MIXER CIRCUITS

A frequency mixer inputs two frequencies—a radio frequency (RF) and a local oscillator (LO) frequency—mixes them, and produces their difference frequency and sum frequency. The output signal is tuned by a filter, and one of the two output frequencies is selected: the difference or the sum. When the output difference frequency is an intermediate frequency (IF), the mixer is usually called a downconversion frequency mixer, and when the output sum frequency is a high frequency, it is usually called an upconversion frequency mixer.

A frequency mixer is fundamentally a multiplier, because the analog multiplier outputs a signal proportional to the product of the two input signals. Therefore, a frequency mixer is represented by the symbol for the multiplier, as shown in Fig. 1.

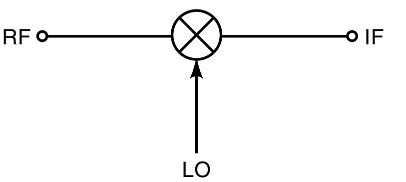


Figure 1. A symbol for a frequency mixer. The symbol for a multiplier is used.

The transfer function of a nonlinear element is expressed as

$$f(u) = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \cdots + a_n u^n + \cdots \quad (1)$$

The product xy of the two input signals x and y can be derived from only the second-order term: a_2u^2, where $u = x + y$, and x and y are the two input signals. The product of the two input signals is produced by a nonlinear element, such as a diode or transistor. For example, single-diode mixers, singly balanced diode mixers, doubly balanced diode mixers, single-transistor mixers, singly balanced transistor mixers, and doubly balanced transistor mixers are usually used as frequency mixers.

APPLICATION TO RECEIVERS

Mixers are used to shift the received signal to an intermediate frequency, where it can be amplified with good selectivity, high gain, and low noise, and finally demodulated in a receiver. Mixers have important applications in ordinary low-frequency and microwave receivers, where they are used to shift signals to frequencies where they can be amplified and demodulated most efficiently. Mixers can also be used as phase detectors and in demodulators, and must perform these functions while adding minimal noise and distortion.

Figure 2 shows, for example, the block diagram of a VHF or UHF communication receiver. The receiver has a single-stage input amplifier; this preamp, which is usually called an RF amplifier, increases the strength of the received signal so that it exceeds the noise level of the following stage; therefore, this preamp is also called a low-noise amplifier (LNA). The first IF is relatively high (in a VHF or UHF receiver, the widely accepted standard has been 10.7 MHz); this high IF moves the image frequency well away from the RF, thus allowing the image to be rejected effectively by the input filter. The second conversion occurs after considerable amplification, and is used to select some particular signal within the input band and to shift it to the second IF. Because narrow bandwidths are generally easier to achieve at this lower frequency, the selectivity of the filter used before the detector is much better than that of the first IF. The frequency synthesizer generates the variable-frequency LO signal for the first mixer, and the fixed-frequency LO for the second mixer.

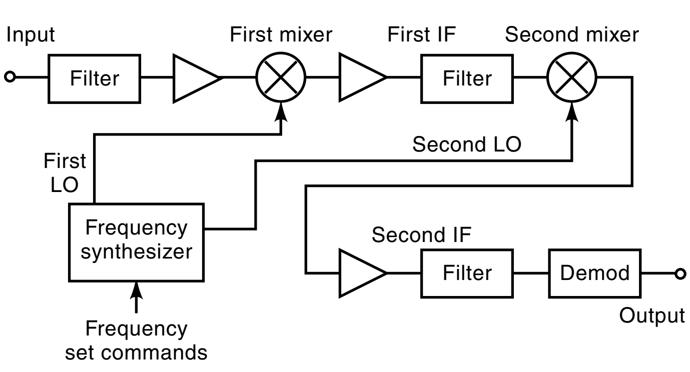


Figure 2. Double superheterodyne VHF or UHF communication receiver.

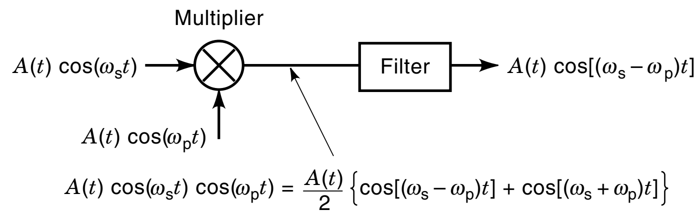


Figure 3. A mixer is fundamentally a multiplier. The difference frequency in the IF results from the product of sinusoids.

Figure 3 illustrates an ideal analog multiplier with two sinusoids applied to it. The signal applied to the RF port has a carrier frequency ω_s and a modulation waveform $A(t)$. The other, the LO, is a pure, unmodulated sinusoid at frequency ω_p.

Applying some basic trigonometry to the output is found to consist of modulated components at the sum and difference frequencies. The sum frequency is rejected by the IF filter, leaving only the difference.

Fortunately, an ideal multiplier is not the only device that can realize a mixer. Any nonlinear device can perform the multiplying function. The use of a nonideal multiplier results in the generation of LO harmonics and in mixing products other than the desired one. The desired output frequency component must be filtered from the resulting chaos.

Another way to view the operation of a mixer is as a switch. Indeed, in the past, diodes used in mixers have been idealized as switches operated at the LO frequency. Figure 4(a) shows a mixer modeled as a switch; the switch interrupts the RF voltage waveform periodically at the LO frequency. The IF voltage is the product of the RF voltage and the switching waveform.

Another switching mixer is shown in Fig. 4(b). Instead of simply interrupting the current between the RF and IF ports,

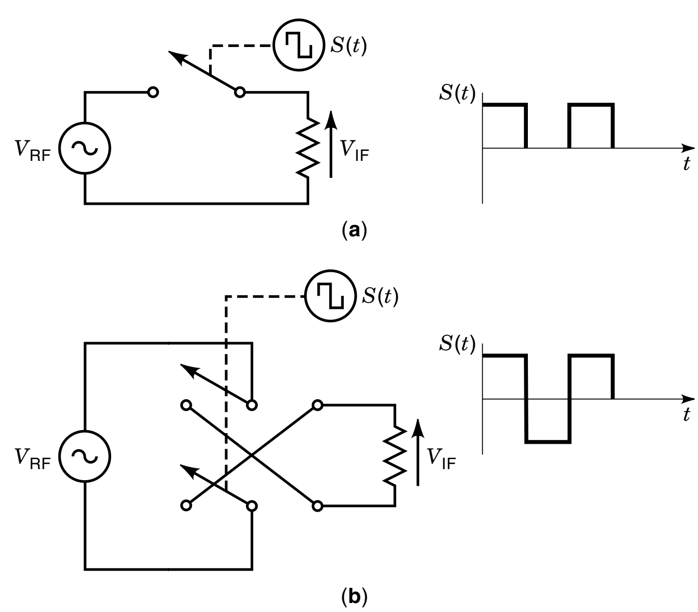


Figure 4. Two switching mixers: (a) a simple switching mixer: (b) a polarity-switching mixer. The IF is the product of the switching waveform $s(t)$ and the RF input, making these mixers a type of multiplier.

the switch changes the polarity of the RF voltage periodically. The advantage of this mixer over the one in Fig. 4(a) is that the LO waveform has no dc component, so the product of the RF voltage and switching waveform does not include any voltage at the RF frequency. Thus, even though no filters are used, the RF and LO ports of this mixer are inherently isolated. Doubly balanced mixers are realizations of the polarity-switching mixer.

SEMICONDUCTOR DEVICES FOR MIXERS

Only a few devices satisfy the practical requirements of mixer operation. Any device used as a mixer must have strong nonlinearity, electrical properties that are uniform between individual devices, low noise, low distortion, and adequate frequency response. The primary devices used for mixers are Schottky-barrier diodes and field-effect transistors (FETs). Bipolar junction transistors (BJT) are also used occasionally, primarily in Gilbert-cell multiplier circuits [see Fig. 6(d)], but because of their superior large-signal-handling ability, higher frequency range, and low noise, FET devices such as metal–oxide–semiconductor FETs (MOSFET), gallium arsenide (GaAs) metal–semiconductor FETs (MESFET), and high-electron-mobility transistors (HEMTs) have been usually preferred.

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

Since the 1950s, bipolar transistors have dominated mixer applications as single-transistor mixers in AM radio and communication receivers. In particular, an analog multiplier consisting of a doubly balanced differential amplifier, called the *Gilbert cell,* was invented in the 1960s. Since then, the Gilbert-cell mixer has been used as a monolithic integrated circuit (IC) for AM radio receivers and communication equipment. Silicon BJTs are used in mixers because of their low cost and ease of implementation with monolithic ICs. These bipolar devices are used as mixers when necessary for process compatibility, although FETs generally provide better overall performance. Silicon BJTs are usually used in conventional single-device or singly and doubly balanced mixers. Progress in the development of heterojunction bipolar transistors (HBT), which use a heterojunction for the emitter-to-base junction, may bring about a resurgence in the use of bipolar devices as mixers. HBTs are often used as analog multipliers operating at frequencies approaching the microwave range; the most common form is a Gilbert cell. Silicon–germanium (Si–Ge) HBTs are a new technology that offers high performance at costs close to that of silicon BJTs.

A variety of types of 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 frequency, GaAs MESFETs are often used. The LO and RF signals can be applied to separate gates of dual-gate FETs, 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.

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 to GaAs, silicon MOSFETs do have some advantages. The primary one is low cost, and the performance of silicon MOSFET mixers is not significantly worse than GaAs in the VHF and UHF range. The high drain-to-source resistance of silicon MOSFETs gives 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 with 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 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. GaAs is decidedly superior to silicon for high-frequency mixers because of its higher electron mobility and saturation velocity. The gate 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.

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

Passive Diode Mixers

Figure 5 shows the most common form of the three diode-mixer types: a single-device diode mixer, a singly balanced diode mixer, and a doubly balanced diode mixer. Conversion loss of 6 to 8 dB is usually accepted in these passive mixers.

Active Transistor Mixers

Active transistor mixers have several advantages, and some disadvantages, in comparison with diode mixers. Most sig-

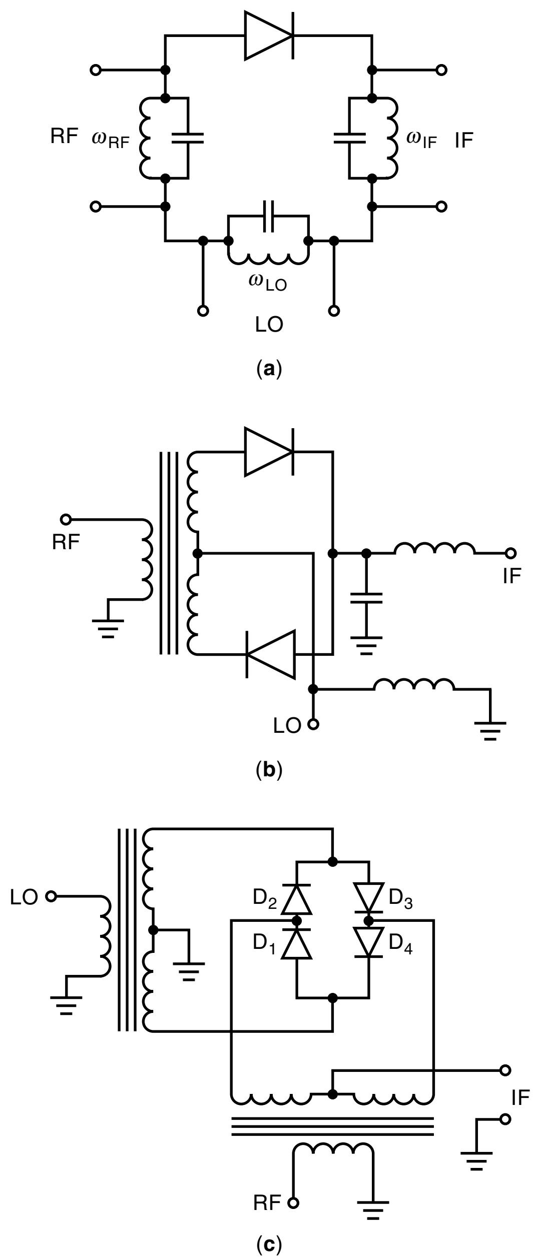


Figure 5. The three most common diode-mixer types: (a) single-device, (b) singly balanced, (c) doubly balanced.

nificantly, 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 transistor 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.

It is usually easy to achieve good conversion efficiency in active mixers. Thus, active transistor mixers have gained a reputation for low performance. Nevertheless, achieving good overall performance in active transistor mixers is not difficult.

Because transistors cannot be reversed, as can diodes, balanced transistor 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.

Bipolar Junction Transistor Mixers. Figure 6 shows BJT mixers: a single-device BJT mixer, a singly balanced BJT mixer, a differential BJT mixer, and a doubly balanced BJT mixer.

In a single-device BJT mixer [Fig. 6(a)], the input signals are introduced into the device through the RF LO diplexer, which consists of an RF bandpass filter, an LO bandpass filter, and two strips, $\lambda/4$ long at the center of the RF and LO frequency ranges; the square-law term of the device's characteristic provides the multiplication action. A single-device BJT mixer achieves a conversion gain of typically 20 to 24 dB, a noise figure of typically 4 to 5 dB (which is about 3 dB more than that of the device in the amplifier at the RF), and a third intercept point near 0 dBm. The IM product from this type of single-device BJT mixer usually depends on its collector current, but when the supplied collector-to-emitter voltage, V_{CE}, is not enough (typically, below 1.2 V), the IM product increases as V_{CE} decreases.

A singly balanced BJT upconversion mixer [Fig. 6(b)] consists of two BJTs interconnected by a balun or hybrid. The two collectors are connected through a strip, $\lambda/2$ long at the center of the LO frequency range, for reducing the LO leakage. This upconversion mixer exhibits 16 dB conversion gain and 12 dB LO leakage suppression versus the wanted RF output level at 900 MHz.

A singly balanced BJT differential mixer [Fig. 6(c)] consists of an emitter-coupled differential pair. The RF is superposed on the tail current by ac coupling through capacitor C_2, and the LO is applied to the upper transistor pair, where capacitive degeneration and ac coupling substantially reduce the gain at low frequencies. Note that the circuit following C_2 is differential and hence much less susceptible to even-order distortion.

A multiplier circuit [Fig. 6(d)] conceived in 1967 by Barrie Gilbert and widely known as the Gilbert cell (though Gilbert himelf was not responsible for his eponymy; indeed, he has noted that a prior art search at the time found that essentially the same idea—used as a "synchronous detector" and not as true mixer—had already been patented by H. Jones) is usually used as an RF mixer and sometimes as a microwave mixer.

Ignoring the basewidth modulation, the relationship between the collector current I_C and the base-to-emitter voltage V_{BE} for a BJT is

$$I_C = I_S \exp\left(\frac{V_{BE}}{V_T}\right) \tag{2}$$

where $V_T = kT/q$ is the thermal voltage, k is Boltzmann's constant, T is absolute temperature in kelvin, and q is the charge of an electron. I_S is the saturation current for a graded-base transistor.

Assuming matched devices, the differential output voltage of the Gilbert cell is

$$V_{IF} = -R_L I_{EE} \tanh\left(\frac{V_{RF}}{2V_T}\right) \tanh\left(\frac{V_{LO}}{2V_T}\right) \tag{3}$$

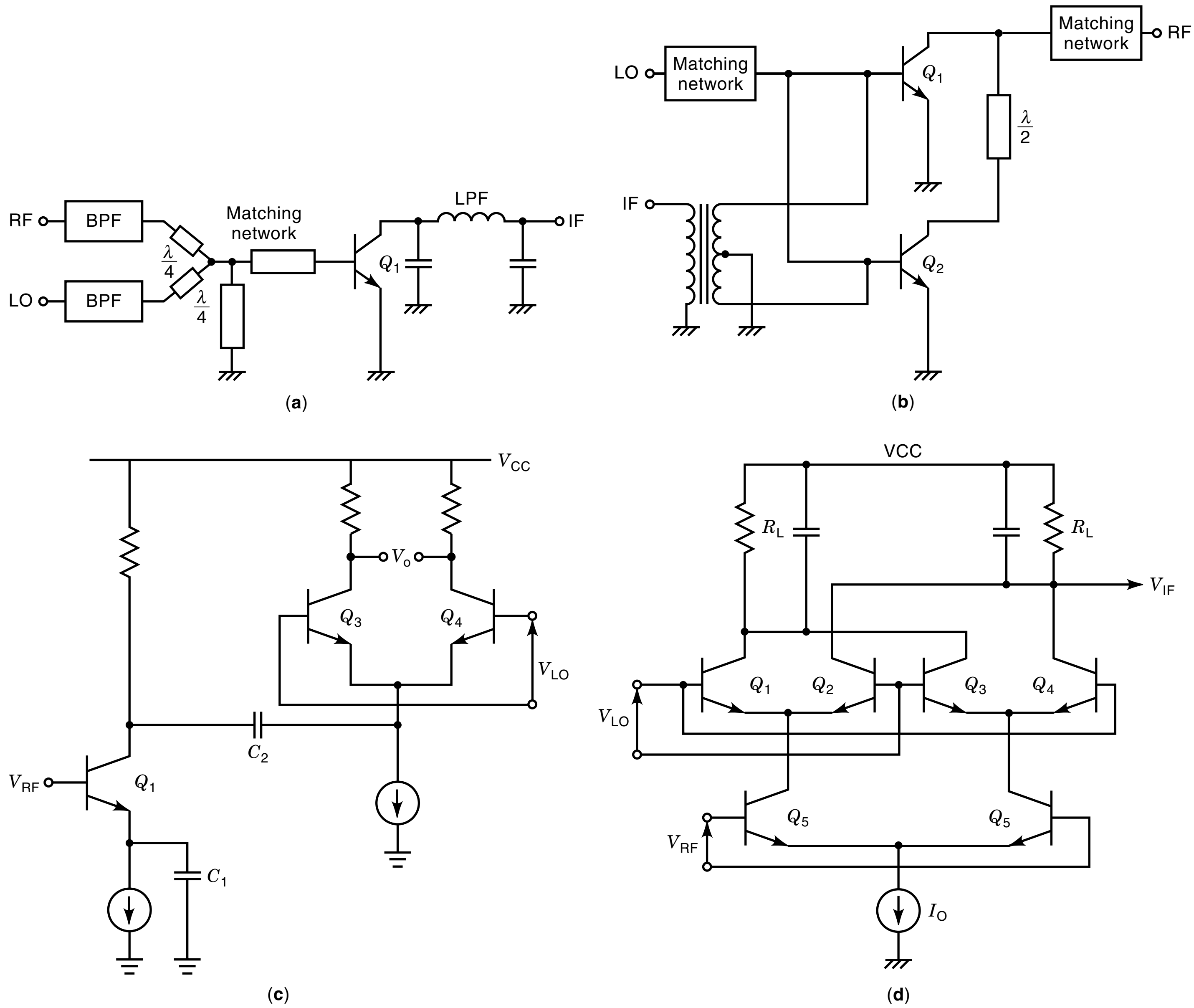


Figure 6. BJT mixers: (a) a single-device BJT mixer, (b) a singly balanced BJT upconversion mixer, (c) a singly balanced BJT differential mixer, (d) a doubly balanced BJT mixer consisting of a Gilbert cell.

For small inputs,

$$V_{\mathrm{IF}} \approx -\frac{R_{\mathrm{L}} I_{\mathrm{EE}}}{4V_T^2} V_{\mathrm{RF}} V_{\mathrm{LO}} \tag{4}$$

The product $V_{\mathrm{RF}}V_{\mathrm{LO}}$ is obtained by the Gilbert cell at small signals.

FET Mixers. Figure 7 shows FET mixers: a single-device FET mixer, a dual-gate FET mixer, a singly balanced FET mixer, a differential FET mixer, and a doubly balanced FET mixer.

In a single-device FET mixer [Fig. 7(a)], 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 and must short-circuit the drain at the RF and especially at the LO frequency and its harmonics.

The configuration of a dual-gate mixer [Fig. 7(b)] 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 LO–RF filter is used at the drain. A dual-gate mixer is usually realized as two single-gate FETs in a cascade connection.

A singly balanced FET mixer [Fig. 7(c)] 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. The IF filters provide the requisite short circuits to the drains at the LO and RF frequencies, and additionally pro-

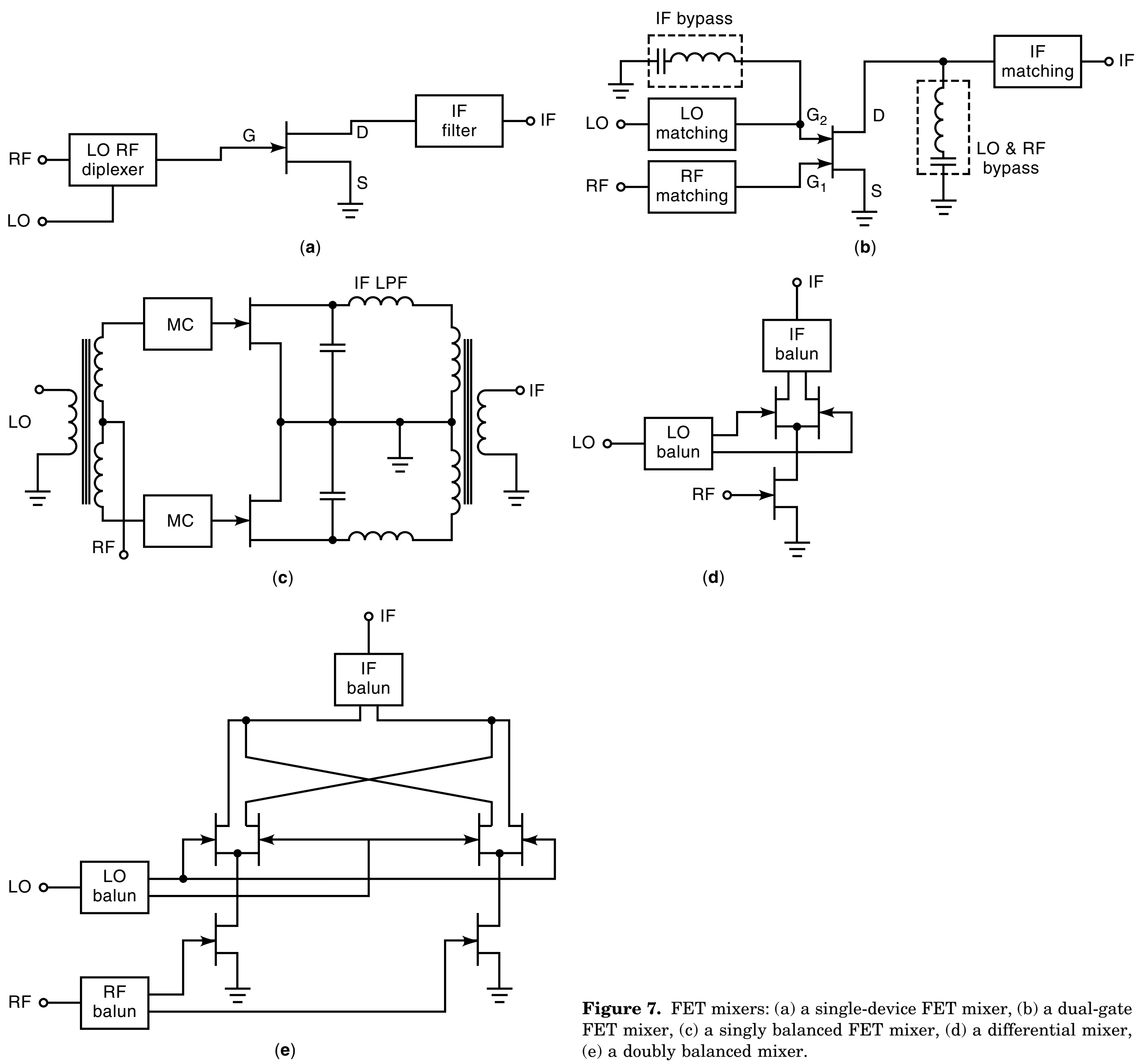


Figure 7. FET mixers: (a) a single-device FET mixer, (b) a dual-gate FET mixer, (c) a singly balanced FET mixer, (d) a differential mixer, (e) a doubly balanced mixer.

vide IF load impedance transformations. The singly balanced mixer of Fig. 7(c) is effectively two single-device mixers interconnected by hybrids.

In a differential FET mixer [Fig. 7(d)], 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. 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 between effective LO and RF matching. Similarly, the lower FET can be matched effectively at the RF. An IF filter is necessary to reject LO current.

A doubly balanced FET mixer [Fig. 7(e)] is frequently used as an RF or microwave mixer. Like many doubly balanced mixers, this mixer consists of two of the singly balanced mixers shown in Fig. 7(d). Each half of the mixer operates in the same manner as that of Fig. 7(d). The interconnection of the outputs, however, causes the drains of the upper four FETs to be virtual grounds for both LO and RF, as well as for even-order spurious responses and IM products.

IMAGE-REJECTION MIXERS

The image-rejection mixer (Fig. 8) 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 sig-

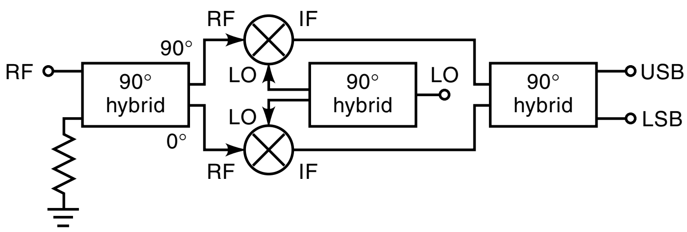


Figure 8. Image-rejection mixer.

nals applied to the RF ports have 90° phase difference. A 90° IF hybrid is used to separate the RF and image bands. A full discussion of the operation of such mixers is a little complicated.

The most difficult part of the design of an image-rejection mixer is the IF hybrid. If the IF 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.

MIXING

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

$$S_{\mathrm{RF}}(t) = a(t)\sin(\omega_s t) + b(t)\cos(\omega_s t) \tag{5}$$

is multiplied by the LO signal function

$$f_{\mathrm{LO}}(t) = \cos(\omega_p t) \tag{6}$$

to obtain the IF signal

$$\begin{aligned} S_{\mathrm{IF}}(t) = {} & \tfrac{1}{2}a(t)\sin[(\omega_s + \omega_p)t] + \sin[(\omega_s - \omega_p)t]\} \\ & + \tfrac{1}{2}b(t)\cos[(\omega_s + \omega_p)t] + \cos[(\omega_s - \omega_p)t] \end{aligned} \tag{7}$$

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.

Even if the LO 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 and eliminate the other (so-called *spurious*) responses.

Every mixer, even an ideal one, has a second RF 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 $2f_{\mathrm{LO}} - f_{\mathrm{RF}}$. 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 nonlinearity. Distortion and spurious responses other than the sum and difference frequency, though often severe in mixers, are not fundamentally required by the frequency-shifting operation that a mixer performs.

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 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. The conversion gain (CG) is defined as

$$\mathrm{CG} = \frac{\text{IF power available at mixer output}}{\text{RF power available to mixer input}} \tag{8}$$

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 decibels, is best.

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 a 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 (NF) is defined by the equation

$$\mathrm{NF} = \frac{\text{input signal-to-noise power ratio}}{\text{output signal-to-noise power ratio}} \tag{9}$$

The sensitivity of a receiver is usually limited by its internally generated noise. However, other phenomena sometimes affect the performance of a mixer front end more severely than noise. One of these is the AM noise, or *amplitude noise,* from the LO source, which is injected into the mixer along with the LO signal. This noise may be especially severe in a single-ended mixer (balanced mixers reject AM LO noise to some degree) or when the LO signal is generated at a low level and amplified.

Phase noise is also a concern in systems using mixers. LO sources always have a certain amount of phase jitter, or phase noise, which is transferred degree for degree via the mixer to the received signal. This noise may be very serious in communications systems using either digital or analog phase modu-

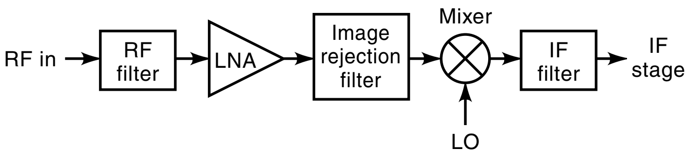


Figure 9. RF front end.

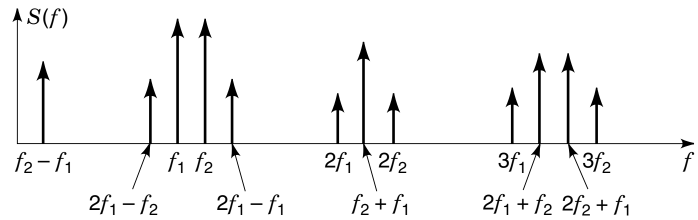


Figure 10. IF spectrum of intermodulation products up to third order. The frequencies f_1 and f_2 are the excitation.

lation. Spurious signals may also be present, along with the desired LO signal, especially if a phase-locked-loop frequency synthesizer is used in the LO source. Spurious signals are usually phase-modulation sidebands of the LO signal, and, like phase noise, are transferred to the received signal. Finally, the mixer may generate a wide variety of intermodulation products, which allow input signals—even if they are not within the input passband—to generate spurious output at the IF. These problems must be circumvented if a successful receiver design is to be achieved.

An ideal amplifier would amplify the incoming signal and incoming noise equally and would introduce no additional noise. From Eq. (9) such an amplifier would have a noise figure equal to unity (0 dB).

The noise figure of several cascaded amplifier stages is

$$\mathrm{NF} = \mathrm{NF}_1 + \frac{\mathrm{NG}_2 - 1}{G_1} + \frac{\mathrm{NF}_3 - 1}{G_1 G_2} + \cdots + \frac{\mathrm{NF}_n - 1}{\prod_1^n G_n} \tag{10}$$

where NF is the total noise figure, NF_n is the noise figure of the nth stage, and G_n is the available gain of the nth stage.

From Eq. (10), the gain and noise figure of the first stage of a cascaded chain will largely determine the total noise figure. For example, the system noise figure (on a linear scale) for the downconverter shown in Fig. 9 is

$$\begin{aligned}\mathrm{NF} &= \frac{1}{L_{\mathrm{RF}}} + \frac{\mathrm{NF}_{\mathrm{LNA}} - 1}{L_{\mathrm{RF}}} + \frac{1}{L_{\mathrm{RF}} G_{\mathrm{LNA}}}\left(\frac{1}{L_{\mathrm{IM}}} - 1\right) \\ &\quad + \frac{\mathrm{NF}_{\mathrm{M}} - 1}{L_{\mathrm{RF}} G_{\mathrm{LNA}} L_{\mathrm{I}}} + \cdots = \frac{1}{L_{\mathrm{RF}}}\left(\mathrm{NF}_{\mathrm{LNA}} + \frac{\mathrm{NF}_{\mathrm{M}} - L_{\mathrm{I}}}{G_{\mathrm{LNA}} L_{\mathrm{I}}} + \cdots\right)\end{aligned} \tag{11}$$

where L_{RF} and L_{I} are the insertion losses of the RF filter and the image-rejection filter, respectively, $\mathrm{NF}_{\mathrm{LNA}}$ and NF_{M} are the noise figures of the LNA and the mixer, respectively, and G_{LNA} is the power gain of the LNA. This equation assumes that the noise figures of the filters are the same as their insertion losses.

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 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 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 IM distortion (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 Fig. 10. Figure 10 shows all intermodulation products up to third order; by nth order, we mean all n-fold combinations of the excitation tones (not including the LO frequency). In general, an nth-order nonlinearity gives rise to distortion products of nth (and lower) order.

An important property of IMD is that the level of the nth-order IM product changes by n decibels for every decibel of change in the levels of the RF excitations. The extrapolated point at which the excitation and IMD levels are equal is called the nth-order IM intercept point, abbreviated IP_n. This dependence is illustrated in Fig. 11. In most components, the intercept point is defined as an output power: in mixers it is traditionally an input power.

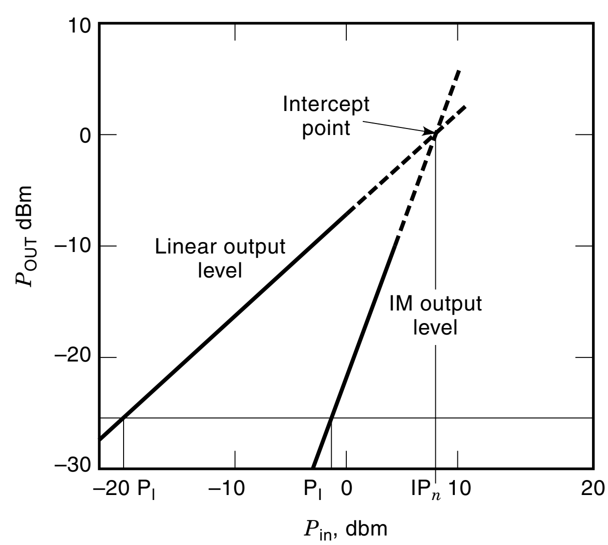


Figure 11. The output level of each nth-order IM product varies n decibels for every decibel change in input level. The intercept point is the extrapolated point at which the curves intersect.

Given the intercept point IP_n and input power level in decibels, the IM input level P_I in decibels can be found from

$$P_I = \frac{1}{n}P_1 + \left(1 - \frac{1}{n}\right) IP_n \tag{12}$$

where P_l is the input level of each of the linear RF tones (which are assumed to be equal) in decibels. By convention, P_l and P_I are the input powers of a single frequency component where the linear output level and the level of the nth-order IM product are equal; They are not the total power of all components. For example, P_l is the threshold level for a receiver. The fluctuation of the IMD level is rather small in spite of the fluctuations of P_l and IP_n.

Spurious Responses

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

$$f_{IF} = f_{RF} - f_{LO} \tag{13}$$

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

$$f_{IF} = mf_{RF} - nf_{LO} \tag{14}$$

where m and n are integers. If an RF signal 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. Most singly balanced mixers reject some, but not all, products where m or n (or both) are even.

Harmonic Mixer

A mixer is sensitive to many frequencies besides those at which it is designed to operate. The best known of these is the image frequency, which is found at the LO sideband opposite the input, of the RF frequency. The mixer is also sensitive to similar sidebands on either side of each LO harmonic. These responses are usually undesired; the exception is the harmonic mixer, which is designed to operate at one or more of these sidebands.

When a small-signal voltage is applied to the pumped diode at any one of these frequencies, currents and voltages are generated in the junction at all other sideband frequencies. These frequencies are called the small-signal mixing frequencies ω_n and are given by the relation

$$\omega_n = \omega_0 + n\omega_p \tag{15}$$

where ω_p is the LO frequency and

$$n = \cdots, -3, -2, -1, 0, 1, 2, 3, \ldots \tag{16}$$

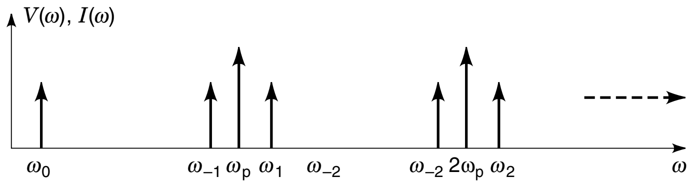


Figure 12. Small-signal mixing frequencies ω_n and LO harmonics $n\omega_p$. Voltage and current components exist in the diode at these frequencies.

These frequencies are shown in Fig. 12. The frequencies are separated from each LO harmonic by ω_0, the difference between the LO frequency and the RF.

MODULATION AND FREQUENCY TRANSLATION

Modulation

Modulation is the process by which the information content of an audio, video, or data signal is transferred to an RF carrier before transmission. Commonly, the signal being modulated is a sine wave of constant amplitude and is referred to as the carrier. The signal that varies some parameter of the carrier is known as the modulation signal. The parameters of a sine wave that may be varied are the amplitude, the frequency, and the phase. Other types of modulation may be applied to special signals, e.g., pulse-width and pulse-position modulation of recurrent pulses. The inverse process—recovering the information from an RF signal—is called demodulation or detection. In its simpler forms a modulator may cause some characteristic of an RF signal to vary in direct proportion to the modulating waveform: this is termed analog modulation. More complex modulators digitize and encode the modulating signal before modulation. For many applications digital modulation is preferred to analog modulation.

A complete communication system (Fig. 13) consists of an information source, an RF source, a modulator, an RF channel (including both transmitter and receiver RF stages, the antennas, the transmission path, etc.), a demodulator, and an information user. The system works if the information user receives the source information with acceptable reliability. The designer's goal is to create a low-cost working system that complies with the legal restrictions on such things as transmitter power, antenna height, and signal bandwidth. Since modulation demodulation schemes differ in cost, bandwidth, interference rejection, power consumption, and so forth, the choice of the modulation type is an important part of communication system design.

Modulation, demodulation (detection), and heterodyne action are very closely related processes. Each process involves generating the sum and/or difference frequencies of two or more sinsuoids by causing one signal to vary as a direct function (product) of the other signal or signals. The multiplication of one signal by another can only be accomplished in a nonlinear device. This is readily seen by considering any network where the output signal is some function of the input signal e_1, for example,

$$e_0 = f(e_1) \tag{17}$$

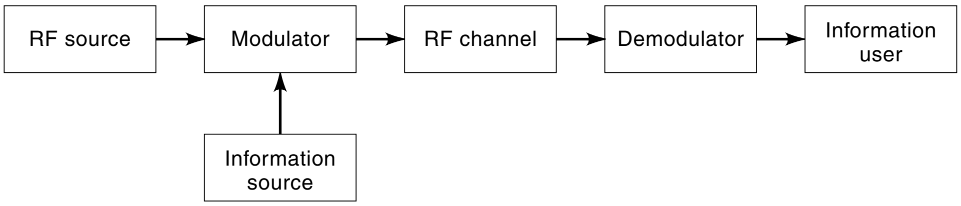


Figure 13. Conceptual diagram of a communication system.

In any perfectly linear network, this requires that

$$e_0 = ke_1 \tag{18}$$

and, assuming two different input signals,

$$e_0 = k(E_a \cos \omega_a t + E_b \cos \omega_b t) \tag{19}$$

where k is a constant. In this case the output signal contains only the two input-signal frequencies. However, if the output is a nonlinear function of the input, it can, in general, be represented by a series expansion of the input signal. For example, let

$$e_0 = k_1 e_1 + k_2 e_2^2 + k_3 e_3^3 + \cdots + k_n e_n^n \tag{20}$$

When e_1 contains two frequencies, e_0 will contain the input frequencies and their harmonics plus the products of these frequencies. These frequency products can be expressed as sum and difference frequencies. Thus, all modulators, detectors, and mixers are of necessity nonlinear devices. The principal distinction between these devices is the frequency differences between the input signals and the desired output signal or signals. For example, amplitude modulation in general involves the multiplication of a high-frequency carrier by low-frequency modulation signals to produce sideband signals near the carrier frequency. In a mixer, two high-frequency signals are multiplied to produce an output signal at a frequency that is the difference between the input-signal frequencies. In a detector for amplitude modulation, the carrier is multiplied by the sideband signals to produce their different frequencies at the output.

To understand the modulation process, it is helpful to visualize a modulator as a black box (Fig. 14) with two inputs and one output connected to a carrier oscillator producing a sinusoidal voltage with constant amplitude and frequency f_{RF}. The output is a modulated waveform

$$F(t) = A(t)\cos[\omega_s t + \Theta(t)] = A(t)\cos\Phi(t) \tag{21}$$

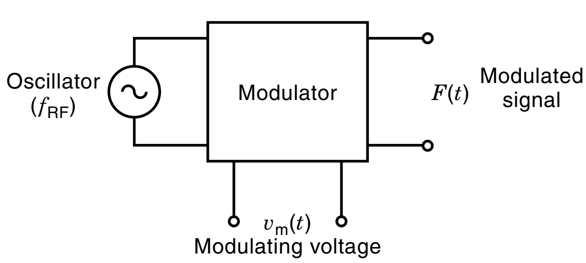


Figure 14. Black-box view of a modulator.

whose amplitude $A(t)$ or angle $\Phi(t)$, or both, are controlled by $v_m(t)$. In amplitude modulation (AM) the carrier envelope $A(t)$ is varied while $\Theta(t)$ remains constant; in angle modulation $A(t)$ is fixed and the modulating signal controls $\Phi(t)$. Angle modulation may be either frequency modulation (FM) or phase modulation (PM), depending upon the relationship between the angle $\Phi(t)$ and the modulation signal.

Although the waveform (21) might be called a modulated cosine wave, it is not a single-frequency sinusoid when modulation is present. If either $A(t)$ or $\Theta(t)$ varies with time, the spectrum of $F(t)$ will occupy a bandwidth determined by both the modulating signal and the type of modulation used.

Amplitude Modulation. Amplitude modulation in the form of on–off keying of radio-telegraph transmitters is the oldest type of modulation. Today, amplitude modulation is widely used for those analog voice applications that require simple receivers (e.g., commercial broadcasting) and require narrow bandwidths.

In amplitude modulation the instantaneous amplitude of the carrier is varied in proportion to the modulating signal. The modulating signal may be a single frequency, or, more often, it may consist of many frequencies of various amplitudes and phases, e.g., the signals constituting speech. For a carrier modulated by a single-frequency sine wave of constant amplitude, the instantaneous signal $e(t)$ is given by

$$e(t) = E(1 + m\cos\omega_m t)\cos(\omega_c t + \phi) \tag{22}$$

where E is the peak amplitude of unmodulated carrier, m is the modulation factor as defined below, ω_m is the frequency of the modulating voltage (radians per second), ω_c is the carrier frequency (radians per second), and ϕ is the phase angle of the carrier (radians).

The instantaneous carrier amplitude is plotted as a function of time in Fig. 15. The modulation factor m is defined for asymmetrical modulation in the following manner:

$$m = \frac{E_{max} - E}{E} \qquad \text{(upward or positive modulation)} \tag{23}$$

$$m = \frac{E - E_{min}}{E} \qquad \text{(downward or negative modulation)} \tag{24}$$

The maximum downward modulation factor, 1.0, is reached when the modulation peak reduces the instantaneous carrier envelope to zero. The upward modulation factor is unlimited.

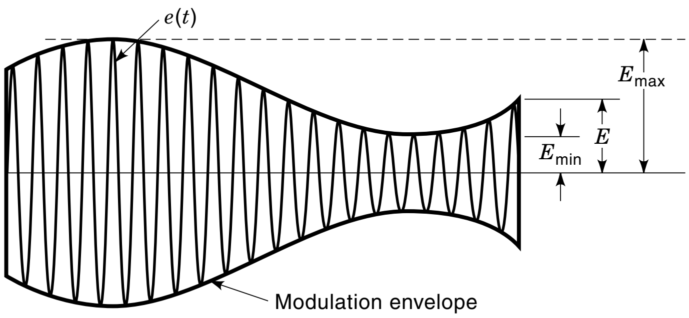


Figure 15. Amplitude-modulated carrier.

The modulation carrier described by Eq. (22) can be rewritten as follows:

$$\begin{aligned} e(t) &= E(1 + m\cos\omega_{\rm m}t)\cos(\omega_{\rm c}t + \phi) \\ &= E\cos(\omega_{\rm c}t + \phi) + \frac{mE}{2}\cos[(\omega_{\rm c} + \omega_{\rm m})t + \phi] \\ &\quad + \frac{mE}{2}\cos[(\omega_{\rm c} - \omega_{\rm m})t + \phi] \end{aligned} \tag{25}$$

Thus, the amplitude modulation of a carrier by a cosine wave has the effect of adding two new sinusoidal signals displaced in frequency from the carrier by the modulating frequency. The spectrum of the modulated carrier is shown in Fig. 16.

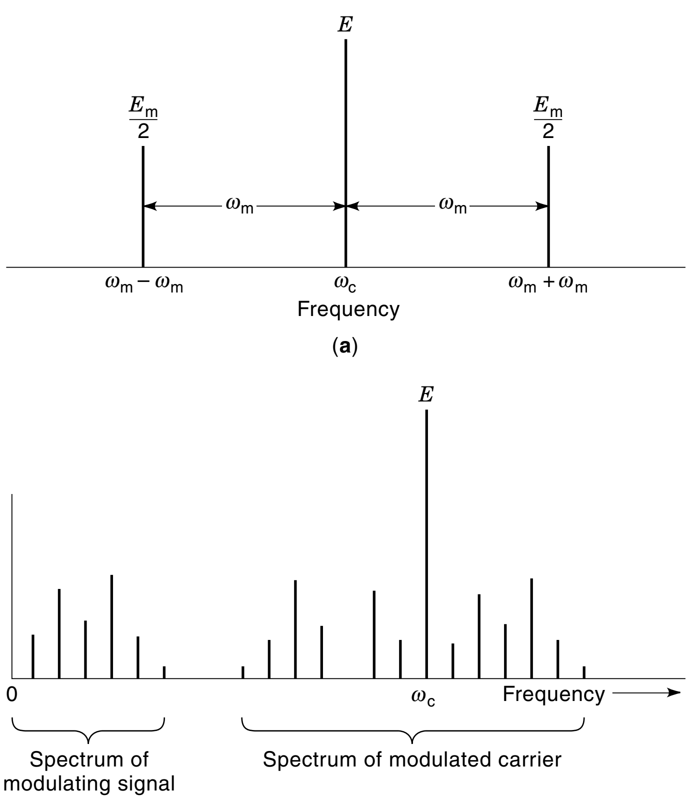


Figure 16. Frequency spectrum of an amplitude-modulated carrier: (a) carrier modulated by a sinusoid of frequency ω_m, (b) carrier modulated by a complex signal composed of several sinusoids.

Angle Modulation. Information can be transmitted on a carrier by varying any of the parameters of the sinusoid in accordance with the modulating voltage. Thus, a carrier is described by

$$e(t) = E_{\rm c}\cos\theta \tag{26}$$

where $\theta = \omega_c t + \phi$.

This carrier can be made to convey information by modulating the peak amplitude E_c or by varying the instantaneous phase angle θ of the carrier. This type of modulation is known as angle modulation. The two types of angle modulation that have practical application are phase modulation (PM) and frequency modulation (FM).

In phase modulation, the instantaneous phase angle θ of the carrier is varied by the amplitude of the modulating signal. The principal application of phase modulation is in the utilization of modified phase modulators in systems that transmit frequency modulation. The expression for a carrier phase-modulated by a single sinusoid is given by

$$e(t) = E_{\rm c}\cos(\omega_{\rm c}t + \phi + \Delta\phi\cos\omega_{\rm m}t) \tag{27}$$

where $\Delta\phi$ is the peak value of phase variation introduced by modulation and is called the phase deviation, and ω_m is the modulation frequency (radians per second).

In frequency modulation, the instantaneous frequency of the carrier, that is, the time derivative of the phase angle θ, is made to vary in accordance with the amplitude of the modulating signal. Thus,

$$f = \frac{1}{2\pi}\frac{d\theta}{dt} \tag{28}$$

When the carrier is frequency-modulated by a single sinusoid,

$$f = f_{\rm RF} + \Delta f\cos\omega_{\rm m}t \tag{29}$$

where Δf is the peak frequency deviation introduced by modulation. The instantaneous total phase angle θ is given by

$$\theta = 2\pi\int f\,dt + \theta_0 \tag{30}$$

$$\theta = 2\pi f_{\rm RF}t + \frac{\Delta f}{f_{\rm m}}\sin 2\pi f_{\rm m}t + \theta_0 \tag{31}$$

The complete expression for a carrier that is frequency-modulated by a single sinusoid is

$$e(t) = E_{\rm c}\cos\left(\omega_c^t + \frac{\Delta f}{f_{\rm m}}\sin 2\pi f_{\rm m}t + \theta_0\right) \tag{32}$$

The maximum frequency difference between the modulated carrier and the unmodulated carrier is the frequency deviation Δf. The ratio of Δf to the modulation frequency f_m is known as the modulation index or the deviation ratio. The degree of modulation in an FM system is usually defined as the ratio of Δf to the maximum frequency deviation of which the system is capable. Degree of modulation in an FM system is therefore not a property of the signal itself.

In digital wireless communication systems, Gaussian-filtered minimum-shift keying (GMSK) is the most popular, and four-level frequency-shift keying (4-FSK) and $\pi/4$-shifted differential encoded quadriphase (or quadrature) phase-shift keying (π/4-DQPSK) are also used. GMSK and 4-FSK are both frequency modulation, but π/4-DQPSK is phase modulation.

Pulse Modulation. In pulse-modulated systems, one or more parameters of the pulse are varied in accordance with a modulating signal to transmit the desired information. The modulated pulse train may in turn be used to modulate a carrier in either angle or amplitude. Pulse modulation provides a method of time duplexing, since the entire modulation information of a signal channel can be contained in a single pulse train having a low duty cycle, i.e., ratio of pulse width to interpulse period, and therefore the time interval between successive pulses of a particular channel can be used to transmit pulse information from other channels.

Pulse-modulation systems can be divided into two basic types: pulse modulation proper, where the pulse parameter which is varied in accordance with the modulating signal is a continuous function of the modulating signal, and quantized pulse modulation, where the continuous information to be transmitted is approximated by a finite number of discrete values, one of which is transmitted by each single pulse or group of pulses. The two methods are illustrated in Fig. 17.

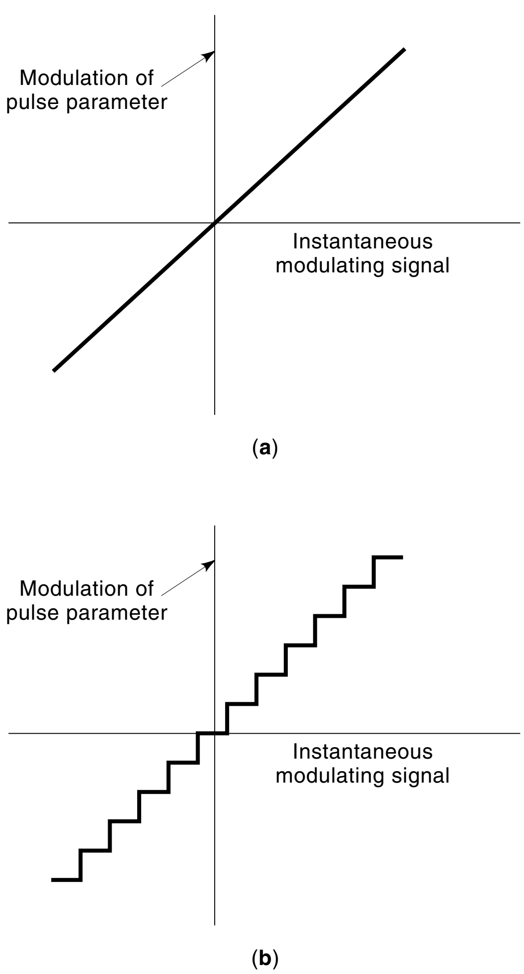


Figure 17. Input versus output relationships of quantized and unquantized pulse-modulation systems: (a) unquantized modulation system, (b) quantized modulation system.

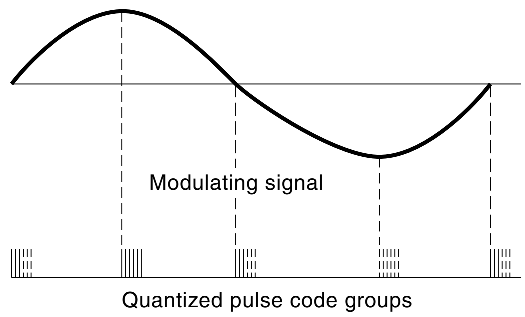


Figure 18. Example of a quantized pulse-modulation system.

In quantized pulse modulation systems, the input function can be approximated with arbitrary accuracy by increase of the number of discrete values available to describe the input function. An example of a quantized pulse modulation system is shown in Fig. 18: the information is transmitted in pulse code groups, the sequence of pulses sent each period indicating a discrete value of the modulating signal at that instant. Typically, the pulse group might employ a binary number code, the presence of each pulse in the group indicating a 1 or 0 in the binary representation of the modulating signal.

The principal methods for transmitting information by means of unquantized pulse modulation are pulse-amplitude modulation (PAM; see Fig. 19), pulse-width modulation (PWM), and pulse-position modulation (PPM).

Frequency Translation

The most common form of radio receiver is the superheterodyne configuration shown in Fig. 20(a). The signal input, with a frequency ω_s, is usually first amplified in a tunable bandpass amplifier, called the RF amplifier, and is then fed into a circuit called the mixer along with an oscillator signal, which is local to the receiver, having a frequency ω_p. The LO is also

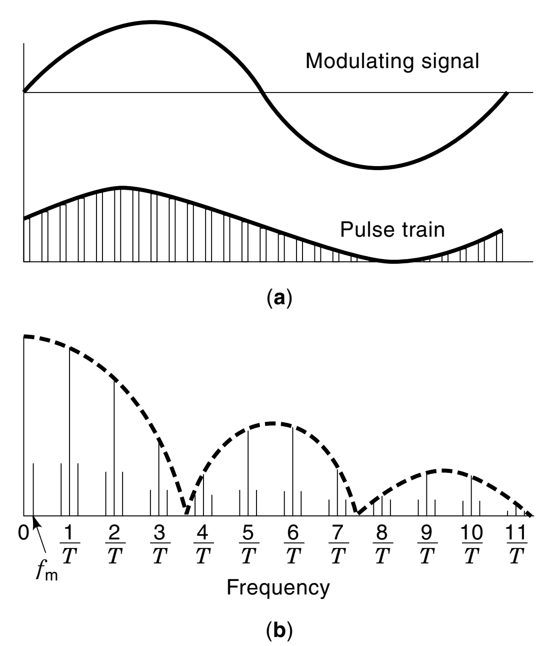


Figure 19. Pulse-amplitude modulation: (a) amplitude-modulated pulse train, (b) frequency spectrum of the modulated pulse train.

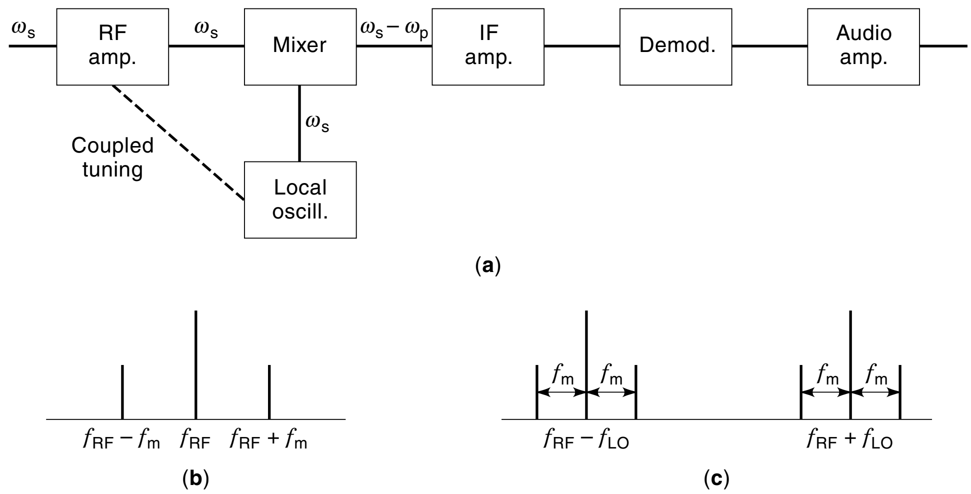


Figure 20. (a) The superheterodyne configuration; frequency spectra of (b) the input and (c) the multiplier output.

tunable and is ganged with the input bandpass amplifier so that the difference between the input signal frequency and that of the LO is constant.

In operation, the mixer must achieve analog multiplication. With multiplication, sum and difference frequency components at $\omega_s \pm \omega_p$ are produced at the output of the mixer. Usually, the sum frequency is rejected by sharply tuned circuits and the difference frequency component is subsequently amplified in a fixed-tuned bandpass amplifier. The difference frequency is called the intermediate frequency (IF), and the fixed-tuned amplifier is called the IF amplifier. The advantage of this superheterodyne configuration is that most amplification and outband rejection occurs with fixed-tuned circuits, which can be optimized for gain level and rejection. Another advantage is that the fixed-tuned amplifier can provide a voltage-controlled gain to achieve automatic gain control (AGC) with input signal level. In high-performance and/or small-size receivers, the filtering in the IF amplifier is obtained with electromechanical crystal filters.

To formalize the mixer operation, assume that both the input signal and the local oscillator output are unmodulated, single-tone sinusoids:

$$V_s = E_s \cos(\omega_s t) \tag{33}$$

$$V_p = E_p \cos(\omega_p t) \tag{34}$$

If the multiplier (mixer) has a gain constant K, the output is

$$V_o = \frac{K}{2} E_s E_p [\cos(\omega_s - \omega_p)t + \cos(\omega_s + \omega_p)t] \tag{35}$$

The difference frequency, $\omega_s - \omega_p$, is denoted by ω_{if}.

If the input is a modulated signal, the modulation also is translated to a band about the new carrier frequency, ω_{if}. For example, if the input is amplitude-modulated,

$$\begin{aligned} V_s &= E_s(1 + m \cos \omega_m t) \cos \omega_s t \\ &= E_s \cos(\omega_s t) + \frac{m}{2} E_s \cos(\omega_s - \omega_m)t \\ &\quad + \frac{m}{2} E_p \cos(\omega_s + \omega_m)t \end{aligned} \tag{36}$$

The input can be represented as in Fig. 20(b), with the carrier frequency term and an upper and a lower sideband, each sideband containing the modulation information.

For a linear multiplier, each of the input components is multiplied by the LO input, and the output of the multiplier contains six terms, as shown in Fig. 20(c): the difference-frequency carrier with two sidebands and the sum-frequency carrier with two sidebands. The latter combination is usually rejected by the bandpass of the IF amplifier.

ANALOG MULTIPLICATION

An analog multiplier can be used as a mixer. A multiplier inputs two electrical quantities, usually voltages but sometimes currents, and outputs the product of the two inputs, usually currents but sometimes voltages. The product of two quantities is derived from only the second-order term of the transfer characteristic of the element, because the product xy can be derived from only the second term of $(x + y)^2$. The second-order term is, for example, obtained from the inherent exponential law for a bipolar transistor or the inherent square law for a MOS transistor.

There are three methods of realizing analog multipliers: the first is by cross-coupling two variable-gain cells, the second is by cross-coupling two squaring circuits, and the third is by using a multiplier core. Block diagrams of these three multiplication methods are shown in Fig. 21(a–c). For example, the bipolar doubly balanced differential amplifier, the so-called *Gilbert cell,* is the first case, and utilizes two-quadrant analog multipliers as variable-gain cells. The second method has been known for a long time and is called the quarter-square technique. The third method is also based on the quarter-square technique, because a multiplier core is a cell consisting of the four properly combined squaring circuits.

Multipliers Consisting of Two Cross-Coupled Variable-Gain Cells

The Gilbert Cell. The Gilbert cell, shown in Fig. 22, is the most popular analog multiplier, and consists of two cross-coupled, emitter-coupled pairs together with a third emitter-coupled pair. The two cross-coupled, emitter-coupled pairs form a multiplier cell. The Gilbert cell consists of two cross-coupled

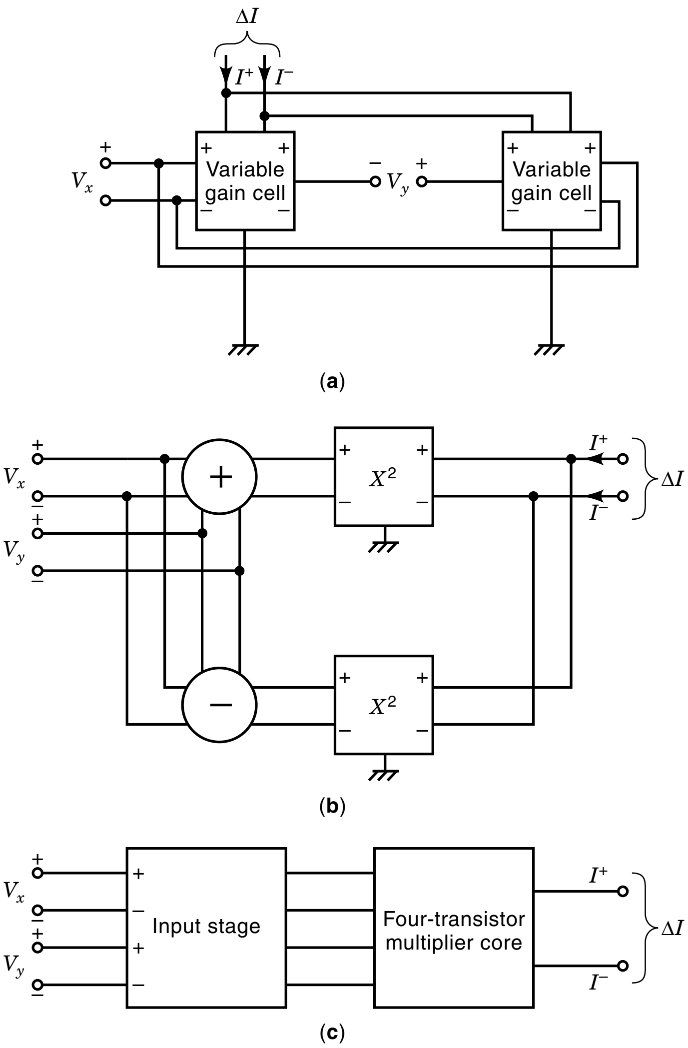


Figure 21. Multiplier block diagrams: (a) built from two cross-coupled variable-gain cells, (b) built from two cross-coupled squaring circuits, (c) built from a multiplier core and an input system.

variable-gain cells, because the lower emitter-coupled pair varies the transconductance of the upper cross-coupled, emitter-coupled pairs.

Assuming matched devices, the differential output current of the Gilbert cell is expressed as

$$\begin{aligned}\Delta I = I^{+} - I^{-} &= (I_{C13} + I_{C15}) - (I_{C14} + I_{C16}) \\ &= \alpha_F^2 I_0 \tanh\left(\frac{V_x}{2V_T}\right) \tanh\left(\frac{V_y}{2V_T}\right)\end{aligned} \tag{37}$$

where α_F is the dc common-base current gain factor.

The differential output current of the Gilbert cell is expressed as a product of two hyperbolic tangent functions. Therefore, the operating input voltage ranges of the Gilbert cell are both very narrow. Many circuit design techniques for linearizing the input voltage range of the Gilbert cell have been discussed to achieve wider input voltage ranges.

In addition, the Gilbert cell has been applied to ultra-high-frequency (UHF) bands of some tens of gigahertz using GaAs heterojunction bipolar transistor (HBT) and InP HBT technologies. The operating frequency of the Gilbert cell was 500 MHz at most in the 1960s.

The series connection of the two cross-coupled, emitter-coupled pairs with a third emitter-coupled pair requires a high supply voltage, more than 2.0 V. Therefore, many circuit design techniques for linearizing the low-voltage Gilbert cell have also been discussed.

Modified Gilbert Cell with a Linear Transconductance Amplifier. The modified Gilbert cell with a linear transconductance amplifier in Fig. 23 possesses a linear transconductance characteristic only with regard to the second input voltage V_y, because it utilizes a linear transconductance amplifier for the lower stage. Low-voltage operation is also achieved using the differential current source output system of two emitter-follower-augmented current mirrors. The general structure of the mixer is a Gilbert cell with a linear transconductance amplifier, since the cross-coupled emitter-coupled pairs that input the LO signal possess a limiting characteristic. To achieve the desired low distortion, the differential pair normally used as the lower stage of the cell is replaced with a superlinear transconductance amplifier. In practice, the linear input voltage range of the superlinear transconductance amplifier at a 1.9 V supply voltage is 0.9 V peak to peak for less than 1% total harmonic distortion (THD) or 0.8 V for less than 0.1% THD.

The differential output current of the modified Gilbert cell with a linear transconductance amplifier is

$$\begin{aligned}\Delta I = I^{+} - I^{-} &= (I_{C1} + I_{C3}) - (I_{C2} + I_{C4}) \\ &= 2G_y V_y \tanh\left(\frac{V_x}{2V_T}\right)\end{aligned} \tag{38}$$

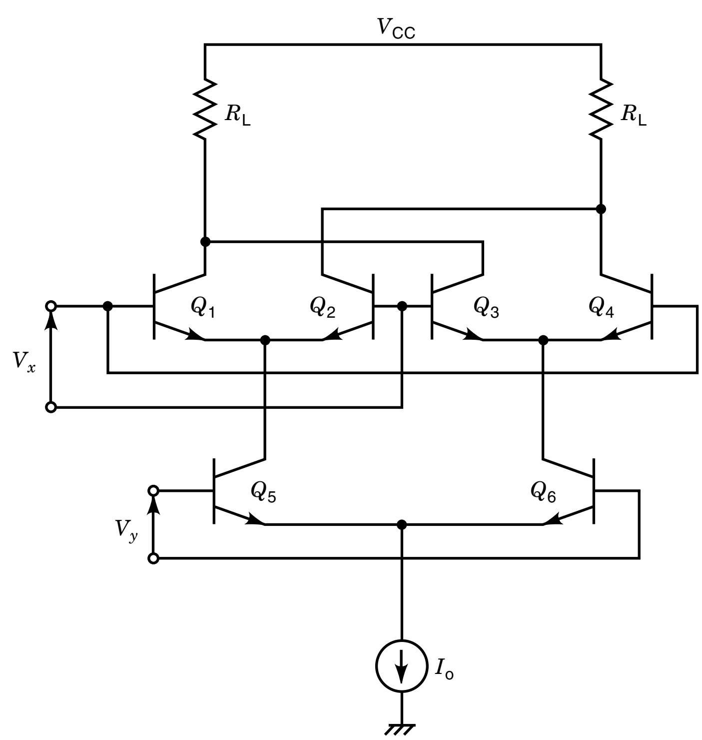


Figure 22. Gilbert cell.

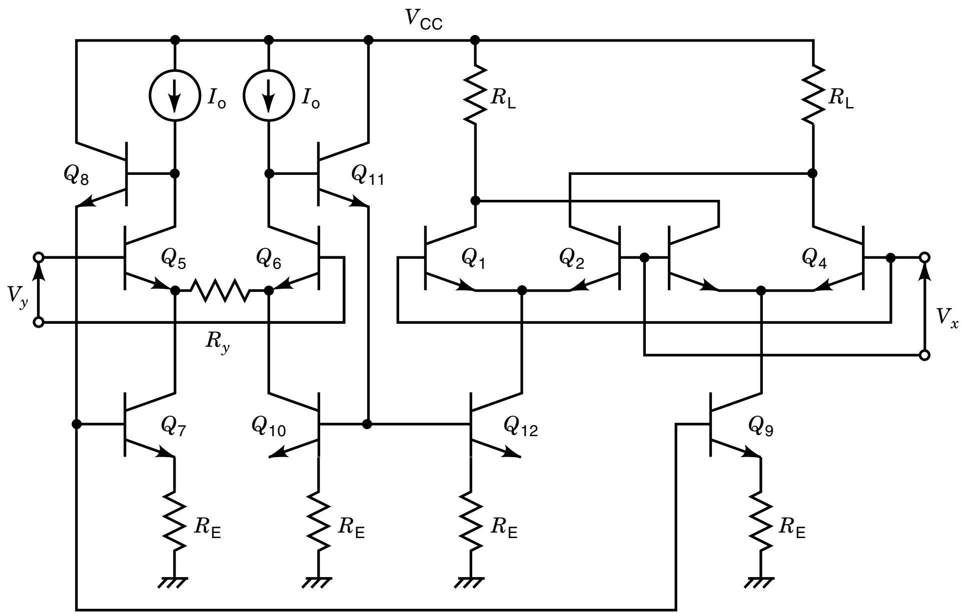


Figure 23. Modified Gilbert cell with a linear transconductance amplifier.

where $G_y = 1/R_y$ and the dc common-base current gain factor α_F is taken as equal to one for simplification, since its value is 0.98 or 0.99 in current popular bipolar technology.

The product of the hyperbolic tangent function of the first input voltage and the second input voltage of the linear transconductance amplifier is obtained.

Quarter-Square Multipliers Consisting of Two Cross-Coupled Squaring Circuits

To realize a multiplier using squaring circuits the basic idea is based on the identity $(x + y)^2 - (x - y)^2 = 4xy$ or $(x + y)^2 - x^2 - y^2 = 2xy$. The former identity is usually expressed as

$$\tfrac{1}{4}[(x+y)^2 - (x-y)^2] = xy \tag{39}$$

The quarter-square technique based on the above identity has been well known for a long time.

The two input voltage ranges and the linearity of the transconductances of the quarter-square multiplier usually depend on the square-law characteristics of the squaring circuits and sometimes depend on the linearities of the adder and subtractor in the input stage. A quarter-square multiplier does not usually possess limiting characteristics with regard to both inputs.

Four-Quadrant Analog Multipliers with a Multiplier Core

The multiplier core can be considered as four properly combined square circuits. The multiplication is based on the identity

$$\begin{aligned}(ax+by)^2 &+ \left[(a-c)x + \left(b - \frac{1}{c}\right)y\right]^2 \\ &- [(a-c)x+by]^2 - \left[ax + \left(b-\frac{1}{c}\right)y\right]^2 = 2xy\end{aligned} \tag{40}$$

where a, b, and c are constants.

If each squaring circuit is a square-law element with another parameter z, the identity becomes

$$\begin{aligned}(ax+by+z)^2 &+ \left[(a-c)x + \left(b-\frac{1}{c}\right)y + z\right]^2 \\ &- [(a-c)x+by+z]^2 \\ &- \left[ax + \left(b-\frac{1}{c}\right)y + z\right]^2 = 4xy\end{aligned} \tag{41}$$

In Eqs. (40) and (41), the parameters a, b, c, and z can be canceled out.

MOS transistors operating in the saturation region can be used as square-law elements. Four properly arranged MOS transistors with two properly combined inputs produce the product of two inputs in accordance with Eq. (22). Also, four properly arranged bipolar transistors with two properly combined inputs produce the product of the hyperbolic functions of the inputs. A cell consisting of four emitter- or source-common transistors biased by a single cell tail current can be used as a multiplier core.

Bipolar Multiplier Core. Figure 24(a) shows a bipolar multiplier core. The individual input voltages applied to the bases of the four transistors in the core can be expressed as $V_1 = aV_x + bV_y + V_R$, $V_2 = (a - 1)V_x + (b - 1)V_y + V_R$, $V_3 = (a - 1)V_x + bV_y + V_R$, $V_4 = aV_x + (b - 1)V_y + V_R$. The differential output current is expressed as

$$\begin{aligned}\Delta I = I^+ - I^- &= (I_{C1} + I_{C2}) - (I_{C3} + I_{C4}) \\ &= \alpha_F I_0 \tanh\left(\frac{V_x}{2V_T}\right)\tanh\left(\frac{V_y}{2V_T}\right)\end{aligned} \tag{42}$$

The parameters a and b are canceled out. The transfer function of the bipolar multiplier core is expressed as the product of the two transfer functions of the emitter-coupled pairs. The difference between Eq. (42) and Eq. (38) is only in whether the tail current value is multiplied by the parameter α_F or by

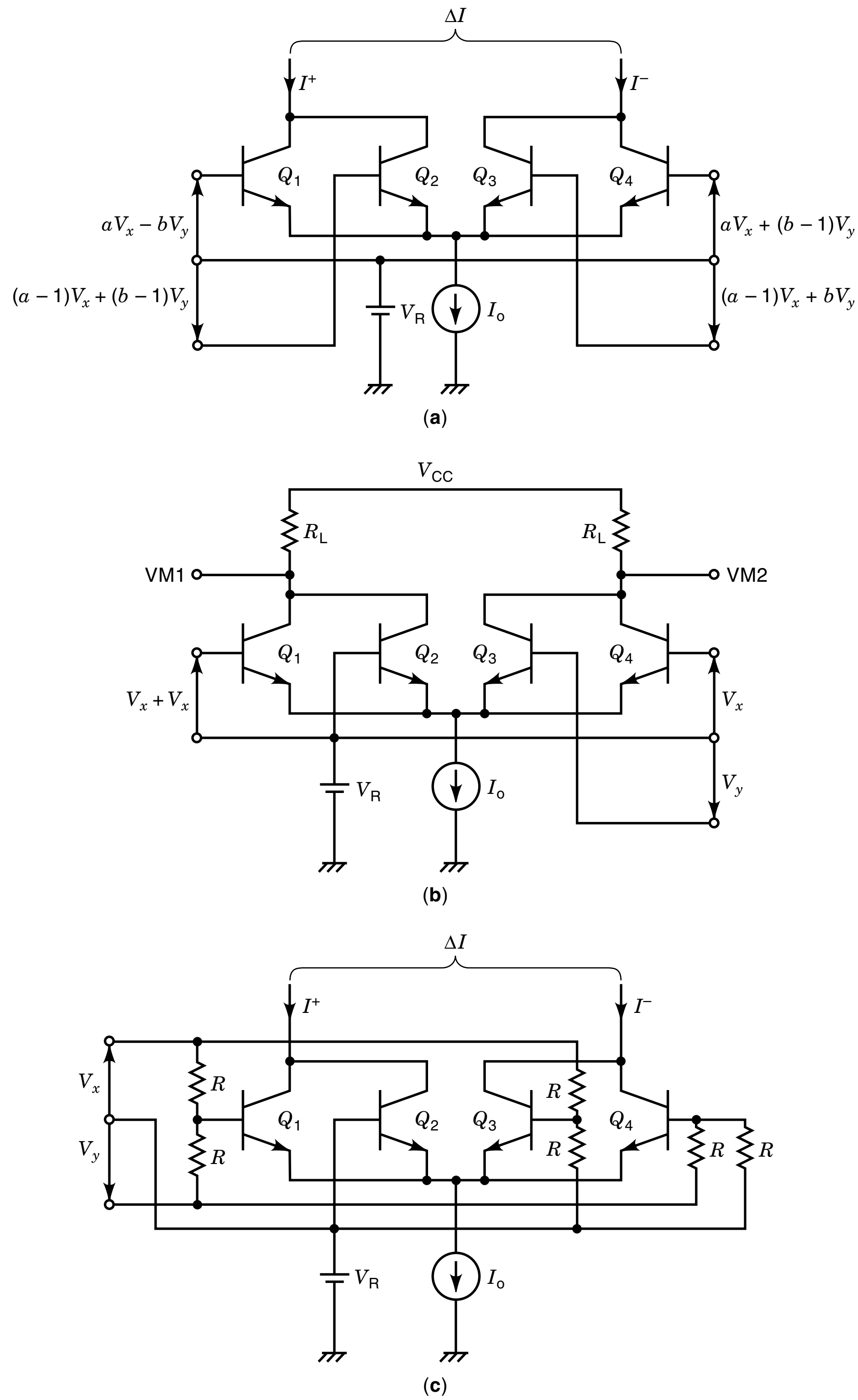
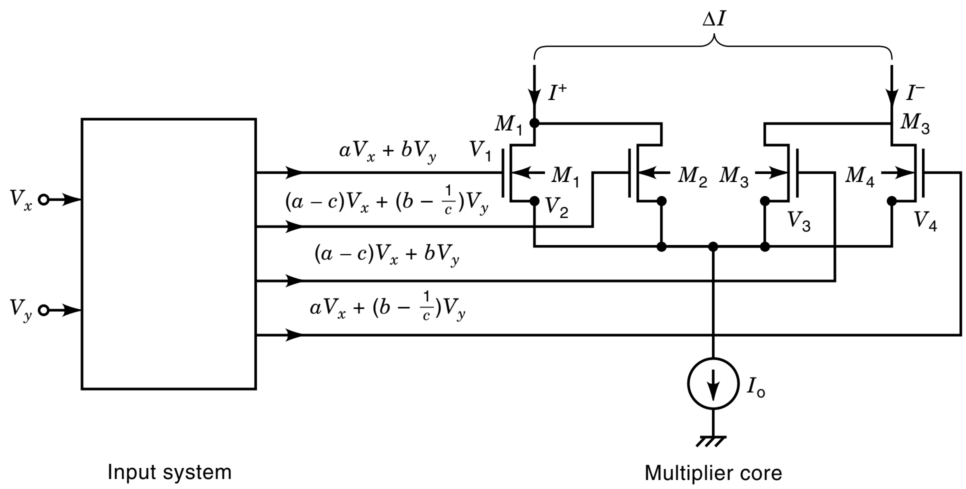
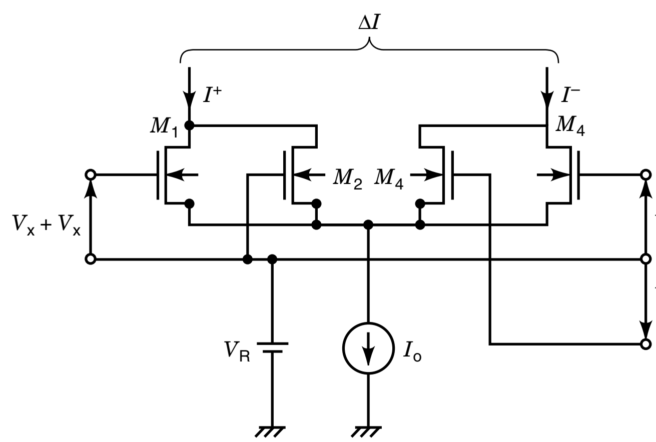


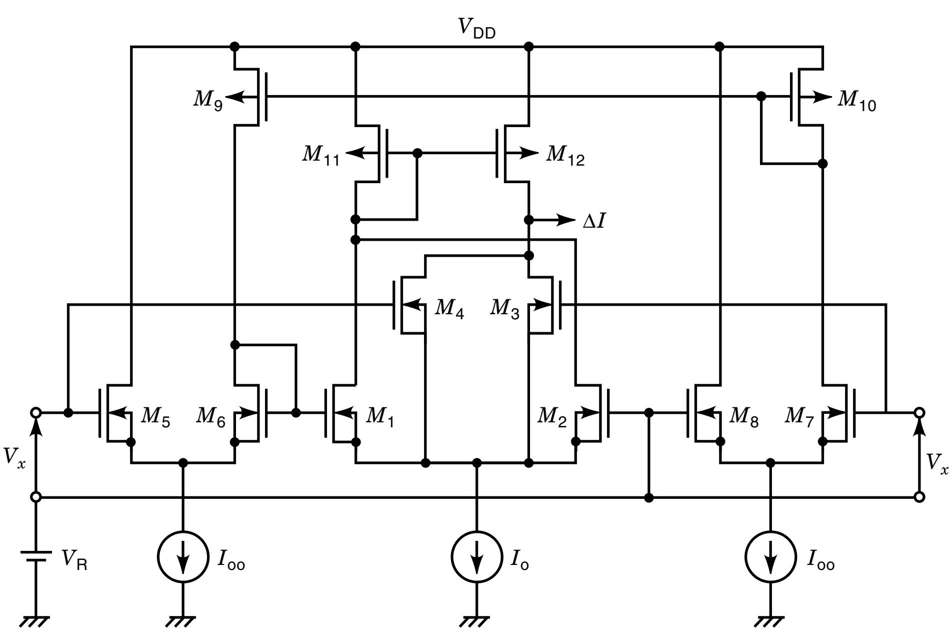
Figure 24. Bipolar multiplier: (a) general circuit diagram of core, (b) the core with the simplest combination of the two input voltages, (c) the bipolar multiplier consisting of a multiplier core and resistive dividers.



(**a**)



(**b**)



(**c**)

Figure 25. MOS multiplier: (a) general circuit diagram of core (b) the core with the simplest combination of the two input voltages, (c) MOS multiplier consisting of the multiplier core and an active voltage adder.

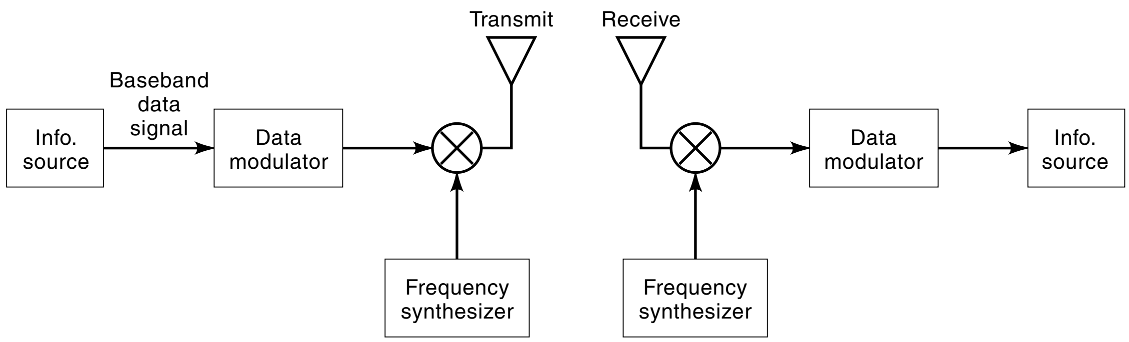


Figure 26. Block diagram of communications system, showing modulation and demodulation.

its square. Therefore, a bipolar multiplier core consisting of a quadritail cell is a low-voltage version of the Gilbert cell.

Simple combinations of two inputs are obtained when $a = b = \frac{1}{2}$, $a = \frac{1}{2}$ and $b = 1$, and $a = b = 1$ as shown in Fig. 24(b). In particular, when $a = b = 1$, resistive voltage adders are applicable because no inversion of the signals V_x and V_y is needed [Fig. 24(c)].

MOS Multiplier Core. Figure 25(a) shows the MOS four-quadrant analog multiplier consisting of a multiplier core. Individual input voltages applied to the gates of the four MOS transistors in the core are expressed as $V_1 = aV_x + bV_y + V_R$, $V_2 = (a - c)V_x + (b - 1/c)V_y + V_R$, $V_3 = (a - c)V_x + bV_y + V_R$, $V_4 = aV_x + (b - 1/c)V_y + V_R$. The multiplication is based on the identity of Eq. (41).

Ignoring the body effect and channel-length modulation, the equations for drain current versus drain-to-source voltage can be expressed in terms of three regions of operation as

$$I_D = 0 \tag{43a}$$

for $V_{GS} \leq V_T$, the off region,

$$I_D = 2\beta\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right)V_{DS} \tag{43b}$$

for $V_{DS} \leq V_{GS} - V_T$, the triode region, and

$$I_D = \beta(V_{GS} - V_T)^2 \tag{43c}$$

for $V_{GS} \geq V_T$ and $V_{DS} \geq V_{GS} - V_T$, the saturation region, where $\beta = \mu(C_O/2)(W/L)$ is the transconductance parameter, μ is the effective surface carrier mobility, C_O is the gate oxide capacitance per unit area, W and L are the channel width and length, and V_T is the threshold voltage.

The differential output current is expressed as

$$\begin{aligned} \Delta I = I^+ - I^- &= (I_{D1} + I_{D2}) - (I_{D3} + I_{D4}) \\ &= 2\beta V_x V_y \qquad (V_x^2 + V_y^2 + |V_x V_y| \leqq I_0/2\beta) \end{aligned} \tag{44}$$

The parameters a, b, and c are canceled out. Four properly arranged MOS transistors with two properly combined inputs produce the product of two input voltages. Simple combinations of two inputs are obtained when $a = b = \frac{1}{2}$ and $c = 1$, $a = \frac{1}{2}$ and $b = c = 1$, and $a = b = c = 1$ as shown in Fig. 25(b).

Figure 25(c) shows a CMOS four-quadrant analog multiplier consisting of only a multiplier core and an active voltage adder.

In addition, a multiplier consisting of the multiplier core in Fig. 25(a) and a voltage adder and subtractor has been implemented with a GaAs MESFET IC, and a useful frequency range from dc to UHF bands of 3 GHz was obtained for a frequency mixer operating on a supply voltage of 2 or 3 V.

RADIO-FREQUENCY SIGNAL AND LOCAL OSCILLATOR

Figure 26 shows a block diagram of a communication system, showing modulation and demodulation. A wireless communication system will usually consists of an information source, which is modulated up to RF or 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 Fig. 26. The rate at which information can be sent over the channel is determined by the available bandwidth, the modulation scheme, and the integrity of the modulation–demodulation process.

Frequency synthesizers are ubiquitous building blocks in wireless communication systems, since they produce the precise reference frequencies for modulation and demodulation of baseband signals up to the transmit and/or receive frequencies.

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

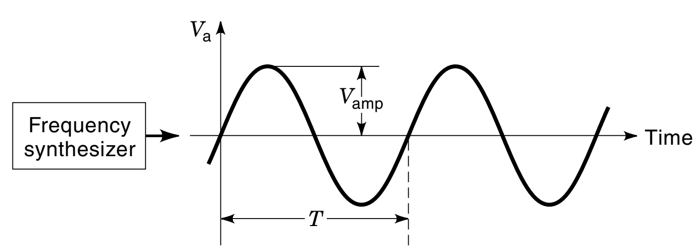


Figure 27. Block diagram of frequency synthesizer producing single-tone sinusoidal output.

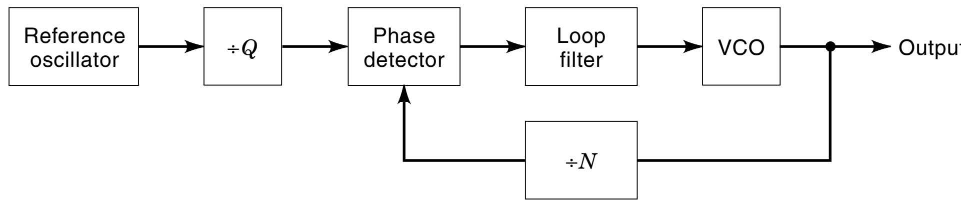


Figure 28. Indirect frequency synthesizer using a phase-locked loop.

rely on variations in the characteristics of a resonant circuit to set the frequency. These circuits can then be embedded in phase-locked loops (PLLs) to broaden their range of operation and further enhance their performance.

A representative view of a frequency synthesizer is given in Fig. 27 which shows a generic synthesizer producing a single tone of a given amplitude that has a delta-function-like characteristic in the frequency domain.

Indirect frequency synthesizers rely in feedback, usually in the form of the PLL, to synthesize the frequency. A block diagram of a representative PLL frequency synthesizer is shown in Fig. 28. 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). Thus, the PLL can act as a broadband frequency synthesizer.

FREQUENCY SYNTHESIZER FIGURES OF MERIT

An ideal frequency synthesizer would produce a perfectly pure sinusoidal signal, which would be 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. Instead, it is distributed around it, and the spectral distribution on either side of the carrier is known as the spectral sideband. This is illustrated schematically in Fig. 29. This noise can be represented as

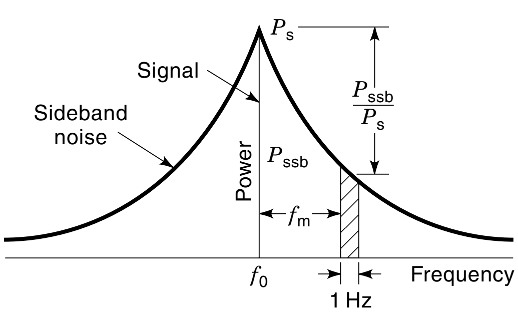


Figure 29. Phase noise specification of frequency source. The noise is contained in the sidebands around the signal frequency at f_0.

modulation of the carrier signal, and resolved into AM and 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 "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.

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 megahertz per volt. 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 achieve some minimum specified value.

Frequency Stability

Frequency stability of an oscillator is typically specified in parts per million per degree centigrade (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.

Harmonics

Harmonics are output 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 in "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 in "dBc" below the carrier.

BIBLIOGRAPHY

1. A. A. Abidi, Low-power radio-frequency IC's for portable communications, *Proc. IEEE,* **83**: 544–569, 1995.
2. L. E. Larson, *RF and Microwave Circuit Design for Wireless Communications,* Norwood, MA: Artech House, 1996.

3. N. Camilleri et al., Silicon MOSFETs, the microwave device technology for the 90s. In *1993 IEEE MTT-S Int. Microw. Symp. Dig.,* June 1993, pp. 545–548.
4. C. Tsironis, R. Meierer, and R. Stahlman, Dual-gate MESFET mixers, *IEEE Trans. Microw. Theory Tech.,* **MTT-32**: 248–255, March 1984.
5. S. A. Maas, *Microwave Mixers,* 2nd ed., Norwood, MA: Artech House, 1993.
6. J. M. Golio, *Microwave MESFETs & HEMTs,* Norwood, MA: Artech House, 1991.
7. F. Ali and A. Gupta (eds.), *HEMTs & HBTs: Devices, Fabrication, and Circuits,* Norwood, MA: Artech House, 1991.
8. D. Haigh and J. Everard, *GaAs Technology and its Impact on Circuits and Systems,* London: Peter Peregrinus, 1989.
9. D. O. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication—Principles, Simulation and Design,* Norwell, MA: Kluwer Academic, 1991.
10. W. Gosling, *R·A·D·I·O Receivers,* London: Peter Peregrinus, 1986.
11. K. Murota and K. Hirade, GMSK modulation for digital mobile telephony, *IEEE Trans. Commun.,* **COM-29**: 1044–1050, 1981.
12. Y. Akaiwa and Y. Nagata, Highly efficient digital mobile communications with a linear modulation method, *IEEE J. Selected Areas Commun.,* **SAC-5** (5): 890–895, June 1987.
13. J. Eimbinder, *Application Considerations for Linear Integrated Circuits,* New York: Wiley, 1970.
14. H. E. Jones, *Dual output synchronous detector utilizing transistorized differential amplifiers,* U.S. Patent No. 3,241,078, March 15, 1966.
15. B. Gilbert, A precise four-quadrant analog multiplier with subnanosecond response, *IEEE J. Solid-State Circuits,* **SC-3** (4): 365–373, 1968.
16. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits,* New York: Wiley, 1977, pp. 667–681.
17. K. W. Kobayashi et al., InAlAs/InGaAs HBT X-band double-balanced upconverter, *IEEE J. Solid-State Circuits,* **29** (10): 1238–1243, 1994.
18. F. Behbahani et al., A low distortion bipolar mixer for low voltage direct up-conversion and high IF frequency systems. *Proc. IEEE 1996 Bipolar Circuits Technol. Meet.,* Sept. 1996, pp. 50–52.
19. H. Song and C. Kim, An MOS four-quadrant analog multiplier using simple two-input squaring circuits with source-followers, *IEEE J. Solid-State Circuits,* **25** (3): 841–848, 1990.
20. K. Kimura, A unified analysis of four-quadrant analog multipliers consisting of emitter and source-coupled transistors operable on low supply voltage, *IEICE Trans. Electron.,* **E76-C** (5): 714–737, 1993.
21. K. Bult and H. Wallinga, A CMOS four-quadrant analog multiplier, *IEEE J. Solid-State Circuits,* **SC-21** (3): 430–435, 1986.
22. K. Kimura, An MOS four-quadrant analog multiplier based on the multitail technique using a quadritail cell as a multiplier core, *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.,* **42**: 448–454, 1995.
23. Z. Wang, A CMOS four-quadrant analog multiplier with single-ended voltage output and improved temperature performance, *IEEE J. Solid-State Circuits,* **26** (9): 1293–1301, 1991.
24. K. Kimura, A bipolar very low-voltage multiplier core using a quadritail cell, *IEICE Trans. Fundam.,* **E78-A** (5): 560–565, May 1995.
25. K. Kimura, Low voltage techniques for analog functional blocks using triple-tail cells, *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.,* **42**: 873–885, 1995.
26. R. Siferd, A GaAs four-quadrant analog multiplier circuit, *IEEE J. Solid-State Circuits,* **28** (3): 388–391, 1993.
27. B. Razavi, Challenges in the design of frequency synthesizers for wireless applications. *Proc. IEEE 1997 Custom Integrated Circuits Conf.,* May 1997, pp. 395–402.

KATSUJI KIMURA
NEC Corporation

MIXERS. See MULTIPLIERS, ANALOG; MULTIPLIERS, ANALOG CMOS.

MMIC AMPLIFIERS. See MICROWAVE LIMITERS.

MOBILE AGENTS. See MOBILE NETWORK OBJECTS.