INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
INTERNATIONAL CENTRE FOR SCIENCE AND HIGH TECHNOLOGY
c/o INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS 34100 TRIESTE (ITALY) VIA GRIGNANO, 9 (ADRIATICO PALACE) P.O. BOX 586 TELEPHONE 040-22572 TELEFAX 040-224575 TELEX 460449 APH I

H4.SMR/537-18

SECOND COLLEGE ON THEORETICAL AND EXPERIMENTAL RADIOPROPAGATION PHYSICS
(7 January - 1 February 1991)

Co-sponsored by ICTP, URSI, ICSU and with the participation of ICS

ANALOG MODULATION

S. C. Dutta Roy
Indian Institute of Technology, Delhi
New Delhi, India

ANALOG MODULATION

· Notes prepared by Professor S Prasad of the Indian Institute of Technology, New Delhi, India and used, with permission, by Professor S.C. Dutta Roy, for the Second College on Radio Propagation Physics (ICTP/URSI sponsored) at the ICTP, 7 Jan - 1 Feb 1991.

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

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

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 $\theta = 0$, $y(t) = km(t)\cos \delta wt$ 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

$$\begin{aligned} y(t) &= a_1[A \cos w_c t + m(t)]^2 + a_2[A \cos w_c t + m(t)]^2 \\ &\quad - a_1[A \cos w_c t - m(t)]^2 - a_2[A \cos w_c t - m(t)]^2 \\ &= 2a_1 m(t) + 4a_2 m(t) A \cos w_c t \end{aligned}$$

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

$$v(t) = k_0 m(t) + k_1 m(t) \cos w_c t + k_3 m(t) \cos 3w_c t + \ldots$$

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

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

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 $\beta \ll 1$ (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.

<u>Bandwidth:</u> 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 <u>voltage controlled oscillator</u> 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:

$$
\begin{aligned}
x(t) &= A\cos(w_c t + \phi(t)) \\
&= A\cos w_c t\cos\phi(t) - A\sin w_c t\sin\phi(t) \\
&\approx A\cos w_c t - A\phi(t)\sin w_c t
\end{aligned}
$$

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

$$e_1(t) = A\cos(w_c t + \beta\sin w_m t)$$

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.

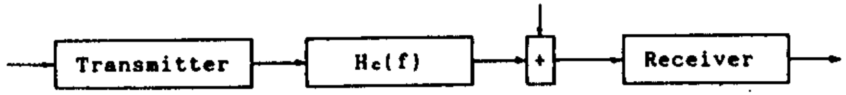

Fig. 1: Block diagram of a communication system

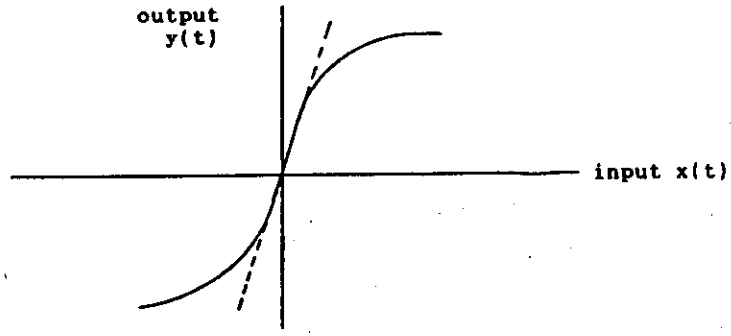

Fig. 2: Saturating nonlinearity

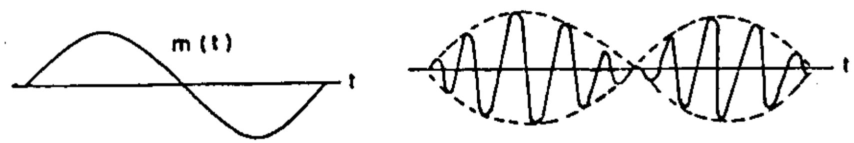

Fig.3: Double Sideband Modulation

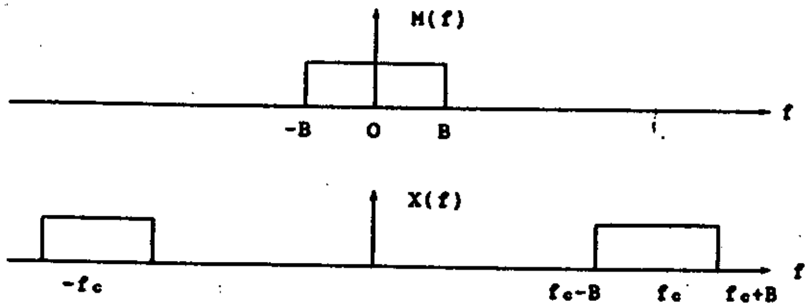

Fig.4: Spectrum of DSB-SC Signal

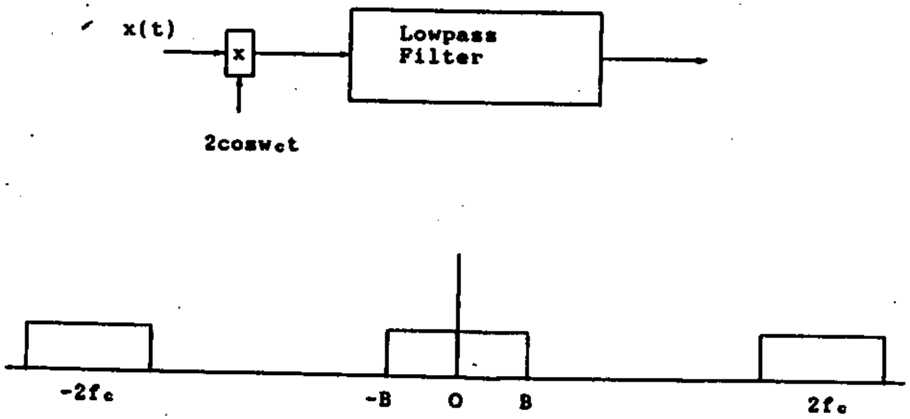

Fig.5: Synchronous Demodulation

Fig.6: A squaring synchroniser

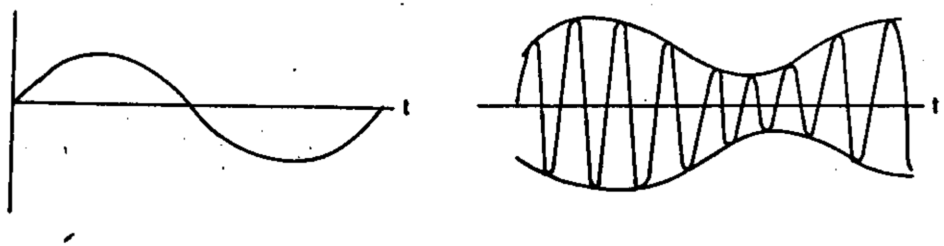

Fig.7: AM Waveforms

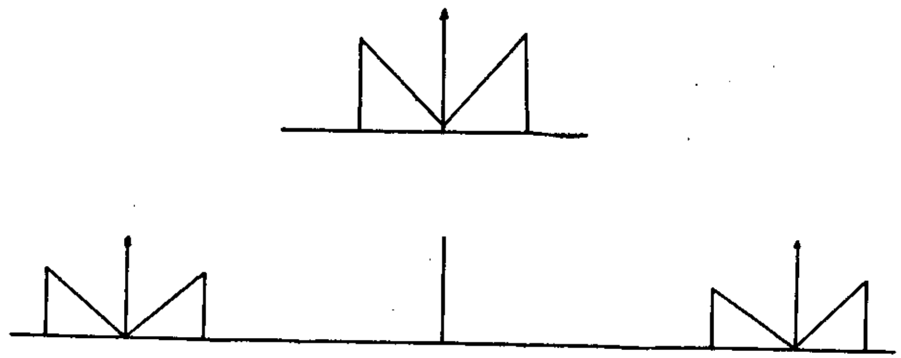

Fig.8: Spectra of Message and Modulated Signals

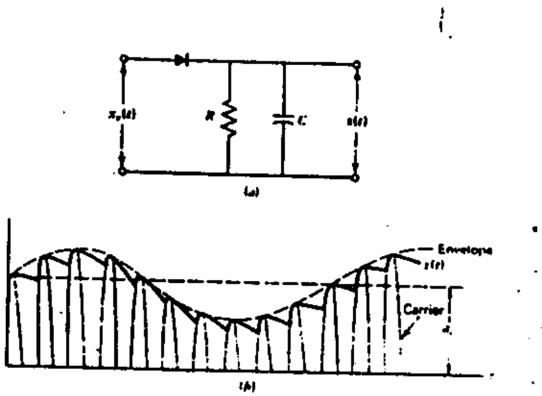

Fig.9: Envelope Demodulation of AM Signals

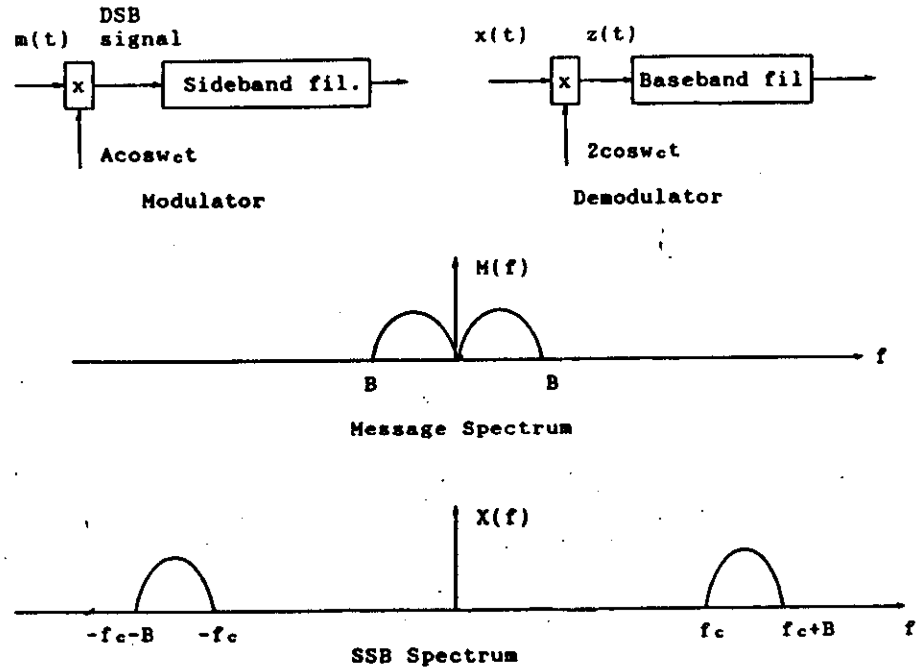

Fig.10: Single Sideband Modulation

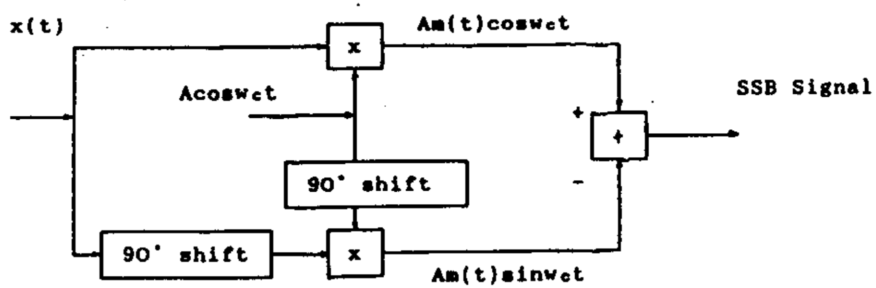

Fig.11: Phase-shift SSB Modulator

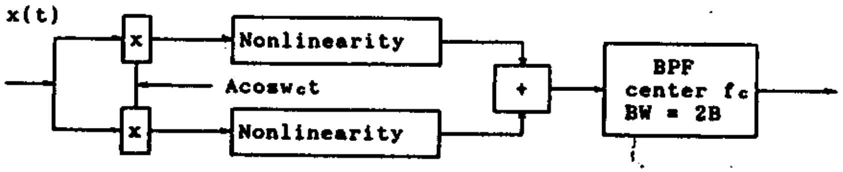

Fig.12: Balanced Modulator

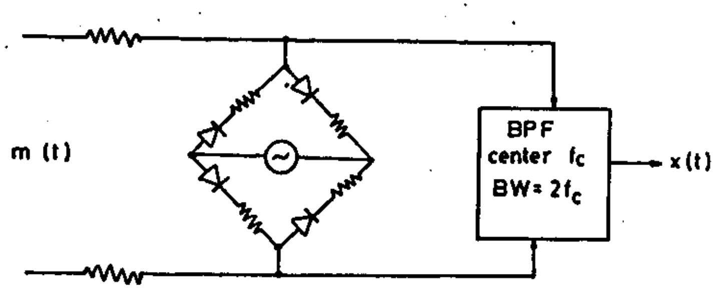

Fig.13: A Switching Modulator

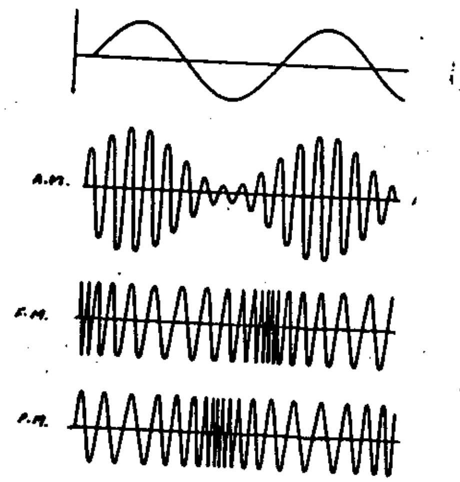

Fig.14: AM and FM Waveforms

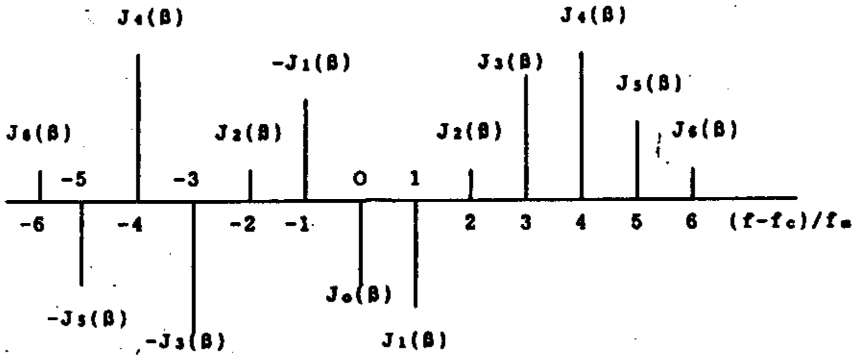

Fig.15: Spectrum of an FM Signal

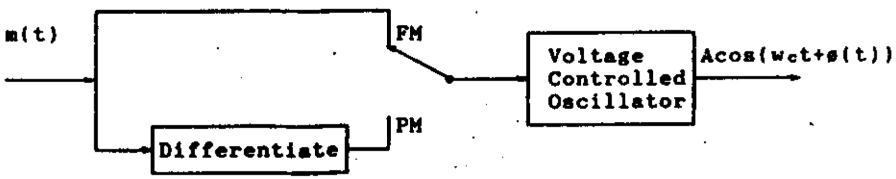

Fig.16: Direct method of FM Generation

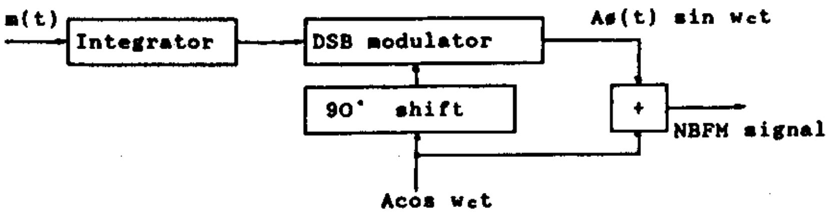

Fig.17: Generation of NBFM signal

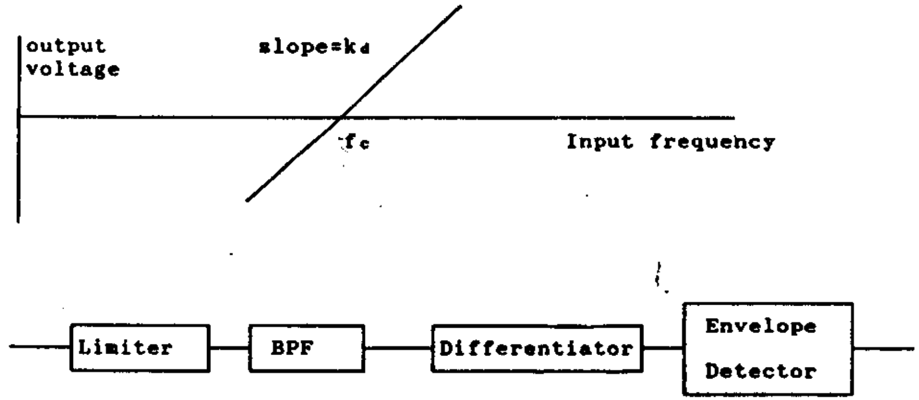

Fig.18: FM demodulation

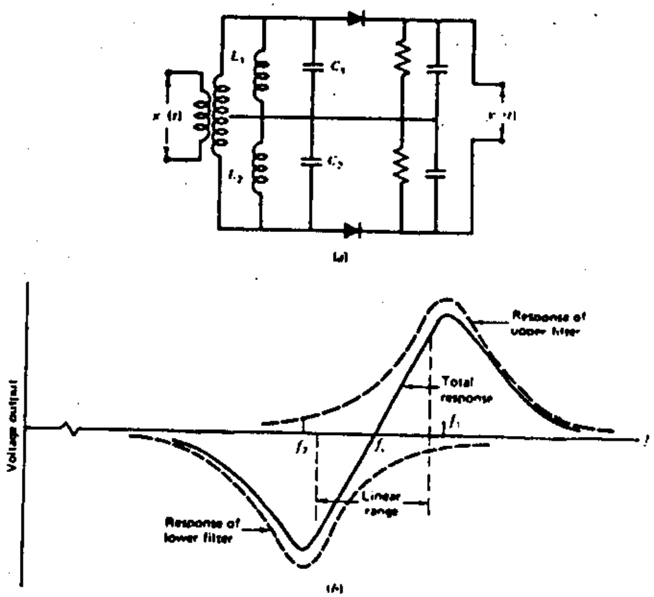

Fig.19: A Practical Discriminator