

Principles of Digital Communication

Robert G. Gallager

January 5, 2008

Preface: introduction and objectives

The digital communication industry is an enormous and rapidly growing industry, roughly comparable in size to the computer industry. The objective of this text is to study those aspects of digital communication systems that are unique to those systems. That is, rather than focusing on hardware and software for these systems, which is much like hardware and software for many other kinds of systems, we focus on the fundamental system aspects of modern digital communication.

Digital communication is a field in which theoretical ideas have had an unusually powerful impact on system design and practice. The basis of the theory was developed in 1948 by Claude Shannon, and is called information theory. For the first 25 years or so of its existence, information theory served as a rich source of academic research problems and as a tantalizing suggestion that communication systems could be made more efficient and more reliable by using these approaches. Other than small experiments and a few highly specialized military systems, the theory had little interaction with practice. By the mid 1970's, however, mainstream systems using information theoretic ideas began to be widely implemented. The first reason for this was the increasing number of engineers who understood both information theory and communication system practice. The second reason was that the low cost and increasing processing power of digital hardware made it possible to implement the sophisticated algorithms suggested by information theory. The third reason was that the increasing complexity of communication systems required the architectural principles of information theory.

The theoretical principles here fall roughly into two categories - the first provide analytical tools for determining the performance of particular systems, and the second put fundamental limits on the performance of any system. Much of the first category can be understood by engineering undergraduates, while the second category is distinctly graduate in nature. It is not that graduate students know so much more than undergraduates, but rather that undergraduate engineering students are trained to master enormous amounts of detail and to master the equations that deal with that detail. They are not used to the patience and deep thinking required to understand abstract performance limits. This patience comes later with thesis research.

My original purpose was to write an undergraduate text on digital communication, but experience teaching this material over a number of years convinced me that I could not write an honest exposition of principles, including both what is possible and what is not possible, without losing most undergraduates. There are many excellent undergraduate texts on digital communication describing a wide variety of systems, and I didn't see the need for another. Thus this text is now aimed at graduate students, but accessible to patient undergraduates.

The relationship between theory, problem sets, and engineering/design in an academic subject is rather complex. The theory deals with relationships and analysis for *models* of real systems. A good theory (and information theory is one of the best) allows for simple analysis of simplified models. It also provides structural principles that allow insights from these simple models to be applied to more complex and realistic models. Problem sets provide students with an opportunity to analyze these highly simplified models, and, with patience, to start to understand the general principles. Engineering deals with making the approximations and judgment calls to create simple models that focus on the critical elements of a situation, and from there to design workable systems.

The important point here is that engineering (at this level) cannot really be separated from theory. Engineering is necessary to choose appropriate theoretical models, and theory is necessary

to find the general properties of those models. To oversimplify it, engineering determines what the reality is and theory determines the consequences and structure of that reality. At a deeper level, however, the engineering perception of reality heavily depends on the perceived structure (all of us carry oversimplified models around in our heads). Similarly, the structures created by theory depend on engineering common sense to focus on important issues. Engineering sometimes becomes overly concerned with detail, and theory overly concerned with mathematical niceties, but we shall try to avoid both these excesses here.

Each topic in the text is introduced with highly oversimplified toy models. The results about these toy models are then related to actual communication systems and this is used to generalize the models. We then iterate back and forth between analysis of models and creation of models. Understanding the performance limits on classes of models is essential in this process.

There are many exercises designed to help understand each topic. Some give examples showing how an analysis breaks down if the restrictions are violated. Since analysis always treats models rather than reality, these examples build insight into how the results about models apply to real systems. Other exercises apply the text results to very simple cases and others generalize the results to more complex systems. Yet others explore the sense in which theoretical models apply to particular practical problems.

It is important to understand that the purpose of the exercises is not so much to get the ‘answer’ as to acquire understanding. Thus students using this text will learn much more if they discuss the exercises with others and think about what they have learned after completing the exercise. The point is not to manipulate equations (which computers can now do better than students) but rather to understand the equations (which computers can not do).

As pointed out above, the material here is primarily graduate in terms of abstraction and patience, but requires only a knowledge of elementary probability, linear systems, and simple mathematical abstraction, so it can be understood at the undergraduate level. For both undergraduates and graduates, I feel strongly that learning to reason about engineering material is more important, both in the workplace and in further education, than learning to pattern match and manipulate equations.

Most undergraduate communication texts aim at familiarity with a large variety of different systems that have been implemented historically. This is certainly valuable in the workplace, at least for the near term, and provides a rich set of examples that are valuable for further study. The digital communication field is so vast, however, that learning from examples is limited, and in the long term it is necessary to learn the underlying principles. The examples from undergraduate courses provide a useful background for studying these principles, but the ability to reason abstractly that comes from elementary pure mathematics courses is equally valuable.

Most graduate communication texts focus more on the analysis of problems with less focus on the modeling, approximation, and insight needed to see how these problems arise. Our objective here is to use simple models and approximations as a way to understand the general principles. We will use quite a bit of mathematics in the process, but the mathematics will be used to establish general results precisely rather than to carry out detailed analyses of special cases.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction to digital communication | 1 |
| 1.1 | Standardized interfaces and layering | 3 |
| 1.2 | Communication sources | 5 |
| 1.2.1 | Source coding | 6 |
| 1.3 | Communication channels | 7 |
| 1.3.1 | Channel encoding (modulation) | 9 |
| 1.3.2 | Error correction | 10 |
| 1.4 | Digital interface | 11 |
| 1.4.1 | Network aspects of the digital interface | 12 |
| 1.5 | Supplementary reading | 14 |
| 2 | Coding for Discrete Sources | 15 |
| 2.1 | Introduction | 15 |
| 2.2 | Fixed-length codes for discrete sources | 16 |
| 2.3 | Variable-length codes for discrete sources | 18 |
| 2.3.1 | Unique decodability | 19 |
| 2.3.2 | Prefix-free codes for discrete sources | 20 |
| 2.3.3 | The Kraft inequality for prefix-free codes | 22 |
| 2.4 | Probability models for discrete sources | 24 |
| 2.4.1 | Discrete memoryless sources | 25 |
| 2.5 | Minimum \bar{L} for prefix-free codes | 26 |
| 2.5.1 | Lagrange multiplier solution for the minimum \bar{L} | 27 |
| 2.5.2 | Entropy bounds on \bar{L} | 28 |
| 2.5.3 | Huffman's algorithm for optimal source codes | 29 |
| 2.6 | Entropy and fixed-to-variable-length codes | 33 |
| 2.6.1 | Fixed-to-variable-length codes | 35 |
| 2.7 | The AEP and the source coding theorems | 36 |
| 2.7.1 | The weak law of large numbers | 37 |

| | | |
|----------|---|-----------|
| 2.7.2 | The asymptotic equipartition property | 38 |
| 2.7.3 | Source coding theorems | 41 |
| 2.7.4 | The entropy bound for general classes of codes | 42 |
| 2.8 | Markov sources | 43 |
| 2.8.1 | Coding for Markov sources | 45 |
| 2.8.2 | Conditional entropy | 45 |
| 2.9 | Lempel-Ziv universal data compression | 47 |
| 2.9.1 | The LZ77 algorithm | 48 |
| 2.9.2 | Why LZ77 works | 49 |
| 2.9.3 | Discussion | 50 |
| 2.10 | Summary of discrete source coding | 51 |
| 2.E | Exercises | 53 |
| 3 | Quantization | 63 |
| 3.1 | Introduction to quantization | 63 |
| 3.2 | Scalar quantization | 65 |
| 3.2.1 | Choice of intervals for given representation points | 65 |
| 3.2.2 | Choice of representation points for given intervals | 66 |
| 3.2.3 | The Lloyd-Max algorithm | 66 |
| 3.3 | Vector quantization | 68 |
| 3.4 | Entropy-coded quantization | 69 |
| 3.5 | High-rate entropy-coded quantization | 70 |
| 3.6 | Differential entropy | 71 |
| 3.7 | Performance of uniform high-rate scalar quantizers | 73 |
| 3.8 | High-rate two-dimensional quantizers | 76 |
| 3.9 | Summary of quantization | 78 |
| 3A | Appendix A: Nonuniform scalar quantizers | 79 |
| 3B | Appendix B: Nonuniform 2D quantizers | 81 |
| 3.E | Exercises | 83 |
| 4 | Source and channel waveforms | 87 |
| 4.1 | Introduction | 87 |
| 4.1.1 | Analog sources | 87 |
| 4.1.2 | Communication channels | 89 |
| 4.2 | Fourier series | 90 |
| 4.2.1 | Finite-energy waveforms | 92 |
| 4.3 | \mathcal{L}_2 functions and Lebesgue integration over $[-T/2, T/2]$ | 94 |

| | | |
|----------|--|------------|
| 4.3.1 | Lebesgue measure for a union of intervals | 95 |
| 4.3.2 | Measure for more general sets | 96 |
| 4.3.3 | Measurable functions and integration over $[-T/2, T/2]$ | 98 |
| 4.3.4 | Measurability of functions defined by other functions | 100 |
| 4.3.5 | \mathcal{L}_1 and \mathcal{L}_2 functions over $[-T/2, T/2]$ | 101 |
| 4.4 | The Fourier series for \mathcal{L}_2 waveforms | 102 |
| 4.4.1 | The T -spaced truncated sinusoid expansion | 103 |
| 4.5 | Fourier transforms and \mathcal{L}_2 waveforms | 105 |
| 4.5.1 | Measure and integration over \mathbb{R} | 107 |
| 4.5.2 | Fourier transforms of \mathcal{L}_2 functions | 109 |
| 4.6 | The DTFT and the sampling theorem | 111 |
| 4.6.1 | The discrete-time Fourier transform | 112 |
| 4.6.2 | The sampling theorem | 112 |
| 4.6.3 | Source coding using sampled waveforms | 115 |
| 4.6.4 | The sampling theorem for $[\Delta - W, \Delta + W]$ | 116 |
| 4.7 | Aliasing and the sinc-weighted sinusoid expansion | 117 |
| 4.7.1 | The T -spaced sinc-weighted sinusoid expansion | 117 |
| 4.7.2 | Degrees of freedom | 118 |
| 4.7.3 | Aliasing — a time domain approach | 119 |
| 4.7.4 | Aliasing — a frequency domain approach | 120 |
| 4.8 | Summary | 121 |
| 4A | Appendix: Supplementary material and proofs | 122 |
| 4A.1 | Countable sets | 122 |
| 4A.2 | Finite unions of intervals over $[-T/2, T/2]$ | 125 |
| 4A.3 | Countable unions and outer measure over $[-T/2, T/2]$ | 125 |
| 4A.4 | Arbitrary measurable sets over $[-T/2, T/2]$ | 128 |
| 4.E | Exercises | 132 |
| 5 | Vector spaces and signal space | 141 |
| 5.1 | The axioms and basic properties of vector spaces | 142 |
| 5.1.1 | Finite-dimensional vector spaces | 144 |
| 5.2 | Inner product spaces | 145 |
| 5.2.1 | The inner product spaces \mathbb{R}^n and \mathbb{C}^n | 146 |
| 5.2.2 | One-dimensional projections | 146 |
| 5.2.3 | The inner product space of \mathcal{L}_2 functions | 148 |
| 5.2.4 | Subspaces of inner product spaces | 149 |
| 5.3 | Orthonormal bases and the projection theorem | 150 |

| | | |
|----------|--|------------|
| 5.3.1 | Finite-dimensional projections | 151 |
| 5.3.2 | Corollaries of the projection theorem | 151 |
| 5.3.3 | Gram-Schmidt orthonormalization | 153 |
| 5.3.4 | Orthonormal expansions in \mathcal{L}_2 | 153 |
| 5.4 | Summary | 156 |
| 5A | Appendix: Supplementary material and proofs | 156 |
| 5A.1 | The Plancherel theorem | 156 |
| 5A.2 | The sampling and aliasing theorems | 160 |
| 5A.3 | Prolate spheroidal waveforms | 162 |
| 5.E | Exercises | 164 |
| 6 | Channels, modulation, and demodulation | 167 |
| 6.1 | Introduction | 167 |
| 6.2 | Pulse amplitude modulation (PAM) | 169 |
| 6.2.1 | Signal constellations | 170 |
| 6.2.2 | Channel imperfections: a preliminary view | 171 |
| 6.2.3 | Choice of the modulation pulse | 173 |
| 6.2.4 | PAM demodulation | 174 |
| 6.3 | The Nyquist criterion | 175 |
| 6.3.1 | Band-edge symmetry | 176 |
| 6.3.2 | Choosing $\{p(t-kT); k \in \mathbb{Z}\}$ as an orthonormal set | 178 |
| 6.3.3 | Relation between PAM and analog source coding | 179 |
| 6.4 | Modulation: baseband to passband and back | 180 |
| 6.4.1 | Double-sideband amplitude modulation | 180 |
| 6.5 | Quadrature amplitude modulation (QAM) | 181 |
| 6.5.1 | QAM signal set | 182 |
| 6.5.2 | QAM baseband modulation and demodulation | 183 |
| 6.5.3 | QAM: baseband to passband and back | 184 |
| 6.5.4 | Implementation of QAM | 185 |
| 6.6 | Signal space and degrees of freedom | 187 |
| 6.6.1 | Distance and orthogonality | 188 |
| 6.7 | Carrier and phase recovery in QAM systems | 190 |
| 6.7.1 | Tracking phase in the presence of noise | 191 |
| 6.7.2 | Large phase errors | 191 |
| 6.8 | Summary of modulation and demodulation | 192 |
| 6.E | Exercises | 193 |

| | | |
|----------|--|------------|
| 7 | Random processes and noise | 199 |
| 7.1 | Introduction | 199 |
| 7.2 | Random processes | 200 |
| 7.2.1 | Examples of random processes | 201 |
| 7.2.2 | The mean and covariance of a random process | 202 |
| 7.2.3 | Additive noise channels | 203 |
| 7.3 | Gaussian random variables, vectors, and processes | 204 |
| 7.3.1 | The covariance matrix of a jointly-Gaussian random vector | 206 |
| 7.3.2 | The probability density of a jointly-Gaussian random vector | 207 |
| 7.3.3 | Special case of a 2-dimensional zero-mean Gaussian random vector | 209 |
| 7.3.4 | $\mathbf{Z} = \mathbf{A}\mathbf{W}$ where \mathbf{A} is orthogonal | 210 |
| 7.3.5 | Probability density for Gaussian vectors in terms of principal axes | 210 |
| 7.3.6 | Fourier transforms for joint densities | 211 |
| 7.4 | Linear functionals and filters for random processes | 212 |
| 7.4.1 | Gaussian processes defined over orthonormal expansions | 214 |
| 7.4.2 | Linear filtering of Gaussian processes | 215 |
| 7.4.3 | Covariance for linear functionals and filters | 215 |
| 7.5 | Stationarity and related concepts | 216 |
| 7.5.1 | Wide-sense stationary (WSS) random processes | 217 |
| 7.5.2 | Effectively stationary and effectively WSS random processes | 219 |
| 7.5.3 | Linear functionals for effectively WSS random processes | 220 |
| 7.5.4 | Linear filters for effectively WSS random processes | 220 |
| 7.6 | Stationary and WSS processes in the Frequency Domain | 222 |
| 7.7 | White Gaussian noise | 224 |
| 7.7.1 | The sinc expansion as an approximation to WGN | 226 |
| 7.7.2 | Poisson process noise | 227 |
| 7.8 | Adding noise to modulated communication | 227 |
| 7.8.1 | Complex Gaussian random variables and vectors | 229 |
| 7.9 | Signal to noise ratio | 233 |
| 7.10 | Summary of Random Processes | 235 |
| 7A | Appendix: Supplementary topics | 236 |
| 7A.1 | Properties of covariance matrices | 236 |
| 7A.2 | The Fourier series expansion of a truncated random process | 238 |
| 7A.3 | Uncorrelated coefficients in a Fourier series | 239 |
| 7A.4 | The Karhunen-Loeve expansion | 242 |
| 7.E | Exercises | 244 |

| | | |
|----------|---|------------|
| 8 | Detection, coding, and decoding | 249 |
| 8.1 | Introduction | 249 |
| 8.2 | Binary detection | 251 |
| 8.3 | Binary signals in white Gaussian noise | 254 |
| 8.3.1 | Detection for PAM antipodal signals | 254 |
| 8.3.2 | Detection for binary non-antipodal signals | 256 |
| 8.3.3 | Detection for binary real vectors in WGN | 257 |
| 8.3.4 | Detection for binary complex vectors in WGN | 260 |
| 8.3.5 | Detection of binary antipodal waveforms in WGN | 261 |
| 8.4 | M -ary detection and sequence detection | 264 |
| 8.4.1 | M -ary detection | 265 |
| 8.4.2 | Successive transmissions of QAM signals in WGN | 266 |
| 8.4.3 | Detection with arbitrary modulation schemes | 268 |
| 8.5 | Orthogonal signal sets and simple channel coding | 271 |
| 8.5.1 | Simplex signal sets | 271 |
| 8.5.2 | Bi-orthogonal signal sets | 272 |
| 8.5.3 | Error probability for orthogonal signal sets | 273 |
| 8.6 | Block Coding | 276 |
| 8.6.1 | Binary orthogonal codes and Hadamard matrices | 276 |
| 8.6.2 | Reed-Muller codes | 278 |
| 8.7 | The noisy-channel coding theorem | 280 |
| 8.7.1 | Discrete memoryless channels | 280 |
| 8.7.2 | Capacity | 282 |
| 8.7.3 | Converse to the noisy-channel coding theorem | 283 |
| 8.7.4 | noisy-channel coding theorem, forward part | 284 |
| 8.7.5 | The noisy-channel coding theorem for WGN | 287 |
| 8.8 | Convolutional codes | 288 |
| 8.8.1 | Decoding of convolutional codes | 290 |
| 8.8.2 | The Viterbi algorithm | 291 |
| 8.9 | Summary | 292 |
| 8A | Appendix: Neyman-Pearson threshold tests | 293 |
| 8.E | Exercises | 298 |
| 9 | Wireless digital communication | 305 |
| 9.1 | Introduction | 305 |
| 9.2 | Physical modeling for wireless channels | 308 |
| 9.2.1 | Free space, fixed transmitting and receiving antennas | 309 |

| | | |
|-------|--|-----|
| 9.2.2 | Free space, moving antenna | 311 |
| 9.2.3 | Moving antenna, reflecting wall | 311 |
| 9.2.4 | Reflection from a ground plane | 313 |
| 9.2.5 | Shadowing | 314 |
| 9.2.6 | Moving antenna, multiple reflectors | 314 |
| 9.3 | Input/output models of wireless channels | 315 |
| 9.3.1 | The system function and impulse response for LTV systems | 316 |
| 9.3.2 | Doppler spread and coherence time | 319 |
| 9.3.3 | Delay spread, and coherence frequency | 321 |
| 9.4 | Baseband system functions and impulse responses | 323 |
| 9.4.1 | A discrete-time baseband model | 325 |
| 9.5 | Statistical channel models | 328 |
| 9.5.1 | Passband and baseband noise | 330 |
| 9.6 | Data detection | 331 |
| 9.6.1 | Binary detection in flat Rayleigh fading | 332 |
| 9.6.2 | Non-coherent detection with known channel magnitude | 334 |
| 9.6.3 | Non-coherent detection in flat Rician fading | 336 |
| 9.7 | Channel measurement | 338 |
| 9.7.1 | The use of probing signals to estimate the channel | 339 |
| 9.7.2 | Rake receivers | 343 |
| 9.8 | Diversity | 346 |
| 9.9 | CDMA; The IS95 Standard | 349 |
| 9.9.1 | Voice compression | 350 |
| 9.9.2 | Channel coding and decoding | 351 |
| 9.9.3 | Viterbi decoding for fading channels | 352 |
| 9.9.4 | Modulation and demodulation | 353 |
| 9.9.5 | Multiaccess Interference in IS95 | 355 |
| 9.10 | Summary of Wireless Communication | 357 |
| 9A | Appendix: Error probability for non-coherent detection | 358 |
| 9.E | Exercises | 360 |