CODING THEORY

Submitted by

SELMATH K. I

Register No:170021032428

ATHULYA MARY

Register No:170021032402

ANU A.B

Register No: 170021032400

Under the guidance of

Dr. Savitha K S

In partial fulfillment of the requirement for the award of
BACHELOR DEGREE OF SCIENCE in MATHEMATICS

2017- 2020

ST. PAUL'S COLLEGE KALAMASSERY, HMT COLONY P.O - 683503

CERTIFICATE

This is to certify that the project report title “**CODING THEORY** ” submitted by **ANU A.B**(Reg no.170021032400),**ATHULYA MARY** (Reg no.170021032402),**SELMATH K.I**(Reg no.170021032428) towards partial fulfillment of the requirements for the award of degree of bachelor of science in mathematics is a bonafide work carried out by them during the academic year 2017-2020

Supervising guide

Dr.Savitha K S

Assistant Professor
Department of mathematics

Head of the Department

Dr.Savitha K S

Assistant professor
Department of mathematics

Place : Kalamassery

Date :

DECLARATION

We, **ANU A.B** ,(Reg No:170021032400),**ATHULYA MARY** (Reg no.170021032402),**SELMATH K I**(Reg no.170021032428) hereby declare that this project entitled "**CODING THEORY**" is an original work done by us under the supervision and guidance of Dr. Savitha K S , faculty, Department of Mathematics in St. Paul's college Kalamassery in partial fulfillment for the award of The Degree of Bachelor of Science in Mathematics under Mahatma Gandhi University. We further declare that this project is not partly or wholly submitted for any other purpose and the data included in the project is collected from various sources and are true to the best of our knowledge.

SELMATH K.I
ATHULYA MARY
ANU A.B

Place: Kalamassery

Date:

CODING THEORY

ACKNOWLEDGEMENT

For any accomplishment or achievement the prime requisite is the blessing of the Almighty and it’s the same that made this world possible. We bow to the lord with a grateful heart and prayerful mind.

It is with great pleasure that we express our sincere gratitude to our beloved teacher Dr. Savitha K S , Department of Mathematics, St. Paul’s College, for her overwhelming support, motivation and encouragement.

We would like to thank all the faculty members of the department and my friends who helped us directly and indirectly through their valuable suggestions and self-criticisms, which came a long way in ensuring that this project becomes a success.

ABSTRACT

Coding theory – theory of error correcting codes – is one of the most interesting and applied part of mathematics and informatics.

All real communication systems that work with digitally represented data, as CD players, TV, Fax machines, internet ,satellites, mobiles, requires to use error correcting codes because all real channels are ,to some extent ,noisy – due to interference caused by environment

Coding theory problems are therefore among the very basic and most frequent problems of storage and transmission of information.

Coding theory results allow to create reliable systems out of unreliable systems to store and /or to transmit information.

Coding theory methods are often elegant applications of very basic concepts and methods of (abstract) algebra.

Basic concepts, methods, results are discussed in this project

Coding Theory

INTRODUCTION

Coding theory sometimes called algebraic coding deals with the design of errors correcting codes for the reliable transmission of information across noisy channel. It is one of the most interesting and applied part of mathematics and information. It has connections with other areas of discrete mathematics, especially number theory and theory of experimental designs.

The reliable transmission of information over noisy channel is one of the basic requirements. Transmission is understood both as transmission in space (over mobile radio channels) and as transmission in time by storing information in appropriate storage media. Because of this requirement communication systems rely heavily on powerful channel coding methodologies. A communicating system which sends information from one place to another.

Eg: Telephone networks, computer networks, storage systems (such as magnetic and optical disc driver) are systems for storage and later retrieval of information. Coding theory is mainly concerned with exploit methods for efficient and reliable data transmission or storage which can be roughly divided into data compression and error control techniques former attempts to compress.

The data from a source in order to transmit or storage them more efficiently (as it is found in internal where data are usually transformed into the zip format to make files smaller and reduce the network load) while the latter adds extra data bits to make transmission of data more robust to channel disturbances.

CHAPTER 1

WORKING OF CODES IN COMMUNICATION

1.1 BASIC OPERATIONS IN COMMUNICATIONS

Communication and storage system can be regarded as examples of information processing system and may be represented abstractly by block diagram in the fig 1.1.

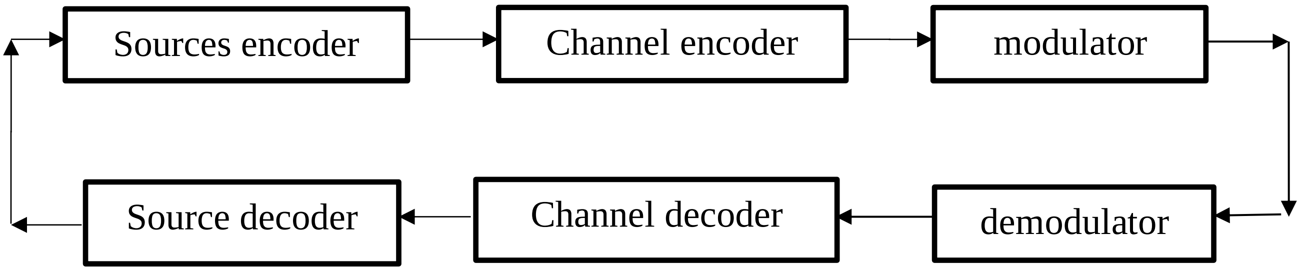


Fig 1.1

In all cases, there is a source from which the information originates and examples include a book, a video, a music, etc. The source output is processed by the encoder to facilitate the transmission (the storage) of the information. In general, three basic operations are executed in the encoder; source coding, channel coding, and modulation. For source coding, the encoder maps the source output into digital format. The mapping is one -to- one and it reduces the redundancy. By channel coding, the encoder introduces extra redundant data so as to combat the noisy environment in which the information must be stored or transmitted. Also we need proper modulation to convert the data to wave fronts that are suitable for transmitted or recording the output of the encoder is then transmitted through some physical communication channel or stored in some physical storage medium. Wireless radio transmission magnetic storage devices are examples. Each of these examples contains various types of noise disturbance. In a telephones call, It may come from thermal noise, switching noise or cross talk from other lines while in a magnetic discs, the surface defects and dust particles are the noise disturbances. Information conveyed and processes to restore its original form. This is the task of the decoder. The signal processing performed by the decoder can be viewed as the inverse of the function performed by the encoder which includes demodulation, channel decoder and source decoder. The output of the decoder is then presented to the final user, which we call the information sink.

The physical channel usually produces a received signal which differs from the original input signal. This is because of signal distortion and noise introduces by the channel consequently the decoder can only produce an estimate of the original information message. All well designed system aim at reproducing as reliably possible per unit time or per unit storage.

1.2 INFORMATION SOURCE

Nature supplies information in continues form. However digital signals in which both amplitude and time taken on discrete values are preferred in modern communication systems. Main reason for the use of digital signal is that they can be transmitted more reliably than analog signals. When the inevitable corruption of the transmission of the system begins to degrade the signal, the digital pulses can be detected, reshaped and amplified to standard form before relaying them to their final destination. Fig 1.2 illustrate an ideal binary digitals pulse propagating along a transmission line, Where the pulse shape is degraded as a function of line length. At a propagating distance where the transmitted pulse can still be reliably identified, the pulse is amplified by a digital amplifier that recovers its original ideal shape. The pulse is thus regenerated on the other hand; analog signals cant be reshaped since they taken infinite variety of shapes. Hence the further the signal is sent and the more it is processed, the more degradation it suffers from small errors.

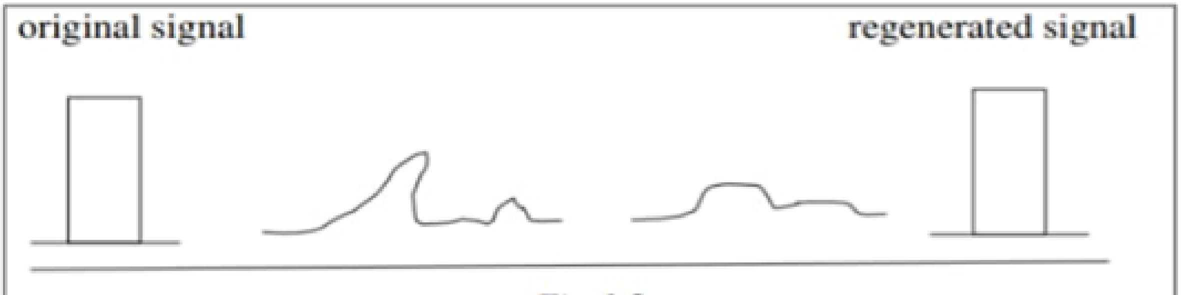


Fig 1.2

Modern practices for transforming analog signals into digital forms is no sample the continuous signals at equally spaced intervals of time and then to quantize the observed value but as a result of sampling and quantizing operations errors are introduce into the digital signal. These errors are non-reversible in that it is not possible to produce an exact replica of the original analog signal from its digital representation. However the errors are under the designers control. Indeed by proper selection of the sampling rate and the number of quantization levels, the errors due to sampling and quantization can be made so small that the difference between the analog signal and its digital reconstruction is not discernible by a human observer.

1.3 INFORMATION ENTROPY

Information entropy is the average rate at which information is produced by a stochastic source of data. The measure of information entropy associated with each possible data value is the negative logarithm of the probability mass function for the value.

$$S = -\sum Pi \log Pi = -Ep\ (\log P)$$

Where $Ep(X) = \sum PiXi$ is the expectation defined by the probability p .

The entropy provides an absolute limit on the shortest possible average length of a lossless compression encoding of the data produced by a source, and if the entropy of the source is less than the channel capacity of the communication channel, the data generated by the source can be reliably communicated to the receiver (at least in theory, possibly neglecting some practical considerations such as the complexity of the system needed to convey the data and the amount of time it may take for the data to be conveyed).

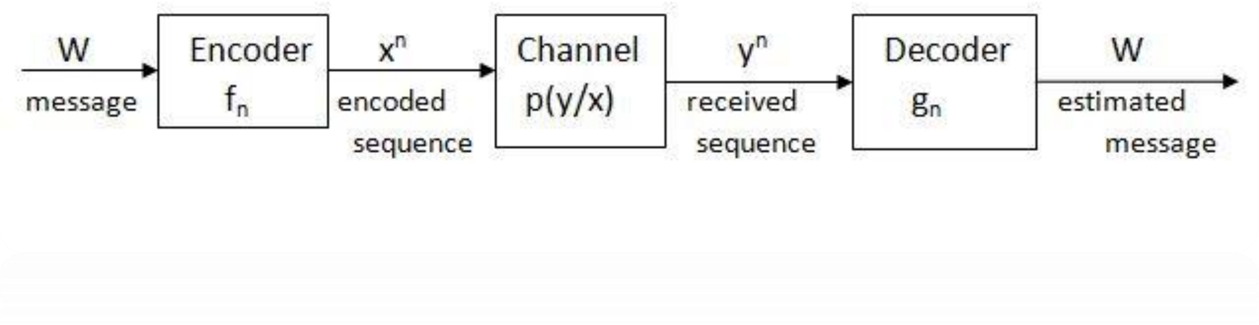
Information entropy is typically measured in bits (alternatively called “shannons”) or sometimes in “natural units”(nats) or decimal digits (called ‘dits’ , ‘ban’ or ‘hartleys’). The unit of the measurement depends on the base of the logarithm that is used to define the entropy.

English text, treated as a string as a string of characters has fairly low entropy that is fairly predictable. If we do not know exactly what is going to come next, we can be fairly certain that for example ‘e’ will be far more common than ‘z’ that the combination ‘qu’ will be much more common than any other combination with a ‘q’ in it, and that the combination ‘th’ will be more common than ‘z’,’q’, or ‘qu’. After the first few letters one can often guess the rest of the word. English text has between 0.6 and 1.3 bits of entropy per character of the message. If a compression scheme is lossless – one in which you can always recover the entire original message by decompression -then a compressed message has the same quantity of information as the original, but communicated in fewer characters. It has more information (higher entropy) per character. A compressed message has less redundancy.

1.4 CHANNEL CAPACITY

Channel capacity, in electrical engineering, computer science and information theory is the tight upper bound on the rate at which information can be reliably transmitted over a communication channel.

Following the terms of the noisy-channel coding theorem, the channel capacity of a given channel is the highest information rate (in units of information per unit time) that can be achieved with arbitrarily small error probability. The basic mathematical model for a communication system is the following.



Where,

- W is the message to be transmitted.
- X is the channel input symbol (X^n is a sequence of n symbols) taken in an alphabet X.
- Y is the channel output symbol (y^n is a sequence of n symbols) taken in an alphabet Y.
- $\hat{W}$ is the estimate of the transmitted message
- fn is the encoding function for a block of length n
- P(y/x) = PY/x(y/x) is the noisy channel which is modeled by a conditional probability distribution and length n.
- gn is the decoding function for a block of length n.

Let X and Y be modeled as random variables.Further more, let PY/x(y/x) be the conditional probability distributed function of Y given X , which is an inherent fixed property of the communication channel. Then the choice of the marginal distributed PX (x) completely determines the joint distribution PX,Y(x,y) due to the identity.

PX,Y(x,y)= PY/X(y/x)PX(x)

Which ,in turn includes a mutual information I(x;y). The channel capacity is defined as c=Sup I(x;y)

PX(x)

Where the supremum is taken overall possible choices of Px(x)

1.4.1 Binary Symmetric Channel

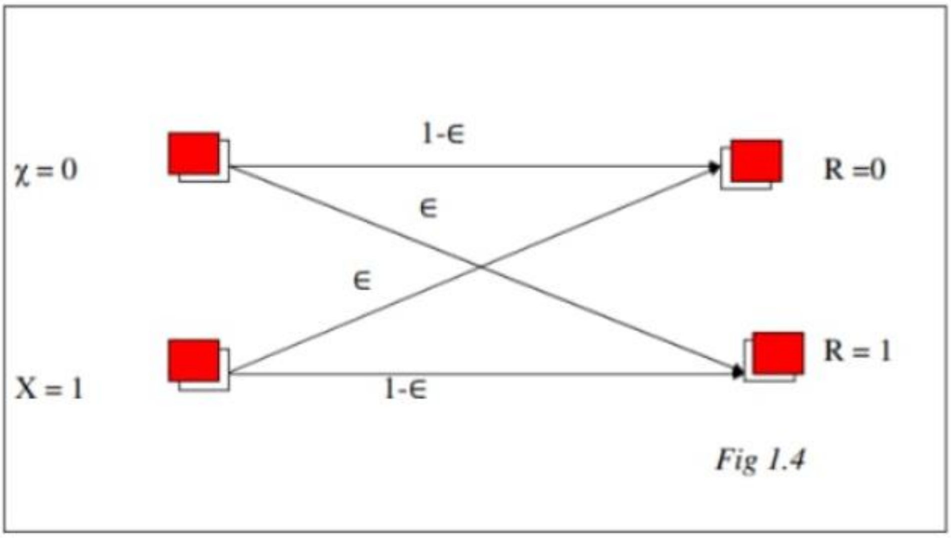


Fig 1.4

Fig 1.4 shows the channel diagram of the binary symmetric channel with bit error probability ε. This channel transmits the binary symbol x=0 or x=1 correctly with probability 1-ε. Whenever the incorrect binary symbol R-1 or R=0 is emitted with probability ε.

By maximizing the mutual information I(X:R), the channel capacity of a binary symmetric channel is obtained according to

$$C=1+\varepsilon\log_2\varepsilon+(1-\varepsilon)\log_2(1-\varepsilon)$$

This channel capacity is equal to 1 if ε =0 or ε=1 for ε =1\4,the cahnnel capacity is 0.The channel capacity is 0.

1.4.2 AWGN Channel

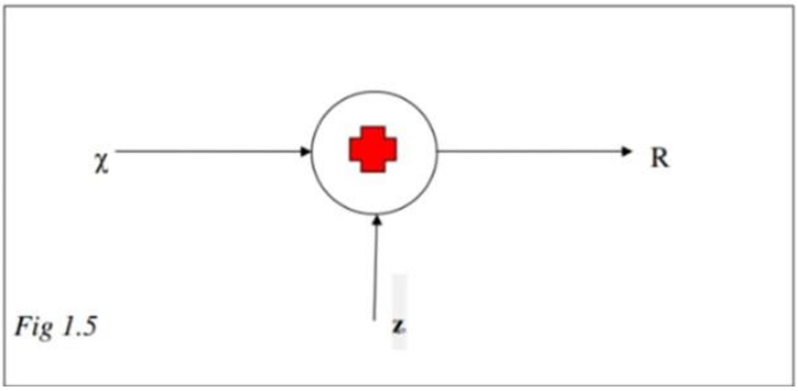


Fig 1.5

Up to now we have exclusively considered descrete valued symbols. The concept of entropy can be transferred to continuous real valued random variables by introducing the so-called differential entropy. It turns out that a channel with real valued input and output symbols can again be characterized with the help of the mutual information I and its maximum, the channel capacity c. In Fig 1.5 the so called AWGN channel is illustrated which is described by the additive white Gaussian noise term with the help of signal power.

S={X^2} and Noise power ,N={x}

The channel capcity of the AWGN channel is given by

C=1\2 log (1+S\N)

Here the channel capacity exclusively depends on the signal to noise ratio S\N.

1.5 CODE PARAMETERS

Channel code are characterized by the so called code parameters . With the help of there code parameters, the efficiency of the encoding process and the error detection and error correction capabilities can be evaluated for a given (n,k) block code.

Code Rate

In telecommunication and information theory, the code rate(or information rate) of a forward error correction code is the proportion of the data stream that is useful (non-redundant).That is, if the code rate is k\n for every k bits of useful information,the coder generater a total of n bits of data,of which n-k are redundant.

Minimum Weight

In error,correcting coding , the minimum Hamming weight,commonly referred to as the minimum weight W_{min} of a code is the weight of the lowest weight of non zero code word. The weight W of a code word is the number of 1s in the word.

For eg:The word 11001010 has a weight of 4.

Hamming Distance

In information theory, the Hamming distance between two strings of equal length. Is the number of positions at which the corresponding symbols are different.In other words ,it measures the minimum number of substitutions required to change one string into the other, or the minimum number of errors that could have transformed one string into the other.

Eg: hamming distance between karolin and kathrin is 3.

1.6 CODING GAIN

In many practical applications, ona has to choose B and W ,where B equals the number of user bits per second that must be transmitted reliably through a noisy channel using a power of atmost W walt.A well known example is mobile telephony, where B determines the speech quality and W is related to life-time of batteries. Another example is in deep space transmissions,where B determines the number of pictures that can be transmitted. While W is the power that is available from solar panels. In all these cases, the transmitter has an average energy of E_b = W/B Joule per user bit available to generate signals to be sent to the receiver. Coding may influence the choices. The effect of coding is often expressed as ‘coding gain’ which is defined as follows;

The ratio between SNR (uncoded) and SNR^1(coded) for equal probability after decoding is called the coding gain.

Here SNR stands for signal to Noise ratio which is the ratio E_b/σ^2, where E_b is the available energy and σ^2 is the variance of noise drawn from Gaussian distribution.

1.7 SIMPLE CHANNEL CODE

As an introductory example of a simple code we consider the transmission of the binary information sequence.

$$00\underline{1}0\underline{1}110$$

Over a binary symmetric channel with bit error probability on average, every fourth binary symbol will be received incorrectly. In this example we assume that the binary sequence

$$00\underline{0}0\underline{0}110$$

is received at the output of the binary symmetric channel. In order to implement a simple error correction scheme, we make use of the so-called binary triple repetition code. This simple code is used for the encoding of binary data, If the binary symbol 0 is to be transmitted the encoder emits the code word 000 alternatively the code word 111 is issued by the encoder. When binary symbol 1 is to be transmitted for the binary operation information sequence given above we obtained binary code sequence

000 000 111 000 111 111 111 000

At the output of the encode. If we again assume that on average every fourth binary symbol is incorrectly transmitted by the binary symmetric channel, we may obtained the received sequence.

$$0\underline{1}0 \;\; 000\underline{0}11 \; 0\underline{1}0 \;\; 111\underline{0}1\underline{0} \;\; 111 \;\; 0\underline{1}0$$

The decoder tries to estimate the original information sequence with help of majority decision. If the number 0's with in a received 3 bit word is larger than the number of 1's , the decoder emits the binary symbol 0, otherwise a 1 is decode with this decoding algorithm, we obtain the decoded information sequence.

$$00101\underline{0}10$$

1.8 ENCODING A SOURCE ALPHABET

We would assume without loss of generality that an information source generates a finite number of messages. This is undoubtedly true for a digital source. As for an analog source, the analog to digital conversion process makes the assumption feasible. However even through specific messages are actually set, the system designer has no idea in advance, which messages will be chosen for transmission. We thus need to think of a source as a random source of information and ask how we may encode, transmit and recover the original information. Generally all the messages are one to one mapping, no matter which encoding methods are employed. The original message can always be recovered from its present sequence. Given an encoding method, Let l_i denote the length of the output sequence, called the code word. Corresponding to u_i $1<i<r$ from the view point of some coding for data compression. An optimal encoding should minimize the average length of cord word is defined by,

$$L_{av}=\sum P_iL_I$$

As for channel coding a good encoding method should be able to protect the source message against the invertible noise corruption hence the encoding of the messages to be transmitted over the channel odd's redundancy to combat channel noise on the other hands, the source encoding usually removes redundancy contained in the massage to be compressed.

1.9 DECODING THE DATA RECEIVED

The major decoding methods of convolution codes are described below.

1.9.1 Viterbi Decoding

It is the most widely used decoding technique. It is an efficient technique for searching all possible paths to find the most likely transmitted code sequence.

1.9.2 Sequential Decoding

In sequential coding, codes with large constraints length can be used yielding large coding gains. It is more suitable than Viterbi decoding when the low bit error rate are required.

1.9.3 Threshold Decoding

Several parity checks may be calculated for each message bit and if they exceed a threshold, a decision on correctness is made.

The major decoding methods of block words are listed below. The first step of decoding process involves recoding the received information bit to obtain a new parity sequence.

Syndrome difference between this parity sequence and original parity sequence. If no error is occurred syndrome is zero or otherwise the syndrome is processional further for error correction. The syndrome is processional using following methods.

1.9.4 Table Look Up Decoding

There is unique correspondence between the Z_n-k distinct syndrome and the correctable error patient codes with small redundancy n-k, all correctable error patients can be stored in ROM, with the syndrome of the received word forming the ROM address.

1.9.5 Algebraic Decoding

The basic idea is to compute the error locator polynomial and solve its roots. The complexity of this algorithm increases only as the square of the number of the error to be corrected.

1.9.6 Majority Large Decoding

It is a simple form of decoding applicable to both block and convolution codes. Because of the special form of their parity check equator are majority logic decodable.

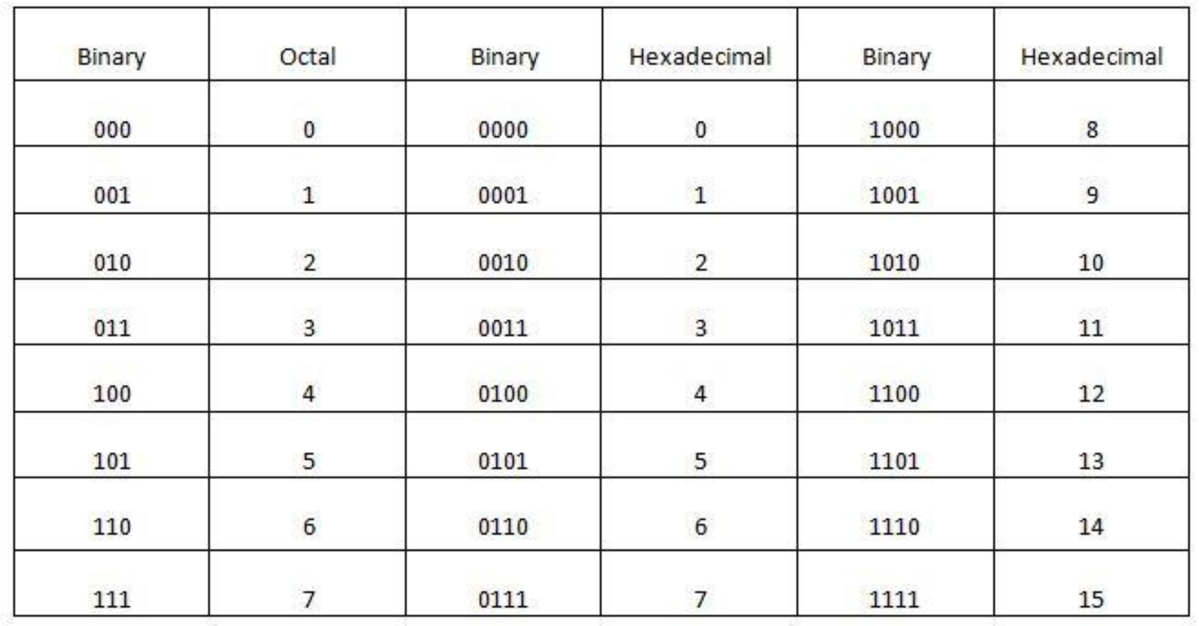
CHAPTER 2

TYPES OF CODES

This chapter aims at giving brief knowledge about the various types of codes that are currently implied in data transformation process. Let's have a look at each of these Simple codes.

2.1 VARIOUS CODES

Although the messages of an information source are usually encoded as binary sequences, the binary code is sometimes inconvenient for human to use. People usually prefer to make a single discrimination among many things. Evidence for this is the size of the common alphabets. For example the English alphabet has 26 letters, the Chinese alphabet has 37 letters, Phoenician alphabet has 22 letters, the Greek has 24 letters, Russian alphabet has 33 letters, cyclic alphabet has 44 letters, etc. Thus, for human use, it is often convenient to group the bits into groups of three at a time and call the octal code (base 8). This is given in table 2.1. When using the octal representation, numbers are often enclosed in parenthesis with a following subscript 8 for example decimal number 81 is written in octal as (121) as $81 = 1*8^2+2*8^1+1*8^0$. This transformation from octal to binary is so immediate that there is little problem in going either way. Again binary digits are sometimes grouped into four to make the hexadecimal code (table 2.1).



Binary	Octal	Binary	Hexadecimal	Binary	Hexadecimal
000	0	0000	0	1000	8
001	1	0001	1	1001	9
010	2	0010	2	1010	10
011	3	0011	3	1011	11
100	4	0100	4	1100	12
101	5	0101	5	1101	13
110	6	0110	6	1110	14
111	7	0111	7	1111	15

Table 2.1.

In this section of codes, we consider the linear codes but the linear codes are divided into three, the block codes, convolutional codes and their hybrid Turbo codes. Some examples of the linear codes are described below:-

2.1.1 Block Codes

A code is called a block code if the coded information can be divided into blocks of n symbols which can be decoded independently. These blocks are called the code words and n is called the block length or word length.

2.1.2 Hamming Codes

If C is an [n, k] code we define C' by

$$C' = \{y \in R / \forall\, x \in C\ [< x, y> = 0]\}.$$

If C is a code of length n over the alphabet F_q. We define extended code by

$$C''' = \{ (C_1, C_2, \ldots, C_n, C_{n+1}) / (C_1, C_2, \ldots, C_n) \in C_i \textstyle\sum_{i=1}^{n+1} \mathrm{Ci} = 0 \}.$$

Let G be the k by n generator matrix of [n, k] code C over F_q. If any two columns of G are linearly independent, that is they represent distinct points of PG (k - 1, q) then C is called a projectile code.

Let n = $(q^k - 1)/(q-1)$ then [n, n - k] Hamming code over F_q is a code for which the parity check matrix has columns that are pairwise independent that is the columns are maximal set of pair wise linearly independent vectors.

Note: Hamming codes are also perfect code.

Working of Hamming codes

Create a byte of data: 10011010

Create the data word, leaving space for parity bits _ _1_001_1010.

Calculate the parity for each parity bit.

Position 1 check bits 1,3,5,7,9,11

?_1_001_1010 even parity so set position 1 to a 0.

0?1_001_1010 even parity so set position 2 to a 1.

=0_1_001_1010

Position 2 check bits 2,3,7,10,11

0?1_001_1010 odd parity so set position 2 to a 1.

= 011_001_1010

Position 4 check bits 4,5,6,7,12

011?001_1010 odd parity so set position 4 to a 1.

= 0111001_1010

Position 8 check bits 8, 9,10,11,12

0111001?1010 even parity so set position 8 to a 0.

= 011100101010

Code word = 011100101010

2.1.3 Hadamard Codes

Hadamard codes are obtained from Hadamard Matrix. A Hadamard Matrix of order n is an I_n matrix of 1s and -1s in which HHT = I_n (in refers to the n × n identity matrix). Hadamard code based on H_2 as

$$\begin{pmatrix} H & \cdots & -H \\ \vdots & & \vdots \\ -H & \cdots & H \end{pmatrix}$$

and so on.

Hadamard code based on H_{16} the rate is 5/16 or about 0.31. But it is not very good. But it can correct 3 errors in any 16 bit encoded code and detect a fourth.

Hadamard codes based on H_{32} the generator matrix is 6 × 32. This code has even worse rate: 6/32 or about 0.19. But it can correct 7 errors in any 32 bit encoded block. This code is used on space craft Mariner in 1969; to broadcast pictures back to earth.

Working of Hadamard codes

To encode: we put all $\sigma_i \rightarrow C_i$ where σ_i are alphabets and C_i are elements of Hadamard codes.

To decode: If we receive the n-vector, v we search for a row of C_i of C that differs from V in almost n/4-1 positions.

(i) If C_i exists, then we decode σ_i (assuming at most n/4errors occur, this is always right).

(ii) Otherwise we detect an error (assuming most n/4errors occur, an error that is not corrected will always be detected). This code has 2n code words of length n. The minimum distance between any distinct code words is n/2. Under "minimal distance decoding", we can always correct n/4-1errors in n-bit encoded block and detect n/4errors.

2.1.4 Arithmetic Codes

The arithmetic operations are carried out with numbers represented in the number system with base r ($r \in N$, $r \geq 2$). For practical purposes the binary case ($r = 2$) and the decimal case ($r =10$) are most important. The first thing we have to do is to find a suitable distance function. One error in an addition can cause many incorrect digits in the answer because of carry. We need a distance function that corresponds to arithmetical error in the same way as hamming distance corresponds to misprints in words.

We shall consider code C of the form.

$C = \{AN/ N\in Z, 0 \leq N \leq B\}$, where A and B are fixed positive integers. Such codes are called AN codes.

Suppose we wish to add two integers N_1 and N_2 (both positive and small compared to B). These are encoded as AN_1 and AN_2 and then these two integers are added. Let S be the sum. If no errors have been made, then we find $N_1+ N_2$ by dividing by A. If S is not divisible by A i.e., errors have been made. We look for the code word AN_3 such that $d(S, AN_3)$ is minimal. The most likely value of N_1+N_2 is N_3. In order to be able to correct all possible patterns of at most e errors it is again necessary and sufficient that the code C has maximum weight at least distance $\geq 2e + 1$. As before, that is equivalent to requiring that C has minimum weight at least 2e+1. These properties of code C are based on the resemblance of C to the subgroup $H = \{AN /N\in Z\}$ of Z.

Working of an Arithmetic code

A code with A= 3, the operation of adding 15 and 16 will start by encoding both operands. This results in the operation R = 45 + 48 = 93. Then to find the solution, we divide 93 / 3 = 31. As long as B > 31, this will be a possible operation under the code. Suppose an error occurs in each of the binary representation of the operands such that 45 = 101101 → 10111 and 48 = 110000 → 110001, then P = 101111+110001+1100000. Notice that since 93 is 1011101, the hamming weight between the received word and the correct solution is 5 just after 2 errors. To compute arithmetic weight we take1100000 - 1011101 = 11 which can be represented as 11 = 2^0 + 2^1 or 11 = 2^2- 2^0. In either case, arithmetic distance is 2 as it is the number of errors that were made. To correct this error, an algorithm could be used.

2.1.5 Huffman Coding

Huffman coding or Data coding theory is an entropy encoding algorithm used for lossless data compression. For a set of symbols with a uniform probability distribution and a number of members which is a power of two. Huffman coding is equivalent to simple binary block encoding.

E.g. ASCII coding.

The term refers to the use of variable-length code table for encoding a source symbol (such as character in file) where the variable-length code table has been derived. In a particular way based on the estimated probability of occurrence for each possible value of the source symbol. It was developed by David A Huffman.

Steps in Huffman coding algorithm

1. The source symbols are arranged in order of decreasing probability. Then the two of the lower probabilities are assigned bit 0 and 1.

2. Then combine last two symbols and move the combined symbol as high as possible.

3. Repeat the above step until end.

4. Code for each symbol is found by moving backwards.

5. Calculation.

Efficiency is calculated using the equation

efficiency $\eta = H/ (L \log_2 r)$

where r is based on the system we are using

i.e., for binary r = 2 (0, 1)

for ternary r = 3 (0, 1, 2)

for quaternary r = 4(0, 1, 2, 3)

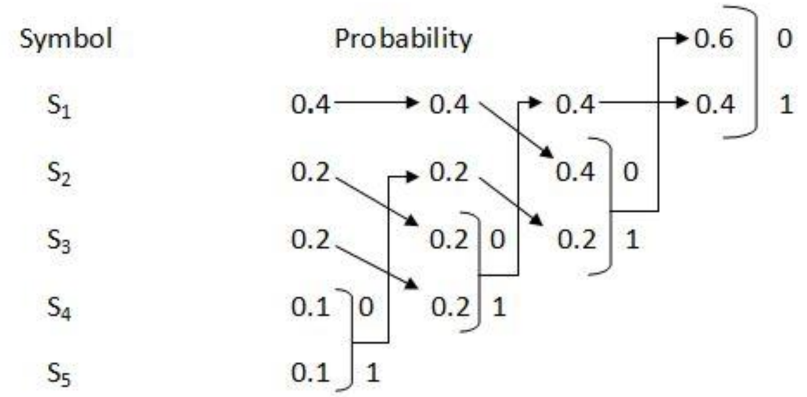
L is the average code word = $\sum_{i=1}^{n} p_i n_i$

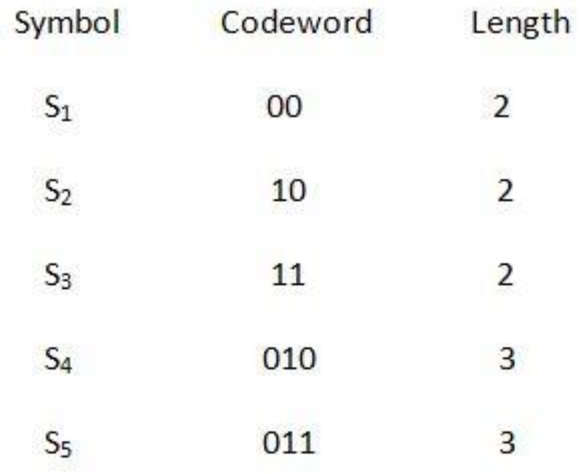
H is the entropy == $\sum_{i=1}^{n} p_i \log_2 (1/ p_i)$

Variance $\sigma^2 = \sum_{i=1}^{n} p_i (n_i - L)^2$

Example:

Alphabet with probability = {0.4, 0.2, 0.2, 0.1, 0.1} for symbols (s_1, s_2,..., s_5). Find the Huffman code and also find efficiency and variation.





Symbol	Codeword	Length
S1	00	2
S2	10	2
S3	11	2
S4	010	3
S5	011	3

First we want to calculate the entropy

Entropy H =$\sum$ $p_i \log_2 (1/p_i)$

= 0.4 $\log_2$ (1/0.4) +2×0.2 $\log_2$ (1/0.2) + 2×0.1 $\log_2$ (1/0.1)

= 2.1216 bits/symbol.

Average code word length L = $\sum$ $p_i\, n_i$

= 2×0.4+2 (2×0.2) +3(0.1×2)

= 2.2 bits/symbol.

Efficiency η = H/ (L $\log_2$ r) = 2.1216/(2.2 ×$\log_2$ 2) =96.4%

Variance $\sigma^2 == \sum \quad p_i (n_i - L)^2$

$$= 0.4\,(2-2.2)^2 + 2\times0.2\,(2-2.2)^2 + 2\times0.1\,(3-2.2)^2$$

= 0.16

Variance should be as low as possible.

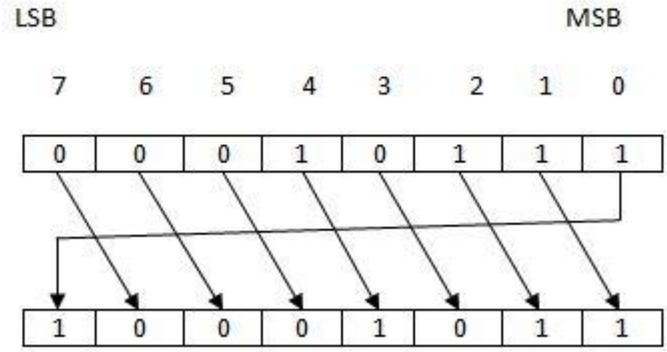
2.1.6 Convolutional Codes

The codes which we consider here are not block codes, i.e., the words do not have constant length. Although there are and analogies and connections to the block codes, there is one big difference that the mathematical theory of convolution codes is not well developed. This is one of the reasons that mathematicians find it difficult to become interested in this code.

At present it is one of the main tool used in area is convolution coding.

2.1.7 Cyclic Codes

In coding theory, a cyclic code is a block code, where the circular shifts of each code word gives another word that belongs to the code. They are error-correcting codes that have algebraic properties that are convenient for efficient error detection and correction.



If 00010111 is a valid code, applying a right circular shift give the string 10001011. If the code is cyclic then 10001011 is again a valid code word. In general, applying a right circular shift moves the least significant bit (LSB) to the leftmost position, so that it becomes the most significant bit (MSB) the other positions are shifted by 1 to the right.

Let C be a linear code over a finite field (also called Galois field) GF (q) of block length n

C is called a cyclic code if for every code word C = (C_1,C_2..., C_n) from C, the word (C_n, C_1..., C_{n-1}) in GF(q)^n obtained by a cyclic right shift of components is again a code word. Because of one cyclic right shift is equal to n -1 cyclic left shifts, a cyclic cord may also be defined via cyclic left shifts. Therefore the linear code C is cyclic precisely when it is invariant under all cyclic shifts.

Cyclic codes have some additional structural constraint on the codes. They are based on the Galois field and because of their structural properties they are very useful for error controls. Their structure is strongly related to Galois field because of which the encoding and decoding algorithms for cyclic codes are computationally efficient.

The Idempotent of C is a code word e such that $e^2 = e$ (that is e is an idempotent element of C) and e is an identity for the code, that is e C = C for every code word C. If n and q are co-prime such a word always exists and is unique. It is generator of a code.

Cyclic code follows following properties:

1. Linearity property.

2. Cyclic shifting.

Now let’s have a brief glimpse at these properties.

1. Linearity property

If we have two codes C_i and C_j then $C_p = C_i + C_j$ where C_p should be a code word.

2. Cyclic shifting

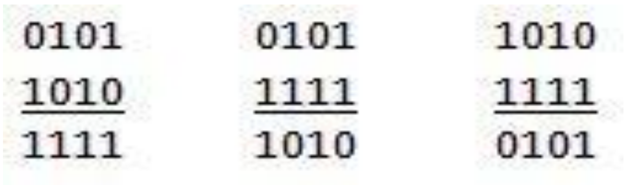
Code word = (C_1, C_2, C_3 ... C_n). After shifting left or right by any number of bits, resulting code should be a code word.

If code words follow above two properties then only the codes will be cyclic codes.

Examples:

1) A = {0000, 0101, 1010, 1111}. Is it a cyclic code?

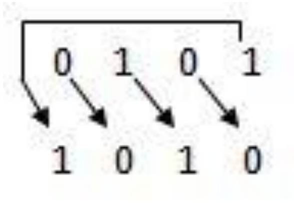
- Check the property of linearity.



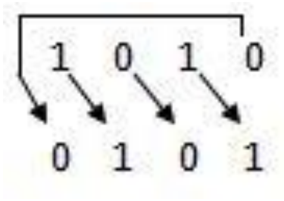
0101	0101	1010
1010	1111	1111
1111	1010	0101

As the answers belong to the set A, it follows the property of linearity.

- Check shifting property.



It is a codeword.



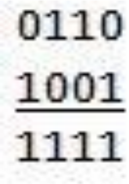
It is a codeword.

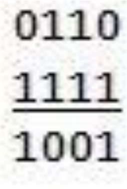
It follows shifting property.

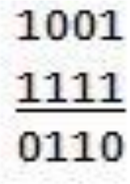
Therefore it is cyclic code.

2) A = {0000, 0110, 1001, 1111}. Is it cyclic code?

- Check the property of linearity.

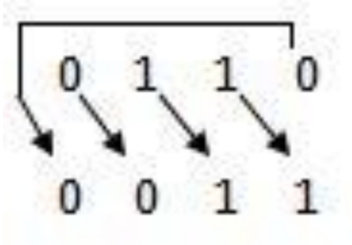

$$\begin{array}{c} 0110 \\ \underline{1001} \\ 1111 \end{array}$$


$$\begin{array}{c} 0110 \\ \underline{1111} \\ 1001 \end{array}$$


$$\begin{array}{c} 1001 \\ \underline{1111} \\ 0110 \end{array}$$

It follows the property of linearity.

- Check shifting property.



It is not a codeword, as it do not belongs to the set A.

Hence it does not follow shifting property.

So the above codes are not cyclic.

2.1.8 BCH Codes

In coding theory, the BCH codes or Bose Chaudhuri Hoquenghem codes form a class of cyclic error correcting codes that are constructed using polynomials over a finite field (also called Galois field). BCH codes were invented in 1959 by French mathematician Alexis Hoquenghem and independently in 1960 by Raj Bose and DK Ray Chaudhuri, and the name arises from initials of the inventor’s surname.

One of the key features of BCH code is that during code design, there is a precise control over the number of symbol errors correctable by the code. In particular, it is possible to design binary BCH codes that can correct multiple bit errors. Another advantage of BCH codes is the ease with which they can be decoded. These codes are using in many applications.

As there are many codes we have discussed some of the codes and their features.

CHAPTER – 3

APPLICATIONS OF CODING THEORY

3.1 APPLICATIONS

Although information theory was formed soon after 1948, when Claud Shannon published “ A mathematical theory of communication”. Information theory have its influence all over the world in the field of science and technology. As a conclusion to the project work it would be considered the best to point out the major classifications of application of the coding theory which is described as follows:

3.1.1 In Satellite Communications

Two of the most treasured resources in the satellite communication are power and bandwidth. Conserving this resource has been a major answer ever since Information Technology has its transmission via satellite became a reality. Error controlling coding often used to improve the transmission duality which is otherwise compromised by indifference and power limitations. Furthermore because of the considerable propagation delay in geostationary satellite links, FEC techniques tend to be more widely used than ARQ technique, which require data transmission. One of the remarkable feature of satellite communication system is that bit error occurs randomly which is considered to be significant advantages when applying FEC codes. In communicating systems, convolutional codes with constraint length 7 are widely used. Block codes are also applied in some satellite systems. Example of using block Court includes at (30, 15) Rs code for joint practical information distribution systems a(127, 112) BCH code for INTELSATV system and a (7, 2) code for airforce satellite communication wideband channel. Higher rates codes are expected to be widely utilised.

3.1.2 In Mobile communication

Mobile communications are defined as Communications involving mobile vehicles such as automobiles, trains, aeroplanes and marine results. Mobile Communications relay on wireless communication technology. In mobile environment burst error due to multipath fading are dominant. Since the band length available to each channel is strictly limited the code rate must be high. The MDS program has adopted a (11, 15) Rs code which may soon be adopted as a standard for tactical military communication links in NATO countries. The RS (16, 12) is adopted by North America rail roads for advanced train control system (ATCS) and it provides the best trade off between throughput, delay and implementation complexity. Mobile communications are expected to evolve to point where satellite access is possible allowing global personal communication ultimately, mobile satellite systems should be capable of providing basic communication services such as the voice and low rate to a very small terminal including handle held units. It is anticipated that power limitations will be even more important in future mobile satellite systems.

3.1.3 In Broadcasting

Conventional broadcasting such as AM/FM radio and IV is based on analogue transmission. However digital technology has been rapidly advancing along with IC technology. Microcomputers and their associated peripheral equipments 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 instantaneous properties.

3.1.4 In Genetics

Occurring in the cell's noisy environment DNA replication process is not error free. DNA is replicated several million times in a lifetime of a species and if there were no error correction mechanisms, the accumulation of errors during its lifetime and on a large scale, over millions of years of evolution would simply make genetic communication and hence life impossible. Capability of any error correction mechanism is limited and if the number of errors increase a certain threshold (denoted as the code's minimum distance) the error could not be detected or corrected. Therefore mutation could be viewed as an uncorrected errors in this regards.

CONCLUSION

In project, the development of basic coding theory and the state of art coding techniques have been reviewed. Application of coding to communications systems and future trends are also discussed. The theory of error recording is a very active area of our search. Error control systems have 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 designing of modern communicating systems. Over the past 10 years, VL 31 technology has reduced the cost best of coding systems by many orders of magnitude. Future generations of Technology are expected to continue this trend. Indeed more complicated scheme of coding will certainly become an economic necessity.

BIBLIOGRAPHY

Neubauer, A. Freudenberger, J. Kuhn, V. Coding Theory: Algorithms, Architecture and Applications, John Wiley & sons, Chichester, 2007.

Van Lint, J. H. Introduction to Coding Theory, 3rd ed., Springer, Heidelberg, 1991.