Chapter 11

Mixers, Modulators and Demodulators

At base, radio communication involves translating information into radio form, letting it travel for a time, and translating it back again. Translating information into radio form entails the process we call *modulation*, and *demodulation* is its reverse. One way or another, every transmitter used for radio communication, from the simplest to the most complex, includes a means of modulation; one way or another, every receiver used for radio communication, from the simplest to the most complex, includes a means of demodulation.

Modulation involves varying one or both of a radio signal's basic characteristics — amplitude and frequency (or phase) — to convey information. A circuit, stage or piece of hardware that modulates is called a *modulator*.

Demodulation involves reconstructing the transmitted information from the changing characteristic(s) of a modulated radio wave. A circuit, stage or piece of hardware that demodulates is called a *demodulator*.

Many radio transmitters, receivers and transceivers also contain *mixers* — circuits, stages or pieces of hardware that combine two or more signals to produce additional signals at sums of and differences between the original frequencies. Amateur Radio textbooks have traditionally handled mixers separately from modulators and demodulators, and modulators separately from demodulators.

This chapter, by David Newkirk, ex-W9VES, and Rick Karlquist, N6RK, examines mixers, modulators and demodulators together because the job they do is essentially the same. Modulators and demodulators translate information into radio form and back again; mixers translate one frequency to others and back again. All of these translation processes can be thought of as forms of frequency translation or frequency shifting — the function traditionally ascribed to mixers. We'll therefore begin our investigation by examining what a mixer is (and isn't), and what a mixer does.

THE MECHANISM OF MIXERS AND MIXING

What is a Mixer?

Mixer is a traditional radio term for a circuit that shifts one signal's frequency up or down by combining it with another signal. The word *mixer* is also used to refer to a device used to blend multiple audio inputs together for recording, broadcast or sound reinforcement. These two mixer types differ in one very important way: A radio mixer makes new frequencies out of the frequencies put into it, and an audio mixer does not.

Mixing Versus Adding

Radio mixers might be more accurately called "frequency mixers" to distinguish them from devices such as "microphone mixers," which are really just signal *combiners*, *summers* or *adders*. In their most basic, ideal forms, both devices have two inputs and one output. The combiner simply *adds* the instantaneous voltages of the two signals together to produce the output at each point in time (**Fig 11.1**). The mixer, on the other hand, *multiplies* the instantaneous voltages of the two signals together to produce its output signal from instant to instant (**Fig 11.2**). Comparing the output spectra of the combiner and mixer, we see that the combiner's output contains only the frequencies of the two inputs, and nothing else, while the mixer's output contains *new* frequencies. Because it combines one energy with another, this process is sometimes called *heterodyning*, from the Greek words for *other* and *power*.

Mixing as Multiplication

Since a mixer works by means of multiplication, a bit of math can show us how they work. To begin with, we need to represent the two signals we'll mix, A and B, mathematically. Signal A's instantaneous amplitude equals

$$A_a \sin 2\pi f_a t \quad (1)$$

in which A is peak amplitude, f is frequency, and t is time. Likewise, B's instantaneous amplitude equals

$$A_b \sin 2\pi f_b t \quad (2)$$

Since our goal is to show that multiplying two signals generates sum and difference frequencies, we can simplify these signal definitions by assuming that the peak amplitude of each is 1. The equation for Signal A then becomes

$$a(t) = A \sin\left(2\pi f_a t\right) \quad (3)$$

and the equation for Signal B becomes

$$b(t) = B \sin\left(2\pi f_b t\right) \quad (4)$$

Each of these equations represents a sine wave and includes a subscript letter to help us keep track of where the signals go.

Merely combining Signal A and Signal

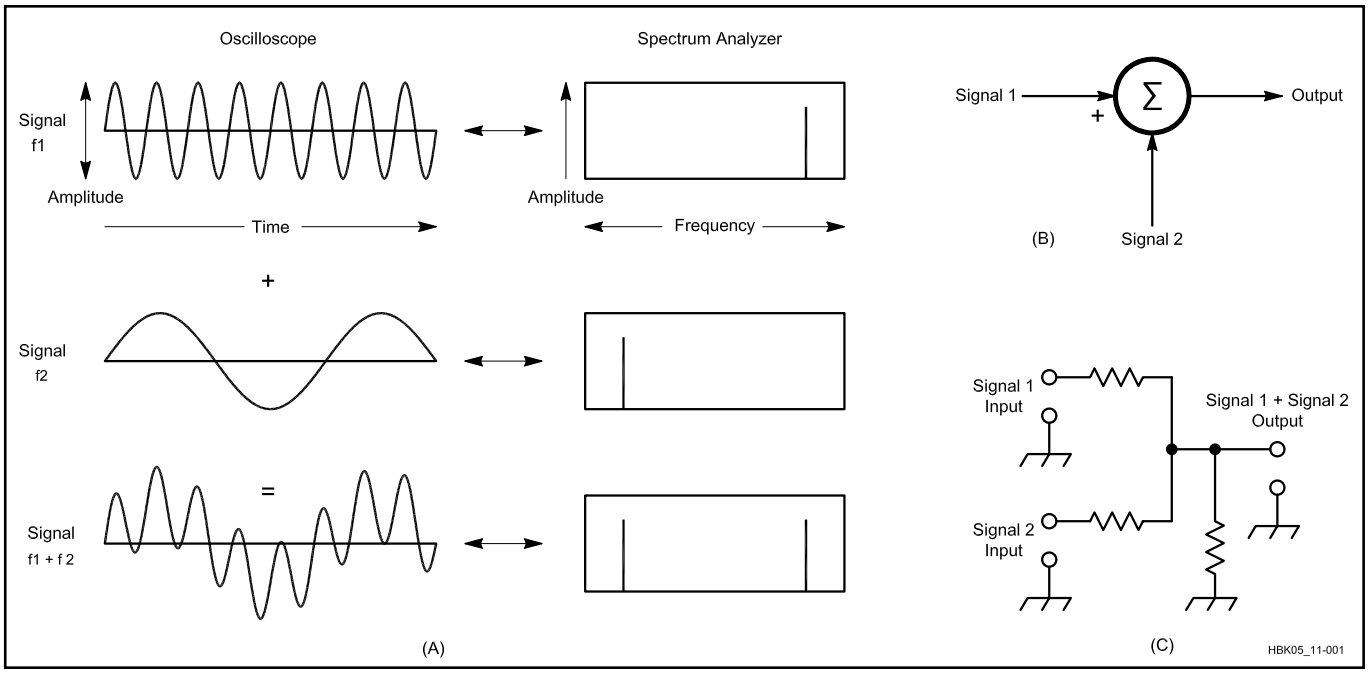


Fig 11.1 — *Adding or summing* two sine waves of different frequencies (f1 and f2) combines their amplitudes without affecting their frequencies. Viewed with an *oscilloscope* (a real-time graph of amplitude versus time), adding two signals appears as a simple superimposition of one signal on the other. Viewed with a *spectrum analyzer* (a real-time graph of signal amplitude versus frequency), adding two signals just sums their spectra. The signals merely coexist on a single cable or wire. All frequencies that go into the adder come out of the adder, and no new signals are generated. Drawing B, a block diagram of a summing circuit, emphasizes the stage's mathematical operation rather than showing circuit components. Drawing C shows a simple summing circuit, such as might be used to combine signals from two microphones. In audio work, a circuit like this is often called a mixer — but it does not perform the same function as an RF mixer.

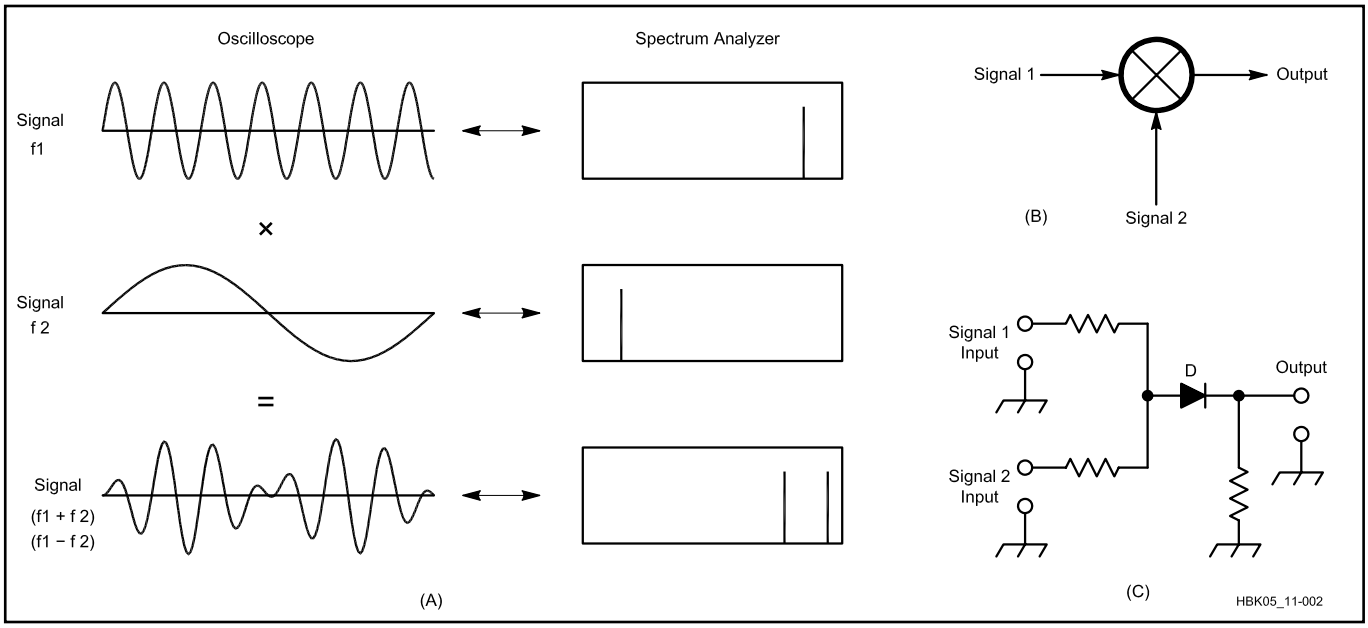


Fig 11.2 — *Multiplying* two sine waves of different frequencies produces a new output spectrum. Viewed with an oscilloscope, the result of multiplying two signals is a composite wave that seems to have little in common with its components. A spectrum-analyzer view of the same wave reveals why: The original signals disappear entirely and are replaced by two new signals — at the *sum* and *difference* of the original signals' frequencies. Drawing B diagrams a multiplier, known in radio work as a mixer. The *X* emphasizes the stage's mathematical operation. (The circled *X* is only one of several symbols you may see used to represent mixers in block diagrams, as Fig 11.3 explains.) Drawing C shows a very simple multiplier circuit. The diode, D, does the mixing. Because this circuit does other mathematical functions and adds them to the sum and difference products, its output is more complex than f1 + f2 and f1 – f2, but these can be extracted from the output by filtering.

B by letting them travel on the same wire develops nothing new:

$$a(t) + b(t) = A \sin(2\pi f_a t) + B \sin(2\pi f_b t) \quad (5)$$

As needlessly reflexive as equation 5 may seem, we include it to highlight the fact that multiplying two signals is a quite different story. From trigonometry, we know that multiplying the sines of two variables can be expanded according to the relationship

$$\sin x \sin y = \frac{1}{2}\left[\cos(x - y) - \cos(x + y)\right] \quad (6)$$

Conveniently, Signals A and B are both sinusoidal, so we can use equation 6 to determine what happens when we multiply Signal A by Signal B. In our case, $x = 2\pi f_a t$ and $y = 2\pi f_b t$, so plugging them into equation 6 gives us

$$a(t) \bullet b(t) = \frac{AB}{2}\cos(2\pi[f_a - f_b]t) - \frac{AB}{2}\cos(2\pi[f_a + f_b]t) \quad (7)$$

Now we see two momentous results: a sine wave at the frequency *difference* between Signal A and Signal B $2\pi(f_a - f_b)t$, and a sine wave at the frequency *sum* of Signal A and Signal B $2\pi(f_a + f_b)t$. (The products are cosine waves, but since equivalent sine and cosine waves differ only by a phase shift of 90°, both are called *sine waves* by convention.)

This is the basic process by which we translate information into radio form and translate it back again. If we want to transmit a 1-kHz audio tone by radio, we can feed it into one of our mixer's inputs and feed an RF signal — say, 5995 kHz — into the mixer's other input. The result is two radio signals: one at 5994 kHz (5995 – 1) and another at 5996 kHz (5995 + 1). We have achieved modulation.

Converting these two radio signals back to audio is just as straightforward. All we do is feed them into one input of another mixer, and feed a 5995-kHz signal into the mixer's other input. Result: a 1-kHz tone. We have achieved demodulation; we have communicated by radio.

The key principle of a radio mixer is that in mixing multiple signal voltages together, it adds and subtracts their frequencies to produce new frequencies. (In the field of signal processing, this process, *multiplication in the time domain,* is recognized as equivalent to the process of *convolution in the frequency domain.* Those interested in this alternative approach to describing the generation of new frequencies through mixing can find more information about it in the many textbooks available on this fascinating subject.) The difference between the mixer we've been describing and any mixer, modulator or demodulator that you'll ever use is that it's ideal. We put in two signals and got just two signals out. *Real* mixers, modulators and demodulators, on the other hand, also produce *distortion* products that make their output spectra "dirtier" or "less clean," as well as putting out some energy at input-signal frequencies and their harmonics. Much of the art and science of making good use of multiplication in mixing, modulation and demodulation goes into minimizing these unwanted multiplication products (or their effects) and making multipliers do their frequency translations as efficiently as possible.

Putting Multiplication to Work

Piecing together a coherent picture of how multiplication works in radio communication isn't made any easier by the fact that traditional terms applied to a given multiplication approach and its products may vary with their application. If, for instance, you're familiar with standard textbook approaches to mixers, modulators and demodulators, you may be wondering why we didn't begin by working out the math involved by examining *amplitude modulation*, also known as *AM*. "Why not tell them about the *carrier* and how to get rid of it in a *balanced modulator*?" A transmitter enthusiast may ask "Why didn't you mention *sidebands* and how we conserve spectrum space and power by getting rid of one and putting all of our power into the other?" A student of radio receivers, on the other hand, expects any discussion of the same underlying multiplication issues to touch on the topics of *LO feedthrough*, *mixer balance* (*single* or *double*?), *image rejection* and so on.

You likely expect this book to spend some time talking to you about these things, so it will. But *this* radio-amateur-oriented discussion of mixers, modulators and demodulators will take a look at their common underlying mechanism *before* turning you loose on practical mixer, modulator and demodulator circuits. Then you'll be able to tell the forest from the trees. **Fig 11.3** shows the block symbol for a traditional mixer along with several IEC symbols for other functions mixers may perform.

It turns out that the mechanism underlying multiplication, mixing, modulation and demodulation is a pretty straightforward thing: Any circuit structure that *nonlinearly distorts* ac waveforms acts as a multiplier to some degree.

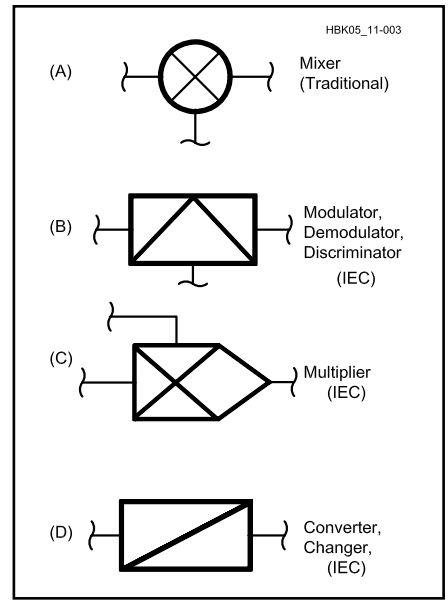


Fig 11.3 — We commonly symbolize mixers with a circled *X* (A) out of tradition, but other standards sometimes prevail (B, C and D). Although the converter/changer symbol (D) can conceivably be used to indicate frequency changing through mixing, the three-terminal symbols are arguably better for this job because they convey the idea of two signal sources resulting in a new frequency. (*IEC* stands for *International Electrotechnical Commission*.)

Nonlinear Distortion?

The phrase *nonlinear distortion* sounds redundant, but isn't. Distortion, an externally imposed change in a waveform, can be linear; that is, it can occur independently of signal amplitude. Consider a radio receiver front-end filter that passes only signals between 6 and 8 MHz. It does this by *linearly distorting* the single complex waveform corresponding to the wide RF spectrum present at the radio's antenna terminals, reducing the amplitudes of frequency components below 6 MHz and above 8 MHz relative to those between 6 and 8 MHz. (Considering multiple signals on a wire as one complex waveform is just as valid, and sometimes handier, than considering them as separate signals. In this case, it's a bit easier to think of distortion as something that happens to a waveform rather than something that happens to separate signals relative to each other. It would be just as valid — and certainly more in keeping with the consensus view — to say merely that the filter attenuates signals at frequencies below 6 MHz and above 8 MHz.) The filter's output waveform certainly differs from its input waveform; the waveform has been distorted.

But because this distortion occurs independently of signal level or polarity, the distortion is linear. No new frequency components are created; only the amplitude relationships among the wave's existing frequency components are altered. This is *amplitude* or *frequency* distortion, and all filters do it or they wouldn't be filters.

Phase or *delay distortion*, also linear, causes a complex signal's various component frequencies to be delayed by different amounts of time, depending on their frequency but independently of their amplitude. No new frequency components occur, and amplitude relationships among existing frequency components are not altered. Phase distortion occurs to some degree in all real filters.

The waveform of a non-sinusoidal signal can be changed by passing it through a circuit that has only linear distortion, but only *nonlinear distortion* can change the waveform of a simple sine wave. It can also produce an output signal whose output waveform changes as a function of the input amplitude, something not possible with linear distortion. Nonlinear circuits often distort excessively with overly strong signals, but the distortion can be a complex function of the input level.

Nonlinear distortion may take the form of *harmonic distortion*, in which integer multiples of input frequencies occur, or *intermodulation distortion (IMD)*, in which different components multiply to make new ones.

Any departure from absolute linearity results in some form of nonlinear distortion, and this distortion can work for us or against us. Any so-called linear amplifier distorts nonlinearly to some degree; any device or circuit that distorts nonlinearly can work as a mixer, modulator, demodulator or frequency multiplier. An amplifier optimized for linear operation will nonetheless mix, but inefficiently; an amplifier biased for nonlinear amplification may be practically linear over a given tiny portion of its input-signal range. The trick is to use careful design and component selection to maximize nonlinear distortion when we want it, and minimize it when we don't. Once we've decided to maximize nonlinear distortion, the trick is to minimize the distortion products we don't want, and maximize the products we desire.

Keeping Unwanted Distortion Products Down

Ideally, a mixer multiplies the signal at one of its inputs by the signal at its other input, but does not multiply a signal at the same input by itself, or multiple signals at the same input by themselves or by each other. (Multiplying a signal by itself — squaring it — generates harmonic distortion [specifically, *second-harmonic* distortion] by adding the signal's frequency to itself per equation 7. Simultaneously squaring two or more signals generates simultaneous harmonic and intermodulation distortion, as we'll see later when we explore how a diode demodulates AM.)

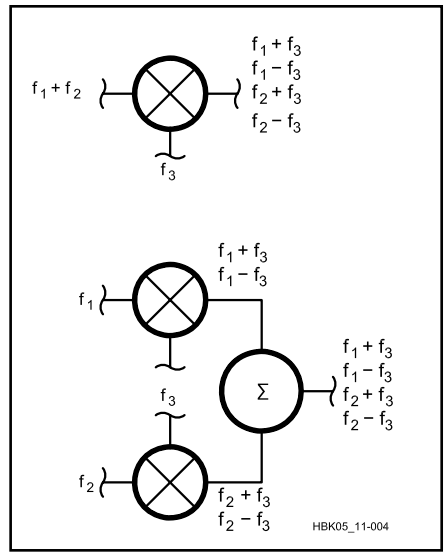


Fig 11.4 — Feeding two signals into one input of a mixer results in the same output as if f_1 and f_2 are each first mixed with f_3 in two separate mixers, and the outputs of these mixers are combined.

Consider what happens when a mixer must handle signals at two different frequencies (we'll call them f_1 and f_2) applied to its first input, and a signal at a third frequency (f_3) applied to its other input. Ideally, a mixer multiplies f_1 by f_3 and f_2 by f_3, but does not multiply f_1 and f_2 by each other. This produces output at the sum and difference of f_1 and f_3, and the sum and difference of f_2 and f_3, but *not* the sum and difference of f_1 and f_2. **Fig 11.4** shows that feeding two signals into one input of a mixer results in the same output as if f_1 and f_2 are each first mixed with f_3 in two separate mixers, and the outputs of these mixers are combined. This shows that a mixer, even though constructed with nonlinearly distorting components, actually behaves as a *linear frequency shifter.* Traditionally, we refer to this process as mixing and to its outputs as *mixing products*, but we may also call it frequency *conversion*, referring to a device or circuit that does it as a *converter*, and to its outputs as *conversion products.*

Real mixers, however, at best act only as *reasonably* linear frequency shifters, generating some unwanted IMD products — spurious signals, or *spurs* — as they go. Receivers are especially sensitive to unwanted mixer IMD because the signal-level spread over which they must operate without generating unwanted IMD is often 90 dB or more, and includes infinitesimally weak signals in its span. In a receiver, IMD products so tiny that you'd never notice them in a transmitted signal can easily obliterate weak signals. This is why receiver designers apply so much effort to achieving "high dynamic range."

The degree to which a given mixer, modulator or demodulator circuit produces unwanted IMD is often *the* reason why we use it, or don't use it, instead of another circuit that does its wanted-IMD job as well or even better.

Other Mixer Outputs

In addition to desired sum-and-difference products and unwanted IMD products, real mixers also put out some energy at their input frequencies. Some mixer implementations may *suppress* these outputs — that is, reduce one or both of their input signals by a factor of 100 to 1,000,000, or 20 to 60 dB. This is good because it helps keep input signals at the desired mixer-output sum or difference frequency from showing up at the IF terminal — an effect reflected in a receiver's *IF rejection* specification. Some mixer types, especially those used in the vacuum-tube era, suppress their input-signal outputs very little or not at all.

Input-signal suppression is part of an overall picture called *port-to-port isolation*. Mixer input and output connections are traditionally called *ports*. By tradition, the port to which we apply the shifting signal is the *local-oscillator (LO)* port. The convention for naming the other two ports (one of which must be an output, and the other of which must be an input) is usually that the higher-frequency port is called the *RF (radio frequency)* port and the lower-frequency port is called the *IF (intermediate frequency)* port. If a mixer's output frequency is lower than its input frequency, then the RF port is an input and the IF port is an output. If the output frequency is higher than the input frequency, the IF port may be the input and the RF port may be the output. (We hedge with *may be* because usage varies. When in doubt, check a diagram carefully to determine which port is the "gozinta" and which port is the "gozouta.")

It's generally a good idea to keep a mixer's input signals from appearing at its output port because they represent energy that we'd rather not pass on to subsequent circuitry. It therefore follows that it's usu-

ally a good idea to keep a mixer's LO-port energy from appearing at its RF port, or its RF-port energy from making it through to the IF port. But there are some notable exceptions.

Mixers and Amplitude Modulation

Now that we've just discussed what a fine thing it is to have a mixer that doesn't let its input signals through to its output port, we can explore a mixing approach that outputs one of its input signals so strongly that the fed-through signal's amplitude at least equals the combined amplitudes of the system's sum and difference products! This system, *amplitude modulation*, is the oldest means of translating information into radio form and back again. It's a frequency-shifting system in which the original unmodulated signal, traditionally called the *carrier*, emerges from the mixer along with the sum and difference products, traditionally called *sidebands*.

We can easily make the carrier pop out of our mixer along with the sidebands merely by building enough *dc level shift* into the information we want to mix so that its waveform never goes negative. Back at equations 1 and 2, we decided to keep our mixer math relatively simple by setting the peak voltage of our mixer's input signals directly equal to their sine values. Each input signal's peak voltage therefore varies between +1 and –1, so all we need to do to keep our modulating-signal term (provided with a subscript m to reflect its role as the modulating or information waveform) from going negative is add 1 to it. Identifying the carrier term with a subscript c, we can write

$$\text{AM signal} = \left(1 + \text{m} \sin 2\pi f_m t\right) \sin 2\pi f_c t \quad (8)$$

Notice that the modulation ($2\pi f_m t$) term has company in the form of a coefficient, m. This variable expresses the modulating signal's varying amplitude — variations that ultimately result in amplitude modulation. Expanding equation 8 according to equation 6 gives us

$$\begin{aligned}\text{AM signal} = {} & \sin 2\pi f_c t \\ & + \frac{1}{2}\text{m} \cos\left(2\pi f_c - 2\pi f_m\right)t \\ & - \frac{1}{2}\text{m} \cos\left(2\pi f_c + 2\pi f_m\right)t\end{aligned} \quad (9)$$

The modulator's output now includes the carrier ($\sin 2\pi f_c t$) in addition to sum and difference products that vary in strength according to m. According to the conventions of talking about modulation, we call the sum product, which comes out at a frequency higher than that of the carrier, the *upper sideband (USB)*, and the difference product, which comes out a frequency lower than that of the carrier, the *lower sideband (LSB)*. We have achieved amplitude modulation.

Why We Call It Amplitude Modulation

We call the modulation process described in equation 8 *amplitude modulation* because the complex waveform consisting of the sum of the sidebands and carrier varies with the information signal's magnitude (m). Concepts long used to illustrate AM's mechanism may mislead us into thinking that the *carrier* varies in strength with modulation, but careful study of equation 9 shows that this doesn't happen. The carrier, $\sin 2\pi f_c t$, goes into the modulator — we're in the modulation business now, so it's fitting to use the term *modulator* instead of *mixer* — as a sinusoid with an unvarying maximum value of |1|. The modulator multiplies the carrier by the dc level (+1) that we added to the information signal ($\text{m} \sin 2\pi f_m t$). Multiplying $\sin 2\pi f_c t$ by 1 merely returns $\sin 2\pi f_c t$. We have proven that the carrier's amplitude does not vary as a result of amplitude modulation. The carrier is, however, used by many circuits as a reference signal.

Overmodulation

Since the audio we are transmitting in AM shows up entirely as energy in its sidebands, it follows that the more energetic we make the sidebands, the more information energy will be available for an AM receiver to "recover" when it demodulates the signal. Even in an ideal modulator, there's a practical limit to how strong we can make an AM signal's sidebands relative to its carrier, however. Beyond that limit, we severely distort the waveform we want to translate into radio form.

We reach AM's distortion-free modulation limit when the sum of the sidebands and carrier at the modulator output *just reaches zero* at the modulating waveform's most negative peak (**Fig 11.5**). We call this condition *100% modulation*, and it occurs when m equals 1. (We enumerate *modulation percentage* in values from 0 to 100%. The lower the number, the less information energy in the sidebands. You may also see modulation enumerated in terms of a *modulation factor* from 0 to 1, which directly equals m; a modulation factor of 1 is the same as 100% modulation.) Equation 9 shows that each sideband's voltage is half that of the carrier. Power varies as the square of voltage, so the power in each sideband of a 100%-modulated signal is therefore $(1/2)^2$ times, or 1/4, that of the carrier. A transmitter

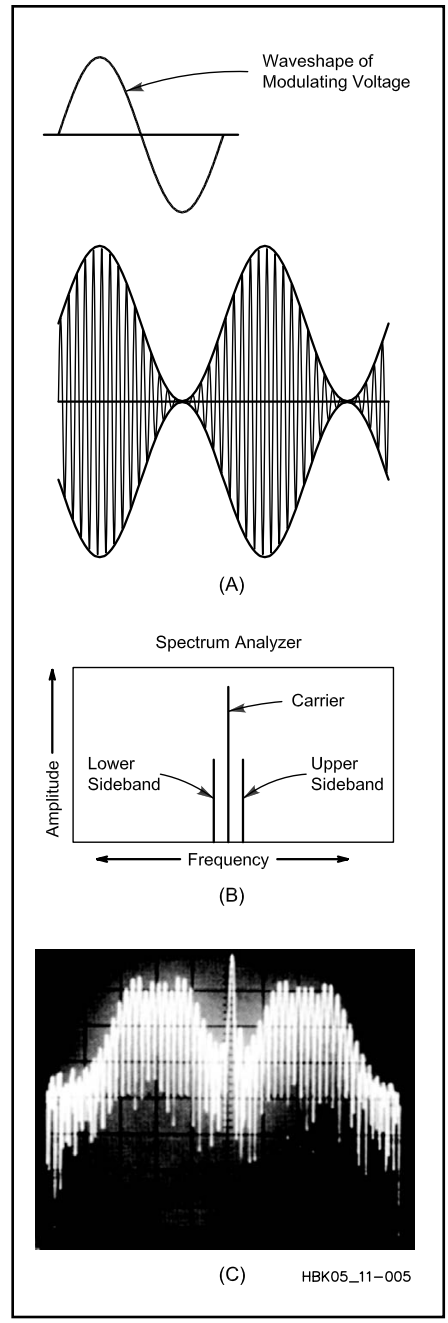


Fig 11.5 — Graphed in terms of amplitude versus time (A), the *envelope* of a properly modulated AM signal exactly mirrors the shape of its modulating waveform, which is a sine wave in this example. This AM signal is modulated as fully as it can be — 100% — because its envelope *just* hits zero on the modulating wave's negative peaks. Graphing the same AM signal in terms of amplitude versus frequency (B) reveals its three spectral components: Carrier, upper sideband and lower sideband. B shows sidebands as single-frequency components because the modulating waveform is a sine wave. With a complex modulating waveform, the modulator's sum and difference products really do show up as *bands* on either side of the carrier (C).

capable of 100% modulation when operating at a carrier power of 100 W therefore puts out a 150-W signal at 100% modulation, 50 W of which is attributable to the sidebands. (The *peak* envelope power [PEP] output of a double-sideband, full-carrier AM transmitter at 100% modulation is four times its carrier PEP. This is why our solid-state, "100-W" MF/HF transceivers are usually rated for no more than about 25 W carrier output at 100% amplitude modulation.)

One-hundred-percent modulation is a brick-wall limit because an amplitude modulator can't reduce its output to less than zero. Trying to increase modulation beyond the 100% point results in *overmodulation* (**Fig 11.6**), in which the modulation envelope no longer mirrors the shape of the modulating wave (Fig 11.6A). An overmodulated wave contains more energy than it did at 100% modulation, but some of the added energy now exists as *harmonics of the modulating waveform* (Fig 11.6B). This distortion makes the modulated signal take up more spectrum space than it needs. In voice operation, overmodulation commonly happens only on syllabic peaks, making the distortion products sound like crashy, transient noise we refer to as *splatter*.

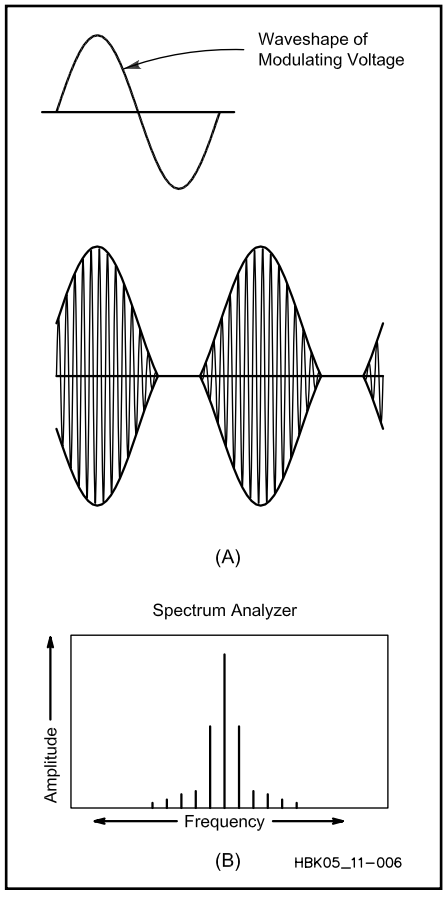


Fig 11.6 — Overmodulating an AM transmitter results in a modulation envelope (A) that doesn't faithfully mirror the modulating waveform. This distortion creates additional sideband components that broaden the transmitted signal (B).

Modulation Linearity

If we increase an amplitude modulator's modulating-signal input by a given percentage, we expect a proportional modulation increase in the modulated signal. We expect good *modulation linearity*. Suboptimal amplitude modulator design may not allow this, however. Above some modulation percentage, a modulator may fail to increase modulation in proportion to an increase its input signal (**Fig 11.7**). Distortion, and thus an unnecessarily wide signal, results.

Using AM to Send Morse Code

Fig 11.8A closely resembles what we see when a properly adjusted CW transmitter sends a string of dots. Keying a carrier on and off produces a wave that varies in amplitude and has double (upper and lower) sidebands that vary in spectral composition according to the duration and envelope shape of the on-off transitions. The emission mode we call *CW* is therefore a form of AM. The concepts of modulation percentage and overmodulation are usually not applied to generating an on-off-keyed Morse signal, however. This is related to how we copy CW by ear, and the fact that, in CW radio communication, we usually don't translate the received signal all the way back into its original pre-modulator (*baseband*) form, as a closer look at the process reveals.

In CW transmission, we usually open and close a keying line to make dc transitions that turn the transmitted carrier on and off. See Fig 11.8B. CW reception usually does not entirely reverse this process, however. Instead of demodulating a CW signal all the way back to its baseband self — a shifting dc level — we want the presences and absences of its carrier to create long and short audio tones. Because the carrier is RF and not AF, we must mix it

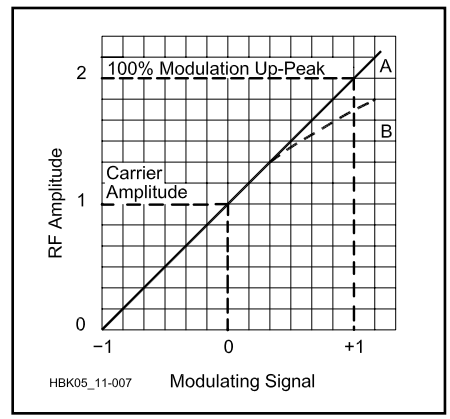


Fig 11.7 — An ideal AM transmitter exhibits a straight-line relationship (A) between its instantaneous envelope amplitude and the instantaneous amplitude of its modulating signal. Distortion, and thus an unnecessarily wide signal, results if the transmitter cannot respond linearly across the modulating signal's full amplitude range.

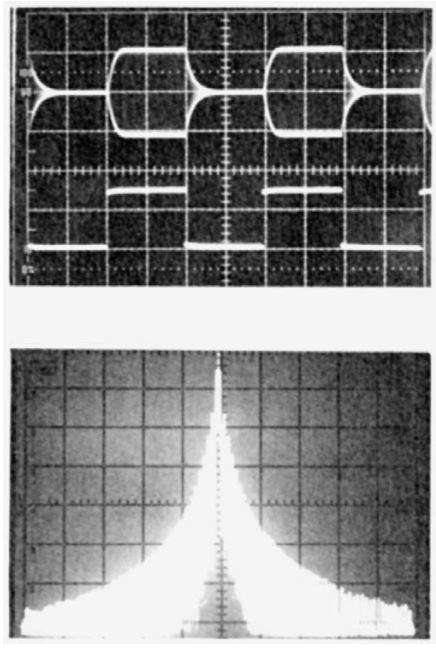


Fig 11.8 — Telegraphy by on-off-keying a carrier is also a form of AM, called *CW* (short for *continuous wave*) for reasons of tradition. *Waveshaping* in a CW transmitter often causes a CW signal's RF envelope (the amplitude-versus-time display at A) to contain less harmonic energy than the faster transitions of its keying waveform (B) suggest should be the case. C, an amplitude-versus-frequency display, shows that even a properly shaped CW signal has many sideband components.

with a locally generated RF signal — from a *beat-frequency oscillator (BFO)* — that's close enough in frequency to produce a difference signal at AF. What goes into our transmitter as shifting dc comes out of our receiver as thump-delimited tone bursts of dot and dash duration. We have achieved CW communication.

The dots and dashes of a CW signal must start and stop abruptly enough so we can clearly distinguish the carrier's presences and absences from noise, especially when fading prevails. The keying sidebands, which sound like little more than thumps when listened to on their own, help our brains be sure when the carrier tone starts and stops.

It so happens that we always need to hear one or more harmonics of the fundamental keying waveform for the code to sound sufficiently crisp. If the transmitted signal will be subject to propagational fading — a safe assumption for any long-distance radio communication — we *harden* our keying by making the transmitter's output rise and fall more quickly. This puts more energy into more keying sidebands and makes the signal more copiable in the presence of fading — in particular, *selective fading*, which linearly distorts a modulated signal's complex waveform and randomly changes the sidebands' strength and phase relative to the carrier and each other. The appropriate keying hardness also depends on the keying speed. The faster the keying in WPM, the faster the on-off times — the harder the keying — must be for the signal to remain ear-readable through noise and fading.

Instead of thinking of this process in terms of modulation percentage, we just ensure that a CW transmitter produces sufficient keying-sideband energy for solid reception. Practical CW transmitters rarely do their keying with a modulator stage as such. Instead, one or more stages are turned on and off to modulate the carrier with Morse, with rise and fall times set by R and C values associated with the stages' keying and/or power supply lines. A transmitter's CW *waveshaping* is therefore usually hardwired to values appropriate for reasonably high-speed sending (35 to 55 WPM or so) in the presence of fading. As a result, we generally cannot vary keying hardness at will as we might vary a voice transmitter's modulation with a front-panel control. Rise and fall times of 1 to 5 ms (5 ms rise and fall times equate to a keying speed of 36 WPM in the presence of fading and 60 WPM if fading is absent) are common.

The faster a CW transmitter's output changes between zero and maximum, the more bandwidth its carrier and sidebands occupy. See Fig 11.8C. Making a CW signal's keying too hard is therefore spectrum-wasteful and unneighborly because it makes the signal wider than it needs to be. Keying sidebands that are stronger and wider than necessary are traditionally called *clicks* because of what they sound like on the air. There is a more detailed discussion of keying waveforms in the **Receivers and Transmitters** chapter of this *Handbook*.

The Many Faces of Amplitude Modulation

We've so far examined mixers, multipliers and modulators that produce complex output signals of two types. One, the action of which equation 7 expresses, produces only the frequency sum of and frequency difference between its input signals. The other, the amplitude modulator characterized by equations 8 and 9, produces carrier output in addition to the frequency sum of and frequency difference between its input signals. Exploring the AM process led us to a discussion on-off-keyed CW, which is also a form of AM.

Amplitude modulation is nothing more and nothing less than varying an output signal's amplitude according to a varying voltage or current. All of the output signal types mentioned above are forms of amplitude modulation, and there are others. Their names and applications depend on whether the resulting signal contains a carrier or not, and both sidebands or not. Here's a brief overview of AM-signal types, what they're called, and some of the jobs you may find them doing:

- *Double-sideband (DSB), full-carrier AM* is often called just *AM*, and often what's meant when radio folk talk about just *AM*. (When the subject is broadcasting, *AM* can also refer to broadcasters operating in the 525- to 1705-kHz region, generically called *the AM band* or *the broadcast band* or *medium wave*. These broadcasters used only double-sideband, full-carrier AM for many years, but many now use combinations of amplitude modulation and *angle modulation*, which we'll explore shortly, to transmit stereophonic sound.) Equations 8 and 9 express what goes on in generating this signal type. What we call *CW* — Morse code done by turning a carrier on and off — is a form of DSB, full-carrier AM.
- *Double-sideband, suppressed-carrier AM* is what comes out of a circuit that does what equation 7 expresses — a sum (upper sideband), a difference (lower sideband) and no carrier. We didn't call its sum and difference outputs upper and lower sidebands earlier in equation 7's neighborhood, but we'd do so in a transmitting application. In a transmitter, we call a circuit that suppresses the carrier while generating upper and lower sidebands a *balanced modulator*, and we quantify its *carrier suppression*, which is always less than infinite. In a receiver, we call such a circuit a *balanced mixer*, which may be *single-balanced* (if it lets either its RF signal or its LO [carrier] signal through to its output) or *double-balanced* (if it suppresses both its input signal and LO/carrier in its output), and we quantify its *LO suppression* and *port-to-port isolation*, which are always less than infinite. (Mixers [and amplifiers] that afford no balance whatsoever are sometimes said to be *single-ended*.) Sometimes, DSB suppressed-carrier AM is called just *DSB*.
- *Vestigial sideband (VSB), full-carrier AM* is like the DSB variety with one sideband partially filtered away for bandwidth reduction. Commercial television systems that transmit AM video use VSB AM.
- *Single-sideband, suppressed-carrier AM* is what you get when you generate a DSB, suppressed carrier AM signal and throw away one sideband with filtering or phasing. We usually call this signal type just *single sideband (SSB)* or, as appropriate, *upper sideband (USB)* or *lower sideband (LSB)*. In a modulator or demodulator system, the *unwanted sideband* — that is, the sum or difference signal we don't want — may be called just that, or it may be called the *opposite sideband*, and we refer to a system's *sideband rejection* as a measure of how well the opposite sideband is suppressed. In receiver mixers not used for demodulation and transmitter mixers not used for modulation, the unwanted sum or difference signal, or the input signal that produces the unwanted sum or difference, is the *image*, and we refer to a system's *image rejection*. A pair of mixers specially configured to suppress either the sum or the difference output is an *image-reject mixer (IRM)*. In receiver demodulators, the unwanted sum or difference signal may just be called the opposite sideband, or it may be called the *audio image*. A receiver capable of rejecting the opposite sideband or audio image is said to be capable of *single-signal* reception.
- *Single-sideband, full-carrier AM* is akin to full-carrier DSB with one sideband missing. Commercial and military communicators may call it *AM equivalent*

(AME) or *compatible AM (CAM) — compatible* because it can be usefully demodulated in AM and SSB receivers and because it occupies about the same amount of spectrum space as SSB.)

- *Independent sideband (ISB) AM* consists an upper sideband and a lower sideband containing different information (a carrier of some level may also be present). Radio amateurs sometimes use ISB to transmit simultaneous slow-scan-television and voice information; international broadcasters sometimes use it for point-to-point audio feeds as a backup to satellite links.

Mixers and AM Demodulation

Translating information from radio form back into its original form — demodulation — is also traditionally called *detection*. If the information signal we want to detect consists merely of a baseband signal frequency-shifted into the radio realm, almost any low-distortion frequency-shifter that works according to equation 7 can do the job acceptably well.

Sometimes we recover a radio signal's information by shifting the signal right back to its original form with no intermediate frequency shifts. This process is called *direct conversion*. More commonly, we first convert a received signal to an *intermediate frequency* so we can amplify, filter and level-control it prior to detection. This is *superheterodyne* reception, and most modern radio receivers work in this way. Whatever the receiver type, however, the received signal ultimately makes its way to one last mixer or demodulator that completes the final translation of information back into audio, or into a signal form suitable for device control or computer processing. In this last translation, the incoming signal is converted back to recovered-information form by mixing it with one last RF signal. In heterodyne or *product* detection, that final frequency-shifting signal comes from a BFO. The incoming-signal energy goes into one mixer input port, BFO energy goes into the other, and audio (or whatever form the desired information takes) results.

If the incoming signal is full-carrier AM and we don't need to hear the carrier as a tone, we can modify this process somewhat, if we want. We can use the carrier itself to provide the heterodyning energy in a process called *envelope detection*.

Envelope Detection and Full-Carrier AM

Fig 11.5 graphically represents how a full-carrier AM signal's *modulation envelope* corresponds to the shape of the modulating wave. If we can derive from the

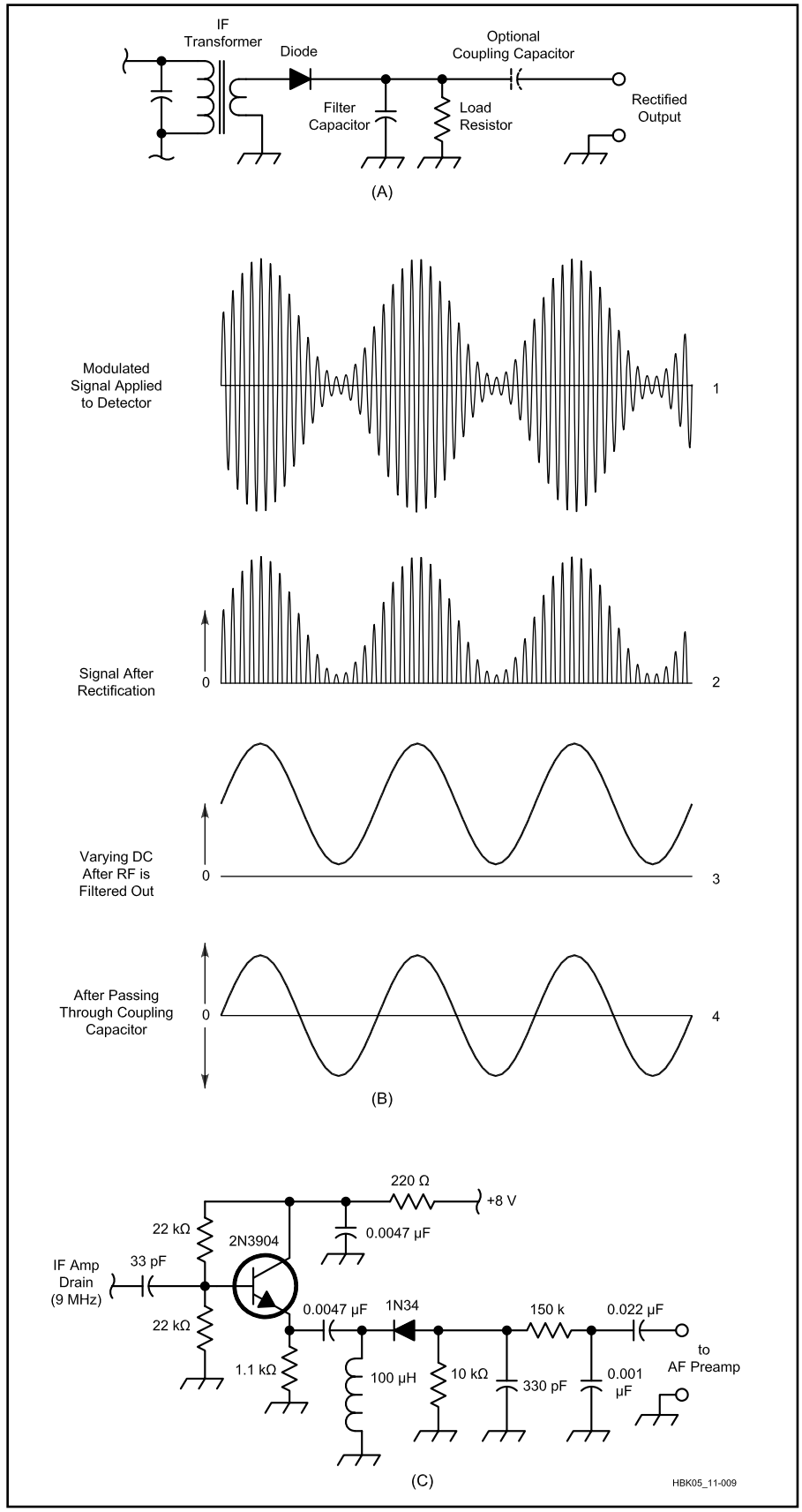


Fig 11.9 — Radio's simplest demodulator, the diode rectifier (A), demodulates an AM signal by multiplying its carrier and sidebands to produce frequency sums and differences, two of which sum into a replica of the original modulation (B). Modern receivers often use an emitter follower to provide low-impedance drive for their diode detectors (C).

modulated signal a voltage that varies according to the modulation envelope, we will have successfully recovered the information present in the sidebands. This process is called envelope detection, and we can achieve it by doing nothing more complicated than half-wave-rectifying the modulated signal with a diode (**Fig 11.9**).

That a diode demodulates an AM signal by allowing its carrier to multiply with its sidebands may jar those long accustomed to seeing diode detection ascribed merely to "rectification." But a diode is certainly nonlinear. It passes current in only one direction, and its output voltage is (within limits) proportional to the square of its input voltage. These nonlinearities allow it to multiply.

Exploring this mathematically is tedious with full-carrier AM because the process squares three summed components (carrier, lower sideband and upper sideband). Rather than fill the better part of a page with algebra, we'll instead characterize the outcome verbally: In "just rectifying" a DSB, full-carrier AM signal, a diode detector produces

- Direct current (the result of rectifying the carrier);
- A second harmonic of the carrier;
- A second harmonic of the lower sideband;
- A second harmonic of the upper sideband;
- Two difference-frequency outputs (upper sideband minus carrier, carrier minus lower sideband), each of which is equivalent to the modulating waveform's frequency, and both of which sum to produce the recovered information signal; and
- A second harmonic of the modulating waveform (the frequency difference between the two sidebands).

Three of these products are RF. Low-pass filtering, sometimes little more than a simple RC network, can remove the RF products from the detector output. A capacitor in series with the detector output line can block the carrier-derived dc component. That done, only two signals remain: the recovered modulation and, at a lower level, its second harmonic — in other words, second-harmonic distortion of the desired information signal.

Mixers and Angle Modulation

Amplitude modulation served as our first means of translating information into radio form because it could be implemented as simply as turning an electric noise generator on and off. (A spark transmitter consisted of little more than this.) By the 1930s, we had begun experimenting with translating information into radio form and back again by modulating a radio wave's angular velocity (frequency or phase) instead of its overall amplitude. The result of this process is *frequency modulation (FM)* or *phase modulation (PM)*, both of which are often grouped under the name *angle modulation* because of their underlying principle.

A change in a carrier's frequency or phase for the purpose of modulation is called *deviation*. An FM signal deviates according to the amplitude of its modulating waveform, independently of the modulating waveform's frequency; the higher the modulating wave's amplitude, the greater the deviation. A PM signal deviates according to the amplitude *and frequency* of its modulating waveform; the higher the modulating wave's amplitude *and/or frequency*, the greater the deviation.

An angle-modulated signal can be mathematically represented as

$$\begin{aligned} f_c(t) &= \cos\left(2\pi f_c t + m\sin\left(2\pi f_m t\right)\right) \\ &= \cos\left(2\pi f_c t\right)\cos\left(m\sin\left(2\pi f_m t\right)\right) \\ &\quad - \sin\left(2\pi f_c t\right)\sin\left(m\sin\left(2\pi f_m t\right)\right) \end{aligned} \qquad (10)$$

In it, we see the carrier frequency ($2\pi f_c t$) and modulating signal ($\sin 2\pi f_m t$) as in the equation for AM (equation 8). We again see the modulating signal associated with a coefficient, m, which relates to degree of modulation. (In the AM equation, *m* is the modulation factor; in the angle-modulation equation, m is the *modulation index* and, for FM, equals the deviation divided by the modulating frequency.) We see that angle-modulation occurs as the cosine of the sum of the carrier frequency ($2\pi f_c t$) and the modulating signal ($\sin 2\pi f_m t$) times the modulation index (m). In its expanded form, we see the appearance of sidebands above and below the carrier frequency.

Angle modulation is a multiplicative process, so, like AM, it creates sidebands on both sides of the carrier. Unlike AM, however, angle modulation creates an *infinite* number of sidebands on either side of the carrier! This occurs as a direct result of modulating the carrier's angular velocity, to which its frequency and phase directly relate. If we continuously vary a wave's angular velocity according to another periodic wave's cyclical amplitude variations, the rate at which the modulated wave repeats *its* cycle — its frequency — passes through an infinite number of values. (How many individual amplitude points are there in one cycle of the modulating wave? An infinite number. How many corresponding discrete frequency or phase values does the corresponding angle-modulated wave pass through as the modulating signal completes a cycle? An infinite number!) In AM, the carrier frequency stays at one value, so AM produces two sidebands — the sum of its carrier's unchanging frequency value and the modulating frequency, and the difference between the carrier's unchanging frequency value and the modulating frequency. In angle modulation, the modulating wave shifts the frequency or phase of the carrier through an infinite number of different frequency or phase values, resulting in an infinite number of sum and difference products.

Wouldn't the appearance on the air of just a few such signals result in a bedlam of mutual interference? No, because most of angle modulation's uncountable sum and difference products are vanishingly weak in practical systems, and because they don't show up just anywhere in the spectrum. Rather, they emerge from the modulator spaced from the average ("resting," unmodulated) carrier frequency by integer multiples of the modulating frequency (**Fig 11.10**). The strength of the sidebands relative to the carrier, and the strength and

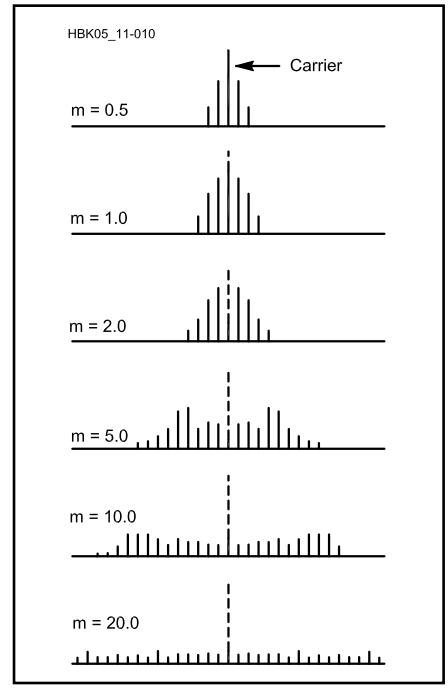


Fig 11.10 — Angle-modulation produces a carrier and an infinite number of upper and lower sidebands spaced from the average ("resting," unmodulated) carrier frequency by integer multiples of the modulating frequency. (This drawing is a simplification because it only shows relatively strong, close-in sideband pairs; space constraints prevent us from extending it to infinity.) The relative amplitudes of the sideband pairs and carrier vary with modulation index, *m*.

phase of the carrier itself, vary with the degree of modulation — the modulation index. (The *overall* amplitude of an angle-modulated signal does not change with modulation, however; when energy goes out of the carrier, it shows up in the sidebands, and vice versa.) In practice, we operate angle-modulated transmitters at modulation indexes that make all but a few of their infinite sidebands small in amplitude. (A mathematical tool called *Bessel functions* help determine the relative strength of the carrier and sidebands according to modulation index. The **Modes and Modulation Sources** chapter includes a graph to illustrate this relationship.) Selectivity in transmitter and receiver circuitry further modify this relationship, especially for sidebands far away from the carrier.

Angle Modulators

Vary a reactance in or associated with an oscillator's frequency-determining element(s), and you vary the oscillator's frequency. Vary the tuning of a tuned circuit through which a signal passes, and you vary the signal's phase. A circuit that does this is called a *reactance modulator*, and can be little more than a tuning diode or two connected to a tuned circuit in an oscillator or amplifier (**Fig 11.11**). Varying a reactance through which the signal passes (**Fig 11.12**) is another way of doing the same thing.

The difference between FM and PM depends solely on how, and not how much, deviation occurs. A modulator that causes deviation in proportion to the modulating wave's amplitude and frequency is a phase modulator. A modulator that causes deviation only in proportion to the modulating signal's amplitude is a frequency modulator.

Increasing Deviation by Frequency Multiplication

Maintaining modulation linearity is just as important in angle modulation as it is in AM, because unwanted distortion is always our enemy. A given angle-modulator circuit can frequency- or phase-shift a carrier only so much before the shift stops occurring in strict proportion to the amplitude (or, in PM, the amplitude and frequency) of the modulating signal.

If we want more deviation than an angle modulator can linearly achieve, we can operate the modulator at a suitable subharmonic — submultiple — of the desired frequency, and process the modulated signal through a series of *frequency multipliers* to bring it up to the desired frequency. The deviation also increases by the overall multiplication factor, relieving the modulator of having to do it all directly. A given FM or PM radio design may achieve its final output frequency through a combination of mixing (frequency shift, no deviation change) and frequency multiplication (frequency shift *and* deviation change).

The Truth About "True FM"

Something we covered a bit earlier bears closer study:

[An FM signal deviates according to the amplitude of its modulating waveform, independently of the modulating waveform's frequency; the higher the modulating wave's amplitude, the greater the deviation. A PM signal deviates according to the amplitude *and frequency* of its modulating waveform; the higher the modulating wave's amplitude *and/or frequency*, the greater the deviation.]

The practical upshot of this excerpt is that we can use a phase modulator to generate FM. All we need to do is run a PM transmitter's modulating signal through a low-pass filter that (ideally) halves the signal's amplitude for each doubling of frequency (a reduction of "6 dB per octave," as we sometimes see such responses characterized) to compensate for its phase modulator's "more deviation with higher frequencies" characteristic. The result is an FM, not PM, signal. FM achieved with a phase modulator is sometimes called *indirect FM* as opposed to the *direct FM* we get from a frequency modulator.

We sometimes see radio gear manufacturers claim that one piece of gear is better than another solely because it generates "true FM" as opposed to indirect FM. We can immunize ourselves against such claims by keeping in mind that direct and indirect FM *sound exactly alike in a receiver* when done correctly.

Depending on the nature of the modulation source, there *is* a practical difference between a frequency modulator and a phase modulator. Answering two questions can tell us whether this difference matters: Does our modulating signal contain a dc level or not? If so, do we need to accurately preserve that dc level through our radio communication link for successful communication? If both answers are *yes*, we must choose our hardware and/or information-encoding approach carefully,

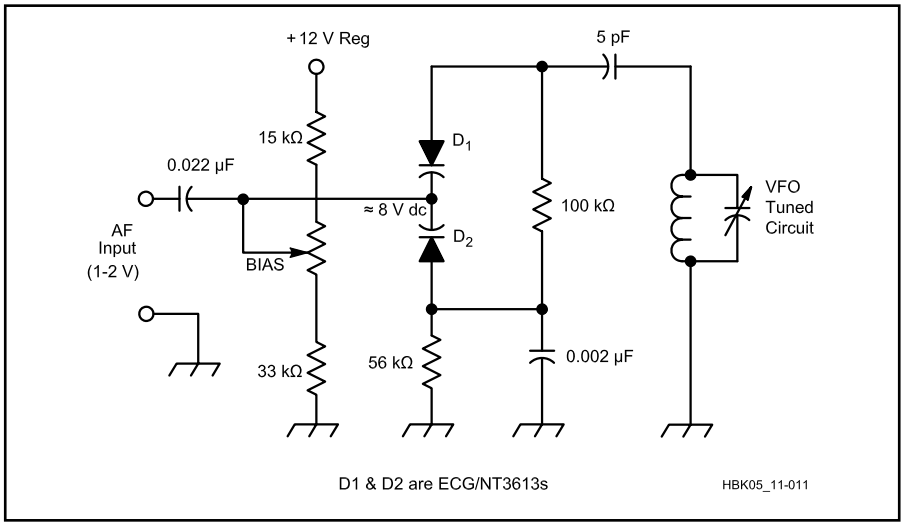


Fig 11.11 — One or more tuning diodes can serve as the variable reactance in a reactance modulator. This HF reactance modulator circuit uses two diodes in series to ensure that the tuned circuit's RF-voltage swing cannot bias the diodes into conduction. D1 and D2 are "30-volt" tuning diodes that exhibit a capacitance of 22 pF at a bias voltage of 4. The BIAS control sets the point on the diode's voltage-versus-capacitance characteristic around which the modulating waveform swings.

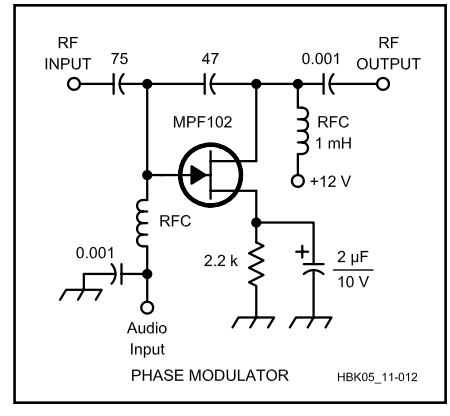


Fig 11.12 — A series reactance modulator acts as a variable shunt around a reactance — in this case, a 47-pF capacitor — through which the carrier passes.

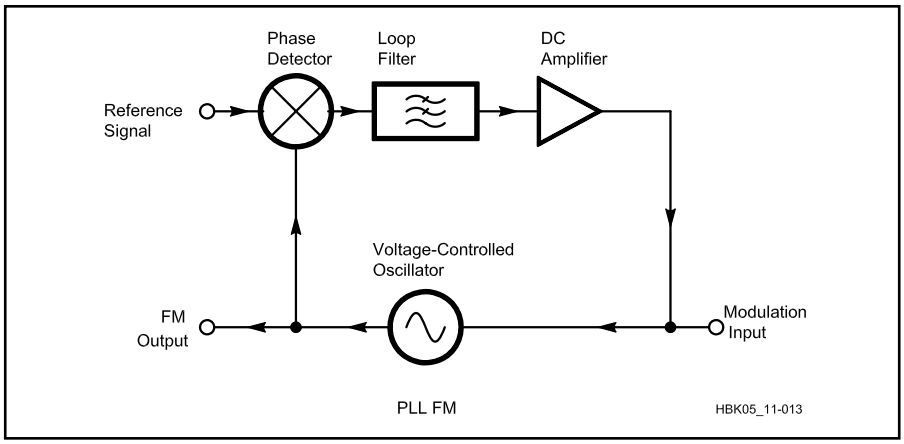


Fig 11.13 — Frequency modulation, PLL-style.

because a frequency modulator can convey shifts in its modulating wave's dc level, while a phase modulator, which responds only to instantaneous changes in frequency and phase, cannot.

Consider what happens when we want to frequency-modulate a phase-locked-loop-synthesized transmitted signal. **Fig 11.13** block-diagrams a PLL frequency modulator. Normally, we modulate a PLL's VCO because it's the easy thing to do. As long as our modulating frequency results infrequency excursions too fast for the PLL to follow and correct — that is, as long as our modulating frequency is outside the PLL's *loop bandwidth* — we achieve the FM we seek. Trying to modulate a dc level by pushing the VCO to a particular frequency and holding it there fails, however, because a PLL's loop response includes dc. The loop sees the modulation's dc component as a correctable error and dutifully "fixes" it. FMing a PLL's VCO therefore can't buy us the dc response "true FM" is supposed to allow.

We *can* dc-modulate a PLL modulator, but we must do so by modulating the loop's *reference*. The PLL then adjusts the VCO to adapt to the changed reference, and our dc level gets through. In this case, the modulating frequency must be *within* the loop bandwidth — which dc certainly is — or the VCO won't be corrected to track the shift.

Mixers and Angle Demodulation

With the awesome prospect of generating an infinite number of sidebands still fresh in our minds, we may be a bit disappointed to learn that we commonly demodulate angle modulation by doing little more than turning it into AM and then envelope- or product-detecting it! But this is what happens in many of our FM receivers and transceivers, and we can get a handle on this process by realizing that a form of angle-modulation-to-AM conversion begins quite early in an angle-modulated signal's life because of distortion of the modulation by amplitude-linear circuitry — something that happens to angle-modulated signals, it turns out, in any linear circuit that doesn't have an amplitude-versus-frequency response that's utterly flat out to infinity.

Think of what happens, for example, when we sweep a constant-amplitude signal up in frequency — say, from 1 kHz to 8 kHz — and pass it through a 6-dB-per-octave filter (**Fig 11.14A**). The filter's rolloff causes the output signal's amplitude to decrease as frequency increases. Now imagine that we linearly sweep our constant-amplitude signal *back and forth* between 1 kHz and 8 kHz at a constant rate of 3 kHz per second (Fig 11.14B). The filter's output *amplitude* now varies cyclically over time as the input signal's *frequency* varies cyclically over time. Right before our eyes, a frequency change turns into an amplitude change. The process of converting angle modulation to amplitude modulation has begun.

This is what happens whenever an angle-modulated signal passes through circuitry with an amplitude-versus-frequency response that isn't flat out to infinity. As the signal deviates across the frequency-response curves of whatever circuitry passes it, its angle modulation is, to some degree, converted to AM — a form of crosstalk between the two modulation types, if we wish to look at it that way. (Variations in system phase linearity also cause distortion and FM-to-AM conversion, because the sidebands do not have the proper phase relationship with respect to

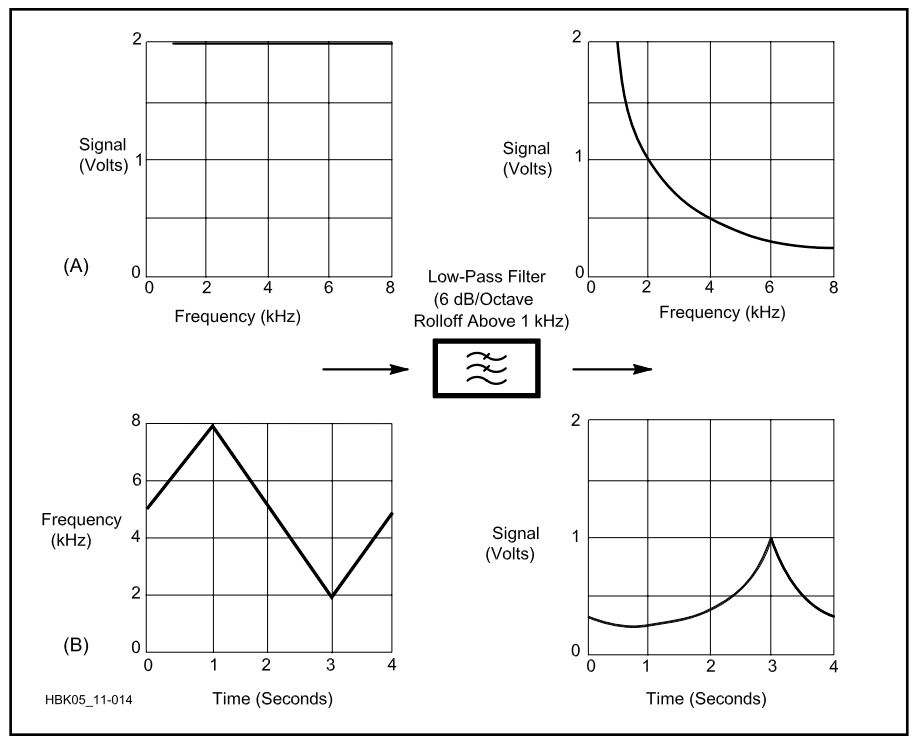


Fig 11.14 — Frequency-sweeping a constant-amplitude signal and passing it through a low-pass filter results in an output signal that varies in amplitude with frequency (A). Sweeping the input signal back and forth between two frequency limits causes the output signal's amplitude to vary between two limits (B). This is the principle behind the angle-demodulation process called *frequency discrimination*.

each other and with respect to the carrier.)

All we need to do to put this effect to practical use is develop a circuit that does this frequency-to-amplitude conversion linearly (and, since more output is better, steeply) across the frequency span of the modulated signal's deviation. Then we envelope-demodulate the resulting AM, and we're in.

Fig 11.15 shows such a circuit — a *discriminator* — and the sort of amplitude-versus-frequency response we expect from it. (It's possible to use an AM receiver to recover understandable audio from a narrow angle-modulated signal by "off-tuning" the signal so its deviation rides up and down on one side of the receiver's IF selectivity curve. This *slope detection* process served as an early, suboptimal form of frequency discrimination.)

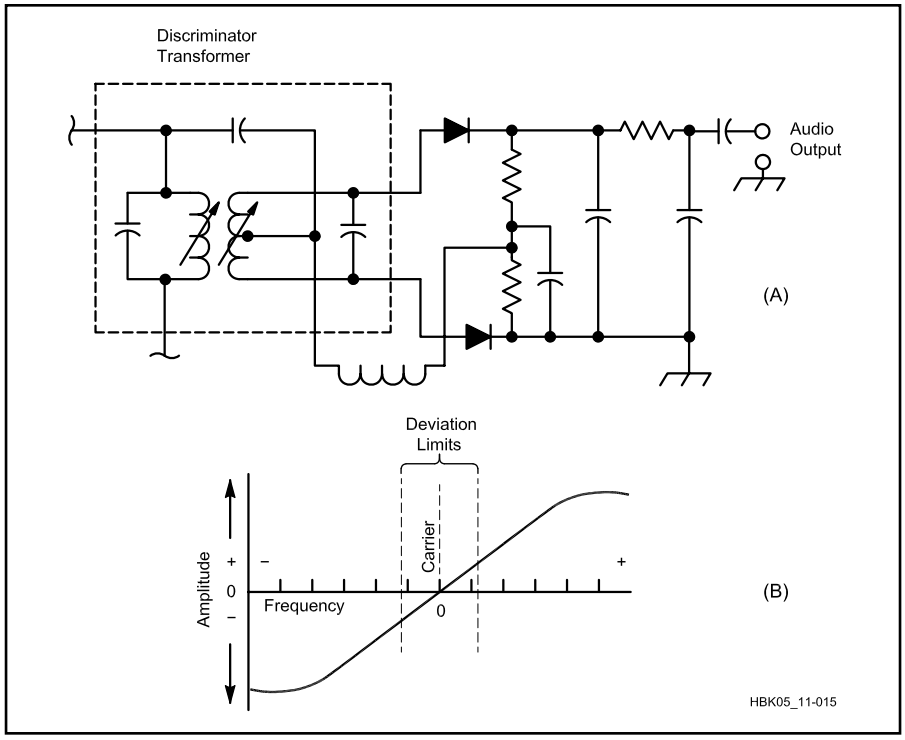


Fig 11.15 — A *discriminator* (A) converts an angle-modulated signal's deviation into an amplitude variation (B) and envelope-detects the resulting AM signal. For undistorted demodulation, the discriminator's amplitude-versus-frequency characteristic must be linear across the input signal's deviation. A *crystal discriminator* uses a crystal as part of its frequency-selective circuitry.

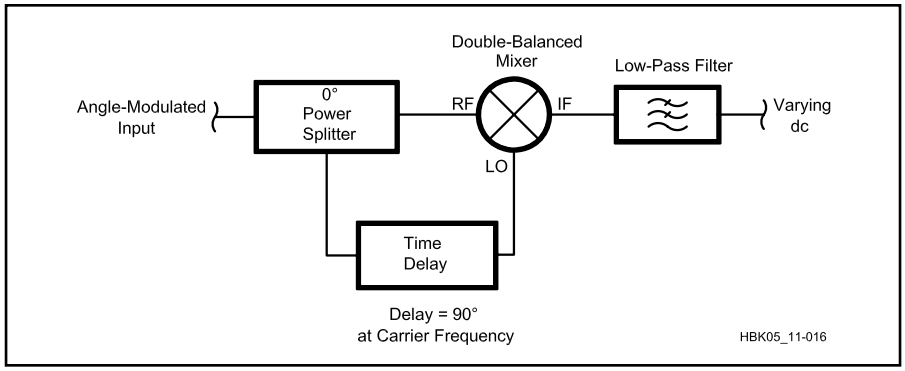


Fig 11.16 — In *quadrature detection*, an angle-modulated signal multiplies with a time-delayed copy of itself to produce a dc voltage that varies with the amplitude and polarity of its phase or frequency excursions away from the carrier frequency. A practical quadrature detector can be as simple as a 0° power splitter (that is, a power splitter with in-phase outputs), a diode double-balanced mixer, a length of coaxial cable ¼-λ (electrical) long at the carrier frequency, and a bit of low-pass filtering to remove the detector output's RF components. IC quadrature detectors achieve their time delay with one or more resistor-loaded tuned circuits (Fig 11.17).

Quadrature Detection

It's also possible to demodulate an angle-modulated signal merely by multiplying it with a time-delayed copy of itself in a double-balanced mixer (**Fig 11.16**). For simplicity's sake, we'll represent the mixer's RF input signal as just a sine wave with an amplitude, *A*

$$A\sin(2\pi ft) \tag{11}$$

and its time-delayed twin, fed to the mixer's LO input, as a sine wave with an amplitude, *A*, and a time delay of d:

$$A\sin[2\pi f(t+d)] \tag{12}$$

Setting this special mixing arrangement into motion, we see

$$\begin{aligned} &A\sin(2\pi ft)\bullet A\sin(2\pi ft+d)\\ &=\frac{A^2}{2}\cos(2\pi fd)-\frac{A^2}{2}\cos(2\pi fd)\cos(2\bullet 2\pi ft)\\ &+\frac{A^2}{2}\sin(2\pi fd)\sin(2\bullet 2\pi ft) \end{aligned} \tag{13}$$

Two of the three outputs — the second and third terms — emerge at twice the input frequency; in practice, we're not interested in these, and filter them out. The remaining term — the one we're after — varies in amplitude and sign according to how far and in what direction the carrier shifts away from its resting or center frequency (at which the time delay, d, causes the mixer's RF and LO inputs to be exactly 90° out of phase — in *quadrature* — with each other). We can examine this effect by replacing f in equations 11 and 12 with the sum term $f_c + f_s$, where f_c is the center frequency and f_s is the frequency shift. A 90° time delay is the same as a quarter cycle of f_c, so we can restate d as

$$d=\frac{1}{4f_c} \tag{14}$$

The first term of the detector's output then becomes

$$\begin{aligned} &\frac{A^2}{2}\cos\left(2\pi(f_c+f_s)\,d\right)\\ &=\frac{A^2}{2}\cos\left(2\pi(f_c+f_s)\frac{1}{4f_c}\right)\\ &=\frac{A^2}{2}\cos\left(\frac{\pi}{2}+\frac{\pi f_s}{2f_c}\right) \end{aligned} \tag{15}$$

When f_s is zero (that is, when the carrier is at its center frequency), this reduces to

$$\frac{A^2}{2}\cos\left(\frac{\pi}{2}\right)=0 \qquad (16)$$

As the input signal shifts higher in frequency than f_c, the detector puts out a positive dc voltage that increases with the shift. When the input signal shifts lower in frequency than f_c, the detector puts out a negative dc voltage that increases with the shift. The detector therefore recovers the input signal's frequency or phase modulation as an amplitude-varying dc voltage that shifts in sign as f_s varies around f_c — in other words, as ac. We have demodulated FM by means of quadrature detection.

An ideal quadrature detector puts out 0 V dc when no modulation is present (with the carrier at f_c). The output of a real quadrature detector may include a small *dc offset* that requires compensation. If we need the detector's response all the way down to dc, we've got it; if not, we can put a suitable blocking capacitor in the output line for ac-only coupling.

Quadrature detection is more common than frequency discrimination nowadays because it doesn't require a special discriminator transformer, and because the necessary balanced-detector circuitry can easily be implemented in IC structures along with limiters and other receiver circuitry. The synchronous AM detector project later in this chapter uses such a chip, the Philips Components-Signetics NE604A, to do limiting and phase detection as part of a phase-locked loop (PLL); **Fig 11.17** shows another example.

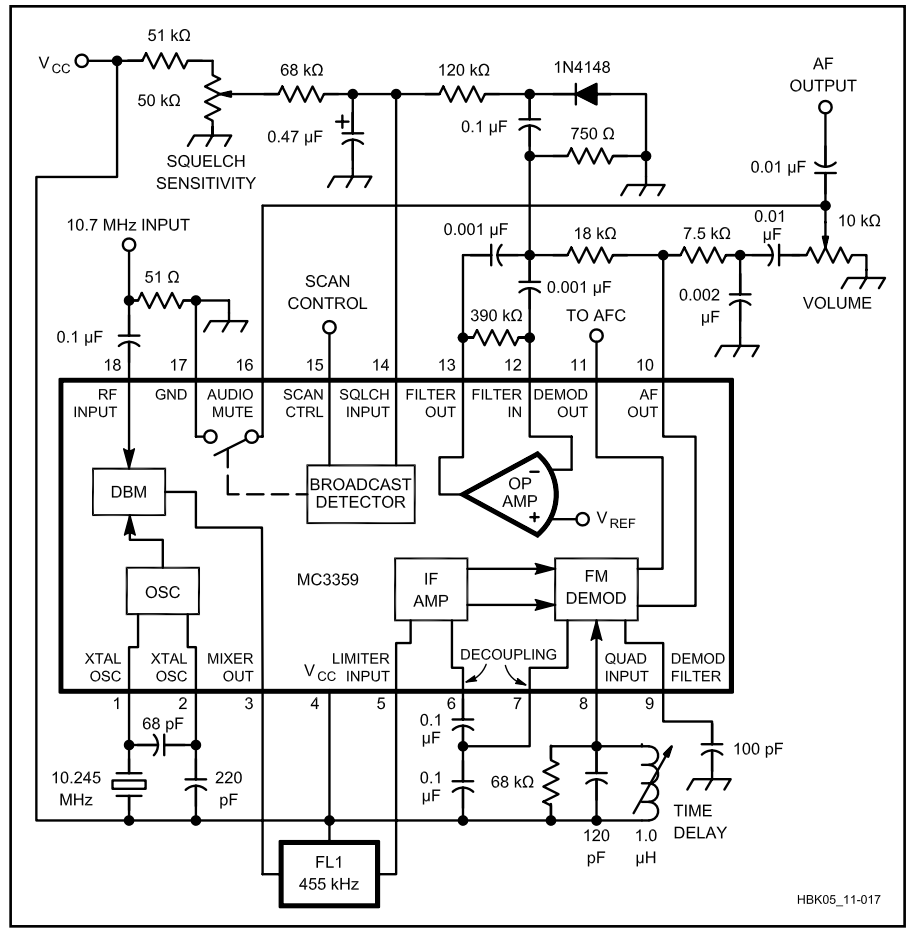


Fig 11.17 — The Motorola MC3359 is one of many FM subsystem ICs that include limiter and quadrature-detection circuitry. The TIME DELAY coil is adjusted for minimum recovered-audio distortion.

PLL Angle Demodulation

Back at Fig 11.14, we saw how a PLL can be used as an angle modulator. A PLL also makes a fine angle *de*modulator. Applying an angle-modulated signal to a PLL keeps its phase detector and VCO hustling to maintain loop lock through the input signal's angle variations. The loop's error voltage therefore tracks the input signal's modulation, and its variations mirror the modulation signal. Turning the loop's varying dc error voltage into audio is just a blocking capacitor away.

Although we can't convey a dc level by directly modulating the VCO in a PLL angle modulator, a PLL demodulator can respond down to dc quite nicely. A constant frequency offset from f_c (a dc component) simply causes a PLL demodulator to swing its VCO over to the new input frequency, resulting in a proportional dc offset on the VCO control-voltage line. Another way of looking at the difference between a PLL angle modulator and a PLL angle demodulator is that a PLL demodulator works with a varying reference signal (the input signal), while a PLL angle modulator generally doesn't.

Amplitude Limiting Required

By now, it's almost household knowledge that FM radio communication systems are superior to AM in their ability to suppress and ignore static, manmade electrical noise and (through an angle-modulation-receiver characteristic called *capture effect*) co-channel signals sufficiently weaker than the desired signal. AM-noise immunity is not intrinsic to angle modulation, however; it must be designed into the angle-modulation receiver.

If we note the progress of A from the left to the right side of the equal sign in equation 13, we realize that the amplitude of a quadrature detector's input signal affects the amplitude of a quadrature detector's three output signals. A quadrature detector therefore responds to AM, and so does a frequency discriminator. To achieve FM's storied noise immunity, then, these angle demodulators must be preceded by *limiting* circuitry that removes all amplitude variations from the incoming signal.

PRACTICAL BUILDING BLOCKS FOR MIXING, MODULATION AND DEMODULATION

So far, we've tended to look at mixing as a process that frequency-shifts one sinusoidal wave by mixing it with another. We need to expand our thinking to other cases, however, since it turns out that many practical mixers work best with *square-wave* signals applied to their LO inputs.

Sine-Wave Mixing, Square-Wave Mixing

Thinking of mixers as multiplying sine waves implies that mixers act like tiny analog computers performing millions of multiplications per second. It's certainly possible to build mixers this way by using an IC mixer circuit (**Fig 11.18**) conceived in 1967 by Barrie Gilbert and widely

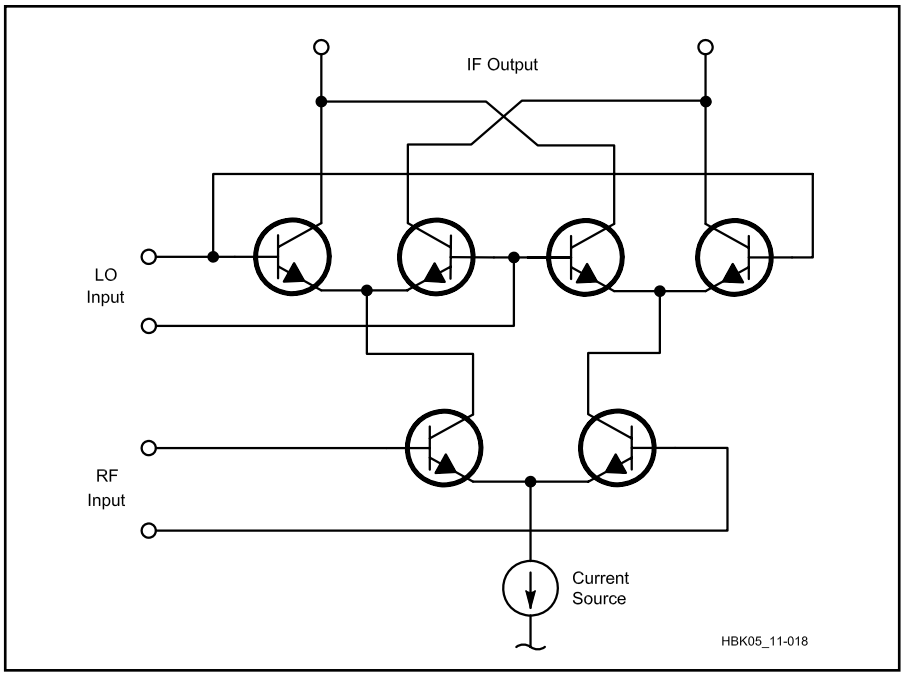


Fig 11.18 — The Gilbert cell mixer. The Motorola MC1496 and Philips Components-Signetics NE602A are based on this circuit.

known as the Gilbert cell. (Gilbert himself was not responsible for this eponym; 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 a true mixer — had already been patented by H. E. Jones.) A Gilbert cell consists of two differential transistor pairs whose bias current is controlled by one of the input signals. The other signal drives the differential pairs' bases, but only after being "predistorted" in a diode circuit. (This circuit distorts the signal equally and oppositely to the inherent distortion of the differential pair.) The resulting output signal is an accurate multiplication of the input voltages.

Early Gilbert-cell ICs, such as the Motorola MC1495 multiplier, had a number of disadvantages, including critical external adjustments, narrow bandwidth, and limited dynamic range. Modern Gilbert cells, such as the Burr-Brown MPY600 and Analog Devices AD834, overcome most of these disadvantages, and have led to an increase in usage of analog multipliers as mixers. Most practical radio mixers do not work exactly as analog multipliers, however. In practice, they act more like fast analog *switches*.

In using a mixer as a fast switching device, we feed its LO input with a single-frequency square wave rather than a sine wave, and feed sine waves, audio, or other complex signals to the mixer's RF input. The RF port serves as the mixer's "linear" input, and therefore must preferably exhibit low intermodulation and harmonic distortion. Feeding a ±1-V square wave into the LO input alternately multiplies the linear input by +1 or –1. Multiplying the RF-port signal by +1 just transfers it to the output with no change. Multiplying the RF-port signal by –1 does the same thing, except that the signal inverts (flips 180° in phase). The LO port need not exhibit low intermodulation and harmonic distortion; all it has to do is preserve the switching signal's fast rise and fall times.

Using square-wave LO drive allows us to simplify the Gilbert multiplier by dispensing with its predistortion circuitry. The Motorola MC1496, still in wide use despite its age, is an example of this. The Philips Components-Signetics NE602A and its relatives, popular with Amateur Radio experimenters, are modern MC1496 descendants. The vast majority of Gilbert-cell mixers in current use are square-wave LO-drive types. In practice, though, we don't have to square the LO signals we apply to them to make them work well. All we need to do is drive their LO inputs with a sine wave of sufficient amplitude to overdrive the associated transistors' bases. This clips the LO waveform, effectively resulting in square-wave drive.

Reversing-Switch Mixers

We can multiply a signal by a square wave without using an analog multiplier at all. All we need is a pair of balun transformers and four diodes (**Fig 11.19A**).

With no LO energy applied to the circuit, none of its diodes conduct. RF-port energy (1) can't make it to the LO port because there's no direct connection between the secondaries of T1 and T2, and (2) doesn't produce IF output because T2's secondary balance results in energy cancellation at its center tap, and because no complete IF-energy circuit exists through T2's secondary with both of its ends disconnected from ground.

Applying a square wave to the LO port biases the diodes so that, 50% of the time, D1 and D2 are on and D3 and D4 are reverse-biased off. This unbalances T2's secondary by leaving its upper wire floating and connecting its lower wire to ground through T1's secondary and center tap. With T2's secondary unbalanced, RF-port energy emerges from the IF port.

The other 50% of the time, D3 and D4 are on and D1 and D2 are reverse-biased off. This unbalances T2's secondary by leaving its lower wire floating, and connects its upper wire to ground through T1's secondary and center tap. With T2's secondary unbalanced, RF-port energy again emerges from the IF port — shifted 180° relative to the first case because T2's active secondary wires are now, in effect, transposed relative to its primary.

A reversing switch mixer's output spectrum is the same as the output spectrum of a multiplier fed with a square wave. This can be analyzed by thinking of the square wave in terms of its Fourier series equivalent, which consists of the sum of sine waves at the square wave frequency and all of its odd harmonics. The amplitude of the equivalent series' fundamental sine wave is $4/\pi$ times (2.1 dB greater than) the amplitude of the square wave. The amplitude of each harmonic is inversely proportional to its harmonic number, so the third harmonic is only $^1/_3$ as strong as the fundamental (9.5 dB below the fundamental), the 5th harmonic is only $^1/_5$ as strong (14 dB below the fundamental) and so on. The input signal mixes with each harmonic separately from the others, as if each harmonic were driving its own separate mixer, just as we illustrated with two sine waves in Fig 11.4. Normally, the harmonic outputs are so widely removed from the desired output frequency that they are easily filtered out, so a reversing-switch mixer is just as good as a sine-wave-driven analog multiplier for most practical purposes, and usually better — for radio purposes — in terms of dynamic range and noise.

An additional difference between multiplier and switching mixers is that a switching mixer's signal flow is reversible. It really only has one dedicated input (the LO input). The other terminals can be

thought of as I/O (input/output) ports, since either one can be the input as long as the other is the output.

Conversion Loss

Fig 11.19B shows a perfect *multiplier* mixer. That is, the output is the product of the input signal and the LO. The LO is a perfect square wave. Its peak amplitude is ±1.0 V and its frequency is 8 MHz. Fig 11.19C shows the output waveform (the product of two inputs) for an input signal whose value is 0 dBm and whose frequency is 2 MHz. Notice that for each transition of the square-wave LO, the sine-wave output-waveform polarity reverses. There are 16 transitions during the interval shown, at each zero-crossing point of the output waveform. Fig 11.19D shows the mixer output spectrum. The principle components are at 6 MHz and 10 MHz, which are the sum and difference of the signal and LO frequencies. Each of these is –3.9 dBm. Numerous other pairs of output frequencies occur that are also spaced 4 MHz apart and centered at 24 MHz, 40 MHz and 56 MHz and higher odd harmonics of 8 MHz. The ones shown are at –13.5 dBm, –17.9 dBm and –20.9 dBm. Because the mixer is lossless, the sum of all of the outputs must be exactly equal to the value of the input signal. As explained previously, this output spectrum can also be understood in terms of each of the odd-harmonic components of the square-wave LO operating independently.

If the mixer were a lossless switching mixer, such as Fig 11.19A, with diodes that are perfect switches, the results would be mathematically identical to the above example. The diodes would commutate the input signal exactly as shown in Fig 11.19C.

Now consider the perfect multiplier mixer of Fig 11.19B with an LO that is a perfect sine wave with a peak amplitude of ±1.0 V. In this case the dashed lines of Fig 11.19D show that only two output frequencies are present, at 6 MHz and 10 MHz (see also Fig 11.2). Each component now has a –6 dBm level. The product of the 0 dBm sine-wave input at one frequency and the ±1.0V sine-wave LO at another frequency (see Eq 6 in this chapter) is the –3 dBm total output.

These examples illustrate the difference between the square-wave LO and the sine-wave LO, for a perfect multiplier. For the same peak value of both LO waves, the square-wave LO delivers 2.1 dB more output at 6 MHz and 10 MHz than the sine-wave LO. An actual diode mixer such as Fig 11.19A behaves more like a switching mixer. Its sine-wave LO waveform is considerably flattened by interaction between

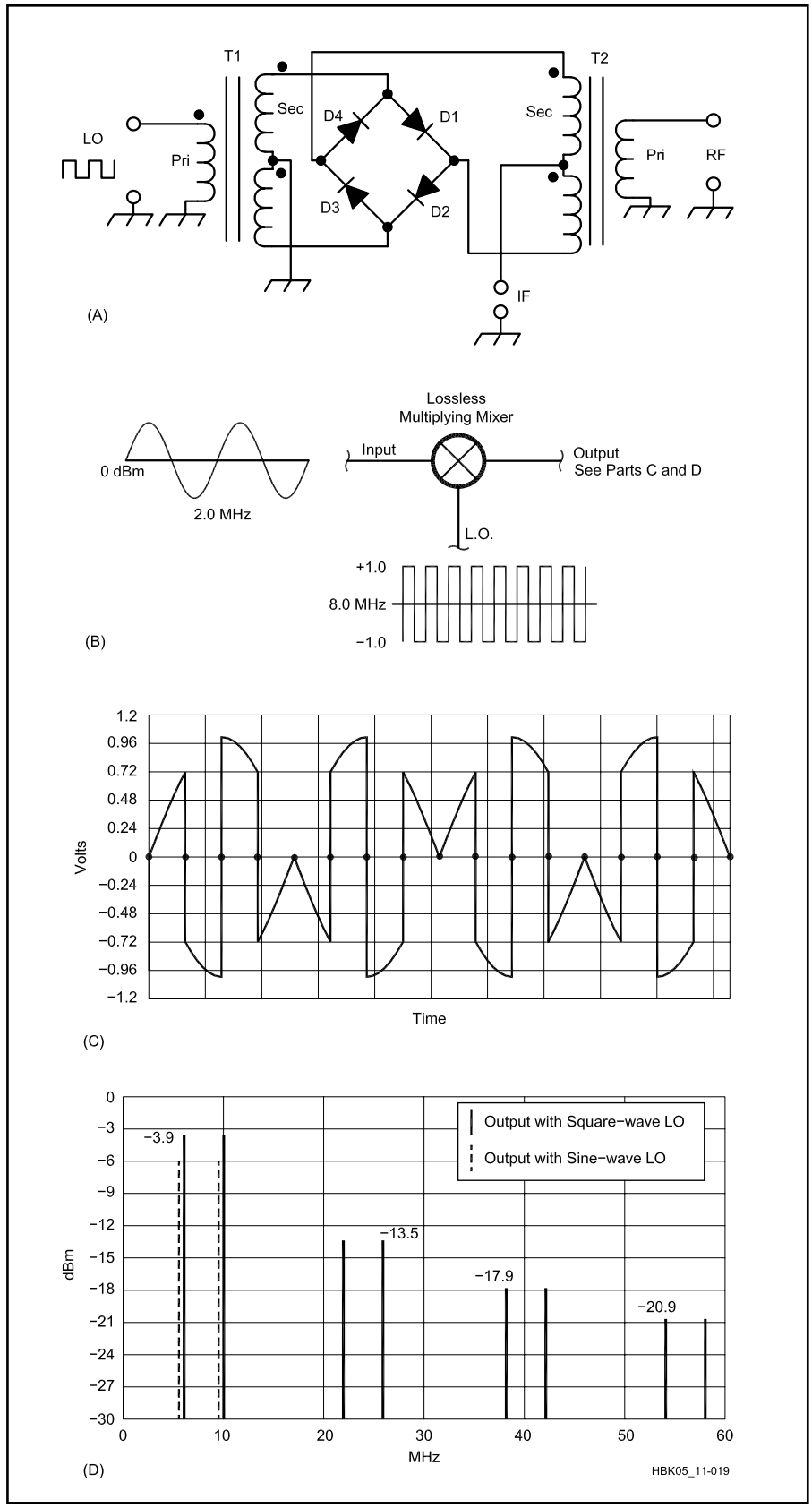


Fig 11.19 — Part A shows a general-purpose diode *reversing-switch* mixer. This mixer uses a square-wave LO and a sine-wave input signal. The action of this mixer is described in the text. Part B is an ideal *multiplier* mixer. The square-wave LO and a sine-wave input signal produce the output waveform shown in part C. The solid lines of part D show the output spectrum with the square-wave LO. The dashed lines show the output spectrum with a sine-wave LO.

the diodes and the LO generator, so that it looks somewhat like a square wave. The diodes have nonlinearities, junction voltages, capacitances, resistances and imperfect parameter matching. (See the **Real-World Component Characteristics** chapter.) Also, "re-mixing" of a diode mixer's output with the LO and the input is a complicated possibility. The practical end result is that diode double-balanced mixers have a conversion loss, from input to each of the two major output frequencies, in the neighborhood of 5 to 6 dB.

The Diode Double-Balanced Mixer: A Basic Building Block

The most common implementation of a reversing switch mixer is the diode *double-balanced mixer* (DBM). DBMs can serve as mixers (including image-reject types), modulators (including single- and double-sideband, phase, biphase, and quadrature-phase types) and demodulators, limiters, attenuators, switches, phase detectors, and frequency doublers. In some of these applications, they work in conjunction with power dividers, combiners and hybrids.

The Basic DBM Circuit

We have already seen the basic diode DBM circuit (Fig 11.19). In its simplest form, a DBM contains two or more unbalanced-to-balanced transformers and a Schottky diode ring consisting of 4 × n diodes, where n is the number of diodes in each leg of the ring. Each leg commonly consists of up to four diodes.

As we've seen, the degree to which a mixer is *balanced* depends on whether either, neither or both of its input signals (RF and LO) emerge from the IF port along with mixing products. An unbalanced mixer suppresses neither its RF nor its LO; both are present at its IF port. A single-balanced mixer suppresses its RF or LO, but not both. A double-balanced mixer suppresses its RF *and* LO inputs. Diode and transformer uniformity in the Fig 11.19 circuit results in equal LO potentials at the center taps of T1 and T2. The LO potential at T1's secondary center tap is zero (ground); therefore, the LO potential at the IF port is zero.

Balance in T2's secondary likewise results in an RF null at the IF port. The RF potential between the IF port and ground is therefore zero — except when the DBM's switching diodes operate, of course!

The Fig 11.19 circuit normally also affords high RF-IF isolation because its balanced diode switching precludes direct connections between T1 and T2. A diode DBM can be used as a current-controlled switch or attenuator by applying dc to its IF port, albeit with some distortion. This causes opposing diodes (D2 and D4, for instance) to conduct to a degree that depends on the current magnitude, connecting T1 to T2.

One extension of the single-diode-ring DBM is a *double* double-balanced mixer (DDBM) with high dynamic range and larger signal handling capability than a single-ring design. **Fig 11.20** diagrams such a DDBM, which uses transmission-line transformers and two diode rings. This type of mixer has higher 1-dB compression point (usually 3 to 4 dB lower than the LO drive) than a DBM. Low distortion is a typical characteristic of DDBMs. Depending on the ferrite core material used (ferrites with a magnetic permeability — μ — of 100 to 15,000), frequencies as low as a few hundred hertz and as high as a few gigahertz can be covered.

Diode DBM Components

Commercially manufactured diode DBMs generally consist of: (A) a supporting base; (B) a diode ring; (C) two or more ferrite-core transformers commonly wound with two or three twisted-pair wires; (D) encapsulating material; and (E) an enclosure.

Diodes

Hot-carrier (Schottky) diodes are the devices of choice for diode-DBM rings because of their low ON resistance, although ham-built DBMs for non-critical MF/HF use commonly use switching diodes like the 1N914 or 1N4148. The forward voltage drop, V_f, across each diode in the ring determines the mixer's optimum local-oscillator drive level. Depending on the forward voltage drop of each of its diodes and the number of diodes in each ring leg, a diode DBM may be categorized as a Level 0, 3, 7, 10, 13, 17, 23 or 27 device. The numbers indicate the mixer's optimal LO drive level in dBm. As a rule of thumb, the LO signal must be 20 dB larger than the RF and IF signals for proper operation. This ensures that the LO signal, rather than the RF or IF signals, switches the mixer's diodes on and off — a critical factor in minimizing IMD and maximizing dynamic range.

Schottky diodes are characterized by loss and contact resistance (R_s), junction capacitance (C_j), and forward voltage drop (V_f) at a known current, typically 1 mA or 10 mA. The lower the diode-to-diode V_f difference in millivolts, the better the diode match at dc. (Some early diode DBM designs used diodes in series with a parallel resistor/capacitor combination for automatic bias-

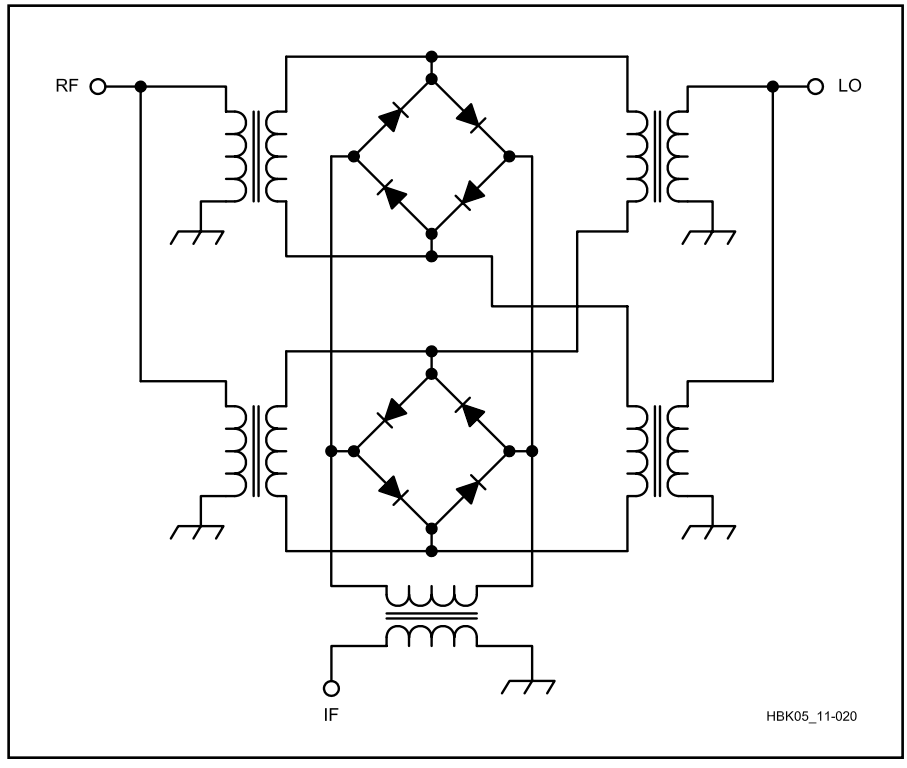


Fig 11.20 — Five transmission-line transformers and two Schottky-quad rings form this double double-balanced mixer (DDBM). Such designs can provide lower distortion, better signal-handling capability and higher interport isolation than single-ring designs.

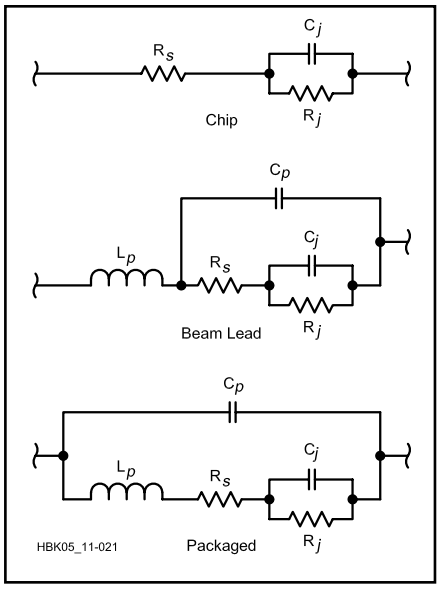


Fig 11.21 — Its semiconductive properties aside, a Schottky diode can be represented as a network consisting of resistance, capacitance and/or inductance. Of these, the junction capacitance (C_j) plays an especially critical role in a double-balanced mixer's high-end response. (R_j = junction resistance, R_s = contact resistance, L_p = parasitic inductance and C_p = parasitic capacitance.)

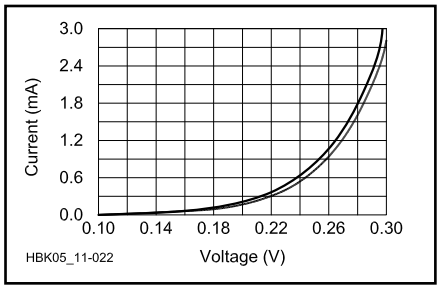


Fig 11.22 — Current-voltage (I-V) characteristic for Schottky diode quad, showing worst-case voltage imbalance (the spread between the two curves) among the four diodes.

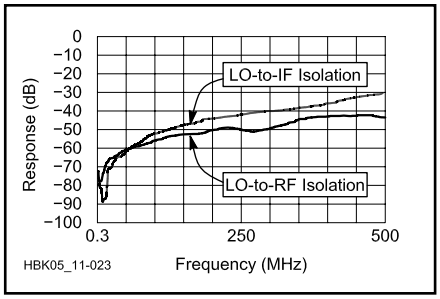


Fig 11.23 — A diode DBM's port-to-port isolation depends on how well its diodes match and how well its transformers are balanced. This graph shows LO-IF and LO-RF isolation versus frequency for a Synergy Microwave CLP-4A3 mixer.

ing.) Better diode matching (in V_f and C_j) results in higher isolation among the ports. Diodes capable of operating at higher frequencies have lower junction capacitance and lower parasitic inductance. **Fig 11.21** shows the equivalent circuit for Schottky diodes of three package types.

Manufacturers of diodes suitable for DBMs characterize their diodes as low-barrier, medium-barrier, high-barrier and very-high-barrier (usually two or more diodes in each leg), with typical V_f values of 220 mV, 350 mV, 600 mV and 1 V or more, respectively. **Fig 11.22** shows a typical current-voltage (I-V) characteristic for a low-barrier Schottky quad capable of operating up to 4 GHz. Note that as current through the diodes increases, the V_f difference among the ring's diodes also increases, affecting the balance.

At higher frequencies, diode packaging becomes critical and expensive. As the frequency of operation increases, the effect of junction capacitance and package capacitance cannot be ignored. Part or all of the capacitance can be compensated at the mixer's highest operating frequency by properly designing the unbalanced-to-balanced transformers. The transformer inductance and diode junction capacitance form a low-pass network with its cutoff frequency higher than the frequency of operation. Compensated in this way, diodes with a junction capacitance of 0.2 pF can be used up to 8 to 10 GHz.

Transformers

From the DBM schematic shown in Fig 11.19, it's clear that the LO and RF transformers are unbalanced on the input side and balanced on the diode side. The diode ends of the balanced ports are 180° out of phase throughout the frequency range of interest. This property causes signal cancellations that result in higher port-to-port isolation. **Fig 11.23** plots LO-RF and LO-IF isolation versus frequency for Synergy Microwave's CLP-4A3 DBM. Isolations on the order of 70 dB occur at the lower end of the band as a direct result of the balance among the four diode-ring legs and the RF phasing of the balanced ports.

As we learned in our discussion of generic switching mixers, transformer efficiency plays an important role in determining a mixer's conversion loss and drive-level requirement. Core loss, copper loss and impedance mismatch all contribute to transformer losses.

Ferrite in toroidal, bead, balun (multihole), or rod form can serve as DBM transformer cores. Radio amateurs commonly use Fair-rite Mix 43 ferrite (μ = 950), but if the mixer will be used over a wide temperature range, the core material must be evaluated in terms of temperature coefficient and curie temperature (the temperature at which a ferromagnetic material loses its magnetic properties). In some materials, μ may change drastically across the desired temperature range, causing a frequency-response shift with temperature. Once a suitable core material and form have been selected, frequency requirements determine the necessary core size. For a given core shape and size, the number of turns, wire size, and the number of twists determine transformer performance. Wire placement also plays an important role.

RF transformers combine lumped and distributed capacitance and inductance. The interwinding capacitance and characteristic impedance of a transformer's twisted wires sets the transformer's high-frequency response. The core's μ and size, and the number of winding turns, determine the transformer's lower frequency limit. Covering a specific frequency range requires a compromise in the number of turns used with a given core. Increasing a transformer's core size and number of turns improves its low-frequency response. Cores may be stacked to meet low-frequency performance specs.

Inexpensive mixers operating up to 2 GHz most commonly use twisted trifilar (three-wire) windings made of a wire size between #36 and #32. The number of twists per unit length of wire determines a winding's characteristic impedance. Twisted wires are analogous to transmission lines, and can be analyzed in terms of distributed interwinding capacitance. Decreasing the number of twists lowers the interwinding capacitance and increases the frequency of operation. On the other hand, using fewer twists per inch than four makes winding difficult because the wires tend to separate instead of behaving as a single cable.

The transmission-line effect predominates at the higher end of a transformer's frequency range. If two impedances, Z_1 and Z_2, need to be matched through a transmission line of characteristic impedance, Z_0, then

$$Z_0 = \sqrt{Z_1 \times Z_2} \qquad (17)$$

Fig 11.24 shows two types of transformers using twisted wires: (A) a three-wire type in which the primary winding is isolated from the secondary winding with a center tap, and (B) a two-wire (transmission-line) type in which two sets of transmission lines are interconnected to form a center tap at the secondary with a direct connection between primary and second-

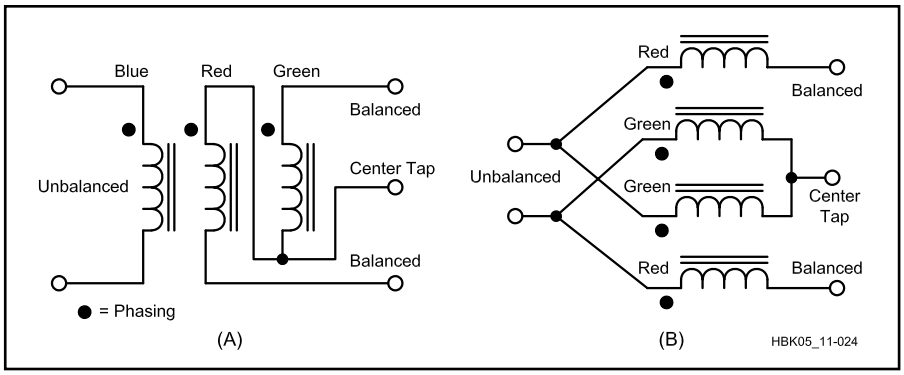


Fig 11.24 — Transformers for DBMs: three-wire (A), and transmission-line (B).

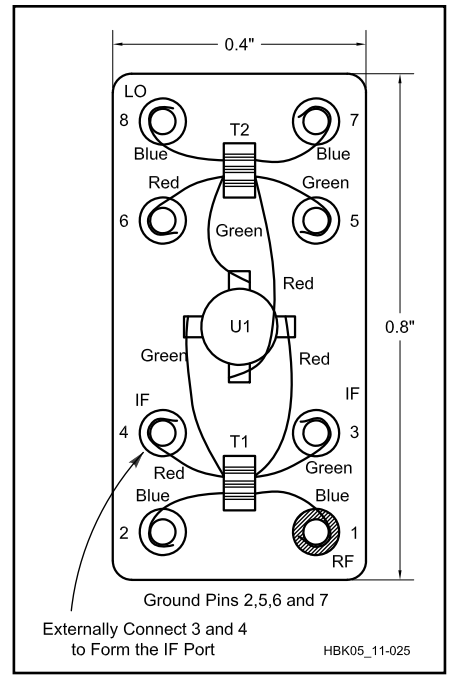


Fig 11.25 — How a typical commercial DBM is wired. The use of different wire colors for the transformers' various windings speeds assembly and minimizes error. U1, a Schottky-diode quad, contains D1-D4 of Fig 11.19.

ary. The primary-to-secondary turns ratio determines the impedance match as shown in equation 17. The properties of these two transformer types can be summarized as follows:

1. By virtue of its construction, the three-wire transformer is less symmetrical at higher frequencies than the transmission-line type.

2. The transformers' lower cutoff frequency (f_L) is determined by the equation

$$\omega L > 4R \qquad (18)$$

where

L = inductance of the winding

R = system impedance; for example, 50 Ω, 75 Ω and so on; and

$\omega = 2\pi f_L$.

3. The transmission-line transformer's upper cutoff frequency (f_H) is determined by the highest frequency at which its wires' twists (that is, the coupling between them) allow it to function as a transmission line of the proper characteristic impedance.

4. Transformers convert one impedance, Z_1 (primary) to another, Z_2 (secondary) according to the relationship

$$Z_2 = Z_1 (N)^2 \qquad (19)$$

where

N = secondary to primary turns ratio. Within certain limits, if Z_1 is varied, Z_2 also varies to a new value multiplied by N^2. Thus, a mixer designed for a 50-Ω system may work in a 75-Ω system with minor modifications.

Diode DBMs in Practice

Fig 11.25 shows the wiring of a typical commercial DBM made with toroidal cores. The wires are wrapped around the package pins and diode leads, and then soldered. In this unit, the LO transformer's primary winding connects across pins 7 and 8; the RF-transformer primary, across pins 1 and 2. The pin pairs 3-4 and 5-6 are connected externally to form the transformers' secondary center taps, one of which (5-6, that of the LO transformer) connects to a common ground point while the other (3-4, that of the RF transformer) serves as the IF port.

The DBM shown in Fig 11.25 has a dc-coupled IF port. If necessary, this DBM can be operated at a particular polarity (positive or negative) by appropriately connecting the LO, RF, IF and common ground points.

DBM Specifications

Most of the parameters important in building or selecting a diode DBM also apply to other mixer types. They include: conversion loss and amplitude flatness across the required IF bandwidth; variation of conversion loss with input frequency; variation of conversion loss with LO drive, 1-dB compression point; LO-RF, LO-IF and RF-IF isolation; intermodulation products; noise figure (usually within 1 dB of conversion loss); port SWR; and dc offset, which is directly related to isolation among the RF, LO and IF ports.

Conversion Loss

Fig 11.26 shows a plot of conversion loss versus intermediate frequency in a typical DBM. The curves show conversion loss for two fixed RF-port signals, one at 100 kHz and the another at 500 MHz, while varying the LO frequency from 100 kHz to 500 MHz.

Fig 11.27 plots a diode DBM's simulated output spectrum. Note that the RF input is –20 dBm and the IF output (the frequency difference between the RF and LO signals) is –25 dBm, implying a conversion loss of 5 dB. This figure also applies to the sum of both signals (RF + LO).

We minimize a diode DBM's conversion loss, noise figure and intermodulation by keeping its LO drive high enough to switch its diodes on fully and rapidly. **Fig 11.28** plots noise figure for the DDBM shown in Fig 11.20. Its 4-dB noise figure assumes ideal transformers and somewhat idealized diodes; typical mixers have a noise figure of 5 to 6 dB. **Fig 11.29** plots conversion gain (loss) for the same mixer circuit.

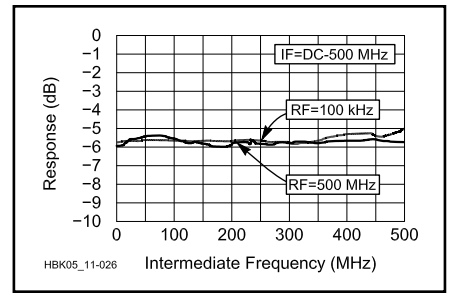


Fig 11.26 — Conversion loss versus frequency for a typical diode DBM. The LO drive level is +7 dBm.

Insufficient LO drive results in increased noise figure and conversion loss. IMD also increases because RF-port signals have a greater chance to control the mixer diodes when the LO level is too low.

Dynamic Range: Compression, Intermodulation and More

The output of a linear stage — including a mixer, which we want to act as a

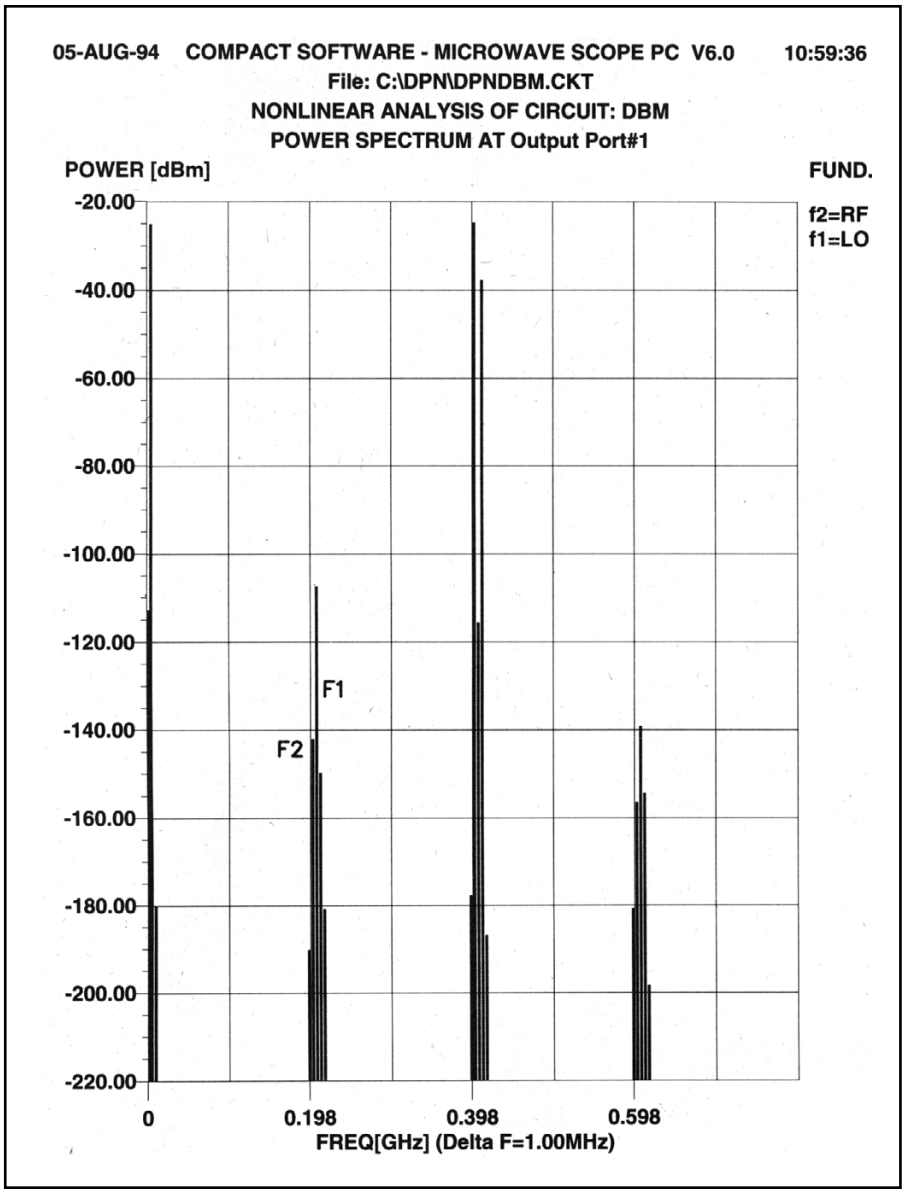


Fig 11.27 — Simulated diode-DBM output spectrum. Note that the desired output products (the highest two products, RF – LO and RF + LO) emerge at a level 5 dB below the mixer's RF input (–20 dBm). This indicates a mixer conversion loss of 5 dB. (*Microwave SCOPE* simulation)

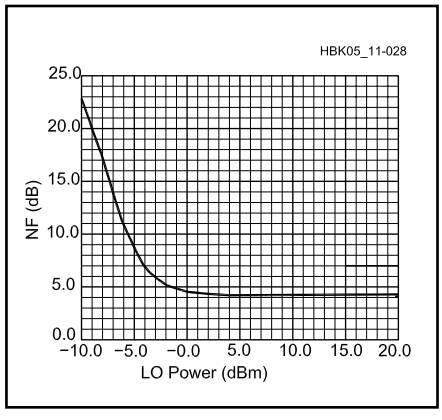


Fig 11.28 — Noise figure versus LO drive for a DDBM built along the lines of Fig 11.20.

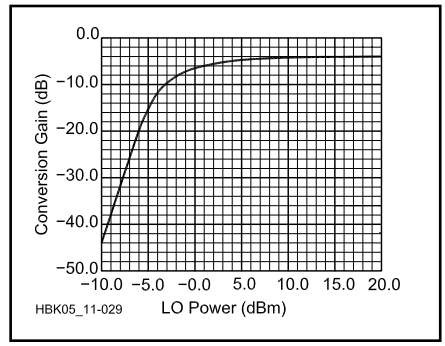


Fig 11.29 — Conversion gain (loss) for a DDBM. Increasing a mixer's LO level beyond that sufficient to turn its switching devices all the way on merely makes them dissipate more LO power and does not improve performance.

linear frequency shifter — tracks its input signal decibel by decibel, every 1-dB change in its input signal(s) corresponding to an identical 1-dB output change. This is the stage's *first-order* response.

Because no device is perfectly linear, however, two or more signals applied to it generate sum and difference frequencies. These IMD products occur at frequencies and amplitudes that depend on the order of the IMD response as follows:

- *Second-order* IMD products change 2 dB for every decibel of input-signal change (this figure assumes that the IMD comes from equal-level input signals), and appear at frequencies that result from the simple addition and subtraction of input-signal frequencies. For example, assuming that its input bandwidth is sufficient to pass them, an amplifier subjected to signals at 6 and 8 MHz will produce second-order IMD products at 2 MHz (8–6) and 14 MHz (8+6).
- *Third-order* IMD products change 3 dB for every decibel of input-signal change (this also assumes equal-level input signals), and appear at frequencies corresponding to the sums and differences of twice one signal's frequency plus or minus the frequency of another. Assuming that its input bandwidth is sufficient to pass them, an amplifier subjected to signals at 14.02 MHz (f_1) and 14.04 MHz (f_2) produces third-order IMD products at 14.00 ($2f_1 - f_2$), 14.06 ($2f_2 - f_1$), 42.08 ($2f_1 + f_2$) and 42.10 ($2f_2 + f_1$) MHz. The subtractive products (the 14.02 and 14.04-MHz products in this example) are close to the desired signal and can cause significant interference. Thus, third-order-IMD performance is of great importance in receiver mixers and RF amplifiers.

It can be seen that the IMD order determines how rapidly IMD products change level per unit change of input level. *N*th-order IMD products therefore change by *n* dB for every decibel of input-level change.

IMD products at orders higher than three can and do occur in communication systems, but the second- and third-order products are most important in receiver front ends. In transmitters, third- and higher-odd-order products are important because they widen the transmitted signal.

Intercept Point

The second type of dynamic range concerns the receiver's *intercept point*, sometimes simply referred to as *intercept*. Intercept point is typically measured by applying two or three signals to the antenna input, tuning the receiver to count the number of resulting spurious

Testing and Calculating Intermodulation Distortion in Receivers

Second and third-order IMD can be measured using the setup of **Fig A**. The outputs of two signal generators are combined in a 3-dB hybrid coupler. Such couplers are available from various companies, and can be homemade. The 3-dB coupler should have low loss and should itself produce negligible IMD. The signal generators are adjusted to provide a known signal level at the output of the 3-dB coupler, say, –20 dBm for each of the two signals. This combined signal is then fed through a calibrated variable attenuator to the device under test. The shielding of the cables used in this system is important: At least 90 dB of isolation should exist between the high-level signal at the input of the attenuator and the low-level signal delivered to the receiver.

The measurement procedure is simple: adjust the variable attenuator to produce a signal of known level at the frequency of the expected IMD product ($f_1 \pm f_2$ for second-order, $2f_1 - f_2$ or $2f_2 - f_1$ for third-order IMD).

To do this, of course, you have to figure out what equivalent input signal level at the receiver's operating frequency corresponds to the level of the IMD product you are seeing. There are several ways of doing this. One way — the way used by the ARRL Lab in their receiver tests — uses the minimum discernible signal. This is defined as the signal level that produces a 3-dB increase in the receiver audio output power. That is, you measure the receiver output level with no input signal, then insert a signal at the operating frequency and adjust the level of this input signal until the output power is 3 dB greater than the no-signal power. Then, when doing the IMD measurement, you adjust the attenuator of Fig A to cause a 3-dB increase in receiver output. The level of the IMD product is then the same as the MDS level you measured.

There are several things I dislike about doing the measurement this way. The problem is that you have to measure noise power. This can be difficult. First, you need an RMS voltmeter or audio power meter to do it at all. Second, the measurement varies with time (it's noise!), making it difficult to nail down a number. And third, there is the question of the audio response of the receiver; its noise output may not be flat across the output spectrum. So I prefer to measure, instead of MDS, a higher reference level. I use the receiver's S meter as a reference. I first determine the input signal level it takes to get an S1 reading. Then, in the IMD measurement, I adjust the attenuator to again give an S1 reading. The level of the IMD product signal is now equal to the level I measured at S1. Note that this technique gives a different IMD level value than the MDS technique. That's OK, though. What we are trying to determine is the *difference* between the level of the signals applied to the receiver input and the level of the IMD product. Our calculations will give the same result whether we measure the IMD product at the MDS level, the S1 level or some other level.

An easy way to make the reference measurement is with the setup of Fig A. You'll have to switch in a lot of attenuation (make sure you have an attenuator with enough range), but doing it this way keeps all of the possible variations in the measurement fairly constant. And this way, the difference between the reference level and the input level needed to produce the desired IMD product signal level is simply the difference in attenuator settings between the reference and IMD measurements.

Calculating Intercept Points

Once we know the levels of the signals applied to the receiver input and the level of the IMD product, we can easily calculate the intercept point using the following equation:

$$IP_n = \frac{n \bullet P_A - P_{IM_n}}{n - 1} \qquad (A)$$

Here, n is the order, P_A is the receiver input power (of one of the input signals), P_{IMn} is the power of the IMD product signal, and IP_n is the nth-order intercept point. All powers should be in dBm. For second and third-order IMD, equation A results in the equations:

$$IP_2 = \frac{2 \bullet P_A - P_{IM_2}}{2 - 1} \qquad (B)$$

$$IP_3 = \frac{3 \bullet P_A - P_{IM_3}}{3 - 1} \qquad (C)$$

You can measure higher-order intercept points, too.

Example Measurements

To get a feel for this process, it's useful to consider some actual measured values.

The first example is a Rohde & Schwarz model EK085 receiver with digital preselection. For measuring second-order IMD, signals at 6.00 and 8.01 MHz, at –20 dBm each, were applied at the input of the attenuator. The difference in attenuator settings between the reference measurement and the level needed to produce the desired IMD product signal level

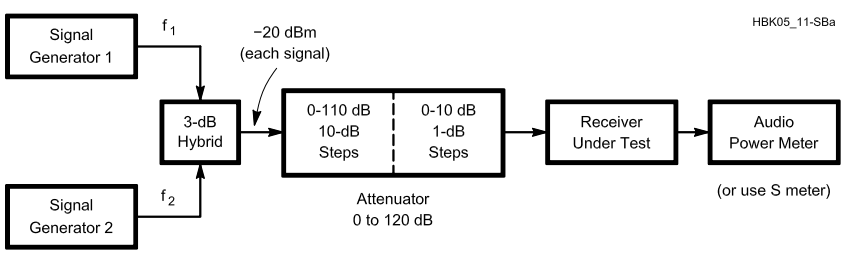


Fig A — Test setup for measurement of IMD performance. Both signal generators should be types such as HP 608, HP 8640, or Rhode & Schwarz SMDU, with phase-noise performance of –140 dBc/Hz or better at 20 kHz from the signal frequency.

was found to be 125 dB. The calculation of the second-order IP is then:

$$IP_2 = \frac{2(-20\,\text{dBm}) - (-20\,\text{dBm} - 125\,\text{dB})}{2-1}$$
$$= -40\,\text{dBm} + 20\,\text{dBm} + 125\,\text{dB} = +105\,\text{dBm}$$

For IP_3, we set the signal generators for 0 dBm at the attenuator input, using frequencies of 14.00 and 14.01 MHz. The difference in attenuator settings between the reference and IMD measurements was 80 dB, so:

$$IP_3 = \frac{3(0\,\text{dBm}) - (0\,\text{dBm} - 80\,\text{dB})}{3-1}$$
$$= \frac{0\,\text{dBm} + 80\,\text{dB}}{2} = +40\,\text{dBm}$$

We also measured the IP_3 of a Yaesu FT-1000D at the same frequencies, using attenuator-input levels of –10 dBm. A difference in attenuator readings of 80 dB resulted in the calculation:

$$IP_3 = \frac{3(-10\,\text{dBm}) - (-10\,\text{dBm} - 80\,\text{dB})}{3-1}$$
$$= \frac{-30\,\text{dBm} + 10\,\text{dBm} + 80\,\text{dB}}{2}$$
$$= \frac{-20\,\text{dBm} + 80\,\text{dB}}{2}$$
$$= +30\,\text{dBm}$$

Synthesizer Requirements

To be able to make use of high third-order intercept points at these close-in spacings requires a low-noise LO synthesizer. You can estimate the required noise performance of the synthesizer for a given IP_3 value. First, calculate the value of receiver input power that would cause the IMD product to just come out of the noise floor, by solving equation A for P_A, then take the difference between the calculated value of P_A and the noise floor to find the dynamic range. Doing so gives the equation:

$$ID_3 = \frac{2}{3}(IP_3 + P_{min}) \qquad (D)$$

Where ID_3 is the third-order IMD dynamic range in dB and P_{min} is the noise floor in dBm. Knowing the receiver bandwidth, BW (2400 Hz in this case) and noise figure, NF (8 dB) allows us to calculate the noise floor, P_{min}:

$$P_{min} = -174\,\text{dBm} + 10\log(\text{BW}) + \text{NF}$$
$$= -174\,\text{dBm} + 10\log(2400) + 8$$
$$= -132\,\text{dBm}$$

The synthesizer noise should not exceed the noise floor when an input signal is present that just causes an IMD product signal at the noise floor level. This will be accomplished if the synthesizer noise is less than:

$$\text{ID} + 10\log(\text{BW}) = 114.7\,\text{dB} + 10\log(2400)$$
$$= 148.5\,\text{dBc/Hz}$$

in the passband of the receiver. Such synthesizers hardly exist.—*Dr. Ulrich L. Rohde, KA2WEU*

responses, and measuring their level relative to the input signal.

Because a device's IMD products increase more rapidly than its desired output as the input level rises, it might seem that steadily increasing the level of multiple signals applied to an amplifier would eventually result in equal desired-signal and IMD levels at the amplifier output. Real devices are incapable of doing this, however. At some point, every device *overloads*, and changes in its output level no longer equally track changes at its input. The device is then said to be operating in *compression*; the point at which its first-order response deviates from linearity by 1 dB is its *1-dB compression point*. Pushing the process to its limit ultimately leads to *saturation*, at which point input-signal increases no longer increase the output level.

The power level at which a device's second-order IMD products equal its first-order output (a point that must be extrapolated because the device is in compression by this point) is its *second-order intercept point*. Likewise, its *third-order intercept point* is the power level at which third-order responses equal the desired signal. **Fig 11.30** graphs these relationships.

Input filtering can improve second-order intercept point; device non-linearities

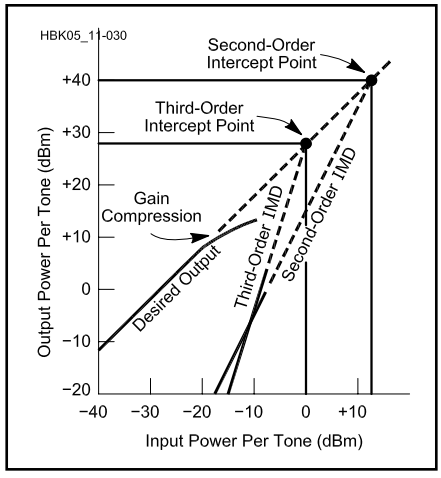
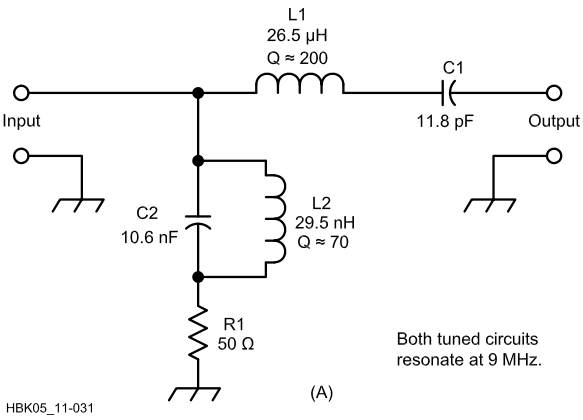
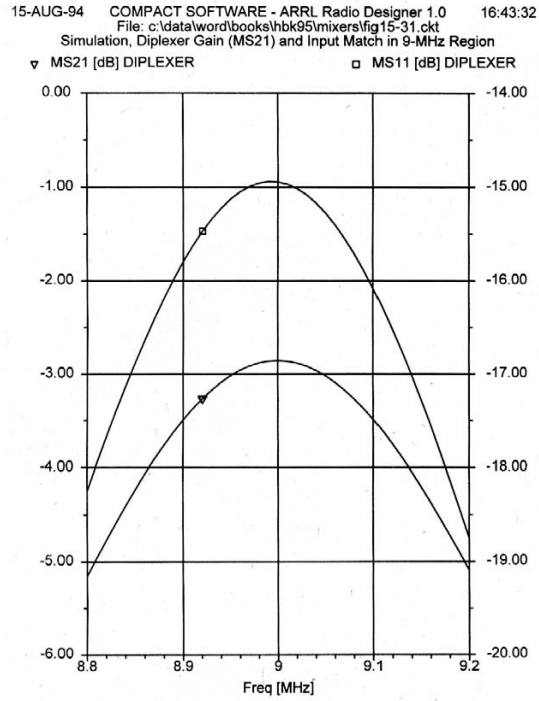
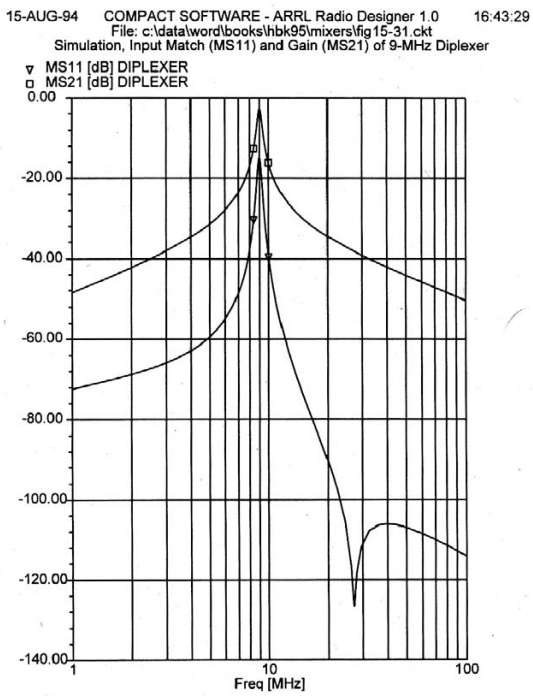


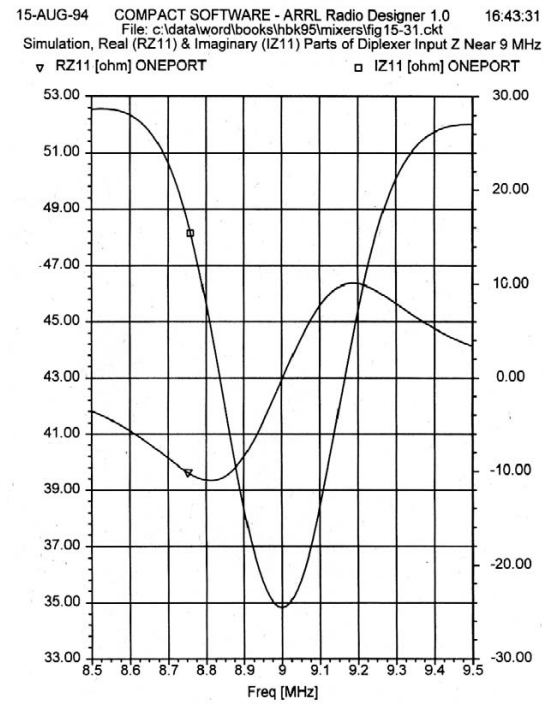
Fig 11.30 — A linear stage's output tracks its input decibel by decibel on a 1:1 slope — its *first-order* response. *Second-order intermodulation distortion (IMD) products* produced by two equal-level input signals ("tones") rise on a 2:1 slope — 2 dB for every 1 dB of input increase. *Third-order IMD products* likewise increase 3 dB for every 1 dB of increase in two equal tones. For each IMD order *n*, there is a corresponding *intercept point* IP_n at which the stage's first-order and *n*th order products are equal in amplitude. The first-order output of real amplifiers and mixers falls off (the device overloads and goes into *compression*) before IMD products can intercept it, but intercept point is nonetheless a useful, valid concept for comparing radio system performance. The higher an amplifier or mixer's intercept point, the stronger the input signals it can handle without overloading. The input and output powers shown are for purposes of example; every device exhibits its own particular IMD profile. (After W. Hayward, *Introduction to Radio Frequency Design*, Fig 6.17)

Fig 11.31 — A diplexer resistively terminates energy at unwanted frequencies while passing energy at desired frequencies. This band-pass diplexer (A) uses a series-tuned circuit as a selective pass element, while a high-C parallel-tuned circuit keeps the network's terminating resistor R1 from dissipating desired-frequency energy. Computer simulation of the diplexer's response with *ARRL Radio Designer 1.0* characterizes the diplexer's insertion loss and good input match from 8.8 to 9.2 MHz (B) and from 1 to 100 MHz (C); and the real and imaginary components of the diplexer's input impedance from 8.8 to 9.2 MHz with a 50-Ω load at the diplexer's output terminal (D). The high-C, low-L nature of the L2-C2 circuit requires that C2 be minimally inductive; a 10,000-pF chip capacitor is recommended. This diplexer was described by Ulrich L. Rohde and T. T. N. Bucher in *Communications Receivers*: *Principles and Design*.









determine the third, fifth and higher-odd-number intercept points. In preamplifiers, third-order intercept point is directly related to dc input power; in mixers, to the local-oscillator power applied.

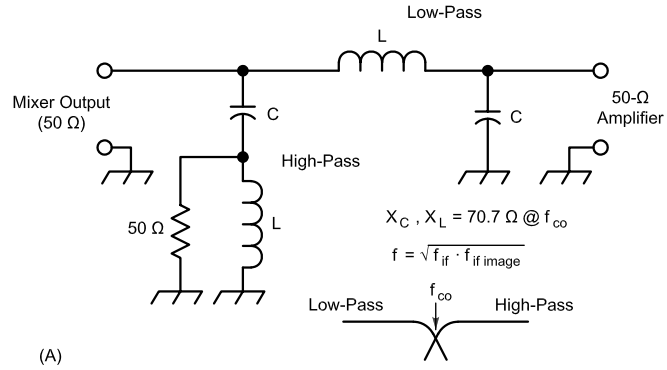
Applying Diode DBMs

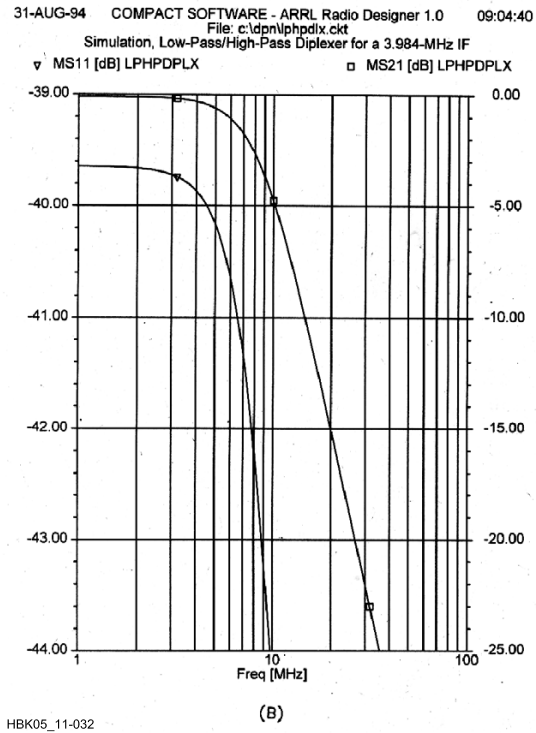
At first glance, applying a diode DBM is easy: We feed the signal(s) we want to frequency-shift (at or below the maximum level called for in the mixer's specifications, such as –10 dBm for the Mini-Circuits SBL-1 and Synergy Microwave S-1, popular Level 7 parts) to the DBM's RF port, feed the frequency-shifting signal (at the proper level) to the LO port, and extract the sum and difference products from the mixer's IF port.

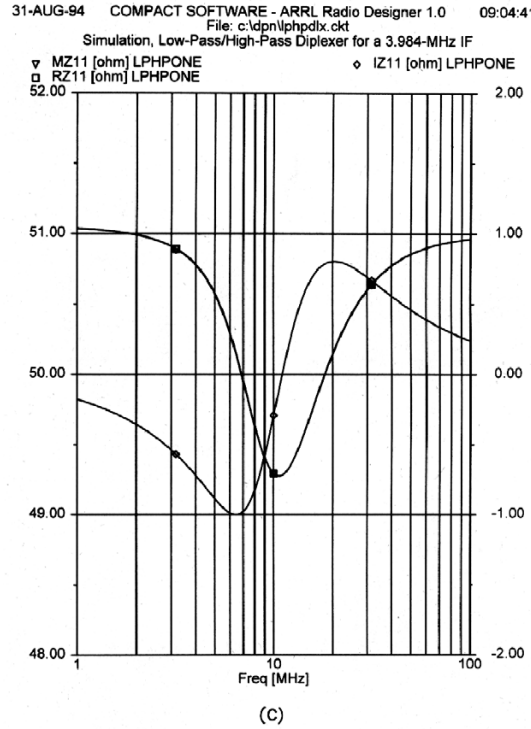
There's more to it than that, however, because diode DBMs (along with most other modern mixer types) are *termination-sensitive*. That is, their ports — particularly their IF (output) ports — must be resistively terminated with the proper impedance (commonly 50 Ω, resistive). A wideband, resistive output termination is particularly critical if a mixer is to achieve its maximum dynamic range in receiving applications. Such a load can be achieved by

- terminating the mixer in a 50-Ω resistor or attenuator pad (a technique usually avoided in receiving applications because it directly degrades system noise figure);
- terminating the mixer with a low-noise, high-dynamic-range *post-mixer amplifier* designed to exhibit a wideband resistive input impedance; or
- terminating the mixer in a *diplexer*, a frequency-sensitive signal splitter that appears as a two-terminal resistive load at its input while resistively dissipating unwanted outputs and passing desired outputs through to subsequent circuitry.

Termination-insensitive mixers are available, but this label can be misleading. Some termination-insensitive mixers are nothing more than a termination-sensitive mixer packaged with an integral post-mixer amplifier. True termination-insensitive mixers are less common and considerably more elaborate. Amateur builders will more likely use one of the







HBK05_11-032

Fig 11.32 — All of the inductors and capacitors in this high-pass/low-pass diplexer (A) exhibit a reactance of 70.7 Ω at its tuned circuits' 3-dB cutoff frequency (the geometric mean of the IF and IF image). B and C show *ARRL Radio Designer* simulations of this circuit configured for use in a receiver that converts 7 MHz to 3.984 MHz using a 10.984-MHz LO. The IF image is at 17.984 MHz, giving a 3-dB cutoff frequency of 8.465 MHz. The inductor values used in the simulation were therefore 1.33 µH (Q = 200 at 25.2 MHz); the capacitors, 265 pF (Q = 1000). This drawing shows idler load and "50-Ω Amplifier" connections suitable for a receiver in which the IF image falls at a frequency *above* the desired IF. For applications in which the IF image falls below the desired IF, interchange the 50-Ω idler load resistor and the diplexer's "50-Ω Amplifier" connection so the idler load terminates the diplexer low-pass filter and the 50-Ω amplifier terminates the high-pass filter.

many excellent termination-sensitive mixers available in connection with a diplexer, post-mixer amplifier or both.

Fig 11.31 shows one diplexer implementation. In this approach, L1 and C1 form a series-tuned circuit, resonant at the desired IF, that presents low impedance between the diplexer's input and output terminals at the IF. The high-impedance parallel-tuned circuit formed by L2 and C2 also resonates the desired IF, keeping desired energy out of the diplexer's 50-Ω load resistor, R1.

The preceding example is called a *band-pass diplexer.* **Fig 11.32** shows another type: a *high-pass/low-pass diplexer* in which each inductor and capacitor has a reactance of 70.7 Ω at the 3-dB cutoff frequency. It can be used after a "difference" mixer (a mixer in which the IF is the difference between the signal frequency and LO) if the desired IF and its image frequency are far enough apart so that the image power is "dumped" into the network's 51-Ω resistor. (For a "summing" mixer — a mixer in which the IF is the sum of the desired signal and LO — interchange the 50-Ω idler load resistor and the diplexer's "50-Ω Amplifier" connection.) Richard Weinreich, KØUVU, and R. W. Carroll described this circuit in November 1968 *QST* as one of several absorptive TVI filters.

Fig 11.33 shows a BJT post-mixer amplifier design made popular by Wes Hayward, W7ZOI, and John Lawson, K5IRK. RF feedback (via the 1-kΩ resistor) and emitter degeneration (the ac-coupled 5.6-Ω emitter resistor) work together to keep the stage's input impedance low and uniformly resistive across a wide bandwidth.

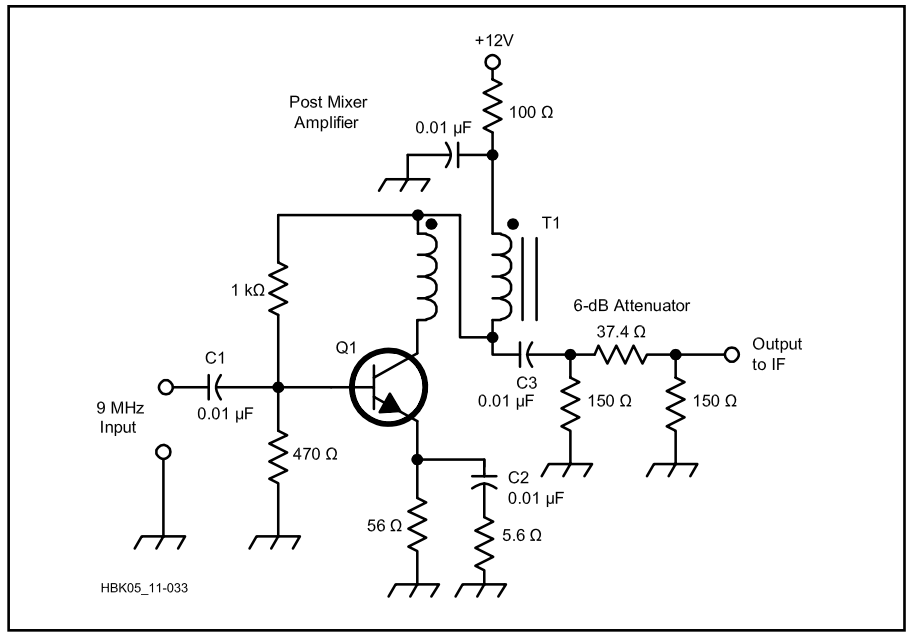


Fig 11.33 — The post-mixer amplifier from Hayward and Lawson's Progressive Communications Receiver (November 1981 *QST*). This amplifier's gain, including the 6-dB loss of the attenuator pad, is about 16 dB; its noise figure, 4 to 5 dB; its output intercept, 30 dBm. The 6-dB attenuator is essential if a crystal filter follows the amplifier; the pad isolates the amplifier from the filter's highly reactive input impedance. This circuit's input match to 50 Ω below 4 MHz can be improved by replacing 0.01-µF capacitors C1, C2 and C3 with low-inductance 0.1-µF units (chip capacitors are preferable).

Q1 — TO-39 CATV-type bipolar transistor, f_T = 1 GHz or greater. 2N3866, 2N5109, 2SC1252, 2SC1365 or MRF586 suitable. Use a small heat sink on this transistor.

T1 — Broadband ferrite transformer, ≈42 µH per winding: 10 bifilar turns of #28 enameled wire on an FT 37-43 core.

Amplitude Modulation with a DBM

We can generate DSB, suppressed-carrier AM with a DBM by feeding the carrier to its RF port and the modulating signal to the IF port. This is a classical *balanced modulator*, and the result — sidebands at radio frequencies corresponding to the carrier signal plus audio and the RF signal minus audio — emerges from the DBM's LO port. If we also want to transmit some carrier along with the sidebands, we can dc-bias the IF port (with a current of 10 to 20 mA) to upset the mixer's balance and keep its diodes from turning all the way off. (This technique is sometimes used for generating CW with a balanced modulator otherwise intended to generate DSB as part of an SSB-generation process.) **Fig 11.34** shows a more elegant approach to generating full-carrier AM with a DBM.

As we saw earlier when considering the many faces of AM, two DBMs, used in

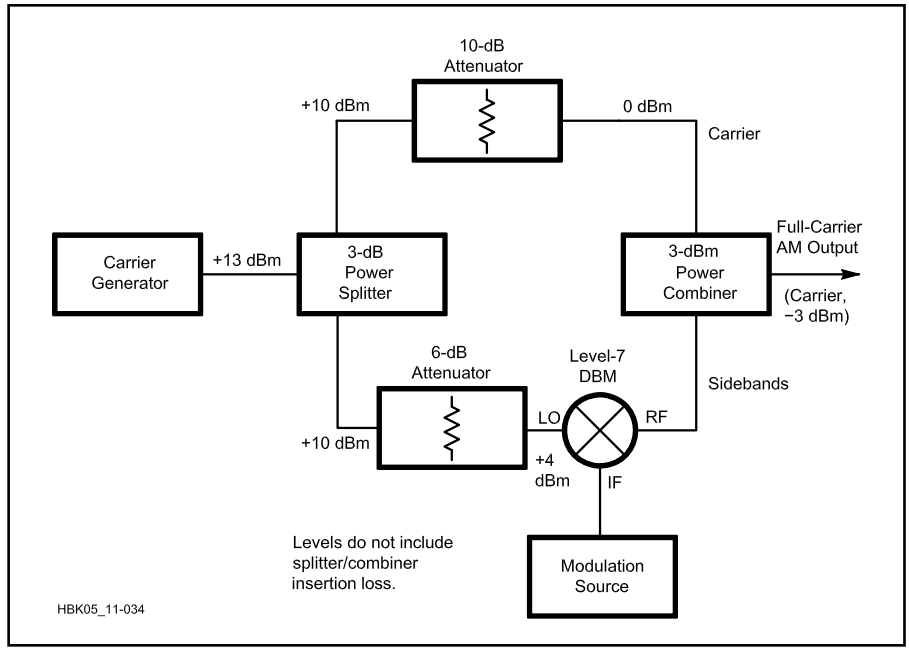


Fig 11.34 — Generating full-carrier AM with a diode DBM.

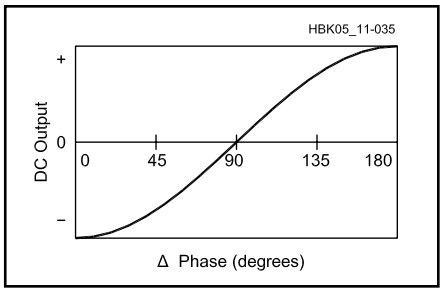


Fig 11.35 — A phase detector's dc output is the cosine of the phase difference between its input and reference signals.

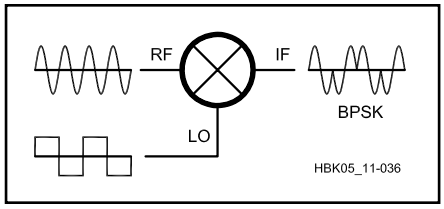


Fig 11.36 — Mixing a carrier with a square wave generates biphase-shift keying (BPSK), in which the carrier phase is shifted 180° for data transmission. In practice, as in this drawing, the carrier and data signals are phase-coherent so the mixer switches only at carrier zero crossings.

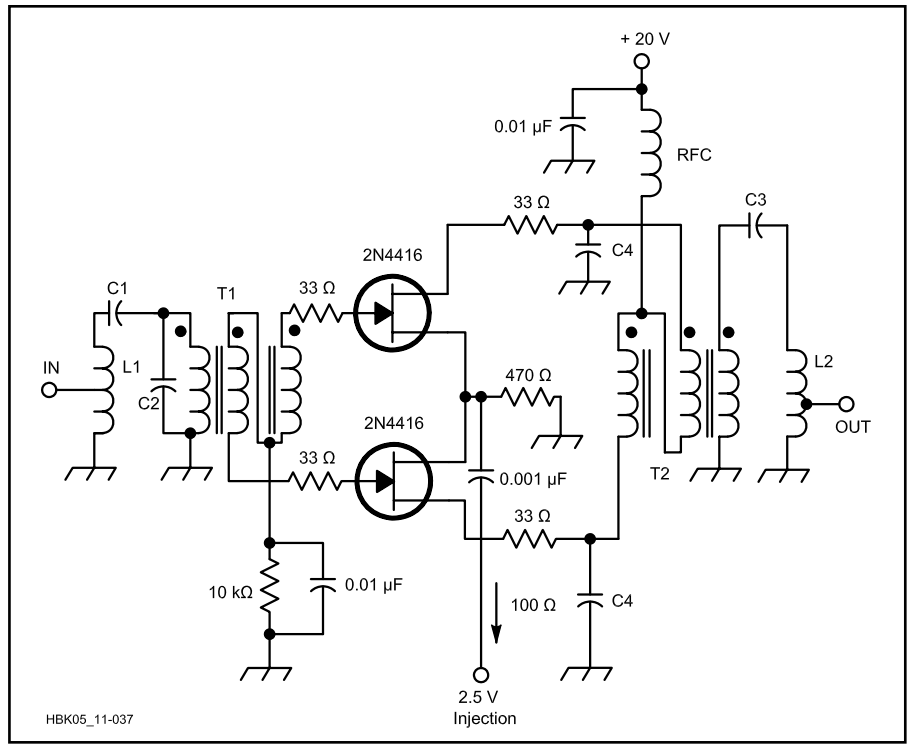


Fig 11.37 — Two 2N4416 JFETs provide high dynamic range in this mixer circuit from Sabin, *QST*, July 1970. L1, C1 and C2 form the input tuned circuit; L2, C3 and C4 tune the mixer output to the IF. The trifilar input and output transformers are broadband transmission-line types.

conjunction with carrier and audio phasing, can be used to generate SSB, suppressed-carrier AM. Likewise, two DBMs can be used with RF and LO phasing as an image-reject mixer.

Phase Detection with a DBM

As we saw in our exploration of quadrature detection, applying two signals of equal frequency to a DBM's LO and RF ports produces an IF-port dc output proportional to the cosine of the signals' phase difference (**Fig 11.35**). (This assumes that the DBM has a dc-coupled IF port, of course. If it doesn't — and some DBMs don't — phase-detector operation is out.) Any dc output offset introduces error into this process, so critical phase-detection applications use low-offset DBMs optimized for this service.

Biphase-Shift Keying Modulation with a DBM

Back in our discussion of square-wave mixing, we saw how multiplying a switching mixer's linear input with a square wave causes a 180° phase shift during the negative part of the square wave's cycle. As **Fig 11.36** shows, we can use this effect to produce *biphase-shift keying (BPSK)*, a digital system that conveys data by means of carrier phase reversals. A related system, *quadrature phase-shift keying (QPSK)* uses two DBMs and phasing to convey data by phase-shifting a carrier in 90° increments.

Transistors as Mixer Switching Elements

We've covered diode DBMs in depth because their home-buildability, high performance and suitability for direct connection into 50-Ω systems makes them attractive to Amateur Radio builders. The abundant availability of high-quality manufactured diode mixers at reasonable prices makes them excellent candidates for home construction projects. Although diode DBMs are common in telecommunications as a whole, their conversion loss and relatively high LO power requirement have usually driven the manufacturers of high-performance MF/HF Amateur Radio receivers and transceivers to other solutions. Those solutions have generally involved single- or double-balanced FET mixers — MOSFETs in the late 1970s and early 1980s, JFETs from the early 1980s to date. Many of the JFET designs are variations of a single-balanced mixer circuit introduced to *QST* readers in 1970!

Fig 11.37 shows the circuit as it was presented by William Sabin in "The Solid-State Receiver," *QST*, July 1970. Two 2N4416 JFETs operate in a common-source configuration, with push-pull RF input and parallel LO drive. **Fig 11.38** shows a similar circuit as implemented in the ICOM IC-765 transceiver. In this version, the JFETs (2SK125s) operate in common-gate, with the LO applied across a 220-Ω resistor between the gates and ground.

Bipolar junction transistors can also work well in switching mixers. In June 1994 *QST*, Ulrich Rohde, KA2WEU, published a medium-frequency mixer well suited to shortwave applications and home-built projects. Shown in **Fig 11.39**, it consists of two transistors in a push-pull, single-balanced configuration. Because of the degenerative feedback introduced by the 20-Ω emitter resistors, the two transistors need not be matched. This mixer's advantage lies in its achievement of a 33-dBm-output intercept with only 17 dBm of local oscillator drive. (Typically, a diode DBM with the same IMD performance requires 25 to 27 dBm of LO drive.) Tests indicate that the upper frequency limit of this mixer lies in the

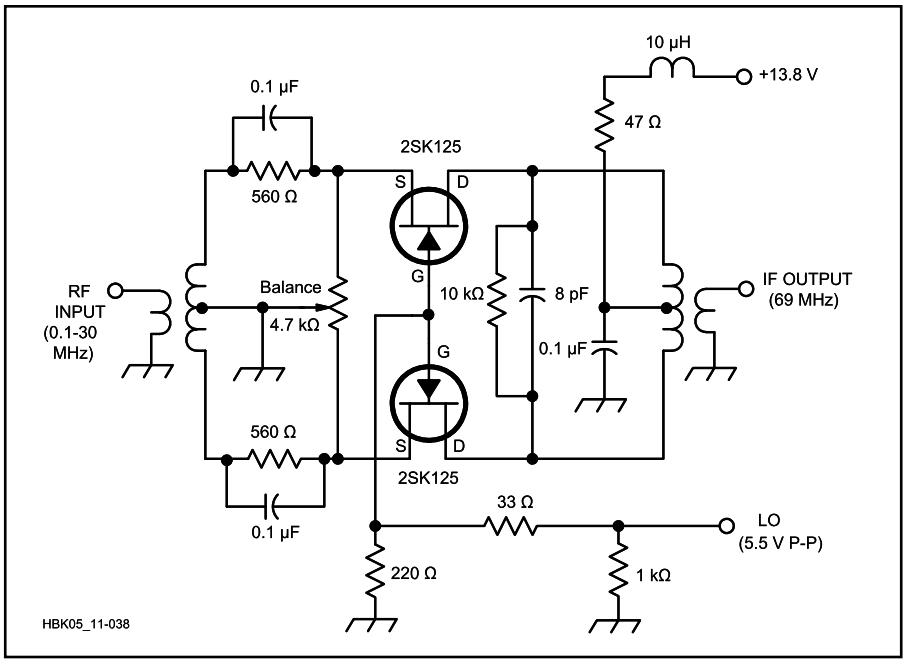


Fig 11.38 — The ICOM IC-765's single-balanced 2SK125 mixer achieves a high dynamic range (per *QST* Product Review, an IP_3 of 10.5 dBm at 14 MHz with preamp off) with arrangement very much like Sabin's. The first receive mixer in many commercial Amateur Radio transceiver designs of the 1980s and 1990s used a 2SK125 pair in much this way.

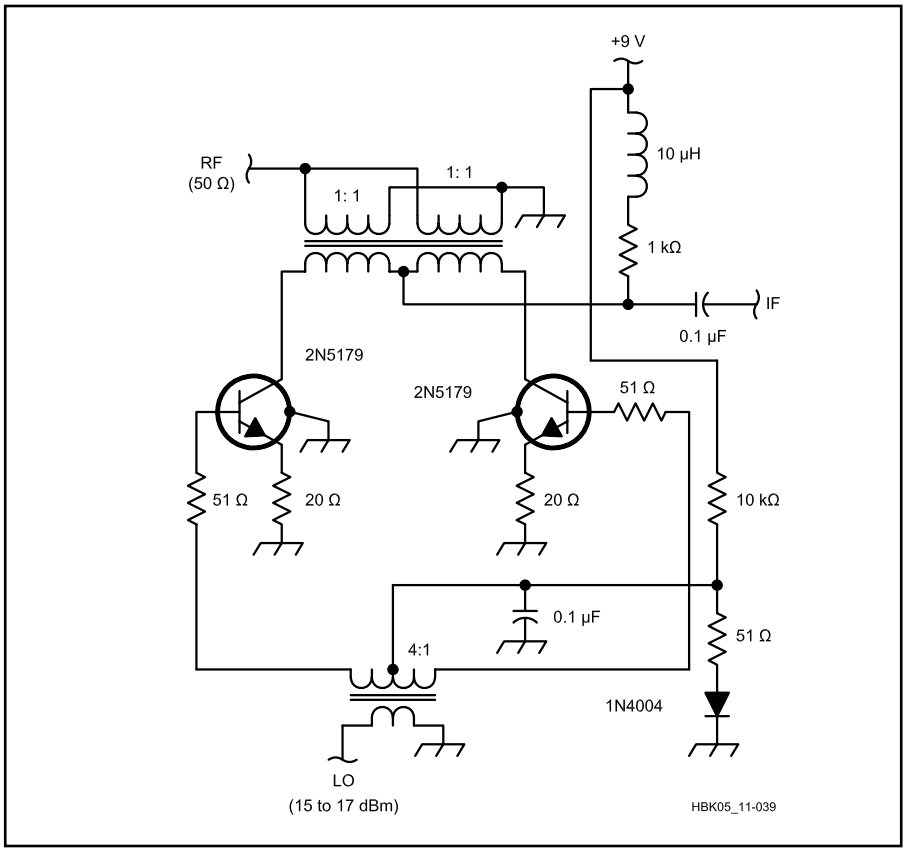


Fig 11.39 — This single-balanced, push-pull BJT mixer achieves a high dynamic range (IP_3 ≈33 dBm) with 15 to 17 dBm of LO drive. Its insertion loss is approximately 6 dB. A diode-ring mixer would require 25 to 27 dBm of LO drive to achieve the same IP_3 .

500-MHz region. The circuit's lower frequency limit depends on the transformer inductances and the ferrites used for the transformer cores.

Examining the state of the art we find that the best receive-mixer dynamic ranges are achieved with quads of RF MOSFETs operating as passive switches, with no drain voltage applied. The best of these techniques involves following a receiver's first mixer with a diplexer and low-loss roofing crystal filter, rather than terminating the mixer in a strong wideband amplifier.

High-Performance Mixer Experiments

Colin Horrabin's (G3SBI) experimentations with variations of an original high-performance mixer circuit by Jacob Mahkinson, N6NWP, led to the development of a new mixer configuration, called an H-mode mixer. This name comes from the signal path through the circuit. (See **Fig 11.40A**.) Horrabin is a professional scientist/engineer at the Science and Engineering Research Council's Darebury Laboratory, which has supported his investigative work on the H-mode switched-FET mixer, and consequently holds intellectual title to the new mixer. This does not prevent readers from taking the development further or using the information presented here.

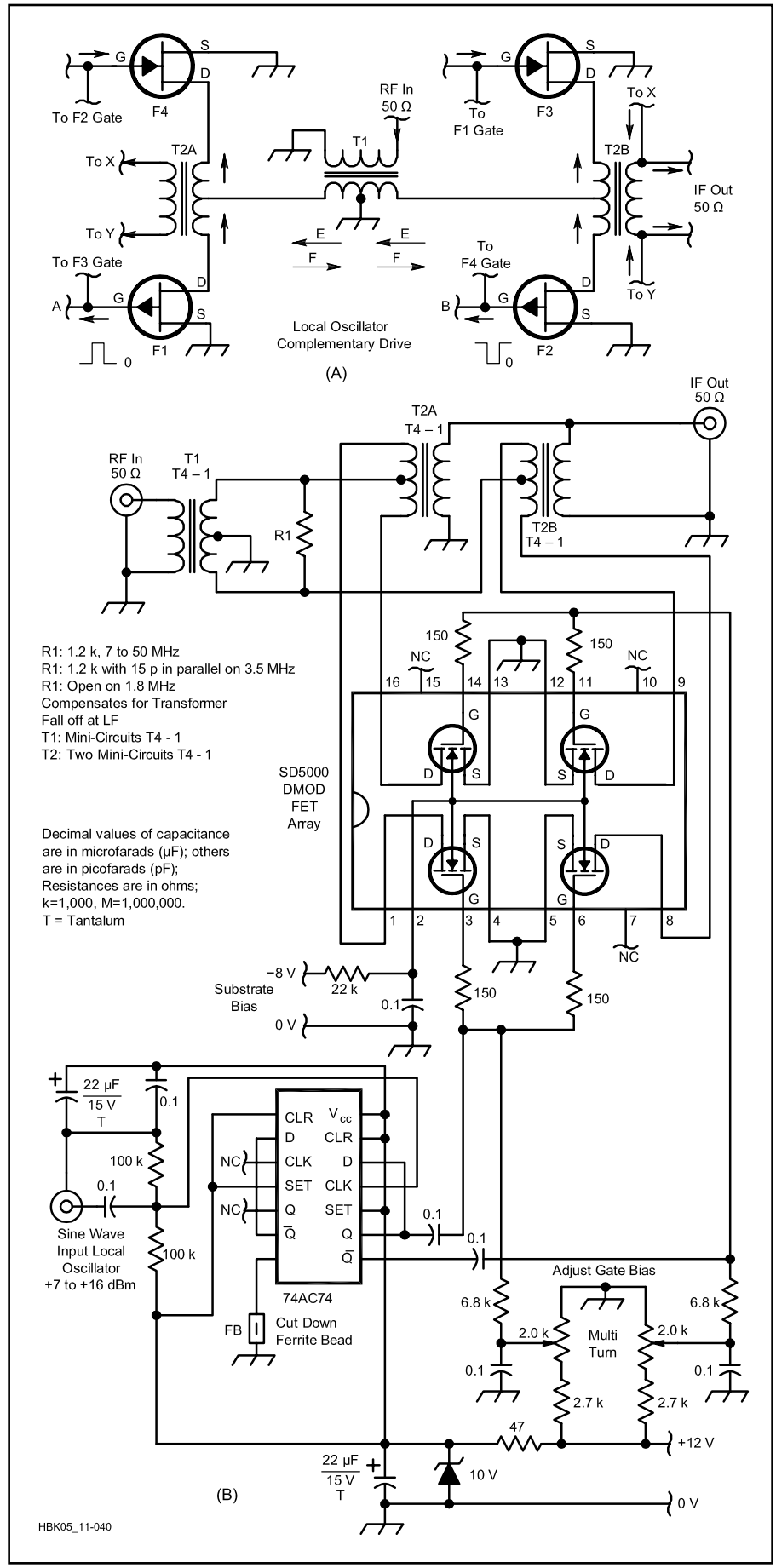
Inputs A and B are complementary square-wave inputs derived from the sine-wave local oscillator at twice the required square-wave frequency. If A is ON, then FETs F1 and F3 are ON and F2 and F4 are OFF. The direction of the RF signal across T1 is given by the E arrows. When B is ON, FETs F2 and F4 are ON and F1 and F3 are OFF. The direction of the RF signal across T1 reverses, as shown by the F arrows.

This is still the action of a switching mixer, but now the source terminal of each FET switch is grounded, so that the RF signal switched by the FET cannot modulate the gate voltage. In this configuration the transformers are important: T1 is a Mini-Circuits type T4-1 and T2 is a pair of these same transformers with their primaries connected in parallel.

The Ubiquitous NE602: A Popular Gilbert Cell Mixer

Introduced as the NE602 in the mid-1980s, the Philips Components-Signetics NE602A mixer-oscillator IC has become greatly popular with amateur experimenters for transmit mixers, receive mixers and balanced modulators. **Fig 11.41** shows its equivalent circuit. The NE602A's typical current drain is 2.4 mA; its supply voltage

Fig 11.40 — A shows the operation of the H mode switched mixer developed by Colin Horrabin, G3SBI. B shows the actual mixer circuit implemented with the Siliconix SD5000 DMOS FET quad switch IC and a 74AC74 flip-flop.



range is 4.5 to 8.0.

As we learned in exploring sine-wave versus square-wave mixing, the NE602's mixer is a Gilbert cell multiplier. Its inputs (RF) and outputs (IF) can be single- or double-ended (balanced) according to design requirements (**Fig 11.42**). Each input's equivalent ac impedance is approximately 1.5 kΩ in parallel with 3 pF; each output's resistance is 1.5 kΩ. The mixer can typically handle signals up to 500 MHz. At 45 MHz, its noise figure is typically 5.0 dB; its typical conversion gain, 18 dB. Considering the NE602A's low current drain, its input IP_3 (measured at 45 MHz with 60-kHz spacing) is usefully good at –15 dBm. Factoring in the mixer's conversion gain results in an equivalent output IP_3 of about 5 dBm.

The NE602A's on-board oscillator can operate up to 200 MHz in LC and crystal-controlled configurations (**Fig 11.43** shows three possibilities). Alternatively, energy from an external LO can be applied to the chip's pin 6 via a dc blocking capacitor. At least 200 mV P-P of external LO drive is required for proper mixer operation.

NE602A Usage Notes

The '602 was intended to be used as the second mixer in double-conversion FM cellular radios, in which the first IF is typically 45 MHz, and the second IF is typically 455 kHz. Such a receiver's second mixer can be relatively weak in terms of dynamic range because of the adjacent-signal protection afforded by the high selectivity of the first-IF filter preceding it. When used as a first mixer, the '602 can provide a two-tone third-order dynamic range between 80 and 90 dB, but this figure is greatly diminished if a pre-amplifier is used ahead of the '602 to improve the system's noise figure.

When the '602 is used as a second mixer, the sum of the gains preceding it should not exceed about 10 dB. NE602 product detection therefore should not follow a high-gain IF section unless appropriate attenuation is inserted between the '602 and the IF strip.

The '602 is generally *not* a good choice for VHF and higher-frequency mixers because of its input noise and diminishing

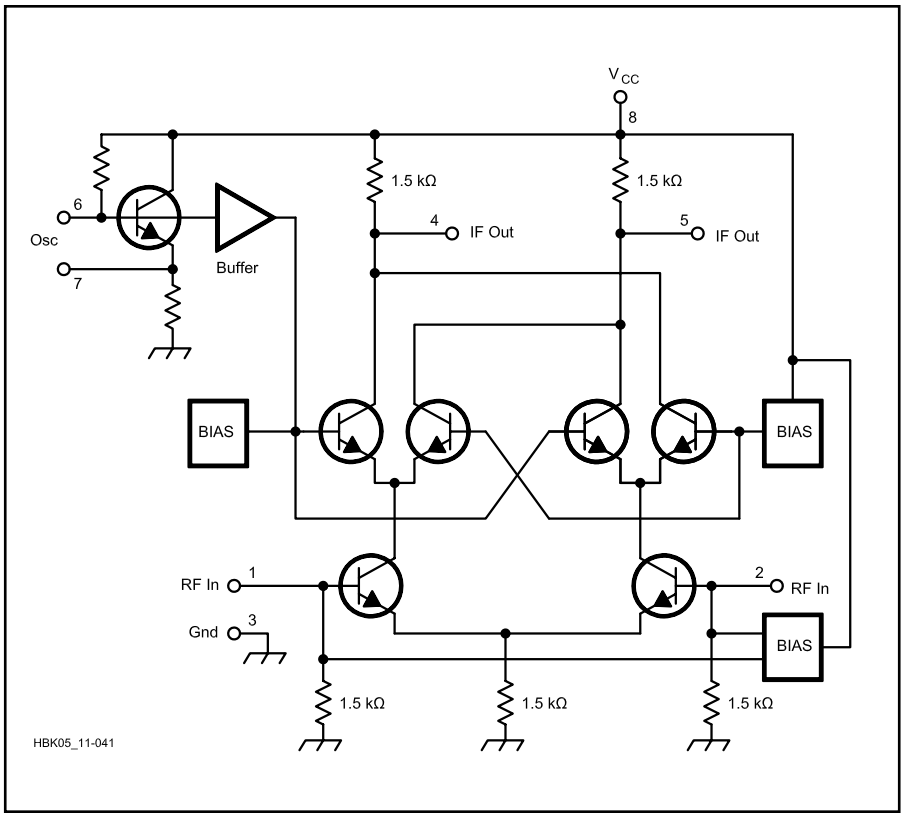


Fig 11.41 — The NE602A's equivalent circuit reveals its Gilbert-cell heritage.

IMD performance at high frequencies. There are applications, however, where 6-dB noise figures and 60- to 70-dB dynamic range performance is adequate. If your target specifications exceed these numbers, you should consider other mixers at VHF and up.

NE602A Relatives

The NE602A (SA602A for operation over a wider temperature range) began life as the NE602/SA602, a part with a slightly lower IP_3 than the A version. The pinout-identical NE612A/SA612A costs less as a result of wider tolerances. All of these parts should nonetheless work satisfactorily in most "NE602" experimenter projects. The same mixer/oscillator topology, modified for slightly higher dynamic range at the expense of somewhat less mixer gain, is also available in the Philips Components-Signetics mixer/oscillator/FM IF chips NE/SA605 (input IP_3, typically –10 dBm) and NE/SA615 (input IP_3, typically –13 dBm).

REFERENCES

J. Dillon, "The Neophyte Receiver," *QST*, Feb 1988, pp 14-18. Describes a VFO-tuned, NE602-based direct-conversion

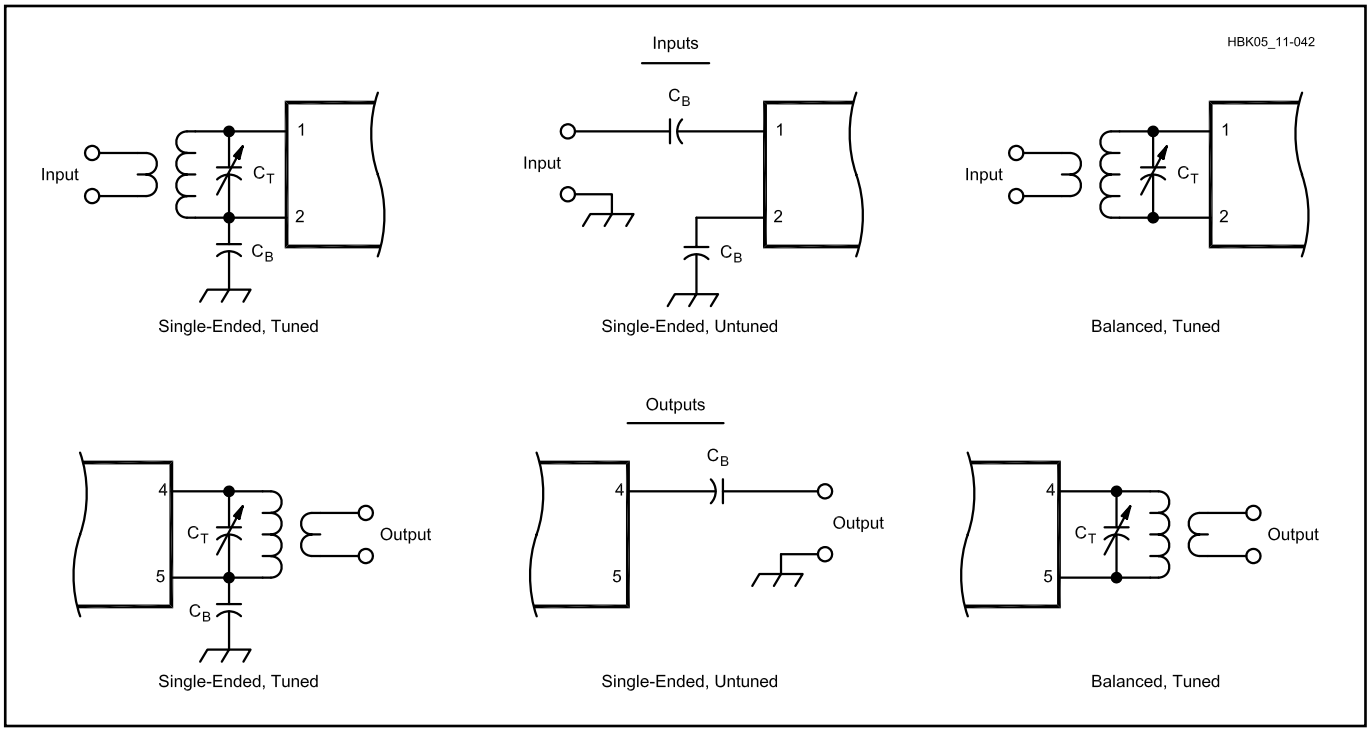


Fig 11.42 — The NE602A's inputs and outputs can be single- or double-ended (balanced). The balanced configurations minimize second-order IMD and harmonic distortion, and unwanted envelope detection in direct-conversion service. C_T tunes its inductor to resonance; C_B is a bypass or dc-blocking capacitor. The arrangements pictured don't show all the possible input/output configurations; for instance, you can also use a center-tapped broadband transformer to achieve a balanced, untuned input or output.

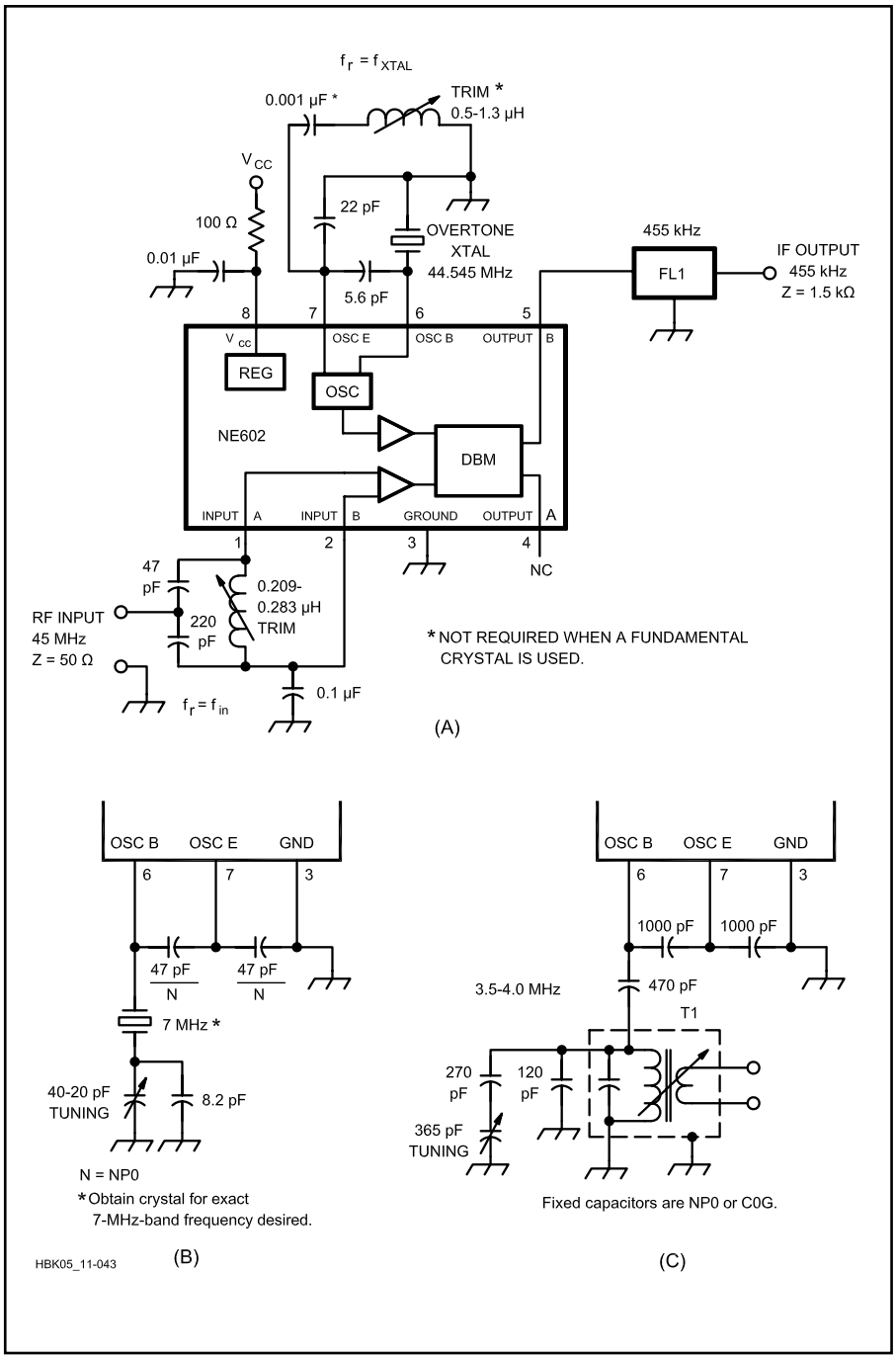


Fig 11.43 — Three NE602A oscillator configurations: crystal overtone (A); crystal fundamental (B); and LC-controlled (C). T1 in C is a Mouser 10.7-MHz IF transformer, green core, 7:1 turns ratio, part no. 42IF123.

receiver for 80 and 40 m.

B. Gilbert, "Demystifying the Mixer," self-published monograph, 1994.

P. Hawker, ed, "Super-Linear HF Receiver Front Ends," Technical Topics, *Radio Communication,* Sep 1993, pp 54-56.

W. Hayward, *Introduction to Radio Frequency Design* (Newington, CT: ARRL, 1994).

P. Horowitz and W. Hill, *The Art of Electronics*, 2nd ed (New York: Cambridge University Press, 1989).

S. Joshi, "Taking the Mystery Out of Double-Balanced Mixers," *QST*, Dec 1993, pp 32-36.

D. Kazdan "What's a Mixer?" *QST*, Aug 1992, pp 39-42.

J. Makhinson, "High Dynamic Range MF/HF Receiver Front End," *QST*, Feb 1993, pp 23-28. Also see Feedback, *QST*, Jun 1993, p 73.

RF/IF Designer's Handbook (Brooklyn, NY: Scientific Components, 1992).

RF Communications Handbook (Philips Components-Signetics, 1992).

L. Richey, "W1AW at the Flick of a Switch," New Ham Horizons, *QST*, Feb 1993, pp 56-57. Describes a crystal-controlled NE602 receiver buildable for 80, 40 and 20 m.

U. Rohde and T. Bucher, *Communications Receivers: Principles and Design* (New York: McGraw-Hill Book Co, 1988).

U. Rohde, "Key Components of Modern Receiver Design," Part 1, *QST*, May 1994, pp 29-32; Part 2, *QST*, Jun 1994, pp 27-31; Part 3, *QST*, Jul 1994, 42-45.

U. Rohde, "Testing and Calculating Intermodulation Distortion in Receivers," *QEX*, Jul 1994, pp 3-4.

W. Sabin, "The Solid-State Receiver," *QST*, Jul 1970, pp 35-43.

Synergy Microwave Corporation product handbook (Paterson, NJ: Synergy Microwave Corp, 1992).

R. Weinreich and R. W. Carroll, "Absorptive Filter for TV Harmonics," *QST,* Nov 1968, pp 20-25.

R. Zavrel, "Using the NE602," Technical Correspondence, *QST*, May 1990, pp 38-39.