MATHEMATICS IN COMMUNICATIONS: INTRODUCTION TO CODING

A Public Lecture to the Uganda Mathematics Society

F F Tusubira, PhD, MUIPE, MIEE, REng, CEng

Abstract

Mathematical theory and techniques play a vital role in modern communications systems. The concept of coding is introduced and used as a vehicle to illustrate this close relationship. Source coding, security coding and channel coding are discussed, with major focus on channel coding.

1. INTRODUCTION

Engineering has always relied heavily on mathematics in the analysis and synthesis of all kinds of systems: the design of engines, structures, bridges, electricity generators and other large and small systems required mathematics to quantify stresses and performance, and also to establish safe working ranges. In all these cases, one however at some stage left the world of mathematics and completed the work in the practical world of experimentation and testing.

The field of communications differs from these traditional systems in that the mathematical and the practical world are intimately linked throughout the process. It is literally impossible to conceive modern communication without the integral linkage with mathematics. Maxwell, Faraday, Gauss, Shannon – and many other great names, gave birth to the principles on which many aspects of modern communication are based.

We range from vector calculus, which is a vital tool in the understanding of wave propagation and designing of transmission systems and antennas, to statistical analysis, statistical distributions, and probability theory, without which communications as we know it would not exist. Even something as basic as the Morse code uses the probability of occurrence of different letters of the alphabet to minimise, on average, the amount of data in transmitting a particular message.

Error correction coding has become so ubiquitous that we never even think about. Our CDs, whether holding music, multi-media entertainment programmes, or software, would not function as they do without the embedded Reed Solomon (RS) code.

This paper focuses on coding, particularly channel coding, to illustrate the use of mathematics to modern communications. It is arranged under the following sub-headings:

- Introduction
- The communication channel: Source, security and channel coding
- Examples from channel coding

2. THE COMMUNICATION CHANNEL: SOURCE, SECURITY AND CHANNEL CODING

The typical communication channel is shown in Figure 1 that shows an information source and an information sink. As the signal moves from the source to the sink, it is corrupted by noise (both internal and external to the system) and interference.
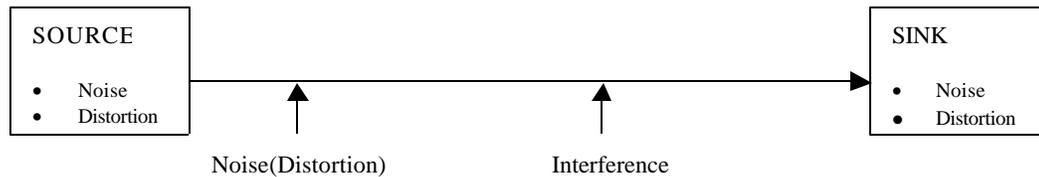
Figure 1: The Communication channel

Noise and interference will lead to corruption of the message. Recall the famous Second World War Story where the transmitted and received messages were, respectively:

"Please send reinforcements. I am going to advance"

"Please send three and four pence. I am going to a dance"

 This example focuses us on to the fundamental problem in communication. This is

"Reproducing at one point either exactly or approximately a message selected at another point" (Shannon, 1948)."

In addition to this, we want to utilise the communication channel efficiently, and, more often than not, to maintain privacy of communication.

The technique used in modern communication systems to achieve the desired performance is known as coding. Coding is can be generally described as the mapping of information to a set of symbols or numbers. The reverse process is decoding. Three major types of coding consequently used in communication systems are:

a) *Source coding*: This is aimed at minimising the amount of data that must be transmitted for acceptably correct message reception. It increases efficiency of utilisation of the channel.
b) *Security coding*: This maintains privacy of communication even if unauthorised people get access to the transmitted data.
c) *Channel coding*: This is aimed at ensuring that any corruption in the transmitted data that occurs in the channel can be detected and corrected. This is the main example vehicle used in this paper.

Shannon's statement of the fundamental problem in communications lays important groundwork for our discussion. It implies that that any message transmitted always comes from a possible or known set of messages. In other words, there is always *a priori* knowledge of the complete message set from which any message is transmitted. To give

a simple illustration, the correct reception problem at the receiving end reduces to the following decision:

If the received message is r, and there are M possible messages, m_i, i=0, M-1, we need to compute the probability that message r is received given that m_i is transmitted: $P(r|m_i)$. By working out this for the received message for all possible m, we can, using some decision guide, select which message m_i was transmitted. We can, for example, select the message that gives the highest probability as the message transmitted.

3. SOURCE CODING AND CHANNEL EFFICIENCY

Channel efficiency refers to the information throughput. All channels are bandwidth limited, and the more efficiently we utilise them, the lower the per unit cost and the better the financial bottom line. Channel efficiency is expressed as the number of bits that can be transmitted per second per unit bandwidth (bits/sec/Hz). To maximise this, we look for

- Ways of minimising the data transmitted for acceptably correct detection of the transmitted message.
- Methods of modulation that transmit as much data as possible per symbol

In minimising the data transmitted, we start by sampling the source (producing discrete time samples of the analog signal), then we quantise the samples (approximation into a set of *known* discrete levels), and finally we encode (represent each level by a code word).

Sampling is based on the sampling theorem:

A signal s(t) whose Fourier transform is zero outside the frequency interval $|f| < W$ can be uniquely represented by a set of samples of the waveform taken at intervals of 1/2W seconds.

Since the Fourier transform gives the energy spectral density, this is the same as saying that an analogue signal can be correctly reconstructed from samples taken at twice the maximum frequency component of the signal (obtained through Fourier analysis).

It is important, for efficiency, to reduce the amount of data transmitted. This can be done through:

- Data reduction: the source encoder removes redundancy in the data stream. If there is correlation between the source outputs (the outputs are not statistically independent), redundancy exists. For example, there is no point in transmitting a *u* after a *q* in the English language since the probability that a u follows a q is 1. Similarly, the picture of a newsreader on TV hardly changes: it is only the small variations of movement that need to be transmitted.
- Data compression: here some tolerable distortion is introduced to reduce the amount of data. Quantising, for example, introduces a quantisation error whose magnitude we can control to acceptable limits.

Source encoding refers to those techniques, which are really mathematical tools, used to minimise the amount of data that must be sent by the source for correct reproduction of information at the receiver.

4. SECURITY CODING

Security coding, or encryption coding, prevents unauthorised users from understanding the message.

As a simple illustration of this [1], we could use Table 1 to transmit a 0 or 1 in a binary symmetric channel (one in which 1 and 0 occur with equal probability):

Table 1
Look up Table for Security Coding

Key/ X:	0	1
A	00	10
B	01	00
C	11	01
D	10	11

The secret key sequence consists of the symbols A, B, C, and D generated using a perfect random generator. To transmit 0 when the key is B, one sends a 01. If the key is D, one transmits a 10.

5. CHANNEL CODING

5.1 General Statement

We accept that we cannot get rid of internal and external noise, interference, and non-linearity in our transmission system. Errors will therefore always occur. Channel coding modifies the code word to be transmitted such that:

- We can know when an error has occurred
- The error can be corrected

The second function is critical in modern communication systems. Where there is a dedicated channel, retransmission can be requested whenever an error occurs (eg Automatic Repeat Request, or ARQ, techniques). Dedicated channels exist in circuit switched environments, but not in packet switched environments. For the later, the channel code used must be such that errors are not only detected but also corrected without reference to the transmitter. This is known as Forward Error Correction (FEC). We shall now look at some introductory concepts with more rigour [2].

Consider the digital transmission system illustrated in Fig 2. We shall assume the channel is memoryless (the current output is only determined by the current input). A

transmitted information vector **i** is coded into a bit vector **c** that is transmitted over the channel. Due to noise and interference, the received vector **r** is different from **c**. We need to compute the probability P(**r|c**), the probability that **r** is received given that **c** was transmitted.
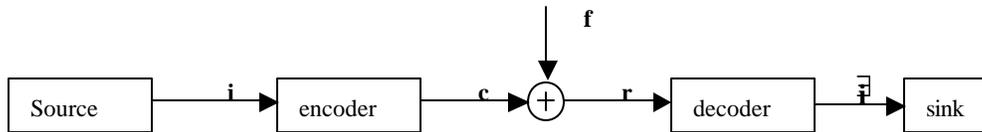

Figure 2: Digital transmission system

Let us examine this conceptually first. Consider a set of messages consisting of three possible messages:

$m_1 = \{0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\}$
$m_2 = \{1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\}$
$m_3 = \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}$

If the received code word is r = {0 0 0 0 0 1 0 0}, we can all visually determine which code word was most probably sent. If r = {0 1 0 0 0 1 0 0}, we can still make a pretty good guess. If r = {1 1 0 1 1 1 0 1}, it is a toss of a coin between m_2 and m_3. Without being rigorous about it, all we are doing is to compare the received and possible transmitted messages and selecting the one that has the least difference from the received message.

Another simple example is the following corrupted message that I am sure those good at Luganda can decode:

"Akantama akatano oyombe kale munkwaqa"

The importance of a priori knowledge about the possible messages is underscored here. We now need to express what we have done conceptually mathematically, including the introduction of some basic definitions.

5.2 Some Fundamental Concepts and Definitions

The ***modulo operation*** is defined for the term *a = cb + d,* such that
$$a = d \mod b, \tag{1}$$
Where $a, c \in N, b \in N$ and $d \in N_o$

In the above definition *Z ,N* and *No* are the sets of integer numbers, natural numbers and natural numbers including zero, respectively.

Block code: A block code uniquely maps a block of information symbols of length k {i_0, i_1, i_2, … i_{k-1}} to a codeword of length n {c_0, c_1, c_2, … c_{n-1}}. The number of *redundant symbols* is n-k, and the ratio k/n is the *code rate*. In a binary block code, a binary information bit stream is divided into independent blocks of bits for encoding.

A simple example of binary block codes is the single parity check code for which n=k+1. The last co-ordinate of the codeword satisfies Equation 2.

$$\sum_{j=0}^{n-2} i_j + c_{n-1} = 0 \bmod 2 \tag{2}$$

The *sum of two codewords* **a** and **c, a+c** is obtained by adding $a_i + c_i$, i=0, n-1 with the mod 2 operation applied to the sum of every co-ordinate.

The Hamming weight of a vector c is defined as the number of non-zero vector coordinates (Equation 3). 0<Hamming Weight<n, where n is the length of the vector c.

$$wt(c) = \sum_{j=0}^{n-1} wt(c_j), where \quad wt(\mathbf{c_j}) = \begin{cases} 0, & c_j = 0 \\ 1, & c_j \neq 0 \end{cases} \tag{3}$$

The Hamming distance between 2 vectors a and c, *dist(a,c)* is the number of coordinates where a and c differ (Equations 4 and 5).

$$dist(\mathbf{a,c}) = \sum_{j=0}^{n-1} wt(\mathbf{a_j + c_j}), where \quad wt(\mathbf{a_j + c_j}) = \begin{cases} 0, & c_j = a_j \\ 1, & c_j \neq a_j \end{cases} \tag{4}$$

$$dist(\mathbf{a,c}) = wt(\mathbf{a + c}) \tag{5}$$

We can now apply some rigour to our conceptual example. Clearly, we were comparing the received codeword with the possible transmitted codewords. We then selected the codeword which, at least in our estimation, has the *minimum distance* to the received codeword.

The minimum distance is easier to calculate through the *Hamming weight,* rather than directly, using Equation 5 (this equation applies to linear codes).

The error correction and detection capability of any coding scheme is determined by the minimum distance d of the code. This is the minimum distance between any two codewords of the code (Equation 6). The minimum weight also gives the minimum distance (Equation 7)

$$d = \min_{\substack{a,c \in C \\ a \neq c}} \{dist(\mathbf{a, c})\} \tag{6}$$

$$d = \min_{\substack{a,c \in C \\ a \neq c}} \{wt(\mathbf{a + c})\} \tag{7}$$

The general concept can be stated as follows:

A code has the ability to correct a received vector **r**=**c**+**f** if the distance between **r** and any other valid codeword a satisfies the condition:

$$\text{dist}(\mathbf{c}, \mathbf{c} + \mathbf{f}) < \text{dist}(\mathbf{a}, \mathbf{c} + \mathbf{f}) \tag{8}$$

$$\text{wt}(\mathbf{f}) < \text{wt}(\mathbf{a} + \mathbf{c} + \mathbf{f}) \tag{9}$$

$$\text{wt}(\mathbf{f}) \leq \left\lfloor \frac{d-1}{2} \right\rfloor \tag{10}$$

In Equations 8 to 10, $\mathbf{f}$ is the error vector. The inherent assumption here is that fewer errors are more probable so that we map the received vector to the nearest codeword.

5.3 The Hamming Bound

Note that the inequality in these equations defines a conceptual space surrounding a valid codeword point. All codewords (which are by implication not valid) within this space can be unambiguously mapped on to the valid codeword. We can extend this easily to a three dimensional concept and a definition of the Hamming bound.

Any binary code defined by (n, k, d) obeys the inequality in Equation 11:

$$2^k \left(1 + \binom{n}{1} + \ldots\ldots + \binom{n}{e} \right) \leq 2^n, where \quad e = \left\lfloor \frac{d-1}{2} \right\rfloor \tag{11}$$

Each sphere defines a correction sphere. The minimum diameter of the correction sphere corresponds to the minimum distance between codewords for any given code.

5.4 Syndromes

A syndrome, in English, means "a concurrence, especially of symptoms; … , characteristic of a particular problem or condition"

The idea behind channel coding is that we set up a mathematical mechanism for detecting symptoms, or the syndrome, of a corrupted codeword. We will give a simple example. Linear block codes obey Equation 12:

$$\mathbf{H.c^T} = 0 \text{ or } \mathbf{c.H^T} = 0 \text{ only for valid codewords} \tag{12}$$

H in this case is the parity check generated specifically for the code used.

The syndrome is obtained by using Equation 11 on the received codeword. A non-zero result indicates that an error has occurred. Further operations indicate the most probable error that would give the detected syndrome. The error is then corrected. For those who have been to the doctor, this is a very familiar process.

5.5 Decoding and Error Probability

In decoding, a decision guide has to be given to the decoding algorithm by the system designer. The algorithm will depend on the nature of the expected errors, the required performance, and other factors. The following are some simple illustrative examples:

Maximum Likelihood decoding: This method selects the codeword c that has the largest probability, $P(r|c)$ as the transmitted codeword. Where two codewords share this property, a random decision is made. This introduces the probability of an error or false decoding.

Symbolwise maximum a posteriori decoding: Each element of the codeword is independently decoded. It should be noted that when the resulting codeword is assembled, it may not be a *valid codeword*. In that case, decoding failure occurs.

Bounded minimum distance decoding: A requirement here is that r must lie within the correction sphere. We can have correct decoding, false decoding, or a decoding failure.

5.6 Code Generation

The modern communication channel contains a lot of computing power, and all the processes of coding and decoding are handled using algorithms programmed in hard or soft form into the channel. There will be a code generating algorithm, which can be a matrix or a polynomial, depending on the selected method.

5.7 Other Types of Channel Codes

It should be noted that we have presented only the most basic examples here as an aid to understanding the concepts. There are several sources that dwell at length on some of the modern and sophisticated coding techniques. All these make very interesting mathematical reading.

6. CONCLUSION

The intimate linkage between mathematics and communications has been demonstrated, using coding theory, specifically channel coding, as a vehicle for this demonstration. It is the hope of the author that this will re-awaken awareness of this important linkage, creating a basis for joint research and training programmes among the Electrical Engineering, Mathematics, and Physics disciplines within Uganda, and particularly within Makerere University.