Chapter 1

Introduction

The advent of cheap high-speed global communications ranks as one of the most important developments of human civilization in the second half of the twentieth century.

In 1950, an international telephone call was a remarkable event, and black-and-white television was just beginning to become widely available. By 2000, in contrast, an intercontinental phone call could often cost less than a postcard, and downloading large files instantaneously from anywhere in the world had become routine. The effects of this revolution are felt in daily life from Boston to Berlin to Bangalore.

Underlying this development has been the replacement of analog by digital communications.

Before 1948, digital communications had hardly been imagined. Indeed, Shannon's 1948 paper [7] may have been the first to use the word "bit."[1]

Even as late as 1988, the authors of an important text on digital communications [5] could write in their first paragraph:

> Why would [voice and images] be transmitted digitally? Doesn't digital transmission squander bandwidth? Doesn't it require more expensive hardware? After all, a voice-band data modem (for digital transmission over a telephone channel) costs ten times as much as a telephone and (in today's technology) *is incapable of transmitting voice signals* with quality comparable to an ordinary telephone [authors' emphasis]. This sounds like a serious indictment of digital transmission for analog signals, but for most applications, the advantages outweigh the disadvantages . . .

But by their second edition in 1994 [6], they were obliged to revise this passage as follows:

> Not so long ago, digital transmission of voice and video was considered wasteful of bandwidth, and the cost . . . was of concern. [More recently, there has been] a complete turnabout in thinking . . . In fact, today virtually all communication is either already digital, in the process of being converted to digital, or under consideration for conversion.

[1] Shannon explains that "bit" is a contraction of "binary digit," and credits the neologism to J. W. Tukey.

The most important factor in the digital communications revolution has undoubtedly been the staggering technological progress of microelectronics and optical fiber technology. For wireline and wireless radio transmission (but not optical), another essential factor has been progress in channel coding, data compression and signal processing algorithms. For instance, data compression algorithms that can encode telephone-quality speech at 8–16 kbps and voiceband modem algorithms that can transmit 40–56 kbps over ordinary telephone lines have become commodities that require a negligible fraction of the capacity of today's personal-computer microprocessors.

This book attempts to tell the channel coding part of this story. In particular, it focusses on coding for the point-to-point additive white Gaussian noise (AWGN) channel. This choice is made in part for pedagogical reasons, but also because in fact almost all of the advances in practical channel coding have taken place in this arena. Moreover, performance on the AWGN channel is the standard benchmark for comparison of different coding schemes.

1.1 Shannon's grand challenge

The field of information theory and coding has a unique history, in that many of its ultimate limits were determined at the very beginning, in Shannon's founding paper [7].

Shannon's most celebrated result is his channel capacity theorem, which we will review in Chapter 3. This theorem states that for many common classes of channels there exists a channel capacity C such that there exist codes at any rate $R < C$ that can achieve arbitrarily reliable transmission, whereas no such codes exist for rates $R > C$. For a band-limited AWGN channel, the capacity C in bits per second (b/s) depends on only two parameters, the channel bandwidth W in Hz and the signal-to-noise ratio SNR, as follows:

$$C = W \log_2(1 + \mathrm{SNR}) \quad \text{b/s}.$$

Shannon's theorem has posed a magnificent challenge to succeeding generations of researchers. Its proof is based on randomly chosen codes and optimal (maximum likelihood) decoding. In practice, it has proved to be remarkably difficult to find classes of constructive codes that can be decoded by feasible decoding algorithms at rates which come at all close to the Shannon limit. Indeed, for a long time this problem was regarded as practically insoluble. Each significant advance toward this goal has been awarded the highest accolades the coding community has to offer, and most such advances have been immediately incorporated into practical systems.

In the next two sections we give a brief history of these advances for two different practical channels: the deep-space channel and the telephone channel. The deep-space channel is an unlimited-bandwidth, power-limited AWGN channel, whereas the telephone channel is very much bandwidth-limited. (We realize that many of the terms used here may be unfamiliar to the reader at this point, but we hope that these surveys will give at least an impressionistic picture. After reading later chapters, the reader may wish to return to reread these sections.)

Within the past decade there have been remarkable breakthroughs, principally the invention of turbo codes [1] and the rediscovery of low-density parity check (LDPC) codes [4], which have allowed the capacity of AWGN and similar channels to be approached in a practical sense. For example, Figure 1 (from [2]) shows that an optimized rate-1/2 LDPC code on an AWGN channel can approach the relevant Shannon limit within 0.0045 decibels (dB) in theory, and within 0.04 dB with an arguably practical code of block length 10^7 bits. Practical systems using block lengths of the order of 10^4–10^5 bits now approach the Shannon limit within tenths of a dB.

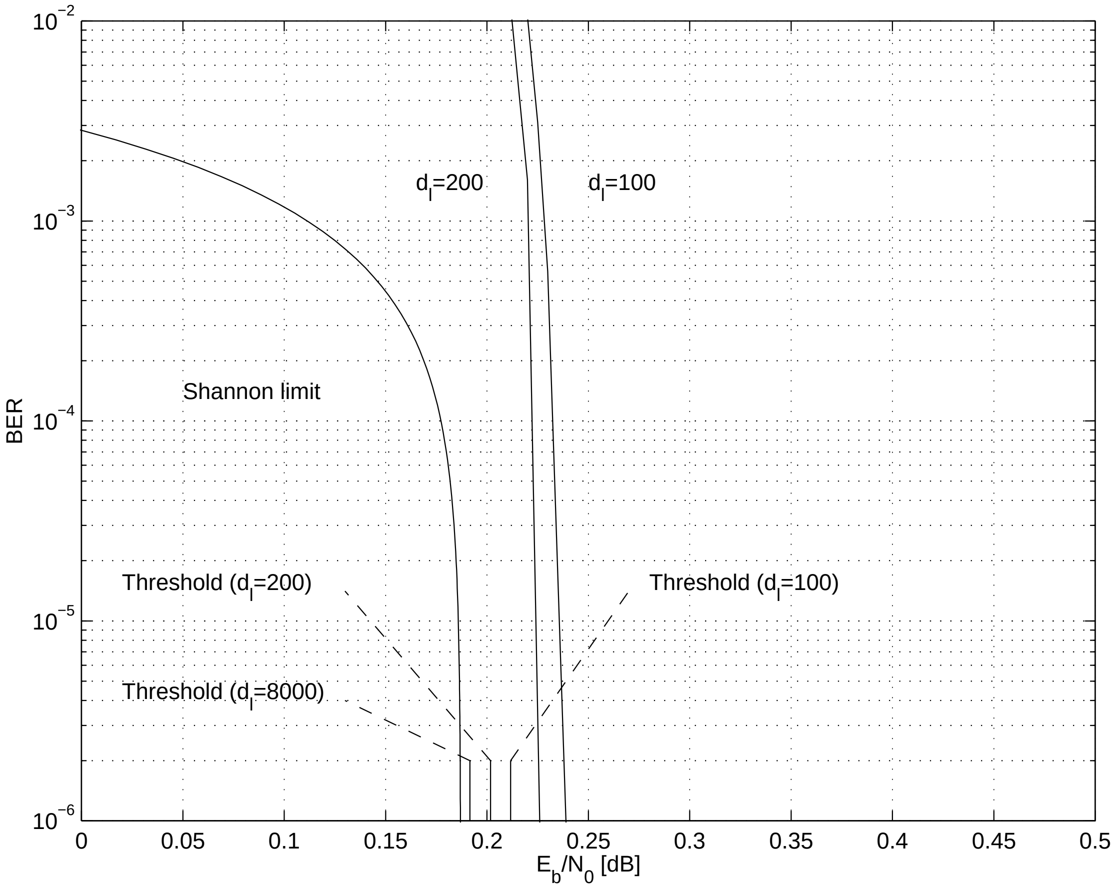

Figure 1. Bit error rate *vs.* E_b/N_0 in dB for optimized irregular rate-1/2 binary LDPC codes with maximum left degree d_l. Threshold: theoretical limit as block length $\to \infty$. Solid curves: simulation results for block length $= 10^7$. Shannon limit: binary codes, $R = 1/2$. (From [2].)

Here we will tell the story of how Shannon's challenge has been met for the AWGN channel, first for power-limited channels, where binary codes are appropriate, and then for bandwidth-limited channels, where multilevel modulation must be used. We start with the simplest schemes and work up to capacity-approaching codes, which for the most part follows the historical sequence.

1.2 Brief history of codes for deep-space missions

The deep-space communications application has been the arena in which most of the most powerful coding schemes for the power-limited AWGN channel have been first deployed, because:

- The only noise is AWGN in the receiver front end;
- Bandwidth is effectively unlimited;
- Fractions of a dB have huge scientific and economic value;
- Receiver (decoding) complexity is effectively unlimited.

For power-limited AWGN channels, we will see that there is no penalty to using binary codes with binary modulation rather than more general modulation schemes.

The first coded scheme to be designed was a simple $(32, 6, 16)$ biorthogonal code for the Mariner missions (1969), decoded by efficient maximum-likelihood decoding (the fast Hadamard transform, or "Green machine;" see Exercise 2, below). We will see that such a scheme can achieve a nominal coding gain of 3 (4.8 dB). At a target error probability per bit of $P_b(E) \approx 5 \cdot 10^{-3}$, the actual coding gain achieved was only about 2.2 dB.

The first coded scheme actually to be launched was a rate-1/2 convolutional code with constraint length $\nu = 20$ for the Pioneer 1968 mission. The receiver used 3-bit soft decisions and sequential decoding implemented on a general-purpose 16-bit minicomputer with a 1 MHz clock rate. At 512 b/s, the actual coding gain achieved at $P_b(E) \approx 5 \cdot 10^{-3}$ was about 3.3 dB.

During the 1970's, the NASA standard became a concatenated coding scheme based on a $\nu = 6$, rate-1/3 inner convolutional code and a $(255, 223, 33)$ Reed-Solomon outer code over $\mathbb{F}_{256}$. Such a system can achieve a real coding gain of about 8.3 dB at $P_b(E) \approx 10^{-6}$.

When the primary antenna failed to deploy on the Galileo mission (*circa* 1992), an elaborate concatenated coding scheme using a $\nu = 14$ rate-1/4 inner code with a Big Viterbi Decoder (BVD) and a set of variable-strength RS outer codes was reprogrammed into the spacecraft computers. This scheme was able to operate at $E_b/N_0 \approx 0.8$ dB at $P_b(E) \approx 2 \cdot 10^{-7}$, for a real coding gain of about 10.2 dB.

Turbo coding systems for deep-space communications have been developed by NASA's Jet Propulsion Laboratory (JPL) and others to get within 1 dB of the Shannon limit, and have now been standardized.

For a more comprehensive history of coding for deep-space channels, see [3].

1.3 Brief history of telephone-line modems

For several decades the telephone channel was the arena in which the most powerful coding and modulation schemes for the bandwidth-limited AWGN channel were first developed and deployed, because:

- The telephone channel is fairly well modeled as a band-limited AWGN channel;
- One dB has a significant commercial value;
- Data rates are low enough that a considerable amount of processing can be done per bit.

To approach the capacity of bandwidth-limited AWGN channels, multilevel modulation must be used. Moreover, it is important to use as much of the available bandwidth as possible.

The earliest modems developed in the 1950s and 1960s (Bell 103 and 202, and international standards V.21 and V.23) used simple binary frequency-shift keying (FSK) to achieve data rates of 300 and 1200 b/s, respectively. Implementation was entirely analog.

The first synchronous "high-speed" modem was the Bell 201 (later V.24), a 2400 b/s modem which was introduced about 1962. This modem used four-phase (4-PSK) modulation at 1200 symbols/s, so the nominal (Nyquist) bandwidth was 1200 Hz. However, because the modulation pulse had 100% rolloff, the actual bandwidth used was closer to 2400 Hz.

The first successful 4800 b/s modem was the Milgo 4400/48 (later V.27), which was introduced about 1967. This modem used eight-phase (8-PSK) modulation at 1600 symbols/s, so the nominal (Nyquist) bandwidth was 1600 Hz. "Narrow-band" filters with 50% rolloff kept the actual bandwidth used to 2400 Hz.

The first successful 9600 b/s modem was the Codex 9600C (later V.29), which was introduced in 1971. This modem used quadrature amplitude modulation (QAM) at 2400 symbols/s with an unconventional 16-point signal constellation (see Exercise 3, below) to combat combined "phase jitter" and AWGN. More importantly, it used digital adaptive linear equalization to keep the actual bandwidth needed to not much more than the Nyquist bandwidth of 2400 Hz.

All of these modems were designed for private point-to-point conditioned voice-grade lines, which use four-wire circuits (independent transmission in each direction) whose quality is higher and more consistent than that of the typical telephone connection in the two-wire (simultaneous transmission in both directions) public switched telephone network (PSTN).

The first international standard to use coding was the V.32 standard (1986) for 9600 b/s transmission over the PSTN (later raised to 14.4 kb/s in V.32*bis*). This modem used an 8-state, two-dimensional (2D) rotationally invariant Wei trellis code to achieve a coding gain of about 3.5 dB with a 32-QAM (later 128-QAM) constellation at 2400 symbols/s, again with an adaptive linear equalizer. Digital echo cancellation was also introduced to combat echoes on two-wire channels.

The "ultimate modem standard" was V.34 (1994) for transmission at up to 28.8 kb/s over the PSTN (later raised to 33.6 kb/s in V.34*bis*). This modem used a 16-state, 4D rotationally invariant Wei trellis code to achieve a coding gain of about 4.0 dB with a variable-sized QAM constellation with up to 1664 points. An optional 32-state, 4D trellis code with an additional coding gain of 0.3 dB and four times (4x) the decoding complexity and a 64-state, 4D code with a further 0.15 dB coding gain and a further 4x increase in complexity were also provided. A 16D "shell mapping" constellation shaping scheme provided an additional shaping gain of about 0.8 dB (see Exercise 4, below). A variable symbol rate of up to 3429 symbols/s was used, with symbol rate and data rate selection determined by "line probing" of individual channels. Nonlinear transmitter precoding combined with adaptive linear equalization in the receiver was used for equalization, again with echo cancellation. In short, this modem used almost every tool in the AWGN channel toolbox.

However, this standard was shortly superseded by V.90 (1998). V.90 is based on a completely different, non-AWGN model for the telephone channel: namely, it recognizes that within today's PSTN, analog signals are bandlimited, sampled and quantized to one of 256 amplitude levels at 8 kHz, transmitted digitally at 64 kb/s, and then eventually reconstructed by pulse amplitude modulation (PAM). By gaining direct access to the 64 kb/s digital data stream at a central site, and by using a well-spaced subset of the pre-existing nonlinear 256-PAM constellation, data can easily be transmitted at 40–56 kb/s (see Exercise 5, below). In V.90, such a scheme is used for downstream transmission only, with V.34 modulation upstream. In V.92 (2000) this scheme has been extended to the more difficult upstream direction.

Neither V.90 nor V.92 uses coding, nor the other sophisticated techniques of V.34. In this sense, the end of the telephone-line modem story is a bit of a fizzle. However, techniques similar to those of V.34 are now used in higher-speed wireline modems, such as digital subscriber line (DSL) modems, as well as on wireless channels such as digital cellular. In other words, the story continues in other settings.

1.4 Exercises

In this section we offer a few warm-up exercises to give the reader some preliminary feeling for data communication on the AWGN channel.

In these exercises the underlying channel model is assumed to be a discrete-time AWGN channel whose output sequence is given by $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{X}$ is a real input data sequence and $\mathbf{N}$ is a sequence of real independent, identically distributed (iid) zero-mean Gaussian noise variables. This model will be derived from a continuous-time model in Chapter 2.

We will also give the reader some practice in the use of decibels (dB). In general, a dB representation is useful wherever logarithms are useful; *i.e.,* wherever a real number is a multiplicative factor of some other number, and particularly for computing products of many factors. The dB scale is simply the logarithmic mapping

$$\text{ratio or multiplicative factor of } \alpha \leftrightarrow 10 \log_{10} \alpha \text{ dB},$$

where the scaling is chosen so that the decade interval 1–10 maps to the interval 0–10. (In other words, the value of α in dB is $\log_\beta \alpha$, where $\beta = 10^{0.1} = 1.2589....$) This scale is convenient for human memory and calculation. It is often useful to have the little log table below committed to memory, even in everyday life (see Exercise 1, below).

α	dB (round numbers)	dB (two decimal places)
1	0	0.00
1.25	1	0.97
2	3	3.01
2.5	4	3.98
e	4.3	4.34
3	4.8	4.77
π	5	4.97
4	6	6.02
5	7	6.99
8	9	9.03
10	10	10.00

Exercise 1. (Compound interest and dB) How long does it take to double your money at an interest rate of $P\%$? The bankers' "Rule of 72" estimates that it takes about $72/P$ years; *e.g.,* at a 5% interest rate compounded annually, it takes about 14.4 years to double your money.

(a) An engineer decides to interpolate the dB table above linearly for $1 \le 1 + p \le 1.25$; *i.e.,*

$$\text{ratio or multiplicative factor of } 1 + p \leftrightarrow 4p \text{ dB}.$$

Show that this corresponds to a "Rule of 75;" *e.g.,* at a 5% interest rate compounded annually, it takes 15 years to double your money.

(b) A mathematician linearly approximates the dB table for $p \approx 0$ by noting that as $p \to 0$, $\ln(1+p) \to p$, and translates this into a "Rule of N" for some real number N. What is N? Using this rule, how many years will it take to double your money at a 5% interest rate, compounded annually? What happens if interest is compounded continuously?

(c) How many years will it actually take to double your money at a 5% interest rate, compounded annually? [Hint: $10 \log_{10} 7 = 8.45$ dB.] Whose rule best predicts the correct result?

Exercise 2. (Biorthogonal codes) A $2^m \times 2^m$ $\{\pm 1\}$-valued Hadamard matrix H_{2^m} may be constructed recursively as the m-fold tensor product of the 2×2 matrix

$$H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix},$$

as follows:

$$H_{2^m} = \begin{bmatrix} +H_{2^{m-1}} & +H_{2^{m-1}} \\ +H_{2^{m-1}} & -H_{2^{m-1}} \end{bmatrix}.$$

(a) Show by induction that:

(i) $(H_{2^m})^T = H_{2^m}$, where ^{T} denotes the transpose; *i.e.,* H_{2^m} is symmetric;

(ii) The rows or columns of H_{2^m} form a set of mutually orthogonal vectors of length 2^m;

(iii) The first row and the first column of H_{2^m} consist of all $+1$s;

(iv) There are an equal number of $+1$s and -1s in all other rows and columns of H_{2^m};

(v) $H_{2^m} H_{2^m} = 2^m I_{2^m}$; *i.e.,* $(H_{2^m})^{-1} = 2^{-m} H_{2^m}$, where $^{-1}$ denotes the inverse.

(b) A biorthogonal signal set is a set of real equal-energy orthogonal vectors and their negatives. Show how to construct a biorthogonal signal set of size 64 as a set of $\{\pm 1\}$-valued sequences of length 32.

(c) A simplex signal set S is a set of real equal-energy vectors that are equidistant and that have zero mean $\mathbf{m}(S)$ under an equiprobable distribution. Show how to construct a simplex signal set of size 32 as a set of 32 $\{\pm 1\}$-valued sequences of length 31. [Hint: The fluctuation $O - \mathbf{m}(O)$ of a set O of orthogonal real vectors is a simplex signal set.]

(d) Let $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ be the received sequence on a discrete-time AWGN channel, where the input sequence $\mathbf{X}$ is chosen equiprobably from a biorthogonal signal set B of size 2^{m+1} constructed as in part (b). Show that the following algorithm implements a minimum-distance decoder for B (*i.e.,* given a real 2^m-vector $\mathbf{y}$, it finds the closest $\mathbf{x} \in B$ to $\mathbf{y}$):

(i) Compute $\mathbf{z} = H_{2^m} \mathbf{y}$, where $\mathbf{y}$ is regarded as a column vector;

(ii) Find the component z_j of $\mathbf{z}$ with largest magnitude $|z_j|$;

(iii) Decode to $\mathrm{sgn}(z_j)\mathbf{x}_j$, where $\mathrm{sgn}(z_j)$ is the sign of the largest-magnitude component z_j and $\mathbf{x}_j$ is the corresponding column of H_{2^m}.

(e) Show that a circuit similar to that shown below for $m = 2$ can implement the $2^m \times 2^m$ matrix multiplication $\mathbf{z} = H_{2^m} \mathbf{y}$ with a total of only $m \times 2^m$ addition and subtraction operations. (This is called the "fast Hadamard transform," or "Walsh transform," or "Green machine.")

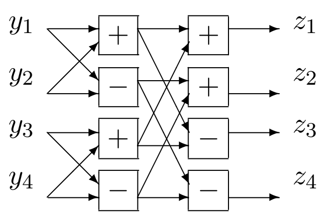

Figure 2. Fast $2^m \times 2^m$ Hadamard transform for $m = 2$.

Exercise 3. (16-QAM signal sets) Three 16-point 2-dimensional quadrature amplitude modulation (16-QAM) signal sets are shown in Figure 3, below. The first is a standard 4×4 signal set; the second is the V.29 signal set; the third is based on a hexagonal grid and is the most power-efficient 16-QAM signal set known. The first two have 90° symmetry; the last, only 180°. All have a minimum squared distance between signal points of $d^2_{\min} = 4$.

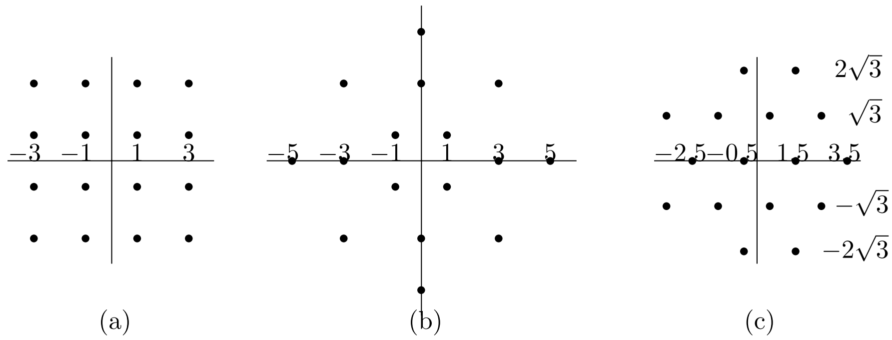

Figure 3. 16-QAM signal sets. (a) (4 × 4)-QAM; (b) V.29; (c) hexagonal.

(a) Compute the average energy (squared norm) of each signal set if all points are equiprobable. Compare the power efficiencies of the three signal sets in dB.

(b) Sketch the decision regions of a minimum-distance detector for each signal set.

(c) Show that with a phase rotation of ±10° the minimum distance from any rotated signal point to any decision region boundary is substantially greatest for the V.29 signal set.

Exercise 4. (Shaping gain of spherical signal sets) In this exercise we compare the power efficiency of n-cube and n-sphere signal sets for large n.

An n-cube signal set is the set of all odd-integer sequences of length n within an n-cube of side $2M$ centered on the origin. For example, the signal set of Figure 3(a) is a 2-cube signal set with $M = 4$.

An n-sphere signal set is the set of all odd-integer sequences of length n within an n-sphere of squared radius r^2 centered on the origin. For example, the signal set of Figure 3(a) is also a 2-sphere signal set for any squared radius r^2 in the range $18 \leq r^2 < 25$. In particular, it is a 2-sphere signal set for $r^2 = 64/\pi = 20.37$, where the area πr^2 of the 2-sphere (circle) equals the area $(2M)^2 = 64$ of the 2-cube (square) of the previous paragraph.

Both n-cube and n-sphere signal sets therefore have minimum squared distance between signal points $d^2_{\min} = 4$ (if they are nontrivial), and n-cube decision regions of side 2 and thus volume 2^n associated with each signal point. The point of the following exercise is to compare their average energy using the following large-signal-set approximations:

- The number of signal points is approximately equal to the volume $V(\mathcal{R})$ of the bounding n-cube or n-sphere region $\mathcal{R}$ divided by 2^n, the volume of the decision region associated with each signal point (an n-cube of side 2).

- The average energy of the signal points under an equiprobable distribution is approximately equal to the average energy $E(\mathcal{R})$ of the bounding n-cube or n-sphere region $\mathcal{R}$ under a uniform continuous distribution.

(a) Show that if $\mathcal{R}$ is an n-cube of side $2M$ for some integer M, then under the two above approximations the approximate number of signal points is M^n and the approximate average energy is $nM^2/3$. Show that the first of these two approximations is exact.

(b) For n even, if $\mathcal{R}$ is an n-sphere of radius r, compute the approximate number of signal points and the approximate average energy of an n-sphere signal set, using the following known expressions for the volume $V_{\otimes}(n, r)$ and the average energy $E_{\otimes}(n, r)$ of an n-sphere of radius r:

$$\begin{aligned} V_{\otimes}(n, r) &= \frac{(\pi r^2)^{n/2}}{(n/2)!}; \\ E_{\otimes}(n, r) &= \frac{nr^2}{n+2}. \end{aligned}$$

(c) For $n = 2$, show that a large 2-sphere signal set has about 0.2 dB smaller average energy than a 2-cube signal set with the same number of signal points.

(d) For $n = 16$, show that a large 16-sphere signal set has about 1 dB smaller average energy than a 16-cube signal set with the same number of signal points. [Hint: $8! = 40320$ (46.06 dB).]

(e) Show that as $n \to \infty$ a large n-sphere signal set has a factor of $\pi e/6$ (1.53 dB) smaller average energy than an n-cube signal set with the same number of signal points. [Hint: Use Stirling's approximation, $m! \to (m/e)^m$ as $m \to \infty$.]

Exercise 5. (56 kb/s PCM modems)

This problem has to do with the design of "56 kb/s PCM modems" such as V.90 and V.92.

In the North American telephone network, voice is commonly digitized by low-pass filtering to about 3.8 KHz, sampling at 8000 samples per second, and quantizing each sample into an 8-bit byte according to the so-called "μ law." The μ law specifies 255 distinct signal levels, which are a quantized, piecewise-linear approximation to a logarithmic function, as follows:

- 1 level at 0;
- 15 positive levels evenly spaced with $d = 2$ between 2 and 30 (*i.e.*, 2, 4, 6, 8, ..., 30);
- 16 positive levels evenly spaced with $d = 4$ between 33 and 93;
- 16 positive levels evenly spaced with $d = 8$ between 99 and 219;
- 16 positive levels evenly spaced with $d = 16$ between 231 and 471;
- 16 positive levels evenly spaced with $d = 32$ between 495 and 975;
- 16 positive levels evenly spaced with $d = 64$ between 1023 and 1983;
- 16 positive levels evenly spaced with $d = 128$ between 2079 and 3999;
- 16 positive levels evenly spaced with $d = 256$ between 4191 and 8031;
- plus 127 symmetric negative levels.

The resulting 64 kb/s digitized voice sequence is transmitted through the network and ultimately reconstructed at a remote central office by pulse amplitude modulation (PAM) using a μ-law digital/analog converter and a 4 KHz low-pass filter.

For a V.90 modem, one end of the link is assumed to have a direct 64 kb/s digital connection and to be able to send any sequence of 8000 8-bit bytes per second. The corresponding levels are reconstructed at the remote central office. For the purposes of this exercise, assume that the reconstruction is exactly according to the μ-law table above, and that the reconstructed pulses are then sent through an ideal 4 KHz additive AWGN channel to the user.

(a) Determine the maximum number M of levels that can be chosen from the 255-point μ-law constellation above such that the minimum separation between levels is $d = 2$, 4, 8, 16, 64, 128, 256, 512, or 1024, respectively.

(b) These uncoded M-PAM subconstellations may be used to send up to $r = \log_2 M$ bits per symbol. What level separation can be obtained while sending 40 kb/s? 48 kb/s? 56 kb/s?

(c) How much more SNR in dB is required to transmit reliably at 48 kb/s compared to 40 kb/s? At 56 kb/s compared to 48 kb/s?

References

[1] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," *Proc. 1993 Int. Conf. Commun.* (Geneva), pp. 1064–1070, May 1993.

[2] S.-Y. Chung, G. D. Forney, Jr., T. J. Richardson and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB from the Shannon limit," *IEEE Commun. Letters*, vol. 5, pp. 58–60, Feb. 2001.

[3] D. J. Costello, Jr., J. Hagenauer, H. Imai and S. B. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2531–2560, Oct. 1998.

[4] R. G. Gallager, *Low-Density Parity-Check Codes.* Cambridge, MA: MIT Press, 1962.

[5] E. A. Lee and D. G. Messerschmitt, *Digital Communication* (first edition). Boston: Kluwer, 1988.

[6] E. A. Lee and D. G. Messerschmitt, *Digital Communication* (second edition). Boston: Kluwer, 1994.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.