Defence Science Journal, Vol 43, No 1, January 1993, pp 59-69

Coding Theory and its Applications in Communication Systems*

Vijay K. Bhargava, Qing Yang and David J. Peterson

Department of Electrical and Computer Engineering, University of Victoria
PO Box 3055, Victoria, BC, Canada V8W 3P6.

ABSTRACT

Error control coding has been used extensively in digital communication systems because of its cost-effectiveness in achieving efficient, reliable digital transmission. Coding now plays an important role in the design of modern communication systems. This paper reviews the development of basic coding theory and state-of-art coding techniques. The applications of coding to communication systems and future trends are also discussed.

1. INTRODUCTION

1.1 The Coding Problem

Error control coding is concerned with methods of delivering information from a source to a destination with a minimum of errors. Error control coding can be categorized as forward error correction (FEC), automatic repeat request (ARQ), or as a combination of FEC and ARQ (hybrid).

The communication system depicted in Fig. 1 employs FEC. The source generates data bits or messages that must be transmitted to a distant user over a noisy channel. Generally speaking, a specific signal is assigned to each of M possible messages that can be emitted by the source. The selection rule that assigns a transmitted signal to each message is the code. The encoder implements the selection rule, while the decoder performs the corresponding inverse mapping. Because of channel noise, the transmitted signals may not arrive at the receiver exactly as transmitted, causing errors to occur at the decoder input. A natural design objective is to select the code such that most of the errors occuring at the decoder input can be corrected by the decoder, thereby providing an acceptable level of reliabilty.

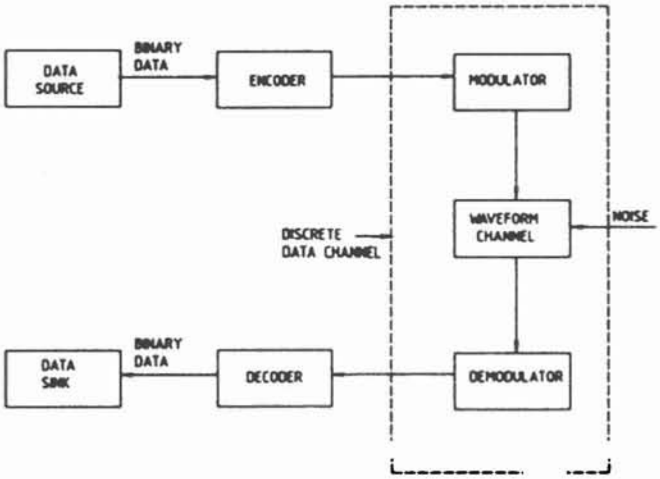

Figure 1. Digital communication using forward error control coding.

Coding is a design technique which can fundamentally change the trade-offs in a digital communication system. The most trivial example of coding is the repetition of the same message on the transmission channel. Here it is clear that redundancy, and therefore reliability, is obtained at the expense of transmission efficiency, or bandwidth utilization. In general, error control coding can increase signal quality from problematic to acceptable levels. If the attendant

Received 14 September 1992

* Research supported by a grant from Canadian Institute for Telecommunications Research under the NCE program of the Government of Canada.

increase in complexity at the transmitter and receiver is economically viable, and bandwidth utilization is not unduly compromised, useful performance improvements may result. For example, with coding, less power may be required to communicate between a satellite and a mobile terminal. Furthermore, coding may result in an increase in the maximum number of mobile terminals per satellite

The study of error control coding began in 1948 with Claude Shannon[1] who demonstrated the existence of codes achieving reliable communication whenever the code rate is smaller than a threshold C called the channel capacity. For the additive white Gaussian noise (AWGN).channel, the channel capacity is given by

$$C = B \log_2 \left[1 + \frac{S}{N} \right]$$

Where B is the channel bandwidth, and S/N is the ratio of signal to noise power falling within the bandwidth. This remarkable result indicates that the ultimate performance limit caused by channel noise is not reliability, as generally believed before Shannon's work, but the rate at which data can be reliably transmitted.

The concept of channel capacity is fundamental to communication theory and is surprisingly powerful and general. It can be applied to a large class of channel models, whether memoryless or not, discrete or nondiscrete. However, Shannon's celebrated coding theorems are only existence theorems; they do not show how promising coding schemes can be constructed. Since the publication of Shannon's result, a considerable amount of research has addressed the design and analysis of practical coding and decoding techniques permitting reliable communication at the data rates promised by the theory [2-4].

1.2 Basic Coding Process

In addition to the FEC/ARQ categorisation mentioned earlier, coding systems have traditionally been separated into block and convolutional error-correction techniques.

In an (n, k) linear block code, a sequence of k information bits is used to obtain a set of n-k parity bits, yielding an encoded block of n bits. Usually modulo–2 arithmetic is used to compute the parity bits. Modulo–2 arithmetic is particularly suited to digital logic; addition corresponds to the EXCLUSIVE–OR operation, while multiplication can be realised as an AND operation. The code rate r is defined as $r = k/n$ where n is called the block length. Linear codes form a linear vector space; two code words can be added (modulo–2) to produce a third code word.

The Hamming weight of a code word c is defined to be the number of nonzero components of c. For example, the code word c =(110101) has a Hamming weight of 4. The Hamming distance between two code words c_1 and c_2, denoted $d(c_1, c_2)$, is the number of positions in which they differ. For example if c_1 = (110101) and c_2 = (111000) then $d(c_1,c_2)$ = 3. The minimum distance d of a linear block code is defined to be the minimum weight of its nonzero code words. A code can correct all patterns of t or fewer random errors and detect all patterns having no more than s errors, provided that $s+2t+1 \leq d$. If the code is used for error correction alone, any pattern of t or fewer random errors can be corrected, provided that $2t+1 \leq d$.

A convolutional code of rate $1/v$ may be generated by a K stage shift register and v modulo–2 adders. Information bits are shifted in at the left, and for each information bit the output of the modulo–2 adders provides two channel bits. The constraint length of the code is defined as the number of shifts over which a single information bit can influence the encoder output. For the simple binary convolutional code, the constraint length is equal to K, the length of the shift register.

Whether block coding or convolutional coding is used, the encoded sequence is mapped to suitable waveforms by the modulator and transmitted over the noisy channel. The physical channel or the waveform channel consists of all the hardware (for example, filtering and amplification devices) and the physical media that the waveform passes through, from the output of the modulator to the input of the demodulator.

The demodulator estimates which of the possible symbols was transmitted based upon an observation of the received signal. Finally, the decoder estimates the transmitted information sequence from the demodulator output. The decoder makes use of the fact that the transmitted sequence is composed of the code words. Transmission errors are likely to result in reception of a noncode sequence.

1.3 Coding Gain

It is often useful to express coding performance not in terms of the error rate reduction for a given signal-to-noise ratio (SNR), but as the SNR difference at a fixed bit error rate. Consider an AWGN channel with one-sided noise spectral density N_0 having no bandwidth restriction. Let E_b denote the received energy per bit. It can be shown that if the SNR E_b/N_0 exceeds -1.6 dB, there exists a coding scheme which allows error-free communications, while reliable communication is not generally possible at lower SNRs. On the other hand, it is well-known that the uncoded phase shift keying (PSK) modulation over the same channel requires about 9.6 dB to achieve a bit error rate of 10^{-5}. Thus, as shown in Fig. 2, a potential coding gain of 11.2 dB is theoretically possible.

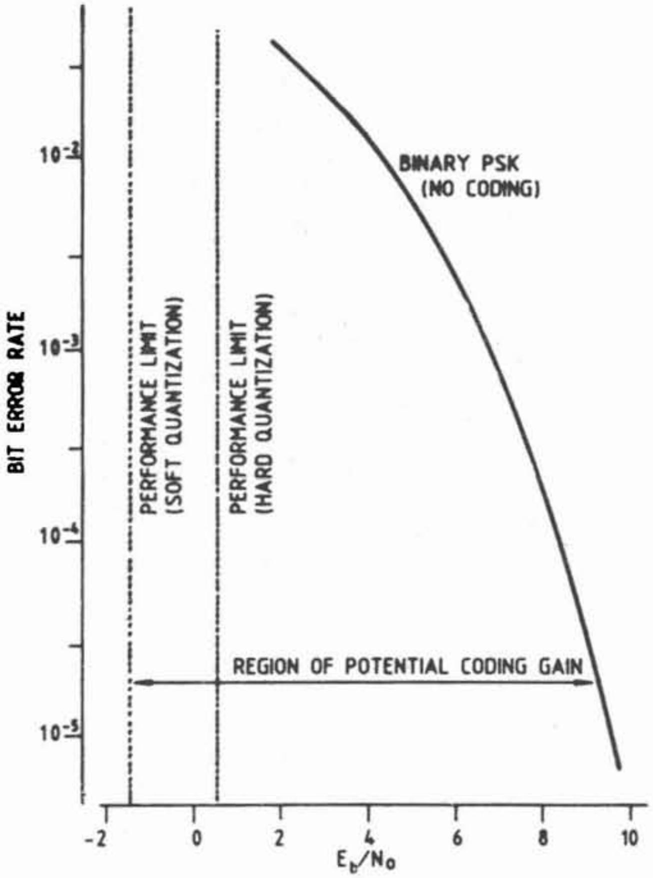

Figure 2. Performance of uncoded PSK over AWGN channel.

Coding gain is defined as the difference in value of E_b/N_0 required to attain a particular error rate with and without coding. Notice that coding gain is obtained at the expense of transmission bandwidth. The bandwidth expansion is the reciprocal of the code rate. Coding schemes delivering 2 to 8 dB coding gain are widely used in modern digital communication systems. This is because of the phenomenal decrease in the cost of digital hardware and the much less significant decrease in the cost of analog components such as power amplifiers, antennas and so on.

Practical communication systems rarely provide the ablity to make full use of the actual analog voltages of the received signal. The normal practice is to quantize these voltages. If binary quantization is used, we say that a hard decision is made at the receiver as to which level was actually sent. For example, in coherent PSK with equally likely transmitted symbols, the optimum threshold is zero. The demodulator output is a one or a zero depending on whether the voltage is above or below the threshold. With coding, it is desirable to maintain an indication of the reliability of this decision. A soft-decision demodulator first decides whether the voltage is above or below the decision threshold, and than computes a 'confidence' number which specifies how far from the decision threshold the demodulator output is. This number in theory could be an analog quantity, but in most practical applications a three-bit (eight-level) quantization is used. It is known that soft decision decoding is about 3 dB more efficient than hard decision decoding at very high E_b/N_0. A figure of 2 dB is more likely at realistic values of E_b/N_0.

2. CODING FOR DIGITAL COMMUNICATIONS

2. Block Codes and their Decoding

The basic idea behind all block codes is illustrated by the following example. We consider a binary code having the eight code words (000000), (001101), (010011), (011110), (100110), (101011), (110101) and (111000). These codes words form a vector space of dimension three, so the code is a (6, 3) linear code. The minimum weight of the seven nonzero code words is 3, so the minimum distance is 3. Thus, the code is a single error correcting code. This code is said to be in systematic form; the first three bits of any code word can be considered as message bits while the last three bits, which are uniquely determined by the first three bits, are the redundant or parity bits.

Many of the important block codes found to date are so-called cyclic codes or are closely related to cyclic codes. For such codes, if an n tuple $c = (c_0, c_1, c_2, \ldots,$

c_{n-1} is a code word, then the n tuple $c' = (c_{n-1}, c_0, c_1, \ldots, c_{n-2})$, obtained by cyclically shifting c one place to the right, is also a code word. This class of codes can be easily encoded using simple feedback shift register circuits. Furthermore, because of their inherent algebraic structure, the decoding of cyclic code is straightforwarJ, both conceptually and in practice. Examples of cyclic and related codes include the Bose-Chaudhuri- Hocquenhem (BCH), Reed-Solomon (RS), Hamming, maximum-length, maximum-distance-separable (MDS), Reed-Muller, Golay, Fire, difference set, quadratic residue, Goppa, and quasicyclic codes. Some of these classes form overlapping sets. For example, RS code are a special class of BCH codes and also belong to the class of MDS codes. The details of these codes can be found in any one of the standard coding references [5-8].

The first ste of the decoding procedure involves re-encoding the received information bits to obtain a new parity sequence. The modulo–2 difference between this parity sequence and the original parity sequence is called the syndrome. If no errors have occurred, the parity bits computed at the decoder will be identical to those actually received, and the syndrome bits will be zero. If the syndrome bits are not zero, errors have been detected.

For error correction, the syndrome is processed further. The algebraic constraints defining a given block code generally yield a decoding technique or algorithm for the code. The decoding algorithm makes further use of the syndrome to calculate the error pattern affecting the received word. Most decoding algorithms require the use of binary quantization (hard decisions) at the demodulator output. The syndrome is processed using any one of the following methods:

2.1.1 Table Look-Up Decoding

There is a unique correspondence between the 2^{n-k} distinct syndromes and the correctable error patterns. Thus, for codes with small redundancy n-k, all correctable error patterns can be stored in a read-only memory (ROM), with the syndrome of the received word forming the ROM address. The error pattern is added modulo-2 to the received sequence to produce the transmitted code word. This procedure is used in some types of error correction hardware for computer memories.

2.1.2 Algebraic Decoding

The most prominent decoding method is the iterative algorithm for BCH codes due to Berlekamp. The basic idea is to compute the error-locator polynomial and solve for its roots. The complexity of this algorithm increases only as the square of the number of errors to be corrected. Thus, it is feasible to decode powerful codes. The use of Fourier-like transtorms has also been proposed to further reduce decoder complexity. The standard version of the algorithm is a bounded-distance algorithm. That is, not all possible error patterns can be corrected. The algorithm does not generalise easily to utilise soft decisions. There are several other algebraic decoding algorithms, some of which utilize soft decisions to improve performance. However, Berlekamp's algorithm is perhaps the deepest and most impressive result, and is straightforward to implement. This algorithm has permitted the use of BCH and Reed-Solomon codes in many applications, from the Voyager mission to compact disks.

2.1.3 Majority Logic Decoding

Majority logic decoding is a simple form of threshold decoding and is applicable to both block and convolutional codes. There are codes that, because of the special form of their parity check equations, are majority logic decodable. Reed-Muller codes are the most important class of codes of this type. A Reed-Muller code was used in the Mariner mission to encode photographs of Mars.

2.2 Convolutional Codes and their Decoding

Conovolutional codes have a much simpler mathematical structure than all but the most trivial block codes. Furthermore, unlike many block codes, it is possible to make use of soft- decision information in their decoding. For these reasons it is not surprising that they have been widely used. Because of the relatively small number of parameters specifying a convolutional code, many good codes have been found by computer search rather than by algebraic construction.

The error-correction capability of a convolutional code is determined in most cases by the free distance of the code. This is defined to be the minimum Hamming distance between any two semi-infinite code sequences generated by the encoder. By linearity, this is simply the minimum Hamming weight of any nonzero code sequence.

Three major decoding methods for convolutional codes are briefly described in the following sections.

2.2.1 Viterbi Decoding

Viterbi decoding is presently the most widely used decoding technique for convolutional codes. The Viterbi decoding algorithm finds the most likely(maximum likelihood) transmitted code sequence by using a structure called a trellis[9]. Each code sequence is represented by a path through the trellis. The degree to which a given code sequence matches the noisy received sequence is measured in terms of a path metric. Paths with high path metrics correspond to the most likely transmitted code sequences. The Viterbi algorithm is an efficient technique for searching all possible paths to find the most likely transmitted code sequence. In fact, the algorithm applies to any trellis code, not just the convolution codes. The significance of the trellis viewpoint is that the transmitted code sequence almost always corresponds to the path with the highest path metric. A major advantage of the Viterbi algorithm is the ease with which soft-decision information may be incorporated into the path metric. Unfortunately, the complexity of the Viterbi algorithm has an exponential dependence on the code's constraint length *K*. In practice, the Veterbi algorithm is rarely used with codes having constraint lengths exceeding 7. Another point worth mentioning is that Viterbi decoding does not perform very well in a bursty channel, making it necessary to use interleaving. Convolutional codes using the Viterbi algorithm are often concatenated with powerful block codes, especially in deep space applications.

2.2.2 Sequential Decoding

Again, code sequences are represented as paths in a trellis. Sequential decoding makes use of the fact that in most cases, there are only a small number of paths with high path metrics. Therefore, by carefully restricting the path search procedure, it is often possible to isolate the maximum likelihood path without keeping track of all possible paths. The complexity of sequential decoders is relatively independent of constraint length, so codes with large constraint lengths (up to 100) can be used, yielding large coding gains. Sequential decoding is more suitable than Viterbi decoding when low bit error rates ($< 10^{-5}$) are required. However, unlike the Viterbi algorithm, the procedure is suboptimum; only a small fraction of the possible code sequences is examined at any one time. The sequential decoder must be capable of detecting situations when the correct path is not in the set of sequences under examination and 'backtracking' to the point where the correct path was most likely lost. The decoder must then examine a different set of paths extending from that point. Several stages of backtracking may be necessary to find the correct path again. A major disadvantage of sequential decoding is that the number of computations is an ill-behaved random variable, necessitating a very large buffer. Consequently, performance is limited by the probability of the buffer overflow.

2.2.3 Threshold Decoding

Some convolutional codes are threshold decodable. Several parity checks may be calculated for each message bit and if they exceed a threshold, a decision on the correctness of the bit is made. Moderate values of coding gain (1-3 dB) can be obtained with relatively inexpensive decoders and limited amount of redundancy.

2.3 ARQ and Hybrid FEC ARQ Schemes

In an automatic repeat request (ARQ) scheme, whenever the receiver detects an error in the transmitted message, it sends a retransmission request to the transmitter over a feedback channel. These requests are repeated until the message is received correctly. Three basic types of ARQ protocols are commonly used-stop-and-wait, go-back-N, and selective-repeat [10-12].

Because of its simplicity, ARQ is used in many data communications systems. However, the technique has a major shortcoming-the throughput efficiency may be highly dependent on channel conditions. At low SNRs, the number of retransmissions required for correct message transmission may be very large. Hence, a successful transmission may involve a very long time

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE