Whitaker, Jerry C. "Chapter 2 – Modulation Systems and Characterisitics"
Power Vacuum Tubes Handbook 2nd Edition.
Ed. Jerry C. Whitaker
Boca Raton: CRC Press LLC, 2000

Chapter

2
Modulation Systems and Characteristics

2.1 Introduction

Radio frequency (RF) power amplifiers are used in countless applications at tens of thousands of facilities around the world. The wide variety of applications, however, stem from a few basic concepts of conveying energy and information by means of an RF signal. Furthermore, the devices used to produce RF energy have many similarities, regardless of the final application. Although communications systems represent the most obvious use of high-power RF generators, numerous other common applications exist, including:

- Induction heating and process control systems
- Radar (ground, air, and shipboard)
- Satellite communications
- Atomic science research
- Medical research, diagnosis, and treatment

The process of generating high-power RF signals has been refined over the years to an exact science. Advancements in devices and circuit design continue to be made each year, pushing ahead the barriers of efficiency and maximum operating frequency. Although different applications place unique demands on the RF design engineer, the fundamental concepts of RF amplification are applicable to virtually any system.

2.1.1 Modulation Systems

The primary purpose of most communications and signaling systems is to transfer information from one location to another. The message signals used in communication and control systems usually must be limited in frequency to provide for efficient transfer. This frequency may range from a few hertz for control systems to a few megahertz for video signals to many megahertz for multiplexed data signals. To facilitate efficient and controlled distribution of these components, an *encoder* generally is required between the source and the transmission channel. The encoder acts to *modulate* the signal, producing at its output the *modulated waveform*. Modulation is a process whereby the characteristics of a wave (the *carrier*) are varied in accordance with a message signal, the modulating waveform. Frequency translation is usually a by-product of this process. Modulation may be continuous, where the modulated wave is always present, or pulsed, where no signal is present between pulses.

There are a number of reasons for producing modulated waves, including:

- *Frequency translation*. The modulation process provides a vehicle to perform the necessary frequency translation required for distribution of information. An input signal may be translated to its assigned frequency band for transmission or radiation.
- *Signal processing*. It is often easier to amplify or process a signal in one frequency range as opposed to another.
- *Antenna efficiency*. Generally speaking, for an antenna to be efficient, it must be large compared with the signal wavelength. Frequency translation provided by modulation allows antenna gain and beamwidth to become part of the system design considerations. The use of higher frequencies permits antenna structures of reasonable size and cost.
- *Bandwidth modification*. The modulation process permits the bandwidth of the input signal to be increased or decreased as required by the application. Bandwidth reduction permits more efficient use of the spectrum, at the cost of signal fidelity. Increased bandwidth, on the other hand, provides increased immunity to transmission channel disturbances.
- *Signal multiplexing*. In a given transmission system, it may be necessary or desirable to combine several different signals into one baseband waveform for distribution. Modulation provides the vehicle for such *multiplexing*. Various modulation schemes allow separate signals to be combined at the transmission end and separated (*demultiplexed*) at the receiving end. Multiplexing may be accomplished by using, among other systems, *frequency-domain multiplexing* (FDM) or *time-domain multiplexing* (TDM).

Modulation of a signal does not come without the possible introduction of undesirable attributes. Bandwidth restriction or the addition of noise or other disturbances are the two primary problems faced by the transmission system designer.

2.1.2 Principles of Resonance

All RF generations rely on the principles of resonance for operation. Three basic systems exist:

- Series resonance circuits
- Parallel resonance circuits

• Cavity resonators

Series Resonant Circuits

When a constant voltage of varying frequency is applied to a circuit consisting of an inductance, capacitance, and resistance (all in series), the current that flows depends upon frequency in the manner shown in Figure 2.1. At low frequencies, the capacitive reactance of the circuit is large and the inductive reactance is small, so that most of the voltage drop is across the capacitor, while the current is small and leads the applied voltage by nearly 90°. At high frequencies, the inductive reactance is large and the capacitive reactance is low, resulting in a small current that lags nearly 90° behind the applied voltage; most of the voltage drop is across the inductance. Between these two extremes is the *resonant frequency*, at which the capacitive and inductive reactances are equal and, consequently, neutralize each other, leaving only the resistance of the circuit to oppose the flow of current. The current at this resonant frequency is, accordingly, equal to the applied voltage divided by the circuit resistance, and it is very large if the resistance is low.

The characteristics of a series resonant circuit depend primarily upon the ratio of inductive reactance ωL to circuit resistance R, known as the circuit Q:

$$Q = \frac{\omega L}{R} \tag{2.1}$$

The circuit Q also may be defined by:

$$Q = 2\pi \left(\frac{E_s}{E_d} \right) \tag{2.2}$$

Where:
E_s = energy stored in the circuit
E_d = energy dissipated in the circuit during one cycle

Most of the loss in a resonant circuit is the result of coil resistance; the losses in a properly constructed capacitor are usually small in comparison with those of the coil.

The general effect of different circuit resistances (different values of Q) is shown in Figure 2.1. As illustrated, when the frequency differs appreciably from the resonant frequency, the actual current is practically independent of circuit resistance and is nearly the current that would be obtained with no losses. On the other hand, the current at the resonant frequency is determined solely by the resistance. The effect of increasing the resistance of a series circuit is, accordingly, to flatten the resonance curve by reducing the current at resonance. This broadens the top of the curve, giving a more uniform current over a band of frequencies near the resonant point. This broadening is achieved, however, by reducing the selectivity of the tuned circuit.

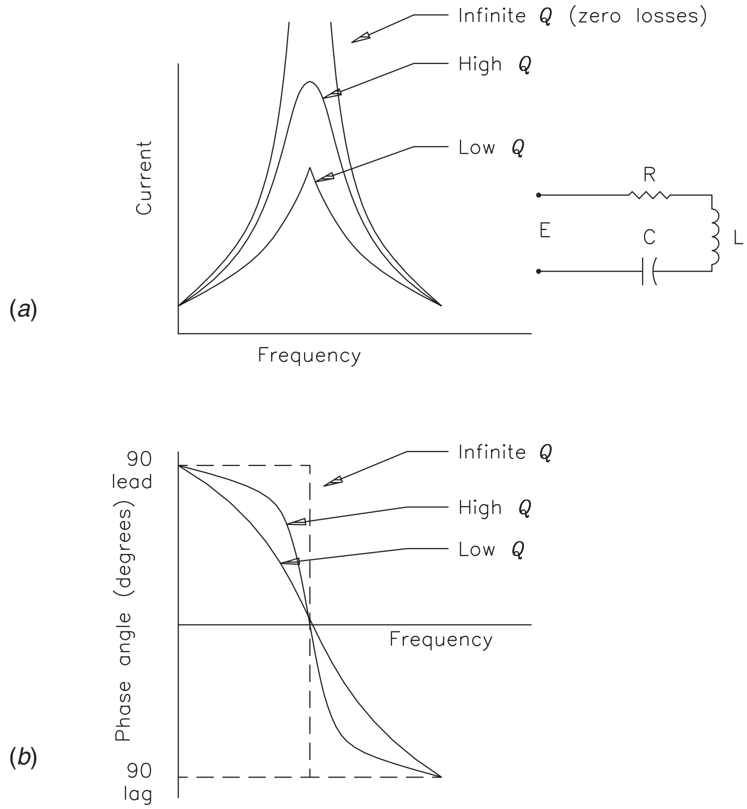


Figure 2.1 Characteristics of a series resonant circuit as a function of frequency for a constant applied voltage and different circuit *Q*s: (*a*) magnitude, (*b*) phase angle.

Parallel Resonant Circuits

A parallel circuit consisting of an inductance branch in parallel with a capacitance branch offers an impedance of the character shown in Figure 2.2. At low frequencies, the inductive branch draws a large lagging current while the leading current of the capacitive branch is small, resulting in a large lagging line current and a low lagging circuit impedance. At high frequencies, the inductance has a high reactance compared with the capacitance, resulting in a large leading line current and a corresponding low circuit impedance that is leading in phase. Between these two extremes is a frequency at which the lagging current taken by the inductive branch and the leading current entering the capacitive branch are equal. Being 180° out of phase, they neutralize, leaving only a small resultant in-phase current flowing in the line; the impedance of the parallel circuit is, therefore, high.

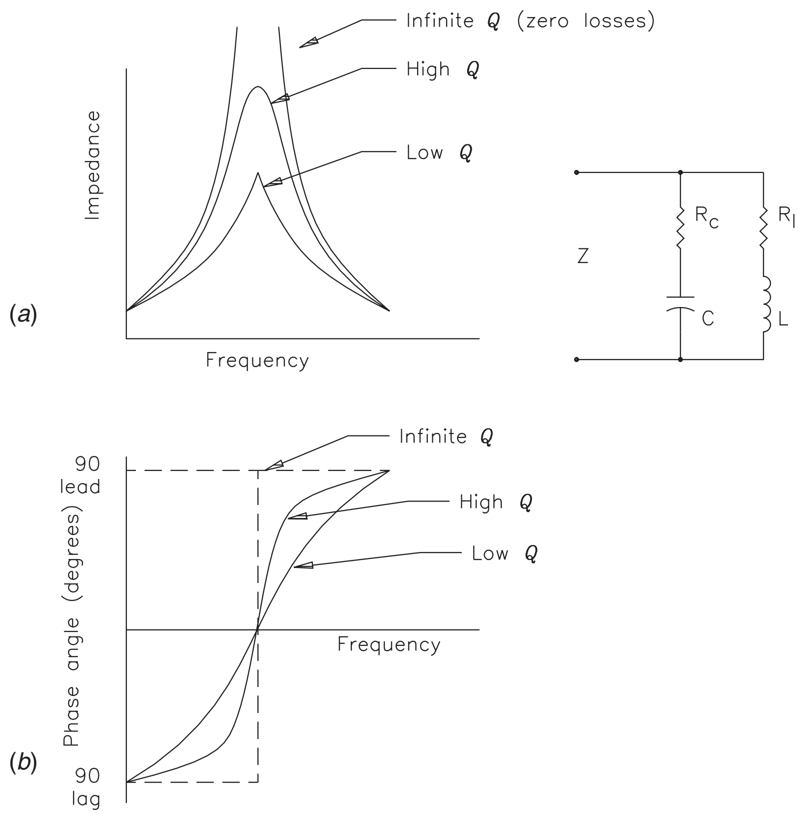


Figure 2.2 Characteristics of a parallel resonant circuit as a function of frequency for different circuit *Q*s: (*a*) magnitude, (*b*) phase angle.

The effect of circuit resistance on the impedance of the parallel circuit is similar to the influence that resistance has on the current flowing in a series resonant circuit, as is evident when Figures 2.1 and 2.2 are compared. Increasing the resistance of a parallel circuit lowers and flattens the peak of the impedance curve without appreciably altering the sides, which are relatively independent of the circuit resistance.

The resonant frequency F_o of a parallel circuit can be taken as the same frequency at which the same circuit is in series resonance:

$$F_0 = \frac{1}{2\pi\sqrt{LC}} \tag{2.3}$$

Where:
L = inductance in the circuit
C = capacitance in the circuit

When the circuit Q is large, the frequencies corresponding to the maximum impedance of the circuit and to unity power factor of this impedance coincide, for all practical purposes, with the resonant frequency defined in this way. When the circuit Q is low, however, this rule does not necessarily apply.

Cavity Resonators

Any space completely enclosed with conducting walls may contain oscillating electromagnetic fields. Such a space also exhibits certain resonant frequencies when excited by electrical oscillations. Resonators of this type, commonly termed *cavity resonators*, find extensive use as resonant circuits at very high frequencies and above. For such applications, cavity resonators have a number of advantages, including:

- Mechanical and electrical simplicity
- High Q
- Stable parameters over a wide range of operating conditions

If desired, a cavity resonator can be configured to develop high shunt impedance.

The simplest cavity resonators are sections of waveguides short-circuited at each end and $\lambda_g/2$ wavelengths long, where is the guide wavelength. This arrangement results in a resonance analogous to that of a 1/2-wavelength transmission line short-circuited at the receiving end.

Any particular cavity is resonant at a number of frequencies, corresponding to different possible field configurations that exist within the enclosure. The resonance having the longest wavelength (lowest frequency) is termed the *dominant* or *fundamental* resonance. In the case of cavities that are resonant sections of cylindrical or rectangular waveguides, most of the possible resonances correspond to various modes that exist in the corresponding waveguides.

The resonant wavelength is proportional in all cases to the size of the resonator. If the dimensions are doubled, the wavelength corresponding to resonance will likewise be doubled. This fact simplifies the construction of resonators of unusual shapes whose proper dimensions cannot be calculated easily.

The resonant frequency of a cavity can be changed through one or more of the following actions:

- Altering the mechanical dimensions of the cavity
- Coupling reactance into the cavity
- Inserting a copper paddle inside the cavity and moving it to achieve the desired resonant frequency

Small changes in mechanical dimensions can be achieved by flexing walls, but large changes typically require some type of sliding member. Reactance can be coupled into the resonator through a coupling loop, thus affecting the resonant frequency. A copper paddle placed inside the resonator affects the normal distribution of flux and tends to raise the resonant frequency by an amount determined by the orientation of the paddle. This technique is similar to the way in which copper can be used to produce small variations in the inductance of a coil.

Coupling to a cavity resonator can be achieved by means of a coupling loop or a coupling electrode. Magnetic coupling is achieved by means of a small coupling loop oriented so as to enclose magnetic flux lines existing in the desired mode of operation. A current passed through such a loop then excites oscillations of this mode. Conversely, oscillations existing in the resonator induce a voltage in such a coupling loop. The combination of a coupling loop and cavity resonator is equivalent to the ordinary coupled circuit shown in Figure 2.3. The magnitude of magnetic coupling can be readily controlled by rotating the loop. The coupling is reduced to zero when the plane of the loop is parallel to the magnetic flux.

The practical application of cavity resonators for VHF frequencies is discussed in Section 4.3; resonators for microwave devices are discussed in Section 7.3.6.

2.1.3 Frequency Source

Every RF amplifier requires a stable frequency reference. At the heart of most systems is a quartz crystal. Quartz acts as a stable high Q mechanical resonator. Crystal resonators are available for operation at frequencies ranging from 1 kHz to 300 MHz and beyond.

The operating characteristics of a crystal are determined by the *cut* of the device from a bulk "mother" crystal. The behavior of the device depends heavily on the size and shape of the crystal, and the angle of the cut. To provide for operation at a wide range of frequencies, different cuts, vibrating in one or more selected modes, are used.

The properties of a piezoelectric quartz crystal can be expressed in terms of three sets of axes. (See Figure 2.4.) The axis joining the points at the ends of the crystal is known as the *optical axis*, and electrical stresses applied in this direction exhibit no piezoelectric properties. The three axes X', X'', and X''', passing through the corners of the hexagon that forms the section perpendicular to the optical axes, are known as the *electrical axes*. The three axes Y', Y'', and Y''', which are perpendicular to the faces of the crystal, are the *mechanical axes*.

If a flat section is cut from a quartz crystal in such a way that the flat sides are perpendicular to an electrical axis, as shown in Figure 2.4 (*b*) (*X-cut*), the mechanical stresses along the Y-axis of such a section produce electric charges on the flat sides of the crystal. If the direction of these stresses is changed from tension to compression, or vice versa, the polarity of the charges on the crystal surfaces is reversed. Conversely, if electric charges are placed on the flat sides of the crystal by applying a voltage across the faces, a mechanical stress is produced in the direction of the Y-axis. This property by which mechanical and electrical properties are interconnected in a crystal is known as the *piezoelectric effect* and is exhibited by all sections cut from a piezoelectric crystal. Thus, if mechanical forces are applied across the faces of a crystal section having its flat sides perpendicular to a Y-axis, as shown in Figure 2.4 (*a*), piezoelectric charges will be produced because forces and potentials developed in such a crystal have components across at least one of the Y- or X-axes, respectively.

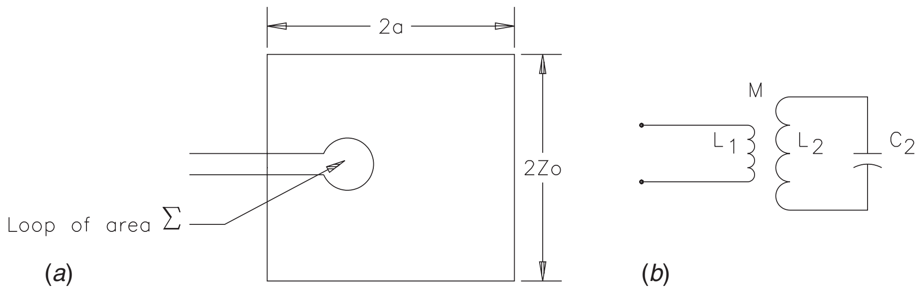


Figure 2.3 Cylindrical resonator incorporating a coupling loop: (*a*) orientation of loop with respect to cavity, (*b*) equivalent coupled circuit.

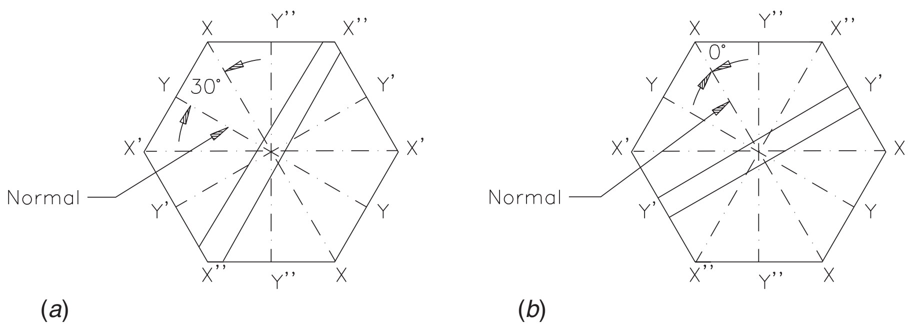


Figure 2.4 Cross section of a quartz crystal taken in the plane perpendicular to the optical axis: (*a*) *Y*-cut plate, (*b*) *X*-cut plate.

An alternating voltage applied across a quartz crystal will cause the crystal to vibrate and, if the frequency of the applied alternating voltage approximates a frequency at which mechanical resonance can exist in the crystal, the amplitude vibrations will be large. Any crystal has a number of such resonant frequencies that depend upon the crystal dimensions, the type of mechanical oscillation involved, and the orientation of the plate cut from the natural crystal.

Crystals are temperature-sensitive, as shown in Figure 2.5. The extent to which a device is affected by changes in temperature is determined by its cut and packaging. Crystals also exhibit changes in frequency with time. Such *aging* is caused by one or both of the following:

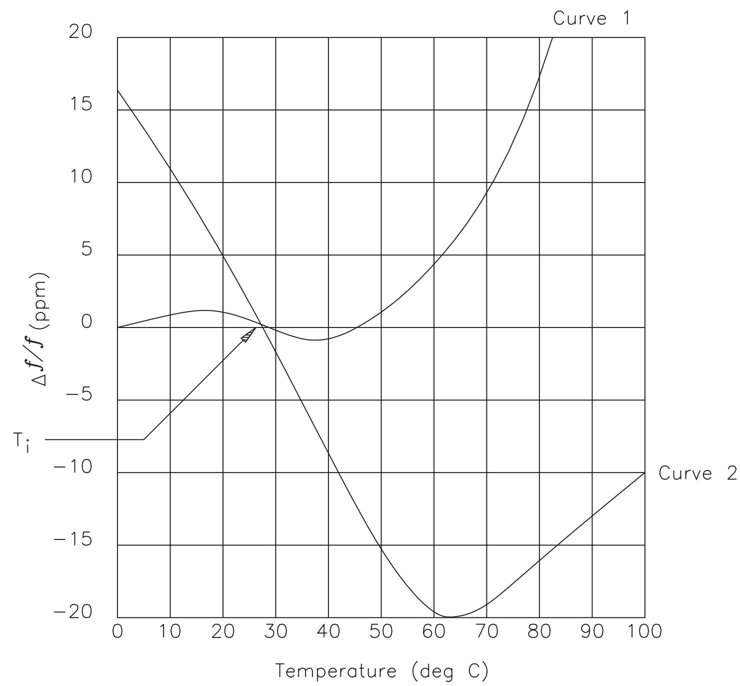


Figure 2.5 The effects of temperature on two types of AT-cut crystals.

- Mass transfer to or from the resonator surface
- Stress relief within the device itself
- Crystal aging is most pronounced when the device is new. As stress within the internal structure is relieved, the aging process slows.

Frequency Stabilization

The stability of a quartz crystal is inadequate for most commercial and industrial applications. Two common methods are used to provide the required long-term frequency stability:

- *Oven-controlled crystal oscillator*—The crystal is installed in a temperature-controlled box. Because the temperature in the box is constant, controlled by a thermostat, the crystal remains on-frequency. The temperature of the enclosure usually is set to the *turnover temperature* of the crystal. (The turnover point is illustrated in Figure 2.5.)
- *Temperature-compensated crystal oscillator* (TCXO)—The frequency vs. temperature changes of the crystal are compensated by varying a load capacitor. A thermistor network typically is used to generate a correction voltage that feeds a

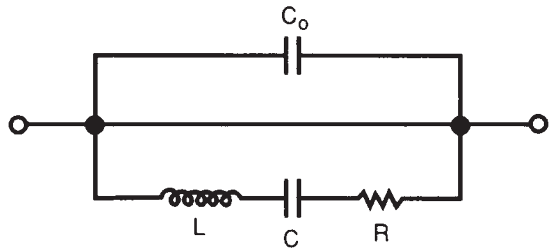


Figure 2.6 Equivalent network of a piezoelectric resonator. (*From* [1]. *Used with permission*.)

voltage-controlled capacitor (*varactor*) to retune the crystal to the desired on-frequency value.

Equivalent Circuit of a Quartz Resonator

Because of the piezoelectric effect, the quartz resonator behaves physically as a vibrating object when driven by a periodic electric signal near a resonant frequency of the cavity [1,2]. This resonant behavior may be characterized by equivalent electrical characteristics, which may then be used to accurately determine the response and performance of the electromechanical system. Dye and Butterworth (1926) [3] developed the equivalent lump electrical network that describes an electrostatic signal driving a mechanically vibrating system (an *electromechanical transducer*). The equivalent electric network that models the piezo resonator was later developed by Van Dyke (1928), Butterworth, and others, and is shown in Figure 2.6. In the equivalent network, the circuit parameters R_1, L_1, and C_1 represent parameters associated with the quartz crystal, while C_0 represents the shunt capacitance of the resonator electrodes in parallel with the container that packages the crystal.

The inductance L_1 is associated with the mass of the quartz slab, and the capacitance C_1 is associated with the stiffness of the slab. Finally, the resistance, R_1 is determined by the loss processes associated with the operation of the crystal. Each of the equivalent parameters can be measured using standard network measurement techniques. Typical parameters are shown in Table 2.1.

Application of an electric field across the electrodes of the crystal results in both face and thickness shear waves developing within the crystal resonator. For a *Y*-cut crystal, face shear modes occur in the X-Z plane, while thickness shear modes develop in the X-Y plane.

The vibration is enhanced for electric signals near a resonant frequency of the resonator. Low frequency resonances tend to occur in the face shear modes, while high frequency resonances occur in thickness shear modes. In either case, when the applied electric signal is near a resonant frequency, the energy required to maintain the vibrational motion is small.

Table 2.1 Typical Equivalent Circuit Parameter Values for Quartz Crystals (*After* [1].)

Parameter	200 kHz Fundamental	2 MHz	30 Mhz 3rd Overtone	90 Mhz 5th Overtone
R_1	2 kS	100 Ω	20 Ω	40 Ω
L_1	27 H	520 mH	11 mH	6 mH
C_1	0.024 pF	0.012 pF	0.0026 pF	0.0005 pF
C_0	9 pF	4 pF	6 pF	4 pF
Q	18×10^3	18×10^3	18×10^3	18×10^3

Temperature Compensation

Compensation for temperature-related frequency shifts can be accomplished by connecting a reactance, such as a variable voltage capacitor (a *varactor*) in series with the resonator [1]. The effect of this reactance is to pull the frequency to cancel the effect of frequency drift caused by temperature changes. An example of a frequency versus temperature characteristic both before and after compensation is shown in Figure 2.7.

Several techniques for temperature compensation are possible. They include:

- Analog temperature compensation
- Hybrid analog-digital compensation
- Digital temperature compensation

For analog compensation, a varactor is controlled by a voltage divider network that contains both thermistors and resistors to manage the compensation. Stability values are in the range 0.5 to 10 ppm. The lowest value requires that individual components in the voltage divider network be adjusted under controlled-temperature conditions. This is difficult over large temperature ranges for analog networks because of the tight component value tolerances required and because of interactions between different analog segments. Analog networks can be segmented in order to increase the independence of the component adjustments required to improve stability.

These problems can be alleviated through the use of some form of hybrid or digital compensation. In this scheme, the crystal oscillator is connected to a varactor in series, as in the analog case. This capacitor provides the coarse compensation (about ±4 parts in 10^7). The fine compensation is provided by the digital network to further improve the stability. The hybrid analog-digital method is then capable of producing stability values of ±5 parts in 10^8 over the range –40°C to +80°C.

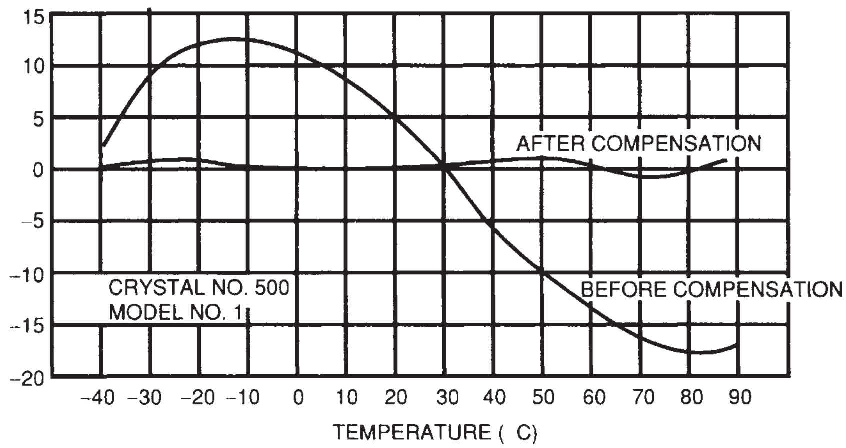


Figure 2.7 Frequency vs. temperature characteristics for a typical temperature-compensated crystal oscillator. (*After* [2].)

Stress Compensation

Stress compensation is employed in order to stabilize the crystal resonator frequency against shifts caused by mechanical stresses [1]. These result from temperature changes or from electrode changes, due to differences in the expansion coefficients, which can occur over time. The frequency variations are caused by alteration of the crystal elastic coefficients with stress bias. The method developed for stress compensation is a special crystal cut called the *stress-compensated* or SC-cut. This cut was developed to obtain a resonator whose frequency would be independent of specified stresses.

The SC-cut has some advantages over the more common AT- and BT-cuts. These include a flatter frequency-temperature response, especially at the turning points, and less susceptibility to aging because of changes in the mounting stresses. The lower sensitivity to frequency changes means that SC-cut resonators can be frequency-temperature tested more rapidly than their AT- and BT-cut counterparts. A disadvantage of the SC-cut is that it is sensitive to air-damping and must be operated in a vacuum environment to obtain optimum Q factors. The frequency-temperature curve is also roughly 70° higher for the SC-cut over the more common AT-cut, which could also be a disadvantage in some applications.

Aging Effects

Aging in crystal oscillators generally refers to any change over time that will affect the frequency characteristics of the oscillator, or the physical parameters that describe the device; such as motional time constant and equivalent circuit parameters [1]. Some factors that influence aging include the following:

- Surface deterioration
- Surface contamination
- Electrode composition
- Environmental conditions

Deterioration on the surface contributes to strain in the crystal, and this type of strain can impact the operating frequency of the oscillator. More importantly, the surface strain is sensitive to humidity and temperature early in the life of the crystal. One method of combating this surface degradation mechanism is to etch away the strained surface layers. The initial slope of the frequency change versus etch time curve is exponential. However, after the strained surface layers have been removed, the slope becomes linear.

Another surface issue is contamination from foreign materials. The presence of foreign substances will load the oscillator with an additional mass, and this can be reflected in the equivalent circuit parameters. The contaminants can be introduced during fabrication processes, from out-gassing of other materials after the device has been packaged, or from contaminants in the environment during storage. The contamination process is often increased in higher temperature environments, and temperature cycling will lead to stability problems for the device.

Aging problems associated with the device electrodes includes delamination, leading to increased resistance or corrosion effects. Typical electrode materials include gold, silver, and aluminum. Of these, silver is the most popular material.

2.1.4 Operating Class

Amplifier stage operating efficiency is a key element in the design and application of an electron tube device or system. As the power level of an RF generator increases, the overall efficiency of the system becomes more important. Increased efficiency translates into lower operating costs and, usually, improved reliability of the system. Put another way, for a given device dissipation, greater operating efficiency translates into higher power output. The operating mode of the final stage, or stages, is the primary determining element in the maximum possible efficiency of the system.

An electron amplifying device is classified by its individual *class of operation*. Four primary class divisions apply to vacuum tube devices:

- *Class A*—A mode wherein the power amplifying device is operated over its linear transfer characteristic. This mode provides the lowest waveform distortion, but also the lowest efficiency. The basic operating efficiency of a class A stage is 50 percent. Class A amplifiers exhibit low intermodulation distortion, making them well suited to linear RF amplifier applications.

- *Class B*—A mode wherein the power amplifying device is operated just outside its linear transfer characteristic. This mode provides improved efficiency at the expense of some waveform distortion. *Class AB* is a variation on class B operation. The transfer characteristic for an amplifying device operating in this mode is, predictably, between class A and class B.
- *Class C*—A mode wherein the power amplifying device is operated significantly outside its linear transfer characteristic, resulting in a pulsed output waveform. High efficiency (up to 90 percent) can be realized with class C operation, but significant distortion of the waveform will occur. Class C is used extensively as an efficient RF power generator.
- *Class D*—A mode that essentially results in a switched device state. The power amplifying device is either *on* or *off*. This is the most efficient mode of operation. It is also the mode that produces the greatest waveform distortion.

The angle of current flow determines the class of operation for a power amplifying device. Typically, the following generalizations regarding conduction angle apply:

- Class A: 360°
- Class AB: between 180 and 360°
- Class B: 180°
- Class C: less than 180°

Subscripts also may be used to denote grid current flow. The subscript "1" means that no grid current flows in the stage; the subscript "2" denotes grid current flow. Figure 2.8 charts operating efficiency as a function of conduction angle for an RF amplifier.

The class of operation is not directly related to the type of amplifying circuit. Vacuum tube stages may be grid- or cathode-driven without regard to the operating class.

Operating Efficiency

The design goal of all RF amplifiers is to convert input power into an RF signal at the greatest possible efficiency. Direct current input power that is not converted into a useful output signal is, for the most part, converted to heat. This heat is wasted energy, which must be removed from the amplifying device. Removal of heat is a problem common to all high-power RF amplifiers. Cooling methods include:

- Natural convection
- Radiation
- Forced convection
- Liquid
- Conduction
- Evaporation

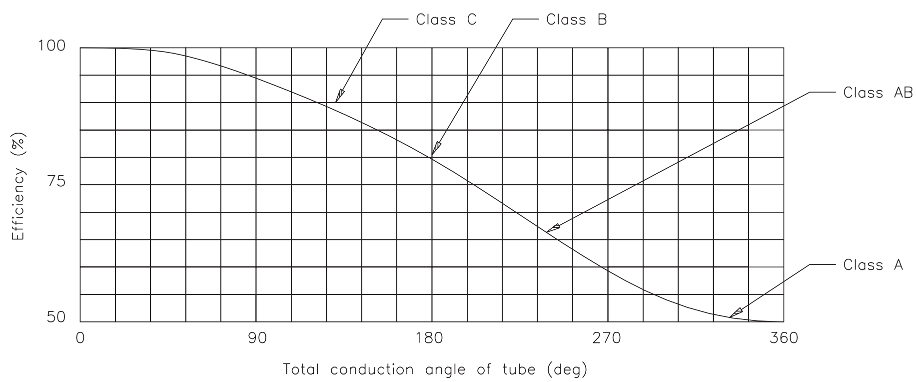


Figure 2.8 Plate efficiency as a function of conduction angle for an amplifier with a tuned load.

The type of cooling method chosen is dictated in large part by the type of active device used and the power level involved. For example, liquid cooling is used almost exclusively for high-power (100 kW) vacuum tubes; conduction is used most often for low-power (10 kW) devices. (Power-device cooling techniques are discussed in detail in Section 8.2.)

2.1.5 Broadband Amplifier Design

RF design engineers face a continuing challenge to provide adequate bandwidth for the signals to be transmitted while preserving as much efficiency from the overall system as possible. These two parameters, although not mutually exclusive, often involve tradeoffs for both designers and operators.

An ideal RF amplifier will operate over a wide band of frequencies with minimum variations in output power, phase, distortion, and efficiency. The bandwidth of the amplifier depends to a great extent on the type of active device used, the frequency range required, and the operating power. As a general rule, a bandwidth of 20 percent or greater at frequencies above 100 MHz can be considered *broadband*. Below 100 MHz, broadband amplifiers typically have a bandwidth of one octave or more.

Stagger Tuning

Several stages with narrowband response (relative to the desired system bandwidth) can be cascaded and, through the use of *stagger tuning*, made broadband. Although there is an efficiency penalty for this approach, it has been used for years in all types of equipment. The concept is simple: Offset the center operating frequencies (and, therefore, peak amplitude response) of the cascaded amplifiers so that the resulting passband is flat and broad.

For example, the first stage in a three-stage amplifier is adjusted for peak response at the center operating frequency of the system. The second stage is adjusted above the center frequency, and the third stage is adjusted below center. The resulting composite response curve yields a broadband trace. The efficiency penalty for this scheme varies depending on the power level of each stage, the amount of stagger tuning required to achieve the desired bandwidth, and the number of individual stages.

Matching Circuits

The individual stages of an RF generator must be coupled together. Rarely do the output impedance and power level of one stage precisely match the input impedance and signal-handling level of the next stage. There is a requirement, therefore, for broadband matching circuits. Matching at radio frequencies can be accomplished with several different techniques, including:

- A 1/4-wave transformer: A matching technique using simply a length of transmission line 1/4-wave long, with a characteristic impedance of $Z_{line} = \sqrt{Z_{in} \times Z_{out}}$, where Z_{in} and Z_{out} = the terminating impedances. Multiple 1/4-wave transformers can be cascaded to achieve more favorable matching characteristics. Cascaded transformers permit small matching ratios for each individual section.
- A *balun* transformer: A transmission-line transformer in which the turns are arranged physically to include the interwinding capacitance as a component of the characteristic impedance of the transmission line. This technique permits wide bandwidths to be achieved in the device without unwanted resonances.

Balun transformers usually are made of twisted wire pairs or twisted coaxial lines. Ferrite toroids may be used as the core material.

Power Combining

The two most common methods of extending the operating power of active devices are *direct paralleling* of components and *hybrid splitting/combining*. Direct paralleling has been used for tube designs, but application of this simple approach is limited by variations in device operating parameters. Two identical devices in parallel do not necessarily draw the same amount of current (supply the same amount of power to the load). Paralleling at UHF and above can be difficult because of the restrictions of operating wavelength.

The preferred approach involves the use of identical stages driven in parallel from a *hybrid coupler*. The coupler provides a constant-source impedance and directs any reflected energy from the driven stages to a *reject port* for dissipation. A hybrid coupler offers a standing-wave-canceling effect that improves system performance. Hybrids also provide a high degree of isolation between active devices in a system.

2.1.6 Thermal and Circuit Noise

All electronic circuits are affected by any number of factors that cause their performance to be degraded from the ideal assumed in simple component models, in ways that can be controlled but not eliminated entirely [4]. One limitation is the failure of the model to account properly for the real behavior of components, either the result of an oversimplified model description or variations in manufacture. Usually, by careful design, the engineer can work around the limitations of the model and produce a device or circuit whose operation is very close to predictions. One source of performance degradation that cannot be easily overcome, however, is noise.

When vacuum tubes first came into use in the early part of the twentieth century, they were extensively used in radios as signal amplifiers and for other related signal conditioning functions. Thus, the measure of performance that was of greatest importance to circuit designers was the quality of the sound produced at the radio speaker. It was immediately noticed that the sound coming from speakers not only consisted of the transmitted signal but also of popping, crackling, and hissing sounds, which seemed to have no pattern and were distinctly different from the sounds that resulted from other interfering signal sources, such as other radio stations using neighboring frequencies. This patternless or random signal was labeled "noise," and has become the standard term for signals that are random and are combined with the circuit signal to affect the overall performance of the system.

As the study of noise has progressed, engineers have come to realize that there are many sources of noise in circuits. The following definitions are commonly used in discussions of circuit noise:

- *White noise*: Noise whose energy is evenly distributed over the entire frequency spectrum, within the frequency range of interest (typically below frequencies in the infrared range). Because noise is totally random it may seem inappropriate to refer to its frequency range, because it is not really periodic in the ordinary sense. Nevertheless, by examining an oscilloscope trace of white noise, one can verify that every trace is different—as the noise never repeats itself—and yet each trace looks the same. Similarly, a TV set tuned to an unused frequency displays never-ending "snow" that always looks the same, yet clearly is always changing. There is a strong theoretical foundation to represent the frequency content of such signals as covering the frequency spectrum evenly. In this way the impact on other periodic signals can be analyzed. The term "white noise" arises from the fact that—similar to white light, which has equal amounts of all light frequencies—white noise has equal amounts of noise at all frequencies within circuit operating range.

- *Interference*: The name given to any predictable, periodic signal that occurs in an electronic circuit in addition to the signal the circuit is designed to process. This component is distinguished from a noise signal by the fact that it occupies a relatively small frequency range, and because it is predictable, it can often be filtered out. Interference usually results from another electronic system, such as an interfering radio source.

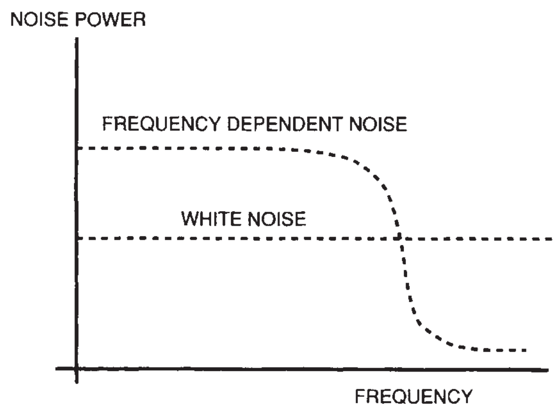


Figure 2.9 A plot of signal power vs frequency for white noise and frequency limited noise. (*From* [4]. *Used with permission*.)

- *Thermal noise*: Any temperature-dependent noise generated within a circuit. This signal usually is the result of the influence of temperature directly on the operating characteristics of circuit elements, which—because of the random motion of molecules as a result of temperature—in turn creates a random fluctuation of the signal being processed.
- *Shot noise*: A type of circuit noise that is not temperature-dependent, and is not white noise in the sense that it tends to diminish at higher frequencies. (See Figure 2.9.) This noise usually occurs in components whose operation depends on a mean particle residence time for the active electrons within the device. The cutoff frequency above which noise disappears is closely related to the inverse of this characteristic particle residence time. It is called "shot noise" because in a radio it can make a sound similar to buckshot hitting a drum, as opposed to white noise, which tends to sound more like hissing (because of the higher frequency content).

Thermal Noise

By definition, thermal noise is internally generated noise that is temperature-dependent [4]. While first observed in vacuum tube devices (because their amplifying capabilities tend to bring out thermal noise), it is also observed in semiconductor devices. It is a phenomenon resulting from the ohmic resistance of devices that dissipate the energy lost in them as heat. Heat consists of random motion of molecules, which are more energetic as temperature increases. Because the motion is random, it is to be expected that as electrons pass through the ohmic device incurring resistance, there should be some random deviations in the rate of energy loss. This fluctuation has the effect of causing variation in the resulting current, which is noise. As the temperature of the device increases, the random motion of the molecules increases, and so does the corresponding noise level.

Noise in Systems of Cascaded Stages

The noise level produced by thermal noise sources is not necessarily large, however because source signal power may also be low, it is usually necessary to amplify the source signal. Because noise is combined with the source signal, and both are then amplified, with more noise added at each successive stage of amplification, noise can become a noticeable phenomenon [4].

2.2 Amplitude Modulation

In the simplest form of amplitude modulation, an analog carrier is controlled by an analog modulating signal. The desired result is an RF waveform whose amplitude is varied by the magnitude of the applied modulating signal and at a rate equal to the frequency of the applied signal. The resulting waveform consists of a carrier wave plus two additional signals:

- An upper-sideband signal, which is equal in frequency to the carrier *plus* the frequency of the modulating signal
- A lower-sideband signal, which is equal in frequency to the carrier *minus* the frequency of the modulating signal

This type of modulation system is referred to as *double-sideband amplitude modulation* (DSAM).

The radio carrier wave signal onto which the analog amplitude variations are to be impressed is expressed as:

$$e(t) = A\,E_c\,\cos(\omega_c t) \tag{2.4}$$

Where:
$e(t)$ = instantaneous amplitude of carrier wave as a function of time (t)
A = a factor of amplitude modulation of the carrier wave
ω_c = angular frequency of carrier wave (radians per second)
E_c = peak amplitude of carrier wave

If A is a constant, the peak amplitude of the carrier wave is constant, and no modulation exists. Periodic modulation of the carrier wave results if the amplitude of A is caused to vary with respect to time, as in the case of a sinusoidal wave:

$$A = 1 + \left(E_m/E_c\right)\cos(\omega_m t) \tag{2.5}$$

Where:
E_m/E_c = the ratio of modulation amplitude to carrier amplitude

The foregoing relationship leads to:

$$e(t) = E_c\left[1 + \left(E_m/E_c\right)\cos(\omega_m t)\cos(\omega_c t)\right] \tag{2.6}$$

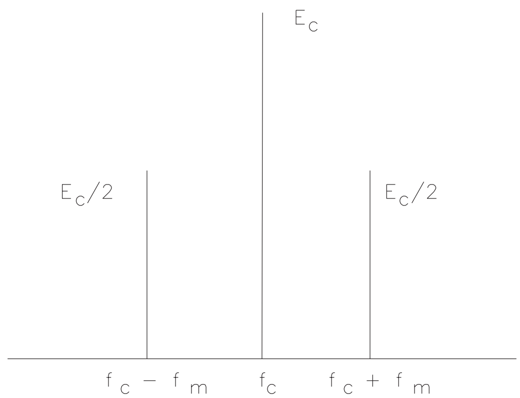


Figure 2.10 Frequency-domain representation of an amplitude-modulated signal at 100 percent modulation. E_c = carrier power, F_c = frequency of the carrier, and F_m = frequency of the modulating signal.

This is the basic equation for periodic (sinusoidal) amplitude modulation. When all multiplications and a simple trigonometric identity are performed, the result is:

$$e(t) = E_c \cos(\omega_c t) + \left(\tfrac{M}{2}\right)\cos(\omega_c t + \omega_m t) + \left(\tfrac{M}{2}\right)\cos(\omega_c t - \omega_m t) \qquad (2.7)$$

Where:
M = the amplitude modulation factor (E_m/E_c)

Amplitude modulation is, essentially, a multiplication process in which the time functions that describe the modulating signal and the carrier are multiplied to produce a modulated wave containing *intelligence* (information or data of some kind). The frequency components of the modulating signal are translated in this process to occupy a different position in the spectrum.

The bandwidth of an AM transmission is determined by the modulating frequency. The bandwidth required for full-fidelity reproduction in a receiver is equal to twice the applied modulating frequency.

The magnitude of the upper sideband and lower sideband will not normally exceed 50 percent of the carrier amplitude during modulation. This results in an upper-sideband power of one-fourth the carrier power. The same power exists in the lower sideband. As a result, up to one-half of the actual carrier power appears additionally in the sum of the sidebands of the modulated signal. A representation of the AM carrier and its sidebands is shown in Figure 2.10. The actual occupied bandwidth, assuming pure sinusoidal modulating signals and no distortion during the modulation process, is equal to twice the frequency of the modulating signal.

The extent of the amplitude variations in a modulated wave is expressed in terms of the *degree of modulation* or *percentage of modulation*. For sinusoidal variation, the degree of modulation m is determined from the following:

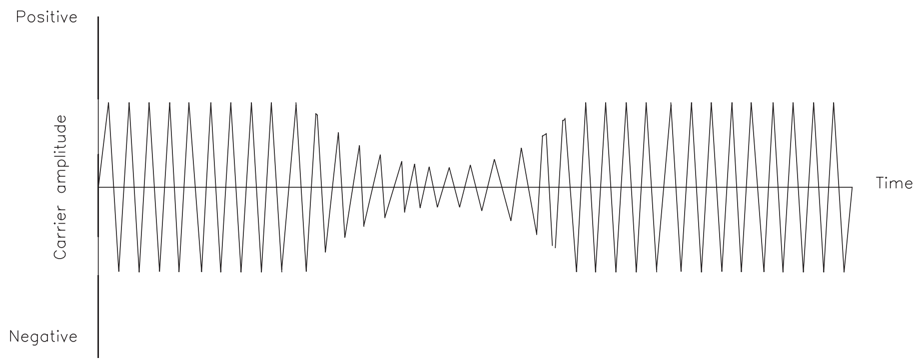


Figure 2.11 Time-domain representation of an amplitude-modulated signal. Modulation at 100 percent is defined as the point at which the peak of the waveform reaches twice the carrier level, and the minimum point of the waveform is zero.

$$m = \frac{E_{avg} - E_{\min}}{E_{avg}} \tag{2.8}$$

Where:
E_{avg} = average envelope amplitude
E_{min} = minimum envelope amplitude

Full (100 percent) modulation occurs when the peak value of the modulated envelope reaches twice the value of the unmodulated carrier, and the minimum value of the envelope is zero. The envelope of a modulated AM signal in the time domain is shown in Figure 2.11.

When the envelope variation is not sinusoidal, it is necessary to define the degree of modulation separately for the peaks and troughs of the envelope:

$$m_{pp} = \frac{E_{\max} - E_{avg}}{E_{avg}} \times 100 \tag{2.9}$$

$$m_{np} = \frac{E_{avg} - E_{\min}}{E_{avg}} \times 100 \tag{2.10}$$

Where:
m_{pp} = positive peak modulation (percent)
E_{max} = peak value of modulation envelope
m_{np} = negative peak modulation (percent)

E_{avg} = average envelope amplitude
E_{min} = minimum envelope amplitude

When modulation exceeds 100 percent on the negative swing of the carrier, spurious signals are emitted. It is possible to modulate an AM carrier asymmetrically; that is, to restrict modulation in the negative direction to 100 percent, but to allow modulation in the positive direction to exceed 100 percent without a significant loss of fidelity. In fact, many modulating signals normally exhibit asymmetry, most notably human speech waveforms.

The carrier wave represents the average amplitude of the envelope and, because it is the same regardless of the presence or absence of modulation, the carrier transmits no information. The information is carried by the sideband frequencies. The amplitude of the modulated envelope may be expressed as follows [5]:

$$E = E_0 + E_1 \sin\left(2\pi f_1 t + \Phi_1\right) + E_2 \sin\left(2\pi f_2 t + \Phi_2\right) \tag{2.11}$$

Where:
E = envelope amplitude
E_0 = carrier wave crest value (volts)
E_1 = 2 × first sideband crest amplitude (volts)
f_1 = frequency difference between the carrier and the first upper/lower sidebands
E_2 = 2 × second sideband crest amplitude (volts)
f_2 = frequency difference between the carrier and the second upper/lower sidebands
Φ_1 = phase of the first sideband component
Φ_2 = phase of the second sideband component

2.2.1 High-Level AM Modulation

High-level anode modulation is the oldest and simplest method of generating a high-power AM signal. In this system, the modulating signal is amplified and combined with the dc supply source to the anode of the final RF amplifier stage. The RF amplifier is normally operated class C. The final stage of the modulator usually consists of a pair of tubes operating class B in a push-pull configuration. A basic high-level modulator is shown in Figure 2.12.

The RF signal normally is generated in a low-level transistorized oscillator. It is then amplified by one or more solid-state or vacuum tube stages to provide final RF drive at the appropriate frequency to the grid of the final class C amplifier. The modulation input is applied to an intermediate power amplifier (usually solid-state) and used to drive two class B (or class AB) push-pull output devices. The final amplifiers provide the necessary modulating power to drive the final RF stage. For 100 percent modulation, this modulating power is equal to 50 percent of the actual carrier power.

The modulation transformer shown in Figure 2.12 does not usually carry the dc supply current for the final RF amplifier. The modulation reactor and capacitor shown provide a means to combine the signal voltage from the modulator with the dc supply to the final RF amplifier. This arrangement eliminates the necessity of having direct current

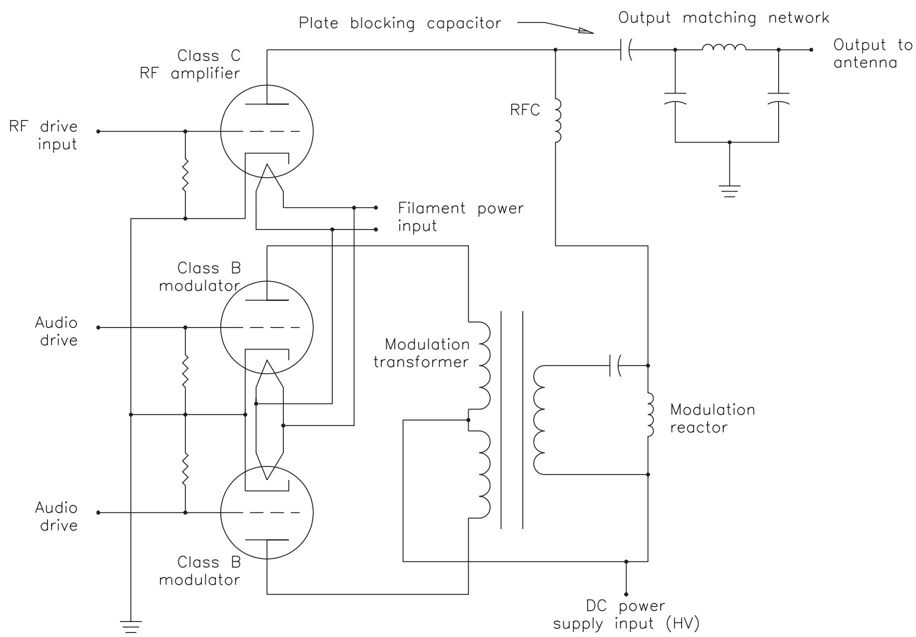


Figure 2.12 Simplified diagram of a high-level amplitude-modulated amplifier.

flow through the secondary of the modulation transformer, which would result in magnetic losses and saturation effects. In some transmitter designs, the modulation reactor is eliminated from the system, thanks to improvements in transformer technology.

The RF amplifier normally operates class C with grid current drawn during positive peaks of the cycle. Typical stage efficiency is 75 to 83 percent. An RF tank following the amplifier resonates the output signal at the operating frequency and, with the assistance of a low-pass filter, eliminates harmonics of the amplifier caused by class C operation.

This type of system was popular in AM applications for many years, primarily because of its simplicity. The primary drawback is low overall system efficiency. The class B modulator tubes cannot operate with greater than 50 percent efficiency. Still, with inexpensive electricity, this was not considered to be a significant problem. As energy costs increased, however, more efficient methods of generating high-power AM signals were developed. Increased efficiency normally came at the expense of added technical complexity.

2.2.2 Vestigial-Sideband Amplitude Modulation

Because the intelligence (modulating signal) of conventional AM transmission is identical in the upper *and* lower sidebands, it is possible to eliminate one sideband and

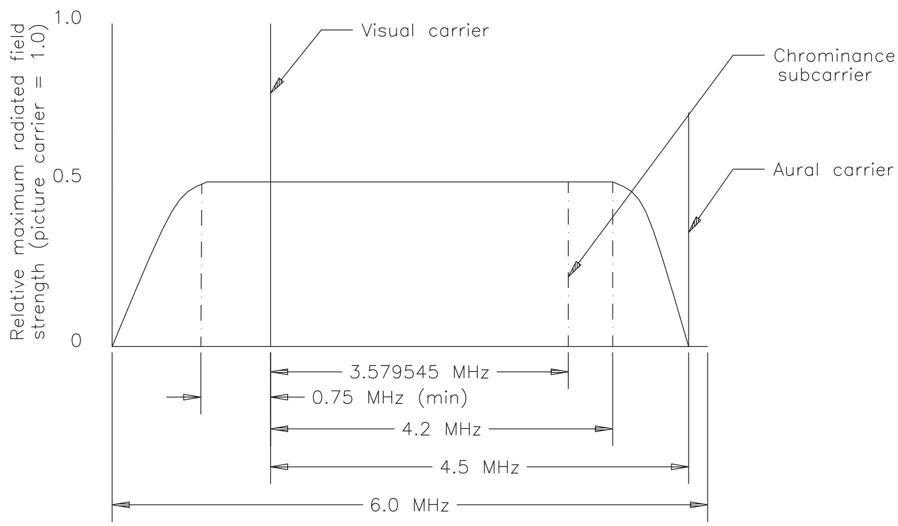


Figure 2.13 Idealized amplitude characteristics of the FCC standard waveform for monochrome and color TV transmission. (*Adapted from*: FCC Rules, Sec. 73.699.)

still convey the required information. This scheme is implemented in *vestigial-sideband AM* (VSBAM). Complete elimination of one sideband (for example, the lower sideband) requires an ideal high-pass filter with infinitely sharp cutoff. Such a filter is difficult to implement in any practical design. VSBAM is a compromise technique wherein one sideband (typically the lower sideband) is attenuated significantly. The result is a savings in occupied bandwidth and transmitter power.

VSBAM is used for television broadcast transmission and other applications. A typical bandwidth trace for a VSBAM TV transmitter is shown in Figure 2.13.

2.2.3 Single-Sideband Amplitude Modulation

The carrier in an AM signal does not convey any intelligence. All of the modulating information is in the sidebands. It is possible, therefore, to suppress the carrier upon transmission, radiating only one or both sidebands of the AM signal. The result is much greater efficiency at the transmitter (that is, a reduction in the required transmitter power). Suppression of the carrier may be accomplished with DSAM and SSBAM signals. *Single-sideband suppressed carrier* AM (SSB-SC) is the most spectrum- and energy-efficient mode of AM transmission. Figure 2.14 shows representative waveforms for suppressed carrier transmissions.

A waveform with carrier suppression differs from a modulated wave containing a carrier primarily in that the envelope varies at twice the modulating frequency. In addition, it will be noted that the SSB-SC wave has an apparent phase that reverses every time the modulating signal passes through zero. The wave representing a single sideband consists of a number of frequency components, one for each component in the

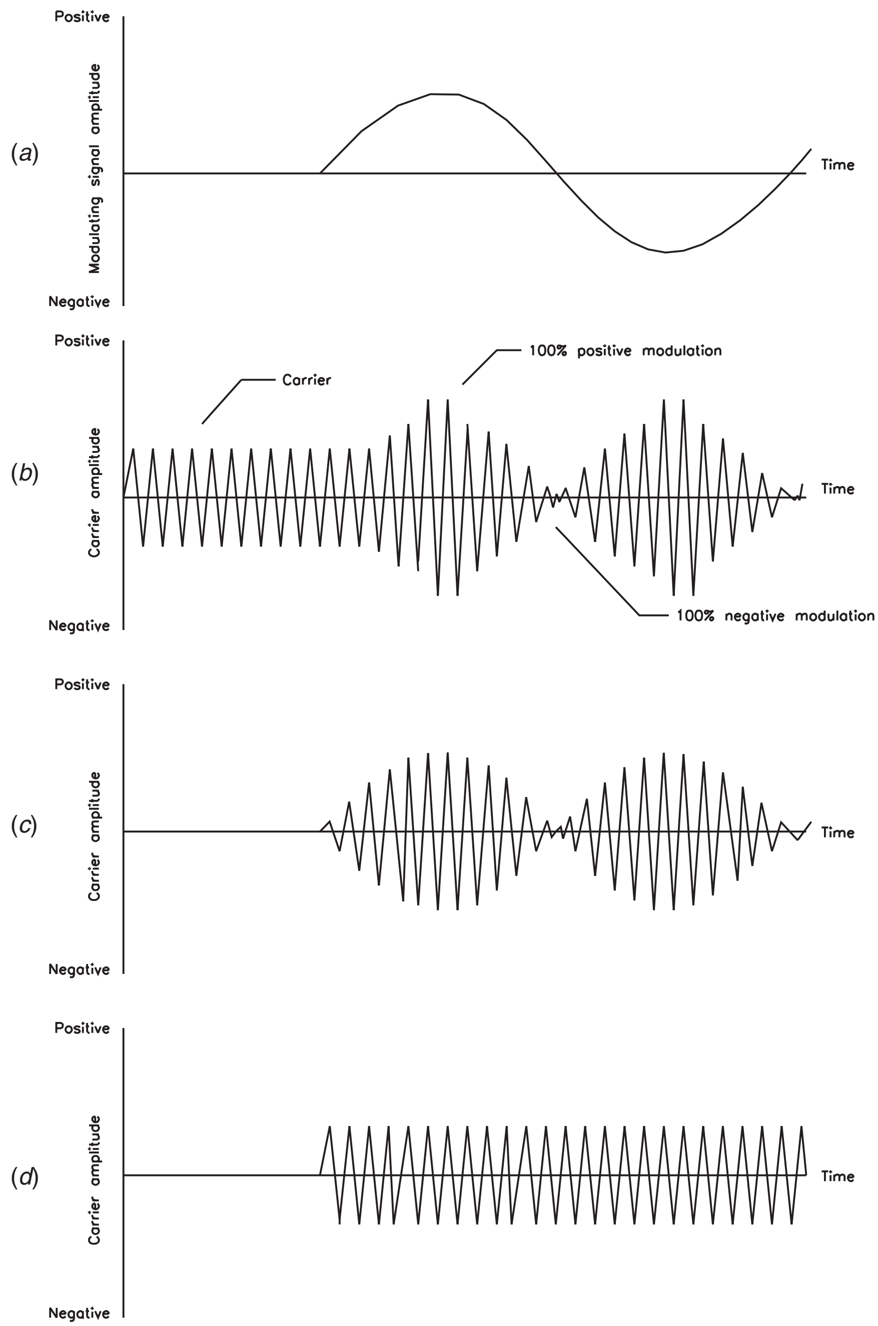


Figure 2.14 Types of suppressed carrier amplitude modulation: (*a*) the modulating signal, (*b*) double-sideband AM signal, (*c*) double-sideband suppressed carrier AM, (*d*) single-sideband suppressed carrier AM.

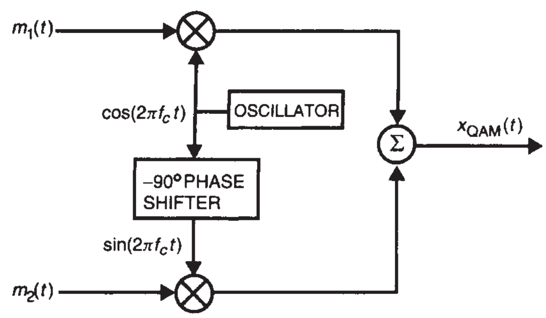


Figure 2.15 Simplified QAM modulator. (*From* [6]. *Used with permission.*)

original signal. Each of these components has an amplitude proportional to the amplitude of the corresponding modulating component and a frequency differing from that of the carrier by the modulating frequency. The result is that, in general, the envelope amplitude of the single sideband signal increases with the degree of modulation, and the envelope varies in amplitude in accordance with the difference frequencies formed by the various frequency components of the single sideband interacting with each other.

An SSB-SC system is capable of transmitting a given intelligence within a frequency band only half as wide as that required by a DSAM waveform. Furthermore, the SSB system saves more than two-thirds of the transmission power because of the elimination of one sideband and the carrier.

The drawback to suppressed carrier systems is the requirement for a more complicated receiver. The carrier must be regenerated at the receiver to permit demodulation of the signal. Also, in the case of SSBAM transmitters, it is usually necessary to generate the SSB signal in a low-power stage and then amplify the signal with a linear power amplifier to drive the antenna. Linear amplifiers generally exhibit low efficiency.

2.2.4 Quadrature Amplitude Modulation (QAM)

Single sideband transmission makes very efficient use of the spectrum; for example, two SSB signals can be transmitted within the bandwidth normally required for a single DSB signal. However, DSB signals can achieve the same efficiency by means of *quadrature amplitude modulation* (QAM), which permits two DSB signals to be transmitted and received simultaneously using the same carrier frequency [6]. A basic QAM DSB modulator is shown schematically in Figure 2.15.

Two DSB signals coexist separately within the same bandwidth by virtue of the 90° phase shift between them. The signals are, thus, said to be in *quadrature*. Demodulation uses two local oscillator signals that are also in quadrature, i.e., a sine and a cosine signal, as illustrated in Figure 2.16.

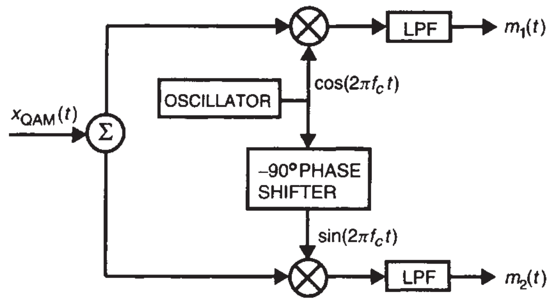


Figure 2.16 Simplified QAM demodulator. (*From* [6]. *Used with permission.*)

The chief disadvantage of QAM is the need for a coherent local oscillator at the receiver exactly in phase with the transmitter oscillator signal. Slight errors in phase or frequency can cause both loss of signal and interference between the two signals (cochannel interference or crosstalk).

The relative merits of the various AM systems are summarized in Table 2.2

2.3 Frequency Modulation

Frequency modulation is a technique whereby the phase angle or phase shift of a carrier is varied by an applied modulating signal. The *magnitude* of frequency change of the carrier is a direct function of the *magnitude* of the modulating signal. The *rate* at which the frequency of the carrier is changed is a direct function of the *frequency* of the modulating signal. In FM modulation, multiple pairs of sidebands are produced. The actual number of sidebands that make up the modulated wave is determined by the *modulation index* (MI) of the system.

2.3.1 Modulation Index

The modulation index is a function of the frequency deviation of the system and the applied modulating signal:

$$MI = \frac{F_d}{M_f} \qquad (2.12)$$

Where:
MI = the modulation index
F_d = frequency deviation
M_f = modulating frequency

Table 2.2 Comparison of Amplitude Modulation Techniques (*After* [6].)

Modulation Scheme	Advantages	Disadvantages	Comments
DSB SC	Good power efficiency; good low-frequency response.	More difficult to generate than DSB+C; detection requires coherent local oscillator, pilot, or phase-locked loop; poor spectrum efficiency.	
DSB+C	Easier to generate than DSB SC, especially at high-power levels; inexpensive receivers using envelope detection.	Poor power efficiency; poor spectrum efficiency; poor low-frequency response; exhibits threshold effect in noise.	Used in commercial AM broadcasting.
SSB SC	Excellent spectrum efficiency.	Complex transmitter design; complex receiver design (same as DSB SC); poor low-frequency response.	Used in military communication systems, and to multiplex multiple phone calls onto long-haul microwave links.
SSB+C	Good spectrum efficiency; low receiver complexity.	Poor power efficiency; complex transmitters; poor low-frequency response; poor noise performance.	
VSB SC	Good spectrum efficiency; excellent low-frequency response; transmitter easier to build than for SSB.	Complex receivers (same as DSB SC).	
VSB+C	Good spectrum efficiency; good low-frequency response; inexpensive receivers using envelope detection.	Poor power efficiency; poor performance in noise.	Used in commercial TV broadcasting.
QAM	Good low-frequency response; good spectrum efficiency.	Complex receivers; sensitive to frequency and phase errors.	Two SSB signals may be preferable.

The higher the MI, the more sidebands produced. It follows that the higher the modulating frequency for a given deviation, the fewer number of sidebands produced, but the greater their spacing.

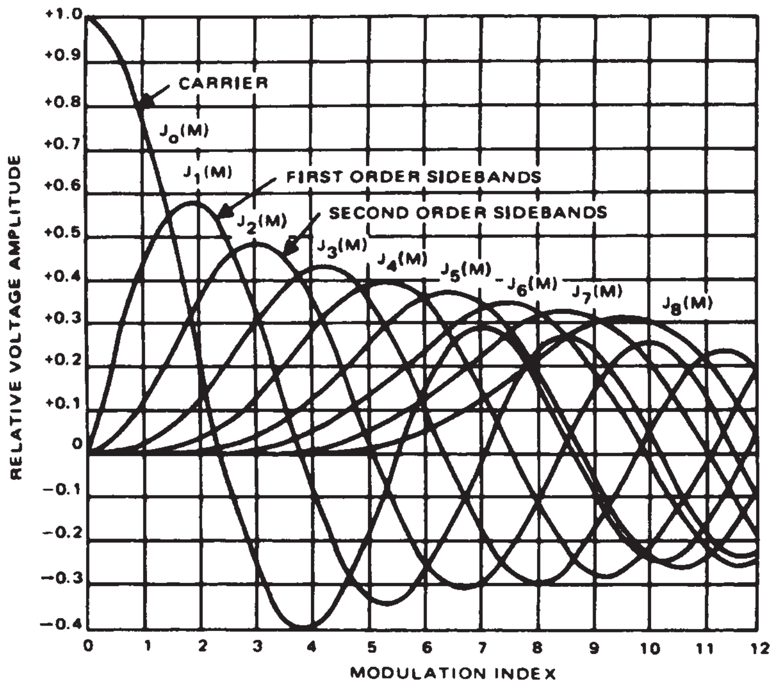


Figure 2.17 Plot of Bessel functions of the first kind as a function of modulation index.

To determine the frequency spectrum of a transmitted FM waveform, it is necessary to compute a Fourier series or Fourier expansion to show the actual signal components involved. This work is difficult for a waveform of this type, because the integrals that must be performed in the Fourier expansion or Fourier series are not easily solved. The result, however, is that the integral produces a particular class of solution that is identified as the *Bessel function*, illustrated in Figure 2.17.

The carrier amplitude and phase, plus the sidebands, can be expressed mathematically by making the modulation index the argument of a simplified Bessel function. The general expression is given from the following equations:

RF output voltage = $E_1 = E_c + S_{1u} - S_{1l} + S_{2u} - S_{2l} + S_{3u} - S_{3l} + S_{nu} - S_{nl}$

Carrier amplitude = $E_c = A\left[J_0(M)\sin\omega c(t)\right]$

First-order upper sideband = $S_{1u} = J_1(M)\sin(\omega c + \omega m)t$

First-order lower sideband = $S_{1l} = J_1(M)\sin(\omega c - \omega m)t$

Second-order upper sideband = $S_{2u} = J_2(M)\sin(\omega c + 2\omega m)t$

Second-order lower sideband = $S_{2l} = J_2(M)\sin(\omega c - 2\omega m)t$

Third-order upper sideband = $S_{3u} = J_3(M)\sin(\omega c + 3\omega m)t$

Third-order lower sideband = $S_{3l} = J_3(M)\sin(\omega c - 3\omega m)t$

Nth-order upper sideband = $S_{nu} = J_n(M)\sin(\omega c + n\omega m)t$

Nth-order lower sideband = $S_{nl} = J_n(M)\sin(\omega c - n\omega m)t$

Where:
A = the unmodulated carrier amplitude constant
J_0 = modulated carrier amplitude
$J_1, J_2, J_3 \ldots J_n$ = amplitudes of the nth-order sidebands
M = modulation index
$\omega c = 2\pi F_c$, the carrier frequency
$\omega m = 2\pi F_m$, the modulating frequency

Further supporting mathematics will show that an FM signal using the modulation indices that occur in a wideband system will have a multitude of sidebands. From the purist point of view, *all* sidebands would have to be transmitted, received, and demodulated to reconstruct the modulating signal with complete accuracy. In practice, however, the channel bandwidths permitted FM systems usually are sufficient to reconstruct the modulating signal with little discernible loss in fidelity, or at least an acceptable loss in fidelity.

Figure 2.18 illustrates the frequency components present for a modulation index of 5. Figure 2.19 shows the components for an index of 15. Note that the number of significant sideband components becomes quite large with a high MI. This simple representation of a single-tone frequency-modulated spectrum is useful for understanding the general nature of FM, and for making tests and measurements. When typical modulation signals are applied, however, many more sideband components are generated. These components vary to the extent that sideband energy becomes distributed over the entire occupied bandwidth, rather than appearing at discrete frequencies.

Although complex modulation of an FM carrier greatly increases the number of frequency components present in the frequency-modulated wave, it does not, in general, widen the frequency band occupied by the energy of the wave. To a first approximation, this band is still roughly twice the sum of the maximum frequency deviation at the peak of the modulation cycle plus the highest modulating frequency involved.

FM is not a simple frequency translation, as with AM, but involves the generation of entirely new frequency components. In general, the new spectrum is much wider than the original modulating signal. This greater bandwidth may be used to improve the *signal-to-noise ratio* (S/N) of the transmission system. FM thereby makes it possible to exchange bandwidth for S/N enhancement.

The power in an FM system is constant throughout the modulation process. The output power is increased in the amplitude modulation system by the modulation process,

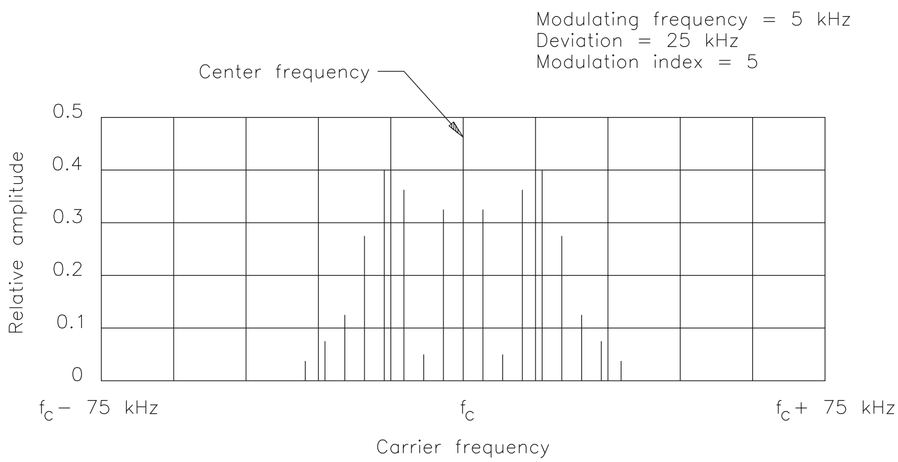


Figure 2.18 RF spectrum of a frequency-modulated signal with a modulation index of 5 and other operating parameters as shown.

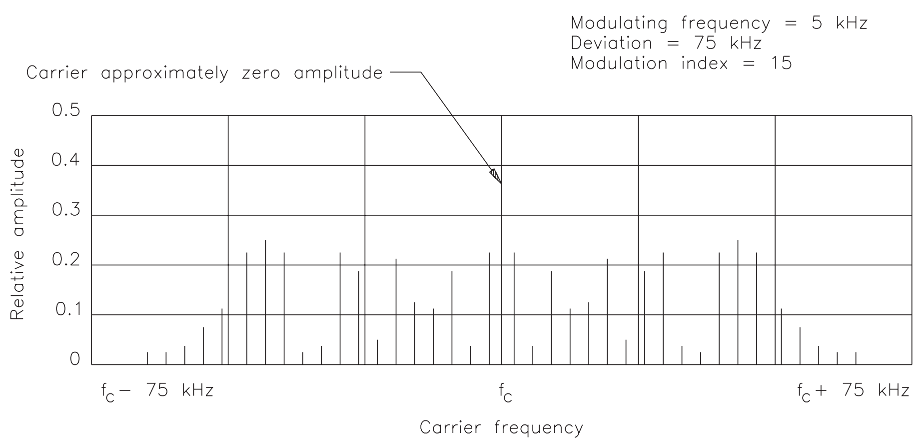


Figure 2.19 RF spectrum of a frequency-modulated signal with a modulation index of 15 and operating parameters as shown.

but the FM system simply distributes the power throughout the various frequency components that are produced by modulation. During modulation, a wideband FM system does not have a high amount of energy present in the carrier. Most of the energy will be found in the sum of the sidebands.

The constant-amplitude characteristic of FM greatly assists in capitalizing on the low noise advantage of FM reception. Upon being received and amplified, the FM signal normally is clipped to eliminate all amplitude variations above a certain threshold. This removes noise picked up by the receiver as a result of man-made or atmospheric signals. It is not possible (generally speaking) for these random noise sources to change the frequency of the desired signal; they can affect only its amplitude. The use of *hard limiting* in the receiver will strip off such interference.

2.3.2 Phase Modulation

In a phase modulation (PM) system, intelligence is conveyed by varying the phase of the RF wave. Phase modulation is similar in many respects to frequency modulation, except in the interpretation of the modulation index. In the case of PM, the modulation index depends only on the amplitude of the modulation; MI is independent of the frequency of the modulating signal. It is apparent, therefore, that the phase-modulated wave contains the same sideband components as the FM wave and, if the modulation indices in the two cases are the same, the relative amplitudes of these different components also will be the same.

The modulation parameters of a PM system relate as follows:

$$\Delta f = m_p \times f_m \tag{2.13}$$

Where:
Δf = frequency deviation of the carrier
m_p = phase shift of the carrier
f_m = modulating frequency

In a phase-modulated wave, the phase shift m_p is independent of the modulating frequency; the frequency deviation Δf is proportional to the modulating frequency. In contrast, with a frequency-modulated wave, the frequency deviation is independent of modulating frequency. Therefore, a frequency-modulated wave can be obtained from a phase modulator by making the modulating voltage applied to the phase modulator inversely proportional to frequency. This can be readily achieved in hardware.

2.3.3 Modifying FM Waves

When a frequency-modulated wave is passed through a harmonic generator, the effect is to increase the modulation index by a factor equal to the frequency multiplication involved. Similarly, if the frequency-modulated wave is passed through a frequency divider, the effect is to reduce the modulation index by the factor of frequency division. Thus, the frequency components contained in the wave—and, consequently, the bandwidth of the wave—will be increased or decreased, respectively, by frequency multiplication or division. No distortion in the nature of the modulation is introduced by the frequency change.

When an FM wave is translated in the frequency spectrum by heterodyne action, the modulation index—hence the relative positions of the sideband frequencies and the bandwidths occupied by them—remains unchanged.

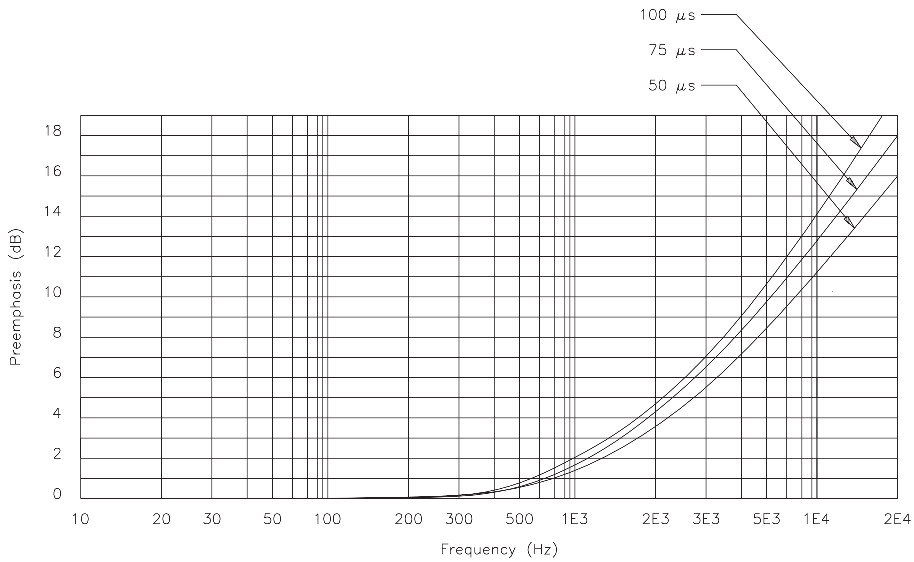


Figure 2.20 Preemphasis curves for time constants of 50, 75, and 100 μs.

2.3.4 Preemphasis and Deemphasis

The FM transmission/reception system offers significantly better noise-rejection characteristics than AM. However, FM noise rejection is more favorable at low modulating frequencies than at high frequencies because of the reduction in the number of sidebands at higher frequencies. To offset this problem, the input signal to the FM transmitter may be *preemphasized* to increase the amplitude of higher-frequency signal components in normal program material. FM receivers utilize complementary *deemphasis* to produce flat overall system frequency response.

FM broadcasting, for example, uses a 75 μs preemphasis curve, meaning that the time constant of the resistance-inductance (*RL*) or resistance-capacitance (*RC*) circuit used to provide the boost of high frequencies is 75 μs. Other values of preemphasis are used in different types of FM communications systems. Figure 2.20 shows three common preemphasis curves.

2.3.5 Modulation Circuits

Early FM transmitters used *reactance modulators* that operated at a low frequency. The output of the modulator then was multiplied to reach the desired output frequency. This approach was acceptable for some applications and unsuitable for others. Modern FM systems utilize what is referred to as *direct modulation*; that is, the

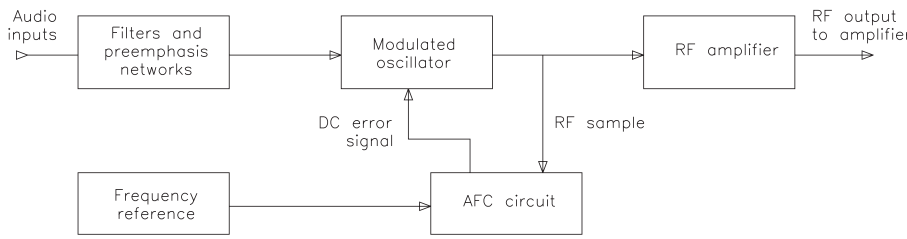


Figure 2.21 Block diagram of an FM exciter.

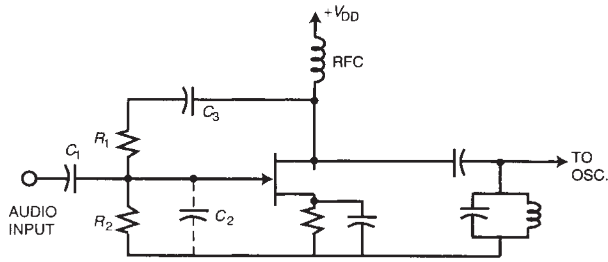
frequency modulation occurs in a modulated oscillator that operates on a center frequency equal to the desired transmitter output frequency.

Various techniques have been developed to generate the direct-FM signal. One of the most popular uses a variable-capacity diode as the reactive element in the oscillator. The modulating signal is applied to the diode, which causes the capacitance of the device to vary as a function of the magnitude of the modulating signal. Variations in the capacitance cause the frequency of the oscillator to change. The magnitude of the frequency shift is proportional to the amplitude of the modulating signal, and the rate of frequency shift is equal to the frequency of the modulating signal.

The direct-FM modulator is one element of an FM transmitter *exciter*, which generates the composite FM waveform. A block diagram of a complete FM exciter is shown in Figure 2.21. Input signals are buffered, filtered, and preemphasized before being summed to feed the modulated oscillator. Note that the oscillator is not normally coupled directly to a crystal, but to a free-running oscillator adjusted as closely as possible to the carrier frequency of the transmitter. The final operating frequency is maintained carefully by an automatic frequency control system employing a *phase locked loop* (PLL) tied to a reference crystal oscillator or frequency synthesizer.

A solid-state class C amplifier typically follows the modulated oscillator and raises the operating power of the FM signal to 20 to 30 W. One or more subsequent amplifiers in the transmitter raise the signal power to several hundred watts for application to the final power amplifier stage. Nearly all high-power FM broadcast transmitters use solid-state amplifiers up to the final RF stage, which is generally a vacuum tube for operating powers of 15 kW and above. All stages operate in the class C mode. In contrast to AM systems, each stage in an FM power amplifier can operate class C; no information is lost from the frequency-modulated signal because of amplitude changes. As mentioned previously, FM is a constant-power system.

Figure 2.22 Simplified reactance modulator. (*From* [7]. *Used with permission*.)



Direct-FM Modulator

Many types of circuits have been used to produce a direct-FM signal [7]. In each case, a reactance device is used to shunt capacitive or inductive reactance across an oscillator. The value of capacitive or inductive reactance is made to vary as the amplitude of the modulating signal varies. Because the reactive load is placed across an oscillator's tuned circuit, the frequency of the oscillator will therefore shift by a predetermined amount, thereby creating an FM signal.

A typical example of a reactance modulator is illustrated in Figure 2.22. The circuit uses a field-effect transistor (FET), where the modulating signal is applied to the modulator through C_1. The actual components that affect the overall reactance consist of R_1 and C_2. Typically, the value of C_2 is small as this is the input capacitance to the FET, which may only be a few picofarads. However, this capacitance will generally be much larger by a significant amount as a result of the *Miller effect*. Capacitor C_3 has no significant effect on the reactance of the modulator; it is strictly a blocking capacitor that keeps dc from changing the gate bias of the FET.

To further understand the performance of the reactance modulator, the equivalent circuit of Figure 2.22 is represented in Figure 2.23. The FET is represented as a current source, gmV_g, with the internal drain resistance, r_d. The impedances Z_1 and Z_2 are a combination of resistance and capacitive reactance, which are designed to provide a 90º phase shift.

Using vector diagrams, we can analyze the phase relationship of the reactance modulator. Referring to Figure 2.22, the resistance of R_1 is typically high compared to the capacitive reactance of C_2. The R_1C_2 circuit is then resistive. Because this circuit is resistive, the current, I_{AB}, that flows through it, is in phase with the voltage V_{AB}. Voltage V_{AB} is also across R_1C_2 (or Z_1Z_2 in Figure 2.23). This is true because current and voltage tend to be in phase in a resistive network. However, voltage V_{C2}, which is across C_2, is out of phase with I_{AB}. This is because the voltage that is across a capacitor lags behind its current by 90°. (See Figure 2.24.)

Figure 2.23 Equivalent circuit of the reactance modulator. (*From* [7]. *Used with permission*.)

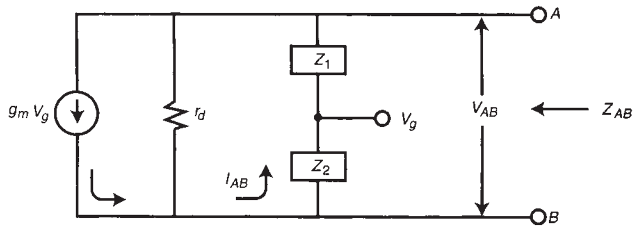


Figure 2.24 Vector diagram of a reactance modulator producing FM. Note: $V_g = V_{C2}$. (*From* [7]. *Used with permission*.)

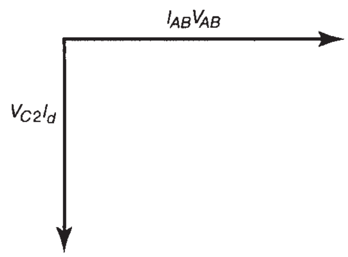
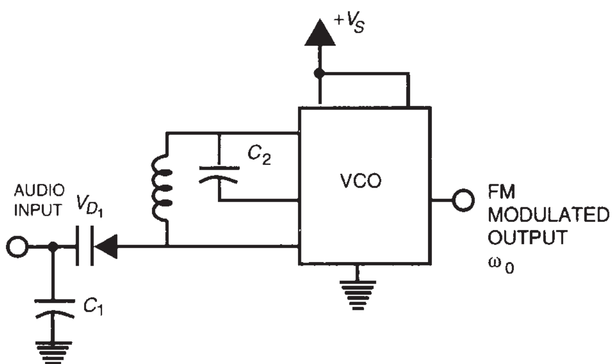


Figure 2.25 Voltage-controlled direct-FM modulator. (*From* [7]. *Used with permission*.)



VCO Direct-FM Modulator

One of the more common direct-FM modulation techniques uses an analog *voltage controlled oscillator* (VCO) in a phase locked loop arrangement [7]. In this configuration, shown in Figure 2.25, a VCO produces the desired carrier frequency that is—in turn—modulated by applying the input signal to the VCO input via a variactor diode. The variactor is used to vary the capacitance of an oscillator tank circuit. Therefore, the variactor behaves as a variable capacitor whose capacitance changes as the signal voltage across it changes. As the input capacitance of the VCO is varied by

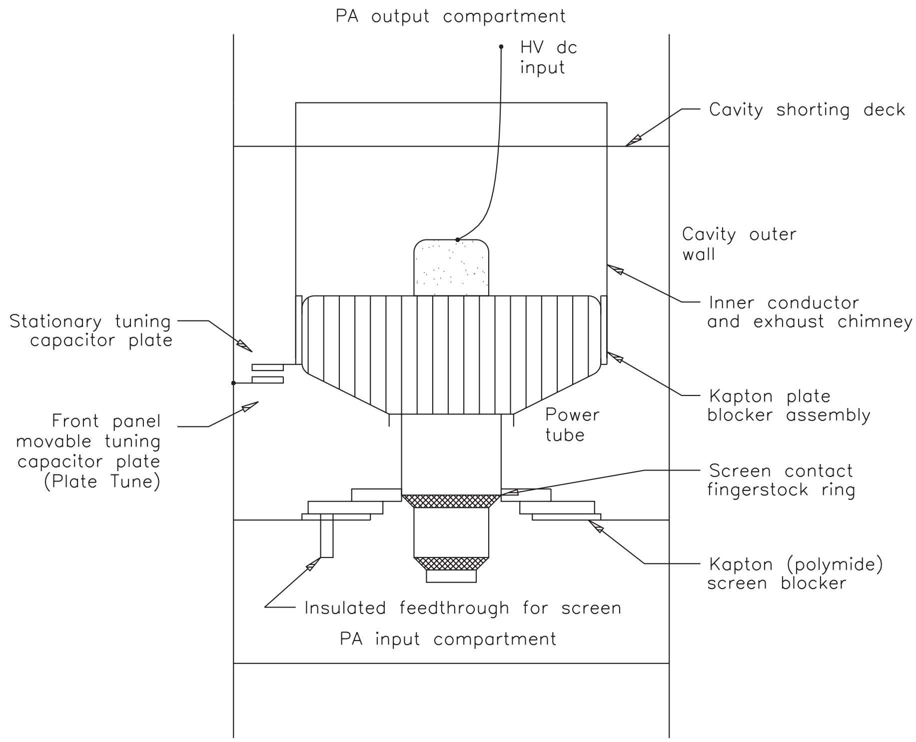


Figure 2.26 Mechanical layout of a common type of 1/4-wave PA cavity for FM service.

the variactor, the output frequency of the VCO is shifted, which produces a direct-FM modulated signal.

FM Power Amplifiers

Nearly all high-power FM transmitters employ cavity designs. The 1/4-wavelength cavity is the most common. The design is simple and straightforward. A number of variations can be found in different transmitters, but the underlying theory of operation is the same. The goal of any cavity amplifier is to simulate a resonant tank circuit at the operating frequency and provide a means to couple the energy in the cavity to the transmission line.

A typical 1/4-wave cavity is shown in Figure 2.26. The plate of the tube connects directly to the inner section (tube) of the plate-blocking capacitor. The blocking capacitor can be formed in one of several ways. In at least one design, it is made by wrapping the outside surface of the inner tube conductor with multiple layers of insulating film. The exhaust chimney/inner conductor forms the other element of the blocking capacitor. The cavity walls form the outer conductor of the 1/4-wave transmission line circuit. The dc plate voltage is applied to the PA tube by a cable routed inside the exhaust chimney and inner tube conductor. In this design, the screen-contact fingerstock ring mounts on a metal plate that is insulated from the grounded cavity deck by a blocking capacitor. This hardware makes up the screen-blocker assembly. The dc screen voltage feeds to the fingerstock ring from underneath the cavity deck using an insulated feedthrough.

Some transmitters that employ the 1/4-wave cavity design use a grounded-screen configuration in which the screen-contact fingerstock ring is connected directly to the grounded cavity deck. The PA cathode then operates at below ground potential (in other words, at a negative voltage), establishing the required screen voltage for the tube.

Coarse-tuning of the cavity is accomplished by adjusting the cavity length. The top of the cavity (the *cavity shorting deck*) is fastened by screws or clamps and can be raised or lowered to set the length of the assembly for the particular operating frequency. Fine-tuning is accomplished by a variable-capacity plate-tuning control built into the cavity. In the example shown in Figure 2.26, one plate of this capacitor—the stationary plate—is fastened to the inner conductor just above the plate-blocking capacitor. The movable tuning plate is fastened to the cavity box, the outer conductor, and is linked mechanically to the front-panel tuning control. This capacity shunts the inner conductor to the outer conductor and varies the electrical length and resonant frequency of the cavity.

The theory of operation of cavity amplifier systems is discussed in Section 4.3.

2.4 Pulse Modulation

The growth of digital processing and communications has led to the development of modulation systems tailor-made for high-speed, spectrum-efficient transmission. In a *pulse modulation* system, the unmodulated carrier usually consists of a series of recurrent pulses. Information is conveyed by modulating some parameter of the pulses, such as amplitude, duration, time of occurrence, or shape. Pulse modulation is based on the *sampling principle*, which states that a message waveform with a spectrum of finite width can be recovered from a set of discrete samples if the sampling rate is higher than twice the highest sampled frequency (the Nyquist criteria). The samples of the input signal are used to modulate some characteristic of the carrier pulses.

2.4.1 Digital Modulation Systems

Because of the nature of digital signals (on or off), it follows that the amplitude of the signal in a pulse modulation system should be one of two heights (present or absent/positive or negative) for maximum efficiency. Noise immunity is a significant advantage of such a system. It is necessary for the receiving system to detect only the presence or absence (or polarity) of each transmitted pulse to allow complete reconstruction of the original intelligence. The pulse shape and noise level have minimal effect (to a point). Furthermore, if the waveform is to be transmitted over long distances, it is possible to regenerate the original signal exactly for retransmission to the next relay point. This feature is in striking contrast to analog modulation systems in which each modulation step introduces some amount of noise and signal corruption.

In any practical digital data system, some corruption of the intelligence is likely to occur over a sufficiently large span of time. Data encoding and manipulation schemes have been developed to detect and correct or conceal such errors. The addition of error-correction features comes at the expense of increased system overhead and (usually) slightly lower intelligence throughput.

2.4.2 Pulse Amplitude Modulation

Pulse amplitude modulation (PAM) is one of the simplest forms of data modulation. PAM departs from conventional modulation systems in that the carrier exists as a series of pulses, rather than as a continuous waveform. The amplitude of the pulse train is modified in accordance with the applied modulating signal to convey intelligence, as illustrated in Figure 2.27. There are two primary forms of PAM sampling:

- *Natural sampling* (or *top sampling*), where the modulated pulses follow the amplitude variation of the sampled time function during the sampling interval.
- *Instantaneous sampling* (or *square-topped sampling*), where the amplitude of the pulses is determined by the instantaneous value of the sampled time function corresponding to a single instant of the sampling interval. This "single instant" may be the center or edge of the sampling interval.

There are two common methods of generating a PAM signal:

- Variation of the amplitude of a pulse sequence about a fixed nonzero value (or *pedestal*). This approach constitutes double-sideband amplitude modulation.
- Double-polarity modulated pulses with no pedestal. This approach constitutes double-sideband suppressed carrier modulation.

2.4.3 Pulse Time Modulation (PTM)

A number of modulating schemes have been developed to take advantage of the noise immunity afforded by a constant amplitude modulating system. *Pulse time modulation* (PTM) is one of those systems. In a PTM system, instantaneous samples of the intelligence are used to vary the time of occurrence of some parameter of the pulsed carrier. Subsets of the PTM process include:

- *Pulse duration modulation* (PDM), where the time of occurrence of either the leading or trailing edge of each pulse (or both pulses) is varied from its unmodulated position by samples of the input modulating waveform. PDM also may be described as *pulse length* or *pulse width* modulation (PWM).
- *Pulse position modulation* (PPM), where samples of the modulating input signal are used to vary the position in time of pulses, relative to the unmodulated waveform. Several types of pulse time modulation waveforms are shown in Figure 2.28.

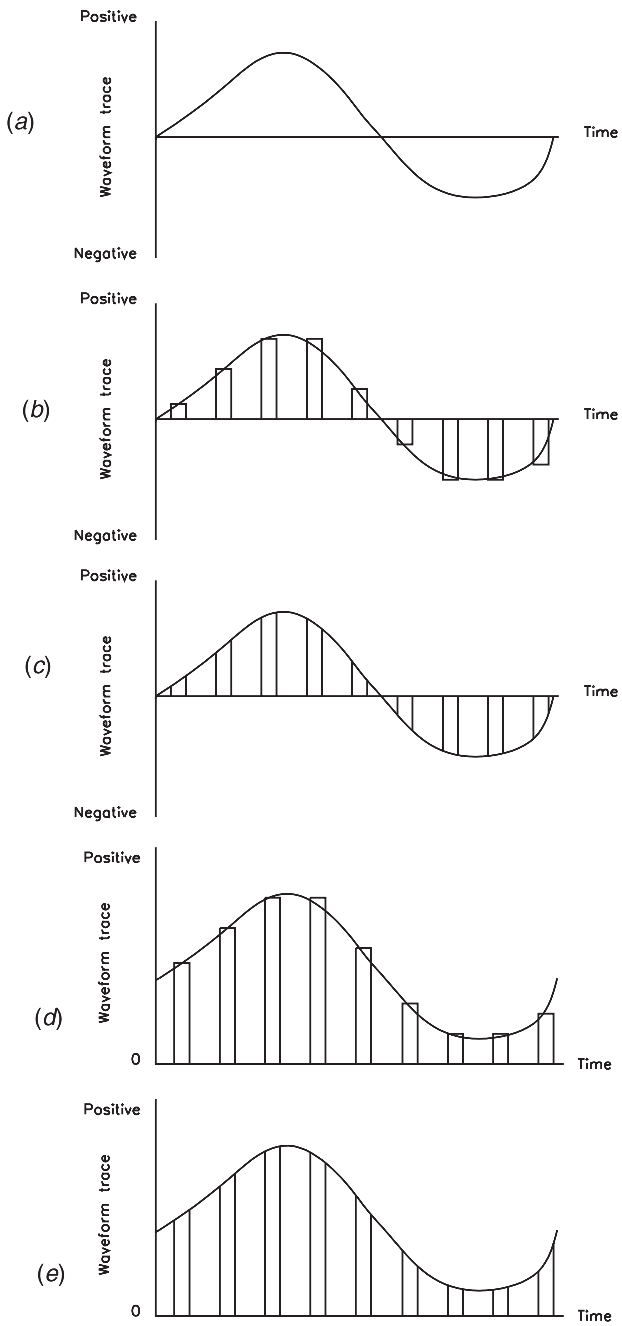


Figure 2.27 Pulse amplitude modulation waveforms: (*a*) modulating signal; (*b*) square-topped sampling, bipolar pulse train; (*c*) topped sampling, bipolar pulse train; (*d*) square-topped sampling, unipolar pulse train; (*e*) top sampling, unipolar pulse train.

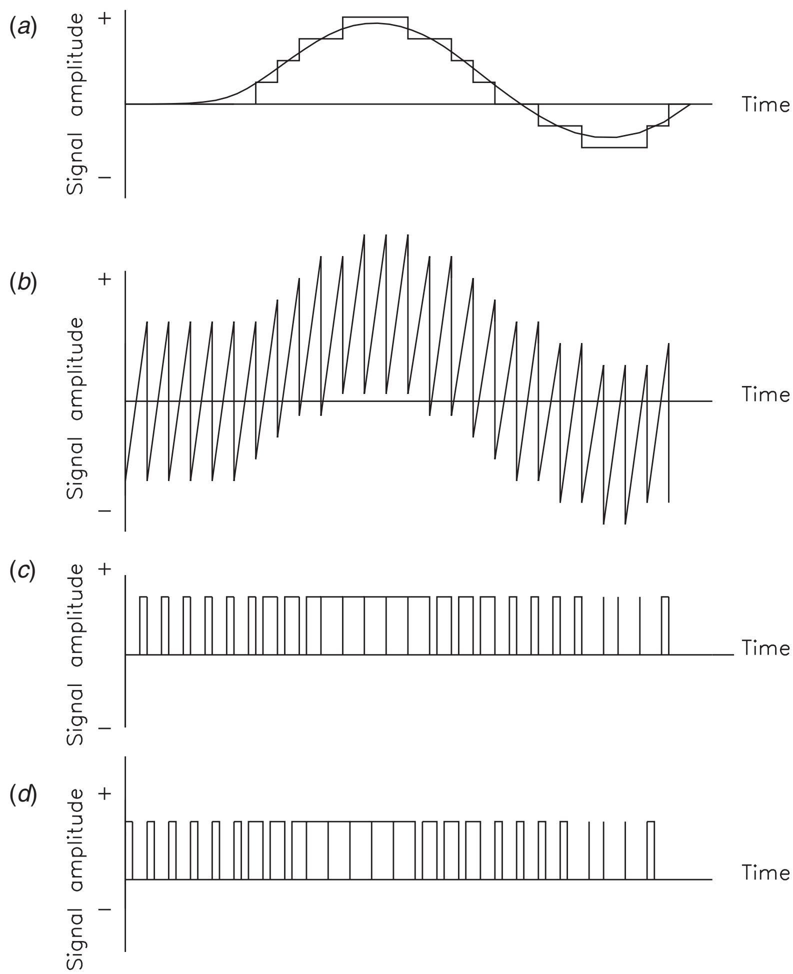


Figure 2.28 Pulse time modulation waveforms: (*a*) modulating signal and sample-and-hold (S/H) waveforms, (*b*) sawtooth waveform added to S/H, (*c*) leading-edge PTM, (*d*) trailing-edge PTM.

- *Pulse frequency modulation* (PFM), where samples of the input signal are used to modulate the frequency of a series of carrier pulses. The PFM process is illustrated in Figure 2.29.

It should be emphasized that all of the pulse modulation systems discussed thus far may be used with both analog and digital input signals. Conversion is required for either signal into a form that can be accepted by the pulse modulator.

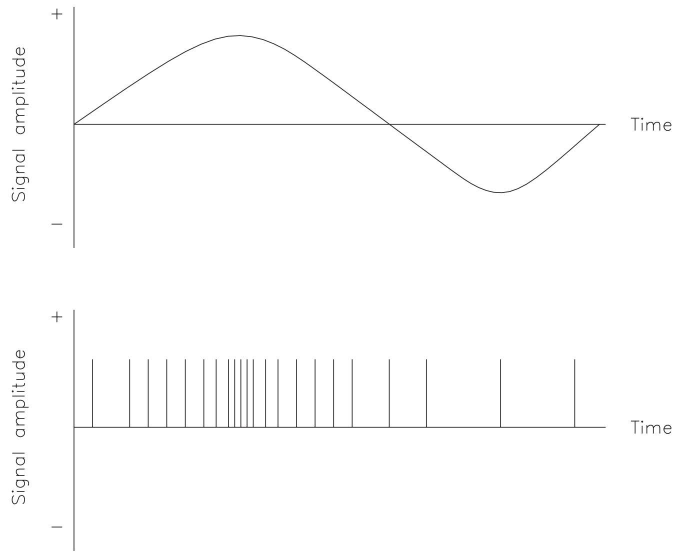


Figure 2.29 Pulse frequency modulation.

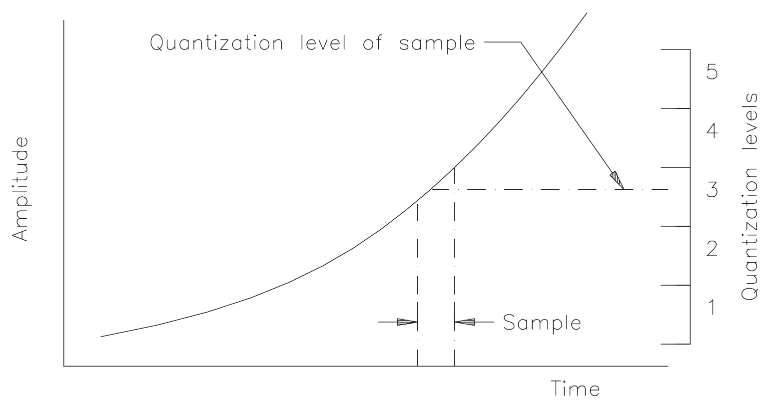


Figure 2.30 The quantization process.

2.4.4 Pulse Code Modulation

The pulse modulation systems discussed previously are *unencoded* systems. *Pulse code modulation* (PCM) is a scheme wherein the input signal is *quantized* into discrete steps and then sampled at regular intervals (as in conventional pulse modulation). In the *quantization* process, the input signal is sampled to produce a code representing the instantaneous value of the input within a predetermined range of values. Figure 2.30 illustrates the concept. Only certain discrete levels are allowed in the quantization process. The code is then transmitted over the communications system as a pattern of pulses.

Quantization inherently introduces an initial error in the amplitude of the samples taken. This *quantization error* is reduced as the number of quantization steps is increased. In system design, tradeoffs must be made regarding low quantization error, hardware complexity, and occupied bandwidth. The greater the number of quantization steps, the wider the bandwidth required to transmit the intelligence or, in the case of some signal sources, the slower the intelligence must be transmitted.

In the classic design of a PCM encoder, the quantization steps are equal. The quantization error (or *quantization noise*) usually can be reduced, however, through the use of nonuniform spacing of levels. Smaller quantization steps are provided for weaker signals, and larger steps are provided near the peak of large signals. Quantization noise is reduced by providing an encoder that is matched to the *level distribution* (*probability density*) of the input signal.

Nonuniform quantization typically is realized in an encoder through processing of the input (analog) signal to compress it to match the desired nonuniformity. After compression, the signal is fed to a uniform quantization stage.

2.4.5 Delta Modulation

Delta modulation (DM) is a coding system that measures changes in the direction of the input waveform, rather than the instantaneous value of the wave itself. Figure 2.31 illustrates the concept. The clock rate is assumed to be constant. Transmitted pulses from the pulse generator are positive if the signal is changing in a positive direction; they are negative if the signal is changing in a negative direction.

As with the PCM encoding system, quantization noise is a parameter of concern for DM. Quantization noise can be reduced by increasing the sampling frequency (the pulse generator frequency). The DM system has no fixed maximum (or minimum) signal amplitude. The limiting factor is the slope of the sampled signal, which must not change by more than one level or step during each pulse interval.

2.4.6 Digital Coding Systems

A number of methods exist to transmit digital signals over long distances in analog transmission channels. Some of the more common systems include:

- *Binary on-off keying* (BOOK), a method by which a high-frequency sinusoidal signal is switched on and off corresponding to 1 and 0 (on and off) periods in the input digital data stream. In practice, the transmitted sinusoidal waveform does not start or stop abruptly, but follows a predefined ramp up or down.
- *Binary frequency-shift keying* (BFSK), a modulation method in which a continuous wave is transmitted that is shifted between two frequencies, representing 1s and 0s in the input data stream. The BFSK signal may be generated by switching between two oscillators (set to different operating frequencies) or by applying a binary baseband signal to the input of a voltage-controlled oscillator (VCO). The

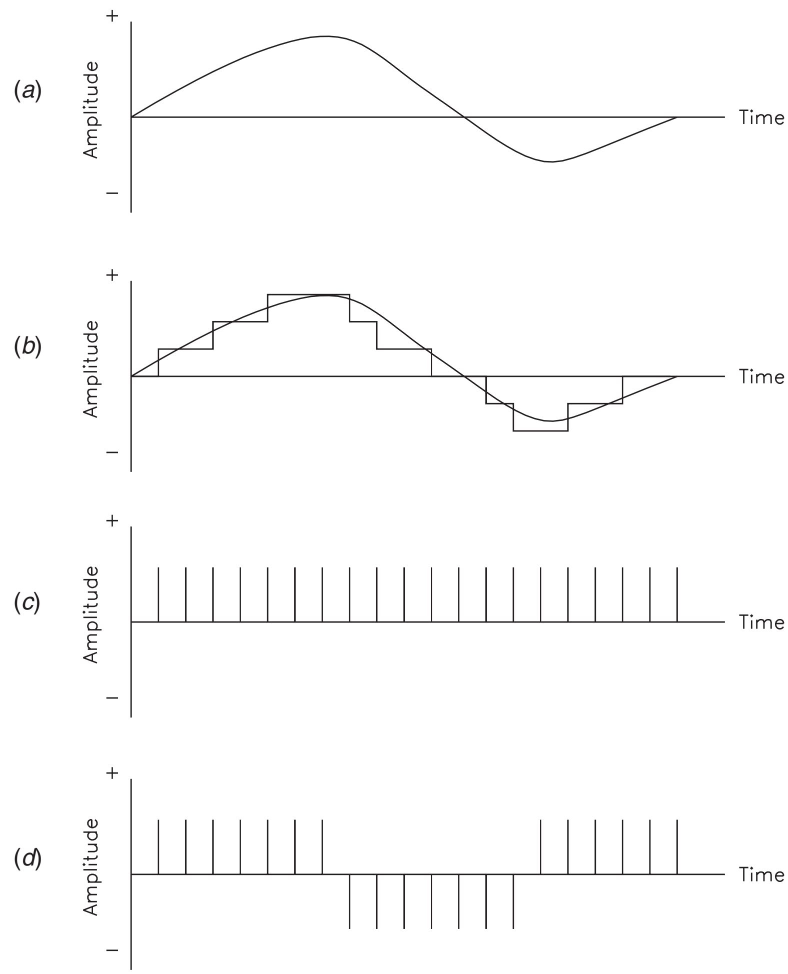


Figure 2.31 Delta modulation waveforms: (*a*) modulating signal, (*b*) quantized modulating signal, (*c*) pulse train, (*d*) resulting delta modulation waveform.

transmitted signals often are referred to as a *mark* (binary digit 1) or a *space* (binary digit 0). Figure 2.32 illustrates the transmitted waveform of a BFSK system.

- *Binary phase-shift keying* (BPSK), a modulating method in which the phase of the transmitted wave is shifted 180° in synchronism with the input digital signal. The phase of the RF carrier is shifted by π/2 radians or –π/2 radians, depending upon whether the data bit is a 0 or a 1. Figure 2.33 shows the BPSK transmitted waveform.

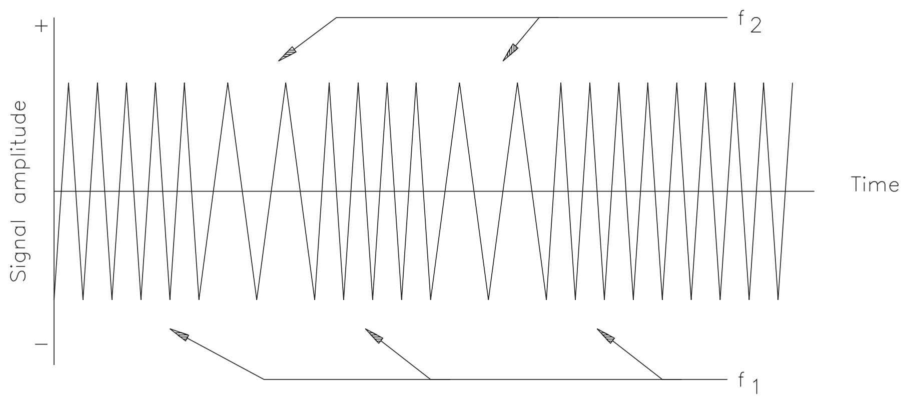


Figure 2.32 Binary FSK waveform.

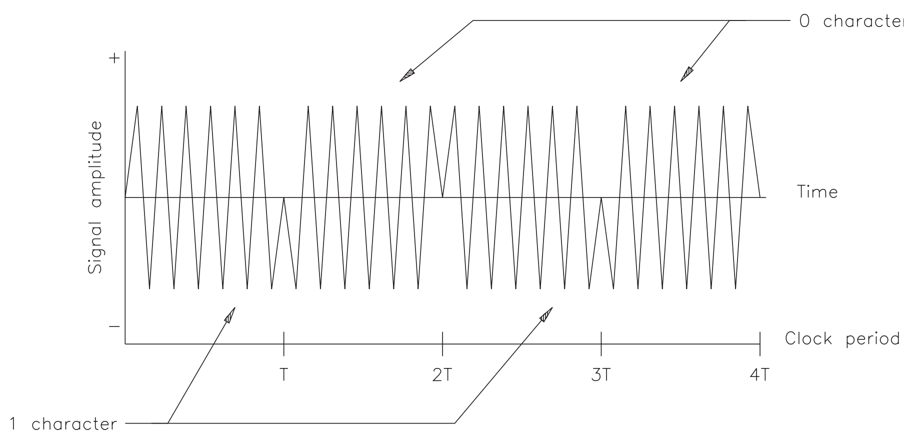


Figure 2.33 Binary PSK waveform.

- *Quadriphase-shift keying* (QPSK), a modulation scheme similar to BPSK except that quaternary modulation is employed, rather than binary modulation. QPSK requires half the bandwidth of BPSK for the same transmitted data rate.

2.4.7 Baseband Digital Pulse Modulation

After the input samples have been quantized, they are transmitted through a channel, received, and converted back to their approximate original form [8]. The format (modulation scheme) applied to the quantized samples is determined by a number of factors, not the least of which is the channel through which the signal passes. A number of different formats are possible and practical.

Several common digital modulation formats are shown in Figure 2.34. The first (*a*) is referred to as *non-return-to-zero* (NRZ) polar because the waveform does not return to zero during each signaling interval, but switches from $+V$ to $-V$, or vice versa, at the end of each signaling interval (NRZ unipolar uses the levels V and 0). On the other hand, the *unipolar return-to-zero* (RZ) format, shown in (*b*) returns to zero in each signaling interval. Because bandwidth is inversely proportional to pulse duration, it is apparent that RZ requires twice the bandwidth that NRZ does. Also, RZ has a nonzero dc component, whereas NRZ does not necessarily have a nonzero component (unless there are more 1s than 0s or vice versa). An advantage of RZ over NRZ is that a pulse transition is guaranteed in each signaling interval, whereas this is not the case for NRZ. Thus, in cases where there are long strings of 1s or 0s, it may be difficult to synchronize the receiver to the start and stop times of each pulse in NRZ-based systems. A very important modulation format from the standpoint of synchronization considerations is NRZ-*mark*, also known as *differential encoding*, where an initial reference bit is chosen and a subsequent 1 is encoded as a change from the reference and a 0 is encoded as no change. After the initial reference bit, the current bit serves as a reference for the next bit, and so on. An example of this modulation format is shown in (*c*).

Manchester is another baseband data modulation format that guarantees a transition in each signaling interval and does not have a dc component. Also known as *biphase* or *split phase*, this scheme is illustrated in (*d*). The format is produced by *OR*ing the data clock with an NRZ-formatted signal. The result is a + to – transition for a logic 1, and a – to + zero crossing for a logic 0.

A number of other data formats have been proposed and employed in the past, but further discussion is beyond the scope of this chapter.

2.4.8 Spread Spectrum Systems

As the name implies, a *spread spectrum* system requires a frequency range substantially greater than the basic information-bearing signal. Spread spectrum systems have some or all of the following properties:

- Low interference to other communications systems
- Ability to reject high levels of external interference
- Immunity to jamming by hostile forces
- Provision for secure communications paths
- Operability over multiple RF paths

Spread spectrum systems operate with an entirely different set of requirements than transmission systems discussed previously. Conventional modulation methods are designed to provide for the easiest possible reception and demodulation of the transmitted intelligence. The goals of spread spectrum systems, on the other hand, are secure and reliable communications that cannot be intercepted by unauthorized persons. The most common modulating and encoding techniques used in spread spectrum communications include:

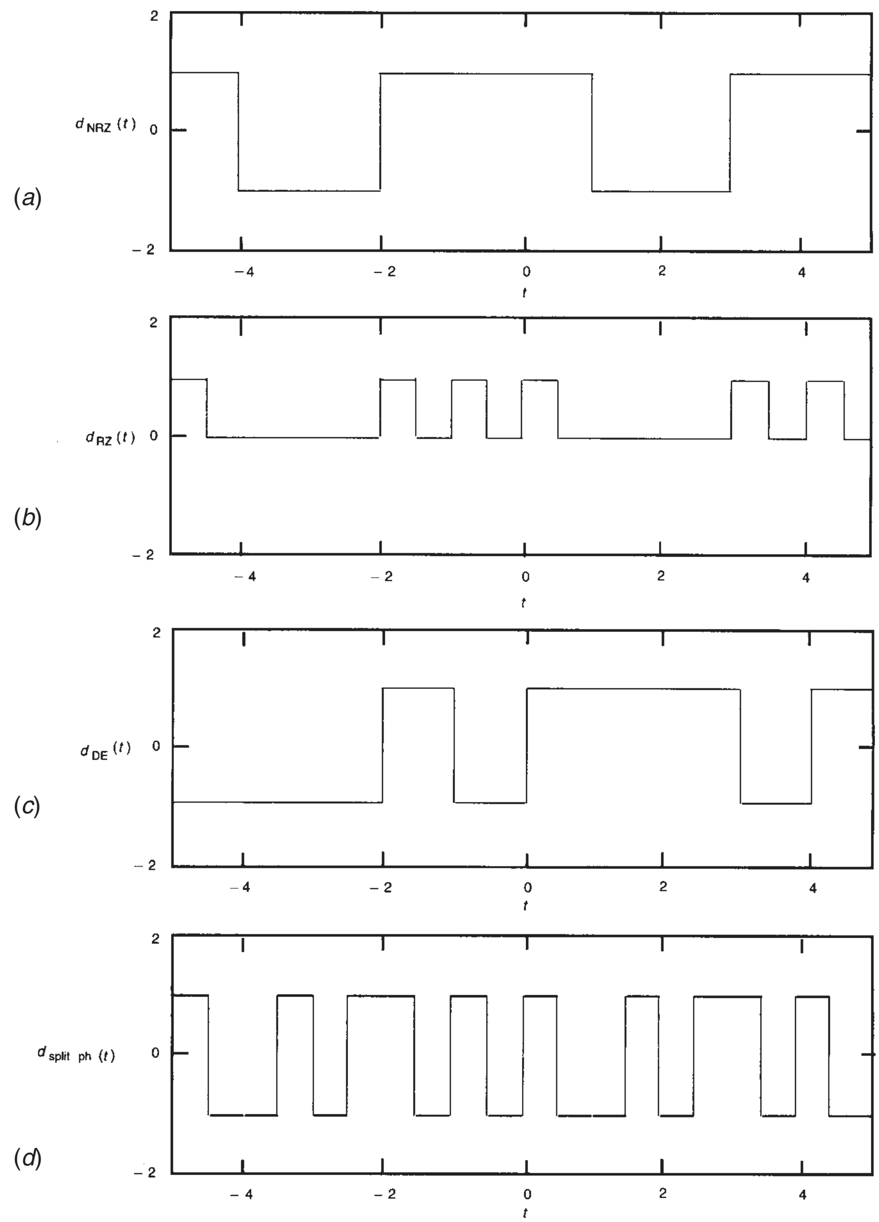


Figure 2.34 Various baseband modulation formats: (*a*) non-return-to zero, (*b*) unipolar return-to-zero, (*c*) differential encoded (NRZ-mark), (*d*) split phase. (*From* [8]. *Used with permission*.)

- *Frequency hopping*, where a random or *pseudorandom* number (PN) sequence is used to change the carrier frequency of the transmitter. This approach has two basic variations: *slow frequency hopping*, where the hopping rate is smaller than the data rate, and *fast frequency hopping*, where the hopping rate is larger than the data rate. In a fast frequency-hopping system, the transmission of a single piece of data occupies more than one frequency. Frequency-hopping systems permit multiple-access capability to a given band of frequencies because each transmitted signal occupies only a fraction of the total transmitted bandwidth.
- *Time hopping*, where a PN sequence is used to switch the position of a message-carrying pulse within a series of frames.
- *Message corruption*, where a PN sequence is added to the message before modulation.
- *Chirp spread spectrum*, where linear frequency modulation of the main carrier is used to spread the transmitted spectrum. This technique is commonly used in radar and also has been applied to communications systems.

In a spread spectrum system, the signal power is divided over a large bandwidth. The signal, therefore, has a small average power in any single narrowband slot. This means that a spread spectrum system can share a given frequency band with one or more narrowband systems. Furthermore, because of the low energy in any particular band, detection or interception of the transmission is difficult.

2.5 References

1. Tate, Jeffrey P., and Patricia F. Mead, "Crystal Oscillators," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, FL, pp. 185–199, 1996.
2. Frerking, M. E., *Crystal Oscillator Design and Temperature Compensation*, Van Nostrand Reinhold, New York, 1978.
3. Dye, D. W., *Proc. Phys. Soc.*, Vol. 38, pp. 399–457, 1926.
4. Douglass, Barry G., "Thermal Noise and Other Circuit Noise," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), pp. 30–36, 1996.
5. Terman, F. E., *Radio Engineering*, 3rd ed., McGraw-Hill, New York, pg. 468, 1947.
6. Kubichek, Robert, "Amplitude Modulation," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, FL, pp. 1175–1187, 1996.
7. Seymour, Ken, "Frequency Modulation," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, FL, pp. 1188–1200, 1996.
8. Ziemer, Rodger E., "Pulse Modulation," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, FL, pp. 1201–1212, 1996.

2.6 Bibliography

Benson, K. B., and Jerry. C. Whitaker (eds.), *Television Engineering Handbook*, McGraw-Hill, New York, 1986.

Benson, K. B., and Jerry. C. Whitaker, *Television and Audio Handbook for Technicians and Engineers*, McGraw-Hill, New York, 1989.
Crutchfield, E. B. (ed.), *NAB Engineering Handbook*, 8th Ed., National Association of Broadcasters, Washington, DC, 1991.
Fink, D., and D. Christiansen (eds.), *Electronics Engineers' Handbook*, 3rd Ed., McGraw-Hill, New York, 1989.
Fink, D., and D. Christiansen (eds.), *Electronics Engineers' Handbook*, 2nd Ed., McGraw-Hill, New York, 1982.
Hulick, Timothy P., "Using Tetrodes for High Power UHF," *Proceedings of the Society of Broadcast Engineers*, Vol. 4, SBE, Indianapolis, IN, pp. 52-57, 1989.
Jordan, Edward C., *Reference Data for Engineers: Radio, Electronics, Computer and Communications*, 7th Ed., Howard W. Sams, Indianapolis, IN, 1985.
Laboratory Staff, *The Care and Feeding of Power Grid Tubes*, Varian Eimac, San Carlos, CA, 1984.
Mendenhall, G. N., "Fine Tuning FM Final Stages," *Broadcast Engineering*, Intertec Publishing, Overland Park, KS, May 1987.
Whitaker, Jerry. C., *Maintaining Electronic Systems*, CRC Press, Boca Raton, FL, 1992.
Whitaker, Jerry. C., *Radio Frequency Transmission Systems: Design and Operation*, McGraw-Hill, New York, 1991.