1

Introduction

1.1 Block Diagram of RF Power Amplifiers

A power amplifier [1–15] is a key element to build a wireless communication system successfully. To minimize interference and spectral re-growth, transmitters must be linear. A block diagram of an RF power amplifier is shown in Figure 1.1. It consists of a transistor (MOSFET, MESFET, or BJT), output network, input network, and RF choke. In RF power amplifiers, a transistor can be operated

- as a dependent-current source;
- as a switch;
- 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}, i.e., $v_{DS} > V_{DSmin} = V_{GS} - V_t$, 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

RF Power Amplifiers Marian K. Kazimierczuk

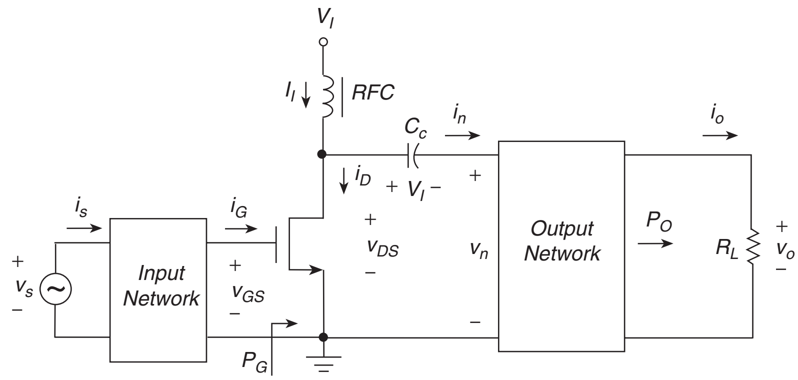

Figure 1.1 Block diagram of RF power amplifier.

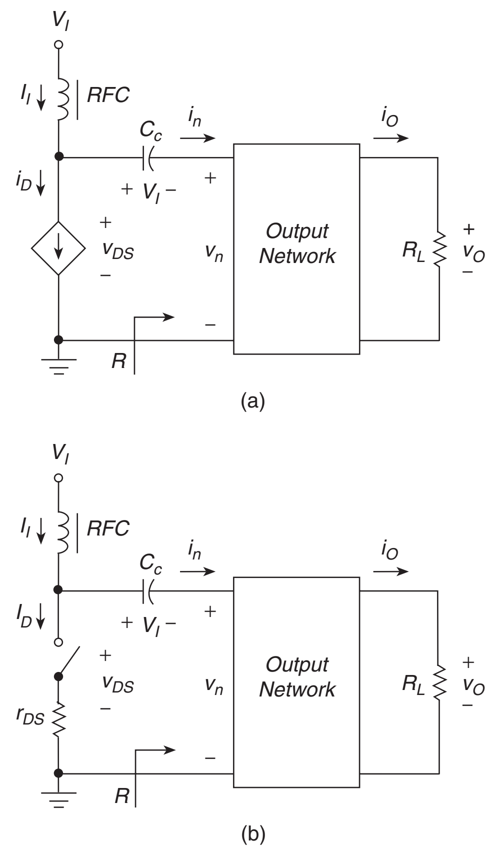

Figure 1.2 Modes of operation of transistor in RF power amplifiers. (a) Transistor as a dependent current source. (b) Transistor as a switch.

of operation is suitable for linear power amplifiers. Amplitude linearity is important for amplitude-modulated (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, also called the pinch-off 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 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 v_{DS}, resulting in low power loss.

If the transistor is driven by 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:

- impedance transformation;
- harmonic suppression;
- filtering of the spectrum of a signal with bandwidth BW to avoid interference with communication signals in adjacent channels.

1.2 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. 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 Figure 1.3. The operating points for various classes of operation are shown in Figure 1.4.

In Class A, the conduction angle 2θ is 360°. The gate-to-source voltage v_{GS} must be higher than the transistor threshold voltage V_t, i.e., $v_{GS} > V_t$. 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 v_{GS}. The dc 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.

In Class B, the conduction angle 2θ 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.

In Class C, the conduction angle 2θ of the drain current 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.

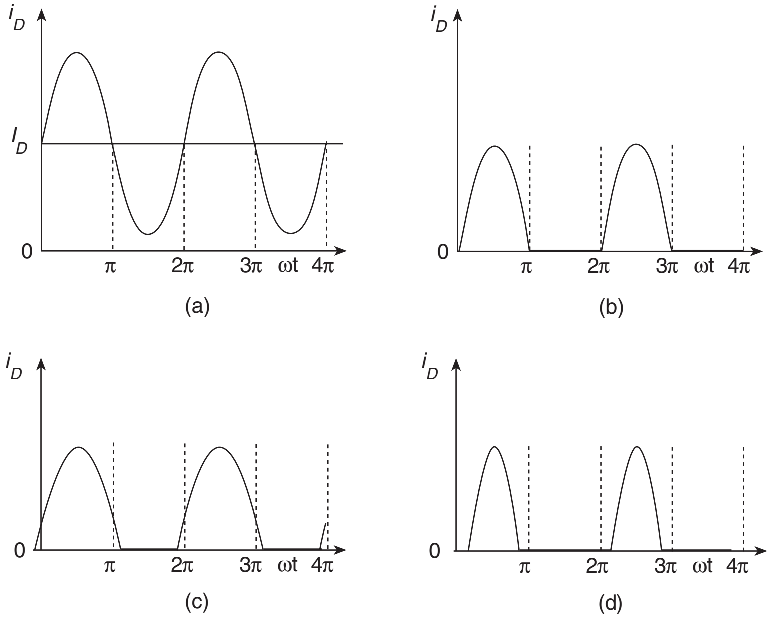

Figure 1.3 Waveforms of the drain current i_D in various classes of operation. (a) Class A. (b) Class B. (c) Class AB. (d) Class C.

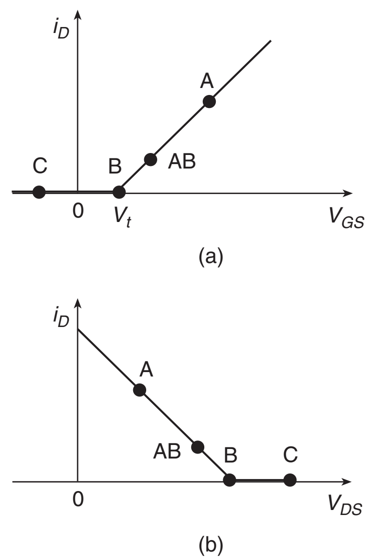

Figure 1.4 Operating points for Classes A, B, AB, and C.

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

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.

1.3 Parameters of RF Power Amplifiers

The dc supply power of an amplifier is

$$P_I = I_I V_I. \tag{1.1}$$

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

$$P_{DS} = \frac{1}{2} I_m V_m = \frac{1}{2} I_m^2 R = \frac{V_m^2}{2R} \tag{1.2}$$

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

$$P_{DS} = \frac{1}{2} I_m V_m \cos\phi = \frac{1}{2} I_m^2 R \cos\phi = \frac{V_m^2 \cos\phi}{2R} \tag{1.3}$$

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

The power level is often referenced to 1 mW and is expressed as

$$P = 10 \log \frac{[P(\mathrm{W})]}{0.001} \text{ (dBm)} = -30 + 10 \log[P(\mathrm{W})](\mathrm{dBW}). \tag{1.4}$$

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

The instantaneous drain power dissipation is

$$p_D(\omega t) = i_D v_{DS}. \tag{1.5}$$

The drain power dissipation is

$$P_D = \frac{1}{2\pi} \int_0^{2\pi} p_D \, \mathrm{d}(\omega t) = \frac{1}{2\pi} \int_0^{2\pi} i_D v_{DS} \, \mathrm{d}(\omega t) = P_I - P_{DS}. \tag{1.6}$$

The drain efficiency is

$$\eta_D = \frac{P_{DS}}{P_I} = \frac{P_I - P_D}{P_I} = 1 - \frac{P_D}{P_I}. \tag{1.7}$$

The gate-drive power is

$$P_G = \frac{1}{2\pi} \int_0^{2\pi} i_G v_{GS} \, \mathrm{d}(\omega t). \tag{1.8}$$

For sinusoidal gate current and voltage,

$$P_G = \frac{1}{2} I_{gm} V_{gsm} \cos\phi_G = \frac{R_G I_{gm}^2}{2} \tag{1.9}$$

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 output power is

$$P_O = \frac{1}{2} I_{om} V_{om} = \frac{1}{2} I_{om}^2 R_L = \frac{V_{om}^2}{2R_L}. \tag{1.10}$$

The power loss in the resonant output network is

$$P_r = P_{DS} - P_O. \tag{1.11}$$

The efficiency of the resonant output network is

$$\eta_r = \frac{P_O}{P_{DS}}. \tag{1.12}$$

The overall power loss on the output side of the amplifier is

$$P_{Loss} = P_I - P_O = P_D + P_r. \tag{1.13}$$

The overall efficiency of the amplifier is

$$\eta = \frac{P_O}{P_I} = \frac{P_O}{P_{DS}}\frac{P_{DS}}{P_I} = \eta_D \eta_r. \tag{1.14}$$

The average efficiency is

$$\eta_{AV} = \frac{P_{O(AV)}}{P_{I(AV)}} \tag{1.15}$$

where $P_{O(AV)}$ is the average output power and $P_{I(AV)}$ is the average dc supply power over a specified period of time. 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 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 $V_m = V_I/2$.

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

$$\eta_{PAE} = \frac{\text{Output power} - \text{Drive power}}{\text{DC supply power}} = \frac{P_O - P_G}{P_I} = \frac{P_O}{P_I}\left(1 - \frac{1}{k_p}\right) = \eta\left(1 - \frac{1}{k_p}\right). \tag{1.16}$$

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

The output power capability is given by

$$c_p = \frac{P_{O(max)}}{NI_{DM}V_{DSM}} = \frac{\eta P_I}{NI_{DM}V_{DSM}} = \frac{1}{2N}\left(\frac{I_I}{I_{DM}}\right)\left(\frac{V_I}{V_{DSM}}\right)$$
$$= \frac{1}{2N}\left(\frac{I_m}{I_I}\right)\left(\frac{I_I}{I_{DM}}\right)\left(\frac{V_m}{V_I}\right)\left(\frac{V_I}{V_{DSM}}\right) \tag{1.17}$$

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}, and N is the number of transistors in an 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

$$P_{O(max)} = c_p I_{DM} V_{DSM}. \tag{1.18}$$

As the output power capability c_p increases, the maximum output power $P_{O(max)}$ also increases. The output power capability is useful for comparing different types or families of amplifiers.

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

Example 1.1

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

Solution. The efficiency of the power amplifier is

$$\eta = \frac{P_O}{P_I} = \frac{10}{20} = 50\,\%. \tag{1.19}$$

The power-added efficiency is

$$\eta = \frac{P_O - P_G}{P_I} = \frac{10-1}{20} = 45\,\%. \tag{1.20}$$

The power gain is

$$k_p = \frac{P_O}{P_G} = \frac{10}{1} = 10 = 10\,\log(10) = 10\,\text{dB}. \tag{1.21}$$

1.4 Conditions for 100 % Efficiency of Power Amplifiers

The drain efficiency is given by

$$\eta_D = \frac{P_{DS}}{P_I} = 1 - \frac{P_D}{P_I}. \tag{1.22}$$

The condition for achieving a drain efficiency of 100 % is

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS}\,\mathrm{d}t = 0. \tag{1.23}$$

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

$$i_D v_{DS} = 0. \tag{1.24}$$

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 Figure 1.5.

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., like in a Class C 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 Figure 1.6(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 Figure 1.6(b).

For the case of Figure 1.6(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

$$v_{DS}(t_o) = \frac{1}{2}\left[\lim_{t\to t_O^-} v_{DS}(t) + \lim_{t\to t_o^-}(t)\right] = \frac{\Delta V}{2}. \tag{1.25}$$

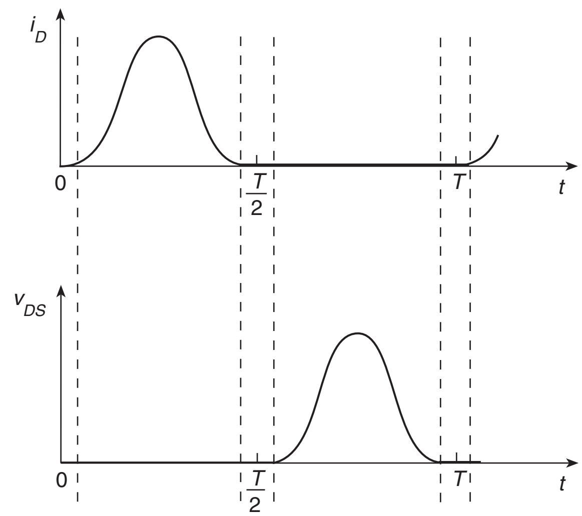

Figure 1.5 Nonoverlapping waveforms of drain current i_D and drain-to-source voltage v_{DS}.

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

$$i_D(t_o) = C\,\Delta V\,\delta(t - t_o). \tag{1.26}$$

Hence, the instantaneous power dissipation is

$$p_D(t) = i_D v_{DS} = \begin{cases} \frac{1}{2}C\,\Delta V^2\delta(t - t_o) & \text{for} \quad t = t_o \\ 0 & \text{for} \quad t \neq t_o, \end{cases} \tag{1.27}$$

resulting in the time average power dissipation

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS}\,\mathrm{d}t = \frac{1}{2} fC\,\Delta V^2 \tag{1.28}$$

and the drain efficiency

$$\eta_D = 1 - \frac{P_D}{P_I} = 1 - \frac{fC\,\Delta V^2}{2P_I}. \tag{1.29}$$

In a real circuit, the switch has a small series resistance, and the current through the switch is an exponential function of time of finite peak value.

Example 1.2

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

Solution. The switching power loss is

$$P_D = \frac{1}{2} fC\,\Delta V_{DS}^2 = \frac{1}{2} \times 2.4 \times 10^9 \times 100 \times 10^{-12} \times 5^2 = 3\,\mathrm{W}. \tag{1.30}$$

Hence, the drain efficiency of the amplifier is

$$\eta_D = \frac{P_O}{P_I} = \frac{P_I - P_D}{P_I} = 1 - \frac{P_D}{P_I} = 1 - \frac{3}{5} = 40\,\%. \tag{1.31}$$

For the amplifier of Figure 1.6(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

$$i_D(t_o) = \frac{1}{2}\left[\lim_{t \to t_o^-} i_D(t) + \lim_{t \to t_o^+} i_D(t)\right] = \frac{\Delta I}{2}. \tag{1.32}$$

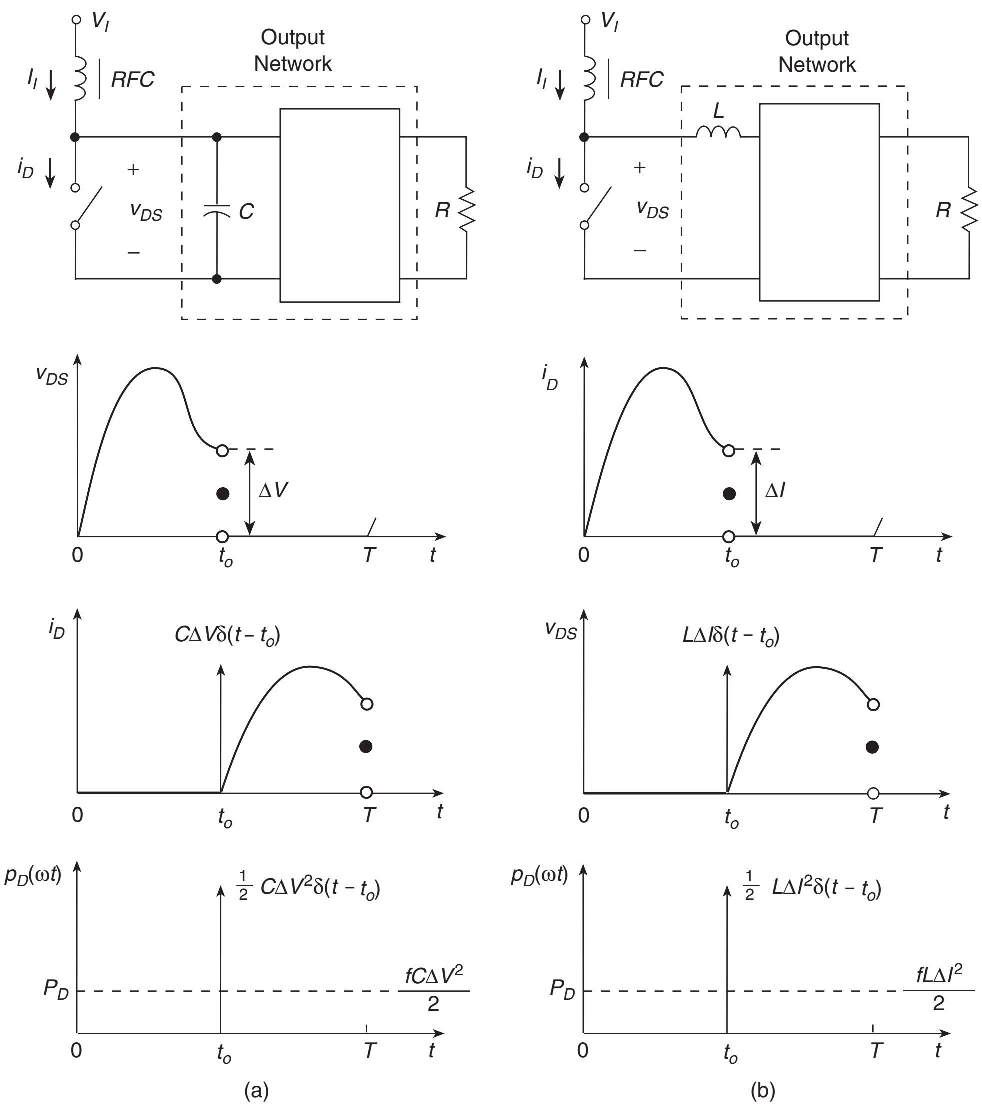

Figure 1.6 Waveforms of drain current i_D and drain-to-source voltage v_{DS} with delta Dirac functions. (a) Circuit with the switch in parallel with a capacitor. (b) Circuit with the switch in series with an inductor.

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

$$v_{DS}(t_o) = L\Delta I\,\delta(t - t_o). \tag{1.33}$$

Hence, the instantaneous power dissipation is

$$p_D(t) = i_D v_{DS} = \begin{cases} \frac{1}{2}L\Delta I^2\delta(t - t_o) & \text{for} \quad t = t_o \\ 0 & \text{for} \quad t \neq t_o, \end{cases} \tag{1.34}$$

resulting in the time average power dissipation

$$P_D = \frac{1}{T}\int_0^T i_D v_{DS}\,\mathrm{d}t = \frac{1}{2}fL\Delta I^2 \tag{1.35}$$

and the drain efficiency

$$\eta_D = 1 - \frac{P_D}{P_I} = 1 - \frac{fL\Delta I^2}{2P_I}. \tag{1.36}$$

In reality, the switch in the off-state has a large parallel resistance and voltage with a finite peak value developed across the switch.

1.5 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 $P_O > 0$ [9, 10]. The drain current i_D and the drain-to-source voltage v_{DS} can be represented by the Fourier series

$$i_D = I_I + \sum_{n=1}^{\infty} i_{dn} = I_I + \sum_{n=1}^{\infty} I_{dn}\sin(n\omega t + \psi_n) \tag{1.37}$$

and

$$v_{DS} = V_I + \sum_{n=1}^{\infty} v_{dsn} = V_I + \sum_{n=1}^{\infty} V_{dsn}\sin(n\omega t + \vartheta_n). \tag{1.38}$$

The derivatives of these waveforms with respect to time are

$$i_D' = \frac{\mathrm{d}i_D}{\mathrm{d}t} = \omega\sum_{n=1}^{\infty} nI_{dn}\cos(n\omega t + \psi_n) \tag{1.39}$$

and

$$v_{DS}' = \frac{\mathrm{d}v_{DS}}{\mathrm{d}t} = \omega\sum_{n=1}^{\infty} nV_{dsn}\cos(n\omega t + \vartheta_n). \tag{1.40}$$

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

$$\frac{1}{T}\int_0^T i_D' v_{DS}'\,\mathrm{d}t = -\frac{\omega^2}{2}\sum_{n=1}^{\infty} n^2 I_{dn}V_{dsn}\cos\phi_n = -\omega^2\sum_{n=1}^{\infty} n^2 P_{dsn} \tag{1.41}$$

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

$$\sum_{n=1}^{\infty} n^2 P_{dsn} = -\frac{1}{4\pi^2 f}\int_0^T i_D' v_{DS}'\,\mathrm{d}t. \tag{1.42}$$

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

$$P_{ds1} = P_{O1} = -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\, \mathrm{d}t. \tag{1.43}$$

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

$$P_{dsn} = P_{On} = -\frac{1}{4\pi^2 n^2}\int_0^T i'_D v'_{DS}\, \mathrm{d}t. \tag{1.44}$$

If the output network is passive and linear, then

$$-\frac{\pi}{2} \le \phi_n \le \frac{\pi}{2}. \tag{1.45}$$

In this case, the output power is nonzero

$$P_O > 0 \tag{1.46}$$

if

$$\frac{1}{T}\int_0^T i'_D v'_{DS}\, \mathrm{d}t < 0. \tag{1.47}$$

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

$$\frac{1}{T}\int_0^T i'_D v'_{DS}\, \mathrm{d}t = 0 \tag{1.48}$$

then

$$P_O = 0 \tag{1.49}$$

for the following cases:

- The waveforms i_D and v_{DS} are nonoverlapping, as shown in Figure 1.5.
- 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 Figure 1.7(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 Figure 1.7(b).
- The waveforms i_D and v_{DS} are adjacent and their both derivatives $i_D(t_j)$ and $V_{DS}(t_j)$ have jumps at the joint time instant t_j, as shown in Figure 1.7(c).

1.6 Output Power of Class E ZVS Amplifier

The Class E zero-voltage switching (ZVS) RF power amplifier is shown in Figure 1.8. Waveforms for the Class E amplifier under zero-voltage switching and zero-derivative switching (ZDS) conditions are shown in Figure 1.9. Ideally, the efficiency of this amplifier is 100 %. Waveforms for the Class E amplifier are shown in Figure 1.9. 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

$$i'_D(t_o) = \Delta I\, \delta(t - t_o) \tag{1.50}$$

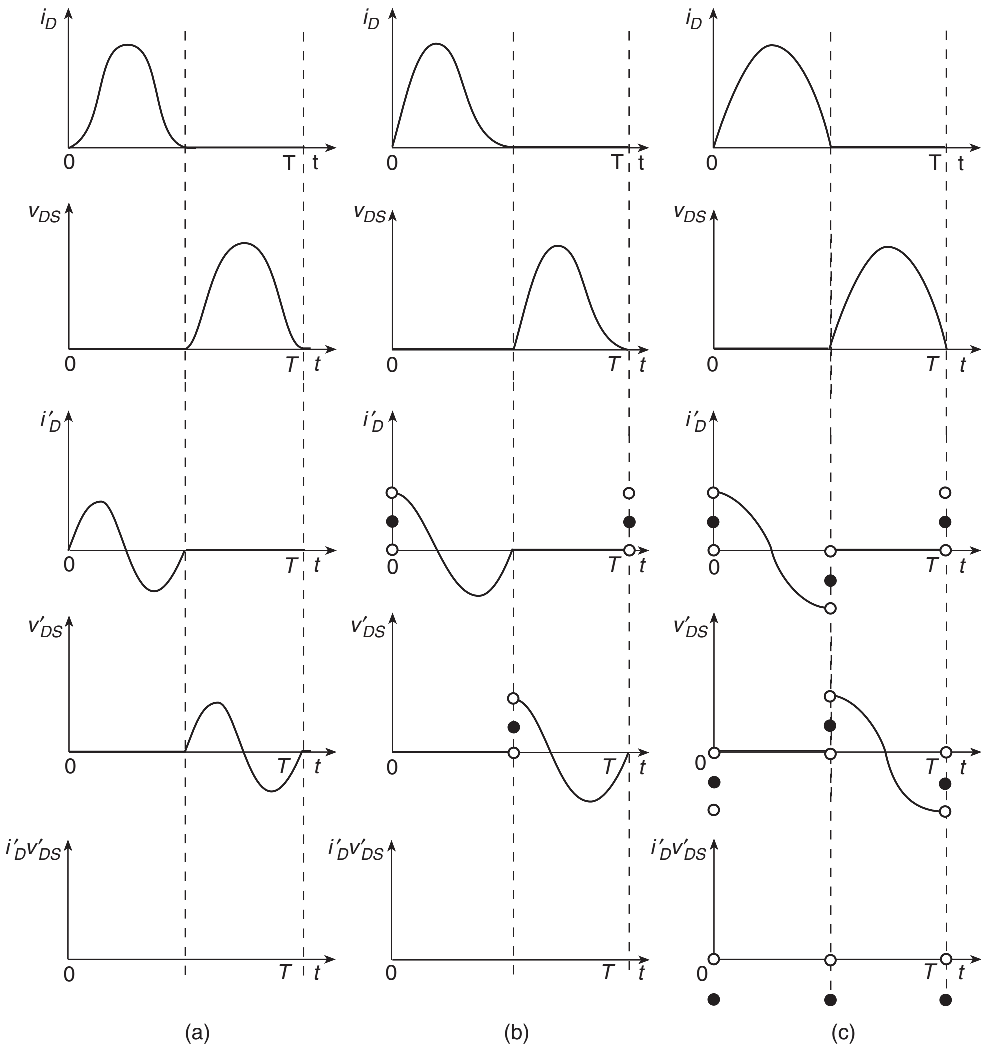

Figure 1.7 Waveforms of power amplifiers with $P_O = 0$.

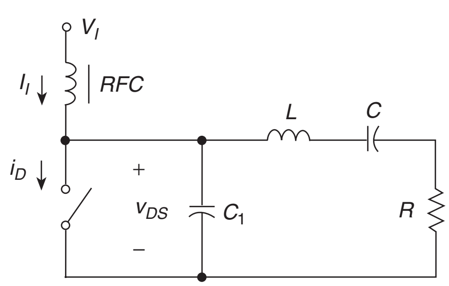

Figure 1.8 Class E ZVS power amplifier.

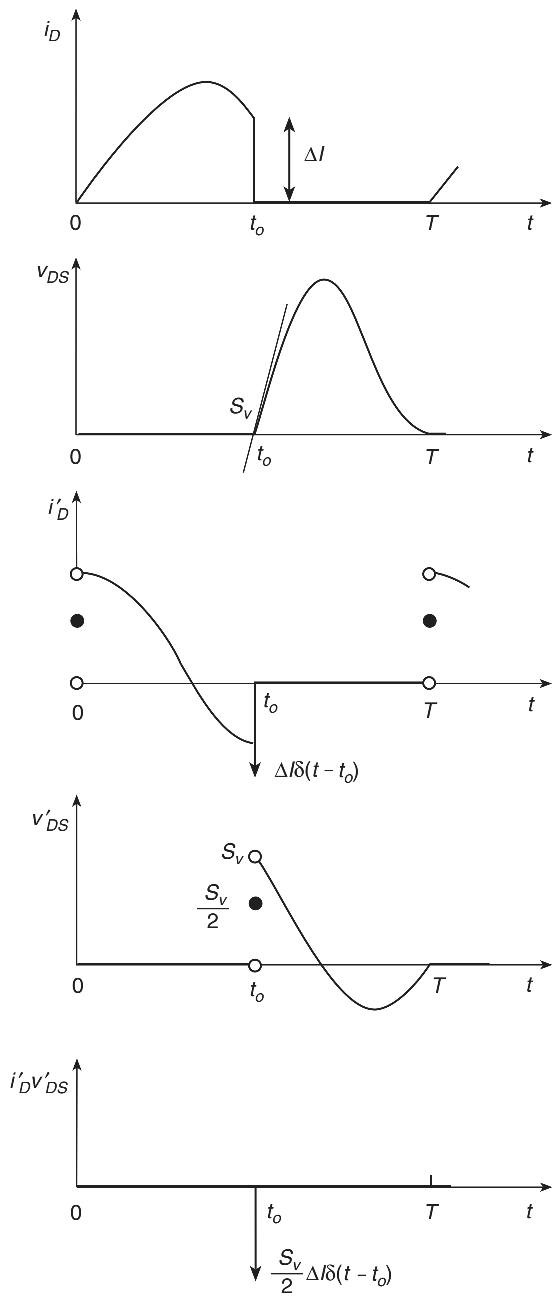

Figure 1.9 Waveforms of Class E ZVS power amplifier.

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

$$v'_{DS}(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} v'_{DS}(t) + \lim_{t\to t_o^+} v'_{DS}(t)\right] = \frac{S_v}{2}. \tag{1.51}$$

The output power of the Class E ZVS amplifier is

$$P_{ds} = P_O = -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\,\mathrm{d}t = -\frac{1}{4\pi^2 f}\int_0^T v'_{DS}\,\Delta I\,\delta(t-t_o)\,\mathrm{d}t$$
$$= -\frac{\Delta I S_v}{8\pi^2 f}\int_0^T \delta(t-t_o)\,\mathrm{d}t = -\frac{\Delta I S_v}{8\pi^2 f} \tag{1.52}$$

where $\Delta I < 0$. Since

$$\Delta I = -0.6988 I_{DM} \tag{1.53}$$

and

$$S_v = 11.08 f V_{DSM} \tag{1.54}$$

the output power is

$$P_O = -\frac{\Delta I S_v}{8\pi^2 f} = 0.0981 I_{DM} V_{DSM}. \tag{1.55}$$

Hence, the output power capability is

$$c_p = \frac{P_O}{I_{DM} V_{DSM}} = 0.0981. \tag{1.56}$$

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 amplifier.

Solution. The output power of the power amplifier is

$$P_O = -\frac{\Delta I S_v}{8\pi^2 f} = -\frac{-1 \times 11.08 \times 10^8}{8\pi^2 \times 10^6} = 14.03\,\mathrm{W}. \tag{1.57}$$

1.7 Class E ZCS Amplifier

The Class E zero-current switching (ZCS) RF power amplifier is depicted in Figure 1.10. Current and voltage waveforms under zero-current switching and zero-derivative switching (ZDS) conditions are shown in Figure 1.11. The efficiency of this amplifier with perfect

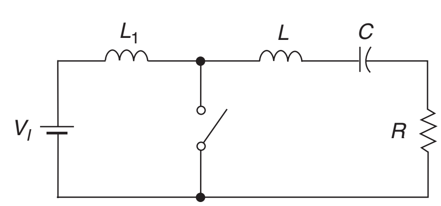

Figure 1.10 Class E ZCS power amplifier.

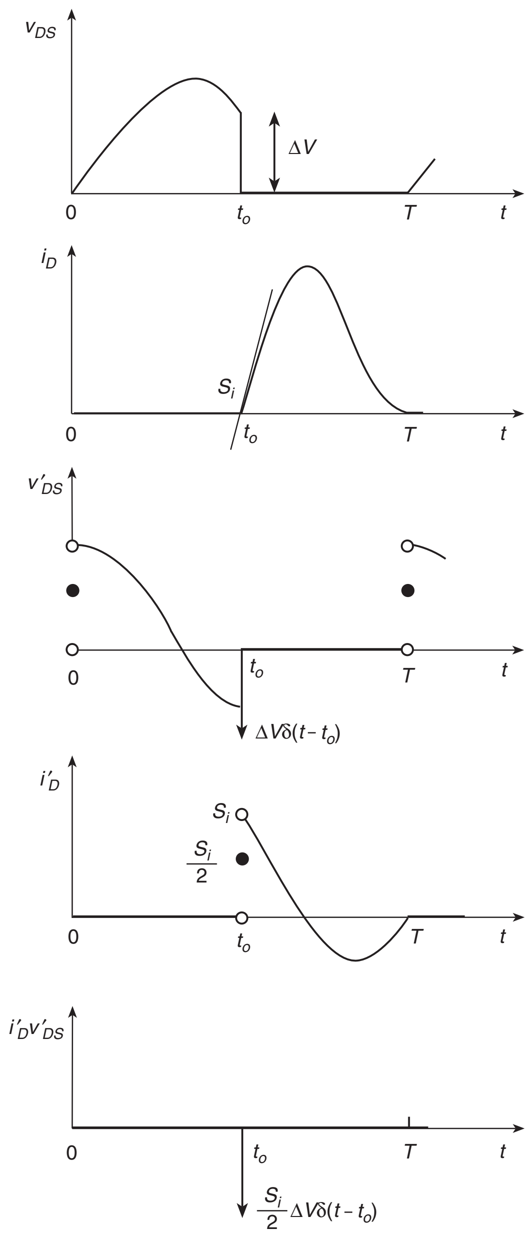

Figure 1.11 Waveforms of Class E ZCS power amplifier.

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

$$v'_{DS}(t_o) = \Delta V\,\delta(t - t_o) \tag{1.58}$$

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

$$i'_D(t_o) = \frac{1}{2}\left[\lim_{t\to t_o^-} i'_D(t) + \lim_{t\to t_o^+} i'_D(t)\right] = \frac{S_i}{2}. \tag{1.59}$$

The output power of the Class E ZCS amplifier is

$$P_{ds} = P_O = -\frac{1}{4\pi^2 f}\int_0^T i'_D v'_{DS}\,\mathrm{d}t = -\frac{1}{4\pi^2 f}\int_0^T i'_D \Delta V\,\delta(t-t_o)\,\mathrm{d}t$$
$$= -\frac{\Delta V S_i}{8\pi^2 f}\int_0^T \delta(t-t_o)\,\mathrm{d}t = -\frac{\Delta V S_i}{8\pi^2 f}. \tag{1.60}$$

Since

$$\Delta V = -0.6988 V_{DSM} \tag{1.61}$$

and

$$S_i = 11.08 f I_{DM} \tag{1.62}$$

the output power is

$$P_O = -\frac{\Delta V S_i}{8\pi^2 f} = 0.0981 I_{DM} V_{DSM}. \tag{1.63}$$

Hence, the output power capability is

$$c_p = \frac{P_O}{I_{DM} V_{DSM}} = 0.0981. \tag{1.64}$$

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 $\Delta V_{DS} = -100\,\mathrm{V}$, a slope of the drain current at the MOSFET turn-on $S_i = 11.08 \times 10^3$ V/s, and the operating frequency is $f = 1$ MHz. Find the output power of the amplifier.

Solution. The output power of the power amplifier is

$$P_O = -\frac{\Delta V_{DS} S_i}{8\pi^2 f} = -\frac{-100 \times 11.08 \times 10^7}{8\pi^2 \times 10^6} = 140.320\,\mathrm{W}. \tag{1.65}$$

1.8 Propagation of Electromagnetic Waves

An antenna is a device for radiating or receiving electromagnetic radio waves. A transmitting antenna converts an electrical signal into an electromagnetic wave. It is a transition structure between a guiding device (such as a transmission line) and free space. A receiving antenna converts an electromagnetic wave into an electrical signal. The electromagnetic wave travels at the speed of light in free space. The wavelength of an electromagnetic wave in free space is given by

$$\lambda = \frac{c}{f} \tag{1.66}$$

where $c = 3 \times 10^8$ m/s is the velocity of light in free space.

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 a transmitter with the output power P_T is given by

$$p(r) = \frac{P_T}{4\pi r^2}. \tag{1.67}$$

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

$$p_r(r) = G_T \frac{P_T}{4\pi r^2}. \tag{1.68}$$

The antenna efficiency is the ratio of the radiated power to the total power fed to the antenna.

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

$$A_e = G_R \frac{\lambda^2}{4\pi} \tag{1.69}$$

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 as

$$P_{REC} = A_e p(r) = G_T G_R P_T \frac{\lambda^2}{(4\pi r)^2} = G_T G_R P_T \left(\frac{c}{4\pi r f_c}\right)^2. \tag{1.70}$$

The received power is proportional to the gain of either antenna and inversely proportional to r^2. For example, the gain of the dish (parabolic) antenna is given by

$$G_T = G_R = 6\left(\frac{D}{\lambda}\right)^2 = 6\left(\frac{Df_c}{c}\right)^2 \tag{1.71}$$

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

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

$$S_L = \frac{P_T}{P_{REC}} = \left(\frac{4\pi r}{\lambda}\right)^2 = 10 \log\left(\frac{P_T}{P_{REC}}\right) = 20 \log\left(\frac{4\pi r}{\lambda}\right). \tag{1.72}$$

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

$$P_{REC} = \frac{L_{syst} G_T G_R P_T}{(4\pi)^2}\left(\frac{\lambda}{r}\right)^2. \tag{1.73}$$

The maximum distance between the transmitting and receiving antennas is

$$r_{max} = \frac{\lambda}{4\pi}\sqrt{L_{syst} G_T G_R \left(\frac{P_T}{P_{REC(min)}}\right)}. \tag{1.74}$$

Antennas are used to radiate the electromagnetic waves and transmit through the atmosphere. The radiation efficiency of antennas is high 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 it should be higher than $\lambda/10$. 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.

There are three groups of electromagnetic waves 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).

Wave propagation is illustrated in Figure 1.12. The 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 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 and therefore 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 and electrons exist in sufficient quantity to affect the wave propagation. The 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 is 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 the skip zone.

The 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 HF band cannot be propagated

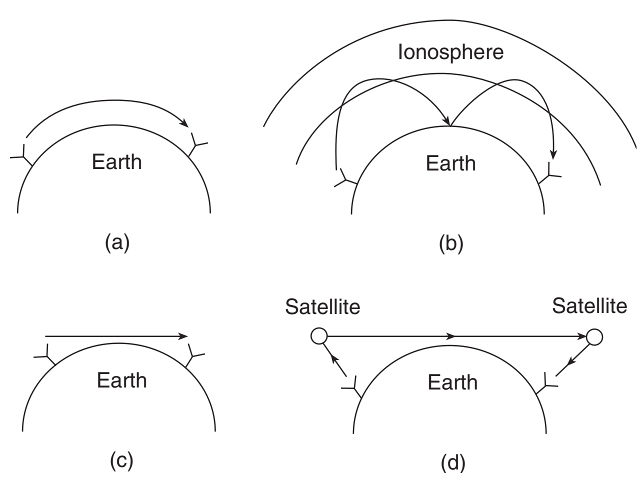

Figure 1.12 Electromagnetic wave propagation. (a) Ground wave propagation. (b) Sky wave propagation. (c) Horizontal wave propagation. (d) Wave propagation in satellite communications.

for long distances along the surface of the Earth. However, it is easy to propagate these signals through free space.

1.9 Frequency Spectrum

Table 1.1 gives the frequency spectrum. In the United States, the allocation of carrier frequencies, bandwidths, and power levels of transmitted electromagnetic waves 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.

Low-frequency (LF) electromagnetic waves are propagated by the 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 amplitude-modulated (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 $f_c = 560$ 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-Ω 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. Aeronautical frequency range starts in MF range and ends in HF range. It is from 2850 to 4063 kHz and is used for short-distance point-to-point communications and ground-air-ground communications. Aircraft flying scheduled routes are allocated specific channels. The high-frequency (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 services. High frequencies are also used for long-distance point-to-point transoceanic ground-air-ground communications. High frequencies 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 frequency-modulated (FM) radio transmission is from 88 to

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	Ultra High Frequencies (UHF)	100–10 cm
3–30 GHz	Super High Frequencies (SHF)	10–1 cm
30–300 GHz	Extra High Frequencies (EHF)	10–1 mm

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	Units
L	1–2	GHz
S	2–4	GHz
C	4–8	GHz
X	8–12.4	GHz
Ku	12.4–18	GHz
K	18–26.5	GHz
Ku	26.5–40	GHz

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 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. The 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 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 to 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 500–1500 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.

Table 1.4 Cellular phone frequency allocation.

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

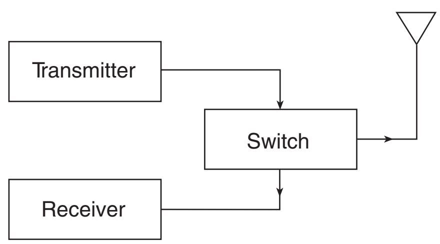

Figure 1.13 Block diagram of a transceiver.

1.10 Duplexing

In two-way communication, a transmitter and a receiver are used. The combination of a transmitter and a receiver is called a *transceiver*. A block diagram of the transceiver is shown in Figure 1.13. 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.

1.11 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 with 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 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.

In the CDMA, each user uses a different code (like a different language). CDMA allows the use of 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.

1.12 Nonlinear Distortion in Transmitters

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. 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

$$v_o = f(v_s) = V_{O(DC)} + a_1 v_s + a_2 v_s^2 + a_3 v_s^3 + a_4 v_s^4 + a_5 v_s^5 + \cdots \quad (1.75)$$

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;
- intermodulation products (IMPs).

Nonlinear distortion components may corrupt the desired signal. Harmonic distortion (HD) occurs when a single-frequency sinusoidal signal is applied to the power amplifier input. Intermodulation distortion (IMD) occurs when two or more frequencies are applied at the power amplifier input. To evaluate the linearity of power amplifiers, we can use (1) a single-tone test and (2) 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 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 (IMPs).

1.13 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

$$v_s(t) = V_m \cos \omega t. \tag{1.76}$$

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

$$v_o(t) = a_1 v_s(t) + a_2 v_s^2(t) + a_3 v_s^3(t). \tag{1.77}$$

The output voltage of the transmitter is given by

$$\begin{aligned} v_o(t) &= a_1 V_m \cos \omega t + a_2 V_m^2 \cos^2 \omega t + a_3 V_m^3 \cos^3 \omega t \\ &= a_1 V_m \cos \omega t + \frac{1}{2} a_2 V_m^2 (1 + \cos 2\omega t) + \frac{1}{4} a_3 V_m (3 \cos \omega t + \cos 3\omega t) \\ &= \frac{1}{2} a_2 V_m^2 + \left(a_1 V_m + \frac{3}{4} a_3 V_m^3 \right) \cos \omega t + \frac{1}{2} a_2 V_m^2 \cos 2\omega t + \frac{1}{4} a_3 V_m^3 \cos 3\omega t. \end{aligned} \tag{1.78}$$

Thus, the output voltage of the power amplifier contains the fundamental components of the carrier frequency $f_1 = f_c$ and harmonics $2f_1 = 2f_c$ and $3f_1 = 3f_c$, as shown in Figure 1.14. The amplitude of the n-th harmonic is proportional to V_m^n. The harmonics may interfere with other communication channels and must be filtered out to an acceptable level.

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

$$f_n = nf_c. \tag{1.79}$$

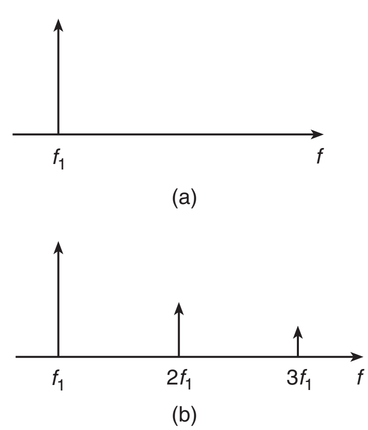

Figure 1.14 Spectrum of the input and output voltages of power amplifiers due to harmonics. (a) Spectrum of the input voltage. (b) Spectrum of the output voltage due to harmonics.

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 output network. For example, the output network of a transmitter may offer 37-dB second-harmonic suppression and 55-dB third-harmonic suppression.

Harmonic distortion is defined as the ratio of the amplitude of the n-th harmonic V_n to the amplitude of the fundamental V_1

$$HD_n = \frac{V_n}{V_1} = 20 \log\left(\frac{V_n}{V_1}\right) \text{ (dB)}. \tag{1.80}$$

The distortion by the second harmonic is given by

$$HD_2 = \frac{V_2}{V_1} = \frac{\frac{1}{2}a_2 V_m}{a_1 V_m + \frac{3}{4}a_3 V_m^3} = \frac{a_2 V_m}{2\left(a_1 + \frac{3}{4}V_m^2\right)}. \tag{1.81}$$

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

$$HD_2 \approx \frac{a_2 V_m}{2a_1}. \tag{1.82}$$

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

The distortion by the third harmonic is given by

$$HD_3 = \frac{V_3}{V_1} = \frac{\frac{1}{4}a_3 V_m^3}{a_1 V_m + \frac{3}{4}a_3 V_m^3} = \frac{a_3 V_m^2}{4a_1 + 3a_3 V_m^2}. \tag{1.83}$$

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

$$HD_3 \approx \frac{a_3 V_m^2}{4a_1}. \tag{1.84}$$

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 n-th harmonic P_n to the power of the carrier P_c is

$$HD_n = \frac{P_n}{P_c} = 10\frac{P_n}{P_c} \text{ (dBc)}. \tag{1.85}$$

The term 'dBc' refers to ratio of the power of a spectral distortion component to the power of the carrer.

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

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

$$= \sqrt{HD_2^2 + HD_3^2 + HD_4^2 + \cdots}. \tag{1.86}$$

Higher harmonics ($n \geq 2$) are distortion terms.

1.14 Intermodulation

Intermodulation 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 *intromodulation products*. The frequencies of the intermodulation products are either the sums or the differences of the 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

$$f_{IM} = nf_1 \pm mf_2 \tag{1.87}$$

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

$$k_{IMP} = n + m. \tag{1.88}$$

If the intermodulation 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 intermodulation products. The third-order intermodulation products usually have components in the system bandwidth. In contrast, the second-order harmonics $2f_1$ and $2f_2$, and the second-order intermodulation 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 excitation test is used to evaluate intermodulation distortion of power amplifiers. In this test, the input voltage of a power amplifier is given by

$$v_s(t) = V_{m1} \cos \omega_1 t + V_{m2} \cos \omega_2 t. \tag{1.89}$$

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

$$\begin{aligned}
v_o(t) &= a_1 v_s(t) + a_2 v_s^2(t) + a_3 v_s^3(t) \\
&= a_1 V_{m1} \cos \omega_1 t + a_1 V_{m2} \cos \omega_2 t + a_2 (V_{m2} \cos \omega_1 t + V_{m2} \cos \omega_2 t)^2 \\
&\quad + a_3 (V_{m1} \cos \omega_1 t + V_{m2} \cos \omega_2 t)^3 \\
&= a_1 V_{m1} \cos \omega_1 t + a_1 V_{m2} \cos \omega_2 t + a_2 V_{m1}^2 \cos^2 \omega_1 t + 2a_2 V_{m1} V_{m2} \cos \omega_1 t \cos \omega_2 t \\
&\quad + a_2 V_{m2}^2 \cos \omega_2 t \\
&\quad + a_3 V_{m1}^3 \cos^3 \omega_1 t + 3a_3 V_{m1}^2 V_{m2} \cos^2 \omega_1 t \cos \omega_2 t + 3a_3 V_{m1} V_{m2}^2 \cos \omega_1 t \cos^2 \omega_2 t \\
&\quad + a_3 V_{m2}^3 \cos^3 \omega_2 t.
\end{aligned} \tag{1.90}$$

Thus,

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

$$+ \frac{3}{4}a_3 V_{m1}^2 V_{m2} \cos(2\omega_1 - \omega_2)t + \frac{3}{4}a_3 V_{m1}^2 V_{m2} \cos(2\omega_1 + \omega_2)t$$
$$+ \frac{3}{4}a_3 V_{m1} V_{m2}^2 \cos(2\omega_2 - \omega_1)t + \frac{3}{4}a_3 V_{m1} V_{m2}^2 \cos(2\omega_2 + \omega_1)t + \cdots. \quad (1.91)$$

The output voltage contains:

- fundamental components f_1 and f_2;
- harmonics of the fundamental components $2f_1, 2f_2\ 3f_1, 3f_2, \ldots$;
- intermodulation products $f_2 - f_1$, $f_1 + f_2$, $2f_1 - f_2$, $2f_2 - f_1$, $2f_1 + f_2$, $2f_2 + f_1$, $3f_1 - 2f_2$, $3f_2 - 2f_1, \ldots$.

The spectrum of the two-tone input voltage of equal amplitudes and several components of the spectrum of the output voltage are depicted in Figure 1.15. If the difference between f_2 and f_1 is small, the intermodulation products appear in the close vicinity of f_1 and f_2. The third-order intromodulation 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. The frequency difference of the IM product at $2f_1 - f_2$ from the fundamental component at f_1 is

$$\Delta f = f_1 - (2f_1 - f_2) = f_2 - f_1. \quad (1.92)$$

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

$$\Delta f = (2f_2 - f_1) - f_2 = f_2 - f_1. \quad (1.93)$$

For example, if $f_1 = 800\,\text{kHz}$ and $f_2 = 900\,\text{kHz}$, then the IMPs of particular interest are $2f_1 - f_2 = 2 \times 800 - 900 = 700\,\text{MHz}$ and $2f_2 - f_1 = 2 \times 900 - 800 = 1000\,\text{MHz}$. To filter out the unwanted components, filters with a very narrow bandwidth are required.

Assuming that $V_{m1} = V_{m2} = V_m$, the amplitudes of the third-order intermodulation product are

$$V_{2f_2 - f_1} = V_{2f_1 - f_2} = \frac{3}{4}a_3 V_m^2. \quad (1.94)$$

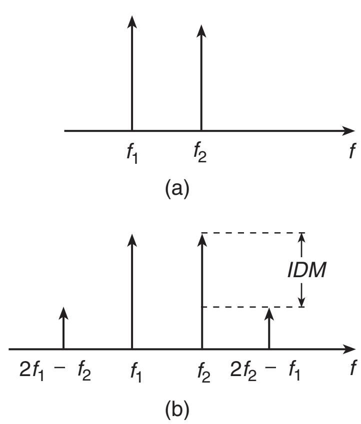

Figure 1.15 Spectrum of the input and output voltages in power amplifiers due to intermodulation. (a) Spectrum of the input voltage. (b) Several components of the spectrum of the output voltage due to intermodulation.

Assuming that $V_{m1} = V_{m2} = V_m$ and $a_3 \ll a_1$, The third-order intermodulation distortion by the IMP component at $2f_1 \pm f_2$ or the IMP component at $2f_2 \pm f_1$ is given by

$$IM_3 = \frac{V_{2f_2-f_1}}{V_{f_2}} = \frac{\frac{3}{4}a_3V_m^3}{\left(a_1 + \frac{9}{4}a_3\right)V_m} \approx \frac{\frac{3}{4}a_3V_m^3}{a_1V_m} = \frac{3}{4}\left(\frac{a_3}{a_1}\right)V_m^2. \tag{1.95}$$

As 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. Therefore, the amplitudes of the IM products increase three times faster 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.

1.15 Dynamic Range of Power Amplifiers

Figure 1.16 shows the desired output power $P_O(f_2)$ and the undesired third-order intermodulation 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 reaches saturation, causing *power gain compression*. The point at which the power gain of the nonlinear amplifier deviates from the fictitious 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

$$P_{O(1\text{dB})}\ (\text{dBm}) = A_{1\text{dB}} + P_{i(1\text{dB})}\ (\text{dBm}) = A_{o(1\text{dB})} - 1\ \text{dB} + P_{i(1\text{dB})}\ (\text{dBm}) \tag{1.96}$$

where A_o is the power gain of an ideal linear amplifier and $A_{1\text{dB}}$ is the power gain at the 1-dB compression point.

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

$$d_R = P_{O(1\text{dB})} - P_{Omin} \tag{1.97}$$

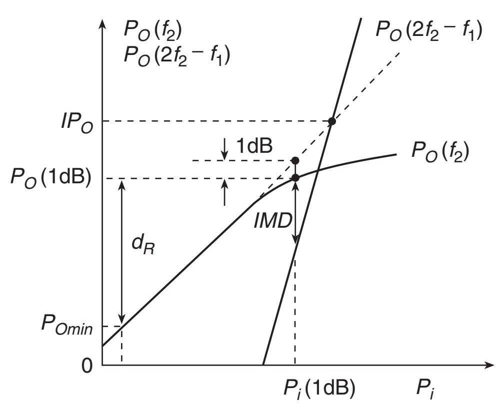

Figure 1.16 Output power $P_O(f_2)$ and $P_O(2f_2 - f_1)$ as functions of input power P_i of power amplifier.

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, e.g., $P_O(f_2) = aP_i$. Assume that the third-order intermodulation product output power $P_O(2f_2 - f_1)$ increases proportionally to the third power, e.g., $P_O(2f_2 - f_1) = (a/8)^3 P_i$. 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 Figure 1.16, where the output power at the IP point is denoted by IP_O.

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

$$IMD \text{ (dB)} = P_O(f_2) \text{ (dBm)} - P_O(2f_2 - f_1) \text{ (dBm)}. \tag{1.98}$$

1.16 Analog Modulation

The function of a communications system is to transfer information from one point to another through a communication link. Block diagrams of a typical communication system depicted in Figure 1.17. It consists of a transmitter whose block diagram is shown in Figure 1.17(a) and a receiver whose block diagram is shown shown in in Figure 1.17(b). In a transmitting system, a radio-frequency 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 bandpass filter is increased (up-conversion) from the intermediate frequency to a carrier frequency

$$f_c = f_{LO} + f_{IF}. \tag{1.99}$$

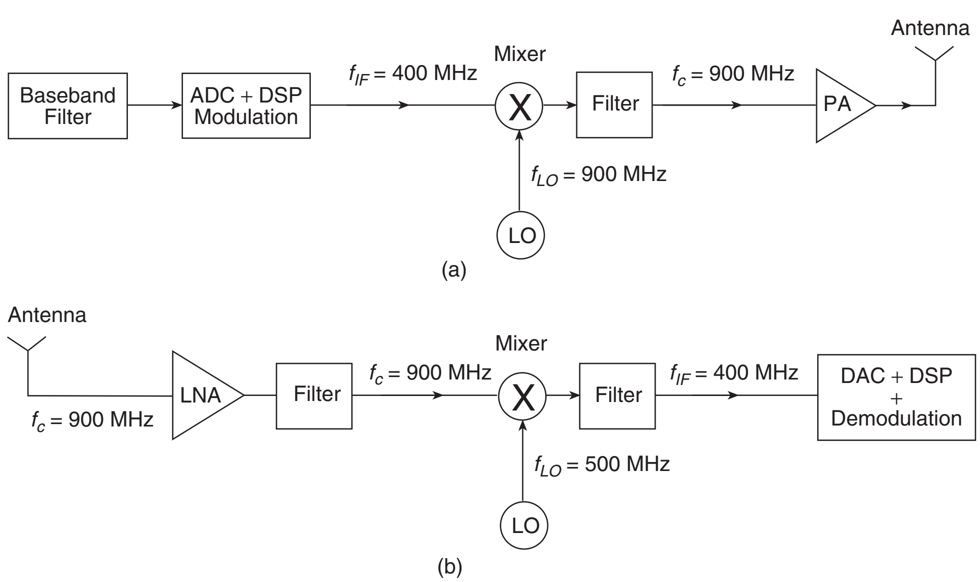

Figure 1.17 (a) Block diagram of a transmitter. (b) Block diagram of a receiver.

The RF current flows through the antenna and produces electromagnetic waves. Antennas produce or collect electromagnetic energy. The transmitted signal is received by the antenna, amplified by a low-noise amplifier (LNA), and applied to a mixer. The frequency of the output signal of the mixer and a bandpass filter is reduced (down-conversion) from the carrier frequency to an intermediate frequency

$$f_{IF} = f_c - f_{LO}. \tag{1.100}$$

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, e.g., 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.

Modulation is the process of placing an information band around a high-frequency 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

$$v_o(t) = A(t)\cos[\omega_c t + \theta(t)] \tag{1.101}$$

where $A(t)$ is the amplitude of the voltage, f_c is the carrier frequency, and $\theta(t)$ is the phase of the carrier. If the amplitude of v_o is varied, it is called amplitude modulation (AM). If the frequency of v_o is varied, it is called frequency modulation (FM). If the phase of v_o is varied, it is called phase modulation (PM).

1.16.1 Amplitude Modulation

In amplitude modulation, the carrier envelope is varied by the modulating signal $v_m(t)$. A single-frequency modulating voltage is given by

$$v_m(t) = V_m \cos \omega_m t. \tag{1.102}$$

The carrier voltage is

$$v_c = V_c \cos \omega_c t. \tag{1.103}$$

The amplitude-modulated (AM) voltage is

$$v_o(t) = V(t)\cos\omega_c t = [V_c + v_m(t)]\cos\omega_c t = (V_c + V_m\cos\omega_m t)\cos\omega_c t$$
$$= V_c\left(1 + \frac{V_m}{V_c}\cos\omega_m t\right)\cos\omega_c t = V_c\,(1 + m\cos\,\omega_m t)\cos\omega_c t \tag{1.104}$$

where the *modulation index* is

$$m = \frac{V_m}{V_c} \leq 1. \tag{1.105}$$

Applying the trigonometric identity,

$$\cos\omega_c t\cos\,\omega_m t = \tfrac{1}{2}[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t] \tag{1.106}$$

we can express v_o as

$$v_o = V_c\cos\omega_c t + \frac{mV_c}{2}\cos(\omega_c - \omega_m)t + \frac{mV_c}{2}\cos(\omega_c + \omega_m)t]. \tag{1.107}$$

The AM voltage modulated by a pure sine wave consists of carrier at f_c, component at $f_c - f_m$, and component at $f_c + f_m$. The AM waveform consists of (1) the carrier component at f_c, (2) the lower side component at $f_c - f_m$, and (3) the upper side component at $f_c + f_m$. If a band of frequencies is used as a modulating signal, we obtain the lower side band and the upper side band. A phasor diagram for amplitude modulation by a single-modulating frequency is shown in Figure 1.18.

The bandwidth of an AM signal is given by

$$BW_{AM} = (f_c + f_m) - (f_c - f_m) = 2f_m. \tag{1.108}$$

The frequency range of voice from 100 to 3000 Hz contains about 95 % of energy. The carrier frequency f_c is much higher than $f_{m(max)}$, e.g., $f_c/f_{m(max)} = 200$.

The power of the carrier is

$$P_C = \frac{V_c^2}{2R}. \tag{1.109}$$

The power of the lower sideband P_{LS} is equal to the power of the upper sideband P_{US}

$$P_{LS} = P_{US} = \frac{m^2V_c^2}{8R} = \frac{m^2}{4}P_C. \tag{1.110}$$

The total transmitted power of the AM signal is given by

$$P_{AM} = P_C + P_{LS} + P_{US} = \left(1 + \frac{m^2}{4} + \frac{m^2}{4}\right)P_C = \left(1 + \frac{m^2}{2}\right)P_C. \tag{1.111}$$

For $m = 1$, the total power is

$$P_{AMmax} = P_C + P_{LS} + P_{US} = \left(1 + \frac{1}{4} + \frac{1}{4}\right)P_C = \frac{3}{2}P_C. \tag{1.112}$$

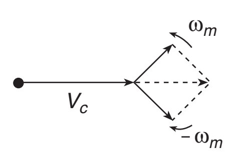

Figure 1.18 Phasor diagram for amplitude modulation by a single modulating frequency.

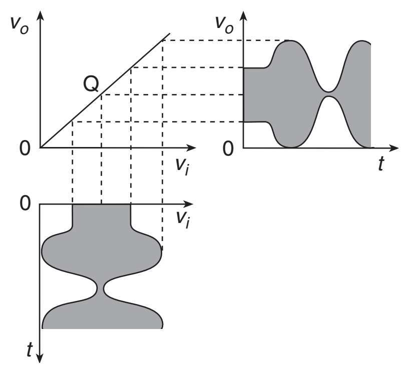

Figure 1.19 Amplification of AM signal by linear power amplifier.

A high-power AM voltage can be generated by the following methods:

- AM signal amplification using linear power amplifiers;
- drain amplitude modulation;
- gate-to-source bias operating point modulation.

Variable-envelope signals, such as AM signals, usually require linear power amplifiers. Figure 1.19 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. The drain AM signal can be generated by connecting a modulating voltage source in series with the drain dc voltage supply V_I. When the MOSFET is operated as a current source like in Class C amplifiers, it should be driven into the ohmic region. If the transistor is operated as a 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 carrier sinusoidal waves. The two waves are out of phase with each other by 90° and therefore are called quadrature carriers. The QAM signal is expressed as

$$v_{QAM} = I(t)\cos\omega_c t + Q(t)\sin\omega_c t \tag{1.113}$$

where $I(t)$ and $Q(t)$ are modulating signals. The received signal can be demodulated by multiplying the modulated signal v_{QAM} by a cosine wave of carrier frequency f_c

$$\begin{aligned} v_{DEM} &= v_{QAM}\cos\omega_c t = I(t)\cos\omega_c t\cos\omega_c t + Q(t)\sin\omega_c t\cos\omega_c t \\ &= \frac{1}{2}I(t) + \frac{1}{2}[I(t)\cos(2\omega_c t) + Q(t)\sin(2\omega_c t)]. \end{aligned} \tag{1.114}$$

A low-pass filter removes the $2f_c$ components, leaving only the $I(t)$ term, which is unef-fected by $Q(t)$. On the other hand, if the modulated signal v_{QAM} is multiplied by a sine wave, and transmitted through a low-pass filter, one obtains $Q(t)$.

1.16.2 Phase Modulation

The output voltage with angle modulation is described by

$$v_o = V_c \cos[\omega_c t + \theta(t)] = V_c \cos \phi(t) \tag{1.115}$$

where the instantaneous phase is

$$\phi(t) = \omega_c t + \theta(t) \tag{1.116}$$

and the instantaneous angular frequency is

$$\omega(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \omega_c + \frac{\mathrm{d}\theta(t)}{\mathrm{d}t}. \tag{1.117}$$

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

- 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 amplitude-modulated signal.

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

$$v_m(t) = V_m \sin \omega_m t. \tag{1.118}$$

The phase of PM signal is given by

$$\theta(t) = k_p v_m(t) = k_p V_m \sin \omega_m t = m_p \sin \omega_m t. \tag{1.119}$$

Hence, the phase-modulated output voltage is

$$\begin{aligned} v_o &= V_c \cos[\omega_c t + k_p v_m(t)] = V_c \cos(\omega_c t + k_p V_m \sin \omega_m t) = V_c \cos(\omega_c t + m_p \sin \omega_m t) \\ &= V_c \cos(\omega_c t + \Delta\phi \sin \omega_m t) = V_c \cos \phi(t) \end{aligned} \tag{1.120}$$

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

$$m_p = k_p V_m = \Delta\phi. \tag{1.121}$$

Hence, the maximum phase modulation is

$$\Delta\phi_{max} = k_p V_{m(max)}. \tag{1.122}$$

The instantaneous frequency is

$$\omega(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \omega_c + \frac{\mathrm{d}\theta(t)}{\mathrm{d}t} = \omega_c + k_p V_m \omega_m \cos \omega_m t. \tag{1.123}$$

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

$$\Delta f = f_{max} - f_c = f_c + k_p V_m f_m - f_c = k_p V_m f_m = m_p f_m. \tag{1.124}$$

The frequency deviation Δf is directly proportional to the modulating frequency f_m. Figure 1.20 shows plots of the frequency modulation Δf as functions of the modulating frequency f_m and the amplitude of the modulating voltage V_m.

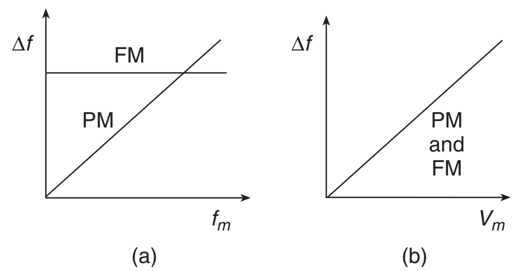

Figure 1.20 (a) Frequency deviation Δf as a function of the modulating frequency f_m. (b) Frequency deviation Δf as a function of the modulating voltage V_m.

1.16.3 Frequency Modulation

The modulating voltage is

$$v_m(t) = V_m \cos \omega_m t. \tag{1.125}$$

The derivative of the phase is

$$\frac{d\theta(t)}{dt} = 2\pi k_f v_m(t). \tag{1.126}$$

Thus, the phase is

$$\theta(t) = 2\pi \int k_f v_m(t)\, dt = k_f V_m \int \cos \omega_m t\, dt = \frac{k_f V_m}{f_m} \sin \omega_m t. \tag{1.127}$$

The frequency-modulated output voltage is given by

$$\begin{aligned} v_o &= V_c \cos[\omega_c t + \theta(t)] = V_c \cos\left(\omega_c t + \frac{k_f V_m}{f_m} \sin \omega_m t\right) \\ &= V_c \cos\left(\omega_c t + \frac{\Delta_f}{f_m} \sin \omega_m t\right) = V_c \cos(\omega_c t + m_f \sin \omega_m t). \end{aligned} \tag{1.128}$$

The instantaneous frequency of an FM signal is given by

$$f(t) = f_c + k_f V_m \sin \omega_m t = f_c + \Delta f \sin \omega_m t \tag{1.129}$$

where the frequency deviation is

$$\Delta f = k_f V_m. \tag{1.130}$$

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

$$\Delta f = f_{max} - f_c = f_c + k_f V_m - f_c = k_f V_m. \tag{1.131}$$

The maximum frequency deviation is

$$\Delta f_{max} = k_f V_{M(max)}. \tag{1.132}$$

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

$$m_f = \frac{\Delta f}{f_m} = \frac{k_f V_m}{f_m}. \tag{1.133}$$

The angle-modulated signal contains components at $f_c \pm nf_m$, where $n = 0, 1, 2, 3, \ldots$. Therefore, the bandwidth of an angle-modulated signal is infinite. However, the amplitudes of components with large values of n is very small. Therefore, the effective bandwidth of the FM signal that contains 98 % of the signal power is given by Carson's rule

$$BW_{FM} = 2(m_f + 1)f_m = 2\left(\frac{\Delta f}{f_m} + 1\right)f_m = 2(\Delta f + f_m). \quad (1.134)$$

If the bandwidth of the modulating signal is BW_m, the frequency-modulation index is

$$m_f = \frac{\Delta f_{max}}{BW_m} \quad (1.135)$$

and the bandwidth of the frequency-modulated signal is

$$BW_{FM} = 2(m_f + 1)BW_m. \quad (1.136)$$

Commercial FM radio stations with an allowed maximum frequency deviation Δf of 75 kHz and a maximum modulation frequency f_m of 15 kHz require a bandwidth

$$BW_{FM} = 2 \times (75 + 15) = 180\,\text{kHz}. \quad (1.137)$$

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

$$\frac{\Delta f}{f_c} = \frac{75}{100 \times 10^6} = 0.075\,\%. \quad (1.138)$$

Standard broadcast FM uses a 200 kHz bandwidth for each station.

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 then used to modulate the phase of the carrier, the FM signal is obtained. This method is used in the Armstrong indirect FM system. The amount of deviation is proportional to the modulating frequency amplitude V_m. If the modulating signal is integrated and then used for phase modulation of the carrier, an FM signal is obtained. This method is used in Armstrong's indirect FM systems.

The relationship between PM and FM is illustrated in Figure 1.21. If the modulating signal amplitude applied to the phase modulator is inversely proportional to the modulating frequency f_m, the phase modulator produces an FM signal. The modulating voltage is given by

$$v_m(t) = \frac{V_m}{\omega_m}\sin\omega_m t. \quad (1.139)$$

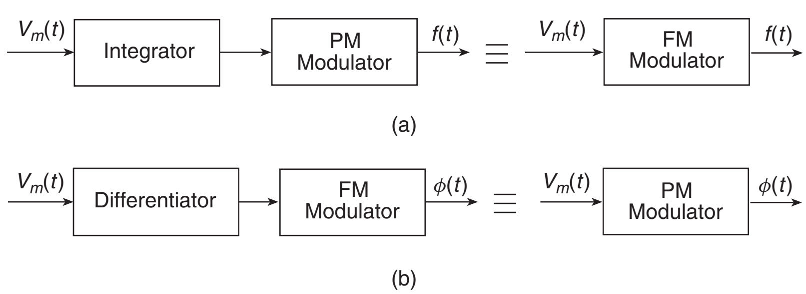

Figure 1.21 Relationship between FM and PM.

The modulated voltage is

$$v_o = V_v \cos\left(\omega_c t + \frac{k_p V_m}{\omega_m} \sin \omega_m t\right) = V_c \cos \phi(t). \tag{1.140}$$

Hence, the instantaneous frequency is

$$\omega(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \omega_c + k_p V_m \cos \omega_m t \tag{1.141}$$

where

$$\Delta f = \frac{k_p V_m}{2\pi}. \tag{1.142}$$

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

The output power of the transmitter with FM and PM modulation 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

$$P_O = \frac{V_c^2}{2R}. \tag{1.143}$$

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

Both FM and PM signals modulated with a pure sine wave and having modulation index $m = m_f = m_p$ can be represented as

$$\begin{aligned} v_o(t) &= V_c \cos(\omega_c t + m \sin \omega_m t) \\ &= V_c \{J_0(m) \cos \omega_c t + J_1(m)[\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t] \\ &\quad + J_2(m)[\cos(\omega_c + 2\omega_m)t - \cos(\omega_c - 2\omega_m)t] \\ &\quad + J_3(m)[\cos(\omega_c + 3\omega_m)t - \cos(\omega_c - 3\omega_m)t] + \cdots\} \\ &= \sum_{n=0}^{\infty} A_c J_n(m) \cos[2\pi(f_c \pm n f_m)t] \end{aligned} \tag{1.144}$$

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

$$\begin{aligned} J_n(m) &= \left(\frac{m}{2}\right) n \left[\frac{1}{n!} - \frac{(m/2)^2}{1!(n+1)!} + \frac{(m/2)^4}{2!(n+2)!} - \frac{(m/2)^6}{3!(n+3)!} + \cdots\right] \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{m}{2}\right)^{n+2k}}{k!(n+k)!}. \end{aligned} \tag{1.145}$$

For $m \ll 1$,

$$J_n(m) \approx \frac{1}{n!}\left(\frac{m}{2}\right)^n. \tag{1.146}$$

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, i.e., at $f_c \pm n f_m$. The amplitude of the sidebands decreases with n, which allows FM transmission within a finite bandwidth. Figure 1.22 shows the amplitudes of spectrum components for angle modulation as functions of modulation index m.

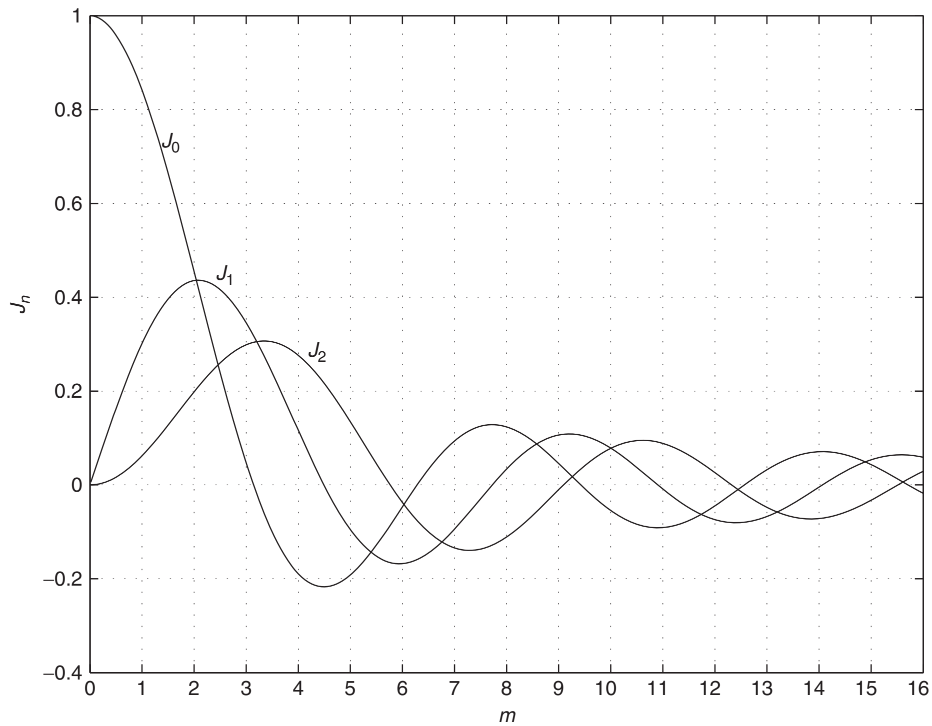

Figure 1.22 Amplitudes of spectrum components J_n for angle modulation as a function of modulation index m.

1.17 Digital Modulation

Digital modulation is 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. Usually, BER $>10^{-3}$. A waveform of a digital binary baseband signal is given by

$$v_D = \sum_{n=1}^{n=m} b_n v(t - nT_c) \tag{1.147}$$

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

1.17.1 Amplitude-shift Keying

The digital amplitude modulation is called the *amplitude-shift keying* (ASK) or ON-OFF keying (OOK). It is the oldest type of modulation used in *radio-telegraph transmitters*. The

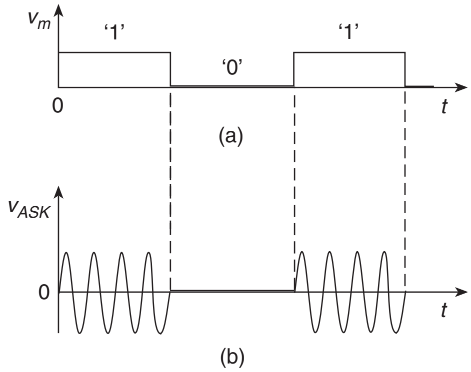

Figure 1.23 Waveform of amplitude-shift keying (ASK).

binary amplitude-modulated signal is given by

$$v_{BFSK} = \begin{cases} V_c \cos \omega_c t & \text{for} \quad v_m = 1 \\ 0 & \text{for} \quad v_m = -1. \end{cases} \tag{1.148}$$

Waveforms illustrating the amplitude-shift keying (ASK) are shown in Figure 1.23. The ASK is sensitive to amplitude noise and therefore is rarely used in RF applications.

1.17.2 Phase-shift Keying

The digital PM modulation is called *phase-shift keying* (PSK). One of the most fundamental types of PSK is the binary phase-shift keying (BPSK), where the phase shift is $\Delta\phi = 180^\circ$. The BPSK has two phase states. Waveforms for BPSK are shown in Figure 1.24. When the waveform is in phase with the reference waveform, it represents a logic '1'. When the

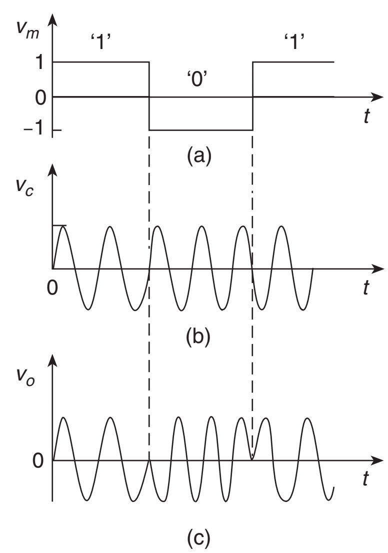

Figure 1.24 Waveforms for binary phase-shift keying (BPSK).

waveform is out of phase with respect to the reference waveform, it represents a logic '0'. The modulating binary signal is

$$v_m = \begin{cases} 1 \\ -1. \end{cases} \quad (1.149)$$

The binary phase-modulated signal is given by

$$v_{BPSK} = v_m \times v_c = (\pm 1) \times v_c = \begin{cases} v_c = V_c \cos \omega_c t & \text{for} \quad v_m = 1 \\ -v_c = -V_c \cos \omega_c t & \text{for} \quad v_m = -1. \end{cases} \quad (1.150)$$

Another type of PSK is the quanternary 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 modulation is given by

$$v_o = V_c \sin\left[\omega_c t + \frac{2\pi(i-1)}{M}\right] \quad (1.151)$$

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

1.17.3 Frequency-shift Keying

The digital FM modulation is called *frequency-shift keying* (FSK). Waveforms for FSK are shown in Figure 1.25. FSK uses two carrier frequencies, f_1 and f_2. The lower frequency f_2

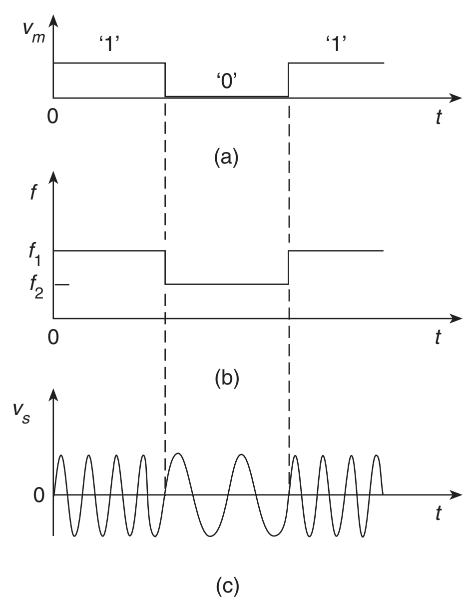

Figure 1.25 Waveforms for frequency-shift keying (FSK).

may represent a logic '0' and the higher frequency f_1 may represent a logic '1'. The binary frequency-modulated signal is given by

$$v_{BFSK} = \begin{cases} V_c \cos \omega_{c1} t & \text{for} \quad v_m = 1 \\ V_c \cos \omega_{c2} t & \text{for} \quad v_m = -1. \end{cases} \tag{1.152}$$

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, high-efficiency switching-mode RF power amplifiers can be used as transmitters.

1.18 Radars

The term 'radar' is derived from *ra*dio *d*etection *a*nd *r*anging. Radars are used in military applications and in a variety of commercial applications, such as navigation, air traffic control, meteorology as weather radars, automobile collision avoidance, law enforcement (in speed guns), astronomy, radio-frequency identification (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 radar antenna transmits a pulse of electromagnetic energy toward a target. The antenna has two basic purposes: it radiates RF energy and provides beam focusing and energy focusing. A wide beam pattern is required for search and a narrow beam pattern is neded for tracking. A typical waveform of the radar signal is shown in Figure 1.26. The duty ratio of the waveform is very low, e.g., 0.1 %. Therefore, the ratio of the peak power to the average power is very large. For example, the peak RF power is $P_{pk} = 200\,\text{kW}$ and the average power is $P_{AV} = 200\,\text{W}$. 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 antennas main beam determines the location of the target. The distance to the target is determined by the time between transmitting and receiving the electromagnetic pulse. The speed of the target with respect to the radar antenna is determined by the frequency shift in the electromagnetic signal, i.e., the Doppler effect. A highly directive antenna is necessary. The radar equation is

$$\frac{P_{rec}}{P_{rad}} = \frac{\sigma_s \lambda^2}{(4\pi)^3 R^4} D^2 \tag{1.153}$$

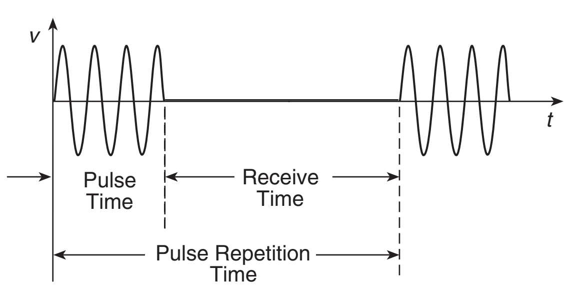

Figure 1.26 Waveform of radar signal.

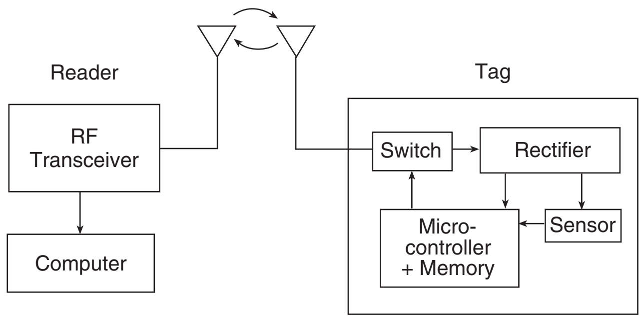

Figure 1.27 Block diagram of an RFID sensor system with a passive tag.

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, and D is the directive gain of the antenna.

1.19 Radio-frequency Identification

A radio-frequency identification (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 to 5.875 GHz. The tags are passive or active. An 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 on-board battery. The reading range is of active tags is much larger that of passive tags and are more expensive. A block diagram of an RFID sensor system with a passive tag is shown in Figure 1.27. The reader transmits a signal to 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 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.

1.20 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 with current-source mode of operation.

- 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.
- Switching power loss can be reduced by using the zero-voltage switching (ZVS) technique or zero-current switching (ZCS) technique.
- In Class A, B, and C 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.
- 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.
- 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 a low-frequency modulating signal on a high-frequency carrier so that the instantaneous changes in the amplitude of the modulating signal produce corresponding changes in the amplitude of the high-frequency carrier.
- FM is the process of superimposing the modulating signal on a high-frequency carrier so that the carrier frequency departs from the 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 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.
- A transceiver is a combination of a transmitter and a receiver.
- Duplexing techniques allow for users to transmit and receive signals simultaneously.
- The major multiple-access techniques are: time-division multiple access (TDMA) technique, frequency-division multiple access (FDMA) technique, and code-division multiple access (CDMA) technique.

1.21 References

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

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

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

[4] M. K. Kazimierczuk and D. Czarkowski, *Resonant Power Converters*, John Wiley & Sons, New York, NY, 1995.

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

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

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

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

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

[10] M. K. Kazimerczuk, 'Generalization of conditions for 100-percent efficiency and nonzero output power in power amplifiers and frequency multipliers,' *IEEE Trans. Circuits and Systems*, vol. CAS-33, pp. 805–807, August 1986.

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

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

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

[14] A. Grebennikov and N. O. Sokal, *Switchmode Power Amplifiers*, Elsevier, Amsterdam, 2007.

[15] M. K. Kazimierczuk, *Pulse-width Modulated DC-DC Power Converters*, John Wiley & Sons, New York, NY, 2008.

1.22 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 ZVS operation of power amplifiers.

1.6 Explain the principle of ZCS operation of 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 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.

1.23 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 2.4 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 dc voltage of an RF transmitter with a Class C power amplifier is 24 V. The minimum voltage of the transistor is $V_{DSmin} = 2\,\text{V}$. Find the maximum value of the index modulation.

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

1.6 An FM signal of a transmitter is applied to a 50 Ω antenna. The amplitude of the signal is $V_m = 1000\,\text{V}$. What is the output power of the transmitter?

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

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

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

1.10 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?