Chapter 2: Basic Modulation Techniques

In most communication systems, the modulated signal has the form

$$x_c(t) = A(t)\cos[\omega_c t + \phi(t)], \tag{2-1}$$

where ω_c is known as the *carrier frequency*, A(t) is the *envelope* and $\phi(t)$ is the *phase*.

Amplitude A(t) and phase $\phi(t)$ *may* depend on message m(t). When A(t) depends linearly on the message, and ϕ is a constant independent of m, we have *linear modulation*. When $\phi(t)$ depends on m(t), we have *nonlinear modulation*.

Linear Modulation

Double Sideband (DSB) is the first form of linear modulation we will consider. The general form of a DSB signal is

$$x_{DSB}(t) = A_c m(t)\cos[\omega_c t + \phi_0], \tag{2-2}$$

where A_c and ϕ_0 are constants. For convenience, we will assume that $\phi_0 = 0$. Figures 2-1a through 2-1c depict a block diagram of a DSB modulator, a sinusoidal message m and the DSB time domain wave form $x_{DSB}(t)$, respectively. Note that every sign change in m(t) results in a 180° phase shift in the transmitted signal $x_{DSB}(t)$.

DSB is very popular when used to transmit digital data. In this application, m(t) is a digital waveform that switches between +1 and -1 volts. Hence, m(t) switches the phase of the transmitted carrier by π radians. For this reason, for a ± 1 binary message, the modulation is called *phase-shift keying*.

The Fourier transform of x_{DSB} is

$$X_{DSB}(j\omega) = \mathcal{F}\left[x_{DSB}(t)\right] = \frac{A_c}{2}\left[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)\right], \tag{2-3}$$

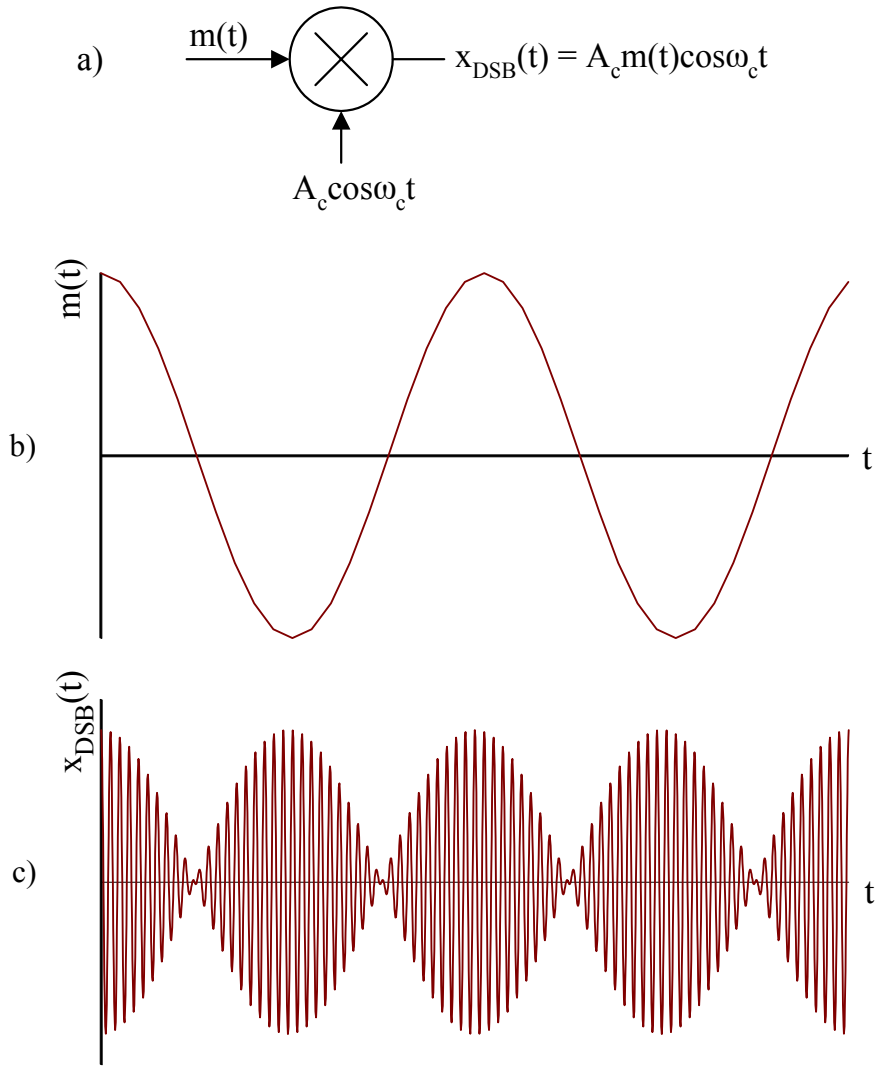


Figure 2-1: a) Block diagram of a DSB modulator. b) Sinusoidal message. c) The resulting $x_{DSB}(t)$.

where $M(j\omega) = \mathcal{F}[m(t)]$ is the Fourier transform of the message. As shown by Figure 2-2, $X_{DSB}(j\omega)$ is a scaled version of the message tranform that has been translated to $\pm\omega_c$. As is usual, we will assume that the message bandwidth is small compared to ω_c, so $x_{DSB}(t)$ is a narrow-band signal.

In general, $X_{DSB}(j\omega)$ contains a discrete carrier component (a *spectral line*) at ω_c, an *upper sideband* (the USB is the portion of $X_c(j\omega)$ that lies above the carrier ω_c) and a *lower sideband* (the LSB is the portion of $X_c(j\omega)$ which lies in the frequency range $0 < \omega < \omega_c$). If m(t) has no DC component, then the carrier in $X_c(j\omega)$ will be *suppressed* (*any* nonzero DC component

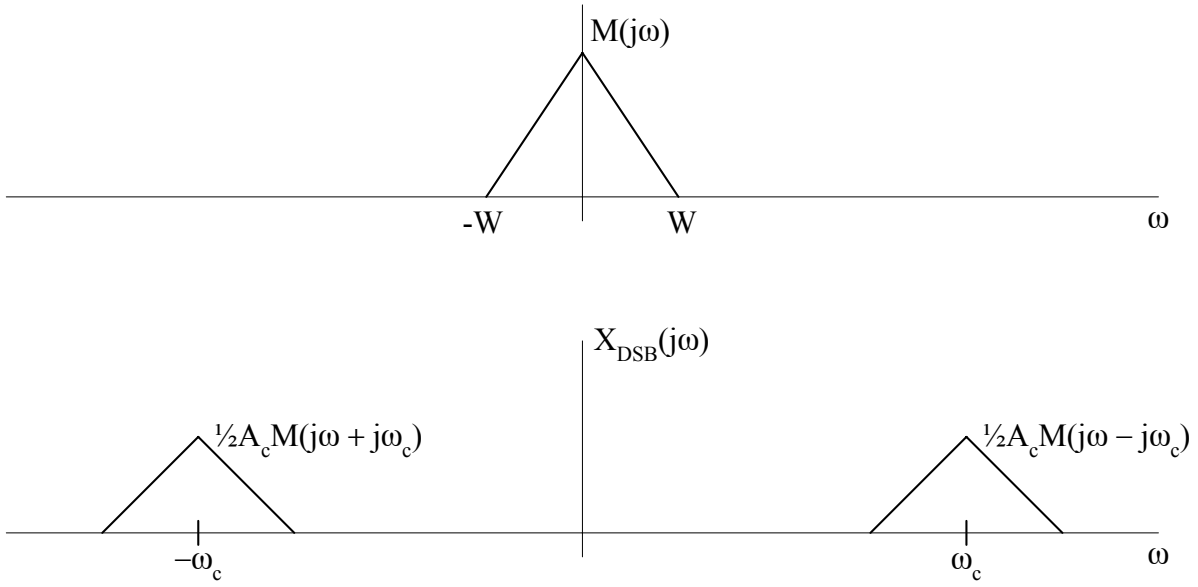


Figure 2-2: Spectrum M(jω) of message and spectrum $X_{DSB}(j\omega)$ of DSB signal.

of m(t) will lead to a nonzero carrier component). In many applications, in order to improve efficiency, we seek to allocate no transmitter power to the carrier (the carrier conveys no information about m(t) so it is desirable to allocated no power to the carrier). Finally, note that the transmission bandwidth is twice the message bandwidth.

DSB Demodulation

We assume that the received signal is a replica of the transmitted signal; that is, the signal

$$x_{DSB}(t) = A_c m(t)\cos\omega_c t \tag{2-4}$$

is received. As shown by Figure 2-3, demodulation involves multiplying x_{DSB} by a *phase coherent* replica of the carrier and then low-pass filtering the product. The output of the demodulator's multiplier is

$$d(t) = [A_c m(t)\cos\omega_c t]2\cos\omega_c t = A_c m(t) + A_c m(t)\cos 2\omega_c t\,. \tag{2-5}$$

The low-pass filter (LPF) following the multiplier filters out all components centered at $2\omega_c$. The output of the LPF is

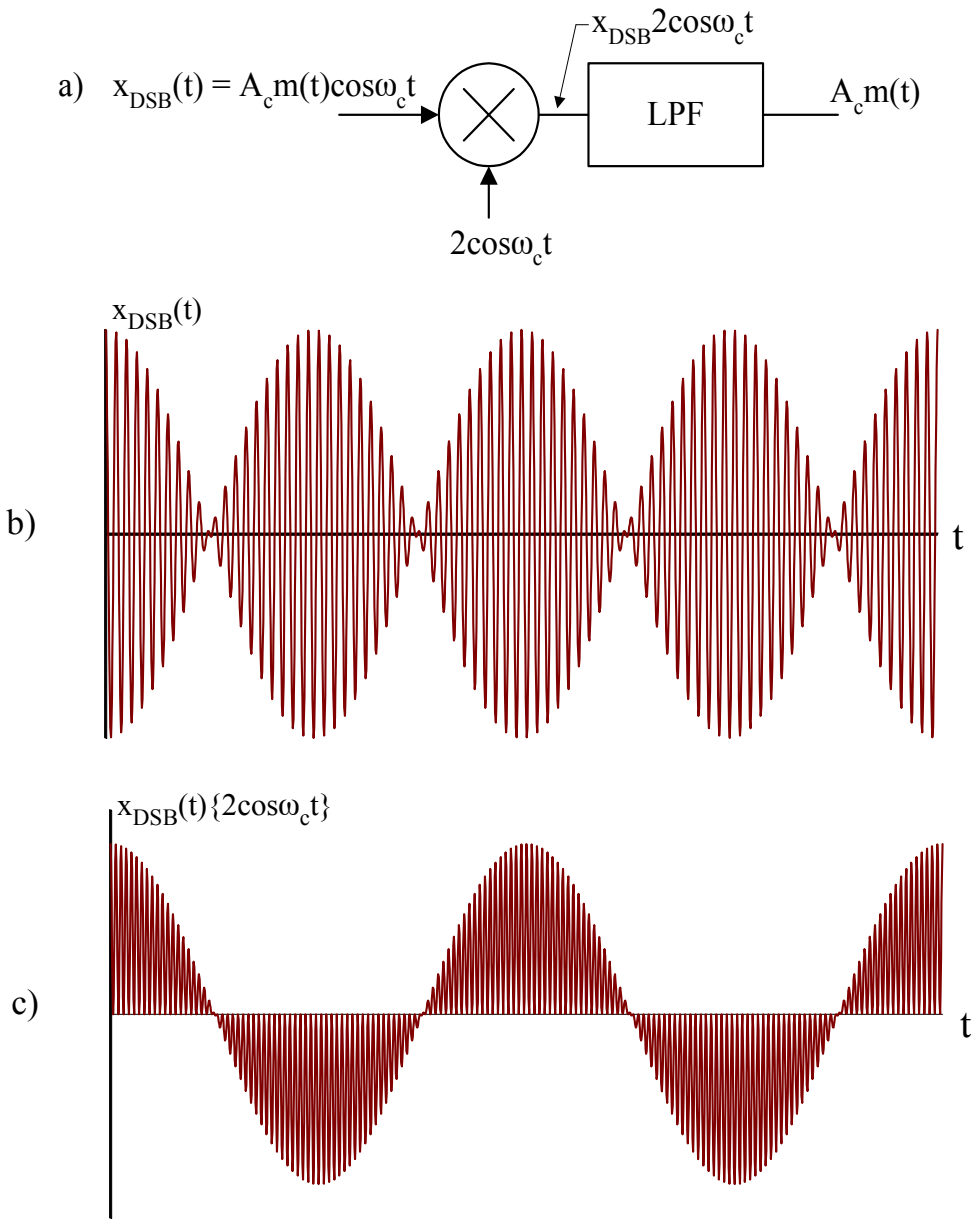


Figure 2-3: a) DSB demodulator, b)x_{DSB} and c) product of x_{DSB} and coherent carrier.

$$y_d(t) = A_c m(t)\,. \tag{2-6}$$

A fundamental problem with DSB is the need for a phase coherent reference (*i.e.*, $A\cos\omega_c t$ on Fig. 2-1a) at the receiver. Complicating this problem is the fact that a carrier may

not be transmitted, in many applications.

Let us analyze the effects of a phase error in the carrier used to demodulate x_r. Assume that our local reference is $2\cos(\omega_c t + \theta(t))$, where $\theta(t)$ is a phase error term. The multiplier's output is

$$d(t) = [A_c m(t)\cos\omega_c t]2\cos(\omega_c t + \theta(t)) = A_c m(t)\cos\theta(t) + A_c m(t)\cos(2\omega_c t + \theta(t)), \quad (2\text{-}7)$$

and the output of the LPF is, at best,

$$y_d(t) = A_c m(t)\cos\theta(t) \quad (2\text{-}8)$$

(we assume this signal is within the pass band of the LPF). In y_d, the time varying term $\cos\theta(t)$ could introduce serious distortion. On the other hand, depending on the application, it many not matter much, if kept small. When m(t) is human voice, we usually can tolerate a small nonzero frequency error $d\theta/dt$ and still make out what is being said. On the other extreme, when m(t) is digital data, and a computer interprets the demodulated y_d, small phase errors can be devastating.

There are ways to *regenerate* a phase coherent carrier at the receiver, even if one is not transmitted. One commonly used method squares the received DSB signal to produce

$$x_r^2(t) = A_c^2 m^2(t)\cos^2\omega_c t = \tfrac{1}{2}A_c^2 m^2(t) + \tfrac{1}{2}A_c^2 m^2(t)\cos 2\omega_c t. \quad (2\text{-}9)$$

If m(t) is a power signal, then $m^2(t)$ has a nonzero DC average. In this case, x_r^2 has a discrete spectral component at $2\omega_c$ which can be extracted by a narrow band filter centered at $2\omega_c$. The extracted $2\omega_c$ component is divided by two in frequency (by a D flip-flop, for example) to generate a coherent reference at the receiver.

Figure 2-4 depicts a block diagram of a DSB demodulator that utilizes a squaring

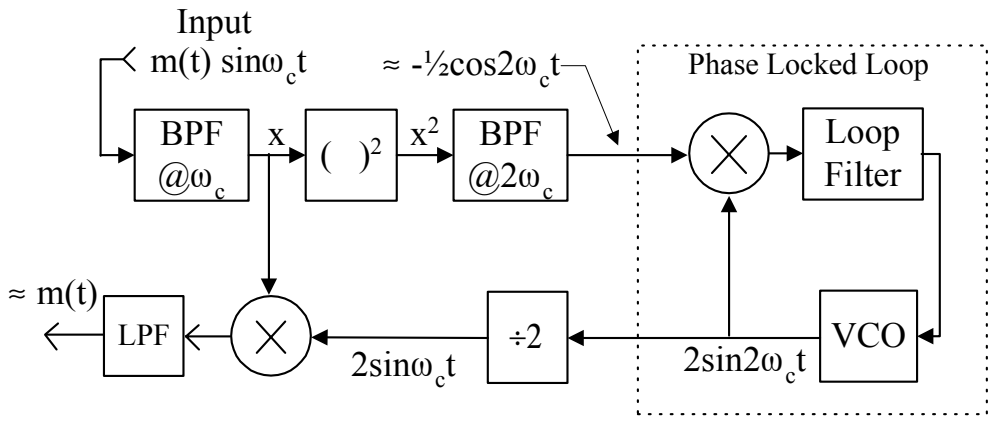


Figure 2-4: Block diagram of a squaring loop DSB demodulator. The VCO output is divided by two in frequency to obtain a phase-coherent reference for coherent demodulation of the input DSB signal.

operation. In this application, a phase lock loop (PLL) serves to recover the $2\omega_c$ component in x^2 (the PLL "locks" onto the $2\omega_c$ component in x^2). That is, the PLL acts like a narrow band-pass filter that extracts the $2\omega_c$ component from its input. Under proper operation (*i.e.,* when the closed loop phase error is small), the phase of the VCO output leads by $\pi/2$ radians the phase of the PLL input. Hence, the VCO output is $2\sin\omega_c t$, a result that is divided by two in frequency to produce a coherent reference for demodulating the DSB input. Since the demodulator relies on the nonlinear operation x^2, the demodulator is often called a *squaring loop*.

Amplitude Modulation

Amplitude modulation was invented by Reginald A. Fessenden, a Canadian, who successfully transmitted, for the first time, the sound of human voice. He first transmitted voice between two 50-foot towers on Cobb Island located in the Potomac River, Washington D.C., December 23rd, 1900. Prior to AM, radio operators used crude spark gap transmitters to send only Morse code.

At the time, few people shared Fessenden's belief that broadcasting the human voice was possible, much less practical. When Fessenden asked the opinion of the great Thomas Edison, Edison replied, "Fezzie, what do you say are man's chances of jumping over the moon? I think one is as likely as the other." Fortunately, Edison was wrong.

It took six years for Fessenden to refine his invention but, on Christmas Eve 1906, Fessenden made the first radio broadcast (of speech and music) in history from Brant Rock Station, Massachusetts. Radio operators on ships in the Atlantic were shocked to hear a human voice emitting from the equipment they used to receive Morse code. Many operators called their Captains to the radio room, where they heard Fessenden make a short speech, play a record, and give a rendition of "O Holy Night" on his violin.

Since the 1920's, AM has been used in commercial broadcasting. Also, it is still used in civil aviation and amateur radio. Most signal generators can be AM modulated by a built-in modulator. Also, other types of test equipment can modulate/demodulate AM.

AM results when a DC bias A is added to message m(t) prior to the DSB modulation process (in what follows, we assume that m(t) has a zero DC component). This results in the transmission of a carrier component if bias $A \neq 0$.

The AM signal is defined as

$$\begin{aligned} x_{AM}(t) &= [A + m(t)]A_c \cos\omega_c t = A_c'[1 + a\, m_n(t)]\cos\omega_c t \\ &= \underbrace{A_c' \cos\omega_c t}_{\text{carrier component}} + \underbrace{A_c' a\, m_n(t)\cos\omega_c t}_{\text{sideband component}} \end{aligned}, \tag{2-10}$$

where

$$A_c' \equiv AA_c,\ m_n(t) \equiv \frac{m(t)}{\left|\min_t\{m(t)\}\right|},\ a \equiv \frac{\left|\min_t\{m(t)\}\right|}{A}. \tag{2-11}$$

$m_n(t)$ is message m(t) normalized *so that the minimum value of* $m_n(t)$ *is* -1. Parameter a, $a \geq 0$, is known as the *modulation index*. The quantity $A_c'[1 + a\, m_n(t)]$ is known as the *envelope* of AM signal $x_c(t)$. *For* $a \leq 1$*, the envelope is never negative, and the message appears to "ride" on top*

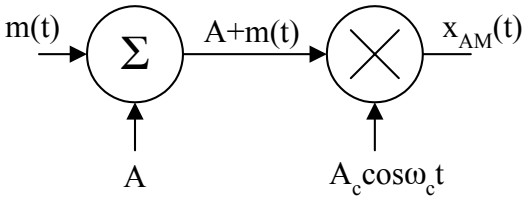


Figure 2-5: AM modulator

of the transmitted signal. For a > 1, the signal experiences a π-radian phase shift at each zero crossing of the envelope (a fact of important significance as discussed below). See Figure 2-5 for a block diagram of an AM modulator and Figure 2-6 for an example of a message and AM modulated signal.

In the frequency domain, the spectrum of AM is

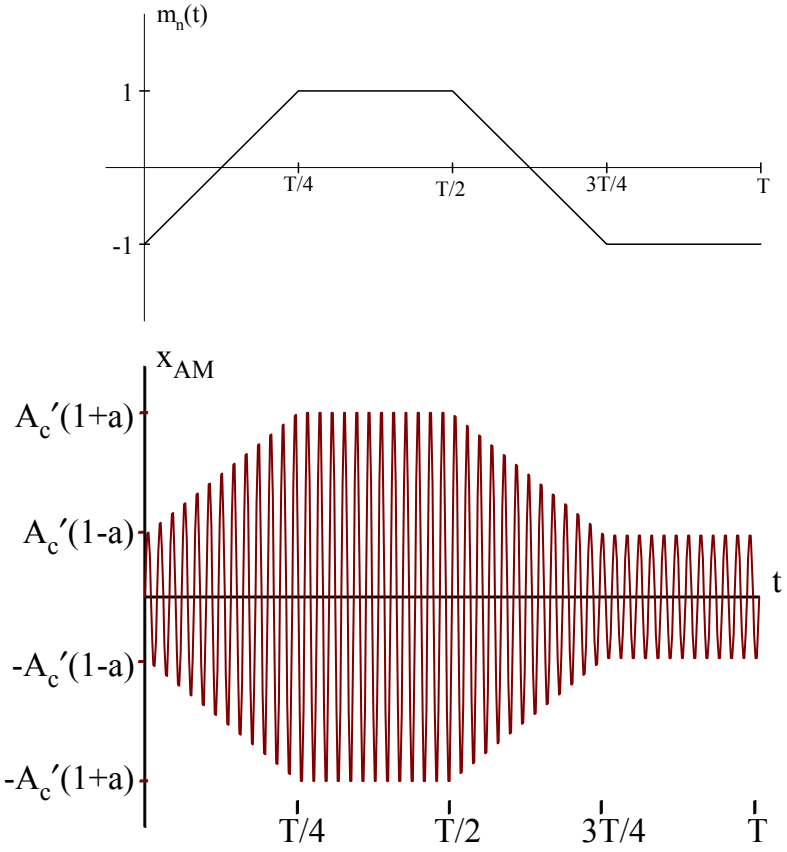


Figure 2-6: Message m_n and AM waveform x_{AM}.

$$X_{AM}(\omega) = \mathcal{F}\left[x_{AM}(t)\right] = \mathcal{F}\left[A_c' \cos\omega_c t + A_c' a\, m_n(t)\cos\omega_c t\right]$$

$$= \underbrace{A_c'\pi\left[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)\right]}_{\text{carrier spectrum}} + \underbrace{\frac{A_c' a}{2}\left[M_n(\omega+\omega_c)+M_n(\omega-\omega_c)\right]}_{\text{sideband spectrum}} \tag{2-12}$$

Note the existence of discrete carrier spectral lines at $\pm\omega_c$. Also, the translated message terms $M_n(\omega\pm\omega_c)$ contain upper and lower sidebands ($M_n(\omega-\omega_c)$, for $\omega > \omega_c$, is an upper side band while $M_n(\omega-\omega_c)$, for $0 < \omega < \omega_c$, is a lower side band). Finally, note that the transmission bandwidth of AM is twice the message bandwidth, just like DSB.

The transmitted signal power is divided between the carrier and information conveying sidebands. Power allocated to the carrier is (in the sense that it does not convey information) wasted. This leads to the notion of efficiency.

Efficiency of AM

The average transmitted power of the AM signal is

$$\left\langle x_c(t)^2 \right\rangle = \left\langle [A+m(t)]^2 A_c^2 \cos^2\omega_c t \right\rangle = \left\langle [A+m(t)]^2 A_c^2 \tfrac{1}{2}(1+\cos 2\omega_c t) \right\rangle. \tag{2-13}$$

If m(t) is slowly varying with respect to $\cos 2\omega_c t$, this last equation leads to the approximation

$$\left\langle x_c(t)^2 \right\rangle \approx \tfrac{1}{2}A_c^2\left[A^2 + 2A\left\langle m(t)\right\rangle + \left\langle m(t)^2\right\rangle\right] = \tfrac{1}{2}A_c^2\left[A^2 + \left\langle m(t)^2\right\rangle\right], \tag{2-14}$$

since $\left\langle m(t)\right\rangle = 0$ by assumption.

Define *efficiency* as the percentage of total power that conveys information. More precisely, *efficiency* is the percentage of total transmitted power that is in the sidebands. From the last equation, we can write

$$\text{Efficiency} = \frac{\left\langle m^2\right\rangle}{A^2 + \left\langle m^2\right\rangle} \times (100\%). \tag{2-15}$$

Since $m(t) = aAm_n(t)$ we have

$$\text{Efficiency} = \frac{a^2\left\langle m_n^2\right\rangle}{1+a^2\left\langle m_n^2\right\rangle} \times (100\%)\,. \tag{2-16}$$

For $a \le 1$ the maximum efficiency is 50% (for a square wave message with a = 1). If m(t) is a sine wave, and a = 1, then efficiency = 33%. For most complex messages, such as voice, efficiency is under 10%.

Example 2-1: Determine the efficiency and output $x_{AM}(t)$ for an AM modulator operating with a modulation index of .5. The carrier power is 50 watts, and the message signal is

$$m(t) = 4\cos[\omega_m t - \tfrac{\pi}{9}] + 2\sin[2\omega_m t]\,, \tag{2-17}$$

a graph of which is depicted by Figure 2-7.

Solution: Observe the message signal shown in Fig. 2.7. The minimum value of m(t) is -4.364, and the minimum falls at $\omega_m t = 2\pi(.435)$. The normalized message signal is given by

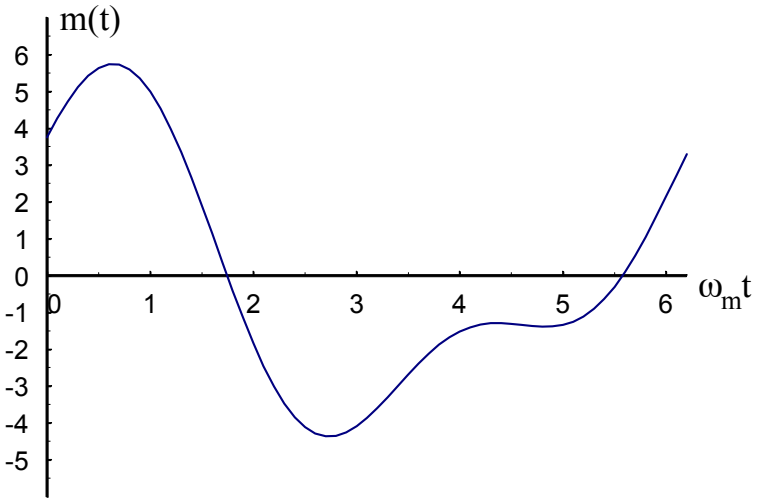


Figure 2-7: One period of a two tone message.

$$m_n(t) = \frac{1}{4.364}\left[4\cos[\omega_m t - \tfrac{\pi}{9}] + 2\sin[2\omega_m t]\right] = .9166\cos[\omega_m t - \tfrac{\pi}{9}] + .45832\sin[2\omega_m t]. \quad (2\text{-}18)$$

The mean-square value of $m_n(t)$ is

$$\left\langle m_n^2(t)\right\rangle = \tfrac{1}{2}(.9166)^2 + \tfrac{1}{2}(.4583)^2 = .5251 \quad (2\text{-}19)$$

Finally, the efficiency is

$$\text{Efficiency} = \frac{.25(.5251)}{1 + .25(.5251)} \times (100\%) = 11.60\% \quad (2\text{-}20)$$

Since the carrier power is 50 watts, we have

$$\tfrac{1}{2}(A_c')^2 = 50 \quad (2\text{-}21)$$

which implies that

$A_c' = 10$. Since sin(x) = cos(x - π/2), we can write

$$x_c(t) = 10\left[1 + .5\left(.9166\cos(\omega_m t - \tfrac{\pi}{9}) + .4583\cos(2\omega_m t - \tfrac{\pi}{2})\right)\right]\cos\omega_c t. \quad (2\text{-}22)$$

Transmitted Power in AM Signal

The transmitted AM signal is given by

$$x_{AM}(t) = A_c'\left[1 + a\, m_n(t)\right]\cos\omega_c t \quad (2\text{-}23)$$

The instantaneous transmitted power is $x_{AM}^2(t)$. The *average power* in x_{AM} is given by

$$P_{AVG} = \left\langle x_{AM}^2(t) \right\rangle = (A_c')^2 \left\langle [1 + a\, m_n(t)]^2 \frac{1}{2}(1 + \cos 2\omega_c t) \right\rangle = \frac{(A_c')^2}{2}\left[1 + a^2 \left\langle m_n^2(t) \right\rangle\right], \qquad (2\text{-}24)$$

watts. To obtain this result, we used the fact that message m(t) has an average value of zero.

Often, power is specified in terms of *peak envelope power.* The envelope $A_c'[1 + am_n(t)]$ is slowly varying with respect to the RF carrier $\cos\omega_c t$. Over every cycle of the RF carrier, the envelope is approximately constant. The *peak envelope power* (PEP) is the instantaneous power $[x_{AM}(t)]^2$ averaged over the RF cycle having the greatest amplitude. Hence, we can write

$$P_{PEP} = \frac{(A_c')^2}{2} \max_t [1 + a\, m_n(t)]^2. \qquad (2\text{-}25)$$

For $m_n = \cos\omega_m t$ and a = 1, we get $P_{AVG} = 3A'^2/4$ and $P_{PEP} = 2A'^2$, so P_{PEP} is about 2.7 times P_{AVG}. For a message consisting of a human voice, the PEP power might be two or three times (or more) the average power.

AM Coherent Demodulation

Amplitude modulation can be demodulated coherently, see Figure 2-8. The demodulate output contains a constant DC term that is usually eliminated by a lack of DC response in the audio stages that follow the demodulator. The coherent reference needed by the demodulator can be supplied by phase locking a PLL onto the carrier component of x_{AM}. The PLL acts like a narrow-band filter that extracts the carrier component of the signal. Note that coherent

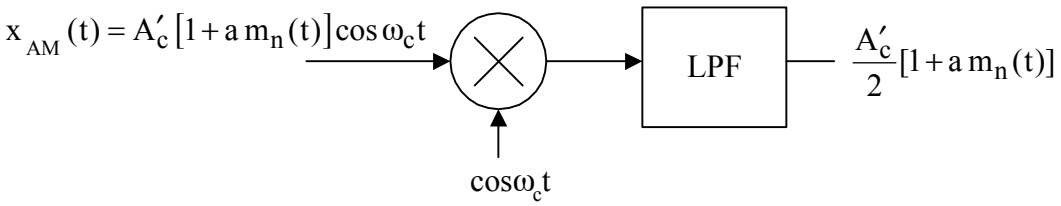


Figure 2-8: Coherent demodulation of AM.

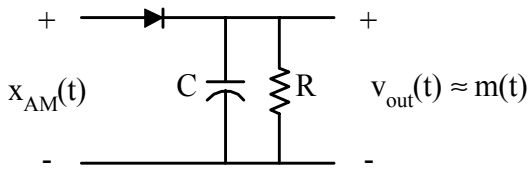


Figure 2-9: A simple envelope detector.

demodulation can be used regardless of the modulation index a.

AM Demodulation - Envelope Detection

If modulation index a is equal to, or less than, unity ($a \le 1$), AM can be demodulated by a very simple technique called *envelope detection*. On the other hand, if $a > 1$, envelope detection will not work; the detector output audio will be highly distorted. The reason for this is simple. For $a > 1$, the signal experiences a 180° phase change at each envelope sign change, and envelope detectors are insensitive to signal phase. So, an envelope detector will not respond to sign changes in the AM signal envelope, and distortion of the recovered audio results. A simple envelope detector will only work if $0 \le a \le 1$. Figure 2-9 depicts a schematic diagram of a simple envelope detector.

As long as envelope $A_c'[1 + a\,m_n(t)]$ is non-negative, message m(t) appears to "ride" on top of half-wave rectified $x_c(t)$. In this case, a close approximation of $A_c'[1 + a\,m_n(t)]$ can be obtained by smoothing the output of the diode with an RC circuit. The time constant of the RC smoothing circuit is not extremely critical. However, as a general rule of thumb, best results can be obtained if

$$1/f_c << RC << 1/f_B, \tag{2-26}$$

where f_c is the carrier frequency in Hz, and f_B is the message bandwidth, in Hz.

The diode is assumed to have a small forward "on" resistance; the charging time constant is extremely small (charging occurs when $x_{AM} > v_{out}$). Except for the drop across the diode, output v_{out} "follows" input x_{AM} when the diode is conducting. When $x_{AM} < v_{out}$, the diode is not conducting, and capacitor voltage v_{out} discharges through the resistor. If the discharging time

constant RC is too small, a severe "saw-tooth-like buzz", at frequency f_c, will be imposed on the demodulated message. If RC is too large, the output will "float" on envelop peaks, and severe distortion will occur. It is important to realize that, due to the nonlinear switching action of the diode, the role of the RC circuit is to smooth the output and form a signal that follows closely the modulation envelope. In this nonlinear circuit, do not think of the RC circuit as just a conventional, single-pole low-pass filter.

A relatively simple upper bound can be obtained on time constant RC for the case of a sinusoidal message. As shown on Figure 2-10, assume that the capacitor discharges from the carrier peak value $E_0 = A_c'\left[1 + a\cos\omega_m t_0\right]$ at time t_0. Note that t_0 is associated with a peak in the carrier, not the message or envelope ($\cos\omega_m t_0$ can be any value between –1 and +1). For a range of t between t_0 and $t_0 + 1/f_c$, the capacitor is discharging, so the capacitor voltage is

$$V_c(t) = E_0 e^{-(t-t_0)/RC}. \tag{2-27}$$

The time interval between two successive carrier peaks is $1/f_c = 2\pi/\omega_c$. Since $RC >> 1/\omega_c$, the quantity t/RC is small for time t between carrier peaks and

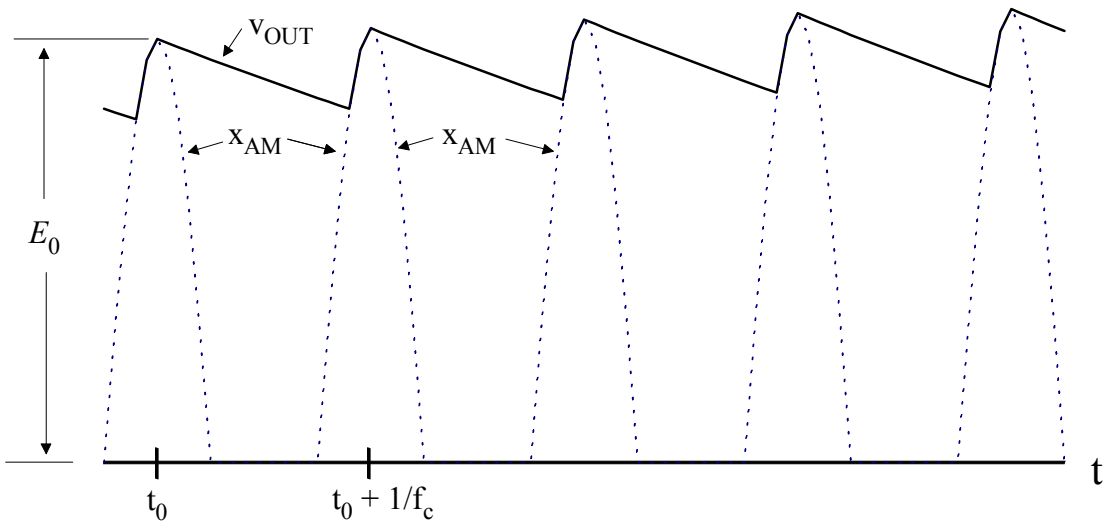


Figure 2-10: Positive half of x_{AM} shown as dotted-line graph. Output $v_{OUT}(t)$ depicted as solid line graph. $E_0 = v_{OUT}(t_0) = A'_c[1+a\cos\omega_m t_0]$, a local peak in the carrier (the message and envelope may not peak at t_0).

$$V_c(t) \approx E_0\left(1-\frac{t-t_0}{RC}\right). \tag{2-28}$$

If $V_c(t)$ is to follow the envelope, then it is required that

$$\left[1+a\cos\omega_m t_0\right]\left(1-\frac{1}{RC\,f_c}\right) \le 1+a\cos\omega_m(t_0+1/f_c). \tag{2-29}$$

Since $\omega_m << \omega_c$, we have (use the identity $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ and the fact that $\cos\beta \approx 1$ and $\sin\beta \approx \beta$ for small β)

$$\begin{aligned} 1+a\cos\omega_m(t_0+1/f_c) &= 1+a\cos(\omega_m t_0+\omega_m/f_c) \\ &= 1+a\cos(\omega_m t_0)\cos\left(\frac{\omega_m}{f_c}\right)-a\sin(\omega_m t_0)\sin\left(\frac{\omega_m}{f_c}\right) \\ &\approx 1+a\cos(\omega_m t_0)-a\left(\frac{\omega_m}{f_c}\right)\sin(\omega_m t_0). \end{aligned} \tag{2-30}$$

Now, the last two equations combine to yield

$$\left[1+a\cos\omega_m t_0\right]\left(\frac{1}{RC\,f_c}\right) \ge a\left(\frac{\omega_m}{f_c}\right)\sin(\omega_m t_0). \tag{2-31}$$

This result can be written as

$$\frac{1}{RC}+\frac{a}{RC}\cos\omega_m t_0 \ge a\,\omega_m \sin\omega_m t_0 \tag{2-32}$$

or

$$a\left(\omega_m \sin\omega_m t_0 - \frac{1}{RC}\cos\omega_m t_0\right) = a\sqrt{\omega_m^2 + \left(\frac{1}{RC}\right)^2}\,\sin\left(\omega_m t_0 - \tan^{-1}(1/\omega_m RC)\right) \le \frac{1}{RC} \quad . \tag{2-33}$$

Since $\omega_m t_o$ is arbitrary, we must have

$$a\sqrt{\omega_m^2 + \left(\frac{1}{RC}\right)^2} \le \frac{1}{RC} \tag{2-34}$$

and

$$RC \le \frac{\sqrt{1-a^2}}{a\omega_m}, \tag{2-35}$$

the desired upper bound on time constant RC.

Matlab Envelope Detector Simulation

The Matlab program listed in Figure 2-11 envelope detects the AM signal

$$v_{in}(t) = [1 + a\sin(t)]\sin(Wt) \tag{2-36}$$

over the time period $0 \le t \le 2\pi$. The results are depicted by Figs. 2-12 through 2-15; these figures show the input v_{in} as the "thin line" plot, and they show the diode-based envelope detector output as a "thick line" plot (riding on top of v_{in}). As given by (2-36), AM signal $v_{in}(t)$ uses ω_m = 1 rad/sec, and (2-35) yields an upper bound of $RC < \sqrt{1-a^2}/a$ for "floating distortion" not to occur.

To aid visualization, the values of $\omega_m = 1$, W = 50 and Dt = $2\pi/1000$ (output time step)

```
%Envelope.m

%Envelope.m detects an AM waveform

global RC, alpha, W, Dt;

t = 0 : 2*pi/1000 : 2*pi;

%Allocate memory for input and output arrays
Vin    = zeros(1,1001);
Vout = zeros(1,1001);

%Define input array
Vin = ( 1 + alpha*sin(t) ).*sin(W*t);

%First point of output is the initial value of the envelope
Vout(1) = 1;

%Compute output over all points
for i = 2:1001
  if Vin(i) > Vout(i-1);
    Vout(i) = Vin(i);
  else
    Vout(i) = Vout(i-1)*exp(-Dt/RC);
  end
end

%Plot input then pause
plot(t, Vin)
axis([0   2*pi   -1-alpha   1+alpha])

pause

%Hit any key to plot output
plot(t, Vout)
axis( [0  2*pi  0  1+alpha] )
```

Figure 2-11: Matlab program for simulation of an envelope detector.

were used in all plots. The carrier frequency is 50 times the message frequency (a ratio of 50 is smaller than what you would normally encounter in practice). Also, a = .5 (50% modulation depth) was used for Figures 2-12 through 2-14. For Fig. 2-12, the RC time constant is $2\pi/10$, a little bit too small. The value RC = $2\pi/10$ = .628 is significantly below upper bound $\sqrt{1-.5^2}\,/.5 = 1.732$ (so no floating distortion occurs). For Fig. 2-13, the RC time constant is $2\pi/5$, a value that is just about right. The value RC = $2\pi/5$ = 1.26 is less than upper bound

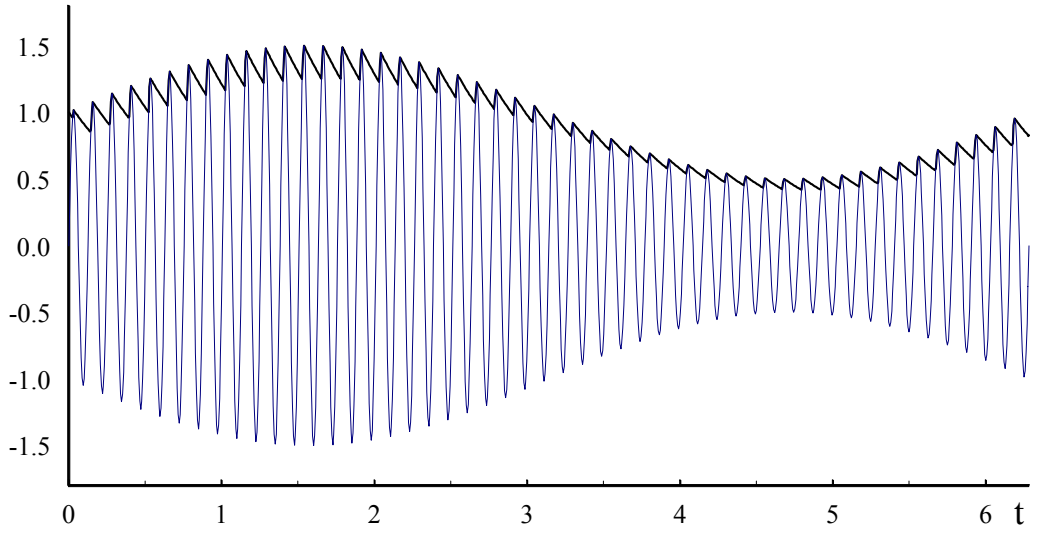


Figure 2-12: Input AM signal (thin line plot) and output of envelope detector (thick line plot). The modulation index is ½. The RC time constant is 2π/10, a value that is a bit too small.

$\sqrt{1-.5^2}/.5 = 1.732$ (so no floating distortion occurs). For Fig. 2-14, the RC time constant is 2π/3, a value that is a bit too large. The value RC = 2π/3 = 2.09 is above upper bound $\sqrt{1-a^2}/a = 1.732$, and floating distortion occurs.

As should be evident by now, a "good" value for RC depends on modulation index a. As index a approaches unity, you must use smaller values of RC to prevent demodulator output "floating" with its associated harmonic distortion (some output distortion is unavoidable for

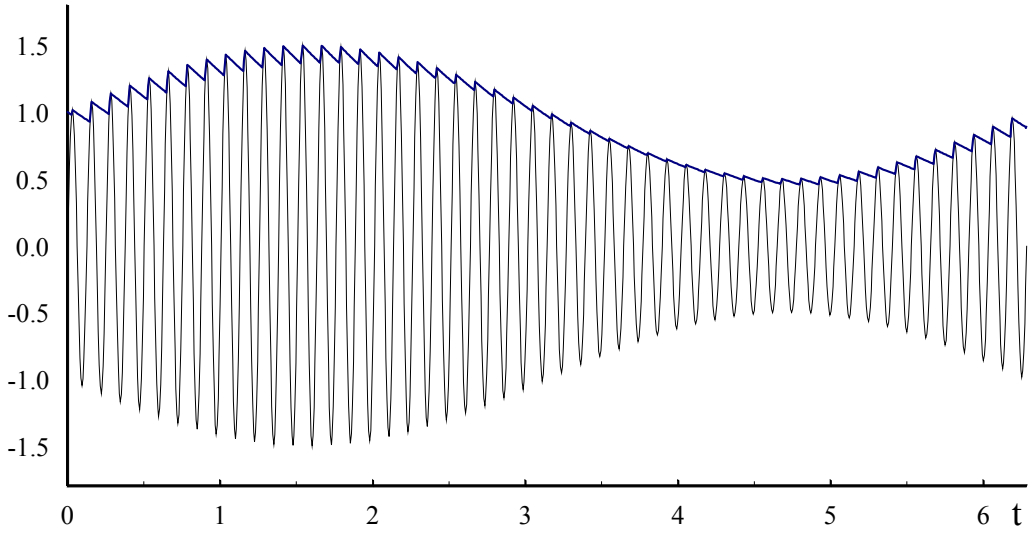


Figure 2-13: Input AM signal (thin line plot) and output of envelope detector (thick line plot). The modulation index is ½. The RC time constant is 2π/5, a value that is just about right.

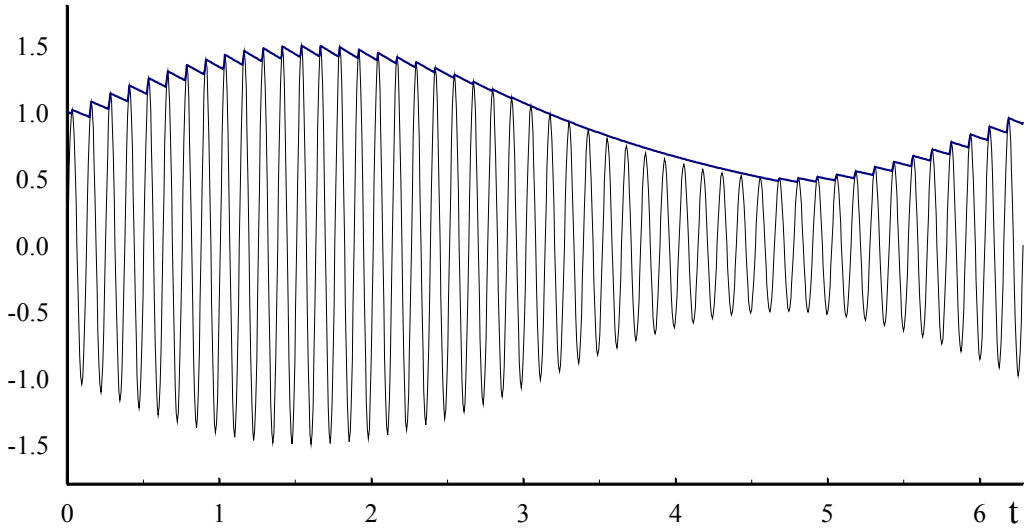


Figure 2-14: Input AM signal (thin line plot) and output of envelope detector (thick line plot). The modulation index is ½. The RC time constant is 2π/3, a value that is a bit too large (the detector output floats" above the true envelope over part of the modulation period).

near-unity index values). For example, Figure 2-13 shows good results with RC = 2π/5 and a = ½. However, significant distortion occurs if the same value of RC is used with a = .95, as can be seen from examining Figure 2-15.

Square-Law Detector

An amplitude modulated signal can be demodulated by a *square law* detector, if the

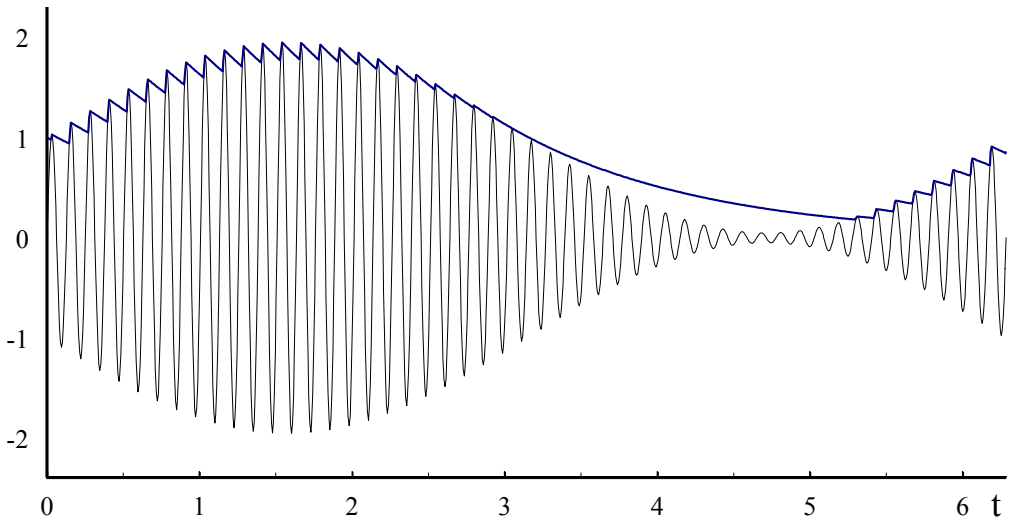


Figure 2-15: Input AM signal (thin line plot) and output of envelope detector (thick line plot). A value of modulation index a = .95 was used to obtain this plot. The RC time constant is 2π/5, a value that is too large (for a = .95) as is evident by the significant amount of detector output "floating".

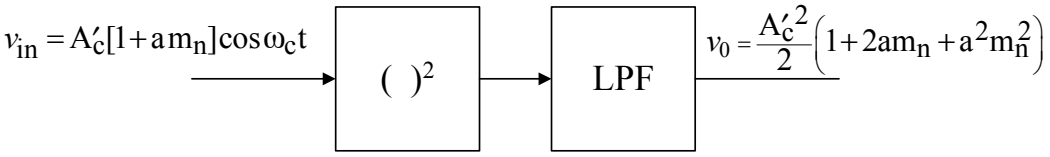


Figure 2-16: A square-law detector.

modulation index is sufficiently small. As depicted by Figure 2-16, a square law detector forms its output v_o by low-pass filtering the square of the input v_{in}. If $v_{in} = A_c'[1 + a\,m_n(t)]\cos\omega_c t$ we get

$$\left(A_c'[1 + a\,m_n(t)]\cos\omega_c t\right)^2 = A_c'^2\left(1 + 2a\,m_n(t) + a^2 m_n^2(t)\right)\tfrac{1}{2}(1 + \cos 2\omega_c t). \tag{2-37}$$

The low-pass filter removes the $2\omega_c$ component to produce the output

$$v_o(t) = \frac{A_c'^2}{2}\left(1 + 2a\,m_n(t) + a^2 m_n^2(t)\right). \tag{2-38}$$

The second-order term $m_n^2(t)$ introduces second-order harmonic distortion that can be severe if modulation index a is not small. On the other hand, if a << 1 and a blocking capacitor is used to remove the DC component, this last result can be approximated by

$$v_o(t) \approx A_c'^2 a\,m_n(t). \tag{2-39}$$

The nonlinear squaring operation can be implemented by a diode that is forward biased

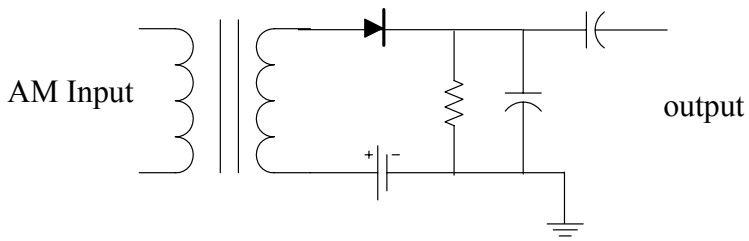


Figure 2-17: Approximation to Square Law Detector

into the "knee" of its *i-v* characteristic, see Fig. 2-17. The DC voltage source (the battery on the figure) serves to forward bias the diode into its "square law" region. It is important to remember that the input signal is small compared to the DC bias so that the diode *always is forward biased* (unlike the envelope detector). Unlike its use in the envelope detector, the RC network serves as a band-pass filter to extract the demodulated message.

To some degree of efficiency, a wide range of nonlinear operations will demodulate AM. In fact, in the presence of a strong transmitted AM signal, it is hard to *prevent* demodulation of the signal by *rectifying connections/junctions* in telephone sets, loudspeaker coils, etc. Often, it is necessary to place by-pass capacitors across devices in order to short-circuit "picked-up" radio frequency (RF) currents and prevent unwanted demodulation of a strong AM modulated signal.

Single Sideband Modulation

In DSB, either sideband contains sufficient information to reconstruct the message m(t). Elimination of one of the sidebands results in *single sideband modulation* (SSB). The signal is known as *lower sideband* (LSB) if the upper sideband is eliminated, and it is known as *upper sideband* (USB) if the lower sideband is eliminated. Figure 2-18 depicts single-sided spectral plots of the message $M(j\omega)$, double sideband $X_{DSB}(j\omega)$, lower sideband $X_{LSB}(j\omega)$, and upper sideband $X_{LSB}(j\omega)$.

In 1915, SSB was patented by John Carson. Originally, in the telephone system, it was used to frequency-division multiplex (FDM) multiple voice channels onto one cable. In radio communication today, SSB is very popular for the transmission of voice information.

SSB has some obvious advantages. First, it requires only half of the transmission bandwidth, as compared to DSB and AM. In this era of government sponsored spectrum auctions, spectrally efficient forms of modulation can improve both system performance *and* one's bottom line. When system design and bandwidth are optimized for a given modulation format, both DSB and SSB offer similar performance in terms of receiver output signal-to-noise (SNR) ratio (for a given received signal power and noise spectral density). However, when compared to AM with its high percentage of power allocated to the carrier, SSB offers much

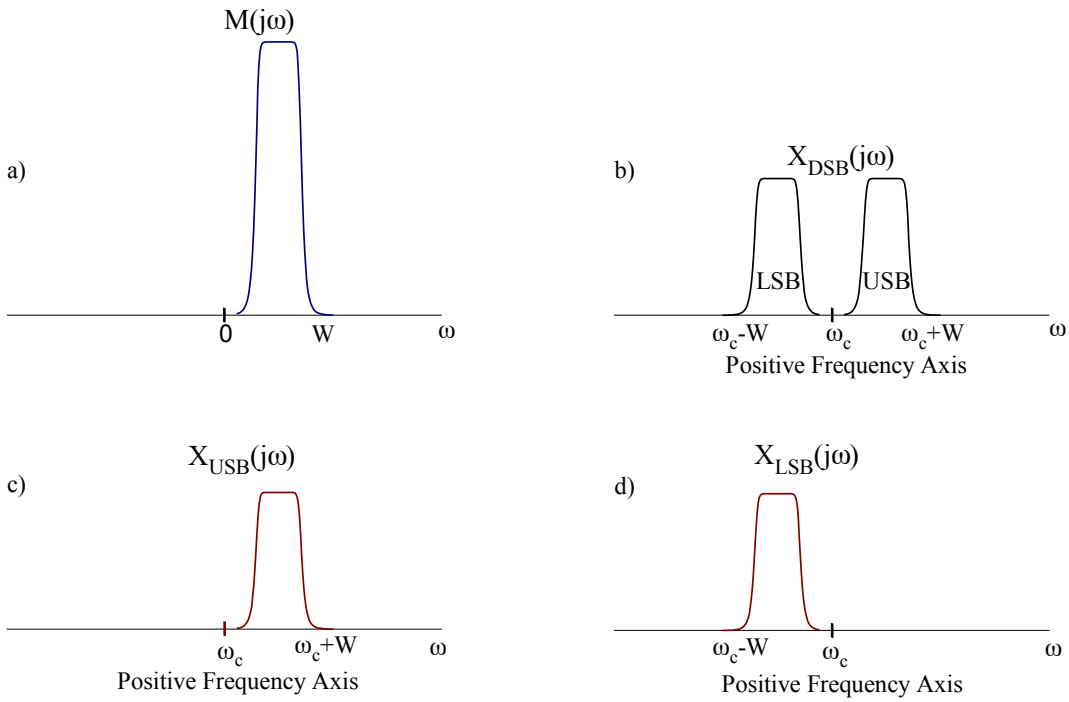


Figure 2-18: Single-sided spectral plots of a) message M(jω), b) double sideband $X_{DSB}(j\omega)$, c) upper sideband $X_{USB}(j\omega)$ and lower sideband $X_{LSB}(j\omega)$.

improvement over AM in terms of receiver output SNR (for a given received signal power and noise spectral density).

There are two commonly used methods to generate SSB. The first is called the *phasing method*, and it gained popularity early in the practical development and use of SSB (primarily in the 1950's). The second method is called the *filter method*. In the early days of SSB development, good sideband filters were expensive and hard to obtain (so the phasing method was dominant). However, in the 1960's and 1970's, significant technical advances were made in the design and manufacture of crystal band-pass filters, and good sideband filters became inexpensive. For this reason, the filter method of SSB generation is dominant today (however, with the advent of powerful DSP technology, the phasing method is making a comeback).

Filter Method of Single Sideband Generation

Figure 2-19 depicts a simplified block diagram of a filter-type SSB generator. First, double sideband is generated. Then, with the aid of a steep-skirt, band-pass filter (known as a

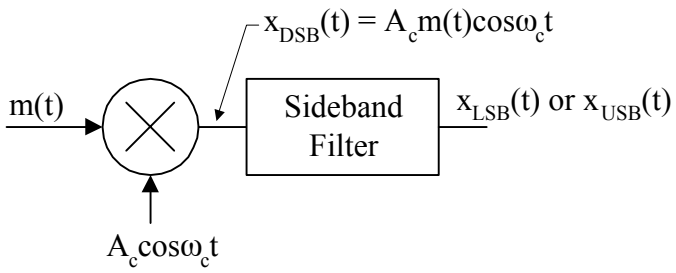


Figure 2-19: Filter method of single sideband generation.

sideband filter in the literature), the desired sideband is selected and the undesired sideband is filtered out.

Of course, in practice, the filtering process is imperfect, and a small amount of the unwanted sideband is transmitted (along with a small amount of unbalanced carrier). However, 40dB (or more) of unwanted sideband suppression is obtained easily with commercially available sideband filters.

Obviously, the sideband signal is generated at a fixed frequency $\omega_c = 2\pi f_c$. Commercially available, quartz crystal-based-technology sideband filters are available at f_c = 9Mhz and other standard frequencies. Also, mechanical filters are available at f_c = 455Khz and other standard frequencies (typically, mechanical filters are under 1Mhz). After generation at a fixed frequency, the single sideband signal is heterodyned (using one-or-more mixer, or frequency translation, stages) to the desired transmit frequency. Then, the signal is amplified in power and sent up the transmission line to the transmitting antenna.

Phasing Method of Single Sideband Generation

We develop the phasing method for generating LSB first. LSB will be generated if a DSB signal is passed through an ideal low-pass filter that extends from $-\omega_c$ to $+\omega_c$, as depicted by Figure 2-20. Filter $H_L(j\omega)$ can be represented as

$$H_L(j\omega) = \frac{1}{2}\left[\text{sgn}(\omega+\omega_c) - \text{sgn}(\omega-\omega_c)\right], \tag{2-40}$$

a result that is depicted by Figure 2-21. Apply the DSB signal to H_L; in the frequency domain,

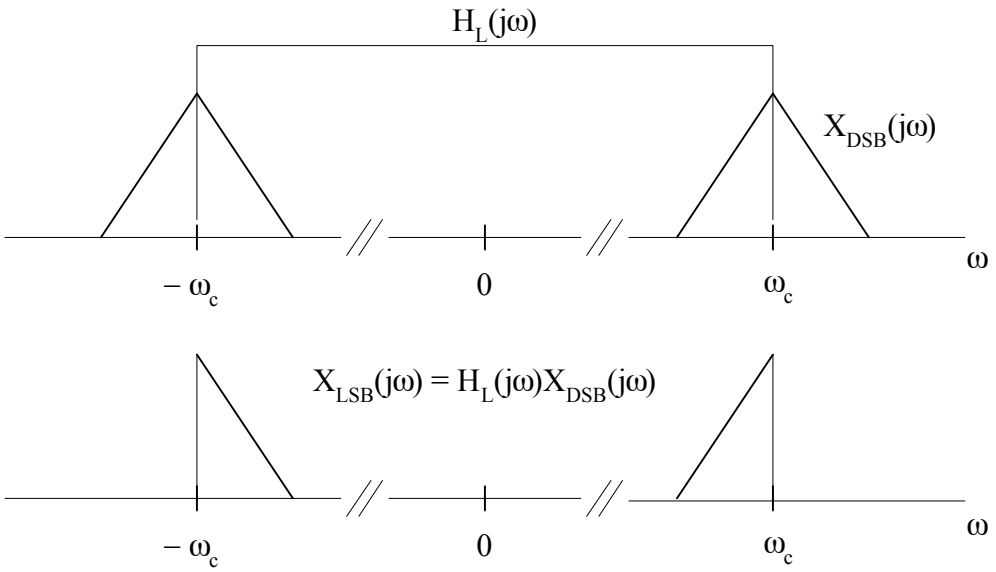


Figure 2-20: Development of lower sideband.

we write

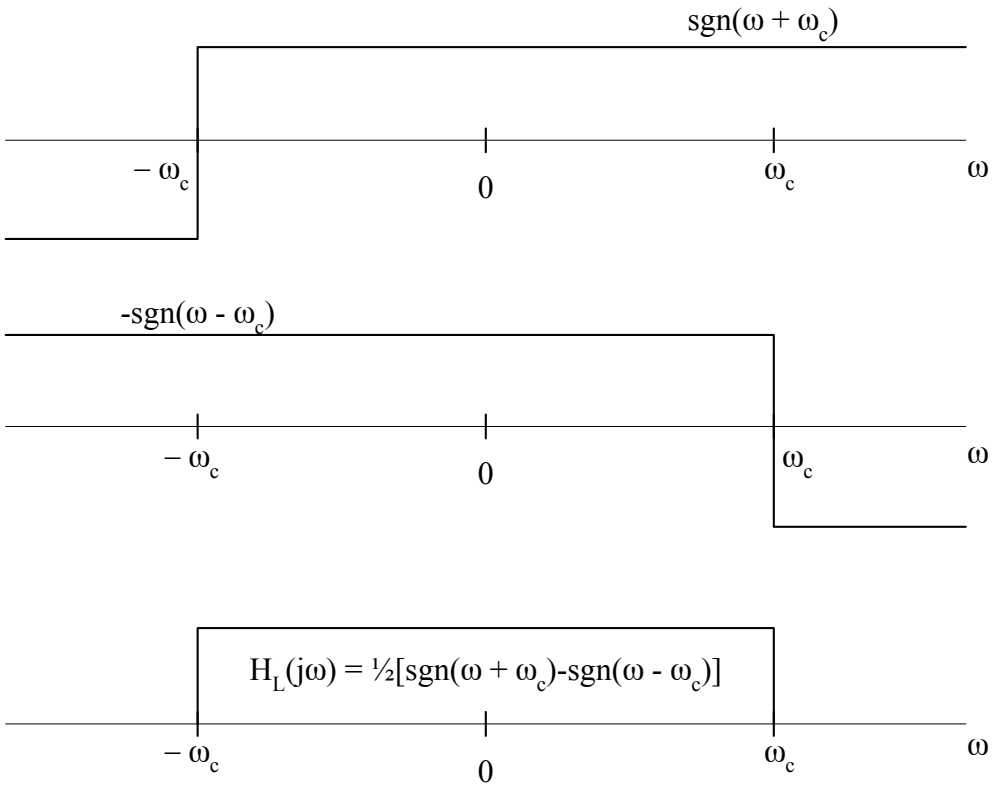


Figure 2-21: Construction of H_L for the generation of lower sideband.

$$X_{LSB}(j\omega) = X_{DSB}(j\omega)H_L(j\omega)$$

$$= \frac{A_c}{4}\left[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)\right]\left[\text{sgn}(\omega + \omega_c) - \text{sgn}(\omega - \omega_c)\right]$$

$$= \frac{A_c}{4}\left[M(j\omega + j\omega_c)\,\text{sgn}(\omega + \omega_c) + M(j\omega - j\omega_c)\,\text{sgn}(\omega + \omega_c)\right] \quad . \qquad (2\text{-}41)$$

$$-\frac{A_c}{4}\left[M(j\omega + j\omega_c)\,\text{sgn}(\omega - \omega_c) + M(j\omega - j\omega_c)\,\text{sgn}(\omega - \omega_c)\right]$$

Note that there are four terms on the right-hand-side of this last equation. The second and third terms combine to form $\frac{1}{4}A_c[M(\omega + \omega_c) + M(\omega - \omega_c)]$. The first and fourth terms combine to produce $\frac{1}{4}A_c[M(\omega + \omega_c)\text{sgn}(\omega + \omega_c) - M(\omega - \omega_c)\text{sgn}(\omega - \omega_c)]$. Hence, we can write

$$X_{LSB}(j\omega) = \frac{A_c}{4}\left[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)\right]$$

$$+\frac{A_c}{4}\left[M(j\omega + j\omega_c)\,\text{sgn}(\omega + \omega_c) - M(j\omega - j\omega_c)\,\text{sgn}(\omega - \omega_c)\right] \quad . \qquad (2\text{-}42)$$

The time-domain LSB signal is just the inverse transform of this last result. First, note that

$$\mathcal{F}^{-1}\left[\frac{A_c}{4}\left[M(j\omega + j\omega_c) + M(j\omega - j\omega_c)\right]\right] = \frac{A_c}{2}m(t)\cos\omega_c t\,. \qquad (2\text{-}43)$$

Further, note that

$$\mathcal{F}\left[\hat{m}(t)\right] = -j\,\text{sgn}(\omega)M(j\omega)$$

$$\mathcal{F}\left[\hat{m}(t)e^{\pm j\omega_c t}\right] = -j\,\text{sgn}(\omega \mp \omega_c)M(j\omega \mp j\omega_c) \qquad (2\text{-}44)$$

so that

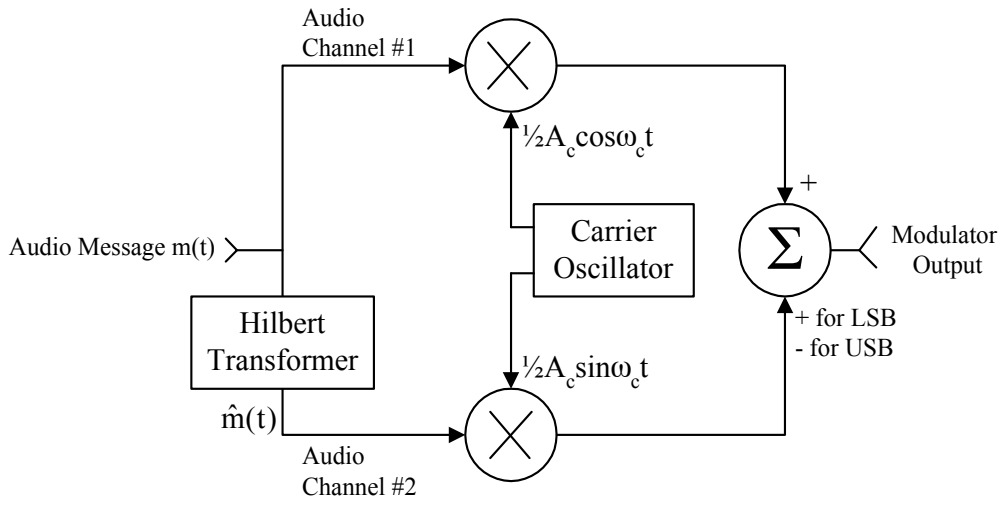


Figure 2-22: Simplified block diagram of a phasing method SSB generator.

$$\mathcal{F}^{-1}\left[\frac{A_c}{4}\left[M(j\omega+j\omega_c)\,\mathrm{sgn}(\omega+\omega_c)-M(j\omega-j\omega_c)\,\mathrm{sgn}(\omega-\omega_c)\right]\right]$$

$$=\frac{j}{4}A_c\left[\hat{m}(t)e^{-j\omega_c t}-\hat{m}(t)e^{+j\omega_c t}\right] \tag{2-45}$$

$$=\frac{A_c}{2}\hat{m}(t)\sin\omega_c t$$

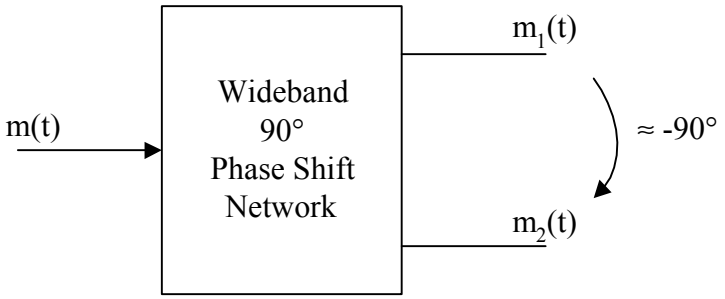
Finally, Equations (2-43) and (2-45) can be used to determine the inverse transform of (2-42); this leads to the desired result

$$x_{LSB}(t)=\mathcal{F}^{-1}\left[X_{LSB}(j\omega)\right]=\tfrac{1}{2}A_c m(t)\cos\omega_c t+\tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t\,, \tag{2-46}$$

a useful formula for LSB. A similar development leads to

$$x_{USB}(t)=\mathcal{F}^{-1}\left[X_{USB}(j\omega)\right]=\tfrac{1}{2}A_c m(t)\cos\omega_c t-\tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t \tag{2-47}$$

for upper sideband. The block diagram depicted by Figure 2-22 illustrates how to implement



$$\frac{|M_1(j\omega)|}{|M(j\omega)|} = \frac{|M_2(j\omega)|}{|M(j\omega)|} \approx \text{constant (over message bandwidth)}$$

$$M_2(j\omega) \approx -jM_1(j\omega) \text{ (over message bandwidth)}$$

Figure 2-23: Wide band 90° phase shift network.

Equations (2-46) and (2-47).

In a practical phasing SSB modulator, the Hilbert transformer would be replaced by a wide-band 90° phase shift network, a network/system that accepts m(t) as input and produces two nearly equal amplitude and nearly orthogonal messages to feed the balanced modulators. Figure 2-23 depicts a block diagram of such a system. The network's magnitude response from input to either output would be nearly constant over the message bandwidth of interest. Also, over the message bandwidth, there would be (nearly) a 90° differential phase shift between the two outputs. The phase relationship between the input and either output is not important. In a practical phasing modulator, the functionality described by Figure 2-23 would be used to produce the two base band audio signals that are fed to the balanced modulators depicted on Figure 2-22.

Alternate Development of SSB

As shown by Figure 2-24, let $M_p(j\omega)$ and $M_n(j\omega)$ denote the positive and negative, respectively, parts of the transform $M(j\omega)$ of message m(t). From inspection of Figure 2-24a, we can write

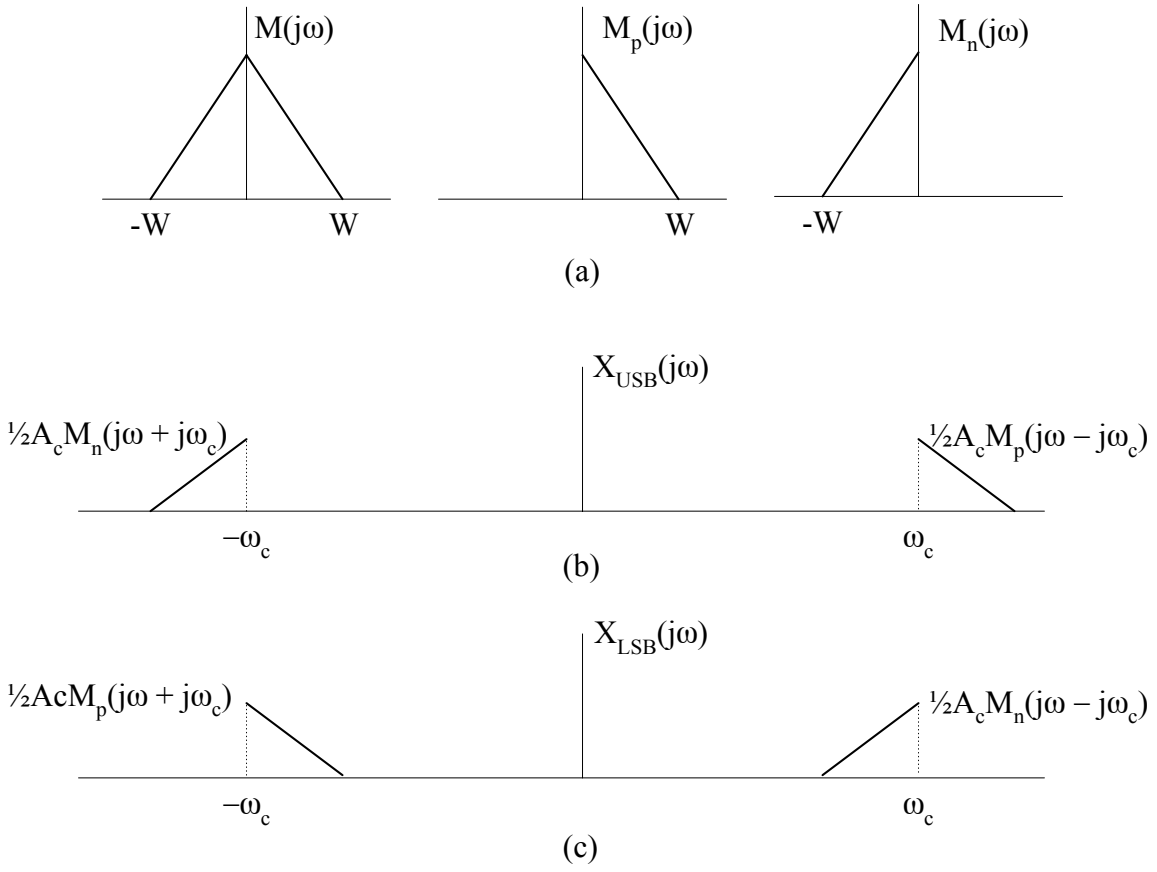


Figure 2-24: Alternate Development of SSB

$$\begin{aligned} M_p(j\omega) &= \tfrac{1}{2}\mathcal{F}\left[m(t) + j\hat{m}(t)\right] \\ M_n(j\omega) &= \tfrac{1}{2}\mathcal{F}\left[m(t) - j\hat{m}(t)\right] \end{aligned} \quad . \tag{2-48}$$

By definition, an USB signal has a frequency domain representation (see Fig. 2-24b)

$$X_{USB}(j\omega) = \tfrac{1}{2}A_c M_p(j\omega - j\omega_c) + \tfrac{1}{2}A_c M_n(j\omega + j\omega_c) \tag{2-49}$$

Take the inverse transform of (2-49) to obtain

$$x_{USB}(t) = \tfrac{1}{4}A_c\left[m(t) + j\hat{m}(t)\right]e^{j\omega_c t} + \tfrac{1}{4}A_c\left[m(t) - j\hat{m}(t)\right]e^{-j\omega_c t}$$

$$= \tfrac{1}{2}A_c m(t)\left[\frac{e^{j\omega_c t} + e^{-j\omega_c t}}{2}\right] - \tfrac{1}{2}A_c\hat{m}(t)\left[\frac{e^{j\omega_c t} - e^{-j\omega_c t}}{2j}\right], \qquad (2\text{-}50)$$

$$= \tfrac{1}{2}A_c m(t)\cos\omega_c t - \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t$$

the desired formula for an upper sideband modulated signal.

The formula for a lower sideband signal can be developed in a similar manner. Inspection of Figure 2-24c reveals

$$X_{LSB}(j\omega) = \tfrac{1}{2}A_c M_p(j\omega + j\omega_c) + \tfrac{1}{2}A_c M_n(j\omega - j\omega_c). \qquad (2\text{-}51)$$

The inverse transform of this signal is

$$x_{LSB}(t) = \tfrac{1}{4}A_c\left[m(t) + j\hat{m}(t)\right]e^{-j\omega_c t} + \tfrac{1}{4}A_c\left[m(t) - j\hat{m}(t)\right]e^{j\omega_c t}$$

$$= \tfrac{1}{2}A_c m(t)\left[\frac{e^{j\omega_c t} + e^{-j\omega_c t}}{2}\right] + \tfrac{1}{2}A_c\hat{m}(t)\left[\frac{e^{j\omega_c t} - e^{-j\omega_c t}}{2j}\right], \qquad (2\text{-}52)$$

$$= \tfrac{1}{2}A_c m(t)\cos\omega_c t + \tfrac{1}{2}A_c\hat{m}(t)\sin\omega_c t$$

the desired formula for a lower sideband modulated signal.

Demodulation of SSB - Product Detectors

As depicted by Figure 2-25, SSB can be demodulated by multiplying it by a phase coherent carrier and filtering the product by a low-pass filter. The product of the SSB signal and the coherent reference yields

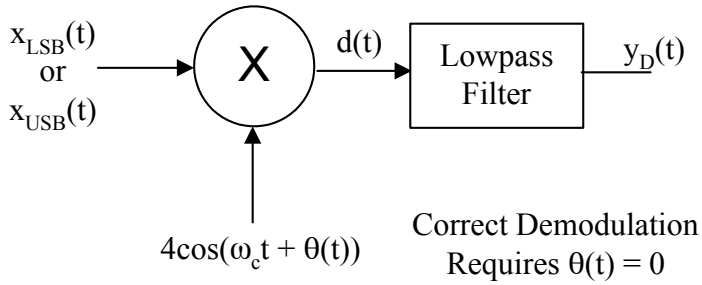


Figure 2-25: SSB demodulator.

$$\begin{aligned} d(t) &= \left[\tfrac{1}{2}A_c m(t)\cos\omega_c t \pm \tfrac{1}{2}A_c \hat{m}(t)\sin\omega_c t\right]\left[4\cos(\omega_c t + \theta(t))\right] \\ &= A_c\left[m(t)\cos\theta(t) + m(t)\cos(2\omega_c t + \theta(t)) \mp \hat{m}(t)\sin\theta(t) \pm \hat{m}(t)\sin(2\omega_c t + \theta(t))\right] \end{aligned} \tag{2-53}$$

Low-pass filtering product d(t) produces

$$y_d(t) = A_c m(t)\cos\theta(t) \mp A_c \hat{m}(t)\sin\theta(t)\,. \tag{2-54}$$

For θ(t) = 0 we obtain the desired result. Depending on the application, if θ(t) ≠ 0, the term $\mp A_c \hat{m}(t)\sin\theta(t)$ *may* introduce serious distortion. In the case of human speech, it is possible to understand the message even if small frequency errors are present (dθ/dt ≠ 0). For speech, dθ/dt is adjusted "manually" by a listener who adjusts a radio tuning dial until he/she can "copy" the transmission. (The ear/brain is not sensitive to static phase errors; it is sufficient to make dθ/dt small.)

Demodulation of SSB - Carrier Reinsertion/Envelope Detection

SSB can be demodulated by the method illustrated by Figure 2-26. The output of the summer operation is

$$\begin{aligned} e(t) &= \tfrac{1}{2}A_c m(t)\cos\omega_c t \pm \tfrac{1}{2}A_c \hat{m}(t)\sin\omega_c t + K\cos\omega_c t \\ &= \left[\tfrac{1}{2}A_c m(t) + K\right]\cos\omega_c t \pm \tfrac{1}{2}A_c \hat{m}(t)\sin\omega_c t \end{aligned} \tag{2-55}$$

In terms of magnitude and phase, Equation (2-55) can be written as

$$e(t) = R(t)\cos(\omega_c t + \theta(t)), \tag{2-56}$$

where

$$R(t) = \sqrt{[\tfrac{1}{2}A_c m(t) + K]^2 + [\tfrac{1}{2}A_c \hat{m}(t)]^2}$$
$$\theta(t) = -\mathrm{Tan}^{-1}\left[\frac{\tfrac{1}{2}A_c \hat{m}(t)}{\tfrac{1}{2}A_c m(t) + K}\right] \tag{2-57}$$

are the envelope and phase, respectively, of e(t). Now, the output of the demodulator depicted by Figure 2-26 is given by

$$y_d(t) = \sqrt{[\tfrac{1}{2}A_c m(t) + K]^2 + [\tfrac{1}{2}A_c \hat{m}(t)]^2}\,. \tag{2-58}$$

If constant K is chosen large enough so that

$$[\tfrac{1}{2}A_c m(t) + K]^2 >> [\tfrac{1}{2}A_c \hat{m}(t)]^2, \tag{2-59}$$

we can approximated $y_d(t)$ as

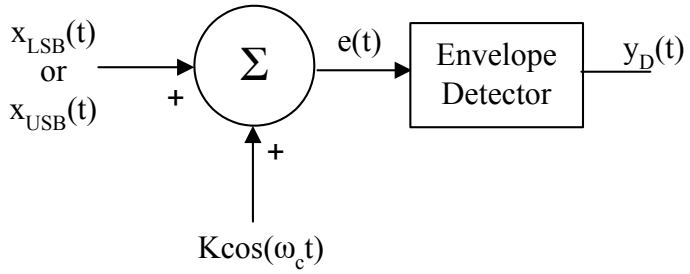


Figure 2-26: Demodulation of SSB using carrier reinsertion

$$y_d(t) \approx \tfrac{1}{2} A_c m(t) + K, \tag{2-60}$$

a result that contains the desired message.

Transmitted Power in SSB Waveform

The SSB signal is represented as

$$x_{SSB}(t) = \tfrac{1}{2} A_c m(t) \cos \omega_c t \pm \tfrac{1}{2} A_c \hat{m}(t) \sin \omega_c t. \tag{2-61}$$

The *instantaneous* transmitted power is $x_{SSB}^2(t)$. The *average* transmitted power is

$$\begin{aligned} P_{AVG} &= \left\langle x_{SSB}^2(t) \right\rangle = \tfrac{1}{4} A_c^2 \left[\left\langle m^2(t) \cos^2 \omega_c t \right\rangle + \left\langle \hat{m}^2(t) \sin^2 \omega_c t \right\rangle \right] \\ &= \tfrac{1}{4} A_c^2 \left[\frac{1}{2} \left\langle m^2(t) \right\rangle + \frac{1}{2} \left\langle \hat{m}^2(t) \right\rangle \right] \\ &= \tfrac{1}{4} A_c^2 \left\langle m^2(t) \right\rangle \end{aligned} . \tag{2-62}$$

The *peak-envelope-power* (PEP) is of interest. The SSB signal can be represented as

$$x_{SSB}(t) = \frac{A_c}{2} \sqrt{m^2(t) + \hat{m}^2(t)} \; \cos(\omega_c t \mp \tan^{-1}(\hat{m}/m)). \tag{2-63}$$

The envelope and phase are slowly varying relative to the carrier $\cos\omega_c t$. Over every RF cycle the envelope and phase are approximately constant. The peak envelope power (PEP) is the instantaneous power averaged over the RF cycle having the greatest amplitude. Hence, the PEP power is

$$P_{pep} = \frac{1}{8} A_c^2 \max_t \left[m^2(t) + \hat{m}^2(t) \right] \tag{2-64}$$

For $m(t) = \cos\omega_m t$, we have $P_{avg} = P_{pep} = A_c^2/8$. For m(t) a human voice, a general *rule of thumb* is that P_{pep} is between two and three times P_{avg}.

Angle Modulation

The general angle modulated signal is described by

$$x_c(t) = A_c \cos[\omega_c t + \phi(t)], \tag{2-65}$$

where A_c and ω_c are constants, and angle ϕ depends on the message m(t). Unlike the modulation methods discussed so far, an angle-modulated signal is a *nonlinear* function of the message. Phase modulation and frequency modulation are two forms of commonly-used angle modulation.

The *instantaneous phase* θ_t of signal x_c is given by

$$\theta_t = \omega_c t + \phi(t). \tag{2-66}$$

Often, angle ϕ is called the *instantaneous phase deviation*.

The *instantaneous frequency* of x_c is

$$\frac{d\theta_t}{dt} = \omega_c + \frac{d\phi}{dt}. \tag{2-67}$$

The quantity $d\phi/dt$ is called the *instantaneous frequency deviation*. The *peak frequency deviation*

$$\omega_{peak} \equiv \max_t \left|\frac{d\phi}{dt}\right| \tag{2-68}$$

is an important parameter in practical FM system design.

The two basic types of angle modulation are 1) *phase modulation* (PM) and 2) *frequency*

modulation (FM). We will consider both; however, we will place most of our emphasis on FM.

In PM the phase

$$\phi(t) = K_p m(t) \tag{2-69}$$

is proportional to message m(t). Constant $K_p > 0$ is the *modulation index* for PM, and it has units of radians/volt. The PM signal is

$$x_c(t) = A_c \cos[\omega_c t + K_p m(t)]. \tag{2-70}$$

In FM, the frequency deviation is proportional to the message so that

$$\frac{d\phi(t)}{dt} = K_f m(t), \tag{2-71}$$

or

$$\phi(t) = K_f \int m(x)dx. \tag{2-72}$$

Positive K_f is the *frequency deviation* constant, expressed in radians/second-volt. Sometimes, frequency deviation is specified in Hz. In this case, the relevant constant is f_d where

$$K_f = 2\pi f_d, \tag{2-73}$$

and f_d is expressed in Hz/volt. The FM signal is

$$x_c(t) = A_c \cos[\omega_c t + K_f \int^t m(x)dx]. \tag{2-74}$$

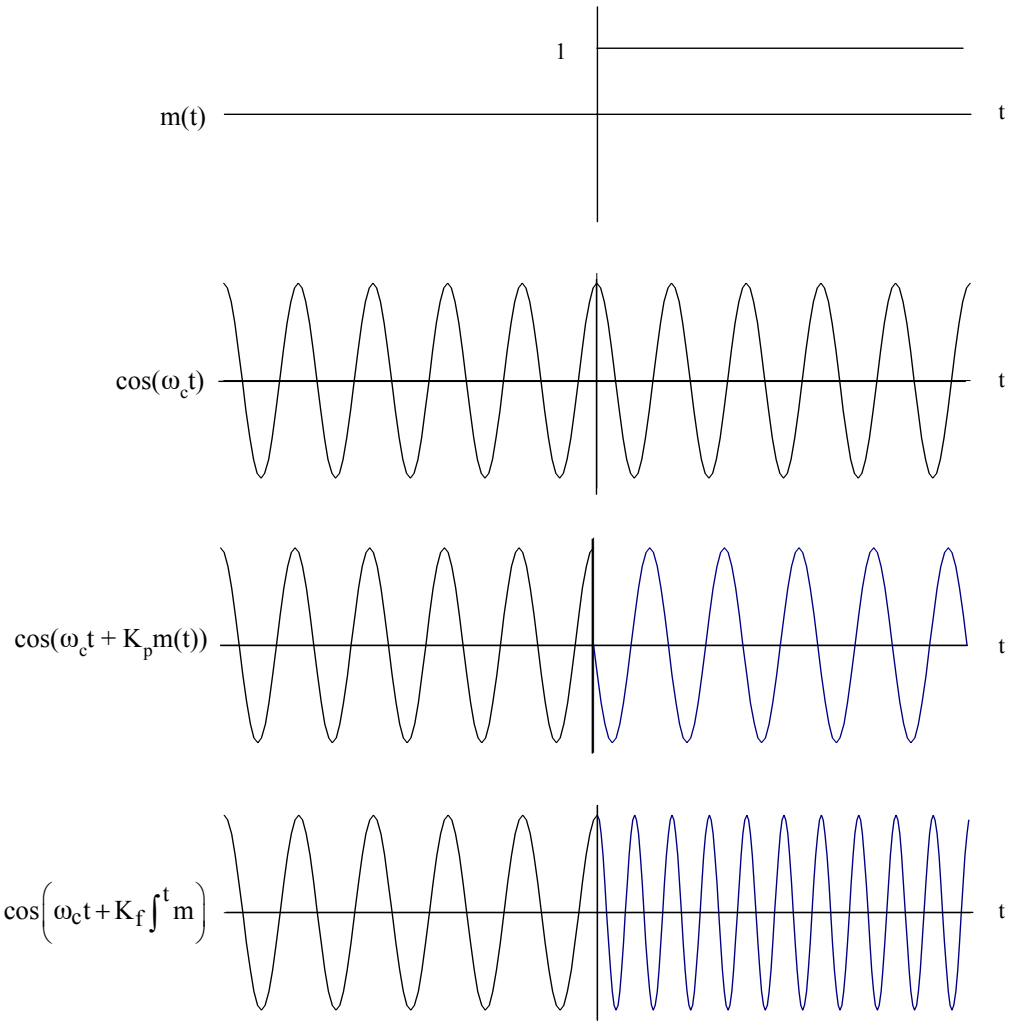


Figure 2-27: Message m(t), unmodulated carrier $\cos(\omega_c t)$, phase modulated carrier $\cos(\omega_c t + K_p m(t))$ (with $K_p = \pi/2$) and frequency modulated carrier $\cos(\omega_c t + K_f \int m)$ ($K_f \approx \omega_c$ in magnitude).

Figure 2-27 depicts a unit step message, an unmodulated carrier, a phase modulated carrier and a frequency modulated carrier.

Note that the transmitted power is constant and independent of message m(t). The transmitted power is

$$\left\langle x_c^2(t) \right\rangle = \left\langle A_c^2 \cos^2[\omega_c t + \phi(t)] \right\rangle = \left\langle A_c^2 \frac{1}{2}(1 + \cos[2\omega_c t + 2\phi(t)]) \right\rangle \approx A_c^2 \frac{1}{2}, \tag{2-75}$$

the approximation (a very good one in practice!) due to the fact that ϕ varies slowly relative to $\cos 2\omega_c t$. Unlike AM, DSB and SSB, the angle modulated transmitter duty cycle is 100%. Hence, an angle-modulated transmitter must have a "huskier" power supply, and more conservatively rated components, than an equivalent-power AM/DSB/SSB transmitter.

Narrow Band Angle Modulation

The angle modulation is said to be *narrow band* if $|\phi(t)| << 1$ for all time. For this case, we can write the transmitted signal as

$$\begin{aligned} x_c(t) &= A_c \cos[\omega_c t + \phi(t)] = A_c \cos\phi(t)\cos\omega_c t - A_c \sin\phi(t)\sin\omega_c t \\ &\approx A_c \cos\omega_c t - A_c\phi(t)\sin\omega_c t \end{aligned} \tag{2-76}$$

$$\begin{aligned} \phi(t) &= K_p m(t), \quad \text{for PM} \\ \phi(t) &= K_f \int m(x)dx, \quad \text{for FM} \end{aligned} . \tag{2-77}$$

In this last result, we have use the fact that $\cos\phi \approx 1$ and $\sin\phi \approx \phi$. Narrow band angle modulation is used by all municipal services (*i.e.*, police, fire, city, state personnel, etc.), the military, amateur radio operators and many other groups.

Consider taking the Fourier transform of the narrow band angle modulate signal. The transform is

$$\begin{aligned} \mathcal{F}[x_c(t)] &= \mathcal{F}[A_c \cos\omega_c t - A_c\phi(t)\sin\omega_c t] \\ &= \pi A_c[\delta(\omega+\omega_c)+\delta(\omega-\omega_c)] + j\frac{A_c}{2}[\Phi(\omega-\omega_c)-\Phi(\omega+\omega_c)] \end{aligned} , \tag{2-78}$$

where

$$\Phi(\omega) = \mathcal{F}\left[\phi(t)\right] \quad = K_p M(j\omega), \qquad \text{for PM}$$
$$= -j\left[\frac{K_f}{\omega}\right] M(j\omega), \quad \text{for FM} \tag{2-79}$$

From this we conclude that if m(t) has message bandwidth W << ω_c (the usual case in application), then $x_c(t)$ will be a band-pass signal with bandwidth 2W. In the amplitude spectrum of FM, note that low frequencies in $M(j\omega) = \mathcal{F}[m(t)]$ are emphasized more than high frequencies, because of the factor K_f/ω in $\Phi(j\omega)$. Figure 2-28 depicts a message amplitude spectrum, the amplitude spectrum of the resulting narrow band PM signal, and the amplitude spectrum of the resulting narrow band FM signal.

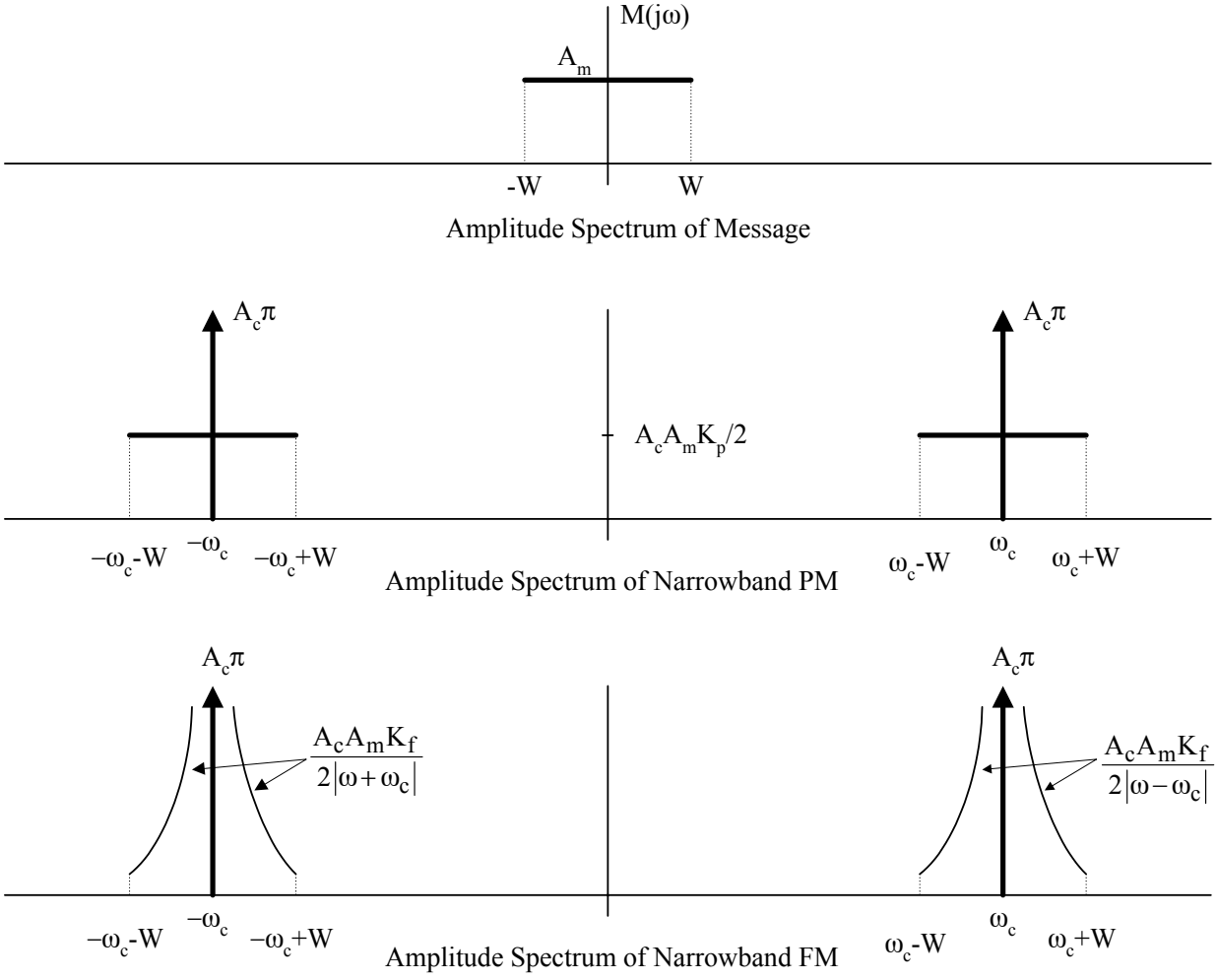


Figure 2-28: a) Message amplitude spectrum, b) amplitude spectrum of narrow band PM, c) amplitude spectrum of narrow band FM.

Tone Modulation

Suppose m(t) is a simple tone. To keep things simple, assume that

$$\begin{aligned} m(t) &= A_m \sin \omega_m t \quad \text{for PM} \\ &= A_m \cos \omega_m t \quad \text{for FM} \end{aligned} . \tag{2-80}$$

For this sinusoidal message we have

$$\begin{aligned} \phi(t) &= K_p m(t) &&= K_p A_m \sin \omega_m t \quad \text{for PM} \\ &= K_f \int m(x)dx &&= K_f \frac{A_m}{\omega_m} \sin \omega_m t \quad \text{for FM} \end{aligned} . \tag{2-81}$$

To summarize this we write

$$\phi(t) = \beta \sin \omega_m t , \tag{2-82}$$

where

$$\begin{aligned} \beta &= K_p A_m \quad \text{for PM} \\ &= K_f \frac{A_m}{\omega_m} \quad \text{for FM}. \end{aligned} \tag{2-83}$$

Constant β is known as the *modulation index* for tone modulation (symbol β is used only when tone modulation is under consideration). Narrow band modulation requires that $\beta << 1$. Finally, note that $K_f A_m$ is the *peak frequency deviation* for narrow band FM. For police, municipal services and for amateur radio, narrow band FM utilizes a peak frequency deviation $K_f A_m$ of around 5 kHz.

Phasor Diagram for Narrow Band Tone FM/PM

For narrow band tone modulation with $\phi = \beta \sin\omega_m t$ we have

$$\begin{aligned} x_c(t) &= A_c \cos\omega_c t - A_c(\beta \sin\omega_m t)\sin\omega_c t \\ &= A_c \cos\omega_c t - \frac{A_c\beta}{2}\cos[(\omega_c - \omega_m)t] + \frac{A_c\beta}{2}\cos[(\omega_c + \omega_m)t] \end{aligned} \quad (2\text{-}84)$$

A phasor diagram for x_c is given by Figure 2-29. The carrier is the reference, so it remains stationary. The component at $\omega_c + \omega_m$ (the USB) is increasing in angle relative to the carrier and is drawn as rotating ω_m radians/second in a counter clockwise direction. The component at $\omega_c - \omega_m$ (the LSB) is decreasing in angle relative to the carrier and is drawn as rotating ω_m radians/second in a clockwise direction.

Note that for narrow-band angle modulation, x_c contains a component that is orthogonal to the carrier. This contrasts with the case for AM. For AM with tone modulation we have

$$\begin{aligned} x_c(t) &= A_c'[1 + a\cos\omega_m t]\cos\omega_c t \\ &= A_c'\cos\omega_c t + \frac{aA_c'}{2}\left[\cos[(\omega_c + \omega_m)t + \cos[(\omega_c - \omega_m)t]\right] \end{aligned} \quad (2\text{-}85)$$

The phasor diagram for AM is given by Figure 2-30. Note that x_c contains no component orthogonal to the carrier. For AM, *the transmitted signal is always in phase with the carrier; the*

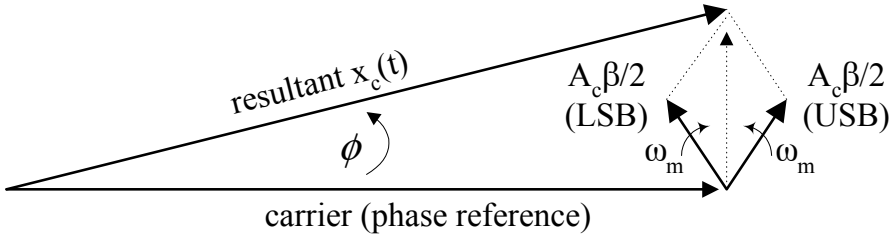


Figure 2-29: Phase diagram for narrow-band angle modulation. Note that x_c has a component that is orthogonal to the carrier.

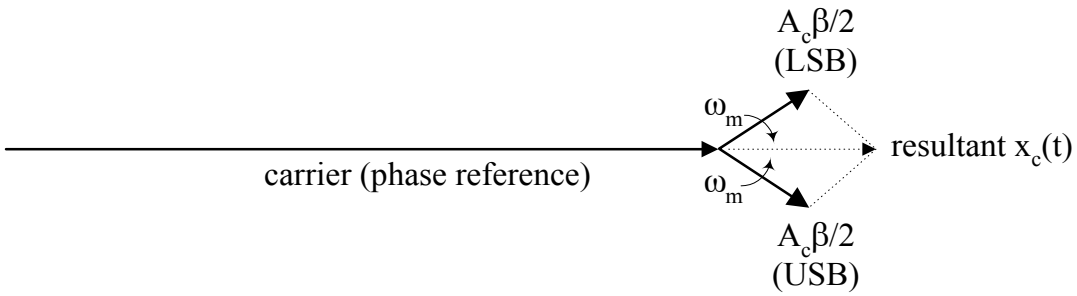


Figure 2-30: Phasor diagram for amplitude modulation. Note that x_c has no component orthogonal to the carrier.

information is contained in the amplitude and not in the phase.

Wide Band Tone Modulation (Large β)

The tone-modulated signal (either FM or PM) can be written as

$$x_c(t) = A_c \cos(\omega_c t + \beta \sin \omega_m t) = A_c \, \mathrm{Re}\left[e^{j\omega_c t} e^{j\beta \sin \omega_m t} \right]. \tag{2-86}$$

The $e^{j\beta \sin \omega_m t}$ term is periodic in t. It can be expanded into an exponential Fourier series. The series expansion has the form

$$e^{j\beta \sin \omega_m t} = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_m t} \tag{2-87}$$

where

$$c_n = \frac{\omega_m}{2\pi} \int_{-\pi/\omega_m}^{\pi/\omega_m} \left[e^{j\beta \sin \omega_m t} \right] e^{-jn\omega_m t} dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j(nx - \beta \sin x)} dx \tag{2-88}$$

However, this last integral is the well-known *Bessel function of the first kind of order n with real argument* β. Hence, the Fourier series can be written as

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t}$$

$$J_n(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j(nx-\beta\sin x)}dx, \tag{2-89}$$

a version of the well-known Jacobi-Anger formula. Use this expansion in the formula for x_c to obtain

$$x_c(t) = A_c\,\mathrm{Re}\left[e^{j\omega_c t}\sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t}\right] = A_c\sum_{n=-\infty}^{\infty} J_n(\beta)\cos[(\omega_c + n\omega_m)t]. \tag{2-90}$$

Note that x_c has a carrier with amplitude $J_0(\beta)$, and it has an infinite number of sidebands, in theory. The n^{th} sideband pair has amplitude $|J_n(\beta)|$. Figure 2-31 depicts an example of an angle-modulated signal containing a carrier and four pairs of modulation sidebands.

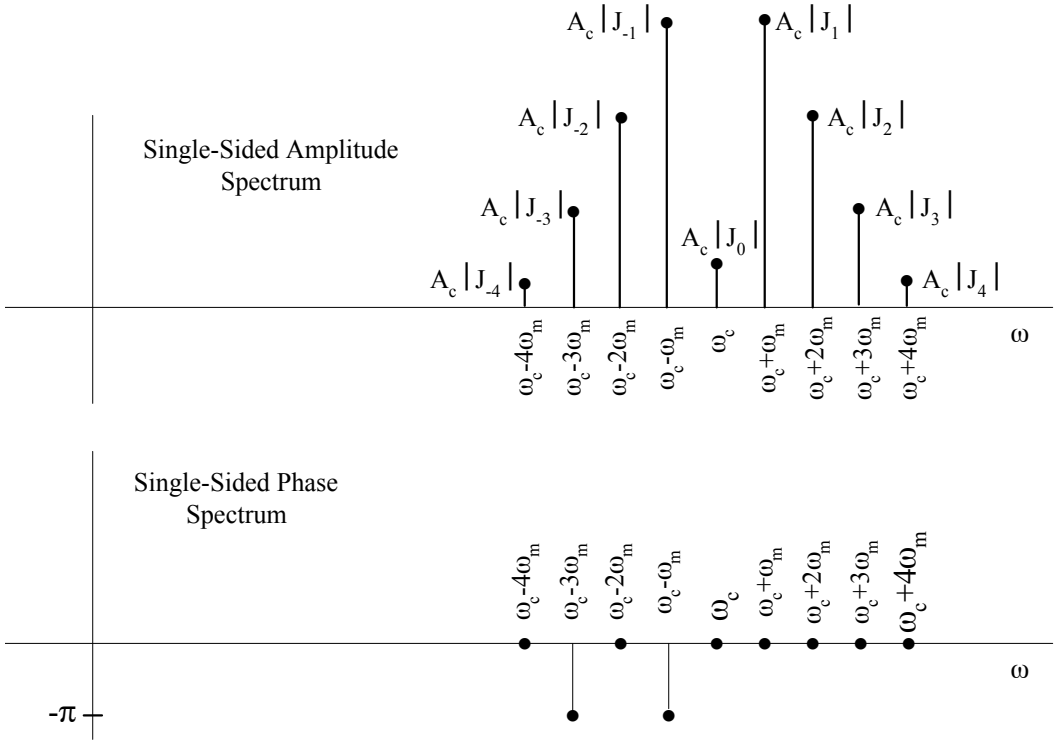


Figure 2-31: Single-sided amplitude and phase spectrum.

Tables and computer programs can provide values of $J_n(\beta)$ for $n \geq 0$. For negative integers n, we can use

$$\begin{aligned} J_{-n}(\beta) &= J_n(\beta), \quad n \text{ even} \\ J_{-n}(\beta) &= -J_n(\beta), \quad n \text{ odd} \end{aligned} \qquad (2\text{-}91)$$

The integer-order Bessel functions satisfy a recursion relationship. This relationship is

$$J_{n-1}(\beta) + J_{n+1}(\beta) = \left(\frac{2n}{\beta}\right) J_n(\beta). \qquad (2\text{-}92)$$

This recursion can be used in a forward direction; given J_n and J_{n-1}, it can be used to compute J_{n+1}. However, the forward recursion is *numerically unstable*. Any initial error will grow rapidly and, after a few iterations of the recursion, make the results unusable. On the other hand, the backward recursion is *numerically stable*. Given J_{n+1} and J_n, the recursion can be used to calculate a very accurate value for J_{n-1}, and the backward recursion can be repeated as often as desired without fear of numerical instability. Figure 2-32 depicts graphs of a few low-order Bessel functions.

A few properties of Bessel functions are of interest.

1. For $\beta << 1$, $J_0(\beta)$ and $J_1(\beta)$ strongly dominate $J_k(\beta)$, $k \geq 2$, and (2-90) can be written as

$$\begin{aligned} x_c(t) &= A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[(\omega_c + n\omega_m)t] \\ &\approx A_c J_0(\beta) \cos \omega_c t + A_c J_1(\beta) \cos(\omega_c + \omega_m)t + A_c J_{-1}(\beta) \cos(\omega_c - \omega_m)t \end{aligned} \qquad (2\text{-}93)$$

This approximation represents the narrow band modulation case.

2. For fixed n, $J_n(\beta)$ oscillates with increasing β. However, the amplitude of oscillation decays

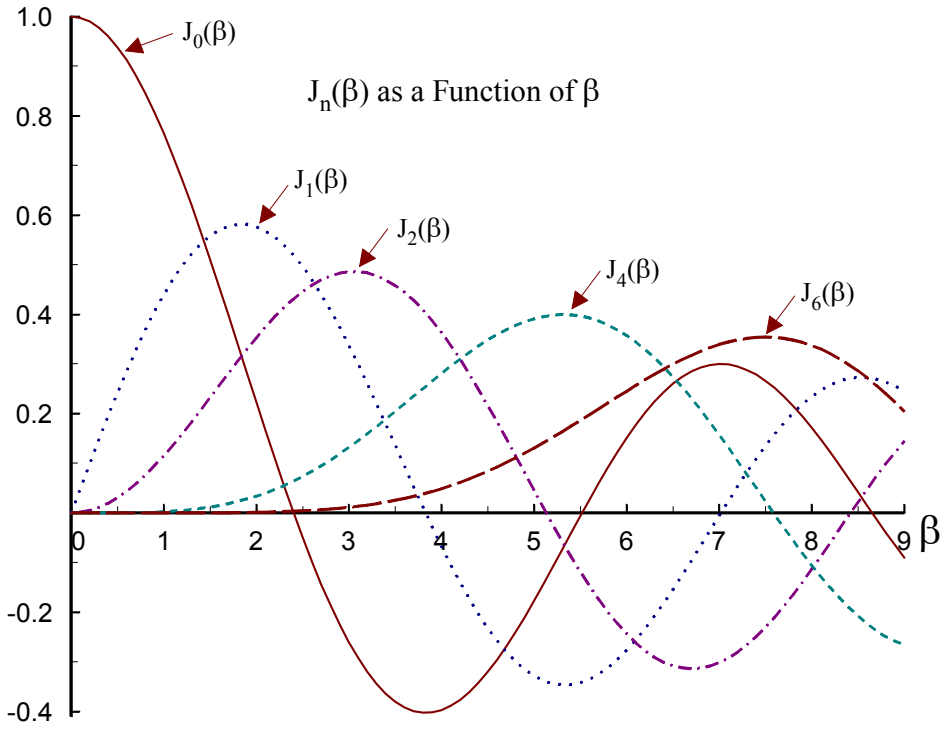


Figure 2-32: Graphs of some Bessel functions.

quickly with large β.

3. For fixed β, the maximum value of $J_n(\beta)$ decreases with increasing n. In fact, for sufficiently large n, we have the asymptotic relationship

$$J_n(\beta) \approx \frac{\beta^n}{2^n n!}. \tag{2-94}$$

4. $J_0(\beta)$ is zero for β = 2.4048, 5.5201, 8.6537, … . The signal $x_c(t)$ will not contain a carrier component for these values of β.

Observation #4 provides a practical method of measuring K_f for an FM transmitter. Feed the output of an audio generator to the FM transmitter under test. Observe the output of the transmitter on a spectrum analyzer. Use any convenient audio frequency (say 1KHz). Increase the audio level until the carrier vanishes and then calculate K_f.

Bandwidth of Angle Modulated Signal - Tone Message Case

Strictly speaking, a signal that is angle modulated with a tone message contains power over an infinite bandwidth of frequencies. However, only a finite number of sideband pairs have significant power, so the bandwidth of the signal is finite, for all practical purposes. The *power ratio* P_r is the ratio of the power contained in the carrier and first k sideband pairs to the total power. For the case of a tone message, we write the power ratio as

$$P_r \equiv \frac{\frac{1}{2}\sum_{n=-k}^{k} A_c^2 J_n^2(\beta)}{\frac{1}{2}A_c^2} = J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta) . \tag{2-95}$$

For a particular application, bandwidth can be determined by 1) defining an acceptable P_r, 2) solving for the required value of k (using a table of Bessel functions), and 3) computing bandwidth $B = 2k\omega_m$. The results of this procedure are given in Table 1 below. The value of k for $P_r = .7$ is indicated by a single underscore in the table; the value of k for $P_r = .98$ is indicated by a double underscore. For example, for $\beta = 2$, a double under score occurs in the fourth row (in the $\beta = 2$ column) of the table. For this case, a carrier and three sideband pairs contain 98% of the total transmitted power; the transmitted bandwidth is $B = 6\omega_m$.

For $P_r = .98$ and integer values of β, it is noted that k is equal to $1 + \beta$. Hence, for $P_r = .98$ and integer β we approximate

$$\text{Bandwidth} \approx 2(\beta + 1)\omega_m , \tag{2-96}$$

where ω_m is the frequency of the tone message. For $P_r = .98$ and non-integer β, this formula gives a useful approximation to bandwidth.

Bandwidth as a Function of ω_m

For FM, note that $\beta = A_m K_f/\omega_m$, the peak frequency deviation $A_m K_f$ divided by the tone

k	β = .1	β = .2	β = .5	β = 1	β = 2	β = 5	β = 8	β = 10
0	.997	.990	.938	.765	.224	-.178	.172	-.246
1	.050	.100	.242	.440	.577	-.328	.235	.043
2	.001	.005	.031	.115	.353	.047	-.113	.255
3				.020	.129	.365	-.291	.058
4				.002	.034	.391	-.105	-.220
5					.007	.261	.186	-.234
6					.001	.131	.338	-.014
7						.053	.321	.217
8						.018	.223	.318
9						.006	.126	.292
10						.001	.061	.207
11							.026	.123

Table 1: Table of $J_k(\beta)$, $0 \le k \le 11$.

frequency ω_m. As ω_m decreases, β increases and so does the number of significant sidebands. However, with decreasing ω_m, the sidebands come closer together in frequency, and the required bandwidth approaches a constant (this is predicted by the formula: for small ω_m, we have $\text{Bandwidth} \approx 2(\beta+1)\omega_m \approx 2A_mK_f$).

Bandwidth for Non-sinusoidal Modulation

For arbitrary m(t) the transmission bandwidth is difficult to define and compute (instead of signal bandwidth, it is better to think in terms of how much bandwidth is required to transmit the signal without excessive distortion). Hence, we resort to a general "rule of thumb" estimate. For arbitrary m(t) we define the *deviation ratio*

$$D \equiv \frac{\text{peak frequency deviation}}{\text{one-sided bandwidth of message m(t)}}. \qquad (2\text{-}97)$$

For FM, this definition is equivalent to

$$D \equiv \frac{K_f \max_t |m(t)|}{\text{one-sided bandwidth of message } m(t)}. \tag{2-98}$$

Deviation ratio D plays the same role for non-sinusoidal modulation as the modulation index β plays for sinusoidal messages. Hence the transmission bandwidth can be approximate by

$$\text{Bandwidth} \approx 2(D+1)W, \tag{2-99}$$

where W is the one-sided message bandwidth. This last approximation is known as *Carson's Rule*. (Also, in 1915, Carson patented SSB.) Experimental data shows that Carson's rule provides good results for $D << 1$ (bandwidth = 2W, which is the narrow band modulation case) *and* for $D >> 1$. (Carson's rule works well for $D > 5$.)

Commercial FM Broadcasting

Commercial FM broadcasting uses wideband FM in the "channelized" FM band that extends from 88Mhz to 108Mhz in the United States. The Federal Communication Commission (FCC) in the U.S. assigns stations to carrier frequencies that are spaced 200KHz apart in the FM band. The peak frequency deviation is limited to 75KHz. The message bandwidth is limited to 15KHz. Hence, for commercial FM

$$D = \frac{75\text{KHz}}{15\text{KHz}} = 5. \tag{2-100}$$

Carson's rule gives a bandwidth of 2(D+1)W = 180KHz. Experimental data shows this to be a little low; a bandwidth of 200KHz is closer to reality.

Figure 2-33: Edward H. Armstrong

Edward H. Armstrong – The Father of Wideband FM

In the early days of commercial broadcasting, amplitude modulation was "king". However, AM is plagued by static caused by atmospheric electrical activity, especially during summertime thunderstorms. Many researchers sought methods for reducing the effect of static on commercial broadcasting. Prior to the mid 1930's, a generally accepted axiom was that you could only reduce radio static by decreasing transmission/reception bandwidth. Because of this, many leading researchers of the time thought (and wrote in scholarly journals) that wideband FM was "worthless".

This belief was not held by Edward H. Armstrong, a leading, if somewhat eccentric, researcher of the time. About 1934, Armstrong showed that, under certain specified conditions, you could "trade" FM transmission/reception bandwidth for FM detector output signal-to-noise ratio. That is, within limits and under specified conditions, you could reduce the influence of

radio noise by *increasing* transmission/reception bandwidth! At the time, this result by Armstrong was hard to "digest", especially by those in the AM broadcasting industry!!

The discovery of the benefits of wide-band FM was not Armstrong's first major contribution to the art/science of radio. While in high school, Armstrong was an avid experimenter/hobbyist in the nascent field of radio (he was a "ham radio" operator). While a junior electrical engineering student at Columbia University (about 1913), Armstrong invented the regenerative detector, the use of which made radio receivers orders of magnitude more sensitive (this was a big discovery at the time). He went on to invent the super regenerative detector. While serving in the US Army during WWI (in Paris France), Armstrong invented the superheterodyne receiver, the dominant receiver architecture used today. So, by the time of his wide-band FM work, Armstrong was a well-known inventor in the field of radio.

To help establish his belief in the superiority of FM for commercial broadcasting, Armstrong started, in 1938, one of the first commercial FM stations at Alpine, New Jersey, across the Hudson River from Yonkers, NY. (His massive and tall transmission tower still stands today!). In addition, throughout New England, he established the *Yankee Network* of FM stations.

After WWII, commercial FM broadcasting was "pushed" (by the FCC in response to a "request" by David Sarnoff, president of RCA and a major backer of TV broadcasting) from its existing 42-to-50Mhz frequency allocation to its current allocation, 88-to-108 MHz. This made existing FM equipment obsolete, and it set FM broadcasting back by many years.

During this time, Armstrong was engaged in many patent infringement law suites, trying to protect his several patents. This long litigation process severely stressed Armstrong, according to his many friends. Unfortunately, health, marital, legal and financial problems drove Armstrong, in 1954, to take his life by jumping out of his River House apartment window in Manhattan.

It was not until the 1970's that FM broadcasting surpassed AM broadcasting in the numbers of listeners and advertising revenue. However, today, FM commercial broadcasting is

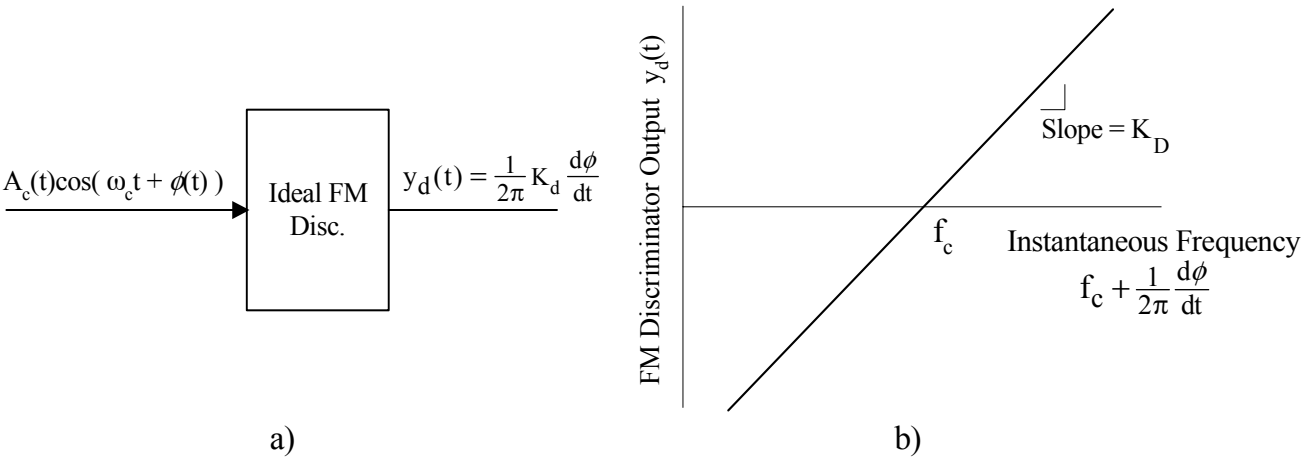


Figure 2-34: Ideal FM discriminator. The constant f_c is the *discriminator center frequency*, and K_D is the *discriminator gain constant*. Constants f_c and K_D completely characterize the ideal FM discriminator.

"king", trumping AM when it comes to broadcasting high-fidelity music and entertainment (most AM stations have survived by changing to an all sports/news/talk format).

Ideal FM Discriminator

The *ideal FM discriminator* yields an output that is proportional to the frequency deviation of the discriminator input, as illustrated by Figure 2-34. If the input to the ideal FM discriminator is

$$x_c(t) = A_c(t)\cos[\omega_c t + \phi(t)], \quad A_c(t) > 0 \text{ for all t}, \tag{2-101}$$

then the output is

$$y_d(t) = \frac{1}{2\pi} K_D \frac{d\phi}{dt}, \tag{2-102}$$

where K_D is the *discriminator gain constant* (units are volts/Hz). Note that the ideal discriminator output is unaffected by changes in the amplitude $A_c(t)$ of input $x_c(t)$ (we assume that $A_c(t) > 0$ for all t). Output $y_d(t)$ is not influenced by any envelope modulation imposed on signal $x_c(t)$ (the ability to reject amplitude modulation is a measure of "goodness" when

comparing discriminator circuits/designs).

Since $\phi = 2\pi f_d \int^t m(x)dx$ for FM, we have

$$y_d(t) = \frac{1}{2\pi} K_D \frac{d\phi}{dt} = K_D f_d m(t) . \quad (2\text{-}103)$$

As shown by Figure 2-34, the ideal FM discriminator has a linear frequency to voltage transfer function that passes through zero at the carrier frequency $\omega_c = 2\pi f_c$. Frequency f_c is called the *center frequency* of the discriminator. Center frequency f_c and discriminator gain K_D completely characterize the ideal FM discriminator.

Ideal PM Discriminator

The *ideal phase modulation discriminator* has a linear phase to voltage characteristic. If the PM input is

$$x_c(t) = A_c(t)\cos[\omega_c t + \phi(t)], \quad A_c(t) > 0 \text{ for all t, , } \quad (2\text{-}104)$$

the output is

$$y_d(t) = K_D \, \phi(t) , \quad (2\text{-}105)$$

where *discriminator gain constant* K_D has units of volts/radian. Frequency f_c is called the *center frequency* of the ideal PM discriminator. Center frequency f_c and discriminator gain K_D completely characterize the ideal PM discriminator.

Band-Pass Limiter

As stated above, an ideal discriminator (either FM or PM) does not respond to amplitude modulation on its input signal (that may be present due to a poorly designed/operated transmitter or a time-varying communication channel). However, this is not true for most practical

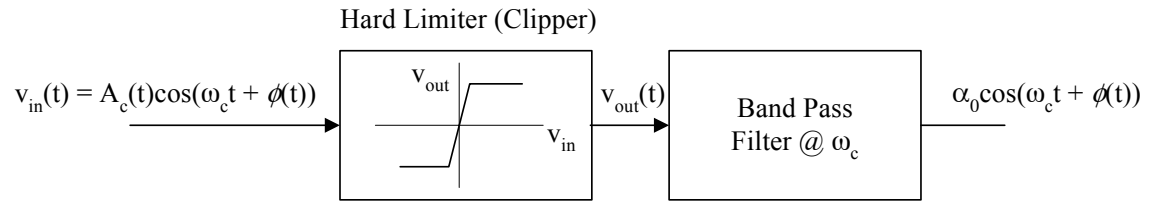


Figure 2-35: Block diagram of a band pass limiter. Voltage v_{out} is $\alpha_0\cos(\omega_c t + \phi(t))$ plus band pass processes centered at $2\omega_c$, $3\omega_c$, etc. Following the hard limiter (clipper) is a band pass filter that passes only $\alpha_0\cos(\omega_c t + \phi(t))$. When operated correctly, α_0 is a constant, and the band pass limiter forms its output by "clipping" envelope amplitude variations from the input signal.

discriminators. To some degree, a practical discriminator does produce an undesirable response to amplitude modulation on its input (a low response to envelope modulation is a measure of "goodness" for practical discriminator circuit designs). Often, this undesirable response causes problems; sometimes, it will distort the desired output signal (the response to the angle modulation). Hence, before an angle modulated input signal is sent to a practical discriminator, it is common to try to "strip off" any amplitude modulation on the signal (there are other reasons for wanting to do this).

To "strip off" small variations in the envelope, the input signal is passed through a *band pass limiter*, a block diagram of which is depicted by Figure 2-35. Here, the input signal is of the form $A_c(t)\cos(\omega_c t + \phi(t))$. We assume that envelope $A_c(t)$ is the sum of a constant and a small time varying term; that is, $A_c(t) = A_0 + A_1(t)$, where $A_0 > 0$ is a constant, and $|A_1(t)| << A_0$ for all time t. Now, the hard limiter stage has an output of $\alpha_0\cos(\omega_c t + \phi(t))$, α_0 a constant, added to a sum of band pass process each of which is centered at a different integer multiple of ω_c. Finally, the band pass filter at ω_c selects the desired output $\alpha_0\cos(\omega_c t + \phi(t))$; effectively, the band pass limiter "strips" undesirable amplitude modulation from the input signal. Some form of band pass limiting is used in almost all FM and PM receivers.

Approximations to the Ideal FM Discriminator

The ideal FM discriminator can be approximated by a differentiation followed by an envelope detector as shown by Figure 2-36. Here, we assume that A_c is constant (there is no amplitude modulation on the signal). The derivative of input $x_c(t) = A_c\cos[\omega_c t + \phi(t)]$ is

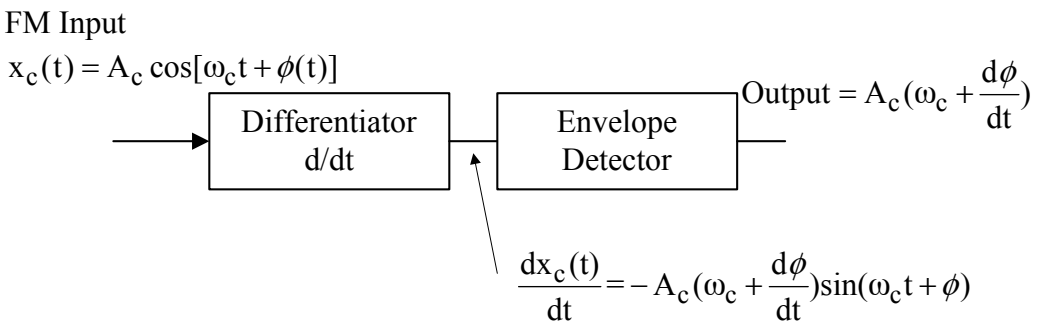


Figure 2-36: Approximation to ideal FM discriminator

$$\frac{dx_c}{dt} = -A_c(\omega_c + \frac{d\phi}{dt})\sin(\omega_c t + \phi)\,. \tag{2-106}$$

The envelope detector extracts the envelope to form the output

$$y_d(t) = A_c(\omega_c + \frac{d\phi}{dt}) = A_c[\omega_c + 2\pi f_d m(t)]\,. \tag{2-107}$$

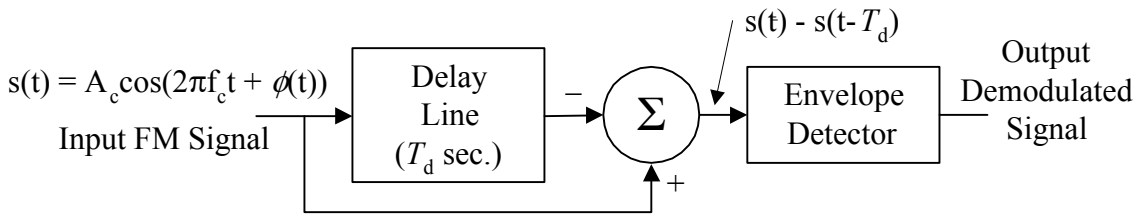
If desired, the DC component $A_c\omega_c$ can be filtered out. Note that $K_D = 2\pi A_c$ for this FM discriminator.

Delay Line FM Discriminator

Consider the practical and frequently-used FM demodulation scheme depicted by Figure 2-37. The input FM signal is passed through a delay line that produces a phase shift of π/2 radians at the carrier frequency f_c (*i.e.,* $2\pi f_c T_d = \pi/2$). The delay-line output is subtracted from the incoming FM signal, and the resulting difference signal is then envelope detected. Assume that the incoming FM wave s(t) is

$$s(t) = A_c \cos[2\pi f_c t + \phi(t)], \text{ where } \phi(t) = K_f \int^t m(\alpha)d\alpha\,. \tag{2-108}$$

Also, assume that $|\phi(t) - \phi(t - T_d)| << 1$ for all time. This assumption is true when the carrier frequency f_c is much larger that the highest frequency in the message (so the message changes



Notes: 1. The envelope detector output does not (in anyway) depend on the phase of its input s(t) - s(t-T_d).

2. Delay T_d is 25% of the carrier period: $2\pi f_c T_d = \pi/2$.

3. $|\phi(t) - \phi(t - T_d)| << 1$.

Figure 2-37: Delay-Line FM demodulator.

very little over a time interval of length T_d), a condition that is true in almost all applications.

We start the analysis of the delay-line discriminator by writing the difference of s(t) and s(t - T_d) as

$$s(t) - s(t - T_d) = A_c \cos[2\pi f_c t + \phi(t)] - A_c \cos[2\pi f_c (t - T_d) + \phi(t - T_d)]$$

$$= -2A_c \underbrace{\sin\left[\frac{2\pi f_c (2t - T_d) + \phi(t) + \phi(t - T_d)}{2}\right]}_{\text{first sin term}} \underbrace{\sin\left[\frac{2\pi f_c T_d + \phi(t) - \phi(t - T_d)}{2}\right]}_{\text{second sin term}} \quad (2\text{-}109)$$

a product of two sine functions. The first sine term is an angle-modulated high-frequency component. The second sine term is a low-frequency, modulation-dependent envelope. This envelope never changes sign because the argument

$$\frac{2\pi f_c T_d + \phi(t) - \phi(t - T_d)}{2}$$

wanders around π/4 (the argument stays in the first quadrant where sine is positive). This is important if the envelope detector is to function properly.

The envelope detector is "**totally deaf**" when it comes to sensing any angle modulation that may exist on its input (*i.e.*, the quantity [φ(t) - φ(t-T_d)] in sine term #1 has no influence on

the envelope detector output). The envelope detector responds to the envelope of its input and nothing else! As a result, the output of the envelope detector is

$$2A_c \sin\left[\frac{\pi/2+\phi(t)-\phi(t-T_d)}{2}\right] = 2A_c \sin\left[\frac{\pi}{4}+\frac{1}{2}\{\phi(t)-\phi(t-T_d)\}\right]$$
$$= \frac{2A_c}{\sqrt{2}}\left[\cos\{\frac{\phi(t)-\phi(t-T_d)}{2}\}+\sin\{\frac{\phi(t)-\phi(t-T_d)}{2}\}\right]. \quad (2\text{-}110)$$

Now, since $|\phi(t) - \phi(t - T_d)| << 1$ for all time, the envelope detector output can be approximated by

$$\text{Envelope Detector Output} \approx \sqrt{2}A_c\left[1+\{\frac{\phi(t)-\phi(t-T_d)}{2}\}\right] \approx \sqrt{2}A_c\left[1+\frac{1}{2}\frac{d\phi}{dt}T_d\right]$$
$$= \sqrt{2}A_c\left[1+\frac{K_f}{2}T_d m(t)\right]. \quad (2\text{-}111)$$

Hence, the envelope detector output contains a scaled version of the message.

Single-Tuned FM Discriminator (Slope Detector)

A practical FM discriminator can be made by using the voltage-to-frequency response characteristics of a high-Q RLC tuned circuit or other resonant network (such as a ceramic pizo-electric resonator or quartz crystal). The basic circuit is illustrated by Figure 2-38, and it

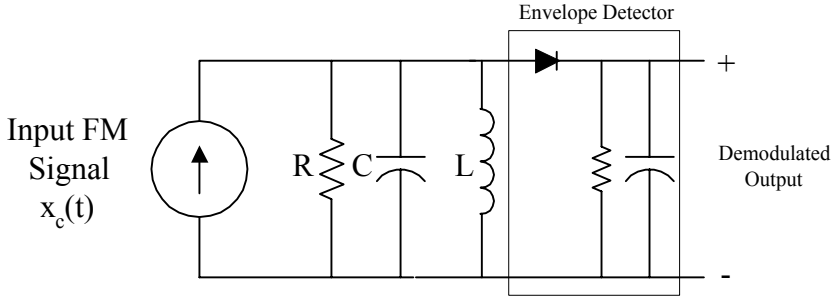


Figure 2-38: Single-tuned FM discriminator (slope detector).

contains a parallel RLC circuit and envelope detector.

Figure 2-39a depicts the band-pass amplitude response curve of the detector's high-Q RLC resonant circuit. The FM signal carrier frequency ω_c is placed slightly above the resonance peak, and the FM signal sweeps out a small frequency range around the carrier, always staying on a fairly straight portion of the response curve. Often, ω_c is placed at an inflection point of the magnitude response curve. (At an inflection point, the magnitude response has a second derivative that vanishes. There is an inflection point both above and below the resonant frequency. Can you find such points?) As depicted by Fig. 2-38, the circuit does a frequency-to-voltage conversion, and the output voltage swings over a range on the vertical axis of the response curve.

The circuit must yield a frequency-to-voltage relationship that is approximately linear for this method to work well. Any non-linearity shows up as harmonic distortion in the demodulated audio output. In practice, this method works best for low-peak-frequency-deviation FM signals. A single tuned resonator slope detector is used extensively in narrow-

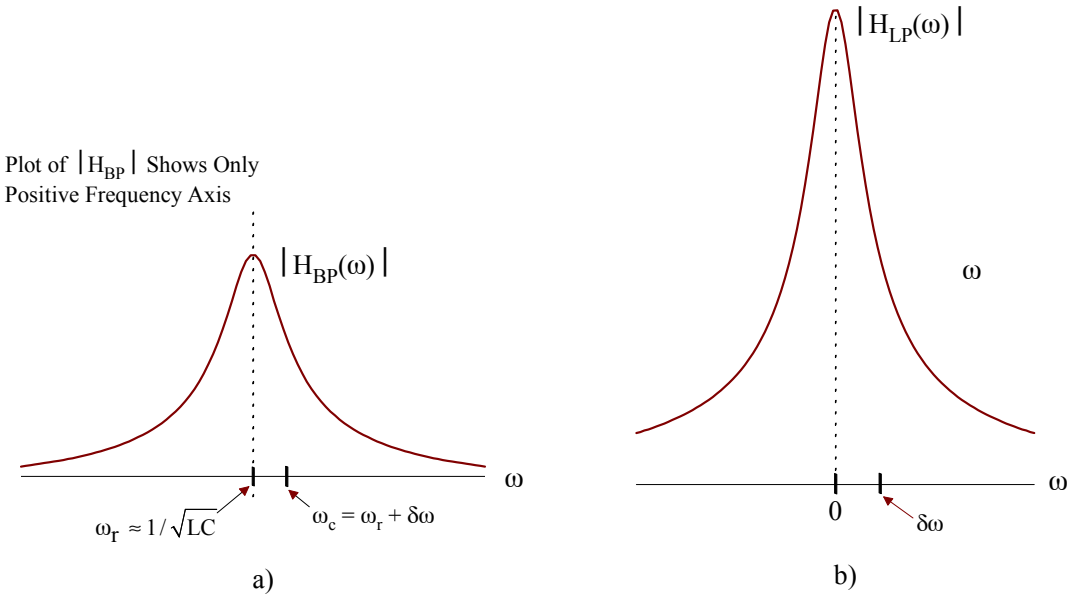


Figure 2-39: In a), $|H_{BP}(\omega)|$ has a resonance peak at $\omega_r \approx 1/\sqrt{LC}$. Place the FM carrier at $\omega_c = \omega_r + \delta\omega$, $\delta\omega > 0$, so that the FM signal sweeps a narrow range of frequencies around ω_c. In b), $|H_{LP}(\omega)|$ is the low pass equivalent; also shown is the center frequency $\delta\omega$ of the FM signal low-pass equivalent.

band FM applications (however, a ceramic pizo-electric resonator or quartz crystal may be used instead of the LC network).

Recall that a high-Q, single-tuned, parallel RLC circuit has a response that can be modeled as being *symmetrical* (review the *symmetrical* band-pass filter theory contained in Chapter 1 of these notes). For the high-Q case, $|H_{LP}(\omega)|$, the magnitude of the low-pass equivalent transfer function, can be modeled as an **even** function of frequency ω. Figure 2-39b depicts $|H_{LP}(\omega)|$.

Figure 2-39b also depicts δω, the center frequency of the FM signal low-pass equivalent. This low-pass equivalent signal sweeps out a small frequency range centered at δω, where the amplitude response curve $|H_{LP}(\omega)|$ is nearly linear.

Since it is an even function of frequency, magnitude $|H_{LP}(\omega)|$ has a power-series expansion that contains only even-order powers of ω; that is, it has an expansion of the *form*

$$|H_{LP}(\omega)| = \alpha_0 - \alpha_2\omega^2 - \alpha_4\omega^4 - \cdots, \tag{2-112}$$

where α_k, $k \geq 0$, are constants (the actual values of the α_k are not important in this discussion).

In the single-tuned discriminator, the transmitted carrier frequency ω_c is positioned on the slope of the response curve, δω radians/second above the resonance peak that occurs at ω_r. That is, $\omega_c = \omega_r + \delta\omega$, where $\omega_r \approx 1/\sqrt{LC}$ is the resonant frequency of the LC circuit. The transmitted signal is a narrow-band signal centered at the carrier frequency $\omega_r + \delta\omega$ radians/second. The instantaneous transmit frequency is $\omega_r + \delta\omega + \dot{\phi}$, where $\dot{\phi}$ is the frequency deviation ($\dot{\phi} = K_f m(t)$ for FM). The instantaneous frequency of the low-pass equivalent FM signal (referenced to ω_r as shown by Fig 2-39b) is $\delta\omega + \dot{\phi}$. As shown by Figure 2-40, the single-tuned discriminator output is approximately proportional to

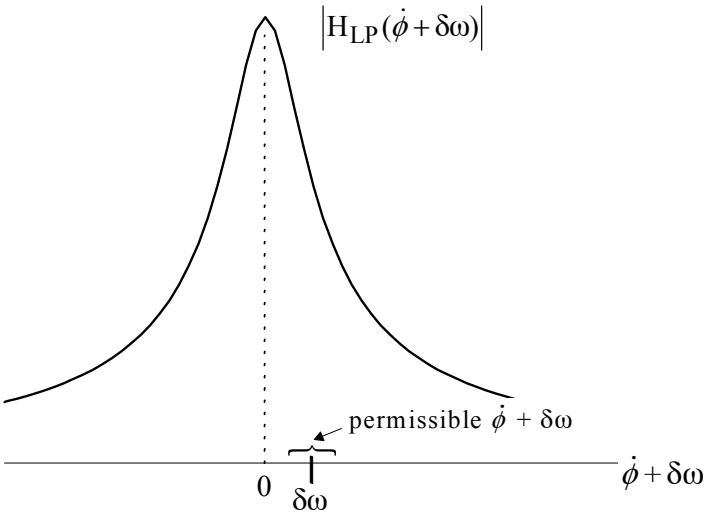


Figure 2-40: Response of singly-tuned discriminator. $\dot{\phi}$ is the instantaneous frequency deviation. It must remain small.

$$\left|H_{LP}(\dot{\phi}+\delta\omega)\right| = \alpha_0 - \alpha_2(\dot{\phi}+\delta\omega)^2 - \alpha_4(\dot{\phi}+\delta\omega)^4 - \cdots$$

$$= \underbrace{\{\alpha_0 - \alpha_2\,\delta\omega^2 - \alpha_4\,\delta\omega^4 - \cdots\}}_{\text{constant}} - \underbrace{\{2\alpha_2\,\delta\omega + 4\alpha_4\,\delta\omega^3 + \cdots\}}_{\text{coefficient of first-order term}}\dot{\phi} \quad , \qquad (2\text{-}113)$$

$$- \underbrace{\{\alpha_2 + 6\alpha_4\,\delta\omega^2 + \cdots\}}_{\text{coefficient of second-order term}}\dot{\phi}^2 + \cdots$$

where the range $|\dot{\phi}| \le \delta\omega$ is of interest. To write (2-113), we have assumed that $\dot{\phi}$ varies slowly enough (relative to the reciprocal of the RLC network bandwidth) to permit the use of *quasi-steady-state* analysis (*i.e.,* (2-113) is an approximation that is valid for slowly varying $\dot{\phi}$).

In general, the expansion (2-113) of $|H_{LP}(\dot{\phi}+\delta\omega)|$ contains all powers of $\dot{\phi}$. Only the linear term (the first-order-in-$\dot{\phi}$ term is linear in m) is of interest. Second and higher-order terms (*i.e.,* $\dot{\phi}^2, \dot{\phi}^3, \ldots$) cause harmonic distortion in the demodulated output. However, harmonic distortion of the demodulated message will not be significant if $|\dot{\phi}|$ remains sufficiently small for all time, as would be the case for narrow-band FM where the peak frequency deviation is small (5 kHz or less in practice).

In (2-113), note that even (alternatively, odd) powers of $\dot{\phi}$ have coefficients that contain only even (alternatively, odd) powers of δω. So, when the sign of δω is changed (*i.e.*, the FM carrier is placed on the *lower side* of the RLC response curve) we obtain the expansion

$$\left|H_{LP}(\dot{\phi}-\delta\omega)\right| = \alpha_0 - \alpha_2(\dot{\phi}-\delta\omega)^2 - \alpha_4(\dot{\phi}-\delta\omega)^4 - \cdots$$

$$= \underbrace{\{\alpha_0 - \alpha_2\,\delta\omega^2 - \alpha_4\,\delta\omega^4 - \cdots\}}_{\text{constant}} + \underbrace{\{2\alpha_2\,\delta\omega + 4\alpha_4\,\delta\omega^3 + \cdots\}}_{\text{coefficient of first-order term}}\dot{\phi} \qquad (2\text{-}114)$$

$$- \underbrace{\{\alpha_2 + 6\alpha_4\,\delta\omega^2 + \cdots\}}_{\text{coefficient of second-order term}}\dot{\phi}^2 + \cdots$$

Compare the right-hand sides of (2-113) and (2-114); note that terms involving odd (alternatively, even) order powers of $\dot{\phi}$ are opposite (alternatively, the same) in sign. This fact allows the cancellation of even powers of $\dot{\phi}$ in the output of the double-tuned slope detector discussed next.

Double-Tuned FM Discriminator (Slope Detector)

The use of two tuned circuits reduces/eliminates even-order harmonic terms in the discriminator output and makes for a more linear frequency-to-voltage relationship. Figure 2-41 depicts a block and circuit diagram of a *doubly-tuned discriminator*. The upper resonant L_1C_1 circuit is tuned δω *above* the carrier frequency ω_c; the lower resonant L_2C_2 circuit is tuned δω *below* the carrier frequency (the individual LC circuits have responses depicted by the dotted-line plots on Fig. 2-42). On the secondary of the discriminator transformer, the two coils are wound so that the "coupling dots" appear on opposite ends of the secondary windings. Hence, the output response is the *difference* of the two responses.

The response of the doubly-tuned FM discriminator is proportional to the *difference* of (2-113) and (2-114). Subtract (2-114) from (2-113) to obtain

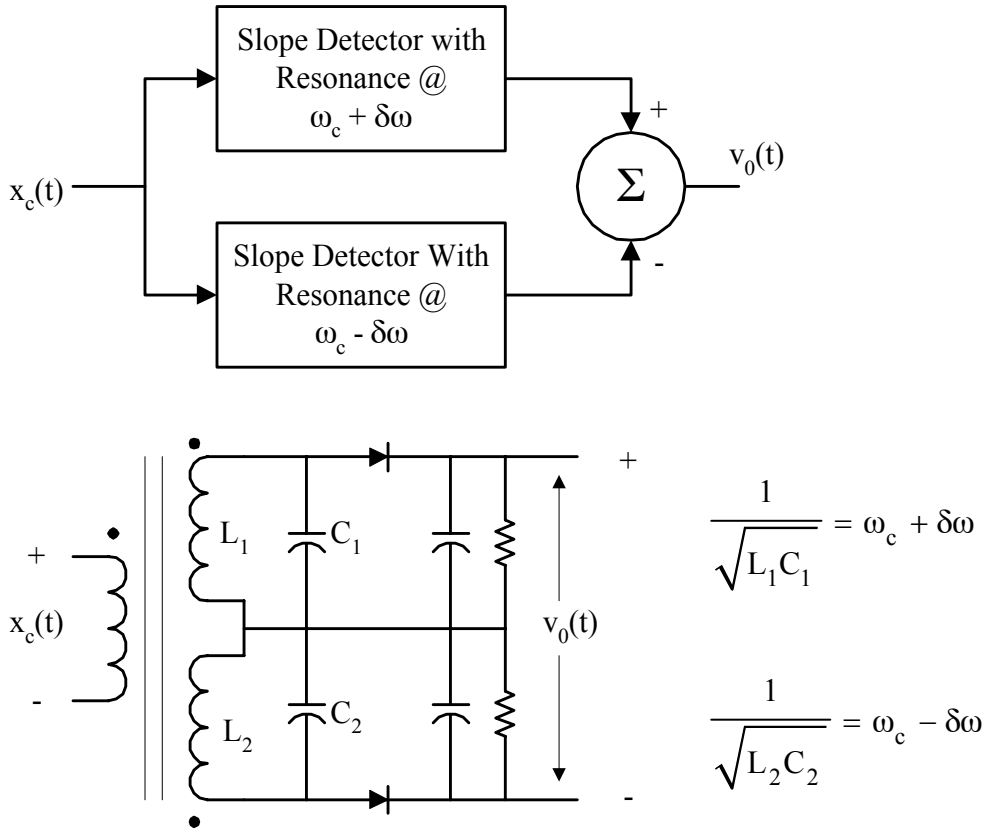


Figure 2-41: Doubly-tuned FM discriminator.

$$\left|H_{LP}(\dot{\phi}-\delta\omega)\right| - \left|H_{LP}(\dot{\phi}+\delta\omega)\right| = 2\{2\alpha_2\,\delta\omega + 4\alpha_4\,\delta\omega^3 + \cdots\}\dot{\phi} + 2\{4\alpha_4\delta\omega + 20\alpha_6\,\delta\omega^3 + \cdots\}\dot{\phi}^3 + \cdots \quad (2\text{-}115)$$

Note that even-order powers of $\dot{\phi}$ have canceled out while odd-order powers have added constructively. For the double-tuned discriminator described by (2-115), this cancellation of even-order terms makes for a more linear response than what is possible with a single RLC circuit.

Figure 2-42 illustrates this cancellation effect. The two LC circuits have responses that are depicted by the dotted line graphs. Over the band $\omega_c - \delta\omega < \omega < \omega_c + \delta\omega$, the solid-line difference curve is significantly "straighter" than either dotted line curve, implying the partial cancellation of non-linearity effects (the output is free of even-order powers of $\dot{\phi}$ and message

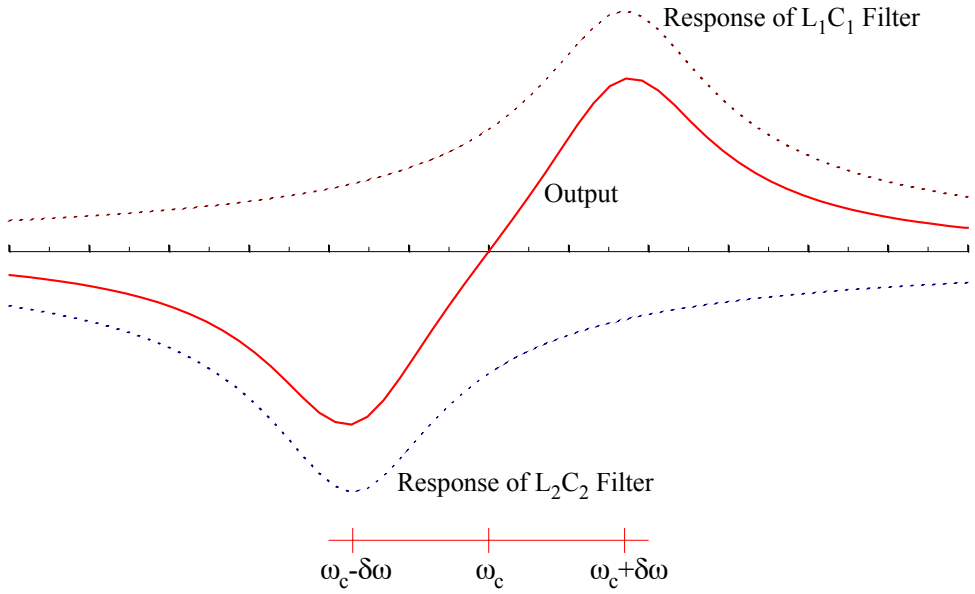


Figure 2-42: Response of a doubly-tuned FM discriminator. Compared to the response of either LC circuit acting alone (the dashed line plots), the output (the solid line plot) is a "more linear" function of frequency for values of ω in a neighborhood of ω_c.

m(t)). Double resonator slope detectors are used extensively in receivers designed for commercial FM broadcast where the peak frequency deviation (on the order of 75 kHz) is much higher than that used in point-to-point, narrow-band FM communication systems (which use a peak frequency deviation of 5 kHz).

Interference - Pre-emphasis/De-emphasis

Assume an input to an ideal PM or FM discriminator is an unmodulated carrier at ω_c plus an interfering, in-band tone at $\omega_c + \omega_i$ (where $\omega_i << \omega_c$). That is, we assume that the received signal is of the form

$$\begin{aligned} x_r(t) &= A_c \cos\omega_c t + A_i \cos[(\omega_c + \omega_i)t] \\ &= A_c \cos\omega_c t + A_i \cos\omega_i t \cos\omega_c t - A_i \sin\omega_i t \sin\omega_c t \\ &= [A_c + A_i \cos\omega_i t]\cos\omega_c t - A_i \sin\omega_i t \sin\omega_c t \\ &= R(t)\cos[\omega_c t + \phi(t)] \end{aligned} \quad , \qquad (2\text{-}116)$$

where

$$R(t) = \sqrt{(A_c + A_i \cos\omega_i t)^2 + (A_i \sin\omega_i t)^2}$$
$$\phi(t) = \text{Tan}^{-1}\left(\frac{A_i \sin\omega_i t}{A_c + A_i \cos\omega_i t}\right) \quad (2\text{-}117)$$

We are interested in the case where the in-band interference is much smaller than the unmodulated carrier. That is, we assume that $A_c >> A_i$ so that

$$R(t) \approx (A_c + A_i \cos\omega_i t)$$
$$\phi(t) \approx \frac{A_i}{A_c}\sin\omega_i t \quad . \quad (2\text{-}118)$$

are the envelope and phase, respectively. Hence, the received signal can be approximated as

$$x_r(t) = R(t)\cos[\omega_c t + \phi(t)] \approx A_c\left(1 + \frac{A_i}{A_c}\cos\omega_i t\right)\cos\left(\omega_c t + \frac{A_i}{A_c}\sin\omega_i t\right). \quad (2\text{-}119)$$

The instantaneous phase of x_r is given in (2-118). For PM, the ideal discriminator output is

$$y_d(t) = K_D\phi(t) = K_D\frac{A_i}{A_c}\sin\omega_i t\,. \quad (2\text{-}120)$$

For FM, the ideal discriminator output is

$$y_d(t) = \frac{1}{2\pi}K_D\frac{d\phi(t)}{dt} = K_D\frac{A_i}{A_c}f_i\cos\omega_i t\,, \quad (2\text{-}121)$$

where f_i is the frequency in Hz of the interfering, in-band tone. Note that the amplitude of y_d is proportional to f_i for FM but not for PM. *The interfering tone has more effect on FM than PM if* f_i *is large*.

On FM systems, the severe effect of large-f_i tone interference can be reduced by placing a filter, called a *deemphasis filter*, at the FM discriminator output. In practice, the filter is typically a simple RC low-pass filter with a 3dB cutoff frequency that is considerably less than the audio message bandwidth. In the USA, the industry standard for wide-band FM commercial broadcasting is $\Omega_b/2\pi \approx 2.1$kHz; for narrow-band FM used in point-to-point communications, the standard is $\Omega_b/2\pi \approx 150$Hz. At frequencies well above cutoff, the filter has a response that is proportional to 1/f. Since the tone interference has a magnitude proportional to f_i for FM, the filter cancels out further increases in interference output.

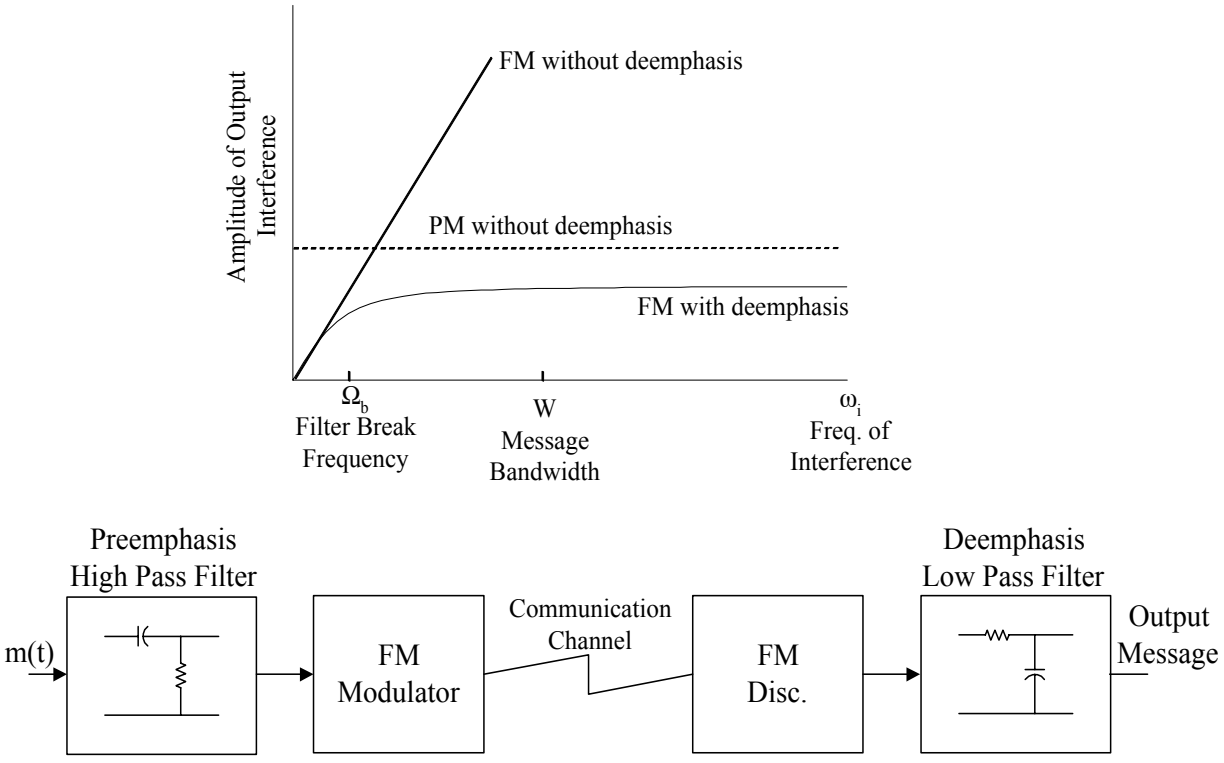


Figure 2-43: Pre- and Deemphasis filtering in an FM communication System.

The deemphasis filter distorts the message signal in addition to reducing the in-band tone interference. However, the distortion can be avoided by passing message m(t) through a high-pass filter, called a *preemphasis filter*, before applying it to the transmitter's audio input. Figure 2-43 depicts how preemphasis and deemphasis filtering are used in an FM communication system.

Matlab Example - In-Band Tone Interference

The previous section described the result of in-band tone interference on an unmodulated carrier. Results were obtained under the assumptions 1) the interfering tone is offset from the carrier by ω_i, and 2) the interfering tone is much weaker than the carrier. Under these conditions, Equation (2-121) predicts a demodulated output consisting of a "weak" sinusoidal tone (its amplitude is proportional to $A_i/A_c << 1$, the ratio of interference to carrier amplitudes) at frequency ω_i. In this section, these results are extended to the case of a strong interfering in-band tone on an unmodulated carrier. According to the Matlab simulation described here, the demodulated output will loose its "weak" sinusoidal nature as the interfering signal grows in amplitude and $A_i/A_c \rightarrow 1$.

According to (2-117), the unmodulated carrier and interfering tone can be modeled as a

```
%MATLAB command line dialog
    r = .5;
    % r = .9;
    % r = .95;
    t = [ [0:2*pi/50:pi-.1],[pi-.1:2*pi/200:pi+.1],[pi+.1:2*pi/50:2*pi] ];
    y=fm2(r,t);
    plot(t,y)
%calls the MATLAB function fm2(r,t) given by
    function y = fm2(r,t)
    % FM demodulated output for in-band sinusoidal interference
    % John Stensby
    % Dec 6, 2000
    s = r*sin(t);
    c = r*cos(t);
    y = (r^2 + c)./( (1+c).^2 + s.^2 );
%to produce the following plots (uncomment one of r = .5, r = .9 or r = .95)
```

Figure 2-44: Matlab script to produce the Figures 2-44 through 2-46.

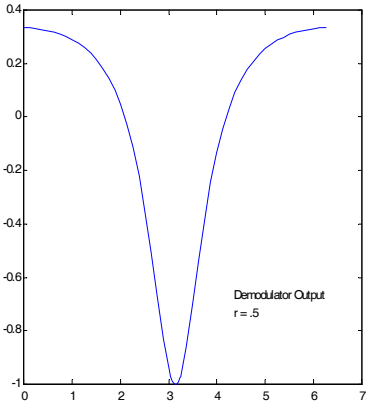


Fig. 2-45: Output for r = .5

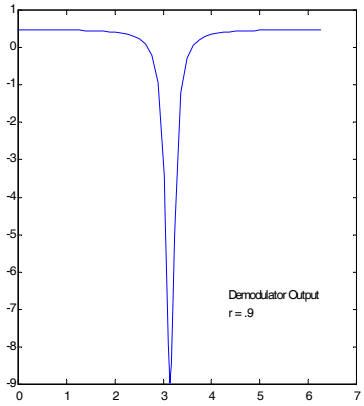


Fig. 2-46: Output for r = .9

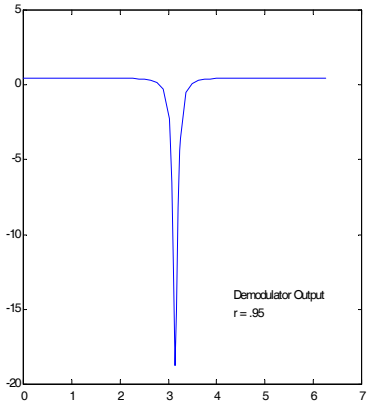


Fig. 2-47: Output for r = .95

signal at ω_c with angle modulation given by

$$\phi(t) = \mathrm{Tan}^{-1}\left(\frac{r\sin\omega_i t}{1+r\cos\omega_i t}\right), \tag{2-122}$$

where $r = A_i/A_c$ is the ratio of interference to carrier amplitudes. The demodulated output can be computed to be

$$y_d(t) = \frac{K_D}{2\pi}\frac{d\phi(t)}{dt} = \frac{K_D}{2\pi}\omega_i\left[\frac{r^2 + r\cos\omega_i t}{(1+r\cos\omega_i t)^2 + (r\sin\omega_i t)^2}\right], \tag{2-123}$$

a periodic signal with fundamental frequency ω_i. For various values of r, one period of this output was computed and plotted by using Matlab and the script listed in Figure 2-44. The results are described by Figures 2-45 through 2-47.

As A_i/A_c approaches unity, the demodulated output becomes a periodic stream of very sharp, narrow spikes (see the three plots given above). Listening to this demodulated signal on a loudspeaker, it would sound like a periodic series of "pops" or "clicks".

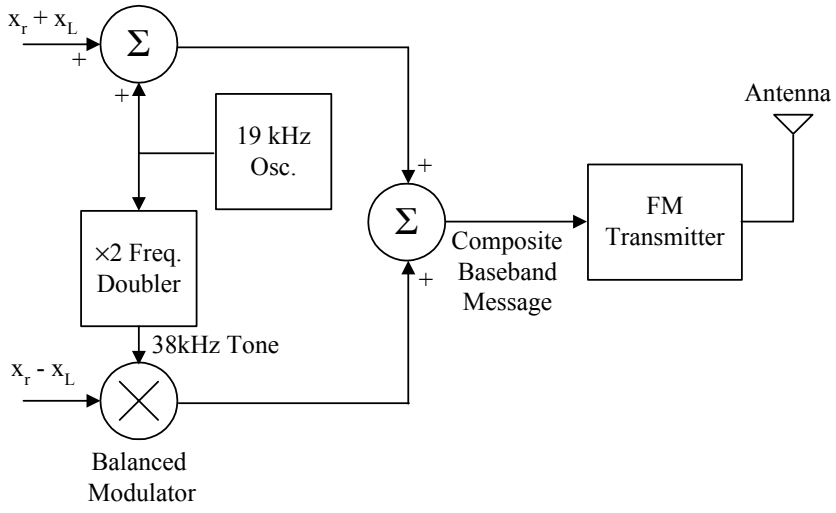
FM Stereo Transmitter and Receiver

In the USA, stereo broadcasts were first authorized in the early 1960's (all FM was

monaural before the advent of stereo). Today, almost all FM stations broadcast in stereo using the scheme discussed in this section.

Figure 2-48a) depicts a simplified block diagram of an FM stereo transmitter. Added to a conventional FM transmitter are components that encode "left" and "right" channel audio signals into a composite base-band audio signal. The signals x_r and x_L denote the right-channel audio and left-channel audio, respectively. Both audio channels are band limited to 15Khz. The system uses the notion of a "sub-carrier" that is double-sideband modulated with the difference of the audio channels.

a)



b)

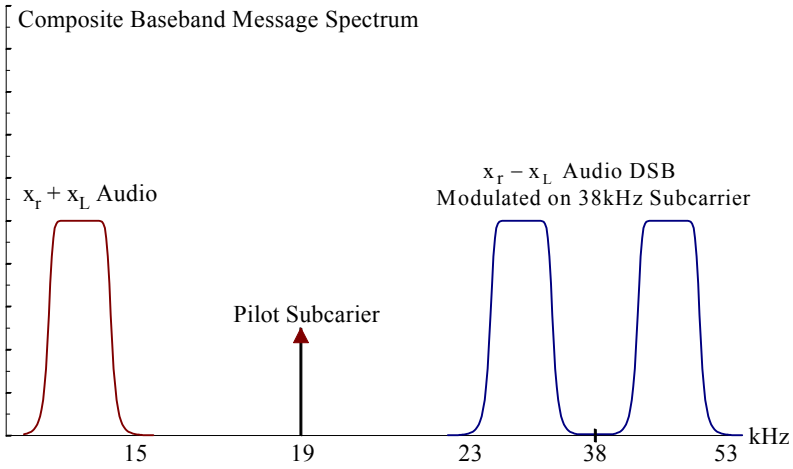


Figure 2-48: a) Simplified block diagram of a stereo FM transmitter. b) Composite base band message spectrum.

Figure 2-48b) depicts a one-sided spectral plot of the composite base-band message signal that is FM modulated on the transmitter carrier. In the base-band signal, the sum $x_r + x_L$ occupy the first 15kHz of spectrum. This sum signal would be recovered by a monaural FM receiver (*i.e.*, one that could not decode a stereo broadcast), so the stereo system is backwards compatible with older monaural receivers. Also in the base-band signal, a 19kHz pilot tone is added to aid channel separation at the stereo receiver. Finally, the audio signal $x_r - x_L$ is used to DSB modulate a 38 kHz subcarrier (the second harmonic of the pilot tone).

Figure 2-49 depicts a simplified block diagram of a stereo FM receiver. A narrow band-pass filter recovers the 19kHz sub-carrier from the composite base-band message signal. Usually, this narrow band-pass filter is implemented by a phase-lock loop (PLL). The PLL locks

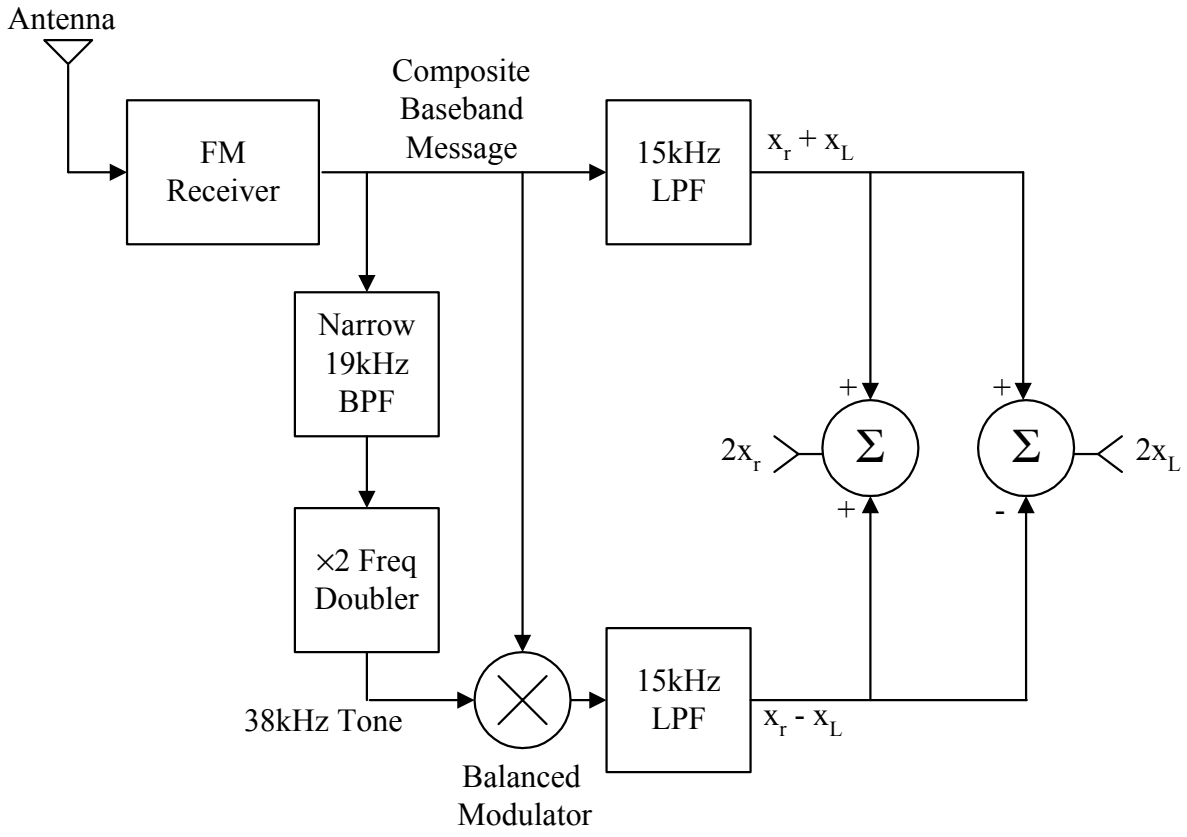


Figure 2-49: FM stereo receiver. Often, the narrow 19kHz BPF is implemented by a phase-lock loop that locks onto the 19kHz subcarrier.

onto the 19kHz subcarrier and extract it from the composite base-band message. A PLL lock detector (not shown on the diagram) would sense a lock condition and illuminate a FM Stereo indicator LED (or other indicator) on the receiver front panel.

Electronic Oscillators

Every modern electronic communication system contains at least one oscillator, a circuit that generates a periodic waveform. In addition, oscillators are used widely though out the field of electronics. Basically, an oscillator is an amplifier (transistor, integrated circuit, etc.) with a feedback path composed of passive, frequency selective components. From output to input of the amplifying device, energy is fed back in such a manner that the circuit is unstable (so it oscillates). At the frequency of oscillation, the small-signal gain of the amplifying device must equal, or exceed, the attenuation of the passive, frequency selective, network. Also, the signal must be fed back in phase to the amplifier. Let $H_a(\omega)$ and $H_n(\omega)$ denote the transfer function of the amplifier and network, respectively. If oscillation is to occur at ω_0, then we must have $|H_a(\omega_0)H_n(\omega_0)| \geq 1$ and $\measuredangle H_a(\omega_0)H_n(\omega_0) = 0$ (modulo 2π). See Figure 2-50.

Oscillators can employ a variety of different technologies. Simple oscillators can utilize RC networks. For example, the frequency of a multivibrator is set by a RC time constant. A wide variety of commonly used oscillators employ the resonant properties of LC networks. Very stable high-frequency (HF) oscillators can be constructed by using quartz crystals. Stable, very-

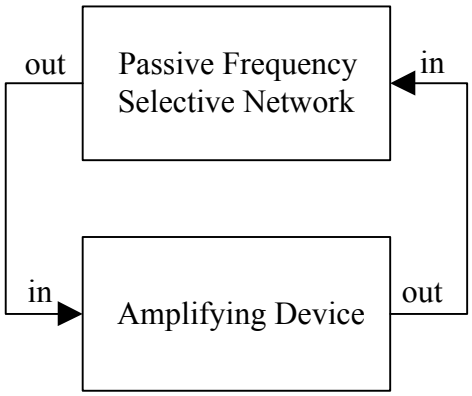


Figure 2-50: Every oscillator contains an amplifying device and a passive, frequency selective network.

high frequency (VHF) oscillators can be constructed using surface acoustic wave (SAW) devices or quartz crystals in an over-tone mode. A resonant cavity can be used to make a microwave oscillator. There are many oscillator technologies, each with advantages and disadvantages.

There are many practical applications where one needs to use an external voltage to control the frequency of an oscillator. Such circuits are known as *voltage controlled oscillators (VCOs),* and they are used as FM generators, in phase-locked loops (PLLs), and in many other applications.

Figure 2-51 depicts a simplified circuit of a *Colpitts* voltage controlled oscillator. The frequency control (*i.e.*, "tuning") voltage consists of a DC bias v_{dc} added to a control signal $e(t)$ so that the sum $v_{dc} + e(t)$ is always positive, and the varactor diodes (also known as *vari-cap diodes*) are always back biased (and their *pn* junctions exhibit capacitance that is inversely related to the back-bias voltage). The frequency of oscillation is determined by L, C, C_1, C_2, the capacitance of the *varactor diodes*, and the reactances inherent in the FET. In addition, the passive LC network provides for impedance matching between the low-impedance point at the FET source (output of the amplifying device) and the high-impedance point at the FET gate (input of the amplifying device). Normally, the circuit is designed so that the frequency of

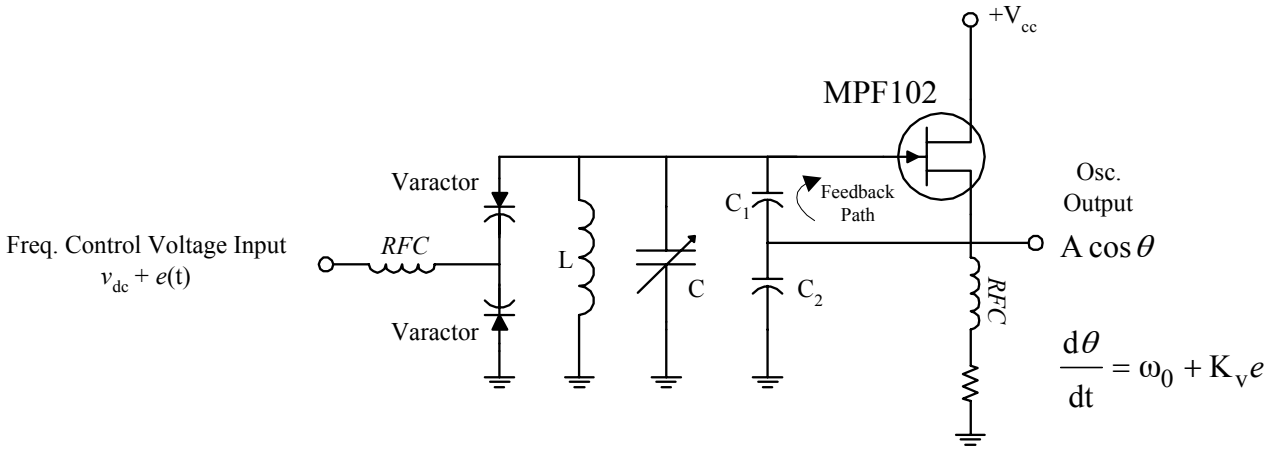


Figure 2-51: Simplified circuit diagram of a Colpitts voltage controlled oscillator. *RFC* stands for *radio frequency choke*. The frequency control voltage consists of a DC bias v_{dc} plus a voltage control signal $e(t)$. Coarse frequency control can be obtained by manually adjusting variable capacitor C. Fine frequency control can be achieved by adjusting tuning voltage $e(t)$.

oscillation is linearly dependent (approximately so) on the control voltage input e(t), at least for small e. Note that the FET drain is at RF ground (in all oscillators, one of the three FET/transistor leads must be at RF ground). Finally, the use of "split capacitor" (*i.e.*, C_1 and C_2) feedback is characteristic of *Colpitts* oscillators (to obtain feedback, oscillators of the *Hartley* type employ a tapped inductor instead of a spit capacitor network).

A simple linear model can be constructed for this VCO. The oscillator output signal can be modeled as $A\cos\theta$, where

$$\frac{d\theta}{dt} \equiv \omega_0 + K_v e , \tag{2-124}$$

assuming that *VCO control voltage* e(t) remains sufficiently small for all time. The constant ω_0 is called the *VCO center frequency*, and K_v is known as the *VCO gain* constant.