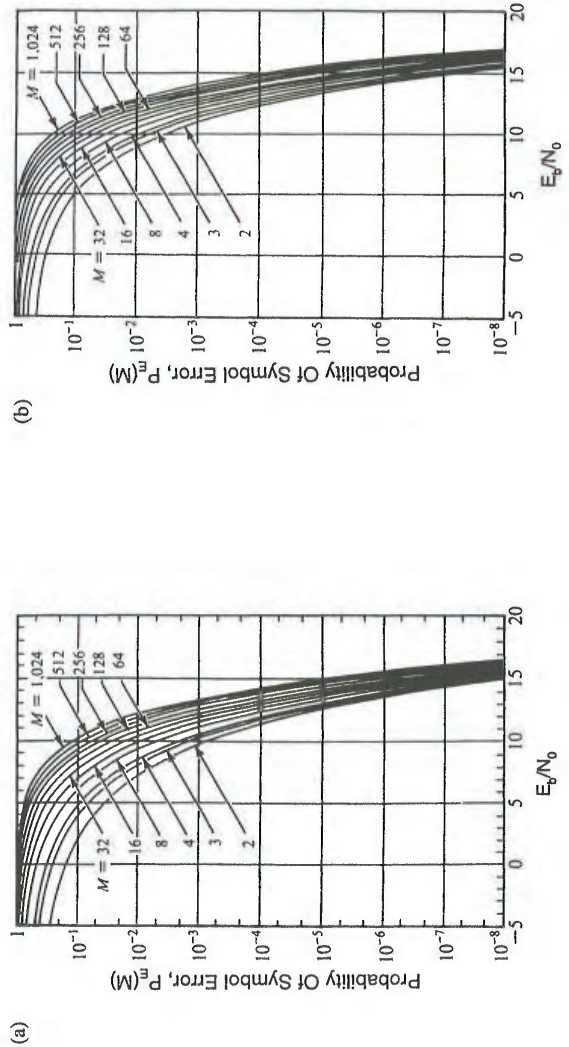


Figure 4.80 MFSK probability of bit error: (a) coherent MFSK, (b) noncoherent MFSK.

$$h(t) = e^{-\frac{1}{2}\left(\frac{t}{\tau}\right)^2} \tag{4.73}$$

Figure 4.81 shows plots of both the time domain and frequency response of a Gaussian lowpass filter. Notice that both curves have the same bell-shaped response.

The Gaussian impulse response is well-behaved, exhibiting no ringing or overshoot. However, the frequency response of the filter tends to fall off rather slowly. In general, the more gradual the frequency response of a filter, the better its time domain response (an exception to this rule are digital FIR filters). Gaussian filters have one of the most gradual frequency responses of any analog filter type.

Notice that the frequency response, given by (4.71) is purely real. This implies that the phase of the Gaussian filter is linear. Linear phase filters can be designed using digital FIR filter techniques or can be implemented with SAW filters. However, it is possible to approximate the response of a Gaussian filter by using standard reactive filters such as LC ladder filters. Figure 4.82 shows a schematic for a passive LC ladder filter with a table of normalized Gaussian filter coefficients for different filter orders. The table is for double-terminated filters.

For example, the design of a Gaussian lowpass filter with a cut-off (or –3 dB) frequency of 500 kHz with a 50-ohm system impedance is presented here [12]. The following equations are used to find the actual component values from the prototype values listed in the table:

$$L = \frac{R}{2\pi f_c} L_{\text{proto}} \tag{4.74}$$

$$C = \frac{1}{2\pi f_c R} C_{\text{proto}} \tag{4.75}$$

From the table, for a third-order filter, we find

$$C_{\text{proto},1} = 0.2624 \tag{4.76}$$

$$L_{\text{proto},2} = 0.8167 \tag{4.77}$$

$$C_{\text{proto},3} = 2.2262 \tag{4.78}$$

This gives the following component values:

$$C_1 = 1.67 \text{ nF} \tag{4.79}$$

$$L_2 = 13.0 \text{ uH} \tag{4.80}$$

$$C_3 = 14.2 \text{ nF} \tag{4.81}$$



Figure 4.81 Gaussian lowpass filter waveforms: (a) impulse response, (b) frequency response.

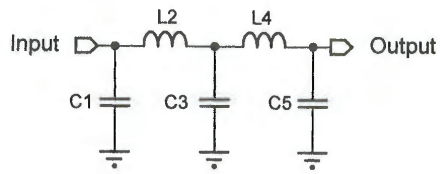


Table Of Normalized Values

N	$C1_{proto}$	$L2_{proto}$	$C3_{proto}$	$L4_{proto}$	$C5_{proto}$
2	0.4738	2.1850			
3	0.2624	0.8167	2.2262		
4	0.1772	0.5302	0.9321	2.2450	
5	0.1312	0.3896	0.6485	0.9782	2.2533

Actual component values given by: $L = \frac{R}{2\pi f_c} L_{proto}$ $C = \frac{1}{2\pi f_c R} C_{proto}$

Figure 4.82 Using lumped-element LC ladder filter to approximate Gaussian lowpass filter.

Figure 4.83 shows the completed, third-order Gaussian lowpass filter. Gaussian filters designed using reactive elements have a more gradual frequency response than other reactive filters, such as Bessel, Butterworth, and Chebyshev. However, Gaussian filters exhibit no overshoot or ringing in the time domain. This smooth, well-behaved impulse response results in very little intersymbol interference (ISI).

Figure 4.84(a) shows a block diagram of a Gaussian filtered MSK (GMSK) modulator [13–15]. The GMSK modulator is identical to the VCO-based version of the MSK modulator, except for the lowpass modulation filter. The bandwidth of the Gaussian filter is often given in terms its relation to the bit rate. Recall

$$T \equiv T_b = \frac{1}{f_b} \tag{4.82}$$

where T or T_b is the bit period of the filter and f_b is the bit rate. If we let B equal the –3-dB bandwidth of the filter, then we can specify the filter response in terms of its *relative bandwidth*, or BT

$$\text{BT} = (\text{Filter Bandwidth}) \cdot (\text{Bit Period}) = \frac{\text{Filter Bandwidth}}{\text{Bit Rate}} \tag{4.83}$$

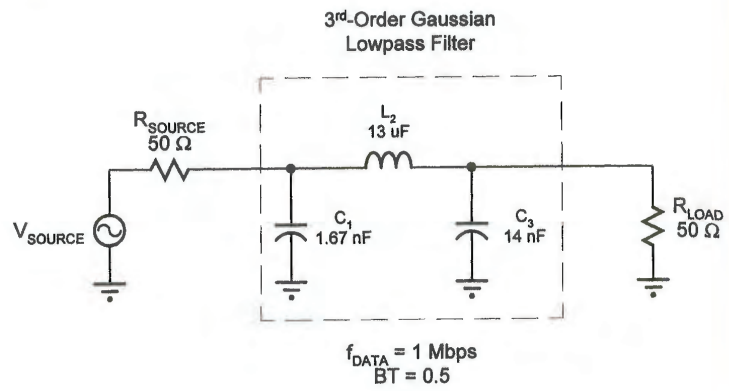


Figure 4.83 Actual third-order Gaussian lowpass filter (BT = 0.5, 50Ω system).

As we shall see shortly, GMSK has a very narrow transmitted spectrum, yet exhibits very little ISI and is tolerant to hardlimiting. Figure 4.84(b) shows a simple noncoherent demodulator, similar to that used in FSK, which can be used to demodulate GMSK. Applications for GMSK include the global system for mobile communication (GSM) with BT = 0.3, and the digital European cordless telephone (DECT) with BT = 0.5.

The time domain response of GMSK with BT = 0.5 (i.e., the filter bandwidth is equal to one-half the bit rate) is shown in Figure 4.85. The top trace of the figure shows the modulated signal. Just as in the case of MSK, GMSK is a constant envelope modulation. In contrast to MSK, the second trace shows that the phase now has a smooth slope. This is caused by the action of the Gaussian lowpass filter. Notice that the filter also smoothes out the input bit stream, removing the sharp transitions in the data.

The action of the filter has a radical effect on the transmitted spectrum. Figure 4.86 shows a plot of the power spectral density for various GMSK signals having different relative filter bandwidths (BT). Recall that the "M" in GMSK stands for minimum, so the modulation index, h, is equal to 0.5 in all cases. Should h not be equal to 0.5, then the modulations is referred to as GFSK. If no filter is used, then the BT = infinity. This is equivalent to MSK, and is shown in the plot for comparison purposes. The figure shows that a filter having a bandwidth equal to half the data rate (i.e., BT = 0.5) rolls off much more quickly in frequency than MSK (i.e., no filter). Using filters with smaller relative bandwidths causes even faster roll-off. Unfortunately, this faster spectral roll-off has a price.

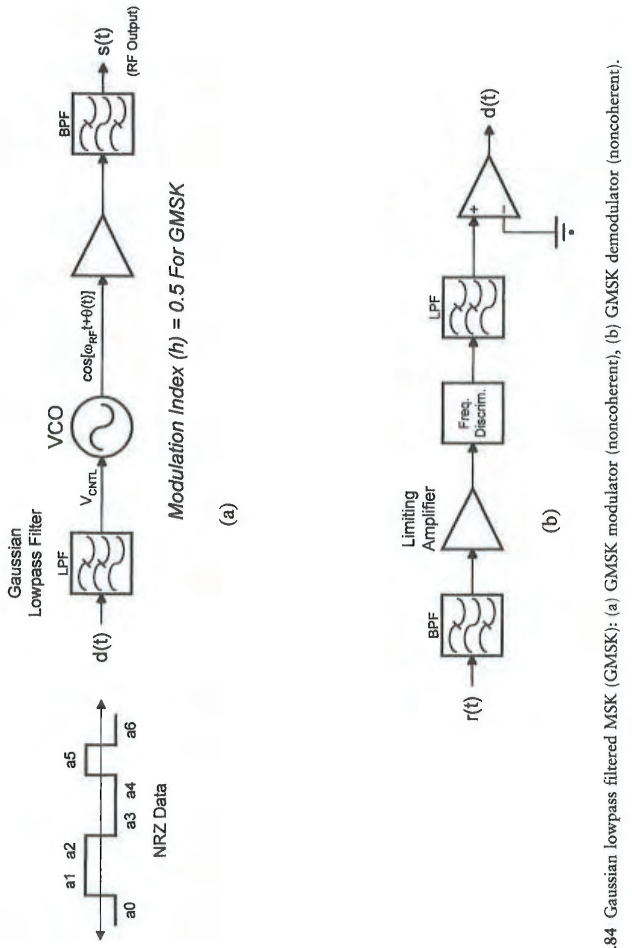


Figure 4.84 Gaussian lowpass filtered MSK (GMSK): (a) GMSK modulator (noncoherent), (b) GMSK demodulator (noncoherent).

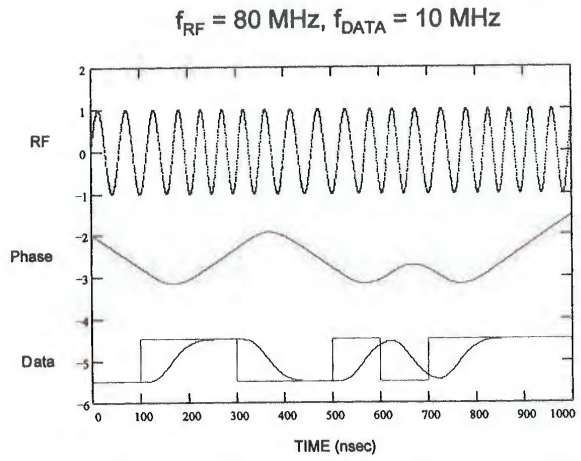


Figure 4.85 GMSK time domain waveforms.

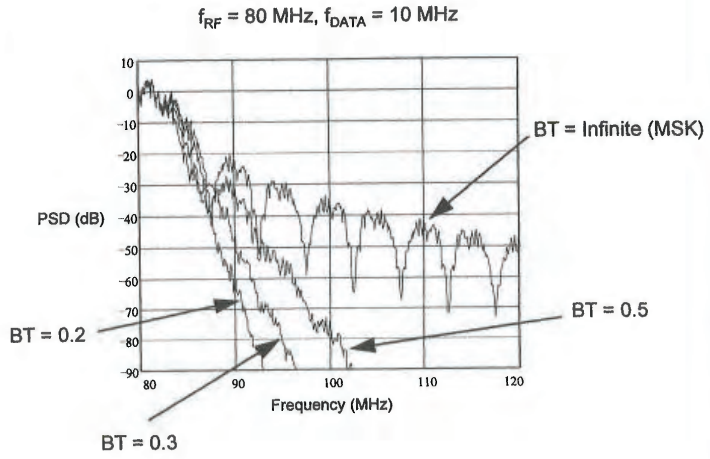


Figure 4.86 GMSK power spectral density for different relative filter bandwidths (BT).

As the relative filter bandwidth is lowered, more and more ISI is imparted to the waveform. One method by which we can get a qualitative indication of the amount of ISI on the waveform is by looking at an eye diagram. Figure 4.87 shows the test setup used to measure eye diagrams. The received, or recovered, data is fed to the vertical input of an oscilloscope. The scope is triggered on the rising edge of the recovered, or local, clock. This action causes multiple traces to overlap on the scope's cathode ray tube (CRT). Eye diagrams get their name by the way the signal resembles an "eye" on the CRT. As noise or ISI increases, the eye begins to close, increasing P_E.

Figure 4.88 shows a collection of waveforms with varying degrees of ISI. In Figure 4.88(a), very little filtering is used and all data transitions on the recovered waveform are rectangular in shape. Long sequences of 0s or 1s result in the horizontal lines at the bottom and top of the traces, respectively. In Figure 4.88(b), a special type of filter was used that rolls off the frequency response of the filter, causing the rectangular pulses to become sinusoidal in shape. However, at the sampling instant,

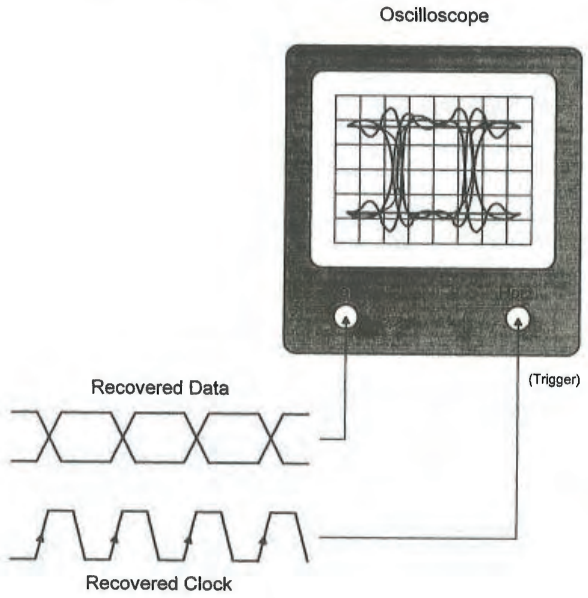


Figure 4.87 Measurement setup for observing eye diagrams.

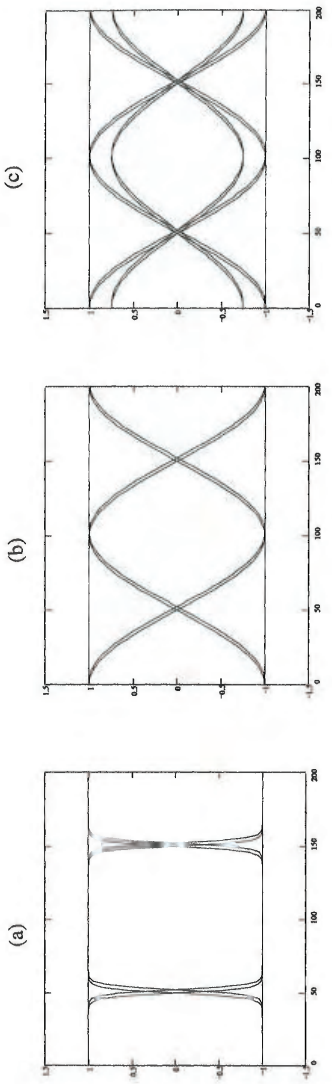


Figure 4.88 Intersymbol interference (ISI): (a) unfiltered eye pattern, (b) filtered eye pattern without ISI, (c) filtered eye pattern with ISI.

the signal has reached its full value, resulting in no ISI at the sampling instant. As we shall see later, Nyquist filters are one type of filter that possess this quality. On the other hand, the trace shown in Figure 4.88(c) *does* exhibit ISI at the sampling instant. ISI is created by data bits in periods preceding the present data period not fully settling out.

Figure 4.89(a) shows the time domain response of a GMSK modulation signal with a BT = 0.5 filter. Note that the amplitude of the modulation waveform is equivalent to the instantaneous frequency deviation of the carrier. Figure 4.89(b) shows an eye diagram of this same signal taken over many more bit periods. Notice that with a filter having a bandwidth of half the bit period (BT = 0.5), very little ISI is seen in the waveform. Figure 4.90(a) shows the time domain response for a BT = 0.25 filter (i.e., one-quarter the bit rate). Notice how alternating strings of 0s and 1s result in waveforms that are nowhere near the full value of the original, unfiltered signal. The eye diagram shows this effect much more clearly. At the sampling instant (near $t = 0$), the eye is almost closed. It is easy to see that very little noise will cause a bit error in this case.

When the GMSK signal is noncoherently demodulated, the output of the demodulator is equivalent to the instantaneous frequency deviation of the signal. We have just seen that with low values of BT, the eye is almost closed. Therefore, a relatively high SNR must be maintained if bit errors are to be kept low. Earlier, we discussed how MSK can be viewed as either phase or frequency modulation. We saw that MSK probability of bit error can equal that of BPSK and QPSK if the signal is processed over two bit periods when the signal is coherently demodulated. We might expect that we should be able to approach this level of performance if we coherently demodulate GMSK—given the fact that GMSK is a derivative of MSK.

Figure 4.91 shows a block diagram of a coherent GMSK demodulator. The demodulator is identical to that of MSK. In many cases, a predetection Gaussian bandpass filter is used at the input of the demodulator. The lowpass filters shown in the figure can be implemented as matched filters for better performance. Figure 4.92 shows a plot of the amount of degradation in performance for a GMSK signal for different values of BT. According to the plot, we need to increase E_b/N_0 by only 1 dB over that of coherently demodulated MSK in order to maintain the same P_E performance when BT = 0.215. This is surprising in view of how bad the eye diagrams look when BT is reduced.

The reason for this result is that when we coherently demodulate the GMSK signal, we are really looking at the *phase* of the signal over two or more bit periods. Figure 4.93 shows eye diagrams for both the I and Q portions of a coherently demodulated GMSK signal for BT = 0.5. As the figure shows, the eye diagrams are relatively open. If we reduce BT to 0.25, we get the eye diagrams shown in Figure 4.94. Although there is some closing of the eye, the waveforms are remarkably good—especially compared to the eye diagram shown for noncoherent demodulation. This

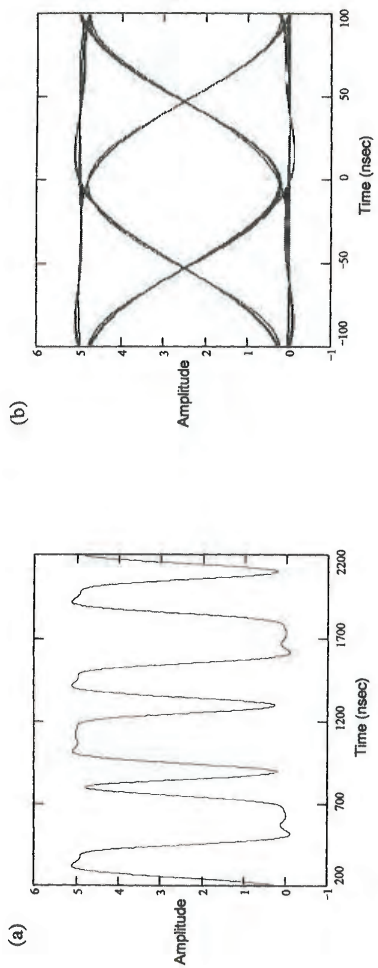


Figure 4.89 GMSK demodulated waveforms (noncoherent): (a) recovered bit stream (BT = 0.5), (b) eye diagram (BT = 0.5).

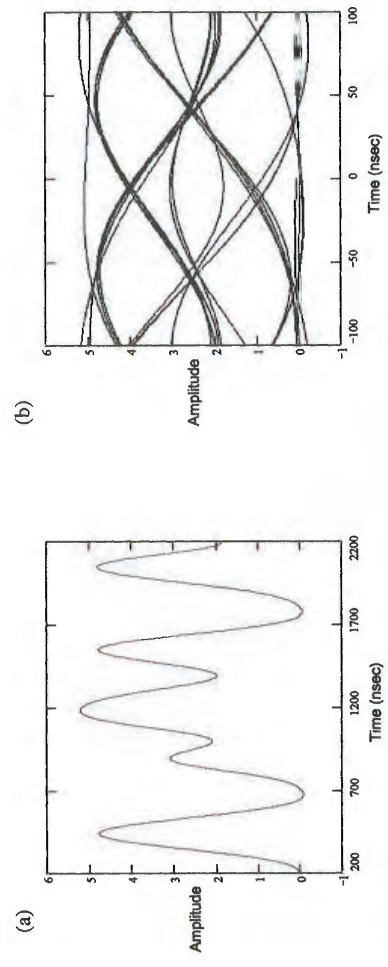


Figure 4.90 GMSK demodulated waveforms (noncoherent): (a) recovered bit stream (BT = 0.25), (b) eye diagram (BT = 0.25).

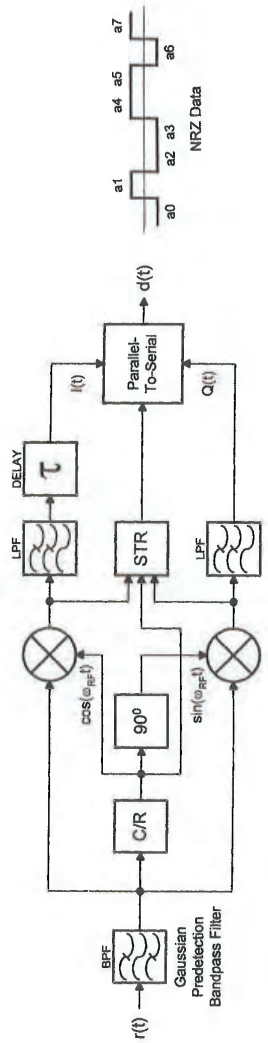


Figure 4.91 Coherent GMSK demodulator.

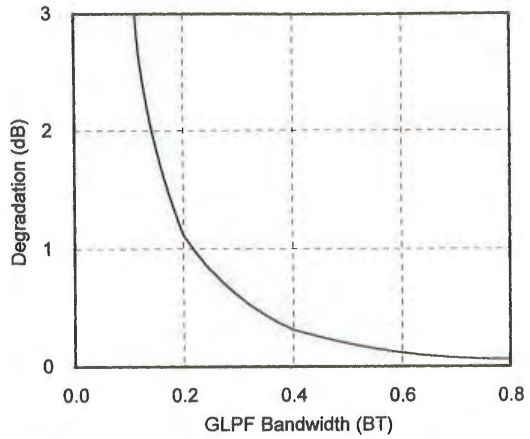


Figure 4.92 Theoretical E_b/N_0 degradation of GMSK as a function of BT compared to ideal coherent demodulation.

(a)

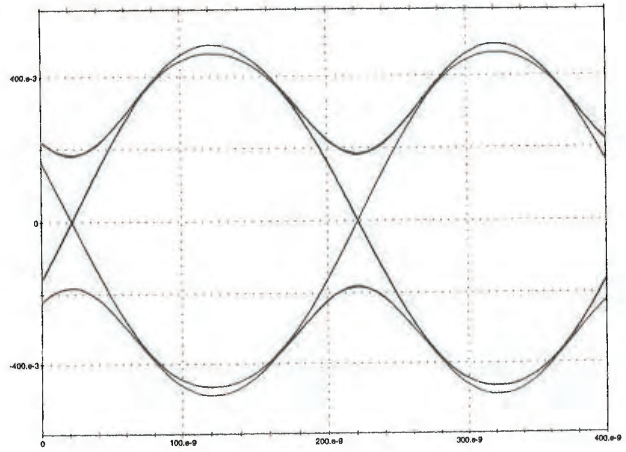


Figure 4.93 Eye diagrams for coherently demodulated GMSK (BT = 0.5): (a) *I*-channel, (b) *Q*-channel.

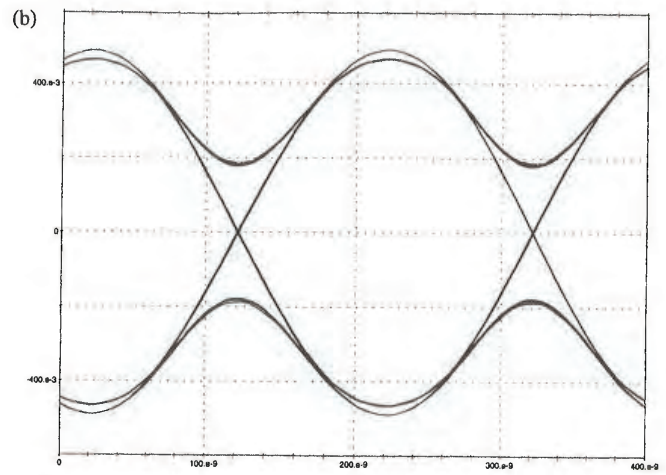


Figure 4.93 (continued)

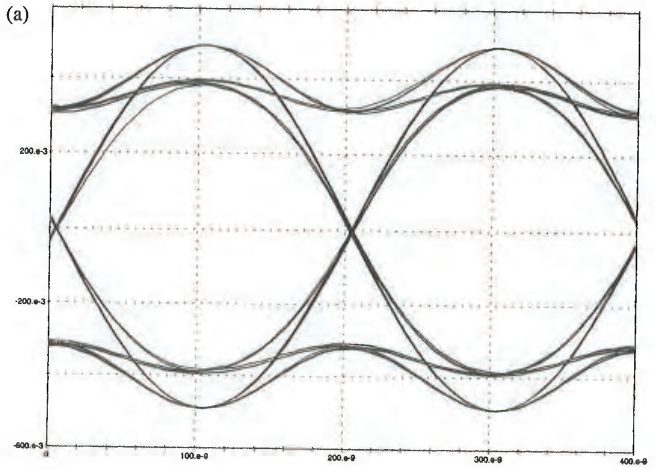


Figure 4.94 Eye diagrams for coherently demodulated GMSK (BT = 0.25): (a) *I*-channel, (b) *Q*-channel.

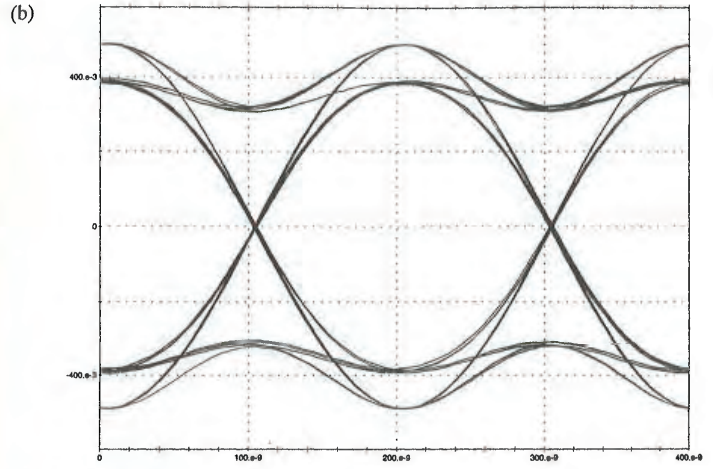


Figure 4.94 (continued)

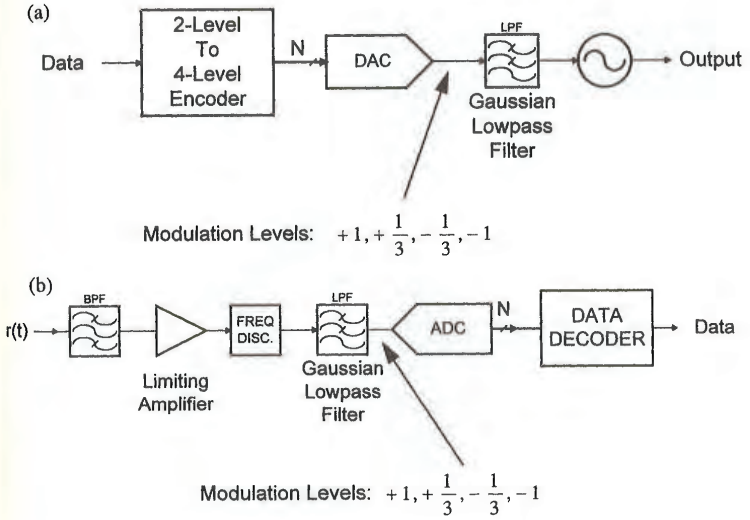


Figure 4.95 4-level GMSK: (a) modulator, (b) demodulator.

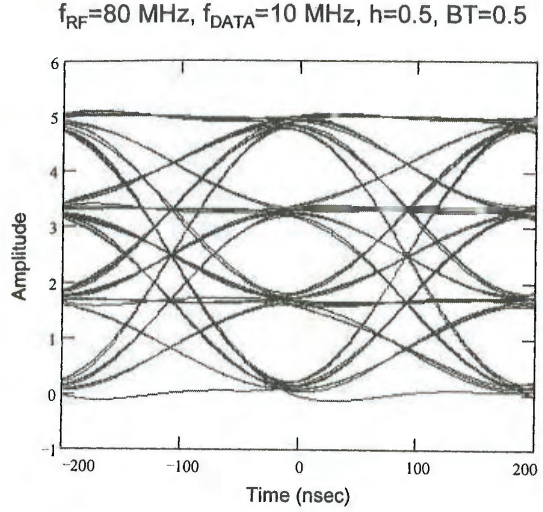


Figure 4.96 4-level GMSK eye diagram.

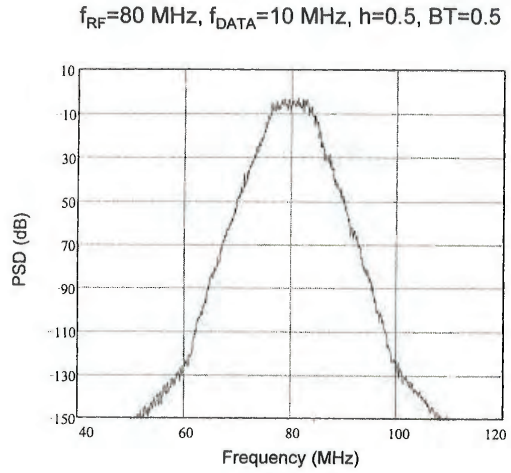


Figure 4.97 4-level GMSK power spectral density (h = 0.5, BT = 0.5).

is the reason the plot in Figure 4.92 shows little degradation over coherent MSK, even for low values of BT.

We can extend GMSK by using multiple frequencies to represent groups of bits in much the same way we did with FSK. Figure 4.95 shows a block diagram for a 4-level GMSK modulator and a 4-level GMSK demodulator. These circuits are identical to those used in FSK, with the exception of the lowpass filters following the DACs. An eye diagram for a 4-level GMSK signal is shown in Figure 4.96 with a BT = 0.5. Note that the relative filter bandwidth is in terms of the symbol rate and not the bit rate. Notice that even with four levels, there is a relatively small amount of ISI at the sampling instant (t = 0).

The transmitted power spectral density for a 4-level GMSK signal with a carrier frequency of 80 MHz and a bit rate of 10 Mbps is shown in Figure 4.97. Again, the action of the lowpass filter caused the transmitted spectrum to fall off very rapidly.

4.4.2 π/4-Shifted Differentially Encoded PSK (π/4-DQPSK)

Variations of π/4-shifted differentially encoded PSK (π/4-DQPSK) have been around for quite some time. Baker [16] first applied this technique to 2,400 bps wired telephone modems. More recently, however, π/4-DQPSK has received attention as the U.S. (North American dual-mode cellular, NADC) and Japanese digital cellular (JDC) cellular telephone standards [17–21]. This modulation allows for a threefold increase in channel capacity over the existing analog FM cellular systems. Receivers can employ either coherent demodulators, differential demodulators, or discriminators (followed by integrate-and-dump filters). The ability to noncoherently demodulate the π/4-DQPSK signal results in cost and complexity reductions. Discriminator detection can also be used to demodulate the existing analog FM signals, allowing dual-mode operation.

π/4-DQPSK is a modulation that is a cross between QPSK and 8-PSK. Like QPSK, π/4-DQPSK separates the incoming bit stream into groups of two-bit symbols. In QPSK, each symbol would correspond to one of four different phase angles: +45 deg, −45 deg, +135 deg, or −135 deg. In π/4-DQPSK, however, every other symbol is part of a second QPSK constellation, which is shifted (or rotated) from the original QPSK constellation by π/4 (or 45 deg) with each symbol lying on either the I or Q axis. The inclusion of this second, π/4-shifted QPSK constellation means that symbols are represented by one of eight equally spaced phase angles, as in

8-PSK. However, unlike 8-PSK, only selected phase angles are available for the next symbol, depending on the phase angle of the present symbol.

Figure 4.98 illustrates the action of $\pi/4$-DQPSK modulation. Let the present symbol correspond to +45 deg (i.e., one of the white dots). In $\pi/4$-DQPSK, the only allowable symbols for the next symbol are *on* the I and Q axes (i.e., one of the black dots). This means that the phase of the next symbol can be either +90 deg, 180 deg, −90 deg, or 0 deg.

Continuing on, let's assume that we do in fact jump to +90 deg from our original symbol, which was at +45 deg. This means that the next symbol (the one after our *now* present value of +90 deg) must lie *off* the I and Q axes. In short, if the present symbol corresponds to one of the white dots in Figure 4.98 (i.e., *off* the I and Q axes), then the next symbol must be one of the black dots in the figure (i.e., *on* the I and Q axes). If the present symbol corresponds to one of the black dots, then the next symbol must correspond to one of the white dots. Even when long strings of 0s or 1s are encountered in the input data stream, a phase transition will always occur in the transmitted signal. This feature makes it easier to generate symbol timing in the receiver.

Figure 4.99 shows the phase-modulated carrier versus I and Q data inputs. No premodulation filters were used in order to show the relative phase changes of the carrier. However, as we will discuss later, special premodulation filters are generally used to bandlimit the transmitted signal.

The maximum phase transition in $\pi/4$-DQPSK is limited to 135 deg (in either direction). Recall that QPSK has maximum phase transitions of 180 deg, while the maximum phase transition of OQPSK is 90 deg (both, again, can be in either direction). In effect, $\pi/4$-DQPSK is a compromise between QPSK and OQPSK. By limiting phase transitions to 135 deg, amplitude variation is reduced compared to QPSK, yet is still larger than the 90-deg transitions of OQPSK. Limiting the maximum phase transition per symbol allows the use of amplifiers with lower output back-off (OBO) without experiencing spectral regrowth. In reality, even though $\pi/4$-DQPSK has less amplitude variation than QPSK, the amount of OBO over QPSK is negligible when reasonably linear amplifiers are used.

The "D" in $\pi/4$-DQPSK comes from the fact that the next symbol phase angle is derived by adding a *differential* phase angle to the present symbol's phase. These differential phase angles are transmitted, rather than absolute phase angles. Using differential encoding in the transmitter allows the use of simpler differential demodulators in the receiver. It has been shown [21] that differential demodulators are more robust in high-Doppler, Rayleigh-faded cellular environments. Table 4.1 shows a mapping of the differential phase angle, $\Delta\phi_k$, which is added to the present phase angle, as a function of A_k and B_k (the parallel symbol bits) for the next symbol. For example, given a phase angle of $\pi/4$ (45 deg) for the present symbol, suppose the next set of bits is $A_k = 1$ and $B_k = 1$. This means $-3\pi/4$ (−135 deg) will be added to the present symbol's phase, resulting in a new symbol having a phase angle of $-\pi/2$ (−90 deg). Note that the values of A_k and B_k result in gray-coded differential phase angles. Again, suppose $A_k = 1$ and $B_k = 1$ and our present symbol is $\pi/4$ (45 deg). This means the next symbol we transmit will have a phase angle of $-\pi/2$ (−90 deg).

As the signal passes through the noisy channel, it is possible that the recovered symbol is estimated to be either 0, $+\pi/2$ (+90 deg), or $+\pi$ (180 deg)—our only choices. The Euclidean distance on the I/Q phase plane between $-\pi/2$ and 0 (or π) is less than the distance between $-\pi/2$ and $+\pi/2$. With Gaussian white noise, should an error occur, it is much more likely that either 0 or π are chosen by the demodulator as the estimate of received phase angle. In other words, it is more likely the resultant vector lands in an adjacent quadrant than in the opposite quadrant when additive noise causes an error. If an error does occur, A_k, B_k will most likely be estimated to be either 1, 0 or 0, 1. Remember, if there were no error, the received values of A_k, B_k would be 1, 1. Therefore, the most likely symbol errors result in only one bit being in error.

Figure 4.100 shows a block diagram for a $\pi/4$-DQPSK modulator. As the figure shows, the modulator is very similar to that used for QPSK or 8-PSK, where analog in-phase and quadrature signals are used to drive a quadrature modulator. However, in the $\pi/4$-DQPSK modulator, the in-phase (I) and quadrature (Q) data streams are first encoded. The values of I_k and Q_k driving the inputs of the quadrature modulator are derived from the present symbol phase angle, ϕ_k, and the differential phase angle, $\Delta\phi_k$ (which is a function of A_k and B_k), by the following relations:

$$I_k = \cos[\phi_k] = \cos[\phi_{k-1} + \Delta\phi_k] \tag{4.84}$$

$$Q_k = \sin[\phi_k] = \sin[\phi_{k-1} + \Delta\phi_k] \tag{4.85}$$

Alternatively, the present values of I_k and Q_k can be written as functions of the previous values of I_k, Q_k (i.e., I_{k-1}, Q_{k-1}) and the differential phase angle, $\Delta\phi_k$, as

$$I_k = I_{k-1}\cos[\Delta\phi_k] - Q_{k-1}\sin[\Delta\phi_k] \tag{4.86}$$

$$I_k = I_{k-1}\sin[\Delta\phi_k] + Q_{k-1}\cos[\Delta\phi_k] \tag{4.87}$$

Notice that a pair of premodulation lowpass filters are used to bandlimit analog inputs I_k and Q_k, before being fed to the quadrature modulator. In addition to specifying the use of differential phase angles, the NADC and JDC standards specify the use of special premodulation filters in the transmitter and receiver. Recall that raised cosine (Nyquist) filters have very sharp roll-off in the frequency domain (approaching ideal, brick-wall response) and exhibit no ISI at sampling intervals, making them ideal for PSK premodulation filters.

A variation of the raised cosine filter is used in $\pi/4$-DQPSK. The raised cosine filter is separated into two elements whose aggregate response, when connected in series, results in the raised cosine filter response. One filter is used in the transmitter and the other filter is used in the receiver. The demodulated signal will then be free of ISI, after passing through both the transmit and receive filters since the series cascade of the two filters has a raised cosine response. If we let the filter response be $S(\omega)$, then we can write

$$|H_{RC}(\omega)| = |S(\omega)| \cdot |S(\omega)| = |S(\omega)|^2 \tag{4.88}$$

where $H_{RC}(\omega)$ is the response of the raised cosine filter. We can then write the frequency response of each of the filters as

$$|S(\omega)| = \sqrt{|H_{RC}(\omega)|} = |S_{SRRC}(\omega)| \tag{4.89}$$

This type of filter is called a square-root raised cosine (SRRC) filter, whose frequency response can be written as the square-root of the raised cosine filter, or [18]

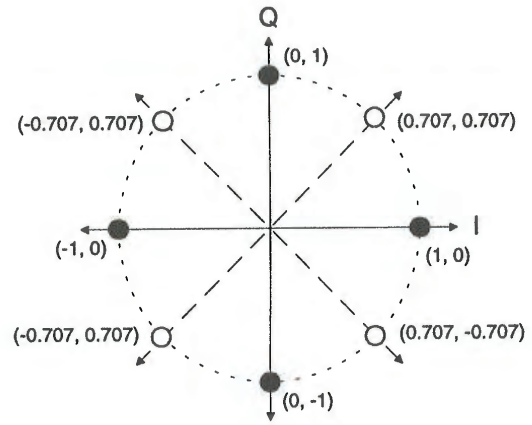


Figure 4.98 $\pi/4$-DQPSK phase domain.

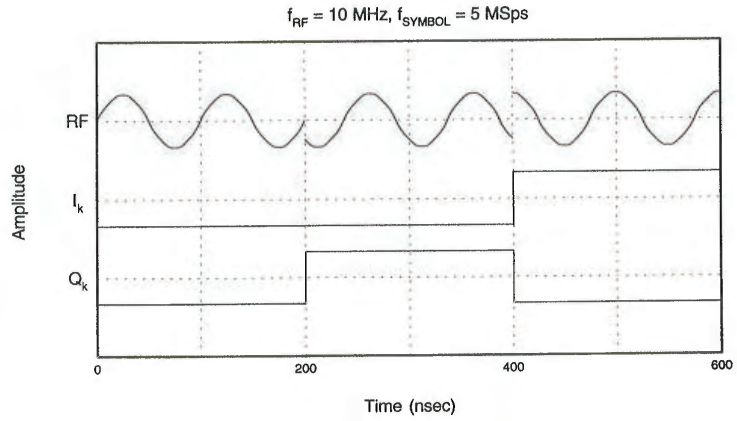


Figure 4.99 $\pi/4$-DQPSK time domain waveforms.

Table 4.1
Differential Phase Angle, $\Delta\phi_k$, Versus Symbol Bits A_k (In-Phase) and B_k (Quadrature)

A_k	B_k	$\Delta\phi_k$
0	0	$+\pi/4$ (+45 deg)
0	1	$3\pi/4$ (135 deg)
1	0	$-\pi/4$ (−45 deg)
1	1	$-3\pi/4$ (−135 deg)

$$|S_{SRRC}(\omega)| = T;\ 0 \le |f| \le \frac{(1-r)}{2T}$$
$$= T\sqrt{\frac{1}{2}\left\{\sin\left[\pi\frac{T}{r}\left(|f| - \frac{1}{2T}\right)\right]\right\}};\ \frac{(1-r)}{2T} \le |f| \le \frac{(1+r)}{2T} \quad (4.90)$$
$$= 0\ ;\ |f| > \frac{(1+r)}{2T}$$

where T is the symbol period and r is the roll-off factor. In terms of its impulse response, the SRRC filter can be expressed as

$$S_{SRRC}(t) = 1 - r + 4\frac{r}{\pi};\ t = 0$$
$$= \frac{r}{\sqrt{2}}\left[\left(1 + \frac{2}{\pi}\right)\sin\left(\frac{\pi}{4r}\right) + \left(1 - \frac{2}{\pi}\right)\cos\left(\frac{\pi}{4r}\right)\right];\ t = \pm\frac{T}{4r} \quad (4.91)$$
$$= \frac{\sin\left[\pi(1-r)\frac{t}{T}\right] + 4r\frac{t}{T}\cos\left[\pi(1+r)\frac{t}{T}\right]}{\pi\frac{t}{T}\left[1 - \left(4r\frac{t}{T}\right)^2\right]};\ \text{for all other } t$$

An advantage to splitting the raised cosine filter into two elements is that the SRRC filter in the receiver now becomes a matched filter. A matched filter is a filter whose response is such that the desired signal passes without attenuation, while all other signals (such as noise and interference) are rejected. Matched-filter-based receivers generally have lower P_Es than other receivers, for the same level of channel noise.

One way to demodulate the $\pi/4$-DQPSK signal is to use a standard coherent I/Q quadrature demodulator as shown in Figure 4.101. However, the analog I and Q signals will alternate between two-level and three-level eye patterns, as shown in the figure. If the received symbol lies on either the I or Q axes, the values of I and

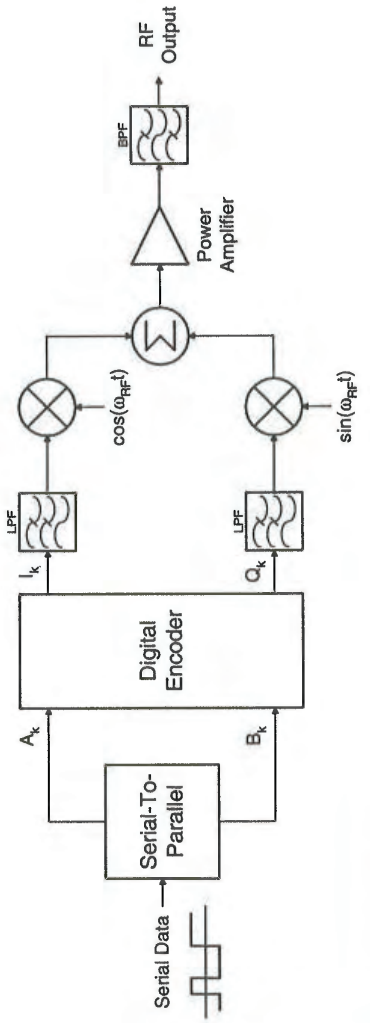


Figure 4.100 $\pi/4$-DQPSK modulator.

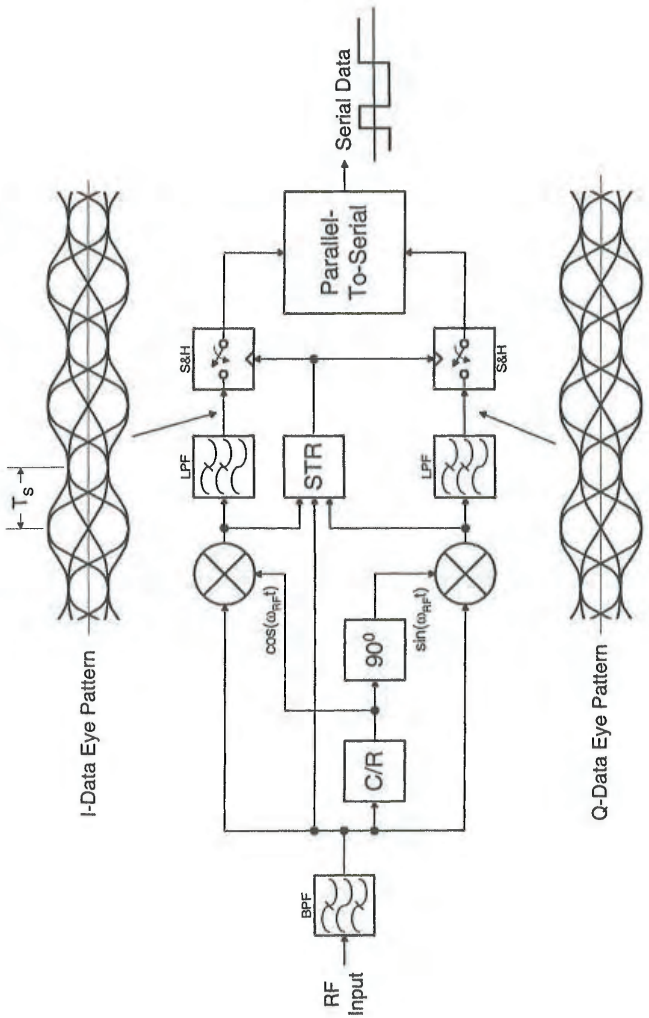


Figure 4.101 $\pi/4$-DQPSK coherent demodulator.

Q can be 0, +1 or −1 (three-level eye pattern). If the received symbol lies off the I and Q axes, the values of I and Q can be ±0.707 (two-level eye pattern). A disadvantage of direct coherent demodulation is that the resulting carrier recovery (C/R) and symbol timing recovery (STR) circuitry becomes rather complex.

Figure 4.102 shows a block diagram of a much simpler differentially coherent $\pi/4$-DQPSK demodulator. Notice that no carrier recovery circuit is needed in this implementation. An analog delay is used to produce a version of the received signal, delayed by one bit period. This delayed "reference" is used as the in-phase double-balanced mixer carrier input, while a $\pi/2$ shifted version is used as the quadrature mixer carrier input. The use of an analog delay greatly reduces the complexity of the demodulator. Lowpass filters are used to filter out high-frequency signals resulting from the mixing operation. An advantage of the differentially coherent demodulator is that the analog I and Q outputs are standard two-level eye patterns, as shown in the figure. Parallel bit streams can be easily obtained from the resulting two-level I and Q outputs via a pair of bit-slicing circuits (analog comparators whose decision level is adjusted to be in the center of the eye). A parallel-to-serial converter (multiplexer) is then used to interleave the I and Q data back into the original bit stream. Although the P_E of the differentially-coherent receiver is worse than the fully coherent demodulator, it is more robust in high-Doppler, fast-fading channels.

A major disadvantage of $\pi/4$-DQPSK is the need for linear amplifiers in the transmit chain. Most power amplifiers exhibit some degree of nonlinearity. In general, maximum efficiency is obtained from power amplifiers when the output signal is close to the maximum power the amplifier can produce (i.e., when the amplifier is at or near saturation). Unfortunately, the closer the output gets to the saturation point of the amplifier, the more distorted the signal becomes.

One way to overcome this problem is to use complex pre or postdistortion circuitry, which cancels out most of the distortion products of the amplifier, even when the power amplifier is run close to saturation. Figure 4.103 shows a block diagram of the predistortion circuitry used by Stapleton et al. [22], which is based on a slowly adapting feedback system. The "clean" $\pi/4$-DQPSK modulated input signal, $y_m(t)$, is applied to a quadrature modulator whose inputs are derived from $y_m(t)$ and a pair of correction signals. The complex modulator produces a distorted version of $y_m(t)$ and feeds this predistorted signal, $v_d(t)$, to the nonlinear power amplifier's input. The distorted transmitted output is applied to another mixer/bandpass filter combination, which extracts the out-of-band component. The action of the feedback loop produces a version of $v_d(t)$ that contains distortion products that minimize out-of-band spectral components out of the power amplifier.

Figure 4.104 shows a plot of the transmitted output spectrum with and without predistortion circuitry in the transmitter. As the figure shows, spectral regrowth can be kept below −60 dBc at the output when distortion canceling circuitry is used. Unfortunately, this solution results in higher complexity and, ultimately, higher cost.

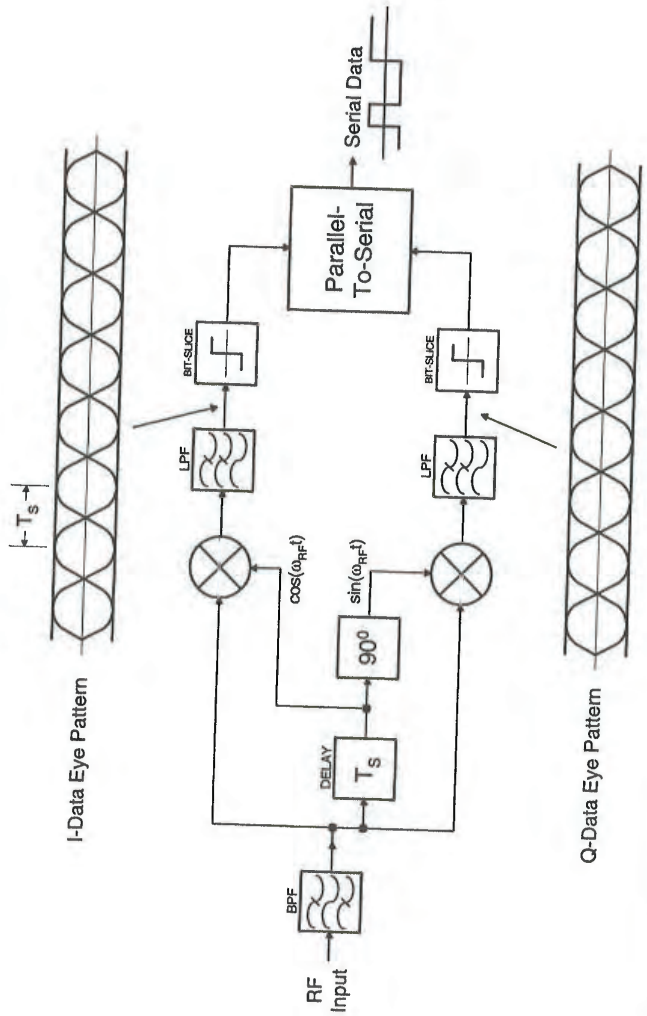


Figure 4.102 $\pi/4$-DQPSK differentially coherent demodulator using analog delay.

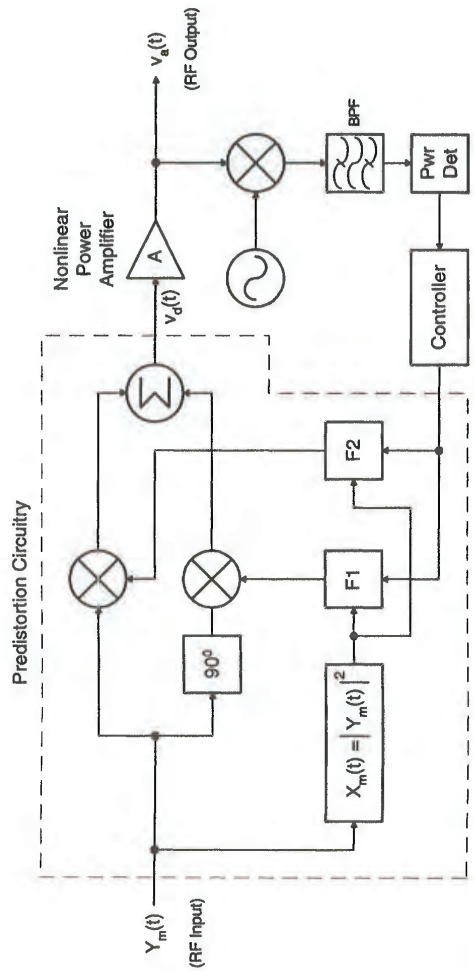


Figure 4.103 Predistortion circuitry to prevent spectral regrowth at power amplifier output for $\pi/4$-DQPSK [22].

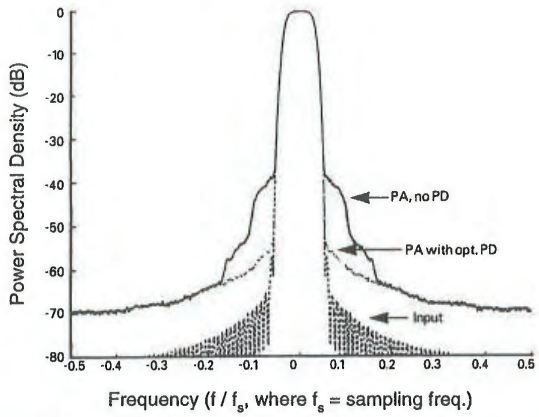


Figure 4.104 $\pi/4$-DQPSK transmit power spectral density with and without predistortion circuitry. (*Source:* Stapleton [22].)

As expected, the P_E of $\pi/4$-DQPSK is worse than that of QPSK. Figure 4.105 shows a plot of the probability of bit error versus E_b/N_0 for coherently demodulated $\pi/4$-DQPSK. For a probability of bit error of 10^{-4}, $\pi/4$-DQPSK is worse than QPSK by 0.7 dB, which is generally considered a small price to pay compared to the advantages $\pi/4$-DQPSK has to offer.

4.4.3 Feher's QPSK (FQPSK)

GMSK possesses the attractive features of a very narrow transmitted output spectrum, little ISI (at least for BT = 0.5), and a constant envelope. Other digital modulations possess these same features as well. One such example *is* Feher's QPSK (FQPSK). FQPSK was developed (and patented) by Dr. Kamilo Feher, now at the University of California, Davis.

Figure 4.106 shows a block diagram of an FQPSK modulator. The FQPSK modulator is very similar to an MSK modulator. The input data stream is demultiplexed into I and Q waveforms, each now at half the original data rate, with the Q signal delayed by a single bit period. Each signal is then sent to special pulse-shaping filters, which are the key to FQPSK modulation. After quadrature modulation, the output of the summing junction is fed to a hardlimiting circuit to produce a constant envelope signal.

The FQPSK pulse-shaping filters have the following impulse response

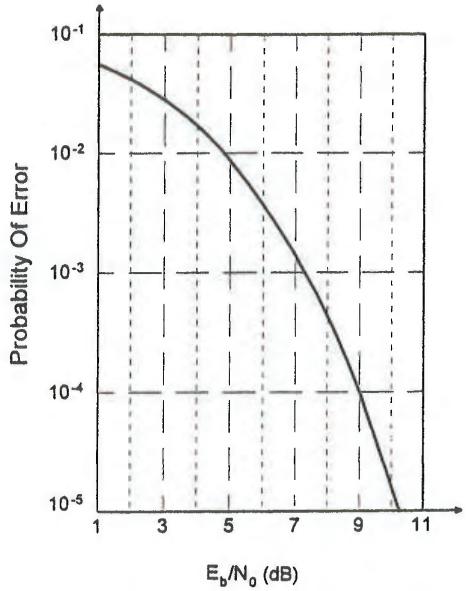


Figure 4.105 $\pi/4$-DQPSK probability of bit error.

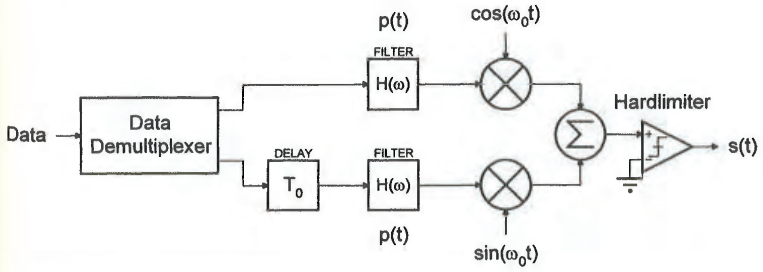


Figure 4.106 FQPSK modulator.

$$p(t) = 0.5 \cdot \left[1 + \cos\left(\frac{\pi t}{T_b}\right)\right]; \; |t| \leq T_b \tag{4.92}$$

$$p(t) = 0; \text{ elsewhere} \tag{4.93}$$

As (4.92) and (4.93) show, the filters produce a cosine shaped pulse response on each data bit. This response is shown in Figure 4.107. Notice that the pulse extends over two bit periods. Figure 4.108 shows a block diagram of the pulse-shaping filter, which is implemented as an FIR lowpass filter having tap weights given by (4.92) and (4.93) above. Since the input to each of the filters is only a single bit, the complexity of the filters is greatly reduced.

Figure 4.109 shows the time domain response for the FQPSK modulator. The upper two traces in the figure are the demultiplexed I and Q data streams after passing through the pulse-shaping filters (and delayed by one bit period for the Q

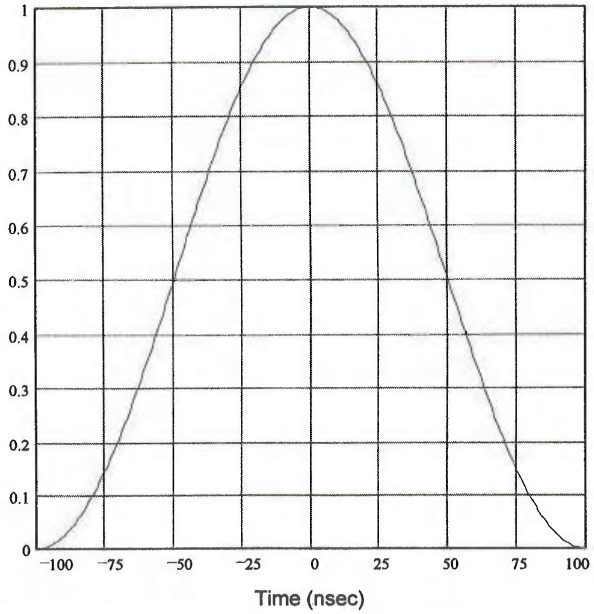
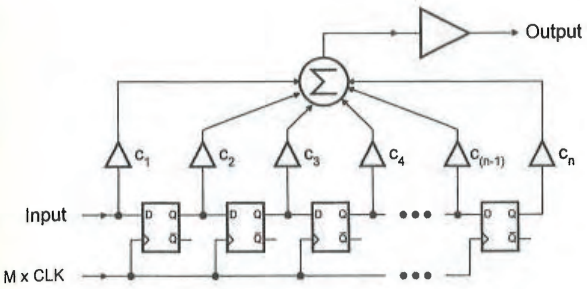


Figure 4.107 FQPSK premodulation filter impulse response (for 10-Mbps data).



FIR-Based FQPSK Filter

Figure 4.108 FIR-based FQPSK premodulation filter.

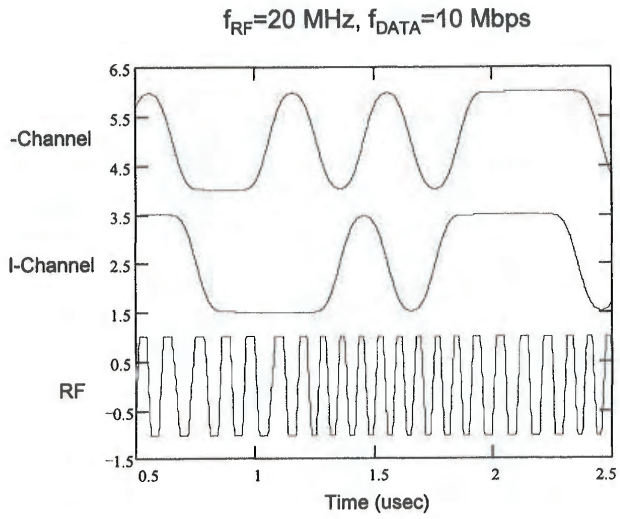


Figure 4.109 FQPSK time domain waveforms.

signal). The lower trace shows the output of the hardlimiter. It is interesting to note that the output of the summing junction is not constant envelope. Therefore, a hardlimiting circuit is required in order to produce a constant envelope signal. However, unlike QPSK, which experiences significant spectral regrowth when nonlinearly amplified, the FQPSK modulator summing junction output does not exhibit spectral regrowth. This is due to the action of the pulse-shaping filters feeding the quadrature modulator. A virtue of having a constant envelope signal at the output of the modulator is that we need not rely on the power amplifier in saturation (compression) to produce a constant envelope signal. This avoids AM-to-PM modulation errors, which can be caused by the power amplifier's nonlinear phase versus input signal-level response.

Figure 4.110 shows eye diagrams for both the I and Q data streams. Notice that not only is there no ISI at the sampling instants, but the zero crossings are jitter-free. The absence of ISI at the sampling instants greatly improves the P_E performance of the received signal. The absence of jitter on the signal causes less uncertainty in the sampling instants and also enhances P_E performance.

Figure 4.111 shows a plot of the power spectral density of FQPSK versus that of GMSK, with BT = 0.5. Notice that FQPSK falls off very quickly in the first 20 to 30 dB of the response. In addition, the main lobe of the FQPSK signal is narrower than that of the GMSK signal. Below the –30 dB point, GMSK falls off more rapidly than FQPSK. FQPSK is a very attractive modulation to consider in applications such as satellite transponders. The rapid roll-off in the first 20 to 30 dB, or so, helps to reduce ACI when all channels are received at approximately the same power level.

Figure 4.112 shows a block diagram of the FQPSK demodulator, which is identical to that of the MSK demodulator. Figure 4.113 shows a plot of probability of bit error versus E_b/N_0 for coherently demodulated FQPSK. The performance of FQPSK is only slightly worse than that of BPSK or QPSK.

FQPSK is a true phase modulation. If a string of 0s or 1s are input to the modulator, the resulting output is a steady-state *phase-shifted* version of the carrier. MSK or GMSK, in contrast, produce steady-state *frequency* differences for long strings of 0s or 1s on the input of the modulator, and can be noncoherently demodulated with simple circuitry. Depending on the particular application, this may or may not be important.

4.5 EXAMPLES OF INTEGRATED CIRCUIT DIGITAL COMMUNICATIONS CIRCUITS

Figure 4.114 shows a block diagram for a typical wireless digital communications system—the GEC/Plessey wireless local area network (WLAN) transceiver [23–25]. This transceiver uses frequency-hopped spread spectrum techniques. It operates in the 2.4-GHz industrial, scientific, and medical (ISM) band at a data rate of 625

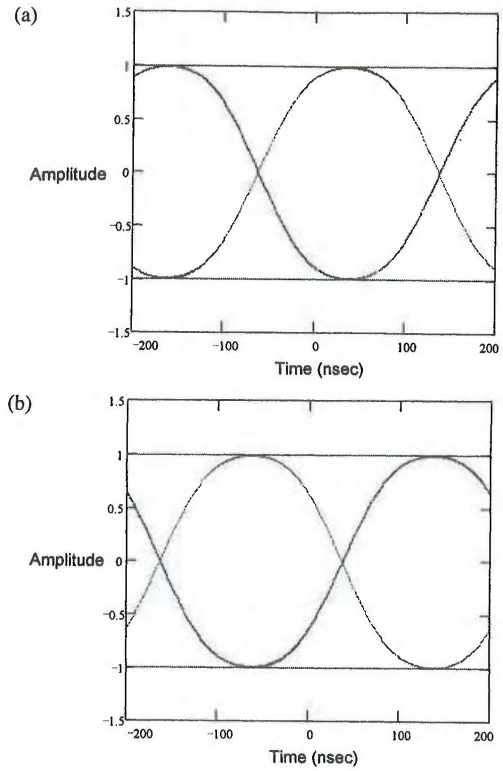


Figure 4.110 Coherently demodulated FQPSK waveforms: (a) I-channel, (b) Q-channel.

Kbps. The transceiver employs GFSK modulation. This transceiver was mentioned earlier in Chapter 3.

The circuit is divided into six main ICs. The first IC contains a low-noise amplifier (LNA), downconverter, power amplifier, upconverting mixer, RF switches, and a 2-GHz oscillator. This IC is fabricated in a Gallium Arsenide (GaAs) MESFET process and performs most of the RF front-end functions. Notice that this chip has a large number of RF I/O lines coming in and out of the chip, which can cause unwanted coupling and crosstalk problems if care is not taken in circuit partitioning.

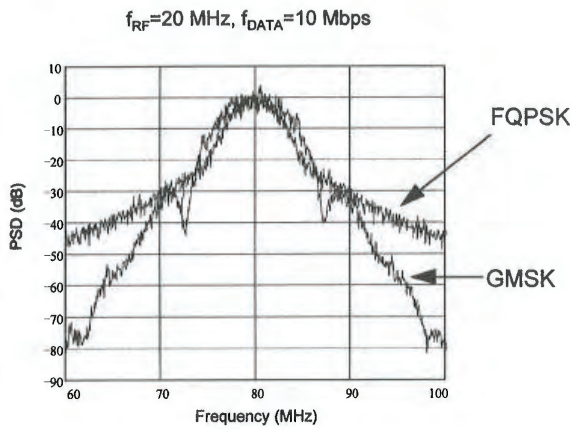


Figure 4.111 FQPSK versus GMSK power spectral density.

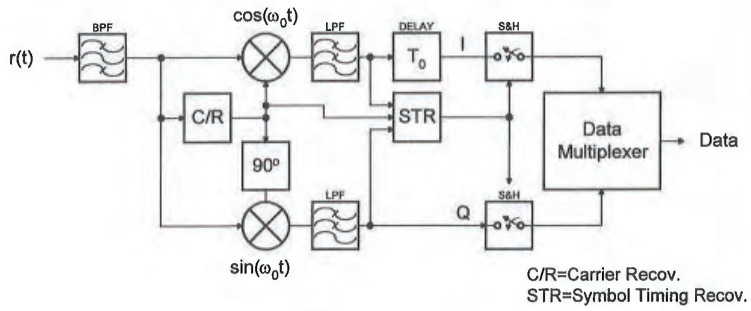


Figure 4.112 FQPSK demodulator.

GaAs is a good choice for this chip as it offers superior noise performance (i.e., low noise figure), as well as high-efficiency power amplifiers and low-loss switches.

After the first downconversion, the signal is passed to the IF and baseband processing circuits. Two additional ICs are fabricated using a high-speed silicon IC process with an f_T of 13 GHz. The first IC contains the second downconverting mixer, IF amplifiers, and a 312-MHz oscillator. This includes a received signal

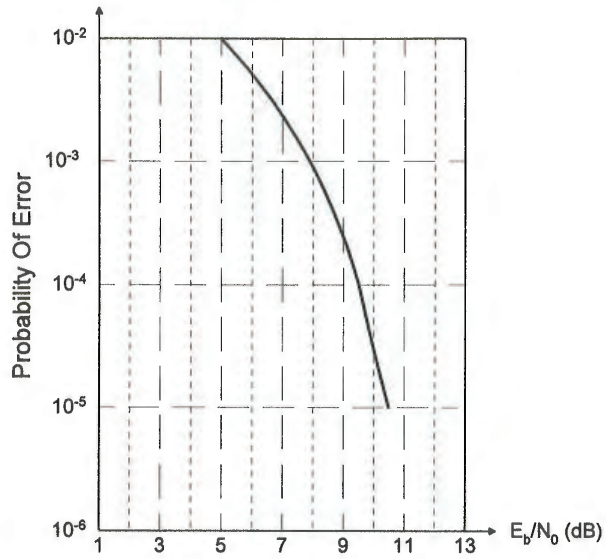


Figure 4.113 FQPSK probability of bit error.

strength indicator (RSSI) for measuring the level of the received signal as well as a quadrature detector to demodulate the GFSK signal. A second silicon bipolar IC contains another oscillator at 700 MHz and a number of high-speed digital dividers and prescalers that interface the synthesizer chips. Silicon bipolar is a good technology to use in the IF and baseband portion of the radio because of its high level of functionality and high yield. The use of a high-speed process greatly reduces the amount of power required for each function.

A number of PLLs are also shown in Figure 4.114. These circuits are fabricated in CMOS. The use of high-speed dividers in the bipolar chips allows the CMOS PLLs to run at much lower speeds, thus conserving power.

Prior to integration, the transceiver required 40 ICs, 200 passive elements and measured 7 inches by 5 inches. After integration (i.e., the system just described), the transceiver requires only 4 ICs and 50 passive elements and measures only 3 inches by 2 inches.

Another example of a digital wireless communication system is shown in Figure 4.115. This is the National Semiconductor DECT transceiver [26, 27]. The

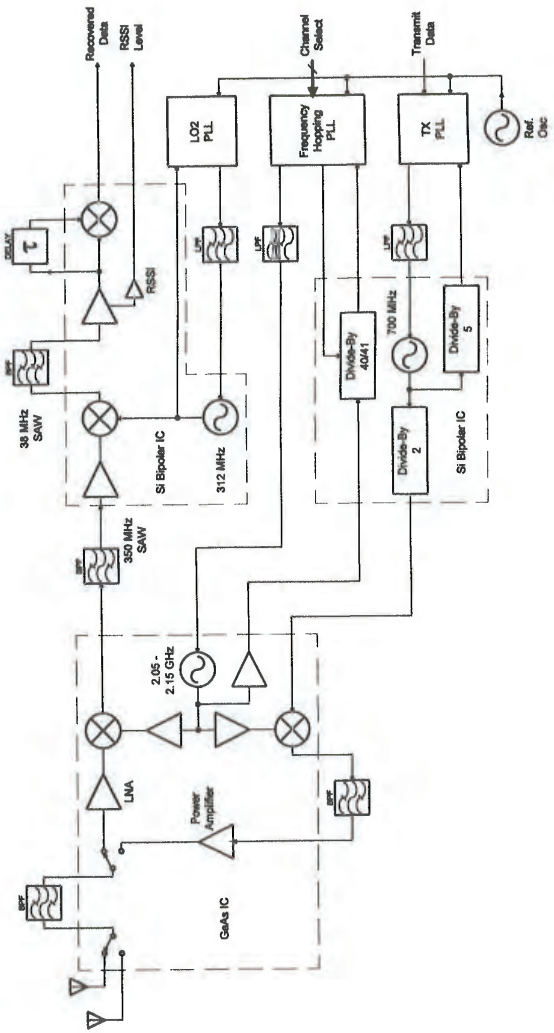


Figure 4.114 GEC Plessey WLAN transceiver.

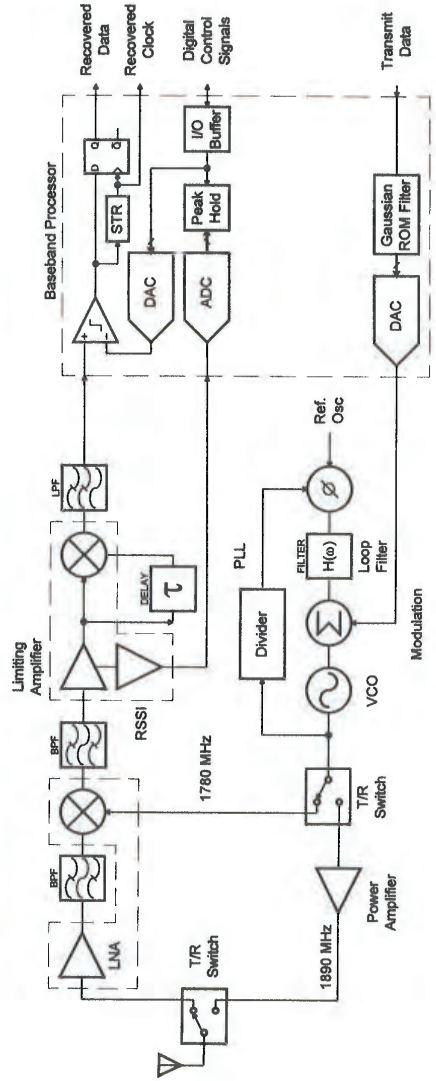


Figure 4.115 National Semiconductor DECT transceiver.

transceiver operates at a carrier frequency of 1.88 to 1.90 GHz with a data rate of 1.152 MHz. It is interesting to note that with the exception of the power amplifier and transmit/receive (T/R) switch, the entire transceiver is fabricated in National Semiconductor's BiCMOS silicon bipolar IC process. Both the power amplifier and T/R switch are fabricated in a GaAs MESFET process.

The downconverter has a relatively high total noise figure of 8.7 dB. However, this is sufficient to meet the DECT specifications. DECT uses GMSK modulation with a relative filter bandwidth, BT, of 0.5. The Gaussian data filter is done on-chip using a read-only memory (ROM) look-up table that drives a DAC, whose output is then filtered with a simple RC lowpass filter. The entire transceiver operates on only 3V.

4.5.1 Design of Modulators and Demodulators for Wireless Communications Systems

Let's turn our attention now to how some of the modulators and demodulators are actually designed in wireless communication systems. Recall that in 8-PSK we represent groups of three bits by eight different phase angles spaced uniformly about the unit circle on 45-deg increments. Figure 4.116 shows how we can use unit increments

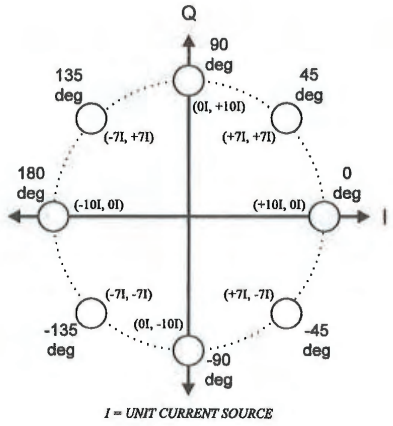


Figure 4.116 8-PSK phase domain showing symbols based on unit-weight currents (unit-weight currents for I, Q shown in parentheses).

of current or voltage to obtain our desired phase increments and yet still lie on the unit circle. Recall from our discussions on quadrature modulation

$$\theta = \arctan\left[\frac{Q}{I}\right] \tag{4.94}$$

$$A = \sqrt{I^2 + Q^2} \tag{4.95}$$

For example, let $I = 7$ and $Q = 7$. Then $q = 45$ deg and $A = 9.899$. If we let $I = 0$ and $Q = 10$, then $q = 90$ deg and $A = 10$. So we see that by using unit increments of current or voltage to derive the I and Q signals, we can get exactly the phase with less than 0.1-dB amplitude error! Furthermore, by using unit increments we can ensure this relation is maintained over process and temperature variations. Symbols that fall on the negative side to either the I or Q axis can be derived by simply inverting the polarity of the I and Q signals.

Figure 4.117 shows a block diagram of the 8-PSK modulator. The NRZ data and clock are applied to the data encoder portion of the chip, which produces the appropriate differential driver signals for the modulator. The modulator is driven

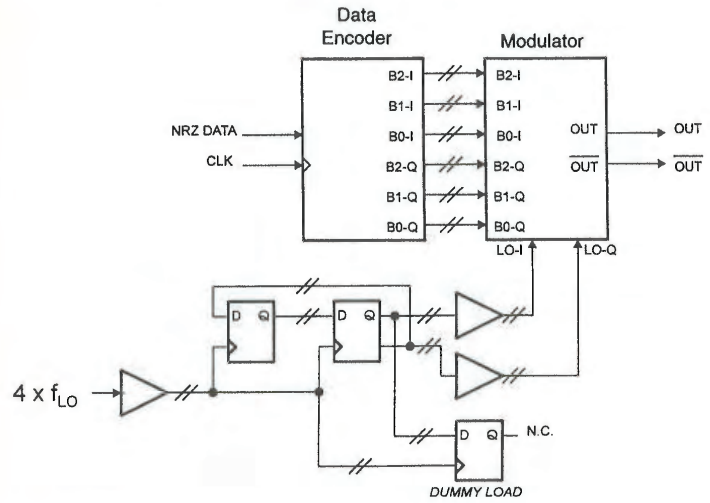


Figure 4.117 Data encoder and digital I/Q generator for 8-PSK IC modulator.

by a digital quadrature generator. Notice that all digital lines are differential in nature. This reduces the amount of digital noise radiated and received within the chip. The quadrature generator uses a 4 × frequency clock. Notice the dummy gates within the digital quadrature generator. These dummy gates ensure that each of the critical nodes is equally loaded to ensure process and temperature invariant operation.

Figure 4.118 shows a schematic of the quadrature modulator itself. Notice the symmetry of the circuit. This helps to ensure tight balance between the *I* and *Q* signals. The *I* and *Q* signals are derived from identical, switched, unit current sources. A common bias line (VREF) ensures that all current sources are equally biased. Matching between each unit current source can be further improved by using a

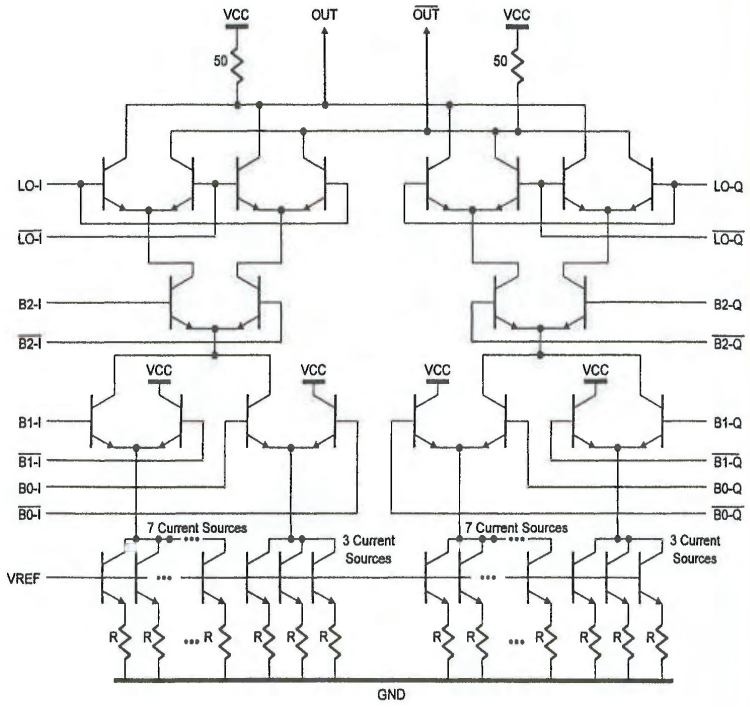


Figure 4.118 Schematic for monolithic 8-PSK modulator.

relatively large voltage drop across each emitter resistor. This helps to cancel out any V_{BE} offset voltage variations. In order to ensure high-speed operation, currents are simply steered from one portion of the circuit to another. All I/O lines are differential in order to improve noise margin and reduce crosstalk. Digital signal swings are kept to 250 mV to reduce slew limiting and unwanted coupling. Current sources are switched in units of 0, 7, or 10. Negative values of *I* and *Q* are produced by simply steering the current to the other side of the circuit. The LO input need only be large enough to fully switch the upper differential pairs. Usually, 100 to 200 mV is sufficient. The summing junction of the quadrature modulator is implemented by simple current summations at the collectors of the LO switching pairs.

Figure 4.119 shows a layout of the modulator. All critical differential pairs are quad-ed to reduce thermal and process gradients. All resistors are made from the same unit size and weight values to ensure matching. The overall layout is completely symmetric. The current sources which are tied together in groups of 7 or 3 (for 3 + 7 = 10) and are interspersed with one another for greater uniformity.

This concept can be taken one step further. The same unit current/voltage approach can be used to design a 16-PSK modulator, as shown in Figure 4.120 [28]. Groups of 4 bits are represented by one of 16 different symbols spaced uniformly about the unit circle at 22.5-deg increments. As mentioned earlier, 16-PSK reduces the symbol rate of BPSK by 4, making the resulting spectrum only one-quarter as wide. Figure 4.121 shows a schematic of the modulator used in Hewlett Packard's [28] version of their 16-PSK modulator, which contains the DACs, *I* and *Q* mixers, and summing junction in one circuit. As in the case of the 8-PSK circuit, the LO signals are derived digitally (though not shown here). This circuit is fabricated in a high-speed silicon bipolar IC process with an f_T of 13 GHz, which is ideal for obtaining the combination of speed, matching, and level of complexity required to meet the design goals. Notice the similarity of this circuit compared to the 8-PSK modulator. Here, currents are grouped in to units of 5, 4, 3, 1, and 0. The phase error of this circuit is only 0.12 deg and the amplitude variation is only 0.18 dB. The modulator is capable of operating at 110 MSymbols/sec with an IF output frequency of 400 MHz.

A more general modulator is shown in Figure 4.122 [29, 30]. This circuit is fabricated in a 0.5-μm GaAs MESFET process. It is a general purpose quadrature modulator that can accept modulation inputs from dc to 500 MHz and LO inputs from 700 MHz to 3 GHz. In this circuit, the 0-deg and 90-deg versions of the LO are derived from the outputs of two passive RC filters. Active FET resistors are used to trim the phase to its final value. Variable amplitude drivers are used to match the *I* and *Q* LO amplitudes before they are applied to the Gilbert-cell mixers. The currents at the mixer outputs are summed together and applied across resistive loads. A differential to single-ended driver is used to provide a back-terminated 50-ohm output.

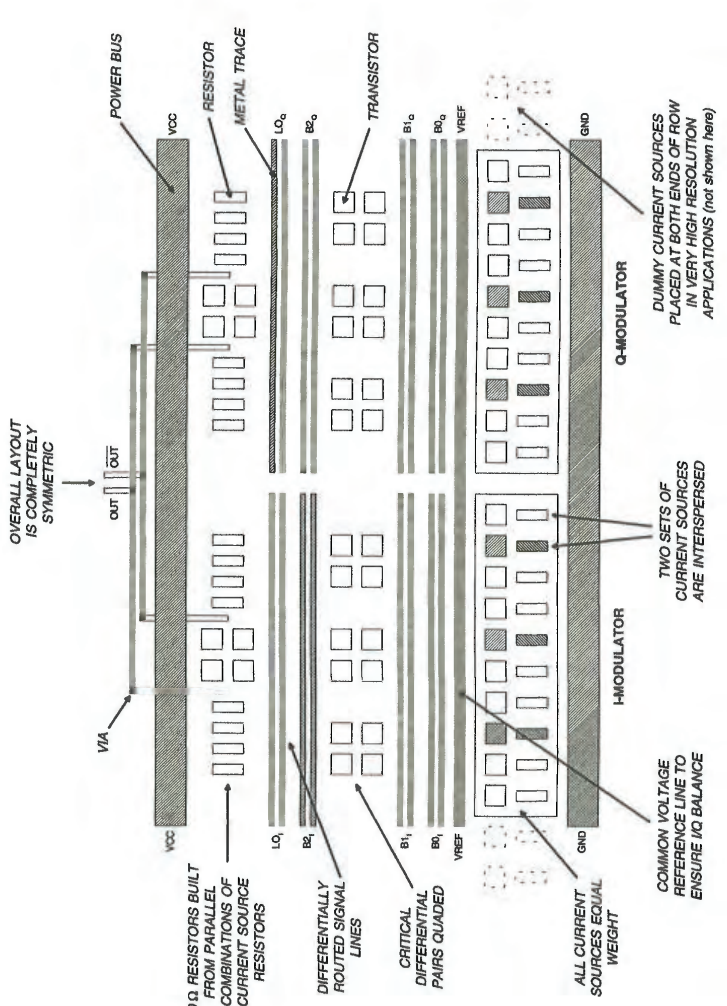


Figure 4.119 Layout for monolithic 8-PSK modulator.

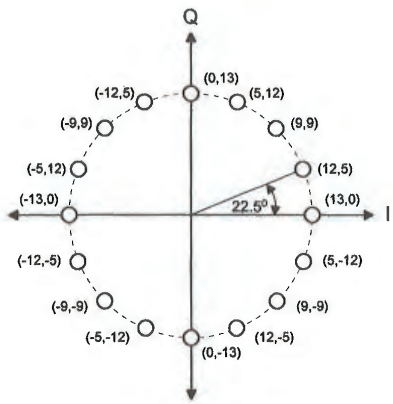


Figure 4.120 16-PSK phase domain showing symbols based on unit-weight currents (unit-weight currents for I, Q shown in parentheses).

4.6 SUMMARY

Modern digital wireless communication systems are requiring increasingly sophisticated modulation/demodulation circuits in order to meet customer and regulatory demands. Many different modulation techniques are available. The best choice is always determined by the constraints of the wireless system designer. We have examined both phase and frequency modulation schemes as well as a few of their derivatives. The use of integrated solutions is becoming more and more prevalent as the level of complexity of today's wireless systems increases and as the demand by the consumer for lower cost continues.

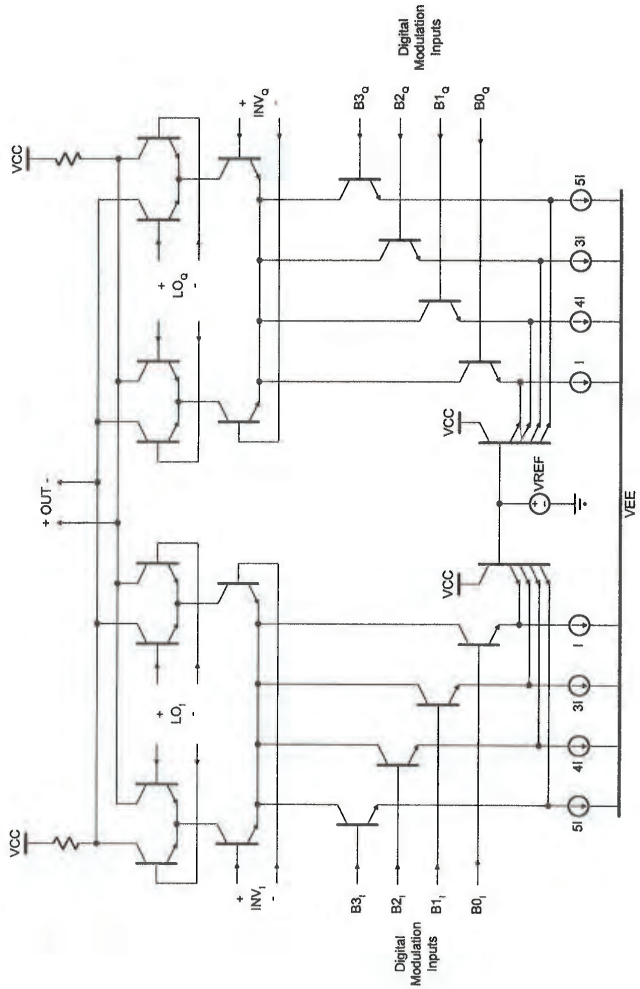


Figure 4.121 Schematic for monolithic 16-PSK modulator (Hewlett Packard).

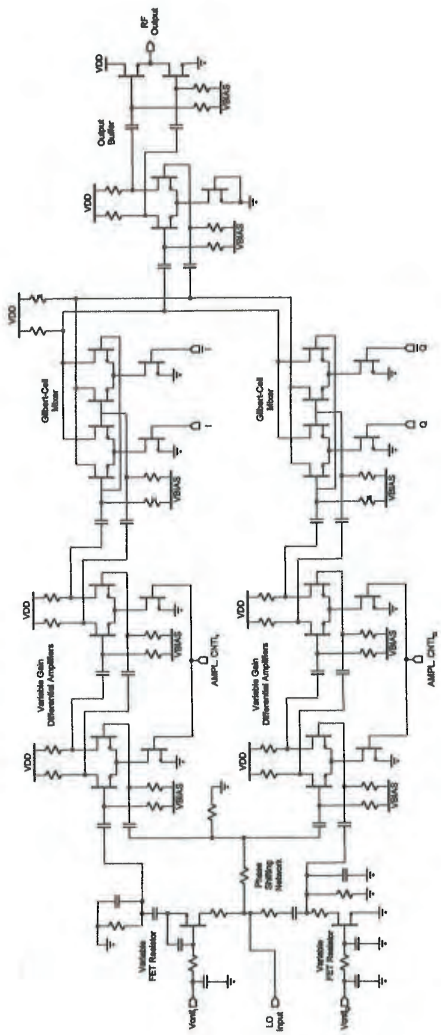


Figure 4.122 Schematic for GaAs I/Q direct IC modulator (Alcatel).

References

[1] Lathi, B. P., *Modern Digital and Analog Communication Systems*, Holt, Rinehart and Winston, New York, NY 1983.

[2] Feher, K., *Digital Communications; Satellite/Earth Station Engineering*, Prentice-Hall, Englewood Cliffs, NJ 1983.

[3] Feher, K., *Advanced Digital Communications; Systems and Signal Processing Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[4] Pasupathy, S., "Minimum Shift Keying: A Spectrally Efficient Modulation," *IEEE Communications Magazine*, July 1979, pp. 14–22.

[5] Samueli, H., "Digital Wireless Transceiver Architectures," ISSCC94 Short Course, San Francisco, CA, February 1994.

[6] Lindsey, W. C. and Simon, S. K., *Telecommunication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[7] Sklar, B., *Digital Communications: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[8] Feher, K., *Wireless Digital Communications: Modulation and Spread Spectrum Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[9] Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*, Wiley, New York, NY, 1965, Chapter 7.

[10] Nyquist, H., "Certain Topics in Telegraph Transmission Theory," *AIEE Trans.*, Vol. 47, April 1928, p. 817.

[11] Jeruchim, M. C., Balaban, P, and Shanmugan, K. S., *Simulation of Communication Systems*, Plenum Press, New York, NY 1992.

[12] Zverev, A. I., *Handbook of Filter Synthesis*, John Wiley & Sons, New York, NY, 1967.

[13] Murota, K., et. al., "GMSK Modulation for Digital Radio Telephony," *IEEE Transactions On Communications*, Vol. COM-29, No. 7, July 1981, pp. 1044–1050.

[14] Socci, J., "GMSK As A Modulation Scheme For A Frequency Hopped PHY," *IEEE 802.11 Wireless Access Method and Physical Layer Specifications*, July 1993.

[15] Rohde and Schwartz, Inc., "Signal Generators For Modern Communication Techniques," August 1990.

[16] Baker, P. A., "Phase-Modulation Data Sets for Serial Transmission at 2000 and 2400 Bits Per Second," Part I, *AIEE Trans. on Communications Electronics*, July 1962, pp. 166–171.

[17] Guo, Y., and Feher, K., "Modem/Radio IC Architectures For ISM Band Wireless Applications," *IEEE Transactions On Consumer Electronics*, Vol. 39, No. 2, May 1993, pp. 100–106.

[18] Chennakeshu, S. and Saulnier, G. J., "Differential Detection Of $\pi/4$-Shifted-DQPSK For Digital Cellular Radio," *IEEE Transactions on Vehicular Technology*, Vol. 42, No. 1, February 1993, pp. 46–53.

[19] Feher, K., and Mehdi, H., "Modulation/Microwave Integrated Digital Wireless Development," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 7, July 1995.

[20] McGrath, F., et. al., "A 1.9-GHz GaAs Chip Set for the Personal Handyphone System," *IEEE Transactions on Microwave Theory and Technique*, Vol. 43, No. 7, July 1995.

[21] Goode, S. H. Kazecki, and D. W. Dennis, "A Comparison of Limiter-Discriminator, Delay, and Coherent Detection for $\pi/4$-QPSK," *Proc. IEEE Veh. Technol. Conf.*, Orlando, FL, May 1990, pp. 687–694.

[22] Stapleton, S. P., Kandola, G. S., and Caver, K., "Simulation and Analysis of an Adaptive Predistorter Utilizing a Complex Spectral Convolution," *IEEE VT*, Vol. 41, Nov. 1992, pp. 387–395.

[23] Williams, D. A., "A Frequency Hopping Microwave Radio System For Local Area Network Communications," 1993 IEEE MTT-S Symposium, Atlanta, GA, June 14–15, 1993.

[24] Williams, D. A., "An Integrated Microwave Radio Transceiver For WLAN Applications," 1993 Wireless Symposium, San Jose, CA.

[25] Devlin, L. M., et. al., "A 2.4 GHz Single Chip Transceiver," IEEE 1993 Microwave and Millimeter-Wave Monolithic Circuits Symposium, Atlanta, GA, June 14–15, 1993.

[26] Madsen, B. and Fague, D. E., "Radios For The Future: Designing For DECT," *RF Design*, April 1993, pp. 48–53.

[27] Fague, D. E., "A Fully Integrated Modulator/Demodulator For DECT," RF Expo West, San Jose, CA, March 17–19, 1993.

[28] Wuppermann, B., et al., "A 16-PSK Modulator With Phase Error Corrections," *ISSCC93*, San Francisco, CA, February 1993.

[29] Boveda, A., et. al., "A 07–3 GHz GaAs QPSK/QAM Direct Modulator," ISSCC93, San Francisco, CA, February 1993.

[30] Boveda, A., et. al., "A 07–3 GHz GaAs QPSK/QAM Direct Modulator," *IEEE Journal Of Solid State Circuits*, Vol. 28, No. 12, December 1993.

INTEL 1005 3 OF 3

CHAPTER 5

MIXERS FOR WIRELESS APPLICATIONS

Stephen A. Maas
Nonlinear Technologies, Inc.
Long Beach, CA

5.1 INTRODUCTION

In the past twenty years, mixer technology has evolved from a major research topic into one of great maturity. The venerable diode mixer is now well understood and diode mixers exhibiting excellent performance over multioctave bandwidths are commonly available at low cost. Although their designs are somewhat less mature, active and passive FET mixers have special advantages in minimizing the complexity of a system or in reducing noise or intermodulation distortion.

Wireless applications create special needs for mixers. Because most wireless systems are commercial, they are manufactured in large quantities and must have low cost. Many wireless components are used in handheld and battery-powered systems, where small size and low dc power are often essential. Electrical performance is usually secondary to cost, although frequently one or two performance characteristics are critical. As with military and aerospace components, thermal stability and mechanical ruggedness are usually also important, but in wireless systems must be achieved at much lower expense. The role of integrated circuits (ICs) is not as clear in wireless as in other systems; certainly, silicon ICs are a proven, low-cost technology, but the cost effectiveness of more advanced technologies such as GaAs and heterojunction monolithics is somewhat less certain.

Most wireless applications are below 2.5 GHz, so the emphasis of this chapter will be largely in this frequency range. Wireless applications in the higher microwave frequencies are not uncommon, however, and we do not neglect them entirely. Low-frequency operation does not necessarily make life easier for the mixer designer; one of the most difficult problems posed by low-frequency operation is to make low-frequency components, especially mixer baluns, small enough for such applications as wireless telephones and handheld transceivers.

This chapter will cover all of the fundamental issues of mixer design, including preferred mixer topologies, choice of semiconductor devices, and passive component design for mixer implementation.

5.2 PROPERTIES OF MIXERS

Mixers have a variety of properties that will be discussed in this section.

5.2.1 Frequency Mixing

A mixer is fundamentally a multiplier. An ideal mixer multiplies a signal by a sinusoid, shifting it to both a higher and lower frequency, and selects one of the resulting sidebands. A modulated narrowband signal, usually called the *RF signal*,[1] represented by

$$S_{RF}(t) = a(t)\sin(\omega_s t) + b(t)\cos(\omega_s t) \tag{5.1}$$

is multiplied by the function called the *local oscillator* (LO) signal

$$f_{LO}(t) = \cos(\omega_p t) \tag{5.2}$$

to obtain the *IF signal*[2]

$$S_{IF}(t) = \frac{1}{2}a(t)(\sin((\omega_s + \omega_p)t) + \sin((\omega_s - \omega_p)t)) + \frac{1}{2}b(t)(\cos((\omega_s + \omega_p)t) + \cos((\omega_s - \omega_p)t)) \tag{5.3}$$

[1]RF was originally an abbreviation for radio frequency. Like many technical abbreviations and acronyms (e.g., radar, laser, FET), the term has evolved into a word in itself, and has lost most of its original meaning. Thus, we regularly use terms like RF frequency to distinguish from RF port, RF voltage, etc. Some of these terms may seem redundant or illogical to grammatical bluestockings who insist that RF have only its original meaning. For better or worse, language evolves; this is an example.

[2]IF originally meant intermediate frequency. The above footnote applies here as well.

In the ideal mixer, two sinusoidal IF components, called *mixing products*, result from each sinusoid in $s(t)$. In receivers, the difference-frequency component is usually desired, and the sum-frequency component is rejected by filters.

Unfortunately, real mixers do not follow this neat description. Even if the local oscillator voltage applied to the mixer's LO port is a clean sinusoid, the nonlinearities of the mixing device distort it, causing the LO function to have harmonics. Those nonlinearities can also distort the RF signal, resulting in RF harmonics. The IF is, in general, the combination of all possible mixing products of the RF and LO harmonics. Filters are usually used to select the appropriate response. The other so-called *spurious responses* must be eliminated, usually by filters.

Every mixer, even an ideal one, has a second RF frequency that can create a response at the IF. This is a type of spurious response, and is called the *image*; it occurs at the frequency $2f_{LO} - f_{RF}$. For example, if a mixer is designed to convert 10 GHz to 1 GHz with a 9-GHz LO, the mixer will also convert 8 GHz to 1 GHz at the same LO frequency. Although none of the types of mixers we shall examine inherently reject images, it is possible to create combinations of mixers and hybrids that do reject the image response.

It is important to note that the process of frequency shifting, which is the fundamental purpose of a mixer, is a linear phenomenon. Although nonlinear devices are invariably used for realizing mixers, there is nothing in the process of frequency shifting that requires a nonlinearity. Distortion and spurious responses other than the sum and difference frequency, which are often severe in mixers, are not fundamentally required by the frequency-shifting operation that a mixer performs.

5.2.2 Practical Characteristics of Mixers

Conversion Efficiency

Mixers using Schottky-barrier diodes are passive components and consequently exhibit conversion loss. This loss has a number of consequences: the greater the loss, the higher the noise of the system, and the more amplification that is needed. High loss contributes indirectly to distortion because of high signal levels that result from the additional preamplifier gain required to compensate for this loss. It also contributes to the cost of the system since the necessary low-noise amplifier stages are usually expensive.

Mixers using active devices often (but not always) exhibit conversion gain. High mixer gain is not necessarily desirable because it reduces stability margins and can increase distortion. Usually, a mixer gain of unity, or at most a few dB, is ideal.

Noise

In a passive mixer whose image response has been eliminated by filters, the noise figure is usually equal to, or only a few tenths of one decibel above, the conversion

loss. In this sense, the mixer behaves as if it were an attenuator having a temperature equal to or slightly above the ambient.

In active mixers, the noise figure cannot be related easily to the conversion efficiency; in general, it cannot even be related qualitatively to the device's noise figure when used as an amplifier. The noise figure of an active FET mixer depends strongly on the characteristics of the design; these are discussed in Section 5.4.

Bandwidth

The bandwidth of a diode mixer is limited by the external circuit, especially by the hybrids or baluns used to couple the RF and LO signals to the diodes. In active mixers, bandwidth can be limited either by the device or by hybrids or matching circuits that constitute the external circuit; much the same factors are involved in establishing active mixers' bandwidths as amplifiers' bandwidths.

Distortion

It is a truism to say that everything is nonlinear to some degree and generates distortion. Unlike amplifiers or passive components, however, mixers often employ strongly nonlinear devices to provide mixing. Because of these strong nonlinearities, mixers generate relatively high levels of distortion. A mixer is usually the dominant distortion-generating component in a receiver.

Distortion in mixers, as with other components, is manifested as intermodulation distortion, (IM or IMD), which involves mixing between multiple RF tones and harmonics of those tones. If two RF excitations f_1 and f_2 are applied to a mixer, the nonlinearities in the mixer will generate a number of new frequencies, resulting in the IF spectrum shown in Figure 5.1. Figure 5.1 shows all intermodulation products up to third order; by nth order, we mean all n-fold combinations of the excitation

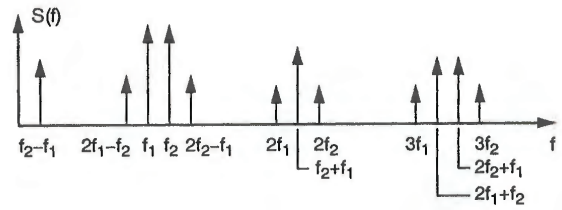


Figure 5.1 IF spectrum of intermodulation products up to third order. The frequencies f_1 and f_2 are the excitations.

tones (not including the LO frequency). In general, an nth order nonlinearity gives rise to nth (and lower) order distortion products.

An important property of intermodulation distortion is that the level of the nth-order IM product varies n decibels for every decibel change in the levels of the RF excitations. The extrapolated point at which the excitation and IM levels are equal is called the nth-order IM intercept point, abbreviated IP_n. This dependence is illustrated in Figure 5.2. In most components, the intercept point is defined as output power; in mixers it is traditionally input power.

Given the intercept point IP_n and input power level, the IM level P_{IM} can be found from

$$P_{IM} = nP_{lin} - (n - 1)IP_n \tag{5.4}$$

where P_{lin} is the level of each of the linear RF tones (which are assumed to be equal). By convention, P_{lin} and P_{IM} are the powers of a single frequency component, not the total power of all components. The intercept point can be defined in terms of either input or output power; usually, input power is used for mixers and output power for other types components.

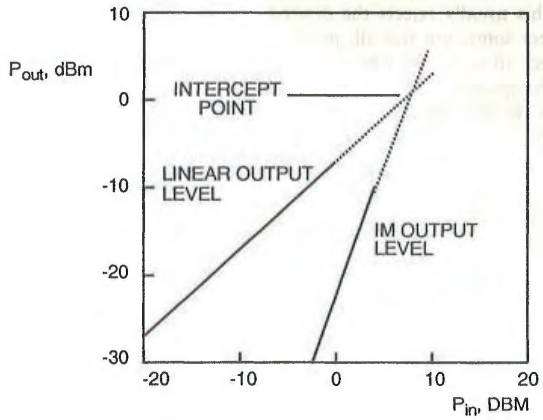


Figure 5.2 The output level of each nth-order IM product varies n dB for every dB change in input level. The intercept point is the extrapolated point at which the curves intersect; it can be specified in terms of either input or output power.

Spurious Responses

A mixer converts an RF signal to an IF signal. The most common transformation is

$$f_{IF} = f_{RF} - f_{LO} \tag{5.5}$$

although others are frequently used. Our discussion of frequency mixing indicated that harmonics of both the RF and LO could mix. The resulting set of frequencies is

$$f_{IF} = mf_{RF} + nf_{LO} \tag{5.6}$$

where m and n are integers. If an RF signal (which may or may not be in-band) creates an in-band IF response other than the desired one, it is called a *spurious response*. Usually the RF, IF, and LO frequency ranges are selected carefully to avoid spurious responses, and filters are used to reject out-of band RF signals that may cause in-band IF responses. IF filters are used to select only the desired response.

Many types of balanced mixers reject certain spurious responses where m or n is even. (It is possible to create a circuit that rejects responses where m or n are odd, but this usually rejects the desired response as well.) Most singly balanced mixers reject some, but not all, products where m or n is even; doubly balanced mixers reject all responses where m or n (or both) are even.

We designate a spurious response by its (m, n) indices, where m and n are by convention the RF and LO harmonics, respectively. If a mixer rejects a certain response having an even m or n, it will reject all other responses having the same m or n even. For example, if a mixer rejects the (2, 1) it will also theoretically reject the (4, 1), (4, 3), (6, 1), and so forth. For simplicity, in this chapter we will describe the (2, 1), (1, 2), or (2, 2) rejection of various types of mixers, and the reader should understand that all similar responses are theoretically rejected as well.

Other Rejection Phenomena

Balanced mixers reject AM noise from the LO, some even-order IM products, and provide port-to-port isolation. These phenomena depend on the cancellation of IF current or voltage components that are equal in magnitude and have a phase difference of 180 deg. Phase and magnitude errors are introduced by imbalance in the circuit, and these cause the cancellation to be imperfect.

In spurious response or IM cancellation, phase errors occurring at the RF or LO frequencies are multiplied by the harmonics involved in the spurious response. For example, if there is a 5-deg error in the phase balance of the RF hybrid, the

(6,1) spurious-response components will have a 30-deg error in the IF. For this reason, high-order spurious responses and IM products, although theoretically rejected, are not canceled by balanced mixers as effectively as low-order ones.

Spurious Signals

It is important to distinguish spurious *signals* from spurious *responses*. Spurious signals usually result from unwanted frequency components in the LO signal that are converted to the IF or simply leak through the mixer into the IF circuit. Mixers rarely create spurious signals; they often create spurious responses. Frequency synthesizers and phase-locked oscillators are notorious generators of spurious signals. Removing spurious signals is usually the job of the designer of the LO source.

dc Power

Mixers used in battery-powered applications must use minimal dc power. Of course, passive diode and FET mixers use no dc power, but their loss necessitates the use of more amplifier stages and balanced mixers require more LO power than single-device mixers; these indirectly affect dc power consumption. Some types of single-device mixers have performance characteristics that are similar to those of balanced mixers, and do not require high LO power.

Thermal and Mechanical Characteristics

Diode mixers and resistive FET mixers have very good thermal stability. The thermal stability of active FET mixers is similar to that of FET amplifiers. In many applications, mixers are subject to considerable mechanical and thermal stress as well as environmental dust and moisture. Modern low-cost fabrication techniques and device packages address these concerns adequately.

Cost

Most wireless applications require low-cost, large-quantity manufacturing. In these, design cost is usually a small portion of the cost of each component, and the cost of parts and assembly are dominant. These costs can be minimized by minimizing parts count and handling, and through automated assembly. Most types of mixers use few components and are amenable to automated assembly techniques.

5.3 SEMICONDUCTOR DEVICES FOR MIXERS

The primary devices used for mixers are Schottky-barrier diodes and FETs. Bipolar-junction transistors (BJTs) are also used occasionally, primarily in Gilbert-cell multi-

plier circuits (see Section 5.6.5), but because of their superior large-signal handling ability, higher frequency range, and lower noise, FET devices are usually preferred. Schottky-barrier diodes have the advantage of low cost and do not need dc bias. Unlike FETs and bipolar transistors, diodes are two-terminal devices and thus can be reversed; this allows them to be used in configurations that are impractical for three-terminal devices.

The Schottky-barrier diode is the dominant device used in mixers. Because Schottky diodes are inherently capable of fast switching and have very small reactive parasitics, they can be used in very broadband mixers. Schottky-diode mixers usually do not require matching circuits, so no tuning or adjustment are needed.

Although mixers using Schottky-barrier diodes always exhibit conversion loss, transistor mixers are capable of conversion gain. This helps simplify the architecture of a system, often allowing the use of fewer amplifier stages than necessary in diode-mixer receivers. The local oscillator (LO) and RF signals can be applied to separate gates of dual-gate MOSFETs, allowing good RF-to-LO isolation to be achieved in a single-device mixer. Dual-gate devices can be used to realize self-oscillating mixers, in which a single device provides both the LO and mixer functions.

5.3.1 Schottky-Barrier Diodes

A Schottky-barrier diode consists of a rectifying metal-to-semiconductor junction. The semiconductor consists of a thin epitaxial layer grown on a heavily doped substrate. The metal contact is the anode, and an ohmic cathode contact is made to the substrate either directly to the bottom of the chip or to the top side by etching away the epitaxial layer. Figure 5.3 shows the structure of a typical Schottky-barrier diode.

Equivalent Circuit

The equivalent circuit of the intrinsic diode is shown in Figure 5.4; real diodes may require additional elements to account for lead inductance, intermetallic capacitance, or package parasitics. The junction I/V characteristic, $I(V)$, is the dominant nonlinearity of the diode. The junction capacitance is more weakly nonlinear, and its nonlinearity rarely has a significant effect on the mixer's operation. The series resistance, R_s, is the resistance of the undepleted epitaxial layer under the junction. The resistance of the substrate and the ohmic contact may also be a minor contributor to R_s.

The I/V characteristic is given by the exponential expression

$$I(V) = I_0\left(\exp\left(\frac{qV}{\eta KT}\right) - 1\right) \tag{5.7}$$

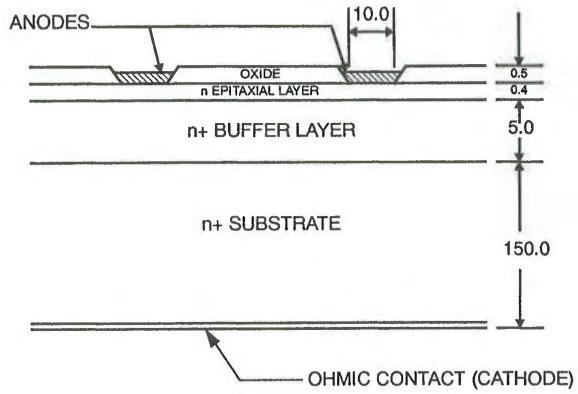


Figure 5.3 Cross section of a Schottky-barrier diode chip. Dimensions (in microns) and doping levels are typical for microwave diodes; low-frequency diodes may have larger junction areas and thicker epitaxial layers.

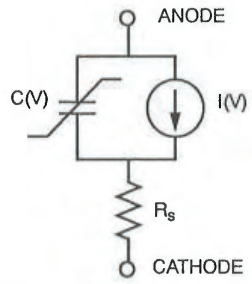


Figure 5.4 Equivalent circuit of a Schottky junction. The equivalent circuit of a practical diode often must include other parasitics, such as lead inductance and package capacitance.

where q is electron charge, K is Boltzmann's constant, and T is absolute temperature in kelvins. η is the ideality factor, usually in the range of 1.05 to 1.25, and accounts for the nonideality of the junction. I_0 is the current parameter, proportional to junction area. It is affected in a complex manner by temperature and the materials used in the junction.

The capacitance $C(V)$ is given by the expression

$$C(V) = \frac{C_{j0}}{\left(1 - \frac{V}{\phi}\right)^{\gamma}} \tag{5.8}$$

where C_{j0} is the zero-voltage junction capacitance and ϕ is the built-in potential of the junction. γ depends on the doping profile of the epitaxial layer; it is 0.5 if the epilayer is uniformly doped.

The series resistance R_s depends strongly on the geometry of the junction and the doping level of the epilayer: the longer or narrower the current path through the semiconductor, or the lower the doping level, the greater will be R_s. The value of R_s is lowest in diodes that use a cathode connection to the underside of the substrate; it is greatest in diodes that have the cathode connection on the top of the substrate. The value R_s is approximately inversely proportional to junction area (but C_{j0} is proportional to junction area). Because of the higher mobility in GaAs, GaAs diodes have lower R_s than silicon. R_s is usually on the order of a few ohms.

A figure of merit for a Schottky diode is its cutoff frequency, f_c:

$$f_c = \frac{1}{2\pi R_s C_{j0}} \tag{5.9}$$

where R_s is measured at dc and C_{j0} at any convenient low frequency. f_c is usually a few hundred to several thousand GHz, depending on the type of diode and semiconductor material. As a general rule, f_c should be at least five and preferably ten times the RF frequency of the mixer.

Diode Types

Schottky diodes are manufactured in many configurations. Chip diodes, similar to the structure in Figure 5.3, are used frequently. The diode chip is usually soldered or epoxy-bonded to a mounting surface, which serves as the cathode connection. The anode is connected to the circuit via a thermocompression-bonded gold wire or ribbon. Chip diodes can be mounted in a wide variety of packages, from miniature "pill" packages, consisting of a ceramic cylinder with metal end caps, to inexpensive epoxy-filled packages with ribbon leads.

One of the most useful structures is a beam-lead diode, in which the diode chip and its gold ribbon leads are an integral part of the structure. Figure 5.5 shows a beam-lead diode. The cathode lead is connected to the substrate through the top of the chip via an etched hole in the epilayer. The epilayer is removed entirely in the vicinity of the cathode connection, allowing the cathode to contact the substrate directly. The anode is insulated from the substrate by an oxide layer.

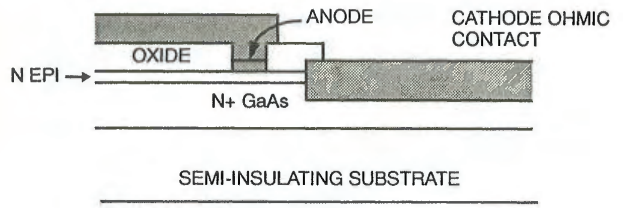


Figure 5.5 Cross section of a conventional beam-lead diode. The "overlay" capacitance between the anode lead and the epilayer is a significant parasitic; the cathode ohmic contact surrounds the anode to minimize series resistance.

The overlay capacitance between the anode lead and the substrate is a significant parasitic reactance. To minimize it, the anode lead is often located near the edge of the chip, resulting in a fragile structure. Other methods can be used to minimize the overlay capacitance while allowing a more rugged anode connection. These include air gaps, boron implants (which reduce the conductivity of the substrate), and glass-filled trenches under the anode lead.

A structure often used in monolithic circuits is an air-bridge mesa diode (Figure 5.6). The use of an air bridge to the anode minimizes overlay capacitance. The cathode is connected to the top of the structure; series resistance is minimized by wrapping the cathode ohmic contact around the anode.

5.3.2 Field-Effect Transistors

A variety of types of field-effect transistors (FETs) are used in mixers. Since the 1960s, silicon MOSFETs (often dual-gate devices) have dominated mixer applications in communication receivers up to approximately 1 GHz. At higher frequencies, GaAs MESFETs are often used.

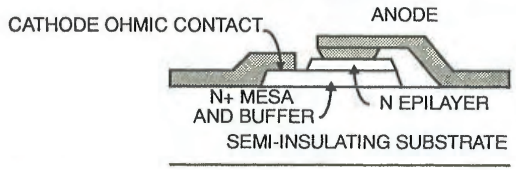


Figure 5.6 Cross section of a mesa diode used for monolithic circuits.

Although FETs can perform frequency mixing in several ways, the most common is to use the time-varying transconductance that results from applying the LO to the gate. Thus, in a FET used as a mixer, the transconductance is of primary importance. Most resistive and reactive parasitics are of secondary interest; the exception is the gate-to-channel capacitance, which limits the frequency range over which the device can be used.

Figure 5.7 shows the equivalent circuit of a FET. The topology is the same for all types of FETs; the differences are in the values of the elements, and especially in the *I/V* and *C/V* characteristics of the nonlinear elements. (The functional forms of these expressions will be presented in the sections describing the individual devices.) All FETs have resistances associated with the ohmic contacts to their drains and sources, R_d and R_s, respectively, and the resistance of the gate, R_g. The value R_{ds} is the drain-to-source resistance and R_i is the resistance of the undepleted epitaxial layer at the source end of the channel (R_i is usually negligible in MOSFETs and in all FETs used as resistive mixers; see Section 5.7.4). The value C_{ds} is the drain-to-source capacitance and C_{gd} is the drain-to-gate capacitance, caused primarily by the proximity of the gate and drain metalizations.

Silicon MOSFETs

Although silicon devices have distinctly lower transconductance than GaAs, they are useful up to at least the lower microwave frequencies. In spite of the inherent inferiority of silicon in comparison to GaAs, silicon MOSFETs do have some advan-

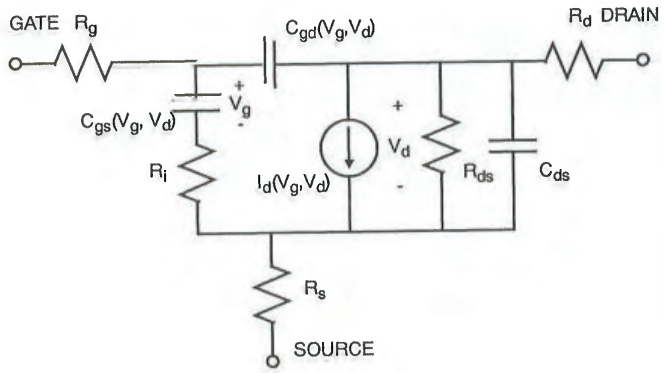


Figure 5.7 Equivalent circuit of a FET. Most types of FETs are well characterized by this equivalent circuit; only the values of the elements and their *I/V* or *C/V* characteristics are different.

tages. The primary advantage of silicon over GaAs is low cost, and the performance of silicon MOSFET mixers is not significantly worse than GaAs in the VHF and UHF ranges. The high drain-to-source resistances (R_{ds}) of silicon MOSFETs give them higher voltage gain than GaAs devices; in many applications this is a distinct advantage. Additionally, the positive threshold voltage (in an *N*-channel enhancement MOSFET), in comparison to the negative threshold voltage of a GaAs FET, is very helpful in realizing low-voltage circuits and circuits requiring only a single dc supply. Mixers using enhancement-mode silicon MOSFETs often do not require gate bias, and dual-gate MOSFETs offer convenient LO-to-RF isolation (when the LO and RF are applied to different gates).

A MOSFET is sometimes called an *insulated-gate FET*, because the gate is insulated from the channel by a very thin oxide layer. In an *enhancement-mode* device, a voltage applied to the gate generates carriers at the oxide-semiconductor interface, and the resulting channel becomes conductive. Increasing the gate voltage increases the surface charge and the channel current. A *depletion-mode* device has a lightly doped layer under the gate, so channel current is possible at zero gate voltage, and a reverse voltage is needed to deplete the layer and to turn off the channel. MOSFETs can be produced in either *N*- or *P*-channel form.

Figure 5.8 shows a cross section of a modern single-gate MOSFET [1]. This device is suitable for use up to at least 6 GHz but is intended primarily for 900-MHz applications. The device is built on a *P*+ substrate having a *P*– epilayer. A *P*– base is provided to adjust the threshold voltage, and *P*+ "sinkers" connect the source regions through the *P*+ substrate to the underside of the substrate for a ground

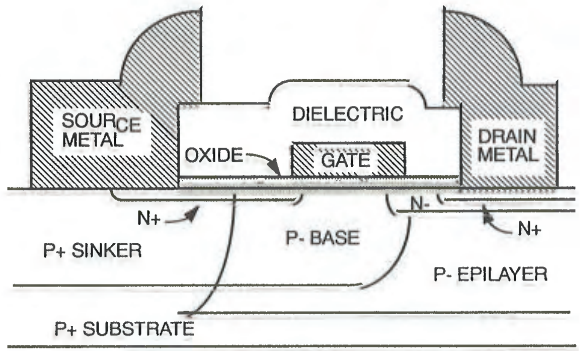


Figure 5.8 Cross section of a silicon MOSFET adequate for use in microwave mixers. (*After:* [1], used with permission).

connection. The gate length is 1.5 μm; gates of 0.6-μm length have been fabricated. A dual-gate version of this device exhibits 16-dB conversion gain and 1.5-dB noise figure at 900 MHz.

A simple and reasonably accurate expression for the channel current in a MOSFET is

$$I_d(V_g, V_d) = \beta\left((V_g - V_t)V_d - \frac{V_d^2}{2}\right) \tag{5.10}$$

where V_t is the threshold voltage, β is a constant having units of V^{-2}, and other voltages are as shown in Figure 5.7. Equation (5.10) is valid in the so-called *linear region* of the device, when $V_d < V_g - V_t$ and $V_g > V_t$. At higher drain voltages, the current is largely independent of V_d, and is given by (5.10) with $V_d = V_g - V_t$. This is called the *saturation region* of the device, because the current is "saturated" in the sense that it does not increase with gate voltage.[3] The current is therefore

$$I_d(V_g, V_d) = \frac{\beta}{2}(V_g - V_t)^2 \tag{5.11}$$

Equations (5.10) and (5.11) result from a relatively simple approximate analysis of the device. More accurate (and more complex) models have been developed, and are a subject of ongoing research. Equations (5.10) and (5.11) nevertheless describe one of the most important characteristics of a silicon FET: the square-law dependence of its drain current on gate voltage. This is responsible for the device's low third-order distortion.

Differentiating (5.11) gives the transconductance when $V_d > V_g - V_t$:

$$G_m(V_g) = \beta(V_g - V_t) \tag{5.12}$$

The transconductance is proportional to V_g and can be made to vary with time by the application of an LO signal to the gate. The magnitude of the transconductance variation is proportional to β; a simple expression for β is

$$\beta = \mu C_{ox}\frac{W}{L} \tag{5.13}$$

where μ is the mobility of holes or electrons (in $cm^2/V\,s$), C_{ox} is the oxide capacitance (the gate-to-channel capacitance per unit area, in F/cm^2), W is the gate width, and

[3]In GaAs devices, current saturation occurs primarily because electrons in the channel reach saturated drift velocity. Although velocity saturation also occurs in silicon devices, the effect on the *I/V* characteristic is much less, and current saturation is caused primarily by the pinch-off of the drain end of the channel.

L is the gate length, both in centimeters. Equation (5.13) shows that the peak transconductance is proportional to gate width and mobility and is inversely proportional to gate length. Because the peak transconductance is also proportional to C_{ox}, a thin oxide layer results in high transconductance.

GaAs MESFETs

A MESFET is a junction FET having a Schottky-barrier gate. Although silicon MESFETs have been made, they are now obsolete and all modern MESFETs are fabricated on GaAs. The gate's length is usually less than 0.5 μm, and may be as short as 0.1 μm; this short gate length, in conjunction with the high electron mobility and saturation velocity of GaAs, results in a high-frequency, low-noise device.

Figure 5.9 shows a cross section of a GaAs MESFET. The channel is a moderately doped epitaxial layer 1,000Å to 2,500Å thick, grown on an undoped substrate. Because N GaAs has higher mobility than P GaAs, all conventional GaAs MESFETs use N material for the channel. A thick, undoped buffer layer forms the lower boundary of the channel. This layer prevents substrate impurities from diffusing into the epitaxial layer during processing. Like the MOSFET, the drain and source metalizations form ohmic contacts at each end of the channel. The gate is often T-shaped in cross section to minimize its resistance while minimizing its length.

A dc voltage applied to the channel creates a longitudinal electric field. In normal operation, the field is strong enough to accelerate the electrons to their saturated drift velocity, creating an electron current from the source to drain. The Schottky barrier formed by the gate creates a depletion region that extends part way into the channel at zero gate bias. Varying the gate-to-source voltage modulates the depletion depth and hence the thickness of the conductive channel; the channel

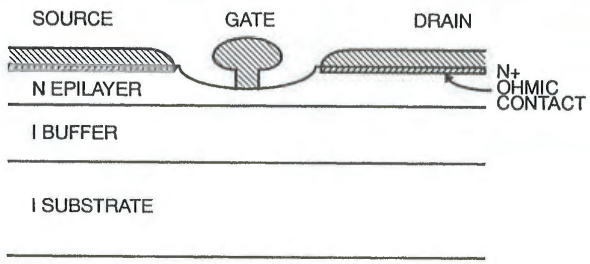


Figure 5.9 Cross section of a modern, recessed-gate MESFET. The recessed gate reduces the parasitic source resistance and minimizes the effective gate length. The mushroom-shaped gate cross section minimizes the resistance of the very short (< 0.5 μm) gate.

current varies accordingly. As in the Schottky diode, the gate voltage also varies the depletion capacitance. A negative gate-to-source voltage is necessary to fully deplete the channel, causing it to "pinch off," preventing channel current.

Because of the complex velocity-field characteristic of the GaAs FET and the short channels of modern MESFETs, it is not possible to develop a simple and accurate expression for the GaAs FET's *I/V* characteristic from the physics of the device. Instead, empirical formulae are usually employed. A popular expression is that of Curtice [2]:

$$I_d(V_g, V_d) = (A_0 + A_1V_1 + A_2V_1^2 + A_3V_1^3)\tanh(\alpha V_d) \tag{5.14}$$

where A_n and α are constant coefficients and V_1 is an intermediate variable,

$$V_1 = V_g(t - \tau)(1 + \beta(V_{d0} - V_d)) \tag{5.15}$$

β is an empirical constant, and τ is the time delay. V_{d0} is the value of V_d at which the A_n coefficients were determined.

Because GaAs MESFETs are considerably more expensive than silicon devices, they are usually used only at frequencies that are too high for silicon devices. GaAs devices can be used at lower frequencies if necessary; for example, as IF amplifier stages in monolithic microwave mixers.

High Electron-Mobility Transistors

A high electron-mobility transistor (HEMT) is a junction FET that uses a heterojunction (a junction between two dissimilar semiconductors), instead of a simple epitaxial layer, for the channel. The discontinuity of the band gaps of the materials used for the heterojunction creates a layer of charge at the surface of the junction; the charge density can be controlled by the gate voltage. Because the charge in this layer has a very high mobility, high-frequency operation and very low noise are possible. It is not unusual for HEMTS to operate successfully as low-noise amplifiers above 100 GHz. HEMTs require specialized fabrication techniques, such as molecular-beam epitaxy, and thus are very expensive to manufacture. HEMT heterojunctions are invariably realized with III-V semiconductors; AlGaAs and InGaAs are common.

HEMTs are used for mixers in the same way as conventional GaAs FETs. Because the gate *I/V* characteristic of a HEMT is generally more strongly nonlinear than that of a MESFET, HEMT mixers usually have greater intermodulation distortion (IM) than FETs. The noise figure (NF) of a HEMT mixer usually is not significantly lower than that of a GaAs FET, however.

Figure 5.10 shows the cross section of one type of HEMT. There are many degrees of freedom in the design of such devices; for example, the number of hetero-

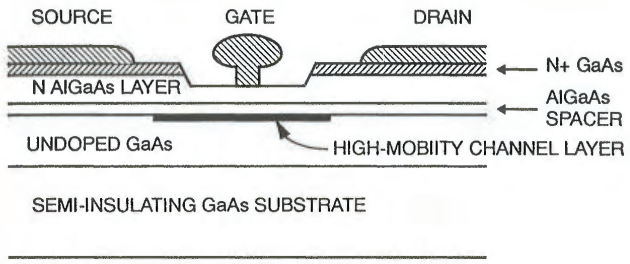


Figure 5.10 Cross section of a simple HEMT. The two-dimensional electron gas resides in the GaAs layer, adjacent to the AlGaAs spacer.

junctions, the thicknesses of layers, and the fraction of Al or In in AlGaAs or InGaAs devices are all variable, and can be used to optimize the device.

Like that of the GaAs FET, the *I/V* characteristic of a HEMT usually is described empirically. Because the *I/V* characteristics for HEMTs are very different from those of GaAs devices, *I/V* models appropriate for GaAs devices are usually inadequate for HEMTs. An expression that works well for many types of small-signal HEMTs is [3]

$$I_d(V_g, V_d) = I_p(1 + \tanh(\Psi))(1 + \lambda V_d)\tanh(\alpha V_d) \tag{5.16}$$

where

$$\Psi = P_1(V_g - V_p) + P_2(V_g - V_p)^2 + \ldots \tag{5.17}$$

The V_p is the gate voltage at peak transconductance, I_p is the current at the same point, and λ and α are constants.

Bipolar Transistors

Silicon bipolar-junction transistors (BJTs) find limited use in mixers. Because FETs generally provide better overall performance, silicon BJTs are rarely used in conventional single-device or balanced mixers. Bipolar devices are used as mixers when necessary for process compatibility, not for any inherent technical advantages. For this reason, we will be concerned with them only minimally.

Heterojunction bipolar transistors (HBTs), BJTs that use a heterojunction for the emitter-to-base junction, are also useful in wireless circuits. They are most useful as high-efficiency power amplifiers and low-distortion small-signal amplifiers.

Because they require III-V materials and special fabrication techniques, they are relatively expensive. Silicon-Germanium (Si-Ge) HBTs are a new technology that offers high performance at costs close to those of silicon BJTs.

5.4 FREQUENCY MIXING

The operation of all practical mixers is based on frequency conversion in a time-varying device parameter, usually conductance or transconductance. This time-varying conductance is created by applying an LO waveform to a nonlinear device. In a diode mixer, the junction conductance can be found from the derivative of (5.7)

$$g_j(t) = \frac{\partial}{\partial V} I(V)\bigg|_{V=V_{LO}(t)} = \frac{qI_0}{\eta KT}\exp\left(\frac{qV_{LO}(t)}{\eta KT}\right) \approx \frac{qI(V_{LO}(t))}{\eta KT} \tag{5.18}$$

where $V_{LO}(t)$ is the junction voltage of the LO signal. The conductance waveform $g_j(t)$ can be expressed as a Fourier series

$$g_j(t) = \sum_{k=-\infty}^{\infty} G_k \exp(jk\omega_p t) \tag{5.19}$$

where ω_p is the LO radian frequency. If a signal voltage $V_s\exp(j\omega_s t)$ is applied to this conductance, the resulting current is

$$i(t) = \sum_{k=-\infty}^{\infty} I_k \exp(j(k\omega_p + \omega_s)t) \tag{5.20}$$

where the resulting frequencies $k\omega_p + \omega_s$ are the mixing products. In general, however, the voltage will not have only a single frequency, ω_s; the source impedance will generate junction-voltage components at the same mixing frequencies as the current components. Also, we customarily express the mixing products in (5.20) as $k\omega_p + \omega_0$, where ω_0 is the lowest mixing frequency; usually $\omega_0 = |\omega_s - \omega_p|$. With a change of the index k, this results in the same set of frequencies as in (5.20). Thus, we have

$$\begin{aligned} i(t) &= \sum_{k=-\infty}^{\infty} I_k \exp(j(k\omega_p + \omega_0)t) \\ v(t) &= \sum_{k=-\infty}^{\infty} V_k \exp(j(k\omega_p + \omega_0)t) \\ g_j(t) &= \sum_{k=-\infty}^{\infty} G_k \exp(jk\omega_p t) \end{aligned} \tag{5.21}$$

where $v(t)$ is the small-signal junction voltage and $i(t)$ is the small-signal junction current. Substituting these into the relation

$$i(t) = g_j(t)\, v(t) \tag{5.22}$$

gives the result

$$\sum_{l=-\infty}^{\infty} I_l \exp(j(\omega_0 + l\omega_p)t) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} G_m V_n \exp(j(\omega_0 + (n+m)\omega_p)t) \tag{5.23}$$

which can be expressed by the matrix equation

$$\begin{bmatrix} I_N^* \\ I_{N+1}^* \\ \cdots \\ I_1^* \\ I_0 \\ I_1 \\ \cdots \\ I_{N-1} \\ I_N \end{bmatrix} = \begin{bmatrix} G_0 & G_{-1} & G_{-2} & \cdots & \cdots & G_{-N} & \cdots & G_{-2N} \\ G_1 & G_0 & G_{-1} & \cdots & \cdots & G_{-N+1} & \cdots & G_{-2N+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_{N-1} & G_{N-2} & G_{N-3} & \cdots & \cdots & G_{-1} & \cdots & G_{-N-1} \\ G_N & G_{N-1} & G_{N-2} & \cdots & \cdots & G_0 & \cdots & G_{-N} \\ G_{N+1} & G_N & G_{N-1} & \cdots & \cdots & G_1 & \cdots & G_{-N+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_{2N-1} & G_{2N-2} & G_{2N-3} & \cdots & \cdots & G_{2N-1} & \cdots & G_{-1} \\ G_{2N} & G_{2N-1} & G_{2N-2} & \cdots & \cdots & G_N & \cdots & G_0 \end{bmatrix} \begin{bmatrix} V_N^* \\ V_{N+1}^* \\ \cdots \\ V_1^* \\ V_0 \\ V_1 \\ \cdots \\ V_{N-1} \\ V_N \end{bmatrix} \tag{5.24}$$

where the infinite series in (5.24) is truncated, for obvious practical reasons, to $2N + 1$ terms. The subscript n refers to the frequency component

$$\omega_n = n\omega_p + \omega_0;\ \omega_{-n} = -\omega_n \tag{5.25}$$

In (5.24), negative-frequency components have been represented as the conjugate of conventional positive-frequency phasors. Figure 5.11 shows the frequencies in (5.23) and (5.24).

The matrix in (5.24) relates the voltages to the currents at each mixing frequency. It can be treated in all ways like a multiport admittance matrix. Indeed, one can think of the junction of the mixing diode as a type multiport network in which the "ports" are voltages and currents at the various mixing frequencies instead of physically separate ports. The external circuit can then be treated as a set of circuits, one at each mixing frequency, connected to the diode. These circuits at the various mixing frequencies are interconnected via the diode.

A consequence of this representation is that the characteristics of the external circuit (often called the *embedding network*) at all mixing frequencies affect the

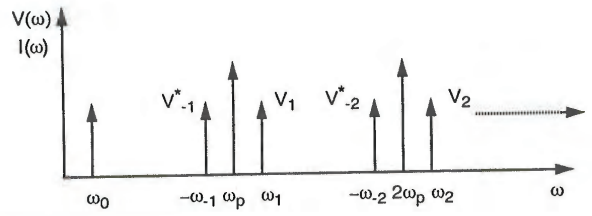


Figure 5.11 Mixing frequencies in a pumped nonlinear element.

properties of the mixer at the RF and IF frequencies. For this reason, mixer analysis must take into account the characteristics of the embedding network at all mixing frequencies. In theory, this is an infinite set of frequencies, so practical analysis requires truncating the set to some reasonable size. In most diode-mixer analyses, for example, a value of $N = 4$ (which requires 8 LO harmonics plus dc) is usually adequate.

Although we used diodes as an example, the same approach to analysis is applicable to active mixers. However, in active mixers the RF excitation voltage and IF output current are at separate ports that are only weakly coupled (e.g., the gate and drain of a FET), so it is often possible to simplify the analysis.

5.5 DIODE MIXERS

Semiconductor diode mixers are the most common of all mixer types. This section will discuss their design, implementation, and performance.

5.5.1 Diode Mixer Characteristics

Diode mixers are an example of a universal component in RF and microwave engineering: one that does nothing especially well, but does everything at least adequately. For this reason, diodes have been used as mixers from the earliest days of radio up to the present. It is safe to predict that they will be in common use far into the future.

In contrast to transistor mixers, which can provide conversion gain, diode mixers always exhibit conversion loss, usually 6 to 10 dB, depending on the frequency and type of design. Conversion loss and input/output impedance depend on LO level: conversion loss usually decreases monotonically with LO level and port impedance decreases, until the diode is very strongly pumped. At this point, further increases in LO level have progressively less effect on performance, until avalanche breakdown occurs.

Diode mixers can be very broadband. Because diodes can have very small junction capacitances (as little as 0.005 pF, if necessary) the junction capacitance rarely limits the bandwidth of a diode mixer; instead, bandwidth is usually established by the embedding circuit. The general approach to the design of a diode mixer is to avoid the use of a reactive matching circuit. Usually the diode is selected so that its junction capacitance is negligible, and the embedding circuit is designed to match the resistive diode.

Most practical diode mixers are balanced. Single-diode mixers are occasionally used for simple, low-cost applications where performance need not be high. Balanced mixers offer a number of advantages over single-diode mixers:

1. Balanced mixers usually have some degree of inherent port-to-port isolation.
2. AM noise on the LO signal is rejected by a balanced mixer.
3. Certain even-order spurious responses and intermodulation products are rejected by many types of balanced mixers.

Unfortunately, balanced mixers do not offer improved conversion loss or noise figure compared to single-device structures. These characteristics depend strongly on the balance of the hybrids or baluns used to realize the mixer; achieving good balance in a broadband hybrid is difficult. Finally, certain types of single-device mixers (for example, the series-switch FET mixer examined in Section 5.7.1) can offer performance that is comparable to that of balanced diode mixers.

5.5.2 Single-Diode Mixers

Although single-diode mixers have serious limitations, there are two good reasons for examining them. First, in spite of their limitations, single-diode mixers are occasionally used, and in some technologies (e.g., submillimeter applications) they are used exclusively. Second, all balanced diode mixers can be reduced to a single-diode equivalent circuit; thus, the single-diode mixer serves as a prototype for more complex mixers.

Figure 5.12 shows a simple circuit of a single-diode mixer. It illustrates a number of requirements for the design of practical mixers. First, the input and output circuits must be *noninteracting*; that is, they should present either a short or open circuit, depending on the configuration, to the other ports. In Figure 5.12, the parallel L-C resonators are ideal; they are open circuits at their resonant frequencies and short circuits at all other frequencies. Because of this property, at each frequency (RF, IF, and LO) the diode is connected to the respective port, and is isolated from all the others. Thus, at the RF frequency the LO and IF ports are disconnected; similarly, at the IF and LO frequencies the RF/LO and RF/IF ports, respectively, are disconnected.

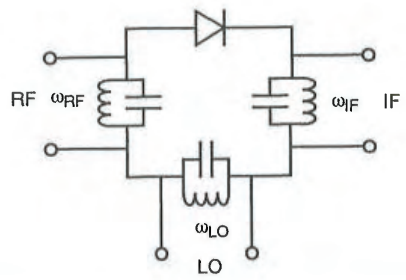


Figure 5.12 An ideal single-diode mixer. The parallel L-C resonators are ideal; each is a short circuit at all frequencies except the resonant frequency.

A second requirement is that the diode be matched at all important frequencies. As long as the terminations at all mixing frequencies are not reactive, the diode's impedance will be resistive. In this manner, the port's impedance can be selected to match the diode (transformers or other such structures can be used to obtain a standard port impedance (for example, 50Ω), if necessary). The diode's junction capacitance, if not negligible, can be absorbed into the capacitances of the resonators.

Third, the diode must be appropriately terminated at all unwanted mixing frequencies. In this case, the diode is short-circuited at all frequencies except the RF, IF, and LO. In general, the best overall performance is obtained when the diode is short-circuited; other terminations (for example, an open circuit) are theoretically possible, but are often difficult to implement or result in poor performance in some respects. In particular, an open-circuit at unwanted mixing frequencies results in greater intermodulation distortion than a short-circuit.

When the diode is short-circuited at all mixing frequencies except the RF and LO, (5.24) can be reduced to the form

$$\begin{bmatrix} I_{IF} \\ I_{RF} \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_{IF} \\ V_{RF} \end{bmatrix} \tag{5.26}$$

This admittance matrix can be used in the same way as any two-port admittance matrix to find conversion loss and input or output impedance.

Fourth, conversion efficiency must be reasonably high. In most systems, the mixer's conversion efficiency need not be as high as possible; indeed, high conversion gain in an active mixer is often deleterious. Conversion loss of 6 to 8 dB is usually adequate for passive mixers, and approximately unity or a few decibels is usually optimum for active mixers. Only rarely should conversion efficiency be optimized at the expense of other performance characteristics.

The single-diode mixer in Figure 5.12 meets these requirements through the properties of the L-C resonators. This is usually the case; the performance of a single-device mixer depends strongly upon the characteristics of filters. Because practical filters have limited Qs and are often uncomfortably large, single-diode mixers often exhibit only modest performance. Indeed, even with perfect filters, single-diode mixers cannot be realized if the RF, IF, or LO bands overlap. For these reasons, balanced mixers, which employ the properties of hybrids or baluns to circumvent these difficulties, are more frequently used.

Figure 5.13 shows a practical microstrip implementation of a single-diode mixer. This mixer uses a ring resonator for LO injection. Such narrowband resonators are usually suitable only if the LO is fixed-frequency. The ring is $\lambda/2$ in circumference at the resonant frequency and is loosely coupled to the RF and LO microstrips. This coupling can be varied to adjust the Q of the resonator; for minimum LO loss, the bandwidth should be as wide as possible consistent with obtaining adequate RF-to-LO isolation. The ring generates a wave on the RF microstrip propagating toward the diode; if the diode is not well matched, LO power is reflected toward the RF port, and the LO-to-RF isolation is equal to the diode's LO reflection coefficient. The termination on the LO microstrip is necessary to ensure a low, broadband VSWR at the LO input; this is necessary for stable operation of the LO source. The termination is decoupled from the LO at the ring's resonant frequency, and therefore absorbs no power.

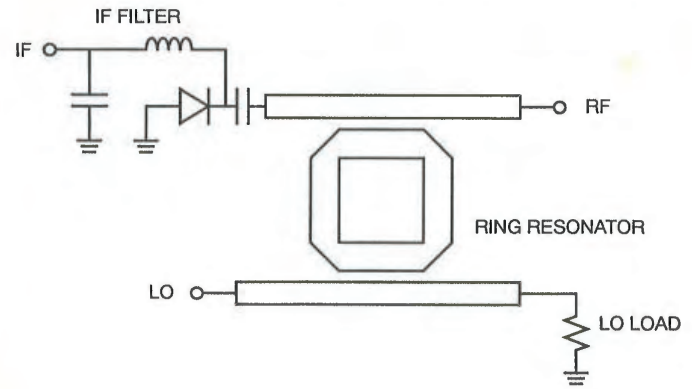


Figure 5.13 A practical microstrip realization of a single-diode mixer. A ring resonator is used to inject the LO; because this resonator is very narrow, only a single-frequency LO is practical. The LO load provides a good out-of-band match to the LO; it is decoupled at the ring's resonant frequency.

5.5.3 Baluns and Hybrids for Wireless Mixers

Before we discuss balanced mixers, we must examine one of the most important parts of the mixer: the balun or hybrid used in its realization.

Baluns and Hybrids

Although both baluns and hybrids are often used interchangeably in mixers, it is important to distinguish between the two. A balun is a structure that interconnects balanced and unbalanced transmission lines (*balun* is a contraction of the words *bal*anced and *un*balanced). One terminal of a balun's input port is grounded; both output terminals are ungrounded. Ideally, the outputs are "floating": the voltage between either and ground is undefined (just as the voltage between ground and the wires of a balanced, ungrounded transmission line are undefined). If an even-mode excitation is applied to the output, no power is transmitted to the input, and the even-mode output-port impedance is either an open or short circuit (most practical baluns exhibit an open circuit to even-mode excitation).

A hybrid is a special type of four-port junction. If a signal is applied to any port, it emerges from two of the other ports at half power; the fourth port is isolated. All four ports are matched. Only two fundamental types of hybrids are possible: the 180-deg and 90-deg forms. The ideal hybrids have the scattering matrices

$$S_{90} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & -j & 1 \\ 0 & 0 & 1 & -j \\ -j & 1 & 0 & 0 \\ 1 & -j & 0 & 0 \end{bmatrix} \tag{5.27}$$

for the 90-deg hybrid and

$$S_{180} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix} \tag{5.28}$$

for the 180-deg hybrid.

If the input and output ports of a 180-deg hybrid are chosen so that the outputs are in phase, the input port is called the *sigma* or *sum* port, designated by the Greek letter Σ. The remaining isolated port is the *difference* or *delta* port, designated by Δ, because a signal applied to it will generate out-of-phase outputs. Because a 180-deg phase difference can be obtained at the output ports, a 180-deg hybrid often can be used in the same manner as a balun.

Indeed, certain components can be used as either baluns or hybrids. Figure 5.14 shows a transformer used as both. In Figure 5.14(a) the transformer is connected as a hybrid; in Figure 5.14(b), as a balun. This type of transformer is usually realized as a trifilar winding on a magnetic core; such a transformer can have a frequency range from a few megahertz to several hundred megahertz. The upper frequency response is limited by magnetic leakage flux and capacitance between the windings. Transmission-line transformers, which operate as transformers at low frequencies but as transmission lines at high frequencies, avoid this problem.

RF Transformers

RF transformers are used frequently in wireless components. Properly designed and configured, these can be used as either baluns or hybrids, and often have bandwidths of several decades. We use the term transformer somewhat loosely, as such compo-

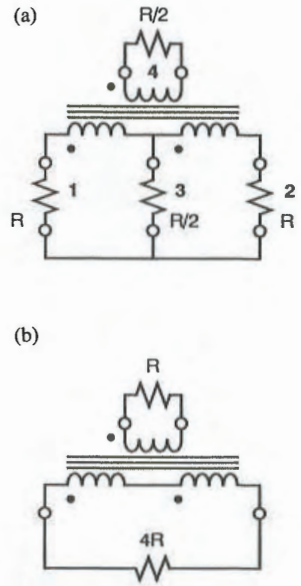


Figure 5.14 A transformer connected as (a) a hybrid and (b) a balun. The port numbers correspond to (5.28).

nents often operate as magnetic transformers at low frequencies, but as transmission-line structures at high frequencies.

A magnetic transformer consists of two inductors coupled by their magnetic fields. The inductors have self inductances L_1, and L_2, the inductance of one inductor when the other inductor's current is zero, and mutual inductance, M, loosely defined as the inductance common to both inductors. When the entire magnetic field generated by one inductor is contained by the other, the coupling is perfect and

$$M = \sqrt{L_1 L_2} \tag{5.29}$$

Figure 5.15 shows a transformer connected to a load, treated as a two-port. When (5.29) is satisfied, the ratio of the port voltages is

$$\frac{V_2}{V_1} = \sqrt{\frac{L_2}{L_1}} \tag{5.30}$$

which, for solenoidal inductors, is equal to the ratio of the turns of wire on the two coils. The current ratio is

$$\frac{I_1}{I_2} = -\frac{R + j\omega L_2}{j\omega M} \tag{5.31}$$

which is the negative inverse of (5.30) when $R \ll \omega L_2$.

In mixers, transformers are often used for impedance transformation. The input impedance at L_1 is

$$Z_1 = R\frac{L_1}{L_2} \tag{5.32}$$

when $R \ll \omega L_2$.

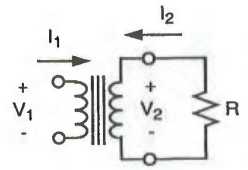


Figure 5.15 An RF transformer, configured as a two-port, with a terminating resistance R.

Equation (5.29) represents an idealization. In practice, there is always some leakage inductance, so the inductors are not perfectly coupled. In this case, the coupling coefficient k is defined as

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{5.33}$$

Clearly, when the inductors are perfectly coupled, $k = 1$; $k = 0$ implies $M = 0$, and there is no coupling. When $k < 1$ and $R \ll \omega L_2$, the input impedance at port 1 becomes

$$Z_1 = R\frac{L_1}{L_2} + j\omega L_1(1 - k^2) \tag{5.34}$$

The imperfect coupling causes the input impedance to be reactive. This reactance limits the bandwidth of the transformer. The bandwidth limitation of a magnetic transformer becomes apparent: on one hand, to satisfy the requirement that $\omega L_2 \gg R$, the inductance of the coils must be large; however, imperfect coupling causes part of that large inductance to appear in series with the transformer.

Another problem that limits the bandwidth of a transformer is the stray capacitance between the inductors. This capacitance degrades the performance of the transformer in a number of ways, affecting the phase characteristics and introducing spurious resonances. Because of these effects, simple magnetic transformers are limited in bandwidth to a few hundred megahertz.

Even so, simple magnetic transformers are adequate for many applications. Customarily, they consist of two or three bifilar or trifilar windings on a toroidal magnetic core. Leakage inductance can be minimized by the use of cores that completely enclose the windings.

Bandwidth can be increased through the use of a structure that is equivalent to a transformer at low frequencies, but is equivalent to a transmission-line transformer at high frequencies. In this way, at high frequencies the stray capacitance between the windings becomes part of the transmission-line capacitance and, because the transformer is not operating as a magnetic transformer, magnetic leakage is inconsequential. Such transformers may have bandwidths above 2 GHz, and are limited primarily by the practical difficulties in making low-parasitic connections at microwave frequencies. Many of the transmission-line transformers in the following section consist of twisted-pair transmission lines wound on a magnetic core to realize such transformers.

Transmission-Line Baluns and Hybrids

A wide variety of baluns and hybrids can be realized with transmission lines. Most of these require transmission-line sections one-quarter or one-half wavelength long.

This makes them uncomfortably large for use in wireless applications at frequencies below 2,400 MHz. Often, however, the structures can be modified to make them practical at low frequencies by either meandering the strips or loading them with dielectric material.

Figure 5.16 shows a simple transmission-line balun. It consists of a $\lambda/4$ section of transmission line whose characteristic impedance Z_0 is chosen to act as a transformer between the source and load impedances:

$$Z_0 = \sqrt{Z_s Z_L} \tag{5.35}$$

This balun is very sensitive to stray capacitance between the transmission line and ground. To account for this capacitance, we treat the conductors as coupled transmission-line sections, and their even-mode impedance accounts for the stray capacitance. Ideally, the capacitance is zero, causing the even-mode impedance to be infinite. The odd-mode impedance is half the transmission-line impedance; that is,

$$Z_{0o} = 0.5Z_0 = 0.5\sqrt{Z_s Z_L} \tag{5.36}$$

Analysis shows that the even-mode impedance should be at least ten times the odd-mode. This is often accomplished in microwave mixers by the use of a suspended substrate for the balun. At low frequencies, the even-mode impedance can be increased by using a twisted pair of conductors for the transmission line and inserting the conductors into a magnetic core to increase the even-mode inductance. This has two effects: first, it reduces the size of the balun; second, the balun operates as a 1:1 transformer at low frequencies, improving the low-frequency response. (Of course, to have a good VSWR in this low-frequency range, the source and load impedances must be equal.)

The transmission-line balun can be shortened by loading the input and output ends with capacitors. The optimum capacitance depends on the mode impedances and is best determined by numerical optimization. As an initial value, the line length should be $\lambda/8$ and the capacitors' reactances should be $0.707Z_0$.

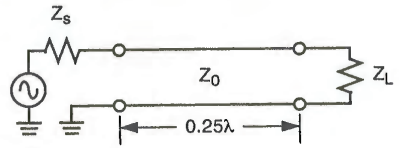


Figure 5.16 A simple transmission-line balun.

Other transmission-line transformer combinations can be used as baluns. Figure 5.17 shows a balun having a 1:4 impedance ratio with $Z_0 = Z_s$. This balun is not fundamentally different from that of Figure 5.16; it consists of two such baluns connected in parallel at the input and in series at the output. By selecting $Z_0 = 0.5Z_s$, one could obtain a 1:1 transformation at the high-frequency end of its bandwidth. If the balun were wrapped on a magnetic core to improve its low-frequency response, the low-frequency impedance transformation would still be 1:4.

Figure 5.18 shows another 1:4 balun. Both outputs of this balun are at ground potential for dc, and provide a convenient dc return for the mixer diodes.

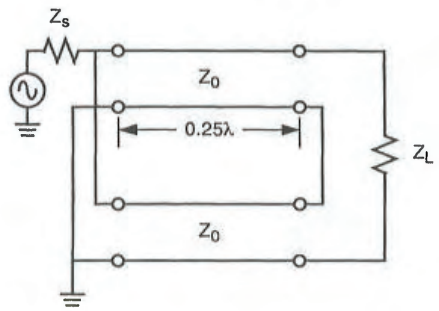


Figure 5.17 A 1:4 impedance-transforming transmission-line balun; $Z_0 = Z_s$, $Z_L = 4\ Z_s$.

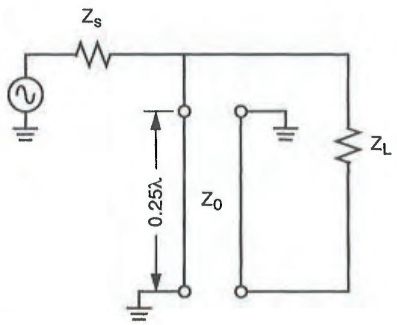


Figure 5.18 A 1:4 impedance-transforming balun; $Z_0 = Z_s$.

Lumped-Element Hybrids

It is possible to create a hybrid from lumped elements. Figure 5.19 shows such a 180-deg hybrid. The relation for the element values is very simple:

$$\omega L = \frac{1}{\omega_C} = 1.414Z \tag{5.37}$$

where Z is the port impedance. This hybrid has a bandwidth of approximately 20%.

Ninety-degree lumped hybrids are usually very narrowband. Figure 5.20 shows two circuits having bandwidths of approximately 1 to 2%, depending on the balance required. Figure 5.20(a) consists entirely of L-C elements and can be realized in an integrated circuit. Its design equations are as follows:

$$C_1 = \frac{1}{\omega Z}$$

$$L = \frac{R}{\omega\sqrt{1 + \omega C_1 Z}}$$

$$C_0 = \frac{1}{\omega^2 L} - C_1 \tag{5.38}$$

where Z again is the port impedance. Figure 5.20(b) shows a quadrature hybrid using a transformer [4]. The transformer is a 1:1 realization, consisting of a twisted pair wound on a magnetic core. It is assumed that coupling is perfect, so the self-inductance of the windings L is equal to the mutual inductance. Then,

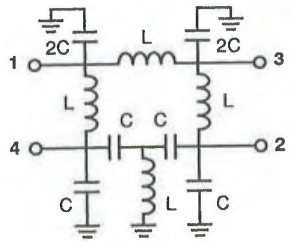


Figure 5.19 A lumped-element 180-deg hybrid.

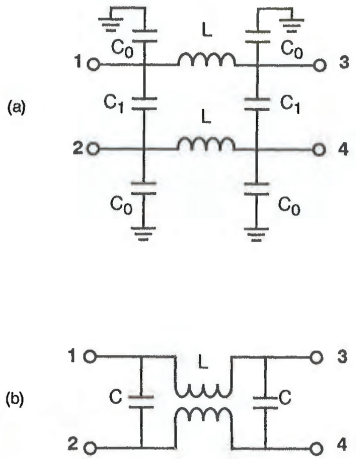


Figure 5.20 Two realizations of a narrowband quadrature hybrid: (a) using lumped L-C elements, (b) using a transformer and capacitors. The bandwidth is only 1 to 2%.

$$Z = \sqrt{\frac{L}{C}}$$

$$f_0 = \frac{Z}{2\pi L} \tag{5.39}$$

In theory, the hybrid is well matched at all frequencies, but the coupling is relatively narrowband and is 3 dB at the center frequency, f_0.

Multiple sections of the coupler shown in Figure 5.20(b) can be cascaded with either intervening transmission lines or lowpass filter sections to improve bandwidth [4]. However, the resulting coupler is relatively complex, requires high Q elements to achieve reasonably low loss, and is not much smaller than a coupled-line hybrid.

A set of uncoupled transmission lines loaded by capacitors can also create a quadrature hybrid that is only $\lambda/8$ long. The reactances of the capacitors are equal to the transmission-line impedances, which are also equal to the port impedances. The bandwidth of this hybrid is approximately 10%. The circuit is shown in Figure 5.21.

5.5.4 Singly Balanced Diode Mixers

A singly balanced diode mixer consists of two diodes interconnected by a balun or hybrid. Two fundamental realizations are possible: the 90-deg and 180-deg forms.

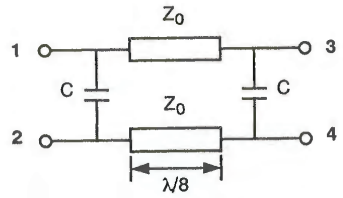


Figure 5.21 A quadrature hybrid realized from two uncoupled transmission lines and capacitors. The length of the line is λ/8, and the bandwidth is approximately 10%.

Figure 5.22 shows both types of mixers. Each consists of two diodes, a hybrid, and a lowpass IF filter. The diodes are connected to mutually isolated ports of the hybrids and the RF and LO are applied to the other pair of isolated ports. Because there is no inherent isolation between the RF or LO and IF ports, a filter is needed.

One of the most important reasons for using a balanced mixer is its inherent rejection of spurious responses (see Section 5.2.2). Most types of balanced mixers reject certain spurious responses where m or n (or both) are even. This phenomenon depends on the even or odd character of m and n; for example, if a mixer rejects the (2, 1) response, it will also reject the (4, 1) and (6, 1) responses.

For several reasons, the 180-deg balanced mixer is most practical for wireless applications:

1. As we have seen, 180-deg hybrids are usually more broadband than 90-deg.
2. The 180-deg-hybrid mixers have better even-order spurious response and IM rejection, and 180-deg mixers reject all (2, 2) responses. They reject the (2, 1)

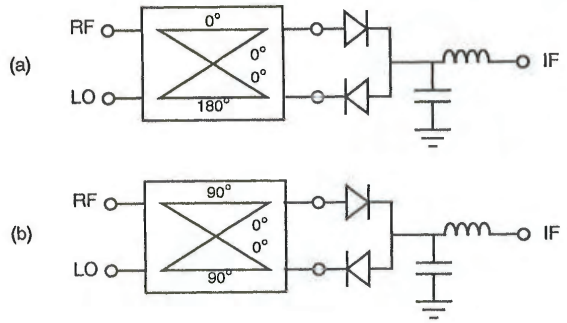


Figure 5.22 A (a) 180-deg and a (b) 90-deg singly balanced mixer.

spurious response if the Δ port is the LO, and the (1, 2) if the Σ port is the LO. The 90-deg hybrid mixer rejects only the (2, 2); it has no appreciable rejection of the (2, 1) or (1, 2).

3. The 90-deg hybrid mixers have a number of odd characteristics. Most important is the fact that the performance depends critically on both the RF and LO ports being well matched over both the RF and LO bands, and the diodes being well matched to the ports. If these requirements are not met, the balance suffers and most of the benefits of a balanced structure—AM noise rejection, spurious-response rejection, and port-to-port isolation—are lost.

Beyond these concerns, the design of the two types of mixers is virtually identical. Thus, we concern ourselves only with the 180-deg mixer.

Figure 5.23 shows a 180-deg mixer that uses a transformer as a hybrid. In this configuration, the RF uses the Δ port of the hybrid and the LO uses the Σ port. The diodes are connected to the remaining pair of mutually isolated ports. The IF port is connected to the diodes through a lowpass filter. This filter is necessary for rejecting the LO; it should also present a short circuit to the diodes at the LO frequency, so the right ends of the diodes are effectively connected to ground. Because the RF voltages at the diodes have a 180-deg phase difference, the diodes' connection point is a virtual ground for the RF. The inductor at the LO port is a current return for the IF; without this inductor, the IF is open-circuited. If the frequency separation between the IF and LO frequency is not large, a more complex filter may be necessary.

An important and often overlooked consideration in balanced mixers is the diodes' dc voltage and current. A dc bias is sometimes used in mixers to improve conversion loss and to reduce LO power requirements. A dc bias is most often used in single-diode millimeter wave mixers; it is often impractical to apply dc bias to a balanced mixer, so no external dc bias is used in the mixer of Figure 5.23. Even

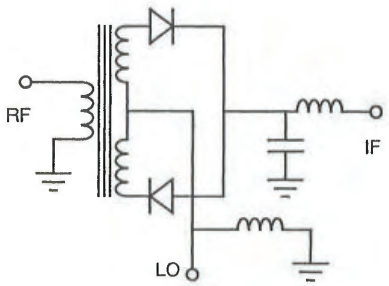


Figure 5.23 A singly balanced, 180-deg hybrid mixer using a transformer as a hybrid.

when possible, use of dc bias in balanced mixers is usually not advisable because inadvertent differences in bias at the two diodes can unbalance the mixer; in most wireless applications, maintaining the mixer's balance is usually more important than achieving a minor improvement in conversion loss.

In Figure 5.23 the diodes are in an antiparallel configuration at dc, so dc current circulates through them and the transformer unimpeded, and the dc voltage across the diodes must be zero.

The mixer of Figure 5.23 includes no RF or LO matching circuits. As with dc bias, matching circuits introduce a means whereby the mixer could become unbalanced, and therefore are usually avoided. Instead, the diodes are selected so that their junction capacitances are negligible, and the transformer, diode *I/V* characteristics, and LO level are selected to achieve a match. Silicon diodes having various barrier heights are available and can be selected to help achieve a match at a satisfactory LO level. The barrier heights of such diodes are adjusted by the selection of doping levels and anode metals, which result in various values of I_0 (5.7). Through these techniques, the knee voltage of the diode can be made as low as 0.3V, although these "low-barrier" diodes have relatively low breakdown voltages and high series resistances.

Diodes for both singly and doubly balanced mixers (which we shall consider in the next section) are often realized as two or four diodes on a single chip. Diodes fabricated in this manner have well-matched characteristics and identical parasitics. The use of such diodes helps to maintain the balance of the mixer.

The diodes' input impedance at the RF or IF port is defined as

$$Z_n = \frac{V_n}{I_n} \tag{5.40}$$

where n is the frequency component at mixing frequency ω_n in (5.25) (if $n < 0$, Z_n is the conjugate of the input impedance). Usually $n = 0$ for the IF and $n = 1$ for the RF. The RF input impedance of the mixer depends, as we have noted, on the IF termination as well as the diode characteristics and LO level. Even so, it is possible to define simultaneous conjugate match values for this impedance; the relation is the same as for conventional two-ports [5]. In general, the optimum IF output impedance is greater than the RF input impedance. Both impedances are high at low LO levels and decrease as LO level is increased.

Because the diodes are in parallel at the IF, each diode sees a 100Ω IF load impedance. Fortunately, as long as adequate LO power is available, this is usually near the optimum value.

5.5.5 Doubly Balanced Diode Mixers

Doubly balanced mixers use at least four diodes, most often in a ring configuration but occasionally in a star. Doubly balanced mixers have several advantages over singly balanced mixers:

1. Doubly balanced mixers reject all (m, n) spurious responses where m, n, or both are even.
2. The IF, RF, and LO ports are mutually isolated in all doubly balanced mixers. Therefore, with properly designed baluns, doubly balanced mixers can have overlapping RF, IF, and LO bands.

Because they use twice as many diodes, doubly balanced mixers require twice the LO power and have 3-dB greater IM intercept points (IP) than singly balanced mixers.

The Ring Mixer

Figure 5.24 shows the most common configuration for a doubly balanced diode mixer, a ring mixer. It is best understood as a switching mixer, in which the diodes are viewed as switches controlled by the LO. In Figure 5.24 points A and A' are virtual grounds for the LO; points B and B' are similarly virtual grounds for the RF. For one-half of the LO cycle, while point B' is positive and B is negative, diodes D1 and D2 are turned on and D3 and D4 are off. During the second half of the LO cycle, the voltage reverses and D3 and D4 are on, while D1 and D2 are off.

When D1 and D2 are switched on, the left half of the RF transformer's secondary is connected to the IF port through the secondary of the LO transformer. The current is in opposite directions in the two halves of the secondary, so no RF voltage is generated across the LO transformer's secondary; it is effectively a short circuit. When diodes D3 and D4 are turned on, the right half of the transformer is connected to the IF, and the polarity of the RF-to-IF connection is reversed. In this manner, the RF is connected to the IF but the polarity is switched every half-cycle.

This process can be described as

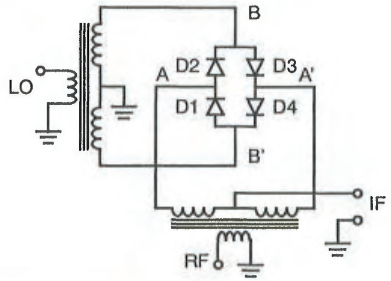


Figure 5.24 A doubly balanced diode ring mixer.

$$V_{IF} = s(t)\,V_{RF}(t) \tag{5.41}$$

where $s(t)$ is a symmetrical square-wave switching function. As such, $s(t)$ has no dc component, so the IF contains no RF component. Furthermore, $s(t)$ has no even harmonics, so mixing with even harmonics of the LO is impossible; similar considerations can be used to show that mixing with even RF harmonics also cannot occur. Mixing occurs between the fundamental frequency of $s(t)$ and $V_{RF}(t)$; the resulting IF current excites both transformers in an even mode, so no IF voltage is developed across their secondaries and the transformers are "invisible" at the IF.

Ring mixers are often realized precisely as shown in Figure 5.24, with two transformers and four diodes. Such transformers can be realized in practice at frequencies up to about 500 MHz or, with care, perhaps 1 GHz.

For design purposes, the ring mixer can be reduced to a single-diode equivalent circuit [5]. When this is done, the source impedance seen by each diode is found to be n^2Z_{RF}, where n is the turns ratio of the transformer (in a conventional trifilar transformer, $n = 2$). At the IF, the diode is terminated in 4 Z_{IF}, where Z_{IF} is the IF load impedance. Finally, at odd LO harmonics the impedance seen by each diode is n^2Z_{LO}, at even harmonics it is zero. This analysis assumes, of course, that the ideality of the transformer is maintained at all harmonics.

In an untuned, adequately pumped single-diode mixer, the optimum RF impedance is usually in the range 50 to 100Ω and the optimum IF load impedance is usually 100 to 200Ω. Thus, if a 50Ω source and load are used, and the diode's junction capacitance is negligible, the ring mixer with a 1:1 transformer usually has port VSWRs below 2.0. Conversion loss in this case is typically 6 to 8 dB.

Microwave Ring Mixer

Above 1 GHz, transformer hybrids are impractical, and distributed (transmission-line) baluns must be used. Microwave baluns differ from the ideal transformers in that they present an open circuit, not a short circuit, to IF currents. Therefore, microwave baluns have nothing that approximates the IF center tap of the transformer, and some other structure must be used to extract the IF. One common technique is simply to use IF blocking capacitors in the baluns and inductive connections from the diodes to the IF [5]; this may be adequate if wide IF bandwidth is not needed.

A better approach is to use a four-wire IF balun [6]. This allows the IF to overlap the RF or LO bands, and provides a very broad IF bandwidth [6]. Figure 5.25 shows a ring mixer using a pair of transmission-line baluns similar to those of Figure 5.16. No IF blocking capacitors are shown at the LO or RF ports because the balun should reject the IF inherently. If the IF bandwidth extends to very low frequencies, however, blocking capacitors may be necessary.

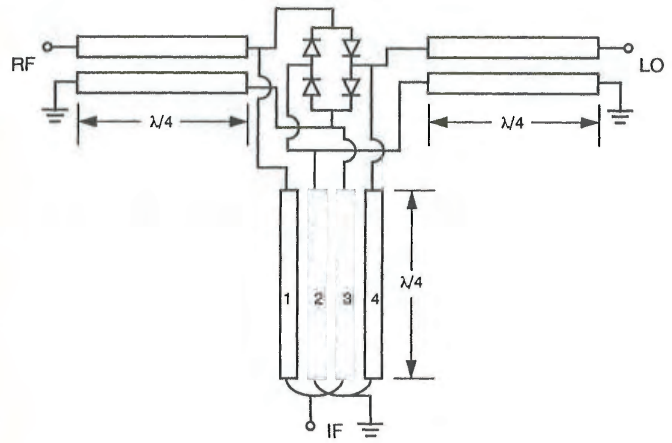


Figure 5.25 A realization of the doubly balanced ring mixer using transmission-line baluns and a four-wire IF balun. Balun lengths are λ/4 at the center of the RF/LO band. The four-wire balun allows the IF to overlap the RF or LO bands.

The IF balun consists of four coupled strips, λ/4 long at the center of the RF and LO frequency ranges. It operates as a transmission line for IF excitation, but as a set of stubs for RF and LO. Strips 1 and 3 in Figure 5.25 realize one side of the IF transmission line, and strips 2 and 4 realize the other. Thus, the odd-mode impedance, which is half the transmission-line impedance, should be equal to half the IF load impedance. Usually, the IF load is 50Ω, so the odd-mode impedance is 25Ω; as with other baluns, the even-mode impedance should be as high as possible. A Lange-coupler analysis can be used to determine the strip width and spacing.

Strips 1 and 3 are excited in an odd mode by the RF; strips 2 and 4 are similarly excited by the LO. Because these strips are shorted at the load end, they behave as high-impedance shorted stubs, their input impedance is high, and the IF balun has little effect on the RF or LO.

Effects of Imperfect Balance

Cancellation of AM LO noise and spurious responses, and the inherent port-to-port isolation of balanced mixers, occur because the currents generated by these phenomena ideally have equal amplitudes and phase differences of 180 deg. These currents cancel at the IF port. Because of imbalances in the hybrid, differences in

the diodes' characteristics, or asymmetries in the layout or physical structure of the mixer, the amplitudes or phases may differ from the ideal. The currents then do not cancel completely.

The rejection R in dB under these circumstances is [5]

$$R = -10\log\left(\frac{1 - 2\sqrt{G}\cos(\theta) + G}{1 + 2\sqrt{G}\cos(\theta) + G}\right) \tag{5.42}$$

where G is the gain imbalance (a ratio) and θ is the phase imbalance. This expression is valid for all rejection phenomena.

5.5.6 Image-Rejection Mixers

The image-rejection mixer (Figure 5.26) is realized as the interconnection of a pair of balanced mixers. It is especially useful for applications where the image and RF bands overlap, or the image is too close to the RF to be rejected by a filter. The LO ports of the balanced mixers are driven in-phase, but the signals applied to the RF ports have a 90-deg phase difference. A 90-deg IF hybrid is used to separate the RF and the image bands. A full discussion of the operation of such mixers is a little complicated; such a discussion can be found in [5]. Equation (5.42) can be used to determine the image rejection.

The most difficult part of the design of an image-rejection mixer is the IF hybrid. If the IF frequency is fairly high, a conventional RF or microwave hybrid can be used. However, if the mixer requires a baseband IF, the designer is placed in the problematical position of trying to create a Hilbert-transforming filter, a theoretical impossibility. Fortunately, it is possible to approximate the operation of such a filter over a limited bandwidth; for specific design information see [7].

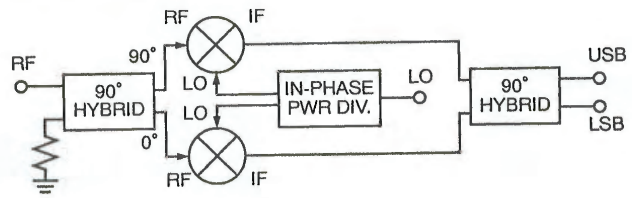


Figure 5.26 Image-rejection mixer.

5.6 ACTIVE FET MIXERS

Active FET mixers have several advantages, and some disadvantages, in comparison to diode mixers. Most significantly, an active mixer can achieve conversion gain, while diode and other passive mixers always exhibit loss. This allows a system using an active mixer to have one or two fewer stages of amplification; the resulting simplification is especially valuable in circuits where small size and low cost are vital. A precise comparison of distortion in diode and active FET mixers is difficult to make because the comparison depends on the details of the system. Generally, however, it is fair to say that distortion levels of well-designed active mixers are usually comparable to those of diode mixers.

Paradoxically, it is usually easy to achieve good conversion efficiency in active mixers even when the design is fundamentally poor. Thus, many research papers and trade journal articles have reported active mixers having good conversion gain but poor noise figures or high distortion. As a result, active FET mixers have gained a reputation for low performance. Nevertheless, achieving good overall performance in active FET mixers is not difficult; it involves only the observation of a few simple rules. We will discuss these in Section 5.6.2.

Because FETs cannot be reversed, as can diodes, balanced FET mixers invariably require an extra hybrid at the IF. This can be avoided only by using a *P*-channel device instead of an *N*-channel device, or vice versa; however, this is possible only in silicon circuits, and even then the characteristics of *P*- and *N*-channel devices are likely to be significantly different. We examine this further in Section 5.6.4.

5.6.1 Single-Device Active FET Mixers

As with diode mixers, the single-device FET mixer is minimally useful, but is a valid prototype for a balanced mixer. In a single-device FET mixer, filters must be used to separate the RF, IF, and LO, and it is rarely possible to have overlapping bands. As we shall see, LO-to-IF isolation is a special concern in FET mixers, and requires careful IF design.

Transconductance Mixers

Although a number of configurations are possible, the best mode of operation for a FET mixer is as a transconductance mixer. In this mode, the LO, applied to the gate of the device, modulates the FET's transconductance. In a conventional downconverter, the fundamental-frequency component of the transconductance is used for mixing. Maximizing this frequency component of the transconductance optimizes the gain, noise figure, and IM intercept points.

When the peak transconductance is limited, as it is in any FET, the fundamental LO-frequency transconductance component is maximized when the transconductance waveform is a rectangular pulse train having a peak value equal to the maximum transconductance and a 50% duty cycle. To realize this waveform, the FET must remain in current saturation throughout the LO cycle. This requires that (1) the drain voltage remain invariant at the dc value throughout the cycle, and (2) the FET's gate be biased at pinchoff. The first condition dictates that the drain termination for the LO be a short circuit; the second is easily satisfied through the use of external dc bias. When these conditions are met, the transconductance waveform will approximate a half-sinusoidal pulse train, which is as close to the optimum waveform as one can achieve in practice.

The RF signal is combined with the LO and applied to the gate. Mixing then occurs between the transconductance $g_m(t)$ and the RF signal. The small-signal drain current is

$$i_d(t) = g_m(t)v_{RF}(t) \tag{5.43}$$

Usually the fundamental-frequency component of the transconductance is used.

Equation (5.43) is identical in form to (5.22), so one might expect a full conversion matrix to be necessary to describe an active FET mixer. In general, this is the case; however, in active FET mixers it is often possible to obtain a reasonably accurate analysis without so much analytical complexity. Unlike the diode, the FET inherently separates $v_{RF}(t)$ at the gate and $i_d(t)$ at the drain; the FET's drain-to-gate coupling is usually very weak, especially when the input and output are tuned to different frequencies. Consequently, one can assume with reasonable accuracy that the small-signal gate voltage consists only of $v_{RF}(t)$ and that the device is unilateral. The drain current $i_d(t)$ is then given directly by (5.43), and the IF component of the current can be found in a straightforward manner.

If feedback effects are ignored, the FET mixer's conversion gain is

$$G_c = \frac{G_{m,\max}R_L}{16\omega_{RF}^2 C_{gs}^2(R_s + R_i + R_g)} \tag{5.44}$$

where $G_{m,\max}$ is the peak transconductance, R_L is the IF load resistance, and the other parameters are as shown in Figure 5.7. In the derivation of (5.44), we have assumed that the FET's gate is conjugate matched at the RF frequency. The minimum LO power required to pump the mixer when the gate is conjugate matched to the LO is

$$P_{LO} = 0.5V_p^2\omega_p^2 C_{gs}^2(R_s + R_i + R_g) \tag{5.45}$$

We have assumed in deriving (5.45) that the gate is biased at the pinchoff voltage, V_p, and is pumped to a peak voltage of 0V. It is often difficult to match the gate successfully, especially at low frequencies where the gate's input impedance is very high. In this case, the required LO power may be greater than the value given by (5.45).

Equation (5.44) implies that the FET mixer's gain can be made arbitrarily large by selecting a high value of R_L. This conclusion is valid, within certain limits. The pumped FET's output impedance is remarkably high, on the order of a few thousand ohms, and for all practical purposes the FET's drain can be considered a current source. The value of R_L is limited by practical considerations, primarily realizability, bandwidth, and stability.

A practical difficulty in using FET mixers is that the FET is, after all, an amplifying device. As such, it amplifies the LO. To prevent high levels of LO leakage into the IF, the IF filter in a FET mixer must have considerably more rejection than would an LO filter in a diode mixer.

Another practical concern is amplifier-mode gain at the IF. Again, because the FET is inherently an amplifier, it can amplify an IF-frequency signal applied to or leaking into the input. If this signal is out-of-band noise from a low-noise amplifier, or even noise from a bias circuit or some other spurious noise source, the noise will be amplified and transmitted to the IF, making the mixer appear to have an extraordinarily high noise figure. Amplifier-mode gain is one of the major causes of high noise figures in FET mixers; it can be prevented by careful design of the input circuit.

Other Types of FET Mixers

Sometimes the LO is applied to the source of the FET instead of the gate. This reduces the level of the LO at the IF port slightly (although not as much as one might hope) and provides a modest degree of LO-to-RF isolation. It also provides a better LO match, especially at low frequencies, because the LO input impedance is approximately $1/\langle G_m \rangle$, where $\langle G_m \rangle$ is the time-average transconductance. This configuration requires careful bypassing of the source at the IF frequency or instability may result. See Section 5.6.2 for an example of a source-driven FET mixer.

The final option is, of course, to apply the LO to the drain. This is a distinctly different mode of operation, because varying the drain voltage of a FET in current saturation does not vary its transconductance. In order to achieve mixing, the LO voltage must be great enough to drive the drain voltage very low over part of the LO cycle, below the knee of the drain I/V characteristic, reducing both transconductance and drain-to-source resistance. This low average drain-to-source resistance reduces the FET's conversion gain significantly.

Because a large LO voltage is impressed on the drain, LO-to-IF isolation is likely to be poor. Furthermore, the IF filter must present an open circuit to the drain, not the optimum short circuit. This has a number of deleterious effects: it exacerbates

intermodulation distortion, degrades stability, and creates amplifier-mode gain at the IF. This is not a good way to use a FET as a mixer.

5.6.2 Single-Device Mixer Circuits

Gate-Driven Transconductance Mixer

Figure 5.27 shows a single-device FET mixer. The RF/LO diplexer must combine the RF and LO and also provide matching between the FET's gate and both ports. The IF filter must provide an appropriate impedance to the drain of the FET at the IF frequency, and must short-circuit the drain at the RF and especially at the LO frequency and its harmonics.

Especially if the RF and LO are well separated in frequency, it may be virtually impossible to match the gate effectively at both the RF and LO frequencies. In this case, it is usually better to match the gate at the RF and accept a mismatch at the LO; the only cost will be increased LO power. Usually, it is simplest to combine the RF and LO with a separate filter diplexer. To prevent amplifier-mode gain, the RF matching circuit must reject the IF; an RF-frequency shorted stub at the FET's gate is adequate. The gate bias circuit often can provide the IF short.

As well as short-circuiting the drain at unwanted frequencies, the IF circuit must transform the IF load impedance (usually 50Ω) to an appropriate value at the FET's drain. At IFs below approximately 1 GHz, a 2:1 transformer provides an IF load of 200Ω; this provides good gain even with low-transconductance silicon MOSFETs, and, if stray capacitances are minimized, broad bandwidth. At higher frequencies, lumped or distributed matching circuits are necessary; with these, a high IF load impedance is more difficult to achieve. The RF/LO short circuit at the drain must be designed carefully. A single microstrip stub usually does not have adequate Q to provide an effective short circuit; a more complex filter circuit is necessary. Textbook filter designs are also inadequate; even if they provide high rejection out-of-band, the port impedance may be reactive, not necessarily a short circuit. A good

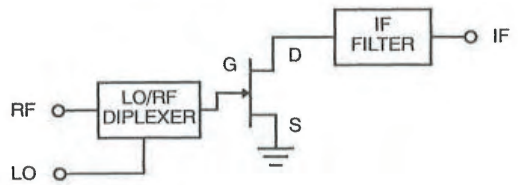


Figure 5.27 A single-device FET mixer.

approach is to use a textbook filter design as a starting point, and to optimize it numerically to achieve the desired terminations.

The only remaining dilemma in the design of a single-gate FET mixer is the determination of the input reflection coefficient. This can be found from nonlinear analysis or can be estimated. A good initial estimate of the input reflection coefficient can be found from S parameters of the device at approximately 50% I_{dss}, with a short-circuit load. The input reflection coefficient Γ_{in}, in terms of the FET's S parameters, is

$$\Gamma_{in} = S_{11} - \frac{S_{12}S_{21}}{1 + S_{22}} \tag{5.46}$$

Source-Driven Mixer

Figure 5.28 shows a single-device source-driven mixer that has been used successfully at frequencies below 2 GHz. The FET has a relatively large source resistor, R_s, that keeps the FET biased just above pinchoff when no LO power is applied. The value of this resistor is not critical; because the LO input impedance at the FET's source is much lower than R_s, the resistance has little effect on LO-port VSWR. The resistor R_g keeps the dc gate voltage at ground potential. The negative-going half-cycle of the LO waveform drives the FET's source negative, causing the FET to become active; the positive half-cycle keeps the FET turned off. In this way, the FET is switched in much the same manner as the gate-driven mixer.

In this circuit, the LO port is in series with the RF circuit; as a result, there is no significant RF-to-LO isolation. LO-to-RF isolation, however, is inherent in the

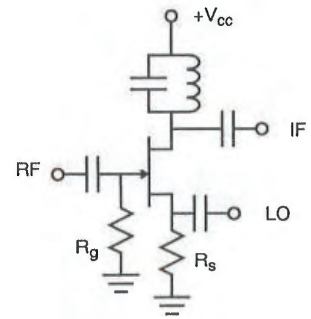


Figure 5.28 A single-device FET mixer having the LO applied to the source. This results in improved LO-to-RF isolation.

circuit. As with the gate-driven mixer, LO-to-IF isolation is inherently poor; the drain resonator must be designed carefully to reject LO leakage. Because the source is not bypassed for RF currents, the conversion gain of this mixer is lower than that of the gate-driven mixer.

5.6.3 Dual-Gate FET Mixer

Although several configurations and modes of operation of dual-gate mixers are possible [8], the configuration we shall examine, shown in Figure 5.29(a), usually provides the best performance in most receiver applications. In this circuit, the LO is connected to the gate closest to the drain (gate 2), while the RF is connected to the gate closest to the source (gate 1). An IF bypass filter is used at gate 2, and an

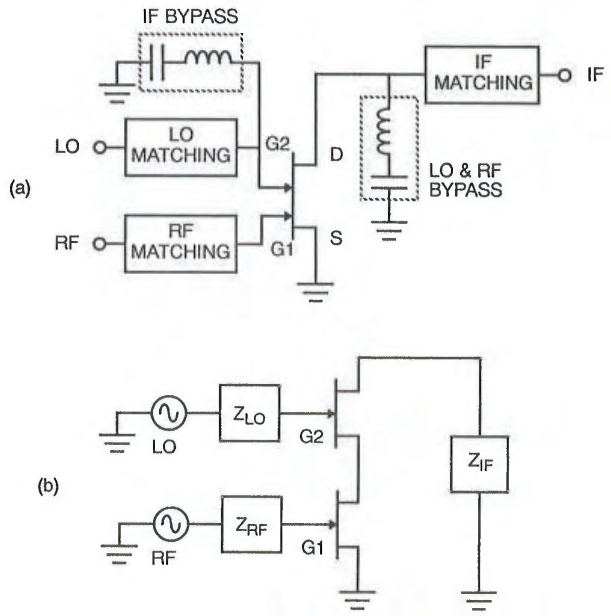


Figure 5.29 (a) A dual-gate FET mixer; (b) the FET mixer's equivalent circuit, in which the dual-gate FET is represented by two single-gate FETs in a cascode connection.

LO/RF filter, similar to those used in single-gate mixers, is used at the drain. Although not shown in the figure, dc bias must be provided to both gates and to the drain.

The operation of a dual-gate mixer is very different from that of a single-gate transconductance mixer; it is similar to that of a drain-pumped single-gate mixer. To understand the operation, we view the dual-gate FET as two single-gate FETs in a cascode connection (Figure 5.29(b)). In this view, it is evident that the upper FET acts as a source follower, applying LO voltage to the drain of the lower FET. The lower FET then provides mixing, operating as a drain-pumped mixer. We noted in Section 5.6.1 that mixing occurs in the drain-pumped mixer because the LO drives the drain voltage low enough, over part of the LO cycle, to push the FET into its linear region, and the FET provides mixing by a combination of time-varying transconductance and drain-to-source resistance. The average drain-to-source conductance is a parasitic resistance in parallel with the lower FET's drain, reducing the conversion efficiency. Because of this parasitic element, the dual-gate FET mixer is inferior in terms of conversion gain and noise figure to the active single-gate FET mixer. This inferiority is evident in published results: few reported dual-gate FET mixers have noise figures or gain as good as single-gate. The dual-gate mixer should be viewed as a tradeoff between a modest noise and gain performance and the low cost that results from the use of a single-device, unbalanced circuit, and fewer filters.

Even with lower gain and higher noise, properly designed dual-gate mixers have achieved third-order intercept points that modestly exceed those of single-gate mixers. It is not clear whether this is inherent in the dual-gate device or simply good fortune on the part of the designers. A reason often given is that the dual-gate mixer's *I/V* characteristic is more linear than the single-gate; however, dual-gate FET mixers usually do not operate in the most linear part of the characteristic. Whatever the reason, it is fair to state that dual-gate mixers have distortion at least as low as single-gate designs.

The IF current generated in the channel of the lower FET must pass through the channel of the upper FET; the latter acts as an inefficient IF amplifier stage. To operate as efficiently as possible, the gate of this FET must be grounded at the IF frequency; this is the purpose of the IF bypass filter in Figure 5.29(a). Furthermore, the drain voltage must remain high enough to keep the FET in its saturation region throughout the LO cycle; this requires the use of a filter to short-circuit the drain at the LO frequency. Of course, to provide good RF-to-IF as well as LO-to-IF isolation, this filter should short-circuit the RF as well.

5.6.4 Singly Balanced FET Mixers

In a singly balanced diode mixer, the LO and RF are applied to mutually isolated ports of a hybrid and the diodes are connected to the other pair of ports. No hybrid is needed in the IF, because the diode polarities are reversed. Unfortunately, no such

reversal is possible in most types of FETs, so an IF hybrid is required. The only way to realize the analog of a reversed diode is to use a *P*-channel device for one of the FETs and an *N*-channel device for the other. Even in technologies where both *P*- and *N*-channel devices can be realized, this may not be a good thing to do. Because of the differences in *P* and *N* materials, *P*- and *N*-channel devices can never be identical, and in fact are often substantially different. Unfortunately, to achieve the benefits of a balanced mixer, the characteristics of the FETs must be closely matched.

Figure 5.30 shows one type of singly balanced FET mixer. Although this mixer uses a transformer hybrid for the LO and RF, any appropriate type of hybrid can be used. A matching circuit is needed at the gates of both FETs; to prevent unbalancing of the mixer, the circuits should not be tuned independently. Although the bias circuitry is not shown in the figure, dc bias must also be applied to the gates. The IF filters provide the requisite short circuits to the drains at the LO and RF frequencies, and additionally provide IF load impedance transformations.

The singly balanced mixer of Figure 5.30 is effectively two single-device mixers interconnected by hybrids. Thus, the design considerations for single-device mixers apply equally to this type of doubly balanced mixer.

Differential Mixer

Figure 5.31 shows a mixer that is similar in topology to a differential amplifier. The RF is applied to the lower FET, and the LO is applied through a balun or hybrid to the upper FETs.

This mixer operates as an alternating switch, connecting the drain of the lower FET alternately to the inputs of the IF balun. The node connecting the sources of

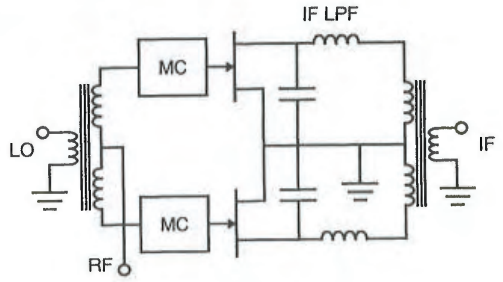


Figure 5.30 A singly balanced FET mixer. MC is the RF/LO matching circuit, and IF LPF is the IF output lowpass filter. In contrast with the singly balanced diode mixer, a transformer or hybrid is needed at the output.

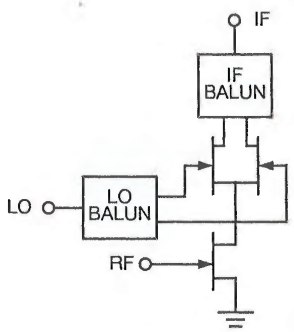


Figure 5.31 Differential mixer. This is a type of singly balanced mixer. Matching circuits and an IF filter may be necessary; these are not shown.

the upper FETs is a virtual ground for the LO, and therefore the LO simply switches the two upper FETs on and off on alternate half cycles. The upper FETs are biased slightly above pinchoff, so each is turned on over at least 50% of the LO cycle. In this way one of the upper FETs is always on, so the lower FET remains in its saturation region throughout the LO cycle. When on, the upper FETs remain in their linear regions. To accomplish this, the lower FETs must be biased between pinchoff and approximately 0.3 I_{dss}; higher bias may cause the lower FET to drop into its linear region over part of the LO cycle.

An LO matching circuit may be needed. Because the RF and LO circuits are separate, the gates of the upper FETs can be matched at the LO frequency and there is no tradeoff, as there often is in single-device mixers, between effective LO and RF matching. Similarly, the lower FET can be matched effectively at the RF. An IF filter, not shown in the figure, is necessary to reject LO current; in theory, the output balun rejects the RF, but in practice its bandwidth may not be great enough to do so. Because the upper FETs are not biased in current saturation, the LO short circuit at the FETs' drains, which is extremely important in other active mixers, is not quite as critical in this circuit.

The design of this mixer is straightforward. Because the source node of the upper FETs is a virtual ground, the input impedance is simply that of the input loop of the FET. The LO input impedance Z_{LO} is approximately

$$Z_{LO} \cong R_s + R_g + R_i + \frac{1}{\omega_{LO} C_{gs}} \tag{5.47}$$

where ω_{LO} is the LO radian frequency and the other variables are defined in Figure 5.7. The lower FET is effectively an amplifier. Its load impedance is the impedance looking into the sources of the upper FETs, approximately the "on" resistance of the channels. Its input impedance can be found from *S* parameters. Finally, the IF circuit is designed in the usual manner: to present a moderately high IF impedance to the drains of the FETs, but a short circuit to LO and RF frequencies. The design of these matching and filtering structures is discussed in Section 5.6.2.

A version of this mixer that does not require an LO balun is shown in Figure 5.32. In this circuit, the source node of the upper FETs is no longer a virtual ground, so, to prevent imbalance, the lower FET must have a high drain-to-source resistance. This is usually not the case in GaAs FETs, but in MOSFETs and silicon junction FETs the resistance is usually adequate.

5.6.5 Doubly Balanced FET Mixers

Gilbert Multiplier

The venerable Gilbert analog multiplier [9] is frequently used as an RF or microwave mixer.[4] The classical Gilbert mixer depends on a cancellation between the exponential *I/V* and logarithmic *V/I* characteristics of bipolar transistors to produce an accurate

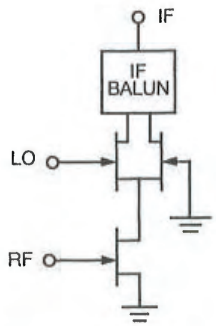


Figure 5.32 A version of the differential mixer that does not require an LO balun. Because the source node of the upper FETs is no longer a virtual ground, this works well only if the lower FET has a high drain-to-source resistance.

[4]Although Gilbert's original paper described a multiplier using bipolar transistors, FET implementations having the same circuit topology are in common use, and are named in honor of the original invention.

product characteristic. However, any circuit having the same topology, using either bipolar transistors or FETs, cancels the even-degree parts of the devices' *I/V* characteristics and preserves the odd-degree parts, producing multiplication. Any such circuit is therefore a balanced mixer.

Figure 5.33 shows a doubly balanced FET mixer based on the Gilbert multiplier. Like many doubly balanced mixers, this mixer consists of two of the singly balanced mixers shown in Figure 5.31. Each half of the mixer operates in the same manner as that of Figure 5.31 and is designed in exactly the same manner. The interconnection of the outputs, however, causes the drains of the upper four FETs to be virtual grounds for both the LO and RF, as well as for even-order spurious responses and IM products.

5.7 PASSIVE FET MIXERS

Passive FET mixers, often called *FET resistive mixers*, are used frequently in wireless applications. They have a number of advantages over diode and active FET mixers.

In a FET resistive mixer, the time-varying channel resistance is used for frequency conversion. The LO is applied to the gate, but no dc bias is applied to the channel (the gate may be dc biased, usually near pinchoff). The FET then operates entirely in its linear region, where the channel is effectively a gate-voltage-controlled

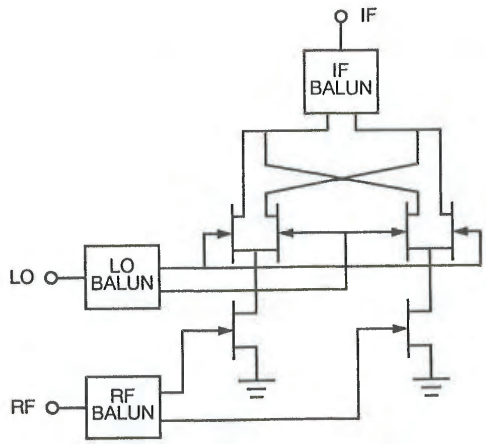


Figure 5.33 A doubly balanced mixer based on the Gilbert multiplier. It consists of two of the singly balanced mixers shown in Figure 5.31.

resistor. The RF signal is applied to the drain, and IF currents are filtered from the drain or source.

The primary advantage of the FET resistive mixer is its very low levels of intermodulation distortion. In diode or active FET mixers, the same nonlinearity that provides mixing causes intermodulation distortion. A strong nonlinearity is needed for efficient mixing, but that same nonlinearity causes distortion. In a FET resistive mixer, no such nonlinearity is employed; the channel resistance can be perfectly linear and mixing is still achieved. In fact, the channel resistance at low signal levels is very linear, so low distortion results; it is not unusual to achieve third-order input intercept points above 30 dBm in such mixers. Conversion loss and LO power are comparable to diode mixers.

The noise generated in the channel of a resistive FET is entirely thermal in origin, causing the mixer's single-sideband noise figure in decibels to be equal at room temperature to its conversion loss. This contrasts with diode mixers, which have substantial shot noise, and whose SSB noise figures usually exceed the conversion loss by at least a few tenths of a decibel.

Single-device resistive FET mixers are very practical components, especially at frequencies below approximately 2 GHz. At these low frequencies, the gate-to-channel capacitance may be negligible and LO-to-RF and LO-to-IF isolation is very good. Unfortunately, the gate-to-drain capacitance in an unbiased FET is as great as the gate-to-source capacitance, so at high frequencies the LO-to-RF isolation of a single-device mixer may not be adequate. Furthermore, the LO voltage on the drain increases the mixer distortion. Balanced FET resistive mixers also exist. In the best of these, both the drain and source terminals of each FET are virtual grounds, so the coupling problem is inherently solved.

5.7.1 Single-Device FET Resistive Mixers

Figure 5.34 shows the basic FET resistive mixer. As with other mixers, filters are used to separate the various mixing frequencies, to short-circuit unwanted mixing

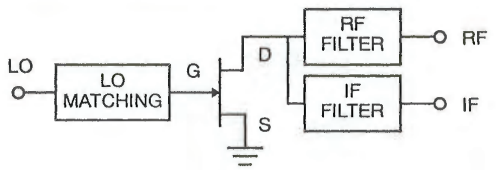


Figure 5.34 A FET resistive mixer. The drain is unbiased and mixing occurs because of the time-varying channel conductance.

products, and, if necessary, to provide matching. The design of a FET resistive mixer has a number of now-familiar requirements:

1. The drain of the FET must be short-circuited at the LO frequency. Short-circuiting it at LO harmonics is less important, because the capacitive nonlinearity generates only weak LO harmonics at the drain. The short circuit prevents LO voltages from appearing at the drain. This minimizes distortion by keeping the drain voltage at zero (except, of course, for the small-signal RF and IF voltages) where it is most linear.
2. The RF circuit must provide filtering to prevent IF leakage into the RF port, as well as matching. Matching is not particularly difficult; the real part of the RF input impedance at the FET's drain of a conventional 0.5 μm by 250 μm Ku-band FET, biased at pinchoff, is close to 50Ω.
3. Similarly, the IF must provide both filtering and matching. As with diode mixers, the IF output impedance is usually higher than the RF input impedance, approximately 100Ω.
4. In order to minimize LO power, the gate must be matched. If the LO short at the drain is well realized, the input impedance is given approximately by (5.47). Ideally, the gate should be short-circuited at the RF frequency, although at low frequencies RF-to-LO coupling may be low enough that this may not matter.
5. The optimum dc gate bias is usually at or below pinchoff. Distortion is lowest when the LO gate voltage variation is as great as possible and swings from the threshold of avalanche breakdown to the threshold of gate conduction. Such a large gate voltage may adversely affect reliability; fortunately, the voltage can be reduced a few decibels with minimal effect on distortion.
6. The drain must be short circuited at dc. A dc return path at the drain is usually required. This can be implemented as part of the RF or IF filter.

Figure 5.35 shows an example of a practical FET resistive mixer for use at frequencies below approximately 1 GHz. The RF is applied to the drain, and the IF is filtered from the source (although, in an unbiased FET there is little difference between the source and the drain). Because of the low frequency, the real parts of the input and output impedances are close to 50Ω and the reactive parts are negligible. In a monolithic circuit, the device width can be adjusted to optimize the port impedances. The parallel L-C tuned circuits ω_{RF} and ω_{IF} provide a short circuit to both the drain and source at the LO frequency; they also help to guarantee good LO-to-RF and LO-to-IF isolation.

The LO input impedance at the gate is virtually impossible to match at low frequencies. The value R_g, which is necessary for proper dc bias, can be selected to provide a good VSWR. If VSWR is not important, R_g can have a large value; in this

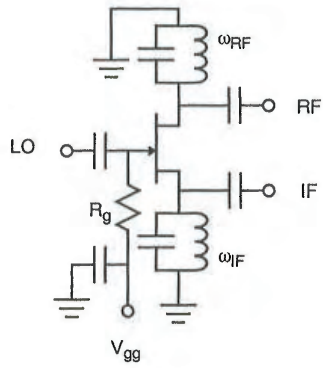


Figure 5.35 A series-switch FET resistive mixer.

case, the required gate voltage can be achieved with up to 6 dB less available LO power.

Even though this is a single-device mixer, its performance is in many ways similar to a balanced mixer. Because it uses a resistive FET, its distortion is very low; even-order distortion is often lower than that of a balanced diode mixer. AM LO noise is largely rejected by the weak gate-to-drain coupling. Port-to-port isolation, also attributed to the weak coupling, is inherently very good.

5.7.2 Balanced FET Resistive Mixers

Balanced FET resistive mixers have essentially the same advantages mixers as balanced diode or active-FET mixers: inherent port isolation, rejection of AM LO noise, and rejection of certain spurious responses and IM products. Figure 5.36 shows a singly balanced FET resistive mixer. At low frequencies, the LO balun can be realized by a transformer, as shown; at higher frequencies, a 180-deg transmission line or one of the structures in Figures 5.16 to 5.19 could be used. The LO pumps the two FETs 180-deg out of phase, but the RF is applied in phase at the drains. The IF currents in the FETs' channels therefore have a 180-deg phase difference, and an output hybrid or balun is necessary to subtract them. Of course, the two FETs generally require twice the LO power of a single FET, and have 3-dB greater IM intercept points. The mixer rejects all (2, 2) and (1, 2) spurious responses; (2, 1) responses are not rejected.

As with other types of gate-driven FET mixers, the LO input impedance at low frequencies is virtually an open circuit, and is thus almost impossible to match.

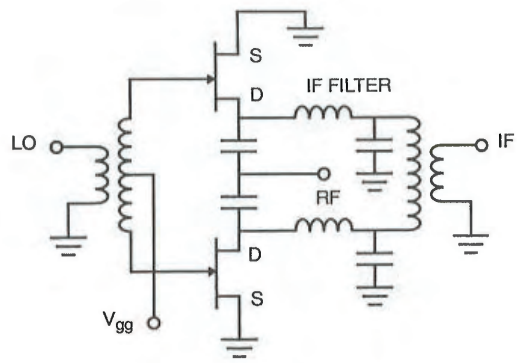


Figure 5.36 Singly balanced FET resistive mixer. The RF connection is a virtual ground for both the LO and the IF.

It is important to note in this case that the task of the LO circuit is not simply to provide a low LO VSWR, but also to maximize the LO gate voltage for a given available power. If matching is impossible, the next best option is to transform the source impedance to a high value, increasing the open-circuit LO voltage. The increase in source impedance is limited primarily by the gate capacitance of the FETs and stray capacitances of the circuit, and practical limitations in the realization of the transformer. Another advantage of an LO transformer is the short circuit it provides to the RF at the gate. This is the optimum termination.

The RF input impedance is half that of the single-device mixer. This may be uncomfortably low, on the order of 25Ω. To achieve an adequate VSWR, an RF matching circuit (not shown in Figure 5.36) may be necessary. The FETs are effectively in series at the IF, and the common 2:1 transformer (i.e., 4:1 impedance ratio) usually terminates the IF quite well.

An attractive feature of this mixer is that the drain-connection node is a virtual ground for the LO and IF; thus, no drain RF filter is needed. Similarly, the RF drives the IF transformer in an even mode, so RF leakage is inherently rejected from the IF port. An IF filter is still needed, however, to prevent LO leakage and to prevent the IF transformer from short-circuiting the RF signal at the FETs' drains.

5.7.3 FET Resistive Ring Mixer

The ring mixer in Figure 5.37 is a type of commutating mixer and operates in a manner analogous to the diode-ring mixer (Figures 5.24 and 5.25). As with the

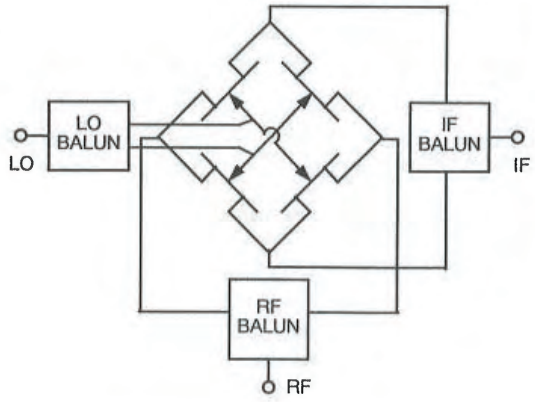


Figure 5.37 Commutating FET ring mixer. All ports are inherently isolated and no matching circuits are needed at low frequencies.

diode ring, the LO causes alternate pairs of devices to turn on, switching the polarity of the RF connection to the IF. This causes the RF to be multiplied by the switching waveform, and mixing between the RF signal and the fundamental frequency component of the switching waveform generates the IF output.

An important property of the FET ring mixer is that the nodes of the ring are all virtual grounds for the LO, the IF connection nodes are virtual grounds for the RF, and the RF connections for the IF. The gates are also virtual grounds for both the RF and IF. As with the diode mixer, the topology of the circuit isolates the ports, making filters unnecessary. The result is that the mixer's bandwidth is limited primarily by the RF, IF, and LO baluns.

The impedances of this mixer are not the same as those of the diode ring. The RF input and IF output impedances at the terminals of the ring are the same as those of a single device; thus, if transformers are used, a 1:1 is best for the RF and IF. (A slightly higher turns ratio might be better for the IF, but transformers having noninteger turns ratios usually have poor performance.)

Optimizing the distortion of a FET ring mixer is more or less automatic: the FETs effectively short-circuit each other at the unwanted mixing frequencies, inherently providing ideal terminations. No filtering or special efforts are required to achieve the optimum terminations. If GaAs devices are used, dc gate bias is necessary, and the gates should be biased near pinchoff. In enhancement MOSFETs having a threshold voltage around 1V, dc bias may be unnecessary.

5.7.4 Modeling FETs for Resistive Mixer Analysis

Conventional FET models, discussed in Section 5.3.2, are adequate for conversion-loss analysis of resistive FET mixers, but because the *tanh* function does not model the drain nonlinearities adequately, these models should not be used for IM analysis. The *I/V* function should be fit to measured data near zero drain-to-source voltage, because this is where the device is used; a poor fit to the *I/V* characteristic in the current-saturation region is tolerable. To achieve good LO power estimates, the equivalent-circuit element values used in (5.45) and the pinchoff voltage should be well characterized.

The equivalent circuit of a resistive FET need not be as complex as an active FET. Figure 5.38 shows the equivalent circuit of a FET used as a resistive mixer, including only the most important elements. A number of the elements in Figure 5.7 have been eliminated as unnecessary: R_{ds} is included implicit in I_d, and C_{ds}, which was part of the FET's biased channel, does not exist in the unbiased device (a small parasitic capacitance may remain; however, this is completely swamped by the low average value of R_{ds}). Similarly, R_i is usually unnecessary. The diodes used to model the gate-to-channel conduction in junction-gate devices should be included, as these present an important limit to achievable LO level. Note that the FET's gate-to-drain capacitance C_{gd}, when the drain is unbiased, is much greater than the biased value, and C_{gs} is somewhat less.

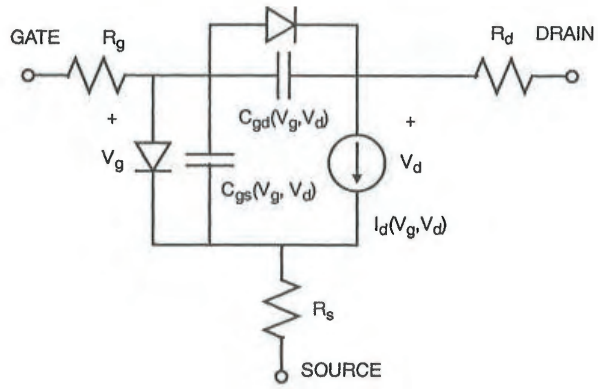


Figure 5.38 Equivalent circuit of a FET for resistive mixer analysis. A number of elements from Figure 5.7 have been eliminated; these are not necessary in resistive FET mixer analysis.

5.7.5 Mixer Comparison Table

Table 5.1 contains a brief comparison of mixer types and a qualitative description of their characteristics. Although such comparisons contain only a limited amount of information, they apparently are useful to readers. A similar table appears in [5].

5.8 HARMONIC-BALANCE SIMULATION OF MIXERS

A very important development for mixer design is the availability of general-purpose nonlinear-circuit simulators. These simulators, which use harmonic-balance analysis, are capable of analyzing and sometimes optimizing mixers having arbitrary circuit topology. As valuable as they are, their use requires some judgment, or invalid or inaccurate calculations may result.

Classical harmonic-balance analysis is applicable only to a circuit having a period excitation; the waveforms in a mixer arise from the combination of two noncommensurate frequencies, the RF and LO, and clearly the combination is not periodic. Fortunately, it is possible to modify the classical method to allow its application to mixers. Two approaches are possible. The first, called *multitone harmonic balance*, involves the use of aperiodic Fourier transforms to allow a limited number of noncommensurate excitations. The second, called *conversion-matrix analysis* (outlined in Section 5.4), is considerably more accurate and orders of magnitude faster for conversion efficiency and port-impedance calculations.[5] Most simulators offer both methods; because it is much faster and more accurate, only the conversion-matrix method should be used in mixer design.

Harmonic-balance analysis is an iterative method. It estimates a solution and iteratively improves the solution until the errors are acceptably low. As such, it is subject to convergence difficulties; the stronger the nonlinearity and the stronger the excitation, the greater the likelihood of convergence failure. Simulators often use a "source stepping" technique, wherein the excitation is increased stepwise, allowing convergence to be achieved more easily at each level. More sophisticated methods for improving convergence have been developed, but few have been implemented in commercial products.

Errors in models and circuit descriptions are a common cause of convergence failure. Chief among these are open-circuiting a diode at dc and using a value of threshold voltage in FET mixers that is not precisely the zero of $I_d(V_g)$. Any discontinuity in an I/V characteristic is likely to result in poor convergence.

The manner in which the design process is implemented does much to guarantee the success of a harmonic-balance simulation. In the first stage of the design, the circuit should be as simple as possible: ideal hybrids or transformers should be used,

[5]The author recognizes that this is a controversial statement, and is subject to strong objections, primarily from the marketing departments of software companies.

Table 5.1
Mixer Comparison Table

Mixer Type	VSWR			Port-to-Port Isolation			LO AM Noise Rej.	LO Spur. Sig. Rej.
	RF	LO	IF	RF/IF	LO/RF	LO/RF		
Single-diode	Depends on matching circuits	Depends on matching circuits	Depends on matching circuits	Depends on filters	Depends on filters	Depends on filters	None	None
Singly balanced diode (180 deg)		Same as single-diode		Depends on filters	Good; equal to hybrid isolation	Depends on filters	Good	Good
Singly balanced diode (90 deg)	Good	Good	Same as single-diode	Depends on filters	Poor	Depends on filters	Good	Good
Doubly balanced diode	Good	Good	Good	Good	Good	Good	Good	Good
Image rejection (diode)	Good	Good	Good	Good	Good	Good	Good	Good
Resistive FET ring	Good	Good	Good	Good	Good	Good	Good	Good

Mixer Type	Low-Order Spurious-Response Rejection	LO Power Requirements	Third-Order IM Intercept	1-dB Compression
Single-diode	None	Low	Low	Low
Singly balanced diode (180 deg)	(2, 2): good (2, 1): good if Δ port is the LO (1, 2): good if Σ port is the LO	Moderate	Moderate	Moderate
Singly balanced diode (90 deg)	(2,2): good (2, 1): none (1, 2): none	Moderate	Moderate	Moderate
Doubly balanced diode	(2,2): good (2, 1): good (1, 2): good	High	High	High
Image rejection (diode)	Same as doubly balanced mixer	High to very high (depends on type)	High	High
Resistive FET ring	Very good	High	Very good	Very high

Table 5.1 (continued)

Mixer Type	VSWR			Port-to-Port Isolation			LO AM Noise Rej.	LO Spur. Sig. Rej.
	RF	LO	IF	RF/IF	LO/RF	LO/RF		
Single-device active FET	Depends on matching circuits	Depends on matching circuits	Very high	Depends on filters	Depends on filters	Depends on filters; usually poor	None	None
Singly balanced active FET	Depends on matching circuits	Depends on matching circuits	Very high	Good (if RF port is Δ)	Good (if LO port is Δ)	Good (if LO port is Δ)	Good	Good
Single-device dual-gate FET	Depends on matching circuits	Depends on matching circuits; usually poor	Very high	Depends on filters	Good	Depends on filters; usually poor	None	None
Gilbert-cell FET	Depends on matching circuits	Depends on matching circuits	High	Good	Good	Good	Good	Good

Mixer Type	Low-Order Spurious-Response Rejection	LO Power Requirements	Third-Order IM Intercept	1-dB Compression
Single-device active FET	None	Low	Low	Low
Singly balanced active FET	(2, 2): good (2, 1): good if Δ port is the RF (1, 2): good if Σ port is RF	Moderate	Moderate	Moderate
Single-device dual-gate FET	None	Low	Moderate	Low
Gilbert-cell FET	(2., 2): good (2, 1): good (1, 2): good	High	High	High

bias circuits should be isolated by high-value RF chokes, and so forth. Real elements then can be substituted one by one for the ideal ones. Trying to begin with the analysis of a complete design usually is unsuccessful because the large number of circuit elements makes it difficult to debug a failed analysis.

References

[1] N. Camilleri, J. Costa, D. Lovelace, and D. Ngo, "Silicon MOSFETs, the Microwave Device Technology for the 90's," *1993 IEEE MTT-S International Microwave Symposium Digest*, June 1993, p. 545.

[2] W. R. Curtice, "A MESFET Model for Use in the Design of GaAs ICs" *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, 1980, p. 448.

[3] I. Angelov, H. Zirath, and N. Rorsman, "Validation of a Nonlinear HEMT Model by Power Spectrum Characteristics," *1994 IEEE MTT-S International Microwave Symposium Digest*, June 1994, p. 1571.

[4] C. Y. Ho, "Design of Wideband Quadrature Couplers for UHF/VHF: Part I," *RF Design*, Nov. 1989, p. 58.

[5] S. A. Maas, *Microwave Mixers* (Second ed.), Norwood, MA: Artech House, 1993.

[6] D. Neuf, "M/W Mixer Using 4-Wire Transmission Line," *Microwave J.*, May, 1974, p. 48.

[7] M. G. Ellis, *Electronic Filter Analysis and Synthesis*, Norwood, MA: Artech House, 1994.

[8] C. Tsironis, R. Meierer, and R. Stahlman, "Dual-Gate MESFET Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, March 1984, p. 248.

[9] B. Gilbert, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid-State Circuits*, Vol. SC-3, Dec. 1968, p. 365.

CHAPTER 6

RF AND MICROWAVE FREQUENCY SYNTHESIZER DESIGN

Lawrence Larson and Steven Rosenbaum
Hughes Research Laboratories
Malibu, CA

6.1 INTRODUCTION

Frequency synthesizers are ubiquitous building blocks in wireless communications systems, since they provide the precise reference frequencies for modulation and demodulation of baseband signals up to the transmit and/or receive frequencies. Because of their crucial nature in the implementation of these systems, considerable effort has been expended over the years in an attempt to improve their performance and reduce their cost, to the point where they are now standard cells in completely integrated implementations.

A simple frequency synthesizer might consist of a transistor oscillator operating at a single frequency determined by a precise crystal circuit. Tunable transistor frequency sources rely on variations in the characteristics of a resonant circuit to set the frequency. These circuits can then be embedded in PLLs to broaden their range of operation and further enhance their performance. More recently, direct digital synthesis (DDS) has been developed as a technique for synthesizing a broad range of frequencies and waveforms using digital techniques. All of these different approaches will be covered in this chapter.

A representative view of a frequency synthesizer is illustrated in Figure 6.1, which shows a generic synthesizer producing a single tone of a given amplitude that has a delta-function-like characteristic in the frequency domain. However, the frequency characteristics of a realistic signal source will have a finite width in the frequency domain, as well as significant harmonic content at frequencies higher than the fundamental tone. The results of these imperfections in the frequency domain can result in degraded overall system performance, as the following simple example will illustrate.

A wireless communication system will usually consist of an information source, which is modulated up to RF/microwave frequencies and then transmitted. A receiver will take the modulated signal from the antenna, demodulate it, and send it to an information "sink," as illustrated in Figure 6.2. The rate at which information can be sent over the channel is determined by the available bandwidth (which is strictly regulated by national and international regulatory agencies), the modulation scheme, and by the integrity of the modulation/demodulation process.

This latter effect can be heavily influenced by the quality of the frequency synthesis, the noise floor of the synthesized signal, and the absence of distortion and

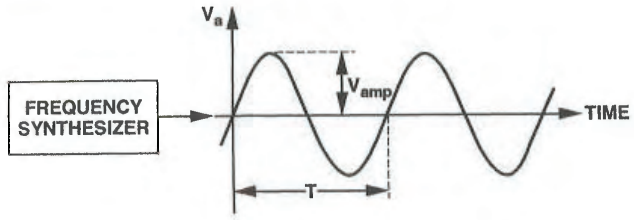


Figure 6.1 Block diagram of frequency synthesizer producing single-tone sinusoidal output.

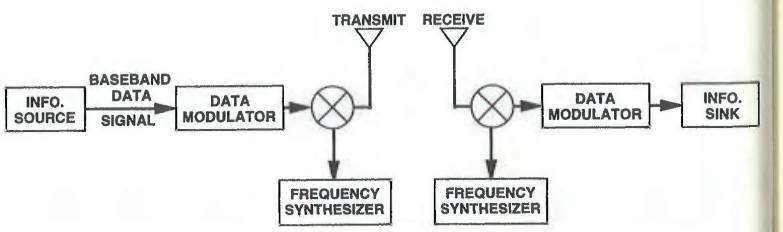


Figure 6.2 Block diagram of communications system, showing modulation and demodulation.

other spurious responses. For example, the bit error probability in a typical phase-shift keyed (PSK) system is approximately

$$P_e \approx \frac{1}{2}\text{erfc}(\sqrt{S/N}) \tag{6.1}$$

where S is the signal power and N is the noise power in the frequency band of interest. The noise power is heavily influenced by integrity of the reference oscillator used in the transmitter and receiver.

The transmitted signal power (S) is usually fixed by the system. For example, a satellite communications system typically has a fixed transponder power limit. Thus, the most straightforward way to improve the bit error rate (BER) of the system is to reduce the noise power (N) in the band of interest. This may require improved noise performance of the frequency synthesizer portion of the receiver.

This chapter will cover the essential elements of frequency synthesizer design, utilizing both analog and digital techniques. The emphasis will be on practical design considerations, rather than theoretical depth. Where appropriate, we will refer the reader to other publications that cover the material more thoroughly.

6.2 FREQUENCY SYNTHESIZER FUNDAMENTALS

The multiplicity of frequency synthesizers makes straightforward classification difficult, but they can be generally categorized into direct and indirect architectures. Direct frequency synthesizer architectures utilize no feedback to generate the appropriate output frequencies. These direct synthesizers can be further subdivided into digital and analog approaches. Analog approaches utilize fixed precise frequency sources, combined with frequency multipliers, dividers, switches, filters, and mixers to generate the necessary output frequencies. A schematic of this generic approach is given in Figure 6.3. Direct analog synthesis is very straightforward conceptually, but has a number of drawbacks if multiple frequency and/or fine frequency control are required.

In particular, the physical size of the components required to implement the multipliers, dividers, mixers and filters can be daunting. Furthermore, these synthesizers can generate a menagerie of spurious responses, which must be carefully filtered if a clean output is required. One advantage of the direct synthesizer approach is that the fundamental waveform can be generated using an ultraclean source—for example, generated by a crystal source with a high Q—and the phase noise characteristics of the resulting higher frequency waveform can be excellent. This will be discussed in more detail in Section 6.3.

By contrast, direct *digital* frequency synthesizers (DDSs) rely on digital components to generate a variety of frequency and waveforms from a single system reference

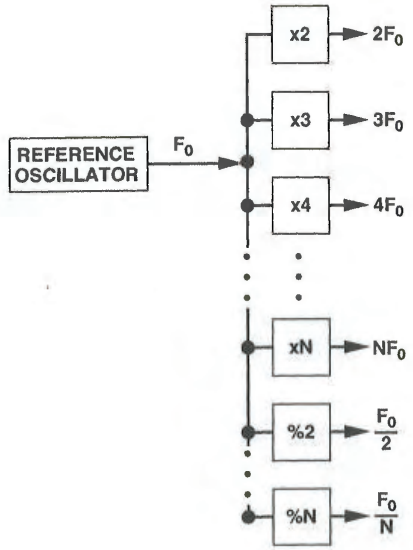


Figure 6.3 Direct analog frequency synthesizer utilizing a reference oscillator, along with frequency multipliers and dividers.

clock. The advantage of this approach is the precision available from an all-digital approach due to the lack of tuning required, as well as absence of drift and scalability with constantly improving digital IC technology. A block diagram of a typical direct digital frequency synthesizer is shown in Figure 6.4. There are four significant blocks

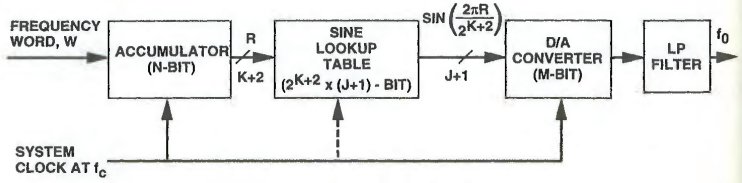


Figure 6.4 Direct digital synthesizer (DDS), utilizing an accumulator, sine lookup table, DAC, and smoothing filter.

for a typical DDS: a phase accumulator, sine lookup table (these first two together are often referred to as a numerically controlled oscillator (NCO)), digital-to-analog converter (DAC), and antialiasing filter.

By contrast, indirect frequency synthesizers rely on feedback, usually in the form of a PLL, to synthesize the frequency. A block diagram of a representative PLL frequency synthesizer is shown in Figure 6.5. Most PLLs contain three basic building blocks: a phase detector, an amplifier/loop filter, and a voltage-controlled oscillator (VCO). During operation, the loop will acquire (or lock) onto an input signal, track it, and exhibit a fixed phase relationship with respect to the input. The output frequency of the loop can be varied by altering the division ratio (N) within the loop, or by tuning the input frequency with an input frequency divider (Q). As such, the PLL can act as a broadband frequency synthesizer. Details of the design of the PLL will be developed in Section 6.4.

As with direct frequency synthesis techniques, indirect techniques can be implemented in both digital and analog forms. Purely digital PLLs are often implemented for clock generation on digital VLSI chips.

6.2.1 Frequency Synthesizer Figures of Merit

An ideal frequency synthesizer will produce a perfectly pure sinusoidal signal, which is tunable over some specified bandwidth. The amplitude, phase, and frequency of the source would not change under varying loading, bias, or temperature conditions. Of course, such an ideal circuit is impossible to realize in practice, and a variety of performance measures have been defined over the years to characterize the deviation from the ideal.

Noise

The output power of the synthesizer is not concentrated exclusively at the carrier frequency alone. Instead, it is distributed around it, and the spectral distribution on either side of the carrier is known as spectral sidebands. This is illustrated schematically in Figure 6.6. This noise can be represented as modulation of the carrier signal,

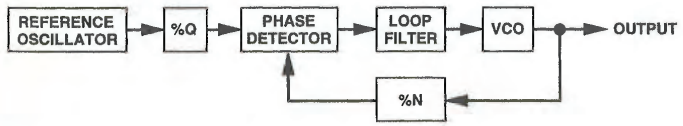


Figure 6.5 Indirect frequency synthesizer using a phase-locked loop.

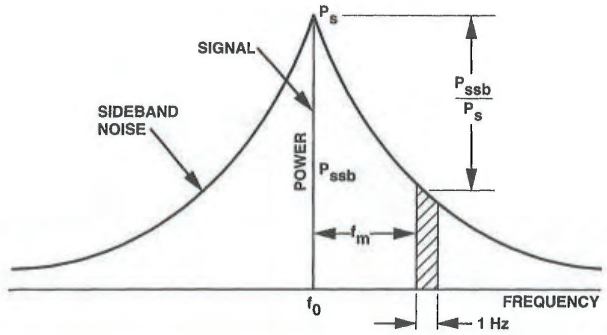


Figure 6.6 Phase noise specification of frequency source. The noise is contained in the sidebands around the signal frequency at f_0.

and resolved into amplitude modulation (AM) and frequency modulation (FM) components. The AM portion of the signal is typically smaller than the FM portion.

FM noise power is represented as a ratio of the power in some specified bandwidth (usually 1 Hz) in one sideband to the power in the carrier signal itself. These ratios are usually specified in the quantity of dBc/Hz at some frequency offset from the carrier. The entire noise power can be integrated over a specified bandwidth to realize a total angular *error* in the output of the oscillator, and oscillators are often specified this way. The calculation of the angular error from the FM noise performance is relatively straightforward, and a simple technique for accomplishing this was detailed in Chapter 2.

Frequency Pulling

Frequency pulling is the change in output frequency of an oscillator when the phase of the load VSWR is rotated by 360 deg. The allowed frequency range determines the amount of isolation required at the output of the oscillator. A load VSWR of 1.5:1 is typically used for test purposes.

A simplified expression for the total frequency deviation from the ideal can be given by [1]

$$\Delta\omega_T \cong \frac{\omega_0}{2Q_{ext}}(\text{VSWR} - \text{VSWR}^{-1}) \tag{6.2}$$

where Q_{ext} is the quality factor of the oscillator, and ω_0 is the radian frequency of oscillation. This expression demonstrates the importance of high Q in minimizing frequency pulling in oscillator design.

Power Output and Stability

The power output and stability of the oscillator are usually specified as a power output range in dBm at a given frequency and over a temperature range, say from 0°C to 60°C.

Tuning Range

The tuning range of an oscillator specifies the variation in output frequency with input voltage or current (usually voltage). The slope of this variation is usually expressed in MHz/V. In particular, the key requirements of oscillator or synthesizer tuning are that the slope of the frequency variation remain relatively consistent over the entire range of tuning and that the total frequency variation achieves some minimum specified value.

Frequency Stability

Frequency stability of an oscillator is typically specified in terms of parts per million per degree centigrade (PPM/°C) or ±PPM/°C. This parameter is related to the Q of the resonator and the frequency variation of the resonator with temperature. In a free-running system, this parameter is particularly important, whereas in a PLL it is less so since an oscillator that drifts may be locked to a more stable oscillator source.

Frequency Pushing

Frequency pushing is the effect on the output frequency due to changes in the dc power supply voltage, and is typically specified in terms of percent change in output frequency per volt. Values of less than 0.001 percent per volt are not uncommon for high Q oscillators.

Harmonics

Harmonics are outputs from the oscillator/synthesizer that occur at integral multiples of the fundamental frequencies. They are typically caused by nonlinearities on the

transistor or other active device used to produce the signal. They can be minimized by proper biasing of the active device and design of the output matching network to filter out the harmonics. Harmonics are typically specified as a dBc below the carrier.

Spurious Outputs

Spurious outputs are outputs of the oscillator/synthesizer that are not necessarily harmonically related to the fundamental output signal. As with harmonics, they are typically specified as a dBc below the carrier.

Spurious responses are a major limitation in DDS, since the output of the DDS is a essentially a periodic repetition of an approximation to the ideal waveform. This effect, combined with aliasing and nonidealities in the DAC can create a significant spurious output. This will be covered in more detail in Section 6.5.

6.3 VOLTAGE-CONTROLLED OSCILLATORS

6.3.1 Introduction

Voltage-controlled oscillators are key components of all analog communications systems. They are essentially dc-to-RF converters; they produce an RF signal output with a only a dc signal input. An oscillator consists of two chief components: an active device that acts as an amplifier and a feedback network to provide positive feedback in the system. The feedback network is frequency-sensitive and includes some type of resonator to set the operating frequency. In addition, some type of variable reactance element must be present for control of the frequency. Normally, the variable reactance is controlled by a dc voltage, hence the term *voltage-controlled oscillator*. The typical design emphasis is on low noise, high efficiency, temperature stability, bandwidth, linear and wideband tunability, low cost, and reliability.

6.3.2 General Oscillator Design

A VCO acts as a sinusoidal frequency generator with a specified frequency versus voltage behavior. The types of oscillators under discussion here employ a positive feedback loop consisting of an amplifier and a frequency-selective feedback network. Such oscillators are known as *linear oscillators*. Although the oscillation amplitude is limited due to nonlinear effects in the amplifying device, the term linear oscillator is used to denote that the onset of oscillation is due to the embedded linear amplifier and to distinguish this from those oscillators that generate square, triangular, and other shaped waveforms (i.e., *nonlinear oscillators*).

A general block diagram of a linear oscillator is shown in Figure 6.7. Using standard notation for feedback amplifiers (of which this is a special case) the following characteristic equations describe the oscillator [2]:

$$x_o = Ax_i \tag{6.3}$$

$$x_f = \beta x_o \tag{6.4}$$

$$x_i = x_s + x_f \tag{6.5}$$

$$A_f \equiv \frac{x_o}{x_s} \tag{6.6}$$

Combining (6.3) through (6.5) with (6.6), we derive the well-known expression for the closed-loop gain

$$A_f = \frac{A}{1 - A\beta} \tag{6.7}$$

The *loop gain* is defined as

$$L(s) \equiv A(s)\beta(s) \tag{6.8}$$

where $s = j\omega$. If, at some frequency, ω_o, the loop gain is unity, then the closed-loop gain of the system will be infinite. In other words, there will be a finite output signal with zero input signal (implying oscillation). So, for a linear oscillator at some frequency ω_o, the magnitude of the loop gain is equal to one, the phase of the loop gain is equal to zero, and the system will oscillate at this frequency.

A generalized oscillator circuit is shown in Figure 6.8, where Q is a three-terminal amplifying device (transistor) in some configuration (common-base, -collec-

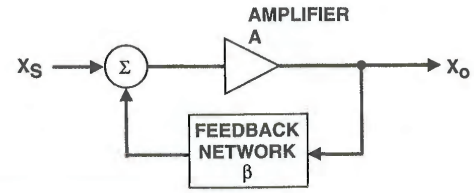


Figure 6.7 General block diagram of a linear oscillator.

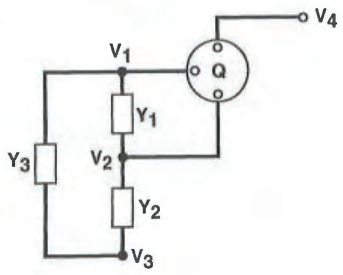


Figure 6.8 Generalized oscillator circuit.

tor, or -emitter, or common-gate, -drain, or -source). The transistor can be replaced with a voltage-controlled current source as in Figure 6.9. This configuration represents a generalization of the most commonly used circuit types: common-base, common-gate, common-emitter, and common-source. A slight variation in the configuration of the transistor model will enable analysis of the common-collector and common-drain circuit types. The circuit is shown without feedback; there must be some interconnection of nodes in order for it to oscillate. The exact feedback necessary to achieve oscillation will be derived next.

The phase of the feedback network necessary to cause oscillation is dependent on the phase shift through the amplifying device. For example, in the case of a common-emitter (or common-source) configuration, the phase of the feedback network must be 180 deg since the phase shift through the transistor is 180 deg, which yields a total loop phase shift of 0 deg. However, in the case of a common-base (or common-gate) configuration, the phase of the feedback network must be 0 deg since there is no phase shift through the transistor in that configuration.

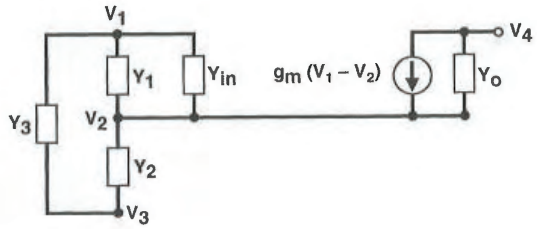


Figure 6.9 Generalized oscillator circuit with Q replaced by voltage-controlled current source.

Writing the nodal equations of the circuit of Figure 6.9 results in the following set of homogeneous equations. If a solution exists other than the trivial solution (i.e., all currents equal zero), then currents arise in the circuit with no forcing voltage, which implies the existence of oscillation. Therefore, as a condition of oscillation:

$$\begin{bmatrix} Y_1 + Y_3 + Y_{in} & -(Y_1 + Y_{in}) & -Y_3 & 0 \\ -(Y_1 + Y_{in} + g_m) & Y_1 + Y_2 + Y_{in} + Y_o + g_m & -Y_2 & -Y_o \\ -Y_3 & -Y_2 & Y_2 + Y_3 & 0 \\ g_m & -(Y_o + g_m) & 0 & Y_o \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = 0 \tag{6.9}$$

Equation (6.9) is the most general formulation of the oscillator circuit. For specific implementations, the oscillator matrix can be simplified. The rules for simplifying (6.9) are

1. If a node is connected to ground, then the corresponding row and column of that node are eliminated. For example, if node V_2 was connected to ground, row 2 and column 2 from the above matrix would be eliminated. We would also eliminate V_2 from the voltage matrix.
2. If two nodes are connected together, then the corresponding rows and columns are added together. For example, if node V_3 was connected to V_4, then rows 3 and 4 and columns 3 and 4 would be added together. V_3 and V_4 would be combined into one V in the voltage matrix.

For any system of homogeneous equations, the determinant of the system must be equal to zero to assure the existence of nontrivial solutions. Once (6.9) has been simplified based on the implementation desired, one need merely set the determinant equal to zero to find the conditions for oscillation.

6.3.3 Colpitts and Hartley Oscillators

A number of different configurations may be used to realize oscillator circuits; two of the most common are the *Colpitts* and *Hartley oscillators*. These oscillators may use a bipolar transistor (BJT) or a field-effect transistor (FET) as the active device. The common node may be any of the three transistor terminals or another circuit node. We will consider two example configurations below.

Example: Bipolar Transistor Common-Emitter Oscillator

An example of a bipolar transistor oscillator in the common-emitter configuration proceeds next. Since we are considering a common-emitter configuration, ground is placed at V_2 of Figure 6.9.

For a bipolar transistor, it is useful to make the following simplifying assumptions for the ideal case:

1. The output admittance Y_o is equal to 0.
2. The input admittance is purely real ($Y_{in} = G_{in}$).

For the common-emitter case, the feedback must be between the collector and the base, which will be achieved by connecting V_4 to V_3. As a result, the phase shift through the feedback network (collector-to-base) must be 180 deg, as required by the common-emitter configuration. By using the rules outlined in the previous section and setting the determinant equal to zero we have

$$\begin{vmatrix} Y_1 + Y_3 + G_{in} & -Y_3 \\ -Y_3 + g_m & Y_2 + Y_3 \end{vmatrix} = 0 \tag{6.10}$$

Y_1, Y_2, and Y_3 are typically purely imaginary (inductors or capacitors), so that

$$Y_n = jB_n \tag{6.11}$$

where $n = 1, 2, 3$. Also, the transconductance g_m is typically purely real. Since both the real and imaginary parts of the determinant must individually be equal to zero, the real part reduces to

$$B_1B_2 + B_2B_3 + B_1B_3 = 0 \tag{6.12}$$

by dividing both sides of the equation by $B_1B_2B_3$ and using the relationship $X_n = 1/B_n$,

$$X_1 + X_2 + X_3 = 0 \tag{6.13}$$

The imaginary part of the determinant reduces to

$$B_2 + \left(1 + \frac{g_m}{G_{in}}\right)B_3 = 0 \tag{6.14}$$

which becomes

$$B_2 + (1 + \beta)B_3 = 0 \tag{6.15}$$

for a bipolar transistor (since $\beta = g_m/G_{in}$). By dividing both sides of (6.15) by B_2B_3, combining with (6.13), and again using the relationship $X_n = 1/B_n$, (6.15) further simplifies to

$$X_1/X_2 = \beta \tag{6.16}$$

Since β is real and positive, X_1 and X_2 must be of the same sign. Therefore, (6.13) shows that X_3 must be of opposite sign. In the case where X_1 and X_2 are negative (capacitors) and X_3 is positive (inductor), the configuration is called a *Colpitts oscillator*. In the case where X_1 and X_2 are positive (inductors) and X_3 is negative (capacitor), the configuration is called a *Hartley oscillator*. In fact, Colpitts and Hartley designations describe any possible configuration regardless of where ground in the circuit has been placed.

For a Colpitts oscillator, we can substitute the actual impedances at some frequency ω_o due to the capacitances C_1 and C_2 as well as that due to the inductance L into (6.13) to find

$$\omega_o = 1/\sqrt{L\left(\frac{C_1C_2}{C_1 + C_2}\right)} \tag{6.17}$$

and by substitution into (6.16)

$$\beta = C_2/C_1 \tag{6.18}$$

If a Hartley configuration was employed for this oscillator, (6.13) would have yielded

$$\omega_o = 1/\sqrt{C(L_1 + L_2)} \tag{6.19}$$

and (6.16) would have reduced to

$$\beta = L_1/L_2 \tag{6.20}$$

Equations (6.18) and (6.20) seem to indicate that $C_2 \gg C_1$ and $L_1 \gg L_2$ since β is large for transistors. However, at high frequencies, β rolls off and it is common to find that $C_1 = C_2$ and $L_1 = L_2$. The final common-emitter bipolar Hartley oscillator is shown in Figure 6.10 with a bias network included [2].

Example: FET Common-Gate Oscillator

FET oscillators are often operated in the common-gate configuration, so ground is placed at V_1 of Figure 6.9. For a FET, it is useful to make the following simplifying assumptions for the ideal case:

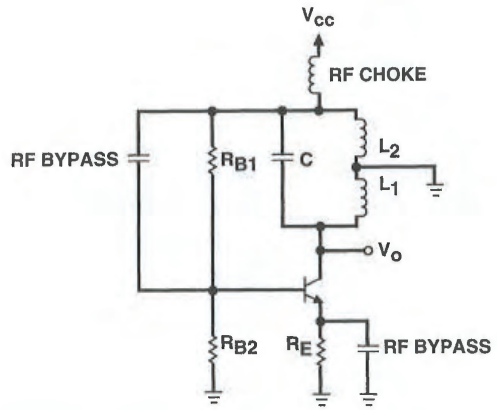


Figure 6.10 Common-emitter Hartley oscillator with bias network included.

1. The input admittance G_{in} is equal to 0.
2. The output admittance is purely real ($Y_o = G_o$).

Also, for a common-gate design, the feedback must be between the drain and the source, which will be achieved by connecting V_4 to V_3. One may also readily see that the phase shift through the feedback network (drain-to-source) is 0 deg, as required by the common-gate configuration. By using the rules outlined in the previous section, the determinant of (6.9) reduces to

$$\begin{vmatrix} Y_1 + Y_2 + G_o + g_m & -(Y_2 + G_o) \\ -(Y_2 + G_o + g_m) & Y_2 + Y_3 + G_o \end{vmatrix} = 0 \tag{6.21}$$

In solving (6.21), Y_1, Y_2, and Y_2, are once again restricted to be purely imaginary. By setting the real part to zero and using the reciprocal relationship between B_n and X_n, the real part of (6.21) yields

$$X_1 + X_2 + X_3 = 0 \tag{6.22}$$

which leads to (6.17) and (6.19) regarding the value of the frequency of oscillation ω_o. However, the imaginary part of the solution results in a slightly different equation from (6.18) and (6.20), due to the different characteristics of the FET, i.e.,

$$X_2/X_1 = g_m/G_o \tag{6.23}$$

or, in the case of a Colpitts oscillator

$$C_1/C_2 = g_m/G_o \tag{6.24}$$

and for the Hartley oscillator

$$L_2/L_1 = g_m/G_o \tag{6.25}$$

As in the bipolar case, (6.24) and (6.25) indicate that $C_1 \gg C_2$ and $L_2 \gg L_1$ since g_m/G_0 is large for transistors. However, at high frequencies, g_m/G_0 rolls off and it is common to find that $C_1 = C_2$ and $L_1 = L_2$. The final common-gate FET Colpitts oscillator is shown in Figure 6.11 with a bias network included.

6.3.4 Clapp Oscillator

An additional configuration, called the *Clapp oscillator* [3] (shown in Figure 6.12), is very similar to the Colpitts in design and operation, except for the extra capacitor in series with the inductor. At the intended operating frequency, the series capacitor-inductor pair will present the same inductive reactance as the Colpitts, but if the frequency were to drift, that equivalent inductive reactance would change rapidly (much faster than a single inductor.) Figure 6.13 shows the resonant tank circuit for a Colpitts and a Clapp oscillator. At ω_o, the total series reactances are equal in the two tanks. At a frequency slightly offset from ω_o, however, the Clapp oscillator

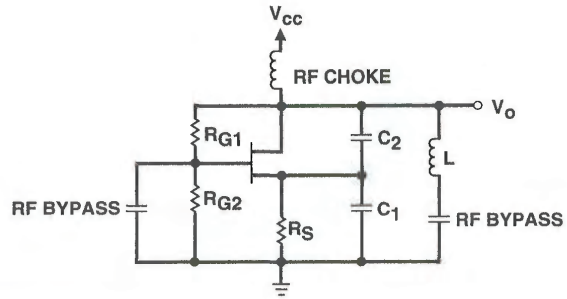


Figure 6.11 Common-gate Colpitts oscillator with bias network included.

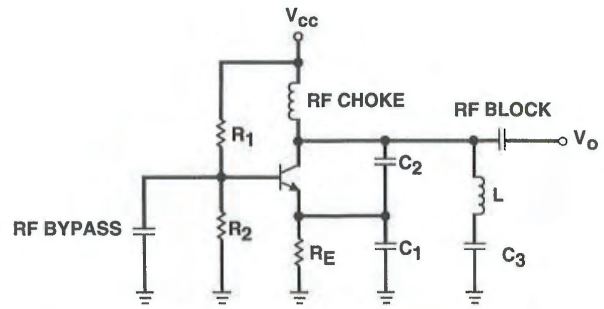


Figure 6.12 Common-base Clapp oscillator with bias network included.

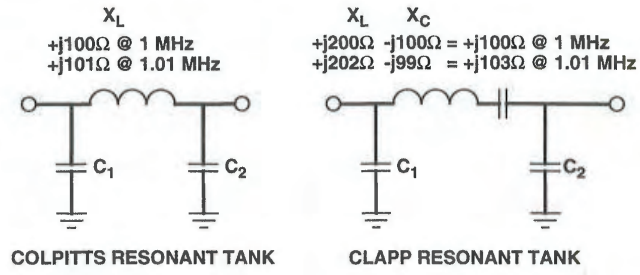


Figure 6.13 Resonant tank circuit for a Colpitts and a Clapp oscillator [3].

tank series reactance changes more than the Colpitts series reactance. As a result, the Clapp oscillator is more stable with frequency. It should be noted, however, that the operation is dependent upon a stable inductor. If the inductor drifts with temperature, the Clapp oscillator will drift at a *faster* rate than the Colpitts. So, the Clapp should only be used where it is possible to construct high-quality inductors and keep them at a constant temperature.

6.3.5 Crystal Oscillators

Additional frequency stability may be obtained by the use of a quartz crystal in the oscillator circuit. The quartz crystal is a high-Q resonator with unloaded Qs up to 100,000 and temperature drift of less than 0.001%/°C. The crystal has an equivalent

circuit as shown in Figure 6.14. It should be apparent that the crystal has both a series and parallel resonant mode, which are near each other in frequency. The impedance of the crystal between resonances is inductive, with a rapid rate of change of impedance with frequency, which stabilizes the oscillator frequency because any significant change in operating frequency will cause a large change in the feedback loop phase shift to prevent off-frequency oscillations [4].

Crystals may be used as either parallel or series resonant circuits in oscillators [5]. The series resonant frequency of the crystal occurs when L_1 and C_1 are in resonance, which corresponds to the frequency

$$\omega_s = 1/\sqrt{L_1 C_1} \tag{6.26}$$

ω_s is referred to as the *series resonant frequency*. The crystal impedance will be infinite at the frequency

$$\omega_a = 1/\sqrt{L_1\left(\frac{C_o C_1}{C_o + C_1}\right)} \tag{6.27}$$

ω_a is referred to as the *antiresonant frequency* of the crystal.

Oscillators using crystal resonators are frequently designed in a parallel-mode configuration, where the crystal is used in the antiresonant mode. In this mode, the crystal actually serves as an inductor. A commonly used crystal oscillator circuit, called a *Pierce oscillator* (based on the common-emitter Colpitts oscillator), is shown in Figure 6.15. The crystal is used in place of the inductor in the regular Colpitts oscillator.

The main drawback to the crystal oscillator is its relatively small tuning range. One can introduce tunability into the crystal oscillator circuit by adding a capacitor in parallel with the crystal. In this mode, the antiresonant frequency can be decreased until it is equal to the series resonant frequency. The resultant so-called *pulling range* of the crystal oscillator is expressed by the difference between the antiresonance and series resonance of the crystal and is equal to

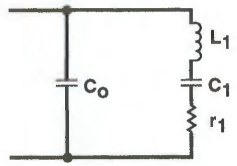


Figure 6.14 Equivalent circuit representation of crystal.

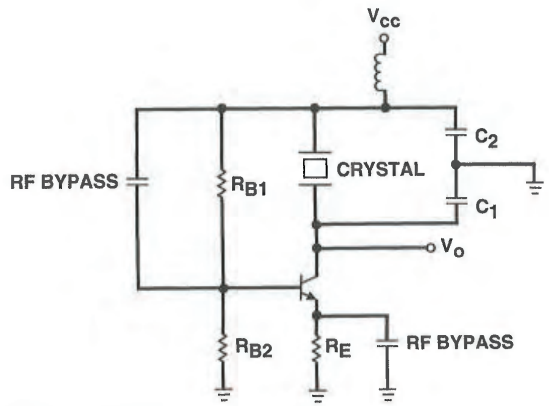


Figure 6.15 Pierce oscillator with bias network included.

$$f_{pull} = f_a - f_s = \frac{C_1}{2C_o} f_s \tag{6.28}$$

6.3.6 Oscillator Configurations for Microwave Frequencies

A more general set of oscillator circuit configurations are often used for high-microwave frequencies, and they are shown in Figure 6.16 [6]. The first three are resonant circuits with series feedback. The second three are resonant circuits with shunt feedback. Typically, the ground node is at the load resistor. The element values in each circuit can be found in terms of the Z- or Y-parameters of the transistor by use of the equations below each circuit in the figure.

A is defined as the complex voltage ratio between the output voltage and the input voltage of the active device in the oscillator. The optimal value of A for maximum power transfer out of the transistor has been determined to be [7]

$$A = -\frac{y_{21} - y_{12}^*}{2Re(y_{22})} \tag{6.29}$$

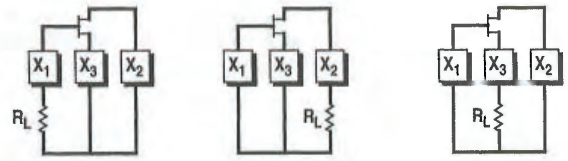
by setting the magnitude of the voltage at the input of the transistor to be a constant and optimizing for the highest power flowing out of the device. The value of A has been further optimized by determining the value that would give the maximum two-



G_L	$C_1+A_rC_3+A_iC_4$	$(C_1+A_rC_3+A_iC_4)/\|A\|^2$	$(C_1+A_rC_3+A_iC_4)/\|1+A\|^2$
B_1	$C_2+(1-A_r)(C_4+C_3A_r/A_i)$	$C_1(A_r-1)/A_i+C_2$	$(C_1+C_3)A_r/A_i+C_2+C_4$
B_2	$C_3(A_r-1)/A_i+C_4$	$[C_1(A_r-\|A\|^2)/A_i-A_iC_3+A_rC_4]/\|A\|^2$	$-(C_1+C_3)/A_i$
B_3	$-C_3A_r/A_i-C_4$	C_1/A_i	$[(A_r-1)(C_4+(C_1+C_3)A_r/A_i)+A_iC_1]/\|1+A\|^2$

WHERE $C_1 = -Re[Y_{11}+AY_{12}]$ $C_3 = -Re[Y_{21}+AY_{22}]$ $A = A_r+jA_i$

$C_2 = -Im[Y_{11}+AY_{12}]$ $C_4 = -Im[Y_{21}+AY_{22}]$



R_L	$D_1+F_rD_3+F_iD_4$	$(D_1+F_rD_3+F_iD_4)/\|F\|^2$	$(D_1+F_rD_3+F_iD_4)/\|1+F\|^2$
X_1	$D_2-(1-F_r)(D_4+D_3F_r/F_i)$	$D_1(1+F_r)/F_i+D_2$	$(D_1-D_3)F_r/F_i+D_2-D_4$
X_2	$-D_3(1+F_r)/F_i-D_4$	$[D_1(F_r+\|F\|^2)/F_i-F_iD_3+F_rD_4]/\|F\|^2$	$(D_1-D_3)/F_i$
X_3	D_3F_r/F_i+D_4	$-D_1/F_i$	$[(1+F_r)D_4-(D_1-D_3)F_r/F_i)-F_iD_1]/\|1+F\|^2$

WHERE $D_1 = -Re[Z_{11}+FZ_{12}]$ $D_3 = -Re[Z_{21}+FZ_{22}]$ $F = F_r+jF_i$

$D_2 = -Im[Z_{11}+FZ_{12}]$ $D_4 = -Im[Z_{21}+FZ_{22}]$ $= (Z_{21}-AZ_{11})/(AZ_{12}-Z_{22})$

Figure 6.16 Oscillator circuit configurations for high-microwave frequencies [6].

port added power for the oscillating system. That value, written here as A_{opt}, is given by [8]

$$A_{opt} = \frac{-2g_{11}y_{21}}{\sqrt{N} + 2g_{11}g_{22} + j(g_{12}b_{21} + g_{21}b_{12})} \tag{6.30}$$

where

$$N = 4g_{11}g_{22}(g_{11}g_{22} + b_{12}b_{21} - g_{12}g_{21}) - (g_{12}b_{21} + g_{21}b_{12})^2 \tag{6.31}$$

A_{opt} differs from A in that its determination is independent of the input voltage to the transistor, whereas the determination of A assumes a constant input voltage magnitude, which is unlikely.

In general, it can be shown that a larger transistor must be used to achieve higher power outputs, due to its higher current handling capabilities. This must be balanced, however, by the resultant increase in transistor capacitances and conductances, which may begin to make the realization of the correct matching networks difficult at higher frequencies. A tradeoff may be made, however, by designing a lower power oscillator and following it with a higher power amplifier to act as a power boosting stage.

6.3.7 Dielectric Resonator Oscillators (DROs)

Dielectric resonators (DRs) are made of low-loss, temperature-stable, high-permittivity, and high-Q ceramic materials. They resonate in various modes determined by their material composition and their dimensions and shielding and are used in oscillators that employ transmission lines rather than lumped element components. In practice, the DR is placed near a microstrip transmission line so as to enable magnetic coupling between the two structures, as is shown in Figure 6.17 [9]. The lateral distance between the resonator and the microstrip conductor is the parameter that determines the coupling between the DR and the line. The coupling is accomplished by orienting the magnetic moment of the resonator perpendicular to the microstrip plane so that the magnetic lines of the resonator link with those of the microstrip line, as shown in the figure. The DR acts as a cavity, which reflects the RF energy at the resonant frequency. The resonant frequency of the DR is given by

$$f_r(\text{GHz}) = \frac{34}{a\epsilon_r}\left(\frac{a}{H} + 3.45\right) \tag{6.32}$$

where a (in millimeters) is the radius of the resonator and H is its height. Equation (6.32) is valid to about 2% accuracy in the range

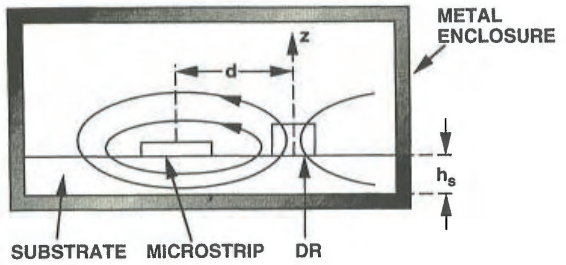


Figure 6.17 Coupling between a microstrip line and a dielectric resonator [9].

$$0.5 < \frac{a}{H} < 2 \tag{6.33}$$

$$30 < \epsilon_r < 50 \tag{6.34}$$

The equivalent circuit of the DR coupled to the microstrip line is shown in Figure 6.18. Typically, the manufacturer supplies the equivalent-circuit values of the DR. The resonant frequency and the unloaded Q of the resonator in terms of the equivalent circuit parameters are given by

$$\omega_o = 1/\sqrt{L_rC_r} \tag{6.35}$$

$$Q_u = \frac{\omega_o L_r}{R_r} \tag{6.36}$$

The circuit may be further simplified to the one shown in Figure 6.19, where

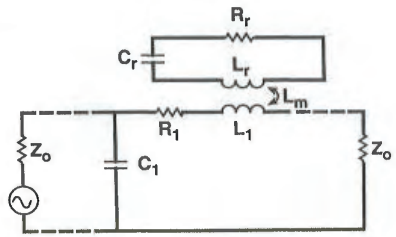


Figure 6.18 Equivalent circuit of a dielectric resonator coupled to a microstrip line [9].

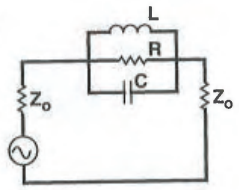


Figure 6.19 Simplified equivalent circuit of dielectric resonator coupled to a microstrip line [9].

$$L = \frac{L_m^2}{L_r} \tag{6.37}$$

$$C = \frac{L_r}{\omega_0^2 L_m^2} \tag{6.38}$$

$$R = \omega_0 Q_u \frac{L_m^2}{L_r} \tag{6.39}$$

The coupling coefficient γ at the resonant frequency is defined by

$$\gamma = \frac{R}{R_{ext}} \tag{6.40}$$

where R_{ext} is the total resistance seen on both sides of the DR.

A FET common-source DRO is shown in Figure 6.20 [10]. The value X_3 of the series feedback element may be determined by the equations in Figure 6.16. The transmission line is terminated in its characteristic impedance. This transmission

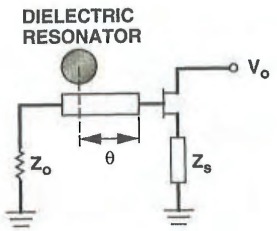


Figure 6.20 FET common-source DRO [10].

structure has a bandstop characteristic so the gate will see a reflection coefficient equal to unity at the resonant frequency and the characteristic impedance (reflection coefficient equal to zero) at all other frequencies. The transmission structure must present the correct impedance X_1 to the gate at resonance, which may be achieved by adjusting the phase of reflection coefficient at resonance by precisely placing the DR at some point θ (in electrical length at the resonant frequency) from the gate. The parameters for the placement of the DR are determined from the following equations:

$$\Gamma_t = \frac{\gamma}{\gamma + 1 + 2jQ_u\delta}\exp(-2j\theta) \tag{6.41}$$

where Γ_t is the desired reflection coefficient seen by the transistor and δ is the fractional bandwidth given by

$$\delta = (\omega - \omega_o)/\omega_o \tag{6.42}$$

γ is the coupling coefficient between the DR and the transmission line and Q_u is the unloaded Q of the DR.

Since the general series feedback oscillator circuit diagrams are shown in terms of X rather than Γ, the following transformation is included for completeness:

$$X_3 = Z_3 = Z_o\frac{1 + \Gamma_t}{1 - \Gamma_t} \tag{6.43}$$

By using the above equations, we may determine the correct values for γ and θ to achieve the desired value of X_3.

Figure 6.21 shows another possible topology for the DRO that is similar to the previous example except the resonator is placed in the feedback path so that

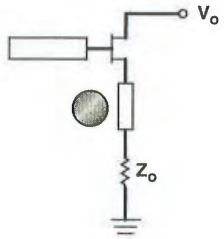


Figure 6.21 Alternate FET common-source DRO [10].

feedback is allowed only at the resonant frequency. This configuration ensures stability at all frequencies other than resonance. The circuit in Figure 6.22 uses the DR to complete a shunt feedback path between the drain and gate. In each of the above cases, as well as multiple other possible DRO configurations, the DR is used to satisfy one set of equations in Figure 6.16 to achieve the proper conditions for oscillation.

6.3.8 Practical Considerations for Oscillator Design

The previous sections have analyzed ideal oscillator cases only. In reality, there are a number of practical issues that must be considered. For example, in the preceding analyses, the elements Y_1, Y_2, and Y_3 were assumed to be purely reactive, while there will always be some finite Q associated with any capacitor or inductor, causing the admittance (or impedance) to be complex. Furthermore, the assumptions made about the bipolar transistor—that Y_{in} is purely real and Y_o is equal to zero—are only approximations. Likewise, the assumptions made about the FET are simplifications. In addition, the oscillator dependence on the output load was not considered in the above analyses, which could prove to be significant for high conductance loads. Finally, the bias networks and surrounding circuitry may impact the performance of the oscillator in complicated ways that will cause nonideal responses.

The previous assumptions need only be used to determine a starting point for the design, which may then be optimized considering the issues raised in this section. In particular, computer aided design (CAD) software may be employed to perform the computation required to include all nonideal effects in the full circuit. Most manufacturers of microwave components include full characterization data for each component that will be used in the circuit. This information, in the form of equivalent circuit models, S-parameters, or some large-signal model, can easily be incorporated into most time or frequency domain simulators for final simulation of the oscillator circuit.

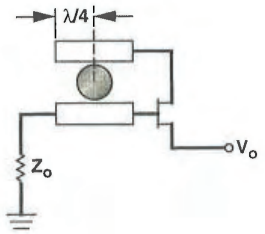


Figure 6.22 FET DRO using drain-to-gate feedback [10].

6.3.9 Voltage Control of Oscillators

In most configurations shown here, the frequency of oscillation can be varied with the addition or subtraction of capacitance in the resonant circuit. Diodes whose capacitance can be varied with application of dc voltage are called *varactor diodes* (or *varactors*) and serve to function as the controlling capacitance in oscillators. One commonly used VCO circuit technique is to shunt the inductor of a Colpitts oscillator with a varactor. The voltage applied changes the amount of capacitance in parallel with the inductor, which changes the frequency of oscillation. If the capacitor is in series with the inductor, the circuit is identical to the Clapp configuration, with the fixed capacitor replaced by the varactor. The advantage of this configuration is that a smaller capacitance can be used to change the frequency of oscillation. The transistor capacitance may also be used as the voltage-varying frequency-controlling reactance [11, 12].

A key measure of the quality of a VCO is the linearity of the voltage-frequency relationship, as will be seen in the next section on phase-locked loops. As such, it is imperative that the designer consider the voltage-capacitance relationship in the diode carefully when designing the VCO. For a resonator, ω is expressed by the relationship

$$\omega = \frac{1}{\sqrt{LC}} \tag{6.44}$$

If C is a function of voltage, as in the case of a varactor, then linear tuning with voltage is achieved when

$$C = \frac{K}{v^2} \tag{6.45}$$

where K is a constant and v is the tuning voltage. The capacitance versus voltage for varactors may be expressed by

$$C(v) = \frac{C_j(0)}{v^\gamma} \tag{6.46}$$

where

$C_j(0)$ = zero volt junction capacitance
γ = capacitance exponent

Hyperabrupt junction varactors have a capacitance exponent that is close to 2 over a wide range of tuning voltages, which make them ideal for linear voltage controlled oscillators.

6.3.10 Noise in Oscillators

Any noise sources within the active device in the oscillator will modulate the output signal such that noise sidebands are produced around the fundamental signal. These sidebands can be divided into three classes: AM noise, due to amplitude modulation by the noise sources; FM noise, due to frequency modulation by the noise sources; and PM noise due to phase modulation by the noise sources. FM and PM noise produce the same effects in the spectrum. These three noise sources produce spectra of the form shown in Figure 6.23.

The two primary types of noise in transistors are low-frequency $1/f$ noise due to trapping in the semiconductor material and thermal (white) noise. These noise sources act to modulate the fundamental oscillator signal (carrier) and produce AM and FM noise sidebands about the center frequency.

The following noise analysis is proposed by Sweet [13]. Consider the noise circuit shown in Figure 6.24. The thermal noise source is represented in the circuit by a current source at the input of the transistor, which acts as a randomly varying injected microwave signal. This signal produces both amplitude and frequency modulation. The low-frequency $1/f$ noise sources are not shown explicitly in the circuit,

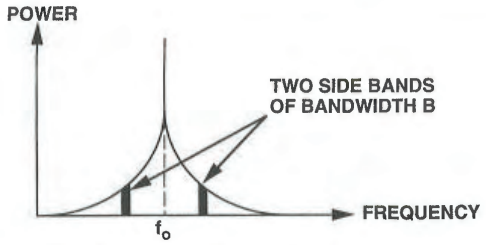


Figure 6.23 Oscillator output spectrum including effects of AM and PM noise [13].

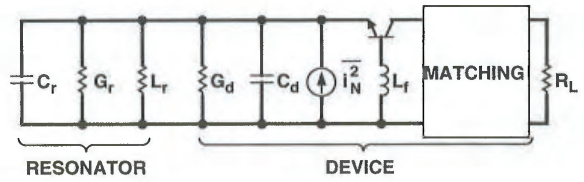


Figure 6.24 Noise equivalent circuit of transistor oscillator [13].

but they act to cause random variations in the transistor's equivalent-circuit elements, which randomly modulate the amplitude and frequency of the oscillator.

FM Noise

In analogy to diode negative resistance oscillators [14], the oscillator's mean square FM frequency deviation noise due to thermal noise of a transistor oscillator is

$$\overline{\Delta\omega^2} = \frac{\omega_o^2 k T_n B}{4Q_r^2 P_{\text{out}}} \tag{6.47}$$

where

ω_o is the oscillator's center frequency;
k is Boltzmann's constant (1.38×10^{-23} J/K);
T_n is the transistor equivalent noise temperature in K;
B is the measurement bandwidth;
Q_r is the unloaded Q of the resonator;
P_{out} is the oscillator output power.

This noise is dominant for frequencies far from the carrier frequency.

The upconverted $1/f$ noise is dominant at frequencies near the carrier and the resulting mean square FM frequency deviation is given by

$$\overline{\Delta\omega^2} = \left[\frac{\omega_o^2\left(\dfrac{\partial C_d}{\partial V_o}\right)^2}{4Q_r G_r}\right] S_{\Delta V_o}(\omega_m) B \tag{6.48}$$

where

$\left(\dfrac{\partial C_d}{\partial V_o}\right)$ is the active device's capacitance change with dc voltage;
$S_{\Delta V_o}$ is the spectrum of the low-frequency noise fluctuation in V^2/Hz;
G_r is the resonator loss in terms of conductance;
ω_m is the low frequency from where the noise fluctuations are originating.

The spectrum of low-frequency noise fluctuations obey a pattern that may be expressed by

$$S_{\Delta V_o}(f) B = \frac{N}{f^\alpha} \tag{6.49}$$

where

N is an empirically derived strength parameter;
α is a spectrum shape coefficient (usually close to unity).

Since α is usually close to unity, the spectrum has a shape proportional to $1/f$.

The FM noise mean square frequency deviation may be converted to a noise-to-carrier ratio by using small-signal FM theory:

$$\frac{N}{C} = \frac{\overline{\Delta\omega^2}}{\omega_m^2} \tag{6.50}$$

AM Noise

The AM noise-to-carrier ratio in a bandwidth B at a given frequency ω, is

$$\frac{N}{C} = \frac{kT_n BG_r^3/P_{out}\left(\frac{\partial G_d}{\partial V_1}\right)}{1 + (\omega/\omega')^2} \tag{6.51}$$

where

$$(\omega')^2 = \frac{\omega_o^2 P_{out}\left(\frac{\partial G_d}{\partial V_1}\right)}{8Q_r^2 G_r^3} \tag{6.52}$$

ω' is referred to as the *AM corner frequency*

$\left(\frac{\partial G_d}{\partial V_1}\right)$ is the device's input terminal conductance sensitivity to RF signal level, also called the *oscillator nonlinearity factor.*

The AM upconverted $1/f$ noise is given by the expression

$$\frac{N}{C} = \left(\frac{1}{4}P_{out}^2\right)\left(\frac{\partial P_{out}}{\partial V_o}\right)^2 S_{\Delta V_o}(\omega)\frac{B}{1 + (\omega/\omega')^2} \tag{6.53}$$

where $\left(\frac{\partial P_{out}}{\partial V_o}\right)$ is the oscillator's power sensitivity to changes in dc voltage.

Additional analyses have produced models of oscillator noise based on the small-signal equivalent circuit of the transistor [15] and methods of designing for lowest phase noise [16, 17]. Various techniques have been used to lower oscillator noise, including a method to suppress $1/f$ noise that would be upconverted to near the carrier [18].

Figure 6.25 shows the typical shape of the AM and FM components of oscillator noise. Figure 6.26 shows a comparison of the $1/f$ noise spectra between Si bipolar transistors and GaAs FETs. Since the bipolar $1/f$ noise is so much lower than the FET's, bipolar oscillators have a distinct advantage where upconverted close to the carrier noise is important. When noise far from the carrier is more important, the GaAs FET is preferred due to its lower thermal noise.

Oscillator noise is typically specified in terms of decibels below carrier per Hertz (dBc/Hz) at some specified offset frequency.

6.3.11 Ring Oscillator VCO Designs

One drawback to the designs presented in the previous sections is the difficulty in realizing a monolithic resonator capable of sufficiently high Q to realize an acceptable phase noise. Some GaAs monolithic circuits are capable of realizing reasonably high-

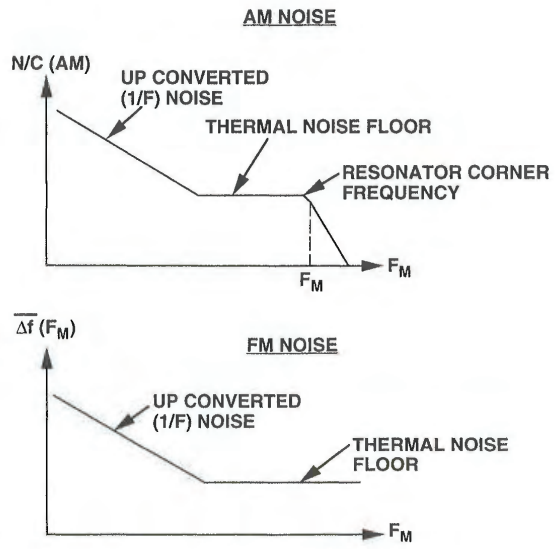


Figure 6.25 AM and FM oscillator noise spectra [13].

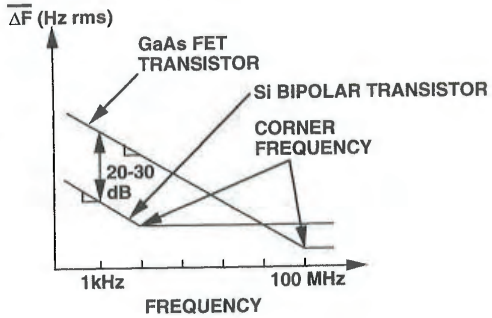


Figure 6.26 A comparison of (1/*f*) FM noise spectra in Si bipolar transistors and GaAs FETs [13].

quality on-chip resonators, but most high-performance designs typically employ an off-chip resonator of some kind. Even in designs that do employ on-chip resonator circuits, the physical size required by the resonant circuit is often a significant portion of the overall chip area.

In order to avoid this limitation, a number of groups have explored the use of ring-oscillator-based VCOs as alternative approaches for the realization of voltage-variable oscillators [19, 20]. These circuits have the advantage that they do not require any kind of resonator, but their frequency and phase characteristics are somewhat poorer than those of resonator-based approaches. An illustrative example of a ring-oscillator VCO is shown in Figure 6.27.

The oscillation frequency of the circuit is determined by the gate delay in the inverting stages, which can be controlled by the current through the gate, and the number of gates in the oscillator, and is given by

$$f_{osc} = \frac{1}{2n\tau_d} \tag{6.54}$$

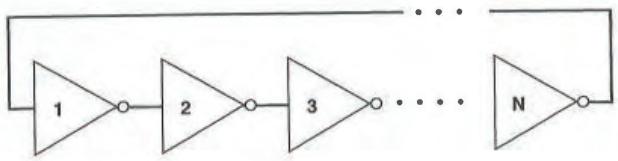


Figure 6.27 Ring oscillator voltage-controlled oscillator.

where n is the number of inverters in the ring oscillator, and is necessarily odd, and τ_d is the gate delay per stage.

One limitation of this approach is that maximum frequency of oscillation is typically limited by the minimum number of gates (three, typically) and the minimum gate delay, which is process dependent. A number of techniques have been developed to circumvent this limitation. One approach (shown in Figure 6.28) relies on the fact that at any given time in the ring oscillator, the output of $(n + 1)/2$ stages are high and $(n - 1)/2$ stages are low, or the output of $(n + 1)/2$ stages are low and $(n - 1)/2$ stages are high [21]. The oscillator toggles between these two conditions at a frequency $1/(2\tau_d)$. By sensing these transitions, we can generate a frequency of $1/(2\tau_d)$. The design of Figure 6.28 employs transconductance stages to convert the inverter output voltage to a current, and then the resulting currents are summed at a single output node. This approach has been employed to realize an 8-GHz oscillator in a 20-GHz BiCMOS process.

Another approach utilizes a four stage ring-oscillator, where one of the stages of the ring is noninverting. The outputs of the circuit are mixed with each other,

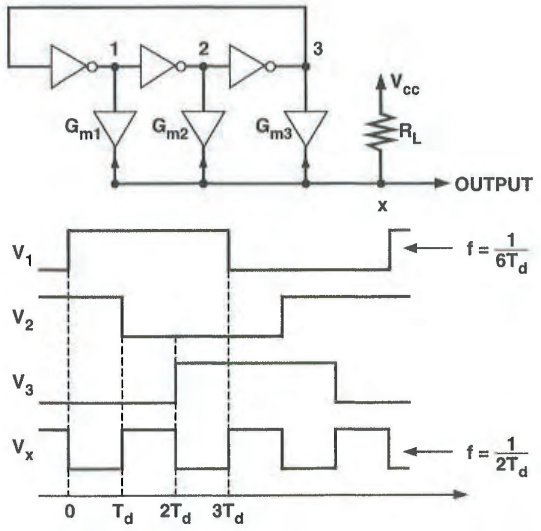


Figure 6.28 Ring oscillator VCO operating at $1/(2\tau_d)$ by sensing the transitions in the output of the inverters [21].

and twice the output frequency results. See Figure 6.29 for an example [22]. The advantage of this approach is that in-phase (I) and quadrature phase (Q) output are available simultaneously from the structure, which is a major consideration for quadrature modulation and demodulation in communication systems design (see Chapter 4). Furthermore, an output frequency of four times the ring is available if the I and Q signals are themselves multiplied. As a result, the final available output frequency is $1/2\tau_d$.

Noise in Ring-Oscillator VCOs

The noise in ring-oscillator VCOs is difficult to calculate exactly, because of the highly nonlinear nature of the circuit. However, the noise power spectral density can be estimated using techniques similar to those described in [23] for the analysis of a multivibrator, and are extended here for the case of a ring oscillator.[1] When this derivation is carried out, the resulting phase noise, in dBc/Hz, for the ring-oscillator VCO is given by

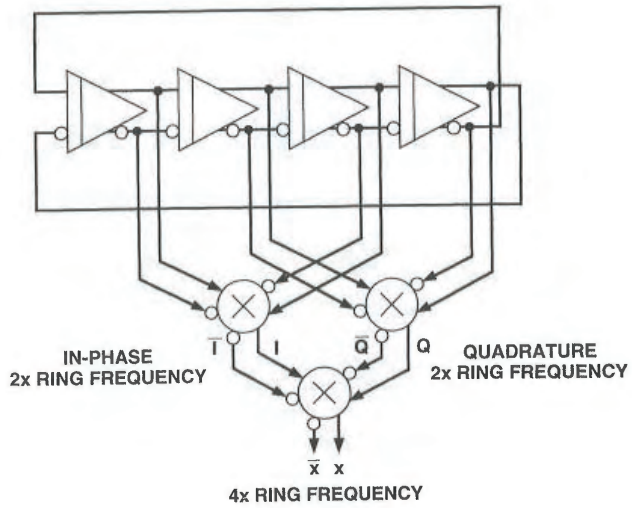


Figure 6.29 Frequency doubling and quadrupling ring-oscillator VCO [22].

[1]This analysis was provided to the author by Dr. Cynthia Baringer of the Hughes Research Laboratories.

$$S_\phi(f_m) \cong \frac{f_o}{f_m^2}\left(\frac{\Delta\tau_{\text{rms}}}{T_0}\right)^2 \tag{6.55}$$

where f_0 is the fundamental frequency of oscillation, T_0 is the fundamental oscillation period ($1/f_0$), and f_m is the offset frequency from the carrier.

The value $\Delta\tau_{\text{rms}}$ is the rms variation in the output waveform of the VCO. The jitter in the output waveform is defined as the ratio of the mean μ and rms variation of the output; that is,

$$\text{jitter} = \Delta\tau_{\text{rms}}/\mu \tag{6.56}$$

and is usually expressed in ppm.

A useful estimate for this rms variation can be given by

$$\Delta\tau_{\text{rms}} = \left(N/2\left(C_P/R_L I_{\text{EE}}^2\right)\nu_{\text{in}}^2\right)^{1/2} \tag{6.57}$$

where N is the number of stages in the ring oscillator, I_{EE} is the bias current in the oscillator, R_L is the load resistance in the inverter, and C_P is the parasitic resistance at the output of the circuit. The input referred noise voltage, ν_{in}, is in volts/Hz$^{1/2}$, which can be calculated from a simple linear analysis of the circuit using a SPICE simulator or its equivalent.

These last two expressions demonstrate that the noise of the oscillator can be reduced by increasing the bias current and decreasing the parasitic capacitances, which has the overall effect of maximizing the slew rate of the circuit and minimizing the time period where the circuit is in the transition region, where it is amplifying the noise of the differential pair.

The noise of a ring-oscillator VCO can be very acceptable for some applications. The circuit described in [21] exhibited a phase noise of −75 dBc/Hz at a 1-kHz offset from the carrier at 6 GHz. The circuit described in [22] exhibited a phase noise of −100 dBc/Hz at a 100-kHz offset from the carrier at 6.8 GHz.

6.3.12 Multivibrator VCO Designs

Another approach for the design of resonator-less VCOs is the traditional multivibrator, an example of which is shown in Figure 6.30. This design has been used extensively for a number of lower frequency (<500 MHz) monolithic VCOs [23], and is a specific example of a class of oscillators known as *relaxation oscillators*.

The circuit operates with transistors Q1 and Q2 alternately turning on and off, with capacitor C being discharged through current sources I. Diodes D1 and

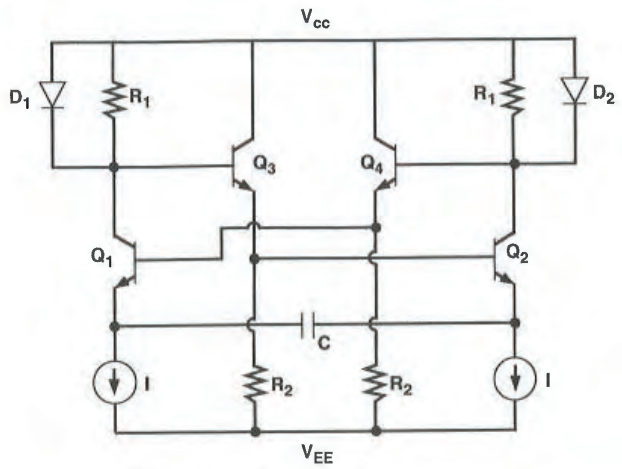


Figure 6.30 Multivibrator VCO. The circuit functions by charging and discharging capacitor C through current sources I.

D2 set the voltage swings at the collectors of Q1 and Q2, and transistors Q3 and Q4 are level-shifting devices. A square-wave output appears at the collectors of Q1 and Q2, and a triangle wave appears across the capacitor, whose frequency is given by

$$f_o = \frac{I}{4CV_{be(on)}} \tag{6.58}$$

A variety of techniques have been developed to improve the temperature stability and drift of the circuit, but they are all functionally equivalent to that of Figure 6.30. The oscillation frequency is controlled by varying the current I through the current sources, and a wide frequency tuning range can be achieved.

Another version of this oscillator type is shown in Figure 6.31. In this case, the clamping diodes D1 and D2 have been eliminated, and the circuit operates as a fixed oscillator of frequency

$$f_o = \frac{1}{8RC} \tag{6.59}$$

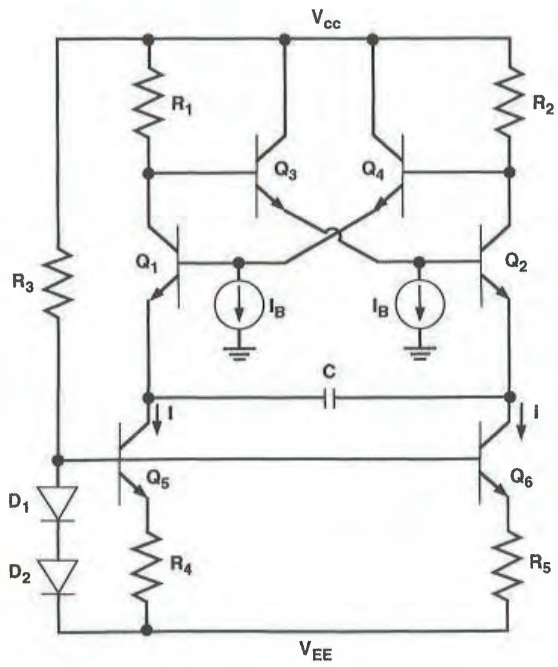


Figure 6.31 Multivibrator oscillator that oscillates at a fixed frequency $f = 1/8R_1C$.

Noise in Multivibrator VCOs

The noise process in relaxation oscillators was discussed in [23], and the results are summarized here. Like ring oscillators, the noise in a multivibrator is intrinsically higher than that of an oscillator that employs a high-Q resonator.

For a general relaxation oscillator, the jitter in the output waveform is given by

$$\text{jitter} = \frac{K\nu_{nrms}}{S} \tag{6.60}$$

where K is a constant, S is the slope of the waveform at a triggering point, and v_{nrms} is the equivalent input noise voltage to the circuit. For this particular multivibrator design, the jitter can be expressed as

$$\sigma(T) = \alpha \frac{RC}{I_o - I_R} \sigma(I_n) \tag{6.61}$$

where I_R is $V_T/2R$ ($V_T = kT/q$), α is a constant approximately equal to 1, and $\sigma(I_n)$ is the standard deviation of the equivalent noise current generator at the output of the multivibrator (the collector of transistor Q2 of Figure 6.30). This last quantity can be calculated from SPICE modeling of the circuit over the desired current and temperature range of operation.

Equation (6.61) can be used to calculate the jitter of the resulting waveform, or it can be used in conjunction with (6.55) to calculate the theoretical phase noise of the output waveform. These results demonstrate that the output waveform jitter can be reduced by increasing the voltage swing in the circuit (increasing the slew rate S), and by reducing the noise sources within the circuit itself.

6.4 PHASE LOCK-LOOP DESIGN

The phase-locked loop concept has a long history [24] and has been used extensively for a wide variety of communications system designs, particularly in the satellite communications area [25]. One of the advantages of the PLL architecture is its ability to realize excellent phase noise performance over a wide tuning range, while simultaneously achieving exacting standards of frequency accuracy. This, combined with its ease of implementation in monolithic form make it nearly ideal for design in wireless communications systems.

This section will discuss the basic elements of PLL design, with an emphasis on analog techniques. There is an extensive body of literature on the design and implementation of PLLs, and the reader is encouraged to consult these for a more thorough treatment of the subject [26].

6.4.1 PLL Design Fundamentals

A basic block diagram of a linearized PLL architecture is shown in Figure 6.32. During operation, the loop acquires—or locks—to an input signal, and its output displays a fixed phase relationship to the input signal. A basic loop consists of a phase detector, amplifier/loop filter, and voltage-controlled oscillator. Phase detectors produce an output voltage whose dc level is proportional to the phase difference between the two input signals. The linearized small-signal gain of the phase detector is K_d, whose units are typically volts/radian.

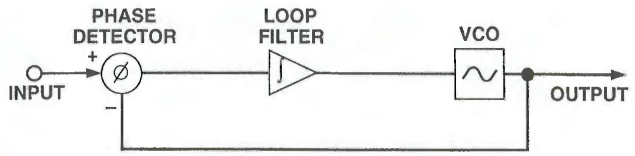


Figure 6.32 Fundamental PLL architecture.

The VCO produces an output frequency in response to an applied input voltage, and it is designed so that its output frequency range is at least as large as that required by the overall system. The ratio of change in output frequency to change in input frequency is known as the gain constant of the VCO, K_0.

The loop filter is typically a lowpass filter whose dc gain is K_f, and whose overall transfer function can be given by

$$F(s) = \frac{K_f(1 + s\tau_z)}{(1 + s\tau_p)} \tag{6.62}$$

The loop filter is typically realized as an active filter, using operational amplifiers, resistors, and capacitors to set the transfer function of the circuit.

So, the overall *phase* transfer function of the loop is given by (recall that frequency is the time derivative of phase):

$$H_\theta(s) = \frac{\theta_{out}(s)}{\theta_{in}(s)} = \frac{K_d K_0 F(s)/s}{1 + K_d K_0 F(s)/s} \tag{6.63}$$

which can be rewritten as [22]

$$H_\theta(s) = \frac{1 + \frac{s}{\omega_n}\left(2\xi - \frac{\omega_n}{\Omega_k}\right)}{1 + \frac{s}{\omega_n}2\xi + \left(\frac{s}{\omega_n}\right)^2} \tag{6.64}$$

where $\Omega_k = K_d K_f K_0$, $\omega_n = \frac{\Omega_k}{\tau_p}$, and $\xi = \frac{1}{2}\left[\frac{\omega_n}{\Omega_k} + \omega_n \tau_z\right]$.

The value ω_n is known as the loop bandwidth, or natural frequency, of the PLL. The damping factor ξ determines the transient response of the system. Both factors are crucial in determining the overall frequency and time domain response of the system. Figure 6.33 plots the normalized step response as a function of ξ and ω [27].

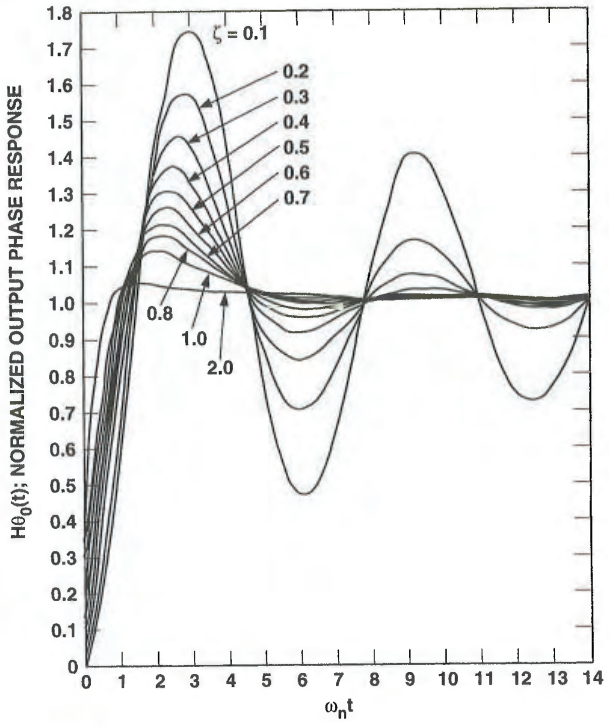


Figure 6.33 Normalized transient response of PLL.

The frequency response of the loop is identical for frequency-modulated signals (as opposed to phase-modulated signals) except for a constant term; that is,

$$H_f(s) = \frac{K_\phi}{K_0} H_\theta(s) \tag{6.65}$$

Equation (6.65) is a lowpass transfer function, which indicates that the output phase (and frequency) track the input phase and frequency up to the loop bandwidth.

Conversely, the phase error transfer function (the output phase error as a function of input phase) is given by

$$H_\theta(s) = \frac{\frac{s}{\omega_n}\left(\frac{s}{\omega_n} + \frac{\omega_n}{\Omega_k}\right)}{1 + \frac{s}{\omega_n} 2\xi + \left(\frac{s}{\omega_n}\right)^2} \tag{6.66}$$

which is a highpass transfer function whose bandwidth is ω_n.

When configured as shown in Figure 6.32, the output frequency of the PLL simply tracks the input frequency. This feature of the loop can be used to great advantage in clock recovery circuits in digital communications receivers [28, 29]. A discussion of clock recovery in digital communications circuits appears in Chapter 4.

The PLL can also be used as a technique for locking an imprecise and noisy microwave oscillator to a less noisy, but lower frequency and more precise, signal source, as shown in Figure 6.34.

In this case, a frequency divider N has been inserted between the output of the VCO and the input to the phase detector. This circuit divides its input frequency down by a factor of N. The transfer function of the loop is the same as described in (6.65), except that K_0 has been replaced by K_0/N. The output frequency transfer function is then

$$H_f(s) = N\frac{K_\phi}{K_0} H_\theta(s) \tag{6.67}$$

The output frequency is now an integer multiple N of the input frequency. These frequency dividers can be fixed or programmable. If they are programmable, then the output frequency can be varied in discrete steps, and the PLL can function as a true broadband frequency synthesizer. This technique will be discussed in Section 6.4.2.

There are a variety of nonideal effects that can degrade the performance of the PLL. These effects can limit both the noise and dynamic range of the overall circuit.

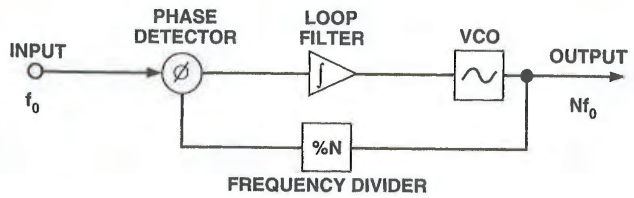


Figure 6.34 Phase-locked loop with frequency divider inserted within the loop.

Dynamic Range Limitations

The dynamic range of the PLL is limited by the tuning range of the VCO as well as the output voltage range of the loop amplifier. Since a significant amount of overshoot is possible because of the closed-loop operation of the system, the design of the loop amplifier must be able to account for the extra voltage required to accommodate the entire control voltage range. For example, if a 3V swing is required to tune the PLL over its steady-state frequency range, then the operational amplifier should be capable of producing at least a 6V swing to accommodate the transient overshoot conditions. At the same time, the VCO must also be able to accommodate a 6V swing at its input without experiencing a degradation in its control gain constant K_0.

The loss of loop gain due to transient overshoot can be a serious problem because it can decrease the damping factor ξ, which can result in even more overshoot. Major loop instability can result from this condition. Furthermore, in systems employing digital phase detectors, the transition pulses at the output of the phase detector can feed through the loop filter, causing an even larger voltage excursion at the input to the VCO. This last problem can be handled by additional filtering within the loop at a much higher frequency than the natural loop frequency [27].

Loop Lock Range and Acquisition Time

The PLL will begin its operation in an "unlocked" condition, where the input and output phases are not locked to each other. A simplified expression for loop lock-in range—the change in frequency that can be "pulled in" by the self acquisition of the loop—is given by Gardner [26]:

$$\Delta\omega_p \cong \Omega_k\sqrt{2\left(\tau_z/\tau_p\right)} \tag{6.68}$$

This expression assumes linear operation of all of the elements in the loop, and the actual available lock-in range can be substantially degraded due to nonlinearities within the loop itself.

Using a similar analysis, the time required to achieve lock assuming an initial frequency error $\Delta\omega_p$ was determined to be [22]

$$T_{\text{lock}}(\Delta\omega_L) \cong 2\tau_p\left(\frac{\Delta\omega_L}{\Delta\omega_p}\right)^2 \tag{6.69}$$

which demonstrates that the loop acquisition time is dependent on the square of the initial frequency error and the pole time constant of the loop filter.

Noise and Spurious Outputs of PLLs

There are a variety of factors that influence the noise and spurious outputs of a PLL, most of which are related to noise sources that modulate the VCO input in some way.

The first is the phase noise within the VCO itself. This can be modeled as an additive noise source at the input of the VCO, as shown in Figure 6.35. In addition to this noise source, there are noise sources associated with prescalers, both within and outside the loops, as well as the phase detector. Based on the linearized small-signal model of the PLL, the output noise power spectral density within the loop bandwidth is given by [30, 31]

$$S_o(f) = S_r(f)[N/Q]^2 + [S_{DQ}(f) + S_{DN}(f) + \{S_{PD}(f) + S_{LF}(f)\}/K_d^2(f) + S_{VCO}(f)/(K_d^2(f)K_v^2(f)K_f^2(f))]N^2 \tag{6.70}$$

where

$S_o(f)$ is the phase noise density (in dBc/Hz) at the output of the PLL;
$S_r(f)$ is the phase noise density at the input to the PLL;
N is the division ratio within the loop;
Q is the division ratio preceding the loop;
$S_{DQ}(f)$ is the phase noise density at the output of the $1/Q$ divider, assuming an ideal input;
$S_{DN}(f)$ is the phase noise density at the output of the $1/N$ divider, assuming an ideal input;
$S_{PD}(f)$ is the phase noise density at the output of the phase detector;
$S_{LF}(f)$ is the phase noise density at the input of the loop filter;
$S_{VCO}(f)$ is the phase noise density at the output of the VCO;

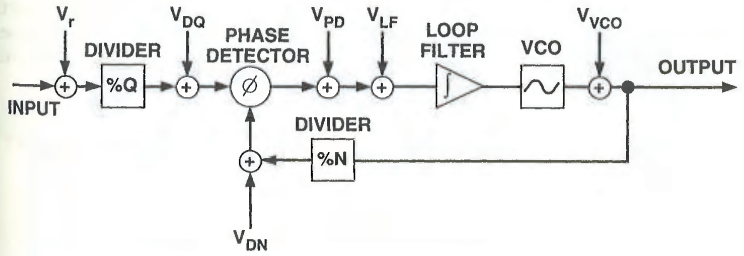


Figure 6.35 Linearized small-signal model of PLL showing noise sources.

and

K_d is the phase detector gain (V/rad);
K_f is the loop filter gain (V/V);
K_V is the VCO gain (rad/sec/V).

This expression demonstrates the importance of considering all of the noise sources when predicting loop performance, and the excellent noise-shaping properties of the loop at frequencies less than the loop bandwidth.

There are a variety of additional spurious outputs that can affect the performance of the PLL. These are principally introduced by noise sources on the VCO control line, and may arise from the power supply, package, or the pulse components of the phase detector. This latter effect can be especially troublesome because the phase detector transitions often occur near the extremes of the VCO input during a rapid excursion in the output frequency, causing even more noise to be coupled through during a transient response.

The relative level of spurious sideband components can be estimated using the following expression:

$$\frac{\text{sidebands}}{\text{carrier}} \cong \frac{V_{\text{spurious}} K_o}{2\omega_{\text{out}}} \tag{6.71}$$

where V_{spurious} is the level of spurious voltages at the input to the VCO.

6.4.2 PLL Applications

There are many different applications of the PLL presented in the previous sections. One example of their use is in "locking" a relatively noisy high-frequency oscillator (such as a typical VCO) to a relatively quiet lower frequency oscillator, such as a quartz oscillator. The resulting output has the phase noise characteristics of the quiet input reference, but the output frequency of the high-frequency VCO. This technique can be more generally employed to synthesize multiple output frequencies, as shown in Figure 6.36.

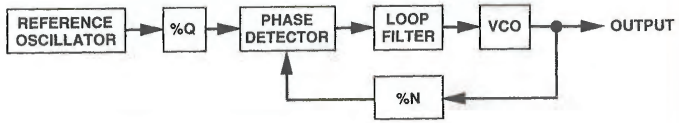


Figure 6.36 Multiple frequency synthesizer using PLL architecture.

In this case, the input frequency is divided by a factor Q, and the output frequency of the VCO is divided by N. Hence the output of the PLL is given by

$$f_{\text{out}} = \frac{Q}{N} f_{\text{in}} \tag{6.72}$$

If Q and N are programmable, then a wide variety of output frequencies is available at the output of the system.

This concept has been extended and elaborated on in several novel ways to realize an even greater range of output frequencies and programmability. Most of these approaches involve multiple loop frequency synthesizers. An excellent discussion of these circuits appears in [26].

Another example of the use of PLLs is in the area of demodulation of digital communications signals. Most digital modulation techniques require that the carrier frequency and phase be known precisely prior to demodulation. However, transmission of the carrier frequency along with the data is an inefficient use of bandwidth, and so a variety of techniques have been devised to recover the carrier and clock signals from the input. These circuits are known as *signal synchronizers*, and are discussed in more detail in Chapter 4.

For example, in binary phase-shift keyed (BPSK) data, the phase of the carrier waveform is shifted by +90 deg for a data bit = 1, and −90 deg for a data bit = 0. The resulting waveform can be represented by

$$V(t) = m(t)\sin(\omega_c t + \theta_i) \tag{6.73}$$

where $m(t)$ has an average value equal to zero if an equal number of zeros and ones is transmitted.

Figure 6.37 shows one technique for recovering the carrier from the resulting waveform. After squaring the input waveform, the resulting signal is

$$S(t) = V^2(t) = m^2(t)[1 - \cos 2(\omega_c t + \theta_i)] \tag{6.74}$$

This waveform contains a frequency component at twice the original input frequency, and a phase component of twice the original input phase. Now, the two input phases at ±90 deg are translated to ±180 deg, and they lie precisely on top of



Figure 6.37 BPSK Demodulator using PLL for frequency tracking.

each other. So, the squaring of the input signal eliminates the digital information in the original input signal, leaving only twice the carrier frequency remaining.

The PLL functions as a tracking filter, acquiring the input waveform and the divide-by-two functions to recover the original carrier signal. One drawback of this approach is that the phase of the carrier is ambiguous (it is impossible to distinguish between a 1 and a 0). Special encoding techniques can be used to recover the original phase of the carrier, and a different demodulation scheme is required (see [32] for some examples).

Fractional-N PLL Synthesizers

The conventional approach to frequency synthesis using PLLs employs a frequency divider where the division ratio is programmable between 1 and an integer N, where the step size is equal to the reference frequency. This approach requires a very small reference frequency if fine frequency resolution is desired, and this will inevitably result in an increase in the loop settling time.

An alternative approach is to have the division ratio N take on fractional values so that the output frequency could be stepped in fractions of the reference frequency. The class of PLLs that accomplish this is known as fractional-N PLLs, and we will describe their operation very briefly here. The reader is directed to [33] and [34] for a more thorough treatment of the subject.

The most common method of accomplishing the fractional division is to have the divider divide the frequency down by $N + 1$ every $M \cdot K$ cycles (where the desired fraction is $1/M$) and divide by N the remaining $(1/M) \cdot K$ cycles. The effective division ratio is then $(N + 1/M)$. For example, if a PLL was desired with an output frequency of 835 MHz and a reference frequency 100 MHz, the integral part of the division would be N = 8/9, $1/M$ would be 0.35, with $M = 2.86$. So, the VCO output could be divided by 9 every 286 cycles (for $K = 1{,}000$) and divided by 8 every 714 cycles.

This approach is typically implemented using a *phase accumulation register*, as shown in Figure 6.38. In the above case, the quantity $1/M$ (0.35) would be added to an accumulator, which would overflow when its contents exceeded 1. Each time the accumulator overflowed, the divider would divide by 9, rather than 8. The accumulator itself holds the phase error that accumulates over time as the loop operates.

This accumulated phase error is itself useful in improving loop operation. The phase detector produces a sawtooth output, whose value increases each period. This voltage variation will modulate the VCO, and create an undesirable frequency modulation in the output. The contents of the phase accumulation register contain the negative of this error and, when summed with the phase detector output, produce a "corrected" error voltage to the loop filter and VCO free of frequency modulation.

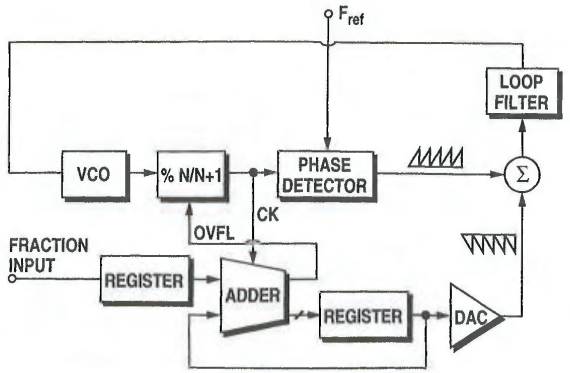


Figure 6.38 Block diagram of fractional-N frequency synthesizer.

The sawtooth generated by the D/A converter must be scaled to provide the correct voltage cancellation required by the VCO and loop filter.

6.4.3 Implementation of PLLs

Phase-locked loops are implemented using four basic components: VCOs, phase detectors, frequency dividers, and loop filters. The design of VCOs was discussed in detail in Section 6.3. This section will discuss the design of the other three components, beginning with phase detectors.

Phase Detectors

An ideal phase detector produces an output voltage proportional to the difference in phases of two input signals; that is,

$$V_{out} = K_d(\theta_1 - \theta_2) \tag{6.75}$$

Note that this output voltage is independent of the amplitude of the input voltage, so the inputs to the phase detector are usually digital signals.

This particular ideal transfer function is difficult to implement in practice, and a variety of circuit architectures have been developed to approximate it. They can be generally classified as exclusive-OR designs (sometimes referred to as "analog" designs) or digital edge-triggered designs.

Exclusive-OR designs typically utilize a balanced Gilbert multiplier, shown in Figure 6.39. The output waveforms of the circuit are also shown in the diagram. Through simple integration, the dc component of this waveform can be shown to be [35]

$$V_{dc} = I_{EE}R_C(2\phi/\pi) \tag{6.76}$$

where ϕ is the phase difference between the two input signals.

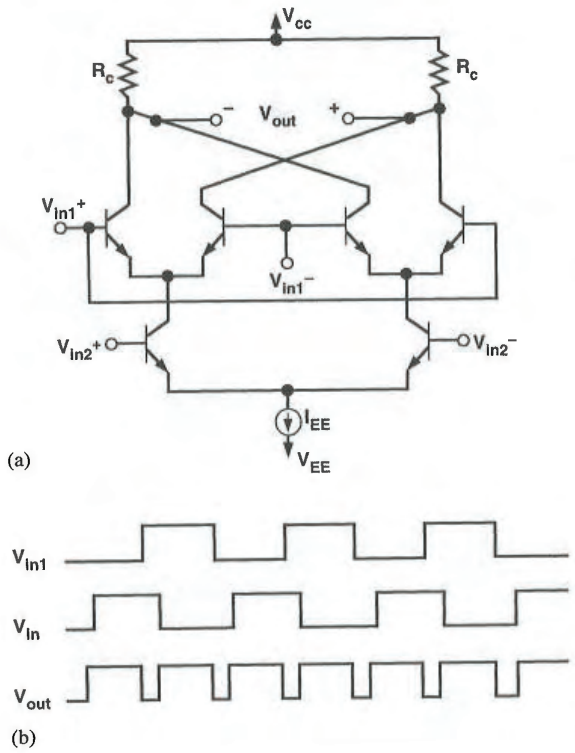


Figure 6.39 Gilbert-cell-based phase detector: (a) circuit diagram, (b) waveforms.

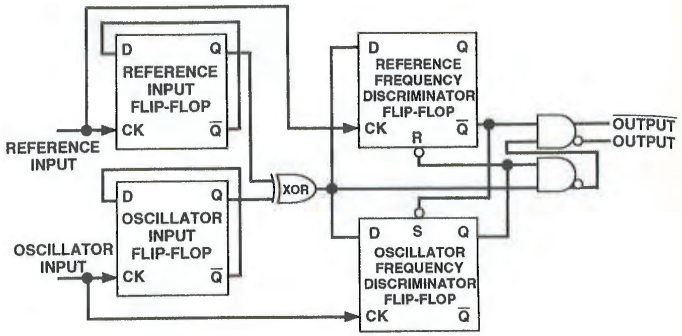
There are several drawbacks to this implementation of the phase detector. The major difficulty is in the region of 0 (2π) rad. In this case, the output voltage is symmetric about the origin, and the error voltage produced by the phase detector is the same whether the phase difference is 0.1 rad or –0.1 rad. In order to avoid this problem, PLLs using exclusive-OR type phase detectors are often operated near an input phase difference of $\pi/2$ rad. This can be accomplished by applying a low-noise dc offset voltage, equal to minus half-peak phase detector output, to the input of the integrator following the phase detector.

Another solution to this drawback is the use of so-called edge-triggered phase detector designs. A commercial example of this approach is the Analog Devices AD9901, whose block diagram is shown in Figure 6.40 along with a representative timing diagram. The phase detector is locked when the input and reference waveforms are 180-deg out of phase. When the two inputs are near lock, then only the input flip-flops and the exclusive-OR gate are active. The input flip-flops divide the input frequency and the reference frequency by a factor of two. This moves the region where the output voltage shifts abruptly from its most negative to most positive point from 180 deg (at lock) to 0 deg. The resulting transfer curve is illustrated in Figure 6.41. When the input and reference voltages are 180-deg out of phase, the output has a 50% duty cycle, and any shift in the phase between the signals will cause a change in the output duty cycle.

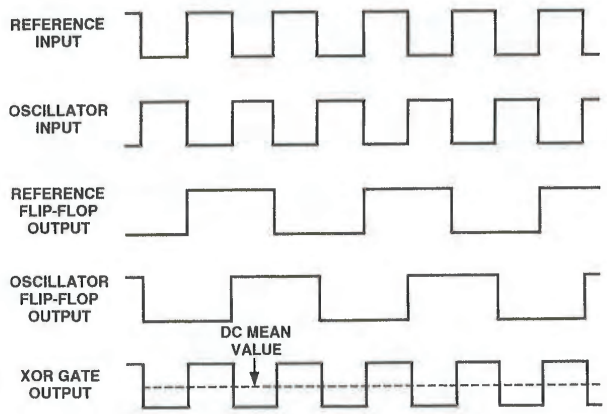
When the input and reference frequencies differ, then the circuit transitions to the frequency discriminator mode. Whenever the reference or oscillator input transitions twice before the other input (indicating that its frequency is higher), then the exclusive-OR gate is overridden and the set and reset inputs to the flip-flop push the output toward the correct frequency. When the frequency returns to the appropriate range, then the phase detector portion of the circuit takes over again. In this way, locking to a harmonic of the reference signal is avoided.

Frequency Dividers

Frequency dividers within PLLs are key components for frequency translation from high frequencies to lower frequencies. Most implementations of frequency dividers utilize digital flip-flops to generate the requisite signals, as shown in Figure 6.42. These circuits typically utilize master/slave flip-flops, where the Q-bar output is fed to the data input to divide the clock frequency by a factor of two. Digital frequency dividers can be classified into static or dynamic types. Static frequency dividers operate over a broad frequency range and use the latching action of the circuit to set the output voltage. Dynamic circuits typically utilize charge storage at critical nodes in the circuit to "hold" the output voltage [36]. These circuits typically have a narrower range of operating frequency (since charge storage at high-impedance nodes has a limited lifetime) although they can often operate at higher frequencies

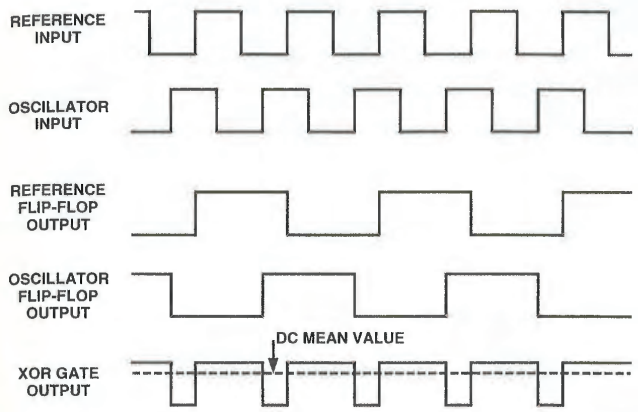


(a)

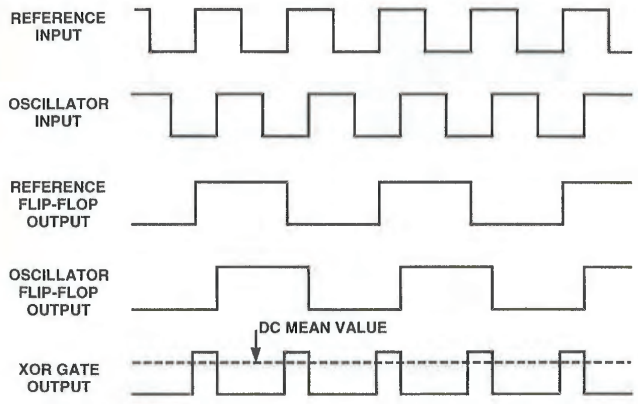


(b)

Figure 6.40 Analog Devices AD9901 edge-triggered phase detector. (a) Circuit diagram, (b) timing waveform at lock, (c) timing waveform (ϕ_{out} leads ϕ_{in}), and (d) timing waveform (ϕ_{out} lags ϕ_{in}).



(c)



(d)

Figure 6.40 (continued)

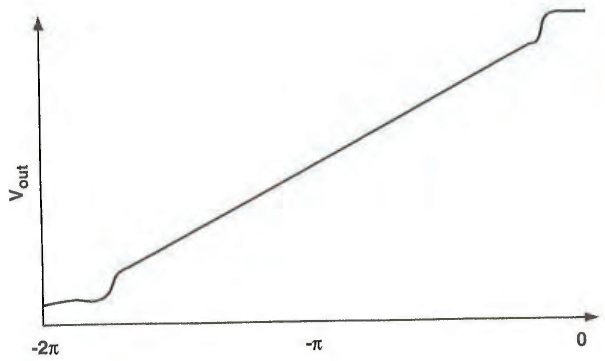


Figure 6.41 Transfer function of edge-triggered phase detector.

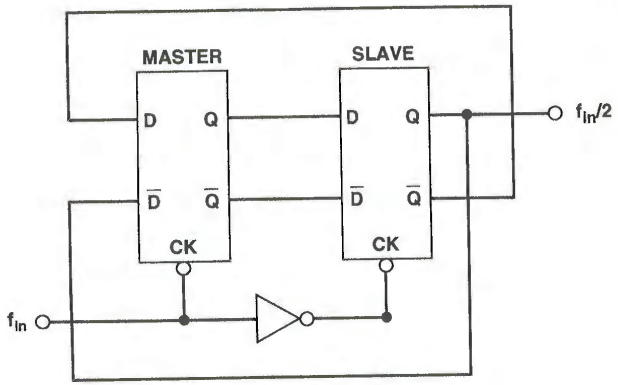
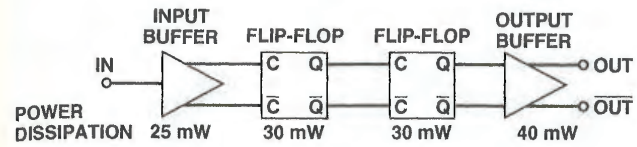


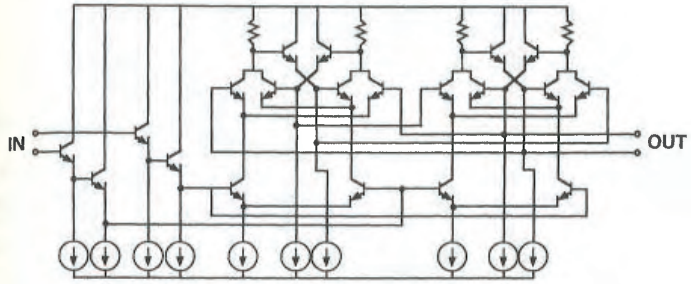
Figure 6.42 Master/slave D-type flip-flop used for frequency division. The circuit is configured as a divide-by-two.

than static dividers. Due to their reputation for improved robustness, static frequency dividers are typically employed more often than dynamic frequency dividers.

An example of a monolithic static D-type flip-flop divide-by-four is shown in Figure 6.43 [37]. This circuit operated at frequencies in excess of 5 GHz, and dissipated less than 30 mW per flip-flop. The maximum frequency of operation of



(a)



(b)

Figure 6.43 Divide-by-four frequency divider with input and output buffering for communications applications [37]. (a) Block diagram, and (b) flip-flop schematic.

these circuits is typically limited by the f_T and f_{MAX} of the transistors, with an upper limit on the clock frequency lying somewhere between $f_T/2$ and $f_T/3$, depending on the circuit design.

Flip-flops can be embedded in more complicated digital circuits to realize *programmable dividers*, whose output frequency can be altered by changing a digital input word. The division ratios can be integers, as well as fractional integers. A variety of schemes have been proposed to implement these programmable features. An example of a commercially available programmable dividers from Motorola is shown in Figure 6.44 [38], which realizes an 8/9 divider. Other typical circuits include 10/11, as well as 5/6 counters.

A second class of frequency dividers utilizes analog techniques for the realization of frequency division. These so-called *regenerative frequency dividers* use mixing, filtering, and amplification to implement the frequency division function, and a

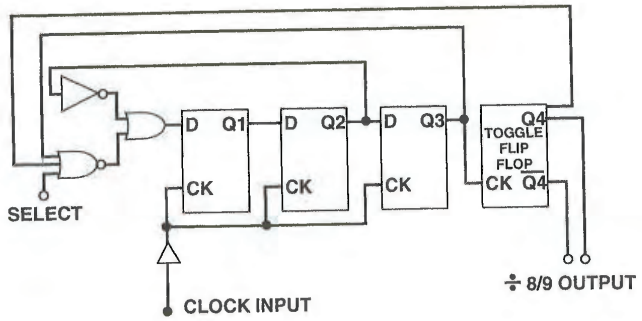


Figure 6.44 Schematic diagram of programmable 8/9 divider [38].

block diagram of a typical implementation appears in Figure 6.45 for the general case of a divide-by-N circuit.

In the case of a divide-by-two circuit, the output can simply be fed back to the mixer input without any intermediate multiplication. The interstage bandpass filter is used to reject the $f_{in} + f_{out}$ component of the mixer output, and pass the $f_{in} - f_{out}$ component. The mixer can often be combined with the amplifier to create a very simple circuit design, as shown in Figure 6.46 [39]. The circuit operated up to 28 GHz in a 40-GHz silicon bipolar process. The filter is accomplished by the addition of some capacitors within the circuit, creating a lowpass response.

Loop Filters

The final component to consider in the implementation of the PLL is the loop filter. Although this is considered the most straightforward, some subtle issues are involved that require consideration.

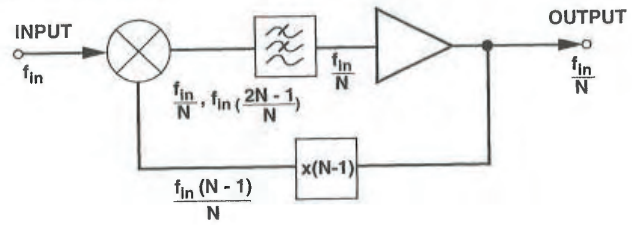


Figure 6.45 Regenerative frequency divider.

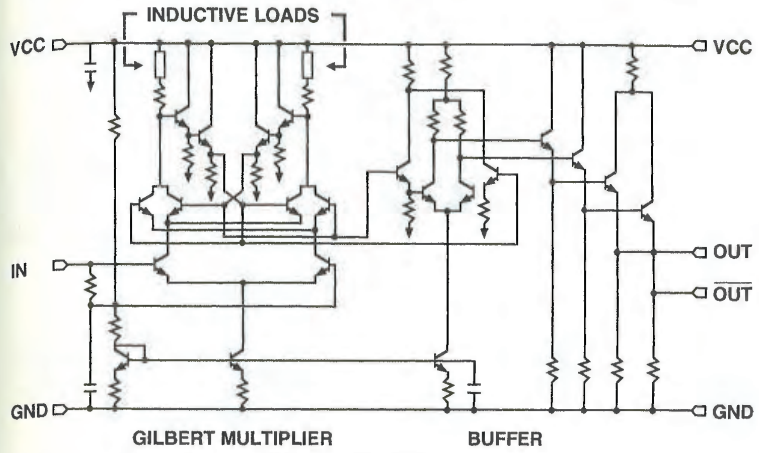


Figure 6.46 Schematic diagram of silicon bipolar regenerative frequency divider [39].

Loop filters are typically implemented using operational amplifiers (op amps), combined with resistors and capacitors to implement the desired transfer function. A typical single-pole, single-zero lowpass filter is shown in Figure 6.47. Referring back to the filter transfer function of (6.62), the necessary element values are defined as follows:

$$K_f = \frac{-R_2}{R_1} \tag{6.77(a)}$$

$$\tau_Z = C_1 R_1 \tag{6.77(b)}$$

$$\tau_P = C_2 R_2 \tag{6.77(c)}$$

One of the major limitations in the design of the loop filter is its bandwidth and dynamic range under high slew-rate conditions; the output of the amplifier must be able to supply the VCO the full voltage required during the clock cycle, without entering slew-rate limited operation.

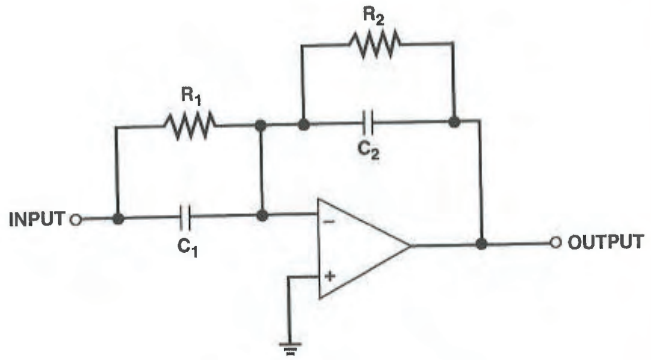


Figure 6.47 Schematic diagram of active op-amp based PLL lowpass filter.

6.5 DIRECT DIGITAL SYNTHESIS (DDS)

An alternative approach for the realization of frequency synthesizers is the use of digital techniques to synthesize the sine wave, and to follow the digital synthesis with a digital-to-analog converter (DAC). The benefits of this approach are due to the well-known advantages of digital signal processing: accuracy, absence of drift, and scalability to increasingly high levels of integration. At the same time, DDS-based systems have historically exhibited higher levels of spurious outputs than their analog counterparts, and these levels must be carefully understood and analyzed before a DDS is used in a system. This section will analyze the architectures, performance tradeoffs, and limitations of the DDS approach to frequency synthesis.[2]

The block diagram of a typical DDS appears in Figure 6.48. It consists of an n-bit accumulator, sine lookup table ROM, DAC, and lowpass filter. The accumulator performs a running calculation of the instantaneous phase angle of the desired sinusoidal wave and delivers its register value R to the sine table. The sine lookup table uses R as an address to find the appropriate sine-wave amplitude, which is stored in the ROM; that is,

$$\sin(\phi) = \sin\left(\frac{2\pi R}{2^n}\right) \tag{6.78}$$

[2]Drs. Ronald Lundgren and Viktor Reinhardt of Hughes Electronics contributed to some of the material in this section.

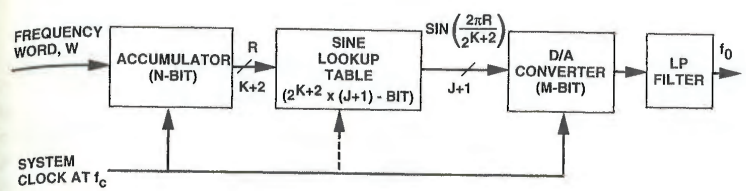


Figure 6.48 Direct digital synthesizer (DDS), utilizing an accumulator, sine lookup table, DAC, and smoothing filter.

The output of the sine table is then sent to an M-bit DAC, which converts the digital waveform into an analog one. The lowpass filter is used to remove the out-of-band harmonics and spurs, and smooth the output waveform. The output of the DAC is given by

$$V(n) = A \bmod_m[\sin(2\pi nF)] \tag{6.79}$$

where $\bmod_m(x)$ truncates x to J bits.

As a result, the ideal output of the DDS is given by

$$V_{\text{out}} = A \sin\left(\frac{2\pi Wt}{T_c 2^N}\right) \tag{6.80}$$

where W is the N-bit input word, t is time, and T_c is equal to the clock period $(1/f_c)$. As a result, the output frequency can theoretically have 2^N different unique values.

These last two expressions illustrate some of the unique advantages and limitations of the DDS approach to frequency synthesis. The advantage is that the output frequency is discretely variable over a very wide frequency range, from $f_c/2^N$ to $f_c/2$, and this frequency range is set digitally and is solely dependent on the clock frequency, which is intrinsically very accurate.

One disadvantage of the DDS approach is that the finite phase resolution of the input to the sine lookup table creates spurious degradation in the output signal. Generating a sine table to full resolution directly would require a prohibitively large ROM, and spurious errors are created by the periodic excitation of the approximation to an ideal sine wave. In addition, the *output* of the sine ROM is itself a quantized version of an ideal sine wave, further adding to spurious levels. The DAC is not ideal and can exhibit considerable "glitch" energy at major transitions, and this can also degrade the integrity of the output waveform significantly.

6.5.1 DDS Architectures and Design Considerations

The architecture outlined in Figure 6.48 forms the baseline for most basic DDS implementations. It has been noted, however, that there is a theoretical tradeoff between spurious performance and sine ROM and DAC complexity, with higher complexity resulting in better theoretical spurious performance. One major limitation is that the size of the sine lookup ROM must be $2^K \cdot J$ bits, where K is the phase resolution (in bits) and J is the DAC resolution. Various alternative architectures have been proposed that should reduce the complexity of the sine lookup ROM, and these are briefly discussed here.

Hutchinson has proposed a sine-table scheme for reducing the ROM size necessary for a given level of table accuracy [40]. The scheme is based on the approximation

$$\begin{aligned} \sin(X_K) &= \sin(X_A + X_B) \\ &\approx \sin(X_A) + \sin(X_B)\cos(X_A) \end{aligned} \tag{6.81}$$

where the input word X_K is separated into the A most significant bits X_A and the least significant bits X_B (i.e., $K = A + B$).

The sine lookup table can now be partitioned into two separate ROM circuits, ROM1 and ROM2, where ROM1 is used to generate $\sin(X_A)$ and ROM2 is used to generate $\sin(X_B)\cos(X_A)$. The output of the two ROMs are then added together to produce a complete output. The size of the ROM table is reduced, because ROM1 must store 2^A words of J-bit length, and ROM2 must store 2^B words of $(J - A + 1)$ bit length. Therefore, the total storage requirements of the ROM are now reduced from $2^K \cdot J$ to $(2^A J + 2^K(J - A + 1))$ bits.

The total phase error that results from this scheme is $(\pi^2/2(2A + 3))$, so for a J-bit accurate phase, $A \geq (J + 0.3)/2$, and

$$A_{\min} = K - \log_2(J\ln(2)) \tag{6.82}$$

For example, in the case where $K = 10$ and $J = 10$, $A_{\min} = 7$, and the number of bits required is 5,376, half of the 10,240 bits required for the brute-force approach. A block diagram of an architecture that implements this algorithm is shown in Figure 6.49.

Spurious Responses

Spurious responses arise from a number of factors, and are a significant limitation to the overall performance of DDS systems. They can be further subdivided into phase quantization errors, amplitude quantization errors, and DAC nonlinearities.

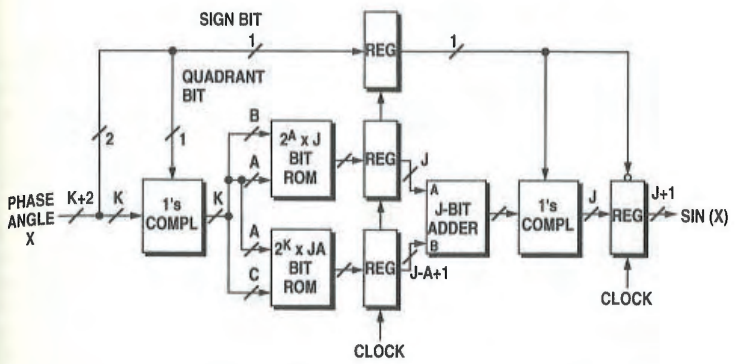


Figure 6.49 Block diagram of memory structure used to implement Hutchinson algorithm.

Phase quantization errors arise from the finite resolution of the input address for the sine lookup ROM. In particular, when the frequency select word is not an integer power of 2, the accumulator does not overflow to zero for each cycle of the output wave. As a result, residual phase increments add up in the accumulator until an overflow occurs after multiple output cycles. This long-term phase error accumulation produces spurious responses at a fraction of the clock frequency.

A variety of techniques have been proposed to minimize this effect by introducing a pseudo-random code into the accumulator with the intent of creating a small amount of phase randomization in the input phase function, with the hope of "spreading out" the resulting spur energy into a wider bandwidth [41, 42].

Amplitude quantization errors result from the finite resolution of the DAC. As a result, it occurs even when the DAC is "perfect," since even an ideal DAC will have finite resolution. This quantization noise is often assumed to be "white" (i.e., uniformly distributed in the frequency domain), but in fact is often harmonically related to the input signal, especially when the input signal is a submultiple of the clock frequency. A rough estimate for the worst case dynamic range with a perfect DAC is 6 dB/bit. So, for example, an 8-bit DAC could have a relatively modest 48-dB dynamic range.

Finally, there are nonlinearities associated with the DAC, which are different from quantization error. These nonlinearities include gain errors, offset errors, nonmonotonicity, differential nonlinearity, integral nonlinearity, and major transition glitches. An excellent discussion of the definition of all of these appears in [43], but their effect on DDS performance is not simple to analyze. In effect, the errors can be viewed as generators of harmonics and intermodulation products. Any nonlinearities in the transfer curve will produce these to varying degrees.

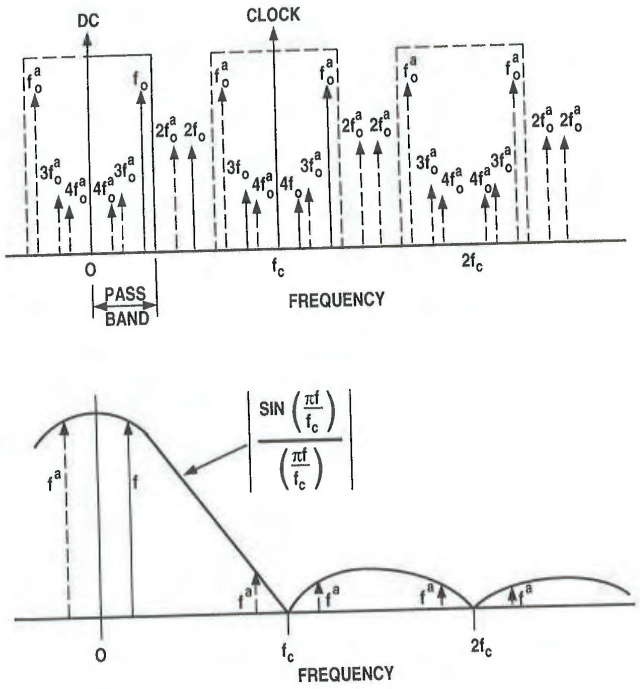


Figure 6.50 (a) Harmonics and aliased harmonics of sampled DDS output waveform. (b) The effect of lowpass filter on the spurious components of DDS output.

Major carry "glitches" create particular problems for DDS systems. These occur when the input word to the DAC changes from, say, 011111 to 100000. In typical implementations of the DAC, this will cause all of the $n - 1$ LSB current sources to turn off and the single MSB current source to turn on. In such a case, there may be a brief period where *all* of the current sources are off, or on, in which case the output swings dramatically before settling to its correct value. One measure of this problem is so-called *glitch energy* which is the integrated area under the voltage curve during the transition period. It is often specified in $pS - V$, or $nS - V$, with a value less than $1pS - V$ being a very good result for a high-speed DAC.

Spurious responses are especially a problem for high-speed DDS systems. For example, if a 500-MHz clock is used to produce a 150-MHz signal, then a second and third harmonic will also be generated at 300 and 450 MHz, which are aliased back to 200 MHz and 50 MHz, creating an in-band spurious response.

Aliasing

A DDS System that incorporates a DAC can be viewed as the convolution of a sampling process and a hold process. The sampling of a signal at frequency f at f_{clock} causes the spectrum to be duplicated at $mf_{clock} \pm f$, where m is an integer from $-\infty$ to $+\infty$. An example of this is shown in Figure 6.50(a). The hold multiplies the amplitude of these frequencies by $(\sin(\pi f/f_{clock})/(\pi f/f_{clock}))$, as shown in Figure 6.50(b). This process would not be a problem if it were not for the harmonics of the fundamental signal, which are aliased back into the band of interest.

References

[1] J. Obregon, and P. S. Khanna, "Exact Derivation of the Non-Linear Negative Resistance Oscillator Pulling Figure," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, July, 1982, pp. 1109–1111.
[2] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, Philadelphia, PA: Saunders College Publishing, 1991, pp. 841–844.
[3] J. K. Hardy, *High Frequency Circuit Design*, Reston, VA: Reston Publishing, 1979, pp. 259–260.
[4] H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York, NY: John Wiley and Sons, 1980, pp. 151–152.
[5] J. Smith, *Modern Communication Circuits*, New York, NY: McGraw-Hill, 1986, pp. 252–261.
[6] K. L. Kotzebue and W. J. Parrish, "The use of large signal S-parameters in microwave oscillator design," *Proc. of the Int'l Microwave Symposium on Circuits and Systems*, 1975.
[7] M. Vehovec, L. Houselander, and R. Spence, "On oscillator design for maximum power," *IEEE Trans. on Circuit Theory*, Sept. 1968, pp. 281–283.
[8] B. K. Kormanyos and G. M. Rebeiz, "Oscillator design for maximum added power," *IEEE Microwave and Guided Wave Letters*, Vol. 4, No. 6, June 1994, pp. 205–207.
[9] A.P.S. Khanna, "Oscillators," book chapter in *GaAs MESFET Circuit Design*, ed. R. Soares, Norwood, MA: Artech House, 1988, pp. 361–364.
[10] S. A. Maas, *Nonlinear Microwave Circuits*, Norwood, MA: Artech House, 1988, pp. 463–467.
[11] J. Y. Lee and U. S. Hong, "Voltage controlled dielectric resonator oscillator using three-terminal MESFET varactor," *Electronic Letters*, Vol. 30, No. 16, Aug. 1994, pp. 1320–1321.
[12] E. A. Allamando, "X-band MMIC voltage-controlled oscillator using HEMTs," *Microwave and Optical Technology Letters*, Vol. 5, No. 8, July 1992, pp. 395–398.
[13] A. A. Sweet, *MIC and MMIC Amplifier and Oscillator Circuit Design*, Norwood, MA: Artech House, 1990, pp. 183–189.
[14] A. A. Sweet, "A general analysis of noise in Gunn oscillators," *Proc. of IEEE*, Vol. 60, Aug. 1972.
[15] J. S. Yuan, "Modeling the bipolar oscillator phase noise," *Solid State Electronics*, Vol. 37, No. 10, Oct. 1994, pp. 1765–1767.
[16] U. Rohde, "Oscillator design for lowest phase noise," *Microwave Engineering Europe*, May 1994, pp. 35–40.

[17] A. Howard, "Simulate oscillator phase noise," *Microwaves and RF*, Vol. 32, No. 11, Nov. 1993, pp. 65–70.
[18] L. D. Cohen, "Low phase noise oscillator with flicker (1/*f*) noise suppression circuit," IEEE MTT-S Int'l Microwave Symposium Digest, June 1992, pp. 1081–1084.
[19] A. Buchwald, and K. Martin, "High-Speed Voltage Controlled Oscillator with Quadrature Outputs," *Electronics Letters*, Vol. 27, No. 4, 1991, pp. 309–310.
[20] S. K. Enam and A. A. Abidi, "NMOS ICs for Clock and Data Regeneration in Gb/s Optical-Fiber Receivers," *IEEE J. of Solid State Circuits*, Vol. 27, No. 12, 1992, pp. 1763–1774.
[21] R. Razavi, and J. J. Sung, "A 60 GHz 60 mW BiCMOS Phase-Locked Loop," *IEEE J. of Solid-State Circuits*, Vol. 29, No. 12, Dec. 1994, pp. 1560–1565.
[22] A. Buchwald, K. Martin, A. Oki, and K. Kobayashi, "A 6-GHz Integrated Phase-Locked Loop Using AlGaAs/GaAs Heterojunction Bipolar Transistors," *IEEE J. of Solid-State Circuits*, Vol. 27, No. 12, Dec. 1992, pp. 1752–1762.
[23] A. Abidi, and R. Meyer, "Noise in Relaxation Oscillators," *IEEE J. of Solid-State Circuits*, Vol. SC-18, No. 6, Dec. 1983, pp. 794–802.
[24] H. Bellescize, "La Reception Synchrone," *Onde Electr.*, Vol. 11, June, 1932, pp. 230–240.
[25] W. Lindsey, and M. Simon, (Eds.), *Phaselocked Loops and Their Application*, New York, NY: IEEE Press, 1978.
[26] F. Gardner, *Phaselock Techniques*, 2nd Edition, Wiley Interscience, 1979.
[27] Motorola MECL Data Book, Chapter 6, "Phase-Locked Loops," 1993.
[28] B. Sklar, *Digital Communications: Fundamentals and Applications*, New York, NY: Prentice-Hall, 1988.
[29] J. G. Proakis, *Digital Communications*, New York, NY: McGraw-Hill, 1983.
[30] A. Hodisan, Z. Hellman, and A. Brillant, "CAE Software Predicts PLL Phase Noise," *Microwaves and RF*, Nov. 1994, pp. 95–100.
[31] V. F. Kroupa, "Noise Properties of PLL Systems," *IEEE Trans. on Communications*, Oct. 1982, pp. 2244–2251.
[32] U. L. Rohde, and T.T.N. Bucher, *Communications Receivers: Principles & Design*. New York, NY: McGraw-Hill, 1988.
[33] U. L. Rohde, *Digital PLL Frequency Synthesizers—Theory and Design*, New York, NY: Prentice-Hall, 1983.
[34] W. F. Egan, *Frequency Synthesis by Phase Lock*, Wiley Interscience, 1981.
[35] P. Gray, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York, NY: John Wiley and Sons, 1989.
[36] H. Ichino et al., "18 GHz 1/8 Dynamic Frequency Divider using Si Bipolar Technologies,"
[37] I. Kipnis et al., "A Wideband Low-Power, High Sensitivity And Small-Size 5.5 GHz Static Frequency Divider," *IEEE 1988 Bipolar Circuits and Technology Mtg.*, pp. 88–91.
[38] Motorola MECL Data Book, 1993.
[39] M. Kurisu et al., "A Si Bipolar 28-GHz Dynamic Frequency Divider," *IEEE J. Solid-State Circuits*, Vol. 27, No. 12, Dec. 1992, pp. 1799–1804.
[40] B. H. Hutchison, "Contemporary Frequency Synthesis Techniques," *Frequency Synthesis and Applications*, Ch. 1, New York, NY: IEEE Press, 1975.
[41] V. S. Reinhardt, "Direct Digital Synthesizers," *Proc. of 17th Annual Precise Time and Time Interval Applications and Planning Mtg.*, Washington, DC, Dec., 1985.
[42] C. E. Wheatley, "Digital Frequency Synthesizer with Random Jittering for Reducing Discrete Spectral Spurs," U.S. Patent 4,410,954 (Oct. 1983).
[43] *Analog Devices Data Converter Reference Manual*, Vol. 1, 1993.

CHAPTER 7

AMPLIFIERS FOR WIRELESS COMMUNICATIONS

Cynthia Baringer
Hughes Research Labs
Malibu, CA

Christopher Hull
Rockwell International
Newport Beach, CA

7.1 INTRODUCTION

Signal amplification is a fundamental function of the radio portion in all wireless communication systems. The receiver must take the smallest possible signal level at its input and amplify that signal to the minimum level that can be detected by the demodulator. Correspondingly, the transmitted signal level must be amplified sufficiently so that despite loss inherent in wireless transmission, the signal can be received by a nearby base station.

There are several types of amplifiers that serve specific functions in the radio. Amplifiers in the receiving chain that are closest to the antenna contribute directly to the noise performance. These low-noise amplifiers (LNAs) must provide enough gain so that the noise contributions from the rest of the receiving components are minimal. The amplifiers closest to the antenna in the transmit path are power

amplifiers (PAs). They must amplify the signal linearly and efficiently. Noise is not a primary consideration.

Along the receiver chain, there are variable-gain amplifiers (VGAs). The function of these amplifiers is to help keep the receiver output constant over the range of expected signal intensity. The noise and linearity requirements of VGAs are not as strict as for LNAs and PAs, respectively. However, VGAs must have reasonable dynamic range, which is a measure of an amplifier's ability to handle both small and large signals.

Higher frequency bands are being sought to avoid congestion. Therefore, wireless systems are requiring higher frequency and wider bandwidth amplification. The bandwidth of an amplifier may ultimately be limited by the technology. Low-cost objectives and higher integration often prohibit the use of higher performing technologies. Techniques used to maximize the bandwidth of an amplifier, for a given technology, are introduced in Section 7.2.

In the next three sections of this chapter, amplifiers widely used in current wireless systems are presented. Section 7.3 discusses the benefits of using a simple cascode for use in low-noise applications. Section 7.4 compares the dynamic range of several VGAs. The design of a PA is covered in Section 7.5. Issues of practical concern, such as package limitations, are emphasized.

7.2 WIDEBAND AMPLIFIER TECHNIQUES

Operating frequencies of many commercial wireless systems are in the low gigahertz range. Amplifiers must provide linear gain at these frequencies.

Feedback is a common technique used in wideband amplification to provide gain stability over process and supply variations, wider bandwidth, and lower distortion. Other techniques for broadbanding, not discussed in this section, include a Gilbert-cell amplifier approach [1] and feedforward techniques such as the Cascomp circuit [2].

The bandwidth of amplifiers is fundamentally limited by the transistor gain-bandwidth product. Applying feedback to an amplifier does not change this product, but rather, allows the designer to trade gain for bandwidth. Compound stages, which replace single transistors with multiple transistors that behave like a single transistor, can exceed the single device gain-bandwidth product. This section will conclude with a discussion of practical design issues in wideband feedback amplifiers, including a discussion of parasitic capacitance and impedance matching.

7.2.1 Gain-Bandwidth Product

A useful measure of the performance of a wideband amplifier is its gain-bandwidth product. It can be used to compare the effective utilization of topologies for a given application. In an N-stage amplifier, the gain-bandwidth is arbitrarily defined as

$$(\text{GBW})_N = (\text{Gain})^{1/N}\text{BW} \tag{7.1}$$

where N is the number of stages, Gain is the total midband gain, and BW is the overall bandwidth.

In general, amplifiers with fewer gain stages will have better gain-bandwidth because the bandwidth shrinks as stages are added. Consider an amplifier that is a cascade of N identical stages each with a single-pole role-off response given as

$$\frac{v_o}{v_i} = \frac{G}{1 + j\left(\frac{\omega}{B}\right)} \tag{7.2}$$

The gain-bandwidth product of each stage is $G \cdot B$. (G is the gain and B is the bandwidth of each individual stage.) The overall transfer function is

$$H(j\omega) = \frac{G^N}{\left(1 + j\frac{\omega}{B}\right)^N} \tag{7.3}$$

The total midband gain is Gain = G^N and the overall bandwidth is when $|H(j\omega)| = \frac{G^N}{\sqrt{2}}$. This gives

$$\text{BW} = B\sqrt{2^{\frac{1}{N}} - 1} \tag{7.4}$$

and using (7.1), the amplifier gain-bandwidth product is

$$(\text{GBW})^N = \text{GB}\sqrt{2^{\frac{1}{N}} - 1} \tag{7.5}$$

So, as the number of stages increases, the amplifier gain-bandwidth is reduced.

Only single-stage and two-stage amplifiers are considered in the following sections. The overall gain-bandwidth product of single-stage and two-stage amplifiers are denoted $(\text{GBW})_1$ or $(\text{GBW})_2$, respectively. In the discussion of two-stage amplifiers, the individual stage gain-bandwidth products are denoted GB_1 and GB_2 for the first and second stages, respectively.

Single-Stage Amplifiers

Amplifier gain-bandwidth products are ultimately limited by the unity-gain frequency f_T of the transistor. The device f_T is the frequency at which the transistor current

gain equals one. This can be found from a high-frequency model of a transistor, shown in Figure 7.1. The input impedance is predominantly capacitive and the transconductance g_m is approximately independent of frequency. The unity-gain frequency can be found to be

$$f_T = \frac{g_m}{2\pi C_{in}} \tag{7.6}$$

To first-order, the expression predicts the unity-gain frequency of a FET when $C_{in} = C_{gs}$ and of a bipolar device when $C_{in} = C_\pi$. Analysis with a more accurate transistor model yields a two-pole transfer function. In most cases, there is a dominant pole and while the presence of the second pole does not ideally degrade the gain-bandwidth product, it does contribute excess phase in the circuit. For low source impedance, the gain-bandwidth is degraded by resistive parasitics [3].

The amplifier gain-bandwidth of a simple single-stage amplifier, such as a common-emitter stage, is identical to the transistor gain-bandwidth product. That is, $(GBW)_1 = 2\pi f_T$. By reducing the load resistance, voltage gain can be exchanged for bandwidth, which is a technique referred to as resistive broadbanding. Lower impedances also reduce the interaction between stages by minimizing the Miller effect on device feedback parasitics. These parasitic capacitances can lead to narrowbanding. For these reasons, wideband circuits are always low-impedance circuits.

The gain-bandwidth product of a single-pole feedback amplifier is also identical to the transistor gain-bandwidth product. Though there are only two circuit topologies to consider, series feedback (as in emitter-degeneration) and shunt feedback, a general approach will be taken to find the gain-bandwidth product. In a feedback amplifier, the closed-loop gain is given in terms of the *open-loop gain* and *loop gain*. Open-loop gain is the gain of the basic amplifier when the feedback network is not present and the loop gain is the product of the open-loop gain and the feedback. The closed-loop gain of the single-pole feedback amplifier is [4]

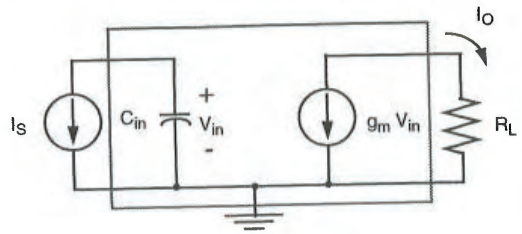


Figure 7.1 Simplified high-frequency small-signal transistor model.

$$A(s) = \frac{a(s)}{1 + T(s)} \tag{7.7}$$

For an amplifier with a single pole and gain-bandwidth of ω_T, the open-loop gain is expressed as $a(s) = \dfrac{a_o}{1 + a_o\dfrac{s}{\omega_T}}$.

If the feedback network is made up of resistors, then the feedback factor f_o is independent of frequency and the loop-gain can be expressed as $T(j\omega) = f_o a(s)$. Substituting the expressions for open-loop gain and loop-gain into (7.7) gives (after some manipulation)

$$A(s) = \frac{\dfrac{a_o}{1 + a_o f_o}}{1 + \left(\dfrac{s}{\omega_T}\right)\left(\dfrac{a_o}{1 + a_o f_o}\right)} \tag{7.8}$$

Referring to (7.8), the low-frequency closed-loop gain is obtained by setting $s = 0$ and the bandwidth is obtained by finding the frequency at which the denominator equals zero. The gain-bandwidth is calculated to be

$$G \cdot B = \left(\frac{a_o}{1 + a_o f_o}\right)\left(\frac{1 + a_o f_o}{a_o}\omega_T\right) \tag{7.9}$$

or

$$G \cdot B = \omega_T \tag{7.10}$$

Thus, ideal resistive feedback does not change the gain-bandwidth product of a single-pole amplifier.

Two-Stage Amplifiers

Improved performance can often be obtained by applying feedback around two stages rather than a single stage. The drawback is that amplifier stability must be addressed in the design. The negative feedback at low frequencies becomes positive feedback at high frequencies, reducing or eliminating any phase margin in the system.

The problem of inadequate phase margin can be overcome by either dominant-pole compensation or by creation of a feedback zero. In the dominant-pole technique, a single pole is introduced at a sufficiently low frequency that the magnitude of the

loop gain drops to less than unity before the phase shift reaches 180 deg [4]. The disadvantage of dominant-pole compensation is a significant gain-bandwidth reduction. For wideband amplification, the feedback-zero technique is preferred since little or no gain-bandwidth is lost.

An example of a feedback configuration with a feedback zero is shown in Figure 7.2. The amplifier is shown schematically as an operational amplifier. The capacitor in parallel with R_F increases feedback at high frequencies, giving extra phase lead to compensate for the phase lag in the amplifier. Excessive peaking can be observed in the two-stage amplifier without the capacitor. The feedback zero can eliminate the peaking and eventually cause an over-damped response characteristic.

The closed-loop response of the two-stage feedback amplifier may be written as

$$A(s) = \frac{K}{s^2 + \frac{\omega_o}{Q}s + \omega_o^2} \tag{7.11}$$

where K is the product of the low-frequency open-loop gain $a(0)$ and the two open-loop pole frequencies p_1 and p_2. The closed-loop response is characterized by a resonance frequency, $\omega_o = \sqrt{p_1 p_2 (T_o + 1)}$ and a quality factor, $Q = \dfrac{\omega_o}{p_1 + p_2 + \dfrac{p_1 p_2 T_o}{z}}$ where T_o is the low-frequency loop gain and z is the frequency of the feedback zero. The value of z can be chosen to give a desired value of Q. For $Q = \dfrac{1}{\sqrt{2}}$, the bandwidth *is* the resonant frequency and is given as

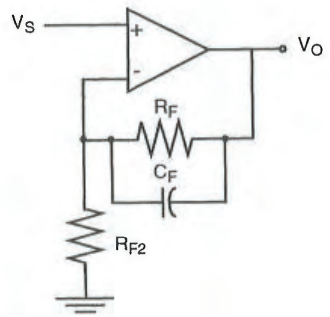


Figure 7.2 Amplifier with feedback zero.

$$\text{BW} = \sqrt{(T_o + 1)p_1 p_2} \tag{7.12}$$

The low-frequency gain is

$$A_v = \frac{a(0)}{1 + T_o} \tag{7.13}$$

Combining (7.12) and (7.13) yields the expression

$$\sqrt{\text{Gain}} \cdot \text{BW} = \sqrt{p_1 p_2 a(0)} \tag{7.14}$$

If there are two gain stages with gains $a_1(j\omega)$ and $a_2(j\omega)$, then (7.14) can be rewritten as

$$\sqrt{\text{Gain}} \cdot \text{BW} = \sqrt{p_1 p_2 a_1(0) a_2(0)} \tag{7.15}$$

$$= \sqrt{\text{GB}_1 \cdot \text{GB}_2} \tag{7.16}$$

where $\text{GB}_1 = p_1 a_1(0)$ and $\text{GB}_2 = p_2 a_2(0)$.

The product $\sqrt{\text{Gain}} \cdot \text{BW}$ is conserved and is independent of the loop gain. Because a two-pole amplifier generally is made up of two gain stages, a reduction in bandwidth by a given factor increases the gain by the square of that factor. The reason for this is that each of the two stages increases in gain when the bandwidth is reduced. Hence $\sqrt{\text{Gain}} \cdot \text{BW}$, as described in the beginning of this section, is the proper performance measure for a two-stage amplifier.

The gain-bandwidth will only be achieved if a sufficiently high Q is obtained. A necessary and sufficient condition for $Q > \dfrac{1}{\sqrt{2}}$ is

$$T_o > \frac{1}{2}\left(\frac{p_1}{p_2} + \frac{p_2}{p_1}\right) \tag{7.17}$$

Thus, the loop-gain must be at least unity. For increasing loop gains, further initial pole position spreads are permitted. Because of the unwanted feedback zero introduced through capacitive parasitics (such as the collector-base junction capacitance of a bipolar transistor), the loop gain must be larger than the right-hand side of (7.17) for optimal gain-bandwidth.

Compound-Transistor Stages

The use of compound transistors in place of single bipolar transistors is common in wideband amplifier design because they can achieve higher current gain for a

given frequency. A common configuration is common-collector/common-emitter, referred to as a Darlington configuration. The first transistor can be biased either resistively as shown in Figure 7.3(a) or with a diode as shown in Figure 7.3(b). The former has seen more implementation in circuit design so far, but the latter configuration exhibits superior phase characteristics [5, 6].

The current gain of the resistive Darlington has a two-pole characteristic, causing the compound transistor to appear more like a two-stage amplifier. There is a zero in the transfer characteristic, which can be used to achieve extremely wideband designs [7]. Substituting the resistive Darlington for a single transistor in a parasitic-sensitive circuit is not advantageous, since the base resistance and output capacitance is double that of the single transistor.

At low frequencies, the diode Darlington (Figure 7.3(b)) appears as a single transistor with twice the current gain and twice the effective f_T. At very high frequencies, the pole-zero pair will add extra phase to the current gain. The input impedance is simply double that of a simple transistor and the transconductance is identical to a simple transistor at low frequencies, though it is slightly lower at high frequencies (near f_T). The impact of degraded transconductance is even more evident when the source impedance is low.

The use of the diode Darlington over the resistive Darlington is preferable. The resistive Darlington has excellent current-gain performance but the diode Darlington maintains double the current gain of a single transistor to very high frequencies. The resistive Darlington exhibits poor phase characteristics; phase shifts of 180 deg at frequencies one decade below f_T may occur. Another key advantage of the diode Darlington is that it can be substituted for a simple transistor with only minor

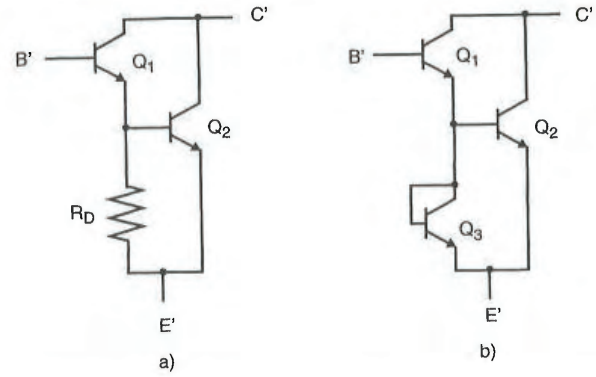


Figure 7.3 Resistive Darlington (a) with resistor feedback, (b) with transistor feedback.

changes in bias and compensation for the slight extra phase shift while the resistive Darlington may require an entire redesign of the circuit.

7.2.2 Gain-Bandwidth of Practical Circuits

Gain-bandwidth concepts are applied to practical circuits in this section. Although the circuit equations given in this section are formulated for bipolar transistors, the results apply equally well to FETs with little modification. Because of the low transconductance of FETs, at lower bias currents some configurations are not suitable to use with FETs.

Transistor Model

To simplify the analysis of the practical amplifier designs presented in the following sections, a simplified equivalent circuit of a bipolar transistor is shown in Figure 7.4. The elements in Figure 7.4 can be related to the elements of the small-signal hybrid-π model [3] shown in Figure 7.5. The small-signal model is equivalent to

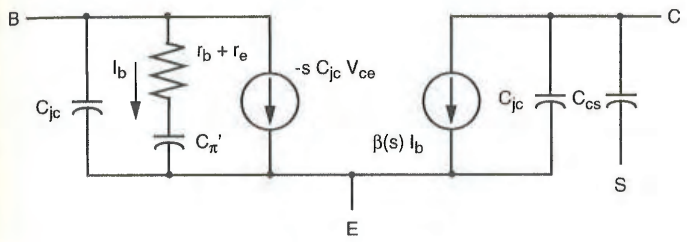


Figure 7.4 Simplified model of bipolar transistor.

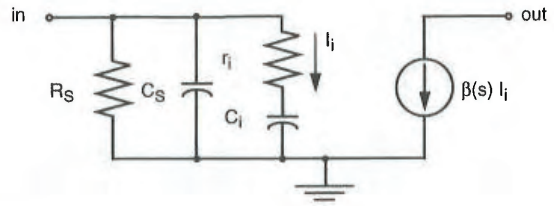


Figure 7.5 Equivalent circuit hybrid-pi model.

the commonly used version, with the exception that feedforward through the collector-base junction capacitance has been neglected. The mapping of the amplifier elements to a bipolar transistor is given as

$$C_s = C_{jc} + C_{source} \tag{7.18}$$

$$R_s = R_{source} \tag{7.19}$$

for input stages and for other stages

$$C_s = C_{jc} + C_{out} \tag{7.20}$$

$$R_s = R_{out} \tag{7.21}$$

where $C_{out} = C_{cs} + C_{jc}$ is the output capacitance of the previous stage and R_{out} is the output resistance of the previous stage. For all stages,

$$C_i = C'_\pi \tag{7.22}$$

or

$$C_i = \frac{C_\pi}{1 + g_m r_e} \tag{7.23}$$

$$r_i = r_b + r_e \tag{7.24}$$

The emitter resistance r_e is typically small and $C_i \approx C_\pi$ and $r_i \approx r_b$.

The ideal feedback source ($-j\omega V_{ce} C_{jc}$) in Figure 7.5 is dealt with separately in the mapping from the multistage amplifier to the equivalent hybrid-π model. Usually, this ideal feedback creates a feedback zero, which helps to stabilize the amplifier. However, if the magnitude of the feedback zero becomes too low, the Q will be reduced and the bandwidth may drop from its optimum value, as discussed in Section 7.2.1. In the terminal stage of an amplifier, the ideal feedback source has the effect of Miller-multiplication of the feedback capacitance without the compensating benefit of pole-splitting. Hence, a large feedback capacitance in the output stage reduces the gain-bandwidth product.

An FET small-signal model can be manipulated into a form similar to the equivalent model in Figure 7.5. The element and node names change, but the structure is identical. This is quite convenient since circuit equations developed for bipolar transistors generally apply equally well to field-effect transistors.

Current-Feedback Pair

One of the most commonly used wideband amplifiers is the current-feedback pair shown in Figure 7.6. Shunt feedback is used to give a low input impedance (typically 50Ω). The output can be taken from either the emitter or collector of Q_2. Taking the output from the collector gives the feedback loop a better degree of isolation from the loading of subsequent stages. A shunt feedback capacitance stabilizes the amplifier by introducing a feedback zero.

In the first stage, the important pole is contributed by the base capacitance of Q_1. The parasitic capacitance at the base of Q_1 is equal to the sum of the source capacitance, feedback capacitance, and collector-base capacitance. The collector-base capacitance has no effect (ideally) on the gain-bandwidth product but, rather, it adds to the shunt feedback capacitance C_F and lowers the feedback-zero frequency. The gain-bandwidth product of the first stage is

$$GB_1 = \frac{\omega_{T_1}}{1 + \dfrac{r_{b_1}}{R_F} + \dfrac{C_s + C_{jc_1} + C_F}{C_{\pi_1}}} \tag{7.25}$$

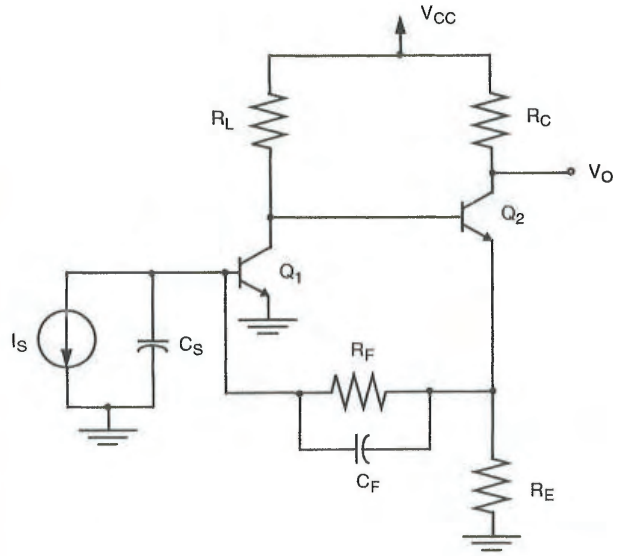


Figure 7.6 Bipolar amplifier current-feedback pair.

where $\omega_{T_1}(= 2\pi f_{T_1})$ is the gain-bandwidth product of transistor Q1. The pin and input bonding pad capacitance can be lumped into the source capacitance C_s.

The form of the second-stage gain-bandwidth product is similar to the first stage. The differences are that there is emitter degeneration by R_E and Miller multiplication of C_{jc2}. Both affect the gain-bandwidth product, which is given as

$$GB_2 = \frac{\omega_{T_2}}{1 + \dfrac{R_E + r_{b_2}}{R_L} + \dfrac{C_x}{C_{\pi_2}}(1 + g_{m_2}R_E)} \tag{7.26}$$

where $\omega_{T_2} = (2\pi f_{T_2})$ is the gain-bandwidth product of transistor Q2 and $C_x = C_{cs_1} + C_{jc_1} + C_{jc_2}\left(1 + \dfrac{R_{cc}}{R_E}\right)$.

Since the current-feedback pair is a two-stage feedback amplifier, the analysis in the previous section can be applied. The overall gain-bandwidth performance is equal to the geometric-mean of the gain-bandwidth product of each of the two individual stages as expressed in (7.16) and is given as

$$(GBW)_2 = \sqrt{GB_1 \cdot GB_2} \tag{7.27}$$

Parasitic effects can limit the useful bandwidth of the circuit. The existence of base ohmic resistance in series with the input capacitance gives one parasitic pole for each stage. These poles do not appreciably affect the magnitude of the open-loop gain characteristic, but they do add excess phase. To compensate for the excess phase, the feedback-zero location must be lowered by increasing C_F. For extremely wideband amplifiers with bandwidths approaching the device f_T, it may be impossible to compensate for the extra phase shift added by the parasitic poles.

Although bipolar transistors are shown, the current-feedback pair can be designed in an FET technology. In a BiCMOS process, better noise performance can be achieved through replacement of transistor Q_1 with a MOSFET [8]. The equations can be adapted to a different technology by correct substitution of device parameters (e.g., C_{gs} for C_π).

Current-Feedback Pair With Darlington

An improved version of the current-feedback pair is shown in Figure 7.7. By replacing transistor Q_2 with a diode Darlington, a significant increase in the gain-bandwidth product can be observed. The diode Darlington configuration can be used with only minor design changes since it adds less than 10 deg of extra phase to the current gain response.

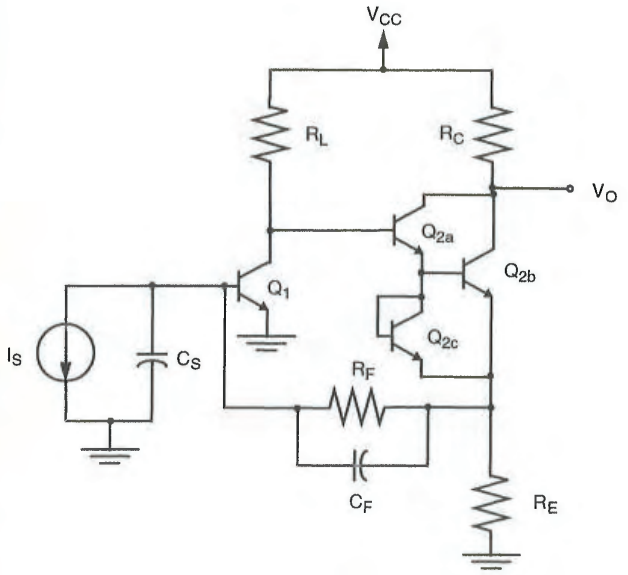


Figure 7.7 Bipolar amplifier current-feedback pair with mirror Darlington.

The diode Darlington has a couple of advantages that help to improve the performance of the current-feedback pair. As shown previously, the compound transistor has a higher current gain than the single transistor. The Darlington configuration has the additional advantage that the extra V_{be} drop in Q_2 provides an extra voltage drop across the collector-emitter of Q_1. The collector-base capacitance of Q_1 is then lowered, which increases the device f_T. The extra voltage drop also allows for a larger output swing.

There is usually little advantage in using a compound in the first stage. The impedance at the input node at high frequencies is fairly low due to the source capacitance and the gain-bandwidth of the stage is more dependent on resistive parasitics. A disadvantage to using a compound stage is the increase in noise. Finally, the advantages of having an extra diode drop in the second stage are lost if the first stage is also a Darlington.

If the excess phase in the current gain of the diode Darlington (due to the pole-zero pair near f_T) can be neglected, then the diode Darlington appears to be identical

to a single transistor with twice the current gain (and twice the input impedance). The gain-bandwidth of the improved second stage is

$$GB_2 = \frac{2\omega_{T_2}}{1 + 2\left[\dfrac{R_E + r_{b_2}}{R_L} + \dfrac{C_x}{C_{\pi_2}}(1 + g_{m_2}R_E)\right]} \tag{7.28}$$

The gain-bandwidth product is doubled but the effects of the parasitics are also doubled. In a situation where parasitics dominate, little advantage is gained by the use of the Darlington compound structure. In practice, however, about 35% improvement in gain-bandwidth can be expected.

The resistive Darlington is also widely used as the second stage in the current-feedback pair [9]. Because of the high collector impedance of Q_1, the Darlington is slightly modified to avoid any Miller multiplication of the junction capacitance of Q_2. (The collector of the input transistor Q_{2a} is connected to V_{CC}.) Unfortunately, current gain near f_T drops off and this, combined with the large phase shift (in the absence of dominant-pole compensation), causes the resistive Darlington/current feedback pair configuration to be less than ideal for applications requiring bandwidths beyond about 10% of f_T.

Voltage-Feedback Pair

While the current-feedback pair provides a low input impedance and a high output impedance, the voltage-feedback pair shown in Figure 7.8 does the reverse. Using series-shunt feedback, the voltage-feedback pair gives a high input impedance and a low output impedance. The circuit approximates an ideal voltage gain stage.

The gain-bandwidth of the stages are

$$GB_1 = \left(\frac{R_{CC}}{R_S}\right)\frac{\omega_{T_1}}{1 + \dfrac{r_{b_1} + R_E}{R_S} + \dfrac{C_{jc_1} + C_F}{C_{\pi_1}}} \tag{7.29}$$

and

$$GB_2 = \frac{\omega_{T_2}}{1 + \dfrac{r_{b_2}}{R_L} + \dfrac{C_{cs_1} + C_{jc_1} + C_{jc_2}\left(1 + \dfrac{R_{CC}}{R_E}\right)}{C_{\pi_2}}} \tag{7.30}$$

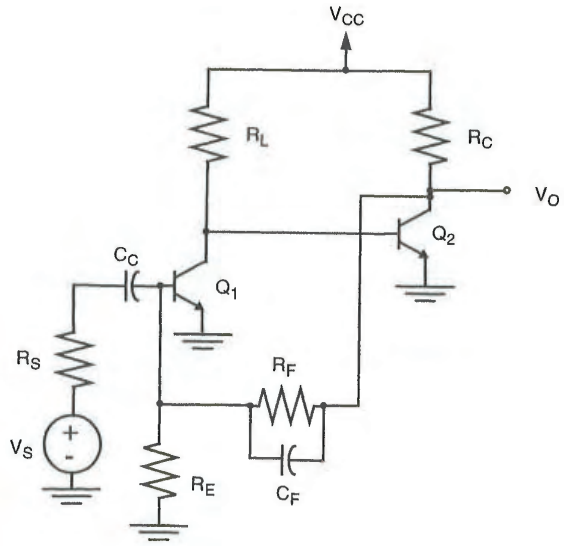


Figure 7.8 Bipolar amplifier voltage-feedback pair.

and, again, (7.16) can be used to find the system gain-bandwidth product.

There are some drawbacks to this design. The frequency response of the voltage-feedback pair is quite sensitive to the source resistance. Unless the source resistance is well-known, and the parasitics at the input are well-controlled, the amount of peaking becomes difficult to control. Additionally, the high input impedance of this configuration makes it unsuitable as an input stage for bandwidths beyond a few hundred megahertz. The high input return ratio can be lowered by applying impedance-matching techniques, as discussed in the next section.

The voltage-feedback pair need not employ bipolar transistors. Because the circuit is not suitable for ultrawideband applications, there is little to gain by using GaAs FETs. However, the circuit is suitable for MOSFETs, JFETs, and even vacuum tube triodes.

7.2.3 Wideband Matched-Impedance Amplifiers

There are advantages to matching integrated circuit terminal impedances as well as designing for matched impedances internal to the chip. Terminal impedances matched

to a specified system impedance (often 50Ω or 75Ω) will prevent standing waves in transmission lines that connect amplifiers separated by many wavelengths. A major advantage to matching impedances on the integrated circuit is to reduce sensitivity to parasitics. Using impedances that are too high makes a circuit sensitive to capacitive parasitics, and using impedances that are too low makes a circuit sensitive to series resistive and inductive parasitics.

Practical Single-Stage Design

The most straightforward way to achieve matching is by using a shunt resistor at the input and output of a high-impedance amplifier. While sometimes used, this brute-force approach reduces gain and power output capability, and degrades noise performance. A better approach to impedance matching is to use shunt feedback (as was used in the current-feedback pair) since it causes only slight degradation in noise performance and power output capability.

A simple matched-impedance amplifier is shown in Figure 7.9. The design uses shunt feedback around a voltage gain stage to achieve the match. Series feedback in the form of emitter degeneration is used to control the gain. The bipolar transistor may be replaced by an FET, though no source resistance should be included in the FET design because the device transconductance is already low and would be further

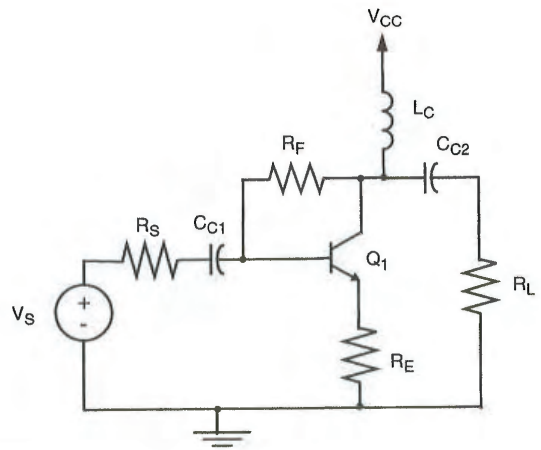


Figure 7.9 Simple matched-impedance bipolar transistor amplifier.

reduced. To achieve reasonable gain, the FET must be run at very high currents (usually tens of milliamps). For this reason, bipolar transistors are usually preferred for amplifiers of this type. GaAs FETs are used only when the widest bandwidth is required.

A small-signal analysis with a simplified model can be applied to the circuit to arrive at the characteristic equations. The effect of finite beta and loading of the feedback network have been ignored. The amplifier is matched if the loop gain of the shunt feedback is unity. The condition for matching is then

$$R_F \approx G_m R_S R_L \tag{7.31}$$

where $G_m = \dfrac{g_m}{1 + g_m R_E} - \dfrac{1}{R_F}$

The gain and gain-bandwidth product under matching conditions are

$$A_v = -\frac{1}{2} G_m R_L \tag{7.32}$$

$$\text{GBW} = \frac{\omega_T}{1 + \dfrac{r_b + R_E}{R_S \| R_F} + \dfrac{C_x}{C_{in}}} \tag{7.33}$$

where $C_x = C_{jc}(1 + G_m R_L)$ and $C_{in} = G_m/\omega_T$.

Two-Stage Matched-Impedance Amplifiers

While one-stage matched-impedance amplifiers are very simple, they often do not provide adequate gain. Two-stage amplifiers offer superior gain for a given bandwidth and have flatter bandpass characteristics. For FET designs, it is even more essential to use two gain stages since the low transconductance of the FET limits the gain that can be achieved in a single stage.

The most straightforward way to create a two-stage matched-impedance amplifier is with a cascade of single stages, as shown in Figure 7.10. The shunt feedback around the first stage is effectively global since it encompasses two fundamental poles (one at the base of Q_1 and the other at the base of Q_2). The global feedback has a loop gain of unity (required for matching) and, as discussed in Section 7.2.1, the optimal gain-bandwidth is difficult to meet for loop gains less than or equal to unity.

Figure 7.11 is a design that is a variation of two cascaded single stages [10]. The design, referred to as the Kukielka configuration, contains a voltage-gain stage (common-emitter/common-collector) with shunt feedback around it. Series-series

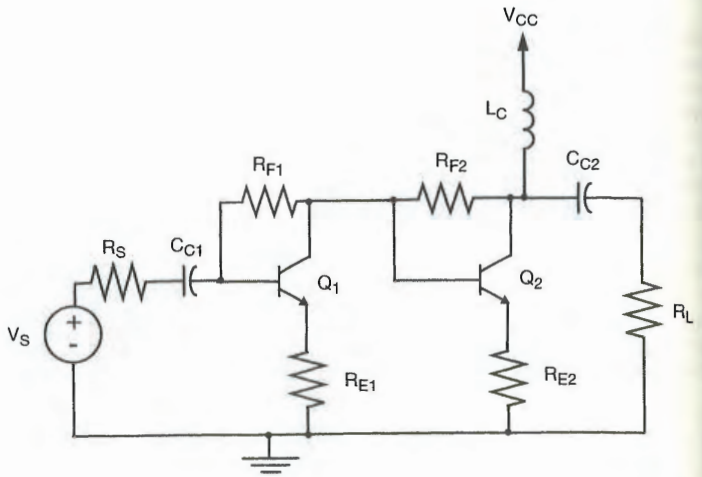


Figure 7.10 Two-stage matched-impedance bipolar transistor amplifier.

feedback, in the form of emitter degeneration in Q_1, gives a high input impedance (in the absence of R_{F1}). Local shunt feedback around Q_2 provides a low input impedance node at its base. Then shunt feedback is applied around this voltage amplifier via R_{F_1} to achieve matching.

The circuit can be designed to achieve optimal gain-bandwidth by appropriate positioning of the poles. As before with the single stage, the loop gain of the shunt feedback must be unity to achieve matching. Therefore, to meet the condition for optimum gain-bandwidth as specified in (7.17), the two open-loop pole positions must be nearly equal for unity loop gain. Under the assumption that $G_{m_2} \gg 1$, $R_{E_1} \gg R_S$ and $R_L C_{jc2} \gg \tau_2$, the condition for equal pole positions is

$$R_{F_2} = G_{m_1} R_S R_L \tag{7.34}$$

The condition for matching is

$$G_{m_2} R_{F_1} = G_{m_1} R_{F_2} \tag{7.35}$$

and the gain when matched is

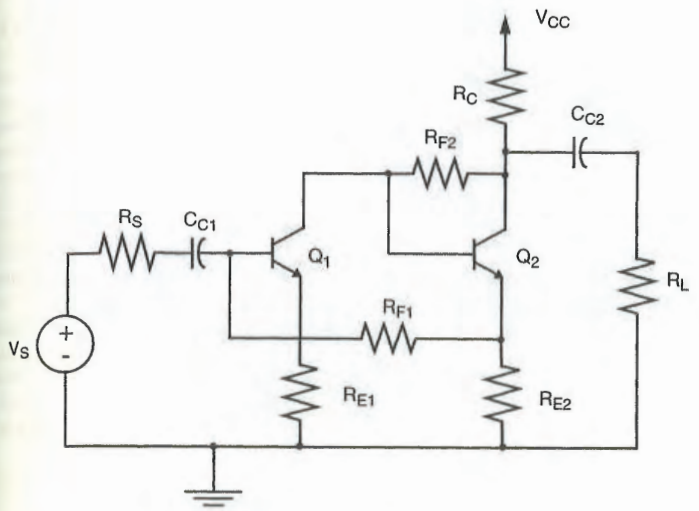


Figure 7.11 Kukielka configuration.

$$A_v = \frac{1}{2} G_{m_1} R_{F_2} \tag{7.36}$$

$$= \frac{1}{2} G_{m_2} R_{F_1} \tag{7.37}$$

Combining the conditions for matching and pole position gives the first-order design equations

$$R_{E_1} \approx \sqrt{R_L R_S} - \frac{1}{g_{m_1}} \tag{7.38}$$

$$R_{F_2} \approx \sqrt{2 A_V R_L R_S} \tag{7.39}$$

$$R_{F_1} \approx \frac{2 A_V}{G_{m_2}} \tag{7.40}$$

The choice of R_{E_2} is unconstrained by these conditions. A smaller value will increase the bandwidth because it reduces sensitivity to parasitics at the base of Q_2. On the other hand, a large value of R_{E_2} will give better noise performance since R_{F_1} will be forced to be larger by the third design equation.

The gain-bandwidth performance of the Kukielka configuration is essentially identical to that of the current-feedback pair. The optimum gain-bandwidth is not easily met because of parasitics. The circuit tends to give an over-damped response since parasitics decrease Q from its maximum value of $\frac{1}{\sqrt{2}}$.

A diode Darlington can replace Q_2. The extra phase shift will help raise the Q of the circuit and increase the bandwidth. Use of the resistive Darlington has also given good results [7].

Another interesting two-stage variant first proposed by Meyer [11] is depicted in Figure 7.12. The circuit can be viewed as a hybrid between the voltage-feedback pair and the current-feedback pair. Like the Kukielka configuration, the Meyer configuration is a voltage-gain stage (a voltage-feedback pair in this case) with shunt feedback around it for matching. The ability to adjust the circuit Q through use of a feedback zero or by design is a significant advantage of this design.

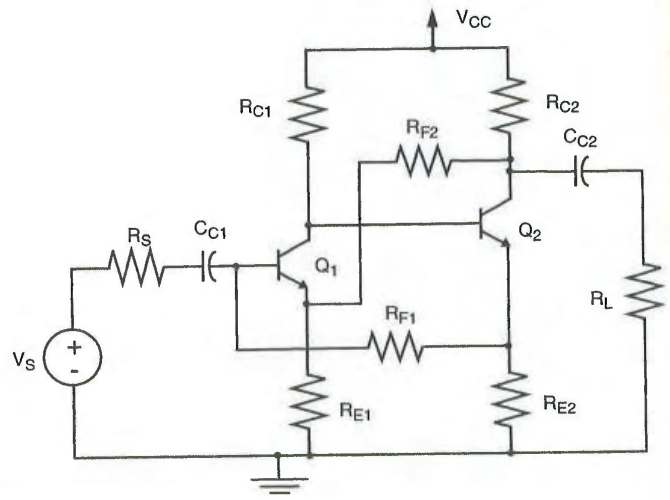


Figure 7.12 Meyer configuration.

7.3 LOW-NOISE AMPLIFIERS

Noisy components in the beginning of the receiver chain have the most significant impact on the overall noise performance. For this reason, a large emphasis is placed on low-noise operation of the first RF amplifier, also referred to as the low-noise amplifier (LNA).

The design of an LNA involves numerous tradeoffs. The amplifier must have sufficient gain to overcome mixer noise, but not so much to cause mixer overload. Good noise characteristics are desired while achieving desired input and output matching. The matching networks at the input and output of the LNA are often a compromise between optimum gain matching and optimum matching for noise figure. Selecting a technology and package involves tradeoffs of cost and level of integration.

The apparent simplicity of an LNA is misleading—the design should be easy because there are relatively few components but the tradeoffs complicate the design. A handful of guidelines are provided in the next section that may facilitate the design. The design of an LNA commonly used in wireless systems is presented.

7.3.1 Guidelines for Low-Noise Amplifier Design

Noise figure (NF) is the most common method used to specify the noise performance of an amplifier. Noise figure is defined as ten times the logarithm (base 10) of the noise factor, F. The noise factor is defined as the ratio of the total output noise power of an amplifier to the output noise power due to the source alone.

A two-port network with noise sources can be represented by the same network with the noise sources removed and with a noise voltage $\overline{v_t^2}$ and current generator $\overline{i_t^2}$ connected at the input. The noise factor can be expressed in terms of these equivalent input noise generators. It is

$$F = 1 + \frac{\overline{v_t^2}}{4kTR_s\Delta f} + \frac{\overline{i_t^2}}{4kT\frac{1}{R_s}\Delta f} \tag{7.41}$$

The design of an LNA is strongly influenced by the source resistance, R_S. For high source resistance, $\overline{i_t^2}$ is the dominant equivalent noise source and, conversely, for low source resistance, $\overline{v_t^2}$ is the dominant equivalent noise source. When the source resistance is large, FET devices offer superior performance to bipolar devices at low frequencies because the noise-current generator is determined by gate leakage current, which is small. The noise-voltage generator of a bipolar device is typically smaller than that of an FET, so for low source impedances, a bipolar transistor often has noise performance superior to that of an FET.

From (7.41), a value of R_S giving minimum F can be found. It is

$$R_{S\text{opt}} = \frac{\overline{v_i^2}}{\overline{i_i^2}} \tag{7.42}$$

The optimum source resistance is unlikely to be the actual source resistance (typically, 50Ω or 75Ω), so a matching network is implemented using integrated or off-chip passive components. (The package parasitics can be incorporated in the matching network if the package has been well-characterized.) Often, the optimum source resistance is much larger than the actual source resistance and the matching network cannot easily be realized. Therefore, minimum noise figure cannot be achieved.

For a bipolar transistor, the optimum source impedance and the minimum noise factor are given as

$$R_{S\text{opt}} = \frac{\sqrt{\beta}}{g_m}\sqrt{1 + g_m r_b} \tag{7.43}$$

$$F_{\min} \cong 1 + \frac{1}{\sqrt{\beta}}\sqrt{1 + g_m r_b} \tag{7.44}$$

Increasing the current will increase g_m and decrease the optimum source impedance. Optimum matching conditions are more likely achievable, however, the noise figure increases from its minimum value as the current increases. Advantages for both optimum source resistance and minimum noise figure come from using larger input devices. The base resistance r_b decreases, thereby reducing the optimum source impedance and lowering the minimum noise figure.

For a FET transistor, the optimum source impedance and the minimum noise factor are given as

$$R_{S\text{opt}} \cong \frac{1}{\omega C_{gs}} \tag{7.45}$$

$$F_{\min} \cong 1 + \frac{4}{3}\frac{\omega C_{gs}}{g_m} \tag{7.46}$$

It has been assumed that the frequency of operation is such that flicker noise and gate leakage current are negligible. Choosing a larger device will increase C_{gs} and reduce $R_{S\text{opt}}$. (Note that at low frequencies, $R_{S\text{opt}} \to \infty$.) The effect on minimum noise of a larger device is negligible since both capacitance and transconductance

increase. Biasing the input transistor at high currents reduces the minimum noise figure because of an increase in the transconductance.

In general, whether FET or bipolar technology, larger input transistors biased at large drain (or collector) currents are commonly used in low-noise amplifiers. Often, the designs include just the single transistor with input and output matching networks because fewer active components means fewer noise sources. For ultralow noise applications, the gate (or base) is biased through an RF choke.

Feedback is used in many LNAs to set bias and achieve matching. Shunt-resistive feedback is commonly used though the resistor adds noise at the input. If the feedback is minimal, as in [12], the noise performance is not degraded by much.

The use of reactive feedback for matching is preferable. A technique used in some LNA designs [13] is adding inductance (e.g., from the bond wire) at the emitter of a bipolar transistor. The small-signal input impedance becomes small at high frequencies due to the shunting effects of the input capacitance C_π and the Miller effect due to C_μ. Ultimately, the input impedance approaches the base resistance, which is small (~10Ω), because a large input device is used. Emitter inductance can increase the input impedance by the amount $Z_i = \omega_T L_E$. An input match of 50Ω can be achieved without raising the noise figure. The practical design discussed in the next section uses this technique.

7.3.2 Practical Design of a Low-Noise Amplifier

A low-noise amplifier design that uses a cascode arrangement as shown in Figure 7.13 offers performance advantages in wireless applications over other configurations. A single common-emitter amplifier will typically not provide the gain required to overcome the noise figure of successive stages. A cascade of two common-emitter stages may be used, however, the required dc bias current on the second stage will have to be quite high to handle amplified versions of the input blocking signal (unless additional filtering is used between the two stages). A cascode LNA provides high gain and low noise. Also, the cascode is a more power-efficient approach, as the same bias current flows through both Q_1 and Q_2.

The gain of the cascode amplifier is higher than a common-emitter stage because Miller feedback (due to the collector-base junction capacitance of Q_1) is eliminated. The cascode arrangement limits the voltage gain from the base to the collector of Q_1 to unity. The effect is to raise input and output impedances without lowering the current gain. Hence, increased power gain can be obtained.

An additional advantage of the cascode is its ability to handle larger supply voltages. A common-emitter stage can generally be operated only up to the *collector-emitter* breakdown voltage of the process (typically only 4V in a 25-GHz f_T process). However, the cascode arrangement with its common-base output can be operated up to the *collector-base* breakdown of the process (typically greater than 10V for a 25-GHz f_T process).

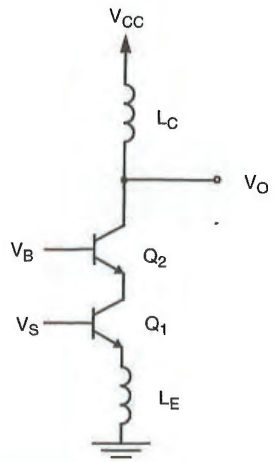


Figure 7.13 Cascode bipolar transistor amplifier.

The cascode arrangement has three main disadvantages. First, the available output-voltage swing is decreased by a diode drop (or at least a collector-emitter saturation voltage). This loss does not limit the cascode performance much; the circuit can still be operated on voltage supplies as low as 1.8V (2 AA cells at end-of-life) for a bias current around 4 mA and a collector impedance of 200Ω. The collector impedance should be no more than ~200Ω to easily match to an external 50Ω impedance.

The second disadvantage of the cascode is a somewhat higher noise figure than a cascade of two common-emitter stages. For example, an achievable noise figure at 900 MHz for a cascode amplifier is 1.5 dB compared to the cascaded common-emitter with 1.2-dB noise figure.[1] For systems where noise is more critical than blocking performance, such as GPS, a cascade arrangement is preferable to the cascode. For handheld cellular systems, the cascode offers an excellent tradeoff between gain, noise, and blocking performance.

The third main disadvantage of the cascode is that the input 1-dB compression point and third-order intercept point will be lower than that of a single common-emitter stage because of the increased gain.

An example of a practical circuit that uses a cascode is shown in Figure 7.14. The circuit splits the signal current between Q_3 and Q_5. Transistor Q_3 is much larger

[1]The process for comparison has f_T = 2.5 GHz.

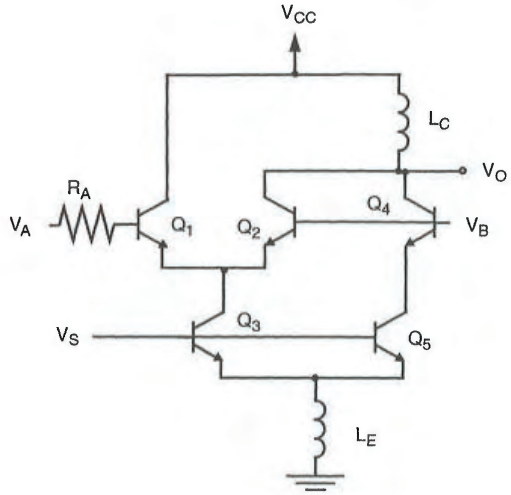


Figure 7.14 Low-noise bipolar transistor amplifier.

than Q_5 and carries most of the signal current. When the voltage V_A is low, the signal current in Q_3 flows through Q_2 and to the output. This is the high-gain state of the amplifier. When V_A is high, all of the signal current in Q_3 is diverted from Q_2 to Q_1 and only the signal current that flows from Q_5 through Q_4 makes it to the output. This is the low gain state [14].

The ability to change gain stages is quite useful for cordless and cellular systems. The feature prevents overload of the mixer and subsequent IF stages in the presence of a large input signal. For example, typical radio systems can have input signals as high as −15 dBm incident at the antenna. If the LNA has 15-dB gain, the mixer would be required to handle 0 dBm of signal (neglecting filter losses). However, with a gain reduction in the LNA from 15 dB to 3 dB (a 12-dB step), the mixer need only handle −12 dBm at the input (a much easier design task). In addition, the voltage swing at the collector will be reduced in the low-gain setting. Furthermore, different gain stages in the LNA reduce the required gain-control range of the variable-gain amplifiers.

7.4 VARIABLE-GAIN AMPLIFIERS

In a variable-gain amplifier (VGA), an applied dc control voltage causes the amplifier gain to change. The gain varies monotonically with increasing (or decreasing) control

voltage. A common application of a VGA is within an automatic gain control (AGC) environment such as used in receivers. In many systems, the receiver output signal strength is monitored. A control voltage, which is a function of the output signal strength, is fed back to the VGA to vary the gain for a constant receiver output. Digitally controlled VGAs are becoming increasingly popular. Rather than varying over a continuum of gains, the amplifiers take on either a high-gain or low-gain state. Cascading several VGAs will give finer increments of gain variation for the receiver chain.

The vast number of designs can be classified in a handful of circuit configurations. Most VGAs can be categorized as either emitter-driven, where a current signal is injected at the emitters of diff pairs (differential pairs), or base-driven, where a voltage signal is applied to the bases of the diff pairs. In emitter-driven designs, current is steered between pairs of diff pairs, called a quad pair. Unlike emitter-driven designs, which are inherently linear, local feedback is used in base-driven designs to linearize the circuit. Gain is typically varied in these designs by effectively altering the feedback.

VGAs used in wireless communication systems typically operate at IF signal frequencies and must attain a large gain-control range (30 to 40 dB). Currently, the technology most often used in commercial systems is silicon bipolar and, consequently, is emphasized in this section. Several of the designs can easily be translated to FET technologies. Some attention is given to BiCMOS, which is the technology probably best suited to the design of variable-gain amplifiers; the bipolar devices have good bandwidth and gain characteristics while the FETs can be used as variable resistors.

How well a circuit can handle large and small signals is a measure of its dynamic range. The dynamic range of the emitter- and base-driven circuits are explored first. Drawing from the results of the emitter-driven circuit, three VGA designs are compared. The last part of this section discusses the merits of several gain-control circuits (not restricted to emitter-driven and base-driven configurations).

7.4.1 Dynamic Range

An important performance metric of a variable-gain amplifier is its dynamic range. A definition for dynamic range is the ratio of maximum allowable input signal to minimum detectable signal. The minimum detectable signal is the signal at which the signal-to-noise ratio (SNR) is deemed acceptable (typically $\cong$ 10 dB). The maximum allowable input signal is often limited by distortion.

In gain-control amplifiers, the output SNR is a more useful measure of circuit noise than the noise figure because the noise figure depends on the circuit gain, which is variable. At maximum gain, the amplitude of the input signal reaches its lowest value but it does not imply that the SNR is at a minimum. In many circuits

this is the case, but often the minimum SNR is found at half-gain when all devices in the amplifier are on and contributing to the noise level. For a specific minimum requirement in SNR over the whole dynamic range, only the amount of noise at the worst case (e.g., either maximum gain or half-gain) has to be examined.

Of particular concern in variable-gain amplifiers used in wireless communication systems is third-order intermodulation distortion. Intermodulation distortion can severely degrade receiver performance because unwanted signals falling within the RF passband encounter nonlinearities in the circuit. A third-order product can be created that for all practical purposes appears to be a legitimate signal because it falls within the desired signal bandwidth.

There is no single satisfactory measure of intermodulation performance. The ratio of the third-order intermodulation product magnitude to the fundamental product magnitude is third-order intermodulation distortion (IM_3). The problem with this is that the value specified for IM_3 is signal-level dependent. Instead, the concept of intercept point is widely used as a measure of amplifier performance. The intercept point is signal independent, but it is a theoretical not actual measure of performance.

The third-order intercept point (IP_3) is the power level (either input or output is specified) at which the extrapolated curves of the linear response and the third-order intermodulation product intersect [15]. Extrapolation is necessary since as the amplifier approaches compression, the actual distortion curves no longer follow the predicted cubic behavior. Another problem with the intercept point is that it only has meaning for amplifiers that can be described as a cubic system. The power in the third harmonic does not increase at three times the rate of that of the fundamental in all systems.

7.4.2 Emitter-Driven and Base-Driven Gain Control

Often the linearity and noise performance of a gain-control amplifier are constrained by the driver design. Two possibilities are the base-driven and emitter-driven designs [16]. In the base-driven design of Figure 7.15(a), the signal is applied to the bases of the diff pair Q_1 and Q_2. A gain-control voltage V_C varies the amount of tail current I_{EE} flowing through the pair. Interchanging the positions of input signal and gain control of the base-driven circuit results in an emitter-driven circuit, shown in Figure 7.16(a). The gain-control voltage steers the current between Q_1 and Q_2.

Intermodulation distortion performance of the emitter-driven circuit is superior to the base-driven circuit. The output current of the base-driven circuit is a nonlinear function (tanh) and thus, the output is distorted with respect to the input. On the other hand, in the emitter-driven circuit the output current is linearly related to the input signal current i. Thus, to first-order, no distortion is produced by the diff pair in the emitter-driven circuit.

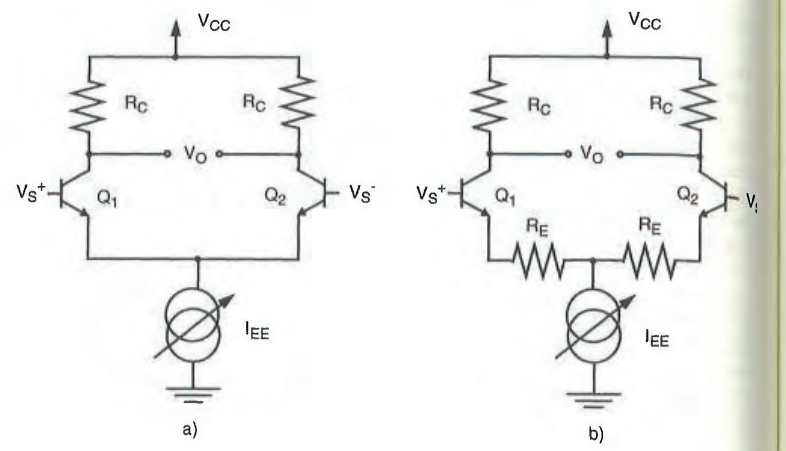


Figure 7.15 Base-driven transistor amplifier: (a) basic circuit approach, (b) practical design.

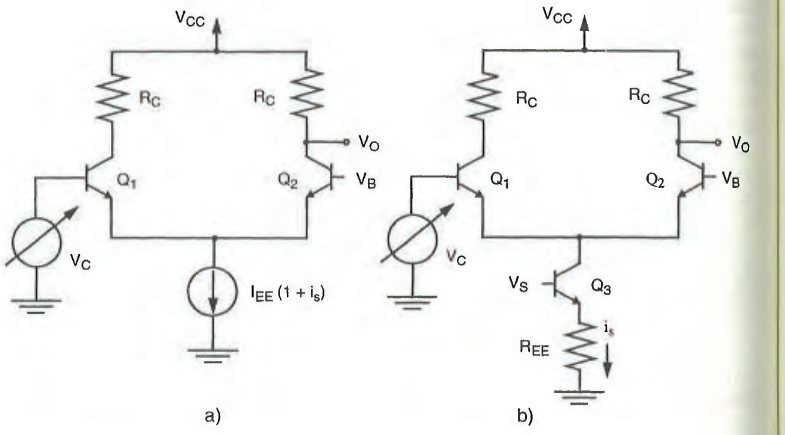


Figure 7.16 Emitter-driven transistor amplifier: (a) basic circuit approach, (b) practical design.

More practical versions of emitter- and base-driven circuits are shown in Figures 7.15(b) and 7.16(b), respectively. When the practical versions are compared, the results are not as dramatic, though the performance of the emitter-driven circuit still exceeds that of the base-driven circuit. Nonlinearities in the emitter-driven pair introduced as the input voltage v_s are converted to current i_s. Since the distortion performance of the base-driven circuit is not satisfactory for most wireless communication systems, the linear range is extended by the presence of emitter and base resistors, either inherent in the transistor or added intentionally. These resistors introduce local negative feedback. The output currents flow in these resistors and are converted by them into voltages that subtract from the input voltage. The smaller effective input signal reduces the distortion.

At low frequencies, a power series can be used to estimate the amount of distortion for a given output signal level [17]. For the base-driven circuit with emitter degeneration,

$$\mathrm{IM}_3 = \left(\frac{v_o}{I_{EE}R_L}\right)^2\left(\frac{1}{1+\dfrac{I_{EE}R_E}{2V_T}}\right) \tag{7.47}$$

The effect of emitter degeneration is a reduction in third-order intermodulation of approximately the loop gain, A_L, of the circuit. For a given input, the third-order intermodulation reduces by $(1 + A_L)^3$, which appears as a significant reduction in distortion; however, the gain is reduced by $(1 + A_L)^2$. Distortion in the emitter-driven circuit is

$$\mathrm{IM}_3 = \left(\frac{v_o}{I_{EE}R_L}\right)^2\left[\frac{\left(1-\dfrac{2I_{EE}R_E}{V_T}\right)}{8\left(1+\dfrac{I_{EE}R_E}{V_T}\right)^2}\right] \tag{7.48}$$

It has been assumed that the circuit is operating at maximum gain and Q_2 is acting as a cascode. Equations (7.47) and (7.48) are expressions for third-order intermodulation distortion (the ratio of the third-order intermodulation product and the fundamental component). The output third-order intercept point is found by setting $\mathrm{IM}_3 = 1$, solving for v_o, and converting the resulting signal voltage to an output power level. A numerical example shows that the emitter-driven circuit has better distortion than the base-driven with emitter degeneration. An output signal of 325 mV can be tolerated in the emitter-driven circuit before IM_3 reaches 1%. When the base-driven circuit reaches 1% third-order intermodulation, the output

is only 110 mV. These calculations were arrived at using $R_L = 250\Omega$, $I_{EE} = 2$ mA, and $R_E = 100\Omega$.

When noise is compared in these two circuits, it is found that again the emitter-driven circuit has better performance than the base-driven circuit with emitter degeneration. Dominant noise sources for both circuits are the base resistance and collector current noise of transistors Q_1 and Q_2. However, the emitter resistors used for linearization further increase the peak noise in the base-driven circuit. The input noise voltage for the emitter-driven circuit is

$$v_i = \sqrt{4kT\Delta f\left(2r_b + \frac{2V_T}{I_{EE}}\right)} \tag{7.49}$$

and for the base-driven circuit it is

$$v_i = \sqrt{4kT\Delta f\left(2r_b + \frac{2V_T}{I_{EE}} + 2R_E\right)} \tag{7.50}$$

Another numerical example will illustrate the noise performance advantage of the emitter-driven circuit. With the parameter values used in the distortion example and $r_b = 35\Omega$, $\frac{v_i}{\sqrt{\Delta f}} = 1.3nV/\sqrt{\text{Hz}}$ and $2.2nV/\sqrt{\text{Hz}}$ for the emitter-driven and base-driven designs, respectively. The output noise of a base-driven circuit with emitter degeneration may actually be lower than the emitter-driven circuit because the gain is less in the former.

7.4.3 Basic Circuit Configurations

Three variable-gain amplifiers commonly used as AGC circuits are referred to as the *AGC amplifier, multiplier*, and as *Gilbert's quad*. Each of these designs offers wideband performance and wide dynamic range operation. Input and output ports are differential and input and output impedances are independent of amplifier gain. A summary of the comparisons are provided in this section. A more detailed analysis can be found in [18].

The AGC amplifier and the multiplier are realized using the emitter-driven pair described in the previous section. They are formed by balancing two emitter-driven circuits. In the AGC amplifier (Figure 7.17), the control voltage V_C causes the currents in Q_4 and Q_5 to rise or fall together. Maximum gain (no attenuation) occurs at a large positive control voltage and causes all the tail current I_{EE} to flow through these transistors. Minimum gain (maximum attenuation) occurs at a large negative control voltage.

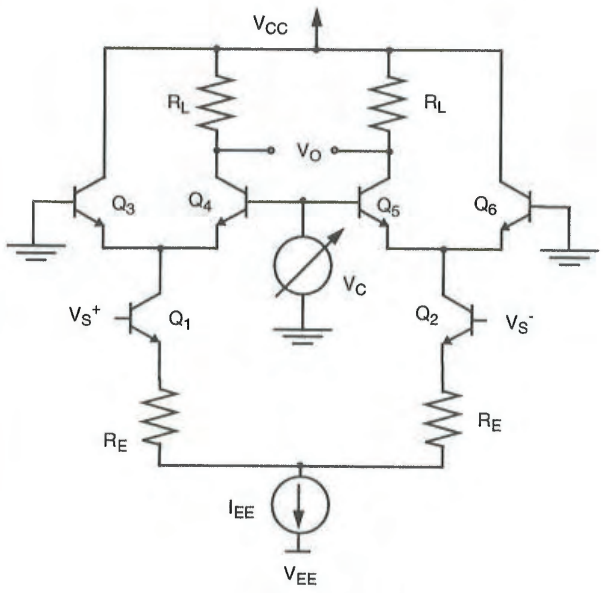


Figure 7.17 AGC transistor amplifier.

There are only slight differences between the AGC amplifier and the multiplier (Figure 7.18). The outputs of the multiplier are cross-coupled and maximum attenuation occurs when $V_C = 0$, at which point the currents through the quad pair (Q_3–Q_6) are equal. Minimum attenuation occurs at both large positive and large negative values of the control voltage.

A different variable-gain quad is based on Gilbert's wideband amplifier technique [1]. The quad is shown in Figure 7.19. Under ideal conditions, there exists a perfectly linear relationship between the input current i_i and the output current i_o. A technique known as *predistortion* is applied to achieve the linear relationship. The nonlinear relationship of current to voltage in transistors Q_1 and Q_2 is complimentary to the nonlinear relationship of voltage to current in transistors Q_3 and Q_4. Thus, ideally, a linear relationship between input and output currents exists. The nonlinearity in the circuit is introduced by converting the input voltage to the input current.

For the AGC amplifier and multiplier circuit configurations, gain is varied by steering the tail current away from some quad transistors and towards others. In

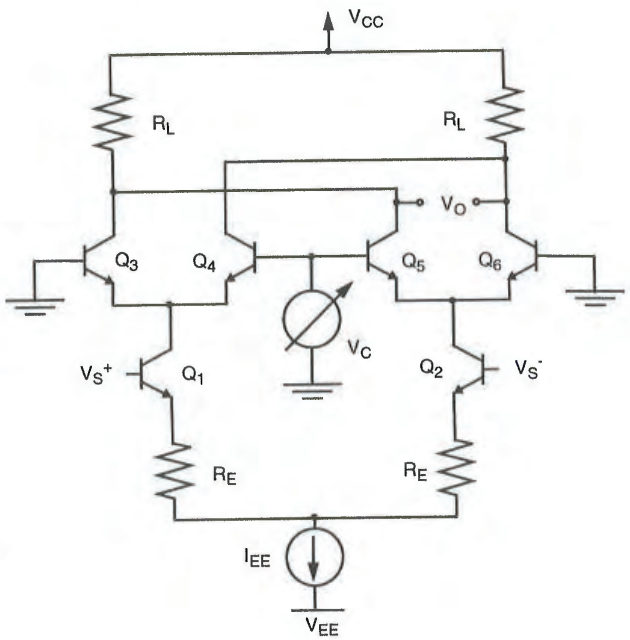


Figure 7.18 Bipolar multiplier.

Gilbert's quad, gain variation is achieved by varying the ratio of the tail currents I_{E1} and I_{E2}. The voltage gain of each configuration is summarized:

$$A_{v_{\text{AGC}}} = \frac{R_L}{R_E}\left(\frac{1}{1 + \exp - \dfrac{V_C}{V_T}}\right) \tag{7.51}$$

$$A_{v_{\text{mult}}} = \frac{R_L}{R_E}\left(\frac{1 - \exp - \dfrac{V_C}{V_T}}{1 + \exp - \dfrac{V_C}{V_T}}\right) \tag{7.52}$$

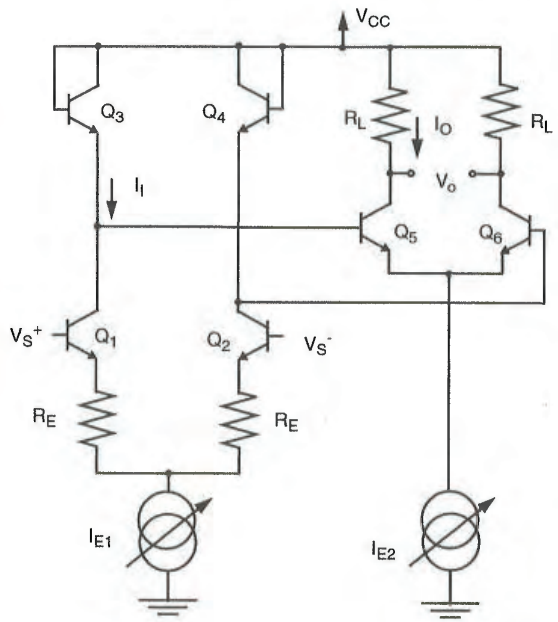


Figure 7.19 Gilbert's quad.

$$A_{v_{\text{quad}}} = \frac{R_L}{R_E}\frac{I_{E2}}{I_{E1}} \tag{7.53}$$

The mechanisms of distortion in the variable-gain amplifiers are the same as in the emitter-driven circuit. Distortion occurs due to base and emitter resistance in the quad devices, causing deviation from exact exponential characteristics. Linearity is preserved if the voltage drops across the base resistors are equal and the voltage drops across the emitter resistors are equal. The effect is observed as a null in the distortion at specific attenuation levels. A null exists at half-gain in the AGC amplifier and multiplier. When the tail currents I_{E1} and I_{E2} of Gilbert's quad are equal, distortion is significantly reduced.

In practice, the AGC amplifier is found to have the lowest distortion. The performance at low attenuation is limited by the nonlinearity introduced by the base resistance. The performance of Gilbert's quad is affected more by the base resistance

[16] than the AGC amplifier. At high attenuation, the multiplier theoretically has the best performance because the currents in all transistors are nearly equal and the effect of base resistance is reduced because of more equal voltage drops. However, the imbalances in a practical multiplier circuit implementation cause the fundamental to cancel at a different point than the distortion. Thus, distortion increases significantly at high attenuation and becomes worse than the AGC circuit.

The noise results of the emitter-driven pair presented in the previous section can be applied to the three variable-gain circuits. The total output noise power in the AGC amplifier and multiplier is twice that for the emitter-driven pair. The noise peaks at half-gain (6-dB attenuation) in the AGC circuit and decreases for both increasing and decreasing attenuation. Though the multiplier has the same peak noise power as does the AGC amplifier, the noise does not decrease for higher attenuation, and so has the disadvantage that the worst SNR is the value obtained over most of the attenuation range. In Gilbert's quad, the maximum output noise voltage due to base resistance equals $2\sqrt{2}$ times the noise peak for the other circuits if base resistance and quad current are assumed to be equal. Peak noise occurs at maximum gain and decreases with increasing attenuation.

The choice of gain-control circuit configuration is dependent on application. For large gain-control variation (~40 dB), the AGC amplifier is found to have the greatest dynamic range performance. The multiplier and quad have good performance for small gain-control ranges (~10 dB) and offer some advantages. The multiplier has a constant dc output voltage and polarity inversion of the control voltage, which is attractive in some applications. Maximum distortion at high frequencies in Gilbert's quad increases only linearly with frequency whereas in the AGC amplifier, the maximum distortion increases twice as fast. For high frequencies, Gilbert's quad has a better distortion performance than the AGC amplifier. However, direct feedthrough in Gilbert's quad is worse than in the AGC amplifier.

7.4.4 Variable-Gain Circuit Configurations

Variable-gain amplifiers used in wireless systems are not restricted to the class of designs based on the emitter-driven circuit. Many designs extend the dynamic range of the base-driven circuit by employing variable resistance. Current trends in VGA design include the use of FETs as variable resistors in the feedback path of the circuit. A sampling of these designs is now discussed. The first designs are improvements to the AGC amplifier and Gilbert's quad.

A proposed improvement [18] to the AGC amplifier increases the current in the lower differential pair and decreases the current in the upper quad transistors. The circuit is shown in Figure 7.20. More current through the degeneration resistors improves distortion. Noise will also improve. The optimum dynamic range and flat transfer characteristics can only be realized at fairly low quad currents. Additional dynamic range at high attenuation is achieved through proper sizing of the quad

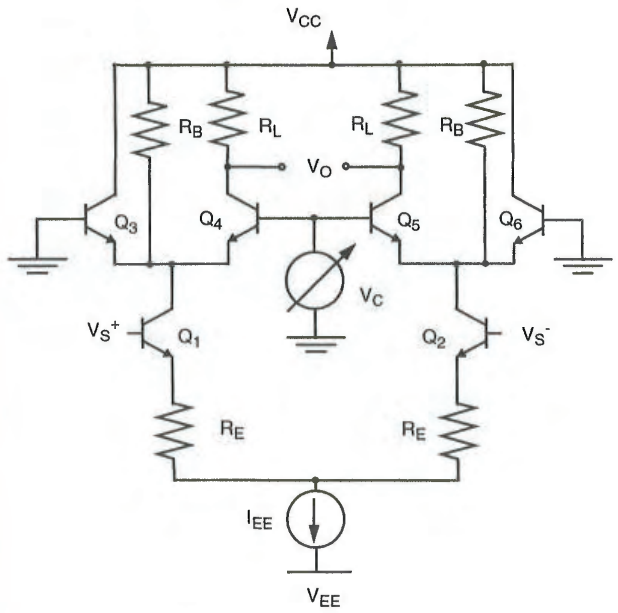


Figure 7.20 Improved AGC circuit.

devices. Making Q_3 and Q_6 a matched pair of large devices minimizes r_b and creates a more ideal exponential junction relation. The other quad devices, Q_4 and Q_5, should be a matched pair of small devices to minimize feedthrough.

Feedback is often used to linearize a circuit at the expense of increased circuit noise. This was observed in the base-driven circuit. Alternative techniques, such as *predistortion* used in Gilbert's quad, achieve linearity by canceling circuit nonlinearities with complimentary nonlinearities. Another example of this is a variation on Gilbert's quad shown in Figure 7.21 [1].

The circuit can achieve a slightly larger gain-bandwidth product. The current gain for this circuit is $A_v = 1 + \frac{I_E}{I_B}$.

Further advancing an alternative to emitter degeneration is the *multitanh* doublet shown in Figure 7.22 [19]. It is comprised of two differential pairs with their inputs and outputs in parallel. The junction area of one emitter in each pair is

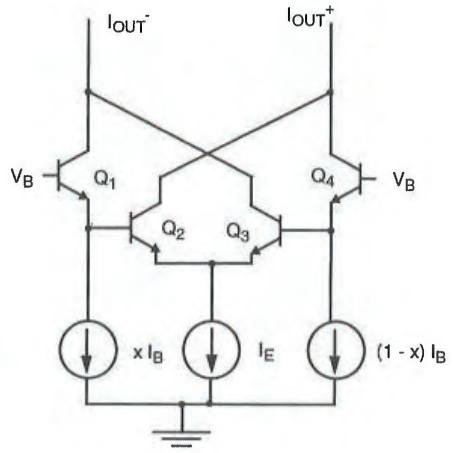


Figure 7.21 Variation on Gilbert's quad.

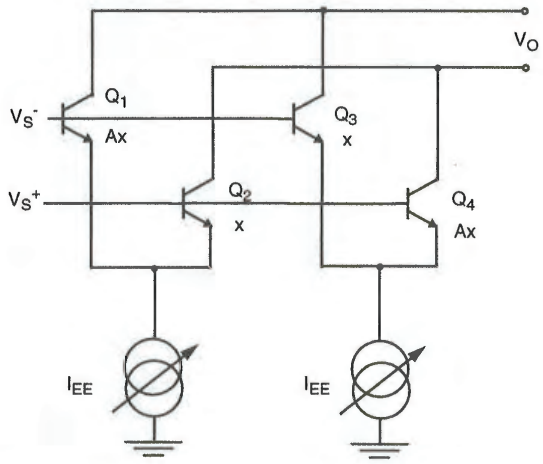


Figure 7.22 Multitanh doublet.

arranged to be A times larger than the opposite emitter. The gain of the circuit is the gain of a basic differential pair reduced by $\frac{4A}{(1+A)^2}$ where A is the ratio of area of Q_1 to Q_2 and Q_4 to Q_3. The gain can be varied by simultaneously increasing or decreasing the equal tail currents of the pairs.

The effect of the emitter area ratio is to shift the peak of each transconductance by an equivalent offset voltage, resulting in an overall g_m that is much flatter than the simple diff-pair. This leads to improved linearity. There is an optimum value of A; if too low, the g_m curve is still a hump and if too high, there is a double hump. For best performance, A is taken to be equal to 4. The gain is thus 0.64 times that of the basic diff pair. Even though the input-referred noise of the multitanh doublet is 25% higher than the basic differential pair, the dynamic range is much better. There is also a significant improvement in dynamic range of the multitanh doublet over the differential pair with emitter degeneration.

The multitanh is a base-driven circuit that uses transconductance offset to linearize the circuit. The next few circuits revisit the idea of base-driven variable-gain circuit configurations. In the dual-gain-state differential amplifier shown in Figure 7.23 [20], current is switched between two diff pairs with connected bases and collectors. The pairs are degenerated differently with emitter resistors R_L and R_H of low and high values, respectively. At maximum attenuation, all current flows through R_H, and conversely for minimum attenuation, all current flows through R_L. Therefore, using this topology, the lowest gain state allows a large undistorted input signal, while the highest gain state has a lower input noise, resulting in a very large dynamic range. This particular design is an example of a digitally controlled VGA as discussed earlier.

A similar approach to achieving wide dynamic range as was used in the previous design is used in the diode-bridge variable-gain amplifier shown in Figure 7.24 [21]. It is another base-driven design that operates with small emitter resistance for high-gain, low-noise operation and large emitter resistance for low-gain, low-distortion operation. Unlike the dual-gain-state amplifier, there is only one diff pair and the emitter resistance is varied by controlling the current, I_C, through the diode bridge. For minimum current, the diode bridge with the additional resistors R_D has a very high impedance, roughly 10 times greater than $2R_E$. Therefore, the amplifier has minimum gain for minimum I_C. With maximum current, the diode bridge has minimum impedance about four times smaller than $2R_E$; therefore, the amplifier has maximum gain.

A BiCMOS technology is desirable when designing a variable-gain amplifier. The bipolar devices can be used to achieve high gain and the FETs can be used as resistors in feedback paths. For example, the diode-bridge network in the circuit just described can be replaced by a single MOSFET. The control voltage can be applied to the gate voltage to vary the transconductance of the transistor. Another

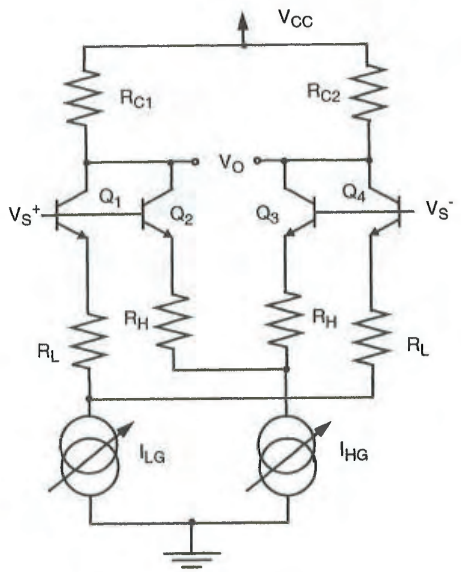


Figure 7.23 Dual-gain-state differential amplifier.

BiCMOS variable-gain amplifier is shown in Figure 7.25 [22]. FETs are used as local shunt feedback in the first and second stage. The advantage of this design above the other designs that vary feedback is that the input impedance remains constant over gain variation. Local shunt feedback with a high impedance load makes the input impedance independent of the feedback. Dynamic range in this circuit is achieved by low base resistance and high collector current in the first stage. Increasing current along the signal path will help to minimize distortion.

7.5 POWER AMPLIFIERS

Power amplifiers (PAs) are in the transmitting chain of a radio in a wireless system. They are the final amplification stage before the signal is transmitted, and therefore must produce sufficient output power to overcome transmission loss between the transmitting and receiving units. The amount of output power is highly dependent on the application cell size (how closely units are arranged), ranging from milliwatts in indoor picocell applications to watts for broader regional coverage.

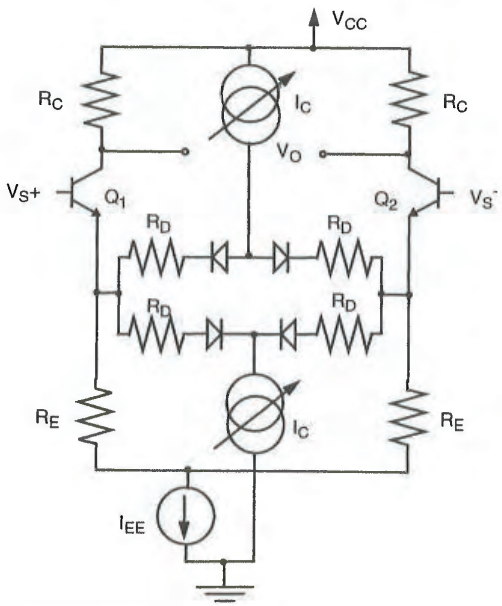


Figure 7.24 Diode-bridge gain control.

The PA is the primary consumer of power in the radio portion of a wireless unit. Thus, a major design issue is how efficiently the PA can convert dc input power to RF output power. Another important characteristic of a PA is linearity; the input-output relation must be linear to preserve signal integrity. A primary consideration in the design of a PA involves tradeoffs of efficiency and linearity.

The choice in system and corresponding modulation method defines the RF waveform and directly affects the linearity requirements of the PA. For highly linear applications, efficiency may be sacrificed. Modulation methods (such as DQPSK in direct sequence systems) that have waveforms with high peak-to-average ratios must have greater linearity. Often, this is at the expense of efficiency. On the other hand, there are modulation techniques (such as GFSK in frequency-hopping systems) that result in more constant envelope signals, allowing the PA to operate in saturation (a nonlinear regime) so higher efficiency can be achieved.

Power amplifiers can generally be classified by mode of operation. A single power device operates continuously in a Class A amplifier. In Class B designs, the

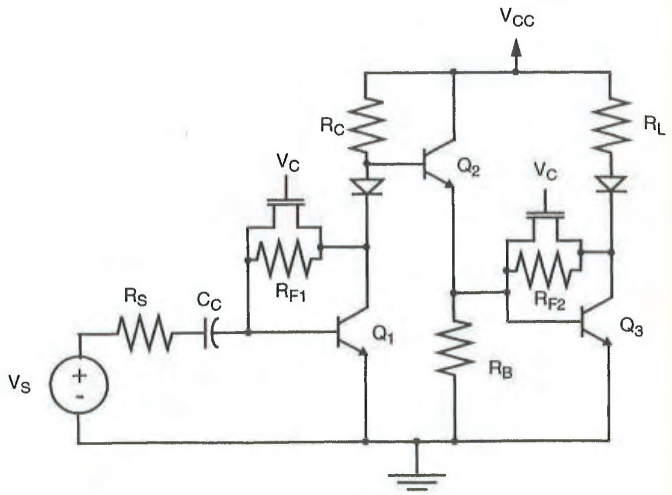


Figure 7.25 BiCMOS RF variable-gain amplifier.

power device is operating for half of the time. A common implementation of a Class B design is in a push-pull fashion, where two power devices operate half of the time in an alternating fashion. Class AB amplifiers fall between Class A and Class B PAs; the efficiency is better than Class A and worse than Class B operation. There are also Class AB push-pull power amplifiers that operate like Class B push-pull PAs, except there is a small overlap of time when both devices are on. Higher class power amplifiers distinguish themselves either by how often the transistors are on, which affects efficiency, or by how hard they are driven (i.e., some act like switches).

This section covers Class A and Class B operation. Design fundamentals are presented as well as practical issues in PA design. In particular, the effect of the package is treated in the design of a Class A power amplifier. Amplifier stability, input-output matching, and package thermal conductivity are also addressed.

7.5.1 Class A Power Amplifier Design

At RF frequencies, Class A amplifiers are usually designed with a common-emitter structure (or common-source for FETs), as shown in Figure 7.26. The collector is biased with a choke to allow the collector to swing from ground to twice the supply rail. The result is that the maximum efficiency doubles from 25% to 50%.

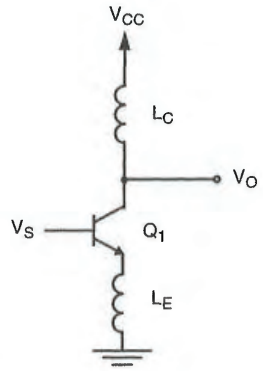


Figure 7.26 Basic Class A amplifier.

A common figure of merit used to compare different PA configurations is efficiency. Efficiency, denoted η, is defined as the ratio of output power to dc input power. Power-added efficiency (PAE) is another measure of efficiency for transistors that discriminates against low-gain devices in which a significant fraction of the output power comes from the RF input power [23]. The latter measure of efficiency is more specific to FET devices, which have lower gain performance than bipolar transistors. Expressions for efficiency and PAE are

$$\eta = \frac{P_{RF}}{P_{DC}} \tag{7.54}$$

$$\text{PAE} = \left(1 - \frac{1}{G}\right)\frac{P_{RF}}{P_{DC}} \tag{7.55}$$

where G is the RF gain of the device at the particular value of P_{RF} being achieved.

Maximum efficiency in Class A amplifiers can easily be found by making simplifying assumptions. First of all, the bias current I_C is approximately equal to the average supply current I_{CC}. Next, the maximum current delivered to the load equals the bias current. Finally, the peak voltage delivered to the load is the supply voltage V_{CC}. Consequently, the input power is

$$P_{DC} = I_C V_{CC} \tag{7.56}$$

the optimum load to present to the collector of the power device is

$$R_{Lopt} = \frac{V_{CC}}{I_C} \tag{7.57}$$

the rms power delivered to the load is then

$$P_{RF} = \frac{1}{2} V_{CC} I_{CC} \tag{7.58}$$

and from (7.54), the efficiency is

$$\eta = 50\% \tag{7.59}$$

Maximum efficiency in PAs can rarely be achieved in practice. The output swing is limited by collector-emitter saturation. As the collector falls below ~200 mV, the output signal is no longer linear. Another major limitation is that the emitter is not at an ideal ground; voltage drops develop across emitter resistance and package inductance, which further limit the output swing.

The problem of emitter voltage moving from 0V is commonly known as "ground-bounce" and is a major design issue in PAs. The ground-bounce voltage, or the voltage at the emitter referred to ground, is given by

$$V_{GB}(\omega) = j\omega L_E I_C(\omega) + R_E I_C(\omega) \tag{7.60}$$

The inductance introduced by the package (L_E) can be as little as 100 pH for an RF power package and as much as 4 nH for an inexpensive plastic package such as a thin quad flat pack (TQFP). Emitter resistance (R_E) increases ground bounce, but is useful for temperature compensation (an issue discussed later in this section).

Note that the ground bounce due to the inductor is in quadrature to the signal current, while the ground bounce due to emitter resistance will be in phase with the signal current.

An expression for the efficiency of a Class A amplifier in the presence of ground bounce is readily found. For maximum output power, the collector current will have an amplitude equal to the bias current. If the collector current is approximately sinusoidal, it can be written as

$$I_C(t) = I_C(1 + \sin\omega t) \tag{7.61}$$

The collector-emitter potential of the device can be written as

$$V_{ce}(t) = V_{CC} - R_L I_C \sin\omega t - \omega L_E I_C \cos\omega t - R_E I_C(1 + \sin\omega t) \tag{7.62}$$

The optimum load to present to the collector of the power device is

$$R_L = \frac{V_O}{I_C} \tag{7.63}$$

where V_O is the maximum output voltage amplitude. An expression for the maximum output voltage is found by assuming that the square root of the sum of the squares of inductive ground bounce and collector-emitter potential plus resistive ground bounce must not be larger than the supply voltage. It is given as

$$V_o = \sqrt{(V_{CC} - V_{CE_{sat}} - 2I_C R_E)^2 - (\omega L_E I_C)^2} \tag{7.64}$$

Assuming that all of the signal current is delivered to the load, the rms output power is

$$P_{RF} = \frac{1}{2} V_O I_C \tag{7.65}$$

and the efficiency is given by

$$\eta \cong \left(\frac{V_O}{V_{CC}}\right) 50\% \tag{7.66}$$

The output amplitude is limited by ground bounce and collector-emitter saturation voltage. If both effects are negligible, V_O approaches V_{CC} and a maximum efficiency of 50% is achieved.

The value of the optimum load impedance is invariably different (much lower) from the value required to achieve a good impedance match, which is typically 50Ω or 75Ω. An impedance transformation network can be used to meet both requirements. However, for low supply voltages and high output powers, the required step-up ratio can be quite large.

A numerical example will illustrate how to approach the design of a Class A power amplifier. The goal is to design a 1W power amplifier operating at 1 GHz with a supply voltage of V_{CC} = 5V. It is easiest to begin by neglecting ground inductance, thus simplifying (7.64). The bias current can be calculated from

$$I_C = \frac{2P_{out}}{V_O} \tag{7.67}$$

$$\approx \frac{2P_{out}}{V_{CC} - V_{CEsat} - 2I_C R_E} \tag{7.68}$$

When the resistive drop $I_C R_E$ is taken to be 0.1V (for ease of calculation) and the saturation voltage is V_{CEsat} = 0.2V, the bias current is approximately 435 mA.

(When ground inductance is considered, the bias current will have to increase to achieve the desired output power. The output voltage decreases and therefore, by (7.67), bias current must increase.)

The presence of ground inductance will reduce the efficiency of the amplifier. A reduction in efficiency as a function of inductance is given as

$$\frac{\eta|_{\text{with L}}}{\eta|_{\text{w/o L}}} = \sqrt{1 - \left(\frac{\omega L_E I_C}{V_{CC} - V_{CEsat} - 2I_C R_E}\right)^2} \tag{7.69}$$

Rewriting (7.69) for L_E will give a maximum inductance that can be tolerated for a given reduction in efficiency. In this example, if the efficiency is to be reduced by no more than 4%, the inductance can be no larger than 146 pH. A low-inductance RF package would be needed to meet this requirement.

Using (7.64), the output voltage is found to be ~4.6V and the efficiency from (7.66) is ~46% (as expected by design). The output power is found from

$$P_{RF} = \frac{1}{2} V_O I_C \tag{7.70}$$

and is calculated to be 978 mW. The output power is slightly lower than the target of 1W because of the inductive ground bounce. The output power can be increased to meet the specification by increasing the bias current by roughly 5%. The optimum load impedance is found (by (7.63)) to be 10.6Ω. This impedance can be matched to 50Ω with a single or two-stage L-section matching network. As the output power increases, the optimum load impedance will drop and an impedance match will become more difficult to achieve. Increasing the supply voltage is a potential solution if the device used is not limited by breakdown or there are no battery voltage constraints as would be the case in a handheld unit.

The design example was optimistic. In practice, a Class A amplifier will typically achieve less than 40% efficiency since the design must tolerate process, temperature, and supply variations.

7.5.2 Higher Efficiency Power Amplifiers

In low power receivers, Class B operation has significant advantages over Class A operation. One major advantage is that there is essentially zero power dissipation with zero input signal in a Class B amplifier. Unlike with Class A operation, the power dissipated in a Class B power amplifier is not wasted when the unit switches to a standby (or idle) mode. Another advantage of Class B amplifiers is that the efficiency is much higher than Class A amplifiers.

The difference between a single-ended Class B amplifier with resistive load and the preceding Class A amplifier is that the quiescent bias point is changed. In the Class B amplifier, the device is biased on the edge of conduction. The resulting output current waveform is a half sine-wave. When Fourier-analyzed, the output current waveform is found to contain a dc component that is a function of the RF power output, a component at the input signal frequency, and only even-order harmonics.

A Class B amplifier with resistive load is rarely used in practice because all the power in the harmonics is also dissipated in the load. An improved design has a tuned load. The load is a shunt parallel-tuned circuit that provides a short-circuit path for all the harmonics so that only the fundamental current component flows into the load. This results in the voltage across the device being a pure sinusoid at the fundamental frequency. The downside of this design is that optimum Class B performance from this single-ended circuit is only achieved over a narrow bandwidth.

For wider bandwidth performance, a commonly used Class B amplifier is in a push-pull configuration, as shown in Figure 7.27. Each device conducts for alternate half-cycles. The two transistors are driven 180-deg out of phase from each other. A Fourier expansion of the current waveforms shows that the even-order harmonics (including dc) are in phase and that the fundamental components are out of phase. The outputs can be combined in such a way that the even-order harmonics cancel and the fundamental components sum.

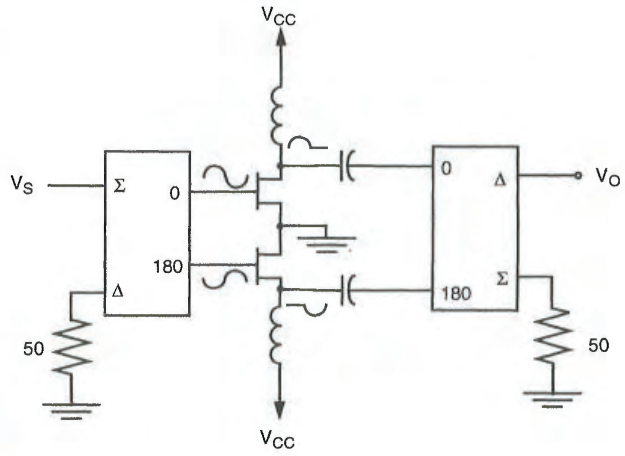


Figure 7.27 Push-pull Class B amplifier.

The power delivered to the load is the same in Class A and Class B operation. However, efficiency improves in Class B amplifiers because the average power drawn from the supplies is smaller than in Class A amplifiers.

The average supply current in each of the transistors is found by integrating the half-sinusoid over one period. It is found to be

$$I_D = \frac{1}{\pi} I_O \tag{7.71}$$

$$= \frac{1}{\pi} \frac{V_O}{R_L} \tag{7.72}$$

where I_O and V_O are the peak output current and voltage amplitudes. Since each supply delivers the same current magnitude, the total average power drawn from the two supplies is

$$P_{DC} = 2 I_O V_{CC} \tag{7.73}$$

$$= \frac{2}{\pi} \frac{V_O}{R_L} V_{CC} \tag{7.74}$$

Note that in the case of the Class B amplifier, the average power drawn from the supplies does vary with signal level and is directly proportional to V_O. The average power delivered to the load is

$$P_{RF} = \frac{1}{2} \frac{V_O^2}{R_L} \tag{7.75}$$

and, from the definition of circuit efficiency,

$$\eta = \frac{\pi}{4} \left(\frac{V_O}{V_{CC}} \right) \tag{7.76}$$

$$= \left(\frac{V_O}{V_{CC}} \right) 78.5\% \tag{7.77}$$

Many recently reported designs have been targeted at the commercial wireless market. In [24], a GaAs Class B power amplifier is designed for use in cellular phone systems. The amplifier achieves a PAE of 65% for an input power of 20 dBm from a 3.5V supply. The output power delivered to the load is 1.3W. Another design targeted for the cellular market was reported in [25]. This design is a two-stage

Class AB power amplifier achieving a PAE of 60% from a 6V supply with an input power of 7 dBm and an output of 1.2W. A third design recently reported was developed for portable wireless LAN (local area network) applications [26]. The design is a two-stage Class B amplifier. The output power is 24 dBm for an input power of 7 dBm. The supply voltage is 5V and the operating frequency is at 2.4 GHz.

7.5.3 Maximum Power Gain and Stability

The maximum power gain of a packaged device is typically limited by bond-wire inductance. The bond wire creates series-series feedback in the emitter and the parasitic collector-base capacitance of the device creates shunt-shunt feedback. The two effects combined will generate a narrowband matched-impedance condition that is analogous to the series-series/shunt-shunt resistive broadbanding technique for matched impedance discussed in Section 7.2. With simultaneous input and output match, the optimum power gain is

$$A_p = \frac{1}{\omega^2 L_E C_{jc}} \tag{7.78}$$

where L_E is the bond-wire inductance and C_{jc} is the parasitic collector-base capacitance. The power gain decreases by 20-dB per decade change in frequency.

The collector-base junction capacitance will increase with larger devices that are required for more output power. Therefore, bond-wire inductance L_E must scale inversely with power to keep power gain constant. Recall also that the inductance has to decrease linearly with increasing bias to keep ground bounce constant. A decrease in supply voltage must be accompanied by a decrease in package inductance to maintain power gain and ground bounce.

In choosing an appropriate technology, a compromise must be made between breakdown performance and junction capacitance. As discussed in the next section, low breakdown voltages present difficult problems for the designer. However, it is desirable to have a process with low C_{jc} for higher power gain. For higher frequency applications, a technology with lower parasitic junction capacitance and a low-inductance package should be used. To avoid breakdown problems, the supply voltage should also be fairly low. Applications at lower frequencies can tolerate a higher breakdown technology and a less expensive package.

As output power and frequency of operation increase, distributed effects need to be considered. Higher power levels require multiple devices spread out over large physical areas. The distributed effects of interconnect inductance and device input capacitance cause a phase shift at the input of each device. A similar phase shift can occur at the output of each device. Care must be taken in the device layout so that

these phase shifts do not become excessive; otherwise, power output and efficiency will be lost. Specifically, the outputs should be on the opposite side of the inputs so that the phase shifts at the input and output tend to cancel rather than add. This approach is analogous to a traveling wave amplifier, which is widely used in GaAs applications for frequencies in the millimeter wave region [27].

Stability is another crucial issue when designing PAs; the power amplifier must be stable under all possible load impedances. The combination of capacitive shunt feedback and series feedback, which will include both a resistive and inductive part, can lead to instability when driving inductive loads. Instability caused by variations in source impedance is less of an issue in monolithic implementations, because the driver impedance can more easily be controlled.

The source of the stability problem can be examined by calculating the input impedance as a function of load impedance. The input impedance of a device with both series and shunt feedback is given by

$$Z_{\text{in}} \cong \frac{\left(Z_e + \frac{1}{g_m}\right)Z_F}{Z_L} \tag{7.79}$$

where $Z_E = R_E + j\omega L_E$ and $Z_F = \frac{1}{j\omega C_{jc}}$. The load impedance Z_L includes the collector-substrate capacitance and the Early resistance $\left(\frac{V_A}{I_C}\right)$, as well as the impedance presented externally. The angle of Z_F equals –90 deg (purely capacitive feedback) and the angle of $Z_E + \frac{1}{g_m}$ is less than +90 deg. Therefore, the angle of Z_{in} will be more negative than –90 deg if the load is highly inductive. This implies a negative input resistance and possible instability.

Simple modifications to the PA of Figure 7.26 can be made to prevent instability. The improved design is shown in Figure 7.28. The resistor reduces the angle of the purely capacitive feedback and can be chosen to guarantee stability for any source or load. Resistive shunt feedback has the additional benefit of stabilizing the gain of the device over process, temperature, and supply variations. The shunt feedback will draw power from the output, so R_F should be much larger than R_L.

7.5.4 Practical Design Issues

Device breakdown is a major issue in the design of PAs. As process geometries shrink, the breakdown voltages of the devices decrease (in exchange for higher speed) and this creates a difficult problem for PA design. For example, a bipolar device with an f_T = 25 GHz will have a collector-base breakdown voltage of ~10V. As a

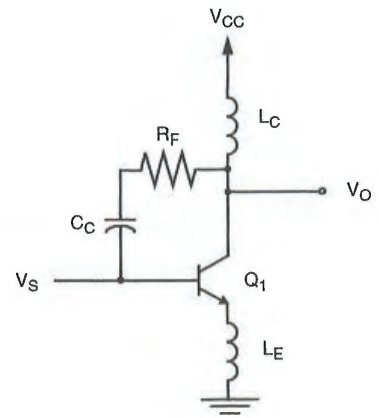


Figure 7.28 Stable Class A power amplifier.

general rule, the breakdown voltage of a power device should be three times the supply voltage. (Although the collector normally doesn't swing more than twice the supply voltage, it may swing higher under transient conditions.) Therefore, the supply voltage is limited to 3.3V. For a lower supply voltage, the bias current must be increased to deliver sufficient power to the output. The impact of increased bias current is more ground bounce and a decrease in the optimum load impedance, making matching more difficult.

The maximum power available from a PA in a bipolar technology is typically less than 1W from a 3V supply. Most cellular applications require PAs to produce 2 to 3W. The amplifiers operate off of four or five batteries, which corresponds to an end-of-life voltage of 4 to 5V. They are constructed in a process with larger geometry. PAs commonly use GaAs MESFETs or silicon DMOS power devices with custom geometries optimized for a high breakdown voltage.

Thermal properties of the package can limit the amount of power dissipated in the amplifier. A critical number for thermal design is the package $\theta_{J\text{-}A}$, or thermal conductivity. This figure of merit indicates how poorly the package dissipates heat by specifying the increase in junction temperature from ambient temperature per watt dissipated in the circuit. A good practice is to keep the junction temperature of the power device no more than 50°C above ambient temperature. An inexpensive plastic package may have $\theta_{J\text{-}A}$, = 100° C/W, which would limit the amount of power dissipated by the circuit to 500 mW. Assuming maximum efficiency of the Class A power amplifier, the package is required to dissipate at least as much power as

delivered to the load. Alternatives are to use a massive heat sink structure to reduce $\theta_{J\text{-}A}$, to as low as 2°C/W [4] or use and a more expensive package, which could have a $\theta_{J\text{-}A}$, of approximately 40°C/W. For even more output power, arrays of packaged devices may be used and the output power may be combined through a power-combining network. For TDMA cellular communications, the thermal requirements are ameliorated somewhat because the transmit power is only required for a certain fraction of the time.

Another temperature concern in the design of power amplifiers is thermal runaway. When a bipolar device self-heats, its intrinsic base-emitter voltage drops. However, if the bias voltage forces a constant V_{BE}, then the collector current will increase. This increase in current will produce more self-heating and at some point the thermal feedback causes instability or thermal runaway and the device burns up.

Thermal runaway problems can be reduced in two ways. First, the bias voltage should be developed from a reference device that is at thermal equilibrium with the power device. Second, each power device should have some emitter degeneration resistance. The degeneration will act as negative feedback both electrically and thermally. Using separate resistors in each of the multiple power devices ensures that no single device will consume all of the current and create a hot spot. This technique is known as emitter ballasting [28].

References

[1] B. Gilbert, "A New Wide-Band Amplifier Technique," *IEEE Journal of Solid-State Circuits*, Dec. 1968.

[2] P. A. Quinn, Feed-Forward Amplifier, United States Patent #4,146,844, March 1979.

[3] C. Hull, and R. G. Meyer, "Principles of Monolithic Wideband Feedback Amplifier Design," *Int'l Journal of High Speed Electronics*, March 1992.

[4] P. R. Gray, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Second Edition. New York, NY: John Wiley & Sons, 1984.

[5] C. R. Battjes, Monolithic Wideband Amplifier, United Stages Patent #4,323,119, Nov. 1980.

[6] B. W. Lai, Amplification Beyond f_T Using Non-Inductive Circuit Techniques, Master's Thesis UCB/ERL, University of California, Berkeley, CA, May 1983.

[7] I. Kipnis, J. F. Kukielka, J. Wholey, and C. P. Snapp, "Silicon Bipolar Fixed and Variable Gain Amplifier MMIC's for Microwave and Lightwave Applications up to 6 GHz." *IEEE MTT-S Digest*, 1989, pp. 109–112.

[8] D. Lee, A wideband Low-Noise BiCMOS Transimpedance Amplifier, Master's Thesis UCB/ERL, University of California, Berkeley, CA, May 1990.

[9] R. G. Meyer, and R. A. Blauschild. "A Wide-Band Low-Noise Monolithic Transimpedance Amplifier," IEEE Journal of Solid-State Circuits, Aug. 1986.

[10] J. Kukielka and C. Snapp, *Wideband Monolithic Cascadable Feedback Amplifiers using Silicon Bipolar Technology. Monolithic Microwave Integrated Circuits*, IEEE Press, 1985, pp. 330–331.

[11] R. G. Meyer, and R. A. Blauschild, "A 4-terminal Wide-Band Monolithic Amplifier," *IEEE Journal of Solid-State Circuits*, Aug. 1986.

[12] Avantek Application Note. An-S012: Magic Low Noise Amplifiers.

[13] R. G. Meyer, and W. D. Mack. "A 1-GHz BiCMOS RF Front-End IC," *IEEE Journal of Solid-State Circuits*, March 1994.

[14] I. A. Koullias, et al. "A 900 MHz Transceiver Chip Set for Dual-Mode Cellular Radio Mobile Terminals," *ISSCC 1993 Digest of Technical Papers*, Feb. 1993, pp. 140–141.

[15] H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York, NY: Wiley & Sons, 1976.

[16] W.M.C. Sansen, and R. G. Meyer, "Distortion in Bipolar Transistor Variable-Gain Amplifiers," *IEEE Journal of Solid-State Circuits*, Aug. 1973, pp. 275–282.

[17] D. O. Pederson, and K. Mayaram, *Analog Integrated Circuits for Communication*, Kluwer, 1991.

[18] W.M.C. Sansen, and R. G. Meyer, "An Integrated Wide-Band Variable-Gain Amplifier with Maximum Dynamic Range," *IEEE Journal of Solid-State Circuits*, Aug. 1974, pp. 159–166.

[19] B. Gilbert, Demystifying the Mixer, Analog Devices Internal Document, 1994.

[20] I. A. Koullias, S. L. Forgues, and P. C. Davis, "A 100 MHz IF Amplifier/Quadrature Demodulator for GSM Cellular Radio Mobile Terminals," *IEEE 1990 Bipolar Circuits and Technology Mtg.*, 1990.

[21] J. Fenk, et al. "An RF Front-End for Digital Mobile Radio." *IEEE 1990 Bipolar Circuits and Technology Mtg.*, 1990.

[22] P. Piriyapoksombut, and R. G. Meyer, "RF Variable Gain Amplifier," *UCB Infopad Retreat*, July 1994.

[23] J.L.B. Walker (editor), *High-power GaAs FET Amplifiers*, Norwood, MA: Artech House, 1993.

[24] M. Maeda, et al., "A 3.5 V, 1.3 W GaAs Power Multi Chip IC for Cellular Phones," *IEEE GaAs IC Symposium*, 1983.

[25] J. Naber, J. Griffiths, and J. Salvey, "A 1.2 W, 60% Efficient Power Amplifier IC for Commercial Applications," *IEEE GaAs IC Symposium*, 1994.

[26] T. Quach, and J. Staudinger, "A High Efficiency Commercial GaAs MESFET Power Amplifier for PCM/CIA Applications at 2.45 GHz," *IEEE GaAs IC symposium*, 1994.

[27] E. Strid, and R. Gleason, "A DC-12 GHz Monolithic GaAs FET Distributed Amplifier," *IEEE Trans. of Microwave Theory and Techniques*, July 1982.

[28] P. R. Gray, "A 15 Watt Monolithic Power Operational Amplifier," *IEEE Journal of Solid-State Circuits*, Dec. 1972, p. 474.

▼▼▼

GLOSSARY

ADPCM	adaptive differential pulse code modulation
AFC	automatic frequency control
AGC	automatic gain control
AM	amplitude modulation
AMPS	advanced mobile phone service
ARDIS	advanced radio data information service
AWGN	additive white Gaussian noise
BER	bit error rate
BPSK	bi-phase shift keying
BT	signal bandwidth × bit time
CDMA	code division multiple access
CDPD	cellular digital packet data
CMOS	complementary metal-oxide semiconductor
CNR	carrier over noise ratio
CIR	carrier over interference ratio
CRC	cyclic redundancy check
CT	cordless telephone (interim ETSI standard)
dBc/Hz	dB with respect to carrier level per hertz

DDS	direct digital synthesis
DECT	digital European cordless telecommunications
DRO	dielectric resonator oscillator
DSMA	digital sensed multiple access
EIA	Electronic Industries Association
EMC	electromagnetic compatibility
ERC	European Radio Commission
ETSI	European Telecommunication Standard Institute
FCC	Federal Communications Commission
FDD	frequency division duplex
FDMA	frequency division multiple access
FET	field effect transistor
FM	frequency modulation
FSK	frequency shift keying
GEO	geosynchronous satellite
GMSK	Gaussian filtered minimum shift keying
GSM	(Groupe Special Mobile) global system for mobile communication
HBT	heterojunction bipolar transistor
HEMT	high-electron mobility transistor
HIPERLAN	high-performance radio LAN
IEEE	Institute of Electrical and Electronics Engineers
IF	intermediate frequency
I and Q	in-phase and quadrature phase
IC	integrated circuits
IP	intermodulation product
ISM	industrial, scientific, and medical
JDC	Japanese digital cellular standard
KTB	Boltzmann's constant × temperature × bandwidth
LAN	local area network
LEO	low-Earth-orbit satellite
LNA	low noise amplifier
LO	local oscillator
MEO	medium-Earth-orbit satellite
MESFET	metal semiconductor field-effect transistor
MOSFET	metal oxide semiconductor field-effect transistor

MSS	mobile satellite service
NF	noise figure
NMT	Nordic Mobile Telephone
NTT	Nippon Telephone and Telegraph
PHS	personal handyphone system (formerly PHP)
PLL	phase lock loop
QCELP	qualcomm coded excited linear predictive coding
QDM	quadrature demodulator
QPSK	quadrature phase shift keying
RES	Radio Expert Systems Group
RF	radio frequency
RMS	root mean square
RPE-LTP	regular pulse excitation-long term predictor
RTMS	radio telephone mobile system
SAW	surface acoustic wave
SFDR	spur free dynamic range
SINAD	signal over noise and distortion
SMR	specialized mobile radio
TACS	total access communication system
TDD	time division duplex
TDMA	time division multiple access
TETRA	trans-European trunked radio system
TIA	Telecommunications Industry Association
UDPC	universal digital personal communications
VCO	voltage controlled oscillator
VSAT	very small aperture (satellite ground) terminal
VSELP	vector sum excited linear predictive coding
WER	word error rate
WLAN	wireless local area network
WLL	wireless local loop

▼▼▼

About the Editor

Lawrence E. Larson received his B.S. and M.S. degrees in Electrical Engineering from Cornell University, Ithaca, NY, in 1979 and 1980, and his Ph.D. from the University of California - Los Angeles, in 1986, also in Electrical Engineering. He is currently head of the Telecommunications Technology Department at Hughes Research Laboratories, Malibu, CA, where he directs research and development into advanced hardware solutions for emerging communications systems. His past experience includes the development of high-frequency III-V and silicon-based integrated circuits, improved microwave semiconductor devices, and semiconductor device modeling. In the past, Dr. Larson was an Adjunct Associate Professor of Electrical Engineering at UCLA, and has taught numerous short courses in Electrical Engineering at UCLA, UC-Berkeley, Oregon Graduate Research Institute, and Polytechnic University. He is a Senior Member of the IEEE.

▼▼▼

INDEX

4-FSK, 169–71, 172, 173, 174, 175
4-level GMSK, 189, 194, 195, 196
8-PSK, 137–41, 143–46, 214–17, 218ACI (adjacent channel interference), 100
16-PSK, 141, 146, 217, 219, 220

Active FET mixers, 263
  circuits for, 266–68
  doubly balanced, 272–73
  dual-gate, 268–69
  single-device, 263–66
  singly balanced, 269–72
Additive white Gaussian noise (AWGN), 110, 112
Adjacent channel interference (ACI), 100
Advanced mobile phone service (AMPS), 60–61
  frequencies for, 4
AGC amplifier, 374–79
  improved, 378–79
Air-bridge mesa diode, 233
Alcatel, direct-conversion single-chip transceiver by, 70–71, 72, 75–76
Aliasing, in direct digital synthesis, 343
Altair system, 7
AM corner frequency, 312
AM noise, in oscillators, 312, 313
Amplifier(s)
  AGC, 374–78
  balanced input CMOS RF, 82, 83
  cascode bipolar transistor, 367–69
  Class A, 384–88
  Class B, 388–91
  compound-transistor stages in, 351–52
  current-feedback pair, 355–56
  current-feedback pair with Darlington, 356–57
  dual-gain-state differential, 381, 382
  gain-bandwidth of practical circuits in, 353–59
  gain-bandwidth product for, 346–53
  low-noise, 345, 365–69
  power, 63–65, 345–46, 382–94
  requirements for, 28–29
  single-stage, 347–49
  transistor model of, 353–54
  two-stage, 349–51
  variable-gain, 346, 369–82
  voltage-feedback pair, 358–59
  wideband matched-impedance, 359–64
  wideband techniques for, 346–64
Amplitude quantization errors, in direct digital synthesis, 341
AMPS (advanced mobile phone service), 60–61
  frequencies for, 4
Analog cellular telephones, standards for, 18–19
Analog techniques, digital vs., 2
Antipodal modulation, 102
Antiresonant frequency, 301

Applications, of wireless communications, 2–8, 18–22
AWGN (additive white Gaussian noise), 110, 112

Balanced mixers
  doubly, 258–62
  FET resistive, 276–77
  singly, 255–58
Baluns, for wireless mixers, 248–53
Bandwidth, of mixer, 228
Base-driven gain control, 371–74
Beam-lead diode, 234, 235
Bellboy™ paging receiver, 57
BER. *See* Bit error rate (BER)
BiCMOS RF variable-gain amplifier, 381–82, 383
Binary phase shift keying (BPSK), 101–15
Bipolar-junction transistors (BJTs), in mixers, 241–42
Bipolar multiplier, 374–78
Bipolar transistor, gain-bandwidth of, 353–54
Bipolar transistor common-emitter oscillator, 295–97
Bit energy requirement, 33–34
Bit error, 101
Bit error rate (BER)
  of BPSK, 112–13
  of 8-PSK, 138, 145
  of FQPSK, 211
  of FSK, 167, 171
  of MSK, 165, 166
  of multilevel FSK, 171, 176
  of $\pi/4$-DQPSK, 204, 205
  of QPSK, 130, 131
  with frequency synthesizer, 286–87
Bit slicer, 109
BJTs (bipolar-junction transistors), in mixers, 241–42
Boxcar response, of raised cosine filter, 153–55
BPSK (binary phase shift keying), 101–15
Brick-wall lowpass filter, frequency response and impulse response of, 147–48
Broadcast radio receivers, ICs in, 51–57

Capacitance, of Schottky-barrier diode, 233–34
Cascode bipolar transistor amplifier, 367–68
CCI (cochannel interference), 100
Cellular telephone(s)
  digital, 67–73
  standards for, 18–22
Cellular telephone transceivers, ICs in, 60–67
Chip(s), 78
Chip diodes, 234
Circuit technology options, 8–10
Clapp oscillator, 299–300
Class A amplifiers, 384–88
Class B amplifiers, 388–91
Closed-loop gain
  of amplifier, 348–49
  of oscillator, 293
CMOS RF amplifier, balanced input, 82, 83
Cochannel interference (CCI), 100
Colpitts oscillators, 295–99
Compound-transistor stages, gain-bandwidth product for, 350–53
Compression point, 47
Conversion efficiency, 64
  of mixer, 227
Conversion-matrix analysis, 242–44, 280
Cordless telephones, digital, 67–73
Crystal oscillators, 300–1, 302, 303
CT-2, standards for, 68
Current-feedback pair, 355–56
  with Darlington, 356–58
Cutoff frequency, of Schottky-barrier diode, 234

DAC (digital-to-analog converter) nonlinearities, in direct digital synthesis, 341
Darlington, current-feedback pair with, 356–58
Data encoder, for 8-PSK IC modulator, 215–16
Data slicer, 109
DBPSK (differentially encoded/decoded BPSK), 113–15
dc bias, in mixers, 257–58
dc offset, 75–76
DDFSs (direct-digital frequency synthesizers), 84–85, 86
DDS (direct digital synthesis), 287–89, 338–43
DECT, standards for, 68
DECT transceiver, National Semiconductor, 211–14
Delay spread, 100
Delta port, 248
Demodulators
  8-PSK, 138, 144
  4-FSK, 171, 174, 175BPSK, 108–9
  design for wireless communications systems of, 214–19, 220, 221
  FQPSK, 208, 210
  FSK, 167, 169, 170
  GMSK, 185–89, 190
  MSK, 159, 163
  OQPSK, 136, 140
  $\pi/4$-DQPSK, 198–201, 202
  QPSK, 127, 128
Demultiplexer

  OQPSK, 136
  QPSK, 126
Dielectric monoblock filter, 10, 11
Dielectric resonator oscillators (DROs), 304–8
Difference port, 248
Differentially encoded/decoded BPSK (DBPSK), 113–15
Differential mixer, 270–72
Differential phase angle, in $\pi/4$-DQPSK, 192–93, 198
Digital cellular and cordless telephones
  block diagram for, 68
  ICs in, 67–73
  standards for, 68
Digital communications, requirements for, 33–38
Digital communications circuits, examples of integrated circuit, 208–14
Digital modulation
  advantages of, 99–101
  by binary phase shift keying, 101–15
  by 8-PSK and 16-PSK phase modulation, 137–45
  by Feher's QPSK, 204–8
  by frequency shift keying, 166–73, 174, 175, 176
  by Gaussian-filtered MSK, 175–89, 190–96
  by minimum shift keying, 156–66
  Nyquist filtering in PSK modulation, 145–55
  by offset or staggered phase-shift keying, 134–37
  by $\pi/4$–shifted differentially encoded PSK, 189–204, 205
  by quadriphase (or quadrature) phase-shift keying, 115–34
Digital standards, 19–21
Digital techniques, analog vs., 2
Digital-to-analog converter (DAC) nonlinearities, in direct digital synthesis, 341
Diode-based BPSK modulator, 104, 105
Diode-bridge variable-gain amplifier, 381, 383
Diode mixers
  baluns and hybrids for, 248–55
  characteristics of, 244–45
  doubly balanced, 258–62
  image-rejection, 262
  Shottky-barrier, 232–35
  single-, 245–47
  singly balanced, 255–58
Direct-conversion transceivers, 73–78
Direct-digital frequency synthesizers (DDFs), 84–85, 86
Direct digital synthesis (DDS), 282–85, 338–43
Direct-sequence modulation, 78–79
Distortion, 99–100
  in mixers, 228–29
Dominant-pole technique, 349–50
Doubly balanced diode mixers, 258–62
Doubly balanced FET mixers, 272–73
Down conversion mixer, based on direct sampling of RF, 82–84
DROs (dielectric resonator oscillators), 306–8
Dual-gain-state differential amplifier, 381, 382
Dual-gate FET mixer, 268–69
Dual Gilbert-cell QPSK modulation, 123, 124
Duplexer, requirements for, 28
Dynamic range
  of receiver, 47, 48
  of variable-gain amplifier, 370–71

Edge-triggered phase detector, 331, 332–34
Efficiency, of power amplifiers, 385–86, 388
Electromagnetic compatibility (EMC), in wireless devices, 38–40
Embedding network, 243
Emitter-driven gain control, 371–74
Equivalent circuit
  of FET, 236
  of resistive FET, 279
  of Schottky-barrier diode, 233–34
Eye diagrams, measurement setup for observing, 183

Fast frequency shift keying (FFSK), 164
Feedback, in low-noise amplifiers, 367
Feedback zero, 350
Feher's QPSK (FQPSK), 204–8, 209, 210, 211
FET common-gate oscillator, 297–99
FET mixers
  active, 263–73
  doubly balanced, 272–73
  dual-gate, 268–69
  passive, 273–80
  singly balanced, 269–72
FET resistive mixers, 273–75
  balanced, 276–77
  modeling FETs for analysis of, 279
  ring, 277–78
  single-device, 274–76
FFSK (fast frequency shift keying), 164
Field-effect transistors (FETs)
  bipolar, 241–42
  equivalent circuit of, 236
  GaAs MESFETs, 239–40
  high electron-mobility, 240–41
  insulated-gate, 237
  silicon MOSFETs, 236–39
Filter(s)
  dielectric monoblock, 10, 11

matched, 159
requirements for, 29–30
Flip-flops, as frequency dividers, 331–35
FM noise
with frequency synthesizer, 290
in oscillators, 311–12, 313, 314
Four-FET commutating switch mixer, 63, 64
FQPSK (Feher's QPSK), 204–8, 209, 210, 211
Fractional-*N* PLL synthesizers, 328–29
Frequencies, for wireless communications, 2–8
Frequency dividers, in phase-locked loop design, 323, 331–36, 337
Frequency hopping, 79
Frequency mixing, 226–27, 242–44
Frequency modulation, 155
by frequency shift keying, 166–73, 174, 175, 176
by minimum shift keying, 156–66
Frequency pulling, 290–91
Frequency pushing, 291
Frequency response
of brick-wall lowpass filter, 147–48
of Gaussian filter, 177, 178
of Nyquist filter, 148–49
Frequency shift keying (FSK), 166-73, 174, 175, 176
4-, 169–71, 172, 173, 174, 175
fast, 164
Frequency stability, of oscillator, 291
Frequency synthesizer(s)
bit error rate with, 286–87
block diagram of, 286
in communications system, 286
direct, 287–89
direct-digital, 84–85, 86, 287–89, 338–43
figures of merit for, 289–92
for frequency-hopped transceiver, 84–85, 86
fundamentals of, 287–92
indirect, 289
multiple, 326–27
phase locked-loop design, 320–38
requirements for, 30
voltage-controlled oscillators, 292–320
Frequency tracking, PLL for, 327–28
FSK (frequency shift keying), 166–73, 174, 175, 176
4-, 169–71, 172, 173, 174, 175
fast, 164

GaAs Class B power amplifier, 390
GaAs *I/Q* direct IC modulator, 219, 223
GaAs MESFETs, in mixers, 239–40
GaAs MMIC, for cellular telephones, 62, 63
Gain-bandwidth product, 346–53
for compound-transistor stages, 251–53
for current-feedback pair, 355–56
for current-feedback pair with Darlington, 356–58
for single-stage amplifiers, 347–49
for transistor model, 353–54
for two-stage amplifiers, 349–51
for voltage-feedback pair, 358–59
Gain constant, of VCO, 321
Gain control, emitter-driven and base-driven, 371–74
Gate-driven transconductance mixer, 266–67
Gaussian-filtered MSK (GMSK), 175–89
4-level, 189, 194, 195, 196
GEC/Plessey wireless local area network (WLAN) transceiver, 208–11, 212
Geosynchronous (GEO) mobile satellite services, 5
Gilbert-cell-based phase detector, 329–30
Gilbert-cell BPSK modulator, 106–8
Gilbert multiplier, 272–73
Gilbert's quad, 374–78
variation of, 379, 380
Glitch energy, 342
Global system for mobile communication (GSM)
direct-conversion receivers for, 75–76
frequencies for, 4
standards for, 68
GMSK (Gaussian-filtered MSK), 175–89
4-level, 189, 194, 195, 196
Gray coding, in QPSK, 130–31
"Ground-bounce," 386–87

Hardlimiting, MSK signal with, 166, 168
Harmonic-balance simulation, of mixers, 280–82
Harmonics
with direct digital synthesis, 342, 343
with frequency synthesizer, 291–92
Hartley oscillators, 295–99
Heterojunction bipolar transistors (HBTs), 241–42
High electron-mobility transistors (HEMTs), 240–41
HIPERLAN, 20
Homodyne, 73
Hybrids, for wireless mixers, 248–55

ICs. *See* Integrated circuits (ICs)
IEEE 802.11, 20
IF (intermediate frequency), 46–47, 50
IF (intermediate frequency) signal, 226
IM (intermodulation distortion), 228–29
in variable-gain amplifier, 371

Image, 225
Image rejection, 46–47, 50
Image-rejection mixer, 54, 55, 260
IMD (intermodulation distortion), 226–27
in variable-gain amplifier, 369
Impulse response
of brick-wall lowpass filter, 147–48
of Gaussian filter, 177, 178
of Nyquist filter, 149–50, 153
Industrial, scientific, and medical (ISM) bands, 6, 7, 78
Input-referred third-order intercept (P3), 47
Insulated-gate FET, 237
Integrated circuits (ICs)
in broadcast radio receivers, 51–57
in cellular telephone transceivers, 60–67
in digital cellular and cordless telephones, 67–73
in direct-conversion transceivers, 73–78
to enable future transceivers, 81–87
examples of digital communications circuits using, 208–14
monolithic microwave, 51
signal-processing issues in wireless transceivers with, 45–51
for spread-spectrum wireless transceivers, 78–81
in wireless paging receivers, 57–60
Intermediate frequency (IF), 46–47, 50
Intermediate frequency (IF) signal, 226
Intermodulation distortion (IM, IMD), 228–29
in variable-gain amplifier, 371
Intersymbol interference (ISI), 100, 183–85
IP3 (input-referred third-order intercept), 47
I/Q generator, digital, for 8-PSK IC modulator, 215–16
IS-19 cellular telephone, block diagram of, 23
IS-55 cellular telephone
block diagram of, 24
practical performance limits for, 22–27
ISI (intersymbol interference), 100, 183–85
ISM (industrial, scientific, and medical) bands, 6, 7, 78
I/V characteristic, of Schottky-barrier diode, 232–33

Japanese Personal Handy Phone (J-PHP), standards for, 68

Kukielka configuration, 361–64

Large-signal analysis, 33
LEO (low Earth orbit) mobile satellite services, 5
LNAs (low-noise amplifiers), 345, 365–69
Local area networks (LANs), 5–8
standards for, 20–21
Local oscillator (LO), with quadrature output, 76–78
Local oscillator (LO) signal, 226
Loop filters, in phase-locked loop design, 336–38
Loop gain
of amplifier, 348–49
of oscillator, 293
Low Earth orbit (LEO) mobile satellite services, 5
Low-noise amplifiers (LNAs), 345, 365–69
Lumped-element hybrids, 254–55

Markets, for wireless communications, 2–8, 18–22
Matched filters, 159
Matched-impedance amplifiers, 259–64
Matching networks, in analog cellular telephones, 64–65
Maximum power gain, of power amplifier, 391–92
Medium earth orbit (MEO) mobile satellite services, 5
Mesa diode, 235
MESFETs, GaAs, in mixers, 239–40
Meyer configuration, 364
MicroTac, 66
Microwave frequencies, oscillator configurations for, 301–304
Microwave integrated circuits, monolithic, 57
Microwave ring mixer, 260–61
Microwave wireless communications circuits, recent examples of, 10–13
Minimum shift keying (MSK), 156–66
Gaussian-filtered, 175–89
Mixers
active FET, 263–73
baluns and hybrids for, 248–55
bandwidth of, 228
based on direct sampling of RF, 82–84
comparison table of, 281–82
conversion efficiency of, 227
cost of, 231
dc power use by, 301
differential, 270–72
diode, 244–62
distortion in, 228–29
FET resistive, 273–80
field-effect transistors in, 235–42
gate-driven transconductance, 266–67
Gilbert multiplier as, 272–73
harmonic-balance simulation of, 280–82
image-rejection, 262
microwave ring, 260–61
noise with, 227–28

passive FET, 273–79
practical characteristics of, 227–31
properties of, 226–31
rejection phenomena with, 230–31
requirements for, 29, 225
ring, 259–60
Schottky-barrier diodes in, 232–35
semiconductor devices for, 231–42
source-driven, 267–68
spurious responses with, 227, 230
spurious signal with, 231
thermal and mechanical characteristics of, 231
transconductance, 263–65
Mixing, frequency, 226–27, 242–44
Mixing products, 227
MMIC (monolithic microwave integrated circuits), 51
GaAs, for cellular telephones, 62, 63
Mobile radios, specialized, 21–22
Mobile satellite services (MSS), 5
Mobile telephony, frequencies for, 4
Mobile wide area wireless data services, summary of, 21
Modulation. *See* Digital modulation
Modulation accuracy, 38
Modulators
BPSK, 102–8
4-FSK, 173
8-PSK, 138, 144, 214–17, 218
16-PSK, 217, 219, 220
design for wireless communications systems of, 214–19, 220, 221
FQPSK, 204, 205
FSK, 166–67, 169
GaAs I/Q direct IC, 217, 221
GMSK, 179–80, 181
MSK, 159, 161, 162
OQPSK, 136, 139
π/4-DQPSK, 194–95, 199
QPSK, 118–27
quadrature, 119–21
Monolithic microwave integrated circuits (MMICs), 51
GaAs, for cellular telephones, 62, 63
MOSFETs, silicon, in mixers, 236–39
MSK (minimum shift keying), 156–66
Gaussian-filtered, 175–89
MSS (mobile satellite services), 5
Multiple frequency synthesizer, 326–27
Multiplier, 374–78
Multitanh doublet, 379–81
Multitone harmonic balance, 278
Multivibrator voltage-controlled oscillators, 317–20

NADC (North American Digital Cellular), standards for, 68
National Semiconductor DECT transceiver, 211–14
NEC, paging receivers from, 59–60
Noise
in BPSK, 110–15
with frequency synthesizer, 289–90
with mixer, 227–28
in multivibrator VCOs, 319–20
in oscillators, 310–13, 314
in phase-locked loops, 325–26
in ring oscillator VCOs, 316–17
Noise figure (NF), 30
of amplifier, 365, 366–67
Noise matching, series-feedback inductor for, 63
Noise power spectral density, with BPSK, 110–12
Noise spike, in QPSK, 130–31
Noise variance, 35
Nonreturn-to-zero (NRZ) data signal, in BPSK, 109
North American Digital Cellular (NADC), standards for, 68
Nyquist filtering, applied to PSK modulations, 145–55

Offset quadriphase phase-shift keying (OQPSK), 134–37, 138–42
Open-loop gain, of amplifier, 348–49
Oscillator(s), voltage-controlled. *See* Voltage-controlled oscillators (VCOs)
Oscillator nonlinearity factor, 312

PA(s). *See* Power amplifiers (PAs)
PAE (power-added efficiency), 385
Paging receivers, ICs in wireless, 57–60
Passive devices, 10
Passive FET mixers, 273–74
balanced, 276–77
modeling FETs for analysis of, 279
ring, 277–78
single-device, 274–76
Personal communication services (PCS)
frequencies for, 4–5
standards for, 18–22
Phase accumulation register, 328, 329
Phase detectors, in phase-locked loop design, 329–31, 332–34
Phase-locked loop (PLL)
acquisition time of, 324
applications of, 326–29
block diagram of, 320–21
design fundamentals of, 320–26
dynamic range of, 324
frequency dividers in, 323, 331–36, 337
frequency response of, 322
frequency synthesizer with, 289, 326–329
implementation of, 329–38
loop bandwidth of, 321
loop filters in, 336–38
loop lock range of, 324
natural frequency of, 321
noise in, 325–26
normalized transient response of, 321–22
phase detectors in, 329–31, 332–34
phase error transfer function of, 322–23
spurious outputs of, 325–26
Phase noise specifications, 34, 35–38
Phase quantization errors, in direct digital synthesis, 340–41
Phase shift keying (PSK)
8- and 16-, 137–45
binary, 101–15
Nyquist filtering in, 145–55
offset or staggered, 134–37, 138–42
π/4-shifted differentially encoded, 189–204, 205
quadriphase (or quadrature), 115–34
Phase trellis, MSK, 156, 158
Phasing method, 73–75
Philips, pager ICs from, 58–59
Pierce oscillator, 301, 303
π/4-shifted differentially encoded PSK (π/4-DQPSK), 189–204, 205
Plessey, spread-spectrum transceiver by, 79–81
PLL. *See* Phase-locked loop (PLL)
PN (pseudo-noise) sequence, 78–79
Power-added efficiency (PAE), 385
Power amplifiers (PAs), 345–46, 382–84
for analog cellular telephones, 63–65
Class A, 384–88
Class B, 388–91
higher efficiency, 388–91
maximum power gain and stability of, 391–92, 393
practical design issues with, 392–94
Power gain, of power amplifier, 391–92
Power output, of oscillator, 291
Power spectral density
4-FSK, 173
8-PSK, 141, 146
BPSK, 109–10
FSK, 169, 172
GMSK, 180, 182
MSK, 165
QPSK, 127, 131–32, 133
Predistortion, 375, 379
Predistortion circuitry, in π/4-DQPSK, 201–4
Premodulation filter, FQPSK, 204–6, 207
Programmable dividers, 334–35, 336
Pseudo-noise (PN) sequence, 78–79
PSK. *See* Phase shift keying (PSK)
Pulling range, 301
Push-pull Class B amplifier, 389

Quadrature amplitude modulation (QAM), 121, 141
Quadrature modulation, 74, 119–21
Quadrature outputs, local oscillators with, 76–78
Quadriphase (or quadrature) phase-shift keying (QPSK), 115–34
Feher's, 204–8, 209, 210, 211
offset or staggered, 134–37, 138–42

Radiofrequency (RF) components, for transceivers, 27–30
Radiofrequency (RF) signal, 226
Radiofrequency (RF) transformers, in mixers, 249–51
Radiofrequency (RF) wireless communications circuits, recent examples of, 10–13
Radio transceiver. *See* Transceiver(s)
Raised cosine filter, 148–55, 156
square-root, 195–98
Receiver(s)
broadcast, 51–57
ICs in, 51–60
wireless paging, 57–60
Receiver sensitivity, required, 34–35
Regenerative frequency dividers, 335–36, 337
Relaxation oscillators, 317
Resistive broadbanding, 348
Rf components, for transceivers, 27–30
RF signal, 226
Rf transformers, in mixers, 249–51
RF wireless communications circuits, recent examples of, 10–13
Ring mixer, 259–60
FET resistive, 277–78
microwave, 260–61
Ring oscillator voltage-controlled oscillators, 313–17
Roll-off factor, 148–49

Satellite communications, frequencies for, 5
SAW (surface acoustic wave) devices, 10, 11
Schottky-barrier diodes, for mixers, 232–35
Semiconductor devices, 8–10

for mixers, 231–42
Series-feedback inductor, for noise matching, 63
Series resistance, of Schottky-barrier diode, 234
Series resonant frequency, 301
Siemens, receiver and transmitter ICs by, 67–70, 72–73
Sigma port, 248
Signal bandwidth, 34
Signal processing, in wireless transceivers, 45–51
Signal sideband noise, 37
Signal synchronizers, 325
Signal vector
  BPSK, 115
  QPSK, 129–30
Silicon MOSFETs, in mixers, 236–39
Silicon substrates, spiral inductors on, 81–82
Single-device active FET mixers, 263–66
  circuits for, 266–68
Single-device FET resistive mixers, 274–76
Single-diode mixers, 245–47
Single-sideband synthesizers, direct-conversion, 73–75
Single-stage amplifiers, gain-bandwidth product of, 347–49
Single-stage matched-impedance amplifier, 360–61
Singly balanced diode mixers, 255–58
Singly balanced FET mixers, 269–72
Small-signal analysis, 32
Sony, broadcast radio receivers by, 54
Source-driven mixer, 267–68
Source resistance, in low-noise amplifier, 365–66
Specialized mobile radios (SMRs), 21–22
Spectral regrowth, 49, 134, 135
Spiral inductors, on silicon substrates, 81–82
Spread-spectrum wireless transceivers, ICs for, 78–81
Spurious responses
  in direct digital synthesis, 340–42
  in frequency synthesizer, 292
  in mixers, 227, 230
  in phase-locked loops, 325–26
Spurious signals, 231
Square-root raised cosine (SRRC) filter, 195–98
Stability
  of oscillator, 291
  of power amplifier, 392, 393
Staggered quadriphase phase-shift keying, 134–37, 138–42
Standards, for wireless communications, 18–22
Sum port, 248
Surface acoustic wave (SAW) devices, 10, 11
Symbols, in QPSK, 115–18
Systematic analysis, of transceiver design, 30–33

Terrestrial mobile telephony, frequencies for, 4
Thermal runaway, 394
Tracking filter, PLL, 327–28
Transceiver(s)
  cellular telephone, 60–67
  direct-conversion, 73–78
  ICs in, 60–67, 73–87
  requirements for, 22–27
  RF component requirements for, 27–30
  signal processing in, 45–51
  spread-spectrum wireless, 78–81
  systematic analysis of, 30–33
Transconductance mixers, 263–65
  gate-driven, 266–67
Transmission-line baluns and hybrids, 251–53
Tuning range, of oscillator, 291
Two-stage amplifiers, gain-bandwidth product of, 347–51
Two-stage matched-impedance amplifiers, 361–64

UCLA frequency-hopped spread-spectrum transceiver, 82
Unit increments, in design of modulators and demodulators, 214–15

Varactor diodes, 309
Variable-gain amplifiers (VGAs), 346, 369–70
  BiCMOS RF, 381–82, 383
  circuit configurations for, 374–82
  diode-bridge, 381, 383
  dynamic range of, 370–71
  emitter-driven and base-driven gain control in, 371–74
Voltage-controlled oscillators (VCOs), 292
  bipolar transistor common-emitter, 295–97
  Clapp, 299–300
  Colpitts, 295–99
  crystal, 300–301, 302, 303
  dielectric resonator, 304–8
  FET common-gate, 297–99
  general design of, 292–95
  Hartley, 295–99
  linear, 292–93
  for microwave frequencies, 300–304
  as MSK modulator, 164
  multivibrator, 317–20
  noise in, 310–13
  nonlinear, 292
  Pierce, 301, 303
  practical considerations for design of, 308
  relaxation, 317
  ring oscillator, 313–17
  voltage control of, 309
Voltage-feedback pair, 358–59

Wideband matched-impedance amplifiers, 359–64
Wireless communications
  applications of, 2–8, 18–22
  frequencies for, 2–8
  standards for, 18–22
Wireless communications circuits
  circuit technology options for, 8–10
  recent examples of RF and microwave, 10–13
Wireless communications systems, design of modulators and demodulators for, 214–19, 220, 221
Wireless devices, electromagnetic compatibility in, 38–40
Wireless local area network (WLAN) transceiver, GEC/Plessey, 208–11, 212
Wireless local loop (WLL), 22

Recent Titles in the Artech House Mobile Communications Series

John Walker, Series Editor

Advances in Mobile Information Systems, John Walker, editor

An Introduction to GSM, Siegmund M. Redl, Matthias K. Weber, and Malcolm W. Oliphant

CDMA for Wireless Personal Communications, Ramjee Prasad

CDMA RF System Engineering, Samuel C. Yang

CDMA Systems Engineering Handbook, Jhong S. Lee and Leonard E. Miller

Cell Planning for Wireless Communications, Manuel F. Cátedra and Jesús Pérez-Arriaga

Cellular Communications: Worldwide Market Development, Garry A. Garrard

Cellular Mobile Systems Engineering, Saleh Faruque

The Complete Wireless Communications Professional: A Guide for Engineers and Managers, William Webb

GSM and Personal Communications Handbook, Siegmund M. Redl, Matthias K. Weber and Malcolm W. Oliphant

GSM Networks: Protocols, Terminology, and Implementation, Gunnar Heine

GSM System Engineering, Asha Mehrotra

Handbook of Land-Mobile Radio System Coverage, Garry C. Hess

Handbook of Mobile Radio Networks, Sami Tabbane

High-Speed Wireless ATM and LANs, Benny Bing

Introduction to Mobile Communications Engineering, José M. Hernando and F. Pérez-Fontán

Introduction to Radio Propagation for Fixed and Mobile Communications, John Doble

Introduction to Wireless Local Loop, William Webb

IS-136 TDMA Technology, Economics, and Services, Lawrence Harte, Adrian Smith, and Charles A. Jacobs

Mobile Communications in the U.S. and Europe: Regulation, Technology, and Markets, Michael Paetsch

Mobile Data Communications Systems, Peter Wong and David Britland

Mobile Telecommunications: Standards, Regulation, and Applications, Rudi Bekkers and Jan Smits

Personal Wireless Communication With DECT and PWT, John Phillips and Gerard Mac Namee

Practical Wireless Data Modem Design, Jonathan Y.C. Cheah

Radio Propagation in Cellular Networks, Nathan Blaunstein

RDS: The Radio Data System, Dietmar Kopitz and Bev Marks

Resource Allocation in Hierarchical Cellular Systems, Lauro Ortigoza-Guerrero and A. Hamid Aghvami

RF and Microwave Circuit Design for Wireless Communications, Lawrence E. Larson, editor

Spread Spectrum CDMA Systems for Wireless Communications, Savo G. Glisic and Branka Vucetic

Understanding Cellular Radio, William Webb

Understanding Digital PCS: The TDMA Standard, Cameron Kelly Coursey

Understanding GPS: Principles and Applications, Elliott D. Kaplan, editor

Universal Wireless Personal Communications, Ramjee Prasad

Wideband CDMA for Third Generation Mobile Communications, Tero Ojanperä and Ramjee Prasad

Wireless Communications in Developing Countries: Cellular and Satellite Systems, Rachael E. Schwartz

Wireless Technicians's Handbook, Andrew Miceli

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
46 Gillingham Street
London SW1V 1AH UK
Phone: +44 (0)20 7596-8750
Fax: +44 (0)20 7630-0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at:
www.artechhouse.com