ETC 006

# A brief introduction to Digital Data Transmission Techniques

1

B. M. Smith, B.E., Ph.D.

Australian Post Office, Engineering Training.

# ACKNOWLEDGEMENT

Reproduced from the Telecommunications Journal of Australia Volume 21 No. 2 dated June, 1971 with the kind permission of the Telecommunications Society of Australia.

# INTRODUCTION.

Data transmission has become a rapidly expanding component of the raffic in the telephone network since nany large public and private institutions have the need for a rapid transfer of information between widely scatered computing and data processing nstallations. Data transmission is not of course restricted to the telebhone network and is used widely in nilitary and space communication 'acilities.

- (a) a bank with a central computer linked directly with the tellers in
- the bank's branches;
  (b) a Stock Exchange quotation service for the rapid transfer of the prices of securities to stockbrokers and financial institutions;
- (c) time-sharing computers where several independent operators have access simultaneously by telephone lines to a large computer;
- (d) reservation networks for airlines, hotels and motels;
- (e) transmission of meteorological data.

Digital data transmission refers to he transmission of information in a ligital rather than an analogue form over a communication channel. The nformation to be transmitted may already be in digital form (e.g. the autput from computers) or may be an unalogue signal that is sampled at discrete time instants and then conzerted into digital form.

In this paper a discussion of some of the basic concepts underlying data ransmission is given and these concepts are illustrated by examples from lata transmission systems which use the telephone network. A list of definitions of data transmission terms is given in Table 1.

#### BITS, BAUDS AND BANDWIDTH.

A large number of data processing levices (e.g. computers) operate with binary notation in which information hat is fed into and out from the comouter must be coded into a sequence of elements, with each element in the sequence having only two possible itates, which are often referred to as r or 1. For example, the state of hese elements may be whether there s or is not a hole in a particular

Modem	: a unit for modulating the data into a form suitable for transmission over an analogue transmission channel and for the demodulation of the received signal.
Binary	: having two possible states.
Bit	: the elementary unit of information associated with the outcome of a binary event.
Symbol, Digit	: the basic signalling element of a data transmis- sion system; it will have two or more states.
Baud	: the number of symbols transmitted per unit time.
Intersymbol Interference	: overlapping of neighbouring symbols making their correct detection more difficult.

location in a paper tape or computer card.

Since data transmission is concerned with the digital transmission of information, a brief introduction to the unit of information is appropriate. Consider a given element in a binary sequence which can be either 0 or 1; then when this element has been received the uncertainty about its state has been removed, or alternatively some information has been received. If the two states are equally likely, the amount of information that has been received is defined to be one bit, which is a corruption of the words 'binary digit.' In short, a bit is the information associated with the outcome of an equally likely binary event.

Most simple communication systems transmit this binary data in binary form, as, for example, a telegraph system. Thus the communication rate can be expressed in bits/sec. With the advent of computers and other terminal equipment which can produce a very large volume of data in a given time, the need to increase the information rate of a channel occurs and initially this can be done just by increasing the repetition rate of the binary data sequence.

However, in a channel with a limited frequency bandwidth (e.g., a telephone channel occupying 300 Hz to 3400 Hz) there is some upper limit to the signalling rate after which the binary data elements or symbols interfere with each other and cause errors. This phenomenon occurs because the bandwidth of the signals no longer fits inside the channel bandwidth and is known as inter-symbol interference. The maximum rate at which the symbols can be transmitted without any inter-symbol interference at the sampling instants has been shown by Nvquist to be twice the bandwidth of the channel for an ideal channel with

Page 3.

no distortion, although in practice one must be satisfied with rates less than this.

Thus for an ideal channel of bandwidth W Hz, with no delay distortion, the maximum symbol rate that can be sustained without any intersymbol interference at the sampling instants is given by

Consequently to achieve higher data rates in a channel of limited bandwidth, more than two states or levels must be allowed in each symbol to be transmitted. Assuming that the total power of the transmitter is limited, this means that as the number of possible states in a symbol is increased, the states come closer together and are more sensitive to an error due to ncise. Hence, summing up the preceding discussion, the maximum symbol rate is determined by the channel bandwidth and the number of possible states in each symbol that can be reliably detected is limited by the ratio of the signal power to noise power. The signalling or symbol rate is specified in bauds and is only equal to the bit rate for a binary system.

To see how the information rate (in bits/sec) of a transmission channel with multilevel symbols can be calculated, consider the following simple examples:—

(a) if binary data is grouped into pairs, called dibits, then each dibit can have four possible states (see below), which can be identified with four level transmission:

TABLE 1: DATA TRANSMISSION TERMS.



From these simple examples it can be inferred that a symbol with m possible levels is equivalent to q bits where

 $m = 2^{q}$  (2) Taking logarithms to base '2' of both sides of (2) gives

 $q = \log_2 m$  . (3)That is, the information q, in bits of an m-level symbol, with each level equally likely, is given by (3). The logarithmic expression in (3) can be demonstrated to advantage by considering n consecutive symbols each of m possible levels; then intuitively we would expect n  $\log_2 m = nq$  bits of information to be contained in these symbols. But n symbols of m levels can have mn possible states and hence contains log<sub>2</sub> m<sup>n</sup> bits, which is identical to n log<sub>2</sub>m, as above.

The received signal after having been distorted and filtered by the communication channel will typically appear as in Fig. 1, if it is a binary signal. It is often useful to form an 'eye-pattern' from this signal by superimposing segments of the received signal many times on an oscilloscope, where each trace is triggered at the same point in time in relation to the symbol boundaries. The formation of a binary eye pattern is also shown in Fig. 1. In general the more open the eye-pattern the greater the immunity from error, although in the next section an exception to this is given. As expected, non-binary symbols give multi-level eye-patterns with the number of vertical eye-openings being one fewer than the number of levels in the symbol.

#### MODULATION.

Where there is a direct physical line with a d.c. path between the transmitter and receiver it is possible to transmit the data directly as a sequence of pulses whose rate and number of levels are determined by the bandwidth of the circuit and the noise on the line. The d.c. path is





necessary since the power spectral density of a sequence of random pulses extends to d.c. (Ref. 1, p. 320). For example, the power spectral density of rectangular binary pulses of random sign is shown in Fig. 2.

However, in many cases the communication channel may have a bandwidth that does not extend down to d.c.; for example, a telephone channel extends from about 300 Hz to 3400 Hz. This precludes the direct transmission of the baseband pulses which have a frequency spectrum extending down to d.c. and to circumvent this problem the frequency spectrum of the signal must be shifted up above the lower cut-off frequency of the channel. This operation is called modulation

This operation is called modulation and is performed by modifying some parameter (e.g. amplitude, phase, frequency) of a sinusoidal carrier by the

Page 4.

data sequence, where the frequency of the carrier is located in the passband of the channel. The location of the carrier frequency and the maximum symbol rate are determined by the passband of the channel.

At the receiver, the reverse process called demodulation is performed to obtain the transmitted data. Often the functions of modulation and demodulation are contained in the one unit of equipment, which is then referred to as a modem.

The characteristics of different types of modulation are now discussed in the context of data transmission, with special reference to data transmission over the telephone network. For several of the modulation techniques, relations between the error probability and the signal-to-noise ratio for gaussian (normal) noise are presented. Although other forms of noise are experienced on communication channels (especially on the telephone network where impulse noise predominates) these relations are widely used for the following reasons. Gaussian noise is easy to generate and has simple statistics, whereas impulsive noise has no such simple model. Furthermore, because it is the large peaks of the gaussian noise which cause the data transmission errors, the ranking of modems with respect to their error performance in the presence of gaussian noise often does not change when impulsive type noise is encountered.

# Frequency Modulation (FM).

This is a widely used method of modulation in data transmission, especially for low data rates up to 1200 bit/sec. in the telephone network. It is also called frequency-shift keying (FSK).

Digital frequency modulation can be carried out in two ways; firstly there can be a number of oscillators, one at each signalling frequency, to which the transmitter output is switched depending on the input data. The second technique switches the frequency of a single oscillator between the different signal frequencies, whilst maintaining continuous phase in the oscillator output. The second technique has an advantage in that the continuous phase in the transmitted signal gives a narrower frequency spectrum than the first method with its attendant phase discontinuities.

For a random binary data sequence the power spectral density of the cutput of an FM transmitter with continuous phase has been calculated (see Bennett and Davey, 1, chapter 18). Furthermore, it can be shown from this result that a peak-to-peak frequency deviation of about 0.6 to 0.7 times the bit rate leads to a well-



Fig. 3 — The Relative Power Spectral Density of a Random Binary FM Signal at 1200 Baud, with Signalling Frequencies of 1300 Hz and 2100 Hz.

shaped spectrum, resulting in a high efficiency of transmission in terms of occupied bandwidth. For example, the modems used in the A.P.O. DATEL service at 1200 baud have two signal-ling frequencies at 1300 Hz and 2100 Hz, giving a ratio of frequency difference to bit rate of 800/1200 = 0.67. Similarly at 600 baud, the carrier frequencies are 1300 Hz and 1700 Hz and the ratio is 400/600 = 0.67. Fig. 3 shows the relative power spectral density of the 1200 baud FM signal with random binary data and signalling

frequencies of 1300 Hz and 2100 Hz. It may be noted that there are no discrete frequency components at either 1300 Hz or 2100 Hz, but only a finite power spectral density at these frequencies. This explains why this signal sounds noise-like, with no discrete frequency components.

If the signalling frequencies of 1300 and 2100 Hz are retained but the data rate is reduced to 600 baud, a radically different spectrum as shown in Fig. 4 is obtained and which still requires about the same bandwidth.









Fig. 5. - Functional Diagram of a Digital FM Receiver.

A functional diagram of a typical receiver for this type of modulation is shown in Fig. 5. The bandpass filter has the important role of removing noise from the signal outside the signal's spectrum, without significantly distorting the signal. That is, the filter needs to have as narrow a bandwidth as possible without filtering too severely an FM signal with a spectrum as in Fig. 3.

The limiter has the function of hard limiting the input signal and any noise that passes through the bandpass filter. Thus the instantaneous frequency of the signal fed to the discriminator is the frequency of the sum of the FM signal plus the filtered noise. The resulting instantaneous frequency is not a linear function of the noise and consequently the calculation of the error rates due to the noise is a relatively difficult problem.

However, there is a simple approximate approach to the error behaviour of a binary FM system which gives a result that is relatively accurate and gives some insight into what causes an error.

This approach is based on the 'capture effect' in FM receivers whereby the average frequency of the sum of two sinusoidal signals is equal to that of the stronger signal (Ref. 2, Section 5.4). Hence if the noise envelope exceeds the signal envelope it will to a large extent determine the instantaneous frequency in the discriminator. Given a bandpass filter symmetrical about the two binary signal frequencies, we will assume an equal probability of the instantaneous frequency of the signal plus noise being closer to a given one of these frequencies when the noise envelope exceeds the signal envelope. If it is closer to the frequency opposite to that transmitted, an error will occur. For white gaussian noise the probability that the envelope of the noise exceeds the signal can be shown to be  $\exp(-\rho)$ where  $\rho$  is the signal-to-noise power ratio (SNR). Hence for this form of noise the probability of a bit error is given by

More detailed analysis gives a result that is very close to the above relation (see Ref. 3, Section 8.2).

Equation (4) has been plotted on Fig. 6, from which it can be seen that a SNR of 8.4 (9.2 dB), the error probability is  $10^{-4}$ . Note that the SNR is not plotted in dB. It is important to realise that the error probability is highly dependent on the FM signal envelope amplitude, although this parameter does not appear on the eye-pattern' of the discriminator when there is no noise. Thus when assessing the likelihood of errors in an FM signal, both the signal envelope and eve-pattern should be observed.

#### Phase Modulation (PM).

In data transmission over the telephone network at 2400 bit/sec, it is possible to use binary FM at 2400 baud, but with twice the bandwidth requirements relative to 1200 baud operation. However, it has been found in practice that it is better to use phase modulation at 1200 baud with four possible phases to give a data rate of 2400 bit/sec. This is done by grouping the incoming binary data sequence to be transmitted into dibits which are then encoded into four pos-



Fig. 6. - The Probability of Error as a Function of the Signal-to-Noise Ratio of a Binary FM System.

Page 6.

TABLE 2: ALTERNATIVE CODING CONVENTIONS.

Dibit Pattern –	Phase Change	
	Alternative A	Alternative B
00	0 0	45 °
01	$90\circ$	135°
11	$180^{\circ}$	225 °
10	270 °	315 °

sible phases depending on the particular dibit.

To avoid the problem of having to transmit a reference carrier phase against which the phase of the signal can be compared, the information is transmitted as a phase change and this form of modulation is often referred to as differentially coherent phase shift keying (DCPSK). The two internationally agreed alternative coding conventions used in these modems are shown in Table 2.

A phase modulated signal with the instantaneous phase jumps as set out in Table 2 will have a wide frequency spectrum and it is normal to adopt measures to limit the spectrum of the signal. For example, after phase modulation filtering may be used or alternatively the carrier may be amplitude modulated and the phase jumps applied when the carrier is a minimum. A typical power spectral density using the latter technique for a 1200 baud four-phase modem (2400 bit/sec.) is shown in Fig. 7.

There are several techniques that can be used in demodulating the phasemodulated signal, but when demodulating for example, a four-phase signal, the basic procedure is to divide the

Å

RELATIVE POWER SPECTRAL DENSITY

circle representing 360 deg. into four quadrants as shown in Fig. 8. The correct phase vectors are situated at the centres of the quadrants or 'decision regions' and if the demodulated phase falls inside a given quadrant, the phase vector associated with that quadrant is assigned to be the decision of the receiver.

The added noise can be represented by the vector, n(t), as shown in Fig. 8. As long as the resulting phase vector falls inside the quadrant, the correct decision will be made. For gaussian noise, the probability of an incorrect decision for four-phase modulation is plotted in Fig. 9 as a function of the SNR.

The coding of the phase changes shown in Table 2 has been chosen so that the two phase shifts on either side of a given phase shift corresponds to dibits differing in only one digit



Fig. 8. - The Decision Regions of a Four-phase Data Modem.

1-0 0-8 0-6 0-4 0-2 0 400 800 1200 1600 2000 2400 2800 3200 FREQUENCY (Hz)

Fig. 7. — The Relative Power Spectral Density of a Particular Four-Phase Data Modem Operating at 2400 bit/sec.

Page 7.

from the dibit corresponding to the given phase shift. Normally if the noise causes an error in the detection of the phase shift, the incorrectly detected phase shift will be one on either side of the correct one rather than 180 deg. different, and hence the bit error rate is minimised for a given symbol error rate.

A further interesting point emerges from the differentially coherent phase modulation system in that the comparison of these two successive phases can be made before or after these phases are quantised in the receiver. If the phases are first quantised and then successive phases subtracted to give the phase change, an error in detection in a given phase will affect two successive subtractions and hence two dibits will be in error with probably only one bit in error per dibit due to coding of the phase changes (see Table 2). This means that errors will tend to occur in pairs. However, if



the successive phases of each dibit symbol are first subtracted and then quantised to the changes shown in Table 2, then the likelihood of double errors is considerably reduced.

### Amplitude Modulation (AM).

Although amplitude modulation is perhaps the easiest form of modulation to understand and analyse, it has not been widely used for data transmission on the telephone network, especially at the slower speeds where frequency and phase modulation are more reliable and economical. However, at higher speeds multilevel AM is preferred, especially as the linear nature of this form of modulation facilitates the design of automatic equalisers for correcting the effects of frequency distortion.

In AM data transmission it is normal to conserve bandwidth by transmitting only one sideband of the AM signal (SSB) or one sideband with just a vestigial part of the other sideband (VSB). Alternatively quadrature AM can be used where two independent carriers, 90 deg. apart, are independently amplitude modulated, retaining both upper and lower sidebands, and then synchronously demodulated by two carriers 90 deg. apart as used by the transmitter. Double sideband transmission is necessary to prevent inter-channel interference between the two carriers. The capacities of SSB and quadrature AM are the same, because, although the necessity for double sidebands halves the baud rate of the latter system, there are effectively two parallel channels giving the same overall data rate.

When one considers Fig. 5 for fourphase modulation it can be seen that a given phase vector can be resolved into two components in quadrature; so that each of the four possible phase vectors can be made up of two quadrature two-level AM signals. Consequently four-phase modulation can be regarded as two-level quadrature AM; for example, the spectrum shown in Fig. 7 follows from considering the phase-modulated signal as quadrature AM.

A typical high-speed data modem for the telephone channel uses VSB AM with a carrier frequency of about 2.8 kHz and a modulation rate of 4800 baud with four levels giving a data

Page 8.

rate of 9600 bit/sec. With this carrier frequency the lower sideband will be the main one transmitted.

AM is a linear form of modulation and the output from an AM receiver can be regarded as a linear sum of the individual signal symbols transmitted. This fact is used in adaptive equalisers (see next section) which are often associated with this type of data modem.

Other variations on the above forms of modulation are used. For example it is possible to combine amplitude and phase modulation to give a modem with four phase states and two amplitude levels giving effectively eight possible levels per symbol.

#### CHANNEL IMPAIRMENTS.

When transmitting data over a communication channel there may be a wide variety of impairments which can degrade the channel and increase the likelihood of errors in the received data. However, it is possible to classify these impairments into several categories based on their effects rather than their causes. The most important channel impairments experienced in the telephone network fall into the following classifications.

Additive Disturbances. — These include:

(i) Gaussian (background) Noise. In general this is at a very low level and is not usually significant in data transmission on the telephone network. However, in some situations such as receivers for spacecraft communication this type of noise is dominant. This form of noise is widely used by designers and analysts of digital communication systems for the reasons given in the previous section on modulation.

(ii) Impulse Noise. This is due to switching transients in the telephone exchanges, lightning, and other similar disturbances. Its characteristics are far less defined than gaussian noise and there is not at the present time any simple model of this form of noise. The usual technique of measurement is to count the number of pulses in a given time that exceed fixed levels.

(iii) Sinusoidal and other periodic disturbances.

Linear Distortion. — As its name suggests this form of distortion linearly transforms the transmitted signal on its passage through the communication channel. The characteristics of linear distortion are usually specified in the frequency domain by amplitude and delay (phase) curves. Linear distortion can be caused by the attenuation of cables, the delay in carrier filters, echoes due to mismatches, etc.

Since speech is not sensitive to delay distortion, the telephone network has developed without this parameter being specified, but when data is transmitted over the network the delay distortion can very seriously degrade the signal especially for high data rates. Some form of equalisation of the delay will then be needed. The effect of linear distortion on data transmission is to cause interference between adjacent symbols of the received data. Normally this inter-symbol interference is insufficient to cause errors by itself but rather it makes the data more susceptible to noise disturbances.

The most common form of equalisation used on the network brings the attenuation and group delay of the channel within some given range over the bandwidth of interest. For example, Figs. 10 and 11 show the C.C.I.T.T. recommendation M102 attenuation and group delay characteristics for special quality telephone circuits.

As linear distortion can cause intersymbol interference, it is better to adjust the equalisation to minimise this interference at the sampling times of the receiving data modem. rather than attempt to fit the frequency response of the channel into some arbitrary criterion. Because the inter-symbol interference is more easily represented in the time domain than the frequency domain, the form of the equaliser that best minimises the inter-symbol interference is a transversal filter which gives weighted versions of the original signal at a range of delays. (See Fig. 12.) In fact, it is this form of equaliser that lends itself to being made adaptive by observing the inter-symbol interference in the data and adjusting the weighting at the various delays to minimise this interference.

Other Impairments: The communication channel may distort the signal in a non-linear fashion, which can give rise to inter-modulation products and harmonic distortion.

Other important classes of impairments experienced on a communication channel are frequency and phase varia-On the telephone network tions. these effects are the result of using carrier systems. A constant frequency shift due to a difference between the modulator and demodulator oscillator frequencies of a carrier system is not important in data transmission provided it is small. However, above about 10 Hz some data modems cannot track the frequency shift and hence cannot function.



Fig. 10. — The Bounds on Attenuation as Specified by the C.C.I.T.T. (Recommendation M102).





Page 9.



#### Fig. 12. — A Transversal Filter.

Channel phase variations, either random, periodic or isolated jumps, are thought to be a major impairment to the transmission of data at the higher rates on the telephone network. The effect of these phase variations commonly called phase jitter — on data transmission will depend on the type of modulation being used as well as the statistics of the phase jitter on a channel carrying a frequency modulated signal appears at the output from the FM receiver as additive noise.

As well as phase variations, a communication channel may be susceptible to amplitude variations such as fading and breaks. Once again the effect of these variations on data transmission will depend on the type of modulation being used, as well as the statistics of the amplitude variations.

#### TESTING DATA TRANSMISSION SYSTEMS.

In a synchronous data transmission, where the signal at the receiver is not fully regenerated after retiming, but is a squared up version of the demodulated analogue waveform, it is usual to characterise the effect of the communication channel on the received signal in terms of telegraph distortion. Basically this measures the variations of the data transition instants from the undistorted times (see Fig. 13).

It is conventional to allot 100 per cent. as the total bit duration and then as can be seen in Fig. 13 an error will occur when the distortion exceeds 50 per cent., assuming an ideal sampling clock at the middle of the bit period. More telegraph distortion can be tolerated in electronic equipment than in mechanical equipment which effectively samples the bit over a longer duration.

In most higher speed data modems, a clock waveform is extracted from the received signal and used to retime the data with resulting low telegraph distortion. However, the data may still contain errors due to noise and hence the criterion of performance of these modems in a given channel is their error rate. This error rate is specified, for example, as one error per 10,000 received bits or an error probability of  $10^{-4}$ . A block error rate is often measured where a block error is said to have occurred if there is one or more errors in a predetermined length of data called a block. The comparison of the block error rate and the bit error rate can give some idea of how the errors are distributed. For example, if the bit errors are widely distributed, then the bit and block



error rates are close together, while if the bit errors cluster together, then many occur in a given block and the two rates will differ widely.

Although this is a simple technique, it is relatively crude and for a better understanding of the distribution of bit errors a more detailed recording and analysis of the error events is needed. Also it may be noted that if multi-level modulation is used and then decoded back to binary, the error events occur in the multi-level symbols and this will then impose certain properties on the binary error distribution. For example, the four-phase differential phase-keying system often gives double errors because a comparison between the present and the previous symbols is used. Hence an error in the present symbol will affect this decision and the following one, when the present symbol becomes the previous symbol.

For testing data transmission systems, a data sequence with the following properties is needed. It must be easily generated and have characteristics similar to that of random data; furthermore it should have a finite length and be predictable so that the same sequence can be generated at the receiver, synchronised to the incoming sequence and matched bit by bit to detect errors in transmission. Α maximal length shift register sequence commonly referred to as a binary pseudo-random signal fulfils these reuirements and is widely used in A.P.O. data tests. The usual sequence length is 511 bits generated by a 9-bit feedback shift-register.

Although much information about the performance of a data transmission system can be obtained from the distribution of errors, a more complete understanding of the causes of these errors is only obtained when the analogue waveforms of the signal and noise are observed, especially at the point where the decision is made. For example, observation of the eye-pattern may yield considerable information about the distortion of the signal and its immunity to noise.

# CONCLUSIONS.

Data transmission in the telephone network is growing rapidly in both quantity and system complexity as users aim for higher rates of reliable information transfer. This paper briefly outlines some of the charac-teristics of currently used data trans-miss on systems, especially in regard to the different modulation and de-

modulation techniques that are possible. Of course not all techniques have been discussed and the reader should refer to the references cited if further information is required.

## ACKNOWLEDGEMENT.

The author wishes to thank Mr. J. Semple for his constructive criticism during the preparation of this paper.

# REFERENCES.

REFERENCES. 1. W. R. Bennett and J. R. Davey, 'Data Transmission'; McGraw-Hill Book Co., New York, N.Y., 1965. 2. J. J. Downing, 'Modulation Sys-tems and Noise'; Prentice-Hall, Engle-wood Cliffs, N.Y., 1964. 3. R. W. Lucky, J. Salz and E. J. Weldon, 'Principles of Data Communi-cation'; McGraw-Hill Book Co., New York, N.Y., 1968.

Page 11.

