Angle Modulation

Phase and Frequency Modulation

Consider a signal of the form $\mathrm{x}_c(t) = A_c \cos(2\pi f_c t + \phi(t))$ where A_c and f_c are constants. The envelope is a constant so the message cannot be in the envelope. It must instead lie in the variation of the cosine argument with time. Let $\theta_i(t) \triangleq 2\pi f_c t + \phi(t)$ be the **instantaneous phase**. Then

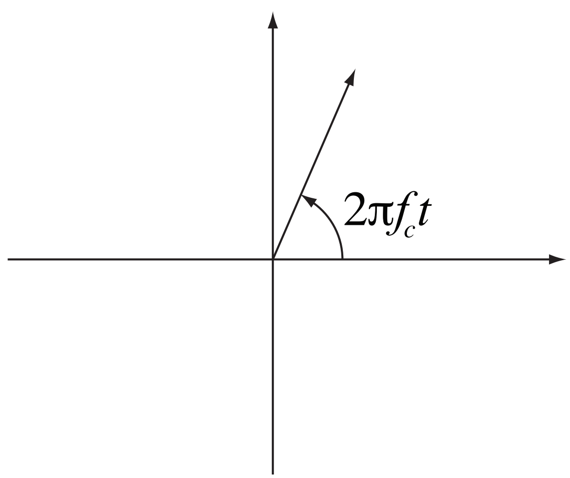
$$\mathrm{x}_c(t) = A_c \cos(\theta_i(t)) = A_c \operatorname{Re}\left(e^{j\theta_i(t)}\right).$$

$\theta_i(t)$ contains the message and this type of modulation is called **angle** or **exponential** modulation. If $\phi(t) = k_p \,\mathrm{m}(t)$ so that $\mathrm{x}_c(t) = A_c \cos(2\pi f_c t + k_p \,\mathrm{m}(t))$ the modulation is called **phase modulation (PM)** where k_p is the **deviation constant** or **phase modulation index**.

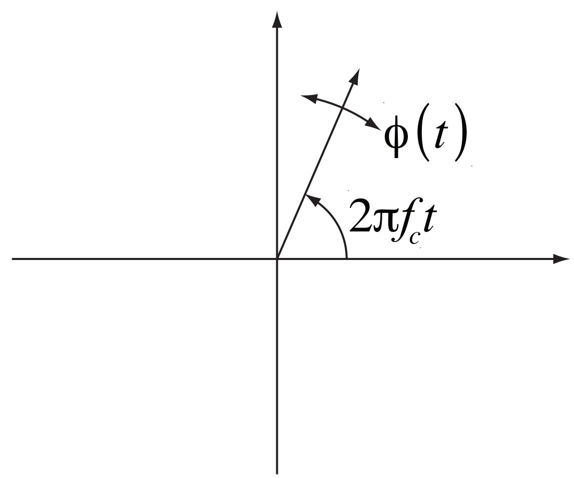
Phase and Frequency Modulation

Think about what it means to modulate the phase of a cosine. The total argument of the cosine is $2\pi f_c t + \phi(t)$, an angle with units of radians (or degrees). When $\phi(t) = 0$, we simply have a cosine and the angle $2\pi f_c t$ is a linear function of time. Think of this angle as the angle of a phasor rotating at a constant angular velocity. Now add the effect of the phase modulation $\phi(t)$. The modulation adds a "wiggle" to the rotating phasor with respect to its position when it is unmodulated. The message is in the variation of the phasor's angle with respect to the constant angular velocity of the unmodulated cosine.

Unmodulated Cosine



Modulated Cosine



Phase and Frequency Modulation

The total argument of an unmodulated cosine is $\theta_c(t) = 2\pi f_c t$ in which f_c is a cyclic frequency. The time derivative of $2\pi f_c t$ is $2\pi f_c$. We could also express the argument in radian frequency form as $\theta_c(t) = \omega_c t$. Its time derivative is ω_c. Therefore one way of defining the cyclic frequency of an unmodulated cosine is as $\frac{1}{2\pi}\frac{d}{dt}\left(\theta_c(t)\right)$. Now let's apply this same idea to a modulated cosine whose argument is $\theta_c(t) = 2\pi f_c t + \phi(t)$. Its time derivative is $2\pi f_c + \frac{d}{dt}\left(\phi(t)\right)$. Now we define **instantaneous frequency** as $\mathrm{f}(t) \triangleq \frac{1}{2\pi}\frac{d}{dt}\left(\theta_c(t)\right) = \frac{1}{2\pi}\left[2\pi f_c + \frac{d}{dt}\left(\phi(t)\right)\right] = f_c + \frac{1}{2\pi}\frac{d}{dt}\left(\phi(t)\right)$. It is important to draw a distinction between instantaneous frequency $\mathrm{f}(t)$ and spectral frequency f. They are definitely not the same. Let $\mathrm{x}_c(t) = \cos\left(2\pi f_c t + \phi(t)\right)$. It has a Fourier transform $\mathrm{X}_c(f)$. Spectral frequency f is the independent variable in $\mathrm{X}_c(f)$ but $\mathrm{f}(t) = f_c + \frac{1}{2\pi}\frac{d}{dt}\left(\phi(t)\right)$. Some Fourier transforms of phase and frequency modulated signals later will make this distinction clearer.

Phase and Frequency Modulation

If we make the variation of the instantaneous frequency of a sinusoid be directly proportional to the message we are doing **frequency modulation** $(\mathbf{FM})$. If $\frac{d\phi}{dt} = k_f \,\mathrm{m}(t)$ then k_f is the **frequency deviation constant** in radians/second per unit of $\mathrm{m}(t)$. In frequency modulation $\mathrm{f}(t) = f_c + f_d \,\mathrm{m}(t)$, where $f_d = \frac{k_f}{2\pi}$ is the frequency deviation constant in Hz per unit of $\mathrm{m}(t)$. In frequency modulation

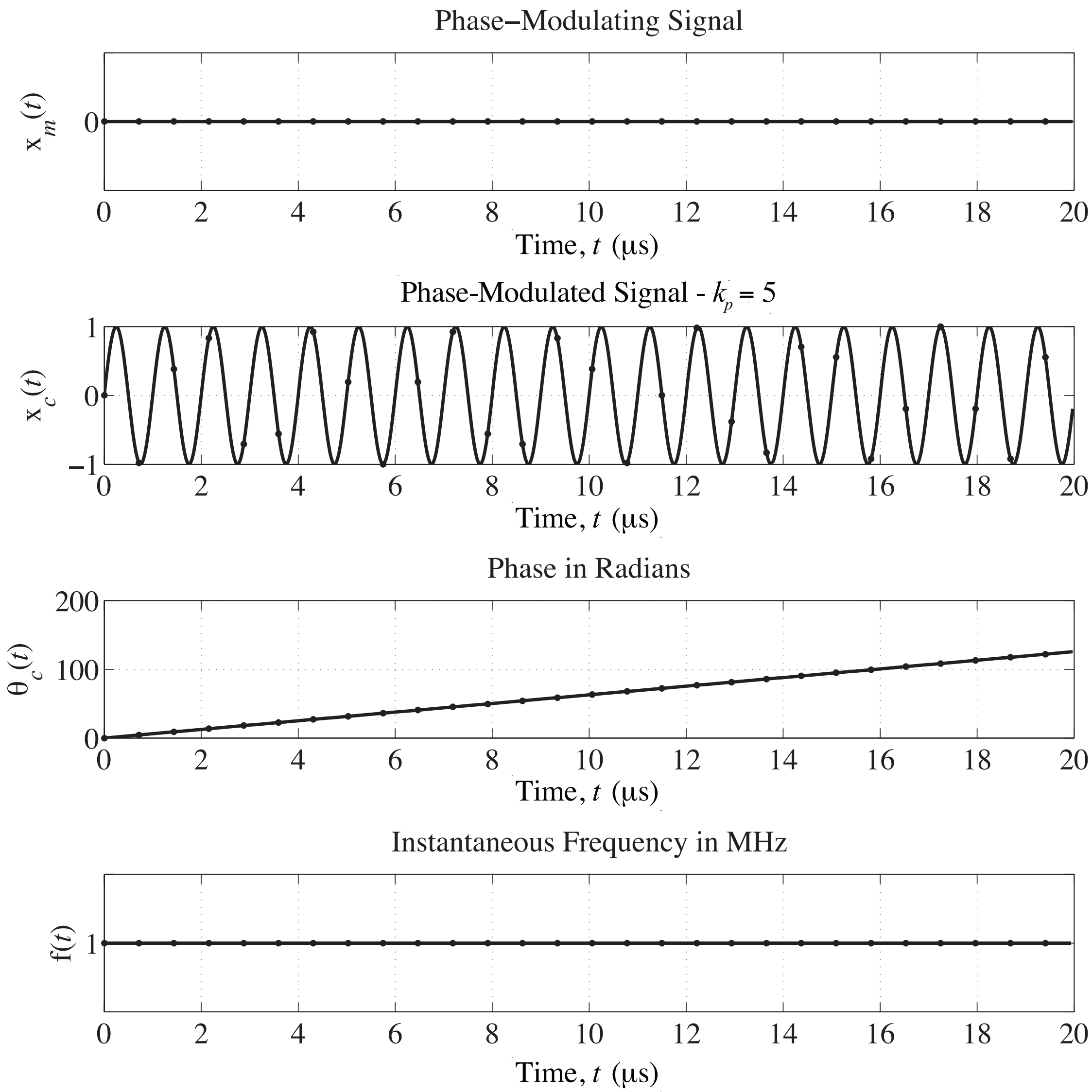
$$\phi(t) = k_f \int_{t_0}^{t} \mathrm{m}(\lambda)\, d\lambda + \phi(t_0) = 2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\, d\lambda + \phi(t_0) \quad , \; t \geq t_0$$

therefore

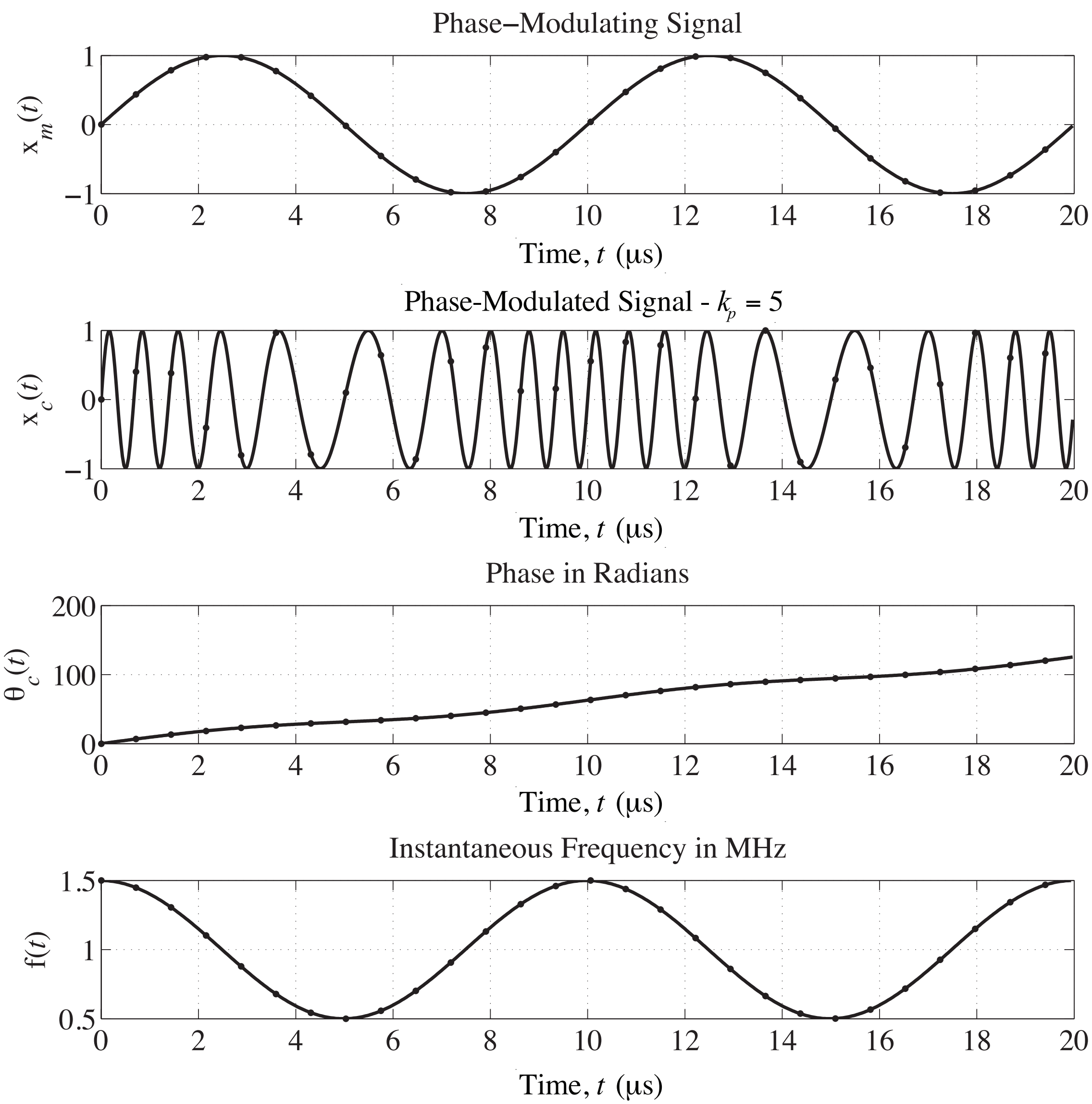
$$\mathrm{x}_c(t) = A_c \cos\left(2\pi f_c t + 2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\, d\lambda + \phi(t_0) \right).$$

So PM and FM are very similar. The difference is between integrating the message signal before phase modulating or not integrating it.

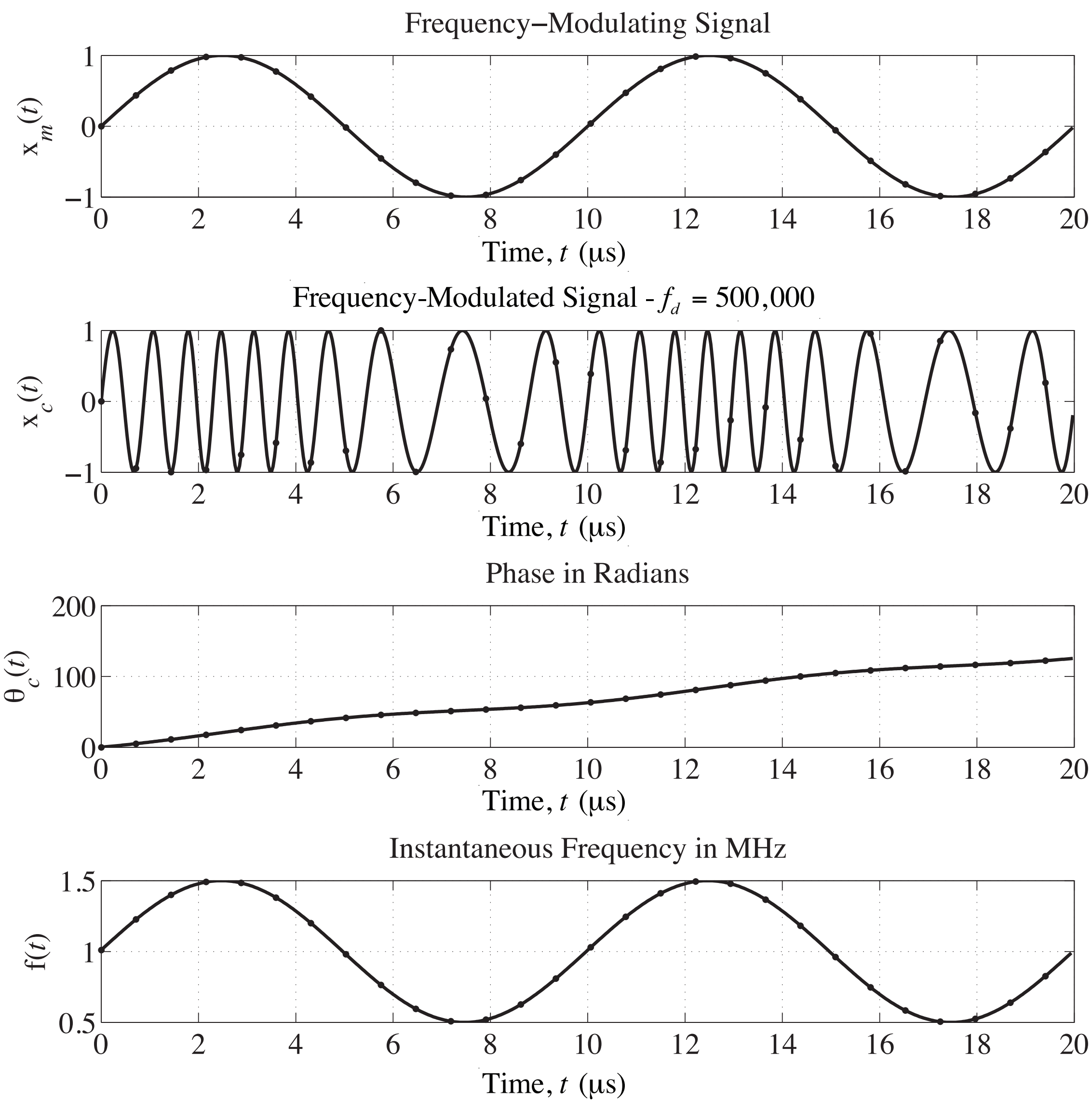
Phase and Frequency Modulation



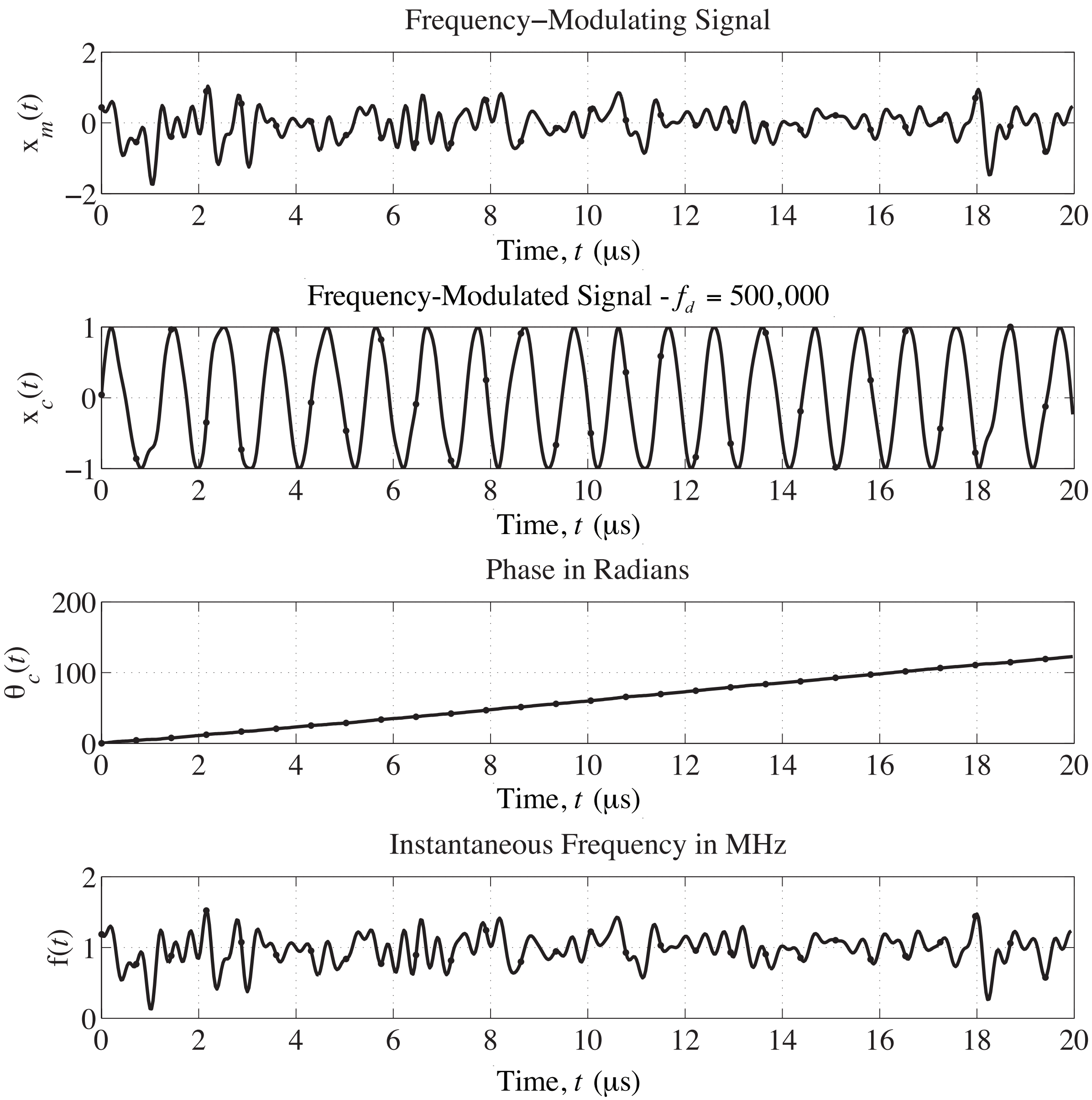
Phase and Frequency Modulation



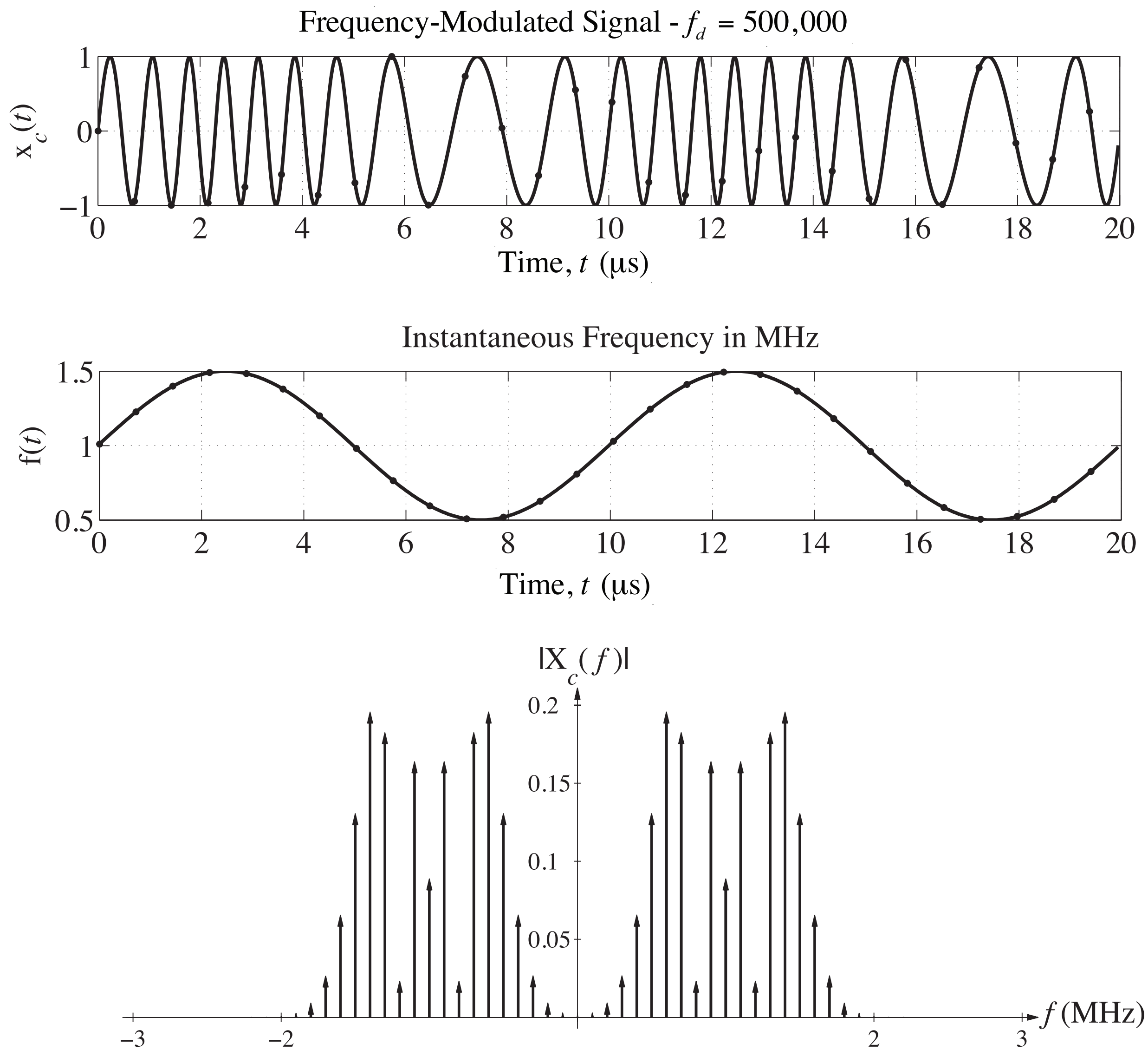
Phase and Frequency Modulation



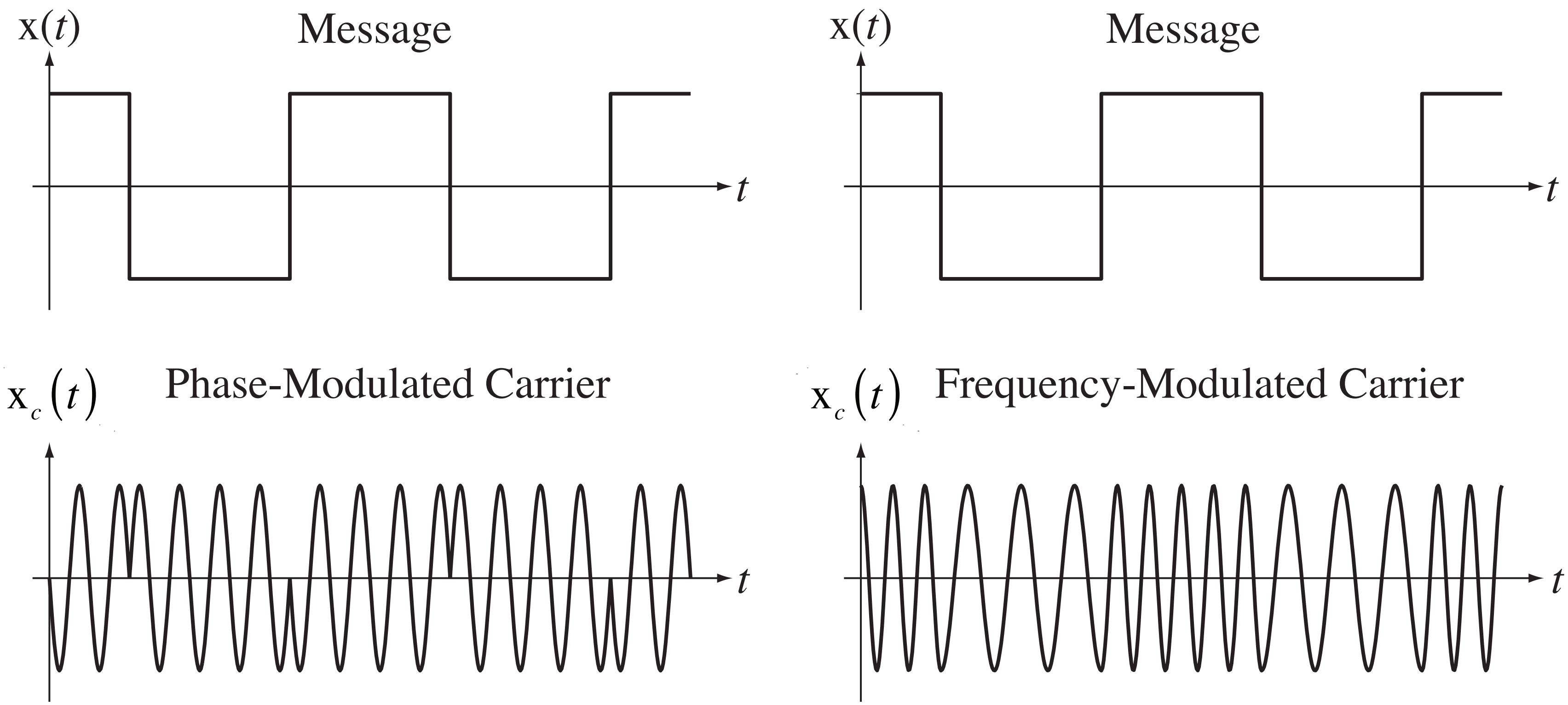
Phase and Frequency Modulation



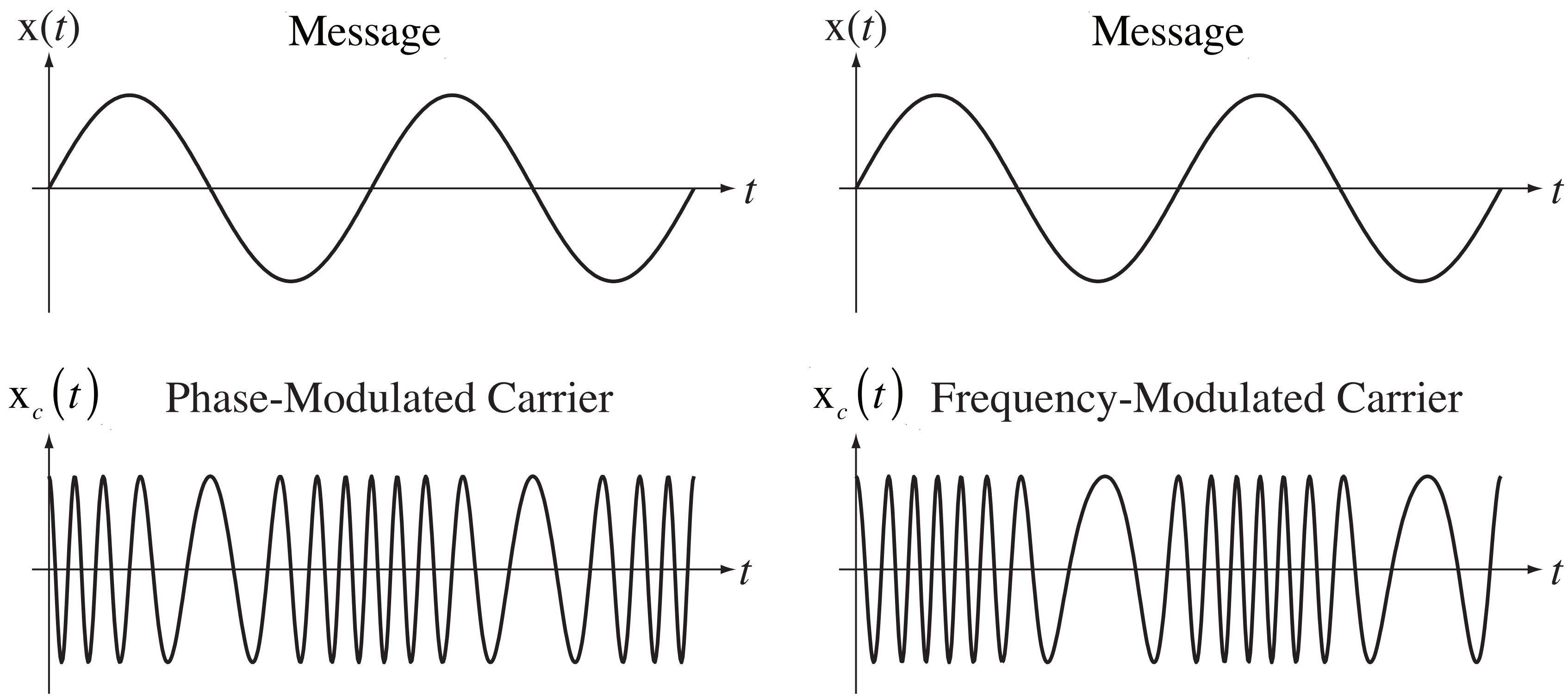
Phase and Frequency Modulation



Phase and Frequency Modulation



Phase and Frequency Modulation



Phase and Frequency Modulation

For phase modulation $\mathrm{x}_c(t) = A_c \cos\left(2\pi f_c t + k_p\, \mathrm{m}(t)\right)$

For frequency modulation $\mathrm{x}_c(t) = A_c \cos\left(2\pi f_c t + 2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\, d\lambda\right)$

There is no simple expression for the Fourier transforms of these signals in the general case. Using $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$

we can write for PM $\mathrm{x}_c(t) = A_c\left[\cos(2\pi f_c t)\cos\left(k_p\, \mathrm{m}(t)\right) - \sin(2\pi f_c t)\sin\left(k_p\, \mathrm{m}(t)\right)\right]$

and for FM $\mathrm{x}_c(t) = A_c\left[\cos(2\pi f_c t)\cos\left(2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\, d\lambda\right) - \sin(2\pi f_c t)\sin\left(2\pi f_d \int_{t_0}^{t} \mathrm{x}(\lambda)\, d\lambda\right)\right]$

(under the assumption that $\phi(t_0) = 0$).

Phase and Frequency Modulation

If k_p and f_d are small enough, $\cos\left(k_p\,\mathrm{m}(t)\right) \cong 1$ and $\sin\left(k_p\,\mathrm{m}(t)\right) \cong k_p\,\mathrm{m}(t)$

and $\cos\left(2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\,d\lambda\right) \cong 1$ and $\sin\left(2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\,d\lambda\right) \cong 2\pi f_d \int_{t_0}^{t} \mathrm{m}(\lambda)\,d\lambda$.

Then for PM $\quad \mathrm{x}_c(t) \cong A_c\left[\cos(2\pi f_c t) - k_p\,\mathrm{m}(t)\sin(2\pi f_c t)\right]$

and for FM $\quad \mathrm{x}_c(t) \cong A_c\left[\cos(2\pi f_c t) - 2\pi f_d \sin(2\pi f_c t)\int_{t_0}^{t} \mathrm{m}(\lambda)\,d\lambda\right]$

These approximations are called **narrowband PM** and **narrowband FM**.

Phase and Frequency Modulation

If the information signal is a sinusoid $\mathrm{m}(t) = A_m \cos(2\pi f_m t)$

then $\mathrm{M}(f) = (A_m / 2)\left[\delta(f - f_m) + \delta(f + f_m)\right]$ and, in the narrowband approximation,

For PM,

$$\mathrm{x}_c(t) \cong A_c\left[\cos(2\pi f_c t) - k_p A_m \cos(2\pi f_m t)\sin(2\pi f_c t)\right]$$

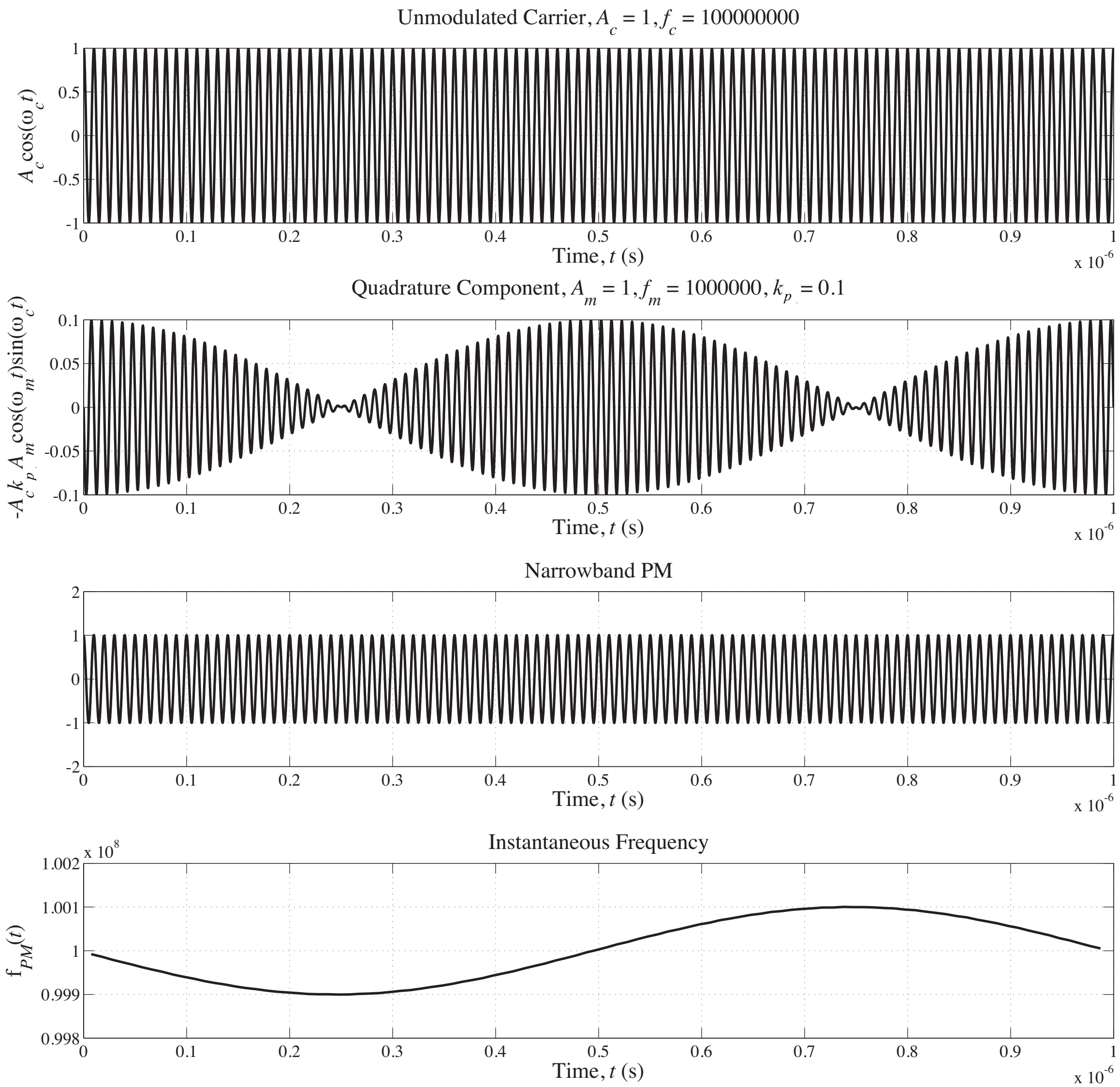
$$\mathrm{X}_c(f) \cong (A_c / 2)\left\{\left[\delta(f - f_c) + \delta(f + f_c)\right] - \frac{jA_m k_p}{2}\begin{bmatrix}\delta(f + f_c - f_m) + \delta(f + f_c + f_m) \\ -\delta(f - f_c - f_m) - \delta(f - f_c + f_m)\end{bmatrix}\right\}$$

For FM,

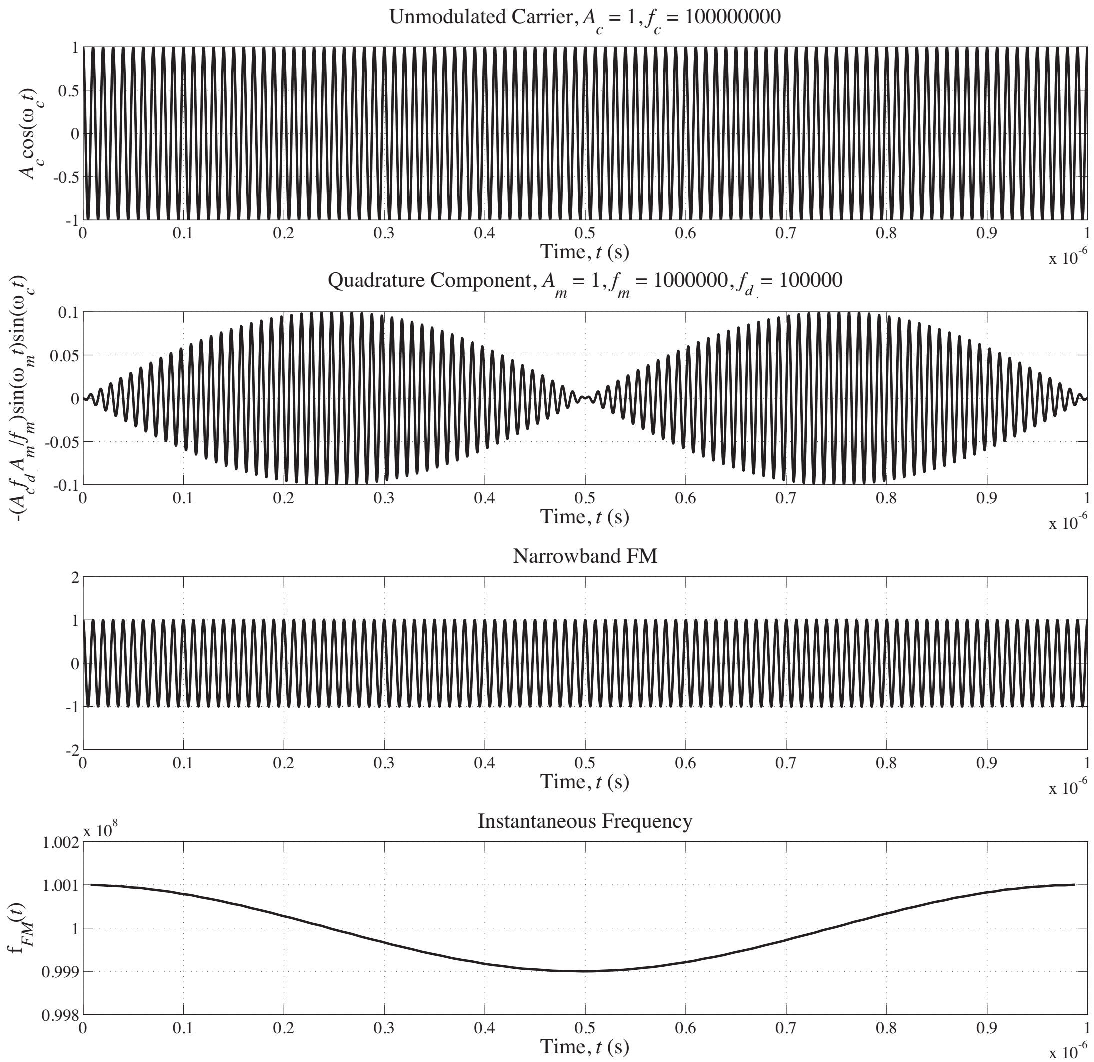
$$\mathrm{x}_c(t) \cong A_c\left[\cos(2\pi f_c t) - \frac{2\pi f_d A_m}{2\pi f_m}\sin(2\pi f_c t)\sin(2\pi f_m t)\right]$$

$$\mathrm{X}_c(f) \cong (A_c / 2)\left\{\left[\delta(f - f_c) + \delta(f + f_c)\right] - \frac{A_m f_d}{2 f_m}\begin{bmatrix}\delta(f + f_c - f_m) - \delta(f + f_c + f_m) \\ -\delta(f - f_c - f_m) + \delta(f - f_c + f_m)\end{bmatrix}\right\}$$

Phase and Frequency Modulation



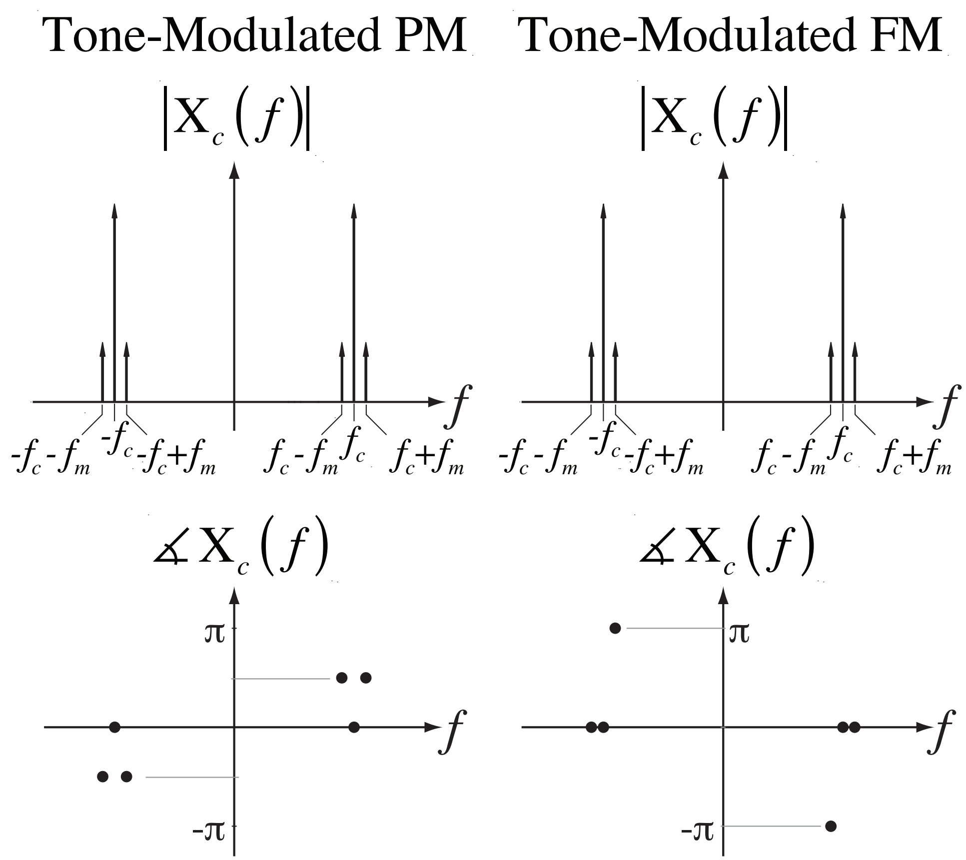
Phase and Frequency Modulation



Phase and Frequency Modulation

Narrowband PM and FM Spectra

for Tone Modulation



Phase and Frequency Modulation

If the information signal is a sinc, $\mathrm{x}(t)=\mathrm{sinc}(2Wt)$ then $\mathrm{X}(f)=(1/2W)\mathrm{rect}(f/2W)$ and, in the narrowband approximation,

For PM,

$$\mathrm{X}_c(f)\cong(A_c/2)\left\{\left[\delta(f-f_c)+\delta(f+f_c)\right]-j\frac{k_p}{2W}\left[\mathrm{rect}\big((f+f_c)/2W\big)-\mathrm{rect}\big((f-f_c)/2W\big)\right]\right\}$$

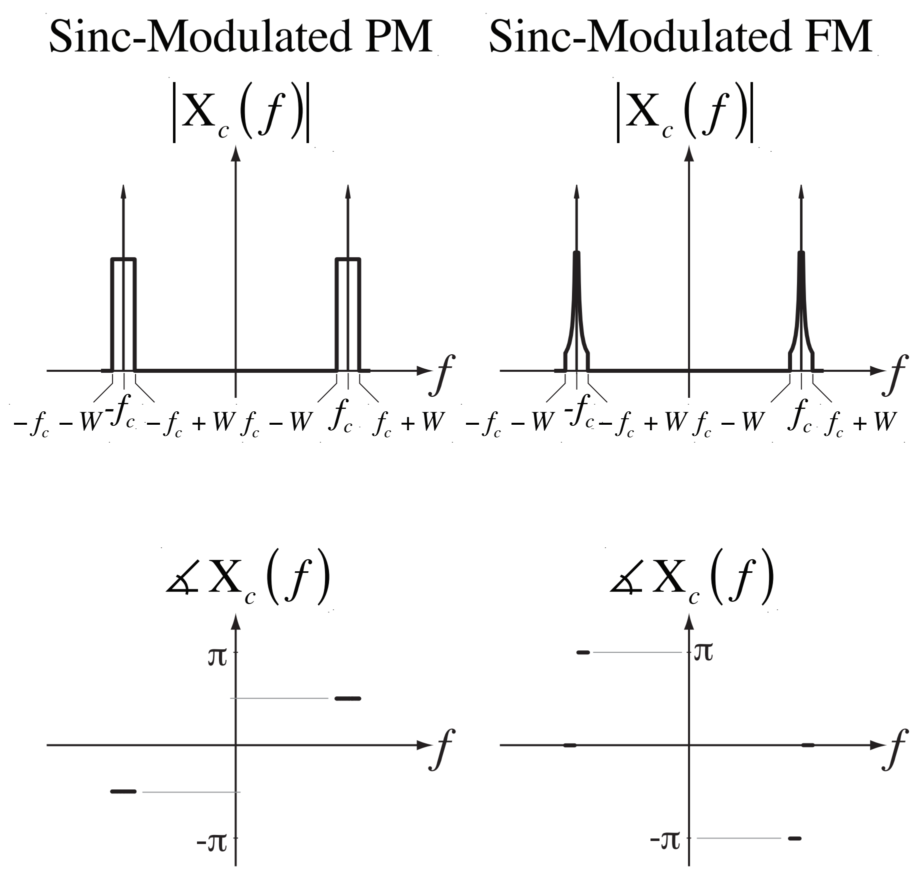
For FM,

$$\mathrm{X}_c(f)\cong(A_c/2)\left\{\left[\delta(f-f_c)+\delta(f+f_c)\right]-\frac{f_m k_f}{2W}\left[\frac{\mathrm{rect}\big((f+f_c)/2W\big)}{f+f_c}-\frac{\mathrm{rect}\big((f-f_c)/2W\big)}{f-f_c}\right]\right\}$$

Phase and Frequency Modulation

Narrowband PM and FM Spectra

for a Sinc Message



Phase and Frequency Modulation

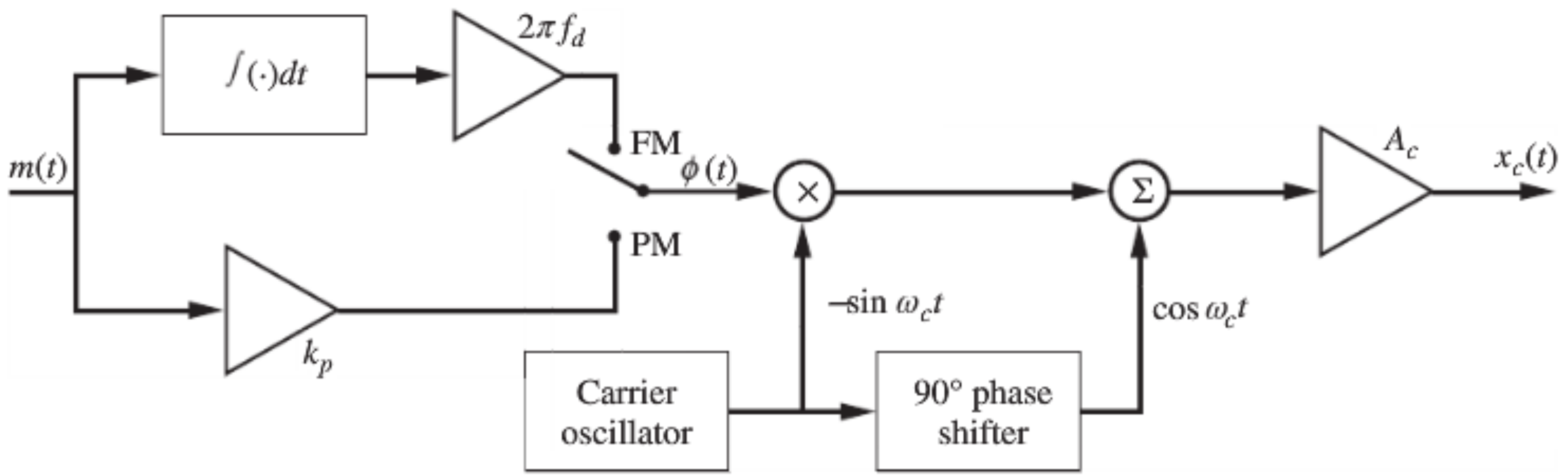


Figure 4.3
Generation of narrowband angle modulation.

Phase and Frequency Modulation

If the narrowband approximation is not adequate we must deal with the more complicated wideband case. In the case of tone modulation we can handle PM and FM with basically the same analysis technique if we use the following conventions:

$$\mathrm{x}(t)=\begin{cases} A_m \sin(2\pi f_m t) & , \text{ PM} \\ A_m \cos(2\pi f_m t) & , \text{ FM} \end{cases}$$

For FM, $\phi(t)=2\pi f_d \int_{t_0}^{t} \mathrm{x}(\lambda)\,d\lambda = 2\pi f_d \int_{t_0}^{t} A_m \cos(2\pi f_m \lambda)\,d\lambda$

$$\phi(t)=2\pi \frac{A_m}{\omega_m} f_d \sin(2\pi f_m t)=\frac{A_m}{f_m} f_d \sin(2\pi f_m t)$$

Then, for PM and FM, $\phi(t)=\beta \sin(2\pi f_m t)$, where $\beta \triangleq \begin{cases} k_p A_m & , \text{ PM} \\ (A_m / f_m) f_d & , \text{ FM} \end{cases}$

Then $\mathrm{x}_c(t)=A_c\left[\cos(\beta \sin(2\pi f_m t))\cos(2\pi f_c t)-\sin(\beta \sin(2\pi f_m t))\sin(2\pi f_c t)\right]$

Phase and Frequency Modulation

In $x_c(t) = A_c\left[\cos\left(\beta\sin(2\pi f_m t)\right)\cos(2\pi f_c t) - \sin\left(\beta\sin(2\pi f_m t)\right)\sin(2\pi f_c t)\right]$

$\cos\left(\beta\sin(2\pi f_m t)\right)$ and $\sin\left(\beta\sin(2\pi f_m t)\right)$ are periodic with fundamental period $1/f_m$. We can now use two results from applied mathematics (Abramowitz and Stegun, page 361)

$$\cos\left(z\sin(\theta)\right) = J_0(z) + 2\sum_{k=1}^{\infty} J_{2k}(z)\cos(2k\theta) = J_0(z) + 2\sum_{\substack{k=1\\ k\ \text{even}}}^{\infty} J_k(z)\cos(k\theta)$$

$$\sin\left(z\sin(\theta)\right) = 2\sum_{k=0}^{\infty} J_{2k+1}(z)\sin\left((2k+1)\theta\right) = 2\sum_{\substack{k=1\\ k\ \text{odd}}}^{\infty} J_k(z)\sin(k\theta)$$

Adapting them to our case

$$\cos\left(\beta\sin(2\pi f_m t)\right) = J_0(\beta) + 2\sum_{\substack{k=1\\ k\ \text{even}}}^{\infty} J_k(\beta)\cos(2k\pi f_m t)$$

$$\sin\left(z\sin(2\pi f_m t)\right) = 2\sum_{\substack{k=1\\ k\ \text{odd}}}^{\infty} J_k(\beta)\sin(2k\pi f_m t)$$

Phase and Frequency Modulation

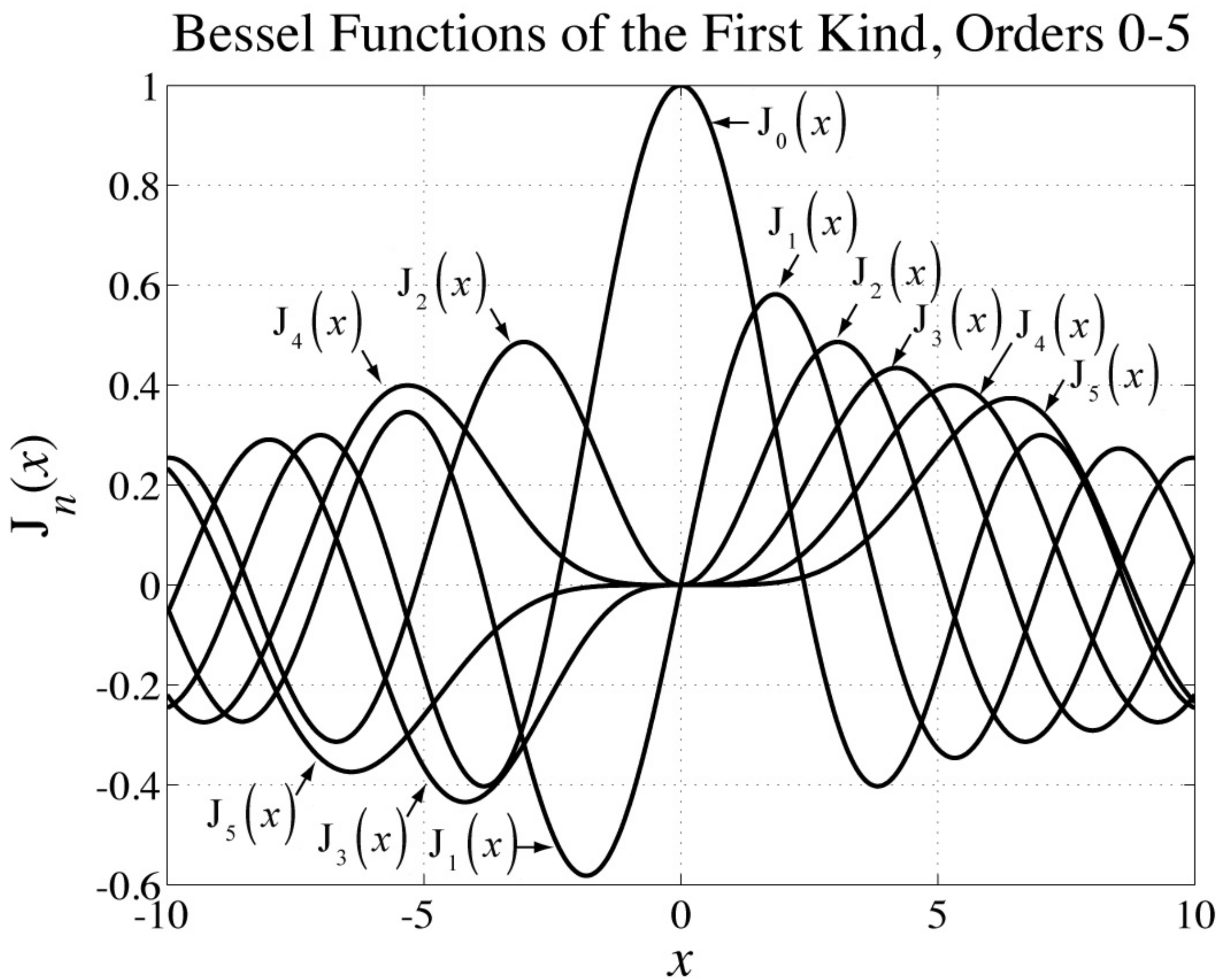
$$x_c(t) = A_c\left\{\left[J_0(\beta) + 2\sum_{\substack{k=1 \\ k \text{ even}}}^{\infty} J_k(\beta)\cos(2k\pi f_m t)\right]\cos(2\pi f_c t) - \left[2\sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} J_k(\beta)\sin(2k\pi f_m t)\right]\sin(2\pi f_c t)\right\}$$

$$x_c(t) = A_c\left\{\begin{array}{l} J_0(\beta)\cos(2\pi f_c t) + 2\sum\limits_{\substack{k=1 \\ k \text{ even}}}^{\infty} J_k(\beta)\cos(2\pi f_c t)\cos(2k\pi f_m t) \\ -2\sum\limits_{\substack{k=1 \\ k \text{ odd}}}^{\infty} J_k(\beta)\sin(2\pi f_c t)\sin(2k\pi f_m t) \end{array}\right\}$$

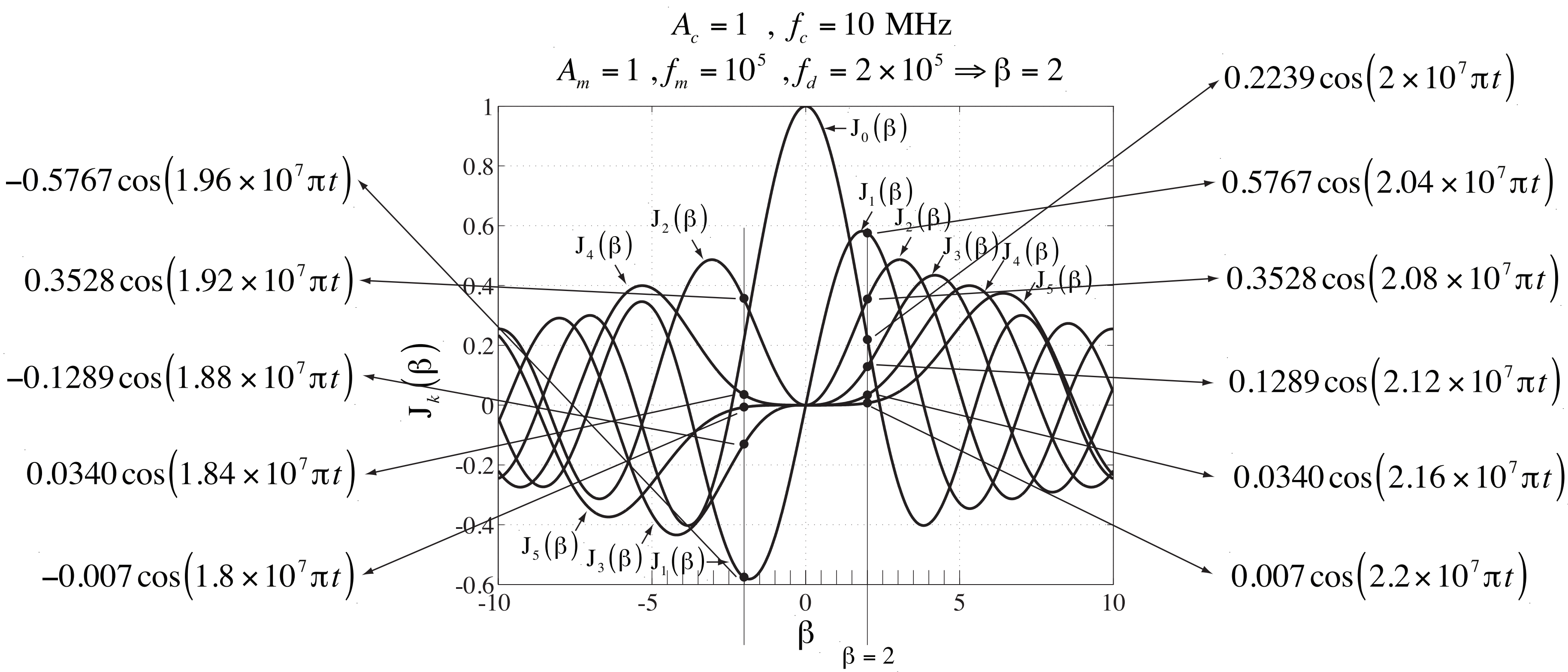
$$x_c(t) = A_c\left\{\begin{array}{l} J_0(\beta)\cos(2\pi f_c t) + \sum\limits_{\substack{k=1 \\ k \text{ even}}}^{\infty} J_k(\beta)\left[\cos(2\pi(f_c - kf_m)t) + \cos(2\pi(f_c + kf_m)t)\right] \\ -\sum\limits_{\substack{k=1 \\ k \text{ odd}}}^{\infty} J_k(\beta)\left[\cos(2\pi(f_c - kf_m)t) - \cos(2\pi(f_c + kf_m)t)\right] \end{array}\right\}$$

This can also be written in the more compact form, $x_c(t) = A_c\sum_{k=-\infty}^{\infty} J_k(\beta)\cos(2\pi(f_c + kf_m)t)$

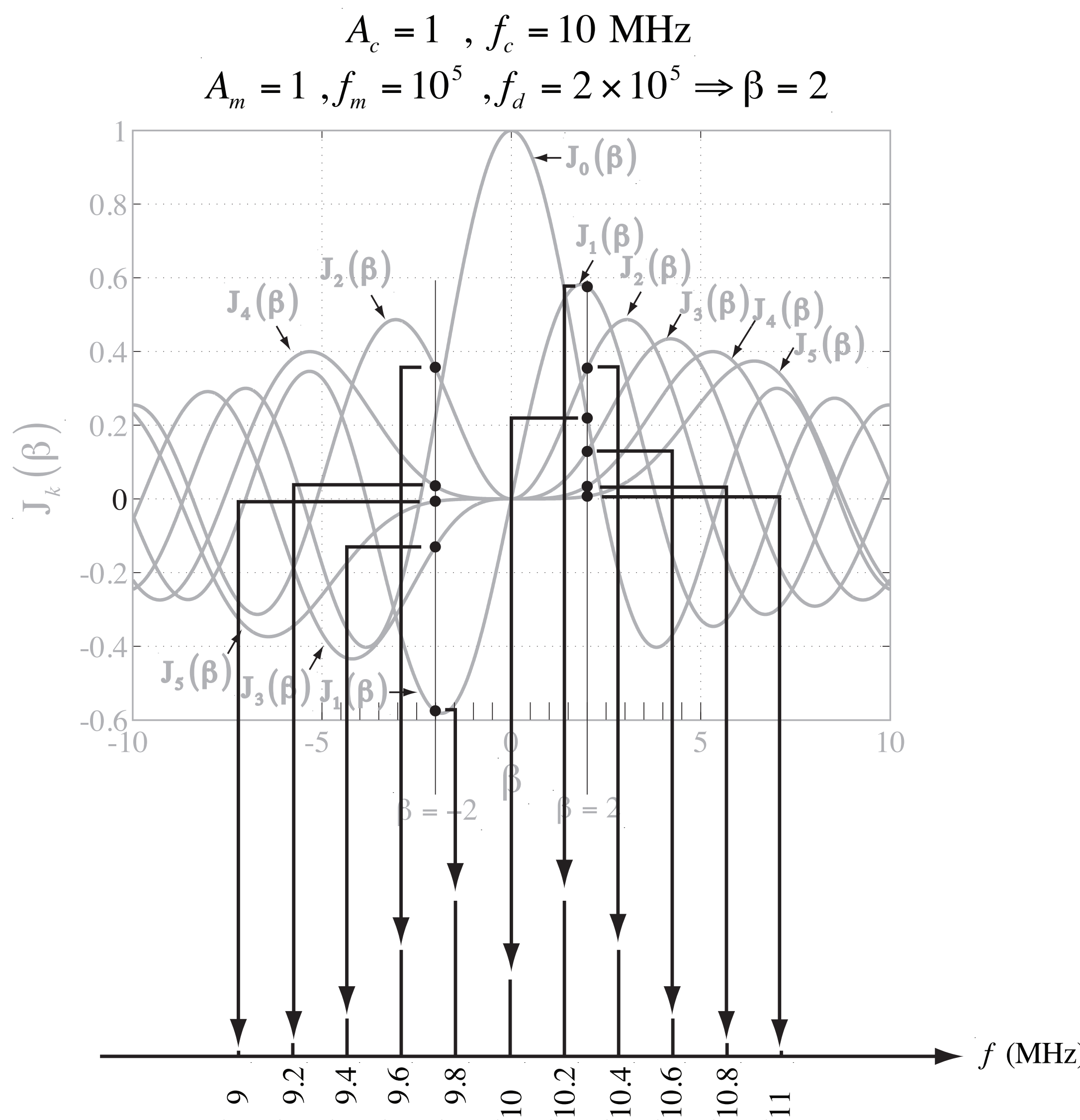
Phase and Frequency Modulation



Phase and Frequency Modulation



Phase and Frequency Modulation



Phase and Frequency Modulation

Now, to find the spectrum of $\mathrm{x}_c(t)$ take the Fourier transform of $\mathrm{x}_c(t)$.

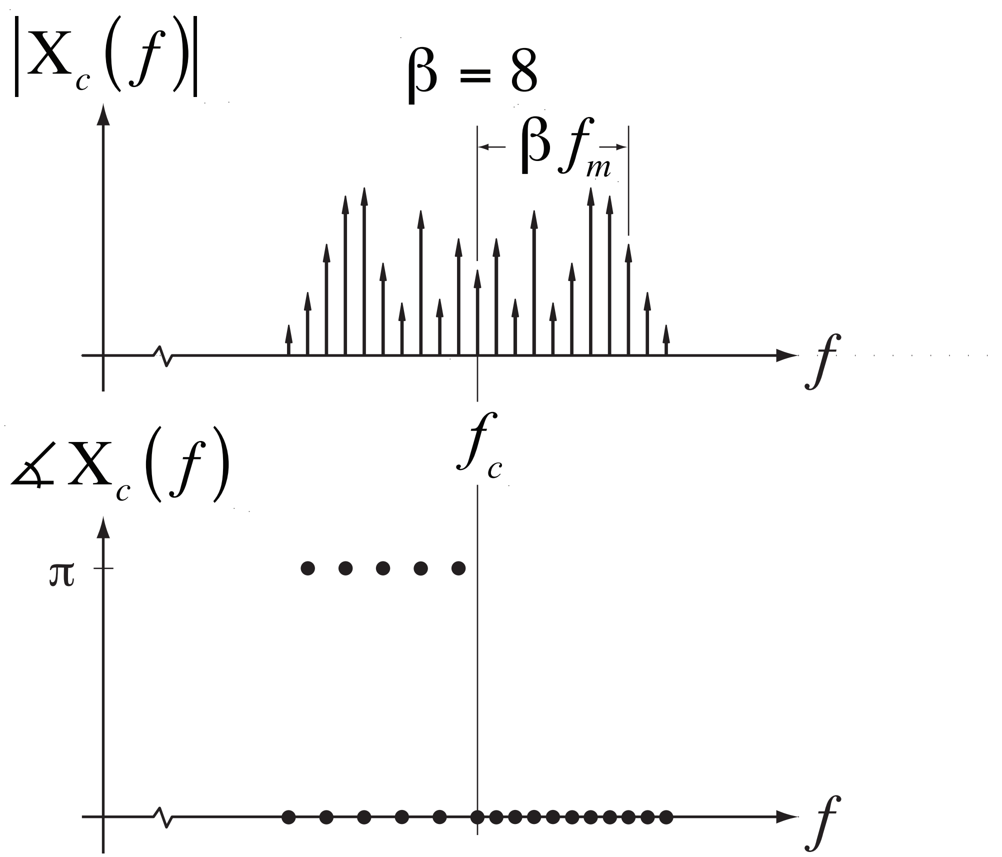
$$\mathrm{X}_c(f) = (A_c/2)\sum_{k=-\infty}^{\infty} \mathrm{J}_k(\beta)\left[\delta\left(f-(f_c+kf_m)\right)+\delta\left(f+(f_c+kf_m)\right)\right]$$

The impulses in the spectrum extend in frequency all the way to infinity. But beyond βf_m the impulse strengths die rapidly. For practical purposes the bandwidth is approximately $2\beta f_m$.

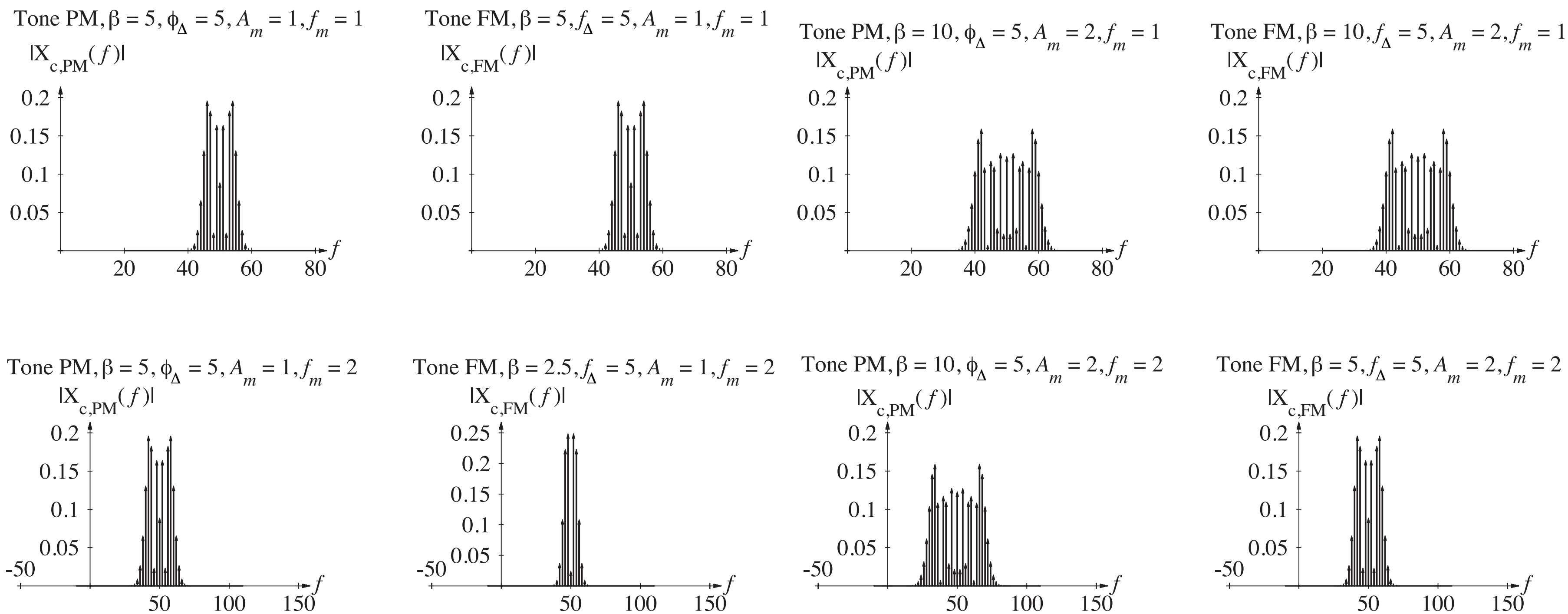
Phase and Frequency Modulation

Wideband FM Spectrum

for Cosine-Wave Modulation



Phase and Frequency Modulation



Transmission Bandwidth

The bandwidth required for transmitting an FM signal is theoretically infinite. That is, an infinite bandwidth would be required to transmit an FM signal *perfectly*, even if the modulating signal is bandlimited. Fortunately, in practical systems, perfection is not required and we can get by with a finite bandwidth. With tone modulation, the bandwidth required depends on the modulation index β. The spectral line magnitudes fall off rapidly at positive frequencies for which $|f - f_c| > \beta f_m$. So for tone modulation the bandwidth required for transmission would be approximately $2\beta f_m$. In the narrowband case when β is very small we cannot exactly follow this rule because we would have no modulation at all. So there is a "floor" of at least $2f_m$.

Transmission Bandwidth

For the general case, **Carson's rule** is a handy approximation that says $B \cong 2(D+1)W$, where

$$D = \frac{\text{peak frequency deviation}}{\text{bandwidth of m}(t)} = \frac{f_d}{W}\left|\mathrm{m}(t)\right|_{\max}$$

If $D << 1$, then $B \cong 2W$. This is the narrowband case.

If $D >> 1$, then $B \cong 2DW = f_d\left|\mathrm{m}(t)\right|_{\max}$. This is the wideband case.

Generation and Detection of FM and PM

The most direct and straightforward way of generating FM is to use a device known as a **voltage-to-frequency converter** (**VCO**). One way this can be done is by varying with time the capacitance in an LC parallel resonant oscillator. Let the capacitance be the capacitance of a varactor diode in parallel with another capacitor forming $\mathrm{C}(t) = C_0 - C\,\mathrm{x}(t)$. The time-varying LC resonant frequency is

$$\mathrm{f}(t) = \frac{1}{2\pi}\frac{d}{dt}\big(\theta(t)\big) \Rightarrow \frac{d}{dt}\big(\theta(t)\big) = \frac{1}{\sqrt{L\mathrm{C}(t)}} = \frac{1}{\sqrt{LC_0}}\frac{1}{\sqrt{1 - \dfrac{C}{C_0}\mathrm{x}(t)}}$$

We can use the formula (Abramowitz and Stegun, page 15),

$$(1+x)^{\alpha} = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \cdots$$

to write

$$\frac{d}{dt}\big(\theta(t)\big) = \frac{1}{\sqrt{LC_0}}\left[1 - \frac{C}{C_0}\mathrm{x}(t)\right]^{-1/2} = \frac{1}{\sqrt{LC_0}}\left[1 + \frac{1}{2}\frac{C}{C_0}\mathrm{x}(t) + \frac{3}{8}\left(\frac{C}{C_0}\mathrm{x}(t)\right)^2 + \cdots\right]$$

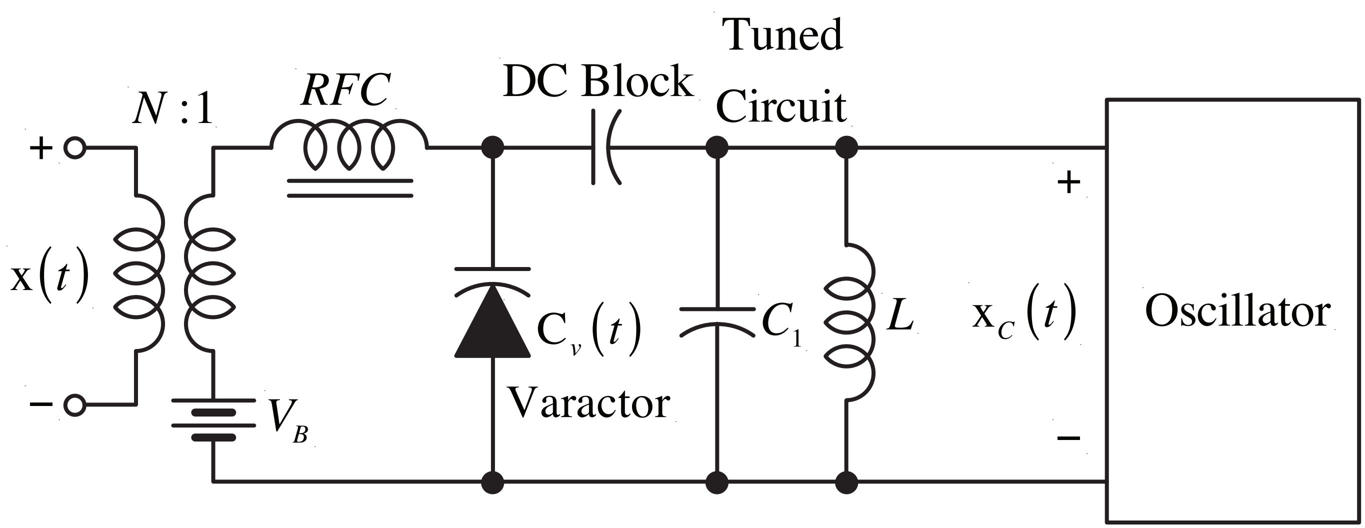
Generation and Detection of FM and PM

If $C\,\mathrm{x}(t)$ is "small enough", then $\dfrac{d}{dt}(\theta(t)) \cong \dfrac{1}{\sqrt{LC_0}}\left[1+\dfrac{1}{2}\dfrac{C}{C_0}\mathrm{x}(t)\right]$ and

$\theta(t) = 2\pi f_c t + 2\pi \dfrac{C}{2C_0} f_c \int^t \mathrm{x}(\lambda)\,d\lambda$. This is in the form of FM with $f_d = \dfrac{C}{2C_0} f_c$.

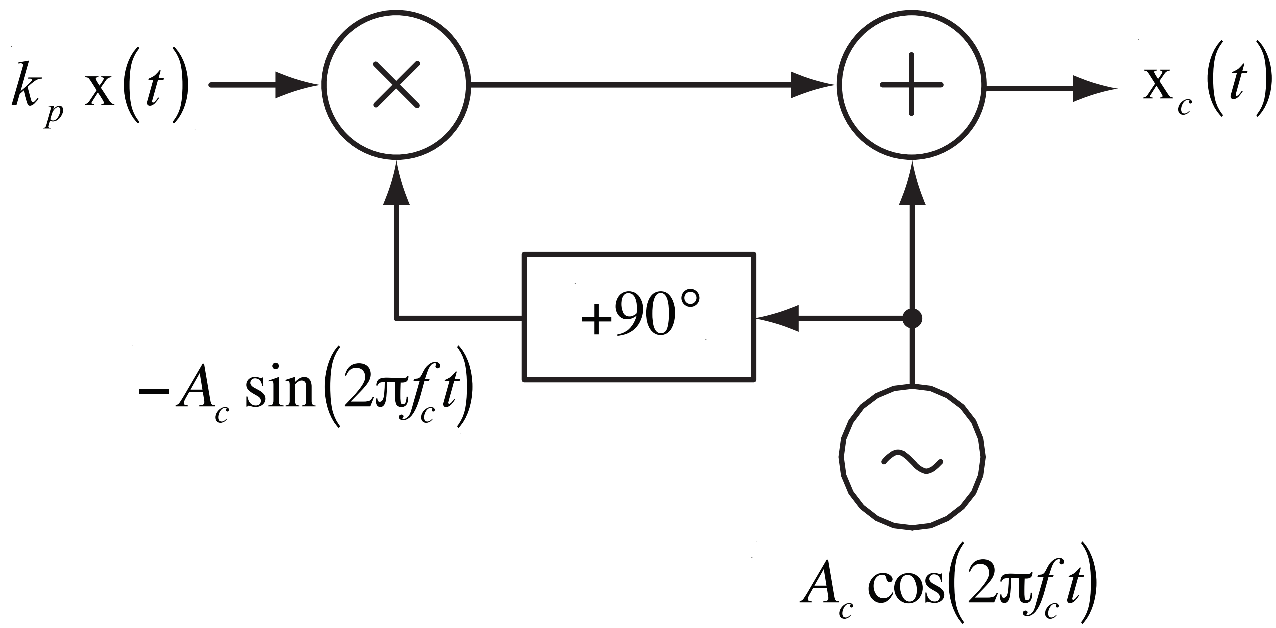
Since $|\mathrm{x}(t)| \le 1$, the approximation is good to within one percent if $C / C_0 < 0.013$.

So, taking that as an upper limit, $f_d = \dfrac{C}{2C_0} f_c \le 0.006 f_c$. This is a practical result that usually causes no design problems.



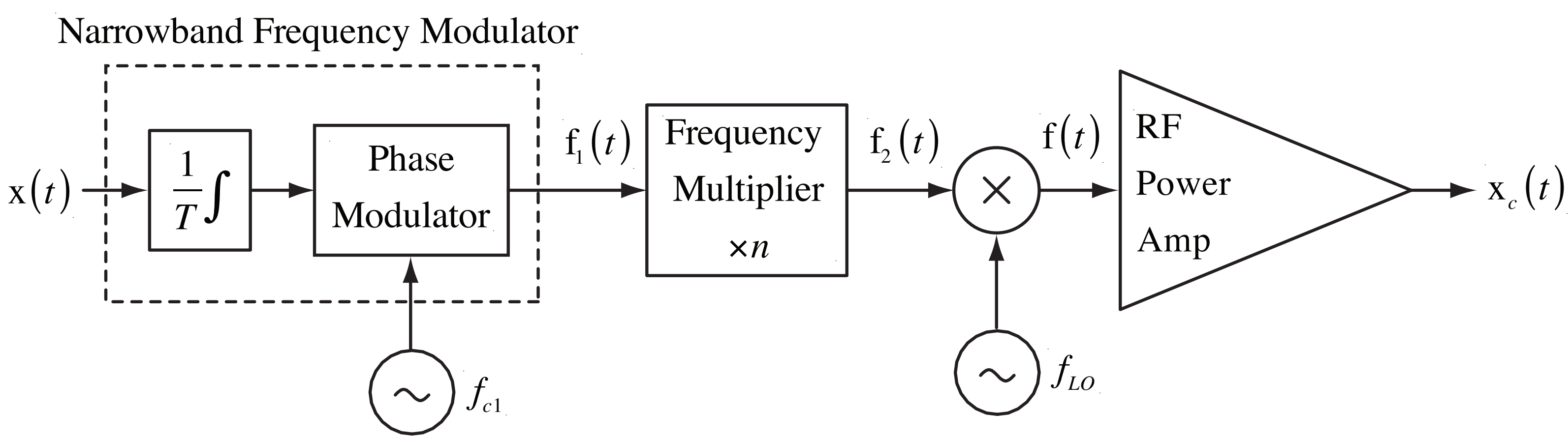
Generation and Detection of FM and PM

Another method for generating FM is to use a phase modulator, which produces PM, but integrate the message before applying it to the phase modulator. A narrowband phase modulator can be made by simulating the narrowband approximation $\mathrm{x}_c(t) = A_c\cos(2\pi f_c t) - A_c k_p\,\mathrm{x}(t)\sin(2\pi f_c t)$.



Generation and Detection of FM and PM

A third method for generating FM is called **indirect FM**. First, integrate the message $\mathrm{x}(t)$. Then use the integral of the message $\frac{1}{T}\int^{t}\mathrm{x}(\lambda)d\lambda$ as the input signal to a narrowband phase modulator with a carrier frequency f_{c1}. This produces a signal with instantaneous frequency $\mathrm{f}_1(t) = f_{c1} + \frac{k_p}{2\pi T}\mathrm{x}(t)$.

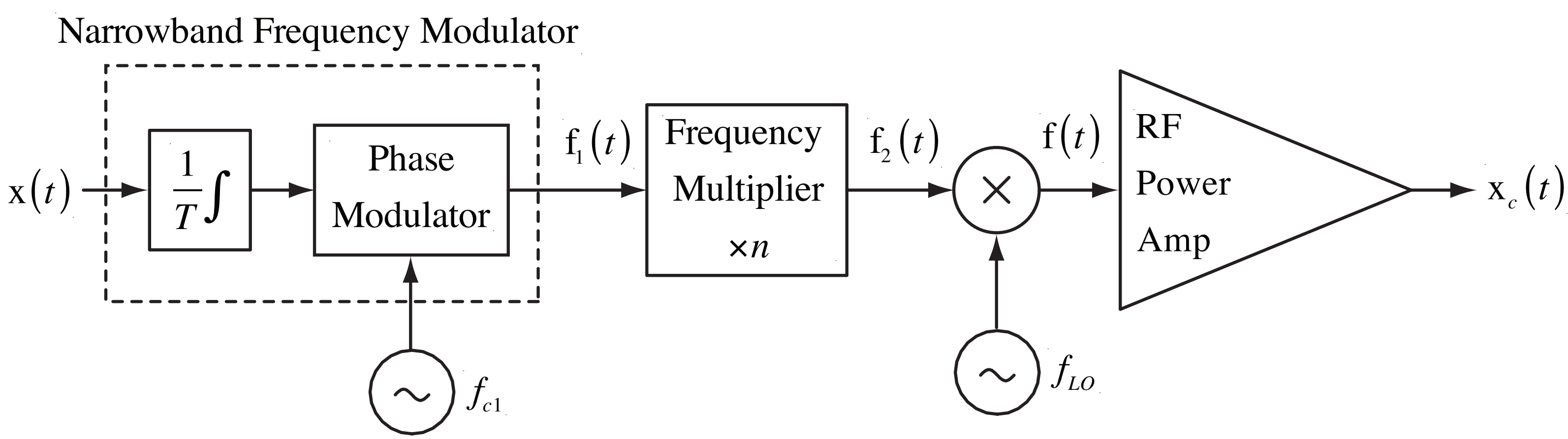


Generation and Detection of FM and PM

Next frequency-multiply the narrowband FM signal by a factor of n. This moves the carrier frequency to nf_{c1}, creating a signal with instantaneous frequency

$\mathrm{f}_2(t) = nf_{c1} + n\dfrac{k_p}{2\pi T}\mathrm{x}(t)$. The effective value of the frequency deviation is now

$f_d = n\dfrac{k_p}{2\pi T}$. This changes the range of frequency variation but not the rate of frequency variation. Then, if needed, shift the entire FM spectrum to whatever carrier frequency is required and amplify for transmission.



Generation and Detection of FM and PM

There are four common methods of detecting FM:

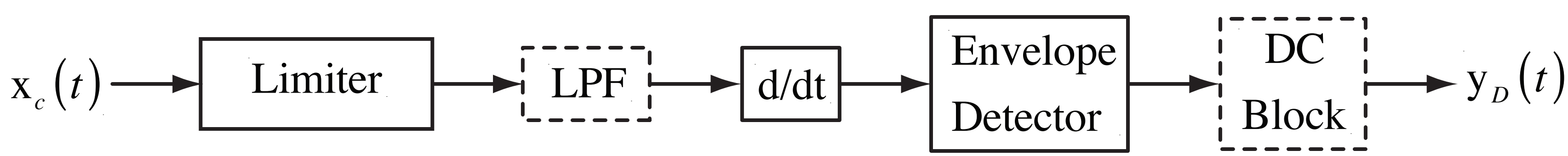
1. FM-to-AM Conversion Followed by Envelope Detection
2. Phase-Shift Discrimination
3. Zero-Crossing Detection
4. Frequency Feedback

FM-to-AM conversion can be done by time-differentiating the modulated signal.

Let $\mathrm{x}_c(t) = A_c \cos(\theta_c(t))$ with $\dot{\theta}(t) = 2\pi[f_c + f_d \mathrm{x}(t)]$. Then

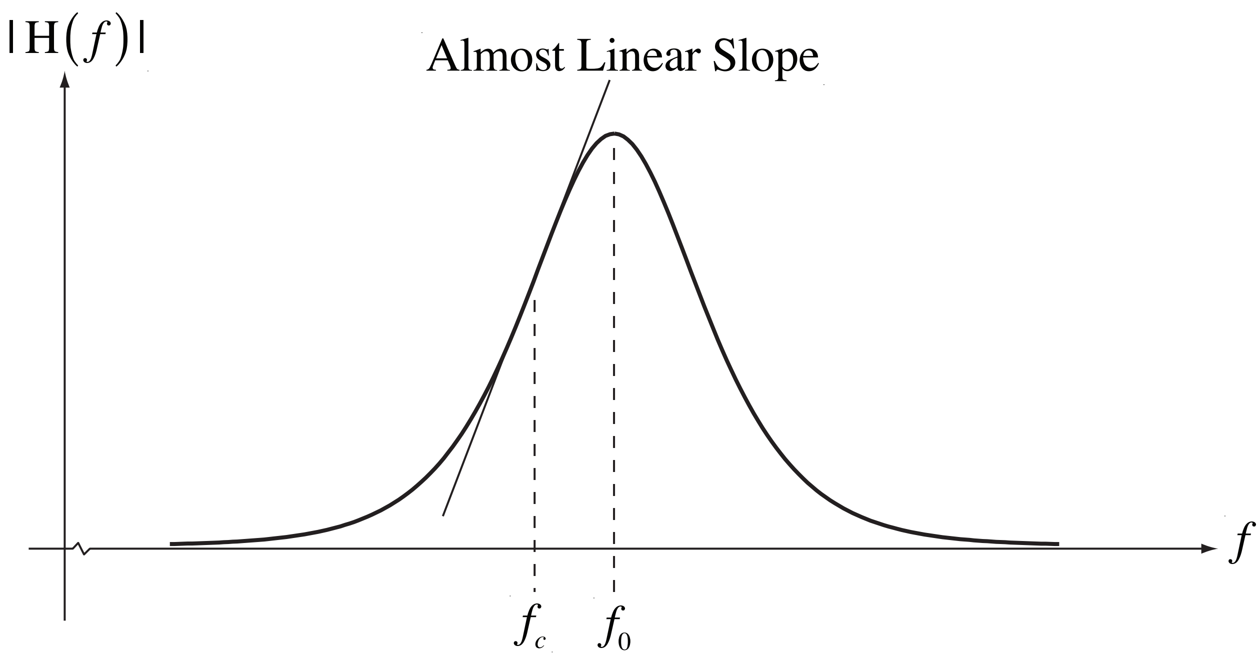
$\dot{\mathrm{x}}_c(t) = -A_c\dot{\theta}(t)\sin(\theta_c(t)) = 2\pi A_c[f_c + f_d \mathrm{x}(t)]\sin(\theta_c(t) \pm 180°)$.

The message can then be recovered by an envelope detector.



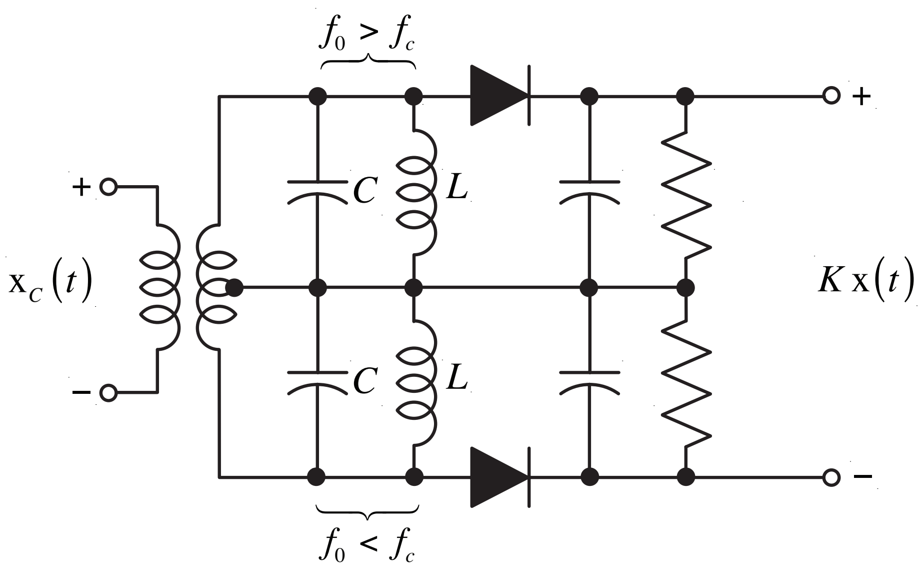
Generation and Detection of FM and PM

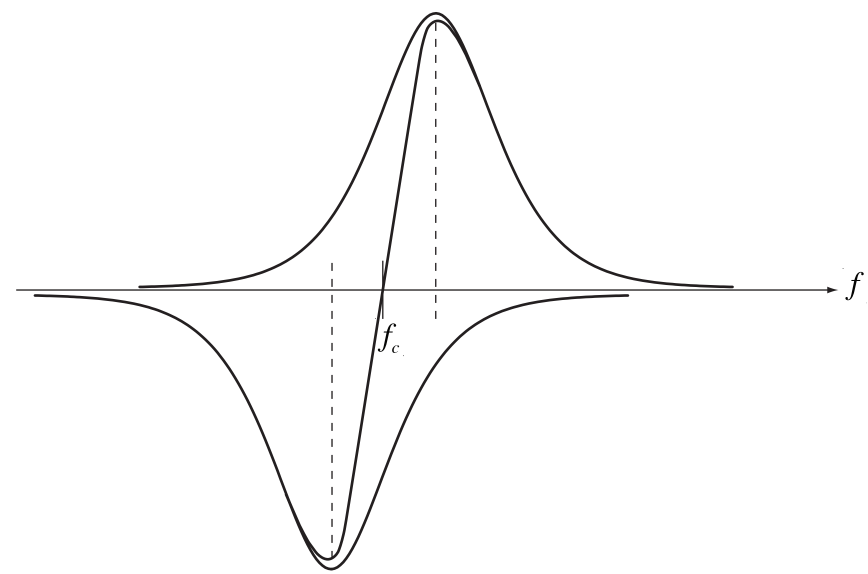
The "differentiator" in FM-to-AM detection need not be a true differentiator. All that is really needed is a frequency response magnitude that has a linear (or almost linear) slope over the bandwidth of the FM signal. Just below and just above resonance a tuned circuit resonator has an almost linear magnitude dependence on frequency. This type of detection is commonly called **slope detection**.



Generation and Detection of FM and PM

The linearity of slope detection can be improved by using two resonant circuits instead of only one. This type of circuit is called a **balanced discriminator**.





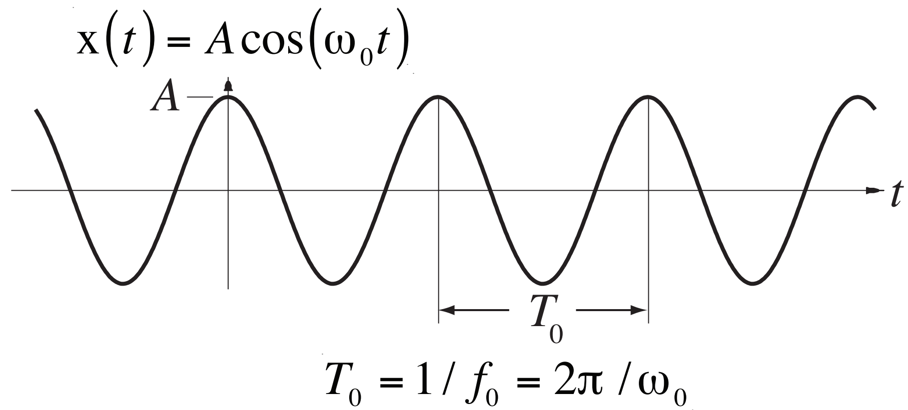
Phase and Frequency

Consider a cosine of the form $\mathrm{x}(t) = A\cos(2\pi f_0 t + \phi(t))$. The **phase** of this cosine is $\theta(t) = 2\pi f_0 t + \phi(t)$ and $\phi(t)$ is its **phase shift**.

First consider the case $\phi(t) = 0$.

Then $\mathrm{x}(t) = A\cos(2\pi f_0 t)$

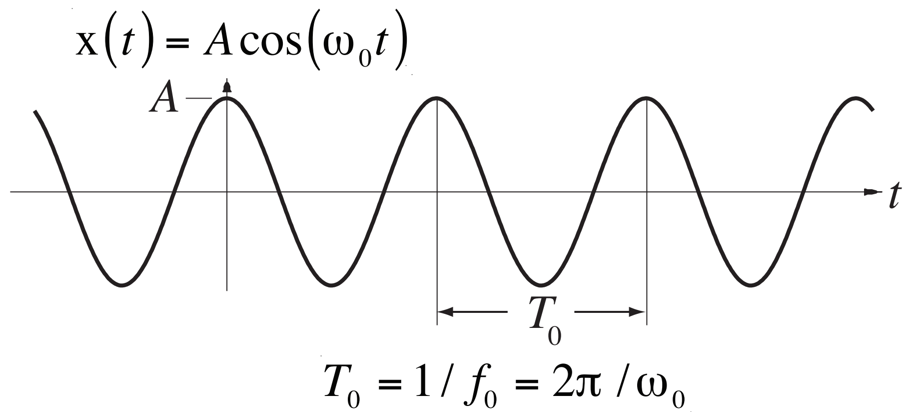
and $\theta(t) = 2\pi f_0 t$.

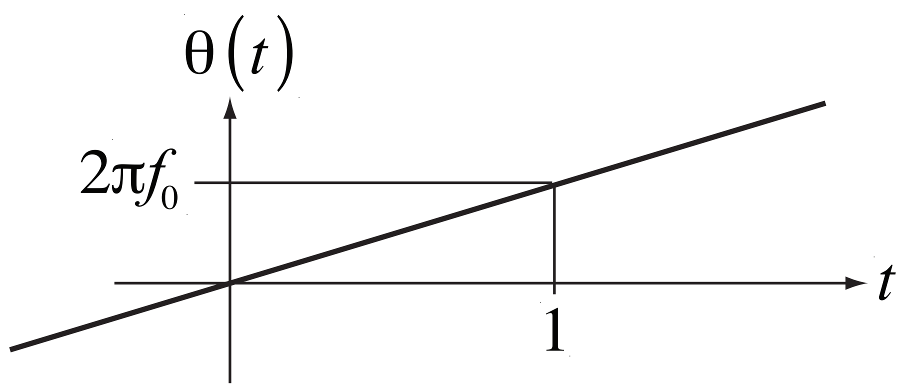


The cyclic frequency of this cosine is f_0. Also, the first time derivative of $\theta(t)$ is $2\pi f_0$. So one way of defining cyclic frequency is as the first derivative of phase, divided by 2π. It then follows that phase is the integral of frequency.

Phase and Frequency

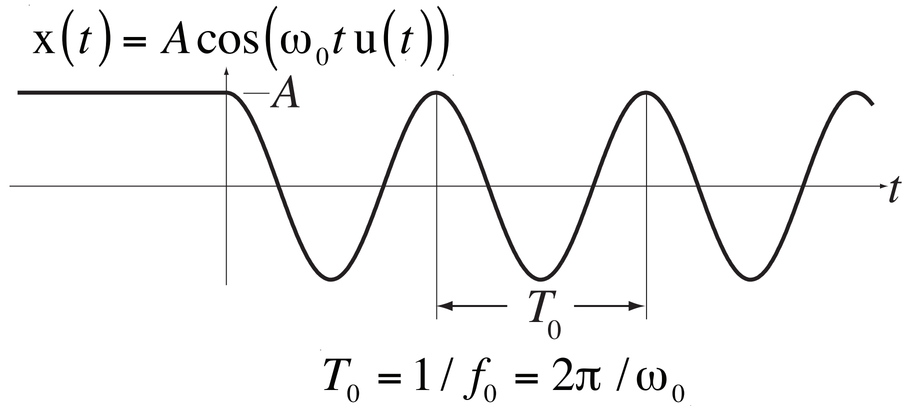
If $x(t) = A\cos(2\pi f_0 t)$ and $\theta(t) = 2\pi f_0 t$. Then a graph of phase versus time would be a straight line through the origin with slope $2\pi f_0$.

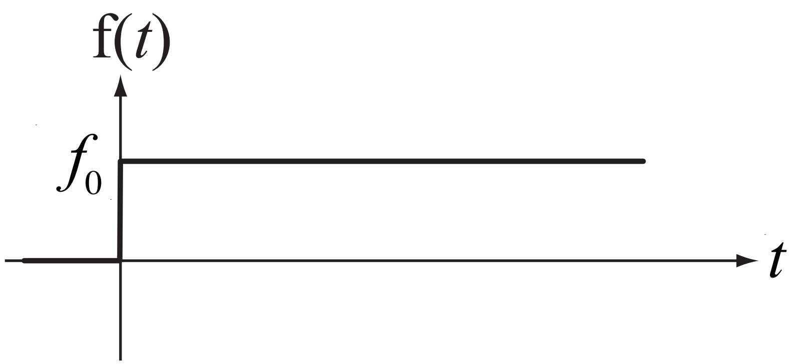


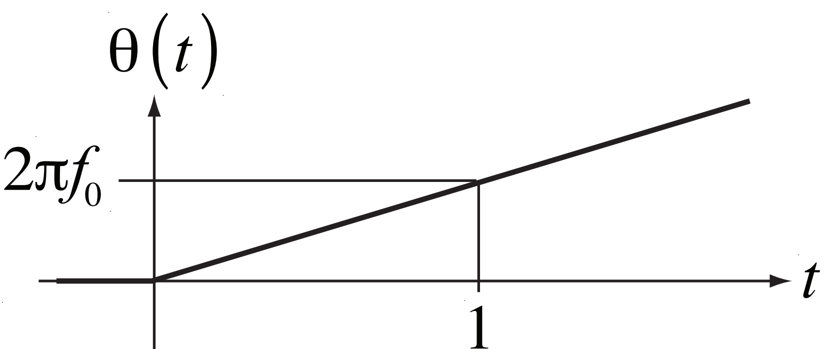


Phase and Frequency

Let $\mathrm{x}(t) = A\cos(\theta(t))$ and let $\theta(t) = 2\pi f_0 t\,\mathrm{u}(t) = 2\pi f_0\,\mathrm{ramp}(t)$. Then the cyclic frequency is $\dfrac{1}{2\pi}\dfrac{d}{dt}\theta(t) = f_0\,\mathrm{u}(t)$. Call its **instantaneous cyclic frequency** $\mathrm{f}(t)$.

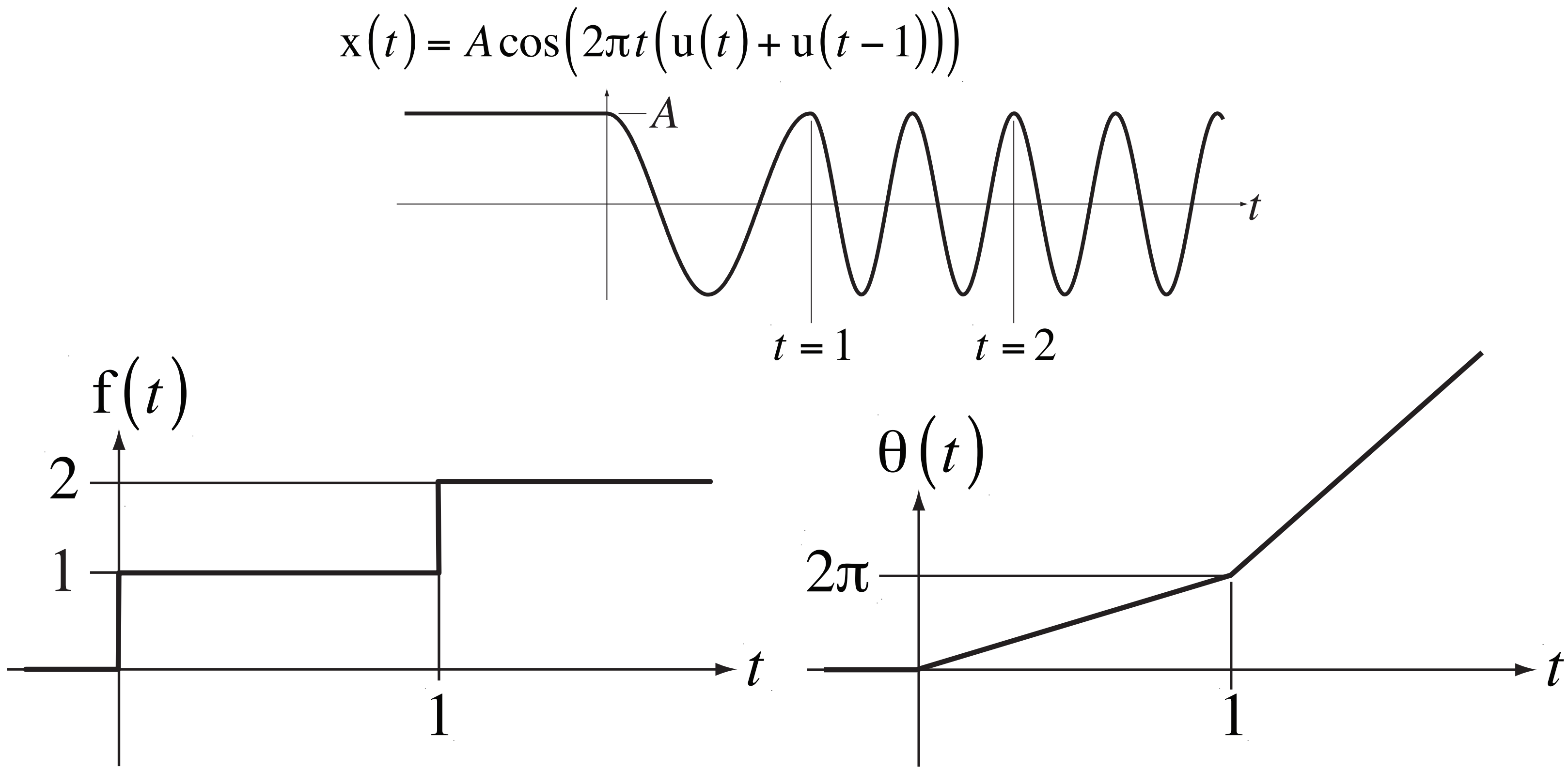






Phase and Frequency

Now let $\mathrm{x}(t) = A\cos\left(2\pi t\left(\mathrm{u}(t)+\mathrm{u}(t-1)\right)\right)$. Then the instantaneous cyclic frequency is $\mathrm{f}(t) = \mathrm{u}(t)+\mathrm{u}(t-1)$ and the phase is $\theta(t) = 2\pi\left(\mathrm{ramp}(t)+\mathrm{ramp}(t-1)\right)$.



Phase Discrimination

Let $\mathrm{x}_1(t) = A_1 \sin(2\pi f_0 t + \theta(t))$ and let $\mathrm{x}_2(t) = A_2 \cos(2\pi f_0 t + \phi(t))$.

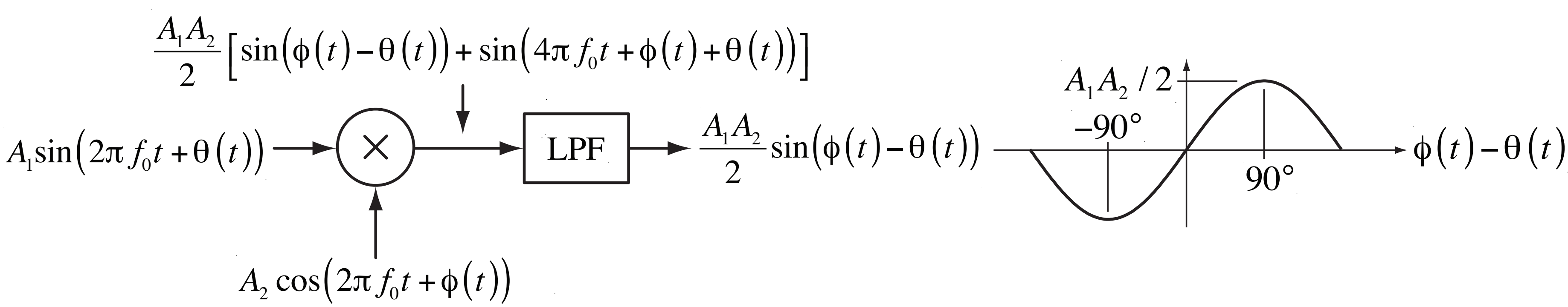
The product is $\mathrm{x}_1(t)\mathrm{x}_2(t) = A_1 A_2 \sin(2\pi f_0 t + \theta(t))\cos(2\pi f_0 t + \phi(t))$.

Using a trigonometric identity,

$$\mathrm{x}_1(t)\mathrm{x}_2(t) = \frac{A_1 A_2}{2}\left[\sin(\phi(t) - \theta(t)) + \sin(4\pi f_0 t + \phi(t) + \theta(t))\right]$$

and

$$\left\langle \mathrm{x}_1(t)\mathrm{x}_2(t) \right\rangle = \frac{A_1 A_2}{2}\sin(\phi(t) - \theta(t))$$

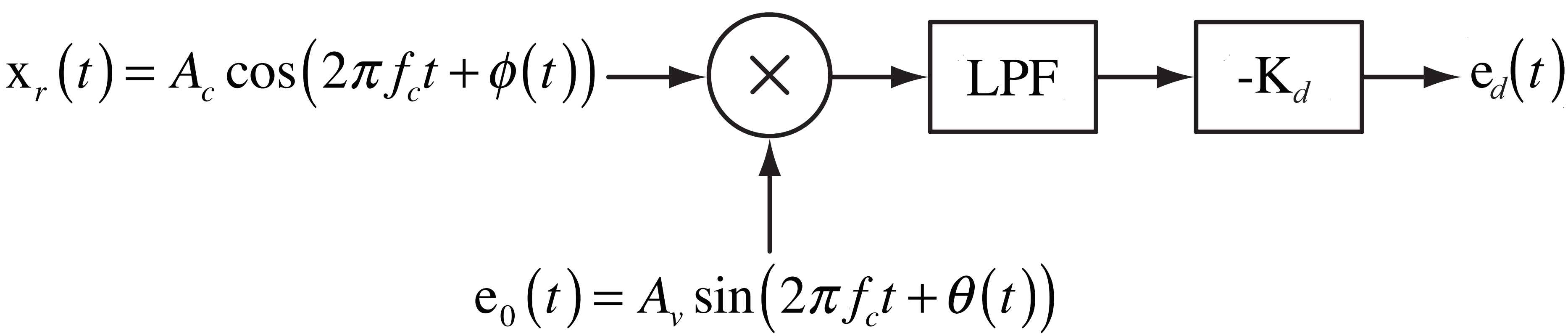


Voltage-Controlled Oscillators

A **voltage - controlled oscillator** (**VCO**) is a device that accepts an analog voltage as its input and produces a periodic waveform whose fundamental frequency depends on that voltage. Another common name for a VCO is "voltage-to-frequency converter". The waveform is typically either a sinusoid or a rectangular wave. A VCO has a free-running frequency f_{v}. When the input analog voltage is zero, the fundamental frequency of the VCO output signal is f_{v}. The output frequency of the VCO is $f_{VCO} = f_{\mathrm{v}} + K_{\mathrm{v}}\, \mathrm{v}_{in}$ where K_{v} is a gain constant with units of Hz/V.

Phase-Locked Loops

A **phase - locked loop** (**PLL**) is a device used to generate a signal with a fixed phase relationship to the carrier in a bandpass signal. An essential ingredient in the locking process is an **analog phase comparator**. A phase comparator produces a signal that depends on the phase difference between two bandpass signals. One system that accomplishes this goal is an analog mulitplier followed by a lowpass filter. Let the two bandpass signals be $\mathrm{x}_r(t) = A_c \cos\left(2\pi f_c t + \phi(t)\right)$ and $\mathrm{e}_0(t) = A_v \sin\left(2\pi f_c t + \theta(t)\right)$ and let the output signal from the phase comparator be $\mathrm{e}_d(t)$.

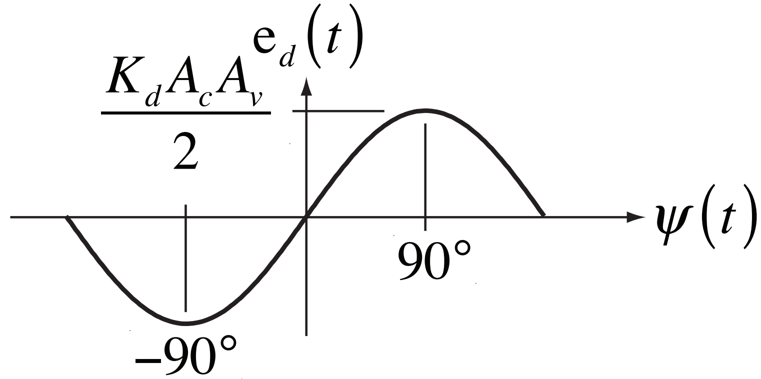


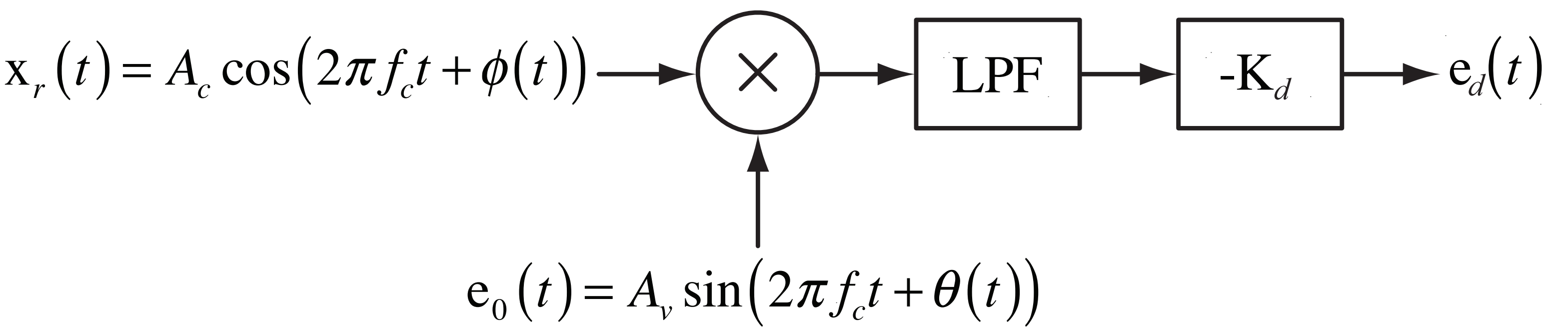
Phase-Locked Loops

$$\mathrm{x}_r(t)\mathrm{e}_0(t) = -K_d A_c A_v \cos\left(2\pi f_c t + \phi(t)\right)\sin\left(2\pi f_c t + \theta(t)\right)$$

$$\mathrm{x}_r(t)\mathrm{e}_0(t) = -\frac{K_d A_c A_v}{2}\left[\sin\left(\theta(t) - \phi(t)\right) + \sin\left(4\pi f_c t + \phi(t) + \theta(t)\right)\right]$$

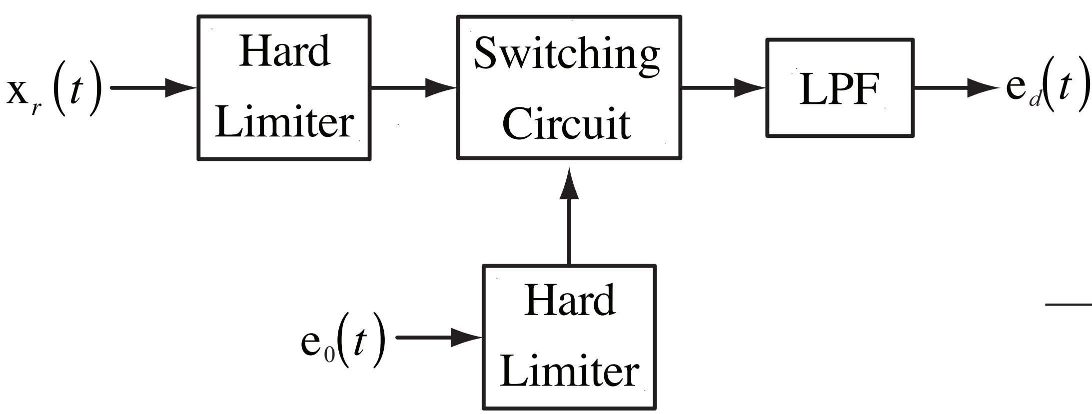
$$\mathrm{e}_d(t) = \frac{K_d A_c A_v}{2}\sin\left(\phi(t) - \theta(t)\right) = \frac{K_d A_c A_v}{2}\sin\left(\psi(t)\right)$$

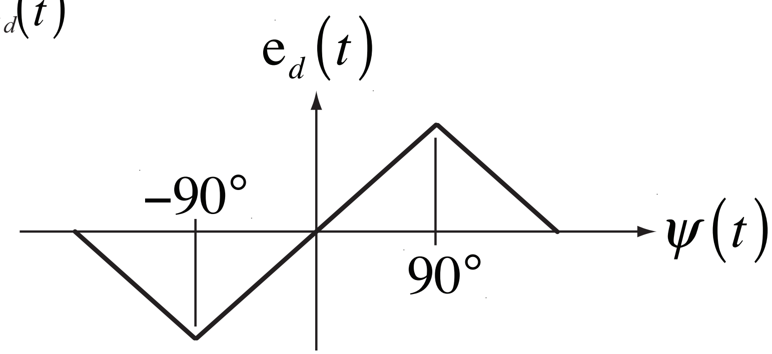




Phase-Locked Loops

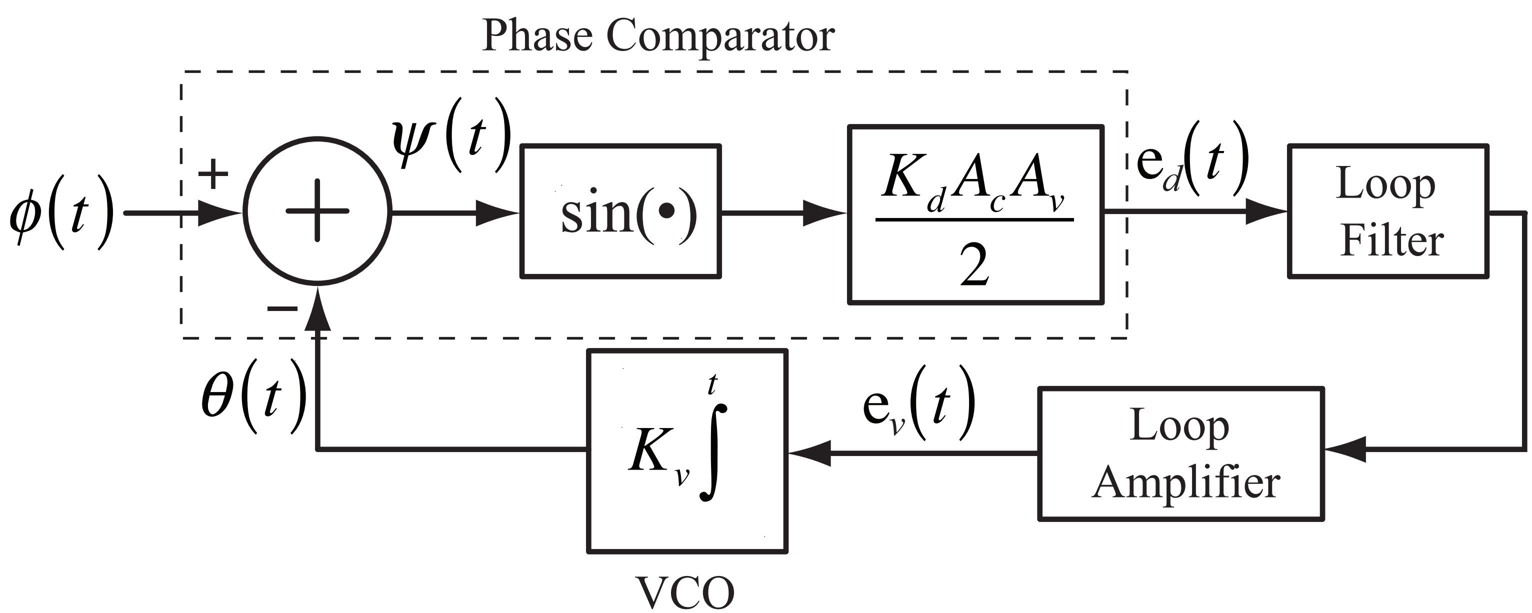
$e_d(t)$ depends on both the phase difference and A_c and A_v. We can make the dependence on these amplitudes go away if we first **hard limit** the signals, turning them into fixed-amplitude square waves. Another benefit of hard-limiting is that the multiplication becomes a switching operation and the error signal $e_d(t)$ is now a linear function of $\psi(t)$ over a wider range.





Phase-Locked Loops

From the block diagram of the phase-locked loop below it is clear that $\mathrm{E}_v(s) = \mathrm{F}(s)\mathrm{E}_d(s)$, where $\mathrm{F}(s)$ is the transfer function of the loop-filter-loop-amplifier combination. The VCO converts voltage to frequency and phase is the integral of frequency. That is why the VCO is represented as an integrator with voltage in and phase out.

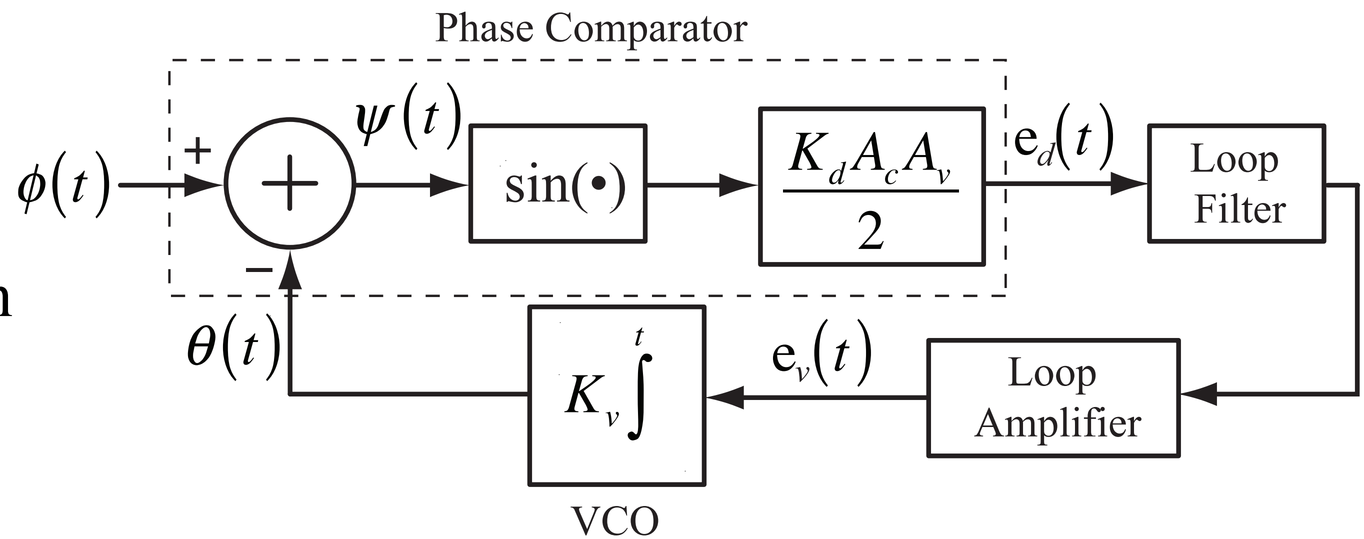


Phase-Locked Loops

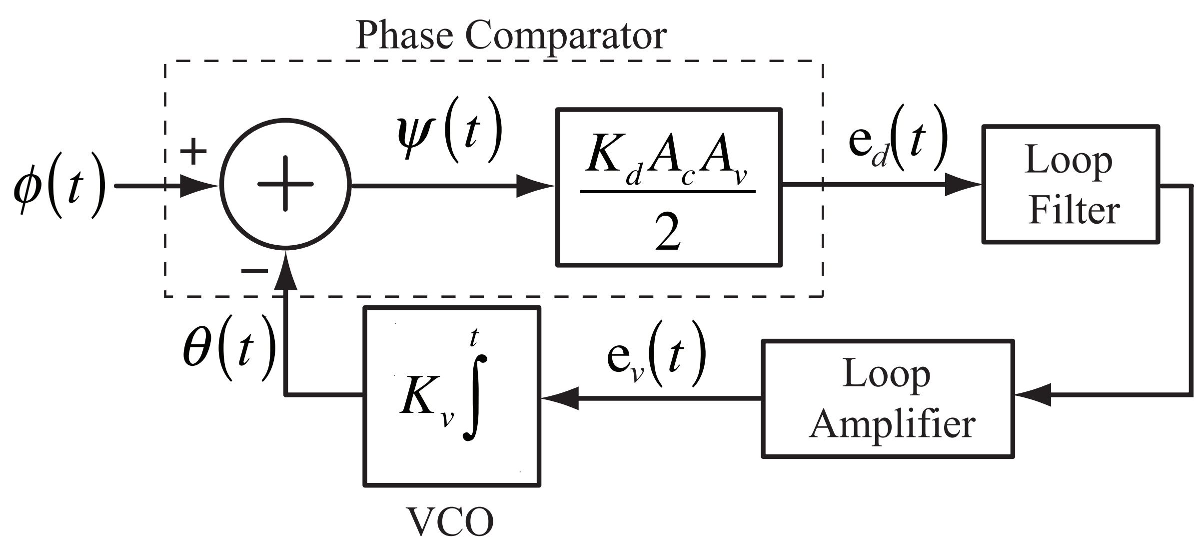
Phase-locked loops operate in two modes, **acquisition** and **tracking**. When a PLL is turned on it must first acquire a phase lock and thereafter it must track the phase changes in the incoming signal. The acquisition of a phase lock must be described by the non-linear model of the PLL in which the phase discriminator has a sine transfer function. In the tracking mode, the phase error is typically small, the sine function can be approximated by its argument and the model of the PLL becomes linear.

Phase-Locked Loops

Non-linear PLL Model for Acquisition



Linear PLL Model for Tracking

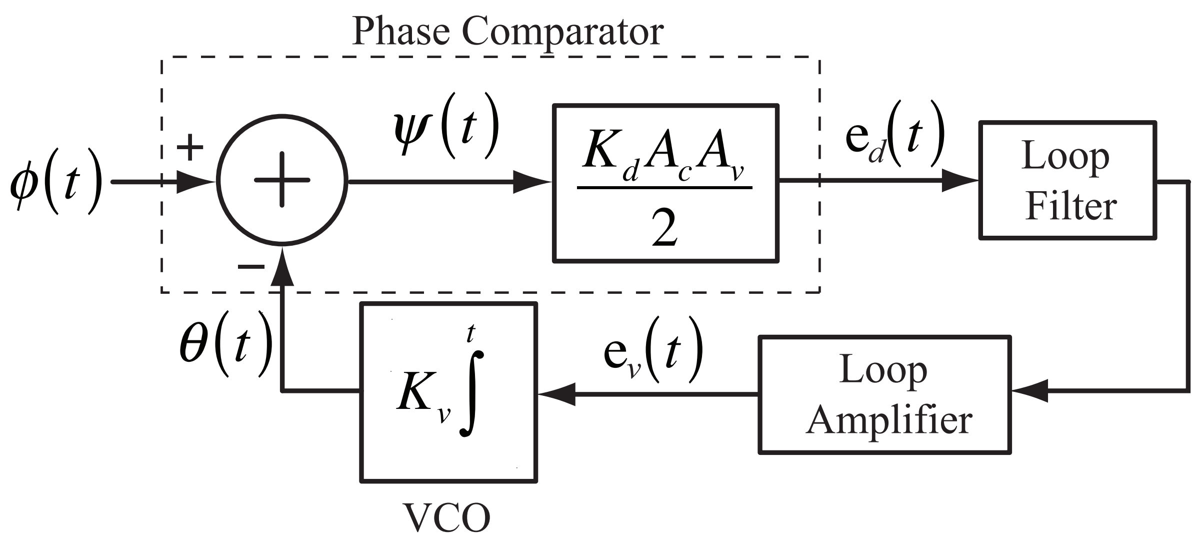


Phase-Locked Loops

In the tracking mode $\Theta(s) = \frac{K_d A_c A_v}{2}\left[\Phi(s) - \Theta(s)\right]\mathrm{F}(s)\frac{K_v}{s}$.

It follows that $\mathrm{H}(s) = \frac{\Theta(s)}{\Phi(s)} = \frac{K_t \mathrm{F}(s)}{s + K_t \mathrm{F}(s)}$ where $K_t = \frac{K_d A_c A_v K_v}{2}$.

$\Psi(s) = \Phi(s) - \Theta(s)$, therefore $\mathrm{G}(s) = \frac{\Psi(s)}{\Phi(s)} = \frac{\Phi(s) - \Theta(s)}{\Phi(s)} = 1 - \mathrm{H}(s)$.



Phase-Locked Loops

Let the phase deviation of the incoming signal $\phi(t)$ be of the general form $\phi(t) = \left[\pi R t^2 + 2\pi f_\Delta(t) + \theta_0\right]\mathrm{u}(t)$. Then $\dfrac{1}{2\pi}\dfrac{d\phi}{dt} = (Rt + f_\Delta)\mathrm{u}(t)$, a frequency ramp plus a frequency step. Then $\Phi(s) = \dfrac{2\pi R}{s^3} + \dfrac{2\pi f_\Delta}{s^2} + \dfrac{\theta_0}{s}$. Using the final value theorem of the Laplace transform, the steady state phase error between the incoming signal and the VCO output signal is $\lim\limits_{t\to\infty}\psi(t) = \lim\limits_{s\to 0} s\left[\dfrac{2\pi R}{s^3} + \dfrac{2\pi f_\Delta}{s^2} + \dfrac{\theta_0}{s}\right]\mathrm{G}(s)$. Now let $\mathrm{F}(s) = \dfrac{s^2 + as + b}{s^2}$.

Then $\mathrm{H}(s) = \dfrac{K_t\left(s^2 + as + b\right)}{s^3 + K_t\left(s^2 + as + b\right)}$ and $\mathrm{G}(s) = \dfrac{s^3}{s^3 + K_t\left(s^2 + as + b\right)}$.

Phase-Locked Loops

Then the steady-state phase error is $\lim\limits_{t\to\infty}\psi(t)=\lim\limits_{s\to 0} s\dfrac{\theta_0 s^2+2\pi f_\Delta s+2\pi R}{s^3+K_t\left(s^2+as+b\right)}$.

Its value depends on the form of the input signal's phase deviation and the order of the loop filter.

Steady State Error, $\lim\limits_{t\to\infty}\psi(t)$

PLL Order $\downarrow$	$\theta_0 \neq 0$, $f_\Delta = 0$, $R = 0$	$\theta_0 \neq 0$, $f_\Delta \neq 0$, $R = 0$	$\theta_0 \neq 0$, $f_\Delta \neq 0$, $R \neq 0$
$1(a=0,b=0)$	0	$\dfrac{2\pi f_\Delta}{K_t}$	∞
$2(a\neq 0,b=0)$	0	0	$\dfrac{2\pi R}{t}$
$3(a\neq 0,b\neq 0)$	0	0	0

Phase-Locked Loops

So a first-order PLL can track a phase step with zero error and a frequency step with a finite error. A second-order PLL can track a frequency step with zero error and a frequency ramp with a finite error. A third-order PLL can track a frequency step and a frequency ramp with zero error. When the error is finite, its size can be made arbitrarily small by making K_t large. However, this also increases the loop bandwidth, making the signal-to-noise ratio worse.
(More in Chapter 8.)

Phase-Locked Loops

A first-order PLL can be used for demodulation of angle-modulated signals but a second-order PLL has some advantages and is more common in practice. Therefore, in $\mathrm{F}(s) = \dfrac{s^2 + as + b}{s^2}$, make $b = 0$.

Then $\mathrm{F}(s) = 1 + \dfrac{a}{s}$. This can be implemented as the signal plus a times the integral of the signal. With this $\mathrm{F}(s)$, $\mathrm{H}(s) = \dfrac{\Theta(s)}{\Phi(s)} = \dfrac{K_t(s+a)}{s^2 + K_t(s+a)}$ and

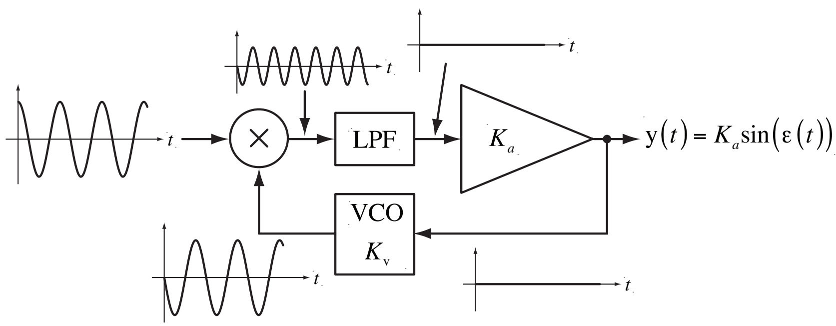
$\mathrm{G}(s) = \dfrac{\Psi(s)}{\Phi(s)} = \dfrac{s^2}{s^2 + K_t(s+a)}$, a second-order transfer function. Expressing this transfer function in a standard second-order system form,

$\mathrm{G}(s) = \dfrac{s^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$, where $\zeta = \dfrac{1}{2}\sqrt{\dfrac{K_t}{a}}$ is the damping factor and $\omega_0 = \sqrt{K_t a}$ is the radian natural frequency.

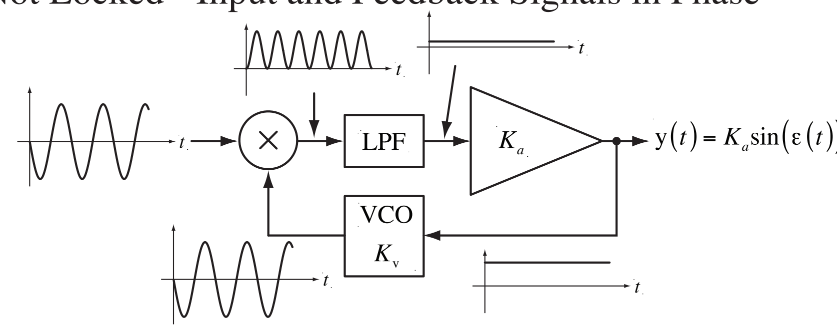
Phase-Locked Loop States

Input and Feedback Signals at Same Frequency

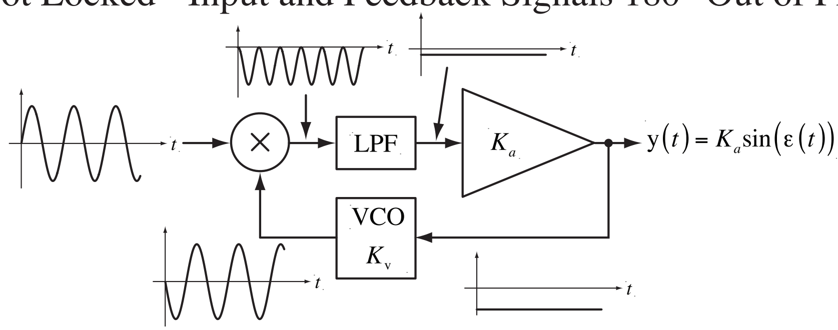
Locked in Quadrature



Not Locked - Input and Feedback Signals in Phase



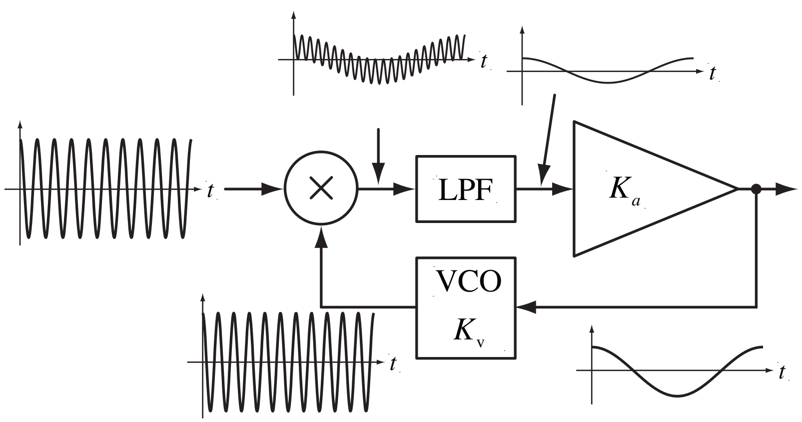
Not Locked - Input and Feedback Signals 180° Out of Phase



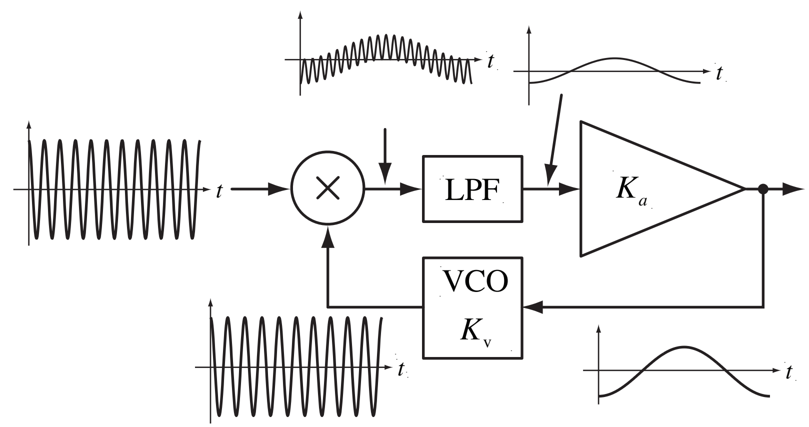
Phase-Locked Loop States

Input and Feedback Signals at Different Frequencies

Input Frequency > Feedback Frequency

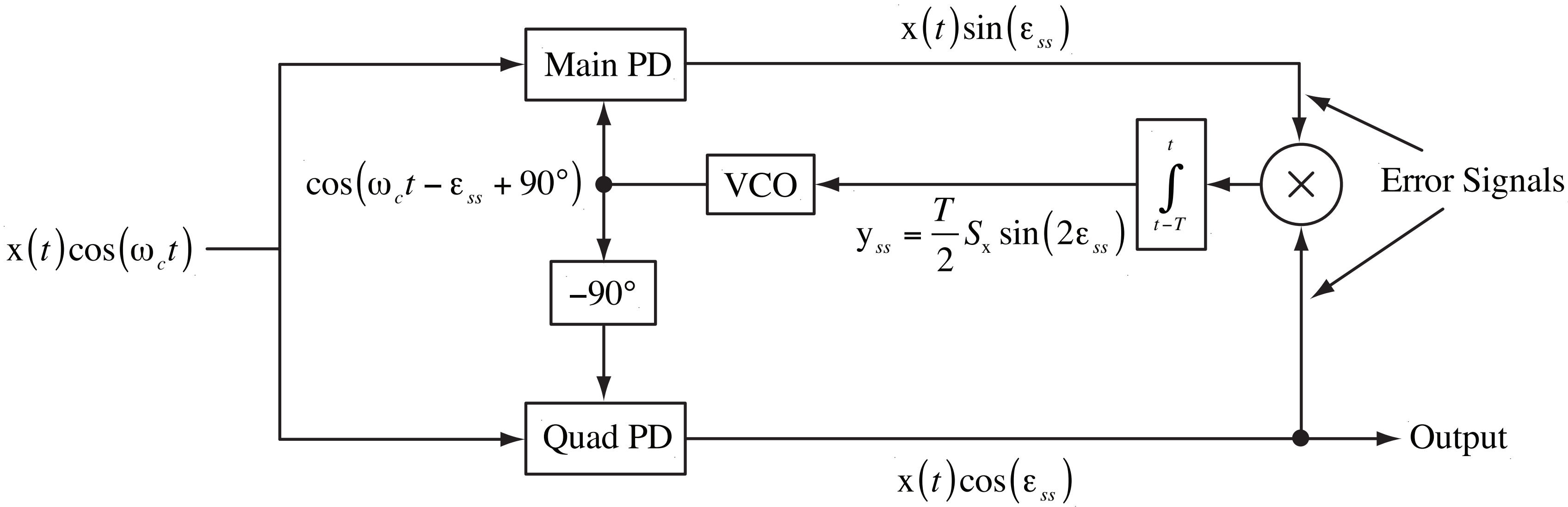


Input Frequency < Feedback Frequency



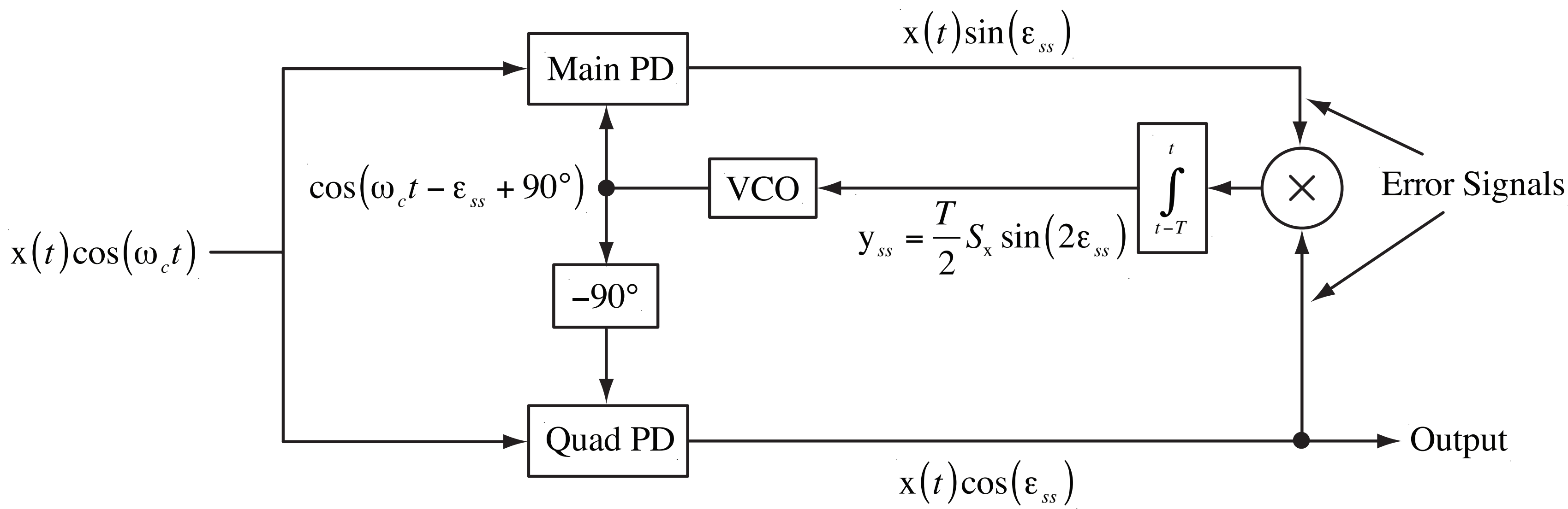
Phase-Locked Loops

For DSB signals, which do not have transmitted carriers, Costas invented a system to synchronize a local oscillator and also do synchronous detection. The incoming signal is $x_c(t) = x(t)\cos(\omega_c t)$ with bandwidth $2W$. It is applied to two phase discriminators, main and quad, each consisting of a multiplier followed by a LPF and an amplifier. The local oscillators that drive them are 90° out of phase so that the output signal from the main phase discriminator is $x(t)\sin(\varepsilon_{ss})$ and the output signal from the quad phase discriminator is $x(t)\cos(\varepsilon_{ss})$.



Phase-Locked Loops

The VCO control voltage y_{ss} is the time average of the product of $x(t)\sin(\varepsilon_{ss})$ and $x(t)\cos(\varepsilon_{ss})$ or $y_{ss} = \int_{t-T}^{t} x^2(\lambda)\cos(\varepsilon_{ss})\sin(\varepsilon_{ss})\,d\lambda$ which is $y_{ss} = \frac{T}{2}\left\langle x^2(t)\left[\sin(0)+\sin(2\varepsilon_{ss})\right]\right\rangle = \frac{T}{2}S_x\sin(2\varepsilon_{ss})$. When the angular error ε_{ss} is zero, y_{ss} does not change with time, the loop is locked and the output signal from the quad phase discriminator is $x(t)\cos(\varepsilon_{ss}) = x(t)$ because $\varepsilon_{ss} = 0$.



Interference

Let the total received signal at a receiver be

$$\mathrm{v}(t) = A_c \cos(\omega_c t) + A_i \cos((\omega_c + \omega_i)t + \phi_i)$$

where the first term represents the desired signal and the second term represents interference. Also define $\rho \triangleq A_i / A_c$ and $\theta_i(t) \triangleq \omega_i t + \phi_i$. Then

$$\mathrm{v}(t) = A_c\left[\cos(\omega_c t) + \rho\cos(\omega_c t + \theta_i)\right] = A_c\left\{\cos(\omega_c t) + \rho\begin{bmatrix}\cos(\omega_c t)\cos(\theta_i(t)) \\ -\sin(\omega_c t)\sin(\theta_i(t))\end{bmatrix}\right\}$$

$$\mathrm{v}(t) = A_c\left\{\left[1 + \rho\cos(\theta_i(t))\right]\cos(\omega_c t) - \rho\sin(\theta_i(t))\sin(\omega_c t)\right\}$$

The in-phase component is $A_c\left[1 + \rho\cos(\theta_i(t))\right]\cos(\omega_c t)$ and the quadrature component is $-A_c\rho\sin(\theta_i(t))\sin(\omega_c t)$. The envelope is

$\mathrm{A_v}(t) = A_c\sqrt{\left[1 + \rho\cos(\theta_i)\right]^2 + \rho^2\sin(\theta_i(t))} = A_c\sqrt{1 + \rho^2 + 2\rho\cos(\theta_i(t))}$. The phase relative to the desired signal is $\phi_{\mathrm{v}}(t) = \tan^{-1}\left(\dfrac{\rho\sin(\theta_i(t))}{1 + \rho\cos(\theta_i(t))}\right)$.

Interference

The envelope and phase of the total received signal

$$\mathrm{A_v}(t) = A_c\sqrt{1+\rho^2+2\rho\cos(\theta_i(t))} \text{ and } \phi_{\mathrm{v}}(t) = \tan^{-1}\left(\frac{\rho\sin(\theta_i(t))}{1+\rho\cos(\theta_i(t))}\right)$$

show that the effect of the interference on the received signal is to create both amplitude and phase modulation. If $\rho << 1$, then

$$\mathrm{A_v}(t) \cong A_c\sqrt{1+2\rho\cos(\theta_i(t))} \cong A_c\left[1+\rho\cos(\theta_i(t))\right] \text{ and } \phi_{\mathrm{v}}(t) \cong \tan^{-1}\left(\rho\sin(\theta_i(t))\right) \cong \rho\sin(\theta_i(t))$$

or

$$\mathrm{A_v}(t) \cong A_c\left[1+\rho\cos(\omega_i t+\phi_i)\right] \text{ and } \phi_{\mathrm{v}}(t) \cong \rho\sin(\omega_i t+\phi_i)$$

This result has the form of AM tone modulation with $\mu = \rho$ and simultaneous PM or FM tone modulation with $\beta = \rho$. If $\rho >> 1$, then

$$\mathrm{A_v}(t) = \rho A_c\sqrt{1+2\rho^{-1}\cos(\omega_i t+\phi_i)} \cong \rho A_c\left[1+\rho^{-1}\cos(\omega_i t+\phi_i)\right] \text{ and } \phi_{\mathrm{v}}(t) = \omega_i t+\phi_i$$

Interference

In the weak interference case

$$\mathrm{A_v}(t) \cong A_c\left[1+\rho\cos(\omega_i t+\phi_i)\right] \text{ and } \phi_{\mathrm{v}}(t) \cong \rho\sin(\omega_i t+\phi_i)$$

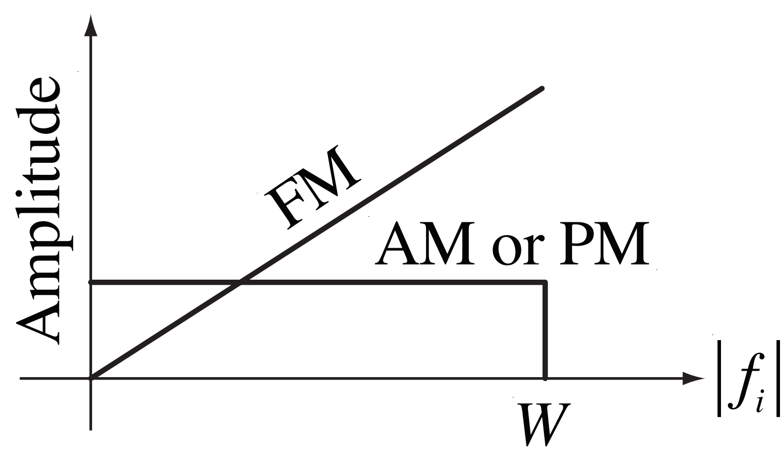
if we demodulate with an envelope, phase or frequency demodulator we get $(\text{with } \phi_i = 0)$

Envelope Detector: $K_D\left[1+\rho\cos(\omega_i t)\right]$

Phase Detector: $K_D\rho\sin(\omega_i t)$

Frequency Detector: $K_D\rho f_i\cos(\omega_i t)$

For AM or PM demodulation the demodulated signal strength is proportional to ρ. For FM demodulation the demodulated signal strength is porportional to the product of ρ and f_i.



Interference

The effects of interference on FM signals increases with frequency. So one way to reduce the effect is to lowpass filter the demodulated output. Of course this also lowpass filters the message, an undesirable outcome. To avoid the lowpass filtering effect on the message a technique called **preemphasis** is often used. The higher frequency parts of the message are preemphasized before transmission by passing them through a preemphasis filter with frequency response $\mathrm{H}_{pe}(f)$ that amplifies the higher frequencies more than the lower frequencies. Then, after transmission and frequency demodulation, the demodulated signal is passed through a **deemphasis** filter whose frequency response is $\mathrm{H}_{de}(f) = \dfrac{1}{\mathrm{H}_{pe}(f)}$.

Interference

A typical deemphasis filter has a frequency response $\mathrm{H}_{de}(f) = \dfrac{1}{1 + jf / B_{de}}$ in which B_{de} is less than the cutoff frequency of the normal sharp-cutoff lowpass filter that determines the bandwidth. That makes the corresponding preemphasis filter have a frequency response $\mathrm{H}_{pe}(f) = 1 + jf / B_{de}$.

Interference

A phenomenon that most people have experienced in receiving FM signals is the so-called **capture effect**. Suppose there are two FM stations, both transmitting in the same bandwidth and of approximately equal signal strength at the receiver. Their signal strengths will fluctuate some causing one to be stronger for a time and then the other. The stronger signal will "capture" the receiver for a short time and will dominate the demodulated signal. But then later the other signal will dominate and capture the receiver. The two stations switch back and forth and the listener hears a time-multiplexed version of both signals. To keep the math simple, assume we have one unmodulated carrier and one modulated carrier. This is exactly the "interfering sinusoid" case we analyzed earlier with the results

$$\mathrm{A}_{\mathrm{v}}(t) = A_c\sqrt{1+\rho^2+2\rho\cos(\theta_i(t))} \text{ and } \phi_{\mathrm{v}}(t) = \tan^{-1}\left(\frac{\rho\sin(\theta_i(t))}{1+\rho\cos(\theta_i(t))}\right)$$

with $\theta_i(t) = \phi_i(t)$, the phase modulation of the interfering signal.

Interference

$$\mathrm{A_v}(t) = A_c\sqrt{1+\rho^2+2\rho\cos(\theta_i(t))} \text{ and } \phi_{\mathrm{v}}(t) = \tan^{-1}\left(\frac{\rho\sin(\theta_i(t))}{1+\rho\cos(\theta_i(t))}\right)$$

The demodulated signal is then

$$\mathrm{y}_D(t) = \phi_{\mathrm{v}}(t) = \frac{d}{dt}\left(\tan^{-1}\left(\frac{\rho\sin(\phi_i(t))}{1+\rho\cos(\phi_i(t))}\right)\right).$$

Using $\frac{d}{dz}\left(\tan^{-1}(z)\right) = \frac{1}{1+z^2}$ and the chain rule of differentiation,

$$\mathrm{y}_D(t) = \frac{1}{1+\left(\frac{\rho\sin(\phi_i(t))}{1+\rho\cos(\phi_i(t))}\right)^2} \times \frac{\left[1+\rho\cos(\phi_i(t))\right]\rho\cos(\phi_i(t))\dot{\phi}(t)+\rho\sin(\phi_i(t))\rho\sin(\phi_i(t))\dot{\phi}(t)}{\left[1+\rho\cos(\phi_i(t))\right]^2}$$

$$\mathrm{y}_D(t) = \frac{\rho\cos(\phi_i(t))+\rho^2}{\left[1+\rho\cos(\phi_i(t))\right]^2+\rho^2\sin^2(\phi_i(t))}\dot{\phi}(t)$$

$$\mathrm{y}_D(t) = \frac{\rho\left[\rho+\cos(\phi_i(t))\right]}{1+\rho^2+2\rho\cos(\phi_i(t))}\dot{\phi}(t) = \alpha(\rho,\phi_i)\dot{\phi}(t)$$

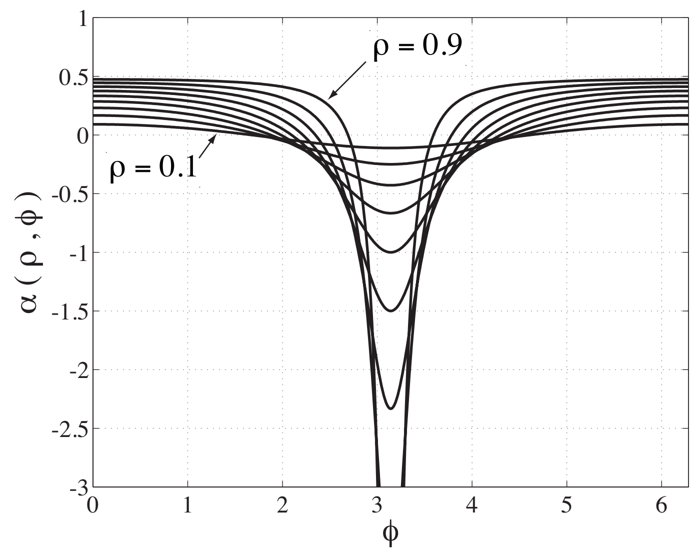
where $\alpha(\rho,\phi_i) = \frac{\rho\left[\rho+\cos(\phi_i(t))\right]}{1+\rho^2+2\rho\cos(\phi_i(t))}$

Interference

$$y_D(t) = \alpha(\rho, \phi_i)\dot{\phi}(t)$$

The $\dot{\phi}(t)$ factor suggests that the interference may be intelligible if $\alpha(\rho, \phi_i)$ is relatively constant with time. If $\rho >> 1$, then $\alpha(\rho, \phi_i) \cong 1$ and $y_D(t) \cong \dot{\phi}(t)$. But we wish to examine the case in which the two signals are approximately equal in strength, implying that $\rho \cong 1$.

$$\alpha(\rho, \phi_i) = \frac{\rho\left[\rho + \cos(\phi_i(t))\right]}{1 + \rho^2 + 2\rho\cos(\phi_i(t))} = \begin{cases} \rho/(1+\rho) & , \ \phi_i = 0 + 2n\pi \\ \rho^2/(1+\rho^2) & , \ \phi_i = \pi/2 + n\pi \\ -\rho/(1-\rho) & , \ \phi_i = \pi + 2n\pi \end{cases} \quad , \ n \text{ an integer}$$



Interference

$$y_D(t)=\alpha(\rho,\phi_i)\dot{\phi}(t) \quad , \quad \alpha(\rho,\phi_i)=\frac{\rho\left[\rho+\cos(\phi_i(t))\right]}{1+\rho^2+2\rho\cos(\phi_i(t))}$$

As $\rho \to 1$, $\alpha \to 0.5$ and $y_D(t) \to 0.5\dot{\phi}(t)$.

For $\rho < 1$, the strength of the demodulated interference depends mostly on the peak-to-peak value of α

$$\alpha_{p-p}=\alpha(\rho,0)-\alpha(\rho,\pi)=\frac{2\rho}{(1-\rho)^2}$$

The interference effect is small-to-negligible for $\rho < 0.7$ and the interference captures the demodulated output signal when $\rho > 0.7$.