

Springer Undergraduate Mathematics Series

Advisory Board

M.A.J. Chaplain *University of Dundee*

K. Erdmann *University of Oxford*

A. MacIntyre *Queen Mary, University of London*

L.C.G. Rogers *University of Cambridge*

E. Süli *University of Oxford*

J.F. Toland *University of Bath*

For other titles published in this series, go to
www.springer.com/series/3423

Norman L. Biggs

Codes: An Introduction to Information Communication and Cryptography

Norman L. Biggs
Department of Mathematics
London School of Economics
Houghton Street
London WC2A 2AE, UK

Maple is a trademark of Waterloo Maple Inc.

Springer Undergraduate Mathematics Series ISSN 1615-2085
ISBN: 978-1-84800-272-2 e-ISBN: 978-1-84800-273-9
DOI: 10.1007/978-1-84800-273-9

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008930146

Mathematics Subject Classification (2000): 94A, 94B, 11T71

© Springer-Verlag London Limited 2008

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Preface

Many people do not realise that mathematics provides the foundation for the devices we use to handle information in the modern world. Most of those who do know probably think that the parts of mathematics involved are quite ‘classical’, such as Fourier analysis and differential equations. In fact, a great deal of the mathematical background is part of what used to be called ‘pure’ mathematics, indicating that it was created in order to deal with problems that originated within mathematics itself. It has taken many years for mathematicians to come to terms with this situation, and some of them are still not entirely happy about it.

This book is an integrated introduction to Coding. By this I mean replacing symbolic information, such as a sequence of bits or a message written in a natural language, by another message using (possibly) different symbols. There are three main reasons for doing this: Economy (data compression), Reliability (correction of errors), and Security (cryptography).

I have tried to cover each of these three areas in sufficient depth so that the reader can grasp the basic problems and go on to more advanced study. The mathematical theory is introduced in a way that enables the basic problems to be stated carefully, but without unnecessary abstraction. The prerequisites (sets and functions, matrices, finite probability) should be familiar to anyone who has taken a standard course in mathematical methods or discrete mathematics. A course in elementary abstract algebra and/or number theory would be helpful, but the book contains the essential facts, and readers without this background should be able to understand what is going on.

There are a few places where reference is made to computer algebra systems. I have tried to avoid making this a prerequisite, but students who have access to such a system will find it helpful. In particular, there are occasional specific references to MAPLETM (release 10), by Maplesoft, a division of Waterloo Maple Inc., Waterloo, Canada.

The book has been developed from a course of twenty lectures on Information, Communication, and Cryptography given for the MSc in Applicable Mathematics at the London School of Economics. I should like to thank all those students who have contributed to the development of the course materials, in particular those who have written dissertations in this area: Rajni Kanda, Ovijit Paul, Arunduti Dutta-Roy, Ana de Corbavia-Perisic, Raminder Ruprai, James Rees, Elisabeth Biell, Anisa Bhatt, Timothy Morill, Shivam Kumar, and Carey Chua. I owe a special debt to Raminder Ruprai, who worked through all the exercises and helped to sort out many mistakes and obscurities.

Finally, I am grateful to Aaron Wilson, who helped to produce the diagrams, and especially to Karen Borthwick, who has been very helpful and supportive on behalf of the publishers.

Norman Biggs
January 2008

Contents

Preface	v
1. Coding and its uses	1
1.1 Messages	1
1.2 Coding	3
1.3 Basic definitions	4
1.4 Coding for economy	7
1.5 Coding for reliability	8
1.6 Coding for security	9
2. Prefix-free codes	13
2.1 The decoding problem	13
2.2 Representing codes by trees	16
2.3 The Kraft-McMillan number	18
2.4 Unique decodability implies $K \leq 1$	21
2.5 Proof of the Counting Principle	24
3. Economical coding	27
3.1 The concept of a source	27
3.2 The optimization problem	30
3.3 Entropy	32
3.4 Entropy, uncertainty, and information	34
3.5 Optimal codes – the fundamental theorems	38
3.6 Huffman’s rule	40
3.7 Optimality of Huffman codes	44

4. Data compression	47
4.1 Coding in blocks	47
4.2 Distributions on product sets	49
4.3 Stationary sources	52
4.4 Coding a stationary source	55
4.5 Algorithms for data compression	58
4.6 Using numbers as codewords	59
4.7 Arithmetic coding	62
4.8 The properties of arithmetic coding	65
4.9 Coding with a dynamic dictionary	67
5. Noisy channels	73
5.1 The definition of a channel	73
5.2 Transmitting a source through a channel	76
5.3 Conditional entropy	78
5.4 The capacity of a channel	81
5.5 Calculating the capacity of a channel	83
6. The problem of reliable communication	89
6.1 Communication using a noisy channel	89
6.2 The extended BSC	94
6.3 Decision rules	96
6.4 Error correction	100
6.5 The packing bound	102
7. The noisy coding theorems	107
7.1 The probability of a mistake	107
7.2 Coding at a given rate	111
7.3 Transmission using the extended BSC	113
7.4 The rate should not exceed the capacity	117
7.5 Shannon's theorem	119
7.6 Proof of Fano's inequality	120
8. Linear codes	123
8.1 Introduction to linear codes	123
8.2 Construction of linear codes using matrices	126
8.3 The check matrix of a linear code	128
8.4 Constructing 1-error-correcting codes	131
8.5 The decoding problem	135

9. Algebraic coding theory	141
9.1 Hamming codes	141
9.2 Cyclic codes	145
9.3 Classification and properties of cyclic codes	149
9.4 Codes that can correct more than one error	153
9.5 Definition of a family of BCH codes	155
9.6 Properties of the BCH codes	158
10. Coding natural languages	163
10.1 Natural languages as sources	163
10.2 The uncertainty of english	165
10.3 Redundancy and meaning	168
10.4 Introduction to cryptography	170
10.5 Frequency analysis	174
11. The development of cryptography	179
11.1 Symmetric key cryptosystems	179
11.2 Poly-alphabetic encryption	180
11.3 The Playfair system	183
11.4 Mathematical algorithms in cryptography	185
11.5 Methods of attack	187
12. Cryptography in theory and practice	191
12.1 Encryption in terms of a channel	191
12.2 Perfect secrecy	195
12.3 The one-time pad	197
12.4 Iterative methods	198
12.5 Encryption standards	201
12.6 The key distribution problem	203
13. The RSA cryptosystem	207
13.1 A new approach to cryptography	207
13.2 Outline of the RSA system	209
13.3 Feasibility of RSA	212
13.4 Correctness of RSA	215
13.5 Confidentiality of RSA	217
14. Cryptography and calculation	221
14.1 The scope of cryptography	221
14.2 Hashing	222
14.3 Calculations in the field \mathbb{F}_p	224
14.4 The discrete logarithm	226

14.5 The ElGamal cryptosystem	228
14.6 The Diffie-Hellman key distribution system	230
14.7 Signature schemes	232
15. Elliptic curve cryptography	237
15.1 Calculations in finite groups	237
15.2 The general ElGamal cryptosystem	239
15.3 Elliptic curves.....	241
15.4 The group of an elliptic curve	245
15.5 Improving the efficiency of exponentiation	248
15.6 A final word	250
Answers to odd-numbered exercises	255
Index	271