Basics of communications and coding

William G. Chambers

Basics of communications and coding

William G. Chambers

Department of Mathematics,
Westfield College,
University of London

CLARENDON PRESS · OXFORD
1985

Oxford University Press, Walton Street, Oxford OX2 6DP
London New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Beirut Berlin Ibadan Mexico City Nicosia

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
Chambers, William G.
Basics of communications and coding.
1. Digital communications 2. Coding theory
I. Title
621.38'0413 TK5103
ISBN 0-19-853195-8
ISBN 0-19-853194-X Pbk

Library of Congress Cataloging in Publication Data
Chambers, William G.
Basics of communications and coding.
Bibliography: p.
Includes index.
1. Digital communications. 2. Coding theory.
3. Statistical communication theory. I. Title.
TK5103.7.C53 1985 621.38'0413 84-23006
ISBN 0-19-853195-8
ISBN 0-19-853194-X (pbk).

Set by Joshua Associates, Oxford
Printed in Great Britain by
St. Edmundsbury Press Ltd,
Bury St. Edmunds, Suffolk.

Preface

This book grew out of a course taught at the University of London, intended for third-year and graduate students of electronics, computer science, mathematics, and physics. It thus attempts to reach a wider audience than a traditional text on communications. None the less it covers what is normally regarded as the basic material of traditional communications theory, although not always in such detail as a typical engineering text.

In keeping with modern trends the emphasis is totally on digital communications. Analogue communication systems such as FM radio are not discussed since (a) they are well-covered in the classical texts, (b) they are of decreasing importance in comparison with digital systems, (c) the concepts do not seem to be so interesting as those of digital communications, so that the limited space available is better used on the latter. None the less much of the modulation theory described can easily be adapted for dealing with analogue signalling.

The book starts in a fairly traditional manner, with chapters on Fourier theory, digital signalling methods, probability theory and noisy channels. However it then goes on to cover several topics not usually covered at this level. (a) There is a fairly thorough discussion of Shannon's revolutionary discovery that with suitable coding information transmission over a noisy channel at a given rate not set too high can be carried out with a probability of error as small as one pleases. This is first illustrated using a simple form of modulation known as pulse-position modulation. Shannon's theorem is then discussed as it applies to the band-limited Gaussian channel. (b) Error-correcting codes are introduced in the context of Gaussian channels as forms of modulation which lower the probabilities of error. (c) In an attempt to allay the idea that Shannon's work is rather remote from everyday life a family of powerful and useful error-correcting codes is introduced, capable of approaching the Shannon limit. (d) The advent of optical communications means that the limitation imposed by quantum theory on information rates is now an important topic. This topic is discussed from an elementary point of view, but without sacrificing accuracy of presentation. (e) Cipher systems are becoming increasingly important in communications systems, because of the risk of interception and tampering to which signals sent by radio are prone. Moreover there have been some exciting theoretical developments in this field in the last few years.

The topics have been selected so that the mathematics is both conceptually straightforward and also useful in other contexts. Three chapters cover the mathematical background. They have been spaced out so that applications of the techniques follow the introduction of the techniques as soon as possible. Of necessity the treatments are brief, so that some familiarity is helpful, but they are also intended to be as far as possible complete for the purpose in hand.

This is to minimize the number of gaps in the main lines of development needing to be filled by appeal to external authority or by the use of material in other texts. Of the mathematical topics, most third-year students will have some familiarity with Fourier transforms and probability theory. The third topic, the theory of finite fields, is not usually found in elementary texts on communications theory. However it seems that it is becoming as much part of the repertoire of communications theory as the other topics. Although coding has been introduced fairly early on as a form of digital modulation and as a natural extension of modulation techniques like signal-constellations, none the less the use of algebraic structures in coding has been left until later. Algebraic codes are then treated in reasonable detail. Fortunately the algebra used here closely mimics ordinary polynomial algebra, and so the reader is never very far from what is familiar to him.

It may be worth drawing the reader's attention to the index. A lot of trouble has gone into preparing it. If a technical term or acronym is used without definition or cross-reference, then the index should indicate where it is first used.

Acknowledgements

I would like to express my gratitude above all to Paddy Farrell for his advice and encouragement, and also to many friends, colleagues, and visitors to Westfield College for their help in the preparation of this work. In particular I would like to mention Fred Piper, Thomas Beth, Dwijendra Ray-Chaudhuri, Gary Gledhill, Terry Parker, and Peter Williams. I would also like to thank the students whose comments on the presentation have proved very helpful, and the staff of the Westfield computer unit for their assistance in the preparation of the manuscript and of the diagrams.

London
May 1984

W. G. C.

Glossary

$A(f)$	two-sided noise spectral density, p. 79
$a(t)$	autocorrelation function of noise, p. 79
B	bandwidth equivalent, p. 95
C	channel capacity (Ch. 5), (5.42) on p. 113
C'	scaled channel capacity, $= C\ln 2$, (5.43) on p. 113
$\mathrm{cov}(x)$	covariance of random variable x, (4.5) on p. 72
d	Hamming distance (Ch. 6), p. 133
E	energy of signal, p. 88
E_b	energy per bit of information, (6.14) on p. 135
$\mathrm{erfc}(x)$	error-function complement, (4.18) on p. 76
$\mathrm{erf}(x)$	error function, p. 228
f	frequency, p. 6
GHz	frequency unit, $= 10^9$ Hz, p. 7
h	Planck's constant, p. 187
$H(f)$	filter transfer function, p. 22
$h(t)$	filter impulse-response function, p. 21
h_{ij}	parity-check matrix (Chs. 6, 7), (6.18) on p. 137
I	information, (3.1) on p. 40
J	information rate, p. 88
J'	scaled information rate, $= J\ln 2$, (4.56) on p. 88
kHz	frequency unit, $= 10^3$ Hz, p. 6
k_B	Boltzmann's constant, p. 10
MHz	frequency unit, $= 10^6$ Hz, p. 6
mod	mod symbol, p. 123, p. 162
$\bar{n}$	photons per normal mode (Ch. 8), p. 200
P	signal power, p. 87
$P(A)$	probability of event A, p. 70
$P(A\|B)$	probability of event A given event B, p. 71
P_b	bit-error probability, p. 143
P_e	error probability, (4.52) on p. 87
P_W	word-error probability, p. 142
R	code-rate, $= k/n$, (6.12) on p. 135
$R(t)$	raised-cosine function, (2.6) on p. 15
$\mathrm{sinc}(x)$	sinc function, $= \sin(\pi x)/(\pi x)$, (2.5) on p. 14

t	number of errors (Chs. 6, 7), p. 135
T	period, p. 26, or long interval of time, p. 45 n.
T_b	duration of pulse or bit, p. 45 n.
T_W	duration of codeword, p. 45 n.
THz	frequency unit, $= 10^{12}$ Hz, p. 7
var(x)	variance of random variable x, (4.4) on p. 72
W	bandwidth, p. 26
w	Hamming weight (Ch. 6), p. 134
z	polynomial argument (Ch. 7), p. 160
$z(t)$	positive-frequency representation, (3.17) on p. 61
α	field element obeying $\alpha^4 = \alpha + 1$ (Ch. 7 only), p. 156
α	subscript denoting member of ensemble (Ch. 4), p. 72
γ_b	$= E_b/\eta$, (6.13) on p. 135
$\delta(t)$	delta function, p. 19
δ_{kl}	Kronecker delta symbol, (5.6) on p. 96
η	(η_{NE} in Ch. 8) white-noise density, (4.36) on p. 80
η_Q	(Ch. 8 only) quantum efficiency, p. 189
θ_N	noise-temperature p. 82
$\phi(t)$	phasor, (3.21) on p. 62
$\Pi(t)$	rectangular-pulse function, (2.3) on p. 14
$\rho(x')$	probability density, p. 73
$\rho(x', y')$	joint probability density, p. 73
$\equiv$	congruence symbol, p. 162
$*$	convolution symbol, p. 18
(n, k) code	p. 123
$\binom{n}{s}$	combinatorial factor, p. 71
$\langle x \rangle, \bar{x}$	mean of random variable x, p. 72
Ш(t)	sha function, (2.54) on p. 35

Contents

1. Introduction

1.1 Aims and outline

1.1.1 *Introductory comments*

This book is planned around a theme, the theme of reliable digital communications. This theme is used as a guide for selecting topics from communications theory. It is not rigidly pursued however, and side-tracks of interest are often investigated as well. In consequence the text covers both the basics of digital communications and the basics of the theory of error-correcting codes as a single topic. It provides an introduction to coding theory in the context of traditional communications theory. This has been done by introducing coding as a form of digital modulation. In consequence it will be found that this book, although it starts in the manner of most textbooks on communications theory, brings in considerations of reliability and coding at a fairly early stage. The words 'communications', 'digital', and 'reliable' are examined further in the next three subsections.

The second part of the title refers to error-correcting codes. These are becoming more widely used in communications for at least the following reasons:

(a) The electronic hardware for carrying out the necessary computations for encoding and decoding is becoming cheaper and more reliable, so that what a few years ago was not economically feasible is easily carried out now.

(b) Some applications, such as the sending of data by *cipher*, need a level of reliability higher than usual. Ciphers are designed to detect tampering with the data-stream and so even a slight error may well leave a block of data indecipherable.

(c) The alternative of using error-checking, followed if necessary by a request for retransmission, may be clumsy where the round-trip time is rather long, as on a satellite link, or it may be inconvenient or even impossible, as with a recording or a broadcast sent to several receivers.

(d) Modern developments such as fibre-optic cables provide channels which can bear the extra load imposed by error-correcting codes. These codes always increase the amount of data to be sent since they work by introducing redundant checks by means of which errors are not only detected but also corrected.

1.1.2 *Communications*

The word 'communications' will refer to the point-to-point transmission of information, usually by electrical means, either by wire, coaxial cable, waveguide, or by radio. The overall system consists of several parts:

(a) The first component is the source, which provides the information to be sent. There are basically two types of information-signal, analogue and digital. Analogue signals are in the form of a continuously varying function of time. Typical sources of such signals are microphones and television cameras, which translate sound- and light-intensities respectively into electrical signals. Digital sources may be computer files or impulse streams used for driving remote teletypes or visual display units. Here the information is *coded* into a stream of pulses, whose heights can take only a discrete set of values, usually just two.

(b) The output from the source is then processed by a modulator in preparation for sending it over the channel.

(c) The channel is the link to the remote location where the information is to be sent. It may be a wired electrical connection, such as a telephone or telex line, or a radio link, where the transmitter radiates electromagnetic waves into space which are picked up by a receiver at the remote location. Other examples of channels are information-storage systems such as magnetic discs and tapes, which can be regarded as linking the sender who puts in a message at one time to the receiver who retrieves it, perhaps months later. The channel is the part of the communication system not under the control of the sender or receiver, and it tends to degrade the signal. Two main forms of such degradation will be considered. The channel can distort the signal in a predetermined way, which can be compensated for. It can also introduce interference, or *noise*, into the received signal in a manner which is unpredictable except in a statistical sense.

(d) The channel delivers the signal to the receiver. The degradation can cause errors at the receiver, and quite sophisticated means may be needed to combat its effects.

(e) After any necessary processing the signal is passed on from the receiver to its destination, the *information sink*. What happens to it after that is not the concern of this book.

1.1.3 *The robustness of digital communications*

Digital signalling has an innate capacity for countering interference which gives it a quality of robustness or strength. We illustrate this by contrasting digital and analogue signalling.

Suppose that a remotely situated instrument has to send back the values of some physical variable such as wind-speed or temperature to a central station along a signal-wire. A very simple method is to put a voltage on the wire proportional to the instrument reading. The voltage is then an analogue of the physical quantity being measured and so this system of sending the results is called *analogue* transmission. The possible values of the physical quantity and hence the corresponding voltage form a continuous range. Hence if there is any interference on the wire which changes the voltage as measured at the receiving end there is no simple way of knowing that the received value is false.

Now suppose the transmitter is designed to send voltages from a discrete set of values, say, 0, 1, . . . , 9. If the interference on the wire is very unlikely to cause the output to deviate by more than 0.5 V then the receiver can tell quite reliably what actually was sent. Thus if 7.21 V is received then it is almost certain that 7 V was sent. Only very occasionally when the interference happens to be too large is the receiver fooled. A transmission system, in which the possible signals fed into the channel form a discrete set, is said to use ***digital signalling***. As noted above, this discreteness provides a capacity for correcting the effects of interference and hence can be used to make the signalling very reliable. This advantage of digital signalling over analogue signalling may not be very great in a simple system, but if the signal has to go through a lot of complicated processing then the tendency of analogue signals to 'drift' makes them difficult to deal with. In contrast digital signals can be corrected or re-formed at each stage.

There are two main disadvantages of digital signalling. The first is that the electronics may be rather complicated. This is no longer necessarily a serious disadvantage, since developments in electronics have made available hardware of incredible complexity combined with remarkable cheapness and reliability against breakdown. The second disadvantage is that digital signalling makes greater demands on the channel. (In particular it needs a greater bandwidth.) Thus suppose that we wish the remote instrument to send its readings accurate to three figures. Sending a reading like 7.52 is no harder than sending a single integer for an analogue system, but the digital system just described has to send three distinct digits, 7, 5, and 2, either in succession (serially) along the wire, or in parallel along three wires. Again, developments in the techniques of communication have reduced the seriousness of this problem.

The points made in this section can be illustrated by considering the trend towards digital sound-recording (Philips Technical Review 1982). In the old system the sound-signal was represented in an analogue manner by the continuously variable magnetization of a magnetic tape or by the lateral displacement of the groove on a record. Digital recording involves sampling the signal very frequently (at 44.1 thousand times a second), and representing the sampled values digitally to a sufficiently high precision. (Sixteen binary digits are used rather than the three decimal digits in the example above.) The ensuing stream of digits is then written on to a medium such as a video disc or video tape, originally developed for analogue television signals, and which in consequence can provide the extra resources (in bandwidth) needed. This technique can obviate not only the hiss and crackle caused by imperfections in the record, but also the effects of unsteadiness in the turntable speed, and frequency and intermodulation distortion in the recording process.

1.1.4 *Reliable communications*

If the reliability of a signalling system is too low, then one way of getting a message through is to send it several times, so that the receiver can piece it together from the various garbled versions received. Such a strategy evidently slows up the transmission rate, the rate at which messages are getting through, and it is easy to see that the achievement of very high reliability involves a great number of repetitions. It seems reasonable to expect that, in general, arbitrarily high reliability can only be achieved at the cost of an excessively low transmission rate. However, Shannon (1948) proved that this is not so, and that arbitrarily high reliability can be achieved while keeping the transmission rate up at a finite value. His work spurred on the development of both information theory and of coding theory. Information theory provides precise definitions of the amount of information and of information rates. Shannon proved that arbitrarily high reliability can be achieved if and only if the rate of transmission of information is below a rate called the *channel capacity*. This reliability is achieved by enhancing the robust quality of digital signalling just described. In practice error-correcting codes are used and the attainment of very high reliabilities is part of the province of coding theory. Unfortunately, Shannon's proof failed to demonstrate a practical coding scheme, and in consequence it was believed for a while that any code achieving the levels of reliability promised by Shannon would be impossibly complicated and totally impractical to use, employing huge lookup tables whose entries might well exceed the total number of atoms in the earth. However, as we shall see, families of codes have been found which can achieve arbitrarily high reliability without excessive complexity. In consequence, some of these codes are useful in practice; they are not too hard to decode, and yet they can attain reliabilities as high as one may reasonably demand in this uncertain world.

1.1.5 *Outline*

The rest of this chapter will be used to fill in background material.

Chapter 2 contains a brief summary of Fourier theory, one of the mathematical tools needed in communications theory. One may remind oneself of how basic it is by noting that the concepts of frequency and of bandwidth in radio signalling are derived from Fourier theory.

Chapter 3 discusses ways in which digital signalling is carried out. It draws heavily on the Fourier theory of Chapter 2 in the discussion of bandwidths of pulsed signalling. Two examples of the robustness of digital signalling are given. In the first example the probability of error is considered when a signal has to be relayed by a long line of repeaters, and in the second it is shown how adaptive equalizers can be made to track frequency and phase errors in the transmission. Another important topic is how analogue signals can be faithfully sampled for digital transmission or recording. It seems reasonable to expect that sampling would cause significant degradation, but in fact this need not be so. The topics of information theory and data-compression are touched on, but only superficially since that is sufficient for the discussions at a later stage.