ELECTRONIC COMPONENTS, INSTRUMENTS, AND TROUBLESHOOTING

DANIEL L. METZGER

O. CONTRERAS

ELECTRONIC COMPONENTS, INSTRUMENTS, AND TROUBLESHOOTING

DANIEL L. METZGER

Monroe County Community College
Monroe, Michigan

PRENTICE-HALL, INC., *Englewood Cliffs, New Jersey 07632*

CONTENTS

PREFACE

Here is a book which bridges the gap between electronics as it is taught in the classroom and electronics as it is practiced in the industry. The author, drawing upon his own experience in four different industries, shares with the reader a wealth of tips and insights which mark the difference between the inexperienced beginner and the practiced professional.

The approach is strictly practical—none of the nasty realities are omitted because they are not theoretically pretty, and none of the theoretical niceties are included unless there is a solid practical reason for them. Here you will find frank and numeric discussions of stray inductance and skin effect in hookup wire, leakage in various types of capacitors, turn-on surge in power transformers, and response times of various photodetectors. You will also find quick and easy methods for:

- statistical sampling of batches of components
- converting decibels to voltage ratios and back, instantly, in your head
- analyzing and designing attenuators and filters
- designing modern power supplies, including IC regulators, foldback short-circuit protection, and switching regulators
- troubleshooting electronic systems without getting more involved than you have to.

In scope, the book is designed for the person who has completed a study of basic electronics (including power supplies, amplifiers, oscillators, and transistor fundamentals), but who recognizes that there is a considerable gap between this elementary knowledge and current industry practice. The mathematical level of the discussions is held to a minimum. Basic algebra and right-angle trigonometry are adequate as prerequisites. The topical coverage is broken into three units:

- Unit I provides an in-depth look at the real properties of modern electronic components—from hookup wire to large-scale integrated circuits—gleaned from manufacturers' literature and from numerous lab tests revealing limitations which manufacturers often choose not to publish.
- Unit II starts with the basics of a course in electronic instrumentation and adds chapters on high-performance amplifiers, op-amp applications, and modern power-supply circuits.
- Unit III outlines an approach to troubleshooting that is specific as to methodology yet applicable to any type of equipment. Notes on selection and use of test equipment and wiring practices are included.

Special attention has been given to make the Table of Contents at the front and the Index at the back of this book as complete as possible. Take the time now to thumb through and become familiar with them. A tremendous amount of information has been compressed between these two covers, and if you learn to use the Index and Table of Contents well, you will find yourself reaching frequently for this volume as a reference in solving practical problems in the art of electronics.

It has been my good fortune to have the cooperation of numerous industries in preparing the material for this book. In addition to the acknowledgements appearing with figures throughout the text, I would like to express appreciation to the following individuals who reviewed portions of the manuscript in their areas of specialty, and in many cases made important corrections and additions: A. P. Gabriel, Alpha Wire Corporation, Chapter 1 on Wire and Cable; Jack R. Polakowski, Allen-Bradley Co., Chapter 2 on Resistors; F. Woody Brooking, Kemet Division of Union Carbide Corp., Chapter 3 on Capacitors; S. Vincent Burns, Caddell Burns Mfg. Co., Chapter 4 on Inductors; B. G. Merritt, Eveready Division of Union Carbide Corp., Sections 6.1 and 6.2 on Batteries; H. A. Fishel, S T Semicon, Inc., Chapter 7 on Diodes; Robert M. Mendelson, RCA Corporation, Chapter 8 on Transistors; Jack Takesuye, Motorola Semiconductor Group, Chapter 9 on Triggered Semiconductors; Kathleen Emery, Statistical Analyst, Blue Cross of Northwest Ohio, Chapter 10 on Statistics; Barry Siegel, National Semiconductor, Chapter 19 on Signal Processing.

Thanks are also due to three others who proofread the entire manuscript: Professor Emeritus Robert C. Carter of Atlanta, Georgia, whose mastery of both the Language and the technology might have given me an inferiority complex had he been less of a gentleman; Professor Leon Heselton of Fulton-Montgomery

Community College in Johnstown, New York, whose attention to detail uncovered several errors in mathematics and notation; and my good friend and colleague at Monroe County Community College, Tim Maloney, whose final reading of the page proofs eliminated errors that had escaped all others.

Readers having knowledge of errors or new developments in areas covered in the text are encouraged to communicate with the author through the Editor for Electronics Technology at Prentice-Hall, Englewood Cliffs, NJ 07632.

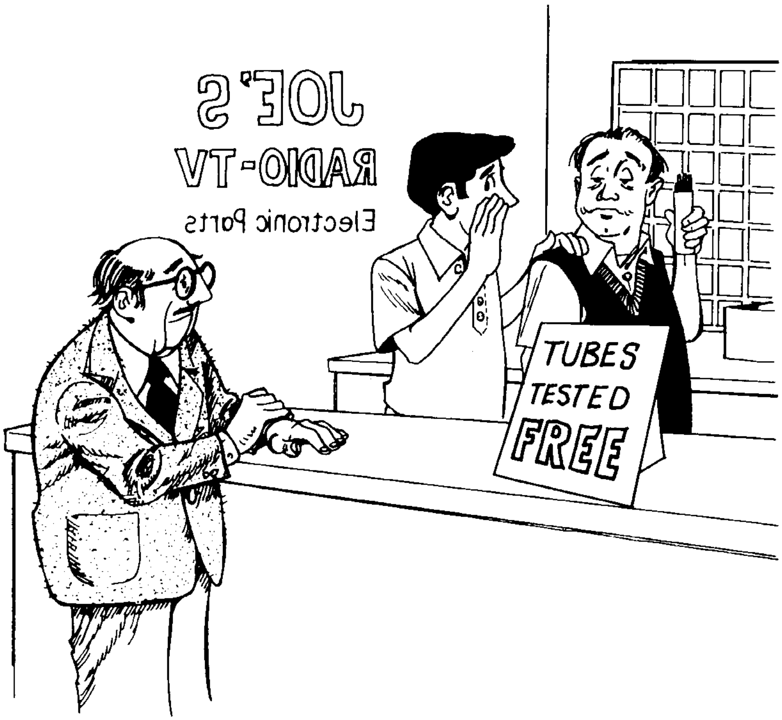
Finally, lest there be any misunderstanding, let me say that I am delighted at the recent entry of women in significant numbers to my chosen field. At the same time a respect for the grace of the Language prevents me from littering the text with *he and/or she* constructions. Until we develop a neutral personal pronoun, please understand *he* to refer to both sexes.

Daniel L. Metzger

Temperance, Michigan

PART 1

REAL VERSUS IDEAL COMPONENTS



"HE SAYS HE WANTS AN IDEAL VOLTAGE SOURCE, A ONE-FARAD CAPACITOR, AND A ONE-THOUSAND MHO CONDUCTOR. WHAT SHOULD I TELL HIM?"

Major Edwin Armstrong, who was undoubtedly the most prolific and down-to-earth of the radio "pioneers," was fond of repeating a saying which went something like this:

> *It ain't the things you don't know that get you into trouble; it's the things you know for sure—that ain't so.*

Armstrong went on to prove the wisdom of that proverb by singlehandedly inventing FM radio, even though the theoretical experts already "knew for sure" that FM radio wouldn't work.

Electronics technicians today often get into trouble because there are too many things that they "know for sure" about electronic components that in reality "ain't so." Ideally, for example, hookup wire has zero resistance, and a coil whose inductance is measured as 10 μH at 1 MHz will also be 10 μH at 10 MHz. Such idealizations are commonly applied in the practice of electronics, and indeed a good case could be made that the practice of the art as we know it would be impossible without such simplifying assumptions. Eventually, however, the technician is bound to run afoul of his assumptions and he will have to recall that his initial analysis was based on an idealized mathematical model of a certain device which cannot possibly reflect all the properties of the actual physical component.

Then he will have to ask himself: "What are the real properties of this component? How are they different from the mathematical model? Under what conditions is the model essentially valid, and when must I discard the model and look for the nonideal effects?" He may find that the near-zero resistance of hookup wire is not near-zero enough in the case of 5-A current pulses, and that the distributed capacitance in the windings of a 10-μH coil cancels out the majority of the coil's reactance at 10 MHz, even though it was negligible at 1 MHz.

The purpose of this unit is to give the technician some awareness of the nonideal effects in common electronic components. In many cases, these concepts will not be needed, as the behavior of the real component will be close enough to the ideal to make the idealized model valid. However, when an "impossible" situation comes up in troubleshooting—say a substantial voltage drop across a supposedly zero-resistance wire, or a transistor that refuses to saturate even though the beta formula says that it has more-than-enough base drive—then it will be time to remember the difference between the ideal and the real.

1

PROPERTIES OF REAL WIRE AND CABLE

1.1 CONDUCTIVITY OF WIRE

The actual dc resistance of various sizes of solid copper wire is given in Fig. 1-1. For resistance to ac, be sure to read Section 1.4 on skin effect. Stranded wire is sized to have a similar cross-sectional area, and hence equivalent dc resistance, as equivalent-sized solid wire. For example, 10 strands of No. 30 wire would have an area of 1000 circular mils, and would be called No. 20 stranded wire, since No. 20 wire has about the same area (1020 circular mils, to be exact).

Other Metals: The resistance of other metals relative to copper is given in Fig. 1-3. Notice, for example, that a length of aluminum wire has 1.64 times the resistance of a copper wire of the same length and diameter.

Circular Mils: The cross-sectional areas of the wires in Fig. 1-1 are given in *circular mils*. One circular mil is the area of a circle 1 mil in diameter. As illustrated in Fig. 1-2, circular-mil area is then simply the square of wire diameter in mils (thousandths of an inch).

$$A_{\mathrm{cmil}} = t_{\mathrm{mil}}^2 \tag{1-1}$$

Calculating Wire Resistance: If necessary, the resistance of a length of wire can be

Gage	Diameter at 20°C		Area at 20°C		Resistance at 2!	
AWG	Mil	cm	cmil	cm^2	Ω/1000 ft	Ω/m
0	324.9	0.825	105,600	0.535	0.100	0.00032
2	257.6	0.654	66,360	0.337	0.159	0.00052
4	204.3	0.519	41,740	0.212	0.253	0.00083
6	162.0	0.411	26,240	0.133	0.403	0.0013
8	128.5	0.326	16,510	0.0837	0.64	0.0021
10	101.9	0.259	10,380	0.0526	1.02	0.0033
12	80.8	0.205	6,530	0.0331	1.62	0.0053
14	64.1	0.163	4,110	0.0208	2.57	0.0084
16	50.8	0.129	2,580	0.01307	4.10	0.0134
18	40.3	0.102	1,620	0.00823	6.51	0.0214
20	32.0	0.081	1,020	0.00519	10.3	0.0339
22	25.3	0.064	640	0.00324	16.5	0.0542
24	20.1	0.051	404	0.00205	26.2	0.0859
26	15.9	0.040	253	0.00128	41.8	0.137
28	12.6	0.032	159	0.000800	66.6	0.219
30	10.0	0.025	100	0.000507	106	0.347
32	8.0	0.020	64	0.000320	165	0.542
34	6.3	0.0160	39.7	0.000201	266	0.874
36	5.0	0.0127	25.0	0.000127	423	1.39
38	4.0	0.0102	16.0	0.000080	661	2.17
40	3.1	0.0079	9.61	0.000049	1100	3.61

***FIGURE 1-1** Resistance and standard sizes of copper wire.*

found directly from the following formula:

$$R = \frac{\rho l}{A} \tag{1-2}$$

where R is the wire resistance in ohms, ρ the resistivity of the wire material in ohms·cmil/ft (Fig. 1-3), l the length of the wire in feet, and A the cross-sectional area in circular mils (square of diameter in mils, or from Fig. 1-1).

EXAMPLE 1-1

What is the dc resistance of 5280 ft of No. 26 stranded copper wire at 25°C?

Solution

From Fig. 1-1, each foot has 0.0418 ohm (Ω):

$$R = 5280 \text{ ft} \times 0.0418\ \Omega/\text{ft} = \mathbf{220.7\ \Omega}$$

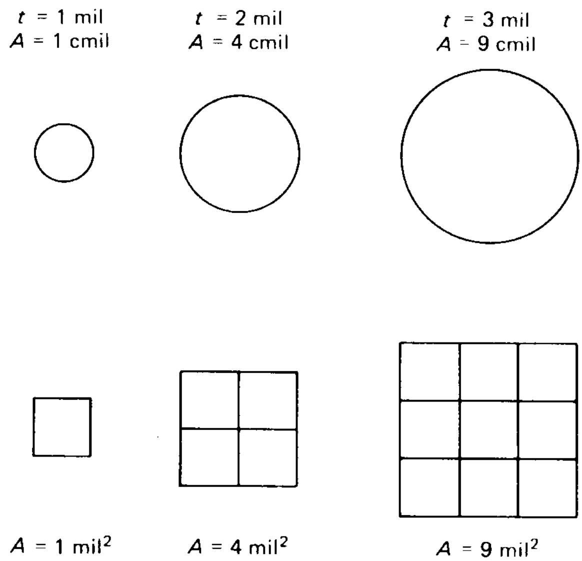


***FIGURE 1-2** Area in circular mils equals the square of wire diameter in mils (thousandths of an inch).*

Material	Resistivity Relative to Copper	Resistivity Ω·cmil/ft. at 20°C	Resistivity $\mu\Omega$·cm^2/cm at 20°C	Temperature Coefficient per °C at 20°C
Aluminum	1.64	17	2.83	+ 0.0040
Brass	3.58	37	6.17	+ 0.0015
Copper	1.00	10.37	1.724	+ 0.00393
Gold	1.42	14.7	2.44	+ 0.0034
Iron	5.59	58	9.64	+ 0.0055
Lead	11.86	123	20.4	+ 0.0039
Mercury	100	—	172	—
Nichrome	65.09	675	112	+ 0.00017
Silver	0.945	9.80	1.63	+ 0.0038
Steel (soft)	9.24	95.8	15.9	+ 0.0016
Tin	6.68	69.3	11.5	+ 0.0042
Tungsten	3.20	33.2	5.52	+ 0.005
Carbon (graphite)	450	—	775	− 0.0005

***FIGURE 1-3** Relative resistance of various conductors (copper = 1.000) and temperature coefficients of resistance.*

EXAMPLE 1-2

What is the resistance of 10 ft of No. 22 Nichrome wire?

Solution

From Fig. 1-1, 10 ft of No. 22 *copper* wire has a resistance of 0.165 Ω. From Fig. 1-3, Nichrome wire has 65.09 times the resistance of copper, so

$$R = 0.165\ \Omega \times 65.09 = \mathbf{10.7\ \Omega}$$

EXAMPLE 1-3

The diameter of a solid steel conductor measures 0.25 in. What is the resistance of 500 ft of this conductor?

Solution

$$A_{\text{cmil}} = d^2 = (250)^2 = 62{,}500 \text{ cmil}$$

$$R = \frac{\rho L}{A} = \frac{95.8 \times 500}{62{,}500} = \mathbf{0.766\ \Omega}$$

1.2 TEMPERATURE EFFECTS AND CURRENT CAPACITY OF WIRE

Figure 1-3 shows that the resistance of metal conductors increases at higher temperatures. Copper wire, for example, undergoes an increase of 0.393% for each degree Celsius rise in temperature. For a 100°C rise (the rise from freezing to boiling of water) this would amount to a resistance increase of 39%.

Temperature-Stable Alloys: For applications where highly temperature-stable resistances are required, special alloys (usually of copper with nickel) are available which have temperature coefficients below 0.00002 per °C. For such metals, resistances can be held stable to better than 0.1% over the normal range of human-environment temperatures.

Semiconductor Temperature Coefficients: Nonmetals, such as carbon, generally have a negative temperature coefficient of resistance, meaning that their resistance goes down as temperature goes up.

The semiconductor materials, "doped" silicon and germanium, have temperature coefficients which are large and negative, but are otherwise difficult to predict because of their extreme dependence on very small impurity contents. The resistance versus temperature curve for a typical slab of silicon semiconductor is given in Fig. 1-4. The rapid change of resistance with temperature is exploited in the manufacture of *thermistors* for temperature measurement and control.

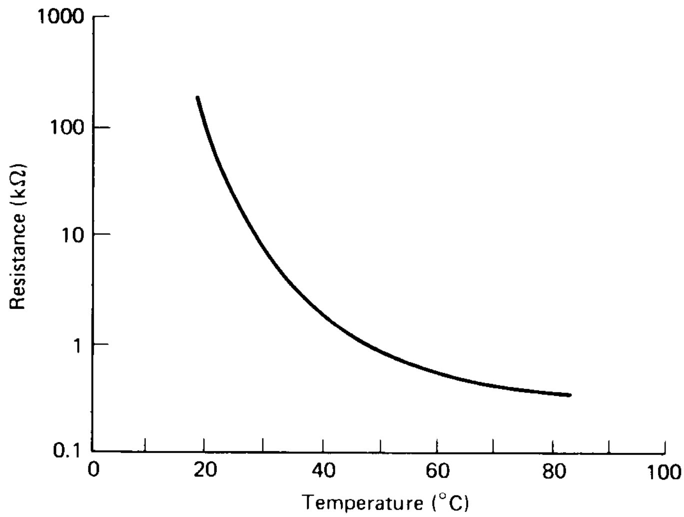
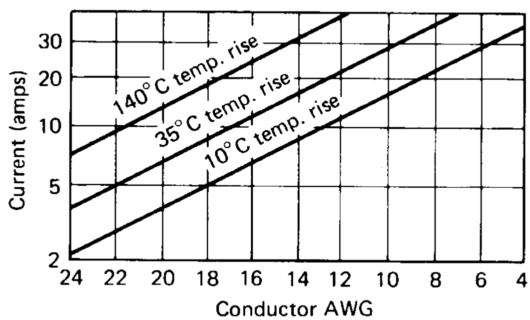


FIGURE 1-4 Resistance versus temperature for a typical silicon thermistor.

Maximum Current versus Wire Size: The recommended maximum currents for various sizes of copper wire may be obtained from Fig. 1-5. The lowest line (10°C rise) should be used for systems that must run cool. The center (35°C) should be thought of as the maximum for most common types of insulated wire, since it represents operating temperatures in the neighborhood of 150°F. The top (140°C rise) line represents final wire temperatures of about 320°F, and should be used only for bare wire suspended by high-temperature insulating supports.

The basic chart is for two or three insulated wires in a bundle, and the allowable currents should be reduced by the factors given in the table accompanying Fig. 1-5 for larger bundles.



Correction factors for current-carrying capacity

No. of conductors in bundles	Multiplying factor
1	1.6
2-3	1.0
4-5	0.8
6-15	0.7

FIGURE 1-5 Maximum current versus wire size of copper wire for various maximum temperature-rise limits. (Courtesy Potter and Brumfield Division, AMF Inc., Princeton, Ind.)

1.3 INSULATION AND VOLTAGE RATING OF WIRE

Any insulating material can be made to conduct if enough voltage is applied across it to rip the normally well-bound electrons loose from their atoms. If a fairly large current is allowed to flow through such an insulation breakdown, enough heat may be generated to permanently destroy the insulator. Figure 1-6 shows that not all materials are alike in their ability to resist voltage breakdown, and that just about any solid insulator is better than air. The values given do not include a safety factor, so the wise designer will assume the worst-case insulating strength to be one-fifth to one-tenth of the value given.

Material	Breakdown voltage kV/mm	Breakdown voltage V/mil (0.001 in.)
Air, vacuum	3.0	75
Porcelain	3.9	100
Bakelite	12	300
Oil	12	300
Vinyl plastic	16	400
Rubber	16	400
Phenolic	18	450
Polystyrene	20	500
Plexiglas	35	900
Paper, waxed	47	1200
Teflon	59	1500
Mica	157	4000

***FIGURE 1-6** Breakdown voltages for various common insulators.*

In choosing an insulating material, it is well to consider the possibility of environmental deterioration as well as the original insulating ability. Rubber can become cracked and brittle with age, oil can become contaminated, polyvinyl chloride materials can embrittle under exposure to sunlight, and air can be laden with salt spray and humidity.

Arcing from a Sharp Point: Electric charge accumulates at a sharp point, but spreads evenly over a smooth surface. The results of this are depicted in Fig. 1-7. Four different conductors were held 2 mm from a metal plate, and the voltage required to arc the air between them was measured. When the conductor was a smooth sphere, a voltage of 8 kV was required. When a 90° bend in a No. 24 wire was used, 5 kV caused an arc. But when a sharply pointed No. 24 wire was used, arcing started at only 2.5 kV. A piece of insulating sleeving on the wire stopped all

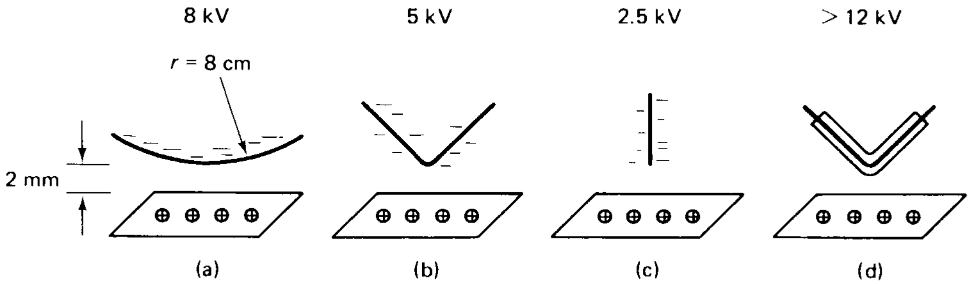


FIGURE 1-7 *Test results: arcing is less likely to occur from a smooth or insulated surface than from a sharp point.*

arcing. Therefore, when constructing high-voltage systems, the technician should:

1. Cut or file off all sharp metal and wire points.
2. Leave smooth, rounded solder joints with no sharp protruding solder "tails."
3. Use insulating sleeving on any bare high-voltage conductors.
4. Cover, coat, or paint surfaces where arcing is likely to be a problem. Special "corona dope" compounds are available for this purpose.

1.4 SKIN EFFECT AND STRAY INDUCTANCE OF WIRE

At high frequencies, the magnetic fields within a wire become such as to force all charge carriers to flow at the surface of the wire, leaving the center of the wire useless as a conductor. This phenomenon, known as the *skin effect*, means that the resistance of a wire conductor increases at high frequencies.

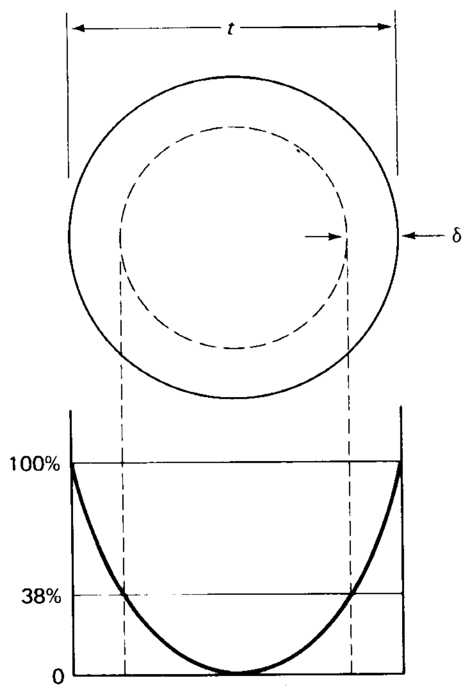
At dc and low ac frequencies the conductivity of a wire is proportional to its circular-mil area (diameter squared), and heavier wire really does have much better conductivity. However, at radio frequencies conductivity is proportional to surface area (which is directly proportional to diameter), and the advantage of heavier wire is less pronounced. Larger-sized wire begins to suffer from the skin effect at a lower frequency because its central core is farther removed from the skin.

Calculating Skin Effect: Although the current density under the skin effect actually drops off exponentially from the surface toward the center of the wire, the effect is the same as if a uniform current density were confined to a conductive surface with a thickness δ, called the *skin depth*. This is illustrated in Fig. 1-8(a).

Actually, the current density is 38% of its full value at the skin depth. For round copper conductors, skin depth can be calculated by

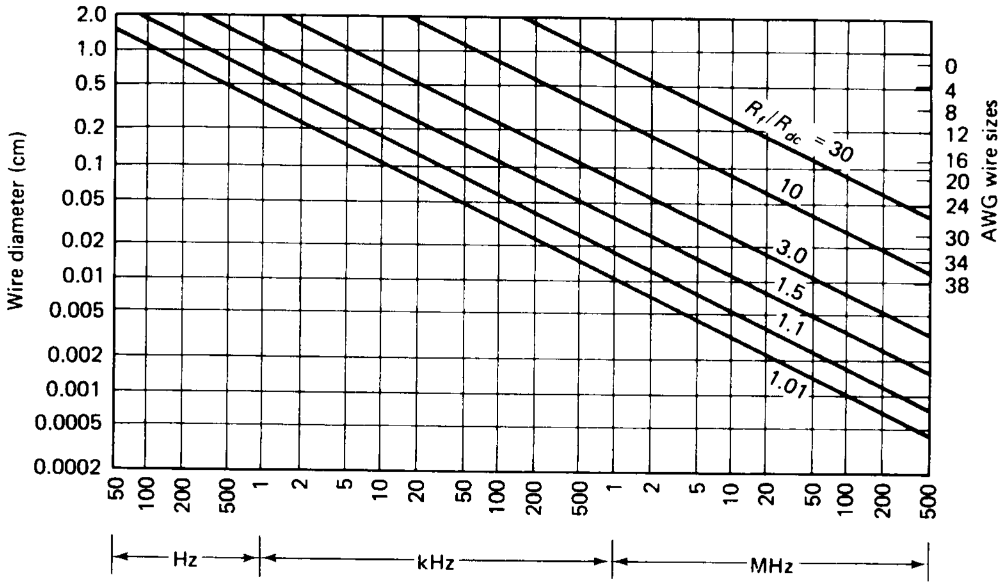
$$\delta = \frac{6.6}{\sqrt{f}} \tag{1-3}$$

where δ is in centimeters and f is in hertz. For other metals, the depth can be determined by multiplying by the square root of the relative resistivity from the table of Fig. 1-3. Aluminum, for example, has a skin depth of $\sqrt{1.64} = 1.28$ times that of copper. Skin depth is important because it determines the heaviest solid wire that can be effectively employed at a given frequency. If the wire has a diameter greater than 2δ, the center of the wire is relatively useless.



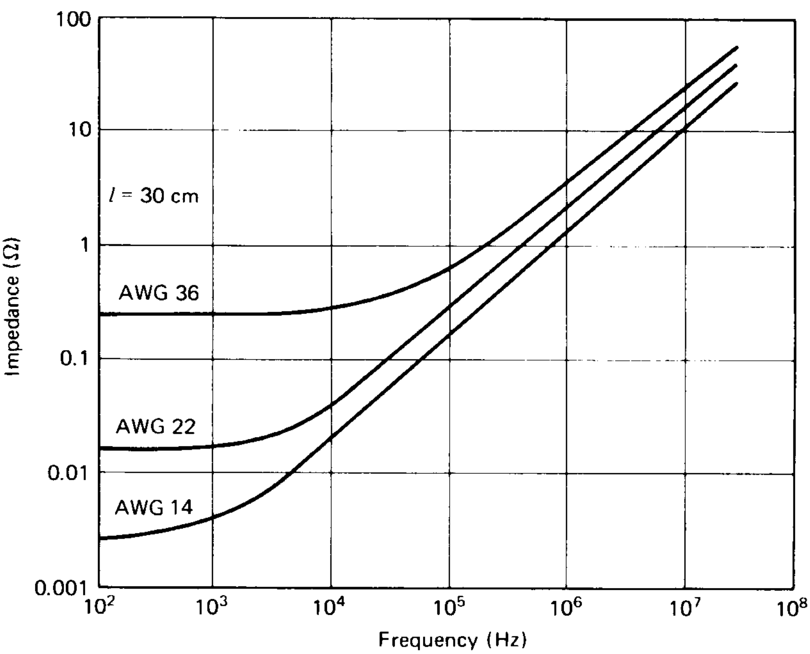
(a)

FIGURE 1-8 Wire impedance increases at high frequencies: (a) skin depth δ and current distribution in a round wire at high frequency.



(b)

***FIGURE 1-8** (b) Skin effect for straight conductors versus wire size and frequency.*



(c)

***FIGURE 1-8** (c) Measured values of impedance versus frequency for 30-cm lengths of various wire sizes.*

EXAMPLE 1-4

What is the maximum effective diameter of a copper wire for 60-Hz power transmission?

Solution

$$\delta = \frac{6.6}{\sqrt{f}} = \frac{6.6}{\sqrt{60}} = 0.85 \text{ cm}$$

$$t = 2\delta = 1.7 \text{ cm} \tag{1-3}$$

The actual resistance of a copper wire under skin effect can be determined from

$$R = 8.3 \times 10^{-8} \frac{\sqrt{f}}{t} l \Big|_{t \gg 2\delta} \tag{1-4}$$

where R is in ohms, f is in hertz, and t and l are thickness and length, respectively, in any commensurate units (both inches, both centimeters, or whatever). For other conductors, multiply by the square root of the relative-resistance value in Fig. 1-3. Silver, for example, has a skin-effect resistance of $\sqrt{0.945}$ or 0.972 times that of copper. Figure 1-8(b) shows the ratio of skin-effect resistance to dc resistance as a function of wire size and frequency, and may be used in place of Equation 1-4 for obtaining rough estimates.

EXAMPLE 1-5

What is the voltage loss across a 200-ft length of RG-58 coax feeding a 50-Ω load at 30 MHz?

Solution

Assuming the loss to be primarily in the 0.033-in. diameter center conductor, we find its skin-effect resistance.

$$R_w = 8.3 \times 10^{-8} \frac{\sqrt{f}}{t} l = 8.3 \times 10^{-8} \frac{\sqrt{30 \times 10^6}}{0.033} \times 200 \times 12 = \mathbf{33\ \Omega}$$

$$\frac{V_o}{V_{in}} = \frac{R_L}{R_w + R_L} = \frac{50}{33 + 50} = \mathbf{0.60}$$

This situation was set up in the laboratory where a V_o/V_{in} ratio of 0.58 was measured at 30 MHz.

Proximity Effect: The skin-effect formulas are accurate only for round conductors running in a straight line. When the wire is wound into coils, additional magnetic fields cause further increases in wire resistance. This *proximity effect* is difficult to predict, but it is by no means negligible. The following factors will serve as a rough guide to the proximity effect to be expected.

Coil Configuration	$\dfrac{R_{\text{proximity}}(\text{actual})}{R_{\text{skin}}(\text{calculated})}$
Single-layer, spaced 5 to 20 diameters between windings	1.5 to 3
Single-layer, close-wound	5 to 15
Multi-layer winding	15 to 100

A summary of the most common situations may be helpful. At 60 Hz, skin effect does not become evident even in transformer coils except in wire sizes larger than about AWG 4. In audio transformers up to 10 kHz, skin effect is negligible for wire sizes smaller than about AWG 24. Such large wire sizes are quite uncommon, making dc resistance satisfactory for most power and audio transformer calculations. For most common wire sizes skin effect or proximity effect is evident by 100 kHz, and there are few practical situations above 1 MHz where skin-effect resistance is not many times dc resistance.

Combatting Skin Effect: Stranded wire suffers from skin effect to about the same degree as does equivalent-sized solid wire. (The advantage of stranding is the ability of the wire to flex without breaking.) However, if each strand is separately insulated from the others, with the strands tied together at each end, some relief from skin effect is obtained. Such wire is called *litz wire* and is effective for winding high-Q coils in the 0.1- to 2-MHz range. Below 0.1 MHz, stray capacitance between the strands washes out the benefits of litz wire.

Where energy loss is critical and strength is important, copper-clad steel wire is often employed. The copper coating, one or two skin depths thick, provides conductivity while the steel provides strength. For large-diameter conductors, hollow tubing is lighter, cheaper, and just as conductive, provided that its thickness is two skin depths or more. Flat copper strap also provides greater surface area with less volume of metal. At high radio frequencies a very thin silver plating on a metal or plastic support provides as much conductivity as possible, if the plating is several times the skin depth.

Stray Inductance of Wire: In practice, skin effect in hookup wire is usually masked by the stray inductance of the wire, which may be anticipated to be about 0.04 μH/cm or 1 μH/25 cm. Skin-effect resistance in hookup wire is typically an order of magnitude (factor of 10) less than the reactance caused by stray inductance. Figure 1-8(c) shows measured values of total impedance for three sizes of solid copper wire from low audio to high radio frequencies. Notice that a simple 1-meter test lead of No. 22 wire has an impedance in the vicinity of 60 Ω at 10 MHz.

1.5 STRAY WIRING CAPACITANCE

Any two conductors separated by an insulator form a capacitor, and adjacent lengths of hookup wire fill this requirement in every respect. The following list will be helpful in estimating the stray capacitance between various types of conductors.

Conductors	*Capacitance*
Two No. 22 solid PVC-insulated wires adjacent in a bundle	40 pF/m
Two No. 22 wires twisted 1 turn/cm	46 pF/m
One No. 22 wire held flat against a metal chassis	65 pF/m
Two No. 22 wires held 2 cm apart	12 pF/m
Two conductors of No. 18 lamp cord	50 pF/m
Shielded wire; No. 22 stranded center conductor to braid	300 pF/m
RG-58/U coax cable (7 mm diameter)	88 pF/m
300 Ω TV twin lead	20 pF/m
Tubular capacitor, 2 cm diameter, 5 cm long; outside foil to metal chassis	10–20 pF

High-Frequency Loss on a Microphone Line: Cable capacitances may prove far from negligible in many practical cases. For example, at 5 kHz, 20 m of heavy shielded wire may present a reactance to ground of 5 kΩ. The popular low-cost crystal microphones typically have an internal impedance of 50 kΩ. Figure 1-9(a) illustrates how the loading effect of the cable capacitance may cause the higher audio frequencies to be attenuated (by a factor of about 10 in this case) before they reach the amplifier input. Of course, the problem can be readily solved in a number

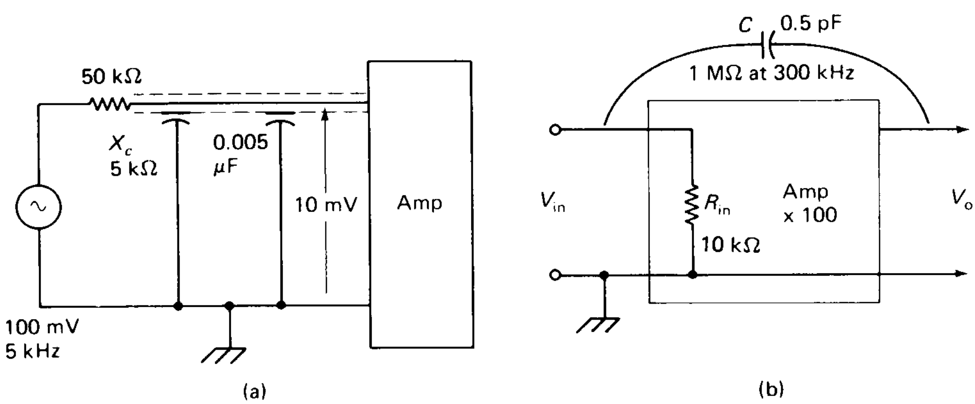
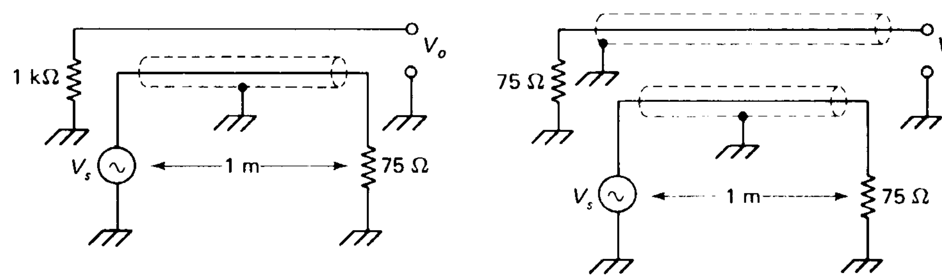


***FIGURE 1-9** Common problems from stray capacitance; (a) shielded microphone cable attenuates audio highs; (b) small stray feedback capacitance causes self-oscillation in an amplifier.*

of ways—use a low-impedance microphone; use a step-down transformer at the "mic" and a step-up transformer at the amp; put an emitter follower or FET amplifier stage in the mic case; use special hi-Z shielded cable; and so on. But the first step in solving any problem must be to identify its cause.

Self-oscillation in an Amplifier: Figure 1-9(b) illustrates another common problem caused by stray capacitance. The input and output leads of the high-frequency amplifier depicted pass within a few centimeters of each other for a short distance, resulting in a feedback capacitance of 0.5 pF. At about 300 kHz this capacitance has a reactance of 1 MΩ and the voltage division of X_C with R_{in} provides a feedback voltage equal to the input voltage. If the phase shift within the amplifier is right, the result will be self-oscillation. To avoid this problem, the input and output wires of any amplifier should be kept completely away from each other and the components on the circuit board should be laid out to implement this. Vertical shield plates can be mounted across a transistor and soldered directly to the chassis to screen the input from the output. Keeping the amplifier input impedance low reduces the feedback ratio and thus helps prevent oscillation.

Coupling between Shielded Cables: Shielded wire and coaxial cable can greatly reduce unwanted coupling between wire runs, but you should guard against the feeling that nothing can get into or out of a shielded cable. Figure 1-10 shows the results of some laboratory tests using RG-59/U coax cables. In Fig. 1-10(a) an unshielded length of hookup wire run along the outside of the coax cable picks up as much as 25% of the cable's signal at certain frequencies. In Fig. 1-10(b), two lengths of coax couple as much as 3% of the signal from one to the other at high frequencies when held close together.



f (MHz)	V_o/V_s
1	0.0001
10	0.001
30	0.25

(a)

f	V_o/V_s
1 MHz	0.00005
10 MHz	0.001
20 MHz	0.01
50 MHz	0.03

(b)

***FIGURE 1-10** Lab test: (a) up to 25% of the voltage on a coax cable is coupled to an adjacent high-impedance wire; (b) up to 3% is coupled between two coax lines 1 m long.*

2

PROPERTIES OF REAL RESISTORS

2.1 FIXED-RESISTOR TYPES

Resistors are available in a variety of physical forms, each with its advantages and drawbacks. However, the vast majority of the market is covered by three basic types, shown in Fig. 2-1.

Carbon Composition: These are the bread-and-butter resistors used in the majority of noncritical applications. They are universally tubular in shape, with their value marked by three color bands according to the standard color code. They are formed of a mixture of carbon powder with nonconductive fillers and resin binders encapsulated in an insulating tube. The ratio of fillers to carbon is the prime determiner of resistance. Carbon-composition resistors are normally available in tolerances of 5% and 10% and in power ratings of $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1, and 2 W. Figure 2-2 shows the actual sizes of the various wattages of carbon-composition resistors.

Composition resistors can be expected to change value by several percent with ambient temperature changes and internally caused heating. Figure 2-3 documents this effect for a wide range of resistance values. The principal advantage of carbon resistors is cost, which may be as low as 2 or 3¢ each for $\frac{1}{2}$-W 10% types purchased in quantities of 1000.

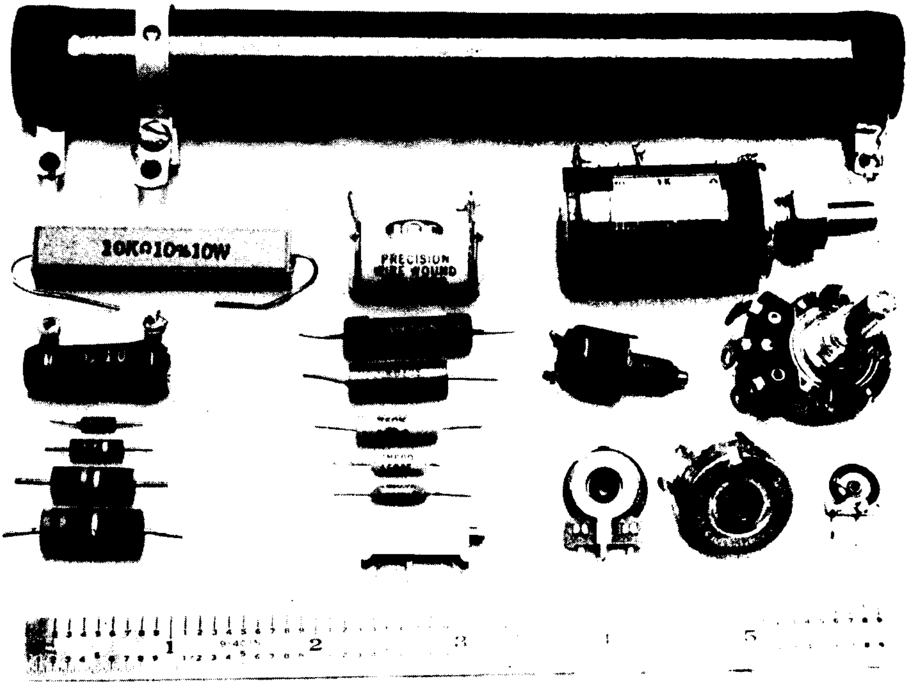
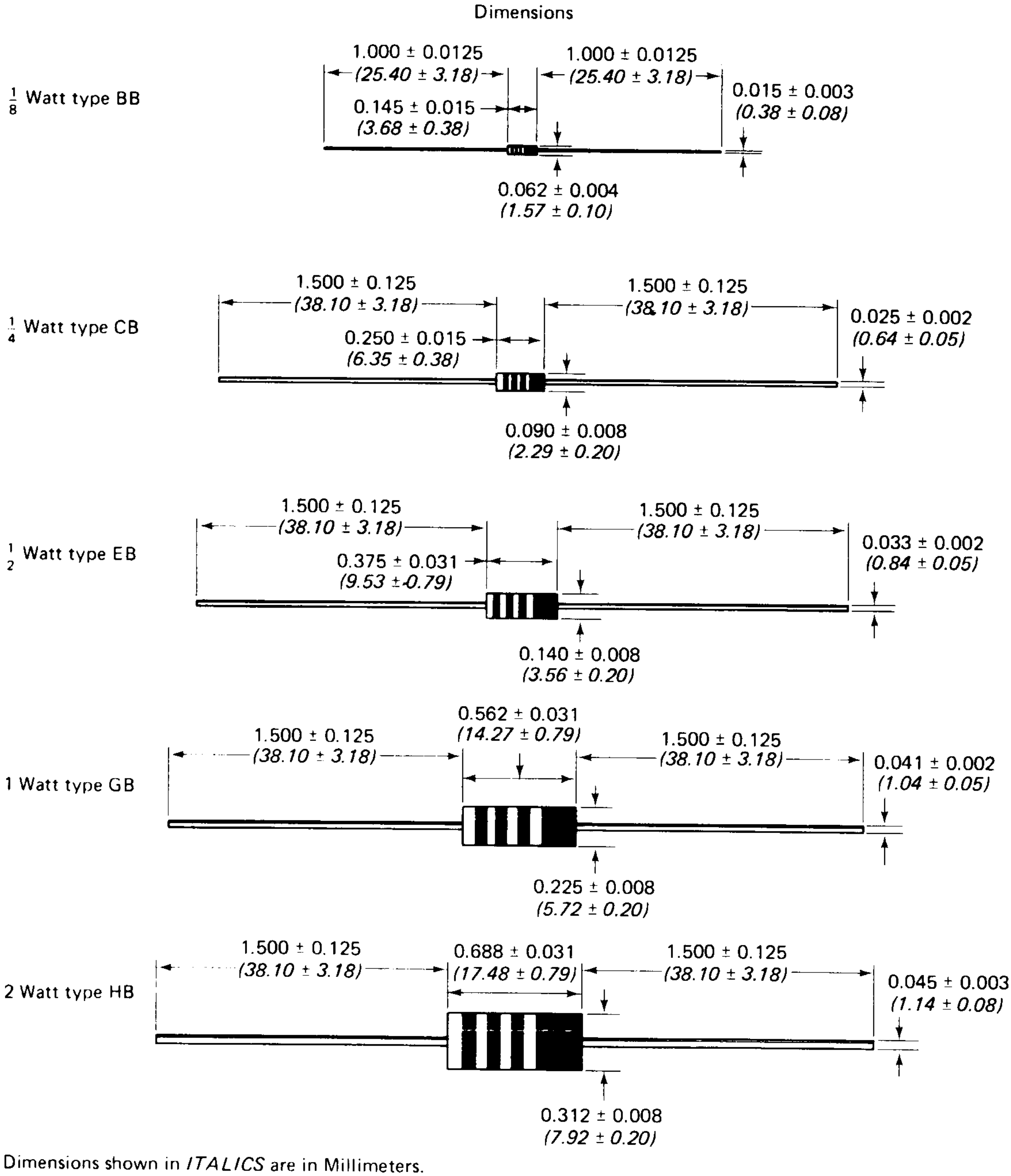


FIGURE 2-1 Representative fixed and variable resistors. Top: *100-W adjustable wirewound;* left: *10-W and 5-W wirewound, followed by $\frac{1}{4}$-, $\frac{1}{2}$-, 1-, and 2-W carbon-composition types;* center: *precision wirewound with five precision metal -film types from 1 W to $\frac{1}{8}$ W below;* top right: *precision 10-turn potentiometer with miniature and standard single-turn pots below;* bottom center to right: *10-turn trim pot, two PC-mount single-turn trimmers, and an inexpensive carbon-track adjustable resistor.*

Metal Film: These resistors are available from stock with a tolerance of ±1%, and can be special ordered with tolerances as close as 0.05%. They are commonly used where accurately known and tightly held values of resistance are required. They can be purchased with guaranteed resistance change with temperature in the range 25 to 100 ppm/°C (parts per million per degree Celsius), as compared to typical coefficients of 200 to 500 ppm/°C for carbon-composition types. Their internal noise generation is also quite low, an important factor in handling low-level audio and radio signals. Metal film resistors are commonly available in $\frac{1}{8}$-, $\frac{1}{4}$-, $\frac{1}{2}$-, and 1-W sizes, and are stocked in values listed in Fig. 2-4. Prices are typically 5 to 10 times the cost of carbon-composition units. Carbon-film resistors with many of the noise and temperature-stability advantages of metal films are becoming available in 5% tolerances at a cost competitive with composition types.



(a)

***FIGURE 2-2** (a) Actual sizes of 2-, 1-, ½-, ¼-, and ⅛-W carbon-composition resistors. (Courtesy of Allen-Bradley Co.)*

INDUSTRIAL GRADE → EB 5145

TYPE OR STYLE			
Allen-Bradley	Rating (Watts) (at 70°C)	MIL-R-11 Style	MIL-R-39008 Style
BB	1/8	RC05	RCR05
CB	1/4	RC07	RCR07
EB	1/2	RC20	RCR20
GB	1	RC32	RCR32
HB	2	RC42	RCR42
GM	3	–	–
HM	4	–	–

TYPE

Fixed Composition Resistors, Insulated, Established Reliability

RESISTANCE

Expressed in ohms and identified by a three-digit number. First two digits represent significant figures. Last digit specifies the number of zeros to follow, except below 10 ohms (see below)

FOR VALUES BELOW 10 OHMS	
ALLEN-BRADLEY DESIGNATION	MIL DESIGNATION
The letter "G" is substituted in place of the third digit. The desired resistance value is then the first two digits multiplied by 0.1	The letter "R" is substituted in place of a significant digit to represent the decimal point. The following digits represent significant figures.

EXAMPLE		
Allen-Bradley	Resistance	MIL Designation
27G	2.7 Ohm	2R7

RESISTANCE TOLERANCE		
Allen Bradley	Tolerance	MIL Designation
5	±5%	J
1	±10%	K
2	±20%	–

FAILURE RATE LEVEL

At 50% rated wattage expressed in %/1000 hrs.

M = 1.0%
P = 0.1%
R = 0.01%
S = 0.001%

MIL-R-39008 TYPE DESIGNATION → RCR 20 G 514 JS

MIL-R-11 TYPE DESIGNATION → RC20 GF514 J

MAXIMUM AMBIENT TEMPERATURE AND RESISTANCE TEMPERATURE CHARACTERISTIC
(Refer to MIL-R-11 or MIL-R-39008 Specifications as Applicable)

Standard Color Code and Preferred Number Series

First Band – 1st Digit

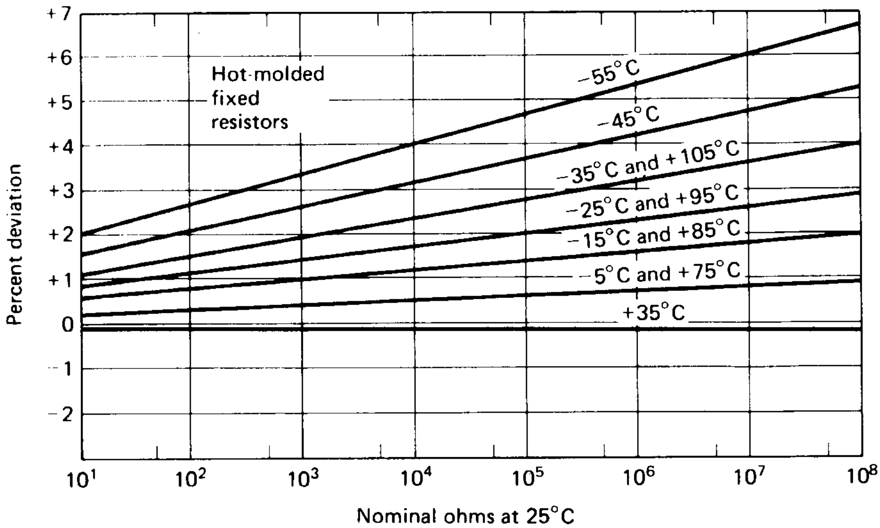
Second Band – 2nd Digit

Color	Digit	Multiplier	Tolerance	Reliability Level ■ (Percent Per 1000 Hours)
Black	0	1	–	–
Brown	1	10	–	M = 1.0%
Red	2	100	–	P = 0.1%
Orange	3	1000	–	R = 0.01%
Yellow	4	10,000	–	S = 0.001%
Green	5	100,000	–	–
Blue	6	1,000,000	–	–
Violet	7	10,000,000	–	–
Gray	8	–	–	–
White	9	–	–	–
Gold	–	0.1	±5%	–
Silver	–	–	±10%	–
No color	–	–	±20%	–

■ When Applicable

Preferred Number Series						
±5% Tolerance	10 11 12 13	15 16 18 20	22 24 27 30	33 36 39 43	47 51 56 62	68 75 82 91
±10% Tolerance	10 12	15 18	22 27	33 39	47 56	68 82
±20% Tolerance	10	15	22	33	47	68

FIGURE 2-2 (b) Standard values and marking system for carbon-composition resistors. (Courtesy Allen-Bradley Co.)



Percent resistance deviation from 25°C value for various nominal resistance values and temperatures

***FIGURE 2-3** Resistance change as a function of temperature for representative carbon-composition resistors. (Courtesy Allen-Bradley Co.)*

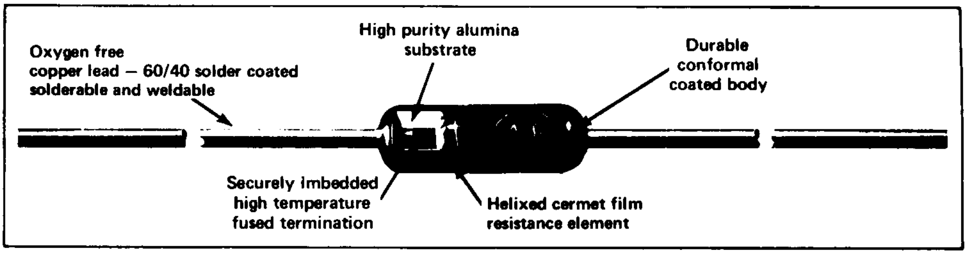
Wirewound: Wirewound resistors are commonly used in high-power applications (above 2 W). Some readily available sizes are as follows:

Power (W)	*Length* (cm)	*Diameter* (cm)
5	1.6	0.6
12	4.5	0.8
25	5.1	1.4
50	10.2	1.4
100	16.5	1.9
225	27	2.9

For color-coded wirewound resistors, the first band is double width, to distinguish wirewound from composition types. Typical tolerances for wirewound power resistors are 5 to 10%, with 1% available in the lower-power ratings. Costs range from approximately 30¢ for 5-W types to about $2 for 225-W types in 10% tolerance. Temperature stability for power resistors is typically on the order of 200 ppm/°C.

Precision-value wirewound resistors are manufactured by the simple expedient of trimming the length of a resistance wire wound on a bobbin. These resistors are not generally meant to dissipate high power, however.

Resistor Networks: In applications where a number of resistors of the same value are required, size and assembly time can be minimized by using resistors packaged in the dual-in-line package (DIP) developed for integrated circuits. These resistors



RN55D

Style and Size — RN55; Characteristic — D. For Allen-Bradley D = ±100 ppm/° C

1962 F

Resistance Value — 1962. First three digits are significant, 4th. digit "number of zeros".
1962 = 19,600 ohms
For values below 100 ohms, the letter R represents the decimal point, with all digits significant.

Tolerance — F. F = 1%

Example: 10R5 = 10.5 ohms
1022 = 10,200 ohms

STANDARD DECADE — 1%											
10.0	12.1	14.7	17.8	21.5	26.1	31.6	38.3	46.4	56.2	68.1	82.5
10.2	12.4	15.0	18.2	22.1	26.7	32.4	39.2	47.5	57.6	69.8	84.5
10.5	12.7	15.4	18.7	22.6	27.4	33.2	40.2	48.7	59.0	71.5	86.6
10.7	13.0	15.8	19.1	23.2	28.0	34.0	41.2	49.9	60.4	73.2	88.7
11.0	13.3	16.2	19.6	23.7	28.7	34.8	42.2	51.1	61.9	75.0	90.9
11.3	13.7	16.5	20.0	24.3	29.4	35.7	43.2	52.3	63.4	76.8	93.1
11.5	14.0	16.9	20.5	24.9	30.1	36.5	44.2	53.6	64.9	78.7	95.3
11.8	14.3	17.4	21.0	25.5	30.9	37.4	45.3	54.9	66.5	80.6	97.6

To obtain standard resistance values, multiply preferred number from decade table by powers of 10.

***FIGURE 2-4** Physical construction, standard values, and marking system for 1% film resistors. (Courtesy Allen-Bradley Co.)*

are manufactured either by thin-film technology, which involves depositing a strip of metal a few millionths of an inch thick on a glass or ceramic base, or by thick-film techniques, in which resistive material is painted on an insulating surface, usually by silk-screen printing.

2.2 VARIABLE-RESISTOR TYPES

Variable resistors come in the following basic types:

- Carbon-composition potentiometer
- Wirewound potentiometer
- Multiturn potentiometer
- Linear slide pots
- Single-turn trimmers
- Gear-actuated trimmers
- Adjustable resistor

These are illustrated in Fig. 2-1. The term *potentiometer* derives from the early use of the component in a potential-measuring circuit. Verbally, the term is almost always shortened to *pot*. Trimmer pots are usually small, and almost always designed to be adjusted only occasionally by a screwdriver rather than frequently by a front-panel knob. High-power pots, especially those having only two terminals, are sometimes called *rheostats*.

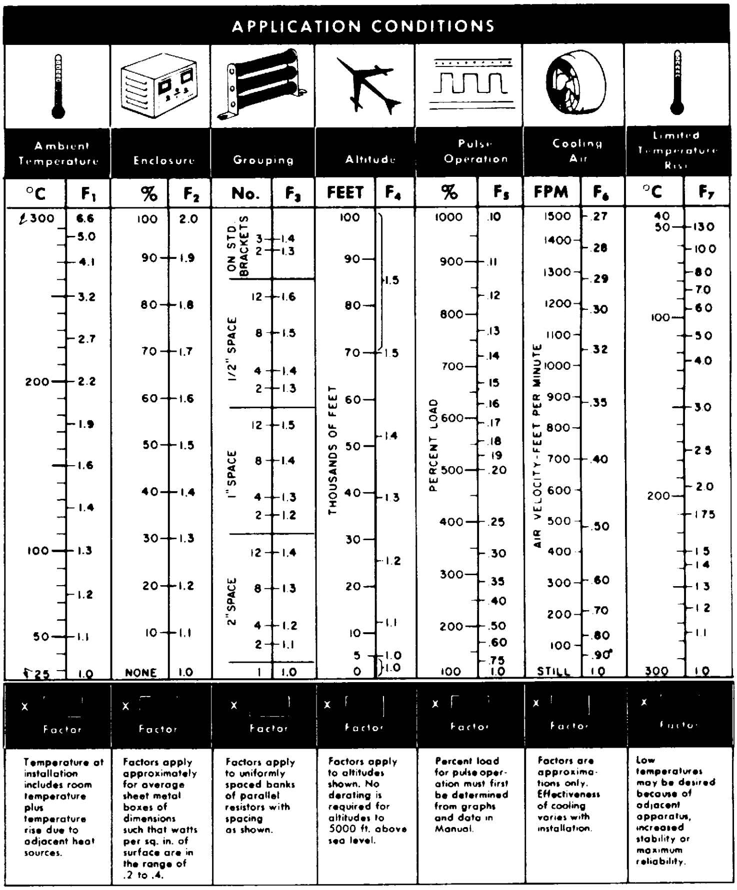
The wiper of a wirewound pot divides the resistance neatly into two parts, so a 100-Ω pot at midrotation will measure 50 Ω from either end to the wiper. The wiper of a carbon-track pot provides only a point (or two points) contact on the track, and there is extra resistance associated with this incomplete contact to the cross section of resistance. A 100-Ω carbon pot at midrotation will typically measure 55 Ω from each end to the wiper.

2.3 RESISTOR POWER RATINGS

Manufacturers' power ratings for resistors apply for "standard" conditions which are seldom realized in practice. Therefore, it is wise to adjust the resistor power rating as calculated from $P = IV$ to accommodate actual conditions of operation. Standard conditions for wirewound power resistors are:

- Ambient temperature of 25°C (77°F)
- Resistor suspended in free air (no enclosure)
- Single resistor suspended away from other heat-producing devices
- Air pressure equivalent to sea level (76 cm Hg)
- 100% duty cycle
- Still air; no forced-air cooling
- Allowable temperature rise of 300°C

Figure 2-5 shows the recommendations of the Ohmite Manufacturing Company for adjusting wirewound power resistor ratings to accommodate deviations from these conditions. This chart is given only as a guide, since the factors involved are seldom accurately known. The percent load for pulse operation, in particular, is a complex function of duty cycle (% on-time), cycle time, and resistor size. For cycle times of less than 0.1 s the percent load is nearly the reciprocal of duty cycle, but for longer cycle times the resistor will have a chance to heat up during the on-time, and the percent load will be much less than 1/duty cycle. In practice, most resistor users settle on a safety factor (usually 2, 3, or 4) for resistor power ratings which they have found from experience to be suitable to the service conditions of their products.



EXAMPLE: Four resistors, each dissipating 115 watts, are to be mounted in a group. Spacing is to be 2" surface to surface. Ambient to be 50 C (122 F). Enclosure to be total. Other factors standard. Determine Watt Size required.

Operation (1) On Ambient Temperature scale locate 50°C. Note and record F_1 = 1.1 as shown. Locate and record the other factors.

F_1	F_2	F_3	F_4	F_5	F_6	F_7
50°	100%	4@2"	Standard Conditions			
1.1 x	2.0 x	1.2 x	1 x	1 x	1 x	1

Operation (2) Multiply the factors together = 2.64

Operation (3) 115 watts x 2.64 = 314 Watts Free Air Watt Size Rating required for each resistor.

OHMITE

***FIGURE 2-5** Resistor power rating calculation chart. (Courtesy Ohmite Manufacturing Co.)*

EXAMPLE 2-1

A totally enclosed instrument is to operate at an ambient temperature of 75°C. The instrument contains four resistors, each dissipating 6 W, mounted with 1-in. space between resistors. A fan provides forced-air cooling at a velocity of 500 ft/min, and the resistors may not exceed a temperature of 250°C because of insulation used within the instrument. The altitude is sea level and the duty cycle is 100%. Find the watt-size resistor required.

Solution

From the chart, the factors involved are:

F_1 for 75°C	1.2
F_2 for total enclosure	2.0
F_3 for 4 at 1-in. spacing	1.3
F_4 and F_5 normal	1.0
F_6 for 500 ft/min	0.5
F_7 for 175°C rise (250° final − 75° ambient)	2.5

The minimum watt size required for each resistor is

$$6 \times 1.2 \times 2.0 \times 1.3 \times 1.0 \times 0.5 \times 2.5 = \mathbf{23.4\ W}$$

Where available resistors fall short of power requirements, a heat-sinking clamp and/or thermally conductive adhesive to the chassis can be used to approximately double the power-handling capability of 1-W and 2-W carbon resistors. For laboratory bench tests, immersing small resistors in a bath of distilled or deionized water will increase their capability by approximately a factor of four.

2.4 RESISTOR TOLERANCE AND RELIABILITY RATINGS

A manufacturer's tolerance specification on a resistor is a guarantee that the resistor will be within the specified percentage of its marked value. For example, a 470-Ω, 10-% resistor will be within the range $470 \pm 47\ \Omega$, or 517 to 423 Ω. For some manufacturers, this means that all resistors shipped have been tested and those whose values lie outside the tolerance range have been rejected. For others, it simply means that the manufacturer will replace any lot of resistors that is found to contain an excessive number of out-of-tolerance pieces (assuming that all 50,000 of them have not been soldered into printed-circuit boards by the time the problem is noticed). Of course, 100% testing makes the resistors cost more, but troubleshooting a finished system to find a defective resistor also costs money.

Reliability Levels: Reliability ratings for resistors have been established by a military specification (MIL-R-39008), which specifies a test procedure for determining the failure rate of the resistors. The procedure involves subjecting a large number of resistors to a 10,000-h life test at 50% rated power at 70°C. The results of this test are indicated by a fifth color band on the resistor. The

MIL-SPEC designations for resistors, including type, tolerance, reliability, and other ratings, are given below:

MIL-STD Resistor Designations

EXAMPLE :

$\underset{\mathbf{a}}{\underline{RC}}\ \underset{\mathbf{b}}{\underline{07}}\ \underset{\mathbf{c}}{\underline{C}}\ \underset{\mathbf{d}}{\underline{8R2}}\ \underset{\mathbf{e}}{\underline{J}}\ \underset{\mathbf{f}}{\underline{R}}$

a (type) and b (power):

RA: Wirewound variable, precision:

Code:	20	25	30
Wattage:	2	3	4

RB: Wirewound fixed precision:

Code:	08	16	17	18	19	52	53	55	56	57	58	70	71
Wattage:	$\frac{1}{2}$	$\frac{2}{3}$	1	$1\frac{1}{2}$	2	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	1	2	$\frac{1}{2}$	$\frac{1}{4}$

RC: Composition fixed:

Code:	05	07	20	32	42
Wattage:	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2

RD: Noninductive film power:

Code:	31	33	35	37	39	60	65	70
Wattage:	7	13	25	55	115	1	2	4

RE: Wirewound power with heat sink:

Code:	60	65	70	75	77	80
Wattage:	$7\frac{1}{2}$	20	25	50	100	250

RL: Film:

Code:	07	20	32	42
Wattage:	$\frac{1}{4}$	$\frac{1}{2}$	1	2

RN: Film, high stability:

Code:	05	50	55	60	65	70	75
Wattage:	$\frac{1}{8}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1

RP: Wirewound variable, power:

Code:	10	11	15	16	20	25	30	35	40	45	50	55
Wattage:	25	12	50	25	75	100	150	225	300	500	750	1000

RV: Composition variable:

Code:	01	04	05	06
Wattage:	$\frac{1}{4}$	2	$\frac{1}{2}$	$\frac{1}{3}$

RW: Wirewound, power:

Code:	55	56	67	68	69	70	74	78	79	80	81
Wattage:	5	10	5	10	$2\frac{1}{2}$	1	5	10	3	$2\frac{1}{4}$	1

c (temperature coefficient):

(this position used alternately to specify temperature limit or lead structure)

$$J, E = \pm 25; H, C = \pm 50; K, O = \pm 100 \text{ ppm}/°\text{C}$$

d (value):

three or four digits; first two or three significant, last gives number of zeros; R = decimal point with last digit significant.

e (tolerance):

$$F(\pm 1\%), G(\pm 2\%), J(\pm 5\%), K(\pm 10\%)$$

f (failure rate or reliability):

M brown (1%), *P* red (0.1%), *R* orange (0.01%), *S* yellow (0.001%), all per 1000 h

The third letter *R* inserted after the two letters in part **a** of the code indicates that the resistor has a reliability level established by a military (MIL-STD) specification.

Considerable effort is expended to document the reliability of components according to exacting procedures laid down in military specifications. Military Reliability Prediction Handbooks give detailed procedures for estimating mean time between failures for systems containing large numbers of specified components. Unless these rather involved procedures are actually applied, reliability ratings should be used for comparison purposes only. It should be kept in mind that a reliability rating *P* means that a component type has been tested under specified conditions with fewer than 0.1% failures per thousand hours of operation. It may not be assumed that the failure rate in service will be this high—only that it is very unlikely to be any higher. Operation at lower power dissipation under less stringent environmental conditions than those used in the test procedure also contributes to reliabilities in service which are much better than the specified rating would appear to indicate.

2.5 RESISTORS AT HIGH FREQUENCIES

Composition and Film Types: The wire leads of a standard $\frac{1}{2}$-W composition or film-type resistor are sufficiently close to provide about 0.5 pF of capacitance in parallel with the resistance value. The circuit leads will introduce at least another 0.5 pF, even if care is taken to keep the leads short and direct. Figure 2-6(a) shows the reactance of 1 pF versus frequency. Notice that, at 1 MHz, a voltage divider composed of a 900-kΩ resistor in series with 100 kΩ would be completely swamped out by the reactance of the stray capacitance.

Where resistors must be used to handle high-frequency ac, it is advisable to limit the resistance value to one-tenth or less of the reactance of 1 pF at the frequency in use. For the circuit of Fig. 2-6(b), values of 9 kΩ and 1 kΩ would hold stray-capacitance effects at 1 MHz to an acceptable value.

A related problem appears when high-value resistors are used to handle fast-rise pulses. Figure 2-6(c) shows the output that would be observed if a 10-ns-rise pulse were applied to the voltage divider of Fig. 2-6(b). To avoid this problem, the highest resistor value should be chosen so that the circuit time constant is one-tenth or less of the pulse rise time.

EXAMPLE 2-2

What value resistors should be used in the circuit of Fig. 2-6(b) if the input pulse has a rise time of 10 ns and the pulse waveform is to be preserved?

Solution

The time constant should be $\frac{1}{10}$ of 10 ns or less.

$$\tau = RC$$
$$R = \frac{\tau}{C} = \frac{1\text{ ns}}{1\text{ pF}} = 1\text{ k}\Omega$$

The resistor values should be 900 and 100 Ω.

Variable Resistors: Potentiometers exhibit the same problems of stray capacitance as do fixed resistors, but to a greater extent. Typical stray-capacitance values for a standard carbon-track potentiometer are as follows:

End to end	2–5 pF
Wiper to end	3–5 pF
Wiper to case	8–15 pF

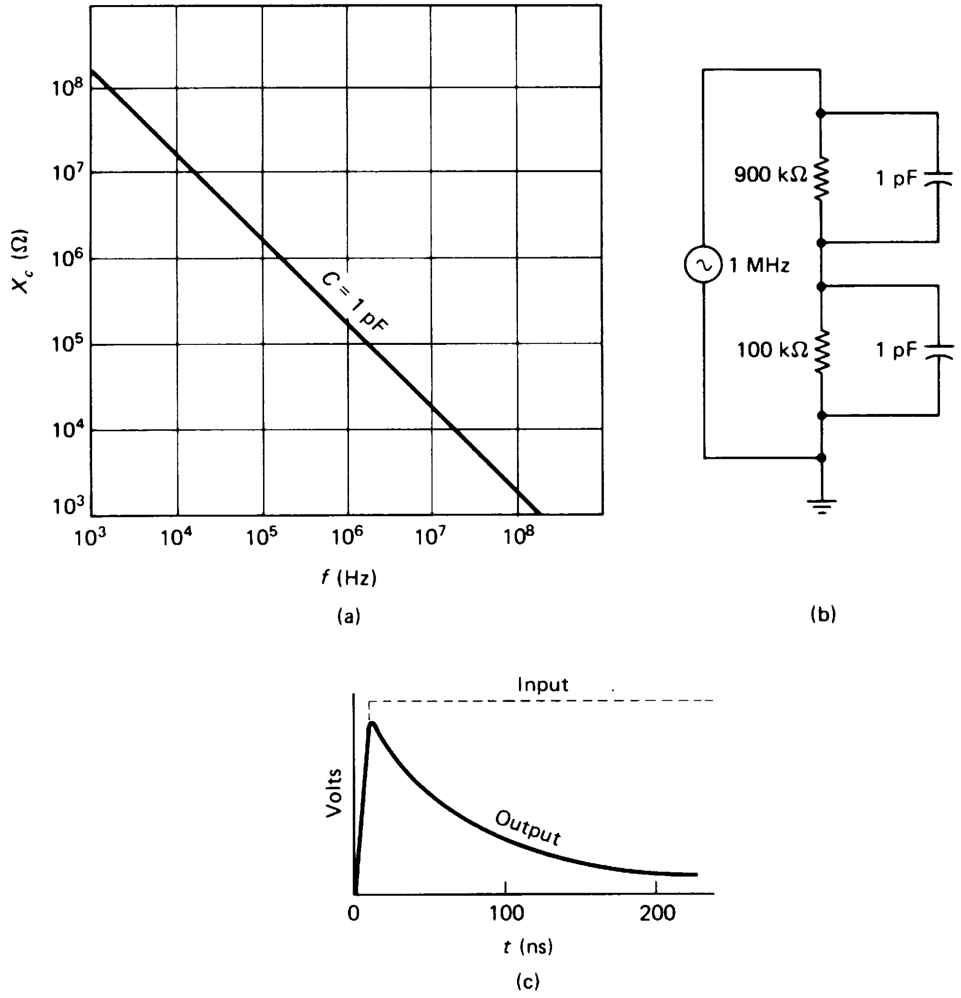


***FIGURE 2-6** (a) The reactance of the typical 1 pF of stray capacitance across a resistor drops to 159 kΩ at 1 MHz. (b) A high-resistance voltage divider is completely swamped out by this reactance. (c) The shape of a square pulse is distorted by the stray capacitive shunts.*

Wirewound Resistors: Wirewound resistors designed exclusively for dc and low-frequency use may have an effective series inductance of 10 to 100 μH, enough to present significant reactance at high frequencies. However, nonreactive resistors using an Ayrton-Perry winding (two windings in opposite directions to cancel inductance) are readily available, and these display only a microhenry or so of residual inductance. Figure 2-7 shows the effects of frequency on composition, film, and wirewound resistors.

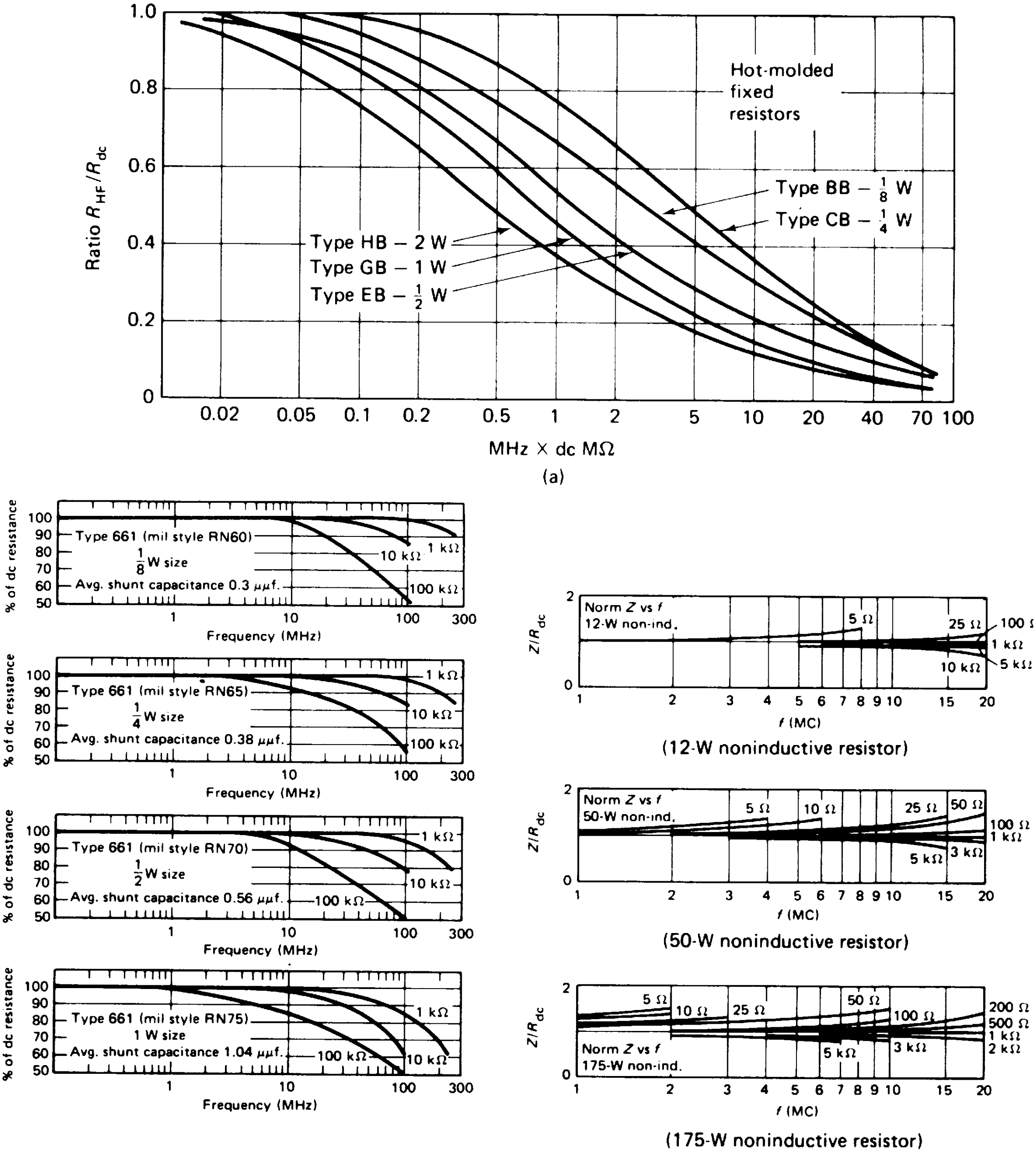


***FIGURE 2-7** (a) Impedance versus frequency for composition resistors. (Courtesy Allen-Bradley Co.) (b) Impedance versus frequency for metal-film resistors. (c) Normalized impedance versus frequency for wirewound resistors. (Courtesy Ohmite Manufacturing Co.)*

Wirewound resistors also have a shunt stray capacitance, typically 5 to 10 pF. For low ohmic values (below 500 Ω) residual series inductance begins to increase the resistor's impedance at frequencies above a few MHz, but for higher values (1000 Ω and above) stray shunt capacitance lowers the impedance above a few megahertz. There is an optimum resistance range in the vicinity of 1 kΩ for which resistance is essentially constant to 10 or 20 MHz [see Fig. 2-7(c)].

3

PROPERTIES OF REAL CAPACITORS

3.1 FIXED-CAPACITOR BASICS

Energy Storage: A capacitor is formed whenever two conductive surfaces or plates are separated by an insulator (called the *dielectric*). If a potential difference exists between the surfaces, there will be an electric field in the dielectric representing a stored energy of

$$W = \tfrac{1}{2}CV^2 \tag{3-1}$$

where W is the energy in joules (1 joule = 1 watt for 1 second), C the capacitance in farads, and V the potential difference in volts.

Charge Storage: The potential difference between the plates exists because of a difference in charge on them, and if a conductive path is provided between the plates, a charge (quantity of electrons) will flow between them to equalize their potentials. The amount of this stored charge is

$$Q = CV \tag{3-2}$$

where Q is the charge in coulombs (1 coulomb = 1 ampere for 1 second), and C and V are in farads and volts as above.

Calculating Capacitance: The amount of stored energy and stored charge per volt is thus directly proportional to capacitance C, and this property is dependent on three factors:

1. C is directly proportional to plate area. Doubling the plate area doubles the capacitance, assuming, of course, that the same dielectric is maintained.
2. C is inversely proportional to plate spacing. Increasing the plate spacing weakens the field strength in the dielectric and thus reduces the amount of charge that can be held on the plates per volt of potential difference. Doubling the plate spacing halves the capacitance.

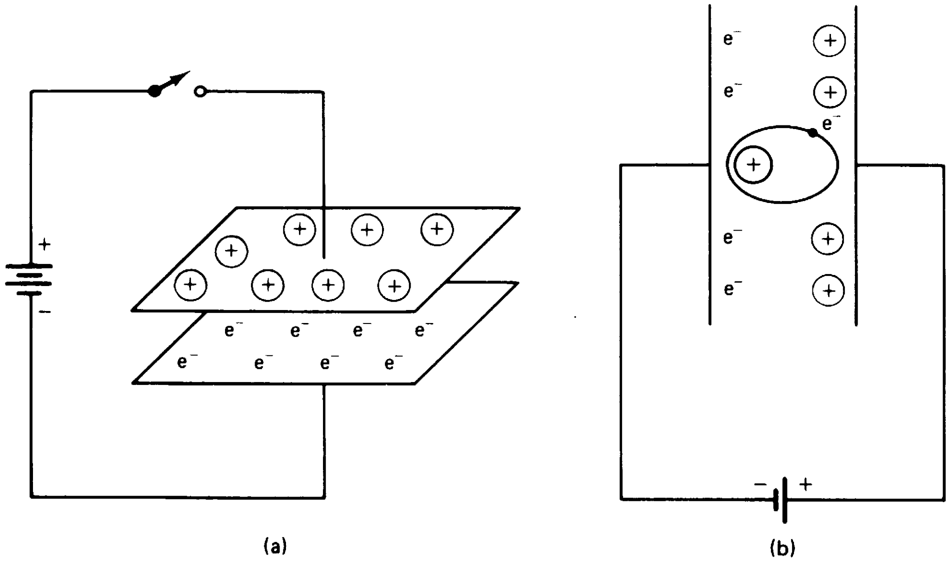


***FIGURE 3-1** (a) Opposite charges are held on the plates of a capacitor by mutual attraction. (b) The atoms in the dielectric become polarized (distorted) when the plates are charged.*

3. C depends upon the dielectric material. As depicted in Fig. 3-1(b), the atoms of a dielectric material are distorted by the electric field, the electrons being pulled closer to the positive plate. This deformation represents additional stored energy in the dielectric, and the polarized atoms attract the charges on the plates, holding more charge per volt as polarization becomes more pronounced. Various dielectric materials are susceptible to polarization in different degrees, as indicated in the table of dielectric constants, Fig. 3-2. C is directly proportional to dielectric constant k.

For a capacitor constructed of two parallel plates, the capacitance is given by

$$C = 8.85 \times 10^{-10} \frac{kA}{d} \tag{3-3}$$

Material	Dielectric Constant	Material	Dielectric Constant
Vacuum	1.0000	Oil	4
Air	1.0006	Mica	5
Teflon	2	Ceramic (lo-k)	6
Paraffin	2.2	Bakelite	6
Paper	2.5	Water	81
Rubber	3	Ceramic (hi-k)	7000

***FIGURE 3-2** Typical values of dielectric constant k. Exact values depend upon specific composition of material.*

where C is in farads, k is the dielectric constant, A is the area between the plates in cm^2, and d is the plate spacing in centimeters.

Capacitor Construction: Capacitors are physically constructed as single plates, stacked plates, or rolled tubes, as illustrated in Fig. 3-3. The single-plate structure is common in the low-cost ceramic-disc capacitors, where values up to 1 μF can be obtained with high-dielectric-constant ceramic. Otherwise, this structure is limited to low capacitance values. Stacked plates are used with rigid dielectric materials such as mica and ceramic, whereas the more economical rolled tubular structure is used with flexible dielectrics such as paper and Mylar.

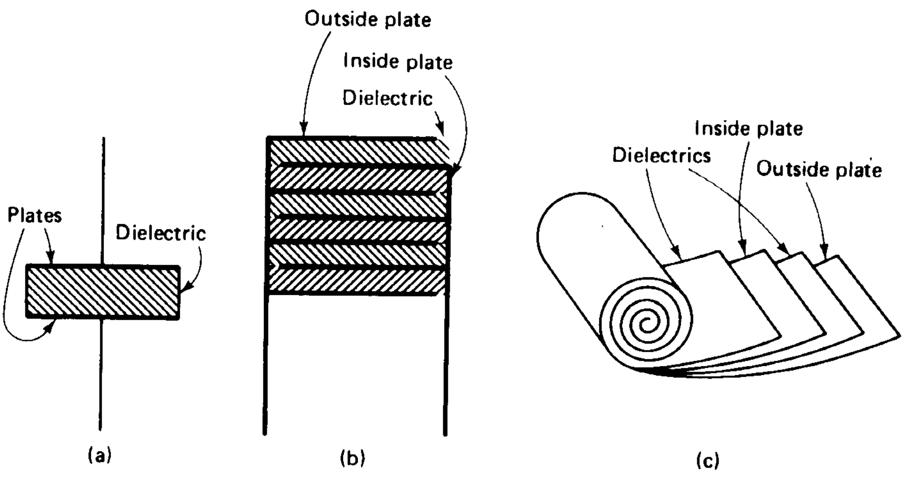


***FIGURE 3-3** Capacitor construction: (a) single plate; (b) stacked plate; (c) rolled tubular.*

Notice that some capacitors are constructed so that the outside "plate" completely surrounds and enshields the inside plate. This outside foil is then marked with a band or stripe on the case, and this lead should be connected to the circuit ground in noise-critical bypass applications. Figure 3-4 shows a selection of fixed-nonelectrolytic-capacitor types.

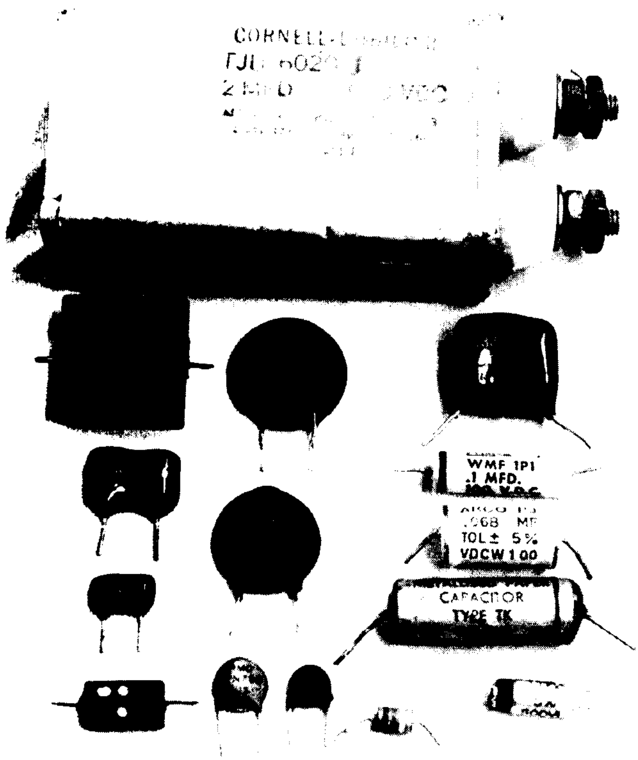


FIGURE 3-4 Representative types of fixed, nonelectrolytic capacitors. Top: *a relatively small oil-filled type;* left: *four mica types ranging from 0.01 μF to 47 pF;* center: *high-k and medium-k ceramic discs, with temperature-compensating and zero-temperature-coefficient ceramic discs below;* right: *dip-coated and tubular Mylars, plastic tubular, metallized paper, tubular ceramic;* bottom right: *polystyrene dielectric capacitor.*

3.2 NONIDEAL EFFECTS IN CAPACITORS

Real capacitors exhibit a number of nonideal effects which must be measured against cost in selecting a capacitor type for a given application. This section presents a brief catalog of the nonideal effects to be expected. The following sections of this chapter detail the types of capacitors available and their real-life characteristics. Figure 3-18 at the end of the chapter provides a tabular summary of this information.

DC Leakage: Ideally, a capacitor should be capable of holding a charge forever. In practice, of course, current leakage through the dielectric and through the case material discharges the capacitor in time. For general-purpose capacitors of 0.1 μF and below, the case resistance generally limits the discharge time constant to a few seconds or less, decreasing for lower values. For high-value capacitors, however, the discharge time is primarily a function of the dielectric-material resistivity, as follows:

Polystyrene	~3 days
Paper	~3 h
Tantalum	~1 h
Ceramic (high k)	~5 min
Electrolytic	~10 s

Dielectric Loss: If an ac voltage is applied across a capacitor with a dielectric other than air or vacuum, the successive polarizations of the dielectric will cause some molecular friction and loss of energy. This loss is not easy to predict, as it varies with frequency and temperature and is different for each material. Some materials exhibit severe loss peaks at certain frequencies. Polyvinyl chloride (PVC), which is widely used as an insulator, is unsuitable as a dielectric because of its loss characteristics, as shown in Fig. 3-5(a). Some typical effects of temperature on dielectric loss are shown in Fig. 3-5(b).

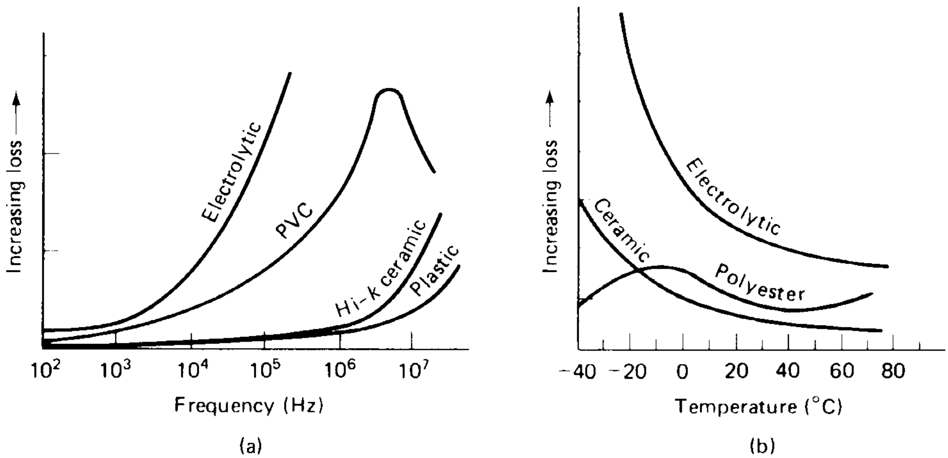


FIGURE 3-5 (a) Dielectric loss generally increases with frequency, but loss peaks occur in some materials. (b) Loss varies with temperature, generally increasing at low temperature for high-k materials.

Dissipation Factor: The cumulative effect of resistive shunt leakage, dielectric loss, and whatever series resistance may be present in the capacitor leads and plates is represented by the dissipation factor of the capacitor. This is defined as

$$D = \frac{R_s}{X_C} \tag{3-4}$$

where R_s is an equivalent series resistance (called ESR by the manufacturers) and X_C is the equivalent series reactance of the capacitor at the frequency in question. For reasonably low values (10% or less), D is approximately equal to the power factor, which is the ratio of real power dissipated (heat loss) to apparent power (volt-amps) for the capacitor. The dissipation factor can be measured on most impedance bridges, and D or ESR is often specified in manufacturers' literature. It should be kept in mind that these values are valid only at the frequency and temperature specified, and will generally increase (sometimes drastically) at higher frequencies or lower temperatures.

Tolerance and Temperature Stability: Most capacitor types are manufactured to a tolerance of ± 10 or $\pm 20\%$ of the marked value at 25°C. However, certain types, such as aluminum electrolytics and high-k ceramics, are intended only to provide large amounts of capacitance in a small package at a low cost, and may have tolerances as wide as -20 and $+100\%$. Most often, the manufacturers do not make a point of advertising these poor tolerances too loudly, so it is up to the technician to be aware of them. On the other hand, mica, polystyrene, low-k ceramic and Mylar capacitors are available (for a price) with tolerances as narrow as 1%.

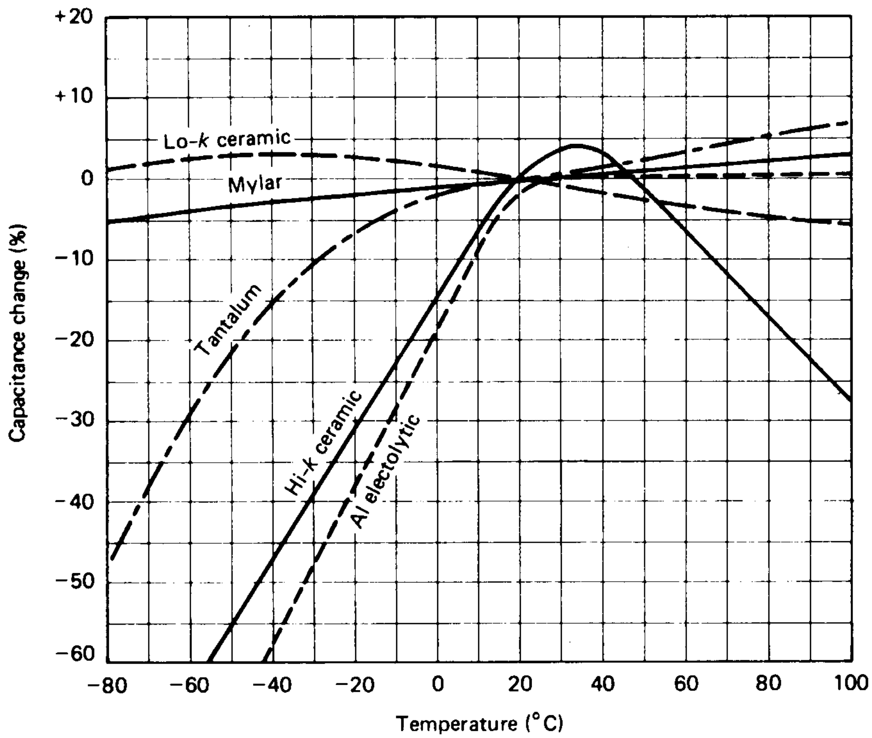


FIGURE 3-6 Capacitance changes with temperature vary markedly for different dielectrics.

General statements about the effects of temperature on capacitance are quite difficult to make, since the effect varies, not only between different dielectrics, but also between different versions of essentially the same dielectric. Figure 3-6 gives a rough idea of the effect to be expected for a few common dielectrics. Note in particular the drastic drop in capacitance at low temperatures for aluminum electrolytic and high-k ceramic capacitors.

Lead Inductance and Resonance: As noted in Chapter 1, hookup wire possesses a stray inductance of approximately 0.04 μH/cm, so a 0.1-μF capacitor with two 2.5-cm leads is really 0.1 μF in series with 0.2 μH, a combination which resonates at slightly above 1 MHz, with X_L and X_C of about 1.4 Ω. Above this resonant frequency, the reactance of the "capacitor" is actually inductive, rising to about 14 Ω at 10 MHz, 140 Ω at 100 MHz, and so on.

The table of Fig. 3-7(a) shows the measured resonant frequencies of various capacitors, and Fig. 3-7(b) shows the measured impedance of two of them as a function of frequency.

The reactance of the 0.01-μF "capacitor" in Fig. 3-7(a) is definitely inductive above 7 MHz, and this fact is demonstrated with surprising effects when a 470-pF capacitor is placed in parallel with it. A parallel resonance occurs at 23 MHz, resulting in a total impedance rise to 36 Ω, four times higher than the impedance of either capacitor alone. This parallel-resonance effect should be kept in mind, as it is common practice to use separate audio and RF bypass capacitors in parallel in some circuits. The effect is much less pronounced if the larger capacitor is in the microfarad range, however. Here the reactance at resonance is likely to be a fraction of an ohm, and the total parallel-resonant impedance will not rise above an ohm or two.

Real Capacitor Equivalent Circuit: The real behavior of a capacitor is represented in Fig. 3-8. Series inductance L_s and dielectric loss R_s become important at high frequencies, while leakage R_p is important for dc storage over long time periods.

Capacitor Type	Lead Length (cm)	Resonant Frequency (MHz)
1-μF, Mylar	4.0	0.54
0.1-μF Mylar	3.0	2.5
0.01-μF Mylar	3.0	7.0
0.001-μF ceramic	3.0	23
0.001-μF ceramic	0.5	43
100-pF ceramic	3.0	74
100-pF ceramic	0.5	140

(a)

FIGURE 3-7 (a) Real capacitors series resonate with the stray inductance of their leads and plates. Above f_r, reactance is inductive and increases.

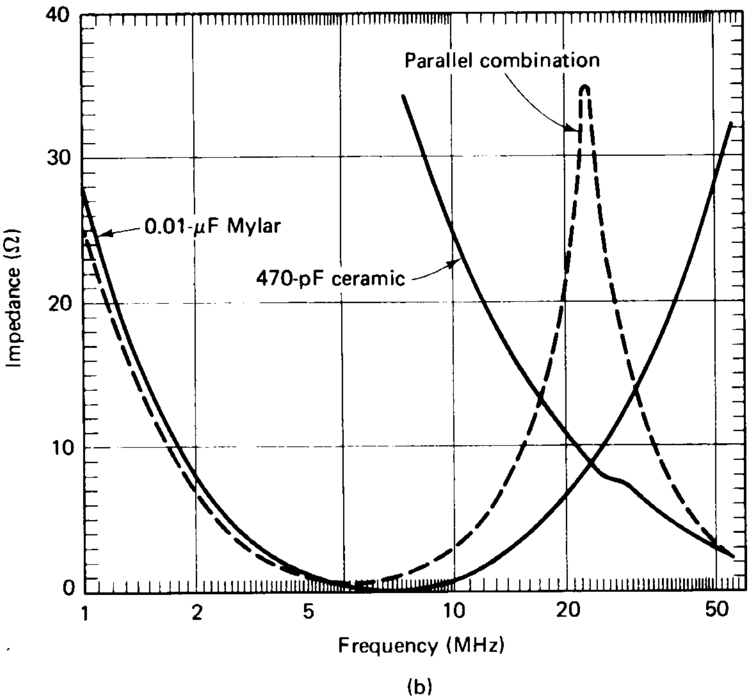


FIGURE 3-7 (b) Parallel resonance of a low- and high-value capacitor produces an impedance rise.

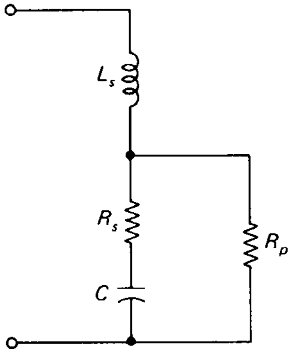


FIGURE 3-8 Equivalent circuit for a real capacitor. L_s is the lead and plate inductance, R_s represents dielectric loss and lead resistance, and R_p is the dc leakage through the dielectric and case.

3.3 MICA- AND CERAMIC-DIELECTRIC CAPACITORS

Mica Capacitors: Mica is a mineral that occurs naturally in thin, flat sheets. As a dielectric material it is quite temperature-stable and has a low dissipation factor, even at high frequencies. It has been used since the early days of radio in stacked-plate capacitors up to about 0.01 μF, although its popularity is finally giving way to ceramic and plastic-film types. Constructed as alternate sheets of metal foil and mica, stacked, bound, and encapsulated in plastic, these capacitors achieve a temperature stability of about ±300 ppm/°C. The silver-mica variety, constructed by depositing a thin silver layer directly on the mica sheets, provides a temperature coefficient not greater than ±100 ppm/°C and are often called zero-temperature-coefficient mica capacitors. Both types have dissipation factors of 0.1% or lower. They are available in tolerances ranging from ±20% to ±1%.

Ceramic Capacitors: Ceramics are a class of mineral compositions which are molded and then fired at high temperature. Ceramic capacitors are occasionally seen in the form of a hollow cylinder, but more often as a disc, as shown in Fig. 3-4, or as a block containing a stack of plates, as shown in Fig. 3-3(b). They are potentially the most confusing things, because they all appear so similar, yet they have widely different properties. The discs invariably have a diameter of between 0.5 and 2 cm, and a single conductive plate on either side of the dielectric which ranges in thickness not much beyond 0.2 to 1 mm. Yet their capacitance values range from 1 pF to 1 μF, a span of six orders of magnitude. This is possible because ceramics can be produced with dielectric constants from 5 to 10,000 or more. However, the quality of the dielectric is severely reduced as its dielectric constant is increased.

Although the range is actually continuous, it might be helpful to think of ceramic capacitors as being divided into two classifications—low-*k* and high-*k* dielectric.

Low-*k* Ceramics include the values to approximately 0.01 μF in disc and 0.2 μF in monolithic (stacked plate) form. Their dissipation factor at 25°C is typically less than 1% at 1 kHz, rising to perhaps 10% at 10 MHz for the larger values. Insulation resistance exceeds 10 GΩ. Tolerances of ±10 and ±20% and capacitance changes of ±2% from −20 to +85°C are common.

High-*k* Ceramics include values from 0.01 to 1 μF in disc and from 0.2 to 6.8 μF in monolithic form. Dissipation factors at 25°C are typically 2 to 5% at 1 kHz, rising to 50% at 1 MHz for the larger values. Tolerances of ±10% are available in monolithic ceramic, but for the larger value discs −20, +80% is common. Capacitance changes of −50% and dissipation factor increases of 500% may be encountered at −20°C in the larger value discs. Insulation resistance also varies drastically, from more than 1 GΩ to less than 10 kΩ depending upon type, value, and voltage rating. In general the high-value discs have the poorest characteristics and the low-value monolithics have the best characteristics. A thorough check of the capacitor's specifications is recommended when high-*k* ceramics are to be used in critical applications. Capacitors for which no tolerance or temperature specifications are given should be regarded with suspicion.

High-*k* ceramic capacitors may exhibit a decrease in capacitance with age on the order of −1% after 1 h, −2% after 10 h, −3% after 100 h, up to −5% at 10,000 h of operation.

EIA Standards for Ceramic Capacitors: In an attempt to bring some order to the confusion over ceramic-capacitor properties, the Electronics Industries Association (EIA) has devised a system of standard ratings, which is reproduced in Fig. 3-9. As an example, a capacitor marked 680 K Y5E would have a capacitance of 680 pF ±10%, with a change of not more than ±4.7% over the temperature range −30 to +85°C.

Temperature Characteristics						←Color→	Tolerance at 25°C (%)	
Low (°C)		High (°C)		ΔC (%)				
X	−55	5	+85	*A*	±1	Gold		–
Y	−30	7	+125	*B*	±1.5	Silver		–
Z	+10			*C*	±2.2	Black	*M*	±20
				D	±3.3	Brown	*F*	±1
				E	±4.7	Red	*G*	±2
				F	±7.5	Orange	*H*	±3
				P	±10	Yellow	*P*	+100-0*
				R	±15	Green	*J*	±5
				S	±22	Blue		–
				T	+22-23	Violet		–
				U	+22-56	Gray	*Z*	+80-20
				V	+22-82	White	*K*	±10

*Tolerance +100-0%, also known as GMV (guaranteed minimum value)

***FIGURE 3-9** EIA standard markings for ceramic capacitors. Example: 0.47M, Y5F indicates 0.47 μF ± 20%, with maximum ±7.5% change from −30 to +85°C.*

Temperature-Compensating Ceramics: By controlling the composition of the ceramic dielectric, it is possible to produce capacitors whose value varies in a predictable and fairly linear manner with temperature (from +100 to about −5000 ppm/°C.) Note that −5000 ppm/°C amounts to a −10% change for a 20°C temperature rise. The table in Fig. 3-10 lists the EIA codes for the standard temperature coefficients available. The most common among these are the zero-temperature-coefficient NPO types, which are often used in critical tuning and filtering circuits, and the N750 (−750 ppm/°C), which is used, often in parallel with NPO or variable types, to offset the positive temperature coefficients of coils in rf tuned circuits. With proper implementation of temperature-compensating capacitors, an rf oscillator can be produced whose frequency drift with temperature is an order of magnitude less than that of an uncompensated oscillator.

NPO ceramics are available from 1 pF to about 0.18 μF in voltage ratings of 50 to 1000 V. N750 types are available to about 1000 pF. These types have dissipation factors of 0.1% or less at 1 MHz and are available in tolerances from ±20% to ±5%. Ceramics with temperature coefficients up to N4700 have slightly higher dissipation factors (0.4% typical).

EIA Code	Temperature Coefficient (ppm/°C)	Temperature Coefficient Tolerance (ppm/°C, +25 to +85°C)	EIA Color
C0	NP0	±30	Black
S1	N030	±30	Brown
U1	N080	±30	Red
P2	N150	±30	Orange
R2	N220	±30	Yellow
S2	N330	±60	Green
T2	N470	±60	Blue
U2	N750	±60	Violet
P3	N1500	±250	Orange - orange
R3	N2200	±500	Yellow - orange
S3	N3300	±500	Green - orange
T3	N4700	±1000	Blue - orange

***FIGURE 3-10** EIA letter and color codes for temperature-compensating capacitors, with typical available tolerances on TC. Example: U2 or violet code indicates nominal negative 750 ppm/°C capacitance change, but may be actually −690 to −810 ppm/°C.*

3.4 PAPER AND PLASTIC-FILM CAPACITORS

Paper Capacitors: Paper-dielectric capacitors are constructed as a cylindrical roll of metal foil and paper, as shown in Fig. 3-3(c). Generally, two papers are used in each layer to minimize the effects of carbon spots that appear in the paper. The dielectric is soaked in various waxes or oils, called *impregnates*, to stabilize the dielectric constant and prevent the intrusion of moisture. Values from 0.001 to 1 μF are commonly available, with voltage ratings from 50 to 1500 V and tolerances of ± 10% or ±20%. Insulation resistance is typically above 10 GΩ, and the dissipation factor is in the vicinity of 1% at 25°C and 1 kHz.

Metallized Paper: Methods have been devised for depositing a very thin metal coating directly on the paper dielectric. This has the advantage of reducing the size and weight of the capacitor by about 30%. Metallized capacitors are also self-healing in many cases; a spark across the dielectric will evaporate the metal film at the defect point, leaving no conductive short circuit. Metallized paper capacitors have characteristics similar to standard paper units, except that their insulation resistance is about an order of magnitude lower.

Plastic-Film Capacitors: Paper dielectrics have been largely replaced in the last two decades by a variety of plastic films: Mylar, metallized Mylar, polyester, polycarbonate, and polystyrene. Mylar and polyester capacitors typically have power factors, insulation resistances, and tolerances which are all a factor of 2 better than their paper counterparts. Polystyrene capacitors are considerably better still, with dissipation factors as low as 0.02%, insulation resistance in excess of 1000 GΩ, and tolerances of ±1%.

Oil-filled Paper Capacitors: These are the "brutes" of the capacitor world, commonly reaching sizes of $10 \times 10 \times 15$ cm and prices of \$50 each. They consist of a steel case containing a paper-dielectric capacitor immersed in liquid oil and hermetically sealed. Their use is dictated when capacitances above 1 μF at voltages above 500 V are required, or when dissipation factor, tolerance, or reliability requirements eliminate the low-cost electrolytic capacitor as an option. Their dissipation factor and insulation resistance are comparable to those of standard paper capacitors.

3.5 ELECTROLYTIC CAPACITORS

Construction: The electrolytic capacitor is formed as a cylindrical roll of aluminum foil and paper layers, much like the paper capacitor. However, the paper is impregnated, not with a dielectric, but with a conductive paste called the *electrolyte*. The dielectric is an oxide of this electrolyte which forms in a very thin layer against the foil when voltage is applied across the capacitor. This forming process may take several minutes for a new capacitor, during which time the capacitor will conduct current and may become noticeably warm. Electrolytics left uncharged for several months may again require a short re-forming period, during which time they are dc-conductive. Their great advantage is their high capacitance per unit volume—electrolytics can typically squeeze 10 times the capacitance into a given case size as a paper type of the same voltage rating, and at low voltages the advantage swells to a factor of 100 or more.

Tantalum electrolytics have become popular for military and industrial use in recent years. They offer the same advantage as aluminum electrolytics, and feature better tolerance, greater reliability, and higher cost, all by about a factor of 4. Tantalums can be formed, not only as a roll, but also as a porous cylindrical slug of tantalum (anode) with an electrolyte filling the pores and conductive to the case (cathode). Figure 3-11 shows a representative selection of electrolytic capacitors.

Polarity: Most electrolytic capacitors are polar and care must be taken to ensure that significant reverse voltage is never applied across them. Where the markings are obscure, the metal case is almost always the negative lead. Momentary reverse voltage will not usually destroy the capacitor, but may require its re-forming. Sustained large reverse voltages will cause current conduction, overheating, and, in the larger sizes, actual explosion of the capacitor, sometimes quite violently.

Nonpolar electrolytics are available, although they are not nearly so popular as the polar type. The letters *NP* after the voltage rating designate a nonpolar electrolytic. Regular polar electrolytic capacitors can be used safely with reverse voltages of 1 or 2 V, but the leakage current goes up rapidly for reverse voltages much beyond this. For higher voltages, two polar electrolytics can be connected in series back to back to provide a single nonpolar unit.

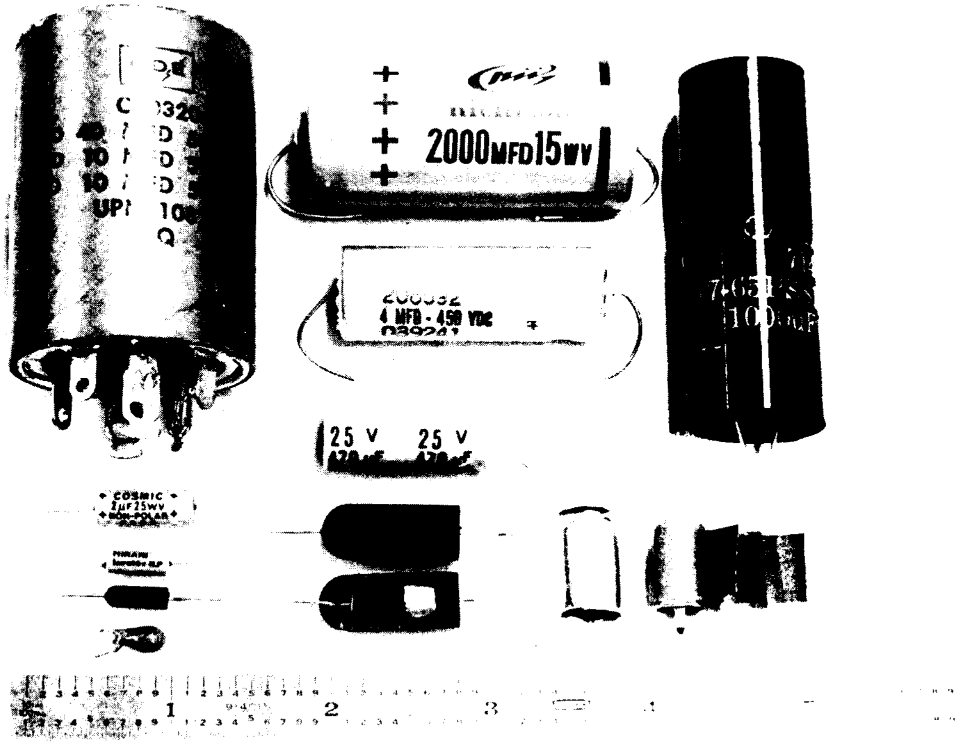


FIGURE 3-11 Representative electrolytic capacitors. Top left: *three-section FP (flat plate) chassis-mount type;* top center: *low- and high-voltage axial-lead types, with radial-lead (also called vertical or PC-mount) types below and at right;* bottom left: *two nonpolar electrolytics with two small tantalums beneath;* bottom center: *tantalum slug type with broken-away unit showing interior;* bottom right: *case and foil-paper wrappings from an aluminum electrolytic.*

Voltage Rating: Aluminum and tantalum-foil electrolytics are limited to voltage ratings of 550 V or less by the nature of the dielectric. Slug-type tantalums are generally confined to applications requiring a rating of 100 V or less. Electrolytics are generally rated for continuous use at their *working* voltage, and for short-term (such as turn-on or warm-up) use at their *surge* voltage. *Surge* rating is typically 130% of *working* rating (e.g., 8 MFD, 150 WVDC, 200 V SURGE). Where higher-voltage ratings are required, two or more electrolytics may be placed in series and their voltage ratings added (their capacitance being halved, of course). It is advisable in this case to place equalizing resistors across each capacitor, as shown in Fig. 3-12, to ensure that the dc voltage divides equally across each unit. The resistors should be calculated to draw 10 to 100 times the expected leakage current of the capacitors (see below), but a value of 100 kΩ is common.

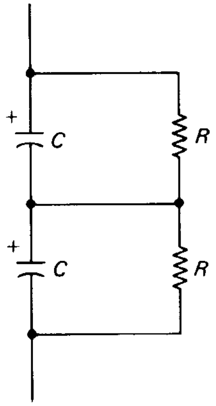


FIGURE 3-12 Equalizing resistors are used across series electrolytics to ensure equal division of high dc voltages.

Nonpolar electrolytics intended for motor-starting applications are often rated for rms ac voltage. The peak-conversion factor of 1.41 can be applied to convert this rating to dc voltage.

Reliability: Electrolytic capacitors have a relatively high failure rate, so their use is avoided in high-reliability designs, except where their high capacitance-to-size ratio makes them the only choice. The best tantalums have expected failure rates of 0.1% per 1000 h (compared to 0.01 or 0.001% for typical carbon resistors), and low-cost aluminum electrolytics are much worse. In view of this, the wise designer will choose electrolytics with a working-voltage rating 25 to 50% higher than the intended application and will take steps to keep the operating temperature of the capacitors as low as possible.

Tolerance and Temperature Effects: Aluminum electrolytics are marketed with tolerances between -10 $+50\%$ and -20 $+100\%$. They are intended for filtering, coupling, and bypassing and not for timing and resonance applications. Tantalum electrolytics can be purchased with tolerances as good as $\pm 10\%$, making them suitable for many applications where aluminum will not do. It should be noted that all these tolerances are specified at $+25°C$. The capacitance of aluminum electrolytics will typically drop by 20% at $-20°C$ and by 50% at $-40°C$. Tantalum generally fares somewhat better, dropping by typically 20% at $-40°C$ and 50% at $-55°C$. Capacitance change at high temperatures is positive, typically $+10\%$ at 85°C for aluminum and $+5\%$ at 125°C for tantalum. Figure 3-13(a) and (b) documents this.

Leakage Current and Dissipation Factor: Electrolytic capacitors generally exhibit dc leakage currents which amount to a shunt resistance between 1 MΩ and 10 MΩ at 25°C. This leakage resistance is relatively constant for various voltage ratings and for values up to 100 μF, but begins to drop inversely with capacitance values above 100 μF. Figure 3-13(c) shows how this leakage increases with voltage and

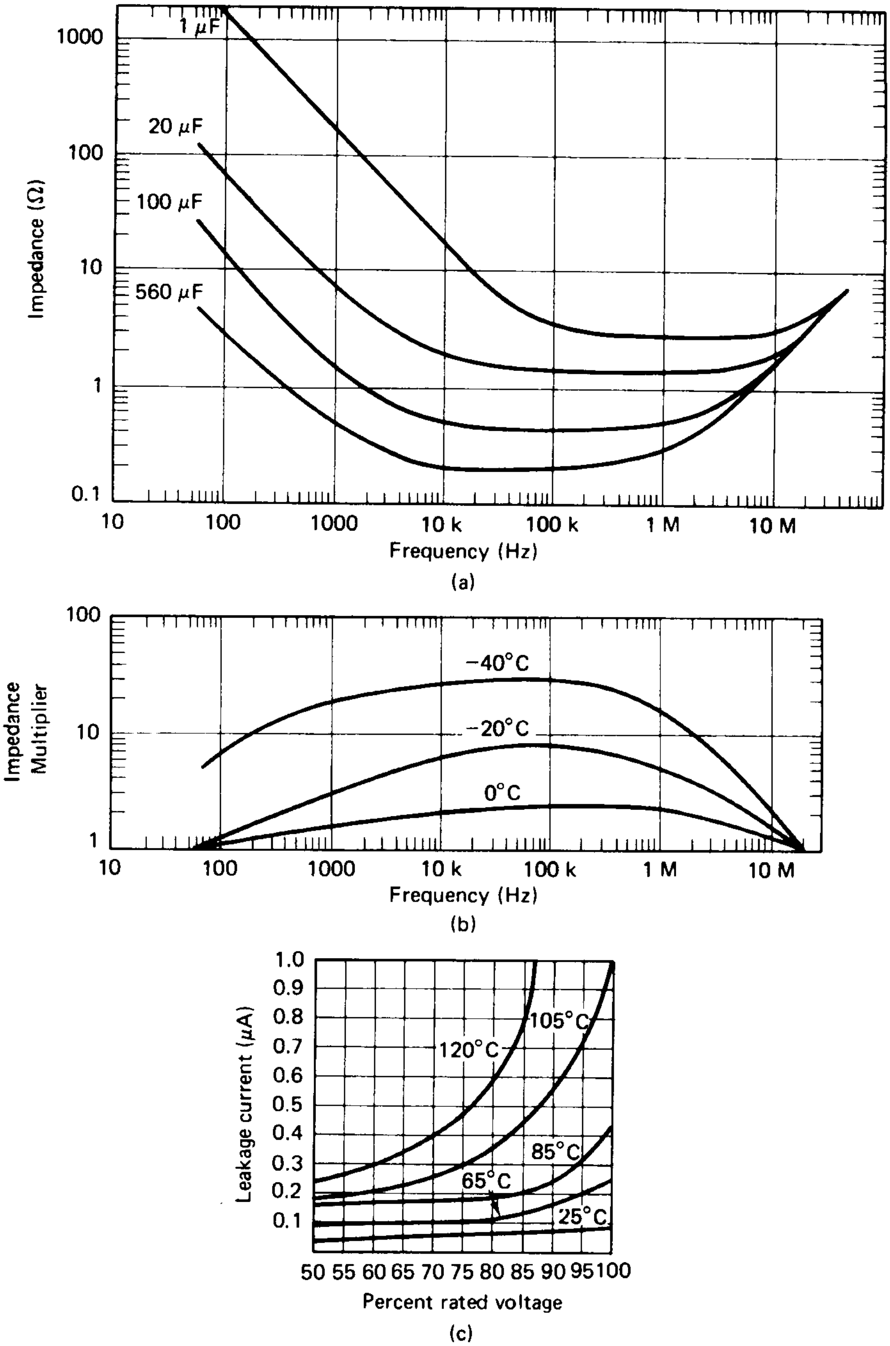


FIGURE 3-13 *(a) Impedance versus frequency for typical aluminum electrolytics at 25°C. (b) Correction factors, showing effective capacitance drop at low temperatures. (c) Leakage current in an electrolytic increases rapidly above rated voltage, resulting in heat generation. Higher temperatures further increase leakage, so the heat generation can quickly avalanche. (Courtesy General Electric Company.)*

Operating Life: Electrolytic capacitors require a moist electrolyte, and if this is allowed to dry out, the dissipation factors will increase and the capacitance will decrease, eventually to the point of failure. At 25°C, an electrolytic capacitor with a simple rubber seal may be expected to last for 10 to 20 years, but, as shown by Fig. 3-14, the life at 85°C may be reduced to 1 year. Operation at low atmospheric pressures further accelerates dry-out. Improved seals and encapsulation increase dry-out time, the ultimate being the hermetically sealed unit of Fig. 3-15, which anticipates a life in excess of 50 years in a vacuum at 125°C.

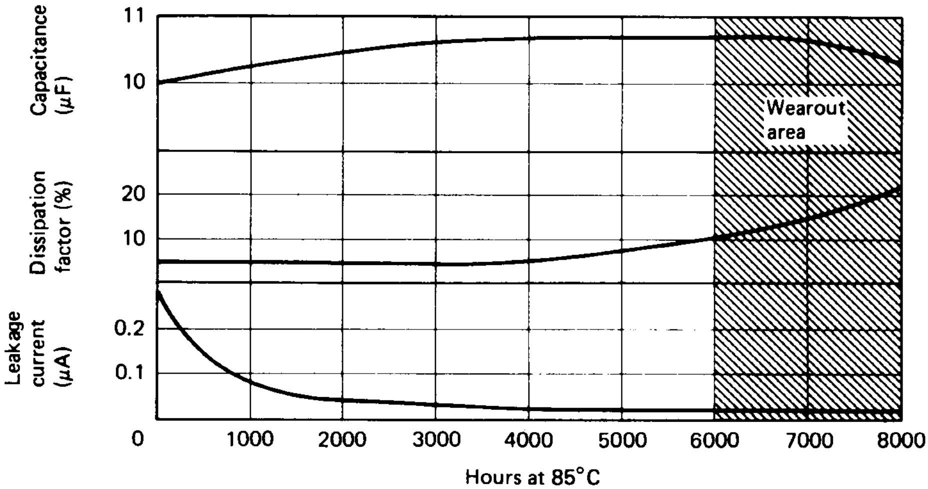


FIGURE 3-14 Expected load life at 85°C. The electrolyte drys out with age, giving the typical electrolytic capacitor a limited life span. High temperature greatly reduces life expectancy. (Courtesy General Electric Company.)

Residual Charge: Figure 3-1 shows how the dielectric is polarized by the electric field between the plates of a charged capacitor. In an electrolytic, the dielectric requires a considerable period of time (0.1 to 10 s) to depolarize after discharge, and a residual charge of perhaps 10% can be expected to appear after a brief discharge, as shown in Fig. 3-16. A second discharge, or a very long first discharge, will be required to establish true zero voltage across a charged electrolytic. Paper, Mylar, and ceramic capacitors will recharge in this manner to only about 0.1% of original charge, so the problem is generally negligible in their case.

3.6 VARIABLE CAPACITORS

Variable Capacitors: These capacitors are most commonly used for tuning radio-frequency circuits. They are available with maximum capacitance from about 5 to 500 pF, and in voltage ratings from 300 to several thousand volts. Minimum capacitance is generally $\frac{1}{10}$ to $\frac{1}{50}$ of the maximum value. Their structure, illustrated

Tubular Tantalum-foil Capacitors

KEY TO MILITARY TYPE DESIGNATION (FOR MIL-C-3965E)

Example: CL31 B J 200 M P G

Style	Description
CL20	Polar, etch foil, uninsulated
CL21	Polar, etch foil, insulated
CL22	Non polar, etch foil, uninsulated
CL23	Non polar, etch foil, insulated
CL30	Polar, plain foil, uninsulated
CL31	Polar, etch foil, insulated
CL32	Non polar, plain foil, uninsulated
CL33	Non polar, plain foil, insulated
CL51	Polar, plain foil, KSR®
CL52	Non polar, plain foil, KSR
CL53	Polar, etch foil, KSR
CL54	Non polar, etch foil, KSR
CL55	Polar, wet-slug, package
CL64	Polar, wet-slug, uninsulated
CL65	Polar, wet-slug, insulated
CL66	Polar, wet-slug, hermetic seal, uninsulated
CL67	Polar, wet-slug, hermetic seal, insulated
CL70	Polar, high-etch foil, uninsulated
CL71	Polar, high-etch foil, insulated
CL72	Non polar, high-etch foil, uninsulated
CL73	Non polar, high-etch foil, insulated

Characteristic
−55 C to 85 C
(Will operate up to 125 C when properly derated)

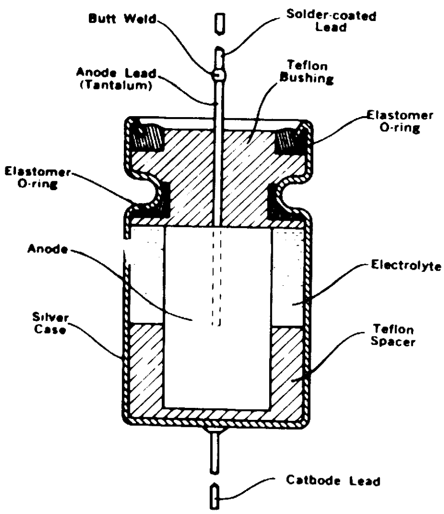
Capacitance — Expressed in (μf)
First two digits represent significant figures; third digit represents number of zeroes to follow. "R" indicates decimal point.

Voltage Symbol	Volts 85 C/125 C	Voltage Symbol	Volts 85 C/125 C
A	3/2	N	100/65
B	6/4	O	250/165
C	8/5	P	125/85
D	10/7	Q	150/100
E	15/10	R	180/120
F	20/13	S	270/180
G	25/15	T	360/240
H	30/20	U	450/300
I	112/75	V	540/360
J	50/30	W	630/420
K	60/40	X	300/200
L	75/50	Y	375/250
M	90/60	Z	200/150

Seal
G—Hermetic (Glass)
E—Non hermetic (Elastomer)

Construction
P—Polar
N—Non polar

Capacitance Tolerance Symbol	%
J	±5
K	±10
L	±15
M	±20
S	±30-15
T	±50-15
U	±75-15



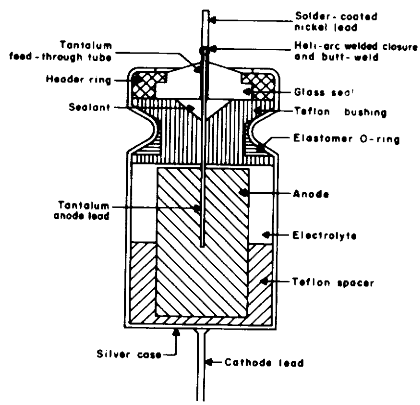


***FIGURE 3-15** (a) MIL-STD designations for electrolytic capacitors. (b) Typical rubber seal to retard drying of electrolyte. (c) Hermetic seal which greatly increases life span. (Courtesy General Electric Company.)*

in Fig. 3-17, consists of a number of half-circle-shaped *rotor* plates (1 to 20, depending on capacitance) which are turned on a shaft to mesh in or out of a number of *stator* plates. The rotor plates are generally connected to a U-shaped metal support, while the stator plates remain insulated. High-quality variables use ballbearings at two points on the shaft and ceramic insulation. The mechanical rigidity of the plates has much to do with the ultimate stability of a radio-frequency tuning system, and thick, stout plates should be sought for critical applications.

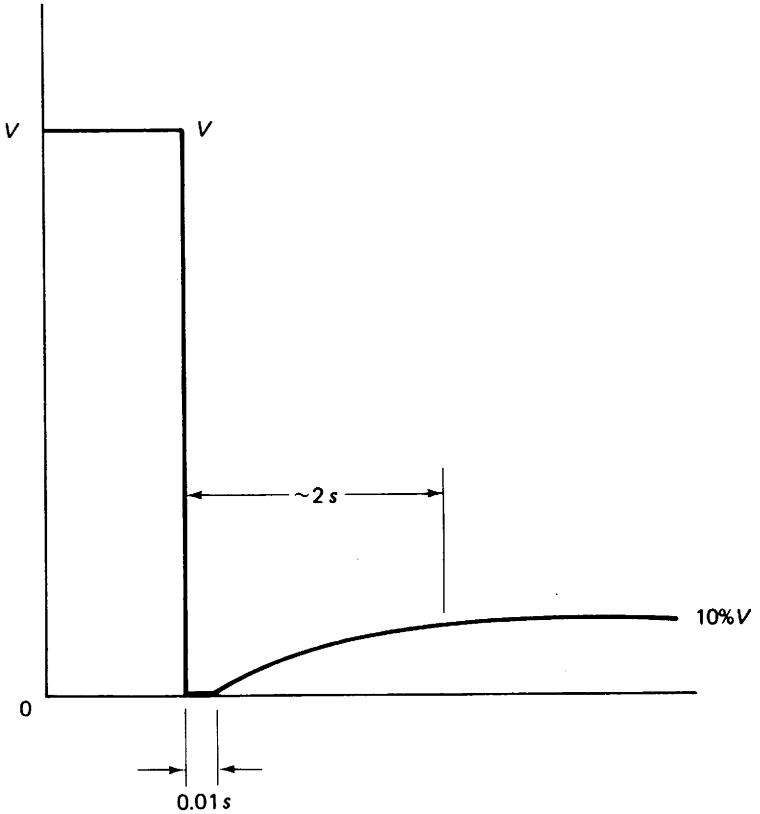


FIGURE 3-16 A charged electrolytic that is discharged to zero for a brief period will recharge to as much as 10% of its initial voltage due to polarization in the dielectric.

If the shaft is placed at the center of the plate, a straight-line curve of capacitance versus shaft rotation is obtained. However, if the shaft is offset to one side, the curve can be made nearly linear for frequency versus shaft rotation, which is more desirable for radio-tuning applications.

Ganged Capacitors: Often, several circuits must be tuned simultaneously to the same frequency or to related frequencies. Variable capacitors are available with two, three, or four sections ganged together on the same shaft, as shown in Fig. 3.17. The outside plate is often split in four or five places, with the intent that it be bent out or in to make that section *track* with the other sections, or with the tuning dial. The rotors of all the sections are already electrically connected, and often the stators are also paralleled to produce a high-capacitance variable.

Butterfly capacitors, illustrated in Fig. 3.17, provide a symmetrically balanced two-gang variable which covers its range in $\frac{1}{4}$ turn. The differential variable capacitor is a two-gang unit in which one section increases as the other decreases, the sum being essentially constant.

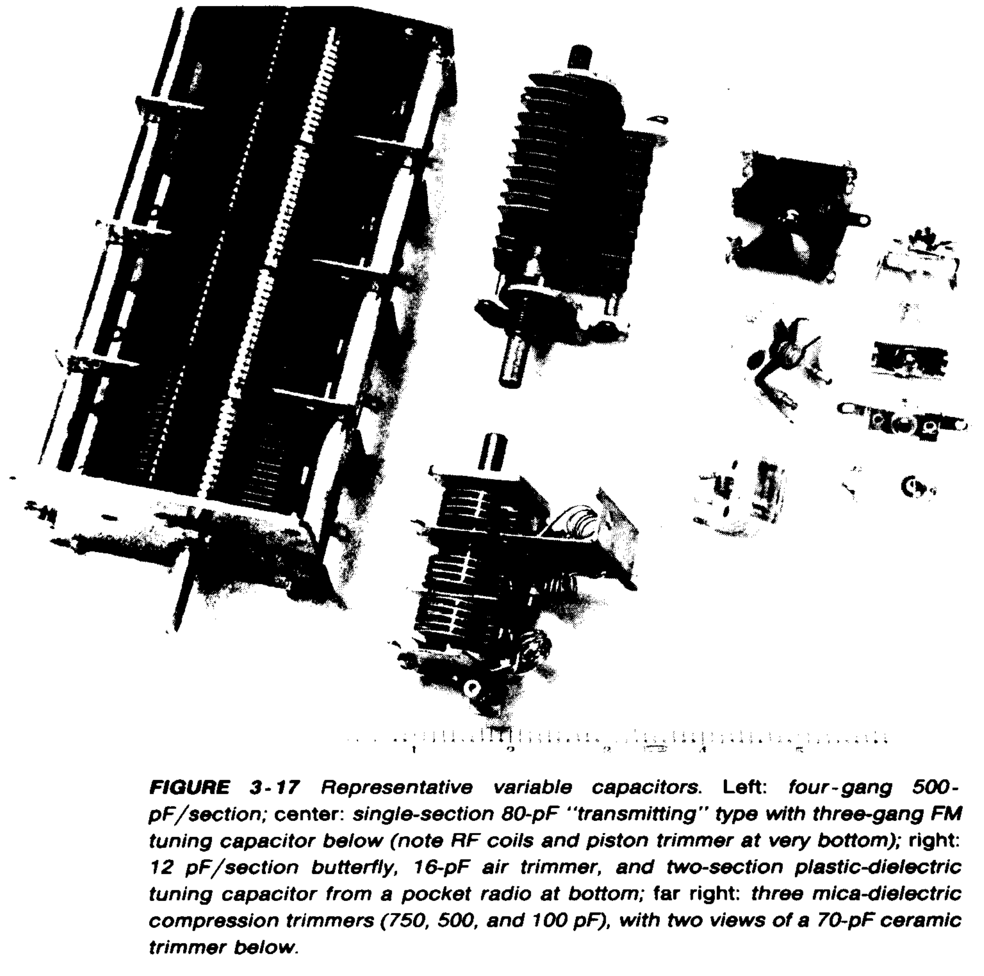


FIGURE 3-17 Representative variable capacitors. Left: *four-gang 500-pF/section;* center: *single-section 80-pF "transmitting" type with three-gang FM tuning capacitor below (note RF coils and piston trimmer at very bottom);* right: *12 pF/section butterfly, 16-pF air trimmer, and two-section plastic-dielectric tuning capacitor from a pocket radio at bottom;* far right: *three mica-dielectric compression trimmers (750, 500, and 100 pF), with two views of a 70-pF ceramic trimmer below.*

Trimmer and Padder Capacitors: Trimmer capacitors are variables that are intended to be adjusted with a screwdriver rather than a dial. They are available as air/plate units, piston type, ceramic, and compression types, all of which are illustrated in Fig. 3-17. Compression trimmers with capacitance from 500 to 5000 pF are generally called *padders*. The following table and Fig. 3-18 lists their relative advantages.

Type	*Minimum Size* (pF)	*Maximum Size* (pF)	*Comments*
Air/plate	4.2–1.2	25–2.5	Stable, expensive, 180° rotation
Piston	3–0.5	50–5	10–40 turns to cover range
Ceramic	7–1.5	125–20	$Q \approx 500$ at 1 MHz; available with temperature coefficient NPO to N 650
Compression padder	12–1	3055–1400	Inexpensive; severe capacitance drift with time and temperature as plates warp.

	Normal range min/max (µF, pF)	Normal voltage min/max (V)	Tolerance at 25°C low cost/best (%)	Capacitance change from 25°C (%) −20°C/+80°C	Operating temperature min/max (°C)	Dissipation factor at 25°C 1 kHz/1MHz (%) *120 Hz	Frequency limit at 25°C (MHz)	Insulation resistance (25°C/85°C)(GΩ-Fd)	Cost per standard unit ($)	Volume per standard unit (cm^3)	Standard unit
Air/Vacuum	1 pF 1000 pF	300 V 10 kV	–	–	–	0.001 0.001	10,000	> 10,000	1.00	16	100pF/500 V
Mica-foil	2 pF 0.02 µF	100 V 1 kV	± 20 ± 5	−0.2 +0.2	−55 + 150	0.1 0.1	1000	100 20	0.13	0.35	
Silver mica	2 pF 1000 pF	100 V 1 kV	± 10 ± 1	−0.1 +0.1	−55 + 150	0.05 0.05	1000	100 20	0.13	0.35	
Low-*k* disc ceramic	2 pF 0.01 µF	500 V 6 kV	+ 80 − 10 ± 5	+5 +5	− 55 + 100	0.5-2 1-5	200	50 1	0.06	0.12	
Low-*k* monolithic ceramic	1 pF 0.5 µ	50 200	± 1 ± 10	+5 −2	−55 +125	0.2−1 0.5-4	200	100 10	0.16	0.04	
Polystyrene	5 pF 0.01 µF	50 V 1 kV	± 10 ± 1	+0.1 −0.4	− 55 + 85	0.05 0.1	100	1000 75	0.09	0.21	
Temp comp ceramic	1 pF 1000 pF	500 V 6 kV	± 20 ± 5	to +20 to −30	− 55 + 125	0.1 0.1	1000	100 2	0.07	0.12	
Paper-foil	0.001 µF 1 µF	50 6 kV	± 20 ± 2	−2.0 +10	− 20 + 125	1 10	2	2 0.02	1.00	6.0	1.0 µF/50 V
Metalized paper	0.01 µF 1 µF	50 V 500 V	± 20 ± 5	−1.5 +8	− 55 + 125	1.5 15	1	0.2 0.005	0.60	4.0	
Mylar-foil	200 pF 10 µF	50 V 1 kV	± 20 ± 1	−0.3 +3	− 55 + 125	0.5 5	2	20 1	0.45	5.0	
High-*k* disc ceramic	0.01 µF 2.2 µF	3 V 500 V	+ 80-10 ± 20	−70 −35	− 55 + 85	1-5 10-70	1-20	1 to 0.0001	1.00	0.80	
High-*k* monolithic ceramic	0.5 µ 6.8 µ	50 200	± 10 ± 20	− 10 −20	−55 +125	0.5−2 2-6	1-20	10 0.2	0.70	0.22	
Aluminum electrolytic	0.5 µF 180,000 µF	3 V 525 V	+ 80-10 ± 20	−25 +20	− 20 + 85	10-30* > 100	0.1	$15M \parallel \frac{250}{C}$ $4M \parallel \frac{50}{C}$	0.20	0.30	
Tantalum (wet)	0.1 µF 1000 µF	3 V 100 V	± 20 ± 10	−2 +6	− 40 + 125	5-20* > 100	0.05	$25M \parallel \frac{500}{C}$	0.50	0.10	
Tantalum (dry)	0.1 µF 500 µF	3 V 100 V	± 20 ± 5	−2 +5	− 55 + 125	5-20* > 100	0.05	$10M \parallel \frac{200}{C}$	0.35	0.15	
	Range	Volts	Tol	TC	Temp	DF	f max	IR	Cost	Vol.	

***FIGURE 3-18** Summary of real properties of fixed capacitors.*

4

PROPERTIES OF REAL INDUCTORS

4.1 BASIC INDUCTANCE AND REACTANCE

Analysis of real inductor and transformer behavior requires a mastery of basic ac-circuit concepts. A condensed treatment of these topics is given in this section and the ones following, both as a review and as a ready reference.

Inductance: If a current is run through a coil of wire, a magnetic field is set up whose strength is proportional to the current and the number of turns. This is the principle of the electromagnet. If a magnetic field is moved through a coil of wire, a voltage is induced across the coil whose magnitude is proportional to the speed of movement and the number of turns. This is the principle of the generator. Figure 4-1(a) and (b) illustrates these effects. Combining these facts, if ac is passed through a coil, a changing (moving) magnetic field is set up, and this field induces a voltage in the coil that caused it. The direction of this self-induced voltage is always such as to oppose the initially applied voltage. Thus a coil (we call it an inductor—never use one syllable when you can use three) opposes any change in current.

Reactance: Notice that the inductor's opposition to ac is by means of a reverse voltage, not a restriction in the conductive path as in the case of a resistor. We call this kind of opposition *reactance* as opposed to resistance. The main difference is

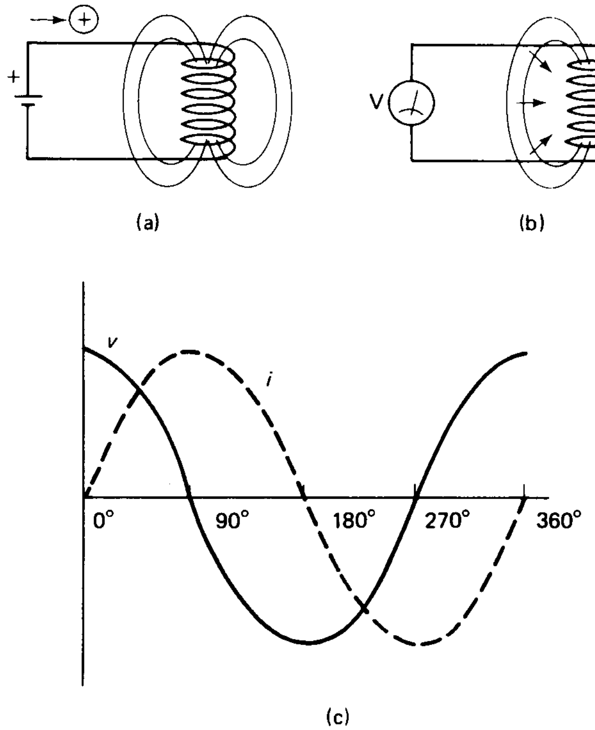


FIGURE 4-1 (a) Current through a coil sets up a magnetic field. (b) Moving a magnetic field through a coil causes a voltage to be induced across it. (c) Self-induced voltage causes current to flow back against the direction of applied voltage from 90° to 180° and again from 270° to 360°.

that reactance does not burn energy—it merely stores it on one quarter-cycle and then bounces it back into the source on the next quarter-cycle.

This bounce-back results in a phase shift between current and voltage for an inductor. Current through the coil reaches a maximum just as the voltage across the coil falls back to zero. We say that current in an inductance lags voltage by 90°. This is shown in Fig. 4-1(c).

The amount of reactance that a coil presents to ac depends on the value of the inductance and the speed with which the current alternates (frequency):

$$X_L = 2\pi f L \tag{4-1}$$

where X_L is the reactance in ohms, f the frequency in hertz, and L the inductance in henrys.

Series and Parallel Reactances: Two inductors in series have a total reactance equal to their sum (just like series resistors), provided that their magnetic fields are not close enough to interact. This is illustrated in Fig. 4-2(a).

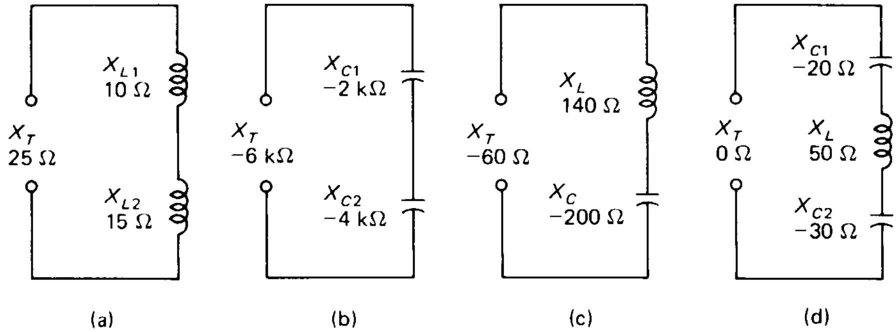


***FIGURE 4-2** Reactances series combine by simple algebraic addition, with capacitive reactance considered negative.*

Capacitors store energy in an electric field, and when connected to an ac source, simply bounce this energy back to the source on alternate quarter-cycles. Thus they present a reactance to ac, much like an inductor. However, there are two important differences. First, the current in a capacitance *leads* the voltage by 90°; just the opposite of the inductor. We give a negative sign to the reactance of a capacitor:

$$X_C = \frac{-1}{2\pi f C} \tag{4-2}$$

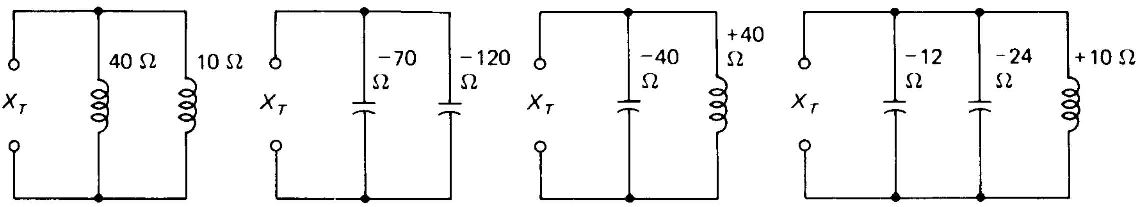
where C is the value of the capacitance in farads.

Second, as shown by equation 4-2, capacitive reactance varies inversely with frequency f and capacitance C, again just the opposite of the inductor. This means that as frequency increases, inductive reactance goes up while capacitive reactance goes down. Calculations involving reactances are, therefore, valid only at the frequency for which reactance has been calculated. Capacitive and inductive reactances in series are subtractive, as illustrated in Fig. 4-2(c) and (d).

Reactances in parallel are governed by the same formulas as resistors, namely, the product over the sum and the reciprocal of the reciprocals. In this case, however, attention must be paid to the signs. Figure 4-3 gives some examples.

4.2 BASIC CONCEPTS OF IMPEDANCE

Series Resistance and Reactance: When resistance and reactance are placed in series, the result is termed *impedance*. Part of the energy delivered by the source is then dissipated as heat (in the resistance) and part is reflected back to the source (by the reactance). Current will lag applied voltage if the net reactance is inductive, and lead if the net reactance is capacitive, but by an angle between zero and 90°. The total value of the impedance Z and the exact phase angle θ are determined by phasor addition of X and R, as shown in Fig. 4-4. The calculations proceed as



(a)

$$X_T = \frac{X_1 X_2}{X_1 + X_2}$$

$$X_T = \frac{40 \times 10}{40 + 10}$$

$$X_T = 8\ \Omega$$

(b)

$$X_T = \frac{X_1 X_2}{X_1 + X_2}$$

$$X_T = \frac{(-70)(-120)}{-70 - 120}$$

$$X_T = -44\ \Omega$$

(c)

$$X_T = \frac{X_1 X_2}{X_1 + X_2}$$

$$X_T = \frac{(-40)(40)}{-40 + 40}$$

$$X_T \to \infty$$

(d)

$$X_T = \frac{1}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3}}$$

$$X_T = \frac{1}{\frac{1}{-12} + \frac{1}{-24} + \frac{1}{10}}$$

$$X_T = -40\ \Omega$$

FIGURE 4-3 ***Reactances in parallel combine by the product over the sum or the reciprocal of the reciprocals as do resistors. Again, capacitive reactance is negative.***

follows:

$$Z = \sqrt{X_s^2 + R_s^2} \tag{4-3}$$

$$= \sqrt{30^2 + 40^2} = \mathbf{50\ \Omega}$$

$$\theta = \text{Arctan}\ \frac{X_s}{R_s} \tag{4-4}$$

$$= \text{Arctan}\ \frac{30}{40}$$

$$= \text{Arctan}\ 0.75 = \mathbf{37°}$$

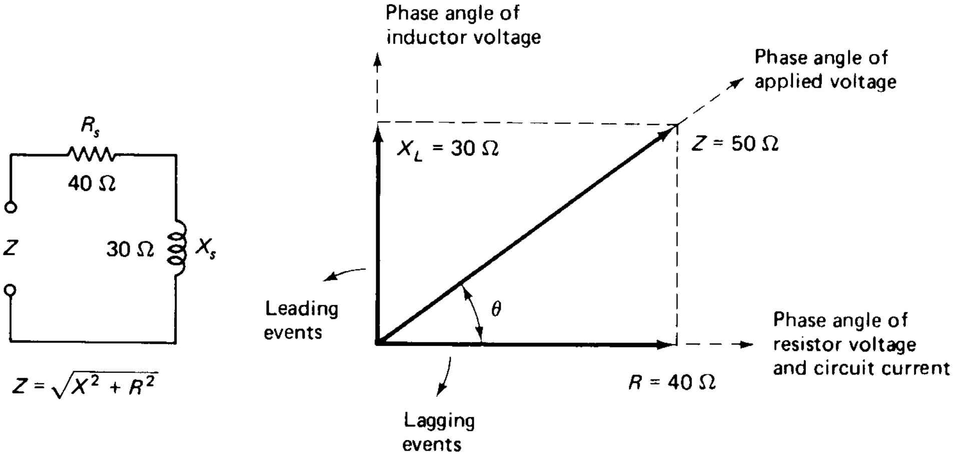


FIGURE 4-4 ***Resistance and reactance cannot be added or subtracted directly. The combination, called impedance (Z), is obtained by phasor addition.***

Notice that the total circuit current is in phase with the resistor voltage and lags the applied voltage by 37°.

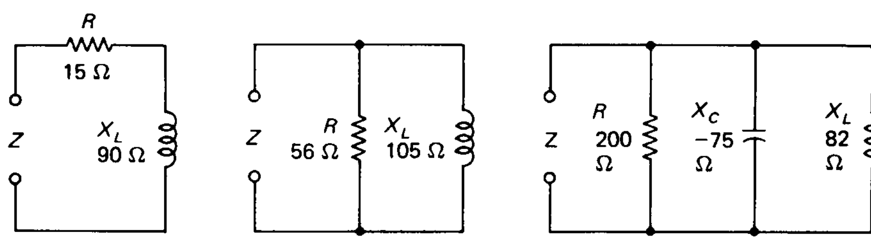
A few trial calculations with equation 4-3 will show that, if X_s is three times R_s or more, the total impedance is only 5% greater than X_s. Likewise, if $R_s > 3X_s$, Z is greater than R_s by only 5%. We will often find it helpful to neglect the lower of X_s or R_s in these cases, so $Z \approx R_s$ if X_s is small, or $Z_s \approx X_s$ if R_s is small.

Parallel Resistance and Reactance: When resistance and reactance appear in parallel, the equivalent impedance is always less than the smaller of the two, but never smaller than 0.707 times the lower one. The phase shift is characteristic of the reactance involved: total current lagging applied voltage for inductance, leading for capacitance. The exact impedance and phase angle can be calculated:

$$Z = \frac{X_p R_p}{\sqrt{X_p^2 + R_p^2}} \tag{4-5}$$

$$\theta = \text{Arctan}\, \frac{R_p}{X_p} \tag{4-6}$$

In the parallel circuit it can be shown that, if X_p is at least three times R_p (or vice versa), the impedance of the parallel combination is less than the lower value (X_p or R_p) by only 5%. Here, again, we will often find it appropriate and helpful to disregard the higher parallel value (X_p or R_p), so $Z \approx X_p$ if R_p is large, or $Z \approx R_p$ if X_p is large. Such simplifications can be of great help in analyzing real inductor and transformer equivalent circuits. A few examples are given in Fig. 4-5.



(a)

$Z = \sqrt{X^2 + R^2}$

$= \sqrt{90^2 + 15^2}$

$= 91.2\ \Omega$

$\approx X_L$

(b)

$Z = \dfrac{X R}{\sqrt{X^2 + R^2}}$

$= \dfrac{105 \times 56}{\sqrt{105^2 + 56^2}}$

$= 49.4\ \Omega$

(c)

$X_C \| X_L = \dfrac{82(-75)}{82 - 75}$

$= -879\ \Omega$

$X > 3R$; therefore,

$Z \approx R = 200\ \Omega$

FIGURE 4-5 (a) Series resistance is negligible if reactance is three or more times larger. (b) Equation 4-5 always gives parallel impedance lower than the smaller of X or R. (c) The parallel reactances are combined first, giving a net reactance much larger than R, so that $Z \approx X_p$. The exact Z, by equation 4-5, is 195 Ω.

Equivalent Series and Parallel Circuits: Any parallel R–X circuit can be replaced by an equivalent series R–X circuit using the following equations:

$$R_s = X_p \frac{X_p R_p}{X_p^2 + R_p^2} \tag{4-7}$$

$$X_s = R_p \frac{X_p R_p}{X_p^2 + R_p^2} \tag{4-8}$$

Conversely, any series R–X circuit can be replaced by an equivalent parallel circuit using the equations

$$R_p = \frac{X_s^2 + R_s^2}{R_s} \tag{4-9}$$

$$X_p = \frac{X_s^2 + R_s^2}{X_s} \tag{4-10}$$

Of course, these equivalencies are valid only at the particular frequency for which the reactance has been calculated. Also note that all of the reactances are to be placed into these formulas as absolute (unsigned) quantities. Although capacitive reactance has been termed negative, it can be subtracted only from inductive reactance ($X_L - X_C$ valid), not from resistance ($R - X_C$ invalid).

Using the reactance combination methods illustrated in Figs. 4-2 and 4-3 and the equivalency formulas 4-7 through 4-10, it is possible to reduce most circuits containing numerous R, X_L, and X_C components to a single series or parallel R–X equivalent.

EXAMPLE 4-1

Figure 4-6(a) shows a series-parallel RLC circuit. It happens to represent a real transformer, but we'll get to that later. For now, determine the power delivered to R_2, which represents the load.

Solution

First it is necessary to calculate the reactances:

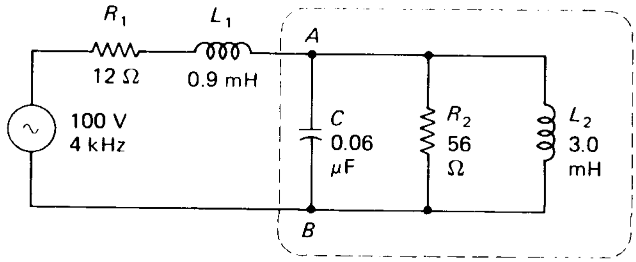
$$X_{L1} = 2\pi f L_1 \qquad X_{L2} = 2\pi f L_2 \qquad X_C = \frac{-1}{2\pi f C}$$

$$= 2\pi 4000 \times 0.9 \times 10^{-3} \qquad = 2\pi 4000 \times 3 \times 10^{-3} \qquad = \frac{-1}{2\pi 4000 \times 0.06 \times 10^{-6}}$$

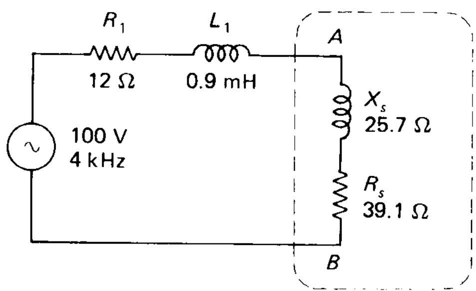
$$= 22.6\ \Omega \qquad = 75.4\ \Omega \qquad = -663\ \Omega$$

X_{L2} and X_C can be parallel-combined:

$$X_p = \frac{X_{L2} X_C}{X_{L2} + X_C} = \frac{75.4(-663)}{75.4 - 663} = +85.1\ \Omega$$



(a)



(b)

***FIGURE 4-6** Circuit for Example 4-1, illustrating parallel-to-series X–R conversion to obtain a solution: (a) initial transformer representation; (b) reduced to a simple series equivalent.*

Now the parallel combination of R_2 and X_p can be converted to its series equivalent:

$$R_s = X_p \frac{X_p R_2}{X_p^2 + R_2^2} = 85.1 \times 0.459 = 39.1\ \Omega$$

$$X_s = R_p \frac{X_p R_2}{X_p^2 + R_2^2} = 56 \times 0.459 = 25.7\ \Omega$$

The equivalent circuit now stands as in Fig. 4-6(b), and consists of a total series resistance and inductive reactance:

$$R_{sT} = R_1 + R_s = 12 + 39.1 = 51.1\ \Omega$$

$$X_{sT} = X_{L1} + X_s = 22.6 + 25.7 = 48.3\ \Omega$$

$$Z_T = \sqrt{X_{sT}^2 + R_{sT}^2} = \sqrt{48.3^2 + 51.1^2} = 70.3\ \Omega$$

The total current is

$$I = \frac{V}{Z} = \frac{100}{70.3} = 1.42\ \text{A}$$

This current all flows through the impedance of C, R_2, and L_2 in parallel, producing a voltage across R_2:

$$Z_p = \frac{X_p R_2}{\sqrt{X_p^2 + R_2^2}} = \frac{85.1 \times 56}{\sqrt{85.1^2 + 56^2}} = 46.8\ \Omega$$

$$V_{R2} = IZ_p = 1.42 \times 46.8 = 66.5\ V$$

$$P_{R2} = \frac{V^2}{R} = \frac{66.5^2}{56} = \mathbf{78.9\ W}$$

4.3 BASIC RESONANCE CONCEPTS

This section discusses the effects of resonance and tuned circuits. After that we will be equipped to start talking about real inductors.

Series Resonance: You may have noticed that in the example of Fig. 4-2(d), the inductive and capacitive reactances exactly canceled, leaving a net reactance of zero. This short-circuit condition prevails at only one frequency, called the *resonant* frequency, which is given by

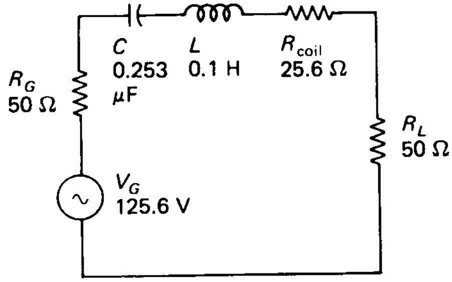
$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{4-11}$$

where f_r is in hertz, L is in henrys, and C is in farads. At lower frequencies the capacitive reactance predominates and at higher frequencies the inductive reactance predominates. We therefore have a *series tuned circuit* which can be placed in series with a line to pass only the resonant frequency or in shunt across the line to trap out the resonant frequency.

Series Circuit Q: Of course, the series tuned circuit is not a perfect short circuit at resonance. The coil has some resistance, perhaps increased by the skin effect (Section 1.4) if the frequency is high, and at very high frequencies the capacitor may have an appreciable dissipation factor, which can be represented as additional series resistance according to equation 3-4. The ratio of the reactance of either the coil or the capacitor at resonance to the total effective series resistance is termed the *quality factor* or Q of the series tuned circuit:

$$Q = \frac{X_{Ls}}{R_s} = \frac{X_{Cs}}{R_s} \quad \text{at resonance} \tag{4-12}$$

When a series tuned circuit is used in series with a load resistance and/or a source resistance, the entire series resistance must be included as R_s in determining Q. This is illustrated in Fig. 4-7(a).



$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

$$= \frac{1}{2\pi\sqrt{0.1 \times 0.253 \times 10^{-6}}}$$

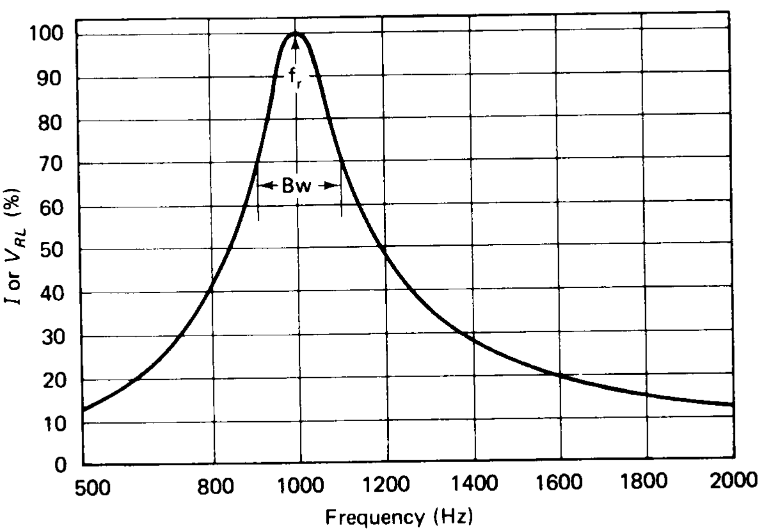
$$= 1000 \text{ Hz}$$

$$X_L = 2\pi f L$$

$$= 2\pi \times 1000 \times 0.1$$

$$= 628\ \Omega$$

(a)



$$Q = \frac{X_L}{R_T} = \frac{628}{125.6} = 5$$

$$\text{Bw} \approx \frac{f_r}{Q} = \frac{1000}{5} = 200 \text{ Hz}$$

(b)

***FIGURE 4-7** Q (quality) of a series-tuned circuit is defined as X/R, where X is either X_C or X_L (equal at resonance) and R is the total resistance in series at resonance: (a) the resonant circuit; (b) response at frequencies off resonance. Bandwidth Bw is defined as the span between the −3 dB (0.707 voltage) points.*

Bandwidth: The current through a series tuned circuit is maximum at resonance, and decreases above or below that frequency as net inductive or capacitive reactance appears in the circuit. Series-circuit analysis, as illustrated in Fig. 4-4, can be used to determine the current response at any frequency, and a graph of the results appears in Fig. 4-7(b). Of particular interest are the frequencies at which the response falls by 3 dB (to 0.707 of maximum). These are commonly termed the lower (f_1) and upper (f_2) cutoff frequencies or half-power points. The frequency span between them is called the *bandwidth* and is approximated by

$$\mathrm{Bw} = \frac{f_r}{Q} \tag{4-13}$$

The approximation grows better as Q goes higher, being about 5% off when $Q = 5$. Bandwidth is easy to measure experimentally, and narrow bandwidth (high Q) is generally a desirable feature in a tuned circuit. At frequencies beyond f_1 and f_2, the response can be determined from the following table:

N (*bandwidths*)	*Response* (%)
0	100
0.5	89.8
1	70.7
2	47.9
3	35.3
4	28.0
5	23.3
10	13.2
25	4.2
50	2.0
100	1.0

$$f_1 = \frac{f_r}{N(\mathrm{Bw}/2) + 1}$$

$$f_2 = f_r\left(N\frac{\mathrm{Bw}}{2} + 1\right)$$

Voltage Rise at Resonance: Notice that in the circuit of Fig. 4-7(a) the total impedance seen by the source at resonance is 125.6 Ω purely resistive, and the current is $V/R = 126.5\ \mathrm{V}/126.5\ \Omega = 1\ \mathrm{A}$. This current flows through X_C and X_L, producing voltages across each of them, even though their reactances cancel. The voltages produced are

$$V_{XL} = V_{XC} = IX_C = 1\ A \times 628\ \Omega = \mathbf{628\ V}$$

These voltages are larger than the source voltage by exactly a factor of Q (five times in this case). In general,

$$V_{XL} = V_{XC} = QV_G \tag{4-14}$$

for a series-resonant circuit, where V_G is the open-circuit generator voltage and Q is calculated using the entire circuit resistance. Notice that with $Q = 50$ and $V_G = 100$ V, the voltage rating required of the capacitor becomes 5000 V. This is no idle theory! The voltage rise at resonance must be considered when selecting components for series-resonant circuits.

Of course, since V_{XL} leads the circuit current by 90° and V_{XC} lags by 90°, the two voltages are 180° out of phase and exactly cancel each other. Thus the voltages across the circuit elements R_G, C, L, R_c, and R_L in Fig. 4-7(a) add up to 125.6 V, which is the generator voltage.

Parallel Resonance: In the circuit of Fig. 4-3(c), equal capacitive and inductive reactances were placed in parallel, and the resulting reactance rose to infinity. This was because the two currents ($I_L = V/X_L$ and $I_C = V/X_C$) were equal and exactly 180° out of phase, resulting in zero total current. Figure 4-8(a) shows a real parallel-resonant circuit with the inevitable series resistance of the inductor. Equations 4-9 and 4-10 can be used to change X_{Ls} and R_s to their parallel equivalent, but if Q is reasonably high, we can operate with the much simpler approximations:

$$X_{Lp} = X_{Ls} \tag{4-15}$$

$$R_p = \frac{X^2}{R_s} \tag{4-16}$$

The error incurred with $Q_{\text{coil}} = 5$ is 4%, dropping to 1% at $Q = 10$.

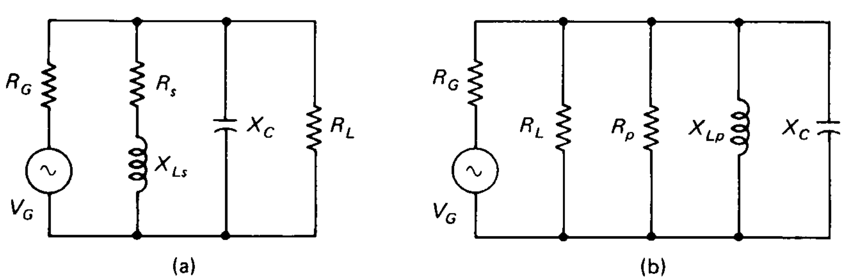


***FIGURE 4-8** Parallel tuned circuit: (a) as found in practice with generator, load, and series coil resistances; (b) equivalent with all resistances appearing in parallel.* $Q_p = R_p(\text{tot})/X_p$.

The Q of the entire circuit can now be determined from the equivalent circuit of Fig. 4-8(b) with R_G, R_L, and R_p appearing in parallel:

$$Q_p = \frac{R_{p(\text{tot})}}{X_{Lp}} = \frac{R_{p(\text{tot})}}{X_C} \tag{4-17}$$

The maximum output voltage at resonance is

$$V_{RL} = V_G \frac{R_p \| R_L}{R_G + R_p \| R_L} \tag{4-18}$$

The curve of Fig. 4-7(b) can be used to determine the output at other frequencies.

Notice that on high-impedance lines (R_G and R_L above 100 Ω or so) the highest Q is obtained with a parallel-resonant circuit across the line, while low-impedance lines are more sharply tuned with a series-tuned circuit in series with the line. Also notice that with the parallel circuit, a low L/C ratio (small inductance, large capacitance) will produce lower reactances at resonance, and hence higher Q, according to equation 4-17, while the series circuit will tune more sharply with a high L/C ratio, according to equation 4-12. Some of the advantage of increasing L in the series circuit is offset by the inevitable increase in R_s of the coil when its number of turns increases. However, L increases by the square of the turns, while R_s increases linearly, so substantial benefits may be realized.

Resonant Traps: Figure 4-9(a) and (b) show, respectively, a series-resonant trap across the line and a parallel-resonant trap in the line, together with formulas for their Q and minimum output at the null point. Again, these formulas are approximations whose accuracy falls off markedly below a Q of 5. It is also assumed that R_s is much less than either R_G or R_L. Notice that the depth of the null is completely dependent upon R_s, whereas the sharpness of the null is dependent upon the ratio of reactance at resonance to line impedance. As with the peaking tuned circuits, series-resonant traps tune most sharply with a high L/C ratio on a low-impedance line, whereas parallel-resonant traps have highest Q with a low L/C ratio on a high-impedance line.

EXAMPLE 4-2

Design a tuning circuit with a -3 dB passband from 110 to 140 kHz. The source resistance is 40 kΩ and the load impedance is essentially infinite. A 2.5-mH coil with a Q of 8.0 at 125 kHz is available.

Solution

Solving equation 4-11 for C gives the required capacitance:

$$C = \frac{1}{4\pi^2 f^2 L} = \frac{1}{4\pi^2 \times (125 \times 10^3)^2 \times 2.5 \times 10^{-3}} = \mathbf{648\ pF}$$

The Q required if the parallel-resonant circuit of Fig. 4-8(a) is used is obtained from equation 4-13:

$$Q = \frac{f_r}{\text{Bw}} = \frac{125\ \text{kHz}}{30\ \text{kHz}} = 4.17$$

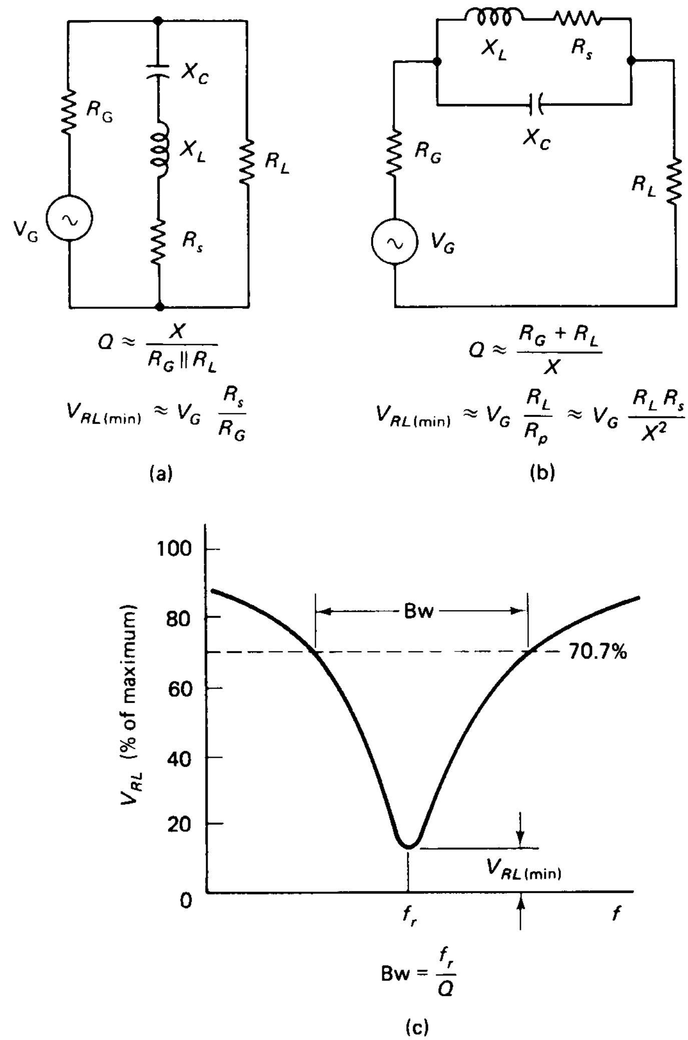


***FIGURE 4-9** Resonant traps: (a) series tuned across the line; (b) parallel tuned in series with the line; (c) response curve showing bandwidth and null depth.*

The parallel resistance needed to produce this is obtained by equation 4-17:

$$X_L = 2\pi fL = 1960\ \Omega$$

$$R_p = QX = 4.17 \times 1960 = 8.17\ k\Omega$$

The coil's series resistance and its parallel equivalent, by equations 4-12 and 4-16, are

$$R_s = \frac{X}{Q} = \frac{1960}{8.0} = 245\ \Omega$$

$$R_P = \frac{X^2}{R_s} = 15.68\ k\Omega$$

The effective parallel resistance of R_G and R_p is 40 kΩ||15.68 kΩ, or 11.26 kΩ. A shunt resistance, sufficient to bring the parallel total down to the necessary 8.17 kΩ, must now be added in the R_L position of Fig. 4-8(a). The value is calculated by the product over the difference:

$$R_L = \frac{R_E R_T}{R_E - R_T} = \frac{11.26 \times 8.17}{11.26 - 8.17}\ \text{k}\Omega = 29\ \text{k}\Omega$$

The practice of shunting a tuned circuit to lower its Q and widen the bandwidth is in common use.

4.4 REAL INDUCTORS

For the sake of discussion and analysis we will place inductors into four categories, all of which are shown in Fig. 4-10.

Air-Core Solenoid-Wound Coils: These may commonly range in value from 0.1 μH to several mH, having one to several hundred turns. They may be self-supporting in air, or wound on a ceramic, plastic, or cardboard form. Their diameter depends upon inductance required and power level, but commonly ranges from 0.5 to 5 cm, with still larger sizes appearing in high-power transmitters. When used in frequency-sensitive tuning and filtering circuits, they are called coils, and their exact inductance and its stability are likely to be critical. When used simply to block high frequencies they are called *rf chokes*, and wider tolerances and lower Q's are generally acceptable.

Coils with Open-Path Magnetic Cores: These are wound on forms commonly between 0.5 and 2 cm in diameter, and have powdered iron or metal oxides (ferrites) in the core to increase the inductance. The core may be an adjustable slug on screw threads, providing variable inductance. Their values range to a few hundred millihenrys.

Coils with Closed-Path Magnetic Cores: These are wound on a form consisting of strips (laminations) of iron, as shown in Fig. 4-14, and are often enclosed (potted) in a steel cover. They contain several hundred to several thousand turns of wire, and common values range from a few millihenrys to several hundred henrys. Common sizes are from 2 to 10 cm on a side, although much larger units are available for high-power applications. They are generally used to block the flow of relatively low-frequency ac (either audio or line-frequency ripple in a power supply) while passing dc. Tolerance is generally rather broad ($-20, +50\%$ typical) and Q is relatively low. They can be designed with an adjustable air gap in the core to provide variable inductance.

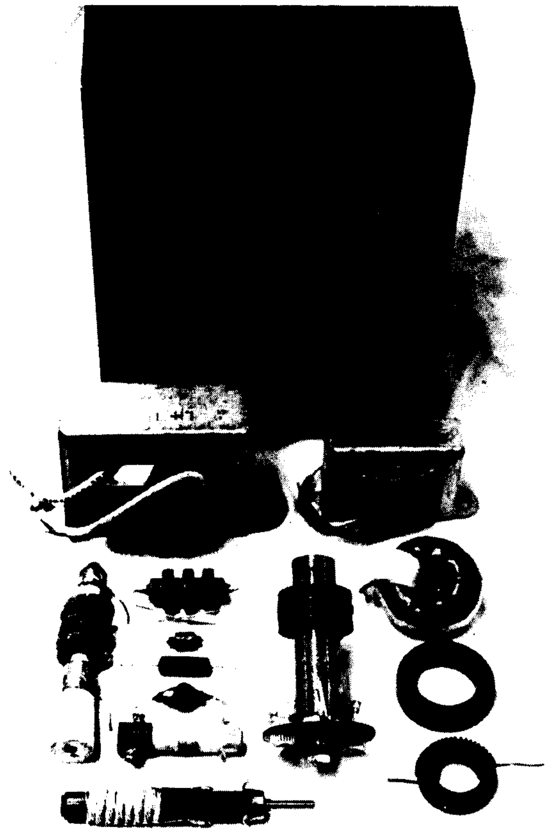


FIGURE 4-10 Representative inductors. Top: *large potted power-supply choke with two smaller lower-current 0.5-H open-frame units below, all wound on E-I laminated iron cores;* left: *2.5-mH and 1-mH air core RF chokes, with small 1.8-mH ferrite-core choke below, followed by 6.8-μH and 100-μH air-core chokes, and two ferrite slug-tuned coils;* center: *ferrite-slug-adjustable inductor of 20 mH;* right: *three toroids, 70 mH coated, blank core, and 1.8 mH wound of AWG 18 wire.*

Toroidal Inductors: Toroids are wound on donut-shaped cores of magnetic materials specially formulated for temperature stability, low loss, and high permeability. Nonmagnetic cores may be used for microhenry values. Because the turns cannot be wound on a simple rotating drum or bobbin, toroids are considerably more expensive than af chokes of equal inductance. However, by proper selection of core material and winding, they can achieve a remarkably high Q, often above 100 in the frequency range 1 kHz to 1 MHz. Their tolerance, linearity, temperature stability, and Q are generally far superior to inductors wound on laminated iron cores.

4.5 AIR-CORE COIL CALCULATIONS

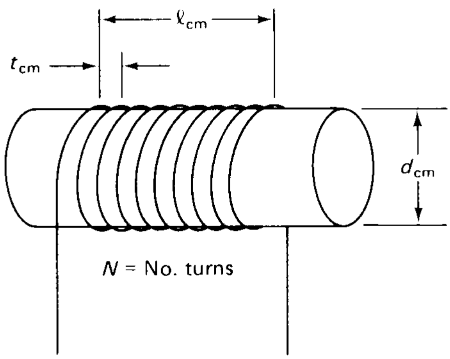
Predicting Inductance: The inductance of an air-core coil of known dimensions can be predicted to within about 10% by the formula

$$L_{\mu\text{H}} = \frac{d^2N^2}{46d + 101l} \qquad [\text{cm}] \tag{4-19}$$

where L is in microheneries, d is diameter in centimeters, l is length in centimeters, and N is the number of turns. The equivalent with d and l in inches is

$$L_{\mu\text{H}} = \frac{d^2N^2}{18d + 40l} \qquad [\text{in.}] \tag{4-20}$$

Figure 4-11 illustrates the application of the formula. The formula is most accurate for single layer coils of 10 turns or more, but is still useful for smaller or multilayer coils. For iron-or ferrite-core coils, be sure to see Section 4.6 on core effects.



$$L_{\mu\text{H}} = \frac{d^2N^2}{46d + 101\ell}$$

$$N = \frac{\sqrt{(50.5\,L_{\mu\text{H}}\,t)^2 + 46\,d^3\,L} + 50.5\,Lt}{d^2}$$

***FIGURE 4-11** The inductance L or turns required N for an air-core coil diameter d and length l. Turn spacing center to center is designated t.*

Usually, the question is not "calculate the inductance of a given coil," but rather "wind me a coil of a given value." Solving the foregoing equations for N with tN substituted for l, gives the somewhat unwieldy turns-winding formula for single-layer close-wound coils:

$$N = \frac{\sqrt{(50.5Lt)^2 + 46d^3L} + 50.5Lt}{d^2} \qquad [\text{cm}] \tag{4-21}$$

where N is the number of turns required, L is the desired inductance *in microhenries*, d is coil diameter in centimeters, and t is the wire thickness (including insulation) in centimeters. The same formula, with d and t in inches, becomes

$$N = \frac{\sqrt{(20Lt)^2 + 18d^3L} + 20Lt}{d^2} \qquad [\text{in.}] \tag{4-22}$$

For multilayer coils wound on a bobbin of a given length, the required number of turns is given by direct transposition of equations 4-19 and 4-20:

$$N = \frac{\sqrt{L(46d + 101l)}}{d} \qquad [\text{cm}] \tag{4-23}$$

$$N = \frac{\sqrt{L(18d + 40l)}}{d} \qquad [\text{in.}] \tag{4-24}$$

Stray Capacitance and Resonance: Adjacent turns of wire in a coil have a certain amount of capacitance between them, the total effect being very much as if a single capacitor were placed across an ideal inductor, as in Fig. 4-12(a). This, of course, gives the coil a *natural parallel-resonant frequency*, above which its reactance is capacitive and decreases with frequency. Using a coil at one-fifth of its resonant frequency will result in an apparent inductance increase of 4%. At one-third the resonant frequency the increase is 11%, and at $\frac{1}{2} f_r$ it is 25%. The coil is still useful as a choke up to and beyond parallel resonance, however, since its impedance is still very high.

As frequency is increased above the parallel-resonant point, the impedance (capacitive) of the coil will drop until a point is reached at which the stray capacitance from the bulk of the coil becomes series-resonant with the first turn or so on the end of the coil. Impedance at this *natural series-resonant frequency* drops to near zero as shown in Fig. 4-12(b). There may be several more series-resonant points above the first one, but the coil is useless even as a choke above the first one. When a wide range of frequencies is to be blocked, two chokes in series may be used. The first, placed at the "hot" end, is small and has a very high series-resonant point. The second, placed at the "cold" or ac ground end, is large and has its series-resonant point at a frequency where the reactance of the first choke is high.

A grid-dip meter coupled to the open-circuited coil will detect the parallel-resonant point. With the terminals shorted, the dip meter will respond to the series-resonant points. The shorting wire may have to be looped once around the dip meter's coil to provide adequate coupling. Figure 4-12(c) shows four techniques for reducing stray capacitance and hence raising the natural resonant frequency of a coil. The first is simply to keep the turns spaced apart, since capacitance varies inversely with conductor spacing. Of course, this spacing also decreases the inductance by increasing length (see equation 4-19), but if the spacing is limited to one-half the wire diameter or less, significant improvement can be realized. Spacing

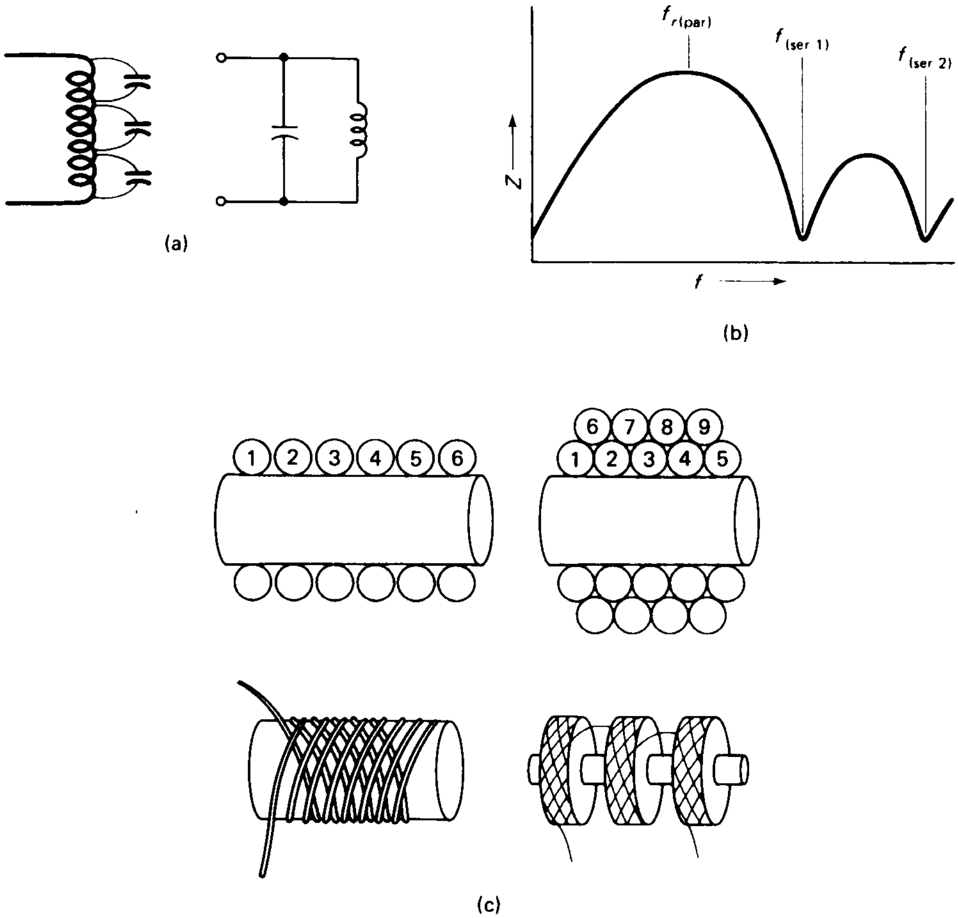


***FIGURE 4-12** (a) Stray capacitance between turns gives a coil a natural parallel-resonant frequency. (b) Impedance of a coil versus frequency showing parallel-resonant and (higher) series-resonant frequencies. (c) Four methods for reducing stray capacitance in a coil (see the text).*

can be achieved by using specially grooved forms, by winding a string along with the wire to hold the space, or by simply using Teflon-insulated wire in place of the usual thin enamel insulation. Common PVC (polyvinyl chloride) insulated hookup wire should never be used for winding coils. PVC has severe dielectric loss at certain frequencies which are a function of temperature, and serious reduction in the Q of the coil will result. Bakelite coil forms should be avoided at radio frequencies for similar reasons.

The other techniques for raising f_r illustrated in Fig. 4-12(c) are (2) keep the starting and finishing coils of a two-layer coil at opposite ends of the form, (3) use a diagonal-pitched winding, alternating direction with each layer so the wires are close only at the crossing points, and (4) wind the coil in separate sections (called pies), and build up the required inductance by series-connecting the required number of these low-value high-f_r coils. This last technique has the drawback of requiring more wire than a single bulk coil because there is no mutual inductance between windings of different sections. The result is higher winding resistance and lower Q.

Estimating the stray capacitance of a coil is difficult because it depends upon so many factors: insulation thickness and dielectric constant, coil spacing, layering, winding pattern, wire thickness, and coil diameter primary among them. However, for single-layer close-wound coils using No. 32 to No. 18 enameled wire on 1- to 3-cm forms, capacitance will generally be in the vicinity of 0.05 pF/turn. Multi-layered coils may have less than half that capacitance. Spaced winding and cross-pitch winding can reduce the capacitance by about 25%.

Skin Effects and Q: In the absence of nearby objects, such as shields or other components, the Q of an air-core coil depends almost entirely on the resistance of the wire at the frequency in question. The formulas for skin-effect resistance of copper wire are therefore reproduced from Chapter 1:

$$\delta = \frac{6.6}{\sqrt{f}} \qquad [\text{cm}] \tag{1-3}$$

$$R = 8.3 \times 10^{-8} \frac{\sqrt{f}}{t} l \tag{1-4}$$

where δ is skin depth in centimeters, t is wire diameter (thickness) and ℓ is wire length, both in the same units, and f is in hertz. In winding coils for high frequencies it is necessary to consider the proximity effect as explained in Section 1-4.

Some examples from the laboratory will serve to demonstrate the utility of the inductance and skin-effect formulas.

EXAMPLE 4-3

The core was removed from an audio choke. The remaining air winding had an average diameter of 3.0 cm and a length of 2.2 cm, and was wound of wire with a thickness measured as 0.014 cm. The dc resistance of the coil was 460 Ω. Predict the coil's inductance and Q at 1000 Hz.

Solution

Interpolating from the wire table (Fig. 1-1), this wire has a resistance of about 0.011 Ω/cm. The resistance of one turn is then

$$R_1 = \pi d r = 3.14 \times 3 \times 0.011 = 0.104\ \Omega$$

The number of turns is then found from the total resistance:

$$N = \frac{R_T}{R_1} = \frac{460}{0.104} = 4420 \text{ turns}$$

The inductance, from equation 4-19, is

$$L = \frac{d^2N^2}{46d + 101l} = \frac{3^2 \times 4420^2}{(46 \times 3) + (101 \times 2.2)} = \mathbf{0.488\ H}$$

The skin depth at 1000 Hz is calculated from equation 1-3:

$$\delta = \frac{6.6}{\sqrt{f}} = \frac{6.6}{\sqrt{1000}} = 0.21 \text{ cm}$$

This is 30 times the radius of the wire, so skin effect is negligible and the dc resistance is used:

$$Q = \frac{X}{R} = \frac{2\pi f L}{R} = \frac{6.28 \times 1000 \times 0.488}{460} = \mathbf{6.67}$$

The inductance and Q of the actual coil, as measured on a commercial bridge at 1000 Hz, were **0.47 H** and **6.6**, respectively.

EXAMPLE 4-4

Find the inductance and Q at 3 MHz of a 30-turn close-wound coil of No. 22 enameled wire on a 1.9-cm-diameter plastic form.

Solution

The thickness of the wire, from the table in Fig. 1-1, is 0.064 cm. Allowing 0.002 cm for the enamel, the length occupied by 30 turns is calculated:

$$l = Nt = 30 \times 0.068 = 2.04 \text{ cm}$$

The inductance, from equation 4-19, is

$$L_{\mu H} = \frac{d^2N^2}{46d + 101l} = \frac{1.9^2 \times 30^2}{(46 \times 1.9) + (101 \times 2.04)} = \mathbf{11\ \mu H}$$

The skin depth, by equation 1-3, is

$$\delta = \frac{6.6}{\sqrt{f}} = \frac{6.6}{\sqrt{3 \times 10^6}} = 0.00038 \text{ cm}$$

This is 84 times less than the wire radius, so the skin-effect resistance is calculated by equation 1-4:

$$\text{Wire length} = l_w = \pi dN = 3.14 \times 1.9 \times 30 = 179 \text{ cm}$$

$$R = 8.3 \times 10^{-8} \frac{\sqrt{f}}{t} l = 8.3 \times 10^{-8} \frac{\sqrt{3 \times 10^6}}{0.064} \times 179 = 0.403\ \Omega$$

Checking the table in Section 1-4, we make a not-too-confident estimate of the proximity-effect resistance as ten times the calculated skin-effect resistance, or 4 Ω.

$$Q = \frac{X_L}{R} = \frac{2\pi f L}{R} = \frac{2\pi \times 3 \times 10^6 \times 11 \times 10^{-6}}{4} = \mathbf{52}$$

The measured values for L and Q of the actual coil at 3 MHz were **9.6 μH** and **41**, respectively.

The Q of a coil is not a constant, but varies with frequency in a generally predictable way, as shown in Fig. 4-13. At low frequencies, X_L is low, although the

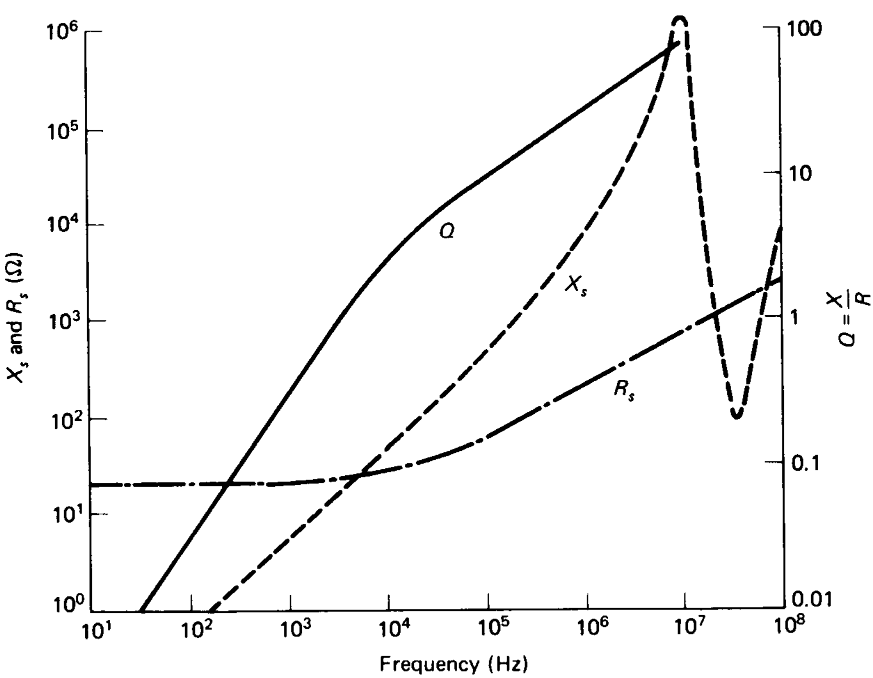


FIGURE 4-13 Actual reactance, resistance, and Q for a 1-mH air-core three-pie cross-pitch choke. R holds at its dc value until skin effect causes it to rise as $\sqrt{f}$. X increases as f until parallel resonance (9 MHz) brings it up and series resonance (35 MHz) brings it down. Q rises as f until skin effect begins, whereupon Q rises as $\sqrt{f}$ until parallel resonance, where it is maximum. Above f_r the 1-mH coil becomes swamped by the stray effects and is no longer accessible. Q readings above f_r therefore apply to the stray effects and not to the coil as such.

resistance of the wire is equal to its dc value. Q is therefore relatively low. Q increases linearly with frequency until the skin effect begins to raise the value of R. Since X_L goes up as f, while R goes up only as $\sqrt{f}$, Q now rises as $\sqrt{f}$:

$$Q = \frac{X}{R} \propto \frac{f}{\sqrt{f}} = \sqrt{f}$$

At the self-resonant frequency, Q is supposedly at a maximum, since the reactance of X_L and X_C in parallel approaches infinity. However, the coil is not purely inductive at this point. The maximum usable Q should generally be read at $\frac{1}{2}$ to $\frac{1}{5}$ the frequency of the peak at f_r.

4.6 MAGNETIC-CORE INDUCTORS

Core Effect on Inductance: A magnetic field can pass much more readily through a magnetic metal such as iron or nickel than through air. This fact is demonstrated by the common trick of holding up a string of a dozen straight pins from a single magnet at the top. We say that the magnetic metals have a higher *permeability* than air. To the extent that we can provide a high-permeability medium within and around a coil, we can increase its inductance for a given number of turns. Although the amount of increase depends upon the exact shape and composition of the core, we can estimate that a slug of magnetic material in the center of the coil will provide an inductance increase of seven times over an air core. Brass slugs are occasionally used in rf coils to provide an inductance *decrease* of up to 25%.

Calculating Core Effects: The highest inductances are obtained by providing a complete magnetic flux path through and around the coil. Figure 4-14 shows the common methods for achieving this. The CI- and EI-shaped stacks of silicon-iron strips insulated with varnish have long been popular and still prevail at power-line and audio frequencies, while the newer ferrite materials in the pot core and toroid ring structures generally offer superior performance in the ranges above a few kilohertz.

The inductance of a cored coil with no air gap can be calculated with reasonable accuracy from the formula

$$L_{\mu H} = 0.012 N^2 \mu \frac{A}{\ell} \tag{4-25}$$

where N is the number of turns, μ the relative permeability of the material, A the cross-sectional area of the core in cm^2, and ℓ the effective magnetic path length in centimeters. Note that in the E-I structure there are two parallel paths if the coil is wound on the center post, so the effective length is

$$WX + \frac{XZ + ZY + YW}{2}$$

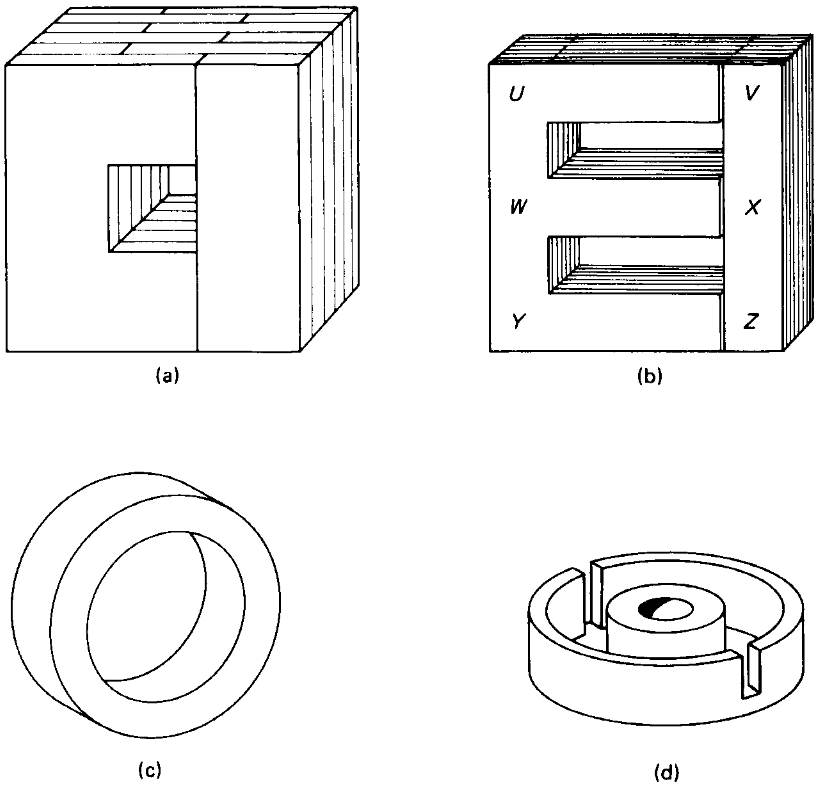


FIGURE 4-14 Popular magnetic-core configurations: (a) alternating layers of C-I-shaped silicon-iron strips; (b) the more popular E-I-shaped laminated core; (c) toroid core of ferrite material; (d) ferrite pot core (one of two identical halves); coil is wound on a plastic bobbin and slipped over center post.

For the pot-core structure, the effective length and area are generally given by the manufacturer.

Air-Gap Considerations: All of the structures in Fig. 4-14 except the toroid contain an air gap in the flux path. It may be less than 0.01 cm, due primarily to the insulating varnish coat on the laminations; or it may be intentionally made much larger, or made variable to allow adjusting inductance. The inductance of a coil having a magnetic core with an air gap can be calculated from

$$L_{\mu \mathrm{H}} = 0.012 N^2 \frac{A}{\ell_g + \ell_c/\mu} \tag{4-26}$$

where ℓ_g is the length of the gap in centimeters, ℓ_c the effective core length in centimeters, and the other terms are as above. Values of μ for various popular core materials are given in the table of Fig. 4-15.

Material Type	Permeability μ_i (Low Level)	Permeability μ_m (High Level)	Saturation, B (kGauss)	Maximum Frequency, f (MHz)	Curie Temperature, T_{max} (°C)
Silicon iron 0.5-mm laminations	400	40,000	15	0.003	700
0.05-mm laminations	400	40,000	15	0.03	700
Iron/nickel alloy	3000	20,000	10	0.003	460
Powdered iron	125	127	3	0.1	–
Ferrites: 3B7, 3C8, TI, W-03	2300	1900	4	0.3	170
3E2A	5000	1800	3	0.2	170
3E3	12,500	1900	4	0.1	125
3D3	750	1500	3	2.5	150
4C4	125	600	2	20	300
1Z2	15	–	–	1000	–

FIGURE 4-15 *Properties of common magnetic-core materials.*

Besides lowering the inductance, the presence of an air gap has the important effect of *linearizing* the core. Notice from Fig. 4-15 that for most materials μ varies markedly from low- to high-level magnetization. This is further illustrated in the magnetization curve of Fig. 4-16(a). The air gap adds a constant and relatively large *reluctance* (the magnetic equivalent of resistance) to the magnetic path, tending to swamp out the reluctance changes of the core material. Nevertheless, it

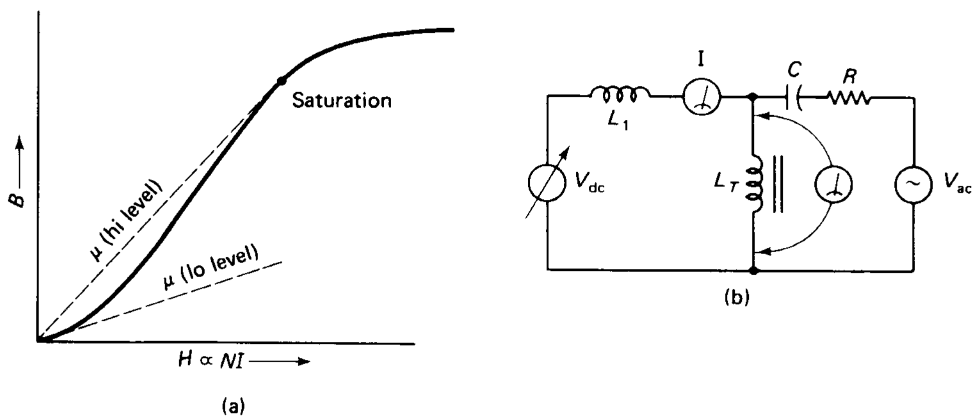


FIGURE 4-16 *(a) Magnetic flux density B versus magnetizing force H for a typical magnetic material. Initial permeability μ_i is generally lower than maximum-level permeability μ_m. (b) Circuit for measuring the dc saturation current of an inductor.*

is common to find that an iron-core inductor measures 1 H at low signal levels and 2 H at higher levels. Waveform distortion can also be expected due to the nonlinear response of magnetic-core inductors.

Core Saturation: There is a limit to the degree of magnetization (molecular alignment) that a material can obtain, and this places a limit on the peak current that can pass through a magnetic-core coil without causing *saturation*. If this limit is exceeded with ac, the coil suddenly stops generating inductive reactance to the ac and becomes (almost) a simple piece of wire. The result could be excessive current, overheating, and the destruction of the coil.

The slope of the magnetization curve, Fig. 4-16(a), equals permeability, which is proportional to inductance. Notice that this slope drops to near zero as current I is increased. The value of flux density B in the core of an ac-excited coil can be calculated from

$$B = \frac{V_{\mathrm{rms}}}{4.4 \times 10^{-8} fNA} \tag{4-27}$$

where B is in gauss, V is in volts, f is in hertz, N is turns, and A is the cross-sectional area in cm^2. Notice that μ and ℓ do not enter this equation, since their effect on flux density via reluctance is exactly offset by a reciprocal effect on current via inductance. The value of B must be kept less than the saturation value given in Fig. 4-15.

The flux density for a dc-excited coil is

$$B = \frac{1.26NI}{\ell_g + \ell_c/\mu} \tag{4-28}$$

where B is in gauss, N is turns, I is dc amperes, ℓ_g is air-gap length in centimeters, ℓ_c is effective core length, and μ is core permeability.

EXAMPLE 4-5

An inductor is wound on a C-I-shaped core of silicon-iron laminations. The legs of the lamination stack are 2 cm wide by 2 cm thick, and the overall dimensions of the core are 6 cm × 6 cm. An air gap of 0.05 cm is left in the core. The winding consists of 900 turns of AWG 24 enameled wire. Calculate the inductance at high and low excitation levels, and determine the maximum dc current before saturation.

Solution

Low-level μ is 400, high-level μ is 40,000, ℓ_c is 4 × 4 or 16 cm, and A is 4 cm^2. Using equation 4-26, we obtain

$$L_{\mathrm{lo}} = 0.012N^2 \frac{A}{\ell g + \ell_c/\mu} \mu\mathrm{H} = 0.012 \times 900^2 \frac{4}{0.05 + 16/400} \mu\mathrm{H} = \mathbf{0.43\ H}$$

$$L_{\mathrm{hi}} = 0.012 \times 900^2 \frac{4}{0.05 + 16/40{,}000} \mu\mathrm{H} = \mathbf{0.77\ H}$$

Transposing equation 4-28 to find current for the 15,000-G saturation level yields

$$I = \frac{B(\ell_g + \ell_c/\mu)}{1.26N} = \frac{15{,}000(0.05 + 16/40{,}000)}{1.26 \times 900} = \mathbf{0.67\ A}$$

Measuring Saturation: Figure 4-16(b) shows an experimental circuit for measuring the dc saturation current of a choke. L_T is the inductor under test, and the saturation current is measured on dc ammeter I. R must have a value of at least $3X_{LT}$ and X_C must be $\frac{1}{3}X_{LT}$ or less. L_1 must have a higher current rating than L_T and X_{L1} must be $3X_{LT}$ or more. L_1 can be replaced by a resistor if these conditions are met. In operation, V_{ac} is set to produce a small voltage, say 1 V rms, across L_T. V_{dc} is then increased until the measured ac voltage drops by a specified amount (usually 10%), indicating that X_{LT} has dropped by 10%.

Useful Saturation: Core saturation can be used to advantage in some applications. A *swinging choke* in a power supply is designed to saturate at maximum dc load current. The resulting decrease in inductance allows more charging current to reach the first filter capacitor, thus largely offsetting the output voltage drop under heavy load.

The *magnetic amplifier* uses a relatively small dc current through a many-turn coil to saturate a core upon which is wound an inductor carrying a heavy ac current. A small increase in the dc can thus cause a large increase in the ac.

Core Effect on Q: The type of core material used in an inductor must be selected for the frequency range of intended use. Thin strips (laminations) of silicon iron are used at power-line and audio frequencies, with powdered iron, and then various compositions of metal oxides (ferrites) being preferred at progressively higher frequencies. The reason for the frequency dependence of the core is this: In magnetizing the core with ac, we are requiring the molecules of the core material to realign their positions every half-cycle. It takes a rather specially composed material to follow these realignments into the high kilohertz and megahertz ranges, and there is always some frictional loss (due to hysteresis) in the process. The Q of a magnetically cored coil is therefore likely to be higher than its air-core counterpart because of its higher X_L, but only up to the maximum frequency of the core material, beyond which Q drops sharply due to hysteresis loss.

Predicting Q for a magnetic-core coil is quite difficult, since permeability and hysteresis vary with frequency and material composition. Measurement at the intended frequency is the most practical recourse.

5

PROPERTIES OF REAL TRANSFORMERS

5.1 THE IDEAL TRANSFORMER

The Basic Idea: If two strands of wire are used to wind two coils simultaneously on the same form, a *transformer* is produced. This is illustrated in Fig. 5-1(a). If an ac voltage is impressed across one of these windings, which is then called the *primary*, an ac current will flow in the amount $I_P = V_G/X_P$, where X_P is the reactance of the primary coil at the generator frequency. Since the reactance is inductive, the current, and hence the strength of the magnetic field produced, will lag the applied voltage by 90°, as shown in Fig. 5-1(b). The other winding, called the *secondary*, experiences fully the effects of this changing magnetic field, and a voltage is induced in it which is proportional to the rate of change of that field. For a sine-wave input, this rate of change turns out to be another sine wave, which is exactly in phase with the input, as shown in Fig. 5-1(b). If the number of turns on the primary exactly equals the number of turns on the secondary ($N_P = N_S$), and if the windings are kept perfectly close together so that all of the magnetic field produced by the primary reaches the secondary, the secondary voltage will be an exact reproduction of the primary voltage.

Isolation: You may ask: "What good is reproducing the same voltage?" The answer is: *isolation*. Notice in Fig. 5-1(a) that the secondary circuit is not actually connected to the primary circuit or to chassis ground. The entire V_G source may be

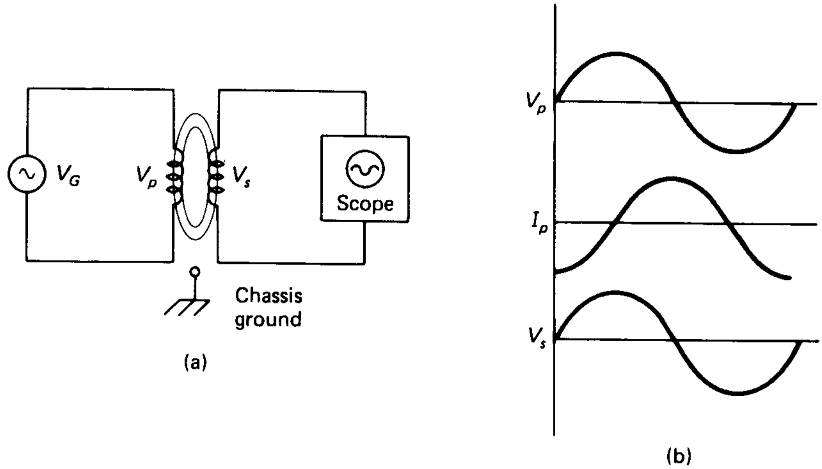


FIGURE 5-1 *Basic transformer: (a) the primary creates a changing magnetic field which induces a voltage in the secondary; (b) the secondary voltage is proportional to the rate of change of primary current, placing it in phase with the input sine wave.*

held 1000 V above ground by another supply, but the secondary reproduction of V_G can be safely tied one side to ground. This use of an *isolation transformer* to allow ac signals to be shifted up or down with respect to their original ground reference is quite common.

Voltage Transformation: If the primary winding of an ideal transformer contains 1000 turns and the secondary contains 2000 turns, the voltage induced in the secondary will be twice the applied primary voltage. In general,

$$\frac{V_S}{V_P} = \frac{N_S}{N_P} = n \tag{5-1}$$

where n is called the turns ratio. Now, if a load resistance is connected across the secondary, as in Fig. 5-2(a), a secondary current $I_S = V_S/R_L$ will flow and a power $P = I_S V_S$ will be delivered to the load. This power must first be delivered to the

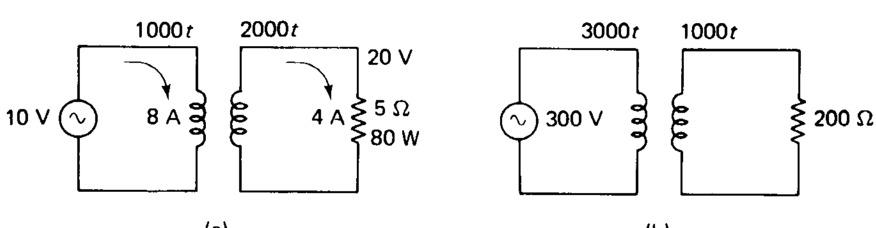


FIGURE 5-2 *A transformer can step up voltage at the expense of current (a), or step down voltage to get greater current (b), but power in must equal power out.*

primary side of the transformer by the generator, and requires a current $I_P = P/V_P$, assuming an ideal transformer with no losses. Using the numbers in Fig. 5-2(a), the secondary voltage of 20 V causes 4 A through the 5-Ω load, resulting in 80 W of power. This requires 8 A of primary current. Thus the transformer has stepped voltage up by a factor of 2, but has stepped current down by a factor of 2. In general, for a transformer the current and voltage ratios are reciprocals:

$$n = \frac{N_S}{N_P} = \frac{V_S}{V_P} = \frac{I_P}{I_S} \tag{5-2}$$

Notice that we are using capital P and S subscripts to identify *primary* and *secondary*, respectively. We use lower case p and s to refer to *parallel* and *series*.

Of course, there must also be a magnetizing current flowing through the primary inductance, but in an ideal transformer it is assumed that this is negligible in the face of the resistive current I_P which is dealt with in this formula. Remember from Fig. 4-4 that a reactive current that is one-third or less of a resistive current is negligible because of phasor addition properties.

EXAMPLE 5-1

Find the primary current in the circuit of Fig. 5-2(b).

Solution

$$V_S = V_P \frac{N_S}{N_P} = 300 \frac{1000}{3000} = 100 \text{ V}$$

$$I_S = \frac{V_S}{R_L} = \frac{100}{200} = 0.5 \text{ A}$$

$$I_P = I_S \frac{N_S}{N_P} = 0.5 \frac{1000}{3000} = \mathbf{0.167\ A}$$

Reflected Resistance: If the inductive primary current is neglected, the source in Example 5-1 appears to be driving a resistance $R_{ref} = V_P/I_P = 300/0.167 = 1800\ \Omega$. This is exactly 3^2 (or 9) times R_L. In general, the apparent or reflected resistance at the primary is

$$R_{P(\text{ref})} = \left(\frac{N_P}{N_S}\right)^2 R_L = \frac{R_L}{n^2} \tag{5-3}$$

We say that *the impedance ratio varies as the square of the turns ratio.*

Looking at it from the other side, the load seems to be driven through a resistance which is the actual generator resistance multiplied by the square of the turns ratio:

$$R_{S(\mathrm{ref})} = R_G n^2 \qquad (5\text{-}3a)$$

The reflected generator voltage is multiplied by the turns ratio n only. Figure 5-3 illustrates the reflected resistance as seen from the primary and secondary sides.

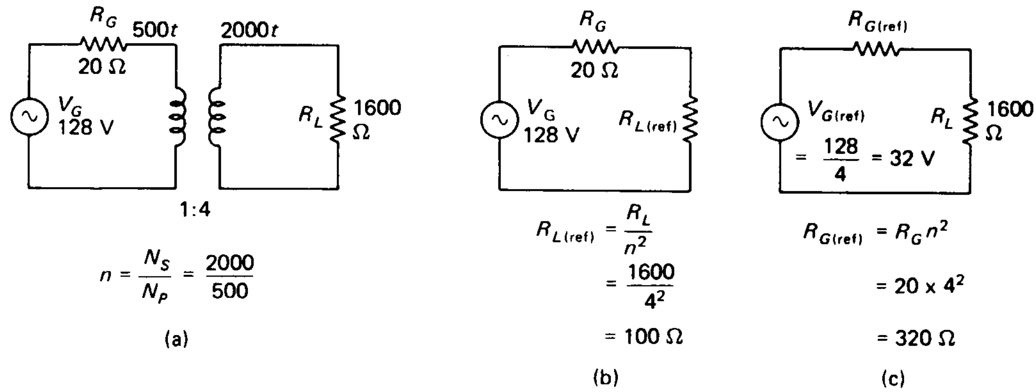


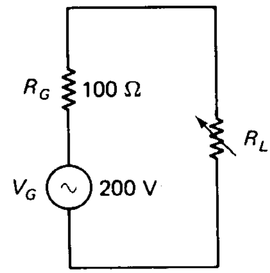
***FIGURE 5-3** Resistance is reflected across a transformer by the square of the turns ratio: (a) original circuit; (b) load resistance referenced to primary; (c) source voltage and resistance referenced to secondary.*

Impedance Matching: Signal sources invariably have some source resistance R_G through which they must push current to a load. This is illustrated in Fig. 5-4(a). If the load resistance is made very low, the load current will be maximized, but the power in the load will be low because the voltage across it will be low and $P = IV$. If the load resistance is made high, the voltage across it will be maximized but the current will be low, again resulting in low power in the load. For any fixed R_G, *the greatest power is delivered to the load when* $R_L = R_G$. We then say that the source and load impedances are matched. Figure 5-4(b) graphs the effects of matching and mismatching.

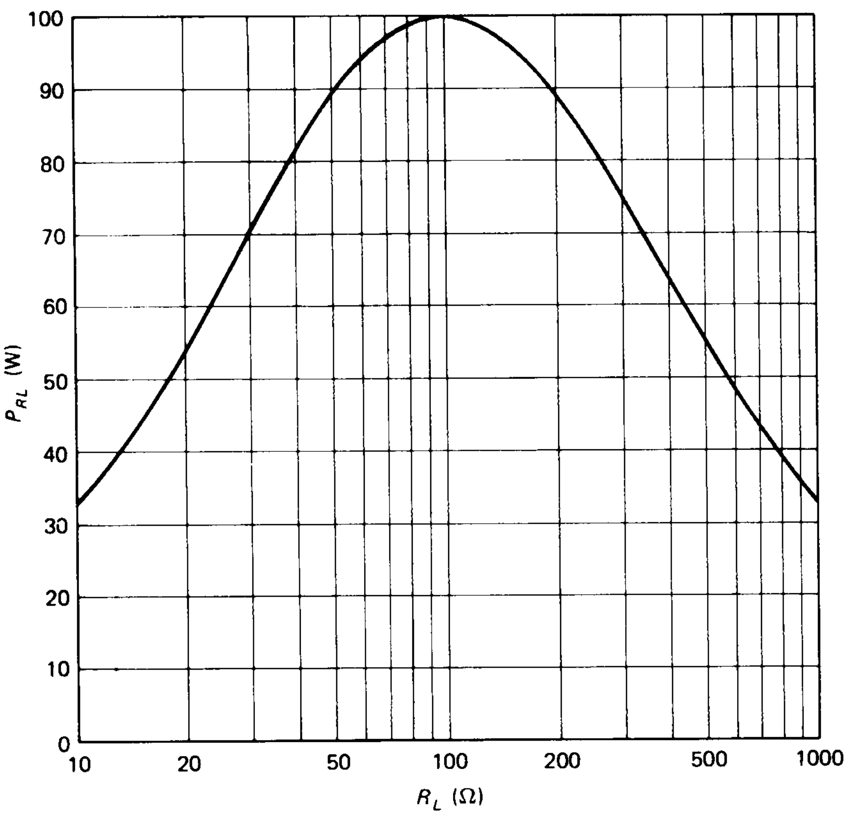
Transformers are often used to deliver maximum power to a load when the available generator resistance differs markedly from the available load resistance. The following example will illustrate the method.

EXAMPLE 5-2

An experimenter wishes to drive an 8-Ω speaker with a 600-Ω signal generator in connection with an audio test. The maximum open-circuit output is 40 V p-p. Determine the power delivered with a direct connection, and then find the transformer ratio for maximum power transfer and calculate the new power delivered.



(a)



(b)

FIGURE 5-4 *Impedance matching: (a) circuit with source resistance and variable load; (b) power in the load as a function of load resistance.*

Solution

Without the transformer, as in Fig. 5-5(a):

$V_{rms} = \dfrac{V_{PP}}{2 \times \sqrt{2}}$

$$I_L = \frac{V_G}{R_T} = \frac{40 \text{ V p-p}}{2.82 \times 608\ \Omega} = 0.0233 \text{ A}$$

$$P_L = I^2R = 0.0233^2 \times 8 = \mathbf{4.3\ mW}$$

With the transformer, as in Fig. 5-5(b), an impedance ratio of 600:8 must be matched:

$$\left(\frac{N_P}{N_S}\right)^2 = \frac{R_G}{R_L} = \frac{600}{8} = 75$$

$$\frac{N_P}{N_S} = \sqrt{75} = 8.66$$

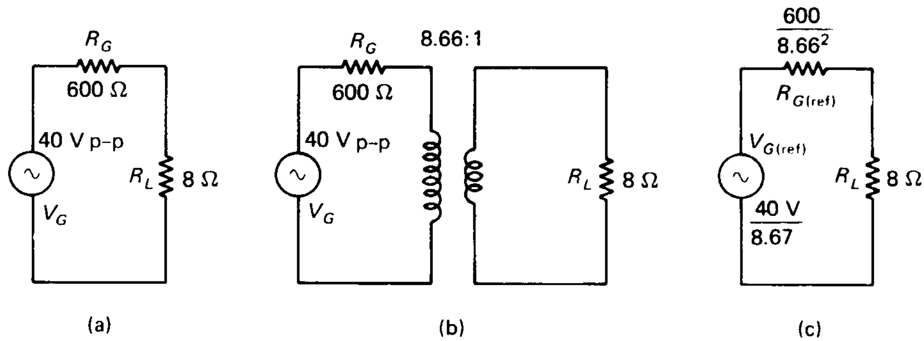


FIGURE 5-5 *Impedance matching to maximize power to a load: (a) mismatched circuit; (b) transformer provides matching; (c) equivalent circuit referenced to secondary.*

The transformer should have an 8.66-to-1 step-down ratio. Note from Fig. 5-4(b), however, that this is not an especially critical number. Impedance ratio differences by a factor of 2 up or down (turns-ratio differences by $\sqrt{2}$ or 1.41 up or down) will cause only an 11% loss of load power. Assuming an ideal transformer with an 8.66:1 primary-to-secondary turns ratio, the equivalent circuit referenced to the load appears in Fig. 5-5(c).

$$V_{\text{ref}} = V_P\frac{N_S}{N_P} = \frac{40 \text{ V p-p}}{8.66 \times 2.82} = 1.64 \text{ V rms}$$

$$R_{\text{ref}} = R_G\left(\frac{N_S}{N_P}\right)^2 = \frac{600}{8.66^2} = 8.0\ \Omega$$

$$I_L = \frac{V_{\text{ref}}}{R_T} = \frac{1.64 \text{ V}}{16\ \Omega} = 0.102 \text{ A}$$

$$P_L = I^2R_L = 0.102^2 \times 8 = \mathbf{83\ mW}$$

The output power is increased 84/4.3 = **19.5 times** by the transformer impedance match.

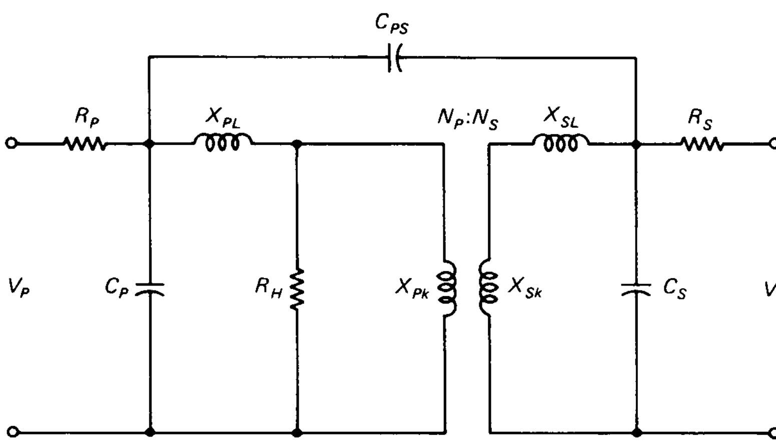
5.2 REAL TRANSFORMERS

A complete equivalent circuit for a real transformer and a photograph of some representative transformer types appear in Fig. 5-6. Fortunately, we do not have to deal with all the components shown in any single application of any single transformer. The next four sections are devoted to sorting out which parts are essential and which can be neglected, case by case.

X_{Pk} and X_{Sk} are the magnetically *coupled* parts of the primary and secondary inductive reactances at the operating frequency. These parts of the real transformer behave just like the ideal transformers of Figs. 5-1 to 5-3. X_{PL} and X_{SL} are the remaining parts of the primary and secondary reactances, termed *leakage* reactances. Leakage reactance results from the inability to wind the primary and secondary coils perfectly close together, or from the specific intention to keep them from being too close. Thus not absolutely all of the magnetic field produced by the primary is intercepted by the secondary. We say that the *coefficient of coupling k* is less than unity. The actual value is given as $k = X_{Pk}/(X_{Pk} + X_{PL})$. If leakage reactance becomes larger than the reflected load impedance across X_{Pk}, the input voltage V_P will be prevented from reaching coupling reactance X_{Pk}.

R_P and R_S are the winding resistances of the primary and secondary, respectively, and may be governed by skin effect if the wire size is large or the frequency is high. R_H represents losses due to hysteresis and induced currents in the core material.

C_P and C_S represent the effective stray capacitance across the primary and secondary windings, while C_{PS} represents the stray capacitive coupling between the two windings.



(a)

***FIGURE 5-6** (a) Real transformer equivalent circuit.*

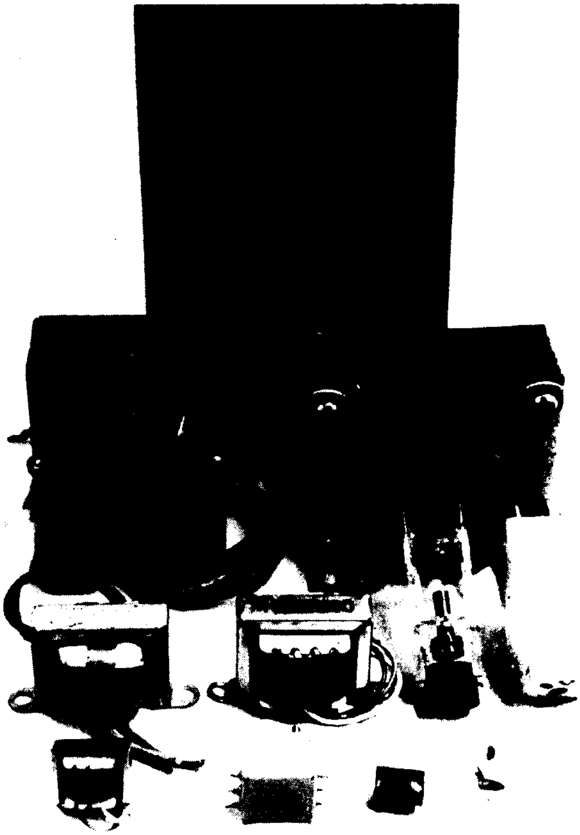


FIGURE 5-6 (b) Some representative types of transformers: potted, bell-end, and open-frame power transformers (rear), 2-W power and 0.5-W audio transformers (center left), miniature audio (foreground left), two pulse transformers (foreground center), and three 455-kHz parallel-tuned transformers, one with the shield removed (foreground right).

5.3 POWER TRANSFORMERS

The following characteristics are assumed for power transformers in this discussion:

- Operation at a relatively low fixed frequency (50 or 60 Hz for land-based, 400 Hz for airborne equipment). C_P, C_S, and C_{PS} in Fig. 5-6(a) are therefore negligible.
- Coefficient of coupling k equals unity, so X_{PL} and X_{SL} are negligible. (Typical k values for power transformers actually center around 99.8%.)
- Core losses small enough that R_H is negligible, except for determining the primary current with no secondary load.

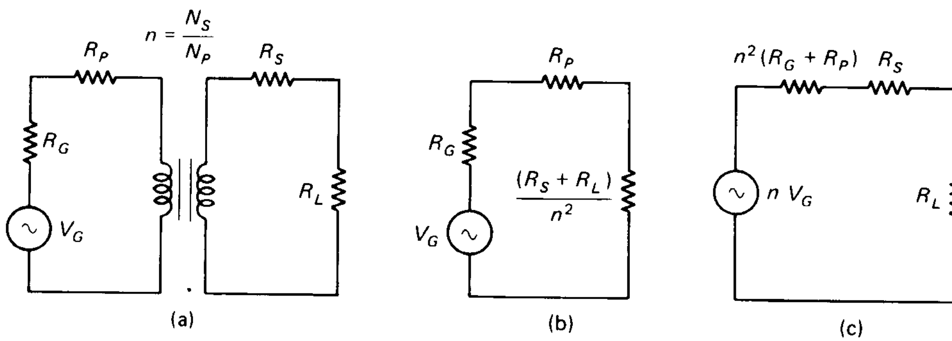


FIGURE 5-7 (a) Power transformer equivalent circuit. (b) Referenced to primary. (c) Referenced to secondary.

The resulting power transformer equivalent circuit is given in Fig. 5-7(a). The equivalents with reflected resistances referenced to the primary and secondary are shown in Fig. 5-7(b) and (c), respectively.

Skin Effect: The skin depth for copper at 60 Hz was calculated in Example 1-4 as 0.85 cm (0.33 in.). If the wire radius is three or four times smaller than this, the skin effect can be neglected and the dc winding resistance can be used in calculation. This is the case for all 60-Hz power transformers that are likely to appear in electronic equipment.

Power-Transformer Ratings: Unfortunately, manufacturers' specifications for electronic transformers generally include only the rms secondary voltage at the maximum rated secondary current with the specified line voltage at the primary. This leaves numerous questions unanswered. The most useful pieces of data for power transformers are the true turns ratio and the primary and secondary winding resistances. The true turns ratio of a power transformer can be determined by simply measuring the ratio of primary to secondary voltages *with no load connected to the secondary*. The primary and secondary winding resistances can be measured with a dc ohmmeter unless the transformer is a real brute, in which case skin-effect resistance will have to be determined.

The IEEE has established norms for power-transformer voltage regulation η which, to the extent that they are followed, permit an approximate calculation of true turns ratio n and Thévenin resistance R_{th} from the manufacturer's specifications of primary voltage V_P, full-load secondary voltage V_O, and secondary current I_O. The formulas for making the conversions are

$$n = \frac{N_S}{N_P} = \frac{V_O(1+\eta)}{V_P} \tag{5-4}$$

$$R_{th} = \frac{\eta V_O}{I_O} \tag{5-5}$$

The recommended values of η, which expresses rise in output voltage from full load to no load, are given below.

Power (W) $[I_O V_O]$	η (%) $\left[\frac{V_{O(NL)} - V_{O(FL)}}{V_{O(FL)}}\right]$
10	10 to 20
100	5 to 7
1000	2 to 4
10,000	2.5

For example, a transformer rated as 117-V primary, 48-V, 2-A secondary, would have a regulation η of about 6%. The true turns ratio and Thévenin resistance would be approximately

$$n = \frac{V_O(1+\eta)}{V_P} = \frac{48(1+0.06)}{117} = 0.435$$

$$R_{th} = \frac{\eta\, V_O}{I_O} = \frac{0.06 \times 48}{2} = 1.44\ \Omega$$

Output-Voltage Calculation: The output voltage for a power transformer supplying sinusoidal current to a load can be predicted by reflecting the primary winding resistance to the secondary side and analyzing a simple series circuit.

EXAMPLE 5-3

A power transformer rated at 35 V, 1 A is tested, with the results appearing in Fig. 5-8(a). Predict the actual output voltage while supplying 1 A to a resistive load, and determine the power loss in the transformer.

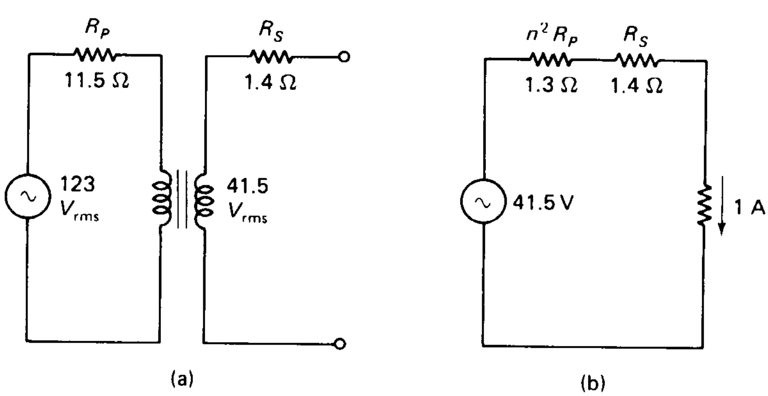


FIGURE 5-8 (a) Transformer and parameters for Example 5-3. (b) Equivalent circuit referenced to secondary.

Solution

The actual turns ratio is

$$n = \frac{N_S}{N_P} = \frac{V_{S(OC)}}{V_P} = \frac{41.5}{123} = 0.337$$

The line voltage and primary resistance reflected to the secondary are

$$V_{\text{ref}} = nV_P = 0.337 \times 123 = 41.5\ V$$
$$R_{P(\text{ref})} = n^2 R_P = 0.337^2 \times 11.5 = 1.3\ \Omega$$

The equivalent circuit is shown in Fig. 5-8(b), and is easily analyzed by Ohm's law:

$$V_{\text{drop}} = IR = 1(1.3 + 1.4) = 2.7\ \text{V}$$
$$V_{\text{load}} = V_{\text{ref}} - V_{\text{drop}} = 41.5 - 2.7 = \mathbf{38.8\ V}$$
$$P_{\text{trans}} = I^2 R = 1^2(1.3 + 1.4) = \mathbf{2.7\ W}$$

The output voltage at any other load current can be predicted in a similar manner. Large overcurrents can be drawn for brief periods with no ill effects other than the great drop in output voltage. (Excessive load currents do not cause core saturation—this is explained in a subsequent section.) Of course, the power burned in the transformer goes up as the *square* of load current, and the heat-dissipation rule of thumb (20 mW/cm^2 of surface area for 20°C rise) presented in Section 3.5 should be kept in mind. Also keep in mind that, for large power transformers, the resistance of the ac line feeding the transformer may be significant. The exact value depends entirely upon the type of service available, but ranges typically from 0.1 to 1 Ω. This value should be added to the primary resistance in the analysis.

Pulsed Load Currents: The most common use of a power transformer in electronic equipment is to drive a rectifier, which in turn feeds a filter capacitor, as in Fig. 5-9(a). The current from the transformer secondary is not sinusoidal but flows in large brief pulses at the peak of the sine wave, as shown in Fig. 5-9(b). There are two important consequences of this current waveshape:

1. The voltage drop across the transformer winding resistances must be calculated using the *peak secondary charging current*, not merely the dc load current.
2. The power dissipated in the transformer winding resistances is, of course, I^2R_W, but *I is the true rms value of the charging current peaks*, not merely the dc load current.

To illustrate these points, consider that the hypothetical rectangular charging pulses of Fig. 5-9(c) apply to the circuit of Fig. 5-9(a). The total effective winding resistance of the transformer is the primary resistance reflected down by n^2, plus

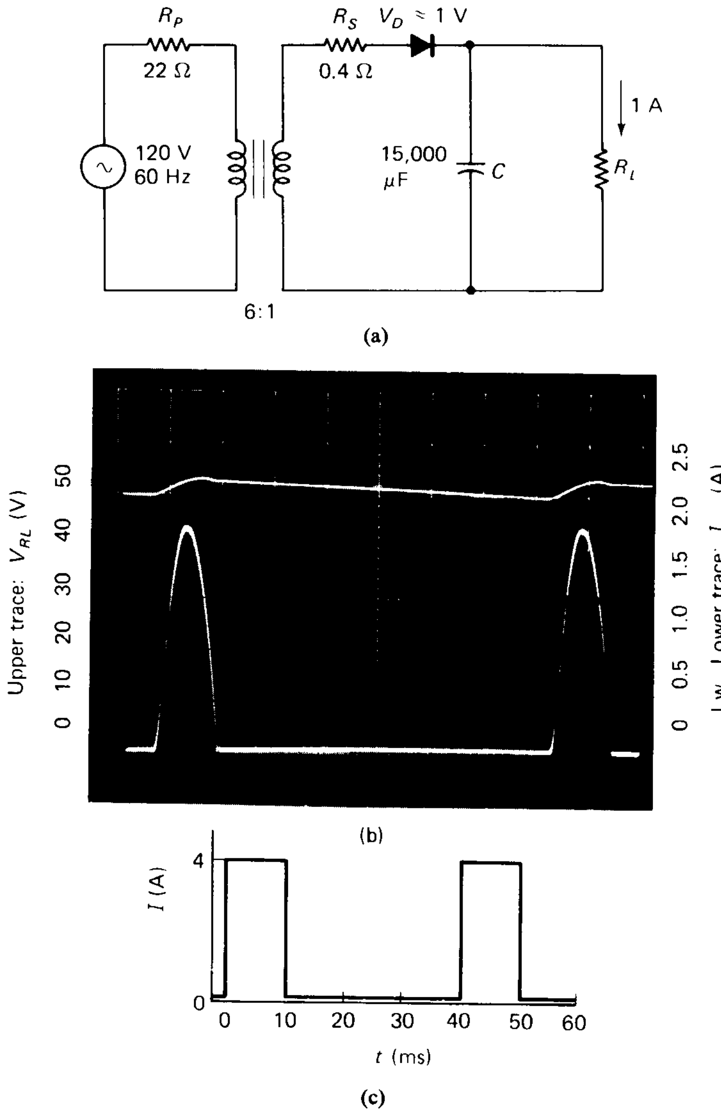


FIGURE 5-9 *(a) Standard rectifier circuit with capacitor filter. (b) Secondary current (lower trace) is not a sine wave, but rather consists of brief pulses. (c) Hypothetical rectangular charging pulses (see the text).*

the secondary resistance:

$$R_W = \frac{22}{6^2} + 0.4 = 1\ \Omega$$

The *average* current in the secondary is 4 A × 10 ms/40 ms = 1 A, which is the load current. However, the voltage drop across R_W is due to the *peak* charging current, and this results in a lower peak voltage to the filter capacitor:

$$V_{\text{drop}} = I_{\text{pk}} R_W = 4\ \text{A} \times 1\ \Omega = 4.0\ \text{V}$$

$$V_{RL(\text{pk})} = V_{\text{sec}(\text{pk},NL)} - V_{\text{drop}} - V_{\text{diode}}$$

$$= \left(\frac{120}{6} \times 1.41\right) - 4.0 - 1.0 = 23.2\ \text{V}$$

The average power burned in the winding resistance is calculated by the following steps:

1. *Square.* Find the power during the 10-ms charging period by $P = I^2 R$:

$$P_{\text{chg}} = I_{\text{pk}}^2 R_W = 4^2 \times 1 = 16\ \text{W}$$

2. *Mean.* Find the average (mean) power during the total 40-ms period:

$$P_{\text{av}} = P_{\text{chg}} \times \text{duty cycle} = 16 \times \frac{10\ \text{ms}}{40\ \text{ms}} = 4\ \text{W}$$

3. *Root.* The power-burning equivalent current can now be calculated by $I = \sqrt{P/R}$:

$$I_{\text{rms}} = \sqrt{\frac{P_{\text{av}}}{R_W}} = \sqrt{\frac{4}{1}} = 2\ \text{A}$$

The application of this *root of the mean of the square* technique for finding the power-burning dc equivalent of a current or voltage is relatively simple for rectangular pulses, but for most waveforms it requires calculus or computer-assisted analysis. For sine waves, the familiar results $I_{\text{rms}} = 0.707 I_{\text{pk}}$ and $I_{\text{rms}}/I_{\text{av}} = 1.11$ are obtained. In general, the more narrow and high-peaked the waveform, the higher the ratio $I_{\text{rms}}/I_{\text{av}}$. Figure 5-10(a) and (b) gives the ratios of peak and rms currents, respectively, to average current in full-and half-wave rectifier circuits with capacitor-input filters. It is assumed that C is large enough so that it discharges to not less than 75% of its peak voltage between charging pulses. Notice that full-wave rectifiers and higher values of R_W result in lower transformer current requirements for a given R_L.

EXAMPLE 5-4

Calculate the dc voltage at the load and the power dissipated in the winding resistance for the circuit of Fig. 5-9(a).

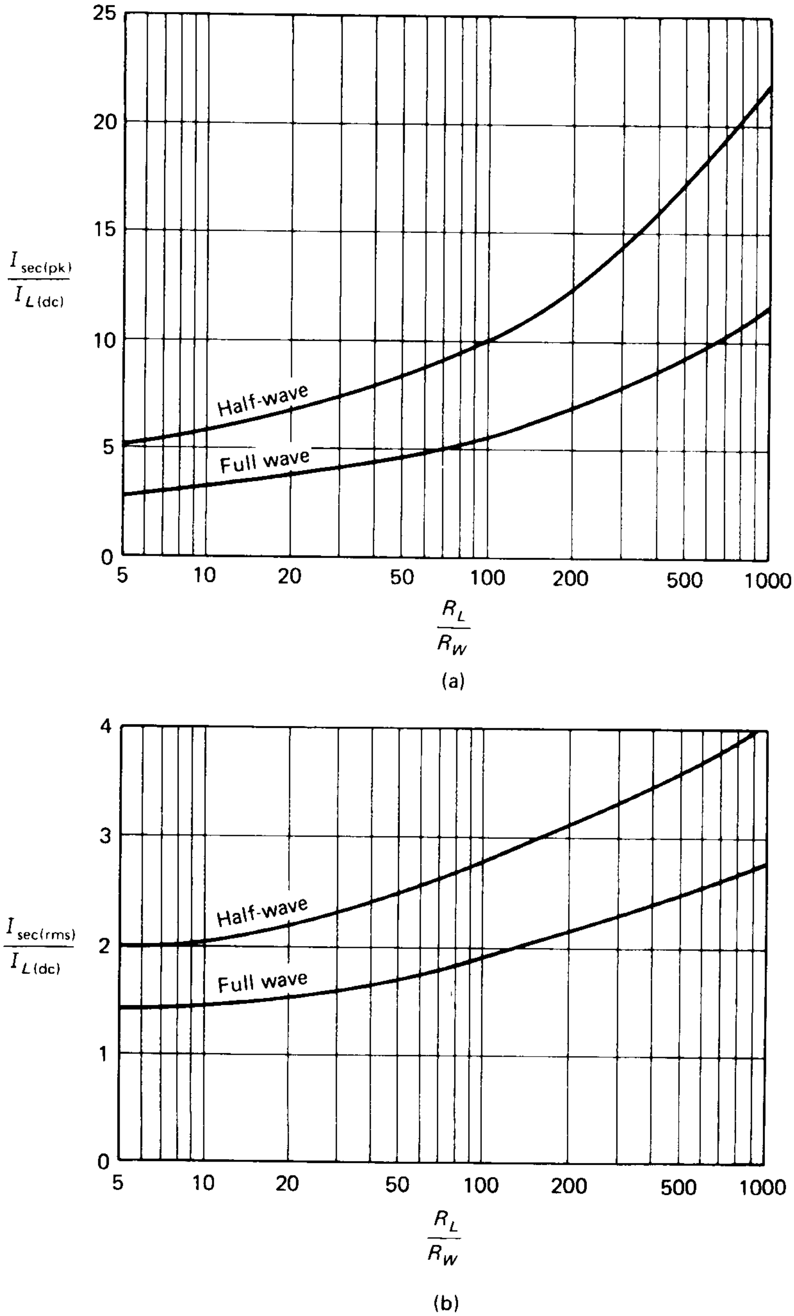


***FIGURE 5-10** (a) Peak secondary current can be 2 to 20 times dc load current for a capacitor-input filter. (b) The transformer secondary must be rated for the true rms value of charging current, which may be from 1.5 to 4 times the dc load current.*

Solution

The effective winding resistance R_W has already been found to be 1 Ω. An estimate of V_{out} is needed to calculate R_L and hence R_L/R_W. We will estimate $V_{out} \approx V_{sec(rms)} = 20$ V:

$$R_L = \frac{V_{out}}{I_L} \approx \frac{20\text{ V}}{1\text{ A}} = 20\ \Omega \qquad \frac{R_L}{R_W} = \frac{20}{1} = 20$$

From Fig. 5-10(a), the peak charging current will be $6.6I_L$ or 6.6 A. The voltage drop across R_W is therefore that due to 6.6 A through 1 Ω. The load voltage is the peak reflected voltage minus the drop across R_W, minus the diode drop, which we estimate as 1.0 V:

$$\begin{aligned} V_{RL(pk)} &= V_{sec(pk)} - V_{RW} - V_{diode} \\ &= 20 \times 1.41 - 6.6\text{ A} \times 1\ \Omega - 1.0\text{ V} = \mathbf{20.6\ V} \end{aligned}$$

For the sake of completeness we will calculate the minimum (valley) voltage of the rippled dc, using the constant-ratio-of-discharge formula $CV = It$:

$$V_{droop} = \frac{I_L t}{C} = \frac{1\text{ A} \times 16.7\text{ ms}}{15{,}000\ \mu F} = 1.1\text{ V}$$

$$V_{valley} = V_{RL(pk)} - V_{droop} = 20.6 - 1.1 = 19.5\text{ V}$$

From Fig. 5-10(b), the rms secondary current (which must not exceed the transformer rating) is 2.1 times I_L, or 2.1 A. The power in the winding resistance is

$$P_W = I^2 R_W = 2.1^2 \times 1 = \mathbf{4.4\ W}$$

The utilization of the secondary winding must be considered when applying this analysis technique to other rectifier circuits. Secondary ratings for the half-wave and the full-wave bridge circuits can be obtained directly as above. A full-wave center-tapped transformer rated at 1 A can be thought of as two 1-A windings, equivalent to a single 2-A winding, since each half-winding is used only half the time. Both the full- and half-wave voltage-doubling circuits can be viewed as two half-wave circuits, each feeding $\frac{1}{2} R_L$, so the rms current for one half-wave circuit feeding $\frac{1}{2} R_L$ from actual R_W must be found and then doubled to find the secondary rating required. A choke-input filter allows nearly constant charging current if the choke is large, eliminating the peaks of charging current which produce high rms currents. Figure 5-11 summarizes the various rectifier circuits and the common ratios of $I_{sec(rms)}/I_{L(dc)}$. The low ratios given are for $R_L/R_W = 10$ and the high numbers are for $R_L/R_W = 100$. The ratio of 1000 is not considered because it is likely to occur only with high R_L when currents are relatively low anyway.

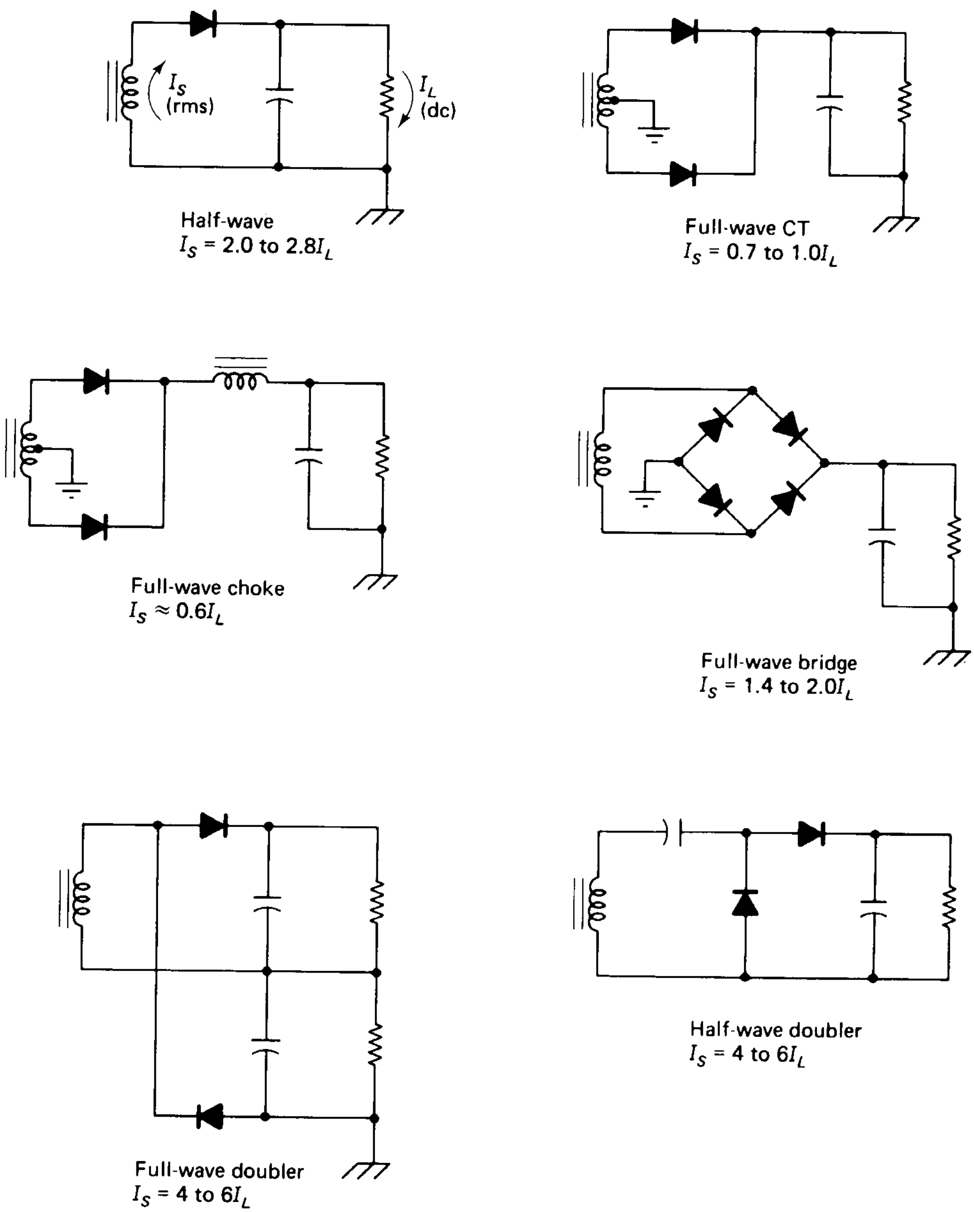
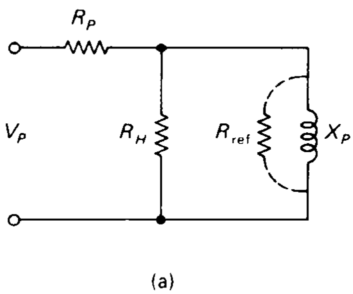


FIGURE 5-11 *Recommended secondary rms current ratings for various rectifier circuits.*

No-load Currents: Although they have no effect on the ability of a transformer to supply current to a load, it may be well to consider the *magnetization currents* that flow in the primary of a power transformer regardless of load. Figure 5-12 shows the equivalent-circuit components involved, and the accompanying phasor diagram gives an idea of the relative magnitudes of their currents. Transformers for 60-Hz electronics equipment generally have primary inductances ranging from about 1 H for 500-W units, up to 10 H for 2-W units. The smaller transformers require higher inductance to limit the magnetizing current to avoid saturating the tiny core. The Q of the primary inductance is generally between 1.5 and 3. Most of the resistive component of primary impedance is due to losses in the core material, represented as shunt resistance R_H. Series winding resistance R_P is very much smaller than R_H, except in the smallest transformers, where the fine primary winding may have a resistance as much as $\frac{1}{5} R_H$. R_H is calculated as QX_P, where Q is measured on an impedance bridge, calculated from tangent of the angle between no-load current and voltage, or simply estimated as 2 for standard power transformers.



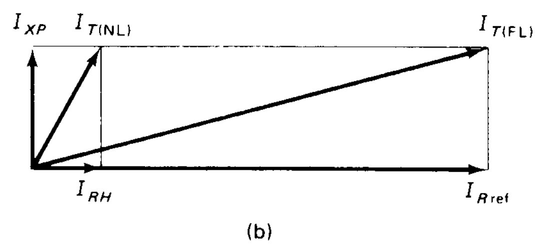


FIGURE 5-12 (a) Transformer primary equivalent circuit showing winding resistance R_p and core-loss resistance R_H. (b) Phasor diagram showing relation between no-load and full-load primary currents.

EXAMPLE 5-5

A nominal 24-V 2-A 60-Hz power transformer has the following parameters:

$$V_P = 120\text{ V} \qquad R_P = 12\ \Omega$$
$$L_P = 3\text{ H} \qquad R_S = 1.5\ \Omega$$
$$Q_P = 2.5 \qquad n = 0.278$$

Find the power dissipated within the transformer at full load.

Solution

The core loss is computed first, using X_{LP} and Q_P to find R_H:

$$X_{LP} = 2\pi fL = 2\pi \times 60 \times 3 = 1130\ \Omega$$
$$R_H = Q_P X_P = 2.5 \times 1130 = 2825\ \Omega$$
$$P_{\text{core}} = \frac{V_P^2}{R_P} = \frac{120^2}{2825} = 5.1\text{ W}$$

Next, the winding loss is found by reflecting R_P to the secondary:

$$V_{\text{ref}} = nV_P = 0.278 \times 120 = 33.3 \text{ V}$$

$$R_{S(\text{ref})} = n^2 R_P = 0.278^2 \times 12 = 0.93 \ \Omega$$

$$P_{\text{winding}} = I_L^2 (R_{P(\text{ref})} + R_S) = 2^2(0.93 + 1.5) = 9.7 \text{ W}$$

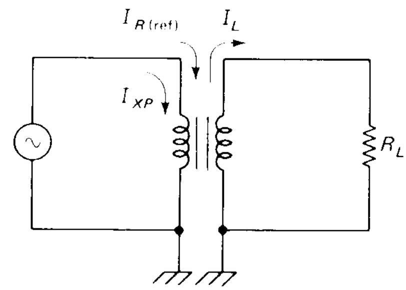
$$P_{\text{total}} = P_{\text{core}} + P_{\text{winding}} = 5.1 + 9.7 = \mathbf{14.8} \text{ W}$$

Core Saturation: The current in the magnetizing inductance X_P increases with higher primary voltages or lower line frequencies. The transformer's number of primary turns and core structure have been designed for a particular voltage and frequency, and higher voltages or lower frequencies are likely to cause core saturation (see Fig. 4-16).

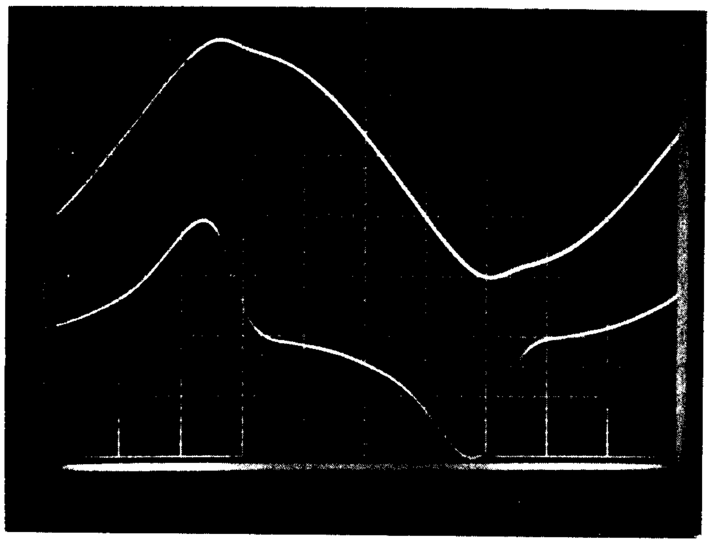
Load current does not contribute to core saturation because the primary current flowing into the reflected load resistance produces a magnetomotive force which exactly cancels that of the secondary load current, as illustrated in Fig. 5-13(a). You may ask why higher-power transformers always have larger iron cores if load current does not cause saturation. The answer is that high load currents require heavier wire to minimize copper losses. This means that a larger winding form is required and that fewer turns can be fit on the form, so X_{LP} is lower and magnetizing current is higher for larger transformers. Transformers for 400 Hz can achieve high X_{LP} with 2.5 times fewer turns, and do in fact use about 2.5 times less iron than that used by 60-Hz transformers. Power transformers for dc inverter-type supplies operate in the kilohertz range and are much smaller still.

Since the ratio of voltage to frequency at the threshold of saturation is constant for any given transformer winding, a signal generator and oscilloscope can be used to check the saturation limit. Set the generator to a low frequency, say 6 Hz, and connect it to the primary. Observe the secondary waveform on the oscilloscope, watching for signs of saturation as the primary voltage is increased. What you see at 15 V at 6 Hz is what you can expect at 150 V at 60 Hz, or at 125 V at 50 Hz. The low-frequency test is a little less destructive, however. Figure 5-13(b) shows primary and secondary voltages for an extreme case of saturation. The primary is being fed from a 50-Ω source. Notice that, when saturation is entered, the primary voltage is loaded down from its usual sine shape, and the secondary voltage simply disappears. The primary current at this time is limited only by the dc resistance in the primary circuit, and at 120 V it would be destructively high.

Core saturation often causes a tremendous current surge when a power transformer is switched across the ac line. Figure 5-13(c) shows why. At steady state (dotted line) the inductive current lags the voltage by 90°, rising from −1 A to +1 A during the positive voltage half-cycle in Fig. 5-13(c). However, if the switch happens to close at the start of a voltage half-cycle, the current rises from an initial value of zero to +2 A *provided that the core doesn't saturate first* (dashed line.) In fact, manufacturers almost never put in the extra iron and copper to prevent turn-on saturation, so primary current is limited only by primary winding resistance in the last part of the first half-cycle. Figure 5-13(d) shows the turn-on



(a)



(b)

FIGURE 5-13 *(a) Load current does not cause core saturation because $I_{R(ref)}$ cancels its magnetizing effect. (b) Primary (upper trace) and secondary voltages for a transformer in saturation.*

surge for a 200-W transformer reaching 45 A, while the inset trace shows a steady-state peak less than 1 A.

Turn-on surge may cause trouble with fast circuit breakers or line noise. The simplest way to alleviate the problem is to leave the primary connected permanently to the line and switch the secondary.

5.4 BROADBAND TRANSFORMERS

The following characteristics are assumed for broadband transformers in this discussion:

- Operation over a frequency range spanning at least one decade, possibly as much as five decades
- Magnetic core with coefficient of coupling of 0.75 or greater

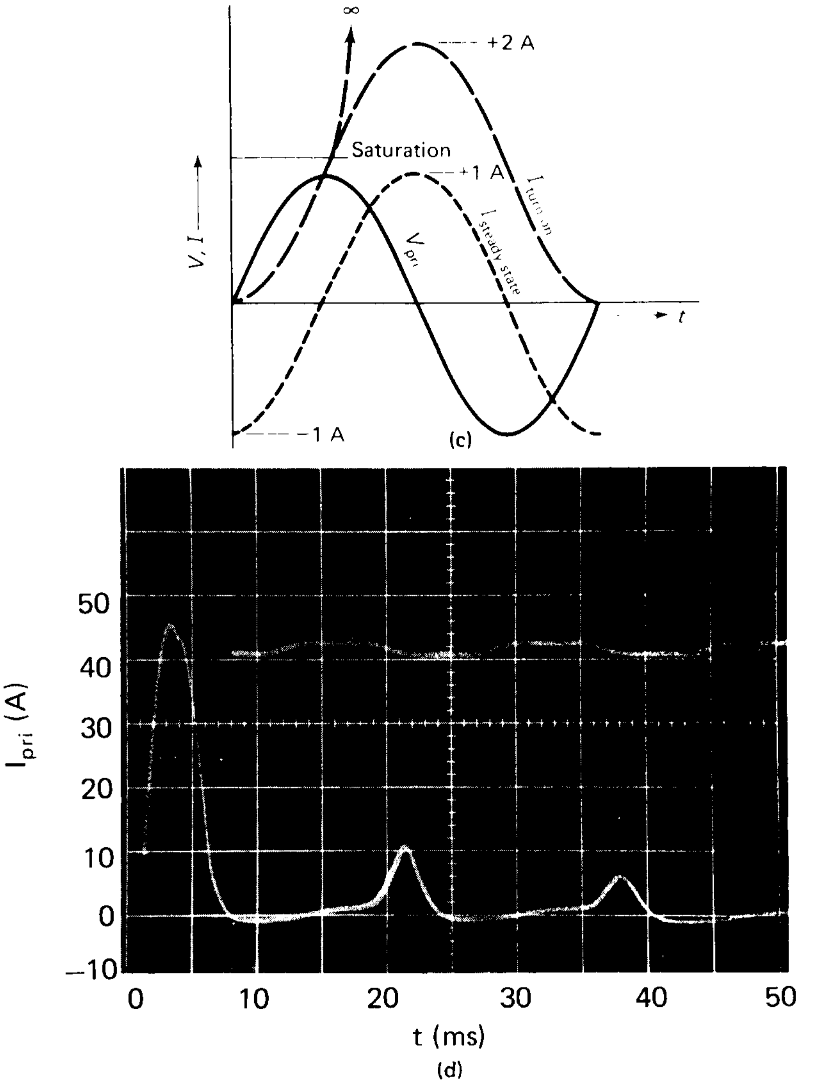


***FIGURE 5-13** (c) Turn-on current starts from zero and exceeds the saturation level if the primary is switched on at the start of a voltage half-cycle. (d) Actual turn-on surge for a 200-W transformer with steady-state current (inset) for comparison.*

The chief representative of this class is the audio transformer, wound on a core of laminated silicon-iron strips. The more recent development of high-frequency ferrite materials has made possible the construction of broadband video and radio-frequency transformers, even into the UHF range.

All of the components shown in the equivalent circuit of Fig. 5-6 may become involved in the analysis of broadband transformers, with the additional considerations of skin-effect resistance, series resonance of the windings, and saturation and high-frequency loss in the core. We can keep the analysis reasonably simple, however, by concentrating on two main questions: low-frequency cutoff and high-frequency cutoff.

Low-Frequency Cutoff: A transformer can cease to function at low frequencies for one of two reasons: core saturation or inductive loading.

Core saturation can be calculated from equation 4-27 if the number of primary turns is known and the core area can be determined. It must be remembered that core saturation for any given transformer depends upon primary voltage and frequency only, as illustrated in Fig. 5-13. The result of low-frequency core saturation is not simply reduction in output, but rather waveform distortion. The cure can be a larger transformer (higher X_{LP} or bigger A_{core}), lower primary voltage, or limiting the low-frequency response before the signal gets to the transformer.

Inductive Loading: As signal frequency is lowered, the inductive reactance of the primary drops. At some frequency X_{LP} will equal the Thévenin equivalent source resistance through which it is driven, and the voltage across it will then drop to 0.707 of maximum. This is illustrated in Fig. 5-14. Combining the expressions for X_P, R_{ref}, and R_{th} yields a formula for low-frequency cutoff:

$$f_{\text{lo}} = \left(\frac{R_G + R_P}{2\pi L_P}\right) \frac{R_L/n^2}{R_G + R_P + \dfrac{R_L}{n^2}} \tag{5-6}$$

where $n = N_S/N_P$, f is in hertz, L is in henrys, and all R's are in ohms. Examination of this equation reveals that raising L_P is the most effective way of obtaining the lowest possible operating frequency. Lowering R_G and R_P, and to a lesser extent raising R_L, will also help.

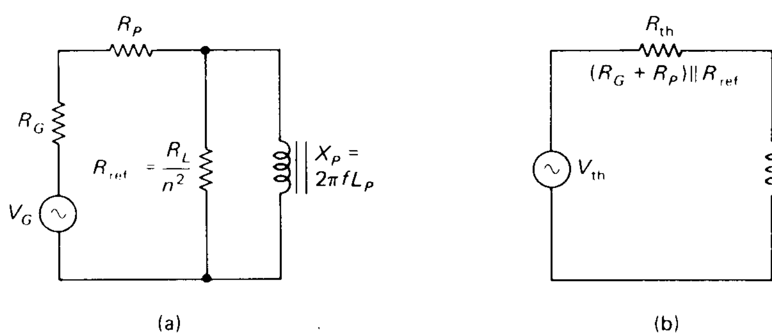


FIGURE 5-14 *Low-frequency cutoff occurs when primary reactance drops lower than effective driving resistance: (a) equivalent primary circuit; (b)* Thévenin equivalent.

High-Frequency Cutoff: The response of a transformer at high frequencies can be limited by any of three factors: leakage reactance, core losses, and series resonance.

Leakage Reactance: The coefficient of coupling of broadband transformers ranges from above 99.5% for audio types down to 80% or so for some of the low-μ ferrite toroids. In any case, the leakage reactance in the primary X_{PL} is not zero, and it appears in series between the source and the load, as shown in Fig. 5-15. As frequency increases, X_{PL} will increase until it becomes large enough to limit the voltage across R_{ref} to 0.707 of its midband value. In most cases, core loss R_H and coupled reactance X_{Pk} are negligible compared with the low value of R_{ref}, so the simple circuit of Fig. 5-15(b) applies and

$$f_{hi} = \frac{R_G + R_P + \dfrac{(R_S + R_L)}{n^2}}{2\pi(1-k)L_P} \tag{5-7}$$

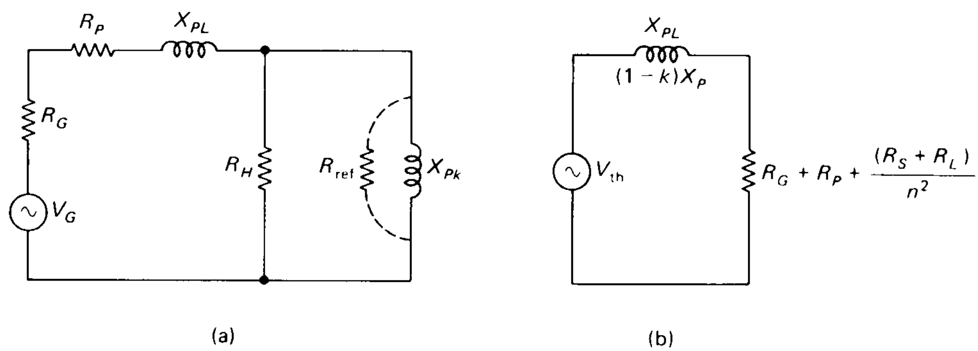
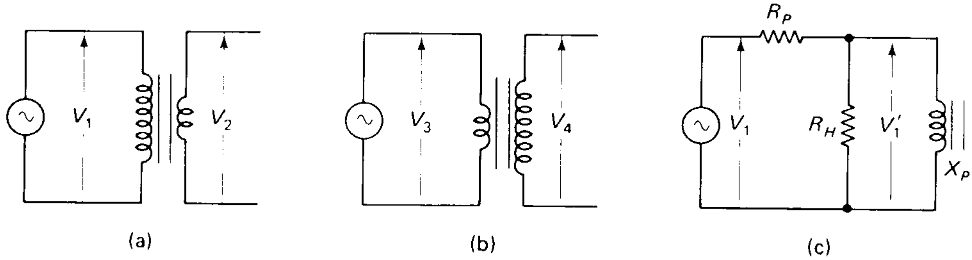


***FIGURE 5-15** Leakage reactance can be great enough to cause high-frequency cutoff: (a) primary equivalent circuit; (b) Thévenin equivalent.*

Notice that most of the things that will improve high-frequency response (lower L_P, higher R_G, higher R_P) are precisely the things that will degrade low-frequency response—such are the frustrations of life. Raising R_L will aid both causes, but driving a load is usually a prerequisite, unless we can contrive to feed the secondary into an emitter follower or FET amplifier.

The real key to high-frequency response (at least this one-third of it) is a coefficient of coupling k that approaches unity, which then results in a leakage reactance approaching zero. High-k values are obtained by keeping both coils close together on a high-μ form with minimum or zero air gap. Measurement of k is a fairly simple matter and is illustrated in Fig. 5-16(a) and (b). A test voltage at a frequency toward the high end of the transformer midband range is applied to the first winding and the primary (V_1) and secondary open-circuit (V_2) voltages are measured. The transformer is then reversed, and the test voltage is applied to the second winding as the new primary, and input (V_3) and open-circuit output (V_4) are measured. The value of k is then found as

$$k = \sqrt{\frac{V_2 V_4}{V_1 V_3}} \tag{5-8}$$



$$k = \sqrt{\frac{V_2 V_4}{V_1 V_3}}$$

FIGURE 5-16 Measuring coefficient of coupling k *is relatively simple: (a) and (b) the basic measurements; (c) correction technique for use when* k *is near unity and high precision is required.*

The true turns ratio of the transformer can also be calculated from these data:

$$n = \sqrt{\frac{V_2 V_3}{V_1 V_4}} \tag{5-9}$$

If k is within a percent or two of unity, n can be found more simply as V_{out}/V_{in}, but as k becomes lower this calculation loses accuracy and equation 5-9 must be used.

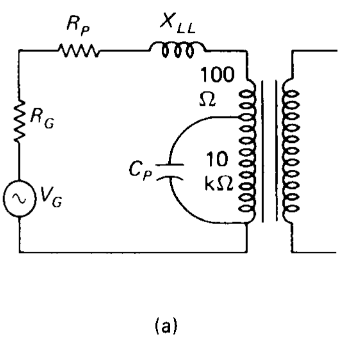
Figure 5-16(c) shows that the V_1 and V_3 input voltages measured may not exactly equal the voltage across the primary reactance V_1'. By making f_{test} relatively high, we make X_P large and minimize the drop across R_P, but if it is desired to calculate k to within 0.1% it will be necessary to find the Thévenin equivalent of V_1, R_P, and R_H, and calculate the true value of V_1'. Similar treatment must be given to V_3 on the second measurement set, and of course the voltmeter used must be precise to at least $3\frac{1}{2}$ or 4 digits.

Core Losses: At higher frequencies the hysteresis losses in a magnetic material begin to increase, lowering the Q of the coil and lowering the value of R_H in Fig. 5-15(a). The methods of defense against this mode of high-frequency cutoff are:

- Use a core material with a higher maximum frequency than the application calls for (if possible).
- Use a low generator source resistance R_G.
- Use heavy primary wire to lower R_P.
- Keep k as high as possible by placing the windings close together on a high-μ core with no air gap.

At some frequency, generally considerably above the frequency where hysteresis loss becomes severe, the permeability of the core material will begin to drop. Above this frequency the magnetic coupling between the windings decreases (lower k value) and no combination of circuit parameters will elicit a full response from the transformer.

Series Resonance: The most common cause of high-frequency cutoff in broadband transformers is series resonance of the primary winding. Parallel resonance of the winding with its stray capacitance presents no problem since the reactance simply increases, but when the first few turns of the primary coil resonate with the series stray capacitance from the center of the coil, a very low impedance results, which tends to load down the source. The worst part of it is that the capacitance shunts out the bulk of the primary winding, effectively increasing the turns ratio N_S/N_P of the transformer. An attempt to illustrate this phenomenon is made in Fig. 5-17(a).



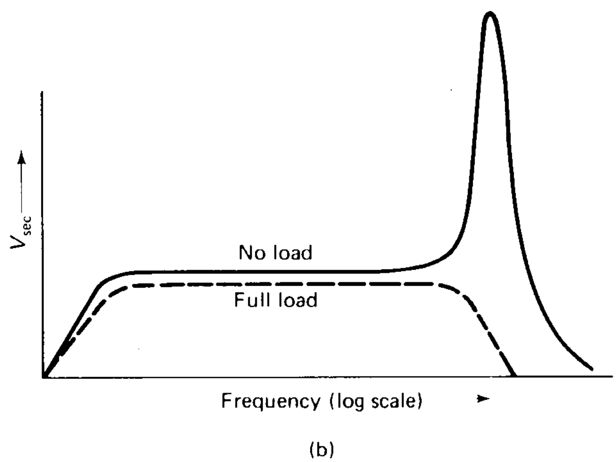


FIGURE 5-17 *(a) Series resonance of the first few primary turns with the bulk stray capacitance is the most common cause of high-frequency cutoff in an audio transformer. (b) Dangerously high voltages can be generated at the series-resonant frequency if the transformer is operated without a secondary load.*

With no load on the secondary, this new turns ratio will result in a severe peaking of the secondary output voltage at the series-resonant frequency, as shown in Fig. 5-17(b). The peak voltages attained can easily reach 10 times their normal values. Large signals containing high-frequency components should never be applied to an unloaded transformer, or insulation breakdown could result. The peaking problem can be completely eliminated by making sure that the secondary reflects back even a moderate load to the primary. The Q of the primary inductance and consequently the series-resonant voltage rise will thereby be reduced.

Increasing the upper-frequency limit imposed by series resonance is another matter, however. The best solution is to limit stray capacitance by keeping the number of turns low. This can be done without lowering L_P and thus ruining low-frequency response if a low-reluctance gapless core of high-μ material is used.

Spaced and cross-diagonal winding techniques for reducing capacitance are difficult to implement physically and the gains to be realized in this way are small.

Calculating the series-resonant point is chancy at best, but an estimate can be given as 3 to 10 times the parallel-resonant frequency, which can be calculated from the inductance and the estimated stray capacitance of 0.05 pF/turn.

EXAMPLE 5-6

A 50-turn to 50-turn transformer is wound of No. 22 enameled wire on a 3B7 toroid core with average diameter 2 cm and cross-sectional area 0.25 cm^2. $V_G =$ 20 V rms, $R_G = 50\ \Omega$, and $R_L = 50\ \Omega$. Coupling k is measured as 98% per Fig. 5-16. Predict the upper and lower cutoff frequencies.

Solution

Consulting Fig. 4-15 for μ and using equation 4-25, we obtain

$$L_P = 0.012 N^2 \mu \frac{A}{\ell} = 0.012 \times 50^2 \times 2000 \times \frac{0.25}{2\pi} = \mathbf{2400\ \mu H}$$

The low cutoff due to inductive loading, by equation 5-6, is

$$f_{lo} = \left(\frac{R_G + R_P}{2\pi L_P} \right) \frac{R_L/n^2}{R_G + R_P + \dfrac{R_L}{n^2}}$$

Since $n = 1$ and R_P is negligible, this becomes

$$f_{lo} = \frac{50}{2\pi \times 0.0024} \frac{50}{50 + 50} = 1.7\ \text{kHz}$$

Of the 20-V source, 10 V is dropped across R_G and 10 V appears across R_{ref}. The low cutoff due to core saturation, by equation 4-27, is

$$f_{lo} = \frac{V_{rms} \times 10^8}{4.4 B_{sat} N A} = \frac{10 \times 10^8}{4.4 \times 4000 \times 50 \times 0.25} = \mathbf{4.5\ kHz}$$

Thus core saturation governs at the 10-V level, but f_{lo} could be as low as 1.7 kHz if V_G were limited.

High-frequency cutoff due to leakage reactance, by equation 5-7, is

$$f_{hi} = \frac{R_G + R_P + \dfrac{R_S + R_L}{n^2}}{2\pi(1 - k) L_P} = \frac{50 + 50}{2\pi \times 0.02 \times 0.0024} = \mathbf{330\ kHz}$$

This is about equal to the limit imposed by the core material (Fig. 4-15).

A wild guess at the series-resonant point may be ventured:

$$C_p \approx 0.05\ \text{pF/turn} \times 50\ \text{turns} = 2.5\ \text{pF}$$

$$f_{r(\text{parallel})} \approx \frac{1}{2\pi LC} = \frac{1}{2\pi\sqrt{2.4\ \text{mH} \times 2.5\ \text{pF}}} = 2\ \text{MHz}$$

The series-resonant point is therefore likely to be above 6 MHz, which is safely above the previously calculated 330 kHz.

A final calculation of the skin-effect resistance of the coil will be made to ensure that it is negligible in the face of R_G and R_L. The thickness of AWG 22 wire is 0.064 cm. The total length, assuming that the core is 0.5 cm on a side, is 2 cm/turn × 50 turns, or 100 cm. Using equation 1-4 yields

$$R = 8.3 \times 10^{-8} \frac{\sqrt{f}}{t} l = 8.3 \times 10^{-8} \frac{\sqrt{330{,}000}}{0.064} \times 100 = 0.075\ \Omega$$

Audio-Transformer Specifications: Audio-transformer ratings generally include:

- *Primary and secondary impedances*. These are the load resistance and source resistance with which the transformer is designed to operate over its specified frequency range, and do not directly designate the inductance or impedance of the windings themselves. The true turns ratio n can be calculated as $\sqrt{Z_S/Z_P}$.

- *Audio-power rating*. This limit is generally imposed by core saturation at the low-frequency limit, and can be increased considerably if the audio lows are attenuated before reaching the transformer.

- *Winding dc resistances*. These are often high enough to enter into dc bias calculations. They also impose a power limit on the transformer due to the I^2R loss in each winding, and the resultant temperature rise.

- *Maximum dc winding current*. Transformers frequently carry dc as well as ac, and the dc contributes its own I^2R heating. If the winding is single-ended, dc also contributes toward reaching the saturation limit. In "push-pull" center-tapped transformers the dc flows into the center and equally out each end, so the magnetizing effects cancel.

- *Frequency response*. This is given for the primary and secondary impedances (resistive R_G and R_L) specified. If R_G and R_L are lower than specified, a downward shift in the response range may be expected. Raising R_G and R_L will raise the lower cutoff but probably not the upper cutoff, since it is most likely to be due to series resonance of the winding.

The Audio Range: Considerable research has gone into determining the range of frequency response required for audio reproduction, with the following fairly well-accepted results:

Range (Hz)	*Application*
300–3500	Understandable voice communication
80–8000	Natural voice, reasonable music
50–15,000	Good-fidelity music
15–20,000	Extremely high fidelity

Analysis of typical audio waveforms shows that they tend to have higher peaks with lower effective (rms) values than do simple sine waves. This means that saturation and other peak-clipping phenomena are likely to be more troublesome than power-dissipation limitations in audio systems. A two-tone test—two signals added, for example, 400 Hz in series with 1 kHz—is often used to simulate an audio waveform.

5.5 RADIO-FREQUENCY TRANSFORMERS

In dealing with RF transformers we can expect to encounter a rather wide range of physical device forms and a variety of design objectives. Among the devices:

- Ferrite toroid-core transformers; k values from 0.80 to 0.98; generally unsuitable at high power levels because of core saturation
- Powdered-iron slug-tuned transformers; k values around 0.75; also susceptible to saturation at high levels
- Air-core transformers; no saturation or nonlinearity problems; k values depending on structure:

bifilar winding (two strands wound together side by side)	$k \sim 0.95$
primary over secondary (single layers, 1:1 ratio)	$k \sim 0.90$
link coupling (2 or 3 turns over 10 or more turns)	$k \sim 0.60$
end-to-end coils (length $= \frac{1}{2}$ diameter)	$k \sim 0.35$
end-to-end coils (length = 2 diameters)	$k \sim 0.10$
in-line separated by one length (length = 2 diameters)	$k \sim 0.02$

Among the objectivities:

- Transfer power from a source to a load, possibly with impedance step up or step down
- Tune (select) a band of frequencies while rejecting others

In general, the more we try to transfer power, the less sharply we will be able to tune; and the more sharply we wish to tune, the less will we be able to transfer power.

Untuned RF Transformer: For an untuned rf transformer to be successful, the coefficient of coupling must be high (say, above 60%), so it is necessary to use toroid cores at low power levels and winding-over-winding air-core structures at high power levels. Figure 5-18 shows the transformer equivalent circuit, and simplifications referenced to the primary and secondary, respectively. The latter is most useful for analysis of results at the load. The "prime" marks indicate reflected values; voltages are scaled by n and impedances by n^2. If $n = 1$, the original values apply, of course.

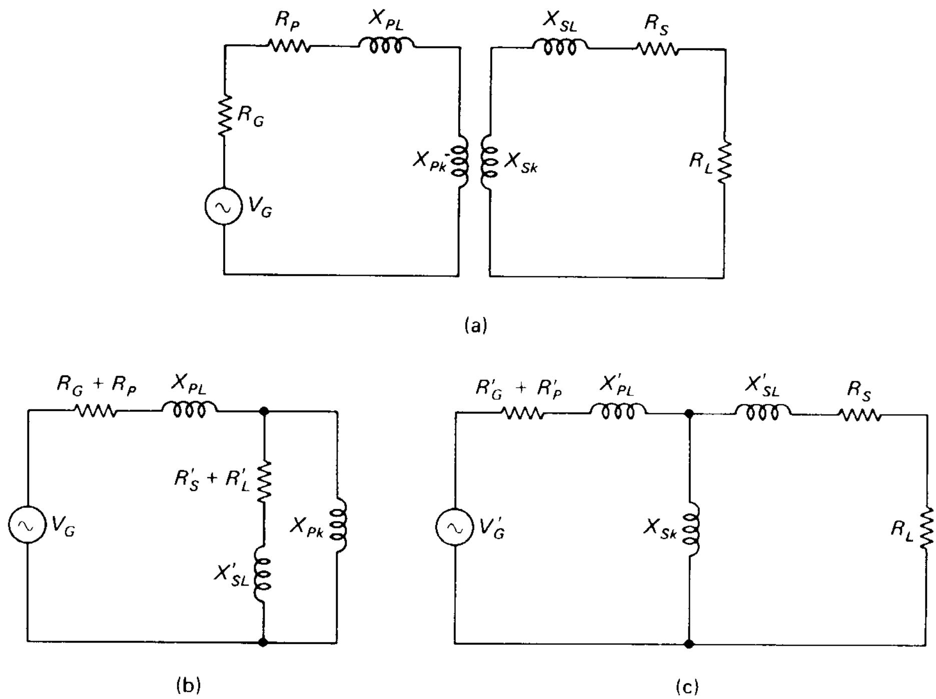


***FIGURE 5-18** Untuned RF transformer: (a) basic equivalent circuit; (b) referenced to primary; (c) referenced to secondary. Leakage reactance is usually the factor that limits output.*

Examination of Fig. 5-18(c) will show that maximum power is delivered to the load when the primary and secondary leakage reactances are kept small in comparison to the generator and load resistances, respectively, and when the corresponding coupled reactances are kept large in comparison to R_G and R_L. A high-k value achieves both of these conditions at once, so winding a coil with the closest possible coupling is of prime importance. Beyond this, a decision on the number of turns required for each coil must be made. Much calculation with Fig. 5-18 can be done to find the optimum coil reactances, but it turns out that if we choose $X_P = R_G$ and $X_S = R_L$, very little improvement in power transfer can be made, and this simple rule of thumb is recommended where k is below 90%. For higher k values, $X_P = 2R_G$ and $X_S = 2R_L$ will give noticeably greater output.

Above the design frequency, X_{PL} and X_{SL} will become large enough to block current to R_L. Below the design frequency X_{Pk} (or X_{Sk}, depending on how you look at it) will become small enough to shunt the output, causing all the voltage to drop across R_G. If k is large (0.90 or so), there will be quite a bit of room between these two cutoff points, and bandwidths spanning more than one decade may be achieved easily. Low-k values will result in narrower bandwidths, however.

EXAMPLE 5-7

An untuned rf transformer consists of a 40-turn primary wound over an 80-turn secondary, both of AWG 32, both spaced over 2 cm on a 1-cm-diameter form. Measured k is 0.80. $R_G = 75\ \Omega$ and $R_L = 300\ \Omega$. V_G(open circuit) is 2.5 V rms. Find V_{RL} at the optimum frequency f.

Solution

First the inductance of the primary and the frequency where $X_P = R_G$ are found:

$$L_P = \frac{d^2N^2}{46d + 101l} = \frac{40^2}{46 + 202} = 6.45\ \mu\text{H} \qquad (4\text{-}19)$$

$$f = \frac{X}{2\pi L} = \frac{75\ \Omega}{2\pi \times 6.45\ \mu\text{H}} = 1.85\ \text{MHz}$$

The inductive reactances, coupled and leakage, are found next. X_S is related to X_P by $(N_S/N_P)^2$

$$X_P = 2\pi fL = 75\ \Omega \qquad X_S = n^2X_p = 2^2 \times 75 = 300\ \Omega$$

$$X_{Pk} = 0.80 \times 75 = 60\ \Omega \qquad X_{PL} = 75 - 60 = 15\ \Omega$$

$$X_{Sk} = 0.80 \times 300 = 240\ \Omega \qquad X_{SL} = 300 - 240 = 60\ \Omega$$

Then skin-effect winding resistances are found:

$$\ell_P = \pi dN = \pi \times 1 \times 40 = 126\ \text{cm}$$

$$\ell_S = 2\ell_P = 252\ \text{cm}$$

$$r_P = 8.3 \times 10^{-8}\frac{\sqrt{f}}{t}\ell = 8.3 \times 10^{-8}\frac{\sqrt{1.85\ \text{MHz}}}{0.020} \times 126 = 0.7\ \Omega \qquad (1\text{-}4)$$

Considering the proximity effect, the primary resistance is estimated as ten times the calculated value or 7 Ω. Secondary resistance is 14 Ω. Figure 5-19(a) shows the circuit with the values calculated above. The equivalent circuit with primary values referenced to secondary (marked with a prime) is given in Fig. 5-19(b). The method of solution for V_{RL} will be to obtain the parallel equivalent of X_{SL} and $R_S + R_L$, combine the resulting parallel X_L with X_{Sk}, and convert back to a series equivalent, shown in Fig. 5-19(c). Solution for Z_T and V_{AB} is straightforward in this circuit, and knowledge of V_{AB} in Fig. 5-19(b) leads directly to I_L and V_{RL}. The parallel

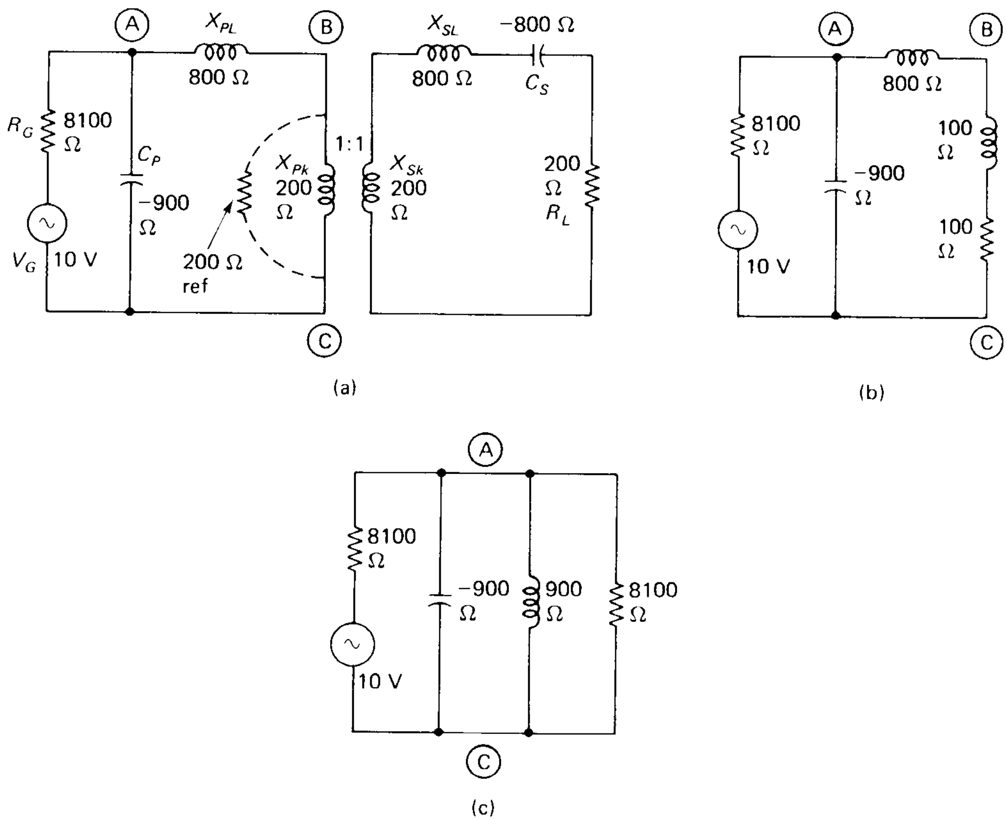


***FIGURE 5-19** Example 5-7; analysis of an untuned RF transformer circuit: (a) equivalent circuit; (b) referenced to secondary; (c) simplified to a simple series circuit.*

equivalent of 60 Ω inductive and 314 Ω resistive in series:

$$R_p = \frac{X_s^2 + R_s^2}{R_s} = 325\ \Omega \tag{4-9}$$

$$X_p = \frac{X_s^2 + R_s^2}{X_s} = 1703\ \Omega \tag{4-10}$$

$$X_p \| X_{Sk} = \frac{1703 \times 240}{1703 + 240} = 210\ \Omega$$

The series equivalent of 210 Ω inductive and 325 Ω resistive in parallel:

$$R_s = X_p \frac{X_p R_p}{X_p^2 + R_p^2} = 96\ \Omega \tag{4-7}$$

$$X_s = R_p \frac{X_p R_p}{X_p^2 + R_p^2} = 148\ \Omega \tag{4-8}$$

Now, looking at Fig. 5-19(c), we obtain

$$Z_T = \sqrt{X^2 + R^2} = \sqrt{208^2 + 424^2} = 472\ \Omega$$

$$I_T = \frac{V_G'}{Z_T} = \frac{5.0}{472} = 0.0106\ \text{A}$$

$$V_{AB} = IZ_{AB} = 0.0106\sqrt{148^2 + 96^2} = 1.87\ \text{V}$$

Returning to Fig. 5-19(b), we find that

$$I_L = \frac{V_{AB}}{Z} = \frac{1.87}{\sqrt{60^2 + 314^2}} = 0.00585\ \text{A}$$

$$V_{RL} = I_L R_L = 0.00585 \times 300 = \mathbf{1.75\ V}$$

Anyone willing to repeat these calculations for a few other frequencies will find the following results:

f (MHz)	V_o(V)
0.925	1.37
1.85	1.75
3.70	1.81
5.55	1.71
9.25	1.36

Series-Tuned RF Transformer: When coefficient of coupling is low and R_G is less than X_{PL}, it will be advantageous to cancel the primary leakage reactance with a series capacitance. Cancellation will occur only for a narrow band of frequencies, so the broadband characteristic of the transformer will be lost if this technique is employed.

On the secondary side, a series-tuning capacitor to cancel leakage reactance X_{LS} will also increase output, but again, R_L must be lower than X_{LS} if the increase is to be significant.

With both the primary and secondary series tuned, much lower coefficients of coupling can be tolerated with good power transfer than in the untuned transformer. However, very low values of k produce very low coupling reactances X_{Pk}, and this tends to short the source. Attempts to raise the input impedance by increasing the primary inductance may be helpful to an extent, but will require more windings and result in larger losses in winding resistances R_P and R_S. The best approach with series-tuned transformers is to keep k at least reasonably high (say, 0.10 minimum).

The required values of the primary and secondary tuning capacitors are interactive because of the reflected impedances that appear across the coupled reactances. With reference to Fig. 5-20, let us assume that $X_{CS} = X_{SL}$, so that the

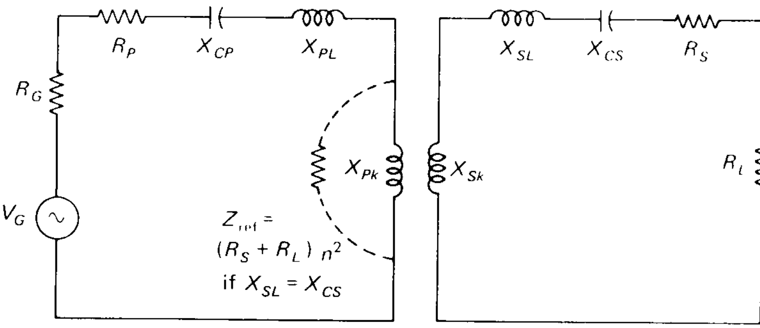


FIGURE 5-20 *Series-tuned rf transformer: capacitors cancel leakage reactances, giving effectively unity coupling.*

impedance reflected across X_{Pk} is purely resistive; $R_{ref} = (R_S + R_L)/n^2$ to be exact. The parallel combination of R_{ref} and X_{Pk} has a series equivalent (call it R_s and X_s) and it is the sum of X_s and X_{PL} which should be canceled by X_{CP} to achieve maximum current to the secondary and maximum power to the load. Smaller values of secondary capacitance will result in a net capacitive reactance in series with the reflected R_L appearing across X_{Pk}, which will raise the effective reactance and require a smaller tuning capacitance X_{CP}. Likewise, larger C_S will require larger C_P values. All of this could be analyzed by methods similar to Example 5-7, but it is probably not worth the effort, since C_P and C_S are generally variable and easily optimized.

The overall Q and hence the bandwidth of the circuit of Fig. 5-20 depends upon the ratios of reactance (X_C or X_L) to resistance in the primary and secondary circuits. If k is large, Q is approximately the average of primary and secondary Q. If k is small, the separate response curves of the primary and secondary are multiplied, resulting in an overall response curve that is narrower at the top than a single response curve, and drops more quickly to zero.

Note that if R_L is higher than X_{SL}, series secondary tuning will provide little improvement in selectivity or power output. Similarly, if R_G is higher than X_{PL}, series primary tuning will not be necessary.

Parallel Primary Tuning: If the generator resistance is higher than the primary inductive reactance, it will be advantageous to parallel-resonate this reactance with a capacitor across the primary so that the generator sees a higher-impedance load. This technique will be especially effective if k is low or if R_L is high, because in either of these cases the primary impedance will be largely inductive, with only a small resistive component.

Figure 5-21 shows an example of parallel primary tuning used to effect an impedance match between an 8.1-kΩ source and a 200-Ω load with a simple 1:1-ratio transformer having a k of 20%. The −800-Ω secondary capacitance cancels the secondary leakage reactance, placing the 200-Ω resistive load directly

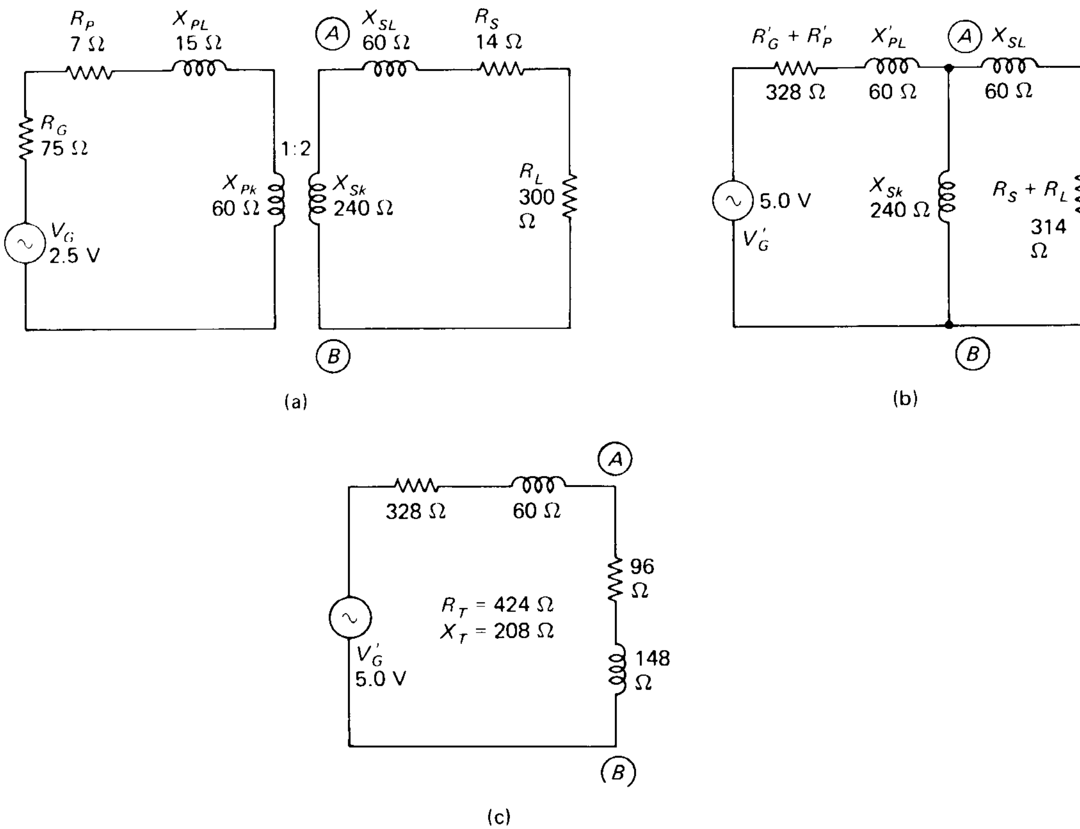


FIGURE 5-21 *Parallel-primary series-secondary tuning prevents loading of high-impedance source and allows match of an 8100-Ω source to a 200-Ω load with a 1:1 transformer: (a) equivalent circuit; (b) referenced to primary with R_{ref} and X_{pk} converted to series equivalent; (c) simplified parallel equivalent.*

across the secondary. The 1:1 ratio reflects this 200-Ω resistance across the 200-Ω primary reactance, a parallel combination that is equivalent to 100-Ω X_L and 100-Ω R in series. The series branch ABC in Fig. 5-21(b) is converted to its parallel equivalent in Fig. 5-21(c). The 900-Ω primary reactance is parallel-tuned to infinity by C_P, leaving only the parallel equivalent of the 100-Ω R_S, which (quite fortuitously) happens to equal the source impedance of 8100 Ω. Reference to equation 4-16 will verify this.

The voltage across the primary is thus $\frac{1}{2} V_G$, or 5 V. The voltage across points BC is found by phasor voltage division:

$$V_{BC} = V_{AC} \frac{\sqrt{100^2 + 100^2}}{\sqrt{900^2 + 100^2}} = \mathbf{0.78\ V}$$

This voltage is transformed undiminished to the load. Winding resistances have been neglected in this example.

Notice that the inductive reactance of the primary would be low enough to seriously load down the source were it not parallel-tuned by C_P. We are thus able to use fewer turns in the rf transformer coils if parallel tuning is employed, and this in turn cuts down on winding-resistance loss and cost.

Parallel Secondary Tuning: If the load resistance is very high and the coefficient of coupling is low, the secondary of an rf transformer may be resonated with a parallel capacitor to achieve an impressive gain in output voltage, even with a 1:1-ratio transformer. Figure 5-22 illustrates the technique. Although the term *parallel resonance* is used, notice from Fig. 5-22(b) and (c) that *series resonance* and the accompanying voltage rise across the capacitive reactance (Section 4.3) are actually involved.

An analysis of the circuit with the values given requires finding the Thévenin equivalent of the reflected V_G and 60 Ω as shunted by the 200-Ω X_L. Finding the

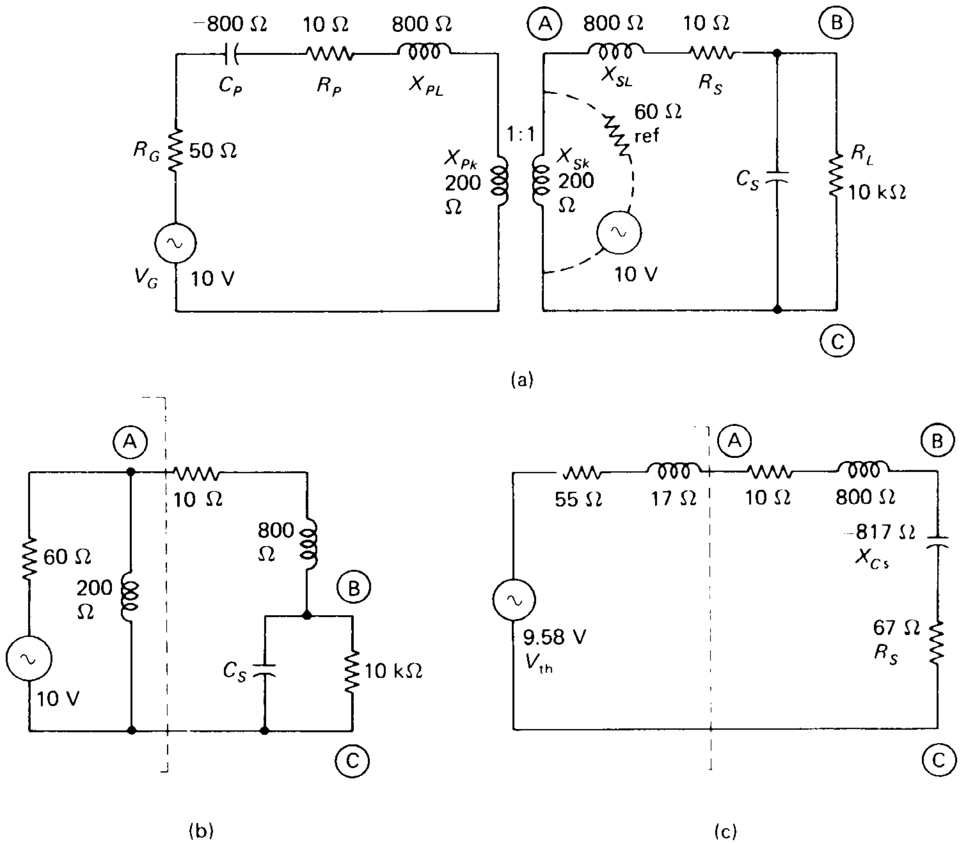


FIGURE 5-22 Series-primary parallel-secondary tuning. Secondary is actually series-resonant, giving voltage increase across output: (a) equivalent circuit; (b) referenced to secondary; (c) reduced to simple series equivalent.

series equivalent of C_S and finding the output V_{BC} becomes a simple matter:

$$V_{\text{th}} = 10\frac{200}{\sqrt{200^2 + 60^2}} = 9.58 \text{ V}$$

$$R_{\text{th}} = R_s = X_p\frac{X_p R_p}{X_p^2 + R_p^2} = 200\frac{200 \times 60}{200^2 + 60^2} = 55\ \Omega \tag{4-7}$$

$$X_{\text{th}} = X_s = R_p\frac{X_p R_p}{X_p^2 + R_p^2} = 60\frac{200 \times 60}{200^2 + 60^2} = 17\ \Omega \tag{4-8}$$

Using the approximation equations 4-15 and 4-16, we let C_S cancel the total inductive reactance of Fig. 5-22(c):

$$X_{Cp} \approx X_{Cs} = 817\ \Omega$$

$$R_s = \frac{X^2}{R_p} = \frac{817^2}{10{,}000} = 67\ \Omega$$

$$I = \frac{V}{R_T} = \frac{9.58}{132\ \Omega} = 0.0726 \text{ A}$$

$$V_{BC} = IZ \approx IX_C = 0.0726 \times 817 = \mathbf{59\ V}$$

The resonate voltage rise (nearly six times in this example) will be less if R_G is higher or if R_L is lower. Of course, very low R_G, very high R_L, and higher leakage inductance X_{SL} can produce a circuit with a Q, and accompanying voltage rise, of 100 or more.

Double Parallel-Tuned Transformer: If the generator and source resistance are both very high, an effective voltage transfer can be achieved with a loose-coupled transformer parallel tuned at both the primary and the secondary. This method of tuning is widely used to achieve selectivity in radio and TV receivers.

The primary and secondary tuning are interactive because of reflected reactances across the primary, as is true in the previously discussed double-tuned circuits. However, in this case the reflected reactances can cause a *double-peaking* effect which is not present in the other circuits. Neither of the peaks occurs at the actual resonant frequency [Fig. 5-23(a)]. The high-frequency peak [Fig. 5-23(b)] occurs when X_{SL} equals X_{CS} and a very low impedance is reflected back across X_{Pk}, leaving X_{PL} to resonate with X_{CP}. The low-frequency peak [Fig. 5-23(c)] occurs when the capacitive reactance reflected back across X_{Pk} raises the total primary reactance just enough to cause parallel resonance with X_{CP}. In practice, the low-frequency peak is always somewhat stronger because of the greater winding-resistance losses incurred with the low-impedance series-resonant condition of the high-frequency peak.

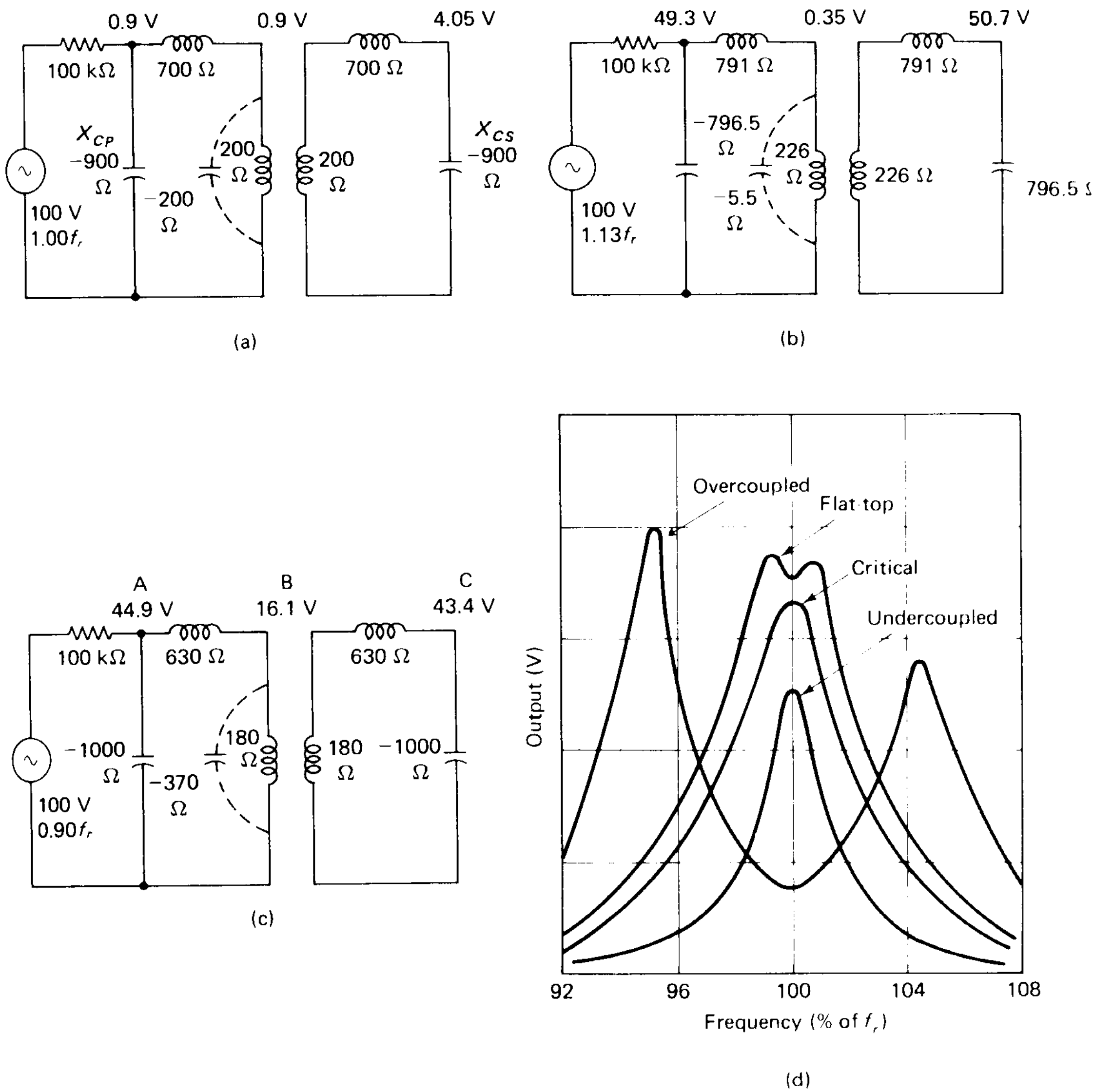


FIGURE 5-23 *Analysis of a double-parallel-tuned transformer: (a) at resonance; (b) above resonance with* X_{SL} *nearly equal* X_{CS}*; (c) below resonance, with reflected capacitive reactance causing primary reresonance; (d) typical response curves for the double-parallel-tuned transformer, showing effect of varying* k.

The calculations for Fig. 5-23(c) are reproduced below to demonstrate the method of analysis. Calculation at a number of other frequencies (an ideal job for a programmable calculator) produces the curve of Fig. 5-23(d). Analysis of the circuit of Fig. 5-23 has been simplified by assuming that R_L is infinite and R_W is zero. This results in an indeterminate division of zero by zero in the calculations at the exact peak points, so Fig. 5-23(b) and (c) show data at frequencies closely approaching the peak points.

For Fig. 5-23(c), the reactance looking out of the secondary is

$$X_S = -1000 + 630 = -370\ \Omega$$

This is reflected across X_{Pk}, giving the parallel result:

$$X_p = \frac{-370 \times 180}{-370 + 180} = +350.5\ \Omega$$

This inductive reactance adds to X_{PL}:

$$X_{L(\text{tot})} = 350.5 + 630 = +980.5\ \Omega$$

The parallel $X_{L(\text{tot})}$ is nearly resonant with X_{CP}:

$$X_L \| X_C = \frac{(980.5)(-1000)}{+980.5 - 1000} = +50.3\ \text{k}\Omega$$

The voltages are, by phasor voltage division:

$$V_A = V_G \frac{X}{\sqrt{X^2 + R^2}} = 100 \frac{50.3}{\sqrt{50.3^2 + 100^2}} = 44.9\ \text{V}$$

$$V_B = V_A \frac{X_{\text{par}}}{X_{L(\text{tot})}} = 44.9 \times \frac{350.5}{980.5} = 16.05\ \text{V}$$

$$V_C = V_B \frac{X_{CS}}{X_S} = 16.05 \times \frac{1000}{370} = 43.4\ \text{V}$$

The effect of these reflected reactances diminishes as coupling k is reduced and X_{Pk} becomes much smaller than X_{PL}. The peaks become closer together and the dip between them less severe. If k is made variable, a point can be found where the two peaks merge to form a fairly flat-topped response curve with steep-sloped sides. This curve is ideal for broadband signals such as TV or FM radio. At a still lower k value, a point is reached where only a single peak exists. The exact value of this *critical* coupling depends upon the Q of the primary and secondary circuits:

$$k_{\text{crit}} = \frac{1}{\sqrt{Q_P Q_S}} \tag{5-10}$$

$$Q_P = \frac{R_G \| (X_C^2 R_W)}{X_C} \tag{5-11}$$

$$Q_S = \frac{R_L \| (X_C^2 R_W)}{X_C} \tag{5-12}$$

Below critical coupling, voltage output drops off and bandwidth decreases. The Q (selectivity) of the total transformer at critical coupling is

$$Q_T = \frac{1}{k}$$

An examination of the foregoing equations will show that the highest output with sharpest tuning is obtained by making Q_P and Q_S as high as possible, and then adjusting k for critical coupling, or slightly more if flat-topped response is desired. Highest Q is obtained with a high L/C ratio, so the coils are often resonated with only their own stray capacitance. Typical values encountered in practice are $Q = 100$ and $k = 0.01$. This can mean a physical spacing between the coils that is several times their length.

Stagger tuning (tuning the primary to a slightly different frequency from the secondary) can also be used in an undercoupled circuit to achieve a flat-topped response curve. Overcoupling is to be preferred, however, because stagger tuning results in several times less output voltage.

5.6 PULSE TRANSFORMERS

Pulse transformers are used to handle rectangular rather than sinusoidal waveforms. Although they may provide voltage step-up or step-down, they are most often used to provide ground-reference isolation between the primary and secondary circuits. They are generally ferrite-core devices, and must have an inductance and saturation level appropriate to the period and voltage of the pulse they are intended to carry.

Inductance Requirement: Figure 5-24 shows a pulse-transformer circuit. Coefficient of coupling k and turns ratio n are both assumed to be unity. When the generator applies a voltage step, the current in L_P will build up and the voltage across L_P will fall off according to the familiar time-constant curves of Fig. 5-24(b). The voltage across L_S will be an exact reproduction of that across L_P. If this

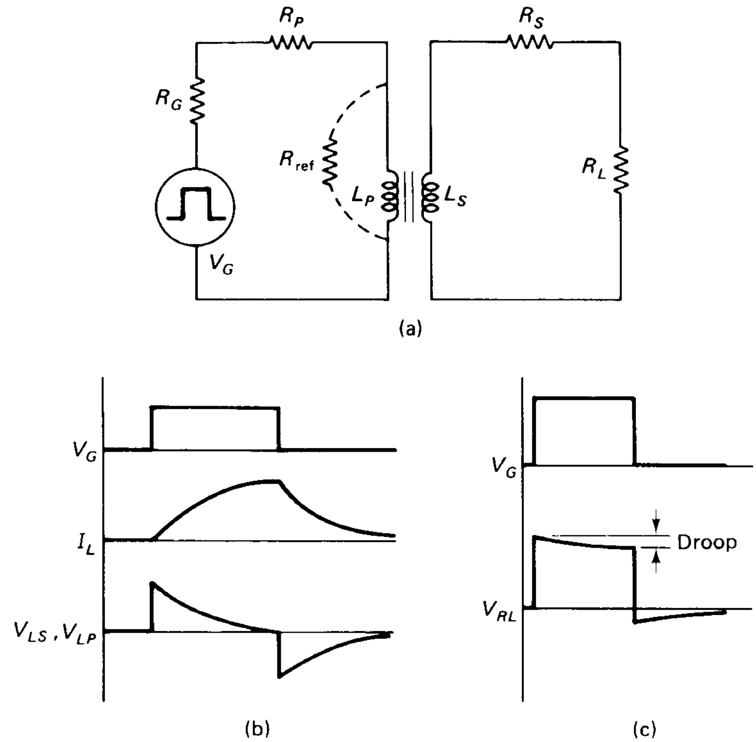


FIGURE 5-24 *(a) Pulse-transformer circuit. (b) Voltage and current waveforms in response to a pulse from V_G. (c) Droop becomes more severe if primary inductance is insufficient.*

voltage is to maintain a reasonably square shape, the active pulse time t (when the generator voltage is nonzero) must be much less than the time constant τ. If $t = 0.1\tau$, the output waveform will droop by about 9%. If $t = 0.01\tau$, the droop will be 1%. The resistance involved in this time-constant determination is the Thévenin equivalent driving L_P with R_S and R_L reflected to the primary. The primary inductance requirement is therefore:

$$L_P = 10tR \qquad \text{(for 9\% droop)} \tag{5-13}$$

$$L_P = 100tR \qquad \text{(for 1\% droop)} \tag{5-14}$$

where

$$R = (R_G + R_P) \| (R_S + R_L)/n^2 \tag{5-15}$$

Volt-Microsecond Limit: Once we have assured that the primary inductance is adequate for the pulse length, we must also be sure that the primary current does not become large enough to saturate the core. Notice from Fig. 5-24(b) that

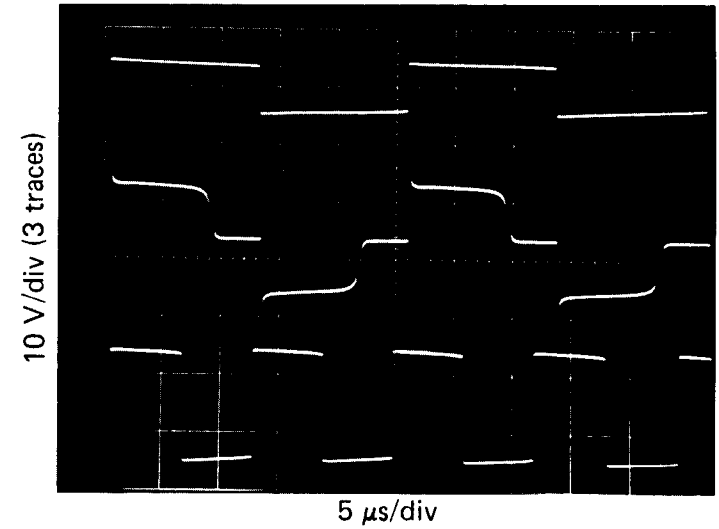


FIGURE 5-25 Low-level pulse passes through transformer (upper trace), but high-level pulse exceeds Vt *constant and is cut off (center trace). Reducing pulse period allows full pulse to pass without saturating core (lower trace).*

primary current increases with time while secondary current decreases. The secondary current does not cancel the primary current magnetizing force for square waves as it does for sine waves. This difference between primary and secondary current increases quite linearly with time during the first portion of the time constant, and the net magnetizing force may eventually become high enough to saturate the core. If the generator voltage is increased, these two currents, and hence the difference between them, will increase, resulting in earlier saturation. For any magnetic-core transformer the maximum product of $V_G t$ before saturation is a constant:

$$V_G t_{\text{pulse}} = constant \tag{5-16}$$

Figure 5-25, top trace, shows a long low-level pulse at the secondary of a pulse transformer. In the center trace the voltage has been increased, exceeding the Vt constant, and the last half of the pulse is missing. In the bottom trace the pulse period t has been reduced, bringing the Vt product back below the saturation level, and the stronger, shorter pulses are passed completely.

Transistor-Switched Pulse Transformer: The circuit of Fig. 5-26, or something like it, appears frequently enough to merit special attention, especially since its behavior is not quite as might at first be expected. In order to simplify the analysis we will make the following assumptions:

1. The transistor is a perfect switch: zero resistance when on, infinite resistance when off.
2. The transformer has unity coefficient of coupling and does not saturate.

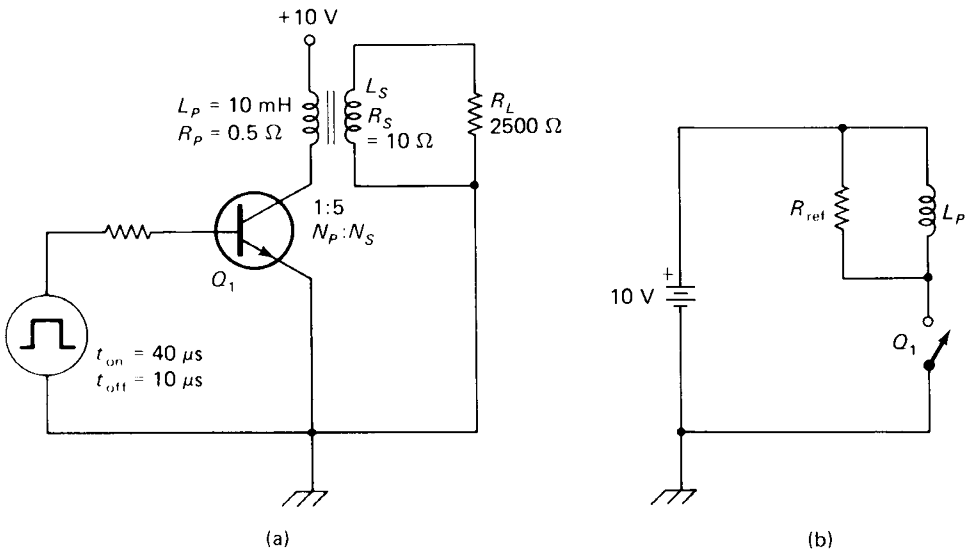


FIGURE 5-26 *(a) Transistor-switched pulse-transformer circuit. (b) Equivalent circuit with* R_L *reflected to primary.*

3. The transformer winding resistances are low enough to be negligible compared to R_L.
4. The charging time constant L_P/R_P is much longer than the *on* time set by the pulse generator.
5. The discharge time constant $L_S/(R_S + R_L)$ is much longer than the *off* time set by the pulse generator.

None of these assumptions is unreasonable except 5 and, in some cases, 3. The effect of high winding resistance (assumption 3) is simply a reduced load voltage. The amount of this effect can be predicted by simple voltage division. The effect of discharge occurring over an appreciable part of the time constant (assumption 5) can be estimated graphically, as will be shown later.

Using our assumptions, we draw the equivalent circuit of Fig. 5-26(b) and the graph of Fig. 5-27(a). Assumption 2 guarantees that the waveshape across R_{ref} (the primary) will be identical with that across R_L (the secondary.) Assumptions 4 and 5 guarantee that the inductor current will rise and fall linearly.

The voltage across L_P and R_{ref} is obviously V_{CC} during the *on* time. The fact that the average (or dc) voltage across an inductor must be zero demands that the voltage across R_{ref} during discharge is determined by the ratio of *on* to *off* time:

$$V_{CC}t_{on} = V_{pri}t_{off} \tag{5-17}$$

Graphically, this is illustrated in Fig. 5-27(a) by the equal areas A_1 and A_2.

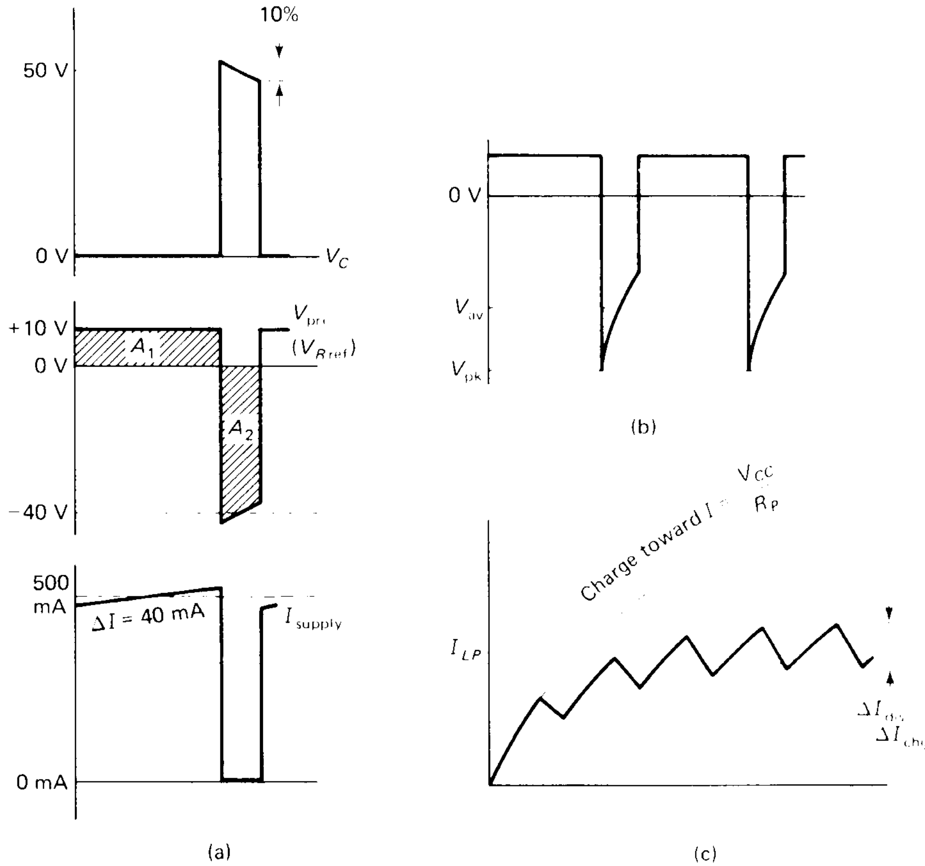


FIGURE 5-27 *(a) Waveforms for Fig. 5-26: collector voltage kicks up to 50 V during short off interval; primary voltage must average zero, making $A_1 = A_2$; collector current shows I during on interval and I = 0 during off interval. (b) Lower values of L_p produce spiked kickback voltages as L discharges more quickly. (c) Charging rate stays relatively constant at turn-on, while discharge rate increases from zero until equilibrium is reached.*

Qualitatively, this means that a long *on* time and a short *off* time allows L_P to build up a large current during *on*, which then produces a large kickback voltage during *off*. Quantitatively, the current change can be calculated from the charging conditions:

$$\Delta I_{chg} = \frac{V_{CC} t_{on}}{L_P} \tag{5-18}$$

The ΔI_{disch} during t_{off} must be the same at steady state.

The average value of I_{LP} can be determined in the following manner:

1. Calculate the average power in R_{ref} from V_{on} and V_{off}.
2. Calculate the average current from the supply using P_{av} and V_{CC}.

EXAMPLE 5-8

Calculate the load voltage and supply current for the pulse-transformer circuit of Fig. 5-26(a).

Solution

The load resistance is first reflected to the primary:

$$R_{ref} = \left(\frac{N_P}{N_S}\right)^2 R_L = \left(\frac{1}{5}\right)^2 2500 = 100\ \Omega$$

The 10-Ω R_S and 0.5-Ω R_P alter this value by less than 1% and are therefore neglected (assumption 3).

The charging and discharging time constants are found:

$$\tau_{chg} = \frac{L_P}{R_P} = \frac{10\text{ mH}}{0.5\ \Omega} = 20\text{ ms}$$

$$\tau_{dis} = \frac{L_P}{R_{ref}} = \frac{10\text{ mH}}{100\ \Omega} = 100\ \mu\text{s}$$

Assumptions 4 and 5 are valid, although τ_{dis} is only ten times t_{off} and some distortion of the output wave will be observed.

The voltage across R_{ref} during t_{on} is 10 V. The voltage during t_{off} is determined by the t_{on}/t_{off} ratio:

$$V_{pri} = V_{CC}\frac{t_{on}}{t_{off}} = 10\text{ V}\frac{40\ \mu\text{s}}{10\ \mu\text{s}} = 40\text{ V} \tag{5-17}$$

This is a kickback voltage produced by the inductor current decreasing from the value it attained during t_{on}. Its polarity is negative at the top in Fig. 5-26(b). The power in R_{ref} is thus $10^2/100 = 1$ W for the 40 μs of t_{on} and $40^2/100 = 16$ W for the 10 μs of t_{off}. The average power is then:

$$P_{av} = \frac{t_{on}P_{on} + t_{off}P_{off}}{t_{on} + t_{off}} = \frac{40 \times 1 + 10 \times 16}{40 + 10} = 4\text{ W}$$

Theoretically, there are no power-dissipating elements in the circuit except R_{ref}, so the average current from the supply must be:

$$I_{av} = \frac{P_{av}}{V_{CC}} = \frac{4}{10} = \mathbf{400\ mA}$$

In practice, we must expect I_{av} to be 10% to 50% greater than this calculation because of transistor switching-time and saturation limitations and transformer losses. Also, the supply current is zero during t_{off}, so the ideal t_{on} supply current is:

$$I_{on} = I_{av}\frac{t_{on} + t_{off}}{t_{on}} = 400\frac{40 + 10}{40} = 500\text{ mA}$$

The change of current (rise of I_P during t_{on}) is:

$$\Delta I_{chg} = \frac{V_{CC} t_{on}}{L_P} = \frac{10\ \mathrm{V} \times 40\ \mu\mathrm{s}}{10\ \mathrm{mH}} = 40\ \mathrm{mA} \qquad (5\text{-}16)$$

The load voltage V_{RL} is simply N_S/N_P times $V_{R\mathrm{ref}}$, namely 50 V positive and 200 V negative. Note from Fig. 5-27(a), however, that the negative voltage peaks approximately 5% greater than its 200-V average and droops by 10% during t_{off}. This is because t_{off} occupies about 10% of τ_{dis}. Figure 5-27(b) shows how V_{pk} will greatly exceed V_{av} during t_{off} if a small inductor, high R_L, or long t_{off} makes τ_{dis} shorter than t_{off}.

Figure 5-27(c) shows how I_{av} is approached over several cycles after initial turn-on. ΔI_{chg} is essentially the same each t_{on} period, but ΔI_{dis} increases as I_{LP} increases at each t_{off} period.

5.7 STRAY TRANSFORMER EFFECTS

There are a number of instances where the transformer effect is evident, although the physical structure involved is not a transformer in the usual sense.

Stray Coupling of Chokes and Transformers: The slight amount of leakage magnetism from an iron-core choke or transformer can induce an undesired voltage in a transformer or choke that is mounted nearby. In one test at 60 Hz, 10 V across a 0.5-H choke induced 50 mV across the open terminals of a second choke mounted adjacent to it. The following techniques can be used to combat this problem.

- Keep high- and low-level inductors away from each other on the chassis. Especially watch power-supply filter chokes and audio transformers.
- Mount chokes and transformers with the axes of their windings at right angles if interference is a potential problem. This is especially effective with RF air-core coils which have high leakage reactance.
- Do not operate low-level inductors with very high load impedances unless necessary. Load current will cause stray signals to be dropped across the high leakage reactance.
- Use magnetically shielded transformers.

Current Loops in Nearby Conductors: Shield cans, metal components, or the chassis, if within range of an inductor's magnetic field, will act like loops of wire, and currents will circulate in these little "secondary windings." The resulting I^2R loss will lower the Q of the coil. Air-core coils should therefore be kept several diameters away from metal objects if highest Q is desired.

Magnetic-Field Cancellation: Iron-core transformers sometimes have a heavy copper strap bound around the laminated core. Stray magnetic fields from the transformer's leakage inductance will induce a current in this loop, and this current will set up a field which tends to cancel the original stray field.

Variable Inductors called Variometers are made by setting up one coil which can be rotated within the other so that their magnetic fields either aid or oppose.

6

BATTERIES, FUSES, SWITCHES, AND LAMPS

6.1 PRIMARY BATTERIES

Battery Types: Batteries are portable devices for supplying direct current through chemical action. The simplest complete unit is called a *cell*, and will typically deliver a voltage between 1 and 2 V. Higher voltages are obtained by series connecting a number of cells, and the term *battery* more properly refers to a package of interconnected cells.

Cells are broadly classified as *primary*, in which the cell is discarded when the chemical action is complete, and *secondary*, in which the chemical action is reversible and the cell can be recharged by the application of reverse current from a battery charger.

Battery Ratings: Batteries are rated by their terminal voltage, their internal resistance, and their ampere-hour capacity. A simple measurement of open-circuit terminal voltage is not a good test of a battery, because even a nearly-dead battery will measure full voltage under no load. The internal resistance of a battery goes up sharply at discharge, however, limiting the maximum current it can supply. Batteries should therefore be tested for an acceptable output voltage (usually 60 to 80%) while they are delivering the current demanded by the specific application. A good C or D flashlight cell, for example, should deliver 0.5 A with a terminal voltage of 1.0 V or more.

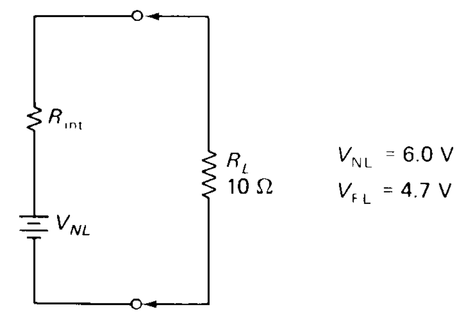


***FIGURE 6-1** A battery should be tested under load (Example 6-1).*

EXAMPLE 6-1

Figure 6-1 shows a battery-test setup. What voltage will this battery maintain with a load current of 150 mA?

Solution

First the internal resistance is calculated:

$$I = \frac{V_L}{R_L} = \frac{4.7}{10} = 0.47 \text{ A}$$

$$R_{\text{int}} = \frac{V}{I} = \frac{6.0 - 4.7}{0.47} = 2.8\ \Omega$$

Now the drop at 150 mA is found:

$$V_{\text{drop}} = IR = 0.15 \times 2.8 = 0.4 \text{ V}$$

$$V_{\text{out}} = V_{NL} - V_{\text{drop}} = 6.0 - 0.4 = \mathbf{5.6} \text{ V}$$

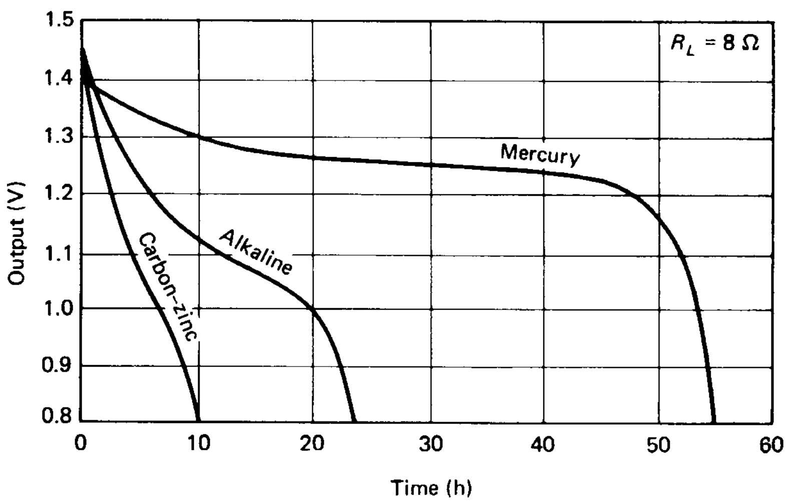
The ampere-hour rating is simply the product of the current drain times the life until discharge of a battery. If the current drain is kept low enough so that the battery life is a few hours or more, the product of amperes times hours is relatively constant. However, for heavy currents and fast discharges (under an hour) the ampere-hour rating decreases progressively, down to half or less of its rated value.

Primary-Battery Comparison: The commercially most important primary batteries are the carbon-zinc, the alkaline, and the mercury types. The table and graph of Fig. 6-2 show comparisons of the three types.

The carbon-zinc is the so-called dry cell (although the chemical filler is actually a moist paste). Its chief advantage is low cost and commercial availability. The alkaline battery is more expensive per cell, but it has a much higher ampere-hour capacity than the carbon-zinc cell and therefore it may be more cost-effective. It also has a lower internal resistance, making it more suitable for heavy-current loads. The mercury cell has an even higher capacity (and higher cost) and for many applications is the most cost-effective of the readily available primary cells. Its special advantages are long shelf life and relatively constant output voltage under light load.

Characteristic	Carbon-Zinc	Alkaline	Mercury
Voltage (V)	1.5	1.5	1.4
Internal resistance	0.3-0.5	0.05-0.2	0.25-10
Capacity (Wh/in^3)	2	6	7
Capacity (A·h/D cell)	0.6-3	3-10	4-14
Cost (cents/Wh)	3	2	3-7
Operation temperature (°F)	0 to +100	−40 to +120	−20 to +100

(a)

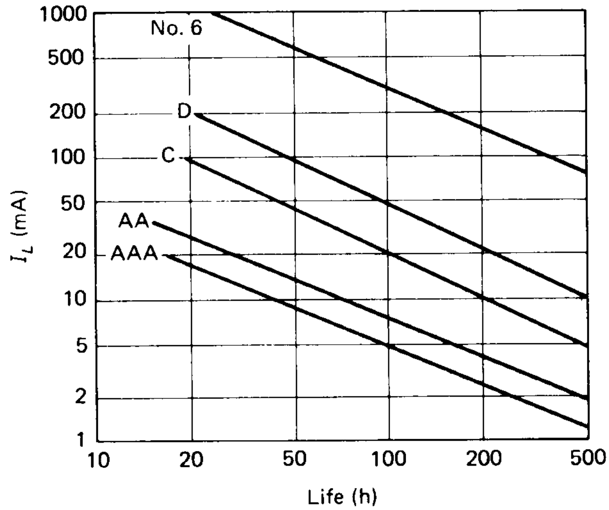


(b)

Cell Type	Volts, No Load	Weight, "C" cell	Capacity: I_L = 100 mA Continuous	Capacity: I_L = 50 mA 4 h/day
Carbon-zinc	1.5	40 g	1.5 A·h	3.3 A·h
Alkaline	1.5	60 g	4.0 A·h	4.9 A·h
Mercury	1.4	100 g	10.5 A·h	11.0 A·h

(c)

***FIGURE 6-2** Comparison of three popular primary battery types: (a) general characteristics; (b) continuous discharge curves for a C-size cell; (c) capacities for a C cell under continuous and intermittent loads. Note the carbon-zinc cell's improvement under intermittent load.*



(a)

COMMON DRY BATTERIES*

	AAA	AA	C	D	No. 6	Lantern	Transistor
Voltage	1.5	1.5	1.5	1.5	1.5	6	9
Ht (mm)	42.9	47.8	46.0	57.2	152.4	111	49.2
W/dia (mm)	9.9	13.5	23.9	31.8	63.6	66.7	26.2
D (mm)	–	–	–	–	–	66.7	17.5
NEDA**	24	15	14	13	905	908	1604
IEC	R-03	R-6	R-14	R-20	R-40	4R23	6F22
Eveready *	912 **E 92**	915 **E 91**	935 **E 93**	950 **E 95**	1S6	509	216 **522**
Burgess	7	Z	1	2	6	F4M	2U6
Mallory	MN2400	MN1500	MN1400	MN1300	M905	M908	MN1604
RCA	VS074	VS034	VS035	VS036	VS0065	VS040C	VS323
Mercury		X502 XX9		XX42			146X

*Boldface indicates alkaline.

***A* on NEDA number indicates alkaline

(b)

***FIGURE 6-3** (a) Life of various size carbon-zinc cells as a function of continuous load current. (b) Physical size and manufacturers' designations for popular primary batteries.*

Figure 6-3(a) shows the life that can be expected from various sizes of carbon-zinc cells at various load currents, and the accompanying table lists the physical sizes and manufacturers' numbers for the most common battery types.

Primary Battery Recharging: Recharging of carbon-zinc batteries is possible but should not be attempted on completely dead cells where the working voltage has dropped to less than 1.0 V per cell. Charging should be done over a 15-h period for 150% of the A·h capacity of the cell. A recharged carbon-zinc cell will have very poor shelf life and should be used immediately. Recharging should not be done more than twice.

6.2 SECONDARY BATTERIES

Lead-Acid Batteries: Lead-acid cells have long been used in automobiles and are still the most common choice where relatively large amounts of energy must be stored and rechargeability is required. The nominal output voltage of this cell is 2.0 V. Figure 6-4 shows a typical fluctuation in terminal voltage as the cell is discharged and then recharged. Lead-acid cells can be recharged typically 500 times from the full-discharge state, although this number increases sharply if recharge is initiated after discharge of 50% or less. The A·h capacity of a lead-acid battery decreases for very fast discharge rates. The following schedule is typical:

Discharge Time	*Percent Full A · h Capacity*
20 h	100
6 h	90
3 h	80
1 h	60
2 min	40

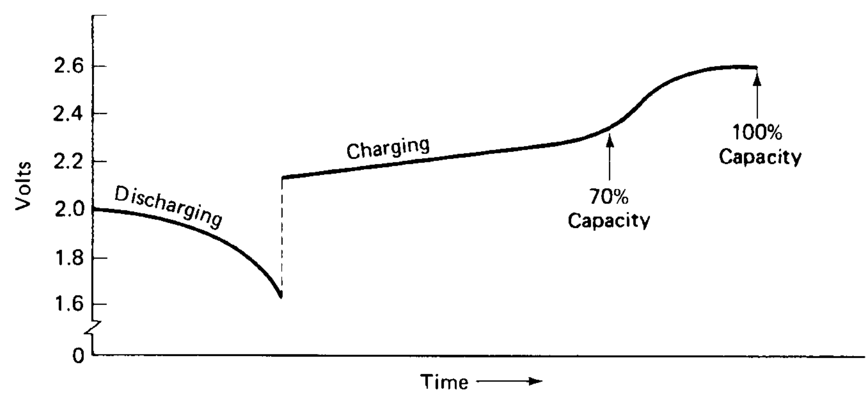


FIGURE 6-4 Typical discharge and recharge curve for a lead-acid cell. Discharge current is about twice charge current in this example.

As with most cells, the A·h capacity decreases at lower temperatures. For a lead-acid cell at 0°F the capacity of 60% of its value at 80°F.

Nickel-Cadmium Batteries: For lighter applications requiring rechargeability, the nickel-cadmium (ni-cad) battery has become popular. Its energy capacity (1.5 W·h/in.^{3}) is considerably less than any of the primary cells but is still better than the lead-acid cell (1 W·h/in.^{3}). The nominal output voltage of a ni-cad cell is 1.3 V. It displays a discharge–recharge curve similar to the lead-acid cell (Fig. 6-5). The A·h capacity of a ni-cad cell is somewhat lessened for fast discharges. For example, the capacity for a 1/2-h discharge is typically 80% of that for a 4-h discharge. These cells are normally recharged over a 12 to 15-h period, but may be recharged at any rate that does not result in cell temperatures above 49°C (120°F). Overcharging nonsealed cells will result in water loss from the electrolyte.

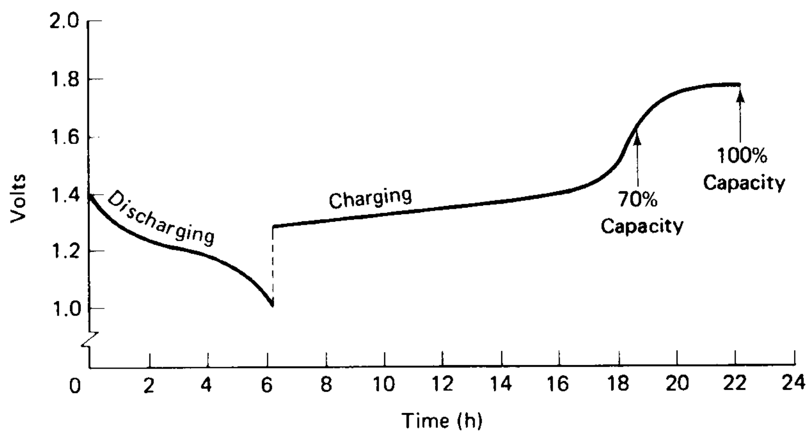


FIGURE 6-5 Discharge–recharge curve for a ni-cad cell. Charge current is about one-half discharge current.

Secondary Cell Charging: The terminal voltage of a secondary cell fluctuates by 20 to 40% during charge, and it is desirable to employ a charger that supplies a relatively constant current in spite of this. The usual approach is to design a charger with a source voltage perhaps double the nominal cell voltage, and limit the current by means of a resistor. The total ampere-hour input required to recharge a secondary cell is typically 150% of its rated A·h capacity. Trickle or maintenance chargers, delivering perhaps 5% of normal charging current, are often left connected to seldom-used devices (such as emergency lanterns) to assure full charge when needed. Ripple filters on ac-operated chargers are not absolutely essential, but adequate filtering will allow ripple-sensitive devices to operate directly from the charger and battery in parallel, which is often a convenience.

EXAMPLE 6-2

Design a charger for a 12-V, 60-A·h battery operating from the 60-Hz line. Charge time is to be 16 h. Include ripple filtering to 1 V p-p and a 5% trickle-charge function.

Solution

Figure 6-6 shows the charger. The current required is

$$I = \frac{60\ \text{A}\cdot\text{h}}{16\ \text{h}} \times 150\% = \mathbf{5.6\ A}$$

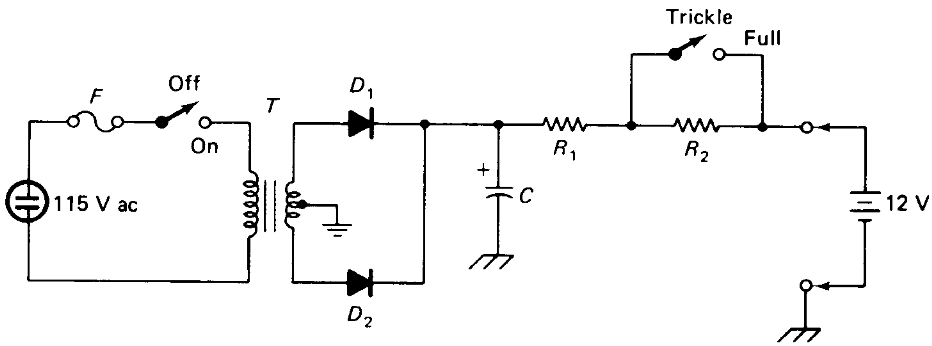


***FIGURE 6-6** Battery charger with filtering and trickle charge (Example 6-2).*

The voltage across C is chosen as $2V_{\text{battery}}$, or 24 V. The value and power dissipation of R_1 are calculated:

$$R_1 = \frac{V}{I} = \frac{24\ \text{V} - 12\ \text{V}}{5.6\ \text{A}} = \mathbf{2.1\ \Omega}$$

$$P_{R_1} = IV = 5.6 \times 12 = \mathbf{67\ W}$$

Additional resistance to limit the charge to 5% is supplied by R_2:

$$R_1 + R_2 = \frac{V}{I} = \frac{12\ \text{V}}{0.28\ \text{A}} = 42.9\ \Omega$$

$$R_2 = \mathbf{40.8\ \Omega}$$

$$P_{R_2} = I^2R = 0.28^2 \times 40.8 = \mathbf{3.2\ W}$$

The capacitor value can be calculated assuming a constant rate of discharge between the charging pulses, which are $\frac{1}{120}$s or 8.3 ms apart.

$$Q = CV$$

$$= It$$

$$CV = It$$

$$C = \frac{It}{V}\ \frac{5.6\ \text{A} \times 8.3\ \text{ms}}{1\ \text{V}} = \mathbf{46\ mF}$$

This is about \$30 worth of capacitor and should make the designer think seriously about some alternatives—electronic filtering or a pi-section choke filter perhaps.

The transformer must deliver a secondary voltage of 25 V *peak* on each side of the center tap to supply 24 V to charge the capacitor (the extra volt is for the drop across the silicon diode). Let us assume that the available transformers have a winding resistance ($\frac{1}{2}$ secondary plus reflected primary) of 0.2 Ω. Using the techniques of Section 5.3 and Fig. 5-10, the no-load secondary voltage required is calculated:

$$R_L = \frac{V}{I} = \frac{24}{5.6} = 4.29\ \Omega$$

$$\frac{R_L}{R_W} = \frac{4.29}{0.2} = 22$$

$$I_{S(\text{pk})} = 4I_L = 4 \times 5.6 = 22.4\ \text{A} \quad [\text{Fig 5-10(a)}]$$

$$V_{\text{drop}(\frac{1}{2}S)} = I_{S(\text{pk})}R_W = 22.4 \times 0.2 = 4.5\ \text{V}$$

$$V_{NL(\frac{1}{2}S)} = 25\ \text{V} + 4.5\ \text{V} = 29.5\ \text{V pk}$$

$$V_{NL(S)} = 2\ V_{pk} \times 0.707 = \mathbf{41.7\ V\ rms\ CT}$$

Next, the secondary current requirement is determined:

$$I_{\frac{1}{2}S(\text{rms})} = 1.5I_L = 1.5 \times 5.6 = 8.4\ \text{A} \quad [\text{Fig. 5-10 (b)}]$$

$$I_S = \tfrac{1}{2} \times 8.4 = \mathbf{4.2\ A\ rms}$$

The diodes should have a rather high surge current rating if it is decided to use the 46-mF filter, because the initial surge charging current on each half-cycle, when the supply is first turned on, will be limited only by the winding resistance:

$$I_{\text{surge}} = \frac{V_{\frac{1}{2}}}{R_W} = \frac{20.9}{0.2} = \mathbf{104\ A\ each\ half}$$

The time duration of this charging surge current is estimated as

$$\tau = R_WC = 0.2 \times 46\ \text{mF} = \mathbf{9.2\ ms}$$

6.3 PROPERTIES OF FUSES

A fuse is a low-melting-point conductor which is placed in series with a line and is designed to break the circuit by melting away when a specified current is exceeded. The cartridge fuses generally used in electronics equipment are usually marked with three items of information as in the following examples:

3 AG	$\frac{1}{4}$ A	250 V
3 AB	10 A	250 V
MDL	$2\frac{1}{2}$	125 V

The first set of characters refers to the manufacturers' case size and case material. The A recalls that the cartridge fuse was originally designed for automotive applications; the G refers to a glass case and the B to a Bakelite, fiber, or ceramic case. The second number is the maximum continuous current in amperes that the fuse can handle without blowing. The third set of characters indicates the voltage that the fuse can safely withstand between its two ends without arcing after it has blown, in the face of a prospective short-circuit current of 10,000 A. In electronic equipment the prospective short-circuit currents are usually limited to a few amperes or less by transformers or house fuses, so it is common (and recommended by fuse manufacturers) to use 125-V-rated fuses to protect secondary circuits handling 500 V or more.

Slow, Normal, and Fast Fuses: Cartridge fuses are available in normal, slow-blow, and fast-blow (instrument) types. Figure 6-7 shows how "slo-blo" fuses can withstand overloads up to five times their rating for a few seconds but blow on a 120% overload of extended duration. Slow-blow fuses can usually be recognized by their coiled internal construction. They are about twice as expensive as

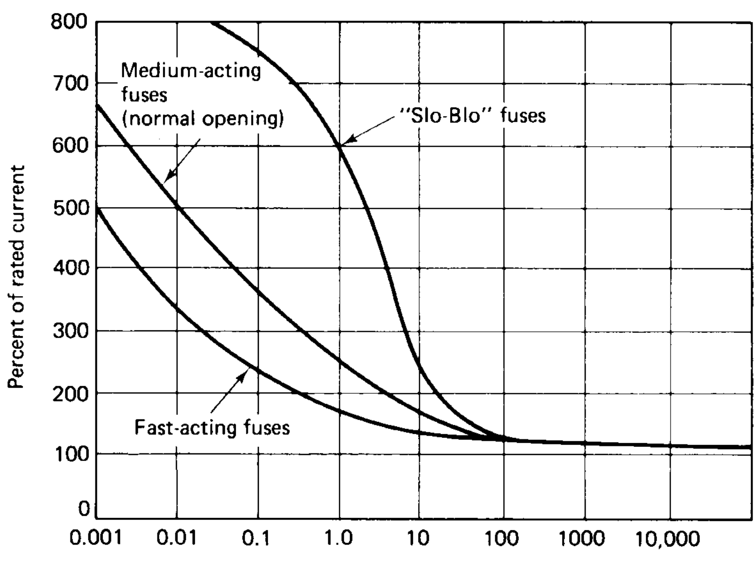


FIGURE 6-7 Typical response times of slow, normal, and fast-blowing fuses. (Courtesy Littlefuse, Inc.)

normal-acting fuses but are used extensively in electronic equipment because of the current surges which are encountered when the power is first turned on and the filter capacitors charge up.

Figure 6-8 shows the current inrush when a small solid-state lab oscilloscope is turned on. Notice that the peak surge current is about three times the steady-state current, which is not reached for about 12 cycles, or 0.2 s. A normal-acting fuse can withstand a 300% overload for about 0.3 s, but this is cutting it pretty close. A slow-blow fuse could withstand such an overload for about 6 s, and would be more likely to stand up to the abuses of high ambient temperatures and idiots who keep flicking the power on and off.

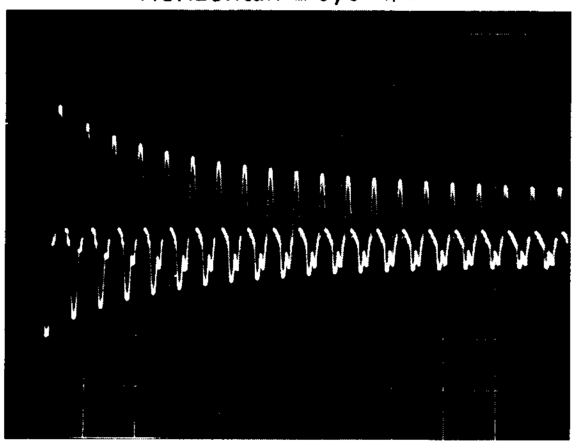


FIGURE 6-8 Line-current inrush due to filter-capacitor charging when an instrument is first turned on.

Fast-acting fuses which blow in as little as 1 ms under a 500% overload are available (at somewhat higher cost) for the protection of meter movements, solid-state diodes, and other semiconductors which can be destroyed in a few milliseconds by overcurrents. Notice from Fig. 6-7 that even these require nearly 0.5 s to react to a 200% overload. For the ultimate in fast overcurrent protection, all-electronic (no-moving-part) circuit breakers can be designed to trip in a few microseconds for even a 110% overload.

Fuse Selection: Manufacturers generally recommend that a fuse be operated at no more than 75% of its rated current to avoid unnecessary shutdowns. Ambient temperature should also be considered when selecting fuse ratings. Figure 6-9 gives one manufacturer's recommendations.

Remember that fuse ratings are in amperes rms, not amperes average. For an instrument whose entire current drain is a highly filtered rectifier circuit, it is quite likely that the true rms line current will be two or three times the value indicated on an average-reading ammeter. The article on pulsed load currents in Section 5.3 explains this in detail. See, in particular, Fig. 5-10(b).

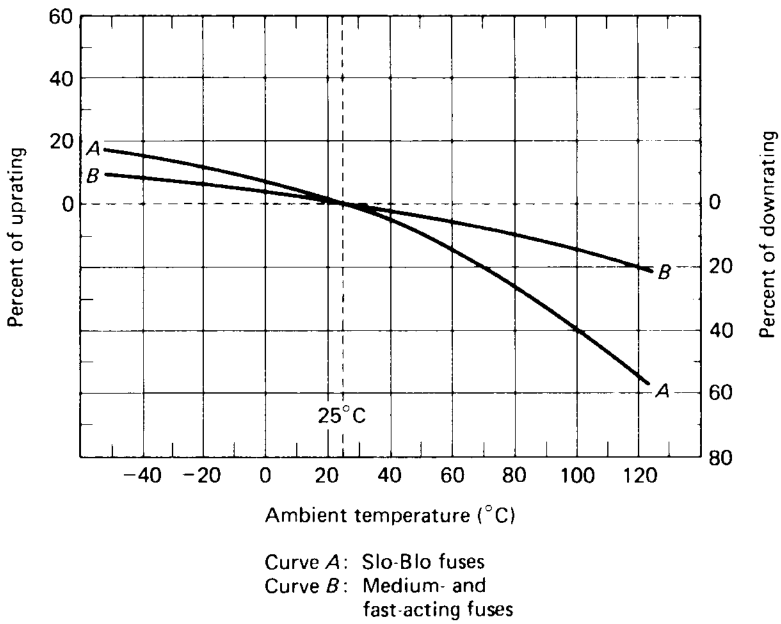


FIGURE 6-9 Effect of ambient temperature on required fuse ratings. (Courtesy Littlefuse, Inc.)

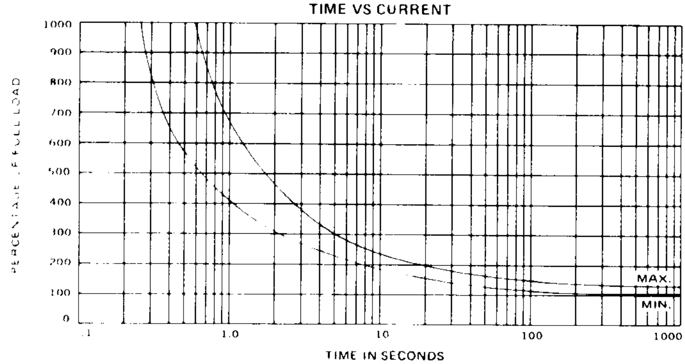
Fuse Resistance: The resistance of a fuse element generally causes a voltage drop of a tenth to several tenths of a volt at full current, and the fuse case may become noticeably warm, especially with high-current fuses. Slow-blow fuses generally have slightly higher resistances than do quick-blowing types. Some typical resistances for electronic fuses are as follows:

$\frac{1}{16}$ A	fast . . .	4.3 Ω
$\frac{1}{4}$ A	fast . . .	3.2 Ω
1 A	fast . . .	0.2 Ω
1 A	slow . . .	0.3 Ω
3 A	slow . . .	0.1 Ω

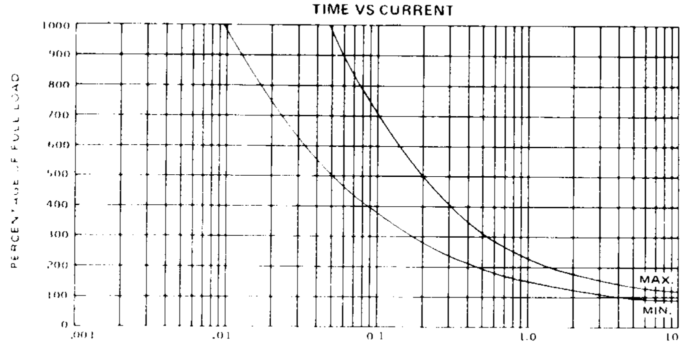
These resistances remain quite constant from very low to maximum current.

Indicating Fuses: Fuses are available with a spring-loaded pin which protrudes from one end of the case upon separation of the fuse element to provide a visible indication of fuse failure. A low-current lamp rated for the supply voltage can be connected across a fuse to provide a similar function. (An NE-2 lamp in series with a 100-kΩ, $\frac{1}{2}$-W, resistor is fine for 110-V lines.) Such indicators are valuable where several fuses are located together and identification of the failed device might otherwise be difficult.

Thermal-Type Breakers: Circuit breakers perform the same basic function as fuses, but they can be reset rather than replaced after they are tripped by an overcurrent. There are two basic types: thermal and electromagnetic. In the thermal types, excessive current heats a bimetallic strip, which bends, releasing a trip lever and opening the contacts. Thermal breakers are at a disadvantage in environments where temperature is not controlled because their trip point varies with ambient temperature. Figure 6-10(a) shows trip time versus current for a typical thermal breaker. Figure 6-10(b) shows a similar curve for a fast-acting "heater wire" type of thermal breaker. Note that a 500% overload may still require 0.2 s to trip this relatively fast thermal breaker.

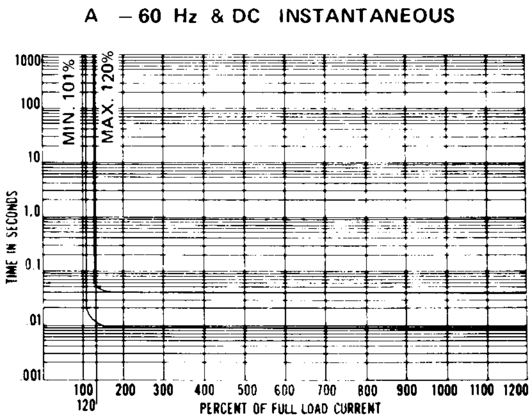


(a)

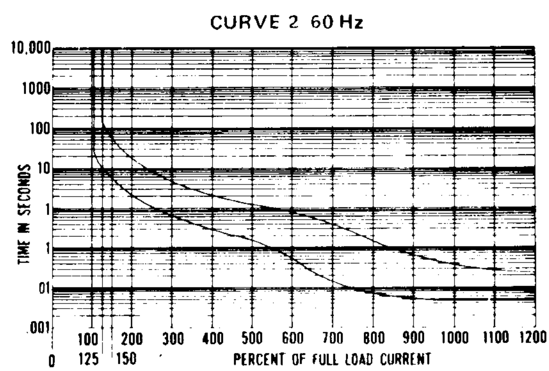


(b)

***FIGURE 6-10** Thermally operated circuit breakers (a) have a response time similar to slow-blow fuses. Hot-wire type breakers respond about as fast as quick-blow fuses (b).*



(c)



(d)

FIGURE 6-10 Magnetic breakers (c) are very fast. Magnetic-hydraulic breakers (d) are slow for moderate overloads but very fast for extreme overloads. (Courtesy Potter & Brumfield division of AMF Corp.)

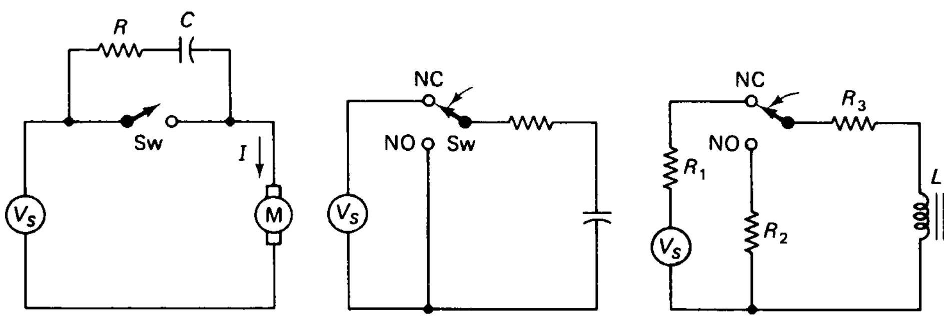
Magnetic and Hydraulic Breakers: Magnetic circuit breakers utilize the magnetic field created by overcurrent in a coil of wire to attract an armature which trips the release mechanism. As shown in Fig. 6-10(c), such a breaker can respond in about 10 ms. Magnetic-hydraulic breakers have two trip mechanisms. The first is identical to the armature described above but is designed to trip only on extreme overloads of about 800%. The second is activated by a piston which is pulled slowly through a cylinder of hydraulic fluid by lesser overloads. Figure 6-10(d) shows the time versus current curve for a magnetic-hydraulic breaker. Such a curve is ideal for many applications because a 200% load (such as might be encountered in the normal starting of a motor) can be tolerated for a few seconds, but a really catastrophic overload (such as a short circuit) will blow the breaker in a few milliseconds.

6.5 PROPERTIES OF SWITCHES AND RELAYS

In selecting switches and relays, the following considerations should be made, in addition to the number of poles and positions required:

1. Current and voltage rating of the contacts
2. Shorting or nonshorting contacts
3. Contact bounce
4. Actuate time
5. Life span

Current and Voltage Ratings: Relay and switch contacts are generally rated for resistive loads at dc or 60-Hz ac. When switching inductive loads (transformers, magnet coils, motors, etc.), the contacts should be operated very conservatively (i.e., 2-A or 5-A contacts to carry 1 A to an inductive load). This is because the inductor will produce a large kickback voltage when the switch opens, thereby causing arcing and subsequent pitting of the contacts. A suppressor circuit can be wired across the switch to minimize this problem, as depicted in Fig. 6-11(a). In the formula for C, τ is the time required for the contacts to open sufficiently to prevent arcing. This may be estimated as 1 ms for small relays and toggle switches and 10 ms for heavy power contacts. If the source is ac, C may have to be lowered considerably to raise its reactance to the point where current leakage through it is no longer objectionable.



$$R \approx 2\frac{V_S}{I} \qquad C \approx \frac{\tau}{R}$$

(a)

Nonshorting required

(b)

Shorting required

(c)

FIGURE 6-11 (a) An RC snubber reduces switch arcing from inductive kickback. (b) In some circuits a make-before-break switch would short the supply or a capacitor during transition. (c) In other circuits a break-before-make switch would interrupt an inductor current.

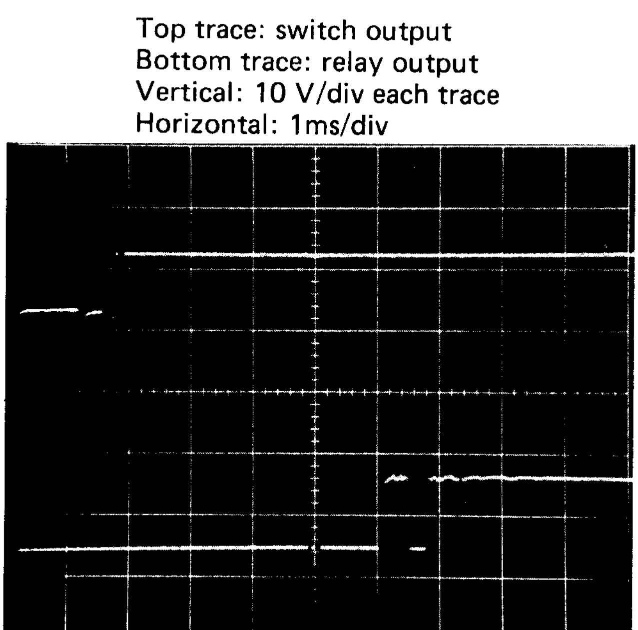
Shorting versus Nonshorting: Most toggle switches and relays have *nonshorting* contacts. This means that the armature (common) pole leaves the normally closed pole before it touches the normally open pole, and vice versa. This is also called *break-before-make* or *form C* contacts. However, there are switches (especially rotary switches) whose contacts *make-before-break*, and these are called *shorting* switches or *form D* contacts. Often the difference between the two is not important, but Fig. 6-11(b) and (c) indicate how it can be critical.

In Fig. 6-11(b), if the NC (normally closed) and NO (normally open) poles are shorted for even an instant during switching, the source will be shorted, with catastrophic results. In Fig. 6-11(c), if the *C* pole is left floating for an instant during its transition from NC to NO, the inductor current will be interrupted and the resulting kickback voltage will arc the switch, resulting in pitted contacts.

Contact Bounce and Actuate Time: Switches and relays do not actuate instantly, nor do they always make contact with a single clean motion. The following table indicates the behavior that can be expected from various types of contactors.

Switch Type	*Actuate Time (ms)*	*Bounce?*
Toggle switch	2–10	Yes
Rotary switch	2–20	No
Coil and armature relay	5–15	Yes
Reed relay	0.5–1	No

Figure 6-12 shows scope traces of contact bounce for a toggle switch and a coil-and-armature relay.



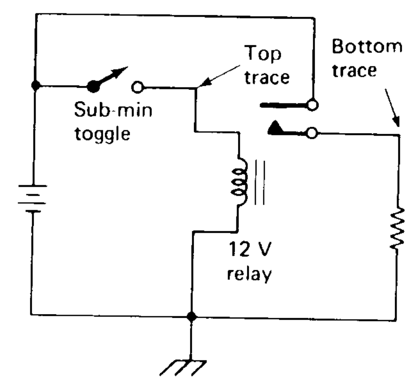


***FIGURE 6-12** Photographic evidence of switch bounce, relay closure time, and relay-contact bounce. A new high-quality toggle switch and small "telephone-type" relay were used for this test.*

Life Expectancy: Good-quality rotary and toggle switches generally guarantee a life of 50,000 to 100,000 mechanical cycles. Inexpensive switches have a life span much shorter than this. Such limitations cannot be ignored when designing equipment for industrial or military use. Consider the case of a rotary switch in a piece of test equipment on a factory assembly line which is operated once every minute, 8 h/day, 5 days/week, for 1 year. The total is 124,800 operations! The cost in downtime and engineering time when that switch fails could easily exceed $1000. The false economy of using a $5 switch instead of a $35 switch in such an application is quite obvious.

6.6 PROPERTIES OF LAMPS

Incandescent Lamps produce light by using an electric current to heat a tungsten filament to a temperature in the vicinity of 2000 K. The great majority of energy produced, however, is not visible light but infrared (radiant heat), as documented in Fig. 6-13(a). Typical efficiencies (visible light energy/electrical input energy) for various types of light sources are as follows:

Neon glow	< 1%
Incandescent	1.5–3%
Mercury	≈6%
Fluorescent	7–12%
High-pressure sodium	16–22%

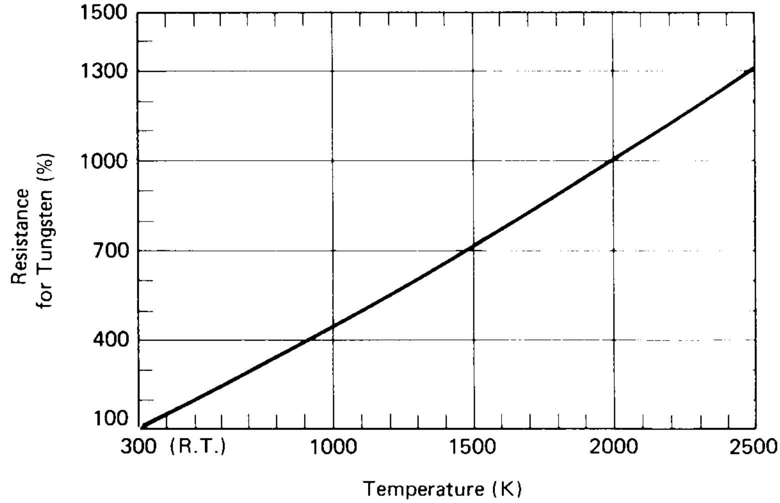
Typical energy distribution of some of the more popular subminiature incandescent lamps

Electrical Characteristics						Energy Emission Characteristics (%)		
Lamp No.	Volts	Amps	MSCP	Avg. Life (h)	Approx. Filament Temperature Kelvin	Ultra-violet: 0.3–0.4 μm	Visible: 0.4–0.7 μm	Infrared* Radiation: 0.7 μm and Up
680	5.0	0.060	0.03	100,000+	1850	0.0005	0.4095	99.59
683	5.0	0.060	0.05	100,000	1950	0.001	0.649	99.35
713	5.0	0.075	0.088	25,000	2100	0.003	1.157	98.84
715	5.0	0.115	0.15	40,000	2125	0.004	1.246	98.75
328	5.0	0.18	0.34	3,000	2275	0.010	1.94	98.05
327	28.0	0.040	0.34	7,000	2200	0.006	1.554	98.44

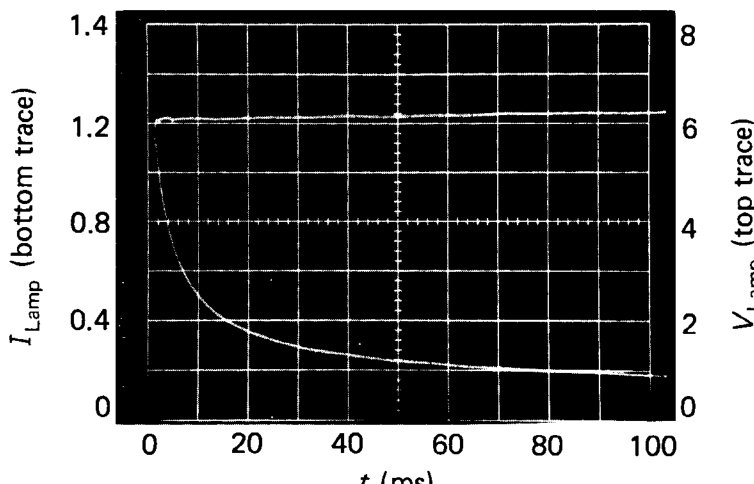
*Includes heat loss by conduction

(a)

FIGURE 6-13 (a) A very small percentage of an incandescent lamp's energy appears as visible light.



(b)



(c)

FIGURE 6-13 (CONTINUED) (b) Resistance of a tungsten filament increases by about ×10 from cold to hot filament. (Courtesy Chicago Miniature Lamp Works.) (c) Current inrush for a No. 47 pilot lamp on a 6.3-V dc source.

Inrush Current: A tungsten filament at 2000 K has a resistance that is about 10 times its room-temperature resistance, as shown in Fig. 6-13(b). It follows that the current surge when a cold lamp is first turned on will be nearly 10 times the final operating current. This inrush phenomenon [illustrated in Fig. 6-13(c)] has implications for lamp life and circuit fusing. If a system can be devised to turn the lamp on slowly, the thermal shock due to inrush current will be reduced and life will be increased. Also, if inrush can be avoided, fast-acting low-threshold circuit breakers can be employed for more complete circuit protection.

Flashover at Burnout: Incandescent lamps usually fail from thermally induced stress just after the power is switched on. Everyone has witnessed the brilliant flash of light which an incandescent lamp makes when the filament parts. This is due to ionization of residual gases, which starts at the point of rupture and spreads to the space between the filament supports. Ionization current in 50- to 150-W 120-V lamps commonly reaches 100 to 200 A and lasts from 1 to 5 ms.

This flashover current can trip a circuit breaker or fuse, causing unnecessary shutdown when one lamp of a parallel group fails. In lamp circuits containing diode, SCR, or triac series pass elements, flashover current can destroy the semiconductor.

A small resistance in series with the lamp is the simplest way of limiting flashover current. For example, 10 Ω in series with a 60-W 120-V lamp will dissipate only 2.5 W and drop the line voltage by only 5 V, yet it will limit flashover current to 12 A, which should be within the surge-current rating of a 1-A semiconductor. As shown under *lamp rerating* below, this will result in a 14% decrease in light output and a 67% increase in lamp life.

Lamp Deterioration: Several modes of deterioration are evident in most incandescent lamps. Filament movement is common in coil-filament lamps and is caused by relaxation of stresses in the filament and its mounting with time and temperature. The result may be a current increase on the order of 10% and a brightness increase on the order of 20% after 1000 h.

Filament evaporation thins down the filament wire, increasing its resistance. The result can be a 10% current reduction and a 35% brightness reduction after 1000 h. The products of a filament evaporation can be deposited on the inner glass surface, causing lamp darkening, which may reduce lamp brightness by several percent.

Lamp Rerating: Manufacturers' specifications for incandescent lamp light output (mean spherical candlepower; MSCP), rated life, and filament current are all given for some arbitrarily selected filament voltage. If it is necessary to increase lamp life by a factor of 3, this can be done simply by reducing the lamp voltage by a certain amount. Of course, the output (MSCP) and current drain will be reduced by this change. Three formulas for rerating incandescent lamps at other than specified voltages are:

$$\text{new output (MSCP)} = \left(\frac{V_{\text{new}}}{V_{\text{rated}}}\right)^{3.5} \times \text{rated MSCP} \tag{6-1}$$

$$\text{new life} = \left(\frac{V_{\text{rated}}}{V_{\text{new}}}\right)^{12} \times \text{rated life} \tag{6-2}$$

$$\text{new current} = \left(\frac{V_{\text{new}}}{V_{\text{rated}}}\right)^{0.55} \times \text{rated current} \tag{6-3}$$

EXAMPLE 6-3

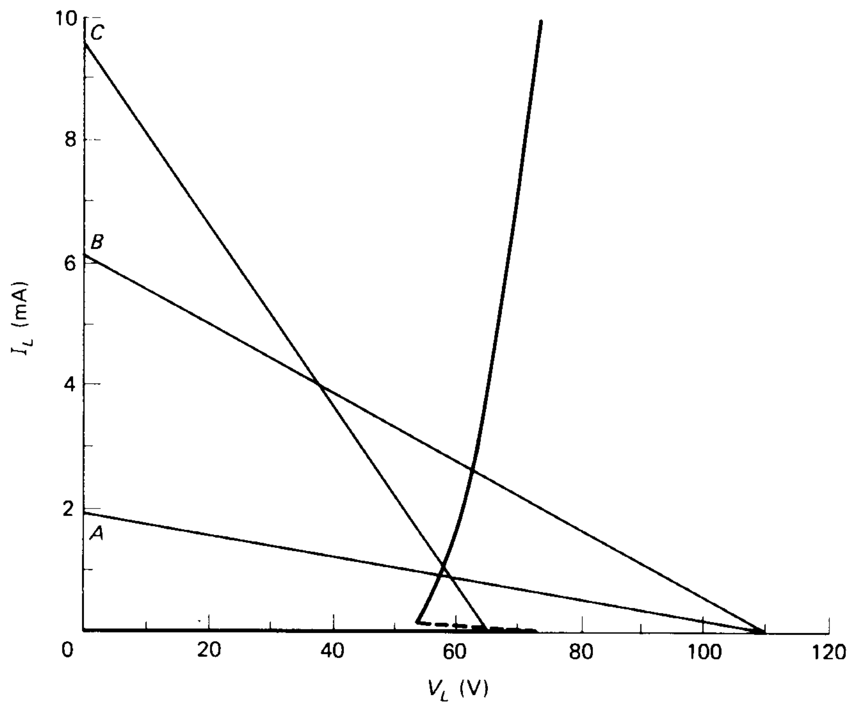
A type-327 lamp [see Fig. 6-13(a)] is to be operated at 24 V rather than 28 V. What will be its new otuput, life, and current?

Solution

$$\text{new MSCP} = \left(\frac{24}{28}\right)^{3.5} \times 0.34 = \mathbf{0.20} \tag{6-1}$$

$$\text{new life} = \left(\frac{28}{24}\right)^{12} \times 7000 = \mathbf{44{,}000\ H} \tag{6-2}$$

$$\text{new current} = \left(\frac{24}{28}\right)^{0.55} \times 0.040 = \mathbf{0.037\ A} \tag{6-3}$$



$$\text{Curve } A\text{: } I = \frac{V}{R} = \frac{110\ \text{V}}{56\ \text{k}\Omega} = 1.96\ \text{mA}$$

$$\text{Curve } B\text{: } I = \frac{V}{R} = \frac{110\ \text{V}}{18\ \text{k}\Omega} = 6.11\ \text{mA}$$

$$\text{Curve } C\text{: } I = \frac{V}{R} = \frac{65\ \text{V}}{6.8\ \text{k}\Omega} = 9.56\ \text{mA}$$

FIGURE 6-14 *Characteristics for an NE-2 neon lamp with several load lines. Line C will not ignite the lamp but will hold it in conduction once it is turned on.*

Notice that a small change in lamp voltage (−14%) causes a severe drop in output (−41%) but a dramatic improvement in lamp life (×6.3).

Gas Discharge Lamps: Certain gases can have one of the outer electrons ripped loose from their atoms by the application of a moderate voltage. The neon glow lamp makes use of this ionization phenomenon. Note from Fig. 6-14 that no current at all flows in this lamp until the firing voltage is reached, whereupon an excessively large current would flow if not limited by an external resistor. Once ignited, the lamp operates at a voltage typically 25 to 50% below the firing voltage. The voltage in the operating range is relatively constant with current change, leading to the occasional use of neon lamps as voltage regulators and peak-voltage limiters. Neon glow lamps are available with nominal firing voltages from 65 to 150 V, although actual voltages may vary by ±25% from unit to unit and by several volts with age and temperature for the same unit. The life span for neon lamps is not unlimited, being typically in the range of 25,000 h at full rated current. Figure 6-15 gives the characteristics of the more popular neon and incandescent lamps used for electronic applications.

Lamp	Volts	Amperes	Base	Lamp	Volts	Amperes	Base
PR2	2.38	0.50	Flange	55	7.0	0.41	Bayonet
PR12	5.95	0.50	Flange	222	2.25	0.25	Screw
14	2.47	0.30	Screw	313	28	0.17	Bayonet
40	6.3	0.15	Screw	1490	3.2	0.16	Bayonet
41	2.3	0.50	Screw	1815	14.0	0.20	Min bay
43	2.5	0.50	Bayonet	1819	28	0.10	Min bay
44	6.3	0.25	Bayonet	1847	6.3	0.15	Min bay
45	3.2	0.35	Bayonet	1891	14.0	0.23	Min bay
46	6.3	0.25	Screw	NE 2	90*	0.003	Wire
47	6.3	0.15	Bayonet	NE2H	135*	0.002	Wire
48	2.0	0.06	Screw	NE45	90*	0.002	Cand scr
49	2.0	0.06	Bayonet	NE48	90*	0.002	2-button bay
51	7.5	0.22	Bayonet				
53	14.4	0.12	Bayonet	NE51	90*	0.0003	Min bay

*Maximum dc firing voltage.

FIGURE 6-15 Characteristics of popular lamps used in electronic devices.

7

DIODE TYPES AND CHARACTERISTICS

7.1 SEMICONDUCTOR MATERIALS AND JUNCTIONS

A diode is a two-terminal device whose current response to an applied voltage is different for the two possible directions of applied voltage. The original vacuum-tube diodes were used almost exclusively as ac-to-dc converters, but today's semiconductor diodes have properties and varieties which invite a much wider range of applications. A few moments spent in studying the nature of semiconductor materials may bring some order to the bewildering range of characteristics exhibited by the different diode types, since the internal structure is similar for most of them.

Intrinsic Semiconductors: The electronic properties of a material are determined primarily by the outer layer of electrons (called the valence shell) of its atoms. Conductors have many mobile electrons which are free to flow from atom to atom, while insulators have a place for each electron, and each electron remains in its place. Semiconductors have four electrons in the valence shell, although there is room for eight. This allows the atoms to pack together with four neighbors around each one, sharing electrons so that each inner atom has all eight spaces filled. Figure 7-1 gives a two-dimensional representation of this crystal-lattice structure, although a. box packed with marbles would give a better idea of the actual three-dimensional form.

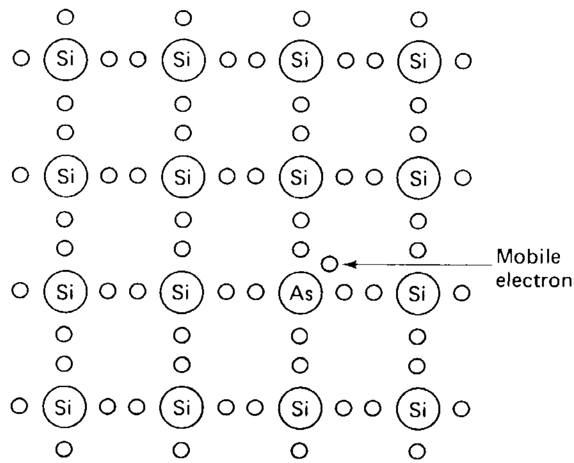


FIGURE 7-1 Representation of a silicon crystal with N-type impurity (mobile electrons).

Intrinsic (perfectly pure) semiconductor material is nonconductive at low temperatures, since each electron is held in a fixed space. Heat, which is essentially vibration of the atoms, can cause some of the valence electrons to break out of the lattice, thus rendering the materials conductive.

***N*-doped Semiconductors:** Notice that the silicon crystal lattice represented in Fig. 7-1 is not perfectly pure. One of the silicon atoms has been replaced by arsenic, which has five electrons in its valence shell. There is no fixed place in the crystal lattice for this electron. It is mobile, and is termed a negative charge carrier. We say that the silicon is *N*-doped, rather than intrinsic.

Two points should be noted concerning this impurity doping. First, the introduction of negative *charge carriers* does not mean that the material is given a negative *charge*, since the arsenic atoms bring an extra proton for each extra electron. Second, impurity concentrations of a few parts per million are all that is required to transform the material from an insulator to a fairly good conductor.

***P*-doped Semiconductors:** An equally effective method for rendering a crystal lattice conductive is *P* doping, in which an impurity such as boron, which has three electrons in its valence shell, is added. The lattice now has a vacant spot or hole, as shown in Fig. 7-2. If a positive voltage were applied at the right side of this lattice, an electron from the second column may be attracted to fill in the third column, leaving a new hole in the second column. The hole has thus moved away from the positive voltage, as a positively charged particle would. We term the hole a positive charge carrier.

***P*-*N*-Junction Behavior:** When *P*- and *N*-type materials are joined, a basic diode is formed. With no applied voltage, the charge carriers are distributed essentially as

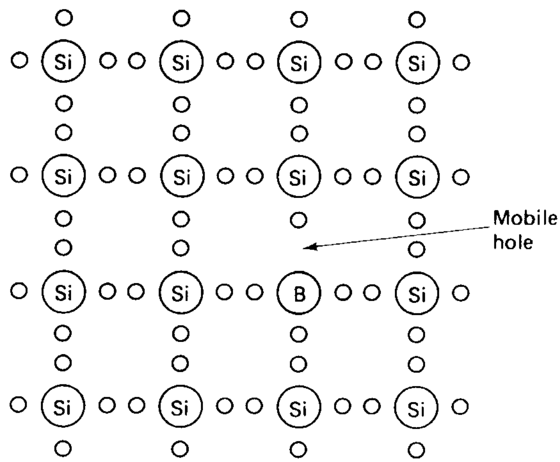


FIGURE 7-2 Representation of a silicon crystal with P-type impurities (mobile holes).

shown in Fig. 7-3(a). Notice that only the mobile charge carriers are shown. The lattice electrons and the inner atoms are not shown because they are immobile.

If an external voltage is applied as shown in Fig. 7-3(b), the charge carriers will be attracted toward the potential of opposite sign, leaving the region around the *P-N* junction depleted of all charge carriers. This depletion zone is insulative and no current can flow in this direction. The diode is said to be reverse-biased.

The opposite voltage polarity (positive to *P*, negative to *N*) results in charge carriers being pulled across the junction, and current flows easily, as depicted in Fig. 7-3(c). The diode is now said to be forward-biased.

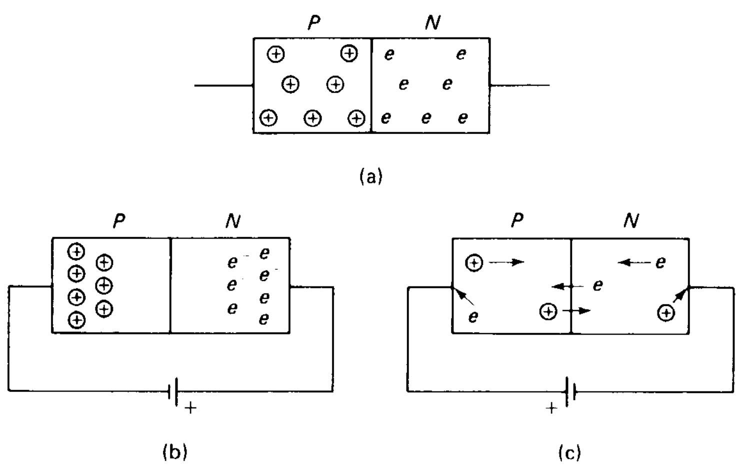


FIGURE 7-3 (a) P-N junction or semiconductor diode. (b) Reverse bias forms an insulative depletion zone around the junction. (c) Forward bias attracts charge carriers across the junction.

7.2 FORWARD-VOLTAGE CHARACTERISTICS

Threshold Voltage: We have seen that a semiconductor diode will not conduct under reverse bias but will conduct readily under forward bias. It must not be assumed that any arbitrarily small forward voltage makes the diode appear as a perfect short circuit, however. There is a tendency for the electron charge carriers near the junction to migrate across to fill the holes on the other side. To the extent that this occurs, a negative charge appears on the *P* side, with a corresponding positive charge on the *N* side of the junction. These *minority* charge carriers produce a potential across the junction which must be overcome by the forward bias before current can begin to flow. This turn-on or *threshold* voltage is quite different for different materials, as shown in Fig. 7-4(a). For silicon diodes, which are by far the most common type, the forward current rises by an approximate factor of 10 for each 80-mV increase in voltage. This produces a linear curve on semilog paper, as in Fig. 7-4(b). For a typical 1-A silicon diode, the following table provides a good estimate of V versus I:

V_{forward} (V)	I_{forward}
0.44	1 μA
0.52	10 μA
0.60	100 μA
0.68	1 mA
0.76	10 mA
0.84	100 mA
0.92	1 A

Small-signal diodes may show 0.1 V greater drop, and heavy-current rectifiers may show 0.1 V less. In addition, there will be a small amount of ordinary resistance in series with the diode because of the semiconductor material and connecting-lead resistivity. This resistance may reach 1 Ω or so for small-signal diodes, but is much smaller for power diodes.

Variations in Forward Voltage: Diodes with identical type numbers may commonly show voltage drops which are different by 0.1 V. Put another way, identical-type diodes wired in parallel may carry forward currents that differ by a factor of 10. Parallel connection of diodes to increase current capacity is therefore not recommended. Matched diodes, which were formed side by side from the same wafer of silicon, are available for applications where it is absolutely necessary that the forward voltage-versus-current curves track. The diodes in commercially available bridge-rectifier packs track typically to better than 10 mV.

Temperature Variations: The forward voltage across a silicon diode may be expected to decrease by typically 0.1 V as operating temperature rises from 25°C to 100°C. For diode applications requiring closely tracking voltages, matched diodes physically mounted on a common heat sink should be employed.

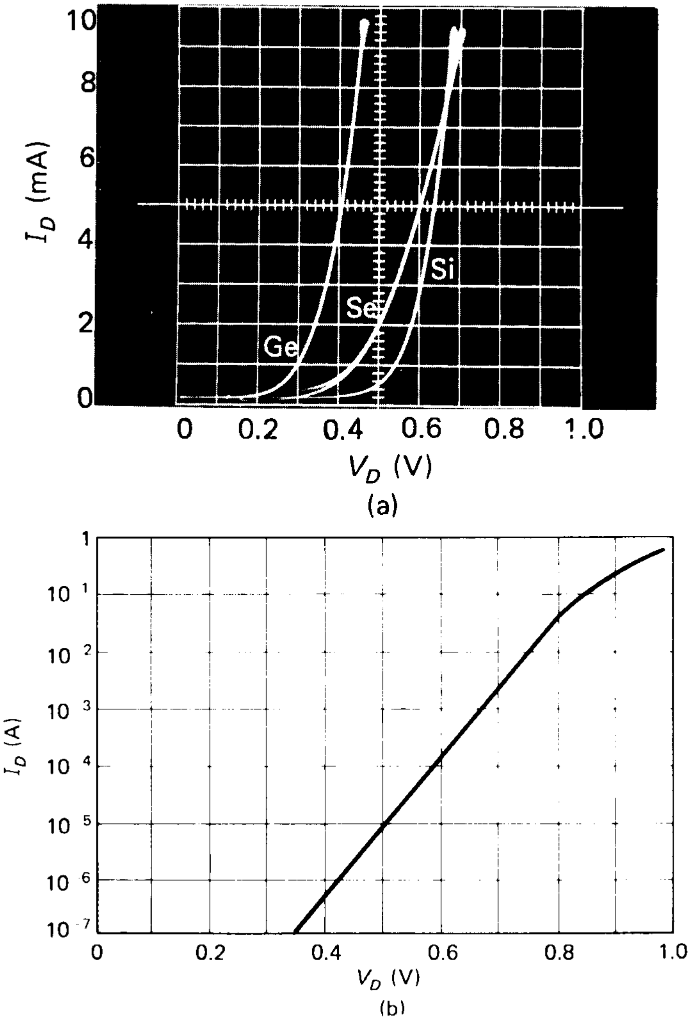


FIGURE 7-4 *(a) Threshold voltage differs for different diode materials. (b) Straight-line curve for silicon diode on semilog paper.*

Diode Current Ratings: Diodes are generally rated for their maximum permissible forward current (average half-wave rectified to a resistive load, which is nearly the same as the dc rating). This limitation is imposed by the power product, $P = IV$, where V is the forward-voltage drop, and is a function of the physical size and heat sinking of the diode. The basic rating applies at or below the specified free-air temperature, and must be reduced according to a chart or derating formula if higher operating temperatures are anticipated. As a rule of thumb when manufacturers' ratings are unknown, lead-mounted diodes should be limited to case temperatures of 100°C, and stud-mounted types should be limited to 150°C. Paints and stick-on labels are available which melt or change color at specific temperatures to test case temperatures. Lead-mounted diodes can be kept slightly cooler if their leads are kept short (2 or 3 mm) and soldered to a relatively large-area binding post or copper pad.

Surge Currents: Diodes are capable of handling extremely large nonrepetitive pulses of current if their duration is too short to allow the diode junction to overheat. Surge current ratings 10 to 20 times I_{av} are typical for 8.3-ms pulses (half-cycle on a 60-Hz line), increasing to 50 to 100 times I_{av} for 2-ms pulses. Such pulses are commonly encountered when a power supply is turned on and the filter capacitors charge up from zero. The amplitude of the pulse can be calculated from the effective transformer resistance.

EXAMPLE 7-1

Determine the peak surge current for the diode in Fig. 7-5. Also estimate the number of cycles until the current settles to normal.

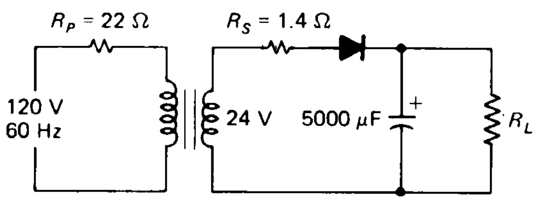


FIGURE 7-5 Surge current is limited only by transformer winding resistance at turn-on (Example 7-1).

Solution

The primary resistance reflected to the secondary is

$$R_{ref} = n^2 R_P = \left(\tfrac{1}{5}\right)^2 \times 22 = 0.9\ \Omega$$

The total charging resistance, peak voltage, and peak current are

$$R_T = R_{ref} + R_S = 0.9 + 1.4 = 2.3\ \Omega$$

$$V_{pk} = 24 \times 1.41 = 34\ \text{V pk}$$

$$I_{pk} = \frac{V_{pk}}{R_T} = \frac{34}{2.3} = \mathbf{15\ A}$$

The charging time constant gives an estimate of surge-time requirements.

$$\tau = RC = 2.3\ \Omega \times 5000\ \mu\text{F} = 11.5\ \text{ms}$$

This is considerably more than one half-cycle, so the manufacturer's chart would be consulted to find a diode with a 15-A surge rating for 2 cycles' duration.

7.3 REVERSE-VOLTAGE CHARACTERISTICS

Leakage Current: Silicon diodes are remarkably insulative in the reverse direction, but some leakage current does flow, as shown in Fig. 7-6. Notice that the leakage increases markedly with temperature.

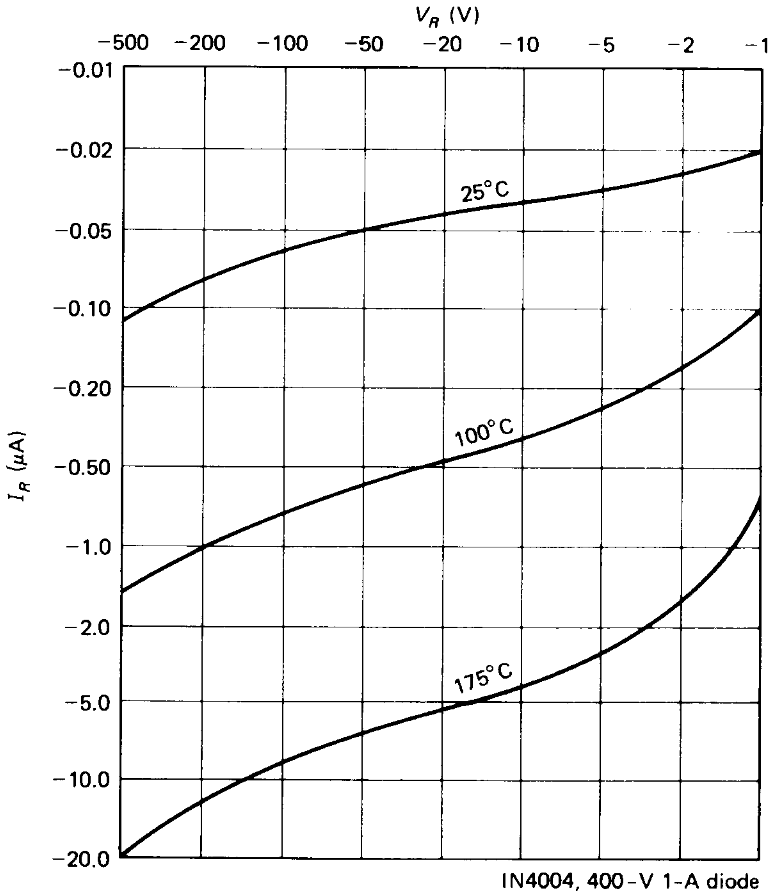


FIGURE 7-6 *Diode reverse leakage current increases at high temperatures.*

Reverse-Voltage Breakdown: If the reverse voltage across a silicon diode is increased, the charge carriers for the leakage current will be accelerated to higher velocities as they cross the junction. Some of these charge carriers will collide with atoms in the crystal lattice, and eventually a voltage will be reached which gives them enough energy to knock electrons loose from the valence band of the parent atom. The resulting hole–electron pair contributes to greater leakage and may itself be accelerated sufficiently to dislodge more electrons, and so on in an avalanche of conduction, as shown in Fig. 7-7.

This avalanche condition is not in itself destructive, but two factors dictate that extreme care be taken to limit the current that is permitted to flow in the reverse direction. First, the current necessary to produce destructive temperatures is much less at the relatively high avalanche voltages than at low forward voltages. A diode carrying 1 A at 1 V forward dissipates 1 W, but at reverse avalanche of −500 V it need only carry 2 mA to dissipate the same power. Second, although it is

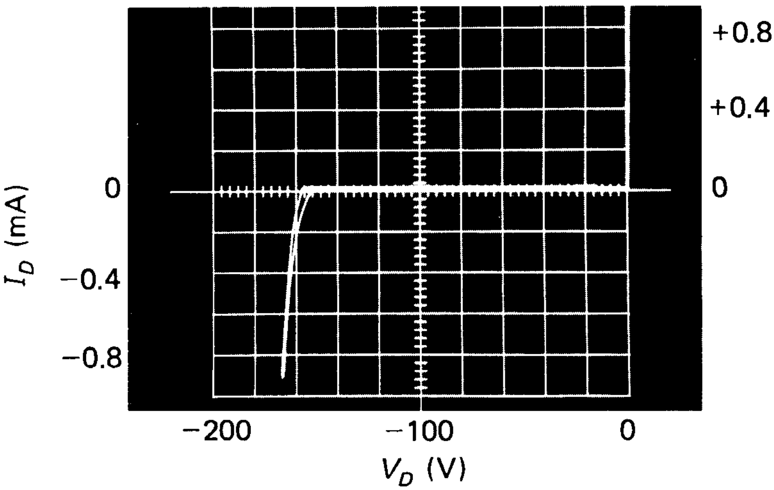


***FIGURE 7-7** Beyond the rated PRV (peak-reverse voltage) of a diode an avalanche point will be encountered at which the diode begins to conduct heavily.*

possible to control the doping of the *P-N* junction so that all portions go into avalanche at the same voltage, not all manufacturers do so. Diodes without controlled avalanche characteristics may experience hot spots which seriously reduce their power-handling capability in the reverse direction. As might be expected, high-current diodes with their large junction areas are more troubled by hot spots than low-current diodes.

Diodes in Series: Low-current diodes (1 A or less) operating from a 60-Hz sine-wave line may generally be connected in series to obtain higher peak-reverse-voltage (PRV) ratings. The calculated PRV should always be multiplied by a safety factor of at least 2, since the ac line is notorious for occasional transient spikes of high voltage. The diode having the highest back resistance will, of course, hog the first of the rising reverse voltage, but this is not a problem since it will simply go into nondestructive avalanche when its limit is reached, throwing the remaining voltage to the other diodes in the string.

With high-current diodes and with high frequency or square-wave input voltages, the fact that diodes cannot turn off instantly becomes important, and shunting capacitors are required to protect the fast diode in the string, as shown in the next section.

7.4 DYNAMIC CHARACTERISTICS

Reverse Recovery Time: A diode that has been conducting in the forward direction has free charge carriers all along its length, in a density directly proportional to the forward current. To become nonconducting, all these charge carriers must be cleared from the junction area, leaving a depletion zone. This takes time, and the diode is capable of conducting in the reverse direction until it is accomplished. Figure 7-8 shows a test circuit and the recovery times measured for several power and signal diodes. Figure 7-9 shows that the recovery times of two identical-type-number 3-A diodes may differ by more than 3 μs.

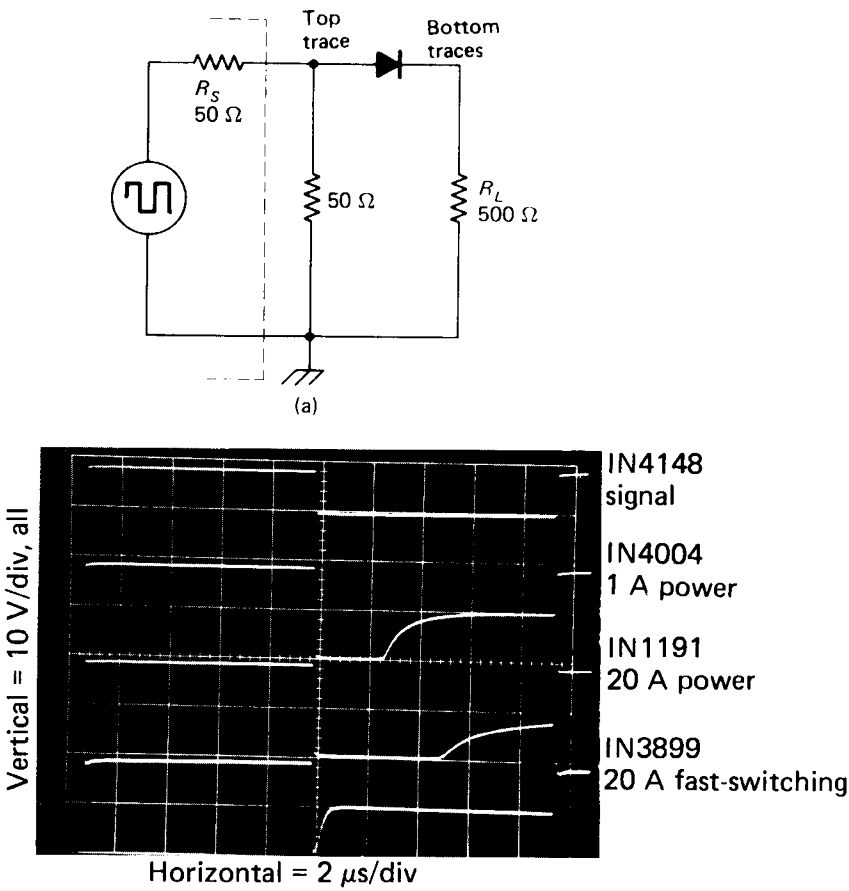


FIGURE 7-8 A forward-conducting diode requires a certain time t_s to return to the nonconducting state: (a) test circuit; (b) the storage time t_s is greater for higher-current diodes, although fast-recovery power diodes are available for a price.

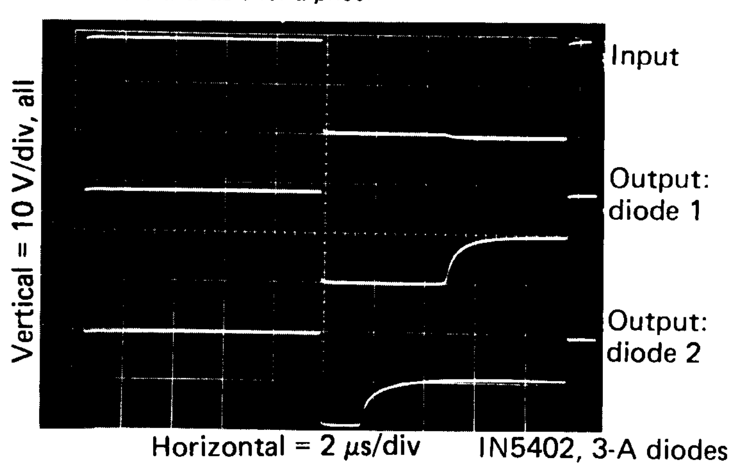


FIGURE 7-9 Identical diode types may have widely different storage times.

Fast Switching of Series Diodes: If one diode in a series string turns off 1 μs faster than any of the others, and if the applied voltage is a square wave, so that it rises to full reverse value instantly, the fast diode will go into avalanche and will experience a current limited only by the series resistance of the source. The resulting instantaneous diode power can be alarmingly high. For 60-Hz sine-wave rectification, the reverse voltage rise is gradual and the diodes have ample time to turn off, but for switching inverters that have rise times on the order of 1 μs and frequencies of several kHz, it may be necessary to slow the voltage rise across the fast diode with a capacitor, as shown in the following example.

EXAMPLE 7-2

In the circuit of Fig. 7-10, it is possible that the diodes may have reverse recovery times which differ by as much as 1 μs. The inverter provides a 1500 V p-p 60-Hz square wave through a series resistance (transformer secondary plus reflected primary) of 100 Ω. The diodes are rated at 1 A, 1.0 V forward, maximum. Determine whether shunt capacitors are needed for the diodes, and if so, calculate their value.

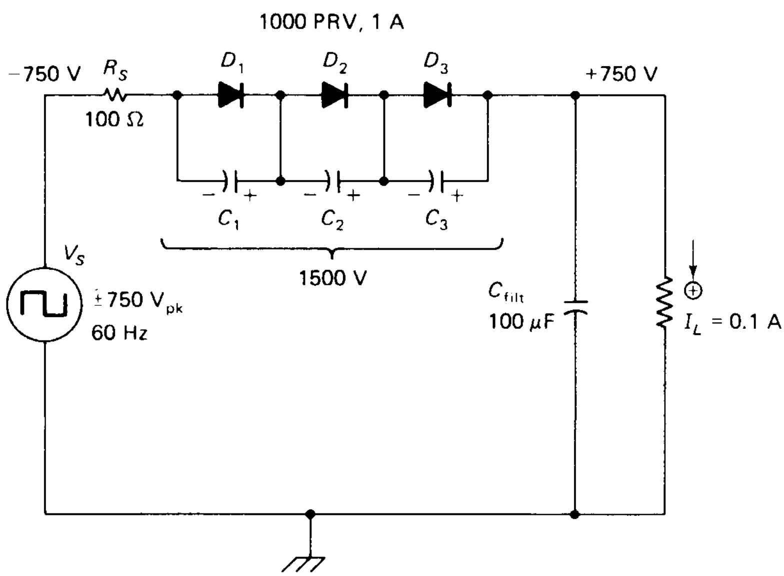


***FIGURE 7-10** Fast-switching high-voltage diodes in series may have capacitors (C_1–C_3) placed across them to give all diodes a chance to turn off before applying reverse voltage to them (Examples 7-2 and 7-3).*

Solution

Assume that the third diode has turned off while the first two remain on for 1 μs. The filter capacitor maintains +750 V at the right end of the diode string. With the source internally applying −750 V, we have a total of 1500 V across R_S, two turned-on diodes, and a diode in avalanche at 1000 V. The voltage and current for R_s and then the instantaneous power for D_3 are calculated:

$$V_{Rs} = V_{\text{applied}} - V_{D3} = 1500 - 1000 = 500 \text{ V}$$

$$I_{Rs} = I_{D3} = \frac{V_{Rs}}{R_s} = \frac{500}{100} = 5 \text{ A}$$

$$P_{D3(\text{inst})} = IV = 5 \times 1000 = 5000 \text{ W}$$

$$P_{D3(\text{av})} = \frac{1\ \mu\text{s}}{16.7 \text{ ms}} \times 5000 \text{ W} = \mathbf{0.3\ W}$$

The average power due to the forward current is about 0.1 A × 1.0 V or 0.1 W, so the 1-W limit of the diode is not being exceeded. However, in the absence of any assurances that the diodes possess controlled avalanche characteristics, we should anticipate hot spots and be reluctant to accept a safety factor less than 10. Determination of the shunt-capacitor value follows.

With D_1 and D_2 conducting, we will ensure that R_s does not charge C_3 to more than one-third of the reverse voltage across the diode string during the 1 μs before D_1 and D_2 turn on. This maintains the PRV safety factor of two. In general:

$$C = n\frac{\Delta t_s}{R}, \quad \text{where } n \text{ is the number of diodes and } \Delta t_s \text{ is the difference in storage times} \tag{7-1}$$

$$C_3 = 3\frac{1\ \mu\text{s}}{100} = 0.03\ \mu\text{F}$$

Of course, equal-value capacitors are normally used for C_1 and C_2 since, in a production situation, it is not known just which of the three diodes is the fastest. Note that these capacitors function as RC charging-rate limiters, and are not ac voltage dividers as they appear at first to be. Relatively broad-tolerance ceramic types are therefore quite acceptable.

Diode reverse-recovery characteristics also have consequences for the power transformer if the positive-to-negative voltage transition is faster than the diode turn-off time. For the few microseconds while the diodes are reverse-biased but still conductive, the transformer resistance R_s receives the sum of the filter capacitor voltage and the reverse transformer voltage. The average power dissipation in the transformer can be significant at high frequencies.

EXAMPLE 7-3

In the circuit of Fig. 7-10, the diodes may have reverse recovery times of 5 μs. Calculate the added power dissipation in the transformer due to reverse-current pulses.

Solution

$$P_{\text{inst}} = \frac{V^2}{R_s} = \frac{(750 + 750)^2}{100} = 22{,}500 \text{ W pk}$$

The disadvantage of using overly large shunt capacitors is now apparent, as they add to the reverse conduction time:

$$\tau = R_s C_T = 100\ \Omega \times 0.01\ \mu\text{F} = 1\ \mu\text{s}$$

$$t_{\text{rev}} = t_s + \tau = (5 + 1)\ \mu\text{s} = 6\ \mu\text{s}$$

$$P_{\text{av}} = 22{,}500\ \text{W}\frac{6\ \mu\text{s}}{16.7\ \text{ms}} = \mathbf{8.1\ W}$$

A final consequence of shunt capacitors across series rectifiers is the added ripple across the filter capacitor when the charge stored in their series combination is dumped into C_{filt} at the start of each conduction cycle. The peak-to-peak ripple thus produced is

$$V_{\text{p-p}} = 2V_{o(dc)}\frac{C_{T(\text{shunt})}}{C_{\text{filt}}} \tag{7-2}$$

For the example of Fig. 7-10, this amounts to only 0.15 V p-p, but for light current supplies with low filter-capacitor values, the effect could be severe.

A few summary comments regarding diode storage time may now be in order:

1. Silicon power rectifiers (50-A rating or so) may be expected to remain conductive for typically 20 μs after they are switched from forward to reverse bias. As long as the switching transition time (square wave) is several times longer than this (say, 50 μs), or the sine-wave frequency produces transitions similarly slow (say, 10 kHz maximum), reverse conduction should be no problem. One-ampere diodes can be expected to be 5 or 10 times faster than this, while silicon signal diodes have recovery times on the order of 1 to 100 ns.

2. If the switching time cannot be reduced by *LC* or *RC* circuit arrangements, fast-switching diodes are available for a price. For example, the 1N3913 switches 30 A in 100 ns, but costs about \$5.

3. High-voltage rectifiers can be placed in series to obtain higher PRV, but they are vulnerable to high avalanche currents if subjected to fast turn-offs. Generous safety factors (PRV rating two or three times actual expected) and line-transient suppressors (thyrector diodes or *RC* networks) are recommended to reduce the possibility of failure. Shunt capacitors are recommended for series high-current diodes where inductive loads or square-wave switching are involved. Such capacitors are not particularly helpful where the input is 60-Hz sine wave and the load is noninductive.

Forward Recovery Time: Diode turn-on time is typically at least an order of magnitude less than turn-off time, and the effects are generally not harmful in any case. In fact, forward turn-on is often masked by the charge stored in the diode's junction capacitance. This capacitance may be on the order of 10 pF for 1-A or 100 pF for 50-A diodes. It is charged on the reverse half-cycle, and this potential is added to the source during the first portion of the conduction cycle, as shown in Fig. 7-11.

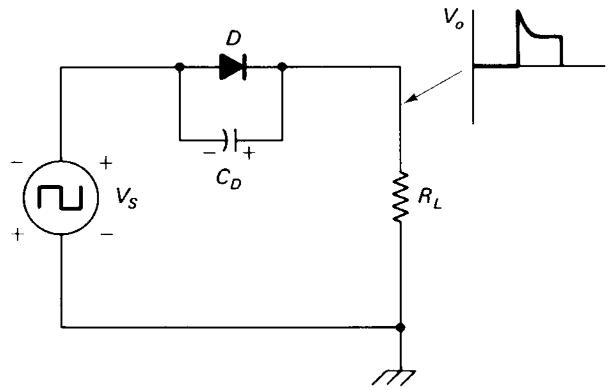


FIGURE 7-11 Charge stored on the diode junction capacitance C_D in the reverse half-cycle may add to V_S on the forward half-cycle, producing an overshoot on the square-wave output.

7.5 ALTERNATIVES TO THE SILICON DIODE

Although the silicon diode captures more than 90% of the market for rectifiers, there are applications where other types are used.

Schottky Barrier or Hot-Carrier Diodes: These diodes employ a silicon-to-metal junction to produce a rectifier with a forward drop typically half that of comparable silicon *P-N*-junction diodes, with virtually zero reverse-recovery time. Power Schottkys are available that will handle 40 A forward with a drop of 0.5 V, so they are an attractive alternative for minimizing power waste in low-voltage/high-current supplies. Their zero turn-off time makes them ideal for high-frequency inverters and switching regulators. In this application, maximum frequency is limited, not by storage time, but by junction capacitance which commonly exceeds 0.001 μF for power Schottkys at low reverse voltages. Hot-carrier (Schottky) rectifiers also have relatively high reverse-leakage currents (as much as 0.01 of the rated I_F at maximum operating temperature) and are limited, at present, to reverse voltages less than 100 V. Figure 7-12 shows the forward and reverse characteristics of a low-power Schottky diode at 25 and 150°C.

Schottky signal diodes are widely employed as microwave detectors and mixers because of their extremely fast turn-off characteristics. Additionally, a whole subfamily of logic elements (the 74S00 series) has been developed using integrated Schottky diodes to improve the switching speed of conventional TTL circuits.

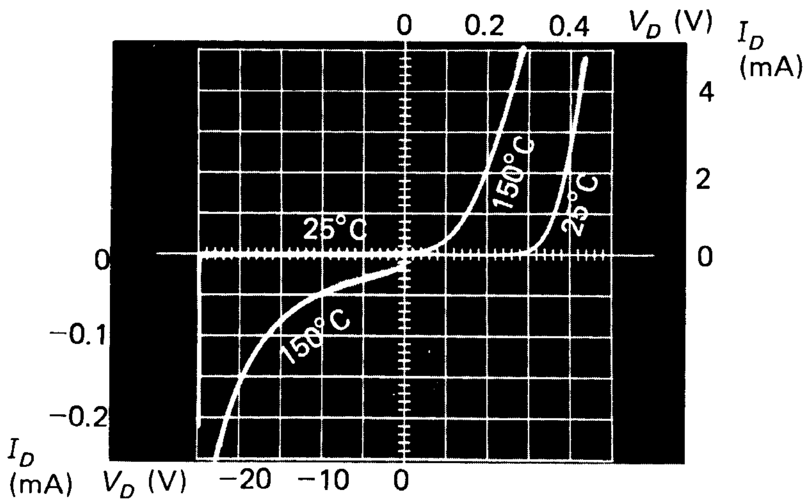


FIGURE 7-12 The Schottky diode has a lower turn-on voltage than the silicon diode, but leakage is severe at high temperatures.

Germanium Diodes: Like silicon, germanium is an element with four electrons in its valence shell, and germanium *P-N* junctions can be formed in a similar manner. Germanium diodes start to turn on at a lower voltage than silicon, but the advantage becomes less important at high currents. More important is the temperature limitation: germanium devices generally have a maximum junction temperature specification from 80 to 110°C, compared to a specification of 170 to 200°C maximum for silicon devices. Silicon diodes can, therefore, handle higher currents for a given package size, even though they drop slightly more voltage. Reverse leakage current and storage time are also notably inferior to comparable silicon units.

Germanium point-contact diodes have been widely used as detectors in receiver circuits because of their low cost and low turn-on voltage. At currents above a few milliamperes, however, the resistance of the internal contact causes a voltage drop greater than that of silicon.

Selenium Rectifiers were popular in the late 1950s, but are seldom seen today. They have a limited life proportional to voltage and current stress, and they smell horrible when overheated. They drop about 1 V forward per cell, but each cell has a reverse-voltage limit of only a few tens of volts, making a stack of cells mandatory for high-voltage rectification.

7.6 ZENER AND AVALANCHE DIODES

Avalanche Diodes: The avalanche phenomenon, wherein high-speed charge carriers collide with and dislodge valence electrons, has been described in Section 7.3. The result is the rapid transition of the reverse-biased diode from nonconduction to

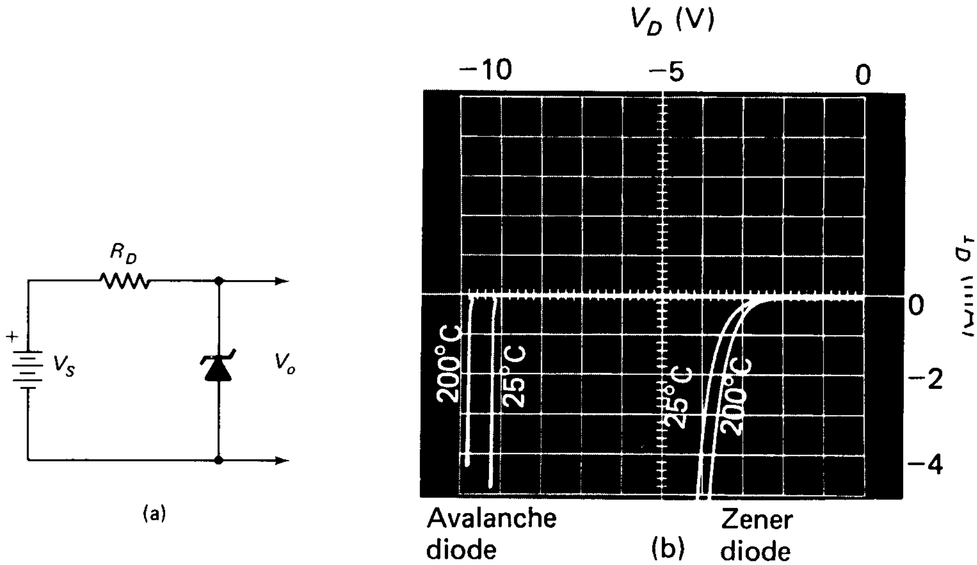


FIGURE 7-13 (a) Basic zener (or avalanche) diode circuit. Changes in V_S are taken by R_D, leaving V_o constant. (b) Avalanche and zener characteristics showing opposite temperature coefficients.

full conduction at the avalanche voltage. By changing the doping level and junction width, the reverse breakdown voltage can be controlled and brought down as low as 6 to 8 V. Such a diode, supplied from a higher-voltage source through a suitable dropping resistor, as in Fig. 7-13, will maintain a nearly constant voltage across its terminals, regardless of variations in the voltage source.

Avalanche Noise: The inherhent violence of the collisions involved in the avalanche mode of conduction results in a good deal of random electron motion, which shows up as noise across the diode. This noise increases slightly at lower avalanche currents, but becomes extremely severe at the threshold of conduction, as seen in Fig. 7-14. The energy of the avalanche noise covers a wide frequency range, but it is the higher-frequency components which are most likely to cause problems by stray-capacitive coupling to other circuits. These can be filtered to an acceptable level by simply placing a 0.1-μF capacitor across the diode. If elimination of the low-frequency noise is required, the zener will have to be followed with an *RC* filter, and the poorer load regulation due to the added source resistance will have to be tolerated.

Zener Diodes: The avalanche phenomenon does not occur below about 6 V, but it is possible to construct a regulator diode with similar behavior based on a different principle. If the *P-N* junction is made extremely narrow, the voltage gradient (volts/cm) across the depletion zone will be high enough to pull some electrons out of their valence bands, even though the total voltage is only on the order of 5 V. This high-field breakdown is properly termed *zener conduction*, and is shown in Fig. 7-13 in comparison to avalanche conduction. Notice that the zener *knee* (start of conduction) is much less abrupt than the avalanche knee. Note also that the zener voltage varies somewhat more with current than does the avalanche

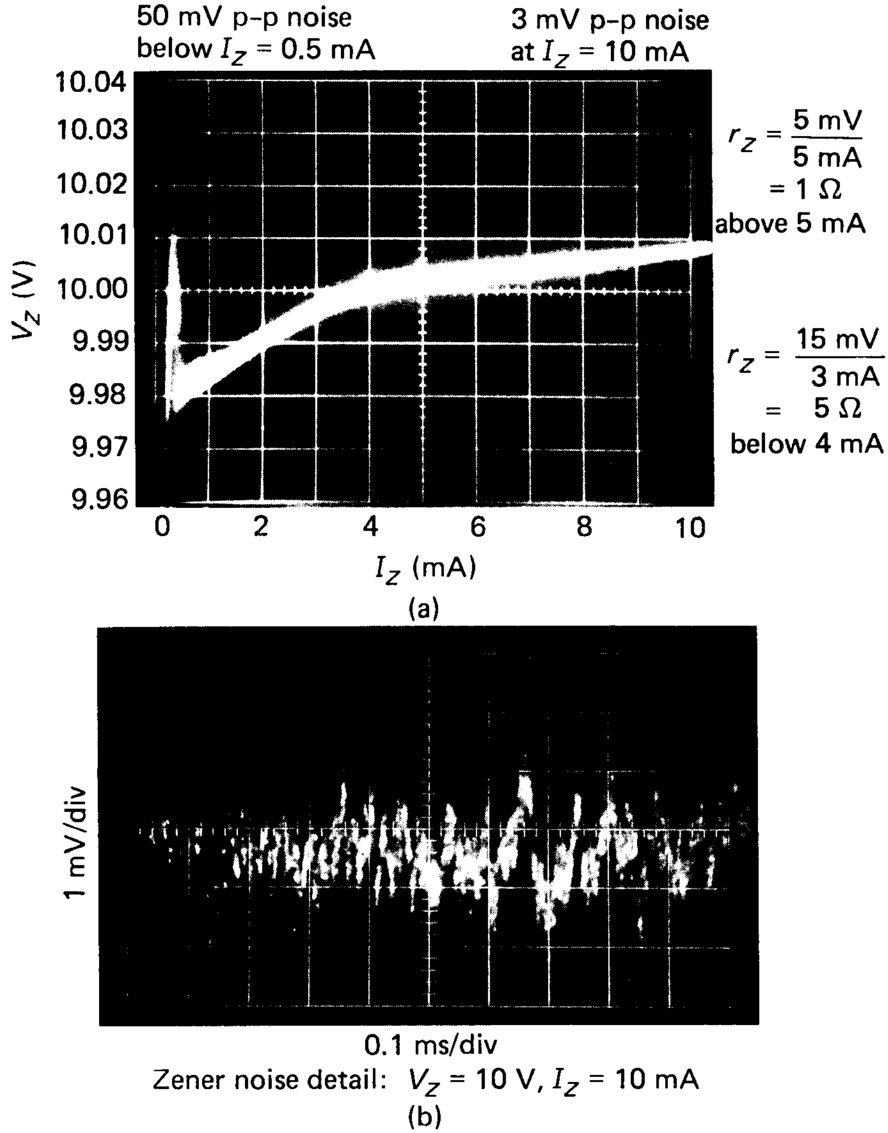


FIGURE 7-14 *Dc flowing through an avalanche diode produces random noise which becomes more severe at the threshold of conduction.*

voltage, as indicated by the steeper vertical slope of the avalanche-mode curve. Noise generated by the zener diode is typically an order of magnitude less than for the avalanche diode.

Temperature Coefficients: The distinction between avalanche and zener diodes is not often made by manufacturers, who tend to sell them both under the names *breakdown diode*, *regulator diode*, or *zener diode*. Besides the better voltage regulation of the avalanche diode and the lower noise of the zener diode, there is, however, a distinct difference in the effect of temperature on breakdown voltage for the two devices. Thermal agitation of the crystal lattice makes it easier for the high field to pull loose some electrons, so zener breakdown voltage decreases at higher temperatures (negative temperature coefficient). However, such agitation makes it difficult for an electron to find a long-enough clear path to acquire the velocity necessary to initiate avalanche collisions, so avalanche breakdown voltage increases at higher temperatures (positive temperature coefficient). The voltage change involved can easily be as high at 5 to 10% for 0 to 100°C temperature shifts. Regulator diodes in the 6 to 7-V range exhibit a near-zero temperature coefficient because of a balance between the zener and avalanche effects.

Since the forward turn-on voltage of an ordinary silicon diode decreases with temperature, it is possible to produce a higher-voltage zero-temperature-coefficient regulator by placing an avalanche diode in series back to back with an ordinary diode having an equal but opposite temperature coefficient. They can be recognized by the fact that they do not conduct at 0.6 V in the "forward" or nonbreakdown direction as uncompensated regulator diodes do.

Self-heating within the diode is usually the major cause of wide temperature variations, so it follows that better voltage regulation can be obtained by operating the diode at the lowest possible current levels. Cascading one regulator circuit after another will keep the current to the second diode nearly constant in spite of source voltage changes, thus improving output-voltage stability.

Dynamic Resistance of Regulator Diodes: Increasing the current through a zener or avalanche diode does produce a slight voltage increase across the device, as illustrated in Fig. 7-15. This $\Delta V/\Delta I$ ratio is the dynamic resistance of the device, Z_{ZT} or R_{ZT}, which is useful in calculating its response to ripple or to changes in input voltage and load current.

Quality avalanche diodes operating at 500 mW can be expected to have R_{ZT} in ohms equal to or less than the numerical voltage rating (i.e., R_{ZT} of 10 Ω for a 10-V 500-mW diode). Zener diodes typically have a dynamic resistance several times lower than given by this rule of thumb. Dynamic impedance can be expected to be roughly inversely proportional to dc current. For example, a 10-V diode at 50 mA dc might exhibit $R_{ZT} = 10\ \Omega$, rising to 100 Ω at 5 mA dc. A power avalanche diode operating at 10 V, 500 mA, would be expected to have $R_{ZT} = 1\ \Omega$. There is so much variation in this parameter, however, that it is wise to trust nothing but the manufacturer's guaranteed specifications on R_{ZT}, where they are available.

Voltage-Regulating Circuits: The basic regulator circuit of Fig. 7-13 provides ripple reduction and stabilization against line-voltage and load-current changes. Its limitations are (1) fixed output voltage, (2) limited output current, (3) temperature sensitivity, and (4) avalanche noise. Transistor or IC amplifier circuits can be used to remedy the first two problems. Temperature effects are best combatted by limiting diode current to reduce self-heating to an absolute minimum. Temperature-compensated regulator diodes may also be employed. Noise can be limited with a 0.1-μF filter or by using the lower-voltage zener diodes.

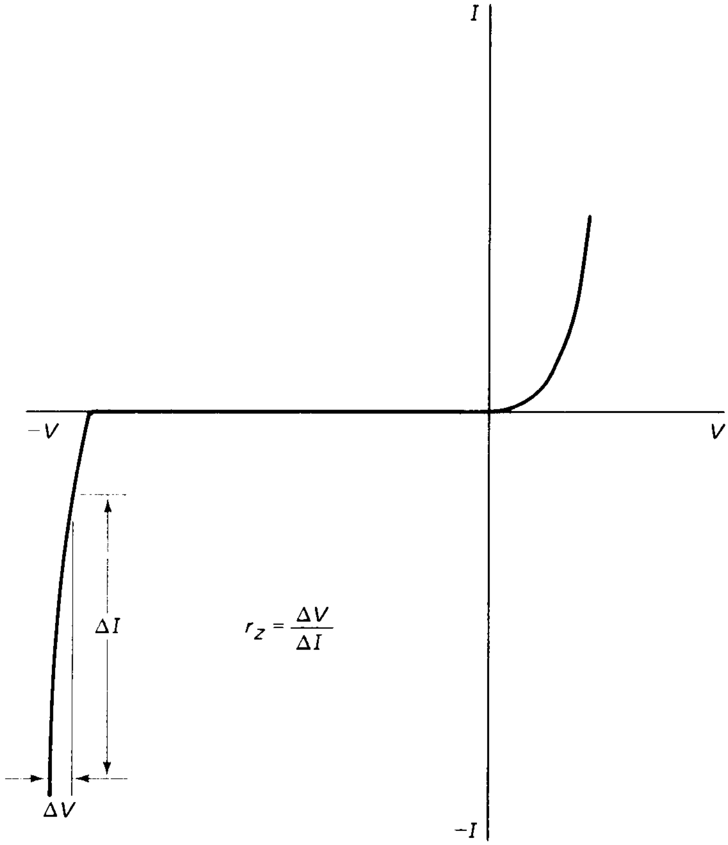


FIGURE 7-15 *Dynamic resistance (resistance to changes in voltage) is given by the slope of the curve ΔV/ΔI.*

EXAMPLE 7-4

In the circuit of Fig. 7-16, an unregulated 18-V source containing 0.2-V p-p ripple feeds a regulator diode with the following parameters:

$$V_Z = 10\text{ V} \qquad T_{\text{ambient}} = 25°\text{C}$$
$$P_{Z\max} = 1\text{ W} \qquad TC = 7\text{ mV/°C}$$
$$T_{\max} = 175°\text{C} \qquad R_{ZT} = 7\ \Omega\max\text{ at }25\text{ mA}$$

The load current of 30 mA may drop to zero.

(a) Find R_D to operate the diode at 250 mW.

(b) Find the minimum input voltage that will maintain $I_Z = 1$ mA, and the output ripple at this point.

(c) Find the maximum input voltage that will keep P_Z below maximum ratings, and the ripple at this point.

(d) Find the dc output voltage change from (b) to (c).

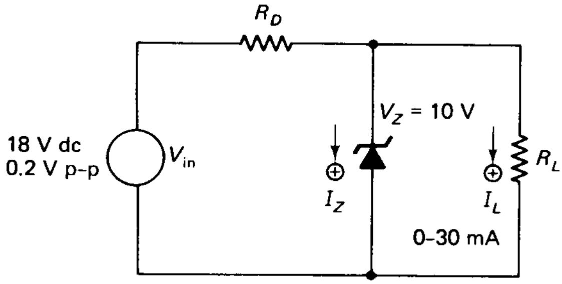


FIGURE 7-16 Zener regulator circuit with load resistance (Example 7-4).

Solution

(a) For a diode power of 0.25 W at 10 V, I_Z is 25 mA:

$$R_D = \frac{V_{in} - V_Z}{I_Z + I_L} = \frac{18 - 10}{25 \text{ mA} + 30 \text{ mA}} = \mathbf{145\ \Omega}$$

(b) If I_Z drops to 1 mA:

$$V_{RD} = (I_Z + I_L)R_D = 31 \text{ mA} \times 145 = 4.5 \text{ V}$$

$$V_{in} = V_Z + V_{RD} = 10 + 4.5 = \mathbf{14.5\ V}$$

The value of R_{ZT} at 1 mA is estimated as 25 times that given at 25 mA:

$$R_{ZT} = 25 \times 7 = 175\ \Omega$$

Voltage division of 0.2 V across R_D and the parallel combination of R_Z and R_L gives the output ripple:

$$R_L = \frac{V}{I_L} = \frac{10}{30 \text{ mA}} = 333\ \Omega$$

$$R_L \| R_Z = 333 \| 175 = 115\ \Omega$$

$$V_o = 0.2 \frac{115}{145 + 115} = \mathbf{0.088\ V\ p\text{-}p}$$

(c) For a diode power of 1 W at 10 V, I_Z is 100 mA. The worst case occurs when load current drops to zero:

$$V_{RD} = I_Z R_D = 100 \text{ mA} \times 145 = 14.5 \text{ V}$$

$$V_{in} = V_Z + V_{RD} = 10 + 14.5 = \mathbf{24.5\ V}$$

R_{ZT} is estimated as $\frac{1}{4}$ of the 7 Ω specified at 25 mA, or 1.75 Ω. R_L is negligible in calculating V_o:

$$V_o = 0.2 \frac{1.75}{145 + 1.75} = \mathbf{0.0024\ V\ p\text{-}p}$$

(d) From near-zero power at (b) to full-rated power at (c), the diode will experience a temperature rise from near 25°C to 175°C, or 150°C.

$$\Delta V = TC\,\Delta T = 7\ \text{mV}/°\text{C} \times 150°\text{C} = \mathbf{1.05\ V}$$

The foregoing analysis should make it clear that it is wise to avoid low zener currents in the interest of better ripple reduction and to avoid wide swings in zener current in the interest of better voltage regulation.

Thyrectors and Varistors: A variety of bidirectional voltage-clamping devices are available for suppressing transient overvoltages across ac or dc supply lines or sensitive semiconductors. The devices may consist of back-to-back avalanche diodes, or selenium diodes, or ceramic-composition materials, and may be marketed as thyrectors, varistors, VDRs, or other trade names. They are available with limiting voltages from ± 10 to over ± 1000 V. Energy ratings in joules are specified rather than power ratings in watts, because the transients they are expected to suppress are too brief to allow heat dissipation to the environment.

EXAMPLE 7-5

It is desired to suppress transients on a 120-V ac line which may be as high as 2000 V pk through a 150-Ω source impedance and may last as long as 100 μs. Choose an appropriate thyrector.

Solution

The peak ac line voltage may approach 190 V on a nominal 120-V-rms line, and the thyrector absolutely must not conduct under this voltage. A 250-V thyrector is chosen, giving a 30% safety margin.

$$I = \frac{V}{R} = \frac{2000 - 250}{150} = 11.7\ \text{A}$$

$$W = Pt = IVt = 11.7 \times 250 \times 0.707 \times 100\ \mu\text{s} = \mathbf{0.2\ J}$$

A $\frac{1}{2}$-J, 250-V unit would be selected.

EXAMPLE 7-6

Select a thyrector to protect the switching transistor in Fig. 7-17 from inductive transients produced by the relay coil. Switching speed may be as high as 10 pulses per second.

Solution

The thyrector voltage must be greater than 24 V and less than 65 V. A rating of 40 V is chosen. The energy expended by the coil at each pulse is

$$W = \tfrac{1}{2}LI^2 = \tfrac{1}{2} \times 0.5 \times 1^2 = \mathbf{0.25\ J}$$

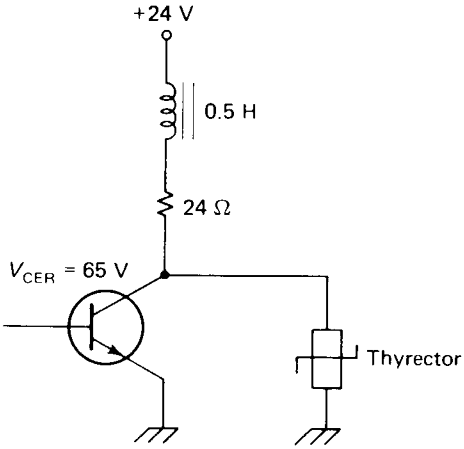


FIGURE 7-17 *A thyrector diode prevents voltage surges which could be damaging to other semiconductors.*

The average power may be as high as

$$P = \frac{W}{t} = \frac{0.25\ \mathrm{J/pulse} \times 10\ \mathrm{pulse}}{1\ \mathrm{s}} = \mathbf{2.5\ W}$$

Part of this energy is delivered to the coil's internal resistance, so the ratings given above are somewhat conservative for the thyrector.

7.7 TUNNEL DIODES

The tunnel or Esaki diode is a two-terminal device capable of switching, oscillating, and amplifying at low power levels into the gigahertz region. Its unique characteristic curve, shown in Fig. 7-18(a), cannot be accounted for by traditional conceptions of atomic structure, and must be explained by the newer and more complex theories of quantum mechanics. The devices are formed as silicon, germanium, or gallium arsenide *P-N* junctions that are very heavily doped to produce extremely narrow depletion-zone barriers. The unexpected currents that were observed were facetiously said to be due to electrons tunneling under the barrier, since they supposedly did not have enough energy to "climb over" it.

A Switching Circuit is produced by biasing the diode from a relatively high-voltage source through a high-value resistance, as shown in Fig. 7-18(b). Two stable points (1 and 2) are thus available, and the diode can be switched between them by externally applied pulses.

An Oscillator is produced by biasing the diode from a low-voltage source through a low-value resistor, so that only one bias point (3) is possible. Since the source voltage required is generally 0.2 to 0.3 V, it is most convenient to use a voltage divider having the required Thévenin-equivalent voltage and resistance, as shown

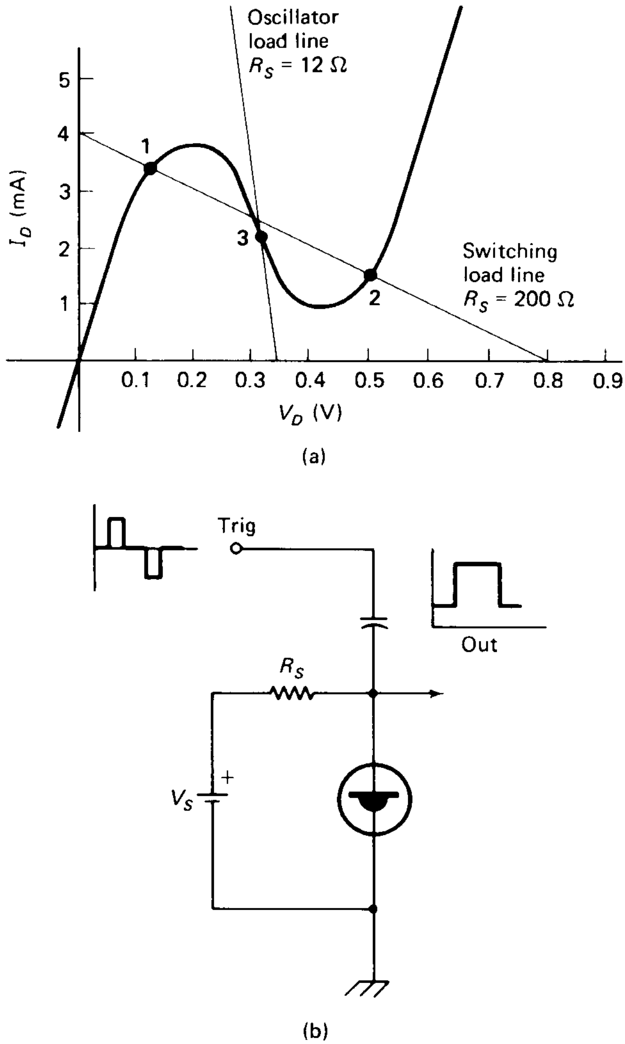
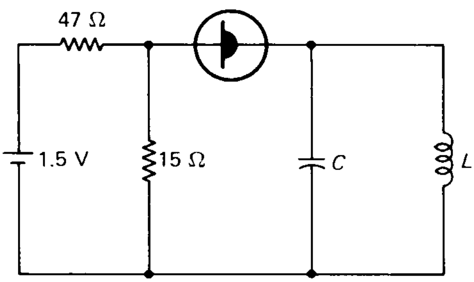


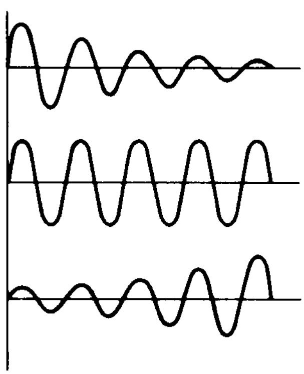
***FIGURE 7-18** (a) A tunnel diode can be biased either as a switch or an oscillator. (b) Tunnel-diode switch circuit.*

in Fig. 7-19(a). The parallel-tuned circuit L and C determines the frequency of oscillation and must be wired with an absolute minimum of stray inductance and capacitance if VHF self-oscillation is to be avoided. Oscillation below 1 MHz will be difficult to control in any case.

The oscillation mechanism is most interesting. It is well understood that the oscillations in an LC tuned circuit will die away due to the inherent resistive losses of the components unless energy is added in the proper phase from the outside. Ideal L and C components would have zero resistance and would allow the



(a)



(b)

FIGURE 7-19 (a) Tunnel-diode oscillator circuit. (b) Damped wave decreasing as energy is lost in circuit resistance, steady wave in an ideal lossless LC circuit, and increasing wave with energy supplied by negative resistance.

oscillations to sustain themselves indefinitely, as in Fig. 7-19(b). The tunnel diode has a negative resistance at the oscillator bias point, since increasing voltage causes decreasing current. A tunnel diode suitably biased can thus be placed in series with a tuned circuit to cancel its positive resistance, allowing the ideal of perpetual oscillation to be realized. The amplitude of the oscillation is limited by the range of the negative-resistance region of the diode, which generally lies between 0.1 and 0.4 V.

Back Diode: Tunnel diodes with peak-point currents of only a few μA are called tunnel rectifiers or back diodes, and are used to rectify signals as low as 20 mV pk. The conductive direction is the "reverse" direction by conventional diode terminology, and the nonconductive direction is "forward" direction. The terms "back direction" and "tunnel direction," respectively, are recommended to avoid confusion. Of course, if signals above about 0.4 V pk are encountered, the tunnel rectifier will begin to conduct in both directions, limiting the rectifier output to 0.5 V dc. A Schottky diode can be used to improve the rectifier's response at higher voltages, as shown in Fig. 7-20.

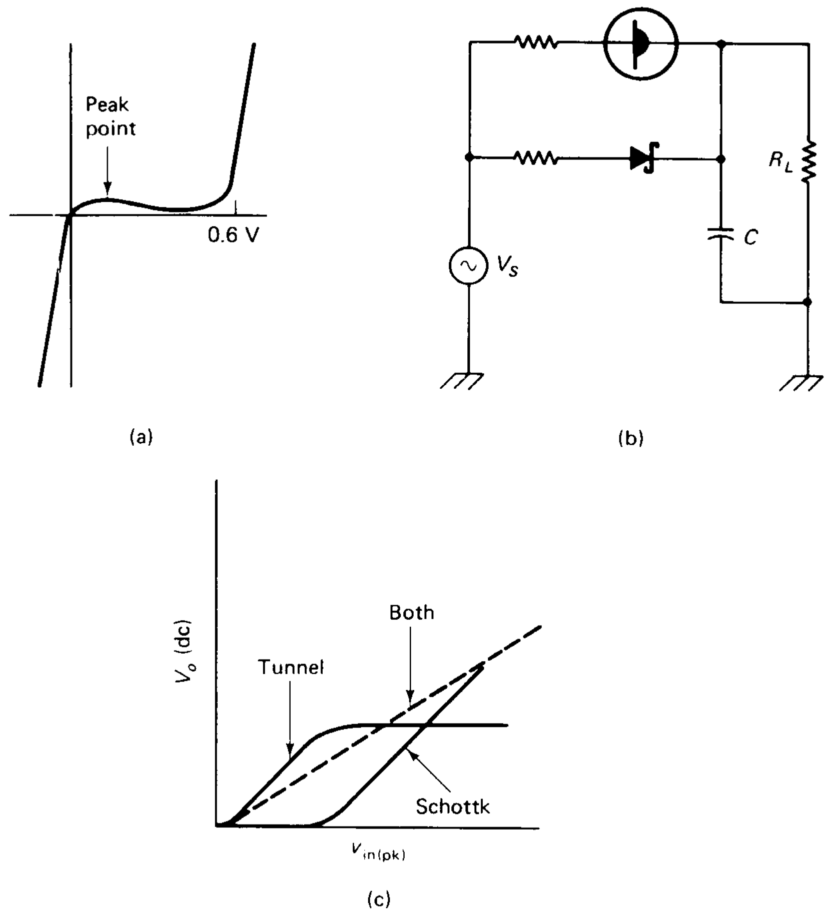


FIGURE 7-20 *(a) Back-diode curves. (b) Low-level detector using back diode for 0-to-100-mV range and Schottky diode for range above 100 mV. (c) Output of detector circuit.*

7.8 VARACTOR DIODES

As was shown in Fig. 7-3(b), a reverse-biased diode consists of two conductive *N*- and *P*-doped zones, separated by an insulative depletion zone. These regions form, respectively, the plates and dielectric of a capacitor. Increasing the reverse bias widens the depletion zone, effectively increasing the plate separation and lowering the capacitance. Thus any *P-N* junction diode is a voltage-variable capacitor. Typical power diodes have capacitances of a few tens of picofarads at low reverse bias. Diodes manufactured to exploit this variable-capacitance effect are called varactors and are available with maximum capacitance from a few picofarads into the 1000-pF range. Capacitance tolerance of ± 10% for a specified reverse voltage is common. Changes on the order of +1 or +2% for a 25°C temperature rise may be expected.

Maximum-to-minimum capacitance ratios may be as low as 2:1 or as high as 20:1. Device Q (reciprocal of dissipation factor) is generally a few hundred at low reverse bias, increasing to several thousand at high reverse bias.

Used as part of a tuned circuit in a radio-frequency oscillator, varactors can be fed by audio voltages to produce frequency modulation, or by dc feedback voltages to produce automatic frequency control or automatic color control in radio and TV circuits. As replacements for mechanical variable capacitors, they are smaller, more rugged, less affected by vibration, and often less expensive.

Figure 7-21 shows a typical voltage-variable tuned circuit and the V-versus-C curve for a representative varactor. Notice that maximum specified capacitance is achieved at −0.3 V across the varactor, dictating a rather small rf signal (0.6 V p-p) to avoid forward-biasing the diode on positive signal swings. For a 6-V-p-p signal, the control voltage would have to be held greater than −3 V, making the maximum usable varactor capacitance 40 pF.

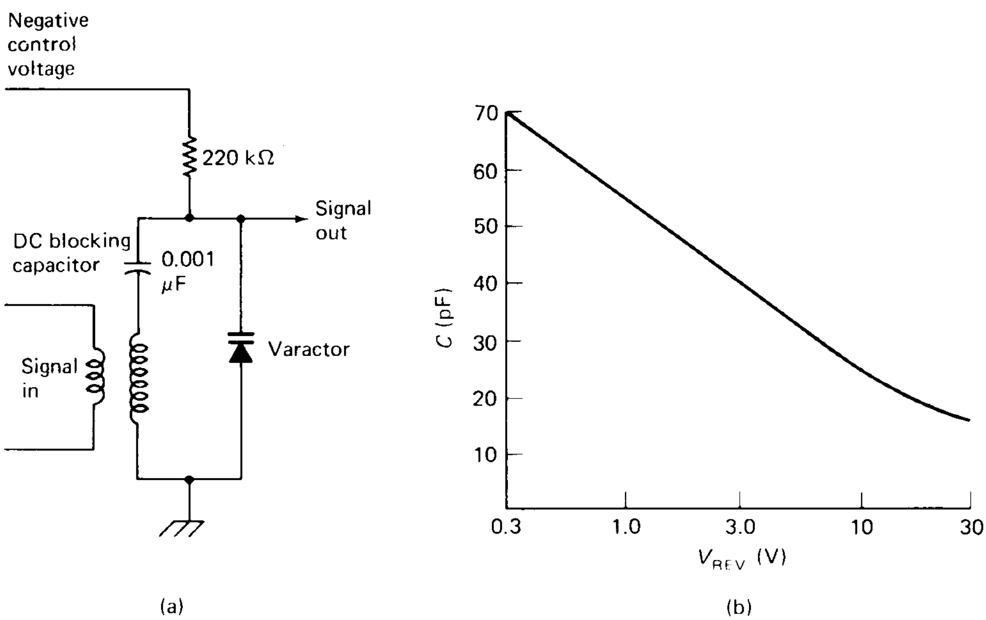


FIGURE 7-21 (a) Circuit diagram for varying the frequency of a tuned circuit with a dc voltage to a varactor diode. (b) Typical C versus V_R curve for a varactor diode.

Varactor Multipliers: Power varactor diodes are useful for producing several watts of output power at frequencies in the 1-to-3-GHz range, where power transistors are unavailable or prohibitively expensive. Efficiencies on the order of 60 to 70% are obtainable in the frequency doubler or tripler mode, so that 10 W at

500 MHz from a conventional transistor rf amplifier can be used to produce 6 W at 1500 MHz.

The principle of operation of the varactor multiplier is easy to grasp if the following basic facts are understood.

1. Nonlinear elements tend to distort sine-wave signals.
2. Distorted sine waves are really nothing more than fundamental sine waves with harmonic sine-wave components added (Fourier's theorem).
3. A diode, ideally at least, burns no power since it is always fully on (zero voltage) or fully off (zero current).
4. A capacitor burns no power, since it is a reactance.
5. A varactor is nonlinear, both as a diode and as a voltage-variable reactance.

Figure 7-22 shows an elementary varactor doubler. The input network is tuned to the fundamental while the outputs are tuned to the second harmonic. The 100-kΩ resistor keeps the varactor biased so that it is always driven just slightly into conduction by the input signal positive peaks.

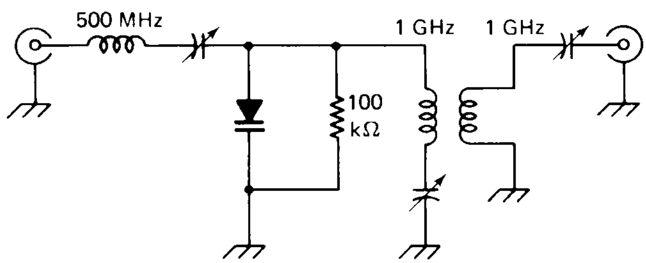


FIGURE 7-22 Circuit diagram for a frequency doubler using a varactor diode to generate harmonics.

7.9 LIGHT-EMITTING DIODES

In a forward-biased *P-N* diode, hole–electron pairs are constantly being formed and recombining. The recombination process releases energy, which for certain materials is manifested as visible light. Although red is by far the most common, LEDs can produce yellow, green and infrared radiation.

Forward turn-on voltage for the most popular gallium-arsenide LEDs varies between 1.0 and 2.0 V. Forward current requirements vary with size and desired intensity, but generally fall between 2 and 20 mA. Reverse voltage must generally be limited to 4 V or less to prevent reverse breakdown.

Actual light power output from an LED is less than 1% of the electrical power input at 25°C. This efficiency nearly doubles for each 25°C reduction in temperature, and halves for each 25°C increase. At any given temperature, light-output intensity varies directly and quite linearly with forward current. Response speed for LEDs is quite good, 0.1-μs rise times being typical.

Laser Diodes: Common light sources actually produce a confused jumble of radiation at many frequencies and in random phase. Laser light, by comparison, is generated at a single frequency and phase. The two types of radiation are termed incoherent and coherent, respectively. "Static" and "snow" are examples of incoherent signals in the RF range, as opposed to a radio-transmitter carrier which is coherent. By giving special attention to the precision of the optical surfaces during manufacture, LEDs can be produced which generate coherent light at forward currents in the range 10 to 100 A. The heating effect of such high currents makes desirable a pulsed mode of operation with a duty cycle on the order of a fraction of 1% to avoid overheating. Cooling the LED junction can reduce the required current for laser action by approximately half for each 50°C reduction.

8

TRANSISTOR TYPES AND CHARACTERISTICS

8.1 BASIC TRANSISTOR ACTION

The original transistor is sometimes called a bipolar or bijunction transistor (BJT), to distinguish it from the several other types of "transistors" which were developed subsequently. Its essential feature is current gain: a small input current from base to emitter acts to control a much larger current from collector to emitter. The existence of a dc collector supply capable of delivering the required current is assumed. The ratio of collector current to base current is the current gain, beta, of the transistor:

$$\beta = h_{FE} = \frac{I_C}{I_B} \tag{8-1}$$

Notice that, to a good approximation at least, collector current does not depend upon collector–emitter voltage, but only on base current.

Transistors are available in two polarities. Assuming that the emitter is at ground, the *NPN* operates with positive potentials at its base and collector, and *PNP* operates with negative potentials, as shown in Fig. 8-1.

Characteristic Curves: The ideal transistor would be completely described by its current gain, $I_C = \beta I_B$. Real transistors place many restrictions on this idealism, however, and the art of using transistors is largely the art of learning to operate

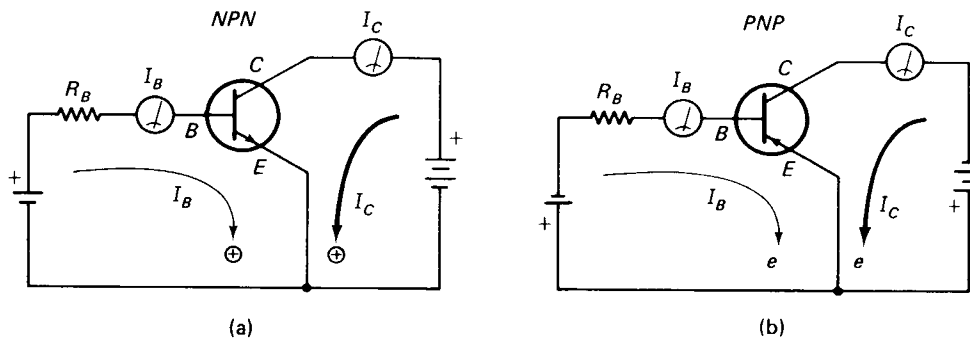


FIGURE 8-1 *The basic transistor action is to use a small base current to control a large collector current.*

within these restrictions. A graph of collector current versus collector–emitter voltage for a series of fixed-base-current values turns out to be remarkably helpful in assessing and visualizing the limitations of a transistor. Such graphs are called collector characteristic curves and can be displayed on a cathode-ray tube using a curve tracer or an inexpensive curve tracer adaptor for a conventional oscilloscope.

Figure 8-2(a) shows an idealized set of characteristic curves for a transistor having a β of 100. Notice that I_C is 100 times I_B, anywhere on the curves, regardless of collector–emitter voltage. Figure 8-2(b) shows, in exaggerated form, several of the limitations of a real transistor. Beta is not constant, but varies with I_C and with V_{CE}. There is a minimum voltage, called $V_{CE(\text{sat})}$, below which the transistor's collector current falls off rapidly and beta becomes meaningless. There is a maximum voltage, called $V_{CE(\text{sus})}$, above which the collector current increases

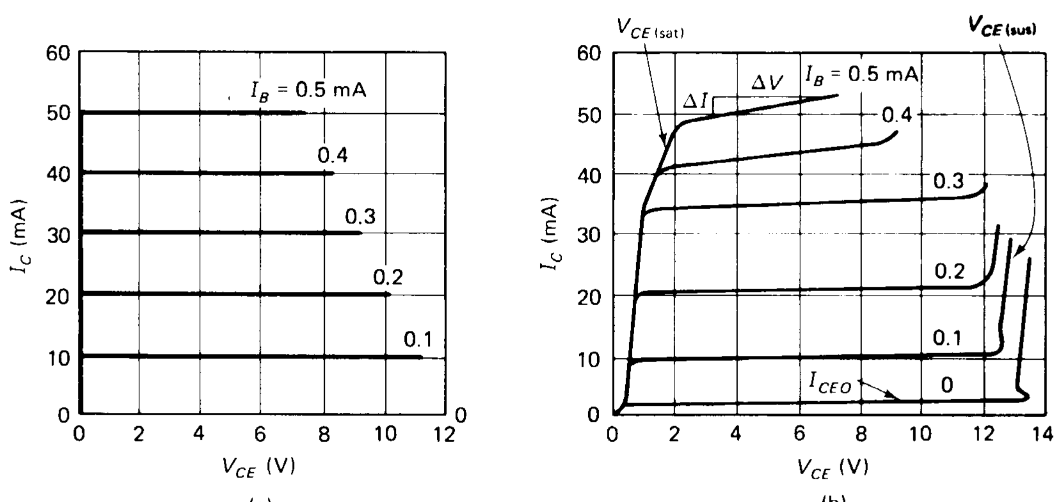


FIGURE 8-2 *Ideal (a) and real (b) transistor characteristic curves. Some of the limitations are exaggerated in (b) for clarity.*

almost without limit. In addition, I_C does increase slightly with V_{CE}, a condition that becomes quite severe at high I_C and low V_{CE}. The next several sections will detail these and other properties of real transistors.

8.2 VARIATIONS IN BETA

Unit-to-Unit Variations: Transistor-specification sheets generally include the minimum and maximum guaranteed beta for a specified collector current and voltage at a temperature of 25°C. Minimum-beta specs may range as low as 10 for high-current types to several hundred for high-gain types. Guaranteed maximum beta may be less than 100 for some types and in excess of 1000 for others, but often no upper limit is specified. It is common to find that the range of possible betas for the same type of transistor spans a factor of 10. A factor-of-2 span (at a fixed current, voltage, and temperature) can be achieved by purchasing specially selected lots, but this is expensive, and the replacement problem causes premature baldness among field-service technicians.

Common practice is to design circuits that are independent of beta, as long as it exceeds some critical minimum value. The designer or service technician then need only select a transistor type such that any unit will have this minimum beta at any current, voltage, temperature, and frequency encountered in service.

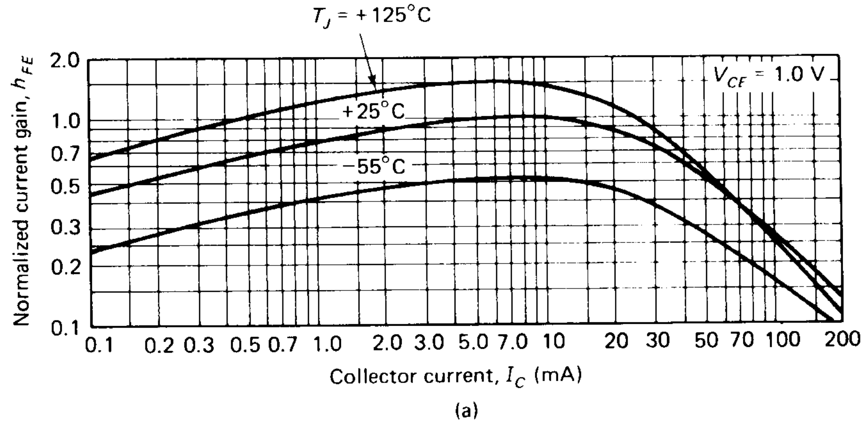
Variations with Temperature: Transistor beta typically increases by 50 to 100% as temperature rises from 25°C to the maximum operating temperature (approximately 175°C for silicon or 85°C for germanium at the semiconductor junction). By the same token, a decrease to 50 to 70% of specified beta may be expected at very low temperatures (−55°C). Figure 8-3 documents the effect for two popular silicon transistors.

Variations with Collector Current: A transistor with a beta of 100 at $I_C = 10$ mA may have a beta of only 50 at $I_C = 0.1$ mA, as illustrated in Fig. 8-3. Even more drastic drops in beta are evident at collector currents near the maximum ratings of the transistor. Specifications for minimum beta are often given only for a collector current near the peak of the β versus I_C curve, and it must not be assumed that this minimum applies at significantly higher or lower currents.

Dynamic or AC Beta: When used as ac-signal amplifiers, transistors translate small changes in base current to large changes in collector current. The relevant property is not then total I_C over total I_B, but change in I_C over change in I_B. Thus we define ac beta:

$$\beta_{ac} = h_{fe} = \frac{\Delta I_C}{\Delta I_B} \tag{8-2}$$

Except at very high values of I_C or very low values of V_{CE}, ac beta h_{fe} and dc beta



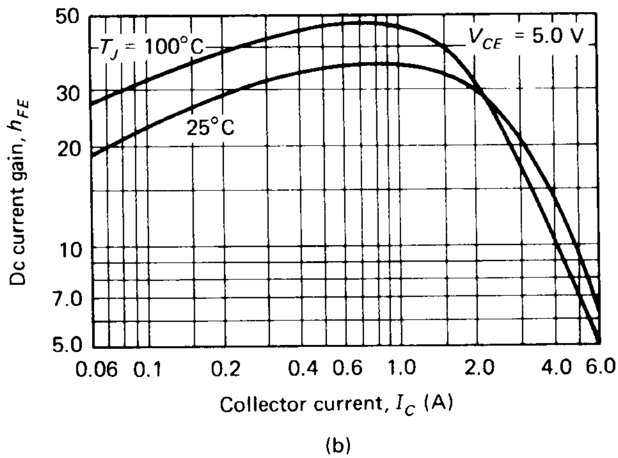


FIGURE 8-3 *Variation of β with I_C for typical low-power (a) and high-power (b) transistors.*

h_{FE} seldom differ by more than 20%, and we do not usually bother to make a distinction between their numerical values. Figure 8-4 shows how ac beta drops drastically where the characteristic curves are pinched together.

8.3 COLLECTOR SATURATION VOLTAGE

When a transistor's collector current is limited not by the transistor itself ($I_C = \beta I_B$) but by the collector supply voltage and resistance ($I_C = V_{CC}/R_C$), the transistor is said to be fully turned on, or *saturated*. Another way of saying this is that an increase in I_B will not result in further increase in I_C. The ideal transistor would have $V_{CE} = 0$ at saturation. Actual transistors will have V_{CE} from 0.1 V or less at collector currents of a few milliamperes up to 2 V and more at collector currents above 10 A.

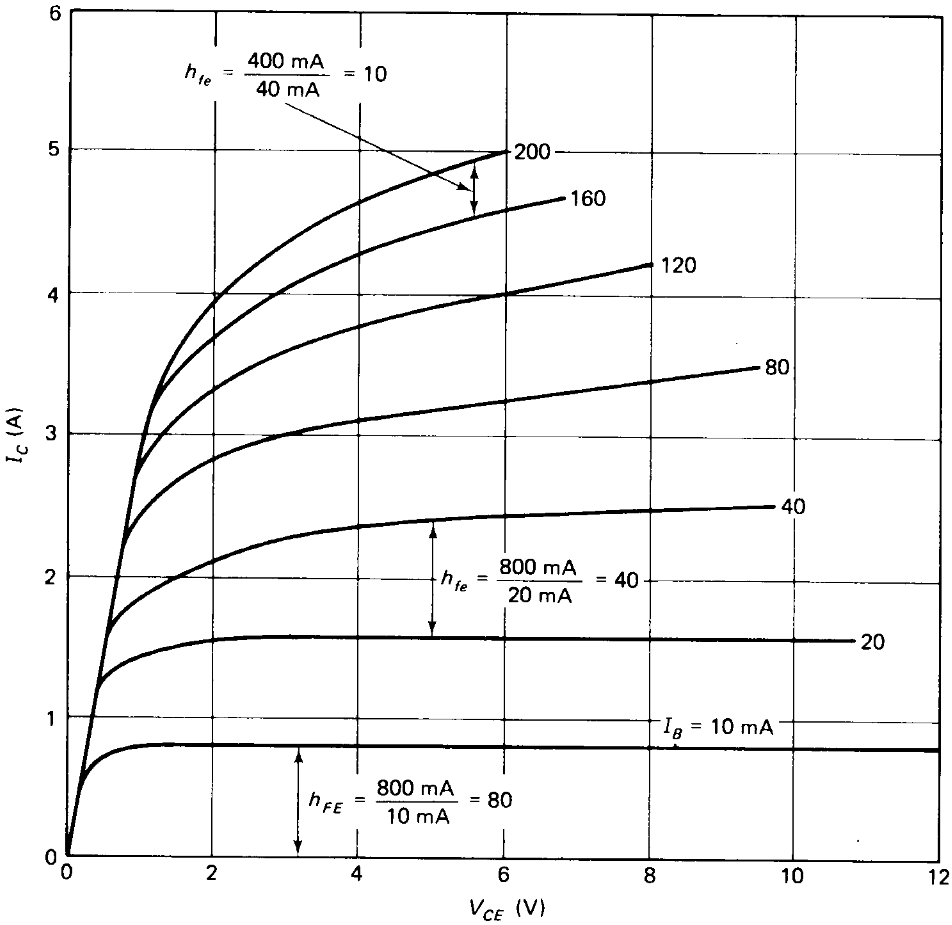
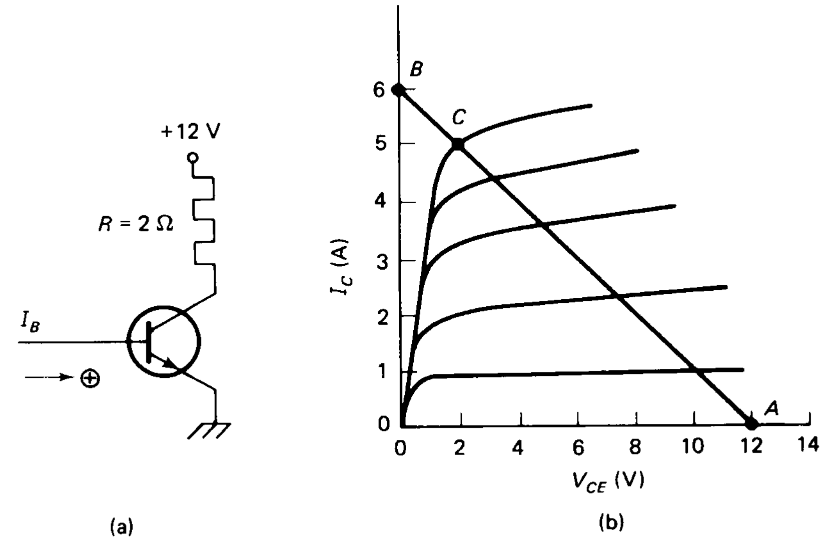


FIGURE 8-4 *Variation of ac beta (h_{fe}) with collector current for a typical power transistor.*

Low-saturation voltage at high collector current is important in a switching transistor, because most of the transistor's power is burned in this state. Figure 8-5 shows a transistor switching in a 2-Ω heating element on a 12-V supply, and the characteristic curves of the transistor. With the transistor turned off, no collector current flows and the full 12 V appears at the collector (point A). No power is burned in this state since there is no current. If the transistor could be made a perfect short from collector to emitter, the full 12 V would appear across R. The collector current would be

$$I_C = I_R = \frac{V_{CC}}{R} = \frac{12\ \text{V}}{2\Omega} = 6\ \text{A}$$

This is represented as point B on the curves. All possible conditions of voltage and current for this circuit must lie on straight line AB. Because of the saturation



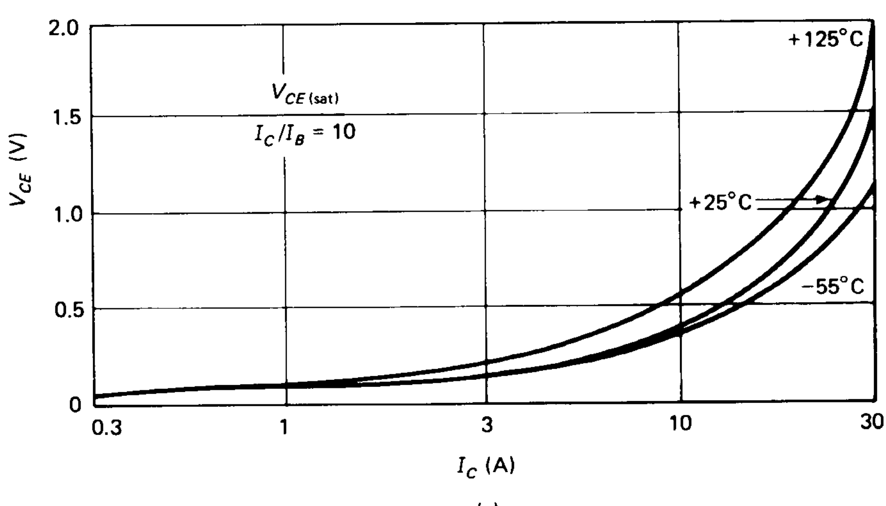


***FIGURE 8-5** (a) Transistor used to switch a heating element. (b) Load line for circuit (a) showing the minimum 2-V drop due to collector saturation voltage $V_{CE(sat)}$. (c) Variation of $V_{CE(sat)}$ with I_C and temperature.*

characteristics of the transistor, any reasonable base current leaves $V_{CE} = 2$ V and the remaining 10 V across R. Thus

$$I_C = I_R = \frac{V_R}{R} = \frac{10\text{ V}}{2\ \Omega} = 5\text{ A}$$

$$P_R = IV = 5 \times 10 = 50\text{ W}$$

$$P_Q = IV = 5 \times 2 = 10\text{ W}$$

The subscript Q is used to identify a transistor, every reasonable letter having been used elsewhere before their invention.

The definition of saturation voltage is somewhat arbitrary since continued increases in base current can always produce slight decreases in $V_{CE(\text{sat})}$. Most manufacturers use $I_B = 0.1 I_C$ and specify minimum $V_{CE(\text{sat})}$ at some stated collector current. $V_{CE(\text{sat})}$ generally remains uniformly low until a critical collector current is reached, whereupon it increases rapidly, as shown in Fig. 8-5. Increases in temperature generally cause increases in $V_{CE(\text{sat})}$, an effect that becomes more pronounced and more critical at high collector currents, again as illustrated in Fig. 8-5.

8.4 COLLECTOR-VOLTAGE LIMITS

Transistor collector-voltage limits may be given by any of more than a dozen different symbols. This section will attempt to sort out the confusion.

Off-State Limits: One method of setting the specification views the transistor not as an active device but as a pair of diodes. The avalanche voltage of the collector–base diode, with the remaining (emitter) lead left open, is then specified as the "collector-voltage limit" V_{CBO}. This specification method produces higher numbers than the active-state methods given below, and is therefore fairly popular among manufacturers seeking a competitive edge. A proliferation of specifications has grown up which supposedly involve both junctions, but in fact short the base to emitter (V_{CES}), effectively short it with a low-value resistor (V_{CER}), or even reverse bias it (V_{CEV}). All these specifications produce numbers that are about equal to V_{CBO}, and all are likewise misleadingly high.

To be sure, a transistor with a solidly turned-off base–emitter junction can withstand the collector–emitter voltages of these specifications, but at these voltages there is no smooth transition from the turned-off to the conducting state, as indicated by the dotted "negative-resistance" line in Fig. 8-6. Attempts to traverse the negative-resistance region may result in self-oscillation at some high frequency determined by stray inductance and capacitance. If the transistor and circuit parameters are tightly controlled, exceptionally fast on–off switching can be achieved by operating the transistor across this region in what is termed the *avalanche mode*.

Active-State Limits: The most commonly specified transistor voltage limit is V_{CEO}, the collector–emitter breakdown voltage with the base left open. This specification is typically about one-half of V_{CBO} or V_{CES} for any given transistor. These breakdown ratings may be given by various manufacturers as $V_{CEO(\max)}$, $V_{(\text{BR})CEO}$, or BV_{CEO}, with the same meaning. A smooth transition from *off* to *on* state can generally be made if the supply voltage is not greater than V_{CEO}, although

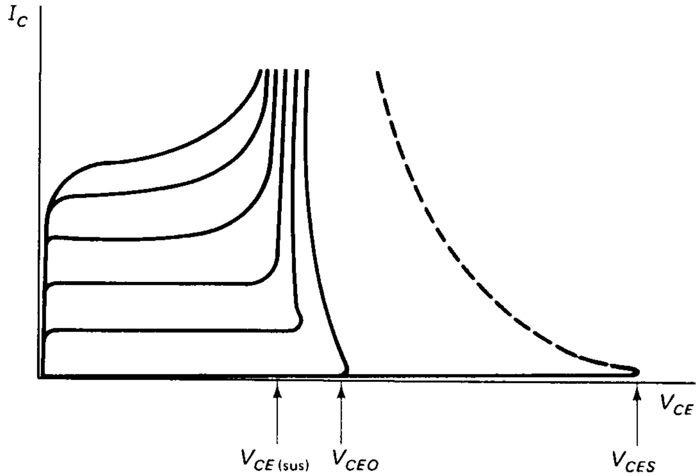


FIGURE 8-6 *Three collector limiting voltages shown in approximate proportion: $V_{CE(sus)}$ with base forward-biased, V_{CEO} with base open, and V_{CES} with base shorted to emitter.*

slight negative resistance and nonlinearity may be encountered at the high-voltage end if a low-value load resistance is used.

The highest voltage for which a transistor's characteristics are a reasonable replica of the ideal curves of Fig. 8-2(a) is termed the *sustaining voltage*. To be complete, the collector current and the base–emitter conditions should be specified. The most commonly given sustaining voltage specs for low collector currents are $V_{CEO(sus)}$ with the base open, or $V_{CER(sus)}$ with the base effectively shorted to the emitter. $V_{CEO(sus)}$ is typically about 80 to 90% of V_{CEO}. $V_{CER(sus)}$ may be about equal to or slightly greater than V_{CEO}. Notice from Fig. 8-6 that at higher currents the ideal shape of the curves gives way at less than half of V_{CEO}. High-current voltage limits are, unfortunately, almost never specified by transistor manufacturers.

Choosing Transistor Voltage Limits: If a transistor is used strictly as a switch and the base is grounded or reverse biased at turn off, it will withstand a collector supply equal to V_{CBO}, V_{CES}, V_{CER}, or V_{CEV}, all of which are about equal. If the base is left unclamped to ground, the transistor will withstand a collector voltage V_{CEO}.

If linear operation with a resistive load is required, the collector-supply limit should be $V_{CEO(sus)}$. If the load contains inductance, such as a choke, transformer, solenoid coil, or motor, the transistor collector-voltage rating should be double the supply voltage, since the inductive voltage adds to the supply voltage during turn-off.

Published voltage limits generally apply at 25°C and should be reduced to 75% of this value to cover maximum operating temperatures. Depending upon the reliability level required, an additional safety factor (up to a factor of 2) is recommended.

Secondary Breakdown, also called second breakdown, is a destructive effect caused by localized heating at certain points along the base–emitter junction. The problem is most often noticed when a high-voltage/high-current load is switched off rapidly, especially if the load is inductive and the pulsing is repetitive. Since the edges of the junction turn off more quickly than the center portion, excessive current concentration and overheating occur in the center. Figure 8-7 shows a representative safe-operating-area curve for avoiding secondary breakdown. As long as all portions of the load line lie within the specified area for the pulse conditions given, the manufacturer guarantees that secondary breakdown will not occur.

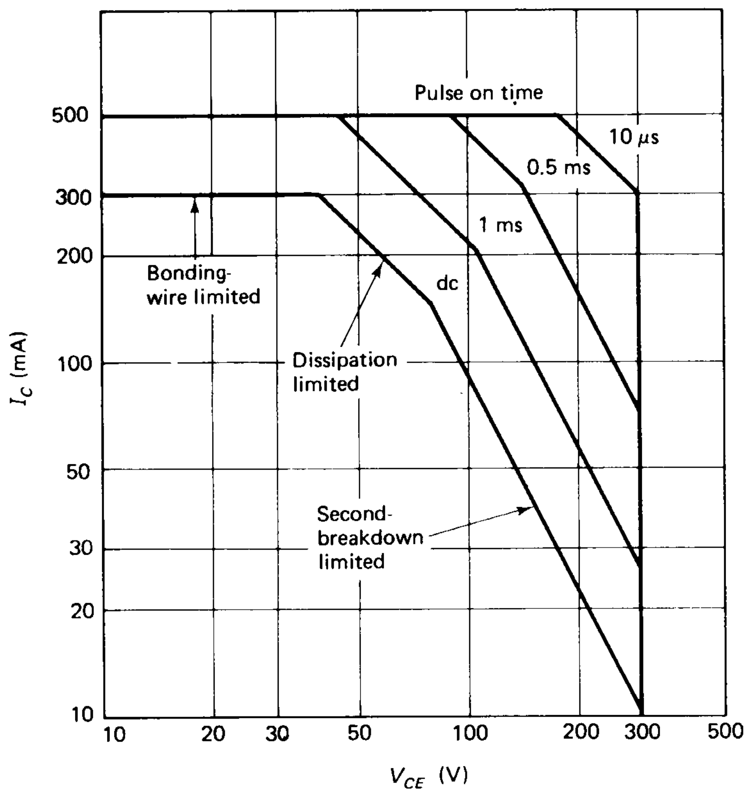


FIGURE 8-7 Safe operating areas for a high-voltage transistor.

8.5 POWER-DISSIPATION LIMITS

The power generated within a transistor is nearly equal to $I_C V_{CE}$, since base current and voltage are usually negligible by comparison. Maximum-power specifications for low-power types generally assume free-air mounting at room temperature ($T_A = 25°C$). A second specification may be given assuming that the case is attached to a heat sink held at room temperature ($T_C = 25°C$). Derating factors are then given (in mW/°C) for ambient temperatures above 25°C. It is assumed that, if the transistor is attached to a heat sink, the temperature of the sink is not significantly raised by the heating effect of the transistor.

Heat Sinking: Power transistors are often rated for power dissipation at a case temperature of 25°C, with derating factors for higher temperatures. It is generally not practical to neglect the temperature rise of the heat sink for transistors dissipating more than 1 W, however. Figure 8-8 shows how the power limit can be calculated from the maximum junction temperature and the thermal resistance from junction to ambient. Heat, expressed in watts, can be visualized as flowing through thermal resistance, producing a temperature drop across the resistance. This is analogous to electrical current flowing through a resistance, producing a voltage drop. The higher the temperature differential, the greater the heat flow. The higher the thermal resistance, the less the heat flow.

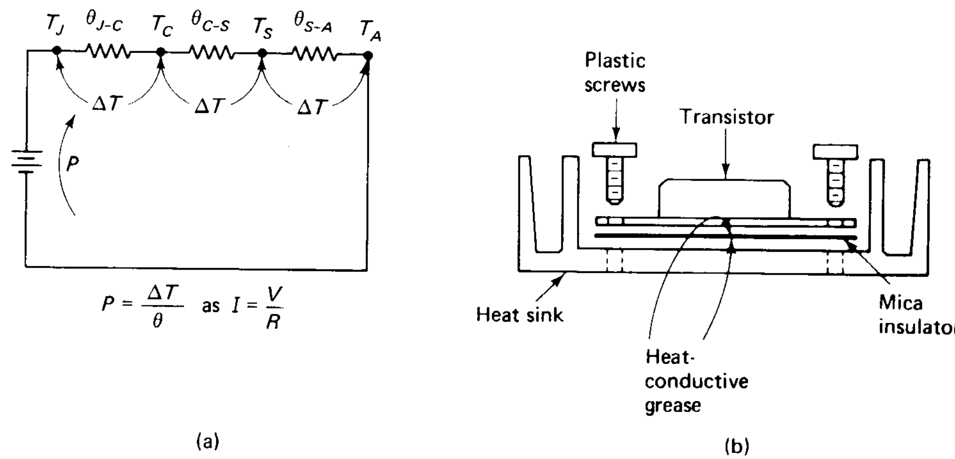


FIGURE 8-8 (a) Thermal power (watts) flows through thermal resistance (θ), producing a temperature differential ΔT. The subscripts refer to junction, case, sink, and ambient. The use of an electric-circuit schematic to represent the process is common. (b) Typical mounting of a transistor to a heat sink with insulation of the transistor case.

Power-transistor specifications usually include maximum junction temperature T_J and junction-to-case thermal resistance θ_{J-C}. Manufacturers of heat-sinking devices supply case-to-sink resistance θ_{C-S} figures for their insulators and sink-to-air resistance θ_{S-A} for their heat sinks. Calculation of maximum power is then quite straightforward.

EXAMPLE 8-1

A transistor has a maximum junction temperature of 180°C and θ_{J-C} of 1.5°C/W. It is mounted through a 0.5°C/W insulator to a 3.5°C/W heat sink. Ambient temperature is 55°C maximum. Find the transistor power limit.

Solution

Total resistance and temperature differential are found first:

$$\theta_T = \theta_{J-C} + \theta_{C-S} + \theta_{S-A}$$
$$= 1.5 + 0.5 + 3.5 = 5.5°\text{C/W}$$
$$\Delta T_T = T_J - T_A = 180 - 55 = 125°\text{C}$$
$$P = \frac{\Delta T_T}{\theta_T} = \frac{12.5}{5.5} = 2.3 \text{ W}$$

In many cases the transistor will be mounted directly to the heat sink, making insulator resistance θ_{C-S} equal to zero. Some data sheets may specify the maximum case temperature T_C, rather than T_J, thus eliminating junction-to-case resistance θ_{J-C} from consideration. If both of these possibilities are true, Ohm's law for thermal circuits reduces to

$$P = \frac{\Delta T_{C-A}}{\theta_{S-A}} \tag{8-3}$$

This illustrates that the law is valid for part of the series circuit as well as for the entire circuit.

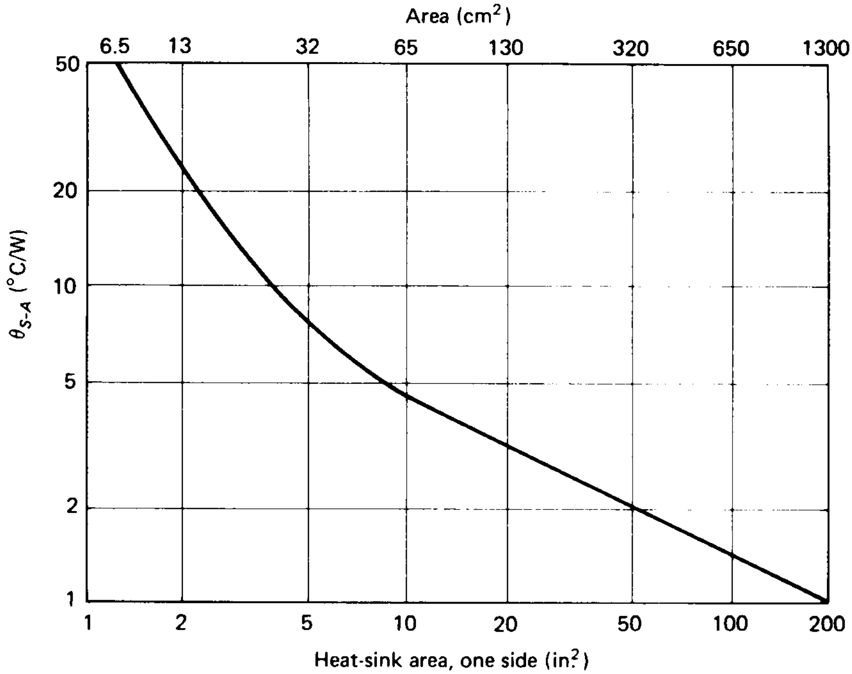
Thermal Resistance: In cases where a transistor is operated without a heat sink, the following table of case-to-air resistances may be used:

Case Style	θ_{C-A}(°C/W)
TO-92 (plastic)	350–200
TO-18 (mini TO-5)	300
TO-5 (standard)	150
TO-60 (stud mount)	70
TO-66 (mini TO-3)	60
TO-220 (power tab)	50
TO-3 (standard power)	30
TO-36 ($1\frac{1}{4}''$ round)	25

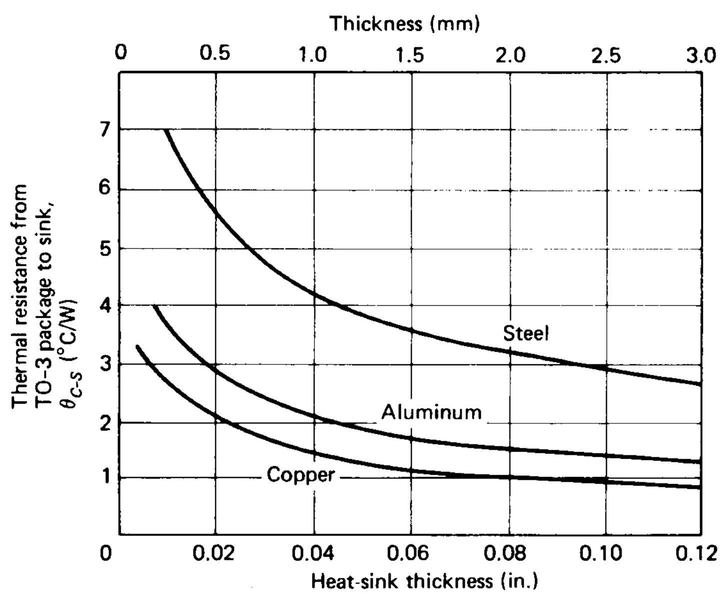
Heat-sink insulators of the mica, plastic, or anodized-aluminum type typically have a thermal resistance of 0.35°C·in.^{2}/W (2.3°C·cm^2/W). Note that smaller areas produce higher resistances, so

$$\theta \approx \frac{0.35(°\text{C}\cdot\text{in.}^2/\text{W})}{A(\text{in.}^2)} \tag{8-4}$$

Silicon grease or a similar heat-sink compound should be applied to all the mating surfaces to prevent air spaces and consequent higher thermal resistance.



(a)



(b)

***FIGURE 8-9** (a) Thermal resistance versus total surface area for a square sheet painted black and mounted vertically in free air. (b) Thermal resistance versus sheet-metal thickness for a TO-3 case mounted to a flat metal surface.*

Commercial heat sinks are often more expensive than the transistor they serve, so it is common to use a simple square of aluminum or a part of the chassis or cabinet as a heat sink. The graph of Fig. 8-9(a) can be used to estimate the thermal resistance of such sinks. The values are approximate for a square aluminum sheet mounted vertically in free air to encourage convection and painted matte black to maximize radiation. For other conditions the following factors apply:

Condition	*Multiply θ by*
Horizontal mounting	1.3
Shiny aluminum surface	1.5
Fan-forced cooling	0.5–0.3

Figure 8-9(b) gives the thermal resistance due to heat conduction from TO-3 package to a thin metal plate. For manufactured heat sinks the metal is always chosen thick enough to make this resistance negligible compared to the dissipation resistance of Fig. 8-9(a). Where a chassis or cabinet is used to sink tens of watts of power, however, inability of the thin metal to conduct the heat to its large surface area may be the limiting factor.

Thermal Time Constant: If a transistor is pulsed with a large amount of power, the thermal capacitance of the case will absorb energy and prevent overheating of the junction for a short time until thermal equilibrium is reached. If it is assumed that the junction temperature rises according to the familiar time-constant curve of Fig. 8-10, the maximum power for a given pulse length can be readily determined, as shown in the example below. One should retain a healthy respect for the kind of trouble he is courting by operating beyond the steady-state limits, however. A relatively-harmless failure in a low-power drive circuit should never be allowed to avalanche into a catastrophic failure in the output stage, load, or power supply. Note also that the cooling time constant is somewhat longer than the heating time constant. If the interval between pulses is not at least three cooling time constants, allowance must be made for cumulative heating.

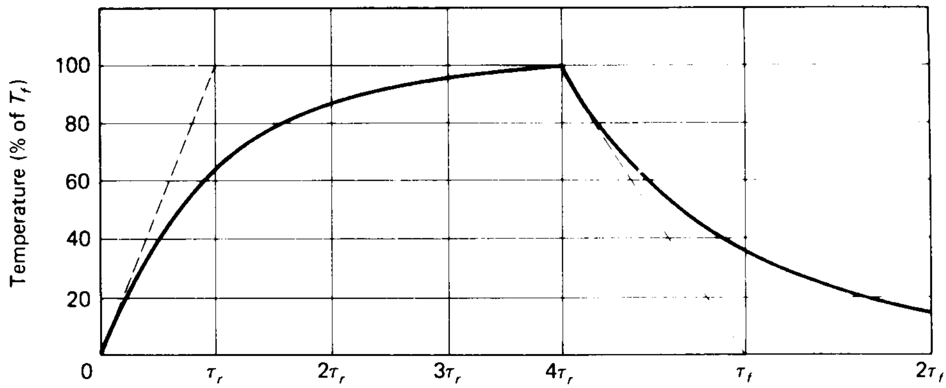
EXAMPLE 8-2

A TO-3 package transistor is rated $T_J = 200°\text{C}$, $\theta_{J-C} = 1.5°\text{C/W}$ and $V_{CE(\text{sat})} \leq 2.0$ V at $I_C = 15$ A. How long can a 15-A *on* pulse be sustained without a heat sink, and how closely may such pulses be spaced?

Solution

Assuming that $T_A = 50°\text{C}$ and referring back to the table of thermal case resistances:

$$P_{\max} = \frac{\Delta T}{\theta} = \frac{200 - 50}{30 + 1.5} = 4.76 \text{ W}$$



Package	Heating τ_r	Cooling τ_f
TO-5	60 s	85 s
TO-66	120 s	160 s
TO-3	120 s	210 s

***FIGURE 8-10** Heating and cooling time constants for various transistor packages.*

The pulse power is considerably greater than this:

$$P_{\text{pulse}} = IV = 15 \times 2 = 30 \text{ W}$$

$$\frac{P_{\text{max}}}{P_{\text{pulse}}} = \frac{4.76}{30} = 16\%$$

The 30-W temperature-rise curve must be stopped at 16% of its full value to avoid overheating. The time-constant curve is in the linear region giving about $0.16\tau_r$ as the time for this rise. Using the τ_r value from Fig. 8-10:

$$t_{\text{max}} = 0.16\tau_r = 0.16 \times 120 = \mathbf{19\ s}$$

To allow a safety margin, an 8-s limit would be specified. Cooling time between pulses should be

$$t_{\text{min}} = 3\tau_{f(\text{cool})} = 3 \times 210 \text{ s} = \mathbf{10\ min}$$

A thermally-operated shutdown circuit is strongly advised to protect against driver failure.

Average Power: When a transistor or other device receives pulses of power at a regular rate much faster than its thermal time constant, it is common practice to compute the average power and treat it in the same manner as dc. Thus a 10-W pulse of 1-s duration recurring every 5 s would be treated as a 2-W continuous dissipation.

When a transistor is switched from the *on* to the *off* state, the power product IV becomes very large during the transition. If the switching rate is low, the time spent in transition will be negligible, and this will present no problem. However, at high switching rates the power burned in transition when averaged may far exceed the saturation-state power. Exact calculation of transition power is well nigh impossible, since switching waveshapes are usually irregular and stray capacitance and inductance causes time lags between voltage and current waveforms. However, for the idealized transition waveform of Fig. 8-11, the average power during transition time t_r can be shown to be

$$P = \frac{IV}{6} \tag{8-5}$$

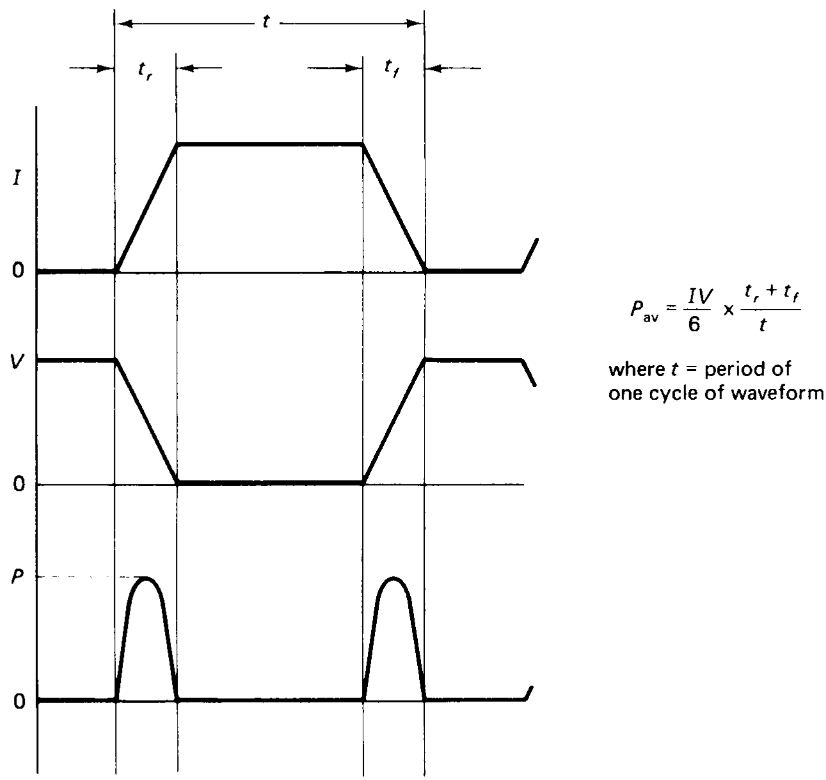


***FIGURE 8-11** Average power developed in an idealized switching device transitioning from full conduction to no conduction.*

When a transistor is operated not as a switch but as a linear amplifier, calculation of transistor dissipation from the IV product at the collector is impractical for all but the simplest waveforms. Neither can average or true rms voltage at the collector be used, since the transistor does not have a constant value of resistance. The power input to the entire amplifier stage can be easily determined, however, since the V_{CC} supply is usually dc. The power output to the load is also fairly easy to measure if the load is resistive. Neglecting losses in any coils, transformers, and resistors in the collector circuit, the transistor power is simply the difference between the two, as illustrated in Fig. 8-12.

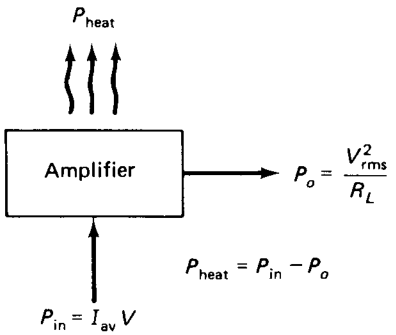


FIGURE 8-12 Relationship of dc input power, signal output power, and heat dissipation in an amplifier.

8.6 COLLECTOR LEAKAGE CURRENTS

In normal operation, a transistor's collector–base junction is reverse-biased. A small leakage current flows, nevertheless, and is termed the *collector cutoff current* with emitter open, I_{CBO}. This leakage current must be stated for a specific voltage and junction temperature.

The transistor treats this leakage current entering the base just as it would a current injected through the base lead—it amplifies it by beta in the collector. Thus the collector–emitter current, with the base left open (termed I_{CEO}), is beta times I_{CBO}. Remember that beta at such low currents may be quite a bit less than beta at normal current levels.

In germanium transistors I_{CEO} is often a noticeable addition to the normal collector bias current at 25°C, and is likely to completely overrun the bias current at 80°C if stabilizing circuitry is not employed. In silicon transistors I_{CEO} is usually negligible except at the upper limit of operating temperature range. Manufacturers generally like to specify I_{CBO} rather than I_{CEO} because it is a more fundamental (although less useful) piece of data, and because it yields lower and therefore more competitive-looking numbers. Some typical values of I_{CEO} are as follows:

I_{CEO}	300-mW *type*	50-W *type*
Germanium		
25°C	5 μA	2 mA
100°C	10 mA	2 A
Silicon		
25°C	50 nA	100 μA
150°C	10 μA	5 mA

Manufacturers sometimes specify the collector leakage with the base shorted to the emitter (I_{CES}), or with the base–emitter junction reverse-biased at a given voltage (I_{CEV}), or with some other base condition (I_{CEX}). Each of these techniques has the effect of draining the I_{CBO} current out of the base, thus preventing its amplification and lowering the value of the published "leakage" figure. In fact, these very techniques are used in operating circuits to combat leakage problems.

8.7 INPUT CHARACTERISTICS

The base of a transistor, which is most commonly used as the input, appears essentially as a forward-biased diode to the emitter and as an open circuit (reverse-biased diode) to the collector. The base–emitter voltage must therefore stay in the range 0.5 to 1.0 V for silicon and 0.1 to 0.6 V for germanium, unless the transistor is turned off. The following approximations are accurate enough for most calculations:

$V_{BE}(Si, I_C = 10\text{ A})$	0.9 V
$V_{BE}(Si, I_C = 10\text{ mA})$	0.6 V
$V_{BE}(Ge, I_C = 10\text{ A})$	0.5 V
$V_{BE}(Ge, I_C = 10\text{ mA})$	0.2 V
Unit-to-unit variation	±0.1 V
Hi temp.	−0.1 V
Lo temp.	+0.1 V

Reverse Base Voltage: Most modern silicon transistors are made by the planar process which requires a very heavy doping of the emitter region, on the order of that used in avalanche-breakdown diodes. As a result, these transistors will conduct heavily, typically at reverse base–emitter voltages of 6 or 7 V. A suitable resistance or a forward diode may be placed in series with the base if larger reverse voltages are expected. The term V_{EB} or V_{EBO} specifies the maximum allowable reverse base voltage.

Dynamic Input Characteristics: Resistance is defined as the ratio of voltage to current: $R = V/I$. Dynamic or ac resistance in a semiconductor presumes the presence of appropriate dc bias levels and examines the ratio of ac voltages and currents superimposed as small fluctuations on these dc levels, as illustrated in Fig. 8-13. William Shockley, who originally conceived the transistor, came to the theoretical conclusion that this dynamic resistance for either a silicon or germanium junction was a function of junction current, being at room temperature

$$r_j = \frac{0.026\text{ V}}{I_J} \tag{8-6}$$

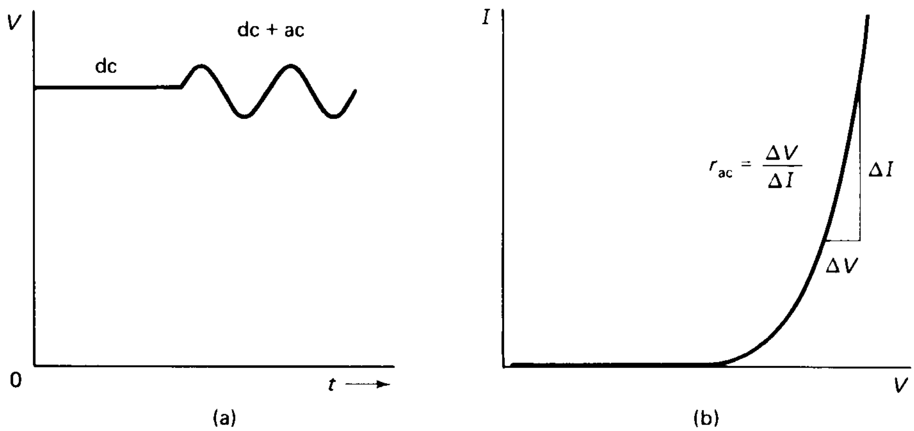


***FIGURE 8-13** (a) Ac signal superimposed on a dc level. (b) The resistance a diode presents to the ac component equals the slope ΔV/ΔI, which varies drastically with the dc level on which the ac rides.*

This ideal is not reached in actual transistors, whose dynamic emitter-junction resistance is more accurately given with values from 0.03 to 0.05 V substituted for Shockley's value in equation 8-6. I_J, of course, is the dc emitter bias current. To help develop a feeling for this relationship it may be well to contemplate that, to a small ac signal, the base–emitter junction of a transistor presents a resistance of about 40 Ω when the dc emitter current is 1 mA, dropping to 4 Ω if I_E is raised to 10 mA. The base supplies only $1/\beta$ of the total junction current, so the ac resistance presented from base to emitter is βr_j. For a transistor with $\beta = 100$ this means a base input resistance to ac of 4000 Ω for $I_E = 1$ mA, dropping to 400 Ω at $I_E = 10$ mA.

A stab at this concept may appear on the data sheet as the somewhat futile h_{ib} specification, which is hardly a function of the transistor at all, but rather of its emitter current. h_{ib} is the junction resistance which we have termed r_j, and h_{ie} is the base input resistance βr_j.

Remember that these "resistances" appear as such to small ac signals only. To dc and large ac signals, the base–emitter is a diode, dropping a fairly constant fraction of a volt.

8.8 OUTPUT RESISTANCE AND OBSCURE PROPERTIES

The ideal transistor has a collector current which does not depend upon collector voltage, but in practice I_C always increases a little with V_{CE}—more so in germanium than in silicon, and more at high currents than low currents. This voltage/current ratio represents a dynamic resistance from collector to emitter, which appears in parallel with the regular load resistance, and may be a real limiting factor in very high gain circuits. Data sheets specify the reciprocal of collector resistance as h_{ob}

for the common-base circuit or h_{oe} for the common emitter. The relationships between them are

$$r_c = \frac{1}{h_{oe}} = \frac{1}{h_{fe}h_{ob}} \tag{8-7}$$

The parameters vary widely for different collector voltages and currents, and can be compared only at identical operating points. Figure 8-14 illustrates r_c.

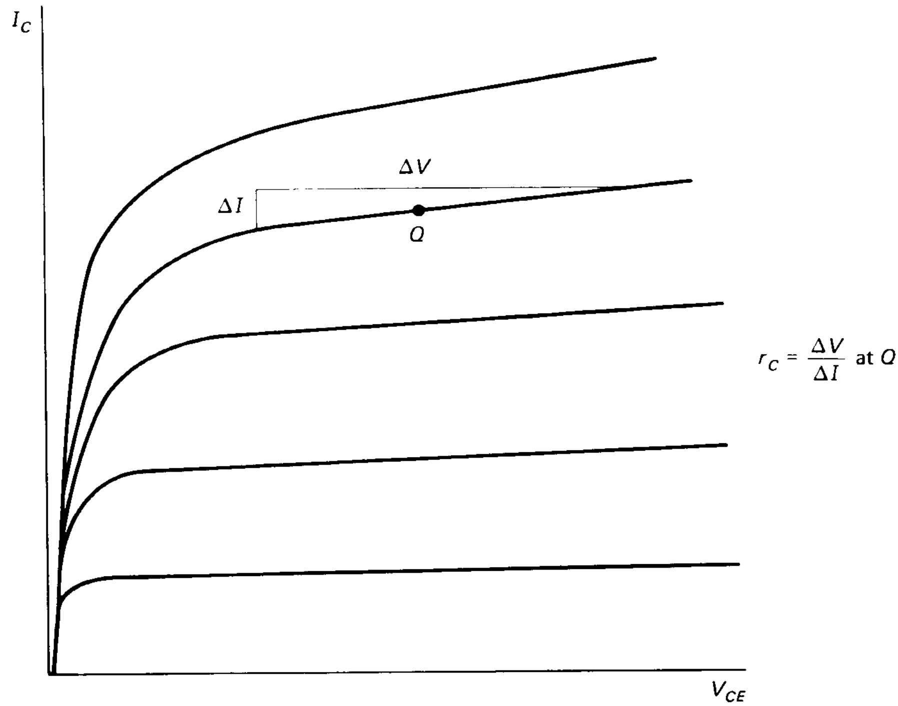


FIGURE 8-14 *Transistor collector dynamic resistance generally decreases at higher I_C and lower V_{CE}.*

Reverse-Voltage Transfer: Ideally, a transistor is a one-way device—input signals are fed to the output, but signals at the output are not fed back to the input. In fact, however, a change in collector voltage does reflect a slight in-phase change back to the input in the common-emitter and common-base modes. Typically, the reflected input voltage is on the order of 1/1000 to 1/10,000 of the output voltage, making the fed-back voltage negligible in almost all cases. The ratios are given by the parameters h_{re} and h_{rb}, respectively, in common-emitter and common-base configurations.

Odd-Quadrant Operation: Silicon *NPN* planar transistors are, of course, normally operated with positive voltages on both collector and base, but they will function with reduced performance when negative collector voltages are applied, as shown

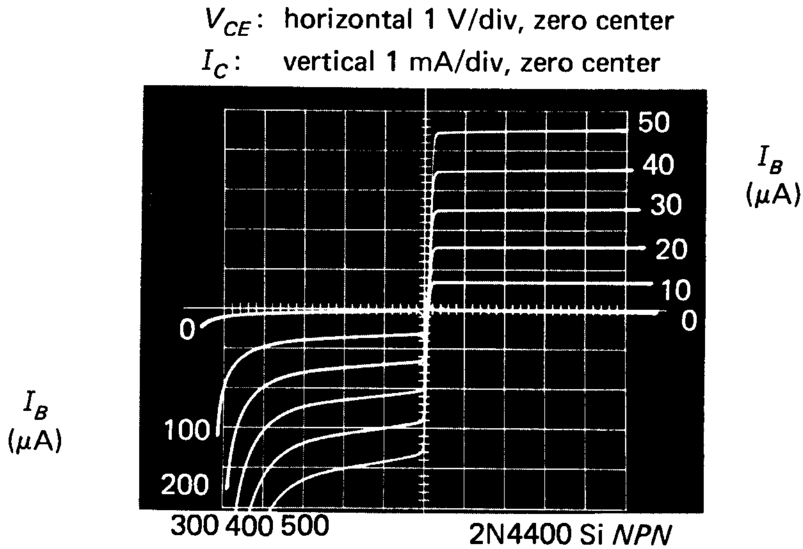


FIGURE 8-15 Quadrant 3 operation of an NPN *transistor: positive* I_B *but negative* V_{CE} *and* I_C. *Notice the* I_B *scale change.*

in Fig. 8-15. Beta is reduced from about 80 to 8, and the negative collector breakdown voltage is below 10 V, less than one-fifth of its forward value. The base voltage must still be positive. Negative base voltages simply turn the transistor off for both collector polarities.

8.9 SWITCHING-TIME LIMITATIONS

Transistors are often employed as high-speed switches, and the speed with which they respond to an input switching pulse is therefore of considerable importance. Switching-time limits are divided into four parts, as illustrated in Fig. 8-16. Rise time always refers to the increasing currents at turn-on, fall time to decreasing currents at turn-off, regardless of the slope of the voltage waveform involved. Rise and fall times are defined as the transition periods between 10% and 90% of full current. Hopefully, the rise and fall times of the input pulse will be much shorter than those of the transistor under test, but if not, it will be necessary to pick beginning and end points on the input pulse for reference.

Delay Time, t_d, is the time from the start of the input pulse until the collector current reaches 10% of its final value. To start the turn-on process, the base–emitter voltage must be raised from its turn-off level to approximately 0.6 V. This requires that the input current first charge the emitter–base capacitance, and any associated stray wiring capacitance, to this level. Shorter delay times can therefore be achieved by:

1. Choosing transistors and circuit arrangements for low base capacitances
2. Using low-resistance drive circuits (low R_1)

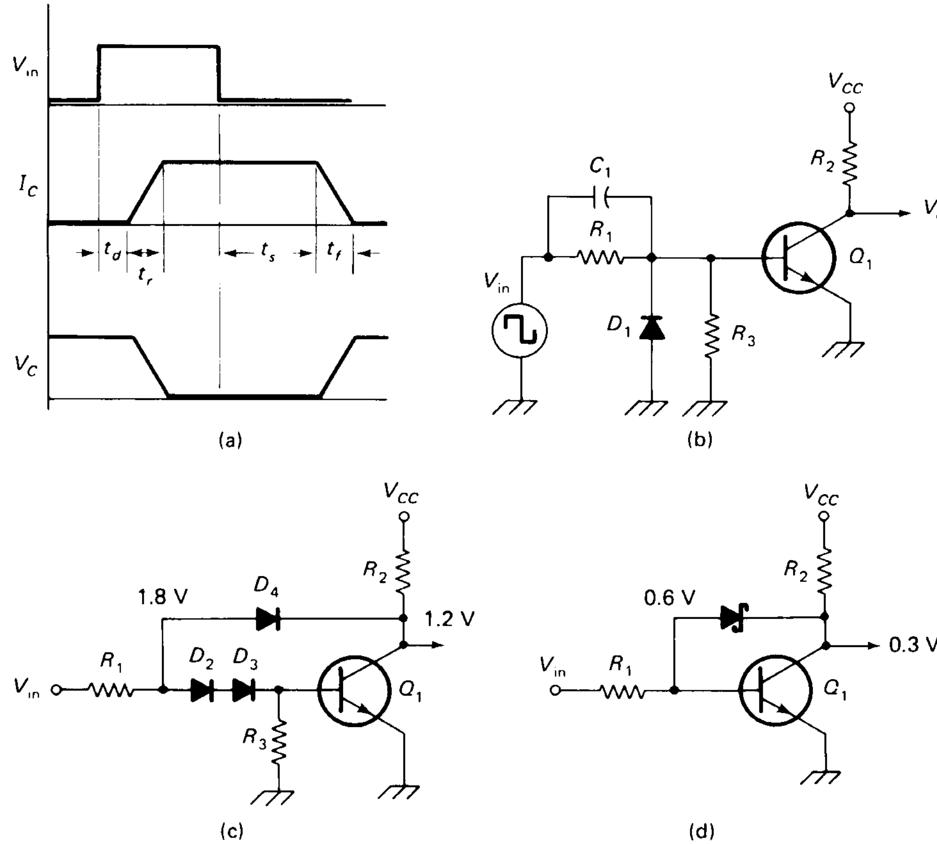


***FIGURE 8-16** (a) Transistor switching-time definitions. (b) Capacitor C_1 and resistor R_3 drain charge carriers from the base, reducing storage time. (c) Three-diode clamp to eliminate base saturation and stored charge. (d) Schottky-diode clamp.*

3. Shunting driving resistors with capacitors to increase initial charging of base (C_1)
4. Limiting the level of the *off* voltage to about 0.4 V, or at least preventing it from going negative (D_1)

Rise Time: When the transistor begins to turn on, the collector voltage will drop, and the collector–base capacitance will couple this drop back to the base, limiting its voltage rise. This capacitance must be discharged from $V_{C(\text{off})}$ to $V_{C(\text{on})}$ by the base input current before turn-on can be completed. Rise time can be shortened by:

1. Selecting transistors and circuit layouts for low collector–base capacitance
2. Using low-resistance drive circuits (R_1) and speedup capacitors (C_1)

Storage Time: In a simple transistor switching circuit, storage time t_s is usually much greater than the other components of switching time. Fortunately, storage time can be dramatically reduced by fairly simple circuit improvements. Storage time is a delay between the end of the base pulse and the beginning of collector turn-off. It is due to the fact that the base region of the transistor stores the extra charge carriers injected by base currents above the saturation requirement ($I_B = I_C/\beta$). These carriers must be diffused or drained away before the transistor can even begin to turn off. One way to minimize storage time is to provide a low-impedance path to ground, or better yet to a negative source, to drain off the stored charge. This can be accomplished by:

1. Keeping drive resistor R_1 (Fig. 8-16), and the off-state impedance of the source low.
2. Shunting the drive resistor with a speed-up capacitor, C_1.
3. Providing a resistance R_3 from the base to ground or to a negative supply.
4. Using a source pulse which swings negative, rather than simply zero in the off state. The resulting base discharge current is often termed I_{B2}, as opposed to the base turn-on current I_{B1}. I_{B2} can be calculated as $(V_{in(off)} + V_{BE})/R_1$.

Saturation Clamping: An even more effective method for reducing storage time is to eliminate the stored charge carriers by preventing the base current from exceeding the saturation requirements in the first place. This cannot be done reliably by choosing V_{in} and R_1 to produce exactly the required current I_C/β, because β varies so drastically from unit to unit and with temperature. Base current available must always be somewhat more than the maximum anticipated requirement, making it far in excess of the average requirement.

Figure 8-16(c) shows a three-diode saturation clamp which diverts all extra drive current through D_4. Assuming that each diode junction drops 0.6 V, D_4 begins to conduct when V_{CE} equals 1.2 V, routing all further drive current through the collector. Turn-off can therefore begin immediately upon removal of drive current. The minimum V_{CE} is, of course, 1.2 V—considerably higher than in an unclamped circuit. Eliminating D_3 will reduce this to 0.6 V, but this should be attempted only in low-current applications where $V_{CE(sat)}$ of the transistor is 0.3 V or less. In low-current applications where $V_{CE(sat)}$ of the transistor is 0.2 V or less, a Schottky diode can be used to clamp the collector at about 0.3 V, as shown in Fig. 8-16(d). This clamp is the basis for the 74S00 series of high-speed TTL integrated circuits.

Fall Time: t_f is controlled by the same factors as rise time and responds to the same treatments given previously.

Specifying Switching Times: Transistor manufacturers often combine t_d and t_r into a single specification t_{on}. Less commonly, t_s and t_f may be combined as t_{off}. In any case, the values given are valid only for the given test circuit, which must be specified down to the stray capacitances and inductances of the test fixture. Manufacturers seldom use a single standard test circuit even within the same company, let alone between companies, so switching-time specifications are generally worthless for comparison purposes. Anyone seriously interested in switching times is forced to build a test fixture representative of his application, obtain a fast pulse generator and 'scope, and start testing.

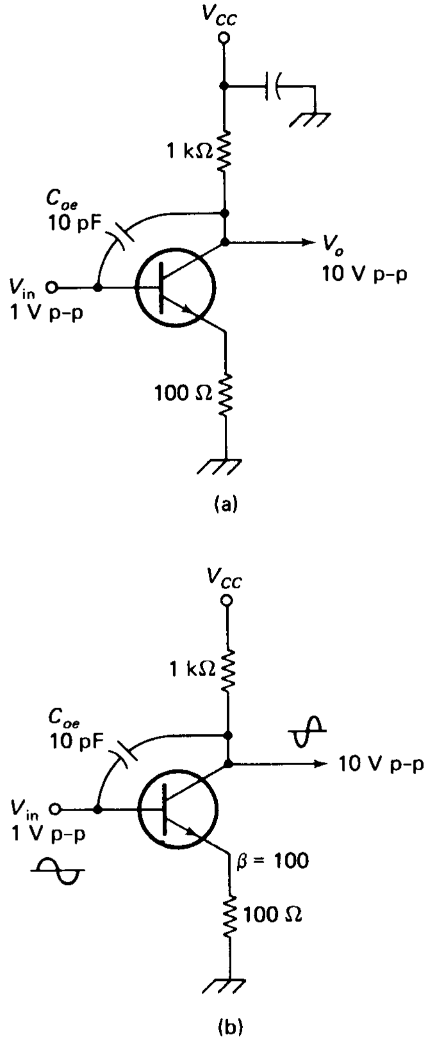
8.10 JUNCTION CAPACITANCE

Each of the leads of a transistor has a small capacitance to the other leads. The exact value depends greatly upon the bias voltages across the junctions, being perhaps 3 to 10 times greater at zero bias than at high reverse bias voltages. Again, manufacturers have established no standard bias voltages to facilitate comparisons, but seem to specify test voltages according to the phase of the moon. Two parameters are of predominant importance.

Input Capacitance: This is measured between the base and emitter at the reverse bias voltage specified. The test circuit is common base or common emitter, and the collector is an ac short or open circuit, depending upon the subscript letters of the symbol: C_{ibs}, C_{ibo}, C_{ies}, or C_{ieo}. The most commonly given of these, C_{ibo} or simply C_{ib}, is essentially the base–emitter junction capacitance. This is generally low enough to have a negligible reactance compared with the shunting resistance of the base–emitter junction.

Output Capacitance: This is measured between the collector and ground at the specified collector–base voltage. The test circuit is common base or common emitter and the base is an ac short or open to ground, depending upon the symbol: C_{obs}, C_{obo}, C_{oes}, or C_{oeo}. These parameters essentially represent the collector–base junction capacitance, which may determine the upper-frequency limit of an amplifier. Figure 8-17(a) shows how C_{oe} effectively shunts the load resistor (since the base voltage is near signal ground) and lowers the amplifier's gain at high frequencies.

Miller Effect: The collector–base junction capacitance is especially troublesome in high-frequency common-emitter amplifiers because its value is effectively multiplied by $(A_v + 1)$, as shown in Fig. 8-17(b). In the illustration A_v is 10, so the output signal is inverted from and 10 times larger than the input signal. The total voltage across C_{oe} is thus $(A_v + 1)$ or 11 times V_{in}, and the current through it is 11 times what would be expected if C_{oe} were connected from the input to ground. Another way of expressing this is to say that the effective C_{in} of the amplifier is $C_{oe}(A_v + 1)$, which in this example is 110 pF. At low frequencies this capacitance



$$f = \frac{1}{2\pi CX} = \frac{1}{2\pi \times 10\text{pF} \times 1\text{ k}\Omega} = 16\text{ MHz}$$

$$Z_{in}(\text{lo-}f) \approx \beta R_E \approx 100 \times 100 \approx 10\text{ k}\Omega$$

$$Z_{in}(1\text{ MHz}) = 10\text{ k}\Omega \| X_{C_{in}}$$

$$C_{in} = C_{oe}(A_v + 1) = 110\text{ pF}$$

$$X_{C_{in}} = 1.4\text{ k}\Omega$$

$$Z_{in}(1\text{ MHz}) \approx 1.4\text{ k}\Omega$$

FIGURE 8-17 *(a) Collector output capacitance loads the output when $X_{C_{oe}} = R_C$. (b) The Miller effect makes C_{oe} look $(A_v + 1)$ times larger than it is, lowering amplifier Z_{in}.*

has a negligibly high reactance, but at higher frequencies it lowers the amplifier's input impedance, making it difficult to drive.

8.11 TRANSISTOR FREQUENCY LIMITS

At high frequencies the current gain of a transistor begins to drop off in a very predictable way, as illustrated by the curve of Fig. 8-18. There are only two variables that distinguish one transistor's curve from another's: low-frequency

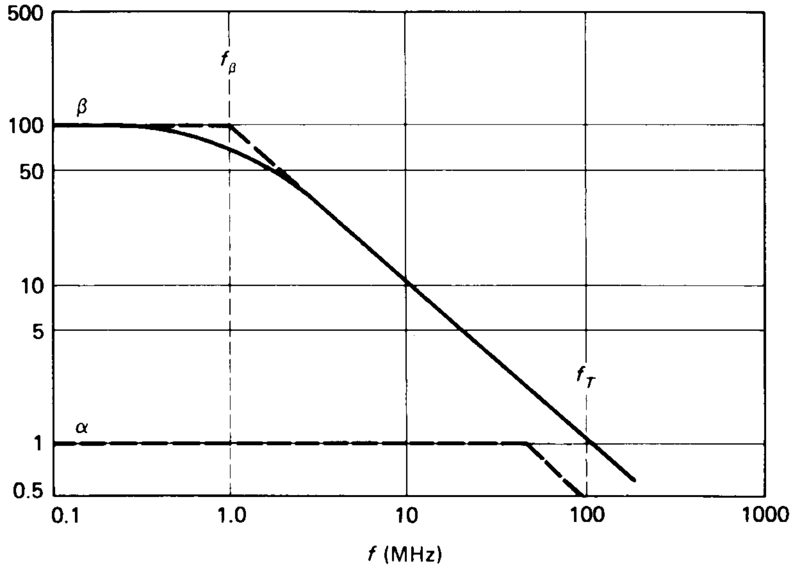


FIGURE 8-18 A transistor's gain begins to drop at f_β, which equals f_T/β, and reaches unity at f_T.

current gain β and unity-gain frequency f_T. The slope of the curve in the roll-off region is always -20 dB/decade, which is to say that the gain drops by a factor of 10 for every increase in frequency by a factor of 10. The product of current gain and frequency is constant and equal to f_T in the roll-off region, so f_T is termed "gain-bandwidth product." If f_T is known, current gain h_{fe} at any frequency f can be calculated using this relationship:

$$h_{fe} = \frac{f_T}{f} \tag{8-8}$$

provided that h_{fe} can never exceed the low-frequency current gain β.

The frequency at which h_{fe} begins to drop (the -3-dB point) is termed f_β or f_{hfe}, and can be calculated as

$$f_\beta = \frac{f_T}{\beta} \tag{8-9}$$

The upper-frequency limit of a transistor in a common-emitter circuit lies somewhere between f_β and f_T and depends primarily upon the gain of the stage and the impedance of the driving source. As h_{fe} drops off, the effect is not an actual decrease in A_v but a decrease in Z_{in} of the amplifier. As long as Z_{in} can be maintained larger than the driving impedance Z_s, overall gain will not suffer. This can be achieved by using an emitter follower to provide a "stiff" driving source or by leaving a large unbypassed emitter resistance to keep Z_{in} high.

Common-Base Frequency Limits: The gain of a common-base amplifier depends not upon β, but upon α, which is defined as

$$\alpha = \frac{I_C}{I_E} = \frac{\beta}{1+\beta} \tag{8-10}$$

Alpha (α) is near unity at low frequencies and is not reduced to 0.707 (-3 dB) until β drops to about 2.4. The alpha cutoff frequency (f_α or f_{hfb}) for common-base amplifiers can be expected to be on the order of $f_T/2.4$. This is generally quite a bit higher than the cutoff for common-emitter amplifiers and accounts in part for the popularity of the common-base circuit at high frequencies.

8.12 FIELD-EFFECT-TRANSISTOR ACTION

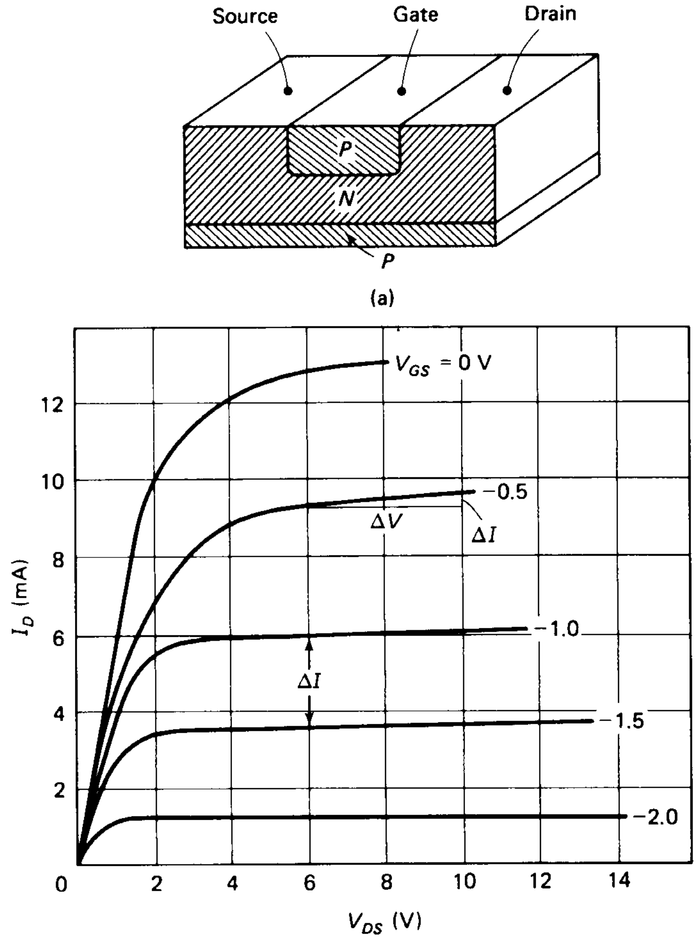
Junction FETs: Field-effect transistors use an input voltage to control an output current. There is practically no input current except that due to stray capacitive reactance at high frequencies. The operation of an FET is quite easy to understand. A channel of *N*-doped silicon is left beneath a *P*-doped covering bar called the gate, as shown in Fig. 8-19(a). Normally, the channel has a resistance of a few hundred ohms, until a saturation current is reached, beyond which voltage increases produce little increase in channel current. However, if a negative voltage is applied to the gate, the *P-N* junction is reverse-biased, and the area around the junction becomes depleted of charge carriers as described in Section 7.1. The *N* channel thus becomes partially constricted; its resistance increases and its saturation current decreases. Further reverse bias causes further current restriction until the channel is finally "pinched off" entirely. Figure 8-19(b) shows a typical set of FET characteristic curves representing this process.

Notice that the elementary FET is symmetrical—the choice of source and drain is arbitrary and reversible. More important, notice that the input element is a reverse-biased diode, so the only input current is that due to leakage—typically 1 nA at room temperature and 1 μA at 150°C.

Like most semiconductors, the junction FET is available in complementary forms, in this case *N*-channel and *P*-channel. The *P*-channel devices operate with negative drain voltages and positive gate voltages with respect to the source.

MOSFETs: If a very thin layer of silicon dioxide, which is a nearly perfect insulator, is placed between the gate electrode and the channel of an FET, the leakage current is reduced by a factor of 100 to 1000, and in addition, it becomes possible to forward-bias the gate without causing gate current. When forward-biased, the gate increases or *enhances* current from source to drain. This is in contrast to the *depletion* effect of reverse bias. These insulated-gate FETs may be termed IGFETs, or MOSTs (metal-oxide silicon transistor), but the most common verbal appellation is MOSFET (say *moss-fet*).

MOSFETs are available in two types. Depletion-enhancement or Type B MOSFETs are normally partially conductive and can made to conduct more or less



$$y_{fs} = \frac{\Delta I_D}{\Delta V_{GS}} = \frac{(6-3.5)\text{mA}}{(1.5-1.0)\text{V}} = 5 \text{ mS}$$

$$r_{oss} = \frac{\Delta V_{DS}}{\Delta I_D} = \frac{(10-6)\text{V}}{(9.6-9.2)\text{mA}} = 10 \text{ k}\Omega$$

(b)

***FIGURE 8-19** (a) Junction-FET structure; (b) Typical characteristic curves.*

current by forward or reverse gate bias, respectively. Enhancement-mode or Type C MOSFETs are normally nonconductive from source to drain and must be rendered conductive by forward gate bias. The Type A or depletion-mode-only FET is not designed for forward gate bias and is represented primarily by the junction FET. Figure 8-20 shows the schematic symbols for various types of FETs. The substrate lead of the MOSFETs is generally connected to the source. The dual-gate FETs are especially suited for use as signal mixers.

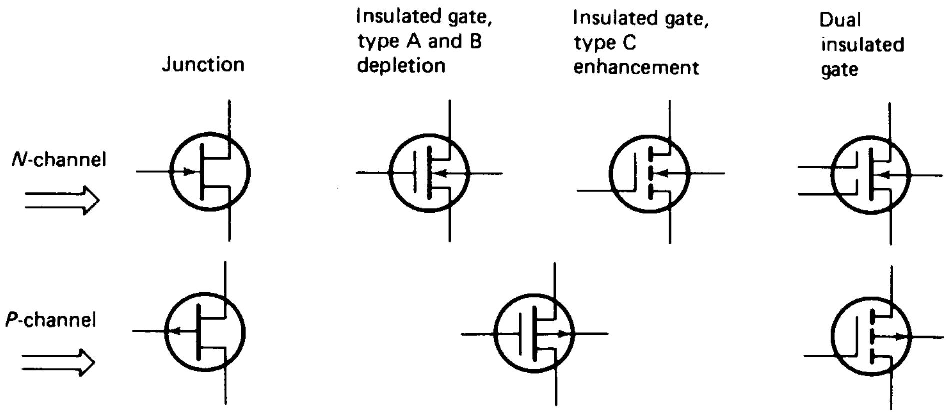


FIGURE 8-20 Various types of FET.

FET versus Bipolar: Both the FET and the bipolar transistor have their advantages, and the choice between them will depend upon the characteristics required. On the plus side for the FET:

1. It has a much higher input resistance (> 100 MΩ compared to < 100 kΩ in typical circuits). Thus it is the first choice for voltmeter and oscilloscope input stages. This advantage disappears above a few MHz, where stray capacitance takes over.

2. It generates fewer and quite predictable distortion products when subjected to large input-signal swings. This has made it popular in radio-receiver RF amplifiers and mixers.

3. It is much less sensitive to radiation, making it highly desirable in military and space applications.

4. It is easier to produce in integrated-circuit form—requires fewer production steps and generally takes less chip area.

5. At low voltage levels the drain–source channel is essentially a linear resistor whose value varies with gate bias. No input (gate) current is injected into the output current path. This makes the FET more suitable for signal-chopping and control applications.

In favor of the bipolar transistor:

1. Bipolars are still less expensive and more readily available in a wide range of power levels, voltage levels, and package styles.

2. Bipolars can operate easily from a single supply of 1.5 V or less. FET amplifiers can be made to work on 5-V supplies, but they are much more comfortable with 10 or 20 V.

3. Bipolars can turn on "harder." Output saturation voltages of a few tenths of a volt versus a few volts for FETs are typical at 10- to 100-mA output currents.

4. Bias points are easier to stabilize with bipolars. V_{BE} is not likely to lie outside the range 0.5 to 0.9 V for any but brute-power transistors. By contrast, a typical low-power JFET lists V_{GS} for I_D of 0.4 mA as -1.0 V min, -7.5 V max. This means that we must tolerate poor bias stability, operate with high supply voltages, use current-source or dc-feedback circuitry to stabilize bias, or resort to selecting FETs for the desired V_{GS}.
5. Voltage gains of 100 or more per stage are easily realized with resistive-coupled bipolar amplifiers. A similar FET stage would yield a gain of perhaps 10.
6. An FET's drain current is a function of the *square* of gate voltage. This means that gain is higher at higher bias currents [note the unequal spacing of the V_{GS} lines in Fig. 8-19(b)]. The result is distortion of large-swinging signals.
7. Some MOSFETs can be destroyed by a touch of the finger due to static electric charges. The caution required around these devices is explained in the next section. Many MOSFETs are protected against this hazard by integrated avalanche diodes.

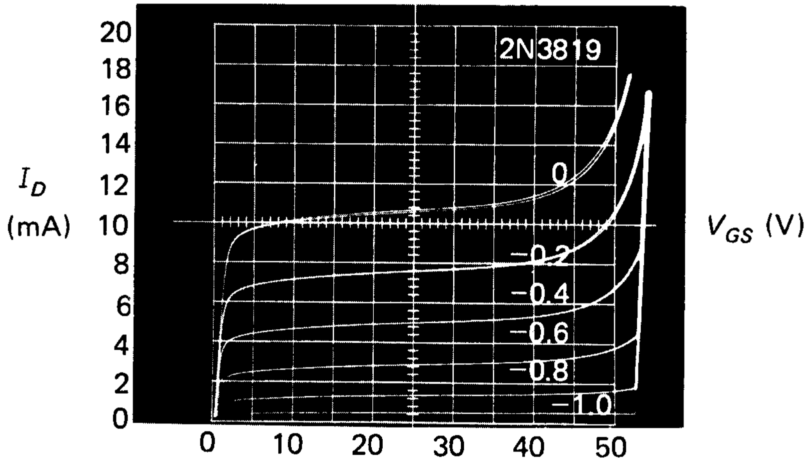
8.13 FET PARAMETERS

Maximum Power and Voltage: FET ratings of maximum internal temperature T_{OPR}, thermal resistance to case θ_{J-C}, and device dissipation P_T are similar to those for bipolar devices.

Avalanche breakdown is guaranteed not to occur below BV_{DS} for drain to source. Figure 8-21(a) shows avalanche breakdown, which is nondestructive if drain current is limited by external resistance.

For a junction FET the maximum forward gate–source voltage is that of a conducting silicon diode—about 0.6 V. The reverse limit, V_{GSS}, is the avalanche voltage of the gate–channel junction and is typically between -25 and -75 V. Care must be taken that this limit is not exceeded, or that a very high impedance appears between the gate and any possible higher-voltage source, as a few milliamperes at this voltage could destroy the gate.

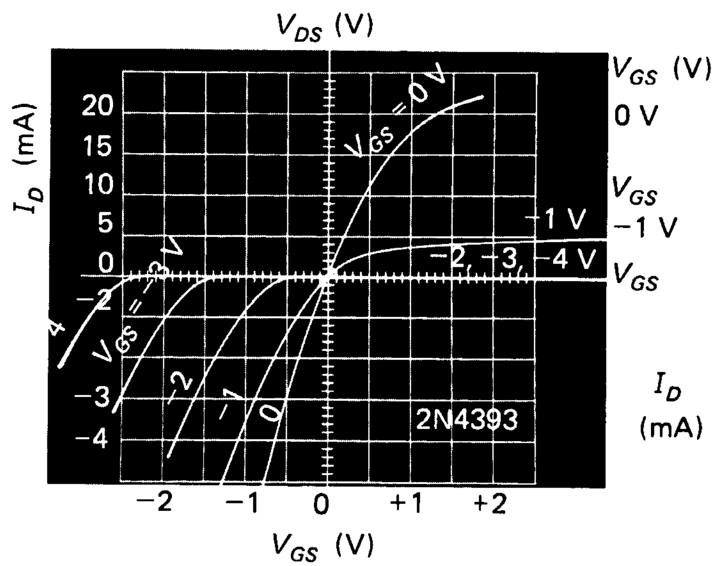
MOSFETs have a somewhat different gate-voltage limit. The gate-insulating layer is of necessity extremely thin, and can be arced by voltage on the order of 100 V or less. Static voltages far beyond this are commonly present on ungrounded metal objects, clothing, plastic table and chair surfaces, and so on. The destructive mechanism here is an actual capacitive-dielectric arc, not a semiconductor avalanche, and the damage is generally irreversible. Some MOSFETs have a pair of back-to-back zener diodes integrated from the gate to the source, but this protective measure adds stray capacitance, and the leakage current of the zeners reduces the input resistance of the insulated gate (although it is still many orders of magnitude higher than bipolar input resistances). Unprotected MOS devices are shipped with protective clips or pads shorting the leads to prevent destruction of



$$\text{At } V_{DS} = 25\text{ V}, I_D = 6\text{ mA: } y_{fs} = \frac{\Delta I_D}{\Delta V_{GS}} = \frac{(7.6\text{–}5.0)\text{mA}}{(0.4\text{–}0.2)\text{V}} = 13\text{ mS}$$

$$r_s = \frac{\Delta V_{DS}}{\Delta I_D} \quad V_{GS} = \text{k} \quad = \frac{20\text{ V}}{0.5\text{ mA}} = 40\text{ k}\Omega$$

(a)



(b)

***FIGURE 8-21** (a) FET characteristic curves showing breakdown voltage BV_{DS}. (b) At low reverse drain voltages, r_{DS} is controllable by V_{GS}.*

the gates by static charges during handling. These should be left in place until the device is in the circuit ready for operation. Servicing of instruments containing such devices should be done on grounded metal benches with grounded soldering pencils. Operators should wear a discharging strap to ensure that their bodies do not acquire a static charge. For safety to the operator there should be a 1-MΩ resistor between the strap and earth ground.

DC Parameters: Leakage current from the gate to the source with the drain grounded is termed I_{GSS}. A maximum specification is made at a stated temperature and gate-source voltage.

$V_{GS(\mathrm{off})}$, also termed pinchoff voltage V_p, denotes the voltage that must be applied from gate to source to effectively "pinch off" the channel. The point of zero drain current is rather indefinite, so an effective *off* current of 10 or 100 μA is commonly used. This parameter often varies drastically from unit to unit, so a maximum guaranteed value is always specified, with typical and minimum values sometimes included. V_{DS} is specified well above the saturation region for this test.

I_{DSS} is the drain current with the gate shorted to the source. V_{DS} is specified well above saturation. A maximum value for this zero-bias drain current is always given, with typical and minimum values sometimes included.

Drain current I_D and gate bias V_{GS} are related as follows:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(\mathrm{off})}}\right)^2 \tag{8-11}$$

Note that drain current is a function of the square of gate voltage, as indicated by the progressively closer spacing of the curves in Fig. 8-21(a).

Enhancement-mode MOSFETs use a turn-on threshold-voltage specification in lieu of $V_{GS(\mathrm{off})}$ and I_{DSS}. This is V_{TH}, the forward gate-source voltage which produces a specified threshold drain current (usually 10 or 100 μA) at a given V_{DS}.

Gain Specifications: The drain-current change caused by a given small gate-voltage change (V_{DS} being held constant) is a measure of the gain of an FET and is termed mutual conductance g_m, forward transconductance g_{fs}, or forward transadmittance y_{fs}. This specification is always made with a high V_{DS}, as its value decreases severely in the saturation region. The value of y_{fs} also depends on bias point, being largest at high drain currents and approaching zero at pinchoff. In particular:

$$y_{fs} = \frac{2I_{DSS}}{V_{GS(\mathrm{off})}}\left(1 - \frac{V_{GS}}{V_{GS(\mathrm{off})}}\right) \tag{8-12}$$

Figure 8-19(b) shows how y_{fs} can be measured from a characteristic-curve display. This parameter is also liable to vary considerably from unit to unit, and minimum guaranteed values are generally given on the data sheet.

Output Resistance: The drain of an FET presents a resistance to any ac signal appearing there because a drain-voltage change causes a drain-current change, V_{GS} being held constant. This is illustrated in Fig. 8-19(b). This dynamic output resistance, called r_{oss}, appears in parallel with whatever external load is connected to the drain. If the load resistance is very high, r_{oss} may become the limiting factor in amplifier gain.

The output resistance decreases drastically as V_{DS} approaches the saturation region. At $V_{DS} = 0$ the parameter is called $r_{ds(\text{on})}$.

Stray Capacitance: A capacitance of typically a few pF appears from the gate to source (C_{gs}) and from the gate to drain (C_{gd}) of an FET. Manufacturer's data sheets often specify total input capacitance C_{iss}, which is the sum of $C_{gd} + C_{gs}$, and reverse-transfer capacitance C_{rss}, which equals C_{gd}. Actual input capacitance of an amplifying circuit may be many times larger than C_{iss} because of the Miller effect, as explained in Section 8.10.

Control Applications: As shown in Fig. 8-21(b), FETs have a drain-to-source resistance which transitions smoothly from a few hundred to many thousands of ohms if V_{DS} is kept in the range of ± 1V or so. The FET can thus be used as an electrically-operated variable resistor, with gate voltage determining resistance. Figure 8-22(a) shows a simple automatic gain control for an audio preamp utilizing this property. The control input is dc-rectified and filtered from the audio output. The dc becomes more negative at high audio levels, thus automatically turning the gain down. In another application, the control signal could be derived from a microphone picking up crowd noise in a lobby. Less negative dc would be delivered for higher crowd noise levels, thus automatically increasing the gain of the PA amplifier.

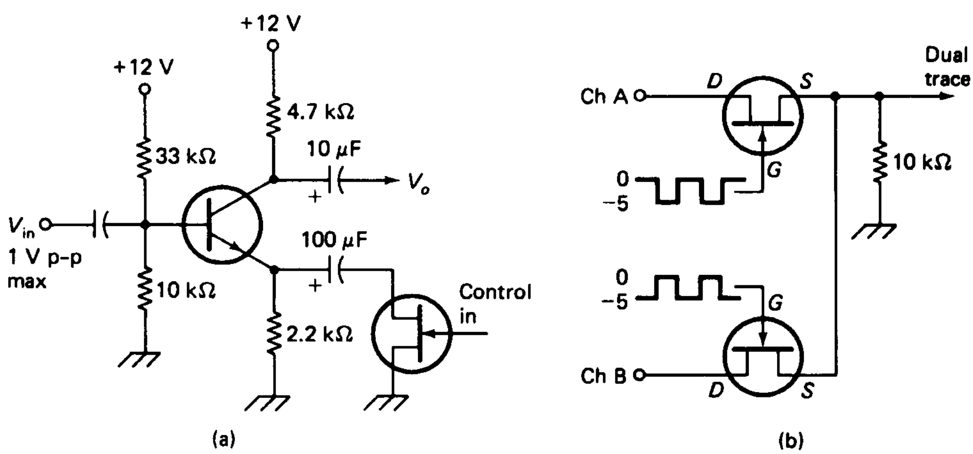


FIGURE 8-22 (a) FET used as an electrically variable resistance to control the gain of a bipolar amplifier. (b) Pair of FETs used to switch two inputs alternately to a single line.

Figure 8-22(b) shows a simplified dual-trace oscilloscope chopper. The gates are fed from the oppositely phased outputs of a multivibrator, turning on alternately the channel A and channel B FETs. Since the gates are never forward-biased, the chopping square wave is not passed to the output. The chopping frequency must be kept reasonably low, however, to prevent C_{GS} from passing the chopping signal to the 10-kΩ resistor. The A and B input signals must not become large enough to interfere with the on–off switching signal. For JFETs, this limits the inputs to a few tenths of a volt. MOSFETs will handle much larger signals, since the switching square waves can be of both polarities (say, ± 10 V) without causing the gate to conduct. Integrated circuits containing several of these MOS bilateral analog switches are available.

9

TRIGGERED SEMICONDUCTOR DEVICES

9.1 THE UNIJUNCTION TRANSISTOR

Next to the diode and the bipolar transistor, the UJT is the oldest of the semiconductor devices. It is a switching device with two stable states: *on* and *off*. The main current path is from emitter to base 1, as shown in Fig. 9-1. This path is normally an open circuit, but it switches to a near short circuit if the emitter–base-1 voltage reaches a critical percentage of the interbase voltage V_{B2-B1}. This percentage is determined by the intrinsic standoff ratio η (eta), a property of the UJT which generally lies between 0.85 and 0.55:

$$V_{E-B1(\mathrm{crit})} = \eta V_{B2-B1} + V_D \tag{9-1}$$

where V_D is the silicon-diode drop, approximately 0.6 V.

The similarity of the UJT symbol to the junction FET symbol is unfortunate, because the two devices have almost no similarities from an operational point of view. The UJT is not usable as an amplifier, since it triggers abruptly from *off* to *on* and has no continuous range in between as do bipolars and FETs.

UJT Oscillator: The circuit of Fig. 9-1, in one form or another, probably encompasses more than three-fourths of all UJT applications. It is an *RC* oscillator whose advantages include simplicity, frequency stability with temperature and

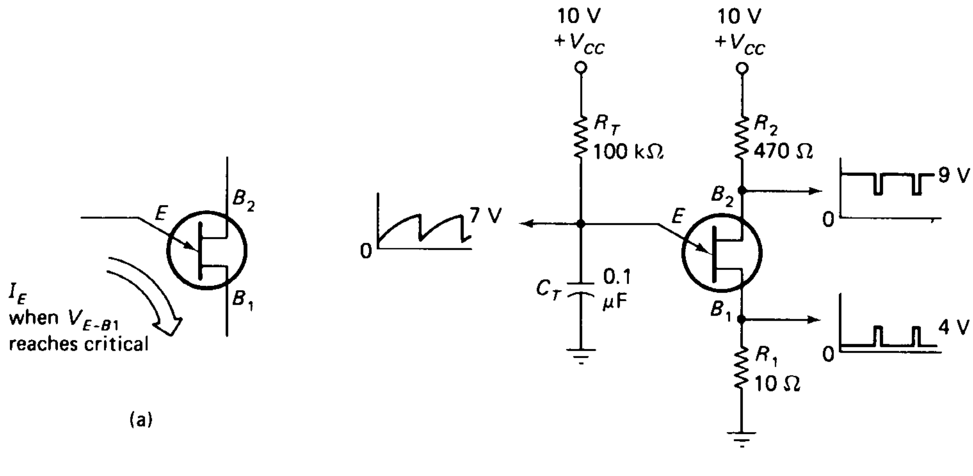


FIGURE 9-1 (a) Schematic symbol, lead identification, and main current path for the unijunction transistor. (b) Basic UJT oscillator with representative waveforms at the three terminals.

supply-voltage change, and sawtooth and pulse-shaped outputs. The operation of the circuit is as follows:

1. Timing resistor R_T charges C_T toward V_{CC} according to the familiar RC-time-constant formula. The UJT is *off*, so the emitter draws no current.
2. When the emitter voltage reaches the critical level, the emitter to B_1 path becomes a very low resistance, and C_T discharges through R_1, producing a positive pulse across R_1. The voltage across C_T drops much more quickly than it rose, because it discharges through R_1 which is much smaller than R_T.
3. When the UJT turns *on*, the interbase resistance drops, causing a negative-going pulse at B_2. For a typical UJT in the circuit shown, this may consist of a drop from 9 V to 6 V.
4. When the capacitor is nearly discharged, the emitter current drops below a minimum hold-on value, and E–B_1 returns to its normal open-circuit state, allowing another charging cycle to begin.

Outputs may be taken from any of the three UJT leads, provided that the impedance of the load is several times the value of the associated circuit resistor. This imposes a stringent requirement on the sawtooth output at the emitter because R_T is usually a high value.

It may be difficult to obtain more than a few volts at B_1 because of internal-device resistance. R_1 can be eliminated if the positive pulse output is not required.

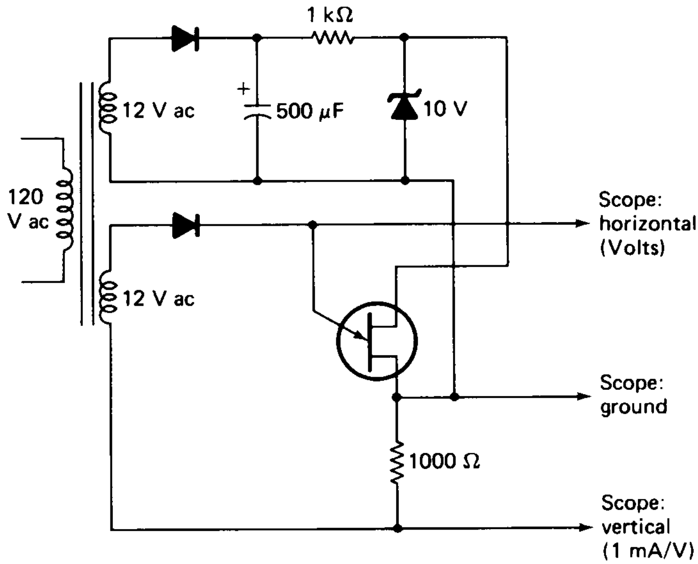


FIGURE 9-3 *UJT curve tracer adapter for use with a conventional dc-coupled oscilloscope.*

minimum or valley current required to maintain the device in the *on* state. This may be 10 mA or so for bar-type down to 1 mA for planar-type UJTs. In the oscillator circuit it is essential that the timing resistor not supply enough current to hold the UJT in the *on* state with V_{E-B1} at the valley voltage.

Using these facts, we can develop formulas for the maximum and minimum values of the timing resistor in the UJT oscillator circuit:

$$R_{T(\max)} = \frac{V_{CC} - V_{\text{pk}(\max)}}{I_{\text{pk}(\max)}} \tag{9-3}$$

$$R_{T(\min)} = \frac{V_{CC} - V_{val(\min)}}{I_{val(\min)}} \tag{9-4}$$

V_{CC} is the supply voltage. Note that these formulas contain no safety factors. Good design practice would keep R_T between half the maximum and twice the minimum calculated values.

Programmable UJTs or PUTs offer certain advantages over the bar and planar UJTs. Figure 9-4 shows the comparison. Internally, they are four-layer *PNPN* devices similar to the SCR (Section 9.3). Externally, they behave like a conventional UJT, with these differences:

1. There is no intrinsic η factor. The device turns *on* when V_A becomes greater than V_G by an amount specified as offset voltage V_T, a value that may lie between 0.2 and 1.6 V.

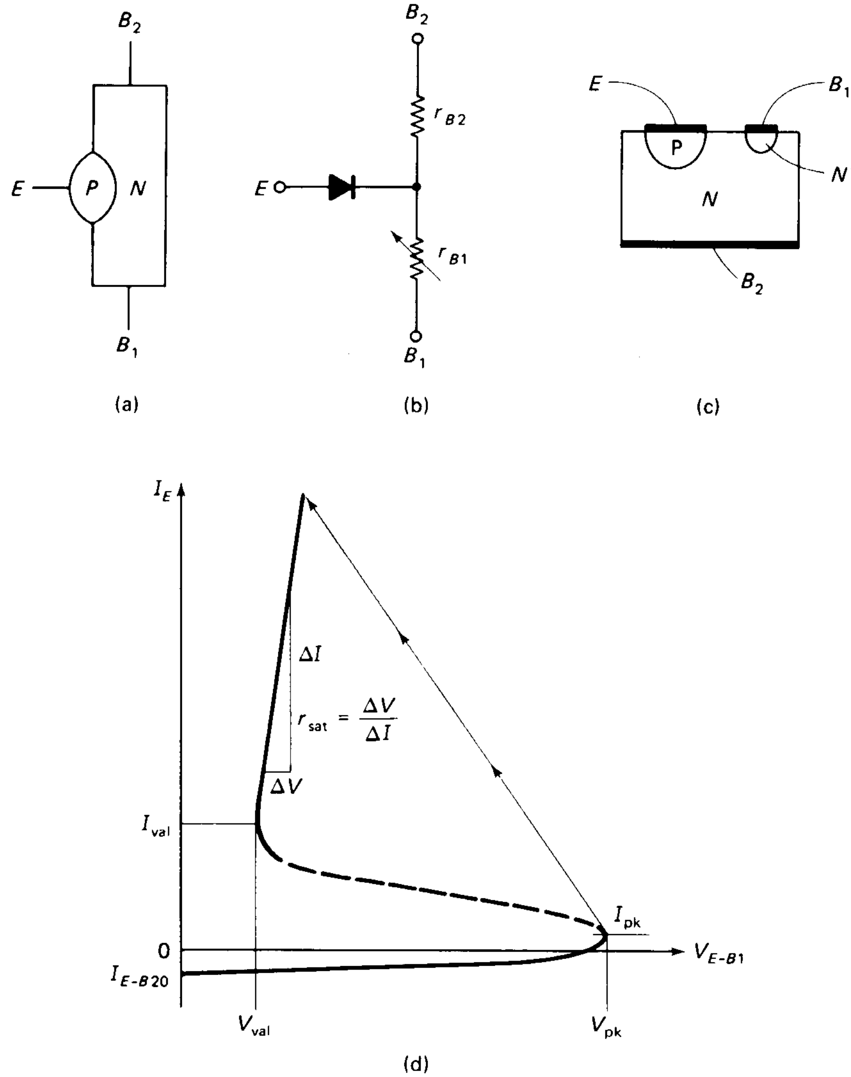


***FIGURE 9-2** (a) Bar-structure UJT. (b) UJT equivalent circuit. (c) Planar-structure UJT. (d) UJT characteristic curves.*

supply adequate firing current (I_{pk}) with the emitter at the threshold of firing (V_{pk}). To maintain timing accuracy and waveform purity at the emitter, I_{RT} must be many times larger than I_{pk}.

Upon firing, the UJT switches very rapidly to the *on* state, as indicated by the dashed line. This transition cannot be displayed accurately by the curve tracer, but that is not important. V_{E-B1} is typically between 1 and 2 V in this region, rising slightly with rising emitter current. The critical parameter of this *on* region is the

The circuit will function without R_2, but it has two other functions besides producing a negative-pulse output. The failure rate for semiconductors is much higher than for resistors, and without R_2 an interbase short amounts to a destructive power-supply short. R_2 also provides frequency stability against temperature variations by compensating for decreases in the V_{E-B1} diode drop at elevated temperature. The interbase resistance r_{BB} rises at high temperature, increasing V_{B2-B1}, which tends to increase the firing voltage. With proper selection of R_2, this effect will just offset the decrease in emitter-diode drop, keeping the firing voltage constant. R_2 is determined by

$$R_2 = \frac{0.7r_{BB}}{V_{CC}} + \frac{(1-\eta)R_1}{\eta} \tag{9-2}$$

UJT Characteristics and Limitations: Figure 9-2(a) shows the structure of a bar-type UJT. The bar is N-doped to have a resistance generally between 1 and 10 kΩ. A P-doped pellet is alloyed to the side of the bar, its position determining the firing voltage-to-interbase voltage ratio η. A schematic representation is given in Fig. 9-2(b). The ratio $r_{B1}/(r_{B1}+r_{B2})$ equals η. Upon firing, r_{B1} decreases drastically. The emitter diode has a drop of about 0.55 V at room temperature for typical UJT firing currents. Figure 9-2(c) shows a planar-UJT structure which generally has lower reverse leakage in the emitter diode than the bar type.

A complete description of a UJT's parameters is given by its characteristic curve of I_E versus V_{E-B1}, shown in Fig. 9-2(d) with the low-current range expanded for clarity. This curve can be displayed on a conventional transistor curve tracer if a separate +10-V supply is connected from B_2 to B_1. The "collector" terminal feeds the UJT emitter, and the "base" step output is unused. The circuit of Fig. 9-3 can be constructed to display the curve on a standard oscilloscope. The curves will be upside down, but this should present no problem.

Following the curve from the lower left, I_{E-B2O} is the reverse leakage current through the emitter diode when V_E is zero and V_{B2} is the +10 V. Its value is typically a few nanoamperes at 25°C, remaining less than 1 μA even at maximum operating temperature for planar UJTs. Bar-structure UJTs may have several μA leakage, even at 25°C. This current must be small compared to the current supplied by the external timing resistor if timing and waveshape at the emitter are to be unaffected by leakage.

As emitter voltage is increased, the emitter diode becomes forward-biased and forward current begins to flow. The current required to fire the unijunction is called the peak current I_{pk}, and V_{pk} is the firing voltage. The standoff ratio can be obtained from V_{pk} as

$$\eta = \frac{V_{pk} - V_D}{V_{BB}}$$

Peak current is typically in the neighborhood of 10 μA for bar UJTs, down to 1 μA or less for planar types. It is essential that the timing resistor be able to

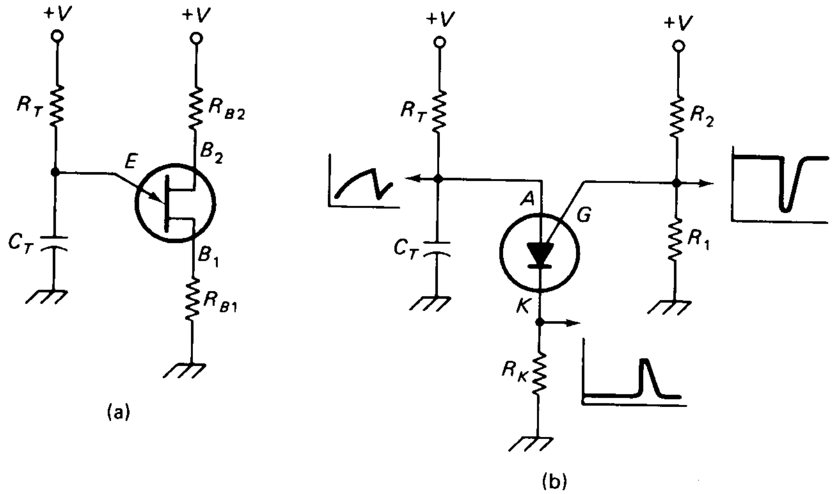


FIGURE 9-4 *Comparison of the basic oscillator for the conventional UJT (a) and PUT (b). R_1 and R_2 determine the firing voltage $V_{A(crit)}$.*

2. The firing voltage is selectable or "programmable" by adjusting the gate voltage with a voltage divider R_1–R_2. (Actually, the term *program* refers to a sequence of events, not a selection of a single event, but we seem to be stuck with this perversion of the word.) The firing-voltage variation thus depends upon resistor tolerance, not semiconductor tolerance, and can be held at 10% or 2% from unit to unit if you are willing to pay for the precision resistors (5% or 1% tolerance, respectively). This is a great improvement over the conventional UJT. The gate leakage current I_{GAO} is typically 10 nA at 25°C, increasing to around 100 nA at maximum temperature.

3. Valley current I_V is a function of the Thévenin equivalent of $R_1 \| R_2$ (termed R_G) and, to a lesser extent, of gate voltage V_G (also termed V_S). There seems to be no formula relating I_V to R_G, but for one PUT type I_V was 300 μA at $R_G = 10$ kΩ, decreasing to $I_V = 18$ μA at $R_G = 1$ MΩ. Other PUTs have I_V as low as 25 μA at $R_G = 10$kΩ. These values are in general much lower than for conventional UJTs, dictating a much higher minimum value of R_T.

4. The turn-on time is generally somewhat less for a PUT than for a UJT. PUT oscillators are nevertheless limited to frequencies below about 50 kHz because of the low hold-on current and consequent high minimum R_T. Conventional UJTs can oscillate up to about 500 kHz.

5. The *on*-state voltage drop (V_{AK}, also termed V_F) is typically about half the $V_{E\text{-}B1}$ of a conventional UJT.

6. The gate is brought forcibly to within a few tenths of a volt of ground in the *on* state, producing a very respectable negative-going pulse output. By

contrast, the conventional UJT brings B_2 to only a few volts above ground through an impedance of perhaps a few hundred ohms.

7. Leakage and firing currents I_{GAO} and I_{pk} are similar or slightly lower than those of the UJT.

Complementary Unijunction Transistors (CUJT) are available, although their popularity seems much less than that of UJTs and PUTs. Internally, they consist of a four-layer *NPNP* sandwich (complementary to the *PNPN* layers of the PUT) with integrated resistors R_1 and R_2 determining the intrinsic standoff ratio η. CUJTs can be applied in the standard circuit on a negative supply, or inverted on a positive supply, as shown in Fig. 9-5.

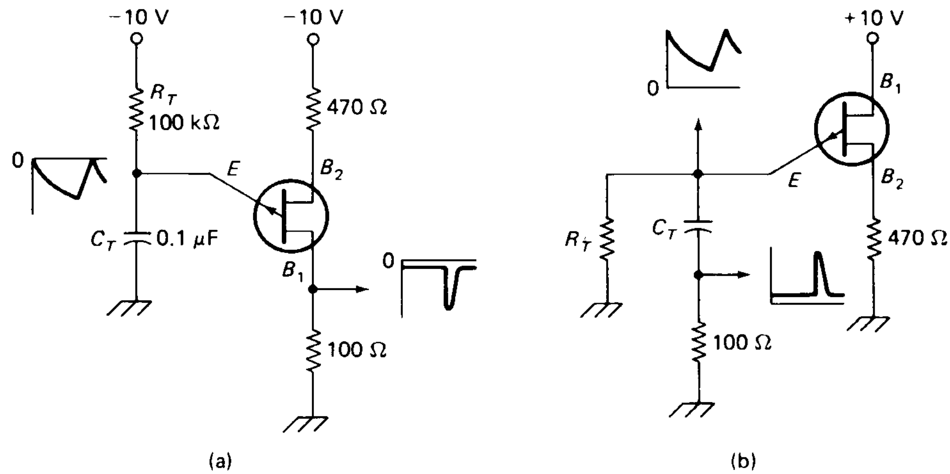


***FIGURE 9-5** Oscillator circuits for the complementary UJT: (a) with negative supply; (b) positive supply.*

UJT Circuit Applications: UJTs are used as oscillators and sweep generators with frequencies from one cycle every few minutes up to a few hundred kHz, as timers over the same range, as voltage-sensitive triggers, as pulse generators (especially for triggering power devices such as SCRs), and as simple frequency dividers (by relatively small integers, such as 7 kHz to 1 kHz). Figure 9-6 shows a number of these applications.

9.2 THE 555 TIMER

Since its introduction in 1972, the *NE 555* timer IC has achieved such widespread industry acceptance that we feel justified in treating it as a standard component. Although called a timer by most manufacturers, the *555* is a general-purpose device with additional applications as an oscillator, free-running or triggered pulse generator, ramp generator, frequency divider, and frequency, pulse-width, or pulse-position-modulated oscillator, to name a few.

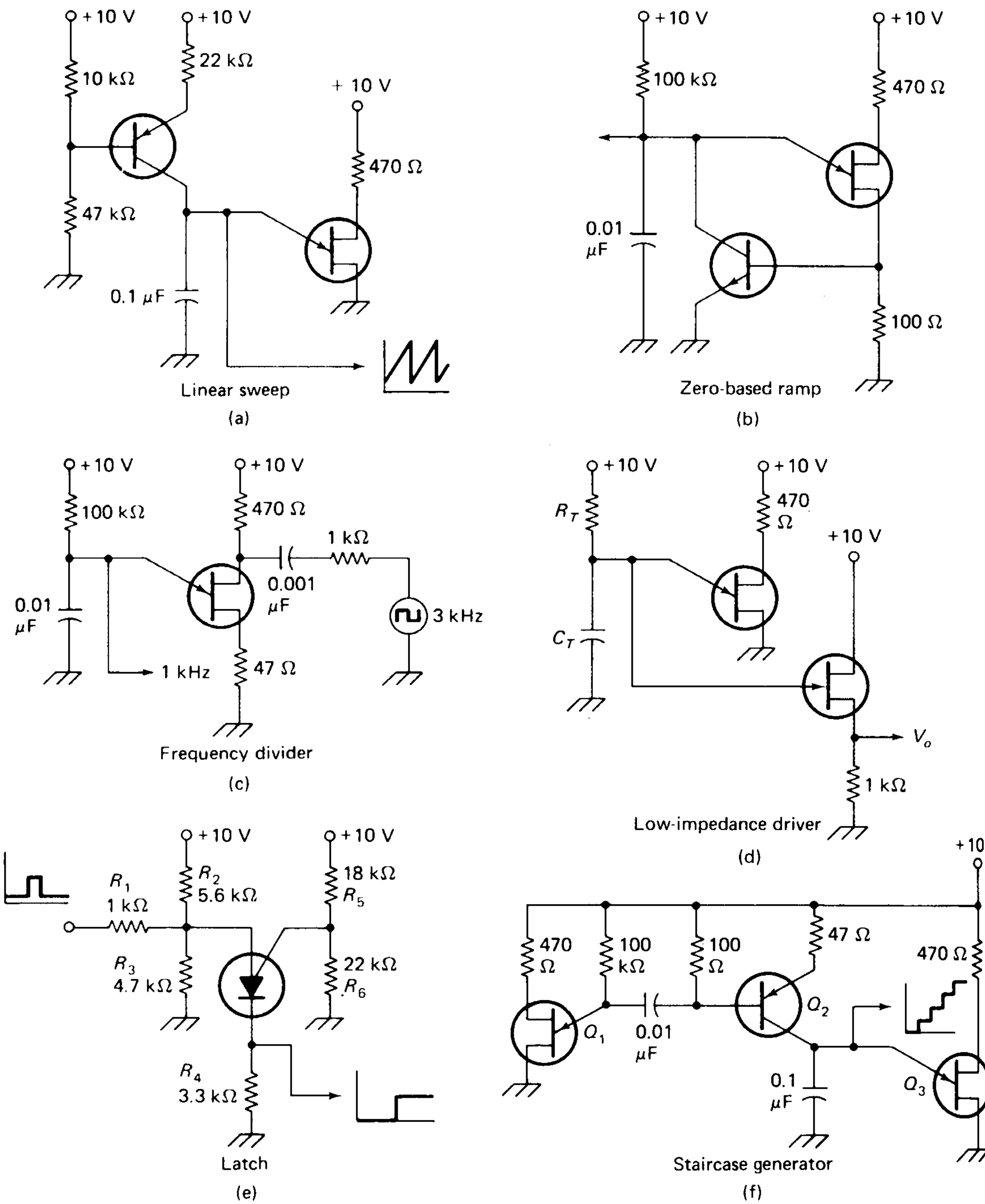


FIGURE 9-6 *Variety of UJT applications.*

often connected to the timing capacitor to reset the device when the capacitor voltage rises to $\frac{2}{3}V_{CC}$.

Reset and Control Functions are not used in elementary timer applications, but these input pins must be properly connected in any case. A *low* voltage (below 0.4 V) at reset directly turns *on* the discharge transistor, ending the cycle prematurely and bringing the output low. Normally, reset is connected directly to the V_{CC} pin *hi* level.

The voltage at the control input is normally held at $\frac{2}{3}V_{CC}$ by the three internal 5-kΩ voltage-dividing resistors. Varying this voltage by either dc or ac coupling alters the trigger-*on* and threshold-*off* voltages. A more positive control lengthens the *on* time, and increases the trigger level. Voltage control of oscillator frequency, pulse duration, and pulse position are possible via this input. Normally, *control* is bypassed to ground with 0.01 μF to prevent noise pickup.

Applications: Figure 9-8 shows the two basic *555* circuits upon which most others are based. The triggered one-shot of Fig. 9-8(a) is the basis of most time-delay,

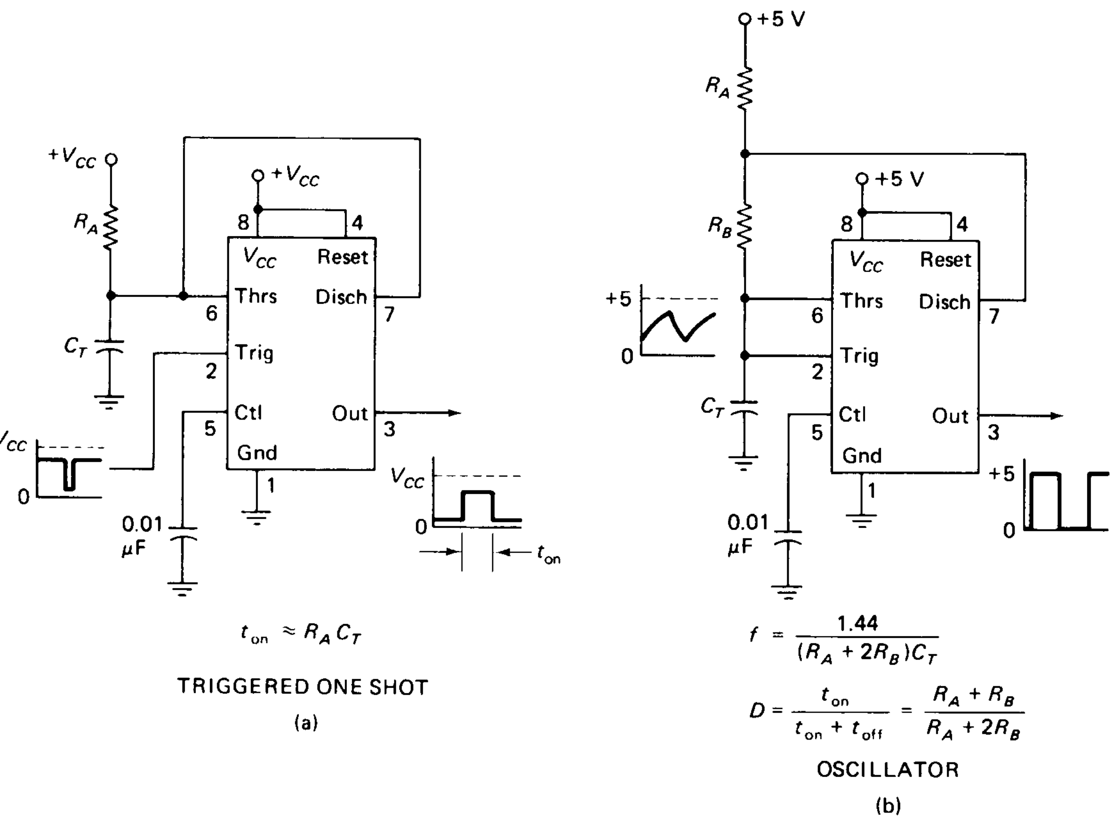


FIGURE 9-8 *The two basic 555 circuits upon which most others are developed: (a) one-shot; (b) oscillator.*

Outputs: The pin connections for the 555 timer and pin function names are shown in Fig. 9-7. The device operates on a positive supply between 4.5 and 16 V. It has a low-impedance (totem-pole) *pulse output* which switches between V_{CC} and ground, less 0.1 V or so for transistor saturation.

A second *discharge output* is connected to an open-collector transistor which pulls low when the output goes low. This is generally used to discharge the timing capacitor at the end of a cycle.

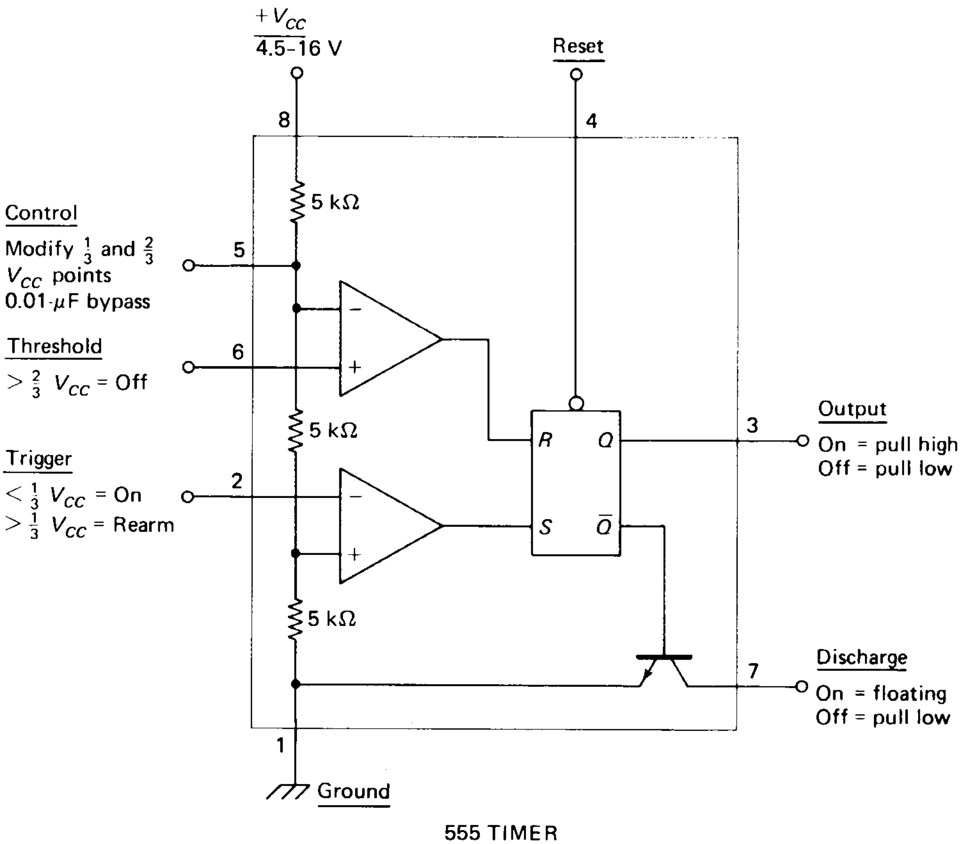


***FIGURE 9-7** Pin connections and representation of internal functions of the popular 555 timer IC. Inputs are on the left, outputs on the right.*

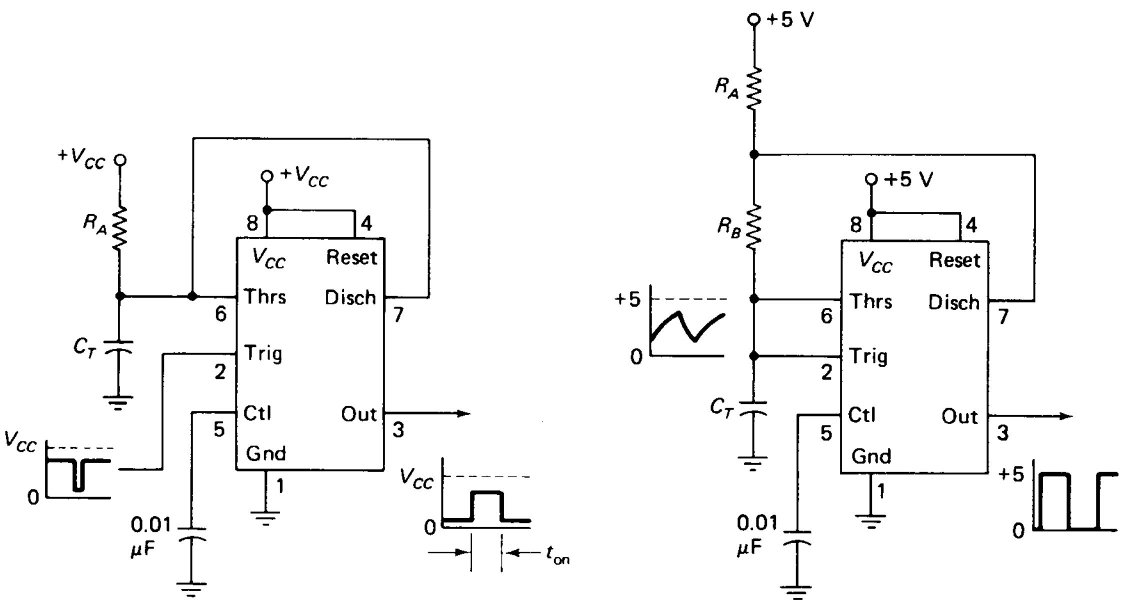
Inputs: There are two primary inputs, one to begin and one to end the cycle. The *trigger input* initiates a high output when its input is driven below $\frac{1}{3}V_{CC}$. The output then remains high, regardless of any further changes in trigger voltage. Notice that the trigger input voltage must be returned to a level above $\frac{1}{3}V_{CC}$ to re-arm the trigger function. Successive triggering cannot be accomplished by simply holding the trigger input low.

The second input (*threshold*) resets the output to *low* and turns *on* the discharge transistor when its voltage level goes above $\frac{2}{3}V_{CC}$. This input is most often connected to the timing capacitor to reset the device when the capacitor voltage rises to $\frac{2}{3}V_{CC}$.

Reset and Control Functions are not used in elementary timer applications, but these input pins must be properly connected in any case. A *low* voltage (below 0.4 V) at reset directly turns *on* the discharge transistor, ending the cycle prematurely and bringing the output low. Normally, reset is connected directly to the V_{CC} pin *hi* level.

The voltage at the control input is normally held at $\frac{2}{3}V_{CC}$ by the three internal 5-kΩ voltage-dividing resistors. Varying this voltage by either dc or ac coupling alters the trigger-*on* and threshold-*off* voltages. A more positive control lengthens the *on* time, and increases the trigger level. Voltage control of oscillator frequency, pulse duration, and pulse position are possible via this input. Normally, *control* is bypassed to ground with 0.01 μF to prevent noise pickup.

Applications: Figure 9-8 shows the two basic *555* circuits upon which most others are based. The triggered one-shot of Fig. 9-8(a) is the basis of most time-delay,



$$t_{on} \approx R_A C_T$$

TRIGGERED ONE SHOT

(a)

$$f = \frac{1.44}{(R_A + 2R_B)C_T}$$

$$D = \frac{t_{on}}{t_{on} + t_{off}} = \frac{R_A + R_B}{R_A + 2R_B}$$

OSCILLATOR

(b)

***FIGURE 9-8** The two basic 555 circuits upon which most others are developed: (a) one-shot; (b) oscillator.*

pulse-stretching, frequency-dividing, and sweep-generating applications. An input pulse less than $\frac{1}{3}V_{CC}$ at pin 2 initiates the output. Pin 2 may be held at $\frac{1}{2}V_{CC}$ with two 100-kΩ resistors between V_{CC} and ground, and the pulse coupled in via a capacitor to increase sensitivity. When R_A charges C_T to $\frac{2}{3}V_{CC}$, threshold pin 6 turns pin 7 *on* to ground, and the output goes to zero until the next trigger pulse. A positive voltage at the control input will lengthen the output pulse. An ac signal several times lower in frequency than the trigger rate will produce pulse-width modulation. If the time between trigger pulses is less than t_{on}, the circuit will lock to a submultiple of the trigger rate, serving as a frequency divider.

Figure 9-8(b) is a free-running oscillator. The circuit triggers itself because the trigger input is connected to the timing capacitor. The voltage across C_T oscillates between $\frac{1}{3}C$ (the trigger-*on* level) and $\frac{2}{3}C$ (the threshold-*off* level). The discharge of C_T can be made almost instantaneous by reducing R_B to zero. Frequency modulation can be achieved by coupling a low-frequency signal to the control input through a capacitor. A voltage-controlled oscillator is produced if a dc level is connected directly to pin 5.

Specifications: The following table summarizes the chief characteristics for the 555 IC timer:

Parameter	*Minimum*	*Maximum*	*Units*
Supply current			
$V_{CC} = +5$ V		6	mA
$V_{CC} = +15$ V		15	mA
Trigger input current		0.5	μA
Threshold input current		0.25	μA
Reset voltage	0.4	1.0	V
Reset current		0.1	mA
Low-level output			
$V_{CC} = +5$ V, $I_{sink} = -5$ mA		0.35	V
$V_{CC} = +15$ V, $I_{sink} = -10$ mA		0.25	V
High-level output			
$V_{CC} = +5$ V, $I_{source} = 100$ mA		2.75	V
$V_{CC} = +15$ V, $I_{source} = 100$ mA		12.75	V
Output rise and fall times		0.1	μs
Trigger pulse width: $V_{trig} = 0.3V_{CC}$	60		ns

9.3 SILICON-CONTROLLED RECTIFIERS

In the chapter on transistors we made quite a point of the fact that the power burned in any device is zero if the voltage across it is zero, regardless of the current it may carry. Likewise, the power is zero if the current is zero, regardless of applied voltage. A device that switches instantly from a perfect short circuit to a perfect open circuit can thus be used to control large amounts of power without burning any power itself. The mechanical SPST switch is the perfect example of this, but it is terribly slow and suffers from wear of the moving parts.

An SCR is like a very fast electronically actuated switch in series with a diode. This switch is turned on by applying a brief current to the gate lead, as shown in Fig. 9-9. Turn-off cannot be accomplished by gate control; rather it results from interrupting the forward current through the device.

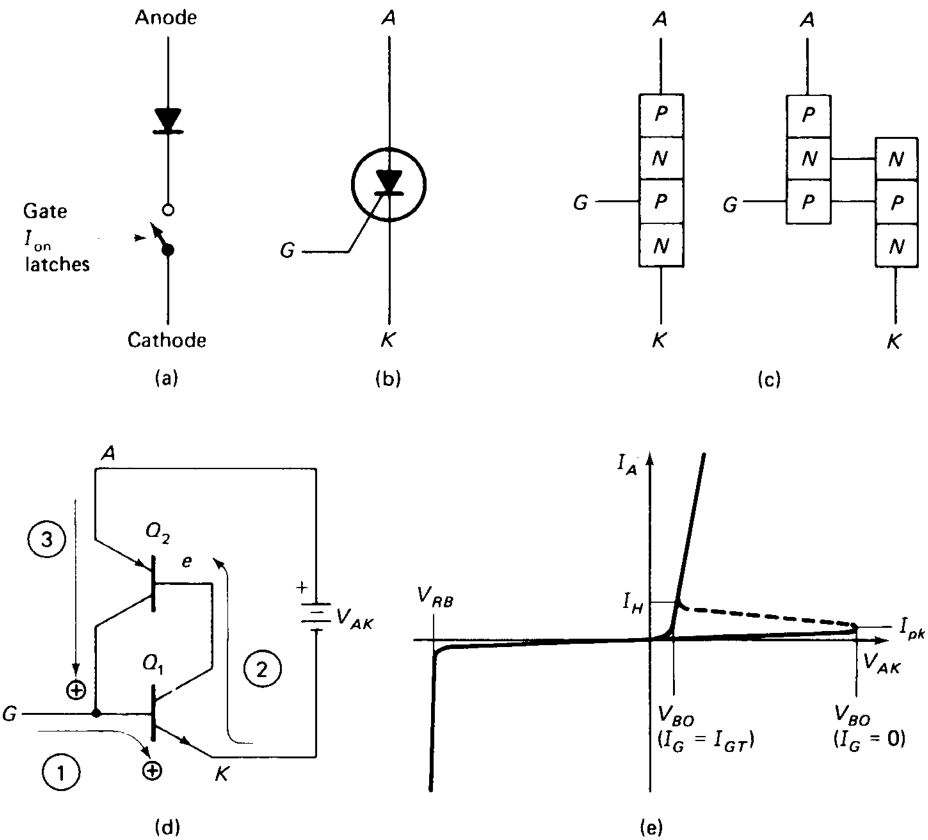


FIGURE 9-9 *(a) Representation of the SCR function. (b) SCR symbol and lead identification. (c) Internal SCR structure and similarity to coupled* PNP–NPN *transistors. (d) Two-transistor equivalent of the SCR showing self-latching current paths. (e) SCR* V *versus* I *characteristic curve.*

Two-Transistor Equivalent: Internally, an SCR consists of a four-layer sandwich of *PNPN*-doped silicon [Fig. 9-9(c)]. This is equivalent to two complementary transistors interconnected as in Fig. 9-9(d). Notice that once a gate turns Q_1 *on*, the Q_1 collector supplies turn-on current for the base of Q_2. Once turned on, Q_2 supplies turn-on current for the Q_1 base, holding it on regardless of any current from the gate. Reducing or even reversing the gate current will not turn an SCR off once it has fired.

SCR Applications most commonly make use of phase control of an ac line. This is illustrated in Fig. 9-10. If the SCR is fired early in the ac positive half-cycle, the average load power is one-half of that obtained by connecting the load directly to the line. However, if the SCR is gated *on* later in the half-cycle, the load power is reduced—to one-fourth at 90°, and eventually to zero as the "firing angle" approaches the end of the positive half-cycle at 180°.

Most of the fun of designing SCR circuits lies in finding more clever ways to trigger the gate at the desired point of the cycle. The objectives of the game are:

1. Trigger with a fast-rise pulse rather than with a ramp or sine wave. The gate firing current varies widely among units and with temperature, making pulse triggering the only stable way.
2. Make the pulses strong and brief. High gate currents make faster turn-ons, and hence less power is dissipated in the device. The average gate power limit must not be exceeded, however, so large gate currents must come and go quickly. Negative gate voltages waste gate power and are a definite no-no.
3. Cover the entire range of firing angles from 0° to 180°, but do not waste too much of the control range on the 0 to 30° delay angles, since there is little power difference between them.

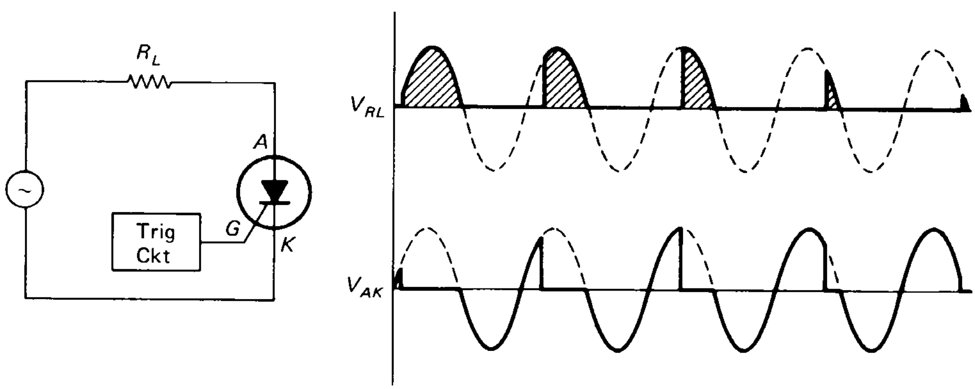
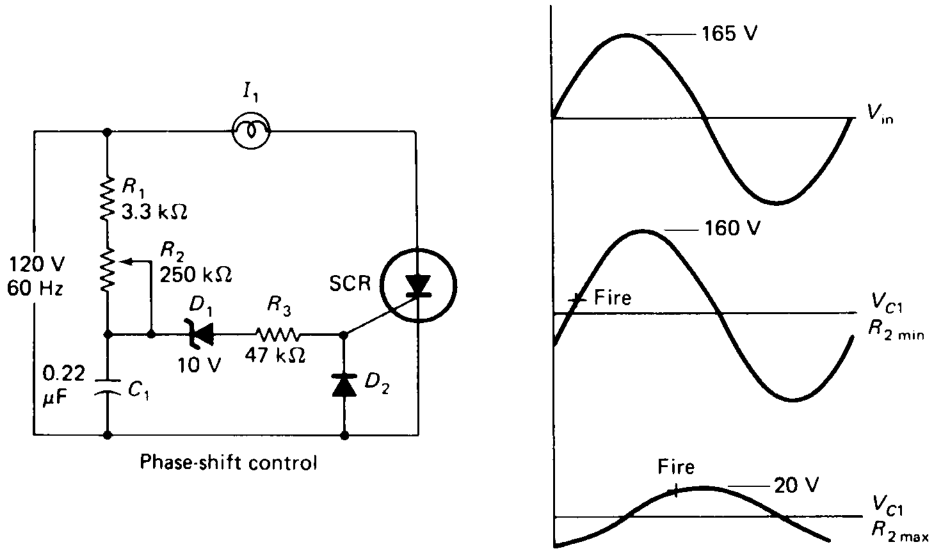
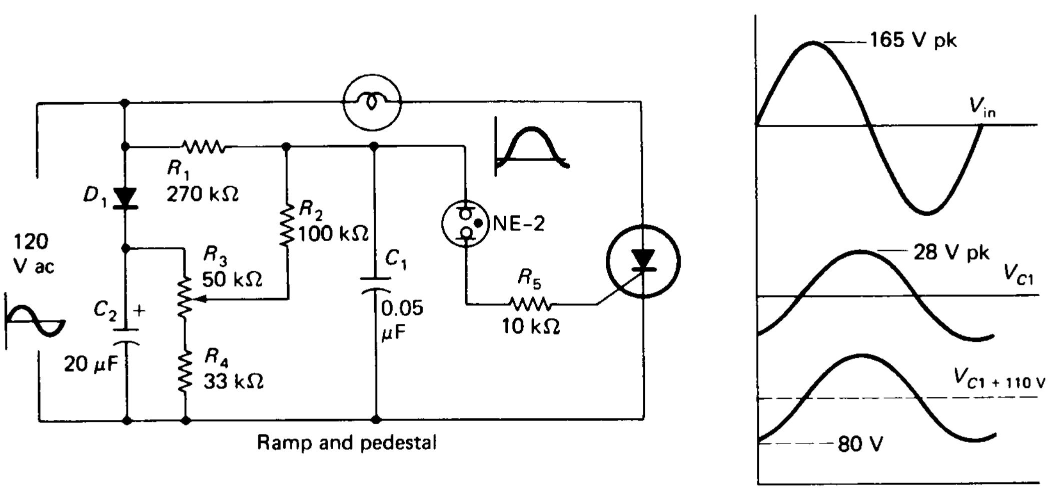


FIGURE 9-10 Decreasing average load power as SCR firing-delay angle progresses from 0° to 180°. Real-life progression generally takes hundreds of cycles.

Figure 9-11(a) shows a phase-shift firing circuit which violates most of the foregoing rules. Triggering is not done with pulses, but with a sine wave phase-shifted so that it rises sufficiently positive to trigger the gate at the desired point. R_1, R_2, and C are chosen to provide a capacitor phase lag adjustable from about 15° to 75°. With R_2 at its minimum, the voltage rises quickly and the SCR triggers before 20°, giving nearly 50% of full-wave ac power to the load. With R_2 at its maximum, the voltage is delayed by 75° and rises more slowly, so triggering may be delayed to

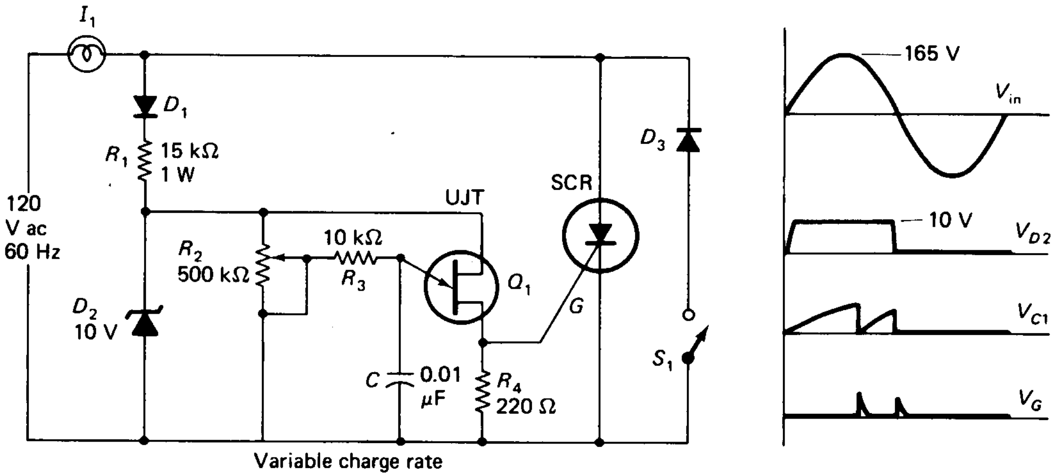


(a)



(b)

***FIGURE 9-11** (a) Simple but rather unreliable method of triggering an SCR by varying the phase and amplitude of the voltage across C_1. (b) More stable and reliable SCR trigger circuit with constant phase shift across C_1 but a variable dc voltage tapped from C_2 by R_3.*



(c)

***FIGURE 9-11** (c) Highly stable SCR trigger using a UJT and variable charge time on* C.

about 160°, giving less than 5% of full output. The circuit requires a sensitive-gate SCR (0.5 mA or less trigger current) and is quite sensitive to temperature and unit-to-unit variations. The diode on the gate prevents negative gate current and equalizes the load in the positive and negative directions.

Figure 9-11(b) shows an improved SCR lamp control using a neon lamp to generate a fast-rise gate-trigger pulse. R_1, R_2, and the 120-V line have a Thévenin equivalent of approximately 60 V through 75 kΩ, feeding a 70° lagging sine wave of 28 V pk to C_1. This wave is on the rise (-28 to $+28$ V) during most of the positive half-cycle of the line. The NE2 lamp fires at about 80 V, dropping immediately to 60 V, and the 20-V difference generates a 2-mA pulse through the 10-kΩ resistor. The 80-V level is achieved by adding a variable dc level to the 28-V-pk sine wave across C_1. D_1, C_2, R_3, and R_4 comprise a +64- to +160-V dc supply, which, through division with R_2 and R_1, supplies a maximum of +80 V dc to C_1. Varying the amount of dc imposed varies the point at which the sine wave will bring the total voltage to the critical 80-V level.

Figure 9-11(c) shows a different approach to triggering an SCR which allows the load to be placed directly in series with the entire SCR circuit. This is quite an advantage, since it means that an existing remotely located light switch can be replaced with the "dimmer" control. Assuming an 8-kΩ r_{BB} for Q_1, D_1 and R_1 bring the top of D_2 up to + 10 after the first 10° of the positive line cycle. Zener action holds this point at + 10 V for the remaining 170° of positive line.

R_2 and R_3 charge C to the UJT firing voltage (about 7 V) after a time that can be varied from 0.2 to 10 ms. This spans essentially the entire positive half-cycle (0 to 8.3 ms for a 60-Hz line). When the UJT fires, a short fast-rise pulse (duration $\approx R_4C$ or about 2 μs) triggers the SCR gate. At the end of the positive half-cycle the UJT B_2 voltage falls to zero, causing the emitter to fire and discharge the capacitor in preparation for the next cycle.

This circuit is not sensitive to variations in gate trigger current, and trigger pulses are brief, keeping average gate power low. The *RC* network is isolated from the SCR gate by the UJT, so that there is no hysteresis effect. A power diode D_3 is shown switchable across the SCR to allow the negative half-cycle through the load, providing variable control from 50% to 100% of full power. This same technique can be applied to the previous two circuits.

Figure 9-12 gives several additional SCR applications. The first is a full-wave version of the UJT-triggered lamp dimmer. A diode bridge converts the ac line to all-positive pulses before application to the SCR and control circuit. This technique is not generally applicable to phase-shift triggering circuits such as those in Fig. 9-11(a) and (b). The load current is ac with the circuit wired as shown, but dc load current can be achieved by placing the load in series with the SCR anode.

Figure 9-12(b) is an automobile battery charger with automatic shutoff. When the battery voltage charges up to about 13.5 V, the gate-to-cathode voltage becomes too low (0.5 V) to trigger the SCR.

Figure 9-12(c) is a speed control for universal (series field and armature) motors, such as are common in hand drills, sewing machines, and food mixers (but not in household fans, record players, and bench tools). The circuit balances the induced voltage produced by the motor against the wiper voltage from R_2. As the motor slows down due to increased load, induced voltage is reduced, and the SCR fires earlier in the positive half-cycle. Raising the wiper voltage raises the required induced voltage and hence motor speed.

Figure 9-12(d) provides a 30-second time delay before turning on the SCR, which is used as a rectifier in a high-voltage supply. Such a circuit is useful in systems containing a CRT or other heated-filament device, to allow warmup time before applying high voltage. The UJT fires the low-power SCR_1 after a delay time set by R_1C_1. SCR_1 then remains *on* as long as the +5-V supply remains on. This turns *on* the LED, which is optically coupled to the photoresistive cell, lowering its resistance and allowing the high-power SCR_2 to fire on each positive half-cycle. Note that the photocell must be able to withstand the peak secondary voltage until C_2 becomes charged.

Figure 9-12(e) is a low-voltage dc supply operating directly from the ac line without transformers or power-wasting series-dropping resistors. Since one side of the line is connected directly to the negative output of this circuit, adequate precautions should be taken to ensure that contact between the circuit and earth ground cannot occur. On the positive half-cycle, D_1, R_1, and D_2 bring the B_2 reference of the UJT

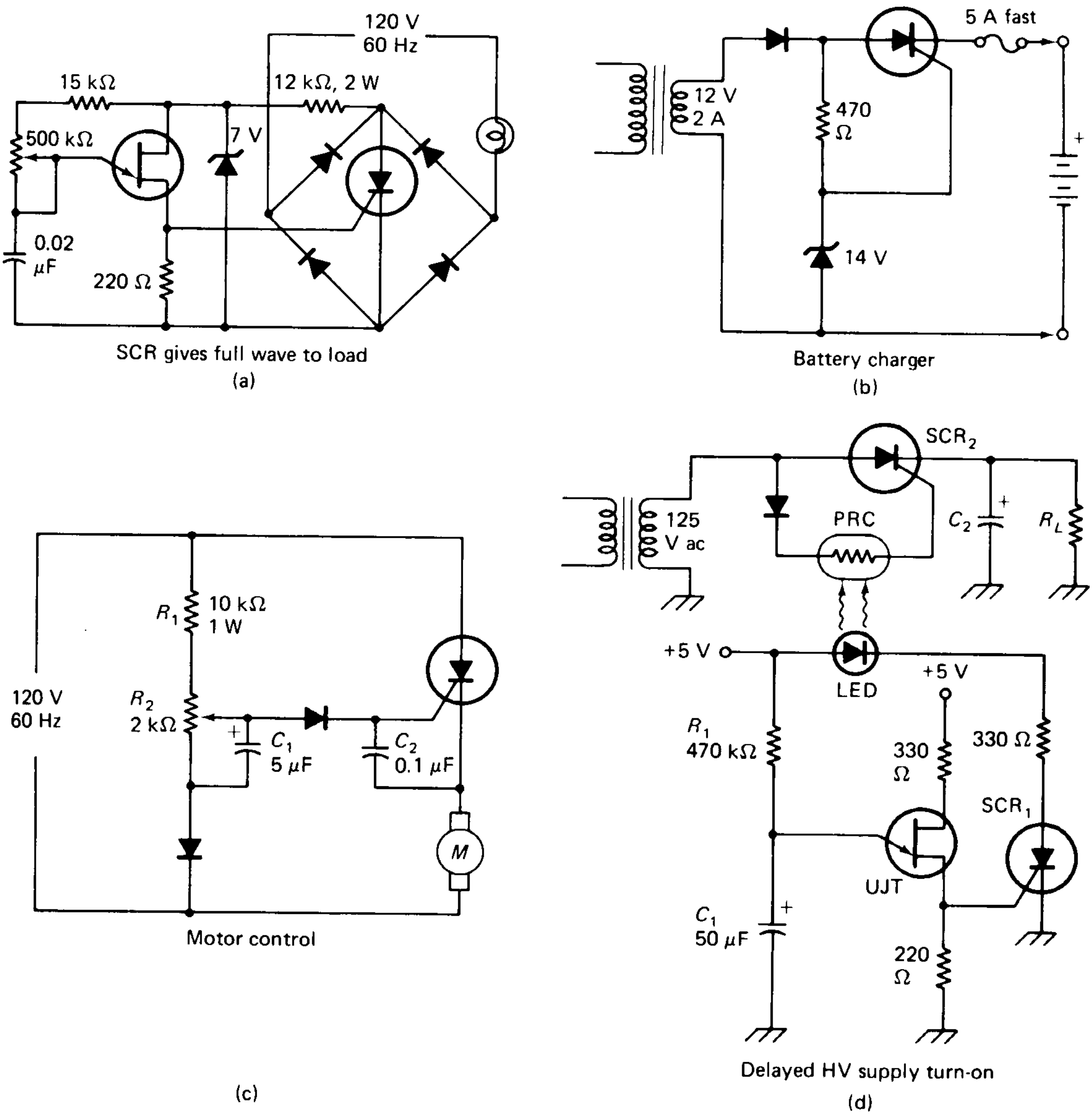
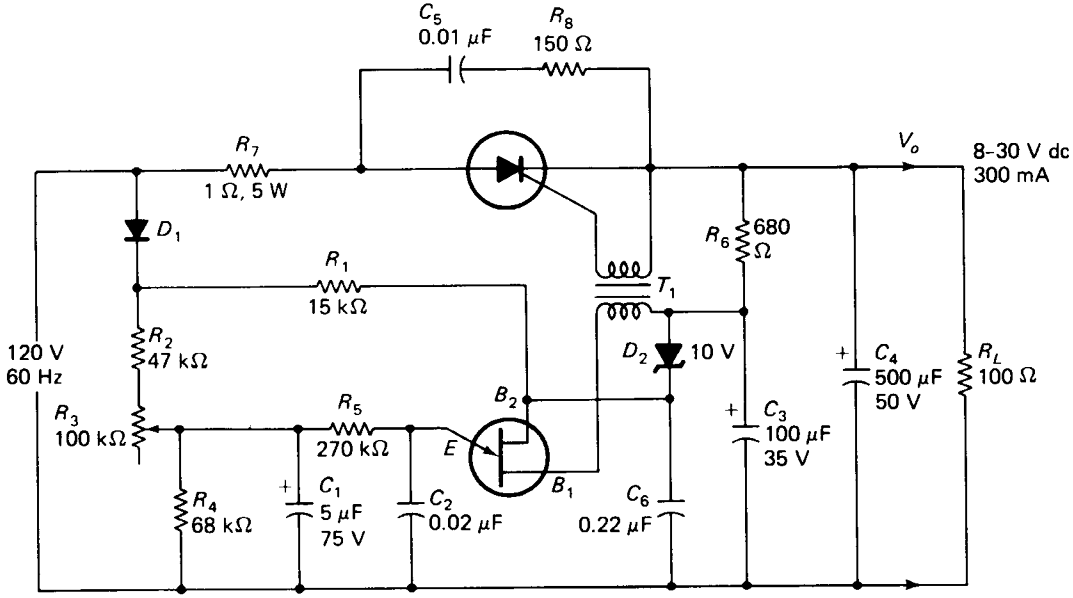


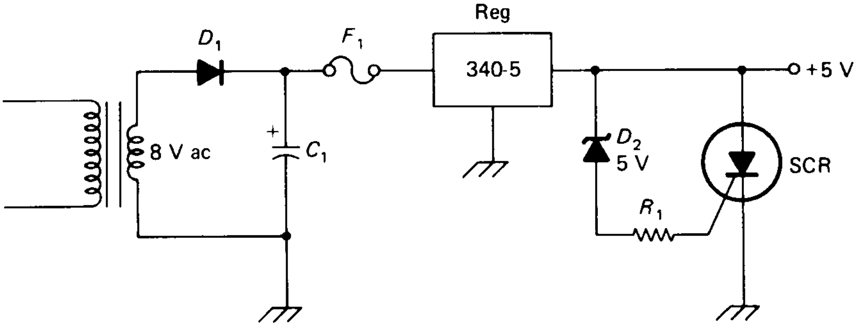
FIGURE 9-12 *Six SCR applications.*

rapidly to + 10 V above V_O, which is initially zero. R_2, R_3, and R_4 form a variable voltage divider charging C_1 to a positive value between 20 and 50 V. The time constant for charging C_1 is about 0.25 s. This slow rise is important because it raises the output voltage demand slowly, preventing destructive voltage and current surges at turn-on. R_5 charges C_2 toward V_{C1}, reaching the firing point of the UJT just as the positive half-cycle of the line is nearing its end. The discharge of C_2 through pulse transformer T_1 fires the SCR, charging C_4 after several cycles, to (let



Transformerless low-voltage supply

(e)



Crowbar overvoltage protection

(f)

FIGURE 9-12 *(Continued).*

us say) + 20 V. R_6 and C_3 filter the ripple from V_O and D_2 now clamps the B_2 voltage at V_O + 10, or 30 V. V_{BB} is still 10 V, since B_1 is referenced to V_O.

If V_O rises (due perhaps to loss of load) R_5 takes longer to fire the UJT because of the lesser slope of the charge curve at higher voltages and the fact that the R_5 source voltage is referenced to ground, not V_O. This later firing means that the line voltage drops nearer to zero before it charges C_4, tending to bring V_O back down. Load regulation of 10% for a 200-mA load at V_O = 20 V is achieved by this feedback. Line regulation could be added by zener regulating the voltage from C_1 before applying it to R_5. C_6 maintains control voltage at B_2 for a millisecond or so after the line drops below $V_O + 10$ V, allowing the UJT to be emitter-triggered even at delay angles approaching 180°. R_7 limits surge current at turn-on. Conduction angles are on the order of 10 to 15°, and current pulses are about $20I_O$ for this circuit. C_5 and R_8 suppress line transients which could cause false triggering of the SCR. A trigger at the early part of the positive half-cycle would apply a 170-V peak to C_4 and R_L, which could be quite bad.

Figure 9-12(f) is a "crowbar" circuit, which has become quite popular in spite of its crude approach to circuit protection. If V_O rises above +5.6 V (due perhaps to failure of the 5-V *IC* regulator), the entire system it supplies is threatened with destruction. The SCR saves the system by shorting V_O dead to ground, blowing the fuse F_1.

SCR Ratings: There are well over a dozen commonly used specifications describing SCR performance. The following table of subscript usage should be of some help in sorting out the parameter symbols:

Subscript	*Meaning*
(AV)	Average
(BO)	Breakover
D	Off state
F	Forward
G	Gate
H	Hold-on
M	Maximum peak
O	Gate open
R	Reverse (first subscript) Repetitive (second subscript)
S	Surge (nonrepetitive)
T	On state (First subscript) Trigger (second subscript)

Now we are prepared to examine the various ratings in detail.

Anode–Cathode Voltage

- V_{RSOM} and V_{RROM} are the maximum reverse peak surge voltage and peak repetitive voltage which the device can block without drawing reverse current. The gate must not have a positive voltage applied while a large negative voltage exists on the anode, as this causes reverse anode current of a magnitude on the order of I_G. Thus a 20-mA gate current during the negative half-cycle of a 120-V line could cause an average additional power dissipation of over one watt:

$$P = I_{(\mathrm{AV})}V_{(\mathrm{AV})} \approx 0.5 \times 0.02 \times 0.9 \times 120 = \mathbf{1.1\ W}$$

Note that the current flows for 0.5 cycle and that $V_{\mathrm{AV}} = 0.9\ V_{\mathrm{rms}}$. A diode in series with the anode will stop reverse leakage if the gate must be positive during the negative half-cycle.

- V_{DSOM} and V_{DROM} are the maximum forward blocking voltages, nonrepetitive surge, and repetitive or continuous, respectively.
- $V_{F(\mathrm{BO})\mathrm{O}}$ is the forward breakover or firing voltage with the gate open. This firing voltage decreases for small forward gate currents below the gate trigger current, but little use is made of this fact because of extreme sensitivity to temperature and unit-to-unit variations.
- V_{TM} is the forward on-state voltage, anode to cathode, for a specified forward current and junction temperature. It may be as low as 0.8 V for low currents, rising to around 2.0 V at $I_{T(\mathrm{AV})}$ and 3.0 V at I_{TSM}.

Anode Current

- I_{ROM} and I_{DOM} are the maximum reverse and forward off-state anode leakage currents at a specified voltage and temperature.
- I_{TSM} is the maximum rated forward on-state surge current for a specified short time, usually one cycle. I_{TSM} is typically at least 10 times $I_{T(\mathrm{AV})}$.
- I^2t is a nonrepetitive surge-current rating intended for use in calculating fuse sizes for the protection of the SCR. If the product of fuse-blow current squared times blow time is less than I^2t, the fuse will blow before the SCR.
- $I_{T(\mathrm{AV})}$ and $I_{T(\mathrm{rms})}$ are the average and rms maximum on-state currents allowable. Values from 0.5 A to over 2000 A are available. These ratings are generally dependent upon conduction angle and line frequency, and are detailed in charts supplied by the manufacturer. They are given in lieu of total device power-dissipation limits, although power dissipation can sometimes be obtained from additional charts. All of this is quite a bit of trouble, so it may be assumed in most cases that the maximum average load current at the 180° conduction angle equals the maximum average anode current. Average power is then approximately $I_{\mathrm{AV}}V_{\mathrm{TM}}$. For circuits such as those shown in Fig. 9-12(e), where conduction angles are never longer than a small fraction of a cycle, the average-current specification will have to be derated because of junction heating on high-current peaks. The exact derating depends upon the thermal time constant of the SCR in use, but the following schedule is typical for 1- to 50-ampere SCRs.

Conduction Angle, α	$\dfrac{I_{(AV)}\alpha \text{ rating}}{I_{(AV)}180° \text{ rating}}$ (%)
dc	120
180°	100
90°	80
60°	70
30°	40

- I_{HO} is the anode holding current, below which the SCR reverts to the *off* state. I_L, which is typically about 50% greater than I_{HO}, is the forward latching current which is necessary for the SCR to initially assume the *on* state.

Gate Characteristics

- V_{GT} is the dc gate voltage required to trigger the device. It is generally between 0.7 and 1.5 V positive with respect to the cathode.
- V_{GDM}, sometimes called V_{GNT}, is the maximum gate voltage guaranteed *not* to trigger the device, and is generally specified at a low value, around 0.25 V. Some spec sheets list $V_{GT(\min)}$ for this value.
- V_{GRM} is the maximum reverse voltage guaranteed not to cause reverse current in the gate. It is typically in the vicinity of 5 V.
- I_{GT} is the dc gate current required to trigger the device, and I_{GTM} is the peak triggering current required for a specified pulse width.
- $P_{G(AV)}$ is the maximum allowable average gate power ($I_G V_G \times$ duty cycle). P_{GM} is the maximum peak gate power for a specified short pulse time.

Time Considerations

- t_{gt} is the gate-controlled turn-on time for a specified gate current and anode current. Values of a microsecond or two are typical. Time is measured from rise of the gate pulse to fall of the anode voltage. t_{gt} is sometimes split into delay time t_d (gate pulse to 10% anode current) and rise time t_r (10% to 90% anode current). Turn-on time can be reduced significantly by increasing gate trigger current.
- t_w is the gate pulse width required to trigger the SCR for a specified gate pulse current. The relationship between t_w and I_{GTM} is generally shown in a manufacturer's chart.

- t_q is circuit-commutated turn-off time for a specified I_T (on current), V_R (off voltage), and V_D (forward blocking voltage). The SCR is triggered carrying anode current I_T. Then a reverse (or zero) anode voltage is suddenly applied. Finally, a forward voltage V_D is applied. The test is repeated to determine the minimum time the forward reapplied voltage can be delayed to allow the SCR to turn off and gain forward-blocking capability. This time t_q has a typical range of values between 5 and 100 μs, depending upon SCR type.

- Critical di/dt is the maximum allowable rate of rise in forward current at turn-on. Because the junction area near the gate is turned *on* before the rest of the junction area, localized hot spots can appear if the initial current at turn-on is allowed to rise too quickly.

- Critical dv/dt is the maximum rate of change of forward anode voltage that the SCR can withstand without false triggering due to gate currents coupled in by the device's own stray capacitance. Transients on the ac line are often responsible for such false triggering. The accepted solution for this problem is to use an RC "snubber" across the SCR, as shown in Fig. 9-13. The values of R and C for a 120-V line can be obtained from the formulas

$$R = 3\frac{dv/dt}{I_L} \tag{9-5}$$

$$C = 0.1\frac{I_L}{(dv/dt)^2} \tag{9-6}$$

where R is in kΩ, C is in μF, I_L is in amperes, and dv/dt is in V/μs.

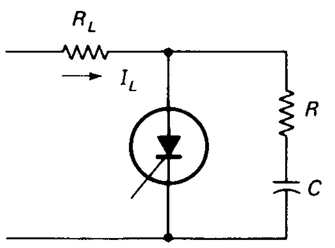


FIGURE 9-13 *RC snubber to prevent false triggering of the SCR on fast-rise line transients. Component values are given by equations 9-5 and 9-6.*

SCR Power Calculations: In dealing with the chopped-up sine waves produced by SCRs, we often need to know the average and rms values of current involved in order to calculate SCR power and load power, respectively. The graph of Fig. 9-14 gives these values for conduction angles from 0 to 180°, assuming a 1-V-pk source applied to a 1-Ω load, giving a 1-A-pk current at full conduction. Other peak voltage and current sources can be scaled accordingly. For do-it-yourselfers, the formulas used to obtain the chart are given:

$$V_{(\mathrm{av})\mathrm{half}\,\lambda} = V_{\mathrm{pk}} \frac{\cos\theta + 1}{2\pi} \tag{9-7}$$

$$V_{(\mathrm{rms})\mathrm{half}\,\lambda} = \frac{V_{\mathrm{pk}}}{2} \sqrt{1 - \frac{\theta}{180} + \frac{1}{2\pi}\sin 2\theta} \tag{9-8}$$

$$V_{(\mathrm{av})\mathrm{full}\,\lambda} = V_{\mathrm{pk}} \frac{\cos\theta + 1}{\pi} \tag{9-9}$$

$$V_{(\mathrm{rms})\mathrm{full}\,\lambda} = \frac{V_{\mathrm{pk}}}{\sqrt{2}} \sqrt{1 - \frac{1}{180} + \frac{1}{2\pi}\sin 2\theta} \tag{9-10}$$

where θ is the firing-delay angle in degrees from the zero-crossing point (0° in half wave, 0° and 180° in full wave) and V_{pk} is the peak value of the applied line voltage.

EXAMPLE 9-1

An SCR is to supply 300 W to a 10-Ω load on a 120-V-rms line by controlled half-wave rectification. Determine the firing-delay angle and the SCR power. The SCR drops 2.0 V in conduction.

Solution

The required rms current as a fraction of peak current is found:

$$I_{\mathrm{pk}} = \frac{V_{\mathrm{pk}}}{R_{\mathrm{L}}} = \frac{1.41 \times 120}{10} = 17.0\ \mathrm{A}$$

$$I_{\mathrm{rms}} = \sqrt{\frac{P}{R}} = \sqrt{\frac{300}{10}} = 5.48\ \mathrm{A}$$

$$\frac{I_{\mathrm{rms}}}{I_{\mathrm{pk}}} = \frac{5.48}{17.0} = 0.322$$

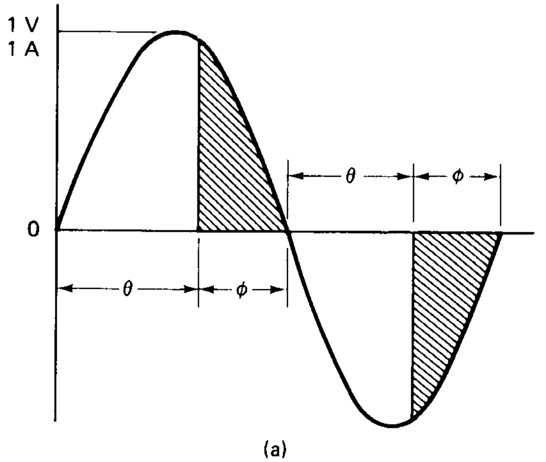


FIGURE 9-14 *(a) Firing-delay angle θ and conduction angle φ.*

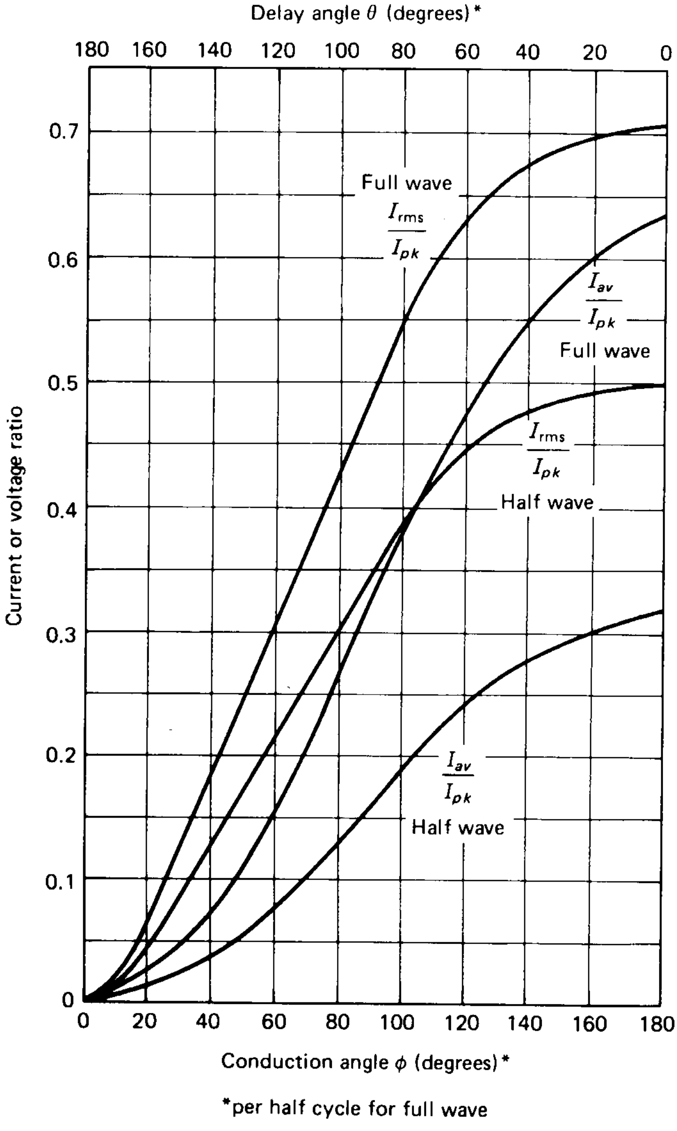


***FIGURE 9-14** (b) Ratios of rms/peak and average/peak voltage or current for SCR waveforms as a function of conduction angle.*

This value appears in Fig. 9-14 at a delay angle of 97°, and the corresponding average current ratio is 0.14.

$$I_{av} = \left(\frac{I_{av}}{I_{pk}}\right) I_{pk} = 0.14 \times 17.0 = 2.4 \text{ A}$$

$$P_{SCR} \approx I_{av}V = 2.4 \times 2.0 = \mathbf{4.8\ W}$$

9.4 TRIACS AND OTHER THYRISTORS

The SCR was the first of the family of semiconductor devices collectively known as thyristors. None of the later additions has near the power-handling capability or the popularity of the SCR, with the exception of the triac, which is available with ratings up to 40 A and is widely used in full-wave lighting and motor controls. Figure 9-15 summarizes the lesser-used thyristors.

Amplified-Gate SCRs: The ratio of maximum average anode current to gate triggering current is several hundred to a thousand for most SCRs, but is increased to several thousand in the amplified-gate SCR. Major operational characteristics remain unchanged.

Gate-Controlled Switch (GCS) or turn-off SCR: The anode current of this device can be turned off by a negative gate pulse while forward anode voltage is still applied. However, the negative gate current required is on the order of one-tenth the anode current, and only low-power devices are available.

Complementary SCR: The CSCR has the gate near the anode and fires with a negative pulse applied between gate and anode. It is a true complement to the SCR, operating on negative voltages as the SCR does on positive voltages.

Silicon-Controlled Switch: The SCS combines the features of the GCS and CSCR. Both anode and cathode gate leads are accessible. It can be turned on by a positive voltage from G_K to K or by a negative voltage from G_A to A. The cathode gate is extremely sensitive ($I_{GTK} \sim 1\ \mu A$), the anode gate less so ($I_{GTA} \sim 100\ \mu A$ for a 200-mA device). The SCS can also be turned off by a negative current of approximately $0.1 I_A$ from the cathode gate. False triggering due to high dv/dt at the anode can be completely eliminated without use of an RC snubber by the simple expedient of connecting G_A to the anode supply through a 100-kΩ resistor and using the SCS as a conventional low power SCR. One notable application of the SCS is the latching high-voltage driver of Fig. 9-15(b). A pulse of a few microamperes at G_K turns the neon lamp on, and it latches until I_A is removed. Alternatively, I_A can be pulse-width-modulated at 100 Hz or so to vary the brightness of the lamp. I_{GK} must be continuous in this case.

Silicon Unilateral Switch: The SUS is a complementary SCR with an integrated zener diode connected from cathode to gate so that the device triggers without a gate signal when V_{AK} exceeds the zener voltage. Commonly available SUSs are set to trigger at 8 V. The gate lead is brought out so that triggering below 8 V is possible with a negative voltage from gate to anode, although this is infrequently done. The primary application for the SUS is as a gate trigger in SCR circuits, in the way the neon lamp was used in Fig. 9-11(b). The advantages of the SUS trigger level of 8 V over the NE-2 level of 70 V are obvious.

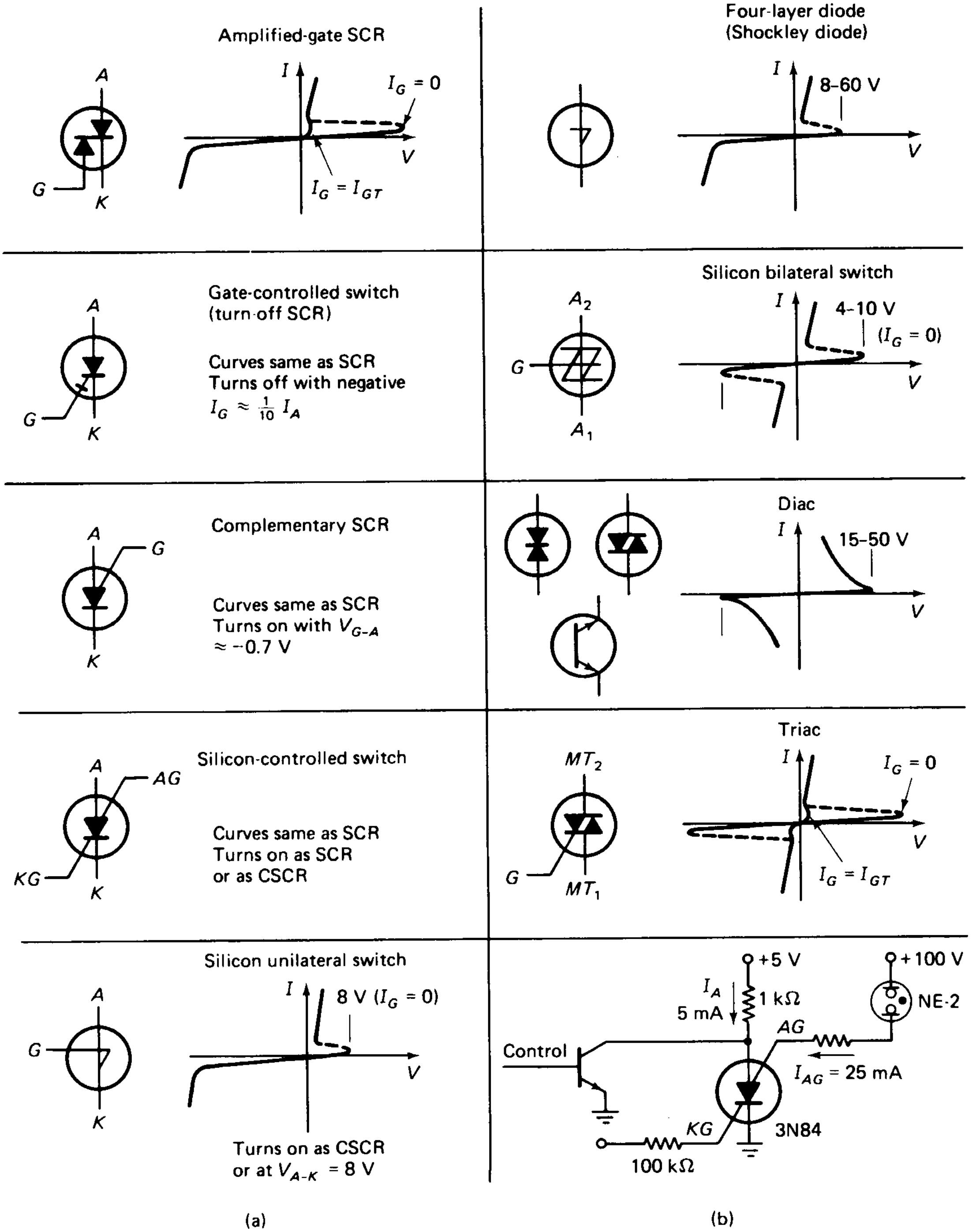


FIGURE 9-15 *Thyristor family portrait.*

Shockley or Four-Layer Diode: This device is similar to the SUS and finds similar applications. It has no gate control and is available in a range of breakdown voltages from 8 to 60 V. It is faster than the SUS, switching in about 0.1 μs.

Silicon Bilateral Switch: The SBS is a parallel connection of two oppositely facing SUS devices. It triggers at the specified voltage in either direction if the gate is left open, but can be triggered earlier by a negative voltage from the gate to the currently positive anode. Self-triggering voltages from 4 to 10 V are common.

Diac: This is a three-layer silicon trigger device similar to SBS but without a gate lead. Commonly available trigger voltages range from ± 15 to ± 50 V. The device voltage does not necessarily drop to a low value upon triggering, but "switches back" at a specified rate $\Delta V/\Delta I$ as shown in Fig. 9-15. There seems to be quite a fight over what symbol to give the diac, so all three contenders are shown in the figure. The primary use for the diac is as a trigger for a triac.

Triac: As the name implies, the triac is a three-lead device designed to handle ac. It might also be called a bidirectional SCR, because that is exactly how it functions. The triac can be triggered by either a positive or negative voltage from gate to main terminal 1 for either a positive or negative voltage on main terminal 2. The four triggering modes are referred to as the four quadrants of the triac, from a presumed graph of I_{MT2} versus I_{GMT1}, as shown in Fig. 9-16. Such a graph is seldom given in practice, but the concept helps keep things straight. Some manufacturers refer to quadrants 2 and 4 as I− mode and III+ mode, respectively, so these labels are given, too. Triggering current I_{GT} is quite different for the various quadrants, being generally lowest in quadrant I and highest in IV. Where a single polarity of gate trigger pulse is used to trigger on both polarities of the line, negative gate pulses are generally used to avoid the high current demand of quadrant IV and the inequality between quadrants I and IV. Some triacs are not guaranteed to trigger at all in quadrants II and IV. Gate voltages and currents, on-state voltage, and response times for triacs are similar to corresponding values for comparable-size SCRs.

Triac Applications: Figure 9-17(a) shows an elementary triac light dimmer. With R_2 at maximum, V_C lags the line by about 85°, but never becomes large enough to trigger D, and the triac remains off. As R_2 decreases, V_C increases, until at its peak (84° + 90°) it is high enough to trigger D and fire the triac. The line voltage now appears across the lamp, the RC voltage being clamped to zero by the triac. The capacitor voltage thus drops as the diac discharges it, and holds until the negative half-cycle of the line, when the triac turns off and voltage again appears across RC.

The discharge of C by the triggering of D has brought V_C nearer to its next (negative) triggering point, however, and triggering occurs quite a bit earlier than it did on the positive half-cycle, as shown in Fig. 9-17(b). The negative trigger likewise brings V_C nearer to the positive trigger level, resulting in conduction angles

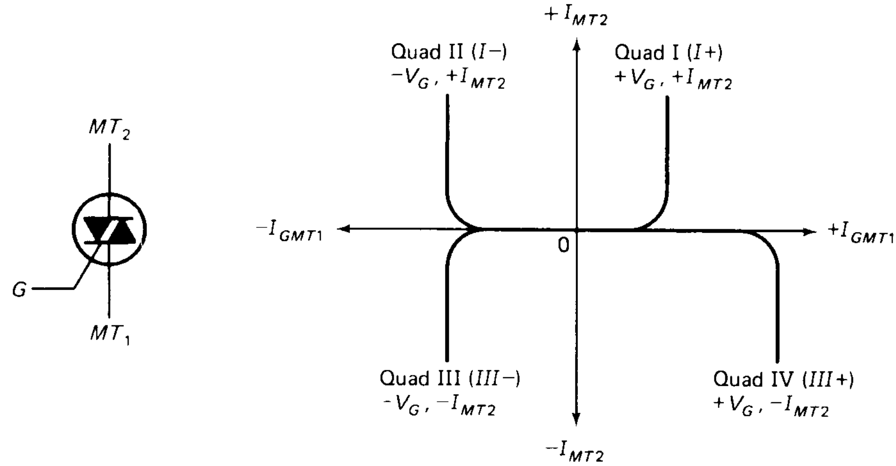


FIGURE 9-16 *Triac symbol and the two competing systems for naming the four possible firing modes. Note representative magnitudes of firing current I_G in each mode.*

between 30° and 40° per half-cycle. Rotating R_2 thus produces nothing at first, until the lamp suddenly pops on at about one-tenth brightness. Backing off on R_2 then produces lower brightness. This effect is called hysteresis.

Figure 9-17(c) shows a simple way of reducing hysteresis to an insignificant level. R_3 and C_2 provide isolation so that C_1 is not seriously discharged when D_1 triggers. R_3 and C_2 also provide an additional 20° lag in triggering voltage so that V_{C2} is on a steeper rise at the point of trigger.

Switching a triac or SCR *on* in the middle of the ac half-cycle causes very abrupt current changes which produce interference in nearby AM radios. The circuit of Fig. 9-17(c) has a hash filter L_1C_3 to suppress this.

Figure 9-18(a) shows a simple scheme for eliminating sparking and pitting when switch contacts are used to control an inductive load. The 1-kΩ resistor turns the triac on near the begining of each half-cycle (or in the middle of a half-cycle if the switch is closed then), but when the switch is opened the triac waits until the current alternates through zero before turning off.

Figure 9-18(b) shows a scheme for turning the power on and off only at the zero-voltage crossings of the ac line. The initial current inrush to a cold lamp filament is 5 or 10 times its hot-filament value, so the thermal shock if the switch happens to close when the line is near its peak voltage is quite severe. Synchronous switching at the zero-voltage point thus extends lamp life, and if the lamp is part of a flashing sign on an 80-ft pole in midwinter, the cost of the circuitry will seem amply justified. R_1, R_2, and D_1 feed turn-on current to the gate of the triac on positive half-cycles. The 3-mA I_{GT} is reached at about 25 line volts. However, if the switch is closed, R_1 D_2, and R_3 provide the 2-mA I_{GT} for the SCR at about 12 line volts, firing the SCR and diverting the triac gate current. D_3 and R_4 charge C on

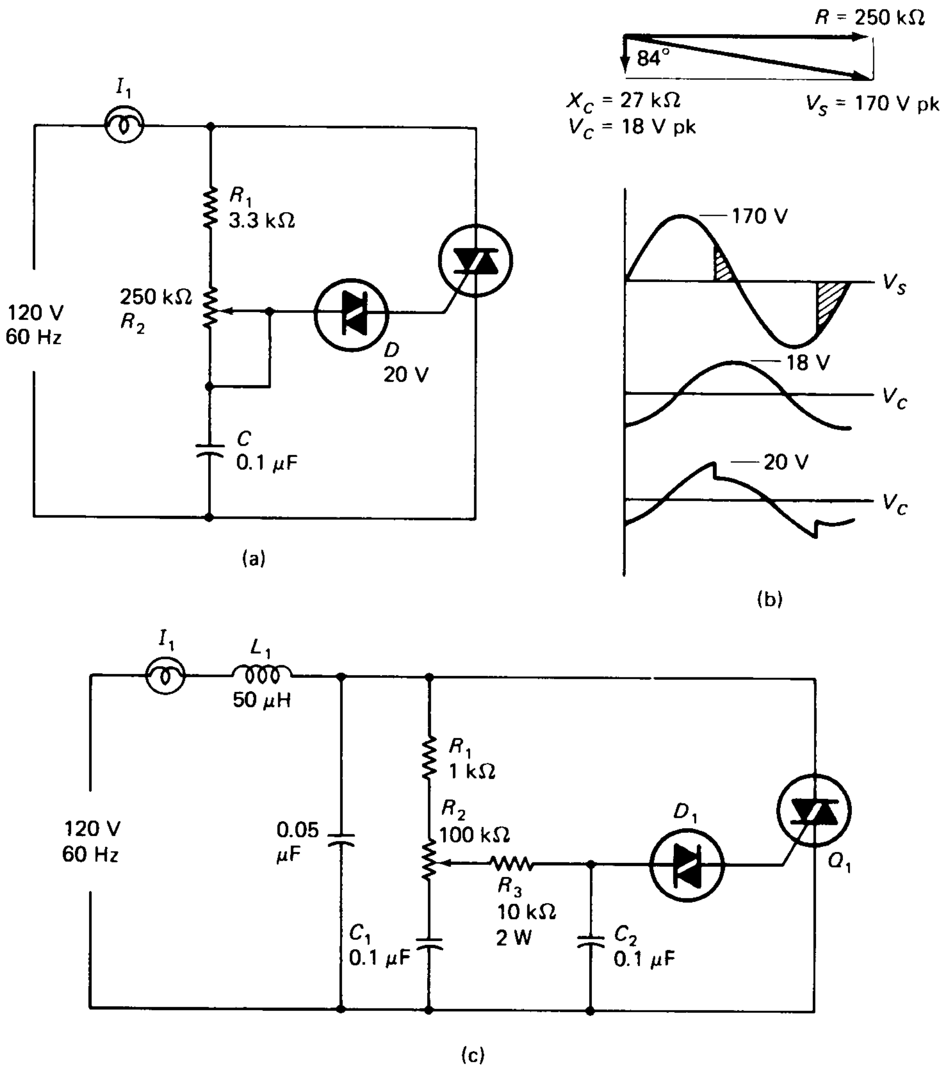
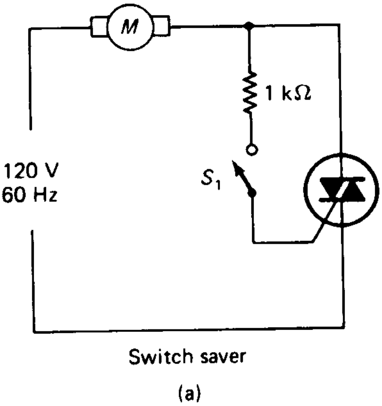
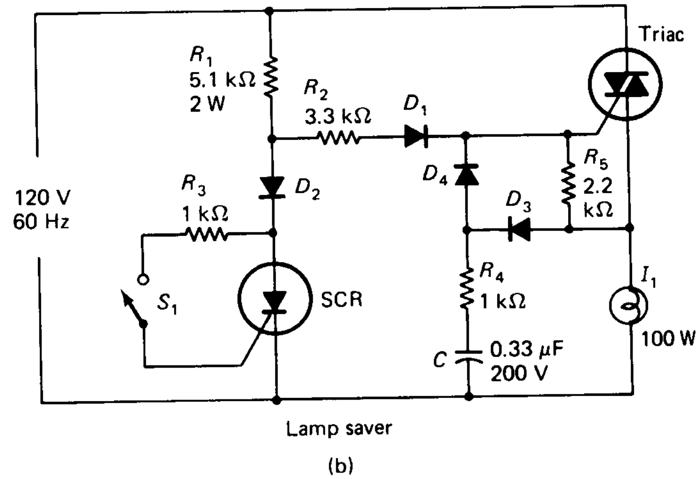


FIGURE 9-17 (a) Simple triac trigger using a diac and single RC phase-shift network. (b) Phase-shift diagram and waveforms at threshold of firing, indicating cause of "pop-on" hysteresis. (c) Two-RC network with diac trigger minimizes hysteresis.

positive half-cycles if the triac fires, and D_4 discharges C through the triac gate on the subsequent negative half-cycle, providing quadrant IV triggering. R_5 speeds the turn off of the triac at the zero-voltage crossings.

Figure 9-18(c) shows a triac solution to the problem of regulating high-voltage power supplies. Consider a 3000-V 1-A supply for a medium-power radio transmitter. Assuming a ±15% line-voltage variation, we would require a 900-V





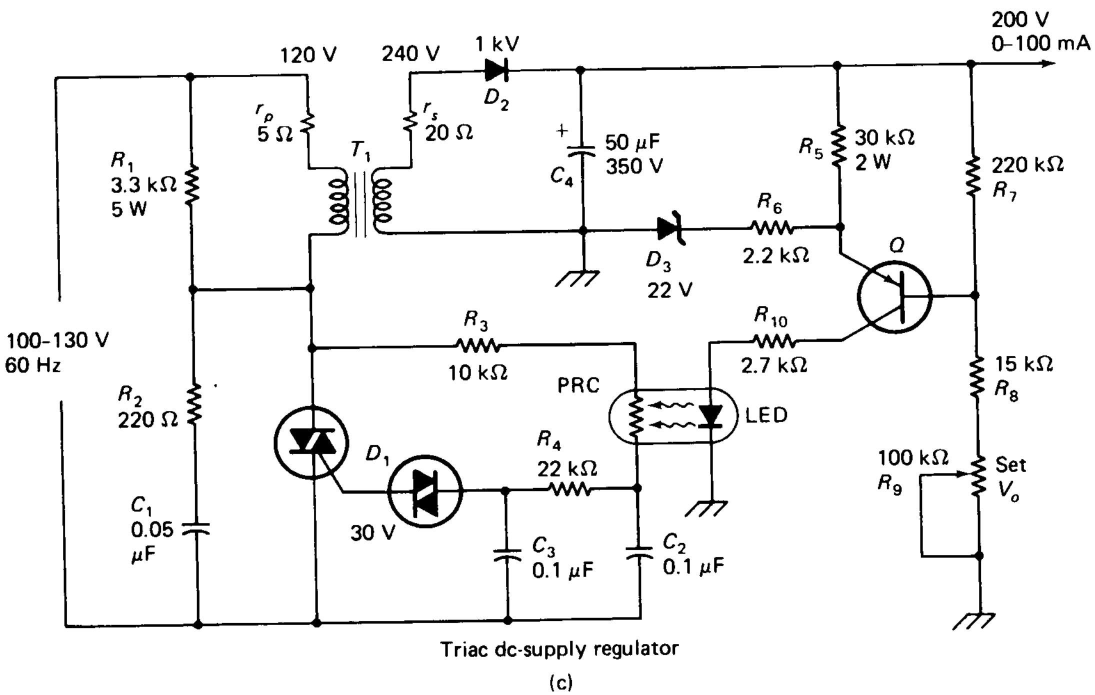


FIGURE 9-18 *(a) Turning off an inductive load at the zero-line-voltage point to avoid kickback. (b) Turning a lamp on and off at the zero-line-voltage points to minimize thermal shock. (c) High-voltage regulated supply using a triac on the low-voltage side of the transformer and optical coupling to the trigger circuit.*

1-A series pass transistor and a 900-W heat sink using conventional techniques. By regulating on the primary rather than the secondary side of the power transformer, we can get by with a 350-V 12-A triac on a 25-W heat sink. The circuit shown and tested is somewhat more modest because of the cowardice of the author, but the same principles are evident.

R_1 supplies hold-on current for the triac after the brief charging pulse of rectifier D_2C_4 is over. R_2 and C_1 form a snubber preventing false triggering of the triac. R_3, C_2, R_4, C_3, and D_1 form a triggering-delay network as in Fig. 9-18(c), except that the variable resistance is a photoresistive cell driven by an LED which senses the output voltage. R_5 and D_3 fix the emitter voltage of Q while R_6 provides degeneration to prevent self-oscillation. R_7, R_8, and R_9 form a voltage divider applying a fraction of V_O to the transistor base. If the output voltage drops, because of added load or drop in line voltage, V_B drops below V_E, turning Q more on, lowering the resistance of the PRC and increasing the triac conduction angle.

Output ripple is 30 V p-p at full load for this circuit, so an LC filter may be required between C_4 and R_5. Full-wave rectification would be of no avail because the triac inevitably triggers a bit earlier on the positive half-cycle, and the charge held on C_4 obscures the smaller pulse from the negative half-cycle. A filter choke between D_2 and C_4 could eliminate this problem, but its value must be kept small and the load current must not drop near zero if regulation is to be maintained.

10

STATISTICS OF COMPONENT VARIATION

10.1 OBJECTIVES AND TERMINOLOGY OF STATISTICS

Descriptive statistics attempts to summarize a mass of data in a few meaningful numbers. Omega University may have 5000 students receiving 20,000 individual grades, but we can speak of the average grade as being 2.72, and the increase in average grade as averaging 0.02 per year over the past 10 years, and the number of E grades as being 6% of the total.

Inferential Statistics attempts to make generalizations about a large group by gathering and analyzing data from a relatively small sample of that group. Thus, although there is no central data bank containing the height measurements of every 21-year-old female in the United States, and we do not have the resources to gather such data, we can nevertheless state with almost complete certainty what the average value is, what percentage is below 62 in., or above 66 in., or whatever other number we choose. This is possible because we have data from a small but carefully chosen sample of this unmanageably large group. Most of this chapter is devoted to techniques for drawing inferences about large batches of electronic components from analysis of smaller samples.

Population is the name given to the entire group under concern. Often it is an indefinite large number, as in the case of components under continuous production. We denote the number of items in the population with the letter N.

Sample is the part of the population for which data are gathered. A random sample (the only useful kind) is one in which each member of the population is equally likely to be selected for the sample. Obtaining a random sample is often more difficult than it would seem. A political survey would be nonrandom or biased if it were taken in a single city or region or if it were taken by phone (certain types of people tend to have unlisted numbers) or by mail (certain types tend not to respond) or in person (certain types don't answer the door for strangers). In industry, random sampling generally requires that components be selected from different boxes, at different times of the day, from the beginning, middle, and end of a production run. One of our major objectives will be to determine the minimum sample size required to adequately represent the population. We denote sample size by n.

Score is the measured value of a parameter, be it ohms, volts, microfarads, or whatever. We term them $X_1, X_2, \ldots, X_n$, and, where two samples are involved, $Y_1, Y_2, \ldots, Y_n$.

Mean is the statistical term for average. It is calculated by summing all the scores and dividing by the number of such scores. We denote population mean by μ and sample mean by $\overline{X}$ and $\overline{Y}$.

$$\mu = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{\Sigma X}{N} \tag{10-1}$$

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\Sigma X}{n} \tag{10-2}$$

Level of Confidence is our degree of certainty about an inference that is drawn from a sample. Say that we wish to know whether Ajax or Zeus brand resistors have a tighter tolerance. We select samples of 10 resistors from each company and find that the Ajax sample has an average deviation of 4%, whereas the Zeus sample averages 5% deviation. The Zeus representative complains that it was just bad luck that some of the worst of his company's resistors were selected and compared to some of the best from Ajax. In other words, the sample was not typical of what would be expected if we drew several such samples. So we pull 10 times the sample size—a total of 100 from each company—and find Ajax averaging 4.6% and Zeus averaging 4.3% deviation. Now the Ajax rep complains that the luck of the draw was against him. But it is harder to argue bad luck in such a large sample. If a man loses at poker one night, he may have had bad luck, but if he loses nine out of 10 nights, we become quite sure that he is simply a poor player.

We can calculate and express our level of confidence precisely as a percentage. A confidence level of 95% means that if we repeated our sampling test 100 times on two populations which were in fact identical, five of these tests would judge the populations to be different, just because the samples drawn would be atypical. A 95% level of confidence is sometimes spoken of as 0.05 *level of significance*. Figure 10-1 illustrates these concepts.

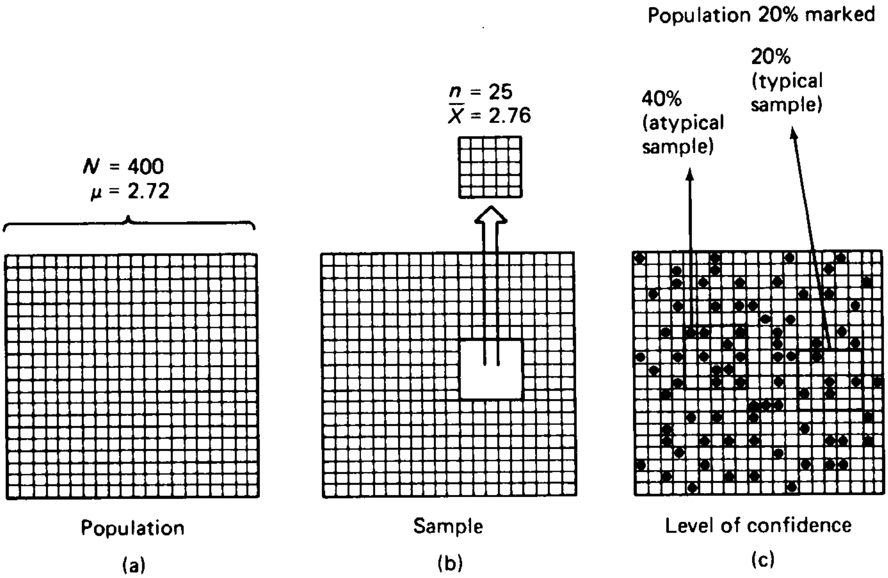


FIGURE 10-1 (a) Illustration of a population of 400 and its mean score μ. (b) Illustration of a sample of 25 from the population of 400, and the mean score of the sample $\bar{X}$. (c) Illustration of the possibility of drawing a sample that is not typical of the population.

Normal Distribution: Quantities of measured data can be summarized by constructing a bar graph (called a histogram) as shown in Fig. 10-2(a), which purports to represent the distribution of capacitance values of a sample of 24 nominal 1.0-μF capacitors. Class intervals of 0.1 μF are used. There are, for example, 5 capacitors in the interval between 1.1 and 1.2 μF. If a very large sample were taken, we might wish to use smaller intervals to obtain a more accurate picture of the distribution, and a histogram like the one in Fig. 10-2(b) might be obtained.

If the sample size is made indefinitely large and the class intervals are made indefinitely small, we obtain a smooth frequency-distribution curve, as shown in Fig. 10-2(c). In a remarkable number of real-life applications, the shape of this distribution curve approaches a mathematical ideal called the normal curve, or somewhat loosely, the bell curve. Grade-point averages, IQ test scores, height of army inductees, values of resistors produced by a given machine, and number of defective parts in a day's production run, to take a few examples, all tend to be normally distributed. Most of the techniques presented in this chapter assume that the population in question is reasonably close to a normal distribution.

Standard Deviation is a measure of dispersion or central tendency which is essential in dealing with the normal curve. Data that are widely scattered above and below the mean have a large standard deviation. Closely grouped data have a small

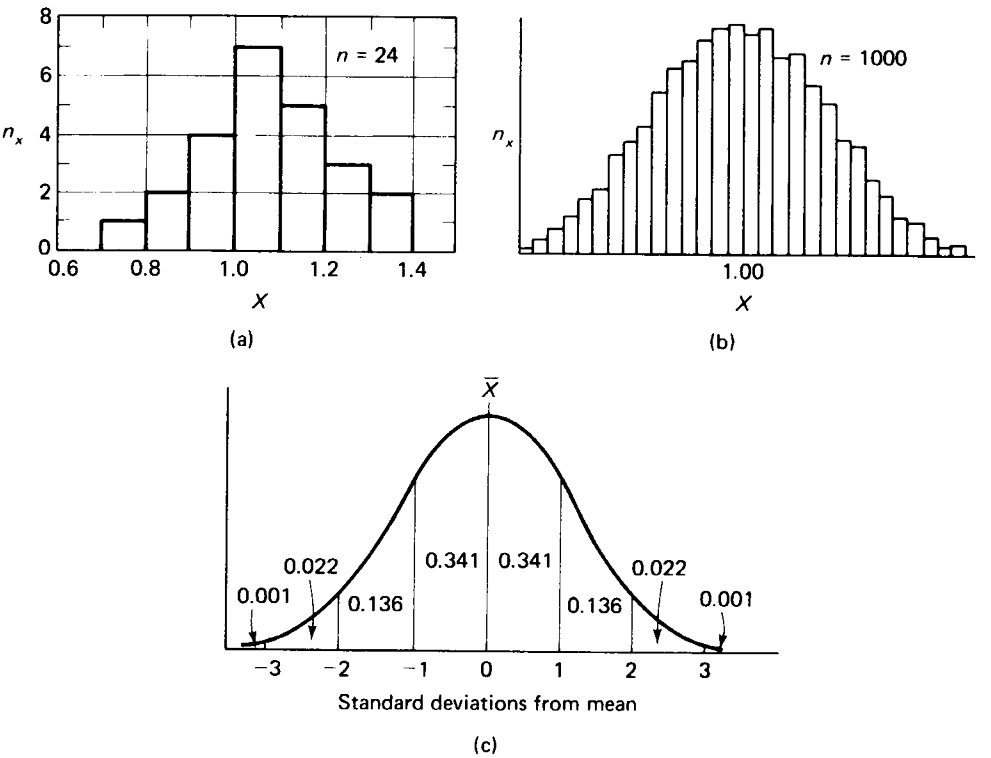


FIGURE 10-2 Histograms showing the distributions of scores: (a) small sample size; (b) large sample size; (c) theoretical normal distribution, showing the fraction of scores lying 1, 2, 3, and beyond 3 standard deviations from the mean. The total area adds to 1.000.

standard deviation. Standard deviation for an entire population is calculated as

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N^2}} = \sqrt{\frac{\Sigma X^2}{N} - \mu^2} \qquad (10\text{-}3)$$

where N is the total number of scores in the population. For a sample of n scores, standard deviation is

$$s = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}} = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\frac{\Sigma X^2 - n\overline{X}^2}{n-1}} \qquad (10\text{-}4)$$

The first forms given are the defining equations, but they may prove inconvenient for calculation since they require subtraction of the mean and squaring for each score. The second forms given are for calculation from raw score data, and the third forms are for calculation from raw scores once the mean has been obtained.

Areas under the Normal Curve: The total area under the normal curve is 1.000. The number of scores included within ±1 standard deviation is 0.6826 of the total. The fraction lying within ±2 standard deviations is 0.9544, while 0.9972 of all scores fall within ±3 standard deviations. This is illustrated in Fig. 10-2(c). Table 10-1 gives "tail areas," that is, the proportion of scores lying beyond the given number of standard deviations from the mean in one direction.

TABLE 10-1 *Tail areas under the normal curve, one side. Where deviation from the mean $X - \bar{X} = x$, x/σ = standard deviations above or below the mean. A = area from x/σ to infinity, one side.*

$\frac{x}{\sigma}$	A	$\frac{x}{\sigma}$	A	$\frac{x}{\sigma}$	A	$\frac{x}{\sigma}$	A
0.0	0.5000	1.0	0.1587	2.0	0.0228	3.0	0.00135
0.1	0.4602	1.1	0.1357	2.1	0.0179	3.1	0.00097
0.2	0.4207	1.2	0.1151	2.2	0.0139	3.2	0.00069
0.3	0.3821	1.3	0.0968	2.3	0.0107	3.3	0.00048
0.4	0.3446	1.4	0.0808	2.4	0.0082	3.4	0.00034
0.5	0.3085	1.5	0.0668	2.5	0.0062	3.5	0.00023
0.6	0.2743	1.6	0.0548	2.6	0.0047	3.6	0.00016
0.7	0.2420	1.7	0.0446	2.7	0.0035	3.8	0.00007
0.8	0.2119	1.8	0.0359	2.8	0.0026	4.0	0.00003
0.9	0.1841	1.9	0.0287	2.9	0.0019	4.2	0.00001

10.2 DESIGN OF STATISTICAL EXPERIMENTS

Now that you have the basic terminology in hand, you may feel ready to set up a statistical experiment and analyze the results. However, it is often said that a little knowledge is a dangerous thing, and you should be aware that the information contained in this chapter is only enough to make you dangerous. Where matters of importance are involved it is recommended that you consult a full-fledged statistics text or, better still, a full-fledged statistician. The primary benefit of this chapter should be to make you aware of the possibilities of statistical analysis and to allow you to converse intelligently with those who can handle the analysis properly. This section consists primarily of a list of thou-shalt-nots in designing your experiment.

Randomize Your Data: This almost always requires that a special sample be drawn for the experiment, which involves staying late to catch the second-shift foreman, and digging down to the bottom rear of a warehouse stack for a sealed carton that has no other reason for being opened. The boss may object and tell you to pull your samples all from one already-available box, but the results cannot then be supported. Each member of the population must have an equal chance to be selected for the sample, so discriminating against hard-to-obtain data cannot be tolerated.

Every dog has its day, and every problem-ridden production line will, by luck, eventually produce a crate of gems. In a competitive economy the temptation to zero in on such good fortune and produce a "statistical study" is almost irresistible. Such tactics are pernicious, but they are perpetrated every day on the gullible.

Nonnormal distributions can destroy the validity of sampling tests that assume a normal distribution. The values of the resistors in a particular shipment, for example, may be normally distributed except that all values above and below 2.5 standard deviations from the mean have been eliminated from the population by the vendor [Fig. 10-3(a)]. Bimodal distributions [Fig. 10-3(b)] may be observed where two machines are used to produce components, which are then batched together. Drastically skewed distributions [Fig. 10-3(c)] are occasionally encountered.

A histogram plot with 8 to 20 class intervals and a sample size of 30 to 100 is probably the most practical device for verifying that a distribution is essentially normal. Mathematical "goodness of fit" tests do exist, but they are quite complex.

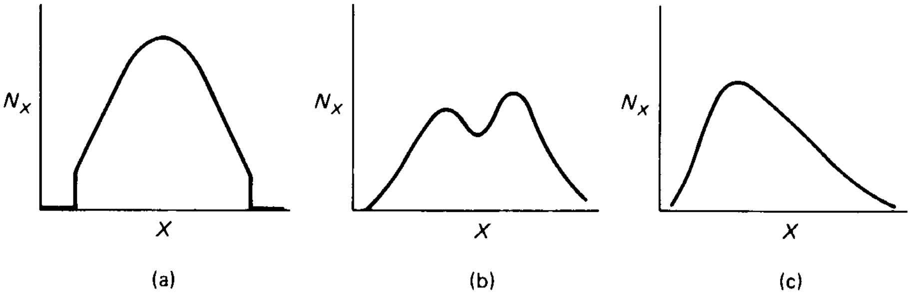


***FIGURE 10-3** Nonnormal distributions: (a) selective elimination; (b) bimodal; (c) skewed.*

Hawthorne Effect: Several decades ago industrial engineers at a major telephone-equipment assembly plant decided to investigate the effect of lighting on worker productivity. They found that when lighting levels were increased production went up—and that when lighting levels were decreased production went up—and that when they simply stated their intent to change the lighting, production went up. Of course, the workers were reacting not to the lighting but to the fact that they were being studied. It may have made them feel "under the gun" or it may have made them feel especially important, either feeling tending to increase production.

The serendipitous lesson of the Hawthorne study should not be overlooked, but where a statistical study of some variable other than worker attitude is being undertaken, care must be taken to see that worker attitude does not bias the results. One tactic is to hide from the workers the fact that they or their products are under study, at least until after the samples are obtained. Where this is not possible, it is advisable to form two groups of workers for comparison, and tell both that they are part of a study. Under no circumstances should the performance of people who knew they were under study be compared with the performance of people working under the "same old grind."

Statistical Errors: In sampling statistics the possiblity always exists that we will accept a bad batch of components or reject a good batch just because we have been unfortunate enough to draw an atypical sample. Large sample sizes minimize this possibility, but they do not eliminate it.

10.3 DIFFERENCE BETWEEN MEANS

For our first statistical study we will analyze a sample to determine whether there is a discernible difference between the means of two populations. Consider the case of two competing brands of capacitors. We wish to determine which brand has the higher breakdown voltage. We cannot be confident that samples of one or two of each brand will be representative of the populations, but the test is destructive and fairly time consuming so we don't want to use a larger sample than necessary. How can we determine when we have a large enough sample to establish with a desired level of confidence that one brand is better than the other?

We begin by randomly selecting five capacitors of each brand. We calculate the mean and standard deviation of each sample according to equations 10-2 and 10-4. We then calculate the value of a parameter known as t:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_x^2 + s_y^2}{n}}} \tag{10-5}$$

where n is the number of items in each sample. We then consult Table 10-2 of t values to determine what, if any, significance can be attached to the difference in sample means $\bar{X} - \bar{Y}$. If the level of confidence thus established is unsatisfactory, we can draw five additional capacitors of each brand and repeat the calculations for $\bar{X}$, $\bar{Y}$, s_x, s_y, and t with $n = 10$. If the results are still unsatisfactory, we can proceed to $n = 15$, 20, 30, and 100 in an attempt to establish the desired level of confidence that one population mean really is greater than the other.

***TABLE 10-2** Critical values of t.*

	Level of Confidence					
n	80%	90%	95%	98%	99%	99.9%
3	1.53	2.13	2.78	3.75	4.60	8.61
5	1.40	1.86	2.31	2.90	3.36	5.04
10	1.33	1.73	2.10	2.55	2.88	3.92
15	1.31	1.70	2.05	2.47	2.76	3.67
20	1.30	1.69	2.03	2.43	2.71	3.57
30	1.29	1.67	2.00	2.39	2.67	3.47
100	1.28	1.65	1.97	2.35	2.60	3.33

Table 10-2 is quite abbreviated to suit the purposes of this introductory chapter. More complete tables of t, which may be found in any statistics text, will no doubt list degrees of freedom (df) in place of the sample size (n) column. The concept of degrees of freedom and how it relates to sample size is too involved to attempt an explanation at this time, but understand that the two are not equal.

EXAMPLE 10-1

Five brand X and five brand Y capacitors were randomly selected and destructively tested for breakdown voltage with the following results:

X (V)	Y (V)
450	550
510	520
470	490
580	550
470	570

Can we state with confidence that one brand has a higher average breakdown voltage than the other?

Solution

$$\bar{X} = \frac{\Sigma X}{n_x} = \frac{450 + 510 + 470 + 580 + 470}{5} = 496$$

$$s_x^2 = \frac{\Sigma (X - \bar{X})^2}{n - 1}$$

$$= \frac{(450 - 496)^2 + (510 - 496)^2 + (470 - 496)^2 + (580 - 496)^2 + (470 - 496)^2}{5 - 1}$$

$$s_x = 51.8$$

Similarly, $\bar{Y} = 536$ and $s_y = 31.3$:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_x^2 + s_y^2}{n}}} = \frac{496 - 536}{\sqrt{\dfrac{51.8^2 + 31.3^2}{5}}} = -1.48$$

Consulting Table 10-2, we find that $t = 1.48$ is sufficient to establish an 80% confidence that Y averages a higher breakdown voltage than X. Statisticians consider 95% confidence to be acceptable for general applications, but require 99% confidence where the consequences of a wrong decision would be severe. We would probably want to take an additional five samples of each brand and calculate t with $n = 10$ in an attempt to gain a higher level of confidence in our decision.

Note that the t test, as it is called, merely establishes a level of confidence that one population mean is greater than the other. It does not establish how much difference there is between μ_x and μ_y. In Example 10-1, our best estimate of the population means is $\mu_x = 496$ V and $\mu_y = 536$ V, but we are by no means 80% confident that these are the true population means or that the true means differ by this amount.

If political and economic considerations were such that we would require demonstration of, for example, a 25-V difference in the means before we would consider switching from brand X to brand Y, we could simply subtract 25 V from each brand Y score in the sample before conducting the t test. The result would be our level of confidence that μ_y exceeds μ_x by at least 25 V.

10.4 DIFFERENCE IN DISPERSION

Sometimes we are interested not in the means of two populations, but in the narrowness of their distribution curve. In comparing two resistor-molding machines, a resistor manufacturer would choose the one that produced the smaller spread of values, since this machine would produce fewer out-of-tolerance resistors. To establish whether a discernible difference in dispersion of two populations exists, equal samples are drawn from each population and the standard deviation of each sample is computed. The larger s is then divided by the smaller s, and the quotient is squared, producing a parameter called F. A table of critical values of F is then consulted to determine the significance of the difference in standard deviations (Table 10-3).

$$F = \frac{s_x^2}{s_y^2} \quad \text{or} \quad F = \frac{s_y^2}{s_x^2} \tag{10-6}$$

TABLE 10-3 Critical values of F.

n	Level of Confidence 90%	98%
5	6.4	16.0
10	3.2	5.4
15	2.5	3.7
20	2.1	3.0
30	1.9	2.4
100	1.4	1.6
200	1.3	1.4

EXAMPLE 10-2

Two machines were set up to produce 1000-Ω resistors. A sample of 10 was drawn from each machine's run with the following results:

$X(\Omega)$		$Y(\Omega)$	
1001	1011	959	1002
1027	995	973	991
1035	1017	965	984
1004	1064	989	976
1052	1018	967	972

Can we state with confidence that one machine produces a closer grouping of resistors than the other?

Solution

$$s_x^2 = \frac{\Sigma\,(X-\bar{X})^2}{n-1} = \frac{\Sigma\,(X-1022.4)^2}{9} = 22.4$$

$$s_y^2 = \frac{\Sigma\,(Y-\bar{Y})^2}{n-1} = \frac{\Sigma\,(Y-977.8)^2}{9} = 13.4$$

$$F = \frac{s_x^2}{s_y^2} = \frac{22.4}{13.4} = 1.67$$

This value of F falls short of establishing a 90% confidence that population standard deviation σ_y really is smaller than σ_x, and we would again call for a larger sample. Table 10-3 and equation 10-6 reveal that samples of 200 will give 90% confidence in the existence of a difference when that difference i s only 13%. If this is considered trivial we would decline to take so large a sample and report essentially identical populations if a significant value of F could not be acieved with smaller samples.

10.5 CORRELATION

In many cases the parameter of interest cannot easily be measured directly, and we must seek to establish a correlation between the critical unmeasurable parameter and some other readily observable parameter. Such instances often arise when the critical parameter involves a destructive effect, such as insulation breakdown or fuse-element melting.

The author was once involved in the manufacture of a computer disc memory which was plagued by low-amplitude recording at certain of its 64 record heads. Replacing a defective head was a very expensive procedure since the entire unit was sealed in an inert gas. It was alleged that potential low-amplitude heads could be spotted by measuring their impedances before the memory was sealed and replacing heads with unusually low impedances. In other words, it was alleged that there was a high correlation between record amplitude and head impedance.

The idea of correlation is illustrated very nicely by a scatter diagram, as in Fig. 10-4. If the correlation is obviously strong, the diagram itself will be the most

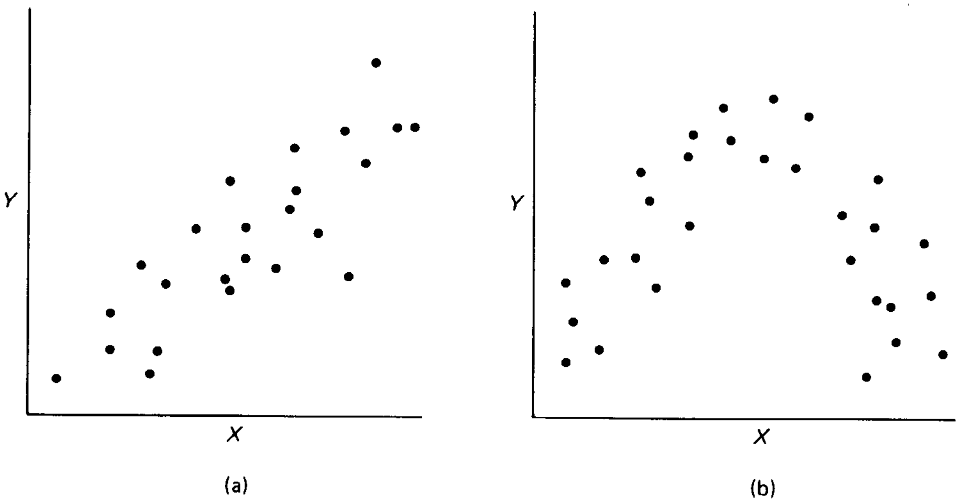


FIGURE 10-4 Strong linear correlation between X and Y is evidenced by the scatter diagram (a). Nonlinear correlations (b) are not amenable to the tests of Section 10.5.

convincing tool, but if the correlation is not so obvious, or if the sample size must be kept as low as possible, we will want to calculate the numerical correlation coefficient r. Note that the r computation assumes a linear relation between X and Y, as in Fig. 10-4(a). A curvilinear relation, as in Fig. 10-4(b) cannot be handled by this technique. The correlation coefficient is calculated as

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{(n-1)s_x s_y} \tag{10-7}$$

The second form of the equation is given to facilitate calculation where the standard deviations of the two samples have already been obtained. The procedure for calculating r using the first form is as follows:

- Obtain pairs of scores for a random sample of n components, X representing the hard-to-measure and Y the easy-to-measure parameters.
- Calculate the means $\bar{X}$ and $\bar{Y}$ according to equation 10-2.
- Form a table of differences from the mean for each pair of scores: $(X-\bar{X})$ and $(Y-\bar{Y})$.
- Extend the table to include $(X-\bar{X})^2$ and $(Y-\bar{Y})^2$ for each component.
- Add to the table the product $(X-\bar{X})(Y-\bar{Y})$ for each component.
- Sum all the $(X-\bar{X})^2$ values, then sum the $(Y-\bar{Y})^2$ values, then sum the $(X-\bar{X})(Y-\bar{Y})$ values for all n components.
- Use equation 10-7 to calculate r from $\Sigma(X-\bar{X})^2$, $\Sigma(Y-\bar{Y})^2$, and $\Sigma(X-\bar{X})(Y-\bar{Y})$.

Interpretation of Correlation Coefficient: The value of r will always lie between ± 1.00, negative r values indicating that high X values are associated with low Y values, and vice versa. The degree of relationship between X and Y is best estimated as r^2. Therefore if $r = 0.707$, our best estimate is that 0.50, or 50%, of the variation in X is predictable from a knowledge of Y. The other 50% variation will be random or at least unpredictable from Y. It should be apparent that the usefulness of correlation coefficients below 0.50 is limited, since this value would leave three-fourths of the variation of Y unaccounted for.

Note also that even where X and Y are highly positively correlated, we speak of a high X value as being *associated* with (not caused by) a high Y value. The statistics do not tell us whether Y causes X, or X causes Y, or X and Y are both caused by a third parameter Z, or whatever. In general, causation is a matter for philosophy, not statistics.

Significance of the Correlation Coefficient: Table 10-4 of critical r values shows the correlation coefficients required before we can state with various levels of confidence that *some correlation* in the direction indicated by the sign of r really does exist in the population. Our best estimate of the correlation in the population is the r value obtained from the sample, but our confidence in *this estimate* is much less than the confidence level obtained from Table 10-4. Indeed, the true r is equally likely to be above as to be below the r of the sample. This uncertainty should make us wary of small sample sizes and low confidence levels in correlation experiments.

TABLE 10-4 *Critical values of r.*

	Confidence Level			
n	90%	95%	98%	99%
5	0.805	0.878	0.934	0.959
10	0.549	0.632	0.716	0.765
15	0.441	0.514	0.592	0.641
20	0.378	0.444	0.516	0.561
30	0.306	0.361	0.423	0.463
100	0.166	0.197	0.232	0.257

EXAMPLE 10-3

Millivolts of record amplitude for magnetic heads in a sealed disc memory and head impedance before sealing are given for 10 randomly selected heads as X and Y, respectively, in Table 10-5. Determine the degree of relatedness between these two parameters.

Solution

Table 10-5 shows the values of the means, differences from the means, their squares and product, and the summation of all 10 squares and products. The

TABLE 10-5 Calculation of correlation coefficient.

X (mV)	Y (Ω)	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
42	630	−17.1	−103	292.4	10,609	1,761.3
68	870	8.9	137	79.2	18,769	1,219.3
53	590	−6.1	−143	37.2	20,449	872.3
81	850	21.9	117	479.6	13,689	2,562.3
47	680	−12.1	−53	146.4	2,809	641.3
51	570	−8.1	−163	65.6	26,569	1,320.3
63	710	3.9	−23	15.2	529	−89.7
58	760	−1.1	27	1.2	729	−29.7
75	820	15.9	87	252.8	7,569	1,383.3
53	850	−6.1	117	37.2	13,689	−713.7
$\bar{X}$	$\bar{Y}$			$\Sigma(X-\bar{X})^2$	$\Sigma(Y-\bar{Y})^2$	$\Sigma(X-\bar{X})(Y-\bar{Y})$
59.1	733			1,406.9	115,410	8,927

calculation of correlation coefficient proceeds:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}} = \frac{8927}{\sqrt{1406.9 \times 115{,}410}} = \mathbf{0.701} \qquad (10\text{-}7)$$

Our best estimate is that r^2, or 49%, of the variation in X is predictable from a knowledge of Y. Checking Table 10-4 for critical values of r with $n = 10$, we see that we can assert with more than 95% confidence that there is some degree of positive correlation between X and Y.

10.6 ESTIMATING THE DISTRIBUTION OF A POPULATION

A frequent industry problem is the elimination of out-of-tolerance components from a shipment. One-hundred-percent testing is an expensive solution, and may not be necessary if we can acquire an adequate level of confidence that the distribution of values about the mean is normal and sufficiently narrow. To do this, we draw a random sample and calculate its standard deviation s. The best estimate of population standard deviation is $\sigma = s$, and it is equally likely that the true σ lies above as below s. However, we would like to establish a high level of confidence that the true σ is not greater than some specific value, which we give as Es. We obtain the value of E for the required level of confidence from Table 10-6, and can then state with that level of confidence that the true σ does not exceed s by a factor greater than E. This is summarized as:

$$\sigma_{max} = Es \qquad (10\text{-}8)$$

Armed with this value of σ_{max}, we can consult Table 10-1 of areas under the normal curve to calculate the percent of the population that lies beyond some critical deviation from the mean. Where x is the critical deviation $X_{min} - \overline{X}$, we look up x/σ_{max} and determine the maximum fraction of cases lying beyond the critical deviation as A. Where upper and lower limits are both important, this last procedure must be repeated for the upper limiting value of the parameter X_{max}.

For this technique to be valid, it is essential that we establish that the population is normally distributed, especially in the "tail" areas of greatest deviation from the mean. This is best done by constructing a histogram (Fig. 10-2) for as large a sample as possible.

TABLE 10-6 *Values of E.*

	Level of Confidence			
n	80%	90%	95%	99%
10	1.19	1.29	1.37	1.52
15	1.15	1.23	1.30	1.43
20	1.13	1.20	1.26	1.37
30	1.11	1.17	1.21	1.30
100	1.06	1.09	1.12	1.16
300	1.03	1.05	1.07	1.10

EXAMPLE 10-4

A manufacturer of precision resistors receives an order for 100,000 resistors per month for the next 20 months, but the customer stipulates that if more than 50 of any month's shipment are found to be out of tolerance, the contract will be canceled. The manufacturer draws a random sample of 100 resistors from the line (nominal value 1000 Ω), measures their values with a four-digit 0.1% ohmmeter, and plots the histogram of Fig. 10-5. Will it be necessary to 100% test the resistors to ensure that fewer than 50 per 100,000 will be out of tolerance?

Solution

The mean and standard deviation of the sample are calculated:

$$\overline{X} = \frac{\Sigma X}{n} = 1001.87 \tag{10-2}$$

$$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}} = 2.003 \tag{10-4}$$

A 99% confidence level is chosen since the manufacturer must pass this test 20 separate times if trouble with the customer is to be avoided. A 1% chance of failure each month would accumulate nearly a 20% chance of hitting a failure after 20 months. Table 10-6 gives an E factor of 1.16 for 99% confidence with $n = 100$, so we express 99% confidence that the population standard deviation is not greater

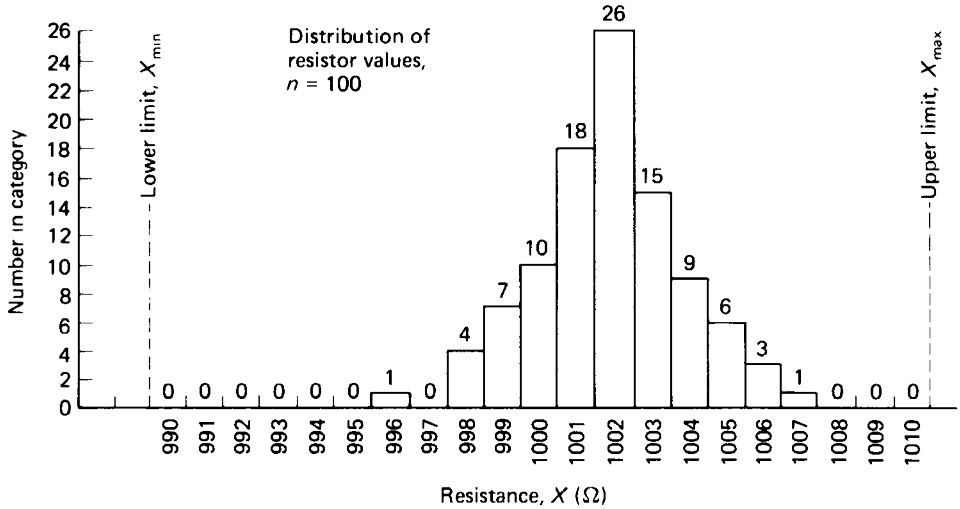


***FIGURE 10-5** Distribution of sample data for Example 10-4.*

than

$$\sigma_{max} = Es = 1.16 \times 2.003 = 2.32 \qquad (10\text{-}8)$$

The upper 1% limit is nearer to the mean. We calculate the number of standard deviations to this limit:

$$\frac{X_{max} - \bar{X}}{\sigma_{max}} = \frac{x}{\sigma_{max}} = \frac{1010 - 1001.87}{2.32} = 3.50$$

Consulting Table 10-1, we see that the number of scores lying beyond 3.5 standard deviations from the mean in one direction is 0.00023 of the total. This means that we are 99% sure that not more than $0.00023 \times 100{,}000 = 23$ units per shipment will be out of tolerance in this direction. In the opposite direction, the limiting X_{min} lies

$$\frac{990 - 1001.87}{2.32} = -5.12$$

standard deviations from the mean, and essentially no failures in this direction are anticipated. The results then establish a 99% confidence that the number of failures per 100,000 will be 23 or less, and it would seem unneccessary to undertake the expense of 100% testing.

Two worries haunt the manager who makes the decision of the foregoing example. First, we are assuming the population distribution to be normal. Figure 10-5 does not show any marked departure from the normal, so the assumption is reasonable, but we cannot be perfectly sure, especially at the tails where sample data are sparse. Second, the mean and deviation may change as the machine ages, new workers are put on the job, new batches of raw materials are used, and so on. This makes it desirable that the sampling test be executed each month and that the samples be drawn after the month's run is completed but before it is shipped, so that any changes on the line will be reflected in the sample.

10.7 CUMULATIVE ERRORS

A component manufacturer's tolerance specification generally takes the form of a guarantee that the components will not deviate from the marked value by more than the percentage specified. No information about the statistical spread of values within this tolerance range is implied, and we would have no grounds for complaint if a shipment of one thousand 100-Ω 5% resistors contained 500 that measured 95 Ω and 500 that measured 105 Ω.

Worst-Case Design assumes that the worst-possible combination of circuit values will eventually come up, and guarantees reliability in spite of it. To take a simple example, a voltage divider composed of two 100-Ω 5% resistors from a 20-V source could put out anything between 9.5 and 10.5 V, as shown in Fig. 10-6(a) and (b).

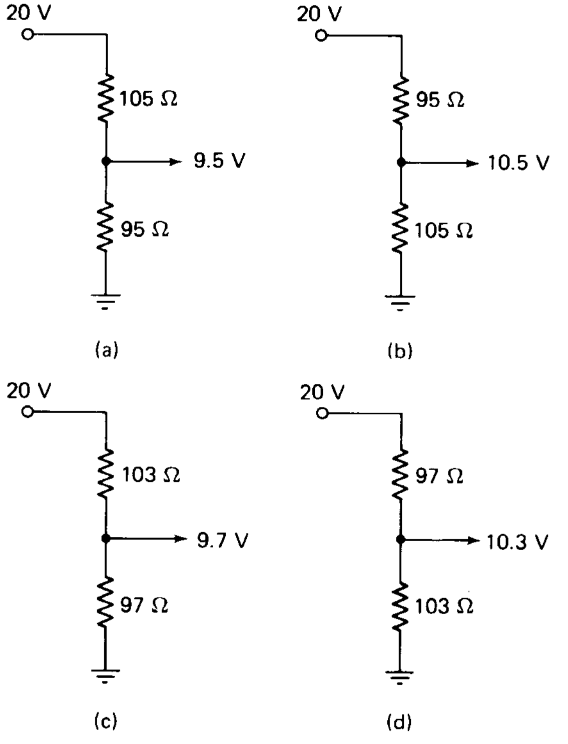


FIGURE 10-6 Voltage divider composed of two nominal 100-Ω 5% resistors: worst-case analysis of minimum (a) and maximum (b) outputs, and statistical analysis with 99.9% confidence of minimum (c) and maximum (d) outputs.

However, if we can determine that the actual distribution of component values is approximately normal, with the values occurring randomly, we may be able to employ a more economical statistical design technique.

Statistical design recognizes that the probability that the worst-case combination will actually occur may be so slight as to be negligible, especially where three, four, or more parameters must all go to their extreme limits together to produce the worst case. To understand why this is so, you must understand that cumulative probabilities multiply. For example, the probability of rolling a six on one die is $\frac{1}{6}$. The probability of rolling another six on a second die is also $\frac{1}{6}$. The probability of rolling a pair of sixes is therefore $\frac{1}{6}\times\frac{1}{6}$, or $\frac{1}{36}$. The probability of rolling four sixes with four dice is $\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6}$, or $\frac{1}{1296}$. The point to be taken here is that four events, each of which has a fairly high probability of occurrence taken alone, have an almost negligible probability of occurring together.

Let us assume that we have used the method of Section 10.6 to determine that the 100-Ω resistors for our voltage divider are normally distributed with a mean of 100.0 Ω and a standard deviation of 1.5 Ω. The 5% limits each lie 5/1.5, or 3.3, standard deviations from the mean, and Table 10-1 indicates that 0.00048 of the resistors will lie beyond the limits in each direction. Actually then, 0.00096, or about 0.1%, of the resistors will be rejects. Another way of saying this is that the probability that any one resistor will be at or beyond one of its worst-case limits is 0.00096. The probability that the other resistors will at the same time be at or beyond its other worst-case limit is the product 0.00096×0.00048, or about 1 chance in 2 million.

We can compute the probability of R_1 of the divider falling more than 2 standard deviations above the mean while R_2 falls more than 2 below. From Table 10-1 the area beyond $x/\sigma=2.0$ is 0.0228, and the product of two such probabilities is about 0.05%. Of course, it is equally probable that R_2 will be high while R_1 is low, so we have a total probability of about 0.1% that one resistor will be 3 Ω (2 standard deviations) high while the other is 3 Ω low. We are thus 99.9% confident that the voltage limits of the divider will be within 9.7 to 10.3 V, as shown in Fig. 10-6(c) and (d). It may be more economical to plan on $\pm3\%$ tolerance for the output voltage and troubleshoot the 0.1% of cases where this limit is exceeded than to purchase precision resistors to ensure a worst-case limit of $\pm3\%$.

Addition of Several Distributed Values: When a number of normally distributed component values are added (as in the case of series resistors), the mean and standard deviation of the total can be computed readily from the means and standard deviations of the components:

$$\mu_T=\mu_1+\mu_2+\mu_3+\mu_4+\cdots \tag{10-9}$$

$$\sigma_T=\sqrt{\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2+\cdots} \tag{10-10}$$

Where parallel combination is involved, resistance values can be converted to conductance values ($G=1/R$) and the conductance distribution determined from the foregoing equations.

EXAMPLE 10-5

Twenty-five 20-MΩ 5% resistors are connected in series to form a 500-MΩ high-voltage probe for a 20-kΩ/V meter. The resistors are randomly selected and are normally distributed about a mean of 20 MΩ, with the 5% limit representing 3 standard deviations. What is the tolerance of the composite probe?

Solution

Extending equation 10-10 for 25 values of σ, each equal to $(5\% \times 20\ \text{M}\Omega)/3$, or 0.33 MΩ:

$$\sigma_T = \sqrt{25\sigma_1^2} = \sqrt{25 \times (0.33\ M\Omega)^2} = 1.65\ \text{M}\Omega \qquad (10\text{-}10)$$

Taking the tolerance limit as 3σ (yielding a 0.25% chance of being out of tolerance), the ohmic limits are $\pm 3 \times 1.65$, or ± 4.95 MΩ. This is a tolerance of

$$\frac{\pm 4.95\ \text{M}\Omega}{500\ \text{M}\Omega} \approx \pm 1\%$$

This same technique can be applied to the subtraction of distributed values. A batch of axles may be required to be randomly fitted into a batch of drilled holes. If the distributions of axle and hole diameters are known, we can predict the probability that a clearance less than some specified value will be encountered. Equations 10-11 and 10-12 now become

$$\mu_{\text{clr}} = \mu_{\text{hole}} - \mu_{\text{axle}} \qquad (10\text{-}11)$$

$$\sigma_{\text{clr}} = \sqrt{\sigma_{\text{hole}}^2 + \sigma_{\text{axle}}^2} \qquad (10\text{-}12)$$

11

ELECTRICAL MEASUREMENT

11.1 DEFINITION OF TERMS

In making measurements we employ a number of terms which need to be accurately defined if they are to be intelligently used.

Sensitivity: a measure of the smallest input signal that will produce an observable output (i.e., *sensitivity* 1 mV rms), often expressed as a ratio of output to input (i.e., *sensitivity* 100 mV rms for full-scale deflection of 100 divisions).

Accuracy: the degree to which the measured value approaches the true value. Usually expressed as the *percent error* that may occur (i.e., *accuracy* $\pm 5\%$ of reading or $\pm 1.5\%$ of full-scale value, whichever is greater). Sometimes expressed as the maximum *amount of error* that may occur (i.e., *accuracy* ± 0.05 V).

Precision: the degree to which a measurement is sharply defined, expressed approximately by the number of significant figures in the reading. If a four-digit digital voltmeter has a precision of ± 1 mV, we will be able to use it to adjust two power supplies to a reading of 5.000 V with assurance that they are no more than 2 mV different in voltage. If the accuracy of the meter is $\pm 0.1\%$, the true voltage of the supplies will be sure to lie within the range 4.995 to 5.005 V.

Linearity: the degree with which a graph of the actual value of the quantity versus the reading of the indicator deviates from an ideal straight line between the end points of the actual curve. Actually, it is a measurement of *non*linearity, calculated as follows:

$$\text{normal linearity} = \frac{\text{max deviation of reading}}{\text{full-scale reading}} \tag{11-1}$$

Figure 11-1 provides an illustration of the properties of sensitivity, accuracy, precision, and linearity. There are many other definitions of linearity, the most common being independent linearity, in which the ideal straight line is adjusted for best fit to the actual curve, and zero-based linearity, in which the zero points are fixed together but the slope of the ideal line is adjusted for best fit.

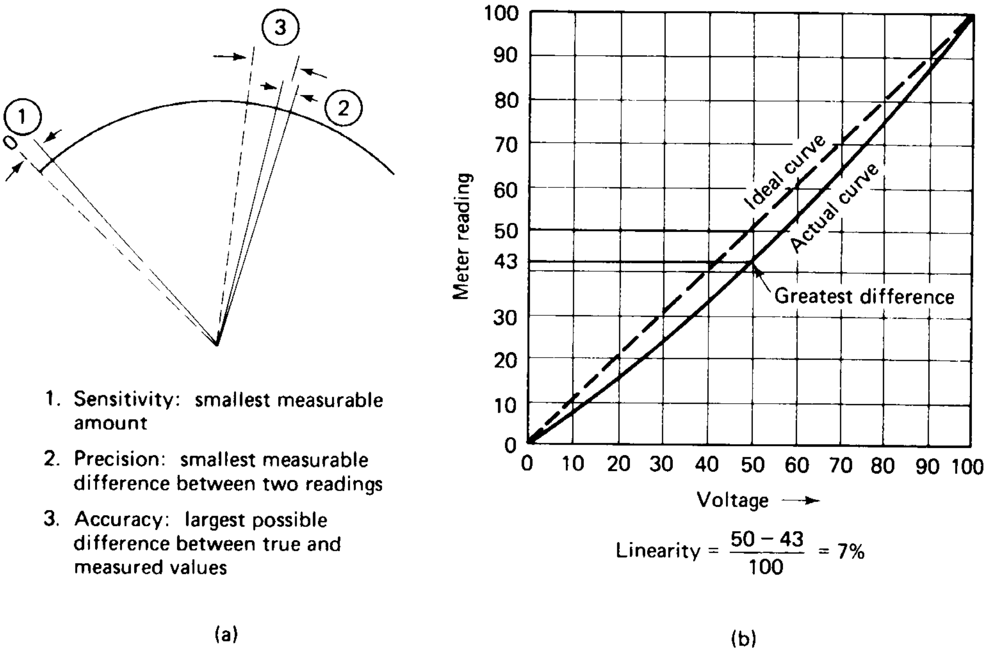


FIGURE 11-1 (a) Illustration of the definitions of sensitivity, precision, and accuracy. (b) Illustration of the definition of normal linearity.

Hysteresis: the amount that a quantity can change before the measured value changes. Often expressed by the graphic words *slop*, *stickiness*, or *dead zone*.

Drift: the amount that a measured value (or a quantity) changes per unit time in the face of temperature, aging, or other variables. For example, an oscillator may be specified as follows: stability, ± 20 Hz/h after $\frac{1}{2}$-h warmup. A standard cell may be specified: stability, $\pm 5\ \mu$V/yr.

11.2 ACCURACY OF ELECTRICAL STANDARDS

Electrical units were originally defined in terms of physical quantities in the CGS (centimeter–gram–second) system. For example, a coulomb of charge was defined in terms of the force (in $g \cdot cm/s^2$) between two plates held a certain distance (in centimeters) apart. With the technology that existed around the turn of the century such physical measurements were difficult to make accurately, so a system of *international units* based on direct electrical measurements was adopted. For example, the volt was defined in terms of a standard cadmium cell of precisely specified construction. The voltage produced by this cell was defined as 1.01830 international volts, and it was hoped that this voltage would prove to be the same as that derived by the CGS system. As technology improved, differences as large as 0.05% became apparent, and in 1948 the international system was replaced by the so-called *absolute* system, which was refined in 1960 to become the *Système International* (SI). A comparison of previous international units and the new SI units follows:

$$\begin{aligned}
1 \text{ international coulomb} &= 0.99985 \text{ C (SI)}\\
1 \text{ international ampere} &= 0.99985 \text{ A (SI)}\\
1 \text{ international volt} &= 1.00034 \text{ V (SI)}\\
1 \text{ international ohm} &= 1.00049\ \Omega \text{ (SI)}\\
1 \text{ international watt} &= 1.00019 \text{ W (SI)}\\
1 \text{ international farad} &= 0.99951 \text{ F (SI)}\\
1 \text{ international henry} &= 1.00049 \text{ H (SI)}
\end{aligned}$$

Accuracies on the order of 10 parts per million (0.001%) in voltage and other electrical units have been obtainable since the early days of electronics. Present standards are accurate to a few parts per million, and precise to better than 1 ppm. Time can be measured to better than 1 part in 80 billion (8×10^{10}). These accuracies are obtained through painstaking work and are available only from *primary standards* maintained at government-operated laboratories, such as the National Bureau of Standards. Commercially available *secondary standards* and measuring devices are available with voltage and resistance accuracies on the order of 5 to 10 ppm. Time signals can be generated independently with accuracies much better than 1 ppm. Timing accuracies approaching that of the primary standard are readily available from the Bureau's radio station WWV, but no one has yet found a way to broadcast a standard volt.

11.3 MEASUREMENT TECHNIQUES

Whenever measurements that you have taken are of any importance, you may be sure that they will be challenged—by a rival within the company seeking to impress the boss at your expense, by a plant or government inspector, by a customer, or by a patent attorney. It is therefore of great importance that you acquire the habit of taking and recording measurements in a form that inspires confidence and stands up under scrutiny.

Format for Data: Every recorded measurement should contain an explanation of what the data refers to. The number 2800 scribbled diagonally across a scrap of paper will probably not impress an FCC engineer investigating an alleged violation. The following format may carry more weight:

> *19 March, 1979, 9:30 A.M. Field strength of transmitter second harmonic measured with Acme test fixture at distance 200 ft N of tower.* $V_2 =$ *2800* μ*V rms.* *John Doe*

This brings up several additional points. Every data record should be dated. It is most frustrating to open a file and find two calibration charts for a test instrument, one of them obviously obsolete, but neither of them dated.

The instrument used in making the measurement should be identified, by serial number if several of the same type are available. This makes it possible to go back and check the calibration of the meter if your results are questioned. It also provides an indication of the degree of accuracy to be expected. A recorded value of 275 V dc (DVM) would be expected to be more accurate than a recorded value of 275 V dc (VOM).

Units should be included after each number recorded. Don't leave the milli- or micro- to be assumed, and always distinguish between rms, peak, and peak-to-peak measure of ac.

The person taking the data or preparing the chart should sign it. This tells the reader whom to see for more details and lends to the legal validity of the document.

Instrument Limitations: The accuracy specification of a meter is not the only limit on its performance. Manufacturers often omit any reference to other limitations, but you should check out your instruments on these points:

- What are the upper and lower cutoff frequencies on the ac ranges? Some VOMs and DVMs lose response above a few hundred Hz, while others are good beyond 100 kHz.
- Does your meter respond to dc components on its ac scales?
- Does your meter respond to the peak, average, or true rms value of a signal on its ac scales? Only special true-rms voltmeters provide valid readings for non-sine waves. Peak- and average-responding meters are calibrated with the rms equivalent for sine waves only.
- Does your meter respond to the peak or average value of an ac signal riding on a dc level on its dc ranges?

Measurement Pitfalls: Here are a number of tips for interpreting measured data:

- Meter accuracy is usually specified by allowable error as a percentage of full scale. Thus a reading of 35 V on the 100-V scale of a 3%-accuracy meter must be interpreted as having limits of 35 ± 3 V, or 32 to 38 V. We may acquire confidence that a particular meter does better than this by comparing it with a lab standard, but the manufacturer's certification is limited to what in this case is a nearly 10% error.
- Meter calibration often changes abruptly with scale changes. A meter may show 2.45 V on the 2.5-V scale, and suddenly jump to an indication of 2.6 V when switched to the 10-V scale, without any actual change in the measured voltage taking place. Where a string of readings from 2.0 to 5.0 V is to be taken, it may be wisest to take all readings to the 10-V scale to avoid the disturbing jump in the curve at 2.5 V due to range change.
- Beware of subtracting two pieces of data. The uncertainty associated with each reading can make the result of a subtraction almost meaningless. For example:

  90 V ±3% *minus* 80 V ±3% could be:

  92.7 V − 77.6 V = **15.1 V**, or at the other extreme:

  87.3 V − 82.4 V = **4.9 V**

  The nominal result is 10 V, but the limits of error are ±51%.
- The numeral 5 appears frequently as the last digit of measured data. If a meter pointer falls between 47 and 48 V, we may record the reading as 47.5 V. If a digital meter keeps alternating between 1020 Hz and 1019 Hz, the data may be recorded as 1019.5 Hz. If such data are later rounded off in the course of calculation, it would be unwise to always round the digit preceding the 5 to the next higher (or the next lower) integer, as this would tend to artificially bias the average of all the data one way or the other. The number 75 lies exactly midway between 70 and 80, and we should not arbitrarily bias our choice by rounding always to one or the other. A commonly accepted rule for rounding off fives is to round higher if the preceding digit is odd, and round lower if the preceding digit is even. This changes the odd numbers to even and keeps the even numbers even in the last digit of the rounded number. On the average, as many numbers will be rounded high as low, and the average of the data will be unbiased. Some examples of rounding fives from 4 to 3 figures follow:

  $4875 \approx 4880 \qquad 3195 \approx 3200$

  $4885 \approx 4880 \qquad 3205 \approx 3200$

11.4 THE DECIBEL SYSTEM

The decibel (dB) system of measure is widely used in the audio, radio, TV, and instrument industries for comparing two signal levels. There are two main points to remember in applying this system:

- A decibel measure is a ratio, not an amount. It tells how many times greater or how many times less a signal is compared to some reference.
- Decibel measure is nonlinear. 20 dB is not twice as much as 10 dB.

The original unit "bel" was defined as the logarithm (to base 10) of the power ratio of two signals. This unit proved too large, and was split into tenths, or decibels:

$$\alpha_{\mathrm{dB}} = 10 \log \frac{P_2}{P_1} \text{ dB} \tag{11-2}$$

where P_1 is the reference power (the input signal level, or perhaps an industry-wide standard-reference level) and P_2 is the signal in question (the output signal). A little reflection, or a few moments with a calculator will reveal that α_{dB} is always *positive* for a signal greater than the reference level, and always negative for a signal less than the reference level. Zero dB means equal levels—no gain, no loss. For reference, the transposition of equation 11-2 is given:

$$\frac{P_2}{P_1} = \log^{-1} \frac{\alpha_{\mathrm{dB}}}{10} = 10^{\alpha_{\mathrm{dB}}/10} \tag{11-3}$$

Choosing a log-based system allows a tremendous range of power ratios to be encompassed using only two-digit numbers, without sacrificing discernment at the low end of the scale:

$$1 \text{ dB} = 1.26{:}1 \text{ power ratio}$$
$$50 \text{ dB} = 100{,}000{:}1 \text{ power ratio}$$

The graphs of Fig. 11-2 illustrate this point. Both graphs show the same data.

Voltage and Decibels: Voltage is much easier to measure than power, so it would be convenient to apply the decibel system to voltage measure. Since $P = V^2/R$, and squaring a number is accomplished by multiplying its logarithm by 2, we can write equations 11-2 and 11-3 as

$$\alpha_{\mathrm{dB}} = 20 \log \frac{V_2}{V_1} \tag{11-4}$$

$$\frac{V_2}{V_1} = \log^{-1} \frac{\alpha_{\mathrm{dB}}}{20} = 10^{\alpha_{\mathrm{dB}}/20} \tag{11-5}$$

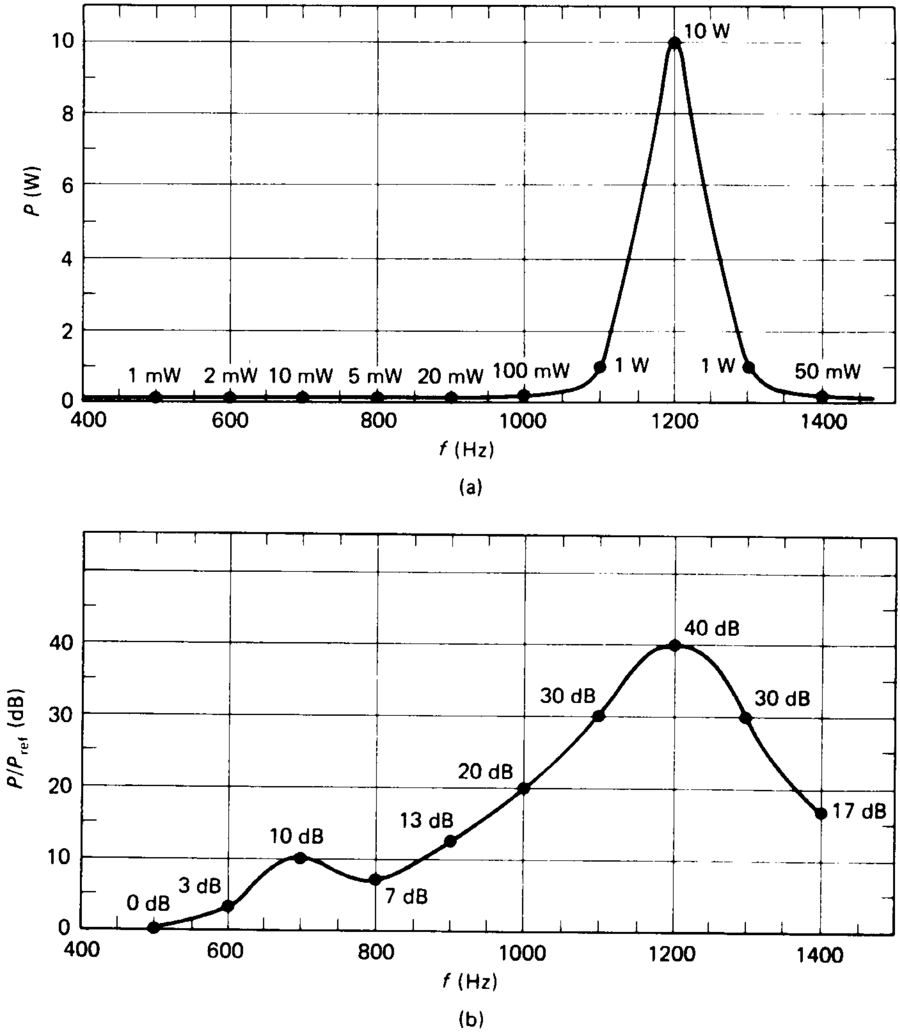


***FIGURE 11-2** (a) A graph of power output versus frequency for a tuned amplifier so compresses the power levels below 100 mW that they cannot be seen. (b) The same data, converted to decibels with 1 mW = 0 dB, shows all power levels clearly.*

These equations rest on the assumption that R_1 for the reference voltage V_1 is the same as R_2 for the measured voltage V_2. This restriction is often ignored in practice, but the results obtained thereby are not correct. If the impedance driven by V_1 is not known to be equal to the impedance driven by V_2, the impedances must be measured, the two power levels calculated, and equation 11-2 applied. If the values of R_1 and R_2 cannot be determined, the decibel system cannot properly be applied.

Decibel Standards: The audio industry has agreed upon a standard reference signal of 1 mW on a 600-Ω load, which equates to 0.775 V rms. Meters calibrated to this standard can only be read directly when used on a 600-Ω line. When used on other lines, the reading must be corrected as follows:

EXAMPLE 11-1

An audio meter reads −4 dB across a 16-Ω speaker. What is the actual level?

Solution

The power error will be as the ratio of the 600-Ω standard to the 16-Ω measurement, since $P = V^2/R$.

$$\frac{P}{P'} = \frac{600}{16} = 37.5$$

Expressed in dB, this ratio is

$$\alpha_{\text{dB}} = 10 \log P/P' \text{ dB} = 10 \log 37.5 \text{ dB} = 15.7 \text{ dB}$$

This number of dB must be added to the reading:

$$x = -4 + 15.7 = \mathbf{11.7 \text{ dB}}$$

The television industry sometimes uses a standard reference of 1 mV rms on a 75-Ω line.

Adding Decibels: Signal transmission in a large system involves multiplying by some factor A_v each time an amplifier is encountered and dividing by a loss factor F_v each time a signal splitter, attenuator, or length of cable is encountered. However, if all the system component voltage ratios (V_o/V_{in}) were specified as decibels (positive for gain, negative for loss), we could determine the total system gain by algebraically *adding* the dB contribution of each component. This is because adding logs of numbers effectively multiplies the numbers, and subtracting logs effectively divides. This is often done in practice, because additions and subtractions are much easier to keep track of than multiplications and divisions.

EXAMPLE 11-2

Figure 11-3 shows a television master-antenna system, with the input signal expressed in dB (reference being 1 mV) and each component expressed by the dB ratio of its V_o/V_{in}. An attenuator is to be placed in the line to TV set 51 to provide a 10-dB signal level at the set. What size of attenuator should be used?

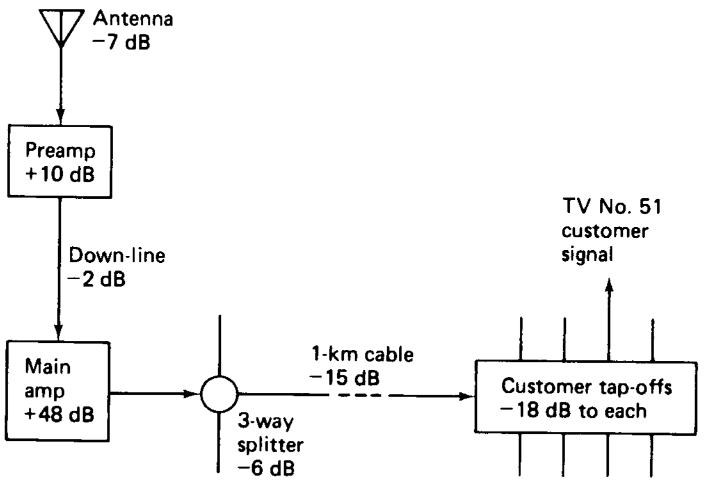


***FIGURE 11-3** Example 11-2 illustrates the advantage of decibel measure in system work.*

Solution

The signal level is $-7 + 10$, or $+3$ dB out of the preamp, $+1$ into the main amp, $+49$ into the splitter, $+43$ out of the splitter, $+28$ into the customer tap-off box, and $+10$ dB out of the tap-off. No attenuator is therefore required.

Anyone not yet convinced of the value of working in dB is invited to work Example 11-2 using voltage ratios. Use 450 μV for the antenna signal, $\times 3.2$ for the preamp, $\div 1.2$ for the down line, $\times 260$ for the main amp, $\div 2$ for the splitter, $\div 5.6$ for the 1-km cable, and $\div 7.9$ for the tap-off box. Shoot for a customer signal of 3000 μV, and try to do it in your head.

Fast Decibel Conversions: People who work frequently with decibels soon commit the following table to memory. With it, any voltage ratio can be converted instantly to dB (or the reverse) without using a pencil or calculator.

$\alpha\,(dB)$	V_2/V_1	$\alpha\,(dB)$	V_2/V_1
0	1	20	10
1	1.12	30	31.6
3	1.41	40	100
6	2.0	50	316
10	3.16	60	1000

To get values not listed in the table, just remember that adding dB is the same as multiplying voltage ratios, and subtracting dB is the same as dividing ratios.

EXAMPLE 11-3

Convert to voltage ratios: 26 dB, -55 dB.

Solution

$$
\begin{aligned}
26\text{ dB} &= 20\text{ dB} + 6\text{ dB} \\
&= 10 \quad \times 2 = \mathbf{20\ voltage\ ratio} \\
55\text{ dB} &= 50\text{ dB} + 6\text{ dB} - 1\text{ dB} \\
&= 316 \quad \times 2 \quad + 1.12 = \mathbf{564\ voltage\ ratio}
\end{aligned}
$$

The minus sign simply indicates a level 564 times below the reference, which we may wish to express as 1/564 voltage ratio.

12

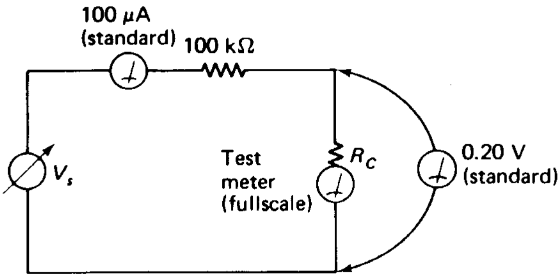
BASIC METERING CIRCUITS

12.1 THE D'ARSONVAL METER

The D'Arsonval meter consists of a moving coil of fine wire suspended between the poles of a permanent magnet, and has long been the most popular analog measuring instrument. The *taut-band* meter, a recent improvement, replaces the jeweled pivot with a torsion band suspension to eliminate hysteresis (stickiness) in the movement of the pointer. Precision meters have mirrored scales to eliminate *parallax*, the error incurred when the eye views the meter from a side angle. Keeping the mirror image of the pointer directly behind the pointer ensures that you are viewing the meter "head-on."

D'Arsonval meters are specified in terms of their full-scale current sensitivity and their coil resistance. Sensitive D'Arsonval meters (often called galvanometers) can be used to produce a voltmeter, ammeter, or ohmmeter of any desired range using the Ohm's law calculations outlined in the next three sections.

The resistance of an unknown meter should *not* be determined by measuring it with an ohmmeter because a VOM may inject enough current to destroy a sensitive meter. Instead, bleed full-scale current though the meter and a high-value resistor ($100\ k\Omega - 10\ k\Omega$), measure the voltage across the meter coil and calculate $R_c = V_c/I$, as illustrated in Fig. 12-1.



$$I_{fs} = 100\ \mu A$$

$$R_C = \frac{V_C}{I} = \frac{0.20}{100\ \mu A} = 2.0\ k\Omega$$

***FIGURE 12-1** Circuit for determining the full-scale current and coil resistance of a meter without damaging it.*

12.2 THE VOLTMETER CIRCUIT

A voltmeter circuit is shown in Fig. 12-2(a). A *multiplying resistor*, R_m, is placed in series with the meter to drop all the input voltage except that required by the meter itself. To calculate the value of the multiplying resistor:

- Calculate the full-scale voltage drop across the meter itself: $V_c = I_{fs} R_c$.
- Select the full-scale voltage input, V_{in}.
- Subtract to find the voltage on the resistor: $V_{Rm} = V_{in} - V_c$.
- Calculate R_m using the full-scale meter current: $R_m = V_{Rm}/I_{fs}$.

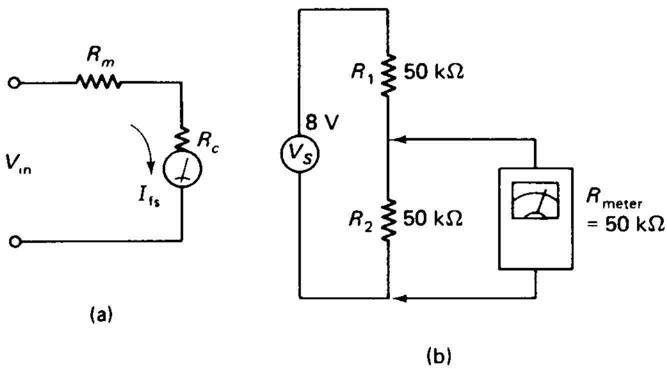


***FIGURE 12-2** (a) The basic voltmeter circuit consists of a multiplying resistance R_m in series with the meter movement. (b) If the total meter resistance $R_m + R_c$ is not many times greater than the circuit Thévenin resistance $R_1 \| R_2$, connecting the meter will seriously load down the voltage being measured.*

EXAMPLE 12-1

A meter with $I_{fs} = 100\ \mu A$ and $R_c = 2\ k\Omega$ is to be used to make a 5-V full-scale metering circuit. Find R_m.

Solution

$$V_c = I_{fs}R_c = 100\ \mu A \times 2\ k\Omega = 0.2\ V$$
$$V_{Rm} = V_{in} - V_c = 5.0 - 0.2 = 4.8\ V$$
$$R_m = \frac{V_{Rm}}{I_{fs}} = \frac{4.8\ V}{100\ \mu A} = 48\ k\Omega$$

Voltmeter Loading is an undesirable effect produced when the meter circuit siphons away an appreciable portion of the current from the circuit under test. The result is an actual lowering of the voltage between the test points when the meter is connected. The extent of voltmeter loading can be calculated by Ohm's law if the meter resistance and circuit resistances are known.

EXAMPLE 12-2

Find the error that will be encountered if the meter of Example 12-1 is used to measure V_{R2} in the circuit of Fig. 12-2(b).

Solution

The voltage before the meter is connected is $\frac{1}{2} \times 8$ V, or 4 V, by inspection. The meter circuit resistance totals 50 kΩ ($R_c + R_m$).

$$R_2 \| R_{meter} = 50\ k\Omega \| 50\ k\Omega = 25\ k\Omega$$
$$R_T = R_1 + R_2 \| R_{meter} = 50\ k\Omega + 25\ k\Omega = 75\ k\Omega$$
$$V_{R2} = V_S \frac{R_2 \| R_{meter}}{R_T} = 8 \times \frac{25\ k\Omega}{75\ k\Omega} = 2.67\ V$$
$$\epsilon = \frac{2.67 - 4}{4} = -33\%$$

Voltmeter loading error can be held to a maximum of 10% if the meter has a resistance at least 10 times the lower of R_1 or R_2 in Fig. 12-2(b), to 1% if R_{meter} is 100 times R_1 or R_2, and 0.1% if R_{meter} is 1000 times R_1 or R_2.

Voltmeter Resistance is, therefore, an important property with which the technician should be familiar. Electronic meters (VTVMs, DVMs, and FET meters) generally have input resistances of about 10 MΩ. The standard oscilloscope input impedance is 1 MΩ, or 10 MΩ when used with a $\times 10$ probe. The common portable VOM has an input sensitivity of 20 kΩ/V, which means that it has a resistance of 200 kΩ on the 10-V scale, 2 MΩ on the 100-V scale, 5 MΩ on the 250-V scale, and so on.

Many instruments also have an appreciable capacitance in shunt with their input resistance, and this adds to the loading effect at high frequencies. For example, a 'scope with 20 pF of input capacitance measuring a 1 MHz signal presents an input impedance of about 8 kΩ capacitive, and its loading effect must be reckoned with this in mind.

If two metering devices are available (e.g., a VOM and an oscilloscope), the loading effect of one can actually be observed on the other as they are alternately connected and disconnected to the points under test. This technique is often more practical than trying to calculate the relative impedances of the circuit versus the meter.

12.3 AMMETER CIRCUITS

A galvanometer is converted to an ammeter of any desired range by connecting a *shunt resistor* across it which bypasses all the measured current except that required by the meter itself. The circuit is shown in Fig. 12-3(a). To calculate the value of the shunt:

- Calculate the full-scale voltage across the meter: $V_{fs} = I_{fs}R_c$.
- Select the full-scale input current I_{in}.
- Subtract the meter's full-scale current to find the current that must be carried by the shunt resistor: $I_{Rsh} = I_{in} - I_{fs}$.
- Calculate R_{sh} using the full-scale voltage which appears across the meter and shunt resistor together: $R_{sh} = V_{fs}/I_{Rsh}$.

EXAMPLE 12-3

A 0 to 1-mA meter movement with a coil resistance of 50 Ω is to be converted to a 0 to 15-mA meter. Find the required shunt resistance.

Solution

$$V_{fs} = I_{fs}R_c = 1 \text{ mA} \times 50\ \Omega = 50 \text{ mV}$$

$$I_{Rsh} = I_{in} - I_{fs} = (15 - 1)\text{mA} = 14 \text{ mA}$$

$$R_{sh} = \frac{V_{fs}}{I_{Rsh}} = \frac{50 \text{ mA}}{14 \text{ mA}} = 3.57\ \Omega$$

Ammeter Loading: Notice that this metering circuit has a resistance of 50 || 3.57, or 3.33 Ω, which must be considered when the meter is inserted in a line. This resistance will cause a line-voltage drop of 50 mV at full-scale current. Generally, these facts will present no serious problem, but in some cases meter voltage V_{fs} may be an appreciable fraction of the source voltage V_S, or changes in the load current may produce objectional output voltage changes due to the IR drop across the

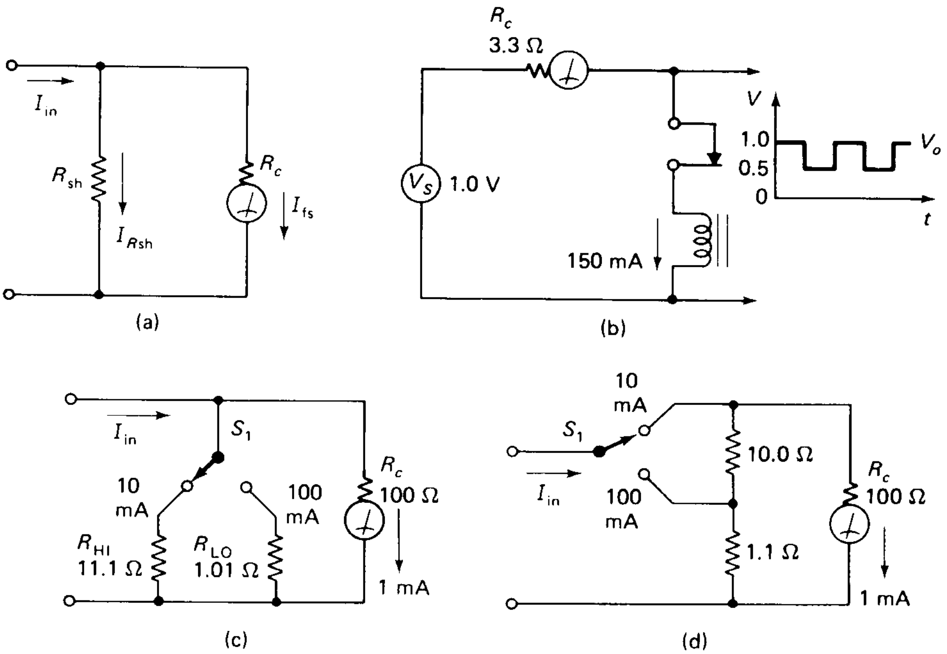


***FIGURE 12-3** (a) Basic ammeter circuit. (b) Ammeter voltage drop or loading. (c) Multirange ammeter. (d) Ayrton shunt to prevent overloading the ammeter during switching.*

meter. Both of these *ammeter loading* problems are represented in extreme degree by Fig. 12-3(b). The meter here is 150 mA with a 3.3-Ω coil.

Multiple-Range Ammeters can be constructed by simply switching in different shunt resistors, as shown in Fig. 12-3(c). If this circuit is used, however, it is essential that a *make-before-break* (also called *shorting*) type of switch be used. Otherwise, the entire input current would flow through the sensitive meter coil while the switch wiper was in transition between positions, and a bent meter pointer or burned-out coil would be the likely result.

If such a switch is not available, an *Ayrton shunt* arrangement, shown in Fig. 12-3(d) can be used. Note that the measured current is interrupted during switch transitions with the Ayrton shunt. This may be objectionable in some applications, such as those involving inductive loads.

12.4 OHMMETER CIRCUITS

A series ohmmeter circuit is shown in Fig. 12-4(a). In operation the probe lines are first shorted together, and zeroing resistance R_z is adjusted for full-scale meter deflection. Then the unknown resistor is connected and its value is read from a specially calibrated scale.

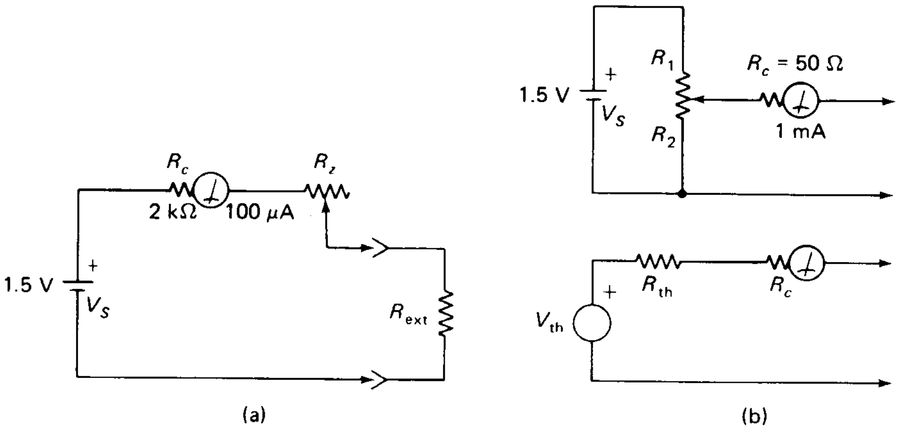


FIGURE 12-4 (a) Basic series ohmmeter circuit. (b) Lower-scale ohmmeter and its Thévenin equivalent for analysis (Example 12-6).

Any ohmmeter constructed according to this circuit will have a right-hand limit of 0 Ω and a left-hand limit of infinity. Nevertheless, ohmmeter sensitivities do vary, and we can conveniently quantify this by specifying the center-scale ohms reading for a particular circuit. Higher-sensitivity ohmmeters (those having higher center-scale ohms readings) can be produced by increasing the battery voltage or by increasing the galvanometer current sensitivity, and conversely, lower-scaled meters can be produced by lowering either of these parameters.

EXAMPLE 12-4

Find the value of R_z and the center-scale reading for the ohmmeter of Fig. 12-4(a).

Solution

First the total series circuit resistance for full-scale current is found:

$$R_T = \frac{V}{I_{fs}} = \frac{1.5}{0.1\ \text{mA}} = 15\ \text{k}\Omega$$

The meter coil has 2 kΩ of resistance, and the battery is assumed to have zero resistance.

$$R_z = R_T - R_c = 15\ \text{k}\Omega - 2\ \text{k}\Omega = \mathbf{13\ k\Omega}$$

The total resistance at half-scale current is

$$R_{T(1/2)} = \frac{V}{I_{1/2}} = \frac{1.5\text{V}}{0.05\ \text{mA}} = 30\ \text{k}\Omega$$

The meter and R_z account for 15 kΩ, so $R_{1/2}$ is found:

$$R_{1/2} = R_{T(1/2)} - R_z - R_c = 30\ \text{k}\Omega - 13\ \text{k}\Omega - 2\ \text{k}\Omega = \mathbf{15\ k\Omega}$$

EXAMPLE 12-5

Find the battery voltage required to produce a center-scale reading of 100 kΩ using the circuit and meter movement of Fig. 12-4(a).

Solution

Although shortcuts are possible, the most general solution is to write equations for full-scale deflection with zeroing resistance only (R_z includes R_c in this example) and partial deflection with the required resistance added:

$$V_S = I_{fs} R_z$$

$$R_z = \frac{V_S}{I_{fs}}$$

$$V_S = I_{ctr}(R_z + R_{ext}) = I_{ctr}\left(\frac{V_S}{I_{fs}} + R_{ext}\right)$$

$$= 0.05 \text{ mA}\left(\frac{V_S}{0.1 \text{ mA}} + 100 \text{ k}\Omega\right)$$

$$= 0.5\ V_S + (100 \text{ k}\Omega \times 0.05 \text{ mA})$$

$$0.5\ V_S = 5 \text{ V}$$

$$V_S = \mathbf{10} \text{ V}$$

Any desired deflection (other than half scale) could have been specified by choosing the appropriate value in place of I_{ctr}.

An expanded low-ohms range can be obtained by lowering the effective source voltage, as shown in Fig. 12-4(b).

EXAMPLE 12-6

Find the values of R_1 and R_2 in Fig. 12-4(b) such that the available 1.5-V battery produces a one-quarter scale deflection for a resistance of 600 Ω.

Solution

First it is necessary to draw the Thévenin equivalent of V_S, R_1, and R_2. Then we proceed as in Example 12-5.

$$V_{th} = I_{fs}(R_{th} + R_c) = 1 \text{ mA}(R_{th} + R_c)$$

$$R_{th} + R_c = \frac{V_{th}}{1 \text{ mA}}$$

$$V_{th} = I_{qtr}(R_{th} + R_c + 600\ \Omega)$$

$$= 0.25 \text{ mA}\left(\frac{V_{th}}{1 \text{ mA}} + 0.6 \text{ k}\Omega\right)$$

$$= 0.25\ V_{th} + 0.15 \text{ V}$$

$$= 0.2 \text{ V}$$

Now it is only necessary to determine R_1 and R_2 to divide the 1.5 V down to 0.2 V while making their Thévenin resistance equal the required zeroing resistance.

$$R_{th} + R_c = \frac{V_{th}}{I_{fs}} = \frac{0.2\text{ V}}{1\text{ mA}} = 200\ \Omega$$

$$R_{th} = 200\Omega - R_c = 200 - 50 = 150\ \Omega$$

$$1.5\text{ V}\frac{R_2}{R_1 + R_2} = 0.2\text{ V}$$

$$\frac{R_1 R_2}{R_1 + R_2} = 150\ \Omega$$

$$R_1 + R_2 = \frac{R_1 R_2}{150}$$

$$= \frac{1.5\text{ V} \times R_2}{0.2\text{ V}}$$

$$\frac{R_1 R_2}{150\ \Omega} = \frac{1.5\text{ V} \times R_2}{0.2\text{ V}}$$

$$R_1 = \frac{150 \times 1.5\text{ V}}{0.2\text{ V}} = 1125\ \Omega$$

$$R_2 = \frac{R_1 R_{th}}{R_1 - R_{th}}$$

$$= \frac{1125 \times 150}{1125 - 150} = 173\ \Omega$$

EXAMPLE 12-7

Note that in the circuit of Fig. 12-4(b), R_1 and R_2 draw current from the battery even with no external resistor connected. Meet the specifications of Example 12-6 by shunting the meter to decrease its sensitivity, thus avoiding this problem. Refer to Fig. 12-5.

Solution

$$R_{ZT} = \frac{V_S}{I_{fs}} = \frac{1.5}{1\text{ mA}} = 1500\ \Omega$$

$$V_{AB} = I_{qtr} R_{ZT} = 0.25\text{ mA} \times 1.5\text{ k}\Omega = 0.375\text{ V}$$

$$I_{Rext} = \frac{V_{Rext}}{R_{ext}} = \frac{1.5 - 0.375}{600} = 1.875\text{ mA}$$

$$R_{sh} = \frac{V_{AB}}{I_{Rsh}} = \frac{0.375}{1.875 - 0.25} = 231\ \Omega$$

The low-ohms circuit offered in Fig. 12-5 is the one that is generally used in commercial instruments, but it does considerably increase the current in the external resistor, whereas the circuit of Fig. 12-4(b) does not.

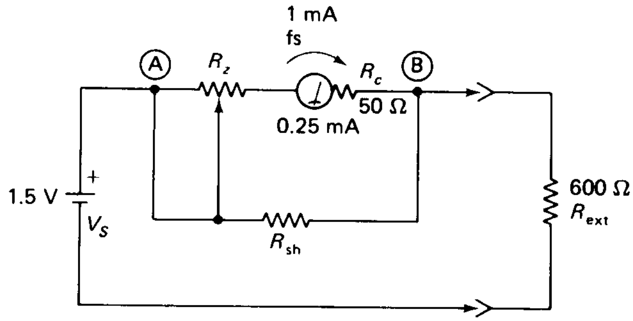


FIGURE 12-5 Low-scale ohmmeter produced by lowering the meter sensitivity.

The Shunt Ohmmeter circuit of Fig. 12-6 is used occasionally where a very low scale ohmmeter is required. It also has the disadvantage of drawing battery current continuously. Its scale has zero on the left, and hence it cannot use the same scale as the higher-range series meters on a $\times 1$, $\times 10$, $\times 100$ basis. The zeroing resistor is adjusted for full scale (infinity) with the probes open circuited. Very low ranges can be realized with this circuit only if the meter-coil resistance is low.

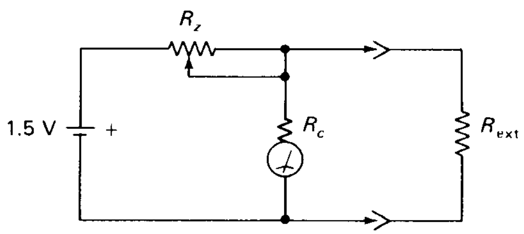


FIGURE 12-6 Shunt ohmmeter for low-ohms measurement.

12.5 METER PROTECTION

D'Arsonval movement meters can typically withstand overloads of $\times 5$ or $\times 10$ without damage. Very fast-acting *instrument fuses* are available in ratings from a few amperes down to a few milliamperes for the protection of meter movements, but this range falls short of being able to protect the very common (and very expensive) meter movements in the 50- to 100-μA range. Solid-state electronic circuit breakers provide the ultimate protection, but a simpler and very effective expedient is shown in Fig. 12-7. The current drawn by a silicon junction diode is less than 1 μA for voltages below 0.4 V, but reaches about 1 A at 1 V forward. Thus if a meter movement has a full-scale voltage between 0.1 and 0.4 V (which is common) a pair of diodes connected across it will not conduct in the meter's normal operating range, but will turn on heavily and shunt large overload currents around it. Of course, serious overloads would destroy the diode, but an inexpensive $\frac{1}{4}$-A fast fuse can be used to open the line before this happens.

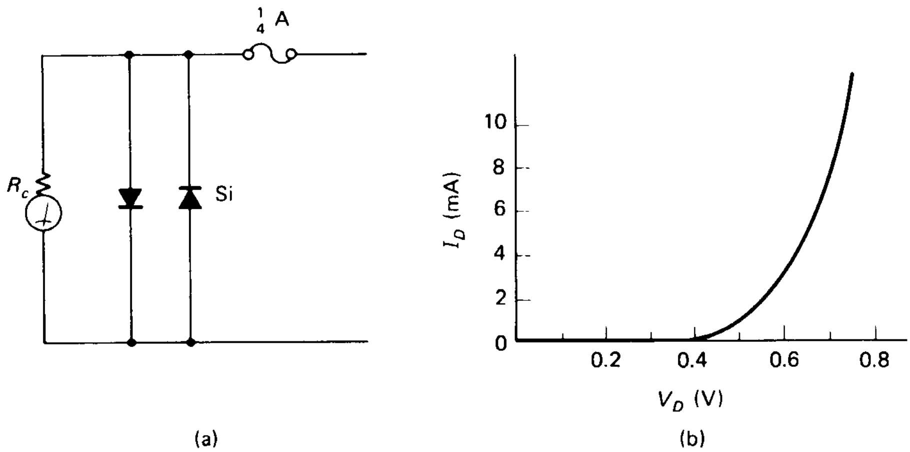


***FIGURE 12-7** (a) Silicon power diodes protect the meter movement by limiting coil voltage to about 1.0 V. (b) Silicon-diode curves.*

12.6 EXPANDED AND COMPRESSED SCALES

Occasionally, it is desired to have a meter scale that does not read linearly from zero to full scale. One example of this is a 117-V line monitor, which may read from 105 to 125 V but will certainly never read in the 0 to 100-V range. The circuit of Fig. 12-8(a) can be used to provide an expanded 100 to 130-V scale for easier reading.

Tuning meters and null detectors are more convenient if they are very sensitive to small signals, yet do not overload and pin on large signals. A silicon diode in the circuit of Fig. 12-8(b) can be used to reduce the sensitivity of the meter at midscale by switching part of the input current around the meter.

EXAMPLE 12-8

What is the full-scale input voltage of the metering circuit of Fig. 12-8(b)?

Solution

$$V_{R2} = I_{fs} R_2 = 0.1 \text{ mA} \times 12 \text{ k}\Omega = 1.2 \text{ V}$$

$$I_{R3} = \frac{V_{R2} - V_D}{R_3} = \frac{1.2 \text{ V} - 0.6 \text{ V}}{1.2 \text{ k}\Omega} = 0.5 \text{ mA}$$

$$V_{R1} = (I_{R2} + I_{R3}) R_1$$

$$= (0.1 \text{ mA} + 0.5 \text{ mA})(8 \text{ k}\Omega) = 4.8 \text{ V}$$

$$V_{in} = V_{R1} + V_{R2} = 4.8 + 1.2 = \mathbf{6.0} \text{ V}$$

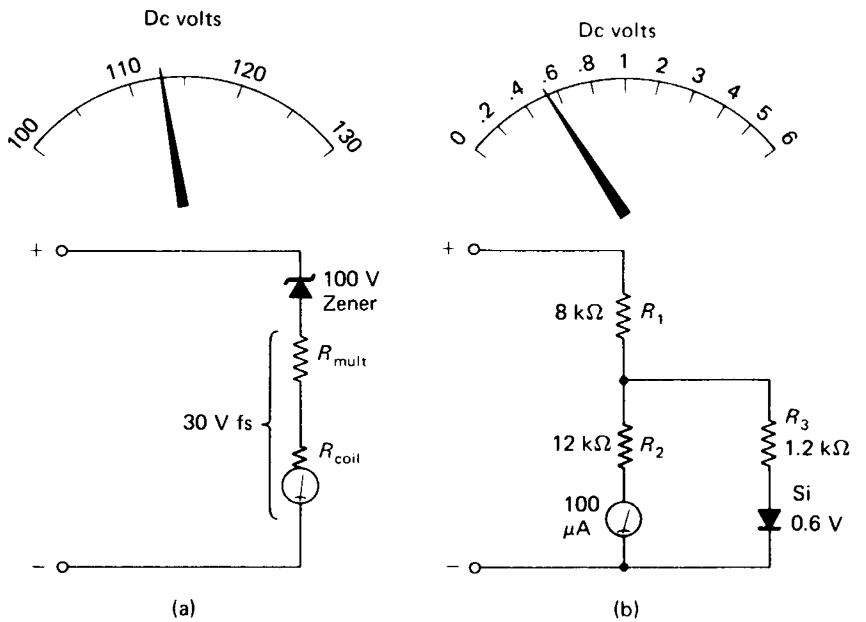


***FIGURE 12-8** (a) High-end-expanded meter circuit. (b) High-end-compressed meter circuit.*

12.7 AC METERS

Ac ammeters and voltmeters are generally constructed using two electromagnets, rather than one electromagnet and one permanent magnet as in the D'Arsonval meter. These two-coil meters are called *electrodynamometers*. The torque that turns the pointer is proportional to the product of the currents in the two coils, which is positive for either current direction if they are connected in series. Torque is then proportional to I^2, so scale linearity is a problem. Sensitivity is also a problem, since the fixed coil cannot generally rival the field strength of the permanent magnet in the D'Arsonval meters. Electrodynamometer ammeters down to 100 mA ac and voltmeters down to 5 V ac full scale are commonly available.

D'Arsonval-movement microammeters are often used in conjunction with solid-state rectifiers to provide ac voltage indication. However, even the special metallic rectifiers that are generally used have forward voltage drops in excess of 0.1 V, resulting in a 0.1 to 0.2-V dead zone at the bottom end of the scale. This dead zone is not significant on high-voltage scales, but causes serious nonlinearities on scales of 5 V and below. The voltage drop is generally held to be too great to tolerate in a current meter, so rectifier-type ammeters are not generally used.

Figure 12-9 shows a simple half-wave-rectifier type ac voltmeter. The second diode, shown in dashed lines, is used to shunt out reverse meter currents when metallic rectifiers, which are very leaky in the reverse direction, are employed. It also allows source current to flow in both directions so that a capacitor can be

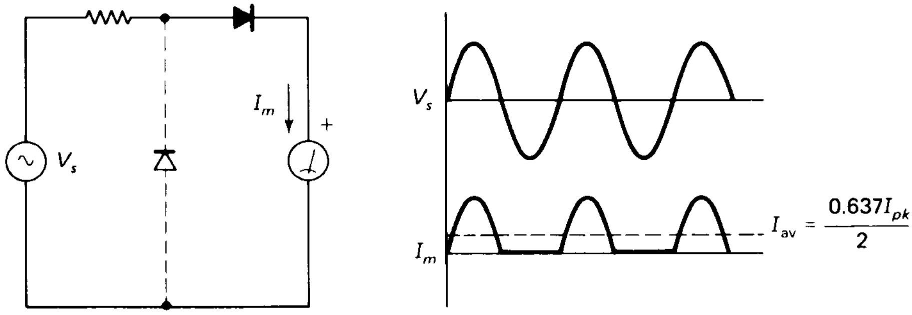


FIGURE 12-9 Rectifier-type ac voltmeter.

inserted in series with the circuit. This will keep any dc that may be present from influencing the ac reading. As illustrated, the meter current flows only on one half-cycle, and varies with the sine-wave voltage. The meter responds to the average value of this current which, for large sine waves, is

$$I_{av} = 0.318 \frac{V_{pk} - V_D}{R} \tag{12-1}$$

where V_D is the diode voltage drop and R includes both the series multiplier resistance and the meter-coil resistance.

EXAMPLE 12-9

The meter in Fig. 12-9 is 100 μA full scale, with a 2-kΩ coil. The diode is silicon with $V_D = 0.6$ V. What value of multiplier resistance will make a 10-V-rms full scale?

Solution

$$R_T = 0.318 \frac{V_{pk} - V_D}{I_{av}} \tag{12-1}$$

$$= 0.318 \frac{(1.41 \times 10 - 0.6)\text{ V}}{0.1\text{ mA}}$$

$$= 42.93\text{ k}\Omega$$

$$R_{mult} = R_T - R_{coil} = (42.93 - 2.00)\text{ k}\Omega = \mathbf{40.93\text{ k}\Omega}$$

12.8 WATTMETERS

The most common type of wattmeter is an electrodynamometer with one fixed coil of heavy wire to sense current and one movable coil of fine wire to sense voltage to the load. The wires may be brought out as shown in Fig. 12-10(a) or (b). Such a meter will respond to true power, even if a reactive load produces a phase difference between V and I, or if V_s contains a dc component or is not a sine wave.

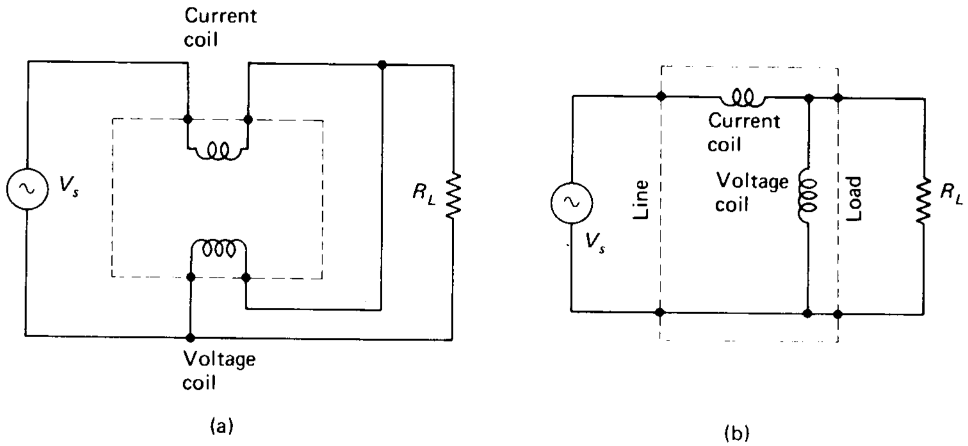


FIGURE 12-10 *Two possible connections of a wattmeter.*

13

POTENTIOMETERS AND BRIDGES

13.1 THE POTENTIOMETER CIRCUIT

The potentiometer circuit has long been popular for the accurate measurement of small dc voltages. The basic idea, as illustrated in Fig. 13-1(a) is to use a variable voltage divider to tap off a fraction of the voltage from a known source, and to balance this against an unknown voltage. When the voltage divider is adjusted for perfect balance, no current flows through the indicator (galvanometer) or the unknown source, and the voltage can be read from the calibrated pot or slide-wire scale. This ability to measure the voltage of a source without drawing current from it is important, because many voltage sources used in instrumentation have a high internal resistance and the meter loading effect (see Section 12.2) would invalidate a simple voltmeter reading.

Figure 13-1(b) shows two additions to the basic potentiometer. The first is a calibrate rheostat to compensate for aging of the working battery. A highly accurate reference cell is switched in place of the unknown and the potentiometer dial is set to this cell's voltage. The rheostat is then adjusted for zero on the galvanometer. The second addition is a high-value resistor in series with the meter to limit the meter and unknown source current to a harmless value while the operator is hunting for a balance. Once an approximate balance is found, the switch is closed to permit maximum sensitivity for a more accurate balance.

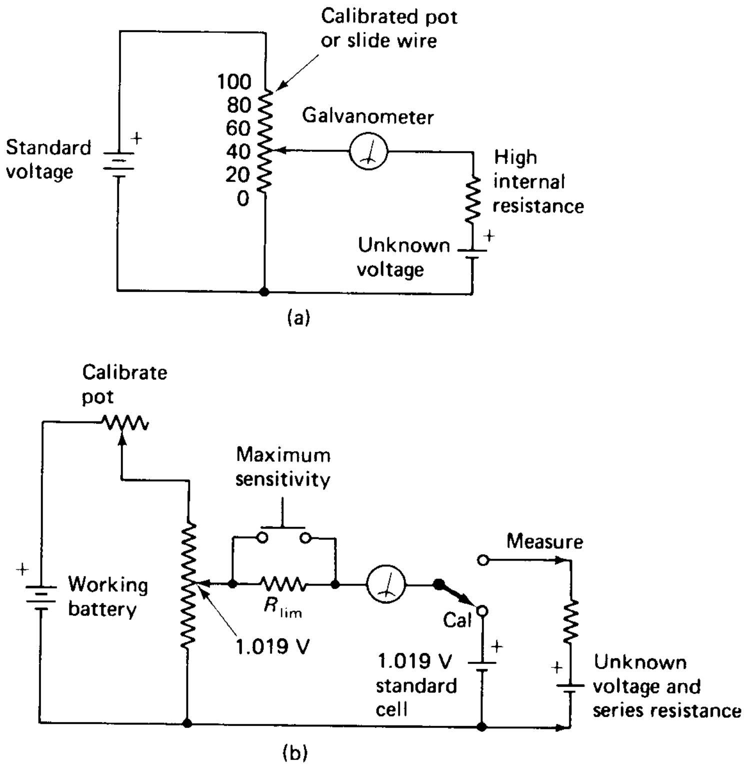


***FIGURE 13-1** (a) Basic potentiometer circuit for measuring an unknown voltage without drawing current from it. (b) Practical potentiometer with a working cell, meter sensitivity switch to assist in finding a null, and calibrating position.*

The industry is literally filled with instruments which are basically potentiometers. Functionally, they can be sorted into three types:

Laboratory Instruments are highly accurate (0.1% or better), manually balanced, and are used for laboratory measurements or for calibration of voltmeters, power supplies, and so on.

Pen-Chart Recorders have a pen attached or geared to the voltage-dividing resistance arm, and a roll of graph paper which moves at a constant rate under the pen. The bridge is automatically balanced by applying the unbalance voltage to an electronic amplifier instead of a galvanometer. The output of this amplifier is then used to drive a small motor which resets the resistance arm until balance is achieved. These units are widely used to monitor temperature, pressure, and other critical factors in industrial processes.

Process Controllers are used not simply to monitor, but to control some factor in an industrial process. They will be treated in detail in Chapter 14.

Calculations for the potentiometer involve nothing more than Ohm's law and a basic application of Thévenin's theorem. An example will serve to illustrate this.

EXAMPLE 13-1

In the potentiometer of Fig. 13-2, the main dial is calibrated in 1-mV steps from 0 to 1200 mV, and can be read to the nearest $\frac{1}{4}$ mV. The galvanometer can detect currents as small as ± 0.2 μA. The voltage to be measured is a standard cell with $V_S = 1.019$ V and $R_S = 600\ \Omega$. Find the value of calibrating resistance R_1, the main dial resistance between wiper and ground, and the resolution of the measurement.

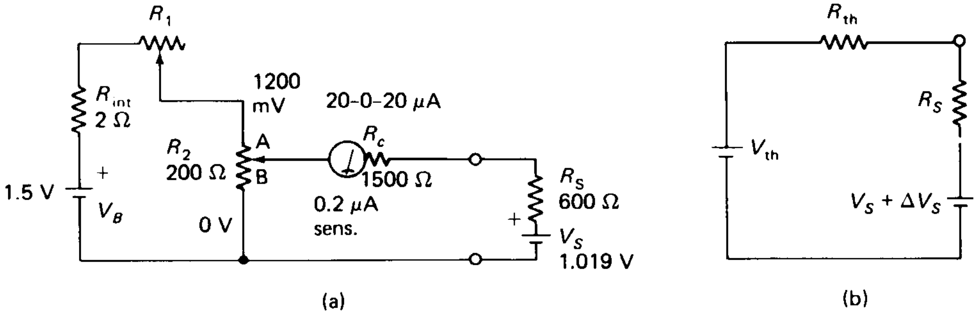


FIGURE 13-2 Example 13-1: analysis of a potentiometer circuit.

Solution

$R_1 + R_{int}$ must take the 0.3 V not taken by R_2;

$$\frac{V_{R2}}{R_2} = \frac{V_B - V_{R2}}{R_1 + R_{int}} \qquad \frac{1.2\ \text{V}}{200\ \Omega} = \frac{0.3\ \text{V}}{R_1 + R_{int}}$$

$$R_1 + R_{int} = 50\ \Omega \qquad \mathbf{R_1 = 48\ \Omega}$$

The setting of R_2 is also found by voltage division:

$$\frac{V_{R2}}{R_2} = \frac{V_{R2B}}{R_{2B}} \qquad \frac{1.2\ \text{V}}{200\ \Omega} = \frac{1.019\ \text{V}}{R_B}$$

$$R_{2B} = \mathbf{169.83\ \Omega}$$

To determine the resolution of the measurements, we will find the Thévenin equivalent of the bridge and then determine what change in V_S would unbalance the bridge by 0.2 μA. Figure 13-2(b) shows the equivalent.

$$V_{th} = 1.019\ \text{V}$$

$$R_{th} = [(R_{int} + R_1 + R_A) \| R_B] + R_C$$

$$= [(2 + 48 + 30.17) \| 169.83] + 1500$$

$$= \mathbf{1554\ \Omega}$$

V_{th} balances out V_S, and ΔV_S causes 0.2 μA in $R_{th} + R_S$.

$$\begin{aligned}\Delta V &= I(R_{th} + R_S)\\ &= (0.2\,\mu A)(1554 + 600)\\ &= \mathbf{0.431}\ \text{mV}\end{aligned}$$

The mechanical imprecision of 0.25 mV from the dial calibration must be added to this electrical imprecision:

$$\begin{aligned}\epsilon &= \Delta V_{\text{mech}} + \Delta V_{\text{elect}}\\ &= (0.250 + 0.431)\ \text{mV} = \pm\mathbf{0.681}\ \text{mV}\end{aligned}$$

13.2 THE WHEATSTONE BRIDGE

The Wheatstone bridge is widely used for the precision measurement of resistance, just as the potentiometer is used for the measurement of voltage. The Wheatstone bridge can also be used as the basis for automatic recording and control systems where the sensing device is resistive rather than voltaic.

A basic Wheatstone bridge is shown in Fig. 13-3(a). Notice that it consists of two voltage dividers: R_1 and R_2 on the left, R_3 and R_4 on the right. When the division ratio of the left divider equals that of the right divider, the bridge is balanced and the indicator reading is null. An easily remembered form of the balance equation is

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \tag{13-1}$$

Balance of the Wheatstone bridge is not affected by the battery voltage or the accuracy of the null meter, and if the resistances are chosen to have the same temperature coefficients, even the small changes in resistance due to temperature change will be balanced out.

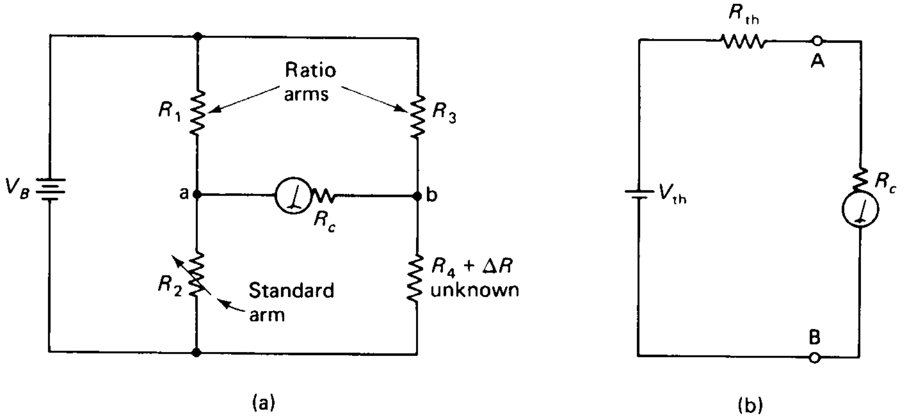


FIGURE 13-3 Basic Wheatstone bridge (a) and equivalent circuit (b) for analysis under slightly unbalanced conditions.

Slightly Unbalanced Bridge: If a Wheatstone bridge is only slightly unbalanced [ΔR no larger than a few percent of R_4 in Fig. 13-3(a)], we can obtain its Thévenin equivalent [Fig. 13-3(b)] and determine the unbalance current through the meter very quickly by use of the following approximations:

$$R_{th} = (R_1 \| R_2) + (R_3 \| R_4) \tag{13-2}$$

$$V_{th} = V_S \frac{R_3 \Delta R}{2R_3R_4 + R_3{}^2 + R_4{}^2} \tag{13-3}$$

If all four resistances are equal, as is often the case, these equations reduce to

$$R_{th} = R \tag{13-4}$$

$$V_{th} = V_S \frac{\Delta R}{4R} \tag{13-5}$$

The precision of a Wheatstone bridge measurement can be increased by increasing the battery voltage, lowering the values of the four resistors proportionately, lowering the galvanometer coil resistance, or increasing the galvanometer current sensitivity. The electrical precision of the bridge can be calculated with the aid of Ohm's law and Thévenin's theorem, but here we must "sneak in the back door" by assuming a small change in the unknown (say 1%) and calculating the current produced in the galvanometer by this upset. This current can then be proportioned against the galvanometer sensitivity to find the precision of the resistance measurement. The following example illustrates the technique.

EXAMPLE 13-2

In the circuit of Fig. 13-3(a) the following values apply at balance:

$V_B = 10$ V	$R_1 = 100\ \Omega$
$R_c = 600\ \Omega$	$R_2 = 40\ \Omega$
$I_c = 0.5\ \mu$A minimum for observable deflection	$R_3 = 500\ \Omega$

Find the value of R_4 and the precision of this determination.

Solution

The value of R_4 is determined quickly from the balance equation:

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \qquad \frac{100}{40} = \frac{500}{R_4} \tag{13-1}$$

$$R_4 = \mathbf{200\ \Omega}$$

The sensitivity will be determined by letting $R_4 = 202\ \Omega$ and finding the resulting unbalance current. The Thévenin equivalent must first be taken with the meter as the external load.

$$R_{th} = (R_1 \| R_2) + (R_3 \| R_4) \tag{13-2}$$
$$= 171\ \Omega$$

$$V_{th} = V_S \frac{R_3 \Delta R}{2R_3R_4 + R_3{}^2 + R_4{}^2} \tag{13-3}$$
$$= 10 \frac{500 \times 2}{(2 \times 500 \times 200) + 500^2 + 200^2}$$
$$= 0.02\ \text{V}$$
$$I_c = \frac{V_{th}}{R_{th} + R_c} = \frac{0.02}{171 + 600} = \mathbf{26\ \mu A}$$

We will now take the ratio of this current over the minimum observable current and set it equal to the ratio of given resistance change over minimum observable resistance change:

$$\frac{I}{I_{min}} = \frac{\Delta R}{\Delta R_{min}} \qquad \frac{26\ \mu\text{A}}{0.5\ \mu\text{A}} = \frac{2\ \Omega}{\Delta R_{min}}$$
$$\Delta R_{min} = \mathbf{0.038\ \Omega}$$

This is the electrical imprecision. Of course, the mechanical imprecision due to dial readability must also be considered.

In commercial bridges, R_2 (Fig. 13-3) is a high-precision potentiometer or a combination of switch-selected precision resistors and a precision pot in series. R_1 and R_3 are precision resistors which are switch-selected to determine the ratio of unknown R_4 to main pot R_2. For example if R_1 is 100 Ω and R_3 is 500 Ω, R_4 will have a value of (500/100) R_2, or 5 R_2.

13.3 VARIATIONS ON THE WHEATSTONE BRIDGE

Kelvin Bridge: If a conventional Wheatstone bridge is used to measure resistance much below 1 Ω, the resistance of the lead wiring within the instrument becomes appreciable compared to the unknown, and the accuracy of the instrument is impaired. The Kelvin bridge of Fig. 13-4 solves this problem and allows accurate measurement in the milliohm range.

Notice from the figure that the right-hand voltage divider R_3–R_4 contains extremely low resistance, and that the wiring resistance R_w between R_3 and R_4 would cause a 10% error if the galvanometer were simply connected at the bottom of R_3. Connecting the meter directly to the top of R_4 would not solve the problem, because R_w would then cause a 5% upset in R_3. The solution is to split the voltage drop across R_w according to the ratio of R_1 over R_2, and this is most conveniently

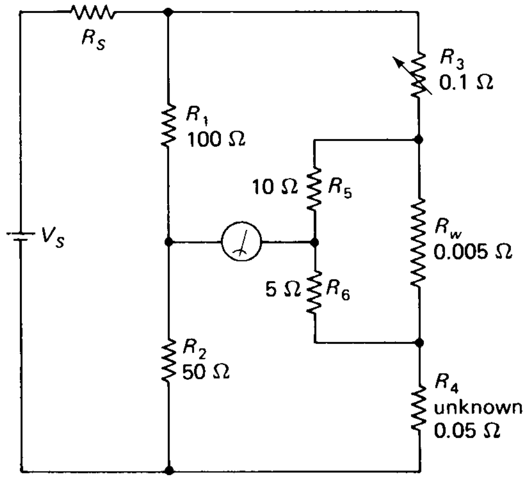


***FIGURE 13-4** The Kelvin bridge compensates for the circuit wiring resistance, allowing measurement of very low values of resistance.*

done with two added resistors R_5 and R_6, selected such that

$$\frac{R_5}{R_6} = \frac{R_1}{R_2} \tag{13-6}$$

If multiple ranges are desired in a Kelvin bridge, R_5 or R_6 must be switched also to satisfy equation 13-6. For example, switching R_1 to 1000 Ω and R_5 to 100 Ω would lower the resistance range of the meter by a factor of 10. Switches should *not* be placed in the low-resistance line containing R_3, R_w, and R_4, since switch contacts have unpredictable and in this case relatively large resistances.

Megohm Bridge: Resistances in the gigaohm range can be measured with the basic Wheatstone bridge circuit if the source voltage is increased to around 1000 V. The range can be extended to the teraohm (10^{12} Ω) range if an amplifier is used to increase the sensitivity of the null detector and if special precautions are taken against leakage currents.

Figure 13-5 shows two of these precautions. T_1 is a terminal with a metal guard band completely surrounding it. Leakage current which otherwise would flow from T_1 across the instrument case to T_2 is intercepted by the guard ring and routed to the other side of the battery. The leakage resistance thus shunts the battery instead of shunting the test resistance and does not affect the measurement.

T_3 is a guard terminal (not the same as T_1, which is a guarded terminal) which is used to intercept leakage currents that may flow around the body of the component under test. The guard terminal is connected to a metal plate on which the test specimen is mounted by stand-off insulators, or to a metal ring around the body of the test specimen if it is desired to eliminate surface leakage effects.

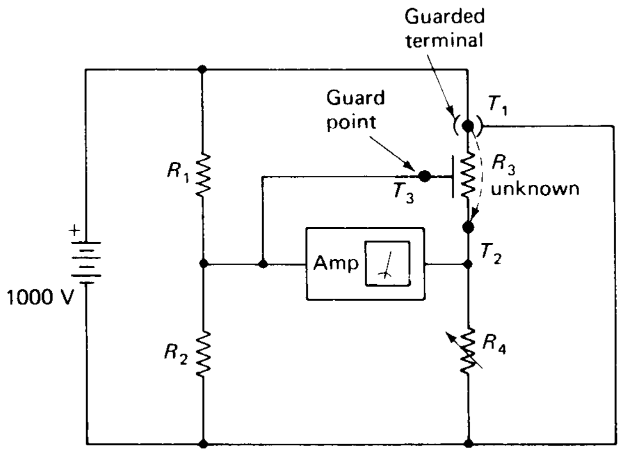


FIGURE 13-5 Very high values of resistance can be measured if special precautions are taken against leakage currents.

13.4 AC BRIDGES

AC bridges have the same four-arm configuration as the Wheatstone bridge, but each arm may contain reactive as well as resistive components. Balance requires that not only the amplitude but also the phase of the voltages at each side of the detector be identical. In general, manual balance of an ac bridge requires adjustment of two variable components, although bridges to measure pure reactance balance with one.

Capacitance-Comparison Bridges can have a wide range of measurable values using only one standard fixed capacitor and a variable resistance as the calibrated readout dial. Figure 13-6(a) shows the simplest such bridge with its balance equation. R_{var} and C_{std} are selected so that X_C is approximately equal to R_{var} at midrotation. This, of course, depends upon the frequency of V_s. Notice that the detector is an ac amplifier driving a rectifier and meter.

When the unknown capacitance contains an appreciable leakage resistance, it will be necessary to similarly shunt the standard capacitance to obtain a balance. Figure 13-6(b) shows a parallel capacitance-comparison bridge and its balance equations. Notice that lower values of shunting resistance indicate lower Q (higher dissipation factor D) in the measured capacitor.

Often the dissipation of the measured capacitor is so low that the required value of R_2 would exceed the available limits for precision variable resistors. In this case the series capacitance comparison bridge of Fig. 13-6(c) is used. Here the low dissipation factor calls for a low value of R_2, which is easily realized. The balance equations yield the series equivalent circuit $C_s + R_s$, whereas we may wish to work with the parallel equivalent. The series-to-parallel conversion equations of Section 4.2 can be used to rectify this problem, although it will be necessary to perform $X_s = 1/(2\pi f C_s)$ before and $C_p = 1/(2\pi f X_p)$ after the conversions.

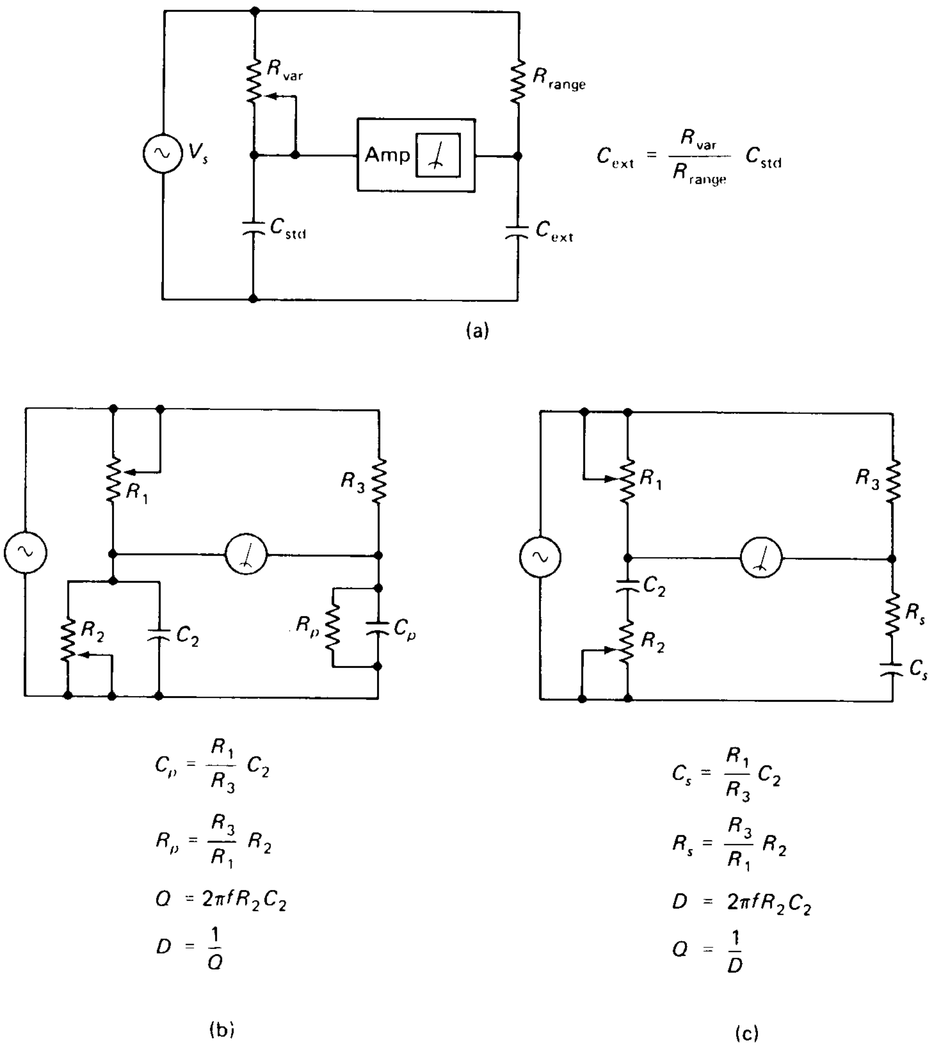
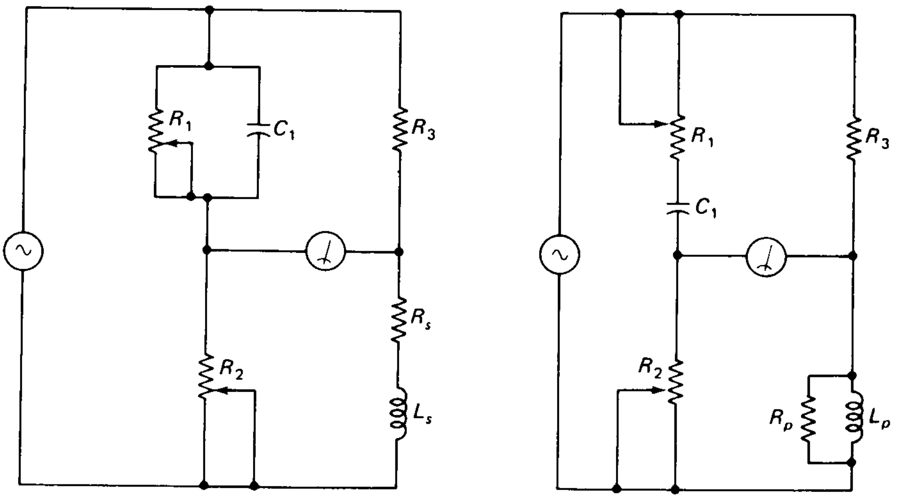


FIGURE 13-6 *Capacitance bridges: (a) simple comparison; (b) parallel resistance for high-D (dissipation) measurements; (c) series resistance for low-D (high-Q) measurements.*

Inductance-Comparison Bridges can be built in a manner similar to the capacitance-comparison bridges, but they are of limited practical value because real standard inductors are expensive, relatively temperature unstable, and plagued by stray effects such as skin-effect resistance, interwinding capacitance, and core hysteresis. Inductance measurement can be handled more effectively by using an "upside-down" C–R voltage divider in the left-hand leg of the bridge, as shown in Fig. 13-7.



$$L_s = R_2 R_3 C_1$$

$$R_s = \frac{R_2 R_3}{R_1}$$

$$Q = 2\pi f R_1 C_1$$

(a)

$$L_p = R_2 R_3 C_1$$

$$R_p = \frac{R_2 R_3}{R_1}$$

$$D = 2\pi f R_1 C_1$$

(b)

FIGURE 13-7 *Inductance bridges: (a) Maxwell bridge, primarily for low-Q inductance measurement; (b) Hay bridge, primarily for high-Q inductors.*

The Maxwell Bridge, Fig. 13-7(a), is suited for measuring relatively low-Q coils. C_1 and R_2 provide a leading phase shift at the left side of the null indicator to match the leading phase at the right side produced by R_3 and L_s. R_1 introduces loss in the standard capacitor to match the loss in the measured inductor. Lower values of R_1 indicate a lower Q of the measured coil. Very-high-Q coils require excessively high values for R_1, making a series R_1C_1 combination desirable. Balance equations for L_s and R_s are given with the figure.

The Hay Bridge, Fig. 13-7(b), is more suited for measuring high-Q coils, since higher Q requires lower values of R_1. The balance equations given with the figure are for a parallel equivalent to the unknown inductance, L_p and R_p. This is inconvenient, since we seldom conceive of a real inductor as being a perfect inductor shunted by a leakage resistance. Also note that the L_p value obtained with the Hay bridge will not exactly equal the L_s value obtained with the Maxwell bridge for the same coil. For these reasons a more complex but perhaps more useful set of formulas for the series-equivalent inductance from the Hay-bridge components is given:

$$L_s = \frac{R_2 R_3 C_1}{1 + 2\pi f R_1{}^2 C_1{}^2} \tag{13-7}$$

$$R_s = \frac{2\pi f R_1 R_2 R_3 C_1{}^2}{1 + 2\pi f R_1{}^2 C_1{}^2} \tag{13-8}$$

Of course, the parallel-to-series-conversion formulas of Section 4.2 could be used to achieve the same results.

Biased-Component Measurement: It is often desirable to measure the inductance of a coil as a function of the dc current it carries to identify core saturation. Similarly, we may be interested in the variations in capacitance and dissipation factor of an electrolytic capacitor as the dc voltage across it is changed. Commercial impedance bridges often have special bias-current and bias-voltage terminals for these applications, but at low frequencies the connections of Fig. 13-8 can be used.

To voltage bias a capacitor, a lab supply can usually be placed in series with the capacitor under test, observing polarity, as in Fig. 13-8(a). Be sure to keep the supply at the LOW side of the bridge terminals, to avoid putting its considerable stray capacitance to ground in parallel with the test capacitor. If the bridge does not provide an internal dc path from the HIGH to LOW terminal (easily checked with an ohmmeter), an external path can be provided as shown. If the connection point shown (arrow) is not accessible, the HIGH terminal can be used, but the dissipation factor will read erroneously high.

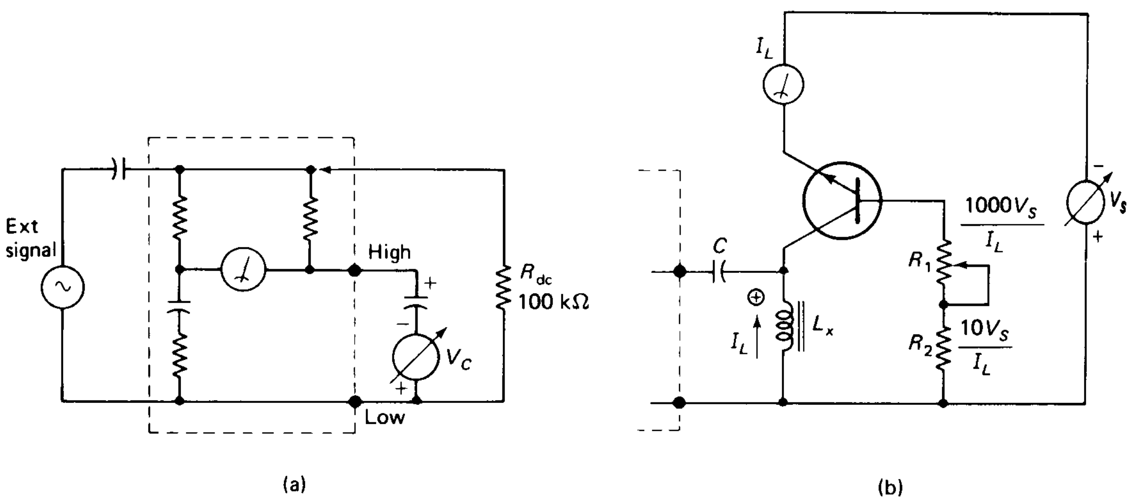


FIGURE 13-8 (a) Measuring capacitance under an externally applied voltage V_C. (b) Measuring inductance under an externally applied current.

To bias an inductor, the circuit of Fig. 13-8(b) is recommended. The transistor acts as a current source, feeding current to L_x without placing appreciable shunt impedance across it. V_S should be a current-limited supply, or if not, a fast fuse should be placed in its negative lead to protect the inductor from accidental overload. V_S should be adjusted so that V_{CE} is only a volt or two above saturation to minimize power in the transistor. Blocking-capacitor C prevents dc from getting back into the bridge, and is not usually necessary. If it is used, it must have a reactance 100 times less than X_L at the test frequency.

Wagner Ground: Stray capacitance on the order of 100 pF can easily accumulate from the amplifier input and from the potentiometers, standard capacitors, switches, and wiring in a wide-range general-purpose impedance bridge. The bridge would be useless at frequencies above a few kilohertz if steps were not taken to compensate for these capacitances. The Wagner ground adjustment, R_w of Fig. 13-9, attempts to place the circuit ground at the potential of the amplifier input by means of a voltage divider R_wC_w. The voltage across the amplifier input capacitances is then zero and no current flows through them. If R_2 is very low (low-D capacitor being measured), this ideal can be approached closely with only a single adjustment, R_w. In operation, switch S_1 is thrown to GROUND and R_w is adjusted for minimum meter reading. The switch is then returned to MEASURE and R_1 and R_2 are adjusted for null.

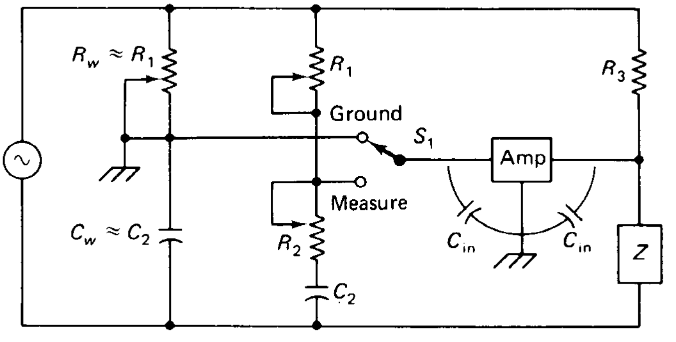


***FIGURE 13-9** Wagner ground connection for eliminating the effects of meter-amplifier input capacitance.*

14

TRANSDUCERS AND CONTROL SYSTEMS

14.1 DEFINITION, CLASSIFICATION, AND SPECIFICATIONS

A transducer is a device that changes energy from one form to another. In electronic instrumentation we are most often interested in changing some physical quantity such as heat, light, or motion into electrical energy for measurement, recording, or processing.

Transducers may be *active*, in which case they will generate an electrical signal without the application of external voltage or current, or they may be *passive*, in which case they will require excitation by an external voltage or current.

Transducers may produce an ac or a dc output signal, and often in the case of passive types, this depends on whether they are excited by ac or dc.

The simplest transducers to use are the voltaic and resistive types, which produce a voltage change or a resistance change in response to a change in some physical property. More difficult in application are transducers that produce a change in capacitance or inductance in response to physical properties.

For any transducer, there will be three primary concerns about its suitability of operation.

1. **Sensitivity:** This is expressed wherever possible in terms of output voltage change per unit input-property change. In measuring temperature, it would be μV/°C, or for displacement, V/cm. For some transducers it may have to be expressed in terms of resistance, inductance, or capacitance change per unit physical property change (i.e., 2.3 pF per lb/in.^{2} change in pressure for a nominal 120-pF unit).

2. **Linearity, Offset, and Dynamic Range:** Ideally, it would be nice to have a transducer that generated zero voltage at absolute zero temperature, and 1 mV for every °C above that, up to and beyond 1000 mV at 1000°C. Few transducers even approach this ideal. Offset (nonzero output for zero input) can be handled easily by the techniques of Chapter 19. Dynamic range generally depends upon material selection and physical structure. Nonlinearity of the input–output curve is a persistent problem, and quite a game can be made of trying to linearize an inherently nonlinear transducer. Solutions may involve anything from simply compressing the meter scale at one end to complex microprocessor programs providing four-digit readout.

3. **Noise Immunity:** Noise is defined as anything other than the desired signal. If a pressure transducer is affected by temperature, or if a light-intensity transducer is affected by nuclear radiation, you have a noise problem.

14.2 TRANSDUCER TYPES

Literally hundreds of transducer types exist, and the technician must become accustomed to an occasional surprise at the clever way someone has devised to sense some physical property. A relatively few types serve as the workhorses of the industry, however, and we will concentrate on descriptions of these.

No attempt is made to classify the devices by function, as this tends to limit rather than expand their range of application. A microphone is not listed as an audio transducer, because with suitable circuitry it can be used to measure the time between two impacts, thus measuring distance, velocity, or density.

Switches are the most common transducers because they are inexpensive and easy to understand. They are commonly applied as:

- Limit switches, to indicate end of travel on elevators, garage-door openers, machine-tool bits, and so on. Two switches may be placed in line, the first indicating slow down, and the second halt. Magnets may be used to actuate the switches if physical contact is undesirable.
- Liquid-level indicators when attached to a float.
- Pressure indicators when attached to a bellows or piston-against-spring assembly, as in Fig. 14-1(a).
- Temperature indicators when attached to a bimetallic strip, as in Fig. 14-1(b).
- Rotational-speed indicators when activated by centrifugal force.

Variable Resistors are often geared to a linear moving element or simply structured as a linear slide wire to provide continuous position readout. They are also formed as a nest of loose-packed carbon granules so that resistance decreases when

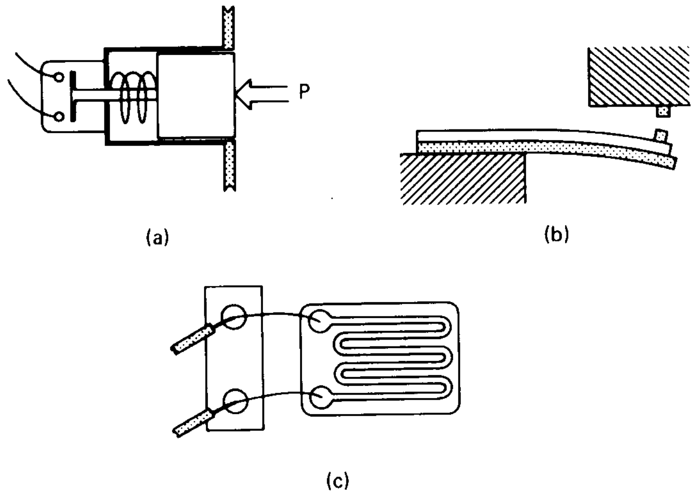


FIGURE 14-1 (a) Pressure-sensitive switch. (b) Switch operated by a temperature-sensitive bimetal strip. (c) Strain gage with terminal mounting block.

pressure on a plug packs them closer together. This is the basis of the familiar telephone microphone.

Strain Gages are most commonly made by etching thin conductive paths of metal on a flexible insulative backing. They are then bonded to a structural element (usually metal) that is placed under mechanical stress. The very slight elongation of the metal stretches the conductors, increasing the resistance of the gage by something on the order of 0.01% for typical stresses in steel. If a strain gage of unknown properties is connected as one arm of a Wheatstone bridge, the variable dial can be calibrated in terms of stress (load in kg/cm^2) for the metal being tested. Figure 14-1(c) shows a strain gage with a terminal block bonded in place to keep the relatively heavy instrument wires from tearing up the delicate strain gage. Strain gages made from semiconductor materials are available with resistance changes on the order 100 times that attainable with metal gages. Temperature sensitivity and nonlinearity have limited the popularity of these gages.

Thermocouples are composed of a simple junction between two dissimilar metals, usually a welded lead enclosed in an inert protective tube. A small voltage which increases with temperature appears across the junction. Typical thermocouples operate up to 3000°F (1650°C) and generate a few tens of millivolts. Iron-constantan thermocouples, for example, generate about 32 mV at 1000°F (538°C).

The output of a thermocouple can be measured directly with a sensitive galvanometer, but voltage drop across the relatively high resistance of the thermocouple wire limits the accuracy of this technique. Some form of self-balancing potentiometer with electronically amplified null detection is generally used. Figure 14-2 shows a thermocouple–potentiometer circuit, and points up the fact that the

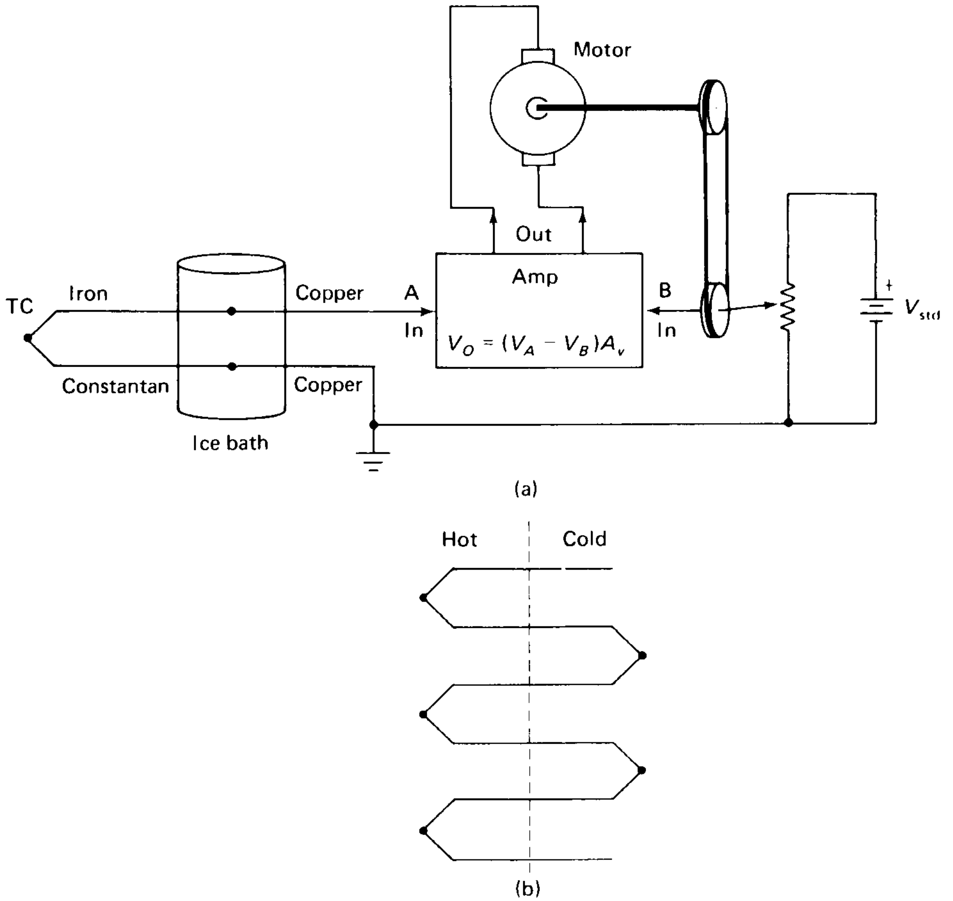


FIGURE 14-2 *(a) Thermocouple feeding a self-balancing potentiometer. (b) Thermopile.*

junctions of the thermocouple wires to the measuring instrument form two additional thermocouple junctions whose voltage must be taken into account. These voltages will be relatively small but will vary with ambient temperature changes. Where highest accuracy is required and some inconvenience can be tolerated, a reference junction held at 0°C by a bath of melting ice can be employed.

Undesired thermoelectric voltage generated across dissimilar metals at sockets or connectors is often the cause of thermal instability in sensitive dc amplifiers. A series of thermocouples, called a thermopile, can be used to obtain higher output voltages, as shown in Fig. 14-2(b). Notice that each hot junction necessitates a cold junction and two wires leading to it. Thermopiles using semiconductor rather than metal elements have been used with some success as direct heat-to-electric power converters.

The Peltier cooling effect should also be mentioned at this point. If a relatively high current is passed through the junctions in Fig. 14-2(b), one set of junctions will be heated and the other will be cooled. This effect has been used to provide spot cooling of critical elements in densely packed equipment without the bulk and moving parts of conventional refrigeration techniques.

Thermistors are resistive elements whose resistance decreases with increasing temperature, in some cases by as much as $-7\%/°C$. Thermistors are available with 25°C resistances from less than 1 Ω to several MΩ and with operating ranges from − 80 to + 1000°C. Physical shapes and sizes vary widely, from beads of 1-mm to discs of 25-mm diameter.

Thermistor curves are inherently nonlinear, as shown in Fig. 14-3, but special three-terminal networks (composed of two thermistors and two resistors in a single package) are available which provide a nearly linear voltage-versus-temperature output over moderate (−30 to +100°C) temperature ranges.

Thermistors are most commonly employed in a Wheatstone bridge or voltage-divider circuit for temperature measurement and control, in which case the IV product must be held to a minimum to prevent self-heating. Deliberately raising the IV product to produce self-heating allows the thermistor to be used to measure other parameters that act to carry away heat. Examples are wind or air-flow velocity, level of immersion in a liquid, viscosity of a liquid, and R value of a thermal insulator.

Thermistors are also used for temperature compensation in electronic circuits. As an example, the thermistor in Fig. 14-3(c) lowers the transistor's bias point at high ambient temperatures, limiting its collector dissipation to a safe value.

Finally, thermistors can be used to produce time delays from less than 1 s to several tens of seconds by operating them in the self-heat mode and taking advantage of their thermal mass, which delays warmup and consequent resistance drop.

Photodetectors are available in a rather bewildering variety. Primary factors for comparison are sensitivity and response speed. Other factors are temperature sensitivity, bulk and fragility, cost, and spectral response peak (some devices respond better to red or infrared or blue or ultraviolet). We will list the five most common photodetectors in order of increasing response time.

- Photomultipliers (Fig. 14-4) are high-vacuum glass tubes requiring a thousand-volt-level power supply. However, they have a response time on the order of 1 ns and a sensitivity to low light levels which is several times better than the best solid-state detectors. They contain a cathode which emits electrons when struck by light, and a series of plates at successively higher positive potentials, to which the electrons are attracted in turn. At each plate the impacting electrons splash up additional electrons, so the total electron current builds up.

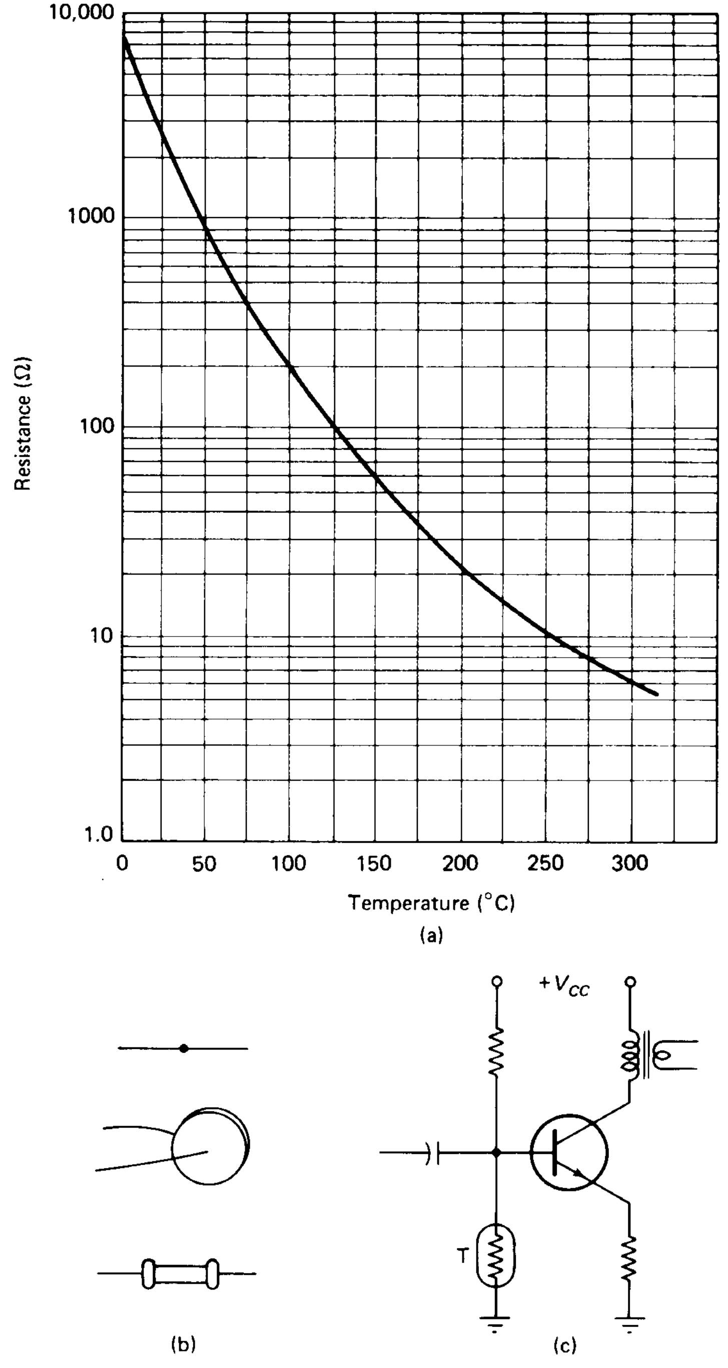


***FIGURE 14-3** Thermistors are temperature-sensitive resistors: (a) typical* R *versus* T *curve; (b) typical thermistor packages; (c) thermistor used to bias-stabilize a power-transistor amplifier.*

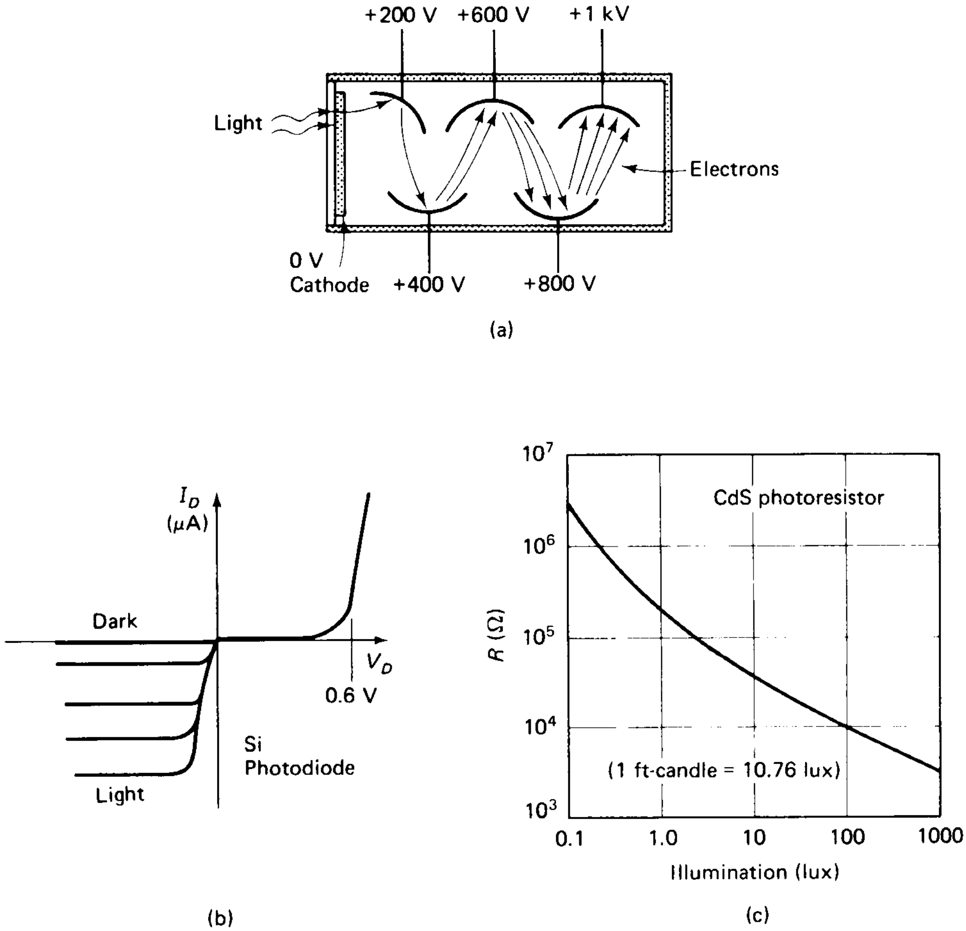


***FIGURE 14-4** (a) Photomultiplier-tube structure. (b) Silicon-photodiode characteristics. (c) Typical CdS photo-resistive cell characteristics.*

- *P-N* and *PIN* silicon photodiodes have response times on the order of 20 ns and 2 ns, respectively. Under forward bias they behave much like conventional diodes, but their reverse leakage current increases with light intensity due to formation of hole–electron pairs near the junction. Currents of 10 μA are typical under room light. The *PIN* photodiode has a layer of intrinsic (pure) silicon between the *P*- and *N*-doped regions to reduce dark-current leakage and junction capacitance.

- Phototransistors, in which base current is replaced by light-generated hole–electron pairs, have response times on the order of 2 μs and collector currents of a few mA at room-light levels. The base lead is usually brought out so the transistor can also be turned on in the usual manner. Light-activated SCRs (LASCR) have similar response times. Photo FETs are somewhat faster, with response times in the vicinity of 0.1 μs.

- Silicon solar cells are large-area *P-N* junctions which produce direct current when struck by light. Conversion efficiency in sunlight is about 10%. A typical cell produces 0.4 V output and provides about 5 mW/cm^2 under bright sunlight. Response times are on the order of 5 μs.
- Resistive photocells of cadmium sulfide (CdS) and cadmium selenide (CdSe) are very popular because of their low cost and relatively high voltage and current capabilities (250 V and 300 mW typical). However, they are extremely slow, exhibiting response times on the order of 1 s at low light levels, down to about 10 ms at bright room-light levels. A wide variety of types with *on* resistances down to the 100-Ω range are available.

Piezoelectric Crystals such as quartz and barium titanate produce a voltage when placed between two electrodes and subjected to mechanical vibration. The voltage is proportional to rate of change of force, so only dynamic forces are recorded. The piezoelectric effect is the basis of the common crystal and ceramic microphones and phonograph pickups and is fairly widely employed in the measurement of vibration and shock. Output voltages from 1 to 100 mV, depending on sound level, and source impedances of approximately 50 kΩ are common for crystal microphones.

Magnetic Transducers take a variety of forms, some of which are shown in Fig. 14-5. The common feature is that the air gap of a coil's magnetic core is varied by mechanical displacement, resulting in a change in inductance of the coil. In some

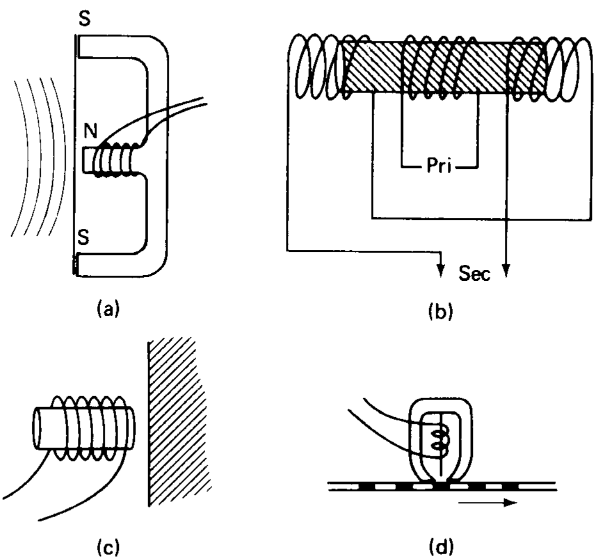


FIGURE 14-5 Magnetic transducers: (a) dynamic microphone and headphone principle; (b) linear variable differential transformer; (c) proximity detector; (d) magnetic-tape pickup.

cases, such as the magnetic headphone of Fig. 14-5(a), the coil is self-excited by the field from a permanent magnet, so that a voltage is produced when air vibration on the diaphragm changes the air gap. In other cases the coil is excited externally. The two secondaries of the linear variable differential transformer (LVDT), Fig. 14-5(b), are connected in series reverse phase, so that, with the movable core centered, the output voltage is zero. Moving the core produces an output in or out of phase with the primary depending on direction of motion. The proximity detector of Fig. 14-5(c) is used in metal detectors and vibration monitors. It is used as the inductor in an LC oscillator at a relatively high frequency (above 100 kHz) so that an inductance change of less than 0.01% will produce a measurable change in oscillator frequency. Figure 14-5(d) gives a brief sketch of a magnetic-tape-player pickup head. Variations in reluctance of the magnetic coating on the tape produce variations in the inductance of the pickup coil as the tape moves past the head gap.

Generator Transducers: Small rotating generators with a permanent-magnetic field are often used as transducers. They produce an output voltage (generally dc) which is directly proportional to rotational speed. Driven by small turbines or rotating vanes, they can be used to measure wind or fluid-flow velocity.

Radiation Transducers: The high-energy radiation associated with nuclear activity is capable of ionizing (stripping away an electron from) the atoms it impinges upon. The Geiger–Müller radiation counter (Fig. 14-6) contains a low-pressure gas and an anode at a very high positive potential. Entering radiation ionizes a single particle of the gas, but the resulting two ions quickly collide with and ionize other gas atoms as they are accelerated by the high potential between the anode and cathode. Thus each entering particle of radiation causes complete ionization and conduction within the tube, regardless of its relative intensity. The ionized condition lasts about 1 ms, so the count rate of the tube is limited to about 1000 per second.

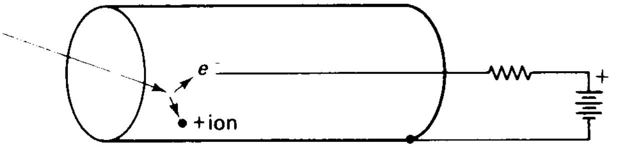


FIGURE 14-6 Geiger–Müller tube structure.

Capacitive Transducers: Changes in capacitance can be produced by varying the plate spacing, dielectric constant, or plate area. Readout can be accomplished by measuring the change in ac current as reactance changes, by sensing the change in voltage across a charged capacitor as its capacitance changes ($V = Q/C$), or by measuring the frequency of an oscillator controlled by the capacitor transducer. Figure 14-7 shows some of the possible applications: (a) a paint-thickness gage, (b) a capacitor microphone, and (c) a dielectric-constant sensor to detect which of two liquids is flowing in a pipe.

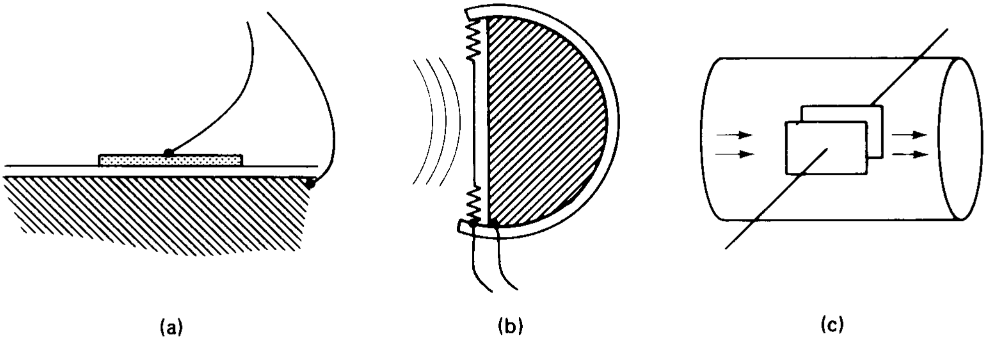


FIGURE 14-7 Capacitive transducers: (a) paint-thickness gage; (b) capacitor microphone; (c) liquid-dielectric sensor.

14.3 FEEDBACK CONTROL SYSTEMS

The usual automobile heater is an *open-loop* system: we turn it on, pick a setting, and hope for comfort. If our choice is wrong, we readjust the setting. When we come off the expressway into the city, heating requirements change and again we adjust the heater manually. By contrast, the usual home heating system is a feedback or *closed-loop* system. We choose a setting on a calibrated dial, and the thermostat leaves the furnace on until that *set point* is reached. Variations in heating demand due to opened doors or outside temperature drop are handled automatically because the *sensing device* (thermometer) constantly measures the *process variable* (room temperature) and compares it to the *demand point* (thermostat setting). If the measured variable is below the demand point, the *controller* (thermostat switch) turns on the *actuator* (gas or oil valve) to restore the process variable to the demand point. Figure 14-8 illustrates this basic feedback control loop for an industrial heat-treating furnace. If the sensing device is a simple switch actuator (such as the bimetal strip used in most thermostats) there need be no electronics involved, but more advanced control systems use an electronic sensing device and a potentiometer or Wheatstone bridge to produce an error or difference signal between the set point and the process variable. An amplifier is then used to allow the difference signal to control the actuator.

14.4 PROBLEMS WITH FEEDBACK CONTROL

Dead Zone or Hysteresis: In general, two-position actuators such as switches and solenoids should be snapped from one position to the other and not "mushed" or "jittered" back and forth. This requires a certain small dead zone between the on and off switching points to screen out noise and irregularities in the transducer output, as shown in Fig. 14-9. Many actuators have hysteresis built in. For example, a relay that pulls in at 100 mA may not drop out until 75 mA, providing desired snap action automatically. In many cases the dead zone is provided by an

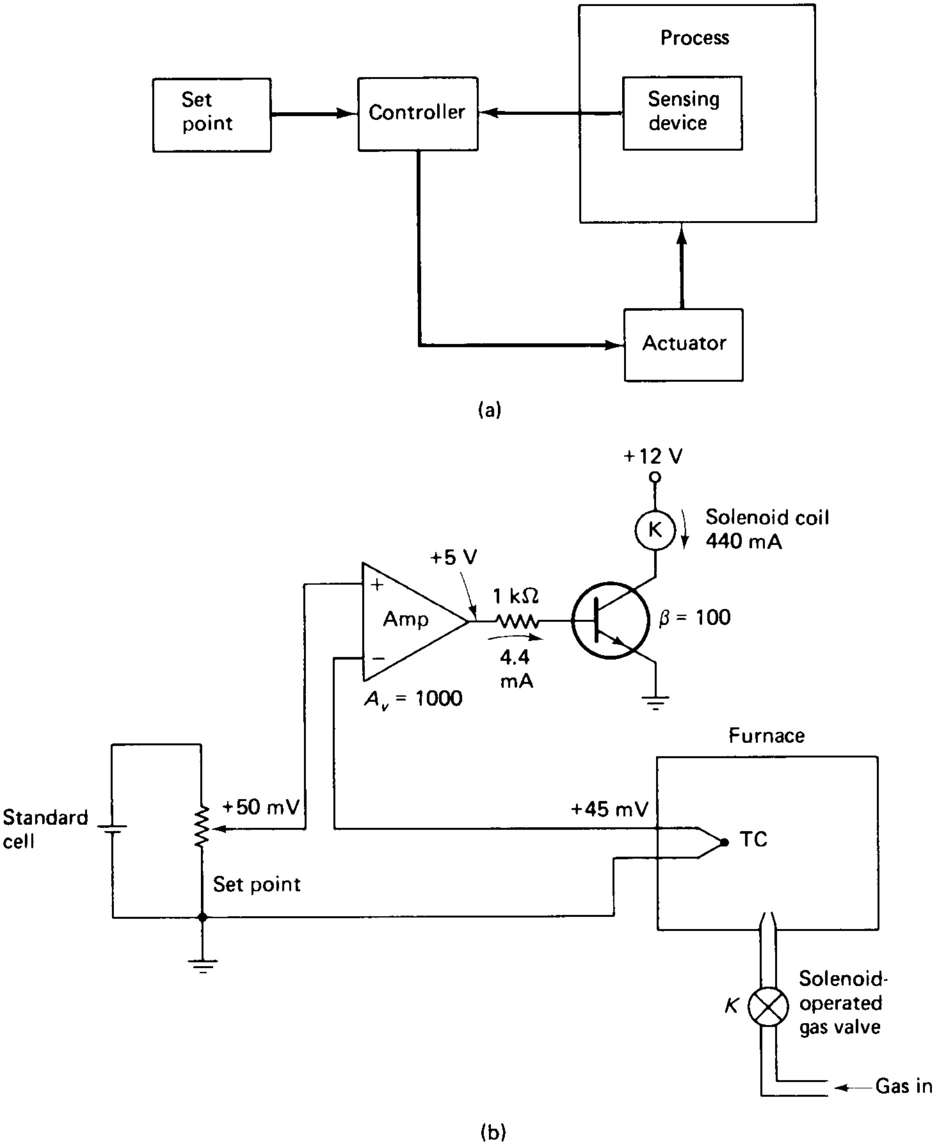


FIGURE 14-8 *(a) Basic feedback or closed-loop control of a process. (b) Implementation of feedback control for a gas-fired furnace with a thermocouple for a sensor, battery and pot for set-point determination, amplifier and transistor coil driver for controller, and solenoid gas valve for actuator.*

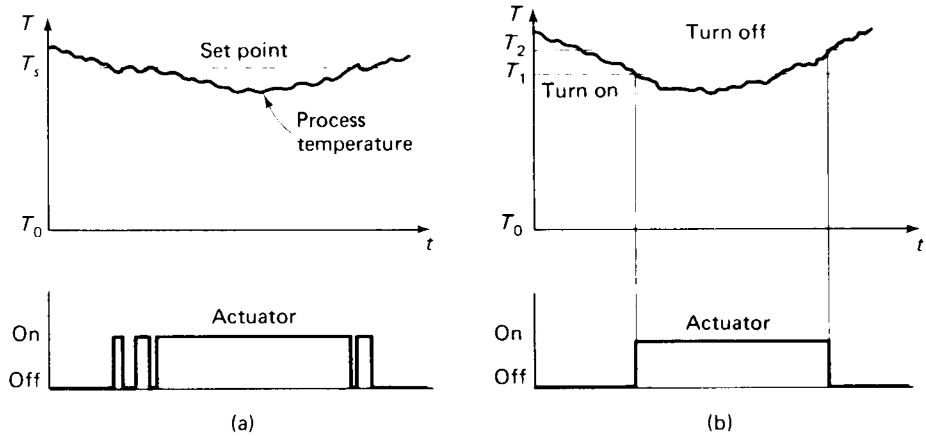


FIGURE 14-9 *(a) Random fluctuation of the sensor output can cause chatter of the actuator near the set point. (b) Introduction of a hysteresis or dead zone between turn-on and turn-off eliminates chatter.*

electronic snap-action switch called a Schmitt trigger. The dead-zone requirement places a limitation on the precision attainable in the control of the process.

Response Time: The time a control system requires to respond to a change in set point or a change in the process requirement is generally of considerable importance. If we change the demand temperature of our furnace from 200 to 250°C, we would like the new temperature to be reached as quickly as possible. Similarly, if we cause a temperature drop by opening the door and inserting an ingot of cold metal, we would like the temperature restored to its set point as soon as possible. Response time depends primarily upon the ratio of process inertia (in this example, the thermal mass of the furnace and its contents) to the strength of the actuating force (in this case, the Btu/h rating of the burner). Figure 14-10 shows the response times of two furnaces with different-size burners.

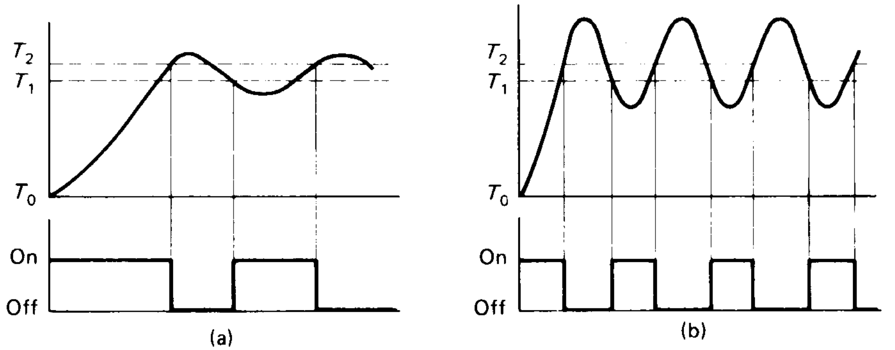


FIGURE 14-10 *Improving response time generally aggravates the tendency to overshoot and oscillate: (a) slow response, moderate overshoot; (b) fast response, excessive overshoot.*

Oscillation and Overshoot: The temperature of our heat-treating furnace is always oscillating between the turn-on and turn-off limits, which define the dead zone of the controller. In addition, there will be some overshoot and undershoot beyond these limits which will increase as response time is decreased. A well-insulated furnace with a small heating element will experience temperatures that rise and fall slowly, and there will be little overshoot, as shown in Fig. 14-10(a). However, if a larger heating element is used to reduce response time, as in Fig. 14-10(b), the temperature within the furnace is likely to rise considerably above the set point as heat spreads from the area of the burner to the area of the sensing device, even after the burner has been shut off. Attempts to increase response time in general have a tendency to increase overshoot and oscillation. Overshoot and undershoot can easily span a range much greater than that of the dead zone itself. If this is the case, attempts to increase the precision of control by narrowing the dead zone will be fruitless.

14.5 ADVANCED CONTROL TECHNIQUES

Overshoot Compensation: A clever and widely used technique for practically eliminating overshoot is illustrated in Fig. 14-11. The basic idea is to reduce the set point slightly whenever the heater comes on. The heater is thus shut off slightly before the process reaches the original turn-off point, and overshoot carries it to precisely the desired point. From then on, it is the set point that oscillates around an almost fixed process temperature. For maximum stability the depression of the set point must exactly equal the overshoot of the process, so the controller compensation must be adjusted to match the system. Changing loads (100 kg versus 1000 kg of ingots in the furnace) and changing activating forces (gas-pressure changes) will require a reset of the compensation control if oscillation is to be minimized.

A common way to achieve set-point reduction is to use a thermal-sensitive resistor as part of the potentiometer or Wheatstone bridge circuit. This resistor is heated by a separate resistive element which is controlled along with the actuator. A variable resistance in series with the heater controls the degree of set-point reduction. The response time of the heater–thermistor pair should approach that of the process itself if cancellation of process oscillation is to be achieved. Thermal mass may be added to the thermistor to lengthen its response time.

Three-Position Control: The controllers we have discussed thus far have been two-position controllers, the two positions being ON and OFF. Some control systems, however, obviously require three positions. In the case of an elevator they would be UP, OFF, and DOWN. For a hydraulic cylinder they would be EXTEND, HOLD, and RETRACT. Even inherently dynamic processes like the heat-treating furnace can benefit from three-position control. Imagine a large quantity of 500°C metal being thrust into a furnace set for 300°C. It may be desirable to employ a fan or cooling-water pump to bring the temperature more quickly back to the set point.

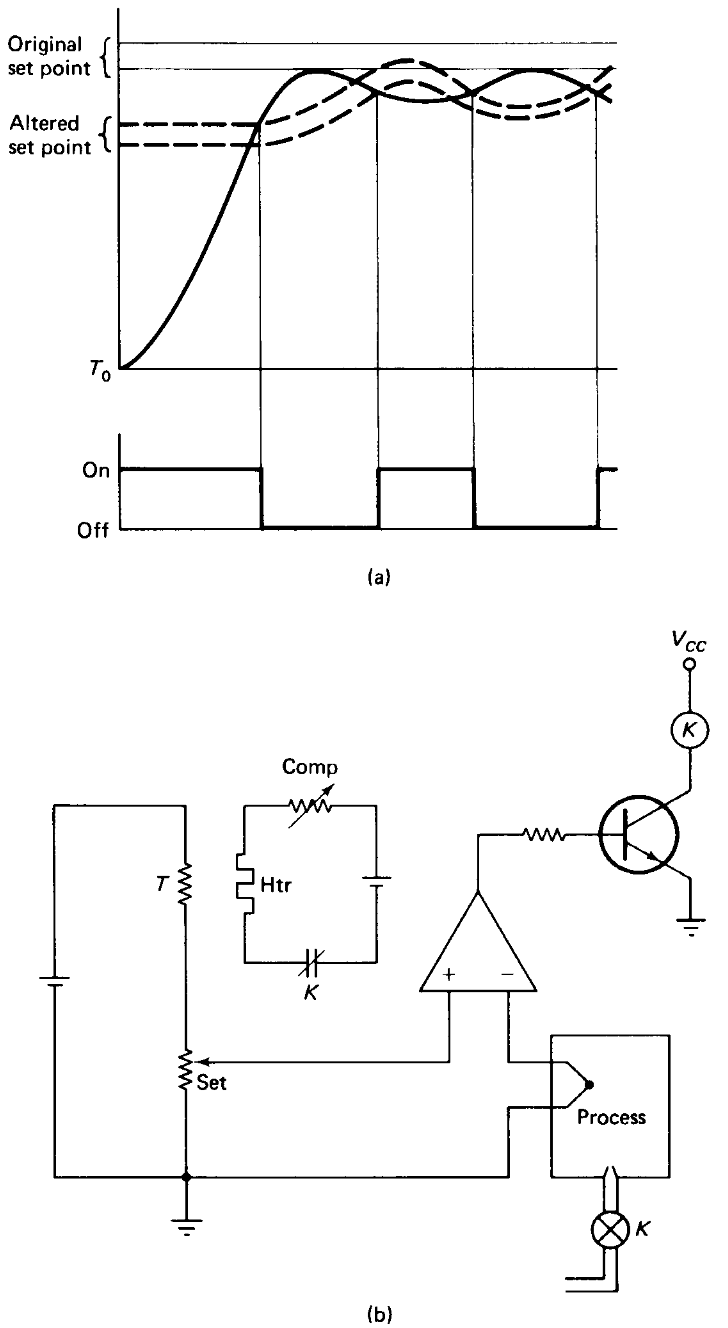
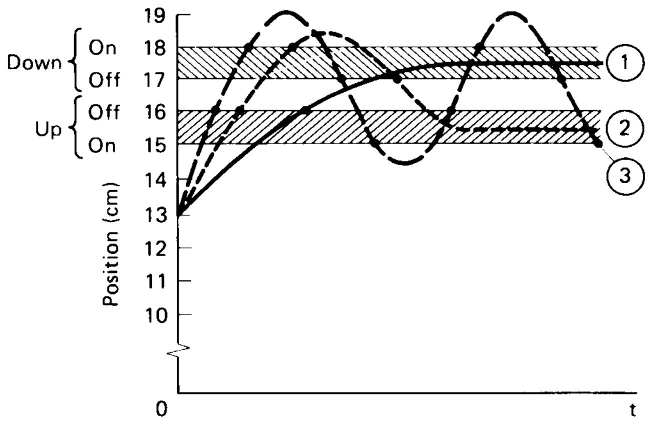
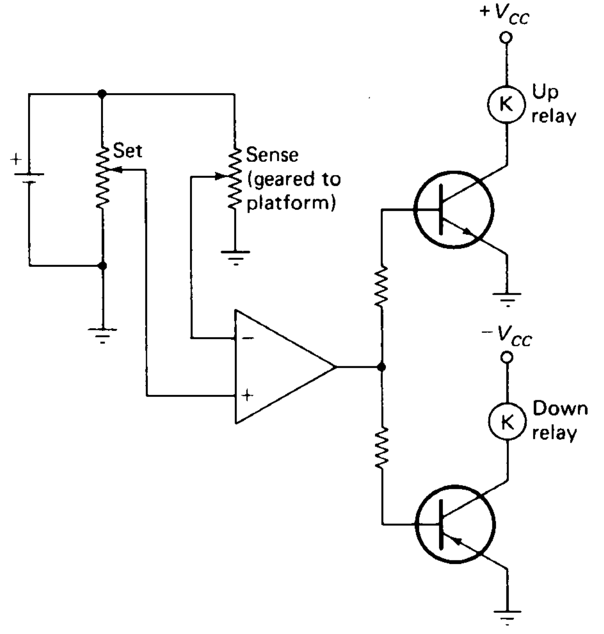


***FIGURE 14-11** Overshoot can be minimized by bringing the set point down to meet the process coming up: (a) graphical analysis; (b) circuit implementation.*



(a)



(b)

***FIGURE 14-12** Three-position control of a platform: (a) curve 1 shows minimal overshoot, curve 2 shows overshoot severe enough to activate the down relay to return the platform to the set zone, and curve 3 shows severe overshoot and undershoot resulting in continuous oscillation; (b) circuit implementation of three-position control.*

Oscillation of a violent sort is a common problem in three-position control systems. Figure 14-12 illustrates the situation using the example of a motor-operated vertical-positioning platform. The sensing device is a slide wire which forms one arm of a Wheatstone bridge. The platform is commanded to move up from its 13-cm initial position to a nominal 16.5-cm position by a rotation of the SET POINT potentiometer. The *up* relay turns on and the platform moves up. At 16 cm the *up* relay turns off, but because of its inertia, the platform does not come to rest until 17.5 cm (curve 1).

Now if the inertia of the platform were a little greater, or the speed of travel a little higher, the platform might overshoot to 18.5 cm, turning on the *down* relay, resulting in a reversal of the motor and a final rest point of perhaps 15.5 cm (curve 2). Notice that any rest point between the *down on* and *up on* limits is possible, depending upon the motor speed, platform load, and direction of approach. In the example, this is a span of 3 cm, which is likely to be objectionable.

The obvious solution is to increase the gain of the amplifier. A factor-of-10 gain increase would reduce the span of possible rest positions to 0.3 cm. However, this would greatly aggravate the overshoot problem, and it is likely that both overshoot and undershoot would exceed the *down on* and *up on* limits, resulting in continual reversals of the motor and oscillation of the platform up and down over a range far exceeding the 0.3-cm neutral zone, as illustrated in curve 3. These oscillations can be eliminated by slowing the system response time down (by a factor of 10), thus reducing inertial overshoot to a point where it remains within the 0.3-cm neutral zone, but such a slow response time may itself be objectionable.

The remainder of this chapter is devoted to resolving this dilemma: How can we build a control system with both minimum response time and minimum neutral zone, while avoiding the oscillation or "hunting" of the process variable which the attempt seems always to produce?

Proportional Control: If the amount of actuator force is continuously variable and is made proportional to the difference between the set point and the sensed process point, a system is produced which brings the output quickly toward the set point when the deviation is large, but slows down the approach as the set point is neared, as illustrated in Fig. 14-13. The actuator cannot be a relay or a solenoid in proportional control; it must be a continuously variable device. Most commonly, the actuator is an SCR or triac circuit, or a power transistor.

Proportional control often results in an objectionable *offset* between set point and process rest point. Offset is a function of process load and actuator force. In addition, mechanical systems have a hysteresis or dead zone caused by friction. The deviation of the process from the set point can be reduced by increasing amplifier gain, but with the same risk as before: overshoot and oscillation.

The term *proportional band* has been coined to express the gain of a proportional control system. It is defined as the change in transducer output as a percent of full scale required to produce a 100% change in actuator output. In effect, it represents the deviation from the set point at which the output force begins to

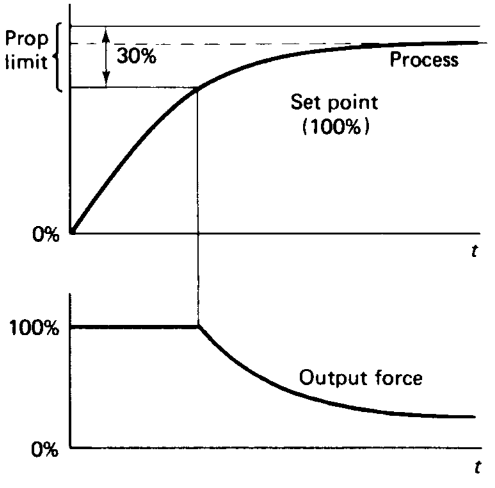


FIGURE 14-13 *In proportional control the restoring force is proportional to the deviation from the set point within the proportional band.*

drop. Notice from Fig. 14-13 that the output force remains at its maximum until the process reaches 70% of its set point, whereupon it begins to decrease. If the set point is at full scale the proportional band is 30%. Smaller proportional bands reduce offset and hysteresis error, but tend to cause overshoot and oscillation.

Reset or Integrating Control is added to proportional control to reduce the offset and hysteresis errors. In this system the output force is proportional to the deviation from the set point plus the integral (accumulation over time) of that deviation. Thus a small error persisting over a long time will eventually build up enough output to move the process to a new rest point. Certain elevators are commonly seen to stop with a floor-leveling error of a centimeter or two and then resettle after about a second as a result of this action.

Rate Action or Derivative Control is added to proportional and reset control to decrease settling time and minimize deviations caused by upsets in system load. It operates within the proportional band to add an output force that is proportional to the negative rate of change (derivative) of the process variable. The output is thus reduced as the set point is approached, allowing the system to be operated with a smaller proportional band without excessive overshoot. In addition, an externally caused upset in the process, if it is fairly rapid, will produce an immediate and strong response from the controller, without waiting for the process to reach the limits of the proportional band.

Figure 14-14 shows the proportional, integral, and derivative outputs for a controller bringing a process from zero to a set point, and readjusting for an externally caused upset. The proportional band is about 40%. Integral and derivative outputs are limited to 10% and 40% of maximum proportional output, respectively. Notice that there is a small offset from the set point until the integral output builds up enough to reset it.

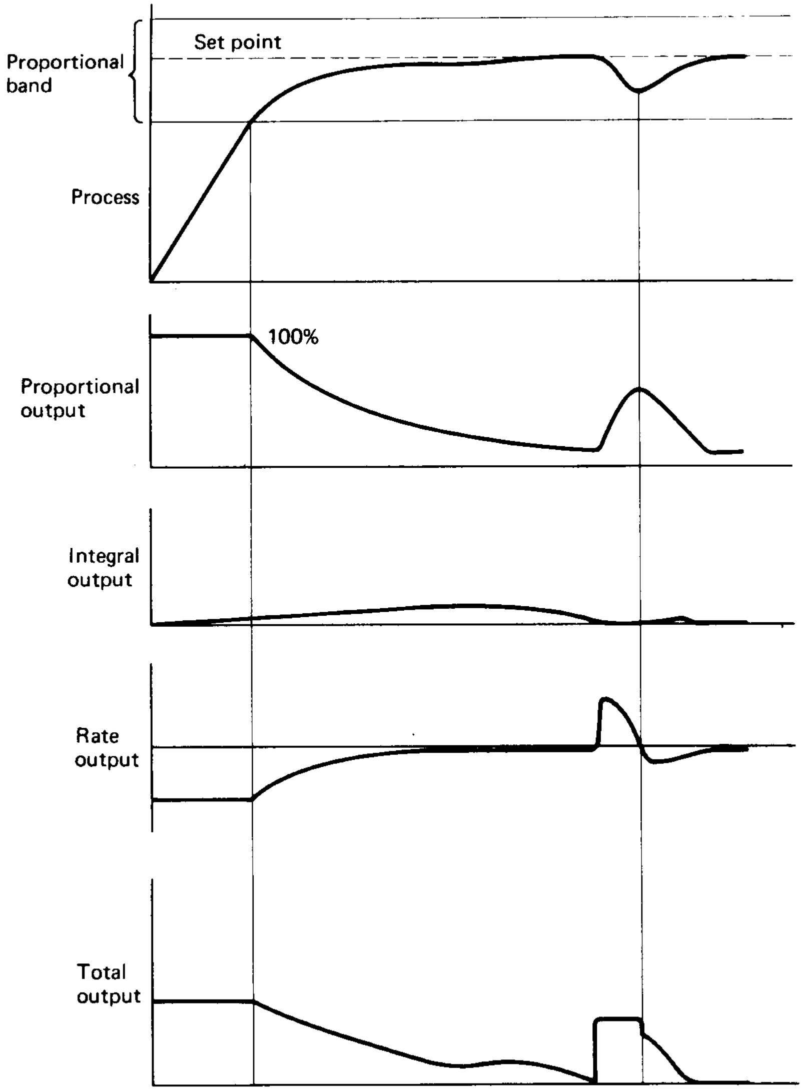


***FIGURE 14-14** Proportional plus integral plus rate-action control brings a process more quickly to the set point without overshoot or offset, and corrects more quickly for process upsets.*

Cascaded Controllers: It is sometimes advantageous to sense one or more secondary variables connected with a process in order to more effectively control the primary variable. In the furnace example, gas-line pressure might be sensed by a secondary controller whose output would be used to modify the overshoot compensation in the primary controller for continued minimum overshoot. The weight and temperature of the ingots arriving on the conveyer might be sensed by a third controller whose output would modify the demand point in anticipation of this external upset, thus keeping the process nearer the original set point. Such controllers are said to be operating in cascade.

Computed Control: The advent in the late 1960s of environmentally rugged minicomputers in the $10,000 price range made the application of computers to large control systems commonplace. Single-chip microprocessors in the $100 price range became available in the late 1970s, and brought computed control to small systems and even to individual instruments. The behavior of a computerized control system depends not so much on its structure or wiring as on the program that is loaded into it. Therefore, it is impossible to be specific about the actions of computerized controllers. Several general advantages can be pointed out, however.

1. Computerized controllers can be programmed to consider the history of a process as well as its current state. A sheet-steel-rolling process might then adjust the temperature and pressure at the second set of rollers based on the results achieved at the first set.
2. Programs can contain decision points altering the action taken or jumping control to an entirely different program. A bank of elevators can have separate programs for up traffic peak, day operation, down traffic peak, evening standby, and emergency evacuation.
3. Computers can handle more complex routines, including sequences and mathematical computations. One good example is transducer linearization, where a nonlinear device such as a thermistor is linearized by means of a lookup table or conversion program stored in memory.
4. Computerized controllers can scan a multiplicity of inputs—literally dozens in 1 ms—to take a variety of factors into account. A furnace controller might monitor temperature at six different places, gas pressure, air flow, temperature, and humidity, incoming ingot weight and temperature, and length of time the door is open—incorporating all of these factors into the control program.
5. Computers can handle several processes at once. A microprocessor in an automobile could control ignition timing and duration, carburetion, nonskid braking, transmission shifting, and passenger environment, and still have time to play spelling games with the kids in the back seat if we wished to so program it.
6. Control programs can be easily modified. Adding a step to a process or meeting a new environmental or safety spec is usually a simple and inexpensive matter of changing a program memory chip rather than buying a new controller or rewiring the old one.

15

HIGH-PERFORMANCE AMPLIFIERS

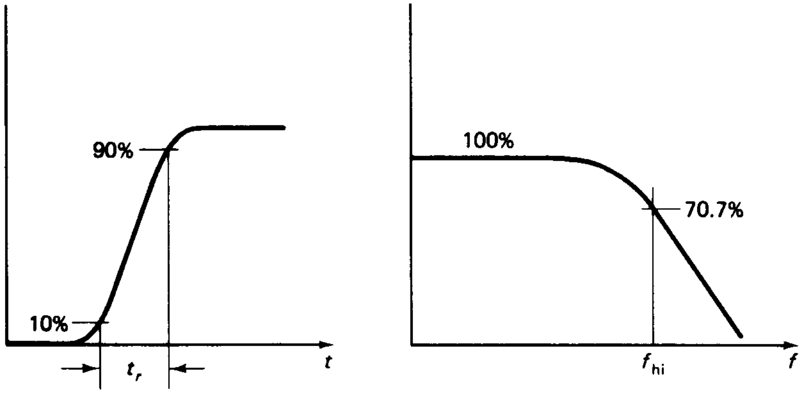
15.1 THE IDEAL AMPLIFIER

Amplification—the ability to use a low-power waveform to reproduce a higher-power copy of that waveform—is unquestionably the feat that gave birth to the art of electronics. The majority of electronic circuitry in use today consists of amplifiers in one form or another. We may be aware that the amplifiers normally treated in a basic electronics course have certain limitations, but it will be helpful to enumerate these in order to define the goal of producing a more nearly ideal high-performance amplifier.

Frequency Response: The ideal amplifier would respond equally to all frequencies from 0 Hz (dc) to any desired high frequency. Elementary amplifiers are capacitively coupled and do not respond to dc or very-low-frequency ac. High-frequency response of real amplifiers is also limited, dropping off somewhere in the range between 100 kHz and 10 MHz for low-power amplifiers of elementary design.

The ability to respond to high-frequency sine waves can also be expressed as the ability to respond quickly to a square or pulse input wave. For an amplifier without serious overshoot or ringing on the square-wave response, the relationship is

$$t_r = \frac{0.35}{f_{hi}} \tag{15-1}$$



$$t_r \approx \frac{0.35}{f_{hi}}$$

FIGURE 15-1 Rise time and frequency response are inversely related if there is no significant overshoot or ringing.

where t_r is the 10% to 90% rise time and f_{hi} is the -3-dB (0.707) high-frequency cutoff, as shown in Fig. 15-1.

Harmonic Distortion: An ideal amplifier would produce an output waveshape that exactly reproduced the waveshape of the input, although the amplitude, phase, and polarity might be different. Real amplifiers always produce some waveshape distortion, even though it may be very slight in modern amplifiers. Consider the case of a 4500-km (2800-mi) transcontinental communications cable with a booster amplifier every 150 km. If each amplifier produced a 0.5% flattening of the waveform peaks (output peak 99.5% of input peak), the total output after 29 amplifiers would be 0.995^{29}, or 0.86, a flattening of 14%.

Amplifier distortion is not commonly measured as described above, but by *percent harmonic distortion*. This is based upon Fourier's theorem that the only single-frequency waveform is the sine wave. All other waveshapes are composed of a fundamental-frequency sine wave with a number of harmonic (integer multiple) sine waves added. Figure 15-2 shows the Fourier spectrum (fundamental and relative strengths of harmonics) for triangle, square, half-wave rectified sine, and sawtooth waves, and gives a graphical addition of the first few harmonic sine waves. The ideal waveforms are approached as more harmonics are added.

Harmonic distortion measurement is accomplished by applying an extremely pure sine wave to the amplifier input, measuring the amplifier output, and adjusting an output meter's sensitivity to read this as 100%. A very sharp filter is then switched in to eliminate the fundamental-frequency sine wave from the output, allowing only the aberrations (harmonics) to pass to the meter. The meter (ideally a true rms meter, but commonly just an average-responding voltmeter) then indicates percent harmonic distortion. It should be emphasized that any departure from the

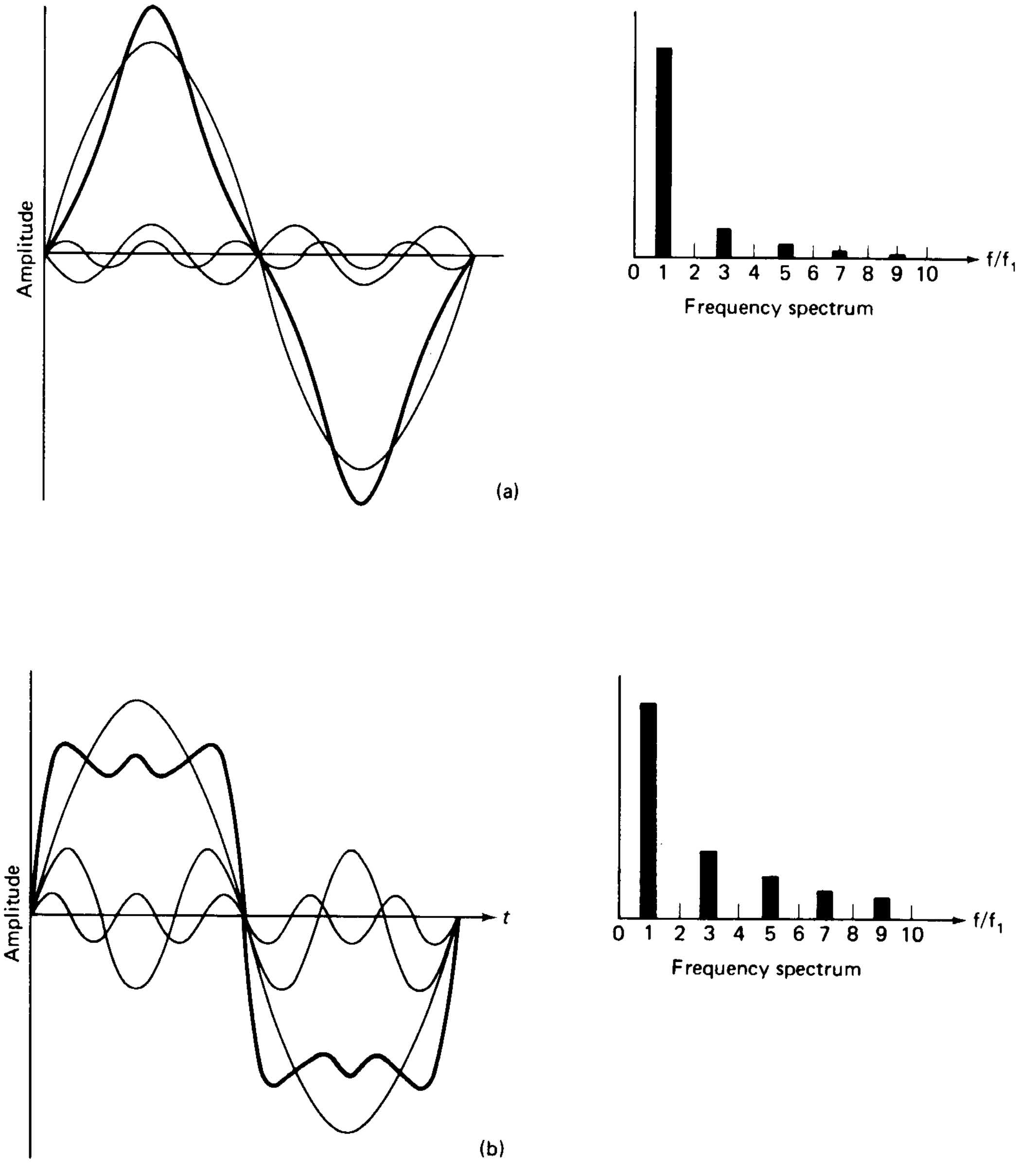


FIGURE 15-2 *Triangle waves (a) and square waves (b), may be synthesized by a fundamental and a series of harmonic sine waves.*

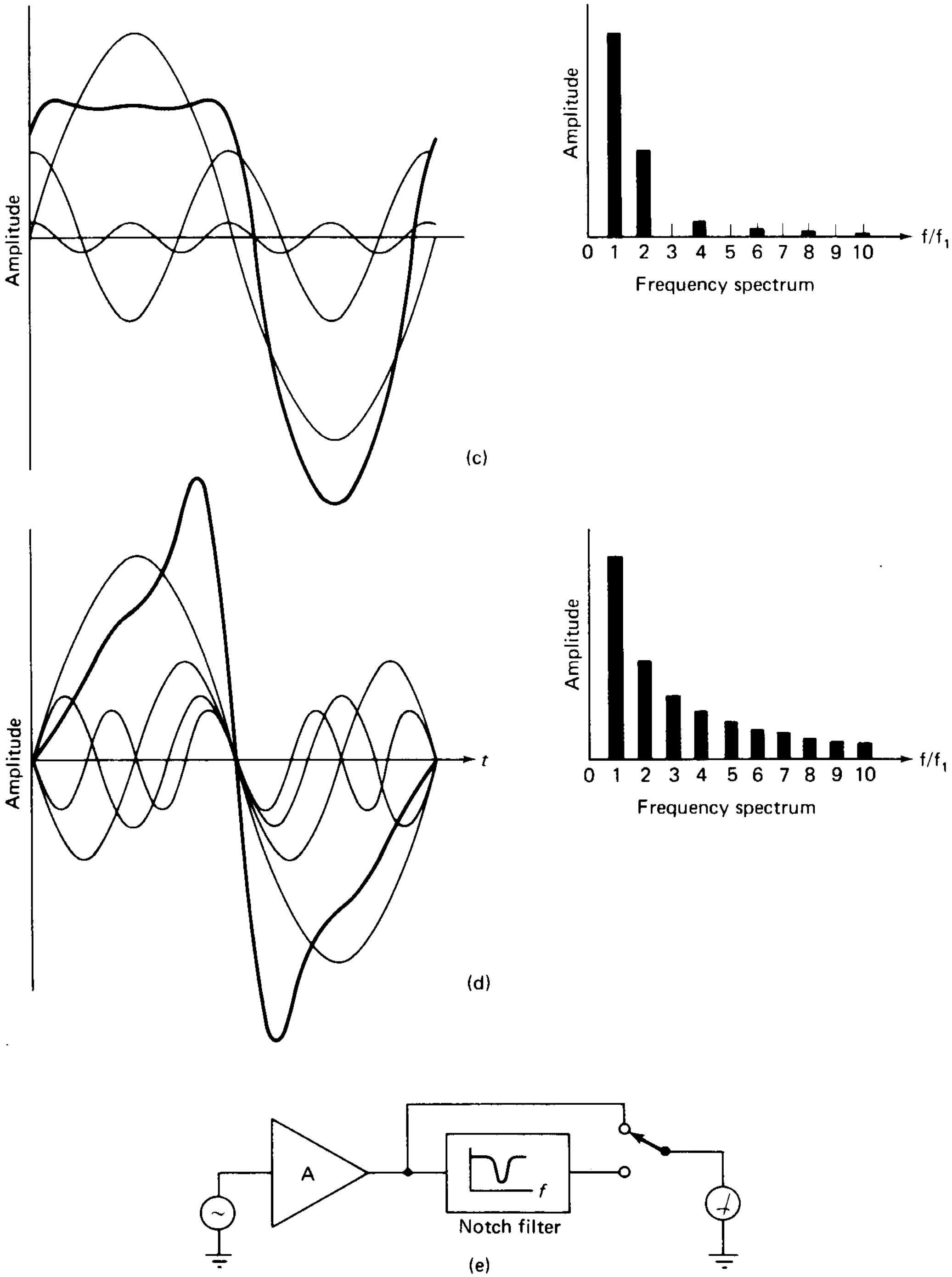


***FIGURE 15-2** Half-wave rectified sine (c) and sawtooth waves (d) synthesized from a fundamental and its harmonics. (e) A Harmonic-distortion meter tunes out the fundamental sine wave and measures the strength of the harmonics.*

pure sine-wave shape, no matter how it is produced, actually consists of various harmonics of various amplitudes and phases added to a fundamental sine wave. Only a pure sine wave is free of harmonics.

Voltage Gain: The ideal amplifier would have any desired voltage gain, and this gain would be variable by manual or electronic control. In practice, this is one of the easiest objectives to meet. The desired gain is built up by cascading several stages: four gain-of-10 stages to produce a gain of 10,000. Variable gain can be achieved in several ways: potentiometer voltage dividers, photoresistive cells, and changing bias points near the cutoff region for FETs can be employed to effect variable gain. As gains become very high (say, around 1000), undesired feedback via stray capacitance from output to input becomes a problem, and self-oscillation may result unless shielding and short straight-line signal paths are used.

Input Impedance: An ideal amplifier would draw absolutely no current from the signal source feeding it, which is to say that it would have infinite input impedance. At low frequencies this ideal is easily realized with an FET input stage, but at high frequencies capacitive reactance becomes a problem. At 30 MHz, for example, the 5-pF input capacitance of the typical FET has a reactance of 1 kΩ.

Output Impedance: The output voltage of the ideal amplifier would not be limited by or reduced by the addition of a load resistance, no matter how much current it drew. Real amplifiers generally show a drop in output voltage when a load draws current from the output. Often there is also a maximum peak output current beyond which gross distortion of the output will become apparent.

Efficiency: Ideally, all the power an amplifier draws from its dc supply would be delivered as signal power to the load resistance. Actually, the amplifier components themselves consume considerable power. Efficiency is defined as

$$\eta = \frac{\text{signal power out}}{\text{dc power supplied}} \tag{15-2}$$

In amplifiers delivering only 1-mW or so of power, the power waste of operating with an efficiency of 25% or even 2% is not objectionable. Audio amplifiers outputting 100 W and radio transmitters outputting several kilowatts are a different matter, however, and special pains are taken to get their efficiencies up to the 50 to 80% range.

15.2 NOISE REDUCTION

The detection of very weak signals—audio, radio, or otherwise—is never limited by lack of amplification. As we have seen, high amplification is simply a matter of cascading low-gain stages while taking reasonable care to isolate the outputs from stray coupling back to the inputs. The problem is always distinguishing the signal from the ever-present background noise. Once the signal level sinks down to the noise level, all attempts to amplify the signal will result in amplification of the noise as well. Anyone who has tried to tune in a weak TV station or listen to an AM radio during a thunderstorm is familiar with the noise problem. Noise is of two basic types: environmental and inherent.

Environmental Noise, more properly called *interference*, can often be eliminated or avoided once its source is identified. Shielding, filtering, and frequency selection are the common techniques employed. Figure 15-3(a) shows how a short length of unshielded microphone line can pick up 60-Hz noise by a stray capacitive coupling of only 10 pF to the 120-V house wiring. You can verify the 23-mV result by calculating the voltage division of 120 V across 265 MΩ capacitive in series with 50 kΩ resistive. This example is not at all exaggerated, and high-impedance instrumentation pickups (1 MΩ or 10 MΩ) are even more susceptible to noise pickup. High-frequency noise (from switching transients, motor brushes, lamp dimmers, etc.) is also present on the ac line, and is coupled in more strongly because of lower capacitive reactance.

Figure 15-3(b) shows how shielding the line and the microphone couples the noise to ground rather than to the amplifier input. Figure 15-3(c) shows some common mistakes in shielding. The shielding enclosure should not simply be placed around the device with the wires protruding through a hole, nor should each enclosure and shielded cable be grounded separately. Rather, as shown in Fig. 15-3(d), the shielding enclosures should completely enclose each device, with the circuit ground connected to the shield at one point internally. Signal wires leaving the enclosure must run through shielded wire, with the shield mated to the metal enclosure to ensure no breaks in the enclosing of the signal wire. The entire system of boxes and cables should be earth-grounded at only one point on one of the cabinets. The continuous shield connections will take care of holding the cabinets at ground, and no extra noise will be introduced thereby.

Figure 15-4(a) shows a motor brush-noise filter, which may keep high-frequency hash from entering the ac line, where it would be radiated all over. A simple capacitor, chosen to have a high reactance at audio frequencies but a low reactance to an interfering radio frequency, can often eliminate interference in tape or phone pickups, as shown in Fig. 15-4(b).

Frequencies to be avoided because of heavy concentrations of interference are:

- 60 Hz and its harmonics up to 300 Hz due to ac line radiation.
- 0.5 to 2.0 MHz because of AM broadcast and LORAN navigation transmitters.
- 27 MHz because of citizen's-band radio.
- 50 to 400 MHz because of broadcast TV, FM, police, and so on.

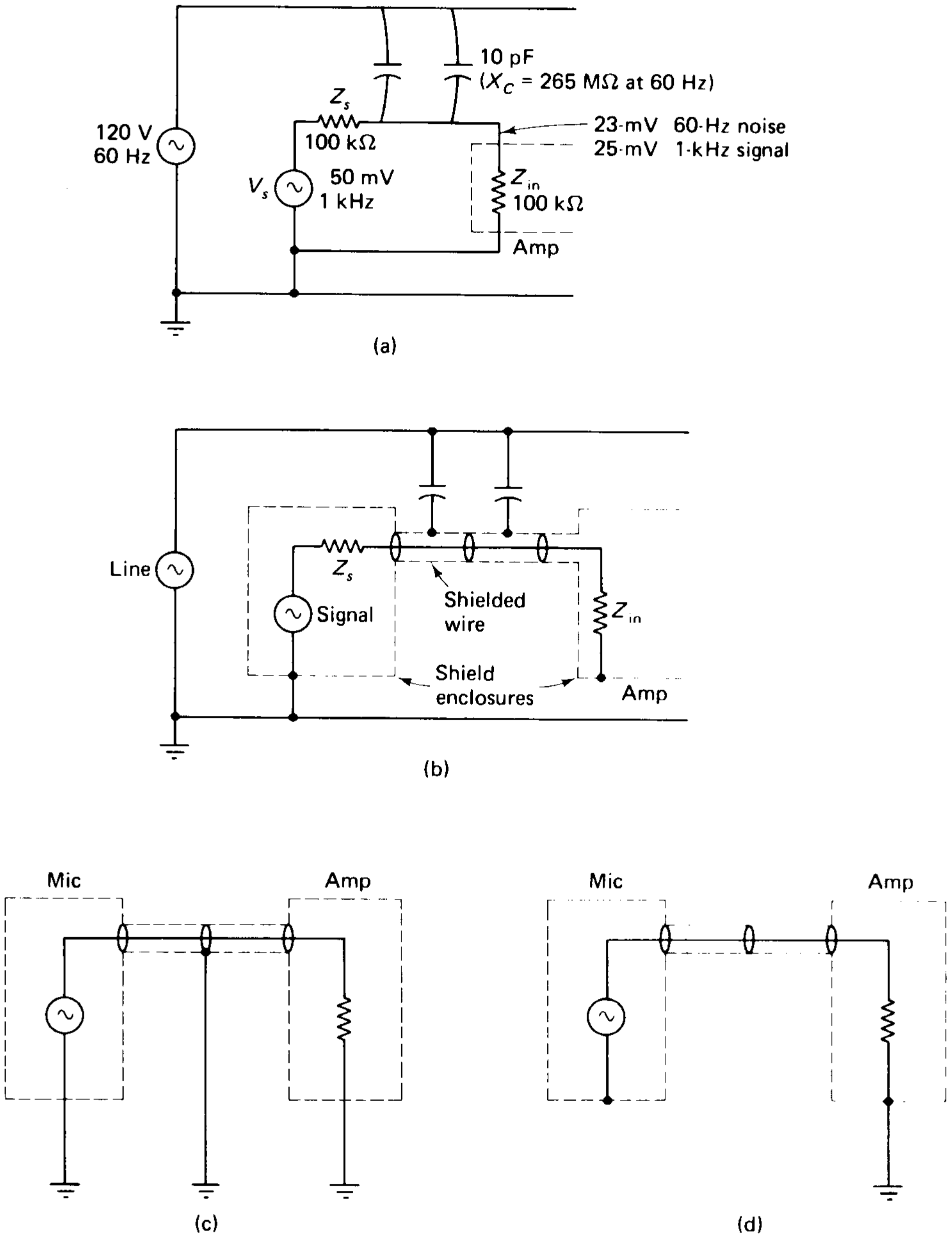


FIGURE 15-3 *(a) Even a small stray capacitance to the ac line causes serious noise pickup on a low-level high-impedance line. (b) Shielding eliminates noise pickup. Improper (c) and proper (d) shielding connections.*

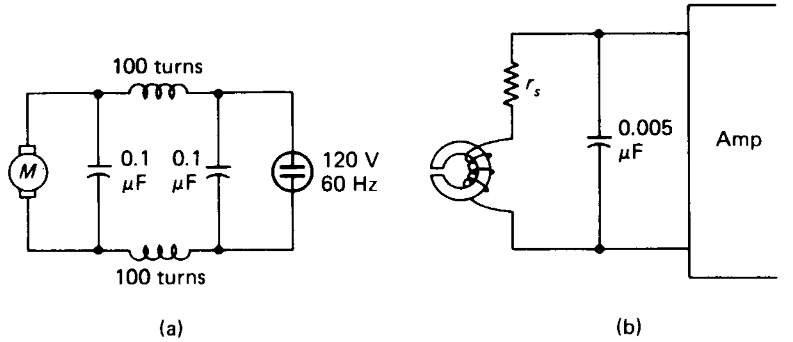


FIGURE 15-4 Noise-suppression techniques for a sparking-brush motor (a) and a magnetic-tape or phono pickup (b).

Thermal Noise is a voltage produced across the terminals of any resistance because of the random thermal vibrations of the atoms constituting it. The frequency spectrum of thermal noise extends from dc to well above the frequency limit of electronic amplification techniques. Its amplitude is given by

$$V_{n(\text{rms})} = \sqrt{(5.5 \times 10^{-23})TBR} \tag{15-3}$$

where T is temperature in Kelvins (Celsius + 273), B is the passband of the amplifier between the upper and lower -3-dB points in hertz, and R is the resistance in ohms. The peaks of the noise waveforms commonly reach four times the rms value. All resistive devices—bias resistors, receiving antennas, strain gages, semiconductors, and so on—produce thermal noise. It can be reduced by reducing the bandwidth of the amplifier or by lowering the temperature of the devices associated with the signal pickup. Indeed, space-probe receivers use very slow data rates (narrow bandwidth) and liquid nitrogen cooling to provide receivers capable of pulling the weakest possible signal from the noise.

EXAMPLE 15-1

What is the thermal noise produced by a 300-Ω television antenna at 27°C? The TV receiver has a bandwidth of 6 MHz.

Solution

$$V_{n(\text{rms})} = \sqrt{5.5 \times 10^{-23}(27 + 273) \times 6 \times 10^{6} \times 300}$$
$$= \mathbf{5.4\ \mu V}$$

This means that there is no possibility of receiving a TV station with an antenna signal of 5.5 μV or less, even with ideal receiving equipment.

Shot Noise is produced at any junction or interface by the passage of the charge carriers. It is given by

$$I_{n(\mathrm{rms})} = \sqrt{(3.2 \times 10^{-19})IB} \tag{15-4}$$

where I is the junction current in amperes and B is the amplifier bandwidth in hertz. This suggests that the input stage of a sensitive amplifier should be operated at a low bias current for lowest noise generation, and this is in fact the case.

EXAMPLE 15-2

What is the shot-noise voltage produced across the base–emitter junction of a transistor operated at an emitter bias current of 0.1 mA? The amplifier bandwidth is 6 MHz.

Solution

$$I_{n(\mathrm{rms})} = \sqrt{3.2 \times 10^{-19} \times 0.0001 \times 6 \times 10^{6}}$$

$$= \mathbf{0.014\ \mu A}$$

The ac resistance of the junction is found:

$$r_j = \frac{0.03}{I_E} = \frac{0.03}{0.0001} = 300\ \Omega$$

$$V_{n(\mathrm{rms})} = I_n r_j = 0.014\ \mu A \times 300\ \Omega = \mathbf{4.2\ \mu V}$$

Flicker Noise or $1/f$ Noise is produced by fluctuations in bias current, and has its strongest components at low frequencies, as the name implies. Above 1 kHz it is generally negligible, but frequencies below 100 Hz are often avoided in sensitive instruments to minimize its effects. It is most severe in semiconductors, so metal-film resistors may be chosen over carbon types in sensitive low-frequency instruments. Low bias currents in the input stage will minimize flicker noise.

Addition of Noise Components from different sources cannot be carried out directly ($V_a + V_b$), since the noise components at any instant may cancel rather than add. Rather, noise adds by the sum of the squares of its components:

$$V_t = \sqrt{V_a^2 + V_b^2} \tag{15-5}$$

where V_a is the rms voltage of the first source, and so on.

15.3 NOISE SPECIFICATIONS

Signal-to-Noise Ratio, S/N, is more properly termed signal-plus-noise to noise ratio, $(S+N)/N$, since it is impossible in practice to achieve a noiseless signal, S. An example will illustrate how the ratio is used.

Let us say that a receiver is tuned to 300 MHz with no signal at the antenna, and the volume is adjusted for a 1-V-rms noise output. A 4-μV signal is then applied (modulated with a 1-kHz tone) and the output rises to 2 V rms. We say that the receiver has a 4-μV sensitivity for a 2 : 1 (6-dB) signal-to-noise ratio.

Noise Figure, NF, is a measure of the deterioration in S/N ratio caused by an amplifier, mixer, or other signal processor. Ideally, it would be unity (0 dB), indicating that the amplifier simply amplified the input noise while contributing none of its own. It can be computed in two ways:

$$\mathrm{NF} = \frac{P_s/P_n\,(\text{input})}{P_s/P_n\,(\text{output})} = \frac{P_{n(\text{out})}}{P_{n(\text{in})}} \cdot \frac{1}{A_p} \tag{15-6}$$

where A_p is the amplifier power gain. Typical noise figures for well-designed transistor amplifiers are in the range 2 to 5 dB at 10 kHz, rising to perhaps 30 dB at 100 Hz because of flicker noise. As a rule of thumb, bipolar transistors give the lowest noise figure when operated with $V_{CE} \approx 2$ V and $I_C \approx 0.1$ mA, with a source impedance around 2 kΩ.

Equivalent Input Noise is sometimes specified as an alternative to noise figure for an amplifier. This is the input noise voltage that would have to be applied to the amplifier input to produce the observed noise output if the amplifier itself were noiseless.

Equivalent Noise Resistance is a second alternative to noise figure and is based on equation 15-3. The noise added by the amplifier is visualized as being due to a second fictitious resistance in series with the actual input source resistance. The thermal noise present at the input and the noise contributed by the amplifier can thus be compared directly as R_i/R_n, where R_n is the equivalent noise resistance.

Noise-Temperature Equivalent is a third alternative to noise figure, and is also based on equation 15-3. The two can be directly converted by the equation

$$\mathrm{NF} = \frac{T_e}{T_o} + 1 \tag{15-7}$$

where T_e is noise temperature, and T_o is ambient temperature, usually 20°C. Here the noise added by the amplifier is visualized as being due to a fictitious rise in the Kelvin temperature of the signal source resistance. Background noise from a radio receiving antenna is also often represented as being due to a fictitious rise in antenna temperature. This system has the advantage that all three noise sources (thermal, antenna background, and amplifier) can be represented in the same terms and compared and added directly.

Dynamic Range is an amplifier specification defining the ratio of maximum to minimum signals which the amplifier can faithfully reproduce. The upper limit is imposed by waveform (harmonic) distortion, 10% being the usually accepted figure. The lower limit is imposed by signal-to-noise ratio, with a S/N voltage ratio of 10 (20 dB) taken as the usual standard. A dynamic range of 40 dB (100 : 1 voltage ratio) is readily achieved in audio amplifiers. A range of 70 dB (3000 : 1 voltage ratio) is considered exceptional.

15.4 VOLTAGE FOLLOWERS

While emitter-follower and source-follower amplifiers do not provide voltage gain, they do have advantages that make them indispensable in designing amplifiers approaching the ideal in performance.

- Their gain is relatively unaffected by transistor parameters.
- They can handle large signal swings without distortion.
- They present a high input impedance (both resistive and capacitive) and therefore offer less loading to the driving source.
- They have a low output impedance. This means they can drive nonlinear loads with less waveform distortion, and capacitive loads with less high-frequency loading than other amplifiers.
- They maintain a relatively more stable bias point in the face of power supply, temperature, and unit-to-unit variations.
- They do not invert the input signal as common-emitter and common-source amplifiers do.

Gain formulas for emitter and source followers are given in Fig. 15-5. Notice that A_v of the emitter follower is to a small extent dependent upon I_E, which varies with v_{in}. As v_{in} increases, I_E increases, decreasing r_j, and increasing A_v. On negative v_{in} swings, I_E decreases, r_j increases, and A_v decreases. The effect (for *NPN* transistors) is to slightly exaggerate the positive signal peaks and flatten the negative peaks. The nonlinearity thus produced is negligible to the extent that the change in r_j is negligible compared to $R_E \| R_L$. Large V_{in} and low R_E or R_L values produce the greatest distortion. For the values shown, distortion approaches 0.1% for $V_{in} = 10$ V p-p.

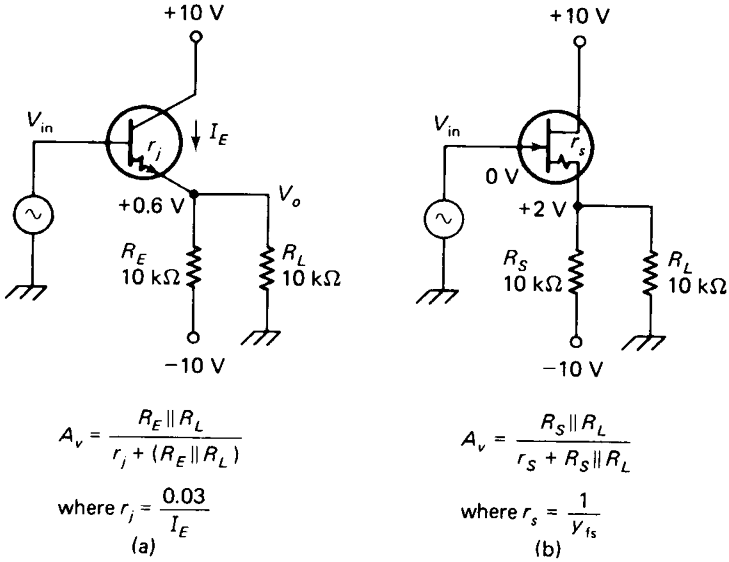


***FIGURE 15-5** Emitter followers (a) and source followers (b) have high Z_{in}, low Z_o, and low distortion.*

A_v for the source follower also increases on positive signal peaks (for an N-channel FET). This is because y_{fs} increases at higher I_D. The distortion is likely to be more severe with the source follower because r_s is generally a much larger value than r_j (250 Ω compared to 32 Ω for the values given in the figure).

Longtailing is a commonly used technique for minimizing voltage-follower distortion by stabilizing the emitter (or drain) current, thus keeping r_j (or r_s) constant. In its simplest form, R_E (or R_S) of Fig. 15-5 would simply be increased to 100 kΩ and the negative supply would be raised to −100 V. The input signal swing (even 10 V p-p) would thus produce a much smaller change in bias current, resulting in less distortion. For longtailing to be effective, the peak current drawn by R_L must be much less than the bias current employed. This means low V_{in}, high I_E, or high R_L. The extra 90 mW lost in the 100-kΩ resistor and the expense of the −100-V supply are the prices one pays for improved linearity.

Current-source bias is the ultimate in longtailing, giving, in effect, an infinite R_E (or R_S). Figure 15-6(a) shows the technique applied to an FET. R_1 and R_2 hold 2 V across R_3, setting I_C of Q_2 at 1 mA regardless of V_C. Again, i_{RL} must be much less than I_C to make this technique effective.

Figure 15-6(b) shows a "short tail" source-follower circuit which *should be avoided* except where V_{in} is only a few millivolts or where linearity is not essential.

Input Protection against accidentally applied overvoltages is good practice and is exemplified in Fig. 15-7. R_1 would be calculated to limit the gate current to a safe value (say, 10 mA) for any anticipated positive V_{in}. P_{R1} should be calculated, as it is

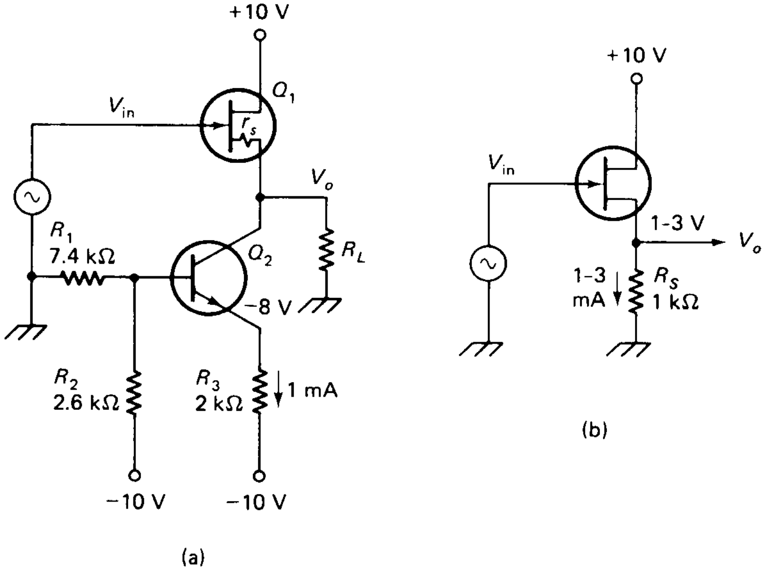


***FIGURE 15-6** (a) Current-source Q_2 stabilizes the bias point and minimizes distortion due to changing r_s. (b) Low voltage across R_S makes the circuit bias-unstable and increases distortion due to changing r_s.*

apt to be surprisingly high. The FET gate junction will avalanche at typically − 40 V, making even − 10 mA unsafe, so D_1 is used to shunt negative overloads to the − 10-V supply. For MOSFETs a diode pointing to the + 10-V supply should be added. The diodes should be fast signal types. Power-supply diodes are too slow and have a large capacitance across their junctions.

R_1 will drop no voltage at dc and low-frequency ac, since the FET gate and back-biased diode have almost infinite resistance. At some high frequency, however, the FET input capacitance will have a reactance equal to R_1, and serious signal loss (×0.707) will result. This happens at 88 kHz for $C_{in} = 6$ pF and $R_1 = 300$ kΩ.

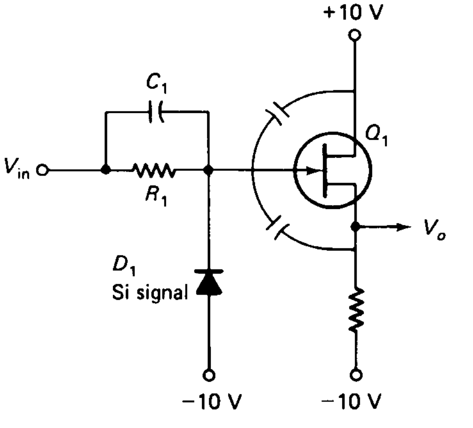


***FIGURE 15-7** R_1 and D_1 protect the gate from accidental overvoltage, but stray capacitance loads R_1 at high frequencies. C_1 prevents this loading, but nullifies protection at high frequencies.*

C_1 is added to prevent this drop, and should be 100 times C_{in} for 1% loss, $50C_{in}$ for 2% loss, and so on.

It must be understood that the addition of C_1 nullifies the input protection action at high frequencies, so special care should be taken to guard against high-frequency instrument overloads. TV horizontal sweep circuits, inverter-type power supplies, and radio transmitters are especially dangerous in this respect. Fortunately, most overloads occur at dc or 60 Hz.

High-Frequency Transients often appear in voltage followers handling sub-microsecond pulses. Figure 15-8(a) shows the reason for this problem. When the +1-V input pulse is first applied, the gate rises by +1 V, but the source (for a few ns) rises by only 0.5 V, because of the voltage division of stray capacitances C_g and C_s. This creates a +0.5-V gate-source signal which demands a rather large increase in I_D. Some of this current surge flows back through C_g developing voltage across r_s which further increases the voltage at the gate and consequently the drain current demand, in a regenerative effect.

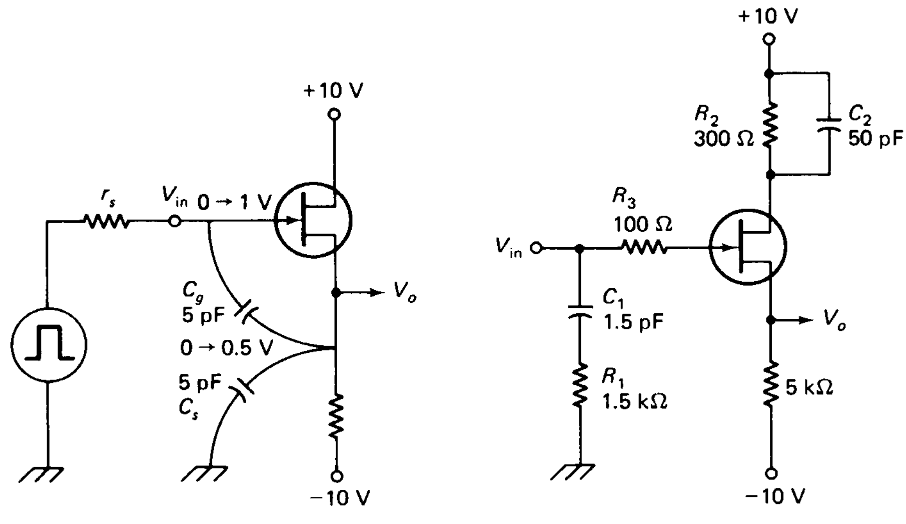


FIGURE 15-8 (a) Voltage division on stray capacitance C_g and C_s causes current surges and ringing on fast-rise pulses. (b) Three techniques for improving fast pulse response.

The three techniques commonly used to suppress this ringing are shown with typical values in Fig. 15-8(b), although a single circuit is not likely to employ all three at once. R_3 limits the rate of rise of gate voltage, and is most effective with amplifiers operating below 10 MHz. C_1 and R_1 absorb the current fed back through C_g and must be sized to match the stray capacitance C_g. C_2 and R_2 allow an out-of-phase voltage to be developed at the FET drain. Stray drain-gate capacitance couples this back to the gate, introducing a degeneration to offset the regenerative ringing.

15.5 DIFFERENTIAL AMPLIFIERS

The differential amplifier has properties that make it an excellent counterpart of the voltage follower in producing high-performance amplifiers.

- It responds to dc as well as ac, allowing it to handle transducer-generated signals directly.
- It can be adjusted for zero offset—that is, zero dc out for zero dc in.
- It contains no capacitors or inductors, and can therefore be fabricated easily in IC form.
- Its voltage gain can be varied without upsetting the bias point.
- Temperature, component, and power-supply variations are mostly cancelled by the balanced nature of the circuit.

On the disadvantage side, the dif amp does not have a particularly high input impedance or low output impedance. Also, its output does not appear from one point to ground, but rather between two points, both at a dc level above ground. This may be an advantage or a disadvantage.

Figure 15-9 shows an FET dif amp with variable gain. Q_1 is a current source, providing a fixed current through the wiper of R_4, which then splits between Q_2 and Q_3. Any increase in I_{D2} (caused by a positive v_{in}) must therefore be accompanied by an equal *decrease* in I_{D3}. If V_{D2} goes down by 1 V, V_{D3} will then go up by 1 V. The output signal v_o is thus formed by mirror-image signals at the two terminals.

Balance: Since V_{GS} may be slightly different for the two FETs, R_4 is varied to compensate for the difference. If R_4 is set to make v_o zero when v_{in} is zero, we are adjusting for zero offset. If R_4 is set to make $V_{R5} = 0$, we are adjusting for *dc balance*, a condition wherein the setting of R_5 varies the gain but not the zero signal dc output of the amp. To achieve both zero offset and dc balance, the gate of Q_3 could be connected to the wiper of a trimpot between $\pm V_{CC}$.

Notice the balanced nature of the circuit. If $+V_{CC}$ increases, V_{D2} will increase, but so will V_{D3}. Net v_o is unchanged. If temperature rises, turning Q_2 on more, Q_3 will experience an identical change, and again there will be no change in v_o.

Gain Calculation: Consider that the right-hand end of r_{s3} (Fig. 15-9) is grounded. Also remember that current source Q_1 presents an infinite resistance, so the wiper of R_4 is connected to an open circuit from a signal point of view. Q_2 is now a conventional common-source amp, and A_v is the ratio of (drain-line resistance)/(source-line resistance), since the same current i_D flows in each. The drain line contains only R_6. The source line contains r_{s2}, in series with R_5 and R_4 in parallel, in series with r_{s3}. The total gain is twice that calculated for Q_2, since Q_3

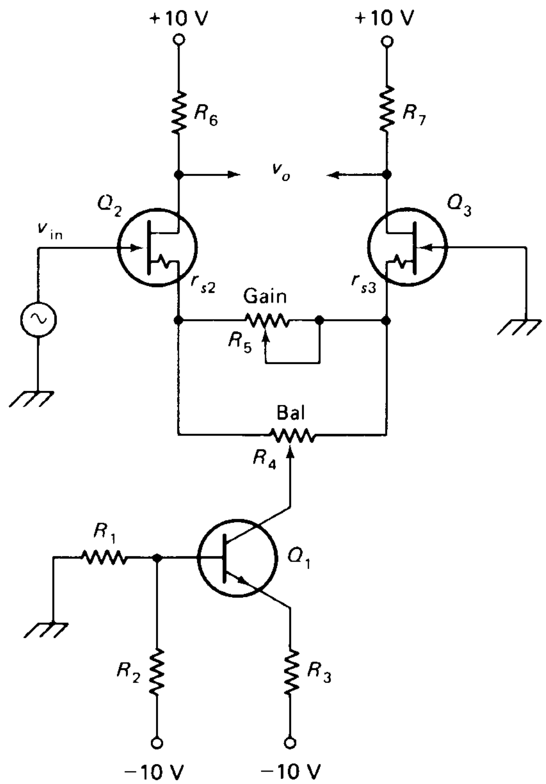


FIGURE 15-9 FET differential amplifier with variable gain.

makes an equal contribution:

$$A_{v(\text{dif})} = \frac{V_o}{V_{\text{in}}} = \frac{R_6 + R_7}{r_{s2} + (R_4 \| R_5) + r_{s3}} \tag{15-8}$$

Of course, in most cases $R_6 = R_7$ and $r_{s2} = r_{s3}$.

For dif amps using bipolar transistors, the source resistances r_s are replaced with emitter junction resistance r_j. We must also be sure that the base of Q_2 is fed from a negligibly low impedance [say less than $\frac{1}{100}$ of β $(r_{j2} + R_4 \| R_5 + r_{j3})$].

Differential Inputs: Instead of grounding the gate of Q_3 (Fig. 15-9) we could have connected one side of the input signal to it, as in Fig. 15-10. The differential amp will now amplify only the difference between the two input gates (V_{dif}) and will ignore any signals appearing in common between the two inputs and ground (V_{com}). This is because the FETs split the current from Q_1 equally as long as their gate voltages are equal. Raising both gates by 1 V simply results in raising V_{C1} by 1 V.

Differential inputs are useful in rejecting common-mode noise picked up by the input wires, especially where these wires must be long or are fed by a

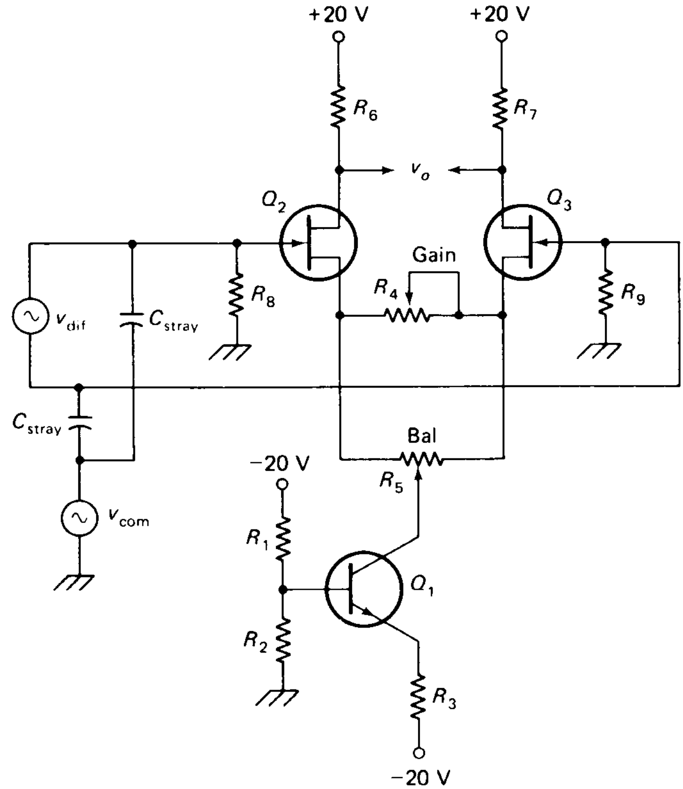


***FIGURE 15-10** Dif amp responds to the difference in voltage between two ungrounded input lines. V_{com} is rejected.*

high-impedance source. As a typical example, V_{dif} may be a remote station speaker used as a microphone on an intercom system. The long wires back to the amplifier input pick up considerable noise from the ac line via C_{stray}, but the noise is picked up in common by both wires. Only the differential signal from the speaker is amplified by the dif amp. Differential amplifiers are used for similar reasons in biomedical sensing and in computer core-memory readout amplifiers.

General-purpose test instruments often have differential inputs so they can monitor the voltage difference between two points, neither one of which is grounded. (Connecting the ground probe of a test instrument to a point other than ground of the circuit under test is generally a poor idea. It shorts the test point to ground if the two devices both have their chassis at ac line ground. Even if the instrument is isolated from the ac line, a great deal of noise will be picked up by the chassis and transformer capacitance to the line, and this noise will then be injected into the circuit under test. Only battery-operated instruments with plastic or isolated cases should have their "common" probe connected to a point other than ground.)

Common-Mode-Rejection Ratio, or CMRR, is the specification detailing how many times more sensitive an instrument is to differential inputs than to common-mode signals. Ideally, the ratio would be infinite, indicating no response at all to common-mode signals, but amplifier and input-attenuator imperfections generally place the ratio between one hundred and several thousand.

To measure CMRR, one input is grounded and an input signal is applied to the other input to produce a certain output (say 1 division deflection on the screen of an oscilloscope). Now the two inputs are tied together and an input is applied in common between both of them and ground, such as to produce the same 1-division deflection. The ratio of the two inputs required is the CMRR.

Common-Mode Voltage Limit: If the voltage applied in common to the gates (Fig. 15-10) becomes too positive, Q_2 and Q_3 will enter the saturation region and be unable to amplify the differential signal. Also, if the common-mode voltage goes too negative, Q_1 will saturate and will cease to function as a current source. These two phenomena define the common-mode voltage limits of the amplifier.

Common-mode voltage limits can be measured with the test setup shown in Fig. 15-11. An output reading from V_{dif} is obtained with V_{com} at 0 V dc. V_{com} is then increased in the positive direction until the accuracy of V_{dif} falls out of spec (−3% for a 3%-accuracy instrument). The test is then repeated to determine the maximum tolerable negative V_{com}.

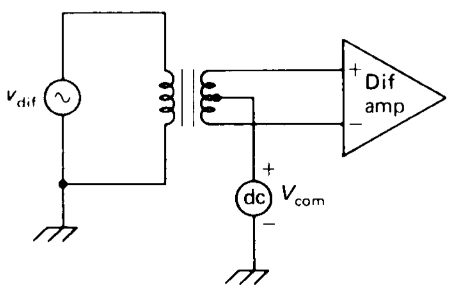
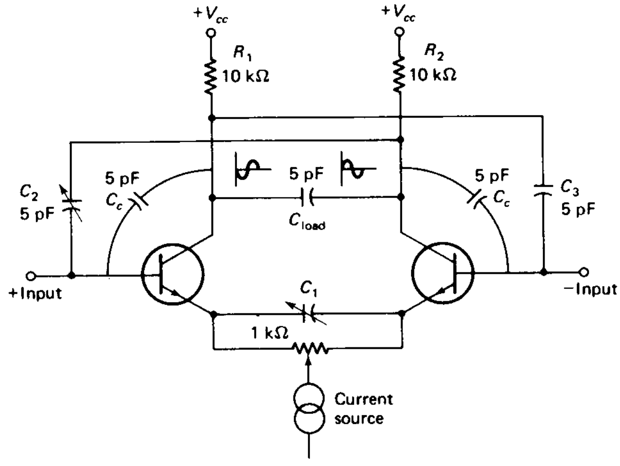
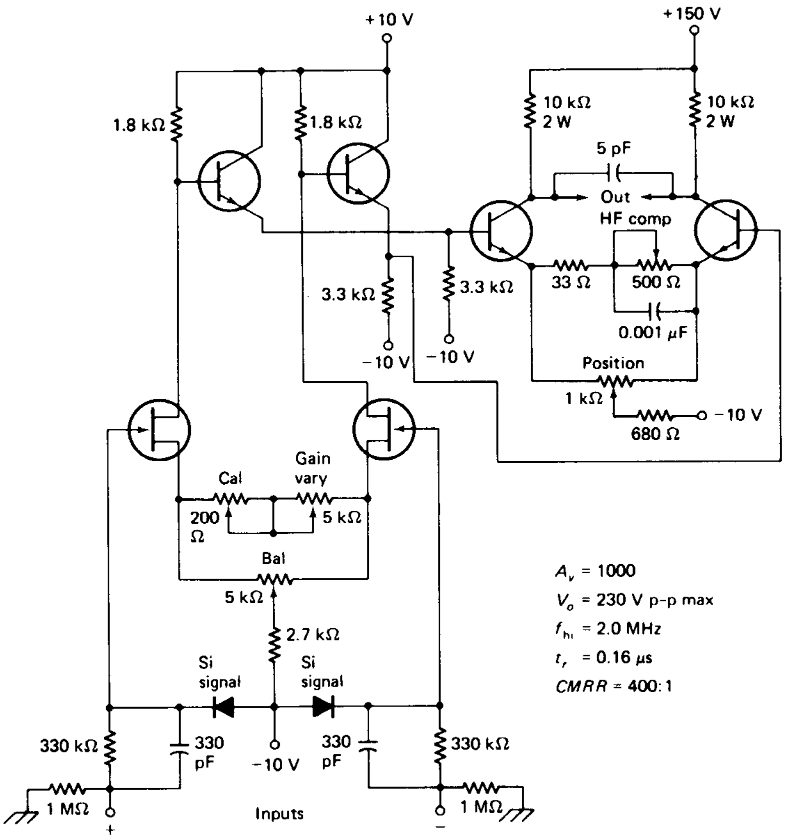


FIGURE 15-11 Circuit for measuring the common-mode voltage limits of an amplifier.

Frequency Compensation: The collector-junction capacitance of a dif amp may typically be 5 pF to ground. The load may present another 5 pF across the outputs, which appears as 10 pF to ground because of the Miller effect with $2V_c$ across it. This is illustrated in Fig. 15-12(a). If the collector resistors are 10 kΩ, the X_C will equal R at about 1 MHz, and the output will be loaded down to 0.707 of full output at that frequency. However, a compensating capacitor C_1 can be placed across the emitter resistance as shown, sized to decrease the emitter line impedance in exact step with the decrease in collector line impedance. The high-frequency limit of the amplifier can generally be extended by a factor of 3 or 4 with this



(a)



(b)

FIGURE 15-12** (a) Stray capacitances become important at high frequencies.* **C_1 compensates for lowered gain caused by C_{load}** ***and **C_2 compensates for lowered Z_{in}** ***caused by*** **C_c. (b) Complete high-performance amplifier: $A_v \approx 1000$, $f_{hi} \approx 2$ MHz, $V_o \approx 230\ V_{p-p}$.**

technique. To determine the value of C_1 required:

$$C_1R_1 = (C_c + 2C_{\text{load}})R_2 \tag{15-9}$$

$$C_1 = \frac{15\text{ pF} \times 10\text{ k}\Omega}{1\text{ k}\Omega} = \mathbf{150\ pF}$$

In practice, C_1 would be adjusted for a high-frequency output (say, 2 MHz) equal to the low-frequency output (say, 1 kHz). A check would then be made to see that no overpeaking was present in the vicinity of 1 MHz. If a waveform can be observed at a later stage, a 200-kHz square wave can be applied and C_1 adjusted for sharpest response without overshoot. However, no extra loading from scope or meter probes can be tolerated at the collectors of the stage being compensated.

Often C_1 must exceed 1000 pF, making a variable unit impractical. In this case, C_1 can be fixed and R_1 can be shunted with a variable resistor which is adjusted to achieve compensation. The gain of the amplifier will then be dictated by stray capacitances which are quite unreliable. If an exact overall gain is required, there must be another stage of amplification which is variable without regard to frequency compensation. Variable gain in a compensated dif amp is seldom attempted, because it would require a variable C_1 to track with a variable resistance in the emitter.

Input Capacitance: The circuit of Fig. 15-12(a) has a gain of 10 to each collector, so the Miller effect makes the 5-pF junction capacitance from base to collector look like (5 pF) (10 + 1) or 55 pF to ground. This is objectionably high, as it causes low input impedance at high frequencies (2.9 kΩ at 1 MHz). A common solution is to drive the dif amp with an emitter or source follower, but it is possible to neutralize the input capacitance by feeding an inverted signal to the base from the opposite collector of the differential pair (C_2 and C_3 in Fig. 15-12). Only C_2 is required if the $-in$ terminal is to be kept at ground. C_2 should be variable. Variable gain can be employed with no problem.

Figure 15-12(b) shows a complete 3-MHz differential amplifier incorporating the features presented in this chapter. Voltage gain can be set as high as 1000, with $Z_{\text{in}} = 1\text{ M}\Omega$, $V_o = 230$ V p-p(max), and Bw = 3 MHz.

15.6 LEVEL SHIFTERS

Differential amplifiers using *NPN* or *N*-channel transistors unavoidably have their outputs at a more positive dc level (with respect to ground) than their inputs. Often, we wish to take our output from one of the differential transistors to ground, and it then becomes necessary to shift the dc level of the output back to zero. There are several ways to achieve this:

- Use a zener diode in series with the output to obtain the required level shift. This works well in digital circuits, but avalanche noise and temperature drift are objectionable in low-level amplifiers.

- Alternate *NPN* and *PNP* stages, letting the *NPN*'s offset the level in the positive direction and the *PNP*'s return it to zero with their negative offset. This can be done with discrete components but is usually avoided in integrated circuits because of the difficulty of producing both types of transistors on the same chip.
- Employ a voltage divider to the negative supply, as in Fig. 15-13(a). The loss of voltage gain ($\frac{10}{16}$ in the example) is not serious, but an emitter follower Q_3 is required to reduce loading of the Q_2 collector.
- Employ a current source (Q_4) and dropping resistor (R_4) as in Fig. 15-13(b). This is similar to the technique of Fig. 15-13(a) except that R_5 has been replaced by a current source that has infinite dynamic resistance, so there is no loss in voltage gain.

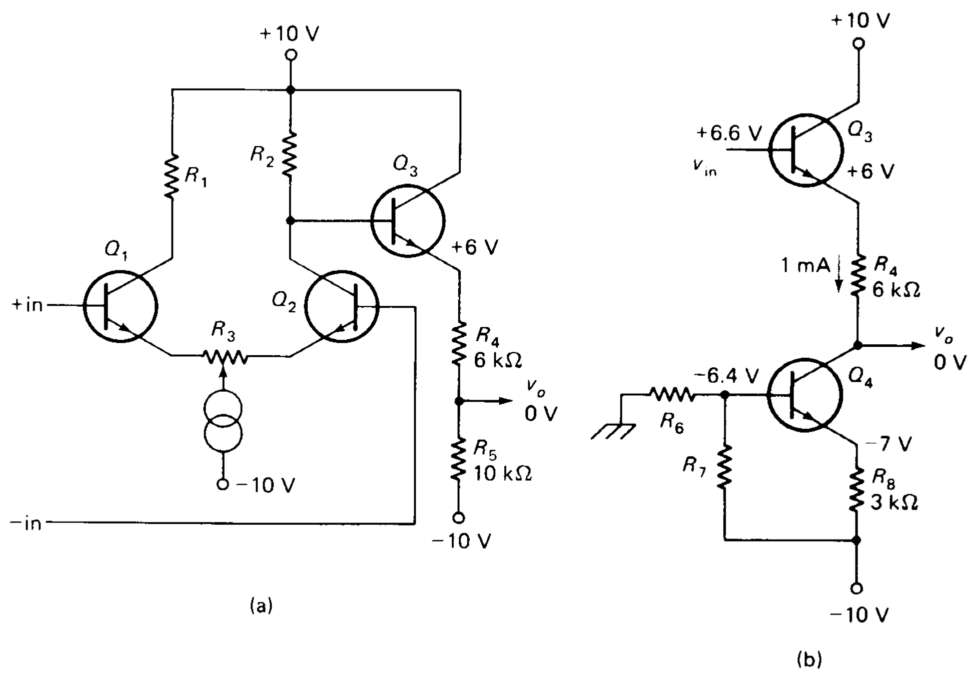


FIGURE 15-13 *(a) Voltage-dividing level shifter. (b) Current-source Q_4 replaces R_5.*

15.7 TOTEM-POLE OUTPUT

Single-ended (as opposed to differential) amplifiers and voltage followers have a slower response to fast pulses in the current-decreasing direction than in the current-increasing direction. This is illustrated for an emitter follower in Fig. 15-14(a). This circuit has an *active pullup* but a *passive pulldown*. The load capacitance C_L is charged through the transistor's collector with a current which may be as large as βI_B. Discharge, however, follows the "negative exponential"

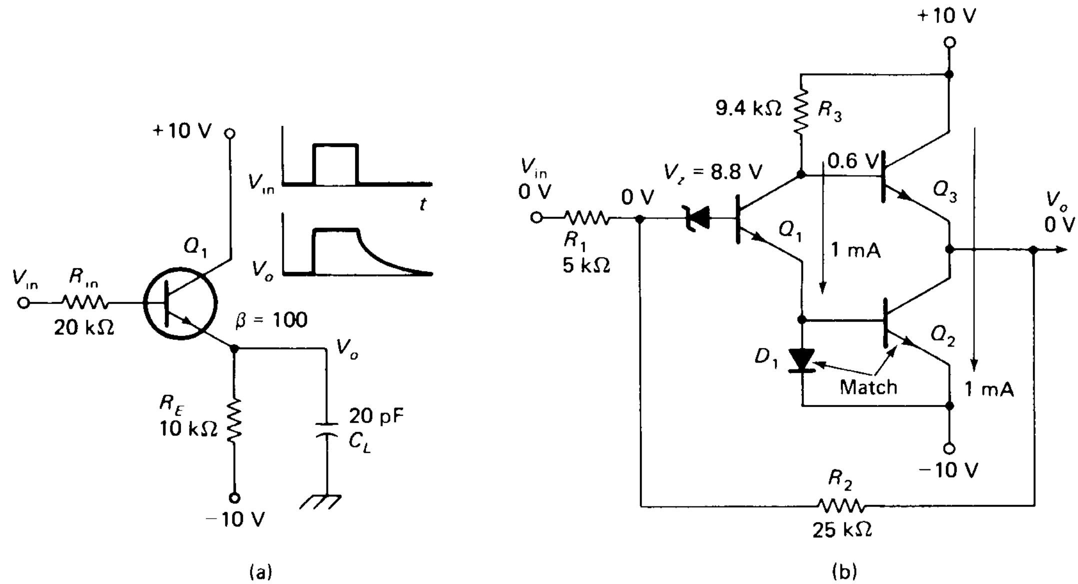


***FIGURE 15-14** (a) An emitter follower has active pullup but passive pulldown. (b) Totem-pole output: Q_3 provides active pullup, Q_2 provides active pulldown.*

time-constant curve, $\tau_{dis} = R_E C_L$. Discharge to 5% will take three time constants, which for the values given amounts to

$$3\tau_{dis} = 3 \times 10\ \text{k}\Omega \times 20\ \text{pF}$$
$$= \mathbf{600\ ns}$$

Charge can be accomplished in a shorter time:

$$3\tau_{chg} = 3 \times \frac{20\ \text{k}\Omega}{100} \times 20\ \text{pF}$$
$$= \mathbf{12\ ns}$$

since the current through R_{in} is multiplied by β.

In a dif amp, one transistor's current is increasing while the other's is decreasing, so there is no lack of symmetry. The response time is also not terribly fast if C_L is appreciable, which is why the *totem-pole* circuit of Fig. 15-14(b) is often used as an output stage. This circuit has active pullup and pulldown, and is similar in many respects to a complementary-symmetry stage.

This circuit uses a *current repeater* D_1Q_2 which can be effective only in integrated-circuit form. The base–emitter diode of Q_2 is identical to D_1, which is only possible since they are on the same chip. They are connected in parallel, so it is certain that they will have the same voltage across them and hence will carry the same current. I_{E1} is therefore equal to I_{E2} and I_{E3}. If V_o attempts to move positive from 0 V, R_2 will turn Q_1 on, which will turn Q_2 on, lowering V_o. If V_o drops below

0 V, Q_1 will turn more off, raising V_{B3}, turning Q_3 on and raising V_o. R_3 sets the idle currents at 1 mA, and R_2/R_1 set the gain of the stage at $\frac{25}{5} = 5$. Level shifting is required to bring the -8.8 V base of Q_1 to zero. A zener is shown for simplicity, but one of the methods shown in Fig. 15-13 would be used in practice.

15.8 CHOPPER-STABILIZED AMPLIFIERS

Sometimes it is necessary to amplify dc signals which are so small that they are lost in the dc bias-point drift of even a well-designed differential amplifier. The solution in this case is to chop the input dc into pulses which are then amplified by a simple ac-coupled amplifier and finally detected and filtered back to dc.

If the dc involved is always of one polarity, we can stop at this point, but most often the dc is actually the output of a potentiometer or bridge circuit and may be of either polarity. In that case we need a detector capable of discerning whether the input signal is positive or negative.

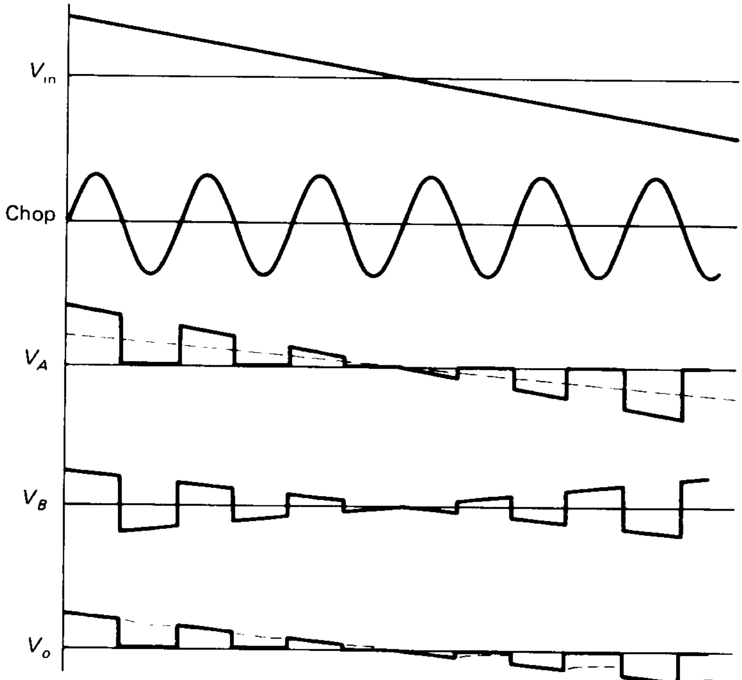
Phase Detector: Figure 15-15(a) represents a dc input signal changing from positive to negative (V_{in}). V_{in} is passed when the chopping signal is positive and blocked when it is negative, producing the signal V_A. V_A is amplified by a capacitor-coupled amp, producing V_B. *V_B is passed when the chopping signal is positive*, and blocked when it is negative, producing V_o. V_o is filtered by a capacitor, reproducing the original transitioning dc waveform.

In practice, the transition would be more likely to take 400 cycles of the chopping signal than the 4 cycles shown, but the principle is the same. The phase of the chopped signal with respect to the chopping signal changes as V_{in} changes from positive to negative, and the detector converts amplified ac of one phase into a positive output while the opposite phase produces negative output.

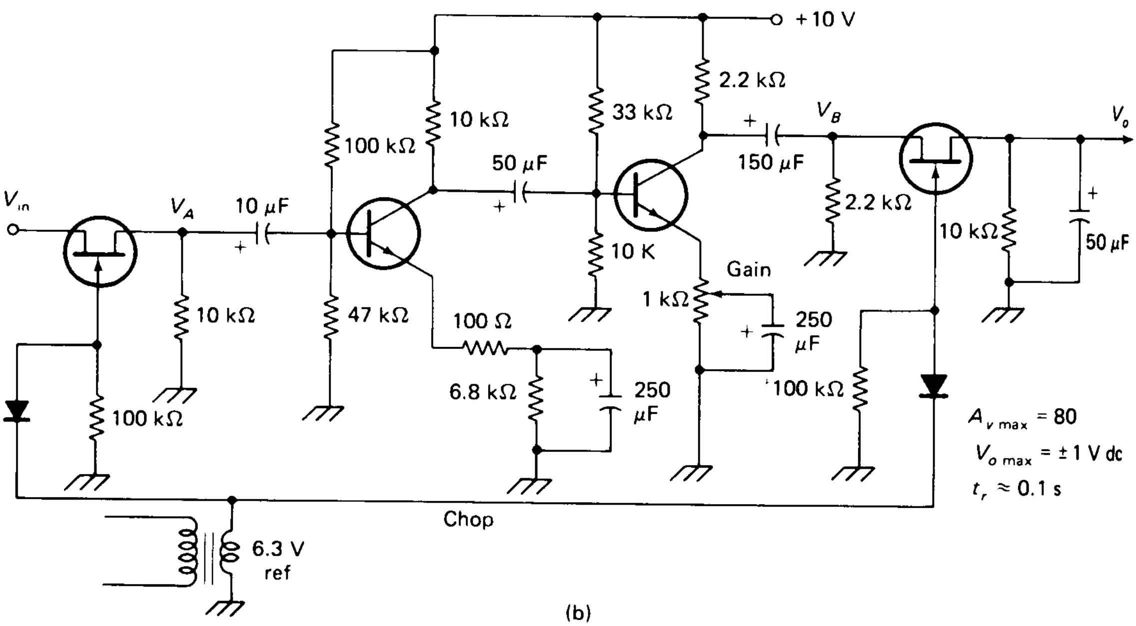
Choppers and phase detectors have taken many forms: mechanically vibrating contacts, light beams with rotating wheels and photocells, transformers and switching diodes, and others. The FET outperforms most of these devices because it is bilateral (source and drain interchangeable), its gate signal does not inject noise to the source or drain, and it offers solid-state reliability.

Figure 15-15(b) shows a chopper, amplifier, and phase detector employing FETs. Notice that no special biasing or temperature-compensating techniques are employed, yet this system will always produce 0 V out for 0 V in, regardless of power supply and temperature variations. The chopping frequency of 60 Hz is common because it is readily available and it makes it possible to cancel noise picked up at the line frequency.

The frequency response of a chopper amp ranges from dc to about $\frac{1}{10}$ of the chopping signal frequency, or 6 Hz in the example circuit. The upper frequency can be extended to the megahertz range by shunting the signals above 6 Hz to an ac-coupled amplifier and combining its output with the phase detector's output. Such a system is properly called a chopper-stabilized amplifier, and has the benefits of high-frequency response and near-zero dc drift.



(a)



(b)

***FIGURE 15-15** A chopper and phase detector amplifies dc with zero offset and drift: (a) pertinent waveforms; (b) circuit diagram.*

16

OPERATIONAL AMPLIFIERS

16.1 OP-AMP CHARACTERISTICS

Operational amplifiers were originally developed in vacuum-tube days to perform the mathematical *operations* of adding, multiplying, and integrating as required in analog computers. The name has stuck but the performance and range of applications for the op amp have grown apace.

There is no strictly accepted definition, but op amps generally have the following characteristics:

- Frequency response from dc to several kHz (although several "fast" op amps operate well into the MHz range).
- Ability to handle both positive and negative dc inputs and outputs.
- Near-zero offset; that is, almost zero input produces zero output voltage.
- Very high voltage gain—typically several tens of thousands.
- High input impedance—typically above 100 kΩ.
- Differential inputs.

Internally, an op amp usually consists of two or three differential-amplifier stages, a level shifter, and a complementary-symmetry or totem-pole output stage.

However, there is no need to be concerned with the internals at this point, because most op amps are integrated-circuit devices with no access to the internals.

The symbol for an op amp is given in Fig. 16-1(a). A positive voltage at the inverting input (− symbol) causes the output to swing negative. A positive voltage at the noninverting input (+ symbol) causes the output to go positive. The gain of the op amp is so high that the voltage between the + and − inputs is *virtually* (very nearly) zero, even for maximum output. This observation forms the basis of much of our analysis of op-amp circuits. Notice that two power supplies are required for the op amp, one positive and one negative with respect to ground. These supply connections are often omitted in circuit diagrams, but of course they cannot be overlooked in wiring the circuit.

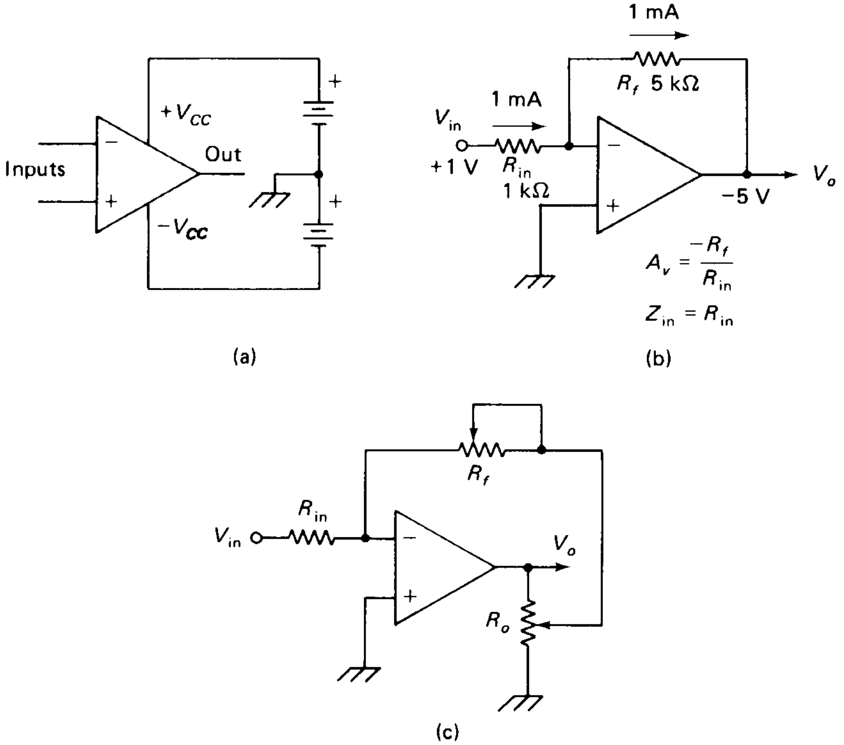


***FIGURE 16-1** (a) Op-amp symbol and power-supply connections. (b) Basic inverting op-amp circuit with example values for gain of −5. (c) Two methods of varying circuit gain.*

16.2 THE INVERTING OP-AMP CIRCUIT

The most popular op-amp circuit is the inverting amplifier of Fig. 16-1(b). The gain of this circuit is controlled by the ratio of R_f/R_{in} and is almost completely unaffected by internal op-amp gain, supply voltage, or temperature. The example numbers given with the figure illustrate why this is true. The +1 V at V_{in} will bring the − input positive, forcing the output negative. The negative output will force a current through R_f in the direction shown. Now it is important to remember that the voltage and current at the − input are virtually zero. (For 5 V out and $A_v = 10{,}000$, $V_{in} = 5/10{,}000 = 0.5$ mV. The input current, for an r_{in} of 1 MΩ would be 0.5 mV/1 MΩ, or 0.5 nA.) Therefore, $I_{R\text{in}}$ is virtually equal to I_{Rf} and the junction of R_{in} and R_f is virtually at ground potential. The only V_o that will produce the required 1 mA in R_f is −5 V. A less-negative V_o would immediately route more of the R_{in} current to the − input, making V_o more negative.

The Input Impedance for this circuit is equal to R_{in}, since the right-hand end of R_{in} is virtual ground. R_f can be a variable or a photosensitive resistor, making A_v variable. Values of R_f above 1 MΩ generally require special op amps or offset-adjust circuits to keep the output at zero for zero input. Figure 16-1(c) shows two methods of gain control for the inverting op-amp circuit.

16.3 THE NONINVERTING OP-AMP CIRCUIT

The noninverting circuit has a much higher input impedance than the inverting circuit, but it requires the input signal to appear as a common-mode voltage on the + and − inputs. This means that common-mode voltage limit and input stray capacitance present problems which do not appear in the inverting circuit where both inputs are kept near ground. The noninverting circuit is therefore less easy to tame than the inverting circuit.

Figure 16-2(a) shows the noninverting circuit and its gain and input-resistance equations. R_f and R_{in} form a voltage divider applying a fraction of V_o back to the

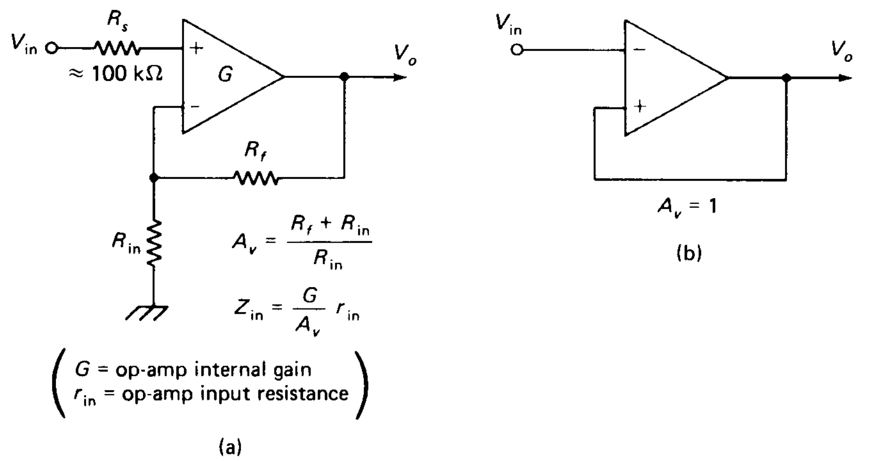


FIGURE 16-2 (a) The noninverting circuit gives much higher input impedance than the inverting circuit. (b) Gain-of-one voltage follower.

− input. The voltage between +in and −in is virtually zero, so this fraction must equal V_{in}.

R_s is not essential, but is recommended to protect the op amp where overvoltages could be applied to the input. Figure 16-2(b) shows a unity-gain voltage follower, which is really just a special case of the noninverting circuit with $R_f = 0$ and $R_{in} \to \infty$.

16.4 OFFSET VOLTAGE—CAUSES AND CURES

An ideal op amp would have zero output voltage with zero input voltage and zero input current. In practice, zero input will cause the output to *offset* (assume some nonzero value) in order to feed back a current and voltage to the − input through R_f as required by input-stage imperfections. This output-voltage offset can be minimized by keeping the value of R_f low, but there are several other methods available.

Input Bias Current is the current required by the bases of the differential-input transistors within the op amp. This current is assumed to be 1 μA for each input in the gain-of-10 circuit of Fig. 16-3(a), although modern op amps are available with bias currents tens and even hundreds of times less than this. Notice that, with the input grounded, + 1 V is required at the output to supply 1 μA through R_f to the − input. The current through R_{in} is zero since the − input is virtually ground. The 1 μA to the + input comes from ground but causes no voltage drop, since there is no resistance in this line.

The + 1-V output-voltage offset could be reduced by decreasing R_f, but this would require a corresponding decrease of R_{in}, resulting in lower input impedance. An op amp with lower bias current could be used, and indeed low-cost/low-speed op amps are available with bias currents on the order of 10 nA, and at a higher price 1 nA is obtainable. High-speed op amps generally have higher bias-current requirements, however.

Figure 16-3(b) shows how a bias-compensating resistor R_b can be added at the + input to eliminate output offset if the input bias currents are equal. The − input bias current is obtained through R_{in} and R_f, which are in parallel if $V_o = 0$ and V_{in} is grounded. R_b is chosen equal to $R_{in} \| R_f$, so the voltage developed at the − input (across $R_{in} \| R_f$) equals the voltage developed at the + input (across R_b). The differential-input voltage is then virtually zero. The value of R_{in} in this determination includes the resistance of the signal source if it is not negligible.

For the noninverting circuit, bias compensation requires that the input protection resistor R_s assume the function of R_b, and have a value $R_{in} \| R_f$, as shown in Fig. 16-3(c). Here R_b includes the resistance of the signal source.

Input Offset Current: The + and − input transistors of an IC op amp are fabricated on the same chip, so their bias-current requirements tend to be equal. To the extent that they are not equal, the compensating resistor R_b will be ineffective

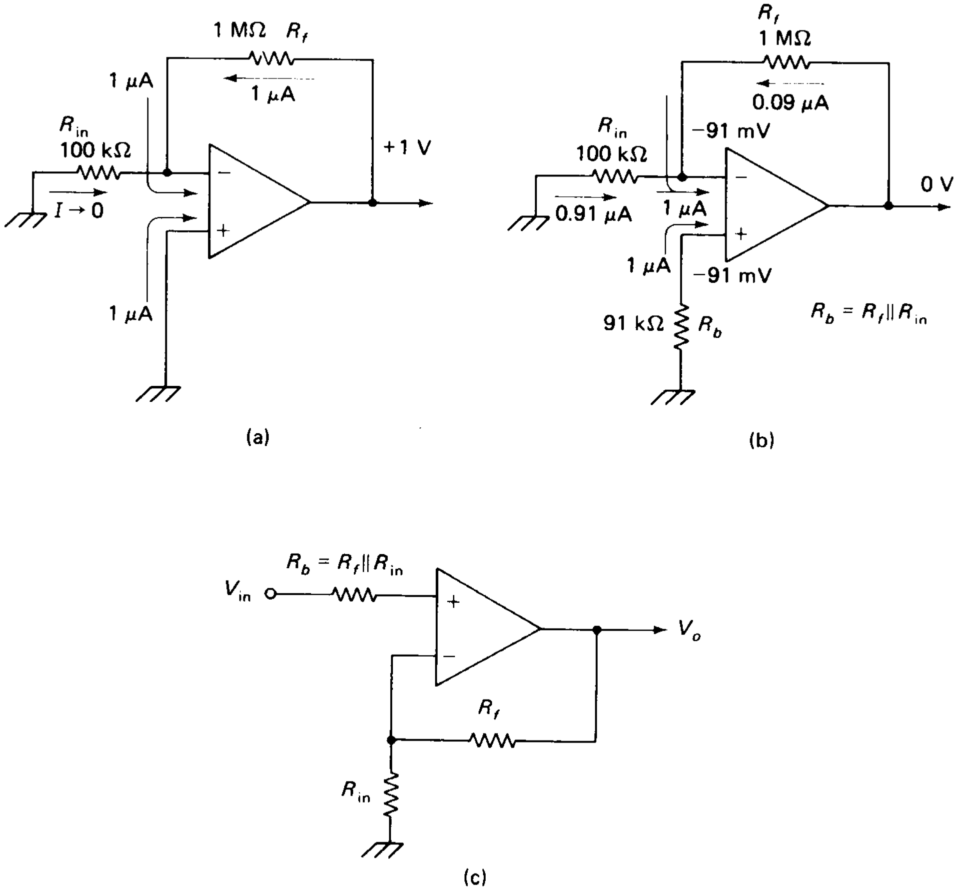


***FIGURE 16-3** (a) Input bias current requirements force the output to assume a nonzero value even when the input is zero. (b) Bias-compensating resistor R_b minimizes output offset. (c) Placement of R_b in noninverting circuit.*

in eliminating output offset. The difference between the input bias currents is called input offset current, and is typically $\frac{1}{10}$ to $\frac{1}{3}$ of bias current for most op amps.

Some op amps have a terminal or pair of terminals provided for connection of an offset null pot, in which case the manufacturer's recommendation should be followed. If no such provision is made, the circuit of Fig. 16-4(a) can be employed. The bias-current-compensating resistor R_b is still recommended, even though R_{os} and R_{null} could supply all the bias required by the − input. This is because the bias currents change with temperature, and the setting required of R_{null} would therefore change with temperature. However, the bias current changes do tend to track with temperature, so the inclusion of R_b makes output offset less sensitive to temperature changes.

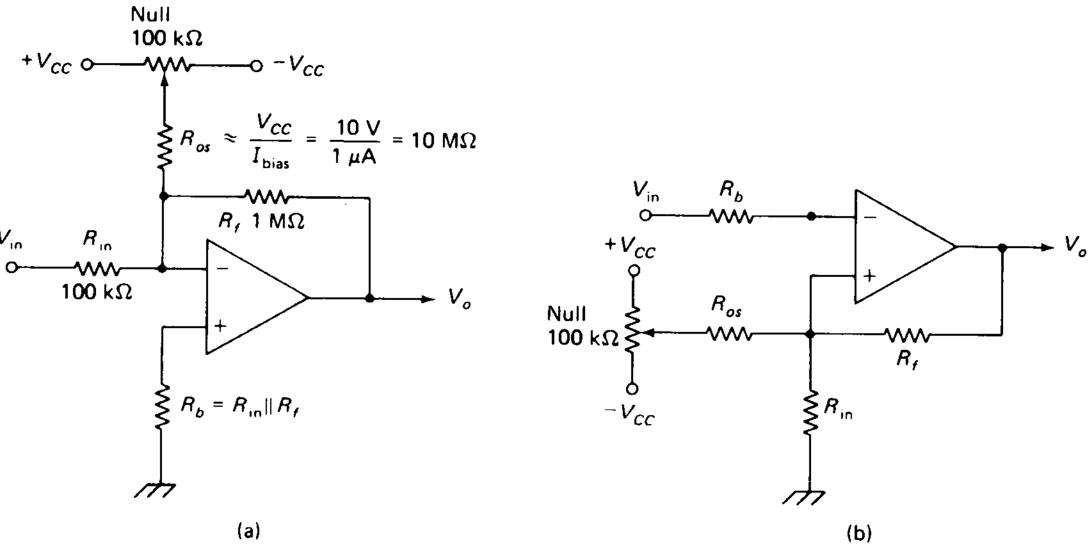


***FIGURE 16-4** Offset current is the difference between the two input bias currents. Offset null circuits for inverting (a) and noninverting (b) amps are shown.*

The null circuit for the noninverting op amp is shown in Fig. 16-4(b). Note that in this case R_{os} is effectively in shunt with R_{in}, and the circuit gain will be increased if R_{os} is not many times larger than R_{in}.

Input Offset Voltage is the voltage difference that must be applied between the two inputs of an op amp to force its output to go to zero. Ideally, both transistors of the input dif amp would turn on at the same voltage, but slight differences do exist. Input offset voltage is specified at 0.5 to 10 mV for commonly available op amps. In Fig. 16-5(a), input offset voltage is represented as a voltage source in series with the + input in an inverting gain-of-10 circuit. Notice that the output voltage must assume a value $V_{os}(A_v + 1)$ in order to bring the ideal + and − inputs to zero differential voltage. (Bias and offset currents are being neglected for the moment.) Decreasing R_f and R_{in} will not reduce the output offset due to V_{os} as it will in the case of I_{bias} and I_{os}. Offset voltage is usually the predominant problem in high-gain circuits, whereas offset current tends to predominate in high-impedance circuits.

Figure 16-5(b) shows a circuit designed specifically to null out input offset voltage in an inverting op amp. Of course, the current null circuit can be used to take care of voltage and current offset in one adjustment. It is not common practice to include both null circuits because it is very difficult to monitor the actual input offsets, and both controls affect the output offset in the same way. An exception to this may occur when gain is variable and the source resistance r_s varies widely. The offset voltage control would first be used to null the output with r_s at minimum (input grounded). Then the offset current control would be adjusted for

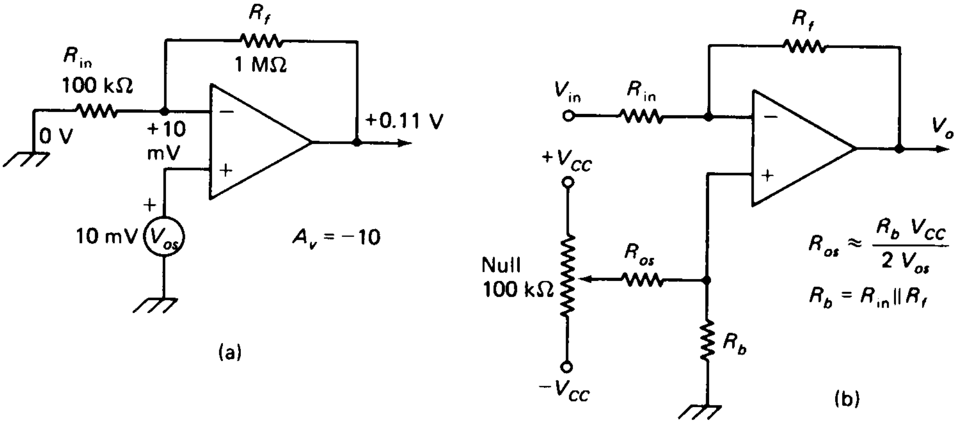


FIGURE 16-5 *(a) Input offset voltage also causes output offset. (b) Input offset voltage null circuit.*

zero output with r_s at maximum. Repeating these two steps, a point should be reached where zero input voltage produces zero output voltage regardless of gain setting or input-resistance conditions.

16.5 HIGH-FREQUENCY EFFECTS

Feedback and Oscillation: When used as linear amplifiers, op amps are connected with negative (or 180°) feedback from output to inverting input. At high frequencies, additional lagging phase shifts are developed as the collector resistance of one stage feeds the base capacitance of the following stage. At some frequency, this internal phase shift will total 180°, resulting in a total phase shift of 360° in the fed-back signal. If the amplifier gain at this frequency is sufficient to make the fed-back signal equal the input signal, the signal will regenerate, and the amplifier will self-oscillate. To prevent this from happening, it is necessary to reduce the high-frequency gain of the amplifier so that the fed-back signal will not be sufficient to regenerate the input signal. This is accomplished by negative feedback or signal shunting of high frequencies through small-value capacitors or *RC* networks. This is termed *frequency compensation.*

External Frequency-Compensation Components are added according to manufacturers' data sheets and application notes, since each op amp has its own requirements. In general, low-A_v circuits have large amounts of feedback (low R_f) and consequently require greater attenuation of high-frequency gain to preserve stability. Too little compensation will produce ringing on square-wave edges and oscillation if it is severe. Too much compensation will simply lower the upper frequency limit of the amplifier. Figure 16-6(a) and (b) shows the compensating components and resulting frequency response for the type 709 op amp. Notice that curve 1 compensation will provide a gain-of-100 amplifier, but the upper frequency

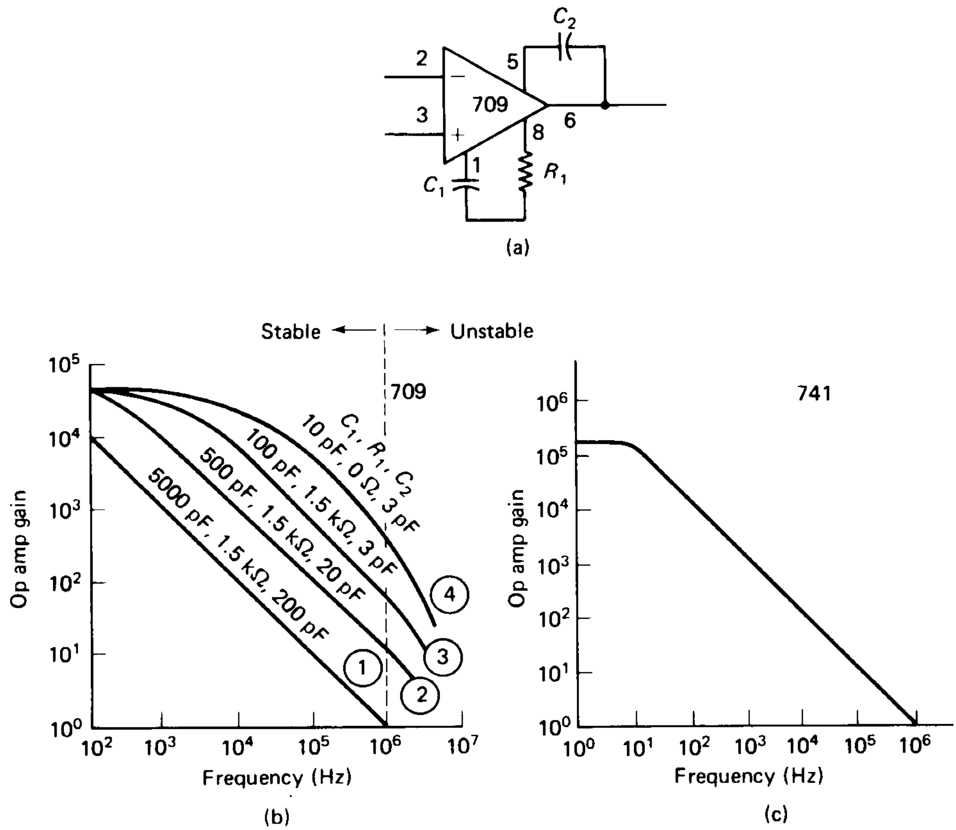


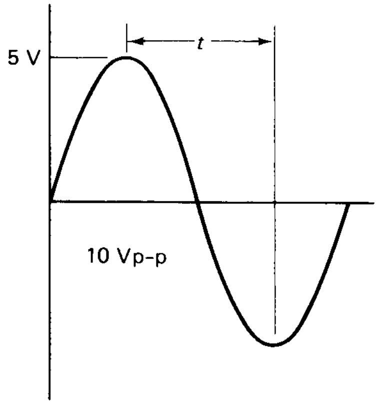
***FIGURE 16-6** Frequency compensation is required to prevent oscillation in op-amp amplifiers: (a) compensation components for the 709 op amp; (b) open-loop gain versus frequency for the compensated 709; (c) open-loop gain drops off rapidly at high frequencies for the internally compensated 741.*

limit will be only 10 kHz. Curve 2 compensation will permit a gain of 100 with a 100-kHz bandwidth. However if curve 2 compensation is used on a gain-of-3 amplifier, oscillation would be the certain result.

Internal Compensation relieves' the user of the burden of selecting compensating components. The compensating capacitor is integrated on the chip and sized for the heaviest feedback possible, which is a unity-gain circuit. Such fixed compensation allows maximum frequency response only at unity gain. For higher gains frequency response is reduced proportionally. Figure 16-6(c) shows frequency response versus gain for the 741 op amp.

Feed-Forward Compensation is a technique of shunting high frequencies around a relatively slow stage of an op amp through an external capacitor of a few hundred picofarads. Frequency response can be improved by as much as a factor of 10 in op amps with this option. Feed forward is applicable only to the inverting circuit.

Slew Rate is the maximum rate of change in output voltage (expressed in V/μs) in response to a much-faster-rising input step. This parameter may impose its limitation at a much lower frequency than the frequency-response curves would indicate if the output signal is large. For example, the curve of Fig. 16-6(c) indicates that the 741 op amp should have an upper frequency limit of 100 kHz in a gain-of-10 circuit. However, the slew-rate spec is only 0.3 V/μs, which, as Fig. 16-7 shows, implies a frequency limit of 15 kHz for a 10-V-p-p output sine wave. Slew rate is generally measured in the voltage-follower circuit of Fig. 16-2(b), which is the worst-case condition.



$$t_{min} = \frac{10\text{ V}}{0.3\text{ V}/\mu\text{s}} = 33\ \mu\text{s}$$

$$p = 2t = 66\ \mu\text{s}$$

$$f_{max} = \frac{1}{p_{min}} = 15\text{ kHz}$$

FIGURE 16-7 Slew-rate limitations cause large voltages to be attenuated at lower frequencies than small voltages.

16.6 OP-AMP APPLICATIONS

FET Inputs: Instrumentation circuits sometimes require a dc-responding amplifier with a near-infinite input impedance and low offset regardless of source resistance. The FET-input op amp in a noninverting circuit fills the requirement nicely. Bias current for the 536 op amp (sometimes called a Bi-FET because of its combination of bipolar and FET circuitry) is specified at 30 pA, maximum. Input offset voltage for the Bi-FETs is several times greater than for bipolar inputs, but this is easily nulled out. Figure 16-8(a) shows the low output offset resulting when a Bi-FET op amp is driven from a high-impedance source. Even the best bipolar inputs will produce offsets 100 times as great.

Figure 16-8(b) shows how a pair of discrete FETs can be used to obtain some of this advantage with a conventional op amp. The circuit is noninverting, with the feedback returned to the FET input. The FETs are differential source followers, with the NULL pot making up for differences in V_{GS}. The FETs should be matched for $V_{GS(\text{off})}$ and thermally bonded together, or if available, should be a matched pair in a single package.

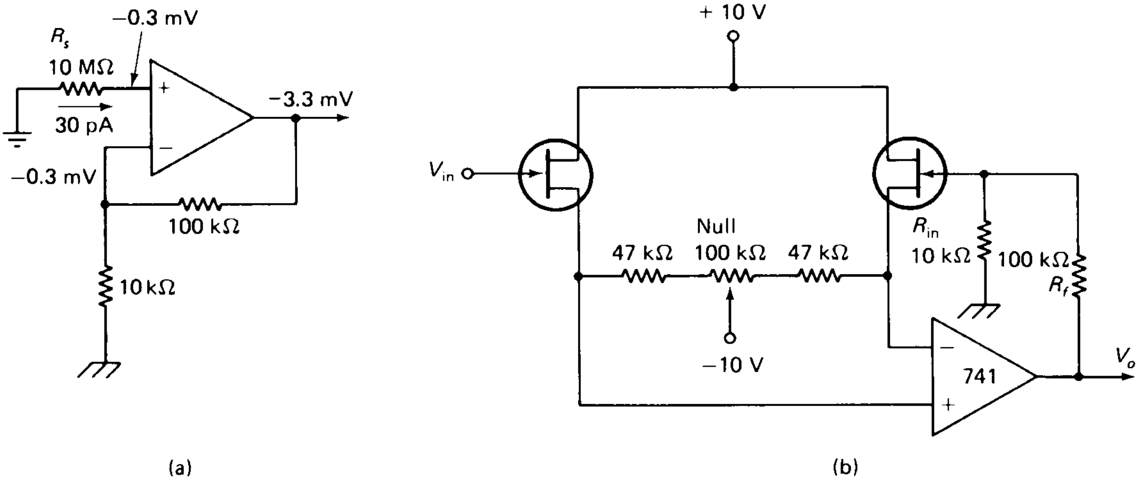


FIGURE 16-8 (a) Bi-FET op amps have extremely low bias currents and can be driven from high-impedance sources without serious offset. (b) FET inputs can be added to a conventional op amp.

Power Output: Most op amps have a total power-dissipation limit of 500 mW and will operate from supply voltages from ±5 to ±18 V. 50-W op amps with 1-A output currents are available, but power output can be realized with a low-power op amp driving a complementary-symmetry output stage, as shown in Fig. 16-9. Notice that the feedback is taken from the transistor outputs, and not from the op amp directly. The first circuit requires the op-amp output to swing to +0.6 and −0.6 V before the *NPN* and *PNP* transistors, respectively, begin to turn on, and crossover distortion will be noticeable at higher frequencies. The added resistors in the second circuit bias the transistors at the threshold of conduction, eliminating crossover distortion. The inverting circuit is shown in the example, but the noninverting circuit can be used as well.

Some op amps are inclined to be unstable if their power-supply lines are not clean. The 10-Ω resistors and 0.1-μF capacitor shown in Fig. 16-9(b) filter the supplies before they reach the op amp, ensuring stability.

Single-Supply Operation: Sometimes it is desirable to utilize an op amp in a system that does not have positive and negative supplies. Figure 16-10 shows how this can be done for noninverting and inverting circuits. Ac coupling is mandatory in these single-supply circuits unless V_{in} is reliably positioned between V_{CC} and ground.

Differential Inputs: Where neither of the input signal wires is grounded, or where it is desired to amplify the difference between two signals (both referenced to ground), or where two signals are to be subtracted, the differential op-amp circuit of Fig. 16-11(a) may be used. Sample resistance values and input voltages are given

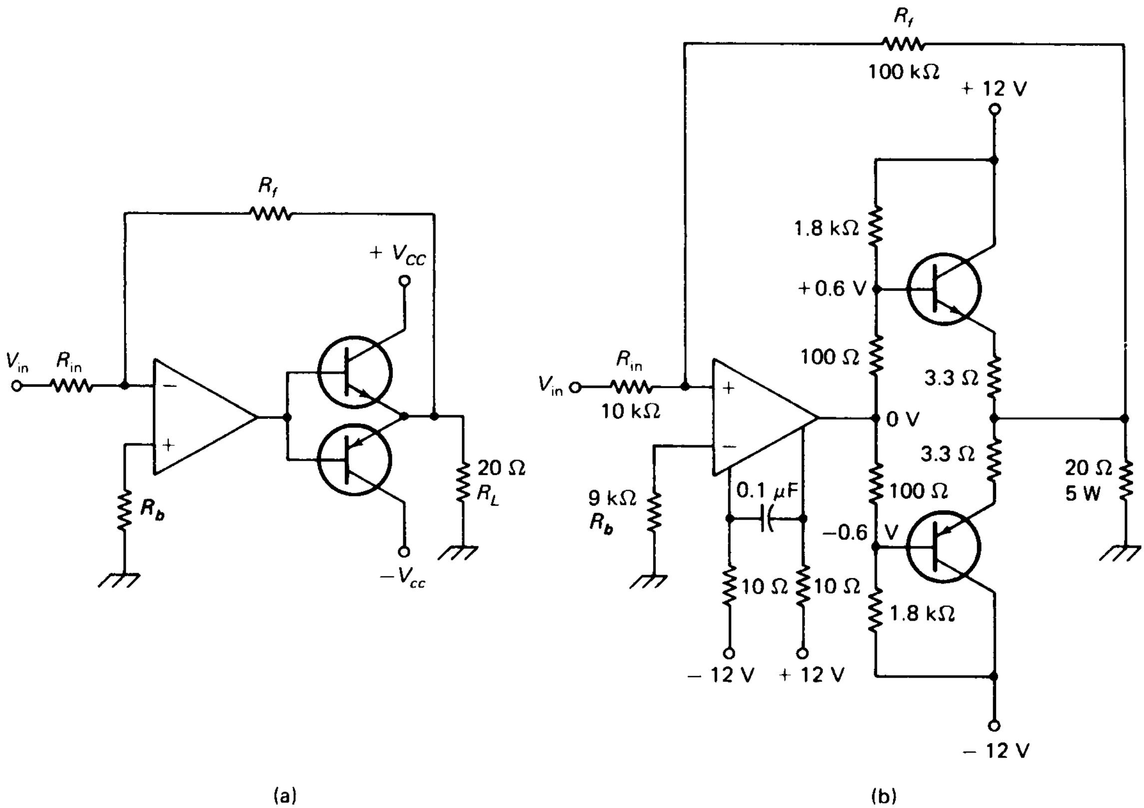


***FIGURE 16-9** Complementary-symmetry power-output stages for op amps: (a) simple; (b) biased and supply-decoupled for minimum high-frequency distortion.*

as an aid to understanding the circuit operation. The noninverting R_{in} and R_f resistors form a simple voltage divider, since the + op-amp input has very high Z_{in}. The rest of the voltages follow directly from the fact that the voltage between the + and − inputs is virtually zero.

The common-mode rejection of this dif amp depends upon the degree of matching of the two R_f/R_{in} ratios, so precision resistors or a trimmer in series with the grounded R_f should be used if high CMRR is desired. The common-mode input voltage should be limited to about half the V_{CC} supplies with this circuit.

The impedance at the inverting input is lower than at the noninverting input, and signals at the noninverting input will cause currents in the inverting input line. It is therefore advisable to drive the circuit of Fig. 16-11(a) through a pair of voltage followers [Fig. 16-2(b)] unless the source impedance to ground is several hundred times lower than R_{in}.

Figure 16-11(b) shows a differential op-amp circuit with identical Z_{in} at each input and a common-mode voltage range of almost 10 times the V_{CC} supplies.

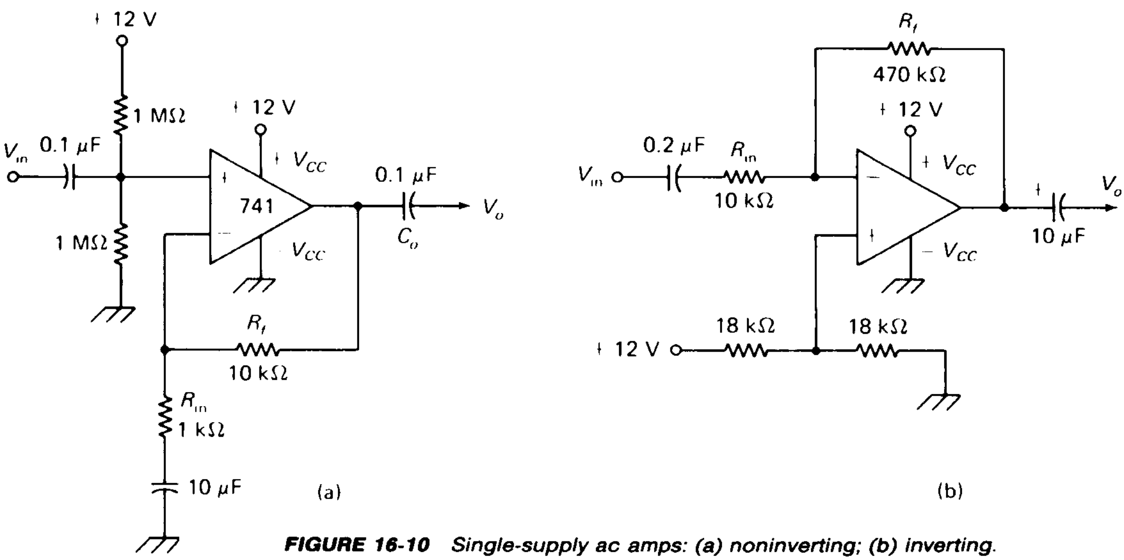
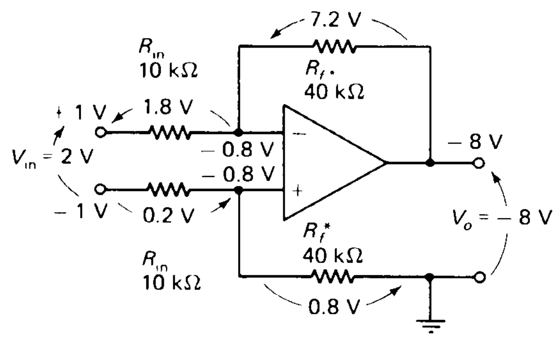


***FIGURE 16-10** Single-supply ac amps: (a) noninverting; (b) inverting.*



(a)

$$A_v = \frac{V_o}{V_{in}} = \frac{R_f}{R_{in}}$$

Z_{in} (between inputs) $= 2R_{in}$

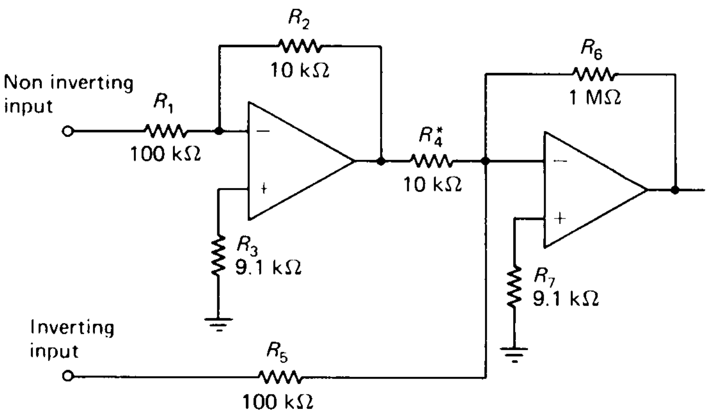
Z_{in} (noninverting input to ground)

$= R_{in} + R_f$

Z_{in} (inverting input to ground)

$= R_{in}$ modified by currents caused by noninverting input signal

*Trim for best CMRR



$$A_v = \frac{R_6}{R_5}$$

Z_{in} (between inputs) $= 2R_1$

Z_{in} (either input to ground) $= R_1$

$R_5 = R_1$

$R_2 = R_4 = \frac{1}{10} R_1$

*Trim for best CMRR.

(b)

***FIGURE 16-11** (a) Basic differential amp. Unequal input impedances, input interaction, and common-mode voltage limit may be troublesome. (b) Improved dif amp.*

17

TRANSMISSION LINES AND ATTENUATORS

17.1 IMPEDANCE MATCHING

When a signal source with a source resistance R_s is used to drive a load R_L, it can be shown that maximum power is delivered to the load when $R_L = R_s$. This was illustrated in Fig. 5-4. Lower values of R_L increase load current, but decrease load voltage, resulting in less load power. Higher values of R_L increase load voltage, but decrease load current, also resulting in less load power. This is one reason for matching the speaker impedance (4, 8, or 16 Ω) to the amplifier output impedance in an audio system. The effect should not be overestimated, however. Notice from Fig. 5-4 that a 2 : 1 mismatch causes only an 11% power loss.

Several other factors concerning impedance matching should be kept in mind:

- At match, the power wasted in the source resistance equals that delivered to the load. In drawing power from a battery we would prefer not to match but to keep R_L much greater than R_S to conserve energy. Mismatching with R_L too high is often preferable to having R_L too low for this reason.
- In driving a bipolar transistor amplifier, the input *current* is what operates the transistor. Designing an amplifier for 10 kΩ Z_{in} because it is to be driven by a 10-kΩ R_s microphone is therefore an error. Z_{in} should be kept low to maximize input current.

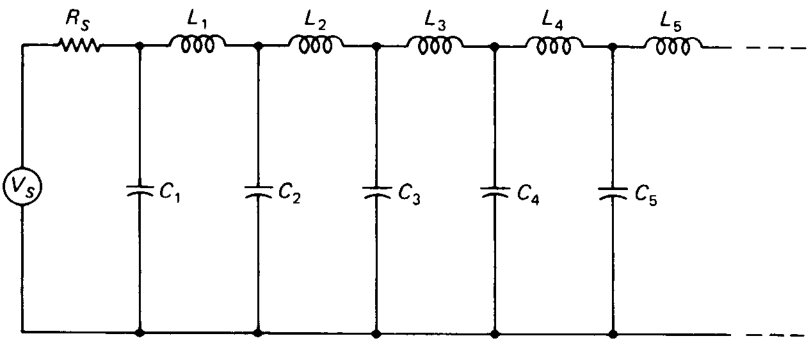


FIGURE 17-1 A transmission line can be represented as a series of infinitely small capacitors and inductors.

- In driving an FET or vacuum tube, the input *voltage* operates the amplifier, and Z_{in} should be kept many times higher than R_s to maximize V_{in}. Attempts to maximize P_{in} are misdirected.

17.2 TRANSMISSION LINES AND REFLECTIONS

A more compelling reason for matching R_L to R_s appears when high-frequency signals are sent over long transmission lines (i.e., when the line is longer than about $\frac{1}{10}$ wavelength at the frequency in use). The lengths involved are about 30 km (20 mi) at audio frequencies, 30 m (100 ft) at 1 MHz, and 0.3 m (1 ft) at 100 MHz. To understand the effect we must first look closely at a transmission line.

Characteristic Impedance of a Line: Figure 17-1 represents an infinitely long transmission line. The stray capacitance from the center conductor to the shield is represented by the series of small capacitors C. The inductance of the center conductor is represented by the series of coils L.

If a dc voltage is suddenly applied by V_S, C_1 will initially accept a charging current through R_S, and L_1 will initially block current from the rest of the line. As C_1 finishes charging, L_1 will begin to pass current to C_2 and L_2 will block current from the rest of the line. This process continues indefinitely, with V_S supplying a constant current I_S, the energy being used to charge the infinite series of L and C components. The effect (a constant voltage producing a constant current) is exactly the same as if the infinite line were a resistance, $R_o = V_S/I_S$.

The value of this effective resistance R_o depends upon the physical structure of the line. It is lower for lines with thick conductors closely spaced, because this makes the C values larger, increasing I_S. Thin conductors increase L, tending to decrease I_S and raise R_o. R_o is termed the *surge resistance* of the line, and is the same as Z_o, the characteristic impedance of the line. Z_o can be shown by network analysis of the reactance of the L and C components to be given by the equation

$$Z_o = \sqrt{\frac{L}{C}} \tag{17-1}$$

where L and C are the inductance and capacitance, respectively, per unit length of line. Note that Z_o is purely resistive for an infinite line.

EXAMPLE 17-1

What is the characteristic impedance of a line having an inductance of 0.2 μH/m and a capacitance of 50 pF/m?

Solution

$$Z_o = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.2 \times 10^{-6}}{50 \times 10^{-12}}} = 63\ \Omega$$

Terminated Lines: Infinite lines, of course, do not exist, but if a long section of line is terminated by a resistance equal to Z_o, the line behaves as if it were infinitely long. This is true because an infinitely long line looks like a resistance of value R_o. The resistance, absorbing energy and dissipating it as heat, cannot be distinguished (from the vantage point of the source) from an infinite line storing the energy in its infinite series of L and C components.

Reflections from an Open Line: If a short pulse is applied to a finite transmission line that is *unterminated* at the far end, the incident (outgoing) energy will have no place to go—no resistance to dissipate it, and no further LC components to pass it on to. The end capacitance of the line will then simply discharge itself back through the inductance that charged it, reflecting the pulse from the end of the line back to the source.

This is a most interesting experiment that can actually be performed if a fairly fast scope and pulse generator are available. The setup is shown in Fig. 17-2. For a 10-MHz scope the line should be about 100 m (300 ft) long, and the pulse generator should be set for 0.1-μs-wide pulses at about 200 kHz repetition rate. For a faster scope the line can be shortened and the pulses speeded up proportionally if desired. The impedance of the generator must equal the Z_o of the line used, or the two ends of the line will play ping-pong with the signal, making the first reflection difficult to see. Since most pulse generators have $R_s = 50\ \Omega$, this means adding 250 Ω in series with the pulse generator for 300-Ω TV twin lead, or 25 Ω for 75-Ω coax cable.

The oscilloscope will display the incident pulse with the reflected pulse following by a time calculated as

$$t = \frac{2d}{v_p} \tag{17-2}$$

where d is the length of the line (the factor of 2 indicates that the pulse travels out and back) and v_p is the propagation velocity of a signal on the line. This is typically

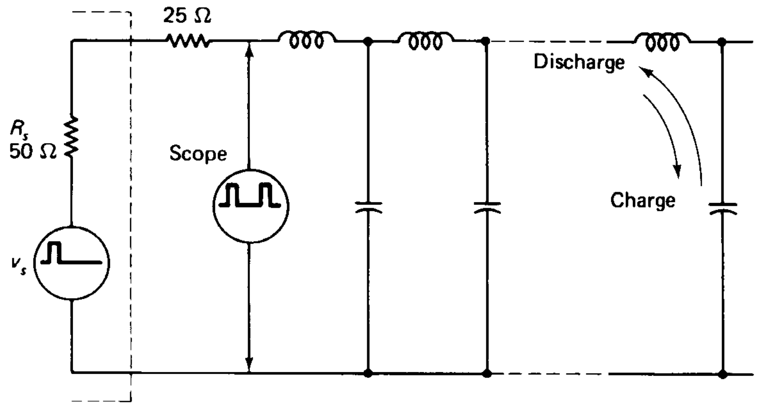


FIGURE 17-2 An open-end transmission line reflects an upright signal back to the source.

200×10^6 m/s for 50- and 75-Ω coax and 250×10^6 m/s for 300-Ω twin lead. Special open-wire lines with ceramic spacers may have v_p of 290×10^6 m/s.

Once v_p for a given type of cable is known, a measurement of t gives distance d to a break in the cable directly from equation 17-2. This technique, called *time-domain reflectometry* (TDR), is of tremendous value in locating faults in underground and undersea cables.

Reflections from a Shorted Line: Figure 17-3 shows a pulse applied to a transmission line which is shorted at the far end. In this case the inductance at the end of the line is caught with no place to send its stored energy, so it discharges into the line that fed it. This inductive kickback is opposite in polarity from the incident pulse, so the reflected pulse is inverted.

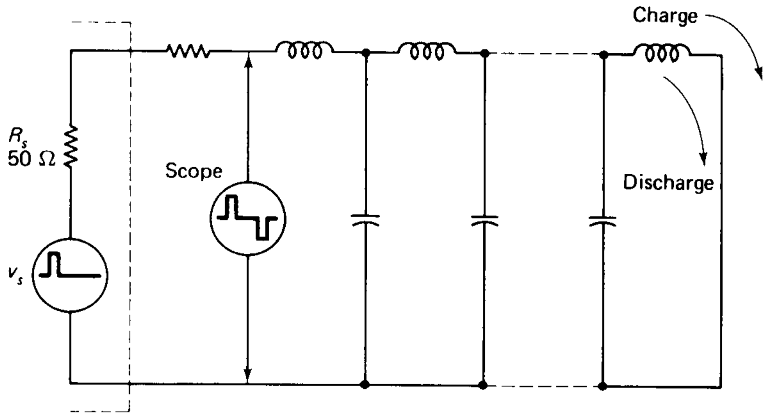


FIGURE 17-3 A shorted-end transmission line reflects an inverted signal back to the source.

The amplitude of the reflected pulse will equal that of the incident pulse if the line is lossless and the termination is a perfect short (or open) circuit. Any termination that is not equal to Z_o will produce some reflections from the termination point which become larger in amplitude as the termination resistance strays away from Z_o and nearer a short or open circuit. In particular, the fraction of voltage reflected is given by a *reflection coefficient* Γ (gamma):

$$\Gamma = \frac{V_r}{V_i} = \frac{R_L - Z_o}{R_L + Z_o} \tag{17-3}$$

where V_r is reflected voltage, V_i is incident voltage, R_L is the load or terminating resistance, and Z_o is the characteristic impedance of the line.

It might be wise to emphasize here that the discrete or *lumped* L and C components represented above are actually *distributed* along the line. The actual representation of even a finite length of line would have to contain an infinitely large number of infinitely small L and C components. Analysis of the LC network on a lumped value per meter (or per foot) would indicate pulse-shape distortions and high-frequency limits which are not present in the real distributed-component line.

17.3 STANDING WAVES

Let's watch a movie of a 100-V-pk sine wave being applied at the start of the positive half-cycle to a line that is one wavelength long. The line is terminated in a resistance higher than Z_o. Let us say that $Z_o = 100\ \Omega$ and $R_L = 300\ \Omega$, so that 50% of V_i is reflected (noninverted) as V_r. A frame of the movie is shown each quarter-cycle in Fig. 17-4. The first four frames show the positive wavefront approaching the end of the line (right) as the source (left) goes through a full + then − cycle and gets ready to go + again.

The fifth frame ($1\frac{1}{4}$-cycle) shows the half-amplitude positive reflected wavefront (dashed line) headed back from the end of the line. The subsequent frames show the incident wave advancing to the right a quarter-wavelength each frame, while the reflected wave advances $\frac{1}{4}$ to the left each frame. The heavy line shows the sum of V_i and V_r, which is the voltage that would actually be observed at each point on the line at each instant pictured.

Consider that you are monitoring the voltage at the $\frac{1}{4}\lambda$ point on the line. As the cycle progresses you will measure 0 V, +100 V, 0 V, −100 V, and of course, all the values that the sine wave assumes between these quarter-cycle points. The same is true of any other point on the line *in the absence of reflections*.

Once the reflected voltage begins combining with the incident voltage, however, the waveforms observed at various points on the line are not identical. At the $\frac{3}{4}\lambda$ point, for example, the voltage passes from 0 V at $1\frac{1}{4}$ cycles to −50 V, to 0 V, to +50 V. The reflected wave always subtracts from the incident wave, producing a peak voltage of only 50 V.

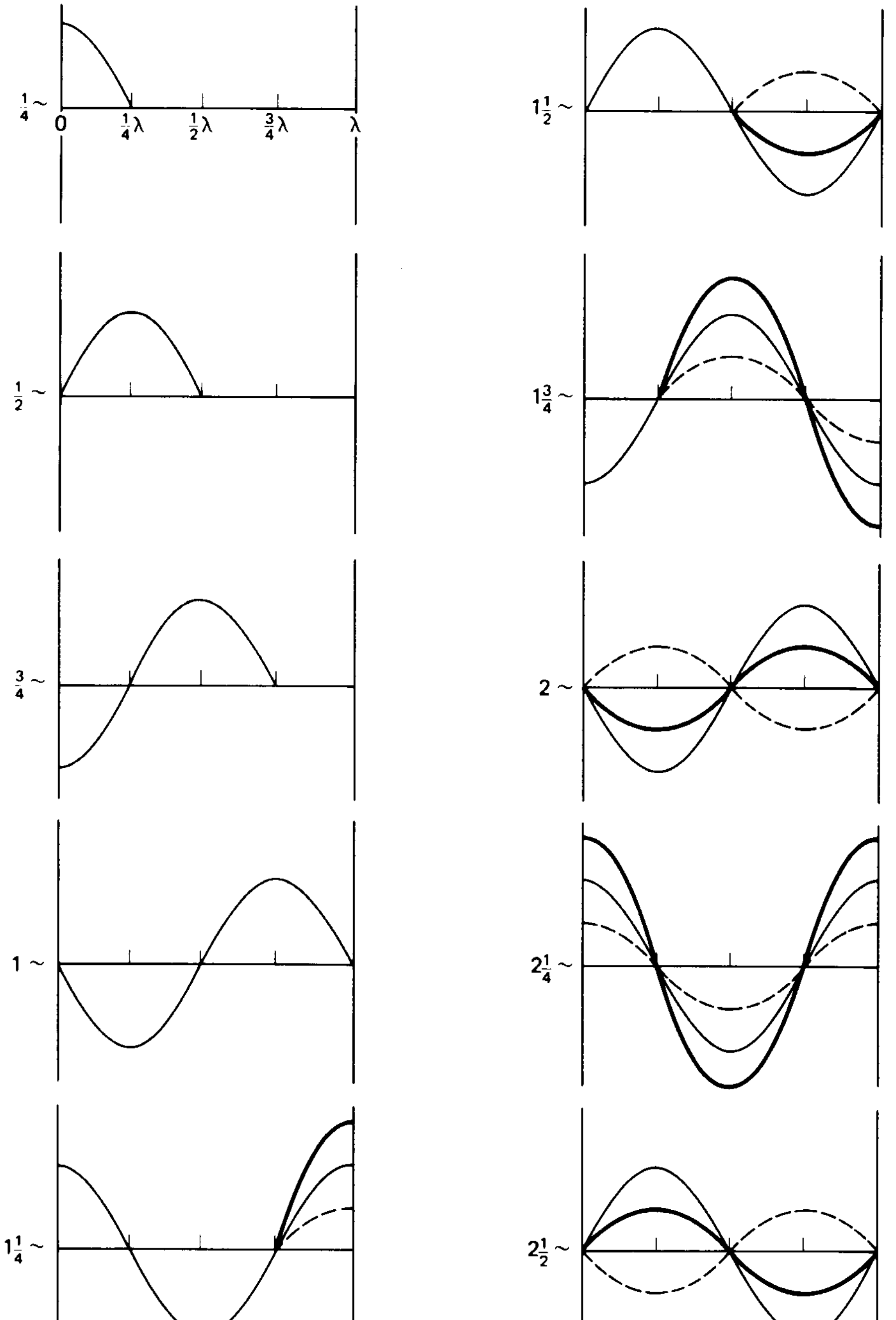


FIGURE 17-4 *A sine wave generated at 0λ heads out a 1λ line (solid line). At the end a mismatch reflects 50% of it back toward the source (dashed line). The combination of incident and reflected wave (heavy line) has larger and smaller peak values at various points along the line, which are called standing waves.*

At the $\frac{1}{2}\lambda$ point the reflected wave always adds to the incident wave, yielding a 150-V peak. At $\frac{1}{4}\lambda$ the wave is 50 V pk, and at 0 λ it is back to 150 V pk again. These voltage *nodes* and *crests* always appear at multiples of $\frac{1}{4}\lambda$ from the point of reflection. They are stationary, or appear to stand in one place, hence the term *standing waves*.

The ratio between the voltages at the crest and node is called the standing-wave ratio (SWR) or voltage standing-wave ratio (VSWR). Current standing-wave ratio ISWR is numerically identical. In our example, the SWR is 150 V/50 V, or 3 : 1.

The current on a line is a maximum at the point where the voltage is minimum, and vice versa. In this case $I_L = V_{RL}/R_L = 150\ \text{V}/300\ \Omega = 0.5$ A. The ratio of currents is the same as the ratio of voltages so the current peak is $0.5\ \text{A} \times 3 = 1.5$ A. Interestingly enough, SWR can be calculated as simply the ratio of impedance mismatch:

$$\text{SWR} = \frac{Z_o}{R_L} \qquad \text{or} \qquad \text{SWR} = \frac{R_L}{Z_o} \tag{17-4}$$

It must be emphasized that this entire discussion has been based on resistive—not reactive—loads. The phase shifts and impedance transformations that result when the load impedance is partly reactive can be predicted with the aid of a great little toy called the Smith chart, but we haven't room to go into that here.

17.4 WHAT'S WRONG WITH STANDING WAVES?

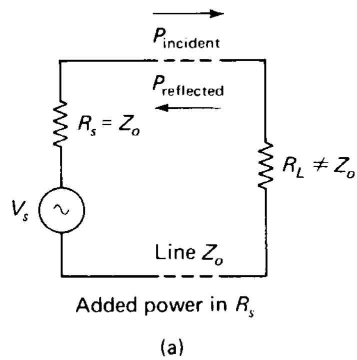
We started the investigation of reflections and standing waves with warnings about "compelling reasons" for avoiding impedance mismatch. We now list some of the reasons.

Smear and Ghosting in TV Pictures: If a TV antenna line is longer than about 24 m (80 ft), reflections on the line will cause a second picture $\frac{1}{4}\ \mu$s or more later than the first. This begins to be a noticeable fraction of the 63-μs horizontal sweep time, so the second picture produces a smear along the edge of the picture details. Where the lines are 10 times this long the second picture will be displaced 1 or 2 cm to the right of the first—the familiar TV *ghost*. (Ghosts on TVs with short antenna lead-ins are caused by reflections of the radiated signal from buildings or other objects along the transmission path, not by line reflections.)

Reflected Power is lost to the load, but is *added* to the signal source. Consider the following example:

EXAMPLE 17-2

A source delivers 100 V rms to a 100-Ω line terminated in 100 Ω. How much power is delivered to the load? The load changes to 300 Ω. How much power is delivered, and what is the effect upon the source dissipation?



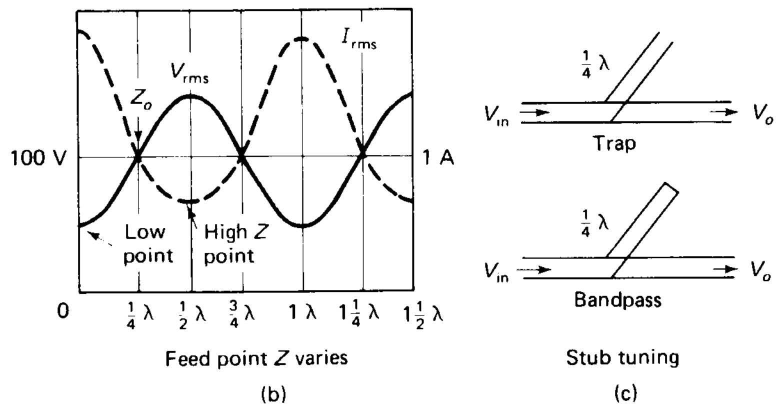


***FIGURE 17-5** (a) Standing waves mean reflected power which is dissipated in the transmitter. (b) Standing waves cause feed-point impedance to vary with length of line. (c) Shorted or open stubs of line can be used as tuned circuits or traps.*

Solution

In the matched case, where SWR = 1 : 1, the line is said to be *flat* and load power is calculated directly:

$$P_{RL} = \frac{V^2}{R_L} = \frac{100^2}{100} = 100 \text{ W}$$

If R_L is changed to 300 Ω, reflection coefficient is

$$\Gamma = \frac{R_L - Z_o}{R_L + Z_o} = \frac{300 - 100}{300 + 100} = 0.5$$

The load voltage is the 100 V incident plus the 50 V reflected:

$$P_{RL} = \frac{V^2}{R_L} = \frac{150^2}{300} = 75 \text{ W}$$

The reflected voltage produces an added power in the source which depends upon the source impedance and whether it is reactive or purely resistive, and upon the length of the feedline which affects the relative phase of the incident and reflected voltages across the source resistance.

Radio transmitters using class C output stages are often designed to dissipate only 30 W while outputting 100 W, so any significant added power due to reflection back to the source could prove disastrous.

Feed-Point Impedance Varies with the length of the transmission line if there is a mismatch at the end. In Fig. 17-4, for example, the impedance seen by the source driving the line at 0λ is $R = V/I = 150\ \text{V}/0.5\ \text{A} = 300\ \Omega$. This is equal to the load R_L, and the same impedance would be seen if the first $\frac{1}{2}\lambda$ of the line were cut off and the remainder driven from the $\frac{1}{2}\lambda$ point. In general, the impedance $\frac{1}{2}\lambda$ or any multiple thereof from a load R_L equals R_L, no matter what the Z_o of the line.

In the same figure the impedance seen if the line were driven at the $\frac{1}{4}\lambda$ or $\frac{3}{4}\lambda$ point would be $R = V/I = 50\ \text{V}/1.5\ \text{A} = 33\ \Omega$. In general, the impedance $\frac{1}{4}\lambda$ or any odd multiple thereof from a load R_L equals Z_o^2/R_L.

The range of impedances that might be presented to the driving source, just by pruning the transmission line is, in this case, 9 : 1. In general, the maximum and minimum possible impedances presented by a transmission line are in the ratio of SWR^2.

This is not all bad. Intelligently cut lines make very efficient and low-cost impedance transformers. Quarter-wavelength stubs of line left open at the end present zero impedance at the frequency for which they are $\frac{1}{4}\lambda$, making very effective null or trap filters. In a similar manner, quarter-wave stubs with shorted ends or half-wave stubs with open ends make excellent tuned circuits. More often than not, however, we would prefer to keep the transmission line flat (1 : 1 SWR), so its length can be varied to suit the physical requirements without worry about electrical consequences.

Voltage and Current Peaks on a high-SWR line can waste power due to excessive I^2R loss and overheat the cable in high-power applications. Dielectric breakdown often appears first at a voltage crest, caused by a higher-than-expected SWR.

17.5 THE L PAD

The L pad is a simple series-resistor voltage divider, so named because it can be (but usually is not) drawn in the shape of a letter L. If one has plenty of signal to spare, an L pad can be used to stabilize the output impedance of an unruly signal generator, as shown in Fig. 17-6(a). For the values given in the figure, Z_o varies only from 48.6 to 48.9 Ω, even though the actual generator impedance varies from 20 to 200 Ω. The price paid for this stability is an output reduction from 10 V to about 0.4 V.

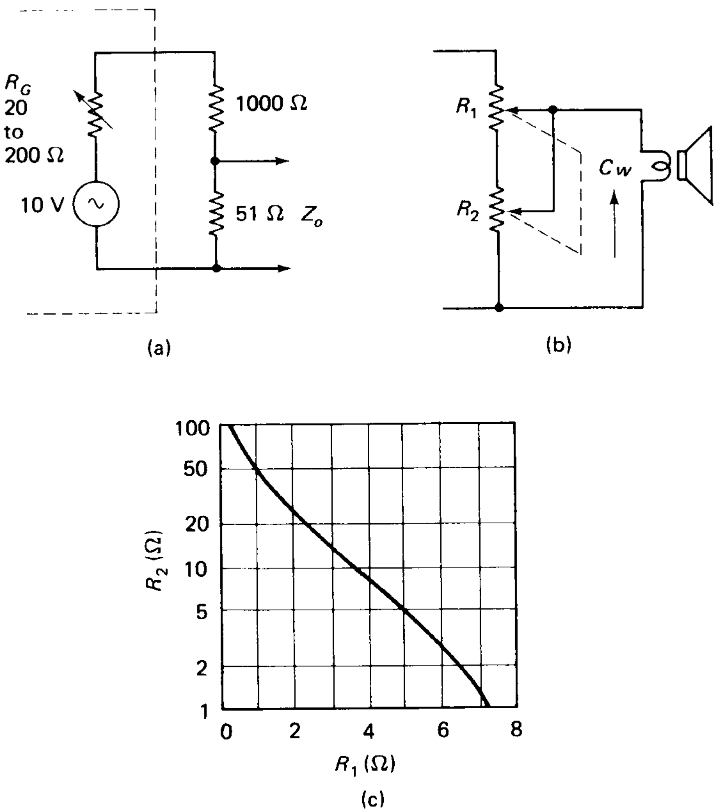


FIGURE 17-6 *(a) An L pad makes the generator source impedance Z_o relatively constant even though R_G varies. (b) Ganged linear and log pots form an L pad which maintains a constant R_{in} while adjusting speaker volume. (c) Chart showing required values of R_1 and R_2 in (b).*

L pads are often used as remote-speaker volume controls in an attempt to keep the total impedance constant. The objective is usually to keep an adjustment of one speaker from changing the volume at another speaker elsewhere in the system. The arrangement for an 8-Ω speaker consists of a linear 8-Ω pot R_1 ganged with a log-taper 50-Ω pot R_2, as shown in Fig. 17-6(b). The required values of R_1 and R_2, shown in Fig. 17-6(c), seldom track exactly, but perfection is not required.

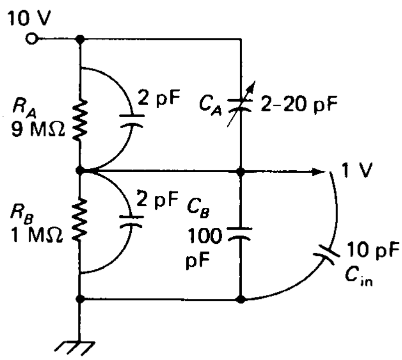
Frequency-Compensated Attenuator: Voltage-divider attenuators are very widely used to select the range of all types of voltmeters and oscilloscopes. These instruments have input resistances standardized at 1 MΩ or 10 MΩ. Each resistor in the voltage divider can be expected to have a stray capacitance of at least 2 pF across it, due to its own structure, wiring, and switch capacitance. Two picofarads has a reactance of 10 MΩ at 8 kHz, making a 10-MΩ resistive voltage divider completely useless at that frequency. Even at 1 kHz the error caused by the capacitance could approach 1%.

High-impedance voltage dividers, even for moderate frequencies, are always compensated with trimmer capacitors, as shown in Fig. 17-7(a). The ratio of resistances R_A/R_B is equal to the ratio of capacitive reactances, X_{CA}/X_{CB}. Another way of expressing this relationship is

$$R_A C_A = R_B C_B = R_T C_T \tag{17-5}$$

The input capacitance of the instrument (10 pF) and the stray capacitances associated with R_A and R_B simply appear in parallel with and become part of C_A and C_B. The input resistance of the instrument is usually infinite since the input amp contains an FET.

An oscilloscope with ranges from 5 mV/division to 20 V/division in a 1–2–5 sequence would require eleven attenuators of the type shown in Fig. 17-7(a). Furthermore, there would be no assurance that the input capacitance would be the



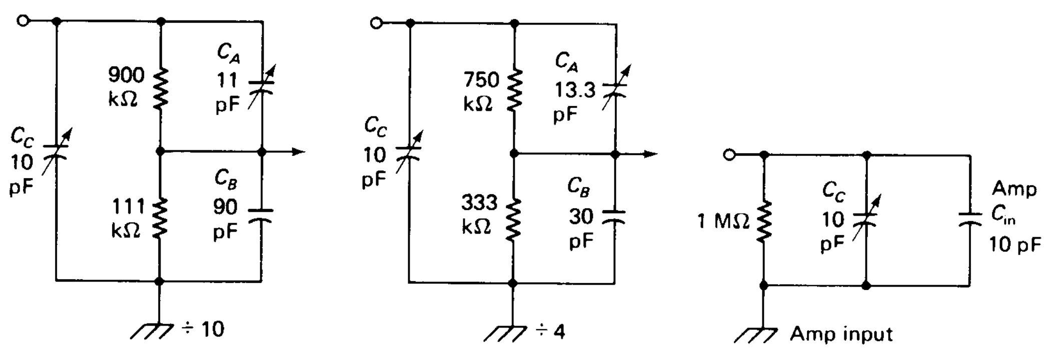


***FIGURE 17-7** (a) ÷10 attenuator with compensation for stray capacitance. (b) Two attenuator sections and an amplifier input designed for mixing and cascading.*

same on each range, since each C_A would be adjusted to compensate for variations in C_B and various stray capacitances. In practice, four standard input attenuators are used for this job. Figure 17-7(b) gives an abbreviated illustration of how the attenuators are used in cascade.

Each attenuator is designed to present a 1 MΩ‖20 pF load to whatever drives it, and in turn to drive a 1 MΩ‖20 pF load. Capacitors C_C are added to each attenuator and the input amp to bring the capacitance up to 20 pF. Resistors R_B are sized to make the required voltage division when shunted by the 1-MΩ load of the following attenuator. Table 17-1 shows the attenuators used on each range.

TABLE 17-1

	mV				V							
Range	5	10	20	50	0.1	0.2	0.5	1	2	5	10	20
	—	2	4	10	10	10	100	100	100	100	100	100
Attenuators					2	4		2	4	10	10	10
											2	4

External low-capacitance probes are often used on oscilloscopes to reduce the loading that appears at high frequencies when the entire input capacitance of the scope (20 to 50 pF) plus the shielded cable (20 to 60 pF/m) is connected across the circuit under test. Figure 17-8 shows a typical ×10-attenuating-probe circuit. The probe reduces the capacitive loading from a 60-pF shunt to 6.7 pF in series with 60 pF or a 6-pF shunt. The resistive loading is reduced from a 1-MΩ shunt to a 10-MΩ shunt. The price for this is reduced sensitivity. The 5-mV/division scale becomes the 50-mV/division scale using the probe.

The cable used in probe applications has a very thin center conductor to minimize capacitance. It therefore has a characteristic impedance of several hundred ohms, but the 1-MΩ termination presented by the scope is still an open circuit by

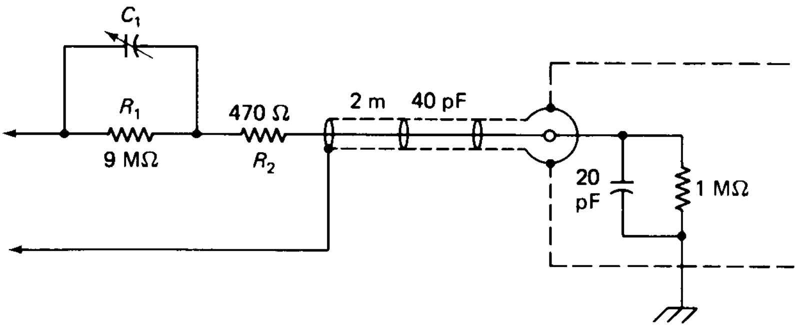


FIGURE 17-8 A × 10 oscilloscope probe. The signal is divided by 10; the scope scale factor is multiplied by 10.

comparison. Nor does R_1 present a proper termination for the reflected signal on its way back. A series of reflections back and forth along the cable must therefore be expected. R_2 is chosen to dissipate the energy of the reflections, although it is negligible in the voltage-division process. On a 2-m line the reflections will be about 15 ns apart. A 10-MHz scope has a rise time of 0.35/10 MHz = 35 ns, so the first several reflections would not be noticed. Very fast scopes require more sophisticated probes and signal-acquisition techniques.

17.6 SYMMETRICAL ATTENUATORS

Simple resistive networks can be designed which voltage-divide an input signal by a desired ratio while presenting a resistance equal to Z_o when viewed from either the source or the load. This is assuming that $R_s = Z_o$ and $R_L = Z_o$. These devices have several advantages over the L pad.

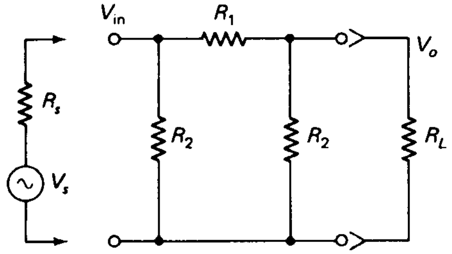
- The input and output are interchangeable. They can be inserted in the line either way, and signals sent over the line in either direction will experience the same loss.
- They preserve a constant line impedance, thus eliminating the problems associated with reflected signals.
- They tend to "cover up" for faults in Z_s and/or Z_o. A shorted or open load seen through a symmetrical attenuator will look like a resistance rather near Z_o regardless of the load fault. This effect is called *padding* and is more pronounced for high-attenuation pads.

Remember, however, that attenuators are essentially power wasters—they are generally used only in low-level applications where the signal source is stronger than required.

The Pi Pad, with its design and analysis equations, is shown in Fig. 17-9. It is essentially a voltage divider shunted by a resistor to keep the input resistance down to Z_o. R_L is shunted directly across the output of the pad and forms part of the voltage-divider circuit. It is assumed that $R_L = Z_o$, and other values of R_L will upset the attenuation ratio. This problem is less severe for higher attenuation factors.

The T Pad appears with relevant equations in Fig. 17-10. It is simply an alternative to the pi pad and provides identical performance.

Padding to preserve line impedance is illustrated in Fig. 17-11. If a break should occur in the line between tap-off boxes B and C, the remaining stub of line would send unwanted reflections back to boxes A and B and may form a tuned stub



Definition:

$$a = \frac{1}{A} = \frac{V_{in}}{V_o}$$

$$Z_o = R_s = R_L$$

Design:

$$R_1 = Z_o \frac{a^2 - 1}{2a}$$

$$R_2 = Z_o \frac{a + 1}{a - 1}$$

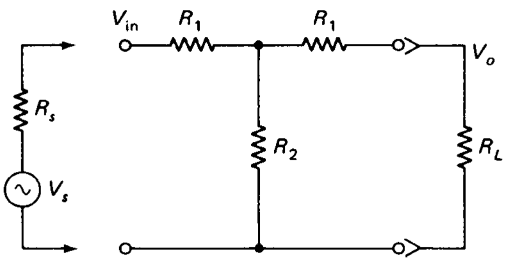
Analysis:

$$Z_o = \frac{R_2}{\sqrt{\dfrac{R_1 + 2R_2}{R_1}}}$$

$$a = \frac{R_1 + R_2 \| R_L}{R_2 \| R_L}$$

$$= 1 + \frac{R_1 + \sqrt{R_1^2 + 2R_1R_2}}{R_2}$$

FIGURE 17-9 *π (pi) attenuator with design and analysis formulas.*



Definitions:

$$a = \frac{1}{A} = \frac{V_{in}}{V_o}$$

$$Z_o = R_s = R_L$$

Design:

$$R_1 = Z_o \frac{a - 1}{a + 1}$$

$$R_2 = Z_o \frac{2a}{a^2 - 1}$$

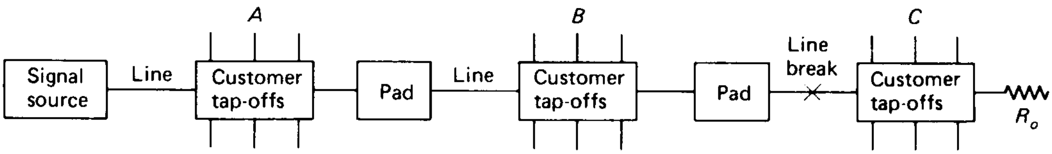
Analysis:

$$Z_o = R_1 \sqrt{\frac{R_1 + 2R_2}{R_1}}$$

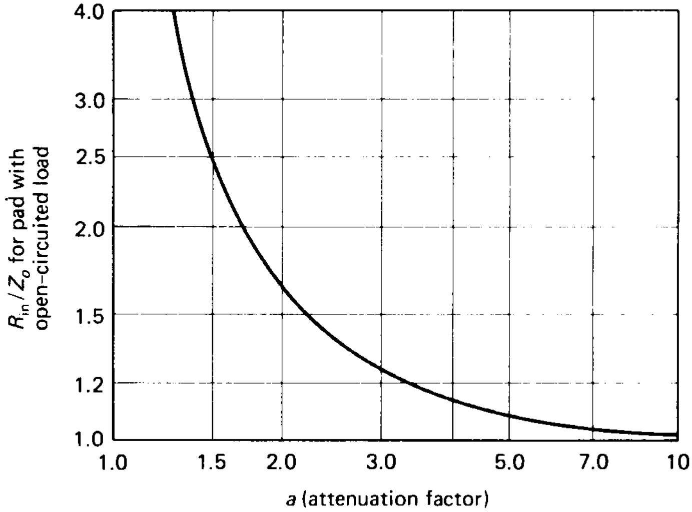
$$a = \frac{R_1 + [(R_1 + R_L) \| R_2]}{(R_1 + R_L) \| R_2}$$

$$= 1 + \frac{R_1 \sqrt{R_1^2 + 2R_1R_2}}{R_2}$$

FIGURE 17-10 *T (tee) attenuator with design and analysis formulas.*



(a)



(b)

***FIGURE 17-11** (a) Padding a line at intervals minimizes reflections from the end in case of a break. (b) Input impedance presented by a pad with an open-circuited load side, as a function of pad attenuation.*

trapping out certain frequencies at boxes A and B. Pads placed at regular intervals along the line will preserve a relatively constant impedance along the intact portion of the line, keeping these customers in good service until the break is repaired. The graph given with the figure shows the impedance seen looking into the pad for an open-circuit load for various values of pad attenuation.

EXAMPLE 17-3

A T pad has $R_1 = 22\,\Omega$ and $R_2 = 100\,\Omega$. Find Z_o and attenuation factor a and tell what R_{line} will be seen looking into the pad if R_L opens up.

Solution

From the T-analysis formulas of Fig. 17-10:

$$Z_o = R_1\sqrt{\frac{R_1 + 2R_2}{R_1}} = 22\sqrt{\frac{22 + 200}{22}} = \mathbf{70\ \Omega}$$

$$a = 1 + \frac{R_1 + \sqrt{R_1^2 + 2R_1R_2}}{R_2} = 1 + \frac{22 + \sqrt{22^2 + (2 \times 22 \times 100)}}{100} = \mathbf{1.92}$$

If R_L is removed, the source sees two resistors of the pad in series:

$$Z_{\text{line}} = R_1 + R_2 = 22 + 100 = \mathbf{122\ \Omega}$$

This is 122/70 or 1.75 times Z_o, and is confirmed by the graph of Fig. 17-11.

Balanced attenuators for use on twin-lead systems are shown in Fig. 17-12, (a) and (b). The circuits of Figs. 17-9 and 17-10 are for unbalanced (coaxial-line) systems. The same equations apply, respectively.

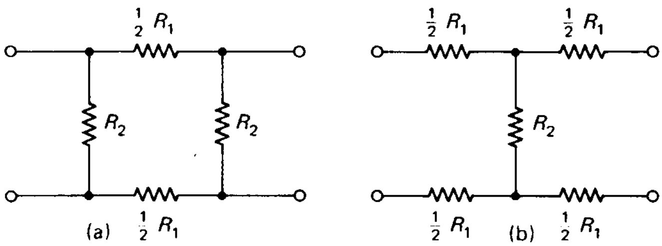


FIGURE 17-12 π and T pads for balanced lines (sometimes called O and H pads, respectively).

Attenuators in cascade, as shown in Fig. 17-13, are often used where a high degree of attenuation ($a > 20$) is required. With the pi design this is so because R_1 values over 10 times Z_o are required for attenuations above 20. The 2 pF or so of stray capacitance around R_1 will thus upset the operation of the pad at high frequencies if too high an attenuation is attempted.

With the T design, $a > 20$ requires values of R_2 less than one-tenth Z_o. For low-Z_o pads (50 Ω) these values may be difficult to obtain and hold accurately.

Pi pads have the advantage for low-impedance work and T pads have the advantage for high-frequency work for these reasons.

Notice that the output R_2 of pad A is simply paralleled with the input R_2 of pad B in cascading pi pads. In cascading T pads, the series equivalent to the two R_1 values is used in a similar manner. Switched attenuators, having pads of 1 dB, 1 dB, 3 dB, 6 dB, 10 dB, 20 dB, and 20 dB are fairly common since they can be placed in or out of the cascade to give attenuations from 0 to 61 dB in 1-dB steps.

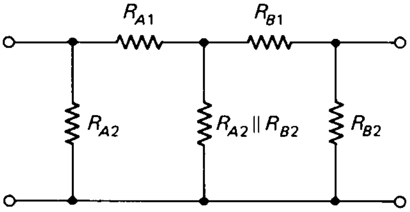


FIGURE 17-13 Cascading pads is sometimes advisable to avoid unreasonably low resistor values, or to keep resistor values lower than stray capacitive reactance.

EXAMPLE 17-4

Find the resistance values required in Fig. 17-13 if the total attenuation required is 36 dB and $Z_o = 600\ \Omega$.

Solution

Each pad will give 18-dB attenuation:

$$a = \log^{-1}\frac{\alpha_{\text{dB}}}{20} = \log^{-1}\frac{18}{20} = 7.9$$

From the pi-pad design formulas:

$$R_{A1} = R_{B1} = Z_o\frac{a^2-1}{2a} = 600\frac{7.9^2-1}{2\times 7.9} = 2345\ \Omega$$

$$R_{A2} = R_{B2} = Z_o\frac{a+1}{a-1} = 600\frac{7.9+1}{7.9-1} = 774\ \Omega$$

$$R_{A2}\|R_{B2} = 387\ \Omega$$

Continuously variable attenuators of the pi or T design would require ganging and odd-curve tracking of three pots. The *bridged-T* attenuator of Fig. 17-14 can be constructed using two conventional log-taper pots ganged together. The relevant equations are given with the figure.

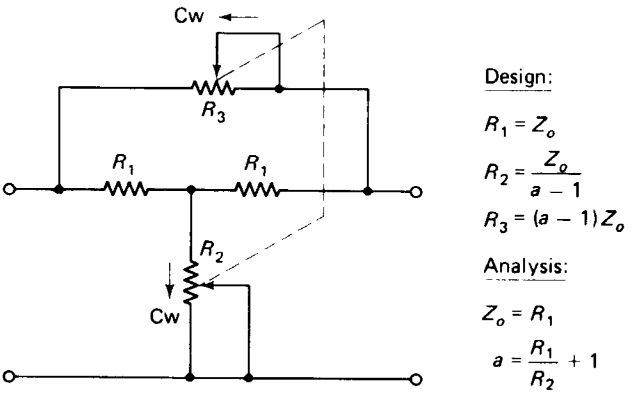


FIGURE 17-14 Bridged-T attenuator with design and analysis formulas.

18

FREQUENCY-SELECTIVE FILTERS

18.1 RC FILTERS AND BODE PLOTS

RC filters occur with great regularity in electronic circuitry, both by design and by accident. By design they are formed from small, light, and inexpensive components with tolerances that can easily be held to 5% over a wide range of temperatures. By accident, they are formed whenever a source resistance drives a load shunted by a stray capacitance, or whenever a resistive load is capacitively coupled to its source.

The Single-Section Low-Pass Filter is shown in Fig. 18-1(a). At low frequencies the reactance of the capacitor is so high that there is negligible drop across the resistor and to a good approximation $V_o = V_{in}$. The phasor diagram of Fig. 18-1(b) shows that even when X_C is only twice R, fully 89% of V_{in} appears at V_o.

The *critical frequency* f_c is the frequency at which $X_C = R$. At this frequency V_o is down by 3 dB (i.e., $V_o = 0.707V_{in}$). Above the critical frequency V_o drops off quite rapidly. Figure 18-1(d) shows $V_o = 0.45V_{in}$ at $2f_c$.

The Bode Diagram (say BO-dee) of Fig. 18-2 summarizes the behavior of an RC filter in two straight lines. It is based on the realization that, at frequencies four or five times f_c, X_c adds little to the total circuit impedance, so the output of the filter is approximately

$$V_o \approx V_{in}\frac{X_C}{R} = V_{in}\frac{1}{2\pi fRC} \qquad [\text{at } f \gg f_c] \tag{18-1}$$

This equation shows that V_o varies in inverse proportion with f above f_c.

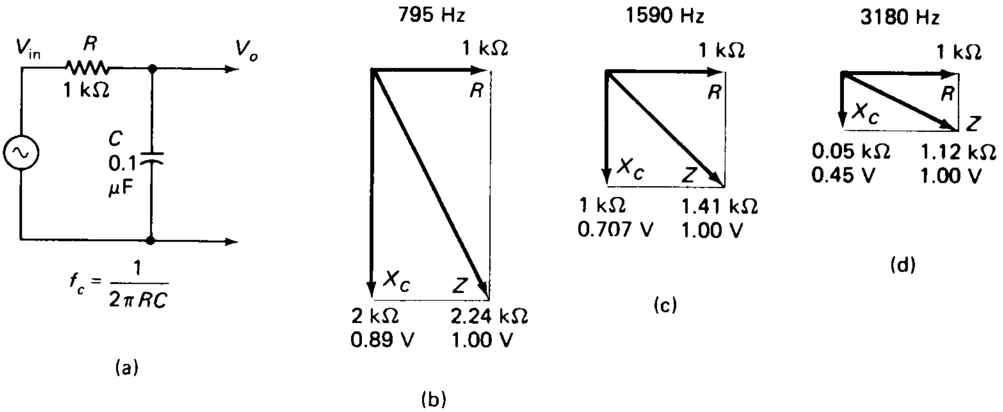


FIGURE 18-1 *Elementary phasor analysis can be used to predict the output of a simple* RC *filter at various frequencies.*

We speak of the filter having a *rolloff of 20 dB per decade*, which is to say that V_o drops by a factor of 10 for each increase in f by a factor of 10. Sometimes this same fact is expressed as a *6-dB per-octave rolloff*, which is a factor-of-2 voltage drop for a factor-of-2 frequency rise (in music an octave—eight-note scale—covers a frequency span of 2 : 1).

To draw the Bode plot for an RC filter it is only necessary to determine the values of R and C, calculate the critical frequency f_c (this is often called the *break*

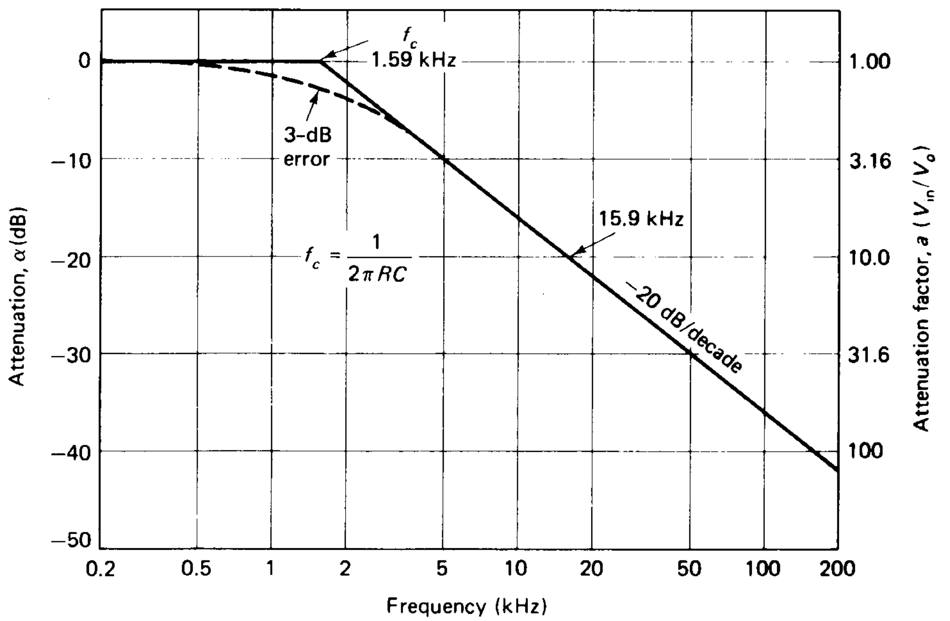


FIGURE 18-2 *The Bode plot summarizes the response of an* RC *filter: 20-dB/decade rolloff, 3 dB down at the break frequency.*

point), and draw a straight line sloping down at 20 dB/decade from f_c. Notice that the graph paper is log-log if the attenuation factor a is being graphed. If α is being expressed (in dB), semilog paper is used. In both cases the frequency must be plotted on a log scale for the straight-line rolloff to be accurate. As a final correction, we draw in a smoothly rounded "actual" curve passing through f_c at -3 dB below the break point to correct for errors in the straight-line approximation curve around f_c.

The Single-Section High-Pass Filter is very similar to the low pass. The calculation of f_c, the 20-dB/decade roll-off, and the -3-dB correction at f_c are all the same. Here the roll-off is to the left rather than the right, however. Figure 18-3 shows the circuit and its Bode plot.

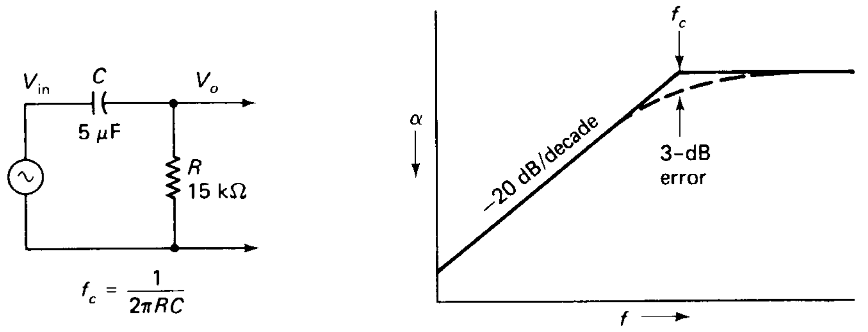


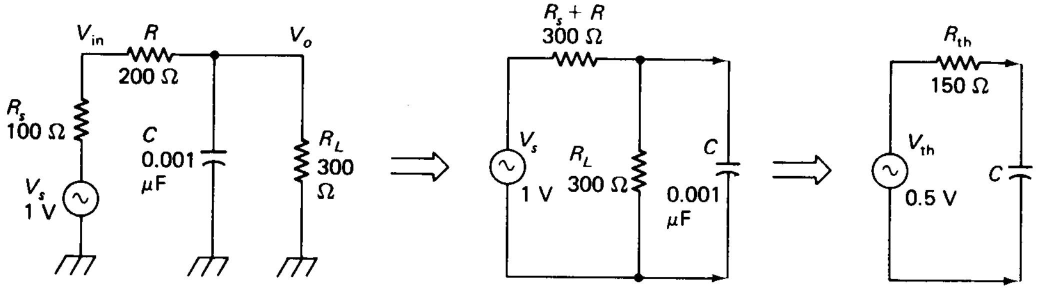
FIGURE 18-3 RC high-pass filter, with Bode plot rolling off at the low end.

Load and Source Resistances are unavoidable parts of most real circuits, but they can be reduced to the simple two-component *RC* circuits of the previous examples without difficulty. Figure 18-4 shows how the load and source resistors can be combined with R so that a calculation of f_c can be made. To determine the output at the horizontal portion of the curve, assume the capacitor to be an open circuit for a low-pass (or short circuit for a high-pass) filter and compute V_o using simple resistive voltage division. Notice that the low-pass filter requires an application of Thévenin's theorem to combine R_L with the other two resistors.

Occasionally, a capacitor is used to drive a load which itself has appreciable shunt capacitance, as in Fig. 18-5. In this case the capacitors are Thévenized, as illustrated by the attendant calculations. It must be emphasized that this technique is valid only if R_s is many times (100 or more) less than R_L, and then is good only up to the frequency where the two capacitors begin to load R_s. Above f_{max}, the circuit becomes a low-pass filter.

18.2 MULTIPLE-SECTION RC FILTERS

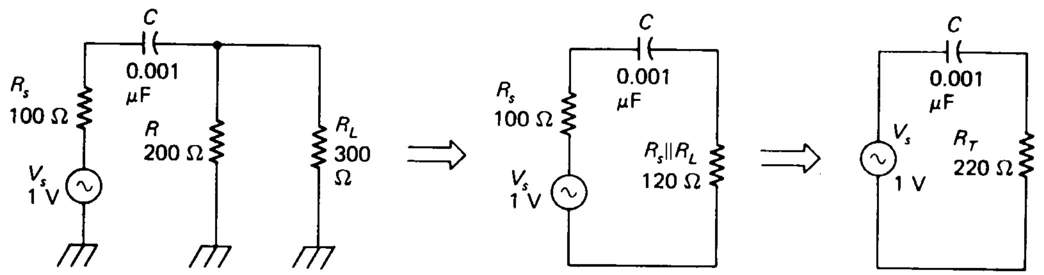
Cascaded *RC* Sections can be used to obtain a steeper rolloff—40 dB/decade for two sections with identical cutoff frequencies, 60 dB/decade for three sections, and so on. The correction when rounding in at the break point is -6 dB for two



$$V_o\,(\text{max}) = V_s \frac{R_L}{R_s + R + R_L} = 1 \frac{300}{100 + 200 + 300} = 0.50\text{ V}$$

$$f_c = \frac{1}{2\pi R_{th} C} = \frac{1}{2\pi \times 150\ \Omega \times 1\text{ nF}} = 1.06\text{ MHz}$$

(a)



$$V_{o\,(\text{max})} = V_s \frac{R \| R_L}{R_s + R \| R_L} = 1 \frac{120}{100 + 120} = 0.55\text{ V}$$

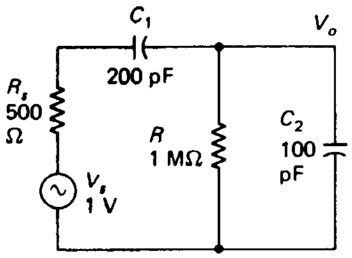
$$f_c = \frac{1}{2\pi R_T C} = \frac{1}{2\pi 220\ \Omega \times 1\text{ nF}} = 0.72\text{ MHz}$$

(b)

***FIGURE 18-4** RC filters with resistive source and load impedance can be reduced to simple RC filters: (a) low pass; (b) high pass.*

sections, -9 dB for three sections, and so on, if the resistor in the second section is several times higher in value than the resistor in the first section. For equal resistor values the correction at the break point for two sections is -10 db (factor of 3). For a section with resistor value R driving a section with resistor value $\frac{1}{2}R$, the correction at the break point is -12 dB (factor of 4) as illustrated by the circuit analysis of Fig. 18-6. The techniques used are those of Section 4.2.

At frequencies well above or below f_c no extra correction to the straight-line plots is required. In Fig. 18-6(a) this is because the reactance of the first capacitor across the line becomes very low (20 Ω at $10f_c$), providing a very low impedance

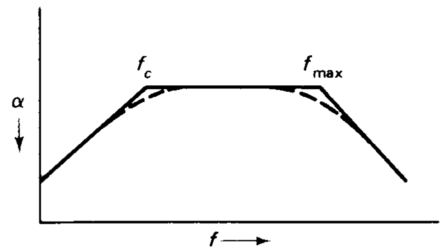


$$V_{o\,(\max)} = V_s \frac{C_1}{C_1 + C_2} = 1 \frac{200}{200 + 100} = 0.67\ \text{V}$$

$$f_c = \frac{1}{2\pi R(C_1 + C_2)} = \frac{1}{2\pi \times 1\ \text{M}\Omega \times 300\ \text{pF}} = 530\ \text{Hz}$$

$$f_{\max} = \frac{1}{2\pi R_s \dfrac{C_1 C_2}{C_1 + C_2}} = \frac{1}{2\pi(500\ \Omega)\left(\dfrac{20000}{300}\ \text{pF}\right)} = 4.8\ \text{MHz}$$

(a)



(b)

***FIGURE 18-5** Commonly encountered* RC *network: a resistive source capacitively coupled (C_1) to a resistive load shunted by an input capacitance (C_2). The circuit has both a high and a low cutoff frequency.*

source which drives the 100-Ω resistor of the second stage with negligible loading. A similar argument prevails for cascaded high-pass sections. Where the two values of f_c are not identical, the Bode plot will roll off at -20 dB/decade between the two f_c points, and then break to -40 dB/decade as shown in Fig. 18-6(f).

In staging cascade sections it is a good idea to keep this rule in mind: small capacitance follows large capacitance. Arranging sections in this order will minimize loading and unwanted attenuation.

Bandpass *RC* Filters can be constructed by following an *RC* high-pass section (or sections) with an *RC* low-pass section (or sections) having substantially higher f_c. The ratio of $f_{c(\text{hi})}/f_{c(\text{lo})}$ should be not much less than 10 : 1 or there will be significant attenuation at midband. Also notice the order of the attenuators: low pass follows high pass; small capacitance follows large capacitance. If the filter were built with high pass following low pass, serious attenuation (over -6 dB) would result at midband. Figure 18-7 shows a two-section *RC* bandpass filter with $f_{c(\text{lo})} = 3.2$ kHz and $f_{c(\text{hi})} = 32$ kHz. The midband and f_c corrections indicated were determined using the techniques of Section 4.2.

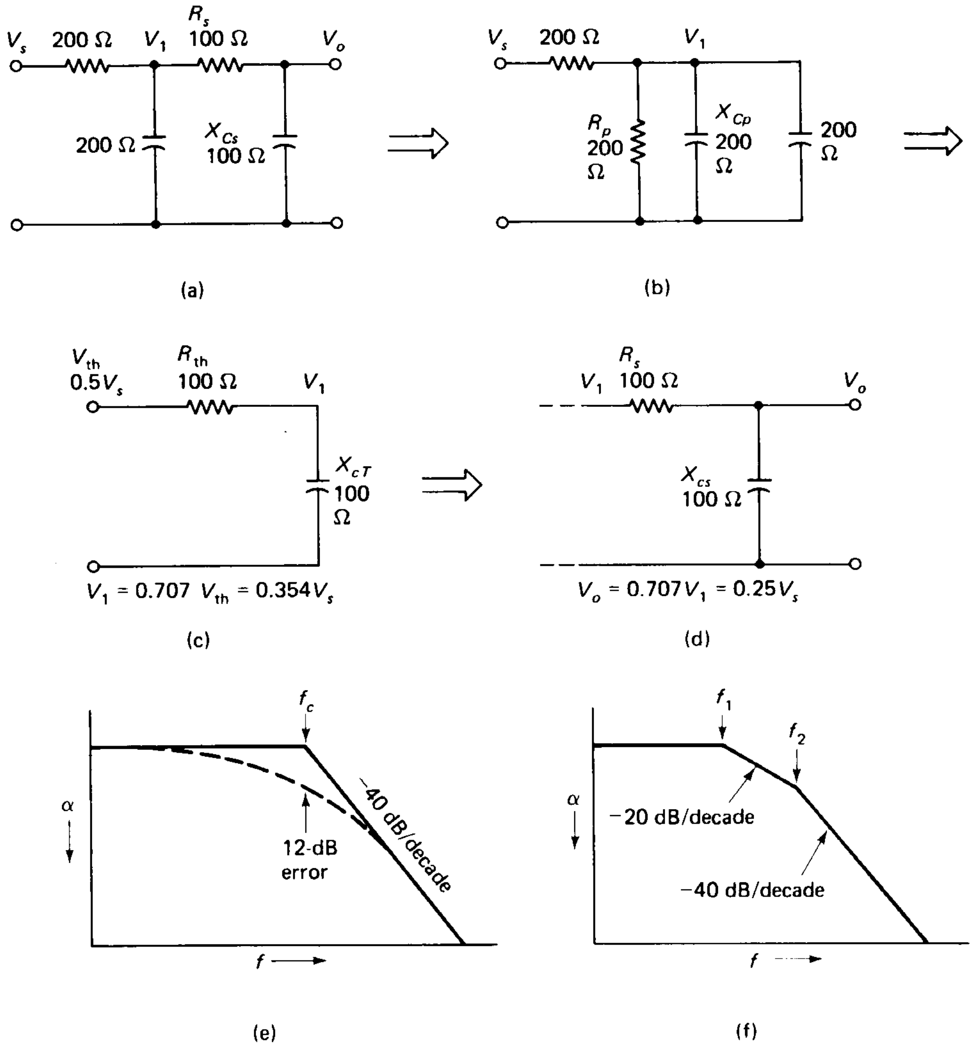


***FIGURE 18-6** Cascading two low-pass sections gives a rolloff of 40 dB/decade. (a)–(d) Reduction of a two-RC circuit at f_c. In the example, a 200-Ω section is loaded by a 100-Ω section, resulting in 12-dB attenuation at f_c (e). (f) Two break points appear when f_c of first and second sections are not equal.*

18.3 RC NOTCH FILTERS

There are two popular *RC* filters which are capable of producing a very high degree of attenuation at a narrow range of frequencies provided that they are operated into a very high impedance load.

The Wien Bridge is a four-arm ac bridge that is balanced at only one frequency. A special case of the Wien bridge is shown in Fig. 18-8, in which the two resistive and

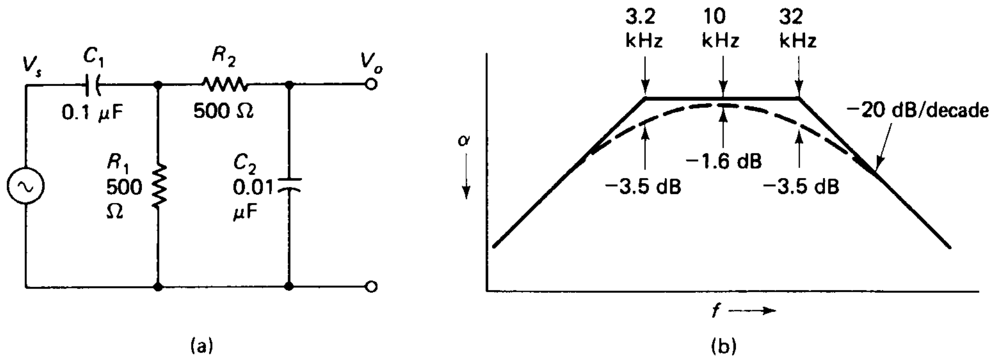


FIGURE 18-7 *Bandpass* RC *filter formed by cascading high- and low-pass sections.*

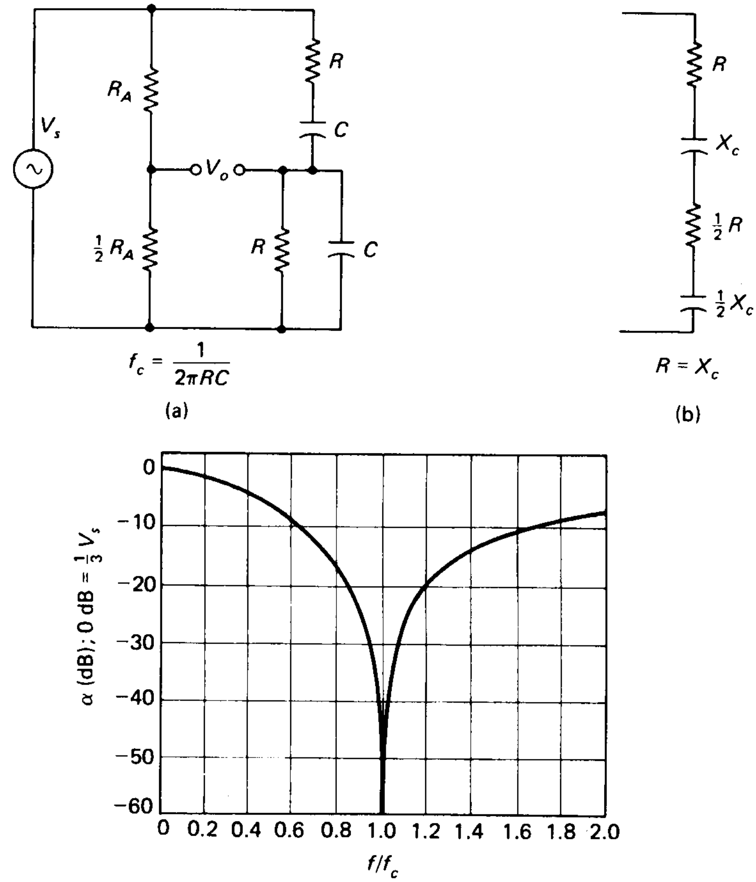


FIGURE 18-8 *Wien bridge notch filter: (a) circuit with equal* R *and* C *values in each arm for easy tuning; (b) equivalent of right-hand arms at* f_c*; (c) frequency-response curve.*

two capacitive frequency-determining elements are equal. This permits tuning of the bridge by ganged identical pots or variable capacitors.

The operation of the bridge is easy to understand. It is based on the equivalency of series and parallel *RC* circuits. At the frequency where $X_{Cp} = R_p$, the parallel *RC* combination is equivalent to the half-value series circuit shown in Fig. 18-8(b). The voltage at the right output terminal (to ground) is therefore $\frac{1}{3}V_s$ and in phase with V_s. The voltage at the left terminal is $\frac{1}{3}V_s$ in phase at all frequencies, so the voltage between the two V_o terminals is zero *at the frequency where* $X_C = R$. At higher frequencies C_p has a low reactance and drops the right-terminal voltage below $\frac{1}{3}V_s$. At lower frequencies C_s has a high reactance and lowers the right-terminal voltage. The right-terminal voltage is also out of phase with V_s at frequencies above or below f_c. All of these factors keep the bridge unbalanced except at f_c.

The attenuation of the Wien bridge (V_{in}/V_o) is theoretically infinite at f_c with no load across V_o. However, there is also substantial loss at frequencies near f_c, as shown in Fig. 18-8(c).

The Twin-T Notch Filter behaves much the same as the Wien bridge, but it has two important advantages:

- The grounded side of the input is common to one side of the output, making the twin-T compatible with single-ended amplifiers and unbalanced lines. The floating outputs of the Wien bridge require a differential amplifier to pick them up.
- The output in the passbands is fully equal to V_s, not $\frac{1}{3}V_s$ as in the Wien bridge.

However, the twin T requires three resistors and three capacitors, making it more difficult to tune with variable capacitors or pots. Figure 18-9 shows one form of the twin-T filter.

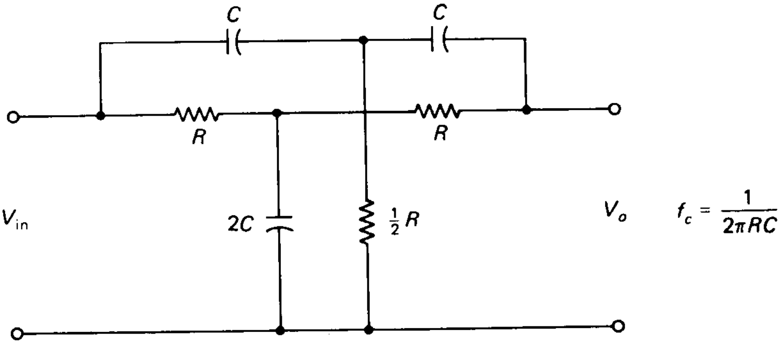


***FIGURE 18-9** The twin-T notch filter has a common ground between input and output, but ganged tuning is difficult.*

18.4 RC SELECTIVE FILTERS

Amplifiers with a selective peak can be built with *RC* notch filters as the feedback elements. Figure 18-10(a) shows a highly selective amplifier using a twin-T notch filter in this manner. The gain of the amplifier will approach infinity, resulting in instability at f_c if a fixed feedback R_f is not shunted around the twin T as shown. The 8-kΩ resistor may be varied slightly to produce oscillation or almost any degree of selectivity.

Figure 18-10(b) shows a simpler but more broadbanded tuned amplifier.

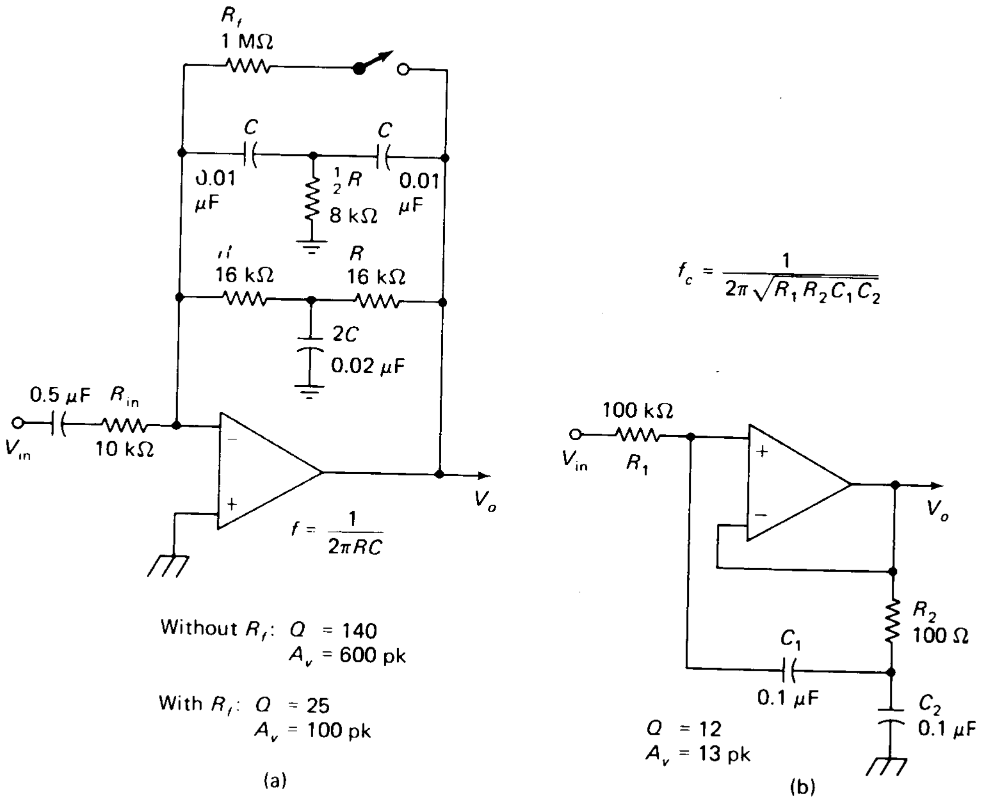


***FIGURE 18-10** (a) Peaking amplifier tuned by T-notch feedback. (b) Broadly peaked amplifier.*

18.5 LC FILTERS

High- and low-pass filters in the pi and T configurations can be constructed using *LC* rather than *RC* components. The advantage is that the rolloff rate is much steeper—60 dB/decade (18 dB/octave) rather than 20 dB/decade since there are three reactive components rather than one per section. The disadvantages include expense and weight of inductors, difficulty in obtaining high-*Q* inductors of high accuracy and stability, stray coupling between inductors, and saturation and nonlinearity of iron-core inductors.

Figure 18-11 shows the basic filter sections and the applicable formulas. All sections are designed to be driven by a source resistance $R_s = Z_o$ and to feed a load resistance $R_L = Z_o$, although these are omitted from the figures. The correction for rounding in the Bode plot at the break point is -3 dB for a single section.

Impedance Mismatch: Notice that the low-pass T has $Z_{\text{in}} \to \infty$ at high frequencies. The other sections have similar mismatch problems: $Z_{\text{in}} \to 0$ at high f for the pi low pass, $Z_{\text{in}} \to \infty$ at low f for the T high pass, and $Z_{\text{in}} \to 0$ at low f for the pi high pass. The SWR produced by this mismatch is 1.25 : 1 at a frequency factor of 2 into the passband, 2 : 1 at $f/f_c = 1.4$, and 5 : 1 at f_c. Well into the pass band the SWR is nearly unity, but it heads for infinity in the stop band.

Where reflections and standing waves on the line back to the source are undesirable, a pad may be used ahead of the filter to preserve Z_o, or the filter can be placed near the source, minimizing the time between incident and reflected signals and the length of high SWR line.

M-Derived End Sections, named for a mathematical constant m used in their design, are used to keep the SWR down to about 1.1 over the passband range to within about 10% of f_c. Figure 18-12(a) shows a pi-section low-pass filter with its m-derived end sections. End sections for the T low pass do exist, but require more inductors and are therefore less popular. Figure 18-12(b) shows a T high pass with end sections, with the pi high pass not shown for the same reason. Notice that the capacitors from the end section and from the filter proper can be combined into a single capacitor in each case. The attenuation at f_c is still about -3 dB for filters with end sections.

Beyond f_c (in the stop band), the end sections do not preserve the constant impedance Z_o. In fact, they comprise tuned circuits resonant 25% beyond f_c which make $Z_{\text{in}} \to \infty$ for the pi-end and $Z_{\text{in}} \to 0$ for the T-end sections. This puts a very sharp notch in the filter's response curve at $1.25f_c$ (or $0.8f_c$ for high pass), but introduces 100% reflection of the out-of-band signal.

Filters in cascade, up to five sections, can be constructed to obtain extremely high roll-off rates beyond f_c. End sections are usually used with these high-performance filters. They are structured by breaking the circuits of Fig. 18-12(a) or (b) along the dashed lines and inserting the required number of pi or T sections from Fig. 18-11(b) or (d), respectively.

The stringent requirements on the components and construction practices for such filters must be emphasized. Coils should be air-core No. 14 or heavier wire for RF, or toroid inductors for AF. Capacitors should be mica at RF, or Mylar at AF. All components should be within 5% of target value. Cylindrical coils should be mounted at right angles to each other, and each section of the filter should be shielded from the other sections by a metal plate (or full enclosure for very high attenuations), with a hole to receive the input wire. The entire filter must be enclosed in a metal box to eliminate stray pickup and radiation. A four-section filter has an attenuation of 72 dB one octave (factor of two) from the cutoff frequency. This is a voltage factor of 4000. If a 4-V input were allowed to couple 1 mV to the output by stray capacitance or inductance, the filter's performance would be ruined.

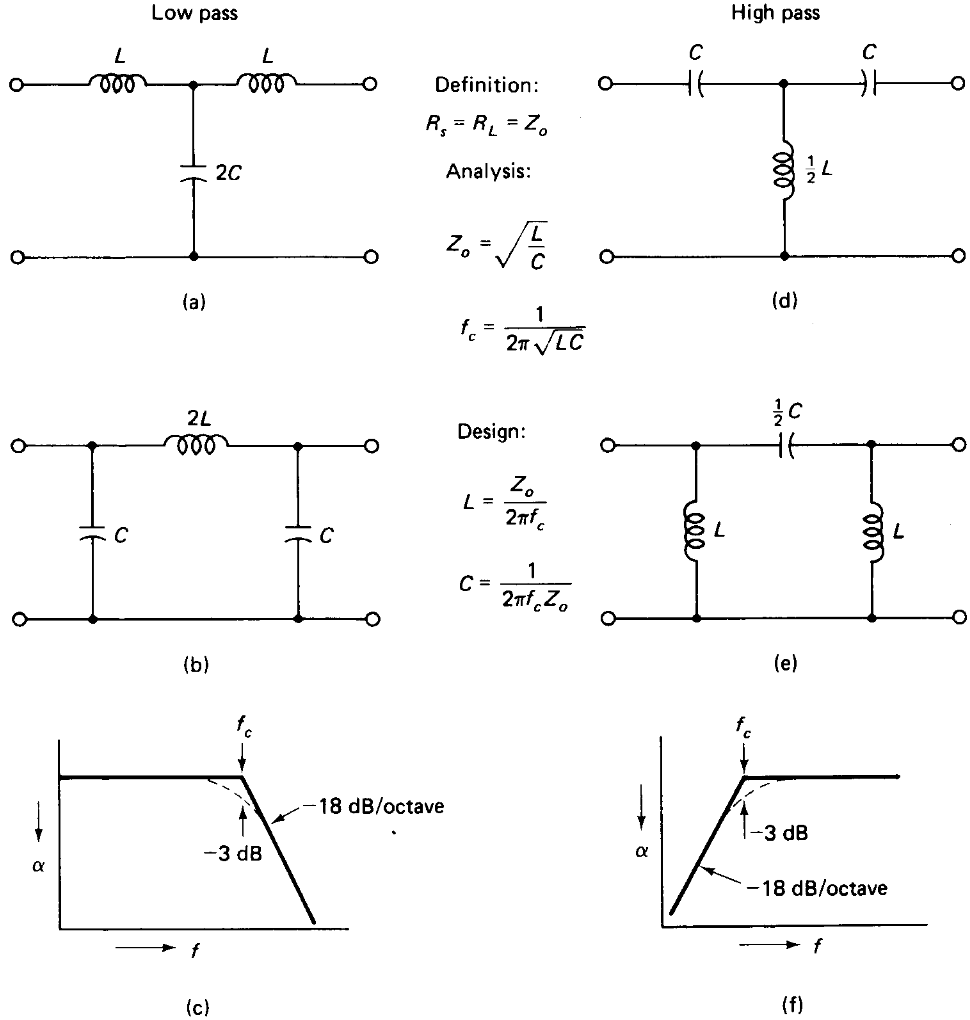


***FIGURE 18-11** LC filters have a basic rolloff rate of 60 dB/decade. Low-pass filters: T section (a), pi section (b), and response curve (c). High-pass filters: T section (d), pi section (e), and response curve (f). Formulas apply to all four types.*

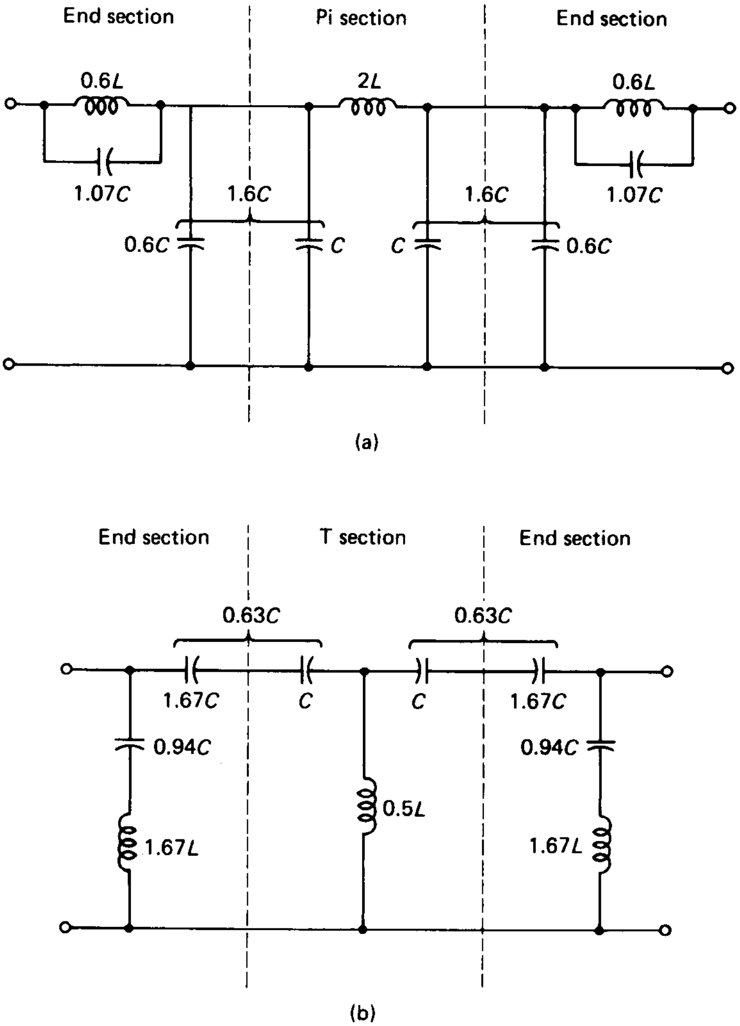


FIGURE 18-12 M-derived end sections preserve impedance match within the pass band and place a sharp attenuation notch just outside it. (a) end sections for low-pass pi; (b) end sections for high-pass T filters.

18.6 CRYSTAL FILTERS

Quartz crystals, when ground into wafers and mounted between two conductive plates, have the characteristics of very-high-Q tuned circuits at a frequency that is fixed primarily by the thickness of the wafer. Frequencies commonly available range from 100 kHz to about 30 MHz. Figure 18-13(a) shows the electrical equivalent of a 1814-kHz crystal. Of course, no real electrical inductor of 0.22 H

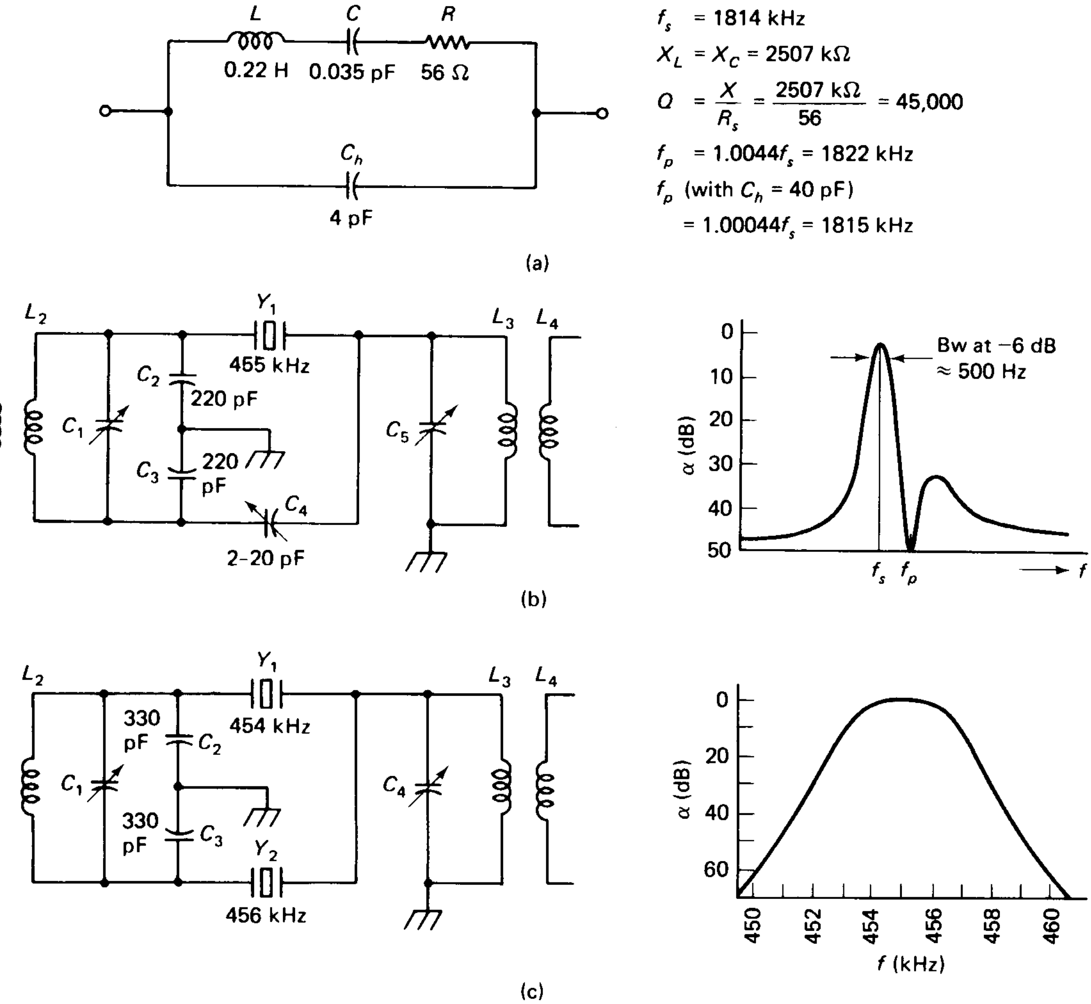


***FIGURE 18-13** (a) Equivalent circuit for a crystal with typical values for an 1814-kHz unit. Holder capacitance C_h places a parallel resonance very near the natural series-resonance point. (b) Crystal gate filter and response curve. Varying C_4 moves notch f_p. (c) Half-lattice circuit and broadband flat-topped response curve.*

could be constructed without incurring a stray capacitance many times higher than 0.035 pF, which is why electrical tuned circuits cannot achieve Q's of 45,000. At the series-resonant frequency of L and C, the impedance of the crystal drops to R (56Ω in the example). At frequencies $\frac{1}{2}$% or more above or below f_s, the shunting reactance of the holder-plate capacitance C_h (about 22 kΩ at this frequency) becomes lower than the reactance of the series L and C, and represents the crystal impedance.

At a frequency slightly less than $\frac{1}{2}\%$ above f_s, the reactance of the series L and C becomes inductive and equal to X_{Ch}. This constitutes a parallel resonance, and the crystal impedance approaches infinity at this frequency f_p. The parallel-resonant frequency can be moved closer to f_s by adding to the shunt capacitance C_h, but this lowers the off-resonance impedance if it is carried too far. Oscillators built to exploit this effect can be "pulled" a small fraction of a percent either way by a trimmer capacitor across the crystal.

The Crystal Gate Filter of Fig. 18-13(b) passes only the series-resonant frequency. C_1 and C_5 tune their coils to the passband frequency and are required to eliminate spurious frequencies which may be passed by the crystal. The signal applied to C_4 is 180° out of phase with that applied to the crystal, so when $C_4 = C_h$ the leakage through C_h is canceled and the attenuation at frequencies off f_s is increased. Varying C_4 moves the parallel-resonant notch with respect to f_s, which may be desirable if a strong nearby signal is to be eliminated.

The Half-Lattice Filter of Fig. 18-13(c) provides a relatively flat bandpass between the two crystal frequencies, with sharply increasing attenuation outside the passband. This is in contrast to the narrow peaked response of the crystal gate and the broad response of the LC tuned circuit, and is ideal for many radio-frequency applications. The passband (difference between the two crystal frequencies) is limited in practice to about $\frac{1}{2}\%$ of f_s. As many as five half-lattice filters may be placed in cascade to obtain extremely sharp cutoff outside the passband.

18.7 MECHANICAL AND CERAMIC FILTERS

Mechanical Filters are used where a relatively broad flat-topped response is required and cost is not a prohibitive factor. Six-decibel bandwidth up to 5% of the passband frequency with roll-off to -60 dB at 6% bandwidth are commonly available. This "straight skirt" selectivity is expressed by the filter's *shape factor*—the ratio of its bandwidth at -60 dB to its bandwidth at -6 dB. In the example above, the shape factor is $\frac{6}{5}$, or 1.2. Shape factors of down to 1.12 are available for special applications. Mechanical filters are commonly available in the range 50 to 500 kHz.

Internally, a mechanical filter consists of an input electromagnet which produces mechanical vibrations in a series of connected resonant metal discs. Each resonant disc alone has a sharp peak (Q over 10,000), but the resonant frequencies are spaced over the passband to form a broad flat response. The output is taken from a second electromagnet coil.

Ceramic Filters, which are similar to lower-Q crystal filters, are inexpensive and require no tuning, and have therefore become popular in AM radios.

Ceramic Ladder Filters contain a series of ceramic elements and provide shape factors nearly as good as those of some mechanical filters at a considerably lower price. Most ceramic filters are designed for 455- or 500-kHz center frequency.

19

SIGNAL PROCESSING

This chapter presents a variety of circuits for generating and processing signals. The circuits were chosen because of their widespread application, so you will be likely to encounter most of them in industrial electronic instruments and control systems. The commercial versions will often be augmented with noise filters, buffer amplifiers, range switches, and temperature-compensating circuitry, but an understanding of the basic circuits will enable you to see through all this to the basic function of the device.

19.1 CURRENT AND VOLTAGE CONVERSIONS

Current-to-Voltage Conversion can be achieved with the circuit of Fig. 19-1(a). The input terminal operates at virtual ground. The circuit is easy to understand if it is remembered that the op-amp input current is virtually zero. I_{in} must therefore be virtually equal to I_{R1}, and $V_o = I_{R1}R_1$. The circuit is useful for currents from $I_{O(\max)}$ to approximately 20 $I_{\mathrm{IN(bias)}}$ of the op amp, which covers ± 5 mA to ± 150 nA for the *308* op amp. FET-input op amps can extend the current down to the picoampere range.

Voltage-to-Current Conversion is accomplished by the circuit of Fig. 19-1(b). Picturing the circuit with the bottom of R_5 grounded, it is easily recognizable as the differential-op-amp circuit with V_o appearing across R_5, so that $I_{R5} = V_o/R_5$. If R_4

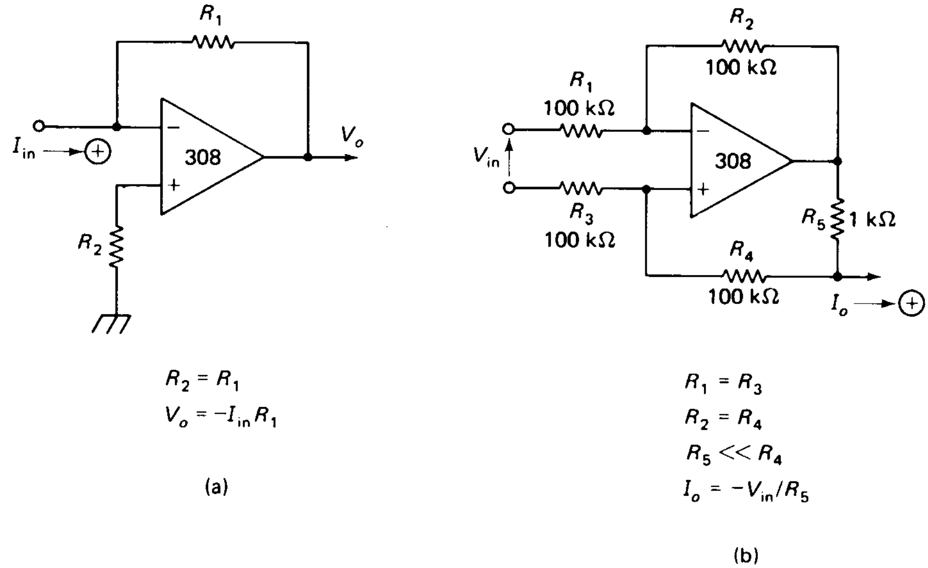


***FIGURE 19-1** Current-to-voltage converter (a) and voltage-to-current converter (b) with applicable formulas.*

is much greater than R_5, then I_o is nearly equal to I_{R5}. Of course, it is not necessary that the bottom of R_5 be at ground. Either input terminal or some point referenced to them can serve as ground, as long as the common-mode input range is not exceeded. Very low currents can be obtained by raising R_1 and R_3 (or lowering R_2 and R_4) by a factor of 10 to lower the dif-amp gain. The linearity of the current source as V_o changes requires a good match of the R_2/R_1 and R_4/R_3 ratios, so a trim pot in one of these positions is advisable.

AC to DC—The Perfect Diode: The circuit of Fig. 19-2(a) provides two outputs equal to the positive and negative half-cycles of an ac input, without the diode drop encountered with a conventional rectifier. Amplification of the output by any factor R_f/R_{in} is also possible. Any value of load resistance can be connected from either output to ground, provided that the op-amp output current limit is not exceeded. Where millivolts of output are important an offset-cancellation circuit is recommended.

Average Value of AC: Most ac voltmeters actually respond to the average value of a rectified ac wave, although this value is scaled to read in rms terms when the input is a sine wave. This is done because average is easier to obtain than true rms, but remains relatively more accurate in the face of nonsine waves and noise spikes than the other alternative, namely peak-reading. Figure 19-2(b) shows an addition to the perfect rectifier which produces a dc output proportional to the average value of the positive half-cycle input. The output impedance is quite high, so a noninverting amplifier is recommended as a buffer. A gain of 2 would fill in for the

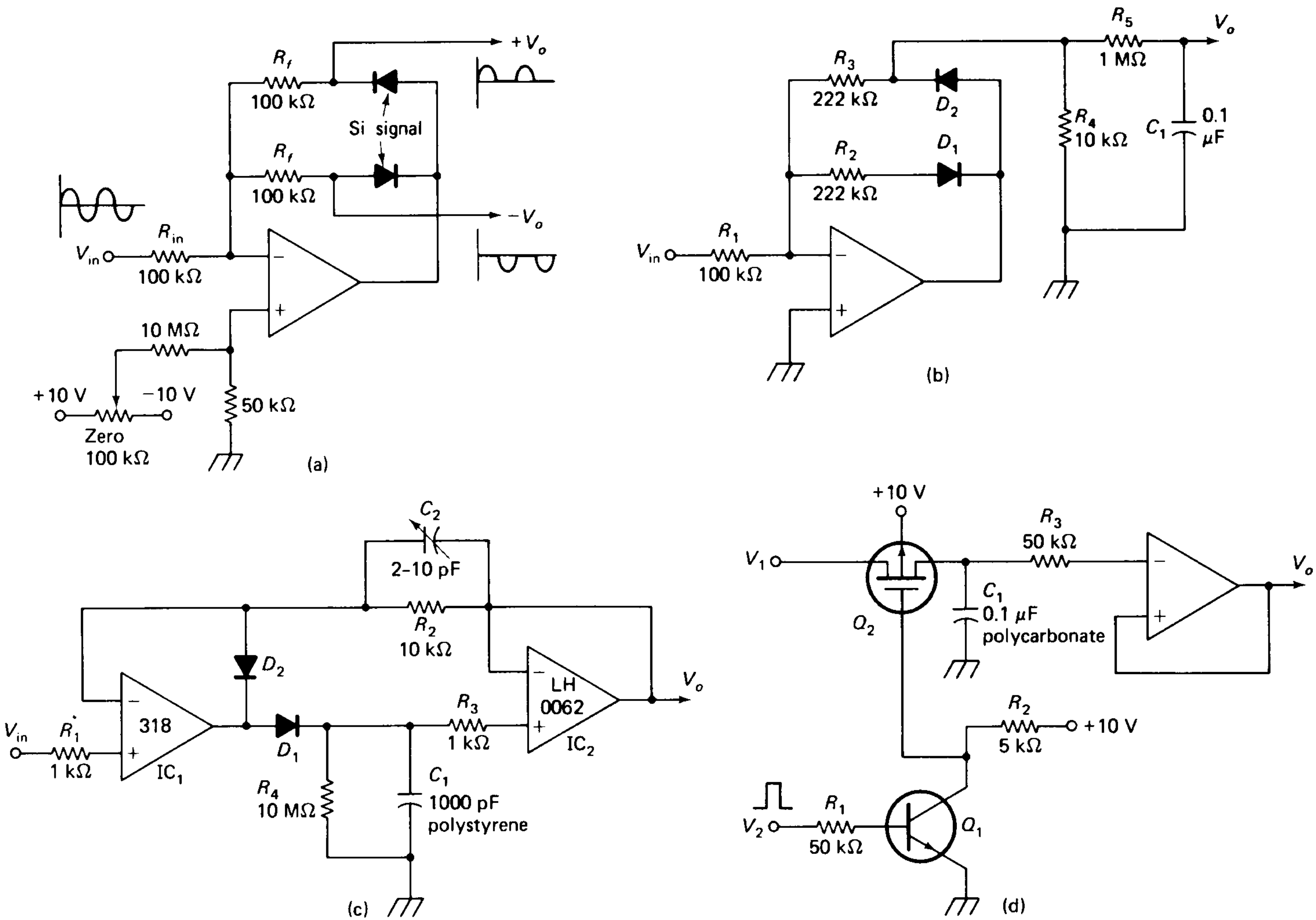


FIGURE 19-2 *Precision rectifier with positive and negative half-cycle outputs (a), and with averaging filter (b). (c) Positive-peak-reading circuit (reverse D_1 and D_2 for negative). (d) Gated sample-and-hold circuit.*

negative half-cycle, so a gain of 2(0.707/0.637), or 2.22, will give a dc output equal to the rms value of an input sine wave. The time constant of averaging filter R_5C_1 is 0.1 s, so the circuit will reach full response in about 0.5 s, which is considered an acceptable lag. Input frequencies below 50 Hz begin to show excessive ripple in V_o, so a two-section filter may be required. R_4 is necessary to preserve a near-zero source impedance for the filter when D_2 turns off.

Peak Reading of AC: The circuit of Fig. 19-2(c) gives a dc output equal to the repetitive positive peak of the input wave. Feedback of V_o through R_2 to the reference input of IC_1 ensures that v_O will equal v_{IN}, regardless of the drop across D_1. D_2 overrides this positive feedback during negative voltage inputs. Reversing D_1 and D_2 allows the circuit to read negative peaks. R_4 discharges C_1 to allow the output to follow decreasing pulse heights. IC_2 should be a low-bias-current FET-input type and IC_1 should be a fast bipolar type for this application.

Peak-holding Circuit: The circuit of Fig. 19-2(c) will hold the maximum peak voltage of V_{in} for several seconds if R_4 is removed, leaving no discharge path for C_1. Peak memory circuits are useful in detecting noise and other one-shot voltages. C_1 should be sized to allow it to be charged by the shortest expected input spike using the familiar $CV = It$ formula. For an op amp with a 5-mA output capability sensing 5-V pulses as short as 1 μs:

$$C_1 = \frac{It}{V} = \frac{5\text{ mA} \times 1\ \mu\text{s}}{5\text{ V}} = 1\text{ nF} \tag{19-1}$$

Output droop rates as low as 20 mV/s are possible with the components specified. C_1 and D_1 should have the lowest possible leakage in this application, and the voltage follower should contain an offset cancellation circuit. A pushbutton switch can be placed in the R_4 position to reset the memory.

Sample and Hold: A common instrument problem is "What was the steam pressure when the pipe broke?" or some similar question requiring memory of a voltage at a particular critical instant of time. The circuit of Fig. 19-2(d) charges C_1 to the input voltage V_1 when the sample command V_2 goes positive. With V_2 at ground, the Q_1 collector is positive, holding P-channel FET Q_2 off, isolating the charge stored on C_1. R_3 is for overload protection.

19.2 ADDING CIRCUITS

The most popular summing circuit is the add-and-invert op-amp circuit of Fig. 19-3(a). Operation is exactly the same as for the inverting amplifier of Section 16.2. Any number of inputs can be used, and there is no interaction between them because the summing point (the −op-amp input) is at virtual ground.

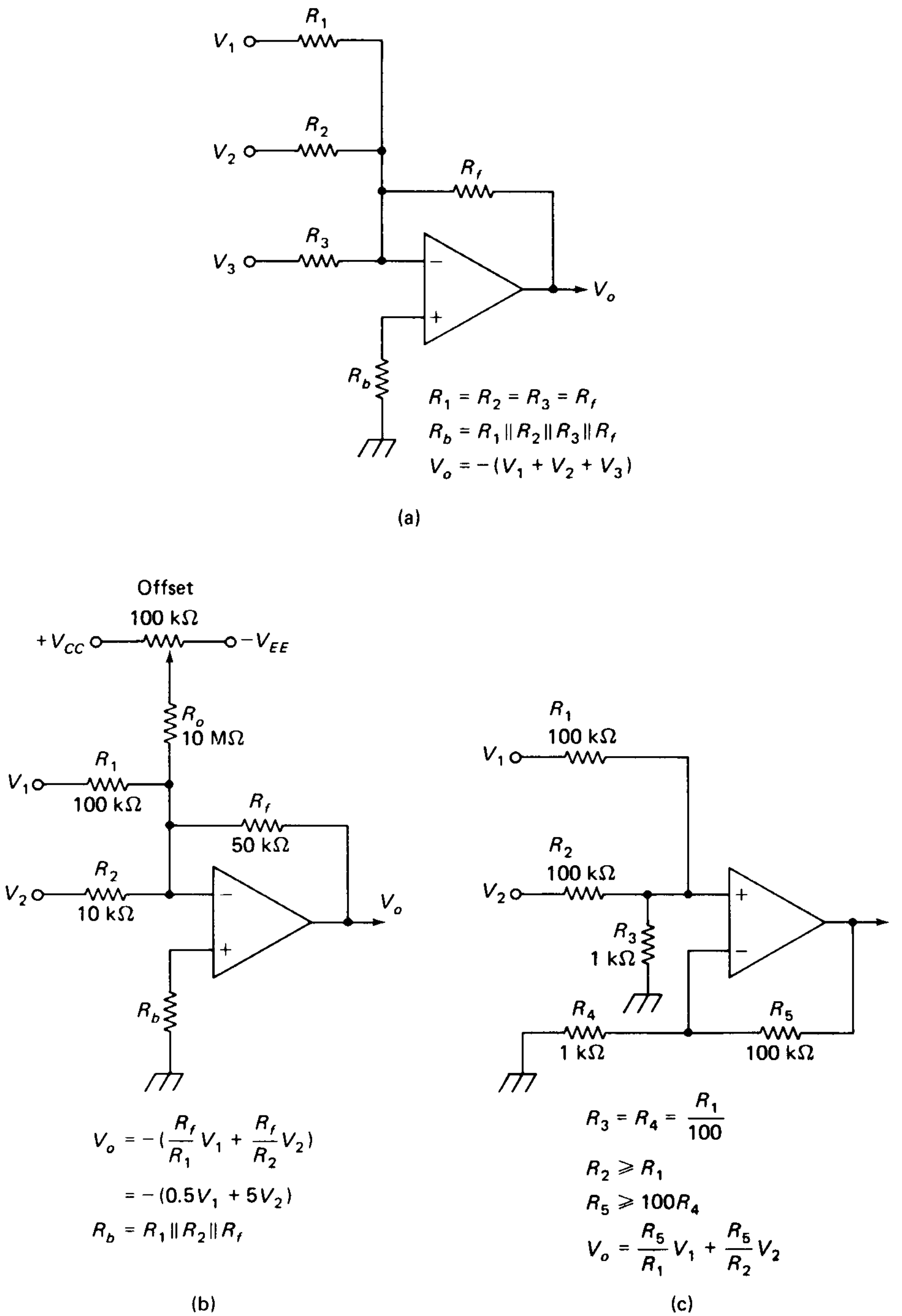


***FIGURE 19-3** (a) Summing circuit. (b) Weighted summing. (c) Noninverting summing circuit.*

Multiply and Add is effected by simply changing the value of the input resistors, as shown in Fig. 19-3(b). Each input is now multiplied by R_f/R_{in} before being summed. The final output is, of course, still inverted. Note that each input now presents a different input resistance. Unless R_s is hundreds of times smaller than each R_{in}, a voltage-following buffer [Fig. 16-2(b)] should be used on each input.

Where a few millivolts output offset would be a problem, or where R_f is in the megohm range, the offset-adjust pot and R_o are recommended. If inversion of the output signal is undesirable, a gain-of-one inverting amplifier can be placed after the summer to reinvert the output. If subtraction is to be performed, the subtracted input can be applied directly to the summer while the added input is first run through an inverter.

Figure 19-3(c) shows a circuit for noninverting addition in a single op amp. Any number of inputs can be used provided that the parallel combination of input resistances is much greater than R_3. The gain factor for all inputs can be made greater or less than unity by making R_5 smaller or larger, respectively. Different gain factors for each input can be realized by making R_2 greater than R_1. The input resistance of this circuit is equal to R_1, R_2, etc., at each input.

19.3 MULTIPLYING AND SQUARING CIRCUITS

Multiplying a voltage by a constant factor is a simple job, requiring only a basic amplifier. Multiplying a voltage by another voltage, both of which are varying, is more of a challenge. Measuring the IV (power) product when an odd-shaped waveform appears across a load that is not a simple resistance presents one application for a multiplying circuit. SCRs and switching transistors often produce such waveforms.

Photocell Multiplier: Figure 19-4(a) shows a multiplier in which V_1 is used to control the gain of the amplifier fed by V_2. When V_1 is $+1$ V, the IC, through emitter-follower Q_1, applies enough negative voltage to lamp I_1 to make the PC_1 resistance 100 kΩ, thus keeping the op-amp input at virtual ground. The photocells are a matched pair, equally coupled to I_1, so PC_2 likewise has $R = 100$ kΩ, making the gain of IC_2 unity. Two volts at V_1 will make $R_{PC} = 50$ kΩ, yielding an IC_2 gain of two, and so on.

The accuracy of this system depends upon the matching of the photocells. Prepackaged pairs matched to within 1% can be obtained. A more serious drawback in most applications is the response speed, which is limited to a few tens of hertz by the photoresistive cells. Notice that this is a *two-quadrant inverting* multiplier. V_2 may be positive or negative, but V_1 must be positive.

Dif-Amp Multiplier: Figure 19-4(b) shows a two-quadrant multiplier with a response speed that can extend into the megahertz range with proper design and selection of the op amps. As in the first circuit, V_1 controls the gain of V_o/V_2, but this time the variable resistance is the emitter-junction resistance r_j of the transistors in the dif-amp Q_1–Q_2, which is approximately $0.03/I_E$.

V_1, which must be positive, supplies a current between 0 and 2 mA to the two emitters of the dif amp, varying their total resistance from infinity to $2(0.03/0.001) = 60\ \Omega$. The gain of the dif amp thus varies from zero to $600/60 = 10$. V_2 is divided down by a factor of 100 to provide a low-impedance low voltage to drive the base of Q_1. Higher base voltages would result in V_2 controlling the gain of the dif amp. IC_2 provides a gain of 10 and restores the output from floating differential to single-ended ground referenced.

Pulse-rate Divider: An approach to analog division that is capable of very high accuracy, but is limited in frequency response, is shown in Fig. 19-4(c). IC_1 free-runs with a period of about 500 μs with $V_A = 5$ V. This is calculated from R_1 current and C_1 charging time:

$$I_{R1} = \frac{V_{R1}}{R_1}$$

$$t_1 = \frac{C_1 V_{C1}}{I_{R1}}$$

Increasing V_A lengthens the period proportionally, since the *Control* voltage equals the *Threshold* voltage to which C_1 must charge to end the cycle.

IC_2 is triggered by IC_1 at $f = 1/t_1$, but the duration of its output pulses is controlled by V_B. For $V_B = 5$ V, t_2 is about 50 μs. The train of pulses at pin 3 of IC_2 is thus approximately 10 V high, with a duration of 50 μs, repeating every 500 μs—an average value of 1 V. Increasing V_B lengthens the pulse duration, thus increasing the average value proportionally. Increasing V_A decreases the repetition rate, decreasing the average value proportionally.

Filter R_3C_3-R_4C_4 converts the pulse train to dc of its average value. The filter's cutoff is 32 Hz, so repetition rates above 1 kHz are attenuated by at least 60 dB. This is the minimum frequency, corresponding to $V_A = 10$ V.

The 100-V supply and 100-kΩ resistors are a simple but crude method of obtaining a fairly constant 1-mA charging current for C_1 and C_2. Substitution of the PNP-transistor current source shown in Fig. 9-6(a) will greatly improve accuracy. The voltage-variable IC current source of Fig. 19-1(b) may be used in place of R_1 to provide a third multiplying input.

Squaring a Voltage can be accomplished by connecting V_1 to V_2 in any of the multiplying circuits described above. In addition, there are nonlinear input elements especially designed to produce $V_o = V_{\text{in}}^2$ when used in place of R_{in} in a simple inverting op-amp circuit. Squaring is useful in obtaining true rms values from a waveform of arbitrary shape. It should be noted that this *mathematical squaring* is not the same as *signal squaring* (i.e., producing a square wave from an analog input signal).

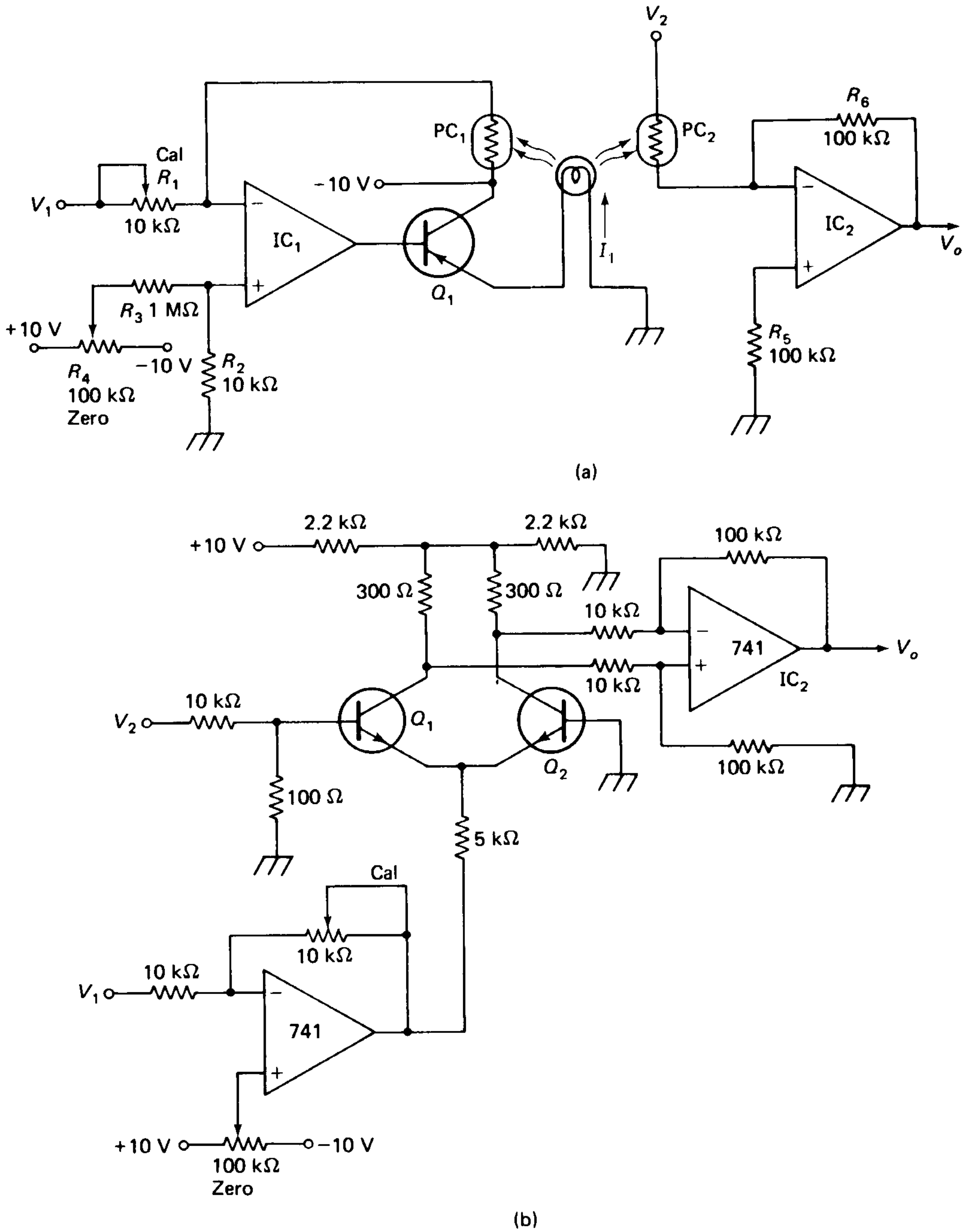


***FIGURE 19-4** Analog multipliers: (a) IC_1 controls gain of IC_2 via matched photocells; (b) IC_1 varies bias, hence emitter r_j, hence gain of dif amp.*

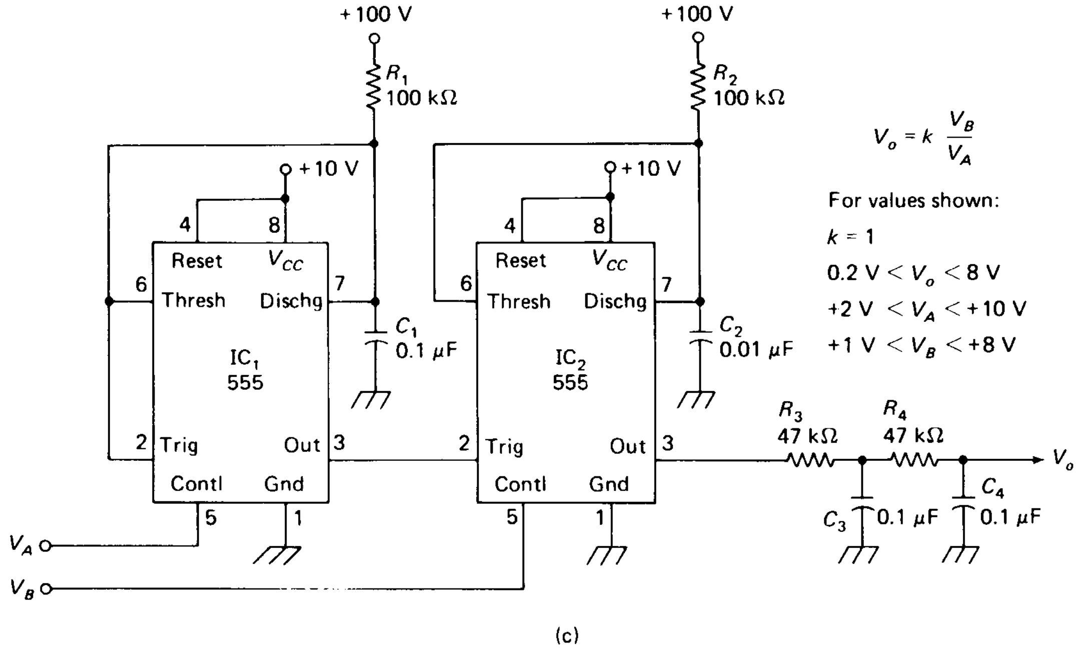


FIGURE 19-4 *(c) Analog division: V_A controls rate and V_B controls duration of pulses out of IC_2. Averaging filter smooths pulses to varying dc.*

19.4 LOG AND ANTILOG GENERATORS

The voltage across a silicon transistor's base–emitter junction is proportional to the logarithm of its junction current to within 1% from 0.05 μA to 500 μA, and this property can be exploited to produce amplifiers with outputs equal to the log (or antilog) of the input. This is useful because adding the logs of two voltages and taking the antilog produces the product of the two voltages. Similarly, multiplying a log by two produces the square, subtracting logs produces the quotient, and dividing logs produces roots of input voltages.

Figure 19-5(a) shows the basis of the log generator. Note that the output voltage is V_{BE} which increases as the log of input current, since $I_E = I_{in}$. There are several problems with this basic circuit:

- There is an output offset voltage V_{BE}.
- The output voltage is proportional to absolute temperature as well as to log V_{in}.
- The output is inconveniently small, covering less than 0.1 V for a factor-of-10 change in V_{in}.

The circuit of Fig. 19-5(b) corrects each of these deficiencies. Q_2 is fed by current-source IC_2 with a current equal to the Q_1 current at $V_{in} = 1$ V. Thus $V_o = 0$

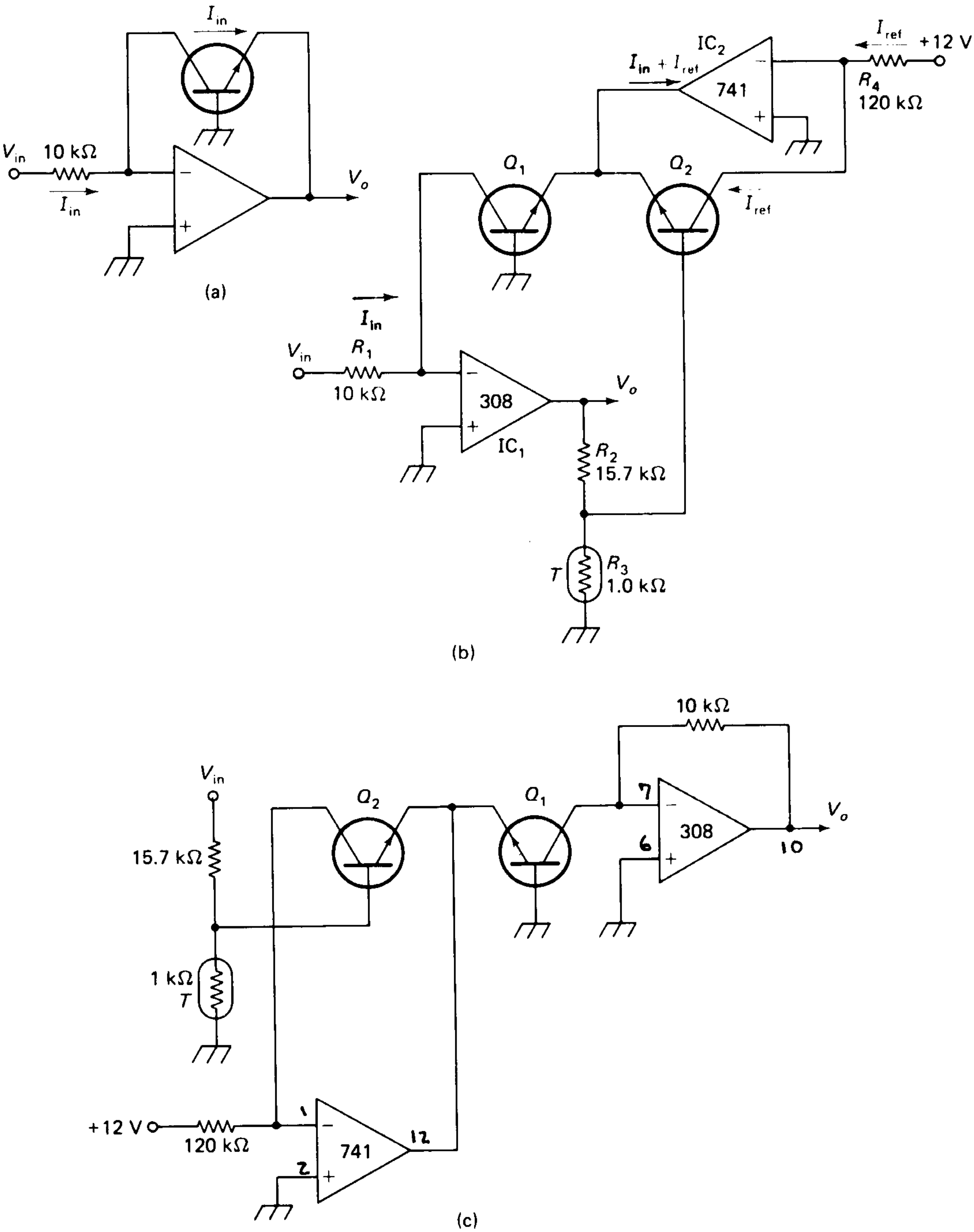


***FIGURE 19-5** (a) Basis of log generator. (b) Practical log generator. (c) Antilog generator.*

when $V_{in} = 1$ V, corresponding to log 1 = 0. Voltage divider $R_2 - R_3$ increases V_o to $16.7(V_{BE1} - V_{BE2})$, and thermistor R_3 increases amplifier gain at higher temperatures, compensating for lower V_{BE}. The output of this circuit is -1 V for $V_{in} = +10$ V, 0 V for $V_{in} = +1$ V, $+1$ V for $V_{in} = +0.1$ V, and so on. Only positive inputs can be handled, and the output is, of course, inverted.

Figure 19-5(c) shows an antilog generator which places the transistor base-emitter junction in the input rather than the feedback loop of the op amp. The input may be positive or negative, but the output will always be positive: $+10$ V out for -1 V in, $+1$ V out for 0 V in, $+0.1$ V out for $+1$ V in, and so on.

True rms Voltmeter: A system for performing the root-of-the-mean-of-the-square operation is given in block-diagram form in Fig. 19-6. The blocks are all circuits that have been described in this chapter. The design is simplified somewhat by the fact that the thermistors can be eliminated when an antilog generator follows a log generator.

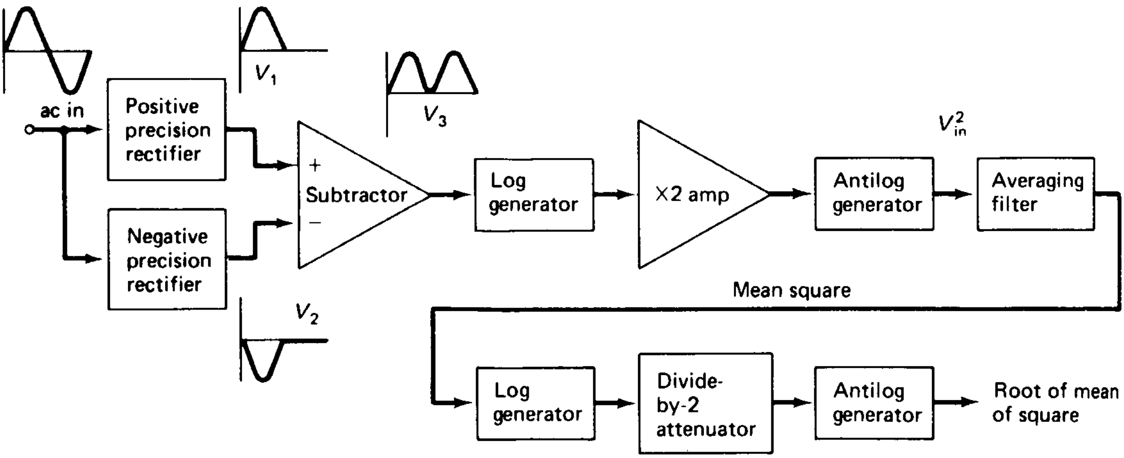


***FIGURE 19-6** Block diagram of a true rms voltmeter based on square and square root by multiplying and dividing logs.*

19.5 INTEGRATION AND DIFFERENTIATION

Integration is a mathematical operation of calculus which is widely used in electronic circuits. Without going into the mathematical details, integration is simply a process of *accumulation*. Thus a 1-V input, integrated, would produce 0 V out initially, 1 V after 1 s, 2 V after 2 s, and so on, as shown in Fig. 19-7(a). A 2-V input produces a faster accumulation, and a negative input produces a negative accumulation, as shown in Fig. 19-7(b) and (c), respectively. It follows that the integral of a ramp ($v = t$) is an exponential rise ($v = t^2$), the integral of a square wave is a triangle, and the integral of a string of pulses is a staircase, as illustrated in the next three parts of Fig. 19-7. A less obvious, but equally true, relationship is that the integral of a sine wave is a negative cosine (90° lagging) wave.

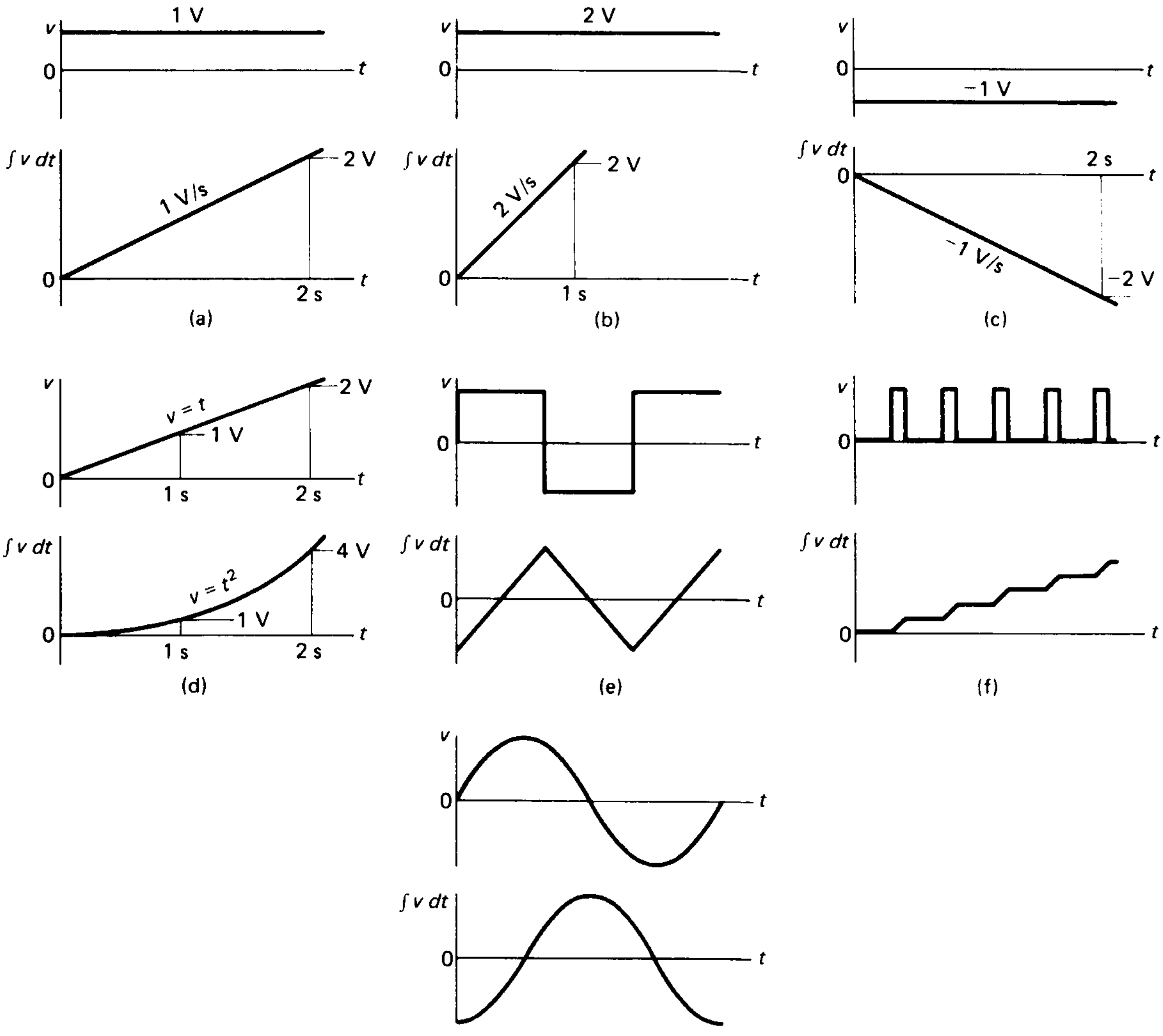


FIGURE 19-7 *Integration is a process of accumulation: seven examples.*

A little reflection will reveal that the output of an integrator depends upon the history of the input—what has been accumulated in the past. Therefore, every integration includes a constant—a dc voltage which may have been accumulated from previous input. In practical integrators the accumulated dc voltage can easily exceed the saturation limits of the circuit. Therefore, it is generally necessary to provide a reset function to return the integrator output to zero.

A Practical Integrator is shown in Fig. 19-8(a). For the example values given, the input resistor's current is 1 μA. Since the op-amp input current is virtually zero, the capacitor must carry virtually this same 1 μA current. The op-amp output will

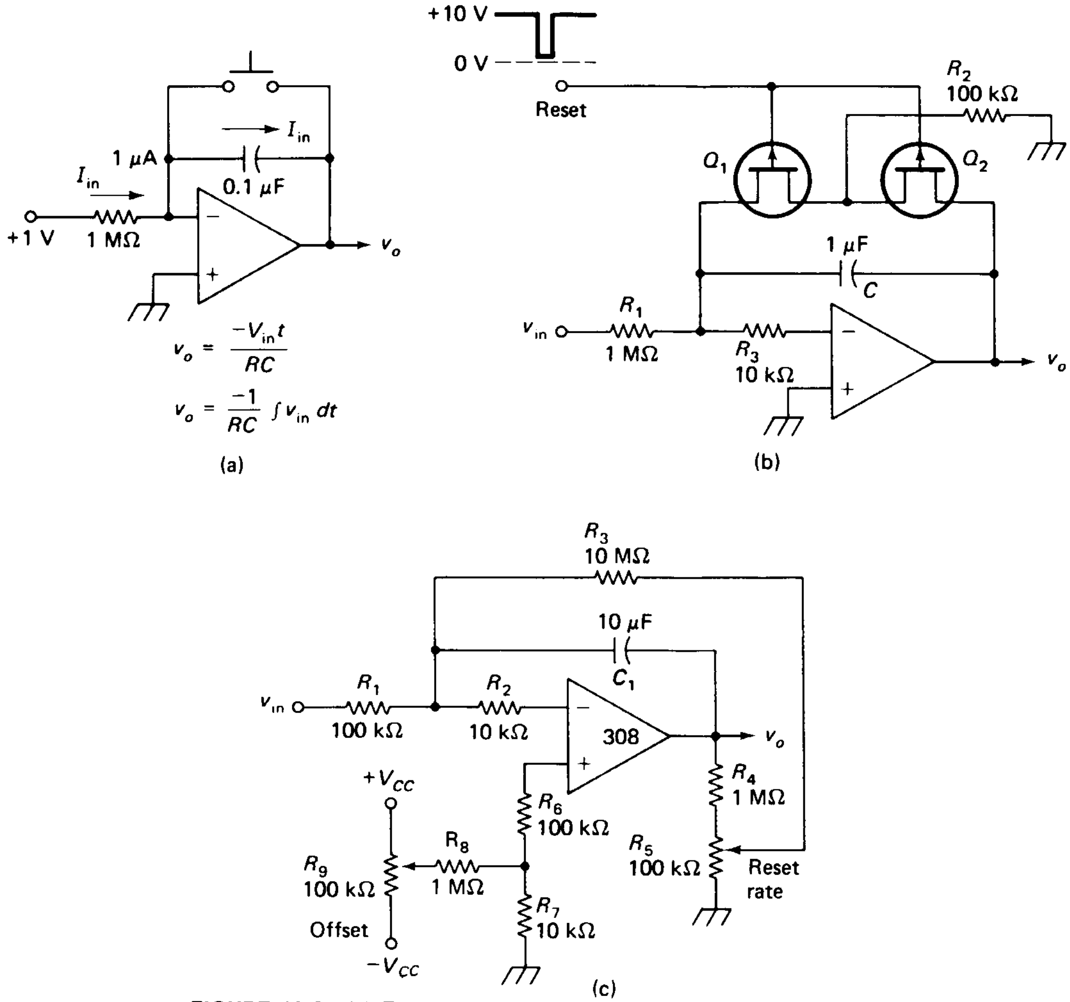


FIGURE 19-8 *(a) Elementary integrator with switch reset. (b) Electronic reset with FETs. (c) Low-drift self-resetting integrator.*

produce any voltage necessary to make $I_c = 1\ \mu A$. The output required is found from the $CV = It$ formula:

$$\frac{V_o}{t} = \frac{I}{C} = \frac{1\ \mu A}{0.1\ \mu F} = 10\ V/s$$

The capacitor voltage (and hence the output voltage, since one end of the capacitor is at virtual ground) must rise at a rate of 10 V/s to keep a current of 1 μA flowing. This corresponds to Fig. 19-7(a): the integral of a constant dc input is a ramp output. Actually, the ramp is negative for a positive dc input because the inverting input of the op amp is used.

After about 1 s the output of this integrator reaches saturation and the capacitor must be discharged through the pushbutton switch to reset the integrator to zero.

Electronic Reset can be achieved by replacing the pushbutton switch with an FET. In some cases the leakage of the FET in the OFF state may be an appreciable fraction of the capacitor current, upsetting the circuit operation. Figure 19-8(b) shows a scheme for greatly reducing the effects of FET leakage. With Q_1 and Q_2 off, R_2 presents a relatively low impedance to ground for the Q_2 leakage current, while Q_1 presents a relatively high impedance back to the $-$input of the op amp. The 10-kΩ resistor R_3 is recommended wherever the feedback capacitor exceeds 0.1 μF to protect the IC in the event power is shut off while C is charged.

A Zero-Drift Integrator is shown in Fig. 19-8(c). No reset circuit is shown because the integrator will eventually return itself to zero if the input is held at zero. The other integrators shown will eventually drift into saturation with zero input and no reset operation because of op-amp bias and offset currents. Offset-cancellation circuits can reduce this drift by two or three orders of magnitude, but they cannot eliminate it entirely.

The circuit of Fig. 19-8(c) minimizes drift with bias compensation R_6 and offset cancellation R_7, R_8, and R_9. It also uses a low-bias-current op amp and the highest practical value of feedback capacitor to ensure that capacitor current is many orders of magnitude greater than input offset and leakage currents. R_3, R_4, and R_5 are then used to introduce a very slight amount of negative feedback to restore the output to zero. This introduces a slight distortion into the integrating function, so that a linear ramp would actually be curved slightly toward zero.

At maximum reset rate with $V_o = 10$ V the feedback current through R_3 is 100 nA. This is enough to completely overcome a 10-mV input or cause a 10% distortion on a 100-mV input at R_1. If a lower reset-rate setting can be used, or if V_o is less than 10 V, the distortion will be correspondingly less. The reset action follows a conventional RC discharge curve with a time constant of about 15 min at maximum reset rate.

In operation, the input is grounded and the R_5 wiper is set to the ground end. R_9 is then adjusted for a V_o that holds as nearly constant as possible on a sensitive scope or meter. R_5 is then brought up slowly until the output holds as nearly as required to zero. The term *slowly* here should be taken to mean a check every minute or so. It goes without saying that C_1 must have the lowest possible leakage for this application.

Differentiation is the inverse operation of integration. Conceptually, the *derivative*, as the result of differentiation is called, can be described as the *rate of change* of the input variable. Looking at the several graphs of Fig. 19-7 it will be seen that in each case the upper curve represents the rate of change (in V/s) of the lower curve. Therefore, if we integrate a function and then differentiate it, we get the original function back again.

Practical differentiators must make some compromise with true mathematical differentiation because reasonably fast-rise pulses can demand unreasonably high output voltages. Furthermore, noise pulses tend to be fast-rise and are greatly overemphasized in true differentiation.

In the differentiator of Fig. 19-9(a), R_1 limits the input current to 0.1 mA/V regardless of the rate of rise of input voltage. C_2 shunts R_2, lowering the gain at

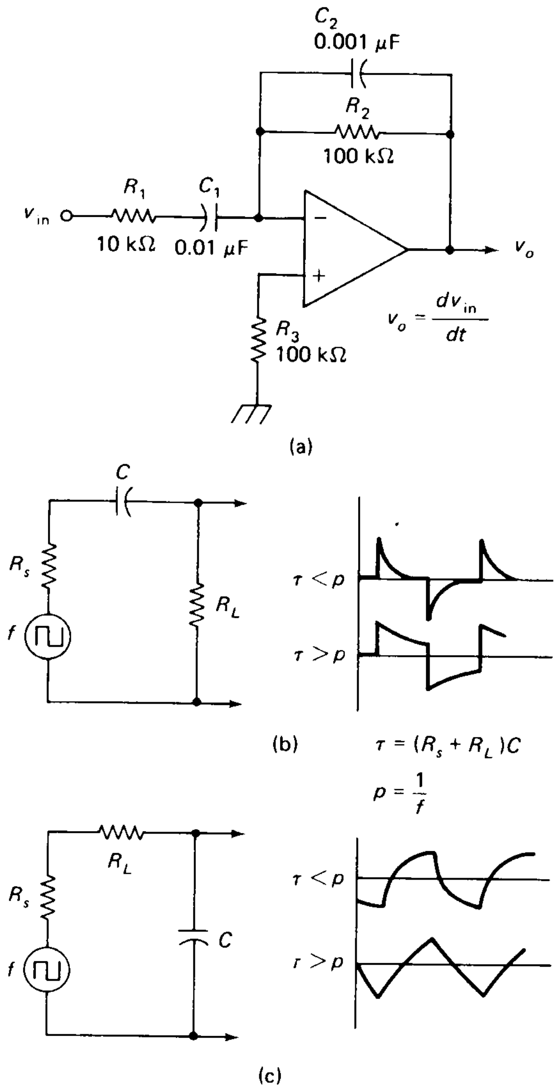


***FIGURE 19-9** (a) Practical op-amp differentiator. (b) Simple RC differentiator with waveforms showing maximum and too-long time constants. (c) Simple RC integrator with waveforms showing too-short and minimum time constants.*

high frequencies. C_1 and R_2 are the basic differentiator components. Noise response can be reduced by increasing R_1 and C_2. Signal response can be increased by increasing C_1 and R_2.

RC Differentiators and Integrators: This seems a good place to note that simple *RC* circuits are often used to approximate differentiators and integrators, as shown in Fig. 19-9(b) and (c). The *RC* differentiator is nearly as good as the op amp version because of the limits that must be placed on the op amp differentiator. The *RC* integrator resets to zero through R_1 and R_s as fast as it accumulates, and therefore produces a poor approximation to true integration. For these reasons, *RC* networks are most often used for differentiating and op amps are most often used for integrating.

19.6 INTEGRATOR APPLICATIONS

Rate Limiter: Figure 19-10(a) shows a circuit in which the output will follow the input voltage, subject to the limitation that the output rate of change cannot exceed 2 V/s, or whatever other limit is selected by integrator components R_3 and C_1. The inversion of IC_2 gives a negative feedback signal to R_4 which is balanced against the positive signal across R_1. IC_1 saturates if the input rate of change exceeds the integrator's limit. Gain can be added by increasing R_4 or lowering R_1. Rate limiters are useful in industrial process controllers, where rapid changes in position, velocity, temperature, and so on, would be damaging to the process.

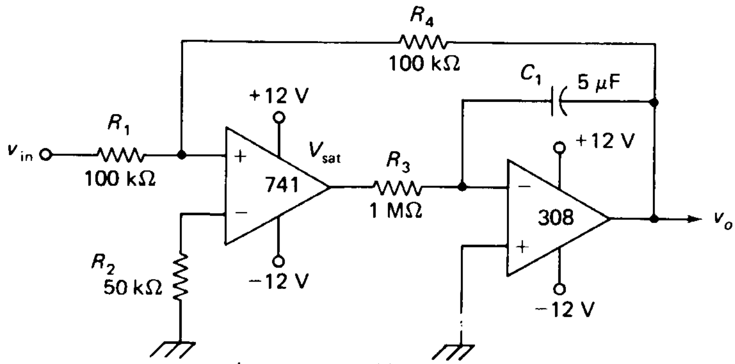
Staircase Generator: A square wave, when differentiated, will produce a string of positive and negative pulses. If the pulses are rectified and integrated, a staircase output will result. Figure 19-10(b) shows the implementation of this concept. A precision rectifier IC_1 is used to avoid the temperature sensitivity of a simple diode. Integrator IC_2 is reset by hysteresis switch IC_3.

19.7 ANALOG AND DIGITAL CONVERSION

Voltage Comparator: The circuit of Fig. 19-11(a) will output a positive saturation level (about +9 V for ±10-V supplies) if V_1 is greater than V_2, and a negative saturation level if V_2 is greater than V_1.

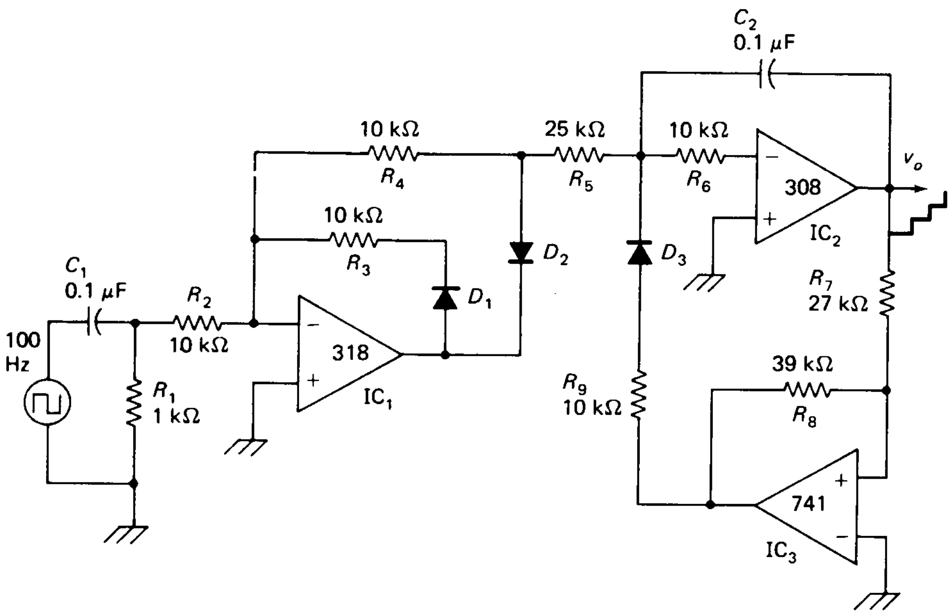
Schmitt Trigger: The circuit of Fig. 19-11(b), also called a *hysteresis switch*, is similar to the voltage comparator except that it tends to remain in its present state until an overvoltage at one input forces it to snap to the other state. The circuit is most often used with V_2 at ground, but in any case V_2 should not be allowed to become a large fraction of the supply voltage. The hysteresis or dead zone is calculated as

$$V_{dz} = \pm V_{\text{sat}}\frac{R_1}{R_3} = \pm 9\ \text{V}\frac{100\ \text{k}\Omega}{1\ \text{M}\Omega} = \pm 0.9\ \text{V} \tag{19-2}$$



$$\frac{dv_o}{dt}\,(\text{max}) = \frac{V_{sat}}{R_3C_1}$$

(a)



(b)

***FIGURE 19-10** (a) Rate limiter: v_o follows v_{in} except that the v_o rate or change cannot exceed the limit set by R_{3L} and C_{L1}. (b) Staircase generator composed of differentiator C_1R_1, perfect rectifier IC_1, integrator IC_2, and reset switch IC_3.*

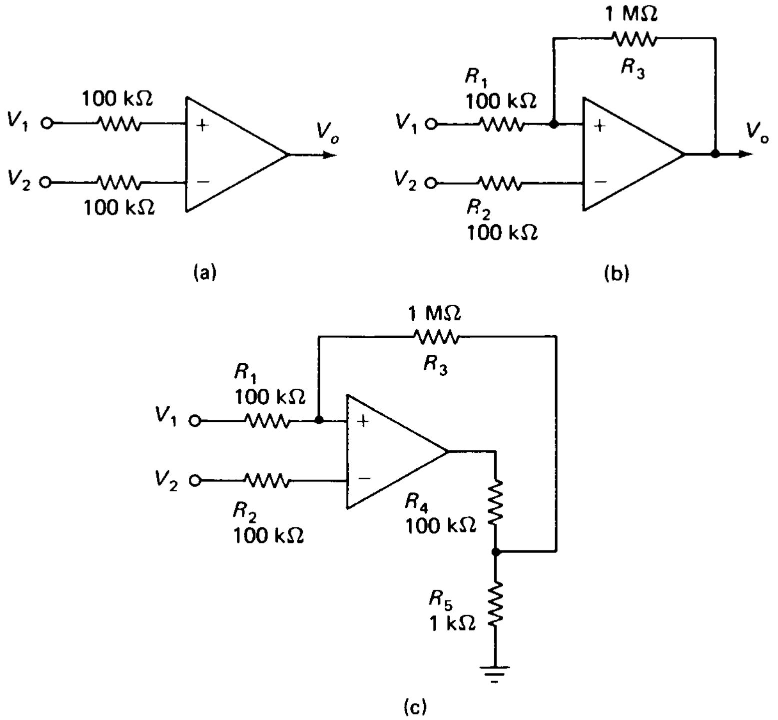


***FIGURE 19-11** (a) Simple voltage comparator. (b) Positive feedback forms a hysteresis "snap-action" switch. (c) Voltage divider* R_4R_5 *gives smaller dead zone with reasonable values of* R_3.

For the values given in Fig. 19-11(b) this means that the output will remain in its current state as long as V_1 does not differ by more than ± 0.9 V from V_2.

Very small dead zones are often desirable, but often result in unrealistic values of R_3. Figure 19-11(c) shows a method for reducing dead zone by a factor of $(R_4 + R_5)/R_5$.

Schmitt triggers find a wide variety of applications. They may be used as:

- Level detectors, to activate a relay or light when a preset voltage is exceeded, and leave it activated until the level drops significantly below the set level
- Snap actuators, to prevent a relay from mushing in uncertainly on levels near the actuate point
- Bounce eliminators in conjunction with an *RC* filter to prevent a separate actuation for each bounce of a switch contact
- Square-wave generators, to produce two-state digital outputs from sine or similar analog inputs

Digital-to-Analog Conversion can be accomplished with the simple summing circuit of Fig. 19-12. The values shown give an output of 1 V/count, although 0.1 V/count could be obtained by making $R_f = 8.0\ \text{k}\Omega$. The 1s through 8s lines are the outputs of a binary or BCD counter, such as the 7490. The 4016 quad analog switch is used, gated by the binary outputs, because the counter outputs themselves have levels which may vary by a few tenths of a volt among outputs and with temperature. The CD4016 actually contains a pair of insulated-gate FETs for each gate. The input-to-output assumes a very low resistance (around 100 Ω) when the control line is high, and a very high resistance (many MΩ) when the control line is low.

More inputs can be added for 16-s, 32-s, and so on, but as the summing resistor values become lower, their tolerance becomes more and more critical. A 3% error in the 32-s resistor will completely swamp out the contribution of the 1-s resistor. Integrated D/A converters are available with 8-bit (256 count) resolution.

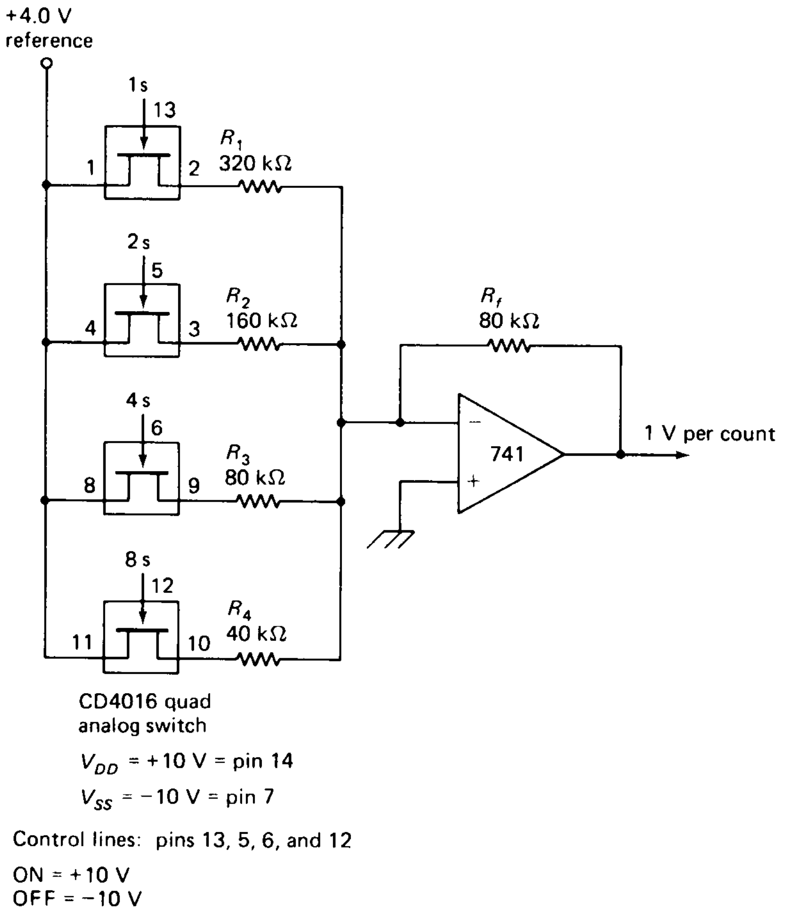


FIGURE 19-12 Digital-to-analog converter using precision resistors and the CD 4016 CMOS IC.

Analog-to-Digital Conversion is a bit more of a challenge than D/A conversion. Many methods have been developed for A/D conversion, but we will describe only two: a simple one and the one most widely used in DVM circuits.

The Continuous-Search A/D Converter, shown in Fig. 19-13(a) incorporates the D/A converter of Fig. 19-12. The analog output is compared to the input voltage, and the binary counter is directed to count up or down, as required to equalize the comparison. Of course, the counter always counts up or down at least one count each clock pulse, and this instability can be quite annoying in a visual display. A timing gate on the clock (1 s hold, 0.1 s count) will alleviate this problem. This converter can be made quite fast, but its accuracy is limited to about $\frac{1}{4}$% by the D/A converter. A slower version eliminates the up/down counter, resetting to zero and counting all the way up each display cycle.

The Dual-Slope A/D Converter of Fig. 19-13(b) is the basis of most commercial digital voltmeters. It can be made highly accurate, since its basic function is to measure the ratio between the input voltage and an internal reference voltage. The tolerance and stability of all the resistors, capacitors, and semiconductors in the circuit do not affect the final accuracy of the conversion—only the accuracy of the reference supply is important.

At the start of the conversion cycle the flip-flop is reset, so Q_2 is turned on by the low level at its gate, while Q_1 is turned off by +5 V at its gate. The integrator thus generates a negative ramp at a rate of 5 V/s for a +0.5-V input. As soon as this ramp crosses zero, the comparator output goes positive, enabling the AND gate and starting the counter. Let us assume that the counter is a three-stage BCD type counting to 999. Also, let the clock rate be 5 kHz. The integrator will thus reach −1 V output in the 0.2 s it takes for the counter to reach maximum count.

The highest-order bit of the counter feeds an overflow flip-flop which is set as the counter advances from 999 to 000. This sets the control flip-flop, disconnecting V_{in} and connecting V_{ref} to the integrator via Q_1.

Note here that the time required for the count from 000 through 999 and back to 000 depends upon the clock frequency and is relatively constant. However, the slope of the integrator output, and hence the voltage reached as the counter rolls back to 000, depends directly upon V_{in}.

Now the counter begins counting again up from zero, but this time the integrator output is rising at a rate of 10 V/s, as determined by V_{ref}. The run back to zero from −1 V will take 0.1 s, and at 5 kHz the counter will read 500 at the zero crossing. The counter will stop at this point because a positive integrator output will set the comparator output negative, disabling the AND gate to the counter. This value will be displayed until the UJT Q_3 fires (after a second or so), resetting the control flip-flop and starting a new cycle.

If the integrator resistor or capacitor should be a little low in value, the negative ramp would be a little steeper, and the peak voltage would be a little higher, but this would be of no consequence because the positive ramp would be a little steeper by the same percentage. Likewise, if the oscillator should run a little

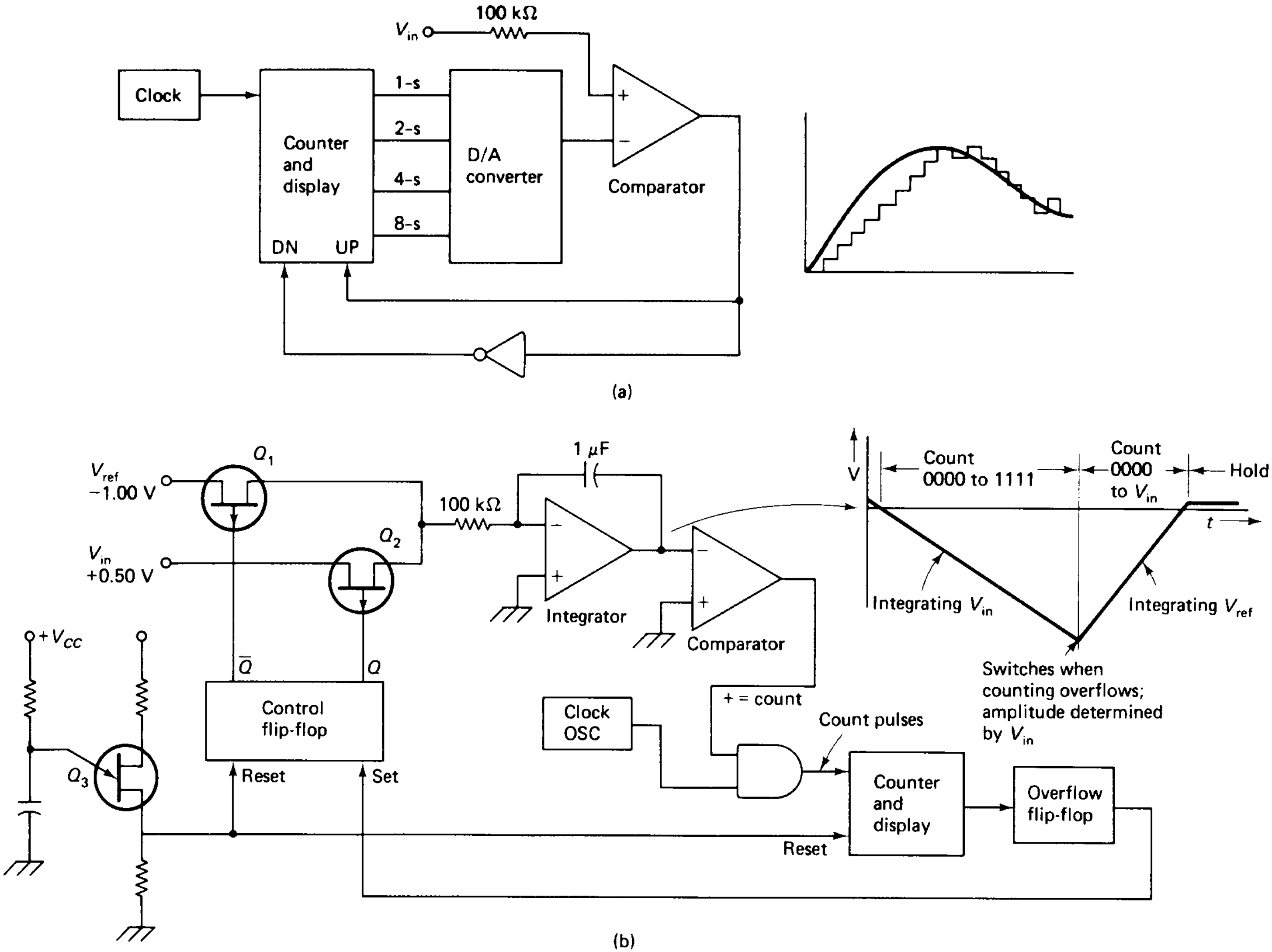


FIGURE 19-13 *Two approaches to analog-to-digital conversion: (a) continuous search using the D/A converter and an up/down counter; (b) dual-slope conversion, which is the basis of most DVMs.*

slow, the time to reach the negative peak would be a little longer, but the time to return to zero would be longer by the same percentage, and the final count would be unaffected.

19.8 VOLTAGE AND FREQUENCY CONVERSIONS

Frequency-to-Voltage Conversion is a simple matter of generating standard-length pulses at a rate equal to the input frequency and averaging the value of the pulses with an *RC* filter. Figure 19-14 shows one approach using a UJT and another using an IC one-shot. F/V converters form the basis for tachometers and frequency meters.

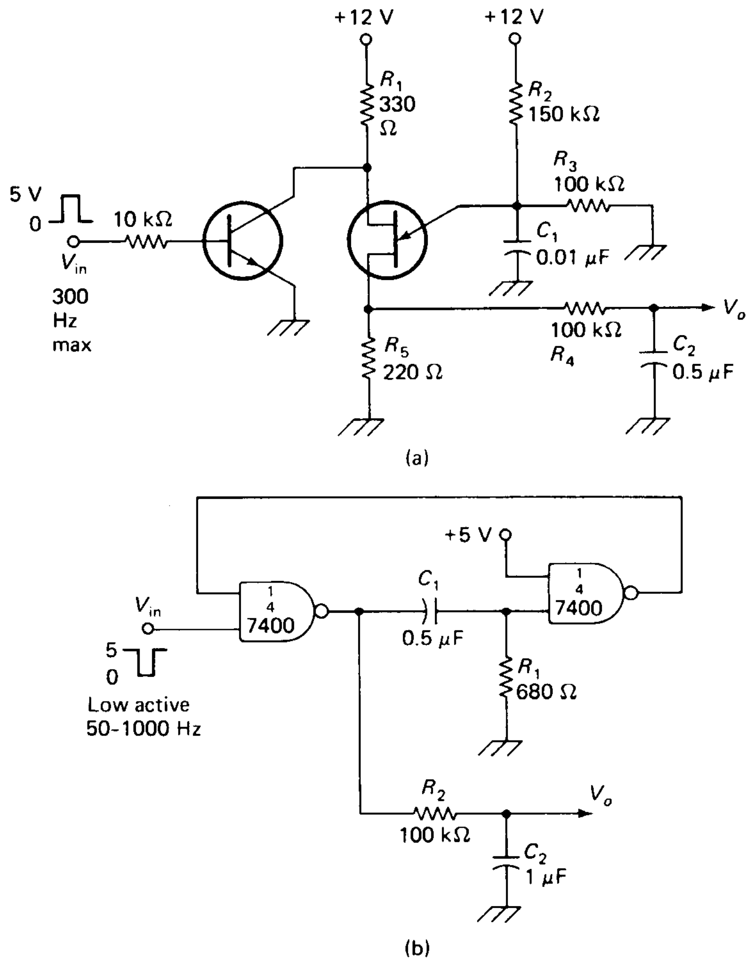


FIGURE 19-14 Frequency-to-voltage conversion: standard pulse widths are generated at the input trigger rate and averaged by an RC filter.

Voltage-to-Frequency Conversion using the *control* input of a *555* oscillator has already been encountered in Fig. 19-4(c). Better linearity can be achieved with an integrator driving a *555* level detector, as shown in Fig. 19-15. For the values given, IC_1 will integrate at $+1$ V/ms for a -1-V input, reaching the *555* threshold in 6.7 ms. The output of IC_2 will then go low, turning Q_1 on and discharging C_1. Discharge will continue until the trigger input of the *555* goes below 3.33 V. R_4 is adjusted so that this will happen as nearly as possible to zero output from the integrator. The discharge time is a negligible fraction of the 6.7-ms charge time.

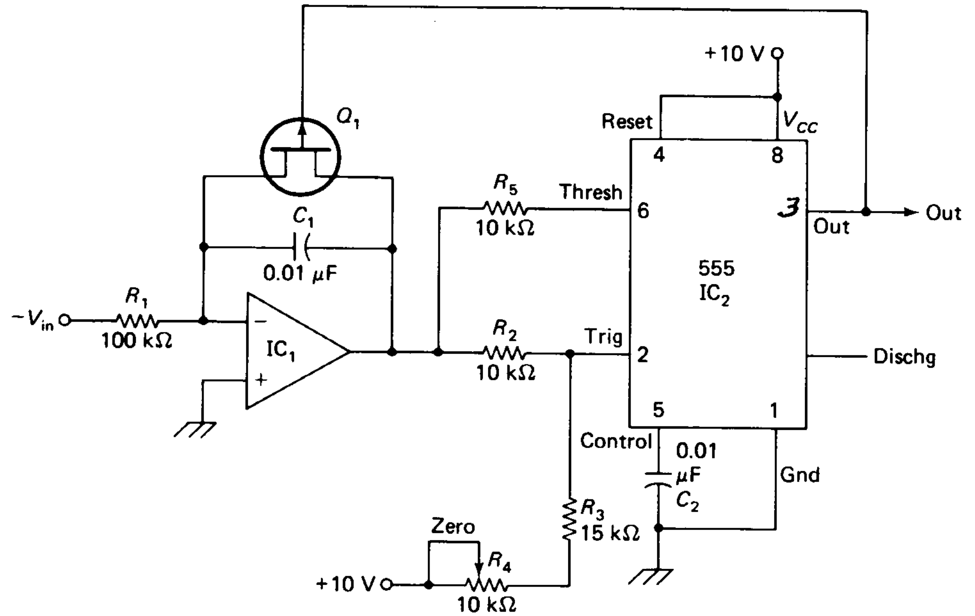


FIGURE 19-15 Voltage-to-frequency conversion: integrator IC_1 reaches threshold 2/3 V_{CC} in a time determined by V_{in}, whereupon Q_1 resets the integrator.

Once pin 2 of IC_2 goes below 3.33 V, the output goes high and Q_1 turns off, initiating another integration period. The integration time to the 6.7-V threshold is directly proportional to V_{in}. In this case, with $V_{in} = -1$ V, the output frequency is 1/6.7 ms, or 150 Hz.

19.9 SIGNAL GENERATION

Oscillators are of three basic types:

- Feedback oscillators feed the output signal in phase back to the input in a regenerative process. The basic requirements here are a total phase shift around the loop of exactly 0° or 360°, and a total loop gain ($A_v \times$ feedback ratio) equal to unity. Loop gains greater than unity produce progressively more severe distortions of the sine-wave output. Frequency is determined by the phase-shifting components, which may include the input and output resistances of the active device in the amplifier. These parameters vary with bias point, making supply voltage a determiner of frequency to a small extent. The phase shift of a tuned circuit varies abruptly from inductive to capacitive around resonance, so transistor parameters cause less frequency shift in a tuned-circuit oscillator than in a phase-shift type. Higher-Q tuned circuits produce better stability. The quartz crystal, with Q's above 20,000, stands as the paragon of stability. FET inputs with emitter-follower buffers before the feedback loop will keep the input resistance near infinity and the output resistance near zero, reducing the effects of the active devices.

- Relaxation oscillators charge a capacitor until some trigger device reaches its threshold, whereupon the capacitor is discharged, and the cycle repeats. Common trigger devices include neon lamps, UJTs, transistor switches, IC logic gates, and op amps. Relaxation oscillators generally produce ramp and square-wave outputs. Frequency is dependent upon the capacitor value, the charging rate, and the trigger threshold. These last two factors are often dependent upon power-supply voltage and/or temperature, making a high degree of stability difficult to obtain.
- Negative-resistance oscillators incorporate an LC tuned circuit and a suitably biased negative-resistance device to cancel the positive resistance representing coil losses and energy coupled to the load. The result is a zero-resistance LC circuit which maintains oscillations without damping over the negative-resistance range of the device. Bipolar transistors, UJTs, diacs, and other devices have negative-resistance regions, but maintaining a stable bias point within them is difficult. The tunnel diode is easily biased in its negative-resistance region, but the dynamic range (and hence oscillator output) is limited to a few tenths of a volt.

A Wien Bridge Oscillator, which retains the frequency stability and sine-wave purity of an LC-tuned oscillator while dispensing with the inductor, is shown in Fig. 19-16. At the frequency where $X_{C1} = R_1$ and $X_{C2} = R_2$, the Wien bridge forms a 21:1 voltage divider with both series and parallel arms having a 45° current-to-voltage phase angle. The feedback is therefore positive, and a gain of 21 will make the circuit oscillate. At other frequencies the series and parallel arms will have differing phase angles, producing a feedback that is not exactly in phase.

R_3 and the drain resistance of Q_1 must have a ratio of 20:1 to produce the required gain. This is accomplished by sensing the output waveform and using it to change the gate bias, and hence the dynamic drain resistance of the FET. Output peaks above -5.3 V charge C_3, applying a negative dc to the gate of the FET, raising its drain resistance and lowering the gain of the amplifier. This tends to lower the output peaks. In fact, the output peaks will be greater than ± 5.3 V by an amount equal to the V_{GS} required to produce a drain resistance of 500 Ω.

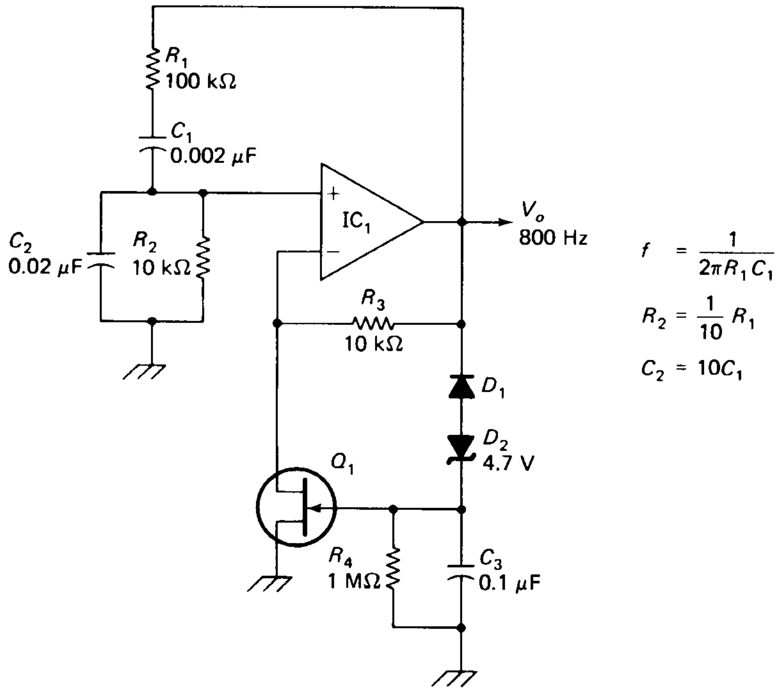


FIGURE 19-16 *Wien-bridge oscillator. R_3 and Q_1 determine the gain of IC_1. Q_1 is biased off by excessive output, lowering gain to prevent distortion.*

A Phase-Shift Oscillator with two sine-wave outputs 90° apart (quadrature outputs, they are called) is shown in Fig. 19-17. IC_1 is an integrator, which assures that the output phase lag is exactly 90°. This is so because the integral of the sine is the negative cosine, which is a 90° lag. The integrator is followed by two *RC* phase-shift networks, each providing 45° lag at the frequency where $X_C = R$. The total 180° lag, combined with the inversion of the integrator, produces an in-phase feedback.

The wave is limited by D_1 and D_2 after the first phase-shifting section to make the feedback adjustment R_5 a little less critical. There are what amounts to two low-pass-filter sections between these limiters and the IC output, which will keep the resulting harmonic distortion at a low level. Without D_1 and D_2 the IC output would go to its saturation limits, producing heaviest distortion at the output.

A Relaxation Oscillator which amounts to a simple function generator is shown in Fig. 19-18(a). It consists of a Schmitt-trigger level detector (Fig. 19-11) driving an integrator (Fig. 19-8). The integrator output is driven positive by the negative saturation level output of IC_1, applied through voltage divider R_3R_4 and current limiter R_5. When the positive ramp output on R_1 overcomes the negative saturation level on R_2, the Schmitt-trigger output switches to positive saturation, ending the positive ramp and starting a negative ramp from the integrator.

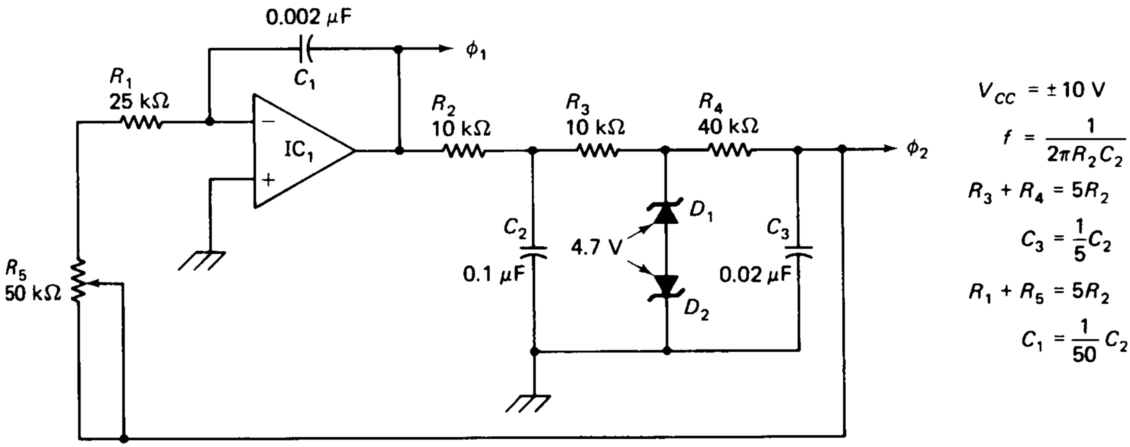


***FIGURE 19-17** A quadrature oscillator has two outputs, 90° out of phase. Integrator IC_1 guarantees exactly 90° shift since the integral of the cosine is the sine. R_5 is adjusted for purest sine wave with D_1 and D_2 just conducting.*

R_3 varies the output frequency over a 20 : 1 range, and C_1 can be switched by decades from 10 μF to 0.001 μF to produce an output range from 0.5 Hz to 100 kHz.

The resistor and diode network from the triangle output places successively heavier loads on R_7 as the triangle reaches its peaks, thus loading it down to an approximate sine-wave shape. With only two break points on each half-cycle the segmentation of the "sine" wave is rather obvious (about 2% harmonic distortion), but commercial units with five or more break points bring the distortion down to 0.2% and less.

A Voltage-Controlled Function Generator is shown in Fig. 19-18(b). Analysis shows that current to the inverting input with Q_1 *on* is equal in magnitude but opposite in sign to the current with Q_1 *off*. The following calculations are done referenced to the −10-V supply as a temporary "zero." Circuit ground thus becomes +10 V in the calculations. This simplifies the arithmetic, since the voltage dividers are all referenced to −10 V:

$$V_{\text{inv}} = V_{n\text{inv}} = 0.50V_c$$

$$V_{\text{th(off)}} = V_c$$

$$I_{\text{(off)}} = \frac{V_{\text{th}} - V_{\text{inv}}}{R_1} = \frac{V_c - 0.50V_c}{R_1} = \frac{+0.50V_c}{100\text{ k}\Omega} = +5V_c\,(\mu\text{A})$$

$$V_{\text{th(on)}} = 0.33V_c$$

$$R_{\text{th(on)}} = R_1 \| R_4 = 100\text{ k}\Omega \| 50\text{ k}\Omega = 33\text{ k}\Omega$$

$$I_{\text{(on)}} = \frac{V_{\text{th}} - V_{\text{inv}}}{R_{\text{th}}} = \frac{0.33V_c - 0.50V_c}{R_1 \| R_4} = \frac{-0.17V_c}{33\text{ k}\Omega} = -5V_c\,(\mu\text{A})$$

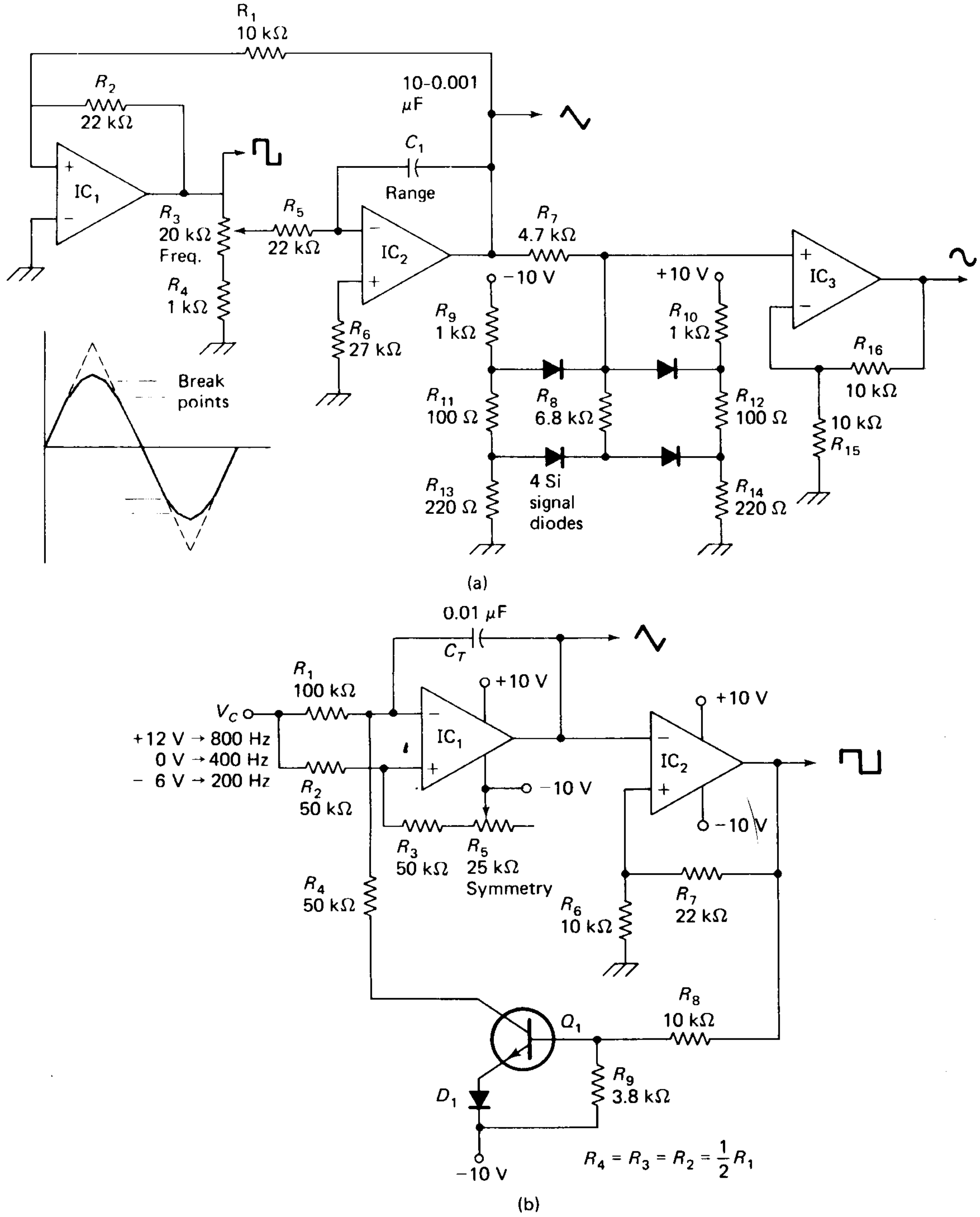


FIGURE 19-18 *(a) Variable-function generator with sine, square, and triangle-wave outputs. (b) Voltage-controlled function generator. V_c determines output frequency.*

Returning the reference to circuit ground, this means that $V_c = -10$ V produces a zero integrator current and zero frequency (ideally), -9 V produces f, -8 V produces $2f$, -5 V produces $5f$, 0 V produces $10f$, and $+5$ V produces $15f$.

19.10 THE PHASE-LOCKED LOOP

The phase-locked loop has appeared in sophisticated communications systems for quite a few years, but its availability as an inexpensive integrated circuit has led to its application in entertainment receivers and a widening variety of noncommunications equipment.

Integrated phase-locked loops (PLLs) have two distinctly different applications:

- They can be used to produce an output voltage level that varies with the frequency of the input signal. In some devices the V-versus-f curve is extremely linear over a limited range, although it is not zero-based (zero frequency does not produce zero voltage). In this mode PLLs are commonly used as FM detectors and FSK (frequency-shift keying) Teletype demodulators.
- They can be used to produce an output frequency that is locked to the frequency of an input signal—even though the input may be modulated, badly distorted, covered with noise, or even interrupted. The output wave, of course, will be perfectly clean, and will follow the input frequency over a limited range if it should vary. In this mode, PLLs are used as frequency synthesizers, automatic frequency-controlled oscillators, and signal reconstructors.

A phase-locked loop contains a voltage-controlled oscillator, a phase detector, and a low-pass filter. Of the three, only the phase detector requires explanation at this point.

A Phase Detector is a circuit designed to produce an output proportional to the phase difference between two input signals. Ideally, if $f_i = f_o$ the output would be dc —positive if f_i leads f_o, negative if f_i lags f_o. The magnitude of the dc would be directly proportional to the phase difference. If f_i were slightly different from f_o, the phase detector output would be a low-frequency ac (frequency $= f_i - f_o$) going positive and negative as f_i and f_o fell in and out of phase.

Real phase detectors are not quite so clean. They are actually a form of multiplier ($v_i v_o$) or mixer, similar to the mixer in a superheterodyne receiver. The outputs actually obtained are the difference ($f_i - f_o$), the sum ($f_i + f_o$), and, if the mixer is not perfectly balanced, the inputs f_i and f_o. To make matters worse, f_o is usually a square wave, and f_i is processed through an amplifier that generally saturates, making f_i a square wave. This means that 3rd, 5th, 7th, and higher odd harmonics and all their possible sums and differences are also present in the output. The only salvation in all this confusion comes from the fact that, if f_i differs from f_o by no more than 10%, none of the harmonics below the 19th and 21st will be able to mix with anything else to produce a frequency as low as $f_i - f_o$. For a 5% difference the lowest harmonics that can cause trouble are the 39th and 41st, and for a 1% difference it becomes the 199th and 201st. A low-pass filter can thus be used to remove the $f_i + f_o$ component, and all the components due to significantly strong harmonics.

An Integrated Phase-Locked Loop, the type *565*, is shown in Fig. 19-19. Comments on the characteristics of each part of the device may be appropriate before explaining some of its applications.

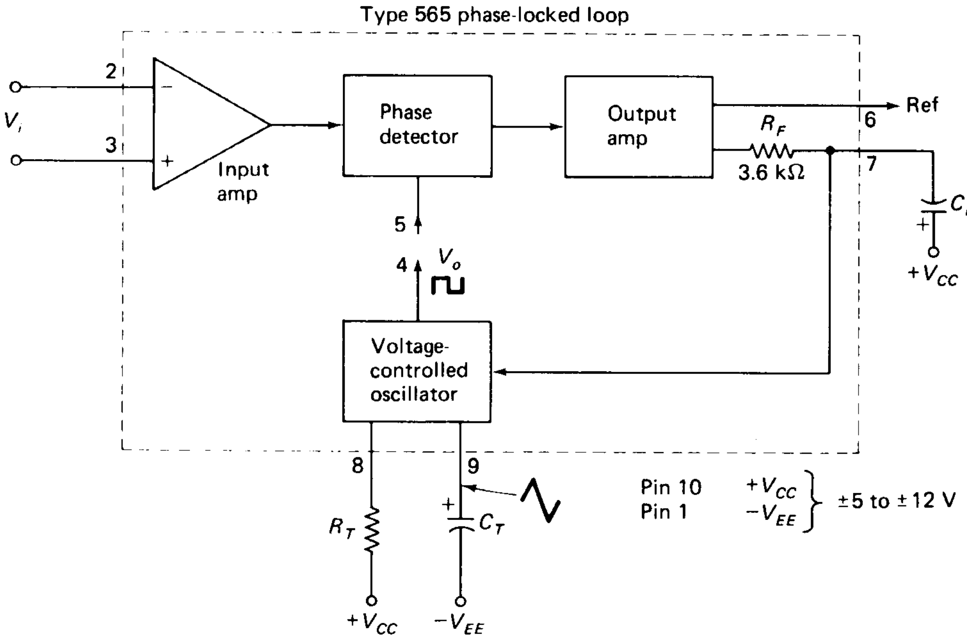


FIGURE 19-19 *The type 565 phase-locked loop: internal functions and pin designations.*

The Input Signal may be differential, with common-mode limits of +1 and −4 V, although pin 3 is usually grounded. Input impedance is typically 10 kΩ between inputs, but much higher to ground. Minimum input is about 25 mV p-p, with saturation a certainty by 200 mV p-p. Bias and offset currents are in the μA range and can easily shift one input more than 100 mV from the other, saturating the amplifier and locking out V_i. High source impedances should therefore be avoided. If ac coupling is used, both inputs should be tied to ground through 4.7 kΩ.

The Phase Detector produces a voltage at pin 7 which is referenced to $+V_{CC}$ and ranges from about 1 to 3 V below this level for V_i to V_o phase differences from 30°

to 150°, respectively. It is driven by square-wave V_o (pin 4 connected to pin 5) which swings from about -0.2 to $+5.0$ V, for a ± 6-V supply.

The Low-Pass Filter is most often a single-section *RC* type. The resistance is an integrated 3.6 kΩ since this allows temperature tracking with the VCO components. The capacitor is external and is selected to filter out $f_i + f_o$ without attenuating $f_i - f_o$. The capacitor is referenced to the $+V_{CC}$ supply because this is the reference for the phase detector and the VCO. Grounding the capacitor would introduce power-supply noise and ripple into the VCO input. More complex filtering is possible by feeding from pin 7 to the filter to a current source to pin 8. Manufacturer's application notes describe this technique.

The Voltage-Controlled Oscillator operates up to 0.5 MHz. (The type *560* PLL operates up to 30 MHz.) Oscillation frequency is given approximately by

$$f_o \approx \frac{0.3}{R_T C_T} \tag{19-3}$$

This frequency will be multiplied by approximately 0.5 with the VCO input (pin 7) 1 V below $+V_{CC}$, and by about 1.5 with pin 7 at 3 V below $+V_{CC}$. This is a total voltage-control range of 3 : 1. It is recommended that the timing resistor be restricted to the range between 2 and 20 kΩ. A 2.5-V-p-p triangle wave is available at pin 9 ($V_{CC} = \pm 6$ V), although loading this output will slow and eventually stall the VCO. The timing capacitor can be referenced to ground, or if an electrolytic is required, to $-V_{CC}$ to avoid reversing polarity voltages on it.

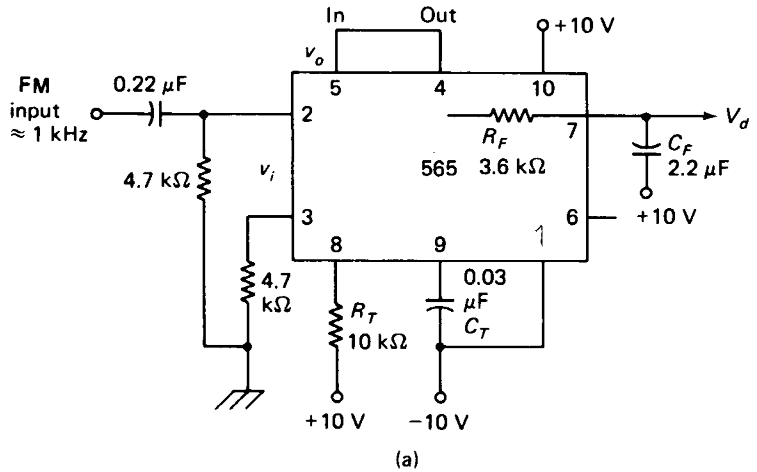
Single-Supply Operation of the *565* PLL is possible if the inputs are biased about 1 V lower than $\frac{1}{2} V_{CC}$ through resistances of 10 kΩ or less. The square- and triangle-wave outputs will not cross, or even come near ground level in this case, but that is often of no consequence.

An FM Detector using a PLL is shown in basic form in Fig. 19-20(a). Capacitor C_T is determined from equation 19-3 to make the VCO oscillate at a center frequency equal to the FM input with R_T chosen as 10 kΩ:

$$C_T = \frac{0.3}{R_T f_o} = \frac{0.3}{10\ \text{k}\Omega \times 1\ \text{kHz}} = \mathbf{0.03\ \mu F}$$

The filter capacitor is chosen to attenuate the $f_i + f_o$ output of the phase detector (2 kHz) by a factor of 100. This places the f_c breakpoint of the Bode plot at 20 Hz, which means that the output of the detector will be severely attenuated if the 1-kHz carrier is modulated at a rate higher than 20 Hz.

$$C_F = \frac{1}{2\pi R_F f_c} = \frac{1}{2\pi \times 3.6\ \text{k}\Omega \times 20} = \mathbf{2.2\ \mu F}$$



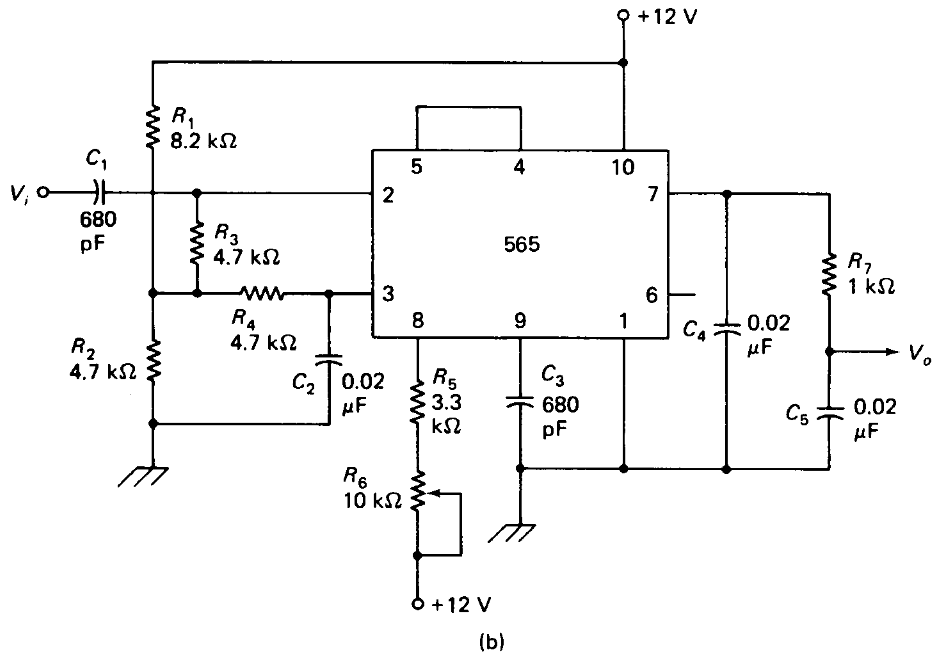


***FIGURE 19-20** (a) FM detector suitable for testing the 565 PLL. (b) FM detector for SCA music subcarrier detection.*

In operation, the VCO will follow the frequency of the FM input. This is so because as f_i increases, putting the phase of V_i ahead of V_o, the phase detector produces a more-negative output, which increases f_o until the two frequencies match. Similarly, a decrease in f_i will raise V_d, lowering the VCO frequency f_o to match f_i. V_d is thus a demodulated output. Note that the output impedance is 3.6 kΩ.

Figure 19-20 is a good circuit to build to become familiar with basic PLL characteristics. It can be observed that the waveshape and amplitude of I_i (down to 10 mV or so) have no effect on V_d. Two signal generators can be placed in series at I_i to demonstrate that noise and interfering signals are ignored as long as they are not too strong and too near f_o. A graph of f_i versus V_d can be constructed to discern the linearity of the FM-detection process.

Lock Range is the span of frequencies above and below f_o for which the VCO will follow f_i. For the *565* this is typically $\pm 50\%$ around f_o. A more precise expression is

$$f_{\text{lock}} \approx \frac{16 f_o}{V_{CC}} \tag{19-4}$$

where f_{lock} is the span between upper and lower lock frequencies, f_o is the VCO center frequency, and V_{CC} is the total supply voltage between pins 1 and 10.

Capture Range is the span between the upper and lower frequencies at which the VCO will snap into lock if f_i is brought toward f_o from far outside the lock range. Capture range is narrower than lock range, and becomes narrower yet as the time constant of the filter $R_F C_F$ is increased:

$$f_{\text{cap}} \approx \sqrt{\frac{10 f_o}{R_F C_F V_{CC}}} \tag{19-5}$$

For the values in Fig. 19-20(a), this yields

$$f_{\text{cap}} \approx \sqrt{\frac{10 \times 1000}{3.6\ \text{k}\Omega \times 2.2\ \mu\text{F} \times 20\ \text{V}}} \approx \mathbf{250\ Hz}$$

This means that VCO center frequency f_o must be within ± 125 Hz of f_{in}, or the VCO will be unable to lock to f_{in}.

Sometimes it is desirable to decrease the lock and capture ranges of a PLL. Tone decoders (touch-tone dialing) must have a narrow lock range. Where several signals are present at the input, a wide capture range makes the PLL apt to lock in on the wrong one. A resistance of 5 kΩ between pins 6 and 7 of the *565* will reduce the lock range by approximately one-half. Shorting pins 6 and 7 will cut the range to about one-fourth.

Figure 19-20(b) shows a practical FM detector for the 67-kHz SCA (subsidiary communications authorization) subcarrier that is broadcast by some FM stations. This subcarrier is available at the detector of a regular FM receiver, but is generally attenuated through the audio-amplifier stages because of its high frequency. This service is sold on a subscription basis and carries background music suitable for offices, shopping areas, waiting rooms, restaurants, and so on. Note that it is illegal

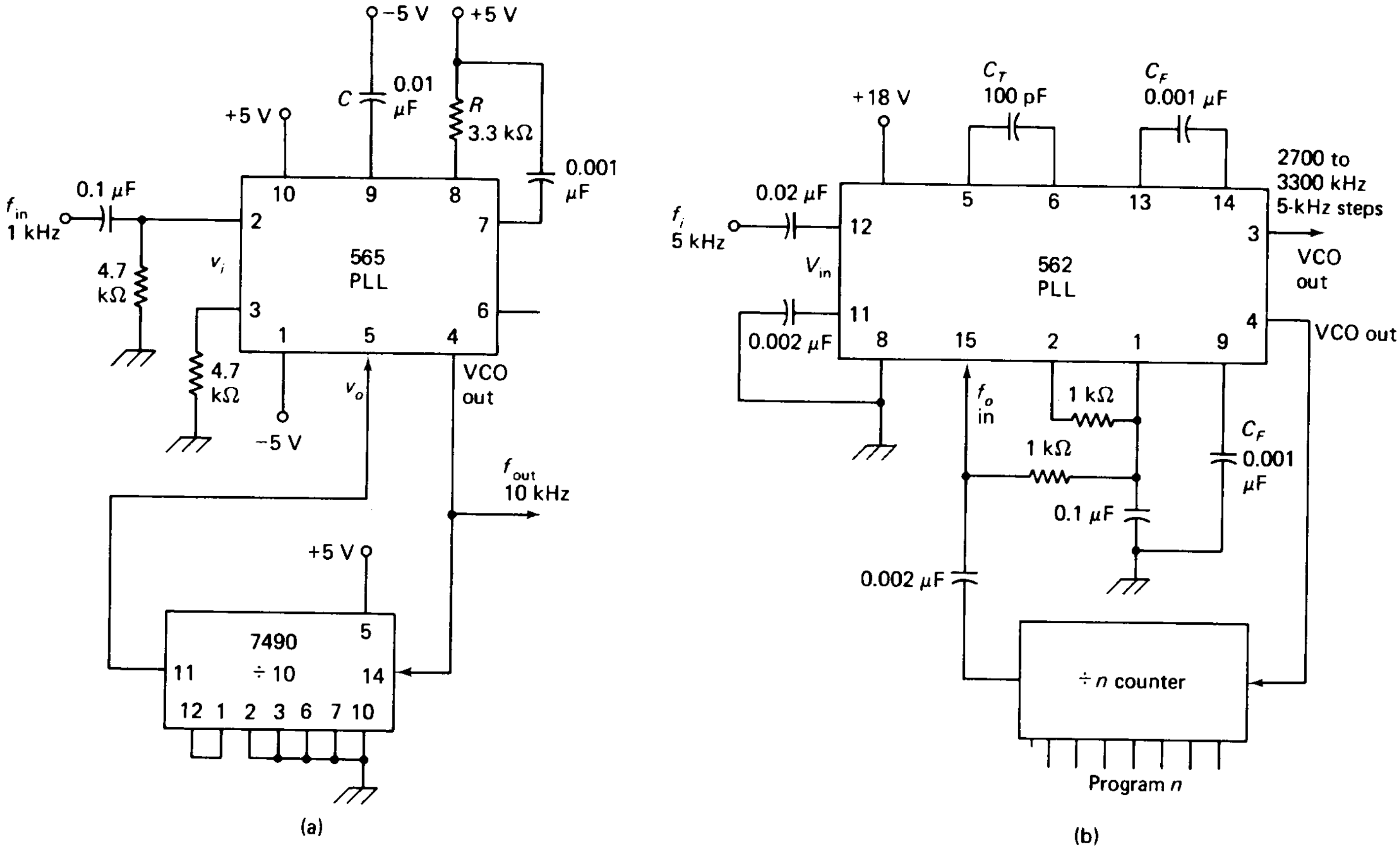


FIGURE 19-21 *(a) Frequency synthesizer suitable for lab tests and demonstration. (b) Frequency synthesizer for covering 2700 kHz to 3300 kHz in 5-kHz steps.*

to use this service on a public or for-profit basis without paying the subscription fee.

C_1 and $R_3 \| R_{in}$ attenuate the high-level main-channel modulation components below 10 kHz. R_1 and R_2 bias the inputs slightly below $\frac{1}{2} V_{CC}$, providing single-supply operation. C_3, R_5, and R_6 are chosen to set the center frequency of the VCO at 67 kHz. C_4 and the internal 3.6-kΩ filter resistor provide a capture range of 27 kHz and at the same time deemphasize the audio frequencies above 2100 Hz in compensation for the preemphasis used in broadcasting. R_7 and C_5 give a high-frequency rolloff starting at 8 kHz, which is the upper limit of the audio output.

A Signal Restorer can be built using the circuit of Fig. 19-20(a) with the output taken from pins 4 and 5 (square wave) or pin 9 (triangle). The output will be a pure waveform in spite of noise, modulation sidebands, or interfering signals at the input. The VCO can also be made to lock at odd submultiples of f_i ($f_o = \frac{1}{3} f_i$, $\frac{1}{5} f_i$, $\frac{1}{7} f_i$, etc.) although the lock range becomes progressively narrower. Locking at odd multiples ($f_o = 3f_i$, $5f_i$, etc.) is also possible if V_i drives the input to saturation, because the resulting square wave contains these odd harmonics. Locking to even harmonics of f_i is possible if the input is nonsymmetrical or if the pin 3 input is dc shifted away from the pin 2 input to produce nonsymmetrical switching on a saturating input. Again, harmonic locking gives a narrower lock range than equal-frequency locking.

Frequency Synthesis is achieved by placing a digital counter in the feedback loop, as shown in Fig. 19-21(a). Note that, in this example, the VCO runs at 10 kHz but the input signal and the phase detector operate at 1 kHz. The output signal has the same frequency stability as the input signal, but is n times the frequency, where n is the division ratio of the counter.

Figure 19-21(b) shows a more elaborate frequency synthesizer using a *562* high-frequency PLL and a divide-by-n programmable counter. The values shown give over 120 frequencies in the range 2.7 to 3.3 MHz all at integer multiples of the 5-kHz input. Additional $\pm 10\%$ bands can be covered by simply changing C_T. Frequency out is always f_i/n. A single-crystal oscillator can, by this means, yield hundreds of crystal-stable frequencies.

20

HIGH-PERFORMANCE POWER SUPPLIES

20.1 POWER-SUPPLY SPECIFICATIONS

Reasonably good power supplies can be built with the simple zener diode, zener follower, and feedback transistor amplifier circuits of Fig. 20-1, and many more-complex discrete designs are also possible. However, the monolithic integrated circuit makes the superior performance of highly sophisticated regulators so inexpensive, reliable, and easy to implement that IC regulators are now more common than discrete types.

Before examining how these regulators can be applied, we will first define some of the terms by which regulators are specified and compared.

Ripple is normally specified by rms ac as a percentage of dc in the output at maximum rated load current:

$$\text{ripple factor} = \frac{V_{o(\text{rms})}}{V_{O(\text{dc})}} \tag{20-1}$$

It is common practice to measure $V_{o(p\text{-}p)}$ on an oscilloscope and divide by 2.828 to get rms, ignoring the error incurred if V_o is not a sine wave.

Ripple Rejection is the factor by which a regulator attenuates ripple in the raw ac fed to it from the rectifier and filter. It is usually specified as V_o/V_{in} in mV/V.

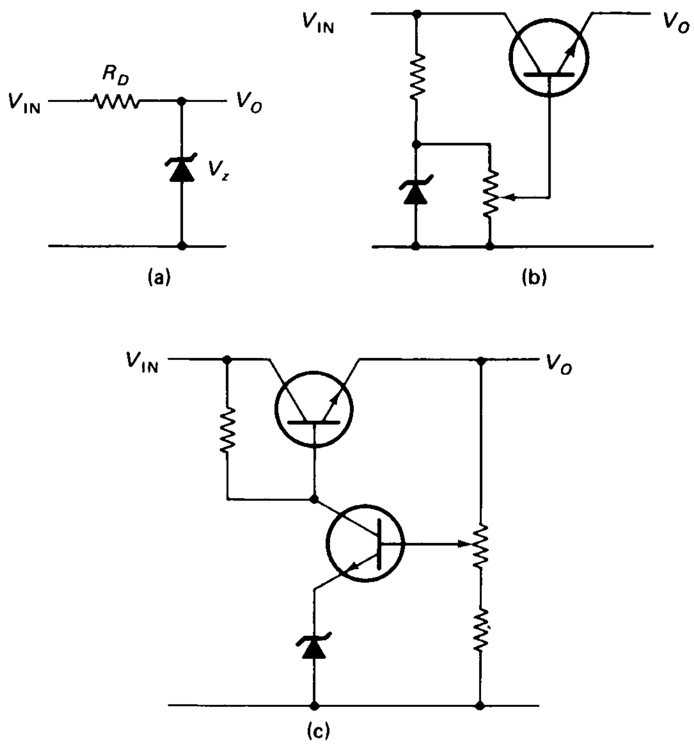


***FIGURE 20-1** Representative discrete regulator circuits: (a) fixed-voltage zener; (b) zener with emitter follower; (c) error amplifier and series pass transistor.*

Load Regulation is the percent rise in dc output voltage when the load is removed. The load is chosen to draw the maximum rated output current at maximum output voltage:

$$\text{load regulation} = \frac{V_{O(\mathrm{NL})} - V_{O(\mathrm{FL})}}{V_{O(\mathrm{FL})}} \tag{20-2}$$

where FL and NL stand for full load and no load, respectively.

Line Regulation is the factor by which the output voltage is stabilized against changes in input voltage. A line regulation of 20% means that a 10% line voltage change will produce only a 2% load-voltage change (10% × 20% = 2%).

$$\text{line regulation} = \frac{\%\ \text{load voltage change}}{\%\ \text{line voltage change}}$$

$$= \frac{(V_{OH} - V_{OL})/V_{OL}}{(V_{IH} - V_{IL})/V_{IL}} \tag{20-3}$$

where O, I, H, and L stand for out, in, high, and low, respectively. For 117-V-input supplies, high and low inputs of 121 V and 110 V are recommended to produce a line change of 10%.

Output Impedance is the ratio of output-voltage change to output-current change for a fixed output setting and a changing load. This is already specified at low frequencies by load regulation, but at high frequencies the response time of the regulator may be too slow to maintain the low-frequency Z_o.

Temperature Coefficient is the change in output voltage per degree temperature change. It may be expressed in mV/°C or %-of-V_O/°C. Alternatively, the total temperature stability may be specified for the entire operating range of temperatures, for example:

Temperature stability: 1.2% max, −25 to +83°C

20.2 GENERAL-PURPOSE IC REGULATORS

Quite a number of IC regulators are available, but their internal operation and external connections are generally similar. Figure 20-2 illustrates the basic regulator philosophy and shows the four sections included in most IC types.

The Voltage Reference is generally derived from a 7-V zener diode because the temperature coefficient is near zero at this voltage. The zener is fed by a transistor current source from the raw V + supply and is buffered by an emitter follower and compensating diode. Most regulators voltage divide the reference down to about 1.7 V before buffering to make it easier to obtain V_O below 7 V.

The Differential Amplifier compares a fraction of the regulated output voltage with the reference voltage and delivers an output signal which corrects for any difference. This dif amp is similar to the op amps of chapter 16 in its internal circuitry. It is powered by the unregulated input as $+V_{CC}$ and ground as $-V_{CC}$, but the balanced nature of the dif-amp circuits allows it to reject the severe variations in this supply voltage.

The output voltage is selected by the ratio of the external voltage divider, $R_A R_B$. In Fig. 20-2, one fifth of V_O is fed back and compared to a 2.0-V reference, so V_O must equal $5V_{ref}$ or 10 V.

The Series Pass Transistor is integrated, but total package dissipation is limited to 0.5 to 0.8 W for most general-purpose regulators. The unregulated V_{raw} is required by the dif amp and pass transistor drive circuitry to be a *minimum* of 3.0 V above V_O. Ripple and line-voltage changes will add to this difference, so that V_{raw} is likely

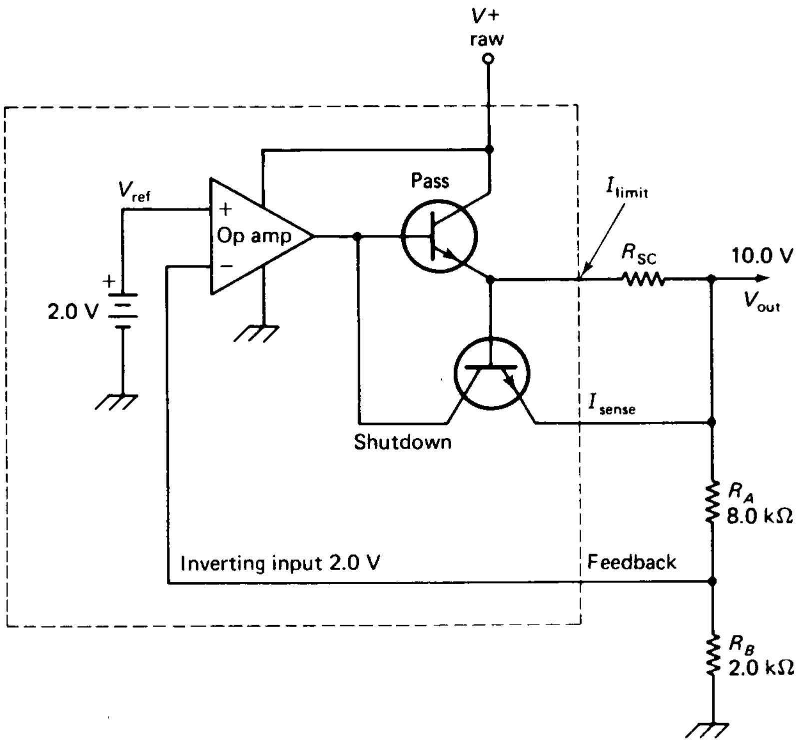


***FIGURE 20-2** Representative IC voltage regulator showing internal elements and basic external circuit.*

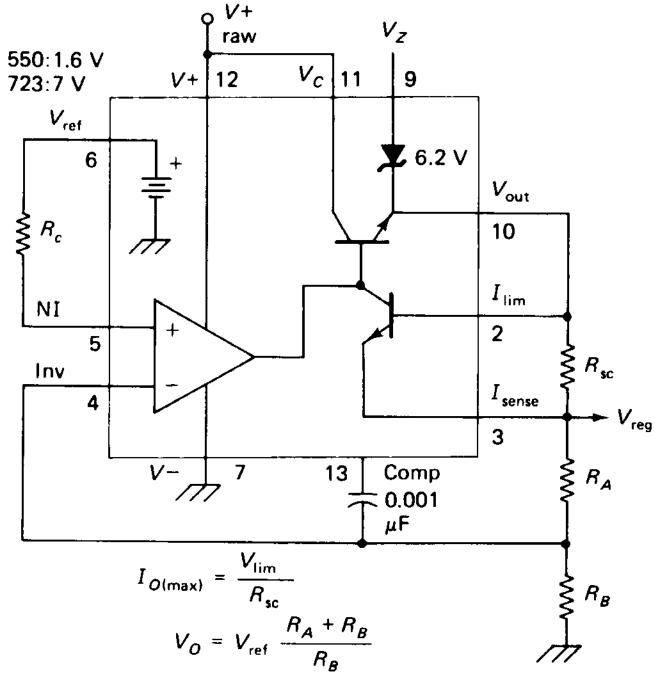
to be 10 V average for $V_O = 5$ V. Maximum current is then

$$I_{\max} = \frac{P}{V_{\mathrm{dif}}} = \frac{0.5}{5} = 100 \text{ mA}$$

Higher currents must be handled by addition of an external series pass transistor with heat sink.

The Current-Limit Transistor is integrated and is connected so that it can turn off the pass transistor by shorting its base to emitter. The current-limit transistor is activated when an external resistor R_{sc} (short circuit) in the output current path drops enough voltage to turn on its base–emitter junction.

Type *723* and *550* Regulators are pin-compatible, and are shown with their basic circuit in Fig. 20-3(a). The *550* outputs 1.6 V at pin 6, and the output of voltage divider $R_A R_B$ is compared to this reference. R_c is a compensating resistor to

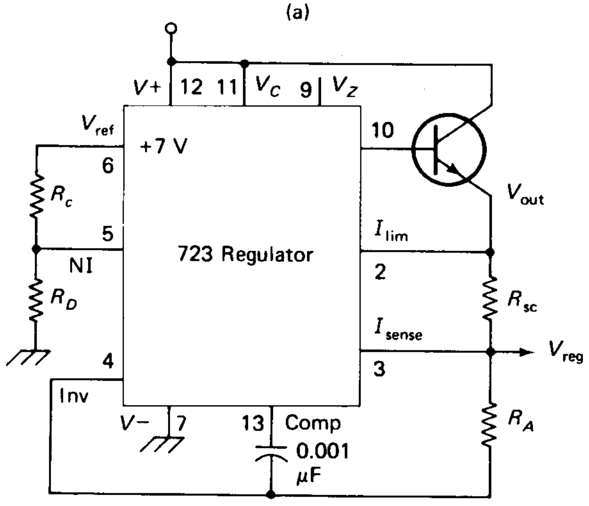


$$I_{O(max)} = \frac{V_{lim}}{R_{sc}}$$

$$V_O = V_{ref}\,\frac{R_A + R_B}{R_B}$$

$$R_c = R_A \| R_B$$

(a)



$$V_O = V_{ref}\,\frac{R_D}{R_C + R_D}$$

$$R_A = R_C \| R_D$$

(b)

FIGURE 20-3 *(a) Internal elements, pin identification, and basic external circuit for the 550 and 723 regulators in 14-pin DIP packages. (b) Placement of* R_C *and* R_D *for output voltages below 7 V with the 723 only, and placement of external pass transistor for greater output current with 723 or 550.*

equalize the voltage drops caused by the bias currents of the two op-amp inputs. R_B should be chosen somewhere near 5 kΩ and R_A and R_c calculated from the formulas given.

The *723* outputs a 7-V reference, making $V_{O(\text{min})} = 7$ V if $R_A = R_c = 0$ and $R_B \to \infty$ in the circuit of Fig. 20-3(a). Lower voltages are obtained by voltage dividing the reference down to the desired voltage and comparing this directly with V_{reg}, as shown in Fig. 20-3(b). Variable output can be obtained by using a pot for the upper arm of the voltage divider in either circuit. Enough fixed resistance should be used in series to avoid voltage demands and power dissipations beyond the rated limits of the IC.

Current Limiting is provided by selecting R_{sc} to drop one V_{BE} junction's voltage at maximum current. This voltage is about 0.6 V at 25°C, dropping to about 0.5 V at 100°C. R_{sc} can be replaced with a short circuit if minimum component count or minimum input/output voltage difference is desired and current limiting is not required.

High-Current Regulators can be built with the simple addition of a series pass transistor connected in a Darlington pair with the internal pass transistor, as shown in Fig. 20-3(b). Assuming a β_{min} of 20 for the power transistor and a drive current of 100 mA from the IC, the load current would be 2 A maximum.

The V_Z pin and the internal 6.2-V zener are used in negative-voltage regulator circuits.

Type *300*, *305A* and *376* regulators are pin-compatible and are shown with their basic circuit in Fig. 20-4(a). All these types use a 1.7-V reference which is internally connected to the noninverting op-amp input, eliminating two package pins. The voltage-dividing resistors R_A and R_B should have a parallel resistance of about 2.2 kΩ to match the internal resistance driving the +input of the op amp.

The reference-defeat input, pin 5, actually accesses the +input of the op amp which is tied to the 1.7-V reference. This pin may be grounded to shut down the supply, or an external voltage may be forced upon it to vary the regulated output voltage.

The current-limit transistor is prebiased in these ICs, reducing the R_{sc} shutdown voltage to about 0.35 V at 25°C, and about 0.30 V at 100°C.

Higher Output Currents can be obtained using the external Darlington-connected transistor, as shown in Fig. 20-3(b), but this increases the $V_{\text{IN}} - V_{\text{OUT}}$ difference requirement by one V_{BE} drop. This family of regulators has a *boost* output which will drive an external *PNP* pass transistor directly, as shown in Fig. 20-4(b). The medium-power *PNP* can in turn drive a high-power *NPN*, as shown in Fig. 20-4(c).

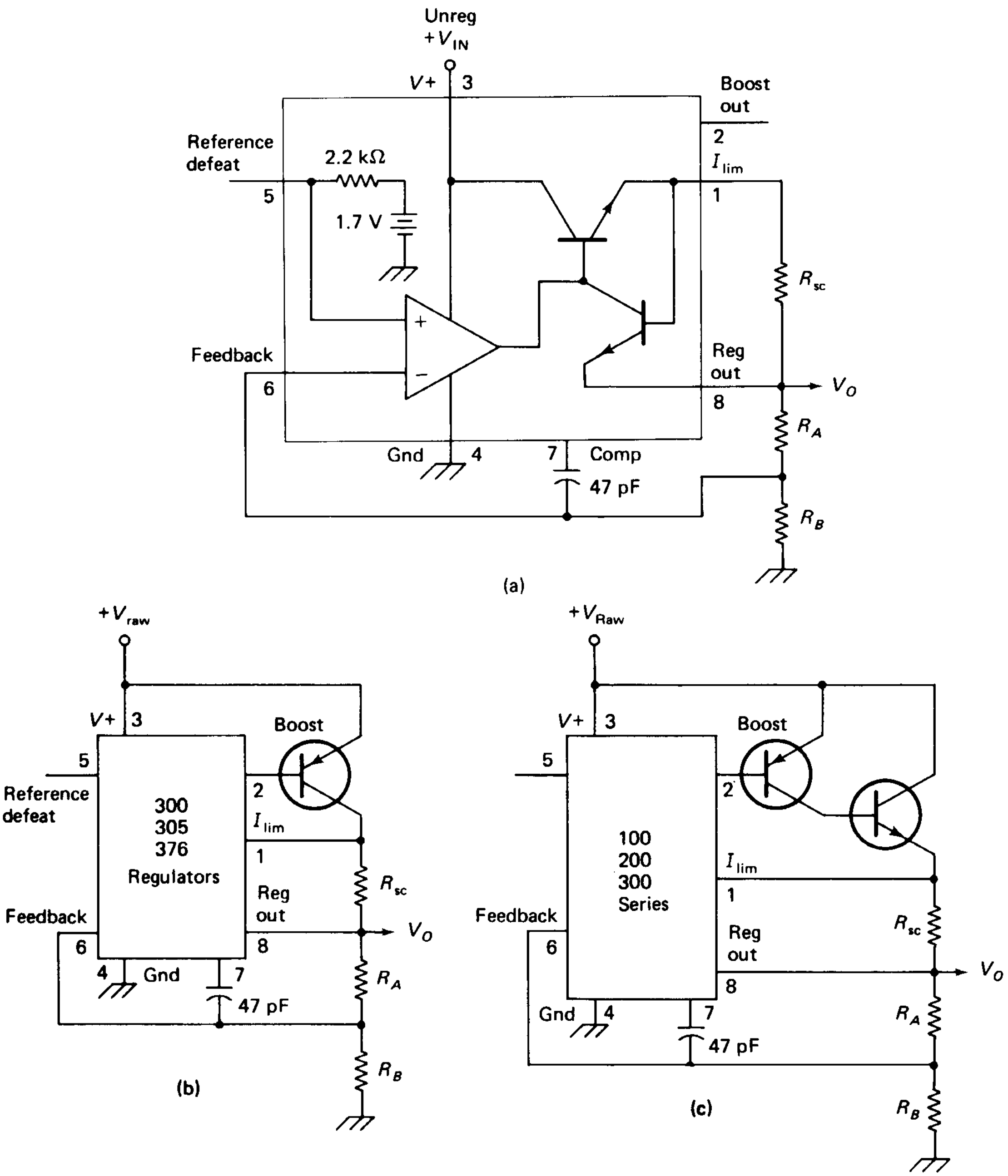


FIGURE 20-4 *(a) Internal elements, pin identification, and basic external circuit for the* 300 *and* 305A *regulators (0 to 70°C operating range) in 8-pin round metal packages. The* 200 *and* 205 *(−25°C to 85°C) and the* 100 *and* 105 *(−55°C to 125°C) are similar. The* 376 *has the same pinout but comes in the low-cost plastic 8-pin DIP package. (b) Adding an external* PNP *pass transistor. (c) External* NPN *pass transistor.*

20.3 FOLDBACK CURRENT LIMITING

In a regulated supply, even with current limiting, the power dissipation in the pass transistor can increase to many times the design limit if the load is short-circuited. This is because the entire input voltage (not just the $V_{IN} - V_{OUT}$ difference) now appears across the pass transistor. To prevent destruction of the regulator in case of a load short, *foldback* current limiting has been developed. Foldback limiting action decreases the current limit as the load resistance decreases, as shown in Fig. 20-5(a).

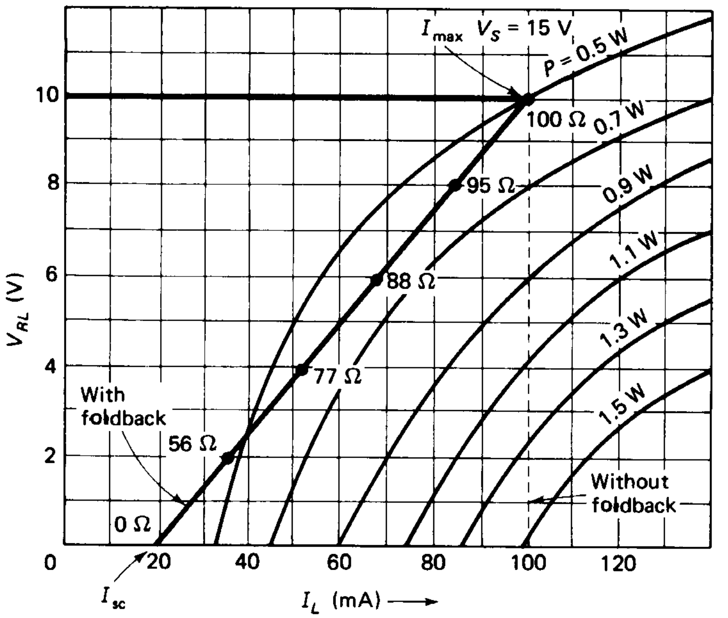
The Foldback Regulation Circuit is basically a Wheatstone bridge with the load resistance as one of its arms, as shown in Fig. 20-5(b). R_{sc} is quite a bit higher in value and drops more voltage than it does in the simple current-limit circuit, but with normal operating R_L values this is compensated for by the drop across R_E in the new voltage divider $R_E R_F$, and the current-limit transistor is not turned on. However, if R_L drops to a short circuit, all of V_O appears across R_{sc}, and most of this appears across the limit transistor. V_O is thus reduced to slightly more than V_{sense}, and the short-circuit current is much less than I_{max} because of the higher value of R_{sc}.

There is a trade-off here in that the input/output voltage difference requirement of the regulator circuit increases as the ratio of $I_{max}/I_{short\ circuit}$ increases. Roughly, folding I_{sc} back to $\frac{1}{5} I_{max}$ requires five V_{sense} drops across R_{sc} instead of one.

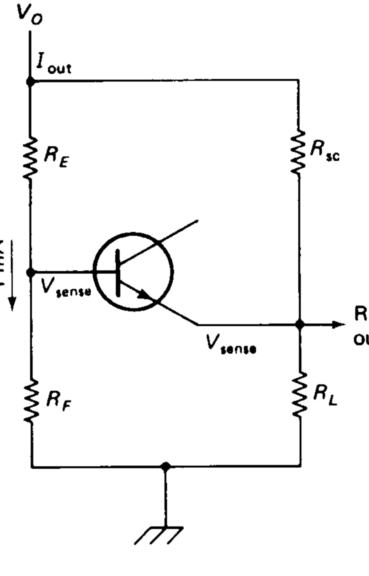
Foldback Circuits for Popular Regulator ICs are given in Fig. 20-6. The circuit for the type *550* deserves special comment because it eliminates the increased R_{sc} drop required with the other types and simplifies the foldback circuitry. This is possible because the *550* uses an integrated SCR instead of a transistor as the current-limiting element. Load currents in excess of the maximum will develop turn-on voltage of 0.6 V across R_{sc} which is fed to the SCR gate (current limit) and cathode (current sense) through R_G. R_G is necessary to prevent the SCR from being shorted by the low-resistance R_{sc}. A value of about 4 kΩ gives an I_{sc} about 15% of I_{max}. R_G can be dropped to 2 kΩ to give I_{sc} about 50% of I_{max}. This may be desirable, because once the SCR switches on, the load current must drop back below I_{sc} to turn it off again. If turn-on surges within the load circuitry exceed I_{max}, and steady-state load current exceeds I_{sc}, the regulator will switch off and not come back on by itself.

20.4 THREE-TERMINAL REGULATORS

Three-terminal regulators have an unregulated input pin, a regulated output pin, and a ground, which is usually the metal case for heat sinking. They are available in a variety of fixed voltages from ±5 to ±24 V, in the plastic TO-92, the power tab TO-220, and the power TO-3 packages with output currents up to 3 A. Their



(a)



$$I_{sc} = \frac{V_{sense}(R_E + R_F)}{R_{sc}\, R_F}$$

$$I_{max} = \frac{V_{sense} + V_O \dfrac{R_E}{R_E + R_F}}{R_{sc}}$$

$$R_{sc} = \frac{V_{RE} + V_{sense}}{I_{max}}$$

$$R_E = \frac{V_{RE}}{1\ \text{mA}} \qquad R_F = \frac{V_{reg} + V_{sense}}{1\ \text{mA}}$$

For $V_{reg} = 10V_{sense}$	
V_{RE}	I_{sc}
V_{sense}	$0.55\, I_{max}$
$2V_{sense}$	$0.39\, I_{max}$
$5V_{sense}$	$0.24\, I_{max}$
$10V_{sense}$	$0.17\, I_{max}$

(b)

FIGURE 20-5 *(a) Typical foldback current limiting: as R_L drops below 100 Ω, load current decreases (diagonal line) instead of holding at I_{max} (dashed line) as it does with conventional short-circuit limiting. Curved lines show pass-transistor power dissipation for $V_{IN} = 15$ V. With foldback $P_{max} = 0.6$ W at $R_L = 88$ Ω. Without foldback $P_{max} = 1.5$ W at $R_L = 0$ Ω. (b) Basic foldback circuit with analysis and design equations, and a chart for estimating I_{sc}.*

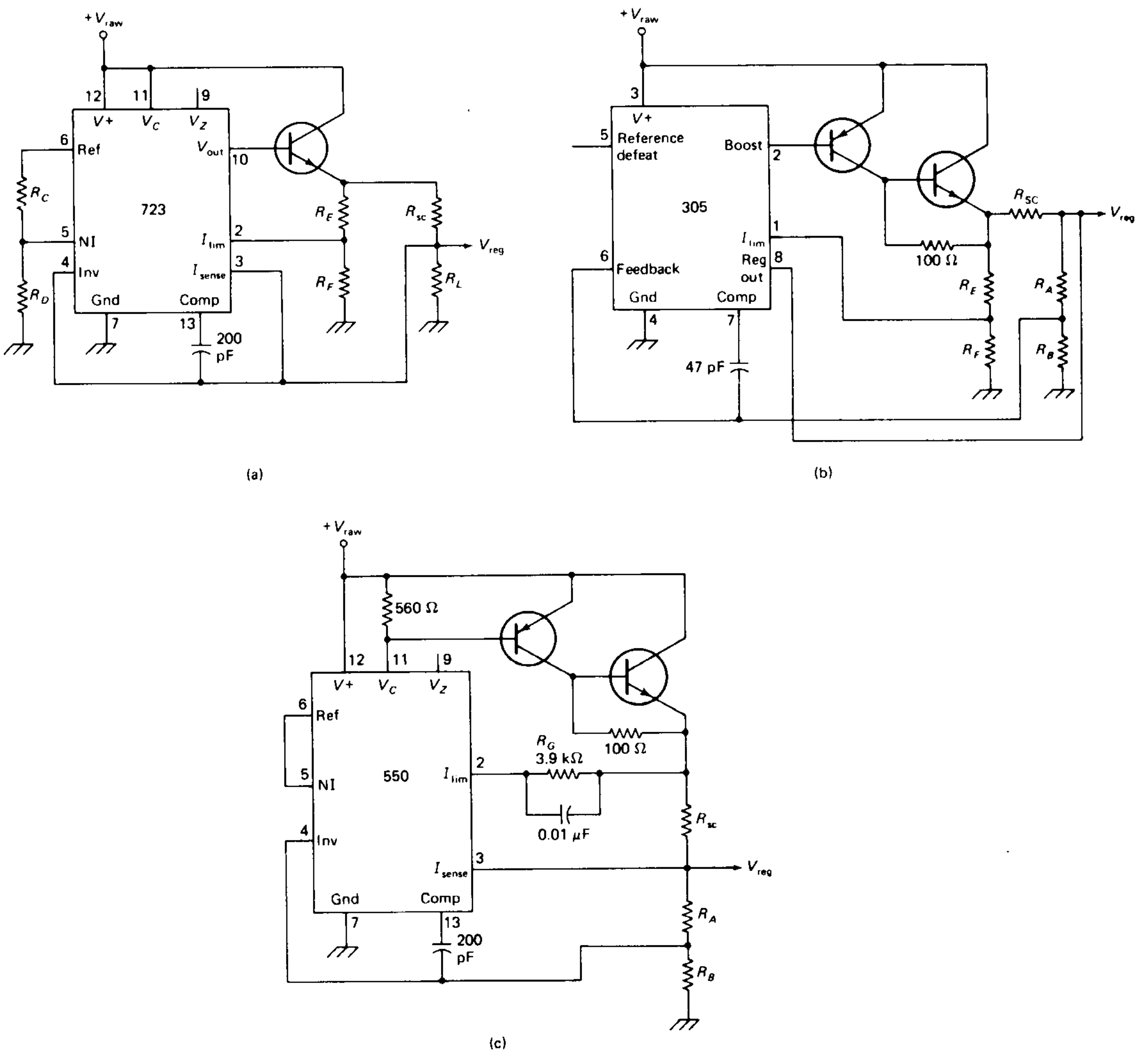


FIGURE 20-6 *Foldback current limiting circuits for the 723 regulator (a) and the 300/305 regulators (b). Relevant equations are given in Fig. 20-5. The 550 regulator achieves foldback with an internal SCR, hence the somewhat simpler circuit (c). External pass transistors are shown since foldback is most commonly used with higher-power regulators.*

low cost and extreme ease of application has brought them immense popularity in a very short time.

On-card regulation has become common since the advent of the three-terminal regulator. A heavy-current bus carrying unregulated voltage from a chassis-mounted supply is run past all the PC cards in an instrument with little regard for noise pickup or line drop. Each card then contains a small three-terminal regulator to supply its own circuitry, practically eliminating noise coupling via the supply lines.

Temperature Stability of the three-terminal regulators is markedly inferior to that of general-purpose regulators—output voltage changes of 1 to 2% over a 100°C temperature range are typical for three-terminal regulators, whereas general-purpose regulators have typical V_O changes of 0.1 or 0.15% over the same range. Guaranteed worst-case temperature stability is three to six times worse than these typical performance listings for general-purpose regulators, but few manufacturers will even put a guaranteed maximum temperature-stability specification on their three-terminal regulators.

Furthermore, a three-terminal regulator is more likely to experience wide temperature swings because it usually contains the series pass transistor. General-purpose regulators can be kept at a relatively cool and even temperature by using an external pass transistor. In spite of all this, three-terminal regulators provide adequate stability for most logic and many op-amp power-supply requirements.

Most three-terminal regulators contain internal short-circuit current-limiting circuitry and will shut down automatically if the junction temperature rises above a safe value for any reason. Figure 20-7 shows some basic three-terminal regulator applications. The input capacitor is required for stability with some regulator types, and is especially recommended if the regulator is located more than a few inches from the power-supply filter capacitor. The output capacitor is seldom essential, but it does keep the output impedance low at high frequencies and improve the output purity when fast pulses of current (such as those produced by TTL) are drawn by the load.

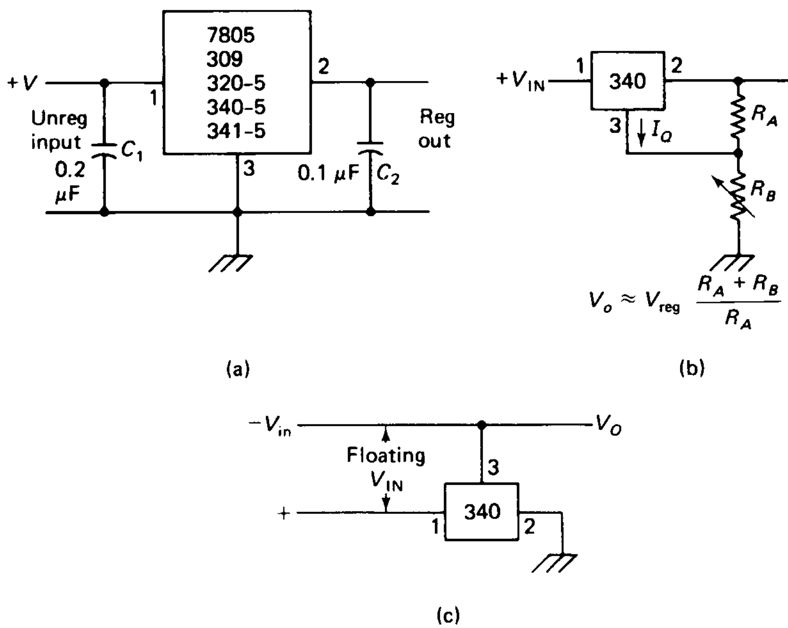


FIGURE 20-7 Three-terminal regulator circuits: (a) basic circuit: C_1 and C_2 optional; (b) variable output voltage: keep $I_{RA} \gg I_Q$; (c) negative-supply regulation with a positive regulator.

The Variable Regulator of Fig. 20-7(b) isolates ground pin 3 from the circuit ground, making heat sinking to the chassis a problem. Load regulation also suffers, because pin 3 delivers a *quiescent current* I_Q, which is typically about 5 mA but which varies with temperature and input voltage. R_A and R_B should therefore be chosen to carry several times this current.

R_B can be replaced by a zener diode to produce a simple regulated supply with V_O greater than the 24 V available in three-terminal regulators.

The negative regulator circuit of Fig. 20-7(c) requires that the input voltage not be referenced to ground (often an unreasonable demand in a large system), and it has its case terminal off circuit ground, presenting a heat-sinking problem. Negative regulators are available, but where only one low-current negative supply is required, Fig. 20-7(c) may make it unnecessary to inventory yet another IC type.

20.5 POWER INVERTERS AND CONVERTERS

A power supply operating at 60 Hz and requiring a 1-kg transformer and 1000 μF of filtering could, ideally at least, be built at 60 kHz with a 32-g transformer (1000 g/$\sqrt{1000}$) and 1 μF of capacitance. High-frequency power supplies are understandably popular where space and weight must be conserved.

An Inverter is an oscillator driven from a dc source (typically a 12-V battery) feeding a transformer which outputs ac of the desired voltage. Figure 20-8(a) shows a simple 6-W inverter which operates at 60 kHz and was designed for a portable fluorescent lantern (fluorescent lamps boast light efficiencies of 12% compared to an incandescent's 3%). The oscillator is an Armstrong type, and the sense of the 4-turn winding must be correct to give positive feedback. The coils are wound on a 1-in. pot core.

Low-frequency inverters are used to produce 60 Hz at 120 V from 12 V dc. Core saturation is more of a problem at low frequencies, so a push-pull oscillator is used to eliminate the dc current in the transformer primary. Frequency stability can be improved by driving the actual switching transistors from a separate low-power oscillator powered from the dc source.

A DC-to-DC Converter is simply an inverter with the ac output feeding a rectifier and filter. As noted above, the filter elements can be quite small if the oscillator frequency is high. Figure 20-8(b) shows a simple dc-to-dc converter using a push-pull oscillator.

Dc-to-dc converters can be regulated by feeding back a portion of the output to alter the bias point of the oscillator. Figure 20-8(c) shows a regulated dc-to-dc converter.

Flyback Power Supplies are step-down converters which operate directly from the rectified ac line voltage and provide line-to-load isolation by means of the high-frequency pulse transformer. They are replacing conventional power supplies in many applications because of the size, weight, and cost advantages of eliminating the 60-Hz transformer. Figure 20-8(d) shows a flyback power supply designed to illustrate the principle as simply as possible. D_1 and C_1 produce 165 V dc which is chopped at 30 kHz by Q_3 and coupled via T_1 to D_4 which rectifies the kickback pulses produced when Q_3 turns off. The LED and photocell provide feedback which shortens the Q_3 on-time if V_O rises. Section 5.6 explains the pulse transformer function in detail. R_6 and D_2 free R_7 from charging C_4, allowing sharp square waves at the collector of Q_2. C_5 prevents the disaster which would otherwise occur

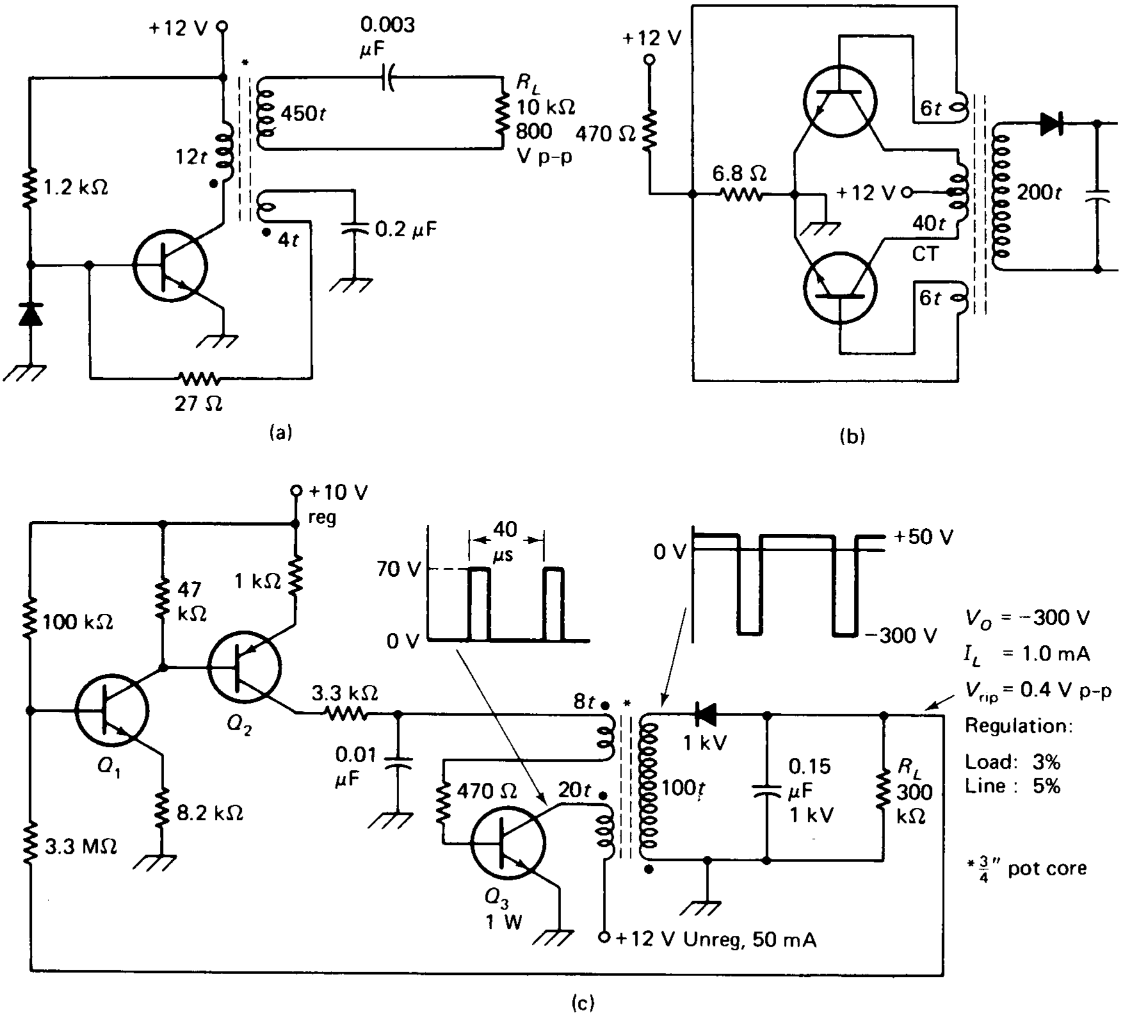
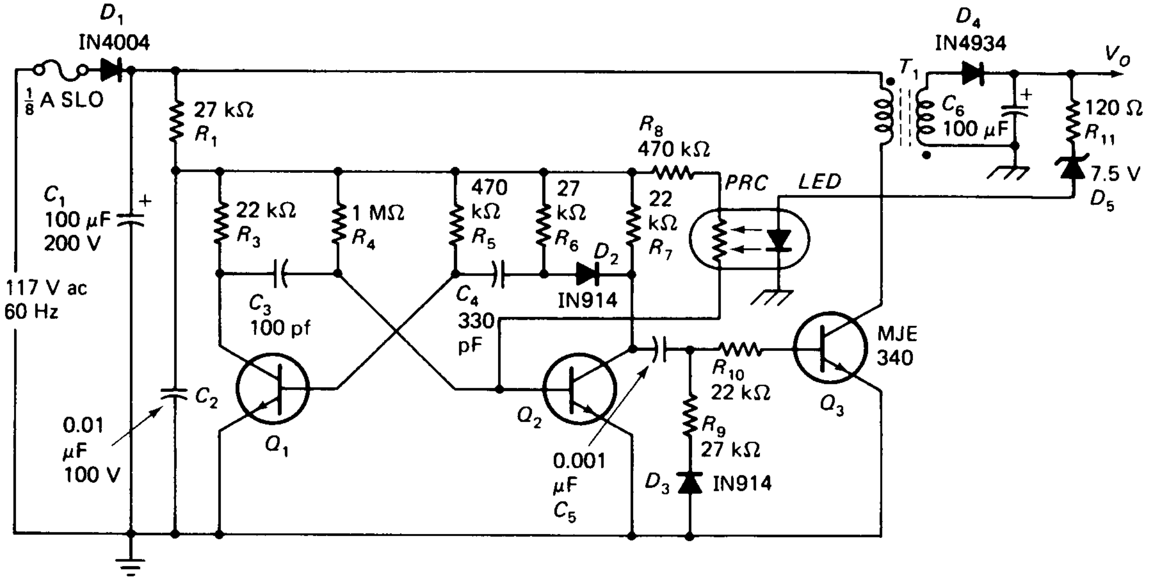


***FIGURE 20-8** Inverters: (a) low-power inverter; (b) high-power push-pull inverter; (c) feedback-regulated dc-to-dc converter with half-wave rectifier and simple capacitor filter.*



V_O = 8.5 V at 220 mA load

Ripple: 60 mV rms, 30 kHz, at 220 mA

Regulation: 0.4 V rise from 220 mA to 40 mA load

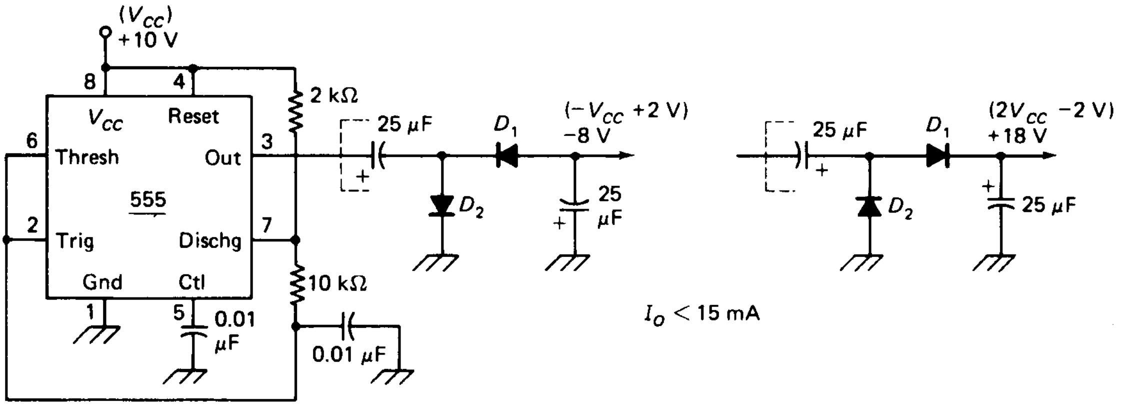
Efficiency: 47% at 220 mA load

Short-circuit toleration: 5 seconds

Open-circuit toleration: Indefinite

Transformer: Amidon PC3019-77 pot core;
70-turn primary, 7-turn secondary

(d)



(e)

***FIGURE 20-8** (d) Flyback supply eliminates 60-Hz transformer; (e) transformerless dc-to-dc converter for supplying negative or boosted positive voltage at low current.*

if multivibrator $Q_1 - Q_2$ should "hang up" with Q_3 turned on. More complex flyback supplies are capable of efficiencies above 80%.

A simple converter which can supply a negative or a boosted positive voltage at about 12 mA is shown in Fig. 20-8(e). The circuit uses no inductors, takes less than 10 cm^2 of circuit board "real estate" and adds a parts cost of about $1. The output current is enough to power five *741* op amps or a dozen *308*s. Used on a +5-V TTL line, it can produce enough bias for FET amplifiers or a CMOS *4016* analog switch. Note that D_1 and D_2 should be moderately fast signal diodes, not conventional power-supply diodes.

20.6 SWITCHING REGULATORS

A switching regulator uses feedback to control the relative on and off times of a transistor, and thereby control the average current delivered to an inductor. This controls the average output voltage, which is filtered to pure dc by the inductor and an output capacitor. Figure 20-9(a) illustrates the basic principle. The diode is necessary to provide a discharge path for the inductor when the transistor turns off.

The inductor and capacitor need not be very large because the switching frequency is high. Typical values for a 1-A, 5-V supply are 0.5 mH, 200 μF, and 10 kHz. The diode, however, must be capable of turning off in less than a microsecond, so standard power rectifiers cannot be used. The filter capacitor must present a very low reactance and low equivalent series resistance at the switching frequency and its low-order harmonics. Low-quality electrolytics will degrade efficiency, reliability, and noise rejection.

Notice that there are no resistances in the circuit—the transistor and diode are essentially on–off switches, and L and C burn no power—so ideally the regulator is 100% efficient. Actual efficiencies up to 90% are practical. Linear regulators, operating from line voltages that vary $\pm 10\%$, or in extreme cases $\pm 15\%$ using 5%-tolerance components, are hard-pressed to achieve a worst-case efficiency of 70%. Low-voltage supplies, such as the common +5-V TTL supply, are lucky to exceed 50% efficiency considering the extra drops for external pass transistors and foldback circuitry. Remember that for 100 W of output, 90% efficiency produces 11 W of dissipation in the regulator, while 70% produces 43 W and 50% produces 100 W of wasted power.

Switching-Regulator Drawbacks: The fast switching of high currents generates noise spikes which can interfere with other parts of the system via stray capacitive or inductive coupling or the input or output supply lines. Shielding and filtering are often required. The consequences of a failure, particularly a shorted switching transistor, can be rather serious, so protective circuitry at the output and input is a good investment. Switching regulators are also slow to adjust to rapid changes in output current, making them less desirable than linear types where the load current contains fast pulses.

Step-Down Regulator Action is best understood by looking closely at the representative circuit of Fig. 20-9(a). It is easy to see that a 100% duty cycle (Q_1 always on) would produce $V_O = V_{IN}$, neglecting losses. Lower duty cycles will produce a rectangular waveform, as shown at the base of Q_1, because D_1 switches on as soon as Q_1 switches off. The average value of this waveform is

$$V_O = V_{IN}\frac{t_{on}}{t_{on} + t_{off}} \tag{20-4}$$

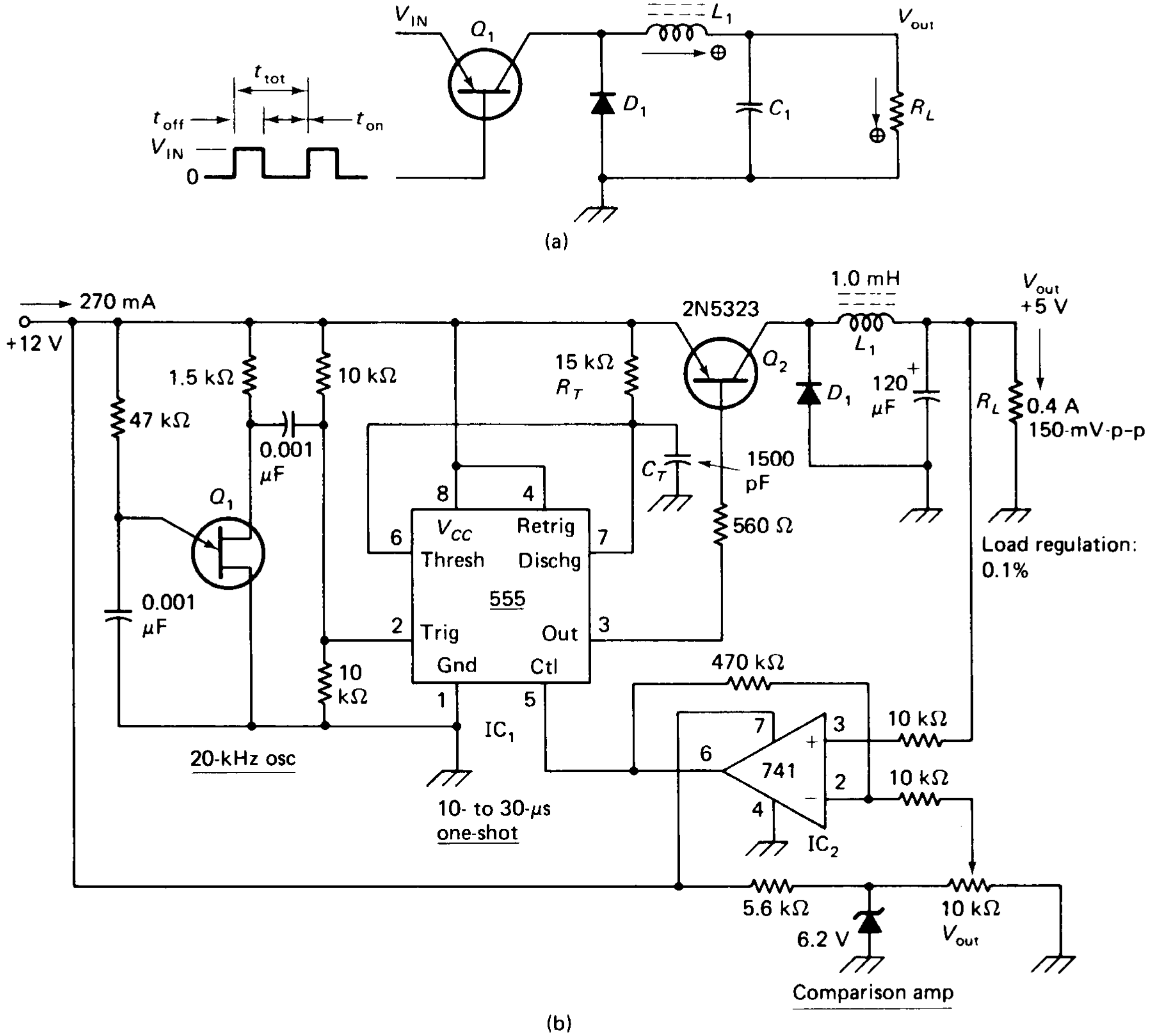
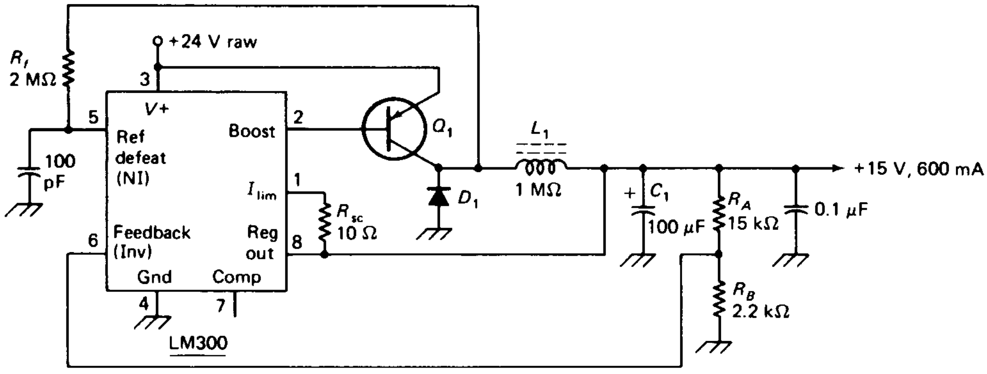
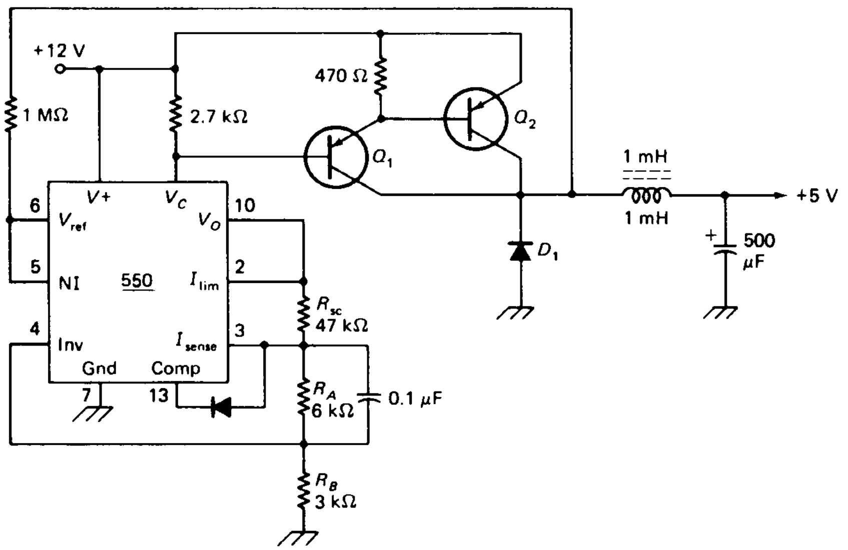


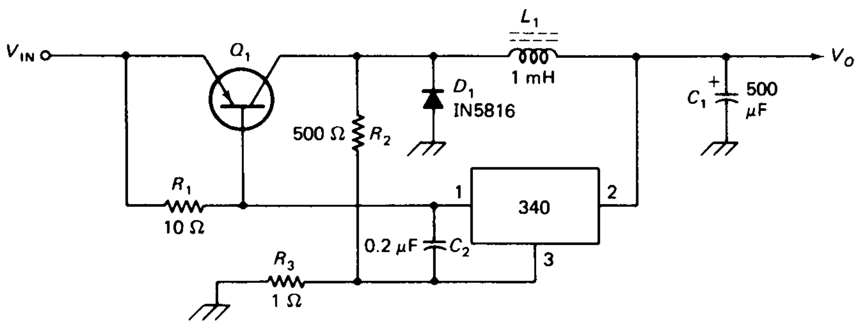
***FIGURE 20-9** Step-down switching regulator: (a) representative circuit; (b) practical circuit with Q_1 oscillator, IC_1 pulse-width-modulated one-shot, driven by IC_2, which compares output with zener reference.*



(c)



(d)



(e)

FIGURE 20-9 *Step-down regulators may use the 300/305/376 IC (c), the 550/723 IC (d), or a three-terminal regulator (e) in a self-oscillating mode.*

The LC filter transforms the pulses into dc of their average value. Load current has no primary effect on the duty cycle required. It is not necessary or even desirable that V_O be as near as possible to V_{IN}, as it is with linear regulators. A V_{IN} of 50 V with V_O of 5 V can be achieved with no loss of efficiency.

The time constant of the inductor and load resistance should be at least twice the off time of the switching transistor:

$$L \approx 2t_{\text{off}} R_L = \frac{2t_{\text{off}} V_O}{I_L} \tag{20-5}$$

Notice that large load currents require *smaller* inductors. Peak inductor current may be 1.5 times the load current, and core saturation must be avoided, however.

The filter capacitor can be determined by considering L_1 and C_1 as a voltage divider with V_{IN} (p-p square wave) across X_L and V_o (p-p ripple) across X_C. There is considerable error in treating the input rectangular waveform as though it were a sine wave, but the approximation is usually adequate, and reduces to

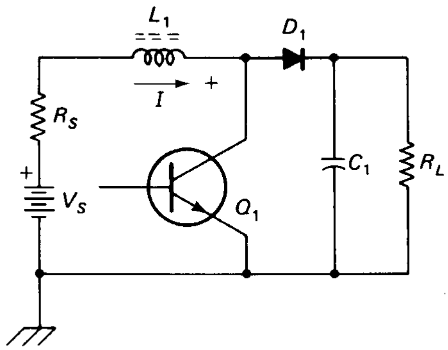
$$C_1 = \frac{V_{\text{in}}}{V_{o(\text{p-p})} 40 L_1 f^2} \tag{20-6}$$

where f is the switching frequency.

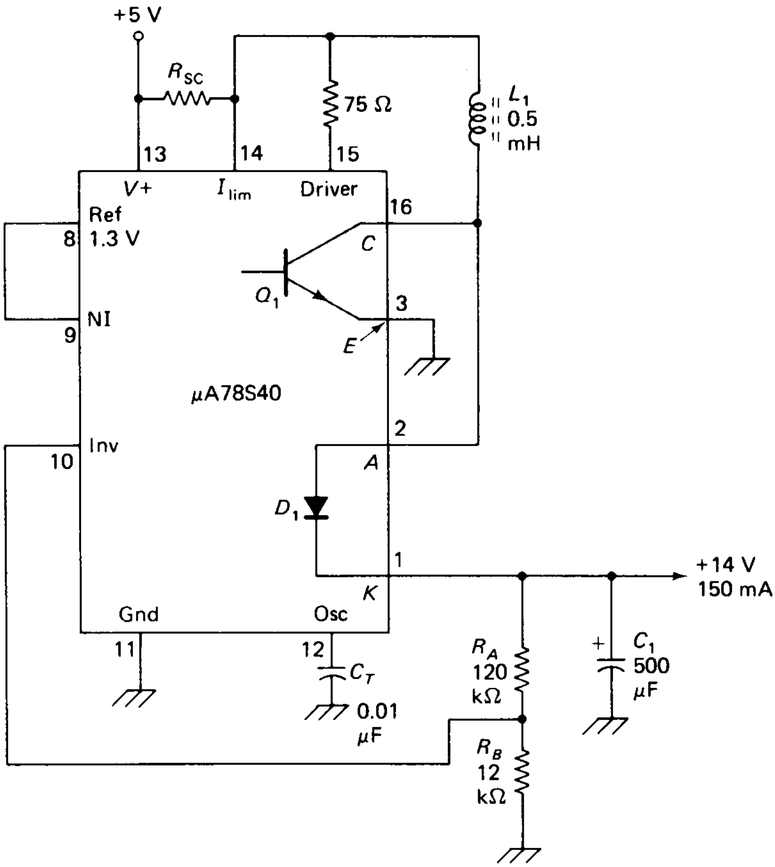
Figure 20-9(b) shows a complete switching regulator using a UJT oscillator to drive a *555* IC pulse generator. The width of the pulses is varied by the output of a *741* op amp whose input compares V_O with a fraction of the voltage obtained from a reference diode.

Figure 20-9(c) shows a simpler regulator using a type-300 regulator IC in a self-oscillating mode. The 2-MΩ resistor provides positive feedback to the noninverting op-amp input, switching Q_1 on with a snap action. This part of the circuit is actually an op-amp Schmitt trigger, with R_f controlling the dead zone. When Q_1 turns on, V_O rises, but relatively slowly due to L_1 and C_1. This rise is fed back to the inverting op-amp input via R_A and R_B, and Q_1 snaps off after the hysteresis is overcome. Some ripple at V_O is necessary to operate the Schmitt trigger. Lowering R_f will increase the ripple (hysteresis) and decrease the switching frequency. A second LC section can be added to remove this ripple. Figure 20-9(d) shows a similar self-switching regulator using a type-*550* IC.

A simple switching regulator using a three-terminal regulator IC is shown in Fig. 20-9(e). Q_1 is turned on by the inrush of current from V_{IN} to pin 1 of the IC. A small part of V_{IN} is applied to reference pin 3 of the IC via voltage divider R_2R_3 when the collector of Q_1 turns on. This decreases the voltage between reference pin 1 and output pin 2, ensuring that the IC will continue to conduct heavily until V_O rises by $1/500$ of V_{IN}. Once V_O does rise by this amount, the IC shuts off, turning Q_1 off. C_1 then discharges through R_L until V_O is again low enough to cause the IC to conduct. Output ripple is in excess of $1/500$ of V_{IN}, and a second LC section may be needed to bring it to an acceptable level. Lowering R_3 reduces ripple but also increases switching frequency while lowering switching speed, resulting in reduced power efficiency.



(a)



(b)

FIGURE 20-10 *Step-up switching regulator: (a) representative circuit; (b) practical circuit using 78S40IC with integrated power transistor, fast power diode, and oscillator.*

A Step-Up Regulator is shown in principle in Fig. 20-10(a), with a practical version given in Fig. 20-10(b). Q_1 is turned on for a small fraction of a time constant L_1/R_S. When Q_1 is turned off, L_1 generates a kickback voltage which adds to V_S and is applied through D_1 to C_1 and the load. There is no hard-and-fast limit to the voltage-step-up-ratio—V_O can be several times V_S.

The practical step-up regulator uses an IC specifically designed for switching regulation. It contains an oscillator, a fast diode and switching transistor capable of handling 1.5 A, and a second op amp in addition to a 1.3-V reference and comparison amplifier.

An Inverting Regulator is shown in principle and in practice in Fig. 20-11(a) and (b), respectively. When Q_1 is turned on, conventional current flows down through L_1 as shown. When Q_1 is turned off, this current continues to flow, producing a kickback voltage which charges C_1 through D_1.

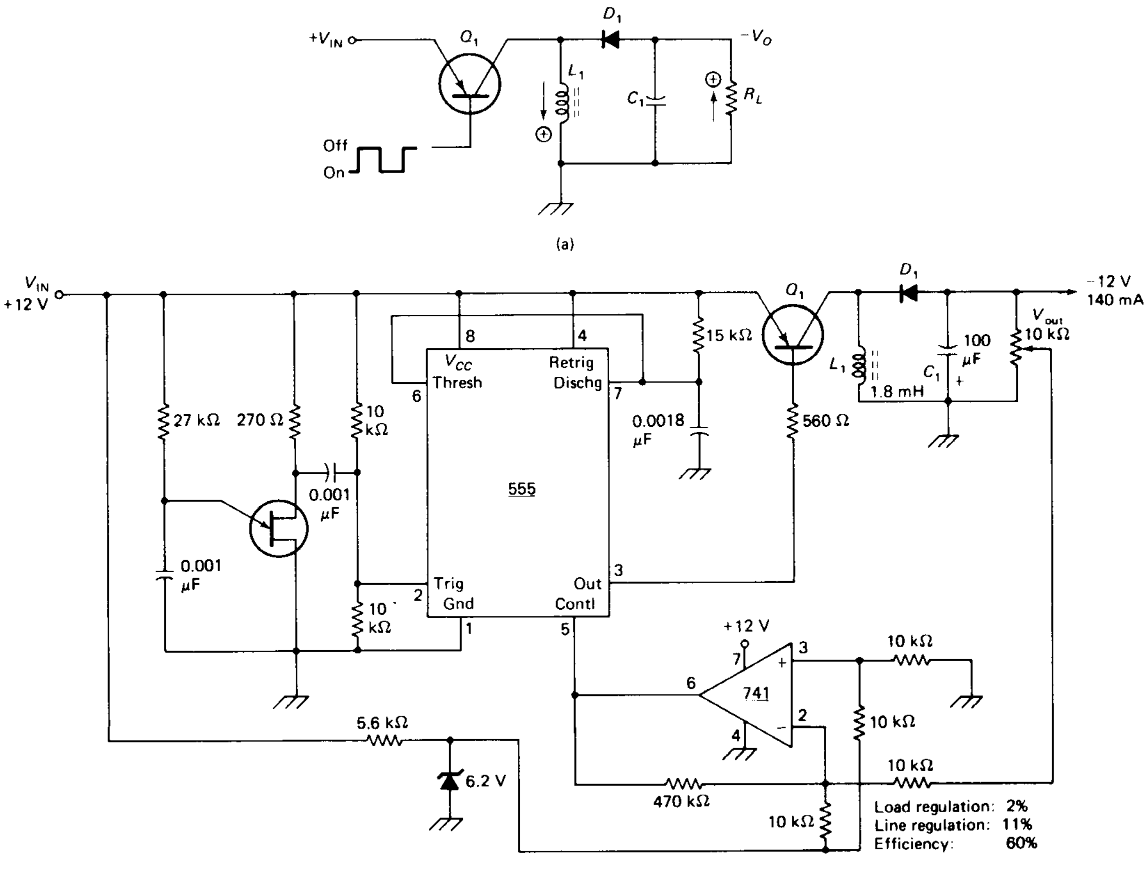


***FIGURE 20-11** Voltage-inverting regulator: (a) representative circuit; (b) practical circuit.*

21

TROUBLE DIAGNOSIS

21.1 AN APPROACH TO TROUBLESHOOTING

We may as well put the main point right up front where it will be noticed: *Effective troubleshooting is a matter of checking out your expectations about a circuit.* The expectation may be general such as: "The pilot light should come on when I pull the ON button," or it may be specific, such as: "If I short pin 2 of the *555* IC to pin 1, pin 3 should go high for about 1 ms." You may not be able to form detailed expectations about every part of the circuit, but neither should you be completely without expectations at this point in your studies. This chapter and the ones following will help you to recall and organize them in a logical way.

You will go on checking the circuit's behavior against your expectations, one by one, until you find one that doesn't match. At that point there are only three possibilities:

1. That part of the circuit is malfunctioning,
2. Your expectation was in error, or
3. Your measurement was in error.

In the first case you will have narrowed the problem down to a small area. In the second case, you will have learned something about circuit behavior which will be of use to you later. In the third case, you will have learned something about measurement technique, or will have discovered a faulty piece of gear. In any case it will be profitable—if you stop long enough to take advantage of the opportunity.

Tremendous mental discipline is often necessary to avoid wasting this opportunity. The temptation is always strong to say, "I don't know if this is the right reading or not," and pass on to an equally fruitless measurement in the next stage. To avoid such aimless troubleshooting it is a good idea to jot down your expectation or speak it aloud *before* you take the measurement. (Contrary to malicious rumors, this is the reason why good technicians are often heard talking to themselves.)

If you don't have a firm expectation of what you *should* see, there is absolutely no point in taking the reading. You won't know what you're seeing anyway, so it's just a waste of time. You would be better advised to cast about for a more clearly held expectation to check out, or when that avenue is exhausted, to look into a book or manual to obtain a better understanding of the circuit in order to form more expectations about its operation.

It bears repeating that once you find a discrepancy between your expectation and the circuit's performance, you must stop at that point in the circuit until you resolve the conflict. If you expect a gain of 20 in the first audio-amplifier stage and you measure a gain of 2, don't check anything in the audio output stage. Don't take any readings in the power supply. Stay with the first audio stage until you find the trouble. You may find an open capacitor or you may find that you forgot about the $\times 10$ probe on the second channel of your scope, but stay with it until your expectation and your measurement match.

Fooling Yourself: There is an insidious tendency in most of us to "read what we need" instead of what our instruments are actually indicating. Our minds are preloaded with an expectation that we want to see confirmed, and as soon as the "right" number pops up on the scope or meter, we jump for it without taking the time to see that the instrument is really set up properly. Here are some common measurement errors:

- Forgetting $\times 10$ probes
- Failing to check *zero* and *calibrate* controls
- Substituting an rms measure for a peak-to-peak expectation
- Measuring line noise picked up on an open probe and calling it signal
- Failing to notice that the meter resistance or capacitance is loading the circuit
- Attempting to measure ac voltages above the maximum rated frequency of the instrument

Clearing the mind of expectations is not the answer to this problem, because expectations are basic to troubleshooting. Careful and methodical technique, experience with the instruments, and a constant skepticism about your readings are the best defense.

21.2 SOURCES OF EXPECTATIONS

Experience: Rapid troubleshooting is almost always the result of experience, not technical competency. Many of us have seen the TV man march into the room, sniff the air, pop the back off the set, whip out his soldering gun, change *C 437*, and put the set back on the air in five minutes. It's easy to be impressed by this sort of thing and to want to emulate it. What the TV man doesn't tell you is that he's changed *C 437* nineteen times in the last week and keeps a handful in his pocket so he won't have to walk back to the truck. The head bench tech spent three days on it the first time somebody brought a set in with that problem. There's nothing wrong with this kind of troubleshooting—it is very effective with mass-produced items. Industrial troubleshooting is seldom this repetitive, and we must look to other sources than experience for our expectations.

Visual Checks lead directly to the trouble often enough to make this the first avenue of attack, even before the voltmeter is brought out. Fuse elements can be inspected and transformer hum can be telegraphed up a large screwdriver to the ear to see if power is reaching the supply. Resistors may have a black charred ring or may be puffed up and cracked in the center. Switch contacts may be corroded or bent. Bare wires may be touching, or a loose nut, wire end, or piece of solder may have fallen and become wedged between two wires or terminals. Printed-circuit tracks may be broken from overstress or shorted by corrosion from water, battery paste, or other foreign material.

Some of these "problems" may actually be only symptoms of a deeper problem. Fuses, resistors, and transformers, for example, seldom fail unless they are overstressed—by a leaky filter capacitor perhaps. When replacing a 1-A fuse which blew for no apparent reason, it would be a good idea to measure the line current (with the fuse out and an ammeter across the fuse holder terminals) to see if the new fuse is also being stressed near its limit. Similarly, the voltage across a newly replaced resistor should be measured and the power calculated to see if it is necessary to look further for the real source of the trouble.

General Expectations can be checked out without a schematic or detailed circuit tracing.

- Silicon power diodes should drop no more than 1.1 V in the forward direction, giving an indication of no more than 0.5 V on the test fixture of Fig. 21-1(a).
- Large electrolytic capacitors (100 μF or more) should have at least half the rated voltage across them unless the designer gave no thought to economy.

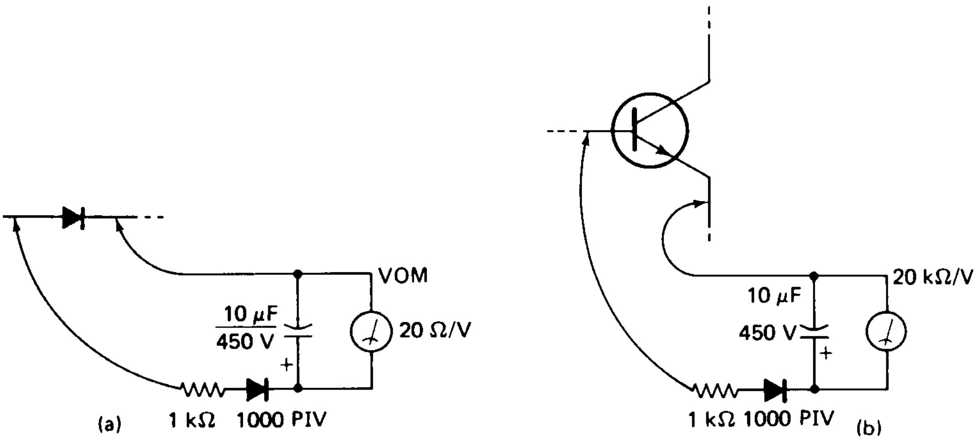


FIGURE 21-1 *(a) A conducting diode should read between 0.1 and 0.5 V on the VOM with this test circuit. Less may mean a shorted diode. More means the diode is open. Start on a high-voltage VOM scale. (b) Transistor forward voltage should produce not more than 0.5 V on the VOM.*

- Transistors should have a forward base–emitter drop of no more than 1.1 V, giving an indication of no more than 0.5 V in the test fixture of Fig. 21-1(b).
- Transistors, even when operated as saturated switches, should show at least a few tens of millivolts from collector to emitter. Absolutely zero volts may mean that the transistor is shorted.
- Power resistors and transistors should get at least moderately warm: if not at idle, then under full signal. Power components would not have been used if cheaper milliwatt components would suffice.
- Generally, any component that boils a drop of water is under strong suspicion.
- Gentle pressure with a plastic probe on components, wire bundles, and circuit boards should not cause abrupt changes in instrument operation.
- Common IC types and families should have dc supply voltage on their V_{CC} pins, as shown in Fig. 21-2.

The Schematic Diagram provides the most valuable and complete source of expectations about the operation of the circuits in an instrument. Often the diagram includes expected scope waveforms and voltmeter readings at specified points in the circuit. These are helpful, but they should not be used blindly. The function of each part of the circuit should be determined first (i.e., preamp, oscillator, Schmitt trigger, etc.) so the waveforms and voltages can be fit into a meaningful pattern.

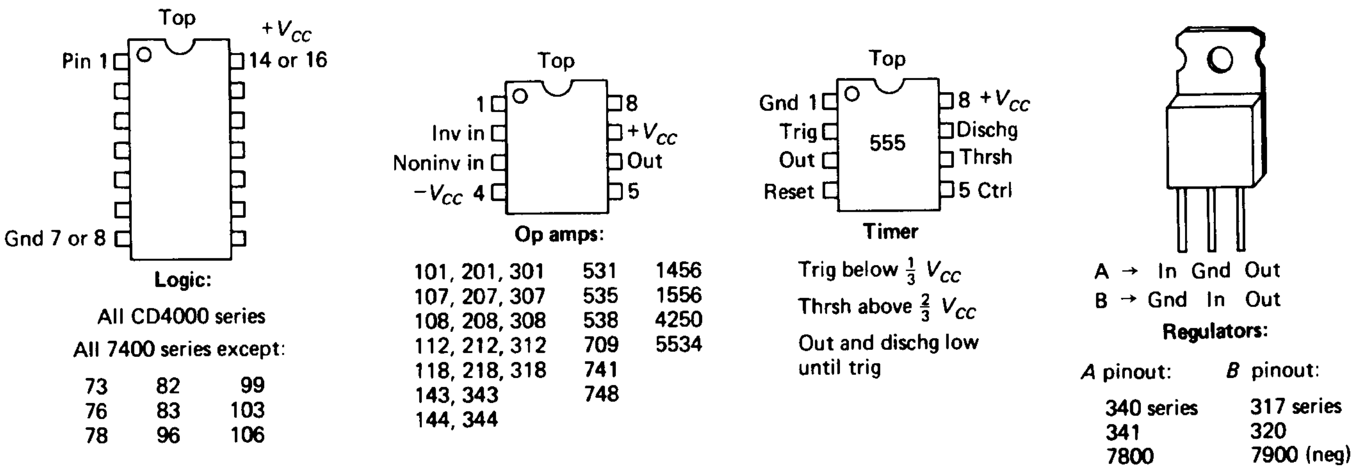


FIGURE 21-2 *Most commonly encountered IC types and voltages to be expected at some of their terminals.*

Service manuals may include block diagrams, circuit descriptions, troubleshooting procedures, and resistance charts. These should be used in conjunction with the schematic diagram, not in place of it.

21.3 FAILURE PROBABILITIES

Certain types of components are observed to cause equipment failure more often than others, and a knowledge of this hierarchy is useful in troubleshooting. The following list, gleaned from experience, proceeds from the most likely to the least likely cause of the problem:

- Operator error
- Cables and connectors external to the instrument
- Switches and relays
- Power semiconductors
- Cables and connectors within the instrument
- Soldering and circuit-board faults
- Small-signal semiconductors
- Electrolytic capacitors
- Power transformers and inductors
- Power resistors
- Variable resistors
- Ceramic capacitors
- Mylar and paper capacitors
- Low-power resistors
- Mica capacitors
- Low-power inductors

A glance at the top of the list will suggest several "quick-fix" approaches to try.

1. More service calls really are due to operators who don't know how to use the instrument than to any of the other causes listed. Make sure that you know how to use it and how to bring it back from a completely misadjusted condition. Find out about all the seldom-used concentric control knobs, back-panel controls, shorting plugs, and so on. Read the operator's

manual—ignominious as that may be. You could find something important hidden among the gobbledegook.

2. External wires and connectors are vulnerable to abuse and are often a source of trouble. Merely examining them or flexing them while the unit is operating will often show up a short or open.
3. Switches and relays often become corroded, pitted or lose contact pressure. Slow operation of the switch handle, or a gentle pressure on the relay armature, may point up an intermittent contact. A shot of contact cleaner may also help.
4. Power transistors and diodes fail frequently enough to make it profitable to disconnect the emitter and base leads (or either end of a diode) so that the junctions can be checked with an ohmmeter.

21.4 HOW TO TROUBLESHOOT

If none of your general expectations about the instrument leads to the defect, you will have to organize your resources for a technical attack on the problem. It will pay to follow some such procedure as the following.

Define the Problem: By this we mean something a little more incisive than the vague statement "It don't work." Of a digital multimeter we might say, "Ohms scales progressively more inaccurate on higher ranges, to -50% error on 10-MΩ range. Dc and ac volts and amps OK." Of a TV receiver we might say, "Picture slants and floats sideways. Horizontal hold sensitive and unable to stabilize. Vertical hold, audio, picture definition OK."

The first benefit of making such a statement is that is concentrates your attention on the defective area. Most electronic instruments are too extensive to be taken in at one bite, so "divide and conquer" must be the rule. The second benefit is that it gives you something intelligent to say when discussing the problem with the factory rep, boss, or senior tech, who might be able to point you toward the solution if you define the problem clearly enough.

Get the Schematic: Troubleshooting without a schematic is like working a crossword puzzle blindfolded. Call the manufacturer, the nearest electronics supplier, the library, or whomever you must, but get a schematic. Without it the situation is all but hopeless and your time is being wasted, so take whatever time is needed to secure it. And when you are through put a copy in a plastic envelope inside the instrument case for the next poor fellow—he may be you.

Clear the Workbench: Signal generator on the left, instrument and schematic in the center, oscilloscope and voltmeter on the right is a good way to organize it. Turn on the soldering pencil and test equipment and let them run so that they will be warmed up when you need them. Have a few parts trays available to keep screws and other small items together as you disassemble the instrument.

Proceed with voltage checks using the oscilloscope in the dc-coupled mode in the following sequence:

- Power supply #1: dc output and ripple; power supply #2: and so on
- Input to stage 1, output; input to stage 2, output; and so on

If you don't know what supply voltage to expect, go back and analyze the circuit to determine what it should be. If you are unable to do this, you're probably in over your head and should go back to a study of power-supply circuits. If you don't understand it, there is little chance that you're going to be able to fix it.

If there is no input to stage 1, use your signal generator to inject one. If you don't know what frequency or amplitude to inject you probably don't understand the circuit well enough to be troubleshooting it, and it's back to the books again. This is not running away from the problem. This is confronting the problem—the real problem—of lack of understanding.

Keep a Record of your expectations and measurements. A looseleaf binder with paper divided vertically down the center is ideal. Date each entry, indicate the instrument by serial number, and when you finally locate the trouble, don't forget in your elation to indicate what it was in the notebook. You or your successor may see that instrument again in two years, and a minute spent now might save an hour then.

And in the name of mercy, if you have to remove a transformer or a connector with two dozen leads on it, keep a very meticulous sketch of how the wires are to be reconnected.

Don't Redesign the Instrument: You may find in the course of your investigations that if *R 542*, which is marked as 820 Ω on the schematic and appears as 820 Ω in the instrument, is changed to 560 Ω, the instrument suddenly begins to work. This is an interesting bit of information, but it does not justify popping in a 560-Ω resistor and closing the cabinet! If it worked once with 820 Ω, it should work now with 820 Ω. Obviously, there is some other problem lurking in there which the change to 560 Ω is able to cover up. Maybe it is a leaky capacitor, or a drying-up electrolytic, or a transistor whose beta is changing, or corrosion on the circuit board, any of which is likely to get worse—so changing *R 542* to 560 Ω would be only a temporary solution anyway.

Watch Out—It May Fry Again: After replacing the guilty component, turn the power on *briefly*, watching for smoke or other signs of malfunction. Next keep it under power for at least five minutes while observing its operation to be sure that heat buildup is not working to cause another failure. Finally, let it cook in an out-of-the way corner for several hours if possible. The failure of one component often overstresses several others, and it is wise to give all latent problems a chance to show themselves before returning the instrument to service.

21.5 SIGNAL TRACING

Divide and conquer is the rule in troubleshooting. Most often we can isolate the fault to one area of the instrument just by the symptoms exhibited at the front panel. Let's take an oscilloscope as an example that every technician will be familiar with.

If there is no light on the screen at all, we would first check the power supplies (there may be four or five). If darkening the room shows a light at the top of the screen which moves with the horizontal-position control but not with the vertical control, we will confine further testing to the vertical amplifier. If we can produce a horizontal line with an external horizontal input but not with the internal sweep, we start troubleshooting in the sweep generator. If we get sweeps but they won't lock to the displayed signal, we start digging into the sweep-trigger circuits.

Always obtain as much information as you can from the front panel before opening up the instrument. If the scope won't lock in on an internal sine wave, what about a square wave? What about external trigger and line trigger, auto trigger and driven mode, different settings of the range and variable controls?

Board Swapping: Many instruments are designed with plug-in circuit boards, each board containing a distinct function. A serviceman armed with a caddy of spare boards and a rudimentary knowledge of the instrument's function can be quite effective in such a situation without expensive test equipment or a detailed knowledge of electronic circuits. Keeping the necessary stock of spare boards is practical only where service work is confined to a few types of instruments, however, and after a few months under these conditions a good serviceman will gain enough experience with the foibles of the equipment to troubleshoot it to the component level anyway.

Instruments often contain two or more identical plug-in boards. If swapping them causes the trouble to move from one area of the instrument to another, one of these boards obviously bears the problem. This technique may be invalid if the boards contain adjustable components. In any case, the boards should be returned to their original positions to preserve the original alignment of the instrument.

ICs may also be swapped in this way to locate problems where an instrument contains several odd types for which no replacement is in stock. Remember, though, that if an IC went bad, the circuit may have overstressed it, so it is best to apply power only briefly when testing the swap. A finger tip on the IC will let you know if it is overheating.

Checking Inputs and Outputs: Let us say that you have a malfunctioning instrument on the bench. Symptoms and front-panel response indicate that the problem lies in an output dc amplifier or preamplifier stage. Your memory produces no recollection of a previous similar problem, visual inspection reveals no obvious damage, and a quick check of the most failure-prone components in the defective area shows them all to be good. You have the schematic and your oscilloscope is warmed up. Now at last you are about to see the fruit of the labor you invested in studying electronic circuits.

Electronic systems are composed of functional blocks: circuits which receive one or several input signals and produce one or several output signals. Trouble-shooting at this stage consists of:

1. Recognizing the functional blocks in the problem area (i.e., generating expectations of what the input and output signals should be)
2. Checking out these expectations with the oscilloscope

Absence of an input signal can be caused either by a defective driving stage or by an input short circuit, as illustrated in Fig. 21-3 (a) and (b), respectively. Absence of an output signal can be caused by a defective stage or by a shorted load. The same two figures serve to illustrate this concept. Determining whether the fault lies in the driving or driven stage may require breaking the path between them. If this causes the driver output to suddenly reappear, the driven stage was apparently shorting the output. To verify this, a signal generator can be used to drive the opened input line. If the signal generator too is shorted, the diagnosis is confirmed.

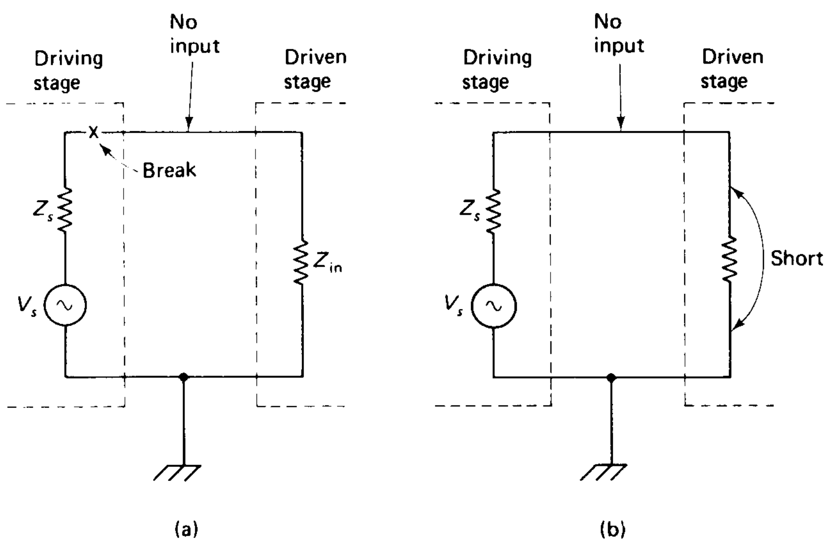


***FIGURE 21-3** Absence of an input signal may be due to defective driving stage (a) or shorted driven stage (b). Breaking the line between them will show which is the case.*

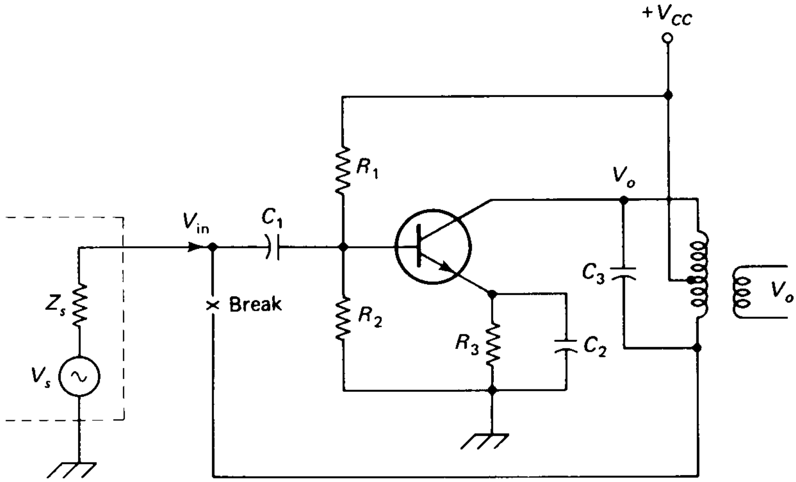


FIGURE 21-4 *If the feedback path of an oscillator is broken and a signal is injected at the proper frequency, the signal will be amplified.*

Oscillators, whose output feeds the input, are a little tougher. However, if the feedback path can be broken and a signal of the proper frequency injected, as in Fig. 21-4, the basic amplifier function can be tested.

Recognizing the Functional Blocks is not always easy; there are at least 50 common ones and designers seem to delight in coming up with new ones and disguising the old ones. However, there is no point in trying to troubleshoot a circuit if we don't know what the input signal should look like, or if we don't know what kind of an output signal it should produce. We could measure the input and output signals, but we wouldn't know if they were right or wrong. This is why we study electronic circuits: to gain an ability to recognize the functional-block circuits and to have an expectation of what the input and output signals should be.

A list of common functional blocks is given in Table 21-1 as a kind of study guide. For each circuit listed, you should ask yourself: "Would I recognize the circuit from its schematic? Can I state what output would be produced for a given input?" Although the list is by no means complete, it may point up areas where further study is required.

TABLE 21-1 *Functional blocks.*

DIODE CIRCUITS	
Detector	Voltage doubler (half-wave)
Voltage clamp	Voltage doubler (full-wave)
Voltage-peak clipper	Voltage tripler
Full-wave center-tapped rectifier	Zener regulator
Bridge rectifier	

TABLE 21-1 (Continued).

AMPLIFIER CIRCUITS	
Common emitter/source	Quasi-complementary symmetry
Emitter/source follower	Class C tuned
Common base	Class B RF linear
Direct-coupled (two-stage CE)	Differential
Phase splitter (paraphase)	Inverting op amp
Push-pull	Noninverting op amp
Complementary symmetry	Totem-pole output
OSCILLATOR CIRCUITS	
Armstrong (transformer feedback)	*RC* phase shift
Hartley (tapped-coil)	Wien bridge
Colpitts (capacitive divider)	UJT relaxation
Clapp (series-tuned)	Tunnel diode
Pierce crystal	Astable multivibrator
DIGITAL CIRCUITS	
Inverter (current-sourcing)	Schmitt trigger
Inverter (current-sinking)	One-shot (monostable)
Lamp/relay driver	Flip-flop (bistable), RS
AND gate	Flip-flop, triggered
OR gate	Flip-flop, type D (data)
NAND gate	Flip-flop, JK
NOR gate	
MISC. CIRCUITS	
Differentiator	SCR firing
Integrator	Frequency doubler
Miller runup	Heterodyne mixer
Current source	Phase detector
Precision rectifier	Phase-locked loop
Sample and hold	Analog switch (*CD 4016*)
Log generator	Timer (*NE 555*)

22

TEST INSTRUMENTS

Once troubleshooting progresses beyond the "quick fix" stage, test instruments become necessary. You will need to know what instruments are required, how to use them effectively, and what their limitations are.

22.1 OSCILLOSCOPE CHARACTERISTICS

The oscilloscope is listed first because, if you have a good one, no other instrument will be used so much. As a minimum, your scope should have a dc-coupled vertical amplifier with a maximum sensitivity of 10 mV/division or less and a bandwidth of several MHz. It should have triggered-sweep ranges covering at least 1 μs to 50 ms/division. Both vertical and sweep should have range-selector switches in a 1–2–5–10 sequence as well as continuously variable controls. Unless your work is confined strictly to audio frequencies (below 20 kHz) you will need a low-capacitance $\times 10$ probe. Other features are listed in order of desirability:

Dual Trace is very handy for comparing two signals, especially to observe time coincidences and phase shifts. A *trigger select*—channel A or B switch—is a useful feature which not all dual-trace scopes possess.

Differential Inputs are necessary to observe the voltage across a component, neither end of which is grounded. Some scopes have this as an option with dual trace: *A and B* (dual) or by switch selection *A minus B* (differential).

A Delay Line is often installed in scopes having a bandwidth of 10 MHz or more. It stalls the vertical signal for about 0.1 μs to give the sweep generator time to get started. This allows the scope to display the events immediately *preceding* the triggering signal, as shown in Fig. 22-1(a).

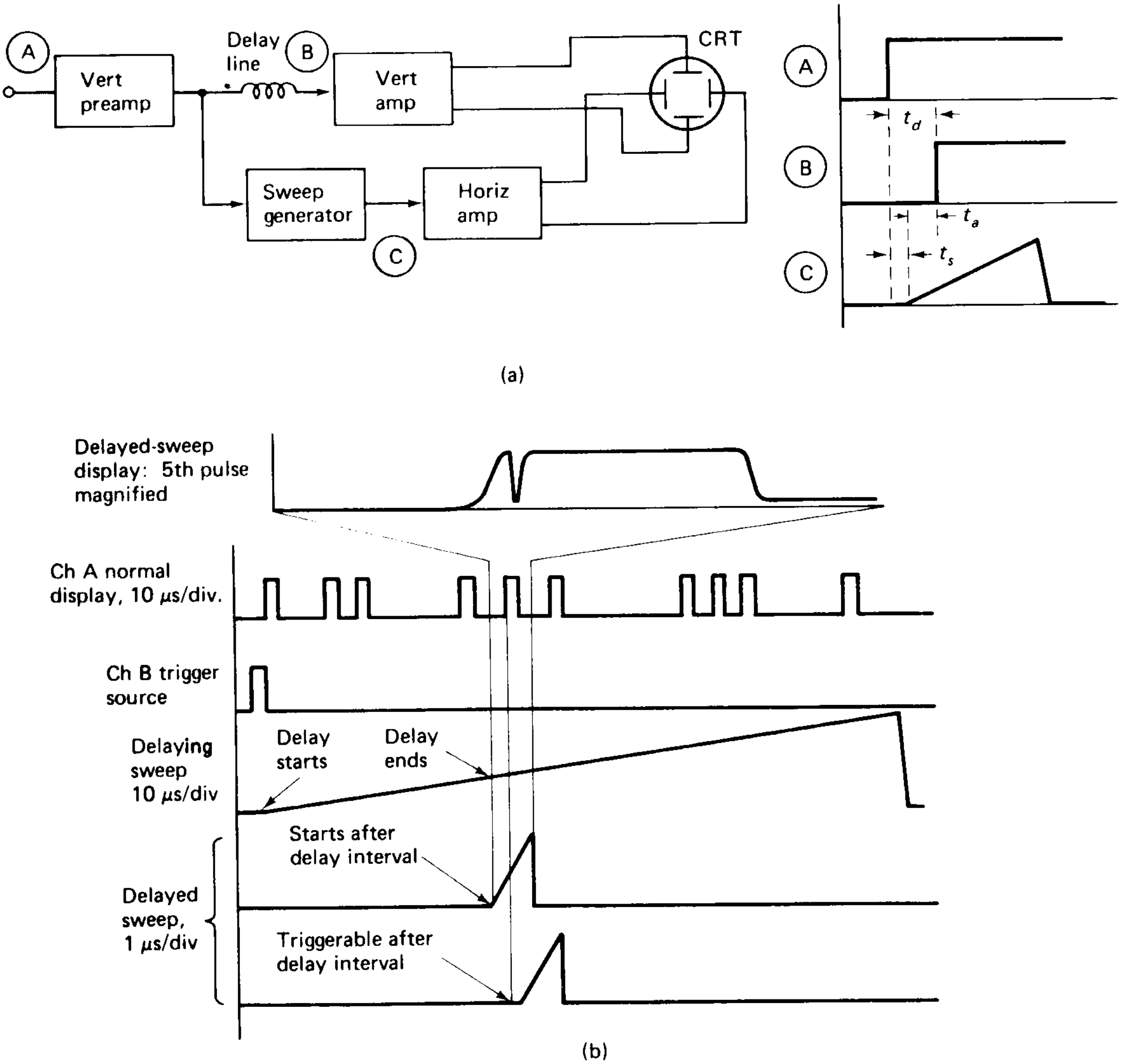


FIGURE 22-1 *(a) A delay line in an oscilloscope allows the sweep to get started before the triggering signal reaches the CRT, allowing the triggering signal to be observed. (b) Delayed sweep allows tremendous expansion of complex waveforms.*

Sweep Magnification is used to expand a display horizontally, so the screen shows only a portion of the total sweep. The horizontal-position control is used to select the portion displayed. Magnification $\times 10$ changes the 10 μs/division range to the

1 μs/division range. Sweep magnification is helpful in displaying complex waveforms whose frequency components are hundreds of times higher than their repetition rates. Digital systems commonly produce such waveforms. The TV composite video signal provides another example.

Delayed Sweep is a more elaborate and more powerful solution to the problem of displaying complex waveforms. Let us say that a digital system generates 10 pulses at irregular times, as in Fig. 22-1(b), on a repetitive basis. A single pulse is also available at the start of each repetition. We wish to examine the fifth pulse more closely. We cannot do this by increasing the sweep speed because the trace would end before we got to the fifth pulse if we triggered on channel B, and the sweep would trigger on all 10 pulses making overlapping images if we triggered on channel A. Magnification would help, but magnification beyond 20 times often results in horizontal "jitter."

The solution is to trigger a slow *delaying sweep* from channel B and set the *sweep delay* for a point just ahead of the fifth pulse. On modern scopes this is facilitated by an intensification of the trace at the delay point. The delayed sweep is set to a much faster rate and can be made to begin automatically at the delay point or to trigger from channel A after the delay point. The figure shows a 10:1 magnification of the fifth pulse, revealing a glitch on the leading edge that could not be seen with normal sweep. Actually, delayed sweep is more commonly used to provide magnifications on the order of 1000, sorting through hundreds of pulses rather than just ten.

Storage is a specialized and fairly expensive feature, used to preserve the display from a single-shot phenomenon. Contact bounce in a toggle switch and strain on a gun barrel when a bullet is fired provide two examples where a storage scope would be desirable. Single-sweep capability is a must in a storage scope. The stored image does not generally have the contrast or sharpness of a conventional display, and the *writing rate* limitation must be considered. A 50-cm/ms limit equals 1/50 ms/cm, or 20 μs/cm. The stored image will not "take" at sweep rates faster than this. Storage CRTs are very expensive and have a limited life in the storage mode. Therefore, they should not be left in the storage mode unnecessarily.

Current Probes are clamp-on devices which use the current-carrying wire as the single-turn primary of a transformer. The secondary voltage is then proportional to the primary current. These passive or transformer-type probes respond to ac above 100 Hz or so only, and simply convert a scope's voltage scale into a current scale—10 mV/division into 100 mA/division, for example.

Active current probes which sense the magnetic field around a conductor and will respond to dc are available at a somewhat higher price.

Sampling Oscilloscopes are used to display signals with rise times below 1 ns and frequencies in the GHz range. Broadband amplifiers supplying CRT deflection-voltage levels to frequencies beyond 1 GHz are all but impossible, but fast-switching diodes at input signal levels are easily obtained. These diodes are switched to obtain a series of samples staggered on successive repetitions of the wave. The samples are then reconstructed into a low-frequency replica of the high-frequency wave which is easily handled by a conventional amplifier. The process is illustrated in Fig. 22-2. It can be seen that the nature of the sampling process precludes the display of single-shot phenomena.

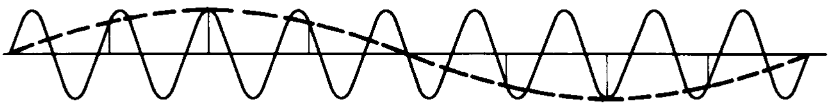


FIGURE 22-2 Fast diodes are switched to obtain samples of a gigahertz wave at successively later points on the cycle. The average of the samples is a low-frequency wave that can be amplified and displayed easily.

22.2 OSCILLOSCOPE LIMITATIONS

Capacitive Loading: A standard $\times 10$ scope probe has an input capacitance of 7 pF. This presents a reactance of 23 Ω at 1 GHz, 230 Ω at 100 MHz, 2.3 kΩ at 10 MHz, and so on. All it takes is money to buy a scope with a 500-MHz bandwidth, but it takes intelligence and skill to make any use of this bandwidth. Loading a 300-Ω source with 100 Ω of capacitive reactance from a scope probe is obviously going to upset the voltage being measured.

Line Reflections: To make matters worse, a 2-m scope-probe cable is nearly a full wavelength long at 100 MHz, and there will be reflections and standing waves on the line unless it is fed by a Z_o source (probe end) and terminated in a Z_o load (scope end). This is almost never the case, so the scope displays the transformed sine-wave voltages and reflected pulses due to the mismatched line, rather than the true waveform at the test point before the probe was connected.

As a rule of thumb, then, the practical limit for conventional scope-probe measurements is about 20 MHz, unless the source is known to have an impedance below about 100 Ω and the probe cable is shorter than $\frac{1}{10}\lambda$.

50-Ω Environment: Accurate measurement of VHF signals is carried out entirely within what is termed a *50-Ω environment*. This means that every output in the system has $R_s = 50\ \Omega$, every line is a coax cable with $Z_o = 50\ \Omega$, and the input to every amp, scope, meter, and so on, presents $R_L = 50\ \Omega$. A meter is not connected directly across a line in such a system, but through a signal splitter, as illustrated in Fig. 22-3.

Bandwidth and Rise Time: A scope's bandwidth is the upper frequency at which a sine wave reads as 0.707 of its actual value. Reasonable accuracy cannot be expected unless the scope bandwidth is at least three times the frequency of the

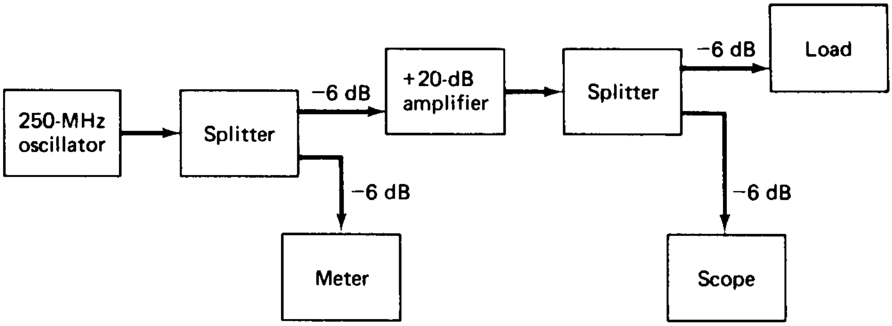


***FIGURE 22-3** High-frequency instruments are designed for 50-Ω inputs and 50-Ω outputs. The meter and scope cannot simply be clipped across a line.*

signal being measured. If the signal is other than a sine wave, the scope must pass all important harmonics of the signal as well, so the scope will require a bandwidth 5 to 10 times the signal frequency.

For square and pulse waveforms, the scope's response speed is measured in terms of rise time, which is the time for the trace to rise from 10% to 90% of full value, given an input pulse with a rise time much faster than that of the scope. Rise time and bandwidth are related approximately by

$$t_r \approx \frac{0.35}{\mathrm{Bw}} \tag{22-1}$$

Where signal rise time is nearly as fast as scope rise time, we can compute signal rise time from

$$t_{\mathrm{sig}} = \sqrt{t_{\mathrm{meas}}^2 - t_{\mathrm{scope}}^2} \tag{22-2}$$

From this it can be seen that a scope rise time three times as fast as the signal rise time will result in a +5% error in the measurement.

22.3 MULTIMETERS

The family of multimeters now includes, in historical order, the VOM (volt-ohm-milliammeter), the VTVM (vacuum-tube voltmeter), its transistorized successor the TVM or FET meter, and the DVM (digital voltmeter). Of these, the VTVM was rendered obsolete by the FET meter, which has itself been all but eclipsed by the DVM. VOMs and DVMs have the following advantages over the oscilloscope:

- They are smaller, lighter, more rugged, and less expensive.
- They are easier to adjust and easier to read.

- They measure continuity and resistance, whereas a scope does not.
- They measure current (dc only for VOMs), which the scope will not do without an expensive special probe.
- They can generally be connected with both probes off ground, whereas the scope generally cannot. (Check your DVM and scope to see if the NEG or GND probe is connected to the chassis or ac-line ground.)

The VOM has an additional advantage over the DVM in continuity testing and tuning for peaks and nulls, in that the relative position of a meter pointer is easier to interpret than the flashing of a digital display. A few DVMs include pointer-type meters on their panels to recapture this convenience in taking quick readings.

The great advantage of the DVM is accuracy, which is commonly 1% to 0.1% on dc volts and 2% to 0.2% on ac, amperes, and ohms ranges, depending on DVM type and cost. This compares quite favorably with the 3%-of-full-scale accuracy which is typical of scopes and VOMs on the dc ranges.

Of course, the great disadvantage of VOMs and DVMs is that they respond strictly to average (on dc) or average absolute (on ac) values, without regard for waveform. A 5-V-dc level and a voltage rippling between 4 V and 6 V will each read 5 V. A 2.83-V p-p sine wave and a 2.22 V p-p square wave will each read 1.00 V rms. Be wary of assuming that a dc multimeter reading is a sine wave. The main reason for preferring the oscilloscope to the multimeter where possible is that it leaves no such ambiguity.

Multimeters do not often have frequency-compensated attenuators, in which case their ac ranges may become totally inaccurate at frequencies as low as 500 Hz (3 pF has $X_C = 10\ \text{M}\Omega$ at 500 Hz).

22.4 SIGNAL GENERATORS

Excepting for the moment some rather expensive special-purpose instruments, the field of signal generators is well represented by AF, RF, pulse, and function generators.

AF Generators most commonly use a Wien-bridge oscillator with some sort of automatic-gain-control circuit to avoid excessive feedback and consequent distortion. Tuning is accomplished by ganged variable capacitors or resistors, and square waves are produced by feeding the sine wave into a Schmitt trigger or saturated amplifier. Frequencies from 10 Hz to 1 MHz are commonly obtained in this manner.

RF Generators use an *LC* oscillator to produce signals from about 100 kHz to 100 MHz. Low-cost units seldom use automatic feedback control, so the output waveshape may be a badly distorted sine wave. Waveshape, amplitude, and output impedance may also change drastically with frequency setting. Provisions for modulating the amplitude or frequency of the output wave are generally made. If wide-band frequency modulation is possible (say, upper and lower limits of 60 MHz and 54 MHz) the instrument is called a sweep generator, and it can be used to display the frequency response curve for an amplifier, as shown in Fig. 22-4. Such a setup is useful in aligning broadband amplifiers such as are found in TV receivers and FM radios.

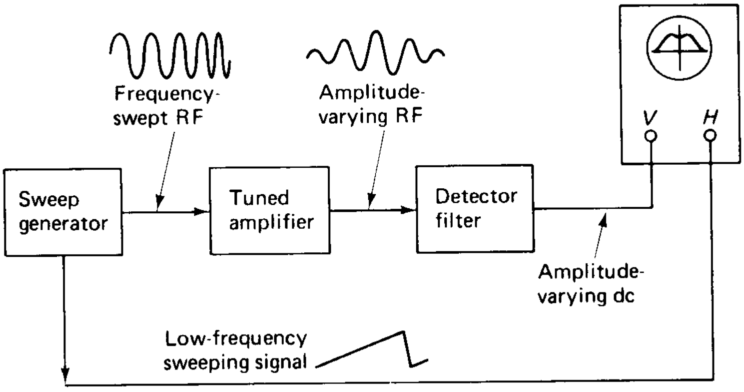


FIGURE 22-4 A sweep generator can be used to display the frequency-response curve of an amplifier on the scope.

Pulse Generators generally use an astable multivibrator to set pulse-repetition rate and a one-shot to set pulse duration. Output amplifiers are used to permit the output pulse to assume variable positive or negative values, mixed with a variable positive or negative dc offset, while maintaining rise and fall times of a few nanoseconds and a constant 50-Ω output impedance. More sophisticated pulse generators provide variable rise and fall times, pretrigger output which provides a short spike about 1 μs before the main pulse output, and double-pulse with variable spacing between the two pulse outputs.

Function Generators use an integrator and hysteresis switch to generate triangle and square waves and a shaping network to produce sine waves from the triangles. Sine-wave purity is generally as good as or better than that of Wien bridge oscillators. Frequencies below 0.001 Hz and above 10 MHz can be produced with one instrument. Constant output amplitude and output impedance is assured. Even midpriced units have dc offset and symmetry (duty-cycle) controls, offering many of the features of the basic pulse generator. Frequency modulation and frequency sweeping is possible via the VCO (voltage-controlled oscillator) input.

Amplitude-modulation capability and frequencies into the VHF range are not commonly available except in high-priced units.

22.5 POWER SUPPLIES

A relatively large number of different power-supply types may be required for development and prototype work, but only the first two in the following list are necessary for service work.

A Battery Eliminator delivers 6 to 15 V dc, roughly filtered and unregulated. It is used to service mobile equipment and must furnish the necessary current, which may run 3 to 20 A, depending on the mobile equipment.

A Variable 60-Hz Autotransformer provides line voltages from 0 to 135 V. It is used to simulate high and low line voltage conditions, since many malfunctions show up only under these extremes. Variable autotransformers are available with current limits from 1 to 10 A and more. Most line-operated equipment is designed to meet its specifications with line voltages between 105 and 125 V. Voltages beyond this should be avoided.

A cheap-and-dirty substitute for a variable autotransformer is shown in Fig. 22-5. It uses a 12.6-V center-tapped transformer to add or subtract 6.3 V or 12.6 V from the prevailing line voltage. The secondary current rating should equal the maximum anticipated load current.

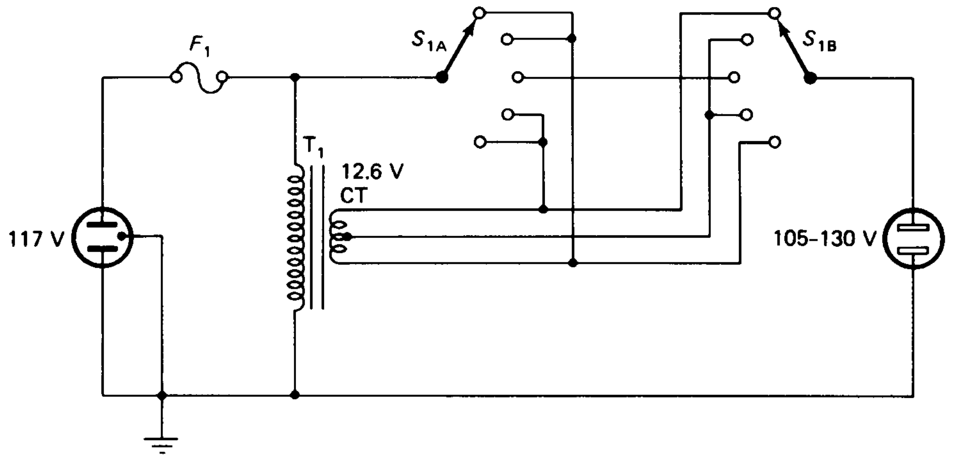


FIGURE 22-5 A 12-V center-tapped transformer can be used to construct high/low line-voltage-test supply.

Low-Voltage Regulated Supplies are commonly used for solid-state development work. Two variable outputs each delivering 0 to 25 V with a maximum current of 0.5 to 2 A will be required, since op amps and many other circuits require balanced ± supplies. A separate +5-V supply for logic ICs is also a great convenience, since this voltage is a widely accepted standard.

Current limiting is a desirable feature wherein the supply voltage drops whenever the load current rises above the set limit. This often prevents an accidental misconnection or a thermal avalanche in the prototype circuit from drawing destructive amounts of current.

Medium-Voltage Regulated Supplies are similar to the low-voltage supplies above, except that they cover the range to about 400 V with a maximum current of typically 300 mA.

High-voltage supplies cover the range to perhaps 20 kV.

22.6 MISCELLANEOUS INSTRUMENTS

The instruments covered so far in this chapter are considered basic to any service bench. This section will describe briefly a number of other instruments with fairly wide ranges of application. In addition, there are literally hundreds of other instruments useful only within a particular specialty (telephone, television, computer, biomedical, automotive, etc.). We will not have space to even mention all of these.

The Electrometer is simply a sensitive voltmeter with a nearly infinite input impedance.

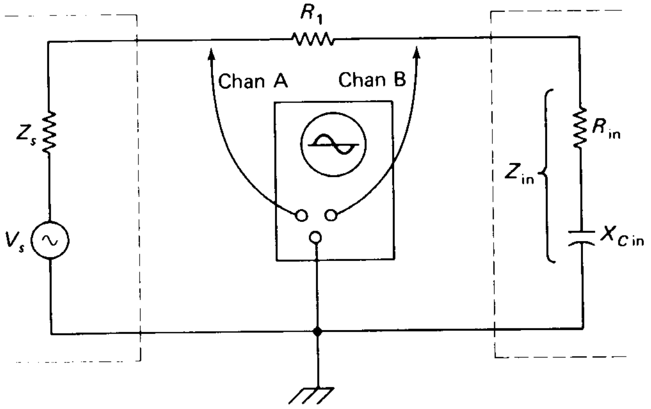
The Vector Voltmeter is a rather complex instrument that measures the phase relationship between two signals as well as their amplitude. A dual-trace scope fills this function at frequencies below a few MHz, but vector voltmeters operate up to several gigahertz.

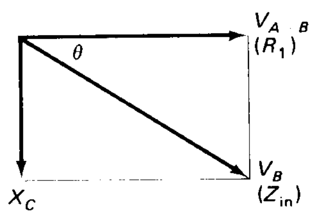
The Vector Impedance Meter is a variable frequency device with two meters, displaying impedance magnitude in ohms and phase in degrees. At low frequencies this information can be obtained by calculation from the voltages and phase angles measured with a dual-trace scope as shown in Fig. 22-6, but the vector Z meter is more accurate and convenient.

The RX Meter is a high-frequency impedance bridge with a built-in variable-frequency oscillator.

The Q Meter provides much the same information as the RX meter, but uses a resonant circuit as shown in Fig. 22-7. An internal capacitor is tuned to resonate an external coil at the input frequency. A voltmeter reads the voltage rise across the capacitor, giving circuit Q. Capacitors are measured by using a standard coil and shunting the variable capacitor with the unknown capacitor. Frequency ranges from 50 kHz to 200 MHz are common for the foregoing two instruments.

Pen-Chart Recorders use a high-gain amplifier to drive a small servomotor which positions a pen on a strip of graph paper. The amplifier is fed by an input voltage and the chart is moved at a constant rate (several cm/day to several cm/s) by a

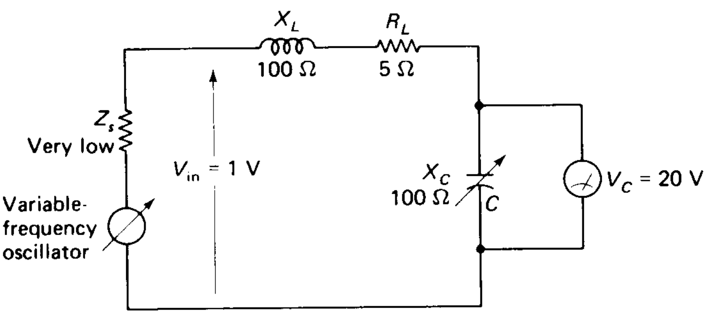




$$\frac{V_{A-B}}{R_1} = \frac{V_B}{Z_{in}}$$

To measure θ, trigger on
Ch B and observe Ch A − B

***FIGURE 22-6** A differential-input scope can be used to obtain input impedance and phase angle at low frequencies.*



$$Q = \frac{V_C}{V_{in}}$$

$$X_L = X_C = \frac{1}{2\pi fC}$$

$$R_L = \frac{X_L}{Q}$$

***FIGURE 22-7** A Q meter measures the voltage rise across C and the value of C at resonance, from which L and Q of a coil can be determined.*

fixed-speed motor and gear box. The pen-chart recorder thus functions as a very-low-sweep-rate oscilloscope.

X-Y Plotters contain two servo systems and can plot voltage against voltage. Current and time scales are often included also.

The Grid-Dip Meter is now built with a transistor or tunnel diode, making the name a bit awkward, but the function is the same. An oscillator constructed with an external plug-in coil is tuned past the frequency of a separate tuned circuit whose resonant frequency is to be measured. The plug-in coil is held as near as possible to the coil of the separate tuned circuit, which will absorb energy from the oscillator when the two are tuned to the same frequency. This shows up as a dip in the oscillator current meter as the frequency is tuned past resonance.

Harmonic-Distortion Analyzers have been mentioned in Chapter 15. They measure the percentage of a voltage that consists of harmonics after the fundamental has been filtered out.

Wave Analyzers selectively tune in the fundamental and each harmonic to provide complete information on the Fourier spectrum of a signal.

Spectrum Analyzers use a sweep oscillator and a CRT display to literally give a picture of the Fourier spectrum of a signal, as in Fig. 15-2.

Intermodulation-Distortion Meters measure the distortion produced on one signal due to the presence of another. You will recall that two signals mixed in a *nonlinear* element will produce sum and difference frequencies by a process known as *heterodyning*. This should not occur in a linear amplifier, but an IM meter measures the extent to which it does occur.

Transistor Testers measure transistor beta and sometimes leakage (I_{CBO}) and cutoff frequency (f_T). Some measure y_{fs} of FETs as well. Some checkers simply look for current gain or signal inversion and give only a good–bad indication.

Curve Tracers, of course, display a complete set of characteristics (transistor of FET) on a CRT. Some curve tracers have a low-duty-cycle pulse mode which allows observation of the transistor at VI levels far above the steady-state maximum limits.

Digital Frequency Counters display the number of zero-line crossings of a voltage on a per second or total basis. For sine or square waves this is frequency in hertz, but for complex waves the display may be $2f$ or $3f$. The upper limit for most counters is 20 to 200 MHz or so. Their great advantage is accuracy, which is typically 5 parts per million (0.0005%). At low frequencies the period (time for one cycle) can be displayed. Some models will also display the ratio between two frequencies or the time between two events.

Logic Probes are small penlike devices with an LED readout: red for logic high, green for logic low, yellow for forbidden voltage between high and low, and blinking for switching between high and low. Most probes are set for standard TTL-logic levels, although other levels are sometimes obtainable. A *logic clip* is actually 16 logic probes built into a single unit which can be clipped to a DIP IC pack to display all pin states at once.

Logic Pulsers supply very short pulses high and low, with the output in a high-impedance or floating state most of the time. The output peak current can thus approach 1 A, while the average current is less than 1 mA. The high peak currents force TTL totem-pole outputs high and low, regardless of what state they are "stuck" in. The driven stages can thus be checked without removing the driver.

The Current Tracer is a probe that senses the magnetic field around a conductor without encircling the conductor. The probe end can be placed on a PC conductor and the probe light will indicate if the track is carrying a switching current of 1 mA or more.

23

TROUBLE LOCATION

Once a malfunction has been isolated to a certain section of an instrument, it becomes necessary to identify the individual faulty component. Checking components once they have been removed from the circuit is simple. The art of troubleshooting at this stage is to locate the defective component without tearing the circuit down to a pile of components.

23.1 SEMICONDUCTOR TESTING

Diodes: Silicon, germanium, Schottky, tunnel, and zener diodes should not have more than 1.1 V forward (positive at the anode) across them. Selenium diodes and HV silicon stacks ($>$ 1 kV PRV) may show 10 or 20 times more, but anything in excess of this indicates an open diode, which should be removed, tested out of circuit, and replaced.

If it can be determined that a diode is carrying current (by a current probe or the voltage drop across a series resistance), but the diode drop is zero or only a few millivolts, the diode is probably shorted. Remove, test, and replace. A shorted power-supply diode may destroy the other diodes, the filter capacitor, and the power transformer, and these should be checked before applying full power.

Transistor In-Circuit Tests: Transistors showing more than 1.1 V base-to-emitter (positive base for *NPN*, negative base for *PNP*) have an open base–emitter junction and must be replaced. Transistors passing any of the following tests can be considered good and should be left alone. Failure to pass one of the tests does not necessarily mean that the transistor is bad, but failure to pass any of them is cause for suspicion.

- Figure 23-1(a): Shorting the base to emitter should cause the collector voltage to rise to V_{CC} and V_{RC} to drop to zero, unless the transistor is normally biased at cutoff.
- Figure 23-1(b): Where the collector load has near-zero resistance, the turnoff may be observed at the emitter resistor. Shorting B to E should cause V_{RE} to drop, unless the transistor is normally biased at cutoff.
- Figure 23-1(c): Where two transistors appear in parallel, both must be turned off to observe the drop in V_{RC}.
- Figure 23-1(d): Where a transistor is already biased off and $V_C = V_{CC}$, a resistor may be added from V_{CC} to base to turn the transistor on. Calculate R to ensure that $I_B < 1$ mA for small-signal and $I_B < 100$ mA for power transistors. Adding R_B should cause V_C to drop.
- Figure 23-1(e): Where the base is driven directly by a transistor, it may be necessary to turn the first transistor (Q_1) off before the second (Q_2) can be tested by methods (a) or (d).
- Figure 23-1(f): In an active transistor circuit, the collector signal appears inverted from the base signal. The base and collector waveforms can be examined on a dual-trace scope to verify this. Disregarding signal levels and distortion, if the collector voltage drops when the base voltage rises, and vice versa, the transistor is basically functioning. It may have high leakage or low beta, but it is not completely dead. If no internally generated signal is present, a signal generator can be used to inject one at the base through 0.1 μF. The frequency may need to be varied widely to find an output signal if the collector circuit is frequency-dependent. If the collector is tied to V_{CC} or ground, it may not be possible to observe any signal there at all.

Transistor Out-of-Circuit Tests: If the in-circuit tests point to a defective transistor, it should be removed and tested out of circuit. A curve tracer is the ideal instrument for this job because it can spot low beta, high leakage, high saturation voltage, low collector breakdown voltage, and thermal instability as well as basic junction failure.

Most transistors faults are simply shorts or opens of one of the junctions, however, and the simple VOM tests of Fig. 23-2 can be used to spot them. Use the $R \times 1$ range for power transistors and the $R \times 100$ or $R \times 1$ kΩ range for small-signal types. Determine with another voltmeter which probe of your VOM outputs a

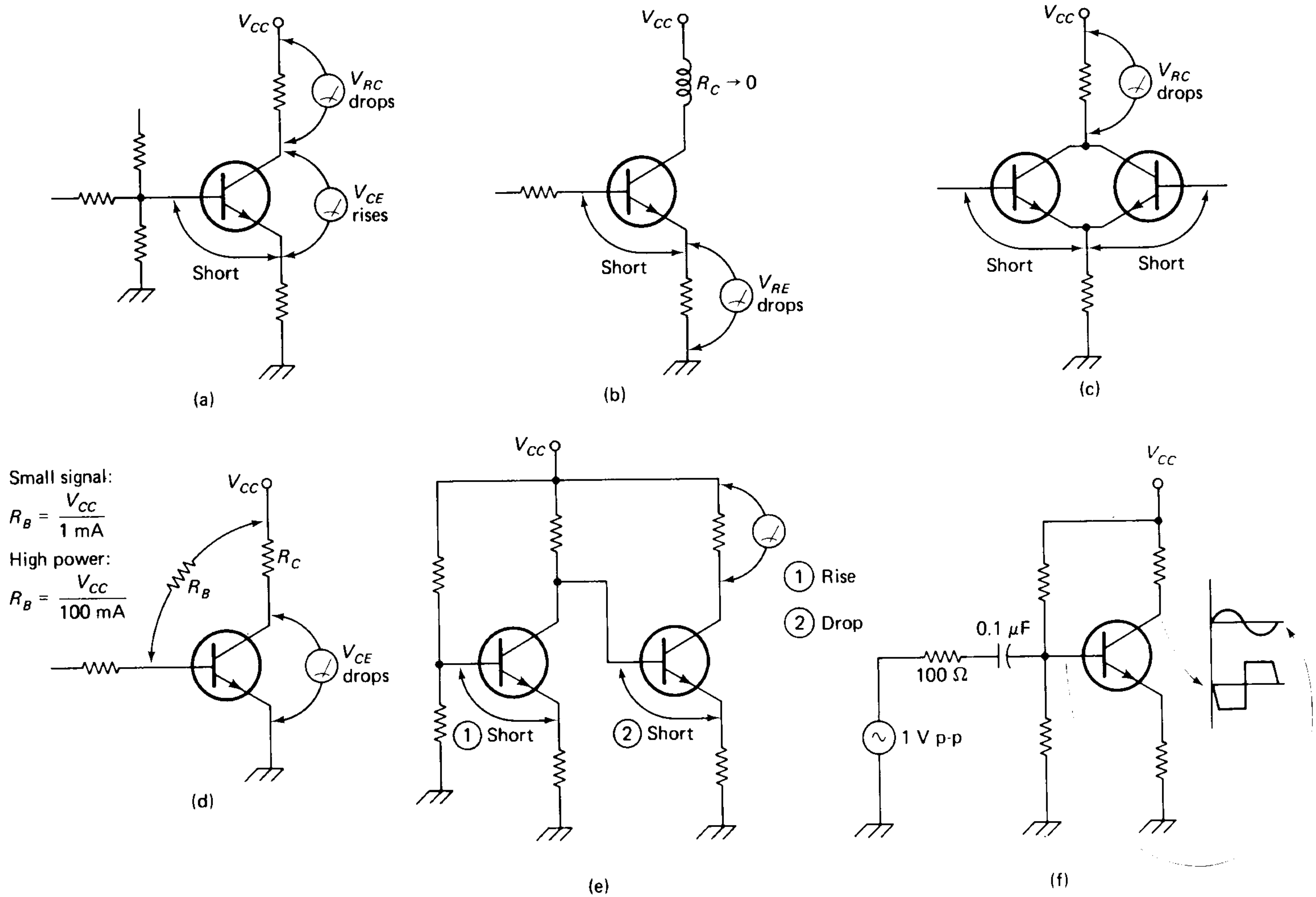


FIGURE 23-1 *In-circuit transistor tests: (a) shorting base to emitter drops collector current to zero, raising V_{CE}; (b) measuring V_E to monitor I_C when $R_C \to 0$; (c) both transistors must be turned off to see V_{RC} drop; (d) turning the transistor on when it is biased off to begin with; (e) turning Q_1 off before testing Q_2; (f) any collector signal inverted from the base signal shows the transistor to be good.*

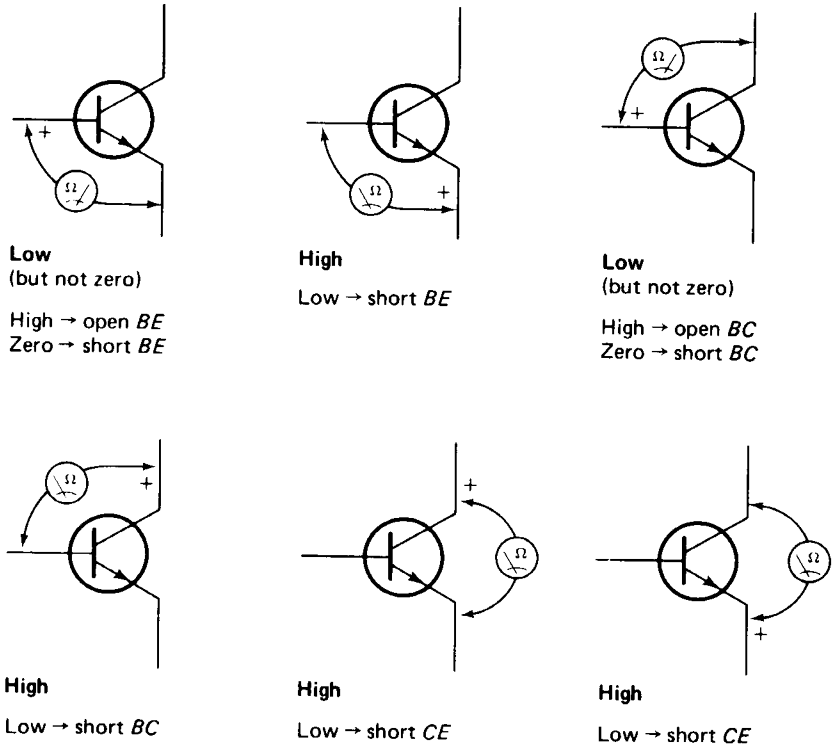


FIGURE 23-2 Out-of-circuit transistor-junction checks. Use a VOM on the R × 100 or R × 1 K scale. Reverse polarities for PNP transistors.

positive voltage on the ohms ranges and mark it. The tests shown in the figure are for *NPN* transistors. For *PNP* types simply reverse the meter probes in each case.

With a little practice, small-signal transistors with abnormally low beta or high leakage can be spotted with the VOM beta test shown in Fig. 23-3. The ohmmeter, on the $R \times 100$ or $R \times 1$ kΩ range, is clipped between the collector and emitter (positive collector for *NPN*, positive emitter for *PNP*). The meter may deflect slightly, reading very high resistance. Large deflections indicate excessive I_{CEO} leakage. Now, without touching the emitter lead, grasp the collector lead with the fingers of one hand and the base lead with the other. The small current bled into the base will allow a large current from collector to emitter, and the meter will deflect greatly, showing low resistance. Small deflection indicates low beta.

FET Tests: Defective FETs can often be spotted in-circuit by the presence of abnormal gate voltage. The gate bias is usually obtained from a simple resistive network, and the expected voltage can be calculated easily, since $I_G = 0$ for a good FET. This is shown in Fig. 23-4(a). Don't forget the loading effect of your meter. Any large deviation from the expected V_G indicates that gate current is flowing. If it is an insulated-gate FET, it is certainly bad in this case. If it is a junction FET, it

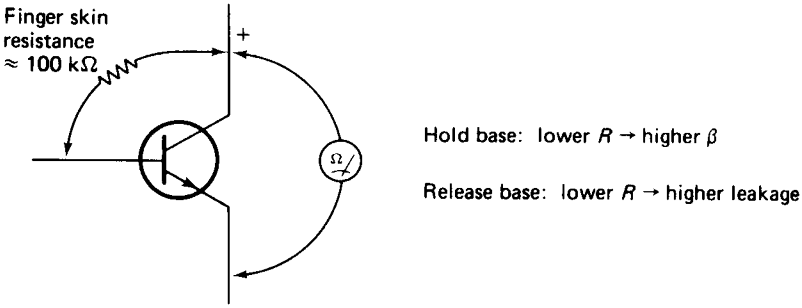


FIGURE 23-3 Ohmmeter beta and leakage test. Reverse polarities for PNP *transistors.*

may be bad, or it may be forward-biased gate-to-source. Check for 0.6-V forward V_{GS}. The phase-shift test of Fig. 23-1(f) can also be used.

Junction FETs can be tested out of circuit with an ohmmeter for diode action between gate and source (high R one way, low R the other way). With the gate shorted to the source, a channel resistance of a few hundred ohms should be measured drain to source, either polarity. These tests are not conclusive but will show up most defects.

Insulated-gate FETs can be checked for the substrate-to-source diode junction and for infinite gate-to-source resistance. Drain-to-source channel resistance (with the gate tied to the source) should range from a few hundred ohms for depletion types to infinity for enhancement types.

SCR Tests: A turned-on SCR should show 0.1 to 1.5 V or so positive anode-to-cathode when conducting. Near-zero voltage indicates a shorted SCR. V_{GK} should never go above about +1.2 V. If it does, suspect an open gate. Shorting the gate to cathode should keep the SCR from triggering, allowing positive voltage to appear from anode to cathode as in Fig. 23-4(b). If no positive voltage appears, suspect an open load or shorted SCR. Out of circuit an SCR should show a diode junction gate-to-cathode, and an open circuit to both polarities anode-to-cathode.

UJT Circuits usually fail to operate because the emitter voltage cannot reach the firing level or because the charging circuit supplies so much current that the UJT holds on once it fires. It is best to unsolder the emitter lead and measure V_C as shown in Fig. 23-4(c). If it does not rise to at least $0.85V_{B2}$, check the charging circuit and C. Next connect a milliammeter from C to B_1. If the current exceeds the valley-current spec of the UJT, the charging circuit is supplying too much current, holding the UJT on.

Integrated Circuits contain, for the most part, transistors, diodes, and resistors which behave the same as their discrete-component counterparts. Do not be afraid to look into the schematic diagram of the internal circuit of an IC. Taken as a whole, it is usually quite formidable, but you can usually penetrate as far as the

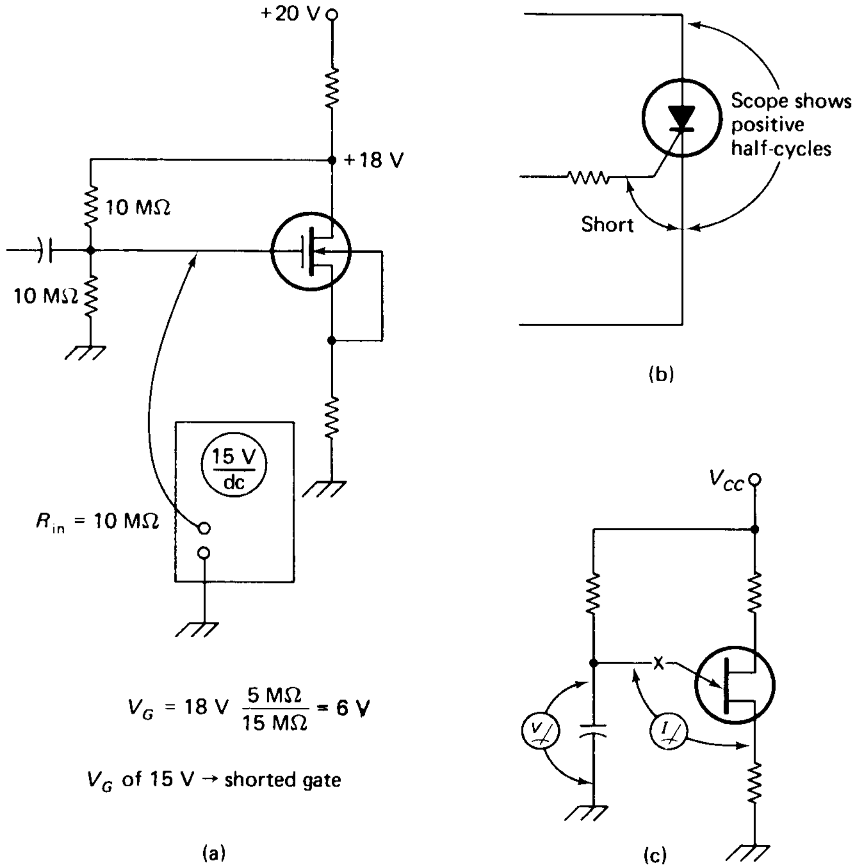


***FIGURE 23-4** (a) Most bad FETs leak current from the channel to the gate. (b) Shorting an SCR gate to cathode should keep it turned off. (c) Breaking a UJT oscillator at the emitter allows measurement of maximum charging voltage and hold-on current.*

base of the first transistor at the input and the collector of the last transistor at the output. This will give you some idea of what to expect and what is completely unreasonable at the input and output terminals.

23.2 CIRCUIT TRACING

Continuity Testing: A great number of problems can be located by checking out the simple expectation that a conductive path should have nearly zero resistance between its ends. A VOM on the $R \times 1$ scale can be used for this, but an audible tester such as the one in Fig. 23-5 allows you to keep your eyes on the circuit. Use needle-point probes to pierce the insulative oxide layer that forms on many conductors, and be sure that the instrument power is off.

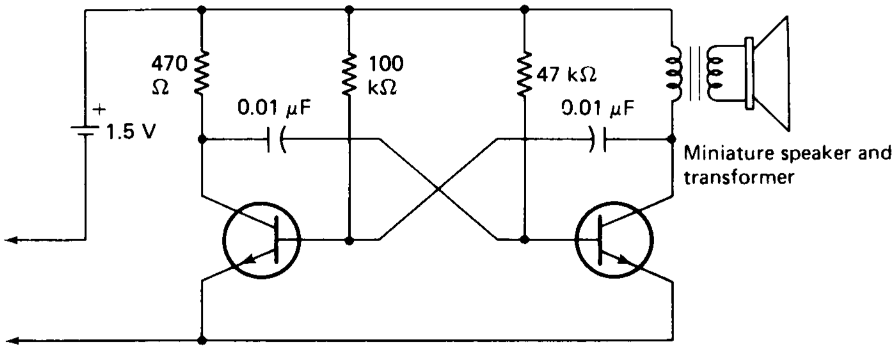


FIGURE 23-5 Audible continuity tester.

When using a scope for troubleshooting with the power on, keep this in mind: If two points in a circuit are connected by a conductor, they will measure the same voltage with respect to ground. If a case arises where two supposedly connected points show different voltages on the scope, you may be sure that the conductive path between them is broken. In this regard, remember that to a 10-ns pulse, 10 cm of wire is not a conductor, but a rather formidable inductor.

Following are some likely places to check for broken continuity:

- The two ends of a cable (broken center conductor or connector joint)
- The actual pin of an IC and the printed-circuit track leading to it (bad connection, especially if IC is in a socket)
- The two ends of a long thin printed-circuit track (broken conductor track or PC board)
- The fixed and moving contacts of a switch or relay (bent, broken, or corroded switch contacts)

Shorts and Opens: Figure 23-6(a) through (c) shows the voltage distribution in a series circuit under normal, short-circuit, and open-circuit conditions, respectively. Notice that the short circuit throws added voltage to the remaining circuit elements, while depriving the shorted elements of any voltage at all. The open circuit deprives all of the circuit elements of voltage, since the entire source voltage appears across the break in the circuit.

To trace a series circuit for shorts or opens, we would start with the scope or voltmeter probe from ground to A, move it on to B, C, D, E, and F, noticing a certain drop in the voltage at each point until zero voltage is observed at F. If there is no voltage dropped across any of the elements until suddenly the entire voltage is dropped, you may be certain there is a break in the circuit [as between D and E in Fig. 23-6(c)].

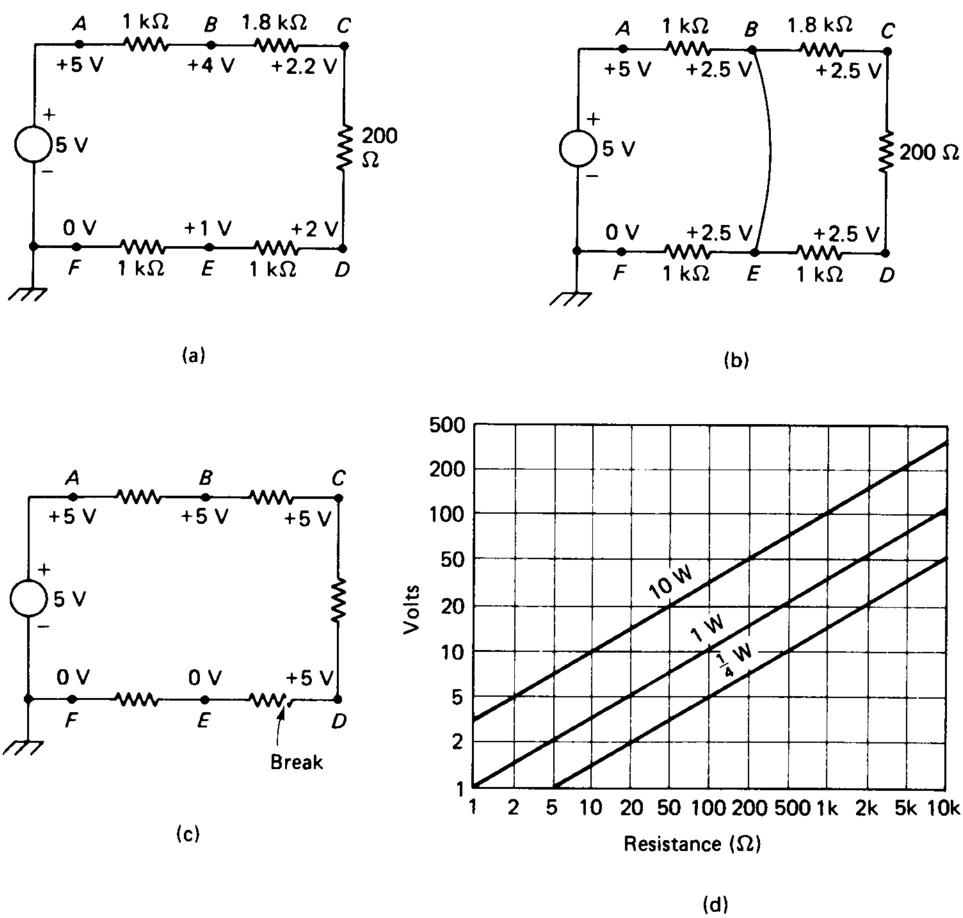


***FIGURE 23-6** (a) Normal series circuit and voltages to ground. (b) Shorted circuit shows no voltage across shorted elements, excessive voltage across others. (c) Open circuit drops all voltage across the break. (d) Chart for quick determination of resistor power.*

If one or more elements are found with little or no voltage drop across them, we may be suspicious of a short circuit, but there may be several other reasons for it:

- Some elements have very low resistance under normal conditions. Fuses, thermistors, and coils may show very little voltage drop.
- Low-value resistors will, of course, drop relatively little voltage, but values in the 100-Ω range are commonly used in series with input and output leads of high-frequency amplifiers to prevent vhf parasitic oscillations. These may show no drop at all at dc and signal frequencies. Power-supply decoupling resistors in the 100-Ω to 1-kΩ range may also show almost no drop at dc.

- Certain resistors may drop almost no voltage under some signal conditions but show a substantial drop under other conditions. The emitter resistor in a class B push-pull or complementary-symmetry power amplifier will show almost no drop at idle, but may drop a volt or more at full signal. Schmitt triggers, one-shots, and flip-flops will show no drop across one collector resistor while dropping almost full V_{CC} across the other. *RC* snubbers used across SCRs and switched inductances will show no voltage across the resistor without switching action.

The chart of Fig. 23-6(d) shows the voltage that will produce a given power in a given resistance. Any resistor found to be dissipating more than its rated power should be suspected of being part of a shorted circuit. Power resistors (1 W or more) dissipating less than one-fourth of their rated power should be investigated for possible involvement in an open circuit.

23.3 TROUBLESHOOTING LOGIC

Often, it will be impossible to isolate a problem down to a single defective component by simply taking measurements on the circuit as it stands. Then it will be necessary to form some expectations of what should happen if the circuit is deliberately altered.

Exercise 1: Power-Supply Failure: Figure 23-7 shows a typical power-supply-input circuit. Let us say that it has blown a fuse. We observe by the black smear inside the fuse glass that its demise was quick and violent, rather than leisurely, so instead of consigning another fuse to the same fate, we connect a 115-V 150-W light bulb across the fuse holder. This will limit the input current to a little more than 1 A, even if the input is shorted. Upon measuring the primary of T_1 we find only 1.1 V ac. The rest is dropped across the lamp. This could be caused by a short in C_1, but it would then be carrying over 1 A at 1.1 V and would be getting quite warm. We feel it (power off first, please), find it at room temperature, and conclude that it is not shorted. We discount as too remote the possibility that R_1 and I_1 both are shorted. That leaves T_1 and T_2.

T_1 is large (12 cm on a side) and has low enough internal resistance that a short at its secondary could reflect as nearly a short at its primary. T_2 is small, (5 cm on a side) and its winding resistance would prevent a secondary short from reflecting back such a low resistance to the primary. The problem could then be due to:

1. T_1 primary short
2. T_1 secondary short
3. D_6 short
4. D_1 and beyond short
5. T_2 primary short.

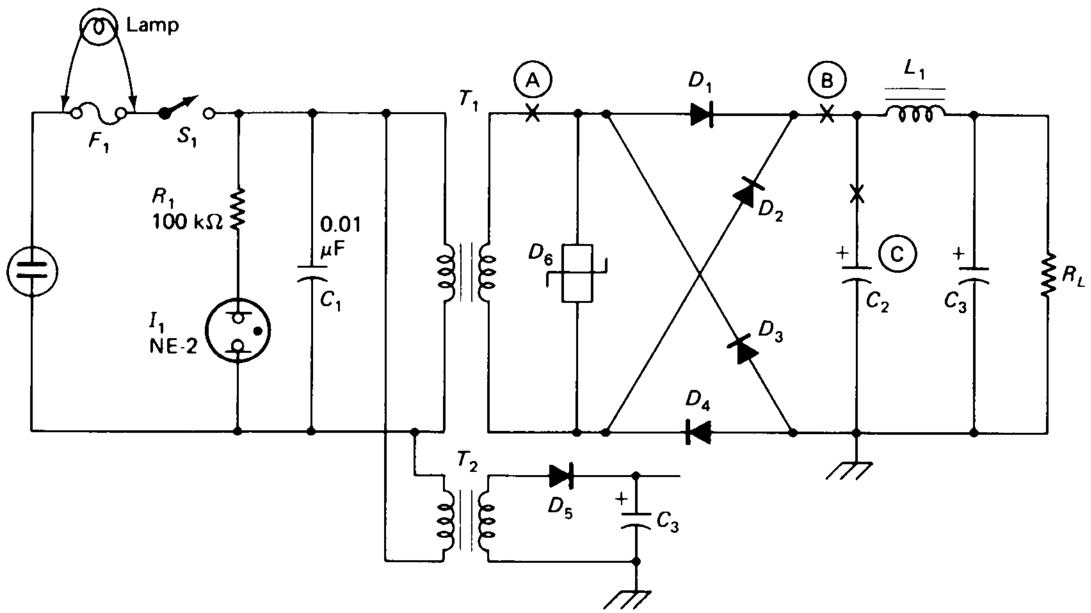


FIGURE 23-7 Power-supply troubleshooting exercise. (See the text.)

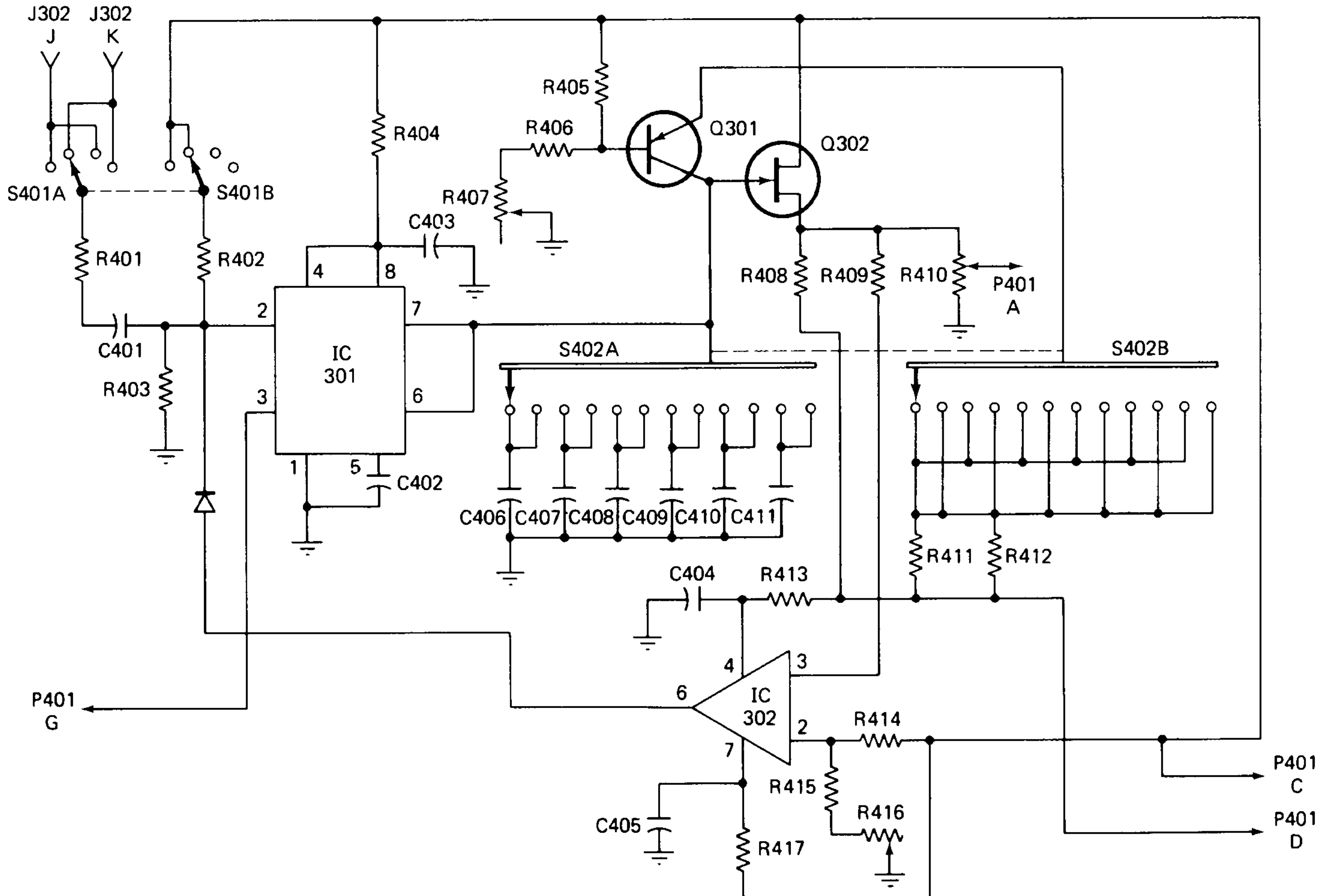
To eliminate among these five possibilities, we decide to split them approximately in half by unsoldering the T_1 secondary lead at A. This causes the 150-W lamp to go dim and we now measure 115 V at the T_1 primary. T_1 and T_2 are obviously good, so the problem lies in D_6 or D_1 and beyond.

Next we move the thyrector D_6 lead to the left of the break at A and notice that the lamp remains dim. That means that D_6 is good. Hoping the diodes are good but C_2 or the load is shorted, we break the circuit at B. Applying power, we see the lamp light again. At least one diode is bad. We unsolder one end of each diode and test it with an ohmmeter. Two are shorted and one is open. We replace them.

Mindful that C_2, C_3, or the load may be shorted, we replace the 150-W lamp at F_1 with a 25-W lamp to avoid blowing out our new diode, reconnect A and B, and again apply power. The 25-W lamp lights and we measure zero voltage from point B to ground. C_2, C_3, or the load must be shorted.

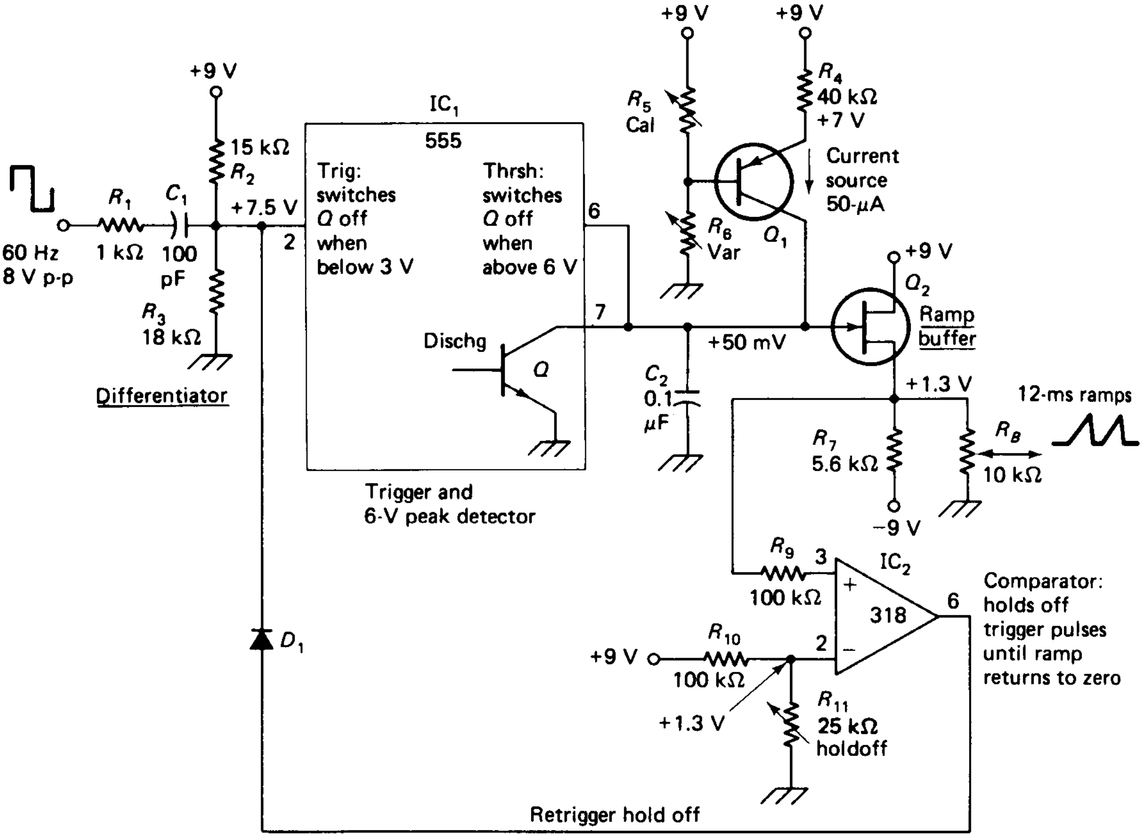
We disconnect C_2 at point C, short it for a moment from force of a well-ingrained habit, and test it with an ohmmeter. Shorted! Replacing C_2, the 25-W lamp glows more dimly, leaving 50 V across T_1, so we graduate to the 150-W lamp. Things still look good, so we replace the fuse and let the instrument cook for a while before returning it to service.

Exercise 2: Sweep-Generator Failure: Figure 23-8(a) shows a sweep generator from a special-purpose oscilloscope. The output voltage hangs at +1 V dc, but cannot be made to produce ramps.



(a)

FIGURE 23-8 *Sweep-circuit troubleshooting exercise. (See the text.)*



(b)

FIGURE 23-8 *(Continued).*

A serious problem is that the schematic seems to have been prepared for purposes of component inventory rather than servicing. Little progress will be made in troubleshooting until the functional relationships in this circuit are clarified. The technician should not hesitate to modify and annotate a copy of the schematic, or even to redraw it in simplified form if that will make its function more clear.

Figure 23-8(b) shows a more usable schematic with voltage measurements included and range-switching and decoupling components omitted. The circuit operation is now clear: C_1 couples the edges of the square wave to the trigger input of the 555, bringing it below +3 V on the negative edges. This causes the discharge transistor to turn off, and current source Q_1 charges C_2 at a linear rate. The source of Q_2 follows this ramp, offset by about +1.25 V.

As long as the source of Q_2 is above $+1.3$ V, op-amp comparator IC_2 holds the *555* trigger input positive forcibly via D_1, so no more negative spikes are allowed to the trigger input. Once the ramp at pin 6 of the *555* reaches 6 V, pin 7 discharges C_2 rapidly, and the FET source brings pin 3 of IC_2 below pin 2. The comparator output then goes negative, turning off D_1 and allowing C_1 to deliver another negative pulse to the trigger input.

A 60-Hz square wave is available at the input, but a scope check at pin 2 of IC_1 reveals neither positive nor negative spikes. This is not surprising, since the time constant of C_1 with R_1, R_2, and R_3 is 1 μs and the pulse rate is one every 16,700 μs.

A check of the *318* op-amp specs shows that the output current is limited to about 20 mA and the output can be short-circuited indefinitely, so we feel safe in shorting trigger pin 2 of the *555* to ground for an instant. A single linear ramp of about 4 V pk is seen on the scope at the output each time this is done. Apparently, everything works except that the *555* is not receiving trigger pulses.

A scope probe at pin 6 of IC_2 shows $+8$ V dc, even when the *555* is artificially triggered. Either IC_2 is bad or it is not being fed the proper inputs. Output pin 6 should go negative if input pin 3 goes below pin 2. Not wishing to upset a calibration setting, we leave trimpot R_{11} alone for a while and short pin 3 to ground. Pin 6 goes negative and ramps appear at the output.

Convinced that the Q_2 source voltage was simply too high for the setting of R_{11}, we readjust R_{11} and proper operation is restored. However, realizing that the malfunction must have been caused by a change in some circuit parameter, we spray clean the wiper of R_{11} and change Q_2 and IC_2. The \$2 component cost is considered a small price to pay for insurance against future failures, which would certainly occur if one of these semiconductors were undergoing parameter changes due to loss of its hermetic seal.

23.4 INTERMITTENT TROUBLES

The most exasperating troubles are the intermittents: bad in the field but working perfectly once you get them on the service bench. Here are some tips for locating intermittents more efficiently:

Check It in the Field: If the problem is intermittent, don't just pick up the instrument and take it back to the bench. If possible, have the operator call you when the trouble is present and let him show it to you. You may find that he isn't using the instrument properly or you may see that he has been abusing it, thereby causing the problem.

If the problem is present, flex all the wires and connectors and tap the control knobs and switch handles while observing the instrument's operation. These externals are most likely to have suffered abuse, and you may be able to locate the trouble or at least narrow it down right in the field.

Try to Induce the Trouble: Once the instrument is on the bench, try to get the malfunction to show up. Removing the cabinet often provides enough ventilation to keep temperature-related problems from occurring, so it is a good idea to use a heat lamp to warm the suspected area of the circuit board. Be careful not to melt any plastic parts.

Aerosol cans containing a spray coolant are readily available and are excellent for locating temperature-sensitive components. The instrument is warmed up until the intermittent problem appears; then each component in the malfunctioning area is sprayed with the coolant until one is found that restores proper operation when cooled. Integrated circuits, transistors, and capacitors are the most likely components to check. Avoid getting the coolant on the hot glass of vacuum tubes or lamps by using a sheet of stiff paper as a shield.

If the trouble cannot be induced by heating or cooling, try high and low line voltages (105 to 130 V) using a variable autotransformer. Let the instrument rest at each extreme for several minutes.

Some problems that appear to be intermittent defects are actually interference from electric motors, CB radios, pocket calculators, and so on.

If the intermittent will not reoccur, tag the instrument in an inconspicuous place, giving the date and nature of complaint, before returning it to service. If it appears again with the same complaint, the tag will at least confirm the validity of the complaint, and you may be able to deduce from the conditions under which it malfunctioned what the causes of the trouble might be. Perhaps both failures occurred under high altitude, or high humidity, or under the same operator.

23.5 CABLE FAULTS

Cables are often flexed, stretched, trampled upon, and subjected to sun, rain, and ice. Small wonder, then, that they are frequently the cause of system malfunctions.

Broken End Connections are very common in test leads and instrument probes. The simple ohmmeter tests of Fig. 23-9(a) will verify shorts and opens, and flexing of the cable and connectors may produce an intermittent connection which will identify the location of the flaw. Often, however, the ohmmeter will reveal that the center conductor is broken, but leave unanswered the question: "At which end?" To avoid removing both end connectors, the test setup of Fig. 23-9(b) can be employed. A break at the far end of the cable will result in several volts of line-noise pickup on the scope, whereas a break at the scope end will give only a few millivolts of noise pickup.

Breaks in Underground Cables can be located quite accurately by measuring the capacitance between two cable conductors with a capacitance bridge, as shown in Fig. 23-10(a). If a check of a sample of the cable shows the conductors to have a capacitance of 43 pF/m and the buried cable measures 0.028 μF at the point of

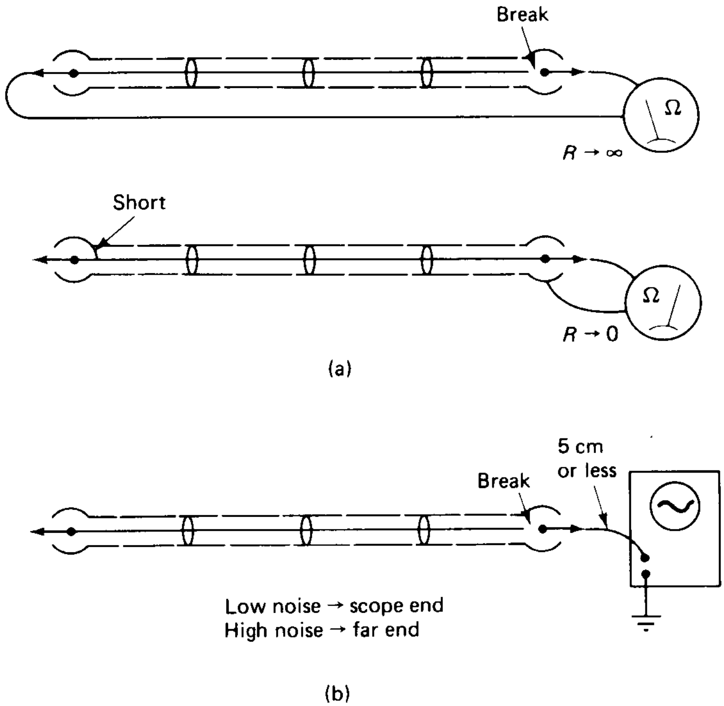


***FIGURE 23-9** (a) Ohmmeter checks for cable shorts and opens. (b) Noise-pickup test to determine which end of the cable contains the break.*

origin, we may expect to find the break if we dig 0.028 μF/43 pF/m or 651 m from the point of origin. Better accuracy might be obtained if the capacitance of a wire pair between two access points were measured before a break occured and the capacitance per unit length calculated and recorded from these data. Pressure and stray capacitance from the packed earth may introduce errors if the capacitance per meter is taken from a short cable sample in air.

Shorts in Underground Cables or insulation breakdowns can be located by means of the *Murray Loop* test shown in Fig. 23-10(b). The setup actually comprises an elementary Wheatstone bridge, with legs R_2 and R_4 consisting of lengths of conductor in the cable. A second conductor, intact but otherwise identical to the faulted conductor, must be available and must be connected to the faulted conductor at the far-access point as shown. The ratio of resistance R_4 (to the fault) to R_2 (out the good conductor and back to the fault) is determined by the bridge, and an equation is developed relating the distance to the fault d_F to the distance between access points d_L:

$$\frac{d_F}{d_L} = \frac{2R_3}{R_1 + R_3} \tag{23-1}$$

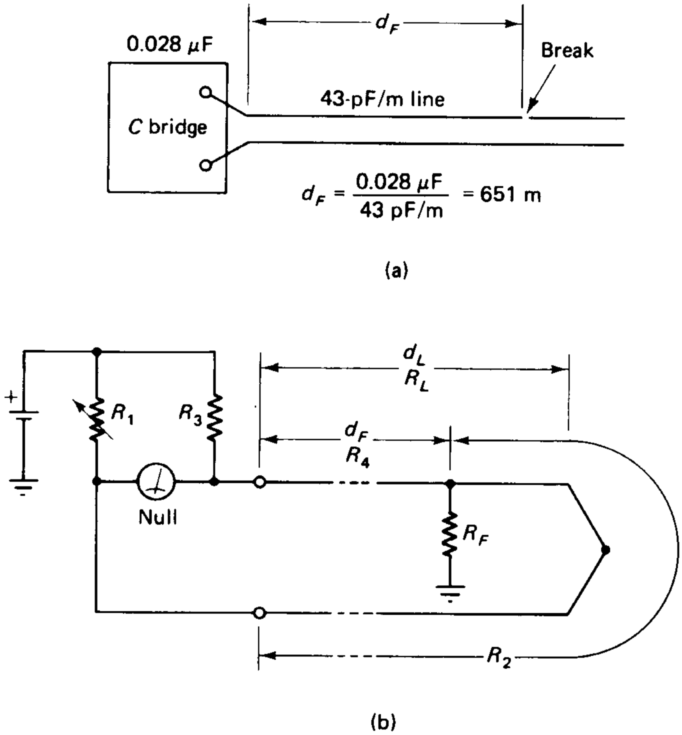


***FIGURE 23-10** (a) Determining the distance to an underground break by interconductor capacitance. (b) Murray loop test to determine distance to an underground short.*

The derivation of this formula makes use of the Wheatstone-bridge-balance equation and the assumption that resistance is proportional to length for the buried conductors.

$$\frac{R_2}{R_4} = \frac{2d_L - d_F}{d_F}$$

$$\frac{R_1}{R_3} = \frac{R_2}{R_4}$$

$$\frac{R_1}{R_3} = \frac{2d_L - d_F}{d_F}$$

$$R_1 d_F = 2R_3 d_L - R_3 d_F$$

$$(R_1 + R_3)d_F = 2R_3 d_L$$

$$\frac{d_F}{d_L} = \frac{2R_3}{R_1 + R_3}$$

If the fault is a short to another conductor in the cable rather than to ground, the negative side of the battery is connected to this other conductor. If the fault is resistive rather than a dead short, R_F simply appears in series with the battery and does not enter into the calculations. If a second conductor identical to the first is not available, it may be worth the trouble to string a temporary one above ground, since the only other alternative might be to dig up the entire length. As a check, it would be wise to repeat the test with the short at the left access point and the bridge at the right access point.

Cable Deterioration is a serious problem, especially with high-frequency coaxial cable. Deterioration is evidenced by characteristic impedance changes and consequent reflected signals, and by dielectric absorption, evidenced by excessive signal attenuation at high frequencies (say, above 100 MHz). Cable life can be extended greatly by proper installation. Here are some tips:

- Support in-air runs from plastic-coated steel wire. Don't let the cable carry its own weight.
- Keep the cable out of the sun if possible, but if it must be in the sun, keep it in free air to avoid heat buildup and standing water.
- Do not expose connector ends or splices to rain or standing water. Keep splices and connectors higher than the cable runs so condensation will run away from—not toward—the break in the cable jacket. Build a birdhouse-size cover over splices if they must be made in the open.
- The smallest break in a cable jacket will soak up water like a sponge if given the chance. Capillary action will send this water far along the inside of the cable, ruining great lengths of it. Tape is completely useless in sealing breaks. It may last a week or a month, but can never be considered permanent. Most paint-on dopes and sealers are also of dubious reliability. Heat-shrinkable tubing properly applied offers the best chance, but keeping breaks out of the weather is the real solution.

24

TROUBLE PREVENTION

Electronic equipment failures can cause economic and social problems ranging from irritated customers to airplane disasters and it is shallow to concentrate on troubleshooting without giving equal attention to trouble prevention. The chain of reliability includes as its links the component manufacturer, the circuit designer, the assembler, the operator, and the service technician. The process of assuring reliability begins at the earliest stages of product development.

24.1 DESIGNING FOR RELIABILITY

Component Specifications: In a high-reliability product, design specifications are drawn up on the desired temperature, vibration, humidity, and altitude limits of the finished product. A guarantee must then be obtained from the manufacturer of every component going into that product that his components will meet or exceed these specifications, or else a program of incoming-component testing must be initiated and continued for as long as the components are being received. This is painstaking and expensive, but product reliability cannot be assured if component reliability has not been assured.

Design Documentation: Responsible design is not a matter of building a prototype with pots for each resistor and then measuring all the values once they have been adjusted for best operation. If the company is building an audio amp specified to deliver 25 W to an 8-Ω load down to 100 Hz, the designer should be able to produce a set of calculations showing that for the lowest guaranteed beta of the transistors at the lowest specified temperature and line voltage, and with all resistors and capacitors at their worst-case limits, these specs will still be met. The number of calculation sets that is required to take each of n components to its upper and lower limit is 2^n. To vary 10 components, this means 1024 calculation sets. In the days of hand calculation this would have been limiting, but ready availability of computer circuit-analysis programs removes all obstacles to its implementation.

Safety Factors: A common and highly recommended practice is to make the worst-case design limits a little wider than the specified operating limits: design for 102 to 130 line volts and publish 105 to 125 line volts, for example. Resistor power and nonelectrolytic capacitor voltage are usually chosen for a safety factor of at least 2. Electrolytic-capacitor voltage, inductor current, and semiconductor power are typically given a 50% safety margin.

24.2 PROTOTYPE CONSTRUCTION

Breadboards, named for the wooden versions that were used several decades ago, are finally available at a reasonable price. They contain hundreds of spring-loaded sockets designed to receive a solid-wire end, component lead, or IC pin. Here are some hints for breadboard prototyping:

- Don't force large leads (like 1-W resistors or big electrolytics) into the sockets, or they will become too loose to hold the smaller leads. Solder on a lighter wire.
- Try to make the physical layout of the breadboard components look as much as possible like the schematic layout. This makes it much easier to follow the circuit and find things during troubleshooting.
- Use stick-on numbers on the ICs and transistors corresponding to numbers on the schematic.
- Do not twist wire ends together—the connection is bound to be intermittent. Use solder or an alligator clip.
- Do twist the insulated leads from power supplies and meters together—about one turn every 2 cm. This minimizes noise coupling and self-oscillation.
- Keep all leads short—about 30 cm maximum. Shielded leads may be longer. This combats self-oscillation.
- Self-oscillation may still be a problem because of the longer leads and lack of shielding in the breadboard prototype. Using medium-power transistors

in place of small-signal types often helps because of their lower f_T. Ferrite beads on the input and supply leads, and a 100-pF capacitor from collector to base may also limit self-oscillation.

24.3 COMPONENT MOUNTING

Tool Selection: Obtaining a quality set of cutters and long-nose pliers for miniature work may require more travel and more money than you like, but it is worth the trouble. Neither should be more than 10 cm (4 in.) long, and the working end should be as narrow as possible. Insulated handles and spring-open jaws are advisable. A good cutter will snip a piece of paper as cleanly as a scissors and nip a hair off the back of your arm without pulling. A good pliers will close at the tip first, then back toward the jaws as more pressure is applied. Try picking up a sheet of paper with the very tip of the pliers.

Have a larger set of tools available for brute work. Don't use your fine cutters for shearing bolts or your good long-nose for a wrench.

The kick when a sharp cutter shears a hard wire can send the wire end flying five meters or more. Point the wire end down, not up or at anyone's face when cutting. This kick has been known to cause component damage from the shock telegraphed up the lead. Placing a temporary bend in the lead between the component and point of cut will keep the shock from reaching the component.

Circuit-Board Mounting: Circuit-board components should be mounted horizontally unless they are designed for vertical mounting. The lead bends should allow plenty of room for a probe hook—do not crush the leads against the side of the component. If the component is operating near its maximum power dissipation, it should be kept a few millimeters above the surface of the board to allow air circulation.

The component leads should be cut leaving 2 or 3 mm on the foil side of the board. The lead end should be clinched back along the PC conductor track to hold the component. In prototype instruments where there is a strong likelihood that some components will have to be changed, it is advisable to cut the lead at 1 or 2 mm and solder it without clinching. When replacing a component, the new lead should be cut at an angle, leaving a sharp point for insertion into the hole, which may be partly clogged with solder. Figure 24-1 illustrates the foregoing concepts. Components weighing more than about ten grams should not be mounted by the leads alone, but should have a clamp to hold them.

Terminal Mounting: Terminals should be used wherever a component lead is not long enough to reach its destination. A hanging splice along a wire run or a midair junction of wires or component leads is not acceptable.

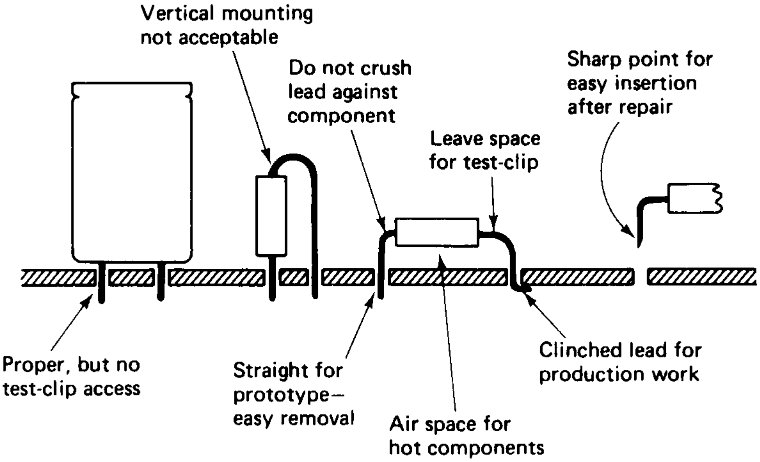


FIGURE 24-1 Proper and improper mounting of components on a printed circuit board. (See the text.)

Wires and leads should not be stretched to a terminal point. A "service loop" relieves strain and permits a reconnection if the wire should break at the terminal.

For production work, the lead end should be wrapped completely around the terminal and clinched, so that the joint is mechanically solid before soldering. A terminal with three or four wrapped connections is very difficult to unsolder, however, so for prototype work you may wish to use a single-bend hook at the terminal. Figure 24-2 illustrates these ideas.

24.4 WIRING

Wire Selection: Stranded wire must be used where there is any possibility that the wire will be flexed in use or during servicing, or in equipment subject to continuous vibration, such as in aircraft. Solid wire is easier to connect to most terminals, but may be used only where flexing will not occur.

Insulation should be selected for the anticipated environment. PVC plastic melts at a relatively low temperature, rubber deteriorates with age, cotton absorbs moisture, and enamel may crack or scrape off. Teflon is excellent but is a little more expensive and difficult to work with.

Wire Stripping: A nick in a wire becomes a stress point which leaves the wire vulnerable to a complete break from relatively mild flexing. Wire stripping with a knife or adjustable-pliers-type strippers is therefore absolutely forbidden. Semiautomatic hand strippers with nonadjustable die-type cutters are acceptable, but care must be taken to use only the die hole for the proper-size wire. Using a

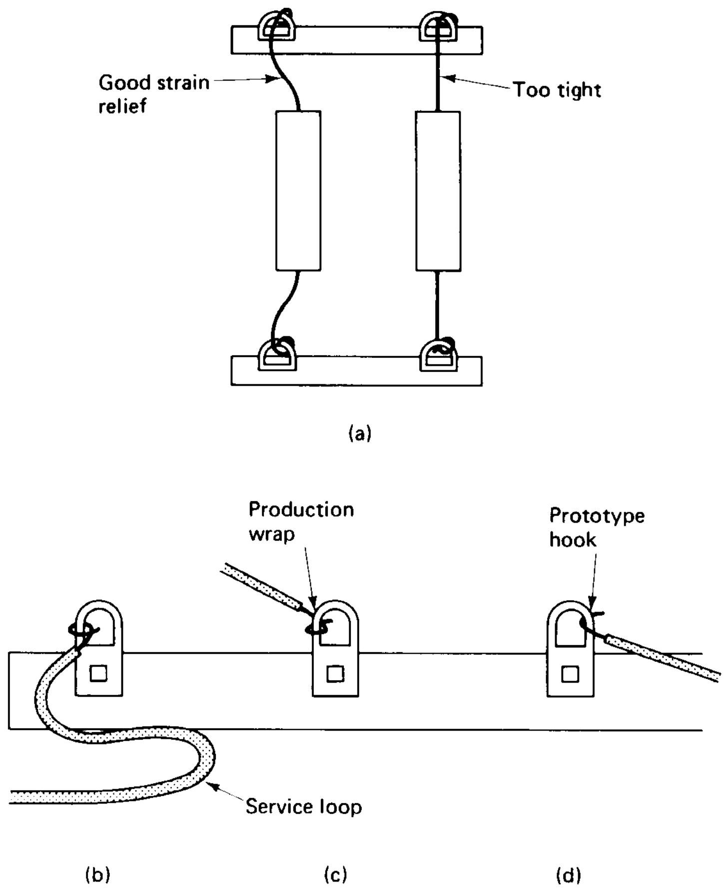


***FIGURE 24-2** Proper wire-lead terminations: (a) leave slack in component leads for strain relief; (b) leave enough extra wire so that it can be cut and reterminated; (c) wrap and crimp wire around terminal on production jobs; but (d) use a simple hook where wire may have to be removed.*

too-small hole will certainly leave a nick. For production work all holes but the proper one should be ground out or single-hole dies should be ordered.

Thermal wire strippers use an unsharpened hot blade to melt rather than cut the insulation. The possibility of a nick is completely eliminated with this method, almost no force is required on the wire to pull off the insulation, and the stripper can operate in closer quarters than most cutting-type strippers. Heat and timing must be skillfully controlled to achieve a clean strip, however.

Preparation for Soldering: The length of the stripped wire end must be carefully controlled so that no more than 2 mm of bare wire is exposed between the solder joint and the insulation. On the other hand, the insulation must not protrude into the solder joint.

Care must be taken to avoid disturbing the lay of the strands after stranded wire has been stripped. It is common practice to *tin* the stripped ends of stranded wires by dipping them first in liquid rosin flux, then in a pot of molten solder. This keeps the strands from fraying and allows the wire end to be bent around a terminal without separated and protruding strands. If a ball of solder forms at the end of the wire after tinning it can be shaken off when hot or cut off when cold.

Stripped wires, etched circuit boards, terminals, and other items to be soldered should be kept in plastic bags if they are to be stored for any length of time before use. This prevents oxidation, which would hinder the soldering process.

24.5 SOLDERING EQUIPMENT

Solder Selection: Solder is an alloy of tin and lead which, when heated above its melting point, is capable of bonding with other metals, such as copper, near their surface. The alloy SN 63 (63% tin, 37% lead) is eutectic (yoo-TEK-tik), which means that it passes immediately from the solid to the liquid state (at 183°C for tin–lead). Other alloy proportions pass through a plastic state when cooling from the liquid state. This is undesirable because a cracked joint can result if the joint is moved while the solder is plastic. Noneutectic alloys also melt at a higher temperature. The following table summarizes the commonly available alloys:

Tin/Lead	*Liquid Above:*	*Solid Below:*
63/37	183°C	183°C
60/40	190°C	183°C
50/50	213°C	183°C
40/60	236°C	183°C

The 60/40 alloy is generally acceptable, but 50/50 and 40/60 should not be used for miniature work. Only the highest-quality solder from a reliable manufacturer should be used. This is no place to look for bargains.

Liquid solder absorbs gases from the air and will oxidize if overheated or held too long in the liquid state. Oxidized solder has a dull granular appearance and forms jagged peaks when the pencil tip is removed. Once solder has oxidized, it must be removed from the joint and new solder applied.

Flux: Pure solder will not bond to most metals because of the thin oxide layer that forms on the surface of a metal exposed to air. Solder will bond to most metals if the oxide can be removed and the solder applied before a new oxide layer forms. Mechanical or ultrasonic cleaning under the molten solder or in an inert-gas atmosphere can be used, but application of a chemical *flux* that removes the oxide when heated is the most practical and widely used method.

Fluxes are available that will permit the soldering of many aluminum and steel surfaces, but these should be used only under the most carefully controlled conditions, as they are likely to leave a residue that will cause corrosion after the work is completed. Solder for electronic use contains one or several hollow cores filled with *rosin flux*, and this is the *only* type that should be used for general soldering work.

In some cases, such as machine soldering, dip soldering, or dip tinning, it may be desirable to apply the flux separately from the solder. In such cases fluxless solder should be used. If solder with core flux is mixed with separately fluxed solder, the flux should be obtained from the same manufacturer, with a guarantee of compatibility to prevent the formation of corrosive residues.

Rosin flux itself is nonconductive, but it tends to collect dust and other foreign material which could accumulate moisture and lead to corrosion and current-leakage paths. Alcohol or a special rosin solvent should be brushed on the joint and wiped off to remove excess flux after soldering.

Soldering Pencils: A joint temperature of 275°C (527°F) for a period of 1 to 2 seconds has been found to produce the most reliable solder joints. For continuous production work a small ruggedly constructed soldering pencil rated at 25 to 30 W and operated from a small variable transformer to control temperature is recommended. Where soldering is done intermittently, as in servicing or prototyping, or where it is not possible to use a variable transformer to reduce tip temperature, a temperature-regulated pencil is recommended. They cost about five times the price of an unregulated pencil, but the ease of producing quality work with them makes the investment well worthwhile.

Unregulated pencils reach excessive temperatures while standing in their cradles, which restrict air movement and reflect heat back to the tip. This makes it likely that the solder will be oxidized and the rosin burned to a black crud which fouls rather than cleans the joint. A microswitch mounted on the soldering pencil cradle can be wired to cut the power to half while the pencil is in the cradle, thus providing some relief for this problem. Figure 24-3 shows the circuit.

A short narrow barrel allows the pencil tip to be guided more surely into dense circuitry without accidental damage to nearby wiring and components. A three-wire grounded-barrel design is recommended for work on circuits containing MOS devices. Soldering guns are not recommended for electronic wiring or component-lead soldering, although they are handy for an occasional job involving a bracket or heavy electrical wire.

Soldering-Pencil Tips: Thermostatically controlled pencils can use iron-plated tips which maintain a smooth surface for a very long time if well cared for. A standard soldering pencil will burn the plating off in a short time, however, so plain copper tips are recommended here.

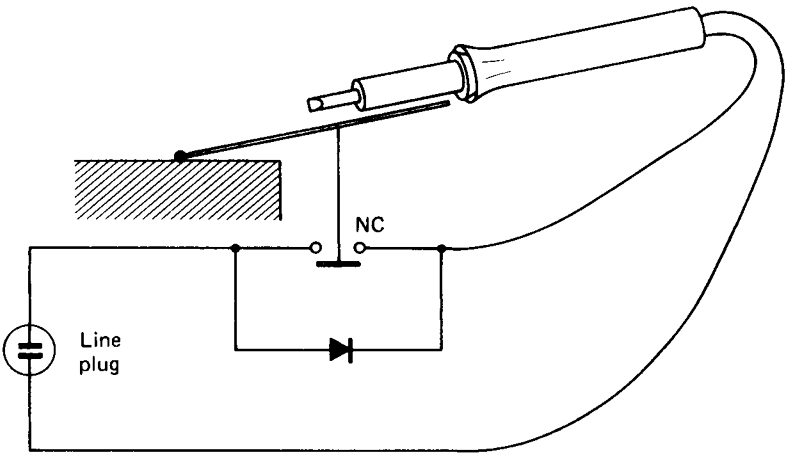


FIGURE 24-3 Microswitch opens reducing power to one-half when soldering pencil is in its cradle.

Copper actually dissolves slowly in molten solder, so it is very important to keep a sponge handy to wipe off the excess solder after each joint is completed. When a copper tip becomes pitted or misshapen, it should be filed back to its original wedge shape. Plated tips should not be filed, of course, but they may be cleaned with a soft wire brush.

A new or freshly cleaned tip should immediately be *tinned* or coated with solder. Take a few inches of solder and wrap it around the cold tip, and then heat up the pencil. Once the solder melts, see that the entire tip is covered and then wipe off the excess solder.

Tips often screw into the barrel, and a loose tip or corrosion between the screw threads can keep heat from flowing to the tip. A tip frozen into the barrel by corrosion can destroy the whole pencil. Antiseize compound applied when the tip is first inserted is the cure for all of these ills.

24.6 HOW TO SOLDER

A good solder joint is made by heating the wire and terminal sufficiently so they will melt solder, applying the solder *to the terminal* (not the pencil tip), and removing the heat quickly.

Heat Application: The soldering pencil should be kept warmed up at the bench all day, not plugged in just before it is needed. Be sure that the tip surface is smooth, shiny, and clean of old solder and flux before starting. Do not attempt to solder with a black, corroded tip.

Press the tip firmly to the terminal and against the wire to be soldered. A light touch will not transfer enough heat. Try to get as much contact between the work and the tip as possible. If the shape of the terminal limits the contact area, you may apply a tiny bit of solder between the tip and the work to serve as a heat-transfer bridge.

Solder Application: Quickly touch the end of the solder to the *opposite* side of the terminal so that the solder does not touch the tip of the pencil. As soon as the terminal becomes hot enough, the solder will melt and immediately flow all around the joint, forming a shiny fillet between the wire and the terminal.

Do not feed in more solder if the entire joint has been wetted. You want to fillet in the space between the wire and terminal—not bury the wire under solder.

Cooling the Joint: Immediately upon seeing the solder flow into a fillet around the terminal, remove the pencil tip. Unless the terminal is unusually massive, it should begin to melt solder within one to three seconds of the application of heat. Solder should flow completely around the terminal within one second of its melting, so the total heat-application time should not exceed 2 to 4 seconds. It is important not to leave the pencil tip on the work longer than necessary. The adhesive that bonds the copper foil to a circuit board is easily destroyed by overheating. Wire insulation may melt and solder may oxidize if excessive heat is applied.

It is essential that the joint be held motionless as it cools, or a cracked joint will result. This is the main reason for clinching wires mechanically tight before soldering. Blowing gently on the joint will greatly accelerate cooling.

A well-made solder joint will appear smooth, shiny, and concave between the wire and the terminal. Convex bulges or beads of solder indicate that the joint was not heated sufficiently or that too much solder was applied. Granular or jagged solder is evidence of excessive heating. Figure 24-4 illustrates soldering technique and joint problems.

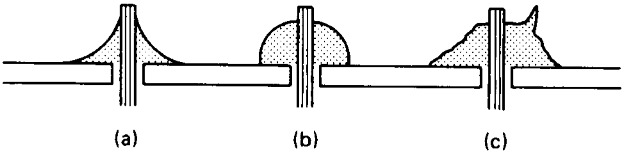


FIGURE 24-4 (a) A good solder joint is concave, smooth, and shiny. (b) Convex bulges indicate too little heat or too much solder. (c) Dull, granular-looking solder which forms peaks has been overheated.

A Rosin Joint sometimes results when the terminal surface refuses to bond with the solder, either because it was insufficiently heated or because the flux was unable to remove its oxide coating. The rosin (an insulator) may then act as a glue, holding the solder to the joint. The worst of it is that the wire will probably be in contact with the terminal, showing electrical connection on an ohmmeter. After a time in service, oxidation and/or vibration will almost certainly bring such a joint to an intermittent contact condition. A suspected rosin joint can be tested by attempting to wedge a knife or thin screwdriver blade between the solder and the terminal.

Rosin joints will usually crack under pressure, whereas the solder will shear before the solder-to-copper bond gives way in a good joint. When probing for rosin joints, be certain not to flex or strain the wires at the joint.

24.7 SOLDERING PROBLEMS

Difficulty in obtaining good solder joints may be caused by insufficient heating of the terminal, overheating of the solder or wire insulation, or inability to remove the metal-oxide coatings. Here are some solutions to these problems.

Cleaning: Wire ends, component leads, and terminal surfaces that are contaminated with a foreign substance (oil is a common one) or heavily corroded from long storage must be cleaned before soldering is attempted. A wire brush, a fine file, emery paper, or a copper kitchen scrub pad can be used. A folded piece of braid from coax cable shield wire makes a fine cleaning pad. Ultrasonic cleaning and chemical solvents may be investigated for special contaminants.

Tinning: Joints that take solder with difficulty may be worked more easily if the wire and the terminal are tinned (coated with solder) separately before the joint is made. Soldering is then simply a matter of melting the two solder coatings, rather than removing the oxide layers.

Abrasion: Mechanical scratching under the molten solder will sometimes break the oxide layer on a difficult soldering job. The pencil tip or a small knife or screwdriver blade may be used for this. It is likely that the solder will be overheated during this process, in which case it should be removed and new solder applied.

Antiwicking: When stranded wire is tinned or soldered, capillary action may draw liquid solder up between the strands a millimeter or two under the insulation. The wire becomes rigid where solder bonds the strands together, so flexing occurs under the insulation at the end of the solder penetration. The wire may actually break at this point, and inspection will be unable to reveal the flaw.

To prevent this problem, an antiwicking tool (Fig. 24-5) is used. This is a tweezerlike tool with a split-cup end for grasping the wire just below the insulation with the end protruding. Heat flowing up the wire is shunted to the tool and the wire under the insulation remains too cool to melt solder.

Heat Sinking: A variety of locking and spring-loaded clips are available to clamp onto a wire or component lead just above the joint during soldering. They may be used when there is danger of heat traveling up the lead and damaging a component or melting wire insulation.

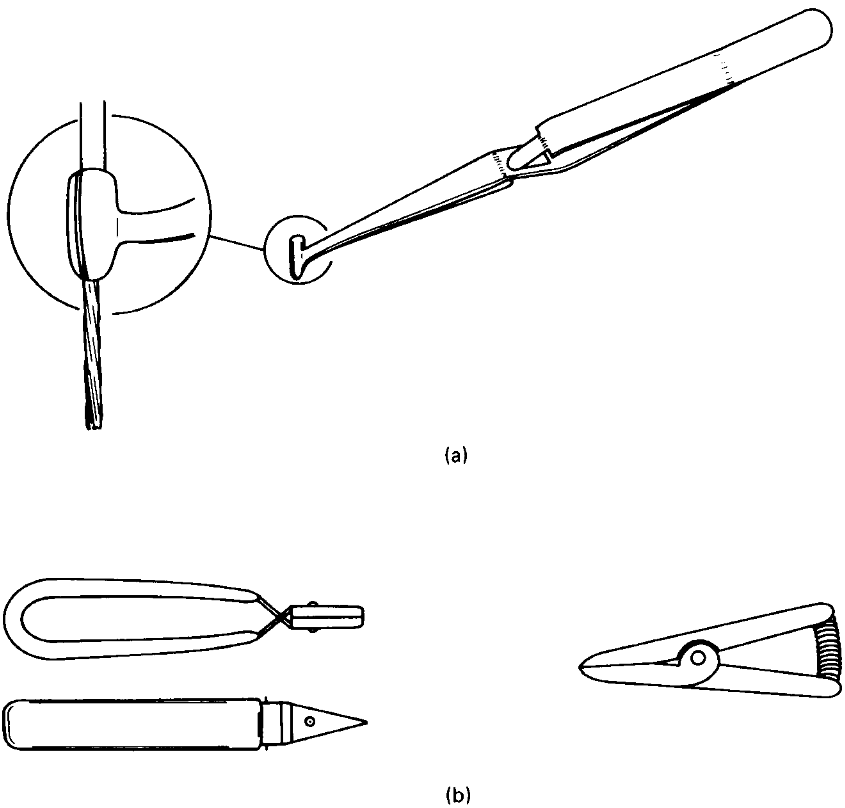


FIGURE 24-5 (a) Antiwicking tool to prevent stiffening of stranded wire under the insulation. (b) Heat-sink clips to prevent wire insulation and component damage during soldering.

24.8 UNSOLDERING AND REWORKING

A joint will have to be unsoldered if the solder has oxidized from overheating, or if a modification is being made, or if a faulty component is being replaced.

Applying Heat: Press the pencil tip firmly to the solder to melt it as quickly as possible. Heat-sink clips should be used on wires whose insulation may be damaged.

The copper tracks will come loose if a printed-circuit board is overheated for more than a second or two. The only defense against this is to get in and get out fast. Be sure that you apply no force to pull a PC pad up with the pencil tip or the component lead while the joint is hot.

Removing Solder: Many types of "solder sucker" devices are available, ranging from simple squeeze bulbs to power-operated vacuums. The more expensive ones work better, but all work reasonably well. Once the solder is melted, the solder sucker is used to vacuum up the liquid solder. Timing is important here: all the solder must be liquefied first because the air flow from the solder sucker will cool the joint. We want to avoid heating the joint a second time if possible.

PC-Component Removal: Likely enough, a small bit of solder will keep a lead or two stuck to the joint, or a clinch in the lead will make it difficult to pull through from the component side of the board. Do not force the lead out by pulling harder. Do not run around to all of the component leads trying to get them all molten at once. These tactics will pull up the foil pads. Unless the component absolutely must be saved, clip the leads on the component side of the board and pull them through *gently* from the foil side, one by one (Fig. 24-6). If you must save the component, use the solder sucker again if necessary, let the pad cool, wedge a knife point under the clinched lead and bend it up, and then remove the component.

Solder wick, which is nothing more than tinned shield braid, can be used to "sponge" up solder from a PC pad as an alternative to the solder sucker. Specially shaped soldering-iron tips are available to melt all the joints on a 16-pin DIP IC pack at once, but it is often difficult to melt all the joints without overheating one of them, and the old solder still must be removed so that the new component can

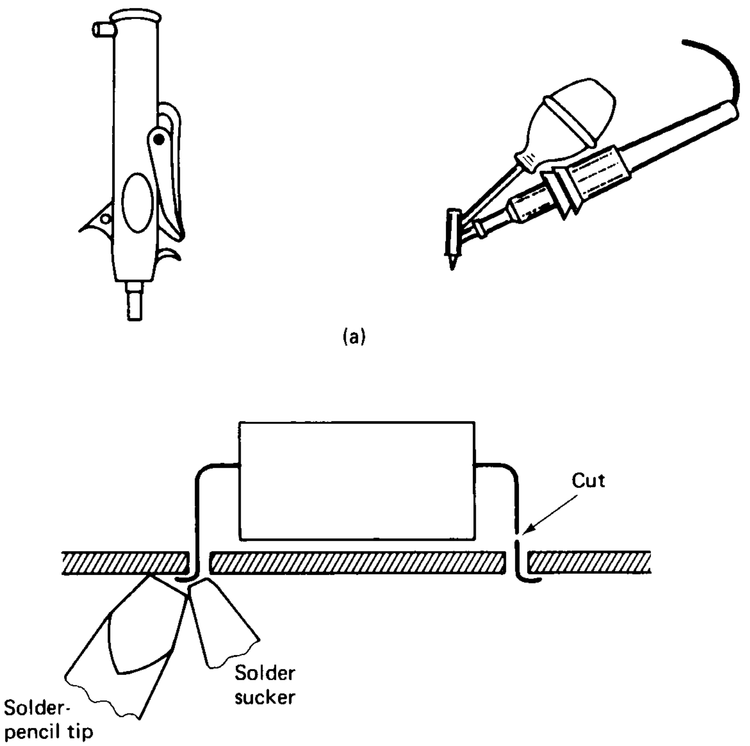


***FIGURE 24-6** (a) Two types of manual solder suckers. (b) Removing a clinched lead from the foil side.*

be inserted. Desoldering tips should not be used where the IC leads have been clinched, as the extraction force required could damage the PC pads.

Wire and Terminal Desoldering: When four wires are wrapped around a terminal it may be impossible to remove one of them without damaging all of them. Now the wisdom of leaving a service loop becomes apparent, for it is often best to simply cut all four wires as close to the end as possible, clean the terminal with the solder sucker, cutters, and long-nose, and reterminate the wires.

24.9 SOLDERING ALTERNATIVES

Resistance Soldering: If a high current is passed through a small terminal, it will heat up from within and melt solder. There can be no possibility of the terminal being too cold to take solder with this method. A step-down transformer is used to provide the high current, which is applied through needle-point or carbon-tip probes. The probes must be held firmly to the terminal before the transformer primary switch (usually a foot switch) is closed and may not be removed until the primary switch is opened. Breaking contact causes arcing and pitting of the terminal.

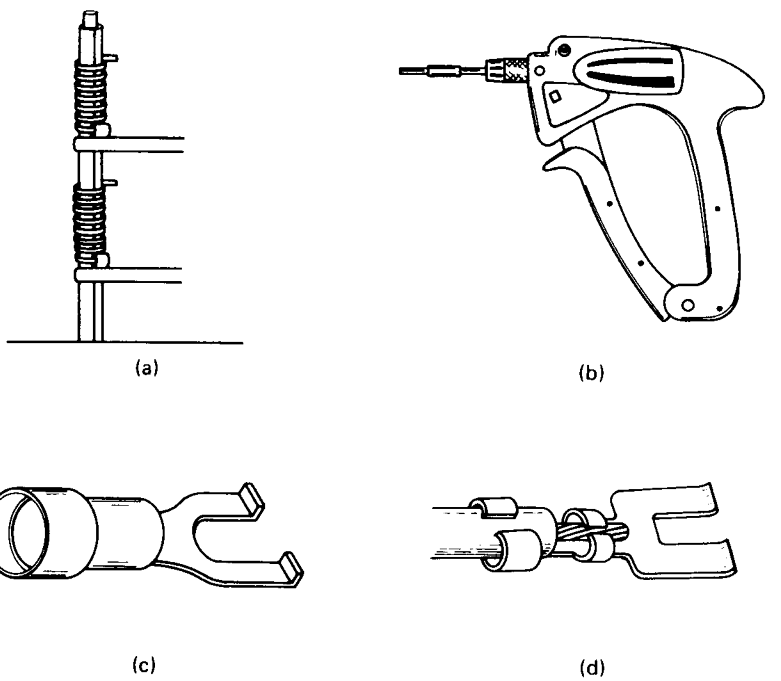


FIGURE 24-7 (a) Wire-wrap terminal. (b) Wire-wrap tool. (c) Solderless crimp terminal. (d) Solder terminal with insulation grip.

Wire Wrap has become very popular for prototyping, one-of-a-kind instrument production, and printed-circuit-card socket interconnections. It is a solderless method for connecting solid-wire ends to special wire-wrap terminals, which are stiff square posts about 2 cm long with sharp corners. The wire is stripped several centimeters back and wrapped around the post six or more turns with a special wire-wrap tool. The pressure at the corners is sufficient to maintain reliable electrical contact in spite of vibration, moisture, and oxidation. As many as three wire terminations can be made to one post, and posts are commonly spaced on 0.1-inch centers. Figure 24-7(a) shows a wire-wrap connection.

Terminals: Solder-type and solderless crimp-type terminals are shown in Fig. 24-7. They are suitable for either solid or stranded wire. Many terminals have a second wrap-around tab which is meant to be crimped lightly around the insulation to prevent flexing of the wire at the point of the stripping cut. With solder-type terminals the insulation should be held away from the terminal during soldering to avoid melting it.

25

TIPS FOR TECHNICIANS

This chapter contains a random collection of hints and shortcuts which the author has acquired over the years.

Checking Power-Supply Current: The current drain on a power supply can be determined without breaking any wires by observing the ripple across the first filter capacitor on a calibrated oscilloscope and calculating:

$$I_L = \frac{CV}{t}$$

where C is the value of the first filter in farads, V is the peak-to-peak ripple voltage, and t is the time between charging pulses in seconds. The accuracy of the determination is limited by the capacitor tolerance, which may be as poor as -20 $+80\%$ for electrolytics.

Is the Zener Conducting? A conducting avalanche diode (8 V and up) will show random noise of about 3 to 10 mV p-p across its terminals. Use a sensitive scope on ac coupling. Absence of noise usually means absence of avalanche current. True zener diodes (6 V and below) show little noise even when conducting.

Sharpening a PC Drill: A dull drill burns the circuit-board material and leaves ridges around the hole which interfere with soldering. Conventional sharpening of such small drills is impractical, but two cuts from an *old* pair of cutting pliers will leave a good cutting edge. Cant the cutters about 20° from a right angle and cut off about 1 mm of the end, then rotate the bit 180° and even up the other side. *Caution:* The end pieces will fly! Wear safety glasses and point the end toward the floor.

Is It Line Noise? If a noise signal or false trigger is encountered when troubleshooting with the scope, switch the trigger source to *line*. If the display holds still, the noise is ac-line-related.

Meter Loading? If a VOM is loading the circuit under test, the voltage reading will change with changing scales (2.0 V on the 2.5-V scale and 3.0 V on the 10-V scale, for example). Of course, if a second meter is available, the loading of the first can be observed on the second as the first is connected and removed.

Tape should never be regarded as a permanent fixture, as it will eventually unwrap itself. A taped joint can be made to hold longer by applying a "flag" over the wraps as shown in Fig. 25-1. The mating adhesive surfaces hold better than the wrapping.

Printed-Circuit Tracing: When troubleshooting a printed-circuit board, shine a light on the bottom (foil) side of the board and view the board from the top (component) side where probe clips can be attached. The light will let you see the conductor tracks through the board.

Measuring Overrange Resistors: Many DVM multimeters have an upper limit of 20 MΩ. Higher values can be measured by paralleling them with a previously measured high-value resistor and measuring the total:

$$\frac{1}{R_2}=\frac{1}{R_T}-\frac{1}{R_1}$$

For example, we measure a nominal 15-MΩ resistor (R_1) as 14.83 MΩ, and parallel it with a nominal 22 MΩ unknown (R_2), reading a total of 9.16 MΩ.

$$\frac{1}{R_2}=\frac{1}{9.16\ \text{M}\Omega}-\frac{1}{14.83\ \text{M}\Omega}$$

$$R_2 = 23.96\ \text{M}\Omega$$

Safety: Keep a $\frac{1}{2}$-m lead with two insulated-handle needle-point probes handy for discharging filter capacitors before starting service work.

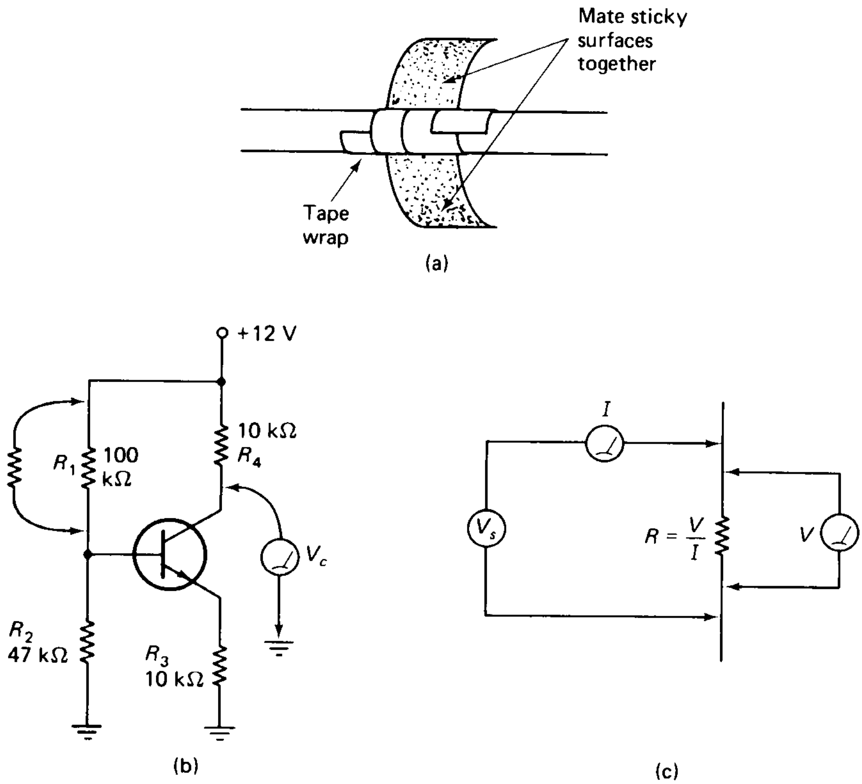


FIGURE 25-1 (a) Tape flag keeps wrapped tape in place longer. (b) Shunting a resistor to observe the circuit response. (c) Separate contacts for measuring low-value resistance: one set applies current, the other measures IR drop.

High Voltage: If you get involved in troubleshooting or prototyping the logic or low-level drivers of a system, disable the high-voltage supplies by disconnecting a transformer wire or shorting the oscillator transistor of an inverter base-to-emitter. This is not only for your personal safety—an accidental slip of the probe shorting the +2000 V to the +5-V supply can do untold amounts of damage.

Self-Oscillation at high frequencies (~100 MHz) will often not show up on the scope, but if moving your hand near the wiring or touching the chassis causes changes in the instrument's operation, self-oscillation is likely to be the problem. You can often tell which stage is self-oscillating by bringing a screwdriver blade or grounded probe near each transistor case in turn. A different transistor may help, or the leads might be dressed closer to the chassis or farther apart from each other.

Changing Resistor Values can be an effective troubleshooting device. In Fig. 25-1(b) for example, shunting a few hundred kΩ across R_1 should lower V_C. Shunting R_2 should raise V_C. Resistors in the 500-kΩ to 10-kΩ range can be lowered by simply touching the fingers of one hand across them. Obviously, this trick is for *low-voltage circuits only*.

If a Failure Has Just Occurred, quickly touch all the transistors and ICs in the suspected area. The bad one will often have gotten very hot just before failure. Watch out for burned fingers and high voltages on metal transistor cases.

Metal Transistor Cases, by the way, are usually electrically connected to the collector, and furnish an easy-to-get-at test point in troubleshooting.

Very Low Value Resistors may be measured with the test setup of Fig. 25-1(c). The resistance of the probe wires will not affect the determination because the voltage across the current-carrying ammeter wires is not measured by the voltmeter, and the voltmeter wires carry almost no current.

Remote-sensing power supplies use the foregoing trick to regulate the voltage *out at the load* rather than at the supply terminals.

A Soldering Gun that won't heat up probably just needs the tip bolts tightened.

Cheap Panel Meters that are off zero and have no zero adjust can often be rezeroed by holding a solder gun near and turning it on and off. The magnetic field from the coil in the gun body does it.

Plastic-Face Panel Meters sometimes get "sticky" and inaccurate because of static charge on the face. Simply moistening the plastic will free them up.

A Sensitive Microammeter Pointer will bounce around when the meter case is rotated if the terminals are open-circuited but will hold quite steady when they are shorted. You can tell if the coil is burned out in this way.

Loudspeaker Cones can be repaired with a piece of tissue paper soaked in nail polish.

Diodes Rated Above 1 kV usually consist of several series silicon diodes in a single case. Three silicon drops is more than the 1.5 V battery in the VOM, so the diode will measure open both ways. Use a 10-V supply and 10-kΩ resistor in series with a 1-mA meter to test HV diodes.

A Short to Ground can sometimes be located with a scope or DVM that has 1-mV sensitivity, as shown in Fig. 25-2(a). A 1-mm-wide copper track has nearly 10 mΩ/cm resistance, so a short-circuit current of 100 mA will give an IR drop of $0.1\ \text{A} \times 0.01\ \Omega$ or 1 mV per cm along the shorted PC track.

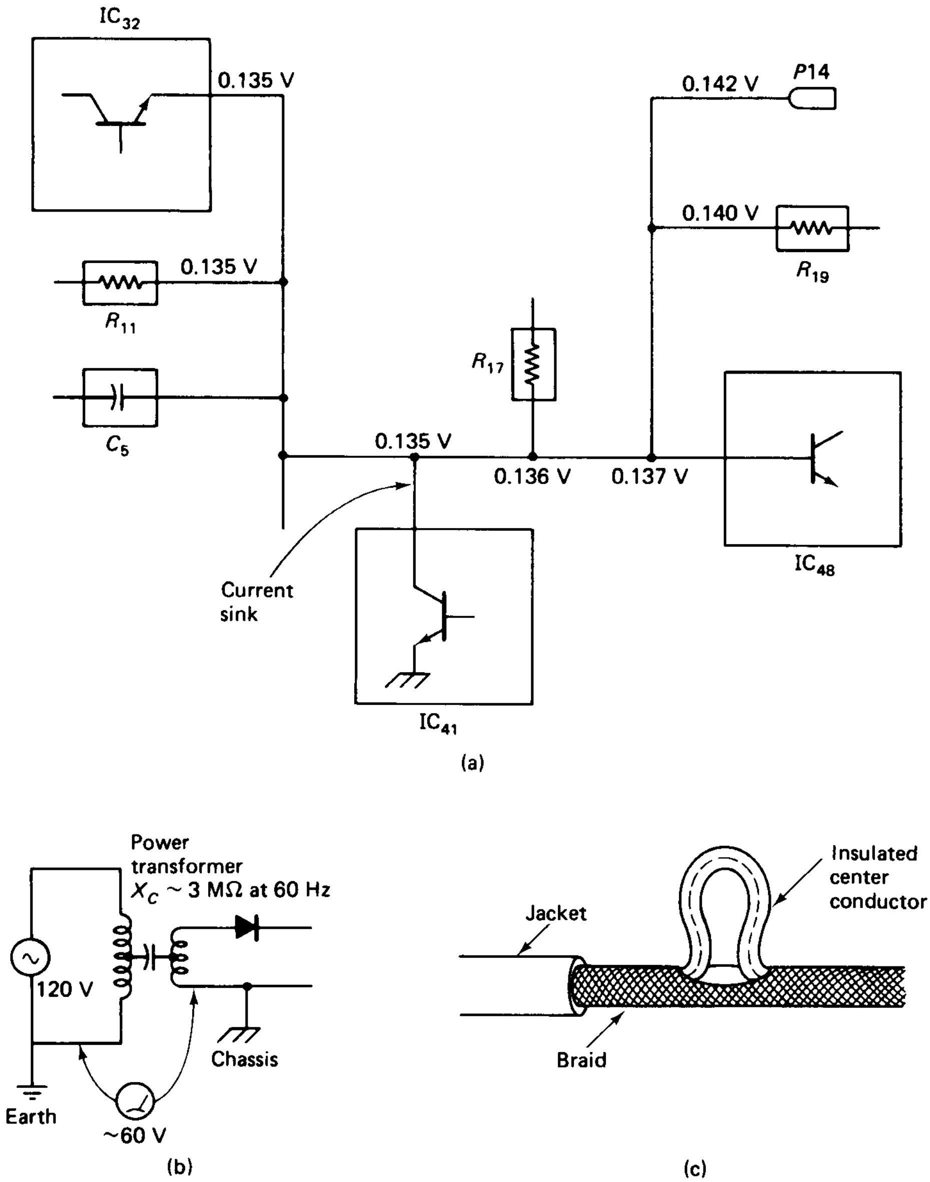


***FIGURE 25-2** (a) Tracing millivoltage to locate a short to ground. (b) Power transformer couples about 60 V ac through a high capacitive reactance to the chassis of an instrument with a two-wire line cord. (c) Pull the center conductor out through a part in the shield braid when stripping coax cable.*

Line Noise on the Chassis of an instrument with a two-wire line cord is coupled through stray capacitance of the power-transformer windings, as shown in Fig. 25-2(b). The instrument chassis will typically be at $\frac{1}{2} V_{\text{line}}$ above earth ground, unless externally grounded. The stray capacitance varies, being larger for large power transformers, but generally lies between 200 pF and 0.003 μF. This represents a reactance range from 13 MΩ to 880 kΩ and a current leakage of 5 to 68 μA. Connecting the instrument chassis to a point other than ground means injecting 5 to 68 μA of 60-Hz noise at the point.

Hot Chassis? Ungrounded instrument chassis will measure 60 V "hot" to ground as explained above, but a 33-kΩ resistor from the chassis to ground will cut the voltage to almost zero unless there really is a short to the ac line.

Stripping Coax: When stripping coax or shielded cable, loop the center wire out through a hole in the braid, as shown in Fig. 25-2(c).

Potentiometers tell a lot by the way they feel as they turn. A bumpy feel indicates a wire-wound type. A single bump indicates a broken wire or carbon track. Irregular bumps indicate a burned carbon track. If noise diminishes when you pull on the knob, wiper tension may be weak or the wiper arm may need cleaning.

Variable-Capacitor Plates are delicate and easily bent. If you listen carefully, you can hear the scraping of bent plates as the shaft is turned.

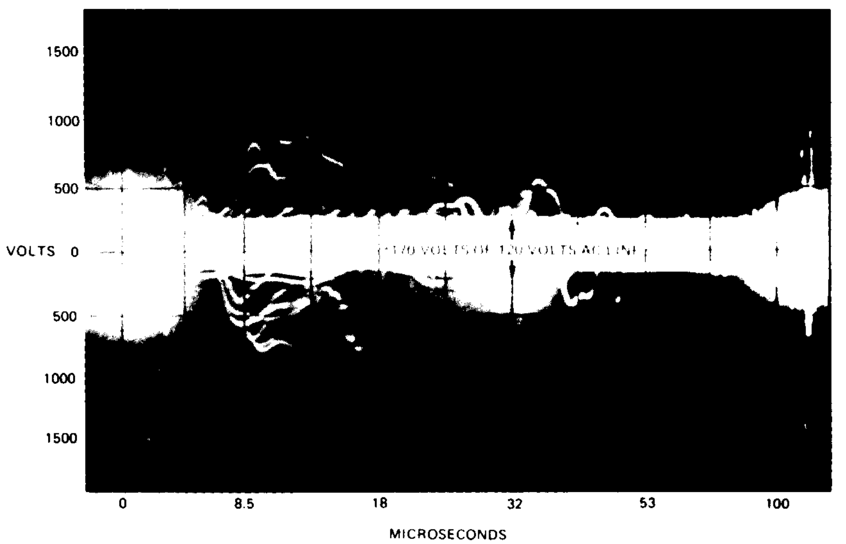


FIGURE 25-3 *Oscilloscope recording of a household power line input (24 h). Potentially damaging transients occur commonly on the ac power line.*

Use a Toothpick to clear PC-board holes of solder when replacing components. Heat the pad and leave the toothpick in until the solder cools.

Transients on the AC Line are documented by Fig. 25-3 to stress the need for generous safety factors on power-supply diodes and for varistor protection of line-operated equipment.

Don't Leave a VOM on the Ohms Range: It could discharge the battery and, over an extended period, could cause the battery to leak corrosive material all over the circuitry inside. Instruments with orange-glowing neon or incandescent-filament digital displays should not be left on indefinitely if maximize display tube life is to be realized.