

Paul Horowitz HARVARD UNIVERSITY

Winfield Hill ROWLAND INSTITUTE FOR SCIENCE, CAMBRIDGE, MASSACHUSETTS



Published by the Press Syndicate of the University of Cambridge The Pitt Building, Trumpington Street, Cambridge CB2 1RP 40 West 20th Street, New York, NY 10011-4211, USA 10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1980, 1989

First published 1980 Second edition 1989 Reprinted 1990 (twice), 1991, 1993, 1994 (twice), 1995

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data is available.

A catalogue record for this book is available from the British Library.

ISBN 0-521-37095-7 hardback

TO (

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:

$$-\stackrel{R_1}{\sim} \stackrel{R_2}{\sim}$$

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.

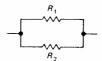


Figure 1.4

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

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

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/nth 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.

mi na iku man nagutiya a inggatunga kati katiwan kilipi. 4 Pilikanasan na kung telah a labu inggatung ni katikan ingkatikan

Sc Ther to w othe signi of ir matt shou (a) tl prec para say, facto circu ished pone You' quic appr rathe pop lr resis thin curr G b I =a lar large appl para resis acro is t The the G =sam deri E cal u of c also

back

the

deve

mos

inste

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 U). 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 $(\sigma=6\times10^{-12} \mathrm{W})^{\circ}\mathrm{K}^4\mathrm{cm}^2)$.

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 However, with just resistors chapter. 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 $V_{\rm in}$ results in a (smaller) output voltage $V_{\rm out}$.

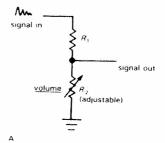
What is $V_{\rm out}$? Well, the current (same everywhere, assuming no "load" on the output) is

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

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

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

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.



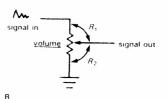


Figure 1.6. An adjustable voltage divider can be made from a fixed and variable resistor, or from a potentiometer.

Voltage dividers are often used in circuits to generate a particular voltage from a larger fixed (or varying) voltage. For instance, if $V_{\rm in}$ 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 voltagedivider 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

A pe blaci drog load that whei term only addi volta serie resis stand like a with a ma amps in a f an o circu term: mear

 $\frac{\downarrow_{+}}{\downarrow_{-}} +1$ \downarrow_{-} vol (battery)

nothi circu

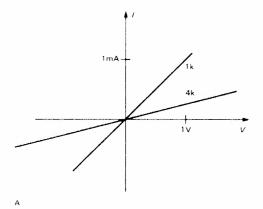
outpi

volta

Figure (dc) o

A termi consta circui

MULT There oscillo or moi trouble



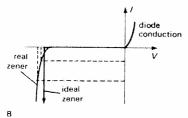


Figure 1.13. *V-I* curves. A. Resistor (linear). B. Zener diode (nonlinear).

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,

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

and

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

SO

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

and finally

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

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 $\Delta V = R_{\rm dyn} \Delta I$, 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, $R_{\rm incr} = \infty$ (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

10mA 5mA

Figur

/sig (r

V_{bart} =

Figur

a tun (Fig. volta

 $v_{
m out}$:

where

the ti

and t small callin this v The t $\Delta V/$ from If $r_{t(zero, provibring)}$

A vices: late i soluti lems. suppo instar

negat neces

devic

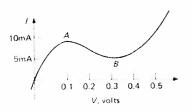


Figure 1.15

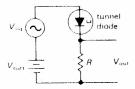


Figure 1.16

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

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

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

$$\Delta V/\Delta I$$
 (or v/i) < 0

from A to B on the characteristic curve. If $r_{t(\text{iner})} \approx R$, 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:

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

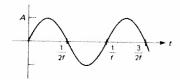


Figure 1.17. Sine wave of amplitude A and frequency f.

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-topeak 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 $V_{\rm rms} = (1/\sqrt{2})A =$ 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

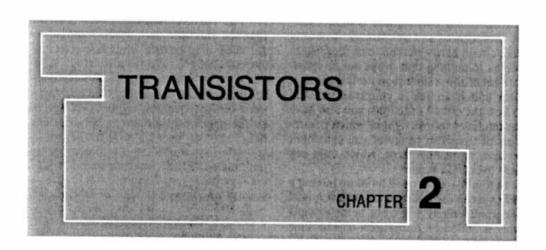
to it 10 ti tenth expra powe

dB =

wher

the to

have wave result form: the d ampl subst Alı to sp are so sure is the signa. other There (whic are u refere (b) d 1mW which 50 oh (the when then (c) th a resi prisin In ad ampli other 0dB S 0.000centin in con dBrn(in free



INTRODUCTION

The transistor is our most important example of an "active" component, a device that can amplify, producing an output signal with more power in it than the input signal. The additional power comes from an external source of power (the power supply, to be exact). Note that voltage amplification isn't what matters, since, for example, a step-up transformer, a "passive" component just like a resistor or capacitor, has voltage gain but no power gain. Devices with power gain are distinguishable by their ability to make oscillators, by feeding some output signal back into the input.

It is interesting to note that the property of power amplification seemed very important to the inventors of the transistor. Almost the first thing they did to convince themselves that they had really invented something was to power a loud-speaker from a transistor, observing that the output signal sounded louder than the input signal.

The transistor is the essential ingredient of every electronic circuit, from the

simplest amplifier or oscillator to the most elaborate digital computer. Integrated circuits (ICs), which have largely replaced circuits constructed from discrete transistors, are themselves merely arrays of transistors and other components built from a single chip of semiconductor material.

A good understanding of transistors is very important, even if most of your circuits are made from ICs, because you need to understand the input and output properties of the IC in order to connect it to the rest of your circuit and to the outside world. In addition, the transistor is the single most powerful resource for interfacing, whether between ICs and other circuitry or between one subcircuit and another. Finally, there are frequent (some might say too frequent) situations where the right IC just doesn't exist, and you have to rely on discrete transistor circuitry to do the job. As you will see, transistors have an excitement all their own. Learning how they work can be great fun.

Our treatment of transistors is going to be quite different from that of many other books. It is common practice to use the h-parameter model and equivalent

circuit. In our opinion that is unnecessarily complicated and unintuitive. Not only does circuit behavior tend to be revealed to you as something that drops out of elaborate equations, rather than deriving from a clear understanding in your own mind as to how the circuit functions; you also have the tendency to lose sight of which parameters of transistor behavior you can count on and, more important, which ones can vary over large ranges.

In this chapter we will build up instead a very simple introductory transistor model and immediately work out some circuits with it. Soon its limitations will become apparent; then we will expand the model to include the respected Ebers-Moll conventions. With the Ebers-Moll equations and a simple 3-terminal model, you will have a good understanding of transistors; you won't need to do a lot of calculations, and your designs will be first-rate. In particular, they will be largely independent of the poorly controlled transistor parameters such as current gain.

Some important engineering notation should be mentioned. Voltage at a transistor terminal (relative to ground) is indicated by a single subscript (C, B, or E): V_C is the collector voltage, for instance. Voltage between two terminals is indicated by a double subscript: V_{BE} is the base-to-emitter voltage drop, for instance. If the same letter is repeated, that means a power-supply voltage: V_{CC} is the (positive) power-supply voltage associated with the collector, and V_{EE} is the (negative) supply voltage associated with the

2.01 First transistor model: current amplifier

Let's begin. A transistor is a 3-terminal device (Fig. 2.1) available in 2 flavors (npn and pnp), with properties that meet the following rules for npn transistors (for pnp simply reverse all polarities):

- I. The collector must be more positive than the emitter.
- 2. The base-emitter and base-collector circuits behave like diodes (Fig. 2.2). Normally the base-emitter diode is conducting and the base-collector diode is reverse-biased, i.e., the applied voltage is in the opposite direction to easy current flow.

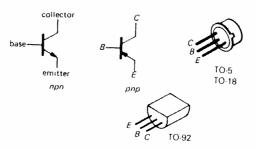


Figure 2.1. Transistor symbols, and small transistor packages.

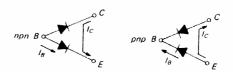


Figure 2.2. An ohmmeter's view of a transistor's terminals.

- 3. Any given transistor has maximum values of I_C , I_B , and V_{CE} that cannot be exceeded without costing the exceeder the price of a new transistor (for typical values, see Table 2.1). There are also other limits, such as power dissipation $(I_C V_{CE})$, temperature, V_{BE} , etc., that you must keep in mind.
- 4. When rules 1-3 are obeyed, I_C is roughly proportional to I_B and can be written as

$$I_C = h_{FE}I_B = \beta I_B$$

where h_{FE} , the current gain (also called beta), is typically about 100. Both I_C and I_E flow to the emitter. Note: The collector current is not due to forward conduction of the base-collector diode;

that diode is it as "transis"

Property 4 ness: A smal controls a muthe collector.

Warning: parameter; for from 50 to 2 given transis the collector voltage, and depends on a bad circuit

Note part This means across the ban enormous is more position about drop). This erating trans $(V_B = V_E)$ normally given for pn

Let us en not try to as diode co collector-ba applied acri Furthermor little with c a not-too-gr ward diode rises very r

SOME BAS

2.02 Trans

Look at the plication, i enables a m other circu From the I derstand.

that diode is reverse-biased. Just think of it as "transistor action."

Property 4 gives the transistor its usefulness: A small current flowing into the base controls a much larger current flowing into the collector.

Warning: h_{FE} is not a "good" transistor parameter; for instance, its value can vary from 50 to 250 for different specimens of a given transistor type. It also depends upon the collector current, collector-to-emitter voltage, and temperature. A circuit that depends on a particular value for h_{FE} is a bad circuit.

Note particularly the effect of property 2. This means you can't go sticking a voltage across the base-emitter terminals, because an enormous current will flow if the base is more positive than the emitter by more than about 0.6 to 0.8 volt (forward diode drop). This rule also implies that an operating transistor has $V_B \approx V_E + 0.6$ volt ($V_B = V_E + V_{BE}$). Again, polarities are normally given for *npn* transistors; reverse them for *pnp*.

Let us emphasize again that you should not try to think of the collector current as diode conduction. It isn't, because the collector-base diode normally has voltages applied across it in the reverse direction. Furthermore, collector current varies very little with collector voltage (it behaves like a not-too-great current source), unlike forward diode conduction, where the current rises very rapidly with applied voltage.

SOME BASIC TRANSISTOR CIRCUITS

2.02 Transistor switch

Look at the circuit in Figure 2.3. This application, in which a small control current enables a much larger current to flow in another circuit, is called a transistor switch. From the preceding rules it is easy to understand. When the mechanical switch is open, there is no base current. So, from

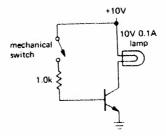


Figure 2.3. Transistor switch example.

rule 4, there is no collector current. The lamp is off.

When the switch is closed, the base rises to 0.6 volt (base-emitter diode is in forward conduction). The drop across the base resistor is 9.4 volts, so the base current is 9.4mA. Blind application of rule 4 gives $I_C = 940 \text{mA}$ (for a typical beta of 100). That is wrong. Why? Because rule 4 holds only if rule 1 is obeyed; at a collector current of 100mA the lamp has 10 volts across it. To get a higher current you would have to pull the collector below ground. A transistor can't do this, and the result is what's called saturation - the collector goes as close to ground as it can (typical saturation voltages are about 0.05-0.2V, see Appendix G) and stays there. In this case, the lamp goes on, with its rated 10 volts across it.

Overdriving the base (we used 9.4mA when 1.0mA would have barely sufficed) makes the circuit conservative; in this particular case it is a good idea, since a lamp draws more current when cold (the resistance of a lamp when cold is 5 to 10 times lower than its resistance at operating current). Also transistor beta drops at low collector-to-base voltages, so some extra base current is necessary to bring a transistor into full saturation (see Appendix G). Incidentally, in a real circuit you would probably put a resistor from base to ground (perhaps 10k in this case) to make sure the base is at ground with the switch open. It wouldn't affect the "on" operation, because it would sink only 0.06mA from the base circuit.

There are certain cautions to be observed when designing transistor switches: I. Choose the base resistor conservatively to get plenty of excess base current, especially when driving lamps, because of the reduced beta at low V_{CE} . This is also a good idea for high-speed switching, because of capacitive effects and reduced beta at very high frequencies (many megahertz). A small "speedup" capacitor is often connected across the base resistor to improve high-speed performance.

- 2. If the load swings below ground for some reason (e.g., it is driven from ac, or it is inductive), use a diode in series with the collector (or a diode in the reverse direction to ground) to prevent collector-base conduction on negative swings.
- 3. For inductive loads, protect the transistor with a diode across the load, as shown in Figure 2.4. Without the diode the inductor will swing the collector to a large positive voltage when the switch is opened, most likely exceeding the collector-emitter breakdown voltage, as the inductor tries to maintain its "on" current from V_{CC} to the collector (see the discussion of inductors in Section 1.31).

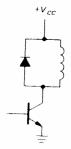


Figure 2.4. Always use a suppression diode when switching an inductive load.

Transistor switches enable you to switch very rapidly, typically in a small fraction of a microsecond. Also, you can switch many different circuits with a single control signal. One further advantage is the possibility of remote cold switching, in which only dc control voltages snake around through cables to reach front-panel switches, rather than the electronically inferior approach of having the signals themselves traveling through cables and switches (if you run lots of signals through cables, you're likely to get capacitive pickup as well as some signal degradation).

"Transistor man"

Figure 2.5 presents a cartoon that will help you understand some limits of transistor

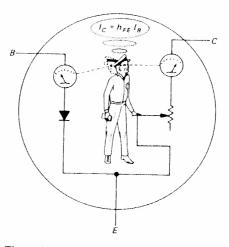


Figure 2.5. "Transistor man" observes the base current, and adjusts the output rheostat in an attempt to maintain the output current h_{FE} times larger.

behavior. The little man's perpetual task in life is to try to keep $I_C = h_{FE}I_B$; however, he is only allowed to turn the knob on the variable resistor. Thus he can go from a short circuit (saturation) to an open circuit (transistor in the "off" state), or anything in between, but he isn't allowed to use batteries, current sources, etc. One warning is in order here: Don't think that the collector of a transistor

looks lik it looks a constantdependir base), pri efforts.

Anoth at any gi cut off (active reg collector of a volt saturation a volt of transistor

2.03 Emi

Figure 2.6 follower. put termi the input $V_E \approx V_B$ The output to 0.7 vol $V_{\rm in}$ must else the creturning supply vo voltage sw no collected



Figure 2.6.

At first useless, un impedance put impeda

looks like a resistor. It doesn't. Rather, it looks approximately like a poor-quality constant-current sink (the value of current depending on the signal applied to the base), primarily because of this little man's efforts.

Another thing to keep in mind is that, at any given time, a transistor may be (a) cut off (no collector current), (b) in the active region (some collector current, and collector voltage more than a few tenths of a volt above the emitter), or (c) in saturation (collector within a few tenths of a volt of the emitter). See Appendix G on transistor saturation for more details.

2.03 Emitter follower

Figure 2.6 shows an example of an *emitter* follower. It is called that because the output terminal is the emitter, which follows the input (the base), less one diode drop:

$$V_E \approx V_B - 0.6 \text{ volt}$$

The output is a replica of the input, but 0.6 to 0.7 volt less positive. For this circuit, $V_{\rm in}$ must stay at +0.6 volt or more, or else the output will sit at ground. By returning the emitter resistor to a negative supply voltage, you can permit negative voltage swings as well. Note that there is no collector resistor in an emitter follower.

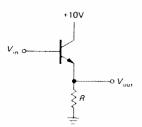


Figure 2.6. Emitter follower.

At first glance this circuit may appear useless, until you realize that the input impedance is much larger than the output impedance, as will be demonstrated shortly. This means that the circuit requires less power from the signal source to drive a given load than would be the case if the signal source were to drive the load directly. Or a signal of some internal impedance (in the Thévenin sense) can now drive a load of comparable or even lower impedance without loss of amplitude (from the usual voltage-divider effect). In other words, an emitter follower has current gain, even though it has no voltage gain. It has power gain. Voltage gain isn't everything!

Impedances of sources and loads

This last point is very important and is worth some more discussion before we calculate in detail the beneficial effects of emitter followers. In electronic circuits, you're always hooking the output of something to the input of something else, as suggested in Figure 2.7. The signal source might be the output of an amplifier stage (with Thévenin equivalent series impedance Zout), driving the next stage or perhaps a load (of some input impedance Z_{in}). In general, the loading effect of the following stage causes a reduction of signal, as we discussed earlier in Section 1.05. For this reason it is usually best to keep $Z_{
m out} \ll Z_{
m in}$ (a factor of 10 is a comfortable rule of thumb).

In some situations it is OK to forgo this general goal of making the source stiff compared with the load. In particular, if the load is always connected (e.g., within a circuit) and if it presents a known and constant $\mathbf{Z}_{\rm in}$, it is not too serious if it "loads" the source. However, it is always nicer if signal levels don't change when a load is connected. Also, if $\mathbf{Z}_{\rm in}$ varies with signal level, then having a stiff source $(Z_{\rm out} \ll Z_{\rm in})$ assures linearity, where otherwise the level-dependent voltage divider would cause distortion.

Finally, there are two situations where $Z_{\rm out} \ll Z_{\rm in}$ is actually the wrong thing to

first amplifier

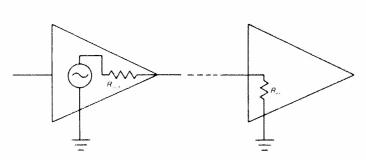


Figure 2.7. Illustrating circuit "loading" as a voltage divider.

do: In radiofrequency circuits we usually match impedances ($Z_{\rm out} = Z_{\rm in}$), for reasons we'll describe in Chapter 14. A second exception applies if the signal being coupled is a current rather than a voltage. In that case the situation is reversed, and one strives to make $Z_{\rm in} \ll Z_{\rm out}$ ($Z_{\rm out} = \infty$, for a current source).

Input and output impedances of emitter followers

As you have just seen, the emitter follower is useful for changing impedances of signals or loads. To put it bluntly, that's the whole point of an emitter follower.

Let's calculate the input and output impedances of the emitter follower. In the preceding circuit we will consider R to be the load (in practice it sometimes is the load; otherwise the load is in parallel with R, but with R dominating the parallel resistance anyway). Make a voltage change ΔV_B at the base; the corresponding change at the emitter is $\Delta V_E = \Delta V_B$. Then the change in emitter current is

$$\Delta I_E = \Delta V_B/R$$
 so
$$\Delta I_B = \frac{1}{h_{fe}+1} \Delta I_E = \frac{\Delta V_B}{R(h_{fe}+1)}$$

(using $I_E = I_C + I_B$). The input resistance is $\Delta V_B/\Delta I_B$. Therefore

$$r_{\rm in} = (h_{fe} + 1)R$$

second amplifier

The transistor beta (h_{fe}) is typically about 100, so a low-impedance load looks like a much higher impedance at the base; it is easier to drive.

In the preceding calculation, as in Chapter 1, we have used lower-case symbols such as h_{fe} to signify small-signal (incremental) quantities. Frequently one concentrates on the changes in voltages (or currents) in a circuit, rather than the steady (dc) values of those voltages (or currents). This is most common when these "small-signal" variations represent a possible signal, as in an audio amplifier, riding on a steady dc "bias" (see Section 2.05). The distinction between dc current gain (h_{FE}) and small-signal current gain (h_{fe}) isn't always made clear, and the term beta is used for both. That's alright, since $h_{fe} \approx h_{FE}$ (except at very high frequencies), and you never assume you know them accurately, anyway.

Although we used resistances in the preceding derivation, we could generalize to complex impedances by allowing ΔV_B , ΔI_B , etc., to become complex numbers. We would find that the same

transforma ances: Z_{in} We cou find that the emitter fo into the er

$$\mathbf{Z}_{\mathrm{out}} = \frac{\mathbf{Z}_{\mathrm{s}}}{h_f}$$

internal in

Strictly spethe circuit resistance impedance inates.

EXERCI Show that tl Hint: Hold the change in output vo voltage is series resis

Because ter follow situations, nal source puts), mak higher-improvoltage div ing signal of subsequ

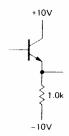


Figure 2.8. plenty of cu sink limited resistor.

transformation rule applies for impedances: $\mathbf{Z}_{\text{in}} = (h_{fe} + 1)\mathbf{Z}_{\text{load}}$.

We could do a similar calculation to find that the output impedance \mathbf{Z}_{out} of an emitter follower (the impedance looking into the emitter) driven from a source of internal impedance $\mathbf{Z}_{\text{source}}$ is given by

$$Z_{\text{out}} = \frac{Z_{\text{source}}}{h_{fe} + 1}$$

Strictly speaking, the output impedance of the circuit should also include the parallel resistance of R, but in practice \mathbf{Z}_{out} (the impedance looking into the emitter) dominates.

EXERCISE 2.1

Show that the preceding relationship is correct. Hint: Hold the source voltage fixed, and find the change in output current for a given change in output voltage. Remember that the source voltage is connected to the base through a series resistor.

Because of these nice properties, emitter followers find application in many situations, e.g., making low-impedance signal sources within a circuit (or at outputs), making stiff voltage references from higher-impedance references (formed from voltage dividers, say), and generally isolating signal sources from the loading effects of subsequent stages.

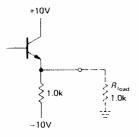


Figure 2.8. An *npn* emitter follower can source plenty of current through the transistor, but can sink limited current only through its emitter resistor.

EXERCISE 2.2

Use a follower with base driven from a voltage divider to provide a stiff source of +5 volts from an available regulated +15 volt supply. Load current (max) = 25mA. Choose your resistor values so that the output voltage doesn't drop more than 5% under full load.

Important points about followers

1. Notice (Section 2.01, rule 4) that in an emitter follower the npn transistor can only "source" current. For instance, in the loaded circuit shown in Figure 2.8 the output can swing to within a transistor saturation voltage drop of V_{CC} (about +9.9V), but it cannot go more negative than -5 volts. That is because on the extreme negative swing, the transistor can do no more than turn off, which it does at -4.4 volts input (-5V output). Further negative swing at the input results in backbiasing of the base-emitter junction, but no further change in output. output, for a 10 volt amplitude sine-wave input, looks as shown in Figure 2.9.

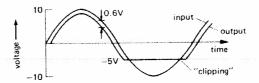


Figure 2.9. Illustrating the asymmetrical current drive capability of the *npn* emitter follower.

Another way to view the problem is to say that the emitter follower has low small-signal output impedance. Its large-signal output impedance is much larger (as large as R_E). The output impedance changes over from its small-signal value to its large-signal value at the point where the transistor goes out of the active region (in this case at an output voltage of -5V). To put this point another way, a low value of small-signal output impedance doesn't

necessarily mean that the circuit can generate large signal swings into a lowresistance load. Low small-signal output impedance doesn't imply large output current capability.

Possible solutions to this problem involve either decreasing the value of the emitter resistor (with greater power dissipation in resistor and transistor). using a pnp transistor (if all signals are negative only), or using a "push-pull" configuration, in which two complementary transistors (one npn, one pnp), are used (Section 2.15). This sort of problem can also come up when the load of an emitter follower contains voltage or current sources of its own. This happens most often with regulated power supplies (the output is usually an emitter follower) driving a circuit that has other power supplies.

2. Always remember that the base-emitter reverse breakdown voltage for silicon transistors is small, quite often as little as 6 volts. Input swings large enough to take the transistor out of conduction can easily result in breakdown (with consequent degradation of h_{FE}) unless a protective diode is added (Fig. 2.10).



Figure 2.10. A diode prevents base-emitter reverse voltage breakdown.

3. The voltage gain of an emitter follower is actually slightly less than 1.0, because the base-emitter voltage drop is not really constant, but depends slightly on collector current. You will see how to handle that later in the chapter, when we have the Ebers-Moll equation.

2.04 Emitter followers as voltage regulators

The simplest regulated supply of voltage is simply a zener (Fig. 2.11). Some current must flow through the zener, so you choose

$$\frac{V_{\rm in} - V_{\rm out}}{R} > I_{\rm out}$$
 (max)

Because $V_{\rm in}$ isn't regulated, you use the lowest value of $V_{\rm in}$ that might occur for this formula. This is called worst-case design. In practice, you would also worry about component tolerances, line-voltage limits, etc., designing to accommodate the worst possible combination that would ever occur.

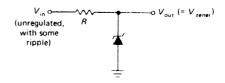


Figure 2.11. Simple zener voltage regulator.

The zener must be able to dissipate

$$P_{\rm zener} = \left(\frac{V_{\rm in} - V_{\rm out}}{R} - I_{\rm out}\right) V_{\rm zener}$$

Again, for worst-case design, you would use $V_{\rm in}$ (max), $R_{\rm min}$, and $I_{\rm out}$ (min).

EXERCISE 2.3

Design a +10 volt regulated supply for load currents from 0 to 100mA; the input voltage is +20 to +25 volts. Allow at least 10mA zener current under all (worst-case) conditions. What power rating must the zener have?

This simple zener-regulated supply is sometimes used for noncritical circuits, or circuits using little supply current. However, it has limited usefulness, for several reasons:

- 1. V_{out} isn't adjustable, or settable to a precise value.
- 2. Zener diodes give only moderate ripple rejection and regulation against changes of

input or l impedance 3. For wi power ze the dissir

By usi the zene shown ir tion is m made rel rent, sinc small, an tion is p $1/h_{FE}$). added to mentary the curren to the en R_C so th less than normal lc

V_{in} ____ (unregulated)

Figure 2.1 for increas transistor b

EXERC Design a + cations as i ter followe in transisto change in : dition to ful circuit.

A nice to elimina (through *i* plying the input or load, owing to their finite dynamic impedance.

3. For widely varying load currents a highpower zener is often necessary to handle the dissipation at low load current.

By using an emitter follower to isolate the zener, you get the improved circuit shown in Figure 2.12. Now the situation is much better. Zener current can be made relatively independent of load current, since the transistor base current is small, and far lower zener power dissipation is possible (reduced by as much as $1/h_{FE}$). The collector resistor R_C can be added to protect the transistor from momentary output short circuits by limiting the current, even though it is not essential to the emitter follower function. Choose R_C so that the voltage drop across it is less than the drop across R for the highest normal load current.

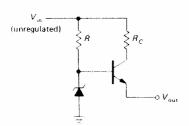


Figure 2.12. Zener regulator with follower, for increased output current. R_C protects the transistor by limiting maximum output current.

EXERCISE 2.4

Design a +10 volt supply with the same specifications as in Exercise 2.3. Use a zener and emitter follower. Calculate worst-case dissipation in transistor and zener. What is the percentage change in zener current from the no-load condition to full load? Compare with your previous circuit.

A nice variation of this circuit aims to eliminate the effect of ripple current (through R) on the zener voltage by supplying the zener current from a current

source, which is the subject of Section 2.06. An alternative method uses a low-pass filter in the zener bias circuit (Fig. 2.13). R is chosen to provide sufficient zener current. Then C is chosen large enough so that $RC \gg 1/f_{\rm ripple}$. (In a variation of this circuit, the upper resistor is replaced by a diode.)

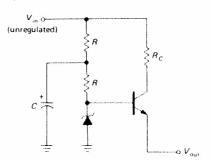


Figure 2.13. Reducing ripple in the zener regulator.

Later you will see better voltage regulators, ones in which you can vary the output easily and continuously, using feedback. They are also better voltage sources, with output impedances measured in milliohms, temperature coefficients of a few parts per million per degree centigrade, etc.

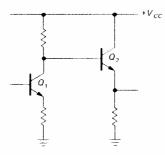


Figure 2.14

2.05 Emitter follower biasing

When an emitter follower is driven from a preceding stage in a circuit, it is usually OK to connect its base directly to the

previous stage's output, as shown in Figure 2.14.

Because the signal on Q_1 's collector is always within the range of the power supplies, Q_2 's base will be between V_{CC} and ground, and therefore Q_2 is in the active region (neither cut off nor saturated), with its base-emitter diode in conduction and its collector at least a few tenths of a volt more positive than its emitter. Sometimes, though, the input to a follower may not be so conveniently situated with respect to the supply voltages. A typical example is a capacitively coupled (or ac-coupled) signal from some external source (e.g., an audio signal input to a high-fidelity amplifier). In that case the signal's average voltage is zero, and direct coupling to an emitter follower will give an output like that in Figure 2.15.

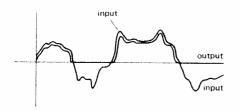


Figure 2.15. A transistor amplifier powered from a single positive supply cannot generate negative voltage swings at the transistor output terminal.

It is necessary to bias the follower (in fact, any transistor amplifier) so that collector current flows during the entire signal swing. In this case a voltage divider is the simplest way (Fig. 2.16). R_1 and R_2 are chosen to put the base halfway between ground and V_{CC} with no input signal, i.e., R_1 and R_2 are approximately equal. The process of selecting the operating voltages in a circuit, in the absence of applied signals, is known as setting the quiescent point. In this case, as in most cases, the quiescent point is chosen to allow maximum symmetrical signal swing

of the output waveform without clipping (flattening of the top or bottom of the waveform). What values should R_1 and R_2 have? Applying our general principle (Section 1.05), we make the impedance of the dc bias source (the impedance looking into the voltage divider) small compared with the load it drives (the dc impedance looking into the base of the follower). In this case,

 $R_1 || R_2 \ll h_{FE} R_E$

This is approximately equivalent to saying that the current flowing in the voltage divider should be large compared with the current drawn by the base.

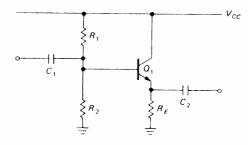


Figure 2.16. An ac-coupled emitter follower. Note base bias voltage divider.

Emitter follower design example

As an actual design example, let's make an emitter follower for audio signals (20Hz to 20kHz). V_{CC} is +15 volts, and quiescent current is to be 1mA.

Step 1. Choose V_E . For the largest possible symmetrical swing without clipping, $V_E = 0.5V_{CC}$, or +7.5 volts.

Step 2. Choose R_E . For a quiescent current of ImA, $R_E = 7.5$ k.

Step 3. Choose R_1 and R_2 . V_B is V_E+ 0.6, or 8.1 volts. This determines the ratio of R_1 to R_2 as 1:1.17. The preceding loading criterion requires that the parallel resistance of R_1 and R_2 be about 75k or less (one-tenth of 7.5k times h_{FE}).

Suitable s $R_2 = 150$ Step 4. C filter with namely th base in pa ing into t assume th is large cor then the i is $h_{FE}R_{E}$ like 70k. about 63k at least 0. be below t 20Hz.

Step 5. (pass filter impedance it is safe to won't be s value for (3dB point now two cathe capaci somewhat (reduction 6dB) at th $C_1 = 0.5\mu$ good choic

Followers :

Because sig is convenie and negati biasing and (Fig. 2.17).

Warning path for ba only to grous assumed path to grous capacitiva resistor to be about or

Suitable standard values are $R_1 = 130$ k, $R_2 = 150$ k.

Step 4. Choose C_1 . C_1 forms a high-pass filter with the impedance it sees as a load, namely the impedance looking into the base in parallel with the impedance looking into the base voltage divider. If we assume that the load this circuit will drive is large compared with the emitter resistor, then the impedance looking into the base is $h_{FE}R_E$, about 750k. The divider looks like 70k. So the capacitor sees a load of about 63k, and it should have a value of at least $0.15\mu F$ so that the 3dB point will be below the lowest frequency of interest, 20 Hz.

Step 5. Choose C_2 . C_2 forms a highpass filter in combination with the load impedance, which is unknown. However, it is safe to assume that the load impedance won't be smaller than R_E , which gives a value for C_2 of at least $1.0\mu F$ to put the 3dB point below 20Hz. Because there are now two cascaded high-pass filter sections, the capacitor values should be increased somewhat to prevent large attenuation (reduction of signal amplitude, in this case 6dB) at the lowest frequency of interest. $C_1 = 0.5\mu F$ and $C_2 = 3.3\mu F$ might be good choices.

Followers with split supplies

Because signals often are "near ground," it is convenient to use symmetrical positive and negative supplies. This simplifies biasing and eliminates coupling capacitors (Fig. 2.17).

Warning: You must always provide a dc path for base bias current, even if it goes only to ground. In the preceding circuit it is assumed that the signal source has a dc path to ground. If not (e.g., if the signal is capacitively coupled), you must provide a resistor to ground (Fig. 2.18). R_B could be about one-tenth of $h_{FE}R_E$, as before.

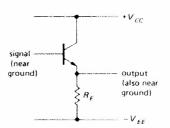


Figure 2.17. A de-coupled emitter follower with split supply.

EXERCISE 2.5

Design an emitter follower with ± 15 volt supplies to operate over the audio range (20Hz-20kHz). Use 5mA quiescent current and capacitive input coupling.

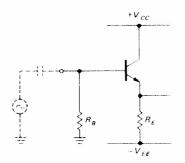


Figure 2.18

Bad biasing

Unfortunately, you sometimes see circuits like the disaster shown in Figure 2.19. R_B was chosen by assuming a particular value for h_{FE} (100), estimating the base current, and then hoping for a 7 volt drop across R_B . This is a bad design; h_{FE} is not a good parameter and will vary considerably. By using voltage biasing with a stiff voltage divider, as in the detailed example presented earlier, the quiescent point is insensitive to variations in transistor beta. For instance, in the previous design example the emitter voltage will increase by only 0.35 volt (5%) for a transistor with $h_{FE}=200$ instead of the nominal

 $h_{FE} = 100$. As with this emitter follower example, it is just as easy to fall into this trap and design bad transistor circuits in the other transistor configurations (e.g., the common-emitter amplifier, which we will treat later in this chapter).

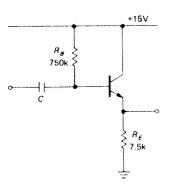


Figure 2.19. Don't do this!

2.06 Transistor current source

Current sources, although often neglected, are as important and as useful as voltage sources. They often provide an excellent way to bias transistors, and they are unequaled as "active loads" for super-gain amplifier stages and as emitter sources for differential amplifiers. Integrators, sawtooth generators, and ramp generators need current sources. They provide widevoltage-range pull-ups within amplifier and regulator circuits. And, finally, there are applications in the outside world that require constant current sources, e.g., electrophoresis or electrochemistry.

Resistor plus voltage source

The simplest approximation to a current source is shown in Figure 2.20. As long as $R_{\rm load} \ll R$ (in other words, $V_{\rm load} \ll V$), the current is nearly constant and is approximately

$$I = V/R$$

The load doesn't have to be resistive. A capacitor will charge at a constant rate, as long as $V_{\rm capacitor} \ll V$; this is just the first part of the exponential charging curve of an RC.

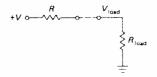


Figure 2.20

There are several drawbacks to a simple resistor current source. In order to make a good approximation to a current source, you must use large voltages, with lots of power dissipation in the resistor. In addition, the current isn't easily programmable, i.e., controllable over a large range via a voltage somewhere else in the circuit.

EXERCISE 2.6

If you want a current source constant to 1% over a load voltage range of 0 to +10 volts, how large a voltage source must you use in series with a single resistor?

EXERCISE 2.7

Suppose you want a 10mA current in the preceding problem. How much power is dissipated in the series resistor? How much gets to the load?

Transistor current source

Fortunately, it is possible to make a very good current source with a transistor (Fig. 2.21). It works like this: Applying V_B to the base, with $V_B > 0.6$ volt, ensures that the emitter is always conducting:

$$V_E = V_B - 0.6 \text{ volt}$$

Sc

$$I_E = V_E/R_E = (V_B - 0.6 \text{ volt})/R_E$$
 But, since $I_E \approx I_C$ for large h_{FE} ,

$$I_C \approx (V_B - 0.6 \text{ volt})/R_E$$

independer tor is not s



Figure 2.21. concept.

Current-sou

The base v number of OK, as lonbefore, the should be m looking into can use a zeor even a discries from emitter supp

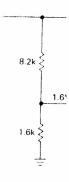


Figure 2.22. transistors singreturned to gre

independent of V_C , as long as the transistor is not saturated ($V_C > V_E + 0.2$ volt).

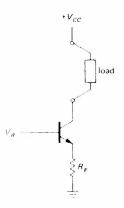


Figure 2.21. Transistor current source: basic concept.

Current-source biasing

The base voltage can be provided in a number of ways. A voltage divider is OK, as long as it is stiff enough. As before, the criterion is that its impedance should be much less than the dc impedance looking into the base $(h_{FE}R_E)$. Or you can use a zener diode, biased from V_{CC} , or even a few forward-biased diodes in series from base to the corresponding emitter supply. Figure 2.22 shows some

examples. In the last example (Fig. 2.22C), a pnp transistor sources current to a load returned to ground. The other examples (using *npn* transistors) should properly be called current sinks, but the usual practice is to call all of them current sources. ["Sink" and "source" simply refer to the direction of current flow: If a circuit supplies (positive) current to a point, it is a source, and vice versa.] In the first circuit, the voltage-divider impedance of ~ 1.3 k is very stiff compared with the impedance looking into the base of about 100k (for $h_{FE} = 100$), so any changes in beta with collector voltage will not much affect the output current by causing the base voltage to change. In the other two circuits the biasing resistors are chosen to provide several milliamps to bring the diodes into conduction.

Compliance

A current source can provide constant current to the load only over some finite range of load voltage. To do otherwise would be equivalent to providing infinite power. The output voltage range over which a current source behaves well is called its output *compliance*. For the preceding transistor current sources, the compliance is set by the requirement that

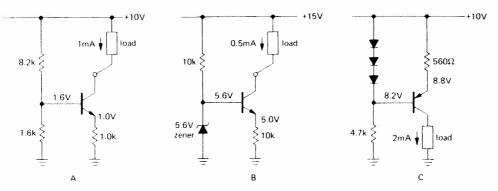


Figure 2.22. Transistor-current-source circuits, illustrating three methods of base biasing; *npn* transistors *sink* current, whereas *pnp* transistors *source* current. The circuit in C illustrates a load returned to ground.

the transistors stay in the active region. Thus in the first circuit the voltage at the collector can go down until the transistor is almost in saturation, perhaps +1.2 volts at the collector. The second circuit, with its higher emitter voltage, can sink current down to a collector voltage of about +5.2 volts.

In all cases the collector voltage can range from a value near saturation all the way up to the supply voltage. For example, the last circuit can source current to the load for any voltage between zero and about +8.6 volts across the load. In fact, the load might even contain batteries or power supplies of its own, carrying the collector beyond the supply voltage. That's OK, but you must watch out for transistor breakdown (V_{CE} must not exceed BV_{CEO} , the specified collector-emitter breakdown voltage) and also for excessive power dissipation (set by I_CV_{CE}). As you will see in Section 6.07, there is an additional safeoperating-area constraint on power transistors.

EXERCISE 2.8

You have +5 and +15 volt regulated supplies available in a circuit. Design a 5mA *npn* current source (sink) using the +5 volts on the base. What is the output compliance?

A current source doesn't have to have a fixed voltage at the base. By varying V_B you get a voltage-programmable current source. The input signal swing $v_{\rm in}$ (remember, lower-case symbols mean variations) must stay small enough so that the emitter voltage never drops to zero, if the output current is to reflect input voltage variations smoothly. The result will be a current source with variations in output current proportional to the variations in input voltage, $i_{\rm out} = v_{\rm in}/R_E$.

□ Deficiencies of current sources

To what extent does this kind of current source depart from the ideal? In

other words, does the load current vary with voltage, i.e., have a finite $(R_{Th} < \infty)$ Thévenin equivalent resistance, and if so why? There are two kinds of effects:

1. Both V_{BE} (Early effect) and h_{FE} vary slightly with collector-to-emitter voltage at a given collector current. The changes in V_{BE} produced by voltage swings across the load cause the output current to change, because the emitter voltage (and therefore the emitter current) changes, even with a fixed applied base voltage. Changes in h_{FE} produce small changes in output (collector) current for fixed emitter current, since $I_C = I_E - I_B$; in addition, there are small changes in applied base voltage produced by the variable loading of the nonzero bias source impedance as h_{FE} (and therefore the base current) changes. These effects are small. For instance, the current from the circuit in Figure 2.22A varied about 0.5% in actual measurements with a 2N3565 transistor. In particular, for load voltages varying from zero to 8 volts, the Early effect contributed 0.5%, and transistor heating effects contributed 0.2%. In addition, variations in h_{FE} contributed 0.05% (note the stiff divider). Thus these variations result in a less-than-perfect current source: The output current depends slightly on voltage and therefore has less than infinite impedance. Later you will see methods that get around this

2. V_{BE} and also h_{FE} depend on temperature. This causes drifts in output current with changes in ambient temperature; in addition, the transistor junction temperature varies as the load voltage is changed (because of variation in transistor dissipation), resulting in departure from ideal current source behavior. The change of V_{BE} with ambient temperature can be compensated with a circuit like that shown in Figure 2.23, in which Q_2 's base-emitter drop is compensated by the drop in emitter follower Q_1 , with similar temperature dependence. R_3 , incidentally, is a

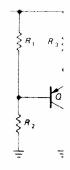


Figure 2.23 compensation

pull-up res sinks curre

☐ Improving

In general. whether ca dence (app: pendence c roughly by can be min voltage to say) so tha millivolts v changes in resistor (re is what is I For instanc applying at cause 10% for 10mV choice V_E 1% current changes. D Remember compliance Using a 5 vc source runr limits the c less than 5 v about $V_E +$ to 10V).



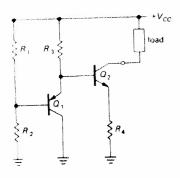


Figure 2.23. One method of temperature-compensating a current source.

pull-up resistor for Q_1 , since Q_2 's base sinks current, which Q_1 cannot source.

I Improving current-source performance

In general, the effects of variability in V_{BE} , whether caused by temperature dependence (approximately -2mV/°C) or by dependence on V_{CE} (the Early effect, given roughly by $\Delta V_{BE} \approx -0.0001 \, \Delta \tilde{V}_{CE}$), can be minimized by choosing the emitter voltage to be large enough (at least 1V, say) so that changes in V_{BE} of tens of millivolts will not result in large fractional changes in the voltage across the emitter resistor (remember that the base voltage is what is held constant by your circuit). For instance, choosing $V_E = 0.1$ volt (i.e., applying about 0.7V to the base) would cause 10% variations in output current for 10mV changes in V_{BE} , whereas the choice $V_E = 1.0$ volt would result in 1% current variations for the same V_{BE} changes. Don't get carried away, though. Remember that the lower limit of output compliance is set by the emitter voltage. Using a 5 volt emitter voltage for a current source running from a +10 volt supply limits the output compliance to slightly less than 5 volts (the collector can go from about V_E + 0.2V to V_{CC} , i.e., from 5.2V to 10V).

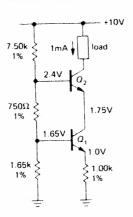


Figure 2.24. Cascode current source for improved current stability with load voltage variations.

Figure 2.24 shows a circuit modification that improves current-source performance significantly. Current source Q_1 functions as before, but with collector voltage held fixed by Q_2 's emitter. The load sees the same current as before, since Q_2 's collector and emitter currents are nearly equal (large h_{FE}). But with this circuit the V_{CE} of Q_1 doesn't change with load voltage, thus eliminating the small changes in V_{BE} from Early effect and dissipationinduced temperature changes. Measurements with 2N3565s gave 0.1% current variation for load voltages from 0 to 8 volts; to obtain performance of this accuracy it is important to use stable 1% resistors, as shown. (Incidentally, this circuit connection also finds use in high-frequency amplifiers, where it is known as the "cascode.") Later you will see current source techniques using op-amps and feedback that circumvent the problem of V_{BE} variation altogether.

The effects of variability of h_{FE} can be minimized by choosing transistors with large h_{FE} , so that the base current contribution to the emitter current is relatively small.

Figure 2.25 shows one last current source, whose output current doesn't

depend on supply voltage. In this circuit, Q_1 's V_{BE} across R_2 sets the output current, independent of V_{CC} :

$$I_{\rm out} = V_{BE}/R_2$$

 R_1 biases Q_2 and holds Q_1 's collector at two diode drops below V_{CC} , eliminating Early effect as in the previous circuit. This circuit is not temperature-compensated; the voltage across R_2 decreases approximately $2.1 \, \mathrm{mV/^\circ C}$, causing the output current to decrease approximately $0.3\%/^\circ C$.

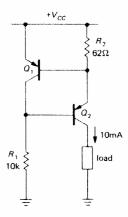


Figure 2.25. Transistor V_{BE} -referenced current source.

2.07 Common-emitter amplifier

Consider a current source with a resistor as load (Fig. 2.26). The collector voltage is

$$V_C = V_{CC} - I_C R_C$$

We could capacitively couple a signal to the base to cause the collector voltage to vary. Consider the example in Figure 2.27. C is chosen so that all frequencies of interest are passed by the high-pass filter it forms in combination with the parallel resistance of the base biasing resistors (the

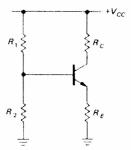


Figure 2.26

impedance looking into the base itself will usually be much larger because of the way the base resistors are chosen, and it can be ignored); that is,

$$C \geq \frac{1}{2\pi f(R_1 || R_2)}$$

The quiescent collector current is 1.0mA because of the applied base bias and the 1.0k emitter resistor. That current puts the collector at +10 volts (+20V, minus 1.0mA through 10k). Now imagine an applied wiggle in base voltage v_B . The emitter follows with $v_E = v_B$, which causes a wiggle in emitter current

$$i_E = v_E/R_E = v_B/R_E$$

and nearly the same change in collector current (h_{fe} is large). So the initial wiggle in base voltage finally causes a collector voltage wiggle

$$v_C = -i_C R_C = -v_B (R_C/R_E)$$

Aha! It's a voltage amplifier, with a voltage amplification (or "gain") given by

$$gain = v_{out}/v_{in} = -R_C/R_E$$

In this case the gain is -10,000/1000, or -10. The minus sign means that a positive wiggle at the input gets turned into a negative wiggle (10 times as large) at the output. This is called a *common-emitter* amplifier with emitter degeneration.



Figure 2.27. with emitter c terminal is the

Input and or common-en

We can eas output imperingut signal and the imput 10k) is abocapacitor that the 3dB point the amplific 8k, which to (well above 8k.

The outp lel with the collector. V ber that if yo for, you're s source. The large (measu output impe collector resi bering that t transistor's c impedance le (as in the en output impe amplifier wil for load resis

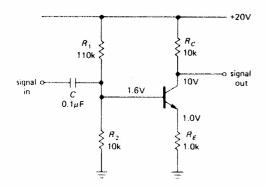


Figure 2.27. An ac common-emitter amplifier with emitter degeneration. Note that the output terminal is the collector, rather than the emitter.

Input and output impedance of the common-emitter amplifier

We can easily determine the input and output impedances of the amplifier. The input signal sees, in parallel, 110k, 10k, and the impedance looking into the base. The latter is about 100k (h_{fe} times R_E), so the input impedance (dominated by the 10k) is about 8k. The input coupling capacitor thus forms a high-pass filter, with the 3dB point at 200Hz. The signal driving the amplifier sees $0.1\mu F$ in series with 8k, which to signals of normal frequencies (well above the 3dB point) just looks like 8k.

The output impedance is 10k in parallel with the impedance looking into the collector. What is that? Well, remember that if you snip off the collector resistor, you're simply looking into a current source. The collector impedance is very large (measured in megohms), and so the output impedance is just the value of the collector resistor, 10k. It is worth remembering that the impedance looking into a transistor's collector is high, whereas the impedance looking into the emitter is low (as in the emitter follower). Although the output impedance of a common-emitter amplifier will be dominated by the collector load resistor, the output impedance of an emitter follower will not be dominated by the emitter load resistor, but rather by the impedance looking into the emitter.

2.08 Unity-gain phase splitter

Sometimes it is useful to generate a signal and its inverse, i.e., two signals 180° out of phase. That's easy to do – just use an emitter-degenerated amplifier with a gain of -1 (Fig. 2.28). The quiescent collector voltage is set to $0.75V_{CC}$, rather than the usual $0.5V_{CC}$, in order to achieve the same result – maximum symmetrical output swing without clipping at either output. The collector can swing from $0.5V_{CC}$ to V_{CC} , whereas the emitter can swing from ground to $0.5V_{CC}$.

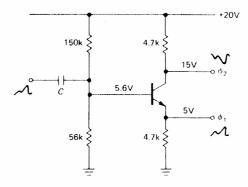


Figure 2.28. Unity-gain phase splitter.

Note that the phase-splitter outputs must be loaded with equal (or very high) impedances at the two outputs in order to maintain gain symmetry.

Phase shifter

A nice use of the phase splitter is shown in Figure 2.29. This circuit gives (for a sine wave input) an output sine wave of adjustable phase (from zero to 180°), but with constant amplitude. It can be best understood with a phasor diagram of voltages (see Chapter 1); representing the input signal by a unit vector along

the real axis, the signals look as shown in Figure 2.30.

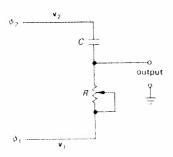


Figure 2.29. Constant-amplitude phase shifter.

Signal vectors \mathbf{v}_R and \mathbf{v}_C must be at right angles, and they must add to form a vector of constant length along the real axis. There is a theorem from geometry that says that the locus of such points is a circle. So the resultant vector (the output voltage) always has unit length, i.e., the same amplitude as the input, and its phase can vary from nearly zero to nearly 180° relative to the input wave as R is varied from nearly zero to a value much larger than Z_C at the operating frequency. However, note that the phase shift also depends on the frequency of the input signal for a given setting of the potentiometer R. It is worth noting that a simple RC high-pass (or low-pass) network could also be used as an adjustable phase shifter. However, its output amplitude would vary over an enormous range as the phase shift was adjusted.

An additional concern here is the ability of the phase-splitter circuit to drive the RC phase shifter as a load. Ideally, the load should present an impedance that is large compared with the collector and emitter resistors. As a result, this circuit is of limited utility where a wide range of phase shifts is required. You will see improved phase-splitter techniques in Chapter 4.

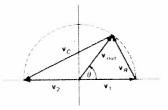


Figure 2.30. Phasor diagram for phase shifter.

2.09 Transconductance

In the preceding section we figured out the operation of the emitter-degenerated amplifier by (a) imagining an applied base voltage swing and seeing that the emitter voltage had the same swing, then (b) calculating the emitter current swing; then, ignoring the small base current contribution, we got the collector current swing and thus (c) the collector voltage swing. The voltage gain was then simply the ratio of collector (output) voltage swing to base (input) voltage swing.

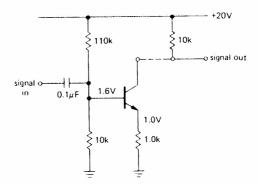


Figure 2.31. The common-emitter amplifier is a transconductance stage driving a (resistive) load

There's another way to think about this kind of amplifier. Imagine breaking it apart, as in Figure 2.31. The first part is a voltage-controlled current source, with quiescent current of 1.0mA and gain of -1mA/ put/input; of current/verse of re (the invers and the in tance) and which used spelled bac gain has u a transcon $I_{\text{out}}/V_{\text{in}}$ is g_m .

Think of transcondu to-current a g_m (gain) which is jus circuit is th that conve resistor cor amplifier, a voltage/cur its quiescer (transresist: is just R_C together giv get the over gains. In th or -10, a ratio (outpi

This is a amplifier, b mance of t example, yo tance part g_m for diff even differe transistors the transres cring gain ' If you are i gain, it is r_m is the tr mately the (current so transresista age gains o

2.10 Improved transistor model: transconductance amplifier

of -1mA/V. Gain means the ratio output/input; in this case the gain has units of current/voltage, or 1/resistance. The inverse of resistance is called conductance (the inverse of reactance is susceptance, and the inverse of impedance is admittance) and has a special unit, the siemens, which used to be called the mho (ohm spelled backward). An amplifier whose gain has units of conductance is called a transconductance amplifier; the ratio $I_{\rm out}/V_{\rm in}$ is called the transconductance,

Think of the first part of the circuit as a transconductance amplifier, i.e., a voltageto-current amplifier with transconductance g_m (gain) of 1mA/V (1000 μ S, or 1mS, which is just $1/R_E$). The second part of the circuit is the load resistor, an "amplifier" that converts current to voltage. resistor could be called a transresistance amplifier, and its gain (r_m) has units of voltage/current, or resistance. In this case its quiescent voltage is V_{CC} , and its gain (transresistance) is 10kV/A ($10k\Omega$), which is just R_C . Connecting the two parts together gives you a voltage amplifier. You get the overall gain by multiplying the two gains. In this case $G = g_m R_C = R_C / R_E$, or -10, a unitless number equal to the ratio (output voltage)/(input voltage).

he

n-

se

er

u-

g-

n.

us

ge

or

lt-

out

is

/e)

ut

it

; a

th

in

This is a useful way to think about an amplifier, because you can analyze performance of the sections independently. For example, you can analyze the transconductance part of the amplifier by evaluating g_m for different circuit configurations or even different devices, such as field-effect transistors (FETs). Then you can analyze the transresistance (or load) part by considering gain versus voltage swing trade-offs. If you are interested in the overall voltage gain, it is given by $G_V = g_m r_m$, where r_m is the transresistance of the load. Ultimately the substitution of an active load (current source), with its extremely high transresistance, can yield one-stage voltage gains of 10,000 or more. The cascode configuration, which we will discuss later, is another example easily understood with this approach.

In Chapter 4, which deals with operational amplifiers, you will see further examples of amplifiers with voltages or currents as inputs or outputs; voltage amplifiers (voltage to voltage), current amplifiers (current to current), and transresistance amplifiers (current to voltage).

Turning up the gain: limitations of the simple model

The voltage gain of the emitter-degenerated amplifier is $-R_C/R_E$, according to our model. What happens as R_E is reduced toward zero? The equation predicts that the gain will rise without limit. But if we made actual measurements of the preceding circuit, keeping the quiescent current constant at 1mA, we would find that the gain would level off at about 400 when R_E is zero, i.e., with the emitter grounded. We would also find that the amplifier would become significantly nonlinear (the output would not be a faithful replica of the input), the input impedance would become small and nonlinear, and the biasing would become critical and unstable with temperature. Clearly our transistor model is incomplete and needs to be modified in order to handle this circuit situation, as well as others we will talk about shortly. Our fixed-up model, which we will call the transconductance model, will be accurate enough for the remainder of the book.

EBERS-MOLL MODEL APPLIED TO BASIC TRANSISTOR CIRCUITS

2.10 Improved transistor model: transconductance amplifier

The important change is in property 4 (Section 2.01), where we said earlier that $I_C = h_{FE}I_B$. We thought of the transistor as a current amplifier whose input circuit behaved like a diode. That's roughly correct, and for some applications it's good enough. But to understand differential amplifiers, logarithmic converters, temperature compensation, and other important applications, you must think of the transistor as a transconductance device – collector current is determined by base-to-emitter voltage.

Here's the modified property 4: **4.** When rules 1-3 (Section 2.01) are obeyed, I_C is related to V_{BE} by

$$I_C = I_S \left[\exp \left(\frac{V_{BE}}{V_T} \right) - 1 \right]$$

where $V_T = kT/q = 25.3 \text{mV}$ at room temperature (68°F, 20°C), q is the electron charge (1.60 × 10⁻¹⁹ coulombs), k is Boltzmann's constant (1.38 × 10⁻²³ joules/°K), T is the absolute temperature in degrees Kelvin (°K = °C + 273.16), and I_S is the saturation current of the particular transistor (depends on T). Then the base current, which also depends on V_{BE} , can be approximated by

$$I_B = I_C/h_{FE}$$

where the "constant" h_{FE} is typically in the range 20 to 1000, but depends on transistor type, I_C , V_{CE} , and temperature. I_S represents the reverse leakage current. In the active region $I_C \gg I_S$, and therefore the -1 term can be neglected in comparison with the exponential.

The equation for I_C is known as the Ebers-Moll equation. It also approximately describes the current versus voltage for a diode, if V_T is multiplied by a correction factor m between 1 and 2. For transistors it is important to realize that the collector current is accurately determined by the base-emitter voltage, rather than by the base current (the base current is then roughly determined by h_{FE}), and that this exponential law is accurate over an enormous range of currents, typically from nanoamps to milliamps. Figure 2.32

makes the point graphically. If you measure the base current at various collector currents, you will get a graph of h_{FE} versus I_C like that in Figure 2.33.

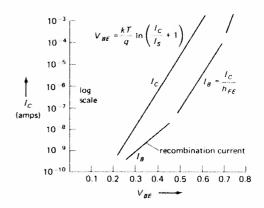


Figure 2.32. Transistor base and collector currents as functions of base-to-emitter voltage V_{BE} .

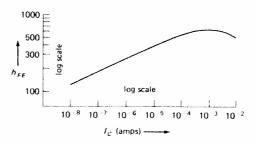


Figure 2.33. Typical transistor current gain (h_{FE}) versus collector current.

Although the Ebers-Moll equation tells us that the base-emitter voltage "programs" the collector current, this property may not be directly usable in practice (biasing a transistor by applying a base voltage) because of the large temperature coefficient of base-emitter voltage. You will see later how the Ebers-Moll equation provides insight and solutions to this problem.

Rules of thumb for transistor design

From the Ebers-Moll equation we can get

several impusing often I. The stee much do we crease I_C 1 Ebers-Moll or 60mV at age increase current. Each where ΔV 2. The sn into the emi voltage. Tak respect to I_C

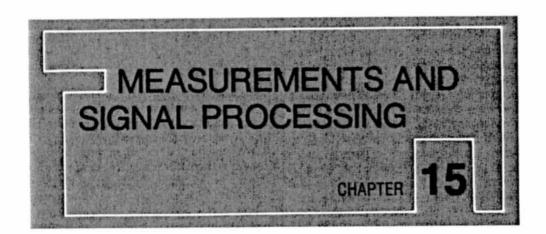
 $r_e = V_T/I_C$

where I_C is value $25/I_C$ intrinsic em: is in series w circuits. It emitter amp lower to hav than unity, a ance of an e zero. Note t grounded en 3. The temp A glance at gests that V_E coefficient. 1 perature dep about 2.1mV to $1/T_{\rm abs}$, w perature.

There is will need on derivable from it the Early e 2.06, and in current-source for example:

4. Early effectioning V_{CI} is caused by cand it is given

 $\Delta V_{BE} = -\alpha.$



OVERVIEW

Perhaps the most exciting (and most useful) area of electronics involves the gathering and manipulation of data from an industrial process or a scientific experiment. Generally speaking, transducers (devices that convert some physical quantity, such as temperature or light level, to a voltage or some other electrical quantity) are used to generate signals that can be manipulated by electronic circuits, quantified by analog-to-digital converters, and logged and analyzed by computers. If the signal you're looking for is masked by noise or interference, powerful "bandwidthnarrowing" techniques such as lock-in detection, signal averaging, multichannel scaling, and correlation and spectral analysis can magically retrieve the sought-after signal. Finally, the results of such physical measurements can be used to control the experiment or process itself, with "on-line" control usually provided by a small computer or micro-processor dedicated to the task. The recent development of powerful and inexpensive microprocessors and support chips has brought about

an explosion in the use of electronics to control and log processes that would not have seemed likely candidates only a decade ago.

In this chapter we will begin with a sampling of quantities that can be measured and the transducers that are normally used for the job. In this area there is plenty of room for ingenuity, and the catalog of transducers we will describe should therefore be considered representative, not exhaustive. We will go into some detail describing the particular problems some of these measurement transducers present and the circuit solutions you might use with them. We will try to cover the most common difficulties, dealing with ultrahigh source impedances (hundreds of megohms in the case of microelectrodes or ionspecific probes), low-level low-impedance transducers (e.g., thermocouples, strain gauges, magnetic pickups), high-impedance ac sensors (capacitance transducers), and others.

The chapter will continue with a look at precision standards (standards of frequency and time, as well as voltage and resistance) and some of the techniques of

987

precision measurements. We will then describe in some detail the whole business of bandwidth-narrowing, "pulling the signal from the noise." These techniques are extremely powerful, and they are mysterious to the uninitiated. Finally, we will conclude the chapter with a brief look at spectrum analysis and Fourier techniques. Readers interested primarily in electronic circuit design may wish to skip this chapter.

MEASUREMENT TRANSDUCERS

In some situations the quantity you want to measure is itself an electrical quantity. Examples might be nerve impulses (voltage), seawater conductivity (resistance), charged-particle fluxes (current), etc. In these cases measurement techniques tend to be relatively straightforward, with most of the difficulties centering around the kind of collection electrode to use and how to handle the signals once they've been collected. You might encounter very high impedances (e.g., with microelectrodes) or very small signals (e.g., a current generated from radioactive decay).

More often, a "transducer" of some sort is necessary to convert some physical quantity to a quantifiable electrical quantity. Examples are measurements of temperature, light level, magnetic field, strain, acceleration, sound intensity, etc. In the following sections we will take a look at some of the more common input transducers to give an idea of what can be measured and how accurately. We will go into somewhat greater detail when describing the more common measurements, such as heat and light, but in a book of this scope we can cover only a fraction of the measurement possibilities.

15.01 Temperature

Temperature transducers illustrate a nice variety of performance trade-offs. Temper-

ature range, accuracy, repeatability, conformity to a universal curve, size, and price are all involved.

Thermocouples

A junction between two dissimilar metals generates a small voltage (with low source impedance!), typically in the millivolt range, with coefficients of about $50\mu V/^{\circ}C$. This junction is called a thermocouple, and it is useful for measuring temperatures over a broad range. By using various pairs of alloys it is possible to span temperatures from -270°C to +2500°C with reasonable accuracy (0.5-2°C). The thermoelectric properties of different alloys are well known, so thermocouple probes in different formats (rods, washers, armored probes, etc.) made from the same alloys can be interchanged without affecting calibration.

The classic thermocouple circuit is shown in Figure 15.1. The particular choice of metals in this figure constitutes what is known as a type J thermocouple (look at Table 15.1 for a listing of the standard choices and their properties). Each couple is made by welding the two dissimilar metals together to make a small junction. (People have been known to get away with twisting the wires together, but not for very long!) The reference junction is absolutely necessary, since otherwise you wind up with additional dissimilar thermocouples where the dissimilar metals join the metal terminals. Those extra thermoelectric voltages produced at uncontrolled places in the circuit would result in erratic and inaccurate results. Even with a pair of thermocouple junctions, you still have thermocouples formed where the leads join the metal terminals. However, this seldom causes problems, since those junctions are at the same temperature.

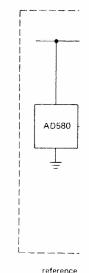
The thermocouple circuit gives you a voltage that depends on the temperatures of both junctions. Roughly speaking, it is proportional to the difference between



millivolt meter

Figure 15.1

the two ju actually v sensing ju handle the Classically at a fixed used to use you can al boxes to d suring ver not even c having the temperatu nique is to that correct



compensati

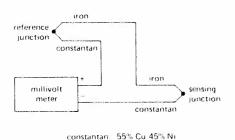


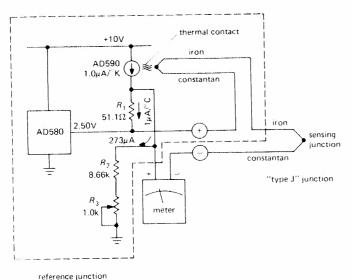
Figure 15.1. Classic thermocouple circuit.

the two junction temperatures. What you actually want is the temperature at the sensing junction. There are two ways to handle the problem of the reference: (a) Classically, you put the reference junction at a fixed temperature, usually 0°C. They used to use ice baths, and you still can, but you can also buy nice little stabilized cold boxes to do the same job. If you are measuring very high temperatures, you may not even care about small errors caused by having the reference junction at "room" temperature. (b) A more modern technique is to build a compensation circuit that corrects for the difference caused by

having the reference junction at a temperature other than 0°C.

Figure 15.2 shows how this is done. The basic idea is to use a temperature-sensing chip and circuitry that adds in a voltage that makes up for the difference between the actual reference junction temperature and the standard 0°C. The AD590 (see the subsequent section on IC temperature sensors) produces an output current in microamps equal to the temperature in degrees Kelvin. R_1 is chosen according to the thermoelectric coefficient, in this case converting $1\mu A/^{\circ}C$ to $51.5\mu V/^{\circ}C$ (see Table 15.1), and the AD580 3-terminal reference (in combination with R_2 and R_3) is used to subtract the AD590 offset of $273\mu A$ at 0°C (273.15°K). Thus, there is no correction made when the reference junction is at 0° C, and 51.5μ V/°C (the thermoelectric coefficient of a type J junction at room temperature) is added to the net output voltage of the pair of junctions when the reference junction is at some other temperature.

The metering circuit deserves a few words of comment. The circuit problems



compensation circuit

Figure 15.2. Thermocouple reference junction compensation.

TABLE 15.1. THERMOCOUPLES

	Alloy	Max temp ^a (*C)	Tempco @20°C (μV/°C)	Output voltage ^b			30 gauge
Туре				100°C (mV)	400°C (mV)	1000°C	lead resistance $^{\mathrm{c}}$ (Ω)
J	{Iron Constantan ^d	760	51.45	5.268	21.846		3.6
κ	{Chromel ^e Alumel	1370	40.28	4.095	16.395	41.269	6.0
T	{Copper Constantand	400	40.28	4.277	20.869	_	3.0
E	{Chromel ^e Constantan ^d	1000	60.48	6.317	28.943	76.358	7.2
s	{Platinum 90%Pt-10%Rh	1750	5.88	0.645	3.260	9.585	1.9
R	{Platinum 87%Pt-13%Rh	1750	5.80	0.647	3.407	10.503	1.9
В	{94%Pt-6%Rh 70%Pt-30%Rh	1800	0.00	0.033	0.786	4.833	1.9

⁽a) thermocouple life is shortened by prolonged operation near maximum temperature. (b) reference junction at 0°C. (c) per double foot; for 24-gauge, multiply values by 0.25. (d) 55%Cu-45%Ni. (e) 90%Ni-10%Cr. (f) 96%Ni-2%Mn-2%Al.

you have with thermocouples stem from their low output voltage $(50\mu\text{V})^{\circ}\text{C}$ or thereabouts), combined with large common-mode ac and radiofrequency interference. The amplifier must have good common-mode rejection at 60Hz and stable differential gain. In addition, the input impedance must be moderately high (of order 10k or more) in order to prevent error from loading, since the thermocouple leads do have some resistance (5 feet of 30 gauge type K junction wire has a resistance of 30Ω , for example).

The circuit shown in Figure 15.3 is a good solution. It is just the standard differencing amplifier with the T connection in the feedback path to get high voltage gain (200 in this case) while keeping the input impedance large enough so that loading of the source impedance doesn't contribute error. The op-amp is a precision low-offset type, with drift of less than $1\mu V/^{\circ}C$ to keep its contribution to the measurement error much less than the $50\mu V$ that corresponds to a $1^{\circ}C$ error. The input bypass

capacitors are a good idea to reduce the common-mode interference at 60Hz and at radiofrequencies (thermocouples and their long connecting cables tend to behave like radio antennas). Since thermocouples respond slowly anyway, you can limit the bandwidth with capacitors across the feedback resistors, as shown. In cases of extreme radiofrequency problems, it may be necessary to shield the input leads and add RF chokes before the input bypass capacitors.

Note that the reference junction compensation circuit in Figure 15.3 acts as the output, rather than the usual method of compensating the voltage from the thermocouple at the input, as in Figure 15.2. This is done to keep the input truly differential, in order to preserve the advantages of the good common-mode rejection of the differencing amplifier. Since the amplifier has a voltage gain of approximately 200, the compensation circuit has to add $200 \times 51.5 \mu V/^{\circ}C$, or $10.3 m V/^{\circ}C$ at the output. Note that the OP-97E's input

sensing junction **∢** (type J)

constanta

offset curre $25\mu V$ of i the $25\mu V$ to zero. A chopper-state $(V_{os} = 5\mu V)$ An instr

ure 7.32, conferencing a case be sur the input.

Thermod Analog De with cold AD594 (tyr monolithic need (inclu produce a put voltage



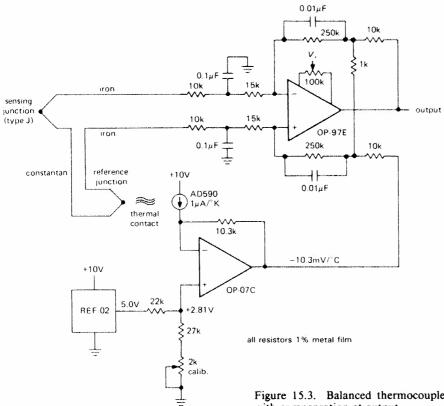


Figure 15.3. Balanced thermocouple amplifier with compensation at output.

offset current of 0.1nA (max) contributes $25\mu V$ of input error, which, along with the $25\mu V$ (max) of V_{os} , can be trimmed to zero. Alternatively, one could use a chopper-stabilized op-amp like the 7652 $(V_{os} = 5\mu V \text{ max}, I_{os} = 40\text{pA max}).$

An instrumentation amplifier, as in Figure 7.32, could be used instead of the differencing amplifier we've shown; in that case be sure to provide a dc bias path at the input.

Thermocouple users should be aware of Analog Devices' "thermocouple amplifier with cold junction compensation," the AD594 (type J) and AD595 (type K). These monolithic devices have everything you need (including an ice-point reference) to produce a temperature-proportional output voltage, or even programmable trip

point, given a thermocouple input. The best grade is accurate (without trimming) to ±1°C at room temperature, increasing to $\pm 3^{\circ}$ C at -25° C and $+75^{\circ}$ C. Linear Technology makes the LT1025 "micropower thermocouple cold-junction compensator," designed to be used with an external precision op-amp. It includes compensation for all thermocouples in Table 15.1 (except type B), with second-order curvature correction to maintain accuracy over a wider temperature range. The best grade (LT1025A) is accurate (without trimming) to $\pm 0.5^{\circ}$ C at room temperature, increasing to $\pm 2^{\circ}$ C at -25° C and $+80^{\circ}$ C.

Complete "smart" temperature-measuring instruments configured for various thermocouple pairs are available commercially. These instruments include computational

circuitry to convert the thermoelectric voltage to temperature. For instance, the digital thermometers manufactured by Analog Devices and Omega Engineering achieve an accuracy of about 0.4°C over a temperature range from -200°C to +1000°C and an accuracy of 1°C at temperatures up to +2300°C.

When compared with other methods of temperature measurement, thermocouples have the advantage of small size and wide temperature range, and they are particularly good for measuring high temperatures.

Thermistors

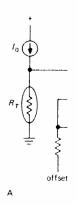
Thermistors are semiconductor devices that exhibit a negative coefficient of resistance with temperature, typically in the neighborhood of -4%/°C. They are available in all sorts of packages, ranging from tiny glass beads to armored probes. Thermistors intended for accurate temperature measurement (they can also be used as temperature-compensation elements in circuits, for instance) typically have a resistance of a few thousand ohms at room temperature, and they are available with tight conformity (0.1-0.2° C) to standard curves. Their large coefficient of resistance change makes them easy to use, and they are inexpensive and stable. Thermistors are a good choice for temperature measurement and control in the range of -50°C to +300°C. It is relatively easy to design a simple and effective circuit for "proportional temperature control" using a thermistor sensing element; see, for example, RCA application note ICAN-6158 or the Plessey SL445A data sheet.

Because of their large resistance change with temperature, thermistors make no great demands on the circuitry that follows. Some simple ways to generate an output voltage are shown in Figure 15.4. The circuit in part A expands the lowtemperature end of the range because of

the thermistor's exponential resistance change, whereas the circuit in part B produces a somewhat more linear variation of output voltage with temperature. For example, Figure 15.5 shows resistance versus temperature for two configurations a bare 10k thermistor (Fenwal UUA41J1 "uni-curve"), and the same thermistor with a 10k resistor in series. The series pair is linear to within 3% from -10° C to $+50^{\circ}$ C. and better than 1% from 0°C to +40°C. The circuit of Figure 15.4B, sensing the drop across R, produces a voltage output of equivalent linearity.

Parts C and D of Figure 15.4 show elaborations of this linearizing idea, using composite matched "thermilinear" thermistors (and corresponding resistor pairs) from Yellow Springs Instrument Company. These 2-thermistor configurations achieve 0.2% linearity from 0°C to 100°C. YSI also makes modules with three thermistors (and three resistors), for even better linearity. The circuit in part E is the classic Wheatstone bridge, balanced when $R_T/R_2 = R_1/R_3$; since it is ratiometric, the null doesn't shift with variations in supply voltage. The bridge circuit, with a high-gain amplifier, is particularly good for detecting small changes about some reference temperature; for small deviations the (differential) output voltage is linear in the unbalance. With all the thermistor circuits you have to be careful about self-heating effects. A typical small thermistor probe might have a dissipation constant of $1 \text{mW}/^{\circ}\text{C}$, meaning that $I^{2}R$ heating should be kept well below 1mW if you would like your reading accurate to better than I degree.

Complete "smart" temperaturemeasuring instruments using curveconforming thermistors are available commercially. These devices include internal computational circuitry to convert resistance readings directly to temperature. As an example, the Omega model 5800 digital thermometer covers -30° C to $+100^{\circ}$ C





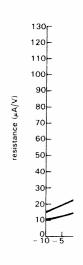
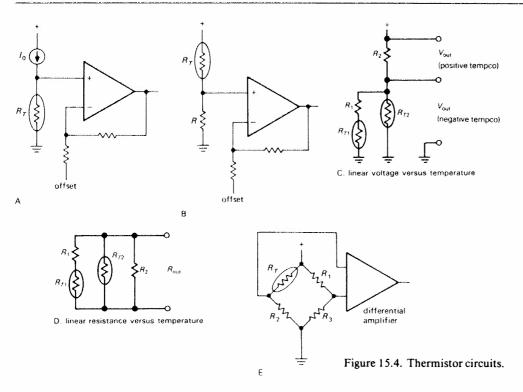


Figure 15.5 thermistor an





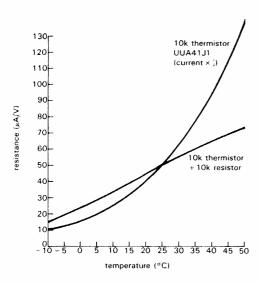


Figure 15.5 Resistance versus temperature for thermistor and thermistor-resistor pair.

in two ranges, reading out in either centigrade or Fahrenheit on a 4-digit LED display. It has 0.5°C accuracy over the full temperature range, with 0.1°C resolution.

When compared with other methods of temperature measurement, thermistors provide simplicity and accuracy, but they suffer from self-heating effects, fragility, and a narrow temperature range.

Platinum resistance thermometers

These devices consist simply of a coil of very pure platinum wire, which has a positive temperature coefficient of about $0.4\%^{\circ}/C$. Platinum thermometers are extremely stable with time and conform very closely $(0.02-0.2^{\circ}C)$ to a standard curve. They are usable over a temperature range of $-200^{\circ}C$ to $+1000^{\circ}C$. They aren't terribly cheap.

IC temperature sensors

As we remarked in Section 6.15, a bandgap voltage reference can be used to generate a temperature-sensing output voltage proportional to absolute temperature, as well as its usual stable zero-tempco reference output. The REF-02, for example, provides a "temp" output with a linear coefficient of +2.1mV/°C. If you buffer this output with an amplifier of adjustable gain and offset for calibration, you can achieve accuracies of about 0.5°C over the range -55°C to +125°C. The LM335 is a convenient 2-terminal temperature sensor that behaves like a zener diode with a voltage of +10mV/°K; e.g., at 25°C (298.2°K) it acts like a 2.982 volt zener (Fig. 15.6). It comes with an initial accuracy as good as ±1°C and it can be externally trimmed. A single point calibration can typically improve its accuracy to ±0.5°C max over a -55°C to +125°C range. After trimming, the output should be accurate to 0.1°C at the test temperature, with an accuracy budget increasing to ±0.5°C at the temperature extremes (Fig. 15.7).

The LM35 also provides a voltage output with +10mV/°C slope, but it behaves like a 3-terminal reference (rather than a 2-terminal zener), powered by +4 volts to +30 volts applied to a third terminal; it has an internal offset such that the output is 0 volts at 0°C. To operate near or below 0°C, you must use a pulldown resistor, as shown. The best grade (LM35A) has a maximum error of 0.5°C, but cannot be trimmed. Its cousin, the LM34A, works similarly, with readout in Fahrenheit (0 volts at 0°F).

Another approach to IC temperature sensors is the AD590, a two-terminal device that acts as a constant-current element, passing a current in microamps equal to the absolute temperature; e.g., at 25° C (298.2°K) it behaves like a constant-current regulator of 298.2μ A ($\pm 0.5\mu$ A). With this simple device you get 1° C accuracy (best grade) over the range -55° C

1µA/°K AD590 2.980V ± 10mV/°C 9.76 0 Va = 67.7 ± 2mV@25°C temp (°K) ± 4% 2.980V ± 10mV/°C 10.0 2.98V@25°C + 10mV/°C LM335A Ø trim C 0.250V@25°C ±0.005V + 10mV/°C

Figure 15.6 IC temperature sensors. Voltages are at 25°C. Methods A and B incur an additional 1% error per 33ppm/°C resistor tempco.

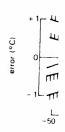


Figure 15.

to +150° cuitry req device. A comparab temperatu The LM3: 6.18) also absolute to resistor ac $R_{\rm set}(\Omega)$ (scludes ≈ 5 ° cludes

Quartz the

The chang quartz crys ploited to 1 thermomet tive in qu lowest post this case y large coeffic high precis surements. mercial ins 2804A, a r mometer w millidegrees +150°C (re range) and microdegree mance, the tion data fc uses in the t



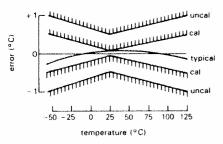


Figure 15.7. LM335 temperature error.

to +150°C. The simplicity of external circuitry required makes this a very attractive device. A plastic version, the AD592, has comparable performance over a smaller temperature range (-25°C to +105°C). The LM334 current-source IC (see Section 6.18) also has an output proportional to absolute temperature, settable with a single resistor according to $I_{\rm out}(\mu A)$ =227T(°K)/ $R_{\rm set}(\Omega)$ (see Fig. 15.6); this formula includes \approx 5% correction for V_- current.

Quartz thermometer

The change of resonant frequency of a quartz crystal with temperature can be exploited to make an accurate and repeatable thermometer. Although the usual objective in quartz-crystal oscillator design is lowest possible temperature coefficient, in this case you choose a crystal cut with a large coefficient and take advantage of the high precision possible in frequency measurements. A good example of a commercial instrument is the Hewlett-Packard 2804A, a microprocessor-controlled thermometer with an absolute accuracy of 40 millidegrees over the range -50°C to +150°C (reduced accuracy over a wider range) and temperature resolution of 100 microdegrees. To get this kind of performance, the instrument contains calibration data for the individual sensor that it uses in the temperature calculation.

Pyrometers and thermographs

An interesting method of "noncontacting" temperature measurement is exemplified by the classic pyrometer, a gadget that lets you sight through a telescope at an incandescent object, comparing its glowing color with that of a filament inside the pyrometer. You adjust the filament's current until its brightness matches the object's, both being viewed through a red filter, then read off the temperature. This is a handy method of measuring the temperatures of very hot objects, objects in inaccessible places like ovens or vacuum chambers, or objects in oxidizing or reducing atmospheres where thermocouples cannot be used. Typical optical pyrometers cover a range of 750°C to 3000°C, with an accuracy of about 4°C at the low end and 20°C at the high end.

The development of good infrared detectors has extended this sort of measurement technique down to ordinary temperatures and for ordinary use. For example, Omega offers a line of digital-readout infrared pyrometers with a temperature range of -30°C to +5400°C. By measuring the intensity of infrared radiation, perhaps at several infrared wavelengths, you can determine with good accuracy the temperature of a remote object. Such "thermography" has recently become popular in quite diverse fields: in medicine, for the detection of tumors, and in the energy business, where a thermograph of your house can tell you where your energy dollars are evaporating.

Low-temperature measurements

Cryogenic (very cold) systems pose special problems when it comes to accurate temperature measurement. What matters there is how close to absolute zero (0°K=273.15°C) you are. Two popular methods involve measuring the resistance of ordinary carbon-composition resistors,

Chapter 15

which soars at low temperatures, and measuring the degree of paramagnetism of some salt. These are really specialty measurement techniques that will not be dealt with here.

Measurement allows control

If you have a way of adjusting some quantity, then the availability of a good measurement technique lets you control that quantity accurately. Thermistors, in particular, provide a nice method for controlling the temperature of a bath or oven.

15.02 Light level

The measurement, timing, and imaging of low light levels are parts of a well-developed field, thanks to the existence of amplification methods that do not rely on conventional circuit techniques. Photomultipliers, channel-plate intensifiers, CCDs (charge-coupled devices), and ISITs (intensified silicon intensifier target) are included in the catalog of high-performance optical detection devices. We will begin with the simplest detectors (photodiodes and phototransistors) and then go on to discuss the exotic and the wonderful.

Photodiodes and phototransistors

A diode junction acts as a photodetector: Light creates electron-hole pairs, and therefore a current through the external circuit. Diodes intended as photodetectors (photodiodes and PIN diodes) are packaged in a transparent case and are designed for high speed, high efficiency, low noise, and low leakage current. In the simplest mode of operation, a photodiode can be connected directly across a resistive load, or current/voltage converter, as shown in Figure 15.8. You get faster response (and the same photocurrent) by reverse-biasing

the junction, as in Figure 15.9. High-speed PIN diodes have response times of a nanosecond or less (1GHz bandwidth) when loaded into a low impedance. It should be noted that the leakage current of good PIN diodes is so low (less than a nanoamp) that the Johnson noise in the load resistance dominates for load resistances less than $100 \mathrm{M}\Omega$ or so, implying a speed/noise trade-off. An additional problem to be aware of is the error caused by the amplifier's input offset voltage, or the applied bias voltage, in combination with the photodiode's "dark resistance," when working at low light levels.

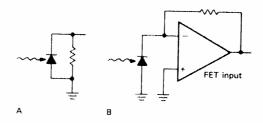


Figure 15.8

Photodiodes are pretty good light detectors when there is plenty of light around, but the output signal can be inconveniently small at low light levels. Typical sensitivities are of the order of 1 microamp per microwatt of incident light. A flux of 1000 photons per second, quite visible with the unaided eye, would cause a photocurrent of 4×10^{-16} amps when focused onto a PIN diode, totally undetectable when compared with the leakage current and noise. No silicon photodetectors are sensitive at the photon level (see the subsequent section on photomultipliers for that), but a device known as a phototransistor has considerably more output current than a photodiode at comparable light levels, bought at the expense of speed. It works like an ordinary transistor, with the base current provided by the photocurrent produced in the base-collector junction.



Figure 15.9

Inexpens MRD701 milliamp of 1mW/cm², of microse like the N roughly 501 of 100 µs or additional of tor or photoits ability to els (its "de limit is set current."

We have the photodi Hamamatsi impressive silicon pn-j GaAs (both and GaP pl tectors. It many sizes photodiode detectors we

Photomultig

For low-ligh ment (and, resolution), tiplier. Thi ton (the sn an electron metal "phot plier then a



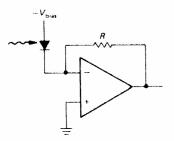


Figure 15.9

Inexpensive phototransistors like the MRD701 have output currents of a milliamp or so at an illumination of 1mW/cm^2 , with rise and fall times of tens of microseconds, and photo-Darlingtons like the MRD711 have photocurrents roughly 50 times higher, but with rise times of $100 \mu \text{s}$ or more. Note, however, that the additional current gain of a phototransistor or photo-Darlington doesn't improve its ability to detect extremely low light levels (its "detectivity"), since the ultimate limit is set by the detector diode's "dark current."

We have had excellent results with the photodiodes from UDT, Siemens, and Hamamatsu. The latter's catalog lists an impressive variety of detectors, including silicon pn-junction, silicon PIN diodes, GaAs (both diffusion and Schottky types) and GaP photodiodes, and avalanche detectors. It includes single detectors of many sizes and shapes, as well as linear photodiode arrays. UDT makes a series of detectors with integral MBC connectors.

Photomultipliers

For low-light-level detection and measurement (and, incidentally, for nanosecond resolution), you can't beat the photomultiplier. This clever device allows a photon (the smallest unit of light) to eject an electron from a photosensitive alkali metal "photocathode." The photomultiplier then amplifies this feeble photocurrent by accelerating the electron onto successive surfaces (dynodes), from which additional electrons are easily ejected. Figure 15.10 suggests the process. This use of "electron multiplication" yields extremely low noise amplification of the initial photocurrent signal. Typically, you use a voltage divider to put about 100 volts between successive dynodes, for a gain of about 10 per stage, or 1 million overall. The final current is collected by the anode, usually run near ground potential (look at Fig. 15.11), and is large enough so that subsequent amplifier noise is negligible.

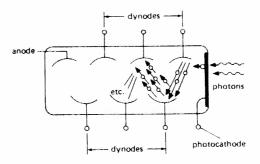


Figure 15.10. Photomultiplier multiplication process.

The most efficient photocathode materials have quantum efficiencies exceeding 25%, and with the large gain provided by the dynodes, individual photoelectron events are easily seen. At low light levels that's how you would use it, following the PMT (photomultiplier tube) with chargeintegrating pulse amplifiers, discriminators (as detailed in Fig. 13.60), and counters. At higher levels the individual photoelectron count rate becomes too high, so you measure the anode current as a macroscopic quantity, instead. PMTs have typical sensitivities of an ampere per microwatt, although you could never operate them at such a high current; maximum PMT anode currents are limited to a milliamp or less. It should be noted that the practical limit to photon counting,

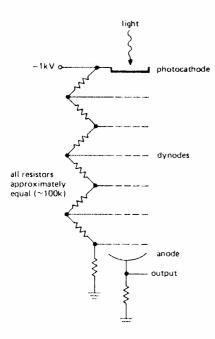


Figure 15.11. Photomultiplier biasing.

something like 1 million counts per second, corresponds to roughly 2 micromicro-watts incident power!

Convenient electronic packages are available for both pulse-counting and output-current-measurement modes. An example is the "quantum photometer" from PAR, a handy gadget with built-in high-voltage supply and both pulse and current electronics. It has 11 ranges of pulse counting (10 pps to 10⁶ pps, full scale) and 11 ranges of anode-current readout (10nA to 1mA, full scale).

Even in total darkness you get a small anode current from a photomultiplier. This is caused by electrons thermally excited from the photocathode and dynodes, and it can be reduced by cooling the PMT down to -25° C or so. Typical dark currents for a sensitive "bialkali" cathode PMT are in the neighborhood of 30 counts per second per square centimeter of cathode area, at room temperature. A cooled

PMT with a small cathode can have dark currents of less than a count per second. It should be pointed out that a powered PMT should never be exposed to ordinary light levels; a PMT that has seen the light of day, even without power applied, may require 24 hours or more to "cool down" to normal dark-current levels. In some applications (e.g., fluorescence measurements) a PMT may be exposed to bright flashes of light at known times. In that case you can minimize the overload recovery time by disabling the accelerating voltage on the first few dynodes during the flash (some manufacturers offer this capability in their PMT/socket combinations).

When compared with photodiodes, PMTs have the advantage of high quantum efficiency while operating at high speed (2ns rise time, typically). They are bulky, though, and they require a stable source of high voltage, since the tube's gain rises exponentially with applied voltage.

It should be emphasized that PMTs are to be used with extremely low light levels. You run them with typical anode currents of a microamp or less, and they can easily see light that you cannot. Photomultipliers are used not only for the detection of light directly, as in astronomy (photometry) and biology (bioluminescence, fluorescence), but also in conjunction with scintillators as particle detectors and x-ray/ gamma-ray detectors, as we will discuss in Section 15.07. Photomultipliers find wide use in spectrophotometry, where they are combined with prisms, gratings, or interferometers to make precise measurements of optical spectra. PMTs are manufactured by RCA (Burle), Hamamatsu, EMI, and EG&G.

CCDs, intensifiers, SITs, ISITs, and image dissectors

It is possible to do *imaging* at the light quantum level, thanks to some clever recent technology; i.e., you can form an image wit low light | imaging) | ventions ; in what a room, the in which ; "snow," al

The ke sifier, an as its outp put image dinary sili or a CCD two-dimen an image a by scannir shifting the register, re have is a to ity is far be it is the tw todiode. you simply in front. F schematica

Intensififirst-genera tive photoc in photomi optics and behind so cathode are



Figure 15.12.



image with the same sort of sensitivity to low light levels that you get with the (non-imaging) photomultiplier. These recent inventions are amazing to see. You can sit in what appears to be a completely dark room, then peer into a television monitor in which are imaged, albeit with plenty of "snow," all the objects in the room.

The key to all this is the image intensifier, an incredible device that produces as its output a brightened replica of an input image. You begin with either an ordinary silicon target vidicon (TV camera) or a CCD array. These are light-sensitive two-dimensional targets that accumulate an image and can be read out electronically by scanning with an electron beam or by shifting the image along as an analog shift register, respectively. At this point, all you have is a television camera whose sensitivity is far below the individual photon level; it is the two-dimensional analog of a photodiode. To bring about the miraculous, you simply put an imaging intensifier tube in front. Figure 15.12 suggests the process schematically.

Intensifiers come in two varieties. The first-generation type consists of a sensitive photocathode surface of the type used in photomultipliers, with electron-focusing optics and a phosphor screen arranged behind so that photoelectrons from the cathode are accelerated by high applied

voltages and hit the phosphor with enough energy to give off a bright flash of light. With this kind of intensifier you can get single-stage light amplification of about 50. with resolution of about 50 lines/mm. Popular types cascade two, three, or four such stages of light amplification to achieve overall light amplifications of a million or more. The input and output may simply be glass surfaces with their internal photosensitive and phosphor coatings, or they may be paved with a dense fiber-optic bundle. Fiber-optics are nice because they let you match a perfectly flat entrance or exit surface to a curved tube surface, and they simplify the external optical system, since you can cascade these devices by just sticking things together, without any lenses.

Second-generation intensifiers using "microchannel plates" allow you to achieve much higher values of single-stage light amplification, and they are better at really low light levels because of fewer "ion events," the result of positive ions being ejected from the phosphor and returning to the cathode, where they make a big splash. In these channel-plate intensifiers the space from cathode to phosphor contains a bundle of microscopic hollow tubes whose insides are coated with a dynode-type multiplication surface. Photoelectrons from the cathode bounce their way down these channels, ejecting secondary

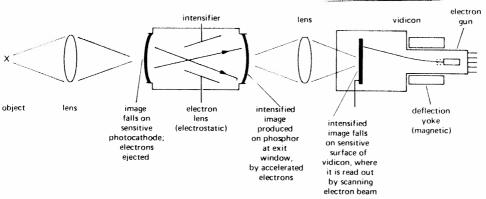


Figure 15.12. Vidicon with single-stage intensifier.

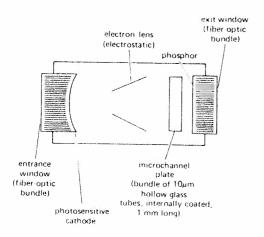


Figure 15.13. Electrostatically focused channel-plate intensifier.

electrons to give light amplification of about 10,000 (Fig. 15.13). You can get resolutions of about 20 lines/mm, and with special configurations ("J-channel," "C-channel," "chevron") the ion-event problem can be eliminated almost entirely. The result is an imaging intensifier with the same sort of quantum efficiency as photomultipliers (20%-30%). The use

of nearly noiseless electron multiplication results in light amplification to a level that the vidicon or CCD can see.

Such an intensifier combined with a silicon-target vidicon in a single tube is called a "SIT" (silicon intensifier target). An ISIT is a SIT with an additional intensifier placed externally in front (Fig. 15.14); this is the sort of gadget that lets you see in the dark. These things are very popular with astronomers and with nightwarfare people.

An interesting variation of the imaging intensifier is the so-called image dissector, a clever device that actually preceded the devices just described. It consists of a sensitive photocathode area, followed by the usual photomultiplier dynode chain. In between is a small aperture and some deflection electrodes, so that any spot on the photocathode can become the active area for electron multiplication by the dynode system. You can think of an image dissector as a photomultiplier with an electronically movable photocathode area. It has the quantum efficiency and gain of a conventional PMT, but it differs from the intensified vidicons, CCDs, and SITs

one-stage intensifier intensifier + vidicon high-voltage leads insulating photocathode "potting vidicon phospha material tube base magentic yoke assembly photocathode magnetic yoke assembly fiber-optic 18-mm-diameter electrostatic entrance focusing electrostatic window focusing scanning readout bundle) (vidicon)

Figure 15.14. Intensified silicon intensifier target (ISIT).

(which are that it does the entire 1 CCD de used witho them from Kodak, T Toshiba. sional ("lir 4096 cells, arrays con more cells (ments"); To an impress chip. Linea for spectro: are used fo example in All CCI which acci until the ar the CCD b with the in serial analo

15.03 Stra

line.

The field variables so own bag of measurer s strain gaug key to all measureme

There ar position, d tion), and s

LVDT. A (linear var. which is all struct a tra excite one sure the inc The second out as two s



(which are all image-integrating devices) in that it does not accumulate the image over the entire field in between readouts.

CCD detector arrays can, of course, be used without intensification. You can buy them from companies like EGG Reticon, Kodak, TI, Tektronix, Thomson, and Toshiba. They are available as 1-dimensional ("linear") arrays containing up to 4096 cells, or as 2-dimensional ("area") arrays containing 256K (512 × 512) or more cells (called "pixels," for "picture elements"); Toshiba has an area detector with an impressive 2 million pixels on a single chip. Linear arrays are handy as detectors for spectroscopy; area detectors, of course, are used for full 2-dimensional images, for example in television cameras.

All CCDs are light-integrating devices, which accumulate charge in each pixel until the array is read out. During readout the CCD becomes an analog shift register, with the image emerging, raster style, as a serial analog waveform at the single output line.

15.03 Strain and displacement

The field of measurements of physical variables such as position and force has its own bag of tricks, and any accomplished measurer should be aware of things like strain gauges, LVDTs, and the like. The key to all of these measurements is the measurement of displacement.

There are several nice ways to measure position, displacement (changes in position), and strain (relative elongation).

LVDT. A popular method is the LVDT (linear variable differential transformer), which is almost self-explanatory. You construct a transformer with a movable core, excite one set of coils with ac, and measure the induced voltage in the second set. The secondary is center-tapped (or brought out as two separate windings) and arranged

symmetrically with respect to the primary, as shown in Figure 15.15. LVDTs come in an enormous variety of sizes, with fullscale displacements ranging from 0.005 inch to 25 inches, excitation frequencies from 50Hz to 30kHz, and accuracies of 1% down to 0.1% or better. A leader in this field is Schaevitz, whose catalog lists a broad selection of linear and angular ("RVDT") transducers, and measurement transducers utilizing LVDT sensors (e.g., pressure, force, acceleration, etc.), as well as LVDT readout electronics. If you become involved with LVDTs, you may want to build your own instrumentation, perhaps using the special ICs designed for the purpose. For example, the monolithic Signetics NE5520/1 "LVDT Signal Conditioner" provides sine-wave excitation signals and includes a synchronous demodulator to provide a voltage output proportional to the LVDT displacement. The 2S54/6 from Analog Devices are LVDT synchronous demodulators with excellent linearity (0.01%) and built-in A/D converters providing direct digital outputs (14 and 16 bits, respectively).

Strain gauge. A strain gauge measures elongation or flexure by subjecting an array of four metal thin-film resistors to deformation. They come as complete assemblies, in sizes from 1/64 of an inch to several inches, and they generally have impedances in the neighborhood of 350 Electrically they look like a ohms/leg. Wheatstone bridge; you apply dc across two of the terminals and look at the voltage difference between the other two, as discussed in Section 7.09. The output voltages are very small, typically 2mV per volt of excitation for full-scale deformation, with accuracies ranging from 1% down to 0.1% of full scale (see Fig. 15.15D).

It is not easy to measure small relative elongations, and strain-gauge specifications are notoriously unreliable. Small differences in the temperature coefficients

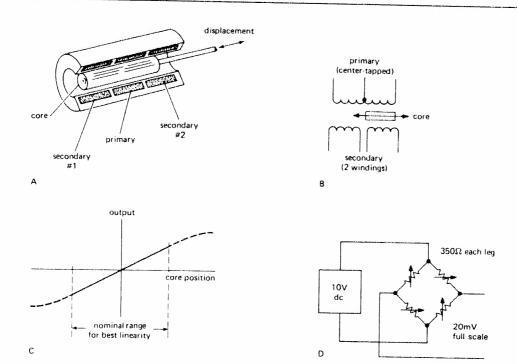


Figure 15.15. Displacement transducers.

- A. Linear variable differential transformer (LVDT) cutaway view.
- B. LVDT schematic.
- C. LVDT output versus displacement.
- D. Strain gauge schematic.

of the bridge elements are responsible for the temperature sensitivity, which limits the performance of the strain gauge. This is a problem, even in controlled-temperature environments, because of self-heating. For example, 10 volts of dc excitation on a 350 ohm bridge will produce 300mW dissipation in the sensor, with a temperature rise of 10°C (or more), causing errors corresponding to a real signal of 0.1% to 0.5% of full scale.

Recently, semiconductor strain gauges have become popular. They have outputs that are 10 times higher than those of the metal-film variety and impedances of a few thousand ohms. It is often necessary to use a current source as excitation, rather than

a voltage source, to minimize temperature sensitivities.

Capacitance transducers. Very sensitive measurements of displacement can be made with a transducer consisting simply of two closely spaced plates, or a plate suspended between a pair of outer plates. By making the capacitor part of a resonant circuit, or by using a high-frequency ac bridge, you can sense or control very small changes in position. Capacitor microphones use this principle to convert acoustic pressure or velocity to an audio signal.

The amplifiers used with capacitor microphones illustrate some interesting circuit ideas, and they are of practical

important ing micro transduce: metallized to a fixed through a to 100 vo of voltage sound fiel

Capacit mously hi capsule ha a reactanc means tha ning the s ble whatsoright at th

R₂ 1000М 2

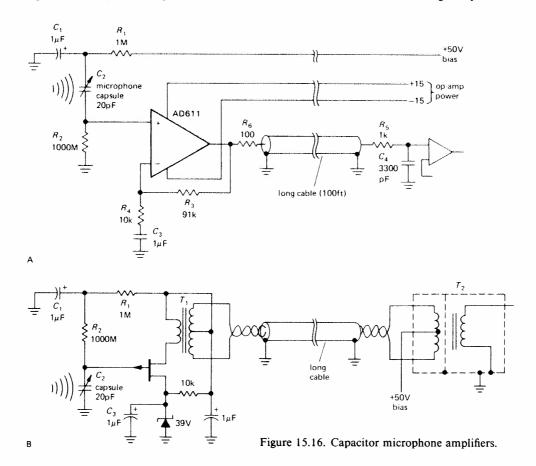
Γ C₁
1μF

В

importance, since many of the best recording microphones are capacitive position transducers, made by supporting a thin metallized plastic foil in close proximity to a fixed plate. You charge the capacitor through a large resistor with a bias of 50 to 100 volts, and you look at the changes of voltage as the diaphragm moves in the sound field.

Capacitor microphones have enormously high source impedances (a typical capsule has about 20pF of capacitance, or a reactance of about 400M at 20Hz), which means that you don't have a chance of running the signal through any length of cable whatsoever without putting a preamp right at the capsule. Figure 15.16 shows

two ways to buffer the voltage from the capsule, which might have an amplitude of 1mV to 100mV (rms) for typical ranges of program material. In the first circuit, a low-noise FET op-amp provides 20dB of gain and the low impedance necessary to drive a single-ended shielded line. Since the amplifier has to be located close to the microphone capsule (within a few inches), it is necessary to supply the operating voltages (bias for the capsule, as well as opamp power) through the microphone cable, in this case on additional wires. Note the trick of floating the microphone capsule in order to simplify biasing of the op-amp. R_1 and C_1 filter the bias supply, and R_2 must be chosen to have a high impedance



compared with the capsule at all audio frequencies. R_5 and C_4 form an RF filter, since the line is unbalanced and therefore prone to radiofrequency interference.

There are a few bad features about this circuit. It requires a 4-wire cable, rather than the industry-standard shielded pair. Also, the floating capsule can create mechanical problems. These drawbacks are remedied in the second circuit, in which the capsule bias is sent on the same lines as the audio itself, which happens to be a balanced 200 ohm pair. One side of the capsule is grounded, and a p-channel JFET \Box is used as a source follower to drive a small audio transformer. A single-ended output is recovered at the far end, where bias is applied via the transformer's center tap. Some would complain that the proliferation of transformers is a poor idea, but in practice they perform admirably.

- Angle. It is possible to convert angles to electronic signals with pretty good precision. There are angular versions of the LVDT, for instance, and popular devices known as resolvers. In both cases you use an ac excitation, and you can easily measure angular position down to an arc-minute. With great care it is possible to measure angles at the arc-second level. There are other techniques, e.g., using light beams looking at a glass disc with gray coded radial stripes.
- Interferometry. Highly accurate position measurements can be made by bouncing laser beams off mirrors attached to the object and counting interference fringes. The ultimate accuracy of such methods is set by the wavelength of light, so you have to work hard to do much better than a half micron (1 micron, or micrometer, is 1/1000 of a millimeter, or 1/25000 of an inch). An example of a commercial laser measurement instrument is the 5527A from Hewlett-Packard, which claims resolution of a microinch or better. Laser

interferometer systems are now used routinely in surveying, in flatness measurements, and in various tasks around research laboratories.

The most precise distance measurements have been made interferometrically by Deslattes at the National Bureau of Standards. Deslattes is a real wizard when it comes to precise physical measurement, routinely measuring spacings to milliangstroms (one ten-millionth of a micron) and angles to milliseconds of arc.

Quartz oscillators. A quartz crystal responds to deformation with a change in its resonant frequency, thus providing a very accurate method for measuring small displacements or changes in pressure. Quartz-oscillator pressure transducers provide the highest resolution presently available (more about this later).

15.04 Acceleration, pressure, force, velocity

The techniques just mentioned allow you to measure acceleration, pressure, and force. Accelerometers consist of strain gauges attached to a test mass, or capacitance-sensing transducers that sense the change of position of the test mass. There are various tricks to damp the system to prevent oscillation, in accelerometers that simply measure the displacement of the test mass to provide an output signal; alternatively, some systems use feedback to prevent the test mass from being displaced relative to the body of the accelerometer, the amount of applied feedback force then being the accelerometer's output signal.

LVDT, strain gauges, capacitance transducers, and quartz oscillators are used for pressure measurements, along with special devices such as a Bourdon gauge, a spiral hollow quartz tube that unwinds when inflated. LVDT transducers, for example, are available with full-scale ranges going

from 1 psi crystal osc resolution able from deliver ac of 0.001% pressure § sensitivity psi.

LVDT measure f the displa Full-scale to 250 to accuracies oratory me will find qu trostatic b esting exa gravimete and Warbi sphere le weight by balanced 1 static sens measure c one part in metric pre effect of ti on local gr

Magnetic 1

The positi talking abc of velocity tive of pos make a dir ploiting the a loop of v field is proj magnetic fl velocity-methat consis netic rods i

Much m velocity tradustry: m

15.04 Acceleration, pressure, force, velocity

from 1 psi to 100,000 psi or more. Quartz-crystal oscillator types provide the highest resolution and accuracy. The types available from Paroscientific, for example, will deliver accuracies of 0.01% and stabilities of 0.001%. Hewlett-Packard has a quartz pressure gauge with 11,000 psi full-scale sensitivity and claimed resolution of 0.01 psi.

LVDT transducers are often used to measure force or weight, although any of the displacement techniques can be used. Full-scale sensitivities go from 10 grams to 250 tons for one popular series, with accuracies of 0.1%. For highly precise laboratory measurements of small forces, you will find quartz-fiber torsion balances, electrostatic balances, and the like. An interesting example of the latter is the clever gravimeter developed by Goodkind and Warburton. It uses a superconducting sphere levitated approximately to zero weight by a persistent magnetic field, then balanced the rest of the way by electrostatic sensing and levitating plates. It can measure changes in gravitational field of one part in a billion, and it easily sees barometric pressure variations because of the effect of the changing overhead air mass on local gravity!

Magnetic velocity transducers

The position transducers we have been talking about can also be used to keep track of velocity, which is just the time derivative of position. However, it is possible to make a direct velocity measurement by exploiting the fact that the voltage induced in a loop of wire moving through a magnetic field is proportional to the rate of change of magnetic flux linked by the loop. There are velocity-measurement gadgets available that consist of long coils of wire with magnetic rods moving through the central bore.

Much more prevalent are the magnetic velocity transducers used in the audio industry: microphones (and their inverse,

loudspeakers), phono cartridges, and analog tape recorders. These devices typically generate signals at very low levels (a few millivolts is typical), and they present unique and interesting circuit challenges. For high-quality sound you have to keep noise and interference down 60dB or more, i.e., at the microvolt level. Since these signals get piped around over large distances in recording studios and radio stations, the problem can become serious.

Figure 15.17 shows how low-level signals from microphones and phono cartridges are usually handled. A dynamic microphone is a loudspeaker in reverse: A coil moves in a magnetic field, propelled by the sound pressure. Typically these things have output impedances of 200 ohms, with signals of $50\mu V$ to 5mV (rms) for quiet speech and concert-hall sound levels, respectively. For any significant length of connecting cable, you always use a balanced and shielded twisted-pair, terminated in the industry-standard Cannon XLR 3-pin audio connector.

At the far end you transform to a terminated impedance level of about 50k with a high-quality audio matching transformer, as indicated. Signal levels are then in the range of 1mV to 100mV (rms) and should be amplified by a low-noise preamp, as shown. Although you will see preamps with 40dB front-end gain, for good overload performance it is best to stick with a gain of 20dB. This is especially true for popular-music recording, where singers often wind up hollering into the microphone at close range.

The use of balanced 200 ohm microphone cable pretty much eliminates interference because of its good common-mode rejection. Good audio transformers for this kind of application have electrostatic shielding between the windings, which further reduces sensitivity to RF pickup. If radiofrequency interference is not suppressed enough with this scheme, as may be the case near transmitting

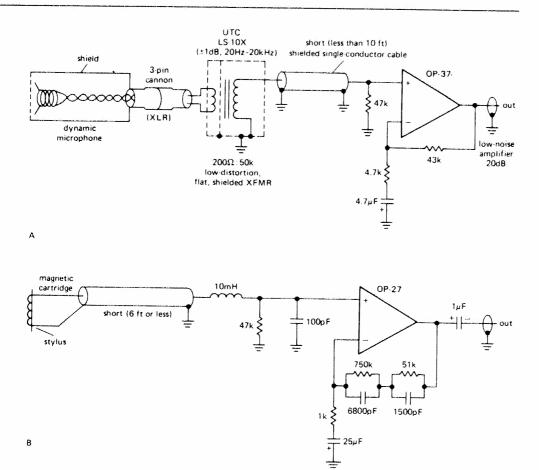


Figure 15.17. Dynamic microphone and phono cartridge amplifiers.

stations, you can add a low-pass filter at the preamp input. A 1k resistor or small RF choke in series at the input, followed by a 100pF capacitor to ground, will usually tame the beast.

Phono cartridges don't require balanced lines, because the cable run to the amplifier is usually very short. The standard method is simply to use single-conductor shielded cable, terminated with the 47k to ground that the cartridge requires for proper frequency response (Fig. 15.17B). We have also shown an input filter to reduce RF interference, because that is such

a common problem in urban areas. RF signals at the inputs of audio equipment pose a particularly insidious problem, because the audio amplifier's nonlinearities at radiofrequencies produce rectification, with consequent interference (audio detection) and distortion. When designing RF filters, be sure to keep the load capacitance small (a maximum of 300pF, including cable capacitance), since otherwise the cartridge's frequency response is changed. The series impedance should not have a resistance greater than a few hundred ohms, to keep the noise low. Quite large values of

inductance cartridge' ry. The standard ing in the

15.05 Ma

Accurate : important nection w magnetic netrons, n vices, etc. ing. For a Hall-eff€ effect is the age in a c ally a sem and comm cover a rar scale. To s field is abo a strong po sand gauss pensive, si example, t is a comp a 3-termin power it fr dc voltage of applied markets a UG3500/3 also used to and panel s tion 9.04.

A meth roots into t turn coil of magnetic firply pulled c voltage or tively. A fl it has the el theory, but old-fashion.

inductance can be safely used, since the cartridge's inductance is typically 0.5 henry. The amplifier circuit shown has the standard RIAA response used for recording in the United States.

15.05 Magnetic field

Accurate magnetic-field measurements are important in the physical sciences, in connection with instrumentation that uses a magnetic field (magnetic resonance, magnetrons, magnetically focused electron devices, etc.), and in geology and prospecting. For measurements at the 1% level, a Hall-effect probe is adequate. The Hall effect is the production of a transverse voltage in a current-carrying conductor (usually a semiconductor) in a magnetic field, and commercial Hall-effect magnetometers cover a range of about 1 gauss to 10kG fullscale. To give an idea of scale, the earth's field is about 0.5 gauss, whereas the field of a strong permanent magnet is a few thousand gauss. Hall magnetometers are inexpensive, simple, small, and reliable. For example, the inexpensive TL173 from TI is a complete linear Hall-effect sensor in a 3-terminal plastic TO-92 package. You power it from +12 volts, and out comes a dc voltage that increases 1.5mV per gauss of applied magnetic field. Sprague also markets a line of linear Hall sensors (the UG3500/3600 series). The Hall effect is also used to make noncontacting keyboard and panel switches, as we remarked in Section 9.04.

A method with considerably greater roots into the past is the flip coil, a multi-turn coil of wire that is either rotated in the magnetic field at some fixed speed or simply pulled out; you measure the induced ac voltage or the integrated current, respectively. A flip coil is simplicity itself, and it has the elegance of pure electromagnetic theory, but it tends to be a bit bulky and old-fashioned looking.

For measurements of minuscule magnetic fields you can't beat the exotic SQUID (superconducting quantum interference device), a clever arrangement of superconducting junctions that can easily measure a single quantum of magnetic flux (0.2 microgauss-cm²). A SQUID can be used to measure the magnetic fields set up in your body when you drink a glass of water, for whatever that's worth. These are fancy devices that require a considerable investment in cryogenic hardware, liquid helium, etc., and shouldn't be considered ordinary circuit items.

For precision magnetic-field measurements in the kilogauss range you can't do better than an NMR (nuclear magnetic resonance) magnometer, a device that exploits the precession of nuclear (usually hydrogen) spins in an external magnetic field. This is the physicist's magnometer, and it effortlessly yields values of magnetic field accurate to a part in a million or better. Since the output is a *frequency*, all the precision of frequency/time measurements can be used (more about this later).

Devices known as flux-gate magnetometers and transductors provide yet another way to measure magnetic fields. They work by exciting a piece of ferrite with an ac excitation field, with the response, as modified by the ambient field, being observed.

15.06 Vacuum gauges

The measurement of vacuum presents no great obstacles, which is fortunate, since it is a quantity of great importance in processes such as transistor and IC manufacture, thin-film evaporation, and the preparation of freeze-dried coffee. The basic device here is the Bayard/Alpert ionization gauge, which looks like an inside-out vacuum tube (Fig. 15.18). A hot filament emits electrons that are collected at a positive wire anode. Along the way the electron beam scatters from residual gas molecules, creating positive ions that are

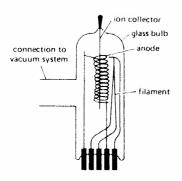


Figure 15.18. Ionization gauge.

collected at a central wire electrode held near ground. The ion current is accurately proportional to the density of gas molecules, i.e., the pressure. Ion gauges are usable at pressures (vacuums!) from about 10^{-3} to 10^{-11} mm Hg (1mm Hg is also known as 1 torr; atmospheric pressure is 760mm Hg). It takes great care to maintain a vacuum of 10^{-10} mm Hg; even a fingerprint on the side of the chamber will frustrate your efforts.

At more mundane levels of vacuum (1mm Hg down to 1μ m Hg, which you get with mechanical "roughing" pumps) the popular choice of measurement device is the thermocouple bonded to a small heater; you run some current through the heater, then measure the temperature with the thermocouple. Residual gas cools the contraption, lowering the thermocouple's output voltage. Thermocouple gauges are usually used so that you know when it's safe to turn on the high-vacuum (diffusion or ion) pumps. Granville-Phillips has an improved heat-loss gauge that works on similar principles. It's called a "convectron" gauge and measures from 1 µm Hg all the way up to atmospheric pressure.

15.07 Particle detectors

The detection, identification, spectroscopy, and imaging of charged particles and energetic photons (x rays, gamma rays)

constitute an essential part of the fields of nuclear and particle physics, as well as numerous fields that make use of radioactivity (medical radiography tracers, forensic science, industrial inspection, etc.). We will treat x-ray and gamma-ray detectors first, then charged-particle detectors.

X-ray and gamma-ray detectors

The classic uranium prospector was a slightly grizzled and shriveled character who went poking around the desert with clicking Geiger counter in hand. The detector situation has now improved considerably. These detectors all have in common the property that they use the energy of an incoming photon to ionize an atom of something, giving off an electron via the photoelectric effect. What they do with the electron depends on the particular detector.

lonization chamber, proportional counter, Geiger counter. These detectors consist simply of a cylindrical (usually) chamber, typically a few inches in size, with a thin wire running down the center. They are filled with some gas or mixture of gases. There's a thin "window" on one side, made of some material that the desired x rays can penetrate (plastic, beryllium, etc.). The central wire is held at a positive potential and connected to some electronics. Figure 15.19 shows a typical configuration.

When an x ray enters, it ionizes an atom by ejecting a photoelectron, which then loses energy by ionizing gas atoms until it is brought to rest. It turns out that the electron loses about 20 volts of energy per electron-ion pair it creates, so the total free charge left after the photoelectron is brought to rest is proportional to the x ray's initial energy. In an ionization chamber that charge is collected and amplified by a charge-sensitive (integrating) amplifier, just as with a photomultiplier. Thus, the output pulse is proportional to

the x ray's of the control of the co

gas-filled (87% Ar 10% Xe 3% CO₂)

0.001" diamete tungsten wire runs down axis

Figure 15.15

As you v device kno lets you co of assorted pulse heigh ergy, you w Thus, with with a Geiq energy spe-

These g the energy about 100k an energy 5.9keV (a provided be inexpensiv



the x ray's energy. The proportional counter works the same way but with the central wire held at a higher voltage, so that electrons drawn toward it cause additional ionization, resulting in a larger signal. This charge-multiplication effect makes proportional counters useful at low x-ray energies (down to a kilovolt, or less) where an ionization counter would be useless. In a Geiger counter the central wire is at a high enough voltage that any amount of initial ionization causes a single large (fixed-size) output pulse. This gives a nice large output pulse, but in the process you lose all information about the x ray's energy.

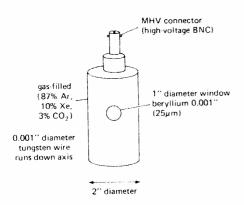


Figure 15.19. Proportional counter.

As you will see in Section 15.16, a clever device known as a pulse-height analyzer lets you convert an input stream of pulses of assorted heights into a histogram. If the pulse heights are a measure of particle energy, you wind up with an energy spectrum! Thus, with a proportional counter (but not with a Geiger counter) you are doing x-ray energy spectroscopy.

These gas-filled counters are usable in the energy range from about 1keV to about 100keV. Proportional counters have an energy resolution of about 15% at 5.9keV (a popular x-ray calibration energy provided by decay of iron 55). They're inexpensive and can be made in very large

or very small sizes, but they require a well-regulated power supply (the multiplication rises exponentially with voltage) and are not terribly fast (25,000 count/s is a rough practical maximum counting rate).

Scintillators. Scintillators work by converting the energy of the photoelectron, Compton electron, or electron-positron pair to a pulse of light, which is then detected by an attached photomultiplier. A popular scintillator is crystalline sodium iodide (NaI) doped with thallium. As with proportional counters, the output pulse is proportional to the incoming x-ray (or gamma-ray) energy, which means that you can do spectroscopy, with the help of a pulse-height analyzer (see Section 15.16). Typically, an NaI crystal will give an energy resolution of about 6% at 1.3MeV (a popular gamma-ray calibration energy provided by decay of cobalt 60) and is usable in the energy range of 10keV to many GeV. The light pulse is about 1μ s long, making these detectors reasonably fast. NaI crystals come in various sizes up to a few inches; they absorb water, though, so they have to be sealed. Since you must keep light out anyway, they're usually supplied in a metal package with thin aluminum or beryllium entrance window and integral photomultiplier tube.

Plastic (organic) scintillator materials are also popular, being very inexpensive. They have poorer resolution than sodium iodide and are used primarily at energies above 1MeV. Their light pulses are very short, roughly 10ns. Liquid scintillation "cocktails" are routinely used in biological studies. In such applications the material being examined for radioactivity is mixed into the scintillator cocktail, and the whole works is put into a dark chamber with a photomultiplier. You'll see handsome instruments in biology labs that automate the whole process, passing one vial after another through the counting chamber and recording the results.

Solid-state detectors. As with the rest of electronics, the great revolution in xray and gamma-ray detection has come about through advances in silicon and germanium technology. "Solid-state" detectors work just like the classic ionization chamber, but with the active volume filled with a nonconducting (intrinsic) semiconductor. An applied potential of about 1000 volts sweeps the ionization out, generating a pulse of charge. In silicon, an electron loses only about 2eV per electron-ion pair created, so many more ions are created for the same incident x-ray energy, as compared with a gas-filled proportional detector, giving better energy resolution through improved statistics. Other subtle effects also contribute to improved performance.

Solid-state detectors come in several varieties, Si(Li), Ge(Li), and intrinsic germanium, or IG (the first two are pronounced "silly" and "jelly"), according to the semiconductor material and dopants used to make it insulating. They are all operated at liquid-nitrogen temperature (-196°C), and the lithium-drifted types must be kept cold at all times (if allowed to warm up, they decay, permanently, with about the same time constant as fresh fish). Typical Si(Li) detectors come in diameters from 4mm to 16mm and are usable for x-ray energies from about 1keV to 50keV. Ge(Li) and IG detectors are used at higher energies, 10keV to 10MeV. Good Si(Li) detectors have energy resolutions of 150ev at 5.9keV (2.5%, six to eight times better than proportional counters), and the germanium detectors have energy resolutions of about 1.8keV at 1.3MeV (0.14%).

In order to illustrate what that extra resolution buys you, we bombarded a random hunk of stainless steel with 2MeV protons and measured the x-ray spectrum produced. This is called PIXE (protoninduced x-ray emission), and it is a powerful technique for determining spatially resolved trace-element distribution. Figure 15.20 shows the energy spectrum (made

with a pulse-height analyzer), with two x-ray lines visible for each element, at least with the Si(Li) detector. You can see iron, chromium, and nickel. A few additional elements are visible if you expand the lower part of the graph. With the proportional counter, all you get is mush.

Figure 15.21 shows the same kind of comparison for gamma detectors. This time it's an NaI scintillator versus a Ge(Li). (We ran out of steam, so we cribbed this one from the friendly folks at Canberra Industries. Many thanks, Mr. Tench.) As before, solid-state detectors win, hands down, for resolution.

Solid-state detectors have the best energy resolution of all the x-ray and gammaray detectors, but they have the disadvantage of a small active area in a large clumsy package (see Fig. 15.22 for an example), relatively slow speed (50 µs or longer recovery time), high price, and high nuisance value (unless you enjoy being a full-time baby-sitter to a liquid-nitrogen guzzler).

Charged-particle detectors

The detectors we've just described are intended for energetic photons (x rays and gamma rays), not particles. Particle detectors have somewhat different incarnations; in addition, charged particles are deflected by electric and magnetic fields according to their charge, mass, and energy, making it much easier to measure particle energies.

Surface-barrier detectors. These germanium and silicon detectors are the analogs of the Ge(Li) and Si(Li) detectors. They don't have to be cooled, which simplifies the packaging enormously. (It also lets you take an occasional vacation!) Surfacebarrier detectors are available in diameters from 3mm to 50mm. They are usable at particle energies of 1MeV to hundreds of MeV, and they have energy resolutions of 0.2% to 1% for 5.5MeV alpha particles (a



Figure 15.2 counter and

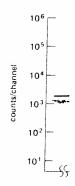


Figure 15.21 detector. [Fr

popular alp by decay of

Cerenkov energies (10

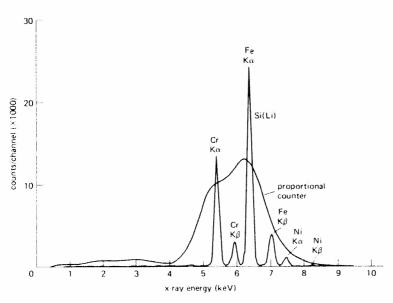


Figure 15.20. An x-ray spectrum from a piece of stainless steel, as seen by an argon proportional counter and a Si(Li) detector.

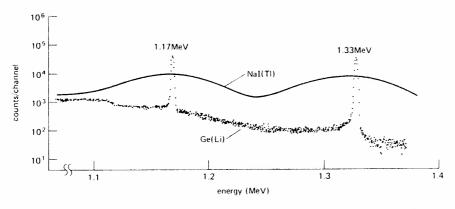


Figure 15.21. Cobalt-60 gamma-ray spectrum, as seen by a sodium iodide scintillator and a Ge(Li) detector. [From Canberra Ge(Li) Detector Systems brochure, Canberra Industries, Inc.]

popular alpha calibration energy provided by decay of americium 241).

 Cerenkov detectors. At very high energies (1GeV and above) even a heavy charged particle can outrun light in material media, giving rise to Cerenkov radiation, a "visible sonic boom." They are used extensively in high-energy physics experiments.

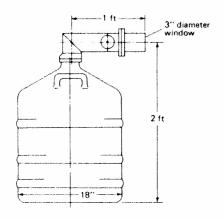


Figure 15.22. Ge(Li) cryostat. (Courtesy of Canberra Industries, Inc.)

- Ionization chambers. The classic gasfilled ionization chamber previously described in connection with x-ray detection can also be used as a detector of energetic charged particles. In its simplest form it consists of a single collecting wire running the length of an argon-filled chamber. Depending on the particle energies involved, the chamber may range from inches to feet in size; variations include the use of multiple collecting wires or plates, and other filling gases.
- ☐ Shower chambers. A shower chamber is the electron equivalent of an ionization chamber. An energetic electron enters a box of liquid argon, where it generates a "shower" of charged particles that are subsequently collected at charged plates. High-energy physicists like to call these things "calorimeters."
- Scintillation chambers. A charged particle can be detected with very good energy resolution by using photomultipliers to detect the ultraviolet-rich scintillations caused by the particle's ionized path in a chamber filled with argon or xenon in gas or liquid form. Scintillation chambers are

delightfully fast, in contrast with the more leisurely response of ionization and shower chambers.

Drift chambers. These are the latest rage in high-energy physics, and they are made possible by advances in high-speed on-line computing. They're simple in conception: a box filled with gas at atmospheric pressure (an argon-ethane mixture is typical) and crisscrossed by hundreds of wires with an applied voltage. The box is full of electric fields, and when a charged particle goes in and ionizes the gas, the ions are swept out by the array of wires. You keep track of the signal amplitudes and timing on all the wires (that's where the computer comes in), and from that information you deduce the particle's path. With an applied magnetic field, that also tells you the momentum.

The drift chamber has become the universal imaging charged-particle detector for high-energy physics. It can deliver spatial resolution of 0.2mm or better over a volume large enough for you to climb into.

15.08 Biological and chemical voltage probes

In the biological and chemical sciences there are many examples of measurement wizardry: electrochemical methods such as electrochemistry with ion-specific electrodes, electrophoresis, voltametry, and polarography, as well as techniques like chromatography, IR and visible spectroscopy, NMR, mass spectroscopy, x-ray spectroscopy, nuclear quadruple spectroscopy, ESCA, etc. It is hopeless in a volume this size to attempt any kind of comprehensive catalog of these sophisticated techniques. Furthermore, these techniques can be characterized as less fundamental than the direct physical measurements cataloged earlier in this chapter.

In order to give an idea of the special problems that arise in chemical and biological only the sindeterminate by a micro and muscle by an ionsure the coionic specimetric elethere are selenges you meaningfu

Microelect

In order to or in the practice to a few hun (1 angstronsize of a hybe easily defined with interfrom the cond 100 MΩ of loading by rolloffs of capacitance

In orde you want performar hertz, or frequency The amy input imp noise. In to commo

The cir good solut trode con measurem pearing as nals. Th as possibl by low-no IC₂, whic

15.08 Biological and chemical voltage probes

biological measurements, we will describe only the simplest sort of measurement: the determination of the potentials generated by a microelectrode (used to explore nerve and muscle signals in biological systems), by an ion-specific electrode (used to measure the concentration of some specific ionic species in solution), and by a voltametric electrochemical probe. As usual, there are some interesting electronic challenges you face, if you want to get anything meaningful out of your measurement.

Microelectrodes

In order to look at the voltage on nerves or in the interiors of cells, it is standard practice to make electrodes that are just a few hundred angstroms in tip diameter (1 angstrom = 10^{-8} cm, approximately the size of a hydrogen atom). That turns out to be easily done by drawing a glass capillary, then filling it with a conductive solution. You wind up with a nice probe, but with interesting circuit problems arising from the electrode's source impedance of $100 M\Omega$ or more. Interference pickup, loading by the circuit, and high-frequency rolloffs of a few hertz due to cable and stray capacitances plague the unwary.

In order to see nerve or muscle signals you want to have decent high-frequency performance, at least out to a few kilohertz, or so (this isn't exactly high frequency in the sense of Chapter 13!). The amplifier must have very high input impedance, and preferably low input noise. In addition, it must be insensitive to common-mode interference.

The circuit in Figure 15.23 represents a good solution. The use of a reference electrode connected near the point of actual measurement keeps interference from appearing as normal (differential) mode signals. The inputs are buffered as close as possible to the actual microelectrode by low-noise FET-input op-amps IC₁ and IC₂, which also bootstrap the guard elec-

trode in order to reduce the effective cable capacitance. Note that the guard is itself shielded. You've got to use FET amplifiers in order to get high input impedance and low input current noise; the particular types shown were chosen for their low input noise voltage $(2.5\mu V)$ pp max 0.1-10Hz), often a problem with FET or MOSFET amplifiers. The pair of buffered signals is applied to the standard differencing amplifier configuration with a low-noise-voltage low-drift op-amp, with 100mV of stable adjustable output offset added via IC₆.

At this point you have an amplifier with a differential gain of 10, suitable noise performance, good common-mode rejection, and low input current (<1pA). However, even with input guarding, the residual input capacitance at the input buffers and at the microelectrode tip will result in poor speed performance. For example, a $100M\Omega$ source impedance driving 20pF has a high-frequency 3dB point of only 80Hz. The solution is active compensation via positive feedback, provided by IC_3 and IC_4 through C_1 and C_2 . In practice, you adjust the voltage gain of amplifiers IC₃ and IC₄ for decent high-frequency performance (or transient response), with response to several kilohertz possible.

ion-specific electrodes

The classic example of an ion-specific electrode is the pH meter, which measures the voltage developed between a reference electrode and a thin-walled glass electrode through which hydrogen ions can diffuse. Once again you're dealing with very high source impedances, although the problems here are less severe than with microelectrodes, because you don't often care about frequency response.

There are more than 20 kinds of ionspecific electrode systems available, e.g., to measure activities of K⁺, Na⁺, NH₄⁺,

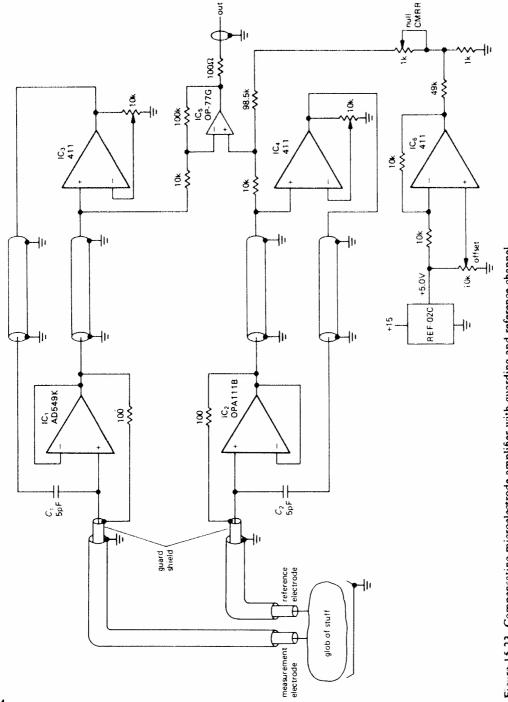


Figure 15.23. Compensating microelectrode amplifier with guarding and reference channel.

CN⁻, Hg Ca⁺⁺, or electrodes silver-coat mersed in sium chlo solution y plug or ge typically mersed in ion you're the soluti that is se of interes an ion-se uid contai ganic mol a voltage volts, with drawing le complicat of as mu voltage p can attem thermisto Conversion to concen total ioni the crossc electrode case, cher this sort some stan ter makin you can s achieve n 1% in solu

Electroch

In the are ble to ma surement: ions by m tion rates tion. By § pass thro cific react

CN⁻, Hg⁺⁺, SCN⁻, Br⁻, Cl⁻, F⁻, I⁻, Ca⁺⁺, or Cu⁺⁺. In general, you have two electrodes: a reference electrode, typically silver-coated with silver chloride and immersed in a concentrated solution of potassium chloride that communicates with the solution you want to measure via a porous plug or gel, and an ion-specific electrode, typically consisting of an electrode immersed in a concentrated solution of the ion you're interested in and separated from the solution under test by a membrane that is selectively permeable to the ion of interest. The membrane is commonly an ion-selective glass or an organic liquid containing mobile ion-transporting organic molecules. Your task is to measure a voltage that is in the range of 0 to 2 volts, with an accuracy of a millivolt, while drawing less than 100pA. The situation is complicated by a temperature coefficient of as much as a few percent change in voltage per degree centigrade, which you can attempt to cancel automatically with thermistor-driven compensation circuitry. Conversion from measured ionic activity to concentration requires attention to the total ionic strength of the sample and to the crossover sensitivity of the ion-specific electrode to other ions present. In any case, chemists say you get best results with this sort of black art if you calibrate on some standard solutions just before and after making your measurements. With care, you can see concentrations of 0.1ppm and achieve measurement accuracies of about

Electrochemical measurements

In the area of electrochemistry, it is possible to make very sensitive analytical measurements of the concentrations of specific ions by measuring electrode currents (reaction rates) versus applied voltage in a solution. By scanning the applied voltage, you pass through the potentials at which specific reactions occur, giving rise to steps or

1% in solutions of moderate concentration.

peaks. Terms such as cyclic voltametry, polarography, and anodic stripping voltametry (ASV) are used to describe various ways of doing such analytical measurements. Among the most sensitive of these techniques is ASV, which uses a hanging drop mercury (hdm) electrode, a renewable electrode onto which you electroplate at a relatively high potential for a while, then reverse the current and strip off each element sequentially. This technique can detect elements like lead and cadmium at the parts-per-billion level, and it should be considered on a par with other traceelement techniques such as neutron activation, flame spectroscopy, and x-ray and ion microprobes.

The technique of measuring a small current while subjecting a system to a fixed voltage is called a "voltage clamp," and it finds application also in nerve and cell physiology. Nerve membranes have voltage-dependent channels through which specific ions can diffuse, and nerve physiologists like to measure the voltages at which such channels open. Again, voltage clamps are used, this time with microelectrodes.

In preparative electrochemistry, the same techniques are used, but with currents measured in amperes rather than microamperes. Once again the idea is to drive a specific reaction product by applying the right voltage.

Figure 15.24 shows a simple potentiostat (or voltage clamp) circuit. The electrolytic cell consists of an electrode to inject current (the counter electrode), a common return electrode (the working electrode), and a small probe to measure the voltage in the solution near the working electrode (the reference electrode). IC₁ maintains a voltage equal to V_{ref} between the reference and working electrodes by varying the current into the counter electrode appropriately (in measurements of membrane potential, the upper two electrodes would be inside the cell, and the working electrode outside). IC₂ holds the

1016 Chapter 15

working electrode at virtual ground, converting the current to an output voltage. The range of voltages encountered is typically ±1 volt; currents in the range of InA to ImA are typical of analytical measurements, whereas currents of 1mA to 10 amps are used in preparative electrochemistry.

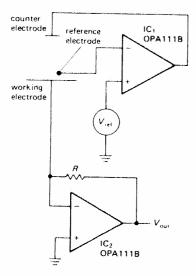


Figure 15.24. Potentiostatic electrochemistry circuit (voltage clamp).

In order to do scanning measurements, $V_{\rm ref}$ would be replaced with a ramp generator. For low-current membrane measurement you would have to shield the input leads carefully, and perhaps use guarding and positive feedback with a capacitor in the manner of Figure 15.23 to maintain some frequency response.

PRECISION STANDARDS AND PRECISION MEASUREMENTS

In Chapter 7 we talked about the circuit methods that are necessary in highaccuracy instruments to maintain small voltage offsets and drifts, e.g., when amplifying very small voltages. We dealt there

only with analog electronics, the amplification of continuously varying voltages and currents. For a number of reasons, it turns out that measurements of digital quantities such as frequency, period, and time intervals can be made to far greater precision than any analog measurement. In the following sections, we will explore the accuracy of electronic standards (time, voltage, resistance), and you will see how to make analog measurements of high precision using such standards as references. We will devote the bulk of the discussion to time/frequency measurements, because of their greater inherent precision and because we have already treated precision analog circuitry in some detail in Chapter 7.

☐ 15.09 Frequency standards

Let's take a look at the high-stability frequency standards you can get, and then discuss how you set and then maintain their frequency.

☐ Quartz-crystal oscillators

Back in Sections 5.12-5.19, we described briefly the stabilities you can expect from frequency standards, going from the simplest RC relaxation oscillator to the atomic standards based on rubidium and cesium. For any serious timing you wouldn't consider anything less stable than a quartzcrystal oscillator. Fortunately, the cheapest crystal oscillators cost only a few dollars and can deliver stabilities of a few parts per million. For about \$50 you can buy a good TCXO (temperature-compensated crystal oscillator), stable to one part per million from 0°C to 50°C. For better performance you need ovenized crystals, with price tags from a couple of hundred dollars to more than \$1000. Once you begin talking about stabilities of a few parts per billion, you have to worry also about "aging," the tendency of crystal oscillators to

drift in frec stant rate o in." The 10 ifies good c of better the full tempera less than 0.5

Uncompe even TCXO a small inst oscillators a dards with a

☐ Atomic star

There are 1 rubidium, c ium has a n 682,608Hz, 9,192,631,7 sorption at quency stan considerably pensive) tha

Rubidium. you have a por, heated cavity with ium lamp s a photocell Meanwhile, nal reference tor is introc lock-in dete transmitted crowave sig onance freq tion of the r microwave tal's freque way to the straightforw quency like several add have glosse Rubidiur

ity than ov though they



"The best self-teaching book and reference book in electronics... The beauty and fun of electronics Radio Communication shows through."

"Tve hunted high and low for good electronics textbooks. This one is the best, bar none. No extraneous math, lots of insider's information ... and a huge range of topics covered clearly and thoroughly. So CoEvolution Quarterly well written that I've had difficulty putting it down."

"Full of clever circuits and sharp insights, but with a surprising minimum of mathematics ... This scheme of making maximum use of the insights of simple approximations described in words, and introducing more formal design only after understanding has gained real headway, is the pedagogical core of this original treatment ... The depth is genuine, as is the richness of examples, data and apt tricks."

"Far and away the finest book on the subject of electronics... in the last decade. I cannot recommend this book highly enough to anyone whose research or experiments require some electronics." Optical Engineering

"This is a delightful book... an excellent teaching text as well as a lab reference... I recently observed a graduate student in physics unfamiliar with op-amps start to build a haywired mess of a circuit. He then took home The Art of Electronics for one night, and the next day the mess was taken apart, to Review of Scientific Instruments be replaced by a correctly designed working circuit."

"This book is filled with a tremendous diversity of valuable information. More importantly, this book is a lov to read ... it's not at all like studying – it's too much fun."

EDN (News Edition) Joy to read ... it's not at all like studying - it's too much fun."

Changes in the new edition:

- Totally rewritten chapters on microcomputers and microprocessors (using the IBM PC and the
- Substantially revised chapters on digital electronics (including PLDs, and the new HC and AC logic families), op-amps and precision design (reflecting the availability of excellent FET-input op-amps), construction techniques (including CAD/CAM).

 • Every table has been revised and many new ones added.
- A rewritten chapter on FETs and an all-new chapter on micropower design.
- New sections on active filter designs, switched-capacitor filters, quadrature oscillators, low-dropout regulators, switching supplies, crowbars, isolation amplifiers, SCR latchup, "ground bounce," dynamic power dissipation, optoelectronics, RS-232 Interfacing, moderns, memory chips, schematic capture, AM detection, battery characteristics, sensor linearization.

CAMBRIDGE UNIVERSITY PRESS

CAMB