}_w^H \tag{7.49}$$

$$= \mathbf{P}_w \tag{7.50}$$

the multipath delays τ_k, $0 \leq k \leq L_p - 1$, correspond to the delays when the time domain MUSIC pseudo-spectrum, defined as

$$\begin{aligned} S_{\text{MUSIC}}(\tau) &= \frac{1}{\|\mathbf{P}_w \mathbf{v}(\tau)\|^2} = \frac{1}{\mathbf{v}(\tau)^H \mathbf{P}_w^H \mathbf{P}_w \mathbf{v}(\tau)} \\ &= \frac{1}{\mathbf{v}(\tau)^H \mathbf{P}_w \mathbf{v}(\tau)} = \frac{1}{\|\mathbf{Q}_w^H \mathbf{v}(\tau)\|^2} \\ &= \frac{1}{\sum_{k=L_p}^{L-1} |\mathbf{q}_k^H \mathbf{v}(\tau)|^2} \end{aligned} \tag{7.51}$$

is maximum.

In practice, the channel frequency response is estimated using the received signal and the super-resolution algorithm then performs the time domain pseudo-spectrum transformation as in Eqn. (7.51). The time of arrival is finally estimated by detecting the first peak of the obtained pseudo-spectrum. However, in a high multipath environment, where non-line-of-sight conditions occur between the transmitter and the receiver, this technique cannot eliminate ranging errors at some locations.

It should be noted that the MUSIC algorithm is only an example and other techniques can be applied to the super-resolution algorithm.

7.4.4 Ultra wideband systems

Finally, the most recent, accurate, and promising technique is the UWB approach. We should note that ranging accuracy depends upon signal bandwidth. The larger the bandwidth the better the accuracy of the estimation of the time of arrival. In fact, as the bandwidth of UWB systems is usually in excess of 2–3 GHz the ranging accuracy is of the order of 1 cm. This fact is clear from Eqn. (7.26).

The large bandwidth of UWB systems means that they are able to resolve multiple paths and combat multipath fading and interference. However, such systems have a limited range and building penetration, due to the high attenuation associated with the high-frequency content of the signal. It was shown by Fontana in [82] and [83] that a set of fixed UWB beacons can provide a subcentimeter ranging accuracy in a multipath scenario.

Another method uses a subsampled version of the received signal as in [84]. A realistic UWB channel can be modeled as

$$h(t) = \sum_{l=0}^{L-1} a_l p_l(t - t_l) \tag{7.52}$$

where t_l denotes a signal delay along the lth path, a_l is a complex propagation coefficient corresponding to this path, and $p_l(t)$ are different pulse shapes that correspond to different paths. Hence, if a signal $s(t)$ is transmitted over this

channel the spectral coefficients of the received signal $y(t)$ are given by

$$Y[n] = S[n] \sum_{l=0}^{L-1} P_l[n] a_l e^{-j\omega_n t_l} + N[n] \tag{7.53}$$

where $N[n]$ represents the spectral coefficients of the noise and $P_l[n]$ are now the unknown parameters. These parameters can be approximated with polynomials of degree $D \leq R-1$, that is

$$P_l[n] = \sum_{r=0}^{R-1} p_{l,r} n^r \tag{7.54}$$

Therefore, Eqn. (7.53) can be rewritten

$$Y[n] = S[n] \sum_{l=0}^{L-1} a_l \sum_{r=0}^{R-1} p_{l,r} n^r e^{-j\omega_n t_l} + N[n] \tag{7.55}$$

Then, if we define

$$c_{l,r} = a_l p_{l,r} \tag{7.56}$$

and

$$Y_s[n] = \frac{Y[n]}{S[n]} \tag{7.57}$$

we get

$$Y_s[n] = \sum_{l=0}^{L-1} \sum_{r=0}^{R-1} c_{l,r} n^r e^{-j\omega_n t_l} + N[n] \tag{7.58}$$

Now, an annihilating filter for $Y_s[n]$ is

$$\begin{aligned} H(z) &= \prod_{l=0}^{L-1} (1 - e^{-j\omega_0 t_l} z^{-1})^R \\ &= \sum_{k=0}^{RL} H[k] z^{-k} \end{aligned} \tag{7.59}$$

with multiple roots at

$$z_l = e^{-j\omega_0 t_l} \tag{7.60}$$

where ω_0 denotes the sampling frequency. Then, noting that received signal $y(t)$ can be modeled as a convolution of L impulses with a known data sequence $g(t)$

we obtain

$$y(t) = \sum_{l=0}^{L-1} a_l p_l(t - t_l) * g(t) \tag{7.61}$$

and

$$Y[n] = \sum_{l=0}^{L-1} a_l P_l[n] G[n] e^{-jn\omega_c t_l} \tag{7.62}$$

where

$$\omega_c = \frac{2\pi}{T_c} \tag{7.63}$$

In the above equation T_c denotes the cycle time. If we use the polynomial approximation of coefficients $P_l[n]$ as defined in Eqn. (7.54) the minimum sampling rate corresponds to the total number of degrees of freedom per cycle (i.e., $2RL$). Therefore, accurate delay estimation can be performed by increasing the sampling rate over the entire cycle.

7.5 UWB POSITIONING AND COMMUNICATIONS

Due to the fine time resolution associated with UWB signals and, therefore, the possibility to have simultaneous timing, location, ranging, and communications integrated in a single UWB system it is an attractive candidate for combined future wireless communication and location positioning systems [85, 86]. Indeed, by transferring information for both location and control, such systems could extend the senses of people or machines into their own environment.

7.5.1 Potential user scenarios

7.5.1.1 Intelligent wireless area network (IWAN) As the current generation of narrowband networks does not enable context-aware services, such as asset tracking, alarm zones, etc., UWB positioning devices in a master-slave topology could be used in an IWAN to enable such context-aware services. A high density of devices (at least five per room) communicating at a medium to low data rate combined with positioning capability is used in an IWAN, which can cover medium to long distances. To be reliable in a smart home or office, such devices have to be very low cost and have very low power consumption as they have to be integrated into all the objects and assets that need to be intelligently controlled. Furthermore, a wireless bridge with the outside world could be implemented providing the capability to remotely control the sensors.

7.5.1.2 Sensor, Positioning, and Identification Network (SPIN) This scenario is more suitable for industrial factories or warehouses because of the very specific interference and propagation environment attributed to these buildings. As this environment continuously changes, adaptive systems and numerous links of reliability are required. Therefore, a SPIN uses a very high density of devices (hundreds per floor) communicating at a low data rate combined with positioning capability. These devices cover medium to long distances to a master station in the usual master-slave topology. However, an ad hoc topology is also possible.

7.5.2 Potential applications

Nowadays, as wireless communications depart from a centralized server system the position location technology associated with communication devices is becoming a requirement for many applications.

7.5.2.1 Personal location Positioning devices could be used in personal location systems when the user needs to precisely locate someone or something. For example, police officers or firefighters in action could know the position of colleagues or drivers could easily locate their cars in a large car park. There exist many other personal location applications where precise positioning would be required.

7.5.2.2 Inventory control In such applications, positioning devices could act as bar code identification tags and could give the precise location of the inventory item at the same time. Therefore, as an object is being moved out of stock, inventory control is instantly notified and, thus, real time information on the stock is possible.

7.5.2.3 Machine control Another application for positioning devices is to combine them with small robotic vehicles which could be used for difficult access infrastructure inspection, such as bridges or sewers, or in a harmful environment, such as nuclear plants. Indeed, due to its accurate location capability a robot could be monitored at a safe, remote location. Another possibility is to use reliable small robots in a house or office environment in order to perform some tasks without the need for supervision.

7.5.2.4 Smart highways Positioning devices could be used to assist autopilots in automobiles. Indeed, vehicles could be guided along a highway by integrated UWB sensors along the road. Furthermore, such sensors placed inside the vehicles might enable them to communicate and, thus, provide real time local intelligence in order to avoid accidents.

7.5.2.5 Smart homes and offices Integrating UWB sensors into home or office appliances, such as televisions, lamps, computers, etc., might enable technologies

that turn the desired appliances on or off by knowing the location of people in the home or the office as well as the location of the appliances.

7.5.2.6 Ad hoc networking An ad hoc network has no infrastructure and consists of several mobile terminals that can communicate with each other without fixed routers, even at the base stations. Each mobile node can act as a terminal and as a router. So, it can find and maintain a suitable route to other nodes in the network dynamically. Such a feature provides a remarkable increase in the level of autonomy compared with the traditional fixed communication infrastructure. However, the location of the mobile, the strict constraints for power consumption of battery-powered terminals, and multipath interferences are the main concerns in ad hoc networks.

When using UWB radio technology, two mobile terminals inside the network can determine their distance within 10 cm and arrival time delays at the receiver as small as a fraction of a nanosecond. Ad hoc networking using UWB technology is a novel application that is able to overcome the main limitations of traditional multihop solutions, such as power constraints, multipath propagation, and location of the mobile terminal.

7.6 SUMMARY

In this chapter the fundamentals of UWB positioning and location were described. The application of various wireless positioning techniques in terms of the operating frequency of the system was explained. Practical considerations, such as antenna and multipath effects, were briefly considered. The standard surveying and navigation system, GPS and differential GPS and their major error sources were addressed.

We also classified positioning systems according to the place where measurements and their evaluation take place, such as network-based, handset-based, or hybrid. Time resolution issues for narrowband and wideband systems as well as super-resolution techniques were discussed.

Finally, the most recent and accurate technique, the UWB approach, was explained. The combination of UWB positioning and communication as well as some potential applications concluded the chapter.

Problems

Problem 1. What are the effects of the following parameters on the performance of a positioning system?

(i) frequency;

(ii) bandwidth; and

(iii) multipath.

Problem 2. How are the following network-based techniques of location finding compared?

(i) received signal strength;

(ii) angle of arrival;

(iii) time of arrival;

(iv) time difference of arrival; or

(v) multipath fingerprinting.

Problem 3. Can you suggest some other possible applications of personal location using UWB systems?

Problem 4. How do UWB positioning systems differ from conventional location systems?

8

Applications using ultra wideband systems

There have been many ultra wideband applications in both the military and government area that have already been developed and shown to be viable products. Although this chapter will only provide an overview of a select few applications, it should give the reader an idea of what has already been accomplished and what applications are planned for the future in the commercial sector. In particular, we want to point out that, following the plan of this book, we do not examine radar applications, which are perhaps the most developed of all UWB applications. Rather, we wish to focus on communications, looking ahead to their use in consumer products. Unfortunately, here we are in a dilemma as the field has only now begun to bear fruit in the form of chipsets and prototype devices.

As test cases for both military and commercial use we examine the precision asset-based location for the military as developed by Multispectral Solutions [83]. For chipsets we look at well-known UWB startups, such as Time Domain and XtremeSpectrum, which have developed functional chipsets.

8.1 MILITARY APPLICATIONS

As with many wireless communication technologies the military has been the major driving force behind the development of UWB. In particular, radar applications have been developed by the military for many years. See, for example, [1] for a large number of UWB radar examples.

8.1.1 Precision asset location system

Using the location and communications aspects of UWB, one extremely interesting military application is the *asset location system* developed by Multispectral Solutions and the U.S. Navy [83]. This system, while military in design, has immediate commercial applications, since it is neither an offensive nor a defensive device, but rather a wireless system to improve logistics by knowing the location of containers and other large objects within a Navy ship at all times.

The major reason behind the development of the UWB precision asset location (PAL) system was the massive mobilization of military forces and goods in Desert Storm. It was reported that 40,000 containers were shipped; however, more than half, approximately 25,000 containers, had to be opened to check the contents due to inaccurate or lost paper invoices.

In Navy ships, narrowband radio frequency identification (RFID) tags have not worked well due to excessive multipath (i.e., large delay spreads) and limited accuracy. To overcome these obstacles a UWB PAL system was developed.

The system consisted of UWB tags which were placed on the devices whose location was to be measured and receivers which were placed at fixed locations in the cargo hold where the object locations were to be measured. The UWB tags consisted of a short-pulse transmitter with a measured peak output power of approximately 250 mW. Each burst of information consisted of 40 bits and was repeated at 5-s intervals. The instantaneous bandwidth of the transmitted pulses was approximately 400 MHz.

In this system the short-pulse width property was used to make fine time-of-arrival measurements of the signals from each of the uniquely identified tags.

Tests were conducted on this system in a military ship's cargo hold, which measured approximately 25 m by 30 m and was 9 m high. Tests were performed both with no cargo and with cargo. The cargo consisted of 20-ft ISO containers stacked singly and doubly and HMMWV(High Mobility Multipurpose Wheeled Vehicle) trucks. Tags were located on the top of the containers and HMMWVs.

The accuracy of the system was between 1–2 m rms for single-stacked containers and approximately 4 m for double-stacked ones. Depending on the particular tag and test, accuracy of less than 1 m was possible.

Further tests to determine the multipath environment showed that the typical delay spread was 3 μs, which is typically an order of magnitude greater than other indoor environments, such as offices. Frequency domain multipath nulls between 30 and 40 dB were also measured.

The PAL system shows the promise of UWB communications and ranging applications because short pulses can give extremely accurate results, even in extreme multipath environments. It was concluded for this particular application that UWB tags outperformed the current narrowband RF identification tags.

Example 8.1

For the PAL system described in this chapter, calculate the average transmitter power of the system (in nW). Assume that transmitting one bit requires 2.5 ns. Describe the output power of the system in W/MHz and dBm/MHz. Is this system above or below FCC Part 15 limits?

Solution

The peak power of the transmitter is 0.25 W, and the 40 bits of information are retransmitted after a 5-s break. The total time T required for 40 bits of information is

$$T = 40 \text{ bits} \times 2.5 \text{ ns} = 100 \text{ ns} \tag{8.1}$$

The amount of time required for transmission T_t is only 100 ns/5 s (i.e., 0.000002%). The average transmit power P_{av} is calculated as

$$P_{\text{av}} \text{ [W]} = P_{\text{peak}} T_t = 0.25 \times 20.0 \times 10^{-9} = 5.0 \times 10^{-9} = 5 \text{ [nW]} \tag{8.2}$$

Since the system has a bandwidth of approximately 400 MHz the output power of the system can be described as

$$P_{\text{av}} \text{ [W/MHz]} = \frac{P_{\text{av}}}{400} = 12.5 \times 10^{-12} \text{ [W/MHz]} \tag{8.3}$$

When converted to decibels

$$\begin{aligned} P_{\text{av}} \text{ [dBm/MHz]} &= 10 \times \log (P_{\text{av}}) + 30 \\ &= -109.03 + 30 \\ &= -79.03 \text{ [dBm/MHz]} \end{aligned} \tag{8.4}$$

Note here that the 30 dB added was to convert from dB to dBm. Since FCC regulations for Part 15 emissions are set at −41 dBm/MHz the power output of the precision location system is approximately 38 dB below FCC limits.

8.2 COMMERCIAL APPLICATIONS

At the time of writing this book the number of applications using UWB technology which have passed beyond the prototype stage to commercialization are limited. Consumer devices are even further ahead. Most UWB devices in actual use are confined to the government or military, and many are UWB radar devices, which in this book have taken a minor role. However, many consumer electronic companies are actively developing chipsets and wireless applications which will use a UWB physical layer.

In this section we list some of these companies and introduce their vision for commercial UWB applications.

8.2.1 Time Domain PulsON 200

Time Domain [4] has developed a UWB chipset commercialized under the PulsON brand. The first generation PulsON 100, the current PulsON 200 and future chipsets form their product range. Time Domain is the first company to pass the FCC certification procedure for a communications product. The product is known as the PulsON 200 Evaluation Kit. It is a general platform to help end developers make consumer products, including wireless communications, tracking, and radar. A photograph of the wireless radios developed for the PulsON 200 Evaluation Kit are shown in Figure 8.1.

Fig. 8.1 PulsON 200 Evaluation Kit UWB radios. [Photograph courtesy of Time Domain.]

The PulsON 200 Evaluation Kit has a maximum raw data rate of 9.6 Mbps over greater than 10 m in an environment free of obstacles and approximately 7 m in a residential or office environment. The detailed specifications of the PulsON 200 are shown in Table 8.1.

Future versions of PulsON are under development specifically for very high-bandwidth and low-power consumption devices with the primary application of wireless multimedia.

8.2.2 Time Domain UWB signal generator

Time Domain has also developed a signal generator which functions as a UWB transmitter. It can be used in conjunction with a digital oscilloscope to characterize the UWB channel. It can also be used with other UWB devices to examine

Table 8.1 Time Domain's PulsON 200 Evaluation Kit specifications.

Parameter	Value
Pulse repetition frequency	9.6 MHz
Data rates	9.6, 4.8, 2.4, 1.2, 0.6, 0.3, 0.15, 0.075 Mbps
Center frequency	4.7 GHz
Bandwidth	3.2 GHz (10 dB radiated)
Effective isotropic Radiated power (EIRP)	−11.5 dBm
Power consumption	12.2 W (transmit) 11.9 W (receive)
FCC compliance	Parts 15.517, 15.209
modulation	bi-phase, quadrature flip-time modulation

the effects of interference and coexistence. A photograph of the PulsON 200 signal generator is shown in Figure 8.2.

8.2.3 XtremeSpectrum

XtremeSpectrum [5] has developed a four-chip lineup for UWB applications. The chipset is named XS100 TRINITY. The four chips are the RF front end (XSI102), RF transceiver (XSI112), digital baseband (XSI122), and media access control (MAC) (XSI141). The chips were developed using low-cost 0.18 μm CMOS and SiGe technology.

The RF front end consists mainly of a low-noise amplifier to boost the received UWB signal. High-gain (20 dB) and low-gain (0 dB) modes are provided. The noise figure is 5.6 dB. The RF transceiver consists of transmitter and receiver circuitry and timing and bias/control circuitry. The digital baseband chip consists of an analog-to-digital converter (ADC), baseband circuitry, and the interface to the MAC chip. A flat, planar design omni-directional antenna 2.5 cm square has also been developed.

The detailed specification of the TRINITY chipset is shown in Table 8.2.

8.2.4 Intel corporation

The Intel Corporation [87] has been involved in UWB both in the standards and research areas. Intel's contributions to technical advancement and general industry knowledge have established Intel research and development (R&D) as a leader in UWB technology. In the company it is believed that UWB is well suited for high-speed, short-range, wireless personal area connectivity for PC and mobile devices. Intel's current efforts focus on the three key hurdles: increased knowledge of UWB

Fig. 8.2 PulsON 200 UWB signal generator. [Photograph courtesy of Time Domain.]

for the target usage models; creation of open standards for high-speed, short-range communications; and the need for worldwide regulatory approval. Intel's researchers are applying their expertise in CMOS radio design to understand the optimum requirements for low-power, low-cost UWB radios. Intel expects initial market deployment of standards-based UWB solutions to be sometime in the 2005-2006 time frame.

8.2.5 Motorola

Motorola's Semiconductor Product Sector [88] has recently announced that it will use the company's UWB technology in its own products. It also backed XtremeSpectrum's UWB proposal at the IEEE 802.15 Working Group.

The two companies have signed a memorandum of understanding to work together to bring UWB technology to market and, eventually, deliver joint products to the marketplace.

Table 8.2 XtremeSpectrum TRINITY chipset specifications.

Parameter	Value
Frequency band	3.1-10.6 [GHz]
MAC protocol	IEEE 802.15.3
Range	10 m
Modulation	Bi-phase
Pulse type	Monocycle
Data rates	25, 50, 75, 100 [Mbps]
Power consumption	200 mW
Output power	<1 mW
Bit error rate	10^{-9}
Coding rates	1, 3/4, 1/2
Network	Peer-to-peer, ad hoc, piconet

Motorola is looking at various alternatives in the UWB market, but has decided to partner with XtremeSpectrum in part because they already have a working silicon solution.

8.2.6 Communication Research Laboratory

In Japan the Communication Research Laboratory (CRL) has recently established a Project Group devoted to UWB in order to promote the R&D of UWB technologies. It has been investigating appropriate specifications for radio regulation on UWB systems [89]. The project focuses on the R&D of UWB systems in microwave and millimeter wavebands as a result of industrial demands and for academic novelty in research.

The Project Group consists of a collaboration among industry, academia, and government. The CRL then established a UWB Consortium together with industrial companies and universities with the support of the Yokosuka Research Park (YRP) in September 2002. The aims of this UWB Consortium are described as follows:

- R&D of all technologies for UWB wireless access systems.
- Implementation and experimental investigation of a test bed using a microwave and submillimeter waveband system (i.e., 960 MHz, 3.1–10.6 GHz, and 22–29 GHz).
- R&D of UWB systems in unused high-frequency bands, such as the millimeter waveband over 60 GHz.

- Establishment of transmission systems based on UWB, with low-cost and high-speed transmission over 100 Mb/s.
- Contribution to the standardization of UWB systems both in Japan and overseas.

8.2.7 General atomics

Although at the time of writing no commercial chipset has been rolled out, General Atomics [90] announced that a UWB chipset would be available at the end of 2003. General Atomics is pursing a communication product based on UWB multi-band technologies. The data rate is expected to be 120 Mbps. The target products are wireless transmission of multiple video streams and high-speed cable replacement.

8.2.8 Wisair

Wisair [91] is a startup company based in Israel which has developed a UWB chipset aimed at wireless indoor communications. Simultaneous multi-streaming of audio and video, broadband multimedia, and quality of service are targeted, to support a wide range of entertainment and communications applications. The name of the chipset is UBLink.

The most interesting technical feature of the Wisair chipset is that it uses the *multi-band* approach, dividing the ultra wideband wireless channel spectrum into several narrower bands. In this chipset, there are 30 possible sub-bands, of which 1–15 can be used. This approach provides flexibility of channels to use in a particular environment. To combat noise at a particular frequency, only a select subset of sub-bands can be used, rejecting low-quality sub-bands. For low bit rate applications, again a lower number of sub-bands can be used to save power or reliability increased by changing the coding rate. However, the disadvantage of this approach is complexity and cost. It will be interesting to see which UWB technologies provide the most benefit at the lowest cost.

Since the focus for this chipset is the consumer product market, low power consumption is a primary target. Thus, a power-saving standby mode has been developed as well. Details of the UBLink chipset are shown in Table 8.3.

Table 8.3 Parameters of the Wisair UBLink chipset.

Parameter	Value
Bit rates	20, 62.5, 83.3, 125 [Mbps]
Range	10–30 [m]
Power consumption	60–200 [mW]
Multi-band	1–15 [sub-bands]

8.2.9 Home networking and home electronics

One of the most promising commercial application areas for UWB technology is the wireless connectivity of different home electronic systems. It is thought that many electronics manufacturers are investigating UWB as the wireless means to connect devices, such as televisions, DVD players, camcorders, and audio systems, together to remove some of the wiring clutter in the living room. This is particularly important when we consider the bit rate needed for high-definition television that is in excess of 30 Mbps over a distance of at least a few meters. An example of a possible home-networking setup using high-speed wireless data transfer of UWB is shown in Figure 8.3.

Of course, UWB wireless connections to and from personal computers are also another possible consumer market area, with products expected in the next few years.

In [92] a proposal is made to use UWB as the wireless link in a ubiquitous "homelink", which consists of an amalgamation of wired and wireless technologies. The wired technology proposed by the authors is based on the IEEE 1394 standard. This is an attempt to effectively integrate entertainment, consumer communications, and computing within the home environment. The reason for the choice of IEEE 1394 is that it provides an *isochronous mode*, in which data are guaranteed to be delivered within a certain time frame after transmission has started. Bandwidth is reserved in advance, which gives a constant transmission

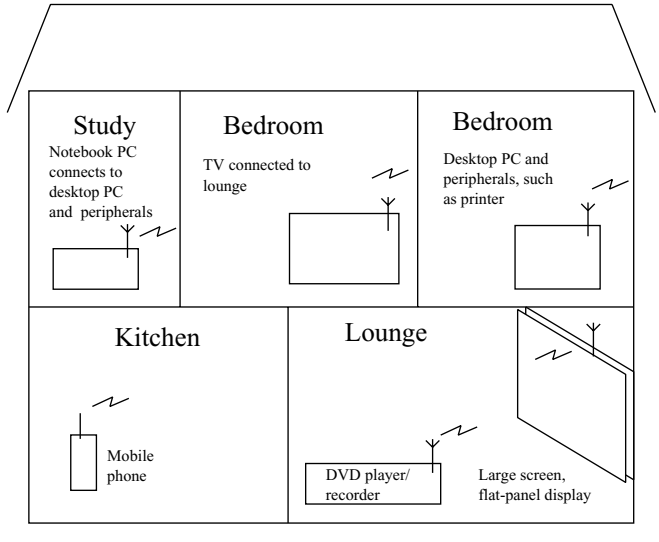


Fig. 8.3 An example of a possible home-networking setup using UWB.

speed. This is important for real time applications, such as video broadcasts, to ensure that there is no break in the movie or television program for the viewer. Some possible services and required data rates are shown in Table 8.4.

Table 8.4 Some possible contents for a home entertainment and computing network, the necessary data rates, and the possible need for real time features.

Service	Data rate [Mbps]	Real time feature
Digital video	32	Yes
DVD, TV	2–15	Yes
Audio	1.5	Yes
Internet access	>10	No
PC	32	No
Other	<1	No

Furthermore, IEEE 1394 has an *asynchronous* mode, in which data are guaranteed to be delivered, but bandwidth is not reserved and no guarantee is made about the time of arrival of the data.

IEEE 1394 provides scalable performance with 100, 200, and 400 Mbps, which is comparable with the target for UWB transmission speeds. IEEE 1394b is under consideration and will support 800 to 1,600 Mbps and may be extended to 3,200 Mbps.

The considerations for home appliances can be described as economy, easy operation, flexibility, and high reliability. Particularly in the area of economy (that is cheap devices) and reliability, UWB can be expected to perform well. However, there are various other wireless systems that are targeting this application. In particular, the established IEEE 802.11a standard [93], which has data rates of up to 54 Mbps, is a strong contender. IEEE 802.11a chipsets have been reported to have speeds of greater than 70 Mbps and are expected to increase past 100 Mbps in the near future. Alternatively, wireless transceivers based on the 802.11g standard may also provide an economical, if slightly lower data rate.

Bluetooth [94], HomeRF [95], and 802.11b standards are not strong contenders, because their maximum data rates are approximately 720 kbps, 1.6 Mbps, and 11 Mbps, respectively. Another possible alternative is Wireless 1394, a wireless extension to the wired 1394 system.

8.2.10 Precision asset location system

In the previous section we looked at the military use of the precision asset location (PAL) system based on UWB devices, developed by Multispectral Solutions, Inc. In [96] the further commercialization and testing of the PAL system was reported. In particular, the PAL system was extended from military ships to use in hospitals

and factories. The product was named PAL650 and was certified according to the current (2003) FCC rules.

It was delivered as a commercial product at the time of writing to such clients as the Washington National Medical Center. The UWB precision asset location system enables hospital administrators to track the utilization of assets and patient flow accurately within the hospital. Tracking of patients and equipment is expected to increase hospital efficiency, particularly in the case of a terrorist attack or mass disasters, where hospitals see four to five times the normal load of patients.

Another client is the National Institute of Standards and Technology (NIST) in the U.S. for precision tracking of robotic vehicles for search and rescue. The system will initially be used to evaluate robotic vehicle performance during international search and rescue tournaments. Future use will involve the tracking of rapidly deployable search and rescue vehicles as they enter buildings, following a disaster.

One technical item of interest for PAL systems is that they do not require a high data rate and, thus, may use significantly higher peak powers than is allowed for high data rate communication systems. The peak measured emission of the PAL650 system was 58.13 dBμV/m with an average emission of 39.01 dBμV/m. All measurements were performed with a 1-MHz resolution bandwidth (RBW) and referenced to a 3-m range. It is interesting to note that average emission measurements were limited by the testing equipment noise floor and the true average effective isotropic radiated power (EIRP) for PAL650 can be calculated to be 16.1 dBμV/m. This value is approximately 37 dB below the FCC limit and over 22 dB below the noise floor. For reference the current (2004) FCC limits are 61.2 dBμV/m peak and 53.9 dBμV/m average.

The commercialization of PAL650 leads to the requirement of long battery life. Operation lifetimes in excess of 3.8 years are expected for these tags using a single lithium cell CR2477 (3.0 V 1A-hr). The tag operates at 3.0 V with a current consumption of approximately 30 μA.

As with the previously described PAL system, PAL650 consists of a set of active tags, UWB receivers, and a central processing hub that processes the received signals from the active tags.

A photograph of an active tag with the polyethylene radome cover is shown in Figure 8.4. The size of the tag is approximately 4.75 cm in diameter and 2.25 cm high when not covered in the housing. With the radome housing the size expands to 5.1 cm and 2.85 cm. The UWB active tag has a center frequency of 6.191 GHz and an instantaneous −10 dB bandwidth of 1.25 GHz.

A photograph of the receiver is shown in Figure 8.5. The size of the receiver boards are approximately 5.5 cm by 9.0 cm and 2.5 cm height. When the housing for the receiver unit is included the total size becomes 5 cm by 8.25 cm by 15.25 cm. The low-noise receiver front end is housed in the gray box in the top left of Figure 8.5.

Receivers obtain power via the central processing unit, and data are transferred by a wired system. The wireless part is between the receivers and the active tags.

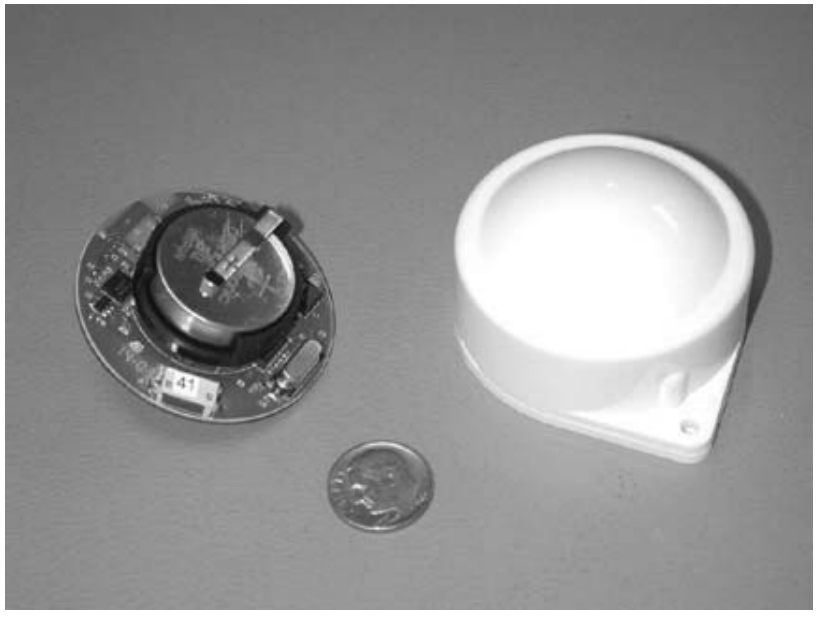


Fig. 8.4 PAL650 UWB active tag with radome. ©IEEE 2003.

The system described in [96] is a daisy-chained one; however, a hub-and-spoke system is under development to eliminate the potential for a single point of failure in the serial communications and power distribution to render the system unusable. The hub-and-spoke system will increase the reliability of the system as a whole.

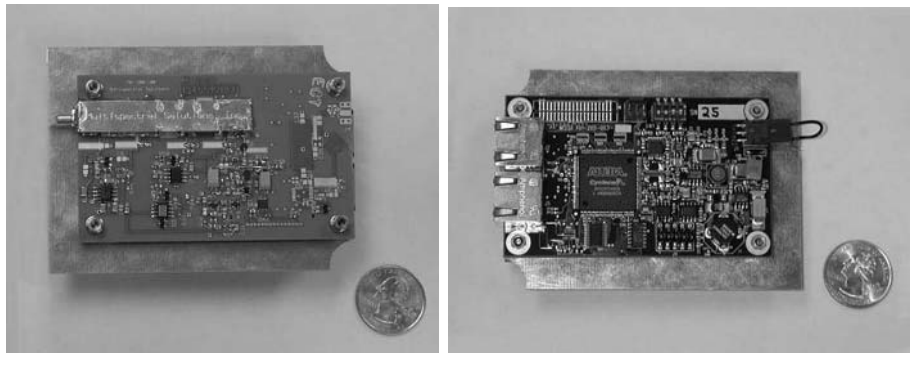


Fig. 8.5 PAL650 receiver RF board (left) and digital board (right). ©IEEE 2003.

A photograph of the central processing hub is shown in Figure 8.6. The central processing unit receives signal from the receivers and calculates the position of the active tags.

Fig. 8.6 PAL650 central processing hub. ©IEEE 2003.

Calibration is performed at startup using a reference tag, whose location is known in advance. The current system receives updates from tags once per second in a burst of 72 pulse (bits) at a 1-Mbps burst rate. These data include synchronization preamble, tag identification, data field, forward error correction, and control bits. Update rates of up to 5,200 per second can be accommodated without exceeding present FCC limits.

8.3 SUMMARY

In this chapter we have investigated a brief selection of current UWB applications, focused on consumer communications. We have presented material about current wireless UWB chipsets, by such companies as Time Domain, XtremeSpectrum, and Wisair. As a test case we examined the precision asset location for the military as developed by Multispectral Solutions. We noted that consumer products are expected to appear within the next few years.

Problems

Problem 1. Survey three current UWB products. Compare and contrast the specifications of each.

Problem 2. Design an original UWB product. What features does it possess to make it a success in the marketplace? Why must your product use UWB, as opposed to other wireless technologies?

Problem 3. Investigate how the ADC in the XtremeSpectrum chip achieves its very high sampling rate.

Problem 4. Investigate the specifications of PAL650. Is it a pulse-based UWB system?

References

1. J. D. Taylor, editor. *Ultra-Wideband Radar Technology.* CRC Press, 2001.

2. J. D. Taylor, editor. *Introduction to Ultra-Wideband Radar Systems.* CRC Press, 1995.

3. G. F. Ross. Transmission and reception system for generating and receiving base-band duration pulse signals without distortion for short base-band pulse communicaton system. *U.S. Patent 3,728,632*, April 1973.

4. Time Domain. http://www.timedomain.com.

5. XtremeSpectrum. http://www.xtremespectrum.com.

6. T. W. Barrett. History of ultra wideband (UWB) radar & communications: Pioneers and innovators. In *Proceedings of Progress in Electromagnetics Symposium 2000 (PIERS2000)*, July 2000.

7. C. L. Bennett and G. F. Ross. Time-domain electromagnetics and its applications. *Proceedings of the IEEE*, 66:299–318, March 1978.

8. J. Williams. The IEEE 802.11b security problem, part 1. *IT Professional*, pages 91–96, November 2001.

9. F. Ramirez-Mireles and R. A. Scholtz. Wireless multiple-access using SS time-hopping and block waveform pulse position modulation, part 2: Multiple-access performance. In *Proceedings ISITA Symposium*, October 1998.

10. M. Z. Win and R. A. Scholtz. Ultra-wide bandwidth time-hopping spread spectrum impulse radio for wireless multiple-access communication. *IEEE Transactions on Communications*, 48(4):679–691, April 2000.

11. D. G. Leeper. Wireless data blaster. *Scientific American*, May 2002.

12. H. Kikuchi. UWB arrives in Japan. *Nikkei Electronics*, pages 95–122, February 2003.

13. R. Mark. XtremeSpectrum rolls out first UWB chipset. InternetNews Website, June 2002.

14. *FCC regulations, 47CFR Section 15.5 (d).* http://ftp.fcc.gov, 1998.

15. R. A. Scholtz. Multiple access with time-hopping impulse modulation. In *IEEE MILCOM93, vol. 2*, October 1993.

16. J. T. Conroy, J. L. LoCicero, and D. R. Ucci. Communication techniques using monopulse waveforms. In *IEEE MILCOM99, vol. 2*, November 1999.

17. M. Ghavami, L. B. Michael, S. Haruyama, and R. Kohno. A novel UWB pulse shape modulation system. *Kluwer Wireless Personal Communications Journal*, 23:105–120, 2002.

18. J. Mills, editor. *Radio Communication Theory and Methods.* McGraw-Hill, 1917.

19. E. Kreyszic. *Advanced Engineering Mathematics.* John Wiley & Sons Inc., 1988.

20. J. B. Martens. The hermite transform – theory. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:1595–1606, 1990.

21. M. R. Walton and H. E. Hanrahan. Hermite wavelets for multicarrier data transmission. In *South African Symposium on Communications and Signal Processing ComSIG 93*, August 1993.

22. J. M. Cramer, R. A. Scholtz, and M. Z. Win. On the analysis of UWB communication channels. In *IEEE MILCOM99*, November 1999.

23. M. Ghavami, L. B. Michael, and R. Kohno. Hermite function-based orthogonal pulses for ultra wideband communication. In *WPMC'01*, September 2001.

24. M. Z. Win and R. A. Scholtz. Impulse radio: How it works. *IEEE Communications Letters*, 2:36–38, 1998.

25. D. Slepian. Prolate spheroidal wave functions, fourier analysis and uncertainty V: The discrete case. *Bell System Technical Journal*, 57, 1978.

26. R. S. Dilmaghani, M. Ghavami, B. Allen, and H. Aghvami. Novel pulse shaping using prolate spheroidal wave functions for UWB. In *IEEE PIMRC 2003 Beijing, China*, 2003.

27. C. Flammer. *Spheroidal Wave Functions*. Stanford University Press, 1957.

28. G. Arfken. *Mathematical Methods for Physicists*, chapter 12: Legendre Functions. Academic Press, 3rd edition, 1985.

29. N. W. Bailey. On the product of two Legendre polynomials. *Proceedings of Cambridge Philos.*, 29:173–177, 1933.

30. J. M. Wilson. Ultra wideband technology update at spring 2003. *Intel Developer UPDATE Magazine*, pages 1–9, 2003.

31. New ultra-wideband technology, white paper. *Discrete Time Communications*, pages 1–8, 2002.

32. H. F. Harmuth. Radio signals with large relative bandwidth for over-the-horizon radar and spread spectrum communications. *IEEE Transactions on Electromagn. Compat.*, 20:501–512, 1978.

33. J. R. Davis, D. J. Baker, J. P. Shelton, and W. S. Ament. Some physical constraints on the use of carrier free waveforms in the radio-wave transmission systems. *Proceedings of IEEE*, 67:884–890, June 1979.

34. H. P. Hsu. *Analog and Digital Communications*. McGraw-Hill, 1993.

35. A. V. Oppenheim, A. S. Willsky, and I. T. Young. *Signals and Systems*. Prentice Hall, 1983.

36. A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.

37. P. P. Newaskar, R. Blazquez, and A. P. Chandrakasan. A/D precision requirements for an ultra-wideband radio receiver. In *SIPS 02*, October 2002.

38. W. Ellersick, C. K. Ken Yang, W. Horowitz, and W. Dally. Gad: A 12gs/s cmos 4-bit A/D converter for an equalized multi-level link. In *Symposium on VLSI Circuits, Digest of Technical Papers*, 1999.

39. M. I. Skolnik. *Introduction to Radar Systems*. McGraw-Hill Book Co, 1962.

40. D. G. Fink and D. Christiansen. *Electronics Engineers Handbook*. McGraw-Hill Book Co, 1975.

41. T. E. McEwan. Ultra-wideband radar motion sensor. *US Patent 5,361,070*, 1994.

42. J. R. Foerster. The effects of multipath interference on the performance of UWB systems in an indoor wireless channel. In *Spring Vehicular Technology Conf.*, May 2001.

43. H. Hashemi. Impulse response modeling of indoor radio propagation channels. *IEEE Journal on Selected Areas in Communications*, 11:967–978, 1993.

44. K. Pahlavan and A. Levesque. *Wireless Information Networks.* John Wiley & Sons, Inc., 1995.

45. M. Z. Win and R. A. Scholtz. On the robustness of ultra-wide bandwidth signals in dense multipath environments. *IEEE Communications Letters*, 2:10–12, 1998.

46. A. A. Saleh and R. A. Valenzuela. A statistical model for indoor multipath propagation. *IEEE Journal of Selected Areas in Communications*, 5:128–137, 1987.

47. H. Suzuki. A statistical model for urban radio propagation. *IEEE Transactions on Communications*, 25:673–680, 1977.

48. R. Ganesh and K. Pahlavan. Statistical modeling and computer simulation of indoor radio channel. *IEE Proceedings, part I(3)*, 138:153–161, 1991.

49. S. S. Ghassemzadeh, R. Jana, C. Rice, W. Turin, and V. Tarokh. A statistical path loss model for in-home UWB channels. In *IEEE UWBST*, May 2002.

50. J. Foerster and Q. Li. UWB channel modeling contribution from Intel. Technical report, IEEE document, 2002.

51. K. Siwiak and A. Petroff. A path link model for ultra wide band pulse transmission. In *IEEE Vehicular Technology Conference 2001*, pages 1173–1175, May 2001.

52. D. Cassioli, M. Z. Win, and A. R. Molisch. The ultra-wide bandwidth indoor channel: From statistical model to simulations. *IEEE Journal on Selected Areas in Communications*, 20:1247–1257, 2002.

53. A. Armogida, B. Allen, M. Ghavami, M. Porretta, and H. Aghvami. Path loss modeling in short-range UWB transmissions. In *International Workshop on UWB Systems, IWUWBS2003*, June 2003.

54. W. C. Stone. Nist construction automation report no. 3: Electromagnetic signal attenuation in construction materials. Technical report, BFRL Publications, 1997.

55. T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

56. W. Turin, R. Jana, S. S. Ghassemzadeh, C. W. Rice, and V. Tarokh. Autoregressive modeling of an indoor UWB channel. In *IEEE UWBST*, May 2002.

57. S. Howard and K. Pahlavan. Autoregressive modeling of wide-band indoor radio propagation. *IEEE Transactions on Communications*, pages 1540–1552, September 1992.

58. S. L. Marple. *Digital Spectral Analysis*. Prentice Hall, 1987.

59. J. G. Proakis. *Digital Communications*. Addison Wesley, 4th edition, 2000.

60. J. McCorkle. Why such uproar over ultrawideband? Communication Systems Design Website, March 2002. http://www.commsdesign.com/csdmag/sections/feature_article/OEG20020301S0021.

61. L. Zhao and A. M. Haimovich. The capacity of a UWB multiple access communication system. In *IEEE International Conference on Communications, ICC '02*, pages 1964–1968, May 2002.

62. K. Eshima, K. Mizutani, R. Kohno, Y. Hase, S. Oomori, and F. Takahashi. Comparison of ultra-wideband (UWB) impulse radio with DS-CDMA and FH-CDMA. In *Proceedings of 24th Symposium on Information Theory and Applications (SITA), Kobe, Japan*, pages 803–806, 2001. In Japanese.

63. T. Ikegami and K. Ohno. Interference mitigation study for UWB impulse radio. In *IEEE PIMRC 2003*, September 2003.

64. M. Hämäläinen, J. Saloranta, J. P. Makela, I. Opperman, and T. Pantana. Ultra wideband signal impact on IEEE 802.11b and bluetooth performance. In *IEEE PIMRC 2003*, September 2003.

65. M. Luo, M. Koenig, D. Akos, S. Pullen, and P. Enge. Potential interference to GPS from UWB transmitters phase II test results accuracy, loss-of-lock, and acquisition testing for GPS receivers in the presence of UWB signals. Technical Report 3.0, Stanford University, March 2001.

66. J. P. Van't Hof and D. D. Stancil. Ultra-wideband high data rate short range wireless links. In *IEEE Vehicular Technology Conference 2002*, pages 85–89, 2002.

67. D. K. Cheng. *Field and Wave Electromagnetics*. Addison-Wesley, 1989.

68. I. I. Immoreev and A. N. Sinyavin. Features of ultra-wideband signals' radiation. In *UWBST 2002 IEEE Conference on Ultra Wideband Systems and Technologies*, May 2002.

69. F. Sabath. Near field dispersion of impulse radiation. In *URSI General Assembly 2002*, August 2002.

70. C. Balanis. *Antenna Theory.* John Wiley & Sons Inc., 1997.

71. K. Y. Yazdandoost and R. Kohno. Ultra wideband antenna, CRL report.

72. Farr Research Inc. http://www.farr-research.com/.

73. S. Ramo and J. Whinnery. *Fields & Waves in Modern Radio.* John Wiley & Sons, Inc., 1962.

74. B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode. Adaptive antenna systems. *Proceedings of IEEE*, 55:2143–2159, December 1967.

75. M. G. M. Hussain. An overview of the principle of ultra-wideband impulse radar. In *CIE 1996 International Conference of Radar*, November 1996.

76. M. G. M. Hussain. Antenna patterns of nonsinusoidal waves with the time variation of a gaussian pulse - part I. *IEEE Transactions on Electromagn. Compat.*, 30:504–512, 1988.

77. CDMA Development Group. CDG: Test plan document for location determination technologies evaluation, 2000.

78. N. Lenihan and S. McGrath. REALM: Analysis of alternatives for location positioning.

79. K. Pahlavan, X. Li, and J. Makela. Indoor geolocation science and technology. *IEEE Communications Society Magazine*, February 2002.

80. M. O. Sunay and I. Tekin. Mobile location tracking for IS-95 using the forward link time difference of arrival techniques and its application to zone-based billing. In *IEEE GLOBECOM Conference 1999.*

81. X. Li. *Super-Resolution TOA Estimation with Diversity Techniques for Indoor Geolocation Applications.* PhD thesis, 2003.

82. R. J. Fontana. Experimental results from an ultra wideband precision geolocation system. *Ultra-Wideband, Short-Pulse Electromagnetics IV.* Kluwer Academic/Plenum Publishers, May 2000.

83. R. J. Fontana and S. Gunderson. Ultra-wideband precision asset location system. In *UWBST 2002 IEEE Conference on Ultra Wideband Systems and Technologies*, May 2002.

84. I. Maravic, M. Vetterli, and K. Ramchandran. Channel estimation and synchronization with sub-Nyquist sampling and application to ultra-wideband systems. In *ISCAS 2004*.

85. R. Fleming and C. Kushner. Low-power miniature distributed position location and communication devices using ultra wideband, nonsinusoidal communication technology. Technical report, AetherwireLocation Inc., July 1995.

86. D. Porcino and W. Hirt. Ultra-wideband radio technology: Potential and challenges ahead. *IEEE Communications Magazine*, July 2003.

87. Intel Corporation. http://www.intel.com/technology/ultrawideband/.

88. Motorola Corporation. http://www.motorola.com/.

89. Communication Research Laboratory. http://www2.crl.go.jp/.

90. General Atomics. http://www.fusion.gat.com/photonics/uwb/.

91. Wisair. http://www.wisair.com.

92. M. Nakagawa, H. Zhang, and H. Sato. Ubiquitous homelinks based on IEEE 1394 and ultra wideband solutions. *IEEE Communications Magazine*, 41(4):74–82, April 2003.

93. 802.11 Standard. Draft supplement to standard for telecommunications and information exchange between systems - LAN/MAN specific requirements-part 11: Wireless MAC and PHY specifications: High speed physical layer in the 5 GHz band. *P802.11a/D6.0*, May 1999.

94. J. C. Harrtsen. The bluetooth radio system. *IEEE Personal Communications*, 7(1):28–36, February 2000.

95. K. J. Negus, A. P. Stephens, and J. Landsford. Homerf: Wireless networking for the connected home. *IEEE Personal Communications*, 7(1):20–27, February 2000.

96. R. J. Fontana, E. Richley, and J. Barney. Commericalization of an ultra wideband precision asset location system. In *UWBST 2003 IEEE Conference on Ultra Wideband Systems and Technologies*, November 2003.

Index