NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

Phase Modulation and Convolutional codes

Graduation Project
EE- 400

Student: khaled M. Rakeb(971391)

Supervisor: Mr. Jamal Fathi

Nicosia- 2002

ACKNOWLEDGMENTS

First I want to thank Mr. JAMAL FATHI to be my advisor .under his guidance, I successfully overcome many difficulties and learn a lot a bout Phase Modulation and Convolutional Code. in each discussion, he explained my questions patiently, and I felt my quick progress from his advises. He always helps me a lot either in my study or my life. I asked him many questions in electrical and communication and he always answered my questions quickly and in detail.

Special thanks to Ramiz and Ibad. With their kind help, thank to faculty of engineering for having such a good computational environment.

And also I want to thank all my friend in NEU: Hamdey. Hamada and Karam being with them make my 3.5 years in NEU full of fun.

Finally I want to thank my family, especially my parents. Without their endless support and love fro me, I would never achieve my current position. I wish my mother lives happily always and my father in the heaven be proud of me and also I would thank some one who joined me and shared me happiness time I spend with .

ABSTRACT

There are, however, many devices that possess a multiplying capability as just one of the many features of a complicated non-linear characteristic. For our application this would mean an output of the form m (t) [cos wt] accompanied by higher-order multiplicative terms.

The analogue-To-Digital conversion introduces an irremovable error known as quantization noise which depends upon the number of levels and hence upon the number of digits in the code. the larger the number of levels the smaller the noise. It will be seen that for a given number of code digits, which is determined by the available bandwidth, the quantization distortion can be minimized by choosing non-uniform spacing of the levels to suit the statistical properties of the signal

This chapter is concerned with establish. In this chapter we will present some recent results on combinations of many-level P5K and CPM with convolutional codes. Covers binary and quaternary CPFSK with rate 1/2 convolutional codes, and presents combinations with the best free Euclidean distance. covers eight-level CPM with rate 2/3 codes, l6-level CPM with rate 3/4 codes, presents some simulations of eight- and 1^6-level CPFSK with rate 2/3 and rate 3/4 codes. Viterbi detection is used throughout.

The optimum pulse shape and filtering of the digital signal, in order that the received pulse sequence may be interpreted with the minimum error. The overall transmission system and the following analysis will deal with the base band signal, i.e. a signal suitable for line communication; however, the basic principles apply equally to radio systems which require, in addition, modulators and demodulators operating at the radio carrier frequency.

CONTENS

CHAPTER ONE

GENERATION AND RECEPTION OF MODULATED SIGNALS

1 .1 Amplitude Modulation

There are, however, many devices that possess a multiplying capability as just one of the many features of a complicated non-linear characteristic. For our application this would mean an output of the form m (t)[cos wt] accompanied by higher-order multiplicative terms. There are, in the main, two solutions to this problem. One is to follow a 'non-perfect' multiplier with a filter that allows only the wanted components to pass, whilst the other is to arrange a pair of non-linear elements, having as near as possible the same characteristics, in a balanced circuit so that the unwanted components cancel out. The degree of suppression depends upon the closeness of matching, and in practice it is usually found necessary to provide further suppression by adding an output filter-this requirement, however, need not be as stringent as that for the first method.

Before a comparison of the two methods can be made, it is necessary to consider a typical non-linear device such as a semiconductor diode, and to establish its input/output relationships for conditions of small and large signal operations. The analyses of the following sections deal

Exclusively with the diode, since both of the above-mentioned possibilities can be realized and, furthermore, the fundamental ideas can be readily applied to modulators using bipolar transistors and MOSFETs.

1.1.1 Small-Signal Modulators

When a diode is operated over a restricted portion of its characteristic (see figure. 1.1) the current i(t), and hence the voltage $v_{out}(t)$ across a resistive load (see figure. 1.2 (a)), is related to the input voltage v(t) by the polynomial

$$V_{out}(t) = i(t)R = av(t) + bv^2(t) + cv^3(t) \qquad (1.1)$$

The coefficients a, b and c may be assumed constant and higher-order terms sufficiently small to be neglected. Consider the input to be made up of a modulating voltage m (t) and a sine wave $E_c \cos w_c t$. The output is

$$V_{out}(t)\ a\ [E_c \cos wct + m(t)]$$

$$+b\ [E_c \cos w_c t + m(t)]^2 \qquad (1.2)$$

$$+c\ [E_c \cos w_c t + m(t)]^3$$

$$= a E_c \cos w_c t + am(t)$$

$$+b\, m^2(t) + 2bm(t) E_c \cos w_c t + bE_c^2 \cos^2 wct$$

$$+ c\, m^3(t) + 3cm^2(t) E_c \cos w_c t + 3cm(t) E_c^2 \cos^2 w_c t + cE_c^3 \cos^3 w_c t$$

We see that the term giving rise to the required upper and lower side bands is $2bm(t) E_c \cos w_c t$. All the other terms produce outputs harmonically related to the modulating signal, carrier and modulated carrier. For example, those of the form $m(t) E_c^2 \cos^2 w_c t$ are responsible for side bands at twice the carrier frequency and although undesirable it is relatively simple to remove them by filtering. The same cannot be said of the outputs due to the term $cm^2(t) E_c \cos w_c t$. For the purpose of illustration let us assume that

$$M(t) = \cos w_1 t + \cos w_2 t \qquad (1.3)$$

Then $cm^2(t) E_c \cos w_c t = C(\cos^2 w_1 t + 2\cos w_1 t \cos w_2 t) E_c \cos w_c t$ and an expansion yields components at frequencies f_c, $f_c \pm 2f_i$, $f_c \pm 2f_2$, $f_c (f_1 \pm f_2)$ and $f_c \pm (f_i - f_2)$. It will be appreciated that these can fall within the bandwidth occupied by the required DSB output and therefore cannot be removed by filtering. We conclude that the single diode modulator is not a practical proposition, since these components, due to intermediation within the base band signal, would cause interference to the information contained in the modulated output.

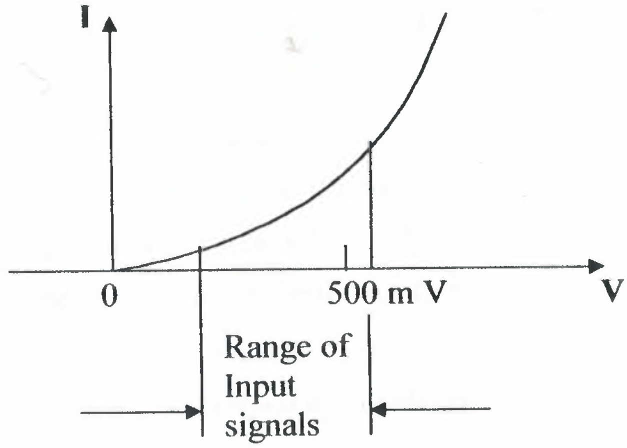


Figure1.1 typical small signal diode characteristic

By using two matched diodes in a circuit that is symmetrical (or balanced) with respect to the carrier input (see figure 1.1) the unwanted intermediation components can be cancelled out. In order to [E_c cos $w_c t$ + m (t) and i_2 (t) is the current through diode 2 due to the applied voltage (E_c cos w_c t – m (t)). Substituting these two conditions into equation. (1.1) and then using equation. (1.4) gives

$$v_{out}(t) \;\alpha\; 2am(t) I - 4\, b\, m(i)\, E_c \cos w_c t$$

$$+ 6cm(t) E_c^2 \cos^2 w_c t + 2cm^2(t)$$

It is to be noted from the above equation that the small-signal balanced modulator has produced a DSBSC output together with unwanted terms that can be filtered out.

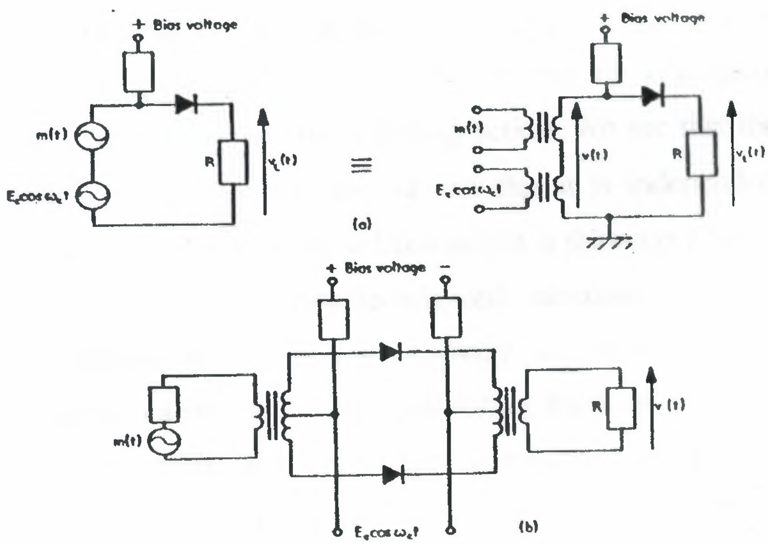


Figure 1.2 (a) signal-diode modulation, and (b) a balanced modulator

1.1.2 Large Signal Modulators (Piecewise-Linear Modulators, balanced)

It is also possible to obtain a method output using the circuit of figure 1.2.by operating above instead of law level carrier (peak value of hindered mill volts) a relatively high level of 3- 4 volts can be used to switch the diodes from a reversed biased conduction to one of forward concretion the d.c biasing no longer being required. If the diodes are driven well into the forward based region each will have a current voltage relationship that is essentiality liner due to the diode acting more like a resistor than a semiconductor under these large signal conditions the overall characteristic is made up of two liner portions see figure (1.3)

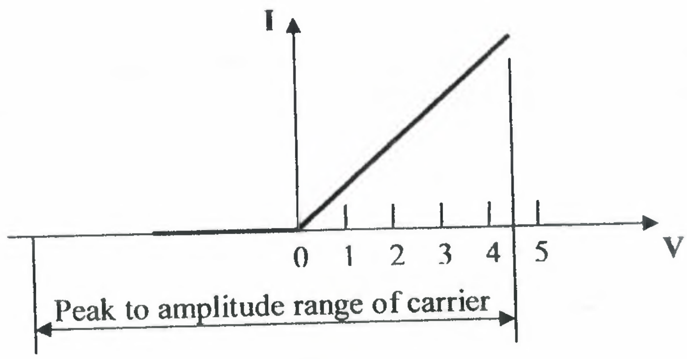


Figure 1.3 typical large-signal diode characteristic.

And hence we have the explanation of the term piecewise linear modulator. The switching of the diodes causes the low-level modulating voltage to appear across the output every other half-cycle of the carriers and therefore the bracketed expression is the Fourier series representation of the switching action. We see that the term 2[m (t) cos w_c t] / π will provide a DSBSC output and that it is independent of the peak amplitude of the carrier E_c. A typical modulated output is shown in Figure. 1.4 (d).

Another version known as a double-balanced modulator has four diodes-the additional two connected in a lattice arrangement as shown in Figure. 1-S. The switching action of the carrier causes the modulating input m(t) to appear across the output with what is effectively a reversal of the connections every other half-cycle of the carrier. The switching waveform is illustrated in Figure. 1.4(e) and the output is

$$v_{out}(t) = m(t) \times \begin{bmatrix} \textit{squraed wave form} \text{ due to switching} \\ \text{of the diodes by the carrier} \end{bmatrix}$$

$$= m(t)\left[\frac{1}{2} + \frac{2}{\pi}(\cos\omega_c t - \frac{1}{3}\cos 3\omega_c t + \frac{1}{3}\cos 5\omega_c t + ...\right] \quad (1.6)$$

$$= \frac{1}{2}m(t) + \frac{2}{\pi}m(t)\cos\omega_c t + \text{terms involing } 3f_c, 5f_{c,} ....etc.$$

$$v_{out}(t) = m(t) \times \begin{bmatrix} \textit{squraed waveeform} \text{ corresponding} \\ \text{to the switshing reversale} \end{bmatrix}$$

$$= m(t)\left[\frac{4}{\pi}(\cos\omega_c t - \frac{1}{3}\cos 3\omega_c t + \frac{1}{3}\cos 5\omega_c t + ...\right] \quad (1.7)$$

$$= \frac{4}{\pi}m(t)\cos\omega_c t + \text{terms involing } 3f_c, 5f_{c,} ....etc.$$

It is seen that the only difference between this output and that of the balanced modulator, given by equation. (1.6), is a component at the modulating frequency m (t).

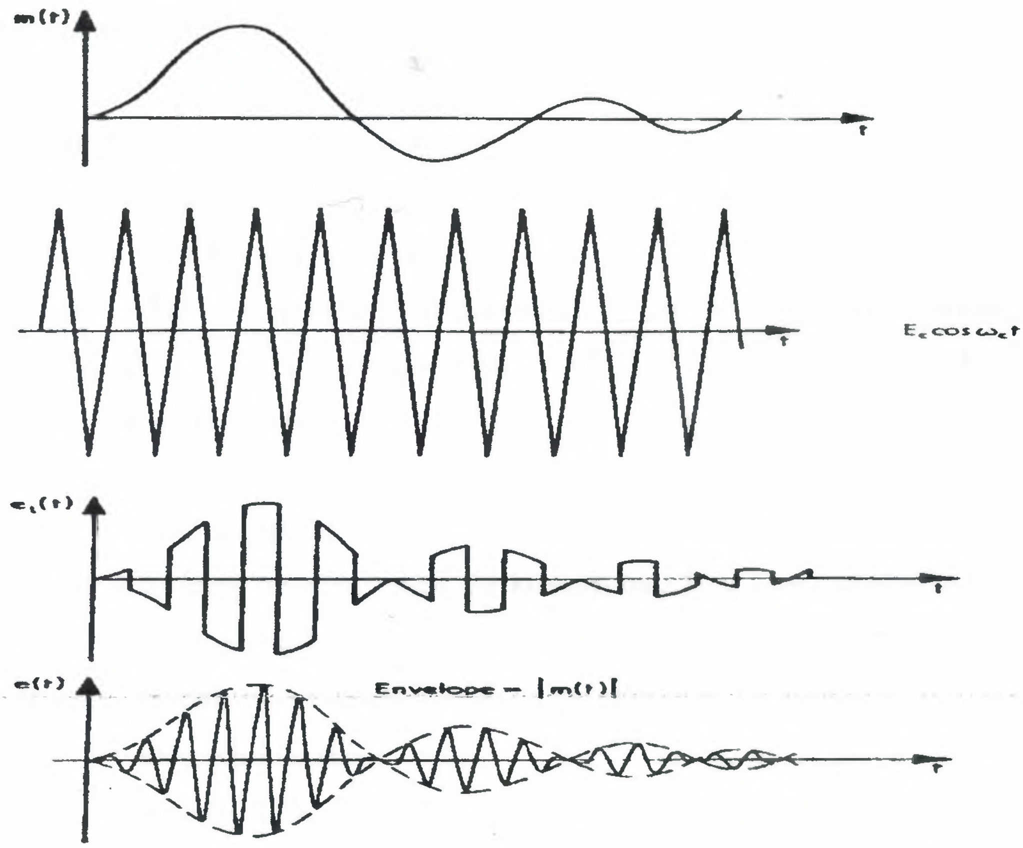


Figure 1.4 Balanced and Double-Balanced Modulator Waveforms. (a) Modulating Signal m (t). (b) Large-signal Carrier E_c cos w_c t. (c) Output from a Double-Balanced Modulator.

(d) DSBSC Filtered Output.

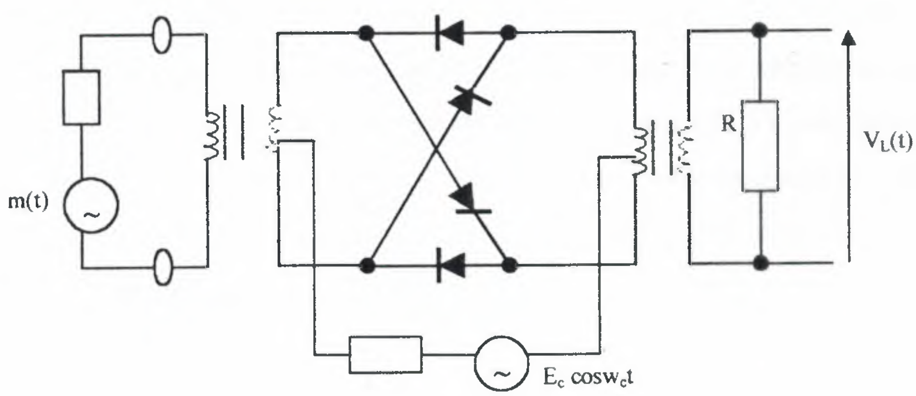


Figure 1.5 Double-Balanced Modulator.

1.1.3 Comparison of Small-Signal and Large-Signal Modulators

After reading the previous two sections, the large-signal modulator would appear to be the obvious choice for generating a DSBSC signal. Before dismissing the small-signal modulator completely it is worth

While to state firstly the conditions under which it can be used and secondly to enumerate the advantages of the large-signal circuit. The latter becomes the obvious choice when diodes of a discrete form are used, as it is not practicable to select pairs having matched coefficients a, b and c necessary requirement to avoid intermediation when using a small-signal modulator. The matching of diodes for large-signal application presents little problem, since the forward bias I- V characteristic does not differ appreciably from diode to diode within a production batch, and the large discrepancies in reverse bias characteristics can be overcome by shunting each diode with a high resistance, say 100 kΩ.

Balanced modulators in the form of integrated circuits using bipolar transistors in place of diodes-simply for convenience of manufacture can be made up of pairs of elements having characteristics that would be considered matched if exhibited by a pair of discrete devices. Nevertheless, these circuits are used solely as large-signal modulators mainly due to the convenience of not having to supply a D.C. Bias and also, of course, of making use of the overriding advantage that spurious outputs due to intermediation are extremely small even when the non-linear elements are not identical. But integrated circuits made up of univocal transistors (FETs and MOSTs) do, however, lend themselves to small-signal modulation. Of these transistors exhibit an extremely good square-law characteristic, i.e. the bv2 (t) term of equation (1.1) is several orders of magnitude greater than the other terms and, therefore, inter-modulation outputs are virtually non-existent. Further reference is made to these devices in the section dealing with frequency changers and demodulators.

1.1.4 High-Power Modulators

Operating an amplifier in a nonlinear mode can generate a full AM signal. The term high-power is used to indicate that the modulated signal can be formed at the required power for transmission and that no further amplification or signal processing is necessary. This method has been used extensively in mobile systems having outputs from tens to Hundreds of watts and also for broadcast and point-to-point radio links with outputs of many kilowatts.

The output stage of the transmitter is used as the modulator and consists of a thermion valve (or valves) biased to operate as a Class C amplifier. Its output power is made to vary in sympathy with the modulating input derived from a linear audio amplifier (see Figure. 1.6). The main attraction of this arrangement is its extreme simplicity compared with other methods of generating a full AM signal, and whilst its use for mobile purposes can be tolerated it is not recommended for the generation of very high powers, because the spurious modulator outputs. Those at 2fc, 3fc, etc.-cannot be filtered out and may cause considerable interference to other users of the radio spectrum.

A full account of Class C amplifiers and high-power modulators is well documented elsewhere

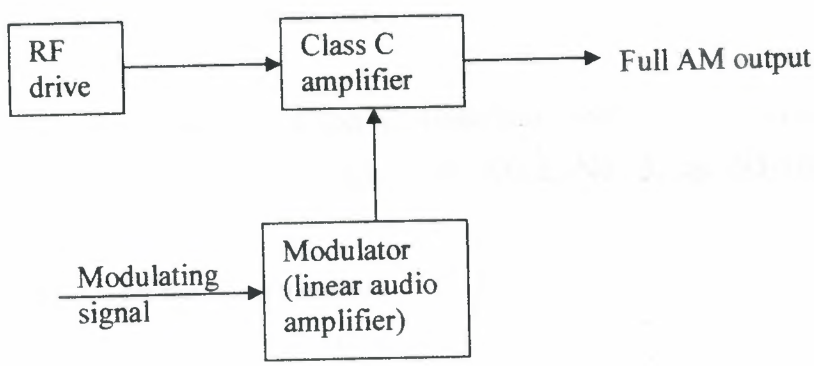


Figure 1.6 High Power Modulators .

1.1.5 SSB Generation

The obvious way to generate a SSB signal is to pass the DSB output from a balanced Modulator through a filter that rejects one of the side USB

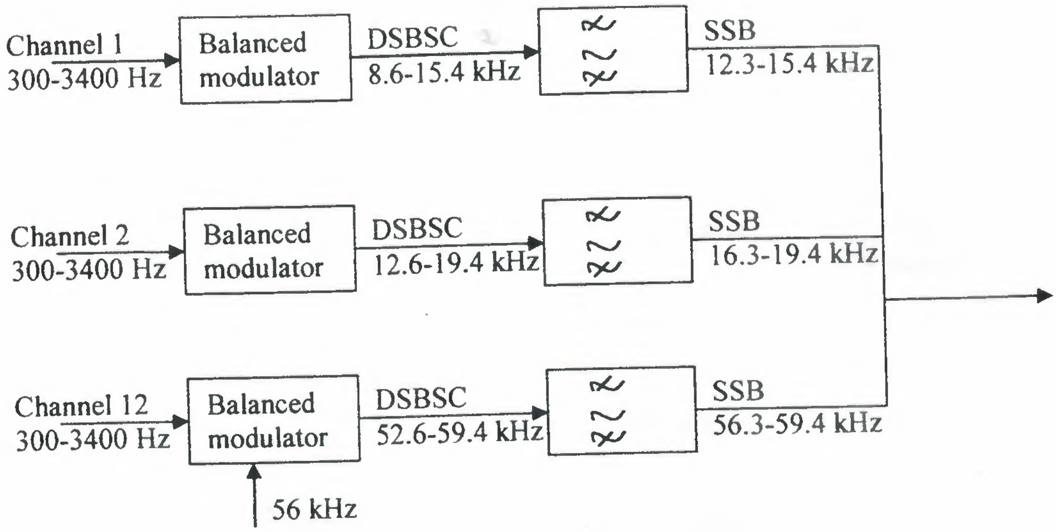


Figure 1.7 Formation of FDM Basic Group.

1. Electronic and Radio Engineering, F. E. Term an, Chapter 13 (McGraw-Hill 1955).

2. A Non-Linear Theory of Class C Transistor Amplifiers and Frequency Multi pliers, R. G. Harrison, Trans. I.E.E.E., Vol. SC-2, No. 3, pp. 93-102 and p. 126, September 1967.

Bands. Its application for the generation of a FDM Basic Group for line communication is illustrated in Figure. 1.7. The outputs from the balanced modulators are filtered to remove the lower sideband of each channel.

It is also common practice to use this method for radio systems. However, the removal of a sideband that is closely spaced to the wanted sideband presents insuperable difficulties at very high carrier frequencies; the separation can only be successfully achieved at a relatively low carrier frequency fc, of less than 2MHz. The technique is illustrated in figure. 3.8, and it is also shown how two independent sidebands (ISB), symmetrically positioned about the carrier, can be formed. It is basically an SSB system, but has the economic advantage of allowing one piece of equipment to handle two messages. The ISB form of signal is used mainly for the transmission of telephony in the HF (330 MHz) band. The propagation characteristics of this part of the spectrum, together with the limited number of channels, preclude the use of wideband systems carrying a large number of messages, whereas the use of ISB to provide two (or four) channels on the same link is feasible and economically desirable.

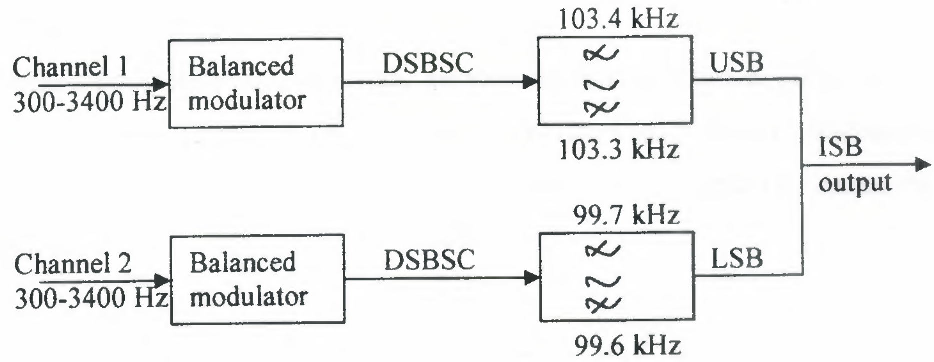


Figure 1.8 Formation of an ISB Signal.

Having formed the SSB or ISB output at a relatively low carrier frequency and low power, usually of the order of 100 mW-1 W, further processing is needed to transpose the signal to a higher frequency and to amplify it to the required output level. It follows that the frequency translation must involve another process of modulation where the modulating signal is the SSB version of the original base band input m (t). This modulator (or frequency changer) will produce a DSB output symmetrically disposed about the new carrier frequency fc, (see Figure1.9). The frequency separation of the sidebands is twice the carrier

Sideband. If the final output is required in the VHF or UHF spectrum, yet another stage of frequency changing is needed-the number of transpositions being determined solely by the final carrier frequency and

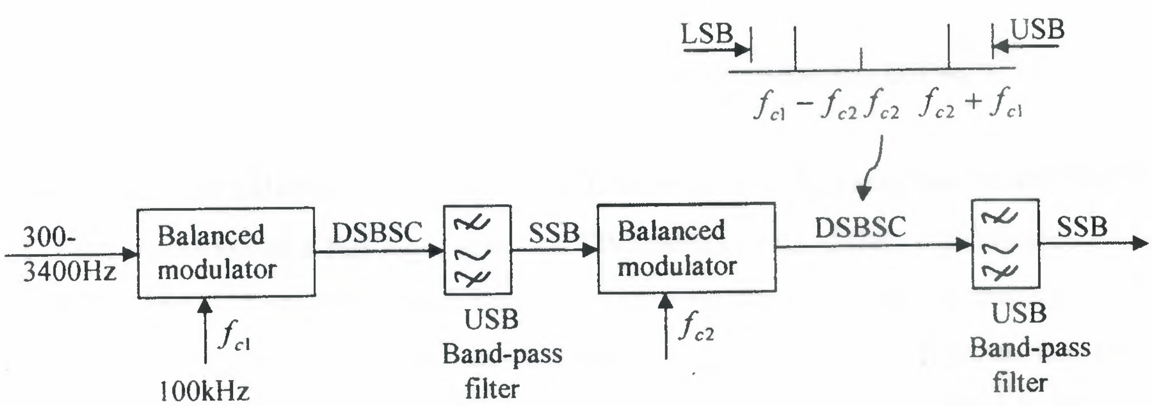


Figure 1.9 Generation of an SSB Signal by Successive Modulation and Filtering.

The ability of the filters to separate the upper and lower sidebands after each stage of modulation.

A phase-shift method of generating an SSB signal (See Figure. 1.10) does not require filters having sharp cut-off frequency characteristics. Two balanced modulators generate DSB signals that, when added together, result in the cancellation of either the lower or upper sidebands.

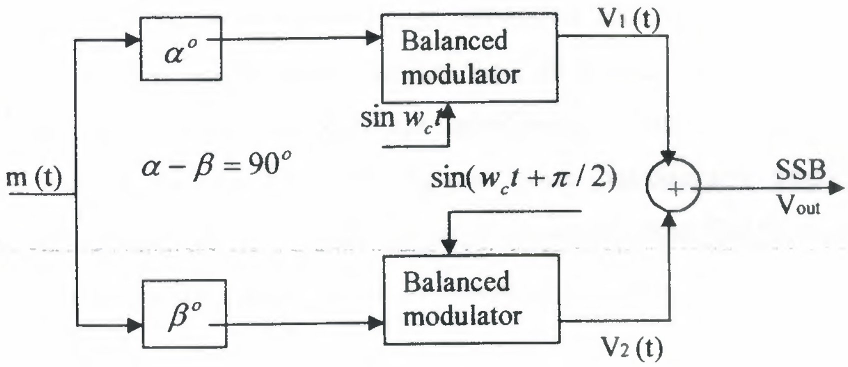


Figure 1.10 Phase-Shift Method of Generating a SSB Signal.

This condition is realized when the modulating inputs and the carriers, each differ by 90g. The output from the quadrate-phase modulator, as it is known, is

$$V_{out}(t) = v,(t) + v_2(t) \tag{1.8}$$

where $v_s(t) = E_m \cos w_m t \sin w\, ct$

and $v_2(t) = E_m \cos(w\, m\, t + \pi/2) \, \text{Sin}\,(w_c t + \pi/2)$

The modulating input m (t) is assumed to be a single sine wave $E_m \cos w_m t$

1.2 Detection (Demodulation) of AM Signals not Accompanied by Noise

The demodulation of a full AM signal, i.e. DSB with carrier, differs considerably from that of a DSBSC or an SSB signal. It can be seen from Figure 2.5 that the message m (t) is contained iii the envelope of the AM signal and therefore the detector can be a rectifier that responds to the peak amplitude of the input. This is referred to as non-coherent detection, since the demodulator does not require any information about the incoming signal. The envelope of a DSBSC or SSB signal is not a replica of m (t) and, consequently, a simple rectifier wilt not serve as a suitable detector. It is shown in section 1-2.3 that for suppressed carrier modulation there is insufficient information for a non-coherent detector to recover the base band information without distortion. The carrier must be inserted, at the receiver, before demodulation can be effected-a condition that defines a coherent detector.

A distinction can be made between the reception of a signal transmitted via a line (or cable) and one by radio. For the latter the received signal power is extremely small ranging from 10^{-10} to 10^{-6} W for short-range broadcast reception to 10^{10} to 10^{-14} W received from a satellite transmitter. Furthermore, it is accompanied by a host of unwanted signals and general background interference. Before demodulation, the wanted signal must be extracted by filtering and, since considerable amplification is also required before the final output can operate a transducer, the filtering and gain are achieved in the same process, namely that of tuned amplification. The problems of line communication are of

A Third Method of Generation and Detection of Single Sideband Signals, D. K. Weaver, Proc. I.R.E., Vol. 44, No. 12, pp. 1703-5. December 1956.

A completely different nature. The received signal power is considerably higher, due, of course, to successive amplification by repeaters spaced along the line. Whilst it is not intended to (Leal with repeater systems, the reader should not gain the impression that line communication is relatively straightforward-perhaps it should be mentioned that an ocean cable system has repeaters providing an overall gain of many thousands of decibels which must be stabilized to within 1 dB.

Before dealing with the different methods of demodulation let us consider the prerequisites of amplification and filtering. A number of high- Q tuned amplifiers in tandem could provide the gain and selectivity if the receiver were for use at a fixed frequency, but for a variable tuning facility the various stages cannot be matched (or ganged) sufficiently closely for the required amplification to be maintained over the tuning range of the receiver. We see, therefore, that high selectivity and high sensitivity are incompatible with this form of receiver. An alternative is to use an amplifier having a high gain over a fixed range of frequencies, i.e. a band-pass amplifier, preceded by a frequency changer that transposes the incoming signal in a linear manner so that it is within the pass-band of the amplifier. This is known as a super heterodyne receiver. We shall not consider a detailed study of its design, since circuit techniques are largely determined by the particular frequencies of the incoming signals. The fundamental principles, however, are common to all receivers whether they be for VLF or SHF reception.

That amplitude modulation can be regarded as a linear process due to the one-to-one correspondence of the input and output components. The frequency translation which must precede the band-pass amplifier-or intermediate-frequency (IF) amplifier as it is known, because the signal is neither at the original input frequency nor at the final base band frequency-can therefore be achieved by a modulator that multiplies the incoming signal with a locally generated sine wave $E_o \cos w_o t$. For the purpose of illustration let us assume that the input frequency is f_i – we shall ignore its bandwidth on the assumption that it is extremely small compared with ft. The output of the frequency changer will consist of upper and lower sidebands, $f_i \pm f_o$, and higher-order components involving $2f_o$, $3f_o$, etc. The intermediate-frequency amplifier is designed to allow one of these side bands to experience full gain and to attenuate effectively the other sideband and higher-order components. Such a receiver is capable of amplifying and selecting incoming signals over a range of frequencies by

Anglo-Canadian Transatlantic Telephone Cable (CANTAT), 1. F. Brampton et al., Proc. LE E. Vol. 110, No. 7, pp. 1115-23, July 1963.

Having a variable-frequency local oscillator fo that can be adjusted to

Satisfy $$f_i - f_o = ft_F \qquad (1.9)$$

When f_{tF}., the receiver frequency of the IF amplifier pass band, is lower than the incoming signal frequency; or

$$F_s \pm f_o = ft_F \qquad (1.10)$$

If the receiver is designed with an IF that is higher than the input frequency. Referring to figure (1.1) and assuming that the receiver is designed for equation (1.9) to apply, the outputs from the frequency changer corresponding to inputs at fm and f2 are ft and f_1+,f_2 respectively, where ft>fo>fs. If the two input frequencies are symmetrically positioned with respect to f_0, i.e. $f_1 = f_o - f_2$, there will be unavoidable mutual interference. The effect of the unwanted input (or second channel) can be reduced by having a high

IF-large separation of f_0 and f_2-and preceding the frequency changer with a stage of tuned amplification cent red on f_0.

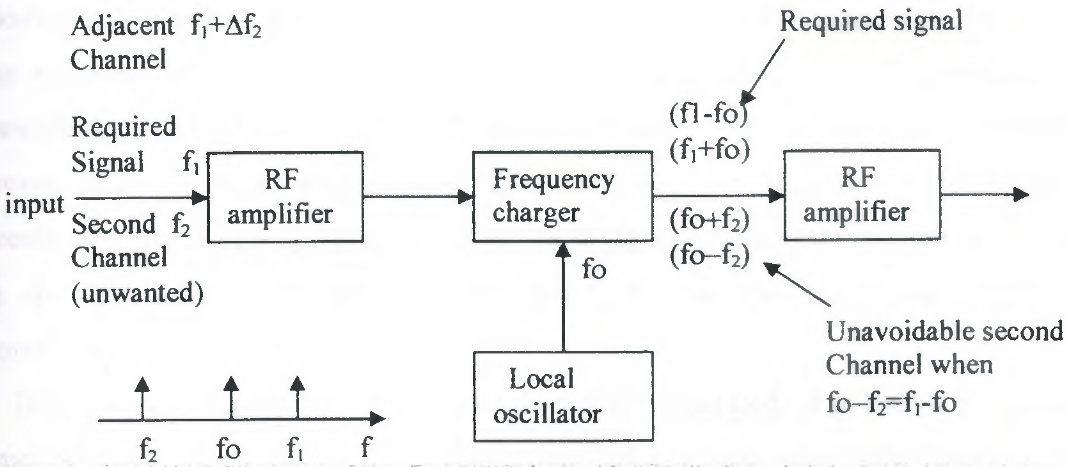


Figure 1.11 Front End of a Super Heterodyne Receiver.

This last statement requires further qualification. For the reception of signals in the LF, MF and lower HF ranges an intermediate frequency of 500 kHz is suitable. This would result in fi-fs = 1 MHz and therefore a single-stage tuned amplifier could be designed to provide considerably more gain at fi than at f_2 . On the other hand, for VHF and UHF reception a 1 MHz separation could not be considered large, since both wanted and second channel inputs would experience the same order of amplification.

Next, let us consider the requirements for removing an adjacent channel $f_1+\Delta f$. The inputs to the IF amplifier are fm- fo due to the wanted signal, and $f_m+ \Delta f- f_o$ corresponding to the adjacent channel. The lower the IF the higher the selectivity of the amplifier and the greater the rejection of the adjacent channel, but unfortunately this is contrary to the requirements for second-channel suppression. A solution is to use two stages of IF amplification, the first at a relatively high frequency for the purpose of second-channel rejection and the second at a low frequency for high selectivity. Typical values for HF reception are 1.6 MHz and 100 kHz for the first and second IF s respectively, whilst for VHF 10.7 MHz and 100 kHz are common and some UHF receivers use three stages at 50 MHz, 5-10 MHz and 100-500 kHz. It is to be stressed that these are only typical values and that many different ones are to be found in common use.

If the need to avoid seeond-chaminel and adjacent-channel interference were the only problems confronting the design engineer, his task would be a relatively simple one, but, as one may imagine, this is not the case. The frequency changer, and also non-linearity's in the tuned amplifier preceding it, gives rise to spurious outputs due to cross modulation between signals having amplitudes that are orders of magnitude greater than the required input. These are similar to the intermediation products generated in a modulator, hut in the case of the frequency changer the interference is considerably worse, since the amplitude range (or dynamitic range) of the input signals is much greater than the tympanic range of a base band signal and therefore higher-order terms in the polynomial expansion of equation (1.1), i.e. dv, etc., make significant contributions.

It has been shown that spurious modulation is not generated, or at least it is extremely small, in a modulator that employs diodes operated in a piecewise-linear manner. There is, however, a drawback to using this form of modulator as a frequency changer. The conversion gain, i.e. the ratio of input signal power to output power at the required frequency, is very much less than unity, and it is the Gaul must be as high as possible in the first and second stages of a receiver or amplifier. Consequently, the bipolar transistor operating as a small-signal common emitter amplifier is used in preference to diodes. The base current conforms to the polynomial expansion of equation (1.1) as for the diode, but the Output, which is proportional to collector current, experiences the current gain of the device. The disadvantage is the generation of cross-modulation components. For this reason the bipolar transistor as a frequency the univocal transistor, such as the Eel and MOST8.9 whose input voltage/output current characteristics conform closely to the ideal square-law response has superseded changer.

Non-Linear Distortion and Mixing Processes in Field-Effect Transistors, J. S. Vogel, Proc. I.E.E.E., Vol. 55, No. 12, pp. 2109-16, December 1967.

Application of dual-gate MOS Field-Effect Transistors in Practical Radio Receivers, Ii. M. Klein Nan, Tran's. I.E.E.E., Vol. BTR-13, No. 2, pp. 72-81, July 1967. There are, of course, spurious contributions due to higher-order terms, but these are several orders of magnitude less than those of a bipolar device.

Having outlined the basic principles of signal amplification and filtering, we are now in a position to consider the demodulation process.

1.2.1 Large-Signal Demodulation of a Full AM Signal ('Linear' Envelope Detection)

It has been suggested that the base band information can be recovered by rectifying the full AM signal to produce an output that is proportional to its envelope. If the peak-to-peak amplitude of the signal from the IF amplifier is of the order of volts, rather than mill volts, a diode used as the rectifier can be considered as operating in a piecewise-linear mode, see Figure. 1.12(a) - refer also to the large-signal modulator of section 1.1.2. It can be seen that the output consists of the envelope of the AM signal supported by half-sinusoids at the carrier frequency. The input to the detector is

$[Ec + m(t)] \cos w_{ct}$

And the output can be expressed as

$$[Ec \pm m\ (t)]^* \ 1/\pi\ [1 + \pi/2 \sin w_c t - 2/1^*3 \cos 2w_c t - 2/3^*5 \cos 4w_c t - \ldots\ldots\ldots \qquad (1.11)$$

Where the Fourier series expansion represents the half-sinusoids. We see that the output contains the base band information m (t) together with higher-frequency terms involving f_c, $2f_c$, $4f_c$, etc. If the carrier frequency is very much higher than the highest frequency in the base band, a simple filter consisting of a capacitor in shunt with the load resistor can remove the unwanted components.

1.2.2 Small-Signal Demodulation of a Full AM Signal (Square-Law Detection)

A second method of recovering the base band information makes use of the non-linear characteristic of the diode described by Equation (1-1), the input must have a peak-to-peak amplitude range of hundreds of mill volts rather than volts. The diode current is given by

$$\begin{aligned} i(t) &= a[E_c + m(t)]\cos w_c t \\ &\quad \pm b\,\{E_c + m(t)]\cos w_c t\}^2 \\ &\quad + c\{E_c + m(t) \cos w_c t\}^3 \qquad (1.12) \\ &= a[E_c + m(t)]\cos w_c t \\ &\quad + b/2(1 + \cos 2w_c t)[E_c 2 + 2tn(t)\, Ec + fli^t(t)1 \\ &+ \text{ terms due to } cv^3(t) \end{aligned}$$

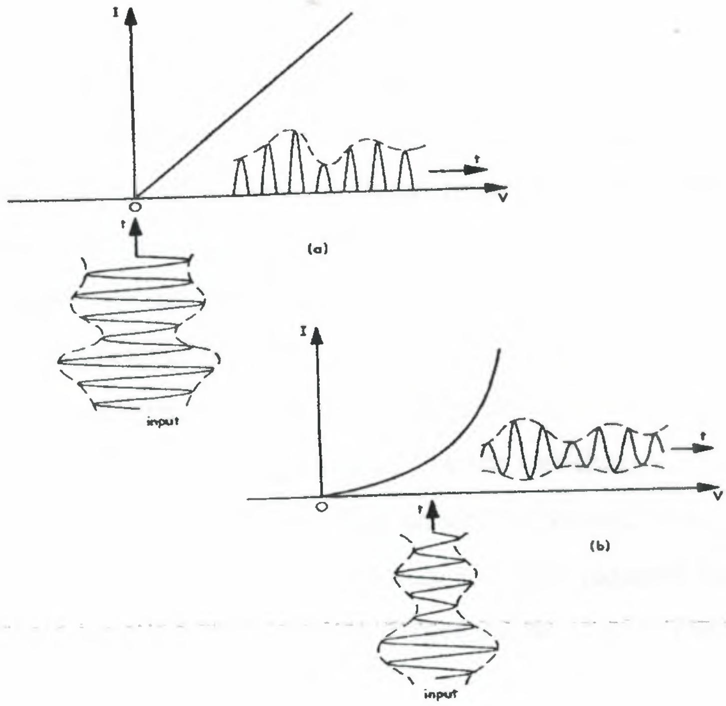


Figure 1.12 Non-Coherent Demodulators. (a) Large-Signal Piecewise-Linear Diode Characteristic, (b) Square-Law Diode Characteristic, and(c) Circuit Arrangement.

The reader should verify that the cv^3 term does not involve terms at base band frequencies and therefore need not be considered any further. The required output is due to the bv2 term-hence the description square law detector-and it can be seen from the above expression that the base band signal is recovered as b E_c m (t). Unfortunately this is not the only contribution within the base band spectrum, since the term bm^2 (t)/2 will give rise to second harmonic distortion and intermediation. To illustrate this point consider m (t) to be made imp of two sine waves $E_t \cos w_t t$ and $E_z \cos w_{2t}$, then

$$bm^2 (t)/2 = b (E_1^2 \cos^2 w_1t+2E_1E_2 \cos w_1 t \cos w_2 t + E_2 \cos^2 w_1t) / 2 \qquad (1.13)$$

And we see that there will be outputs at the second harmonic frequencies $2f_1$ and $2f_2$ and intermediation components at (f_1+f_2) and (f_1-f_2). It is possible for all of these components to fall within the required base band.

1.2.3 Coherent Detection of DSBSC Signals

A locally generated sine wave at the carrier frequency must be available before demodulation of a DSBSC signal can be accomplished. Not only

The output is zero regardless of the magnitude of, m (t) or E_{osc}. We note therefore that satisfactory demodulation of a DSBSC signal relies upon frequency coherence and phase coherence. The full significance of this statement will be appreciated after reading, which deals with the detection of signals in the presence of noise.

In practice the multiplication process can be achieved by any of the methods used for modulation. As an example, consider the chopper demodulator (or modulator) of the form shown in Figure. 1-2 (b), Let $E_{osc} \cos w_c t$ be the switching signal. To simplify the Fourier series representation of the switching action the phase term ϕ_{osc} is not included; the validity of the analysis is in no way affected, since the required output, as given by Equation (1-11), depends upon the phase difference $\phi c - \phi_{osc}$. The output from the chopper demodulator is given by

$$v(t) \alpha \; m(t) \cos (w_c t + \phi) [1/2 + 2/\pi (\cos w_c t + \cos 3w_c \; t + \ldots\ldots\ldots\ldots\ldots\ldots$$

$$= m (t) \cos \phi_c \qquad (1.14)$$

+ terms involving f_c , $3f_c$, etc., which may be filtered out. Note that when phase synchronization exists between the incoming signal and the switching sine wave $E_{ocs} \cos w_c t$, $\phi = 0$ and the output is a maximum.

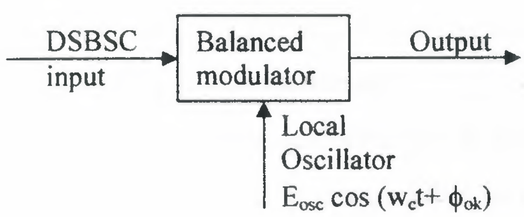


Figure 1.13 Coherent Demodulator for DSBSC signal

1.2.4 Coherent Detection of SSB Signals

It has been shown that the SSB form of modulation is the base band signal translated in frequency by an amount f_c, and therefore the demodulation process need only reverse this spectral transposition is it essential for the sine wave to have the correct frequency, but its that is to say, the demodulator can be considered as a frequency changer. It is left to the student to check this statement by considering an SSB signal due to a base band signal made up of a number of discrete sinusoids. In common with the requirement for demodulating DSBSC signals any of the multipliers described previously would make a suitable detector. However, it is to be stressed that the locally generated sine wave at the receiver must be at the correct frequency, a condition which in unimportant, since the ear is insensitive to this type of distortion. LC tuned oscillators do not have the necessary frequency stability and therefore a receiver using this type of generator could only operate

1.3 Frequency Modulation

A distinction can be made between the frequency modulation of a carrier derived from

(i) a variable frequency oscillator, and

(ii) a crystal controlled source.

For signals of the first classification, and at carrier frequencies of less than 500 MHz, say, LC tuned circuits are invariably used. The modulating signal is made to control the inductance or effective capacitance of the resonant circuit and thereby determine the instantaneous frequency. The generation of FM at higher frequencies can be carried out directly at the carrier frequency by using thermionic devices such as the Klystron Another method is to generate the signal at a lower frequency, i.e. less than 500 MHz, and then adopt a process of frequency translation.

Frequency modulation of a carrier derived from a crystal oscillator can only be achieved by an indirect means, since it is not possible to control the frequency of the resonant circuit by direct application of the modulating signal. Nowadays, nearly all communication systems use crystal-controlled sources for generating the carrier. The formation of a FM signal from a variable frequency oscillator finds application in the field of instrumentation and therefore an account of the basic principle of varying the inductance or capacitance of an oscillator resonant circuit is given below.

The Design of a Reflex-Klystron Oscillator for Frequency Modulation at Centimeter Wavelengths, A. F. Pearce and B. J. Mayo, Proc. I.E.E., Vol. 99, Part IIIA, pp. 445-54, also A. H. Beck and A. B. Cutting, ibid, pp. 357-66, 1952.

1.3.1 Generation of FM by varying the Inductance or Capacitance of an Oscillator Resonant Circuit

The unpopulated resonant frequency of a shunt LC circuit is

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{1.14}$$

If the capacitance is considered to be a variable with a maximum change f±Δ C, then

$$f_{\min} = \frac{1}{2\pi\sqrt{L(C+\Delta C}} \tag{1.15}$$

$$f_{\max} = \frac{1}{2\pi\sqrt{L(C-\Delta C}} \tag{1.16}$$

$$f_0 = \frac{1}{2\pi\sqrt{LC}}\left(1+\frac{\Delta C}{c}\right)^{-\frac{1}{2}} \tag{1.17}$$

$$f_0 = \frac{1}{2\pi\sqrt{LC}}\left(1-\frac{\Delta C}{2C}\right) \tag{1.18}$$

$$= f_0\left(1-\frac{\Delta C}{2C}\right)$$

Therefore the peak frequency deviation, or in general the instantaneous frequency deviation, is directly proportional to the change in capacitance provided that this is small compared with the total capacitance in the resonant circuit. By a similar analysis, but with the, inductance being treated as a variable, the peak frequency deviation is given by

$$f_{\max} - f_0 \cong f_0 \quad \Delta l/2l \tag{1.19}$$

To generate a frequency modulated signal as defined, i.e. a signal whose instantaneous frequency deviation is directly proportional

To the instantaneous amplitude of the modulating signal, a variable reactance device exhibiting a linear reactance/applied voltage characteristic is required. Such a device does not exist, but a reasonable approximation can be obtained by using the variable capacitance properties of a aviator diode operated over a restricted portion of its characteristic. The reactance valve provides a method of simulating either a variable capacitor or inductor and has in the past been used with success. Ferrite-cored inductors have also been considered.

1.3.2 Indirect FM

When the carrier is derived from a crystal-controlled source an indirect means of frequency modulation is necessary. The well-established method known as the Armstrong modulator makes use of the similarity between full AM and narrow-band angle modulation. Consider the DSBSC output from a

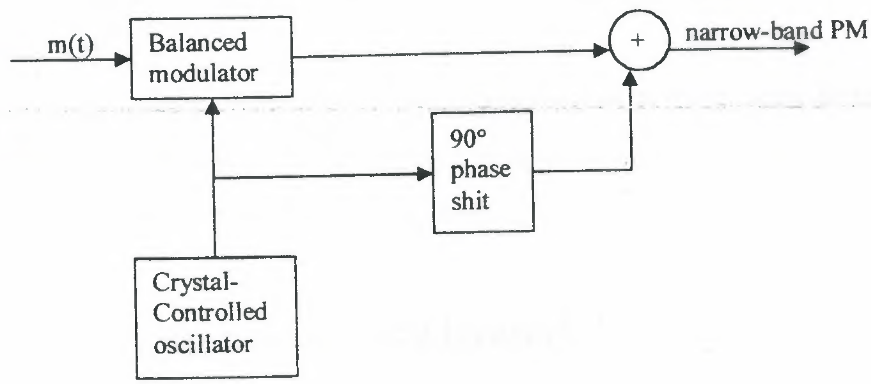


Figure 1.14 Narrow Band Phase Modulation of a Crystal Controlled Carrier

Balanced modulator added to a quadrate carrier component derived from the crystal oscillator (see Figure. 1.14). The output is effectively a narrow-band PM signal, since the phase displacement is directly proportional to the amplitude of the modulating signal m (t). Note that this is only true for phase displacements not exceeding 0.2 rod. A narrow-band FM output is obtained by preceding the balanced modulator with an integrator. This will be appreciated by referring back to Equation. (2.15) and (2.17).

Suppose we wish to form an FM signal conveying analogue information contained in a base band of 50 Hz-IS kHz and to have a maximum frequency deviation of ±75 kHz and a carrier frequency within the VHF band, e.g. a BBC transmission in the range 87.5-100 MHz. The peak phase displacement due to a component at the lowest frequency 50 Hz will be fl = 75 kHz/50 Hz, which is very much greater than the allowable 0.2 rad of the narrow-band signal. It is therefore necessary to convert the latter, which incidentally will also exhibit a small amount of AM into a wider-band version. This can be accomplished by using a non-linear circuit, fed with only the one input as distinct from a modulator or frequency changer requiring two, to generate a series of components harmonically related to the input. For example, if the input is a sine wave of frequency f, the output components from the frequency multiplier are at f, 2f, 3f, 4f, etc. A suitable arrangement would be a diode operated as a switch, i.e. under large-signal conditions, and followed by a band-pass filter to extract the required harmonic. Another and better arrangement would be to use a transistor operating under similar large-signal conditions, since a higher conversion gain can be achieved and also the band-pass filter can be included as a tuned-collector load. These large-signal frequency multipliers also act as amplitude limiters and are therefore able to remove the unwanted AM.

As an illustration considers the narrow-band signal to have been formed at a lower carrier frequency, f_c says, than the finally required output frequency. If this signal forms the, input to a frequency multiplier, the output is

$$V1(t) = E\cos n\left[\omega_c t + 2\pi\int_0^t k_t m(t)dt\right] \quad (1.20)$$

Where n is the frequency multiplication factor and the input is expressed in the form given by Equation (2.15). For ease of explanation, let us consider the modulating signal to be a single sine wave. The output from the multiplier becomes

$$V1\ (t) = E\cos n\ [w_c't + \beta'\sin w_m t] \quad (1.21)$$

and $\beta' \leq 0.2$ rad. We see that the sub-carrier frequency f_c and the peak phase displacement, or modulation index, are multiplied by n. Returning to the numerical example, the peak phase displacement of the required FM signal, corresponding to a 50 Hz component, is β,= 75 x 10'/50

1500 rad/s. Since β=nβ', n 1 500/02 = 7500. If a sub-carrier frequency f_c = 100 kHz is used, the resultant output carrier frequency would be 750 MHz, which is too high for our particular requirement. S may be overcome by interposing two frequency multipliers, producing the required multiplying factor n, i.e. m = n, with a frequency changer that translates the signal to a lower frequency without any

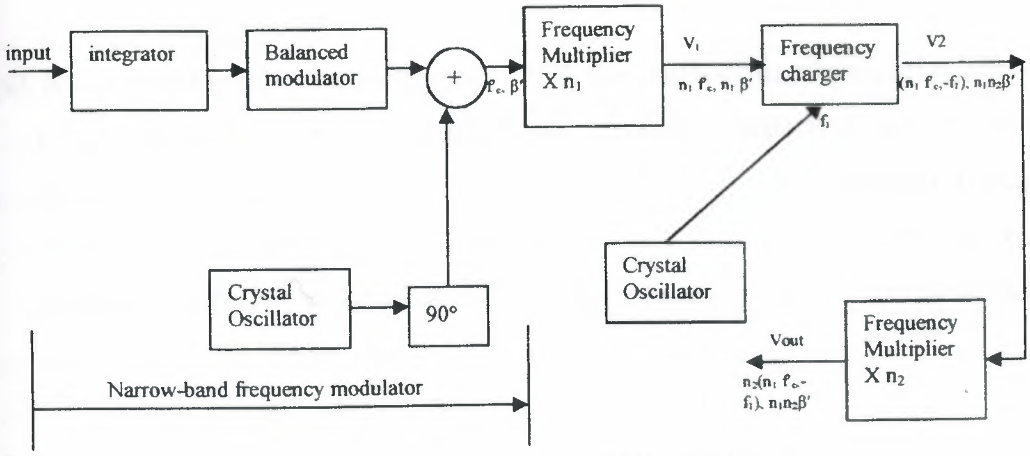


Figure1.15 Armstrong Frequency Modulator

Modification to the peak phase displacement Consider the system shown in Figure. 1.17. The required output from the first frequency multiplier is of the form

$$V_1(t) = E \cos(n_1 w_c{'}t + n_1 \beta \sin w_{mt}) \qquad (1.22)$$

and the output front the following frequency changer is

$$v_2(t) = E \cos(n_1 wc{'}\, t + n_1 \beta{'} \sin w_m t) \cos w_1 t \qquad (1.23)$$

Where w, is the angular frequency of the crystal oscillator. The output at the difference frequency $f_c = (n_1 f_c - f_1)$ can be made the same order as f_c by suitably choosing f_1. Note that in order to simplify the presentation the peak amplitude E is assumed constant throughout the process. The output from the frequency changer is an FM signal at a low sub-carrier frequency f_c", hut with a modulation index. The final output, from the second multiplier, has a frequency $n_2 f_c$" and a modulation index $\beta = n_1 n_2 \beta$

1.4 Demodulation of FM Signals

An FM detector (or discriminator as it is widely known) must produce an output voltage whose instantaneous amplitude is directly proportional to the instamitaneotms frequency of the input, i.e. it must possess the inverse characteristic of the modulator (see Equation (2.9)). There are in general two ways of achieving this:

(i) By converting the frequency modulation into amplitude modulation, followed by envelope detection, or

(ii) By using pulse-counting techniques. With this method the FM signal is converted into a sequence of constant amplitude, constant duration pulses whose rate depends upon the instantaneous frequency of the input. Integration (or averaging) of the pulse train by a simple low-pass filter yields an output having an am amplitude directly proportional to the pulse rate and hence to the instantaneous frequency of the FM signal.

Before de mutilation can be carried out the wanted signal must be amplified and selected from other signals occupying the same transmission medium. The super heterodyne receiver, described earlier for the reception of AM, satisfies these requirements, but for FM reception the IF amplifier must also act as an amplitude limiter to remove any interfering AM which may have been caused by other signals mixing with the wanted input, noise, or distortion of the signal due to band limiting effects of the earlier stages of the receiver. It will be readily

Appreciated from the following sections that most discriminators are sensitive to amplitude changes of the input, and it is to be seen that the advantages of using FM in place of AM, when noise and interference are present, are fully gained when unwanted amplitude modulation is removed prior to demodulation.

1.4.1 Discriminators employing an FM-to-AM Conversion

When an FM signal is applied to a physically realizable network, the relative amplitudes and phases of the sidebands are modified, resulting in an output that is both amplitude and frequency modulated. If this signal is to be rectified by a diode acting as an envelope detector, the amplitude variations should, ideally, be identical to the original modulating signal. If they are not, the output from the detector will be distorted. In practice distortion is unavoidable and furthermore its theoretical assessment for a given network is not straightforward. Let us, however, define the network that will produce an undistorted AM signal from the FM input and hence lead to the faithful reproduction of the message.

Since the frequency of the signal is varying from one instant to the next, care must be exercised if the steady state transfers function H (jw) is to be used in determining the output response. To explain briefly, H (jw) defines the input/output characteristic of a network in terms of its response to sinusoidal inputs of constant frequency. The transfer function can be recorded over any particular frequency range by varying the input frequency slowly, i.e. changing from one value to the next and establishing a steady-state condition at the new frequency before recording the output conditions of amplitude and phase. The response of a network to an input with rapidly changing frequency, as could be the case with FM, cannot be determined from the familiar steady-state relationship

$$V_{out} = H(jw)\, v_{in} \qquad (1.24)$$

Where $\mathbf{V}_{In}$ represents a varying frequency component, since the network may react sluggishly to a rapid change in frequency and this will not be evident from an inspection of

H (jw). It does not mean that the above expression cannot be used, but it does rule out the possibility of an exact solution by a substitution of Equation , namely

$$e(t) = Ec \cos(w_c t + \beta \sin w_m t) \qquad (1.25)$$

For the input yin. However, an exact solution by steady-state analysis'[4] can be obtained from the spectral representation of Equation. Each

Sideband, which is at a constant frequency, is modified by the transfer function defining the network response at that particular frequency. The infinite series of Equation can, be written as

$$e(t) = E_c J_n(\beta) \cos(w_{ct} + nw_m t) \qquad (1\text{-}26)$$

And the transfer function H (jw) of the network as

$$H(jw) = A(w)\, \phi(\omega)$$

With $\quad A(w_c + n\, w_m t) = An$

And $\quad \phi(w_c + n\, w_m) = \phi\, n$

Where the subscript n signifies that the amplitude and phase factors define the response at the frequency corresponding to the n sideband, i.e. at $w = w_c + n\, w_m$. The output is, therefore,

$$v_{out}(t) = \sum_{n=-\infty}^{\infty} E_c A_n J_n(\beta)\cos\left[w_c t + nw_m t + \phi_n\right] \tag{1.27}$$

As an example, consider the network to have a linear amplitude-frequency characteristic and to offer zero phase shifts. Then **A n=** a (w_c + n w_m), where a is the slope of the characteristic, see Figure. 1.16 (a). Equation (1.27) becomes

$$v_{out}(t) = \sum_{n=-\infty}^{\infty} E_c a(w_c + nw_m) J_n(\beta)\cos[w_c t + nw_m t]$$

$$= aw_c E_c \sum_{n=-\infty}^{\infty} Jn(\beta)\cos[w_c t + nw_m t]$$

$$+ aw$$

(1.28)

The first term in the above equation is simply the original FM signal with peak amplitude modified by the constant a w_c. In order to establish the significance of the second term, let us use the relationship.'[5]

$$2nJ_n(B) = \beta\left[J_{n+1}(\beta) + J_{n+1}(\beta)\right] \tag{1.29}$$

Therefore, $\sum_{n=-\infty}^{\infty} nJ_n(\beta)\cos\left[w_c t + nw_m t\right]$

$$= \sum_{n=-\infty}^{\infty} \frac{\beta}{2}\left[J_{n+1}(\beta) + J_{n-1}(\beta)\right]\cos\left[w_c t + nw_m t\right]$$

$$= \sum_{n=-\infty}^{\infty} \frac{\beta}{2} J_{n+1}(\beta)\cos\left[w_c t - w_m t + (n+1)w_m t\right]$$

$$+ \sum_{n=-\infty}^{\infty} \frac{\beta}{2} J_{n+1}(\beta)\cos\left[w_c t + w_m t + (n-1)w_m t\right] \tag{1.30}$$

Combining the (r+ 1) term of the first series with the (r- l) term of the second series reduces the expression to

$$\sum_{n=-\infty}^{\infty} \beta J_n(\beta)\cos\left[w_c t + nw_m t\right]\cos w_m t \tag{1.31}$$

It can be seen that the frequency modulation of the output is the same as the input, but in addition there is amplitude modulation, which gives rise to an envelope that is identical to the original modulating signal.

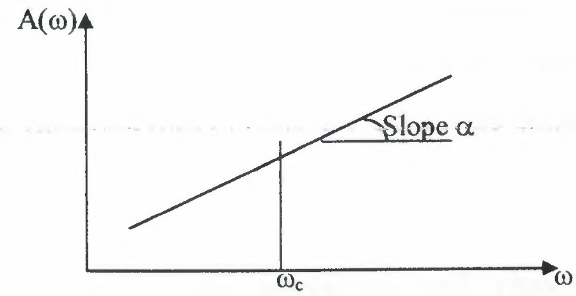


Figure 1.16 Ideal Linear Amplitude-Frequency Characteristic of a Network for FM-to-AM Conversion.

Since β $w_m=2\pi\Delta f$ and $\Delta f = k_t E_m$. Therefore, this simple but nevertheless unrealistic network satisfies the conditions postulated for zero distortion in the FM-to-AM conversion. Assuming a linear phase-frequency characteristic in place of zero phase shifts in no way violates these conditions. If we consider this alone, by making the Amplitude-frequency characteristic a constant and equal to unity, say, the output is

$$v_{out}(t) = \sum_{n=-\infty}^{\infty} EJ_n(\beta)\cos\left[w_c t + nw_m t + \phi(w_c + nw_m)\right]\cos w_m t \tag{1.32}$$

Where ϕ is the slope of the characteristic (see Figure. 1-16). Rearranging gives

$$\begin{aligned} v_{ut}(t) &= \sum_{n=-\infty}^{\infty} EJ_n(\beta)\cos\left[w_c(t+\phi) + nw_m(t+\phi)\right] \\ &= E_c \cos\left[w_c(t+\phi) + \beta \sin\ w_m(t+\phi)\right] \end{aligned} \tag{1.33}$$

And therefore the linear phase-frequency characteristic introduces no distortion but only a time delay.

When physically realizable networks having non-linear transfer functions are considered, the steady-state method of analysis can become extremely complicated, and often it is not possible to establish an exact solution. There are a number of alternative approaches; one is to evaluate an approximate solution by numerical analysis, another is the quasi-stationary method suggested by van der Pol, or thirdly a dynamic analysis using the Convolution Integral. It is felt that the last method should not be covered in an introductory textbook, as it is considered to be genuine post-graduate work; however, a brief description of the quasi-stationary approach is not out of place.

Van der Pol that when the frequency deviation is small compared with the carrier frequency, the sluggishness of the network response to an input of changing frequency can be ignored and the output given by the approximation have showed it

$$V_{out}(t) = E_c A(w_1) \cos[w_c t + \beta \sin w_m t + \phi(wt)] \tag{1.34}$$

Where w_1 defines the amplitude- and phase-frequency characteristics at the instantaneous frequency, i.e. $w_t = 2\pi\ f_c + 2\pi'\ k_t\ E_m \sin w_m t$. In our present application we are only concerned with the amplitude variations of the output, since the envelope detection following the network is insensitive to frequency and phase changes of the carrier. The amplitude-frequency characteristic $A(w_t)$ can be written as

$$A(w_1) = A(w_c + [w_1 - w_c]) \tag{1.35}$$

And because we have imposed the condition $w_1 . w_c < w_c$, it is in order to express $A(w_i)$ as a Taylor series, i.e.

$$A(w_1) = A(w_c) + (w_i - w_c)\frac{dA(w_c)}{dw_1} + \frac{(w_i - w_c)^2}{2!}\frac{d^2A(w_c)}{dw_1^{\ 2}} + \ldots\ldots\ldots\ldots \tag{1.36}$$

If we refer back to the linear amplitude-frequency characteristic of the network used in the example for the steady-state analysis, then

$$A(w_c) = aw_c \ and \ \frac{dA(w_c)}{dw_1} = a \tag{1.37}$$

The slope of the characteristic; all higher derivatives are zero. Therefore

$$v_{out}(t) = E_c[aw_c + 2\pi\ aki\ Em \sin w_m\ t1 \cos[w_e\ t + \beta \sin w_m\ t + \phi(w_i)l \tag{1.38}$$

And we see that the amplitude modulation is identical to that obtained by the steady-state analysis, see Equation (1.22). In this example, of course, Equation (1.24) is not an approximate solution but an exact one, since the idealized transfer function represents a network with no sluggishness and the condition $\Delta f \leq f_c$ need not apply.

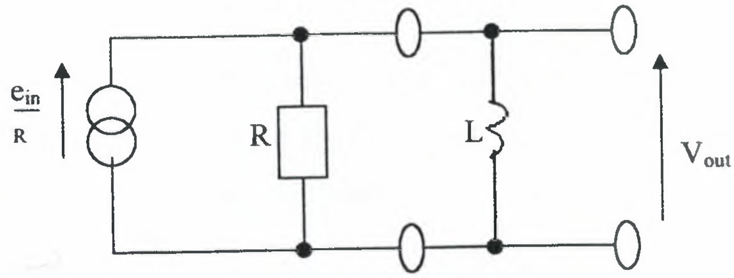


Figure 1.17 Inductor Fed from a Constant-Current Source to Produce an FM-to-AM Conversion.

Let us now look at some physically realizable networks. The simplest arrangement, and, in fact, one that produces in theory zero distortion and in practice only a small amount, is an inductor fed from a constant current source (see Figure. 1.17). The output is

$$v_{out}(t) = \frac{e_{in}(t)}{R} jwL \tag{1.39}$$

Assuming that R > wL, where .R is the source impedance. The amplitude-and phase-frequency characteristics of the voltage transfer function are

$$A(w_i) = \frac{w_i L}{R} \quad and \quad \phi\ (w_i) = \pi / 2 \tag{1.40}$$

And therefore, using Equation (1.39), the output corresponding to an FM input is given by

$$v_{out}(t) \approx E_c(w_c + 2\pi k_i E_m \sin w_m t)\frac{L}{R}\cos(w_c t + \beta \sin w_m t + \pi / 2) \tag{1.41}$$

The disadvantage of this simple network, and the reason for it not being used extensively in practice, is the extremely low efficiency of the FM-to-AM conversion. This can be seen from Equation (1.41), bearing in mind that w_i L<R.

If wL is allowed to be of the same order of magnitude as R,

$$A(w_1) = \frac{w_1 L}{\sqrt{R^2 + w_1^2 L^2}}$$

And

$$v_{out}(t) = \frac{E_c[w_c + 2\pi k_f E_m \sin w_m t]L}{\{R^2 + (2\pi k_f E_m \sin w_m t)^2 L^2\}^{\frac{1}{2}}} \cos[w_c t + \beta \sin w_m t + \phi(w_1)] \quad (1.42)$$

We see that the conversion efficiency is now much higher than that for the previous case, but unfortunately the frequency-dependent term in the denominator signifies distortion of the carrier envelope. A similar

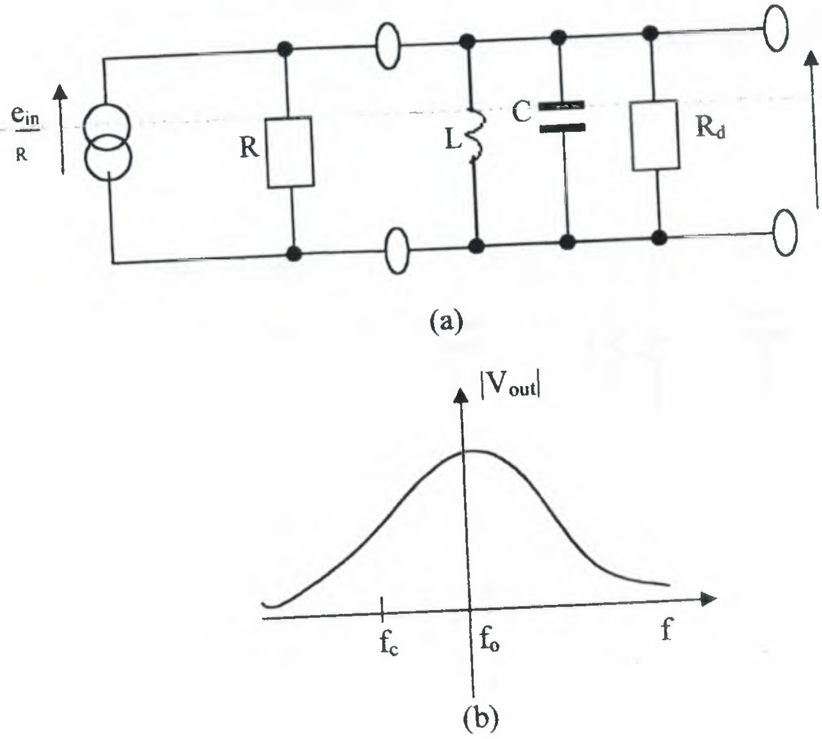


Figure1.18 (a) Off-Resonance FM-to-AM Converter. (b) Amplitude-Frequency Characteristic of the Output.

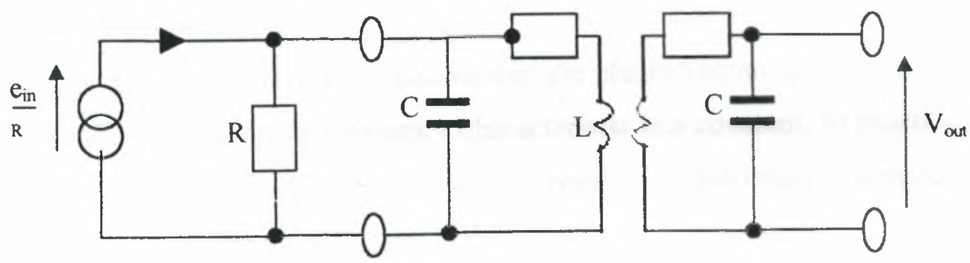
Phenomenon exists for the more widely used off-resonance circuit in which the resonant frequency is different from the carrier frequency f_c (see Figure. 1.18). The output

$$v_{out}(t) = e_{in}(t)\frac{Z(jw)}{R + Z(jw)} \quad (1.43)$$

where $$\frac{|g(T_d)|^2}{N} = \frac{\left| \int_{-\infty}^{\infty} S(jw) R(jw) \exp(jwT_d) df \right|^2}{\frac{\eta}{2} \int_{-\infty}^{\infty} |R(jw)|^2 df}$$

Using the result of the quasi-stationary approach, i.e. Equation (1.40) and the transfer function of Equation (1.43), it can be shown that the FM-to-AM conversion produces a carrier whose envelope variations are a distorted form of the modulating signal. If we is far removed from w_0 and the selectivity Q is high, the distortion can be made small. This, however, is not a practical solution, since the FM-to-AM conversion efficiency also becomes small. It is not intended to carry out a quantitative analysis of distortion, as this should entail a dynamic analysis rather than the quasi-stationary method, and it has been stated previously that such an analysis is beyond the scope of this book.

The most widely used method of achieving a FM-to-AM conversion with high efficiency and relatively low distortion, compared with the



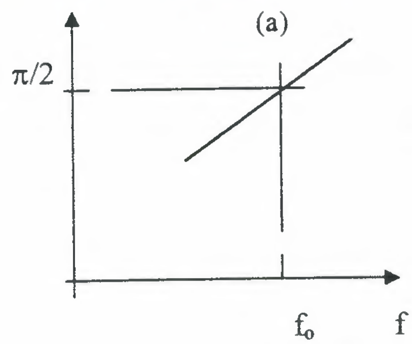


Figure1.19 (a) Mutually Coupled Parallel-Tuned Circuits. (b) Phase-Frequency Characteristic Near Resonance.

Methods described above, makes use of the phase relationship between the input and output of two mutually coupled tuned circuits (see Figure. 1.19). It can be shown by general circuit theory that the input current i (t) and the output voltage Vout (t) are in quadrature and that the phase-frequency characteristic $\phi(w)$ near resonance is approximately linear. The degree of linearity depends upon many factors including the Q, the

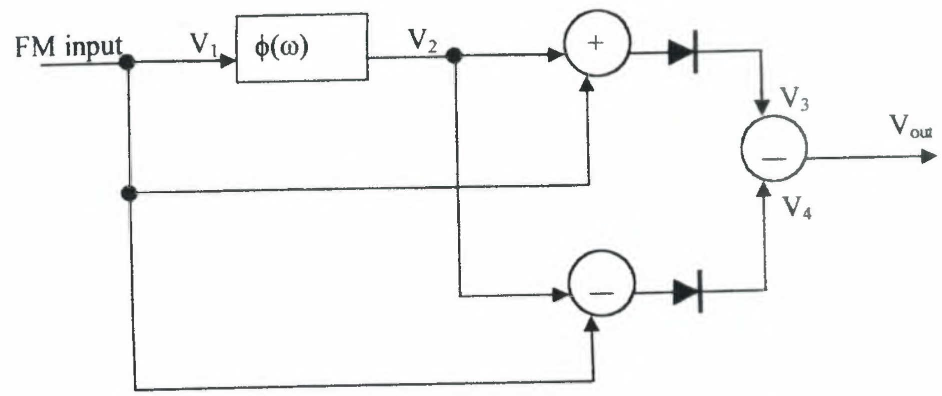


Figure1.20 Basic Arrangement of a Frequency Discriminator Employing a Phase. Dependent Network.

Coupling and, of course, the frequency range over which it is being considered. To simplify the analysis, we shall not only assume that the phase-frequency characteristic is linear but also that the amplitude-frequency characteristic is a constant. In practice any small departure from these Conditions would result in distortion. Consider the schematic arrange. mend shown in Figure. 1.20.

The resonant frequency of the phase-sensitive network is made equal to the carrier frequency of the applied signal, i.e. f/' (wc) = iT/2. The input vt (t) Corresponding to a frequency modulated signal is

$$V_1 (t) = E_c \cos (w_c t + \sin w_m t) \tag{1.44}$$

And the output from the phase-sensitive network is

$$V_2 (t) = Ec \cos [wc\, t + \beta \sin w_m t + (w)] \tag{1.45}$$

$$\text{Where } \phi(w) = \frac{\pi}{2} + k\Delta w \sin w_n t \tag{1.45}$$

And k is the slope of the phase-frequency characteristic. Note that this equation assumes the quasi-stationary analysis to be valid, i.e. the steady-state response and the dynamic response to be the same. Adding V_1 (t) and v_2 (t) gives

$$v_1(t) + v_2(t) = 2E_c \cos\left[\frac{\phi(w)}{2}\right]\cos\left[w_c t + \frac{\phi(w)}{2}\right] \tag{1.46}$$

And therefore the output from the envelope detector is

$$x \log \overline{2} = 2 \int_{1/A}^{1} x^2 p(x)\, dx \tag{1.47}$$

Subtracting Equation (1.47) from Equation (1.46) gives

$$\begin{aligned} v_{out}(t) &= E_c\sqrt{2}\left\{\sqrt{1+\cos\phi(w)} - \sqrt{1-\cos\phi(w)}\right\} \\ &= E_c\sqrt{2}\left\{\sqrt{1+\sin\theta} - \sqrt{1-\sin\theta}\right\} \end{aligned} \tag{1.48}$$

Where $\theta = \pi/2 - \phi$ (w). But ϕ (w) is in the region of 90°, therefore θ is small and the output is given by the approximation

$$\begin{aligned} v_{out}(t) &\approx E_c\sqrt{2}\vartheta \\ &= E_c\sqrt{2}\left[\pi/2 - \phi(w)\right] \end{aligned} \tag{1.49}$$

and using Equation (1.46).

$$v_{out}(t) \approx E_c\sqrt{2}k\Delta w \sin w_n t$$

And therefore the output is directly proportional to the frequency deviation. At carrier frequencies where lumped circuit techniques apply, the summations of v_2+v_1 and V_2-V_1 are usually achieved by parallel

tuned transformers as in the Foster-Scaly and the Ratio-type (discriminators. At higher frequencies, in the microwave region, the same basic principles have been applied using distributed circuits.

1.4.2 Pulse Counting Discriminator

With this form of discriminator the FM signal is converted into a sequence of constant amplitude and constant duration pulses whose repetition rate depends upon the instantaneous frequency of the input. There are a number of ways of achieving this conversion: one such method employs a bitable Schmitt trigger to produce a square wave of varying p.r.f. To obtain pulses of constant duration the output from the Schmitt circuit is differentiated, and either of the resulting positive or negative going pulses are used to trigger a moons table circuit. Family, a simple low-pass filter is used as an integrator to produce an output whose amplitude is directly proportional to the pulse rate and hence to the instantaneous frequency of the FM signal.

One of the main advantages of this arrangement is the wideband characteristic. It is possible to design a discriminator that will operate

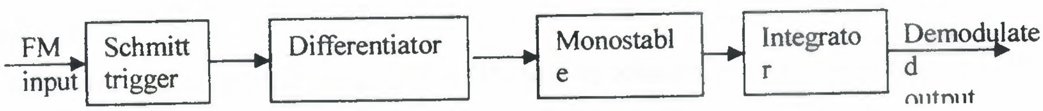


Figure1.21 Pulse-Counter Frequency Discriminator.

Satisfactorily for inputs within the frequency range 200 kHz-10 MHz. The second advantage is the extremely low level of distortion, due to a demodulation process that is essentially linear. Tile disadvantage, compared with discriminators employing an FM-to-AM conversion followed by envelope detection, is the higher output noise resulting from the additional components, some of which are active devices.

CHAPTER TWO

ANALOGUE-TO-DIGITALCONVERSION (QUANTIZATION)

2.1 Overview

First let us look at ways of representing the information digitally; there are in general two possibilities.

One is known as pulse-code modulation (PCM), in which the amplitude range of the sampled data is divided into a finite number of discrete levels; the amplitude of a given pulse is referred to the nearest level and a digital code generated. For example, in a binary system 32 amplitude levels (or quanta) are uniquely specified by a 5.digit code, since $2^5 = 32$. The analogue. To. Digital conversion introduces an irremovable error known as quantization noise which depends upon the number of levels and hence upon the number of digits in the code. the larger the number of levels the smaller the noise. It will be seen that for a given number of code digits, which is determined by the available bandwidth, the quantization distortion can be minimized by choosing non-uniform spacing of the levels to suit the statistical properties of the signal.

The second method of conversion is known as delta modulation (ΔM) or differential PCM and makes use of a 1.digit code. In the basic system, as opposed to delta-sigma modulation (Δ.EM) which is also described, the derivative of the input is transmitted rather than the instantaneous amplitude as in PCM. This is achieved by integrating the digitally encoded signal and comparing it with the analogue input in order to decide which of the two has the larger amplitude; the polarity of the next pulse, either ± V or . V, is chosen to reduce the difference in amplitude between the two waveforms. The receiver need only be an integrator followed by a low-pass filter, which removes the abrupt.

extremely complex and is not undertaken here, but instead the results of recent numerical analyses are assumed and have been found to be in reasonable agreement with previous methods of assessing the signal. To-quantizing noise ratio. Companded delta modulation is discussed and the conclusion reached is that this system offers a comparable specification to PCM for transmitting speech and has the advantage of simplicity of circuitry.

2.2 Pulse Code Modulation (PCM)

2.2.1 Quantization Noise

Misrepresentation cannot be avoided in the conversion of the signal from an analogue to a digital form. It can be seen from Figure 2.1 that the sampled data pulses are referred to the nearest quantization level in order that a binary code may be generated; in this illustration there are 15 levels giving rise to a 4.digit binary code.

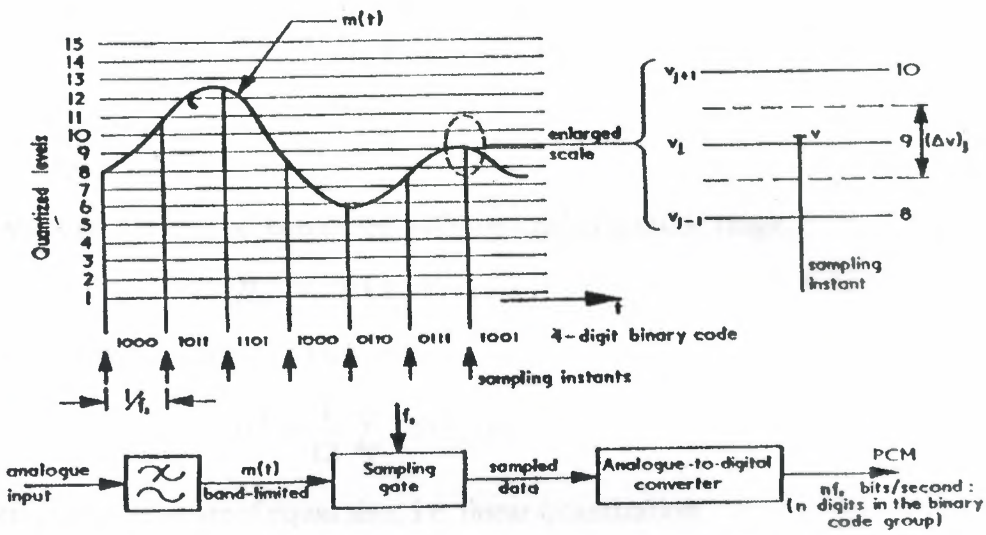


Figure 2.1 Quantization of an Analogue Signal and the Basic Arrangement for Generating PCM.

$$\sigma_j^{\,2} = \int_{v_j - \frac{(\Delta v)j}{2}}^{v_j + \frac{(\Delta v)}{2}} (v - v_j)^2 p(v) dv \tag{2.1}$$

If the step size is small compared with the total amplitude range, we may assume that time signal is uniformly distributed over the step regardless of its statistical distribution over the complete range. Therefore p (v) p (vj) and is a constant in the above integral, hence

$$\sigma_j^{\,2} = \frac{(\Delta v)_j^{\,3}\, p(v_j)}{12} \tag{2.2}$$

The total mean square quantizing noise voltage a^2 is the sum of all the mean square error voltages introduced at each level, i.e.

$$\sigma^2 = \sum_j \sigma_j{}^2 = \frac{1}{12}\sum_j (\Delta v)_j{}^3 p(v_j)$$

$$= \frac{1}{12}\sum_j (\Delta v)_j{}^2 [p(v_j) * (\Delta v)_j] \tag{2.3}$$

The probability of the signal amplitude being within the j[th] step is

$$p_j = \int_{v_j - \frac{(\Delta v)j}{2}}^{v_j + \frac{(\Delta v)}{2}} p(v)dv \tag{2.4}$$

and since (Δ v)j is small compared with the total amplitude range,

$$p_j \approx p(v_j)(\Delta v)_j \tag{2.5}$$

and therefore Equation (2.1) becomes

$$\sigma^2 \approx \frac{1}{12}\sum_j (\Delta v)_j{}^2 p_j \tag{2.6}$$

When all the steps areof equal size, i.e. linear quantization,

$$\sum_j (\Delta v)_j{}^2 p_j = (\Delta v)^2 \tag{2.7}$$

Therefore

$$\sigma lin^2 = \frac{(\Delta v)^2}{12} \tag{2. 8}$$

Example 2.1. To give an indication of the magnitude of the quantizing noise power in relation to a given signal power, consider a sine wave of peak-to-peak amplitude 2Vquantized into M levels. For linear quarintization the step

This ratio is tabulated below for various values of *M*.

Table 2.1 Give an Indication of the Magnitude of the Quantizing

Quantizing Levels M	No. of digits in binary code	r.m.s signal-to-quantizing noise ratio		Approx. % distortion
			dB	
8	3	8.6	18.7	12
16	4	18.4	25.3	5.4
32	5	38.0	31.6	2.6
64	6	77.2	37.7	1.3

When converting speech signals into digital form, a significant reduction in quantizing noise can be realized by employing non-uniform quantization to suit the amplitude distribution of the signal; there is a greater probability of small amplitudes than of large ones (see section 2.1.3), and the probability density function peaks about the average value of zero volts; therefore a tapered step system, as shown in Figure 2.6, provides an improvement over uniform quantization employing the same number of levels. There is also the added advantage that weaker signals, which do not occupy the full amplitude range of + V to - V, will be quantized by considerably more levels; the accommodation of loud and quiet talkers in the same system is dealt with later in the chapter.

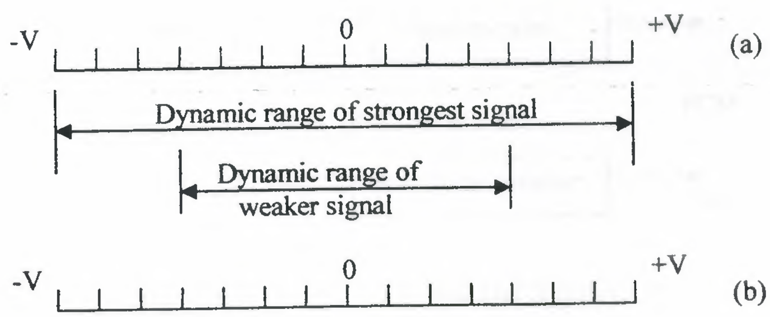


Figure 2.2 (a) Uniform, and (b) Non-Uniform Quantization of the Signal Amplitude Range.

2.2.2 Methods of Achieving Non-uniform Quantization

There are, in the main, two methods of achieving non-uniform quantization. One is to compress the amplitudes of the analogue pulses in a non-linear amplifier followed by a linear encoder, i.e. an encoder that generates coded outputs for equal step inputs: the compression of the signal is known as commanding. It must be stressed that for this application instantaneous compradors are required and they must not be confused with syllabic compradors used to protect weaker signals against Noise over long-haul HF links. To establish the theory of non-uniformly quantized speech signals we need not consider the design and circuitry of the compressor and encoder, but simply refer to the required input-output characteristic-a typical curve is shown in Figure 2.3. Before describing this characteristic, it is desirable to introduce the second method of achieving non-uniform quantization; with this method the cornpre-siohl of the signal is carried out as an integral part of the encoding Found in the references, for our purposes the input-output characteristic can be described by the graph of Figure 2.3, as for the first method of quantizing. Whichever method of quantization is used, a complementary, expand or is required at the receiver to produce an overall linear~, compressor-expand or characteristic.

A compression curve that is reasonably flexible and relatively easy to implement is the logarithmic characteristic first suggested by Cater mole. The normalized output y, i.e. maximum amplitude of unity, and normalized input x are related by the logarithmic expression

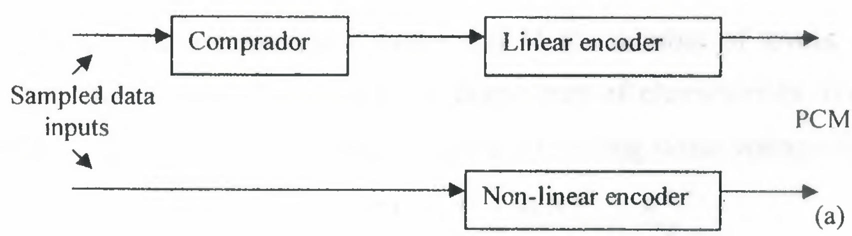


Figure 2.3 (a) Non-Uniform Encoders for PCM, and (b) Input-Output Characteristic of Either the Analogue Comprador or the Non-Linear Encoder (shown for positive inputs only and with an encoded output representing half of a fifteen-level system).

$$y = \frac{1 + \ln Ax}{1 + \ln A} \quad for \ 1 / A \leq x \leq 1 / A \qquad (2.9)$$

and liner expression

$$y = \frac{Ax}{1 + \ln A} \; for \; 0 / A \leq x \leq 1 / A \qquad (2.10)$$

Where A, the compression coefficient, is a constant chosen to suit the amplitude distribution of the signal. The linear expression is required for low levels, i.e. for Ax < 1, to satisfy the condition that y = 0 when x = 0. Furthermore, in practice it is not possible to maintain a logarithm. Characteristic at low levels which also justifies the use of the two. Expressions-note that they are continuous at x = 1/A, as illustrated in Figure 2.4. It is customary to approximate the desired compression curve with a multilinear-segmental characteristic, which can be achieved in the analogue compression process of the first method of non-uniform. Quantization or in the non-linear encoding of the second method. The input and output step sizes Δx and Δy respectively are determined by the encoding process, which is designed according to the required~ number of quantizing levels.

Let us now determine the quantizing noise that would arise from using this combined linear/non-linear quantization characteristic. The number of quantizing levels is sufficiently large for us to write, using Equation (2.9) and (2.10),

For the linear law, $$\Delta y = \frac{A \Delta x}{1 + \ln A} \qquad (2.11)$$

and for the logarithmic law $$\Delta y = \frac{1}{1 + \ln A} \frac{\Delta x}{x} \qquad (2.12)$$

Also $$\Delta y = 2/M \qquad (2.13)$$

Since2 is the normalized amplitude range and M the number of levels. The analogue signal is sufficiently small to occupy the linear part of characteristic only, then from Equation (2.6), (2.7) and (2.9) the mean Square quantizing noise voltage is

$$\sigma lin^2 = \frac{(\Delta x)^2}{12} = \frac{(1 + \ln A)^2}{3 M^2 A^2} = \frac{k^2}{A^2} \qquad (2.14)$$

Where the subscript 'lin' denotes linear quantization and

$$k = (1 \pm \ln A) M_1/3 \qquad (2.15)$$

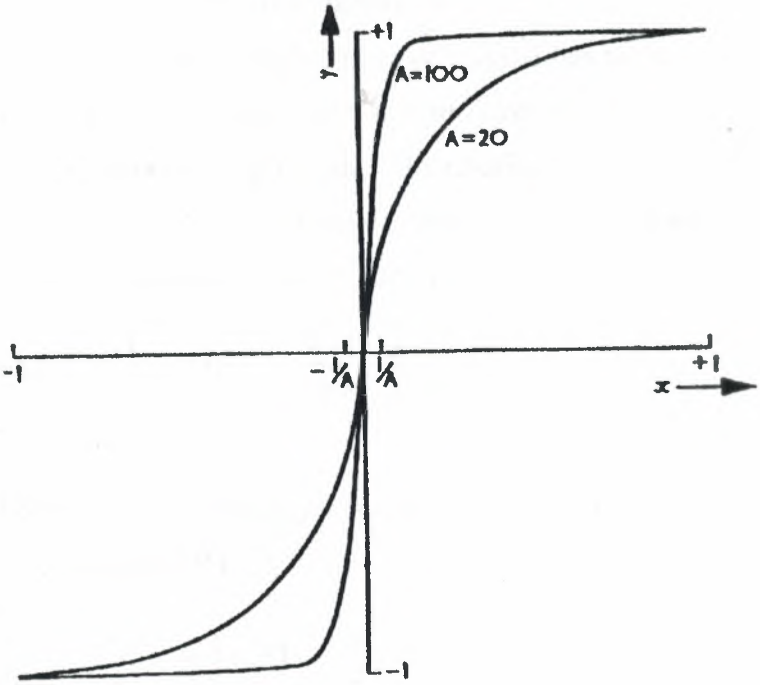


Figure 2.4 Logarithmic Encoding Characteristic.

Next, if the analogue signal occupies the logarithmic portions of the eristic only, a physically unrealizable condition but one which is convenient as an intermediate step in the analysis, then from Equation (2.6), and (2.13) the mean square quantizing noise voltage is

$$(\Delta x)^2 (1 \pm \ln A)^2 \qquad (2.16)$$

A physically realizable signal will give rise to noise from both linear and non-linear quantization; therefore we wish to find

(i) The proportion of the time for which the signal amplitude is within the range .1/A.<x<l/A, for a given r.m.s. value, and

(ii) The effective mean square amplitude of the signal when occupying the amplitude ranges .1 to .1/A and ±1/A to ±1, i.e. x^2 to account for the logarithmic quantizing noise.

We shall consider speech signals in the next section.

2.2.3 Quantization of Speech Signals

The variation of instantaneous level of a speech signal has been studied by Davenport, who suggests that it can be represented as a uniform probability distribution for very low levels plus a negative expdnential distribution for the high-level sounds. A reasonable approximation 6, 7 is possible using only the negative exponential term which is given by the probability density function

$$p(x) = \frac{1}{\sigma s\sqrt{2}} \exp\left(-\frac{\sqrt{2}|x|}{\sigma s}\right) \tag{2.16}$$

Where us is the r.m.s. speech amplitude; note that $p(x)$ is symmetrical about x = 0.

The probability distribution P (.l/A.<x<l/A), is given by

$$p(-1/A < x > 1/A) = \int_{-1/A}^{1/A} p(x)dx$$

$$= 2\int_{0}^{1/A} \frac{1}{\sigma s\sqrt{2}} \exp\left(-\frac{\sqrt{2}|x|}{\sigma s}\right)$$

$$= 1 - \exp\left(-\frac{\sqrt{2}}{\sigma A}\right) \tag{2.17}$$

Note that 1/A and hence x and σ, are normalized amplitudes-the maximum signal range being .1 to +1.

The second requirement, namely to find Xlog, is satisfied by

$$x\log^{2} = 2\int_{1/A}^{1} x^{2} p(x)dx$$

$$= 2\int_{1/A}^{1} \frac{x^{2}}{\sigma s\sqrt{2}} \exp\left(-\frac{\sqrt{2}|x|}{\sigma s}\right)dx \tag{2.18}$$

Integrating by parts and assuming 1/A < 1, which will be shown to be a reasonable approximation, gives

$$x\log^{2} = \left(\frac{1}{A} + \frac{\sqrt{2}\sigma s}{A} + \sigma s\right)\exp\left(-\frac{\sqrt{2}}{\sigma s A}\right) \tag{2.19}$$

And therefore, using Equation (2.14) and (2.15), the total quantization noise is

$$\sigma n^2 = \sigma lin^2 + \sigma \log^2 = k^2\left[\frac{1}{A^2} + p\left(\frac{1}{A} < x < \frac{1}{A}\right) + x\log^2\right] \tag{2.20}$$

The signal-to-quantizing noise power ratio is

$$\frac{S}{Nq} = \left[\frac{\sigma s}{\sigma n}\right]^2 = \frac{\sigma s^2}{k^2\left[\frac{1}{A^2} p\left(-\frac{1}{A} < x < \frac{1}{A}\right) + x\log^2\right]} \tag{2.21}$$

or, for the purpose of illustration (see Figure 2.5), the normalized signal-to-quantizing noise power ratio is

$$\frac{k^2\sigma s^2}{\sigma n^2} = \frac{\sigma s^2}{\frac{1}{A^2} p\left(-\frac{1}{A} < x < \frac{1}{A}\right) + x\log^2} \tag{2.22}$$

For example, when $a_8 = 1/A$,

$$\frac{k^2\sigma s^2}{\sigma n^2} = \frac{1/A^2}{\left(\frac{1}{A^2}(0.76) + \frac{1}{A^2}(1+\sqrt{2}+1)(0.24)\right)} = \frac{1}{1.58} \equiv -2dB \tag{2.23}$$

We see from Figure 2.5 that the signal-to-quantizing noise ratio is constant for r.m.s. signal amplitudes greater than 1/A, ignoring the 2 dB droop at the lower levels, and therefore a system capable of handling a range of inputs should be designed for the r.m.s. amplitude of the quietest talker equal to 1/A. From observations on commercial telephone circuits

Likely range of mean amplitudes in order that peak clipping of the loudest talker may not be troublesome, that the mean speech amplitudes have approximately a Gaussian distribution with a standard deviation between 4 and 6 dB; when this figure is 5.5 dB, 98 % of the talkers have mean amplitudes within ±13 dB of the median talker and 99.8 % are within ± 17 dB. Therefore the PCM system must handle a mean-level range of approximately 30 dB.

Figure 2.6, which is based upon Equation (2.18), indicates that for a given speech signal the probability of exceeding a power level of 13 dB above

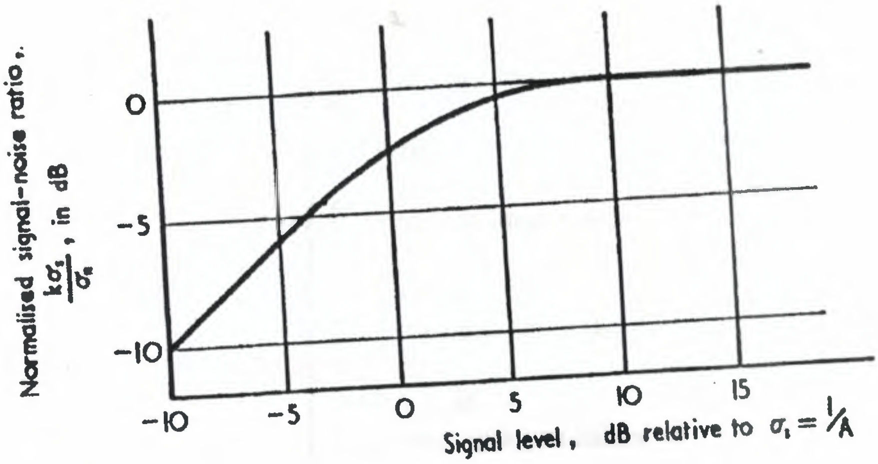


Figure 2.5 Normalized Signal-to-Quantizing Noise Ratio Against r.m.s. Speech Level.

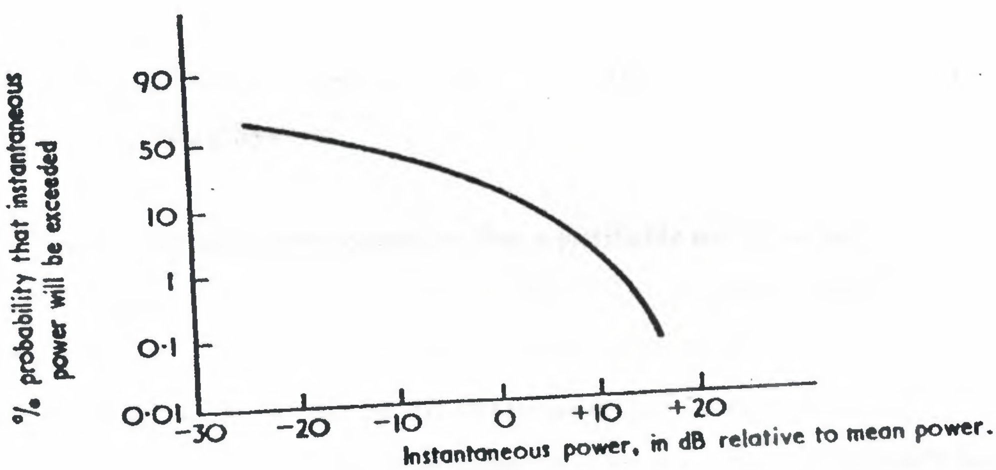


Figure 2.6 Percentage Probabilities That Instantaneous Power Level of a Speech Signal will be exceeded.

The main level is only %1 we shall use the criterion for deter This gives an indication of the required value of *A*, the compression coefficient, since the UVR is to be approximately 30 dB-hence *A* must be of the order of 150.

Having fixed the r.m.s. amplitudes in relation to the normalized range (see Figure 2.7), it is now necessary to estimate the size of the smallest step, which is the step size Δx throughout the linearly quantized range, to satisfy the required 26 dB signal-to-quantizing noise ratio for the quietest talker, whose r.m.s. amplitude is 1/A. Using Equation (2.13), P(x< Il/Al) = 0.75 for $\sigma = 1/A$, indicating that the signal is uniformly quantized for three-quarters of the time. However, to simplify the analysis we shall assume the quietest talker to be linearly quantized all

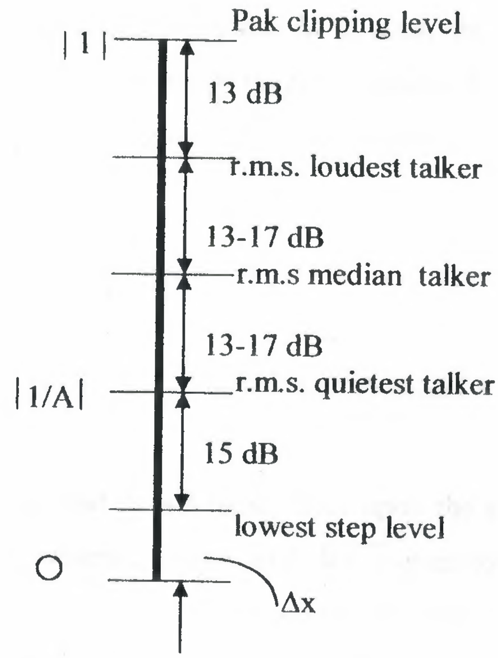


Figure 2.7 r.m.s. Amplitudes of Speech Signals in Relation to the Dynamic Range of the Non-Uniform Encoding System.

of the time, a somewhat loose approximation, but a justifiable one if we are concerned only with estimating the likely size of the smallest step. Let the r.m.s. signal amplitude be $r\Delta x$; using Equation (2.5), the r.m.s. signal-to-quantizing noise ratio is $r\Delta x/(\Delta x/\sqrt{12})$, which must be equal to 20-i.e. the 26 dB requirement-therefore $r = 20/\sqrt{12} = 5.8$, which means that the r.m.s. amplitude of the quietest talker must be 15 dB greater than the lowest step level.

The total amplitude range between the lowest step level and the peak clipping level is therefore 15+30+13 =58 dB. which means that the or for a given compression coefficient A, the minimum number of levels to satisfy the requirement $.\Delta x < 1/800$ is given by

$$M = \frac{1600(1+\ln A)}{A} \tag{2.24}$$

Note that for $A = 1$, that is linear quantization throughout the whole range, the minimum number of levels would have to be 1600, which would require .a 10.unit binary code. But for A = 150, M need only be 64 and hence a 6.unit binary code would suffice.

Finally, let us consider the choice of A and M to meet the 26 dB signal-to-quantizing noise ratio that is required for the whole range of signals. It can be seen from Equation (2.20), which is illustrated in Figure 2.5, that this ratio is approximately equal to a constant and given by

$$\begin{aligned} \frac{\sigma s}{\sigma n} &\approx \frac{1}{k} \, for \, \frac{1}{A} < \sigma s < 1 \\ &= \frac{M\sqrt{3}}{1 + \ln A} \end{aligned} \tag{2.25}$$

The interdependence of M and A and their effect upon the signal-to noise is shown in Figure 2.8; the usable volume range and its connection with the compression coefficient, see Equation (2.22), is also marked on the graph. We conclude that a 30 dB UVR and a minimum signal-to-quantizing noise ratio of 26 dB can only be achieved by a system employing 127 levels and a compression characteristic of not less than 150.

2.3 Delta Modulation (AM)

The essential difference between PCM and ΔM is that in the former information of the instantaneous amplitude of the signal is transmitted as an n-digit code, whereas in the basic AM system the rate of change of signal amplitude is communicated and in delta-sigma modulation (Δ.ΣM) the instantaneous amplitude is transmitted, but both systems employ single-digit codes.

2.3.1 Delta Modulator employing 'Staircase' Integration

As an introduction to the topic it is convenient to consider the ideal system first studied by de Jager and shown in Figure 2.9 (a). The transmitter consists of a comparator/modulator whose input is the difference

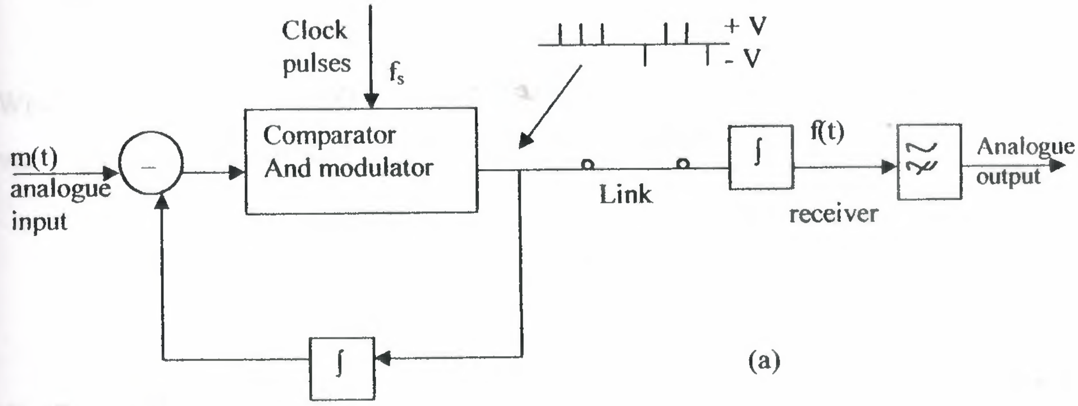


Figure 2.8 (a) Basic Delta Modulation. Analogue Input m (t) and the Corresponding Output f (t) from the Staircase Integrator.

between its integrated output and the input signal *m(t)*. The modulator is triggered by clock pulses and, when the input signal amplitude m (t)
is greater than the intenerated output, a positive output pulse is produced;
Output pulse is generated. It can be seen from Figure 2.9 (b) that the integrated output is of a stepped form which follows the smooth continuous input m (t). Therefore a similar integrator at the receiving terminal, followed by a low-pass filter to smooth out the abrupt changes of the stepped waveform, will produce an output resembling the original message m(t); the original and reconstructed signals will not be identical and we have, as for PCM, quantization noise. Referring again to Figure 2.9 (b) we see that when the input m(t) is zero the digital output consists of alternate positive and negative pulses, which will have a discrete line spectrum with the first line at the pulse repetition frequency fs. It is shown in the following analysis that fs is much greater than the highest frequency in the input signal m(t), and therefore, for this special condition of zero input, the output from the receiver low-pass filter will also be zero.
An important aspect of delta modulation is the problem of overloading which occurs whenever the input changes too rapidly for the stepped waveform to follow it; we see from Figure 2.9 *(b)* that this occurs when the transition from one step to the next does not cross the input waveform *m (t)*. Let the amplitudes of the modulator output pulses be + *V* and . *V*, and the height of each step be a and if overloading is to be avoided

$$\frac{d[m(t)]}{dt} \times ts < \sigma \qquad (2.26)$$

Where $t_s = 1/fs$, the time interval between pulses. For the purpose of illustration, a sine-wave input m(t) $E_m \sin w_m t$ would require the condition

$$\omega m E m \; \frac{1}{fs} < \sigma \qquad (2.27)$$

$$fmEm \; < \frac{\sigma fs}{2\pi}$$

If overloading is to be avoided. At this stage it is not possible to consider the implications of increasing *V* (i.e. increasing *a)* and/or fs in order that a higher frequency or larger amplitude of the input signal may be accommodated, but it would appear that increasing the pulse rate, *fs* should provide the best solution, since an increase in the pulse amplitude *V* would result in an integrated output made up of larger steps, thereby increasing the quantization noise accompanying the reconstructed signal this is pursued later. It has been found experimentally that a JM system can handle speech signals without overloading provided that the instantaneous amplitude of the signal does not exceed the peak amplitude of a 800 Hz sine wave which can also be transmitted without overloading taking place; this is because the energy of the speech signal is not uniformly distributed over the band 300-3400 Hz-the higher frequencies contain less energy than the lower ones.

2.3.2 Delta Modulator employing RC Feedback and Full-width Pulses

Let us now consider a practical system in which an RC network is used in place of a staircase integrator and full-width pulses replace the pseudo-impulses of the previous system. The exponential responses of the RC network to positive and negative step functions of amplitude 2V are shown as AB and CD in Figure 2.10. The slopes of the curves vary over the range ± V to . V, but, since the time constant T= RC is large com-pared with the pulse duration t_s, the output from the RC network, which follows the analogue input, may be considered as a piecewise linear waveform with the slope of each segment determined by the appropriate exponential curve and the amplitude of the waveform in relation to the dynamic range + V to .V (this is illustrated in the figure for both zero and sine-wave inputs). The slope of the exponential curve at amplitude v is (V-v)/T and the slope of the sine wave at the same amplitude is $\omega m\sqrt{Em^2 - v^2}$ the condition for which overloading is avoided is defined as

$$\frac{v-r}{T} > \omega m\sqrt{Em^2 - v^2}$$

or

$$D = \frac{V-v}{T} - \omega m\sqrt{Em^2 - v^2} > 0 \qquad (2.28)$$

i.e. the difference D of the slopes must always be positive and at the limit equal to zero. Differentiating D with respect to v, it is found to be a minimum when v= substituting this expression and the condition D = 0 into Equation (2.28), the overload condition for a sine-wave input is set by

V

$$E\max = \frac{V}{\sqrt{1+\omega m^2 T^2}} \qquad (2.29)$$

Where E_{max} signifies E_m for $D = 0$.

The amplitude and frequency of the sine-wave input illustrated in Figure 2.10 satisfies the overload condition of Equation (2.28) and corresponds to the limit $D = 0$; the integrated and input waveforms do not cross or touch each other at regular intervals, which was the required condition to avoid overloading in the system employing staircase integration (refer to Figure 2.9).

A plot of the overload characteristic of Equation (2.29), which is identical to the steady-state amplitude-frequency characteristic of the RC network, is shown in Figure 2.11 together with the amplitude-frequency spectrum of natural speech.[10] In practice, it is desirable to work as closely as possible to the overload condition in order that the signal-to-noise ratio may be kept as high as possible; we see from the relative positions of the two curves that this is achieved at the expense of some overloading of the middle frequency range.

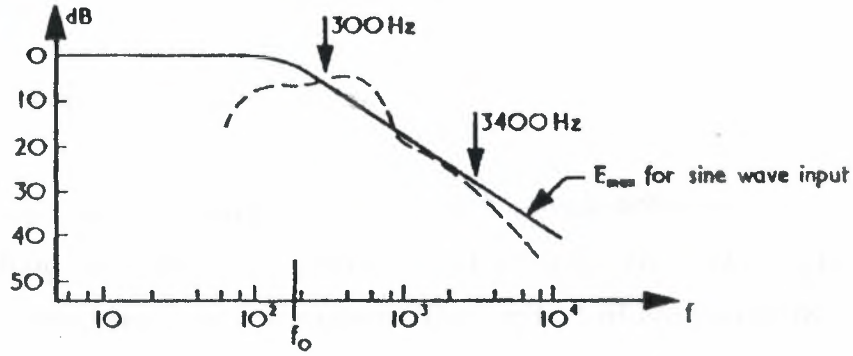


Figure 2.9 Overload Characteristic of a Basic Delta Modulator Employing Single *RC* Integration-Broken Curve Represents the Amplitude Spectrum of Natural Speech.

When the peak amplitude of the sine-wave input is smaller than the central step size Δv the modulator idling pattern of positive and negative pulses is undisturbed and hence zeros information is transmitted-this is referred to as the threshold of coding and is considered in detail in the following section.

2.3.3 Quantization Noise

Unlike PCM, the evaluation of the quantization noise in a delta-modulation system is extremely complex. Rigorous analyses are available, but the student meeting the subject for the first time is advised to consult these references after he is thoroughly conversant with the appendices of this book.

. Quantization noise is defined as the difference between the filtered output signal and the input signal, which according to the definitions illustrated in Figure 2.12 has a spectrum defined by

$$Q(jw) = M(jw) - G(jw) \tag{2.30}$$

Also

$$E(jw) = M(jw) - F(jw)\,M(jw) - S(jw)\,H(jw)$$

Or

$$E(jw)A(jw) = M(jw) - S(jw)H(jw)A(jw) \tag{2.31}$$

And assuming A (jw) is flat over the spectrum of M (jw),

$$E\ (jw)\ A\ (jw) = M\ (jw) - G\ (jw) \text{ and} \qquad (2.32)$$

$$\text{Therefore, } Q\ (jw) = E\ (jw)\ A\ (jw) - \qquad (2.33)$$

The difficulty arises in determining E (jw); even with a single sine-wave input the resulting pulse train s(t) is random and consequently E(jw) is the Fourier transform of a non-deterministic waveform. This method of analysis is not pursued further and has only been introduced to illustrate the complex nature of a rigorous treatment.

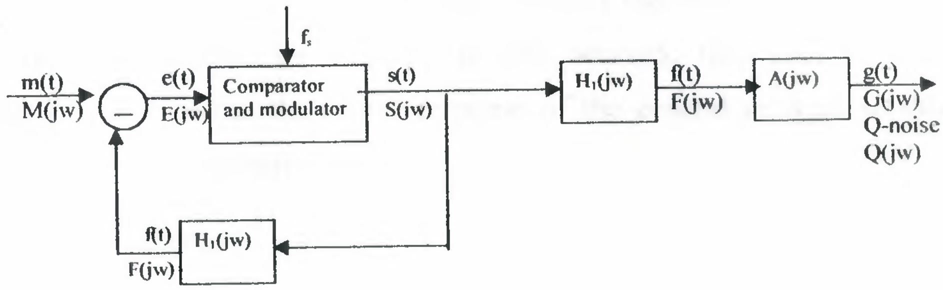


Figure 2.10 Basic Delta Modulation.

Referring again to Figure 2.10, the error waveform e(t), i.e. the quantization noise before filtering, is triangular when the input m(t) is zero. Before inspecting the error corresponding to the sine-wave input, it is necessary to re-draw the integrated and input waveforms with the latter delayed by a clock pulse interval t_3 this is evident from an inspection of the figure, also bearing in mind that the output can only respond to a change of the input after it has taken place-the re-drawn waveforms and their difference e (t) are shown in Figure 2.13. Although the error e (t) is random for the sine-wave input, there is a distinct triangular pattern somewhat similar to that which occurs for zero input. It must be realized that in these illustrations the sine wave represents two extremes; firstly, it has the maximum permissible amplitude set by the conditions to avoid overloading and, secondly, it represents the highest input frequency that would be accommodated in a practical system referring again to Figure 2.10, the clock rate is twenty times the sine-wave frequency and it will be seen later that this is about the smallest allowable figure for the ratio fs :f_m. Thus we may conclude that the triangular of the error waveform will be more pronounced for all other acceptable input.

Let us consider the error waveform e (t) resulting from zero input; the slope of each segment is V/T and the step height Δv, i.e. the peak-to-peak amplitude, is

$$\Delta v \approx \frac{V}{Tfs} = \frac{2\pi V f_0}{fs} \tag{2.34}$$

Where fo = 1/2πT, the 3dB break frequency of the RC network.
A similar expression can be obtained by a different line of reasoning. The digital output corresponding to zero analogue input is a sequence of alternate + *V* and . *V* pulses having a fundamental frequency fs/2. The periodic sequence has a line spectrum, the first being at fs/2, and since fs >fo we may consider that the fundamental sine-wave cornponent is responsible for exciting the RC network, thus producing the error waveform e (t). Therefore the output response of the general network to this digital input is given by the approximation

$$f(t) \approx \frac{4V}{\pi} \sin\left(\frac{2\pi fst}{2}\right) \left| H\left(j\frac{\omega s}{2}\right) \right|$$

Therefore the step height Δv, is

$$f(t)\Big|_{\max} \times \frac{1}{fs} \approx 4V \left| H\left(j\frac{\omega s}{2}\right) \right| \tag{2.35}$$

And for the *RC* network, assuming fo <fs, we have

$$\Delta v \approx 4V \frac{2f_0}{fs} = \frac{4}{\pi} \frac{2\pi V f_0}{fs} \tag{2.36}$$

Which is similar to the previous expression, Equation (2.34), except for the coefficient 4/π.
The mean square error is

$$\overline{e^2(t)} = \frac{2}{ts} \int_0^{t/2} \left[\frac{(\Delta v)t}{ts}\right]^2 dt = \frac{(\Delta v)^2}{12} \tag{2.37}$$

And we have seen previously that this power will be contained in a line spectrum at multiples of f_s. Next consider a sine-wave input; it is reasonable to assume that the mean square error voltage is proportional to $(\Delta v)^2$, since the triangular characteristic of the waveform is similar to that corresponding to zero input. It can also be seen from Figure 2.13 that the error is greater for the sine wave, and it has been established by computer analysis that the mean square error for various sine-wave inputs is $e^2(t)$ $(\Delta v)^2/6$ and that this remains constant over a range of input amplitudes from just above the threshold of coding to the overloading level. But this error power is associated with a random-sequence triangularly shaped polar function which, according to the analysis of section , has a continuous spectrum and a power density that is proportional to $G(jw)^2$. A triangular pulse of peak amplitude A and duration $2t_p$ has a Fourier transform

$$G(j\omega) = At_p\left[\frac{\sin(\omega tp/2)}{\omega tp/2}\right]^2 \tag{2.38}$$

And at $w = 0$, $I\ G\ (jw)\ 2 = G\ (0)\ 2 = A^2t_p^2$. Therefore we can say that the energy contained in the positive (one-sided) frequency spectrum is $A^2t_p/3$ and is equal to the energy that would result from the spectral density at the origin, i.e. $A^2t_p^2$, if it were held constant over a bandwidth from 0 to ± $l/^3$tp. In the present application $l/t_p\, fs$ and hence the unfiltered quantization noise energy may be considered to be uniformly distributed over a bandwidth from 0 to B_m. Therefore, if the receiver low-pass filter has a bandwidth from 0 to B_m, where $B_m < fs$, the mean Quantization noise power Nq is given by

$$Nq \approx \overline{e^2(t)}\frac{3Bm}{fs} = \frac{(\Delta v)^2}{6}\frac{3Bm}{fs} \tag{2.39}$$

And using Equation (2.34), we have

$$Nq \approx \frac{2\pi^2V^2f_0^{\ 2}Bm}{fs^3} \tag{2.40}$$

If we assume that the pulses supplied to the receiver decoder are of amplitude + V and .V, a condition that can easily be satisfied by pulse regeneration, but which may cause some digits to be in error, the mean square signal amplitude of the output corresponding to the limit of non-overloading at the input of the system is, according to Equation (2.33), $Emax^2/2$ $V^2/2(1 .wm^2T^2)$ hence the mean signal-to-quantizing noise power ratio is

$$\frac{S}{Nq} \approx \frac{fs^3}{4\pi^2(1+\omega m^2T^2)f_0^{\ 2}Bm} \tag{2.41}$$

Referring to Figure 2.11, we see that Jo, the RC integrator characteristic frequency, is approximately equal to the lowest frequency of the speech signal and therefore in general fm >f_o, and hence $wm^2T^2>1$, which allows Equation (2.41) to be written as

$$\frac{S}{Nq} \approx \frac{fs^3}{4\pi^2 fm^2 Bm} \approx 0.025 \frac{fs^3}{fm^2 Bm} \qquad (2.42)$$

The form of this expression agrees with that of de Jager, and also with more recent work, except for the coefficient 0.025, which is 0.04 and 0.067 for the respective references. Bearing this in mind, together with the approximations that have been made in the above analysis, the results should not be interpreted too rigidly.

It is tempting to compare AM and PCM for a sine-wave input that, in the former, satisfies the limiting condition of non-overloading and, in the second, has peak-to-peak amplitude occupying the full quantized amplitude range. In fact, this has been carried out on a number of occasions and usually a plot of signal-to-noise ratio against clock rate shows that AM is superior to PCM, when the latter uses a code group of five or less. Unfortunately such a comparison can be misleading if no account is taken of the range of input levels that must be handled by a viable system. The speech signals occupy a 30 dB mean level range from the quietest to loudest talkers and considerable attention has been given to the design of a PCM system capable of handling such a range. Using Equation (2.41) we see that, for the AM system described, an output signal-to-quantizing noise ratio of 34 dB is obtained for a 800 Hz sine-wave input, which we shall assume from previous comments is representative of speech, an output band-width B_m = 3 kHz, and a clock rate of 55 kilobits/second-the latter is the PCM digit rate when using a 7.digit code, i.e. it is equal to 7 times .the sampling rate of the analogue input, which is 8000 per second . It has been stated previously that the signal-to-quantizing noise ratio for commercial telephony should not be less than 26 dB, and. therefore the useful dynamic range of the single *RC* integrator AM system is 34.26 = 8 dB, which is very much smaller than the useful volume range attainable by PCM operating at the same overall digit rate (see Equation (2.25) and Figure (2.8)).

Although this form of delta modulation is unsuitable for commercial

Missible-signal-to-quantizing noise ratio may be less than 26 dB. With this in mind let us pursue the analysis further.

The maximum amplitude of the sine-wave input consistent with the condition of non-overloading has been established; see Equation (2.28). The minimum peak amplitude below which the input fails to excite the modulator occurs when the peak-to-peak amplitude of the signal is smaller than the central step size LI v; this result in a modulator output of alternate positive and negative pulses corresponding to zero input. Let us define this lower amplitude limit (or threshold of coding) as

$E_{min} = (\Delta v)/2$ where E_{min} is the peak amplitude of the sine-wave input. Using Equation (2.28) and (2.27), we have

$$E\min = \frac{\Delta v}{2} = \frac{\pi V f_0}{fs}$$

and

Therefore $$E\min = \frac{\pi f_0}{fs} E\max \sqrt{1+\omega m^2 T^2}$$ and substituting $T = 1/2\pi fo$

Then $$\frac{E\max}{E\min} = \frac{fs}{\pi\sqrt{f_0^2 + fm^2}} \qquad (2.43)$$

The ratio of E_{max}/E_{min} which is usually expressed in dB, is the dynamic input range of the system; for the previous numerical example in which fs = 56 kilobits/second, fm = 800 Hz and fo 150 Hz (see Figure 2.9), 20 $\log_{10}$ *(Emax/Emin)* = 28 dB. The quantizing noise is independent of the input signal amplitude and therefore the signal-to-quantizing noise ratio varies from 34 to 7 dB over the dynamic range; whether or not a signal-to-noise ratio of 7 dB is acceptable would depend upon the particular application.

2.3.4 Delta-sigma Modulation (Δ.EM)

The system so far described exhibits an overload characteristic that is particularly suited to speech when using a high-quality microphone. It is not so attractive when using a carbon microphone or for video signals such as television. A system that provides an overloading characteristic with relatively flat energy spectra-is known as *delta-sigma* modulation. It is similar to the previous system, except that an integrator *H (jw)* is included in the input circuit and a complementary inverse process $H^{-1}(jw)$ in the output circuit; however, it is shown that the latter may be combined with the receiver integrator to allow both to be dispensed with.

The inclusion of an integrator before what has so far been considered a delta modulator means that information of the amplitude of the input and not the derivative of the signal is transmitted. This raises the question of what is the difference between ΔΣM and PCM-simply, the latter involves an ADC process and the generation of a n-digit code, whereas Δ.ΣM is a single-digit code system and furthermore it does not require the complex encoding and decoding associated with PCM; this is discussed in more detail later in the chapter.

Let us pursue an analysis along similar lines to that of basic delta modulation. It is firstly necessary to determine the maximum amplitude of a sine-wave input consistent with the condition of non-overloading. The slope of the output from the first *RC* integrator is

$$\frac{dv}{dt} = \frac{\omega mE \cos \omega mt}{\sqrt{1+\omega m^2 T^2}} = \frac{\omega m}{\sqrt{1+\omega m^2 T^2}}\sqrt{Em^2 - v^2(1+\omega m\ T^2)} \qquad (2.44)$$

And the difference *D* between this slope and the slope of the output from the feedback *RC* integrator is

$$D = \frac{V-v}{T} - \frac{\omega m}{\sqrt{1+\omega m^2 T^2}}\sqrt{Em^2 - v^2(1+\omega m^2 T^2)} \qquad (2.45)$$

This must be greater than or at the limit equal to zero to avoid overloading. The zero limit is found by differentiating *D* with respect to *v* and equating to zero, which gives $v = Em/(1 + w_m^2 T^2)$. Substituting this and $D = 0$ into Equation (2.45) gives the non-overloading condition as

$$E_m = E_{max} = V \qquad (2.46)$$

i.e. a maximum peak amplitude of the input equal to the pulse amplitude-a result that the reader may have realized from an inspection of the block diagram of Figure 2.14 without recourse to the above procedure. Nevertheless, whichever method is adopted it can be seen that the overload condition is independent of the input sine-wave frequency fin.

The quantization noise in a Δ.ΣM has been shown that the error power density at the input to the receiver low-pass filter is approximately

Rearranging the order of the receiver processes-which i_s permissible with linear networks.(see Figure 2.10), second version to receiver decoder, the error power density at the output of the network $H^{1}(jw)$ is

$$\frac{(\Delta v)^2}{6}\frac{3}{fs}\left|H^{-1}(j\omega)\right|^2 = \frac{(\Delta v)^2}{6}\frac{3}{fs}\left(\frac{f^2+f_0^{\,2}}{f_0^{\,2}}\right) watt\ /\ Hz \tag{2.47}$$

Where the symbols have the same meaning as in the previous section. The quantizing noise power Nq at the final output is therefore

$$Nq \approx \int_0^{Bm} \frac{(\Delta v)^2}{6}\frac{3}{fs}\left(\frac{f^2+f_0^{\,2}}{f_0^{\,2}}\right)df$$

$$= \frac{(\Delta v)^2}{2fsf_0^{\,2}}\left(\frac{Bm^3}{3}+Bmf_0^{\,2}\right)$$

Using Equation (2.24) and substituting for Δv,

$$Nq \approx \frac{2\pi^2V^2}{3}\left(\frac{Bm}{fs}+Bmf_0^{\,2}\right)^3 \tag{2.48}$$

We see that the quantization noise is reduced by making Jo as small as possible, i.e. the time constant T of the RC network as large as possible. If we satisfy this requirement, $B\sim$ >fo and

$$Nq \approx \frac{2\pi^2V^2}{3}\left(\frac{Bm}{fs}\right)^3, f_0 << Bm \tag{2.49}$$

Next let us evaluate the mean signal power at the output. Referring to Figure 2.14 and the first version of the receiver decoder, the signal power at the output of the low-pass filter is approximately equal to the signal power at the comparator input, namely $v^2(t)=$ E 2/[2(1 $+w_m^{\,2}T^2$) this line of reasoning was used in the previous section dealing with the basic Δ.ΣM system. Therefore the signal power at the output of the Δ.ΣM receiver is

$$\frac{Em^2}{2(1+\omega m^2T^2)}\left|H^{-1}(j\omega m)\right|^2 = \frac{Em^2}{2(1+\omega m^2T^2)} = \frac{Em^2}{2} \tag{2.50}$$

This is simply the mean power at the input of the system. Hence, the mean output power corresponding to the limit of non-overloading at the input is $E^2_{max}/2 = V^2/2$ (see Equation (2.43)). Thus the mean signal-to-quantizing noise ratio is

$$\frac{S}{N} \approx \frac{3}{4\pi^2}\left(\frac{fs}{Bm}\right)^3, f_0 << Bm \tag{2.51}$$

As an illustration consider a sine-wave input to a system having a bandwidth from d.c. to B_m of 3 kHz-strictly speaking, this is not a system to handle speech, as this would employ a band-pass filter in the receiver in

place of a low-pass unit and consequently the lower limit of the integral leading up to Equation (2.49) would be 300 and not zero. The signal-to-quantizing noise ratio corresponding to $fs = 56$ kilobits/second amid Em 3 kHz, is 28 dB; we see that this is only marginally above the required ratio of 26 dB for commercial telephony. According to the analysis leading tip to Equation (2.49), the quantization noise is independent of the signal level, and since the 27 dB ratio corresponds to the maximum input it may be concluded that there is virtually zero dynamic range available for this application.

Let us pursue the general analysis to establish the dynamic range from the maximum input consistent with non-overloading to the minimum input corresponding to the threshold of coding. For the latter, let the input $m\ (t) = E_{mtn}\ sin\ w_m t;$ the peak output from the first RC network is , and therefore using the definition of the threshold of coding stated in the previous section dealing with basic AM, we obtain

$$\frac{E\min}{\sqrt{1+\omega m^2 T^2}} = \frac{\Delta v}{2} = \frac{\pi V f_0}{fs} \tag{2.52}$$

Combining this with Equation (2.33), we have

$$\frac{E\max}{E\min} = \frac{fs}{\pi\sqrt{f_0^2 + fm^2}} \tag{2.53}$$

This is identical to Equation (2.50) for basic AM.

The main advantage of A.EM is the ability to convey information of d.c. levels. For zero input there are an equal number of positive and negative pulses in the transmitted sequence, and hence the filtered output at the receiver is zero. For a positive d.c. input there are more positive pulses than negative ones, the average number being in direct proportion to the magnitude of the input level; similarly, there are more negative than positive pulses in the digital signal when the d.c. input is negative. This is true for both AM and A.EM, but with the latter the receiver is simply a low-pass filter which allows the average amplitude of the digital signal to be retained in the final output.

The choice of the optimum value of fo, the 3 dB cut-off frequency of the PC network, highlights another difference between the two methods of modulation. In basic ΔM, Jo must be chosen in accordance with the spectrum of the input signal as shown in Figure 2.11, whilst with Δ.EM Equation (2.46) and (2.49) indicate that Jo should be made as small as possible; in practice, the lower limit is set by the instabilities of the comparator the lower fo the smaller the step size.

2.3.5 Companded delta modulation

The main drawback to basic ΔM and Δ.EM is the limited dynamic range over which a signal-to-quantizing noise ratio in excess of 26 dB is obtainable. It has been seen that this is due to the quantization noise being independent of the signal level and set solely by the quantizing step size Δv. The following system known as Companded delta modulation employs a local decoder to modify the step size in accordance with the level of the input signal. There are basically two methods of achieving this. One, is to obtain a control signal whose magnitude is proportional to the mean level of the analogue input and to use this signal to amplitude modulate the digital sequence fed to the integrator in the feedback circuit the step size is directly proportional to the amplitude of the pulses. The second method is similar except that the control signal is derived entirely from the digital output of the encoder; a brief account of a system falling into the latter category is given below.

A block diagram of the scheme is shown in Figure 2.10. The feedback circuit consists of an integrator *Hi(jw)* which performs a similar function to those described previously, except that the input pulses are not of constant amplitude ± *V*, but have a variable level + *C* which is determined by a control signal amplitude modulating the digital output of the encoder. The control voltage is made up of two components, a d.c. bias which is included to prevent *C* falling below a level consistent with stable operation-the smaller C the smaller the step size and the higher must be the gain in the comparator-and the other, the output from a level detector which follows the integrator H_2(jw). The time constant of the level detector is made sufficiently long for the control signal to respond only to syllabic changes. The relevant equations can be obtained from the analysis of basic ΔM. Let the analogue input be a sine wave m (t) = E sin $_{wm}$t and E_{max} represent its peak amplitude corresponding to the limit of non-over-loading. Using Equation (2.32), but expressing the integrator transfer function in its general form, we have $E_{max} = V\{Hi\ (j_w m)\}$. In Companded delta modulation the overload condition is defined similarly except that the amplitude of the pulses fed to the integrator is C; hence we have

$$E_{max} = C \left| H_l\ (jwm) \right|$$

(2 .54)

peak amplitude E to the overload level at the same frequency be and Equation (2.54) may be written as expressed as

$$E = \frac{E}{E\max}, E < E\max \qquad (2.55)$$

Similarly, let the ratio of the signal level contained in the digital output and when C is a maximum,

to that corresponding to the overload level be

$$Y = \left[\frac{E}{E\max}\right]^{n} \qquad (2.56)$$

Therefore A = (1-B/V)/1H$_2$ (Iw$_m$) I

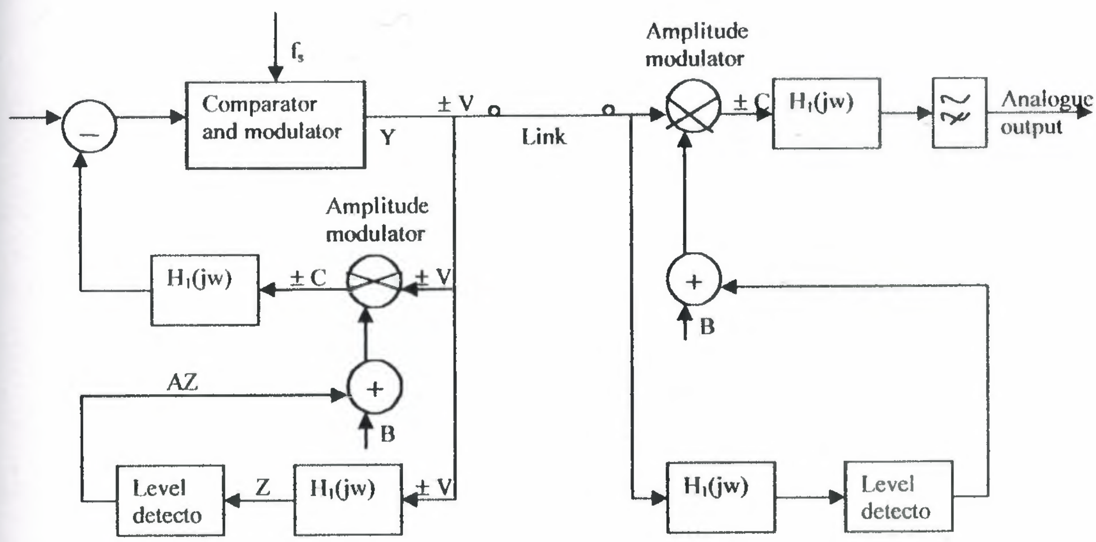


Figure 2.11 Block diagram of a compounded delta modulation system

The reason for introducing the index n will become apparent as we proceed with the analysis. briefly, this is a nonlinear encoder and the Substituting this result in Equation (2.57) and putting B/V = B, gives ratio of signal levels contained in the digital output is not linearly

Related to the ratio of the analogue levels. Note that when $E = E_{max}$, Y = 1.

The output from the integrator H_1 (jw), which must follow the

Input level E, can be written as

$$\text{i.e.} \quad C = B+(1-B)Y \tag{2.57}$$

CHAPTER THREE

DIGITAL SIGNALLING-BASEBAND ANALYSIS

3.1 Overview

Digitally encoded information can tolerate more signal distortion due to Intersymbol interference and noise pick-up than an analogue signal, and in Chapter 2 we have seen how the analogue-to-digital conversion can be effected. This chapter is concerned with establishing the optimum pulse shape and filtering of the digital signal, in order that the received pulse sequence may be interpreted with the minimum error. The overall transmission system is shown in Figure 3.1 and the following analysis will deal with the base band signal, i.e. a signal suitable for line communication; however, the basic principles apply equally to radio systems which require, in addition, modulators and demodulators operating at the radio carrier frequency.

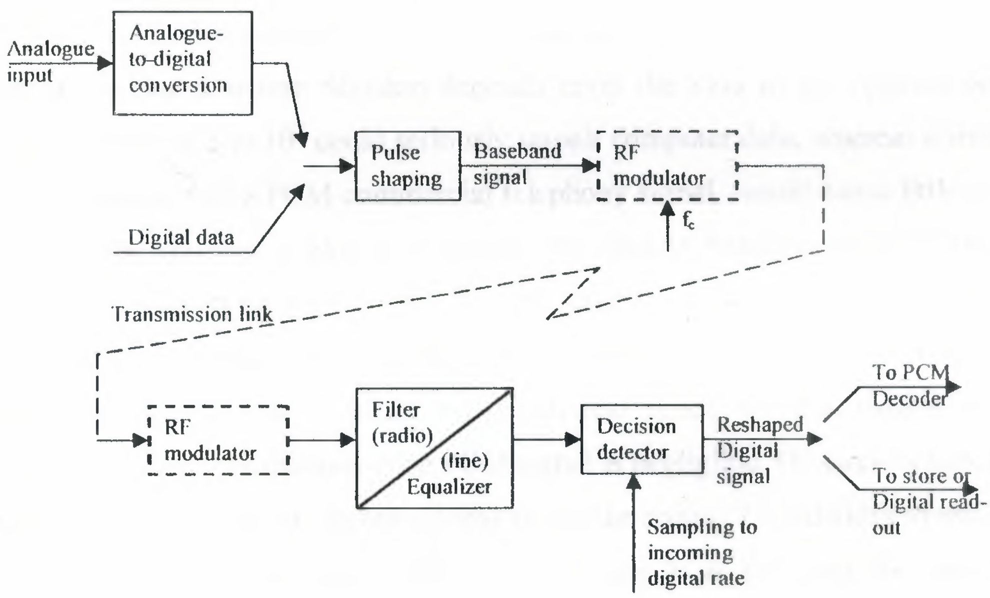


Figure 3.1 Digital Transmission System Employing Decision Detection.

The essential difference between analogue and digital transmission is best illustrated by considering the function of the receiver. The relative amplitudes of the pulses of a PAM analogue transmission directly determine the characteristics of the reconstituted continuous signal, whereas, in a digital system, it is the coding (or grouping) of the digits that determines the analogue characteristics-the amplitudes of the pulses contributing indirectly. That is to say, it is only necessary for the receiver to decide the particular amplitude of a pulse out of a finite number of possibilities; once the decision has been made which, for a binary signal, is accomplished by deciding whether the amplitude of the signal-plus-noise is above or below a certain threshold level, the interference is virtually removed. There is, however, the possibility of introducing an error in the code whenever the interference masks the signal to such an extent that an incorrect decision is made. A method of minimizing the error is to precede the decision detector with a filter (known as a matched filter), which maximizes the signal-to-noise power ratio, and to arrange for the sampling of each pulse to occur at the instant when the signal power is at a maximum. There is also the requirement of zero crosstalk at the sampling instants and ideally the overall transfer function of the system should be chosen accordingly In the design of radio systems handling digital signals both of these requirements need to be considered, whereas in line systems the signal-to-noise ratio is sufficiently high to dispense with the matched filter; instead, the effective bandwidth of the line is widened by equalization in order that a higher digit rate may be achieved.

The effect of making a wrong decision depends upon the kind of information being transmitted; an error of 1 in 10^3 could seriously impair computer data, whereas a similar error rate, associated with a PCM commercial telephony signal, would cause little or no disturbance to the restored analogue message. On circuits handling both PCM and general data the unprotected error rate, i.e. the digit-by-digit error liability due to interference and not the protected error rate of a system that includes error detection and correction, must be low, say 1 in 10^3 or better, and hence the disturbance to the reconstituted analogue information of the PCM signal is negligible. On circuits handling PCM only, the capacity may be increased-and hence the required bandwidth to such an extent that the error rate becomes relatively high, say 1 in 10^2, and the resulting distortion of the analogue output can be significant; that PCM suffers a threshold effect similar to other modulation systems.

Usually, for the transmission of computer data, an error rate of 1 in 10^3 is unacceptably high and the information-carrying digits must be protected with additional digits arranged in a coded sequence. Before deciding on a coding format it must be known whether the errors are likely to occur singly or in clusters. These are considered for a non-fading signal in the presence of band-limited noise of Gaussian distribution and, although a somewhat idealized condition, it serves as a useful introduction to the autocorrelation function.

Apart from determining the optimum pulse shape to minimize crosstalk, there is also the question of deciding upon the polarity of the pulses for line systems. In a binary system, for example, the two states could be represented as pulse or no-pulse, i.e. an on-off signal, or possibly by equal amplitude pulses but of opposite polarity, known as a polar signal. That for a given transmitter powers the pulse polarity representation has a significant effect upon the digit error rate. It will also be seen that there is a marked difference between the frequency spectra of the signals; a random sequence of on-off pulses can have both a continuous and a discrete line spectrum, the latter being particularly useful for synchronizing the decision detection sampling of the incoming digits. On the other hand, a polar pulse train only exhibits a continuous spectrum and the timing information must be extracted by first converting the pulse sequence into an on-off form.

3.2 Pulse Shape and Intersymbol Interference (Crosstalk)

It was shown on page 128 that crosstalk can be eliminated by using sin x/x pulses transmitted at a rate of twice the cut-off frequency of the low-pass filter that forms the pulses. The main drawback to the implementation of this method of signaling is the extreme accuracy to which the transmitting pulse rate and receiver sampling rate would have to be kept if the overlapping oscillatory tails are to cause no concern-there is also the practical difficulty of making an ideal low-pass filter! An alternative arrangement is to use a filter with a more gradual cut-off, thereby reducing the oscillatory tails of the output response whilst maintaining the overlapping zero crosstalk feature of the sin x/x response. A physically realizable filter meeting these requirements can be defined by an expression incorporating the ideal low-pass filter characteristic modified by a sinusoidal amplitude-frequency characteristic having odd-symmetry about the cut-off frequency (see Figure 3-2). The amplitude-frequency response of the resultant transfer function may be defined as[1]

$$|H(j\omega)| = \frac{1}{2}\left[1 + \sin\frac{\pi(\omega_0 - \omega)}{2\omega_x}\right], \quad \omega_0 - \omega \le \omega < \omega_0 + \omega$$
$$= 1, \qquad \omega < \omega_x \tag{3.1}$$
$$= 0, \qquad \omega > \omega_0 + \omega_x$$

And for simplification of the mathematics a linear phase-frequency response is assumed, a condition which, in practice, can be approached more closely with this type of filter than it can for one with extremely

Sharp cut-off. The amplitudes of the oscillatory tails of the output response are smallest when $w_x = w_o$, i.e. the most gradual cut-off (see Figure 3.2) Ideal low-pass and cosine roll-off filter characteristics having the impulse. (The dashed curve of (c) shows the filter characteristic having the impulse Response h (t) when the input is a full-width rectangular pulse.)

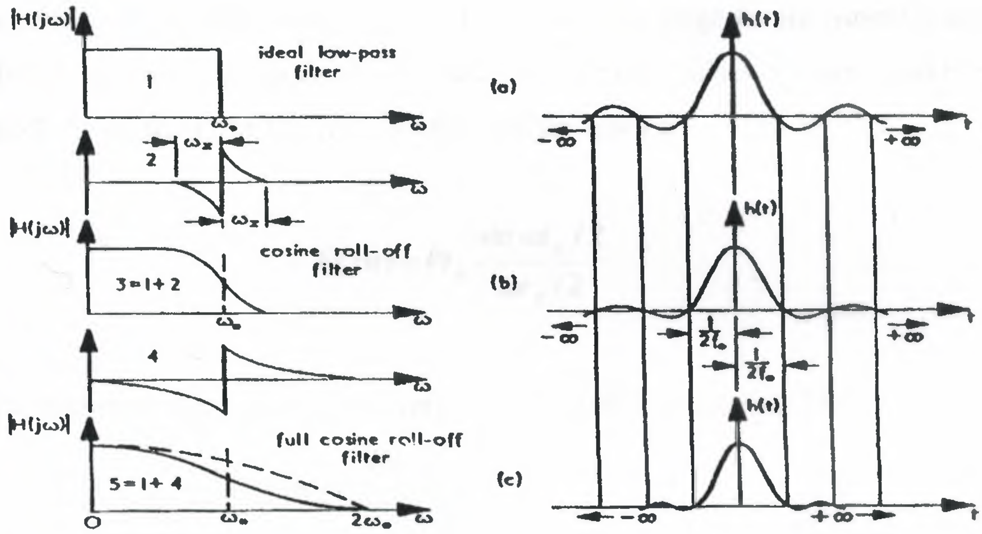


Figure 3.2 (c): This is known as the Full-Cosine Roll-off Characteristic and the Amplitude-Frequency Response of the Transfer Function is

The output response to a pseudo-impulse function with Fourier transform F (jw) k, is

$$\left.\begin{aligned}|H(j\omega)| &= \frac{1}{2}\left[1 + \cos\frac{\pi\omega}{2\omega_0}\right],\\ &= \cos^2\frac{\pi f}{4 f_0}\end{aligned}\right\} \quad 0 < \omega < 2\omega_0 \tag{3.2}$$

The out put response to a pseudo-impulse function with Fourier transform f (jw) =k, is

$$h(t)=\int_{-2f_0}^{2f0}\frac{k}{2}\left[1+\cos\left(\frac{\pi\omega}{2\omega_0}\right)\right]\exp(jwt)df$$
$$=\frac{k\omega_0}{\pi}\frac{\sin 2\omega_0}{2\omega_0 t}\frac{1}{1-(2\omega_0 t/\pi)^2} \quad (3.3)$$

And in Figure 3.2 it can be seen how this response compares with the corresponding low-pass filter response, namely

$$h(t)=\frac{k\omega_0}{\pi}\frac{\sin\omega_0 t}{\omega_0 t} \quad (3.4)$$

For zero crosstalk, the sampling rate$f_3 = 2f_o$.

So far we have considered the filter whose impulse response is the required pulse shape for zero crosstalk. In practice, the pulse would not be generated from an impulse, since the amplitude of the latter would have to be exceedingly large for the output to contain a significant amount of energy. If, instead, rectangular pulses f_0, say, width t_p, and amplitude V are used, the Fourier transform of the input is

$$F(jw)=Vt_p\frac{\sin\omega t_p/2}{\omega t_p/2} \quad (3.5)$$

And the amplitude-frequency characteristic of the required shaping filter is

$$|H(jw)|=\frac{|G(jw)|}{|F(jw)|} \quad (3.6)$$

Where G (jw) is the Fourier transform of the output response given by Equation (3.3), i.e. G (jw) must have an amplitude-frequency characteristic that is identical to the expression of Equation (3.4); hence

$$|H(jw)|=\frac{\frac{1}{2}[1+\cos(\omega\pi/2\omega_0)]}{\sin(\omega t_p/2)/(\omega t_p/2)}, \quad 0<\omega<2\omega_0 \quad (3.7)$$

(Note that Vt_p, and k, both of which have the dimensions of volt-seconds, have been omitted.) As an example, for full-width rectangular pulses, $t_p = 1/fs = 1/2f_o$ and the transfer function of the shaping filter is shown by the dashed curve in Figure 3.2 (c).

It has been explained in the introduction to this chapter that noise can be removed from the received signal by sampling each digit when it has reached its maximum value, which is usually at the centre of the time slot for the digit. A predilection filter must not only maximize the signal-to-noise ratio but also produce an output that will satisfy the zero crosstalk requirements described above. Take, for example, a sequence of uncurbed rectangular pulses which is received with little Signal-to-noise power ratio for the decision detector to produce a reformed signal containing the minimum number of errors? The answer is no, but, to substantiate this comment, let us consider the general transmission system in which pulse shaping (or filtering) is included at both ends of the link (see Figure 3.3). The input f (t) corresponds to a

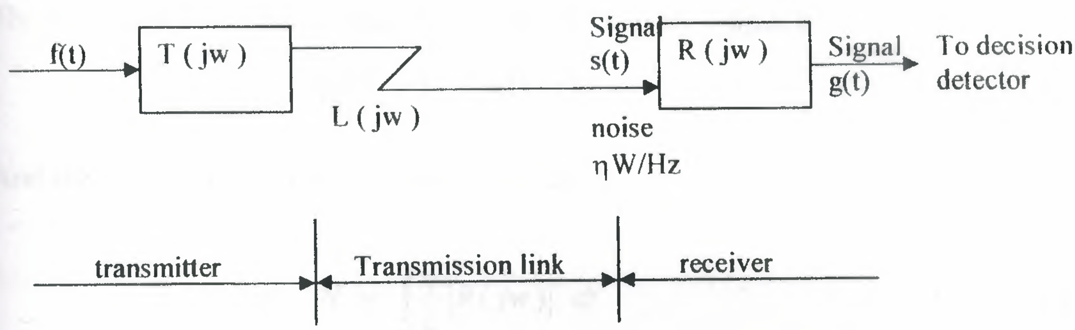


Figure 3.3 Basic Transmission System Defined in Terms of three Transfer Functions: T (jw), L (jw) and R (jw).

Single pulse of the digital signal; the associated output response g (t) must satisfy the condition of zero crosstalk. The Fourier transform of g (t) can be expressed as

$$G\,(jw) = F\,(jw)\; T\,(jw)\; L\,(jw)\; R\,(jw) \qquad (3.8)$$

Where the various transfer functions are defined in Figure 3.3. Since g (t) is specified, G (jw) is known and therefore

If F (jw) = k, i.e. a pseudo-impulse input,

$$T(jw)\, \mathrm{L(jw)}\, \mathrm{R(jw)} \equiv \frac{1}{2}\left[1 + \cos\left(\frac{\pi\omega}{2\omega_0}\right)\right] \qquad (3.9)$$

And if F (jw) = $Vt_p \sin(wt_p/2) / (wt_p/2)$, i.e. a rectangular pulse input

$$T(jw)L(jw)R(jw) \equiv \frac{\frac{1}{2}\left[1+\cos(\pi\omega / 2\omega_0)\right]}{\sin(\omega t_p / 2)/(\omega t_p / 2)} \quad (3.10)$$

We now wish to know if a proportion of the overall amplitude-frequency characteristic of the transmission system can be allocated to R (jw) to maximize the signal-to-noise power ratio at the input to the decision detector. Consider the received signal, not including noise, to have an amplitude-frequency spectrum S (jw)-corresponding to a time-response s (t) which at this stage need not be explicitly known. We have

$$G(jw) = S(jw)\,R(jw) \quad (3.11)$$

The one-sided noise power spectral density of the filter output is

$$\eta |R(jw)|^2 \text{ Watts / Hz} \quad (3.12)$$

And the total output noise $N = \int_0^{\infty} \eta |R(jw)|^2 df$

$$\text{or} \quad N = \int_{-\infty}^{\infty} \frac{\eta}{2} |R(jw)|^2 df \quad (3.13)$$

We require to maximize the ratio g $(t)^2$ /N at some instant in time, say t = T_d, i.e.

$$\frac{|g(T_d)|^2}{N} = \frac{\left|\int_{-\infty}^{\infty} S(jw)R(jw)\exp(jwT_d)df\right|^2}{\frac{\eta}{2}\int_{-\infty}^{\infty} |R()jw|^2 df} \quad (3.14)$$

The maximization of this expression can be accomplished by applying the Schwarz inequality2

$$\int p^* \, pdx \int Q^* \, Qdx \ \geq \left|\int p^* \, Qdx\right|^2 \quad (3.15)$$

Where P is the complex conjugate of P; the equality applies when P = constant x Q. For the present application let

$$\begin{aligned} p^* &= S(jw)\exp(jwt_d) \\ \textit{and} \quad Q &= R(jw) \end{aligned} \quad (3.16)$$

But S (jw) = S*(jw) and R (jw) = R*(-jw) for the input and output time responses to be real, therefore

$$\frac{|g(T_d)|^2}{N} \leq \frac{\int_{-\infty}^{\infty}|R(jw)|^2 df \int_{-\infty}^{\infty}|S(jw)|^2 df}{\frac{\eta}{2}\int_{-\infty}^{\infty}|R(jw)|^2 df} \tag{3.17}$$

$$\frac{|g(T_d)|^2}{N} \leq \frac{\int_{-\infty}^{\infty}|S(jw)|^2 df}{\eta} \tag{3.18}$$

But $f = \int_{-\infty}^{\infty} |S(jw)|^2 df$ k/f is equal to the energy contained in the pulse and the equality applies, i.e. the signal-to-noise power ratio is a maximum at t = T_d, when

$$S^*(jw)\exp(-jwT_d) = cons\tan t \times R(jw) \tag{3.19}$$

And the filter is said to be matched to the input signal-hence the term matched filler. Note that tile constant in Equation (3.20) has the dimensions of volt-seconds.

As a side issue, consider tile impulse response of the matched filter, which is

$$\begin{aligned} h(t) &= \int_{\infty}^{\infty} R(jw)\exp(jwt)df \\ &= \int_{-\infty}^{\infty} S^*(jw)\exp\left[jw(t-T)\right]df \end{aligned} \tag{3.20}$$

When S (jw) is real, i.e. the input s (t) is symmetrical, h (t) = S (t-T_d), which means that the impulse response is a delayed replica of the input signal s (t) to which the filter is matched;

Using Equation (3.10and 3.20),

$$G(jw) = \frac{1}{cons\ \tan t}|S(jw)|^2 \exp(-jwT_d)Vs \tag{3.21}$$

or

$$cons\ \tan t|R(jw)|^2 \exp(-jwT_d)Vs \tag{3.22}$$

Again, note tile dimensions of the constant term and that the equations are dimensionally correct. The amplitude-frequency characteristic of the pulse g (t) having a 1 V peak amplitude and giving zero crosstalk is

$$|G(jw)| = \frac{1}{2}\left[1 + \cos\left(\frac{\pi\omega}{2\omega_0}\right)\right] = \cos^2\left(\frac{\pi f}{4 f_0}\right) \text{ Volt-seconds} \qquad (3.23)$$

And therefore from Equation (3.23) the matched filter must have an amplitude frequency response

$$|R(jw)| = \left[\cos\left(\frac{\pi f}{4 f_0}\right)\right] \qquad (3.24)$$

And, of course, a linear phase-frequency characteristic giving rise to a constant time delay Ta. Having established the characteristics of the receiving filter, it is now possible to determine T (jw) L(jw) by using either Equation (3.9) or Equation(3.10) whether the input f(t) is model, as used in Chapter 6, would result in a complicated expression for T(jw).

We are now beginning to see that the ideal condition of zero crosstalk, which is perfectly feasible when considered in terms of the output response of a network defined by a single transfer function, is of dubious implementation for a communication system made up of many stages in cascade. Nevertheless, the problem of maximizing the signal-to-noise power ratio before decision detection and at the same time minimizing the crosstalk is at the forefront of the design of any digital communication system.

For line systems the problem of designing the complicated filters postulated for zero crosstalk can be alleviated by using non-ideal pulses, e.g. rectangular, setting the digit rate according to the available bandwidth of the equalized line and allowing a finite but acceptable level of inter symbol interference or crosstalk. Regenerative repeaters, which provide the equalization of the line characteristics, also maintain the signal-to-noise ratio at a relatively high level and hence a matched filter is not an essential requirement.

To establish these points consider the following example.

$$S(jw) = F(jw)L(jw) = Vtp\frac{\sin(wtp/2)}{(wtp/2)}\frac{1}{1 + jf / f_{8dB}} \qquad (3.25)$$

The transfer function of the matched filter for this input s(t) can be found from Equation (3-12), but before attempting its evaluation let us examine the impulse response which, for an unsymmetrical input as s(t) is in this example, is a delayed replica of s(t) run backwards in time. Such an impulse response in unrealistic, since it commences at t = -∞ therefore the matched filter cannot be realized and we are thus relieved of the task of determining its transfer function. Instead, assume that a sin x/x filter is used, which is the matched filter for an uncurbed rectangular input when x w_{tp} The Fourier transforms G (jw) of the output g (t) corresponding to the input s (t) is

$$G(jw) = S(jw)\,R(jw)$$

$$= Vtp\frac{\sin(wtp/2)}{(wtp/2)}\frac{1}{1+jf/f_{8dB}}Vtp\frac{\sin(wtp/2)}{(wtp/2)} \tag{3.26}$$

An expression for g (t) must be established in order that the peak power and crosstalk may be determined; the inverse Fourier integral is not recommended as a method of solution, but rather the use of the Convolution Integral,

$$g(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)d\tau \tag{3.27}$$

Referring to Figure 3.4, the received distorted pulse is defined as

$$s(\tau) = \begin{cases} = V[1-\exp(-\tau/RC)] & 0<\tau<t_p \\ = V\exp(-\tau/RC)[\exp(t_p/RC)-1] & \tau > t_p \end{cases} \tag{3.28}$$

Where V is the amplitude of the uncurbed rectangular pulse f (t) and for the RC network simulating the line, f3dB = 1/2πRC. The transfer function of the receiver filter is (Note that the tp term in the denominator of the expression for h (t)) appears as a result of the definition R (jw); h (t) has the dimensions of volts, since F_o (jw) =1 volt-second. In the following analysis the filter time delay Td has been ignored; however, it may be included simply by writing ($t=T_d$) in place of t.

There are four conditions of t for which the Convolution Integral must be evaluated (see Figure 3.4 (b)):

(i) When t<0, g (t) = 0

(ii) When 0< t< t_p,

$$g(t) = \int_0^t V[1-\exp(-\tau/RC)]\frac{1}{tp}d\tau$$
$$= \frac{V}{tp}\{t + RC\}[\exp(tp/RC)-1] \tag{3.29}$$

And is a maximum when t = t_p,

(iii) When $t_p < t < 2t_p$,

$$g(t) = \int^{v} [1 - \exp(-\tau / RC)] \frac{1}{tp} d\tau$$

$$+ \int \text{Vexp(- } \tau\text{/RC) - 2exp(tp/RC) - 1\}} \frac{1}{\text{tp}} \qquad (3.30)$$

$$= \frac{V}{tp} \{2tp - t + RC[\exp(-t / RC) - 2\exp[-(t - tp)] / RC] + 1\}$$

(iv) When t> ^{2}tp,

$$g(t) = \int V \exp(-\tau / RC)[\exp(tp / RC) - 1] \frac{1}{tp} d\tau \qquad (3.31)$$

$$= \frac{V}{tp} \{RC[\exp(tp / RC) - 1][\exp[-()t - tp / RC] - \exp(-t / RC)]\}$$

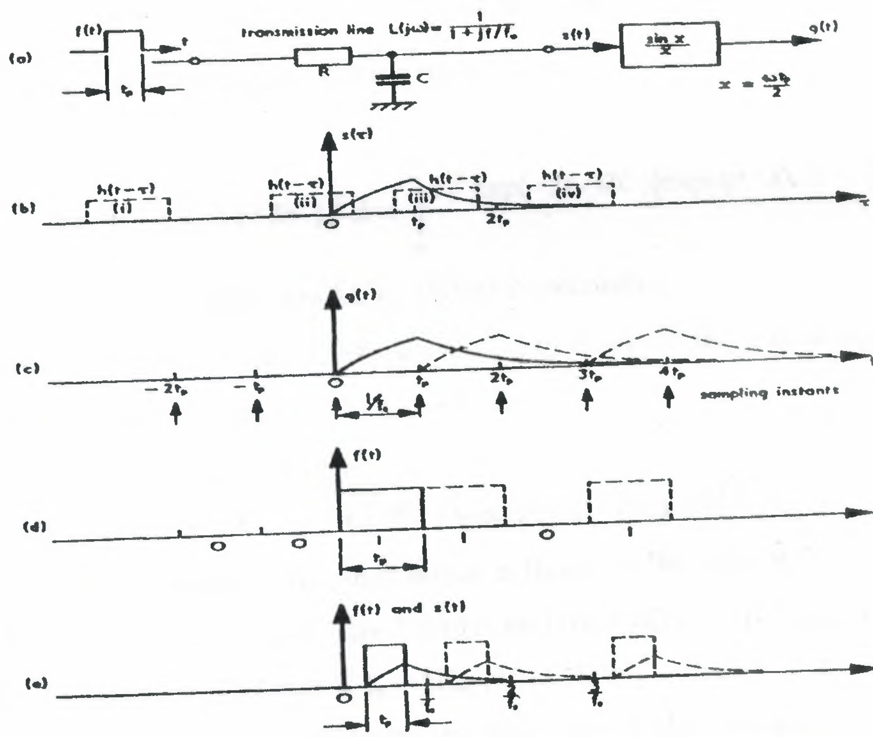


Figure 3.4 (a) The transmission system. (b) Illustration of the impulse response h (t-τ) of the sin x/x filter in relation to s& for the evaluation of g (t) using the Convolution Integral. (c) Output response of the sin xix filter, and (d) The corresponding input f (t) consisting of full-width pulses. (e) Waveforms for half-width (50% duty cycles) pulses.

$$N = \frac{\eta}{\pi tp}\int_{\infty}^{0}\frac{\sin^2 x}{x^2}dx = \frac{\eta}{\pi tp}\frac{\pi}{2} = \frac{\eta}{2tp} \tag{3.32}$$

Let us raise the question, albeit hypothetical, of how this filter compares with the matched filter, had it been physically realizable, in terms of the output signal-to-noise power ratio. To answer this question it is necessary to know the spectral power density of the noise at the input to the filters, which, for simplification of the analysis, can be assumed uniform and equal to η watts/Hz-a condition that is not valid for a practical system, but one which is justifiable for the present comparison.

The mean noise power output from the sin x/x filter is

$$N = \int \eta \left| \frac{\sin(wtp/2)}{(wtp/2)} \right|$$

And putting (3.33)

$$N = \frac{\eta}{\pi tp}\int_{\infty}^{0}\frac{\sin^2 x}{x^2}dx = \frac{\eta}{\pi tp}\frac{\pi}{2} = \frac{\eta}{2tp} \tag{3.34}$$

The energy contained in the input pulse s (t) is

$$E = \int_0^{tp} V^2[1-\exp(-t/RC)]^2\,dt + \int_{tp}^{\infty} V^2\exp(-2t/RC)[\exp(tp/RC)-1]^2\,dt \tag{3.35}$$

$$= V^2\{tp + +RC[\exp(-tp/RC)-1]\}\text{Volt2 - seconds}$$

Using Equation (3.13), (3.33), (3.34) and (3.35) to compare the output signal-to-noise ratios of the matched and sin xix filters, we have

$$\frac{2E}{\eta} \div \frac{[g(t)_{\max}]^2}{N} = \frac{1}{1 + RC/tp[\exp(-tp/RC)-1]} \tag{3.36}$$

For a numerical comparison we must attach a figure to the ratio RC/t assuming that full-width pulses are used (see Figure 3.4 (d)), and the 3 dB cut-off frequency of the line is equal to half the sampling rate fi, $t_p = 1/f_s$, $1/2\pi$ CR

$1/^2$tp and hence $RC/t_p = 1/\pi$. Substituting this figure in the above equation yields a 16 dB degradation of the signal-to-noise ratio for the sin xix filter compared with the match filter. Whether or not this is a significant reduction depends upon the actual signal-to-noise ratio; its effect in terms of the probability of making a wrong decision regarding the state of the signal at the specified instant in time is studied in the next section.

The second important issue is to determine the inter symbol interference (or crosstalk) which for full-width pulses is defined as the ratio of the maximum amplitude-to-amplitude of the pulse t seconds later (see Figure 3.4 (c)). Using Equation (3.35) and substituting $t = 2t_p$ (note the origin $t = 0$ coincides with the commencement of the pulse) and $RC/t_p = 1/\pi$, we get $g(2t_p) = 0.291$ V and using Equation (3.36) the maximum pulse amplitude $g(t)_{max}$ $g(tp) = 0.696$ V; hence the crosstalk ratio is 7.6 dB. Repeating the calculations for half-width

Pulses, but again with $f_{3dB} = fs/2$, $RC/t_p = 2/\pi$, the signal-to-noise ratio reduction when comparing the sin xix and matched filter outputs is 3 dB and the crosstalk ratio, calculated as $g(t)_{max}$: $g(3t_p)$, is 15.6 dB-note, the smaller the inter symbol interference the higher the crosstalk ratio.

It is instructive to evaluate the crosstalk ratios of the pulses before filtering, i.e. for s (t) in place of g (t). Using Equation (3.36), for the full-width pulse and $RC/t_p = 1/\pi$, s (2tp) 0.042 V and $s(t_p) = 0.95$ V giving a crosstalk ratio of 27 dB-cf. 7.6 dB corresponding to the filtered pulse g (t). For half-width pulses and $RC/t_p = 2/\pi$, the crosstalk ratio before filtering is 27 dB, calculated as $s(t_p)$: $s(3t_p)$, compared with the 156 dB figure obtained for the filtered output. We conclude that filtering to maximize signal-to-noise ratio increases the inter symbol interference. On line systems the signal-to-noise ratio can be kept relatively high and therefore the matched filter, or its nearest equivalent, can be dispensed with.

Before leaving this section, let us examine the assumed relationship between the 3 dB cut-off frequency of the RC model of the line and the digit rate, namely f_a dB = $f_s/2$. which is analogous to the relationship between sampling rate and SyQuest bandwidth introduced in Chapter 6. Referring to the insertion loss-frequency characteristic for 20 SWG (5.6 kg/kin) polyethylene insulated wires, it can be seen that the 3 dB cut-off frequency for a 1.83 km (6000 feet) length is in the region of 30 kHz, which, according to the above condition, would impose a maximum limit on the digit rate of 60 kilobits/second. However, it is common practice in 24-channel PCM systems carrying speech information to use digit rates of the order of 1.5 M bits/second over this type of cable by using equalizers in the repeaters that effectively extend the 3 dB (for a F83 km line) cut-off frequency from 30 kHz to the order of 300 kHz.[3] We conclude that in line systems it is not only essential to dispense with matched filters on the grounds of excessive crosstalk, but it is also desirable to extend the bandwidth of the line by equalization- both conditions are dependent upon a high signal-to-noise power ratio being maintained throughout the link. These conclusions do not apply to radio systems in which the signal is more vulnerable to interference during transmission; filtering at the receiver to improve the signal-to-noise ratio is essential. This is considered to be outside the undergraduate curriculum and therefore beyond the scope of this book.

3.3 Decision Detection-Error Probability

The function of the decision detector is to determine the state of each digit out of a finite number of possibilities-two for a binary signal-in order that a re-formed noise-free pulse sequence may be generated; in a regenerative repeater the signal would be processed in this way for further transmission, whereas at the final destination the re-shaped digital signal would be either decoded, if it were a PCM or AM signal, or routed to the appropriate data-handling machine or store if general data. A typical on-off signal before transmission and the corresponding waveform at the input to the detector are shown in Figure 3.5. The amplitude levels of the received signal if unaccompanied by noise would be 0 and V; assuming an equal probability of the two binary states, i.e. p (0) = p (l) = 1/2, a slicing (or threshold) level is set at V/2. At the instant of inspection, which is made to occur when a pulse has reached its maximum amplitude, the detector decides whether the signal plus noise is above or below the slicing level and generates the appropriate output of 1 or 0. We are concerned with the noise masking the signal to such an extent that a wrong decision is made-as illustrated in the figure by the third digit being in error.

There are two conditions to be studied; the reception of no pulse, i.e. noise alone, or a pulse plus noise.

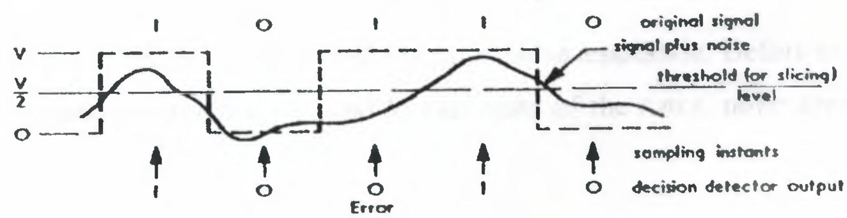


Figure 3.5 Decision Detection-the Effect of Noise on an on-off Pulse Transmission.

(i) No pulse. Noise alone will produce an error whenever its amplitude at the instant of inspection exceeds the slicing level. If we assume the noise to have a Gaussian distribution, the probability of an error is given by the area under the probability density curves (see Figure 3.6 (a)), and is

$$P\left(v > \frac{V}{2}\right) = \int_{\frac{V}{2}}^{\infty} \frac{1}{\sqrt{2\pi N}} \exp\left[-\frac{(v)^2}{2N}\right] dv \tag{3.37}$$

Where N is the mean noise power.

(ii) Pulse plus noise. An error will occur whenever the resultant amplitude is less than V/2 at the instant of inspection; the probability of an error is given by

$$P\left(v < \frac{V}{2}\right) = \int_{\infty}^{\frac{V}{2}} \frac{1}{\sqrt{2\pi N}} \exp\left[-\frac{(v-V)^2}{2N}\right] dv \tag{3.38}$$

And is illustrated in Figure 3-6 (b).

It can be seen from tile symmetry of the curves that

$$P\left(v > \frac{V}{2}\right) \text{ noise alone} = \mathrm{P}\left(\mathrm{v} < \frac{\mathrm{V}}{2}\right) \text{pulse noise} \tag{3.39}$$

And since there is equal probability of 1 or 0 being transmitted, P (v> V/2) can be used to determine the overall error probability P_e. Error function (erf) tables give numerical values of P ($-x_1<x<xi$) where x is in units of the r.m.s. value, or we may write

$$P(x>xi) = j\ \{l-P(-x_1<x<xi)\}$$
$$=1/2\{1-\mathrm{erf}\ x_1\}$$

. In the present application,

$$Pe = \frac{1}{2}\left\{1 - P\left(-\frac{V}{2} < v < \frac{V}{2}\right)\right\} \tag{3.40}$$

Where V/2 is the slicing level set at half the peak pulse amplitude. Before erf tables can be used it is necessary to convert v and V into units of the r.m.s. noise amplitude. We have

1

$$let \frac{v}{\sqrt{2N}} = y$$

$$2\int_0^{\frac{v}{2}} P(v)dv$$

$$= 2\int_0^{\frac{v}{2\sqrt{2N}}} \frac{1}{\sqrt{\pi}} \exp(-y^2)dy$$

Which is a tabulated function as

$$fx = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)dy$$

$$where \quad x = v / 2\sqrt{2N}$$

2

$$let \frac{v}{\sqrt{N}} = y$$

$$2\int_0^{\frac{v}{2}} P(v)dv$$

$$= 2\int_0^{\frac{v}{2\sqrt{2N}}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})dy$$

which is a tabulated function as

$$fx = \sqrt{\frac{2}{\pi}} \int_0^x \exp\left(-\frac{y^2}{2}\right)dy$$

$$where \quad x = V / 2\sqrt{N}$$

$$P_e = 1/2\{1\text{-erf}x\} \tag{3.41}$$

Erf tables of both forms are available and care must be exercised to ensure which set is being consulted; a short form table using substitution 1 is to be found. As a numerical illustration, when the peak signal-to-mean noise power ratio of an on-off signal is 17.4 dB, $V/\sqrt{N}$ 7.45, $x = 7.45/2\sqrt{2}$ and from the tables erf x = 0.9998 resulting in an error probability $P_e = 0.0001$ (1 in 10^4). Before plotting P_e against the signal-to-noise ratio, let us look at another form of signaling in which the two binary states are represented as equal positive and negative amplitude pulses, usually referred to as a polar signal. If a 1 is represented as + V/2 and a 0 as - V/2, the amplitude difference V is the same as for the on-off signal and, with equal probability of 1 and 0 occurring, the slicing level is set at zero volts. Under these conditions the probability of an error can be determined from the above analysis. A difference between the two forms of signaling and the common use of the analysis becomes apparent if we attempt to equate Pe to the mean signal-to-noise power ratio. For the on-off signal the peak pulse power is proportional to V^2 and the mean signal power proportional to k $V^2/2$ where k accounts for the pulse shape and the factor 1/2 arises, since the number of pulses is half the number of digits transmitted, assuming P(l) = 1/2. In the polar system the peak pulse power is proportional to $(V/2)^2$ and, if we assume the same pulse shape as for the on-off signal, the mean signal power is proportional to $kV^2/4$, since there is a pulse for every digit transmitted. We conclude that for equal probability of an error the mean signal power of the polar signal is 3 dB less than that of the on-off signal.

It is often desirable to compare the performance of two or more systems operating under entirely different conditions of available signal power, digit rate and pulse shape; the signal-to-noise power ratios and error probabilities may be known either from measurement or assessed by calculation. The mean signal-to-noise power ratio can be expressed as

$$\frac{S}{N} = \frac{Efs}{\eta B_n} \tag{3.42}$$

Where E is the energy per pulse,

f_s is the digit rate,

B_n is the equivalent noise band width, and

η is the one-sided uniform spectral power density of the noise.

Re-arranging gives, by definition, the normalized signal-to-noise power ratio

$$R\frac{E}{\eta} = \frac{averge\ energy\ per\ pules}{one - sided\ noise\ power\ density} = \frac{S}{N}\frac{B_n}{fs} \tag{3.43}$$

This provides a basis for comparing systems operating under different conditions Re-arranging the right-hand side of Equation (3.43),

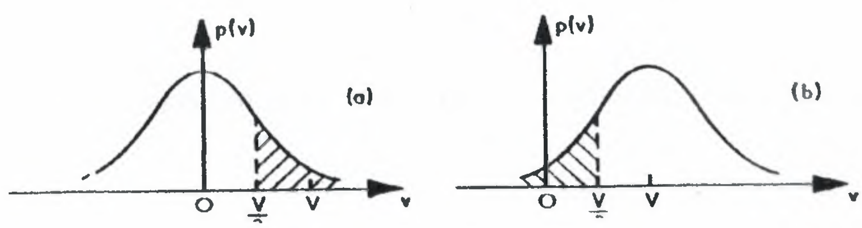


Figure 3.6 Probability Densities in on-off Pulse Transmission-Shaded Areas Indicate Probability of an Error for (a) no Pulse, Arid (b) Pulse Plus Noise.

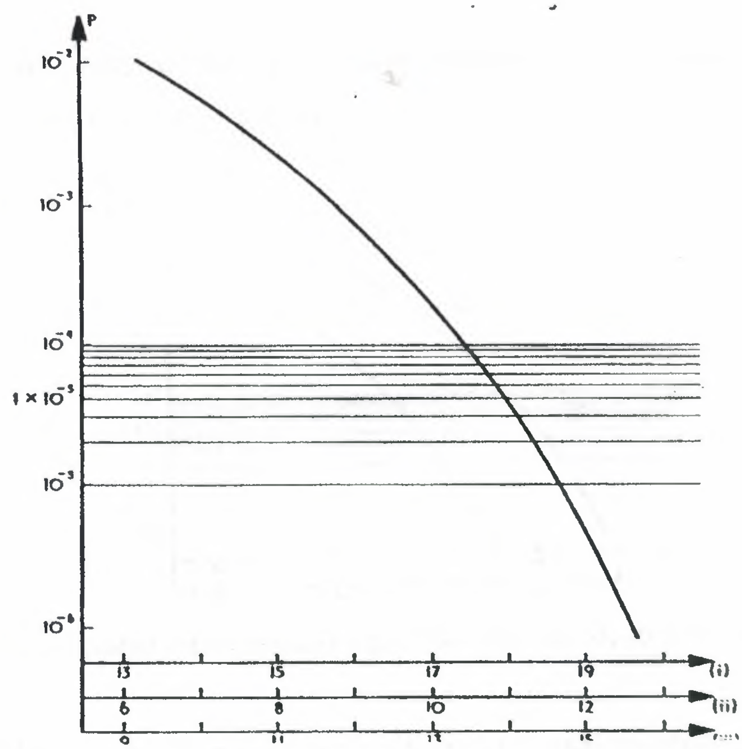


Figure 3.7 Error Probability Against Signal-to-Noise Ratio, in dB, for an on-off Signal Made up of Full Width Rectangular Pulses.

X-axes: (i) peak signal-to-mean noise power ratio, V^2/N;(ii) normalized signal-to-noise power ratio, and mean signal-to-noise power in the bit-rate bandwidth;(iii) Mean signal-to-noise power ratio, V^2/N -4 dB.

Normalized signal-to-noise ratio = mean signal power /noise power in the bit-rate bandwidth (3.44)

And furthermore by substituting $f_s = 2f$ .where f_o is the Nyquist bandwidth, we have

$$\overline{e^2(t)} = \frac{2}{ts}\int_0^{t/2}\left[\frac{(\Delta v)t}{ts}\right]^2 dt = \frac{(\Delta v)^2}{12}$$

$$= \frac{1}{2}\left\{\frac{\textit{mean signal power}}{\textit{noise power in the Nyquist bandwidth}}\right\} \quad (3.45)$$

All three definitions which are to be found in common use are illustrated in Figs. 3.7 and 3.8 where the error probabilities of the on-off and polar forms of signaling are compared; it is assumed that k the pulse shape factor is unity and that optimum pulse shaping for zero crosstalk Prior to detection is used; refer to Equation (3.5). It may be shown that, for this filter, $B_n \approx f_0$ or $f_s/2$.

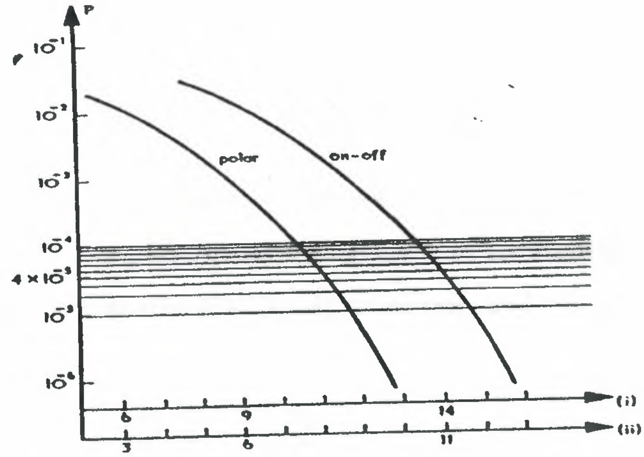


Figure 3.8 Error probabilities against signal-to-noise ratio, in dB, for on-off and polar signals.
X-axes:(i) Mean signal-to-noise ratio, or mean signal-to-noise power in Nyquist bandwidth, and(ii) Mean signal-to-noise power in bit-rate bandwidth.

3.4 Error Probability and its Effect upon the Output Signal-to-noise Ratio of PCM and ΔM Systems

3.4.1 Pulse Code Modulation (PCM)

A PCM encoder quantizes the amplitude range of the analogue input into M states to allow conversion into binary code groups of n digits, where $M = 2^n-1$. The output from a complementary decoder is inevitably a distorted version of the original signal, due to the process of quantization, but in the absence of digital errors this is the only source of noise in the output-interference accompanying the signal during transmission will have been removed by the decision detector. When the fluctuation noise is sufficiently large to cause a digital error in a code group, the decoder generates an incorrect amplitude level, thus giving rise to additional noise (or distortion), the degree of which depends upon the position in the code of the digit in error. Assuming linear quantization, an error in the least significant digit will cause the output to differ from the correct amplitude by the quantized step size Δ v, and an error in the next significant digit would cause an output voltage difference of 2Δ v; in general, an error in the r^{th} digit would give rise to an output error voltage $\Delta v\ 2^{(r-1)}$. If all digits in a code group of n digits are equally likely to be in error with probability P_e, and only one error will occur in the code group, the mean square noise voltage N_e is approximately

$$\sum_{r-1}^{n} p_e\{\Delta v 2^{r-1}\}^2$$

$$N_e = P_e(\Delta v)^2 \sum_{r-1}^{n} 2^{2(r-1)} \tag{3.46}$$

If we consider the analogue signal to be a sine wave of peak amplitude V and occupying the whole of the available amplitude range, then $2V = M\,\Delta v$ and the mean square signal power S is noise ratio, considering fluctuation noise only and not quantization noise, is

$$\frac{S}{N_e} = \frac{M^2}{8P_e \sum_{r-1}^{n} 2^{2(r-1)}} \tag{3.47}$$

This equation is only valid for low error rates, say less than 1 in 100 digits, $P_e < 10^{-2}$; otherwise the signal component cannot be identified separately, as in Equation (3.46), and hence the signal-to-noise ratio becomes the signal plus noise-to-noise ratio.

As a numerical illustration of Equation (3.47) when $P_e = 10^{-3}$ and n 7 (M = 128), the signal-to-noise ratio is 34 dB. The dashed curves of Figure 3.9 show the error probability and output signal-to-noise ratio for on-off and polar signaling plotted against the mean signal-to-noise ratio at the input to the decision detector- i.e. the graph involves the use of Equation (3.36), which is also illustrated in Figure 3.8, and Equation (3.47). The full curves of Figure 3.9 include quantization noise which we see becomes the sole source of interference when the digital error probability becomes very small, 10^{-6} or less; in this example, a sine wave uniformly quantized into 128 levels is assumed which gives rise to a signal-to-quantizing noise ratio of 44 dB

3.4.2 Delta Modulation

The effects of digital errors on the output signal of a delta modulation system can be determined by representing the received re-shaped pulse sequence as an error-free pattern plus a randomly spaced series of rectangular pulses of amplitudes ±2V or -2V, the latter depending upon whether a 1 or a 0 is incorrectly received. A typical sequence is shown in Figure 3.10; it is assumed that the error-free pattern, which is identical to the original transmitted signal, and the randomly spaced error pulses are received simultaneously.

Let the probability of an error be P_e, i.e. error pulses occur on average once every $1/P_e$ digits; half of the errors will give rise to +2 V pulses in waveform of Figure 3.10, and the other half to -2V pulses. For the rectangular pulses. The one-sided error power density of the output from the RC integrator is

$$\frac{S}{N_e} \approx \frac{fs}{16P_e f_0}\left[\tan^{-1}\left(\frac{3400}{f_0}\right) - \tan^{-1}\left(\frac{300}{f_0}\right)\right] \tag{3.48}$$

And the output noise Ne from a band-pass filter having cut-off frequencies $f_{m1} = 300$ Hz and $f_{m2} = 3400$ Hz is given by

$$N_e = 2P_e fsf_0^{\,2} \int_{fm1}^{fm2} |G(jw)|^2 \frac{f_0^{\,2}}{f_0^{\,2} + f^2} df \tag{3.49}$$

And since $f_s > f_{m2}$ we may assume that $G|(jw)|^2 \approx 4\ V^2/f_s^2$, hence

$$N_e \approx \frac{8P_e V^2 f_0^{\,2}}{fs} \int_{300}^{3400} \frac{1}{f_0 + f^2} df$$

$$= \frac{8P_e V^2 f_0}{fs}\left[\tan^{-1}\left(\frac{3400}{f_0}\right) - \tan^{-1}\left(\frac{300}{f_0}\right)\right] \tag{3.50}$$

It follows that the mean signal-to-fluctuation noise power ratio corresponding to a sine wave of peak amplitude V is

$$\frac{S}{N_e} \approx \frac{fs}{16P_e f_0}\left[\tan^{-1}\left(\frac{3400}{f_0}\right) - \tan^{-1}\left(\frac{300}{f_0}\right)\right] \tag{3.51}$$

For example, when $f_o = 150$ Hz and $f_s = 56$ kilobits/second, an error probability $P_e = 10^{-3}$ would result in $S/N_e = 48$ dB-this ratio does not include quantization noise.

A similar expression to Equation (3.51) can be obtained for a basic AM system employing the double integration network of Figure 3.17, Using Equation (3.50) the output noise power N_e due to digital errors is

$$N_e = 2P_e fs \int_{fm1}^{fm2} |G(jw)|^2 \frac{f_2^2 f_1^2}{f^2(f^2+f_2^2)} df \tag{3.52}$$

$$\Delta v \approx 4V \frac{2f_0}{fs} = \frac{4}{\pi} \frac{2\pi V f_0}{fs} \tag{3.53}$$

This equation assumes that the mean output signal power is $V^2/2$. For a signal amplitude consistent with the conditions of non-overloading,

$$\frac{S}{N_e} \approx \frac{fs}{16P_e(1+w^2{}_m T^2)f_0}\left[\tan^{-1}\left(\frac{3400}{f_0}\right) - \tan^{-1}\left(\frac{300}{f_0}\right)\right]^{-1} \tag{3.54}$$

Hence the signal-to-noise power ratio is given by

$$\frac{S}{N_e} \approx \frac{fs}{16P_e f_1^2}\left\{\left(\frac{1}{fm1} - \frac{1}{fm2}\right) - \frac{1}{f_2}\left[\tan^{-1}\left(\frac{3400}{f_0}\right) - \tan^{-1}\left(\frac{300}{f_0}\right)\right]\right\}^{-1}$$

(3.55)

This is illustrated in Figure 3.9 together with a plot of the output signal-to-noise ratio against the input ratio to the pulse regenerator. As we have seen for PCM, Equation (3.54) and (3.55) only apply when the error probability is small, say $P_e < 10^2$. The continuous curve for ΔM assumes that the digital signal is in polar form and made up of full-width rectangular pulses. A system employing double rather than single integration

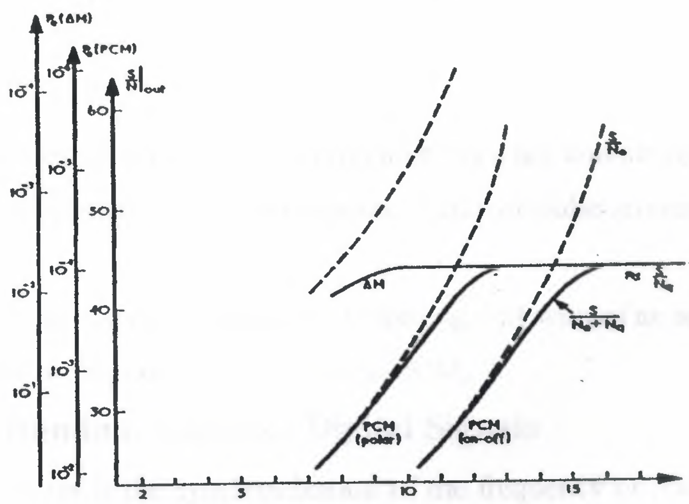


Figure 3.9 Mean Output Signal-to-Total Noise Power Ratio Plotted Against the Signal-to-Noise Power Ratio at the Input to the Decision Detector in the Pulse Regenerator. PCM: sine wave signal uniformly quantized into 12 levels.

ΔM: basic system employing double integration and designed for speech. (Broken lines represent S/N_e.)

Given by

$$S\big|_{line} = \overline{\left[av\{^{k} x(t)\}\right]^{2}} = m_1^{\,2} fs^{2} G^{2}(0) + 2m_1^{\,2} fs^{2} \sum_{r-1}^{\infty} \left|G(jrw_s)\right|^{2} \ watts \qquad (3.56)$$

And the one-sided power density of the

$$S(f)\big|_{continuous} \approx 2 fs \left|G(jw)\right|^{2} \{R(0) - m_1^{\,2} + \sum_{q=1}^{\infty} \left[R(q) - m_1^{\,2}\right] \cos wqT_s\} watts \ / \ Hz \qquad (3.57)$$

Note the factor 2 difference between Equation (3.56) and (3.57)-the latter

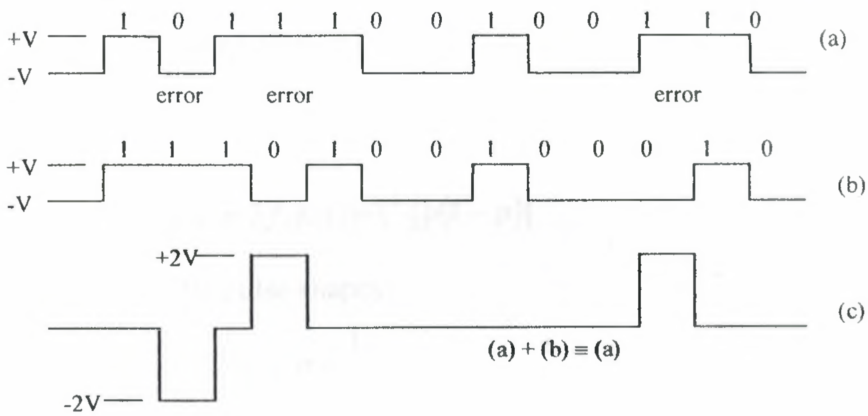


Figure 3.10 Delta Modulations. (a) Received pulse sequence containing errors.(b) Error-free sequence original transmitted sequence.(c)Error-pulse sequence.

is illustrated simply for the convenience of comparing AM with of an are chosen so that the same overall pulse amplitude range is avail- PCM,

3.5 Spectra of Random-sequence Digital Signals

A regenerative repeater is the synchronization of the frequency or phase deviation and hence the occupied bandwidth, locally generated sampling pulses to the incoming digital signal. Which are directly proportional to the amplitude range of the modulating synchronizing information is an inherent characteristic of the incoming signal. Thus with the method of normalization adopted it is possible to signal, but the question that arises is: what form does it take and how compare the various forms of signalling

3.5.1 On-Off Signaling

$a_n = 1$, with probability p

$= 0$, with probability 1-p

Therefore, $m_1 = \text{av } a_n = 1 \times p + 0(1-p) = p$

$m_2 = \text{av } (a_n^2) = 1^2 \, xp + 0^2(1-p) = p$

And $\quad R(0) = m_2 = p$

Since there is zero interdigit correlation, $av(a_n a_m) = av\, a_n\, av\, a_m$

and $R(q) = m_1^{\,2} = p^2$

Substituting these constants into Equation (3.56) and (3.57) gives

$$S\Big|_{\substack{line\,,\\ on-off}} = p^2 f_s^{\,2} G^2(0) + 2p^2 f_s^{\,2} \sum_{r=1}^{\infty} |G(jrw_s)|^2 \tag{3.58}$$

And

$$S(f)\Big|_{\substack{continuous\\ on-off}} \approx 2f_s |G(jw)|^2 \left[p(1-p)\right] \tag{3.59}$$

Let us consider specific pulse shapes

$$t_p = T_s/2 \; and \; p = \frac{1}{2}$$

$$|G(jw)|^2 = \left[t_p \frac{\sin(\frac{wt_p}{2})}{(\frac{wt_p}{2})} \right]^2 = \left[\frac{T_{sa}}{2} \frac{\sin\left(\frac{\pi f T_s}{2}\right)}{\left(\frac{\pi f T_S}{2}\right)} \right]^2$$

And $$|G(jw_s)|^2 = \left[\frac{T_s}{2} \frac{\sin(r\pi/2)}{(r\pi/2)} \right]^2$$

Hence

$$S\,\Big|line = \frac{1}{16} + \frac{1}{8} \sum_{r=1}^{\infty} \left[\frac{\sin\,(\pi r/2)}{(r/2)} \right]^2$$

And

$$S(f)\Big|_{continuous} \approx \frac{T_s}{8} \left[\frac{\sin\,(\pi f T_s/2)}{(\pi f T_s/2)} \right]^2 \; watts\;/HZ \tag{3.60}$$

The evaluation of R (q) must take into account the interdigit correlation which for this form of signaling is not zero. From the j^{th} to the $(j+q)^{th}$ digit inclusive there are (q+l) digits; the total number of possible binary patterns is 2^{q+1} Any pattern with a 0 at the beginning or end would have a product $a_j a_{j+q} = 0$, and therefore we need only consider those patterns starting and ending in 1, of which there are 2^{q-1} The interdigit autocorrelation function may be written as

$$|R(jw)| = \left[\cos\left(\frac{\pi f}{4 f_0}\right)\right] \tag{3.61}$$

Where A is tile number of patterns starting and ending with +1/2 or-1/2 and B is the number of patterns starting with +1/2 and ending with -1/2, or vice versa; this assumes that p(l) = p(0) = 0.5.

When q = 1, A = 0 and B = 1

q = 2, A = 1 and B = 1

In fact, when q > 2, A = B

And therefore

$$R(1) = (-1/4)\frac{1}{2^{1+1}} = -\frac{1}{16}$$

And $\quad R(q) = 0$ for q>2. (3.62)

The average value m_1 is zero and therefore a line spectrum does not exist. When there is equal probability of 1 and 0, i.e. p = 1/2., the one-sided power density of the continuous spectrum is given by

$$S(f)\Big|_{\substack{continuous\\ bipolar}} \approx 2f_s|G(jw)|^2\left[\frac{1}{8} + 2(-\frac{1}{16})\cos wT_S\right]$$

$$= \frac{1}{4T_S}|G(jw)|^2(1-\cos wT_S) \tag{3.63}$$

3.6 Duo binary Signalling

Like the bipolar form of signalling, the duo binary technique tie can also be classified as pseudo-binary-0 is transmitted as no pulse and 1 as either a positive or negative pulse depending upon the previous sequence. Tile choice of polarity is most easily understood by considering the conversion of an on-off sequence into duo binary form. Briefly, if there are an odd number of 0s between the previous 1 and the next to be generated, the polarity of tile latter is reversed; when the number of 0s is even the polarity is unchanged (see Figure 3.11). Unlike the bipolar form, a reversal from a positive to a negative pulse in adjacent time slots is impossible. The statistical constants may be defined as follows:

$$a_n = +\frac{1}{2} \quad = -\frac{1}{2} \quad = 0 \quad \text{with probability} \quad P/2, \quad P/2, \quad 1-P$$

$$\left.\begin{aligned} m_1 &= av\, a_n = 0 \\ m_2 &= av\,(a_n^{\,2}) = \frac{P}{4} \end{aligned}\right\} \text{as for the bioplar form,} \tag{3.64}$$

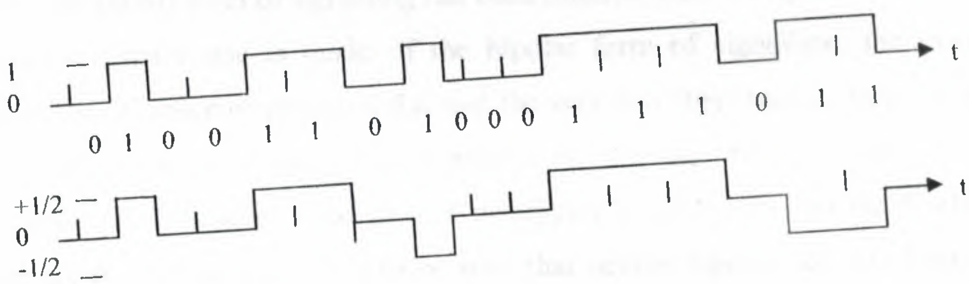


Figure 3.11 A Random on-off Sequence of Full-Width, NRZ, Rectangular Pulses and the Equivalent Duo Binary form.

R (q) may be evaluated by a similar approach to that used for the bipolar signal. Therefore we have

$$R(q) = av\,(a_j a_j + q) = (+\frac{1}{4})\frac{A}{2^{q+1}} + (-\frac{1}{4})\frac{B}{2^{q+1}} \tag{3.65}$$

$$q = 1, A = 1 \; and \; B = 0$$

$$q = 2, A = 1 \; and \; B = 1$$

When $$q = 3, A = 2 \; and \; B = 2$$

$$in \; fact, when \; q \geq 2, A = B \; and \; R(q) = 0. \; This \; leaves$$

$$R(1) = (+\frac{1}{4})\frac{1}{2^{1+1}} = +\frac{1}{16}$$

The average value, ml, is zero and therefore a line spectrum does not exist. When there is equal probability of 1 and 0, p = 1/2. and the one-sided power density of the continuous spectrum is

$$S(f)\Big|_{\substack{continuous \\ duobinary}} \approx 2f_s |G(jw)|^2 \left\{ \frac{1}{8} + 2\left(\frac{1}{16}\right)(1+\cos wT_s) \right\}$$

$$= \frac{1}{4T_s} |G(jw)|^2 \left(1+\cos wT_s\right) \qquad (3.66)$$

A comparison of the four methods of signalling is made in Figure 3-16; the curves illustrate Equation (3.61), (3.63) both for p 1/2. (3.64) and (3.66). The line spectrum associated with the on-off form of signalling has been omitted from the figure.

In line systems extensive use is made of the bipolar form of signalling, the main advantage being the absence of power at d.c. and the very low frequencies, thus easing the bandwidth requirements of the repeater amplifiers. The duo binary technique is favored for radio transmission on account of it occupying a much narrower bandwidth than the other forms of signalling. It is to be seen that neither bipolar nor duo binary signals exhibit a line spectrum, and therefore the timing information necessary for correct sampling at the repeaters and final receiver cannot be extracted linearly with a simple LC resonant circuit. In general, a non-linear operation is required; one method is to convert the signal into the equivalent on-off form at the repeater or receiver, which may then be used to excite the LC circuit.

CHAPTER FOUR

PHASE MODULATION AND CONVOLUTIONAL CODES

In previous chapters we have seen that the memory in the continuous phase of the CPM signal can be utilized to improve the minimum Euclidean distance; so also can the memory introduced by the controlled Intersymbol interference in partial response CPM. Even more memory can be built into the signals by means of multi-h coding (see Chapter 3) or convolution codes. In this chapter we will present some recent results on combinations of many-level P5K and CPM with convolutional codes. Covers binary and quaternary CPFSK with rate 1/2 convolutional codes, and presents combinations with the best free Euclidean distance. covers eight-level CPM with rate 2/3 codes, l6-level CPM with rate 3/4 codes, presents some simulations of eight- and l-level CPFSK with rate 2/3 and rate 3/4 codes. Viterbi detection is used throughout.

4.1 PSK Set Partition Codes

Conventional coding with BPSK and QPSK modulation is discussed in standard textbooks, Optimizing the Euclidean distance is in this case equivalent to optimizing the Hamming distance of the error correcting code. This is not the case with 8PSK or 16PSK modulation. Ungerboeck devised the so-called set partition codes for use with many-level PSK modulators. The idea behind these codes is to start with a large constellation of signal space points in the 1/ Q plane, and then restrict the signaling to a pattern of subsets of this. The choice of this subset is driven by the data symbols and until now the mechanism to do so have always been a Convolutional encoder circuit. The PSK may be either pure or Nyquist PSK. to illustrate the technique. We first give some examples of PSK constellations combined with rate 2/3 convolutuzional codes figure shows the 4PSK (QPSK) and 8PSK signal constellations. The 8PSK constellation is normalized with the rate 2/3 code rate to facilitate comparisons in normalized distance at equal E_b/N_o between uncoded QPSK and rate 2/3 coded 8PSK. Note that the asymptotic difference in E_h/ N_0 between uncoded 8PSK and uncoded 4PSK at high SNRs is

$$10\log_{10}\left(\frac{3d_0^2}{2d_4^2}\right)=10\log_{10}\left[\frac{3}{4}\left(2-\sqrt{2}\right)\right]=-3.57\ dB \tag{4.1}$$

a loss of 3.57 dB. For the uncoded case a Gray code representation of the eight different phase values should be used, as in Figure 4.1.b. For the coded case, both the Gray code and the natural binary code mapping rule have been proposed; Other points are discussed Figures 4.2-4.4 show three examples of coded 8PSK where the decoder has 4, 8, and 16 states, respectively. The natural binary mapper in Figure 4.1 is used. In the scheme of Figure 4.2, the most significant bit of the natural binary mapper word is left uncoded. Thus the free squared Euclidean distance can be no larger than the distance between two antipodal phase values in the signal constellation, i.e., 4. The minimum distance path in the trellis is also shown in Figure 4.2. Larger gains are obtained with the codes shown in Figures 4.3 and 4.4, where the most significant bit of the natural binary mapper is also coded. This is necessary to achieve larger

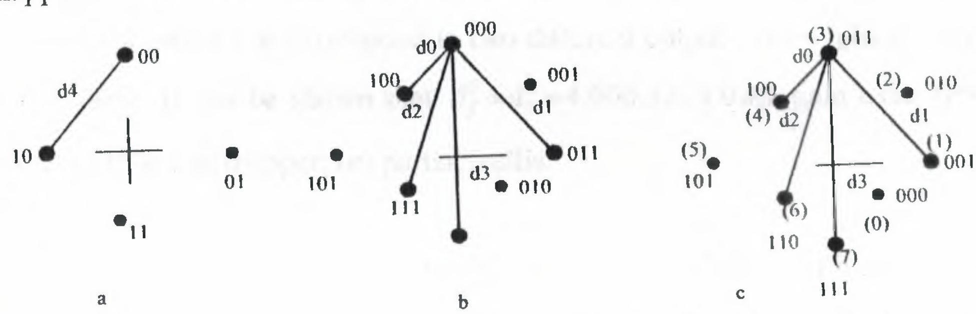


Figure 4.1 Signal constellations and distances for uncoded 4PSK (QPSK) and coded 8PSK with r ate 2/3 coding. All distances are normalized with respect to ²Eh as before. Two different mapping rules (binary representations of the signal points) are shown for 8PSK, viz., Gray mapping and Natural binary mapping. (a) 4PSK = QPSK. $d_4^2=2$. (b) 8PSK.Gray mapping. $d_0^2=2-\sqrt{2}, d_1^2=2, d_2^2=2+\sqrt{2}, d_3^2=4$. (c) 8PSK. Natural binary mapping. The octal number corresponding to each phase is also shown.

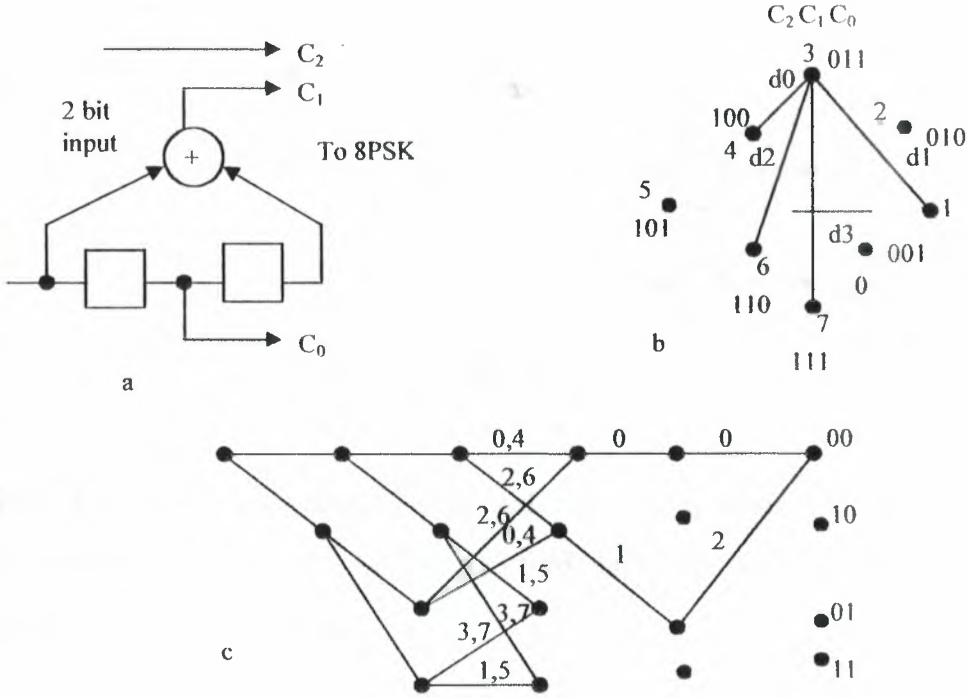


Figure 4.2 Example of a rate 2/3 code with 8PSK, four states. Note that the most significant bit c_2 of the natural binary word is left uncoded in this example. Therefore each transition in the trellis can correspond to two different output values (phase values) at the 8PSK modem. It can be shown that $d_f^2 = d_3^2 = 4.000, i.e., 3.0\,dB$ gain over 4PSK. (a) Encoder; (b) 8PSK and mapper; (c) partial trellis.

coding gains than 3 dB with the concept of 8PSK and rate 2/3 coding in Figure 4.2. For details of codes with high-SNR gains approaching 6 dB, see the tables in Refs. 8,28,7. Wilson, Schottler, and Sleeper list the results of a search for 16PSK codes with rate 3/4 coding. Set partition codes have also found application to high rate (4-6 bits/interval) QAM-style signalling in telephone-line modems, where gains of 3-6 dB have been observed. Here the QAM constellation has 32-128 points and the subset mechanism is in the manner of Figure 4.2. The number of data bits appearing per interval, b, is divided into groups of b_1 and b_2. The group of b_2 bits drives the choice of a sequence of subsets, in which each subset contains 2^{b_1} QAM points; the selection of one of these points carries the remaining b_1 bits.

These QAM codes of course incorporate amplitude modulation, so we will not pursue them further. The set partition idea has not explicitly been

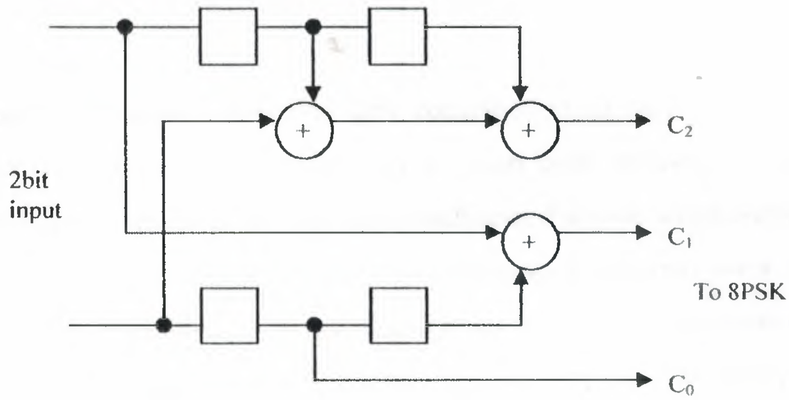


Figure 4.3 Example of a rate 2/3 code with eight states. Natural binary mapper. It can be shown that $d_f^2 = 2d_0^2 + d_0^2 = 4.584, i.e., 3.6dB$ gain over 4PSK. (a)Encoder; (b) 8PSK and mapper; (c) partial trellis.

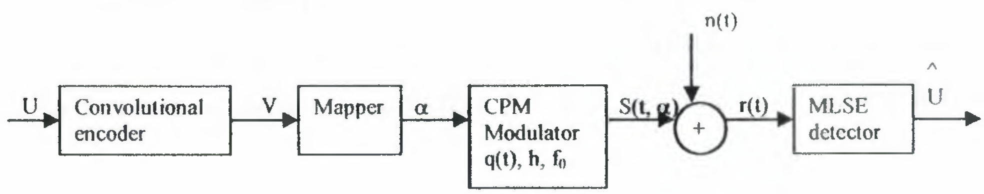


Figure 4.4 Example of a rate 2/3 code with 16 states. Natural binary mapper. It can be shown that for this code, $d_f^2 = 2d_0^2 + 2d_1^2 = 5.171$. This corresponds to a gain of 4.1 dB over 4PSK.applied to continuous phase schemes; we will now turn to coded CPM schemes.

4.2 Coded Binary and Quaternary CPFSK

Now we list good schemes that combine convolution codes and M-ary continuous phase FSK. Memory is introduced in the signal both by means of the convolutional code and by means of the continuous phase FSK modulator. A Viterbi detector is used that takes into account both types of memory. The advantage of CPFSK is its constant envelope and low spectral tails; by introducing convolutional coding, the error probability is improved. We will assume that the detector is coherent, that the rate of the convolutional encoder equals 1/2, and that the CPFSK has 2 or 4 levels. More general cases appear in Section 4.3 The convolutional code implies a bandwidth expansion by 2.

compared to the uncoded CPFSK, which strongly affects the energy-bandwidth trade-off of the combined scheme.

By optimal combination is meant that combination of rate 1/2 convolutional encoder and M level mapping rule which, for a given code memory v and a given modulation index h, gives the largest minimum Euclidean distance when combined with the CPFSK modulator. It is assumed that the observation interval over which the minimum distance is calculated is large enough so that the maximum obtainable minimum distance is reached. Minimum distance growth with the observation interval length will also be studied.

4.2.1 System Description

Let **u** be a sequence of independent binary (0, 1) data symbols. This sequence is encoded by a conventional convolutional encoder (rate k/n) whose output is a sequence of coded binary (0,1) symbols **v**; see Figure 4.5. It is convenient to use the following notation

$\mathbf{u} = (\ldots, u_{-2}, u_{-1}, u_0, u_1, u_2, \ldots) = (\ldots \mathbf{u}^{-2}, \mathbf{u}^{-1}, \mathbf{u}^{0}, \mathbf{u}^{1}, \mathbf{u}^{2}, \ldots)$ where $u_m \in \{0,1\}$, $\mathbf{u}^{i} = (u^i{}_1, u^i{}_2, u^i{}_3, \ldots\ldots, u^i{}_k)$

and $u^i{}_m = u_{ik+1+m}$. Similarly $\mathbf{v} = (\ldots, v_{-2}, v_{-1}, v_0, v_1, v_2, \ldots) = (\ldots, \mathbf{v}^{-2}, \mathbf{v}^{-1}, \mathbf{v}^{0}, \mathbf{v}^{1}, \mathbf{v}^{2}, \ldots)$ where $v_m \in \{0,1\}$, $\mathbf{v}^i = (v^i{}_1, v^i{}_2, v^i{}_3, \ldots v^i{}_n)$ and $v^i{}_m = v_{in-1+m}$. Thus **u** and **v** consist of subsequences $\mathbf{u}^i$ and $\mathbf{v}^i$ of length k and n, respectively. The encoding operation performed.

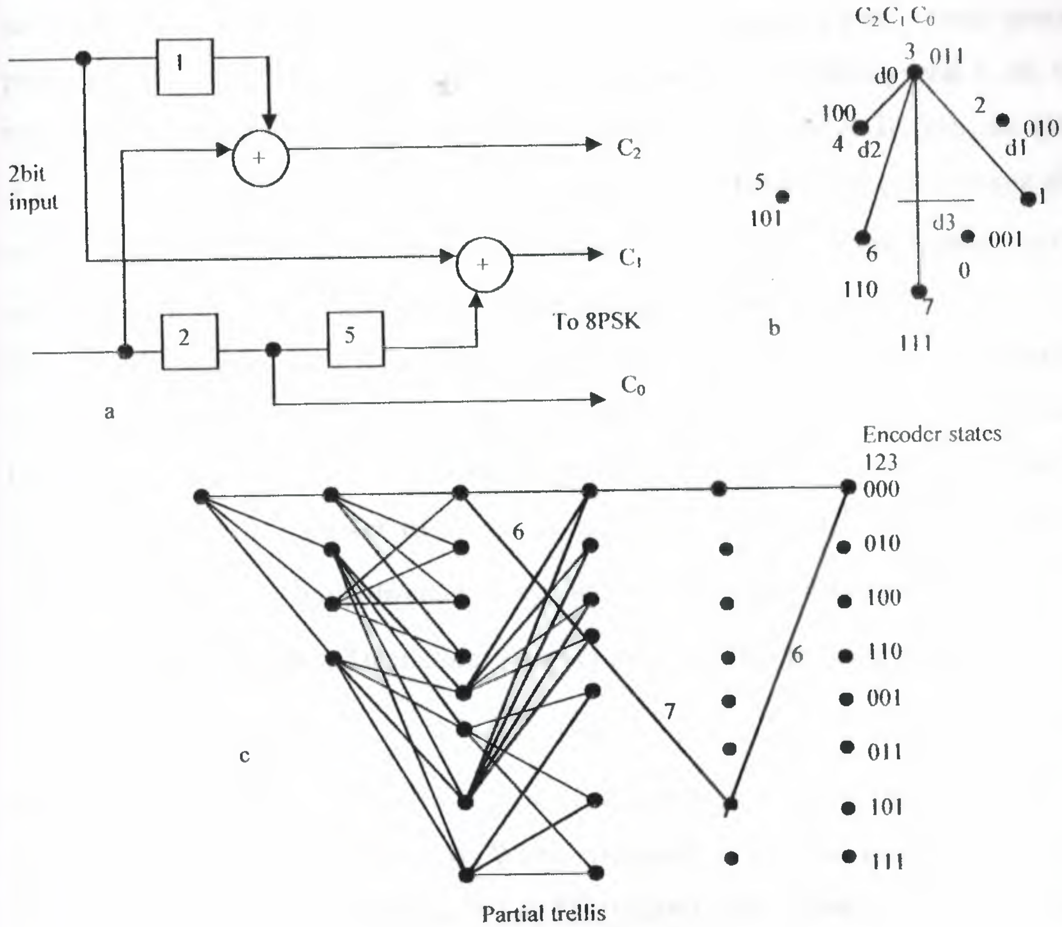


Figure 4.5 The Modulation Scheme, Channel, and Receiver.

by the convolutional encoder can now he written as

$$v_m^l = \sum_{j=1}^{k} \sum_{i=0}^{\mu} q_{mj}(i)\, u_j^{l-i} , \qquad l = \ldots,-2,-1,0,1,2,\ldots, \quad m = 1,2,\ldots, n \tag{4.2}$$

$g_{mj}(i) \in \{0,1\}$ And represents the connection between $u^{l-i}{}_j$ and $v^l{}_m$; the summations are modulo 2 and kμ is the maximum possible number of delay elements in the encoder. This encoder produces blocks of' coded symbols, $\mathbf{v}^l$, of length n each time blocks of data symbols, $\mathbf{u}^l$, of length k enter the encoder. The rate, $\mathbf{R}_c$ of the convolutional encoder is therefore $\mathbf{R}_c$=k/n data symbols/coded symbol. The number of delay elements in the encoder is v and the state of the encoder, $\mathbf{x}$, has v binary components. Another useful way to describe a convolutional encoder is by means of time "code generating polynomials" $G_{ij}(D)$. Using for tile moment the notation of Massey and S_a in, the D-transform of the k input sequences and n output sequences is I_1 (D)...Ik (D) and $T_1(D),\ldots,T_n(D)$. The input-output relationship of the convolutional encoder can then be written as $T(D)=\overline{G}(D)I(D)$, where $\overline{G}(D)$ is the n x k matrix whose th entry is $G_{ij}(D)$, and where T(D) and I(D) are vectors whose components are $T_J(D)$ and $I_J(D)$.

The coded sequence **v** is the input to a mapper which associates levels in an M-ary alphabet with blocks of coded symbols. The output from the mapper is a sequence $\alpha = (\ldots\alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1, \alpha_2,\ldots,)$ of channel symbols according to the mapping rule. It is assumed that **M** is a power of 2 and $\alpha_l \in \{\pm1,\pm3,\ldots\pm(M-1)\}$.The mapper has rate $R_m = \log_2(M)$ code symbols/channel symbol and the overall rate, R, is R=R $_C$R data symbols/channel symbol. The natural M-level mapping rule is defined by

$$\alpha_l = \sum_{i=0}^{s-1} v_{ls+1} \bullet 2^{(2-i)} - M + 1, \qquad 1 = 0, \pm1, \pm2, \ldots \tag{4.3}$$

With s $=\mathbf{R}_m$ The symbols v_{ls} and $v_{(l+1)s-1}$ are referred to as the most significant bit (MSB) and the least significant bit (LSB). When Al equals 2 or 4 the corresponding natural mapping rule will be referred to as B1 or Q1, respectively. Table 4.1 gives the mappers for quaternary schemes considered in Section 4.2.

With the notations introduced above the transmitted energy per information hit, E_b, and the information bit symbol time, T_b, can he written as E_b = E/R and T_b=T/R . The signal available for observation is r(t) =s (t, α) + n(t) where n(t) is a Gaussian random process having zero mean and one-sided power spectral density N_0.A coherent detector which minimizes the probability of erroneous decisions must observe the received signal .

Table 4.1 Four-Level Mapping Rules

Input coded Symbol to The mapper 1st (MSB)	2nd (LSII)	Q1	Q11	Q12	Q13	Q14	Q15	Q2	Q21	Q22	Q23	Q24	Q25
0	0	-3	-3	-3	-3	-3	-3	-1	-1	-1	-1	-1	-1
0	1	-1	-1	-1	1	3	3	-3	-3	1	1	3	3
1	0	-1	3	3	-1	-1	1	1	3	3	-3	-3	1
1	1	1	-1	3	1	-1	3	1	-3	3	1	1	-3

(output channel symbol)

r(t) over the entire time axis and choose the infinitely long sequence ii which minimizes time error probability. This is referred to as maximum likelihood sequence estimation (MLSE).

4.2.2 Error Probability amid Minimum Euclidean Distance

We can over bound P(ε) by the union bound

$$P(\varepsilon)=\frac{1}{S}\sum_{i=0}^{S-1}P(\varepsilon/s_i)\le\frac{1}{S}\sum_{i=0}^{S-1}\sum_{\substack{j=0\\(j\neq i)}}^{S-1}Q\left(\left(d_{ij}^{2}\frac{E_b}{N_0}\right)^{\frac{1}{2}}\right) \qquad (4.4)$$

Where S is time total number of signal alternatives $S_0,S_1,,,,,,,S_{S-1}$. d_{ij}^2 is the squared Euclidean distance (ED) in signal space between S_i and S_j normalized by $2E_b$. When E_b/N_0 is large the union hound becomes increasingly tight and

$$P(\varepsilon)\approx Q\left(\left(d^2{}_{\min}\frac{E_b}{N_0}\right)^{\frac{1}{2}}\right) \qquad (4.5)$$

In our present coding context, $d^2_{\min}$ will be referred to as the normalized squared free Euclidean distance (NSFED)

$$d^2{}_{\min}=\min\frac{1}{2E_b}\int_0^{\infty}\left[s(t,\alpha)-s(t,\beta)\right]^2 dt \qquad (4.6)$$

Where α and β are the output sequences from the mapper when the input data sequences to the convolutional encoder are $u_\alpha = (\ldots.\ 0, 0, 0, u^{-\mu}, u^{-\mu+1}, \ldots., u^{-1}, u^{1}, u^{2} \ldots..)$, respectively. The minimization in (4.6) is carried out for all possible $u^0 \neq u_B{}^0$

We seek modulation schemes which have large d^2_{min} .For comparison, $d^2_{min} = 2$ for uncoded minimum shift keying (MSK), uncoded binary phase shift keying (BPSK), or uncoded quaternary phase shift keying (QPSK). A more general parameter of interest is the minimum normalized squared Euclidean distance (MNSED) $d^2_{min.N}$ defined by

$$d^2{}_{min.N} = \min \frac{1}{2E_b} \int_0^{NT} [s(t,\alpha) - s(t,\beta)]^2\, dt \quad (MNSED) \qquad (11.7)$$

with u_α and u_β the same as in (4.6). $d^2_{min.N}$ is the minimum distance between signals defined up to time NT It is a no decreasing function of N with maximum value equal to d^2_{min} It is interesting to know for which value of N the free Euclidean distance is reached, i.e., the smallest N for which $d^2_{min} = d^2_{min.N}$ because the path memory of the MLSE detector at large E_b/N_0 is closely related to this, It is also of interest to know how fast $d^2_{min.N}$ grows with N.

An example of a coded modulation scheme of the present type is given in Figure 4.6. The rate 1/2 code is defined by means of an octal representation of the upper and lower connection polynomials. This notation will be used throughout this chapter. The code in Figure 4.6 is denoted (13, 4), where 13 and 4 are octal representations of the two polynomials. We are looking for those combinations of a no catastrophic rate 1/2 convolutional encoder [G(D)] and mapping rule such that, for fixed v and h, d^2_{min} is maximized. Formally, this can be written as

$$\max \qquad d^2_{min} \qquad (4.8)$$

4.2.3 Minimum Euclidean Distance: Properties and Calculation

It is always possible to find two data symbol sequences u,, and u_β in (4.6) such that s(t, α) = s(t, β) for I > N_mT where N, is some positive integer; this must yield an upper bound, $d^2_{min,}(h)$, on $d^2_{min,,}(h)$. This bound is the normalized squared ED in signal space between s(t, α) and s(t, β),

$$d^2{}_B(h) = \frac{1}{2E_b} \int_0^{NT} [s(t,\alpha) - s(t,\beta)]^2\, dt \qquad (4.9)$$

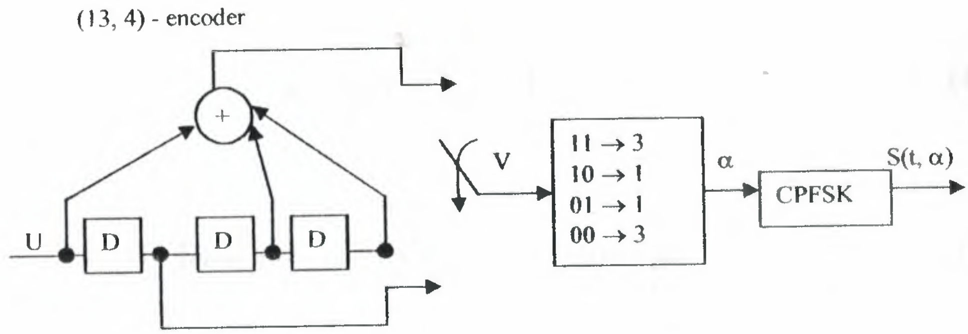


Figure 4.6. An example of a modulation scheme considered in this are the octal representation of the upper and lower polynomial, respectively. The mapping rule used in this example is mapping rule Q1; see Table 4.1.

Clearly, $d^2_{min.N}$ (h) < $d^2_{min,,}$(h) < d^2_B (h) and the bound is tight [d^2_{min} (h)= d^2_B(h)] for properly chosen sequences u_α and u_β. The situation described above is an example of a phase and state merger. The conditions for a phase and state merger can be formulated as

$$\sum_{i=0}^{N-1} Yi = \frac{2l}{h} \textit{ for some } l, l = 0. \pm 1, \pm 2, \ldots\ldots \quad (4.10)$$

where y, = α-β, is the ith element in the difference sequence y = α-β= (...,y_{-2},y_1,y_0,y,,y_2,...) and y_i € {0,±2,±4,...,±2(M—1)}. If l=0 in (4.10) a phase and state merger has occurred independent of h while if l ≠ 0 a phase and state merger has occurred only for specific h values according to (4.10). This is important because in the former case an upper bound can be calculated for all h, while in the latter case an upper bound can only be calculated for those specific h values given by (4.10). Such specific h values are referred to as weak modulation indices. Figure 4.7 shows examples of phase and state mergers for two specific coded binary CPFSK schemes. Calculations show that the phase and state merger given in Figure 4.7a is a minimum ED merger when the modulation index is small. It is also found that the phase and state merger (all h) given in Figure 4 .7b is a minimum ED merger for almost all modulation indices in the interval 0 < h < 0.5; furthermore, for this h interval, there does not exist any shorter minimum ED merger than those two given in Figure 4.7. Note that the phase and state merger (all h) in Figure 4.7b has length 14T, but there are only 10 T intervals giving any distance contribution. Assuming $f_0 >> 1/T$ the expression (4.7) for d^2_{min} (h) can be reduced to

$$d^2_{min,N}(h) = \min R\left(N - \sum_{m=0}^{n-1} C_m\right) \quad (4.11)$$

with

$$C_m = \int_0^1 \cos\left[\pi h \xi_m + 2\pi h \sum_{i=0}^{L-1} Y_{m-iq}((x-i)T)\right] dx \quad (4.12)$$

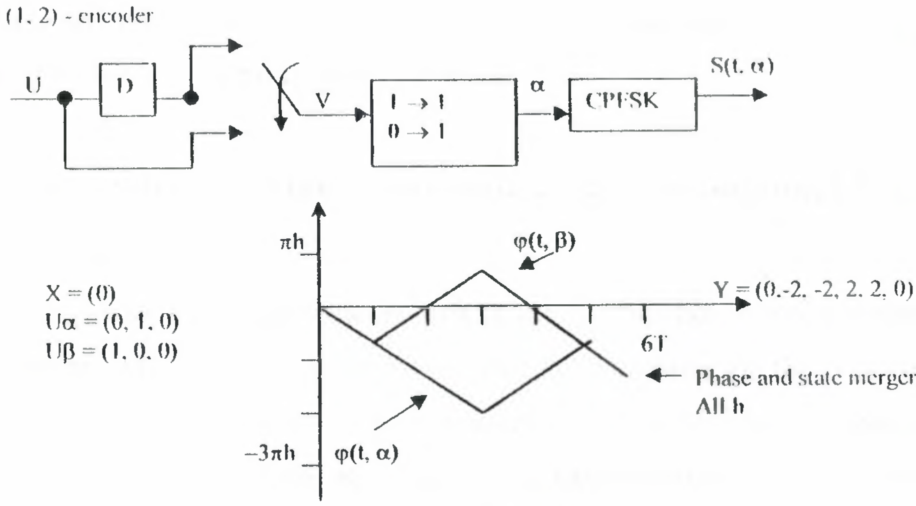


Figure 4.7 Examples of phase and state mergers for binary CPFSK with the (1,2)-encoder.

Since $y_i = 0$ for all i <-1<£ 0 for all m < L -1. When g (t) is LREC [g (t) 1/2LT, 0 < t < LT, further simplification is possible since C_m is

$$C_m = \begin{cases} \dfrac{\sin(y_m + z_m) - \sin(y_m)}{z_m} & z_m \neq 0 \\ \cos(y_m), \quad z_m = 0 \end{cases}$$

with

$$z_m = \frac{\pi h}{L} \sum_{i=0}^{L-1} Y_m - i$$

and

$$Y_m = \pi h\left(\xi_m + \frac{1}{L} \sum_{i=0}^{l-1} i y_{m-i}\right) \quad (4.13)$$

The algorithm used for the calculation of $d^2_{min.N}$ (h) is a natural extension of the algorithm described in Appendix A. The modulation scheme in Figure 4.5 is called catastrophic if there exist two data symbol sequences u_α, and u_β with infinite Hamming distance such that the corresponding ED in signal space between s (t, a) and s (t, β) is finite. This type of scheme must of course be avoided. It is quite easy to show that if the convolutional encoder is no catastrophic, then the modulation scheme is also non-catastrophic for all rational h values.

4.2.4 Symmetries for the Combination of Convolutional Encoder and Mapping Rule

In this section we shall take a closer look at the combination of convolutional encoder and mapping rule. The reason for this is that in the search for the best combinations [largest d^2_{min} (h)] we would like to minimize the number of combinations to be investigated. It is assumed that R_c =1/n, (k = I). Let us first state four conditions for the convolutional encoder G (D). It is also assumed that (1) The convolutional encoder is no catastrophic; (2) $g_{j,1}$ (0) 1 for some j, 1 < j < n; (3) $g_{j1}(v)$ = 1 for some], 1 < j < n; and (4) $g_{j1}(i)$ = 1 for some i, 0 s< i < v, and all j, 1 <j < n. The total only M 1/2 mapping rules need to be investigated. Also, time reversal of phase and state mergers does not change d^2_{min}.

Rate 1/n Convolutional Encoder and Binary Mapping Rule

First we reduce the number of rate 1/n convolutional encoders in the search when M = 2. In this binary case there is only one mapping rule of interest, namely, mapping rule B1. Consider now two convolutional encoders G (D) and G' (D), where

$$g_{j,1}(i) = g_{(n+1-j),1}(v-i), \qquad 1 \le j \le n, \quad 0 \le i \le v \tag{4.14}$$

G(D) and G'(D), together with mapping rule B1 will have the same d^2_{min} (h) but not the same $d^2_{min.N}$ (h). This is now shown.

Let u and u', defined as u = (...,0,0,u_0,u_1,. . ., u_m, , 0, 0,...) and u' = (.... , 0, 0, u'_0, u'_1,...., u'_m,, 0, 0, .. .), be the input data symbol sequences to the convolutional encoders G(D) and G'(D), respectively, and let the corresponding output sequences be t_j and t'_j, j = 1, 2 n. If $g'_{j,1}$ (i) = $g_{j,1}(i) = g_{(n+1-j),1}(v-i), 1 \le j \le n, 0 \le i \le$,and u'_1 =u'_{m-1} then the ith element in t, is $t'_{j,1}$= $t_{(n+v-i)}$. Consequently, after the binary mapping rule, the channel symbol sequence a' is only a time reversed delayed variant of a. Those pairs of data symbol sequences, α, and up, which give d^2_{min} always, be written as u and u'. Suppose now that the convolutional encoder G (D) is used and u_α. And Up causes a state and phase merger. Then, as was shown above, it is always possible to create a new convolutional encoder G' (D), according to (4.13), which together with the two input data symbol sequences u'_α ($u\alpha'_{,1} = u_{\alpha(m-1)}$) and u_β($u_{\beta,1} = u_{\beta,(m-1)}$) produces a state and phase merger which is just a time reversed delayed variant of the phase and state merger created by uα, up and G(D). Since these two state and phase mergers have the same ED it is clear that G (D) and G'(D) are equivalent from the d^2_{min} (h) point of view. However, G (D) and G'(D) have different $d^2_{min.N}$ (h).

Rate 1/n Convolutional Encoder and 2^n level Mapping Rule

Consider two convolutional encoders G (D) and G'(D), where

$$g_{j,1}(i) = g_{j,1}(v-i), \qquad 1 \le j \le n, \quad 0 \le i \le v \tag{4.15}$$

G(D) and G'(D), together with a 2^n level mapping rule, will have the same d^2_{min} (h) but not the same $d^2_{min.N}$ (h). The arguments and notations used in and let the corresponding output sequences be t_j = t'_j = 1,2….,n., then the ith element in $t'_{j,1}$is $t'_{j,1} = t_{j,(m+v-1)}$• Consequently, after any 2" level mapping rule, the channel symbol sequence a' is only a time reversed delayed variant of a. Using the same argument as in the earlier proof, it is clear that G(D) and G'(D) are equivalent from the d^2_{min} (h) point of view, but do not have the same d^2_{min} (h).

We now look at the specific case when a rate 1/2 convolutional encoder is combined with a four-level mapping rule. Consider the mapping rules Q1 and Q2 (defined in Table 4.1). Any combination of a rate 1/2 convolutional encoder and a four-level mapping rule can be transformed to another combination of a rate 1/2 encoder and a four-level rule, where the rule is either Q1 or Q2. Furthermore, the transformed combination will be identical to the original combination. Consequently, only mapping rules Q1 and Q2 are needed in the search. Another useful property is that the combinations {G (D), Q2} and {G (D), Q11} (Q11 defined in Table 4.1) are equivalent from the d^2_{min} (h) point of view if G(D) has an odd number of connections in its lower polynomial. This fact reduces the number of convolutional encoders needed in the search when Q2 is used, since the combination {G (D), Q11} can be transformed to an identical combination using mapping rule Q1.

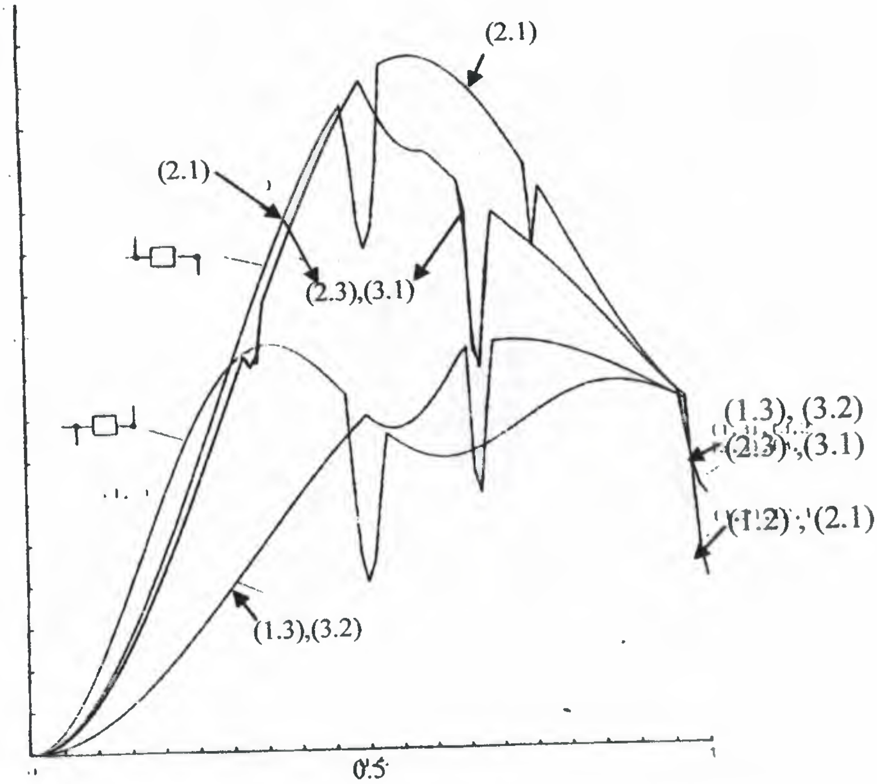


Figure 4.8. $d^2_{min.40}$ (h) as a Function of h for all Noncatastiophic v = 1 Rate ½ Convolutional Encoders. Binary CPFS K Modulation is used.

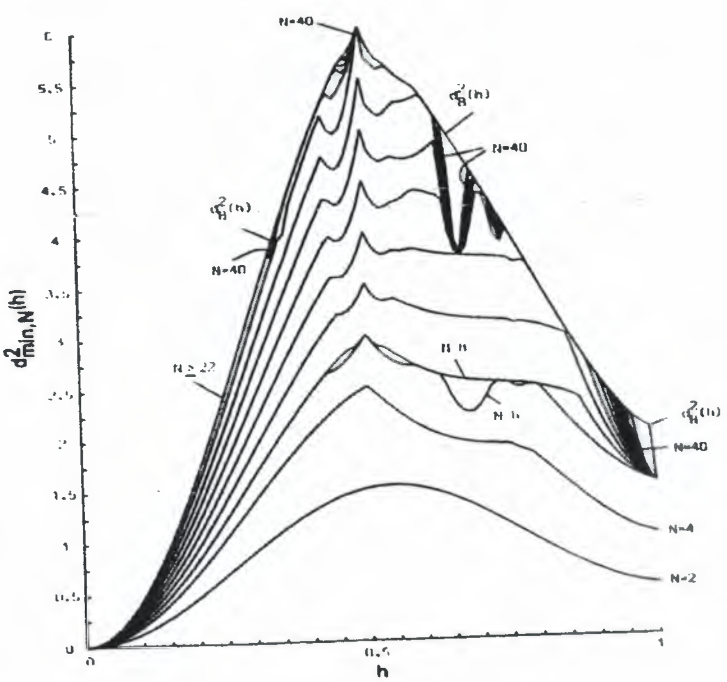


Figure 4.9. $d^2_{min.N}$ (h) as a Function of Jm with **N** as a Parameter when the Rate 1/2 (4, 3)-convolutional Encoder with Binary CPFSK Modulation is used.

suboptimal from the d^2_{min} (h) point of view. Figure 4.11 shows $d^2_{min\ 40}(h)$ for all 24 v = 2 noncatastiophic convolutional encoders and the optimality of the (4, 3)-encoder in the interval 0 < h < 0.75 is clearly demonstrated. When v = 3 there is no significant increase in the optimum d^2_{min} (h) compared to the v = 2 case when 0 < h < 1/2. However, for 1/2 < h < 1 the increase is significant. As an example, when h = 1/2 the optimum d^2_{min} (1/2) = 6, which is the same as in the v = 2 case. The (12, 7)-encoder

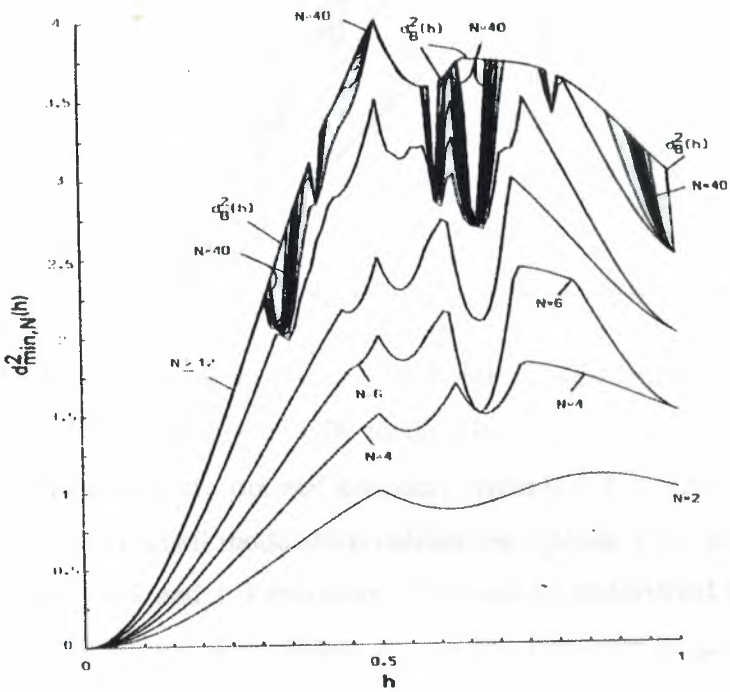


Figure 4.10. d^2_{min} (h) as a function of h with **N** as a parameter when the rate 1/2 (7, 5)-convolutional encoder with binary CPFSK modulation is used.

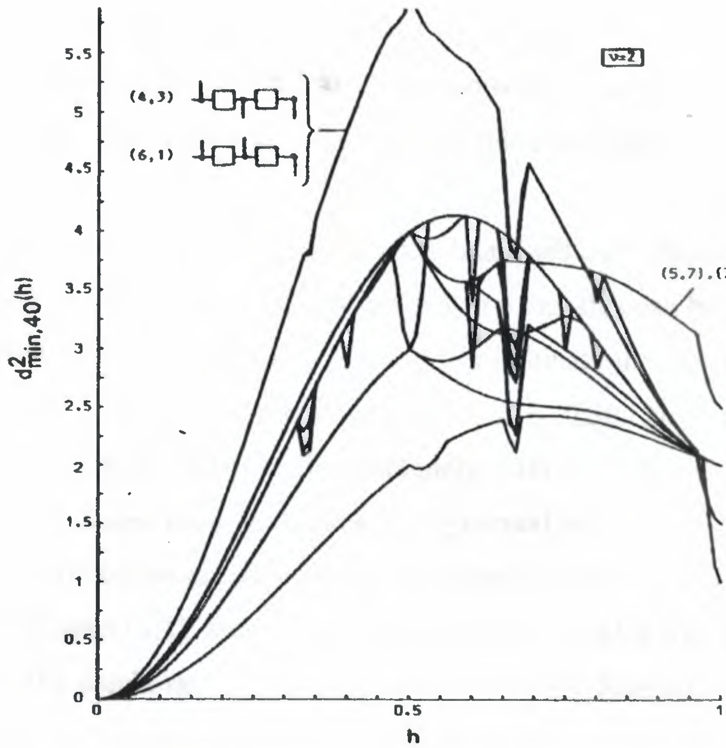


Figure 4.11. $d^2_{min\ 40}(h)$ as a function of h for all noncatastiophic v = 2 rate 1/2 convolutional encoders. Binary CPFSK modulation is used.

And the (10, 3)-encoder are optimal encoders when 0 < h < 0.50 and 0.5 < h < 0.70, respectively. For very small modulation indices the optimal v = 1 encoder is better than both the optimal v = 2 and v 3 encoders. (This can be understood by studying the two mergers (all h) in Figure 4.7.) When v = 4 the optimum d^2_{min} (h) is significantly increased compared to schemes with smaller v values. For example, when h = 1/2, the optimum

Rate 1/2 Convolutional Encoder and Four-Level CPFSK

In this case the search is over all noncatastiophic convolutional encoders and all four-level mapping rules. The modulation indices h = 0.05, 0.10,..., 0.55 and h = 0.60 are considered and 1 < v < 4. The optimum d^2_{min} (h) for different ii and h are given in 85eTables D.3-D.5 in Appendix D together with the corresponding optimal combinations of encoder and mapping rule, and the minimum N for which $d^2_{min.N}$ (h) = d^2_{min}(h)

When v = 1 the optimum d^2_{min} (h) is significantly increased compared to the uncoded four-level case when 0 < h < 0.35. Furthermore, when h = 1/2 the optimum d^2_{min} (1/2) =3 (the corresponding uncoded value is 2), and this is reached when the {(1, 2), Q1}, {(3, 1), Q1}, or {(1, 3), Q2} combinations are used. However, there also exist modulation indices for which coding decreases the optimum d^2_{min} (h) compared to the uncoded case. Note that when $0 < h < 0.45$ no combination using mapping rule Q2 is optimal and when h = 0.55 and h = 0.60 no combination using mapping rule Ql is optimal.

When v = 2 the optimum d^2_{min} (h) is significantly increased over the v = 1 case. When h = 1/2, the optimum d^2_{min} (1/2) is increased from 3 to 5 and this can be reached with the {(1, 7), Q1} combination or with the {(1, 6), Q2} combination. The combinations {(7, 2), Q1} and {(7, 5), Q2} are optimal when $0 < h < 0.30$. Note that whenever a combination {G (D), Ql 1} is optimal the combination {G (D), Q2} is also optimal even when the number of connections in the lower polynomial is even. Calculations indicate that these two combinations are almost (but not exactly) identical from the $d^2_{min.N}$ (h) point of view.tt[91] Figure 11.12 shows $d^2_{min.N}$ (h) as a function of h with N as a parameter for the {(7, 2), Q1} combination. To compare the uncoded four-level CPFSK scheme with the coded v = 2 schemes Figure 11.13 is included. This figure shows the right upper bounds on d^2_{min} (h) for some optimal v = 2 combinations. A significant increase in d^2_{min} (h) can be achieved with coding for almost all modulation indices in the interval $0 < h < 1$. On the other hand, the corresponding N values are significantly larger.

When v = 3 the optimum d^2_{min} (h) again increases compared to the = 2 case. As an example, when h = 1/2 the optimum d^2_{min} (h) is increased from 5 to 6 and this can be reached with the {(1, 15), Q1} combination or the {(1, 14), Q2} combination. When $0 < h < 0.25$ the combinations {(13, 4), Q1} and {(13, 17), Q2} are optimal, and the {(3, 10), Q1} combination is optimal in $0.30 < h < 0.40$. Just as in the v = 2 case, whenever a combination {G(D), Q11} is optimal, the combination {G(D), Q2} is also optimal, even when the number of connections in the lower polynomial is even.

For v = 4 a complete search has only been done for h = 1/4. The optimum d^2_{min} (1/4) = 6.151 and the corresponding optimal combinations

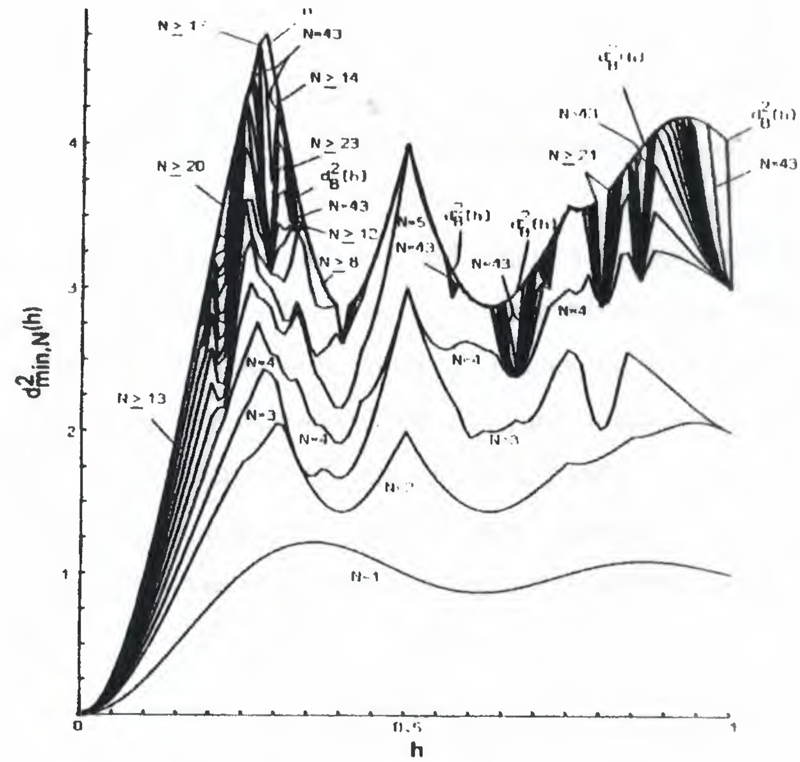


Figure 4.12 d^2_{min} (h) as a function of h with N as a parameter when the rate 1/2 (7, 2)-convolutional encoder with four-level CPFSK modulation is used (mapping rule Q1).

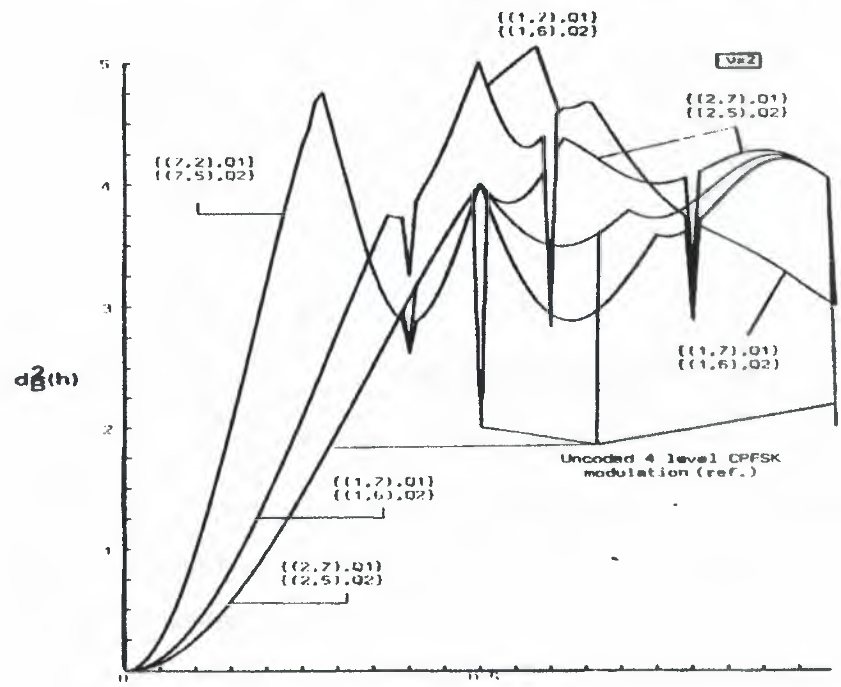


Figure 4.13 Upper bounds on d^2_{min} (h) for some combinations of a rate 1/2 convolutional encoder and a four-level mapping rule. CPFSK modulation is used. These upper bounds are tight for all modulation indices of practical interest.

are {(23, 10), Q1} and {(23, .33), Q2}. For other modulation indices, combinations have been found which significantly increase d^2_{min} (h) compared to the v = 3 case. As an example, when h = 1/2 the {(4, 23), Ql} combination achieves d^2_{min} (1/2) = 7.

Just as the error probability for the combined schemes is dominated by the term Q $((d^2_{min} E_b /N_0)^{1/2})$, the probability for a rate 1/2 code combined with BPSK or QPSK is $Q((d_f E_b / N_o)^{1/2})$, where df is the free Hamming distance. From the coding literature we can construct Table 4.2, which compares the d^2_{min} and d_f from these two approaches. The table shows that the coded CPFSK approach is as good as or even a little better than the coded QPSK method and it can achieve better spectral tails for a fixed signal envelope than can QPSK. The most important fact is not depicted

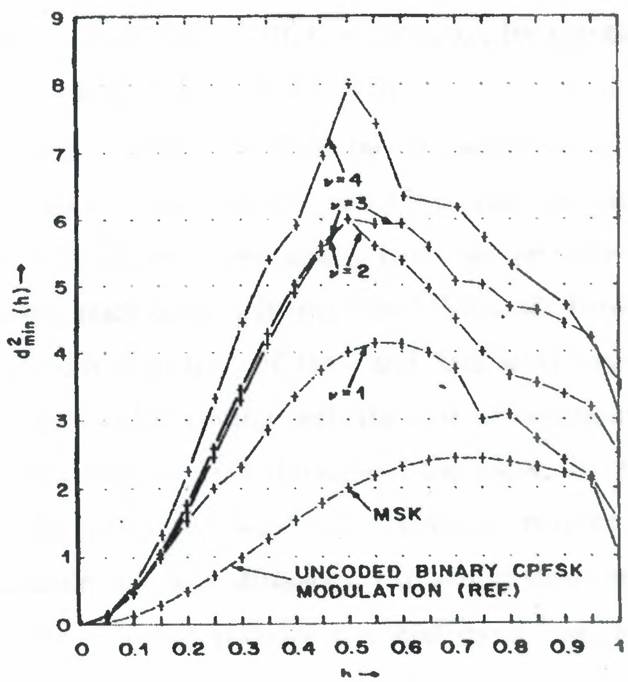


Figure 4.14 Rate 1/2 convolutional encoder and binary CPFSK modulation. The optimum d^2_{min} (h) for different ~ and Ii. See also Tables D.1-D.2. Note that the connection lines do not contain any numerical results.

Table 4.2 Comparison between the Optimal Rate

v	d_f best code	$d^2_{min.}$ (1/2): Optimal combination M = 2	M = 4
1	3	4	3
2	5	6	5
3	6	6	6
4	7	8	7

4.2.6 Energy-Bandwidth Tradeoff

The considerable distance gains of coded CPFSK are bought with an increase in bandwidth over uncoded CPFSK. Denote t_y $2BT_b$ the double-sideband positive-frequency bandwidth of uncoded CPFSK, just as was done. With the introduction of coding, this increases to approximately $2B_cT_b = 2BT_b/R_C$. By comparison, Figure 4.18 gives an energy-bandwidth plot for coded CPFSK in the style of Section 5.1; d^2_{min} relative to MSK is plotted against the 99% power bandwidth $2B_CT_b$. (No separate measurement of the coded spectra was performed.) It is clear that the energy-bandwidth performance of coded CPFSK becomes rapidly better as the convolutional constraint length increases, and that much better schemes than MSK exist. However, a comparison with the energy-bandwidth trajectory of two- and four-level uncoded CPFSK, also plotted in the figure, shows that coding with these short constraint lengths does not really lead to an overall improvement throughout the plane. In the coded two-level schemes particularly, the factor-of-2 bandwidth expansion is quite damaging. But the coding is effective at achieving large distances with simple hardware, and we shall see in the next section that a higher encoder rate and more FSK levels will improve performance.

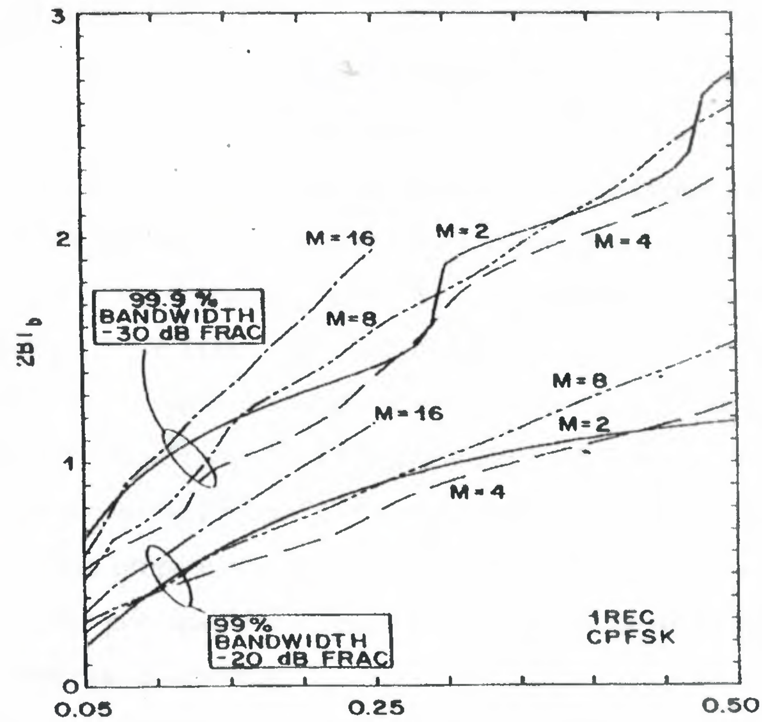


Figure 4.15 The normalized bandwidth $2BT_b$ as a function of the modulation index Ii when M level CPFSK modulation is used, M = 2, 4, 5, and 16. The levels in the fractional out of band power function used in the definitions of bandwidth are -20 dB (99% power within the band) and -30dB (99.9%).

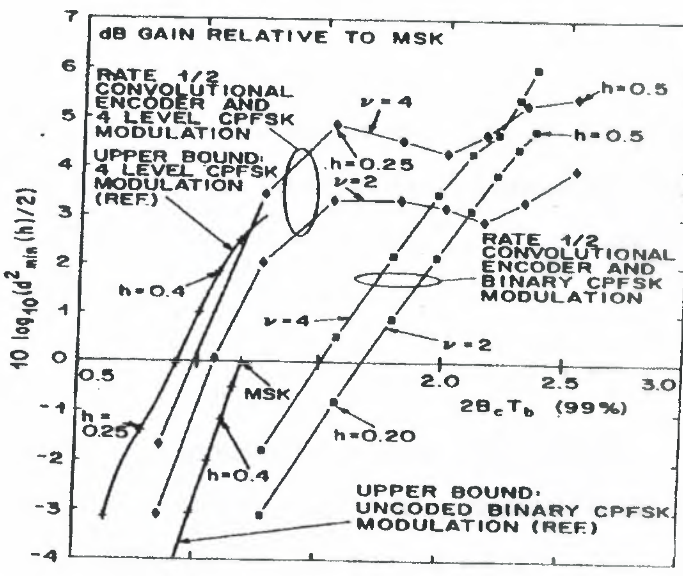


Figure 4.16 Power-bandwidth trade-off for some of the coded CPFSK schemes considered in Section 11.2. Note that the connection lines do not contain any numerical results.

4.3 Coded Multilevel CPM

In this section we will increase the encoder rate and the alphabet size of the FSK in an attempt to find coded schemes with a more attractive energy-bandwidth performance. Following Refs. 3 and 4, we will typically deal with eight- and higher level CPFSK. The encoder rate with eight-level schemes will be 2/3, that is, two information bits per eight-level FSK symbol. We will also give a method generalizing the eight-level results to 16-level CPM with rate 3/4 codes and 32-level CPM with rate 4/5 codes. All of these schemes combine energy and bandwidth efficiency with constant envelope. Significant improvements are obtained over the CPM and multi-h schemes of. The price for this is increased complexity in the FSK and the complexity in the encoder.

We will also propose a way to extend the CPFSK method to smoothed partial response CPM. The principal idea is to use the optimal encoders for CPFSK in a CPM scheme with weak partial response.

Lists of good encoder/modulator combinations appear in Appendix D.

4.3.1 System Description

The general system description in Section 4.2.1 and Figure 4.5 still holds, but in this section we study a restricted class of high rate convolutional encoders. These have k = n - 1, n = 3, 4, 5, 6. Furthermore, the first (n - 2) data symbols in each block are not coded at all.

$$v_m' = u_m' \qquad m = 1,2,\ldots\ldots,n-2, \quad \text{all } l \tag{4.16}$$

The only coding used is a conventional (in general nonsystematic) rate 1/2 convolutional encoder which encodes the data symbol sequence $\ldots u_k^{-1}$, u_k^0, u_k^1,... to produce the two coded symbol sequences ..., v^{-1}_{n-1}, v^{0}_{n-1},... and ..., v^{-1}_{n-1}, v^{0}_{n}, v^{1}_{n} Formally this can be written as

$$v'_m = \sum_{i=0}^{v} g_{m,k}(i) u_k^{l-i}, \quad l = \ldots,-1,0,1,\ldots, \quad m = n-1, n \tag{4.17}$$

where $g_{m,k}$ (i) Є {0, 1} and represents the connection between u^{l-i}_k and v_m,. The summation is modulo 2 and v is the number of delay elements in the encoder. The state x of the encoder has v binary components and there are 2^v states. Since the encoder described above has (n - 1) inputs and n outputs, the rate R_c of the encoder is (n -1)/n data symbols/coded symbol. As before, the coded sequence v is the input to a mapper which associates levels in an M-ary alphabet with blocks of coded symbols. The output from the mapper is α sequence a of channel symbols α Є {±1, ±3, ... , ±(M - 1)} all 1, where M is a power of 2. The mapper has rate R_m = $\log_2$ (M) coded symbols/channel symbol, which is fixed at n. The overall rate R is therefore R = R_cR_m ,= (n - 1) binary data symbols/channel symbol. The (natural) M level mapping rule used is defined by

$$\alpha_1 = \sum_{i=0}^{n-1} v_{\ln+i} . 2^{(n-i)} - M + 1, \qquad l = 0, \pm 1, \pm 2, ... \tag{4.18}$$

The symbols $v_{\ln}$ and $v_{(j+t)\ n-1}$ are referred to as the most significant bit (MSB) and the least significant bit (LSB), respectively. The channel encoding in this section can be viewed as a conventional rate 1/2 convolutional encoder of the least significant symbol in the mapping rule. For further details.

In this section it is assumed that the frequency pulse of the CPM modulator is given by

$$g(t) = \begin{cases} [1 - \cos(\pi t / \varepsilon T)] / 4T & 0 < t < \varepsilon T \\ 1/2T, & \varepsilon T < t > T \\ \{1 + \cos[\pi (t - T) / \varepsilon T]\} / 4T, & T < t < (1 + \varepsilon) T \\ 0, & otherwise \end{cases} \tag{11.19}$$

With 0 < ε < 1. When ε = 0 and e I we have the CPFSK scheme and the 2RC scheme, respectively. Thus (4.17) defines a class of frequency pulses in between CPFSK and 2RC, and a certain degree of smoothness and partial response can be controlled by the parameter e.

For CPFSK modulation (ε = 0) we will search for the optimal rate 1/2 noncatastiophic convolutional encoders such that, for given v and h, d^2_{min} is maximized. Formally this can be written as (4.8). Some of the optimal encoders are then used together with a frequency pulse with a small ε, ε > 0. The idea is to create a modulation scheme which has roughly the same d^2_{min} as the corresponding coded CPFSK scheme but with considerably smaller side lobes in the spectrum because of the smoothing and partial response.

4.3.2 Bounds on the Free Euclidean Distance

As an example of the upper bound defined in Section 4.2 consider the modulation scheme in

2/3 (5, 2)-encoder combined with eight-level CPFSK modulation. Now assume that the start state of the encoder is x = (0, 0). If the two sequences u_α = (01, 10, 00, 01) and t_β = (10, 01, 10, 01) are fed into the encoder the corresponding phase functions are φ (t, α) and φ (t, β). These phase functions are linearly increasing or linearly decreasing depending on the corresponding channel symbols; see the example in Figure 4.20. It is easy to see that the two data symbol sequences u_{α}. and U_{β} given above generate a phase and state merger after four channel symbol intervals. Thus an upper bound on d^2_{min} (h) for the specific coded CPFSK scheme given in Figure 4.19 can be calculated from (4.9), to get the upper bound

$$d^2_{\min}(h) \leq 2\left(4 - \frac{2\sin(2\pi h)}{2\pi h} - \frac{10}{3}\frac{\sin(4\pi h)}{4\pi h}\right) \tag{4.20}$$

Let us now employ a frequency pulse g(t) with the parameter $\varepsilon \neq 0$. Using the sequences u_{α} and u_{β} we again generate phase and state mergers for these cases. We add a common channel symbol at the end of the two channel symbol sequences u_{α} and u_{β} generated by $u\alpha$ and u_{β} and because of the partial response we must also specify a common start-symbol. Figure 4.20 shows the corresponding phase functions φ (t, α) and φ (t, β) when $\varepsilon = 1/2$ and $\varepsilon = 1$; the start symbol is -7 and the added channel symbol is 3. Note that the phase' functions are shown with different delays for different values of the parameter e. This makes Figure 4.20 easier to read. It is seen that a phase and state merger occurs after five channel symbol intervals, and an upper bound on d^2_{min} (h) for the specific coded CPM scheme ($0 < \varepsilon < 1$) given in Figure 4.19 can be calculated from this merger. By increasing the parameter e the phase function φ (t, α) becomes more smooth, and it is this smoothing that reduces the sidelobes in the spectrum. it was shown that the Euclidean distance between any two transmitted signals φ (t, α) and φ (t, β) is a function only of the difference sequence y. By making the k_{th} encoder input 0 (i.e., the second in Figure 4.19), it is possible to generate phase and state mergers corresponding to the difference sequences y = (...., 0, 8 l, -8 l, 0,...), l = 1, 2,..., 2^{n-2} -1 These phase and state mergers are independent of the actual rate 1/2 convolutional encoder used, and thus encoder-independent upper bounds on d^2_{min} can be calculated from them. Taking the minimum of these (2^{n-2} -1)

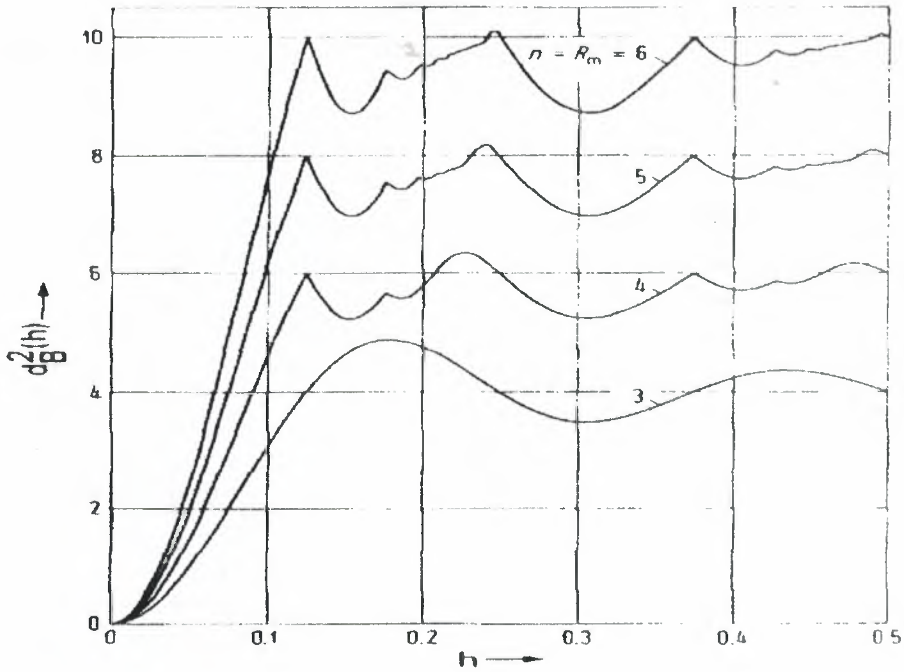


Figure 4.17 Encoder-independent upper bounds on equation (4.21). Coded eight-level to 64-level CPFSK modulation is assumed.

Upper bounds and assuming CPFSK modulation yields the encoder-independent upper bound

$$d^2_{\min}(h) \leq 2(n-1)\min\left(1 - \frac{\sin(18\pi h)}{18\pi h}\right) \tag{4.21}$$

Where A = 2^{n-2} -1. Figure 4.21 shows this upper bound with n as a parameter. Two questions arise: Are the upper bounds tight, and if they are tight, which rate (n - 1)/n convolutional encoders together with 2^{-n} level CPFSK modulation reach this bound? In particular we are interested in encoders with small v values. As we increase n the number of encoder-independent weak modulation indices is also increased, owing to the distance properties of the encoded 2n2 level CPFSK scheme.

4.3.3 Code Search and Tables of Good Codes

For the case of rate 2/3 convolutional encoders with eight-level CPFSK modulation, the encoder consists of an uncoded MSB and a rate 1/2 code on the remaining bit, as in Figure 4.19. The optimum [largest $d^2_{\min}$ (h)] rate 1/2 convolutional encoders for this case have been found when 1 < v < 4 Some of these rate 1/2 encoders can then be used in rate 3/4 and rate 4/5 convolutional encoders combined with 16-level and 32-level CPFSK modulation. The specific modulation indices studied were h = 2/22,

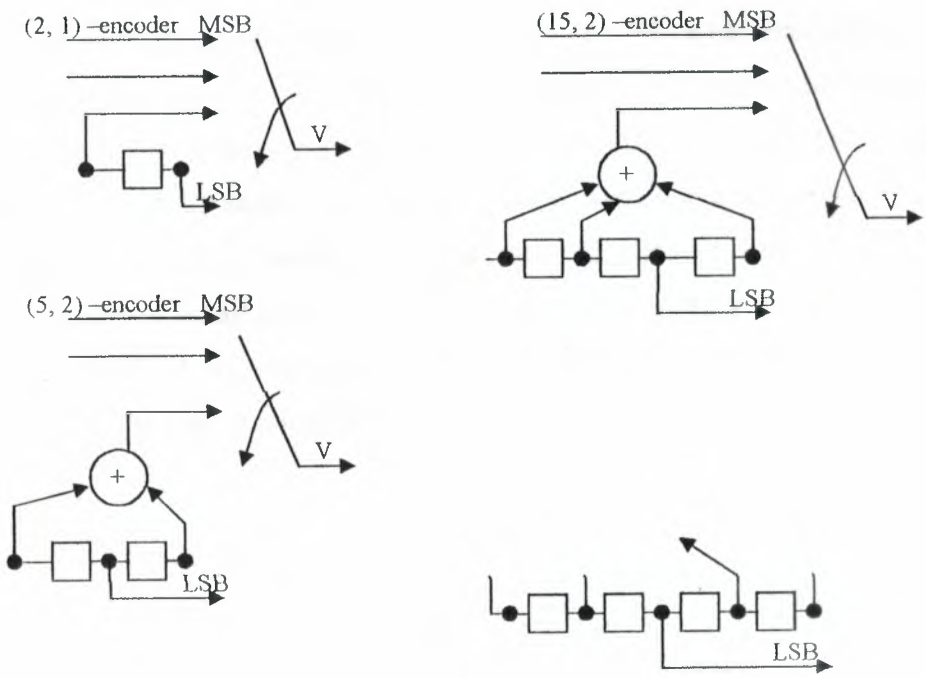


Figure4.18 Examples of good rate 3/4 convolutional encoders when combined with 16-level CPFSK modulation (assuming small modulation indices).

Rate 2/3 Convolutional Encoders and Eight-Level CPFSK

The optimal rate 1/2 convolutional encoders with v = 1, 2, 3, and 4 are listed in Appendix D, Tables D.6-D.9, together with their distances. At v = 1 and h = 1/10, for example, the best code has square distance 0.983 and there are four encoders with this distance. At v = 2, the distances achieved are significantly larger than at constraint 1; at h = 2/9, eight encoders all achieve square distance 4.46, which is also the upper bound value at this h. At v = 3, distances are again somewhat larger. At constraint 4, heavy calculation is required and complete searches have been performed only for a few h; at other indices distances have been calculated for the (33, 4) encoder only. Distance again increases and the upper bound is not reached at a number of modulation indices.

Rate 3/4 and Rate 4/S Convolutional Encoders

For rate 3/4 and rate 4/5 convolutional encoders combined with 16-and 32-level CPFSK modulation no complete search for the optimal rate in- is increased. For small modulation indices (h < 2/17) the upper bound is not reached at v < 4.

Summary of Numerical Results

Very good schemes of combined rate (n - 1)/n convolutional encoders and 2^n level CPFSK modulation, n = 3, 4, 5, have been found when the modulation index is in the interval 0 <h < 1/4. Examples of good rate 1/2 encoders are the (2, 1)-encoder, the (5, 2)-encoder, the (15, 2)-encoder, and the (33, 4)-encoder. These are shown in Figure 4.22 assuming n = 4. When using these four rate 1/2 encoders, the shortest minimum distance mergers at small h are y = (-4, 6, -2), y = (-4,6, -6,4), y = (-4, 8, -6,6, -4), and y = (-4, 4, 2, 0, -2, -4, 4), respectively. If we use these 'y sequences to construct upper bounds on d^2_{min} (h) we obtain (CPFSK modulation)

$$d^2{}_{\min}(h) \le (n-1)\left(3 - \frac{4}{3}\frac{\sin(2\pi h)}{2\pi h} - \frac{5}{3}\frac{\sin(4\pi h)}{4\pi h}\right), \qquad (2,1)-encoder$$

$$d^2{}_{\min}(h) \le (n-1)\left(4 - \frac{2}{3}\frac{\sin(2\pi h)}{2\pi h} - \frac{10}{3}\frac{\sin(4\pi h)}{4\pi h}\right), \qquad (5,2)-encoder$$

$$\sum_{i=0}^{N-1} Yi = \frac{2l}{h} \quad for \quad some \quad l,\ l = 0 . \pm 1, \pm 2, \ldots\ldots$$

$$d^2{}_{\min}(h) \le (n-1)\left(7 - 2\frac{\sin(2\pi h)}{2\pi h} - 4\frac{\sin(4\pi h)}{4\pi h} - \cos(2\pi h)\right), \qquad (33,4)-encoder$$

Some of the encoder-dependent upper bounds on d^2_{min} (h) are shown in Figures 4.23 (n = 3) and 4.24 (n = 4). Also shown are the corresponding encoder-independent upper bound (4.19) and uncoded upper bound. Since many of the upper bounds in Figures 4.23 and 4.24 are tight in the CPFSK case, these figures actually show d^2_{min} (h) for some specific coded CPFSK schemes.

For the same four encoders, we can replace the CPFSK frequency pulse with a smoother one from (4.17) and look at the effect of the smoothness parameter r on the upper bounds. As shown in Figures 4.23-4.24 the bounds do not change much if 0 < ε < 1/2. We can expect that for small r will be similar to the CPFSK case.

bounds. As shown in Figures 4.23-4.24 the bounds do not change much if $0 < \varepsilon < 1/2$. We can expect that for small r will be similar to the CPFSK case.

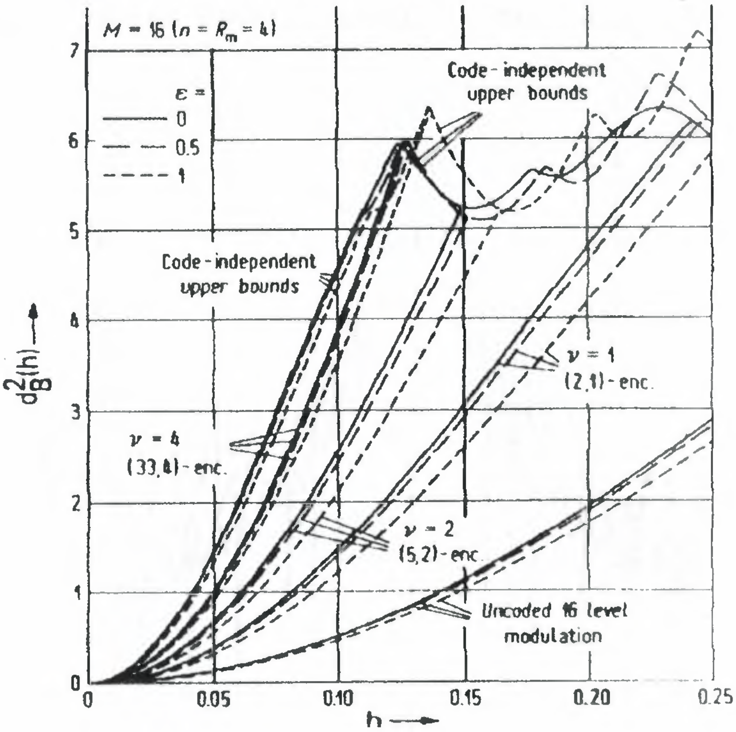


Figure 4.19 Upper bounds on d^2_{min} (h) when a rate 3/4 convolutional encoder and 16-level CPM are used. The frequency pulse g (t) follows equation (4.17).

4.3.4 Energy and Bandwidth Efficiency of Coded CPM

The true power spectral density for coded CPM schemes is not known. Instead we use estimates defined by

$$G_c(\beta) = R_c G(R_c \beta), \qquad \beta = fT_b \tag{4.22}$$

Where G (β) is the normalized power spectral density for the corresponding uncoded 2^n level CPM schemes. Figure 4.19 shows examples of estimated power spectral densities when a rate 2/3 convolutional encoder and eight-level CPM is used. The modulation index in this figure is h = 0.09(1/11), and a reduction of the sidelobes with increasing e is clearly demonstrated. Figure 4.20 shows estimated power spectral densities for two useful coded CPFSK schemes, with the power spectra for conventional uncoded MSK also given as a reference. The rate 2/3 (33, 4)-encoder combined with eight-level CPFSK and h = 1/11 yields a spectrum which is much better than the MSK spectrum, even though the coded scheme has d^2_{min} =2.1615, which is slightly larger than d^2_{min} for MSK. Thus this coded CPFSK scheme saves bandwidth compared to MSK without loss of energy efficiency. The rate 3/4 (33, 4)-encoder combined with 16-level CPFSK and

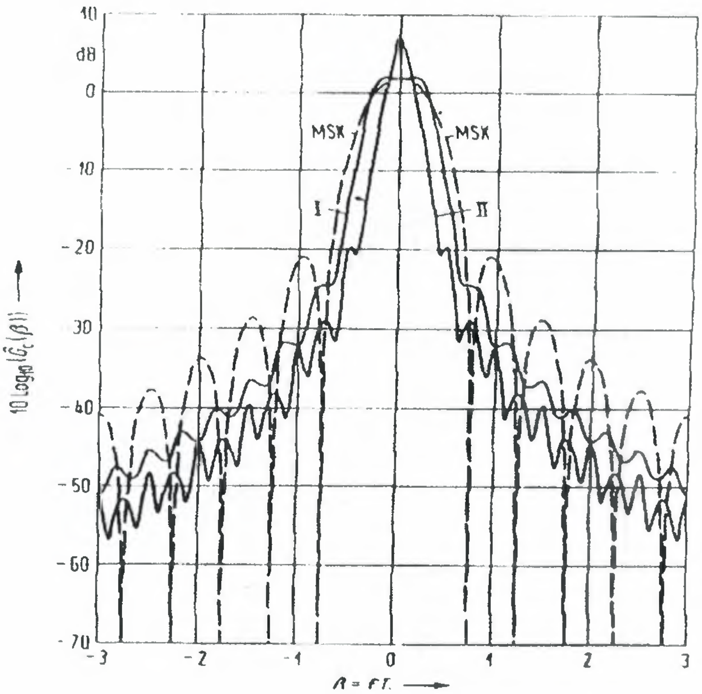


Figure 4.20 Estimated power spectra for some interesting coded CPFSK schemes.

h = 2/15 also yields a better spectrum than MSK, but has d^2_{min} = 5.64! This coded CPFSK schemes saves both power and bandwidth relative MSK. Still more spectrally efficient schemes can easily be obtained by using a small $\varepsilon > 0$ instead of CPFSK.

Figure 4.20 plots bandwidth vs. h using the 99% power in band definition of bandwidth as in Section 5.1. This is estimated by the relation $2B_CT_b = 2BT_b/R_c$, in which BT_b is the single-sideband rf normalized bandwidth of the uncoded CPM scheme. Some uncoded schemes are shown for reference. Also compare Figure 4.17.

Figure 4.21-4.22 show power-bandwidth tradeoffs for some of the schemes considered in Section 11.3. On the vertical axis is the gain in d^2_{min} compared to MSK and on the horizontal axis is the estimated normalized double sideband 99% bandwidth. (The connection lines between schemes with different h in these figures do not contain any numerical results.) Figure 4.19 depicts the rate 2/3 encoders given in Tables D.5-D.9 combined with eight-level CPFSK modulation. Specific schemes are shown as rectangles; note that coded systems with the same h value have the same estimated bandwidth for all v values. The coded schemes become more power efficient with increasing v until the encoder-independent upper bound is reached. Figure 4.29 considers schemes consisting of the rate 3/4 encoders given in Tables D.10 and D.11 combined with 16-level CPFSK

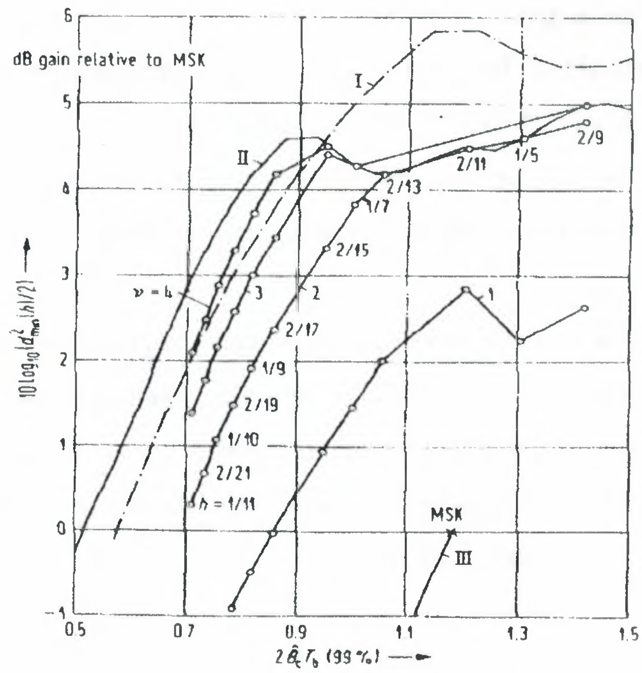


Figure 4.21 Power-bandwidth trade-off for some of the coded 16-level CPFSK schemes considered in this chapter. v = 1, 2, 3, and 4. The 99% power in band definition of bandwidth is used. I is the upper bound for the coded 32-level CPFSK, II is the upper bound for coded 16-level CPFSK, and III is binary CPFSK.

Modulation. These schemes also become more power efficient with increasing v, until the encoder-independent upper bound is reached.

Figure 4.21 considers the best schemes found in Refs. 3 and 4. It is quite obvious from this figure that schemes have been found that are much better than the MSK scheme both in terms of asymptotic error performance and of bandwidth. An example of such a scheme is the rate 3/4 (33, 4)-encoder combined with 16-level CPFSK modulation and h = 2/17. This scheme has an asymptotic gain of 4.2 dB and the bandwidth is only =74% of the MSK bandwidth. Other schemes appear that have the same asymptotic error performance as MSK but with roughly only half the MSK bandwidth; alternately, schemes with the same spectral efficiency as MSK but with an asymptotic performance gain of 5-6 dB have been found. Figure 4.22 shows some of the best available multi-h

CPFSK schemes (denoted A-D) for comparison. these are clearly outperformed by the coded CPFSK schemes considered here.

4.3.5 Conclusions

Good coded CPFSK schemes have been found by utilizing a high rate convolutional code, and a modulation scheme with a large number of levels and a low modulation index. Relative to MSK, these schemes can either reduce bandwidth by a factor of 2 or reduce power by 5 dB. By varying the. System parameters, notably the modulation index, schemes can be obtained with different combinations of bandwidth and power savings.

Code-independent upper bounds on minimum distance are useful because once they are achieved for a certain code constraint length; no increase in this length will give better performance. Upper bounds were often reached in this section. This means that for even better performance, new encoder structures and mapper must be explored. The results for coded CPM schemes in this section have been given in terms of the minimum

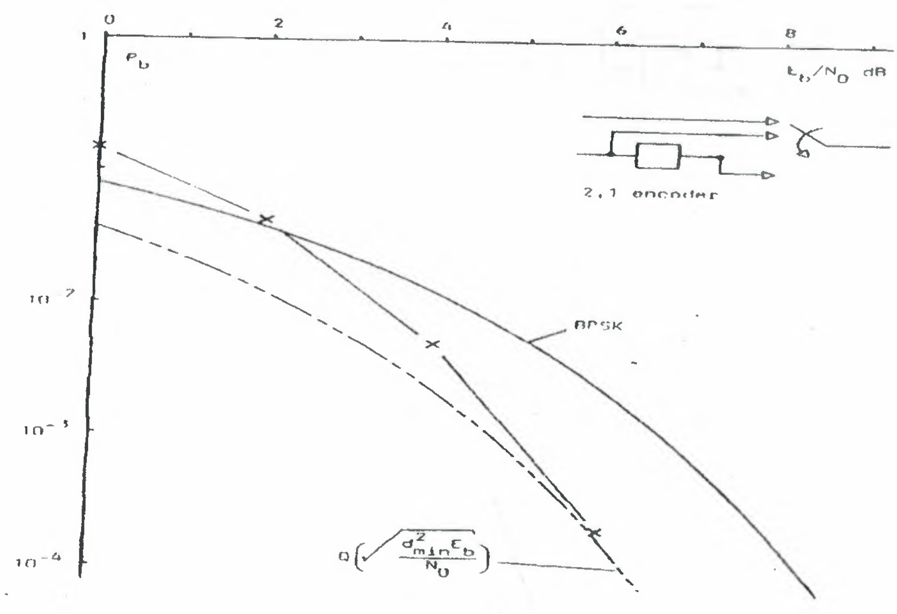


Figure 4.22. Simulated bit error probability for a rate 2/3 encoder with memory **v** =1 and polynomials (2, 1), combined with an eight-level CPFSK scheme with h = 1/5.

Euclidean distance. Recently, tight upper bounds on the bit error probability have been developed. For details, we have discussed only one type of smooth partial response modulation in combination with coding. It had significantly lower side lobes than did the CPFSK schemes, but the spectral main lobe was not much changed. The best encoders for CPFSK were simply carried over to these smoother. Modulations Except for scattered work, little is known about how to combine partial response and coding.

Achieving the good signal properties of coded CPM requires a more complex receiver, and one that employs sequence detection. It is shown in Ref. 4 that the MLSE receiver has

$$S = P \ .2^{v+(L-1)K} \tag{4.23}$$

states in its Viterbi processor, if the modulation index is h = 2i/p (i and p are integers without common divisor) v, L, and K have an exponential effect on S; the small h, which we would like to have, leads to a growth proportional to 1/h. Some of the complexity-reducing approaches.

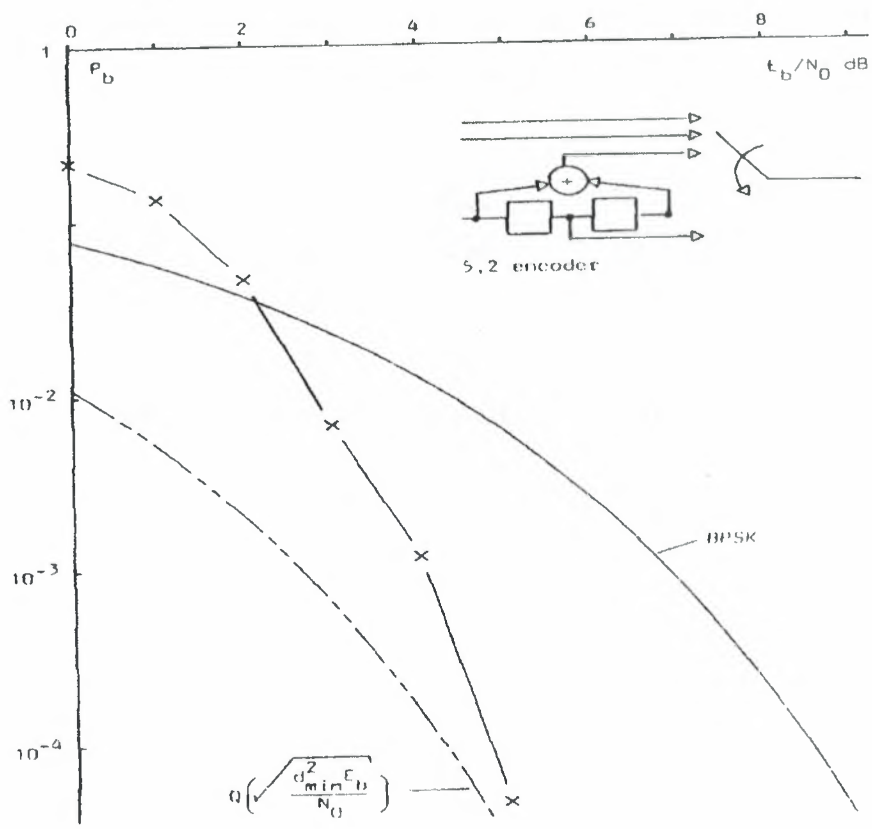


Figure 4.23 Simulated bit error probability for a rate 3/4 encoder with v = 2 and polynomials (5, 2), combined with a 16-level CPFSK scheme with h = 2/13.

4.4 Simulations of Coded CPFSK Schemes

The technique described has been applied to code CPFSK, with results that we will review in this section. Tables 4.3 parameters for the simulated schemes and their energy- bandwidth tradeoffs appear in Figure 4.30. The path memory for the Viterbi detector is 50 intervals. The simulation method is used with $\zeta = 20$ samples per interval. As the reference, the symbol error probability of BPSK, Q $((2E_b / N_0)^{1/2})$, is always given. Simulation results and bounds for CPFSK were presented for coded schemes the multiplicative factor in front of the Q-function is not known, and it has been set to 1.

A rate 3/4 encoder with v = 2 and the polynomials given by (5, 2) is considered in Figure 4.22 A natural binary mapper is used and h is 2/13 for the 16-level CPFSK scheme. The squared free distance is 5.23 and the number of states is 52.1000 bit errors are simulated for 0 and 1 dB, 500 for 2dB, 100 for 3 and 4dB, and 11 for 5dB.

Table 4.3 Parameters of M = 4 Level 2 - h Multi-h Schemes from Ref. 10. $2BT_b$ Is the 99% Bandwidth

scheme	h_1	h_2	d^2_{min}	N	$2BT_b$ 99%	S
A	0.40	0.46	4.80	30	1.13	100
B	0.25	0.29	2.80	30	0.84	200
C	0.25	0.23	2.00	30	0.73	200
D	4/16	3/16	1.60	<30	0.67	32

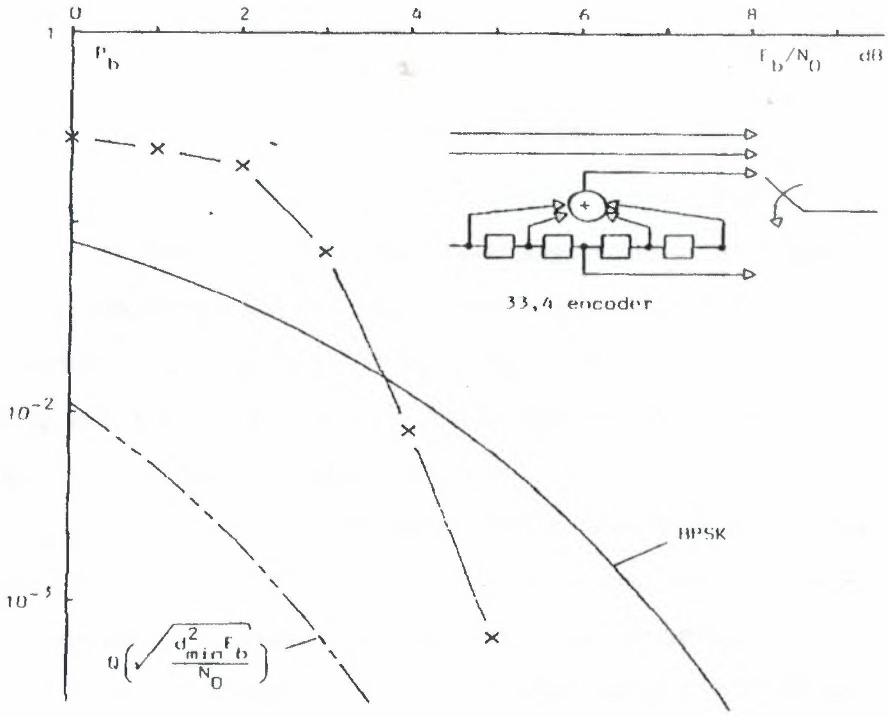


Figure 4.24. Simulated Bit error probability for a rate 2/3 encoder with v = 4 and polynomials (33, 4), combined with an eight-level CPFSK scheme with h = 1/10.

An eight-level CPFSK scheme with h = 2/21 and a rate 2/3 encoder with in. v = 3 and polynomials given by (17,2) is considered in Figure 4.24. A natural binary mapper is used. The squared free distance is 2.00 and the number of states is 168. 1000 bit errors are simulated for 0 to 6 dB, 135 for 8 dB, and 30 for 9 dB. The mapper is a natural binary and the modulation index in the 16-level CPFSK scheme is 2/17. This scheme has squared free distance equal to 5.23 and 272 states. 190 bit errors are simulated for the 5-dB point. Finally, figure 4.24 shows the results for a rate 2/3 code and with v = 4 and polynomials given by (33, 4). A natural binary mapper is used and h = 1/10 for eight-level CPFSK. The squared free distance is 2.59 and the number of states is 320. 2000 bit errors are used for 0 and 2 dB, 1000 bit errors for 4 and 6 dB, and 55 for 8.5 dB.

CONCLUSION

A non-perfect multiplier with a filter is one of two solution that allows only the wanted components to pass, whilst the other is to arrange a pair of non-linear elements, having as near as possible the same characteristics, in a balanced circuit so that the unwanted components cancel out. The degree of suppression depends upon the closeness of matching, and in practice it is usually found necessary to provide further suppression by adding an output filter-this requirement.

One is known as pulse-code modulation (PCM), in which the amplitude range of the sampled data is divided into a finite number of discrete levels; the amplitude of a given pulse is referred to the nearest level and a digital code generated.

Digitally encoded information can tolerate more signal distortion due to Intersymbol interference and noise pick-up than an analogue signal, and how the analogue-to-digital conversion can be effected. It is concerned with establishing the optimum pulse shape and filtering of the digital signal, in order that the received pulse sequence may be interpreted with the minimum error.

CPM signal can be utilized to improve the minimum Euclidean distance; so also can the memory introduced by the controlled Intersymbol interference in partial response CPM. Even more memory can be built into the signals by means of multi-h coding or convolution codes. In this chapter we will present some recent results on combinations of many-level PSK and CPM with convolutional codes. Covers binary and quaternary CPFSK with rate 1/2 convolutional codes, and presents combinations with the best free Euclidean

References

[1]. A. J. Viterbi and i. K. Omura, *Principles of Digital Communications and Coding,* McGraw-Hill, New York (1979).

[2]. C. Clark Jr. and 1. B. Cain, *Error Correction Coding for Digital Communications* Plenum Press, New York (1981).

[3]. G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Transactions on Information Theory* IT-28(1), 55-67 (1982).

[4]. G. Lindell and C-F. Sundberg, "Multilevel Continuous Phase Modulation with High Rate Convolutional Codes," GLOBECOM '83, San Diego, November 1983, Conference Record, pp. 30.2.1-30.2.6.

[5]. G. Lindell, C-F. Sundberg, and A. Svensson, "Error Probability of Multilevel CPM with High Rate Convolutional Codes," International Zurich Seminar on Digital Communications, March 1984, Conference Record, pp. 93-100.

[6]. G. Lindell, C-F. Sundberg, and T. Aulin, "Minimum Euclidean Distance for the Best Combination of Short Rate 1/2 Convolutional Codes and Quaternary CPFSK Modulation," Technical Report TR-170, Telecommunication Theory, University of Lund, August 1982.

[7]. G. Lindell and C-F. Sundberg, "High Rate Convolutional Codes and Multilevel Continuous Phase Modulation," Technical Report TR-180, Telecommunication Theory, University of Lund, November 1983.

[8]. G. Lindell and C-F. Sundberg, "Power and Bandwidth Efficient Coded Modulation Schemes with Constant Amplitude," *Achiv Elektronik und Ubertragungstecknik,* Band 39, Heft 1, 45-56 (1985).

[9]. G. Lindell, C-F. Sundberg, and T. Aulin, "Minimum Euclidean Distance for Combinations of Short Rate 1/2 Convolution Codes and CPFSK Modulation," *IEEE Transactions on Information Theory,* IT-30(3), *509-5* 19 (1984).

[10]. J. B. Anderson and C. P. Taylor, "A Bandwidth-Efficient Class of Signal Space *Codes,"IEEE Transactions on Information Theory* IT-24(6), 703-7 12 (1978).

[11]. J. B. Anderson, C-F. Sundberg, T. Aulin, and N. Rydbeck, "Power-Bandwidth Performance of Smoothed Phase Modulation Codes," *IEEE Transaction on Communications* COM29(3), 187-195 (1981).

[12]. J. L. Massey and M. K. Sam, "Inverses of Linear Sequential Circuits," *IEEE Transactions on Computers* C-17, 330-337 (1968).

[13]. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering,* Wiley, New York *(1965).*
29(11), 1707-1715 (1981).
[14]. R. Blahut, *Error Control Coding,* Addison-Wesley, Reading, Massachusetts (1983).
[15]. S. 0. Wilson, P. J. Schottler, and H. A. Sleeper, "Rate 3/4 16-PSK Phase Codes," IEEE International Conference on Communications, ICC '82, Philadelphia, June 1982, Conference Record, pp. 6F.1.1-6F.1.5.
[16]. S. G. Wilson *et al.,* "Convolutional Coding Combined with Continuous Phase Modulation," University of Virginia, Charlottesville, Report NAG-3-141, November 1982.
[17]. T. Aulin, G. Lindell, and C-F. Sundberg, "Minimum Euclidean Distance for Short Convolution Codes and Continuous Phase Modulation," GLOBECOM '82, Conference Record, pp. E.7.2.2-E.7.2.5.
[18]. T. Aulin and C-F. Sundberg, "On the Minimum Euclidean Distance for a Class of Signal Space Codes," *IEEE Transactions on Information Theory* IT-28(1), *43-55* (1982).
[19]. T. Aulin and C-F. Sundberg, "Minimum Euclidean Distance and Power Spectrum for a Class of Smoothed Phase Modulation Codes with Constant Envelope," *IEEE Transactions on Communications* COM-30(7), 1721-1729 (1982).
[20]. T. Aulin, N. Rydbeck, and C-F. Sundberg, "Continuous Phase Modulations-Part Impartial Response Signaling," *IEEE Transactions on Communications* COM-29(3), 210-225 (1981).
[21]. S. W. Pizzi, "Modulation and Convolutional Coding for Power and Bandwidth Efficiency," thesis, May 1982, Dept. Elec. Eng., University of Virginia.
[22]. T. Aulin, G. Lindell, and C-F. Sundberg, "Minimum Euclidean Distance for the Best Combination of Short Rate 1/2 Convolutional Codes and Binary CPFSI(Modulation," Technical Report TR-160, Telecommunication Theory, University of Lund, November 1981.