Frequency Modulation



Objectives Upon completion of this chapter, the student should be able to:

- Discuss the differences between AM and angle modulation.
- Explain the advantages and disadvantages of the different analog modulation techniques.
- Discuss the relationship between FM and PM.
- Calculate modulation index, signal bandwidth, sideband frequencies and power levels, and be able to sketch the frequency spectra of an FM signal.
- Explain the effect of noise upon an FM signal and relate its effect to deviation and bandwidth.
- Discuss the typical FM transmitter and explain the difference between direct and indirect generation of FM.
- Discuss the differences between AM and FM superheterodyne receivers, including capture effect, noise threshold, and pre-emphasis and de-emphasis.
- Discuss FM demodulation techniques, including the use of a phase-locked loop (PLL).
- Describe FM stereo broadcasting and discuss typical FM-receiver specifications.
- Outline 4.1 An Introduction to the Development of FM
 - 4.2 Frequency-Modulation Theory
 - 4.3 Mathematical Analysis of FM
 - 4.4 FM Signal Bandwidth
 - 4.5 FM Power Relations
 - 4.6 The Effect of Noise on FM

- 4.7 FM Generation
- 4.8 FM Receivers
- 4.9 FM Stereo

Key Terms angle modulation Bessel functions capture effect de-emphasis deviation sensitivity direct FM frequency deviation index of modulation indirect FM modulation sensitivity narrow-band FM

pre-emphasis quieting sensitivity rest frequency threshold effect wideband FM

Introduction

This chapter will introduce the student to the other forms of analog modulation—frequency modulation (FM) and phase modulation (PM)— both of which are commonly known as angle modulation. Both FM and PM are used extensively in communications systems. FM is used in radio broadcasting, for the transmission of the sound signal in standard (NTSC) TV, for private land-mobile radio systems, for direct-satellite broadcasting, and for cordless and cellular telephone systems, just to name a few common applications. PM by itself and in combination with AM is used extensively in modern data-communications systems. Angle modulation has a very important advantage over AM in its ability to provide increased immunity to noise. Angle-modulation systems typically require a larger bandwidth than AM systems, a necessary trade-off for its improved resistance to noise.

Topics covered in this chapter will include:

- theory of FM operation
- frequency deviation
- modulation index
- theory of PM operation
- a comparison of FM and PM
- Bessel functions
- the spectrum of an angle-modulated signal
- signal bandwidth
- sideband power relations
- noise effects on FM

- FM-stereo multiplex operation
- frequency-modulation transmission and reception hardware

A system familiar to most readers of this chapter—standard FM broadcasting, a mature technology—will serve as the vehicle used to present this material. Finally, FM-stereo multiplex operation will be introduced as an example of an early form of combinational modulation used to send more than one signal over the same channel (multiplexing).

4.1 An Introduction to the Development of FM

This chapter will introduce the reader to the second analog form of modulation. This type of modulation scheme is known as angle modulation. **Angle modulation** can be further subdivided into two distinct types: frequency modulation (FM) and phase modulation (PM).

The history and evolution of angle modulation basically revolves around one man, Major Edwin Armstrong. Armstrong, a radio pioneer who invented first the regenerative and then the superheterodyne receiver in the 1910s, worked on the principles of frequency and phase modulation starting in the 1920s. It was not until the 1930s, however, that he finally completed work on a practical technique for wideband frequencymodulation broadcasting. For further information, visit a Web site devoted to Armstrong's work at http://users.erols.com/oldradio/.

As a historical footnote, it should be pointed out that at the turn of the last century, the very early Paulson arc transmitter actually used the simplest form of FM, frequency-shift keying (FSK), to transmit a wireless telegraph signal. With this type of wireless transmitter, a continuous electrical arc would have its fundamental output frequency altered by closing a telegraph key. When the key was closed, it would short out several turns of a tuning inductor, thus changing the transmitter output frequency. For this reason it was a form of FSK.

Despite Armstrong's efforts, the implementation of FM broadcasting was fought by RCA and NBC through 1945, only becoming popular in the United States during the late 1960s and early 1970s when technological advances reduced the cost of equipment and improved the quality of service. Many public-safety departments were early adopters of FM for their fleet communications. AMPS cellular-telephone service, an FM-based system, was introduced in the United States in 1983. Today FM is used for the legacy FM broadcast band, standard TV-broadcasting sound transmission, direct-satellite TV service, cordless telephones, and just about every type of business band and mobile-radio service. FM is capable of much more noise immunity than AM, and is now the most popular form of analog modulation.

4.2 Frequency-Modulation Theory

We will start our discussion of angle modulation by first examining frequency modulation. The classic definition of *FM* is that the instantaneous output frequency of a transmitter is varied in accordance with the modulating signal. Recall that we can write an equation for a sine wave as follows:

$$e(t) = E_P \sin(\omega t + \phi)$$
 4.1

While amplitude modulation is achieved by varying E_p , frequency modulation is realized by varying ω in accordance with the modulating signal or message. Notice that one can also vary ϕ to obtain another form of angle modulation known as phase modulation (PM). Later we will examine the relationship between FM and PM. See Figure 4-1 for a time display of a typical FM signal.





Definitions

An important concept in the understanding of FM is that of **frequency deviation**. The amount of frequency deviation a signal experiences is a measure of the change in transmitter output frequency from the **rest frequency** of the transmitter. The rest frequency of a transmitter is defined as the output frequency with no modulating signal applied. For a transmitter with linear modulation characteristics, the frequency deviation of the carrier is directly proportional to the amplitude of the applied modulating signal. Thus an FM transmitter is said to have a **modulation sensitivity**, represented by a constant, k_{fr} of so many kHz/V,

 k_f = frequency deviation/V = k_f kHz/V

For a single modulating tone of $e_M(t) = e_M \sin(\omega_M t)$, the amount of frequency deviation is given by

 $\delta(t) = k_f \times e_M(t)$

where $\delta(t)$ is the instantaneous frequency deviation and $e_M(t)$ represents the modulating signal. The peak deviation is given by

$$\delta = k_f \times E_M \tag{4.2}$$

where both δ and E_M are peak values.

•—EXAMPLE 4.1

A certain FM transmitter has a modulation sensitivity, k_{f} , of 10 kHz/V. If a 5-kHz sine wave of 2 V_{p-p} is applied to this transmitter, determine the frequency deviation that occurs.

•-Solution The applied modulating sine wave varies between ± 1 V at a rate of 5000 times per second; therefore, the output of the transmitter will appear to deviate ± 10 kHz from its rest frequency at a rate of 5000 times per second. The term *appear* is used here to indicate that one would most likely envision a time display of the signal to visualize this operation. Mathematically, this is represented as

 $\delta = 10 \text{ kHz/V} \times E_M = \pm 10 \text{ kHz}$

It is instructive to pause and think about the FM process at this time. As the example above indicates, for frequency modulation, the amount of frequency deviation is proportional to the amplitude of the applied modulating signal, similarly, the number of times per second that the transmitter signal deviates is equal to the frequency of the applied modulating signal. Note that the amplitude of the FM wave remains constant during the modulation process.

FM versus PM

Let us return to equation 4.1, repeated here:

 $e(t) = E_p \sin(\omega t + \phi)$

As previously mentioned, it is possible to vary both ω (FM) and ϕ (PM) in the above equation. Each term is part of the argument of the sine wave. So, what is the difference between varying one versus the other? A simple answer is that there is no difference, as either one will change the sine wave's frequency. However, closer inspection and the employment of more mathematical rigor reveal that there are some subtle differences between the two

forms of angle modulation. However, due to the complex mathematics involved these differences will not be discussed here. Practically speaking, it is possible to obtain FM from PM, as depicted in Figure 4-2; but most present-day FM systems do not generate FM by this method. This process of generating FM is known as **indirect FM**.



If we compare the two forms of modulation in the time domain, shown in Figures 4-3a and 4-3b on page 145, we would observe that the FM wave and PM wave appear quite similar; however, their timing appears to be out of sync. Indeed, the FM wave has its maximum frequency deviation during the peaks of the input signal, while the PM wave has its maximum frequency deviation during zero crossings of the input signal. Without showing the phase relationship of the input wave to the modulated wave, it would be impossible to tell the difference between the two forms of angle modulation if one simlpy looked at the resulting modulated waveform.



The Subtle Difference between FM and PM

One way to tell the difference between FM and PM is to observe the following: If the instantaneous *frequency* of the signal is directly proportional to







the amplitude of the input signal, it is FM. On the other hand, if the instantaneous *phase* of the signal is proportional to the amplitude of the input signal, it is PM. This last statement, although correct, is unclear because the term *instantaneous phase* is undefined at present. Another way of expressing this last statement is to say that for PM, the transmitter output frequency is at the rest frequency when the input signal is at either its most positive or most negative voltage. The subtle difference between FM and PM is not really very important, for the vast majority of wireless angle modulation transmitters use FM. That being the case, our comments will focus almost entirely on FM from this point on. Several different forms of PM will be discussed in much more detail in chapter 6 in relation to digital modulation.

4.3 Mathematical Analysis of FM

As was done with AM, a mathematical analysis of a high-frequency sine wave, modulated by a single tone or frequency, will be used to yield information about the frequency components in an FM wave, FM power relations, and the bandwidth of an FM signal.

From the definition of frequency deviation, an equation can be written for the signal frequency of an FM wave as a function of time:

$$f_{\text{signal}} = f_C + k_f e_M(t) = f_C + k_f E_M \sin \omega_M t$$
4.3

and substitution of $\delta = k_f \times E_M$ yields:

$$f_{\text{signal}} = f_C + \delta \sin \omega_M t \tag{4.4}$$

But what does this equation indicate? It seems to be saying that the frequency of the transmitter is varying with time. This brings up the same type of problem that was observed when we looked at a time display of AM and then performed a mathematical analysis in an attempt to determine its frequency content. With AM, the signal appeared to be a sine wave thats amplitude was changing with time. At the time, it was pointed out that a sine wave, by definition, has a constant peak amplitude, and thus cannot have a peak amplitude that varies with time. What about the sine wave's frequency? It also must be a constant and cannot be varying with time. As was the case with AM, where it turned out that our modulated wave was actually the vector sum of three sine waves, a similar situation is true for FM. An FM wave will consist of three or more frequency components vectorially added together to give the appearance of a sine wave thats frequency is varying with time when displayed in the time domain.

A somewhat complex mathematical analysis will yield an equation for the instantaneous voltage of an FM wave of the form shown here:

$$e_{\rm FM}(t) = E_C \sin(\omega_C t + m_f \sin \omega_M t)$$
4.5

where E_C is the rest-frequency peak amplitude, ω_C and ω_M represent the rest and modulating frequencies, and m_f is the index of modulation.

This equation represents a single low-frequency sine wave, f_M , frequency modulating another high-frequency sine wave, f_C . Note that this equation indicates that the argument of the sine wave is itself a sine wave.

The Index of Modulation

The index of modulation, $m_{f'}$ is given by the following relationship:

$$m_f = \frac{\delta}{f_M}$$
 4.6

A few more comments about the index of modulation, m_f , are appropriate. As can be seen from the equation, m_f is equal to the peak deviation caused when the signal is modulated by the frequency of the modulating signal; therefore, m_f is a function of both the modulating signal amplitude and frequency. Furthermore, m_f can take on any value from 0 to infinity. Its range is not limited as it is for AM.

Percentage of Modulation

At this point the question might be asked, Is there anything in the FM process similar to the percentage of modulation of an AM signal? The answer is *yes*. However, unlike AM, it has nothing to do with the index of modulation. The practical implementation of FM communication systems in a limited bandwidth-channel environment, such as cellular radio,

requires a limitation upon the maximum frequency deviation to prevent adjacent channel interference. For example, the FCC's Rules and Regulations limit FM broadcast-band transmitters to a maximum frequency deviation of ± 75 kHz. The maximum allowable deviation will be assigned the value of 100% modulation. Therefore, in equation form, the percentage of modulation is given by:

% Modulation =
$$\frac{\delta}{\delta_{\text{max}}} \times 100\%$$
 4.7

EXAMPLE 4.2

An FM broadcast-band transmitter has a peak deviation of ± 60 kHz for a particular input signal. Determine the percentage of modulation.

•-Solution From the equation,

% Modulation = $\frac{\pm 60 \text{ kHz}}{\pm 75 \text{ kHz}} \times 100\%$

% Modulation = $.8 \times 100\% = 80\%$

If one were to visit a local FM radio station, this would be the reading that is monitored in the studio by the on-the-air employee.

Bessel Functions and their Relationship to FM

We now return to our mathematical analysis of the FM wave. Equation 4.5 cannot be solved with algebra or trigonometric identities. However, certain Bessel-function identities are available that will yield solutions to equation 4.5 and allow us to determine the frequency components of an FM wave. As a note of explanation, **Bessel functions** appear as solutions in numerous physical problems, quite often involving cylindrical or spherical geometries.

Without going through the rather detailed mathematics to solve equation 4.5, the resulting equation is shown here. It itemizes the various signal components in an FM wave and their amplitudes.

$$e_{FM}(t) = E_C \begin{cases} J_0(m_f)\sin\omega_C t - J_1(m_f) \left[\sin(\omega_C - \omega_M)t - \sin(\omega_C + \omega_M)t\right] + \\ J_2(m_f) \left[\sin(\omega_C - 2\omega_M)t + \sin(\omega_C + 2\omega_M)t\right] - \\ J_3(m_f) \left[\sin(\omega_C - 3\omega_M)t - \sin(\omega_C + 3\omega_M)t\right] + \\ J_4(m_f) \left[\sin(\omega_C - 4\omega_M)t + \sin(\omega_C + 4\omega_M)t\right] - \dots \end{cases}$$

$$4.8$$

What this equation indicates is that there are an infinite number of sideband pairs for an FM wave. Each sideband pair is symmetrically located about the transmitter's rest frequency, f_C , and separated from the rest frequency by integral multiples of the modulating frequency, $n \times f_M$, where n = 1, 2, 3, The magnitude of the rest frequency and sideband pairs is dependent upon the index of modulation, m_f , and given by the Besselfunction coefficients, $J_n(m_f)$, where the subscript n of J_n is the order of the sideband pair and m_f is the modulation index. Note that $J_n(m_f)$ is all one term and not the product of two numbers.

Several examples might provide some insight to the meaning of $J_n(m_f)$:

- $J_0(1.0)$ represents the rest-frequency amplitude of an FM wave with an index of modulation equal to 1.0.
- $J_1(2.5)$ is the amplitude of the first pair of sidebands for an FM wave with $m_f = 2.5$.
- $J_7(m_f)$ is the amplitude of the seventh pair of sidebands with an unknown index of modulation, m_f .

Because most technology students have never dealt with Bessel functions before, equation 4.8 looks extremely complex and difficult to work with. However, after several examples, dealing with the equation should become a less daunting task.

Before considering the examples, let us see how we determine the value of the term $J_n(m_f)$. From very complex mathematics, the values of the $J_n(m_f)$ terms can be calculated from an "infinite series." Therefore, the results of the numerical computation of the values of $J_0(m_f)$, $J_1(m_f)$, $J_2(m_f)$, and so forth are usually plotted on a graph like the one shown in Figure 4-4.

There are several points that can be made about the Bessel functions plotted on the graph. For small values of m_f , the only Bessel functions with any significant amplitude are $J_0(m_f)$ and $J_1(m_f)$ (the rest frequency and the first sideband pair), while the amplitude of the higher-order (n > 1) sideband pairs is very small. As m_f increases, the amplitude of the rest frequency decreases and the amplitude of the higher-order sidebands increases, which would seem to indicate an increasing signal bandwidth. Further inspection of the graph indicates that as m_f keeps increasing, the sideband pairs are essentially zero amplitude until about $m_f = n$, at which point they increase in amplitude to a maximum and then decrease again. In all cases, as m_f keeps increasing, each Bessel function appears to act like an exponentially decaying sine wave. Therefore, the amplitudes of the higher-order sideband pairs eventually approach zero.

An extremely interesting point is also observed about the Bessel-function amplitudes from the graph. In all cases, including the rest frequency $J_0(m_f)$, the amplitude of the Bessel function goes to zero for numerous values of m_{fr} meaning that the rest-frequency component of the FM wave can disappear. This fact also holds for each pair of sidebands.



Bessel-Function Table

The information from the graph is usually put into table form for integer or fractional values of m_{f} . A table of Bessel-function values is shown in Table 4-1 on page 150. Note that amplitude values with minus signs represent phase shifts of 180 degrees and that amplitude values less than 0.01 have been left out of the table because they represent component frequencies with insignificant power content.

Examples of Different Values of Index of Modulation

At this time, several examples of frequency displays of FM waves with different values of m_f will be shown.

Figure 4-5 on page 151 shows the frequency-domain display of the resultant FM signal resulting from a modulating signal of 10 kHz, an index of modulation of 0.25, and rest frequency of 500 kHz. For this case, there is only one pair of sidebands with appreciable power. This type of FM signal meets the strict definition of **narrow-band FM** (NBFM), where $m_f \leq 0.5$. The bandwidth used by an NBFM signal is approximately equal to that of an AM signal. FM transmitters commonly used by business band and other mobile FM radio services for voice transmission will typically have 150

x	Bessel-function order, n																
m _f	Jo	J ₁	J_2	J ₃	J ₄	J ₅	J ₆	J ₇	J ₈	J,	J ₁₀	J ₁₁	J ₁₂	J ₁₃	J ₁₄	J ₁₅	J ₁₆
0.00	1.00		_														
0.25	0.98	0.12															
0.5	0.94	0.24	0.03														
1.0	0.77	0.44	0.11	0.02													
1.5	0.51	0.56	0.23	0.06	0.01			_						_	_		_
2.0	0.22	0.58	0.35	0.13	0.03												
2.41	0	0.52	0.43	0.20	0.06	0.02											
2.5	05	0.50	0.45	0.22	0.07	0.02	0.01										
3.0	26	0.34	0.49	0.31	0.13	0.04	0.01										
4.0	40	07	0.36	0.43	0.28	0.13	0.05	0.02						_	_		_
5.0	18	33	0.05	0.36	0.39	0.26	0.13	0.05	0.02					_	_		_
5.53	0	34	13	0.25	0.40	0.32	0.19	0.09	0.03	0.01				_	_		_
6.0	0.15	28	24	0.11	0.36	0.36	0.25	0.13	0.06	0.02				_	_		_
7.0	0.30	0.00	30	17	0.16	0.35	0.34	0.23	0.13	0.06	0.02						
8.0	0.17	0.23	11	29	10	0.19	0.34	0.32	0.22	0.13	0.06	0.03					
8.65	0	0.27	0.06	24	23	0.03	0.26	0.34	0.28	0.18	0.10	0.05	0.02				
9.0	09	0.25	0.14	18	27	06	0.20	0.33	0.31	0.21	0.12	0.06	0.03	0.01			
10.0	25	0.04	0.25	0.06	22	23	01	0.22	0.32	0.29	0.21	0.12	0.06	0.03	0.01		_
12.0	0.05	22	08	0.20	0.18	07	24	17	0.05	0.23	0.30	0.27	0.20	0.12	0.07	0.03	0.01

T

 TABLE 4-1
 A table of Bessel Functions of the first kind

MU-04.qxd 10/23/02 10:37 AM Page 150

FIGURE 4-5 An SA display for $m_f = 0.25$ and $f_M = 10$ kHz



maximum frequency deviations of less than 10 kHz and $m_f \ge 0.5$. These applications do not meet the strict definition of NBFM, though they are generally referred to as NBFM systems.

For the example of an FM signal shown in Figure 4-6, with $m_f = 1.0$, there are several pairs of sidebands with appreciable power.

FIGURE 4-6 An SA display of $m_f = 1$ and $f_M = 10$ kHz



4.4 FM Signal Bandwidth

At this time, we should consider how to measure or determine the bandwidth of FM signals. From the SA display shown in Figure 4-6, we saw that there are three pairs of sidebands, each spaced 10 kHz apart. There are many more sideband pairs for this signal, but their amplitudes (and therefore their power content) are negligible. The total bandwidth is given by:

Bandwidth = $f_M \times \#$ of sideband pairs $\times 2$

4.9

or, for this example,

Bandwidth = $10 \text{ kHz} \times 3 \times 2 = 60 \text{ kHz}$.

This is three times the bandwidth of an AM signal having the same modulating tone. The bandwidth of an FM signal is usually determined by the number of significant sidebands. For the example just cited, over 99% of the signal power is contained in the three pairs of sidebands. If one uses the values shown in Table 4-1, any sideband pair with a relative amplitude less than .05 could be ignored, and the other frequency components would still contain at least 99.5% of the total power in the FM signal.

We might be interested in how our SA displays correlate with the mathematical analysis presented earlier. The plot of the Bessel functions shown as Figure 4-4 and the table of Bessel-function values (Table 4-1) are both normalized to 1.0. This would be the relative amplitude of the rest frequency with no modulation present. For a transmitter with no modulation, the output power at the rest frequency would be equal to the power transmitter, P_{trans} , and the signal amplitude (rms value) could be found from

$$E_{\text{signal}} = \left(\frac{P_{\text{trans}}}{R}\right)^{1/2}$$
 4.10

Figure 4-7 was generated by the following sequence of steps: First, the modulating signal amplitude was reduced to zero and the rest-frequency output voltage set to a convenient value (100 mV), which also becomes the reference level for the spectrum analyzer. Then the signal was modulated at $m_f = 1.0$ ($\delta = 10$ kHz and $f_M = 10$ kHz). The sideband and rest-frequency levels were measured using the spectrum-analyzer marker function.

From the display in Figure 4-7 it can be seen that the rest-frequency component has an amplitude of 77 mV, the first pair of sidebands is about 44 mV, and the amplitude of the second pair of sidebands is approximately 11 mV (from marker measurements). Table 4-2 shows the values for $J_n(m_f)$ when n = 0, 1, 2, and 3 and $m_f = 1.0$. Values are shown from Table 4-1 and those measured directly from the SA display. How do these values compare?

The values from the table and those measured with the SA are within measurement error and should help to confirm the theoretical analysis of

FREQUENCY MODULATION • 153

FIGURE 4-7 An SA display of $m_f = 1$ with a linear vertical scale. The reference level of the SA (top line) is set to 100 mV.



TABLE 4-2 Comparison of $J_n(m_i)$ values from Table 4-1 and direct SA measurement

Amplitude	$J_0(m_f)$	$J_1(m_f)$	$J_2(m_f)$	$J_3(m_f)$	
From Table	.77	.44	.11	.02	
From SA	.7703	.4402	.1149	.01956	

FM presented earlier. The next section will show how to calculate the power in a particular sideband or at the rest frequency of an FM wave.

4.5 FM Power Relations

Recall that for an FM wave the amplitude of the signal, and hence the power, remains constant. This means that the power in the individual frequency components of the wave must add up to the transmitter output power. Furthermore, if the modulation index changes, the total power must redistribute itself over the resulting frequency components.

If there is no modulation, then $m_f = 0$ and $J_0 = 1.0$. Mathematically, this can be shown by the following:

 $P_{\text{rest freq}} = J_0^2 \times P_{\text{trans}}$

or

 $P_{\text{rest freq}} = P_{\text{trans}}$ for $m_f = 0.0$.

To determine the power for any individual frequency component, we can use the following relation:

$$P_n = J_n^2(m_f) \times P_{\text{trans}}$$

$$4.11$$

Furthermore, the total signal power will be given by:

$$P_{\text{total}} = (J_0^2 + 2J_1^2 + 2J_2^2 + 2J_3^2 + \dots) \times P_{\text{trans}}$$
4.12

EXAMPLE 4.3

1

An FM transmitter has a power output of 10 W. If the index of modulation is 1.0, determine the power in the various frequency components of the signal.

•-Solution From the row for $m_f = 1.0$ in Table 4-1, we have the following:

 $J_0 = .77$, $J_1 = .44$, $J_2 = .11$, and $J_3 = .02$

Using equation 4.11,

 $P_0 = J_0^2 (P_{\text{trans}}) = (.77)^2 \times 10 = 5.929 \text{ W}$

at the rest frequency,

Similarly,

 $P_1 = 1.936 \text{ W}, P_2 = 0.121 \text{ W}, \text{ and } P_3 = 0.004 \text{ W}$

The total power in the FM wave is the sum of all the powers of the different frequency components, therefore,

$$P_{\text{total}} = P_0 + 2P_1 + 2P_2 + 2P_3 = 5.929 + 2(1.936) + 2(.121) + 2(.004)$$

= 10.051 W

Rounding error accounts for the extra 51 mW.

Let us compare our calculations to an SA display of $m_f = 1.0$. Figure 4-8 is an appropriate SA display to use. After normalizing the rest-frequency power to the reference level, the index of modulation was set to $m_f = 1.0$ and the power of each frequency component was measured relative to the reference level using the SA marker function. This time the vertical scale is logarithmic.

One might want to compare the values measured in Figure 4-8 with those that can be calculated using equation 4.11. In each case, the signal level of the frequency components have been compared to the reference level in terms of their value of dB down from the reference. See Table 4-3.

Again, the calculated and measured values are well within measurement error and rounding error from the table values. This comparison should

FREQUENCY MODULATION • 155



TABLE 4-3Acomparison ofcalculated andmeasured powerlevels below thereference

dB down from the reference	$J_0(m_f = 1)$	$J_{1}(m_{f} = 1)$	$J_2(m_f = 1)$	$J_{3}(m_{f} = 1)$
Calculated dB down	–2.27 dB	–7.13 dB	–19.17 dB	–33.98 dB
Measured dB down	–2.271 dB	–7.145 dB	–18.83 dB	-34.31 dB

help to confirm that the mathematical theory previously presented about FM is indeed an accurate portrayal of what is actually happening in the process know as FM!

Wideband FM

Let us look at an FM wave with a large value of $m_{f'}$ See Figure 4-9 on page 156. For this signal, with $m_f = 20$, there are too many pairs of sidebands for the SA to resolve. This is a wideband signal and would fall into the category of what is known as **wideband FM** (WBFM). The modulating signal is 1.5 kHz and the bandwidth of the FM signal is approximately 60–70 kHz, *if* all frequencies are included that are less than approximately 20 dB down (-20 dB) from the apparent rest-frequency amplitude. What occurs if the modulating signal is a voice or music signal? Figure 4-10 on page 156 shows a typical SA display of an FM signal produced by a modulating music signal. There is a continuous range of frequencies in a music signal



and the resultant FM signal contains, for all practical purposes, a continuous range of sidebands. Figure 4-10 has a bandwidth of approximately 150 kHz. Again, to determine the signal bandwidth we have included all frequency components with an amplitude less than 20 dB down from the

FIGURE 4-10 Ref 0 dBm Atten 10 dB An FM signal Peak spectrum. The Log modulating 10 signal is music. dB/ W1 S2 **S3 FS** ΛΛ Start 300 kHz Stop 700 kHz #Res BW 10 kHz VBW 10 kHz Sweep 27.5 ms (401 pts)

apparent rest-frequency amplitude. One might be tempted to ask, Why is the SA display not symmetrical? The answer lies in the finite time it takes for the analyzer trace to sweep across the screen. During the time required for the trace sweep (27.5 msec, in this case) the frequency content and amplitude of the music has changed significantly, yielding changing sideband pairs during the sweep time when the SA is acquiring its data points.

FM Rest Frequency and Sideband Nulls

As mentioned before, an interesting aspect of FM is that for certain values of modulation index the rest-frequency component of the FM wave can disappear! See Table 4-4, Column 2 for Bessel-function "zeros" of the rest frequency. A term given to the m_f zero values is *eigenvalues*.

TABLE 4-4Zeros ofthe Bessel Functions

Number of Zero	J ₀ (m _f)	J ₁ (m _f)	J ₂ (m _f)	J ₃ (m _f)
0	2.41	3.83	5.14	6.38
1	5.53	7.00	8.42	9.76
2	8.65	10.17	11.62	13.02
3	11.79	13.32	14.80	16.22
4	14.93	16.47	17.96	19.41

This fact also extends to each pair of sidebands, as can be seen by looking at the Bessel-function graph that was shown in Figure 4-3. The zeros of the first three sideband pairs are tabulated in Table 4-4 in the columns to the right of the rest-frequency zeros given by $J_0(m_f)$. Of what importance is this phenomena? This information can be used to calibrate or determine the **deviation sensitivity**, k_f of an FM transmitter using standard test equipment. Recall the relationship between the index of modulation and frequency deviation:

$$m_f = \frac{\delta}{f_M}$$
 or $\delta = m_f \times f_M$

Using a spectrum analyzer, one can set the value of m_f to one of its eigenvalues by adjusting for a null in the amplitude of a frequency component. With an accurate frequency generator or frequency counter, one can set or measure the value of f_M and calculate the deviation sensitivity, k_f , of the transmitter in kHz/V. An example of a rest-frequency null or zero is shown in Figure 4-11 on page 158.



As can be seen in the SA display, the frequency component at $J_0(m_f)$, the rest frequency has been "nulled" out. For this display, the modulation frequency was set to 10 kHz and the frequency deviation was set to 24.06 kHz, hence $m_f \cong 2.406$ and $J_0 \cong 0$. For any FM system, if one starts with $m_f = 0$ and gradually increases the index of modulation by raising the input-signal level, the rest-frequency component will go through a series of nulls as m_f becomes larger. The same effect will happen to the first sideband pair, then the second pair, and so on. See Figure 4-12 for an example of a sideband null.

FM Services

Before looking at the hardware used to produce and receive FM, let us examine several FM services. The first example will be the legacy FM broadcast band. Figure 4-13 is an SA display of the FM broadcast band from 88 to 108 MHz as received off air. This band uses 200-kHz channel assignments starting at 88.1 MHz and ending at 107.9 MHz. Guard bands exist between adjacent channels, and the FCC's rules regarding assignment of station frequencies will preclude any co-channel or adjacent-channel interference.

The next example is the FM business band in the 150- to 174-MHz range. Figure 4-14 on page 160 shows an SA display of the off-air signals in





this band. These channels are only 30-kHz wide and the transmitter deviation is limited. As users activate their microphones, these off-air signals are displayed as blips of RF energy on the SA display.

FIGURE 4-13 The FM broadcast band, off-air spectrum





4.6 The Effect of Noise on FM

Recall AM and the effect of noise on it. Random electrical variations added to the AM signal altered the original modulation of the signal. For FM, noise still adds to the signal, but because the information resides in frequency changes instead of amplitude changes, the noise tends to have less of an effect.

Expanding upon this idea a bit, one notes that the random electrical variations encountered by the FM signal will indeed cause distortion by "jittering" the frequency of the FM signal. However, the change in frequency modulation caused by the jittering usually turns out to be less than the change in the amplitude modulation caused by the same relative amplitude noise variations on an AM signal. Also unlike AM, the effect of the frequency jittering becomes progressively worse as the modulating frequency increases. In other words, the effect of noise increases with modulation frequency.

Pre-Emphasis and De-Emphasis

To compensate for this last effect, FM communication systems have incorporated a noise-combating system of pre-emphasis and de-emphasis. How is this done? **Pre-emphasis** gives added amplitude to the higher modulating frequencies prior to modulation under a well-defined pre-emphasis

(HPF) curve. This added amplitude will serve to make the higher frequencies more immune to noise by increasing their index of modulation. **De-emphasis** is just the opposite operation (using an LPF) and it is done at the receiver. The net effect of the two filtering processes is to cancel one another out. However, the benefit of increased noise immunity is retained. See Figure 4-15 for the filter characteristics for FM broadcasting (the filter time constant = 75 μ sec).



Threshold and Capture Effect

Another interesting effect of noise on FM operation is what is called the **threshold effect**. As long as there is a sufficient signal-to-noise ratio (SNR) at the input to the FM receiver, the FM system has substantially better noise performance than an AM system. However, there is a point below which FM-system performance is no longer better than AM. As a matter of fact, beyond this point performance can be even worse than AM. This effect can be traced to the use of limiters in the FM receiver. The purpose of the limiter is to remove any AM noise on the signal because it contains no information. If the signal strength is high enough, the limiter performs its function. If the input signal strength is not sufficient, the limiter does not perform its function, and the noise performance of the receiver is similar to an AM receiver.

The **capture effect** refers to the case of two co-channel or adjacentchannel FM signals being received at the same time by an FM receiver. When this occurs, the FM receiver will treat the weaker signal as interference and the stronger signal is said to have captured the receiver. The

amplitude limiter in the receiver will tend to suppress the weaker signal in the same manner as it suppressed AM noise. If the two FM signals are of almost identical strength, the FM receiver will switch back and forth between the two signals. If the interfering signal is stronger than the desired signal, the FM receiver will lock onto the interfering signal. This effect can be experienced as one drives down an interstate between two metropolitan areas and receives signals from two stations that are nearby in the frequency spectrum but geographically separate.

4.7 FM Generation

The most common method used to generate FM is called **direct FM**. This method uses an active device that can be set up to implement a voltage-to-frequency (V/F) function. One such device is a varactor diode. Recall that the varactor diode is equivalent to a capacitor when reverse biased. Furthermore, the equivalent capacitance of the varactor diode varies as the reverse-bias voltage applied to it is increased or decreased. Used in conjunction with a tuned circuit, the varactor diode can convert an input-signal voltage into a varying oscillator output frequency. This type of system is often called a frequency-modulated oscillator, or FMO. See Figure 4-16.



Generation of FM with ICs

There are currently off-the-shelf IC chips that can be used to produce lowpower (up to about +20 dB or 100 mW) FM signals over a wide frequency range (typically up to 1 GHz). If power in the range of 1 W is desired, the output power of these ICs can be amplified by class-A linear amplifiers.

As mentioned in chapter 3 when discussing AM IC modulators, the semiconductor industry has evolved to offer application-specific ICs (ASICs)

that are effectively systems on a chip (SOC). Manufacturers of low-power devices like cellular telephones have a host of ICs to choose from that provide complex modulation schemes with onboard amplifiers and frequency synthesizers. If higher final antenna power is desired, amplifier stages can be added. See a typical manufacturer's Web page at either http://www.motorola.com or http://www.rfmd.com for data sheets of their RF products.

Typical High-Power FM Transmitters

Figure 4-17 shows a block diagram of the most common type of FM transmitter designed for output powers in excess of 5 to 10 W. An audio signal is amplified and applied to a pre-emphasis network (HPF). It is then applied directly to a frequency-modulated oscillator (FMO), which provides a small output signal in the mW range. A buffer amplifier and successive driver amplifiers raise the output power to the required level. Radio-frequency power transistors with forced-air cooling can easily supply output powers of over 100 W and air-cooled vacuum tubes can supply kilowatts of output power at frequencies into the UHF range. Refer back to Figure 3-26 for an example. Low-power FM transmitters (approximately 20 to 50 W maximum) are sometimes called FM exciters if they are used to drive high-power amplifiers. In either case, the FM transmitter will use driver or intermediate power amplifiers (IPAs) and a final power amplifier (FPA) to achieve their final output power. These high-power amplifier stages can be run as class-C amplifiers to obtain higher efficiency.





To set the FM-transmitter rest frequency, the FM transmitter will normally incorporate a programmable-PLL frequency synthesizer that can cover the entire band of operation for the particular type of service (i.e. the FM broadcast band is from 88.1 to 107.9 MHz by 200 kHz steps). The average transmitter rest frequency will be compared to the PLL output frequency and an error voltage will be applied to the FMO to correct any frequency drift from the assigned transmitter output frequency.

4.8 Frequency Modulation Receivers

Let us now take a look at a typical FM broadcast-band receiver, as pictured in Figure 4-18. The FM receiver is very similar to the AM superhet receiver, with only slight modifications. The FM receiver will usually always have an RF amplifier (usually a low-noise MOSFET transistor) and a separate mixer and local oscillator. These three subsystems, if constructed with discrete components, will usually be shielded to prevent reradiation of highfrequency energy from the receiver and to prevent oscillations resulting from unwanted feedback. The intermediate frequency (IF) will usually be the standard value of 10.7 MHz with a bandwidth of approximately 200 kHz to allow for a WBFM signal to be amplified and passed on to the FMdetector subsystem. Notice that because of the IF frequency chosen, image frequencies are 21.4 MHz from the desired signals, which puts them outside the 20-MHz WBFM broadcast band. As a result of this, a large part of the superhet-receiver image problem due to other FM stations has been eliminated. The detector will be quite different from the type used in the AM receiver and must reverse the process that occurred at the transmitter. Therefore, a frequency-to-voltage (F/V) converter is needed to demodulate the received signal.





Review of the Superhet's Characteristics

Before examining the FM detector in more detail, we should remind ourselves of some of the characteristics of superhet operation. After selection of the desired signal by the BPFs used for tuning, the signal is amplified by the RF amplifier. All received signals are then translated in frequency to one set, intermediate frequency (10.7 MHz) by the mixer in conjunction with the local oscillator. The signal at 10.7 MHz is amplified by the IF stages and then sent to the detector stage. The detector type depends upon the kind of analog modulation being received. The superhet receiver has specifications of sensitivity, selectivity, distortion, dynamic range, and spurious responses. Superhet receivers often have enhancements such as double conversion, notch filters, squelch, and signal-strength indicators, depending upon the application. For FM receivers, sensitivity is usually given in terms of "useable sensitivity." This is the sensitivity required for a certain value of SINAD. For portable FM equipment, the sensitivity is specified for a 12-dB SINAD. This is sufficient for voice communications, but for FM broadcast receivers, the sensitivity is specified for a 30-dB SINAD. The other specification peculiar to FM receivers is that of quieting sensitivity. This is a measure of the effectiveness of the FM receiver's ability to reduce noise in the presence of a signal.

The FM Detector

A typical modern FM receiver uses a phase-locked loop (PLL) subsystem as the detector. The PLL is insensitive to amplitude variations and can perform the F/V function; it can therefore be used as an FM detector. The FM stereo receiver must also include further demultiplexing circuitry, which we will discuss later. A PLL detector is shown in Figure 4-19.



The PLL detector works in the following manner: The voltagecontrolled oscillator (VCO) free runs at the 10.7-MHz intermediate frequency. The incoming signal, if unmodulated, locks up with the VCO signal, causing there to be no signal (ac) output from the LPF stage.

Now let us assume that the incoming signal has been modulated by a single audio tone. The phase detector will output an error voltage to the VCO in an attempt to drive the VCO into lock-up with the incoming signal. Because the incoming signal frequency is deviating both above and below the 10.7-MHz rest frequency at a certain number of cycles per second, the VCO will do the same, following the input-signal frequency variations. The error voltage from the LPF, which drives the VCO, will be identical to the original modulating signal, and hence is taken as the demodulated output.

FIGURE 4-19 A typical FM detector using a phase-locked loop

4.9 FM Stereo

FM stereo broadcasting was introduced during the early 1960s. The scheme that was adopted was chosen to be compatible with the monaural FM radios that were in existence at the time. Essentially, the system performs the multiplexing of two signals and further combines them into a complex baseband signal that modulates the FM carrier. Figure 4-20 shows a block diagram of the typical legacy analog-stereo generator used to drive an FM transmitter. A left and right source of information are first pre-emphasized and then fed to adder circuits. The output of one adder is the sum of the two signals, or the L + R signal (the monaural signal), and the output of the other adder is the difference of the two signals, or L – R. The L – R signal is applied to a balanced modulator along with a 38-kHz signal. The output of the balanced modulator is a DSBSC AM signal centered at 38 kHz. A portion of the 38-kHz signal is divided in frequency to become 19 kHz, and all three signals are applied to a summer/adder circuit at the output of the generator.

The resulting stereo-generator output-signal spectrum is shown in Figure 4-21. The L + R signal, which contains baseband frequencies from near 0 Hz to 15 kHz, occupies that portion of the frequency spectrum. There are guard bands around the 19-kHz tone, which is called a pilot subcarrier signal, that will be used at the receiver to aid in the demodulation of the received signal. The L – R signal, which has been DSBSC amplitude modulated by a 38-kHz tone, occupies the frequency range from 23 to 53 kHz. A state-of-the-art stereo generator will use digital techniques to synthesize the 19- and 38-kHz tones and the composite final baseband signal. The stereo-demodulator block diagram is shown in Figure 4-22.









The composite signal, as shown in Figure 4-21, is recovered at the FM detector after transmission and reception and then fed to the stereo decoder shown in Figure 4-22. Usually, this is a single IC chip or part of the detector IC chip.

The composite signal is separated by two BPFs and one LPF into three separate signals. The L + R signal, which occupies the 0- to 15-kHz range does not undergo any additional signal processing, with the exception of the addition of a small time delay. The 19-kHz pilot tone is recovered by a narrow BPF centered at 19 kHz. This signal undergoes a frequency doubling to 38 kHz, and is then applied to a balanced modulator. The L – R DSBSC AM signal is recovered by another BPF centered at 38 kHz, and it is also applied to the previously mentioned balanced modulator. The balanced-modulator output consists of the original left-minus-right (L – R) signal. At this point the L + R and L – R signals are applied to adder circuits (sometimes referred to as a matrix) that yield the separate left and right signals. The signals are then fed to identical audio amplifiers.



Let us pause here to reflect upon what has just been explained. The FM stereo system takes an audio signal (L - R) and performs amplitude modulation on it. A composite signal consisting of the unmodulated L + R signal, a 19-kHz tone, and the DSBSC AM L – R signal are now applied to an FM transmitter. The L – R signal has been modulated twice! First, it was DSBSC amplitude modulated, and then it was frequency modulated. Is this a problem? The answer is *no*. To recover the original signal, one must simply reverse the prior steps in the correct order.

At the receiver the FM detector reverses the FM process, yielding exactly the same signal that was produced by the stereo generator at the transmitter. The DSBSC AM L – R signal and a 38-kHz tone are then applied to a balanced modulator and the two signals are effectively mixed together, yielding the difference frequencies, which are the original left-right signal. As explained before, adder circuits then convert the L + R and L – R signals into just the left and right signals. The demodulation process has first reversed the FM process and then the DSBSC AM process to recover the original information. The question might be asked, Is there a limit to the number of times a signal can be modulated and to the type of modulation used? The simple answer is *no*. However, each time a signal undergoes the modulation-demodulation process, noise and nonlinear effects are introduced into the signal. So practically speaking, there is a finite limit to the number of times a signal can be modulated before distortion becomes discernible.

Several manufacturers produce ICs that are complete systems on a chip (SOC). Shown in Figure 4-23 is a simplified block diagram of an IC chip that can be used anywhere on the planet to receive any form of AM or FM broadcasting service presently being used. It also has the ability to receive specialized weather broadcasts and medium-wave transmissions. It contains a dual-conversion AM receiver, an FM receiver, a PLL frequency synthesizer, and many other features.

Summary

The process of producing FM involves the instantaneous variation of the output frequency of a transmitter in accordance with the modulating signal. This process produces an infinite number of sidebands, which are related to the modulating frequency. The number of significant sidebands, and the power distribution over the sidebands, is dependent upon the modulating frequency and the amount of frequency deviation. For specific values of the index of modulation, one can use Bessel-function tables to determine the sideband amplitude distribution and the signal bandwidth. In general, unless the index of modulation is restricted, FM signals

FREQUENCY MODULATION • 169



FIGURE 4-23 Block diagram of a universal IC receiver (From the ISSCC Digest, Volume 41 © 1998 IEEE)

can be many times wider than the baseband signals that performed the modulation in the first place.

The usual method of producing an FM signal is to use a frequencymodulated oscillator (FMO) that converts an input voltage into a frequency deviation. At the heart of these systems (buried on an IC) is the varactor diode. This device is used to perform the voltage-to-frequency conversion needed for the production of FM. If a high-power FM signal is needed, one

simply amplifies the low-power FM signal (repeatedly if necessary) to obtain the desired power. FM transmitters require some form of frequency control for stability. This is usually achieved through the use of a PLL synthesizer and a feedback circuit to correct for any frequency drift. Today, ASICs are used in the design of transmitter/receiver hardware utilizing FM as the modulation technique. If large amounts of transmitting power are needed, power amplifiers are used.

To receive FM, a superhet receiver is used. This receiver will have an RF or low-noise amplifier (LNA) stage, separate mixer and local-oscillator stages, an IF at 10.7 MHz, and an FM detector. The detector is usually a PLL that does a frequency-to-voltage conversion of the received signal. The local oscillator will usually employ a PLL frequency synthesizer to develop the local-oscillator signal and supply front-panel display information that indicates the receiver's current tuning status. Additional features, such as signal-strength indicators, are added on an as needed basis.

FM stereo is a system used with legacy FM broadcasting and TV sound transmission that allows two separate audio signals to be sent between the transmitter and receiver. This multiplexing system allows for the transmission of stereophonic information and enhances the information by adding a spatial quality to the audio signal.

Questions and Problems

Section 4.1

1. Do an Internet search on Major Edwin Armstrong. Write a short paragraph about his legal battles with RCA over the new modulation technology, FM.

Sections 4.2 and 4.3

- **2.** Describe the difference between amplitude modulation and frequency modulation.
- **3.** What does the term *modulation sensitivity* refer to?
- **4.** How does FM embed the modulation signal's information into the FM carrier wave?
- **5.** What is the meaning of the FM index of modulation, m_{fi} and how does it vary from the AM index of modulation?
- **6.** Define the FM *percentage of modulation*.

- **7.** What are Bessel functions and what is their relationship to FM?
- **8.** What is the relationship of Bessel functions to the frequency spectrum of an FM wave?
- **9.** Draw a sketch of the frequency spectra of an FM wave with $m_f = 2.5$.
- **10.** Explain the difference between NBFM and WBFM.

Sections 4.4 and 4.5

- **11.** Draw the spectrum of an FM wave with a modulating signal of 5 kHz and an index of modulation of 2. Determine the bandwidth of this signal.
- **12.** What is different about the rest frequency of an FM wave and the carrier frequency of an AM wave?

FREQUENCY MODULATION • 171

- **13.** What eventually happens to the sideband amplitude of an FM signal as the sidebands become further removed from the rest frequency?
- **14.** At what index of modulation does the second pair of sidebands in an FM wave disappear?
- **15.** Prove that all the individual sideband powers add up to the total power for an FM wave with $m_f = 2$.

Sections 4.6 to 4.8

16. Describe why FM is more resistant to noise than AM.

- **17.** Describe the process of pre-emphasis and de-emphasis for FM modulation. What does it accomplish?
- **18.** Describe several differences between FM and AM superhet receivers. Why are these differences needed?
- **19.** Why is the IF frequency of an FM receiver so high?
- **20.** Explain the purpose of the stereo pilot carrier at 19 kHz.

Chapter Equations

 $e(t) = E_p \sin(\omega t + \phi) \tag{4.1}$

$$\delta = k_f \times E_M \tag{4.2}$$

$$f_{\text{signal}} = f_C + k_f e_M(t) = f_C + k_f E_M \sin \omega_M t$$
4.3

$$f_{\text{signal}} = f_C + \delta \sin \omega_M t \tag{4.4}$$

$$e_{\rm FM}(t) = E_C \sin(\omega_C t + m_f \sin \omega_M t)$$
4.5

$$m_f = \frac{\delta}{f_M}$$
 4.6

% Modulation =
$$\frac{\delta}{\delta_{\text{max}}} \times 100\%$$
 4.7

_

$$e_{FM}(t) = E_C \begin{cases} J_0(m_f)\sin\omega_C t - J_1(m_f) \left[\sin(\omega_C - \omega_M)t - \sin(\omega_C + \omega_M)t \right] + \\ J_2(m_f) \left[\sin(\omega_C - 2\omega_M)t + \sin(\omega_C + 2\omega_M)t \right] - \\ J_3(m_f) \left[\sin(\omega_C - 3\omega_M)t - \sin(\omega_C + 3\omega_M)t \right] + \\ J_4(m_f) \left[\sin(\omega_C - 4\omega_M)t + \sin(\omega_C + 4\omega_M)t \right] - \dots \end{cases}$$

$$4.8$$

Bandwidth =
$$f_M \times \#$$
 of sideband pairs $\times 2$

$$E_{\text{signal}} = \left(\frac{P_{\text{trans}}}{R}\right)^{\frac{1}{2}}$$
4.10

4.9

$$P_n = J_n^2(m_f) \times P_{\text{trans}}$$

$$4.11$$

$$P_{\text{total}} = (J_0^2 + 2J_1^2 + 2J_2^2 + 2J_3^2 + ...) \times P_{\text{trans}}$$
4.12