INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR.379/22

COURSE ON BASIC TELECOMMUNICATIONS SCIENCE

9 January - 3 February 1989

Information Theory and Communication Engineering

P.A. MATTHEWS

University of Leeds, Dept. of Electrical and Electr. Eng., U.K.

INFORMATION THEORY AND COMMUNICATION ENGINEERING

PETER A. MATTHEWS

December 1988

1 INTRODUCTION

Engineering is the pursuit of the ideal, a goal rarely if ever achieved. In this pursuit we need to know what the goal is else we may waste our efforts in trying to attain an impossible objective. Information theory gives us the goal at which to aim in communications engineering but it dose not tell us how to reach that goal.

In engineering we deal with things that can be measured. In communications we must be able to measure the information being transmitted through a system. The amount of information we gain on receiving a message depends on the probability of our knowing the content of the message before it is received. If we receive a message each morning that the sun has risen we gain little information because we are pretty sure that the sun will rise. If we receive a message that there has been an eclipse of the sun we may gain little information because we can accurately predict when eclipses will occur. However if we a driving a car we must watch out for traffic signals because we cannot predict when they will change from green to red. If we miss the information given by the signal there may be a nasty crash.

*Professor, Department of Electrical and Electronic Engineering, University of Leeds, Leeds LS2 9JT, UK

The measure of information depends on our prior knowledge of the message before we receive it. We should note that in a technical sense the written words of the numbers making up a message are not important. It is the change in probability of our knowledge which is important.

2 A DIGITAL COMMUNICATION SYSTEM

Information theory can be applied to both analogue and digital communication systems but it is easier to explain the ideas in relation to digital systems. A block diagram of a simple system is shown in fig.1. The data source produces a data stream which passes through the communication channel to the data receiver. In the channel the data stream is distorted and noise is added to the signal. The job of the receiver is to detect the incoming signal and to produce an output which should reproduce the signal from the source. We would like to know how much information we can pass through the channel and whether we can receive the data without error. We need to define what we mean by information, rate of transmission and the capacity of the channel.

2.1 Rate of transmission

Suppose that the data stream is a stream of multi-level symbols transmitted at a constant rate, for example the output of a p.c.m. coder. If the number of levels in each symbol is n then for two symbols the number of possible combinations is n^2. For r symbols the number of combinations is n^r. If the symbols are transmitted at a rate of r symbols/sec and the message lasts for T seconds the total number of combinations is n^{rT}.

If we have a message which lasts twice as long we expect to be able to transmit twice as much information. If we use a logarithmic measure we get the result we expect. Then the information transmitted is given by

$$Information \propto rT \log n \ .$$

If we use logarithms to the base two we have

$$Information = rT \log_2 n$$

or the rate of transmission is

$$R = r \log_2 n \, .$$

Now $\log_2 n$, where n is the number of levels, is just the number of bits required to express the level as a binary number, e.g. if $n = 16$ then the number of bits required is four. Therefore the information rate is expressed in bits/sec is

$$R = r \log_2 n \, .$$

Note that so far we have not said anything about what limits the rate of transmission R or anything about the possibility of errors occurring in transmision.

2.2 Capacity of a channel

Information when it is transmitted through a channel may not be detected correctly at the receiver. There are errors in the received signal. To overcome these errors we can code the signal at the receiver and then we have to decode the signal at the receiver. This coding usually reduces the rate at which information can be passed through the channel but it reduces the probability of error. In a fixed time with coding more information can be passed through the channel without errors than if coding is not used. There will be a maximum rate of transmission R which the channel will pass without error. This rate is called the capacity of the channel C. The capacity is measured in bits/second.

If the information rate R is less than or equal to the channel capacity, $R \leq C$, then Shannon, [1], showed that the information can be transmitted with an arbitrarily low probability of error. If $R > C$ then the error rate cannot be reduced to zero.

3 INFORMATION CONTENT OF A MESSAGE

In communications engineering we are not concerned with the meaning, the semantics, of a message when we speak of the information content of

a message. What we are concerned with is the probability of being able to predict the symbols making up the message. If at the receiver we can predict the symbols of a message before the message is received then no information is gained on receiving the message. If we cannot predict the symbols then a maximum of information is received.

3.1 Information in a binary message

For simplicity suppose that only two symbols can be transmitted, 0 and 1. If the probability of 0 being transmitted is p then the probability of 1 being transmitted, q, is $q = (1 - p)$. We define the information on transmitting 0 to be $-\log_2(p)$ and the information on transmitting 1 to be $-\log_2(q)$. If the symbol transmission rate is r symbols/sec on the average rp $0's$ and rq $1's$ will be transmitted per second. The average information rate is

$$\begin{aligned} H(r) &= -rp\log_2(p) - rq\log_2(q) \\ &= r\,(p\log_2(1/p) + (1-p)\log_2(1/(1-p)) \end{aligned}$$

The average information per symbol measured in bits is

$$H = p\log_2(1/p) + (1-p)\log_2(1/(1-p)) \, .$$

When $p = 0$ or $p = 1$ then $H = 0$. We know the message before it is sent so no information is gained on receiving the message. If $p = 1/2$ then $H = 1$ and H is a maximum. The information per symbol is one bit which is exactly what the message is as we started by saying the symbols in this example were 0 or 1. The variation in H with p is shown in fig.2.

3.2 Information in a message containing random variables

We can generalise this concept of information for messages containing any number of random variables, [2]. Suppose we are dealing with a random variable X which can take up values from the set of values $a_1, a_2, \ldots, a_k$. We say that X has a sample space $\Omega_X = \{a_1, a_2, \ldots, a_K\}$. The *self-information* when X takes the value a_m is defined to be

$$h(a_m) = -\log_2(p_X(a_m))$$

where $p_x(a_m)$ is the probability that the value a_m will occur. If $p_x(a_m)$ is small then $h(a_m)$ is large as we expect. With this definition the self-information is positive, again as we expect.

We expect that if we have two independent events the total gain in information from the two events should be the sum of the information in the two events taken separately. If the two random variables are X and Y and $\Omega_Y = \{b_1, b_2, \ldots, b_N\}$ the self-information resulting from the observation of the two variables is, as the two variables are independent

$$\begin{aligned} h(a_m, b_n) &= -\log_2(p_{X,Y}(a_m, b_n)) \\ &= -\log_2(p_X(a_m)) - \log_2(p_Y(b_n)) \\ &= h(a_m) + h(b_n) \end{aligned}$$

If we have a sampled and quantized system we can convey the information about the state of the system at the sampling times by binary symbols. If there are only two quantizing levels we need one bit $(0,1)$, to convey the level. With two bits we can convey four levels, with three bits, eight levels and so on. By using $\log_2$ we have a direct relation between the measure of information and the number of bits needed to transmit that information. Even when the information is in the form of a continuous variable we can still use $\log_2$ to measure the information in bits.

When we send a long message made up of many symbols the probabilities of the separate symbols occurring may differ. However we can take an average to get the average information conveyed by the symbols in the message. The average is

$$\begin{aligned} H(X) &= E[-\log_2(p_X(X))] \\ &= -\sum_{x \in \Omega_X} pX(x)\log_2 p_X(x) \end{aligned}$$

The quantity H is called the entropy in X. It is the average information obtained on receiving the message *or* the average uncertainty about X before the message has been received *or* the average uncertainty removed by receiving X. This average information is measured in *bits*.

4 REFERENCES

1. SHANNON, C.E. and WEAVER, W.: "The mathematical theory of communication", 1949, University of Illinois Press, Urbana, USA.

2. LEE,E.A. and MESSERSCHMITT, D.G.: "Digital communication", 1988, Kluwer Academic Publishers, Boston, USA.

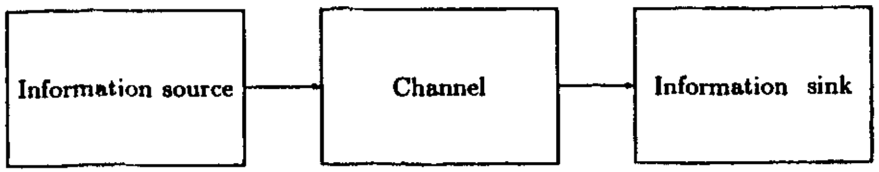

Figure 1: Block diagram of a simple communication system.

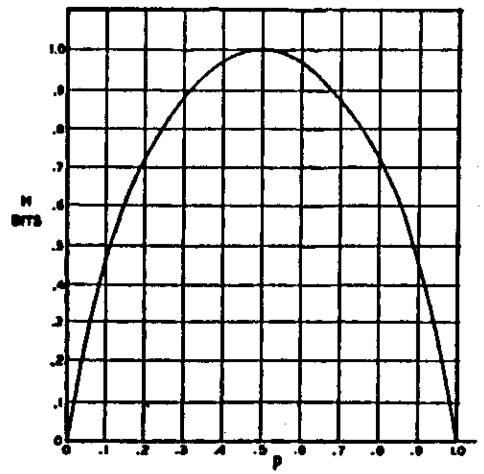

Figure 2: Variation in entropy H with probability for a binary signal.