8 Amplitude Modulation

8.1 Introduction

To modulate means to regulate or adjust, and in the present context it means to regulate some parameter of a high-frequency *carrier* wave with a lower-frequency information signal. The need for modulation first arose in connection with radio transmission of relatively low frequency information signals such as audio signals. For efficient transmission it was found that the antenna dimensions had to be of the same order of magnitude as the wavelength of the signal being transmitted. The relationship between frequency f and wavelength λ for radio transmission is $f\lambda = c$, where $c = 3 \times 10^8$ m/s is the velocity of light in free space. For a typical low-frequency signal of frequency 1 kHz, the wavelength would therefore be on the order of 300 km (or 188 miles), which is obviously impractical.

The problem was overcome by using the low-frequency signal to modulate a much higher frequency signal termed the *carrier* wave, because it effectively carried the information signal. The relatively short wavelength of the high-frequency carrier wave meant that efficient antennas could be constructed.

For the practical implementation of modulation, the carrier frequency has to be very much greater than the highest frequency in the modulating signal, as will be shown later when specific circuits are examined. In practice, the carrier is always sinusoidal and can be described by

$$e_c(t) = E_{c\,\max} \sin(2\pi f_c t + \phi_c) \tag{8.1.1}$$

The parameters that can be modulated are the amplitude $E_{c\,\max}$, the frequency f_c, and the phase ϕ_c. The latter two come under the general heading of *angle modulation* and are the subject of Chapter 10.

A number of different forms of amplitude modulation are in use, and it becomes necessary to distinguish between these. The original concept, which is still widely in use, for example in the medium-wave broadcast band, is referred to simply as amplitude modulation or AM (or sometimes as *standard* AM) and is described in the following sections. As mentioned in Section 2.6, where simple harmonic functions are used, the cosine function is normally taken as reference [see Fig. 2.6.1(c)]. This practice will be followed here, although the term *sinusoidal* will be used when referring to either sine or cosine functions.

8.2 Amplitude Modulation

In amplitude modulation a voltage proportional to the modulating signal is added to the carrier amplitude. Let the added component of voltage be represented in functional notation as $e_m(t)$; then the modulated carrier wave is given by

$$e(t) = [E_{c\,\max} + e_m(t)]\cos(2\pi f_c t + \phi_c) \tag{8.2.1}$$

The term $[E_{c\,\max} + e_m(t)]$ describes the *envelope* of the modulated wave. Figure 8.2.1 shows (a) an arbitrary modulating signal, (b) a carrier wave, and (c) the resulting AM wave, where the envelope is seen to follow the modulating signal waveform. This also illustrates graphically why the term *carrier* is used.

8.3 Amplitude Modulation Index

Referring to Fig. 8.2.1(c), the *modulation index* is defined as

$$m = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \tag{8.3.1}$$

In the case of a periodic modulating signal, such as shown in Fig. 8.2.1, it is easy to identify the maximum and minimum voltages of the modulated wave. With a nonperiodic signal, such as a speech waveform, these quantities will vary, and hence the modulation index will also vary. What is important is that the modulation index must not be allowed to exceed unity. If the modulation index exceeds unity the negative peak of the modulating waveform is clipped, as shown in Fig. 8.3.1. This is bad enough in itself, but, in addition, such clipping is a potential source of interference, as will be shown shortly.

It will be noticed that overmodulation ($m > 1$) occurs when the magnitude of the peak negative voltage of the modulating wave exceeds the peak

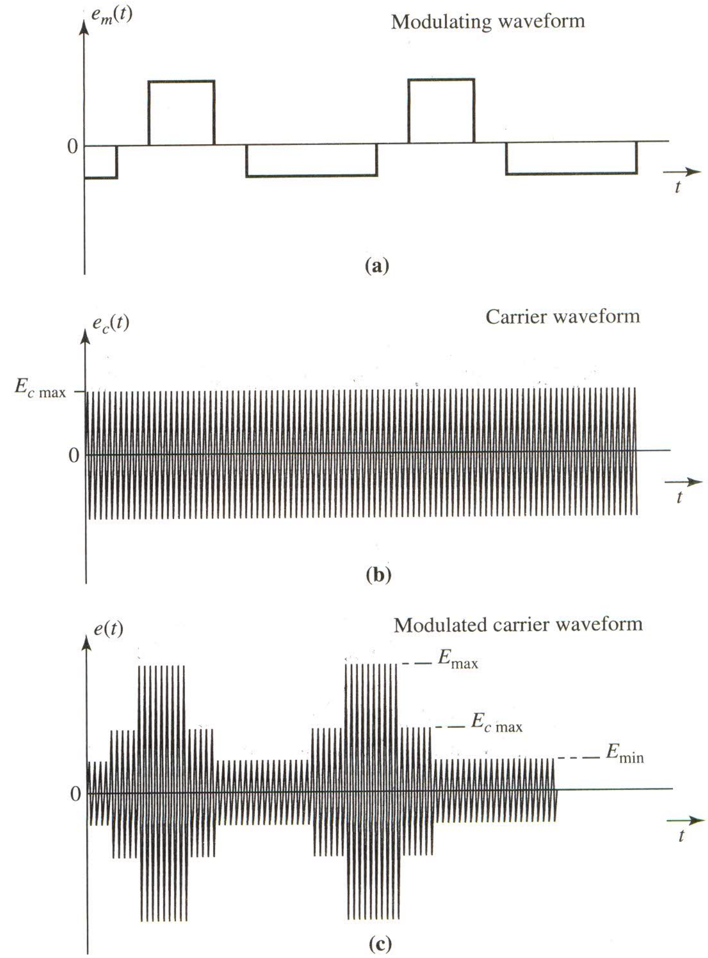

Figure 8.2.1 (a) Modulating voltage waveform. (b) Carrier wave. (c) Modulated waveform, showing the envelope that follows the modulating waveform.

carrier voltage. Under these conditions in practice, E_{min} is clamped at zero, as shown in Fig. 8.3.1(b). The mathematical expression for the modulated wave, Eq. (8.2.1), is not valid under these conditions. The envelope $[E_{c\ max} + e_m(t)]$ goes negative, which mathematically appears as a phase-reversal rather than a clamped level.

Viewing the modulated waveform directly on an oscilloscope is difficult when the modulating waveform is other than periodic because of the problem of synchronizing the sweep to obtain a stationary pattern. The problem can be overcome using the *trapezoidal method* of monitoring the modulation. The trapezoidal method is similar to that used to produce Lissajous patterns. Figure 8.3.2(a) shows the basic Lissajous method, in which the same waveform is applied to both horizontal and vertical plates of the oscilloscope. Assuming that the horizontal and vertical gains are equal and that the spot is initially centered on the screen, the spot will be at the screen center whenever the voltage is zero, as at A.

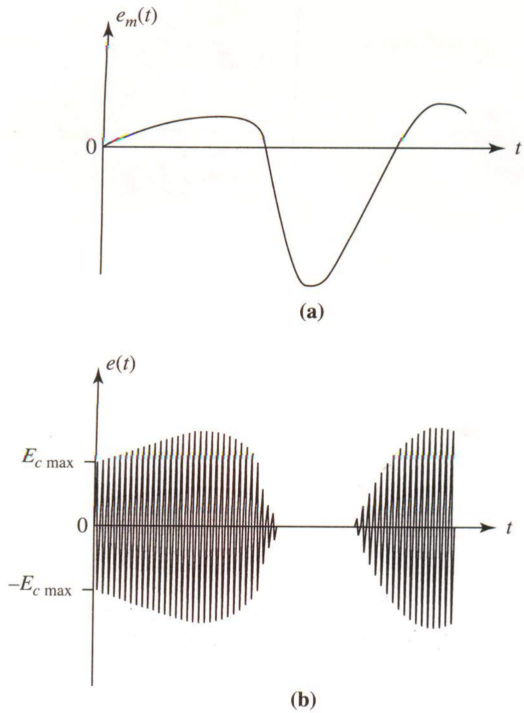

Figure 8.3.1 AM wave for which $m > 1$. (a) Modulating waveform and (b) the modulated waveform, showing clipping on the negative modulation peaks.

Whenever the voltage goes positive, as at B, the spot is deflected vertically upward and horizontally to the right by equal amounts, irrespective of the waveshape, and therefore the spot traces out the upper part of the diagonal line. Likewise, whenever the voltage goes negative, as at C, the downward deflection is equal to that to the left, producing the lower part of the diagonal line.

When the modulated wave is applied to the vertical plates, the spot is deflected vertically by the carrier voltage. For example, at A in Fig. 8.3.2(b), where the modulating voltage is zero, the spot traces out a vertical line centered on the screen, proportional to the peak-to-peak carrier voltage. As the modulating voltage goes positive, as at B, the peak-to-peak voltage of the modulated wave increases while the spot is deflected to the right. The trace is therefore trapezoidal, rather than just a single diagonal line. Likewise, when the modulating voltage goes negative, as at C, the peak-to-peak voltage decreases, while the horizontal deflection is to the left, resulting in the trapezoidal pattern continuing to the left.

Figure 8.3.3 shows a number of the patterns that can be obtained. Figure 8.3.3(a) shows the normal pattern from which the modulation index is easily obtained. Denoting the peak-to-peak voltage by E_{pp}, the longest vertical displacement is $L_1 = E_{pp\ max}$ and the shortest is $L_2 = E_{pp\ min}$. But since

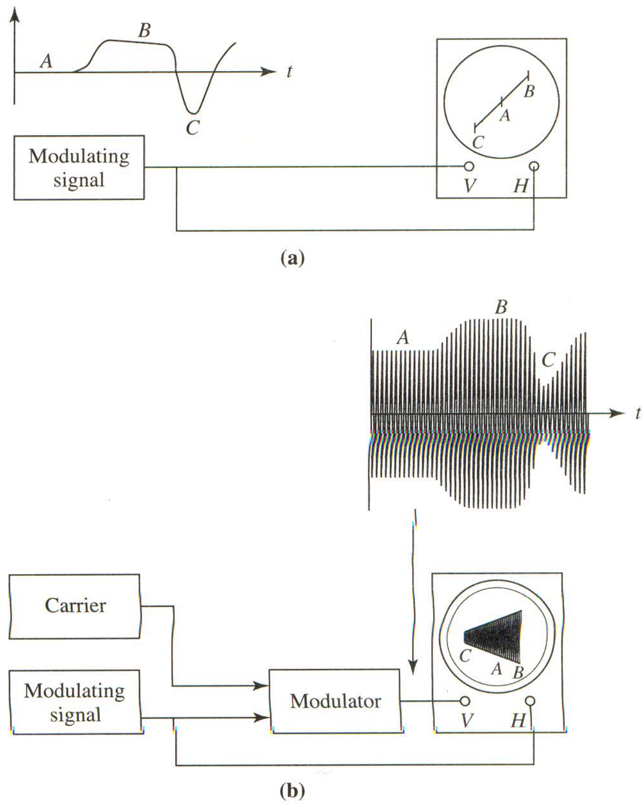

Figure 8.3.2 (a) Basic Lissajous method. (b) Method of obtaining the trapezoidal display.

$E_{pp\ max} = 2E_{max}$ and $E_{pp\ min} = 2E_{min}$, the trapezoidal display gives, on canceling out the common factor 2,

$$m = \frac{E_{max} - E_{min}}{E_{max} + E_{min}}$$
$$= \frac{L_1 - L_2}{L_1 + L_2} \tag{8.3.2}$$

Figure 8.3.3(b) shows the pattern obtained when overmodulation occurs. When E_{min} is zero, the length L_2 becomes zero and the trapezoid reduces to a triangle. Overmodulation results in a spike being produced at the L_2 point of the triangle, since the carrier voltage is cut off completely. (In some practical situations, leakage of the carrier may occur through the circuit, resulting in a blurring of the spike.)

It will be seen that the modulation index is zero when $E_{max} = E_{min} = E_{c\ max}$, and it is unity when $E_{min} = 0$. Thus, in practice, the modulation index should be in the range

$$0 \leq m \leq 1 \tag{8.3.3}$$

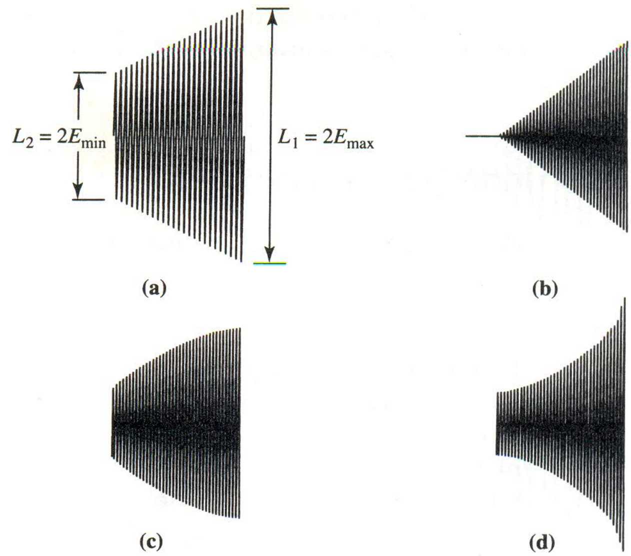

Figure 8.3.3 (a) Normal trapezoidal pattern. (b) Trapezoidal pattern for $m > 1$. (c) Envelope distortion resulting from insufficient RF drive to the modulator. (d) Envelope distortion resulting from nonlinearities in the modulator.

Figures 8.3.3(c) and (d) show two of the patterns obtained when envelope distortion is present. In Figure 8.3.3(c), the modulator output flattens off at high modulation levels, which could be a result of insufficient carrier input (or drive) to the modulator at these levels. Figure 8.3.3(d) shows the effect of a nonlinearity in the modulator, which indicates an accentuation of high levels of modulation relative to low levels.

EXAMPLE 8.3.1

A modulating signal consists of a symmetrical triangular wave having zero dc component and peak-to-peak voltage of 11 V. It is used to amplitude modulate a carrier of peak voltage 10 V. Calculate the modulation index and the ratio of the side lengths L_1/L_2 of the corresponding trapezoidal pattern.

SOLUTION

$$E_{max} = 10 + \frac{11}{2} = 15.5 \text{ V}$$

$$E_{min} = 10 - \frac{11}{2} = 4.5 \text{ V}$$

$$\therefore m = \frac{15.5 - 4.5}{15.5 + 4.5} = \mathbf{0.55}$$

L_1 is proportional to 15.5 V, L_2 to 4.5 V, and therefore $L_1/L_2 = 15.5/4.5 =$ **3.44.**

Fortunately, many of the characteristics of AM can be examined using sinusoidal modulation as described in the following sections.

8.4 Modulation Index for Sinusoidal AM

For sinusoidal AM, the modulating waveform is of the form

$$e_m(t) = E_{m\,\max}\cos(2\pi f_m t + \phi_m) \tag{8.4.1}$$

In general the fixed phase angle ϕ_m is unrelated to the fixed phase angle ϕ_c for the carrier, showing that these two signals are independent of each other in time. However, the amplitude modulation results are independent of these phase angles, which may therefore be set equal to zero to simplify the algebra and trigonometry used in the analysis. The equation for the sinusoidally modulated wave is therefore

$$e(t) = (E_{c\,\max} + E_{m\,\max}\cos 2\pi f_m t)\cos 2\pi f_c t \tag{8.4.2}$$

Since in this particular case $E_{\max} = E_{c\,\max} + E_{m\,\max}$ and $E_{\min} = E_{c\,\max} - E_{m\,\max}$ the modulation index is given by

$$\begin{aligned} m &= \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \\ &= \frac{E_{m\,\max}}{E_{c\,\max}} \end{aligned} \tag{8.4.3}$$

The equation for the sinusoidally amplitude modulated wave may therefore be written as

$$e(t) = E_{c\,\max}(1 + m\cos 2\pi f_m t)\cos 2\pi f_c t \tag{8.4.4}$$

Figure 8.4.1 shows the sinusoidally modulated waveforms for three different values of m.

8.5 Frequency Spectrum for Sinusoidal AM

Although the modulated waveform contains two frequencies f_c and f_m, the modulation process generates new frequencies that are the sum and difference of these. The spectrum is found by expanding the equation for the sinusoidally modulated AM as follows:

$$\begin{aligned} e(t) &= E_{c\,\max}(1 + m\cos 2\pi f_m t)\cos 2\pi f_c t \\ &= E_{c\,\max}\cos 2\pi f_c t + mE_{c\,\max}\cos 2\pi f_m t \times \cos 2\pi f_c t \end{aligned}$$

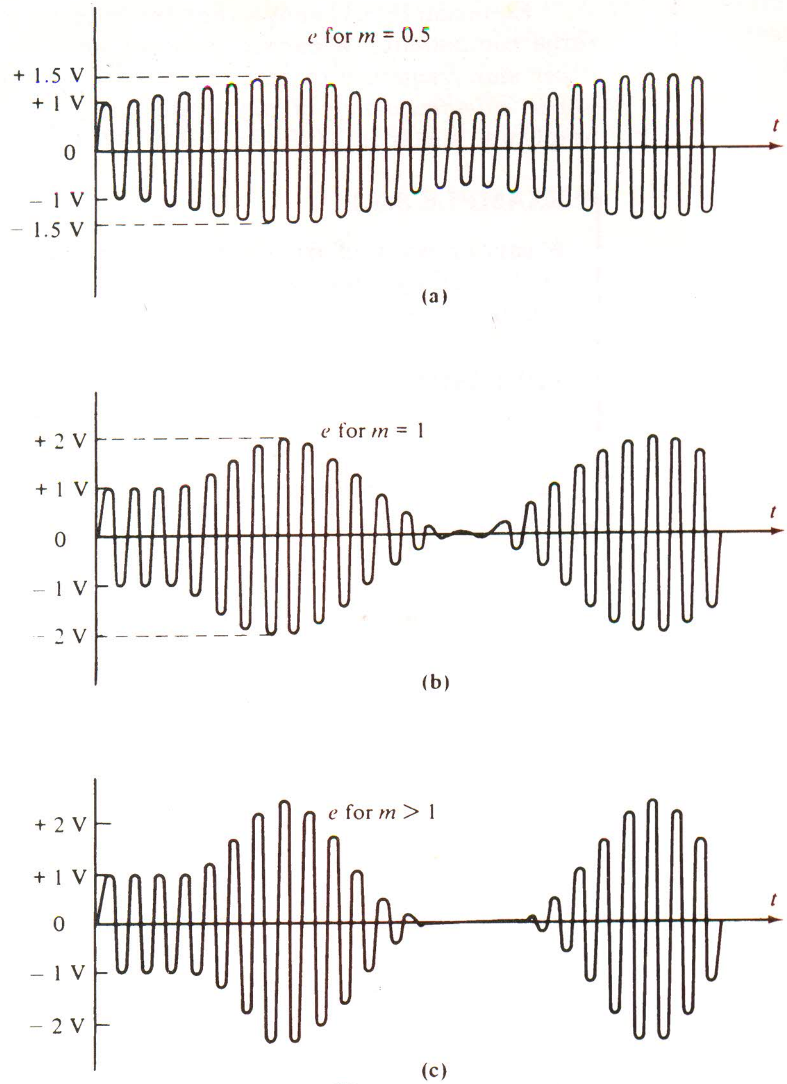

Figure 8.4.1 Sinusoidally amplitude modulated waveforms for (a) $m = 0.5$ (undermodulated), (b) $m = 1$ (fully modulated), and (c) $m > 1$ (overmodulated).

$$= E_{c\,\max} \cos 2\pi f_c t + \frac{m}{2} E_{c\,\max} \cos 2\pi (f_c - f_m)t + \frac{m}{2} E_{c\,\max} \cos 2\pi (f_c + f_m)t \tag{8.5.1}$$

It is left as an exercise for the student to derive this result making use of the trigonometric identity

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B \tag{8.5.2}$$

Equation (8.5.1) shows that the sinusoidally modulated wave consists of three components: a carrier wave of amplitude $E_{c\ \max}$ and frequency f_c, a *lower side frequency* of amplitude $mE_{c\ \max}/2$ and frequency $f_c - f_m$, and an *upper side frequency* of amplitude $mE_{c\ \max}/2$ and frequency $f_c + f_m$. The amplitude spectrum is shown in Fig. 8.5.1.

EXAMPLE 8.5.1

A carrier wave of frequency 10 MHz and peak value 10 V is amplitude modulated by a 5-kHz sine wave of amplitude 6 V. Determine the modulation index and draw the amplitude spectrum.

SOLUTION $m = \dfrac{6}{10} = 0.6$

The side frequencies are $10 \pm 0.005 = 10.005$ and 9.995 MHz. The amplitude of each side frequency is $0.6 \times 10/2 = 3$ V. The spectrum is shown in Fig. 8.5.1(b).

This result is of more than mathematical interest. The three components are present physically, and, for example, they can be separated out by filtering. Use is made of this in single-sideband transmission, discussed in Chapter 9.

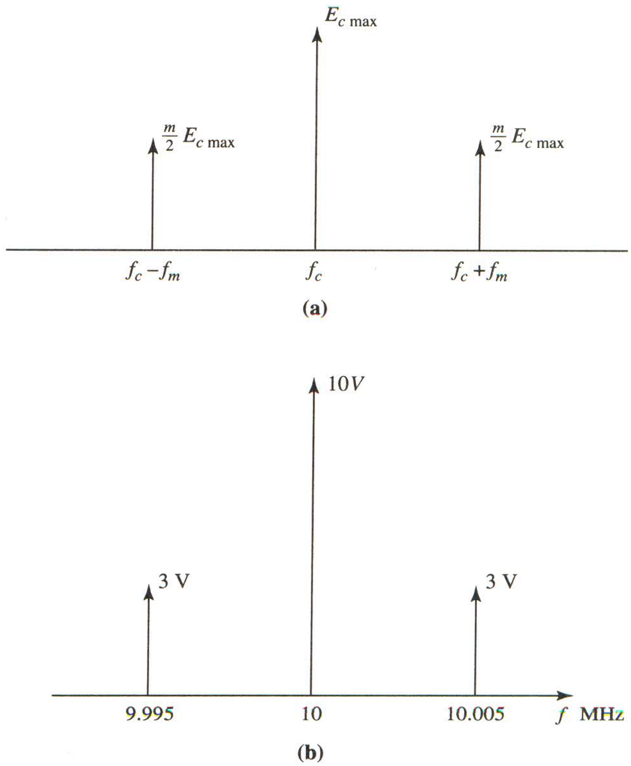

Figure 8.5.1 (a) Amplitude spectrum for a sinusoidally amplitude modulated wave. (b) The amplitude spectrum for a 10-MHz carrier of amplitude 10 V, sinusoidally modulated by a 5-kHz sine wave of amplitude 6 V (Example 8.5.1).

The modulated wave could be considered to be generated by three separate generators, as shown in Fig. 8.5.2(a). Although such an arrangement would be difficult to set up in practice (because of the difficulty in maintaining exactly the right frequencies), it is easily simulated on a computer, and the result of such a simulation obtained using Mathcad is shown in Fig. 8.5.2(b).

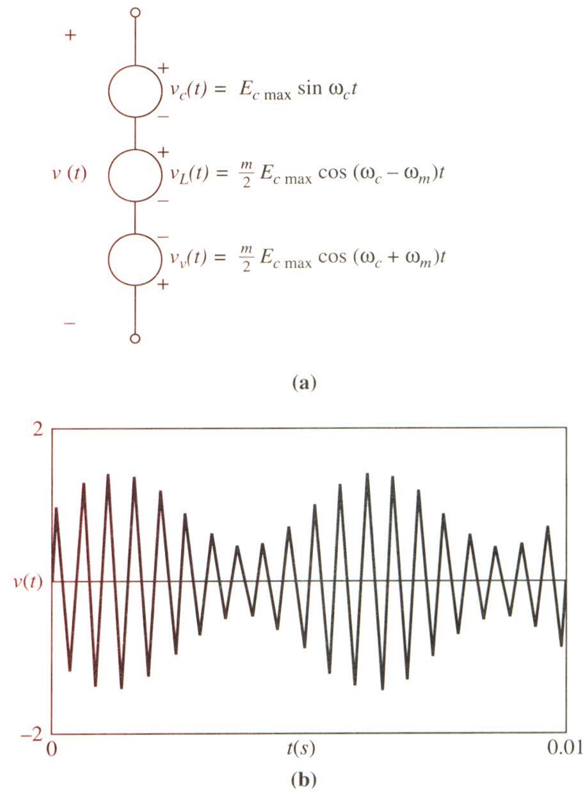

Figure 8.5.2 (a) Generator representation of a sinusoidally amplitude modulated wave. (b) Result of a computer simulation of the three-generator arrangement for $E_{c\,max} = 5$ V, $m = 0.5$, $f_c = 100$ kHz, and $f_m = 10$ kHz.

8.6 Average Power for Sinusoidal AM

Figure 8.5.1(a) shows that the sinusoidally modulated wave can be represented by three sinusoidal sources connected in series. A general result of ac circuit theory is that the average power delivered to a load R by series-connected sinusoidal sources of different frequencies is the sum of the average powers from each source. The average power in a sine (or cosine) voltage wave of peak value E_{max} developed across a resistor R is $P = E^2_{max}/2R$. Applying these results to the spectrum components of the sinusoidally modulated wave gives, for the average carrier power,

$$P_C = \frac{E^2_{c\,max}}{2R} \tag{8.6.1}$$

and for each side frequency

$$P_{SF} = \frac{(mE_{c\,\max}/2)^2}{2R} \qquad \frac{(m\,E_{c\max}/2)^2}{2R}$$

$$= \frac{m^2}{4}P_C \tag{8.6.2}$$

Hence the total average power is

$$P_T = P_C + 2 \times P_{SF}$$

$$= P_C\left(1 + \frac{m^2}{2}\right) \tag{8.6.3}$$

At 100% modulation (m = 1), the power in any one side frequency component is $P_{SF} = P_C/4$ and the total power is $P_T = 1.5P_C$. The ratio of power in any one side frequency to the total power transmitted is therefore 1/6. The significance of this result is that all the original modulating information is contained in the one side frequency, and therefore a considerable savings in power can be achieved by transmitting just the side frequency rather than the total modulated wave. In practice, the modulating signal generally contains a band of frequencies that results in *sidebands* rather than single side frequencies, but, again, single-sideband (SSB) transmission results in more efficient use of available power and spectrum space. Single-sideband transmission is considered in Chapter 9.

8.7 Effective Voltage and Current for Sinusoidal AM

The effective or rms voltage E of the modulated wave is defined by the equation

$$\frac{E^2}{R} = P_T \tag{8.7.1}$$

Likewise, the effective or rms voltage E_c of the carrier component is defined by

$$\frac{E_C^2}{R} = P_C \tag{8.7.2}$$

It follows from Eq. (8.6.3) that

$$\frac{E^2}{R} = P_C\left(1 + \frac{m^2}{2}\right)$$

$$= \frac{E_C^2}{R}\left(1 + \frac{m^2}{2}\right) \tag{8.7.3}$$

from which

$$E = E_C\sqrt{1 + \frac{m^2}{2}} \tag{8.7.4}$$

A similar argument applied to currents yields

$$I = I_C\sqrt{1 + \frac{m^2}{2}} \tag{8.7.5}$$

where I is the rms current of the modulated wave and I_C the rms current of the unmodulated carrier. The current equation provides one method of monitoring modulation index, by measuring the antenna current with and without modulation applied. From Eq. (8.7.5),

$$m = \sqrt{2\left[\left(\frac{I}{I_c}\right)^2 - 1\right]} \tag{8.7.6}$$

The method is not as sensitive or useful as the trapezoidal method described earlier, but it provides a convenient way of monitoring modulation where an ammeter can be inserted in series with the antenna, for example. A true rms reading ammeter must be used, and care must be taken to avoid current overload because such instruments are easily damaged by overload.

EXAMPLE 8.7.1

The rms antenna current of an AM radio transmitter is 10 A when unmodulated and 12 A when sinusoidally modulated. Calculate the modulation index.

SOLUTION $m = \sqrt{2\left[\left(\frac{12}{10}\right)^2 - 1\right]} = \mathbf{0.94}$

8.8 Nonsinusoidal Modulation

Nonsinusoidal modulation has already been illustrated in Fig. 8.2.1 and the modulation index determined as shown in Section 8.3. Sometimes the *modulation depth,* rather than modulation index, is used as a measure of modulation. The modulation depth is the ratio of the downward modulation peak to the peak carrier level, usually expressed as a percentage. As shown in Fig. 8.3.1, overmodulation occurs if the modulation depth exceeds 100%, irrespective of the modulating waveshape. (For sinusoidal modulation, modulation depth is equal to the modulation index. Signal generators generally employ sinusoidal modulation but have meters calibrated in modulation depth.)

Nonsinusoidal modulation produces upper and lower *sidebands,* corresponding to the upper and lower side frequencies produced with sinusoidal modulation. Suppose, for example, that the modulating signal has a line spectrum as shown in Chapter 2 so that it can be represented by

$$e_m(t) = E_{1\max}\cos 2\pi f_1 t + E_{2\max}\cos 2\pi f_2 t + E_{3\max}\cos 2\pi f_3 t + \cdots \tag{8.8.1}$$

As before, the AM wave is

$$e(t) = [E_{c\max} + e_m(t)]\cos 2\pi f_c t \tag{8.8.2}$$

If in general the ith component is denoted by subscript i, then individual modulation indexes may be defined as $m_i = E_{i\max}/E_{c\max}$ and the trigonometric expansion for Eq. (8.8.2) yields a spectrum with side frequencies at $f_c \pm f_i$ and amplitudes $m_i E_{c\max}/2$. This is sketched in Fig. 8.8.1(a). Thus, taken together, the side frequencies form sidebands either side of the carrier component. Again, the practicalities of AM demand that the carrier frequency be much greater than the highest frequency in the modulating wave, so the sidebands are bandlimited about the carrier frequency as shown.

The total average power can be obtained by adding the average power for each component (just as was done for single-tone modulation), which results in

$$P_T = P_c\left(1 + \frac{m_1^2}{2} + \frac{m_2^2}{2} + \frac{m_3^2}{2} + \cdots\right) \tag{8.8.3}$$

Hence an effective modulation index can be defined in this case as

$$m_{\text{eff}} = \sqrt{m_1^2 + m_2^2 + m_3^2 + \cdots} \tag{8.8.4}$$

It follows that the effective voltage and current in this case are

$$E = E_c\sqrt{1 + \frac{m_{\text{eff}}^2}{2}} \tag{8.8.5}$$

$$I = I_c\sqrt{1 + \frac{m_{\text{eff}}^2}{2}} \tag{8.8.6}$$

When the modulating signal is a random power signal such as speech or music, then the concept of power spectral density must be used, as shown in Chapter 2. Thus, if the power spectral density curve is as sketched in Fig. 2.17.1, when used to amplitude modulate the carrier, double sidebands are generated as shown in Fig. 8.8.1(b). Again it is assumed that the modulating signal is bandlimited such that the highest frequency in its spectrum is much less than the carrier frequency.

It will be seen therefore that standard AM produces upper and lower sidebands about the carrier, and hence the RF bandwidth required is double

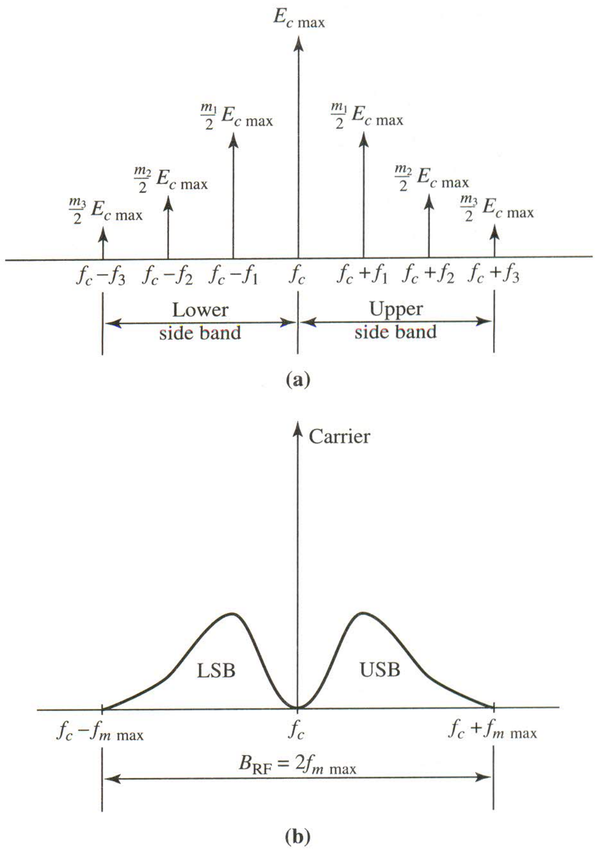

Figure 8.8.1 (a) Amplitude spectrum resulting from line spectra modulation. (b) Amplitude spectrum for a power density modulating spectra.

that for the modulating waveform. From Fig. 8.8.1,

$$\begin{aligned} B_{RF} &= (f_c + f_{m\,max}) - (f_c - f_{m\,max}) \\ &= 2f_{m\,max} \end{aligned} \tag{8.8.7}$$

where $f_{m\,max}$ is the highest frequency in the modulating spectrum. As with the sinusoidal modulation, either sideband contains all the modulating signal information, and therefore considerable savings in power and bandwidth can be achieved by transmitting only one sideband. Single sideband (SSB) transmission is the subject of Chapter 9.

Previously, overmodulation was shown to result in distortion of the modulation envelope (see Fig. 8.3.1). Such distortion also results in sideband frequencies being generated that lie outside the normal sidebands and that may overlap with the sidebands of adjacent channels. This form of interference is referred to as *sideband splatter* and must be avoided. Hence the necessity of ensuring that the modulation index does not exceed unity.

8.9 Double-sideband Suppressed Carrier (DSBSC) Modulation

Certain types of amplitude modulators make use of a multiplying action in which the modulating signal multiplies the carrier wave. The balanced mixer described in Section 5.10 is one such circuit, and in fact these are generally classified as *balanced modulators.* As shown by Eq. (5.10.17), the output current contains a product term of the two input voltages. When used as a modulator, the oscillator input becomes the carrier input, and the signal input becomes the modulating signal input. The output voltage can then be written as

$$e(t) = ke_m(t)\cos 2\pi f_c t \tag{8.9.1}$$

where k is a constant of the multiplier circuit. The expression for standard AM with carrier is

$$\begin{aligned} e(t) &= (E_{c\,\max} + e_m(t))\cos 2\pi f_c t \\ &= E_{c\,\max}\cos 2\pi f_c t + e_m(t)\cos 2\pi f_c t \end{aligned} \tag{8.9.2}$$

The major difference between the multiplier result and this is that carrier term $E_{c\,\max}\cos 2\pi f_c t$ is absent from the multiplier result. This means that the carrier component will be absent from the spectra, which otherwise will be the same as for AM with carrier. The constant multiplier k can be regarded simply as a scaling factor, and it will not materially affect the results. This type of amplitude modulation is therefore known as double-sideband suppressed carrier (DSBSC). The spectra are sketched in Fig. 8.9.1 for sinusoidal modulation and for the general case.

The absence of a carrier component means that DSBSC utilizes the transmitted power more efficiently than standard AM; however, it still requires twice the bandwidth compared to single sideband (SSB). It should be noted that, although the bandwidth is double that required for SSB, the received power is also double that obtained with SSB, and therefore the signal-to-noise ratio is the same. However, conserving bandwidth is an important aim in communications systems, and usually DSBSC represents one step in generating SSB, as described in Chapter 9.

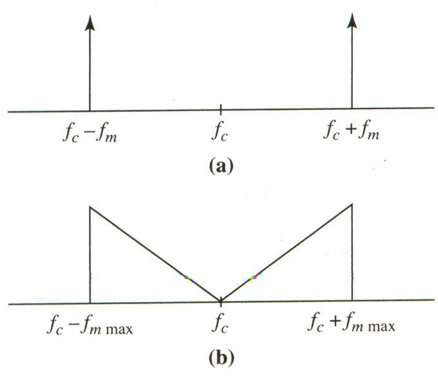

Figure 8.9.1 DSBSC spectrum for (a) sinusoidal modulation and (b) the general case.