Unit 6

Introduction to Electronic Communication Systems

Communication - 2eroon

CHAPTER 3

Amplitude Modulation

OBJECTIVES

After reading this chapter, studying the examples, and solving the problems, you should be able to:

- Write the time-domain equation for an AM signal and describe how the equation relates to the signal itself,
- Define the modulation index, calculate it, and measure it using either an oscilloscope or a spectrum analyzer,
- Describe the effects of overmodulation and explain why it must be avoided,
- Calculate the bandwidth of an AM signal and explain why bandwidth is an important factor in a communications system.
- Calculate power and voltage for an AM signal and for each of its components,
- Calculate the improvements in signal-to-noise ratio that result from the use of suppressed-carrier and single-sideband techniques.
- Analyze full-carrier, suppressed-carrier, and single-sideband suppressed-carrier AM signals in both the time and frequency domains.

ELECTRONICS ...REWIND

Early AM Radio

The simplest and historically earliest form of radio was the transmission of Morse code by simply switching a carrier on and off. The carrier was generated by applying a series of pulses to a tuned circuit by means of a spark gap. This is technically a form of amplitude modulation, but the technique is obviously not suitable for the transmission of audio.

Practical transmission of voice and music using AM radio had to wait for the development of the vacuum tube. An early attempt was made, however, by Reginald Aubrey Fessenden, a prolific radio inventor and engineer. On December 23, 1900, after many unsuccessful tries, Fessenden transmitted a few words using a spark-gap transmitter with a carbon microphone connected in series with the antenna. He used a transmitter that produced about ten thousand sparks per second, producing an approximation of a continuous transmission.

3.1 Introduction

In Chapter 1, we mentioned the several ways in which an information signal can modulate a carrier wave to produce a higher-frequency signal that carries the information. The most straightforward of these, and historically the first to be used, is **amplitude modulation** (AM). AM has the advantage of being usable with very simple modulators and demodulators. It does have some disadvantages, including poor performance in the presence of noise and inefficient use of transmitter power. However, its simplicity and the fact that it was the first system to become established have ensured its continued popularity. Applications include broadcasting in the medium- and high-frequency bands, aircraft communications in the VHF frequency range, and CB radio, among others.

The basic technique of amplitude modulation can also be modified to serve as the basis for a variety of more sophisticated schemes that are found in applications as diverse as television broadcasting and long-distance telephony. Thus, it is essential to understand the process of amplitude modulation in some detail, both for its own sake and as a foundation for further study.

An AM signal can be produced by using the instantaneous amplitude of the information signal (the baseband or modulating signal) to vary the peak amplitude of a higher-frequency signal. Figure 3.1(a) shows a 1 kHz sine wave, which can be combined with the 10 kHz signal shown in Figure 3.1(b) to produce the AM signal in Figure 3.1(c). If the peaks of the individual waveforms of the modulated signal are joined, the resulting envelope resembles the original modulating signal. It repeats at the modulating frequency, and the shape of each "half" (positive or negative) is the same as that of the modulating signal.

The higher-frequency signal that is combined with an information signal to produce the modulated waveform is called the *carrier*. Figure 3.1(c) shows a case in which there are only 10 cycles of the carrier for each cycle of the modulating signal. In practice, the ratio between carrier frequency and modulating frequency is usually much greater. For instance, an AM broadcasting station might have a carrier frequency of 1 MHz and a modulating frequency on the order of 1 kHz. A

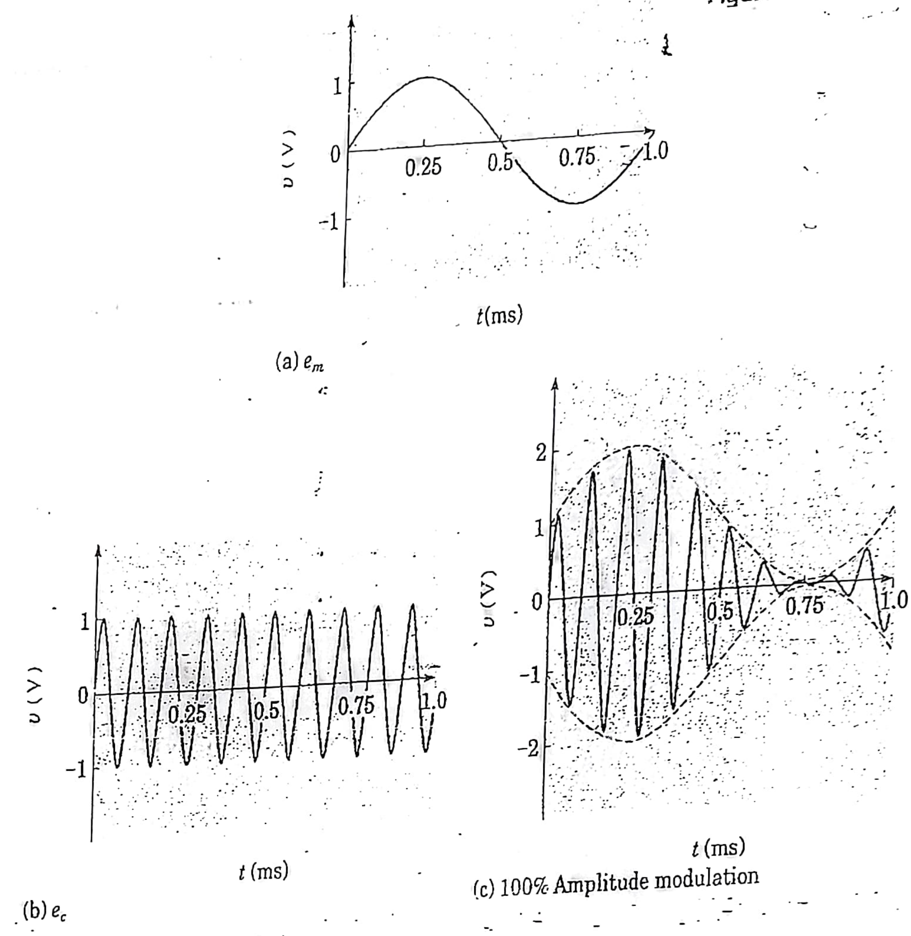


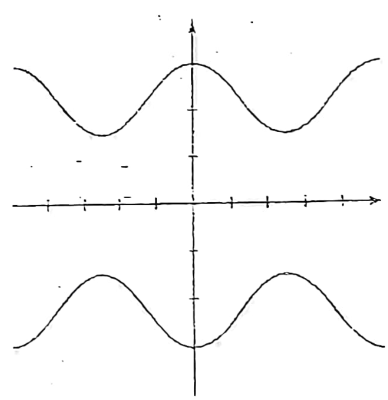
Figure 3.1 Amplitude modulation

waveform like this is shown in Figure 3.2. Since there are 1000 cycles of the carrier for each cycle of the envelope, the individual RF cycles are not visible, and only the envelope can be seen.

Note that amplitude modulation is not the simple linear addition of the two signals. Linear addition would produce the waveforms shown in Figure 3.3. Figure 3.3(a) shows a low-frequency signal, Figure 3.3(b) shows a higher-frequency signal, and Figure 3.3(c) shows the result of adding the two signals.

Amplitude modulation is essentially a nonlinear process. As in any nonlinear interaction between signals, sum and difference frequencies are produced that, in the case of amplitude modulation, contain the information to be transmitted. Another interesting thing about AM is that even though we seem to be varying the amplitude of the carrier (in fact, this is what is implied by the term *amplitude modulation*), a look at the frequency domain shows that the signal component at the carrier frequency survives intact, with the same amplitude and frequency as before! This mystery is easily unravelled with the aid of a little mathematics, as we shall see shortly; for the moment, just remember that AM is a bit of a misnomer, since the amplitude of the carrier remains constant. The amplitude of the entire signal does change with modulation, however, as is very clearly shown in Figure 3.1.

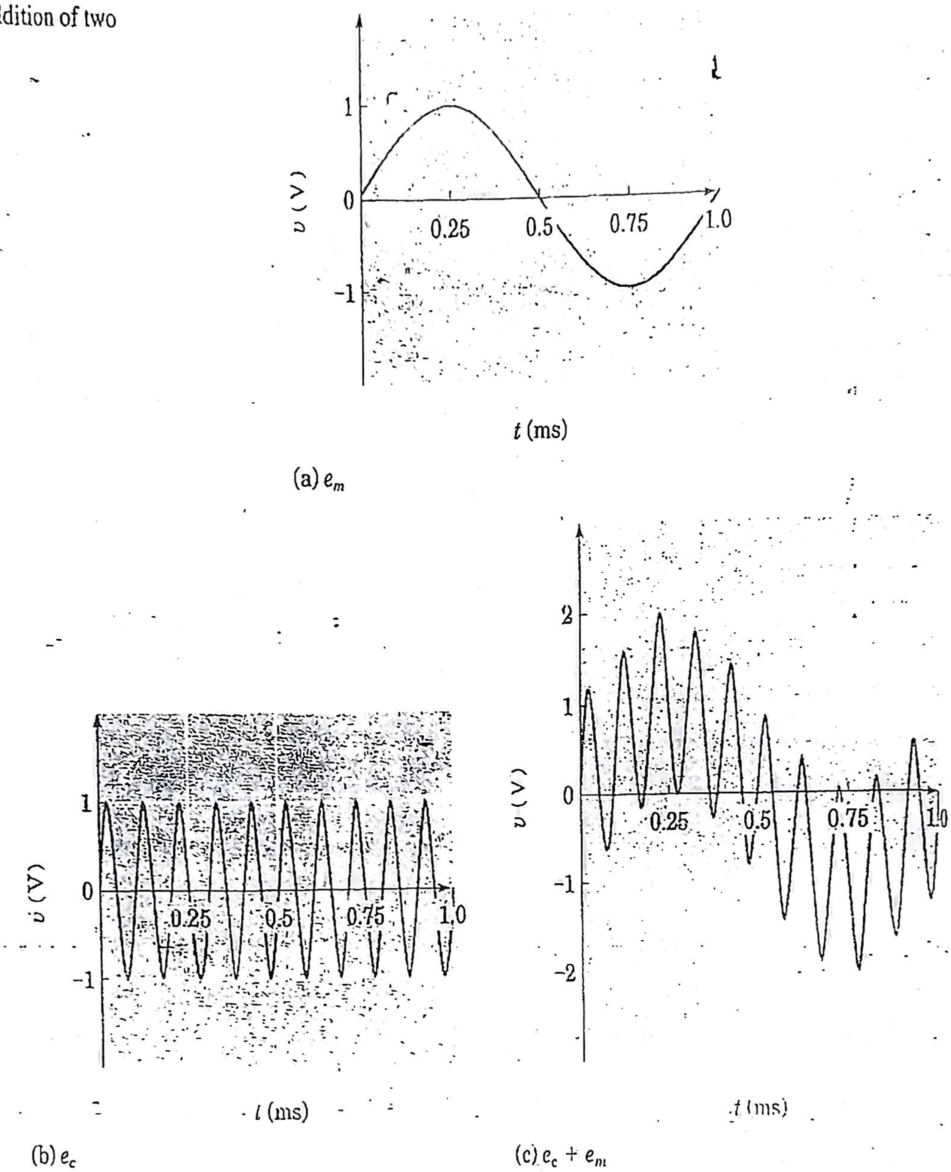
Figure 3.4(a) shows frequency-domain representations for amplitude modulation, and the linear addition of two signals is shown in Figure 3.4(b). The AM signal has no component at the modulating frequency: all the information is transmitted at frequencies near that of the carrier. In contrast, linear addition has accomplished nothing: the frequency-domain sketch shows that the information and carrier signals remain separate, each at its original frequency.



Vertical: 25 mV/division
Horizontal: 200 µs/division

Figure 3.2 Envelope of an AM signal

Figure 3.3 Linear addition of two signals



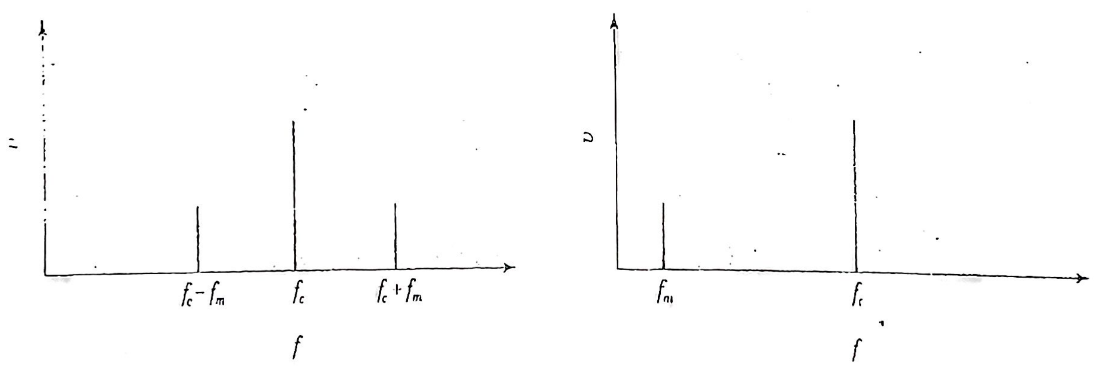


Figure 3.4 AM and linear addition in the frequency domain

3.2 Full-Car

Now that the gener
signal in greater d
the frequency com
characteristics
begin here.

Amplitude
nal voltage to vary th
ways a sine wave
arbitrary waveform,
modulation is very
be expressed as

We can express

where $v(t)$ = inst
E_c =
e_m = inst
ω_c =
t = time

The addition of
lated signal is both m

Once again
simple addition. Jus
equation

which is definite
forms simply adds th
on the other han
to the peak carrier an

If the modu
following form:

where E_m = peak
ω_m = radia

and the other variable

EXAMPLE 3.
quency of 1.5 MHz is
and amplitude of

3.2 Full-Carrier AM: Time-Domain

Now that the general idea of AM has been described, it is time to examine the signal in greater detail. We will look at the modulated signal in both the time and the frequency domains, since each approach emphasizes some of the important characteristics of AM. The time domain is probably more familiar, so let us begin here.

Amplitude modulation is created by using the instantaneous modulating signal voltage to vary the amplitude of the modulated signal. The carrier is almost always a sine wave. The modulating signal can be a sine wave but is more often an arbitrary waveform, such as an audio signal. However, an analysis of sine-wave modulation is very useful, since Fourier analysis often allows complex signals to be expressed as a series of sinusoids.

We can express the above relationship in the equation

$$v(t) = (E_c + e_m) \sin \omega_c t \tag{3.1}$$

where $v(t)$ = instantaneous amplitude of the modulated signal in volts
E_c = peak amplitude of the carrier in volts
e_m = instantaneous amplitude of the modulating signal in volts
ω_c = radian frequency of the carrier
t = time in seconds

The addition of E_c and e_m is algebraic. That is, the peak amplitude of the modulated signal is both increased and decreased by the modulation.

Once again, let us note the difference between amplitude modulation and simple addition. Just adding the carrier and modulating signals would give the equation

$$v'(t) = E_c \sin \omega_c t + e_m \tag{3.2}$$

which is definitely not the same as Equation (3.1). The summing of two waveforms simply adds their instantaneous values at all times. Amplitude modulation, on the other hand, involves the addition of the instantaneous baseband amplitude to the peak carrier amplitude.

If the modulating (baseband) signal is a sine wave, Equation (3.1) has the following form:

$$v(t) = (E_c + E_m \sin \omega_m t) \sin \omega_c t \tag{3.3}$$

where E_m = peak amplitude of the modulating signal in volts
ω_m = radian frequency of the modulating signal

and the other variables are as defined in Equation (3.1).

EXAMPLE 3.1 A carrier wave with an RMS voltage of 2 V and a frequency of 1.5 MHz is modulated by a sine wave with a frequency of 500 Hz and amplitude of 1 V RMS. Write the equation for the resulting signal.

Solution First, note that Equation (3.3) requires peak voltages and radian frequencies. We can easily get these as follows:

$$E_c = \sqrt{2} \times 2 \text{ V}$$
$$= 2.83 \text{ V}$$
$$E_m = \sqrt{2} \times 1 \text{ V}$$
$$= 1.41 \text{ V}$$

The equation also requires radian frequencies:

$$\omega_c = 2\pi \times 1.5 \times 10^6$$
$$= 9.42 \times 10^6 \text{ rad/s}$$
$$\omega_m = 2\pi \times 500$$
$$= 3.14 \times 10^3 \text{ rad/s}$$

So the equation is

$$v(t) = (E_c + E_m \sin \omega_m t) \sin \omega_c t$$
$$= [2.83 + 1.41 \sin (3.14 \times 10^3 t)] \sin (9.42 \times 10^6 t) \text{ V}$$

The Modulation Index

The amount by which the signal amplitude is changed in modulation depends on the ratio between the amplitudes of the modulating signal and the carrier. For convenience, this ratio is defined as the **modulation index** m. It is expressed mathematically as

$$m = \frac{E_m}{E_c} \tag{3.4}$$

Modulation can also be expressed as a percentage, with percent modulation found by multiplying m by 100. For example, $m = 0.5$ corresponds to 50% modulation.

Substituting m into Equation (3.3) gives:

$$v(t) = E_c(1 + m \sin \omega_m t) \sin \omega_c t \tag{3.5}$$

EXAMPLE 3.2 Calculate m for the signal in Example 3.1, and write the equation for this signal in the form of Equation (3.5).

Solution To avoid an accumulation of round-off errors, we should go back to the original voltage values to find m.

$$m = \frac{E_m}{E_c}$$
$$= \frac{1}{2}$$
$$= 0.5$$

It is all right to use the RMS values for calculating this ratio, as the $\sqrt{2}$ factors, if used to find the peak voltages, will cancel.

Now we can rewrite the equation:

$$\begin{aligned} v(t) &= E_c(1 + m \sin \omega_m t) \sin \omega_c t \\ &= 2.83[1 + 0.5 \sin (3.14 \times 10^3 t)] \sin (9.42 \times 10^6 t) \end{aligned}$$

It is worthwhile examining what happens to Equation (3.5) and to the modulated waveform as m varies. To start with, when $m = 0$, $E_m = 0$ and we have the original, unmodulated carrier. As m varies between 0 and 1, the changes due to modulation become more pronounced. Resultant waveforms for several values of m are shown in Figure 3.5. Note especially the result for $m = 1$ ($m = 100\%$). Under

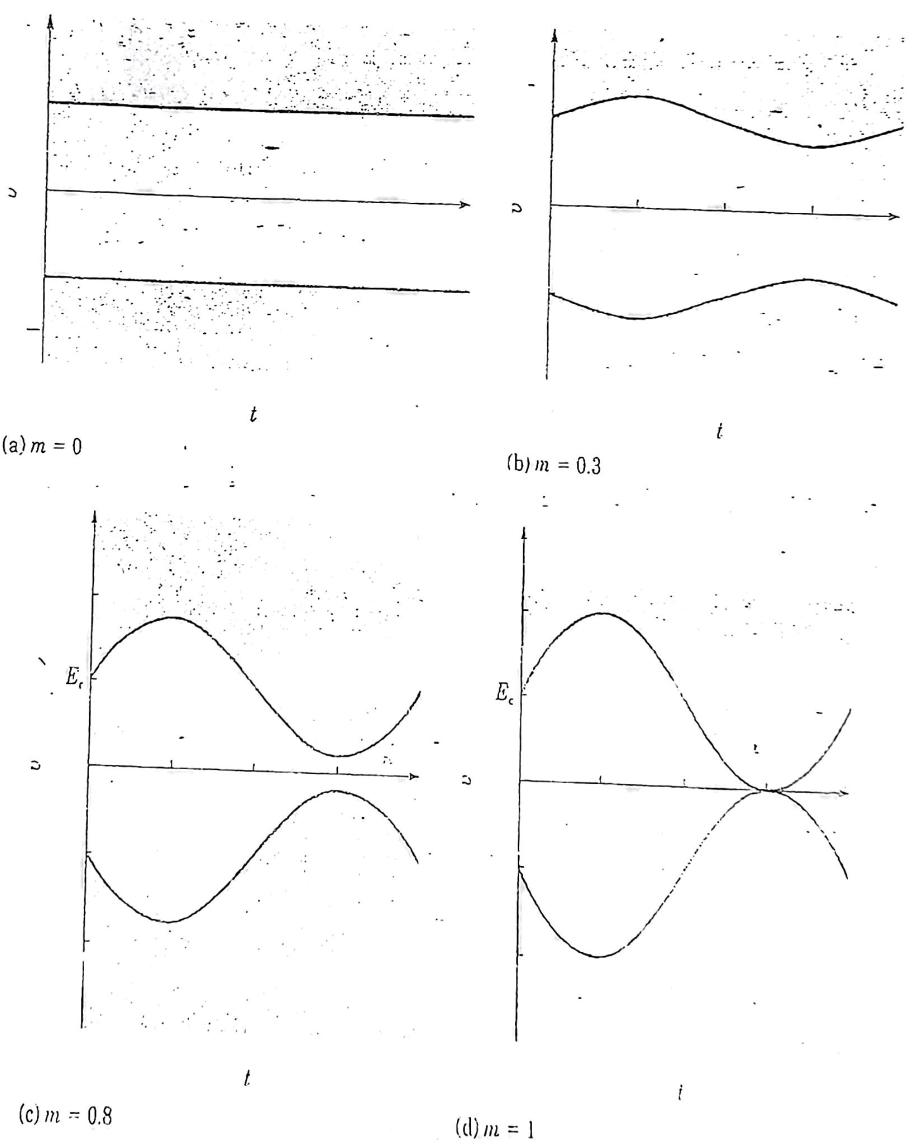
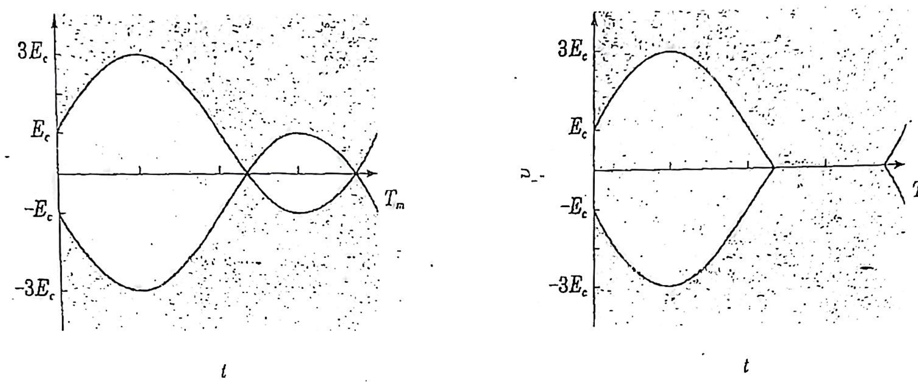


Figure 3.5 AM envelopes

these conditions, the peak signal voltage will vary between zero and twice the unmodulated carrier amplitude.

Overmodulation

When the modulation index is greater than 1, **overmodulation** is said to be present. There is nothing in Equation (3.3) that would seem to prevent E_m from being greater than E_c and m from being greater than 1. There are practical difficulties, however. Figure 3.6(a) shows the result of simply substituting $m = 2$ into Equation (3.5). As you can see, the envelope no longer resembles the modulating signal. Therefore, m must be kept less than or equal to 1.



(a) $m = 2$ in Equation (3.5)

(b) $m = 2$ with practical modulator

Figure 3.6 Overmodulation

Whenever we work with mathematical models, we must remember to keep checking against physical reality. This situation is a good example. It is possible to build a circuit that does produce an output that agrees with Equation (3.5) for m greater than 1. However, most practical AM modulators produce the signal shown in Figure 3.6(b) under these conditions. This is not the waveform predicted by Equation (3.5), but it does have the characteristic that the modulation envelope is no longer an accurate representation of the modulating signal. In fact, if subjected to Fourier analysis, the sharp "corners" on the waveform as the output goes to zero on negative modulation peaks would be found to represent high-frequency components added to the original baseband signal. This type of overmodulation creates side frequencies further from the carrier than would otherwise be the case. These spurious frequencies are known as **splatter**, and they cause the modulated signal to have increased bandwidth.

From the foregoing, we can conclude that for full-carrier AM, m must be in the range from 0 to 1. Overmodulation creates distortion in the demodulated signal and may result in the signal occupying a larger bandwidth than normal. Since spectrum space is tightly controlled by law, overmodulation of an AM transmitter is actually illegal, and means must be provided to prevent it.

Modulation Index for Multiple Modulating Frequencies

Practical AM systems are seldom used to transmit sine waves, of course. The information signal is more likely to be a voice signal, which contains many frequencies. Though a typical audio signal is not strictly periodic, it is close enough to periodic that we can use the idea of Fourier series to consider it as a series of sine waves of different frequencies.

When there are two or more sine waves of different, uncorrelated frequencies (that is, frequencies that are not multiples of each other) modulating a single carrier, m is calculated by using the equation

$$m_T = \sqrt{m_1^2 + m_2^2 + \cdots} \tag{3.6}$$

where m_T = total resultant modulation index

$m_1, m_2, \ldots$ = modulation indices due to the individual modulating components

---

EXAMPLE 3.3 Find the modulation index if a 10 V carrier is amplitude-modulated by three different frequencies with amplitudes of 1 V, 2 V, and 3 V, respectively.

Solution The three separate modulation indices are:

$$m_1 = \frac{1}{10} = 0.1 \qquad m_2 = \frac{2}{10} = 0.2 \qquad m_3 = \frac{3}{10} = 0.3$$

$$\begin{aligned} m_T &= \sqrt{m_1^2 + m_2^2 + m_3^2} \\ &= \sqrt{0.1^2 + 0.2^2 + 0.3^2} \\ &= 0.374 \end{aligned}$$

Measurement of Modulation Index

If we let E_m and E_c be the peak modulation and carrier voltages, respectively, then we can see, either by using Equations (3.3) and (3.5) or by inspecting Figure 3.7, that the maximum envelope voltage is simply

$$E_{max} = E_c + E_m$$

or

$$E_{max} = E_c(1 + m) \tag{3.7}$$

and the minimum envelope voltage is

$$E_{min} = E_c - E_m$$

or

$$E_{min} = E_c(1 - m) \tag{3.8}$$

Note, by the way, that these results agree with the conclusions expressed earlier: for $m = 0$, the peak voltage is E_c, and for $m = 1$, the envelope voltage ranges from $2E_c$ to zero.

Applying a little algebra to the above expressions, it is easy to show that

$$m = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \tag{3.9}$$

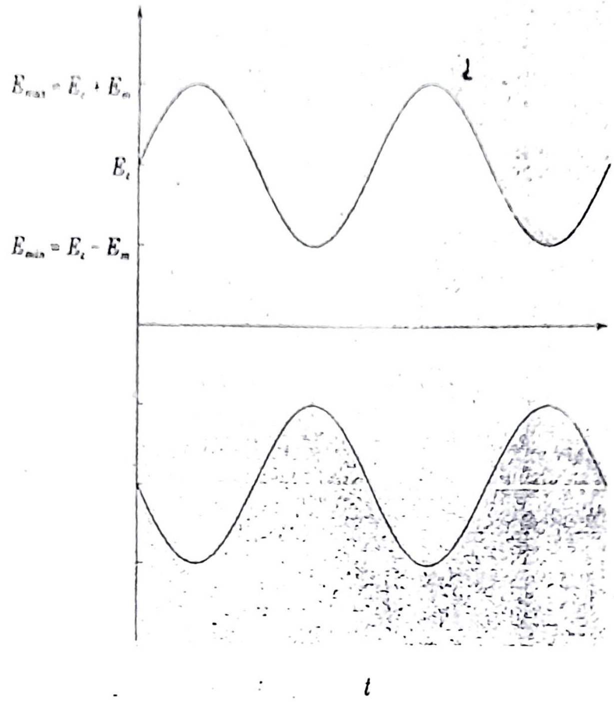


Figure 3.7 Voltage relationships in an AM signal

Of course, doubling both E_{max} and E_{min} will have no effect on this equation, so it is quite easy to find m by displaying the envelope on an oscilloscope and measuring the maximum and minimum peak-to-peak values for the envelope voltage. Other time-domain methods for measuring the amplitude modulation index will be described along with AM transmitters, in Chapter 5.

EXAMPLE 3.4 Calculate the modulation index for the waveform shown in Figure 3.2.

Solution It is easiest to use peak-to-peak values with an oscilloscope. From the figure we see that:

$$E_{max} = 150 \text{ mV} \qquad E_{min} = 70 \text{ mV} \qquad m = \frac{E_{max} - E_{min}}{E_{max} + E_{min}}$$

$$= \frac{150 - 70}{150 + 70}$$

$$= 0.364 \text{ or } 36.4\%$$

SECTION 3.2 REVIEW QUESTION

Why are values of modulation index greater than one not used in full-carrier AM systems?

3.3 Full-Carrier AM: Frequency Domain

RealAudio

So far, we have looked at the AM signal exclusively in the time domain, that is, as it can be seen on an ordinary oscilloscope. In order to find out more about this signal, however, it is necessary to consider its spectral makeup. We could use Fourier methods to do this, but for a simple AM waveform it is easier and just as valid to use trigonometry.

To start, we should note that although both the carrier and the modulating signal may be sine waves, the modulated AM waveform is *not* a sine wave. This can be seen from a simple examination of the waveform of Figure 3.1(c). It is important to remember that the modulated waveform is not a sine wave when, for instance, trying to find RMS from peak voltages. The usual formulas, so laboriously learned in fundamentals courses, do not apply here!

If an AM signal is not a sine wave, then what is it? We already have a mathematical expression, given by Equation (3.5):

$$v(t) = E_c(1 + m \sin \omega_m t) \sin \omega_c t$$

Expanding it and using a trigonometric identity will prove useful. Expanding gives

$$v(t) = E_c \sin \omega_c t + mE_c \sin \omega_m t \sin \omega_c t$$

The first term is just the carrier. The second can be expanded using two trigonometric identities:

$$\sin A \sin B = \frac{1}{2}[\cos (A - B) - \cos (A + B)]$$

and

$$\cos A = \cos (-A)$$

to give

$$v(t) = E_c \sin \omega_c t + \frac{mE_c}{2}[\cos (\omega_c - \omega_m)t - \cos (\omega_c + \omega_m)t]$$

which can be separated into three distinct terms:

$$v(t) = E_c \sin \omega_c t + \frac{mE_c}{2} \cos (\omega_c - \omega_m)t - \frac{mE_c}{2} \cos (\omega_c + \omega_m)t \quad (3.10)$$

We now have, besides the original carrier, two additional sinusoidal waves, one above the carrier frequency and one below. When the complete signal is sketched in the frequency domain, as in Figure 3.8, we see the carrier and two additional frequencies (one to each side), which are called, logically enough, side frequencies. The separation of each side frequency from the carrier is equal to the modulating frequency, and the relative amplitude of the side frequency, compared with that of the carrier, is proportional to m, becoming half the carrier voltage for $m = 1$.

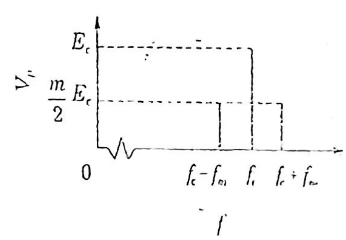


Figure 3.8 AM in the frequency domain

In a realistic situation there is generally more than one set of side frequencies, because there is more than one modulating frequency. Each modulating frequency produces two side frequencies. Those above the carrier can be grouped into a band of frequencies called the upper **sideband**, and the lower sideband looks like a mirror image of the upper, reflected in the carrier.

From now on, we will generally use the term *sideband*, rather than *side frequency*, even for the case of single-tone modulation, because it is more general and more commonly used in practice.

Mathematically, we have:

$$f_{usb} = f_c + f_m \tag{3.11}$$

$$f_{lsb} = f_c - f_m \tag{3.12}$$

$$E_{lsb} = E_{usb} = \frac{mE_c}{2} \tag{3.13}$$

where f_{usb} = frequency of the upper sideband
f_{lsb} = frequency of the lower sideband
E_{usb} = peak voltage of the upper-sideband component
E_{lsb} = peak voltage of the lower-sideband component

EXAMPLE 3.5

(a) A 1 MHz carrier with an amplitude of 1 V peak is modulated by a 1 kHz signal with $m = 0.5$. Sketch the voltage spectrum.

(b) An additional 2 kHz signal modulates the carrier with $m = 0.2$. Sketch the voltage spectrum.

Solution

(a) The frequency scale is easy. There will be three frequency components. The carrier will be at:

$$f_c = 1 \text{ MHz}$$

The upper sideband will be at:

$$\begin{aligned} f_{usb} &= f_c + f_m \\ &= 1 \text{ MHz} + 1 \text{ kHz} \\ &= 1.001 \text{ MHz} \end{aligned}$$

The lower sideband will be at:

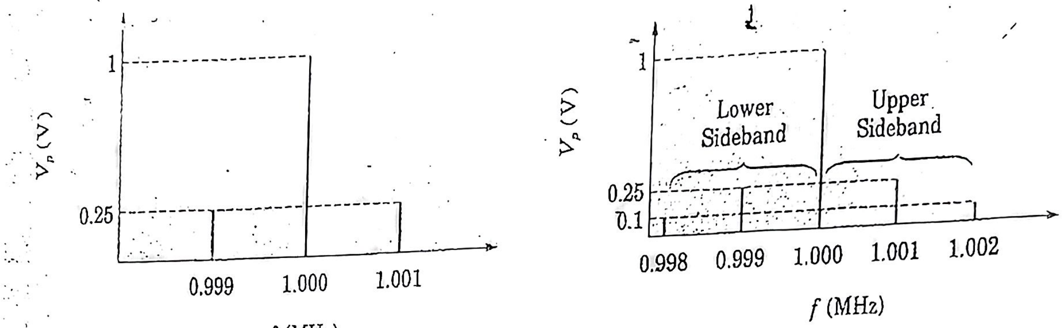
$$\begin{aligned} f_{lsb} &= f_c - f_m \\ &= 1 \text{ MHz} - 1 \text{ kHz} \\ &= 0.999 \text{ MHz} \end{aligned}$$

Next, we have to determine the amplitudes of the three components. The carrier is unchanged with modulation, so it remains at 1 V peak. The two sidebands will have the same peak voltage:

$$\begin{aligned} E_{lsb} = E_{usb} &= \frac{mE_c}{2} \\ &= \frac{0.5 \times 1}{2} \\ &= 0.25 \text{ V} \end{aligned}$$

Figure 3.9(a) shows the solution.

(b) The addition of another modulating signal at a different frequency simply adds another set of side frequencies. It does not change anything that was done in part (a). The new frequency components will be at 1.002 and 0.998 MHz, and their amplitude will be 0.1 V. The result is shown in Figure 3.9(b).



(a) $f_c = 1$ MHz $f_m = 1$ kHz $m = 0.5$ $E_c = 1$ MHz

(b) $f_c = 1$ MHz $E_c = 1$ V $f_{m_1} = 1$ kHz $m_1 = 0.5$ $f_{m_2} = 2$ kHz $m_2 = 0.2$

Figure 3.9

Bandwidth

Signal bandwidth is one of the most important characteristics of any modulation scheme. In general, a narrow bandwidth is desirable. In any situation where spectrum space is limited, a narrow bandwidth allows more signals to be transmitted simultaneously than does a wider bandwidth. It also allows a narrower bandwidth to be used in the receiver. Since ordinary thermal noise is evenly distributed over the frequency domain, using narrower bandwidth in receivers will include less noise, thereby increasing the signal-to-noise ratio. (There is one major exception to the general rule that reducing the bandwidth improves the signal-to-noise ratio. The exception is for wideband frequency modulation, which will be described in Chapter 4. However, the receiver must have a wide enough bandwidth to pass the complete signal including all the sidebands, or distortion will result. Consequently, we will have to calculate the signal bandwidth for each of the modulation schemes we consider.

A glance at Equation (3.10) and Figure 3.8 will show that this calculation is very easy for AM. The signal extends from the lower side frequency, which is at the carrier frequency less the modulation frequency, to the upper side frequency, at the carrier frequency plus the modulation frequency. The difference between these is simply twice the modulation frequency.

If we have a complex modulating signal with more than one modulating frequency, as in Figure 3.9(b), the bandwidth will be twice the *highest* modulating frequency. For telephone-quality voice, for instance, a bandwidth of about 6 kHz would suffice, while a video signal with a 4 MHz maximum baseband frequency would need 8 MHz of bandwidth if transmitted in this way. (Since a television channel is only 6 MHz wide, we can surmise, correctly, that television must actually be transmitted by a more complex modulation scheme that uses less bandwidth.)

Mathematically, the relationship is:

$$B = 2F_m \tag{3.14}$$

where B = bandwidth in hertz

F_m = highest modulating frequency in hertz

EXAMPLE 3.6 CB radio channels are 10 kHz apart. What is the maximum modulation frequency that can be used if a signal is to remain entirely within its assigned channel?

Solution From Equation (3.14) we have

$$B = 2F_m$$

or

$$F_m = \frac{B}{2}$$

$$= \frac{10 \text{ kHz}}{2}$$

$$= 5 \text{ kHz}$$

By the way, many people assume that because AM radio broadcast channe are also assigned at 10 kHz intervals, there is a similar limitation on the audio fr quency response for broadcasting. This is not the case: AM broadcast transmitte typically have an audio frequency response extending to about 10 kHz, giving theoretical signal bandwidth of 20 kHz. This works because adjacent channels a not assigned in the same locality, so some overlap is possible. For instance, your city has a station at 1000 kHz, there will not be one at 990 or 1010 kHz. order to reduce interference from distant stations, many AM receivers do hav narrow bandwidth and consequently limited audio frequency response.

Power Relationships

Power is important in any communications scheme, because the crucial signal-t noise ratio at the receiver depends as much on the signal power being large as t the noise power being small. The power that is most important, however, is n the total signal power but only that portion that is used to transmit informatio Since the carrier in an AM signal remains unchanged with modulation, it contai no information. Its only function is to aid in demodulating the signal at the r ceiver. This makes AM inherently wasteful of power compared with some oth modulation schemes we will describe later.

The easiest way to look at the power in an AM signal is to use the frequen domain. We can find the power in each frequency component and then add the to get total power. We will assume that the signal appears across a resistance R, s that reactive volt-amperes can be ignored. We will also assume that the pow required is average power.

Suppose that the modulating signal is a sine wave. Then the AM signal co sists of three sinusoids, the carrier and two side frequencies (sidebands), as show in Figure 3.8.

The power in the carrier is easy to calculate, since the carrier by itself is sine wave. The carrier is given by the equation

$$e_c = E_c \sin \omega_c t$$

where e_c = instantaneous carrier voltage
E_c = peak carrier voltage

CHAPTER 4

Angle Modulation

OBJECTIVES

After studying the material in this chapter, you should be able to:

- Describe and explain the differences between amplitude and angle modulation schemes and the advantages and disadvantages of each,
- Describe and explain the differences between frequency and phase modulation and show the relationship between the two,
- Calculate bandwidth, sideband frequencies, carrier and sideband voltage and power levels, and modulation index for frequency- and phase-modulated signals,
- Explain the capture effect and noise threshold level for FM signals and calculate the signal-to-noise ratio for simple situations.
- Relate deviation, bandwidth, and signal-to-noise improvement for FM systems,
- Explain the use of pre-emphasis and de-emphasis in FM systems and calculate component values for pre-emphasis and de-emphasis circuits.
- Describe the system used for FM stereo broadcasting and draw a diagram showing the spectrum of an FM stereo signal,
- Perform measurements on FM signals using a spectrum analyzer.

ELECTRONICS REWIND

Historical Development of FM

FM was considered very early in the development of radio communications. At first, it was thought that FM might permit a reduced transmission bandwidth compared with AM. This was refuted by experimental tests and also mathematically by John Renshaw Carson (1887–1940) in 1922.

Carson failed to notice that FM does have an advantage over AM in terms of signal-to-noise ratio. Edwin Armstrong (yes, the same Armstrong who invented the superheterodyne receiver) did notice this, and in 1936 he proposed a practical FM system. FM broadcasting began in the United States in 1939 but suffered a setback in 1944 when its frequency allocation was abruptly shifted from 42–50 MHz to its present range of 88–108 MHz. FM broadcasting gradually became popular because of its noise and fidelity advantages over AM. There are now actually more FM than AM listeners.

Unfortunately, Armstrong did not benefit from the success of FM broadcasting. He spent the remainder of his life involved in lawsuits in an attempt to receive royalties from his inventions, and finally, a broken man, he committed suicide in 1954.

4.1 Introduction

At the beginning of this book, we observed that only three parameters of a carrier wave can be changed or modulated in order for it to carry information: amplitude, frequency, and phase. The last two are closely related, since frequency (expressed in radians per second) is the rate of change of phase angle (in radians). If either frequency or phase is changed in a modulation system, the other will change as well. Consequently, it is useful to group frequency and phase modulation together in the term angle modulation.

Both *frequency modulation (FM)* and *phase modulation (PM)* are widely used in communications systems. FM is more familiar in daily life, since it is used extensively for radio broadcasting. FM is also used for the sound signal in television, for two-way fixed and mobile radio systems, for satellite communications, and for cellular telephone systems, to name only a few of its more common applications.

While PM may be less familiar, it is used extensively in data communications. It is also used in some FM transmitters as an intermediate step in the generation of FM. FM and PM are closely related mathematically, and it is quite easy to change one to the other.

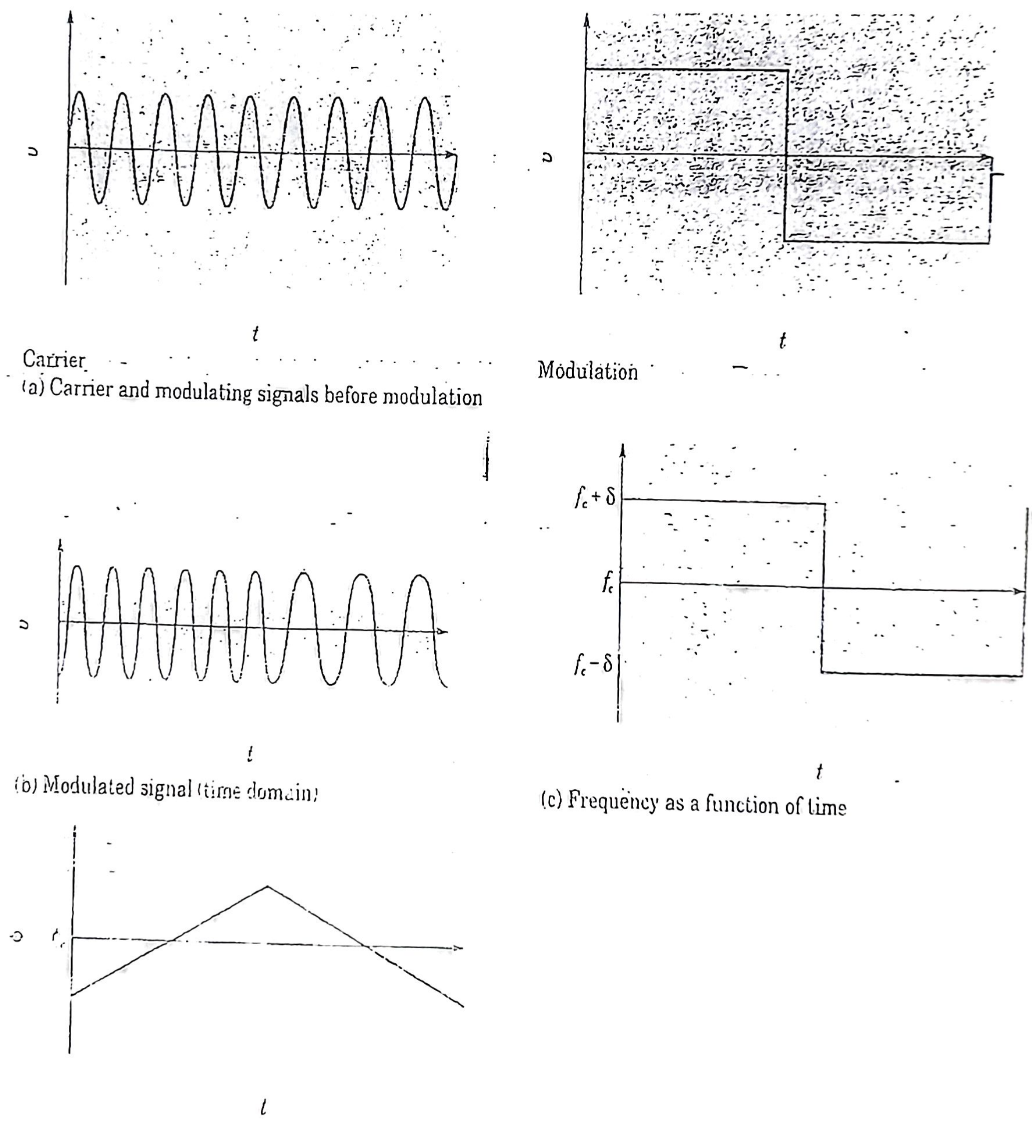
The most important advantage of FM or PM over AM is the possibility of a greatly improved signal-to-noise ratio. A penalty is paid for this in increased bandwidth: an FM signal may occupy several times as much bandwidth as that required for an AM signal. There may seem to be a contradiction here, as we found that for AM, decreasing the bandwidth improved the signal-to-noise ratio. This seeming contradiction will be resolved shortly.

In our discussion of amplitude modulation, we found that the amplitude of the modulated signal varied in accordance with the instantaneous amplitude of the modulating signal. In FM, the *frequency* of the modulated signal varies with the amplitude of the modulating signal. In PM, the *phase* varies directly with the modulating-signal amplitude. It is important to remember that in all types of modulation, it is the amplitude of the modulating signal that varies the carrier wave.

In contrast to AM, the amplitude and the power of an FM or PM signal do not change with modulation. Thus, an FM signal does not have an envelope that reproduces the modulation. This is actually an advantage: an FM receiver does not have to respond to amplitude variations, and thus it can ignore noise to some extent. Similarly, FM transmitters can use Class C amplifiers throughout, since amplitude linearity is not important. Modulation can be accomplished at low power levels.

4.2 Frequency Modulation

Figure 4.1 shows FM with a square wave modulating a sine-wave carrier. Figure 4.1(a) shows the unmodulated carrier and the modulating signal. Figure 4.1(b) shows the modulated signal in the time domain, as it would appear on an oscilloscope. The amount of frequency change has been exaggerated for clarity. The



(a) Carrier and modulating signals before modulation

(b) Modulated signal (time domain)

(c) Frequency as a function of time

(d) Phase as a function of time

Figure 4.1 Frequency modulation of a sine-wave carrier by a square wave amplitude remains as before, and the frequency changes can be seen in the changing times between zero crossings for the waveforms.

The next two sections are interesting. Figure 4.1(c) shows how the signal frequency varies with time in accordance with the amplitude of the modulating signal. Figure 4.1(d) shows how the phase changes with time. For this figure, the phase angle of the unmodulated carrier is used as a reference. When the frequency is greater than the **carrier frequency**, the phase angle gradually moves ahead, and when the frequency is lower than the carrier frequency, the phase begins to lag.

One further inference can be drawn from Figure 4.1. While the unmodulated carrier is a sine wave, the modulated signal is not—a "sine" wave that changes frequency is not really a sine wave at all. In our study of AM, we found that changing the amplitude of a sine wave generated extra frequencies called *side frequencies* or *sidebands.* This happens in FM, too. In fact, for an FM signal, the number of sets of sidebands is theoretically infinite.

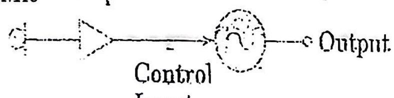


Figure 4.2 Simplified FM generator

There are many ways to generate FM, some of which will be described in the next chapter. The simplest method is to use a voltage-controlled oscillator (VCO) to generate the carrier frequency and to apply the modulating signal to the oscillator's control signal input, as indicated in Figure 4.2. As might be expected, this is a little too simple for most practical transmitters; in particular, the stability of a free-running VCO is not likely to be good enough. It does, however, give us a conceptual model to use for the time being.

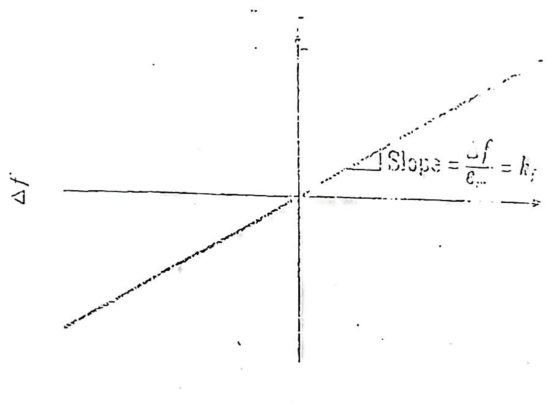
Frequency Deviation

Assume that the carrier frequency is f_c. Modulation will cause the signal frequency to vary (deviate) from its resting value. If the modulation system is properly designed, this deviation will be proportional to the amplitude of the modulating signal. Sometimes this is referred to as *linear modulation,* though no modulation process is really linear. FM can be called linear only in the sense that the graph relating instantaneous modulating-signal amplitude e_m to instantaneous frequency deviation Δf is a straight line. The slope of this line is the ratio $\Delta f/e_m$ and it represents the deviation sensitivity of the modulator, in units of hertz per volt. Let us call this constant k_f. Then

$$k_f = \frac{\Delta f}{e_m} \tag{4.1}$$

Figure 4.3 demonstrates this.

Figure 4.3 Deviation [illegible] FM modulator



Once again, remember that the frequency deviation is proportional to the amplitude, not the frequency, of the modulating signal. The number of times per second that the frequency varies from its lowest to its highest value is, of course, equal to the modulating-signal frequency.

It is possible to write an equation for the signal frequency as a function of time:

$$f_{sig}(t) = f_c + k_f e_m(t) \tag{4.2}$$

where $f_{sig}(t)$ = signal frequency as a function of time
f_c = unmodulated carrier frequency
k_f = modulator deviation constant
$e_m(t)$ = modulating voltage as a function of time

EXAMPLE 4.1 An FM modulator has $k_f = 30$ kHz/V and operates at a carrier frequency of 175 MHz. Find the output frequency for an instantaneous value of the modulating signal equal to:

(a) 150 mV (b) −2 V

Solution Equation (4.2) can be used for both parts of the question.

(a)
$$\begin{aligned} f_{sig} &= (175 \times 10^6 \text{ Hz}) + (30 \times 10^3 \text{ Hz/V})(150 \times 10^{-3} \text{ V}) \\ &= 175.0045 \times 10^6 \text{ Hz} \\ &= 175.0045 \text{ MHz} \end{aligned}$$

(b)
$$\begin{aligned} f_{sig} &= (175 \times 10^6 \text{ Hz}) + (30 \times 10^3 \text{ Hz/V})(-2 \text{ V}) \\ &= (175 \times 10^6 \text{ Hz}) - (30 \times 10^3 \text{ Hz/V})(2 \text{ V}) \\ &= 174.94 \times 10^6 \text{ Hz} \\ &= 174.94 \text{ MHz} \end{aligned}$$

In our study of AM, we found it convenient to assume a sine wave for the modulating signal. This treatment can then be generalized to cover any periodic signal using Fourier techniques. If the modulating signal is a sine wave with the equation

$$e_m(t) = E_m \sin \omega_m t \tag{4.3}$$

then Equation (4.2) becomes

$$f_{sig}(t) = f_c + k_f E_m \sin \omega_m t \tag{4.4}$$

and the peak frequency deviation (on each side of the carrier frequency) will be $k_f E_m$ Hz. Usually, the peak frequency deviation is given the symbol δ. Then

$$\delta = k_f E_m \tag{4.5}$$

where δ = peak frequency deviation in hertz
k_f = modulator sensitivity in hertz per volt
E_m = peak value of the modulating signal in volts

EXAMPLE 4.2 The same FM modulator as in the previous example is modulated by a 3 V sine wave. Calculate the deviation.

Solution Unless otherwise stated, ac voltages are assumed to be RMS. On the other hand, δ is a peak value. Therefore, the modulating voltage must be converted to a peak value before Equation (4.5) can be used.

$$\begin{aligned} E_m &= 3\sqrt{2}\text{ V} \\ &= 4.24\text{ V} \\ \delta &= k_f E_m \\ &= 30\text{ kHz/V} \times 4.24\text{ V} \\ &= 127.2\text{ kHz} \end{aligned}$$

Using frequency deviation, Equation (4.4) becomes

$$f_{sig}(t) = f_c + \delta \sin \omega_m t \tag{4.6}$$

Frequency Modulation Index

Another basic term related to FM is the frequency **modulation index** m_f (not to be confused with f_m, which is the modulating frequency). By definition, for sine-wave modulation,

$$m_f = \frac{\delta}{f_m} \tag{4.7}$$

The reason for this rather peculiar definition, which includes not only the frequency deviation but also the modulating frequency, will become apparent very soon. Meanwhile, note one other peculiarity of m_f: unlike the amplitude modulation index, which cannot exceed one, there are no theoretical limits on m_f. It can exceed one and often does.

EXAMPLE 4.3 An FM broadcast transmitter operates at its maximum deviation of 75 kHz. Find the modulation index for a sinusoidal modulating signal with a frequency of:

(a) 15 kHz (b) 50 Hz

Solution

(a)
$$\begin{aligned} m_f &= \frac{\delta}{f_m} \\ &= \frac{75\text{ kHz}}{15\text{ kHz}} \\ &= 5.00 \end{aligned}$$

(b)
$$\begin{aligned} m_f &= \frac{\delta}{f_m} \\ &= \frac{75 \times 10^3\text{ Hz}}{50\text{ Hz}} \\ &= 1500 \end{aligned}$$

Substituting E

as an equation

Compare the

4.3 Phase

In phase modula
is proportional to the
for FM, we can
phase angle to the

where k_p = phase
ϕ = p
e_m = modul

An equation can
(similar to Equati

Once again it woul

then

and

The peak phase
index. Then

EXAMPLE 4.4
age of a sine wave wou

Solution Remember

Substituting Equation (4.7) into Equation (4.6) gives

$$f_{sig}(t) = f_c + m_f f_m \sin \omega_m t \tag{4.8}$$

as an equation for the frequency of an FM signal with sine-wave modulation.

SECTION 4.2 REVIEW QUESTION

Compare the modulation index for an FM signal to that for an AM signal.

4.3 Phase Modulation

In phase modulation, it is the phase shift, rather than the frequency deviation, that is proportional to the instantaneous amplitude of the modulating signal. As we did for FM, we can define a constant for a phase modulator that relates the change in phase angle to the amplitude of the modulating signal:

$$k_p = \frac{\phi}{e_m} \tag{4.9}$$

where k_p = phase modulator sensitivity in radians per volt
ϕ = phase deviation in radians
e_m = modulating-signal amplitude in volts

An equation can be written for the phase of a PM signal as a function of time (similar to Equation (4.2) for the frequency of an FM signal):

$$\theta(t) = \theta_c + k_p e_m(t) \tag{4.10}$$

Once again it would be useful to express this in terms of sine-wave modulation. If

$$e_m(t) = E_m \sin \omega_m t$$

then

$$\phi = k_p E_m \sin \omega_m t \tag{4.11}$$

and

$$\theta(t) = \theta_c + k_p E_m \sin \omega_m t$$

The peak phase deviation, in radians, is defined as m_p, the *phase modulation index*. Then

$$\theta(t) = \theta_c + m_p \sin \omega_m t \tag{4.12}$$

EXAMPLE 4.4 A phase modulator has k_p = 2 rad/V. What RMS voltage of a sine wave would cause a peak phase deviation of 60°?

Solution Remembering that a circle has 360° or 2π rad, we see that

$$360° = 2\pi \text{ rad}$$

$$60° = \frac{2\pi \text{ rad} \times 60}{360}$$

$$= \frac{\pi}{3} \text{ rad}$$

The voltage to cause this deviation can be found from Equation (4.9):

$$k_p = \frac{\phi}{e_m}$$

$$e_m = \frac{\phi}{k_p}$$

$$= \frac{(\pi/3)\text{ rad}}{2\text{ rad/V}}$$

$$= \frac{\pi}{6}\text{ V}$$

$$= 0.524\text{ V}$$

This is peak voltage. We can find its RMS value in the usual way:

$$V_{RMS} = \frac{V_{peak}}{\sqrt{2}}$$

$$= \frac{0.524}{\sqrt{2}}$$

$$= 0.37\text{ V}$$

The Relationship between Frequency Modulation and Phase Modulation

As mentioned earlier, either FM or PM results in changes in both the frequency and phase of the modulated waveform. It has also been pointed out that frequency (in radians per second) is the rate of change of phase (in radians). That is, frequency is the derivative of phase. This leads to a relatively simple relationship between FM and PM that can make it easier to understand both and to perform calculations with either.

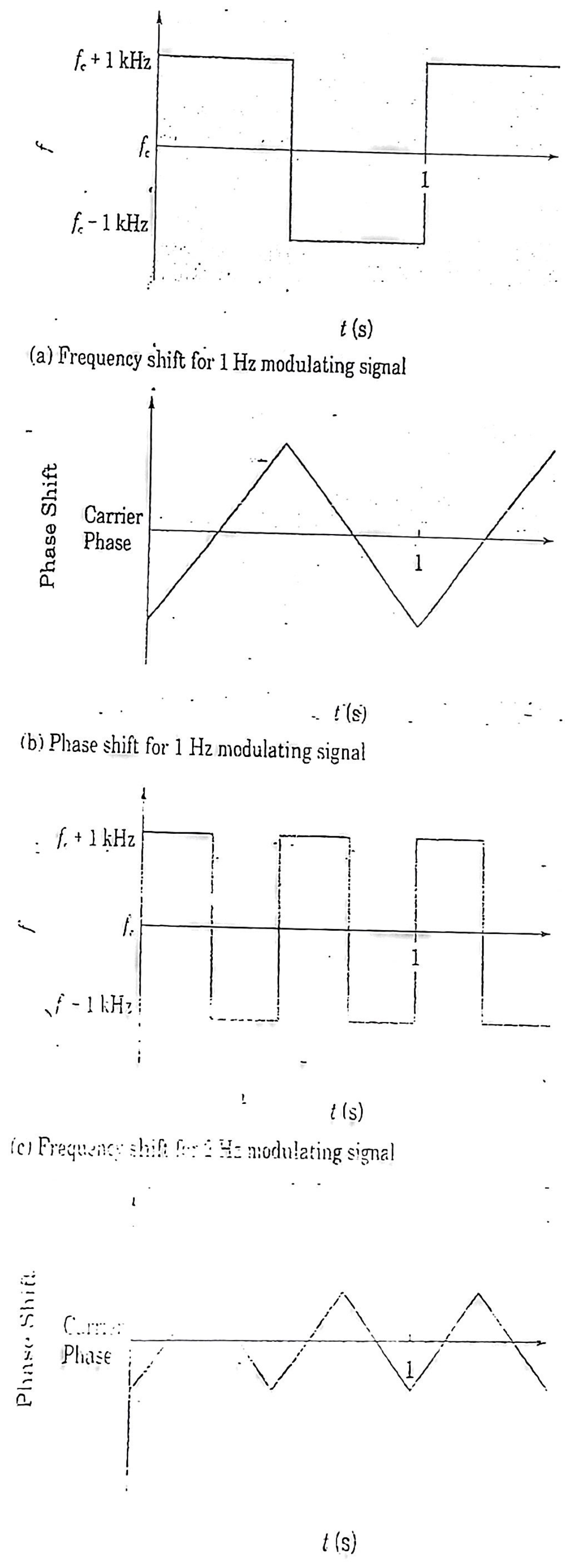
For any angle-modulated signal with sine-wave modulation, the modulation index m_p or m_f represents the peak phase deviation from the phase of the unmodulated carrier, in radians. This is obvious for PM but not quite so obvious for FM. We know that

$$m_f = \frac{\delta}{f_m}$$

That is, the modulation index (which corresponds to peak phase deviation) is proportional to frequency deviation and inversely proportional to modulating frequency. The first part of the statement sounds very reasonable. Suppose that the frequency increases. Then the higher frequency represents a phase angle that changes more quickly than before. The greater the frequency change, the greater the increase in phase angle.

The second part of the statement needs more explanation. Why should an increased modulating frequency result in a reduced change in phase angle? Consider a low modulating frequency, say 1 Hz, that causes a frequency deviation of 1 kHz. For simplicity, suppose that the modulating signal is a square wave. Then the signal frequency will increase to a point 1 kHz above the carrier frequency, stay there for one-half second, then decrease by 2 kHz to a point 1 kHz below the carrier frequency for the next one-half second. Figure 4.4(a) shows the way the instantaneous signal frequency varies.

When the signal frequency is higher than normal, the phase angle with respect to that of the unmodulated carrier steadily increases as the modulated wave gets further and further ahead of the unmodulated signal. This continues until the



(a) Frequency shift for 1 Hz modulating signal

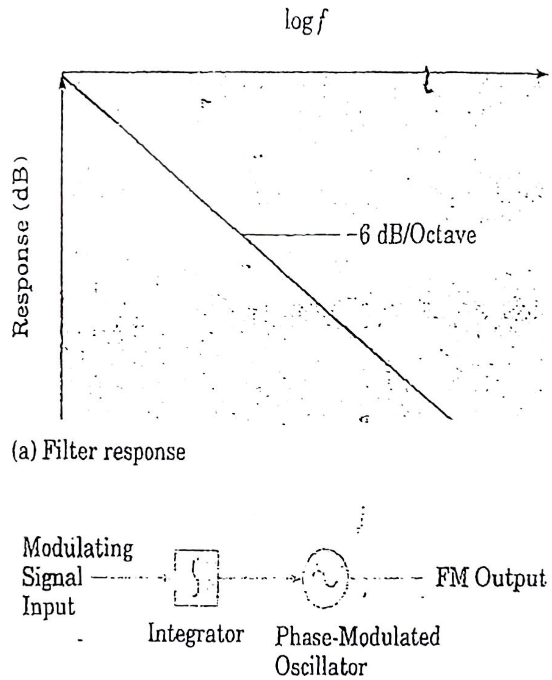
(b) Phase shift for 1 Hz modulating signal

(c) Frequency shift for 2 Hz modulating signal

(d) Phase shift for 2 Hz modulating signal

Figure 4.4 Modulating signal frequency and modulation index

Figure 4.5 Use of an integrator to convert PM to FM



(a) Filter response

(b) Use of filter

frequency decreases. At that point, the signal's phase angle starts to lag, letting the carrier phase angle catch up to and then overtake the angle of the modulated signal.

The amount of phase change is proportional to the length of time the instantaneous frequency stays above the carrier frequency, that is, the phase deviation is proportional to the period of the modulating signal. Another way of saying this is that phase deviation is inversely proportional to the modulating-signal frequency. Figures 4.4(c) and (d) show the frequency and phase change, respectively, for a modulating signal with the same amplitude as before but twice the frequency. The phase deviation is only one-half as much as before.

This simple relationship between FM and PM suggests that it would be easy to convert one to the other, and this is true. For instance, a phase modulator can be used to generate FM. The baseband signal is passed through a low-pass filter with the frequency response shown in Figure 4.5(a). This type of filter is often referred to as an *integrator*. For equal amplitudes at the modulator input, the amplitude of the output from the filter will be inversely proportional to the modulating frequency. Figure 4.5(b) shows the low-pass filtered signal applied to a phase modulator. The phase-modulated output has a modulation index inversely proportional to the modulating frequency, that is, it is identical to FM!

EXAMPLE 4.5 An FM communications transmitter has a maximum frequency deviation of 5 kHz and a range of modulating frequencies from 300 Hz to 3 kHz. What is the maximum phase shift that it produces?

Solution We now know that the peak phase shift in radians is equal to the frequency modulation index m_f. Since, by Equation (4.7),

$$m_f = \frac{\delta}{f_m}$$

m_f will be largest for the lowest possible value of f_m, in this case 300 Hz. For this frequency, the phase shift is

$$\begin{aligned}\phi_{max} &= m_f\\ &= \frac{\delta}{f_m}\\ &= \frac{5000}{300}\\ &= 16.7 \text{ rad}\end{aligned}$$

EXAMPLE 4.6 A phase modulator has a sensitivity of $k_p = 3$ rad/V. How much frequency deviation does it produce with a sine-wave input of 2 V peak at a frequency of 1 kHz?

Solution The maximum phase shift is easily found from Equation (4.11):

$$\phi = k_p E_m \sin \omega_m t$$

The maximum value of ϕ is m_p, and it occurs for the peak modulating voltage.

$$\begin{aligned}m_p &= \phi_{max}\\ &= k_p E_m\\ &= 3 \text{ rad/V} \times 2 \text{ V}\\ &= 6 \text{ rad}\end{aligned}$$

This has the same value as m_f, if the signal is considered as frequency modulation. From Equation (4.7),

$$m_f = \frac{\delta}{f_m}$$

$$\begin{aligned}\delta &= m_f f_m\\ &= 6 \times 1 \text{ kHz}\\ &= 6 \text{ kHz}\end{aligned}$$

SECTION 4.3 REVIEW QUESTION

Suppose the modulating-signal frequency increases but its amplitude remains constant. Explain how the modulation index varies for both FM and PM.

4.4 The Angle Modulation Spectrum

Angle modulation produces an infinite number of sidebands, even for single-tone modulation. These sidebands are separated from the carrier by multiples of f_m, but their amplitude tends to decrease as their distance from the carrier frequency increases. Sidebands with amplitudes less than about 1% of the total signal voltage can usually be ignored, so for practical purposes an angle-modulated signal can be considered to be band-limited. In most cases, though, its bandwidth is much larger than that of an AM signal.

Bessel Functions

FM and PM signals have similar equations. In fact, without a knowledge of the baseband signal, it would be impossible to tell them apart. For the modulation of a carrier with amplitude A and radian frequency ω, by a single-frequency sinusoid, the equation is of the form

$$v(t) = A \sin(\omega_c t + m \sin \omega_m t) \tag{4.13}$$

The factor m can represent m_f for FM or m_p for PM.

This equation cannot be simplified by ordinary trigonometry, as is the case for amplitude modulation. About the only useful information that can be gained by inspection is the fact that the signal amplitude remains constant regardless of the modulation index. This observation is significant, since it demonstrates one of the major differences between AM and FM or PM, but it provides no information about the sidebands.

This signal can be expressed as a series of sinusoids by using Bessel functions of the first kind. Proving this is beyond the scope of this text, but it can be done. The Bessel functions themselves are rather tedious to evaluate numerically, but that, too, has been done. Some results are presented in Table 4.1 and Figure 4.6.

The table and graph of Bessel functions represent normalized voltages for the various frequency components of an FM or PM signal, that is, the numbers in the tables represent actual voltages if the unmodulated carrier has an amplitude of 1 V. Here J_0 represents the amplitude of the component at the carrier frequency (sometimes called the rest frequency). The variable J_1 represents the amplitude of each of the first set of sidebands, which have frequencies of $(f_c + f_m)$ and $(f_c - f_m)$. J_2 represents the amplitude of each of the second pair of sidebands, at $(f_c + 2f_m)$ and $(f_c - 2f_m)$, and so on. Figure 4.7 shows this on a frequency-domain plot. All of the Bessel terms should be multiplied by the voltage of the unmodulated carrier to find the actual sideband amplitudes. Of course, the Bessel coefficients are equally valid for peak or RMS voltages, but the user should be careful to keep track of which type of measurement is being used. There certainly will be a difference when sideband power calculations are involved.

When Bessel functions are used, the signal of Equation (4.13) becomes

$$\begin{aligned} v(t) &= A \sin(\omega_c t + m \sin \omega_m t) \\ &= A\{J_0(m) \sin \omega_c t \\ &\quad - J_1(m)[\sin(\omega_c - \omega_m)t - \sin(\omega_c + \omega_m)t] \\ &\quad + J_2(m)[\sin(\omega_c - 2\omega_m)t + \sin(\omega_c + 2\omega_m)t] \\ &\quad - J_3(m)[\sin(\omega_c - 3\omega_m)t + \sin(\omega_c + 3\omega_m)t] \\ &\quad + \cdots\} \end{aligned} \tag{4.14}$$

With angle modulation, the total signal voltage and power do not change with modulation. Therefore, the appearance of power in the sidebands indicates that the power at the carrier frequency must be reduced below its unmodulated value in the presence of modulation. In fact, the carrier frequency component disappears for certain values of m (for example, 2.4 and 5.5).

This constant-power aspect of angle modulation can be demonstrated using the table of Bessel functions. For simplicity, normalized values can be used. Let the unmodulated signal have a voltage of 1 V RMS across a resistance of 1 Ω. Its power is, of course, 1 W. When modulation is applied, the carrier voltage is