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.

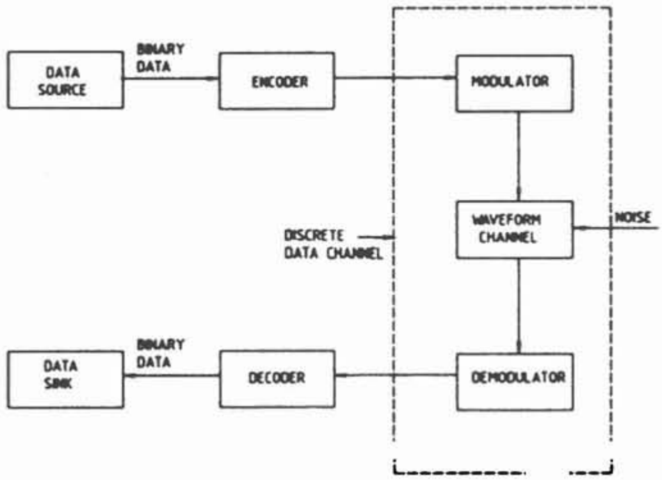

Figure 1. Digital communication using forward error control coding.

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.

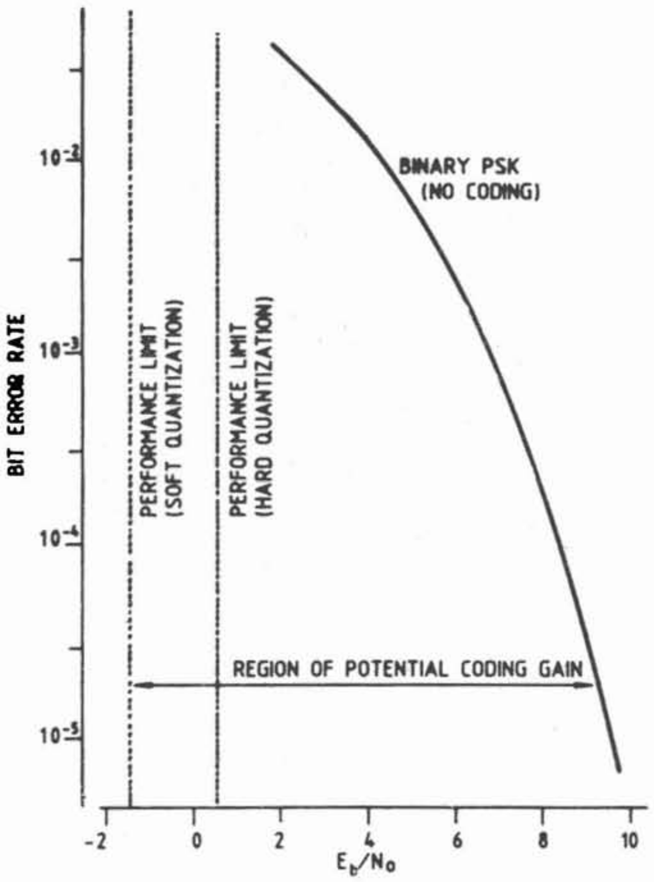

Figure 2. Performance of uncoded PSK over AWGN channel.

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 straightforward, 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 step 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 transforms 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

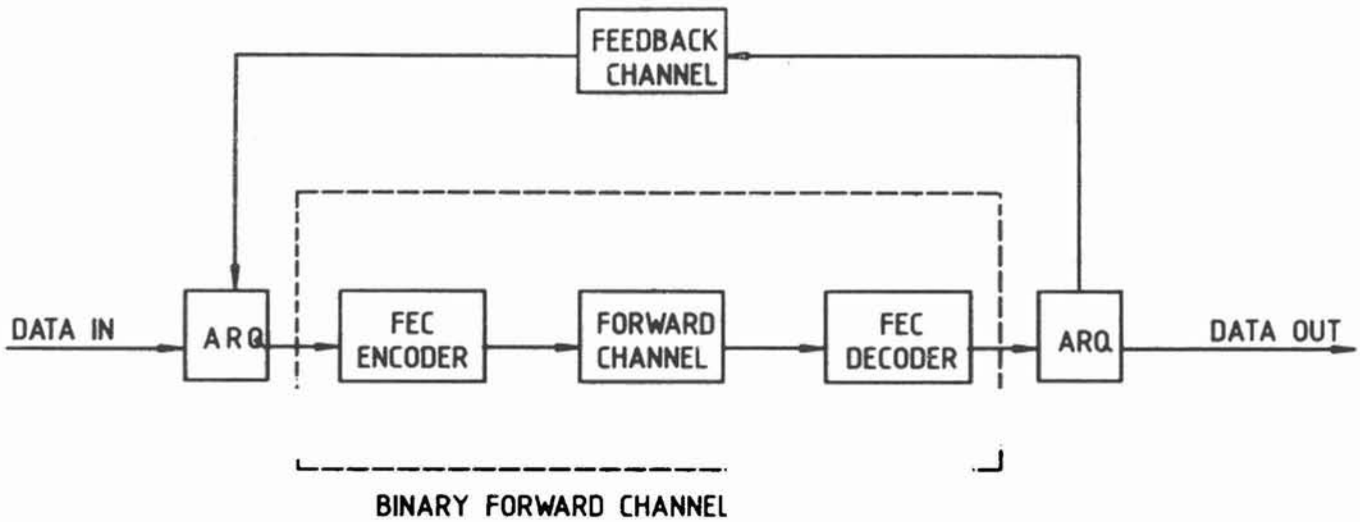

Figure 3. FEC/ARD hybrid error control coding system.

delay. Of course, this behaviour is unacceptable for delay-sensitive applications such as the digital voice communications. One approach to reducing the time delay is the hybrid FEC/ARQ scheme (Fig. 3). Here an ARQ scheme is used to obtain a desired error rate. FEC coding is used to correct low-weight error patterns in each message, reducing the number of retransmission requests.

In a type-I hybrid FEC/ARQ scheme, the message and error detecting parity bits generated by the ARQ system are further encoded with an FEC code[13]. At the receiver, the error correction parity bits are used to correct channel errors. The FEC decoder produces an estimate of the message and the error detection parity bits. The result is then tested by the error detection system to determine if the message should be accepted as error-free, or rejected as containing errors. If the channel signal strength is poor (high bit error rate), or if the message is long, the probability of error-free transmission may approach zero. Under these conditions, the efficiency may be improved by using a type-I hybrid scheme rather than a simple ARQ protocol. However, if signal strength is adequate, the type-I hybrid scheme involves a waste of bandwidth; the error correction parity bits are unnecessary. Thus, a plot of data throughput versus signal strength exhibits a crossover point between standard ARQ and the type-I hybrid schemes.

In a type-II hybrid FEC/ARQ scheme[14] the first transmission of a message is coded with error detection parity bits alone, as in a standard ARQ protocol. If the receiver detects errors in the received block, it saves the erroneous block in a buffer and requests a retransmission. The retransmitted information is not the original coded message, but a block of parity bits derived by applying an FEC code to the message. The receiver uses these parity bits (which may themselves be in error) to correct the block stored in the receiver buffer. If error correction is not successful, subsequent retransmissions are requested, which may consist of the original codeword or another block of error correction parity bits. The retransmission format depends on the strategy and on the error correction code used. The intention of type-II hybrid schemes is to provide the efficiency of standard ARQ under good channel conditions while obtaining the performance improvement of type-I hybrid schemes under poor channel conditions.

2.4 Rate-Adaptive Coding

In order to cope with diverse system requirements and service options, it is very attractive to investigate rate-adaptive FEC protocols, in which the code rate is adapted to suit prevailing transmission conditions[15]. Such a technique is applicable to satellite communication systems, where adaptive signal compensation may provide adequate performance under temporary poor channel conditions (due to rainfall, etc) without a permanent bandwidth reduction caused by excessive coding overhead In the rate-adaptive coding, a high rate code is used under good channel conditions to maintain a high information rate. As the channel condition worsens, the information rate is reduced by applying lower rate codes to maintain the desired BER performance. It is possible to optimize service quality by applying rate-adaptive FEC coding. Traditionally, convolutional codes have only been used in low code rate applications. However, as data transmission rates increase, the need arises for good high rate convolutional codes with practical encoding and decoding techniques. High rate punctured convolutional codes are a significant recent development in which the difficulties usually associated with decoding high rate codes are significantly reduced[16-17]. Viterbi or sequential decoding of rate b/v punctured convolutional codes is no more complex than the decoding of rate *1/v* codes. In fact, the punctured code is obtained in a straightforward manner from a 'parent' rate 1/*v* code, while a Viterbi or sequential decoder for the latter code can be easily modified to decode the punctured code. This property makes punctured convolutional codes attractive for rate-adaptive applications-a single decoder can accommodate the entire family of punctured codes arising from a single parent code.

Among the class of block codes, maximum distance separable (MDS) codes are particularly well suited to rate-adaptaive coding techniques. An (n, k) code for which the minimum distance equals $n-k+1$ is called a MDS code. MDS codes are optimal in the sense that no (n, k) code has minimum Hamming distance exceeding $n-k+1$. The MDS codes most often encountered in practice are RS codes. RS codes are attractive not only because of the ease with which they can be decoded using Berlekamp's decoding algorithm, but also because they feature a wide variety of possible code rates and block lengths. Rate adaptivity arises in the following manner. A low rate parent RS code can be broken down into several high rate subcodes, in much the same way that a parent convolutional code yields high rate punctured codes. The subcodes are themselves linear block codes having the MDS property. With appropriate modifications, the decoder for the parent code can also be used to decode the subcodes. Thus a single decoder may be used in a rate adaptive system where the transmitter employs the subcode most suitable for the current channel state.

3. CODING APPLICATIONS

3.1 Comparison of Block and Convolutional Codes

The comparison of block and convolutional codes presented here is applicable to an AWGN channel. For channels with memory such as a mobile fading channel, the benefits of coding are even more spectacular.

The comparison is based on an uncoded binary PSK system employing coherent detection. The information rate of the coded systems under comparison is assumed to be fixed. The coded systems require more frequency bandwidth. Table 1 has been adopted from[18]. Two bit error rates considered here are 10^{-5} and 10^{-8}. The column labelled 'data rate capability' is taken to be the following: low (<10 kbps), moderate (10 kbps to 1 Mbps), high (1 to 20 Mbps) and very high (>20 Mbps).

At moderate and high data rates, convolutional coding with Viterbi decoding appears to be the most attractive coding technique. This assumes that there is

Table 1: Comparison of major coding techniques on an AWGN channel

Coding technique	Coding gain at 10^{-5} (dB)	Coding gain at 10^{-8} (dB)	Data rate capability
Concatenated (RS and Viterbi)	6.6-7.5	8.5-9.5	Moderate
Sequential decoding (soft decision)	6.0-7.0	8.0-9.0	Moderate
Block codes (soft decision)	5.0-6.0	6.5-7.5	Moderate
Concatenated (RS and short block)	4.5-5.5	6.5-7.5	Very high
Viterbi decoding (hard decision)	4.0-5.5	5.0-6.5	High
Sequential decoding (hard decision)	4.0-5.0	6.0-7.0	High
Block codes (hard decision)	3.0-4.0	4.5-5.5	High
Block codes (threshold decoding)	2.0-4.0	3.5-5.5	High
Convolutional codes (threshold decoding	1.5-3.0	2.5-4.0	Very high

no appreciable interference other than Gaussian noise, that a decoded bit error rate of 10^{-5} is satisfactory, and that long continuous bit streams are transmitted. As mentioned previously, one advantage of the Viterbi algorithm over algebraic block decoding algorithms is its ability to make use of soft-decision information. However, if efficient algorithms for decoding long block codes with soft decisions are developed, they will undoubtedly be quite competitive.

At very high data rates, concatenated RS and short block code systems can provide roughly the same gain with less complexity than Viterbi decoding. For larger coding gains at high speeds, sequential decoding with hard decisions appears to be the most attractive choice. At moderate data rates a better case can be made for using sequential decoding with soft decisions. However, in situations where the system protocols require the transmission of blocks of data (such as TDMA), block codes appear to be more attractive.

In many applications, concatenated coding[19] offers a substantial coding gain. Decoding errors typically manifest themselves as burst errors at the decoder output. These types of bursts can be corrected by superimposing an outer burst-error-correcting code, usually a RS code.

The effect of using the two decoders, each combating a specific type of error, results in significant coding gain improvement. The inner decoder reduces poor quality data to medium quality data and the outer decoder reduces medium quality data to very good quality data. Long codes are often desirable for good error control regardless of channel characteristics, and concetenation is a practical way of creating efficient long codes.

Another important factor to be taken into account is the hardware complexity of the error control coding system. For high- speed transmission systems, selection of the coding scheme is severely restricted by the scarcity of high-speed FEC decoders. Some typical error control coding techniques and associated decoder complexities are listed in Table 2.

In practice, the selection of a coding scheme greatly depends on the data rate, the channel conditions, and implementation complexity. For example, block codes are better suited to applications requiring high data rates. A convolutional decoder would fail to keep up. However, for predominantly Gaussian channels, the convolutional codes tend to outperform most short block codes of interest. Most of the previous applications of FEC coding to digital communications involve convolutional codes. This is mainly because of technological difficulties of implementing decoders for powerful block codes. With recent developments in VLSI technology, the situation has changed. New technology and a demand for integrated packetised data/voice communications have made sophisticated block coding schemes increasingly attractive from an economic viewpoint.

3.2 Applications to Satellite Communications.

Two of the most treasured resources in satellite communications are power and bandwidth. Conserving these resources has been a major concern ever since information transmission via satellite became a reality. Error control coding often used to improve the transmission quality, which is otherwise compromised by interference and power limitations. Furthermore, because of the considerable propagation delay

Table 2. Comlexity of typical error control coding techniques

Type of error	Coding technique	Decoder complexity
	Convolutional codes with threshold decoding	Low
	BCH codes	
Random	onvolutional codes with Viterbi decoding	
	Convolutional codes with sequential decoding	
	Voncatenated (RS and Viterbi)	High
	ARQ with error detection only	Low
	Fire codes	
Burst	Reed-Solomon codes	
	Doubly codes Reed-Solomon codes	
	Concatenated (RS and Viterbi) with interleaving	High

(250-300 ms) in geostationary satellite links, FEC techniques tends to be more widely used than ARQ techniques, which require data retransmission.

One of the remarkable features of satellite communication systems is that bit errors occurs randomly, which is considered to be a significant advantage when applying FEC codes. A recent trend in digital satellite communication systems is to employ powerful FEC codes with large coding gains, making efficient use of limited satellite resources[20]. Since the usable frequency bands are severely limited, it is also desirable to apply high rate FEC codes.

In satellite communication systems, convolutional codes with constraint length 7 are widely used. Block codes are also applied in some satellite systems. Examples of using block codes include a (31, 15) RS code for the joint tactical information distribution system (JTIDS), a (127, 112) BCH code for the INTELSAT V system, and a (7, 2) RS code for the air force satellite communications (AFSATCOM) wideband channels.

Heavy signal attenuation is occasionally experienced in satellite channels due to natural phenomena such as heavy rainfall and scintillation in the atmosphere or ionosphere. It, therefore, seems natural to apply powerful FEC codes only when large signal attenuation occurs. Such an adaptive technique may be even more effective if transmitter power control and/or diversity reception are simultaneously applied.

To meet the requirement for FEC with a large coding gain, soft decision Viterbi decoding will continue to play an important role in future satellite systems. Current technology trends indicate that Viterbi decoders for convolutional codes having constraint lengths of up to 10 will soon be available, yielding a 1 dB coding gain improvement over codes having constraint length 7. Higher rate codes based on punctured coding are expected to be widely utilised. On the other hand, for systems in which a large coding gain is required at the expense of increased signal processing delay, sequential decoding and/or concatenated coding of RS codes and convolutional codes are viable alternatives.

3.3 Applications to Mobile Communications.

Mobile communications are defined as communications involving mobile vehicles such as automobiles, trains, airplanes, and marine vessels. Mobile communications rely on wireless (radio) communication technology. In the mobile environment, burst errors due to multipath fading are dominant. Since the bandwidth available to each channel is strictly limited, the code rate must be high. Furthermore, since each mobile terminal must obviously include a decoder, codes requiring a complex decoder cannot be used. Therefore, burst-error-correcting codes with simple decoding algorithms are preferable for mobile communications. In some systems, BCH codes are used in conjunction with ARQ.

For mobile terminals operating in the presence of large Doppler offset and Doppler rate, as well as multipath and fading, RS codes are extremely attractive to combat the resulting bursty noise. In fact, the JTIDS program has adopted a (31, 15) RS code, which may soon be adopted as a standard for all tactical military communication links in NATO countries. With current technology, the amount of hardware required to implement RS decoders is modest in comparison to the hardware required for routine system functions such as timing, synchronising, buffering and controlling.

A recent example is the coding standard for digital mobile radio proposed by the North American railroads for advanced train control systems (ATCS). The (16, 12) RS code adopted for ATCS provides the best trade-off between throughput, delay, and implementation complexity. The RS code will be used in a hybrid FEC/ARQ system, ensuring a probability of undetected error of less than 10^{-10}.

Powerful random-error-correcting coding combined with bit interleaving is a promosing approach in systems where processing delay is not a concern. In this case, high coding gain is expected because powerful FEC techniques such as convolutional coding with Viterbi decoding can be employed. The application of VLSI technology will make it possible to realise such a coding system with few chips. Because the required data transmission rate is generally low for mobile communications, digital signal processor (DSP) technology can also be profitably applied.

Mobile communications are expected to evolve to a point where satellite access is possible, allowing global personal communications. Ultimately, mobile satellite systems should be capable of providing basic communications services such as the voice and low rate data to very small terminals, including handheld units. It is anticipated that power limitations will be even more strict in future systems. This means that error control coding will become even more important in future mobile satellite communication systems.

3.4 Applications to Broadcasting

Conventional broadcasting such as AM/FM radio and TV is based on analog transmission. However, digital technology has been rapidly advancing along with IC technology. Microcomputers and their associated peripheral equipment have also been greatly popularised. As a result, digital signals are gradually being used more for information services. The broadcasting system itself is evolving into a broader concept including data broadcasting, providing unprecedented variety, selectivity, and instantaneity properties.

The most popular data broadcasting services are code transmission teletext systems under development in Japan, Britain, France, Canada, the United States, and elsewhere. The quality of TV transmission paths are very poor for coded digital signals. In teletext, which transmits digitised characters and figures overlapped with a TV signal, the important types of errors include multipath, random noise, impulse noise, and waveform distortion due to rebroadcasting. Thus a code with large error-correction capability is required. However, a decoder must be provided to each subscriber, necessitating compact decoder hardware. Consequently, a (272, 190) shortened different set cyclic code, which can be decoded by a simple majority logic circuit, has been selected for Japanese teletext. This code has been derived from simulations based on bit-error-rate data collected in field tests.

Another type of data broadcasting service being developed in Japan is PCM sound via direct broadcasting satellite (DBS)[21]. Since the most important bit errors are due to rain attenuation, errors due to multipath, impluse noise, and waveform distortion may be neglected. A (63, 56) BCH code is used as the error-correcting code for this system.

4. CONCLUSION

In this paper, the development of basic coding theory and the state-of-art coding techniques have been reviewed. The applications of coding to communication systems and future trends are also discussed. The theory of error control coding is a very active area of research. An enormous amount of literature is now available on coding and its applications. For the interested reader, we include a brief bibliography. More comprehensive lists of references are available elsewhere[11, 22-24]

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. Over the past ten years, VLSI technology has reduced the cost of coding systems by many orders of magnitude. Future generations of technology are expected to continue this trend. Indeed, more complicated coding schemes will certainly become an economic necessity.

REFERENCES

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **17**, 379-423.
2. Berlekamp, E.R. The technology of error-correcting codes. *Proceedings of the IEEE,* 1980, **68**, 564-93.
3. Bhargava, V.K. Forward error correction schemes for digital communications. *IEEE Commun. Mag.* 1983, **21**, 11-19.
4. Burton, H.O. & Sullivan D.D. Errors and error control. *Proceedings of the IEEE*, 1972, **60**, 1293-1301.
5. Berlekamp, E.R. Algebraic coding theory. McGraw-Hill, New York, 1968.
6. Blahut, R.E. Theory and practice of error control codes. Addison-Wesley, Reading, MA, 1983.
7. MacWilliams, F.J. & Sloane, N.J.A. The theory of error-correcting codes. North-Holland, New York, 1977.
8. Michelson, A.M. & Levesque, A.H. Error-control techniques for digital communication. Wiley, New York, 1985.
9. Viterbi, A. J. Convolutional codes and their performance in communication system. *IEEE Trans. Commun. Technol.* 1971, **COM-19**, 751-72.
10. Comroe, R.A. & Costello, Jr., D.J. ARQ schemes for data transmission in mobile radio systems. *IEEE J. Selected Areas in Commun.*, 1984, **SAC-2**, 472-81.

Lin, S.; Costello, Jr, D.J. & Miller, M.J. Automatic-repeat- request error-control schemes. *IEEE Commun. Mag.*, 1984, **22**, 5-17.

12 Schwartz, M. Telecommunication networks: protocols, modeling and analysis. Addison-Wesley, Reading, MA, 1987.

13. Rocher, E.Y. & Pickholtz, R.L. An analysis of the effectiveness of hybrid transmission schemes. *IBM J. Res. Dev.*, 1970, 426-33.

14 Lin, S. & Yu, P.S. A hybrid ARQ scheme with parity retransmission for error control of satellite channels. *IEEE Trans. Commun.*, 1982, **COM-30**, 1701-19.

15. Mdandelbaum D.M. An adaptive-feedback coding scheme using incremental redundancy. *IEEE Trans. Inform. Theory*, 1974, **IT-20**, 388-89.

16. Cain, J.B.; Clark, G.C. & Geist, J. Punctured convolutional codes of rate (n-1)/n and simplified maximum likelihood decoding. *IEEE Trans. Inform. Theory*, 1979, **IT-25**, 97-100.

Hagenauer, J. Rate compatible punctured convolutional codes and their applications. *IEEE Trans. Commun.*, 1988, **COM-36**, 389-400.

18. Clark, G.C. & Cain, J.B. Error correction coding for digital communications. Plenum Press, New York. 1981.

19. Forney, G.D. Concatenated Codes. MIT Press. Cambridge, MA, 1966.

20. Wu, W.W.; Haccoun, D.; Peile, R.E. & Hirata, Y. Coding for satellite communication. *IEEE J, Select, Areas Commun.*, 1987, **SAC-5**, 724-48.

21 Imai, H. (Ed). Essentials of error-control coding techniques. Academic Press, San Diego, CA, 1990.

22. Berlekamp, E.R.; Peile, R.E. & Pope, S.P. The application of error control to communications. *IEEE Commun. Mag.*, 1987, **25**, 44-57.

23 Bhargava, V.K.; Haccoun, D.; Matyas, P. & Nuspl, P. Digital communication by satellite. Wiley, New York, 1981.

24. Lin, S. & Costello, D.J.Jr. Error control coding: fundamentals and applications. Prentice-Hall, New Jersey, 1983.