Chapter 1

Introduction

1.1 Modern Digital Communications

With the advent of high-speed logic circuits and very large scale integration (VLSI), data processing and storage equipment has inexorably moved towards employing digital techniques. In digital systems, data is encoded into strings of zeros and ones, corresponding to the on and off states of semiconductor switches. This has brought about fundamental changes in how information is processed. While real-world data is primarily in "analog form" of one type or another, the revolution in digital processing means that this analog information needs to be encoded into a digital representation, e.g., into a string of ones and zeros. The conversion from analog to digital and back are processes which have become ubiquitous. Examples are the digital encoding of speech, picture, and video encoding and rendering, as well as the large variety of capturing and representing data encountered in our modern internet-based lifestyles.

The migration from analog communications of the first half of the 20-th century to the now ubiquitous digital forms of communications were enabled primarily by the fast-paced advances in high-density device integration. This has been the engine behind much of the technological progress over the last half century, initiated by the creation of the first integrated circuit (IC) by Kilby at Texas Instruments in 1958. Following Moore's informal law, device sizes, primarily CMOS (Complementary Metal-Oxide Semiconductors), shrink by a factor two every two years, and computational power doubles accordingly. An impression for this exponential growth in computing capability can be gained from Figure 1.1, which shows the number of transistors integrated in a single circuit and the minimum device size for progressive fabrication processes – known as *implementation nodes*.

While straightforward miniaturization of the CMOS devices is becoming increasingly more difficult, transistor designers have been very creative in modifying the designs to stay on the Moore trajectory. As of 2015 we now see the introduction of 3-dimensional

transistor structures such as thin FETs, double-gated FETs, and tunnel FETs, and it is expected that carbon nanotube devices may continue miniaturization well into the sub-10 nm range. In any case, the future for highly complex computational devices is bright.

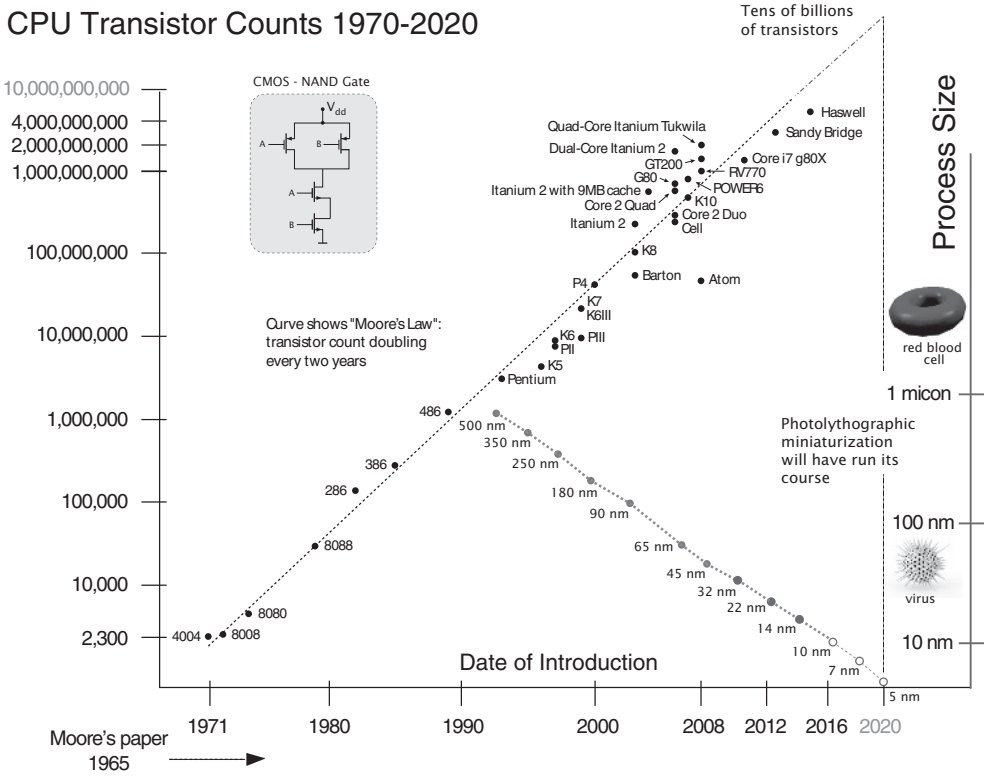

Figure 1.1: Moore's law is driving progress in electronic devices. Top left: A basic CMOS switching structure. Bottom left: Moore observed his "doubling law" in 1965 and predicted that it would continue "at least another 10 years."

One such computational challenge is data communications: in particular data integrity, as discussed in this book. The migration from analog to digital information processing has opened the door for many sophisticated algorithmic methods. Digital information is

treated differently in communications than analog information. Signal estimation becomes signal detection; that is, a communications receiver need not look for an analog signal and make a "best" estimate, it only needs to make a decision between a finite number of discrete signals, say a one or a zero in the most basic case. Digital signals are more reliable in a noisy communications environment; they can usually be detected perfectly, as long as the noise levels are below a certain threshold. This allows us to restore digital data, and, through error correcting techniques, correct errors made during transmission. Digital data can easily be encoded in such a way as to introduce dependency among a large number of symbols, thus enabling a receiver to make a more accurate detection of the symbols. This is the essence of *error control coding*.

Finally, there are also strong theoretical reasons behind the migration to digital processing. Nyquist's sampling theorem, discussed in Section 1.3, tells us that, fundamentally, it is sufficient to know an analog signal at a number of discrete points in time. This opens the door for the discrete time treatment of signals. Then, Shannon's fundamental channel coding theorem states that the values of these discrete time samples themselves, can contain only a finite amount of information. Therefore, only a finite amount of discrete levels are required to capture the full information content of a signal.

The digitization of data is convenient for a number of other reasons too. The design of signal processing algorithms for digital data is much easier than designing analog signal processing algorithms, albeit not typically less complex. However, the abundance of such digital algorithms, including the error control and correction techniques discussed in this book, combined with their ease of implementation in *very large scale integrated* (VLSI) circuits has led to the plethora of successful applications of error control coding we see in practice today.

Error control coding was first applied in deep-space communications where we are confronted with low-power communications channels with virtually unlimited bandwidth. On these data links, convolutional codes (Chapter 4) are used with sequential and Viterbi decoding (Chapter 5), and the future will see the application of turbo coding. The next successful application of error control coding was to storage devices, most notably the compact disk player, which employs powerful Reed-Solomon codes [21] to handle the raw error probability from the optical readout device which is too large for high-fidelity sound reproduction without error correction. Another hurdle taken was the successful application of error control to bandwidth-limited telephone channels, where trellis-coded modulation (Chapter 3) was used to produce impressive improvements and push transmission rates towards the theoretical limit of the channel. Nowadays, coding is routinely applied to satellite communications [41, 49], teletext broadcasting, computer storage devices, logic circuits, semiconductor memory systems, magnetic recording systems, audio-video, and WiFi systems. Modern mobile communications systems like the pan-European TDMA digital telephony standard GSM [35], IS 95 [47], CDMA2000, IMT2000, and the new 4-th generation LTE and LTE-A standards [63, 64] all use error control coding.

1.2 The Rise of Digital Communications

Modern digital communication theory started in 1928 with Nyquist's seminal work on telegraph transmission theory [36]. The message from Nyquist's theory is that finite bandwidth implies discrete time. That is, a signal whose bandwidth is limited can always be represented by sample values taken at discrete time intervals. The sampling theorem of this theory then asserts that the band-limited signal can always be reconstructed *exactly* from these discrete-time samples.[1] Only these discrete samples need to be processed by a receiver since they contain all the necessary information of the entire waveform.

The second pillar to establish the supremacy of digital information processing came precisely from Shannon's 1948 theory. Shannon's theory essentially establishes that the discrete-time samples which are used to represent a bandlimited signal, could be adequately described by a finite number of amplitude samples, the number of which depended on the level of the channel noise. These two theories combined state that a finite number of levels taken at discrete time intervals are completely sufficient to characterize any bandlimited signal in the presence of noise, that is, in any communication system.

With these results, technology has moved towards a complete digitization of communications systems, with error control coding being the key to realize the sufficiency of discrete amplitude levels. We will study Shannon's theorem in more detail in Section 1.5.

1.3 Communication Systems

Figure 1.2 shows the basic configuration of a point-to-point digital communications link. The data to be transmitted over this link can either come from some analog source, in which case it must first be converted into digital format (digitized), or it can be a digital information source. If this data is a speech signal, for example, the digitizer is a speech codec [22]. Usually the digital data is source encoded to remove unnecessary redundancy from the data, i.e., the source data is compressed [14]. Source encoding has the effect that the digital data which enters the encoder has statistics which resemble that of a random symbol source with maximum entropy, i.e., all the different digital symbols occur with equal likelihood, and are statistically independent. The channel encoder operates on this compressed data and introduces controlled redundancy for transmission over the channel. The modulator converts the discrete channel symbols into waveforms which are transmitted through the waveform channel. The demodulator reconverts the waveforms

[1] Since it is not shown elsewhere in this book, we present Nyquist's sampling theorem here. It is given by the following exact series expansion of the function $s(t)$ which is bandlimited to $[-1/2T, 1/2T]$:

$$s(t) = \sum_{i=-\infty}^{\infty} s(iT)\,\mathrm{sinc}\Big(\frac{\pi}{T}(t - iT)\Big); \qquad \mathrm{sinc}(x) = \frac{\sin(x)}{x}.$$

back into a discrete sequence of received symbols, and the decoder reproduces an estimate of the compressed input data sequence, which is subsequently reconverted into the original signal or data sequence.

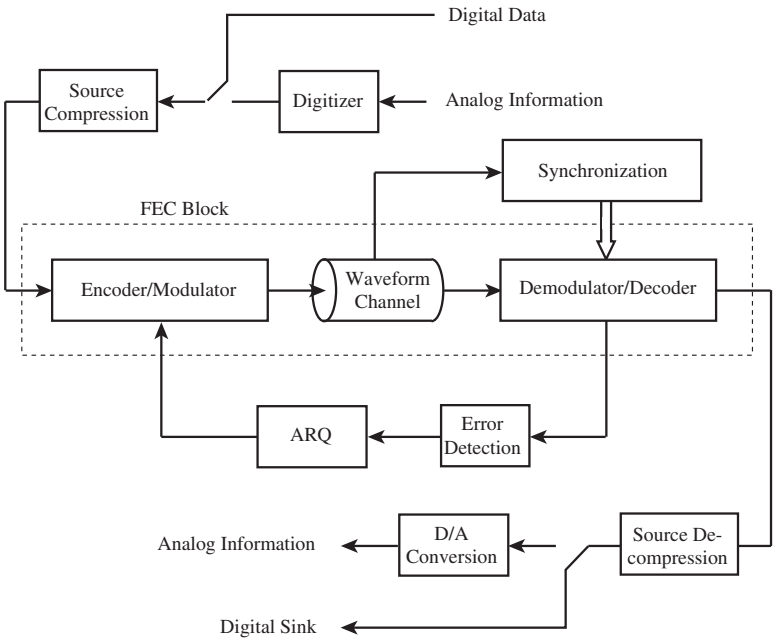

Figure 1.2: System diagram of a complete point-to-point communication system for digital data. The forward error control (FEC) block is the topic of this book.

An important ancillary function at the receiver is the synchronization process. We usually need to acquire carrier frequency and phase synchronization, as well as symbol timing synchronization in order for the receiver to be able to operate. Synchronization is not a topic of this book, and we will assume in most of our discussion that synchronization has been established (with the exception of phase synchronization in the case of rotationally invariant codes in Chapters 4 and 8). References [32] and [33] treat synchronization issues in detail. Since synchronization is a relatively slow estimation process, and data detection is a fast process, we usually have those two operations separated in real receiver implementations as indicated in Figure 1.2. However, we would like to note at this point that novel, iterative receiver designs are increasingly integrating these auxillary functions, for an example of phase synchronization see Chapter 8.

Another important feature in some communication systems is *automatic repeat request* (ARQ). In systems with ARQ the receiver also performs error detection, and, through a return channel, requests retransmission of erroneous data blocks, or data blocks which cannot be reconstructed with sufficient confidence [25]. ARQ can usually improve the data transmission quality substantially, but the return channel needed for ARQ is not always available, or may be impractical. For a deep-space probe on its way to the outer rim of our solar system, ARQ is infeasible since the return path takes too long (several hours!). For speech-encoded signals ARQ is usually impossible for an analogous reason, since only a maximum speech delay is of about 200 ms is acceptable. In broadcast systems, ARQ is ruled out for obvious reasons. Error control coding without ARQ is termed *forward error correction* or *control* (FEC). FEC is more difficult to perform than simple error detection and ARQ, but dispenses with the return channel. Oftentimes, FEC and ARQ are both integrated into hybrid error control systems [25, 24] for data communications.

This book deals primarily with FEC—the dashed block in Figure 1.2. The reason why we can do this relatively easily is due to different functionalities of the various blocks just discussed, and the fact that they operate largely autonomously from each other. Each of them represents a separate entity with its own optimization strategy, and data is simply passed between the different blocks, sometimes with lithe extra mutual interaction. A notable point in Figure 1.2 is that the encoder/modulator and the demodulator/decoder are combined operations. This is done to reflect the fact that error protection and modulation–in the sense of choosing the discrete signal points that represent the digital data–is a process best addressed jointly. This view of joint encoding/modulation was first proposed by Wozencraft and Kennedy [61] and Massey [30] and then realized with stunning results by Ungerböck [50, 51, 52] in the methods of *trellis-coded modulation* of the 1980's.

Since we will assume the encoder input data to be a sequence of independent, identically and uniformly distributed symbols (courtesy of the source compression), the single most important parameter to optimize for the FEC block is arguably the bit and/or symbol error rate, and we will adopt this as our criterion for the quality of an FEC system. Note that this is not necessarily the most meaningful measure in all cases. Consider, for example, pulse-code-modulated (PCM) speech, where an error in the most significant bit is significantly more detrimental than an error in the least significant bit. Researchers have looked at schemes with unequal error protection for such applications (e.g., [18]). However, such methods usually are a variation of the basic theme of obtaining a minimum error rate. Occasionally we may have need for the frame-, or block error rate (FER), which describes the probability that an entire block of data is incorrectly received. Most communications protocols operate on the basis of frame errors, and frame error control is exercised at the upper layers of a communications protocol. While of course tightly connected, the frame-, and bit error rates do sometimes follow different tendencies. This will become important when we discuss error floors for very-low error rate applications, such as those using cable modem or optical fiber communications.

1.4 Error Control Coding

The modern approach to error control in digital communications started with the groundbreaking work of Shannon [45], Hamming [19], and Golay [16]. While Shannon presented a theory which explained the fundamental limits on the efficiency of communications systems, Hamming and Golay were the first to develop practical error control schemes. A new paradigm was born, one in which errors are not synonymous with data which is irretrievably lost; but by clever design, errors could be corrected, or avoided altogether. This new thinking was revolutionary. Even though Shannon's theory promised that large improvements in the performance of communication systems could be achieved, practical improvements had to be excavated by laborious work over half a century of intensive research. One reason for this lies in a fundamental shortcoming of Shannon's theory. While it clearly states theoretical limits on communication efficiency, its methodology provides no insight on how to actually achieve these limits, since it is based on sophisticated averaging arguments which eliminate all detailed system structure. Coding theory, on the other hand, evolved from Hamming and Golay's work into a flourishing branch of applied mathematics [27].

Let us see where it all started. The most famous formula from Shannon's work is arguably the channel capacity of an ideal band-limited Gaussian channel,[2] which is given by

$$C = W \log_2(1 + S/N) \quad [\text{bits/second}]. \tag{1.1}$$

In this formula, C is the channel capacity, that is, the maximum number of bits which can be transmitted through this channel per unit time (second), W is the bandwidth of the channel, and S/N is the signal-to-noise power ratio at the receiver. Shannon's main theorem, which accompanies (1.1), asserts that error probabilities as small as desired can be achieved as long as the transmission rate R through the channel (in bits/second) is smaller than the channel capacity C. This can be achieved by using an appropriate encoding and decoding operation. However, Shannon's theory is silent about the structure of these encoders and decoders.

This new view was in marked contrast to early practices, which embodied the opinion that in order to reduce error probabilities, the signal energy had to be increased, i.e., the S/N had to be improved. Figure 1.3 shows the error performance of QPSK, a popular modulation method for satellite channels (see Chapter 2) which allows data transmission of rates up to 2 bits/symbol. The bit error probability (BER) of QPSK is shown as a function of the signal-to-noise ratio S/N per dimension normalized per bit (see Section 1.3), henceforth called SNR. It is evident that an increased SNR provides a gradual decrease in error probability. This contrasts markedly with Shannon's theory which promises zero(!) error probability at a spectral efficiency of 2 bits/s/Hz, which is the maximum that QPSK

[2]The exact definitions of these basic communications concepts are given in Chapter 2.

can achieve, as long as SNR > 1.5 (1.76 dB), shattering conventional wisdom. The limit on SNR is calculated using (1.1)–see Section 1.5.

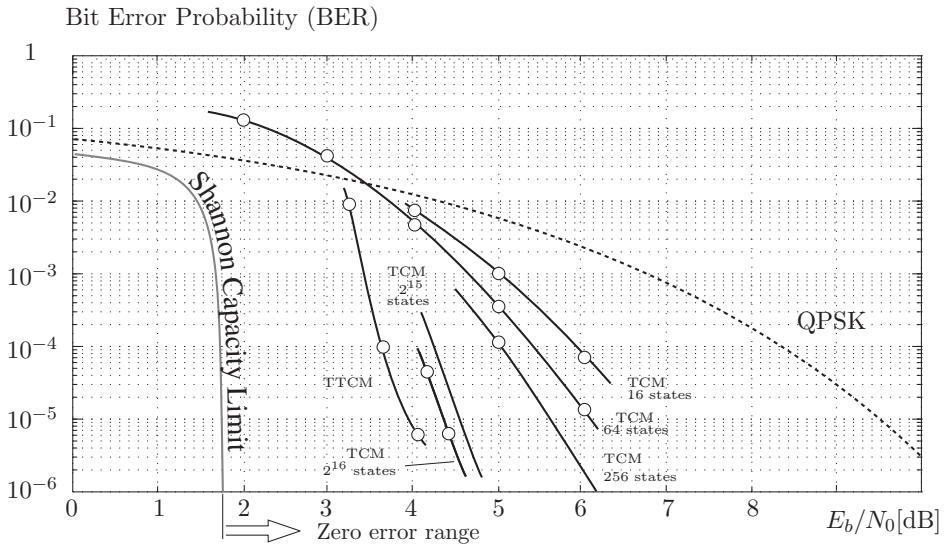

Figure 1.3: Bit error probability of quadrature phase-shift keying (QPSK) and selected 8-PSK trellis-coded modulation (TCM), trellis-turbo-coded (TTCM), and block-turbo-coded (BTC) systems as a function of the normalized signal-to-noise ratio.

Also shown in Figure 1.3 is the performance of several trellis-coded modulation (TCM) and trellis-turbo-coded (TTCM) schemes using 8-ary phase-shift keying (8-PSK) (Chapter 4), and the improvement made possible by coding becomes evident. The difference in SNR for an objective target bit error rate between a coded system and an uncoded system is termed the *coding gain*. Note that the coding schemes shown in Figure 1.3 achieves these gains without requiring more bandwidth than the uncoded QPSK system.

As we will discuss in Chapter 4, a trellis code is generated by a circuit with a finite number of internal states. The number of these states is a direct measure of its decoding complexity if maximum-likelihood decoding is used. Note that the two very large codes are not maximum-likelihood decoded, they are sequentially decoded [55]. The turbo-trellis-coded modulation (TTCM) system is based on turbo coding principles (Chapter 9) using two small concatenated trellis codes. Coding then helps to realize the promise of Shannon's theory which states that for a desired error rate of $P_b = 10^{-6}$ we can gain almost 9 dB in expended signal energy over QPSK. This gain can be achieved by converting required

signal power into decoder complexity, as is done by the TCM and TTCM coding methods.

Incidentally, 1948, the year Shannon published his work, is also the birth year[3] of the transistor, arguably the 20-th century's most fundamental invention, one which allowed the construction of very powerful, very small computing devices. Only this made the conversion from signal energy requirements to (system) complexity possible, giving coding and information theory a platform for practical realizations [21].

Figure 1.4 shows an early feasibility experiment, comparing the performance of a 16-state 8-PSK TCM code used in an experimental implementation of a single channel per carrier (SCPC) modem operating at 64 kbits/second [49] against QPSK and the theoretical performance established via simulations (Figure 1.3). This illustrates the viability of trellis coding for satellite channels. Interestingly, the 8-PSK TCM modem performance comes much closer to the theoretical performance than the original QPSK modem, achieving a practical coding gain of 5 dB. This is due to an effect where system inaccuracies, acting similar to noise, are handled better by a coded system.

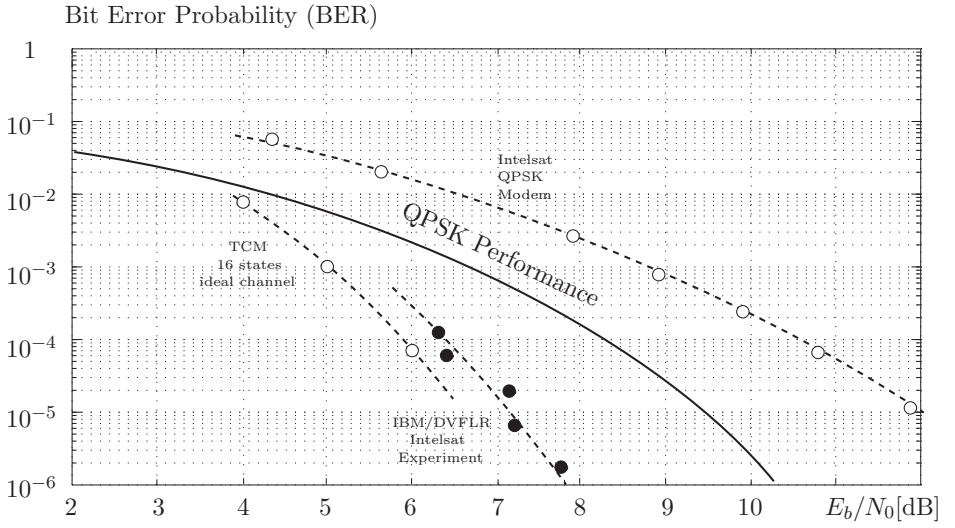

Figure 1.4: Measured bit error probability of (QPSK) and a 16-state 8-PSK (TCM) modem over a 64 kbit/s satellite channel [49]. The discrepancy between the theoretical curves and the implemented BER results are due to non-ideal behaviors of other components in the signaling chain, collectively known as *implementation loss*.

[3]Transistor action was first observed on December 15, 1947, but the news of the invention was not made public until June 30, 1948.

Figure 1.5 shows the performance of selected rate $R = 1/2$ bits/symbol (binary) convolutional and turbo codes codes on an additive white Gaussian noise channel (see also [24, 20, 37, 38, 39]). Contrary to TCM, these codes do not preserve bandwidth and the gains in power efficiency in Figure 1.5 are partly obtained by a power bandwidth trade-off, i.e., the rate 1/2 convolutional codes require twice as much bandwidth as uncoded transmission. This bandwidth expansion may not be an issue in deep-space communications or the application of error control to spread spectrum systems [53, 54]. As a consequence, for the same complexity, a higher coding gain is achieved than with TCM. Note that the convolutional codes reach a point of diminishing returns.

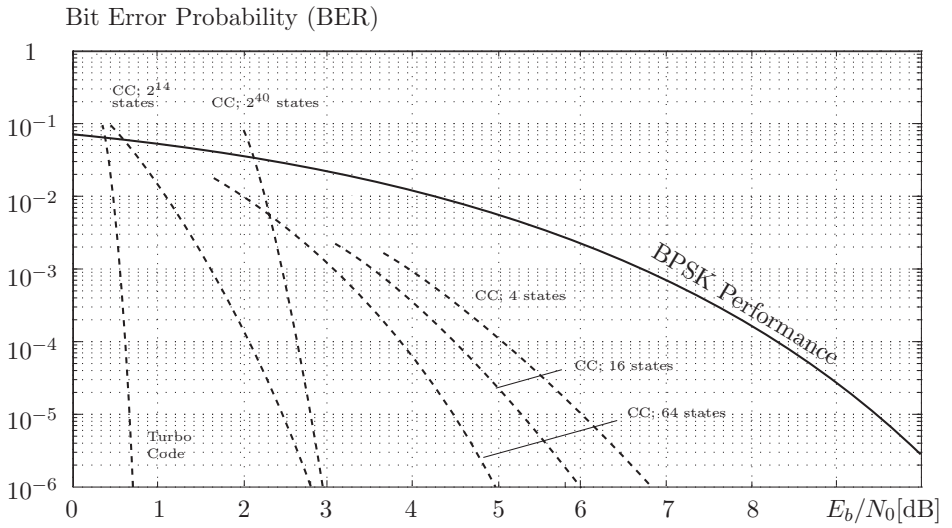

Figure 1.5: Bit error probability of selected rate $R = 1/2$ convolutional and turbo codes as a function of the normalized signal-to-noise ratio. The large-state space code is decoded sequentially, while the performance of all the other convolutional codes is for maximum-likelihood decoding. Simulation results are taken from [20], [38], and [4]. The Shannon limit is at E_b/N_0=0 dB, and that for BPSK is at $E_b/N_0 = 0.19$ dB.

For very low target error probabilities, tandem coding methods, called *concatenated coding*, have become very popular [24, 10, 23, 4]. In *classic concatenation*, as illustrated in Figure 1.6, the FEC codec is broken up into an inner and an outer code. The inner code is most often a trellis code which performs the channel error control, and the outer code is typically a high-rate Reed-Solomon (RS) block code. Its function it is to clean up the residual output error of the inner code. This combination has proven very powerful, since the error mechanism of the inner code is well matched to the error correcting capabilities

of the outer system. Via this tandem construction, very low error rates are achievable.

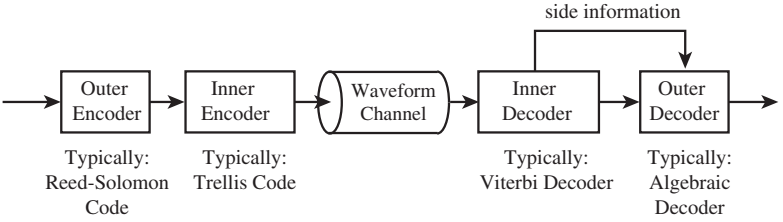

Figure 1.6: Classic concatenated FEC coding using inner and outer codecs.

With the discovery of turbo codes [4] new concatenated structures have appeared, shown in Figure 1.7. *Serial concatenation* is very similar to classic concatenation; however, the interleaver Π and deinterleaver Π^{-1} are fundamental structures, discussed later in Chapters 6 and 8. Furthermore, the outer code in these new structures is usually a very weak error control code such as a simple parity check code. Its function is not to clean up residual errors from the inner decoder, but, through interaction with the inner codes to form a strong error control system, a.k.a. a turbo code.

The original turbo codes, however, used a parallel arrangement of two codes, known as *parallel concatenation*, where both encoders have identical roles. Both arrangements, parallel and serial concatenation, are decoded with the same iterative decoding procedure which alternately invokes soft-output decoders for the two component codes. The structure and workings of these decoders is explored in detail in Chapters 5, 6, 8, and 9.

The field of error control and error correction coding naturally breaks into two disciplines, named somewhat inappropriately block coding and trellis coding. While block coding, which traditionally was approached as applied mathematics, has produced the bulk of publications in error control coding, trellis and turbo coding is favored in most practical applications. One reason for this is the ease with which soft-decision decoding can be implemented for trellis and turbo codes. Soft-decision is the operation whereby the demodulator does not make hard final decisions on the transmitted symbols or bits. Rather, it passes the received signal values directly on to the decoder which derives probabilistic information from those signals, which are then used to generate a final estimate of the decoded bits. There are no "errors" to be corrected; the decoder operates on reliability information obtained by comparing the received signals with the possible set of transmitted signals to arrive at a decision for an entire codeword. This soft-decision processing yields a 2 dB advantage on *additive white Gaussian noise* (AWGN) channels. In many applications the trellis decoders act as "S/N transformers" (e.g., [17]), improving the channel behavior as in concatenated coding. Ironically, many block codes can be de-

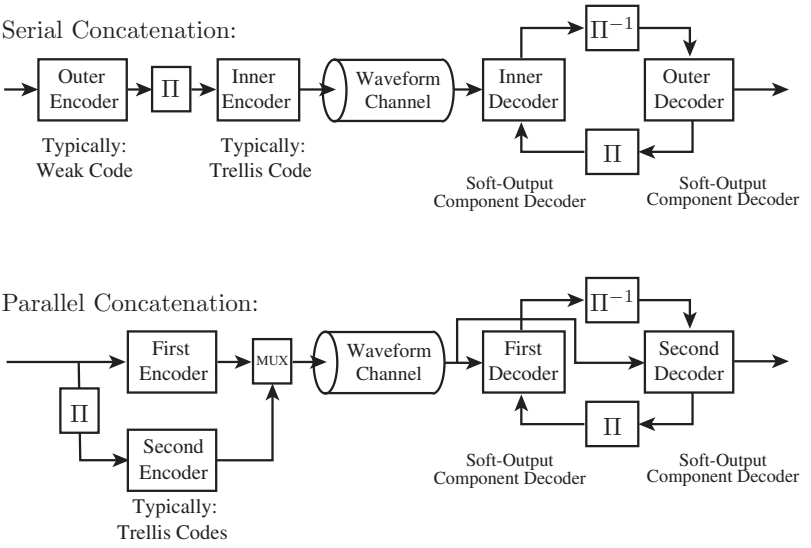

Figure 1.7: Modern concatenated FEC coding using two component codes.

coded very successfully using decoding methods developed for trellis codes (Chapter 4), smearing the boundaries between these two branches of error control coding.

1.5 Bandwidth, Power, and Complexity

Nyquist showed in 1928 [36] that a channel of bandwidth W (in Hz) is capable of supporting PAM signals at a rate of $2W$ samples/second without causing intersymbol interference. In other words, using Nyquist's method of interpolating band-limited functions, there are $2W$ independent *signal dimensions* per second[4]. If two carriers ($\sin(2\pi f_c)$ and $\cos(2\pi f_c)$) are used in quadrature, as in double side-band suppressed carrier amplitude modulation (DSB-SC), we have W pairs of dimensions (or complex dimensions) per second. This is the ubiquitous QAM format popular in digital radio systems (Chapter 2).

The parameter which characterizes how efficiently a system uses its allotted bandwidth is the bandwidth efficiency η, defined as

$$\eta = \frac{\text{Transmission Rate}}{\text{Channel Bandwidth } W} \quad [\text{bits}/s/\text{Hz}]. \tag{1.2}$$

[4]A more general analysis using more sophisticated interpolation functions and a more general definition of bandwidth reveals that there are approximately $2.4W$ independent signal dimensions per second [62].

Using Shannon's capacity formula (1.1) and dividing by W we obtain the maximum bandwidth efficiency for an additive white Gaussian noise channel, the *Shannon limit*, as

$$\eta_{\max} = \log_2\left(1 + \frac{S}{N}\right) \quad [\text{bits}/s/\text{Hz}]. \tag{1.3}$$

In order to calculate η, we must suitably define the channel bandwidth W. This is obvious for some signaling schemes, like Nyquist signaling, which have a rather sharply defined bandwidth (see Chapter 2), but becomes more arbitrary for modulation schemes with infinite spectral occupancy. One commonly used definition is the 99% bandwidth definition, i.e., W is defined such that 99% of the transmitted signal power falls within the band of width W. This 99% bandwidth corresponds to an out-of-band power of -20 dB.

The average signal power S can be expressed as

$$S = \frac{kE_b}{T} = RE_b, \tag{1.4}$$

where E_b is the energy per bit, k is the number of bits transmitted per symbol, and T is the duration of a symbol. The parameter $R = k/T$ is the transmission rate of the system in bits/s. Rewriting the signal-to-noise power ratio S/N, where $N = WN_0$, i.e., the total noise power equals the one-sided noise power spectral density (N_0) multiplied by the width of the transmission band, we obtain the Shannon limit in terms of the bit energy and noise power spectral density, given by

$$\eta_{\max} = \log_2\left(1 + \frac{RE_b}{WN_0}\right). \tag{1.5}$$

Since $R/W = \eta_{\max}$ is the limiting spectral efficiency, we obtain a bound from (1.5) on the minimum bit energy required for reliable transmission at a given spectral efficiency:

$$\frac{E_b}{N_0} \geq \frac{2^{\eta_{\max}} - 1}{\eta_{\max}}, \tag{1.6}$$

also called the *Shannon bound*.

If spectral efficiency is not at a premium, and a large amount of bandwidth is available for transmission, we may choose to use bandwidth rather than power to increase the channel capacity (1.1). In the limit as the signal is allowed to occupy an infinite amount of bandwidth, i.e., $\eta_{\max} \to 0$, we obtain

$$\frac{E_b}{N_0} \geq \lim_{\eta_{\max}\to 0} \frac{2^{\eta_{\max}} - 1}{\eta_{\max}} = \ln(2), \tag{1.7}$$

the absolute minimum bit energy to noise power spectral density required for reliable transmission. This minimum $E_b/N_0 = \ln(2) = -1.59$ dB.

We can cheat on the Shannon limit by allowing a certain number of errors in the following way: Assume that the original information source of rate R is being compressed into a rate $R' < R$. According to source coding theory, this introduces a distorting in the sense that original information can no longer be reconstructed perfectly [14]. If the source is binary, this compression results in a non-zero reconstruction bit error rate and satisfies $R' = R(1 - h(\text{BER}))$, where $h(p) = -p\log(p) - (1-p)\log(1-p)$ is the *binary entropy function*. As long as $R' < C$, the channel coding system will add no extra errors, and the only errors are due to the lossy compression. The source encoder has a rate of $1/(1 - h(\text{BER}))$, and consequently the average power is

$$S = \frac{RE_b}{1 - h(\text{BER})} \tag{1.8}$$

since less energy is used to transport a bit. The actual rate over the channel is R', from which we obtain a modified Shannon bound for non-zero bit error rates, given by

$$\frac{E_b}{N_0} \geq \frac{2^{(1-h(\text{BER})\eta_{\max}} - 1}{\eta_{\max}} (1 - h(\text{BER})). \tag{1.9}$$

This is the bound plotted in Figure 1.3 for a spectral efficiency of $\eta_{\max} = 2$ bits/s/Hz.

The implicit dependence of our formulas on the somewhat arbitrary definition of the bandwidth W is not completely satisfactory, and we prefer to normalize these formulas per signal dimension. Let R_d be the rate in bits/dimension, then the capacity of an additive white Gaussian noise channel per dimension is the maximum rate/dimension at which reliable transmission is possible. It is given by [62]

$$C_d = \frac{1}{2}\log_2\left(1 + 2\frac{R_d E_b}{N_0}\right) \quad [\text{bits/dimension}], \tag{1.10}$$

or as

$$C_c = \log_2\left(1 + \frac{RE_b}{N_0}\right) \quad [\text{bits/complex dimension}], \tag{1.11}$$

if we normalize to complex dimensions, in which case R is the rate per complex dimension. Both (1.10) and (1.11) can easily be derived from (1.1).

Applying similar manipulations as above, we obtain the Shannon bound normalized per dimension as

$$\frac{E_b}{N_0} \geq \frac{2^{2C_d} - 1}{2C_d}; \qquad \frac{E_b}{N_0} \geq \frac{2^{C_c} - 1}{C_c}. \tag{1.12}$$

Equations (1.12) are useful when the question of waveforms and pulse shaping is not a central issue, since it allows one to eliminate these considerations by working with signal dimensions, rather than the signals itself (see also Chapter 2). We will use (1.12) for our comparisons.

The Shannon bounds (1.3) and (1.12) relate the spectral efficiency η to the power efficiency E_b/N_0, and establish fundamental limits in the trade-off between the two primary resources of data communications; power and spectrum. These limits hold regardless of the signal constellation or coding method that is used for transmission.

Figure 1.8 shows (1.12)–as a solid line–as well as similar limits calculated for the cases where the transmitted signal constellations are restricted to BPSK, QPSK, 8-PSK, 16-PSK, and 16-QAM (see Chapter 2). As can be seen, these restrictions lead to various degrees of loss that has to be accepted when a particular constellation is chosen. The figure also shows the power and bandwidth efficiencies of some popular uncoded quadrature constellations as well as that of a number of error control coded transmission schemes. The trellis-coded modulation schemes used in practice, for example, achieve a power gain of up to 6 dB without loss in spectral efficiency, while the more powerful coded modulation methods such as trellis-turbo coding and block turbo coding provide even further gain. The binary coding methods achieve a gain in power efficiency, but at the expense of spectral efficiency with respect to the original signal constellation. The turbo-coded methods come extremely close to capacity for $\eta \leq 1$; we present such methods which can be made to approach the Shannon bound arbitrarily closely in Chapter 8.

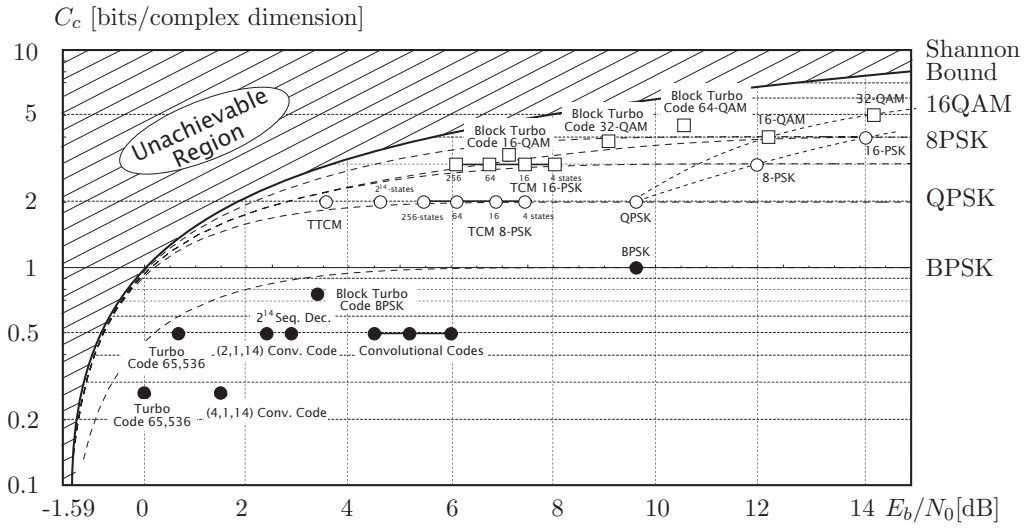

Figure 1.8: Theoretical spectral and power efficiency limits for various signal constellations and spectral efficiencies achieved by coded and uncoded transmission methods.

In all cases a rate reduction (traditional coding) or a signal set expansion (coded modulation) is required in order to give the coding system the required redundancy. This is also clearly obvious from the capacity curves for the different signal constellations.

In order to compare these different communications systems, we also need a parameter expressing the performance level of a system. This parameter is the information bit error probability P_b. For practical systems it typically falls into the range $10^{-3} \geq P_b \geq 10^{-6}$, though it is sometimes required to be lower, e.g., for digital TV, or much lower as for optical and cable modem systems where BER on the order of $P_b \leq 10^{-6}$ are required. The performance points in Figure 1.8 are drawn for $P_b = 10^{-5}$.

Other methods which combine coding with signal shaping exist and exhibit similar gains. For example, coded overlapped quadrature modulation (OC-QPSK) is a combined coding and controlled intersymbol interference method [43] which causes smaller amplitude variations than Nyquist signaling, and is therefore useful for systems with amplifier non-linearities, like satellite traveling wave tube (TWT) amplifiers. Continuous-phase modulation (CPM) is a constant-amplitude modulation format [1], which also has many similarities with TCM, and derives from frequency modulation aiming at improving the bandwidth efficiency by smoothing phase transitions between symbols. CPM has gone through an evolution similar to TCM, and the reader is referred to the book by Anderson, Aulin, and Sundberg [1], the standard reference on the subject. However, with the advent of highly linear amplifiers, CPM has lost one of its main selling points and has seen significantly less deployment then the linear modulation techniques.

The last, and somewhat hidden, player in the application of coding is complexity. While we have shown that power and bandwidth can be captured elegantly by Shannon's theory, measures of complexity are much more difficult to define. First there is what we might term *code complexity*. In order to approach the Shannon bound, larger and larger codes are required. In fact, Shannon et al. [46] proved the following *lower bound* on the codeword error probability P_B:

$$P_B > 2^{-n(E_{\text{sp}}(R)+o(N))}, \quad E_{\text{sp}}(R) = \max_{\boldsymbol{q}} \max_{\rho>1} (E_0(\boldsymbol{q}, \rho) - \rho R)). \tag{1.13}$$

The exponent $E_0(\boldsymbol{q}, \rho)$ is called the Gallager exponent and depends on the symbol probability distribution $\boldsymbol{q}$ and the optimization parameter ρ, it is discussed in more detail in Chapter 5, as well as in Gallager's book on information theory [13]. The most important point to notice is that this bound is exponential in the codelength n.

The bound is plotted for rate $R = 1/2$ in Figure 1.9 for BPSK modulation [44], together with selected turbo-coding schemes and classic concatenated methods. The performance of codes for various lengths follows the tendency of the bound, and we see a diminishing return as the code size exceeds $n \approx 10^4 - 10^5$, beyond which only small gains are possible. This is the reason why most practical applications of large codes target block sizes no larger than this. On the other hand, codes cannot be shortened much below $n = 10^4$ without

a measurable loss in performance, which must be balanced against the overall system performance. Implementations of near-capacity error control systems therefore have to process blocks of 10,000 symbols or more, requiring appropriate storage and memory management.

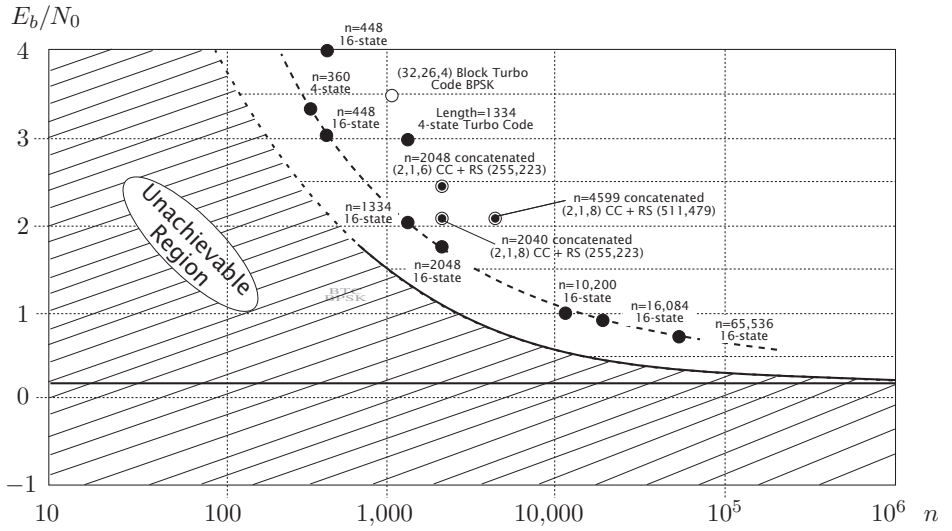

Figure 1.9: Block error rate P_B and sphere-packing lower bound for rate $R = 1/2$ coded example coding systems using BPSK. Turbo codes and selected classic concatenated coding schemes are compared. The solid line is the Shannon limit at $E_b/N_0 = 0.19$ dB at $R = 0.5$.

The other component of our complexity consideration is the *computational complexity*, that is, the amount of processing that has to be performed to decode a codeword. This is notoriously difficult to measure, and parameters like the code state space or number of multiplications may not be relevant. What ultimately counts is the size and power consumption of a VLSI implementation, which is very much technology, architecture, and design dependent. It suffices to say that what a coding theorist considers complex may not be complex for a VLSI circuit designer. An example of this is the "folk theorem" that an APP trellis decoder (Chapter 5) is about four times as complex as a Viterbi decoder for the same code. Some circuit designers would argue that the trace-back implementation of the Viterbi decoder easily compensates for the more complex arithmetic of the APP decoder and that there is no real difference between the two in terms of implementation complexity.

The situation of turbo codes is even more complicated to evaluate, since most of the complexity of the decoder resides in the storage requirements of the large blocks that need to be processed and the efficient management of the this memory. The computational units of a turbo decoder take up less than 10% of the VLSI area. Processing large blocks, however, is inherent in effective error control coding as discussed above.

More recently decoding circuits based on analog processing have emerged [26, 34, 57], using subthreshold CMOS and bipolar multiplier circuits, both based on variations of the well-known *Gilbert cell* [31]. These circuits are ideally matched to the arithmetic requirements of soft-decision decoders, making use of the fundamental exponential transfer characteristics of the transistors. Digital information is processed by small currents in the transistor's off-region, representing probabilities of the various discrete random variables. Contrary to mainstream digital VLSI implementations where the component transistors are used as on-off switches, and are optimized for that operation, analog processing operates the transistors in their off-position, and all computations are accomplished by what digital technology calls leakage currents. Analog decoding is also different from conventional analog processing in that digital information is processed. This leads to a surprising robustness of the analog decoder to typical analog impairments such as current mismatch, body effects, etc.

Several such decoders have already been successfully fabricated and tested, or extensively simulated by a few research groups around the world, see, e.g., [58, 59]. The initial idea was that subthreshold operation would result in substantial power savings of the decoding operation. While operating the transistor in the subthreshold region has indeed a strong potential to save power (see, e.g., [6, 7]), the story regarding analog decoders is more complex.

Zargham et al. [65] for example show that the current mirror circuit which is the base of the Gilbert multiplier used in most of the analog designs is increasingly susceptible to gate threshold voltage variations in the transistors that are paired to make the current mirror. As the fabrication process shrinks, small variations in the gate size translate into exponentially varying current errors in the Gilbert cells. This ultimately leads to a breakdown of the functionality of the decoder as a whole. Zargham et al. [65] argue that analog decoders in sub-100 nm processes will require such large oversizing of the transistors that going to smaller processes is not productive.

In an investigation by Winstead and Schlegel [42], the computational and energy requirements of message passing decoders as discussed in Chapters 6 and 9 are examined from fundamental viewpoints of computational theory and minimum-energy switching principles. Since advanced message-passing decoders have a computational complexity which is linear in the number of bits decoded, a direct energy cost measure per decoded bit can be computed, quite irrespective of the actual code or the code family being used. This leads to predictions of the energy cost per decoded message bit as miniaturization processes continue to advance to ever smaller scales. Figure 1.10 shows the energy per-

formance attained by a number of message-passing decoders that were built by several research and industrial groups. What is evident is that the process miniaturization is the major driver for a reduction in the energy per bit. The ultimate limit for charged-based computation is based on a conceptual single-charge device [66], and would attain a decoding energy consumption anywhere between sub-fempto-joule to about 10 fJ/decoded bit, depending on code and implementation parameters.

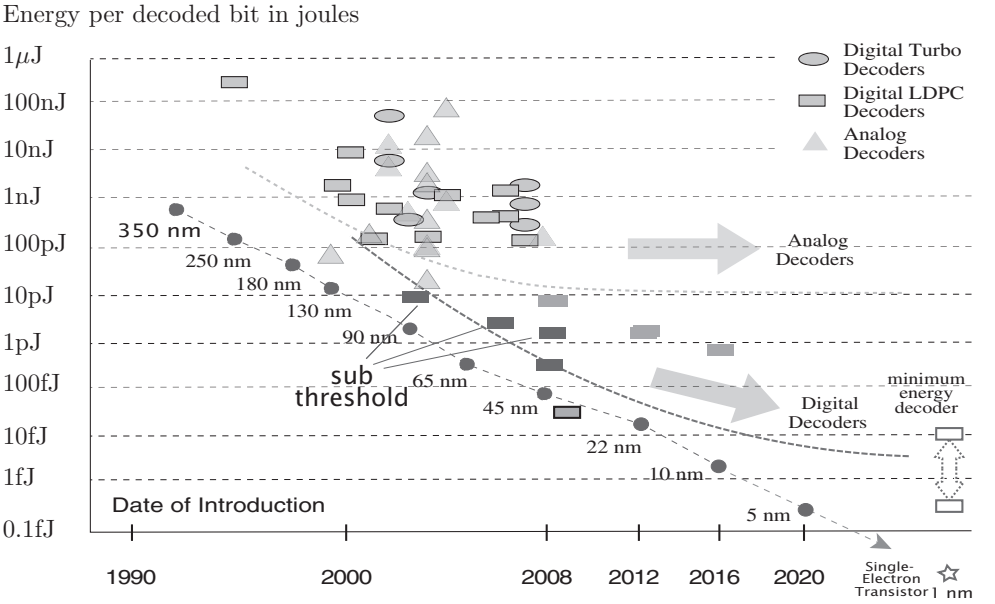

Figure 1.10: Energy consumption of various message-passing decoders and predictions for future developments in joules per decoded bit.

While predictions are notoriously difficult to make with accuracy, it appears that analog decoders based on Gilbert-cell technology would flatten out at around 10 pJ/decoded bit, primarily due to the gate-threshold variations discussed above [65]. Digital decoders, on the other hand, can benefit from the miniaturization, and the calculations in [42] indicate that a minimal energy cost for a digital implementation would be around $30{,}000kT$, where k is Boltzmann's constant, and T is the temperature. This translates to about 0.1 femptojoules. Any further reductions in power consumption would need to rely on speculative technologies such as adiabatic computation and/or quantum computing.

1.6 A Brief History–The Drive Towards Capacity

Forward error control (FEC) coding celebrated its first success in the application of convolutional codes to deep-space probes in the 1960's and 1970's, and for quite a while afterwards, FEC was considered an intellectual curiosity with deep-space communications as its only viable practical application. Deep space communications is a classical case of power-limited communications, and it serves as a picture book success story of error control coding.

If we start with uncoded binary phase-shift keying (BPSK) as our baseline transmission method (see Chapter 2) and assume coherent detection, we can achieve a bit error rate of $P_b = 10^{-5}$ at a bit energy-to-noise power ratio of $E_b/N_0 = 9.6$ dB and at a spectral efficiency of ideally 1 bit/dimension. From the Shannon limit in Figure 1.11, it can be seen that 1 bit/dimension is theoretically achievable with $E_b/N_0 = 1.76$ dB, indicating that a power savings of nearly 8 dB is possible by applying proper coding. 8 dB is an over 6-fold savings in transmit power, antenna size, or other aspect directly linked to the received signal power, as, for example, 2.5 times the transmission distance.

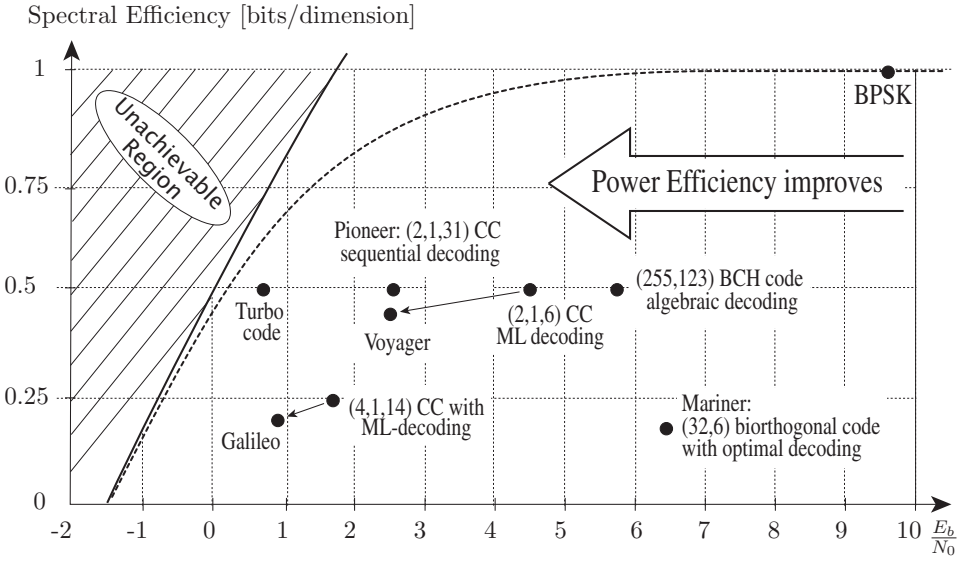

Figure 1.11: Some milestones in the drive towards channel capacity achieved by the space systems which evolved over the past 50 years as answer to the Shannon capacity challenge.

One of the earliest attempts to close the signal energy gap to the Shannon limit was the use of a rate 6/32 biorthogonal (Reed-Muller) block code [27]. This code was used

on the Mariner Mars and Viking missions in conjunction with BPSK and soft-decision maximum-likelihood decoding. This system had a spectral efficiency of 0.1875 bits/symbol and achieved a bit error rate of $P_b = 10^{-5}$ with an $E_b/N_0 = 6.4$ dB. Thus, the (32,6) biorthogonal code required 3.2 dB less power than BPSK at the cost of a five-fold increase in the bandwidth. The performance of the 6/32 biorthogonal code is plotted in Figure 1.11, as are all the other systems discussed below.

In 1967, a new algebraic decoding technique was discovered for the popular Bose-Chaudhuri-Hocquenghem (BCH) codes [2, 28]. This new algorithm enabled the efficient hard-decision decoding of an entire class of block codes–for example, the (255,123) BCH code, which has a code rate of $R_d \approx 0.5$ bits/symbol and achieves a BER of $P_b = 10^{-5}$ at $E_b/N_0 = 5.7$ dB using algebraic decoding.

The next step was taken with the introduction of sequential decoding (see Chapter 5), which could make use of soft-decision decoding. Sequential decoding allowed the decoding of long constraint-length convolutional codes, and it was first used on the Pioneer 9 mission [11]. The Pioneer 10 and 11 missions in 1972 and 1973 both used a long constraint-length (2,1,31), nonsystematic convolutional code (Chapter 4) [29]. A sequential decoder was used which achieved $P_b = 10^{-5}$ with $E_b/N_0 = 2.5$ dB, and $R_d = 0.5$. This is only 2.5 dB away from the capacity of the channel.

Sequential decoding has the disadvantage that the computational load is variable, and this load grows exponentially the closer the operation point moves towards capacity (see Chapter 4). For this and other reasons, the next generation of space systems employed maximum-likelihood decoding. The Voyager spacecraft, launched in 1977, used a short constraint-length (2,1,6) convolutional code in conjunction with a soft-decision Viterbi decoder achieving $P_b = 10^{-5}$ at $E_b/N_0 = 4.5$ dB and a spectral efficiency of $R_d = 0.5$ bits/symbol. The biggest Viterbi decoder built to date [8] found application in the Galileo mission, where a (4,1,14) convolutional code is used, yielding a spectral efficiency of $R_d = 0.25$ bits/symbol at $P_b = 10^{-5}$ and $E_b/N_0 = 1.75$ dB. The performance of this system is 2.5 dB away from the capacity limit. The systems for Voyager and Galileo are further enhanced by the use of concatenation in addition to the convolutional inner code. An outer (255,223) Reed-Solomon code [27] is used to reduce the required signal-to-noise ratio by 2.0 dB for the Voyager system and by 0.8 dB for the Galileo system.

More recently, Turbo-codes [3] using iterative decoding have virtually closed the capacity gap by achieving $P_b = 10^{-5}$ at a spectacularly low E_b/N_0 of 0.7 dB with $R_d = 0.5$ bits/symbol, and longer turbo codes come even closer to capacity, e.g., [48]. It is probably appropriate to say that the half-century effort to reach capacity has been achieved with this latest invention, in particular in the regime of lower spectral efficiencies. With turbo coding then, another quite unexpected step of about 2 dB right to the capacity limit was made possible. Newer space communications systems virtually all use turbo or low-density parity-check codes discussed in this book–see also Chapter 9.

Space applications of error control coding have met with spectacular success, and

for a long time the belief that coding was useful only for improving power efficiency of digital transmission was poplar. This attitude was thoroughly overturned by another spectacular success of error control coding, this time for applications of data transmission over voiceband telephone channels. Here it was not the power efficiency which was the issue, but rather spectral efficiency, i.e., given a standard telephone channel with a fixed bandwidth and SNR, what was the maximum practical rate of reliable transmission?

The first commercially available voiceband modem in 1962 achieved a transmission rate of 2400 bits/s. Over the next 10 to 15 years these rates improved to 9600 bits/s, which was then considered to be the maximum achievable rate, and efforts to push the rate higher were frustrated. Ungerböck's invention of trellis-coded modulation in the late 1970's, however, opened the door to further, unexpected improvements. The modem rates jumped to 14,400 bits/s and then to 19,200 bits/s using sophisticated TCM schemes [12]. The latest chapter in voiceband data modems is the establishment of the CCITT V.34 modem standard [15, 9]. The modems specified therein achieve a maximum transmission rate of 28,800 bits/s, and extensions to V.34 to cover two new rates at 31,200 bits/s and 33,600 bits/s have been specified. However, at these high rates, modems operate successfully only on a small percentage of the connections. It seems that the limits of the voiceband telephone channel have been reached (according to [56]). This needs to be compared to estimates of the channel capacity for a voiceband telephone channel, which are somewhere around 30,000 bits/s. The application of TCM was one of the fastest migrations of an experimental laboratory system to an international standard (V.32 - V.34) [5, 15, 9]. The trellis codes used in these advanced modems are discussed in treated in detail in Chapter 3.

In many ways the telephone voiceband channel is an ideal playground for the application of error control coding. Its limited bandwidth of about 3 kHz (400 Hz - 3400 Hz) implies low data rates by modern standards. It therefore provides an ideal experimental field for high-complexity error control methods, which can be implemented without much difficulty using current DSP technology. It is thus not surprising that coding for voiceband channels was the first successful application of bandwidth efficient error control.

Nowadays, trellis coding in the form of bandwidth-efficient TCM as well as more conventional convolutional coding and higher-order modulation turbo-coding systems are used for satellite communications, both geostationary and low-earth orbiting satellites, for land-mobile and satellite-mobile services, for cellular communications networks, personal communications services (PCS), high-frequency (HF) tropospheric long-range communications, and cable modems, and increasingly also for high-speed optical communications systems. The Shannon capacity is now routinely the goal that is targeted for high-efficiency communications systems. As an example thereof, the IEEE 802.3an standard for Ethernet communications over twisted pair copper cables is discussed in more detail in Chapter 6. To us it seems clear that the age of widespread application of error control coding is upon us and every efficient communications systems will entail some form of FEC.

Bibliography

[1] J.B. Anderson, T. Aulin, and C-E. Sundberg, *Digital Phase Modulation*, Plenum Press, New York, 1986.

[2] E.R. Berlekamp, *Algebraic Coding Theory,* McGraw-Hill, New York, 1968.

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

[4] C. Berrou and A. Galvieux, "Near optimal error-correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.,* vol. COM-44, no. 10, pp. 1261-1271, Oct. 1996.

[5] U. Black, *The V Series Recommendations, Protocols for Data Communications Over the Telephone Network*, McGraw-Hill, New York, 1991.

[6] D. Bol, R. Ambroise, D. Flandre, and J.-D. Legat, "Interests and limitations of technology scaling for subthreshold logic," *IEEE Very Large Scale Integration (VLSI) Syst.*, vol. 17, no. 10, pp. 1508-1519, Oct. 2009.

[7] D. Bol, "Pushing ultra-low-power digital circuits into the nanometer era," Ph.D. thesis, Université catholique de Louvain, École Polytechnique de Louvain, Département d'Électricité, Dec. 2008.

[8] O.M. Collins, "The subtleties and intricacies of building a constraint length 15 convolutional decoder," *IEEE Trans. Commun.,* vol. COM-40, pp. 1810–1819, 1992.

[9] G.D. Forney, L. Brown, M.V. Eyuboglu, J.L. Moran III, "The V.34 high-speed modem standard," *IEEE Commun. Mag.*, pp. 28–33, Dec. 1996.

[10] G.D. Forney, *Concatenated Codes*, MIT Press, Cambridge, Mass., 1966.

[11] G.D. Forney, "Final report on a study of a sample sequential decoder," Appendix A, Codex Corp., Watertown, MA, U.S. Army Satellite Communication Agency Contract DAA B 07-68-C-0093, April 1968.

[12] G.D. Forney, "Coded modulation for bandlimited channels," *IEEE Information Theory Society Newsletter*, Dec. 1990.

[13] R.G. Gallager, *Information Theory and Reliable Communications*, John Wiley & Sons, New York, 1968.

[14] R.M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, Dordrecht, fourth printing, 1997.

[15] CCITT Recommendations V.34.

[16] M.J.E. Golay, "Notes on digital coding," *Proc. IEEE*, vol. 37, p. 657, 1949.

[17] J. Hagenauer and P. Höher, "A Viterbi algorithm with soft-decision outputs and its applications," *Proc. IEEE Globecom'89*, 1989.

[18] J. Hagenauer, "Rate compatible punctured convolutional codes (RCPC-codes) and their application," *IEEE Trans. Commun.,* vol. COM-36, pp. 389–400, April 1988.

[19] R.W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.,* vol. 29, pp. 147–160, 1950.

[20] J.A. Heller and J.M. Jacobs, "Viterbi detection for satellite and space communications," *IEEE Trans. Commun. Technol.*, COM-19, pp. 835–848, Oct 1971.

[21] H. Imai et al. *Essentials of Error-Control Coding Techniques*, Academic Press, New York, 1990.

[22] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[23] K.Y. Lin and J. Lee, "Recent results on the use of concatenated Reed-Solomon/ Viterbi channel coding and data compression for space communications," *IEEE Trans. Commun.,* vol. COM-32, pp. 518–523, 1984.

[24] S. Lin and D.J. Costello, Jr., *Error Control Coding: Fundamentals and Applications,* Prentice Hall, Englewood Cliffs, NJ, 1983.

[25] S. Lin, D.J. Costello, Jr., and M.J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.,* vol. 22, pp. 5–17, 1984.

[26] H.-A. Loeliger, F. Lustenberger, M. Helfenstein, and F. Tarköy, "Probability propagation and decoding in analog VLSI," *IEEE Trans. Inform. Theory*, pp. 837–843, Feb. 2001.

[27] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error Correcting Codes*, North Holland, New York, 1988.

[28] J.L. Massey, "Shift register synthesis and BCH decoding," *IEEE Trans. Inform. Theory,* vol. IT-15, pp. 122–127, 1969.

[29] J.L. Massey and D.J. Costello, Jr., "Nonsystematic convolutional codes for sequential decoding in space applications," *IEEE Trans. Commun. Technol.,* vol. COM-19, pp. 806–813, 1971.

[30] J.L. Massey, "Coding and modulation in digital communications," *Proc. Int. Zürich Sem. Digital Commun.*, Zürich, Switzerland, March 1974, pp. E2(1)–E2(4).

[31] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA,1989.

[32] H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Vol. 1, John Wiley & Sons, New York, 1990.

[33] H. Meyr, M. Moeneclaey, and S.A. Fechtel, *Digital Communication Receivers*, John Wiley & Sons, New York, 1998.

[34] M. Moerz, T. Gabara, R. Yan, and J. Hagenauer, "An analog .25 μm BiCMOS tailbiting MAP decoder," *IEEE Proc. International Solid-State Circuits Conf.*, pp. 356–357, San Francisco, Feb. 2000.

[35] M. Mouly and M-B. Pautet, *The GSM System for Mobile Communications*, sold by the authors, ISBN 2-9507190-0-7. 1993.

[36] H. Nyquist, "Certain topics in telegraph transmission theory," *AIEE Trans.*, pp. 617 ff., 1946.

[37] J.P. Odenwalder, "Optimal decoding of convolutional codes," Ph.D. thesis, University of California, LA, 1970.

[38] J.K. Omura and B.K. Levitt, "Coded error probability evaluation for antijam communication systems," *IEEE Trans. Commun.,* vol. COM-30, pp. 896–903, May 1982.

[39] J.G. Proakis, *Digital Communications,* McGraw-Hill, New York, 1989.

[40] S. Ramseier and C. Schlegel, "On the bandwidth/power tradeoff of trellis coded modulation schemes," *Proc. IEEE Globecom'93* (1993).

[41] S.A. Rhodes, R.J. Fang, and P.Y. Chang, "Coded octal phase shift keying in TDMA satellite communications," *COMSAT Tech. Rev.*, vol. 13, pp. 221–258, 1983.

[42] C. Schlegel and C. Winstead, "From mathematics to physics: Building efficient iterative error control decoders," *International Symposium on Turbo Coding and Iterative Information Processing*, Sweden, Aug. 2012.

[43] C. Schlegel, "Coded overlapped quadrature modulation," *Proc. Global Conf. Commun.* GLOBECOM'91, Phoenix, AZ, Dec. 1991.

[44] C. Schlegel and L.C. Perez, "On error bounds and turbo codes," *IEEE Commun. Lett.*, vol. 3, no. 7, July 1999.

[45] C.E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.,* vol. 27,, pp. 379-423, July 1948.

[46] C.E. Shannon, R.G. Gallager, and E.R. Berlekamp, "Lower bounds to error probabilities for coding on discrete memoryless channels," *Inform. Contr.,* vol. 10, pt. I, pp. 65–103, 1967, Also, *Inform. Contr.,* vol. 10, pt. II, pp. 522-552, 1967.

[47] TIA/EIA/IS-95 interim standard, mobile station–base station compatibility standard for dual-mode wideband spread spectrum cellular systems, Telecommunications Industry Association, Washington, D.C., July 1993.

[48] S. tenBrink, "A rate one-half code for approaching the Shannon limit by 0.1 dB," *Electron. Lett.,* vol. 36, no. 15, pp. 1293–1294, July 2000.

[49] G. Ungerboeck, J. Hagenauer, and T. Abdel-Nabi, "Coded 8-PSK experimental modem for the INTELSAT SCPC system," *Proc. ICDSC, 7th*, pp. 299–304, 1986.

[50] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory,* vol. IT-28, no. 1, pp. 55-67, Jan. 1982.

[51] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets part I: Introduction," *IEEE Commun. Mag.,* vol. 25, no. 2, pp. 5–11, Feb. 1987.

[52] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets part II: State of the art," *IEEE Commun. Mag.,* vol. 25, no. 2, pp. 12–21, Feb. 1987.

[53] A.J. Viterbi, "Spread spectrum communications–myths and realities," *IEEE Commun. Mag.,* vol. 17, pp. 11-18, May 1979.

[54] A.J. Viterbi, "When not to spread spectrum–a sequel," *IEEE Commun. Mag.,* vol. 23, pp. 12-17, April 1985.

[55] F.-Q. Wang and D.J. Costello, "Probabilistic construction of large constraint length trellis codes for sequential decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, Sept. 1995.

[56] R. Wilson, "Outer limits," *Electronic News*, May 1996.

[57] C. Winstead, J. Dai, C. Meyers, C. Schlegel, Y.-B. Kim, W.-J. Kim, "Analog MAP decoder for (8,4) Hamming code in Subthreshold CMOS," *Advanced Research in VSLI Conference ARVLSI*, Salt Lake City, March, 2000.

[58] C. Winstead, J. Dai, S. Yu, C. Myers, R. Harrison, and C. Schlegel, "CMOS analog MAP decoder for (8,4) Hamming code," *IEEE Journal of Solid State Circuits*, vol. 29, no. 1, Jan. 2004.

[59] C. Winstead and C. Schlegel, "Importance Sampling for SPICE-level verification of analog decoders", *International Symposium on Information Theory (ISIT'03)*, Yokohama, June, 2003.

[60] A.P. Worthen, S. Hong, R. Gupta, W.E. Stark, "Performance optimization of VLSI transceivers for low-energy communications systems," *IEEE Military Communications Conference Proceedings, MILCOM 1999*, vol. 2, pp. 1434–1438,1999.

[61] J.M. Wozencraft and R.S. Kennedy, "Modulation and demodulation for probabilistic coding", *IEEE Trans. Inform. Theory,* vol. IT-12, no. 3, pp. 291–297, July, 1966.

[62] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, New York, 1965, reprinted by Waveland Press, 1993.

[63] 3GPP Group Website, Releases, *http://www.3gpp.org/releases.*

[64] 3GPP2 Group Website, *http://www.3gpp2.org.*

[65] M. Zargham, C. Schlegel, J. P. Chamorroc, C. Lahuecc, F. Seguinc, M. Jézéquelc, and V. Gaudet, "Scaling of analog LDPC decoders in sub-100 nm CMOS processes," Integration—the VLSI J., vol. 43, no. 4, Sept. 2010, pp. 365–377.

[66] V. Zhirnov R. Cavin, III, J. Hutchby, and G. Bourianoff, "Limits to binary logic switch scaling: A gedanken model," *Proceedings of the IEEE*, vol. 91, no. 11, Nov. 2003.