1

Introduction

1.1 Radio Transmitters

Radio communication utilizes radio waves as a transmission and receiving medium. A *radio transmitter* consists of information source producing modulating signal, modulator, radio-frequency (RF) power amplifier, and antenna. A *power amplifier* is a circuit that increases the power level of a signal by using energy taken from a power supply [1–28]. Both the efficiency and distortion are critical parameters of power amplifiers. A *radio receiver* consists of an antenna, front end, demodulator, and audio amplifier. A transmitter and receiver combined into one electronic device is called a *transceiver*. A radio transmitter produces a strong RF current, which flows through an antenna. In turn, a transmitter antenna radiates electromagnetic waves (EMWs), called radio waves. Transmitters are used for communication of information over a distance, such as radio and television broadcasting, mobile phones, wireless computer networks, radio navigation, radio location, air traffic control, radars, ship communication, radio-frequency identifications (RFIDs), collision avoidance, speed measurement, weather forecasting, and so on. The information signal is the modulating signal, and it is usually in the form of audio signal from a microphone, video signal from a camera, or digital signal. Modern wireless communication systems include both amplitude-modulated (AM) and phase-modulated (PM) signals with a large peak-to-average ratio (PAR). Typically, the PAR is 6-9 dB for Wideband Code Division Multiple Access (WCDMA) and Orthogonal Frequency Division Multiplexing (OFDM). The main difficulty in transmitters' design is achieving a good linearity and a high efficiency.

An ideal radio transmitter should satisfy the following requirements:

- high efficiency,
- high linearity (i.e., low signal distortion),
- high power gain,
- large dynamic range,

RF Power Amplifiers, Second Edition. Marian K. Kazimierczuk.

- large slew rate,
- low noise level,
- high spectral efficiency,
- wide modulation bandwidth,
- capability of reproducing complex modulated waveforms,
- capability of transmitting high data rate communication,
- capability of transmitting a large diversity of waveforms, and
- portability.

A design of a transmitter with high efficiency and high linearity is a challenging problem. Linear amplification is required when the signal contains both amplitude and phase modulations. Nonlinearities cause imperfect reproduction of the amplified signal.

1.2 Batteries for Portable Electronics

In portable communications, batteries are used as power supplies. The most popular battery technologies are lithium (Li-ion) batteries and nickelcadmium (Ni-Cd) batteries. The nominal output voltage of the Li-ion batteries is 3.6 V. The discharge curve of Li-ion batteries is typically from 4 to 2 V during the period of 5 h of active operation. The nominal output voltage of the Ni-based batteries is 1.25 V. The discharge curve of these batteries is typically from 1.4 to 1 V during the period of 5 h of active operation. These are rechargeable batteries. The energy density of Li-ion batteries is nearly twice that of Ni-based batteries, yielding a smaller battery that stores the same amount of energy. However, the discharge curve of Li-ion batteries is much steeper than that of Ni-based batteries. The slope of the discharge curve of Li-ion batteries is approximately −0.25 V/h, whereas the slope of the discharge curve of Ni-based batteries is approximately −0.04 V/h. Therefore, Li-ion batteries may require a voltage regulator.

1.3 Block Diagram of RF Power Amplifiers

A power amplifier [1–27] is a key element to build a wireless communication system successfully. Its main purpose is to increase the power level of the signal. To minimize interferences and spectral regrowth, transmitters should be linear. A block diagram of an RF power amplifier is shown in Fig. 1.1. It consists of transistor (MOSFET, MESFET, HFET, or BJT), output network, input network, and RF choke. The trend is to replace silicon(Si)-based semiconductor devices with wide band gap (BG) semiconductor devices, such as silicon carbide (SiC) and gallium nitride (GaN) devices. Silicon carbide is also used as a substrate because it has high thermal conductivity, for example, for GaN devices. Gallium nitride semiconductor is used to make high electron mobility transistors (HEMTs). The energy BG of GaN is three times greater than that of silicon, yielding lower performance degradation at high temperatures. The breakdown electric field intensity is six times greater than that of silicon. Also, the carrier saturation velocity is 2.5 greater than that of silicon, resulting in a higher power density.

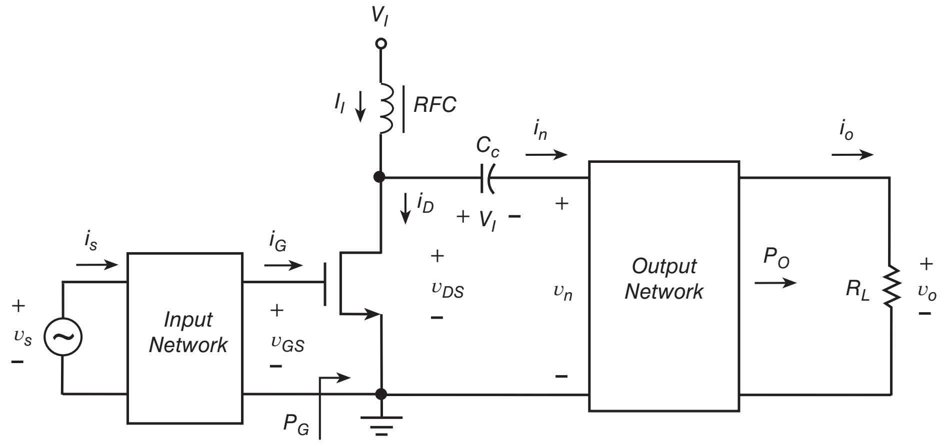


Figure 1.1 Block diagram of RF power amplifier.

In RF power amplifiers, a transistor can be operated

- as a dependent current source,
- as a switch, and
- in overdriven mode (partially as a dependent source and partially as a switch).

Figure 1.2(a) shows a model of an RF power amplifier in which the transistor is operated as a voltage- or current-dependent current source. When a MOSFET is operated as a dependent current source, the drain current waveform is determined by the gate-to-source voltage waveform and the transistor operating point. The drain voltage waveform is determined by the dependent current source and the load network impedance. When a MOSFET is operated as a switch, the switch voltage is nearly zero when the switch is ON and the drain current is determined by the external circuit due to the switching action of the transistor. When the switch is OFF, the switch current is zero and the switch voltage is determined by the external circuit response.

In order to operate the MOSFET as a dependent current source, the transistor cannot enter the ohmic region. It must be operated in the active region, also called the pinch-off region or the saturation region. Therefore, the drain-to-source voltage v_{DS} must be kept higher than the minimum value V_{DSmin}, that is, $v_{DS} > V_{DSmin} = V_{GS} - V_t$, where V_t is the transistor threshold voltage. When the transistor is operated as a dependent current source, the magnitudes of the drain current i_D and the drain-to-source voltage v_{DS} are nearly proportional to the magnitude of the gate-to-source voltage v_{GS}. Therefore, this type of operation is suitable for linear power amplifiers. Amplitude linearity is important for amplification of AM signals.

Figure 1.2(b) shows a model of an RF power amplifier in which the transistor is operated as a switch. To operate the MOSFET as a switch, the transistor cannot enter the active region. It must remain in the ohmic region when it is ON and in the cutoff region when it is OFF. To maintain the MOSFET in the ohmic region, it is required that $v_{DS} < V_{GS} - V_t$. If the gate-to-source voltage v_{GS} is increased at a given load impedance, the amplitude of the drain-to-source voltage v_{DS} will increase, causing the transistor to operate initially in the active region and then in the ohmic region. When the transistor is operated as a switch, the magnitudes of the drain current i_D and the

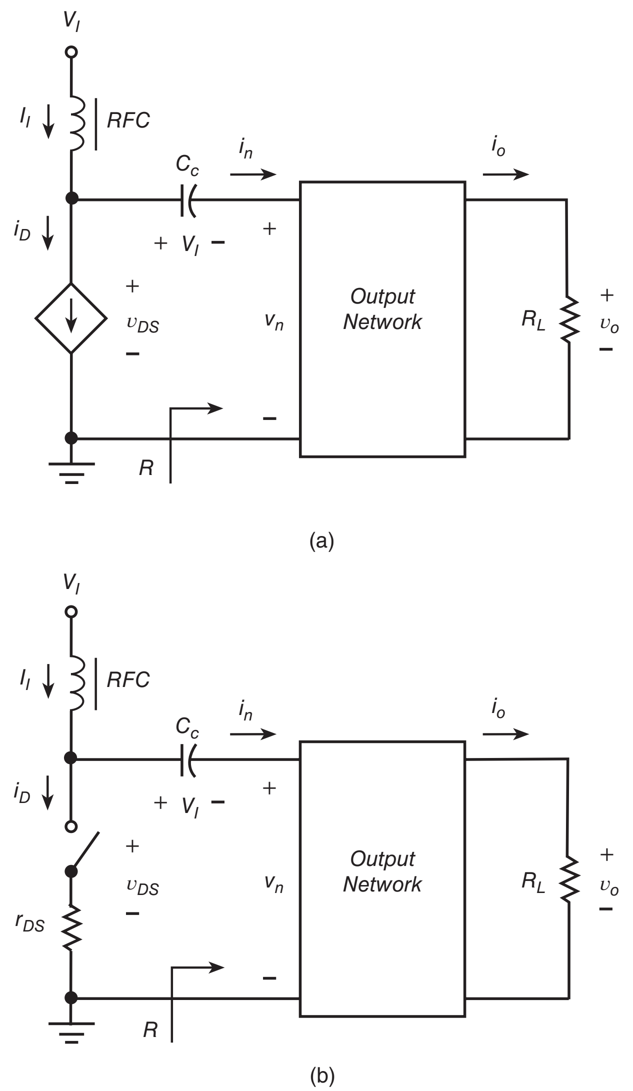


Figure 1.2 Models of operation of transistor in RF power amplifiers: (a) transistor as a dependent current source and (b) transistor as a switch.

drain-to-source voltage v_{DS} are independent of the magnitude of the gate-to-source voltage v_{GS}. In most applications, the transistor operated as a switch is driven by a rectangular gate-to-source voltage v_{GS}. A sinusoidal gate-to-source voltage V_{GS} is used to drive a transistor as a switch at very high frequencies, where it is difficult to generate rectangular voltages. The reason to use the transistors as switches is to achieve high amplifier efficiency. When the transistor conducts a high drain current i_D, the drain-to-source voltage v_{DS} is low, resulting in low power loss.

If the transistor is driven by a sinusoidal voltage v_{GS} of high amplitude, the transistor is overdriven. In this case, it operates in the active region when the instantaneous values of v_{GS} are low and as a switch when the instantaneous values of v_{GS} are high.

The main functions of the output network are as follows:

- Impedance transformation.
- Harmonic suppression.
- Filtering the spectrum of a signal with bandwidth BW to avoid interference with communication signals in adjacent channels.

Modulated signals can be divided into two categories:

- Variable-envelope signals, such as AM and SSB.
- Constant-envelope signals, such as FM, FSK, and CW.

Modern mobile communication systems usually contain both amplitude and phase modulations. Amplification of variable-envelope signals requires linear amplifiers. A linear amplifier is an electronic circuit whose output voltage is directly proportional to its input voltage. The Class A RF power amplifier is a nearly linear amplifier.

1.4 Classes of Operation of RF Power Amplifiers

The classification of RF power amplifiers with a transistor operated as a dependent current source is based on the conduction angle 2θ of the drain current i_D. Waveforms of the drain current i_D of a transistor operated as a dependent source in various classes of operation for sinusoidal gate-to-source voltage v_{GS} are shown in Fig. 1.3. The operating points for various classes of operation are shown in Fig. 1.4.

In Class A, the conduction angle 2θ of the drain current i_D is 360°. The gate-to-source voltage v_{GS} must be greater than the transistor threshold voltage V_t, that is, $v_{GS} > V_t$. This is accomplished by choosing the dc component of the gate-to-source voltage V_{GS} sufficiently greater than the threshold voltage of the transistor V_t such that $V_{GS} - V_{gsm} > V_t$, where V_{gsm} is the amplitude of the ac component of the gate-to-source voltage v_{GS}. The dc component of the drain current I_D must be greater than the amplitude of the ac component I_m of the drain current i_D. As a result, the transistor conducts during the entire cycle. Class A amplifiers are linear, but have low efficiency (lower than 50%).

In Class B, the conduction angle 2θ of the drain current i_D is 180°. The dc component V_{GS} of the gate-to-source voltage v_{GS} is equal to V_t, and the drain bias current I_D is zero. Therefore, the transistor conducts for only half of the cycle.

In Class AB, the conduction angle 2θ is between 180° and 360°. The dc component of the gate-to-source voltage V_{GS} is slightly above V_t, and the transistor is biased at a small drain current I_D. As the name suggests, Class AB is the intermediate class between Class A and Class B. Class AB amplifiers are linear, but have low efficiency (less than 50%).

In Class C, the conduction angle 2θ of the drain current i_D is less than 180°. The operating point is located in the cutoff region because $V_{GS} < V_t$. The drain bias current I_D is zero. The transistor conducts for an interval less than half of the cycle. Class C amplifiers are nonlinear and are only suitable for the amplification of constant-envelope signals, but have a higher efficiency than that Class A and AB amplifiers.

Class A, AB, and B operations are used in audio and RF power amplifiers, whereas Class C is used only in RF power amplifiers and industrial applications.

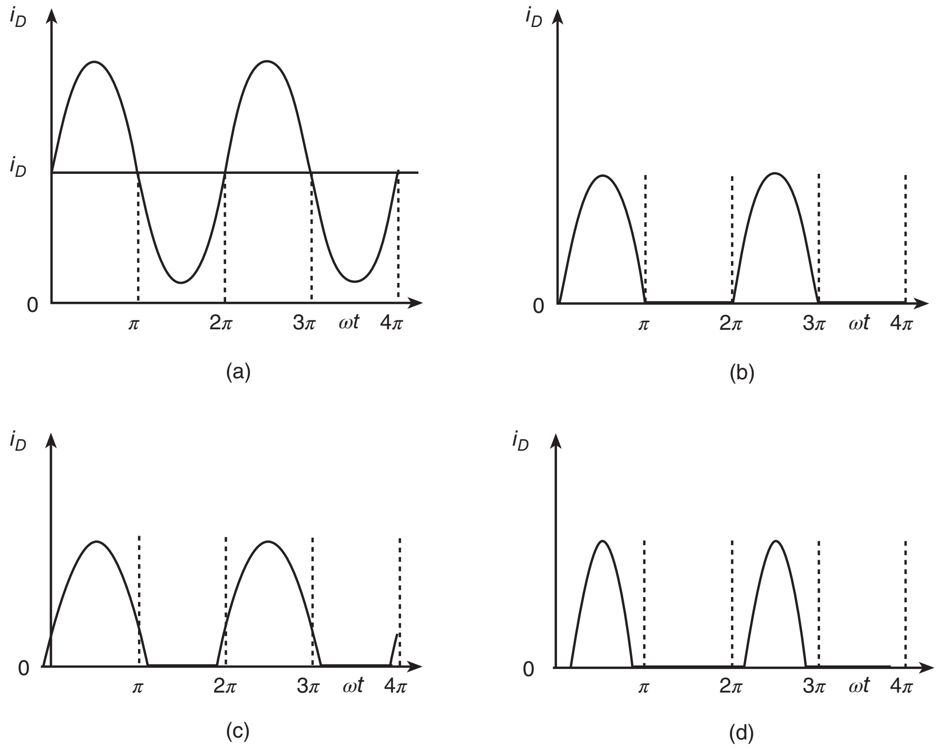


Figure 1.3 Waveforms of the drain current i_D in various classes of operation: (a) Class A, (b) Class B, (c) Class AB, and (d) Class C.

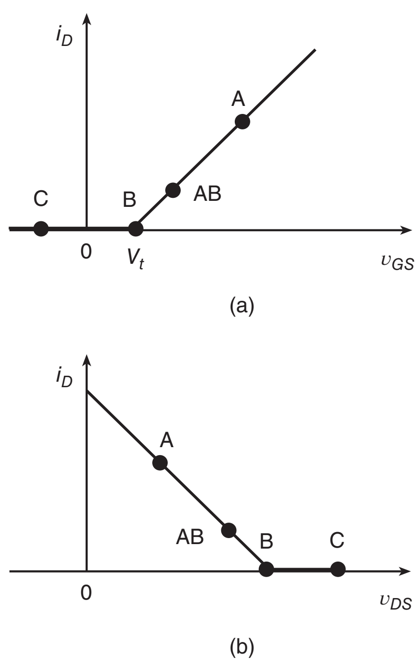


Figure 1.4 Operating points for Classes A, B, AB, and C.

The transistor is operated as a switch in Class D, E, and DE RF power amplifiers. In Class F, the transistor can be operated as either a dependent current source or a switch. RF power amplifiers are used in communications, power generation, and plasma generation.

1.5 Waveforms of RF Power Amplifiers

For steady state, the waveforms of an unmodulated power amplifier are periodic of frequency $f = \omega/(2\pi)$. The drain current waveform can be represented by Fourier series

$$i_D = I_I + \sum_{n=1}^{\infty} I_{mn}\cos(n\omega t + \phi_{in}) = I_{DM}\left[\alpha_0 + \sum_{n=1}^{\infty} \alpha_n \cos(n\omega t + \phi_{in})\right]$$
$$= I_I\left[1 + \sum_{n=1}^{\infty} \frac{I_{mn}}{I_I}\cos(n\omega t + \phi_{in})\right] = I_I\left[1 + \sum_{n=1}^{\infty} \gamma_n \cos(n\omega t + \phi_{in})\right] \tag{1.1}$$

where

$$\alpha_0 = \frac{I_I}{I_{DM}} \tag{1.2}$$

$$\alpha_n = \frac{I_{mn}}{I_{DM}} \tag{1.3}$$

and

$$\gamma_n = \frac{I_{mn}}{I_I} \tag{1.4}$$

The drain-to-source voltage waveform can also be expanded into Fourier series

$$v_{DS} = V_I - \sum_{n=1}^{\infty} V_{mn}\cos(n\omega t + \phi_{vn}) = V_{DSM}\left[\beta_0 - \sum_{n=1}^{\infty} \beta_n \cos(n\omega t + \phi_{vn})\right]$$
$$= V_I\left[1 - \sum_{n=1}^{\infty} \frac{V_{mn}}{V_I}\cos(n\omega t + \phi_{vn})\right] = V_I\left[1 - \sum_{n=1}^{\infty} \xi_n \cos(n\omega t + \phi_{vn})\right] \tag{1.5}$$

where

$$\beta_0 = \frac{V_I}{V_{DSM}} \tag{1.6}$$

$$\beta_1 = \frac{V_{mn}}{V_{DSM}} \tag{1.7}$$

and

$$\xi_n = \frac{V_{mn}}{V_I} \tag{1.8}$$

1.6 Parameters of RF Power Amplifiers

1.6.1 Drain Efficiency of RF Power Amplifiers

When the resonant frequency of the output network f_o is equal to the operating frequency f, the *drain power* (the power delivered by the drain to the output network) is given by

$$P_{DS} = \frac{1}{2}I_m V_m = \frac{1}{2}I_m^2 R = \frac{V_m^2}{2R} \quad \text{for} \quad f = f_o \tag{1.9}$$

where I_m is the amplitude of the fundamental component of the drain current i_D, V_m is the amplitude of the fundamental component of the drain-to-source voltage v_{DS}, and R is the input resistance of the output network at the fundamental frequency. If the resonant frequency f_o is not equal to the operating frequency f, the drain power of the fundamental component is given by

$$P_{DS} = \frac{1}{2} I_m V_m \cos\phi = \frac{1}{2} I_m^2 R \cos\phi = \frac{V_m^2 \cos\phi}{2R} \tag{1.10}$$

where ϕ is the phase shift between the fundamental components of the drain current and the drain-to-source voltage reduced by π.

The *instantaneous drain power dissipation* is

$$p_D(\omega t) = i_D v_{DS} \tag{1.11}$$

The *time-average drain power dissipation* for periodic waveforms is

$$P_D = \frac{1}{2\pi} \int_0^{2\pi} p_D \, d(\omega t) = \frac{1}{2\pi} \int_0^{2\pi} i_D v_{DS} \, d(\omega t) = P_I - P_{DS} \tag{1.12}$$

The dc supply current is

$$I_I = \frac{1}{2\pi} \int_0^{2\pi} i_D \, d(\omega t) \tag{1.13}$$

The dc supply power is

$$P_I = V_I I_I = \frac{V_I}{2\pi} \int_0^{2\pi} i_D \, d(\omega t) \tag{1.14}$$

The *drain efficiency* at a given drain power P_{DS} is

$$\eta_D = \frac{P_{DS}}{P_I} = \frac{P_I - P_D}{P_I} = 1 - \frac{P_D}{P_I} = 1 - \frac{\int_0^{2\pi} i_D v_{DS} \, d(\omega t)}{V_I \int_0^{2\pi} i_D \, d(\omega t)}$$
$$= \frac{1}{2} \left(\frac{I_m}{I_I}\right) \left(\frac{V_m}{I_I}\right) \cos\phi = \frac{1}{2} \gamma_1 \xi_1 \cos\phi \tag{1.15}$$

where $\phi = \phi_{i1} - \phi_{v1}$. When the operating frequency is equal to the resonant frequency $f = f_0$, the drain efficiency is

$$\eta_D = \frac{P_{DS}}{P_I} = \frac{1}{2} \left(\frac{I_m}{I_I}\right) \left(\frac{V_m}{V_I}\right) = \frac{1}{2} \gamma_1 \xi_1 \quad \text{for} \quad f = f_0 \tag{1.16}$$

Efficiency of power amplifiers is maximized by minimizing power dissipation at a desired output power.

For amplifiers in which the transistor is operated as a dependent current source, the highest drain efficiency usually occurs at the *peak envelope power* (PEP). Power amplifiers with time-varying amplitude have a time-varying drain efficiency $\eta_D(t)$. These amplifiers are usually operated below the maximum output power. This situation is called *power backoff*. The peak-to-average power ratio (PAPR) is the ratio of the PEP of the AM waveform PEP to the average envelope power for a long time interval $P_{O(AV)}$

$$PAPR = \frac{\text{Peak Power}}{\text{Average Power}} = \frac{PEP}{P_{O(AV)}} = 10 \log \left(\frac{PEP}{P_{O(AV)}}\right) \text{ (dB)} \tag{1.17}$$

The *power dynamic range* is the ratio of the largest output power P_{Omax} to the lowest output power P_{Omin} defined as

$$DNR = \frac{\text{Maximum Output Power}}{\text{Minimum Output Power}} = \frac{P_{Omax}}{P_{Omin}} = 10 \log \left(\frac{P_{Omax}}{P_{Omin}}\right) \text{ (dB).} \tag{1.18}$$

The *output power* delivered to a resistive load is

$$P_O = \frac{1}{2} I_{om} V_{om} = \frac{1}{2} I_{om}^2 R_L = \frac{V_{om}^2}{2R_L} \tag{1.19}$$

where I_{om} is the amplitude of the output current and V_{om} is the amplitude of the output voltage. The *power loss in the resonant output network* is

$$P_r = P_{DS} - P_O \tag{1.20}$$

The *efficiency of the resonant output network* is

$$\eta_r = \frac{P_O}{P_{DS}} \tag{1.21}$$

The *overall power loss* on the output side of the amplifier (in the transistor(s) and the output network) is

$$P_{Loss} = P_I - P_O = P_D + P_r \tag{1.22}$$

The *efficiency* of the amplifier at a specific output power is

$$\eta = \frac{P_O}{P_I} = \frac{P_O}{P_{DS}} \frac{P_{DS}}{P_I} = \eta_D \eta_r \tag{1.23}$$

The output power level of an amplifier is often referenced to the power level of 1 mW and is expressed as

$$P = 10 \log \frac{P(\mathrm{W})}{0.001} \ (\mathrm{dBm}) = 10 \log P(\mathrm{W}) - 10 \log 0.001 = [10 \log P(\mathrm{W}) + 30] \ (\mathrm{dBm}) \tag{1.24}$$

A dBm or dBW value represents an actual power, whereas a dB value represents a ratio of power, such as the power gain.

1.6.2 Statistical Characterization of Transmitter Average Efficiency

The output power of radio transmitters is a random variable. The average efficiency of a transmitter depends on the statistics of transmitter output power. The statistics is determined by the probability density function (PDF) (or the probability distribution function) of the output power and the dc supply power of a power amplifier. The average output power over a long-time interval $\Delta t = t_2 - t_1$ is

$$P_{O(AV)} = \frac{1}{\Delta t} \int_0^{\Delta t} PDF_{P_O}(t) P_O(t) dt \tag{1.25}$$

and the average supply power over the same time interval $\Delta T = t_2 - t_1$ is

$$P_{I(AV)} = \frac{1}{\Delta t} \int_0^{\Delta t} PDF_{P_I}(t) P_I(t) dt \tag{1.26}$$

Hence, the *long-term average efficiency* is the ratio of the energy delivered to the load (or antenna) E_O to the energy drawn from the power supply E_I over a long period of time Δt

$$\eta_{AV} = \frac{E_O}{E_I} = \frac{E_O/\Delta t}{E_I/\Delta t} = \frac{P_{O(AV)}}{P_{I(AV)}} = \frac{\frac{1}{\Delta t} \int_0^{\Delta t} PDF_{P_O}(t) P_O(t) dt}{\frac{1}{\Delta t} \int_0^{\Delta t} PDF_{P_I}(t) P_I(t) dt} \tag{1.27}$$

This efficiency determines the battery lifetime.

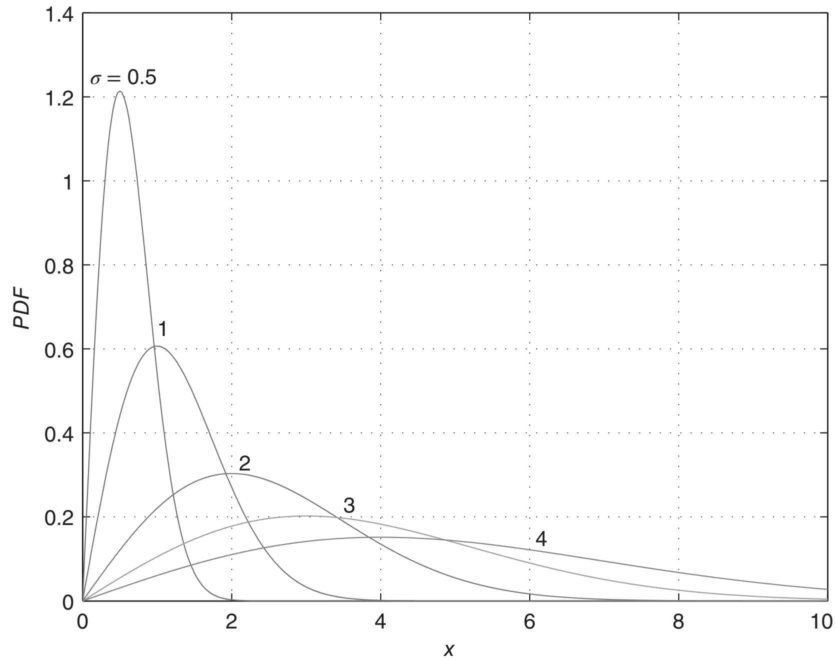


Figure 1.5 Rayleigh's probability density function of the transmitter output power.

The efficiency of power amplifiers in which transistors are operated as dependent current sources increases with the amplitude of the output voltage V_m. It reaches the maximum value at the maximum amplitude of the output voltage, which corresponds to the maximum output power. In practice, power amplifiers with a variable-envelope voltage are usually operated below the maximum output power. For example, the drain efficiency of the Class B power amplifier is $\eta_D = \pi/4 = 78.5\%$ at $V_m = V_I$, but it decreases to $\eta_D = \pi/8 = 39.27\%$ at $V_m = V_I/2$ and to $\eta_D = \pi/16 = 19.63\%$ at $V_m = V_I/4$. The average efficiency is useful for describing the efficiency of radio transmitters with variable-envelope signals, such as AM signals. The probability density function (PDF) of the envelope determines the amount of time an envelope remains at various amplitudes. For multiple carrier transmitters, the PDF may be characterized by Rayleigh's probability distribution.

Rayleigh's PDF of the output power is given by

$$g(P_O) = \frac{P_O(t)}{\sigma^2} e^{-\frac{P_O(t)}{2\sigma^2}} \tag{1.28}$$

where σ is the scale parameter of the distribution. Figure 1.5 shows plots of Rayleigh's PDF for $\sigma = 0.5$, 1, 2, 3, and 4.

1.6.3 Gate-Drive Power

The input impedance of the MOSFET consists of the series combination of the gate resistance r_G and the input capacitance C_i. The input capacitance is given by

$$C_i = C_{gs} + C_{gd}(1 - A_v) \tag{1.29}$$

where C_{gs} is the gate-to-source capacitance, C_{gd} is the gate-to-drain capacitance, and A_v is the voltage gain during the time interval when the drain-to-source voltage v_{DS} decreases. The Miller's capacitance is $C_m = C_{gd}(1 - A_v)$.

The *gate-drive power* is

$$P_G = \frac{1}{2\pi}\int_0^{2\pi} i_G v_{GS}\, d(\omega t) \tag{1.30}$$

For sinusoidal gate current and voltage waveforms, the gate-drive power is

$$P_G = \frac{1}{2} I_{gm} V_{gsm} \cos\phi_G = \frac{r_G I_{gm}^2}{2} \tag{1.31}$$

where I_{gm} is the amplitude of the gate current, V_{gsm} is the amplitude of the gate-to-source voltage, r_G is the gate resistance, and ϕ_G is the phase shift between the fundamental components of the gate current and the gate-to-source voltage. The *total power loss* including the gate-drive power is

$$P_{LS} = P_D + P_r + P_G \tag{1.32}$$

1.6.4 Power-Added Efficiency

The *power gain* of a power amplifier is given by

$$k_p = \frac{P_O}{P_G} = 10\log\left(\frac{P_O}{P_G}\right)\ \text{(dB)} \tag{1.33}$$

The *power-added efficiency* is the ratio of the difference between the output power and the gate-drive power to the dc supply power

$$\begin{aligned}\eta_{PAE} &= \frac{\text{Output Power} - \text{Drive Power}}{\text{DC Supply Power}} = \frac{P_O - P_G}{P_I} = \frac{P_O}{P_I}\left(1 - \frac{P_G}{P_O}\right)\\ &= \frac{P_O}{P_I}\left(1 - \frac{1}{P_O/P_G}\right) = \frac{P_O}{P_I}\left(1 - \frac{1}{k_p}\right) = \eta\left(1 - \frac{1}{k_p}\right)\end{aligned} \tag{1.34}$$

If $k_p = 1$, $\eta_{PAE} = 0$. If $k_p \gg 1$, $\eta_{PAE} \approx \eta$.

For many communication systems, various modulation techniques use variable-envelope voltage and have a very high PAR of the RF output power. Typically, an RF power amplifier achieves a maximum power efficiency at a single operating voltage corresponding to the peak output power. The PAR is usually from 3 to 6 dB for a single-carrier transmitters. For multi-carrier transmitters, the PAR is typically from 6 to 13 dB. The efficiency decreases rapidly as the power is reduced from its maximum value. The *average composite power-added efficiency* is defined as

$$\eta_{AV(PAE)} = \frac{P_{ORF(AV)} - P_{IRF(AV)}}{P_{DC(AV)} + P_{DC(mod)}} = \frac{\displaystyle\int_{V_{min}}^{V_{max}} PDF(V)^*[P_{ORF}(V) - P_{IRF}(V)]dV}{\displaystyle\int_{V_{min}}^{V_{max}} PDF(V)^*[P_{DC(PA)}(V) + P_{DC(mod)}(V)]dV} \tag{1.35}$$

where $PDF(V)^*$ is the PDF of the complex modulated signal, V_{min} and V_{max} are the minimum and maximum voltages of the RF envelope, and $P_{ORF}(V)$, $P_{IRF}(V)$, $P_{DC(PA)}(V)$, and $P_{DC(mod)}(V)$ are all instantaneous power values at a given envelope voltage V.

The overall efficiency of a power amplifier is defined as

$$\eta_{tot} = \frac{P_O}{P_{I(DC)} + P_G + P_{mod}} = \frac{P_O}{P_O + P_O + P_G + P_{mod}} \tag{1.36}$$

where P_{mod} is the power consumption of a modulator.

1.6.5 Output-Power Capability

The *output-power capability* of the RF power amplifier with N transistors is defined as

$$c_p = \frac{P_{Omax}}{NI_{DM}V_{DSM}} = \frac{\eta_{Dmax}P_I}{NI_{DM}V_{DSM}} = \frac{\eta_{Dmax}}{N}\left(\frac{I_I}{I_{DM}}\right)\left(\frac{V_I}{V_{DSM}}\right) = \frac{1}{N}\eta_{Dmax}\alpha_0\beta_0$$
$$= \frac{1}{2N}\left(\frac{I_m}{I_{DM}}\right)\left(\frac{V_m}{V_{DSM}}\right) = \frac{1}{2N}\alpha_{1max}\beta_{1max} \tag{1.37}$$

where I_{DM} is the maximum value of the instantaneous drain current i_D, V_{DSM} is the maximum value of the instantaneous drain-to-source voltage v_{DS}, η_{Dmax} is the amplifier drain efficiency at the maximum output power $P_{O(max)}$, and N is the number of transistors in the amplifier, which are not connected in parallel or in series. For example, a push–pull amplifier has two transistors. The maximum output power of an amplifier with a transistor having the maximum ratings I_{DM} and V_{DSM} is

$$P_{Omax} = c_p N I_{DM} V_{DSM} \tag{1.38}$$

As the output power capability c_p increases, the maximum output power P_{Omax} also increases. The output power capability is useful for comparing different types or families of amplifiers. The larger the c_p, the larger is the maximum output power.

For a single-transistor amplifier, the output-power capability is given by

$$c_p = \frac{P_{Omax}}{I_{DM}V_{DSM}} = \frac{\eta_{Dmax}P_I}{I_{DM}V_{DSM}} = \eta_{Dmax}\left(\frac{I_I}{I_{DM}}\right)\left(\frac{V_I}{V_{DSM}}\right) = \eta_{Dmax}\alpha_0\beta_0$$
$$= \frac{1}{2}\left(\frac{I_{m(max)}}{I_{DM}}\right)\left(\frac{V_{m(max)}}{V_{DSM}}\right) = \frac{1}{2}\alpha_{1max}\beta_{1max} \tag{1.39}$$

1.7 Transmitter Noise

A transmitter contains an oscillator of a carrier frequency. An oscillator is a nonlinear device. It does not generate an ideal single-frequency and constant-amplitude signal. Therefore, the oscillator output power is not only concentrated at a single frequency but also distributed around it. The noise spectra on both sides of the carrier are called noise sidebands. Hence, the voltage and current waveforms contain noise. These waveforms are modulated by noise. There are three categories of noise: AM noise, frequency-modulated (FM) noise, and phase noise. The AM noise results in the amplitude variations of the oscillator output voltage. The FM or PM noise causes the spreading of the frequency spectrum around the carrier frequency. The ratio of a single-sideband noise power contained in 1-Hz bandwidth at an offset from carrier to the carrier power is defined as noise-to-carrier power ratio

$$NCP = \frac{N}{C} = 10\log\left(\frac{N}{C}\right)\left(\frac{\mathrm{dBc}}{\mathrm{Hz}}\right) \tag{1.40}$$

The unit dBc/Hz indicates the number of decibels below the carrier over a bandwidth of 1 Hz. Most of oscillator noise around the carrier is the phase noise. This noise represents the phase jitter. For example, the phase noise is 80 dBc/Hz at 2 kHz offset from the carrier and 110 dBc/Hz at 50 kHz offset from the carrier.

Typically, the output thermal noise of power amplifiers should be below −130 dBm. The purpose of this requirement is to introduce negligible level of noise to the input of the low-noise amplifier (LNA) of the receiver.

Example 1.1

An RF power amplifier has $P_O = 10\,\text{W}$, $P_I = 20\,\text{W}$, and $P_G = 1\,\text{W}$. Find the efficiency, power-added efficiency, and power gain.

Solution. The efficiency of the power amplifier is

$$\eta = \frac{P_O}{P_I} = \frac{10}{20} = 50\% \tag{1.41}$$

The power-added efficiency is

$$\eta_{PAE} = \frac{P_O - P_G}{P_I} = \frac{10 - 1}{20} = 45\% \tag{1.42}$$

The power gain is

$$k_p = \frac{P_O}{P_G} = \frac{10}{1} = 10 = 10\log(10) = 10\text{ dB} \tag{1.43}$$

1.8 Conditions for 100% Efficiency of Power Amplifiers

The drain efficiency of any power amplifier is given by

$$\eta_D = \frac{P_{DS}}{P_I} = 1 - \frac{P_D}{P_I} \tag{1.44}$$

The condition for achieving a drain efficiency of 100% is

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS}\, dt = 0 \tag{1.45}$$

For an NMOS transistor, $i_D \ge 0$ and $v_{DS} \ge 0$; for a PMOS transistor, $i_D \le 0$ and $v_{DS} \le 0$. In this case, the sufficient condition for achieving a drain efficiency of 100% becomes

$$i_D v_{DS} = 0 \tag{1.46}$$

Thus, the waveforms i_D and v_{DS} should be nonoverlapping for an efficiency of 100%. Nonoverlapping waveforms i_D and v_{DS} are shown in Fig. 1.6.

The drain efficiency of power amplifiers is less than 100% for the following cases:

- The waveforms of $i_D > 0$ and $v_{DS} > 0$ are overlapping (e.g., as in a Class C power amplifier).
- The waveforms of i_D and v_{DS} are adjacent, and the waveform v_{DS} has a jump at $t = t_o$ and the waveform i_D contains an impulse Dirac function, as shown in Fig. 1.7(a).
- The waveforms of i_D and v_{DS} are adjacent, and the waveform i_D has a jump at $t = t_o$ and the waveform v_{DS} contains an impulse Dirac function, as shown in Fig. 1.7(b).

For the case of Fig. 1.7(a), an ideal switch is connected in parallel with a capacitor C. The switch is turned on at $t = t_o$, when the voltage v_{DS} across the switch is nonzero. At $t = t_o$, this voltage can be described by

$$v_{DS}(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} v_{DS}(t) + \lim_{t\to t_o^+} v_{DS}(t)\right] = \frac{1}{2}(\Delta V + 0) = \frac{\Delta V}{2} \tag{1.47}$$

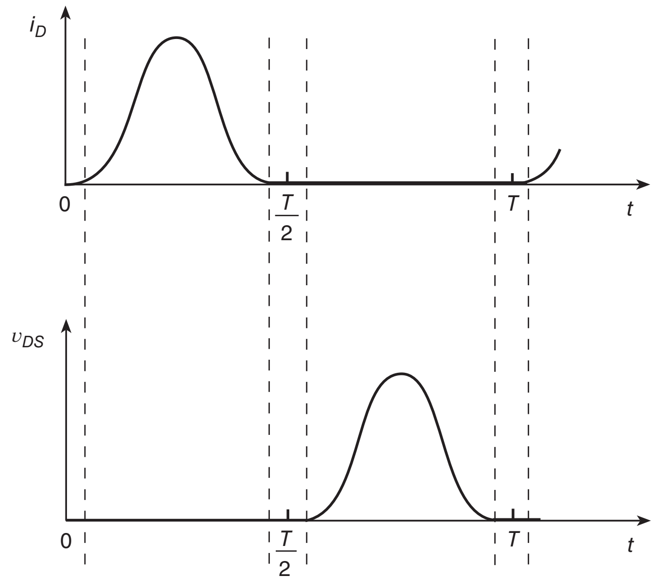


Figure 1.6 Nonoverlapping waveforms of drain current i_D and drain-to-source voltage v_{DS}.

At $t = t_o$, the drain current is given by

$$i_D(t_o) = C\Delta V\delta(t - t_o) \tag{1.48}$$

Hence, the instantaneous power dissipation is

$$p_D(t) = i_D v_{DS} = \begin{cases} \frac{1}{2}C\Delta V^2\delta(t - t_o) & \text{for } t = t_o \\ 0 & \text{for } t \neq t_o \end{cases} \tag{1.49}$$

resulting in the time average power dissipation

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS} dt = \frac{C\Delta V^2}{2T}\int_0^T \delta(t - t_o)dt = \frac{1}{2}fC\Delta V^2 \tag{1.50}$$

and the drain efficiency

$$\eta_D = 1 - \frac{P_D}{P_I} = 1 - \frac{fC\Delta V^2}{2P_I} \tag{1.51}$$

In a real circuit, the switch has a small series resistance, and the current through the switch is an exponential function of time with a finite peak value. Thus, to achieve the efficiency of 100%, either $\Delta V = 0$ or $C = 0$. In a realistic amplifier, the transistor should be turned on at zero drain-to-source voltage v_{DS} so that $\Delta V = 0$. This observation leads to the concept of a zero-voltage switching (ZVS) Class E amplifier.

Example 1.2

An RF power amplifier has a step change in the drain-to-source voltage at MOSFET turn-on $V_{DS} = 5$ V, transistor capacitance $C = 100$ pF, operating frequency $f = 2.4$ GHz, dc supply voltage $V_I = 5$ V, and dc supply current $I_I = 1$ A. Assume that all parasitic resistances are zero. Find the efficiency of the power amplifier.

Solution. The switching power loss is

$$P_D = \frac{1}{2} f C \Delta V_{DS}^2 = \frac{1}{2} \times 2.4 \times 10^9 \times 100 \times 10^{-12} \times 5^2 = 3 \text{ W} \tag{1.52}$$

The dc power loss is

$$P_I = I_I V_I = 1 \times 5 = 5 \text{ W} \tag{1.53}$$

Hence, the drain efficiency of the amplifier is

$$\eta_D = \frac{P_O}{P_I} = \frac{P_I - P_D}{P_I} = 1 - \frac{P_D}{P_I} = 1 - \frac{3}{5} = 40\% \tag{1.54}$$

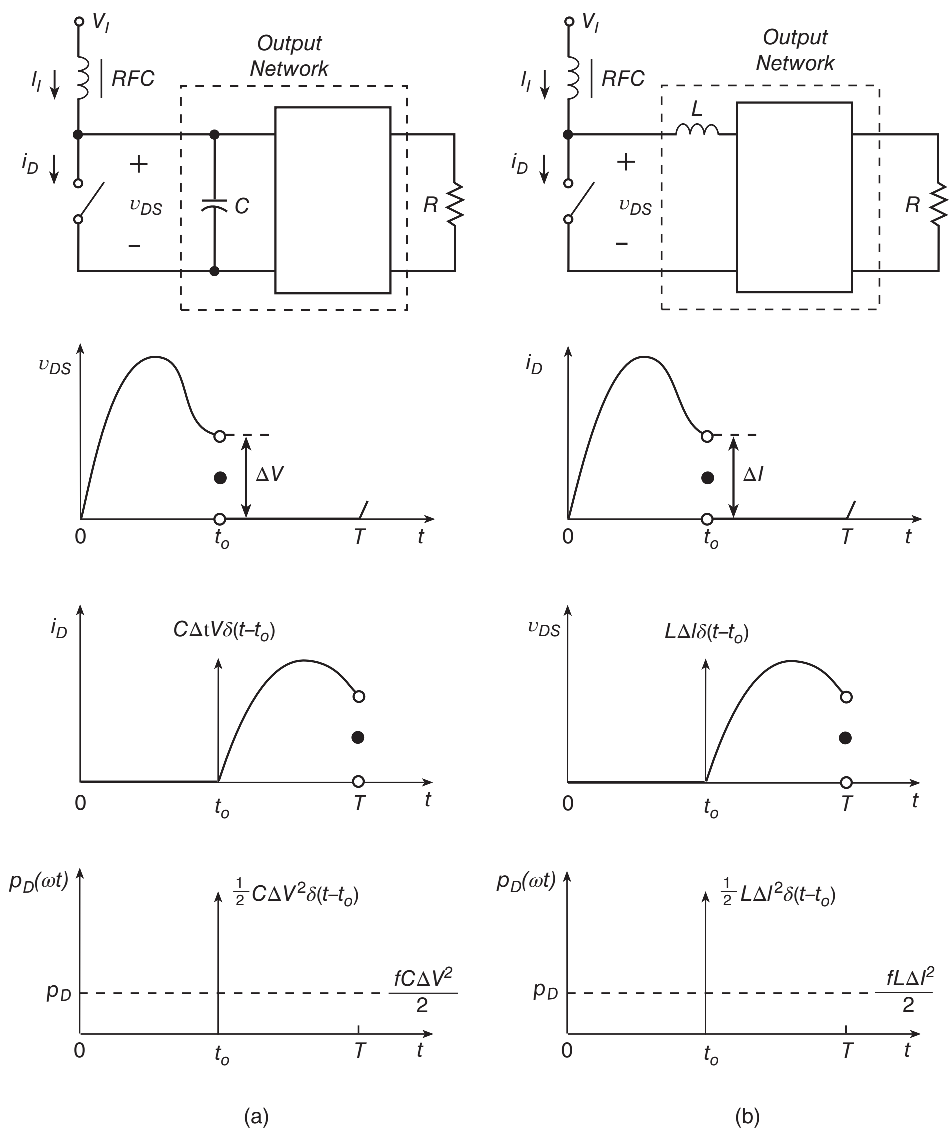


Figure 1.7 Waveforms of drain current i_D and drain-to-source voltage v_{DS} with Dirac delta functions: (a) circuit with the switch in parallel with a capacitor; (b) circuit with the switch in series with an inductor.

For the amplifier of Fig. 1.7(b), an ideal switch is connected in series with an inductor L. The switch is turned on at $t = t_o$, when the current i_D through the switch is nonzero. At $t = t_o$, the switch current can be described by

$$i_D(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} i_D(t) + \lim_{t\to t_o^+} i_D(t)\right] = \frac{1}{2}(\Delta I + 0) = \frac{\Delta I}{2} \tag{1.55}$$

At $t = t_o$, the drain current is given by

$$i_D(t_o) = L\Delta I\delta(t - t_o) \tag{1.56}$$

Hence, the instantaneous power dissipation is

$$p_D(t) = i_D v_{DS} = \begin{cases} \frac{1}{2}L\Delta I^2\delta(t - t_o) & \text{for } t = t_o \\ 0 & \text{for } t \neq t_o \end{cases} \tag{1.57}$$

resulting in the time average power dissipation

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS}\, dt = \frac{1}{2} fL\Delta I^2 \tag{1.58}$$

and the drain efficiency

$$\eta_D = 1 - \frac{P_D}{P_I} = 1 - \frac{fL\Delta I^2}{2P_I} \tag{1.59}$$

In reality, the switch in the off-state has a large parallel resistance and a voltage with a finite peak value developed across the switch. The efficiency of 100% can be achieved if either $\Delta I = 0$ or $L = 0$. This leads to the concept of zero-current switching (ZCS) Class E amplifier [3].

1.9 Conditions for Nonzero Output Power at 100% Efficiency of Power Amplifiers

The drain current and drain-to-source voltage waveforms have fundamental limitations for simultaneously achieving 100% efficiency and $P_O > 0$ [13, 14]. The drain current i_D and the drain-to-source voltage v_{DS} can be represented by the Fourier series as

$$i_D = I_I + \sum_{n=1}^{\infty} i_{dsn} = I_I + \sum_{n=1}^{\infty} I_{dn}\sin(n\omega t + \psi_n) \tag{1.60}$$

and

$$v_{DS} = V_I + \sum_{n=1}^{\infty} v_{dsn} = V_I + \sum_{n=1}^{\infty} V_{dsn}\sin(n\omega t + \vartheta_n) \tag{1.61}$$

The derivatives of these waveforms with respect to time are

$$i'_D = \frac{di_D}{dt} = \omega\sum_{n=1}^{\infty} nI_{dn}\cos(n\omega t + \psi_n) \tag{1.62}$$

and

$$v'_{DS} = \frac{dv_{DS}}{dt} = \omega\sum_{n=1}^{\infty} nV_{dsn}\cos(n\omega t + \vartheta_n) \tag{1.63}$$

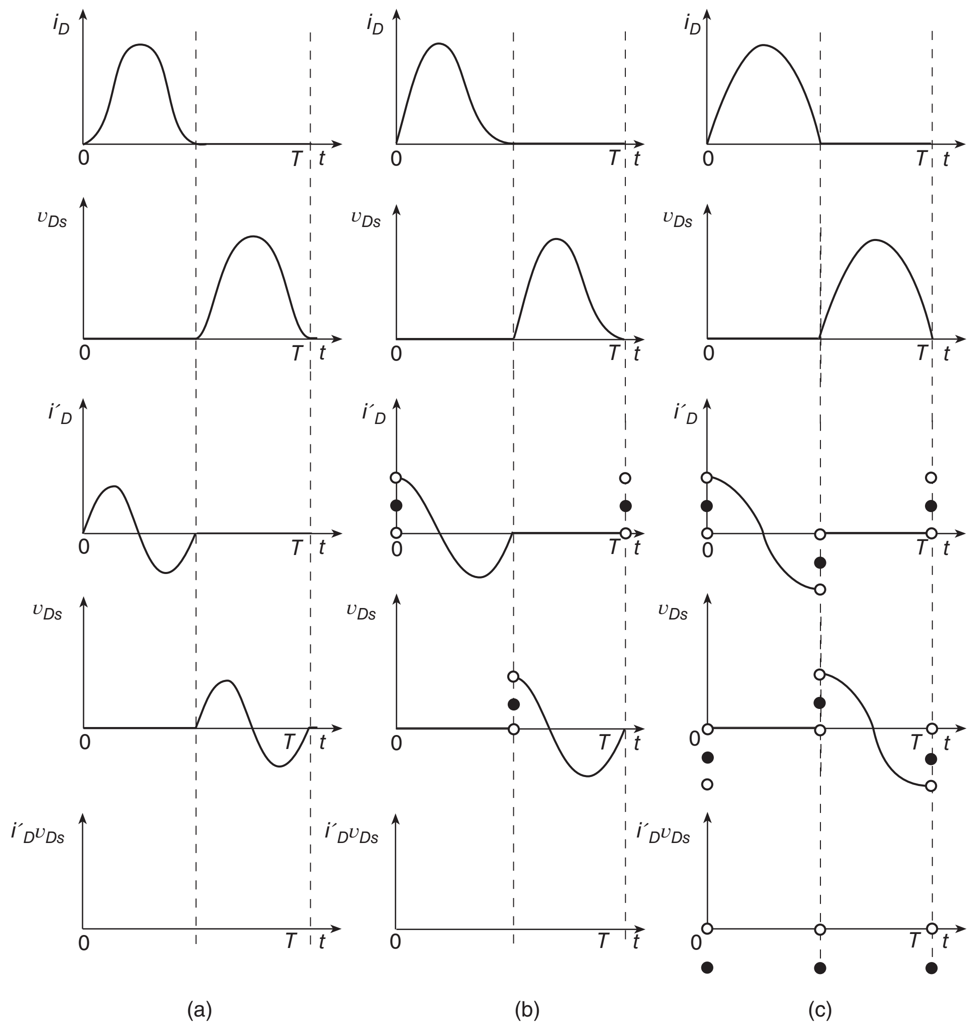


Figure 1.8 Waveforms of power amplifiers with $P_O = 0$.

Hence, the time-average value of the product of the derivatives is

$$\frac{1}{T}\int_0^T i'_D \upsilon'_{DS}\, dt = \omega^2 \int_0^T \sum_{n=1}^{\infty} nI_{dn}\cos(n\omega t + \psi_n) \sum_{n=1}^{\infty} nV_{dsn}\cos(n\omega t + \vartheta_n dt)$$

$$= -\frac{\omega^2}{2}\sum_{n=1}^{\infty} n^2 I_{dn} V_{dsn}\cos\phi_n = -\omega^2 \sum_{n=1}^{\infty} n^2 P_{dsn} \tag{1.64}$$

where $\phi_n = \vartheta_n - \psi_n - \pi$ Next,

$$\sum_{n=1}^{\infty} n^2 P_{dsn} = -\frac{1}{4\pi^2 f}\int_0^T i'_D \upsilon'_{DS}\, dt \tag{1.65}$$

If the efficiency of the output network is $\eta_n = 1$ and the power at harmonic frequencies is zero, that is, $P_{ds2} = 0$, $P_{ds3} = 0$, …, then

$$P_{ds1} = P_{O1} = -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\, dt \tag{1.66}$$

For multipliers, if $\eta_n = 1$ and the power at the fundamental frequency and at harmonic frequencies is zero except that of the nth harmonic frequency, then the power at the nth harmonic frequency is

$$P_{dsn} = P_{On} = -\frac{1}{4\pi^2 n^2}\int_0^T i'_D v'_{DS}\, dt \tag{1.67}$$

If the output network is passive and linear, then

$$-\frac{\pi}{2} \le \phi_n \le \frac{\pi}{2} \tag{1.68}$$

In this case, the output power is nonzero

$$P_O > 0 \tag{1.69}$$

if

$$\frac{1}{T}\int_0^T i'_D v'_{DS}\, dt < 0 \tag{1.70}$$

If the output network and the load are passive and linear and

$$\frac{1}{T}\int_0^T i'_D v'_{DS}\, dt = 0 \tag{1.71}$$

then

$$P_O = 0 \tag{1.72}$$

for the following cases:

- The waveforms i_D and v_{DS} are nonoverlapping, as shown in Fig. 1.6.
- The waveforms i_D and v_{DS} are adjacent and the derivatives at the joint time instants t_j are $i'_D(t_j) = 0$ and $v'_{DS}(t_j) = 0$, as shown in Fig. 1.8(a).
- The waveforms i_D and v_{DS} are adjacent and the derivative $i'_D(t_j)$ at the joint time instant t_j has a jump and $v'_{DS}(t_j) = 0$, or *vice versa*, as shown in Fig. 1.8(b).
- The waveforms i_D and v_{DS} are adjacent, and the derivatives of both waveforms $i'_D(t_j)$ and $v'_{DS}(t_j)$ have jumps at the joint time instant t_j, as shown in Fig. 1.8(c).

In summary, ZVS, zero-voltage derivative switching (ZVDS), and ZCS conditions cannot be simultaneously satisfied with a passive load network at a nonzero output power.

1.10 Output Power of Class E ZVS Amplifiers

The Class E ZVS RF power amplifier is shown in Fig. 1.9. Waveforms for the Class E power amplifier under ZVS and zero-derivative switching (ZDS) conditions are shown in Fig. 1.10. Ideally, the efficiency of this amplifier is 100%. Waveforms for the Class E amplifier are shown in Fig. 1.10. The drain current i_D has a jump at $t = t_o$. Hence, the derivative of the drain current at $t = t_o$ is given by

$$i'_D(t_o) = \Delta I\delta(t - t_o) \tag{1.73}$$

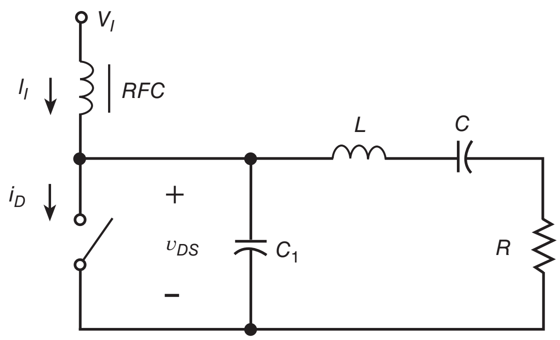


Figure 1.9 Class E ZVS power amplifier.

and the derivative of the drain-to-source voltage at $t = t_o$ is given by

$$v'_{DS}(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} v'_{DS}(t) + \lim_{t\to t_o^+} v'_{DS}(t)\right] = \frac{S_v}{2} \tag{1.74}$$

Assuming that $P_r = 0$ and $\eta_r = 1$, the output power of the Class E ZVS amplifier is

$$\begin{aligned} P_{ds} = P_O &= -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\, dt = -\frac{1}{4\pi^2 f}\int_0^T v'_{DS}(t_o)\Delta I \delta(t - t_o)dt \\ &= -\frac{\Delta I S_v}{8\pi^2 f}\int_0^T \delta(t - t_o)dt = -\frac{\Delta I S_v}{8\pi^2 f} \end{aligned} \tag{1.75}$$

where $\Delta I < 0$. Since

$$\Delta I = -0.6988 I_{DM} \tag{1.76}$$

and

$$S_v = 11.08 f V_{DSM} \tag{1.77}$$

the output power is

$$P_O = -\frac{\Delta I S_v}{8\pi^2 f} = 0.0981 I_{DM} V_{DSM} \tag{1.78}$$

Hence, the output power capability is

$$c_p = \frac{P_O}{I_{DM} V_{DSM}} = 0.0981 \tag{1.79}$$

Example 1.3

A Class E ZVS RF power amplifier has a step change in the drain current at the MOSFET turn-off $\Delta I_D = -1$ A, a slope of the drain-to-source voltage at the MOSFET turn-off $S_v = 11.08 \times 10^8$ V/s, and the operating frequency is $f = 1$ MHz. Find the output power of the Class E amplifier.

Solution. The output power of the Class E power amplifier is

$$P_O = -\frac{\Delta I S_v}{8\pi^2 f} = -\frac{-1 \times 11.08 \times 10^8}{8\pi^2 \times 10^6} = 14.03 \text{ W} \tag{1.80}$$

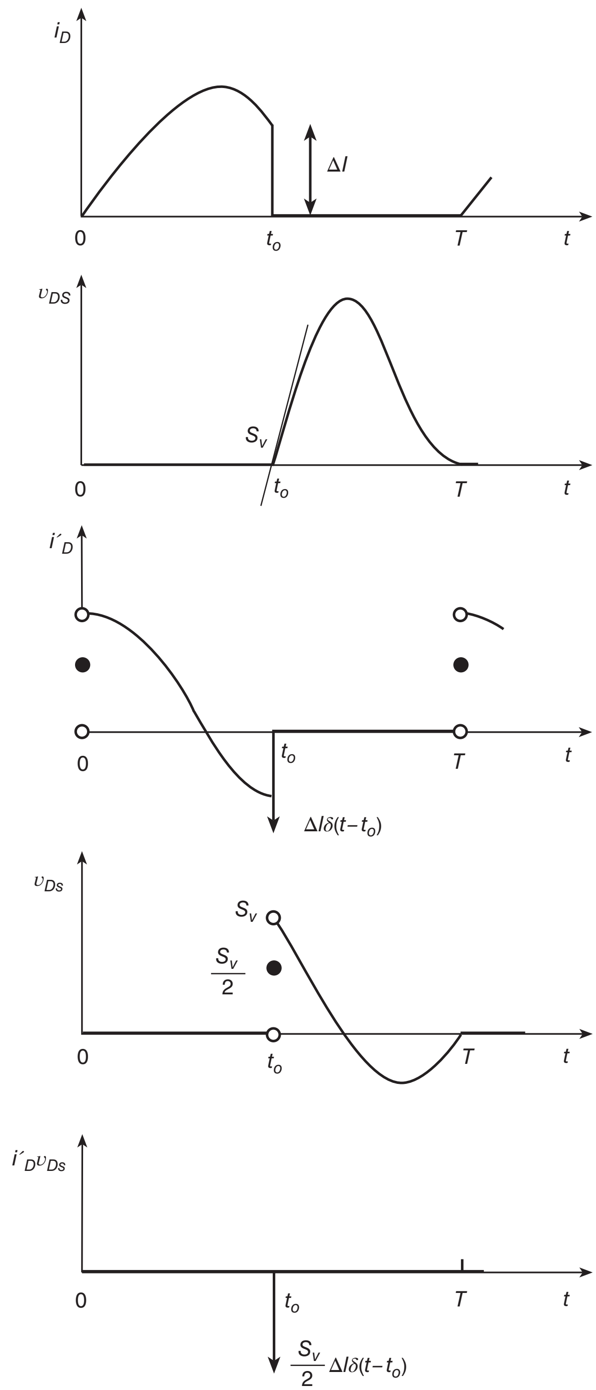


Figure 1.10 Waveforms of Class E ZVS power amplifier.

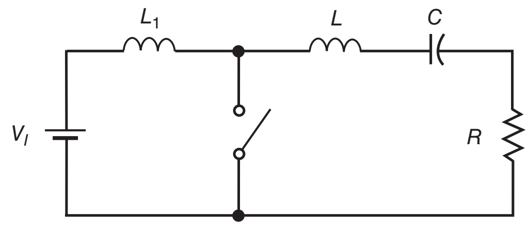


Figure 1.11 Class E ZCS power amplifier.

1.11 Class E ZCS Amplifiers

The Class E ZCS RF power amplifier is depicted in Fig. 1.11. Current and voltage waveforms under ZCS and ZDS conditions are shown in Fig. 1.12. The efficiency of this amplifier with perfect components and under ZCS condition is 100%. The drain-to-source voltage v_{DS} has a jump at $t = t_o$. The derivative of the drain-to-source voltage at $t = t_o$ is given by

$$v'_{DS}(t_o) = \Delta V \delta(t - t_o) \tag{1.81}$$

and the derivative of the drain current at $t = t_o$ is given by

$$i'_D(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} i'_D(t) + \lim_{t\to t_o^+} i'_D(t)\right] = \frac{S_i}{2} \tag{1.82}$$

The output power of the Class E ZCS amplifier is

$$\begin{aligned} P_{ds} = P_O &= -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\, dt = -\frac{1}{4\pi^2 f}\int_0^T i'_D \Delta V \delta(t - t_o) dt \\ &= -\frac{\Delta V S_i}{8\pi^2 f}\int_0^T \delta(t - t_o) dt = -\frac{\Delta V S_i}{8\pi^2 f} \end{aligned} \tag{1.83}$$

Since

$$\Delta V = -0.6988 V_{DSM} \tag{1.84}$$

and

$$S_i = 11.08 f I_{DM} \tag{1.85}$$

the output power is

$$P_O = -\frac{\Delta V S_i}{8\pi^2 f} = 0.0981 I_{DM} V_{DSM} \tag{1.86}$$

Hence, the output power capability is

$$c_p = \frac{P_O}{I_{DM} V_{DSM}} = 0.0981 \tag{1.87}$$

Example 1.4

A Class E ZCS RF power amplifier has a step change in the drain-to-source voltage waveform at the MOSFET turn-on $\Delta V_{DS} = -100$ V, a slope of the drain current at the MOSFET turn-on $S_i = 11.08 \times 10^7$ V/s, and the operating frequency is $f = 1$ MHz. Find the output power of the Class E amplifier.

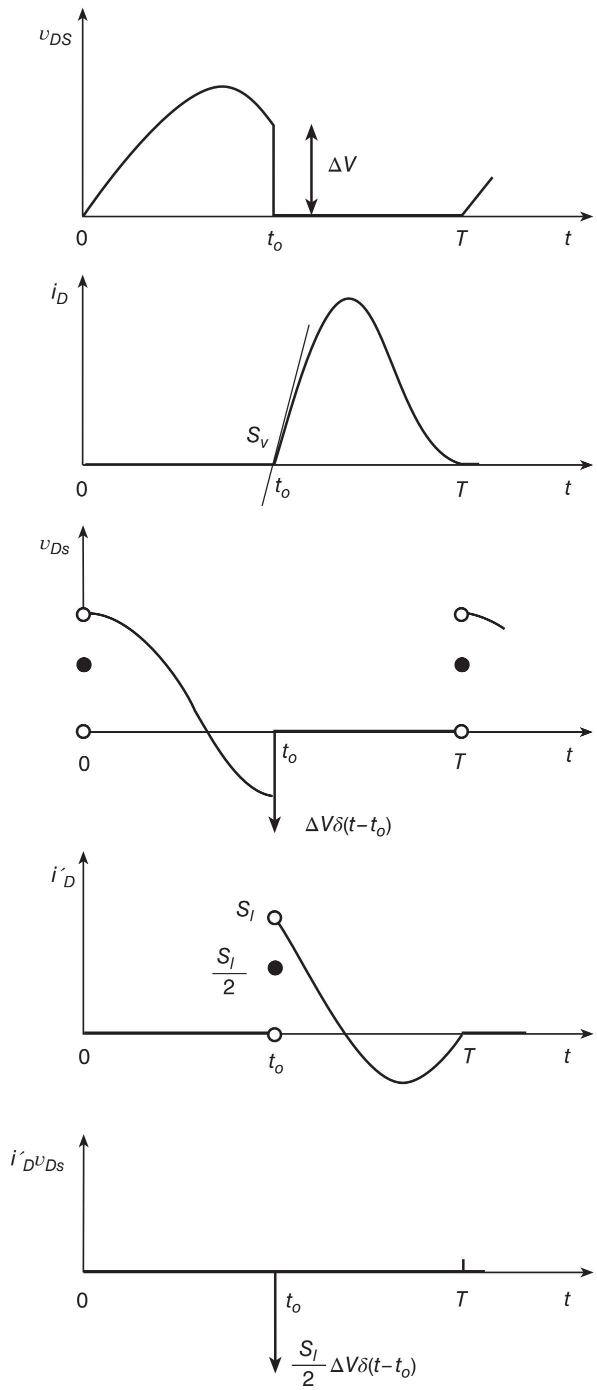


Figure 1.12 Waveforms of the Class E ZCS power amplifier.

Solution. The output power of the Class E power amplifier is

$$P_O = -\frac{\Delta V_{DS} S_i}{8\pi^2 f} = -\frac{-100 \times 11.08 \times 10^7}{8\pi^2 \times 10^6} = 140.320 \text{ W} \tag{1.88}$$

1.12 Antennas

The fundamental principle of wireless communication is based on Ampère–Maxwell's law

$$\Delta \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1.89}$$

Radiation. A transmitting antenna is used to radiate EMWs. The displacement current in the conductor of a transmitting antenna is

$$\frac{\partial \mathbf{D}}{\partial t} = 0 \tag{1.90}$$

Hence, Ampère–Maxwell's law becomes

$$\Delta \times \mathbf{H} = \mathbf{J} \tag{1.91}$$

This equation states that a time-varying magnetic field $\mathbf{H}(t)$ around the transmitting antenna is produced by a time-varying current density $\mathbf{J}(t)$ flowing in the transmitting antenna conductor.

Propagation. The conduction current in the air between the transmitting and receiving antennas is zero

$$\mathbf{J} = 0 \tag{1.92}$$

Hence, Ampère–Maxwell's law becomes

$$\Delta \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \tag{1.93}$$

This law states that magnetic and electric fields form an electromagnetic (EM) wave in the propagation process in the air (or other media).

Receiving of EM Wave. A receiving antenna is used to receive an EM wave and convert it into a current. The displacement current in the conductor of a receiving antenna is

$$\frac{\partial \mathbf{D}}{\partial t} = 0 \tag{1.94}$$

Hence, Ampère–Maxwell's law becomes

$$\mathbf{J} = \Delta \times \mathbf{H} \tag{1.95}$$

This equation states that a current density $\mathbf{J}$ is produced in the receiving antenna by a magnetic field $\mathbf{H}$ present around the receiving antenna.

An antenna is a device for radiating or receiving EM radio waves. A transmitting antenna converts an electrical signal into an EM wave. It is a transition structure between a guiding device (such as a transmission line) and free space. A receiving antenna converts an EM wave into an electrical signal. In free space, the EM wave travels at the speed of light $c = 3 \times 10^8$ m/s. The wavelength of an EM wave in free space is given by

$$\lambda = \frac{c}{f} \tag{1.96}$$

Transmitting antennas are used to *radiate* EMWs. The radiation efficiency of antennas is good only if their dimensions are of the same order of magnitude as the wavelength of the carrier frequency f_c. The length of antennas is usually $\lambda/2$ (a half-dipole antenna) or $\lambda/4$ (quarter-wave antenna) and should be higher than $\lambda/10$. The length of antenna depends on the wavelength of the EM wave. The length of quarter-wave antennas is

$$h_a = \frac{\lambda}{4} = \frac{c}{4f} \tag{1.97}$$

For example, the height of a quarter-wave antenna $h_a = 750$ m at $f_c = 100$ kHz, $h_a = 75$ m at $f_c = 1$ MHz, $h_a = 7.5$ cm at $f_c = 1$ GHz, and $h_a = 7.5$ mm at $f_c = 10$ GHz. Thus, mobile transmitters and receivers (transceivers) are only possible at high carrier frequencies.

An isotropic antenna is a theoretical point antenna that radiates energy equally in all directions with its power spread uniformly on the surface of a sphere. This results in a spherical wavefront. The uniform radiated power density at a distance r from an isotropic antenna with the output power P_T is given by

$$p(r) = \frac{P_T}{4\pi r^2} \tag{1.98}$$

The power density is inversely proportional to the square of the distance r. The hypothetical isotropic antenna is not practical, but is commonly used as a reference to compare with other antennas. If the transmitting antenna has directivity in a particular direction and efficiency, the power density in that direction is increased by a factor called the *antenna gain* G_T. The power density received by a receiving directive antenna is

$$p_r(r) = G_T \frac{P_T}{4\pi r^2} \tag{1.99}$$

The *antenna efficiency* is the ratio of the radiated power to the total power fed to the antenna $\eta_A = P_{RAD}/P_{FED}$.

A receiving antenna pointed in the direction of the radiated power gathers a portion of the power that is proportional to its cross-sectional area. The *antenna effective area* is given by

$$A_e = G_R \frac{\lambda^2}{4\pi} \tag{1.100}$$

where G_R is the gain of the receiving antenna and λ is the free-space wavelength. Thus, the power received by a receiving antenna is given by the Herald Friis formula for free-space transmission [12]

$$P_{REC} = A_e p(r) = \frac{1}{16\pi^2} G_T G_R P_T \left(\frac{\lambda^2}{r}\right)^2 = \frac{1}{16\pi^2} G_T G_R P_T \left(\frac{c}{rf_c}\right)^2 \tag{1.101}$$

The received power is proportional to $(\lambda/r)^2$ and to the gain of either antenna. As the carrier frequency f_c doubles, the received power decreases by a factor of 4 at a given distance r from the transmitting antenna. The gain of the dish (parabolic) antenna is given by

$$G_T = G_R = 6\left(\frac{D}{\lambda}\right)^2 = 6\left(\frac{Df_c}{c}\right)^2 \tag{1.102}$$

where D is the mouth diameter of the primary reflector. For $D = 3$ m and $f = 10$ GHz, $G_T = G_R = 60,000 = 47.8$ dB.

The *space loss* is the loss due to spreading the RF energy as it propagates through free space and is defined as

$$S_L = \frac{P_T}{P_{REC}} = \left(\frac{4\pi r}{\lambda}\right)^2 = 10\log\left(\frac{P_T}{P_{REC}}\right) = 20\log\left(\frac{4\pi r}{\lambda}\right) \tag{1.103}$$

There are also other losses such as atmospheric loss, polarization mismatch loss, impedance mismatch loss, and pointing error denoted by L_{syst}. Hence, the link equation is

$$P_{REC} = \frac{L_{syst} G_T G_R P_T}{(4\pi)^2} \left(\frac{\lambda}{r}\right)^2 \tag{1.104}$$

The maximum distance between transmitting and receiving antennas is

$$r_{max} = \frac{\lambda}{4\pi}\sqrt{L_{syst}G_T G_R\left(\frac{P_T}{P_{REC(min)}}\right)}. \tag{1.105}$$

For example, the Global System for Mobile Communications (GSM) cell radius is 35 or 60 km in the 900 MHz band and 20 km in the 1.8 GHz band.

1.13 Propagation of Electromagnetic Waves

EM wave propagation is illustrated in Fig. 1.13. There are three groups of EMWs based on their propagation properties:

- Ground waves (below 2 MHz).
- Sky waves (2–30 MHz).
- Line-of-sight waves, also called space waves or horizontal waves (above 30 MHz).

Ground waves travel parallel to the Earth's surface and suffer little attenuation by smog, moisture, and other particles in the lower part of the atmosphere. Very high antennas are required for transmission of these low-frequency (LF) waves. The approximate transmission distance of ground waves is about 1600 km (1000 miles). Ground wave propagation is much better over water, especially salt water, than over a dry desert terrain. Ground wave propagation is the only way to communicate into the ocean with submarines. Extremely low-frequency (ELF) waves (30–300 Hz) are used to minimize the attenuation of the waves by sea water. A typical frequency is 100 Hz.

The sky waves leave the curved surface of the Earth and are refracted by the ionosphere back to the surface of the Earth; therefore, they are capable of following the Earth's curvature. The altitude of refraction of the sky waves varies from 50 to 400 km. The transmission distance between two transmitters is 4000 km. The ionosphere is a region above the atmosphere, where free ions

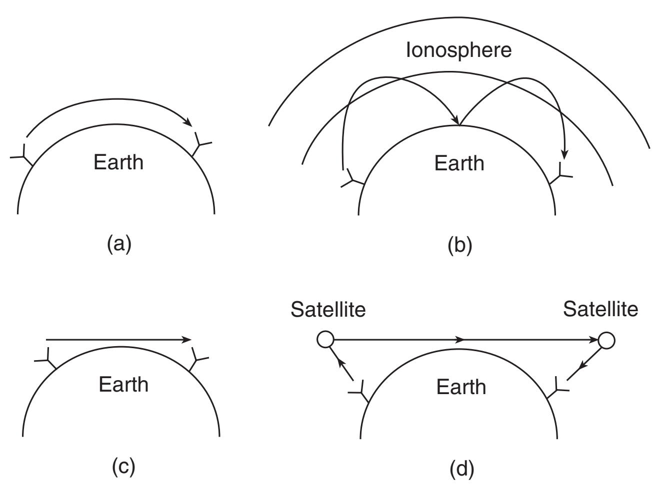


Figure 1.13 Electromagnetic wave propagation: (a) ground wave propagation; (b) sky wave propagation; (c) horizontal wave propagation; and (d) wave propagation in satellite communications.

and electrons exist in sufficient quantity to affect the wave propagation. Ionization is caused by radiation from the Sun. It changes as the position of a point on the Earth with respect to the Sun changes daily, monthly, and yearly. After sunset, the lowest layer of the ionosphere disappears because of rapid recombination of its ions. The higher the frequency, the more difficult the refracting (bending) process. Between the point where the ground wave is completely attenuated and the point where the first wave returns, no signal is received, resulting in a skip zone.

Line-of-sight waves follow straight lines. There are two types of line-of-sight waves: direct waves and ground reflected waves. The direct wave is by far the most widely used for propagation between antennas. Signals of frequencies above the high-frequency (HF) band cannot be propagated for long distances along the surface of the Earth. However, it is easy to propagate these signals through free space. The propagated power depends on distance, orientation of antennas, attenuation by buildings, and multipath [10].

1.14 Frequency Spectrum

Table 1.1 gives the frequency spectrum. In the United States, the allocation of carrier frequencies, bandwidths, and power levels of transmitted EMWs is regulated by the Federal Communications Commission (FCC) for all nonmilitary applications. The communication must occur in a certain part of the frequency spectrum. The carrier frequency f_c determines the channel frequency.

LF EMWs are propagated by ground waves. They are used for long-range navigation, telegraphy, and submarine communication. The medium-frequency (MF) band contains the commercial radio band from 535 to 1705 kHz. This band is used for radio transmission of AM signals to general audiences. The carrier frequencies are from 540 to 1700 kHz. For example, one carrier frequency is at $f_c = 550$ kHz, and the next carrier frequency is at $f_c = 560$ kHz. The modulation bandwidth is 5 kHz. The average power of local stations is from 0.1 to 1 kW. The average power of regional stations is from 0.5 to 5 kW. The average power of clear stations is from 0.25 to 50 kW. A radio receiver may receive power as low as of 10 pW, 1 μV/m, or 50 μV across a $300 - \Omega$ antenna. Thus, the ratio of the output power of the transmitter to the input power of the receiver is on the order of $P_T/P_{REC} = 10^{15}$.

The range from 1705 to 2850 kHz is used for short-distance point-to-point communications for services such as fire, police, ambulance, highway, forestry, and emergency services. The antennas in this band have reasonable height and radiation efficiency. The aeronautical frequency range starts in the MF range and ends in the HF range. It is from 2850 to 4063 kHz and is used for short-distance point-to-point communications and ground–air–ground

Table 1.1 Frequency spectrum.

Frequency range	Band name	Wavelength range
30–300 Hz	Extremely low frequencies (ELF)	10,000–1000 km
300–3000 Hz	Voice frequencies (VF)	1000–100 km
3–30 kHz	Very low frequencies (VLF)	100–10 km
30–300 kHz	Low frequencies (LF)	10–1 km
0.3–3 MHz	Medium frequencies (MF)	1000–100 m
3–30 MHz	High frequencies (HF)	100–10 m
30–300 MHz	Very high frequencies (VHF)	100–10 cm
0.3–3 GHz	Ultrahigh frequencies (UHF)	100–10 cm
3–30 GHz	Superhigh frequencies (SHF)	10–1 cm
30–300 GHz	Extra high frequencies (EHF)	10–1 mm

communications. Aircraft flying scheduled routes are allocated specific channels. The HF band contains the radio amateur band from 3.5 to 4 MHz in the United States. Other countries use this band for mobile and fixed communication services. HFs are also used for long-distance point-to-point transoceanic ground–air–ground communications. HFs are propagated by sky waves. The frequencies from 1.6 to 30 MHz are called short waves.

The very high-frequency (VHF) band contains commercial FM radio and most TV channels. The commercial FM radio transmission is from 88 to 108 MHz. The modulation bandwidth is 15 kHz. The average power is from 0.25 to 100 kW. The transmission distance of VHF TV signals is 160 km (100 miles).

TV channels for analog transmission range from 54 to 88 MHz and from 174 to 216 MHz in the VHF band and from 470 to 890 MHz in the ultrahigh-frequency (UHF) band. The bandwidth of analog TV is 6.7 MHz and digital TV is 10 MHz. The average power is 100 kW for the frequency range from 54 to 88 MHz and 316 kW for the frequency range from 174 to 216 MHz. Broadcast frequency allocations are given in Table 1.2. The UHF and SHF frequency bands are given in Table 1.3. The cellular phone frequency allocation is given in Table 1.4.

The superhigh frequency (SHF) band contains satellite communications channels. Satellites are placed in orbits. Typically, these orbits are 37,786 km in altitude above the equator. Each satellite illuminates about one-third of the Earth. Since the satellites maintain the same position relative to the Earth, they are placed in geostationary orbits. These geosynchronous satellites are called GEO satellites. Each satellite contains a communication system that can receive signals from the Earth or from another satellite and transmit the received signal back to the Earth or to another satellite. This system uses two carrier frequencies. The frequency for transmission from the Earth to the satellite (uplink) is 6 GHz and the transmission from the satellite to the Earth (downlink) is at 4 GHz. The bandwidth of each channel is 500 MHz. Directional antennas are used for radio transmission through free space. Satellite communications are used for TV and

Table 1.2 Broadcast frequency allocation.

Radio or TV	Frequency range	Channel spacing
AM Radio	535–1605 kHz	10 kHz
TV (channels 2–6)	54–72 MHz	6 MHz
TV	76–88 MHz	6 MHz
FM Radio	88–108 MHz	200 kHz
TV (channels 7–13)	174–216 MHz	6 MHz
TV (channels 14–83)	470–806 MHz	6 MHz

Table 1.3 UHF and SHF frequency bands.

Band name	Frequency range (GHz)
L	1–2
S	2–4
C	4–8
X	8–12.4
Ku	12.4–18
K	18–26.5
Ku	26.5–40
V	40–75
W	75–110

Table 1.4 Cellular phone frequency allocation.

System	Frequency range		Channel spacing (MHz)	Multiple access
AMPS	M-B	824–849 MHz	30	FDMA
	B-M	869–894 MHz	30	
GSM-900	M-B	880–915 MHz	0.2	TDMA/FDMA
	B-M	915–990 MHz	0.2	
GSM-1800	M-B	1710–1785 MHz	0.2	TDMA/FDMA
	B-M	1805–1880 MHz	0.2	
PCS-1900	M-B	1850–1910 kHz	30	TDMA
	B-M	1930–1990 kHz	30	
IS-54	M-B	824–849 MHz	30	TDMA
	B-M	869–894 MHz	30	
IS-136	M-B	1850–1910 MHz	30	TDMA
	B-M	1930–1990 MHz	30	
IS-96	M-B	824–849 MHz	30	CDMA
	B-M	869–894 MHz	30	
IS-96	M-B	1850–1910 MHz	30	CDMA
	B-M	1930–1990 MHz	30	

telephone transmission. The electronic circuits in the satellite are powered by solar energy using solar cells that deliver the supply power of about 1 kW. The combination of a transmitter and a receiver is called a transponder. A typical satellite has 12–24 transponders. Each transponder has a bandwidth of 36 MHz. The total time delay for GEO satellites is about 400 ms and the power of received signals is very low. Therefore, a low Earth orbit (LEO) satellite system was deployed for a mobile phone system. The orbits of LEO satellites are from 500 to 1500 km above the Earth. These satellites are not synchronized with the Earth's rotation. The total delay time for LEO satellite is about 250 ms.

1.15 Duplexing

In a two-way communication, a transmitter and a receiver are used. The combination of a transmitter and a receiver is called a *transceiver*. The block diagram of a transceiver is shown in Fig. 1.14. Duplexing techniques are used to allow for both users to transmit and receive signals. The most commonly used duplexing is called time-division duplexing (TDD). The same frequency channel is used for both transmitting and receiving signals, but the system transmits the signal for half of the time and receives for the other half.

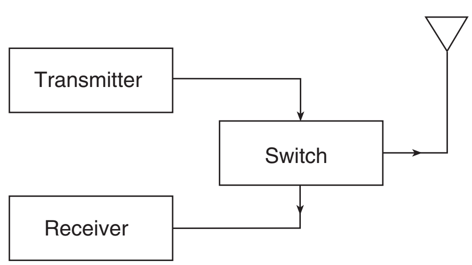


Figure 1.14 Block diagram of a transceiver.

1.16 Multiple-Access Techniques

In multiple-access communications systems, information signals are sent simultaneously over the same channel. Cellular wireless mobile communications use the following multiple-access techniques to allow simultaneous communication among multiple transceivers:

- Time-division multiple access (TDMA).
- Frequency-division multiple access (FDMA).
- Code-division multiple access (CDMA).

In the TDMA, the same frequency band is used by all the users but at different time intervals. Each digitally coded signal is transmitted only during preselected time intervals, called the time slots T_{SL}. During every time frame T_F, each user has access to the channel for a time slot T_{SL}. The signals transmitted from different users do not interfere with each other in the time domain.

In the FDMA, a frequency band is divided into many channels. The carrier frequency f_c determines the channel frequency. Each baseband signal is transmitted at a different carrier frequency. One channel is assigned to each user for a connection period. After the connection is completed, the channel becomes available to other users. In the FDMA, proper filtering must be carried out to provide channel selection. The signals transmitted from different users do not interfere with each other in the frequency domain. The GSM uses both TDMA and FDMA. The uplink is used for mobile transmission and the downlink is used for base station transmission. Each band is divided in 200 kHz slots. Each slot is shared between eight mobiles, each using it in turn. The bandwidth of square pulses is determined by their rise and fall times. Therefore, Gaussian pulse envelope waveforms are generated instead of rectangular pulses. The binary signal is processed by a Gaussian low-pass LP filter to reduce the rise and fall times of pulses, reducing the required bandwidth.

In the CDMA, each user uses a different code (similar to different language). CDMA allows one carrier frequency. Each station uses a different binary sequence to modulate the carrier. The signals transmitted from different users overlap in both the frequency and time domains, but the messages are orthogonal.

There are several standards of cellular wireless communications:

- Advanced Mobile Phone Service (AMPS).
- Global System for Mobile Communications (GSM).
- CDMA wireless standard proposed by Qualcomm.

1.17 Nonlinear Distortion in Transmitters

Linear amplification is required when the signal contains AM. Nonlinearity of amplifiers causes degradation of spectral purity and spectral regrowth. AM signals include multiple carriers. Power amplifiers contain a transistor (MOSFET, MESFET, or BJT), which is a nonlinear device operated under large-signal conditions. The drain current i_D is a nonlinear function of the gate-to-source voltage v_{GS}. Therefore, power amplifiers produce components, which are not present in the amplifier input signal. Another nonlinear mechanism is caused by the saturation of characteristics of the amplifier at a large amplitude of the output voltage. The relationship between the output voltage v_o and the input voltage v_s of a "weakly nonlinear" or a "nearly linear" power amplifier, such as the Class A amplifier, is nonlinear $v_o = f(v_s)$. This relationship can be expanded into Taylor's power series around the operating point Q as

$$v_o = f(v_s) = V_{O(DC)} + a_1 v_s + a_2 v_s^2 + a_3 v_s^3 + a_4 v_s^4 + a_5 v_s^5 + \cdots \tag{1.106}$$

Thus, the output voltage consists of an infinite number of nonlinear terms. Taylor's power series takes into account only the amplitude relationships. Volterra's power series includes both the amplitude and phase relationships.

Nonlinearity of amplifiers produces two types of unwanted signals:

- Harmonics of the carrier frequency $f_h = nf_c$.
- Intermodulation (IDM) products $f_{IDM} = nf_1 \pm mf_2$.

Nonlinear distortion components may corrupt the desired signal, causing errors in signal detection (reproduction) and splitter into adjacent channels. Harmonic distortion (HD) occurs when a single-frequency sinusoidal signal is applied to the power amplifier input. Intermodulation distortion (IMD) occurs when a signal consisting of two or more frequencies is applied at the power amplifier input. To evaluate the linearity of power amplifiers, we can use (1) a single-tone test and (2) and a two-tone test. In a single-tone test, a sinusoidal voltage source is used to drive a power amplifier. In a two-tone test, two sinusoidal sources of different frequencies and connected in series are used as a driver of a power amplifier. The first test will produce harmonics and the second test will produce both harmonics and intermodulation products (*IMP*s). Amplitude-to-phase conversion is caused by nonlinear capacitances. One measure of nonlinearity is the carrier-to-intermodulation (C/I) ratio, which should be 30 dBc or more for communication applications. The traditional measure of nonlinearity is the noise-to-power ratio (NPR).

1.18 Harmonics of Carrier Frequency

To investigate the process of generation of harmonics, let us assume that a power amplifier is driven by a single-tone excitation in the form of a sinusoidal voltage

$$v_s(t) = V_m \cos \omega t \tag{1.107}$$

To gain an insight into generation of harmonics by a nonlinear transmitter, consider an example of a memoryless time-invariant power amplifier modeled by a Taylor series taking a third-order polynomial

$$v_o(t) = a_0 + a_1 v_s(t) + a_2 v_s^2(t) + a_3 v_s^3(t) \tag{1.108}$$

The output voltage of the transmitter is given by

$$\begin{aligned} v_o(t) &= a_0 + a_1 V_m \cos \omega t + a_2 V_m^2 \cos^2 \omega t + a_3 V_m^3 \cos^3 \omega t \\ &= a_0 + a_1 V_m \cos \omega t + \frac{1}{2} a_2 V_m^2 (1 + \cos 2\omega t) + \frac{1}{4} a_3 V_m^3 (3 \cos \omega t + \cos 3\omega t) \\ &= a_0 + \frac{1}{2} a_2 V_m^2 + \left(a_1 V_m + \frac{3}{4} a_3 V_m^3 \right) \cos \omega t + \frac{1}{2} a_2 V_m^2 \cos 2\omega t + \frac{1}{4} a_3 V_m^3 \cos 3\omega t \end{aligned} \tag{1.109}$$

Thus, the output voltage of the power amplifier contains the fundamental components of the carrier frequency $f_1 = f_c$ as well as harmonics $2f_1 = 2f_c$ and $3f_1 = 3f_c, \ldots$, as shown in Fig. 1.15. The amplitude of the nth harmonic is proportional to V_m^n. This harmonics may interfere with other communication channels and must be filtered out to an acceptable low level.

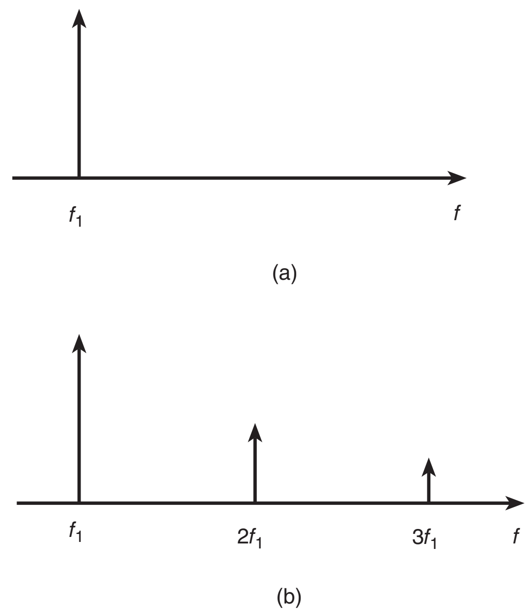


Figure 1.15 Spectrum of the input and output voltages of power amplifiers due to harmonics: (a) spectrum of the input voltage and (b) spectrum of the output voltage because of harmonics.

Even and odd terms produce the following components:

1. The even-order terms produce dc terms and even-order harmonics at $f_h = 2nf$. For example, the second harmonic is produced from the square term. In addition, extra dc terms are produced and the total dc bias increases with the ac signal amplitude V_m. It also varies when the amplitude of the input voltage V_m varies.
2. The odd-order terms produce signals at the fundamental frequency f and at the odd harmonic frequences $f_h = (n+1)f$. For example, the third-order term produces the fundamental component at f and the third harmonic at $3f$. Therefore, there is a nonlinear distortion of the fundamental component by the third-order term.

Let us consider a harmonic distortion in a MOSFET described by the square law. Assume that the total gate-to-source voltage waveform is

$$v_{GS} = V_{GS} + V_{gsm} \sin \omega t. \tag{1.110}$$

The drain current waveform is

$$\begin{aligned} i_D &= K(v_{GS} - V_t)^2 = K[(V_{GS} - V_t)^2 + 2(V_{GS} - V_t)V_{gsm} \sin \omega t + V_{gsm}^2 \sin^2 \omega t] \\ &= K[(V_{GS} - V_t)^2 + 2(V_{GS} - V_t)V_{gsm} \sin \omega t + \frac{1}{2}V_{gsm}^2(1 - \cos 2\omega t)] \\ &= K[(V_{GS} - V_t)^2 + \frac{1}{2}V_{gsm}^2 + 2K(V_{GS} - V_t)V_{gsm} \sin \omega t - \frac{1}{2}V_{gsm}^2 \cos 2\omega t \\ &= I_D + i_{d1} + i_{d2} \quad = I_D + I_{d1m} \sin \omega t - I_{d2} \cos 2\omega t. \end{aligned} \tag{1.111}$$

Hence, the distortion of the drain current waveform by the second harmonic is

$$HD_2 = \frac{I_{dm2}}{I_{dm1}} = \frac{V_{gsm}}{4(V_{GS} - V_t)} = 20 \log \left[\frac{V_{gsm}}{4(V_{GS} - V_t)} \right] = THD. \tag{1.112}$$

Harmonics are always integer multiples of the fundamental frequency. Therefore, the harmonic frequencies of the transmitter output signal with carrier frequency f_c are given by

$$f_h = nf_c \tag{1.113}$$

where $n = 2, 3, 4, \ldots$ is an integer. If a harmonic signal of a sufficiently large amplitude falls within the bandwidth of a nearby receiver, it may cause interference with the reception and cannot be filtered out in the receiver. The harmonics should be filtered out in the transmitter in its bandpass (BP) output network. For example, the output network of a transmitter may offer 37-dB second-harmonic suppression and 55-dB third-harmonic suppression.

The voltage gain of the amplifier at the fundamental frequency f_1 is

$$A_{v1} = \frac{v_{o1}}{v_s} = \frac{\left(a_1 + \frac{3}{4}a_3 V_m^2\right) V_m}{V_m} = a_1 + \frac{3}{4}a_3 V_m^2 \tag{1.114}$$

It can be seen that the voltage gain has not only the linear term a_1 but also an additional term proportional to the square of the input voltage amplitude V_m. In most amplifiers, $a_3 < 0$, yielding

$$A_{v1} = \frac{v_{o1}}{v_s} = a_1 - \frac{3}{4}|a_3| V_m^2 \tag{1.115}$$

and causing the voltage gain A_{v1} to decrease with V_m from the desired linear plot. Therefore, the upper part of the instantaneous output voltage tends to be reduced for large values of V_m. This phenomenon is known as *gain compression* or *amplifier saturation*. The linear part of the voltage gain is limited by the power supply voltage. The range of the input or the output voltage for which the voltage gain is linear is called the *dynamic range*.

Harmonic distortion is defined as the ratio of the amplitude of the nth harmonic V_n to the amplitude of the fundamental V_1

$$HD_n = \frac{V_n}{V_1} = 20 \log \left(\frac{V_n}{V_1}\right) \text{ (dB)} \tag{1.116}$$

The distortion by the second harmonic is given by

$$HD_2 = \frac{V_2}{V_1} = \frac{\frac{1}{2}a_2 V_m^2}{a_1 V_m + \frac{3}{4}a_3 V_m^3} = \frac{a_2 V_m}{2\left(a_1 + \frac{3}{4}V_m^2\right)} \tag{1.117}$$

For $a_1 \gg 3a_3 V_m^2/4$,

$$HD_2 \approx \frac{a_2 V_m}{2a_1} \tag{1.118}$$

The second-harmonic distortion HD_2 is proportional to the input voltage amplitude V_m.

The distortion by the third harmonic is given by

$$HD_3 = \frac{V_3}{V_1} = \frac{\frac{1}{4}a_3 V_m^3}{a_1 V_m + \frac{3}{4}a_3 V_m^3} = \frac{a_3 V_m^2}{4a_1 + 3a_3 V_m^2} \tag{1.119}$$

For $a_1 \gg 3a_3 V_m^2/4$,

$$HD_3 \approx \frac{a_3 V_m^2}{4a_1} \tag{1.120}$$

The third-harmonic distortion HD_3 is proportional to V_m^2. Usually, the amplitudes of harmonics should be -50 to -70 dB below the amplitude of the carrier.

The ratio of the power of an nth harmonic P_n to the carrier power P_c is

$$HD_n = \frac{P_n}{P_c} = 10\log\left(\frac{P_n}{P_c}\right) \text{ (dBc)} \tag{1.121}$$

The term "dBc" refers to the ratio of the power of a spectral distortion component P_n to the power of the carrier P_c.

The harmonic content in a waveform is described by the total harmonic distortion (THD), defined as

$$THD = \sqrt{\frac{P_2 + P_3 + P_4 + \cdots}{P_1}} = \sqrt{\frac{\frac{V_2^2}{2R} + \frac{V_3^2}{2R} + \frac{V_4^2}{2R} + \cdots}{\frac{V_1^2}{2R}}} = \sqrt{\frac{V_2^2 + V_3^2 + V_4^2 + \cdots}{V_1^2}}$$

$$a = \sqrt{\left(\frac{V_2}{V_1}\right)^2 + \left(\frac{V_3}{V_1}\right)^2 + \left(\frac{V_4}{V_1}\right)^2 + \cdots} = \sqrt{HD_2^2 + HD_3^2 + HD_4^2 + \cdots} \tag{1.122}$$

where $HD_2 = V_2/V_1, HD_3 = V_3/V_1, \ldots$ Higher harmonics ($n \geq 2$) are distortion terms.

1.19 Intermodulation Distortion

Intermodulation (IM) occurs when two or more signals of different frequencies are applied to the input of a nonlinear circuit, such as a nonlinear RF transmitter. This results in mixing the components of different frequencies. Therefore, the output signal contains components with additional frequencies, called *IMPs*. The frequencies of the IM products are either the sums or the differences of the carrier frequencies of input signals and their harmonics. For a two-frequency input excitation at frequencies f_1 and f_2, the frequencies of the output signal components are given by

$$f_{IMD} = nf_1 \pm mf_2 \tag{1.123}$$

where $n = 0, 1, 2, 3, \ldots$ and $m = 0, 1, 2, 3, \ldots$ are integers. The order of an IM product for a two-tone signal is the sum of the absolute values of the coefficients n and m given by

$$k_{IMD} = n + m \tag{1.124}$$

If the IM products of sufficiently large amplitudes fall within the bandwidth of a receiver, they will degrade the reception quality. For example, $2f_1 + f_2$, $2f_1 - f_2$, $2f_2 + f_1$, and $2f_2 - f_1$ are the third-order IM products. The third-order IM products usually have components in the system bandwidth. In contrast, the second-order harmonics $2f_1$ and $2f_2$ and the second-order IM products $f_1 + f_2$ and $f_1 - f_2$ are generally out of the system passband and are therefore not a serious problem.

A two-tone (two-frequency) excitation test is used to evaluate IM distortion of power amplifiers. Assume that the input voltage of a power amplifier consists of two sinusoids of equal amplitudes V_m and closely spaced frequencies f_1 and f_2

$$v_s(t) = V_m(\cos\omega_1 t + \cos\omega_2 t) = 2V_m \cos\left(\frac{\omega_2 - \omega_1}{2}\right)\cos\left(\frac{\omega_2 + \omega_1}{2}\right) \tag{1.125}$$

If the power amplifier is a memoryless time-invariant circuit described by a Taylor series using a third-order polynomial at the operating point Q, the output voltage waveform is given by

$$v_o(t) = a_0 + a_1 v_s(t) + a_2 v_s^2(t) + a_3 v_s^3(t)$$
$$= a_0 + a_1 V_m(\cos\omega_1 t + \cos\omega_2 t) + a_2 V_m^2(\cos\omega_1 t + \cos\omega_2 t)^2 + a_3 V_m^3(\cos\omega_1 t + \cos\omega_2 t)^3 + \cdots$$

$$
\begin{aligned}
&= a_0 + a_1 V_m \cos\omega_1 t + a_1 V_m \cos\omega_2 t + \frac{1}{2}a_2 V_m^2(1+\cos 2\omega_1 t) + \frac{1}{2}a_2 V_m^2(1+\cos 2\omega_2 t)\\
&\quad + a_2 V_m^2 \cos(\omega_2-\omega_1)t + a_1 V_m^2\cos(\omega_1+\omega_2)t + a_3 V_m^3\left(\frac{3}{4}\cos\omega_1 t + \frac{1}{4}\cos 3\omega_1 t\right)\\
&\quad + a_3 V_m^3\left(\frac{3}{4}\cos\omega_2 t + \frac{1}{4}\cos 3\omega_2 t\right)\\
&\quad + a_3 V_m^3\left[\frac{3}{2}\cos\omega_1 t + \frac{3}{4}\cos(2\omega_2-\omega_1)t + \frac{3}{4}\cos(2\omega_2+\omega_1)t\right]\\
&\quad + a_3 V_m^3\left[\frac{3}{2}\cos\omega_2 t + \frac{3}{4}\cos(2\omega_1-\omega_2)t + \frac{3}{4}\cos(2\omega_1+\omega_2)t\right]\\
&\quad + \frac{1}{4}a_3 V_m^3(\cos 3\omega_1 t + \cos 3\omega_3 t) + \cdots\\
&= a_0 + a_2 V_m^2 + (a_1 V_m + \frac{9}{4}a_3 V_m^3)\cos\omega_1 t + (a_1 V_m + \frac{9}{4}a_3 V_m^3)\cos\omega_2 t\\
&\quad + \frac{1}{2}a_2 V_m^2(\cos 2\omega_1 t + \cos 2\omega_2 t) + a_2 V_m^2[\cos(\omega_2-\omega_1) + \cos(\omega_1+\omega_2)t]\\
&\quad + \frac{3}{4}V_m^3\cos(2\omega_2-\omega_1) + \frac{3}{4}V_m^3\cos(2\omega_2+\omega_1) + \frac{1}{4}a_3 V_m^3(\cos 3\omega_1 t + \cos 3\omega_2 t) + \cdots
\end{aligned}
\tag{1.126}
$$

Now assume that the input signal of the power amplifier consists of two sinusoids of different amplitudes V_{m1} and V_{m2} and closely spaced frequencies f_1 and f_2

$$
v_s(t) = V_{m1}\cos\omega_1 t + V_{m2}\cos\omega_2 t \tag{1.127}
$$

If the power amplifier is a memoryless time-invariant circuit described by a Taylor series using a third-order polynomial, the output voltage waveform is given by

$$
\begin{aligned}
v_o(t) &= a_0 + a_1 v_s(t) + a_2 v_s^2(t) + a_3 v_s^3(t)\\
&= a_0 + a_1(V_{m1}\cos\omega_1 t + V_{m2}\cos\omega_2 t) + a_2(V_{m2}\cos\omega_1 t + V_{m2}\cos\omega_2 t)^2\\
&\quad + a_3(V_{m1}\cos\omega_1 t + V_{m2}\cos\omega_2 t)^3\\
&= a_0 + a_1 V_{m1}\cos\omega_1 t + a_1 V_{m2}\cos\omega_2 t + a_2 V_{m1}^2\cos^2\omega_1 t + 2a_2 V_{m1}V_{m2}\cos\omega_1 t\cos\omega_2 t\\
&\quad + a_2 V_{m2}^2\cos\omega_2 t + a_3 V_{m1}^3\cos^3\omega_1 t + 3a_3 V_{m1}^2 V_{m2}\cos^2\omega_1 t\cos\omega_2 t\\
&\quad + 3a_3 V_{m1}V_{m2}^2\cos\omega_1 t\cos^2\omega_2 t + a_3 V_{m2}^3\cos^3\omega_2 t
\end{aligned}
\tag{1.128}
$$

Thus,

$$
\begin{aligned}
v_o &= a_0 + \frac{1}{2}a_2(V_{m1}^2 + V_{m2}^2) + \left(a_1 V_{m1} + \frac{3}{2}a_3 V_{m1}V_{m2}^2 + \frac{3}{4}a_3 V_{m1}^3\right)\cos\omega_1 t\\
&\quad + \left(a_1 V_{m2} + \frac{3}{2}a_3 V_{m2}V_{m1}^2 + \frac{3}{4}a_3 V_{m2}^3\right)\cos\omega_2 t + \frac{1}{2}V_m^2\cos 2\omega_1 t + \frac{1}{2}V_m^2\cos 2\omega_2 t\\
&\quad + a_2 V_{m1}V_{m2}\cos(\omega_2-\omega_1)t + a_2 V_{m1}V_{m2}\cos(\omega_1+\omega_2)t\\
&\quad + \frac{3}{4}a_3 V_{m1}^2 V_{m2}\cos(2\omega_1-\omega_2)t + \frac{3}{4}a_3 V_{m1}^2 V_{m2}\cos(2\omega_1+\omega_2)t\\
&\quad + \frac{3}{4}a_3 V_{m1}V_{m2}^2\cos(2\omega_2-\omega_1)t + \frac{3}{4}a_3 V_{m1}V_{m2}^2\cos(2\omega_2+\omega_1)t\\
&\quad + \frac{1}{4}a_3 V_{m1}^3\cos 3\omega_1 t + \frac{1}{4}a_3 V_{m2}^3\cos 3\omega_2 t + \cdots
\end{aligned}
\tag{1.129}
$$

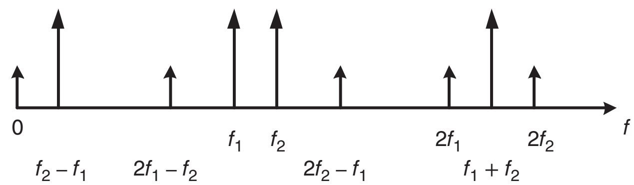


Figure 1.16 Spectrum of the output voltage in power amplifiers due to intermodulation for the two-tone input voltage with $f_2 > f_1$.

Figure 1.16 shows a spectrum of the amplifier output voltage due to intermodulation for two-tone input voltage with $f_2 > f_1$. The output voltage waveform v_o contains the following components:

- A dc component, causing a change in the dc operating current (bias current) of the transistor.
- Fundamental components f_1 and f_2.
- Harmonics of the fundamental components $2f_1, 2f_2, 3f_1, 3f_2, \ldots$
- IM products, which are linear combinations of the input frequencies f_1 and f_2: $nf_1 \pm mf_2$, where $m, n = 0, \pm 1, \pm 2, \pm 3, \ldots$ The IM frequencies are $f_2 - f_1$, $f_1 + f_2$, $2f_1 - f_2$, $2f_2 - f_1$, $2f_1 + f_2$, $2f_2 + f_1$, $3f_1 - 2f_2$, $3f_2 - 2f_1, \ldots$

If the difference between f_2 and f_1 is small, the IM products appear in the close vicinity of f_1 and f_2. The third-order IM products, which are at $2f_1 - f_2$ and $2f_2 - f_1$, are of the most interest because they are the closest to the fundamental components, as illustrated in Fig. 1.17. The frequency

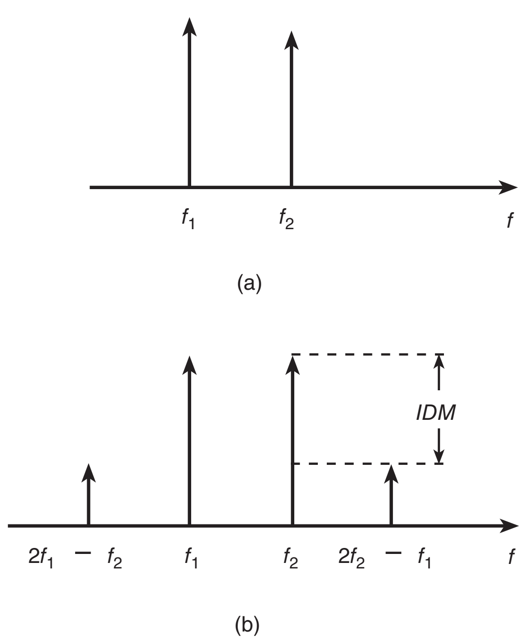


Figure 1.17 Spectrum of the input and output voltages in power amplifiers due to intermodulation: (a) spectrum of the input voltage and (b) several components of the spectrum of the output voltage due to intermodulation.

difference of the IM product at $2f_1 - f_2$ from the fundamental component at f_1 is

$$\Delta f_1 = f_1 - (2f_1 - f_2) = f_2 - f_1 \tag{1.130}$$

The frequency difference of the *IM* product at $2f_2 - f_1$ from the fundamental component at f_2 is

$$\Delta f_2 = (2f_2 - f_1) - f_2 = f_2 - f_1 \tag{1.131}$$

For example, if $f_1 = 800$ kHz and $f_2 = 900$ kHz, then the *IMP*s of particular interest are $2f_1 - f_2 = 2 \times 800 - 900 = 700$ MHz and $2f_2 - f_1 = 2 \times 900 - 800 = 1000$ MHz, resulting in $\Delta f_1 = f_1 - (2f_1 - f_2) = 800 - 700 = 100$ kHz and $\Delta f_2 = (2f_2 - f_1) - f_2 = 1000 - 900 = 100$ kHz. To filter out the unwanted IM components, filters with a very narrow bandwidth are required.

The output voltage at the input signal frequencies f_1 and f_2 is

$$\begin{aligned} v_{o(f_1,f_2)} = {} & \left(a_1 V_{m1} + \frac{3}{2} a_3 V_{m1} V_{m2}^2 + \frac{3}{4} a_3 V_{m1}^3\right) \cos \omega_1 t \\ & + \left(a_1 V_{m2} + \frac{3}{2} a_3 V_{m2} V_{m1}^2 + \frac{3}{4} a_3 V_{m2}^3\right) \cos \omega_2 t \end{aligned} \tag{1.132}$$

which for $V_{m1} = V_{m2} = V_m$ becomes

$$v_{o(f_1,f_2)} = \left(a_1 + \frac{9}{2} a_3 V_m^2\right) V_m (\cos \omega_1 t + \cos \omega_2 t) \tag{1.133}$$

The second-order IM products of the output voltage occur at $f_1 - f_2$ and $f_1 + f_2$ and are given by

$$v_{o(f_1-f_2,f_1+f_2)} = a_2 V_{m1} V_{m2} [\cos(\omega_2 - \omega_1)t + \cos(\omega_1 + \omega_2)t] \tag{1.134}$$

which for $V_{m1} = V_{m2} = V_m$ simplifies to the form

$$v_{o(f_1-f_2,f_1+f_2)} = a_2 V_m^2 [\cos(\omega_2 - \omega_1)t + \cos(\omega_1 + \omega_2)t] \tag{1.135}$$

The third-order IM products of the output voltage occur at $2f_1 - f_2$ and $2f_2 - f_1$ and are expressed as

$$v_{o(2f_1-f_2,2f_2-f_1)} = \frac{3}{4} a_3 V_{m1}^2 V_{m2} \cos(2\omega_1 - \omega_2)t + \frac{3}{4} a_3 V_{m1} V_{m2}^2 \cos(2\omega_2 - \omega_1)t \tag{1.136}$$

yielding for $V_{m1} = V_{m2} = V_m$

$$v_{o(2f_1-f_2,2f_2-f_1)} = \frac{3}{4} a_3 V_m^3 [\cos(2\omega_1 - \omega_2)t + \cos(2\omega_2 - \omega_1)t] \tag{1.137}$$

The amplitudes of the fundamental components of the output voltage are

$$V_{f_1} = V_{f_2} = \left(a_1 + \frac{9}{4} a_3 V_m^2\right) V_m \tag{1.138}$$

the amplitudes of the second-order IM product are

$$V_{f_2-f_1} = V_{f_1+f_2} = a_2 V_m^2 \tag{1.139}$$

and the amplitudes of the third-order IM product are

$$V_{2f_2-f_1} = V_{2f_1-f_2} = \frac{3}{4} a_3 V_m^3 \tag{1.140}$$

Assuming that $V_{m1} = V_{m2} = V_m$ and $a_3 \ll a_1$, the third-order IM distortion by the *IMP* component at $2f_1 \pm f_2$ or the *IMP* component at $2f_2 \pm f_1$ is defined as a ratio of the amplitude of the third-order IM component of the output voltage to the amplitude of the fundamental component of the output voltage

$$IM_3 = \frac{V_{2f_2-f_1}}{V_{f_2}} = \frac{\frac{3}{4} a_3 V_m^3}{\left(a_1 + \frac{9}{4} a_3 V_m^2\right) V_m} = \frac{\frac{3}{4} a_3 V_m^2}{a_1 + \frac{9}{4} a_3 V_m^2} \approx \frac{\frac{3}{4} a_3 V_m^2}{a_1} = \frac{3}{4} \left(\frac{a_3}{a_1}\right) V_m^2 \tag{1.141}$$

for $a_1 \gg \frac{9}{4}a_3V_m^2$. Similarly,

$$IM_2 = \frac{a_2V_m^2}{\left(a_1 + \frac{9}{4}a_3V_m^2\right)V_m} = \frac{a_2V_m}{a_1 + \frac{9}{4}a_3V_m^2} \approx \frac{a_2}{a_1}V_m \tag{1.142}$$

for $a_1 \gg \frac{9}{4}a_3V_m^2$. As the amplitude of the input voltage V_m is increased, the amplitudes of the fundamental components V_{f_1} and V_{f_2} are directly proportional to V_m, whereas the amplitudes of the third-order IM products $V_{2f_1-f_2}$ and $V_{2f_2-f_1}$ are proportional to V_m^3. Therefore, the amplitudes of the IM products increase three times faster on log-log scale than the amplitudes of the fundamentals and have an intersection point.

If the amplitudes are drawn on a log-log scale, they are linear functions of V_m. Usually, the coefficient a_3 is negative, causing the compression of the input voltage versus output voltage characteristic and reducing the amplifier voltage gain. The compression of the voltage gain leads to the compression of the power gain and the characteristic of $P_O = f(P_G)$ saturates. The 1-dB compression point indicates the maximum power value of the linear dynamic power range.

In general, the amplitude of input voltage at which the extrapolated amplitude of the desired output voltage and the nth-order IM product are equal is the nth-order intercept point. Thus, $IM_n = 1$. For the second-order IM component,

$$IM_2 = \frac{a_2}{a_1}V_m = 1 \tag{1.143}$$

which produces

$$V_m = \frac{a_2}{a_1} \tag{1.144}$$

Similarly, for the third-order IM component,

$$IM_3 = \frac{3}{4}\frac{a_2}{a_1}V_m^2 = 1 \tag{1.145}$$

yielding

$$V_m = \sqrt{\frac{4}{3}\frac{a_2}{a_1}} \tag{1.146}$$

Nonlinearities produce power in signal bandwidth. This effect is characterized by the notch power to the total signal power ratio

$$NPR = \frac{\text{Notch power}}{\text{Total signal power}} \tag{1.147}$$

To determine NPR, a power amplifier is driven by Gaussian noise through a notch filter in a portion of its bandwidth.

The effect of the nonlinearities on the adjacent channels is described by the adjacent channel power ratio (ACPR)

$$ACPR = \frac{\text{Power in band outside signal bandwidth}}{\text{Signal power}} = \frac{P_{adjacent-channels}}{P_O} = \frac{\int_{LS} P_O(f)df + \int_{US} P_O(f)df}{\int P_O(f)df} \tag{1.148}$$

The ACPR may be specified for either lower or upper sideband.

1.20 AM/AM Compression and AM/PM Conversion

Assume an AM single-tone (single-frequency) input voltage of an amplifier

$$v_s(t) = V_m(t)\cos\omega t \tag{1.149}$$

In real amplifiers, the voltage gain is a function of the amplitude of the input voltage $k_p(t) = k_p[V_{sm}(t)]$, resulting in an instantaneous voltage gain. In addition, the amplifier transistor contains nonlinear capacitances, dependent on the magnitudes of the transistor voltages. Therefore, the output voltage is

$$v_o(t) = k_p(t)v_s(t) = k_p[V_m(t)]\cos\{\omega t + \phi[V_m(t)]\} \tag{1.150}$$

This results in an AM/AM compression and an AM/PM conversion.

1.21 Dynamic Range of Power Amplifiers

For an ideal amplifier, a plot of the output power as a function of the input power should be a perfect straight line, $P_O = k_pP_i$, where k_p is the power gain and should be constant. Figure 1.18 shows the desired output power $P_O(f_2)$ and the undesired third-order IM product output power $P_O(2f_2 - f_1)$ as functions of the input power P_i on a log-log scale. This characteristic exhibits a linear region and a nonlinear region. As the input power P_i increases, the output power first increases proportionally to the input power and then reaches saturation, causing *power gain compression*. The point at which the power gain of the nonlinear amplifier deviates from that of the ideal linear amplifier by 1 dB is called the *1-dB compression point*. It is used as a measure of the power-handling capability of the power amplifier. The output power at the 1-dB compression point is given by

$$P_{O(1\text{dB})} = k_{p(1\text{dB})} + P_{i(1\text{dB})}\ (\text{dBm}) = k_{po(1\text{dB})} - 1\ \text{dB} + P_{i(1\text{dB})}\ (\text{dBm}) \tag{1.151}$$

where k_{po} is the power gain of an ideal linear amplifier and $k_{p(1\text{dB})}$ is the power gain at the 1-dB compression point.

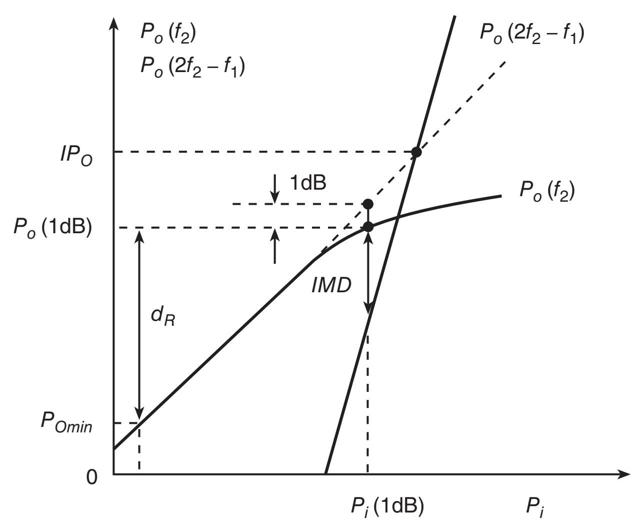


Figure 1.18 Output power $P_O(f_2)$ and $P_O(2f_2 - f_1)$ as functions of input power P_i of power amplifier.

The *dynamic range* of a power amplifier is the region where the amplifier has a linear (i.e., fixed) power gain. It is defined as the difference between the output power $P_{O(1\text{dB})}$ and the minimum detectable power P_{Omin}

$$d_R = P_{O(1\text{dB})} - P_{Omin} \tag{1.152}$$

where the minimum detectable power P_{Omin} is defined as an output power level x dB above the input noise power level P_{On}, usually $x = 3$ dB.

In a linear region of the amplifier characteristic, the desired output power $P_O(f_2)$ is proportional to the input power P_i, for example, $P_O(f_2) = aP_i$. Assume that the third-order IM product output power $P_O(2f_2 - f_1)$ increases proportionally to the third power, for example, $P_O(2f_2 - f_1) = (a/8)^3 P_i$. Projecting the linear region of $P_O(f_2)$ and $P_O(2f_2 - f_1)$ results in an intersection point called the *intercept point* (IP), as shown in Fig. 1.18, where the output power at the IP point is denoted by IP_O.

The IM product is the difference between the desired output power $P_O(f_2)$ and the undesired output power of the IM component $P_O(2f_2 - f_1)$ of a power amplifier

$$IMD = P_O(f_2)\ (\text{dBm}) - P_O(2f_2 - f_1)\ (\text{dBm}) \tag{1.153}$$

Signals with constant envelopes, such as CW, FM, FSK, and GSM, do not require linear amplification.

1.22 Analog Modulation

A *message signal* $v_m(t)$ is usually an LP signal. A signal whose spectrum is in the vicinity $f = 0$ is usually called a *baseband signal*. A baseband signal is usually voice, video, or digital data signal. Modulation converts a message signal $v_m(t)$ from an LP spectrum to a high-pass (HP) spectrum (usually to a BP spectrum) around a carrier frequency f_c. In general, a modulated signal can be described as

$$v(t) = A(t)\cos[2\pi f(t)t + \phi(t)] \tag{1.154}$$

where the amplitude $A(t)$, the frequency $f(t)$, and the phase $\phi(t)$ are modulated. Using the relationship $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$, any narrow band signal can be presented as simultaneous AM and PM waveform

$$\begin{aligned} v_{RF}(t) &= A(t)\cos[2\pi ft + \phi(t)] = A(t)\cos\phi(t)\cos 2\pi ft - A(t)\sin\phi(t)\sin 2\pi ft \\ &= I(t)\cos 2\pi ft - Q(t)\sin 2\pi ft \end{aligned} \tag{1.155}$$

where

$$I(t) = A(t)\cos\phi(t) \tag{1.156}$$

and

$$Q(t) = A(t)\sin\phi(t) \tag{1.157}$$

$$A(t) = \sqrt{I^2(t) + Q^2(t)} = \sqrt{A^2(t)[\sin^2\phi(t) + \cos^2\phi(t)]} \tag{1.158}$$

and

$$\phi(t) = \arctan\left[\frac{Q(t)}{I(t)}\right] \tag{1.159}$$

The function of a communication system is to transfer information from one point to another through a communication link. Block diagrams of a typical communication system are depicted in Fig. 1.19. The system consists of an upconversion transmitter whose block diagram is shown in Fig. 1.19(a) and a downconversion receiver whose block diagram is shown in Fig. 1.19(b).

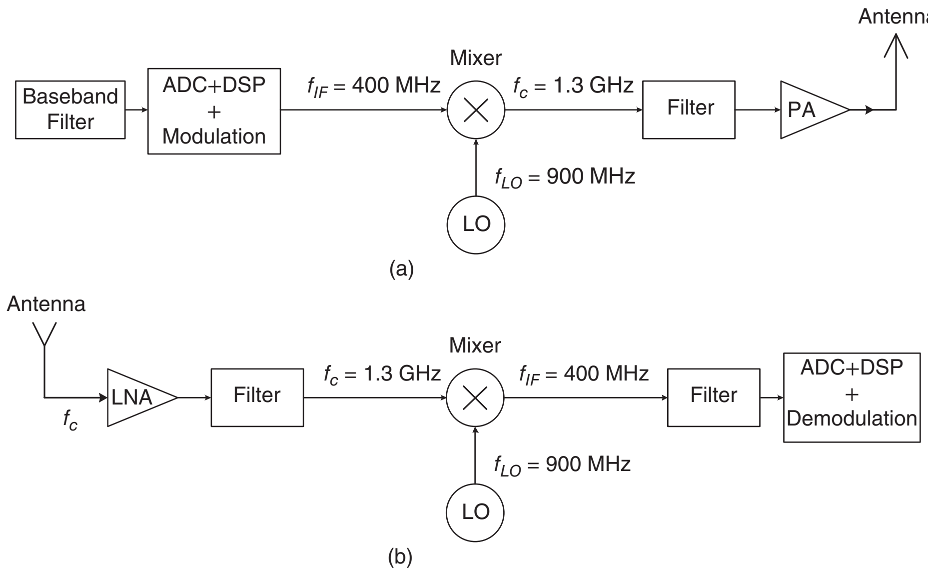


Figure 1.19 Block diagrams of a transmitter and a receiver. (a) Block diagram of an upconversion transmitter. (b) Block diagram of a downconversion receiver.

In a transmitting system, an RF signal is generated, amplified, modulated, and applied to the antenna. A local oscillator generates a signal with a frequency f_{LO}. The signals with an intermediate frequency f_{IF} and the local-oscillation frequency f_{LO} are applied to a mixer. The frequency of the output signal of the mixer and a BP filter is increased (upconversion) from the intermediate frequency f_{IF} to a carrier frequency f_c by adding the local-oscillation frequency f_{LO}

$$f_c = f_{LO} + f_{IF} \tag{1.160}$$

The RF current flows through the antenna and produces EMWs. Antennas produce or collect EM energy. The transmitted signal is received by the antenna, amplified by a LNA, and applied to a mixer. The frequency of the output signal of the mixer and a BP filter is reduced (downconversion) from the carrier frequency f_c to an intermediate frequency f_{IF} by subtracting the local-oscillation frequency f_{LO}

$$f_{IF} = f_c - f_{LO} \tag{1.161}$$

The most important parameters of a transmitter are as follows:

- Spectral efficiency.
- Power efficiency.
- Signal quality in the presence of noise and interference.

A "baseband" signal (modulating signal or information signal) has a nonzero spectrum in the vicinity of $f = 0$ and negligible elsewhere. For example, a voice signal generated by a microphone or a video signal generated by a TV camera are baseband signals. The modulating signal may consist of many, for example, 24 multiplexed telephone channels. In RF systems with analog modulation, the carrier is modulated by an analog baseband signal. The frequency bandwidth occupied by the baseband signal is called the baseband. The modulated signal consists of components with much higher frequencies than the highest baseband frequency. The modulated signal is an RF signal. It consists of components whose frequencies are very close to the frequency of the carrier.

RF modulated signals can be divided into two groups:

- Variable-envelope signals.
- Constant-envelope signals.

Amplification of variable-envelope signals requires linear amplifiers. On the contrary, constant-envelope signals may be amplified by nonlinear amplifiers. However, they require flat frequency response characteristics.

Modulation is the process of placing an information band around a HF carrier for transmission. Modulation conveys information by changing some aspects of a carrier signal in response to a modulating signal. In general, a modulated output voltage is given by

$$v_o(t) = A(t)\cos[2\pi f(t)t + \theta(t)] \tag{1.162}$$

where $A(t)$ is the amplitude of the voltage, and $\theta(t)$ is the phase of the carrier. If the amplitude of the output voltage $A(t)$ is varied with time, it is called amplitude modulation (AM). If the carrier frequency of $f(t)$ is varied with time, it is called frequency modulation (FM). If the phase of $\theta(t)$ is varied, it is called phase modulation (PM). In systems with analog modulation, $A(t), f_c(t)$, and $\theta(t)$ are continuous functions of time. In systems with digital modulation, $A(t), f_c(t)$, and $\theta(t)$ are discrete functions of time.

1.22.1 Amplitude Modulation

In AM, the carrier envelope is varied in proportion to the modulating signal $v_m(t)$. The carrier voltage is usually a sine wave

$$v_c(t) = V_c \cos\omega_c t \tag{1.163}$$

In general, the AM signal is

$$v_{AM}(t) = V(t)\cos\omega_c t = V_c[1+\alpha(t)]\cos\omega_c t = V_c\cos\omega_c t + \alpha(t)V_c\cos\omega_c t \tag{1.164}$$

where the envelope is

$$V(t) = V_c[1+\alpha(t)] \tag{1.165}$$

When the modulating voltage is a single-frequency sinusoid

$$v_m(t) = V_m\cos\omega_m t \tag{1.166}$$

the envelope is given by

$$V(t) = V_c + v_m(t) = V_c + V_m\cos\omega_m t = V_c\left(1+\frac{V_m}{V_c}\cos\omega_m t\right) = V_c(1+m\cos\omega_m t) \tag{1.167}$$

In this case, the AM voltage is

$$\begin{aligned} v_{AM}(t) = v_o(t) &= V(t)\cos\omega_c t = V_c\cos\omega_c t + v_m(t)\cos\omega_c t = [V_c + v_m(t)]\cos\omega_c t \\ &= V_c\cos\omega_c t + V_m\cos\omega_m t\cos\omega_c t = (V_c + V_m\cos\omega_m t)\cos\omega_c t \\ &= V_c\left(1+\frac{V_m}{V_c}\cos\omega_m t\right)\cos\omega_c t = V_c(1+m\cos\omega_m t)\cos\omega_c t \end{aligned} \tag{1.168}$$

where the *modulation index* (or the *modulation depth*) is

$$m = \frac{V_m}{V_c} = \frac{[V_{om(max)} - V_{om(min)}]/2}{[V_{om(max)} + V_{om(min)}]/2} = \frac{V_{om(max)} - V_{om(min)}}{V_{om(max)} + V_{om(min)}} \le 1 \tag{1.169}$$

and

$$\alpha(t) = \frac{V_m}{V_c}\cos\omega_m t = m\cos\omega_m t \tag{1.170}$$

The instantaneous amplitude $V(t)$ is directly proportional to the modulating voltage $v_m(t)$. At $\cos\omega_m t = 1$, $m = m_{max} = 1$, yielding

$$V_{o(MAX)} = V_c(1 + m_{max}) = V_c(1+1) = 2V_c \tag{1.171}$$

Figure 1.20 shows waveforms of modulating signal $v_m(t)$, carrier signal $v_c(t)$, and AM signal $v_{AM}(t)$ at $m = 0.8, f_m = 1$ kHz, and $f_c = 10$ kHz. For $m > 1$, the signal is overmodulated. When $V_m > V_c$, *overmodulation* occurs, causing distortion of the modulated signal. Figures 1.21–1.23 shows the waveforms of AM signals at $m = 0.5, 1,$ and 2.

Applying the trigonometric identity,

$$\cos\omega_c t\cos\omega_m t = \frac{1}{2}[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t] \tag{1.172}$$

we can express the AM signal as

$$\begin{aligned} v_{AM}(t) = v_o(t) &= V_c\cos\omega_c t + mV_c\cos\omega_m t\cos\omega_c t \\ &= V_c\cos\omega_c t + \frac{V_m}{2}\cos(\omega_c - \omega_m)t + \frac{V_m}{2}\cos(\omega_c + \omega_m)t \\ &= V_c\cos\omega_c t + \frac{mV_c}{2}\cos(\omega_c - \omega_m)t + \frac{mV_c}{2}\cos(\omega_c + \omega_m)t \end{aligned} \tag{1.173}$$

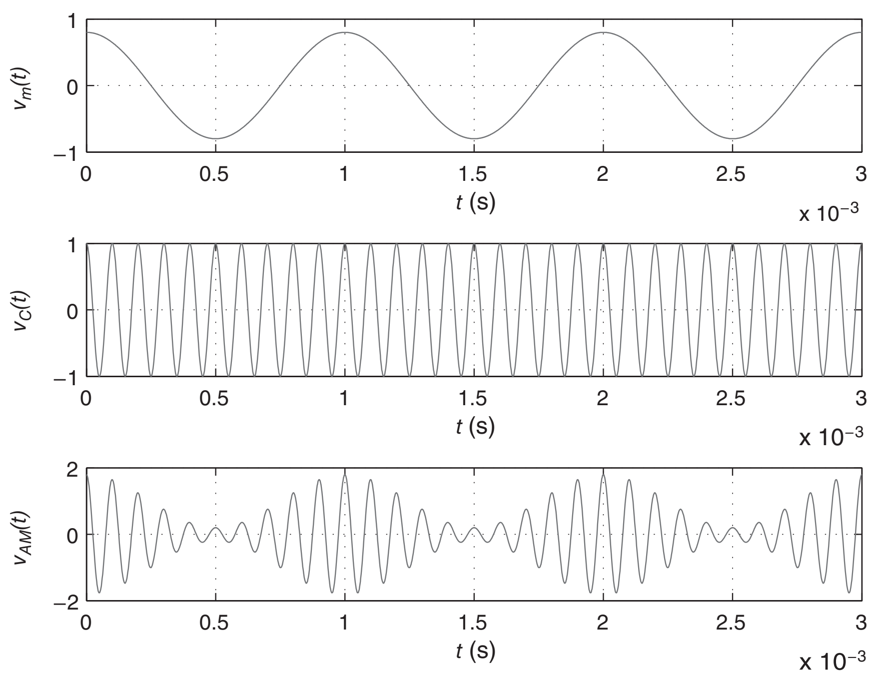


Figure 1.20 Waveforms of modulating signal $v_m(t)$, carrier signal $v_c(t)$, and AM signal $v_{AM}(t)$ at $m = 0.8, f_m = 1$ kHz, and $f_c = 10$ kHz.

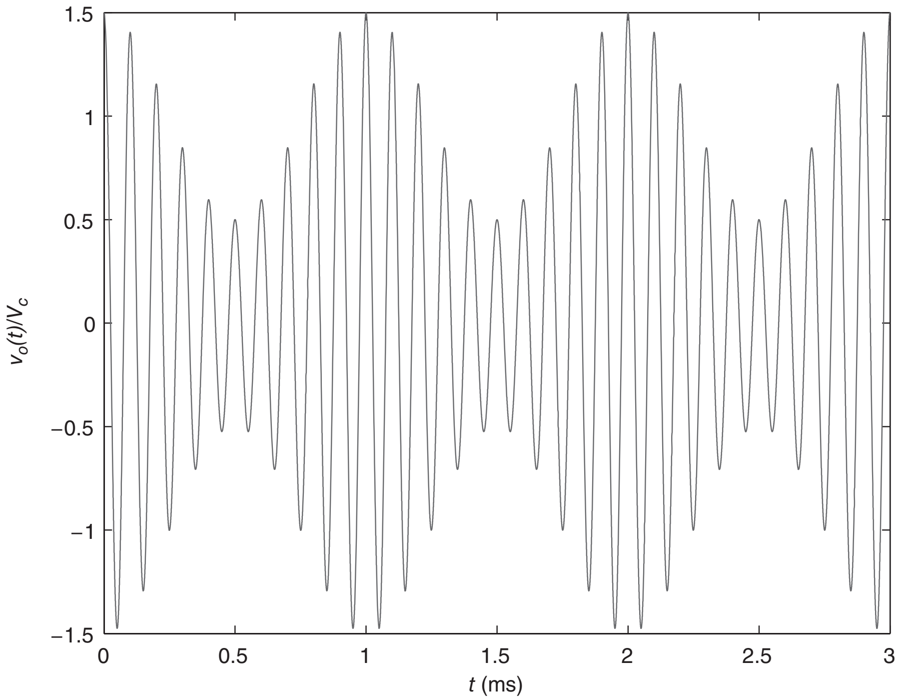


Figure 1.21 AM signal $v_o(t)/V_c$ modulated with a single sinusoid f_m at $m = 0.5$.

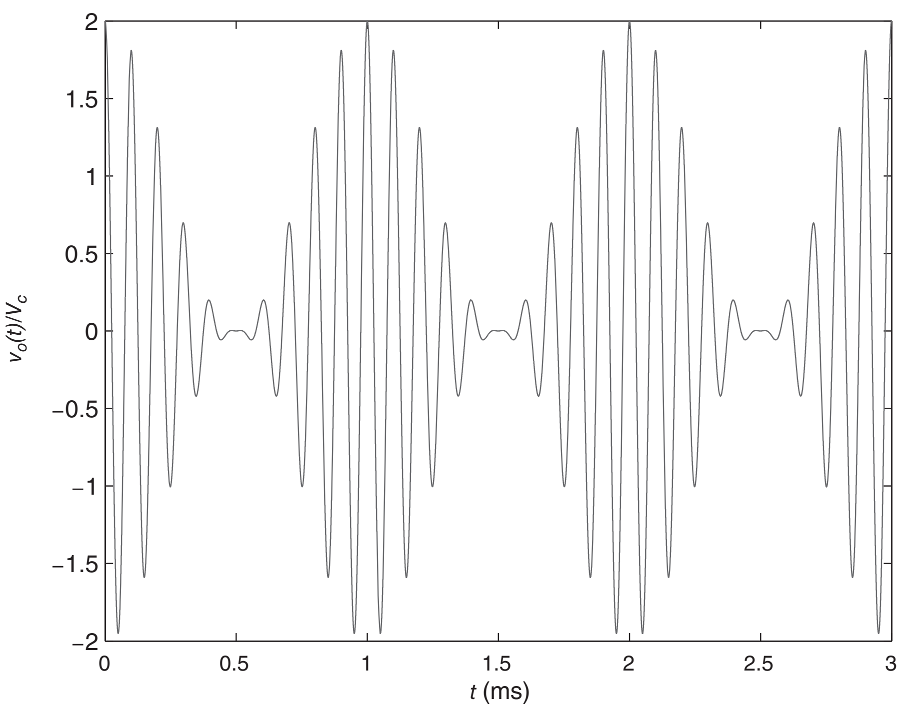


Figure 1.22 AM signal $v_o(t)/V_c$ modulated with a single sinusoid f_m at $m = 1$.

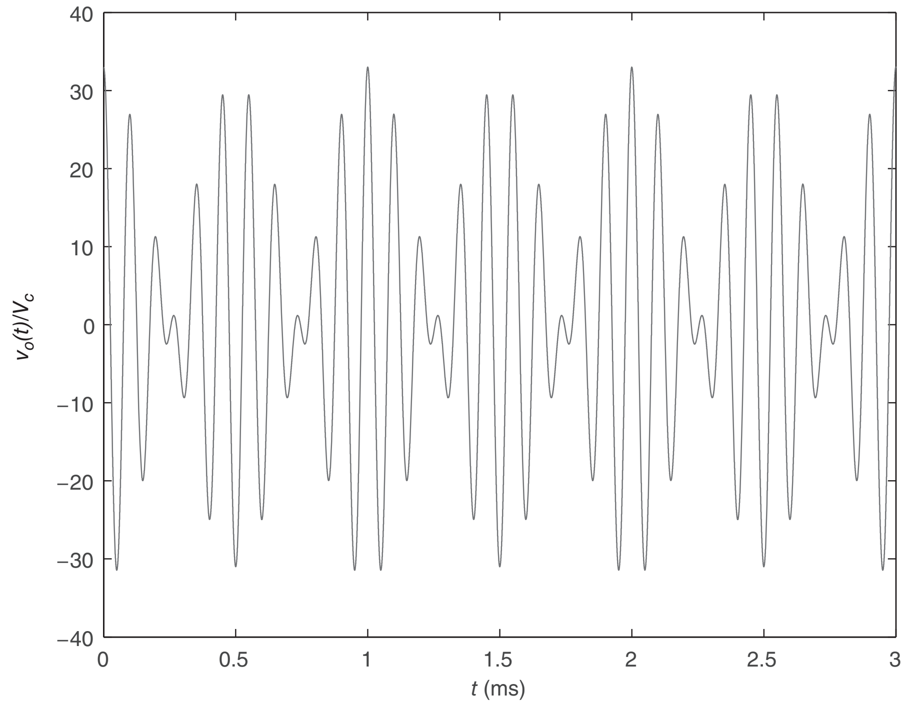


Figure 1.23 AM signal $v_o(t)/V_c$ modulated with a single sinusoid f_m at $m = 2$ (overmodulated signal).

The AM waveform modulated by a single tone consists of the following components:

1. the carrier component at frequency f_c,
2. the lower sideband component at frequency $f_c - f_m$, and
3. the upper sideband component at frequency $f_c + f_m$.

A phasor diagram for AM by a modulating voltage with a single-modulating frequency f_m is shown in Fig. 1.24. It can be seen that the maximum amplitude of the AM signal is $V_c + V_m$, when the sideband components are in phase with the carrier, and the minimum amplitude of the AM signal is $V_c - V_m$, when the sideband components are out of phase with respect to the carrier by 180°. If a band of frequencies is used as a modulating signal, we obtain the lower sideband and the upper sideband.

The bandwidth of an AM signal is given by

$$BW_{AM} = (f_c + f_m) - (f_c - f_m) = 2f_m \tag{1.174}$$

Hence, the maximum bandwidth of an AM signal is

$$BW_{AM(max)} = (f_c + f_{m(max)}) - (f_c - f_{m(max)}) = 2f_{m(max)} \tag{1.175}$$

Thus, the bandwidth of AM signal is determined by the maximum modulating frequency $f_{m(max)}$ and is independent of the minimum modulating frequency. The audio frequency range is typically from 20 Hz to 20 kHz. The frequency range of the human voice from 100 to 3000 Hz contains about 95% of the total energy. The carrier frequency f_c is much higher than $f_{m(max)}$,

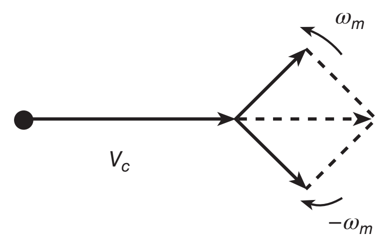


Figure 1.24 Phasor diagram for amplitude modulation by a single modulating frequency.

for example, $f_c/f_{m(max)} = 200$. The AM broadcasting is in the frequency range from 335 to 1605 kHz. It consists of 107 channels. The bandwidth of each channel is 10 kHz.

The time-average power of the carrier is

$$P_C = \frac{V_c^2}{2R} \tag{1.176}$$

The amplitude of modulating voltage is

$$V_m = \frac{mV_c}{2} \tag{1.177}$$

The time-average power of the lower sideband P_{LS} over the cycle of the modulating period $T_m = 1/f_m$ is equal to the average power of the upper sideband P_{US}

$$P_{LS} = P_{US} = \frac{V_m^2}{2R} = \frac{m^2 V_c^2}{8R} = \frac{m^2}{4} P_C \tag{1.178}$$

The total average power of the AM signal is given by

$$P_{AM} = P_C + P_{LS} + P_{US} = P_C + P_m = \left(1 + \frac{m^2}{4} + \frac{m^2}{4}\right) P_C = \left(1 + \frac{m^2}{2}\right) P_C \tag{1.179}$$

where

$$P_m = P_{LS} + P_{US} = \frac{m^2}{2} P_C \tag{1.180}$$

For $m = 1$, the total average power of the AM signal is

$$P_{AMmax} = P_C + P_{LS} + P_{US} = \left(1 + \frac{1}{4} + \frac{1}{4}\right) P_C = \frac{3}{2} P_C \tag{1.181}$$

The typical modulation index is $m = 0.25$, yielding the typical average power of the AM signal

$$P_{AM(typ)} = \left(1 + \frac{m^2}{2}\right) P_C = \left(1 + \frac{0.25^2}{2}\right) P_C = 1.03125 P_C \tag{1.182}$$

The instantaneous output power of the AM voltage modulated by a single sinusoidal voltage is

$$\begin{aligned} p_{o(AM)}(t) &= \frac{v_o^{2(t)}}{2R} = \frac{V_m^2(t)\cos^2\omega_c t}{2R} = \frac{V_c^2(1 + m\cos\omega_m t)^2 \cos^2\omega_c t}{2R} \\ &= P_C(1 + m\cos\omega_m t)^2 \cos^2\omega_c t \end{aligned} \tag{1.183}$$

The PEP is

$$P_{PEP} = P_C(1 + m)^2 \tag{1.184}$$

yielding the maximum PEP

$$P_{PEP(max)} = P_c(1 + m_{max})^2 = P_c(1 + 1)^2 = 4P_C \tag{1.185}$$

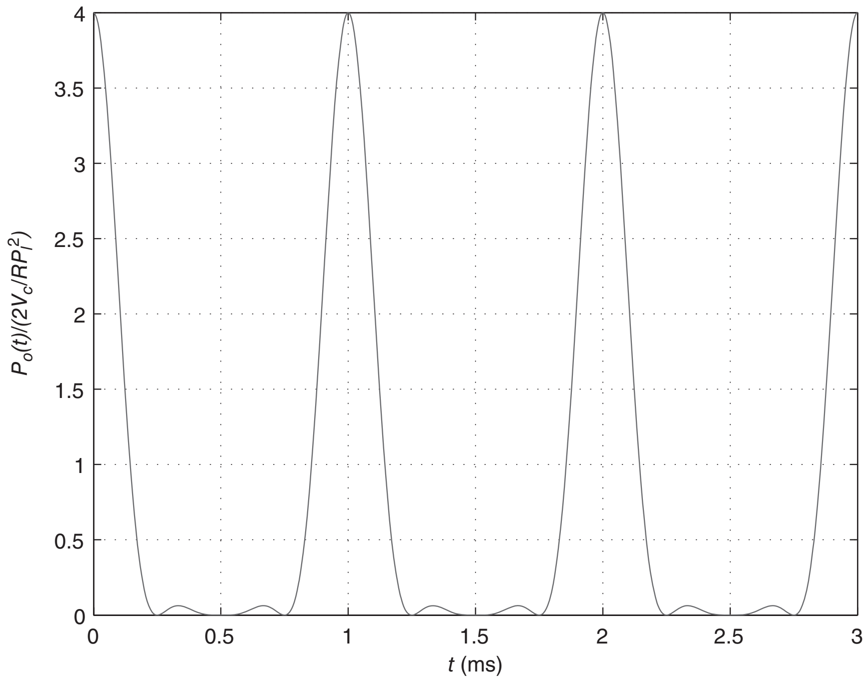


Figure 1.25 Instantaneous power of the AM signal $P_o(t)$ modulated with a single sinusoid f_m at $m = 1$.

Figure 1.25 shows the instantaneous power of the AM signal $P_o(t)$ modulated with a single sinusoid f_m at $m = 1$.

The efficiency of the AM is

$$\eta_{AM} = \frac{m^2}{2 + m^2} \tag{1.186}$$

resulting in

$$\eta_{AM(max)} = \frac{m^2}{2 + m^2} = \frac{m_{max}^2}{2 + m_{max}^2} = \frac{m_{max}^2}{2 + m_{max}^2} = \frac{1}{2 + 1} = \frac{1}{3} \tag{1.187}$$

For $m > 1$, $V_m > V_c$ and *overmodulation* of the AM signal occurs, causing distortion of the envelope. In this case, the shape of the detected signal in the receiver is different from the modulating signal.

Random electrical variations add to the AM signal and alter the original envelope of modulated signal. This is the most important disadvantage of AM systems.

A high-power AM voltage may be generated using the following methods:

- Amplification of the AM signal using a linear RF power amplifier.
- Drain amplitude modulation.
- Gate-to-source bias operating point modulation.

Variable-envelope signals, such as AM signals, usually require linear power amplifiers. Figure 1.26 shows the amplification process of an AM signal by a linear power amplifier. In this case, the carrier and the sidebands are amplified by a linear amplifier. Drain AM signal can be generated by connecting a modulating voltage source in series with the drain dc voltage

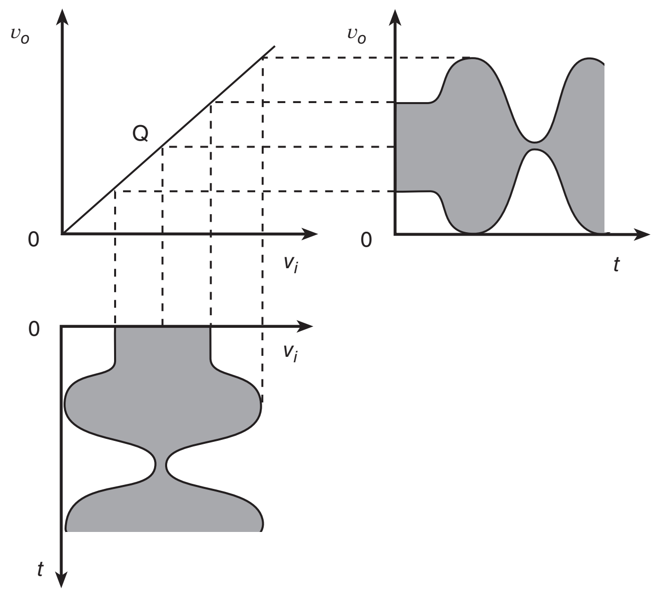


Figure 1.26 Amplification of AM signal by linear power amplifier.

supply V_I. When the MOSFET is operated as a current source such as in Class C power amplifiers, it should be driven into the ohmic region. If the transistor is operated as switch, the amplitude of the output voltage is proportional to the supply voltage and a high-fidelity AM signal is achieved. However, the modulating signal v_m must be amplified to a high power level before it is applied to modulate the supply voltage.

AM is used in commercial radio broadcasting and to transmit the video information in analog TV. Variable-envelope signals are also used in modern wireless communication systems.

Quadrature amplitude modulation (QAM) is a modulation method, which is based on modulating the amplitude of two orthogonal carrier sinusoidal waves. The two waves are out of phase with each other by 90° and, therefore, are called quadrature carriers. The QAM signal (I/Q signal) is expressed as in the Cartesian (or rectangular) form

$$v_{QAM} = I(t)\cos\omega_c t + Q(t)\sin\omega_c t \tag{1.188}$$

where $I(t)$ and $Q(t)$ are modulating signals. The QAM signal can also be expressed in the polar form

$$v_{QAM} = A(t)e^{j\phi(t)} \tag{1.189}$$

where the time-varying amplitude is

$$A(t) = \sqrt{[I(t)]^2 + [Q(t)]^2} \tag{1.190}$$

and the time-varying phase is

$$\phi(t) = \arctan\left[\frac{Q(t)}{I(t)}\right] \tag{1.191}$$

The received signal can be demodulated by multiplying the modulated signal v_{QAM} by a cosine wave of carrier frequency f_c

$$\begin{aligned} v_{DEM} &= v_{QAM}\cos\omega_c t = I(t)\cos\omega_c t\cos\omega_c t + Q(t)\sin\omega_c t\cos\omega_c t \\ &= \frac{1}{2}I(t) + \frac{1}{2}[I(t)\cos(2\omega_c t) + Q(t)\sin(2\omega_c t)] \end{aligned} \tag{1.192}$$

An LP filter removes the $2f_c$ components, leaving only the $I(t)$ term, which is uneffected by $Q(t)$. On the other hand, if the modulated signal v_{QAM} is multiplied by a sine wave and transmitted through an LP filter, one obtains $Q(t)$.

1.22.2 Phase Modulation

The output voltage with angle modulation is described by

$$v_o = V_c \cos[\omega_c t + \theta(t)] = V_c \cos \phi(t) \tag{1.193}$$

where the instantaneous phase is

$$\phi(t) = \omega_c t + \theta(t) \tag{1.194}$$

and the instantaneous angular frequency is

$$\omega(t) = \frac{d\phi(t)}{dt} = \omega_c + \frac{d\theta(t)}{dt} \tag{1.195}$$

Depending on the relationship between $\theta(t)$ and $v_m(t)$, the angle modulation can be categorized as follows:

- Phase modulation (PM).
- Frequency modulation (FM).

In PM systems, the phase of the carrier is changed by the modulating signal $v_m(t)$. In FM systems, the frequency of the carrier is changed by the modulating signal $v_m(t)$. The angle-modulated systems are inherently immune to amplitude fluctuations due to noise. In addition to the high degree of noise immunity, they require less RF power because the transmitter output power contains only the power of the carrier. Therefore, PM and FM systems are suitable for high-fidelity music broadcasting and for mobile radio wireless communications. However, the bandwidth of an angle-modulated signal is much wider than that of an AM signal.

The modulating voltage, also called the message signal, is expressed as

$$v_m(t) = V_m \sin \omega_m t \tag{1.196}$$

The phase of PM signal is given by

$$\theta(t) = k_p v_m(t) = k_p V_m \sin \omega_m t = m_p \sin \omega_m t \tag{1.197}$$

where the PM index is

$$m_p = k_p V_m \tag{1.198}$$

Hence, the PM output voltage is

$$\begin{aligned} v_{PM}(t) = v_o(t) &= V_c \cos[\omega_c t + k_p v_m(t)] = V_c \cos(\omega_c t + k_p V_m \sin \omega_m t) \\ &= V_c \cos(\omega_c t + m_p \sin \omega_m t) = V_c \cos(\omega_c t + \Delta\phi \sin \omega_m t) = V_c \cos \phi(t) \end{aligned} \tag{1.199}$$

where the *index of phase modulation* is the maximum phase shift caused by the modulating voltage V_m.

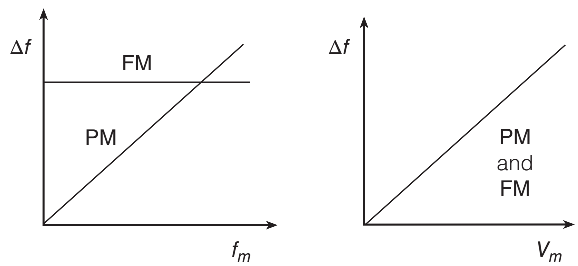
$$m_p = k_p V_m = \Delta\phi \tag{1.200}$$

Hence, the maximum PM is

$$\Delta\phi_{max} = k_p V_{m(max)} \tag{1.201}$$

Since

$$\phi(t) = \omega_c t + \theta(t) \tag{1.202}$$



(a) (b)

Figure 1.27 Frequency deviation Δf as functions of the modulating frequency f_m and the amplitude of the modulating voltage V_m: (a) frequency deviation Δf as a function of the modulating frequency f_m and (b) frequency deviation Δf as a function of the modulating voltage V_m.

the instantaneous frequency is

$$\omega(t) = \frac{d\phi(t)}{dt} = \omega_c + \frac{d\theta(t)}{dt} = \omega_c + k_p V_m \omega_m \cos \omega_m t = \omega_c + m_p \omega_m \cos \omega_m t \tag{1.203}$$

Thus, the phase change results in a frequency change. The frequency deviation is

$$\Delta f = f_{max} - f_c = f_c + k_p V_m f_m - f_c = k_p V_m f_m = m_p f_m \tag{1.204}$$

The frequency deviation Δf is directly proportional to the modulating frequency f_m. Figure 1.27 shows plots of the FM Δf as functions of the modulating frequency f_m and the amplitude of the modulating voltage V_m.

1.22.3 Frequency Modulation

The modulating voltage is

$$v_m(t) = V_m \sin \omega_m t \tag{1.205}$$

The instantaneous frequency is directly proportional to the amplitude of modulating voltage V_m

$$f(t) = f_c + \Delta f_{max} \frac{v_m(t)}{V_{m(max)}} = f_c + \Delta f_{max} \frac{V_m}{V_{m(max)}} \sin \omega_m t = f_c + \Delta f \sin \omega_m t \tag{1.206}$$

where Δf_{max} is the maximum frequency deviation from the carrier frequency (or rest frequency) f_c.

The derivative of the phase is

$$\frac{d\theta(t)}{dt} = 2\pi k_f v_m(t) = 2\pi k_f V - m \cos \omega_m t \tag{1.207}$$

Thus, the instantaneous phase is

$$\theta(t) = 2\pi \int k_f v_m(t) dt = k_f V_m \int \cos \omega_m t \, dt = \frac{k_f V_m}{f_m} \sin \omega_m t \tag{1.208}$$

In general, the waveform of an FM signal is given by

$$v_o(t) = V_c \cos \left(\omega t + \Delta\omega \int_0^t v_m(t) dt \right) \tag{1.209}$$

The FM output voltage is given by

$$v_{FM}(t) = v_o(t) = V_c \cos[\omega_c t + \theta(t)] = V_c \cos\left(\omega_c t + \frac{k_f V_m}{f_m} \sin \omega_m t\right)$$
$$= V_c \cos\left(\omega_c t + \frac{\Delta f}{f_m} \sin \omega_m t\right) = V_c \cos(\omega_c t + m_f \sin \omega_m t). \quad (1.210)$$

The *index of FM modulation* is defined as the ratio of the maximum frequency deviation Δf to the modulating frequency f_m

$$m_f = \frac{\Delta f}{f_m} = \frac{k_f V_m}{f_m} \quad (1.211)$$

Theoretically, the range of m_f is not limited: m_f can take on any value from 0 to infinity. The index of FM modulation m_f is a function of the modulating voltage V_m and the modulating frequency f_m. If $\Delta f = 75$ kHz, $m_f = 3750$ at $f_m = 20$ Hz and $m_f = 3.75$ at $f_m = 20$ kHz. The index of FM modulation m_f is a function of the modulating voltage amplitude V_m and the modulating frequency f_m.

The amplitude of the FM wave and, therefore, the power remain constant during the modulation process. Figure 1.28 shows that the waveform of the FM signal $v_o(t)/V_c$ modulated a single sinusoid f_m at $m_f = 10$ along with the modulating signal v_m.

The instantaneous frequency of an FM signal is given by

$$f(t) = f_c + k_f V_m \sin \omega_m t \quad (1.212)$$

where the frequency deviation of the FM signal is the maximum change of frequency caused by the modulating voltage V_m

$$\Delta f = f_{max} - f_c = f_c + k_f V_m - f_c = k_f V_m \quad (1.213)$$

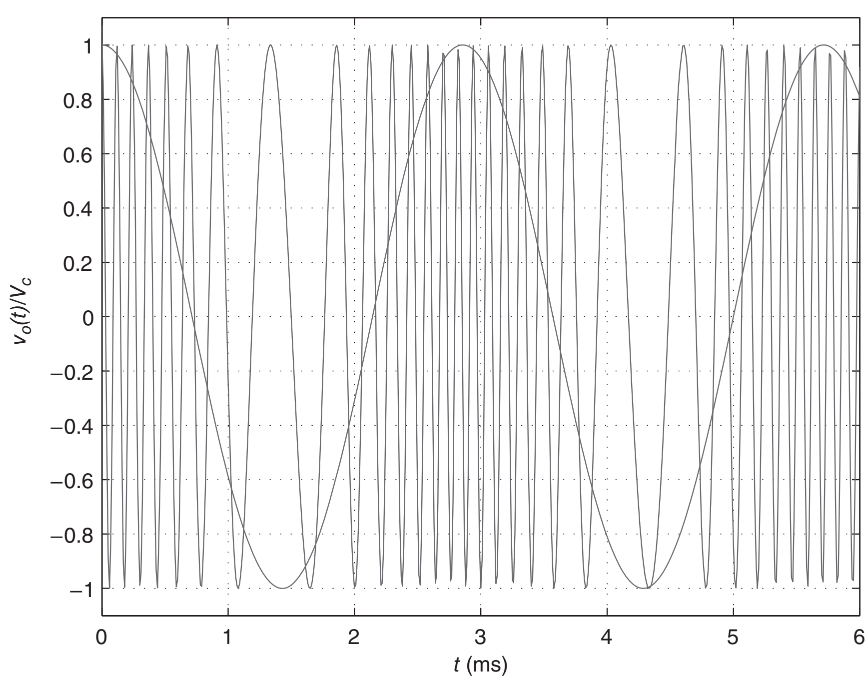


Figure 1.28 Waveforms of the modulating voltage $v_m(t)$ and the FM signal $v_o(t)/V_c$ modulated a single sinusoid f_m at $m_f = 10$.

The maximum frequency deviation is

$$\Delta f_{max} = k_f V_{m(max)} \tag{1.214}$$

The frequency deviation of the FM signal is independent of the modulating frequency f_m.

The angle-modulated signal contains the components at the frequencies

$$f_n = f_c \pm n f_m \qquad n = 0, 1, 2, 3, \ldots \tag{1.215}$$

Therefore, the bandwidth of an angle-modulated signal is infinite. However, the amplitudes of components with large values of n is very small. Hence, the effective bandwidth of the FM signal that contains 98% of the signal power is given by Carson's rule

$$BW_{FM} = 2(\Delta f + f_m) = 2\left(\frac{\Delta f}{f_m} + 1\right) f_m = 2(m_f + 1) f_m \tag{1.216}$$

yielding the maximum bandwidth of the FM signal

$$BW_{FM(max)} = 2(\Delta f + f_{m(max)}) \tag{1.217}$$

If the bandwidth of the modulating signal is BW_m, the FM index is

$$m_f = \frac{\Delta f_{max}}{BW_m} \tag{1.218}$$

and the bandwidth of the FM signal is

$$BW_{FM} = 2(m_f + 1) BW_m \tag{1.219}$$

The FCC's rules and regulations limit the FM broadcast band transmitters to a maximum frequency deviation of 75 kHz to prevent interference of adjacent channels. Commercial FM radio stations with an allowed maximum frequency deviation Δf of 75 kHz and a maximum modulation frequency f_m of 20 kHz require a bandwidth

$$BW_{FM} = 2 \times (75 + 20) = 190 \text{ kHz} \tag{1.220}$$

For $f_c = 100$ MHz and $\Delta f = 75$ kHz,

$$\frac{\Delta f}{f_c} = \frac{75 \times 10^3}{100 \times 10^6} = 0.075\% \tag{1.221}$$

Standard broadcast FM systems use a 200 kHz bandwidth for each station. There are guard bands between the channels.

In general,

$$\phi(t) = \int \omega(t) = 2\pi \int f(t) \tag{1.222}$$

and

$$\omega(t) = 2\pi f(t) = \frac{d\phi(t)}{dt} \tag{1.223}$$

The only difference between PM and FM is that for PM the phase of the carrier varies with the modulating signal and for FM the carrier phase depends on the ratio of the modulating signal amplitude V_m to the modulating frequency f_m. Thus, FM is not sensitive to the modulating frequency f_m, but PM is. If the modulating signal is integrated and used to modulate the carrier, then an FM signal is obtained. This method is used in the Armstrong's indirect FM system. The amount of deviation is proportional to the modulating frequency amplitude V_m.

The relationship between PM and FM is illustrated in Fig. 1.29. If the amplitude of the modulating signal applied to the phase modulator is inversely proportional to the modulating

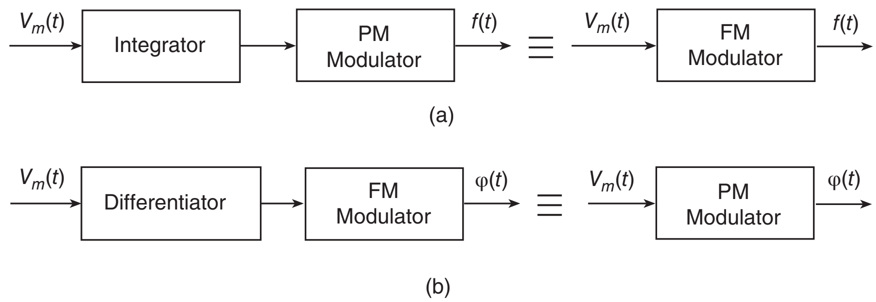


Figure 1.29 Relationship between FM and PM.

frequency f_m, then the phase modulator produces an FM signal. The modulating voltage is given by

$$v_m(t) = \frac{V_m}{\omega_m} \sin \omega_m t \tag{1.224}$$

The modulated voltage is

$$v_o(t) = V_v \cos\left(\omega_c t + \frac{k_p V_m}{\omega_m} \sin \omega_m t\right) = V_c \cos \phi(t) \tag{1.225}$$

Hence, the instantaneous frequency is

$$\omega(t) = \frac{d\phi(t)}{dt} = \omega_c + k_p V_m \cos \omega_m t \tag{1.226}$$

where

$$\Delta f = \frac{k_p V_m}{2\pi} \tag{1.227}$$

This approach is applied in the Armstrong's frequency modulator and is used for commercial FM transmission.

The output power of the transmitter with FM and PM is constant and independent of the modulation index because the envelope is constant and equal to the amplitude of the carrier V_c. Thus, the transmitter output power is

$$P_O = \frac{V_c^2}{2R} \tag{1.228}$$

FM is used for commercial radio and for audio in analog TV.

Both FM and PM signals modulated with a pure sine wave contain a sine as an argument of a cosine, having the modulation index $m = m_f = m_p$, and can be represented as

$$\begin{aligned} v_o(t) = V_c \cos(\omega_c t + m \sin \omega_m t) = V_c \{ & J_0(m) \cos \omega_c t \\ & + J_1(m)[\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t] + J_2(m)[\cos(\omega_c + 2\omega_m)t - \cos(\omega_c - 2\omega_m)t] \\ & + J_3(m)[\cos(\omega_c + 3\omega_m)t - \cos(\omega_c - 3\omega_m)t] + \cdots \} = \sum_{n=0}^{\infty} A_c J_n(m) \cos[2\pi(f_c \pm n f_m)t] \end{aligned} \tag{1.229}$$

where $n = 0, 1, 2, \ldots$ and $J_n(m)$ are the Bessel functions of the first kind of order n and can be described by

$$J_n(m) = \left(\frac{m}{2}\right) n \left[\frac{1}{n!} - \frac{(m/2)^2}{1!(n+1)!} + \frac{(m/2)^4}{2!(n+2)!} - \frac{(m/2)^6}{3!(n+3)!} + \cdots\right] = \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{m}{2}\right)^{n+2k}}{k!(n+k)!} \tag{1.230}$$

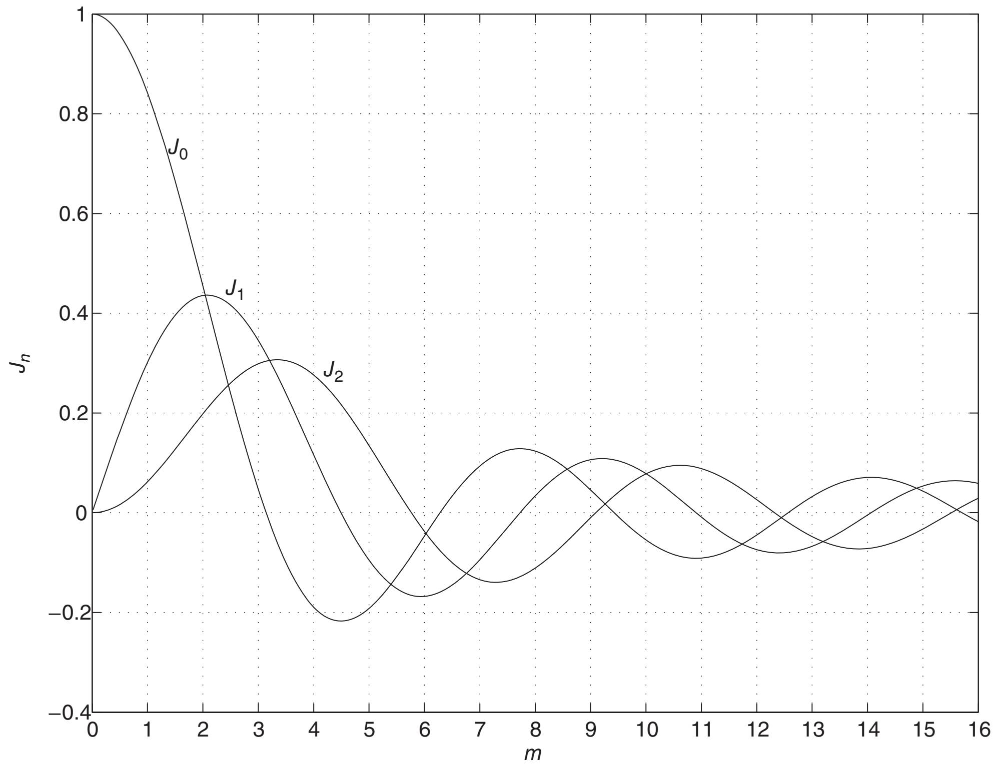


Figure 1.30 Amplitudes of spectrum components J_n for angle modulation as a function of modulation index m.

For $m \ll 1$,

$$J_n(m) \approx \frac{1}{n!}\left(\frac{m}{2}\right)^n \tag{1.231}$$

The angle-modulated voltage consists of a carrier and an infinite number of components (sidebands) placed at multiples of the modulating frequency f_m below and above the carrier frequency f_c, that is, at $f_c \pm nf_m$. The amplitude of the sidebands decreases with n, which allows FM transmission within a finite bandwidth. Figure 1.30 shows the amplitudes of spectrum components for angle modulation as functions of modulation index m. The amplitudes of the carrier (rest) frequency component and the sideband pairs are dependent on the modulation index m and are given by the Bessel function coefficients $J_n(m)$, where n is the order of the sideband pair. J_0 represents the rest-frequency amplitude, J_1 represents the amplitude of the first sideband pairs, and so on. The bandwidth of an FM signal may be written as

$$BW_{FM} = 2(f_{m(max)}) \times (\text{number of significant sideband pairs}) \tag{1.232}$$

The frequency range of the FM broadcasting systems is from 88 to 108 MHz. The FM business band is from 150 to 174 MHz. The average output power of the transmitted FM signal is

$$P_o = (J_0^2 + J_1^2 + J_2)2^2 + J_3^1)P_T \tag{1.233}$$

where P_T is the power of the nonmodulated signal, that is, at $m = 0$.

The angle-modulated signal is

$$v_o(t) = V_c \cos(\omega_c t + m\cos\omega_m t)) = V_c[\cos\omega_c t\cos(m\cos\omega_m t) - \sin\,\omega_c t\sin(m\cos\omega t)] \tag{1.234}$$

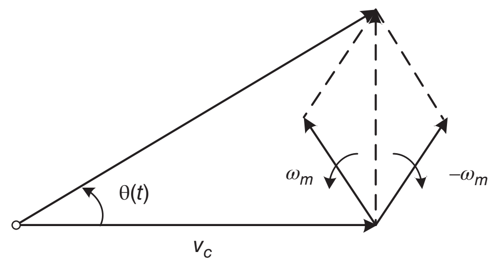


Figure 1.31 Phasor diagram for phase modulation by a single modulating frequency.

For $m \ll 1$, $\cos(m\cos\omega_m t) \approx 1$ and $\sin(m\cos\omega_m t) \approx m\cos\omega t$. Hence,

$$\begin{aligned} v_o(t) &\approx V_c\cos\omega_c t - m\sin\omega_c t\cos\omega_m t \\ &= V_c\cos\omega_c t - \frac{mV_m}{2}\sin[(\omega_c-\omega_m)t] - \frac{mV_m}{2}\sin[(\omega_c+\omega_m)t] \end{aligned} \tag{1.235}$$

Figure 1.31 depicts the phasor diagram for PM by a single modulating frequency.

1.23 Digital Modulation

Digital modulation is actually a special case of analog modulation. In RF systems with digital modulation, the carrier is modulated by a digital baseband signal, which consists of discrete values. Digital modulation offers many advantages over analog modulation and is widely used in wireless communications and digital TV. The main advantage is the quality of the signal, which is measured by the bit error rate (BER). BER is defined as the ratio of the average number of erroneous bits at the output of the demodulator to the total number of bits received in a unit of time in the presence of noise and other interference. It is expressed as a probability of error. Typically, BER $> 10^{-3}$. The waveform of a digital binary baseband signal is given by

$$v_D = \sum_{n=1}^{n=m} b_n v(t - nT_c) \tag{1.236}$$

where b_n is the bit value, equal to 0 or 1. Digital counterparts of analog modulation are amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK).

1.23.1 Amplitude-Shift Keying

Digital AM is called ASK or ON-OFF keying (OOK). It is the oldest type of modulation used in *radio telegraph transmitters*. The binary AM signal is given by

$$v_{BASK} = \begin{cases} V_c\cos\omega_c t & \text{for } v_m = 1 \\ 0 & \text{for } v_m = 0 \end{cases} \tag{1.237}$$

Waveforms illustrating the ASK are shown in Fig. 1.32. The ASK is sensitive to amplitude noise and, therefore, is rarely used in RF applications.

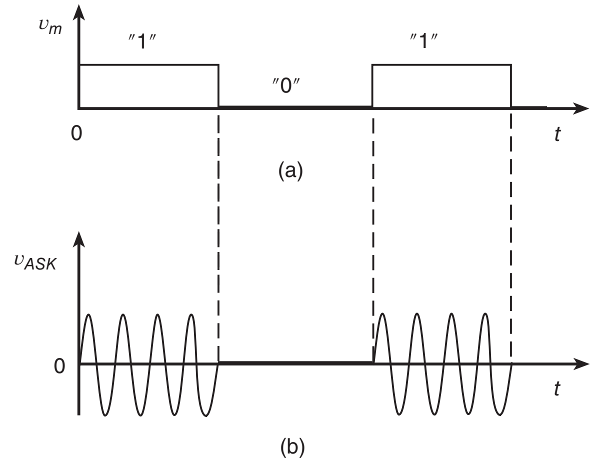


Figure 1.32 Waveform of amplitude-shift keying (ASK).

1.23.2 Phase-Shift Keying

Digital PM is called PSK. One of the most fundamental types of PSK is binary phase-shift keying (BPSK), where the phase shift is $\Delta\phi = 180°$. The BPSK has two phase states. Waveforms for BPSK are shown in Fig. 1.33. When the waveform is in phase with the reference waveform, it represents a logic "1." When the waveform is out of phase with respect to the reference waveform, it represents a logic "0." The modulating binary signal is

$$v_m = \begin{cases} 1 \\ -1 \end{cases} \tag{1.238}$$

The binary PM signal is given by

$$v_{BPSK} = v_m \times v_c = (\pm 1) \times v_c = \begin{cases} v_c = V_c \cos \omega_c t & \text{for } v_m = 1 \\ -v_c = -V_c \cos \omega_c t & \text{for } v_m = -1 \end{cases} \tag{1.239}$$

Another type of PSK is the quaternary phase-shift keying (QPSK), where the phase shift is 90°. The QPSK uses four phase states. The systems 8PSK and 16PSK are also widely used. The PSK is used in transmission of digital signals such as digital TV.

In general, the output voltage with digital PM is given by

$$v_o = V_c \sin \left[\omega_c t + \frac{2\pi(i-1)}{M} \right] \tag{1.240}$$

where $i = 1, 2, \ldots, M$, $M = 2^N$ is the number of phase states, N is the number of data bits needed to determine the phase state. For $M = 2$ and $N = 1$, a BPSK is obtained. For $M = 4$ and $N = 2$, we have a QPSK with the following combinations of digital logic: (00), (01), (11), and (10). For $M = 8$ and $N = 3$, the 8PSK is obtained.

1.23.3 Frequency-Shift Keying

The digital FM is called *FSK*. Waveforms for FSK are shown in Fig. 1.34. FSK uses two carrier frequencies, f_1 and f_2. The lower frequency f_2 may represent a logic "0" and the higher frequency

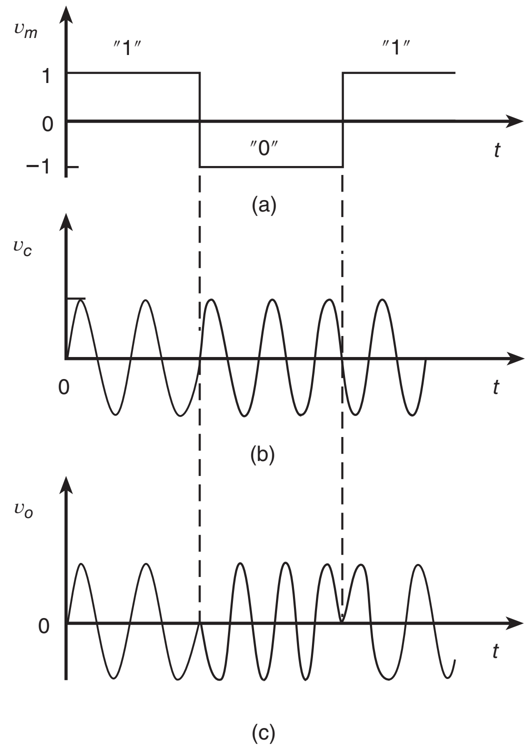


Figure 1.33 Waveforms for binary phase-shift keying (BPSK).

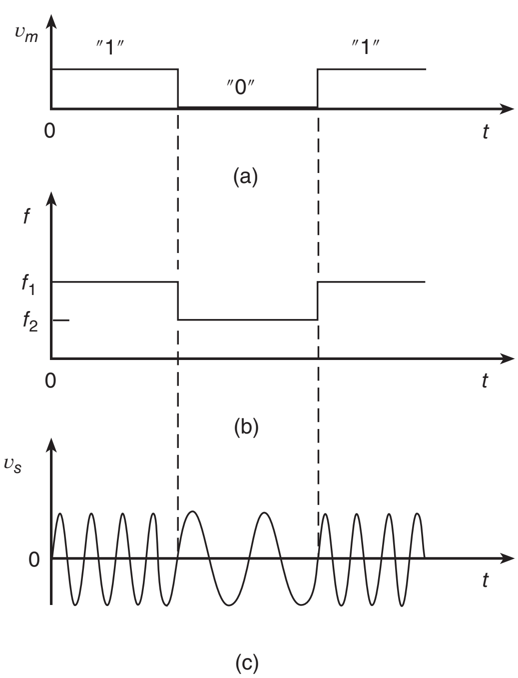


Figure 1.34 Waveforms for frequency-shift keying (FSK).

f_1 may represent a logic "1." The binary FM signal is given by

$$v_{BFSK} = \begin{cases} V_c \cos \omega_{c1} t & \text{for } v_m = 1 \\ V_c \cos \omega_{c2} t & \text{for } v_m = -1 \end{cases} \tag{1.241}$$

For example, GSM uses FSK to transmit a binary signal, where $f_1 = f_c + \Delta f = f_c + 77.708$ kHz and $f_1 = f_c - \Delta f = f_c - 77.708$ kHz, where f_c is the carrier center frequency. The channel spacing is 200 kHz. There are 8 users per channel.

Gaussian frequency-shift keying (GFSK) is used in the Bluetooth. A binary "one" is represented by a positive frequency deviation and a binary "zero" is represented by a negative frequency deviation. The GFSK employs a constant-envelope RF voltage. Thus, a high-efficiency, switching-mode RF power amplifiers can be used in transmitters.

1.24 Radars

Radar was developed during World War II for military applications. The term "radar" is derived from *ra*dio *d*etection *a*nd *r*anging. A radar consists of a transmitter, a receiver, and a transmitting and receiving antenna. A duplexer is used to separate the transmitting and receiving signals. A switch or a circulator can be used as a duplexer. A radar is a system used for detection and location of reflecting targets, such as aircraft, spacecraft, ships, vehicles, people, and other objects. It radiates EMWs (energy) into space and then detects the echo signal reflected from an object. The echo signal contains information about the location, range, direction, and velocity of an reflecting object. A radar can operate in darkness, fog, haze, rain, and snow. Only a very small portion of the transmitted energy is reflected by the object and reradiated back to the radar, where it is received, amplified, and processed.

Today, radars are used in military applications and in a variety of commercial applications, such as navigation, air traffic control, meteorology as weather radars, terrain avoidance, automobile collision avoidance, law enforcement (in speed guns), astronomy, RFID, and automobile automatic toll collection. Military applications of radars include surveillance and tracking of land, air, sea, and space objects from land, air, ship, and space platforms. In monostatic radar systems, the transmitter and the receiver are at the same location. The process of locating an object requires three coordinates: distance, horizontal direction, and angle of elevation. The distance to the object is determined from the time it takes for the transmitted waves to travel to the object and return back. The angular position of the object is found from the arrival angle of the returned signal. The relative motion of the object is determined from the Doppler shift in the carrier frequency of the returned signal. The radar antenna transmits a pulse of EM energy toward a target. The antenna has two basic purposes: radiates RF energy and provides beam focusing. A wide beam pattern is required for search and a narrow beam pattern is needed for tracking. A typical waveform of the radar signal is shown in Fig. 1.35. A radar transmitter typically operates in pulsed mode with an FM (chirped) signal. The amplitude of the pulses is usually constant. Therefore, nonlinear high-efficiency RF amplifiers can be used in radar transmitters. The efficiency of these amplifiers at peak power is usually very high.

The average power of the radar transmitter is

$$P_{AV} = DP_{pk} \tag{1.242}$$

where D is the duty ratio. The duty ratio of the waveform is very low, for example, 0.1%. Therefore, the ratio of the peak power to the average power is very large. For example, the peak RF power is $P_{pk} = 200$ kW and the average power is $P_{AV} = 200$ W. Some portion of this energy is reflected (or scattered) by the target. The echo signal (some of the reflected energy) is received by the radar antenna. The direction of the antenna's main beam determines the location of the

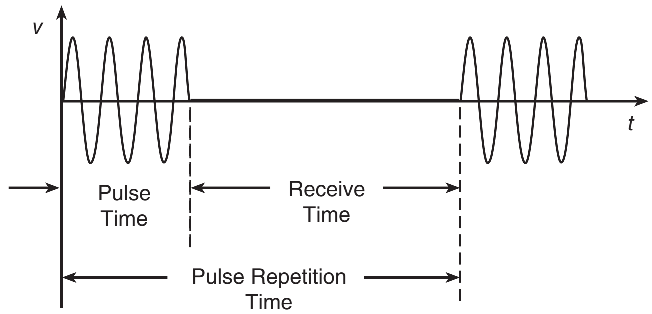


Figure 1.35 Waveform of radar signal.

target. The distance to the target is determined by the time between transmitting and receiving the EM pulse. The speed of the target with respect to the radar antenna is determined by the frequency shift in the EM signal, that is, the Doppler effect. A highly directive antenna is necessary. The radar equation is

$$\frac{P_{rec}}{P_{rad}} = \frac{\sigma_s \lambda^2}{(4\pi)^3 R^4} D^2 \tag{1.243}$$

where P_{rad} is the power transmitted by the radar antenna, P_{rec} is the power received by the radar antenna, σ_s is the radar cross section, λ is the free-space wavelength, and D is the directive gain of the antenna.

1.25 Radio-Frequency Identification

A RFID system consists of a *tag* placed on the item to be identified and a *reader* used to read the tag. The most common carrier frequency of RFID systems is 13.56 MHz, but it can be in the range of 135 kHz–5.875 GHz. The tags are passive or active. The ISO 1443A/B is the international standard for contactless IC cards transmitting at 13.56 MHz in close proximity using a radar antenna. A passive tag contains an RF rectifier, which rectifies a received RF signal from the reader. The rectified voltage is used to supply the tag circuitry. They exhibit long lifetime and durability. An active tag is supplied by an onboard battery. The reading range of active tags is much larger than that of passive tags and are more expensive. A block diagram of an RFID sensor system with a passive tag is shown in Fig. 1.36. The reader transmits a signal to the tag. The memory in the tag contains identification information unique to the particular tag. The microcontroller exports this information to the switch on the antenna. A modulated signal is transmitted to the reader and the reader decodes the identification information.

The tag may be placed in a plastic bag and attached to store merchandise to prevent theft or for checking the store inventory.

Tags may be mounted on windshields inside cars and used for automatic toll collection or checking the parking permit without stopping the car. Passive tags may be used for checking the entire inventory in libraries in seconds or locating a misplaced book. Tags along with pressure sensors may be placed in tires to alert the driver if the tire looses pressure. Small tags may be inserted beneath the skin for animal tracking. IRFD systems are capable of operating in adverse environments.

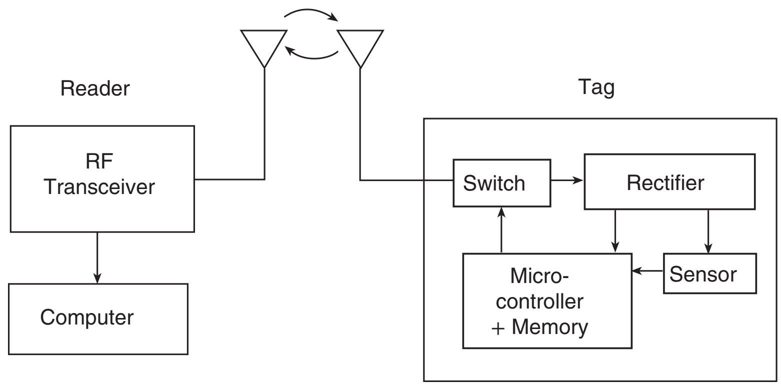


Figure 1.36 Block diagram of an RFID sensor system with a passive tag.

1.26 Summary

- In RF power amplifiers, the transistor can be operated as a dependent current source or as a switch.
- When the transistor is operated as a dependent current source, the drain current depends on the gate-to-source voltage. Therefore, variable-envelope signals can be amplified by the amplifiers in which transistors are operated as dependent current sources.
- When the transistor is operated as a switch, the drain current is independent of the gate-to-source voltage.
- When the transistor is operated as a switch, the drain-to-source voltage in the on-state is low, yielding low conduction loss and high efficiency.
- When the transistor is operated as a switch, switching losses limit the efficiency and the maximum operating frequency of RF power amplifiers.
- Efficiency of a power amplifier is maximized by minimizing power losses at a derided output power.
- Switching power loss can be reduced by using the ZVS technique or ZCS technique.
- In Class A, B, C, and F power amplifiers, the transistor is operated as a dependent current source.
- In Class D, E, and DE power amplifiers, the transistor is operated as a switch.
- When the drain current and the drain-to-source voltage are completely displaced, the output power of an amplifier is zero.
- ZVS, ZVDS, and ZCS conditions cannot be simultaneously satisfied with a passive load network at a nonzero output power.
- The received power P_{REC} is proportional to the square of the ratio $(\lambda/r)^2$. As the carrier's frequency doubles, the received power decreases by a factor of 4 at a given distance r in free space.

- RF modulated signals can be classified as variable-envelope signals and constant-envelope signals.
- The most important parameters of wireless communications systems are spectral efficiency, power efficiency, and signal quality.
- Power amplifiers dissipate more dc power than any other circuit in the transmitter of transceiver.
- The dBm is a method of rating power with respect to 1 mW of power.
- The carrier is a radio wave of constant amplitude, frequency, and phase.
- Power and bandwidth are two scarce resources in RF systems.
- The major analog modulation schemes are AM, FM, and PM.
- AM is the process of superimposing an LF modulating signal on a HF carrier so that the instantaneous changes in the amplitude of the modulating signal produce corresponding changes in the amplitude of the HF carrier.
- In AM transmission, two-thirds of the transmitted power is in the carrier at $m = 1$. No information is transmitted by the carrier.
- The main advantages of AM transmission are narrow bandwidth and simple circuits of transmitters and receivers.
- The main disadvantage of AM transmission is a high level of signal distortion and low efficiency.
- Constant-amplitude signals do not require linear amplification.
- FM is a the process of superimposing the modulating signal on a HF carrier so that the carrier frequency departs from the carrier frequency by an amount proportional to the amplitude of the modulating signal.
- Frequency deviation is the amount of carrier frequency increase or decrease around its center reference value.
- The main advantages of FM transmission are a high signal-to-noise ratio, constant envelope, high efficiency of the transmitters, and low-radiated power.
- Typically, the signal-to-noise ratio is 25 dB lower for FM transmission than that for AM transmission.
- The main disadvantages of FM transmission are a wide bandwidth and the complex circuits of transmitters and receivers.
- The bandwidth of FM transmission is about 10 times wider than that of the AM transmission.
- The major digital modulation schemes are ASK, FSK, and PSK.
- BPSK is a form of PSK in which the binary "1" is represented as no phase shift, and the binary "0" is represented by phase inversion of the carrier signal waveform.
- A transceiver is a combination of a transmitter and a receiver in one electronic device.
- Duplexing techniques allow users to transmit and receive signals simultaneously.
- The major multiple-access techniques are as follows: TDMA technique, FDMA technique, and CDMA technique.

1.27 Review Questions

1.1 What are the modes of operation of transistors in RF power amplifiers?

1.2 What are the classes of operation of RF power amplifiers?

1.3 What are the necessary conditions to achieve 100% efficiency of a power amplifier?

1.4 What are the necessary conditions to achieve nonzero output power in RF power amplifiers?

1.5 Explain the principle of operation of ZVS in power amplifiers.

1.6 Explain the principle of operation of ZCS in power amplifiers.

1.7 What is a ground wave?

1.8 What is a sky wave?

1.9 What is a line-of-sight wave?

1.10 What is an RF transmitter?

1.11 What is a transceiver?

1.12 What is duplexing?

1.13 List multiple-access techniques.

1.14 When are harmonics generated?

1.15 What is the effect of harmonics on communications channels?

1.16 Define THD?

1.17 What are intermodulation products?

1.18 When intermodulation products are generated?

1.19 What is the effect of intermodulation products on communications channels?

1.20 What is intermodulation distortion?

1.21 Define the 1-dB compression point.

1.22 What is the interception point?

1.23 What is the dynamic range of power amplifiers?

1.24 What determines the bandwidth of emission for AM transmission?

1.25 Define the modulation index for FM signals.

1.26 Give an expression for the total power of an AM signal modulated with a single-modulating frequency?

1.27 What is QAM?

1.28 What is phase modulation?

1.29 What is frequency modulation?

1.30 What is the difference between frequency modulation and phase modulation?

1.31 What is FSK?

1.32 What is PSK?

1.33 Explain the principle of operation of a radar.

1.34 Explain the principle of operation of an RFID system.

1.28 Problems

1.1 The rms value of the voltage at the carrier frequency is 100 V. The rms value of the voltage at the second harmonic of the carrier frequency is 1 V. The load resistance is 50 Ω. Neglecting all other harmonics, find *THD*.

1.2 The carrier frequency of an RF transmitter is 4.8 GHz. What is the height of a quarter-wave antenna?

1.3 An AM transmitter has an output power of 10 kW at the carrier frequency. The modulation index at $f_m = 1$ kHz is $m = 0.5$. Find the total output power of the transmitter.

1.4 The amplitude of the carrier of an AM transmitter is 25 V. The antenna input resistance is 50 Ω. The modulation index is $m = 0.5$.

(a) Find the amplitude of the modulating voltage.

(b) Find the output power of the AM transmitter.

1.5 A standard AM broadcast station is allowed to transmit waves with modulating frequencies up to 5 kHz. The carrier frequency is $f_c = 550$ kHz. Determine the following:

(a) the maximum upper sideband frequency,

(b) the minimum sideband frequency, and

(c) the bandwidth of the AM station.

1.6 An intermodulation product occurs at (a) $3f_1 - 2f_2$ and (b) $3f_1 + 2f_2$. What is the order of intermodulation?

1.7 The carrier frequency of an RF transmitter is 2.4 GHz. What is the height of a quarter-wave antenna?

1.8 The peak power of a radar transmitter is 100 kW. The duty ratio is 0.1%. What is the average power of the radar transmitter?

1.9 Find the total output power of the AM voltage when the carrier power is 10 W and $m = 1$.

1.10 Sketch the modulating, carrier, and AM voltage waveforms at $m = 0.5$.

1.11 Find the power of each sideband for a sinusoidal modulating voltage at $m = 0.3$ and the power of the carrier of 1 kW.

1.12 Find the channel bandwidth of an AM transmitter with a maximum modulating frequency of 10 kHz.

1.13 A transmitter applies an FM signal to a 50 Ω antenna. The amplitude of the signal is $V_m = 1000$ V. What is the output power of the transmitter?

1.14 An FM signal is modulated with the modulating frequency $f_m = 10$ kHz and has a maximum frequency deviation $\Delta f = 20$ kHz. Find the modulation index.

1.15 An FM signal has a carrier frequency $f_c = 100.1$ MHz, modulation index $m_f = 2$, and modulating frequency $f_m = 15$ kHz. Find the bandwidth of the FM signal.

1.16 How much bandwidth is necessary to transmit a Chopin's piano sonata over an FM radio broadcasting system with high fidelity?

References

[1] K. K. Clarke and D. T. Hess, *Communications Circuits: Analysis and Design*. Reading, MA: Addison-Wesley Publishing Co., 1971; reprinted Malabar, FL: Krieger, 1994.

[2] H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York, NY, John Wiley & Sons, 1980.

[3] M. Kazimierczuk, "Class E tuned amplifier with shunt inductor," *IEEE Journal of Solid-State Circuits*, vol. 16, no. 1, pp. 2–7, 1981.

[4] E. D. Ostroff, M. Borakowski, H. Thomas, and J. Curtis, *Solid-State Radar Transmitters*, Boston, MA:Artech House, 1985.

[5] M. K. Kazimierczuk and D. Czarkowski, *Resonant Power Converters*, 2nd Ed. New York, NY: IEEE Press/John Wiley & Sons, 2011.

[6] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd Ed. Norwood, MA: Artech House, 2006.

[7] M. Albulet, *RF Power Amplifiers*. Atlanta, GA: Noble Publishing Co., 2001.

[8] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd Ed. New York, NY: Cambridge University Press, 2004.

[9] P. Reynaret and M. Steyear, *RF Power Amplifiers for Mobile Communications*. Dordrecht, The Netherlands: Springer, 2006.

[10] H. I. Bartoni, *Radio Propagation for Modern Wireless Systems*. Englewood Clifs, NJ: Prenice-Hall, 2000.

[11] N. Levanon and E. Mozeson, *Radar Signals*, 3rd Ed. New York, NY: John Wiley & Sons, 2004.

[12] H. T. Friis, "A note on a simple transmission formula," *IEEE IRA*, vol. 41, pp. 254–256, 1966.

[13] B. Molnar, "Basic limitations of waveforms achievable in single-ended switching-mode tuned (Class E) power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. SC-19, pp. 144–146, 1984.

[14] M. K. Kazimierczuk, "Generalization of conditions for 100-percent efficiency and nonzero output power in power amplifiers and frequency multipliers," *IEEE Transactions on Circuits and Systems*, vol. CAS-33, pp. 805–807, 1986.

[15] F. H. Raab, "Intermodulation disrortion in Khan technique transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, pp. 2273–2278, 1996.

[16] M. K. Kazimierczuk and N. O. Sokal, "Cause of instability of power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 19, pp. 541–542, 1984.

[17] B. Razavi, *RF Microelectronics*. Upper Saddle River, NJ: Prentice Hall, 1989.

[18] P. B. Kenington, *High-Linearity RF Power Amplifier Design*. Norwood, MA: Artech House, 2000.

[19] J. Groe and L. Larson, *CDMA Mobile Radio Design*. Norwood, MA: Artech House, 2000.

[20] A. Grebennikov, *RF and Microwave Power Amplifier Design*. New York, NY: McGraw-Hill, 2005.

[21] A. Grebennikov and N. O. Sokal, *Switchmode RF Power Amplifiers*. Amsterdam: Amsterdam, 2007.

[22] P. Colantonio, F. Giannini, and E. Limiti, *High Efficiency RF and Microwave Solid State Power Amplifiers*. Chichester, UK: John Wiley & Sons, 2009.

[23] J. L. B. Walker, Edited, *Handbook of RF and Microwave Power Amplifiers*. Cambridge, UK: Cambridge University Press, 2012.

[24] A. Raghavan, N. Srirattana, and J. Lasker, *Modeling and Design Techniques for RF Power Amplifiers*. New York, NY: IEEE Press, John Wiley & Sons, 2008.

[25] R. Baxley and G. T. Zhou, "Peak savings analysis of peak-to-average power ratio in OFDM," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 3, pp. 792–798, 1998.

[26] B. S. Yarman, *Design of Ultra Wideband Power Transfer Networks*. Berlin, Springer-Verlag, 2008.

[27] M. K. Kazimierczuk, *Pulse-width Modulated DC-DC Power Converters*. 2nd Ed. Chichester, UK: John Wiley & Sons, 2014.

[28] M. K. Kazimierczuk, *High-Frequency Magnetic Components*, 2nd Ed. Chichester, UK: John Wiley & Sons, 2014.