THE ART OF ELECTRONICS

Second Edition

Paul Horowitz HARVARD UNIVERSITY

Winfield Hill ROWLAND INSTITUTE FOR SCIENCE, CAMBRIDGE, MASSACHUSETTS

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

CAMBRIDGE UNIVERSITY PRESS

Cambridge

New York Port Chester Melbourne Sydney

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.

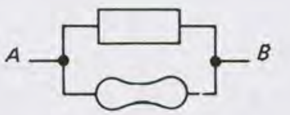

Figure 1.1

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).

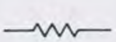

Figure 1.2

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 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.

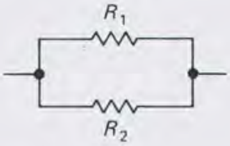

Figure 1.4

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 ($1k\Omega$). Thus, a $10k\Omega$ resistor is often referred to as a 10k resistor, and a $1M\Omega$ 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 ±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 ℧). 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 ($\sigma = 6 \times \cancel{10^{12}}$ W/°K^{4}cm^2). $\sigma \simeq 6e{-12}$ W/K·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.

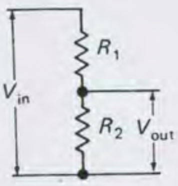

Figure 1.5. Voltage divider. An applied voltage V_{in} results in a (smaller) output voltage V_{out}.

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

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

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

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

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.

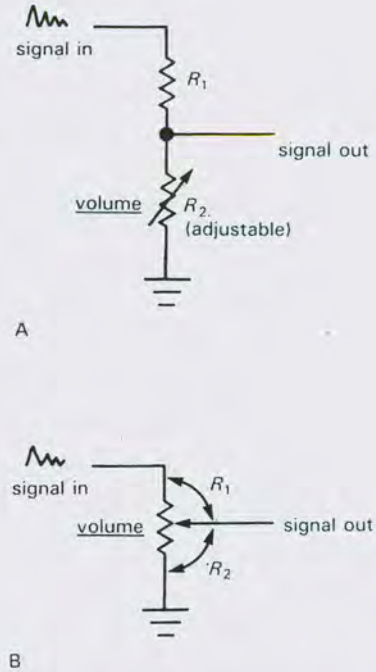

Figure 1.6. An adjustable voltage divider can be made from a fixed and variable resistor, or from a potentiometer.

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_1R_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.

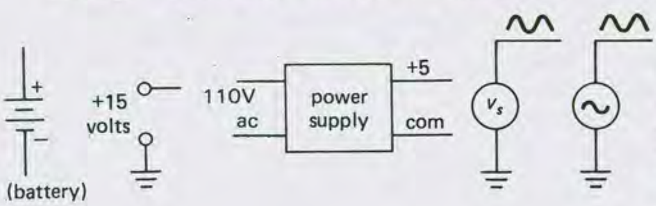

Figure 1.7. Voltage sources can be either steady (dc) or varying (ac).

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.

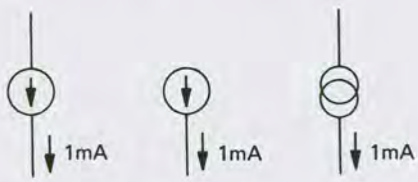

Figure 1.8. Current-source symbols.

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.

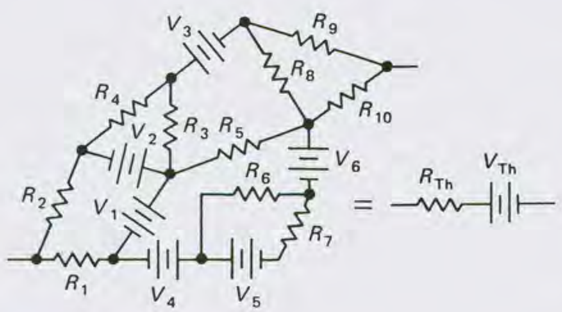

Figure 1.9

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 $V_{\mathrm{Th}}/R_{\mathrm{Th}}$. In other words,

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

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

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

1. The open-circuit voltage is

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

2. The short-circuit current is

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

So the Thévenin equivalent circuit is a voltage source

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

in series with a resistor

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

(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.)

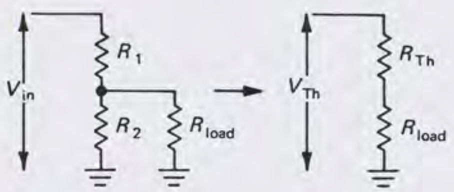

Figure 1.10

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 $V_{\mathrm{in}} = 30\mathrm{V}$ and $R_1 = R_2 = 10\mathrm{k}$, 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.

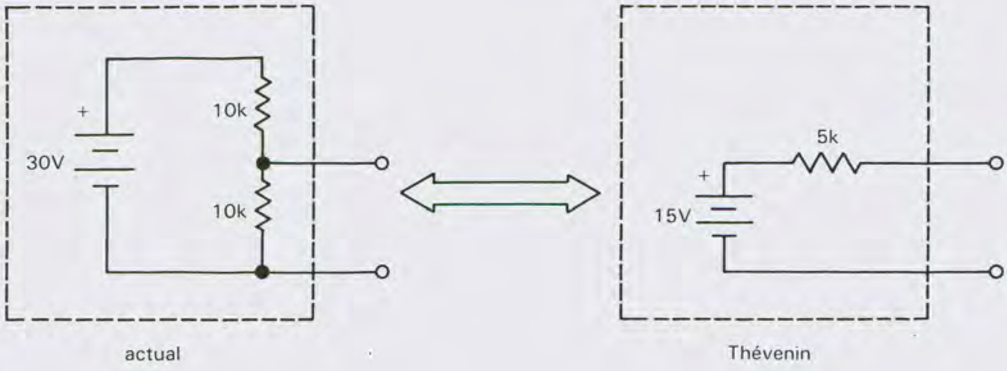

Figure 1.11

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 $R_{\text{load}} \gg R_{\text{internal}}$, 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

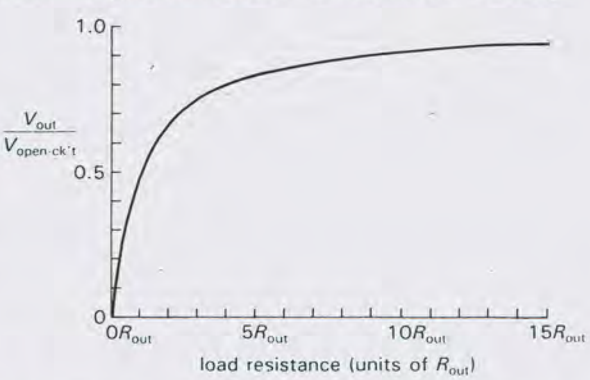

Figure 1.12. To avoid attenuating a signal source below its open-circuit voltage, keep the load resistance large compared with the output resistance.

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.

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:

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

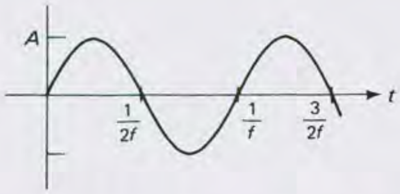

Figure 1.17. Sine wave of amplitude A and frequency f.

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 $V_{rms} = (1/\sqrt{2})A = 0.707A$ (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

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

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:

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

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).

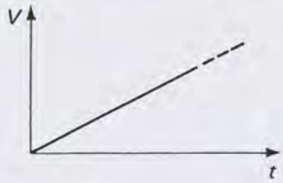

Figure 1.18. Voltage ramp waveform.

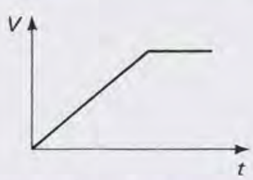

Figure 1.19. Ramp with limit.

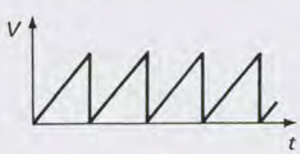

Figure 1.20. Sawtooth wave.

Triangle

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

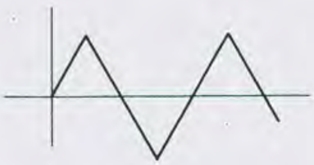

Figure 1.21. Triangle wave.

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.

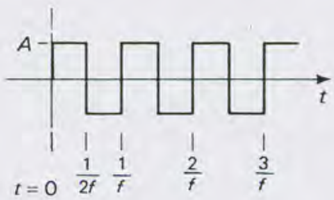

Figure 1.23. Square wave.

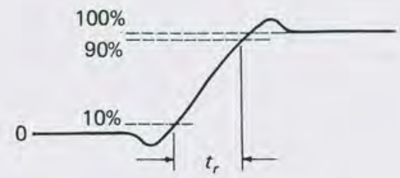

Figure 1.24. Rise time of a step waveform.

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.

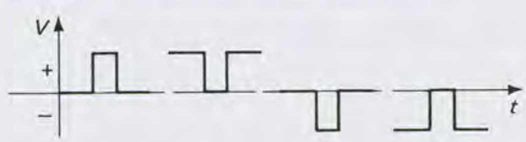

Figure 1.25. Positive- and negative-going pulses of both polarities.

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.

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.

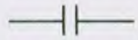

Figure 1.27. Capacitor.

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

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

or

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

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

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

or (two capacitors only)

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

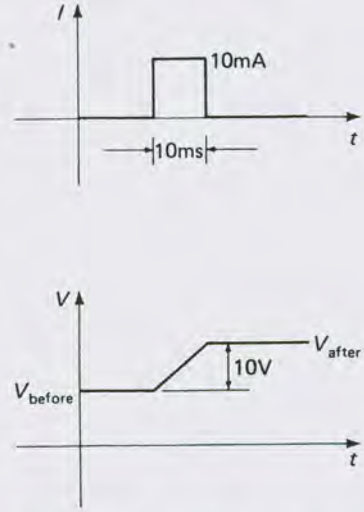

Figure 1.28. The voltage across a capacitor changes when a current flows through it.

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.

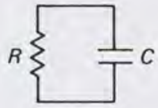

Figure 1.29

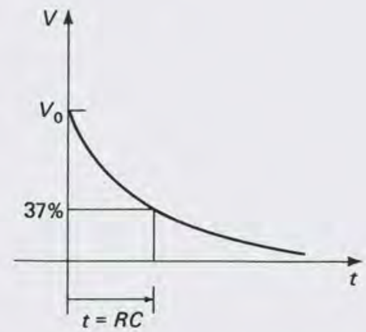

Figure 1.30. *RC* discharge waveform.

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.

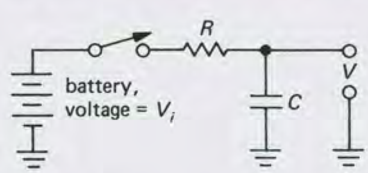

Figure 1.31

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

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

rise time ~ 2 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

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

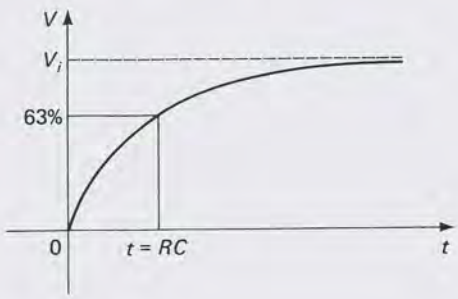

Figure 1.32

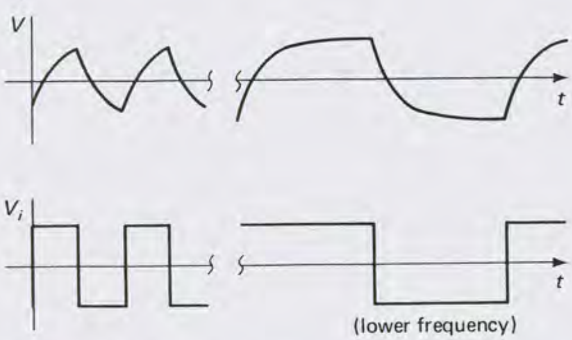

Figure 1.33. Output (top waveform) across a capacitor, when driven by square waves through a resistor.

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.

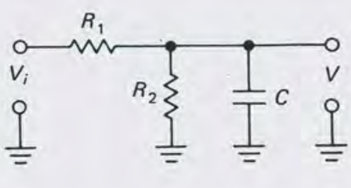

Figure 1.34

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 on RC reaching 50% output = 0.7 RC

$V(t)$ 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.

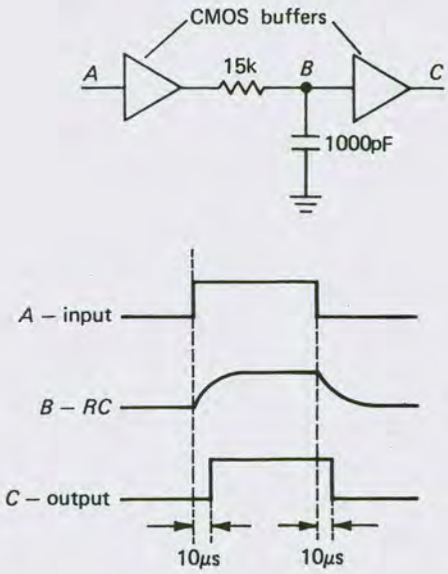

Figure 1.35. Producing a delayed digital waveform with the help of an RC.

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.

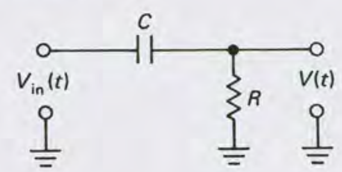

Figure 1.36

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.

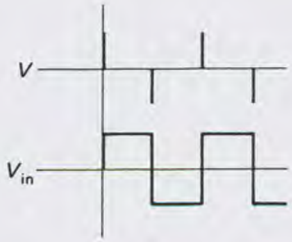

Figure 1.37. Output waveform (top) from differentiator driven by a square wave.

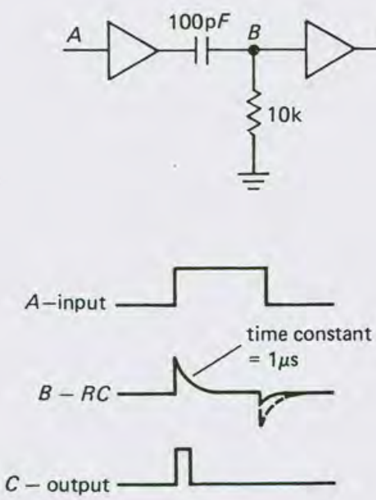

Figure 1.38. Leading-edge detector.

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

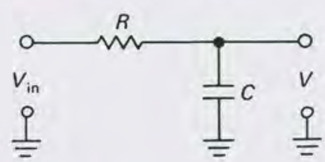

Figure 1.40

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

explained later in chapter

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.

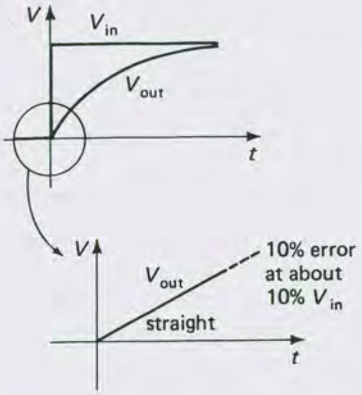

Figure 1.41

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.