Principles of Digital Data Transmission

A. P. Clark

Principles of Digital Data Transmission

A. P. CLARK, MA, PhD, DIC, MIERE, CEng.
Department of Electronic and Electrical Engineering,
Loughborough University of Technology

PENTECH PRESS London: Plymouth

First Published, 1976
by Pentech Press Limited
Estover Road, Plymouth PL6 7PZ

ISBN 0 7273 1603 6 (hard cover)
0 7273 1604 4 (paper cover)

Printed and bound in Great Britain by
Billing & Sons Ltd, Guildford and London

PRINCIPLES OF
DIGITAL DATA
TRANSMISSION

Preface

This book is the outcome of some lecture courses in digital communications presented over the past five years to both undergraduate and postgraduate students in the Department of Electronic and Electrical Engineering of Loughborough University of Technology.

The book is written mainly for private study by practising and student engineers who are interested in the design and development of data-transmission systems. It is concerned primarily with basic principles and techniques rather than with details of equipment design.

Part 1 of the book (Chapters 1-11) is a non-mathematical survey of the properties of the voice-frequency channels formed by telephone circuits and HF radio links, and of the various techniques that have been used or proposed for the transmission of digital data over these channels. The different techniques are compared and descriptions are given of the preferred data-transmission systems. The voice-frequency channels are important not only in their own right but also because they introduce most of the different types of noise and distortion experienced over many other channels. The study of data-transmission over these channels is therefore relevant to many other applications.

Part 2 of the book (Chapters 12-16) is a theoretical analysis and comparison of various different signals that may be used for the transmission of digital data. Both baseband and modulated-carrier signals are studied, with particular emphasis on some of the baseband signals that may be obtained over partial-response channels and by correlative-level coding. Matched-filter detection is studied in some detail as is the related topic of the optimum combination of transmitter and receiver filters.

Part 2 of the book presents the theoretical analysis necessary for the full understanding of the more important topics studied in Part 1. An idealised system is assumed in Part 2, not only to facilitate the theoretical analysis but also to bring out a number of interesting relationships that exist between the different signals studied.

A total of some 474 references, suitably classified according to the topics covered, are provided to enable the reader to pursue in more detail the particular topics that are relevant to the subject matter of the book but are only briefly mentioned here.

The book is concerned primarily with the situation where it is required to transmit digital data as simply and effectively as possible over a given channel but without necessarily achieving an exceptionally high transmission rate. The systems studied use basically simple detection processes and do not involve the more sophisticated techniques such as adaptive equalizers. The proper study of the latter requires rather more specialised mathematics than is considered appropriate for this book, so that a detailed consideration of these techniques has not been included here. The book attempts to establish as clearly as possible the important properties of the different digital signals that may be used for the transmission of data, and to compare these properties in such a way that the most suitable modem design for any given application can be determined.

Part 1 of the book is written at a relatively elementary level and assumes only a limited knowledge of mathematics. Part 2 assumes a basic (but not advanced) understanding of Fourier transforms, probability theory, random variables, random processes and in particular the Gaussian random process, signal transmission through linear systems, the sampling theorem and linear modulation methods. The book is suitable for presentation as part of a course on digital communications, at either a final-year undergraduate or first-year postgraduate level.

Many of the results and conclusions presented in Part 1 of the book are the outcome of a study of practical data-transmission systems carried out by the author from 1956-1965 and again from 1969-1970, while working at Plessey Telecommunications Research Ltd., Taplow, Buckinghamshire. The author is grateful for the opportunity that was provided to undertake this work. The author would also like to thank Professor J. W. R. Griffiths for his encouragement in the later part of the work and for providing the necessary facilities for its completion.

Contents

PART 2 DIGITAL-SIGNAL THEORY

1 Introduction

The most widely used communication channel is undoubtedly the voice-frequency channel designed for the transmission of speech [1-45]. This passes a band of frequencies from approximately 300 to 3000 Hz. Almost all major towns and cities throughout the world are interconnected by the telephone network, using voice-frequency channels [1-31]. With the rapidly increasing use of computers and the corresponding increase in the quantity of digital data transmitted to and from the larger computer installations, there is considerable interest in using the telephone network for the transmission of digital data. It seems likely that over the coming years there will be a steady increase in the ratio of digital data to speech signals, carried by the telephone network.

For communication between isolated locations and over long distances, voice-frequency channels over HF radio links are often used [32-45]. In some important applications these are the most cost effective communication channels for the transmission of digital data.

In Chapters 2 to 11 we shall be concerned with two main types of voice-frequency channel: *telephone circuits* and *HF radio links.* A telephone circuit is, of course, a connection (voice-frequency channel) between one subscriber (user) and another, over the telephone network. We shall be concerned with the problems involved in transmitting digital data over these channels, and with the techniques for achieving satisfactory operation. We shall study the design of the *modem* (modulator-demodulator) which both generates the digital data signals that are transmitted over a voice-frequency channel to another location, and also detects the digital data signals transmitted back from that location, often (but not always) over a different voice-frequency channel. The emphasis throughout will be on basic principles rather than on the detailed designs of particular modems. We shall *not* be concerned with the design of the voice-frequency channel itself or with techniques for improving its characteristics.

An important point that requires to be mentioned here is that the basic types of noise and distortion experienced over typical voice-frequency channels include those experienced over many other quite different transmission paths, so that the study of digital data transmission over voice-frequency channels is in fact relevant to many other applications.

Chapters 2 to 11 present an elementary survey of the more important problems involved in transmitting digital data over voice-frequency channels and of the various conventional techniques that can be used to overcome these problems. The use of mathematical analysis is avoided altogether. For the reader wishing to pursue particular points in more detail, an extensive list of references is provided[1-421].

In Chapters 2 to 11, simple intuitive explanations are provided where possible for the results quoted. Where no simple explanations are available, the results are presented without further justification. The reader must not therefore expect every statement in this part of the book to follow directly or logically from the preceding discussion. The aim of this method of presentation is to collect together the more important results and relationships, without the distraction of detailed mathematical analysis, in the hope that the reader may thereby be enabled to see the wood for the trees! Indeed, the whole purpose of the Chapters 2 to 11 is to give the reader a sense of values, as far as data-transmission is concerned, so that he can discriminate between the things that matter and those that do not. The detailed theoretical analysis, needed for the full understanding of some of the more important topics in Chapters 2 to 11, is given in the Chapters 12 to 16.

2
Data signals

2.1 Power-density spectrum

Any time-varying signal with a given shape or waveform can be considered alternatively as a set of frequency components (separate sine-waves with different frequencies) which occupy a certain range of frequencies. The frequency components form the *spectrum* of the signal and are given by the Fourier transform of the signal waveform. In the spectrum of a given signal waveform, both the amplitudes and phases of the individual frequency components are defined. The range of frequencies occupied by the spectrum gives the *bandwidth* of the signal. In general, the more slowly a signal varies in time, the lower are the frequencies of its frequency components and so the lower are the frequencies occupied by its spectrum. The more rapidly a signal varies in time, the higher are these frequencies. The components of the transmitted signal, carrying the individual digits of the transmitted data, are pulses of suitable shape and (for practical purposes) finite duration. The pulses are known as *signal elements.* As a rule, any reduction in the signal bandwidth tends to lengthen the transmitted signal elements, since, for a given pulse shape, the width of the frequency spectrum is inversely proportional to the duration of the pulse.

The spectrum (Fourier transform) of the transmitted signal, corresponding to a given message (sequence of digits or element values), is not normally the same as the spectrum of the transmitted signal corresponding to any other message. It is assumed here, for the purpose of the comparison, that the message of finite duration is repeated continuously to give a resultant signal of infinite duration. The transmitted signal now has a *line* spectrum, with a set of *discrete* frequency components, whose levels are measured in units of *power* rather than *energy*. As the length of the message increases and assuming that the digital data is random in nature (without repetitive sequences), so the *power* in any given unit bandwidth tends towards a given value, which is the same as that obtained with any other message of the type being considered. In the limit, as the duration of a message tends to infinity, its spectrum becomes *continuous*, with a given relationship between the power per-unit-bandwidth and frequency, that is, with a given *power-density spectrum.* The power at any particular frequency is now vanishingly small, which is why the power *density* and not the power is considered.

It is conventional, on the grounds of mathematical convenience, to evaluate the power-density spectrum over both positive and negative frequencies, to give a two-sided power density spectrum. For the present, however, we shall consider instead the one-sided power density spectrum, which shows directly the variation of power density (power per unit bandwidth) with frequency. This has the property that the area beneath the curve and bounded to the left and right by two different frequencies, gives the average power over the frequency range bounded by the two frequencies. Clearly, the area under the whole of the curve gives the average signal power.

2.2 Baseband signals

The simplest digital data signal contains a sequence of signal elements (units or pulses of the data signal) where each element is binary coded, having the choice of two possible shapes which correspond to the element values 0 and 1. Each signal element has the same duration of T seconds and follows immediately after the preceding element, so that the signal element rate is $1/T$ elements per second or bauds. An example of such a signal is shown in Figure 2.1.

This is a binary antipodal baseband signal, where the element value 0 is represented by a signal value k, and the element value 1 is represented by a signal value $-k$. A baseband signal is one whose spectrum usually extends to zero frequency (d.c.) or to very low frequencies, and which carries

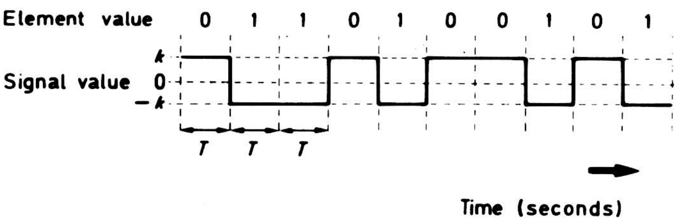

Fig. 2.1. Rectangular binary baseband signal.

information (data) in terms of its *values* at certain points. Thus the above waveform could alternatively be shaped as in Figure 2.2. At the receiver,

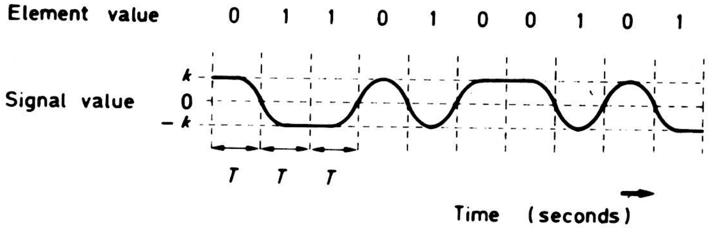

Fig. 2.2. Rounded binary baseband signal

each signal element is here sampled at its mid point in time and each sample has a value $\pm k$.

It can be seen that the second of the above two waveforms does not contain as rapid variations with time as does the first waveform, so that its spectrum does not extend to such high frequencies. Both waveforms, however, contain the same slower variations, so that their spectra do not differ greatly at the lower frequencies. Both spectra extend to zero frequency. Clearly, the second of the two waveforms has a narrower bandwidth, so that it achieves a higher ratio of transmission rate to signal bandwidth, and in this sense it makes a more efficient use of bandwidth.

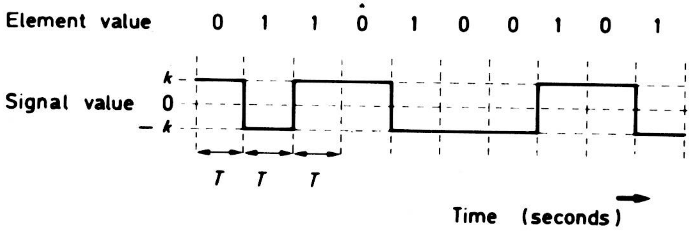

Fig. 2.3. Differentially coded binary signal

Differential coding may alternatively be used to give the signal waveform in Figure 2.3. An element value 1 is represented here by a *change* in value between adjacent signal elements, and an element value 0 by no change. Again, a rounded waveform could be used instead, such that an element value 1 is represented by a change in the sample values of two adjacent elements, and an element value 0 by no change.

2.3 Modulated-carrier signals

In order to obtain satisfactory transmission over a typical voice-frequency channel, signal waveforms of the types just considered could not in general be used. The reasons for this are considered in Section 6.1. The baseband signal now modulates a suitable sine-wave carrier, using amplitude, frequency or phase modulation (AM, FM or PM), and the modulated-carrier signal is then transmitted over the voice-frequency channel. The *signal carrier* is the sine wave which carries the transmitted data in terms of the amplitude, frequency or phase of this sine wave. At the receiver, the baseband data signal is extracted from the received modulated-carrier signal, be a suitable demodulation process, and the demodulated baseband signal is then detected to give the sequence of data element values.

Consider a baseband signal with the power-density spectrum shown in Figure 2.4. The power density here is zero at d.c. and at frequencies above f_m Hz.

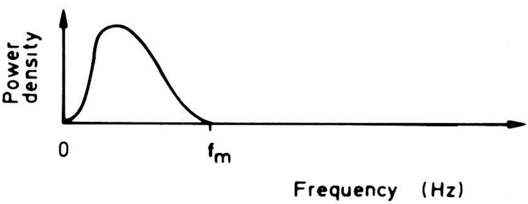

Fig. 2.4. Spectrum of baseband signal

The baseband signal is used to amplitude modulate a sine-wave carrier with frequency f_c Hz, so that the amplitude (level) of the carrier varies as the baseband signal. This gives a double sideband AM signal with the power-density spectrum shown in Figure 2.5.

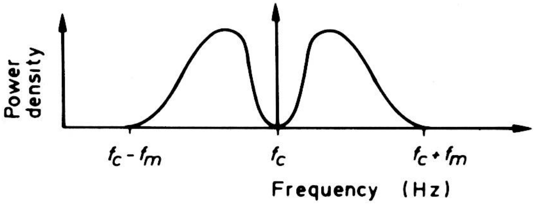

Fig. 2.5. Spectrum of double sideband AM signal

The impulse at f_c Hz (shown by the arrow in Figure 2.5) indicates that there is a non-zero frequency component at f_c. The upper sideband (part of the spectrum above f_c) is the spectrum of the baseband signal, unchanged in shape but shifted up in frequency by f_c Hz. The lower sideband (part of the spectrum below f_c) is the 'reflection' of the upper sideband in the carrier frequency f_c. It is assumed here that $f_c > f_m$.

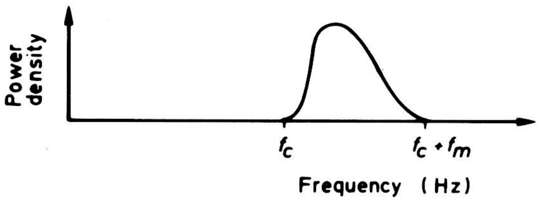

Fig. 2.6. Spectrum of single sideband suppressed carrier AM signal

If now the lower sideband and the carrier frequency are removed by a filter, which passes the upper sideband through unchanged, the signal becomes a single sideband suppressed carrier AM signal, with the spectrum shown in Figure 2.6.

The spectrum of the single sideband suppressed carrier AM signal is simply the spectrum of the baseband signal shifted up in frequency by f_c Hz. Although the *shape* of the spectrum is unchanged, the *levels* of all frequency components may have been increased or decreased by a fixed amount. A single sideband suppressed carrier AM signal may alternatively use the lower sideband of the original double-sideband signal.

The double sideband AM signal and the single sideband suppressed carrier AM signal are each produced by a process of *linear modulation.* If either of these signals is multiplied by a sine-wave carrier with frequency f_c Hz and with the same phase as that of the signal carrier (whether present or suppressed), and if the resultant signal is passed through a filter which removes all frequency components above f_m Hz, without affecting the frequency components below f_m Hz, then the resultant signal is the original baseband signal, with perhaps a change in level. This is a process of *linear demodulation.*

2.4 Signal alphabet and information content

The signal alphabet of a digital signal is the number of *different* digits or element values which may be transmitted. Thus, when there are m different