Chapter 4
Phase and Frequency Modulation

Wireless Information Transmission System Lab.
Institute of Communications Engineering
National Sun Yat-sen University
Outline

- 4.1 Introduction
- 4.2 Basic Definitions
- 4.3 Frequency Modulation
- 4.4 Phase-locked Loop
Chapter 4.1
Introduction

Wireless Information Transmission System Lab.
Institute of Communications Engineering
National Sun Yat-sen University
4.1 Introduction

- In this chapter, we study a second family of continuous-wave (CW) modulation systems, namely, *angle modulation*, in which the angle of the carrier wave is varied according to the baseband signals.

- In this method of modulation, the amplitude of the carrier wave is maintained constant.

- There are two common forms of angle modulation, namely, *phase modulation* and *frequency modulation*.

- An important feature of angle modulation is that it can provide better discrimination against noise and interference than amplitude modulation.
4.1 Introduction

- However, this improvement in performance is achieved at the expense of increased transmission bandwidth.

- Moreover, the improvement in the noise performance with angle modulation is achieved at the expense of increased system complexity in both the transmitter and receiver.

- Such a trade-off is not possible with amplitude modulation.
Chapter 4.2
Basic Definitions
4.2 Basic Definitions

◊ Let $\theta_i(t)$ denote the angle of a modulated sinusoidal carrier at time $t$; it is assumed to be a function of the information–bearing signal or message signal.

◊ We express the resulting angle-modulated wave as

$$s(t) = A_c \cos \left[ \theta_i(t) \right]$$

where $A_c$ is the carrier amplitude.

◊ The average frequency in Hertz over an interval from $t$ to $t+\Delta t$ is given by

$$f_{\Delta t}(t) = \frac{\theta_i(t + \Delta t) - \theta_i(t)}{2\pi \Delta t}$$

◊ The instantaneous frequency of the angle-modulated signal $s(t)$:

$$f_i(t) = \lim_{\Delta t \to 0} f_{\Delta t}(t) = \lim_{\Delta t \to 0} \left[ \frac{\theta_i(t + \Delta t) - \theta_i(t)}{2\pi \Delta t} \right] = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt}$$
4.2 Basic Definitions

◊ For an unmodulated carrier, the angle $\theta_i(t)$ is given by

$$\theta_i(t) = 2\pi f_c t + \phi_c$$

and corresponding phasor rotates with a constant angular velocity equal to $2\pi f_c$. The constant $\phi_c$ is the value of $\theta_i(t)$ at $t=0$.

◊ There are an infinite number of ways in which the angle $\theta_i(t)$ may be varied in some manner with the message (baseband) signal.

◊ We shall consider only two commonly used methods, phase modulation and frequency modulation.
4.2 Basic Definitions

- **Phase modulation (PM)** is that form of angle modulation in which the instantaneous angle $\theta_i(t)$ is varied linearly with the message signal as shown by

$$\theta_i(t) = 2\pi f_c t + k_p m(t)$$  \hspace{1cm} (4.4)

The term $2\pi f_c t$ represents the angle of the unmodulated carrier; $k_p$ represents the *phase sensitivity* of the modulator, expressed in *radians per volt* on the assumption that $m(t)$ is a voltage waveform.

For convenience, we have assumed in Eq. (4.4) that the angle of the unmodulated carrier is zero at $t=0$. The phase-modulated signal $s(t)$ is thus described in the time domain by

$$s(t) = A_c \cos\left[2\pi f_c t + k_p m(t)\right]$$  \hspace{1cm} (4.5)
4.2 Basic Definitions

- **Frequency modulation (FM)** is that form of angle modulation in which the instantaneous frequency \( f_i(t) \) is varied linearly with the message signal \( m(t) \), as shown by

\[
f_i(t) = f_c + k_f m(t)
\]  

(4.6)

\( f_c \): The frequency of the unmodulated carrier  
\( k_f \): The *frequency sensitivity* of the modulator (Hertz per volt)

Integrating Eq. (4.6) with respect to time and multiplying the result by \( 2\pi \), we get

\[
\theta_i(t) = 2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau
\]  

(4.7)

where, for convenience, we have assumed that the angle of the unmodulated carrier wave is zero at \( t=0 \). The frequency-modulated signal is therefore described in the time domain by

\[
s(t) = A_c \cos \left[ 2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau \right]
\]  

(4.8)
4.2 Basic Definitions

a) Carrier wave

b) Sinusoidal modulating signal

c) Amplitude-modulated signal

d) Phase-modulated signal

e) Frequency-modulated signal
Properties of Angle-Modulated Waves

- Property 1: Constancy of Transmitted Power:
  - From both Eqs. (4.4) and (4.7), we readily see that the amplitude of PM and FM waves is maintained at a constant value equal to the carrier amplitude $A_c$ for all time $t$, irrespective of the sensitivity factors $k_p$ and $k_f$.
  - Consequently, the average transmitted power of angle-modulated waves is a constant, as shown by

$$P_{av} = \frac{1}{2} A_c^2$$

where it is assumed that the load resistance is 1 ohm.
Properties of Angle-Modulated Waves

◊ Property 2: Nonlinearity of the Modulation Process

◊ Both PM and FM waves violate the principle of superposition.

◊ For example, the message signal $m(t)$ is made up of two different components, $m_1(t)$ and $m_2(t)$: $m(t) = m_1(t) + m_2(t)$

◊ Let $s(t)$, $s_1(t)$, and $s_2(t)$ denote the PM waves produced by $m(t)$, $m_1(t)$, and $m_2(t)$ in accordance with Eq. (4.4), respectively. We may express these PM waves as follows:

$$s(t) = A_c \cos\left[2\pi f_c t + k_p (m_1(t) + m_2(t))\right]$$

$$s_1(t) = A_c \cos\left[2\pi f_c t + k_p m_1(t)\right]$$

$$s_2(t) = A_c \cos\left[2\pi f_c t + k_p m_2(t)\right]$$

◊ Frequency modulation offers superior noise performance compare to amplitude modulation,
Properties of Angle-Modulated Waves

◊ Property 3: Irregularity of Zero-Crossings

◊ **Zero-crossing** are defined as the instants of time at which a waveform changes its amplitude from positive to negative value or the other way around.

◊ The zero-crossings of a PM or FM wave no longer have a perfect regularity in their spacing across the time-scale.

◊ The irregularity of zero-crossings in angle-modulated waves is attributed to the nonlinear character of the modulation process.
Property 4: Visualization Difficulty of Message Waveform

In AM, we see the message waveform as the envelope of the modulated wave, provided the percentage modulation is less than 100 percent.

(AM: The percentage modulation over 100 percent → phase reversal → distortion)

This is not so in angle modulation, as illustrated by the corresponding waveform of Figures 4.1d and 4.1e for PM and FM, respectively.
Properties of Angle-Modulated Waves

◊ Property 5-Trade-OFF of Increased Transmission Bandwidth for Improved Noise Performance
  ◊ An important advantage of angle modulation over amplitude modulation is the realization of improved noise performance.
  ◊ This advantage is attributed to the fact that the transmission of a message signal by modulating the angle of a sinusoidal carrier wave is less sensitive to the presence of additive noise than transmission by modulating the amplitude of the carrier.
  ◊ The improvement in noise performance is achieved at the expense of a corresponding increase in the transmission bandwidth requirement of angle modulation.
Properties of Angle-Modulated Waves

- Property 5-Trade-OFF of Increased Transmission Bandwidth for Improved Noise Performance
  - The use of angle modulation offers the possibility of exchanging an increase in the transmission bandwidth for an improvement in noise performance.

- Such a trade-off is not possible with amplitude modulation since the transmission bandwidth of an amplitude-modulated wave is fixed somewhere between the message bandwidth $W$ and $2W$, depending on the type of modulation employed.
Example 4.1 Zero-Crossings

Consider a modulating wave \( m(t) \) that increases linearly with time \( t \), starting at \( t=0 \), as shown by

\[
m(t) = \begin{cases} 
  at, & t \geq 0 \\
  0, & t < 0
\end{cases}
\]

where \( a \) is the slope parameter (see Figure 4.2a). In what follows, we study the zero-crossings of the PM and FM waves produced by \( m(t) \) for the following set of parameters:

\[
f_c = \frac{1}{4} \text{Hz} \\
a = 1 \text{ volt/s}
\]
Example 4.1 Zero-Crossings

Fig. 4.2 Starting at time $t = 0$, the figure displays (a) linearly increasing message signal $m(t)$, (b) phase-modulated wave, and (c) frequency-modulated wave.
Example 4.1 Zero-Crossings

◊ Phase Modulation:

◊ Phase-sensitivity factor $k_p = \pi/2$ radians/volt. Applying Eq. (4.5) to the given $m(t)$ yields the PM wave

$$s(t) = A_c \cos\left[2\pi f_c t + k_p m(t)\right] \quad (4.5)$$

$$s(t) = \begin{cases} A_c \cos\left(2\pi f_c t + k_p a t\right), & t \geq 0 \\ A_c \cos\left(2\pi f_c t\right), & t < 0 \end{cases}$$

which is plotted in Figure 4.2b for $A_c = 1$ volt.

◊ Let $t_n$ denote the instant of time at which the PM wave experiences a zero crossing; this occurs whenever the angle of the PM wave is an odd multiple of $\pi/2$:

$$2\pi f_c t_n + k_p a t_n = \pi \left(2f_c + \frac{k_p a}{\pi}\right) t_n = \frac{\pi}{2} + n\pi, \quad n = 0, 1, 2, \ldots$$

$$t_n = \frac{1}{2} + \frac{n}{2f_c + \frac{k_p a}{\pi}}$$

$$t_n = \frac{1}{2} + n, \quad n = 0, 1, 2, \ldots$$
Example 4.1 Zero-Crossings

- Frequency Modulation:
  - Frequency-sensitivity factor, \( k_f = 1 \) Hz/volt. Applying Eq. (4.8) yields the FM wave

\[
s(t) = A_c \cos \left[ 2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau \right]
\]

which is plotted in Figure 4.2c.

- Invoking the definition of a zero-crossing, we can obtain:

\[
2\pi f_c t_n + \pi k_f a t_n^2 = \frac{\pi}{2} + n\pi, \quad n = 0, 1, 2, \ldots
\]

\[
t_n = \frac{1}{ak_f} \left( -f_c + \sqrt{f_c^2 + ak_f \left( \frac{1}{2} + n \right)} \right), \quad n = 0, 1, 2, \ldots
\]

\[
t_n = \frac{1}{4} \left( -1 + \sqrt{9 + 16n} \right), \quad n = 0, 1, 2, \ldots
\]
Example 4.1 Zero-Crossings

- Comparing the zero-crossing results derived for PM and FM waves, we may make the following observations once the linear modulating wave begins to act on the sinusoidal carrier wave:

1. For PM, regularity of the zero-crossings is maintained; the instantaneous frequency changes from the unmodulated value of $f_c = 1/4 \text{ Hz}$ to the new constant value of $f_c + k_p \left( a / 2\pi \right) = 0.5 \text{ Hz}$

2. For FM, the zero-crossings assume an irregular form; as expected, the instantaneous frequency increases linearly with time $t$. 
4.2 Basic Definitions

◊ Comparing Eq. (4.5) with (4.8) reveals that an FM signal may be regarded as a PM signal in which the modulating wave is $\int_0^t m(\tau) d\tau$ in place of $m(t)$.

\[
s(t) = A_c \cos \left[ 2\pi f_c t + k_p m(t) \right] \quad (4.5)
\]

\[
s(t) = A_c \cos \left[ 2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau \right] \quad (4.8)
\]

◊ The FM signal can be generated by first integrating $m(t)$ and then using the result as the input to a phase modulator, as in Figure 4.3a.

◊ Conversely, a PM signal can be generated by first differentiating $m(t)$ and then using the result as the input to a frequency modulator, as in Figure 4.3b.

◊ We may thus deduce all the properties of PM signals from those of FM signals and vice versa. Henceforth, we concentrate our attention on FM signals.
Figure 4.3 Illustrating the relationship between frequency modulation and phase modulation. (a) Scheme for generating an FM wave by using a phase modulator, (b) scheme for generating a PM wave by using a frequency modulator.

<table>
<thead>
<tr>
<th></th>
<th>( \theta_i(t) )</th>
<th>( f_i(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodulated signal</td>
<td>( 2\pi f_c t )</td>
<td>( f_c )</td>
</tr>
<tr>
<td>PM signal</td>
<td>( 2\pi f_c t + k_p m(t) )</td>
<td>( f_c + \frac{k_p}{2\pi} \frac{dm(t)}{dt} )</td>
</tr>
<tr>
<td>FM signal</td>
<td>( 2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau )</td>
<td>( f_c + k_f m(t) )</td>
</tr>
</tbody>
</table>
Chapter 4.3
Frequency Modulation
4.3 Frequency Modulation

◊ The FM signal \( s(t) \) define by Eq. (4.8) is a nonlinear function of the modulating signal \( m(t) \), which makes frequency modulation a nonlinear modulation process.

\[
s(t) = A_c \cos \left[ 2\pi f_c t + 2\pi k \int_0^t m(\tau) d\tau \right]
\]  (4.8)

◊ How then can we tackle the spectral analysis of FM signal? We propose to provide an empirical answer to this important question by proceeding in the same manner as with AM modulation, that is, we consider the simplest case possible, namely, single-tone modulation.

◊ Consider then a sinusoidal modulating signal define by

\[
m(t) = A_m \cos (2\pi f_m t)
\]  (4.10)
4.3 Frequency Modulation

◊ The instantaneous frequency of the resulting FM signal is

\[ f_i(t) = f_c + k_f A_m \cos(2\pi f_m t) = f_c + \Delta f \cos(2\pi f_m t) \quad (4.11) \]

\[ \Delta f = k_f A_m \quad (4.12) \]

◊ The quantity \( \Delta f \) is called the frequency deviation, representing the maximum departure of the instantaneous frequency of the FM signal from the carrier frequency \( f_c \).

◊ A fundamental characteristic of an FM signal is that the frequency deviation \( \Delta f \) is proportional to the amplitude of the modulating signal and is independent of the modulating frequency.

◊ Using Eq. (4.11), the angle \( \theta_i(t) \) of the FM signal is obtained as

\[ \theta_i(t) = 2\pi \int_0^t f_i(t) dt = 2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t) \]

◊ The ratio of the frequency deviation \( \Delta f \) to the modulation frequency \( f_m \) is commonly called the modulation index of the FM signal.
4.3 Frequency Modulation

- The modulation index is denoted by $\beta$: \[
\beta = \frac{\Delta f}{f_m}
\]

\[
\theta_i(t) = 2\pi f_c t + \beta \sin(2\pi f_c t)
\]

- The parameter $\beta$ represents the phase deviation of the FM signal, i.e. the maximum departure of the angle $\theta_i(t)$ from the angle $2\pi f_c t$ of the unmodulated carrier. $\beta$ is measured in radians.

- The FM signal itself is given by

\[
s(t) = A_c \cos\left[2\pi f_c t + \beta \sin(2\pi f_m t)\right]
\] (4.16)

Depending on the value of the modulation index $\beta$, we may distinguish two cases of frequency modulation:

- **Narrow-band FM**, for which $\beta$ is small compared to one radian.
- **Wide-band FM**, for which $\beta$ is large compared to one radian.
4.3 Frequency Modulation

◊ **Narrow-band frequency modulation**

◊ Consider Eq. (4.16), which defines an FM signals resulting form the use of sinusoidal modulating signal. Expanding this relation, we get

\[ s(t) = A_c \cos(2\pi f_c t)\cos[\beta \sin(2\pi f_m t)] - A_c \sin(2\pi f_c t)\sin[\beta \sin(2\pi f_m t)] \]  \hspace{1cm} (4.17)

◊ Assuming that the modulation index \( \beta \) is small compared to one radian, we may use the following two approximations:

\[ \cos[\beta \sin(2\pi f_m t)] \approx 1 \hspace{1cm} \sin[\beta \sin(2\pi f_m t)] \approx \beta \sin(2\pi f_m t) \]

\[ s(t) \approx A_c \cos(2\pi f_c t) - \beta A_c \sin(2\pi f_c t)\sin(2\pi f_m t) \]  \hspace{1cm} (4.18)

\[ s(t) \approx A_c \cos(2\pi f_c t) + \frac{1}{2} \beta A_c \left\{ \cos[2\pi (f_c + f_m) t] - \cos[2\pi (f_c - f_m) t] \right\} \]  \hspace{1cm} (4.19)

\[ \sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)] \]
4.3 Frequency Modulation

- This expression is somewhat similar to the corresponding one defining an AM signal (from Example 3.1):

\[
s_{AM}(t) = A_c \cos(2\pi f_c t) + \frac{1}{2} \mu A_c \left\{ \cos\left[ 2\pi (f_c + f_m) t \right] + \cos\left[ 2\pi (f_c - f_m) t \right] \right\}
\]

where \( \mu \) is the modulation factor of the AM signal.

- Compare Eqs. (4.19) and (4.20), we see that the basic difference between an AM signal and a narrow-band FM signal is that the algebraic sign of the lower side frequency in the narrow-band FM is reversed.

- Thus, a narrow-band FM signal requires essentially the same transmission bandwidth (i.e. \( 2f_m \)) as the AM signal.
4.3 Frequency Modulation

◊ **Wide-band frequency modulation**

◊ The following studies the spectrum of the single-tone FM signal of Eq. (4.16) for an arbitrary value of the modulation index $\beta$.

$$s(t) = A_c \cos\left[2\pi f_c t + \beta \sin(2\pi f_m t)\right] \quad (4.16)$$

◊ By using the complex representation of band-pass signals described in Chapter 2: (Carrier frequency $f_c$ compared to the bandwidth of the FM signal is large enough)

$$s(t) = \text{Re}\left[ A_c \exp\left(j2\pi f_c t + j\beta \sin(2\pi f_m t)\right)\right] \quad (4.21)$$

$$= \text{Re}\left[ \tilde{s}(t) \exp\left(j2\pi f_c t\right)\right]$$

where $\tilde{s}(t) = A_c \exp\left[j\beta \sin(2\pi f_m t)\right] \rightarrow \text{periodic function}$
4.3 Frequency Modulation

◊ Wide-band frequency modulation

◊ We may therefore expend $\tilde{s}(t)$ in the form of complex Fourier series as follows:

$$
\tilde{s}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi nf_m t) \quad (4.23)
$$

$$
c_n = f_m \int_{-\frac{1}{2}f_m}^{\frac{1}{2}f_m} \tilde{s}(t) \exp(-j2\pi nf_m t) dt
$$

$$
= f_m A_c \int_{-\frac{1}{2}f_m}^{\frac{1}{2}f_m} \exp[j\beta \sin(2\pi f_m t) - j2\pi nf_m t] dt \quad (4.24)
$$

$$
c_n = \frac{A_c}{2\pi} \int_{-\pi}^{\pi} \exp[j(\beta \sin x - nx)] dx \quad (4.26)
$$

$$
c_n = A_c J_n(\beta) \quad \therefore J_n(\beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp[j(\beta \sin x - nx)] dx \quad (4.28)
$$

nth order Bessel function of the first kind.

$$
s(t) = A_c \cdot \text{Re} \left[ \sum_{n=-\infty}^{\infty} J_n(\beta) \exp[j2\pi(f_c + nf_m) t] \right] \quad (4.31)
$$
4.3 Frequency Modulation

- Taking the Fourier transforms of both sides of Eq. (4.31)
  \[
  S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta) \left[ \delta(f - f_c - nf_m) + \delta(f + f_c + nf_m) \right]
  \]  
  (4.32)

- In Figure 4.6 we have plotted the Bessel function \( J_n(\beta) \) versus the modulation index \( \beta \) for different positive integer values of \( n \).

**FIGURE 4.6** Plots of Bessel functions of the first kind.
4.3 Frequency Modulation

◊ We can develop further insight into the behavior of the Bessel function $J_n(\beta)$ by making use of the following properties:

1. For $n$ even, we have $J_n(\beta)=J_{-n}(\beta)$; on the other hand, for $n$ odd, we have $J_n(\beta)=-J_{-n}(\beta)$. That is
   \[ J_n(\beta) = (-1)^n J_{-n}(\beta) \quad \text{for all } n \]  
   (4.33)

2. For small values of the modulation index $\beta$, we have
   \[
   \begin{align*}
   J_0(\beta) &\approx 1 \\
   J_1(\beta) &\approx \frac{\beta}{2} \\
   J_n(\beta) &\approx 0, \quad n > 2
   \end{align*}
   \]  
   (4.34)

3. 
   \[ \sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1 \]  
   (4.35)
Thus, using Eqs. (4.32) through (4.35) and the curves of Figure 4.6, we may make the following observations:

1. The spectrum of an FM signal contains a carrier component \((n=0)\) and an infinite set of side frequencies located symmetrically on either side of the carrier at frequency separations of \(f_m, 2f_m, 3f_m, \ldots\) (An AM system gives rise to only one pair of side frequencies.)

2. For the special case of \(\beta\) small compared with unity, only the Bessel coefficients \(J_0(\beta)\) and \(J_1(\beta)\) have significant values (see 4.34), so that the FM signal is effectively composed of a carrier and a single pair of side frequencies at \(f_c \pm f_m\).

(This situation corresponds to the special case of narrowband FM that was considered previously)
3. The amplitude of the carrier component of an FM signal is dependent on the modulation index $\beta$. The physical explanation for this property is that the envelope of an FM signal is constant, so that the average power of such a signal developed across a 1–ohm resistor is also constant, as shown by

$$ P = \frac{1}{2} A_c^2 \quad \text{(Using (4.31) and (4.35))} $$
EXAMPLE 4.3 Spectra of FM Signals

◊ In this example, we wish to investigate the ways in which variations in the amplitude and frequency of a sinusoidal modulating signal affect the spectrum of the FM signal.

◊ Consider first the case when the frequency of the modulating signal is fixed, but its amplitude is varied, producing a corresponding variation in the frequency deviation $\Delta f$.

◊ Consider next the case when the amplitude of the modulating signal is fixed; that is, the frequency deviation $\Delta f$ is maintained constant, and the modulation frequency $f_m$ is varied.
EXAMPLE 4.3 Spectra of FM Signals

FIGURE 4.7 Discrete amplitude spectra of an FM signal, normalized with respect to the carrier amplitude, for the case of sinusoidal modulation of fixed frequency and varying amplitude. Only the spectra for positive frequencies are shown.
We have an increasing number of spectral lines crowding into the fixed frequency interval \( f_c - \Delta f < |f| < f_c + \Delta f \).

When \( \beta \) approaches infinity, the bandwidth of the FM wave approaches the limiting value of \( 2\Delta f \), which is an important point to keep in mind.

**FIGURE 4.8** Discrete amplitude spectra of an FM signal, normalized with respect to the carrier amplitude, for the case of sinusoidal modulation of varying frequency and fixed amplitude. Only the spectra for positive frequencies are shown.
Transmission Bandwidth of FM Signals

- In theory, an FM signal contains an infinite number of side frequencies so that the bandwidth required to transmit such a signal is similarly infinite in extent.
- In practice, however, we find that the FM signal is effectively limited to a finite number of significant side frequencies compatible with a specified amount of distortion.
- Consider the case of an FM signal generated by a single-tone modulating wave of frequency $f_m$.
  - In such an FM signal, the side frequencies that are separated from the carrier frequency $f_c$ by an amount greater than the frequency deviation $\Delta f$ decrease rapidly toward zero, so that the bandwidth always exceeds the total frequency excursion, but nevertheless is limited.
Transmission Bandwidth of FM Signals

We may thus define an approximate rule for the transmission bandwidth of an FM signal generated by a single-tone modulating signal of frequency $f_m$ as follows:

$$B_T = 2\Delta f + 2f_m = 2\Delta f \left(1 + \frac{1}{\beta}\right)$$

This empirical relation is known as Carson’s rule.

For a more accurate assessment of the bandwidth requirement of an FM signal, we may thus define the transmission bandwidth of an FM wave as the separation between the two frequencies beyond which none of the side frequencies is greater than 1% of the carrier amplitude obtained when the modulation is removed.
Chapter 4.4
Phase-locked Loop
4.4 Phase-Locked Loop

- The phase-locked loop (PLL) is a negative feedback system, the operation of which is closely linked to frequency modulation.

- It can be used for synchronization, frequency division/multiplication, frequency modulation, and indirect frequency demodulation.

- Basically, the phase-locked loop consists of three major components: a multiplier, a loop filter (low-pass filter), and a voltage-controlled oscillator (VCO) connected together in the form of a feedback loop, as in Figure 4.16.

- The VCO is a sinusoidal generator whose frequency is determined by a voltage applied to it from an external source.
4.4 Phase-Locked Loop

We assume that initially we have adjusted the VCO so that when the control voltage is zero, two conditions are satisfied:

1. The frequency of the VCO is precisely set at the unmodulated carrier frequency $f_c$.
2. The VCO output has a 90-degree phase-shift with respect to the unmodulated carrier wave.

**FIGURE 4.16** Phase-locked loop.
4.4 Phase-Locked Loop

- Suppose then that the input signal applied to the phase-locked loop is an FM signal defined by

\[
s(t) = A_c \sin \left[ 2\pi f_c t + \phi_1(t) \right]
\]

where \( A_c \) is the carrier amplitude and \( \phi_1(t) = 2\pi k_f \int_0^t m(\tau) d\tau \).

- Let the VCO output in the phase-locked loop be defined by

\[
r(t) = A_v \cos \left[ 2\pi f_c t + \phi_2(t) \right]
\]

where \( A_v \) is the amplitude. With a control voltage \( v(t) \) applied to a VCO input, the angle \( \phi_2(t) \) is related to \( v(t) \) by the integral

\[
\phi_2(t) = 2\pi k_v \int_0^t v(\tau) d\tau
\]

where \( k_v \) is the frequency sensitivity of the VCO, measured in Hertz per volt.
4.4 Phase-Locked Loop

◊ The object of the phase-locked loop is to generate a VCO output $r(t)$ that has the same phase angle (except for the fixed difference of 90 degrees) as the input FM signal $s(t)$.

◊ The time-varying phase angle $\phi_1(t)$ characterizing $s(t)$ may be due to modulation by a message signal $m(t)$ as in Eq. (4.60), in which case we wish to recover $\phi_1(t)$ in order to estimate $m(t)$.

◊ In other applications of the phase-locked loop, the time-varying phase angle $\phi_1(t)$ of the incoming signal $s(t)$ may be an unwanted phase shift caused by fluctuations in the communication channel; in this latter case, we wish to track $\phi_1(t)$ so as to produce a signal with the same phase angle for the purpose of coherent detection (synchronous demodulation).
4.4 Phase-Locked Loop

- To develop an understanding of the phase-locked loop, it is desirable to have a model of the loop.

- In what follows, we first develop a nonlinear model, which is subsequently linearized to simplify the analysis.
Nonlinear Model of the PLL

According to Figure 4.16, the incoming FM signal $s(t)$ and the VCO output $r(t)$ are applied to the multiplier, producing two components:

1. A high- frequency component, represented by the double-frequency term
   
   $$k_m A_c A_v \sin \left[ 4\pi f_c t + \phi_1(t) + \phi_2(t) \right]$$

2. A low- frequency component, represented by the difference-frequency term
   
   $$k_m A_c A_v \sin \left[ \phi_1(t) - \phi_2(t) \right]$$

where $k_m$ is the multiplier gain, measured in volt$^{-1}$.

The loop filter in the phase-locked loop is a low-pass filter, and its response to the high- frequency component will be negligible.
Nonlinear Model of the PLL

Therefore, discarding the high-frequency component (i.e., the double-frequency term), the input to the loop filter is reduced to

\[ e(t) = k_m A_c A_v \sin[\phi_e(t)] \]  \hspace{1cm} (4.63)

where \( \phi_e(t) \) is the phase error defined by

\[ \phi_e(t) = \phi_1(t) - \phi_2(t) \]

\[ = \phi_1(t) - 2\pi k_v \int_0^t \nu(\tau)d\tau \]  \hspace{1cm} (4.64)

The loop filter operates on the input \( e(t) \) to produce an output \( \nu(t) \) defined by the convolution integral

\[ \nu(t) = \int_{-\infty}^{\infty} e(\tau)h(t-\tau)d\tau \]  \hspace{1cm} (4.65)

where \( h(t) \) is the impulse response of the loop filter.
Nonlinear Model of the PLL

Using Eqs. (4.62) to (4.64) to relate $\phi_e(t)$ and $\phi_1(t)$, we obtain the following nonlinear integro-differential equation as descriptor of the dynamic behavior of the phase-locked loop:

$$\frac{d\phi_e(t)}{dt} = \frac{d\phi_1(t)}{dt} - 2\pi K_0 \int_{-\infty}^{\infty} \sin[\phi_e(\tau)]h(t-\tau)d\tau$$

(4.66)

where $K_0$ is a loop-gain parameter defined by

$$K_0 = k_m k_v A_c A_v$$

(4.67)

Equation (4.66) suggest the model shown in Figure 4.17 for a phase-locked loop.

In this model we have also included the relationship between $\nu(t)$ and $e(t)$ as represented by Eqs. (4.63) and (4.65).
Derivation of Eq. 4.66

\[ \phi_e(t) = \phi_1(t) - \phi_2(t) \]

\[ = \phi_1(t) - 2\pi k_o \int_0^t \nu(\tau) d\tau \quad (\nu(t) = \int_{-\infty}^{\infty} e(\tau) h(t-\tau) d\tau, e(t) = k_mA_cA_o \sin[\phi_e(t)]) \]

\[ = \phi_1(t) - 2\pi k_o \int_0^t \int_{-\infty}^{\infty} k_mA_cA_o \sin[\phi_e(k)] h(\tau - k) dk d\tau \]

\[ = \phi_1(t) - 2\pi K_0 \int_0^t \int_{-\infty}^{\infty} \sin[\phi_e(k)] h(\tau - k) dk d\tau \quad (K_0 = k_o k_mA_cA_o) \]

\[ = \phi_1(t) - 2\pi K_0 \int_{-\infty}^{\infty} \sin[\phi_e(k)] \int_0^t h(\tau - k) d\tau dk \]

\[ \frac{\partial \phi_e(t)}{\partial t} = \frac{\partial \phi_1(t)}{\partial t} - \frac{\partial \phi_2(t)}{\partial t} \]

\[ = \frac{\partial \phi_1(t)}{\partial t} - \frac{\partial 2\pi K_0 \int_{-\infty}^{\infty} \sin[\phi_e(k)] \int_0^t h(\tau - k) d\tau dk}{\partial t} \]

(by using the Leibniz integral rule)

\[ \frac{\partial}{\partial \alpha} \int_{a(\alpha)}^{b(\alpha)} f(x, \alpha) dx = \frac{\partial b(\alpha)}{\partial \alpha} f(b(\alpha), \alpha) - \frac{\partial a(\alpha)}{\partial \alpha} f(a(\alpha), \alpha) + \int_{a(\alpha)}^{b(\alpha)} \frac{\partial f(x, \alpha)}{\partial \alpha} dx \]

\[ = \frac{\partial \phi_1(t)}{\partial t} - 2\pi K_0 \int_{-\infty}^{\infty} \sin[\phi_e(k)] \frac{\partial}{\partial t} \int_0^t h(\tau - k) d\tau dk \]

\[ = \frac{\partial \phi_1(t)}{\partial t} - 2\pi K_0 \int_{-\infty}^{\infty} \sin[\phi_e(k)] h(t - k) dk \]
Nonlinear Model of the PLL

FIGURE 4.17 Nonlinear model of the phase-locked loop.

- We see that the model resembles the block diagram of Figure 4.17. The multiplier at the input of the phase-locked loop is replaced by a subtracter and a sinusoidal nonlinearity, and the VCO by an integrator.
- The sinusoidal nonlinearity in the model of Figure 4.17 greatly increases the difficulty of analyzing the behavior of the phase-locked loop. It would be helpful to linearize this model to simplify the analysis.
When the phase error $\phi_e(t)$ is zero, the phase-locked loop is said to be in phase-lock. When $\phi_e(t)$ is at all times small compared with one radian, we may use the approximation
\[
\sin[\phi_e(t)] \approx \phi_e(t)
\] (4.68)
which is accurate to within 4 percent for $\phi_e(t)$ less than 0.5 radians.

We may represent the phase-locked loop by the linearized model shown in Figure 4.18a.
Linear Model of the PLL

According to this model, the phase error $\phi_e(t)$ is related to the input phase $\phi_1(t)$ by the linear integro-differential equation

$$\frac{d\phi_e(t)}{dt} + 2\pi K_0 \int_{-\infty}^{\infty} \phi(\tau) h(t-\tau) d\tau = \frac{d\phi_1(t)}{dt}$$  \hspace{1cm} (4.69)

Transforming Eq. (4.69) into the frequency domain and solving for $\Phi_e(f)$, the Fourier transform of $\phi_e(f)$, in terms of $\Phi_1(f)$, the Fourier transform of $\phi_1(t)$, we get

$$\Phi_e(f) = \frac{1}{1 + L(f)} \Phi_1(f)$$  \hspace{1cm} (4.70)

The function $L(f)$ in Eq. (4.70) is defined by

$$L(f) = K_0 \frac{H(f)}{jf}$$  \hspace{1cm} (4.71)

where $H(f)$ is the transfer function of the loop filter.
Linear Model of the PLL

◊ The quantity $L(f)$ is called the open-loop transfer function of the phase-locked loop.

◊ Suppose that for all values of $f$ inside the baseband we make the magnitude of $L(f)$ very large compared with unity. Then from Eq. 4.70 we find that $\Phi_e(f)$ approaches zero. That is, the phase of the VCO becomes asymptotically equal to the phase of the incoming signal. Under this condition, phase-lock is established, and the objective of the phase-locked loop is thereby satisfied.

◊ From Figure 4.18a we see that $V(f)$, the Fourier transform of the phase-locked loop output $v(t)$, is related to $\Phi_e(f)$ by

$$V(f) = \frac{K_0}{k_v} H(f) \Phi_e(f)$$

(4.72)
Linear Model of the PLL

Equivalently, in light of Eq. (4.71), we may write

\[ V(f) = \frac{jf}{k_v} L(f) \Phi_e(f) \]  
(4.73)

\[ V(f) = \frac{(jf/k_v) L(f)}{1 + L(f)} \Phi_1(f) \]  
(4.74)

\[ L(f) = K_0 \frac{H(f)}{jf} \]

\[ \text{For } |L(f)| >> 1: \]

\[ V(f) \approx \frac{jf}{k_v} \Phi_1(f) \]  
(4.75)

\[ \nu(t) \approx \frac{1}{2\pi k_v} \frac{d\phi_1(t)}{dt} \]  
(4.76)

\[ \text{Time-Domain:} \]

\[ \text{Thus, provided that the magnitude of the open-loop transfer function } L(f) \text{ is very large for all frequencies of interest, the phase-locked loop may be modeled as a differentiator with its output scaled by the factor } \frac{1}{2\pi k_v}, \text{ as in Figure 4.18b.} \]
Therefore, substituting Eq. (4.60) in (4.76), we find that the resulting output signal of the phase-locked loop is approximately

\[ v(t) \approx \frac{k_f}{k_v} m(t) \]  \hspace{1cm} (4.77)

Equation (4.77) states that when the loop operates in its phase-locked mode, the output \( v(t) \) of the phase-locked loop is approximately the same, except for the scale factor \( k_f/k_v \), as the original message signal \( m(t) \).
A significant feature of the phase-locked loop acting as a demodulator is that the bandwidth of the incoming FM signal can be much wider than that of the loop filter characterized by $H(f)$. The transfer function $H(f)$ can and should be restricted to the baseband.

The complexity of the phase-locked loop is determined by the transfer function $H(f)$ of the loop filter.

The simplest form of a phase-locked loop is obtained when $H(f) = 1$; that is, there is no loop filter, and the resulting phase-locked loop is referred to as a first-order phase-locked loop.
The order of the phase-locked loop is determined by the order of the denominator polynomial of the closed-loop transfer function, which defines the output transform $V(f)$ in terms of the input transform $\Phi_1(f)$, as shown in Eq. (4.74).

A major limitation of a first-order phase-locked loop is that the loop gain parameter $K_0$ controls both the loop bandwidth as well as the hold-in frequency range of the loop.

The hold-in frequency range refers to the range of frequencies for which the loop remains phase-locked to the input signal.

It is for this reason that a first-order phase-locked loop is seldom used in practice.
Supplementary Material:
Analysis of PLL Using Laplace Transform
The Phase-Locked Loop

- The PLL basically consists of a multiplier, a loop filter, and a voltage-controlled oscillator (VCO):

- Assuming that the input to the PLL is the sinusoid \( x_c(t) = A_c \cos(2\pi f_c t + \phi) \) and the output of the VCO is \( e_0(t) = -A_v \sin(2\pi f_c t + \phi) \), where \( \phi \) represents the estimate of \( \phi \), the product of two signals is:

\[
e_d(t) = x_c(t) e_0(t) = -A_c \cos(2\pi f_c t + \phi) A_v \sin(2\pi f_c t + \phi)
= \frac{1}{2} A_c A_v \sin(\phi - \hat{\phi}) - \frac{1}{2} A_c A_v \sin(4\pi f_c t + \phi + \hat{\phi})
\]
The Phase-Locked Loop

- The *loop filter* is a low-pass filter that responds only to the low-frequency component $0.5A_cA_v\sin(\phi - \hat{\phi})$ and removes the component at $2f_c$.
- The output of the loop filter provides the control voltage $e_v(t)$ for the VCO.
- The VCO is a *sinusoidal signal generator* with an instantaneous phase given by

$$2\pi f_c t + \hat{\phi}(t) = 2\pi f_c t + K_v \int_{-\infty}^{t} e_v(\tau)d\tau$$

where $K_v$ is a gain constant in rad/s/V.

$$\hat{\phi}(t) = K_v \int_{-\infty}^{t} e_v(\tau)d\tau \quad \text{or} \quad \frac{d\hat{\phi}}{dt} = K_v e_v(t)$$
The Phase-Locked Loop

- By neglecting the double-frequency term resulting from the multiplication of the input signal with the output of the VCO, the phase detector output is:

\[ e_d(\psi) = K_d \sin \psi \]

where \( \psi = \phi - \hat{\phi} \) is the phase error and \( K_d \) is a proportionality constant.

- In normal operation, when the loop is tracking the phase of the incoming carrier, the phase error \( \phi - \hat{\phi} \) is small. As a result,

\[ \sin(\phi - \hat{\phi}) \approx \phi - \hat{\phi} \]

- With the assumption that \( |\psi| \ll 1 \), the PLL becomes linear.
The equations describing loop operation is conveniently obtained by using *Laplace transform* notation.

A loop model using Laplace-transformed quantities and assuming linear operation is shown in the following figure:
The Phase-Locked Loop

- The Laplace-transformed loop equations are:
  \[ E_d(s) = K_d \left[ \Phi(s) - \Theta(s) \right] = K_d \Psi(s) \]
  \[ E_v(s) = F(s) E_d(s) \]
  \[ \Theta(s) = \frac{K_v E_v(s)}{s} \]

- The closed-loop transfer function:
  \[ H(s) \triangleq \frac{\Theta(s)}{\Phi(s)} = \frac{K_v K_d F(s)}{s + K_v K_d F(s)} \triangleq \frac{K F(s)}{1 + K F(s)} / s \]

- The phase error transfer function:
  \[ H_e(s) \triangleq \frac{\Phi(s) - \Theta(s)}{\Phi(s)} = \frac{\Psi(s)}{\Phi(s)} = 1 - \frac{\Theta(s)}{\Phi(s)} = 1 - H(s) = \frac{s}{s + K_v K_d F(s)} \]
The Phase-Locked Loop

◊ The *VCO control-voltage/input-phase transfer function*:

\[
H_v(s) = \frac{E_v(s)}{\Phi(s)} = \frac{sH(s)}{K_v} = \frac{K_d s F(s)}{s + K_v K_d F(s)}
\]

◊ It is convenient to write the closed-loop transfer function in terms of the *open-loop transfer function*, which is defined as:

\[
G_{op}(s) \equiv \frac{K_v K_d F(s)}{s} \quad \Rightarrow \quad H(s) = \frac{G_{op}(s)}{1 + G_{op}(s)}
\]

◊ \(K = K_v K_d\) is the open-loop dc gain.

◊ By appropriate choice of \(F(s)\), any order closed-loop transfer function can be obtained.

◊ For second-order passive loops, the transfer function is:

\[
F(s) = \frac{1 + \tau_2 s}{1 - \tau_1 s} \quad \Rightarrow \quad H(s) = \frac{1 + \tau_2 s}{1 + (\tau_2 + 1/K)s + (\tau_1/K)s^2}
\]
The Phase-Locked Loop

- Second-order phase-locked-loop filters
The Phase-Locked Loop

- Transfer functions and parameters for first- and second-order phase-locked loops

### TABLE A-1. Transfer Functions and Parameters for First- and Second-Order Phase-Locked Loops

<table>
<thead>
<tr>
<th>Loop Filter, $F(s)$</th>
<th>Natural Frequency, $\omega_n$ (rad/s)</th>
<th>Damping Factor $\zeta$</th>
<th>Closed-Loop Transfer Function, $H(s)$</th>
<th>Error Transfer Function, $1 - H(s)$</th>
<th>Single-Sided Noise/Equivalent Bandwidth$^{bc}$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1$ (first order)</td>
<td>$K$</td>
<td>$-$</td>
<td>$\frac{K}{s + K}$</td>
<td>$\frac{s}{s + K}$</td>
<td>$\frac{K}{4}$</td>
</tr>
<tr>
<td>$\frac{s \tau_2 + 1}{s \tau_1 + 1}$ (passive, second order)</td>
<td>$\sqrt{\frac{K}{\tau_1}}$</td>
<td>$\frac{\omega_n}{2} (\tau_2 + K^{-1})$</td>
<td>$\frac{(2\zeta \omega_n - \omega_n^2)Ks + \omega_n^2}{D(s)}$</td>
<td>$\frac{s^2 + \omega_n^2K}{D(s)}$</td>
<td>$\frac{K \tau_2 (1/\tau_2 + K/\tau_1)}{4(K + 1/\tau_2)}$</td>
</tr>
<tr>
<td>$\frac{s \tau_2 + 1}{s \tau_1}$ (active, second order)</td>
<td>$\sqrt{\frac{K}{\tau_1}}$</td>
<td>$\tau_2 \omega_n$</td>
<td>$\frac{2\zeta \omega_n s + \omega_n^2}{D(s)}$</td>
<td>$\frac{s^2}{D(s)}$</td>
<td>$\frac{1}{2} \omega_n \left( \xi + \frac{1}{4\zeta} \right)$</td>
</tr>
<tr>
<td>$\frac{1}{s \tau + 1}$ (lag, second order)</td>
<td>$\sqrt{\frac{K}{\tau}}$</td>
<td>$\frac{1}{2\sqrt{K\tau}}$</td>
<td>$\frac{\omega_n^2}{D(s)}$</td>
<td>$\frac{s^2 + 2\zeta \omega_n}{D(s)}$</td>
<td>$\frac{K}{4}$</td>
</tr>
</tbody>
</table>

$^aK = K_v K_d.$

$^b$The noise equivalent bandwidth of a filter with transfer function $H(f)$ and maximum gain $H_0$ is given by $B_N = (1/H_0) \int_0^\infty |H(f)|^2 df$.

$^c$For a second-order loop with $\zeta = 0.5$, $B_L = 0.5 \omega_n$; with $\zeta = 1/\sqrt{2}$, $B_L = 0.53 \omega_n$. $B_L$ is the single-sided noise equivalent bandwidth in hertz, and the dimensions of $\omega_n$ are rad/s.
Hence, the closed-loop system for the linearized PLL is second-order.

It is customary to express the denominator of $H(s)$ in the standard form:

$$D(s) = s^2 + 2\zeta\omega_n s + \omega_n^2$$

where

- $\xi$: loop damping factor
- $\omega_n$: natural frequency of the loop

$$\omega_n = \sqrt{K/\tau_1} \quad \text{and} \quad \xi = \omega_n \left(\tau_2 + 1/K\right)/2$$

The closed-loop transfer function becomes:

$$H(s) = \frac{\left(2\zeta\omega_n - \omega_n^2 / K\right)s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$
The Phase-Locked Loop

- The frequency response of a second-order loop (with $\tau_1 \gg 1$)

- $\xi = 1 \Rightarrow$ critically damped loop response.
- $\xi < 1 \Rightarrow$ underdamped response.
- $\xi > 1 \Rightarrow$ overdamped response.
In practice, the selection of the bandwidth of the PLL involves a trade-off between speed of response and noise in the phase estimate.

On the one hand, it is desirable to select the bandwidth of the loop to be sufficiently wide to track any time variations in the phase of the received carrier.

On the other hand, a wideband PLL allows more noise to pass into the loop, which corrupts the phase estimate.