1

The Wireless World

On Sunday, April 14, 1912, shortly before midnight, the *RMS Titanic* struck an iceberg off the coast of Newfoundland. The radio operator, John Phillips, repeatedly transmitted the distress call CQD in Morse Code. He also sent the newly established signal $\overline{\text{SOS}}$. Fifty-eight miles away, the *Carpathia* received the messages, and steamed toward the sinking liner. The *Carpathia* pulled 705 survivors out of their lifeboats. Phillips continued transmitting until power failed. He and the other passengers could have been saved if more lifeboats had been available, or if the *California*, which was so close that it could be seen from the deck of the *Titanic*, had had a radio operator on duty. However, this dramatic rescue established the power of wireless communication. Always before, ships out of sight of land and each other were cut off from the rest of the world. Now the veil was lifted. Since the *Titanic* disaster, wireless communications have expanded beyond the dreams of radio pioneers. Billions of people around the world receive radio and television broadcasts every day. Millions use cellular telephones and pagers and receive television programs from satellites. Thousands of ships and airplanes communicate by radio over great distances and navigate by the radio-navigation systems LORAN and GPS.

The enormous increase of wireless communications is tied to the growth of electronics in general, and computers in particular. Often people distinguish between digital and analog electronics. By digital electronics, we mean the circuits that deal with binary levels of voltage, the ones and zeros that are the syllables of computers and calculators. In analog electronics, we deal with voltages and currents that vary continuously. Most systems today actually use a mixture of analog and digital electronics. It is analog electronics that is the focus of this book. We will study the design, construction, and testing of radio circuits and systems. The approach that we take is the progressive construction of a transceiver, the NorCal 40A.

The NorCal 40A transceiver was designed by Wayne Burdick. It is named after the Northern California QRP club. The club is a group of radio amateurs that develop new transceiver designs. QRP is a radio signal indicating a low-power station. Wayne originally designed the NorCal 40 transceiver as a club project, but it proved so popular that the club could not keep up with orders. An improved version, the NorCal 40A, is now available from the Wilderness Radio Company. The NorCal 40A operates in the frequency range from 7.00 MHz to 7.04 MHz, which is the 40-meter amateur band. The NorCal 40A is well suited for learning electronics. It includes an exceptional variety of interesting analog circuits ranging from audio frequencies to radio frequencies and operates at power levels that vary from a picowatt, or 10^{-12} watts, at the receiver input, to two watts at the transmitter output. Because the transmitter was designed for Morse Code, it produces superb sine-wave signals that are extremely useful in receiver testing. In contrast to cellular telephone circuits, the frequencies are low enough that the signals can be conveniently observed on an oscilloscope, and the level of integration is low, so that the detailed behavior of the circuits can be probed. Because of its history as a club project, the circuits are bullet proof, the components are inexpensive, and the design is open.

1.1 Kirchhoff's Laws

The key quantities in electronics are voltage and current. We will study formulas for voltage and current at two levels. At the bottom, we have *components*, such as resistors, capacitors, and inductors, which have particular relationships between voltage and current. We will study these formulas in the next chapter. We connect components to form a *circuit*. In a circuit, two fundamental laws of physics – conservation of energy and charge – govern voltages and currents. Conservation of energy gives us *Kirchhoff's voltage law*, which applies to a loop of components. Conservation of charge gives *Kirchhoff's current law*, which applies to a junction between components. These laws are named for the German physicist Gustav Kirchhoff. Kirchhoff's laws, together with the current–voltage relations for individual components, allow us to predict what circuits will do.

We can define *voltage* by relating electric charge and potential energy. When a charge moves from one place to another, its potential energy changes. The potential-energy change comes from the force exerted by an electric field. We write the charge as Q and the potential-energy change as E. The units of charge are coulombs, with the abbreviation C, and the units of energy are joules, abbreviated J. See Table 1.1 for a list of units and prefixes that we will use in this book. We can write the voltage V as the ratio of the energy change and charge:

$$V = E/Q. \tag{1.1}$$

The units of voltage are volts, abbreviated as V. We can think of the voltage as the potential energy per unit charge. Often the voltage is called the potential, and that may help make the connection to potential energy clearer. In this connection, it is precise to speak of a voltage *difference*, because the voltage depends on the potential-energy difference between the beginning and ending points. Also, the sign of the voltage depends on the order of the points, and we will often add plus and minus signs to indicate the plus and minus terminals. However, it is typical to consider all the voltages relative to a single reference point, the *ground*, which in an electronic circuit is often the metal box. We will often talk about the voltage at some point in a circuit, and by this, we mean the voltage relative to the reference

Table 1.1. Units (a) and Prefixes (b), and Their Abbreviations.

Quantity	**Symbol**	**Unit**	**Abbr.**	**Prefix**	**Multiplier**	**Abbr.**
voltage	V	volt	V	atto	10^{-18}	a
charge	Q	coulomb	C	femto	10^{-15}	f
energy	E	joule	J	pico	10^{-12}	p
current	I	ampere	A	nano	10^{-9}	n
time	t	second	s	micro	10^{-6}	μ
resistance	R	ohm	Ω	milli	10^{-3}	m
conductance	G	siemen	S	kilo	10^{3}	k
power	P	watt	W	mega	10^{6}	M
capacitance	C	farad	F	giga	10^{9}	G
inductance	L	henry	H	tera	10^{12}	T
length	l	meter	m	peta	10^{15}	P
frequency	f	hertz	Hz	exa	10^{18}	E
absolute temperature	T	kelvin	K			
temperature, Celsius	T	degree	°C			
pressure	P	pascal	Pa			

(a) **(b)**

Note: We use the MKS (meter-kilogram-second) system, so that the expressions we derive will be combinations of these units. A few dimensions are also given in centimeters (cm). It is important to distinguish between upper and lower case. For example, one could mistake milli (m) for mega (M), or kilo (k) for kelvin (K). We use both the absolute temperature scale and the Celsius scale. The unit has the same size in each scale, but the zero is different. For absolute temperature, the zero is absolute zero. For the Celsius scale, the zero is the ice point of water, which is 273.15 kelvins on the absolute scale.

ground. We shall also speak of the voltage drop *across* a resistor or capacitor, and this is the voltage difference between the connecting wires.

We can apply our voltage definition to a circuit with components connected in *parallel* (Figure 1.1a). A parallel connection forces the voltage across each component to be the same because we calculate the voltage across each component at the same points:

$$V_1 = V_2. \tag{1.2}$$

It is more interesting to consider the voltages for components connected in a loop (Figure 1.1b). Consider a charge at point P. When we move the charge around the loop, the potential energy changes at each component. If we return to P, the potential energy returns to its initial value. We say that the potential energy is *path independent*. This is true as long as the magnetic-field changes inside the loop are not important. We will consider the effect of magnetic fields when we study inductors. This means that the sum of the voltages around the loop is zero:

$$0 = \sum_i V_i, \tag{1.3}$$

where i is an index for the components. This is *Kirchhoff's voltage law*, and it is a statement of the conservation of energy. You have to be careful about the

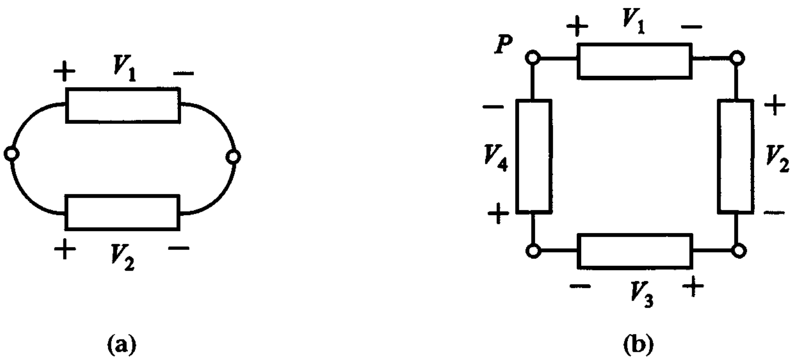


Figure 1.1. Parallel connection (a) and connecting components to form a loop (b). Kirchhoff's voltage law applies to a loop.

signs of voltages when you apply the formula. You might test yourself on your understanding of Kirchhoff's law by considering the parallel circuit in Figure 1.1a as a simple loop.

The *current* is the flow of charge past some point. The units are coulombs per second, or amperes (A). Often this is shortened to "amps." We will talk about the current *through* a wire or a component. Like the voltage, we will need to be careful about signs, and we will add arrows to indicate the positive direction. We can apply our current definition to a series connection (Figure 1.2a). If the current does not leak out along the way, the current in both components must be the same, and we write

$$I_1 = I_2, \tag{1.4}$$

where I is the traditional letter for current. For a junction between several components (Figure 1.2b), if charge does not pile up at the junction, positive current in

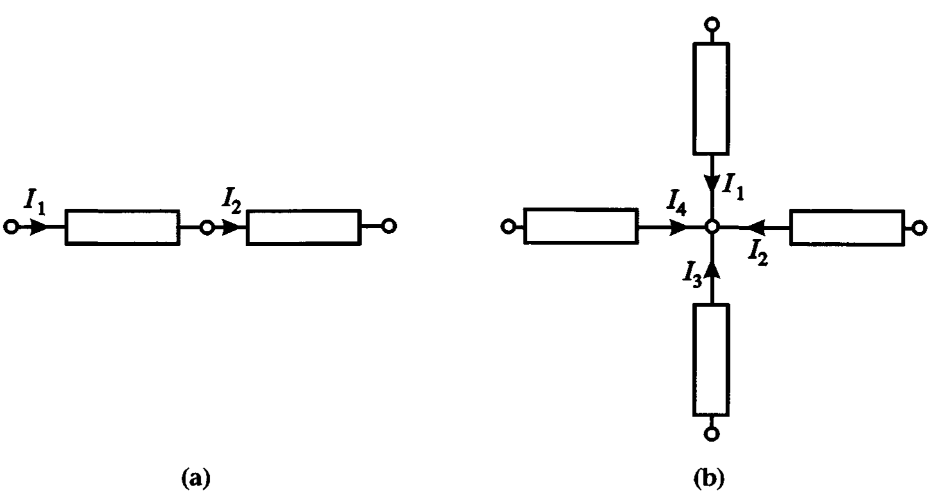


Figure 1.2. Series connection (a) and a junction between several components (b). Kirchhoff's current law applies to a junction between components.

some components must be offset by negative current in others, and we can write

$$0 = \sum_i I_i. \tag{1.5}$$

In words, the net current at a junction must be zero. This is *Kirchhoff's current law*, and it is a statement of conservation of charge. Test yourself on Kirchhoff's current law by considering the series circuit in Figure 1.2a as a simple junction.

Since the voltage is energy per charge, and the current is the rate of charge flow, we can find the power, which is the rate at which work is done, by multiplying voltage and current. We write the power P in watts (W) as

$$P = VI. \tag{1.6}$$

In a radio transmitter, power radiates into the air from an antenna as radio waves. In an audio system, power radiates from a speaker as sound. It may also go into heat, and we will have to watch carefully to see how hot our components get.

1.2 Frequency

In radio engineering, we commonly use voltages and currents that are cosine functions of time. For example, we might write a voltage $V(t)$ and a current $I(t)$ as

$$V(t) = V_p \cos(2\pi f t), \tag{1.7}$$

$$I(t) = I_p \cos(2\pi f t). \tag{1.8}$$

In these expressions f is the *frequency*. The units of frequency are hertz (Hz), after the German physicist Heinrich Hertz who first demonstrated radio waves. The frequency is the number of cycles of the cosine that are completed in one second. Often the zero of the time scale will not be important. In fact, if we shift the time scale by a quarter of a period, the cosine becomes a sine. This means that we could use sine functions if we wanted instead of cosine functions. In these formulas, V_p and I_p are called *peak amplitudes*. In the laboratory, we will often see cosine voltages on an oscilloscope, where the zero of the voltage scale may not be clear. For this reason, people usually measure the peak-to-peak voltage, V_{pp}, shown in Figure 1.3, rather than the peak voltage. The peak-to-peak voltage is twice the

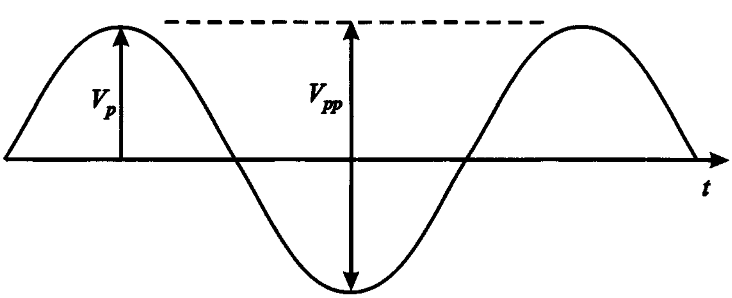


Figure 1.3. Peak and peak-to-peak voltages.

peak voltage. It is important to keep track of whether one is talking about peak voltage, which is easier to use in mathematical formulas, or peak-to-peak voltage, which is easier to measure on an oscilloscope.

We can calculate power by multiplying voltage and current. We write

$$P(t) = V(t)I(t) = V_p I_p \cos^2(2\pi f t). \tag{1.9}$$

The power is not constant as it is for a steady voltage and current. We can separate the power into a steady part and a variable part by writing

$$P(t) = \frac{V_p I_p}{2} + \left(\frac{V_p I_p}{2}\right)\cos(4\pi f t). \tag{1.10}$$

The last term has zero average value, and so we can write the average power P_a as

$$P_a = \frac{V_p I_p}{2}. \tag{1.11}$$

The factor of 2 in the denominator takes into account the fact that the voltage and current vary over a cycle. This formula is for peak amplitudes. In terms of peak-to-peak amplitudes, the average power becomes

$$P_a = \frac{V_{pp} I_{pp}}{8}. \tag{1.12}$$

These formulas depend on the fact that the voltage and current are synchronized so that they are both cosines. This is the situation for a resistor or a properly adjusted antenna. However, for a capacitor or an inductor, one waveform is a cosine and the other is a sine. The power becomes

$$P(t) = V_p I_p \sin(2\pi f t)\cos(2\pi f t) = \left(\frac{V_p I_p}{2}\right)\sin(4\pi f t). \tag{1.13}$$

The average power is zero. This means that an inductor or capacitor does not consume power. However, we will see in the next chapter that they can store energy.

Radio waves travel at the speed of light, and this means that we can write a simple relation between the frequency f and the wavelength λ (the Greek letter *lambda*) as

$$f\lambda = c, \tag{1.14}$$

where c is the velocity of light, given exactly as

$$c = 299{,}792{,}458 \text{ m/s}. \tag{1.15}$$

For our purposes, 3.00×10^8 m/s is close enough. It is useful to be able to convert between wavelength and frequency quickly. To do this, divide the quantity that you know into 300:

$$f\,(\text{MHz}) = 300/\lambda\,(\text{m}) \quad \text{or} \quad \lambda\,(\text{m}) = 300/f\,(\text{MHz}). \tag{1.16}$$

One reason that the wavelength is important is that it determines the antenna size. Transmitting antennas need to be at least an eighth wavelength long to be efficient. If the antenna is much shorter than this, most of the power will heat up the wire rather than radiate as radio waves. For example, consider an AM station transmitting at 1 MHz, where the wavelength is 300 m. A large tower must be used for the transmitting antenna. This wavelength would not be suitable for portable telephones. In contrast, cellular phones use frequencies near 1 GHz, where the wavelength is 30 cm, and the antennas need only be a few centimeters long.

At any time, there are thousands of radio services in use around the world. Radios avoid interfering with each other by transmitting on different frequencies. Television stations do the same thing, but the different frequencies are called channels. In cellular phones the process is more complicated. In some systems, different frequencies are assigned. These frequencies may change as a car moves from the area covered by one antenna to another. In other cellular phone systems, the phones may share a range of frequencies with other users but transmit with different codes that allow them to be distinguished. The frequency bands that can be used for different applications are assigned by the Federal Communications Commission (FCC) in the United States and the corresponding communications authorities in other countries. Radio waves propagate long distances, and for this reason, the different communications authorities cooperate to help keep transmitters in different countries from interfering with each other. One factor in assigning frequencies is that waves at different frequencies travel different distances. For example, frequencies below 30 MHz can propagate around the world by reflecting off the ionosphere. Another factor is that some services like TV need large channel spacings. Each TV channel takes up 6.5 MHz. In contrast, AM stations are 10 kHz apart, and stations communicating by Morse Code may be only 500 Hz apart. Table 1.2 gives the names for the radio bands.

Fundamentally, radios are limited by unwanted signals that are at the same frequency as the signal we want. Sometimes these signals are produced by other

Table 1.2. Naming Radio Bands.

VLF (*very low frequency*, 3–30 kHz, or 100–10 km) – submarine communication (24 kHz)
long wave (30–300 kHz, or 10–1 km) – LORAN navigation system (100 kHz)
medium wave (300 kHz–3 MHz, or 1 km–100 m) – AM radio (500–1,600 kHz)
HF (*high frequency* or *short wave*, 3 MHz–30 MHz, or 100–10 m) – international broadcasting, air and ship communication, amateur communication
VHF (*very high frequency*, 30 MHz–300 MHz, or 10–1 m) – television (channels 2–6: 54–88 MHz), channels 7–13: 174–216 MHz) and FM radio (88–108 MHz)
UHF (*ultra high frequency*, 300 MHz–1 GHz, or 1 m–30 cm) – television (channels 14–69: 470–806 MHz), cellular telephone (824–894 MHz)
microwaves (1–30 GHz, or 30–1 cm) – GPS (1.575 GHz), PCS (personal communications services, 1.85–2.2 GHz), ovens (2.45 GHz), satellite TV (C-band: 3.7–4.2 GHz and Ku band: 10.7–12.75 GHz)
millimeter waves (30–300 GHz, or 10–1 mm) – This is the frontier and many applications are under development: car radar (76 GHz), computer networks inside buildings (60 GHz), aircraft landing in fog (94 GHz).
submillimeter waves (frequencies greater than 300 GHz, wavelengths less than 1 mm) – These frequencies are strongly absorbed by water vapor, and this limits them to scientific applications such as radio astronomy and fusion-plasma diagnostics.

Note: The frequency and wavelength ranges are identified, together with the prominent applications.

transmitters at the same frequency, and we call this *interference*. In addition, there are natural radio waves that are called *noise*. Noise appears at all frequencies, and the noise at the receiver frequency is treated just the same way as the signal we want. We hear the noise from a speaker as *hiss*. In a television receiver, noise makes speckles on the screen that we call *snow*. There are many sources of noise. For example, at the frequency used by the NorCal 40A, 7 MHz, the most important noise source is lightning in tropical thunderstorms. The transistors in a transceiver also produce noise, but typically the power is a thousand times less than the lightning noise. However, if the antenna is small, the lightning noise is not received well, and the transistor noise can dominate. At higher frequencies the noise may be quite different. For example, in the UHF TV channels between 400 and 800 MHz, radio waves from the center of the Milky Way dominate. At 12 GHz, which is used for satellite TV transmissions, the most important noise may be radio waves from the earth's surface.

In the simplest form, a radio communications system consists of a *transmitter*, with its antenna to launch the radio waves, and a *receiver*, with its antenna to receive them. In TV and radio broadcasting, that is all there is, but for two-way communications we need a transmitter and a receiver at each end. Typically these are combined in one box to make a *transceiver*. In the NorCal 40A transceiver, the transmitter and receiver share an oscillator circuit that determines the operating frequency, and they share the antenna.

1.3 Modulation

A cosine waveform by itself does not tell us much. Usually we would like to send a message, which might be an audio signal or a digital signal made up of 1s and 0s from a computer that might represent almost anything. We can include the message in our transmitted signal by varying the amplitude or frequency. This is called *modulation*. In *amplitude modulation*, or AM, we vary the amplitude. This is used in AM radio stations. For example, if we have an audio signal $a(t)$, we get an amplitude-modulated signal by using it as the amplitude:

$$V(t) = a(t)\cos(2\pi f t). \tag{1.17}$$

The signal $a(t)$ is called the *modulating waveform*. Usually the voltage is not transmitted in this form. We will see when we build an AM detector that it is easy to recover the audio voltage if a cosine is added:

$$V(t) = V_c\cos(2\pi f t) + a(t)\cos(2\pi f t). \tag{1.18}$$

Here $V_c\cos(2\pi f t)$ is the *carrier*. AM radio stations use this kind of signal. An AM modulated waveform is shown in Figure 1.4a.

We can also modulate by varying the frequency. This is *frequency modulation*, or FM. We can write an FM signal in the form

$$V(t) = V_c\cos(2\pi(f_c + a(t))t). \tag{1.19}$$

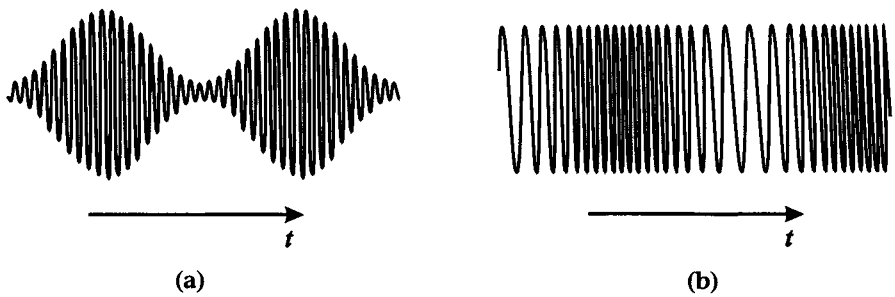


Figure 1.4. AM waveform (a) and FM waveform (b). In each case, the modulating signal is a cosine function. In AM, the amplitude changes, but the frequency stays the same. In FM, the frequency varies, but the amplitude is constant.

An FM waveform is shown in Figure 1.4b. FM signals usually require wider channels than AM signals, but FM systems are usually better at rejecting noise and interference. For this reason FM stations often program music, which is quite sensitive to noise. In contrast, AM stations often emphasize talk shows and news, which are less bothered by noise. Television uses both AM and FM – AM for the picture, and FM for the audio.

For digital signals, a variety of modulation types are used. The oldest is Morse Code, which may properly be considered digital modulation. In Morse Code, the letters are sent as a series of short and long pulses (*dits* and *dahs*), and different lengths of spaces are used to separate letters and words. This could be considered a form of AM. Morse Code is traditionally sent by hand and decoded by ear. A different approach that is better suited for machine reception is to use one frequency for a 1 and another frequency for a 0. We write

$$V(t) = \begin{cases} V_p \cos(2\pi f_1 t) & \text{for a 1,} \\ V_p \cos(2\pi f_0 t) & \text{for a 0.} \end{cases} \tag{1.20}$$

You will also hear these called *mark* and *space* frequencies. This approach is called *frequency-shift keying*, or FSK, and it is a form of frequency modulation. An FSK signal is shown in Figure 1.5a. Another approach is to keep the frequency the same, but to change the sign of the voltage. This is called *phase-shift keying* or

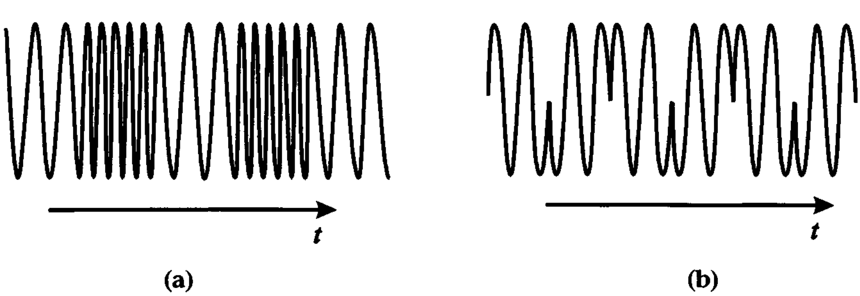


Figure 1.5. An FSK waveform (a) and a PSK waveform (b). The modulating signal in both cases is a series of alternating 1s and 0s.

PSK (Figure 1.5b):

$$V(t) = \begin{cases} +V_p \cos(2\pi f t) & \text{for a 1,} \\ -V_p \cos(2\pi f t) & \text{for a 0.} \end{cases} \tag{1.21}$$

Some of the most sophisticated modulation approaches are found in computer modems, where a combination of amplitude shifts and phase shifts are used.

1.4 Amplifiers

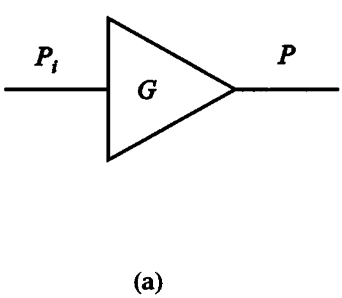
The radio signals that are received by an antenna are often weak, with typical power levels in the range of one picowatt. To make the level convenient for listening, a huge increase in the power is required, in the range of a factor of a billion. The device that does this is called an *amplifier*, and it is key to all electronic systems. Early amplifiers were made with vacuum tubes, but most amplifiers now are made with transistors, which are smaller and lighter. Vacuum tubes can handle much more power than transistors, however, so they are still found in high-power transmitters. We will study several different amplifiers in the NorCal 40A. An amplifier is characterized by the *gain*, which is the ratio of output power to input power. We write the gain G as

$$G = P/P_i, \tag{1.22}$$

where P is the output power and P_i is the input power. Figure 1.6a shows the symbol for an amplifier that is used in circuit diagrams. Amplifiers also appear in transmitters, because the signals are initially generated in low-power oscillators and therefore must be amplified.

1.5 Decibels

In radio systems, power varies by enormous factors. International broadcasters may transmit as much as a megawatt. Receivers work with signals that may be



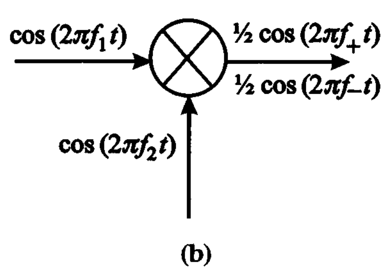


Figure 1.6. (a) Symbol for an amplifier that appears in circuit diagrams. The symbol is a triangle that indicates the direction of power flow. (b) Operation of a mixer. The symbol for a mixer is a circle with a large × inside it to indicate multiplication. The inputs and outputs are distinguished by arrows.

Table 1.3. Comparing the Gain Ratio with the Gain in dB.

G, dB	0	1	2	3	4	5	6	7	8	9	10
G, ratio	1	1.3	1.6	2.0	2.5	3.2	4.0	5.0	6.3	8	10

Note: It is an interesting exercise to construct this table with pencil and paper, using the fact that 3 dB corresponds to a gain ratio of 2, 10 dB corresponds to 10, and doubling and halving the gain in dB are equivalent to squaring and taking the square root of the gain ratio.

less than a femtowatt. In order to deal with such a wide range of numbers, it is convenient to use a logarithmic scale for comparing power levels. It is traditional to use a base-10 logarithm and multiply the result by 10. The units are *decibels*, but people usually say "dB." For an amplifier, we can write the gain in dB as

$$G = 10\log(P/P_i)\,\mathrm{dB}. \tag{1.23}$$

We will write "log" for the base-10 logarithm, to distinguish it from "ln," the natural logarithm. We will use the symbol G for both the gain ratio and the gain in dB; so it is important to include the dB units to distinguish them. For example, if P is twice P_i, then

$$G = 10\log(P/P_i) = 10\log(2) = 3.0\,\mathrm{dB}. \tag{1.24}$$

We say that the amplifier has a gain of 3 dB. Because engineers use dB as commonly as numerical ratios, it is a good idea to learn the conversions. These are given in Table 1.3. Decibels are a relative measure based on the ratio of two powers. However, dBs are also used for absolute powers. For this we use a reference power level in place of P_i. It is common to use 1 watt or 1 milliwatt. For example, 4 W is written as 6 dBw. We add the "w" to dB to indicate that 1 W is the reference level. A femtowatt is written as −120 dBm, with "m" standing for milliwatt.

1.6 Mixers

One thing that you might notice from looking at the frequencies used by the different services is that they are much higher than the ones we hear, which are in the low kilohertz range. In a receiver, we need to be able to convert from a radio frequency down to an audio frequency. In a transmitter, we must go the other way. The device that shifts frequency is called a *mixer*. A mixer effectively multiplies two signals. The output contains two different frequencies that are the sum and difference of the original frequencies. This process is called *heterodyning*. To see how this works, let the two inputs be $\cos(2\pi f_1 t)$ and $\cos(2\pi f_2 t)$. The product is given by

$$V(t) = \cos(2\pi f_1 t)\cos(2\pi f_2 t) = (1/2)\cos(2\pi f_+ t) + (1/2)\cos(2\pi f_- t), \tag{1.25}$$

where f_+ is the *sum frequency*, given by

$$f_+ = f_1 + f_2, \tag{1.26}$$

and f_- is the *difference frequency*, given by

$$f_- = |f_1 - f_2|. \tag{1.27}$$

Now we can identify two frequency components in the output. The sum-frequency component, V_+, is given by

$$V_+ = (1/2)\cos(2\pi f_+ t). \tag{1.28}$$

The other term is the difference-frequency voltage, V_-, given by

$$V_- = (1/2)\cos(2\pi f_- t). \tag{1.29}$$

The operation of a mixer is shown schematically in Figure 1.6b. In a receiver, one frequency might come from the antenna and the other from an oscillator inside the receiver. In practice, we may use either the sum frequency or the difference frequency, depending on whether we want to shift the frequency up or down.

1.7 Filters

The extra frequency that is generated by a mixer is a problem, and we use a device called a *filter* to remove it. A filter lets the frequencies we want through, while blocking the others. For example, if we wanted the difference frequency, but not the sum frequency, we would use a filter with a characteristic like that shown in Figure 1.7a. This is called a *low-pass* filter. In the figure, the output power P is plotted as a function of frequency. At low frequencies, P is at a maximum level P_m, but at high frequencies, the output power is greatly reduced. The lower frequency region where P is large is called the *pass band*. The higher frequencies that are rejected fall into the *stop band*. The response of a filter falls as we move into the

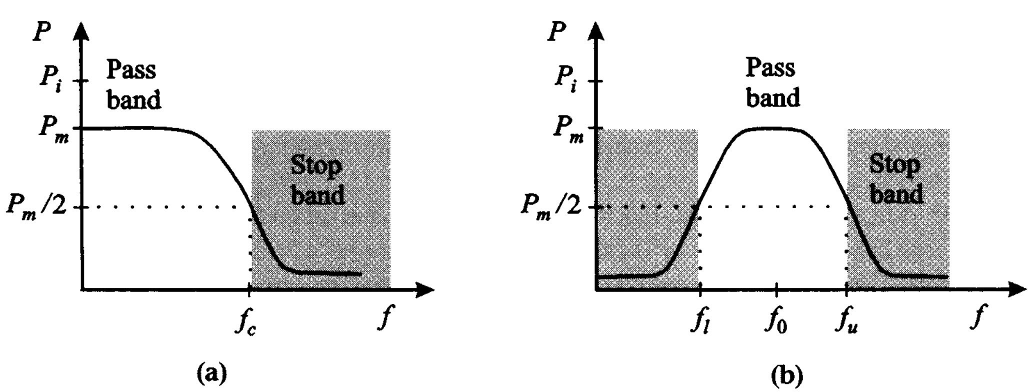


Figure 1.7. Low-pass filter response (a), and band-pass filter response (b). The plots show the output power versus frequency.

stop band, and so people also call this the *roll-off* region. Traditionally the boundary between the pass band and the stop band is taken as the frequency where the output power is reduced to half the maximum value. This half-power frequency is called the *cut-off frequency*, and we write it as f_c. We also refer to it as the 3-dB frequency, because 3 dB corresponds to a 2:1 power ratio.

We characterize a filter by two different numbers. In the pass band we specify the *loss*, which is the ratio of P_i to P_m. We write the loss factor L as

$$L = P_i/P_m. \tag{1.30}$$

We can think of the loss as the inverse of the gain that we defined for amplifiers in Equation 1.22. Often the loss is specified in dB, and we write it as

$$L = 10\ \log(P_i/P_m)\,\mathrm{dB}. \tag{1.31}$$

For example, if the maximum output power is half the input power, we say that the loss is 3 dB. The loss in dB is the negative of the gain in dB. This means that for our example, we could in principle talk about a gain of −3 dB instead of a loss of 3 dB. However, people typically use gain when the gain is positive and loss when the loss is positive. In addition, we specify a *rejection factor*, which shows how well the filter blocks a signal at a particular frequency in the stop band compared to the power in the pass band. We write the rejection factor R as

$$R = P_m/P, \tag{1.32}$$

where P is the output power in the stop band. We can write the rejection in dB as

$$R = 10\ \log(P_m/P)\,\mathrm{dB}. \tag{1.33}$$

For example, if the output power at a particular frequency in the stop band is a million times lower than the power in the pass band, we would say that the filter rejection is 60 dB.

Figure 1.7b shows a band-pass filter that rejects frequencies above and below a particular operating frequency f_0. In this case we need to consider two 3-dB frequencies, f_u and f_l. The bandwidth of the filter is the difference between the upper and lower 3-dB frequencies. We write the bandwidth as

$$\Delta f = f_u - f_l, \tag{1.34}$$

where Δ is the Greek capital *delta*. You might consider what the response curves for *high-pass* filters and *band-stop* filters would look like.

1.8 Direct-Conversion Receivers

Now we consider a heterodyne receiver that converts a radio signal to an audio frequency that we can hear. It is called a *direct-conversion* receiver because there is one mixer (Figure 1.8). The antenna picks up a signal, which is the mixer input. This signal is traditionally called the *RF* signal, with RF standing for radio

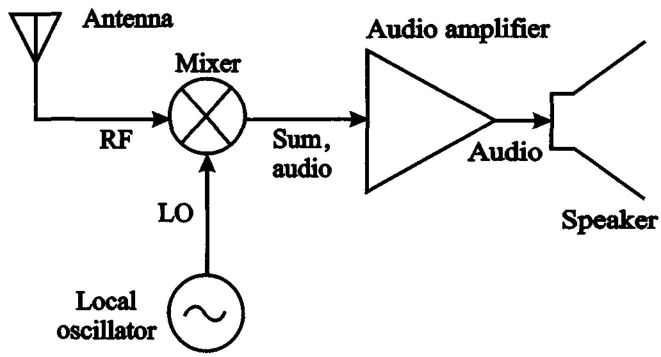


Figure 1.8. Direct-conversion receiver.

frequency. The other input to the mixer is from an oscillator inside the receiver. This oscillator is called the *local oscillator*, or LO for short. The output of the mixer will have a component at both the sum frequency and the difference frequency. The LO frequency is chosen so that the difference frequency is convenient for hearing, and it passes through an amplifier and loudspeaker, and then to our ears. For example, for receiving Morse Code, many operators like a frequency in the range of 600 Hz. For this we could use an LO frequency that is 600 Hz above the RF signal. Typically the sum frequency is out of the range of the amplifier, speaker, and our ears, and so it is ignored. To tune the receiver to pick out a radio station that we want to hear, we adjust the LO frequency until it sounds right. In fact, we will hear the station twice, once when the LO is 600 Hz above the RF frequency, and again when it is 600 Hz below.

Direct-conversion receivers are simple, but they have a fundamental problem: Signals the same distance above and below the LO produce an audio output at the same frequency. The frequency that we do not want to receive is called the *image*. This is shown in Figure 1.9. We can write the relations for the frequencies as

$$f_{rf} = f_{lo} - f_a, \tag{1.35}$$

$$f_i = f_{lo} + f_a, \tag{1.36}$$

where f_i is the image frequency and f_a is the audio frequency. This means that there is a problem if we have a station at the RF frequency and at the image. We hear

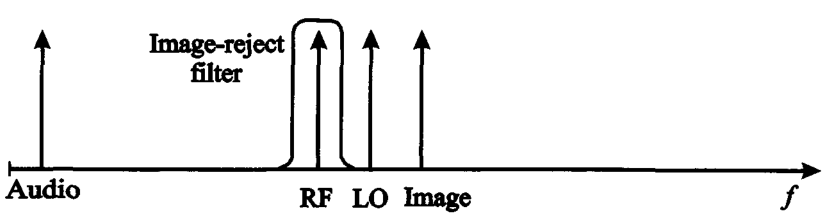


Figure 1.9. Image-frequency problem for a direct-conversion receiver. An RF band-pass filter is needed to remove the image.

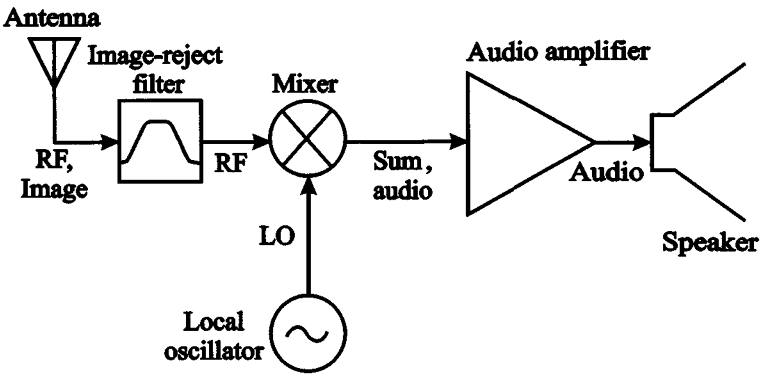


Figure 1.10. Adding an image-reject band-pass filter to a direct-conversion receiver to make a single-signal receiver.

both stations at the same time! This is a major limitation for a direct-conversion receiver.

There are several possible solutions. One is to add a second receiver and shift the oscillator waveform by a quarter of a cycle. It turns out that it is possible to cancel the image by splitting and combining with the correct shifts. This is complicated because it requires a second receiver. Sometimes channel assignments leave gaps, and the image can be put in one of the gaps. The fundamental solution is to put a band-pass filter in front of the mixer to reject the image (Figure 1.10). This requires a band-pass filter with an extremely narrow bandwidth. For example, in the 40-meter band that the NorCal 40A uses, the operating frequency is 7 MHz. The difference between the RF frequency and the image is only one part in 6,000. Surprisingly, it turns out to be quite practical to make a filter like this, with quartz crystals like those used in a wristwatch. However, there is a catch. The pass band in a crystal filter is fixed, and this means that we cannot tune to different stations.

1.9 Superheterodyne Receivers

We can solve the problem of a fixed filter pass band by adding a variable oscillator, another mixer, and another image-reject filter in front (Figure 1.11). This is the *superheterodyne* receiver, arguably the most important invention in the history of communications. The "superhet" is the classic receiver design that has been used in some form in the vast majority of radio and television receivers. The superheterodyne receiver was invented by an American, Howard Armstrong, during the First World War. He wanted to develop a receiver that could be used to intercept German radio transmissions. Armstrong was a brilliant engineer who also invented FM, but he was a tragic figure who spent many years fighting legal battles over patents. He finally committed suicide by jumping off a building in 1954.

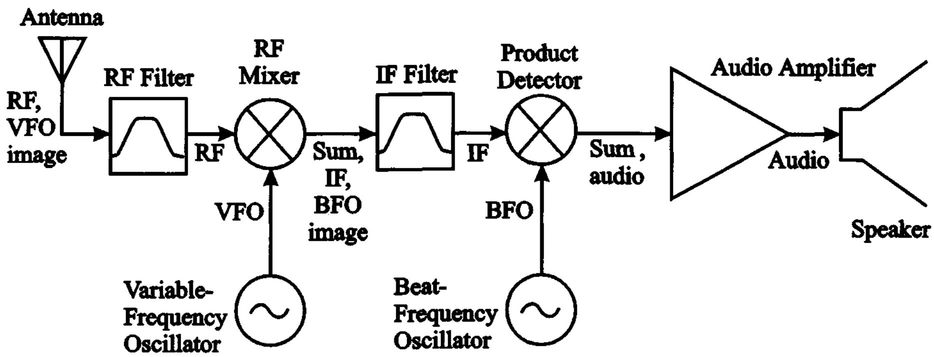


Figure 1.11. The superheterodyne receiver. An additional filter, mixer, and oscillator are added to remove the image.

A superheterodyne receiver is a complicated system, and we need to introduce terminology for the different filters, mixers, and oscillators. The first oscillator can tune in frequency, and it is called the *Variable-Frequency Oscillator*, or VFO. The second, fixed oscillator is the *Beat-Frequency Oscillator*, or BFO. The first mixer is the *RF Mixer*, and the second mixer is the *Product Detector*. The input frequency for the Product Detector is the *Intermediate Frequency*, or IF. The *RF Filter* rejects the image of the VFO, and the *IF Filter* rejects the image of the BFO.

Figure 1.12 shows the relationships among the frequencies in a superheterodyne receiver. These are complicated. We can write

$$f_{rf} = f_{if} + f_{vfo}, \tag{1.37}$$

$$f_{vi} = f_{if} - f_{vfo}, \tag{1.38}$$

$$f_{if} = f_{bfo} - f_a, \tag{1.39}$$

$$f_{bi} = f_{bfo} + f_a, \tag{1.40}$$

where f_{vi} is the VFO image and f_{bi} is the BFO image. We tune the receiver by changing the VFO frequency. The RF-Filter pass band must be broad enough to allow the full range of RF frequencies. In addition, the RF Filter must stop the VFO image. From the figure, we can see that the VFO image is far away from the RF frequency, and this makes it easy for the RF Filter to block it.

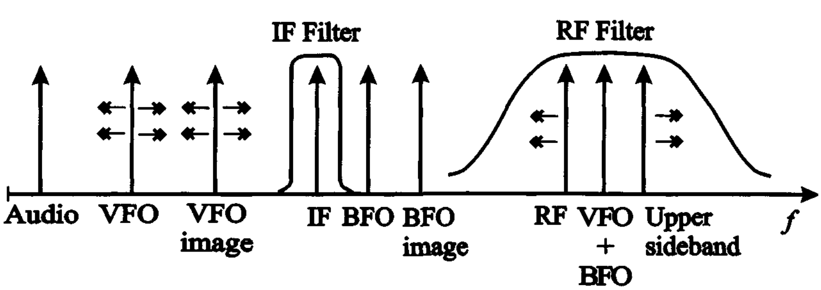


Figure 1.12. Frequencies for a superheterodyne receiver. The variable frequencies are indicated with horizontal arrows.

The BFO-image frequency f_{bi} is blocked by the IF Filter. The frequency that produces a signal at the image frequency f_{bi} is called the *upper sideband*, or USB, and it is given by

$$f_{usb} = f_{vfo} + f_{bi}. \tag{1.41}$$

If we substitute for f_{bi} from Equation 1.40, we find

$$f_{usb} = f_{vfo} + f_{bfo} + f_a. \tag{1.42}$$

The *lower sideband*, or LSB, is given by

$$f_{lsb} = f_{vfo} + f_{bfo} - f_a. \tag{1.43}$$

The lower-sideband frequency is the same as our RF frequency, and so we call this is a *lower-sideband receiver*. Equivalently, we could make an *upper-sideband receiver* by shifting the IF-Filter pass band above the frequency of the BFO so that the upper sideband would pass through, and the lower sideband would be rejected.

1.10 The NorCal 40A

The NorCal 40A transceiver includes a superheterodyne receiver and a transmitter. Figure 1.13 shows a block diagram for the transceiver, with the transmitter on the left, and the receiver on the right. There are two new audio components in the receiver that appear near the bottom. These are the *Automatic Gain Control*, or AGC, and the *AGC Detector*. The purpose of the AGC is to help the receiver adapt to signals with varying power levels. If a signal is strong, the AGC Detector will cause the AGC to reduce the audio signal before it gets to the Audio Amplifier. However, if the signal is weak, the AGC will let the audio signal through to the Audio Amplifier with little loss. This helps to keep the output sound level in a comfortable range for listening.

On the transmitter side, we start at the bottom of the figure. To transmit, a telegraph key activates the *Transmit Oscillator*, which produces a cosine with a frequency near 4.9 MHz. The output of the Transmit Oscillator mixes with the VFO, which is near 2.1 MHz, to give a sum frequency at 7.0 MHz. We write this as

$$f_+ = f_t + f_{vfo} = f_{rf} \approx 7.0\,\text{MHz}, \tag{1.44}$$

where f_t is the frequency of the Transmit Oscillator. We use the same VFO as for the receiver, so that we can tune both the receiver and transmitter together. The difference frequency is given by

$$f_- = f_t - f_{vfo} \approx 2.8\,\text{MHz}. \tag{1.45}$$

Since we do not want this frequency in the output, the Transmit Mixer is followed by the *Transmit Filter*, which passes the sum frequency and removes the difference frequency. At this point, the signal power is low, only about 10 μW. Next there is

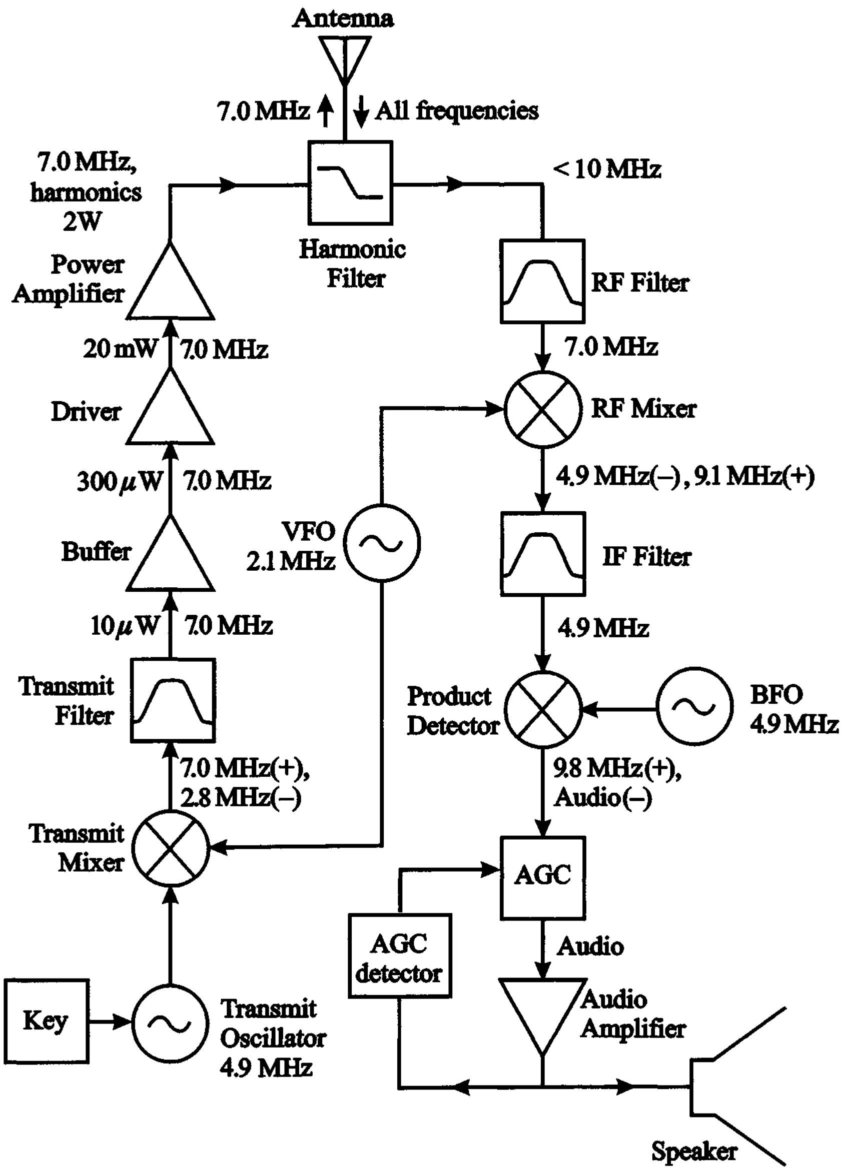


Figure 1.13. Block diagram for the NorCal 40A. Sum frequencies are noted by + signs, and difference frequencies by − signs. Adapted from Appendix C of the *NorCal 40A Assembly and Operating Manual*, by Wayne Burdick, published by Wilderness Radio. Used by permission.

a series of three amplifiers, the *Buffer*, the *Driver*, and the *Power Amplifier*, that raise the power level successively to 300 μW, 20 mW, and finally to 2 W. The output of the Power Amplifier contains the RF frequency, 7.0 MHz, but there is also power at the harmonic frequencies, 14 MHz and 21 MHz. Because these frequencies interfere with other operators, it is important to remove them. This is done with the *Harmonic Filter*, which is a low-pass filter with a 3-dB frequency of 10 MHz.

Inside the front cover is a complete schematic of the NorCal 40A. As you work through the problems, it is helpful to refer to this figure for component values and to see how the different circuits you build go together.

FURTHER READING

Paul Nahin's book, *The Science of Radio*, published by the American Institute of Physics, is a history of early radio, in addition to being a textbook on communications theory. Ken Burns's documentary, *Empire of the Air*, is a fascinating story of Howard Armstrong and the other pioneers of radio. It is available on video cassette through PBS Home Video.

Voltage, current, Kirchhoff's laws, and power all have simple definitions in this chapter, but the underlying ideas in electricity and magnetism are quite subtle and present problems for very fast circuits. My favorite book on this topic is *Electricity and Magnetism*, by the late Edward Purcell, published by McGraw-Hill. We will also explore these issues when we discuss transmission lines in Chapter 4. For broader coverage of radio engineering at a more advanced level, I recommend *Radio-Frequency Electronics*, by Jon Hagen, published by Cambridge University Press.