FOUNDATIONS

CHAPTER 1

INTRODUCTION

Developments in the field of electronics have constituted one of the great success stories of this century. Beginning with crude spark-gap transmitters and "cat's-whisker" detectors at the turn of the century, we have passed through a vacuum-tube era of considerable sophistication to a solid-state era in which the flood of stunning advances shows no signs of abating. Calculators, computers, and even talking machines with vocabularies of several hundred words are routinely manufactured on single chips of silicon as part of the technology of large-scale integration (LSI), and current developments in very large scale integration (VLSI) promise even more remarkable devices.

Perhaps as noteworthy is the pleasant trend toward increased performance per dollar. The cost of an electronic microcircuit routinely decreases to a fraction of its initial cost as the manufacturing process is perfected (see Fig. 8.87 for an example). In fact, it is often the case that the panel controls and cabinet hardware of an instrument cost more than the electronics inside.

On reading of these exciting new developments in electronics, you may get the impression that you should be able to construct powerful, elegant, yet inexpensive, little gadgets to do almost any conceivable task – all you need to know is how all these miracle devices work. If you've had that feeling, this book is for you. In it we have attempted to convey the excitement and know-how of the subject of electronics.

In this chapter we begin the study of the laws, rules of thumb, and tricks that constitute the art of electronics as we see it. It is necessary to begin at the beginning – with talk of voltage, current, power, and the components that make up electronic circuits. Because you can't touch, see, smell, or hear electricity, there will be a certain amount of abstraction (particularly in the first chapter), as well as some dependence on such visualizing instruments as oscilloscopes and voltmeters. In many ways the first chapter is also the most mathematical, in spite of our efforts to keep mathematics to a minimum in order to foster a good intuitive understanding of circuit design and behavior.

Once we have considered the foundations of electronics, we will quickly get into the "active" circuits (amplifiers, oscillators, logic circuits, etc.) that make electronics the exciting field it is. The reader with some background in electronics may wish to skip over this chapter, since it assumes no prior knowledge of electronics. Further generalizations at this time would be pointless, so let's just dive right in.

VOLTAGE, CURRENT, AND RESISTANCE

1.01 Voltage and current

There are two quantities that we like to keep track of in electronic circuits: voltage and current. These are usually changing with time; otherwise nothing interesting is happening.

Voltage (symbol: V, or sometimes E). The voltage between two points is the cost in energy (work done) required to move a unit of positive charge from the more negative point (lower potential) to the more positive point (higher potential). Equivalently, it is the energy released when a unit charge moves "downhill" from the higher potential to the lower. Voltage is also called *potential difference* or *electromotive force* (EMF). The unit of measure is the *volt*, with voltages usually expressed in volts (V), kilovolts ($1\text{kV} = 10^3\text{V}$), millivolts ($1\text{mV} = 10^{-3}\text{V}$), or microvolts ($1\mu\text{V} = 10^{-6}\text{V}$) (see the box on prefixes). A joule of work is needed to move a coulomb of charge through a potential difference of one volt. (The coulomb is the unit of electric charge, and it equals the charge of 6×10^{18} electrons, approximately.) For reasons that will become clear later, the opportunities to talk about nanovolts ($1\text{nV} = 10^{-9}\text{V}$) and megavolts ($1\text{MV} = 10^6\text{V}$) are rare.

Current (symbol: I). Current is the rate of flow of electric charge past a point. The unit of measure is the ampere, or amp, with currents usually expressed in amperes (A), milliamperes ($1\text{mA} = 10^{-3}\text{A}$), microamperes ($1\mu\text{A} = 10^{-6}\text{A}$), nanoamperes ($1\text{nA} = 10^{-9}\text{A}$), or occasionally picoamperes ($1\text{pA} = 10^{-12}\text{A}$). A current of one ampere equals a flow of one coulomb of charge per second. By convention, current in a circuit is considered to flow from a more positive point to a more negative point, even though the actual electron flow is in the opposite direction.

Important: Always refer to voltage *between* two points or *across* two points in a circuit. Always refer to current *through* a device or connection in a circuit.

To say something like "the voltage through a resistor ... " is nonsense, or worse. However, we do frequently speak of the voltage *at a point* in a circuit. This is always understood to mean voltage between that point and "ground," a common point in the circuit that everyone seems to know about. Soon you will, too.

We *generate* voltages by doing work on charges in devices such as batteries (electrochemical), generators (magnetic forces), solar cells (photovoltaic conversion of the energy of photons), etc. We *get* currents by placing voltages across things.

At this point you may well wonder how to "see" voltages and currents. The single most useful electronic instrument is the oscilloscope, which allows you to look at voltages (or occasionally currents) in a circuit as a function of time. We will deal with oscilloscopes, and also voltmeters, when we discuss signals shortly; for a preview, see the oscilloscope appendix (Appendix A) and the multimeter box later in this chapter.

In real circuits we connect things together with wires, metallic conductors, each of which has the same voltage on it everywhere (with respect to ground, say). (In the domain of high frequencies or low impedances, that isn't strictly true, and we will have more to say about this later. For now, it's a good approximation.) We mention this now so that you will realize that an actual circuit doesn't have to look like its schematic diagram, because wires can be rearranged.

Here are some simple rules about voltage and current:

1. The sum of the currents into a point in a circuit equals the sum of the currents out (conservation of charge). This is sometimes called Kirchhoff's current law. Engineers like to refer to such a point as a *node*. From this, we get the following: For a series circuit (a bunch of two-terminal things all connected end-to-end) the current is the same everywhere.

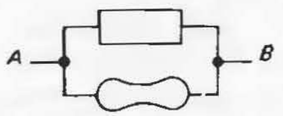


Figure 1.1

2. Things hooked in parallel (Fig. 1.1) have the same voltage across them. Restated, the sum of the "voltage drops" from A to B via one path through a circuit equals the sum by any other route equals the voltage between A and B. Sometimes this is stated as follows: The sum of the voltage drops around any closed circuit is zero. This is Kirchhoff's voltage law.

3. The power (work per unit time) consumed by a circuit device is

$$P = VI$$

This is simply (work/charge) × (charge/time). For V in volts and I in amps, P comes out in watts. Watts are joules per second (1W = 1J/s).

Power goes into heat (usually), or sometimes mechanical work (motors), radiated energy (lamps, transmitters), or stored energy (batteries, capacitors). Managing the heat load in a complicated system (e.g., a computer, in which many kilowatts of electrical energy are converted to heat, with the energetically insignificant by-product of a few pages of computational results) can be a crucial part of the system design.

PREFIXES

These prefixes are universally used to scale units in science and engineering.

Multiple	*Prefix*	*Symbol*
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f

When abbreviating a unit with a prefix, the symbol for the unit follows the prefix without space. Be careful about upper-case and lower-case letters (especially m and M) in both prefix and unit: 1mW is a milliwatt, or one-thousandth of a watt; 1MHz is 1 million hertz. In general, units are spelled with lower-case letters, even when they are derived from proper names. The unit name is not capitalized when it is spelled out and used with a prefix, only when abbreviated. Thus: hertz and kilohertz, but Hz and kHz; watt, milliwatt, and megawatt, but W, mW, and MW.

Soon, when we deal with periodically varying voltages and currents, we will have to generalize the simple equation $P = VI$ to deal with *average* power, but it's correct as a statement of *instantaneous* power just as it stands.

Incidentally, don't call current "amperage"; that's strictly bush-league. The same caution will apply to the term "ohmage" when we get to resistance in the next section.

1.02 Relationship between voltage and current: resistors

This is a long and interesting story. It is the heart of electronics. Crudely speaking, the name of the game is to make and use gadgets that have interesting and useful I-versus-V characteristics. Resistors (I simply proportional to V), capacitors (I proportional to rate of change of V), diodes (I flows in only one direction), thermistors (temperature-dependent resistor), photoresistors (light-dependent resistor), strain gauges (strain-dependent resistor), etc., are examples. We will gradually get into some of these exotic devices; for now, we will start with the most mundane (and most widely used) circuit element, the resistor (Fig. 1.2).

Figure 1.2

Resistance and resistors

It is an interesting fact that the current through a metallic conductor (or other partially conducting material) is proportional to the voltage across it. (In the case of wire conductors used in circuits, we usually choose a thick enough gauge of wire so that these "voltage drops" will be negligible.) This is by no means a universal law for all objects. For instance, the current through a neon bulb is a highly nonlinear function of the applied voltage (it is zero up to a critical voltage, at which point it rises dramatically). The same goes for a variety of interesting special devices – diodes, transistors, light bulbs, etc. (If you are interested in understanding why metallic conductors behave this way, read sections 4.4–4.5 in the *Berkeley Physics Course*, Vol. II, see Bibliography). A resistor is made out of some conducting stuff (carbon, or a thin metal or carbon film, or wire of poor conductivity), with a wire coming out each end. It is characterized by its resistance:

$$R = V/I$$

R is in ohms for V in volts and I in amps. This is known as Ohm's law. Typical resistors of the most frequently used type (carbon composition) come in values from 1 ohm (1Ω) to about 22 megohms (22MΩ). Resistors are also characterized by how

RESISTORS

Resistors are truly ubiquitous. There are almost as many types as there are applications. Resistors are used in amplifiers as loads for active devices, in bias networks, and as feedback elements. In combination with capacitors they establish time constants and act as filters. They are used to set operating currents and signal levels. Resistors are used in power circuits to reduce voltages by dissipating power, to measure currents, and to discharge capacitors after power is removed. They are used in precision circuits to establish currents, to provide accurate voltage ratios, and to set precise gain values. In logic circuits they act as bus and line terminators and as "pull-up" and "pull-down" resistors. In high-voltage circuits they are used to measure voltages and to equalize leakage currents among diodes or capacitors connected in series. In radiofrequency circuits they are even used as coil forms for inductors.

Resistors are available with resistances from 0.01 ohm through 10^{12} ohms, standard power ratings from 1/8 watt through 250 watts, and accuracies from 0.005% through 20%. Resistors can be made from carbon-composition moldings, from metal films, from wire wound on a form, or from semiconductor elements similar to field-effect transistors (FETs). But by far the most familiar resistor is the 1/4 or 1/2 watt carbon-composition resistor. These are available in a standard set of values ranging from 1 ohm to 100 megohms with twice as many values available for the 5% tolerance as for the 10% types (see Appendix C). We prefer the Allen-Bradley type AB (1/4 watt, 5%) resistor for general use because of its clear marking, secure lead seating, and stable properties.

Resistors are so easy to use that they're often taken for granted. They're not perfect, though, and it is worthwhile to look at some of their defects. The popular 5% composition type, in particular, although fine for nearly all noncritical circuit applications, is not stable enough for precision applications. You should know about its limitations so that you won't be surprised someday. Its principal defects are variations in resistance with temperature, voltage, time, and humidity. Other defects may relate to inductance (which may be serious at high frequencies), the development of thermal hot spots in power applications, or electrical noise generation in low-noise amplifiers. The following specifications are worst-case values; typically you'll do better, but don't count on it!

SPECIFICATIONS FOR ALLEN-BRADLEY AB SERIES TYPE CB

Standard tolerance is ±5% under nominal conditions. Maximum power for 70°C ambient temperature is 0.25 watt, which will raise the internal temperature to 150°C. The maximum applied voltage specification is $(0.25R)^{1/2}$ or 250 volts, whichever is less. They mean it! (See Fig. 6.53.) A single 5 second overvoltage to 400 volts can cause a permanent change in resistance by 2%.

	Resistance change		*Permanent?*
	(R = 1k)	(R = 10M)	
Soldering (350°C at 1/8 inch)	±2%	±2%	yes
Load cycling (500 ON/OFF cycles in 1000 hours)	+4%–6%	+4%–6%	yes
Vibration (20*g*) and shock (100*g*)	±2%	±2%	yes
Humidity (95% relative humidity at 40°C)	+6%	+10%	no
Voltage coefficient (10V change)	−0.15%	−0.3%	no
Temperature (25°C to −15°C)	+2.5%	+4.5%	no
Temperature (25°C to 85°C)	+3.3%	+5.9%	no

For applications that require any real accuracy or stability a 1% metal-film resistor (see Appendix D) should be used. They can be expected to have stability of better than 0.1% under normal conditions and better than 1% under worst-case treatment. Precision wire-wound resistors are available for the most demanding applications. For power dissipation above about 0.1 watt, a resistor of higher power rating should be used. Carbon-composition resistors are available with ratings up to 2 watts, and wire-wound power resistors are available for higher power. For demanding power applications, the conduction-cooled type of power resistor delivers better performance. These carefully designed resistors are available at 1% tolerance and can be operated at core temperatures up to 250°C with dependable long life. Allowable resistor power dissipation depends on air flow, thermal conduction via the resistor leads, and circuit density; thus, a resistor's power rating should be considered a rough guideline. Note also that resistor power ratings refer to *average* power dissipation and may be substantially exceeded for short periods of time (a few seconds or more, depending on the resistor's "thermal mass").

much power they can safely dissipate (the most commonly used ones are rated at 1/4 watt) and by other parameters such as tolerance (accuracy), temperature coefficient, noise, voltage coefficient (the extent to which R depends on applied V), stability with time, inductance, etc. See the box on resistors and Appendixes C and D for further details.

Roughly speaking, resistors are used to convert a voltage to a current, and vice versa. This may sound awfully trite, but you will soon see what we mean.

Resistors in series and parallel

From the definition of R, some simple results follow:

Figure 1.3

1. The resistance of two resistors in series (Fig. 1.3) is

$$R = R_1 + R_2$$

By putting resistors in series, you always get a *larger* resistor.

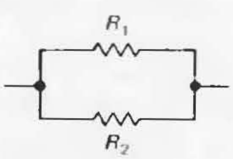


Figure 1.4

2. The resistance of two resistors in parallel (Fig. 1.4) is

$$R = \frac{R_1 R_2}{R_1 + R_2} \quad \text{or} \quad R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}$$

By putting resistors in parallel, you always get a *smaller* resistor. Resistance is measured in ohms (Ω), but in practice we frequently omit the Ω symbol when referring to resistors that are more than 1000Ω ($1\mathrm{k}\Omega$). Thus, a $10\mathrm{k}\Omega$ resistor is often referred to as a 10k resistor, and a $1\mathrm{M}\Omega$ resistor as a 1M resistor (or 1 meg). On schematic diagrams the symbol Ω is often omitted altogether. If this bores you, please have patience – we'll soon get to numerous amusing applications.

EXERCISE 1.1
You have a 5k resistor and a 10k resistor. What is their combined resistance (a) in series and (b) in parallel?

EXERCISE 1.2
If you place a 1 ohm resistor across a 12 volt car battery, how much power will it dissipate?

EXERCISE 1.3
Prove the formulas for series and parallel resistors.

EXERCISE 1.4
Show that several resistors in parallel have resistance

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots}$$

A trick for parallel resistors: Beginners tend to get carried away with complicated algebra in designing or trying to understand electronics. Now is the time to begin learning intuition and shortcuts.

Shortcut no. 1 A large resistor in series (parallel) with a small resistor has the resistance of the larger (smaller) one, roughly.

Shortcut no. 2 Suppose you want the resistance of 5k in parallel with 10k. If you think of the 5k as two 10k's in parallel, then the whole circuit is like three 10k's in parallel. Because the resistance of n equal resistors in parallel is $1/n$th the resistance of the individual resistors, the answer in this case is 10k/3, or 3.33k. This trick is handy because it allows you to analyze circuits quickly in your head, without distractions. We want to encourage mental designing, or at least "back of the envelope" designing, for idea brainstorming.

Some more home-grown philosophy: There is a tendency among beginners to want to compute resistor values and other circuit component values to many significant places, and the availability of inexpensive calculators has only made matters worse. There are two reasons you should try to avoid falling into this habit: (a) the components themselves are of finite precision (typical resistors are $\pm 5\%$; the parameters that characterize transistors, say, frequently are known only to a factor of two); (b) one mark of a good circuit design is insensitivity of the finished circuit to precise values of the components (there are exceptions, of course). You'll also learn circuit intuition more quickly if you get into the habit of doing approximate calculations in your head, rather than watching meaningless numbers pop up on a calculator display.

In trying to develop intuition about resistance, some people find it helpful to think about *conductance*, $G = 1/R$. The current through a device of conductance G bridging a voltage V is then given by $I = GV$ (Ohm's law). A small resistance is a large conductance, with correspondingly large current under the influence of an applied voltage.

Viewed in this light, the formula for parallel resistors is obvious: When several resistors or conducting paths are connected across the same voltage, the total current is the sum of the individual currents. Therefore the net conductance is simply the sum of the individual conductances, $G = G_1 + G_2 + G_3 + \cdots$, which is the same as the formula for parallel resistors derived earlier.

Engineers are fond of defining reciprocal units, and they have designated the unit of conductance the siemens ($S = 1/\Omega$), also known as the mho (that's ohm spelled backward, given the symbol $\mho$). Although the concept of conductance is helpful in developing intuition, it is not used widely; most people prefer to talk about resistance instead.

Power in resistors

The power dissipated by a resistor (or any other device) is $P = IV$. Using Ohm's law, you can get the equivalent forms $P = I^2R$ and $P = V^2/R$.

EXERCISE 1.5

Show that it is not possible to exceed the power rating of a 1/4 watt resistor of resistance greater than 1k, no matter how you connect it, in a circuit operating from a 15 volt battery.

EXERCISE 1.6

Optional exercise: New York City requires about 10^{10} watts of electrical power, at 110 volts (this is plausible: 10 million people averaging 1 kilowatt each). A heavy power cable might be an inch in diameter. Let's calculate what will happen if we try to supply the power through a cable 1 foot in diameter made of pure copper. Its resistance is $0.05\mu\Omega$ (5×10^{-8} ohms) per foot. Calculate (a) the power lost per foot from "I^2R losses," (b) the length of cable over which you will lose all 10^{10} watts, and (c) how hot the cable will get, if you know the physics involved ($\sigma = 6 \times 10^{12}\mathrm{W/^\circ K^4 cm^2}$).

If you have done your computations correctly, the result should seem preposterous. What is the solution to this puzzle?

Input and output

Nearly all electronic circuits accept some sort of applied *input* (usually a voltage) and produce some sort of corresponding *output* (which again is often a voltage). For example, an audio amplifier might produce a (varying) output voltage that is 100 times as large as a (similarly varying) input voltage. When describing such an amplifier, we imagine measuring the output voltage for a given applied input voltage. Engineers speak of the *transfer function* H, the ratio of (measured) output divided by (applied) input; for the audio amplifier above, H is simply a constant (H = 100). We'll get to amplifiers soon enough, in the next chapter. However, with just resistors we can already look at a very important circuit fragment, the *voltage divider* (which you might call a "de-amplifier").

1.03 Voltage dividers

We now come to the subject of the voltage divider, one of the most widespread electronic circuit fragments. Show us any real-life circuit and we'll show you half a dozen voltage dividers. To put it very simply, a voltage divider is a circuit that, given a certain voltage input, produces a predictable fraction of the input voltage as the output voltage. The simplest voltage divider is shown in Figure 1.5.

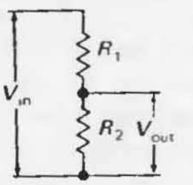


Figure 1.5. Voltage divider. An applied voltage V_{in} results in a (smaller) output voltage V_{out}.

What is V_{out}? Well, the current (same everywhere, assuming no "load" on the output) is

$$I = \frac{V_{in}}{R_1 + R_2}$$

(We've used the definition of resistance and the series law.) Then, for R_2,

$$V_{out} = IR_2 = \frac{R_2}{R_1 + R_2} V_{in}$$

Note that the output voltage is always less than (or equal to) the input voltage; that's why it's called a divider. You could get amplification (more output than input) if one of the resistances were negative. This isn't as crazy as it sounds; it is possible to make devices with negative "incremental" resistances (e.g., the tunnel diode) or even true negative resistances (e.g., the negative-impedance converter that we will talk about later in the book). However, these applications are rather specialized and need not concern you now.

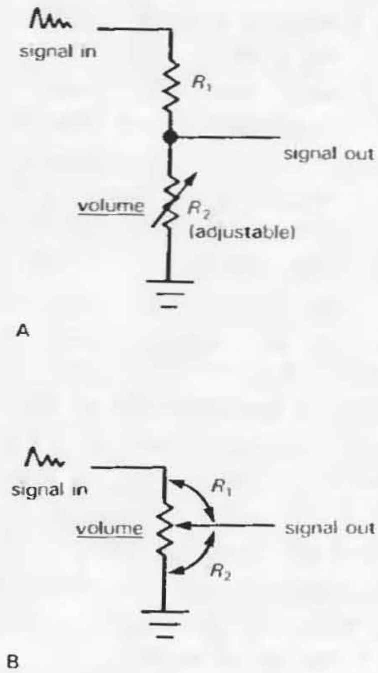


Figure 1.6. An adjustable voltage divider can be made from a fixed and variable resistor, or from a potentiometer.

Voltage dividers are often used in circuits to generate a particular voltage from a larger fixed (or varying) voltage. For instance, if V_{in} is a varying voltage and R_2 is an adjustable resistor (Fig. 1.6A), you have a "volume control"; more simply, the combination $R_1 R_2$ can be made from a single variable resistor, or *potentiometer* (Fig. 1.6B). The humble voltage divider is even more useful, though, as a way of *thinking* about a circuit: the input voltage and upper resistance might represent the output of an amplifier, say, and the lower resistance might represent the input of the following stage. In this case the voltage-divider equation tells you how much signal gets to the input of that last stage. This will all become clearer after you know about a remarkable fact (Thévenin's theorem) that will be discussed later. First, though, a short aside on voltage sources and current sources.

1.04 Voltage and current sources

A perfect voltage source is a two-terminal *black box* that maintains a fixed voltage drop across its terminals, regardless of load resistance. For instance, this means that it must supply a current $I = V/R$ when a resistance R is attached to its terminals. A real voltage source can supply only a finite maximum current, and in addition it generally behaves like a perfect voltage source with a small resistance in series. Obviously, the smaller this series resistance, the better. For example, a standard 9 volt alkaline battery behaves like a perfect 9 volt voltage source in series with a 3 ohm resistor and can provide a maximum current (when shorted) of 3 amps (which, however, will kill the battery in a few minutes). A voltage source "likes" an open-circuit load and "hates" a short-circuit load, for obvious reasons. (The terms "open circuit" and "short circuit" mean the obvious: An open circuit has nothing connected to it, whereas a short circuit is a piece of wire bridging the output.) The symbols used to indicate a voltage source are shown in Figure 1.7.

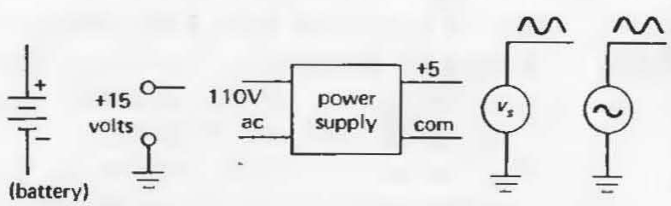


Figure 1.7. Voltage sources can be either steady (dc) or varying (ac).

A perfect current source is a two-terminal black box that maintains a constant current through the external circuit, regardless of load resistance or applied voltage. In order to do this it must be capable of supplying any necessary voltage across its terminals. Real current sources (a much-neglected subject in most textbooks) have a limit to the voltage they can provide (called the *output voltage compliance*, or just *compliance*), and in addition they do not provide absolutely constant output current. A current source "likes" a short-circuit load and "hates" an open-circuit load. The symbols used to indicate a current source are shown in Figure 1.8.

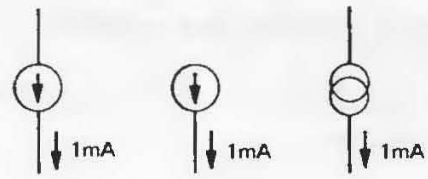


Figure 1.8. Current-source symbols.

A battery is a real-life approximation of a voltage source (there is no analog for a current source). A standard D-size flashlight cell, for instance, has a terminal voltage of 1.5 volts, an equivalent series resistance of about 1/4 ohm, and total energy capacity of about 10,000 watt-seconds (its characteristics gradually deteriorate with use; at the end of its life, the voltage may be about 1.0 volt, with an internal series resistance of several ohms). It is easy to construct voltage sources with far better characteristics, as you will learn when we come to the subject of feedback. Except in devices intended for portability, the use of batteries in electronic devices is rare. We will treat the interesting subject of low-power (battery-operated) design in Chapter 14.

MULTIMETERS

There are numerous instruments that let you measure voltages and currents in a circuit. The oscilloscope (see Appendix A) is the most versatile; it lets you "see" voltages versus time at one or more points in a circuit. Logic probes and logic analyzers are special-purpose instruments for troubleshooting digital circuits. The simple multimeter provides a good way to measure voltage, current, and resistance, often with good precision; however, it responds slowly, and thus it cannot replace the oscilloscope where changing voltages are of interest. Multimeters are of two varieties: those that indicate measurements on a conventional scale with a moving pointer, and those that use a digital display.

The standard VOM (volt-ohm-milliammeter) multimeter uses a meter movement that measures current (typically $50\mu A$ full scale). (See a less-design-oriented electronics book for pretty pictures of the innards of meter movements; for our purposes, it suffices to say that it uses coils and magnets.) To measure voltage, the VOM puts a resistor in series with the basic movement. For instance, one kind of VOM will generate a 1 volt (full-scale) range by putting a 20k resistor in series with the standard $50\mu A$ movement; higher voltage ranges use correspondingly larger resistors. Such a VOM is specified as 20,000 ohms/volt, meaning that it looks like a resistor whose value is 20k multiplied by the full-scale voltage of the particular range selected. Full scale on any voltage range is 1/20,000, or $50\mu A$. It should be clear that one of these voltmeters disturbs a circuit less on a higher range, since it looks like a higher resistance (think of the voltmeter as the lower leg of a voltage divider, with the Thévenin resistance of the circuit you are measuring as the upper resistor). Ideally, a voltmeter should have infinite input resistance.

Nowadays there are various meters with some electronic amplification whose input resistance may be as large as 10^9 ohms. Most digital meters, and even a number of analog-reading meters that use FETs (field-effect transistors, see Chapter 3), are of this type. Warning: Sometimes the input resistance of FET-input meters is very high on the most sensitive ranges, dropping to a lower resistance for the higher ranges. For instance, an input resistance of 10^9 ohms on the 0.2 volt and 2 volt ranges, and 10^7 ohms on all higher ranges, is typical. Read the specifications carefully! For measurements on most transistor circuits, 20,000 ohms/volt is fine, and there will be little loading effect on the circuit by the meter. In any case, it is easy to calculate how serious the effect is by using the voltage-divider equation. Typically, multimeters provide voltage ranges from a volt (or less) to a kilovolt (or more), full scale.

A VOM can be used to measure current by simply using the bare meter movement (for our preceding example, this would give a range of $50\mu A$ full scale) or by shunting (paralleling) the movement with a small resistor. Because the meter movement itself requires a small voltage drop, typically 0.25 volt, to produce a full-scale deflection, the shunt is chosen by the meter manufacturer (all you do is set the range switch to the range you want) so that the full-scale current will produce that voltage drop through the parallel combination of the meter resistance and the shunt resistance. Ideally, a current-measuring meter should have zero resistance in order not to disturb the circuit under test, since it must be put in series with the circuit. In practice, you tolerate a few tenths of a volt drop (sometimes called "voltage burden") with both VOMs and digital multimeters. Typically, multimeters provide current ranges from $50\mu A$ (or less) to an amp (or more), full scale.

Multimeters also have one or more batteries in them to power the resistance measurement. By supplying a small current and measuring the voltage drop, they measure resistance, with several ranges to cover values from an ohm (or less) to 10 megohms (or more).

Important: Don't try to measure "the current of a voltage source," for instance by sticking the meter across the wall plug; the same applies for ohms. This is the leading cause of blown-out meters.

EXERCISE 1.7

What will a 20,000 ohms/volt meter read, on its 1 volt scale, when attached to a 1 volt source with an internal resistance of 10k? What will it read when attached to a 10k–10k voltage divider driven by a "stiff" (zero source resistance) 1 volt source?

EXERCISE 1.8

A $50\mu A$ meter movement has an internal resistance of 5k. What shunt resistance is needed to convert it to a 0–1 amp meter? What series resistance will convert it to a 0–10 volt meter?

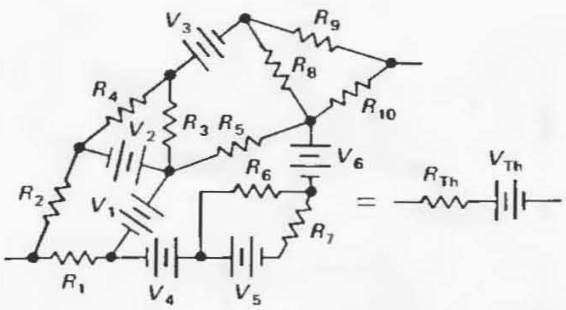


Figure 1.9

1.05 Thévenin's equivalent circuit

Thévenin's theorem states that any two-terminal network of resistors and voltage sources is equivalent to a single resistor R in series with a single voltage source V. This is remarkable. Any mess of batteries and resistors can be mimicked with one battery and one resistor (Fig. 1.9). (Incidentally, there's another theorem, Norton's theorem, that says you can do the same thing with a current source in parallel with a resistor.)

How do you figure out the Thévenin equivalent R_{Th} and V_{Th} for a given circuit? Easy! V_{Th} is the open-circuit voltage of the Thévenin equivalent circuit; so if the two circuits behave identically, it must also be the open-circuit voltage of the given circuit (which you get by calculation, if you know what the circuit is, or by measurement, if you don't). Then you find R_{Th} by noting that the short-circuit current of the equivalent circuit is $V_{\text{Th}}/R_{\text{Th}}$. In other words,

$$V_{\text{Th}} = V \text{ (open circuit)}$$

$$R_{\text{Th}} = \frac{V \text{ (open circuit)}}{I \text{ (short circuit)}}$$

Let's apply this method to the voltage divider, which must have a Thévenin equivalent:

1. The open-circuit voltage is

$$V = V_{\text{in}} \frac{R_2}{R_1 + R_2}$$

2. The short-circuit current is

$$V_{\text{in}}/R_1$$

So the Thévenin equivalent circuit is a voltage source

$$V_{\text{Th}} = V_{\text{in}} \frac{R_2}{R_1 + R_2}$$

in series with a resistor

$$R_{\text{Th}} = \frac{R_1 R_2}{R_1 + R_2}$$

(It is not a coincidence that this happens to be the parallel resistance of R_1 and R_2. The reason will become clear later.)

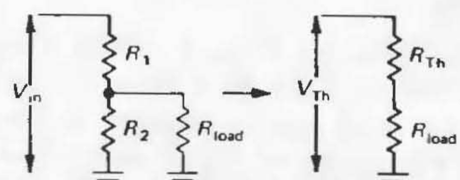


Figure 1.10

From this example it is easy to see that a voltage divider is not a very good battery, in the sense that its output voltage drops severely when a load is attached. As an example, consider Exercise 1.9. You now know everything you need to know to calculate exactly how much the output will drop for a given load resistance: Use the Thévenin equivalent circuit, attach a load, and calculate the new output, noting that the new circuit is nothing but a voltage divider (Fig. 1.10).

EXERCISE 1.9

For the circuit shown in Figure 1.10, with $V_{\text{in}} =$ 30V and $R_1 = R_2 =$ 10k, find (a) the output voltage with no load attached (the open-circuit voltage); (b) the output voltage with a 10k load (treat as voltage divider, with R_2 and R_{load} combined into a single resistor); (c) the Thévenin equivalent circuit; (d) the same as in part b, but using the Thévenin equivalent circuit (again, you wind up with a voltage divider; the answer should agree with the result in part b); (e) the power dissipated in each of the resistors.

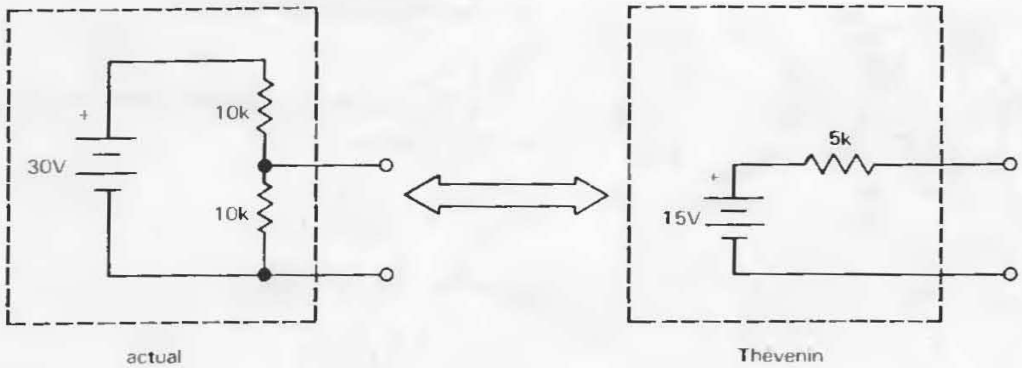


Figure 1.11

Equivalent source resistance and circuit loading

As you have just seen, a voltage divider powered from some fixed voltage is equivalent to some smaller voltage source in series with a resistor; for example, the output terminals of a 10k–10k voltage divider driven by a perfect 30 volt battery are precisely equivalent to a perfect 15 volt battery in series with a 5k resistor (Fig. 1.11). Attaching a load resistor causes the voltage divider's output to drop, owing to the finite *source resistance* (Thévenin equivalent resistance of the voltage divider output, viewed as a source of voltage). This is often undesirable. One solution to the problem of making a stiff voltage source ("stiff" is used in this context to describe something that doesn't bend under load) might be to use much smaller resistors in a voltage divider. Occasionally this brute-force approach is useful. However, it is usually best to construct a voltage source, or power supply, as it's commonly called, using active components like transistors or operational amplifiers, which we will treat in Chapters 2–4. In this way you can easily make a voltage source with internal (Thévenin equivalent) resistance measured in milliohms (thousandths of an ohm), without the large currents and dissipation of power characteristic of a low-resistance voltage divider delivering the same performance. In addition, with an active power supply it is easy to make the output voltage adjustable.

The concept of equivalent internal resistance applies to all sorts of sources, not just batteries and voltage dividers. Signal sources (e.g., oscillators, amplifiers, and sensing devices) all have an equivalent internal resistance. Attaching a load whose resistance is less than or even comparable to the internal resistance will reduce the output considerably. This undesirable reduction of the open-circuit voltage (or signal) by the load is called "circuit loading." Therefore, you should strive to make $R_{\mathrm{load}} \gg R_{\mathrm{internal}}$, because a high-resistance load has little attenuating effect on the source (Fig. 1.12). You will see numerous circuit examples in the chapters ahead. This high-resistance condition ideally

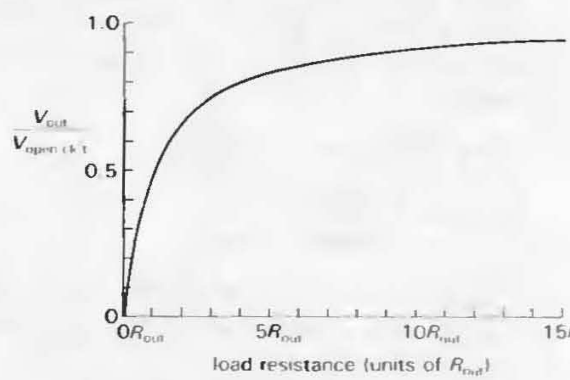


Figure 1.12. To avoid attenuating a signal source below its open-circuit voltage, keep the load resistance large compared with the output resistance.

characterizes measuring instruments such as voltmeters and oscilloscopes. (There are exceptions to this general principle; for example, we will talk about transmission lines and radiofrequency techniques, where you must "match impedances" in order to prevent the reflection and loss of power.)

A word on language: You frequently hear things like "the resistance looking into the voltage divider," or "the output sees a load of so-and-so many ohms," as if circuits had eyes. It's OK (in fact, it's a rather good way to keep straight which resistance you're talking about) to say what part of the circuit is doing the "looking."

Power transfer

Here is an interesting problem: What load resistance will result in maximum power being transferred to the load for a given source resistance? (The terms *source resistance, internal resistance*, and *Thévenin equivalent resistance* all mean the same thing.) It is easy to see that both $R_{\text{load}} = 0$ and $R_{\text{load}} = \infty$ result in zero power transferred, because $R_{\text{load}} = 0$ means that $V_{\text{load}} = 0$ and $I_{\text{load}} = V_{\text{source}}/R_{\text{source}}$, so that $P_{\text{load}} = V_{\text{load}}I_{\text{load}} = 0$. But $R_{\text{load}} = \infty$ means that $V_{\text{load}} = V_{\text{source}}$ and $I_{\text{load}} = 0$, so that $P_{\text{load}} = 0$. There has to be a maximum in between.

EXERCISE 1.10

Show that $R_{\text{load}} = R_{\text{source}}$ maximizes the power in the load for a given source resistance. Note: Skip this exercise if you don't know calculus, and take it on faith that the answer is true.

Lest this example leave the wrong impression, we would like to emphasize again that circuits are ordinarily designed so that the load resistance is much greater than the source resistance of the signal that drives the load.

1.06 Small-signal resistance

We often deal with electronic devices for which I is not proportional to V; in such cases there's not much point in talking about resistance, since the ratio V/I will depend on V, rather than being a nice constant, independent of V. For these devices it is useful to know the slope of the V–I curve, in other words, the ratio of a small change in applied voltage to the resulting change in current through the device, $\Delta V/\Delta I$ (or dV/dI). This quantity has the units of resistance (ohms) and substitutes for resistance in many calculations. It is called the small-signal resistance, incremental resistance, or dynamic resistance.

Zener diodes

As an example, consider the *zener diode*, which has the V–I curve shown in Figure 1.13. Zeners are used to create a constant voltage inside a circuit somewhere, simply by providing them with a (roughly constant) current derived from a higher voltage within the circuit. For example, the zener diode in Figure 1.13 will convert an applied current in the range shown to a corresponding (but narrower) range of voltages. It is important to know how the resulting zener voltage will change with applied current; this is a measure of its "regulation" against changes in the driving current provided to it. Included in the specifications of a zener will be its dynamic resistance, given at a certain current. (Useful fact: the dynamic resistance of a zener diode varies roughly in inverse proportion to current.) For example, a zener might have a dynamic resistance of 10 ohms at 10mA, at its zener voltage of 5 volts. Using the definition of dynamic resistance, we find that a 10% change in applied current will therefore result in a change in voltage of

$$\Delta V = R_{\text{dyn}}\Delta I = 10 \times 0.1 \times 0.01 = 10\text{mV}$$

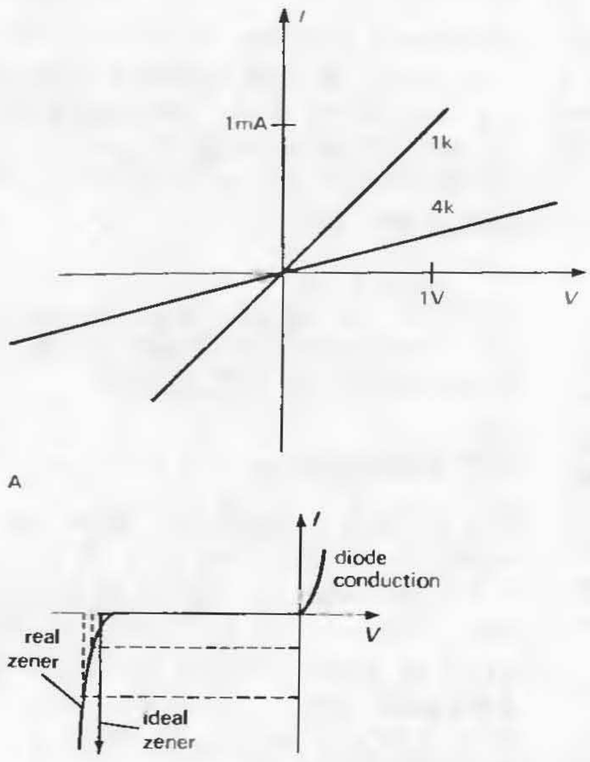


Figure 1.13. *V–I* curves.
A. Resistor (linear).
B. Zener diode (nonlinear).

or

$$\Delta V/V = 0.002 = 0.2\%$$

thus demonstrating good voltage-regulating ability. In this sort of application you frequently get the zener current through a resistor from a higher voltage available somewhere in the circuit, as in Figure 1.14.

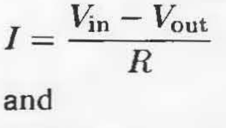


Figure 1.14. Zener regulator.

Then,

$$I = \frac{V_{in} - V_{out}}{R}$$

and

$$\Delta I = \frac{\Delta V_{in} - \Delta V_{out}}{R}$$

so

$$\Delta V_{out} = R_{dyn}\Delta I = \frac{R_{dyn}}{R}(\Delta V_{in} - \Delta V_{out})$$

and finally

$$\Delta V_{out} = \frac{R_{dyn}}{R + R_{dyn}}\Delta V_{in}$$

Thus, for *changes* in voltage, the circuit behaves like a voltage divider, with the zener replaced by a resistor equal to its dynamic resistance at the operating current. This is the utility of incremental resistance. For instance, suppose in the preceding circuit we have an input voltage ranging between 15 and 20 volts and use a 1N4733 (5.1V 1W zener diode) in order to generate a stable 5.1 volt power supply. We choose $R = 300$ ohms, for a maximum zener current of 50mA: (20 − 5.1)/300. We can now estimate the output voltage regulation (variation in output voltage), knowing that this particular zener has a specified maximum dynamic impedance of 7.0 ohms at 50mA. The zener current varies from 50mA to 33mA over the input voltage range; this 17mA change in current then produces a voltage change at the output of $\Delta V = R_{dyn}\Delta I$, or 0.12 volt. You will see more of zeners in Sections 2.04 and 6.14.

In real life, a zener will provide better regulation if driven by a current source, which has, by definition, $R_{incr} = \infty$ (same current regardless of voltage). But current sources are more complex, and therefore in practice we often resort to the humble resistor.

Tunnel diodes

Another interesting application of incremental resistance is the *tunnel diode*, sometimes called the Esaki diode. Its *V–I* curve is shown in Figure 1.15. In the region from *A* to *B* it has *negative* incremental resistance. This has a remarkable consequence: A voltage *divider* made with a resistor and

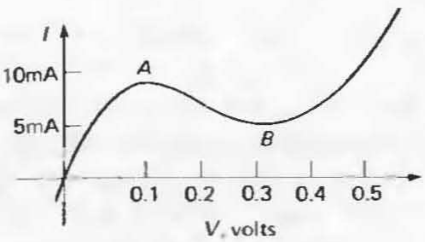


Figure 1.15

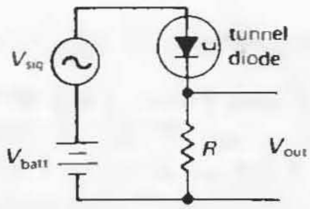


Figure 1.16

a tunnel diode can actually be an *amplifier* (Fig. 1.16). For a wiggly voltage v_{sig}, the voltage divider equation gives us

$$v_{out} = \frac{R}{R + r_t} v_{sig}$$

where r_t is the incremental resistance of the tunnel diode at the operating current, and the lower-case symbol v_{sig} stands for a small-signal variation, which we have been calling ΔV_{sig} up to now (we will adopt this widely used convention from now on). The tunnel diode has $r_{t(incr)} < 0$. That is,

$$\Delta V/\Delta I \quad (\text{or } v/i) \quad < \quad 0$$

from A to B on the characteristic curve. If $r_{t(incr)} \approx R$, the denominator is nearly zero, and the circuit amplifies. V_{batt} provides the steady current, or *bias*, to bring the operating point into the region of negative resistance. (Of course, it is always necessary to have a source of power in any device that amplifies.)

A postmortem on these fascinating devices: When tunnel diodes first appeared, late in the 1950s, they were hailed as the solution to a great variety of circuit problems. Because they were fast, they were supposed to revolutionize computers, for instance. Unfortunately, they are difficult devices to use; this fact, combined with stunning improvements in transistors, has made tunnel diodes almost obsolete.

The subject of negative resistance will come up again later, in connection with active filters. There you will see a circuit called a negative-impedance converter that can produce (among other things) a pure negative resistance (not just incremental). It is made with an operational amplifier and has very useful properties.

SIGNALS

A later section in this chapter will deal with capacitors, devices whose properties depend on the way the voltages and currents in a circuit are *changing*. Our analysis of dc circuits so far (Ohm's law, Thévenin equivalent circuits, etc.) still holds, even if the voltages and currents are changing in time. But for a proper understanding of alternating-current (ac) circuits, it is useful to have in mind certain common types of *signals*, voltages that change in time in a particular way.

1.07 Sinusoidal signals

Sinusoidal signals are the most popular signals around; they're what you get out of the wall plug. If someone says something like "take a 10 microvolt signal at 1 megahertz," he means a sine wave. Mathematically, what you have is a voltage described by

$$V = A \sin 2\pi f t$$

where A is called the amplitude, and f is the frequency in cycles per second, or hertz. A sine wave looks like the wave shown in Figure 1.17. Sometimes it is important to know the value of the signal at some arbitrary time $t = 0$, in which case you may see a *phase* ϕ in the expression:

$$V = A \sin(2\pi f t + \phi)$$

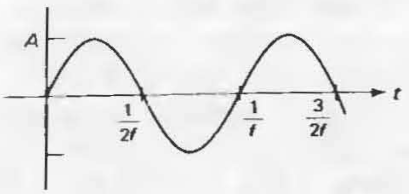


Figure 1.17. Sine wave of amplitude A and frequency f.

The other variation on this simple theme is the use of *angular frequency*, which looks like this:

$V = A\sin\omega t$

Here, ω is the angular frequency in radians per second. Just remember the important relation $\omega = 2\pi f$ and you won't go wrong.

The great merit of sine waves (and the cause of their perennial popularity) is the fact that they are the solutions to certain linear differential equations that happen to describe many phenomena in nature as well as the properties of linear circuits. A linear circuit has the property that its output, when driven by the sum of two input signals, equals the sum of its individual outputs when driven by each input signal in turn; i.e., if $O(A)$ represents the output when driven by signal A, then a circuit is linear if $O(A+B) = O(A) + O(B)$. A linear circuit driven by a sine wave always responds with a sine wave, although in general the phase and amplitude are changed. No other signal can make this statement. It is standard practice, in fact, to describe the behavior of a circuit by its *frequency response*, the way it alters the amplitude of an applied sine wave as a function of frequency. A high-fidelity amplifier, for instance, should be characterized by a "flat" frequency response over the range 20Hz to 20kHz, at least.

The sine-wave frequencies you will usually deal with range from a few hertz to a few megahertz. Lower frequencies, down to 0.0001Hz or lower, can be generated with carefully built circuits, if needed. Higher frequencies, e.g., up to 2000MHz, can be generated, but they require special transmission-line techniques. Above that, you're dealing with microwaves, where conventional wired circuits with lumped circuit elements become impractical, and exotic waveguides or "striplines" are used instead.

1.08 Signal amplitudes and decibels

In addition to its amplitude, there are several other ways to characterize the magnitude of a sine wave or any other signal. You sometimes see it specified by *peak-to-peak amplitude* (pp amplitude), which is just what you would guess, namely, twice the amplitude. The other method is to give the *root-mean-square amplitude* (rms amplitude), which is $V_{\text{rms}} = (1/\sqrt{2})A = 0.707A$ (this is for sine waves only; the ratio of pp to rms will be different for other waveforms). Odd as it may seem, this is the usual method, because rms voltage is what's used to compute power. The voltage across the terminals of a wall socket (in the United States) is 117 volts rms, 60Hz. The *amplitude* is 165 volts (330 volts pp).

Decibels

How do you compare the relative amplitudes of two signals? You could say, for instance, that signal X is twice as large as signal Y. That's fine, and useful for many purposes. But because we often deal with ratios as large as a million, it is easier to use a logarithmic measure, and for this we present the decibel (it's one-tenth as large as something called a bel, which no one ever uses). By definition, the ratio of two signals, in decibels, is

$$\text{dB} = 20\log_{10}\frac{A_2}{A_1}$$

where A_1 and A_2 are the two signal amplitudes. So, for instance, one signal of twice the amplitude of another is +6dB relative to it, since $\log_{10}2 = 0.3010$. A signal 10 times as large is +20dB; a signal one-tenth as large is −20dB. It is also useful to express the ratio of two signals in terms of power levels:

$$\mathrm{dB} = 10\log_{10}\frac{P_2}{P_1}$$

where P_1 and P_2 represent the power in the two signals. As long as the two signals have the same kind of waveform, e.g., sine waves, the two definitions give the same result. When comparing unlike waveforms, e.g., a sine wave versus "noise," the definition in terms of power (or the amplitude definition, with rms amplitudes substituted) must be used.

Although decibels are ordinarily used to specify the ratio of two signals, they are sometimes used as an absolute measure of amplitude. What is happening is that you are assuming some reference signal amplitude and expressing any other amplitude in decibels relative to it. There are several standard amplitudes (which are unstated, but understood) that are used in this way; the most common references are (a) dBV; 1 volt rms; (b) dBm: the voltage corresponding to 1mW into some assumed load impedance, which for radiofrequencies is usually 50 ohms, but for audio is often 600 ohms (the corresponding 0dBm amplitudes, when loaded by those impedances, are then 0.22V rms and 0.78V rms); and (c) the small noise voltage generated by a resistor at room temperature (this surprising fact is discussed in Section 7.11). In addition to these, there are reference amplitudes used for measurements in other fields. For instance, in acoustics, 0dB SPL is a wave whose rms pressure is 0.0002μbar (a bar is 10^6 dynes per square centimeter, approximately 1 atmosphere); in communications, levels can be stated in dBrnC (relative noise reference weighted in frequency by "curve C"). When stating amplitudes this way, it is best to be specific about the 0dB reference amplitude; say something like "an amplitude of 27 decibels relative to 1 volt rms," or abbreviate "27 dB re 1V rms," or define a term like "dBV."

EXERCISE 1.11

Determine the voltage and power ratios for a pair of signals with the following decibel ratios: (a) 3dB, (b) 6dB, (c) 10dB, (d) 20dB.

1.09 Other signals

The ramp is a signal that looks like the signal shown in Figure 1.18. It is simply a voltage rising (or falling) at a constant rate. That can't go on forever, of course, even in science fiction movies. It is sometimes approximated by a finite ramp (Fig. 1.19) or by a periodic ramp, or sawtooth (Fig. 1.20).

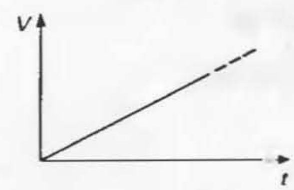


Figure 1.18. Voltage ramp waveform.

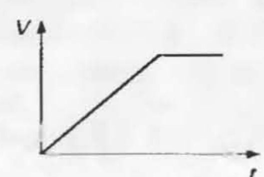


Figure 1.19. Ramp with limit.

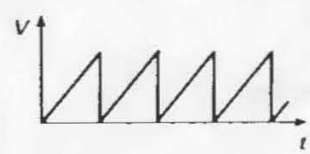


Figure 1.20. Sawtooth wave.

Triangle

The triangle wave is a close cousin of the ramp; it is simply a symmetrical ramp (Fig. 1.21).

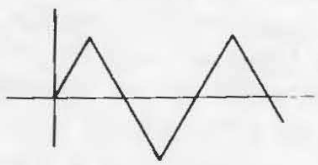


Figure 1.21. Triangle wave.

Figure 1.22. Noise.

Noise

Signals of interest are often mixed with *noise*; this is a catchall phrase that usually applies to random noise of thermal origin. Noise voltages can be specified by their frequency spectrum (power per hertz) or by their amplitude distribution. One of the most common kinds of noise is *band-limited white Gaussian noise*, which means a signal with equal power per hertz in some band of frequencies and a Gaussian (bell-shaped) distribution of amplitudes if large numbers of instantaneous measurements of its amplitude are made. This kind of noise is generated by a resistor (Johnson noise), and it plagues sensitive measurements of all kinds. On an oscilloscope it appears as shown in Figure 1.22. We will study noise and low-noise techniques in some detail in Chapter 7. Sections 9.32–9.37 deal with noise-generation techniques.

Square waves

A square wave is a signal that varies in time as shown in Figure 1.23. Like the sine wave, it is characterized by amplitude and frequency. A linear circuit driven by a square wave rarely responds with a square wave. For a square wave, the rms amplitude equals the amplitude.

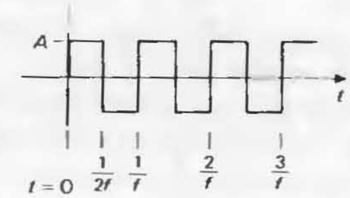


Figure 1.23. Square wave.

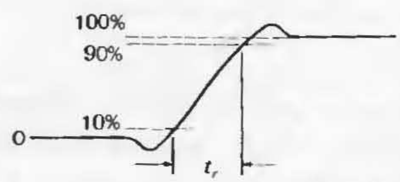


Figure 1.24. Rise time of a step waveform.

The edges of a square wave are not perfectly square; in typical electronic circuits the rise time t_r ranges from a few nanoseconds to a few microseconds. Figure 1.24 shows the sort of thing usually seen. The rise time is defined as the time required for the signal to go from 10% to 90% of its total transition.

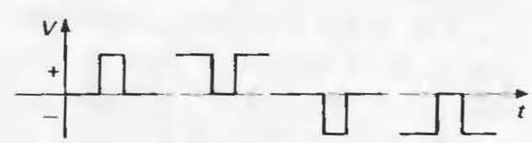


Figure 1.25. Positive- and negative-going pulses of both polarities.

Pulses

A pulse is a signal that looks as shown in Figure 1.25. It is defined by amplitude and pulse width. You can generate a train of periodic (equally spaced) pulses, in which case you can talk about the frequency, or pulse repetition rate, and the "duty cycle," the ratio of pulse width to repetition period (duty cycle ranges from zero to 100%). Pulses can have positive or negative polarity; in addition, they can be "positive-going" or "negative-going." For instance, the second pulse in Figure 1.25

is a negative-going pulse of positive polarity.

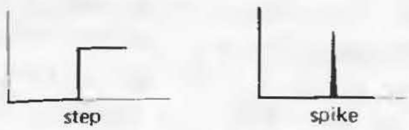


Figure 1.26

Steps and spikes

Steps and spikes are signals that are talked about a lot but are not often used. They provide a nice way of describing what happens in a circuit. If you could draw them, they would look something like the example in Figure 1.26. The step function is part of a square wave; the spike is simply a jump of vanishingly short duration.

1.10 Logic levels

Pulses and square waves are used extensively in digital electronics, where predefined voltage levels represent one of two possible states present at any point in the circuit. These states are called simply HIGH and LOW and correspond to the 0 (false) and 1 (true) states of Boolean logic (the algebra that describes such two-state systems).

Precise voltages are not necessary in digital electronics. You need only to distinguish which of the two possible states is present. Each digital logic family therefore specifies legal HIGH and LOW states. For example, the "74HC" digital logic family runs from a single +5 volt supply, with output levels that are typically 0 volts (LOW) and 5 volts (HIGH), and an input decision threshold of 2.5 volts. Actual outputs can be as much as a volt from ground or +5 volts without malfunction, however. We'll have much more to say about logic levels in Chapters 8 and 9.

1.11 Signal sources

Often the source of a signal is some part of the circuit you are working on. But for test purposes a flexible signal source is invaluable. They come in three flavors: signal generators, pulse generators, and function generators.

Signal generators

Signal generators are sine-wave oscillators, usually equipped to give a wide range of frequency coverage (50kHz to 50MHz is typical), with provision for precise control of amplitude (using a resistive divider network called an *attenuator*). Some units let you *modulate* the output (see Chapter 13). A variation on this theme is the *sweep generator*, a signal generator that can sweep its output frequency repeatedly over some range. These are handy for testing circuits whose properties vary with frequency in a particular way, e.g., "tuned circuits" or filters. Nowadays these devices, as well as many test instruments, are available in configurations that allow you to program the frequency, amplitude, etc., from a computer or other digital instrument.

A variation on the signal generator is the *frequency synthesizer*, a device that generates sine waves whose frequencies can be set precisely. The frequency is set digitally, often to eight significant figures or more, and is internally synthesized from a precise standard (a quartz-crystal oscillator) by digital methods we will discuss later (Sections 9.27–9.31). If your requirement is for no-nonsense accurate frequency generation, you can't beat a synthesizer.

Pulse generators

Pulse generators only make pulses, but what pulses! Pulse width, repetition rate, amplitude, polarity, rise time, etc., may all be adjustable. In addition, many units allow you to generate pulse pairs, with settable spacing and repetition rate, or even coded pulse trains. Most modern pulse generators are provided with logic-level outputs for easy connection to digital circuitry. Like signal generators, these come in the programmable variety.

Function generators

In many ways function generators are the most flexible signal sources of all. You can make sine, triangle, and square waves over an enormous frequency range (0.01 Hz to 10 MHz is typical), with control of amplitude and dc offset (a constant dc voltage added to the signal). Many of them have provision for frequency sweeping, often in several modes (linear or logarithmic frequency variation versus time). They are available with pulse outputs (although not with the flexibility you get with a pulse generator), and some of them have provision for modulation.

Like the other signal sources, function generators come in programmable versions and versions with digital readout of frequency (and sometimes amplitude). The most recent addition to the function-generator family is the synthesized function generator, a device that combines all the flexibility of a function generator with the stability and accuracy of a frequency synthesizer. An example is the HP 8116A, with sine, square, and triangle waves (as well as pulses, ramps, haversines, etc.) from 0.001Hz to 50MHz. Frequency and amplitude (10mV to 16V pp) are programmable, as are linear and logarithmic frequency sweeps. This unit also provides trigger, gate, burst, FM, AM, pulse-width modulation, voltage-controlled frequency, and single cycles. For general use, if you can have only one signal source, the function generator is for you.

CAPACITORS AND AC CIRCUITS

Once we enter the world of changing voltages and currents, or signals, we encounter two very interesting circuit elements that are useless in dc circuits: capacitors and inductors. As you will see, these humble devices, combined with resistors, complete the triad of passive linear circuit elements that form the basis of nearly all circuitry. Capacitors, in particular, are essential in nearly every circuit application. They are used for waveform generation, filtering, and blocking and bypass applications. They are used in integrators and differentiators. In combination with inductors, they make possible sharp filters for separating desired signals from background. You will see some of these applications as we continue with this chapter, and there will be numerous interesting examples in later chapters.

Let's proceed, then, to look at capacitors in detail. Portions of the treatment that follows are necessarily mathematical in nature; the reader with little mathematical preparation may find Appendix B helpful. In any case, an understanding of the details is less important in the long run than an understanding of the results.

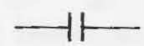


Figure 1.27. Capacitor.

1.12 Capacitors

A capacitor (Fig. 1.27) (the old-fashioned name was *condenser*) is a device that has two wires sticking out of it and has the property

$$Q = CV$$

A capacitor of C farads with V volts across its terminals has Q coulombs of stored charge on one plate, and $-Q$ on the other.

To a first approximation, capacitors are devices that might be considered simply frequency-dependent resistors. They allow you to make frequency-dependent voltage dividers, for instance. For some applications (bypass, coupling) this is almost all you need to know, but for other applications (filtering, energy storage, resonant circuits) a deeper understanding is needed. For example, capacitors cannot dissipate power, even though current can flow through them, because the voltage and current are 90° out of phase.

Taking the derivative of the defining equation above (see Appendix B), you get

$$I = C\frac{dV}{dt}$$

So a capacitor is more complicated than a resistor; the current is not simply proportional to the voltage, but rather to the rate of change of voltage. If you change the voltage across a farad by 1 volt per second, you are supplying an amp. Conversely, if you supply an amp, its voltage changes by 1 volt per second. A farad is very large, and you usually deal in microfarads (μF) or picofarads (pF). (To make matters confusing to the uninitiated, the units are often omitted on capacitor values specified in schematic diagrams. You have to figure it out from the context.) For instance, if you supply a current of 1mA to 1μF, the voltage will rise at 1000 volts per second. A 10ms pulse of this current will increase the voltage across the capacitor by 10 volts (Fig. 1.28).

Capacitors come in an amazing variety of shapes and sizes; with time, you will come to recognize their more common incarnations. The basic construction is simply two conductors near each other (but not touching); in fact, the simplest capacitors are just that. For greater capacitance, you need more area and closer spacing; the usual approach is to plate some conductor onto a thin insulating material (called a dielectric), for instance, aluminized Mylar film rolled up into a small cylindrical configuration. Other popular types are thin ceramic wafers (disc ceramics), metal foils with oxide insulators (electrolytics), and metallized mica. Each of these types has unique properties; for a brief rundown, see the box on capacitors. In general, ceramic and Mylar types are used for most noncritical circuit applications; tantalum capacitors are used where greater capacitance is needed, and electrolytics are used for power-supply filtering.

Capacitors in parallel and series

The capacitance of several capacitors in parallel is the sum of their individual capacitances. This is easy to see: Put voltage V across the parallel combination; then

$$\begin{aligned} C_{\text{total}}V &= Q_{\text{total}} = Q_1 + Q_2 + Q_3 + \ldots \\ &= C_1V + C_2V + C_3V + \ldots \\ &+ (C_1 + C_2 + C_3 + \ldots)V \end{aligned}$$

or

$$C_{\text{total}} = C_1 + C_2 + C_3 + \ldots$$

For capacitors in series, the formula is like that for resistors in parallel:

$$C_{\text{total}} = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots}$$

or (two capacitors only)

$$C_{\text{total}} = \frac{C_1C_2}{C_1 + C_2}$$

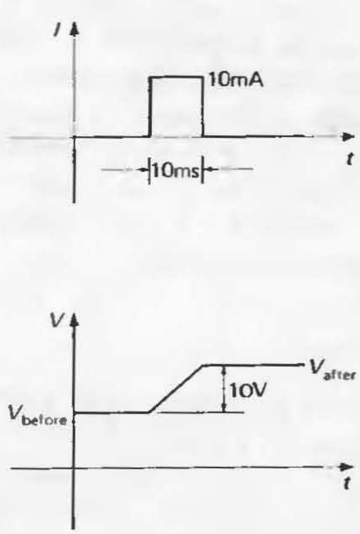


Figure 1.28. The voltage across a capacitor changes when a current flows through it.

CAPACITORS

There is wide variety among the capacitor types available. This is a quickie guide to point out their major advantages and disadvantages. Our judgments should be considered somewhat subjective:

Type	Capacitance range	Maximum voltage	Accuracy	Temperature stability	Leakage	Comments
Mica	1pF–0.01μF	100–600	Good		Good	Excellent; good at RF
Tubular ceramic	0.5pF–100pF	100–600		Selectable		Several tempcos (including zero)
Ceramic	10pF–1μF	50–30,000	Poor	Poor	Moderate	Small, inexpensive, very popular
Polyester (Mylar)	0.001μF–50μF	50–600	Good	Poor	Good	Inexpensive, good, popular
Polystyrene	10pF–2.7μF	100–600	Excellent	Good	Excellent	High quality, large; signal filters
Polycarbonate	100pF–30μF	50–800	Excellent	Excellent	Good	High quality, small
Polypropylene	100pF–50μF	100–800	Excellent	Good	Excellent	High quality, low dielectric absorption
Teflon	1000pF–2μF	50–200	Excellent	Best	Best	High quality, lowest dielectric absorption
Glass	10pF–1000pF	100–600	Good		Excellent	Long-term stability
Porcelain	100pF–0.1μF	50–400	Good	Good	Good	Good long-term stability
Tantalum	0.1μF–500μF	6–100	Poor	Poor		High capacitance; polarized, small; low inductance
Electrolytic	0.1μF–1.6F	3–600	Terrible	Ghastly	Awful	Power-supply filters; polarized; short life
Double layer	0.1F–10F	1.5–6	Poor	Poor	Good	Memory backup; high series resistance
Oil	0.1μF–20μF	200–10,000			Good	High-voltage filters; large, long life
Vacuum	1pF–5000pF	2000–36,000			Excellent	Transmitters

EXERCISE 1.12

Derive the formula for the capacitance of two capacitors in series. Hint: Because there is no external connection to the point where the two capacitors are connected together, they must have equal stored charges.

The current that flows in a capacitor during charging ($I = CdV/dt$) has some unusual features. Unlike resistive current, it's not proportional to voltage, but rather to the rate of change (the "time derivative") of voltage. Furthermore, unlike the situation in a resistor, the power (V times I) associated with capacitive current is not turned into heat, but is stored as energy in the capacitor's internal electric field. You get all that energy back when you discharge the capacitor. We'll see another way to look at these curious properties when we talk about *reactance*, beginning in Section 1.18.

1.13 *RC* circuits: *V* and *I* versus time

When dealing with ac circuits (or, in general, any circuits that have changing voltages and currents), there are two possible approaches. You can talk about V and I versus time, or you can talk about amplitude versus signal frequency. Both approaches have their merits, and you find yourself switching back and forth according to which description is most convenient in each situation. We will begin our study of ac circuits in the time domain. Beginning with Section 1.18, we will tackle the frequency domain.

What are some of the features of circuits with capacitors? To answer this question, let's begin with the simple RC circuit (Fig. 1.29). Application of the capacitor rules gives

$$C\frac{dV}{dt} = I = -\frac{V}{R}$$

This is a differential equation, and its solution is

$$V = Ae^{-t/RC}$$

So a charged capacitor placed across a resistor will discharge as in Figure 1.30.

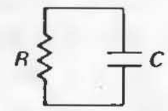


Figure 1.29

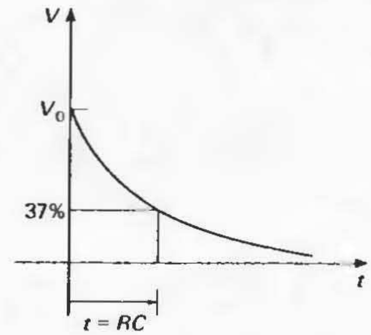


Figure 1.30. RC discharge waveform.

Time constant

The product RC is called the *time constant* of the circuit. For R in ohms and C in farads, the product RC is in seconds. A microfarad across 1.0k has a time constant of 1ms; if the capacitor is initially charged to 1.0 volt, the initial current is 1.0mA.

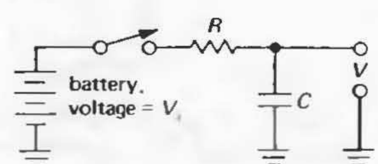


Figure 1.31

Figure 1.31 shows a slightly different circuit. At time $t = 0$, someone connects the battery. The equation for the circuit is then

$$I = C\frac{dV}{dt} = \frac{V_i - V}{R}$$

with the solution

$$V = V_i + Ae^{-t/RC}$$

(Please don't worry if you can't follow the mathematics. What we are doing is getting some important results, which you should remember. Later we will use the results often, with no further need for the mathematics used to derive them.) The constant A is determined by initial conditions (Fig. 1.32): $V = 0$ at $t = 0$; therefore, $A = -V_i$, and

$$V = V_i(1 - e^{-t/RC})$$

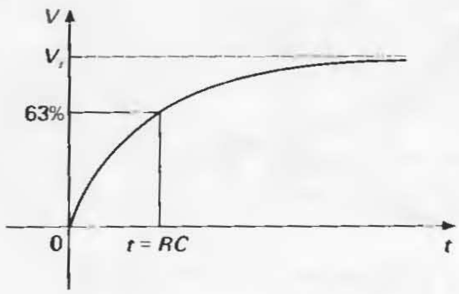


Figure 1.32

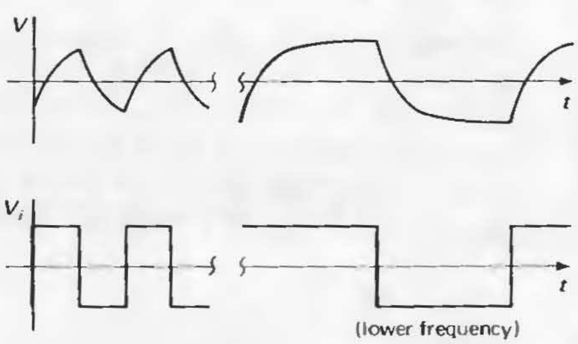


Figure 1.33. Output (top waveform) across a capacitor, when driven by square waves through a resistor.

Decay to equilibrium

Eventually (when $t \gg RC$), V reaches V_i. (Presenting the "$5RC$ rule of thumb": a capacitor charges or decays to within 1% of its final value in 5 time constants.) If we then change V_0 to some other value (say, 0), V will decay toward that new value with an exponential $e^{-t/RC}$. For example, a square-wave input for V_0 will produce the output shown in Figure 1.33.

EXERCISE 1.13

Show that the rise time (the time required to go from 10% to 90% of its final value) of this signal is $2.2RC$.

You might ask the obvious next question: What about $V(t)$ for arbitrary $V_i(t)$? The solution involves an inhomogeneous differential equation and can be solved by standard methods (which are, however, beyond the scope of this book). You would find

$$V(t) = \frac{1}{RC}\int_{-\infty}^{t} V_i(\tau)e^{-(t-\tau)/RC}d\tau$$

That is, the RC circuit averages past history at the input with a weighting factor $e^{-\Delta t/RC}$

In practice, you seldom ask this question. Instead, you deal in the *frequency domain* and ask how much of each frequency component present in the input gets through. We will get to this important topic soon (Section 1.18). Before we do, though, there are a few other interesting circuits we can analyze simply with this time-domain approach.

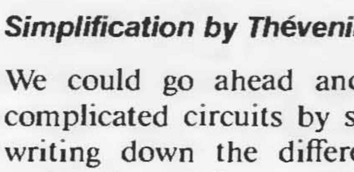


Figure 1.34

Simplification by Thévenin equivalents

We could go ahead and analyze more complicated circuits by similar methods, writing down the differential equations and trying to find solutions. For most purposes it simply isn't worth it. This is as complicated an RC circuit as we will need. Many other circuits can be reduced to it (e.g., Fig. 1.34). By just using the Thévenin equivalent of the voltage divider formed by R_1 and R_2, you can find the output

$V(t)$ produced by a step input for V_0.

EXERCISE 1.14

$R_1 = R_2 = 10\text{k}$, and $C = 0.1\mu\text{F}$ in the circuit shown in Figure 1.34. Find $V(t)$ and sketch it.

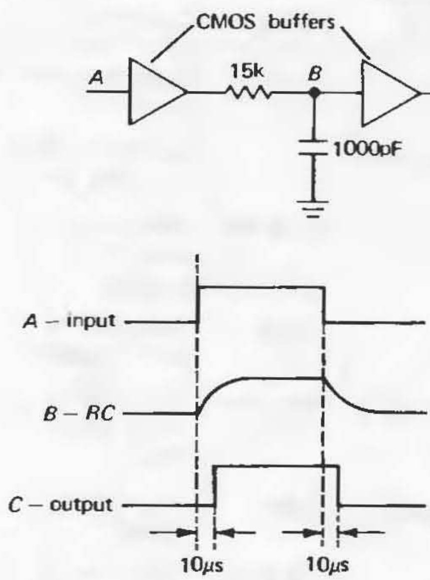


Figure 1.35. Producing a delayed digital waveform with the help of an *RC*.

Example: time-delay circuit

We have already mentioned logic levels, the voltages that digital circuits live on. Figure 1.35 shows an application of capacitors to produce a delayed pulse. The triangular symbols are "CMOS buffers." They give a HIGH output if the input is HIGH (more than one-half the dc power-supply voltage used to power them), and vice versa. The first buffer provides a replica of the input signal, but with low source resistance, and prevents input loading by the *RC* (recall our earlier discussion of circuit loading in Section 1.05). The *RC* output has the characteristic decays and causes the output buffer to switch 10μs after the input transitions (an *RC* reaches 50% output in $0.7RC$). In an actual application you would have to consider the effect of the buffer input threshold deviating from one-half the supply voltage, which would alter the delay and change the output pulse width. Such a circuit is sometimes used to delay a pulse so that something else can happen first. In designing circuits you try not to rely on tricks like this, but they're occasionally handy.

1.14 Differentiators

Look at the circuit in Figure 1.36. The voltage across C is $V_{\text{in}} - V$, so

$$I = C\frac{d}{dt}(V_{\text{in}} - V) = \frac{V}{R}$$

If we choose R and C small enough so that $dV/dt \ll dV_{\text{in}}/dt$, then

$$C\frac{dV_{\text{in}}}{dt} \approx \frac{V}{R}$$

or

$$V(t) = RC\frac{d}{dt}V_{\text{in}}(t)$$

That is, we get an output proportional to the rate of change of the input waveform.

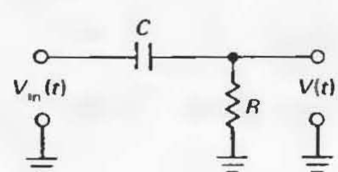


Figure 1.36

To keep $dV/dt \ll dV_{\text{in}}/dt$, we make the product RC small, taking care not to "load" the input by making R too small (at the transition the change in voltage across the capacitor is zero, so R is the load seen by the input). We will have a better criterion for this when we look at things in the frequency domain. If you drive this circuit with a square wave, the output will be as shown in Figure 1.37.

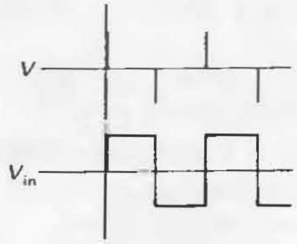


Figure 1.37. Output waveform (top) from differentiator driven by a square wave.

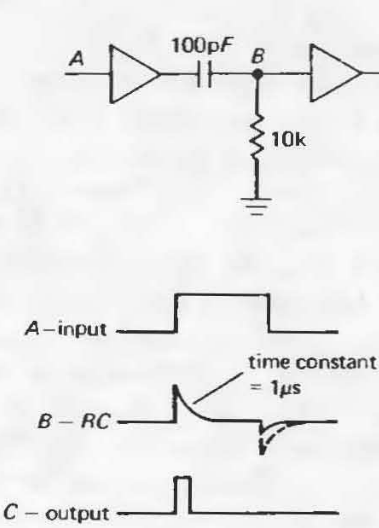


Figure 1.38. Leading-edge detector.

Differentiators are handy for detecting *leading edges* and *trailing edges* in pulse signals, and in digital circuitry you sometimes see things like those depicted in Figure 1.38. The RC differentiator generates spikes at the transitions of the input signal, and the output buffer converts the spikes to short square-topped pulses. In practice, the negative spike will be small because of a diode (a handy device discussed in Section 1.25) built into the buffer.

Unintentional capacitive coupling

Differentiators sometimes crop up unexpectedly, in situations where they're not welcome. You may see signals like those shown in Figure 1.39. The first case is caused by a square wave somewhere in the circuit coupling capacitively to the signal line you're looking at; that might indicate a missing resistor termination on your signal line. If not, you must either reduce the source resistance of the signal line or find a way to reduce capacitive coupling from the offending square wave. The second case is typical of what you might see when you look at a square wave, but have a broken connection somewhere, usually at the scope probe. The very small capacitance of the broken connection combines with the scope input resistance to form a differentiator. *Knowing that you've got a differentiated "something" can help you find the trouble and eliminate it.*

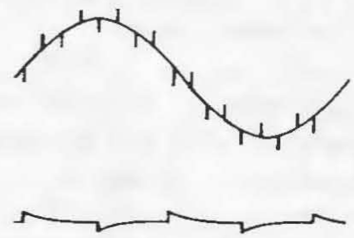


Figure 1.39

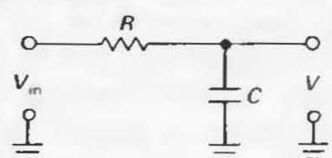


Figure 1.40

1.15 Integrators

Take a look at the circuit in Figure 1.40. The voltage across R is $V_{\text{in}} - V$, so

$$I = C\frac{dV}{dt} = \frac{V_{\text{in}} - V}{R}$$

If we manage to keep $V \ll V_{\text{in}}$, by keeping the product RC large, then

$$C\frac{dV}{dt} \approx \frac{V_{\text{in}}}{R}$$

or

$$V(t) = \frac{1}{RC}\int^{t} V_{\text{in}}(t)\,dt + \text{constant}$$

We have a circuit that performs the integral over time of an input signal! You can see how the approximation works for a square-wave input: $V(t)$ is then the exponential charging curve we saw earlier (Fig. 1.41). The first part of the exponential is a ramp, the integral of a constant; as we increase the time constant RC, we pick off a smaller part of the exponential, i.e., a better approximation to a perfect ramp.

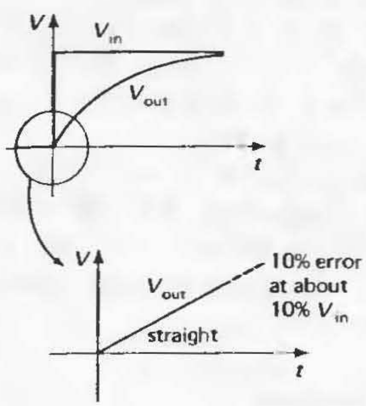


Figure 1.41

Note that the condition $V \ll V_{in}$ is just the same as saying that I is proportional to V_{in}. If we had as input a *current* $I(t)$, rather than a voltage, we would have an exact integrator. A large voltage across a large resistance approximates a current source and, in fact, is frequently used as one.

Later, when we get to operational amplifiers and feedback, we will be able to build integrators without the restriction $V_{out} \ll V_{in}$. They will work over large frequency and voltage ranges with negligible error.

The integrator is used extensively in analog computation. It is a useful subcircuit that finds application in control systems, feedback, analog/digital conversion, and waveform generation.

Ramp generators

At this point it is easy to understand how a ramp generator works. This nice circuit is extremely useful, for example in timing circuits, waveform and function generators, oscilloscope sweep circuits, and analog/digital conversion circuitry. The circuit uses a constant current to charge a capacitor (Fig. 1.42). From the capacitor equation $I = C(dV/dt)$, you get $V(t) = (I/C)t$. The output waveform is as shown in Figure 1.43. The ramp stops when the current source "runs out of voltage," i.e., reaches the limit of its compliance. The curve for a simple RC, with the resistor tied to a voltage source equal to the compliance of the current source, and with R chosen so that the current at zero output voltage is the same as that of the current source, is also drawn for comparison. (Real current sources generally have output compliances limited by the power-supply voltages used in making them, so the comparison is realistic.) In the next chapter, which deals with transistors, we will design some current sources, with some refinements to follow in the chapters on operational amplifiers (op-amps) and field-effect transistors (FETs). Exciting things to look forward to!

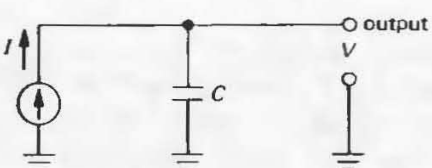


Figure 1.42. A constant current source charging a capacitor generates a ramp voltage waveform.

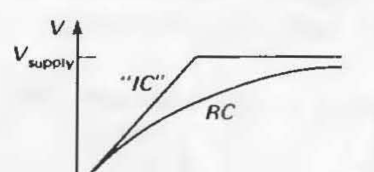


Figure 1.43

EXERCISE 1.15
A current of 1mA charges a 1μF capacitor. How long does it take the ramp to reach 10 volts?

INDUCTORS AND TRANSFORMERS

1.16 Inductors

If you understand capacitors, you won't have any trouble with inductors (Fig. 1.44). They're closely related to capacitors; the rate of current change in an inductor depends on the voltage applied across it, whereas the rate of voltage change in a capacitor depends on the current through it. The defining equation for an inductor is

$$V = L\frac{dI}{dt}$$

where L is called the *inductance* and is measured in henrys (or mH, μH, etc.). Putting a voltage across an inductor causes the current to rise as a ramp (for a capacitor, supplying a constant current causes the voltage to rise as a ramp); 1 volt across 1 henry produces a current that increases at 1 amp per second.

Figure 1.44. Inductor.

As with capacitive current, inductive current is not simply proportional to voltage. Furthermore, unlike the situation in a resistor, the power associated with inductive current (V times I) is not turned into heat, but is stored as energy in the inductor's magnetic field. You get all that energy back when you interrupt the inductor's current.

The symbol for an inductor looks like a coil of wire; that's because, in its simplest form, that's all it is. Variations include coils wound on various core materials, the most popular being iron (or iron alloys, laminations, or powder) and ferrite, a black, nonconductive, brittle magnetic material. These are all ploys to multiply the inductance of a given coil by the "permeability" of the core material. The core may be in the shape of a rod, a toroid (doughnut), or even more bizarre shapes, such as a "pot core" (which has to be seen to be understood; the best description we can think of is a doughnut mold split horizontally in half, if doughnuts were made in molds).

Inductors find heavy use in radio-frequency (RF) circuits, serving as RF "chokes" and as parts of tuned circuits (see Chapter 13). A pair of closely coupled inductors forms the interesting object known as a transformer. We will talk briefly about them in the next section.

An inductor is, in a real sense, the opposite of a capacitor. You will see how that works out in the next few sections of this chapter, which deal with the important subject of *impedance*.

1.17 Transformers

A transformer is a device consisting of two closely coupled coils (called primary and secondary). An ac voltage applied to the primary appears across the secondary, with a voltage multiplication proportional to the turns ratio of the transformer and a current multiplication inversely proportional to the turns ratio. Power is conserved. Figure 1.45 shows the circuit symbol for a laminated-core transformer (the kind used for 60Hz ac power conversion).

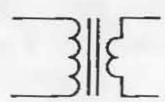


Figure 1.45. Transformer.

Transformers are quite efficient (output power is very nearly equal to input power); thus, a step-up transformer gives higher voltage at lower current. Jumping ahead for a moment, a transformer of turns ratio n increases the impedance by n^2. There is very little primary current if the secondary is unloaded.

Transformers serve two important functions in electronic instruments: They change the ac line voltage to a useful (usually lower) value that can be used by the circuit, and they "isolate" the electronic device from actual connection to the power line, because the windings of a transformer are electrically insulated from each other. *Power transformers* (meant for use from the 110V power line) come in an enormous variety of secondary voltages and currents: outputs as low as 1 volt or so up to several thousand volts, current ratings from a few milliamps to hundreds of amps. Typical transformers for use in electronic instruments might have secondary voltages from 10 to 50 volts, with current ratings of 0.1 to 5 amps or so.

Transformers for use at audiofrequencies and radiofrequencies are also available. At radiofrequencies you sometimes use tuned transformers, if only a narrow range of frequencies is present. There is also an interesting class of transmission-line transformer that we will discuss briefly in Section 13.10. In general, transformers for use at high frequencies must use special core materials or construction to minimize core losses, whereas low-frequency transformers (e.g., power transformers) are burdened instead by large and heavy cores. The two kinds of transformers are in general not interchangeable.

IMPEDANCE AND REACTANCE

Warning: This section is somewhat mathematical; you may wish to skip over the mathematics, but be sure to pay attention to the results and graphs.

Circuits with capacitors and inductors are more complicated than the resistive circuits we talked about earlier, in that their behavior depends on frequency: A "voltage divider" containing a capacitor or inductor will have a frequency-dependent division ratio. In addition, circuits containing these components (known collectively as *reactive* components) "corrupt" input waveforms such as square waves, as we just saw.

However, both capacitors and inductors are *linear* devices, meaning that the amplitude of the output waveform, whatever its shape, increases exactly in proportion to the input waveform's amplitude. This linearity has many consequences, the most important of which is probably the following: *The output of a linear circuit, driven with a sine wave at some frequency* f, *is itself a sine wave at the same frequency (with, at most, changed amplitude and phase).*

Because of this remarkable property of circuits containing resistors, capacitors, and inductors (and, later, linear amplifiers), it is particularly convenient to analyze any such circuit by asking how the output voltage (amplitude and phase) depends on the input voltage, *for sine-wave input at a single frequency*, even though this may not be the intended use. A graph of the resulting *frequency response*, in which the ratio of output to input is plotted for each sine-wave frequency, is useful for thinking about many kinds of waveforms. As an example, a certain "boom-box" loudspeaker might have the frequency response shown in Figure 1.46, where the "output" in this case is of course sound pressure, not voltage. It is desirable for a speaker to have a "flat" response, meaning that the graph of sound pressure versus frequency is constant over the band of audible frequencies. In this case the speaker's deficiencies can be corrected by introducing a passive filter with the inverse response (as shown) into the amplifiers of the radio.

As we will see, it is possible to generalize Ohm's law, replacing the word "resistance" with "impedance," in order to describe any circuit containing these linear passive devices (resistors, capacitors, and inductors). You could think of the subject of impedance and reactance as Ohm's law for circuits that include capacitors and inductors. Some important terminology: Impedance is the "generalized resistance"; inductors

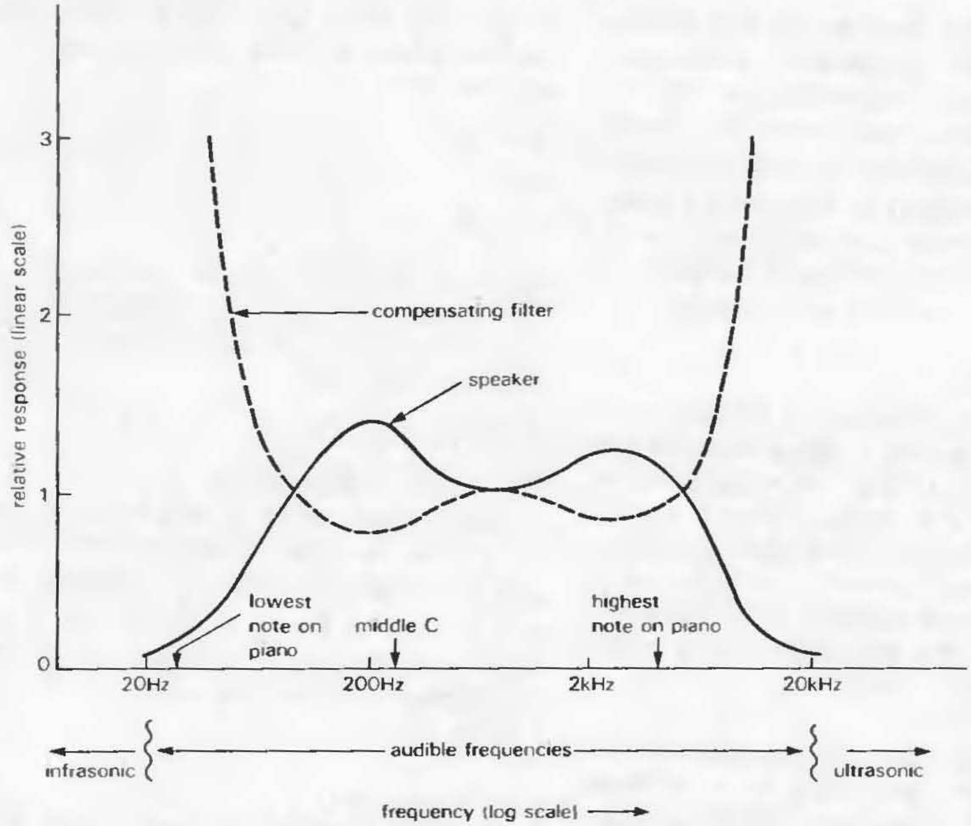


Figure 1.46. Example of frequency analysis: "boom box" loudspeaker equalization.

and capacitors have *reactance* (they are "reactive"); resistors have *resistance* (they are "resistive"). In other words, impedance = resistance + reactance (more about this later). However, you'll see statements like "the impedance of the capacitor at this frequency is ... " The reason you don't have to use the word "reactance" in such a case is that impedance covers everything. In fact, you frequently use the word "impedance" even when you know it's a resistance you're talking about; you say "the source impedance" or "the output impedance" when you mean the Thévenin equivalent resistance of some source. The same holds for "input impedance."

In all that follows, we will be talking about circuits driven by sine waves at a single frequency. Analysis of circuits driven by complicated waveforms is more elaborate, involving the methods we used earlier (differential equations) or decomposition of the waveform into sine waves (Fourier analysis). Fortunately, these methods are seldom necessary.

1.18 Frequency analysis of reactive circuits

Let's start by looking at a capacitor driven by a sine-wave voltage source (Fig. 1.47). The current is

$$I(t) = C\frac{dV}{dt} = C\omega V_0 \cos \omega t$$

i.e., a current of amplitude I, with the phase leading the input voltage by 90°. If we consider amplitudes only, and disregard phases, the current is

$$I = \frac{V}{1/\omega C}$$

(Recall that $\omega = 2\pi f$.) It behaves like a frequency-dependent resistance $R = 1/\omega C$, but in addition the current is 90° out of phase with the voltage (Fig. 1.48).

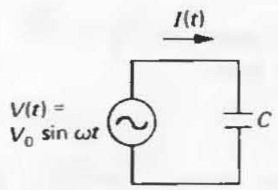


Figure 1.47

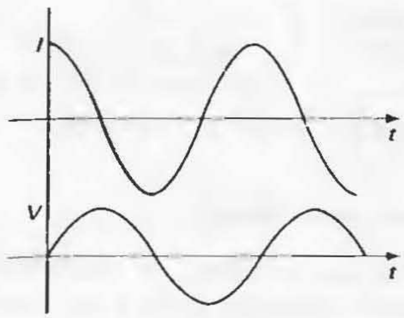


Figure 1.48

For example, a 1μF capacitor put across the 110 volt (rms) 60Hz power line draws a current of rms amplitude

$$I = \frac{110}{1/(2\pi \times 60 \times 10^{-6})} = 41.5\text{mA (rms)}$$

Note: At this point it is necessary to get into some complex algebra; you may wish to skip over the math in some of the following sections, taking note of the results as we derive them. A knowledge of the detailed mathematics is not necessary in order to understand the remainder of the book. Very little mathematics will be used in later chapters. The section ahead is easily the most difficult for the reader with little mathematical preparation. Don't be discouraged!

Voltages and currents as complex numbers

As you have just seen, there can be phase shifts between the voltage and current in an ac circuit being driven by a sine wave at some frequency. Nevertheless, as long as the circuit contains only *linear* elements (resistors, capacitors, inductors), the magnitudes of the currents everywhere in the circuit are still proportional to the magnitude of the driving voltage, so we might hope to find some generalization of voltage, current, and resistance in order to rescue Ohm's law. Obviously a single number won't suffice to specify the current, say, at some point in the circuit, because we must somehow have information about both the magnitude and phase shift.

Although we can imagine specifying the magnitudes and phase shifts of voltages and currents at any point in the circuit by writing them out explicitly, e.g., $V(t) = 23.7\sin(377t + 0.38)$, it turns out that our requirements can be met more simply by using the algebra of complex numbers to *represent* voltages and currents. Then we can simply add or subtract the complex number representations, rather than laboriously having to add or subtract the actual sinusoidal functions of time themselves. Because the actual voltages and currents are real quantities that vary with time, we must develop a rule for converting from actual quantities to their representations, and vice versa. Recalling once again that we are talking about a single sine-wave frequency, ω, we agree to use the following rules:

1. Voltages and currents are *represented* by the complex quantities $\mathbf{V}$ and $\mathbf{I}$. The voltage $V_0\cos(\omega t + \phi)$ is to be represented by the complex number $V_0e^{j\phi}$. Recall that $e^{j\theta} = \cos\theta + j\sin\theta$, where $j = \sqrt{-1}$.

2. *Actual* voltages and currents are obtained by multiplying their complex number representations by $e^{j\omega t}$ and then taking the real part: $V(t) = \mathcal{R}e(\mathbf{V}e^{j\omega t})$, $I(t) = \mathcal{R}e(\mathbf{I}e^{j\omega t})$

In other words,

circuit voltage versus time		complex number representation
$V_0\cos(\omega t + \phi)$	$\rightleftarrows$ multiply by $e^{j\omega t}$ and take real part	$V_0e^{j\phi} = a + jb$

(In electronics, the symbol j is used instead of i in the exponential in order to avoid confusion with the symbol i meaning current.) Thus, in the general case the actual voltages and currents are given by

$$\begin{aligned} V(t) &= \mathcal{R}e(\mathbf{V}e^{j\omega t}) \\ &= \mathcal{R}e(\mathbf{V})\cos\omega t - \mathcal{I}m(\mathbf{V})\sin\omega t \\ I(t) &= \mathcal{R}e(\mathbf{I}e^{j\omega t}) \\ &= \mathcal{R}e(\mathbf{I})\cos\omega t - \mathcal{I}m(\mathbf{I})\sin\omega t \end{aligned}$$

For example, a voltage whose complex representation is

$$\mathbf{V} = 5j$$

corresponds to a (real) voltage versus time of

$$\begin{aligned} V(t) &= \mathcal{R}e[5j\cos\omega t + 5j(j)\sin\omega t] \\ &= -5\sin\omega t \text{ volts} \end{aligned}$$

Reactance of capacitors and inductors

With this convention we can apply complex Ohm's law to circuits containing capacitors and inductors, just as for resistors, once we know the reactance of a capacitor or inductor. Let's find out what these are. We have

$$V(t) = \mathcal{R}e(V_0 e^{j\omega t})$$

For a capacitor, using $I = C(dV/dt)$, we obtain

$$\begin{aligned} I(t) &= -V_0 C\omega \sin\omega t = \mathcal{R}e\left(\frac{V_0 e^{j\omega t}}{-j/\omega C}\right) \\ &= \mathcal{R}e\left(\frac{V_0 e^{j\omega t}}{X_C}\right) \end{aligned}$$

i.e., for a capacitor

$$X_C = -j/\omega C$$

X_C is the reactance of a capacitor at frequency ω. As an example a 1μF capacitor has a reactance of $-2653j$ ohms at 60Hz and a reactance of $-0.16j$ ohms at 1MHz. Its reactance at dc is infinite.

If we did a similar analysis for an inductor, we would find

$$X_L = j\omega L$$

A circuit containing only capacitors and inductors always has a purely imaginary impedance, meaning that the voltage and current are always 90° out of phase – it is purely reactive. When the circuit contains resistors, there is also a real part to the impedance. The term "reactance" in that case means the imaginary part only.

Ohm's law generalized

With these conventions for representing voltages and currents, Ohm's law takes a simple form. It reads simply

$$\mathbf{I} = \mathbf{V}/\mathbf{Z}$$
$$\mathbf{V} = \mathbf{IZ}$$

where the voltage represented by $\mathbf{V}$ is applied across a circuit of impedance $\mathbf{Z}$, giving a current represented by $\mathbf{I}$. The complex impedance of devices in series or parallel obeys the same rules as resistance:

$$\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3 + \cdots \quad \text{(series)}$$
$$\mathbf{Z} = \frac{1}{\frac{1}{\mathbf{Z}_1} + \frac{1}{\mathbf{Z}_2} + \frac{1}{\mathbf{Z}_3} + \cdots} \quad \text{(parallel)}$$

Finally, for completeness we summarize here the formulas for the impedance of resistors, capacitors, and inductors:

$$\mathbf{Z}_R = R \quad \text{(resistor)}$$
$$\mathbf{Z}_C = -j/\omega C = 1/j\omega C \quad \text{(capacitor)}$$
$$\mathbf{Z}_L = j\omega L \quad \text{(inductor)}$$

With these rules we can analyze many ac circuits by the same general methods we used in handling dc circuits, i.e., application of the series and parallel formulas and Ohm's law. Our results for circuits such as voltage dividers will look nearly the same as before. For multiply connected networks we may have to use Kirchhoff's laws, just as with dc circuits, in this case using the complex representations for V and I: The sum of the (complex) voltage drops around a closed loop is zero, and the sum of the (complex) currents into a point is zero. The latter rule implies, as with dc circuits, that the (complex) current in a series circuit is the same everywhere.

EXERCISE 1.16

Use the preceding rules for the impedance of devices in parallel and in series to derive the formulas (Section 1.12) for the capacitance of two capacitors (a) in parallel and (b) in series. Hint: In each case, let the individual capacitors have capacitances C_1 and C_2. Write down the impedance of the parallel or series combination; then equate it to the impedance of a capacitor with capacitance C. Find C.

Let's try out these techniques on the simplest circuit imaginable, an ac voltage applied across a capacitor, which we considered just previously. Then, after a brief look at power in reactive circuits (to finish laying the groundwork), we'll analyze some simple but extremely important and useful RC filter circuits.

Imagine putting a 1μF capacitor across a 110 volt (rms) 60Hz power line. What current flows? Using complex Ohm's law, we have

$$\mathbf{Z} = -j/\omega C$$

Therefore, the current is given by

$$\mathbf{I} = \mathbf{V}/\mathbf{Z}$$

The phase of the voltage is arbitrary, so let us choose $\mathbf{V} = A$, i.e. $V(t) = A\cos\omega t$, where the amplitude $A = 110\sqrt{2} \approx 156$ volts. Then

$$\mathbf{I} = j\omega CA \approx 0.059\sin\omega t$$

The resulting current has an amplitude of 59mA (41.5mA rms) and leads the voltage by 90°. This agrees with our previous calculation. Note that if we just wanted to know the magnitude of the current, and didn't care what the relative phase was, we could have avoided doing any complex algebra: If

$$\mathbf{A} = \mathbf{B}/\mathbf{C}$$

then

$$A = B/C$$

where A, B, and C are the magnitudes of the respective complex numbers; this holds for multiplication, also (see Exercise 1.17). Thus, in this case,

$$I = V/Z = \omega CV$$

This trick is often useful.

Surprisingly, there is no power dissipated by the capacitor in this example. Such activity won't increase your electric bill; you'll see why in the next section. Then we will go on to look at circuits containing resistors and capacitors with our complex Ohm's law.

EXERCISE 1.17

Show that if $\mathbf{A}=\mathbf{BC}$, then $A=BC$, where A, B, and C are magnitudes. Hint: Represent each complex number in polar form, i.e., $\mathbf{A}=Ae^{i\theta}$.

Power in reactive circuits

The instantaneous power delivered to any circuit element is always given by the product $P = VI$. However, in reactive circuits where V and I are not simply proportional, you can't just multiply them together. Funny things can happen; for instance, the sign of the product can reverse over one cycle of the ac signal. Figure 1.49 shows an example. During time intervals A and C, power is being delivered to the capacitor (albeit at a variable rate), causing it to charge up; its stored energy is increasing (power is the rate of change of energy). During intervals B and D, the power delivered to the capacitor is negative; it is discharging. The average power over a whole cycle for this example is in fact exactly zero, a statement that is always true for any purely reactive circuit element (inductors, capacitors, or any combination thereof). If you know your trigonometric integrals, the next exercise will show you how to prove this.

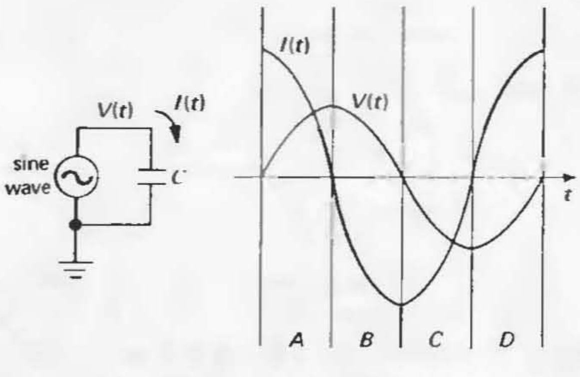


Figure 1.49. When driven by a sine wave, the current through a capacitor leads the voltage by 90°.

EXERCISE 1.18

Optional exercise: Prove that a circuit whose current is 90° out of phase with the driving voltage consumes no power, averaged over an entire cycle.

How do we find the average power consumed by an arbitrary circuit? In general, we can imagine adding up little pieces of VI product, then dividing by the elapsed time. In other words,

$$P = \frac{1}{T}\int_0^T V(t)I(t)\,dt$$

where T is the time for one complete cycle. Luckily, that's almost never necessary. Instead, it is easy to show that the average power is given by

$$P = \mathcal{R}e(\mathbf{V}\mathbf{I}^*) = \mathcal{R}e(\mathbf{V}^*\mathbf{I})$$

where $\mathbf{V}$ and $\mathbf{I}$ are complex rms amplitudes.

Let's take an example. Consider the preceding circuit, with a 1 volt (rms) sine wave driving a capacitor. We'll do everything with rms amplitudes, for simplicity. We have

$$\mathbf{V} = 1$$

$$\mathbf{I} = \frac{\mathbf{V}}{-j/\omega C} = j\omega C$$

$$P = \mathcal{R}e(\mathbf{V}\mathbf{I}^*) = \mathcal{R}e(-j\omega C) = 0$$

That is, the average power is zero, as stated earlier.

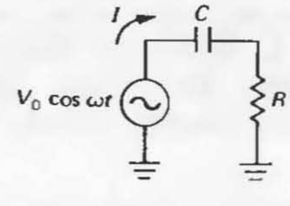


Figure 1.50

As another example, consider the circuit shown in Figure 1.50. Our calculations go like this:

$$\mathbf{Z} = R - \frac{j}{\omega C}$$

$$\mathbf{V} = V_0$$

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{V_0}{R - (j/\omega C)} = \frac{V_0[R + (j/\omega C)]}{R^2 + (1/\omega^2 C^2)}$$

$$P = \mathcal{R}e(\mathbf{V}\mathbf{I}^*) = \frac{V_0^2 R}{R^2 + (1/\omega^2 C^2)}$$

(In the third line we multiplied numerator and denominator by the complex conjugate of the denominator, in order to make the denominator real.) This is less than the product of the magnitudes of $\mathbf{V}$ and $\mathbf{I}$. In fact, the ratio is called the *power factor*:

$$|\mathbf{V}|\,|\mathbf{I}| = \frac{V_0^2}{[R^2 + (1/\omega^2 C^2)]^{1/2}}$$

$$\text{power factor} = \frac{\text{power}}{|\mathbf{V}|\,|\mathbf{I}|} = \frac{R}{[R^2 + (1/\omega^2 C^2)]^{1/2}}$$

in this case. The power factor is the cosine of the phase angle between the voltage and the current, and it ranges from 0 (purely reactive circuit) to 1 (purely resistive). A power factor less than 1 indicates some component of reactive current.

EXERCISE 1.19

Show that all the average power delivered to the preceding circuit winds up in the resistor. Do this by computing the value of V_R^2/R. What is that power, in watts, for a series circuit of a 1μF capacitor and a 1.0k resistor placed across the 110 volt (rms), 60Hz power line?

Power factor is a serious matter in large-scale electrical power distribution, because reactive currents don't result in useful power being delivered to the load, but cost the power company plenty in terms of I^2R heating in the resistance of generators, transformers, and wiring. Although residential users are only billed for "real" power [$\mathcal{R}e(\mathbf{VI}^*)$], the power company charges industrial users according to the power factor. This explains the capacitor yards that you see behind large factories, built to cancel the inductive reactance of industrial machinery (i.e., motors).

EXERCISE 1.20

Show that adding a series capacitor of value $C = 1/\omega^2 L$ makes the power factor equal 1.0 in a series RL circuit. Now do the same thing, but with the word "series" changed to "parallel."

Voltage dividers generalized

Our original voltage divider (Fig. 1.5) consisted of a pair of resistors in series to ground, input at the top and output at the junction. The generalization of that simple resistive divider is a similar circuit in which either or both resistors are replaced by a capacitor or inductor (or a more complicated network made from R, L, and C), as in Figure 1.51. In general, the division ratio $V_{\text{out}}/V_{\text{in}}$ of such a divider is not constant, but depends on frequency. The analysis is straightforward:

$$\mathbf{I} = \frac{\mathbf{V}_{\text{in}}}{\mathbf{Z}_{\text{total}}}$$

$$\mathbf{Z}_{\text{total}} = \mathbf{Z}_1 + \mathbf{Z}_2$$

$$\mathbf{V}_{\text{out}} = \mathbf{Z}_2 = \mathbf{V}_{\text{in}}\frac{\mathbf{Z}_2}{\mathbf{Z}_1 + \mathbf{Z}_2}$$

Rather than worrying about this result in general, let's look at some simple, but very important, examples.

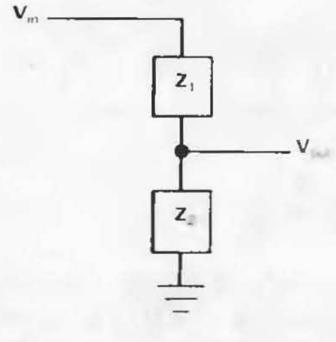


Figure 1.51. Generalized voltage divider: a pair of arbitrary impedances.

1.19 *RC* filters

By combining resistors with capacitors it is possible to make frequency-dependent voltage dividers, owing to the frequency dependence of a capacitor's impedance $\mathbf{Z}_C = -j/\omega C$. Such circuits can have the desirable property of passing signal frequencies of interest while rejecting undesired signal frequencies. In this section you will see examples of the simplest such RC filters, which we will be using frequently throughout the book. Chapter 5 and Appendix H describe filters of greater sophistication.

High-pass filters

Figure 1.52 shows a voltage divider made from a capacitor and a resistor. Complex Ohm's law gives

$$= \frac{\mathbf{V}_{\text{in}}}{\mathbf{Z}_{\text{total}}} = \frac{\mathbf{V}_{\text{in}}}{R - (j/\omega C)}$$

$$= \frac{\mathbf{V}_{\text{in}}[R + (j/\omega C)]}{R^2 + 1/\omega^2 C^2}$$

(For the last step, multiply top and bottom by the complex conjugate of the denominator.) So the voltage across R is just

$$\mathbf{V}_{\text{out}} = \mathbf{Z}_R = R = \frac{\mathbf{V}_{\text{in}}[R + (j/\omega C)]R}{R^2 + (1/\omega^2 C^2)}$$

Most often we don't care about the phase of V_{out}, just its amplitude:

$$\mathbf{V}_{\text{out}} = (\mathbf{V}_{\text{out}}\mathbf{V}^*_{\text{out}})^{1/2}$$
$$= \frac{R}{[R^2 + (1/\omega^2C^2)]^{1/2}} V_{\text{in}}$$

Note the analogy to a resistive divider, where

$$V_{\text{out}} = \frac{R_1}{R_1 + R_2} V_{\text{in}}$$

Here the impedance of the series RC combination (Fig. 1.53) is as shown in Figure 1.54. So the "response" of this circuit, ignoring phase shifts by taking magnitudes of the complex amplitudes, is given by

$$V_{\text{out}} = \frac{R}{[R^2 + (1/\omega^2C^2)]^{1/2}} V_{\text{in}}$$
$$= \frac{2\pi fRC}{[1 + (2\pi fRC)^2]^{1/2}} V_{\text{in}}$$

and looks as shown in Figure 1.55. We could have gotten this result immediately by taking the ratio of the *magnitudes* of impedances, as in Exercise 1.17 and the example immediately preceding it; the numerator is the magnitude of the impedance of the lower leg of the divider (R), and the denominator is the magnitude of the impedance of the series combination of R and C.

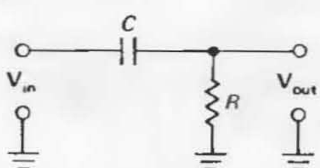


Figure 1.52. High-pass filter.

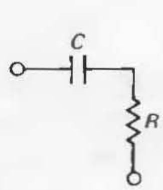


Figure 1.53

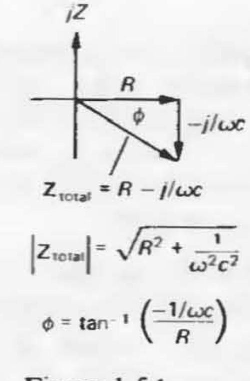


Figure 1.54

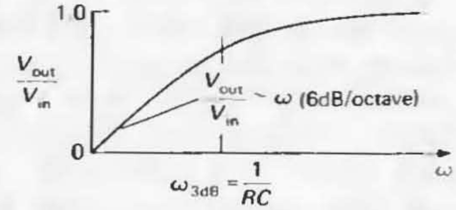
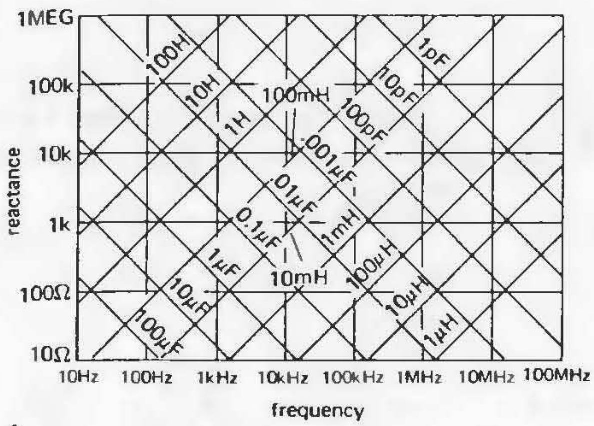


Figure 1.55. Frequency response of high-pass filter.

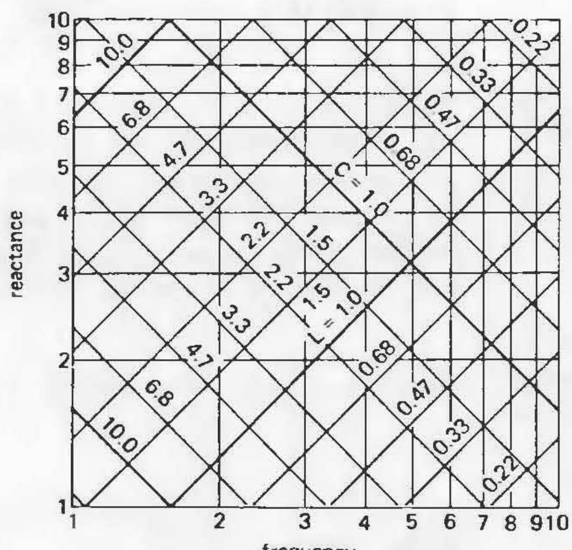
You can see that the output is approximately equal to the input at high frequencies (how high? $\gtrsim 1/RC$) and goes to zero at low frequencies. This is a very important result. Such a circuit is called a high-pass filter, for obvious reasons. It is very common. For instance, the input to the oscilloscope (Appendix A) can be switched to ac coupling. That's just an RC high-pass filter with the bend at about 10Hz (you would use ac coupling if you wanted to look at a small signal riding on a large dc voltage). Engineers like to refer to the -3dB "breakpoint" of a filter (or of any circuit that behaves like a filter). In the case of the simple RC high-pass filter, the -3dB breakpoint is given by

$$f_{3\text{dB}} = 1/2\pi RC$$

Note that the capacitor lets no steady current through ($f = 0$). This use as a dc *blocking capacitor* is one of its most frequent applications. Whenever you need to couple a signal from one amplifier to another, you almost invariably use a capacitor. For instance, every hi-fi audio amplifier has all its inputs capacitively coupled, because it doesn't know what dc level its input signals might be riding on. In such a coupling application you always pick R and C so that all frequencies of interest (in this case, 20Hz–20kHz) are passed without loss (attenuation).



A



B

Figure 1.56. A. Reactance of inductors and capacitors versus frequency; all decades are identical, except for scale.
B. A single decade from part A expanded, with standard 20% component values shown.

You often need to know the impedance of a capacitor at a given frequency (e.g., for design of filters). Figure 1.56 provides a very useful graph covering large ranges of capacitance and frequency, giving the value of $|\mathbf{Z}| = 1/2\pi fC$.

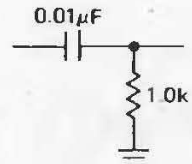


Figure 1.57

As an example, consider the filter shown in Figure 1.57. It is a high-pass filter with the 3dB point at 15.9kHz. The impedance of a load driven by it should be much larger than 1.0k in order to prevent circuit loading effects on the filter's output, and the driving source should be able to drive a 1.0k load without significant attenuation (loss of signal amplitude) in order to prevent circuit loading effects by the filter on the signal source.

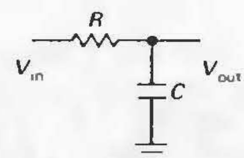


Figure 1.58. Low-pass filter.

Low-pass filters

You can get the opposite frequency behavior in a filter by interchanging R and C (Fig. 1.58). You will find

$$V_{\text{out}} = \frac{1}{(1 + \omega^2 R^2 C^2)^{1/2}} V_{\text{in}}$$

as seen in Figure 1.59. This is called a low-pass filter. The 3dB point is again at a frequency

$$f = 1/2\pi RC$$

Low-pass filters are quite handy in real life. For instance, a low-pass filter can be used to eliminate interference from nearby radio and television stations (550kHz–800MHz), a problem that plagues audio amplifiers and other sensitive electronic equipment.

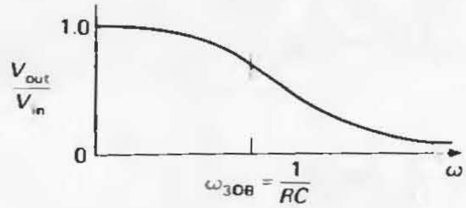


Figure 1.59. Frequency response of low-pass filter.

EXERCISE 1.21

Show that the preceding expression for the response of an RC low-pass filter is correct.

The low-pass filter's output can be viewed as a signal source in its own right. When driven by a perfect ac voltage (zero source impedance), the filter's output looks like R at low frequencies (the perfect signal source can be replaced by a short, i.e., by its small-signal source impedance, for the purpose of impedance calculations). It drops to zero impedance at high frequencies, where the capacitor dominates the output impedance. The signal driving the filter sees a load of R plus the load resistance at low frequencies, dropping to R at high frequencies.

In Figure 1.60, we've plotted the same low-pass filter response with *logarithmic* axes, which is a more usual way of doing it. You can think of the vertical axis as decibels, and the horizontal axis as octaves (or decades). On such a plot, equal distances correspond to equal ratios. We've also plotted the phase shift, using a linear

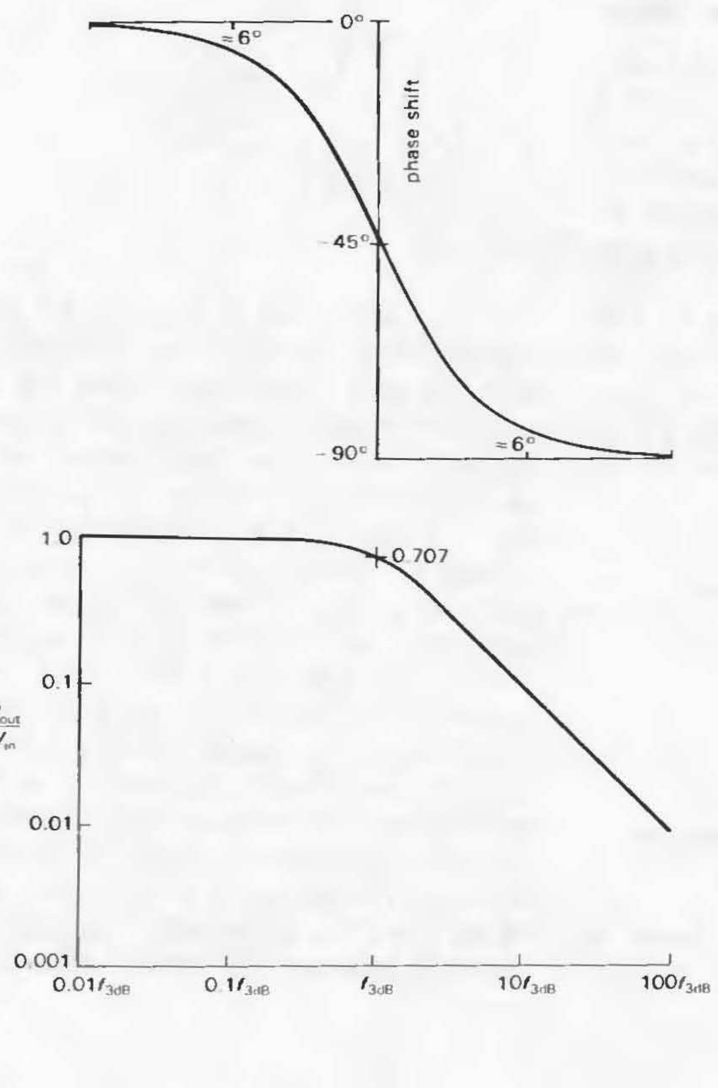


Figure 1.60. Frequency response (phase and amplitude) of low-pass filter, plotted on logarithmic axes. Note that the phase shift is 45° at the 3dB point and is within 6° of its asymptotic value for a decade of frequency change.

vertical axis (degrees) and the same logarithmic frequency axis. This sort of plot is good for seeing the detailed response even when it is greatly attenuated (as at right); we'll see a number of such plots in Chapter 5, when we treat active filters. Note that the filter curve plotted here becomes a straight line at large attenuations, with a slope of -20dB/decade (engineers prefer to say "-6dB/octave"). Note also that the phase shift goes smoothly from 0° (at frequencies well below the breakpoint) to 90° (well above it), with a value of 45° at the -3dB point. A rule of thumb for single-section RC filters is that the phase shift is $\approx 6°$ from its asymptotic value at $0.1f_{3dB}$ and $10f_{3dB}$.

EXERCISE 1.22
Prove the last assertion.

An interesting question is the following: Is it possible to make a filter with some arbitrary specified amplitude response and some other specified phase response? Surprisingly, the answer is no: The demands of causality (i.e., that response must follow cause, not precede it) force a relationship between phase and amplitude response of realizable analog filters (known officially as the Kramers-Kronig relation).

RC differentiators and integrators in the frequency domain

The RC differentiator that we saw in Section 1.14 is exactly the same circuit as the high-pass filter in this section. In fact, it can be considered as either, depending on whether you're thinking of waveforms in the time domain or response in the frequency domain. We can restate the earlier time-domain condition for its proper operation ($V_{out} \ll V_{in}$) in terms of the frequency response: For the output to be small compared with the input, the signal frequency (or frequencies) must be well below the 3dB point. This is easy to check. Suppose we have the input signal

$$V_{in} = \sin \omega t$$

Then, using the equation we obtained earlier for the differentiator output,

$$V_{out} = RC\frac{d}{dt}\sin \omega t = \omega RC \cos \omega t$$

and so $V_{out} \ll V_{in}$ if $\omega RC \ll 1$, i.e., $RC \ll 1/\omega$. If the input signal contains a range of frequencies, this must hold for the highest frequencies present in the input.

The RC integrator (Section 1.15) is the same circuit as the low-pass filter; by similar reasoning, the criterion for a good integrator is that the lowest signal frequencies must be well above the 3dB point.

Inductors versus capacitors

Inductors could be used, instead of capacitors, in combination with resistors to make low-pass (or high-pass) filters. In practice, however, you rarely see RL low- or high-pass filters. The reason is that inductors tend to be more bulky and expensive and perform less well (i.e., they depart further from the ideal) than capacitors. If you have a choice, use a capacitor. One exception to this general statement is the use of ferrite beads and chokes in high-frequency circuits. You just string a few beads here and there in the circuit; they make the wire interconnections slightly inductive, raising the impedance at very high frequencies and preventing "oscillations," without the added resistance you would get with an RC filter. An RF "choke" is an inductor, usually a few turns of wire wound on a ferrite core, used for the same purpose in RF circuits.

□ 1.20 Phasor diagrams

There's a nice graphic method that can be very helpful when trying to understand reactive circuits. Let's take an example, namely the fact that an RC filter attenuates 3dB at a frequency $f = 1/2\pi RC$, which we derived in Section 1.19. This is true for both high-pass and low-pass filters. It is easy to get a bit confused here, because at that frequency the reactance of the capacitor equals the resistance of the resistor; so you might at first expect 6dB attenuation. That is what you would get, for example, if you were to replace the capacitor by a resistor of the same impedance (recall that 6dB means half voltage). The confusion arises because the capacitor is reactive, but the matter is clarified by a phasor diagram (Fig. 1.61). The axes are the real (resistive) and imaginary (reactive) components of the impedance. In a series circuit like this, the axes also represent the (complex) voltage, because the current is the same everywhere. So for this circuit (think of it as an R–C voltage divider) the input voltage (applied across the series R–C pair) is proportional to the length of the hypotenuse, and the output voltage (across R only) is proportional to the length of the R leg of the triangle. The diagram represents the situation at the frequency where the magnitude of the capacitor's reactance equals R, i.e., $f = 1/2\pi RC$, and shows that the ratio of output voltage to input voltage is $1/\sqrt{2}$, i.e., -3dB.

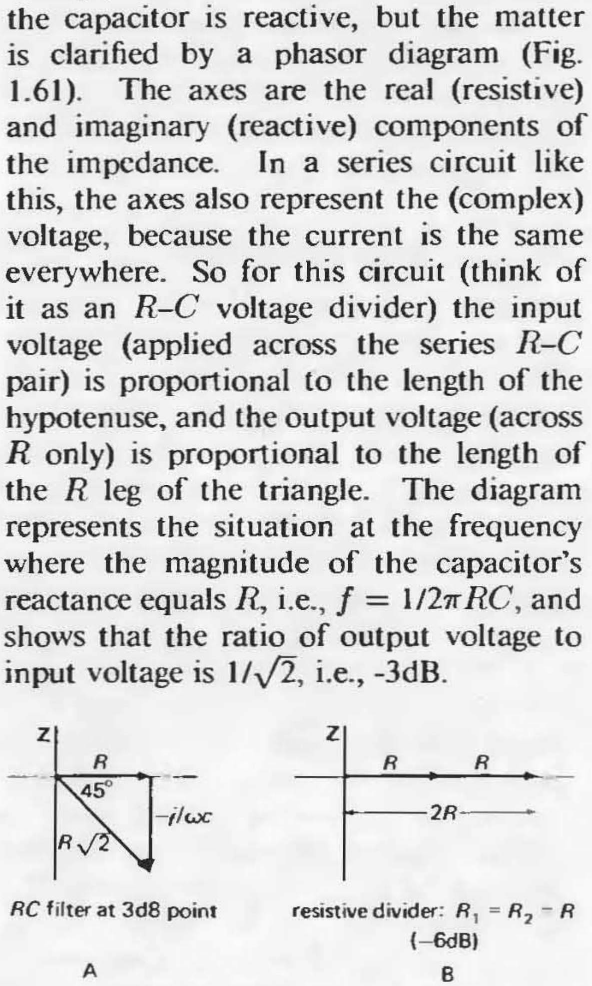


Figure 1.61

The angle between the vectors gives the phase shift from input to output. At the 3dB point, for instance, the output amplitude equals the input amplitude divided by the square root of 2, and it leads by 45° in phase. This graphic method makes it easy to read off amplitude and phase relationships in RLC circuits. For example, you can use it to get the response of the high-pass filter that we previously derived algebraically.

EXERCISE 1.23

Use a phasor diagram to derive the response of an RC high-pass filter:

$$V_{\text{out}} = \frac{R}{[R^2 + (1/\omega^2 C^2)]^{1/2}} V_{\text{in}}$$

EXERCISE 1.24

At what frequency does an RC low-pass filter attenuate by 6dB (output voltage equal to half the input voltage)? What is the phase shift at that frequency?

EXERCISE 1.25

Use a phasor diagram to obtain the low-pass filter response previously derived algebraically.

In the next chapter (Section 2.08) you will see a nice example of phasor diagrams in connection with a constant-amplitude phase-shifting circuit.

1.21 "Poles" and decibels per octave

Look again at the response of the RC low-pass filter (Fig. 1.59). Far to the right of the "knee" the output amplitude is dropping proportional to $1/f$. In one octave (as in music, one octave is twice the frequency) the output amplitude will drop to half, or -6dB; so a simple RC filter has a 6dB/octave falloff. You can make filters with several RC sections; then you get 12dB/octave (two RC sections), 18dB/octave (three sections), etc. This is the usual way of describing how a filter behaves beyond the cutoff. Another popular way is to say a "3-pole filter," for instance, meaning a filter with three RC sections (or one that behaves like one). (The word "pole" derives from a method of analysis that is beyond the scope of this book and that involves complex transfer functions in the complex frequency plane, known by engineers as the "s-plane.")

A caution on multistage filters: You can't simply cascade several identical filter sections in order to get a frequency response that is the concatenation of the individual responses. The reason is that each stage will load the previous one significantly (since they're identical), changing the overall response. Remember that the response function we derived for the simple RC filters was based on a zero-impedance driving source and an infinite-impedance load. One solution is to make each successive filter section have much higher impedance than the preceding one. A better solution involves active circuits like transistor or operational amplifier (op-amp) interstage "buffers," or active filters. These subjects will be treated in Chapters 2 through 5.

1.22 Resonant circuits and active filters

When capacitors are combined with inductors or are used in special circuits called active filters, it is possible to make circuits that have very sharp frequency characteristics (e.g., a large peak in the response at a particular frequency), as compared with the gradual characteristics of the RC filters we've seen so far. These circuits find applications in various audiofrequency and radiofrequency devices. Let's now take a quick look at LC circuits (there will be more on them, and active filters, in Chapter 5 and Appendix H).

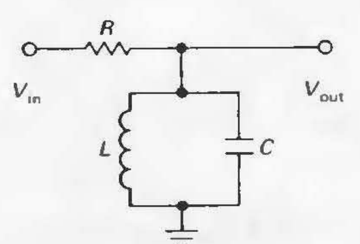


Figure 1.62. LC resonant circuit: bandpass filter.

First, consider the circuit shown in Figure 1.62. The reactance of the LC combination at frequency f is just

$$\frac{1}{Z_{LC}} = \frac{1}{Z_L} + \frac{1}{Z_C} = \frac{1}{j\omega L} - \frac{\omega C}{j}$$
$$= j\left(\omega C - \frac{1}{\omega L}\right)$$

i.e.,

$$Z_{LC} = \frac{j}{(1/\omega L) - \omega C}$$

In combination with R it forms a voltage divider; because of the opposite behaviors of inductors and capacitors, the impedance of the parallel LC goes to infinity at the *resonant frequency* $f_0 = 1/2\pi\sqrt{LC}$ (i.e., $\omega_0 = 1/\sqrt{LC}$), giving a peak in the response there. The overall response is as shown in Figure 1.63.

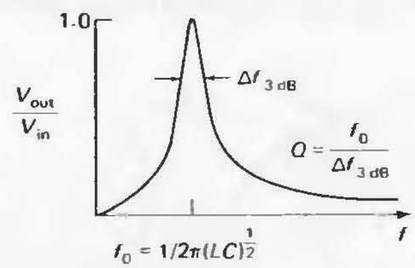


Figure 1.63

In practice, losses in the inductor and capacitor limit the sharpness of the peak, but with good design these losses can be made very small. Conversely, a Q-spoiling resistor is sometimes added intentionally to reduce the sharpness of the resonant peak. This circuit is known simply as a parallel LC resonant circuit or a tuned circuit and is used extensively in radiofrequency circuits to select a particular frequency for amplification (the L or C can be variable, so you can tune the resonant frequency). The higher the driving impedance, the sharper the peak; it is not uncommon to drive them with something approaching a current source, as you will see later. The *quality factor* Q is a measure of the sharpness of the peak. It equals the resonant frequency divided by the width at the −3dB points. For a parallel RLC circuit, $Q = \omega_0 RC$.

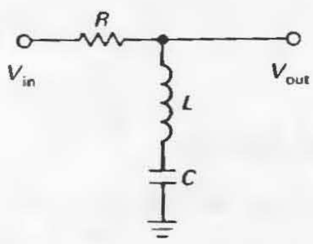


Figure 1.64. LC notch filter ("trap").

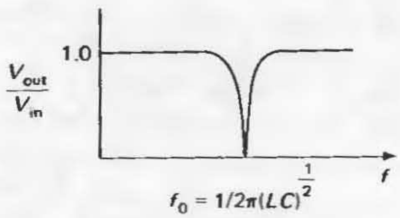


Figure 1.65

Another variety of LC circuit is the series LC (Fig. 1.64). By writing down the impedance formulas involved, you can convince yourself that the impedance of the LC goes to zero at resonance [$f_0 = 1/2\pi(LC)^{1/2}$]; such a circuit is a "trap" for signals at or near the resonant frequency, shorting them to ground. Again, this circuit finds application mainly in radiofrequency circuits. Figure 1.65 shows what the response looks like. The Q of a series RLC circuit is $Q = \omega_0 L/R$.

EXERCISE 1.26

Find the response (V_{out}/V_{in} versus frequency) for the series LC trap circuit in Figure 1.64.

1.23 Other capacitor applications

In addition to their uses in filters, resonant circuits, differentiators, and integrators, capacitors are needed for several other important applications. We will treat these in detail later in the book, mentioning them here only as a preview.

Bypassing

The impedance of a capacitor goes down with increasing frequency. This is the basis of another important application: bypassing. There are places in circuits where you want to allow a dc (or slowly varying) voltage, but don't want signals present. Placing a capacitor across that circuit element (usually a resistor) will help to kill any signals there. You choose the capacitor value so that its impedance at signal frequencies is small compared with what it is bypassing. You will see much more of this in later chapters.

Power-supply filtering

Power-supply filtering is really a form of bypassing, although we usually think of it as energy storage. The dc voltages used in electronics are usually generated from the ac line voltage by a process called *rectification* (which will be treated later in this chapter); some residue of the 60Hz input remains, and this can be reduced as much as desired by means of bypassing with suitably large capacitors. These capacitors really are large – they're the big shiny round things you see inside most electronic instruments. You will see how to design power supplies and filters later in this chapter and again in Chapter 6.

Timing and waveform generation

A capacitor supplied with a constant current charges up with a ramp waveform. This is the basis of ramp and sawtooth generators, used in function generators, oscilloscope sweep circuits, analog/digital converters, and timing circuits. RC circuits are also used for timing, and they form the basis of digital delay circuits (monostable multivibrators). These timing and waveform applications are important in many areas of electronics and will be covered in Chapters 3, 5, 8, and 9.

TABLE 1.1. DIODES

Type	$V_{R(max)}$[a] (V)	$I_{R(max)}$[b] (μA)	Continuous V_F (V)	@ I_F (mA)	Peak V_F (V)	@ I_F (A)	Reverse recovery (ns)	Capacitance (10V) (pF)	Class	Comments
PAD-1	45	1pA@20V	0.8	5	–	–	–	0.8	lowest I_R	Siliconix
FJT1100	30	0.001	–	–	1.1	0.05	–	1.2	very low I_R	1pA@5V, 10pA@15V
ID101	30	10pA@10V	0.8	1	1.1	0.03	–	0.8	very low I_R	Intersil; dual
1N3595	150	3	0.7	10	<1.0	0.2	3000	8.0	low I_R	1nA@125V
1N914	75	5	0.75	10	1.1	0.1	4	1.3	gen purp sig diode	indus std; same as 1N4148
1N6263	60	10	0.4	1	0.7	0.01	0	1.0	Schottky; low V_F	
1N3062	75	50	<1.0	20[b]	–	–	2	0.6	low cap, sig diode	1pF at 0 volts
1N4305	75	50	0.6	1	–	–	4	1.5	controlled V_F	
1N4002	100	50	0.9	1000	2.3	25	3500	15	1-amp rect	indus std; 7-member fam
1N4007	1000	50	0.9	1000	2.3	25	5000	10		
1N5819	40	10000	0.4	1000	1.1	20	–	50	pwr Schottky	lead mounted
1N5822	40	20000	0.45	3000	1.3	50	–	180	pwr Schottky	lead mounted
1N5625	400	50	1.1	5000	2.0	50	2500	45	5-amp rect	lead mounted
1N1183A	50	1000	1.1	40000	1.3	100	–	–	high curr rect	1N1183RA reverse

(a) $V_{R(max)}$ is repetitive peak reverse voltage, 25°C, 10μA leakage. (b) $I_{R(max)}$ is reverse leakage current at V_R and 100°C ambient temperature.

1.24 Thévenin's theorem generalized

When capacitors and inductors are included, Thévenin's theorem must be restated: Any two-terminal network of resistors, capacitors, inductors, and signal sources is equivalent to a single complex impedance in series with a single signal source. As before, you find the impedance and the signal source from the open-circuit output voltage and the short-circuit current.

DIODES AND DIODE CIRCUITS

1.25 Diodes

The circuit elements we've discussed so far (resistors, capacitors, and inductors) are all *linear*, meaning that a doubling of the applied signal (a voltage, say) produces a doubling of the response (a current, say). This is true even for the reactive devices (capacitors and inductors). These devices are also *passive*, meaning that they don't have a built-in source of power. And they are all two-terminal devices, which is self-explanatory.

Figure 1.66. Diode.

The diode (Fig. 1.66) is a very important and useful two-terminal passive *nonlinear* device. It has the *V–I* curve shown in Figure 1.67. (In keeping with the general philosophy of this book, we will not attempt to describe the solid-state physics that makes such devices possible.)

The diode's arrow (the anode terminal) points in the direction of forward current flow. For example, if the diode is in a circuit in which a current of 10mA is flowing from anode to cathode, then (from the graph) the anode is approximately 0.5 volt more positive than the cathode; this is called the "forward voltage drop." The reverse current, which is measured in the nanoamp range for a general-purpose diode (note the different scales in the graph for forward and reverse current), is almost never of any consequence until you reach the reverse breakdown voltage (also called the peak inverse voltage, PIV), typically 75 volts for a general-purpose diode like the 1N914. (Normally you never subject a diode to voltages large enough to cause reverse breakdown; the exception is the zener diode we mentioned earlier.) Frequently, also, the forward voltage drop of about 0.5 and 0.8 volt is of little concern, and the diode can be treated as a good approximation to an ideal one-way conductor. There are other important characteristics that distinguish the thousands of diode types available, e.g., maximum forward current, capacitance, leakage current, and reverse recovery time (see Table 1.1 for characteristics of some typical diodes).

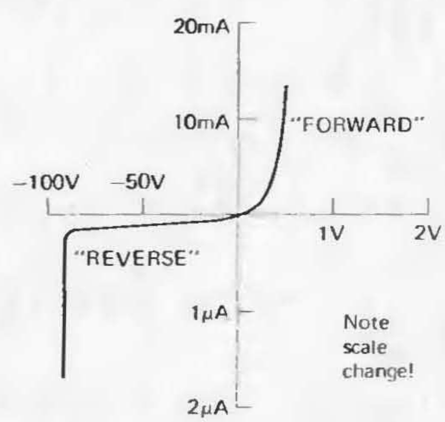


Figure 1.67. Diode *V–I* curve.

Before jumping into some circuits with diodes, we should point out two things: (a) A diode doesn't actually have a resistance (it doesn't obey Ohm's law). (b) If you put some diodes in a circuit, it won't have a Thévenin equivalent.

1.26 Rectification

A rectifier changes ac to dc; this is one of the simplest and most important applications of diodes (diodes are sometimes called rectifiers). The simplest circuit is shown in Figure 1.68. The "ac" symbol represents a source of ac voltage; in electronic circuits it is usually provided by a transformer, powered from the ac power line. For a sine-wave input that is much larger than the forward drop (about 0.6V for silicon diodes, the usual type), the output will look like that in Figure 1.69. If you think of the diode as a one-way conductor, you won't have any trouble understanding how the circuit works. This circuit is called a *half-wave rectifier*, because only half of the input waveform is used.

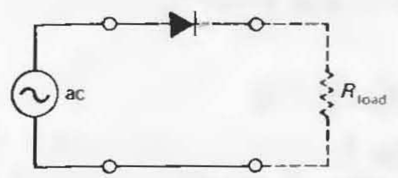


Figure 1.68. Half-wave rectifier.

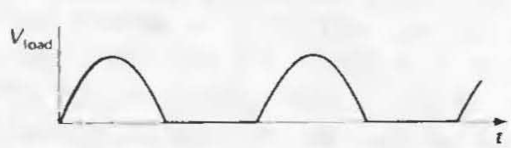


Figure 1.69

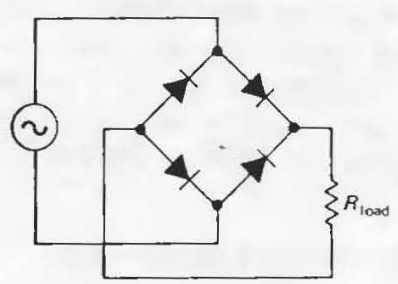


Figure 1.70. Full-wave bridge rectifier.

Figure 1.70 shows another rectifier circuit, a full-wave bridge. Figure 1.71 shows the voltage across the load for which the whole input waveform is used. The gaps at zero voltage occur because of the diodes' forward voltage drop. In this circuit, two diodes are always in series with the input; when you design low-voltage power supplies, you have to remember that.

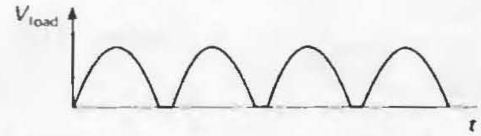


Figure 1.71

1.27 Power-supply filtering

The preceding rectified waveforms aren't good for much as they stand. They're dc only in the sense that they don't change polarity. But they still have a lot of "ripple" (periodic variations in voltage about the steady value) that has to be smoothed out in order to generate genuine dc. This we do by tacking on a low-pass filter (Fig. 1.72). Actually, the series resistor is unnecessary and is always omitted (although you sometimes see a very small resistor used to limit the peak rectifier current). The reason is that the diodes prevent flow of current back out of the capacitors, which are really serving more as energy-storage devices than as part of a classic low-pass filter. The energy stored in a capacitor is $U = \frac{1}{2}CV^2$. For C in farads and V in volts, U comes out in joules (watt-seconds).

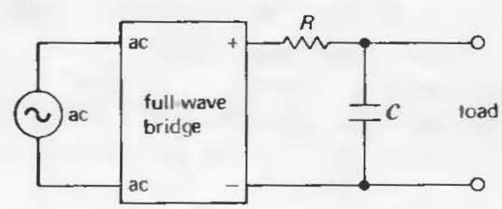


Figure 1.72

The capacitor value is chosen so that

$$R_{load}C \gg 1/f$$

(where f is the ripple frequency, here 120Hz) in order to ensure small ripple, by making the time constant for discharge much longer than the time between recharging. We will make this vague statement clearer in the next section.

Calculation of ripple voltage

It is easy to calculate the approximate ripple voltage, particularly if it is small compared with the dc (see Fig. 1.73). The load causes the capacitor to discharge somewhat between cycles (or half cycles, for full-wave rectification). If you assume that the load current stays constant (it will, for small ripple), you have

$$\Delta V = \frac{I}{C}\Delta t \qquad \left(\text{from } I = C\frac{dV}{dt}\right)$$

Just use $1/f$ (or $1/2f$ for full-wave rectification) for Δt (this estimate is a bit on the safe side, since the capacitor begins charging again in less than a half cycle). You get

$$\Delta V = \frac{I_{\text{load}}}{fC} \qquad \text{(half wave)}$$

$$\Delta V = \frac{I_{\text{load}}}{2fC} \qquad \text{(full wave)}$$

(While teaching electronics we've noticed that students love to memorize these equations! An informal poll of the authors showed that two out of two engineers don't memorize them. Please don't waste brain cells that way – instead, learn how to derive them.)

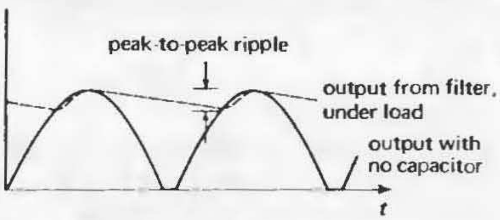


Figure 1.73. Power-supply ripple calculation.

If you wanted to do the calculation without any approximation, you would use the exact exponential discharge formula. You would be misguided in insisting on that kind of accuracy, though, for two reasons:

1. The discharge is an exponential only if the load is a resistance; many loads are not. In fact, the most common load, a *voltage regulator*, looks like a constant-current load.

2. Power supplies are built with capacitors with typical tolerances of 20% or more. Realizing the manufacturing spread, you design conservatively, allowing for the worst-case combination of component values.

In this case, viewing the initial part of the discharge as a ramp is in fact quite accurate, especially if the ripple is small, and in any case it errs in the direction of conservative design – it overestimates the ripple.

EXERCISE 1.27

Design a full-wave bridge rectifier circuit to deliver 10 volts dc with less than 0.1 volt (pp) ripple into a load drawing up to 10mA. Choose the appropriate ac input voltage, assuming 0.6 volt diode drops. Be sure to use the correct ripple frequency in your calculation.

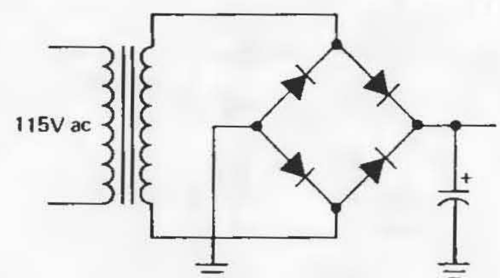


Figure 1.74. Bridge rectifier circuit. The polarity marking and curved electrode indicate a polarized capacitor, which must not be allowed to charge with the opposite polarity.

1.28 Rectifier configurations for power supplies

Full-wave bridge

A dc power supply using the bridge circuit we just discussed looks as shown in Figure 1.74. In practice, you generally buy the bridge as a prepackaged module. The smallest ones come with maximum current ratings of 1 amp average, with breakdown voltages going from 100 volts to 600 volts, or even 1000 volts. Giant bridge rectifiers are available with current ratings of 25 amps or more. Take a look at Table 6.4 for a few types.

Center-tapped full-wave rectifier

The circuit in Figure 1.75 is called a center-trapped full-wave rectifier. The output voltage is half what you get if you use a bridge rectifier. It is not the most efficient circuit in terms of transformer design, because each half of the secondary is used only half the time. Thus the current through the winding during that time is twice what it would be for a true full-wave circuit. Heating in the windings, calculated from Ohm's law, is I^2R, so you have four times the heating half the time, or twice the average heating of an equivalent full-wave bridge circuit. You would have to choose a transformer with a current rating 1.4 (square root of 2) times as large, as compared with the (better) bridge circuit; besides costing more, the resulting supply would be bulkier and heavier.

EXERCISE 1.28

This illustration of I^2R heating may help you understand the disadvantage of the center-tapped rectifier circuit. What fuse rating (minimum) is required to pass the current waveform shown in Figure 1.76, which has 1 amp average current? Hint: A fuse "blows out" by melting (I^2R heating) a metallic link, for steady currents larger than its rating. Assume for this problem that the thermal time constant of the fusible link is much longer than the time scale of the square wave, i.e., that the fuse responds to the value of I^2 averaged over many cycles.

Split supply

A popular variation of the center-tapped full-wave circuit is shown in Figure 1.77. It gives you split supplies (equal plus and minus voltages), which many circuits need. It is an efficient circuit, because both halves of the input waveform are used in each winding section.

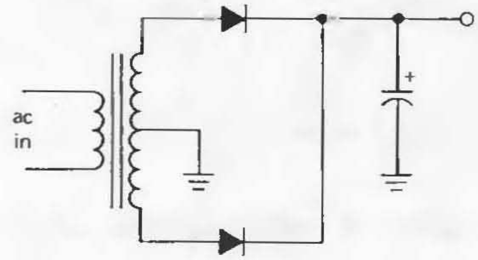


Figure 1.75. Full-wave rectifier using center-tapped transformer.

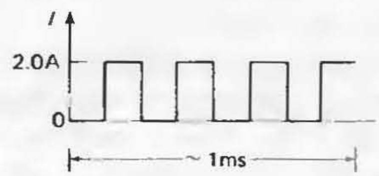


Figure 1.76

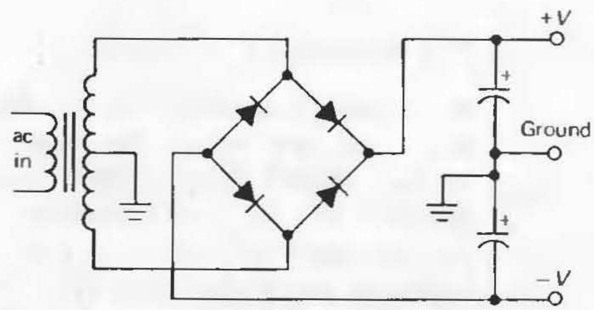


Figure 1.77. Dual-polarity (split) supply.

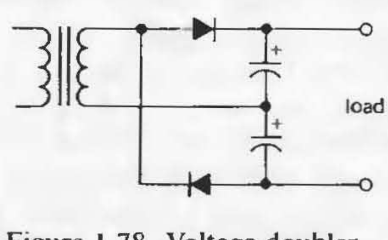


Figure 1.78. Voltage doubler.

□ Voltage multipliers

The circuit shown in Figure 1.78 is called a voltage doubler. Think of it as two half-wave rectifier circuits in series. It is officially a full-wave rectifier circuit, since

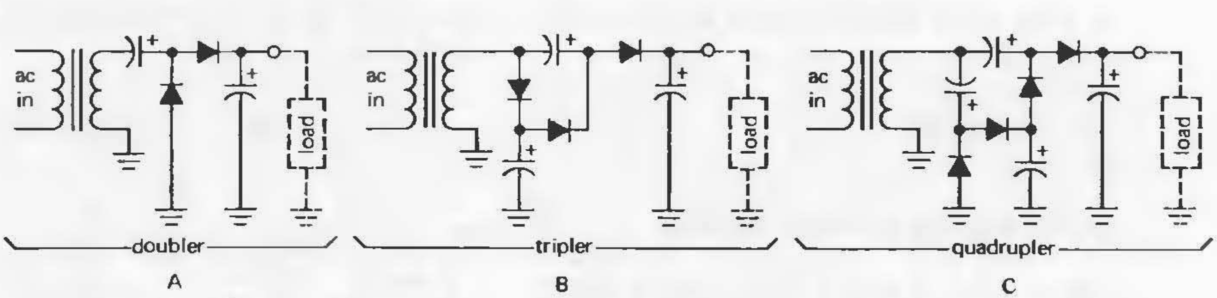


Figure 1.79. Voltage multipliers; these configurations don't require a floating voltage source.

both halves of the input waveform are used – the ripple frequency is twice the ac frequency (120Hz for the 60Hz line voltage in the United States).

Variations of this circuit exist for voltage triplers, quadruplers, etc. Figure 1.79 shows doubler, tripler, and quadrupler circuits that let you ground one side of the transformer.

1.29 Regulators

By choosing capacitors that are sufficiently large, you can reduce the ripple voltage to any desired level. This brute-force approach has two disadvantages:

1. The required capacitors may be prohibitively bulky and expensive.

2. Even with the ripple reduced to negligible levels, you still have variations of output voltage due to other causes, e.g., the dc output voltage will be roughly proportional to the ac input voltage, giving rise to fluctuations caused by input line voltage variations. In addition, changes in load current will cause the output voltage to change because of the finite internal resistances of the transformer, diode, etc. In other words, the Thévenin equivalent circuit of the dc power supply has $R > 0$.

A better approach to power-supply design is to use enough capacitance to reduce ripple to low levels (perhaps 10% of the dc voltage), then use an active *feedback circuit* to eliminate the remaining ripple. Such a feedback circuit "looks at" the output, making changes in a controllable series resistor (a transistor) as necessary to keep the output constant (Fig. 1.80).

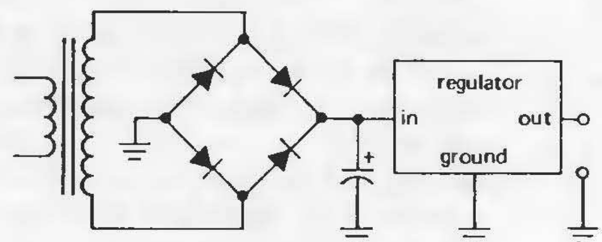


Figure 1.80. Regulated dc power supply.

These voltage regulators are used almost universally as power supplies for electronic circuits. Nowadays complete voltage regulators are available as inexpensive integrated circuits (priced under one dollar). A power supply built with a voltage regulator can be made easily adjustable and self-protecting (against short circuits, overheating, etc.), with excellent properties as a voltage source (e.g., internal resistance measured in milliohms). We will deal with regulated dc power supplies in Chapter 6.

1.30 Circuit applications of diodes

Signal rectifier

There are other occasions when you use a diode to make a waveform of one polarity only. If the input waveform isn't a sine wave, you usually don't think of it as a rectification in the sense of a power supply. For instance, you might want a train of pulses corresponding to the rising edge of a square wave. The easiest way is to rectify the differentiated wave (Fig. 1.81). Always keep in mind the 0.6 volt (approximately) forward drop of the diode. This circuit, for instance, gives no output for square waves smaller than 0.6 volt pp. If this is a problem, there are various tricks to circumvent this limitation. One possibility is to use *hot carrier diodes* (Schottky diodes), with a forward drop of about 0.25 volt (another device called a *back diode* has nearly zero forward drop, but its usefulness is limited by very low reverse breakdown voltage).

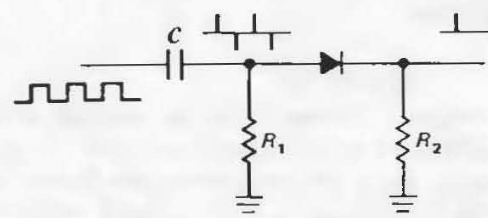


Figure 1.81

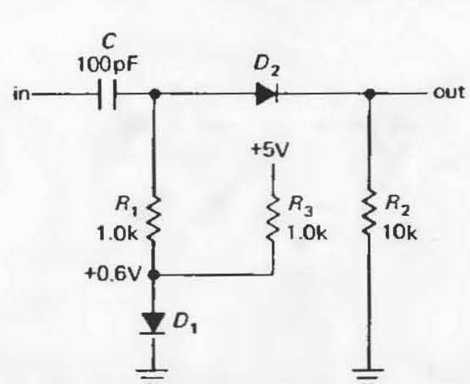


Figure 1.82. Compensating the forward voltage drop of a diode signal rectifier.

A possible circuit solution to this problem of finite diode drop is shown in Figure 1.82. Here D_1 compensates D_2's forward drop by providing 0.6 volt of *bias* to hold D_2 at the threshold of conduction. Using a diode (D_1) to provide the bias (rather than, say, a voltage divider) has several advantages: There is nothing to adjust, the compensation will be nearly perfect, and changes of the forward drop (e.g., with changing temperature) will be compensated properly. Later we will see other instances of matched-pair compensation of forward drops in diodes, transistors, and FETs: it is a simple and powerful trick.

Diode gates

Another application of diodes, which we will recognize later under the general heading of *logic*, is to pass the higher of two voltages without affecting the lower. A good example is *battery backup*, a method of keeping something running (e.g, a precision electronic clock) that must not stop when there is a power failure. Figure 1.83 shows a circuit that does the job. The battery does nothing until the power fails; then it takes over without interruption.

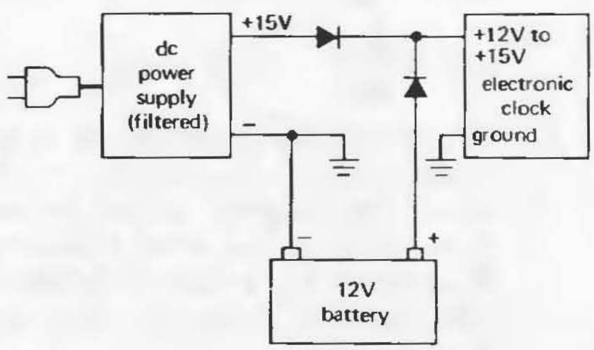


Figure 1.83. Diode OR gate: battery backup.

EXERCISE 1.29
Make a simple modification to the circuit so that the battery is charged by the dc supply (when power is on, of course) at a current of 10mA (such a circuit is necessary to maintain the battery's charge).

Diode clamps

Sometimes it is desirable to limit the range of a signal (i.e., prevent it from exceeding certain voltage limits) somewhere in a circuit. The circuit shown in Figure 1.84 will accomplish this. The diode prevents the output from exceeding about +5.6 volts, with no effect on voltages less than that (including negative voltages); the only limitation is that the input must not go so negative that the reverse breakdown voltage of the diode is exceeded (e.g., −70V for a 1N914). Diode clamps are standard equipment on all inputs in the CMOS family of digital logic. Without them, the delicate input circuits are easily destroyed by static electricity discharges during handling.

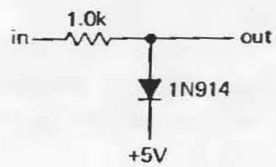


Figure 1.84. Diode voltage clamp.

EXERCISE 1.30

Design a symmetrical clamp, i.e., one that confines a signal to the range −5.6 volts to +5.6 volts.

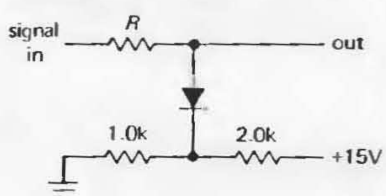


Figure 1.85

A voltage divider can provide the reference voltage for a clamp (Fig. 1.85). In this case you must ensure that the impedance looking into the voltage divider (R_{vd}) is small compared with R, because what you have looks as shown in Figure 1.86 when the voltage divider is replaced by its Thévenin equivalent circuit. When the diode conducts (input voltage exceeds clamp voltage), the output is really just the output of a voltage divider, with the Thévenin equivalent resistance of the voltage reference as the lower resistor (Fig. 1.87). So, for the values shown, the output of the clamp for a triangle-wave input would look as shown in Figure 1.88. The problem is that the voltage divider doesn't provide a stiff reference, in the language of electronics. A stiff voltage source is one that doesn't bend easily, i.e., it has low internal (Thévenin) impedance.

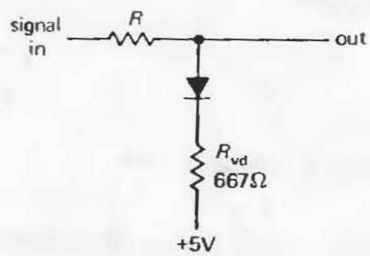


Figure 1.86

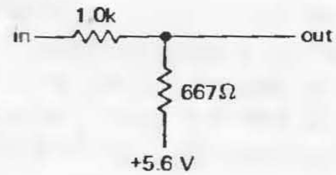


Figure 1.87

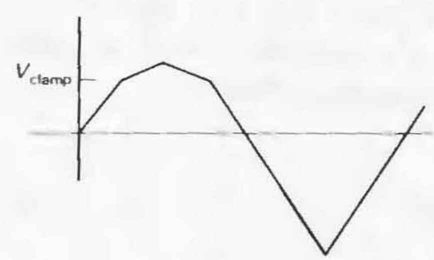


Figure 1.88

A simple way to stiffen the clamp circuit of Figure 1.85, at least for *high-frequency* signals, is to add a bypass capacitor across the 1k resistor. For example, a 15μF capacitor to ground reduces the impedance seen looking into the divider below 10 ohms for frequencies above 1kHz. (You could similarly add a bypass capacitor across D_1 in Fig. 1.82.) Of course, the effectiveness of this trick drops at low frequencies, and it does nothing at dc.

In practice, the problem of finite impedance of the voltage-divider reference can be easily solved using a transistor or operational amplifier (op-amp). This is usually a better solution than using very small resistor values, because it doesn't consume large currents, yet it provides impedances of a few ohms or less. Furthermore, there are other ways to construct a clamp, using an op-amp as part of the clamp circuit. You will see these methods in Chapter 4.

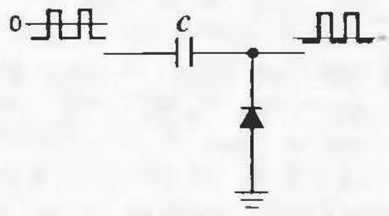


Figure 1.89. dc restoration.

One interesting clamp application is "dc restoration" of a signal that has been ac-coupled (capacitively coupled). Figure 1.89 shows the idea. This is particularly important for circuits whose inputs look like diodes (e.g., a transistor with grounded emitter); otherwise an ac-coupled signal will just fade away.

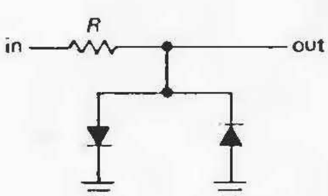


Figure 1.90. Diode limiter.

Limiter

One last clamp circuit is shown in Figure 1.90. This circuit limits the output "swing" (again, a common electronics term) to one diode drop, roughly 0.6 volt. That might seem awfully small, but if the next stage is an amplifier with large voltage amplification, its input will always be near zero volts; otherwise the output is in "saturation" (e.g., if the next stage has a gain of 1000 and operates from ±15V supplies, its input must stay in the range ±15mV in order for its output not to saturate). This clamp circuit is often used as input protection for a high-gain amplifier.

Diodes as nonlinear elements

To a good approximation the forward current through a diode is proportional to an exponential function of the voltage across it at a given temperature (for a discussion of the exact law, see Section 2.10). So you can use a diode to generate an output voltage proportional to the logarithm of a current (Fig. 1.91). Because V hovers in the region of 0.6 volt, with only small voltage changes that reflect input current variations, you can generate the input current with a resistor if the input voltage is much larger than a diode drop (Fig. 1.92).

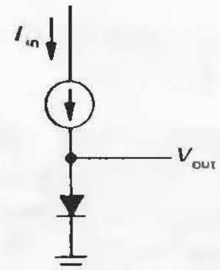


Figure 1.91. Exploiting the diode's nonlinear V–I curve: logarithmic converter.

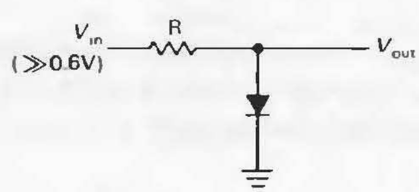


Figure 1.92

In practice, you may want an output voltage that isn't offset by the 0.6 volt diode drop. In addition, it would be nice to have a circuit that is insensitive to changes in temperature. The method of diode drop compensation is helpful here (Fig. 1.93). R_1 makes D_2 conduct, holding point A at about -0.6 volt. Point B is then near ground (making I_{in} accurately proportional to V_{in}, incidentally). As long as the two (identical) diodes are at the same temperature, there is good cancellation of the forward drops, except, of course, for the difference owing to input current through D_1, which produces the desired output. In this circuit, R_1 should be chosen so that the current through D_2 is much larger than the maximum input current, in order to keep D_2 in conduction.

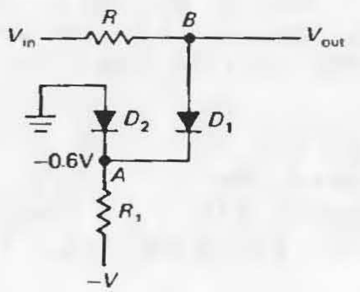


Figure 1.93. Diode drop compensation in the logarithmic converter.

In the chapter on op-amps we will examine better ways of constructing logarithmic converter circuits, along with careful methods of temperature compensation. With such methods it is possible to construct logarithmic converters accurate to a few percent over six decades or more of input current. A better understanding of diode and transistor characteristics, along with an understanding of op-amps, is necessary first. This section is meant to serve only as an introduction for things to come.

1.31 Inductive loads and diode protection

What happens if you open a switch that is providing current to an inductor? Because inductors have the property

$$V = L\frac{dI}{dt}$$

it is not possible to turn off the current suddenly, since that would imply an infinite voltage across the inductor's terminals. What happens instead is that the voltage across the inductor suddenly rises and keeps rising until it forces current to flow. Electronic devices controlling inductive loads can be easily damaged, especially the component that "breaks down" in order to satisfy the inductor's craving for continuity of current. Consider the circuit in Figure 1.94. The switch is initially closed, and current is flowing through the inductor (which might be a relay, as will be described later). When the switch is opened, the inductor "tries" to keep current flowing from A to B, as it had been. That means that terminal B goes positive relative to terminal A. In a case like this it may go 1000 volts positive before the switch contact "blows over." This shortens the life of the switch and also generates impulsive interference that may affect other circuits nearby. If the switch happens to be a transistor, it would be an understatement to say that its life is shortened; its life is ended!

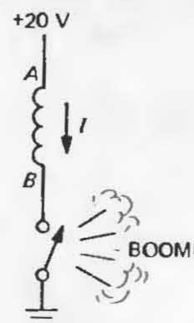


Figure 1.94. Inductive "kick."

The best solution is to put a diode across the inductor, as in Figure 1.95. When the switch is on, the diode is back-biased (from the dc drop across the inductor's winding resistance). At turn-off the diode goes into conduction, putting the switch terminal a diode drop above the positive supply voltage. The diode must be able to handle the initial diode current, which equals the steady current that had been flowing through the inductor; something like a 1N4004 is fine for nearly all cases.

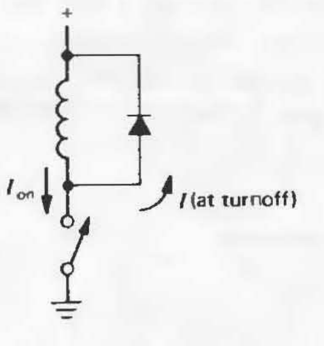


Figure 1.95. Blocking inductive kick.

The only disadvantage of this protection circuit is that it lengthens the decay of current through the inductor, since the rate of change of inductor current is proportional to the voltage across it. For applications where the current must decay quickly (high-speed impact printers, high-speed relays, etc.), it may be better to put a resistor across the inductor, choosing its value so that $V_{supply} + IR$ is less than the maximum allowed voltage across the switch. (For fastest decay with a given maximum voltage, a zener could be used instead, giving a ramp-down of current rather than an exponential decay.)

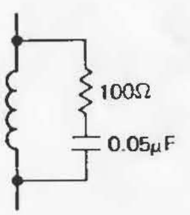


Figure 1.96. *RC* "snubber" for suppressing inductive kick.

For inductors driven from ac (transformers, ac relays), the diode protection just described will not work, since the diode will conduct on alternate half cycles when the switch is closed. In that case a good solution is an RC "snubber" network (Fig. 1.96). The values shown are typical for small inductive loads driven from the ac power line. Such a snubber should be included in all instruments that run from the ac power line, since a transformer is inductive. An alternative protection device is a metal-oxide varistor, or transient suppressor, an inexpensive device that looks something like a disc ceramic capacitor and behaves electrically like a bi-directional zener diode. They are available at voltage ratings from 10 to 1000 volts and can handle transient currents up to thousands of amperes (see Section 6.11 and Table 6.2). Putting a transient suppressor across the ac powerline terminals makes good sense in a piece of electronic equipment, not only to prevent inductive spike interference to other nearby instruments but also to prevent occasional large power-line spikes from damaging the instrument itself.

OTHER PASSIVE COMPONENTS

In the following sections we would like to introduce briefly an assortment of miscellaneous but essential components. If you are experienced in electronic construction, you may wish to proceed to the next chapter.

1.32 Electromechanical devices

Switches

These mundane but important devices seem to wind up in most electronic equipment. It is worth spending a few paragraphs on the subject. Figure 1.97 shows some common switch types.

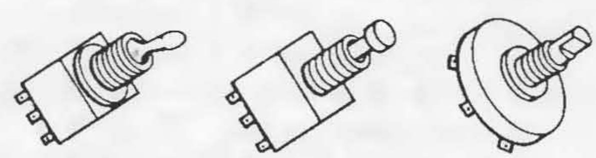


Figure 1.97. Panel switches.

Toggle switches. The simple toggle switch is available in various configurations, depending on the number of poles; Figure 1.98 shows the usual ones (SPDT indicates a single-pole double-throw switch, etc.). Toggle switches are also available with "center OFF" positions and with up to 4 poles switched simultaneously. Toggle switches are always "break before make," e.g., the moving contact never connects to both terminals in an SPDT switch.

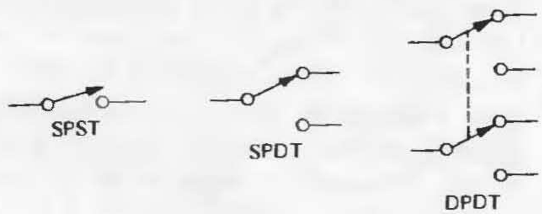


Figure 1.98. Fundamental switch types.

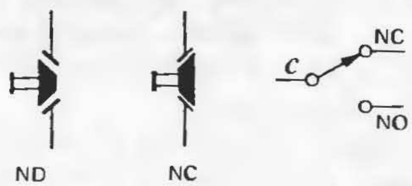


Figure 1.99. Momentary-contact (push-button) switches.

Push-button switches. Push-button switches are useful for momentary-contact applications; they are drawn schematically as shown in Figure 1.99 (NO and NC mean normally open and normally closed). For SPDT momentary-contact switches, the terminals must be labeled NO and NC, whereas for SPST types the symbol is self-explanatory. Momentary-contact switches are always "break before make." In the electrical (as opposed to electronic) industry, the terms form A, form B, and form C are used to mean SPST (NO), SPST (NC), and SPDT, respectively.

Rotary switches. Rotary switches are available with many poles and many positions, often as kits with individual wafers and shaft hardware. Both shorting (make before break) and nonshorting (break before make) types are available, and they can be mixed on the same switch. In many applications the shorting type is useful to prevent an open circuit between switch positions, because circuits can go amok with unconnected inputs. Nonshorting types are necessary if the separate lines being switched to one common line must not ever be connected to each other.

Other switch types. In addition to these basic switch types, there are available various exotic switches such as Hall-effect switches, reed switches, proximity switches, etc. All switches carry maximum current and voltage ratings; a small toggle switch might be rated at 150 volts and 5 amps. Operation with inductive loads drastically reduces switch life because of arcing during turn-off.

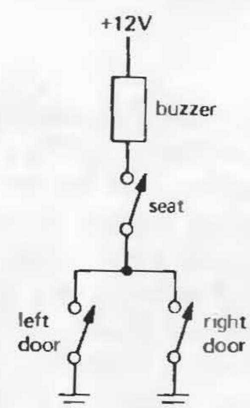


Figure 1.100

Switch examples. As an example of what can be done with simple switches, let's consider the following problem: Suppose you want to sound a warning buzzer if the driver of a car is seated and one of the car doors is open. Both doors and the driver's seat have switches, all normally open. Figure 1.100 shows a circuit that does what you want. If one OR the other door is open (switch closed) AND the seat switch is closed, the buzzer sounds. The words OR and AND are used in a logic sense here, and we will see this example again in Chapters 2 and 8 when we talk about transistors and digital logic.

Figure 1.101 shows a classic switch circuit used to turn a ceiling lamp on or off from a switch at either of two entrances to a room.

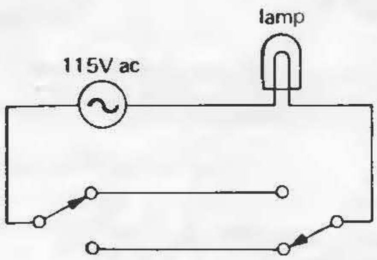


Figure 1.101. Electrician's "three-way" switch wiring.

EXERCISE 1.31

Although few electronic circuit designers know how, every *electrician* can wire up a light fixture so that any of N switches can turn it on or off. See if you can figure out this generalization of Figure 1.101. It requires two SPDT switches and $N-2$ DPDT switches. (Hint: First figure out how to use a DPDT switch to crisscross a pair of wires.)

Relays

Relays are electrically controlled switches. In the usual type, a coil pulls in an armature when sufficient coil current flows. Many varieties are available, including "latching" and "stepping" relays; the latter provided the cornerstone for telephone switching stations, and they're still popular in pinball machines. Relays are available for dc or ac excitation, and coil voltages from 5 volts up to 110 volts are common. "Mercury-wetted" and "reed" relays are intended for high-speed ($\sim$1ms) applications, and giant relays intended to switch thousands of amps are used by power companies. Many previous relay applications are now handled with transistor or FET switches, and devices known as solid-state relays are now available to handle ac switching applications. The primary uses of relays are in remote switching and high-voltage (or high-current) switching. Because it is important to keep electronic circuits electrically isolated from the ac power line, relays are useful to switch ac power while keeping the control signals electrically isolated.

Connectors

Bringing signals in and out of an instrument, routing signal and dc power around between the various parts of an instrument, providing flexibility by permitting circuit boards and larger modules of the instrument to be unplugged (and replaced) – these are the functions of the connector, an essential ingredient (and usually the most unreliable part) of any piece of electronic equipment. Connectors come in a bewildering variety of sizes and shapes.

Single-wire connectors. The simplest kind of connector is the simple pin jack or banana jack used on multimeters, power supplies, etc. It is handy and inexpensive, but not as useful as the shielded-cable or multiwire connectors you often need. The humble binding post is another form of single-wire connector, notable for the clumsiness it inspires in those who try to use it.

Shielded-cable connectors. In order to prevent capacitive pickup, and for other reasons we'll go into in Chapter 13, it is usually desirable to pipe signals around from one instrument to another in shielded coaxial cable. The most popular connector is the BNC ("baby N" connector) type that adorns most instrument front panels. It connects with a quarter-turn twist and completes both the shield (ground) circuit and inner conductor (signal) circuit simultaneously. Like all connectors used to mate a cable to an instrument, it comes in both panel-mounting and cable-terminating varieties (Fig. 1.102).

Figure 1.102. BNC connectors are the most popular type for use with shielded (coaxial) cable. From left to right: A male connector on a length of cable, a standard panel-mounted female connector, two varieties of insulated panel-mounted female connectors, and a BNC "T," a handy device to have in the laboratory.

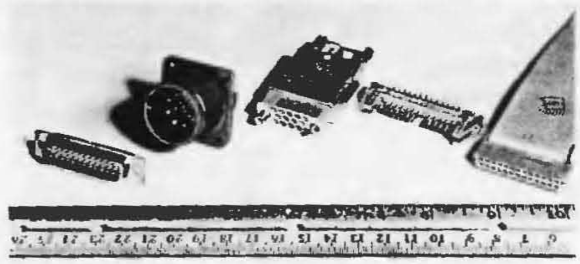


Figure 1.103. A selection of popular multipin connectors. From left to right: *D* subminiature type, available in panel- and cable-mounting versions, with 9, 15, 25, 37, or 50 pins; the venerable MS-type connector, available in many (too many!) pin and mounting configurations, including types suitable for shielded cables; a miniature rectangular connector (Winchester MRA type) with integral securing jackscrews, available in several sizes; a circuit-board-mounting mass-termination connector with its mating female ribbon connector.

Among the other connectors for use with coaxial cable are the TNC (a close cousin of the BNC, but with threaded outer shell), the high-performance but bulky type N, the miniature SMA, the subminiature LEMO and SMC, and the MHV, a high-voltage version of the standard BNC connector. The so-called phono jack used in audio equipment is a nice lesson in bad design, because the inner conductor mates before the shield when you plug it in; furthermore, the design of the connector is such that both shield and center conductor tend to make poor contact. You've undoubtedly *heard* the results! Not to be outdone, the television industry has responded with its own bad standard, the type F coax "connector," which uses the unsupported inner wire of the coax as the pin of the male plug, and a shoddy arrangement to mate the shield.

Multipin connectors. Very frequently electronic instruments demand multiwire cables and connectors. There are literally dozens of different kinds. The simplest example is a 3-wire line cord connector. Among the more popular are the excellent type D subminiature, the Winchester MRA series, the venerable MS type, and the flat ribbon-cable mass-termination connectors (Fig. 1.103).

Beware of connectors that can't tolerate being dropped on the floor (the miniature hexagon connectors are classic) or that don't provide a secure locking mechanism (e.g., the Jones 300 series).

Card-edge connectors. The most common method used to make connection to printed-circuit cards is the card-edge connector, which mates to a row of gold-plated contacts at the edge of the card. Card-edge connectors may have from 15 to 100 pins, and they come with different lug styles according to the method of connection. You can solder them to a "motherboard" or "backplane," which is itself just another printed-circuit board containing the interconnecting wiring between the individual circuit cards. Alternatively, you may want to use edge connectors with standard solder-lug terminations, particularly in a system with only a few cards (see Chapter 12 for some photographs).

1.33 Indicators

Meters

To read out the value of some voltage or current, you have a choice between the time-honored moving-pointer type of meter and digital-readout meters. The latter are more expensive and more accurate. Both types are available in a variety of voltage and current ranges. There are, in addition, exotic panel meters that read out such things as VU (volume units, an audio dB scale), expanded-scale ac volts (e.g., 105 to 130 volts), temperature (from a thermocouple), percentage motor load, frequency, etc. Digital panel meters often provide the option of logic-level outputs, in addition to the visible display, for internal use by the instrument.

Lamps and LEDs

Flashing lights, screens full of numbers and letters, eerie sounds – these are the stuff of science fiction movies, and except for the latter, they form the subject of lamps and displays (see Section 9.10). Small incandescent lamps used to be standard for front-panel indicators, but they have been replaced by light-emitting diodes (LEDs). The latter behave electrically like ordinary diodes, but with a forward voltage drop in the range of 1.5 to 2.5 volts. When current flows in the forward direction, they light up. Typically, 5mA to 20mA produces adequate brightness. LEDs are cheaper than incandescent lamps, they last forever, and they are even available in three colors (red, yellow, and green). They come in convenient panel-mounting packages; some even provide built-in current limiting.

LEDs are also used for digital displays, most often the familiar 7-segment numeric display you see in calculators. For displaying letters as well as numbers (alphanumeric display), you can get 16-segment displays or dot-matrix displays. For low power or outdoor use, liquid-crystal displays are superior.

1.34 Variable components

Resistors

Variable resistors (also called volume controls, potentiometers, pots, or trimmers) are useful as panel controls or internal adjustments in circuits. The most common panel type is known as a 2 watt type AB potentiometer; it uses the same basic material as the fixed carbon-composition resistor, with a rotatable "wiper" contact. Other panel types are available with ceramic or plastic resistance elements, with improved characteristics. Multiturn types (3, 5, or 10 turns) are available, with counting dials, for improved resolution and linearity. "Ganged" pots (several independent sections on one shaft) are also manufactured, although in limited variety, for applications that demand them.

For use inside an instrument, rather than on the front panel, *trimmer pots* come in single-turn and multiturn styles, most intended for printed-circuit mounting. These are handy for calibration adjustments of the "set-and-forget" type. Good advice: Resist the temptation to use lots of trimmers in your circuits. Use good design instead.

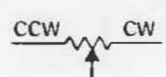


Figure 1.104. Potentiometer (3-terminal variable resistor).

The symbol for a variable resistor, or pot, is shown in Figure 1.104. Sometimes the symbols CW and CCW are used to indicate clockwise and counterclockwise.

One important point about variable resistors: Don't attempt to use a potentiometer as a substitute for a precise resistor value somewhere within a circuit. This is tempting, because you can trim the resistance to the value you want. The trouble is that potentiometers are not as stable as good (1%) resistors, and in addition they may not have good resolution (i.e., they can't be set to a precise value). If you must have a precise and settable resistor value somewhere, use a combination of a 1% (or better) precision resistor and a potentiometer, with the fixed resistor contributing most of the resistance. For example, if you need a 23.4k resistor, use a 22.6k (a 1% value) 1% fixed resistor in series with a 2k trimmer pot. Another possibility is to use a series combination of several precision resistors, choosing the last (and smallest) resistor to give the desired series resistance.

As you will see later, it is possible to use FETs as voltage-controlled variable resistors in some applications. Transistors can be used as variable-gain amplifiers, again controlled by a voltage. Keep an open mind when design brainstorming.

Figure 1.105. Variable capacitor.

Capacitors

Variable capacitors are primarily confined to the smaller capacitance values (up to about 1000pF) and are commonly used in radiofrequency (RF) circuits. Trimmers are available for in-circuit adjustments, in addition to the panel type for user tuning. Figure 1.105 shows the symbol for a variable capacitor.

Diodes operated with applied reverse voltage can be used as voltage-variable capacitors; in this application they're called varactors, or sometimes varicaps or epicaps. They're very important in RF applications, especially automatic frequency control (AFC), modulators, and parametric amplifiers.

Inductors

Variable inductors are usually made by arranging to move a piece of core material in a fixed coil. In this form they're available with inductances ranging from microhenrys to henrys, typically with a 2:1 tuning range for any given inductor. Also available are rotary inductors (coreless coils with a rolling contact).

Transformers

Variable transformers are very handy devices, especially the ones operated from the 115 volt ac line. They're usually "autoformers," which means that they have only one winding, with a sliding contact. They're also commonly called Variacs, and they are made by Technipower, Superior Electric, and others. Typically they provide 0 to 135 volts ac output when operated from 115 volts, and they come in current ratings from 1 amp to 20 amps or more. They're good for testing instruments that seem to be affected by powerline variations, and in any case to verify worst-case performance. Warning: Don't forget that the output is not electrically isolated from the power line, as it would be with a transformer!

ADDITIONAL EXERCISES

(1) Find the Norton equivalent circuit (a current source in parallel with a resistor) for the voltage divider in Figure 1.106. Show that the Norton equivalent gives the same output voltage as the actual circuit when loaded by a 5k resistor.

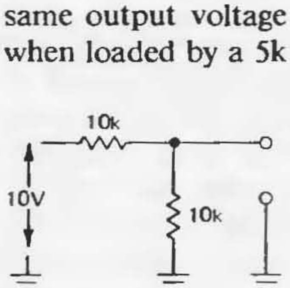


Figure 1.106

(2) Find the Thévenin equivalent for the circuit shown in Figure 1.107. Is it the

same as the Thévenin equivalent for exercise 1?

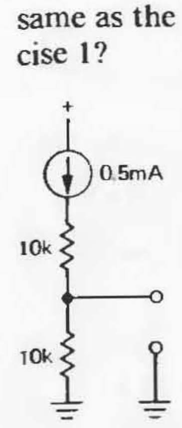


Figure 1.107

(3) Design a "rumble filter" for audio. It should pass frequencies greater than 20Hz (set the −3dB point at 10Hz). Assume zero source impedance (perfect voltage source) and 10k (minimum) load impedance (that's important so that you can choose R and C such that the load doesn't affect the filter operation significantly).

(4) Design a "scratch filter" for audio signals (3dB down at 10kHz). Use the same source and load impedances as in exercise 3.

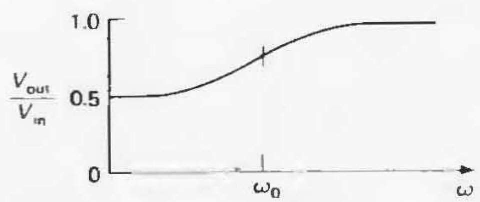


Figure 1.108

(5) How would you make a filter with Rs and Cs to give the response shown in Figure 1.108?

(6) Design a bandpass RC filter (as in Fig. 1.109); f_1 and f_2 are the 3dB points. Choose impedances so that the first stage isn't much affected by the loading of the second stage.

(7) Sketch the output for the circuit shown in Figure 1.110.

(8) Design an oscilloscope "×10 probe"

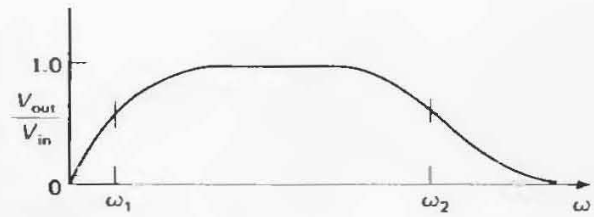


Figure 1.109

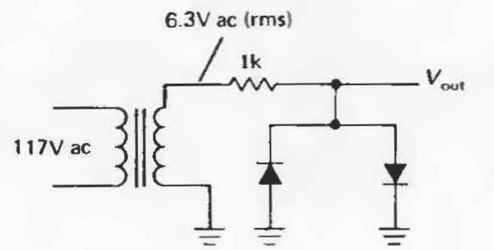


Figure 1.110

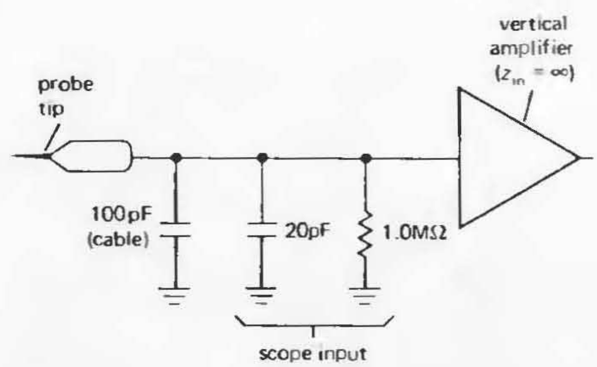


Figure 1.111

(see Appendix A) to use with a scope whose input impedance is 1MΩ in parallel with 20pF. Assume that the probe cable adds an additional 100pF and that the probe components are placed at the tip end (rather than at the scope end) of the cable (Fig. 1.111). The resultant network should have 20dB (×10) attenuation at all frequencies, including dc. The reason for using a ×10 probe is to increase the load impedance seen by the circuit under test, which reduces loading effects. What input impedance (R in parallel with C) does your ×10 probe present to the circuit under test, when used with the scope?