13.15 Amplitude modulation

Let's begin with the simplest form of modulation (AM), taking a look at its frequency spectrum and methods of detection. Imagine a simple carrier, $\cos\omega_c t$, varied in amplitude by a modulating signal of much lower frequency, $\cos\omega_m t$, in the following way:

$$\text{signal} = (1 + m\cos\omega_m t)\cos\omega_c t$$

with m, the "modulation index," less than or equal to 1. Expanding the product, you get

$$\text{signal} = \cos\omega_c t + \tfrac{1}{2}m\cos(\omega_c + \omega_m)t$$
$$+ \tfrac{1}{2}m\cos(\omega_c - \omega_m)t$$

i.e., the modulated carrier has power at frequency ω_c and at frequencies on either side ω_m away. Figure 13.37 shows the signal and its spectrum. In this case the modulation (m) is 50%, and the two "sidebands" each contain $\frac{1}{16}$ of the power contained in the carrier.

If the modulating signal is some complex waveform [$f(t)$], like speech, the amplitude-modulated waveform is given by

$$\text{signal} = [A + f(t)]\cos\omega_c t$$

with the constant A large enough so that $A + f(t)$ is never negative. The resulting spectrum simply appears as symmetrical sidebands around the carrier (Fig. 13.38).

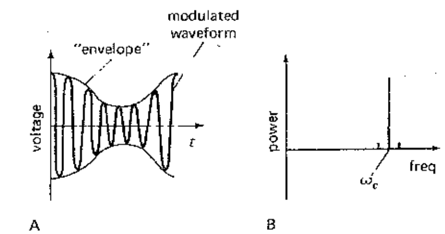

Figure 13.37. Amplitude modulation.

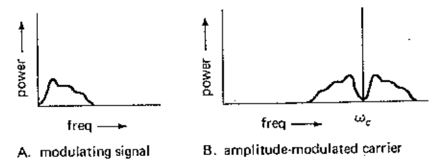

Figure 13.38. AM spectrum for a band of modulating frequencies (speech).

AM generation and detection

It is easy to generate amplitude-modulated RF. Any technique that lets you control the signal amplitude with a voltage in a linear manner will do. Common methods involve varying the RF amplifier supply voltage (if the modulation is done at the output stage) or using a multiplier chip such as the 1496. When the modulation is done at a low-level stage, all following stages of amplification must be linear. Note that in AM the modulating waveform must be biased up so that it never assumes negative values. Look at the graphs in Figure 13.39.

The simplest receiver of AM consists of several stages of tuned RF amplification, followed by a diode detector (Fig. 13.40). The amplifier stages provide selectivity against signals nearby in frequency, and they amplify the input signals (which may be at the microvolt level) for the detector. The latter simply rectifies the RF waveform, then recovers the smooth "envelope" with low-pass filtering. The

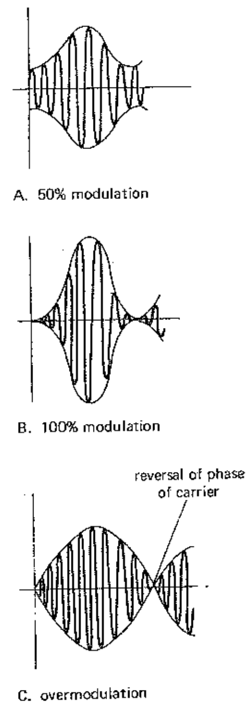

Figure 13.39

low-pass filter should reject RF while passing the audiofrequencies unattenuated. This simple scheme leaves much to be desired, as you will see. It is really just a glorified crystal set.

13.16 Superheterodyne receiver

A receiver consisting of a set of tuned RF amplifiers is undesirable for several reasons. First of all, the individual amplifiers must be tuned to the same frequency, requiring either great coordination by someone with a lot of hands or extremely good tracking of a set of simultaneously tuned LC circuits. Second, since the overall frequency selectivity is determined by the combined responses of the individual amplifiers, the shape of the passband will depend on the accuracy with which the individual amplifiers are tuned; the individual amplifiers cannot have as sharp a response as would be desirable, since tuning would then be practically impossible. And since the signal being received can be at any frequency within the tuning range of the amplifiers, it isn't possible to take advantage of crystal-lattice filters to generate a flat passband with steep falloff on either side (steep "skirts"), a very desirable passband characteristic.

A nice solution to these problems is the superheterodyne ("superhet") receiver shown in Figure 13.41. The incoming signal is amplified with a single stage of tuned RF amplification, then mixed with an adjustable local oscillator (LO) to produce a signal at a fixed intermediate frequency (IF), in this case 455kHz. From then on the receiver consists of a set of fixed-tuned IF amplifiers, including selective elements such as crystal or mechanical filters, finally terminating in a detector and audio amplifier. Changing the LO frequency tunes the receiver, since a different input frequency then gets mixed to the IF passband frequency. The input RF amplifier must be gang-tuned with the LO, but the alignment is not critical. Its purposes are (a) to

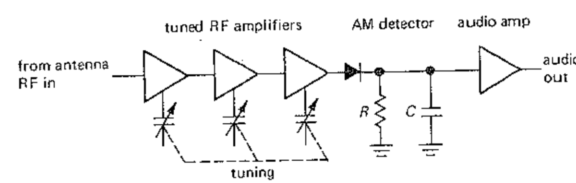

Figure 13.40

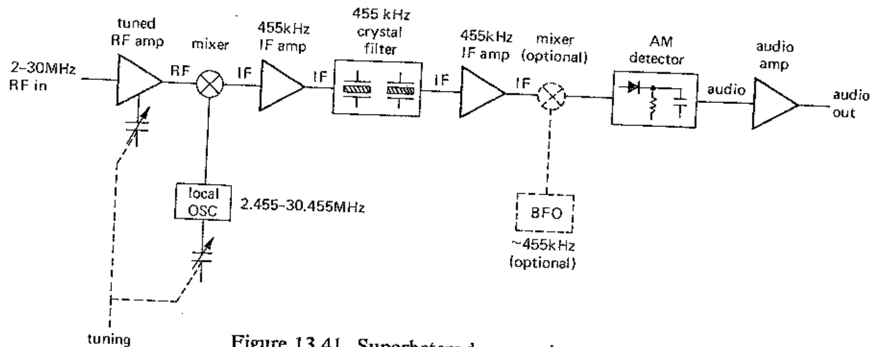

Figure 13.41. Superheterodyne receiver.

improve the sensitivity with a stage of low amplification prior to mixing and (b) to reject signals at the "image" frequency, in this case input signals at a frequency of 455kHz *above* the LO (remember that a mixer generates sum and difference frequencies). In other words, the superheterodyne receiver uses a mixer and local oscillator to shift a signal at the (variable) input frequency over to a fixed intermediate frequency where most of the gain and selectivity are concentrated.

Superhet potpourri

There are some additional features often added to a superheterodyne receiver. In this example a beat frequency oscillator (BFO) is shown; it is used in the detection of some signals with modulation other than AM (telegraphy, suppressed carrier telephony, frequency-shift keying, etc.). It can even be used for AM detection in what is known as a "homodyne" or "synchronous" detector. Receivers often have more than one mixing stage (they're called "multiple-conversion" receivers). By using a high first IF, image rejection is improved (the image is twice the IF frequency away from the actual received signal). A lower second IF makes it easier to use sharp-cutoff crystal filters, and a third IF is sometimes generated to allow the use of audio-type notch filters, low-frequency ceramic or mechanical filters, and "product detectors."

Recently, the use of direct up-conversion (an IF higher than the input signal frequency) in a front-end balanced mixer, with crystal filters at the ~40MHz IF, followed by detection with no further mixing, has become popular. Such a single-conversion scheme offers better performance in the presence of strong interfering signals, and it has become practical with the availability of good VHF crystal-lattice filters and low-distortion wide-range balanced mixers with good noise performance.

Image-reject mixers

The superhet receiver requires a tuned RF amplifier in order to reject the image band, which is separated by twice the IF frequency from the desired in-band RF signal frequency. The RF amplifier must be selective enough to reject the image band (i.e., its response to image-frequency signals must be much less than its response to in-band signals), and it must be tuned in order to keep its bandpass a constant (IF) frequency away from the LO, as the latter is adjusted to tune the receiver.

There is another way to suppress response at the image frequency, without using a tuned RF amplifier. Look at Figure

13.42, which shows an *image-reject* mixer: You begin with a pair of mixers, driven with quadrature LOs ("quadrature" means "differing in phase by 90°"); then combine the IF output signals, once again introducing a 90° phase shift in one path. The pair of 90° phase shifts adds for one sideband, and subtracts for the other, causing cancellation of the image band. Reversing the sign of the final 90° phase shift interchanges image band and signal band. In practice you usually use "4-port quadrature hybrids" to do the phase shifting, resistively terminating the unused output in each case. If you assemble an image-reject mixer from standard broadband components, you can expect something like 20dB suppression of the image sideband, with operation over one or two octaves of frequency. It's sometimes essential to be able to move around rapidly in frequency (called "frequency agility") without having to tune a tracking RF amplifier; in that case image-reject mixers are just what you want.

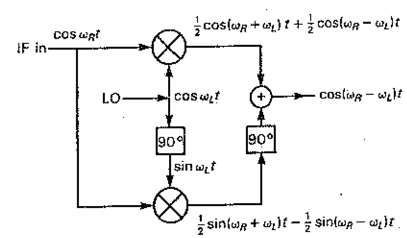

Figure 13.42. Image-reject mixer.

An interesting subtlety: As we remarked in Section 13.12, a mixer can be thought of as a modulator, and vice versa. The language you use to describe it depends on whether you are using the device to translate a low-frequency "baseband" of information up to high frequencies (in which case you call it a "modulator") or using it to translate a modulated RF band down to baseband [or perhaps an intermediate (IF) band along the way], where you demodulate it to extract the original modulating signal (in which case you call it a "mixer"). When you turn things around this way, what we called the image band becomes the other sideband. Our two methods of image rejection (RF filter, image-reject mixer) become the two classic methods of single-sideband modulation, namely the "filter" method and the "phasing" method. This may make more sense after you read the next section (but don't worry if it doesn't; we just couldn't resist trying to explain this unifying idea).

ADVANCED MODULATION METHODS

□ 13.17 Single sideband

From a glance at the spectrum of an AM signal it is obvious that things can be improved. Most of the power (67%, to be exact, at 100% modulation) is in the carrier, conveying no information. AM is at most 33% efficient, and that only when the modulation index is 100%. Since voice waveforms generally have a large ratio of peak amplitude to average amplitude, the modulation index of an AM signal carrying speech is generally considerably less than 100% (although speech-waveform "compression" can be used to get more power into the sidebands). Furthermore, the symmetrical sidebands, by conveying the identical information, cause the signal to occupy twice the bandwidth actually necessary.

With a bit of trickery it is possible to eliminate the carrier [a balanced mixer does the job; note that $\cos A \cos B = \frac{1}{2}\cos(A+B) + \frac{1}{2}\cos(A-B)$], creating what is known as "double-side-band-suppressed carrier," or DSBSC. (This is just what you will get if the audio signal multiplies the carrier directly, without first being biased so that the audio waveform is always positive, as in normal AM.) Then, either by using sharp crystal filters or by using a method known as "phasing," one of the remaining sidebands can be eliminated. The "single-sideband" (SSB) signal that remains forms a highly efficient mode of voice communication and is widely used by radio amateurs and commercial users for long-range high-frequency telephony channels. When you're not talking, there's nothing being transmitted. To receive SSB, you need a BFO and product detector, as shown in the last block diagram, to reinsert the missing carrier.

□ *Modulation spectra*

Figure 13.43 shows representative spectra of voice-modulated AM, DSBSC, and SSB. When transmitting SSB, either sideband can be used. Note that SSB consists simply of the audio spectrum translated upward in frequency by f_c. When SSB is being received, the BFO and mixer combine to translate the spectrum down to audiofrequencies again. If the BFO is slightly mistuned, all audiofrequencies will be offset by the amount of mistuning. This dictates good stability for the LO and BFO in a receiver used for single-sideband.

Note that a mixer (modulator) can always be thought of as a frequency translator, especially when combined with suitable filters to eliminate the undesired outputs: When used as a modulator, a low-frequency band of frequencies is shifted up by the carrier frequency, to form a band centered around f_c. When used as a mixer, a band of frequencies around f_c is shifted down to audiofrequencies ("baseband"), or to a band centered around the IF frequency, by the action of a high-frequency LO.

□ 13.18 Frequency modulation

Instead of modulating the amplitude of a carrier, as in AM, DSBSC, and SSB, it is possible to send information by modulating the frequency or phase of the carrier:

$$\text{signal} = \cos\left[\omega_c t + k\textstyle\int f(t)\,dt\right] \quad \text{frequency modulation (FM)}$$

$$\text{signal} = \cos\left[\omega_c t + k f(t)\right] \quad \text{phase modulation (PM)}$$

FM and PM are closely related and are sometimes referred to as "angle modulation." FM is familiar as the mode used in the 88–108MHz VHF broadcast band, and AM is used in the 0.54–1.6MHz broadcast band. Anyone who has tuned an FM receiver has probably noticed the "quieting" of background noise characteristic of FM reception. It is this property (the steep rise of recovered SNR with increasing SNR of the channel) that makes wideband FM preferable to AM for high-quality transmission.

Some facts about FM: When the frequency *deviation* $kf(t)/2\pi$ is large compared with the modulating frequency [highest frequency present in $f(t)$], you have "wideband FM" as used in FM broadcasting. The modulation index, m_f, equals

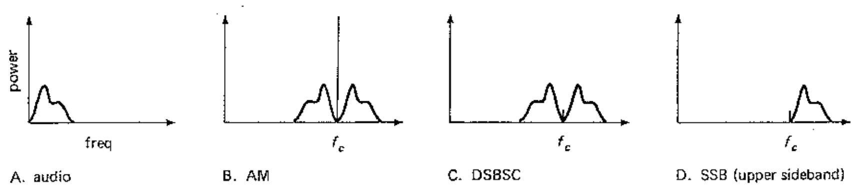

Figure 13.43. Suppressed-carrier spectra.

the ratio of frequency deviation to modulating frequency. Wideband FM is advantageous because under the right conditions the received SNR increases 6dB per doubling of FM deviation. The price you pay is increased channel bandwidth, since a wideband FM signal occupies approximately $2f_{\rm dev}$ of bandwidth, where $f_{\rm dev}$ is the peak deviation of the carrier. FM broadcasting in the 88–108MHz band uses a peak deviation $f_{\rm dev}$ of 75kHz, i.e., each station uses about 150kHz of the band. This explains why wideband FM is not used in the AM band (0.54–1.6MHz): There would be room for only six stations in any broadcasting area.

FM spectrum

A carrier that is frequency-modulated by a sine wave has a spectrum similar to that shown in Figure 13.44. There are

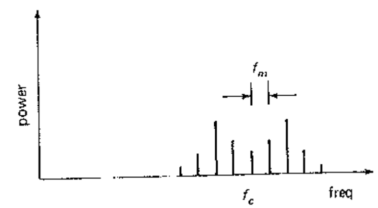

Figure 13.44. Wideband FM spectrum.

numerous sidebands spaced at multiples of the modulating frequency from the carrier, with amplitudes given by Bessel functions. The number of significant sidebands is roughly equal to the modulation index. For narrowband FM (modulation index <1), there is only one component on either side of the carrier. Superficially this looks the same as AM, but when the phase of the sidebands is taken into account, you have a waveform of constant amplitude and varying frequency (FM), rather than a waveform of varying amplitude and constant frequency (AM). With wideband FM the carrier amplitude may be very small, with correspondingly high efficiency, i.e., most of the transmitted power goes into the information-carrying sidebands.

□ *Generation and detection*

FM is easily produced by varying an element of a tuned circuit oscillator; a varactor (a diode used as a voltage-variable capacitor, Section 5.18) is ideal. Another technique involves integrating the modulating signal, then using the result to do phase modulation. In either case it is often best to modulate at low deviation, then use frequency multiplication to increase the modulation index. This works because the *rate* of frequency deviation is not changed by frequency multiplication, whereas the deviation is multiplied along with the carrier.

To detect FM, an ordinary superheterodyne receiver is used, with two differences. First, the final stage of IF amplification includes a "limiter," a stage run at constant (saturated) amplitude. Second, the subsequent detector (called a discriminator) has to convert frequency deviation into amplitude. There are several popular methods of detection:

1. A "slope detector," which is nothing more than a parallel LC circuit tuned off to one side of the IF frequency; as a result, it has a rising curve of response versus frequency across the IF bandwidth, thereby converting FM to AM. A standard envelope detector converts the AM to audio. There are improved versions of the slope detector involving a balanced pair of LC circuits tuned symmetrically to either side of the IF center frequency.

2. The Foster-Seely detector, or its variant, the "ratio detector," using a single tuned circuit in a fiendishly clever diode arrangement to give a linear curve of amplitude output versus frequency over the IF bandpass. These discriminators are superior to the simple slope detector (Fig. 13.45)

3. A "phase-locked loop" (PLL). This is a device that varies the frequency of a voltage-controlled oscillator to match an input frequency, as we discussed in Section 9.31. If the input is the IF signal, the control voltage generated by the PLL is linear in frequency, i.e., it is the audio output.

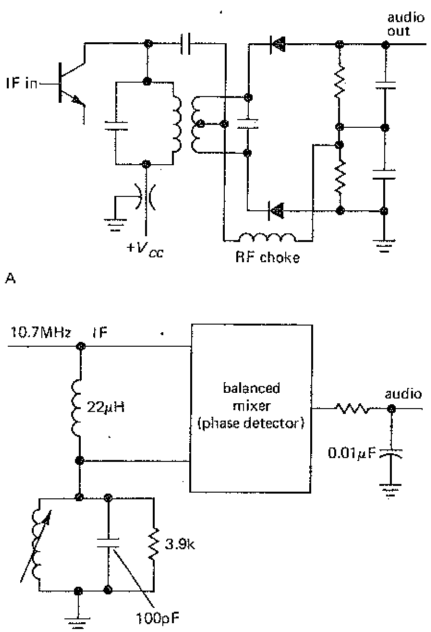

Figure 13.45. FM discriminators.
A. Foster-Seeley.
B. Balanced quadrature detector.

4. An averaging circuit, in which the IF signal is converted to a train of identical pulses at the same frequency. Averaging this pulse train generates an output proportional to IF frequency, i.e., the audio output plus some dc.

5. A "balanced quadrature detector," which is a combination of a phase detector (see Sections 9.27 and 9.31) and a phase-shifting network. The IF signal is passed through a network that produces a shift varying linearly with frequency across the IF passband (an *LC* circuit would do nicely). The resultant signal and the original signal are compared in a phase detector, giving an output that varies with relative phase. That output is the desired audio signal (Fig. 13.45).

It is often pointed out that FM provides essentially noise-free reception if the channel has sufficient SNR, as compared with AM, where the rejection of interference improves only gradually with increasing signal power. This makes sense when you remember that FM signals pass through a stage of amplitude limiting before detection. As a result, the system is relatively insensitive to interfering signals and noise, which appear as amplitude variations added to the transmitted signal.

□ 13.19 Frequency-shift keying

Transmission of digital signals (radioteletype, RTTY) is usually done by shifting a continuous-running carrier in frequency between two closely spaced frequencies according to the 1's and 0's being transmitted; 850Hz of shift is a typical value. The use of frequency-shift keying (FSK), rather than on/off modulation, is extremely effective in the presence of large signal fading from changing propagation conditions. To demodulate FSK, you simply use a differential amplifier looking at the outputs from a pair of filters set at the two detected audiofrequencies. You can think of FSK as digital FM. Narrow-shift FSK has been used to circumvent selective fading between the two signal frequencies. However, the shift cannot be reduced below the information bandwidth of the keyed signal itself, roughly the "baud" rate (number of bit cells per second), or about 100Hz for ordinary radioteletype.

□ 13.20 Pulse-modulation schemes

There are several methods whereby analog signals can be transmitted as pulses.

The basic fact that makes digital transmission of analog signals possible is expressed in the Shannon sampling theorem, which states that a band-limited waveform is fully described by sampling its amplitude at a rate equal to twice the highest frequency present. Thus, a method that conveys the amplitude of a waveform, by digital methods or whatever, at instants of time separated by $1/2f_{\max}$ can be used instead of a continuous modulation scheme. Several methods are shown in Figure 13.46.

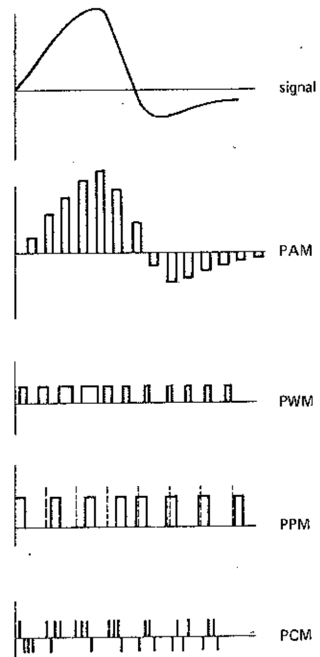

Figure 13.46. Pulse-modulation schemes.

In pulse-amplitude modulation (PAM), a train of pulses of amplitude equal to the signal is transmitted at regular intervals. This scheme is useful for *time multiplexing* of several signals on one information channel, since the time between samples can be used to transmit the samples of another signal (with an increase of bandwidth, of course). In pulse-width modulation (PWM), the width of fixed-amplitude pulses is proportional to the instantaneous signal amplitude. PWM is easy to decode, using simple averaging. In pulse-position modulation (PPM), pulses of fixed width and amplitude are delayed or advanced relative to a set of fixed times, according to the amplitude of the signal.

□ *Pulse-code modulation*

Finally, in pulse-code modulation (PCM) the instantaneous amplitude of the signal is converted to a binary number and transmitted as a serial string of bits. In the illustration, a 4-bit offset binary code corresponding to 16 levels of quantization has been used. PCM excels when error-free transmission is required over noisy channels. As long as 1's and 0's can be identified unambiguously, the correct digital code, and hence a replica of the original signal, can be recovered. PCM is particularly useful in repeater application, e.g., transcontinental telephone channels, where the signal must pass through many stations and be amplified along the way. With any of the linear modulation schemes (AM, FM, SSB) noise accumulated in transit cannot be removed, but with PCM the digital code can be correctly regenerated at each station. Thus the signal starts anew at each station.

There are variations of PCM (known as coded PCM) in which techniques other than simple serial binary sequences are used to encode the quantized samples; for instance, a burst of one of 16 tones could be used in the preceding example. PCM is routinely used for telemetry of images from space vehicles, owing to its error-free properties. It is also used for "compact-disc" digital audio, in which each stereo channel is sampled and converted to a

16-bit number 44,100 times per second. In any PCM application the bit rate must be chosen low enough to ensure a low probability of error in bit recognition. In general, this limits transmission on a given channel to speeds much below what could be used with direct analog modulation techniques.

THE ART OF ELECTRONICS

Second Edition

Paul Horowitz HARVARD UNIVERSITY

Winfield Hill ROWLAND INSTITUTE FOR SCIENCE, CAMBRIDGE, MASSACHUSETTS