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

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

SMR.379/17

COURSE ON BASIC TELECOMMUNICATIONS SCIENCE

9 January - 3 February 1989

Lecture Notes

S. PRASAD

Indian Institute of Technology, Dept. of Electrical Engineering,
New Delhi, India

Analogue Modulation of Carriers

I.Introduction:

The transmission of signals from one place to another requires the use of a carrier, which generally takes the form of a sinusoidal signal of an appropriately chosen frequency. The choice of the frequency depends on the selection of the medium of propagation of the signal. For example, the HF band (30 MHz - 300 MHz) is selected when the signal is to be transmitted over the ionosphere, for long distance propagation, as generally done for short wave radios. On the other hand, the UHF and VHF band are used when a wider bandwidth is needed such as in video and T.V. transmission, or transmission of multiple voice channels via PCM or frequency division multiplexing. Still higher frequencies (in the microwave or millimeter wave frequency bands, or even optical bands), are used to obtain even higher bandwidths, as for example, in satellite communications and other applications.

The underlying information bearing signal (here assumed to be voice, picture or analog telemetry data) is typically a lowpass signal, and must therefore be mapped onto the carrier. This is not only useful but also necessary for the following reasons:

Ease of Radiation: For efficient radiation of electromagnetic waves, the antenna size should be of the order of the wavelength. Clearly, the required antenna sizes would become unmanageable for transmission of raw information bearing signals such as audio and video. The use of a high frequency carrier on to which this information can be embedded, these are much more manageable.

Multiplexing: All voice (or all picture) signals occupy nearly the same frequency band. If a number of voice signals are to be transmitted simultaneously, these need to be located in different spectral bands. This can be accomplished through modulation.

Frequency Assignment: In the same manner several radio or television stations which need to broadcast simultaneously, can do so by using different carrier frequencies.

Signal Processing: The signal processing techniques or technology that can be used depends on the frequency of operation of the receiver. Modulation provides a technique for translating the frequency band to a convenient location from this point of view.

Noise Reduction: Certain types of nonlinear modulations can even provide enhanced protection against the ever present noise in communication systems. This, however, is usually obtained at a sacrifice of transmission bandwith.

II. Fundamentals of Analog Signal Transmission:

Fig.1 shows the block diagram of a typical communication system. The transmitter and the receiver amplify signal power and perform some filtering operations. They may also include the modulator and the demodulator in case of bandpass transmission. Ideally, the output of the receiver should be a faithful replica of the message transmitted. However, noise and distortion introduced in the channel due to its nonideal nature, cause the output signal to be different from the transmitted signal.

Distortionless transmission:

The transmission is distortionless if

$$y(t) = Kx(t-t_d)$$

where K = attenuation and t_d = time delay

Power loss in transmission = 20 $\log_{10}K$

Typical values range from 0.05 dB/km (for twisted pair of wires at low frequencies) to 3-4 dB/km at higher in (twisted pair, coaxial cables waveguides and optical fibers).

Transfer function of channel for distortionless transmission:

$$H_c(f) = K \exp(-j2\pi f t_d) \quad \text{for} \quad |f| < B$$

where B is the bandwidth of the baseband signal.

Real channels usually do not satisfy this condition and some amount of distortion is always encountered. Proper signal, transmitter and receiver design can minimize its effect.

Types of Distortion:

1. Amplitude Distortion: This occurs when the amplitude response is not flat over the desired bandwidth, i.e., when

$$|H_c(f)| \neq K$$

Usually takes the form of excessive attenuation or enhancement of high or low signal frequencies. Its effect is usually negligible, if $|H_c(f)|$ is constant to within a dB in the desired band of frequencies.

2. Phase or Delay Distortion: This occurs if different frequencies of the signal get delayed by different amounts of time. As implied by (1) and (2), there will be no phase distortion if

$$\text{angle of } H(f) = -2\pi t_d f \pm m\pi$$

Any other type of phase response (including a constant phase shift) will cause delay distortion.

Delay distortion is critical in pulse and digital transmission. However, human ear is quite insensitive to delay distortion and hence it is of little importance in analog transmission of information.

Both amplitude and delay distortion are generally referred to as linear distortion. The linear distortion can be removed via the use of an equalizer, which is essentially a filter with a transfer function which is the inverse of the channel transfer function. Thus, the equalizer at the receiver is given by

$$H_{eq}(f) = \frac{K \exp(-j2\pi f t_d)}{H_c(f)} \quad \text{for } |f| < B$$

so that $H_c(f)H_{eq}(f) = K\exp(-2\pi f t_d)$, thus satisfying the condition for distortionless transmission.

3. Nonlinear Distortion: This occurs in many practical systems, particularly when the signal amplitude at the input to a power amplifier etc., is large. A typical example of nonlinear transfer characteristic is the saturating nonlinearity shown in Fig. 2. A typical model for nonlinear characteristics is the polynomial model given by

$$y(t) = a_1x(t) + a_2x^2(t) + a_3x^3(t) + \ldots$$

and a typical effect of this kind of distortion is the generation of the so-called harmonic or intermodulation distortion. This is best understood by realizing that if the input contains two sinusoids at frequencies f_1 and f_2 respectively, then the output will contain not only these frequency components, but also the (generally undesirable) components at $f_1 \pm f_2$, $f_1 \pm 2f_2$,... etc. This kind of distortion can cause severe communication problems, such as cross-talk, in a multiplexed environment.

The nonlinear distortion of the above type can best be tackled via a so-called "compander" device, which is comprised of an amplitude "compressor" at the transmitter (or the appropriate amplifier input) and an "expander" at the receiver or the output of the nonlinear device. A typical set of compression and expander characteristics are the logarithmic characteristics given by $g_c(x) = \ln x$ and $g_e(y) = \exp(y)$, respectively.

III. Linear Modulation of Carriers:

Direct frequency translation of a message spectrum can be done via the linear modulation process given by

$$x_c(t) = Am(t)\cos w_c t$$

where m(t) refers to the message signal. There are a number of important variations of this basic translation process as follows:

A. Double Sideband Suppressed Carrier Modulation (DSB-SC): This refers to the basic equation (5) above. A typical DSB-waveform for a sinusoidally modulated signal is shown in Fig. 3. From basic Fourier transform theory, the spectrum of the modulated signal is given by

$$X_c(f) = (A/2)[X(f+f_c) + X(f-f_c)]$$

where $f_c = w_c/2\pi$. The frequency domain representation of the translated spectrum is shown in Fig. 4.

Here m(t) is the baseband signal.

Multiplication of baseband signal and the carrier signal is called mixing or hetrodyning.

The upper sideband in the translated spectrum contains components corresponding to the positive frequencies of the baseband signal. The lower sideband contains components corresponding to the negative frequencies of the baseband signal.

If B is the baseband message bandwidth, then the bandwidth of the DSB-SC signal is given by $B_T = 2B$.

The average transmitted power in the DSB-SC signal can be seen to be given by

$$P_T = P_c P_m$$

where $P_c = A^2/2$ is the average carrier power.

Recovery of the baseband signal: The baseband signal can be recovered from the received DSB-SC signal by rehetrodyning it at the receiver with a local carrier which is identical (except perhaps for amplitude) to that at the transmitter and low pass filtering the result (Fig. 5). Thus we have

$$z(t) = k[m(t)\cos w_c t][2\cos w_c t]$$

$$= km(t) + km(t)\cos 2w_c t$$

and the corresponding spectrum:

$$Z(f) = kM(f) + (k/2)[M(f-2f_c) + M(f+2f_c)]$$

Since for $B < f_c$, M(f) does not overlap with the spectra $M(f-2f_c)$ or $M(f+2f_c)$, it follows that the low pass filtering of Z(f) will yield the desired message spectrum.

Carrier Recovery: The above method of demodulation is called synchronous or coherent demodulation. This requires the locally generated carrier to be phase synchronous with the carrier of the information signal exactly.

Lack of such synchronism will produce distortion such that the recovered signal will turn out to be

$$y(t) = k\, m(t) \cos(\delta w t + \theta)$$

where δw and θ are the frequency and phase offset respectively, in the local carrier.

If $\delta w = 0$ and $\theta = \pi/2$, the signal is lost completely.

If $\theta = 0$, $y(t) = km(t) \cos \delta w t$ will produce a warbling effect. In voice signals, $\delta f > 30$ Hz is unacceptable.

A phase coherent carrier signal can be recovered from the received signal itself by a squaring circuit followed by a bandpass filter centered at $2f_c$. Since $m^2(t)$ will have a nonzero d.c. component, the modulated signal $x_c^2(t)$ has a carrier component at $2f_c$, which can be extracted by a bandpass filter. Division of this frequency by 2, yields the required phase coherent carrier, which can be used for demodulation. This is shown in Fig. 6.

B. Double Sideband AM (with Carrier):

This is generated by adding a large carrier component to the DSB signal. Thus we have

$$x(t) = A[1 + m(t)]\cos w_c t = e(t) \cos w_c t$$

where $e(t)$ is the so-called envelope of the modulated signal. Provided that $|m(t)| < 1$, the envelope $e(t)$ follows the shape of the message signal, as shown in Fig. 7, and the spectrum $X(f)$ is similar to that of the DSB-SC signal, except for the presence of a carrier component also at f_c, as shown in Fig. 8.

This property of the envelope shape, makes the recovery of the message from the received signal very simple, as will be seen shortly.

The modulation index m of an AM signal is defined as

$$m = \frac{[e(t)]_{max} - [e(t)]_{min}}{[e(t)]_{max} + [e(t)]_{min}}$$

In order for the envelope to remain undistorted, the modulation index should be less than unity.

The bandwidth of the AM signal is identical to that of the DSB-SC signal, viz., $B_T = 2B$. However, since a carrier component is also present, the total transmitted power is given by

$$P_T = P_c + P_c P_m$$

where P_m is the normalised message power. The carrier component of this power does not contain any useful intelligence, and hence is wasted. Thus, we define the power efficiency of the AM signal to be

$$E = \frac{P_c P_m}{P_c + P_c P_m}$$

The maximum efficiency of the AM signal can be at best 50% for an arbitrary signal, and only about 33.3% for a sine wave message signal.

Demodulation:

The demodulation of an AM signal is effected in a very simple manner by using the diode detector circuit shown in Fig. 9. Its working is self-explanatory.

This demodulation does not require any synchronous carrier and hence expensive carrier recovery circuits are not needed at the receiver. This makes it ideally useful for broadcast applications, where the cost of the receiver is a major consideration.

C. Suppressed Sideband Modulations:

Since both the sidebands contain Identical information, it is possible to save on transmission bandwidth and power by suppressing one of them either completely, or at least partially. This leads to the single sideband and vestigial sideband modulations.

(i) SSB Modulation:

Here only one of the two sidebands is transmitted. This can be done either by filtering out one of the two sidebands from the DSB signal obtained after hetrodyning. The frequency domain representation is shown in Fig.10, which also shows the recovery technique of the baseband message via synchronous demodulation. The bandwidth of transmissions well as the average transmitted power is half of that in DSB-SC. Thus, we have

$$B_T = B \quad \text{and} \quad P_T = P_c P_m / 2$$

Practical implementation of the SSB system, is however, quite complex, both at the transmitter as well as the receiver. This is because the modulator filter required for removing the undesired sideband must be an ideal bandpass filter, due to the proximity of the two sidebands. Secondly as in DSB-SC, the demodulation requires a synchronous carrier.

Phase Shift method of SSB Signal Generation:

This method is based on the following representation of the SSB signal:

$$x(t) = m(t) \cos w_c t + \hat{m}(t) \sin w_c t$$

where $\hat{m}(t)$ denotes the Hilbert transform of the signal, which is obtained by shifting all the frequency components of m(t) by (-90°).

For example, if $m(t) = \cos w_m t$, then $\hat{m}(t) = \sin w_m t$ and $x(t) = \cos(w_c - w_m)t$, which is the lower sideband signal. Similarly, we can generate the upper sideband by subtracting the quadrature terms.

A block diagram of the phasing method is shown in Fig.11.

(ii) Vestigial Sideband Modulation:

SSB modulation is suitable for message signals which do not have significant low frequency content, such as the speech signals etc. Absence of low frequency content increases the frequency separation of the two sidebands obtained after hetrodyning with a carrier, thus making their separation possible.

In many instances, baseband signals have both a large bandwidth as well as a significant low frequency content. Large bandwidth makes it necessary to use sideband suppression in some form, but the presence of low frequencies makes the use of SSB quite difficult. Examples of such signals are television, video, facsimile and data signals.

Such signals are best handled via the so-called vestigial sideband modulation, which results in both improved bandwidth and power efficiency.

VSB modulation involves the retaining of most of one sideband as well as a trace or vestige of the other. This is typically done by replacing the sharp cut-off sideband filter with one having a more gradual roll-off. It is important for the transfer function to have an odd symmetry about the carrier frequency, and a relative response of 1/2 at f_c.

The transmission bandwidth of VSB is slightly more than that of SSB but considerably smaller than that of DSB. Thus, we can write

$$B_T = B + \beta, \quad \text{where} \quad \beta < B$$

The VSB signal can be expressed in time domain as follows:

$$x(t) = (A/2)[1 + x(t)] \cos w_c t - (A/2)\, y(t) \sin w_c t$$

where $y(t) = f(m(t), \hat{m}(t))$.

VSB, like all other AM modulations, can be demodulated synchronously. However, it turns out that, if the carrier component is sufficient, it can also be demodulated simply by an envelope demodulator.

D: Methods of Hetrodyning:

The key operation in implementing all the above modulations practically is that of hetrodyning or mixing. This can be done via one of two basic types of mixers, viz., the balanced modulator or the switching modulator.

Balanced Modulator: It consists of two identical nonlinear elements (such as appropriately biased diodes) and some summing devices (e.g., operational amplifiers)., as shown in Fig. 12. Assuming a squaring nonlinearity, we can write

$$y(t) = a_1[A \cos w_c t + m(t)]^2 + a_2[A \cos w_c t + m(t)]^2$$
$$- a_1[A \cos w_c t - m(t)]^2 - a_2[A \cos w_c t - m(t)]^2$$
$$= 2a_1 m(t) + 4a_2 m(t) A \cos w_c t$$

Use of an appropriate bandpass filter to remove the second term yields the desired product signal

$$z(t) = K\, m(t) \cos w_c t$$

Switching Modulator: This is shown in Fig.13. Here the diodes act as switches operating at a rate of f_c. Thus, when the carrier is positive, the output voltage $v(t)$ is present, and when the carrier is negative, the output voltage is zero.

Thus, we have

$$v(t) = m(t) s(t)$$

where $s(t)$ is a switching function with frequency f_c. Assuming the switching function to be a symmetric square wave, the Fourier series expansion of this equation can be written as

$$v(t) = k_o m(t) + k_1 m(t) \cos w_c t + k_3 m(t) \cos 3w_c t + \ldots$$

Using a bandpass filter centered at f_c, we get the desired hetrodyned signal

$$x(t) = k\, m(t) \cos w_c t$$

ANGLE MODULATION

I. Introduction:

An angle or exponentially modulated signal has the general form

$$x(t) = A\cos[w_c t + \phi(t)] = \mathrm{Re}[A\exp(jw_c t + j\phi(t))]$$

The instantaneous phase, say θ_i of the carrier is given by

$$\theta_i(t) = w_c t + \phi(t)$$

The frequency of the carrier also varies and its instantaneous value is given by

$$w_i(t) = d\theta_i/dt = w_c + d\phi(t)/dt$$

$\phi(t)$ and $d\phi/dt$ are called the instantaneous phase and frequency deviations, respectively.

There are essentially two different types of angle modulations, viz.,

Phase Modulation: Here the instantaneous phase deviation of the carrier is made proportional to the message signal, i.e.,

$$\phi(t) = k_p m(t)$$

k_p is called the phase deviation constant.

Frequency Modulation: Here the frequency deviation is proportional to the message signal, i.e.,

$$d\phi/dt = k_f m(t)$$

or

$$\phi(t) = k_f \int_{-\infty}^{t} m(s)\,ds$$

k_f is the frequency deviation constant.

Thus we can write the phase and frequency modulated signals as follows:

PM: $$x(t) = A\cos[w_c t + k_p m(t)]$$

FM: $$x(t) = A\cos\left[w_c t + \int_{-\infty}^{t} m(s)\,ds\right]$$

Typical AM, PM and FM signals are shown in Fig. 14.

II. Spectrum, Bandwidth and Power of FM Signals:

Angle modulation (both PM and FM) are nonlinear processes. The exact calculation of their spectra is very difficult. Some insight can, however, be obtained by considering the case of sinusoidal or tonal message signals.

Assume that $m(t) = A_m \cos w_m t$. It follows that the instantaneous phase deviation is

PM: $\phi(t) = k_p A_m \cos w_m t$

FM: $\phi(t) = \frac{k_f A_m}{w_m} \sin w_m t$

Since the spectral properties of PM and FM signals are similar, we concentrate here on the study of FM signals.

The FM signal is given by

$$x(t) = A \cos(w_c t + \beta \sin w_m t)$$

where the parameter is called the modulation index and is given by

$$\beta = \frac{k_f A_m}{w_m} \quad \text{for FM} \quad \text{and} \quad \beta = k_p A_m \quad \text{for PM}$$

We can write

$$x(t) = A \, \mathrm{Re}\{\exp(jw_c t)\exp(j\beta \sin w_m t)\}$$

Using the Fourier series expansion of $\exp(j\beta \sin w_m t)$, we can write

$$x(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[(w_c + n w_m)t]$$

We can draw the following conclusions regarding the spectrum of FM signals (see Fig.15):

1. The spectrum contains a carrier component with an infinite number of sidebands at frequencies $f_c \pm nf_m$, unlike AM which contains only two sideband components, viz., $f_c \pm f_m$. The relative amplitudes of the various components depend on the modulation index through the function $J_n(\beta)$.

2. The number of significant spectral components is also a function of β. For $\beta << 1$ (known as narrowband FM), only J_0 and J_1 are significant. The resulting spectrum is similar to that of AM, except the associated phase reversal of the lower sideband component.

3. For large values of β, the number of significant components is large, implying a wideband signal. This is wideband FM.

Bandwidth: The transmission bandwidth of an FM signal is defined as the bandwidth which contains 98% of the total FM signal power. It is possible to see that for a sinusoidal modulating signal, this is given by

$$B_T = 2(\beta + 1)f_m$$

Noting that βf_m also denotes the peak frequency deviation of the FM signal, a more useful formula for bandwidth which is also valid for general message signals is given by

$$B_T = 2(\delta f + f_m)$$

where δf denotes the peak frequency deviation.

III. Generation and Demodulation of Angle Modulated Signals:

There are essentially two methods of FM generation. These are the so-called "Direct" and "Indirect" methods.

The "direct" method are based on the use of an appropriately designed voltage controlled oscillator or VCO, which can be implemented via a tuned oscillator with a variable reactance device, or a klystron (at microwave frequencies), or as a relaxation oscillator. The voltage controlled oscillator essentially produces an output signal whose instantaneous frequency is proportional to the input voltage. The variable reactance required for generating FM signals can be obtained using reactance tubes, saturable reactor elements, or reverse biased varactor diodes. The main advantage of direct FM is that large frequency deviations are possible for wideband FM. The main disadvantage is that the carrier frequency tends to drift and additional circuitry is needed for frequency stabilization.

The VCO can also be used for the generation of a PM signal, by inserting a differentiator between the signal and the VCO, in view of the previously discussed relationship between FM and PM. This is schematically shown in Fig.16.

The "indirect" method of generating wideband FM is a two-step process. In the first step, a narrowband signal is generated by using the following approximation for an FM signal:

$$x(t) = A\cos(w_c t + \phi(t))$$

$$= A\cos w_c t \cos\phi(t) - A\sin w_c t \sin\phi(t)$$

$$= A\cos w_c t - A\phi(t)\sin w_c t$$

assuming $\phi(t)$ to be small (which is true for narrowband FM). This can be done easily by using a mixer and a 90° phase shifter as shown in Fig.17. The wideband FM is generated thereafter by using a "frequency multiplier" also shown in the figure. A frequency multiplier is essentially an n'th law device followed by a bandpass filter, which is designed to multiply the frequencies of an input signal by a factor n.

Thus, if the narrowband FM signal is

$$e_i(t) = A\cos(w_c t + \beta \sin w_m t)$$

then the corresponding frequency multiplied signal is given by

$$e_o(t) = A'\cos(n w_c t + n\beta \sin w_m t)$$

thus multiplying both the carrier and the modulation index by a factor of n.

IV. Demodulation of FM Signals:

The devices generally used for the demodulation of FM signals are called frequency discriminators, which produce an output proportional to the input frequency or frequency deviation.

For the FM signal given by

$$x(t) = A\cos[w_c t + k_f \int m(s)ds]$$

the discriminator output is ideally given by

$$y(t) = k_d k_f m(t)$$

where k_d is the discriminator sensitivity. This is illustrated in Fig. 18.

A discriminator is generally built using the principle of slope detector, which consists of a bandpass differentiator (around the carrier frequency) to transform the frequency variations of the input FM signal into amplitude variations followed by an envelope demodulator (fig.18)

A discriminator based on these principles can be built as shown in Fig.19 using two tuned circuits, one resonant to a frequency above f_c and the other to below f_c, resulting in an S-shape curve for the input-output characteristic. A balanced discriminator of this type can provide a linear frequency response over a large range for the detection of wideband FM signals.

Behaviour of Analog Communication Systems in the Presence of Noise

I. Introduction:

Consider the block diagram shown in Fig.20.

Let S_T be the transmitted power, S_i and N_i the signal and noise powers at the receiver input and S_o and N_o the corresponding quantities at the output. The ratio S_o/N_o is the output SNR, which is a fundamental figure of merit for analog communication systems. Typical desirable of the output SNR's for various applications involving voice signals are (i) 5-10 dB for bare intelligibility, 25-35 dB for telephone quality and (iii) 45-55 dB for broadcast quality. Similarly, typical TV signals are desired to have an SNR of 45-55 dB.

II. Baseband Systems:

We start by considering the performance parameters of a baseband system, in which a message is transmitted without modulation of a carrier. A suitable model for the analysis of such a system is shown in Fig.21.

Here $H_p(w)$ denotes the filter used at the transmitter for limiting the message spectrum to the desired bandwidth, $H_c(w)$ models the channel characteristics and $H_d(w)$ is the receiver filter used to eliminate the out of band noise and other channel interference. Alternatively, $H_p(w)$ and $H_d(w)$ can also have additional functions to perform such as pre-emphasis and de-emphasis in order to optimize the SNR at the receiver.

For simplicity, we shall assume the message m(t) to be wide sense stationary and bandlimited to B Hz., and all filters to be ideal Lowpass with bandwidth B. The channel is assumed to be distortionless.

It follows that $S_o = S_i$ and

$$N_o = 2\int_o^B S_n(f)df$$

where $S_n(f)$ is the power spectral density of channel noise.

For white noise with density $S_n(f) = N/2$, we have $N_o = NB$ and hence

$$S_o/N_o = S_i/NB$$

Defining $\alpha = S_i/NB$ as the input SNR in the message bandwidth, we get

$$S_o/N_o = \alpha$$

This result will serve as a standard for comparison of the performance of the various modulation systems.

III. AM Systems:

A. DSB-SC: The system block diagram of interest is as shown in Fig.22. The transmitted signal is denoted here by $\sqrt{2}m(t)\cos w_c t$ and the channel is assumed to be comprised of an ideal bandpass filter of center frequency f_c and bandwidth 2B. The noise at the receiver input can be modelled as narrowband noise of bandwidth 2B and represented as

$$n_i(t) = n_c(t)\cos w_c t + n_s(t)\sin w_c t$$

and the signal at the demodulator by

$$y_i(t) = [\sqrt{2}m(t) + n_c(t)]\cos w_c t + n_s(t)\sin w_c t$$

Synchronous demodulation with a local carrier $\sqrt{2}\cos w_c t$ then yields the demodulator output as

$$y_o(t) = m(t) + (1/\sqrt{2})n_c(t)$$

Thus $S_o = \langle m^2 \rangle = S_i$ and $N_o = (1/2)\langle n_c^2(t) \rangle$. But for white channel noise, $var(n_c(t)) = 2NB = var(n_i(t))$ so that $N_o = NB$. It follows that

$$S_o/N_o = S_i/NB = \alpha$$

B. SSB-SC: It is easy to appreciate that in SSB both the transmitted signal power and the noise power input to the receiver are effectively halved due to the halving of the channel bandwidth. This implies that the output SNR of the SSB system would be identical to that of the DSB-Sc and baseband systems. It is easy to verify this by a formal derivation of the type used for the DSB-SC case.

C. VSB-SC (Synchronous Demodulation): This analysis is omitted here. It can be seen that even in this case the output SNR is approximately given as equal to α.

Thus we have the following general conclusion so far:

$$(S_o/N_o)_{DSB-SC} = (S_o/N_o)_{SSB} = (S_o/N_o)_{BB} = \alpha$$

and

$$(S_o/N_o) \approx \alpha$$

D. AM Systems: The AM signal with a modulation index of m can be demodulated either synchronously or via the envelope detector.

Synchronous Demodulation: When demodulated synchronously as in DSB or SSB, the SNR drops here because the power contained in the carrier does not contribute to the output SNR. The amount of reduction is related to the power efficiency, which in turn depends on the modulation index.

Using the same analysis method, it is straightforward to show that

$$(S_o/N_o)_{syn-AM} = \frac{\langle m^2\rangle a}{A^2 + \langle m^2\rangle}$$

It can also be shown that the maximum value of this SNR $\le a/2$. Thus synchronous AM is at least 3-dB (usually 6 dB in practice) worse than DSB-SC or SSB-SC depending on the modulation index and the signal m(t).

Envelope Detection: This analysis is slightly different. The received signal plus noise can be written as

$$y_i(t) = [A + m(t) + n_c(t)]\cos w_c t + n_s(t)\sin w_c t$$

The input signal power S_i is given by $S_i = \langle[A + m(t)]^2\rangle/2 = (A^2 + \langle m^2\rangle)/2$. The envelope of the received signal is given as

$$e_i(t) = \text{envelope} = [(A + m(t) + n_c(t))^2 + n_s^2(t)]^{1/2}$$

Small Noise Case: Assume that $[A + m(t)] \gg n_i(t)$. We can then write

$$e_i(t) \approx A + m(t) + n_c(t)$$

It follows then that once again the ratio of the desired signal to noise power ratio at the envelope detector output is given by

$$S_o/N_o = \frac{\langle m^2\rangle}{2NB} = \frac{\langle m^2\rangle a}{A^2 + \langle m^2\rangle}$$

which is the same as that for synchronous demodulation of AM signals.

Large Noise Case: The analysis here is more complex and omitted here. The main outcome of this analysis is that the signal quality at the output degrades rapidly (and disproportionately) when the input SNR degrades beyond a certain level, i.e., when a < a certain threshold value. This is called a threshold phenomenon and illustrated in Fig. 23. Typical value of this threshold input SNR is about 10 dB or so.

IV. Angle Modulated Systems:

A typical angle modulated system is shown in Fig.24. For wideband phase or frequency modulation, we note that the variations in the message signal are much slower than the noise fluctuations. This is because for a signal bandwidth of B Hz., noise has a bandwidth of $2(\delta f+B)$ Hz., with $\delta f \gg B$. In particular the phase and frequency variations of the modulated carrier due to m(t) are much slower than variations in the instantaneous phase of the noise input $n_i(t)$ to the demodulator. Intuitively, therefor, we can calculate the output noise of the demodulator by assuming m(t) to be zero over several cycles of the instantaneous carrier frequency.

While the actual calculations are omitted here, it is easy to appreciate that, under small noise conditions, the effect of amplitude and phase variations in the noise waveform can affect the phase of the carrier on a scaled down basis, the scaling factor being dependent on the input SNR. This is because the amplitude of the noise phasor is much smaller than that of the carrier.

More precisely, it can be shown that the output noise power in the message bandwidth in a phase demodulator is given by

$$N_o = 2NB/A^2$$

where A is the carrier amplitude. The desired output signal power by

$$S_o = k_p^2 \langle m^2 \rangle$$

where k_p is the phase modulation index. Writing here the input SNR α as given by

$$\alpha = S_i/NB = A^2/2NB,$$

we can write

$$(S_o/N_o)_{PM} = k_p^2 \langle m^2 \rangle \alpha$$

or, alternately, using $\delta w = k_p m_p'$ where $m_p' = [dm/dt]_{max}$, we can write

$$(S_o/N_o)_{PM} = (\delta w)^2 \alpha \langle m^2 \rangle / m_p'^2$$

thus showing that there is a tremendous improvement in the output SNR which depends on the square of the transmission bandwidth. There is in fact a 6dB SNR advantage for every doubling of bandwidth.

For the FM systems, we note that the demodulation can be carried out via a phase demodulation followed by a differentiator, as shown in Fig. 25. It follows, therefore that

$$S_o = k_f^2 \langle m^2 \rangle \quad \text{and} \quad N_o = (NB/A^2) \int_0^B w^2 dw = 8\pi^2 NB^3/3A^2$$

and

$$(S_o/N_o) = 3\left[\frac{\delta f}{B}\right]^2 \left[\frac{m^2}{m_p^2}\right] \alpha = 3\beta^2 \left[\frac{m^2}{m_p^2}\right]\alpha$$

where we have made use of the fact that $\delta w = k_f m_p$ and that the transmission bandwidth $= 2\delta f$. Once again we notice a significantly improved SNR, which, as for PM, increases by 6dB for every doubling of the transmission bandwidth. Typical performance curves for FM are shown in Fig.26.

<u>The Case of Large Noise-Threshold Effect:</u>

The above picture completely changes when the noise power is less than or equal to that of the carrier. A threshold phenomenon then results, which is much more pronounced than that observed in the envelope detection of AM signals.

For a qualitative appreciation of this phenomenon for angle modulation systems, we note that when noise is small, the noise phasor e_n rotates around the carrier phasor A, as shown in the figure. The angular perturbation $\delta\phi(t)$ of the resultant phasor R produced by noise is thus small, with $|\delta\phi(t)| << 2\pi$.

When e_n is large, R is much more likely to rotate round the origin. This is because $\delta\phi$ is much more likely to go through changes of 2π or larger, more frequently, because the noise varies much faster than the signal. Thus for large noise, spikes of area 2π appear in the derivative of $\delta\phi(t)$, which give rise to a crackling sound in an FM receiver, also known by the name of "<u>clicks</u>". Since this noise is impulsive or spiky in nature, it has a considerable power available at lower frequencies, which can not be filtered out by the final low pass filter.

This causes a considerably large contribution to the output noise power, which increases disproportionately with respect to the decrease in the input SNR as e_n A. This defines the onset of the <u>Threshold Effect</u>, as illustrated in Fig. 27, and works against the otherwise promising potential of AM for reducing the transmission power requirement at the expense of bandwidth. The value of the input SNR at which threshold phenomenon becomes manifest can be approximately written as:

$$\alpha_{thresh} \approx 2(\beta + 2)$$

BANDPASS SIGNALS

I. Definition and Representation:

A bandpass signal is one whose power spectral density is confined to a certain passband. If the signal is also a random one, it is sometimes called a "bandpass random process". The power spectral density of a typical bandpass signal is shown in Fig. 28.

Quadrature Representation: A bandpass signal can be represented in several different ways. However, a very powerful representation is the so-called "quadrature representation". It is given by

$$x(t) = x_c(t)\cos w_c t + x_s(t)\sin w_c t$$

where w_c is the center frequency and $x_c(t)$ and $x_s(t)$ are two lowpass signals.

In order to show this, consider the system shown in Fig. 29. Here $H_o(w)$ represents an ideal low-pass filter of band width B Hz., and with impulse response $h_o(t)$. We first show that the system shown is an ideal bandpass system of bandwidth 2B (fig.29). This is done here by evaluating the impulse response of the system.

If the input to the system is the impulse function $\delta(t-u)$, then the signals at various points in the system are as given below:

At a_1 : $2\cos(w_c u + \theta)\delta(t-u)$

At a_2 : $2\sin(w_c u + \theta)\delta(t-u)$

At b_1 : $2\cos(w_c u + \theta)h_o(t-u)$

At b_2 : $2\sin(w_c u + \theta)h_o(t-u)$

At c_1 : $2\cos(w_c u + \theta)\cos(w_c t + \theta)h_o(t-u)$

At c_2 : $2\sin(w_c u + \theta)\sin(w_c u + \theta)h_o(t-u)$

At d : $2h_o(t-u)\cos[w_c(t-u))]$

Clearly the system is a time invariant system with impulse response $h(t) = 2h_o(t)\cos w_c t$ and $H(w) = H_o(w+w_c) + H_o(w-w_c)$, which represents an ideal bandpass filter.

Clearly, if $x(t)$ with power spectral density (PSD) of Fig. 28 is applied to this system, it remains unaltered and appears as $x(t)$ itself. Denting the lowpass signals at b_1 and b_2 as $x_c(t)$ and $x_s(t)$ respectively, we get the above quadrature representation.

The narrowband or bandpass signal can also be represented via an envelope-phase relationship of the form:

$$x(t) = e(t)\cos[w_c t + \theta(t)]$$

where the envelope $e(t)$ and the phase $\theta(t)$ are given by

$$e(t) = [x_c^2(t) + x_s^2(t)]^{1/2}$$

and

$$\theta(t) = \tan^{-1}[x_s(t)/x_c(t)]$$

II. Characterization of $x_c(t)$ and $x_s(t)$:

$x_c(t)$ is obtained by multiplying $x(t)$ by $2\cos(w_c t + \theta)$ and low pass filtering. The PSD of $2x(t)\cos(w_c t+\theta)$ is given by $[S_x(w+w_c) + S_x(w-w_c)]$ and its low pass filtering results into the spectrum as shown in Fig. 30 for $x_c(t)$ (and similarly for $x_s(t)$). Thus

$$S_{xc}(w) = S_{xs}(w) = \begin{cases} S_x(w+w_c) + S_x(w-w_c) & |w| < 2\pi B \\ 0 & |w| > 2\pi B \end{cases}$$

We note that the areas under the PSD's $S_{xc}(w)$, $S_{xs}(w)$ and $S_x(w)$ are all equal. Hence if $x(t)$ is a deterministic signal, we have

$$S_c = S_s = S_x$$

where S_c and S_s and S_x represent the power (or energy, if applicable) of the corresponding signals. If $x(t)$ is random, we can write

$$\langle x_c^2(t)\rangle = \langle x_s^2(t)\rangle = \langle x^2(t)\rangle = \sigma^2$$

For the case of random signals, it is also possible to show that

$$\langle x_c(t)x_s(t)\rangle = R_{x_cx_s}(0) = 0$$

and, in fact, in general if $S_x(w)$ is symmetrical about w_c, then

$$R_{x_cx_s}(\tau) = 0$$

In particular when x(t) is a Gaussian random process with variance σ^2, it can be shown that both $x_c(t)$ and $x_s(t)$ are also Gaussian with the same variance. Thus

$$p_{x_c}(x) = p_{x_s}(x) = p_x(x) = \sqrt{(1/2\pi\sigma^2)}\exp(-x^2/2\sigma^2)$$

Similarly, in the envelope-phase representation, it can be shown that the probability density function of e(t) and θ(t) are Rayleigh and Uniform, respectively:

$$p_e(e) = (e/\sigma^2)\exp[-r^2/2\sigma^2] \qquad r > 0$$

and

$$p_\theta(\theta) = (1/2\pi), \qquad -\pi \le \theta < \pi$$

III. Hilbert Transformation and Complex Analytic Representation:

Hilbert Transformation and Its Properties:

For a real valued function x(t), its Hilbert transform, denoted by $\hat{x}(t)$ is defined by

$$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau$$

The inverse Hilbert transform is given by

$$x(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\hat{x}(\tau)}{t-\tau}d\tau$$

If X(f) represents the Fourier transform of x(t), then from elementary properties of the Fourier Transform, it follows that

$$\hat{X}(f) = -jX(f)\mathrm{sgn}(f)$$

where sgn(f) is the signum function given by

$$\mathrm{sgn}(f) = \begin{cases} 1 & \text{for } f > 0 \\ 0 & \text{for } f = 0 \\ -1 & \text{for } f < 0 \end{cases}$$

The Hilbert transformation, can therefore be interpreted as a phase shifter such that all the frequency components of the input signal are phase shifted by 90° or $\pi/2$ radians. In particular, if the input signal is $\cos(wt + \theta)$, then the output signal would be $\sin(wt + \theta)$.

It can also be shown that if x(t) is a sample function of a wide-sense stationary random process having an ensemble autocorrelation function $R_x(\tau)$, then

$$R_{\hat{x}}(\tau) = R_x(\tau)$$

Complex Envelope and Analytic Signal Representation:

From the quadrature representation discussed above, it follows that a real signal x(t) can be represented also as

$$x(t) = \mathrm{Re}\ \tilde{x}(t)\exp(jw_ct)$$

where $\tilde{x}(t)$ represents a so-called complex envelope. For narrowband signals, the complex envelope is a slowly changing (or lowpass) function of time. (For the quadrature representation used above, we have $\tilde{x}(t) = x_c(t) + j\,x_s(t)$). It is easy to see that the complex signal $\hat{x}(t)$ has no negative frequency components.

The pre-envelope or analytic representation of a general (not necessarily narrowband) real signal is defined by

$$x_p(t) = x(t) + j\hat{x}(t)$$

whereby the real signal is the real part of the pre-envelope. The envelope of x(t) is the absolute value $|x_p(t)|$ of the pre-envelope.

From the properties of the Hilbert transform ,it is easy to show that the Fourier transform of $f_p(t)$ is

$$F_p(w) = \begin{cases} 2F(jw) & w > 0 \\ F(jw) & w = 0 \\ 0 & w < 0 \end{cases}$$

This representation is general and valid for all types of real signals. In order to verify its validity for narrowband signals in particular, consider the signal

$$x(t) = x_c(t)\cos w_c t + x_s(t)\sin w_c t$$

Its Hilbert transform is given by

$$\hat{x}(t) = x_c(t)\sin w_c t - x_s(t)\cos w_c t$$

It is easy to verify that the pre-envelope is given by

$$x_p(t) = |x_c(t) + jx_s(t)|\exp(jw_c t)$$

and the envelope is therefore

$$e(t) = |x_p(t)| = [x_c^2(t) + x_s^2(t)]^{1/2}$$

PROBABILITY DISTRIBUTIONS

I.Introduction:

The most important probability density function in Communication Theory is the Gaussian or normal density, and the most commonly encountered random process is the Gaussian random process. There are many reasons for this including the well known central limit theory of probability theory. Sometimes, the Gaussian assumption is used even not exactly justifiable. This is done because of its analytic tractability.

A number of other Gaussian related distributions also arise naturally in many applications when Gaussian processes or random variables are processed via linear or nonlinear operations. Here we discuss a few of such important pdf's.

II. Distribution of the Envelope of a Narrowband Gaussian Process:

As already discussed in the chapter on narrowband signals, if n(t) given by

$$n(t) = x(t)\cos w_0 t + y(t)\sin w_c t$$

is a Gaussian process, then its envelope $e(t) = (x^2(t)+y^2(t))^{1/2}$ has a Rayleigh distribution.

Thus, we have

$$p(x,y) = (1/2\pi\sigma^2)\exp[-(x^2+y^2)/2\sigma^2]$$

and

$$p(e) = (e/\sigma^2)\exp[-e^2/2\sigma^2]$$

The normalized distribution for $v = e/\sigma$ is shown in Fig. 31. The important parameters of a Rayleigh distribution are given below:

$$E\{e\} = (\pi/2)^{1/2}\sigma \quad \text{and} \quad E\{e^2\} = 2\sigma^2$$

Also, as discussed earlier, the phase of the narrowband random process has a uniform distribution, and the envelope and phase are independent random variables.

III. Envelope of a Sine Wave Plus Narrowband Noise:

Here we are interested in the pdf of the envelope of the random signal

$$r(t) = A\cos(w_c t + \theta) + n(t)$$

which is often encountered in the detection of signals in analogue and digital communication. Here the envelope can be seen to be given by

$$e(t) = \{[A\cos\theta + x(t)]^2 + [A\sin\theta + y(t)]^2\}^{1/2}$$

Through some tedious but straightforward manipulations, it can be seen that the pdf of e is given by

$$p(e) = (e/\sigma^2)\exp[-(1/2\sigma^2)(e^2+A^2)]I_0(eA/\sigma^2)$$

This density function is referred to as Rician, and also as generalized Rayleigh sometimes. As the amplitude A of the sine wave approaches zero, the density function tends to the Rayleigh function, as might be expected. It is convenient to work with a normalised variable $v = e/\sigma$ and a normalised constant $\alpha = A/\sigma$, to yield the normalised Rician density function:

$$p(v) = v\exp[-(v^2+\alpha^2)/2]I_0(\alpha v)$$

This is also plotted in Fig. 31 for different values of α.

For large values of α, the Rician density function becomes approximately Gaussian.

IV. Decision Threshold and Decision Probabilities:

In digital communication, radar and sonar applications, the receiver is required either to detect the presence or absence of a signal or decide which one of the several possible signals is present. There is some uncertainty in making these decisions correctly due to the presence of noise along with the desired signal at the receiver input. The performance of such receivers is measured in terms of some measure of the error probability in making these decisions.

A typical receiver configuration required to make such a decision is shown in Fig.32. It is comprised of the usual front modulator, a matched filter or correlator, a sampler and a decision device. The decision device works on an input random variable whose pdf depends on whether it contains signal-plus-noise or noise alone. In the case of digital communication systems, it also depends on which of the several possible signals might have been transmitted.

For the case of binary decisions, the decision device is essentially a threshold device, the output y taking one of two possible values (say 0 or 1) depending on whether the input variable x is smaller or larger than a previously selected threshold. The choice of the threshold is crucial in determining the optimum performance of the receiver. In the following discussion, we shall refer to the choices 0 and 1 as the hypotheses H_0 and H_1, respectively, and the corresponding decisions made as D_0 and D_1, respectively.

Fig. 33 shows the decision process in terms of the probability density functions $p_0(x)$ and $p_1(x)$ of x for the case of binary decisions. Since x may take on any value in the interval $(-\infty, \infty)$, the decision process is concerned with the choice of threshold say T, such that the real line is divided into two parts. A sample point (i.e., the observed value of x) which falls in a region R_0 results in the decision H_0 being chosen, and a point which falls in region R_1 results in H_1 being chosen.

There are two types of error which can be made. If we say a signal is present when in fact it is not, an error of the first kind is made. That is, we choose H_1, when in fact, H_0 is true. The probability of this event is denoted as $P(D_1|H_0)$ and is represented by area 1 in Fig.33. In radar and sonar terminology, this is called a false alarm error, equivalent to saying that a target is present when there is none.

If, on the other hand, H_0 is chosen when in fact H_1 is true, an error of the second kind is made. This probability, denoted by $P(D_0|H_1)$, is shown as area 2 in the figure. Alternatively, we define $1 - P(D_0|H_1) = P(D_1|H_1)$ as the probability of detection.

Clearly, each of these probabilities can be calculated in terms of appropriate areas of the two density functions $p_0(x)$ and $p_1(x)$. The average probability of error, denoted by P_e, is

$$P_e = P(D_1|H_0)P(H_0) + P(D_0|H_1)[1 - P(H_0)]$$

In communication problems, it is frequently assumed that $P(H_1) = P(H_0) = (1/2)$. It can be shown that in this case the value of the threshold T to determine the regions 1 and 2 is obtained by ensuring that

$$P(D_1|H_0) = P(D_0|H_1)$$

so that the overall probability of error is minimised.

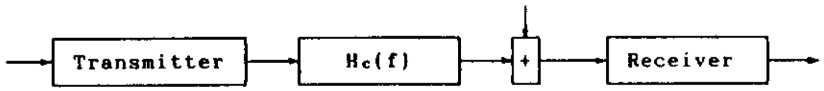

Fig. 1: Block diagram of a communication system

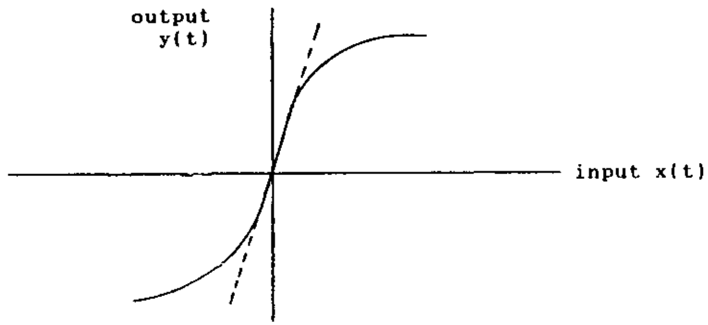

Fig. 2: Saturating nonlinearity

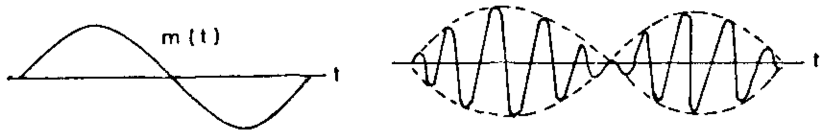

Fig.3: Double Sideband Modulation

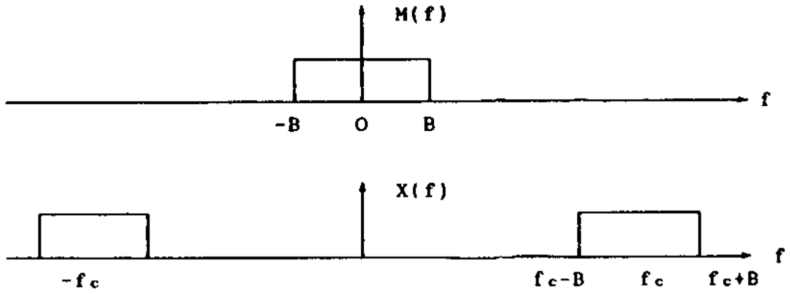

Fig.4: Spectrum of DSB-SC Signal

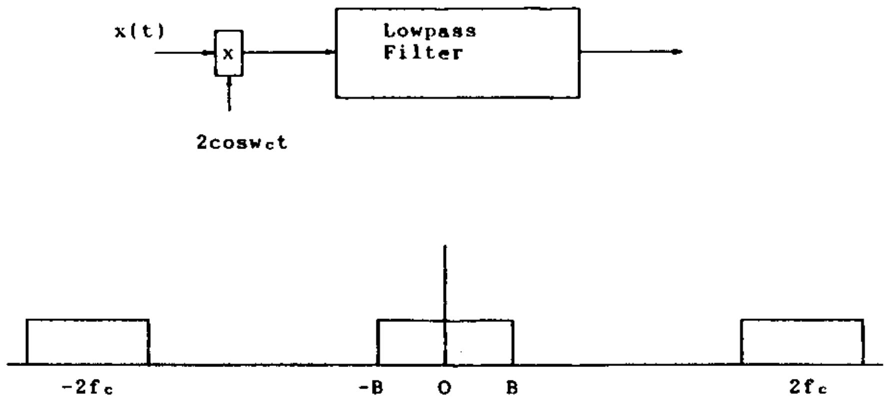

Fig.5: Synchronous Demodulation

Fig.6: A squaring synchroniser

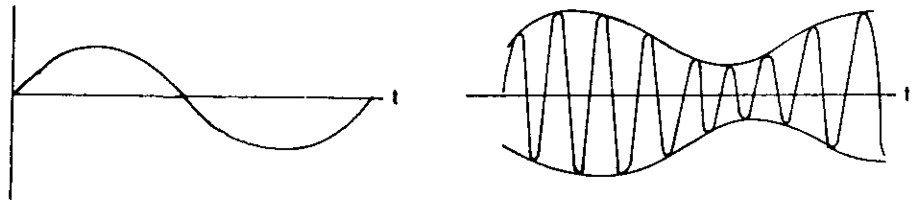

Fig.7: AM Waveforms

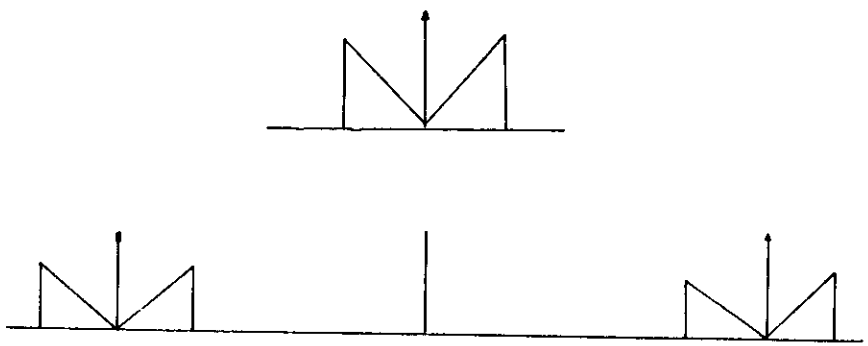

Fig.8: Spectra of Message and Modulated Signals

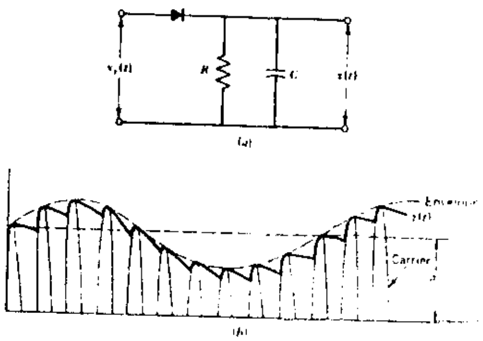

Fig.9: Envelope Demodulation of AM Signals

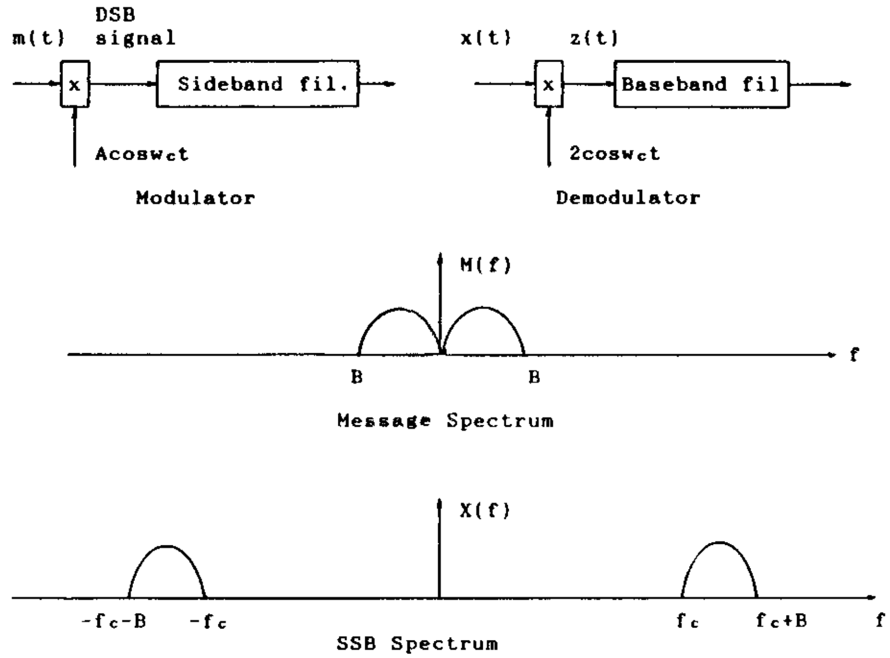

Fig.10: Single Sideband Modulation

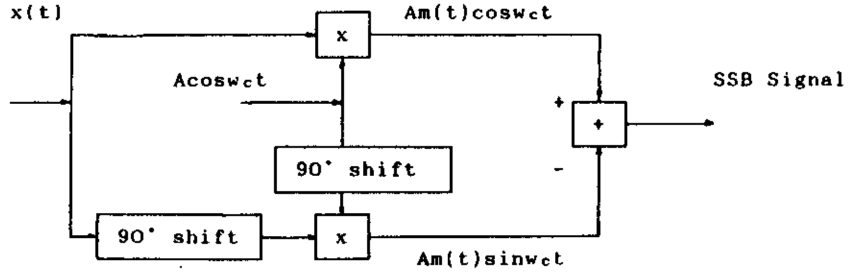

Fig.11: Phase-shift SSB Modulator

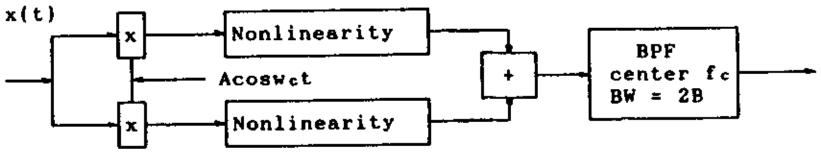

Fig.12: Balanced Modulator

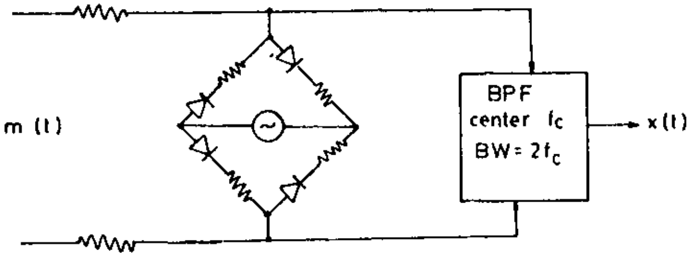

Fig.13: A Switching Modulator

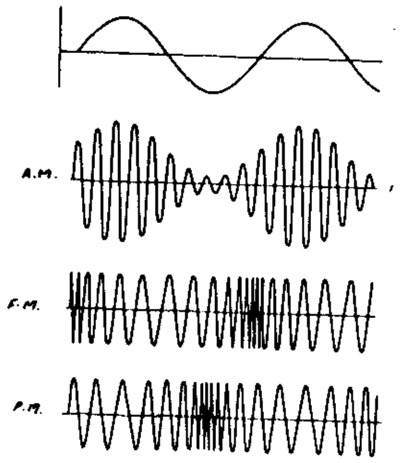

Fig.14: AM and FM Waveforms

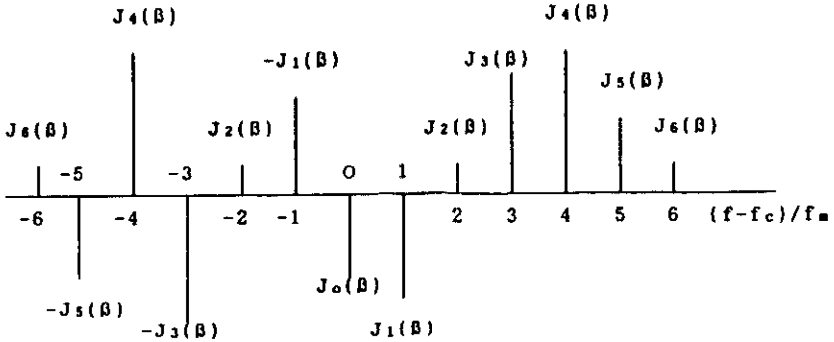

Fig.15: Spectrum of an FM Signal

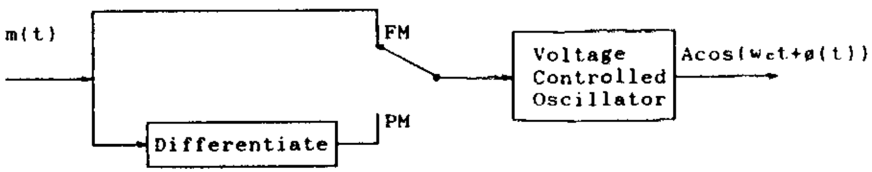

Fig.16: Direct method of FM Generation

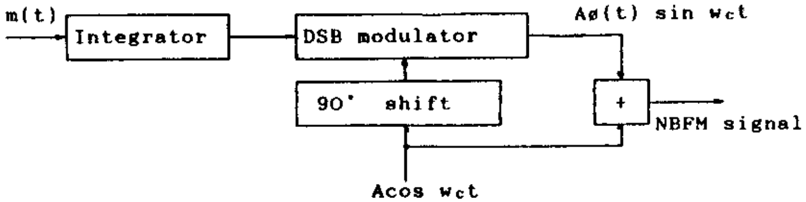

Fig.17: Generation of NBFM signal

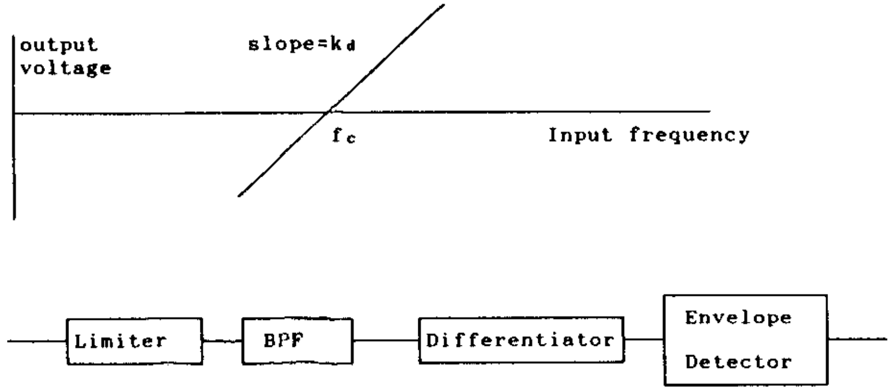

Fig.18: FM demodulation

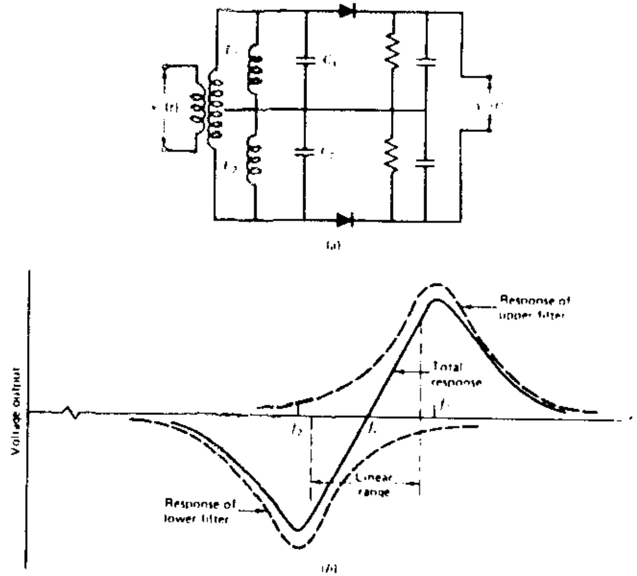

Fig.19: A Practical Discriminator

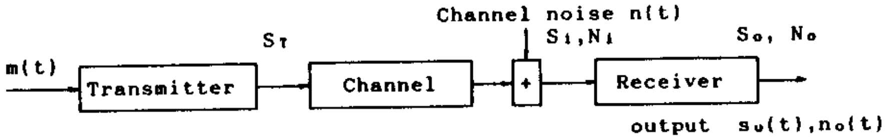

Fig. 20: Noise in a Communication System

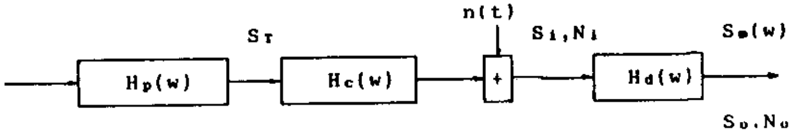

Fig.21: Noise Analysis Model for Baseband Systems

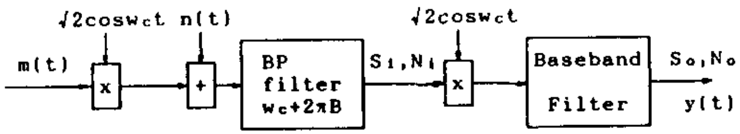

Fig.22: Noise Analysis Model for DSB-SC Systems

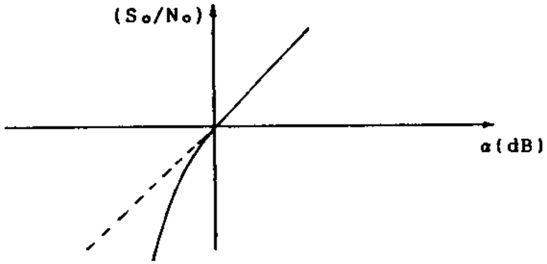

Fig.23: Threshold Effect in Envelope Detection

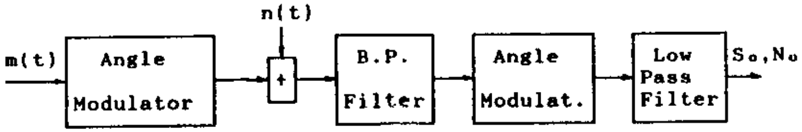

Fig.24: Noise analysis model of an angle modulated system

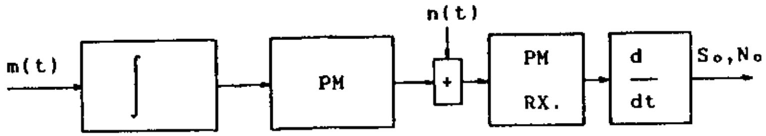

Fig.25: Model for FM Demodulator

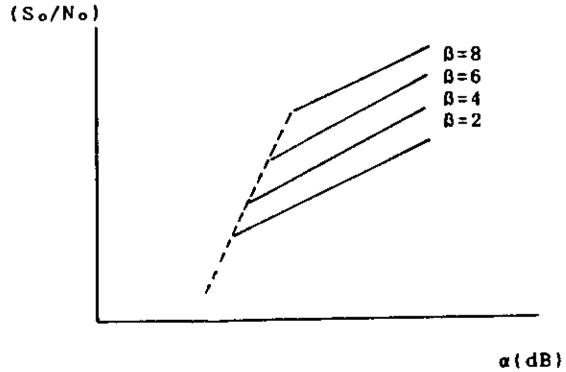

Fig.26: Performance Curves for FM

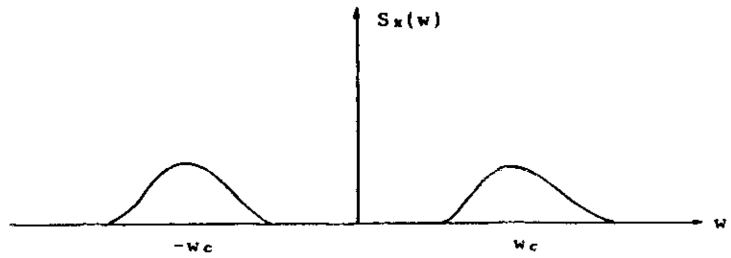

Fig.28: Power Spectrum of a Bandpass Signal

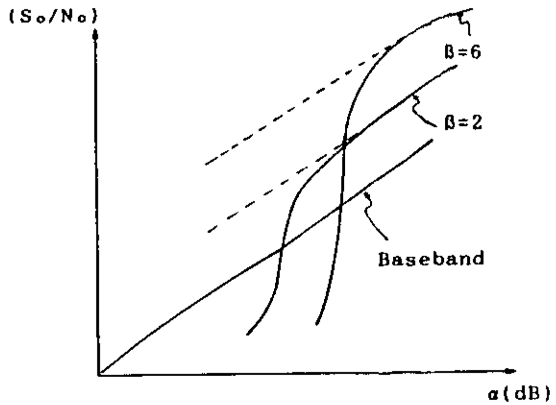

Fig.27: Threshold Effect in FM

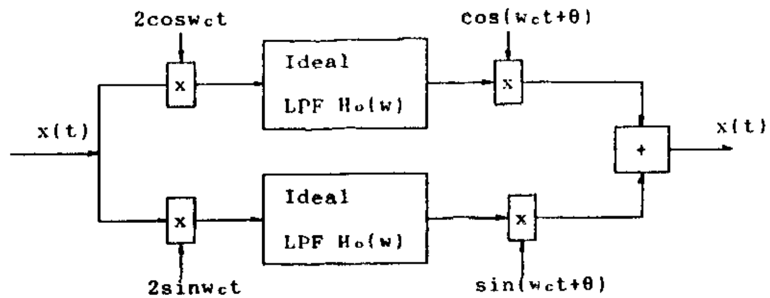

Fig.29: A bandpass system

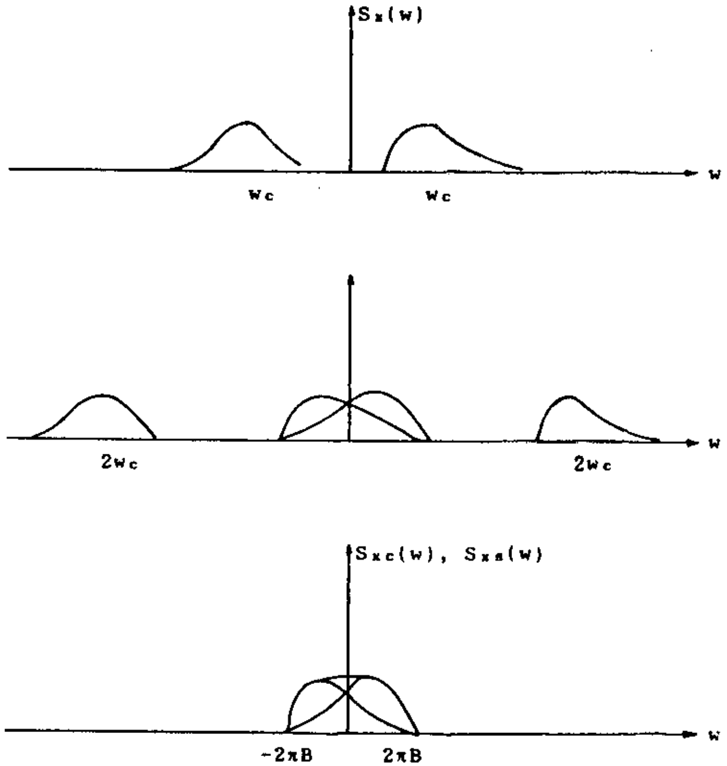

Fig.30: Power Spectra of $x_c(t)$ and $x_s(t)$

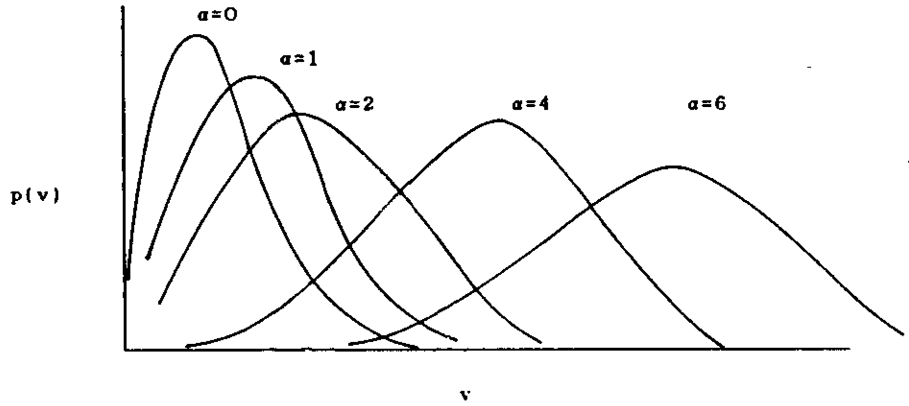

Fig.31: pdf's of envelopes of sinewave plus noise

Fig.32 A Binary Decision Receiver

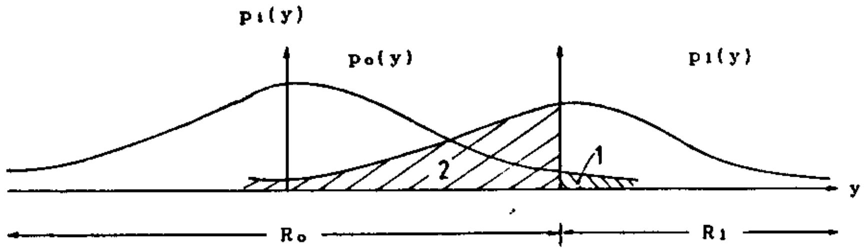

Fig.33: pdf's and Associated Decision Regions