Chapter 3

Angle Modulation

3.1 Introduction

Besides using the amplitude of a carrier to carrier information, one can also use the angle of a carrier to carrier information. This approach is called angle modulation, and includes frequency modulation (FM) and phase modulation (PM). The amplitude of the carrier is maintained constant. The major advantage of this approach is that it allows the trade-off between bandwidth and noise performance. An angle modulated signal can be written as

$$s(t) = A\cos(\theta(t))$$

where $\theta(t)$ is usually of the form

$$\theta(t) = 2\pi f_c t + \phi(t)$$

and f_c is the carrier frequency. The signal $\phi(t)$ is derived from the message signal $m(t)$. If

$$\phi(t) = k_p m(t) \tag{3.1}$$

for some constant k_p, the resulting modulation is called phase modulation. The parameter k_p is called the phase sensitivity. If

$$\phi(t) = 2\pi k_f \int_0^t m(\tau)d\tau \tag{3.2}$$

for some constant k_f, the resulting modulation is called frequency modulation. The parameter k_f is called the frequency sensitivity. Notice that PM and FM differ only in the interpretation of message
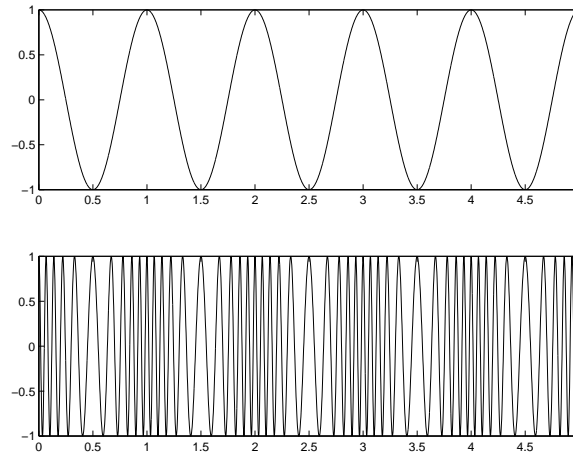
Figure 3.1: Frequency modulation

signals.

The instantaneous frequency $f_i(t)$ of an angle modulated signal is defined by

$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t). \tag{3.3}$$

For PM and FM,

$$f_i(t) = \begin{cases} f_c + \frac{k_p}{2\pi}\frac{d}{dt}m(t) & \text{for PM} \\ f_c + k_f m(t) & \text{for FM} \end{cases}$$

The maximum phase deviation in PM is called the modulation index β_p of PM.

$$\beta_p = k_p \max[|m(t)|] \tag{3.4}$$

Suppose that the message signal is a pure sinusoid $a\cos(2\pi f_m t)$ or $a\sin(2\pi f_m t)$. Then the maximum frequency deviation Δf divided by the message frequency is called the modulation index β_f of FM.

$$\beta_f = \frac{k_f a}{f_m} \tag{3.5}$$

In general, the maximum frequency deviation Δf divided by the message bandwidth W is called of deviation ratio D of FM.

$$D = \frac{k_f \max[|m(t)|]}{W} \tag{3.6}$$

3.2   Frequency Modulation

Narrowband FM

Consider that the message signal is the pure sinusoid $a\cos(2\pi f_m t)$. Then it is easy to check that

$$
\begin{aligned}
s(t) &= A\cos(2\pi f_c t + \beta\sin(2\pi f_m t)) \\
&= A\cos(\beta\sin(2\pi f_m t))\cos(2\pi f_c t) - A\sin(\beta\sin(2\pi f_m t))\sin(2\pi f_c t)
\end{aligned}
$$

Suppose that $\beta << 1$. Then

$$
\begin{aligned}
\cos(\beta\sin(2\pi f_m t)) &\approx 1, \\
\sin(\beta\sin(2\pi f_m t)) &\approx \beta\sin(2\pi f_m t).
\end{aligned}
$$

Therefore,

$$
s(t) \approx A\cos(2\pi f_c t) - A\beta\sin(2\pi f_m t)\sin(2\pi f_c t) \tag{3.7}
$$

Notice that the result is very similar to AM modulation. The main difference is that the message is now modulated onto the quadrature carrier instead of the in-phase carrier. Narrowband FM signals can be approximately generated with this approximation. The envelope of the resultant signal is given by

$$
A\sqrt{1 + \beta^2\sin^2(2\pi f_m t)}
$$

which is not constant. However, if β is small, it would be close to a constant. Notice that for the same modulation index, the amplitude variation of an AM signal is much larger than that of a narrowband FM signal.

General FM

Consider that the message signal is the pure sinusoid. Then the FM signal is of the form

$$
\begin{aligned}
s(t) &= A\cos(2\pi f_c t + \beta\sin(2\pi f_m t)) \\
&= \mathrm{Re}[Ae^{j2\pi f_c t}e^{j\beta\sin(2\pi f_m t)}]
\end{aligned}
$$

Consider the term $e^{j\beta\sin(2\pi f_m t)}$. Since $\sin(2\pi f_m t)$ is periodic with period $T_m = 1/f_m$, $e^{j\beta\sin(2\pi f_m t)}$ is periodic with period T_m. We find its Fourier series representation. The Fourier coefficient c_n is given by

$$
\begin{aligned}
c_n &= \frac{1}{T_m}\int_{-T_m/2}^{T_m/2} e^{j\beta\sin(2\pi f_m t)}e^{-j2\pi n f_m t}dt \\
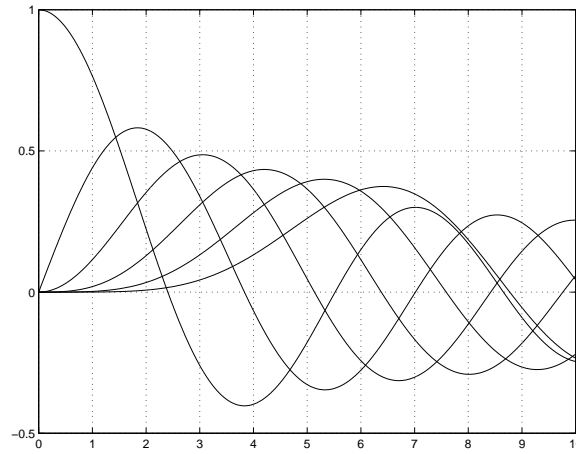&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta\sin u - nu)}du
\end{aligned}
$$

Figure 3.2: Bessel functions

The last integral is the *Bessel function of the first kind of order n* evaluated at β, and is usually denoted by $J_n(\beta)$. Notice that $J_n(\beta)$ is real.

With the Fourier series representation,

$$e^{j\beta \sin(2\pi f_m t)} = \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi n f_m t}$$

Therefore,

$$s(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + nf_m)t] \qquad (3.8)$$

Notice that $s(t)$ contains an infinite number of frequency components of the form $f_c + nf_m$.

The Bessel Functions

There are various equivalent definitions of the Bessel functions. We take the following definition. The Bessel function of the first kind of order n is defined by

$$J_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(x \sin \theta - n\theta)} d\theta. \qquad (3.9)$$

We note the following properties of the Bessel functions:

1. $J_n(x)$ is real.

    Notice that

    $$e^{j(x \sin \theta - n\theta)} = \cos(x \sin \theta - n\theta) + j \sin(x \sin \theta - n\theta).$$

    Since

    $$\sin(x \sin \theta - n\theta) = \sin(x \sin \theta) \cos(n\theta) - \cos(x \sin \theta) \sin(n\theta)$$

is an odd function of θ, its integral from $-\pi$ to π is zero. Notice that $\cos(x \sin \theta - n\theta)$ is an even function of θ. Therefore, $J_n(x)$ can also be expressed as

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin \theta - n\theta) d\theta.$$

2. $J_{-n}(x) = (-1)^n J_n(x)$.

   Using the last result,

   $$J_{-n}(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin \theta + n\theta) d\theta.$$

   Put $\phi = \pi - \theta$.

   $$\begin{aligned} J_{-n}(x) &= \frac{1}{\pi} \int_0^\pi \cos(x \sin \phi - n\phi + n\pi) d\phi \\ &= \frac{1}{\pi} \int_0^\pi [\cos(x \sin \phi - n\phi) \cos(n\pi) - \sin(x \sin \phi - n\phi) \sin(n\pi)] d\phi \\ &= (-1)^n J_n(x) \end{aligned}$$

3. $J_n(x)$ has the following Taylor series expansion (proof omitted):

   $$J_n(x) = \sum_{m=0}^\infty \frac{(-1)^m (x/2)^{n+2m}}{m!(m+n)!}$$

   For small x, the series is dominated by its first term. Therefore,

   $$J_n(x) \approx \frac{x^n}{2^n n!}$$

   When the values of $J_n(x)$ are compared across different n for small x,

   $$\begin{aligned} J_0(x) &\approx 1 \\ J_1(x) &\approx \frac{x}{2} \\ J_n(x) &\approx 0 \quad \text{for } n > 1 \end{aligned}$$

   Towards the other extreme, for large x,

   $$J_n(x) \approx \sqrt{\frac{2}{\pi x}} \cos(x - \frac{\pi}{4} - \frac{n\pi}{2})$$

   which a decaying sinusoid.

4. For all x, $\sum_{n=-\infty}^\infty J_n^2(x) = 1$ (proof omitted).

   A consequence is that for any fixed x, the infinite sum is dominated by a finite number of terms.

Observations on FM signals

With knowledge about the Bessel functions, we can have the following observations on tone-modulated FM signals. Recall that

$$s(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + nf_m)t]$$

1. $s(t)$ contains an infinite number of frequency components of the form $f_c + nf_m$, where $m = 0, \pm1, \pm2, \ldots$

2. The power of $s(t)$ is distributed among the components. The power carried by the component $f_c + nf_m$ is $A^2 J_n^2(\beta)/2$. The total power is $\sum A^2 J_n^2(\beta)/2 = A^2/2$, which is reasonable since the overall FM signal has a constant amplitude A.

3. Most of the power is carried by the components $f_c + nf_m$ for $|n| \le \beta + 1$. (Since $\sum J_n^2(\beta) = 1$, we know that most of the power is carried by a finite number of components. The value $\beta + 1$ is determined empirically.)

4. For small β, using the approximations for the Bessel functions, we have

$$\begin{aligned} s(t) &\approx A\cos(2\pi f_c t) + \tfrac{A\beta}{2}[\cos(2\pi(f_c + f_m)t) - \cos(2\pi(f_c - f_m)t)] \\ &= A\cos(2\pi f_c t) - A\beta\sin(2\pi f_m t)\sin(2\pi f_c t) \end{aligned}$$

which is the narrowband FM approximation we have used before.

In theory, the FM signal occupies a spectrum of infinite bandwidth. However, since most of the power is carried by a finite number of components, we can define an effective bandwidth. For a tone-modulated FM signal, it is defined by

$$B = 2(\beta + 1)f_m = 2(\Delta f + f_m)$$

For a general message signal with bandwidth W, the effective bandwidth of the corresponding FM signal is defined similarly by

$$B = 2(\Delta f + W) = 2(D + 1)W. \tag{3.10}$$

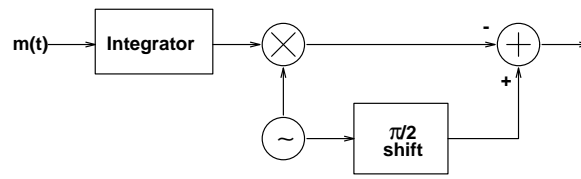This empirical formula is often referred to as Carson's rule.
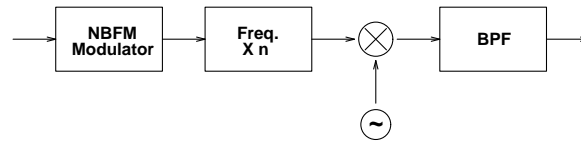
Figure 3.3: Generation of narrowband FM

Figure 3.4: Generation of wideband FM via frequency multiplication

3.3 Modulators and Demodulators

Generation of FM signals

An FM signal can be generated with a voltage controlled oscillator (VCO). Given an input signal $m(t)$, the instantaneous output frequency of a VCO (with nominal frequency f_c) is of the form

$$f_i(t) = f_c + km(t)$$

which is the desired format for FM.

Another approach is to first generate a narrowband FM signal. Notice that a narrowband FM signal can be approximately generated as in Fig. 3.3. Then the frequency of the resulting signal is multiplied to give the desired frequency deviation as shown in Fig. 3.4. A frequency multiplier consists of a non-linear device with input-output relationship of the form

$$y(t) = a_1 x(t) + a_2 x^2(t) + \ldots + a_n x^n(t)$$

followed by an appropriate bandpass filter. Ideally, if the instantaneous input frequency is $f_i(t)$, then the instantaneous output frequency is $n f_i(t)$. When the desired frequency deviation is obtained, the signal is translated to the desired carrier frequency with a mixer and a BPF.

Example: Suppose that the input narrowband FM signal has a frequency deviation of 2.5 kHz with a carrier frequency of 10 MHz, the desired frequency deviation is 50 kHz, and the desired carrier is 100 MHz. Therefore, $n = 20$ should be chosen. As a result, the multiplied carrier frequency is 200

MHz. We can mix the result with a 100 MHz carrier to obtain the desired carrier frequency. Of course, it is also possible to do frequency translation before frequency multiplication. We can translate the narrowband FM signal to 5 MHz. Then multiply the resultant signal in frequency by 20 times.

Demodulation of FM signals

One approach to FM demodulation is to generate an AM signal with amplitude proportional to the instantaneous frequency of the FM signal, and then to recover the message signal with an AM de-modulator. Ideally, FM to AM conversion can be achieved with a differentiator. Consider the FM signal.

$$s(t) = A\cos(2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau)$$

Its derivative is given by

$$\frac{d}{dt}s(t) = -A(2\pi f_c + 2\pi k_f m(t))\sin(2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau) \tag{3.11}$$

Notice that the amplitude varies according to the message signal.

The transfer function of the ideal differentiator is given by

$$H(f) = j2\pi f$$

which is linear. In general, it is hard to realize this linear transfer function over a large range of frequencies. However, since the power of an FM signal concentrates around the carrier frequency f_c, we can try to build a filter with linear transfer function around f_c, i.e.,

$$H(f) = \begin{cases} j2\pi k(f - f_c + C) & \text{for} \quad |f - f_c| < B/2 \\ j2\pi k(f + f_c - C) & \text{for} \quad |f + f_c| < B/2 \end{cases}$$

where B is the bandwidth of the FM signal. The Fourier transform of the output is given by

$$H(f)S(f) = j2\pi k f S(f) - j2\pi k sgn(f)(f_c - C)S(f).$$

In time domain, the output is

$$k\frac{ds(t)}{dt} + 2\pi k(f_c - C)\hat{s}(t) = -2\pi kCA\left[1 + \frac{k_f}{C}m(t)\right]\sin(2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau)$$

with its envelope varying according to $m(t)$. If a sufficiently large C is chosen, the message signal can be recovered with an envelope detector.
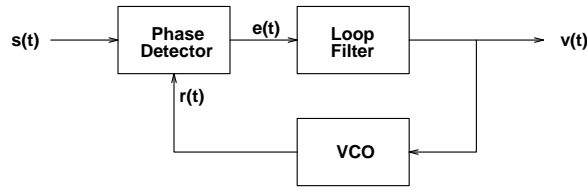
Figure 3.5: The Phase Locked Loop

Phase Locked Loops

The phase locked loop (PLL) finds applications in different areas of communications, including carrier phase synchronization and FM demodulation. A typical PLL is shown in Fig. 3.5. We first consider using a PLL to track the phase of a carrier. The operation of the PLL is as follows:

- The carrier to be "locked" is $s(t) = \cos(2\pi f_c t + \phi)$.

- The reference signal is $r(t) = -\sin(2\pi f t + \phi_r(t))$.

- In the simplest case, the phase detector is just a multiplier followed by a low-passed filter (to remove the double frequency term). Therefore the output of the phase detector is the error signal $e(t) = K_p \sin(\phi - \phi_r(t))$ where K_p is the constant gain of the filter over its passband.

- The loop filter determines the performance of the PLL when noise is present. For simplicity, we assume that it provides a constant gain K_l. Therefore, $v(t) = K_l e(t)$.

- The voltage controlled oscillator adjusts the frequency (and, hence, phase) of its output according to the relations

$$\frac{d}{dt}\phi_r(t) = K_v v(t)$$

where K_v is a constant gain.

Therefore,

$$\frac{d}{dt}\phi_r(t) = K \sin(\phi - \phi_r(t))$$

where the overall gain $K = K_v K_l K_p$. Suppose that $\phi_r(t)$ is close to ϕ. Then, approximately,

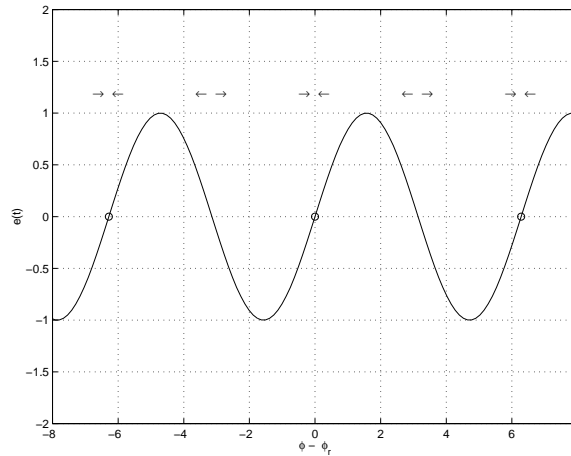$$\frac{d}{dt}\phi_r(t) = K(\phi - \phi_r(t)).$$

Figure 3.6: Drift of the phase reference to a stable equilibrium

The solution of this differential equation is given by

$$\phi_r(t) = \phi - (\phi - \phi_r(0)) \exp(-Kt).$$

Clearly, $\phi_r(t)$ tends to ϕ. Notice that the overall gain controls the speed of convergence. In case that the initial phase reference $\phi_r(t)$ is not close to ϕ. The situation is depicted in Fig. 3.6. $\phi_r(t)$ will drift to the vicinity of $\phi + 2\pi k$ for some integer k. Then locking begins.

Now consider using the PLL for FM demodulation. The FM signal is of the form $s(t) = A\cos(2\pi f_c t + \phi(t))$ where $\phi(t) = 2\pi k_f \int_0^t m(\tau)d\tau$. Suppose that $\phi(t)$ varies so slowly that it is tracked by the PLL, i.e., $\phi_r(t) \approx \phi(t)$. Then

$$v(t) = \frac{1}{K_v}\frac{d}{dt}\phi_r(t) \approx \frac{1}{K_v}\frac{d}{dt}\phi(t) = \frac{2\pi k_f}{K_v}m(t)$$

is the desired message signal.