Module 7:
AM, FM, and the spectrum analyzer.

7.0 Introduction

Electromagnetic signals may be used to transmit information very quickly, over great distances. Two common methods by which information is encoded on radio signals, amplitude and frequency modulation, will be reviewed in this module. Also, the process of retrieving information from encoded signals will be discussed. Finally, the basic components of the spectrum analyzer will be examined.

7.1 Modulation

With the proper equipment, radio signals can be transmitted and received over large distances. Information may therefore be exchanged over large distances by encoding information on radio waves. This is accomplished through *modulation* of radio signals.

– Modulation is the process of encoding information onto a carrier signal which has frequency f_c. This carrier signal is called the *modulated signal*, while the information carrying, or *baseband* signal is referred to as the *modulating* signal.

Two types of modulation will be reviewed in this module.

– *Amplitude modulation* consists encoding information onto a carrier signal by varying the amplitude of the carrier.

– *Frequency modulation* consists of encoding information onto a carrier signal by varying the frequency of the carrier.

Once a signal has been modulated, information is retrieved through a demodulation process.

7.2 Suppressed carrier amplitude modulation (double sideband)

A general sinusoidal signal can be expressed as

$$f(t) = A(t)\cos\theta(t) .$$

where the amplitude A and phase angle θ may, in general, be functions of time. It is convenient to write time varying angle $\theta(t)$ as $\omega_c t + \varphi(t)$, therefore the sinusoidal signal may be expressed as

$$f(t) = A(t)\cos\left[\omega_c t + \varphi(t)\right].$$

The term $A(t)$ is called the *envelope* of the signal $f(t)$, and ω_c is called the carrier frequency.

The process of amplitude modulation consists of the amplitude of the carrier wave being varied in sympathy with a modulating signal. A mathematical representation of an amplitude modulated signal is obtained by setting $\varphi(t)=0$ in the expression for the general sinusoidal

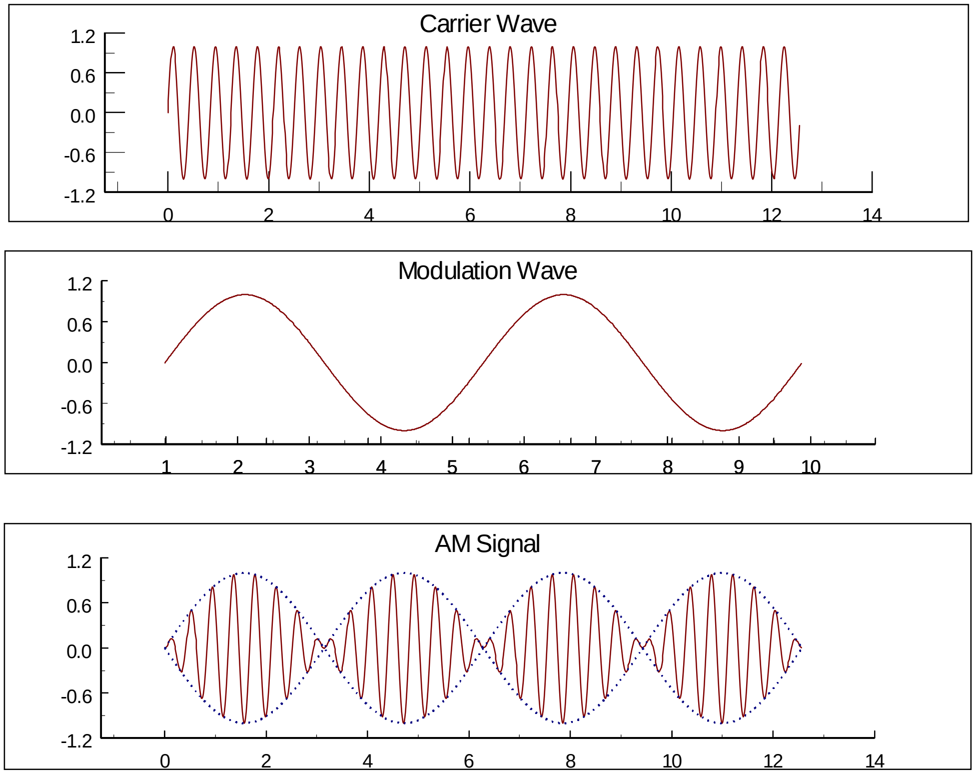

Figure 1. Amplitude modulation.

signal, and letting the envelope $A(t)$ be proportional to a modulating signal $f(t)$. What results is a new (modulated)signal, given by

$$y(t) = f(t)\cos(\omega_c t)\,.$$

The spectrum of the modulated signal $y(t)$ can be found by using the modulation property of the Fourier transform. In Chapter 3, the Fourier transform pair was defined as

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\, e^{j\omega t}\, d\omega$$

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt\,.$$

The Fourier transform of a signal $f(t)e^{j\omega_0 t}$ is then

$$\mathscr{F}\left\{f(t)e^{j\omega_0 t}\right\} = \int_{-\infty}^{\infty} f(t)\, e^{j\omega_0 t} e^{-j\omega t}\, dt$$

$$= \int_{-\infty}^{\infty} f(t)\, e^{-j(\omega - \omega_0)t}\, dt\,.$$

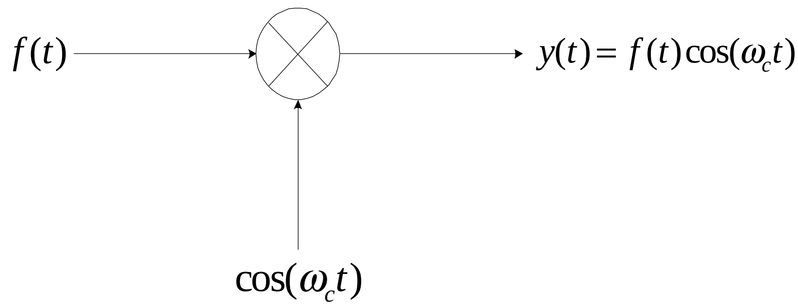

Figure 2: Amplitude modulation (suppressed carrier)

Thus the Fourier transform of $f(t)e^{j\omega_0 t}$ may be expressed

$$\mathscr{F}\left\{f(t)e^{j\omega_0 t}\right\} = F(\omega - \omega_0).$$

The amplitude modulated signal $y(t)$ may be written in terms of complex exponentials

$$y(t) = f(t)\cos\omega_c t = \frac{1}{2}f(t)\left[e^{j\omega_c t} + e^{-j\omega_c t}\right].$$

When $y(t)$ is expressed in this form, and from the example above, it can be seen that the Fourier transform of $y(t)$ is given by

$$\mathscr{F}\left\{f(t)\cos(\omega_c t)\right\} = \frac{1}{2}\left[F(\omega + \omega_c) + F(\omega - \omega_c)\right].$$

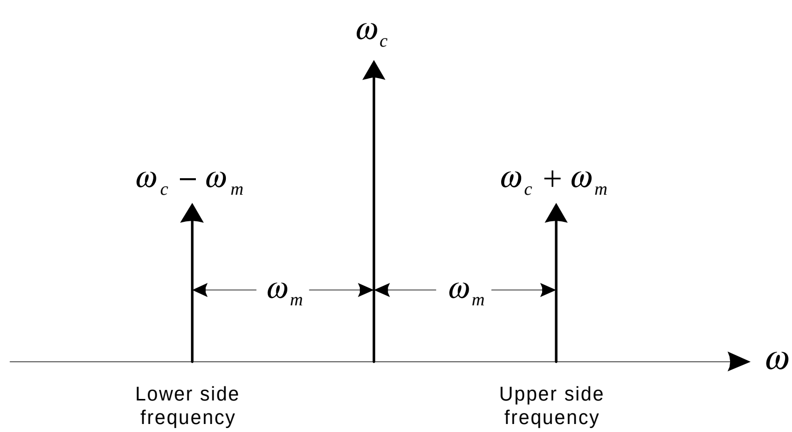

Figure 3. Single modulating frequency AM signal spectrum.

Thus, the spectrum of $f(t)$ is translated by $\pm\omega_c$. It is seen that the modulation process causes the frequencies associated with the modulating signal to disappear. Instead, a new frequency spectrum appears, consisting of two sidebands, known as the upper sideband (USB), and the lower sideband (LSB). The spectrum of the modulated signal $y(t)$ does not contain the spectrum of the original carrier, but is still centered about the carrier frequency ω_c. Thus this type of modulation is referred to as *double-sideband, suppressed-carrier* amplitude modulation. A block diagram of the suppressed carrier amplitude modulation operation is presented in Figure 2.

If the modulating signal contains a single frequency ω_m, then $\omega_{USB}=\omega_c+\omega_m$, and $\omega_{LSB}=\omega_c - \omega_m$ (Figure 3). If modulating signal *f*(*t*) has a bandwidth of ω_{bw} , then F(ω), the spectrum of *f*(*t*), will extend from $-\omega_{bw}$ to $+\omega_{bw}$. The upper sideband of the spectrum of the modulated signal Y(ω) will extend from ω_c to $\omega_c+\omega_{bw}$. Likewise, the lower sideband will extend from $\omega_c-\omega_{bw}$ to ω_c . Both the negative and positive frequency components of the modulating signal *f*(*t*) appear as positive frequencies in the spectrum of the modulated signal *y*(*t*). It is also seen that the bandwidth of *f*(*t*) is doubled in the spectrum of the modulated signal when this type of modulation is employed.

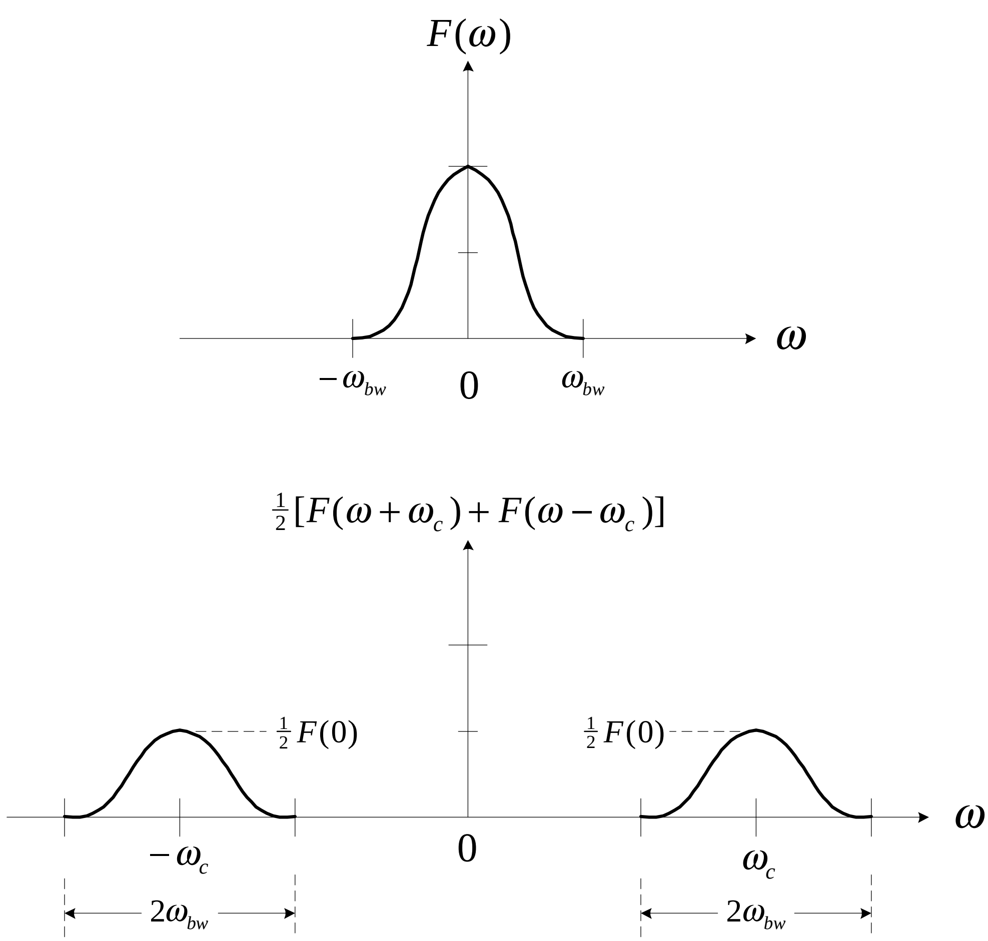

Figure 4. Spectra of modulating wave and resulting AM waveform

7.3 AM demodulation

An AM signal is demodulated by first mixing the modulated signal $y(t)$ with another sinusoid of the same carrier frequency

$$y(t)\cos(\omega_c t) = f(t)\cos^2(\omega_c t) = \tfrac{1}{2} f(t)\big(1+\cos(2\omega_c t)\big).$$

The Fourier transform of this signal is

$$\mathscr{F}\{\,y(t)\cos(\omega_c t)\} = \mathscr{F}\left\{\frac{1}{2} f(t)\big(1+\cos(2\omega_c t)\big)\right\}$$

or

$$\mathscr{F}\{y(t)\cos(\omega_c t)\} = \frac{1}{2}\left\{F(\omega) + \frac{1}{2}\big[F(\omega+2\omega_c)+F(\omega-2\omega_c)\big]\right\}.$$

By using a low-pass filter, the frequency components centered at $\pm 2\omega_c$ can be removed to leave only the $1/2\,F(\omega)$ term. It is obvious that in order to properly recover the original signal it is necessary that $\omega_c > \omega_{bw}$. A block diagram of the demodulation process is shown in Figure 5.

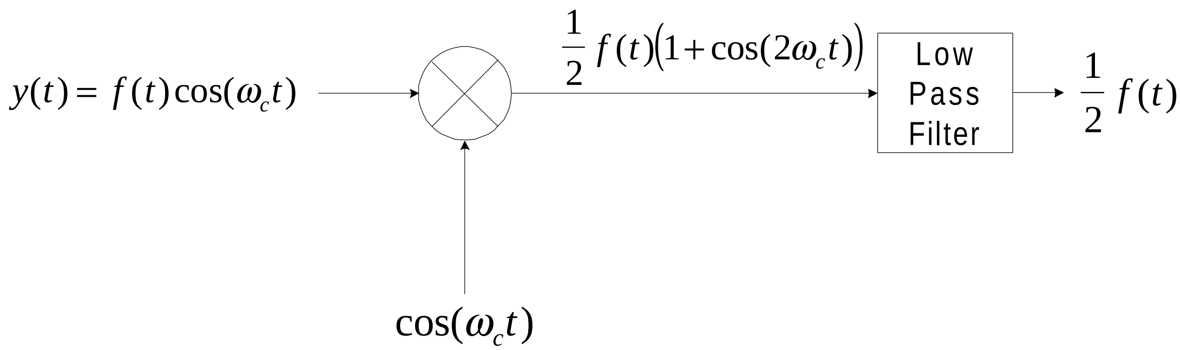

Figure 5: AM demodulation.

7.4 Large carrier amplitude modulation (double sideband)

In practice, the demodulation of suppressed carrier amplitude modulated signals requires fairly complicated circuitry in order to acquire and maintain phase synchronization. A much less complicated (and thus less expensive) receiver can be used if a slightly different modulation scheme is employed.

In *large carrier amplitude modulation*, the carrier wave information is incorporated as a part of the waveform being transmitted. It is convenient to let the amplitude of the carrier be larger than any other part of the signal spectral density. While this makes the demodulation process much easier, low-frequency response of the system is lost. For some signals however, frequency response down to zero is not needed (such as in audio signals).

Consider a carrier wave with amplitude A, and frequency ω_c, represented by

$$c(t) = A\cos(\omega_c t) .$$

The modulated waveform of a large carrier AM signal can be then be expressed mathematically as

$$y(t) = f(t)\cos(\omega_c t) + A\cos(\omega_c t) .$$

The spectrum associated with this modulated signal is given by

$$Y(\omega) = \frac{1}{2}F(\omega + \omega_c) + \frac{1}{2}F(\omega - \omega_c) + \pi A\delta(\omega + \omega_c) + \pi A\delta(\omega - \omega_c) .$$

It is seen that the spectrum of the large carrier AM signal is the same as that of the suppressed carrier AM signal with the addition of impulses at $\pm\omega_c$.

The large carrier AM signal may be rewritten

$$y(t) = \left[A + f(t)\right]\cos(\omega_c t) .$$

In this form, $y(t)$ may be thought of as consisting of a carrier signal $\cos(\omega_c t)$ having amplitude $\left[A + f(t)\right]$. If the amplitude of the carrier A is sufficiently large, then the envelope of the modulated waveform will be proportional to $f(t)$ (hence the name "large carrier" AM). Demodulation in this case is simply involves the detection of the envelope of a sinusoid.

Figure 6. Comparison of large carrier AM and suppressed carrier AM.

7.6 The envelope detector

An envelope detector is any circuit whose output follows the envelope of an input signal. The simplest form of such a detector is a non-linear charging circuit which has a fast charge time and a slow discharge time. This is easily implemented by placing a diode in series with a parallel combination of a capacitor and a resistor. The envelope of an input signal is detected by the following process:

- The input waveform (in this case a large carrier AM signal) charges the capacitor to the maximum value of the waveform during positive half-cycles of the input signal.

- As the input signal falls below maximum, the diode becomes reverse biased, and switches off.

- The capacitor then begins a relatively slow discharge through the resistor until the next positive half-cycle.

- When the input signal becomes greater than the capacitor voltage, the diode becomes forward biased, and the capacitor charges to a new peak value.

For optimum operation, the discharge time constant *RC* is adjusted so that the maximum negative rate of the envelope never exceeds the exponential discharge rate.

- If the time constant is too large, the envelope detector may miss some positive half-cycles of the carrier, and will not correctly reproduce the envelope of the input signal.

- If the time constant is too small, the detector generates a ragged signal.

Figure 7. Envelope detector.

7.5 Frequency modulation

Frequency modulation is a type of *angle* modulation. Angle modulation changes the phase of a signal as well as its amplitude, where amplitude modulation leaves the phase unchanged. Phase modulation is another type of angle modulation very similar to frequency modulation.

The general form of a sinusoidal signal can be written as

$$f(t) = A(t)\cos[\theta(t)].$$

The instantaneous angular frequency of this signal, $\omega_i(t)$, is the derivative of the phase

$$\omega_i(t) = \frac{d\theta(t)}{dt}.$$

This definition of instantaneous frequency suggests two obvious methods of angle modulation:

– If the phase angle of a carrier with fundamental frequency ω_c is varied linearly with a modulating waveform $f(t)$

$$\theta(t) = \omega_c t + k_p f(t) + \theta_0$$

The modulated signal is said to be *phase modulated*. Here ω_c, k_p, and θ_0 are constants. The instantaneous frequency of a phase modulated signal is given by

$$\omega_i = \frac{d\theta}{dt} = \omega_c + k_p \frac{df}{dt}.$$

– If the instantaneous frequency of a carrier with fundamental frequency ω_c is varied in proportion to an input modulating signal $f(t)$, such that

$$\omega_i = \omega_c + k_f f(t)$$

the resulting modulated signal is said to be *frequency modulated*. Here ω_c and k_f, are constants. The phase angle associated with the FM signal is given by

$$\theta(t) = \int_0^t \omega_i(\tau)\,d\tau$$

which may also be expressed

$$\theta(t) = \omega_c t + \int_0^t k_f f(\tau)\,d\tau + \theta_0.$$

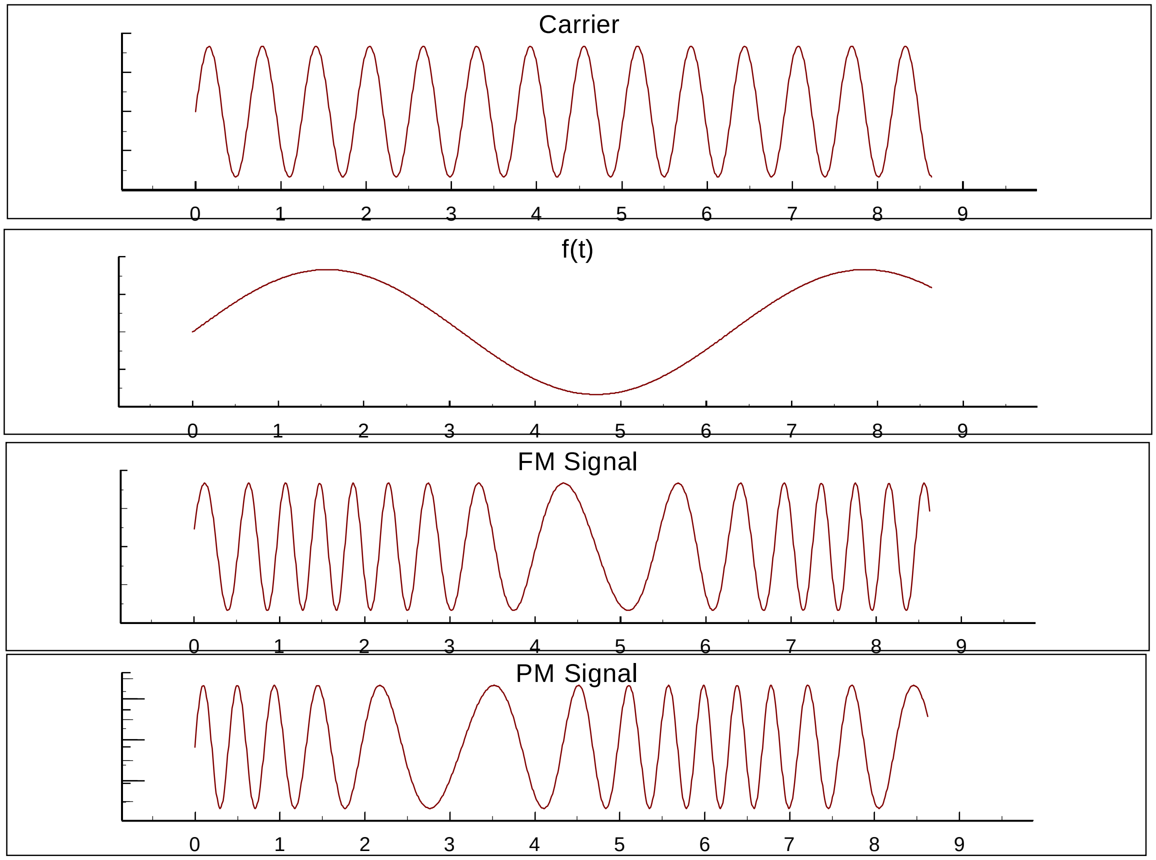

Figure 8. FM and PM signals

It can be seen in Figure 8 that phase and frequency modulation are closely related. For both phase and frequency modulation, the modulating signal causes the carrier to increase and decrease from the fundamental frequency ω_c, while the amplitude remains constant. For the FM signal, the frequency rises with positive modulating amplitudes and falls with negative modulating amplitudes. For the PM signal, the frequency rises with increasing modulating amplitudes, and decreases with decreasing modulating amplitudes.

Frequency modulation has several advantages over amplitude modulation:

– The radiated signal level remains constant, therefore transmitters can be run at a constant power output.

– Amplitude variations due to external interference sources are not interpreted as signals.

– Selective fading does not occur because the amplitude of the carrier is constant.

– It is possible to design systems having better dynamic range and signal-to-noise ratio.

The obvious disadvantage of frequency modulation is that a greater bandwidth is required than for amplitude modulation.

Another significant difference between amplitude and angle modulation has to do with the relationship between the modulated and modulating signals. When signals are amplitude modulated, a one-to-one correspondence exists between the modulated and modulating signals. In this case the modulation is said to be linear. This linear relationship does not always exist for phase and frequency modulation. As a result, the sidebands associated with angle modulation do not obey the principle of superposition.

7.6 Spectral content of FM signals

Unlike amplitude modulation, frequency modulation produces (theoretically) an infinite number of sidebands. It is not possible to evaluate the Fourier transform of a general FM signal, therefore, for the sake of simplicity, the case of a sinusoidal modulating signal is considered. In this case

$$f(t) = A\cos(\omega_m t)$$

and the instantaneous frequency is

$$\omega_i(t) = \omega_c + Ak_f \cos(\omega_m t)$$

which may be expressed

$$\omega_i(t) = \omega_c + \Delta\omega \cos(\omega_m t)$$

where $\Delta\omega$ is called the *peak frequency deviation*. The phase angle of the FM signal may be expressed

$$\theta(t) = \int_0^t \omega_i(\tau)\, d\tau$$

which may be expressed

$$\theta(t) = \omega_c t + \beta \sin(\omega_m t)$$

where $\beta = \Delta\omega / \omega_c$ is the modulation index of the FM signal.

The resulting FM signal may be expressed in phasor notation as

$$y(t) = Re\left\{Ae^{j\theta(t)}\right\}$$

or

$$y(t) = Re\left\{Ae^{j\omega_c t} e^{j\beta \sin\omega_m t}\right\}.$$

The second exponential term in the expression above can be expanded in a Fourier series

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_m t}$$

where

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} e^{j\beta\sin\omega_m t} e^{-jn\omega_m t}\,dt\,.$$

Making a change of variable $\xi = \omega_m t = (2\pi/T)t$ gives

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta\sin\xi - n\xi)}\,d\xi = J_n(\beta)$$

where $J_n(\beta)$ is the Bessel function of the first kind of order n. Using this result gives

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta)\, e^{jn\omega_m t}$$

therefore

$$y(t) = Re\left\{Ae^{j\omega_c t}\sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t}\right\} = A\sum_{n=-\infty}^{\infty} J_n(\beta)\cos(\omega_c + n\omega_m)t\,.$$

From this it is evident that an FM waveform with sinusoidal modulation has an infinite number of sidebands. However, the magnitudes of the spectral components of the higher-order sidebands are negligible. The number of sidebands which are significant depend on the order of the Bessel function n, and the value of β.

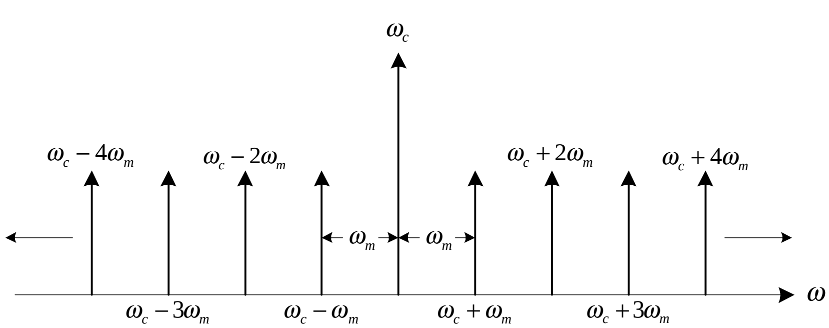

Figure 9. Spectrum of an FM waveform.

7.7 FM demodulation

Many methods of recovering information from an FM waveform exist. One method involves the use of a system that has a linear frequency-to-voltage transfer characteristic. This type of system is referred to as a *discriminator*. The simplest such device is an ideal differentiator.

A general FM waveform may be expressed

$$y(t) = A\cos\left[\omega_c t + k_f \int_0^t f(\tau)\,d\tau\right].$$

If A is a constant then

$$\frac{dy}{dt} = -A\left[\omega_c + k_f f(t)\right]\sin\left[\omega_c t + k_f \int_0^t f(\tau)\,d\tau\right].$$

If $k_f f(t) \ll \omega_c$, then the expression above has the from of an AM signal with envelope

$$A\omega_c\left[1 + \frac{k_f}{\omega_c} f(t)\right]$$

having frequency

$$\omega_c + k_f f(t).$$

Thus, the differentiator has converted the FM signal into an AM signal with a slight frequency variation (assuming that $k_f f(t) \ll \omega_c$). The original waveform may now be recovered by an envelope detector.

7.8 The spectrum analyzer

Spectrum analyzers are instruments that are used to measure the spectrum of periodic waveforms. The most common type of spectrum analyzer contains a superheterodyne receiver. A different type of spectrum analyzer that has had increasing popularity is called a Fast Fourier Transform (FFT) spectrum analyzer. These two types of spectrum analyzers will be discussed in this section.

– Superheterodyne spectrum analyzers

A spectrum analyzer is an example of a superheterodyne receiver. Each frequency component of a signal is mixed down to an intermediate frequency where it can be measured

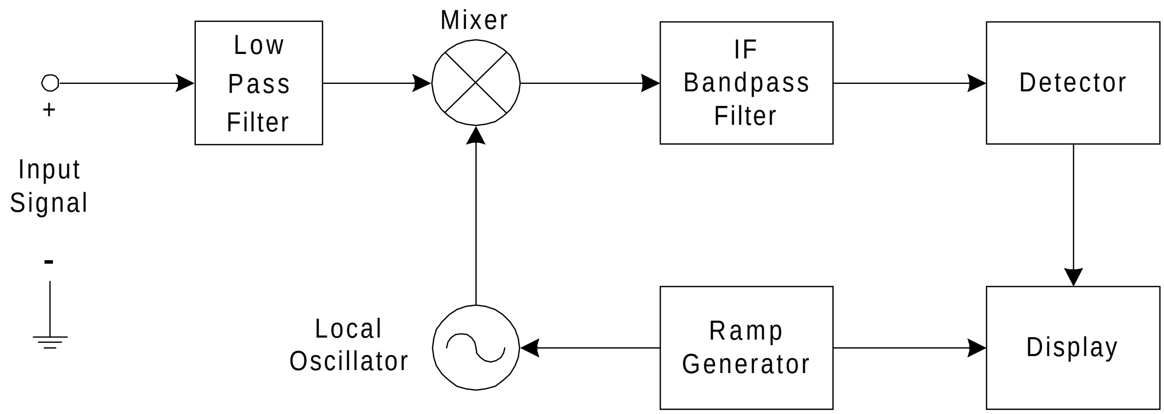

Figure 10. Superheterodyne spectrum analyzer

and manipulated. A simple block diagram of a superheterodyne spectrum analyzer is shown above.

– An input signal passes through a low pass filter to a mixer, where it mixes with a signal from a local oscillator.

– The output of the mixer includes not only the signals from the local oscillator and the low pass filter, but also the harmonics, and the sums and differences of the original signal frequencies and their harmonics.

– Any of the mixed signal that falls within the passband of the intermediate frequency filter is processed further, rectified by a detector, digitized, and applied to the vertical plates of a cathode-ray tube to produce a vertical deflection on the CRT screen.

– A ramp generator deflects the CRT beam horizontally across the screen from left to right. The ramp generator also tunes the local oscillator so that its frequency changes in proportion to the ramp voltage.

The spectrum analyzer shown above bears a strong resemblance to a superheterodyne AM radio that receives ordinary AM radio broadcasts. The output of the spectrum analyzer is the screen of a CRT instead of a speaker, and the local oscillator is tuned electronically instead of mechanically by a selector knob.

– **detectors**

To convert the IF signal from the IF bandpass filter to a constant voltage that the display can correctly process, some sort of detector must be used. Common types of detectors used in spectrum analyzers are peak, quasi-peak, and envelope detectors. Many regulatory agencies specify separate limits for these detectors. This is done because infrequently occurring events will result in a measured quasi-peak level that is smaller than what would be measured with a peak detector. Schematic diagrams of peak and quasi-peak detectors are shown below. These diagrams are simple approximations of actual detectors that may be more complicated.

In the peak detector, the IF signal is rectified by the diode. The positive voltage that reaches the capacitor charges the capacitor to a maximum value. This voltage is then processed and displayed. Therefore, even infrequent events will be measured by the peak detector. The quasi-peak detector has a resistor in parallel with the capacitor. This resistor provides a current path through which the capacitor can discharge. Therefore, infrequent events might give a lower measurement with a quasi-peak detector because the voltage stored in the capacitor decays faster in time than in the peak detector.

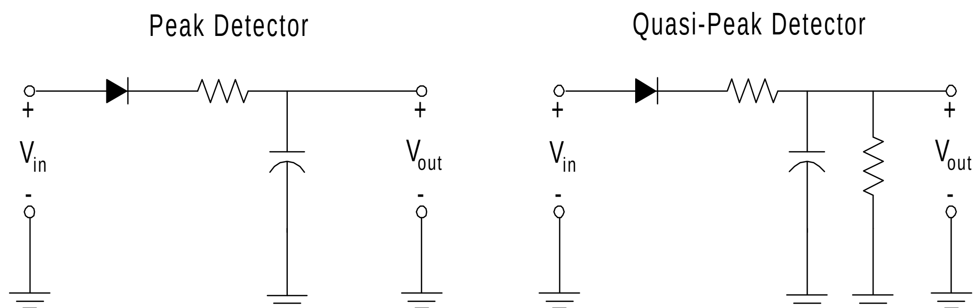

Figure 11. Peak and Quasi-Peak Detectors

The reason for the distinction between detectors is that the intent of the regulatory limits is to prevent interference in wire and radio communications receivers. Infrequent spikes and other short duration noise events do not substantially prevent the reception of desired information.

– **Fast Fourier Transform spectrum analyzers**

Another type of spectrum analyzer is a Fast Fourier Transform (FFT) spectrum analyzer. The FFT spectrum analyzer works by sampling and digitizing the input signal and then performing a discrete FFT on the digitized signal. FFT spectrum analyzers are able to preserve the phase information of a signal, which is difficult for a simple superheterodyne spectrum analyzer. Some of the disadvantages of an FFT spectrum analyzer are that the sensitivity, frequency range, and overall dynamic range are lower than current superheterodyne spectrum analyzers

References

1. Stremler, F., *Introduction to Communications Systems*, Addison Wesley Publishing Company, third edition, 1992.

2. White, G., *Mobile Radio Technology*, Newnes, 1995.

3. HP Application Note 150, *Spectrum Analysis Basics*